

При поддержке ЮНЕСКО  
Международной программы  
по развитию коммуникации (МПРК)

ПРОЕКТ ПРЕДСТОЯЩЕЙ ПУБЛИКАЦИИ ЮНЕСКО  
при поддержке Глобальной инициативы  
МПРК по передовому опыту в области  
журналистского образования и Бюро  
ЮНЕСКО в Алматы в рамках проекта  
МПРК “Центральная Азия:  
журналистика данных”

# Курс по журналистике данных

**Gulmira Sultanbayeva  
Madina Mansurova  
Gulnur Tulepberdinova  
Zarina Buenbava  
Zanaatkar Seidekhanov  
Aytolkyn Ashimova**



При поддержке ЮНЕСКО  
Международной программы по  
развитию коммуникации (МПРК)

ПРОЕКТ ПРЕДСТОЯЩЕЙ ПУБЛИКАЦИИ  
ЮНЕСКО

при поддержке Глобальной  
инициативы МПРК по передовому  
опыту в области журналистского  
образования и Бюро ЮНЕСКО в  
Алматы в рамках проекта МПРК  
“Центральная Азия: журналистика  
данных”

# Курс по журналистике данных

**Gulmira Sultanbayeva  
Madina Mansurova  
Gulnur Tulepberdinova  
Zarina Buenbava  
Zanaatkar Seidekhanov  
Aytolkyn Ashimova**



## ASES KONGRE ORGANİZASYON VE YAYINCILIK

Kanununun hükümlerine göre kitabı yayınlanan yayınevinin izni olmaksızın elektronik, mekanik, fotokopi veya herhangi bir kayıt sistemi ile çoğaltılamaz, kısmen veya tamamen

Yayımlanamaz, depolanamaz.

**Basım Tarihi: 26.04.2025**

ISBN: 978-625-96918-4-8

Серия ЮНЕСКО по образованию журналистов  
2023 г.



**unesco**

МПРК - Международная  
программа развития  
коммуникации

Кафедра ЮНЕСКО по  
журналистике и  
коммуникации,  
Казахский  
национальный  
университет  
им.аль-Фараби,  
Алматы, Казахстан

## Курс по журналистике данных

При поддержке МПРК – Международной программы развития коммуникации

Baskı, Yayım ve Dağıtım

ASES KONGRE ORGANİZASYON YAYINCILIK

Hamidiye Mah. İnönü Cad. Helvacı İş Mrk. No: 43/19 Battalgazi/MALATYA

Tel: 0850 474 30 06 - [www.asescongress.com.tr](http://www.asescongress.com.tr) - e posta: [asescongress@gmail.com](mailto:asescongress@gmail.com)

Sertifika No: 63715

# Серия ЮНЕСКО по образованию журналистов

2023 г.

Опубликовано в 2023 г. Организацией Объединенных Наций по вопросам образования, науки и культуры. 7, Place de Fontenoy, 75352 Paris 07 SP, France и Бюро ЮНЕСКО в Алматы / Казахский национальный университет им. аль-Фараби, Алматы, Казахстан

© UNESCO, 2023

© ЮНЕСКО, Международная программа развития коммуникации (МПРК), КазНУ, 2021



Данная публикация предлагается в открытом доступе под лицензией Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Используя содержание данной публикации, пользователи соглашаются с правилами пользования Репозитория открытого доступа ЮНЕСКО ([www.unesco.org/open-access/terms-use-ccbysa-rus](http://www.unesco.org/open-access/terms-use-ccbysa-rus)).

Публикацию следует цитировать следующим образом: ЮНЕСКО, 2021. Курс по журналистике данных, Алматы.

Использованные названия и представление материалов в данной публикации не являются выражением со стороны ЮНЕСКО какого-либо мнения относительно правового статуса какой-либо страны, территории, города или района или их соответствующих органов управления, равно как и линий разграничения или границ.

Ответственность за взгляды и мнения, высказанные в данной публикации, несут авторы. Их точка зрения может не совпадать с официальной позицией ЮНЕСКО и не накладывает на Организацию никаких обязательств.

Графика: ЮНЕСКО Алматы

Оформление обложки: ЮНЕСКО Алматы

Макет: Казахский национальный университет им. аль-Фараби, Алматы, Казахстан

Отпечатано: Казахский национальный университет им. аль-Фараби, Алматы, Казахстан

Отпечатано в: Казахстане

## Авторы

Координатор проекта ЮНЕСКО/МПРК: Султанбаева Г.С., доктор политических наук, профессор

Гулмира Султанбаева, Казахстан (гл.2,9); Мадина Мансурова, Казахстан (гл. 4, 5, 6); Гульнур Тюлепбердинова, Казахстан (гл.4,5,6); Зарина Буенбава, Казахстан (гл.7); Санатбек Сейдеханов, Казахстан (гл.10); Айтолкын Ашимова, Казахстан (гл. 2,9)

Элира Турдубаева, Кыргызстан (гл.3,11); Аита Султаналиева, Кыргызстан (гл.8), Кымбат Болотбекова, Кыргызстан (гл.8)

Якутхон Маматова, Узбекистан (1 гл.); Фатима Муминова, Узбекистан (1 гл.)

Научный редактор: Ложникова О.П., старший преподаватель КазНУ им.аль-Фараби

## Благодарности

Авторы выражают благодарность *Василию Бурову, Ивану Бегтину («Инфокультура», Россия); Рахиму Ошакбаеву (журналист, Казахстан); Евгению Градову (Microsoft, Россия); Анне Красавиной (Южно-Уральский государственный университет, Россия)* за консультацию и обсуждению темы в рамках Центрально-Азиатских вебинаров по журналистике данных 5–26 мая 2021 года.

# Содержание

|                                                                                                       |    |
|-------------------------------------------------------------------------------------------------------|----|
| Авторы .....                                                                                          | 2  |
| Благодарности.....                                                                                    | 2  |
| Введение.....                                                                                         | 6  |
| О курсе .....                                                                                         | 7  |
| МОДУЛЬ 1: ОСНОВЫ ЖУРНАЛИСТИКИ ДАННЫХ.....                                                             | 8  |
| 1.2. Журналистика данных и Биг-дата-журналистика: общее и специфическое.....                          | 11 |
| 1.3. Становление и развитие дата-журналистики в мировых СМИ .....                                     | 13 |
| 1.4. Журналистика данных в Центральной Азии: позитивные практики, проблемы и тенденции развития ..... | 16 |
| 1.5. Ключевые инструменты журналистики данных.....                                                    | 34 |
| 1.6. Возможности искусственного интеллекта в ЖД.....                                                  | 38 |
| 1.7. Открытость как условие развития журналистики данных: правовой аспект .....                       | 40 |
| Модуль 2. ПРИНЦИПЫ ИССЛЕДОВАНИЯ И СТАТИСТИКА .....                                                    | 48 |
| 2.2. Количественные методы исследования в журналистике .....                                          | 49 |
| 2.3. Как кодировать данные качественных исследований.....                                             | 50 |
| 2.4. Основной алгоритм исследований в журналистике данных .....                                       | 51 |
| 2.5. Источники и тип данных .....                                                                     | 51 |
| 2.6. Основные принципы API.....                                                                       | 53 |
| 2.7. Скрейпинг .....                                                                                  | 53 |
| 2.8. Основные инструменты для обработки данных .....                                                  | 55 |
| 2.9. Статистика в журналистике.....                                                                   | 56 |
| 2.10. Гендерные аспекты журналистики дезагрегированных данных .....                                   | 62 |
| Модуль 3. СБОР ИНФОРМАЦИИ ИЗ ОТКРЫТЫХ ИСТОЧНИКОВ .....                                                | 66 |
| 3.1. Платформизация трансграничной журналистики расследований .....                                   | 67 |
| 3.2. Настройка оповещений о данных .....                                                              | 68 |
| 3.3. База данных оффшорных утечек .....                                                               | 71 |
| 3.4. Открытые данные в Кыргызстане .....                                                              | 72 |
| Модуль 4. ВВЕДЕНИЕ В ЭЛЕКТРОННЫЕ ТАБЛИЦЫ .....                                                        | 75 |
| 4.1. Сложности в развитии журналистики данных как отдельной дисциплины .....                          | 76 |
| 4.2. Перспективы журналистики данных.....                                                             | 76 |
| 4.3. Поиск и сбор информации. Теоретический аспект алгоритма поиска релевантной информации .....      | 77 |

|                                                                                                                              |     |
|------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.4. Подготовка материала для анализа данных .....                                                                           | 80  |
| Модуль 5. ЭЛЕКТРОННЫЕ ТАБЛИЦЫ: ОБРАБОТКА, АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ТАБЛИЧНЫХ ДАННЫХ .....                                       | 84  |
| 5.1. Анализ данных в Microsoft Excel .....                                                                                   | 84  |
| 5.2. Обзор программы Microsoft Excel .....                                                                                   | 84  |
| 5.3. Сортировка.....                                                                                                         | 86  |
| 5.4. Фильтрация .....                                                                                                        | 87  |
| 5.5. Функции .....                                                                                                           | 91  |
| 5.7. Визуализация данных в Excel .....                                                                                       | 96  |
| 5.8. Визуализация данных с помощью диаграмм.....                                                                             | 96  |
| 5.9. Итоговая визуализация статистических данных в Microsoft Excel .....                                                     | 102 |
| 5.10. Объем инновационной продукции.....                                                                                     | 104 |
| Модуль 6. ОБРАБОТКА И АНАЛИЗ ДАННЫХ ПРОГРАММНЫМИ МЕТОДАМИ .....                                                              | 110 |
| 6.1. Python и Jupyter Notebook как инструменты дата - журналиста .....                                                       | 113 |
| 6.3. План работы.....                                                                                                        | 114 |
| 6.4. Установка Pandas .....                                                                                                  | 115 |
| 6.5. Импорт файлов и просмотр содержимого .....                                                                              | 116 |
| 6.6. Основные команды для работы в Jupyter Notebook.....                                                                     | 120 |
| 6.7. Курс по программированию на Python.....                                                                                 | 122 |
| 6.8. Визуализация данных в Jupyter Notebook .....                                                                            | 122 |
| Модуль 7. ВИЗУАЛИЗАЦИЯ В ЖУРНАЛИСТИКЕ .....                                                                                  | 126 |
| 7.1. Создание графики с помощью веб-инструментов .....                                                                       | 126 |
| 7.2. Системы статистического программирования: R (r-project.org) или Pandas (pandas.pydata.org).....                         | 129 |
| 7.3. Географические информационные системы (ГИС): Quantum GIS, ArcGIS, GRASS .....                                           | 131 |
| 7.4. Работа инструментами более сложных графических объектов и карт .....                                                    | 139 |
| Модуль 8. РОЛЬ МЕДИЙНОЙ И ИНФОРМАЦИОННОЙ ГРАМОТНОСТИ (МИГ) В БОРЬБЕ С НЕКОРРЕКТНОЙ ИНФОРМАЦИЕЙ И ДЕЗИНФОРМАЦИЕЙ .....        | 143 |
| 8.1. Влияние некорректной информации и дезинформации на современное общество .....                                           | 143 |
| 8.2. Как страны Центральной Азии борются с распространением ложной и некорректной информации на законодательном уровне ..... | 145 |
| 8.3. Медийная и информационная грамотность как инструмент просвещения в эпоху постправды .....                               | 147 |
| 8.5. Виды и типы фейковой информации .....                                                                                   | 153 |

|                                                                                         |     |
|-----------------------------------------------------------------------------------------|-----|
| 8.6. Борьба с фейками: от медиаграмотности до лишения свободы .....                     | 155 |
| 8.7. Борьба с фейками в Кыргызстане .....                                               | 157 |
| 8.8. Проверка социальных сетей: оценка источников и визуального контента .....          | 159 |
| 8.9. Как искать информацию о пользователях в социальных сетях? .....                    | 161 |
| 8.10. Верификация фотографий .....                                                      | 163 |
| 9.1. SMART -Планирование материала .....                                                | 167 |
| 9.2. Визуализация .....                                                                 | 169 |
| 9.3. Виды инфографики .....                                                             | 171 |
| Модуль 10. СТОРИТЕЛЛИНГ КАК ТРЕНД В ЖУРНАЛИСТИКЕ ДАННЫХ .....                           | 177 |
| 10.1. Типы контента в новых медиа .....                                                 | 177 |
| 10.2. Мультимедийный лонгрид и его место в дата-сторителлинге .....                     | 178 |
| 10.3. Особенности дата-сторителлинга .....                                              | 179 |
| 10.4. Примеры лонгридов в журналистике данных .....                                     | 183 |
| Модуль 11. ЭТИЧЕСКИЕ АСПЕКТЫ ЖУРНАЛИСТИКИ ДАННЫХ .....                                  | 185 |
| 11.1. Специфика этики в журналистике данных .....                                       | 185 |
| 11.2. Этические коллизии в работе с «грязными» данными .....                            | 186 |
| 11.3. Доступ к данным в ракурсе этики .....                                             | 187 |
| 11.4. При представлении данных ключевым является контекст .....                         | 190 |
| 11.5. Этическая ответственность при работе с необработанными данными и их точность .... | 191 |
| ЗАКЛЮЧЕНИЕ .....                                                                        | 193 |
| РЕКОМЕНДАЦИИ: .....                                                                     | 195 |
| ГЛОССАРИЙ .....                                                                         | 196 |
| СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ: .....                                                 | 201 |
| ПОЛЕЗНЫЕ РЕСУРСЫ .....                                                                  | 206 |
| Приложение № 1 Международные источники данных .....                                     | 206 |
| Приложение № 2 Открытые данные - Центральная Азия .....                                 | 211 |
| Приложение № 3 Инструменты по журналистике данных .....                                 | 222 |
| Приложение № 4 Рекомендуемые курсы MOOK по дата журналистике? .....                     | 224 |

## Введение

Кафедра ЮНЕСКО по журналистике и коммуникации представила проект *Центральная Азия: интеграция учебных программ по журналистике данных в университетах* в Международную программу развития коммуникаций (МПРК). 64-я сессия заседания совета МПРК 11-12 июня 2020 г. одобрила проект на своей сессии в Париже.

Проект выбрал следующий тематический приоритет МПРК - *Учебные заведения в области журналистики уполномочены предлагать высококачественные программы подготовки журналистов с использованием инструментов и ресурсов, созданных в рамках Глобальной инициативы МПРК по совершенствованию журналистского образования*. Следующие ресурсы МПРК рекомендованы консорциуму ВУЗов журналистики, координируемому Казахским национальным университетом имени Аль-Фараби:

1. Модуль по журналистике данных из учебной программы для обучения журналистов<sup>1</sup> ЮНЕСКО, 2015.
2. Модуль: Сбор информации и визуализация из «МИГ в журналистике: руководство для преподавателей журналистики» (русский, узбекский)<sup>2</sup>, ЮНЕСКО, 2019.
3. Модуль: Борьба с некорректной информацией и дезинформацией с помощью медийной и информационной грамотности (МИГ), из Журналистика, «фейковые новости» и дезинформация, ЮНЕСКО, 2019, переводы на русский<sup>3</sup>, казахский, кыргызский, таджикский языки.
4. Модуль 6: Проверка социальных медиа: оценка источников и визуального контента<sup>4</sup>, предыдущей публикации, ЮНЕСКО, 2019.

---

<sup>1</sup> [https://unesdoc.unesco.org/ark:/48223/pf0000221199\\_rus/PDF/221199rus.pdf.multi.page=49](https://unesdoc.unesco.org/ark:/48223/pf0000221199_rus/PDF/221199rus.pdf.multi.page=49)

<sup>2</sup> <https://unesdoc.unesco.org/ark:/48223/pf0000374922/PDF/374922rus.pdf.multi.page=56>

<sup>3</sup> <https://unesdoc.unesco.org/ark:/48223/pf0000371785/PDF/371785rus.pdf.multi.page=95>

<sup>4</sup> <https://unesdoc.unesco.org/ark:/48223/pf0000371785/PDF/371785rus.pdf.multi.page=127>

## О курсе

Цель курса - интеграция и адаптация учебной программы ЮНЕСКО по журналистике данных в странах Центральной Азии.

Учебный курс планируется применять в университетах Центральной Азии при обучении журналистике данных и ее внедрения на национальном и региональном уровнях и адаптировать к типовым учебным программам ЮНЕСКО по журналистскому образованию.

Курс содержит сведения по основным этапам становления и развития журналистики данных в мире, а также в странах региона Центральной Азии с учетом объективных условий и специфики деятельности массмедиа. Помимо ключевых инструментов, студенты ознакомятся и с возможностями искусственного интеллекта в развитии журналистики данных. Разработки и рекомендации в представленном курсе нацелены на получение навыков и владение инструментами больших данных, сфокусированных на расследованиях журналистами социальных проблем данного региона в достижении целей устойчивого развития ООН (ЦУР): ЦУР 4 - Качественное образование, ЦУР 5 – Гендерное равенство, ЦУР 16 - Мир, правосудие и эффективные институты.

В курсе разрабатываются практические основы и навыки для работы с журналистикой данных, а также в сочетании со средствами программирования и табличной верстки, которые образуют основу для визуализации, анализа и переработки информации.

Задачами курса служат также раскрытие понятий и терминологии по журналистике данных, постановка задачи и алгоритма выполнения практических упражнений. В пособии рассматриваются поиск, обработка, сортировка данных, работа с данными в MS Excel, с последующим импортом в среду Jupyter Notebook для более продвинутого анализа и визуализации, экспорт визуализации данных, и подготовка материала к публикации.

Предлагаемый курс будет способствовать реализации долгосрочной стратегии Кластерного бюро ЮНЕСКО в Алматы по усилению устойчивого развития журналистского образования в фокусе дата-журналистики.

## МОДУЛЬ 1: ОСНОВЫ ЖУРНАЛИСТИКИ ДАННЫХ

### 1.1. Что такое журналистика данных

Журналистика данных (дата-журналистика, от англ. data – «данные») дословно означает журналистика, использующая статистику, цифры, данные. Данное направление – относительно молодое в журналистике. В силу этого эксперты и ученые ведут активные дискуссии по формулировке точного определения понятия «журналистика данных», а потому вопрос о системной теории дата-журналистики пока остаётся открытым. Тем не менее, эксперты и опытные практики сходятся в одном: дата-журналистика – это междисциплинарная сфера, в рамках которой производство журналистского контента осуществляется на основе анализа данных<sup>5</sup>. Другими словами, дата-журналисты призваны совмещать в своей работе журналистику и навыки работы с открытыми данными, а также с их обработкой – программированием, статистическим анализом и дизайном. Все это усложняет определение теории журналистики данных. Люди же, плохо знакомые с предметом журналистики данных и вовсе склонны думать, что она является надуманным, искусственным концептом. В действительности же журналистика данных существует уже много лет, равно как и контент, производимый дата-журналистами.

Журналистику данных (далее – ЖД) можно рассматривать «как высокоспециализированное подразделение журналистских расследований, но большинство её средств можно также использовать в повседневной журналистике»<sup>6</sup>. Это новое направление подразумевает «новые возможности, которые открываются, когда вы совмещаете традиционный “новостной текст” и способность рассказывать интересную историю с широким масштабом и диапазоном цифровой информации, доступной сейчас. И эти возможности могут использоваться на любом этапе: с помощью программирования для автоматизации процесса сбора и объединения информации из органов власти, правоохранительных структур, управлений статистики, других источников<sup>7</sup>. Деятельность в области журналистики данных предполагает знания и навыки в двух направлениях. «Первое – методология и статистика, поскольку используемые данные обычно являются результатом исследований. Выводы из кучи данных возможны только в том случае, если есть понимание того, что на самом деле измеряется, значений и переменных, уровней измерения и границ погрешностей. Это позволяет сформулировать гипотезу, которая затем тестируется с использованием набора данных. Второе направление – правильное определение интересных данных. Что представляет собой интересное место для поиска данных? Среди наиболее очевидных вариантов – базы данных правительства или организаций (в том числе международных), или веб-страницы этих организаций с данными. После того, как источник найден, следующий шаг – получение данных на ваш компьютер и логическая обработка ифр».

В Предисловии к «Пособию по журналистике данных» преподаватель Бирмингемского университета, известный журналист и блоггер П. Бредшоу очертил понятие ЖД следующим образом: «...это новые возможности, которые открываются, когда вы объединяете традиционное «чутье на новости», умение выведать всё, что случилось, и способность рассказать захватывающую и наглядную историю с огромным количеством разнообразной цифровой информации, которая

---

<sup>5</sup> См.: <https://journalismfund.wordpress.com>

<sup>6</sup> Питер Вервей. Журналистика данных // Модель учебной программы по журналистике: Сборник новых учеб-ных планов, ЮНЕСКО, 2015. – С.49.// [https://unesdoc.unesco.org/ark:/48223/pf0000221199\\_rus/](https://unesdoc.unesco.org/ark:/48223/pf0000221199_rus/)

<sup>7</sup> Питер Вервей. Журналистика данных // Модель учебной программы по журналистике: Сборник новых учеб-ных планов, ЮНЕСКО, 2015. – С.50.// [https://unesdoc.unesco.org/ark:/48223/pf0000221199\\_rus/](https://unesdoc.unesco.org/ark:/48223/pf0000221199_rus/)

теперь стала доступной»<sup>8</sup>. Он предложил модель перевернутой пирамиды для ЖД, которая хорошо объясняет, как работает этот процесс.

Журналистика данных, по мнению П. Брэдшоу, начинается одним из двух способов: либо у вас есть вопрос, для ответа на который нужны данные, либо набор данных, который ставит вопросы. В любом случае сопоставление данных – это то, что превращает процесс в ЖД.

Разобраться в открытых данных непросто. Например, Ева Константарас, журналист-расследователь в сфере журналистики данных, в своём учебном пособии объясняет: при первом просмотре набора данных сложно понять, на что именно обратить внимание<sup>9</sup>. Чтобы найти историю в цифрах, автор предлагает несколько методов для выявления сюжета:

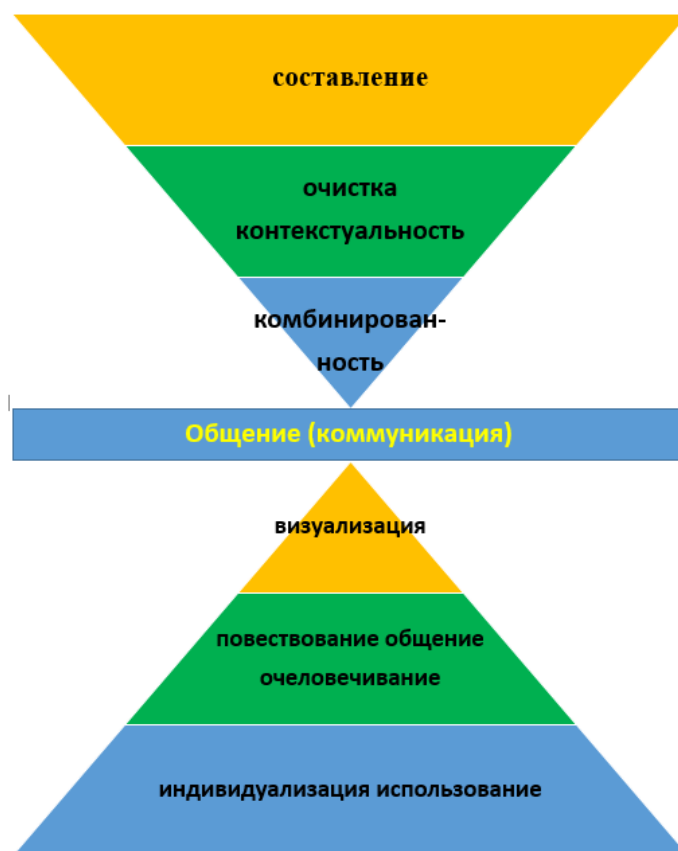


Рисунок № 1. Методы выявления сюжета

- разработать гипотезу, а потом с помощью данных опровергнуть или подтвердить её;
- выявить тренды и контрасты;
- найти слишком высокие или слишком низкие показатели, которые могут указывать на нарушения;

<sup>8</sup> Bradshaw Paul. Data Journalism Handbook.// [https://ria.ru/files/book/\\_site/index.html](https://ria.ru/files/book/_site/index.html)

<sup>9</sup> Ева Константарас. Размышляя как дата-журналист. Региональный центр ПРООН для Европы и стран СНГ учебное пособие по дата-журналистике. OdecaNet.org

- путём сопоставления баз данных найти невидимые связи между различными структурами. По этим методам работают и журналисты в Центральной Азии<sup>10</sup>.

### Базовые компетенции дата-журналиста

Участник и лектор Центральноазиатских вебинаров по журналистике данных, прошедших в мае 2021 года, В. Буров ключевые компетенции дата-журналиста предложил в виде следующей таблицы:



Рисунок № 2. Компетенции дата-журналиста  
(<http://ru.unesco.kz/central-asian-data-journalism-webinars>)

Как видно из таблицы, ЖД объединила в себе техническую составляющую, которая отражается в сборе и обработке больших объёмов данных, и собственно журналистскую составляющую, которая выражается в исследовании темы и представлении её аудитории в понятном для неё виде. Для журналиста необходимым становится умение не только обрабатывать данные, но и ориентироваться в них. Дата-журналистика выдвигает факт как определяющий элемент. Цифры, статистика, карты, диаграммы, таблицы являются не иллюстрирующим материалом, а информационным поводом.

Для работы в направлении ЖД журналист должен обладать аналитическим складом ума, быть уверенным пользователем технических инструментов, а также понимать возможности визуализации данных. Здесь под визуализацией подразумевается работа с текстом и таблицами в стандартных программах, а также использование специализированных сервисов, в том числе сервисов для создания мультимедийных презентаций и инфографики. Из-за того, что ЖД является междисциплинарным явлением, она сочетает в себе гуманитарные и технические знания, а дата-журналист должен обладать формально-логическим и ассоциативно-образным складом ума. В профессиональные компетенции дата-журналиста входят навыки не только анализа информации, но и планирования всего материала от замысла до публикации. По сути, профессиональные компетенции дата-журналиста проявляются уже на этапе составления плана работы, оценивать их содержание и трудоемкость выполнения.

Журналист, анализирующий огромные массивы информации, должен обладать критическим мышлением, которое подскажет ему, как трактовать, анализировать структурированную и неструктурированную информацию, синтезировать, обобщать, трактовать, визуализировать и презентовать ее с определенным эмоциональным зарядом. Он призван эффективно работать с

<sup>10</sup> Дата-журналистика в Центральной Азии: где копают глубже.// <https://newreporter.org/2019/04/18/data-zhurnalistika-v-centralnoj-azii-gde-kopayut-glubzhe/>

системами управления контентом (Content Management Strategy), управления базами данных (Data Management Strategy), управления планированием (Data Management Plan); владеть программным обеспечением и техническими средствами для регулярной коммуникации, мониторинга информации в интернете, в т.ч. в социальных сетях; уметь использовать принципы и технологии управления проектами, методы решения задач по созданию и редактированию контента.

Для дата-журналиста важны и такие компетенции, как владение разнообразными методами поиска, проверки и оценки достоверности информации; принципами сюжетного построения материалов в соответствии с законами определенного жанра; следование праву и принципам профессиональной этики, о чем подробно пойдет речь в отдельных разделах настоящего курса.

Компетенции по управлению базами данных предполагают не только компьютерную грамотность дата-журналиста, но и навыки работы с данными и их структурами, со статистикой, подготовкой данных, их очисткой и представлением, навыки программирования данных, юридические аспекты работы с данными и т.д.

В творческом ракурсе работу дата-журналиста часто уподобляют работе журналиста-расследователя. По мнению эксперта, дата-журналист, как и расследователь, «копается в разных базах данных — недвижимости, деклараций о доходах, владения оффшорными компаниями и т.п. Разница только в объеме данных. Дата-журналист работает не с индивидуальными, конкретными записями, а с большими базами данных»<sup>11</sup>. По его мнению, дата-журналист должен совмещать две важные функции - журналиста и аналитика - и обладать двумя наборами компетенций. Во-первых, владение «языком на уровне лучше, чем у студентов-технарей, и понимание принципов журналистики: как устроены медиа, их экономика, логика, формы существования, что такое спецпроекты, время жизни статьи, и многое другое. Это те вещи, которые большинство айтишников совсем не понимает. Во-вторых, необходимы знания о том, как устроены основные инструменты дата-аналитика (Jupyter Notebook, Python, другой инструментарий). Навыки работы с источниками данных, проектирования разного рода исследований, понимание, как собирать и обрабатывать данные, какие механизмы и подходы можно применить к очистке данных и оценке их достоверности. «Все это, на самом деле, не очень сложно для людей с техническим бэкграундом, но они, как правило, очень плохо пишут. Этому их надо целенаправленно учить. И наоборот — гуманитариям тяжело дается математика и компьютерные (не человеческие) языки. Это такая двойная профессия, в которой могут преуспеть гуманитарии с техническими навыками или технари с некоторым набором гуманитарных навыков»<sup>12</sup>.

## 1.2. Журналистика данных и Биг-дата-журналистика: общее и специфическое

Следует различать **большие данные** и **количественные данные** - гибридный тип данных, являющийся основой большинства публикаций журналистики данных в СМИ.

**Большие данные** - массивы данных огромных объемов и значительного многообразия, которые из-за их количества и сложности не могут быть собраны, сохранены и проанализированы без помощи специальных программных инструментов обработки. Чаще всего такие данные оформляются в базы размером от нескольких петабайт, генерируемые самыми разными источниками: социальными медиа, веб-сервисами, мобильными приложениями и т. д.

**Количественные данные** - наборы данных небольших объемов и разнообразия, которые могут быть обработаны человеком с использованием любого из традиционных статистических

---

<sup>11</sup> Дата-журналистика: «Получается, что журналист не нужен»// <https://polit.ru/article/2020/09/22/begtin/>

<sup>12</sup> Дата-журналистика: «Получается, что журналист не нужен»// <https://polit.ru/article/2020/09/22/begtin/>

инструментов. Это могут быть статистические данные, данные социологических исследований, рейтинги и др., которые являются привычными информационными поводами в журналистике.

В силу такого положения необходимо делать различать ЖД и журналистику больших данных (ЖБД). В основе материалов ЖД лежит обработка любых количественных данных, которые могут служить информационным поводом или быть частью материала, обогащая его. При этом особое значение приобретает визуальная презентация этих данных для аудитории. В основе материалов ЖБД лежит работа именно с большими данными, в соответствии с их пониманием современными специалистами как цифровой информации, которую невозможно обрабатывать традиционными способами.

Британский «Гардиан датаблог» (The Guardian datablog) является показательным примером современной европейской ЖД. Он был создан в 2009 году в качестве отдельного раздела издания «Гардиан» журналистом и впоследствии его редактором С. Роджерсом. Материалы, основанные на крупных цифровых данных, публикуются здесь регулярно. Помимо текстового контента читателям представляется возможность ознакомиться с проанализированными данными через размещение на сайте ссылок на базы данных, послужившие основой публикаций, через визуализацию данных различными способами, в том числе интерактивно.

В 2012 году эта газета запустила проект «Год в графике» (2012: The Year in Graphics), представив главные события года (выборы в США, Олимпиаду, разрушения от урагана и др.) в формате интерактивных визуализаций. А 2013 год «Нью-Йорк Таймс» объявляет годом интерактивного сторителлинга (2013: The Year in Interactive Storytelling), в рамках которого она запускает проект «Визуализация данных» (Data Visualization). Ежегодно журналисты «Нью-Йорк Таймс» создают интерактивные онлайн-проекты, основанные на больших данных в рамках серий публикаций «Год в интерактивном сторителлинге, графике и мультимедиа» (The Year in Interactive Storytelling, Graphics and Multimedia). С апреля 2014 года материалы ЖД размещаются в специализированном новостном разделе — «Апшот» (The Upshot), созданном лауреатом Пулитцеровской премии Д. Леонхардтом.

#### **Важность открытых данных для журналистики данных**

ЖД базируется на открытых данных, которые должны соответствовать следующим критериям:

- открытость доступа. Контент должен быть бесплатным и опубликованным в машиночитаемом виде. Эти данные должен распознавать и считывать не только человек, но и компьютер;
- свобода распространения и повторного использования. Это подразумевает использование свободных лицензий;
- запрет на ограничение использования.

Следовательно, можно сделать вывод, что если данные не соответствуют какому-либо из перечисленных пунктов, например, информацию нельзя повторно использовать, то эти данные мы не можем идентифицировать как открытые.

С развитием информационно-коммуникационных технологий правительство предоставляет социально значимую информацию через электронные платформы. Однако такие платформы не предоставляют достаточные условия для открытости. На самом деле, чтобы перейти от доступности информации к открытости, как утверждают некоторые авторы, информация должна быть понятной и полезной для граждан и других заинтересованных сторон<sup>13</sup>. Тем не менее, доступность информации является важным шагом в достижении действительно открытой государственной политики. Степень прозрачности напрямую связана с такими понятиями, как свобода информации, активное распространение информации, доступ к документам и удобство использования веб-

---

<sup>13</sup> Dawes, 2010; Zuiderwijk, Gasco, Parycek and Janssen, 2014)

сайта<sup>14</sup>. Более подотчетное правительство является более легитимным<sup>15</sup>. Органы государственной власти являются основными создателями и сборщиками данных во многих областях<sup>16</sup>. Эти сферы данных варьируются от трафика, погоды, географического положения, туристической информации, статистики, бюджетирования, уровня эффективности компаний и государственного сектора до всех типов данных о политике и инспекциях (продукты питания, безопасность, качество образования и т.д.). Руководители и другие государственные служащие часто избегают раскрытия контролируемых ими данных, поскольку это позволило бы общественности получить новые знания, которые, в свою очередь, могут вызвать критические вопросы. Это подтверждается и институциональной теорией о том, что открытие данных скорее укрепит существующие структуры, чем изменит их, и позволит им в полной мере использовать преимущества новых разработок<sup>17</sup>. Цифровые технологии становятся обязательной частью политических дискурсов, касающихся мира исследований<sup>18</sup>.

Наибольшую важность для журналистики данных представляют открытые государственные данные. Они позволяют журналистам не только сверять заявления депутатов и власть предержащих с реальными цифрами, но также становятся основой антикоррупционных расследований. Открытые данные помогают не только следить за деятельностью органов государственной власти и управления, но и работают на репутацию этих учреждений.

### **1.3. Становление и развитие дата-журналистики в мировых СМИ**

Журналистика данных на заре своего появления была тесно связана с «компьютерной журналистикой» (computer-assisted reporting-CAR или КЖ) в силу схожести целей. Первый пример «компьютерной журналистики» восходит к 1952 году, когда ставилась цель предсказать результаты выборов в США. С того времени прошло 70 лет. История журналистики данных длиной в 70 лет охватывает три больших периода.

**Первый** – 1952-1990-е годы характеризуется использованием в журналистской деятельности первых компьютеров, а также внедрением первых информационных технологий в сферу. Это дало начало новому этапу коммуникационного и коммуникативного развития мировой журналистики. В основе ее развития лежали диаметрально противоположные по существу концепции использования и представления информации: 1) опора на факты, статистику, исследования – и 2) опора на фактоиды, литературный вымысел, художественные образы, что формировало различия в методологии профессионального творчества. Отличительной особенностью выше отмеченной КЖ стало системное использование компьютера и цифровой информации при подготовке материалов. Функционально-инструментальный уровень использования технических средства для анализа данных позволял повысить качество контента, укрепить доказательную базу, и в первую очередь, в процессе проведения расследований. Однако непосредственно на характеристики самого журналистского сообщения это не повлияло кардинально.

В 1967 году журналист Филип Майер из газеты «Детройт фри пресс» использовал компьютер для систематизации полученных в ходе опроса жителей Детройта ответов для того, чтобы выяснить и объяснить причины серьезных беспорядков, которые разразились в городе тем летом. Проведя

---

<sup>14</sup> Curtin & Mendes, 2011; Reddick, Chatfield, & Puron- Cid, 2017)

<sup>15</sup> Bauhr & Grimes, 2012; Fierro & Gil-Garcia, 2011; Lourenço, 2015; Northrup & Thorson, 2003; Sandoval-Almazan, 2011)

<sup>16</sup> Henschen, Doug. Oracle Releases NoSQL Database, Advances Big Data Plans

<sup>17</sup> Fountain, 2001; Kraemer & King, 2006; West, 2004

<sup>18</sup> <https://www.cairn.info/enseigner-a-l-universite-avec-le-numerique--9782807313309-page-121.htm>

опрос бунтовщиков и составив сводную таблицу в электронном виде, Ф. Майер доказал, что среди вышедших на протест людей, окончивших колледж, было не меньше, чем жителей, не окончивших среднюю школу.

Ф. Майер продолжал использовать этот метод и в 1970-х годах, в частности, сотрудничая с репортёрами газеты «Филадельфия инквайер» Дональдом Барлеттом и Джеймсом Стилом, а также с Ричем Морином из газеты «Майями геральд» для анализа реестра оценки имущества. Позже он написал книгу под названием «Прецизионная журналистика», в которой призывал коллег к активному использованию анализа баз данных и методов социальных исследований в публикациях.

Позже Дональд Барлетт и Джеймс Стил, занимающиеся журналистскими расследованиями, стали активно использовать электронные таблицы для обобщения большого количества данных из отчётов и судебных решений. Благодаря такому методу работы они смогли сделать то, с чем другие никогда бы не справились из-за обилия документов и невозможности обработки всей информации вручную. Сами журналисты, вспоминая своё первое задание с «большими данными» указывали, что редактор газеты попросил их проанализировать и оценить работу Федерального управления жилищного строительства на предмет незаконных операций при реализации программы капитального ремонта старых домов. В результате длительной работы журналисты обнаружили, что у Министерства жилищного строительства и городского развития возникли проблемы с выкупом закладных. Малообеспеченные семьи отказывались покупать дома из-за некачественного ремонта, который никак не отразился на уровне «аварийности» домов. В течение двух месяцев Д. Барлетт и Дж. Стил изучали всевозможные жилищные акты и закладные, проводили интервью с жильцами, выясняли, как работают спекулянты, и кто на самом деле покупает эти дома. В итоге была опубликована серия статей, из-за которой агентства по недвижимости отозвали свою рекламу и прекратили сотрудничество с газетой.

В 1974 году Д. Барлетт и Дж. Стил обратили внимание на «налоговую» проблему в Америке. Проанализировав годовые отчёты Налоговой службы более чем за десятилетие, они обнаружили, что претензии по уплате налогов к крупным конгломератам значительно меньше их реальной задолженности и смогли доказать неравномерность уплаты налогов в американском обществе. Согласно их расследованию, средние и низшие сословия американцев исправно отдавали положенные им налоговые выплаты, в то время как представители богатых слоёв общества платили гораздо меньше того, что на самом деле были должны.

При анализе «больших данных» они применили хронологический принцип: при сопоставлении информации, полученной из различных источников, в хронологическом порядке они получили возможность увидеть полную картину. В итоге за серию расследований «Аудиторская налоговая служба» в 1975 году расследователи получили первую Пулитцеровскую премию. Именно тогда технологию работы с «большими данными» Дональда Барлетта и Джеймса Стила стали постепенно применять другие репортёры. За все время совместной работы журналисты провели более десяти серьёзных крупных расследований, касающихся бизнес- и политических структур, налоговых реформ, фармацевтического бизнеса, злоупотреблений властью со стороны чиновников.

Активное использование такого метода работы стало очевидным, когда в 1990 году профессор Университета штата Индиана Джеймс Браун в Индианаполисе провёл первую конференцию по использованию баз данных в работе журналиста.

Таким образом, первый период развития ЖД характеризуется внедрением в деятельность журналистов нового метода анализа и обработки большого количества информации. Он позволил журналистам не только существенно снижать временные затраты на анализ информации, но и выявлять скрытые ранее факты, подтверждающие совершение преступления.

**Второй период** приходится на начало 1990-х – 2010-е годы. Он характеризуется использованием в качестве основного источника информации электронные базы данных. Журналисты по-прежнему анализировали большое количество документов в печатном виде,

однако с активным развитием Интернета и переводом информации в цифровую форму данных стало в тысячи раз больше и появились специальные базы данных по различным областям, ставшие главным источником информации. Так, практика появления ЖД потребовала от журналистов определённого уровня подготовки, поскольку освоить схему работы самостоятельно было крайне трудно.

В связи с этим в США стали появляться центры по обучению журналистов работе с большими массивами данных в электронном виде. Одним из первых таких центров был Национальный институт компьютерной помощи (НИКП), созданный в 1994 году Дженнифер Лафлер и Брантом Хьюстоном. К 1996 году успехи американских журналистов заставили иностранных репортёров посещать интенсивные семинары в НИКП. В институт стали съезжаться журналисты из Финляндии, Швеции, Новой Зеландии, Венесуэлы, Аргентины, Нидерландов, Норвегии, Бразилии, Мексики, России, Боснии и Герцеговины и Канады, которые смогли увидеть и освоить то, как использовать данные международных или американских ресурсов, а также как получить данные и проанализировать их касательно своей страны.

Между тем, в Лондоне в 1997 году профессор журналистики Мильвертон Уоллас начал проводить ежегодную конференцию под названием «NetМедиа», на которой участникам предлагались занятия по работе с компьютерными отчётами под руководством специалистов из НИКП и датских журналистов.

«С начала XXI века использование информации, основанной на работе с электронными массивами данных, расцвело в основном благодаря семинарам, проводимым в Миссури и во всём мире Содружеством репортёров и редакторов (СРР) и НИКП. В первое десятилетие XXI века Глобальная сеть журналистов-расследователей стала играть решающую роль в этом движении, начиная с первой конференции в 2001 году в Копенгагене, в ходе которой был подготовлен серьёзный доклад о подготовке и практическом обучении журналистов с помощью компьютера.

“В 2001–2002 годы формируется понятие «Big Data», завязанное на увеличении объёмов, типов и скорости поступления новых данных, о чём впервые говорит исследовательская компания «Gartner» в 2001 году. Взрывной рост количества и качества различных видов доступной информации через Google и социальные сети стал результатом «оцифровки информации, которая раньше была спрятана в шкафах и папках, другие получены в результате увеличения ресурсов, выделяемых на маркетинговые исследования»<sup>19</sup>. «Социальные данные – самое ценное сырьё XXI века, новая нефть... К счастью, информация как ресурс радикальным образом отличается от нефти. Запасы нефти на планете конечны, и по мере истощения этого ресурса его эксплуатация обходится всё дороже. Количество же информации, напротив, растёт в геометрической прогрессии, а стоимость технологий её передачи и обработки неуклонно снижается... использование информации практически не ограничено, учитывая её изобилие и цифровую форму»<sup>20</sup>. Это дало возможность американским журналистам проводить расследования на основе анализа баз данных о состоянии государственных школ, больниц, приютов для создания целостной картины, которая помогла бы аудитории сделать выводы о том, как на самом деле обстоит дело в той или иной сфере общественно-политической жизни.

**Третий период** развития ЖД приходится на 2010 годы, когда на международной конференции в Амстердаме, организованной Европейским центром журналистики Data-driven Journalism, она выделяется в самостоятельное направление<sup>21</sup>. С этого времени понятие «дата-

---

<sup>19</sup> Стивенс-Давидовиц С. Все лгут. Поисковики, Big Data и Интернет знают о вас все ; [пер. с англ. Л.И. Степановой]. М.: Эксмо, 2018. –С.27

<sup>20</sup> Вайгенд А. Big Data. Вся технология в одной книге; [пер. с англ. С. Богданова]. М.: Эксмо, 2018.- С.32-33

<sup>21</sup> Вайгенд А. Big Data. Вся технология в одной книге; [пер. с англ. С. Богданова]. М.: Эксмо, 2018.- С.32-33

журналистика» или ЖД повсеместно вводится в профессиональный оборот во всех регионах мира, в том числе и в странах Центральной Азии. Стремительное расширение интернет-пространства и прорывное развитие интернет-технологий дало возможность искушенной аудитории знакомиться с журналистскими публикациями, в которых активно используются большие объемы открытых данных.

В 2010–2011 годы в редакциях СМИ стали проводить так называемые «хакатоны» - многочасовые «мозговые штурмы» команды из журналистов, программистов и хакеров. Их взаимное общение и передача опыта позволили СМИ понять принципы работы в ЖД и начать организовывать собственные базы данных для удобства журналистов и читателей. «Хакатоны» проводили журналисты из Нидерландов, Аргентины, Финляндии. Главной идеей таких мероприятий стал то, что разработчики, которые знают, где искать информацию, не могут придать ей социальной значимости. Журналисты, в свою очередь, зная о возможностях ЖД, не могут представить их в привлекательной форме и не знают, как эти данные обнаружить.

#### **1.4. Журналистика данных в Центральной Азии: позитивные практики, проблемы и тенденции развития**

В 2015 году журналист и медиакритик из Таджикистана Лилия Гайсина сообщила, что три страны нашего региона – Казахстан, Кыргызстан и Таджикистан – оказались в рейтинге Глобального индекса открытых данных (Global Open Data Index). В списке, в который вошли ещё 122 страны, Кыргызстан показал лучший результат в регионе и занял 34 место, Казахстан – 50-е и Таджикистан – 83-е место<sup>22</sup>. Узбекистана в данном рейтинге не было. Однако в последние годы и в Узбекистане появились достойные примеры. С того времени в регионе Центральной Азии ЖД все увереннее «пробивает» себе дорогу. Несмотря на то, что очень «часто государственные и даже независимые источники отказываются предоставлять журналистам нужную информацию (как в Таджикистане), делают это не до конца или искажают данные (как в Казахстане и Кыргызстане), именно работа с открытыми цифрами» становится ныне главной платформой качественной журналистики.

Вот почему важно рассмотреть и то, как происходит становление и развитие ЖД как отдельного направления в государствах Центральной Азии. Тем более, что важным фактором для развития дата-журналистики в регионе, как полагают эксперты АНО «Информационная культура» (Россия), стала доступность множества сервисов для создания медиаматериалов, основанных на данных. С каждым годом появляются новые инструменты для решения разного типа задач (обработка данных, анализ данных, визуализация и т.д.) и разного уровня сложности (онлайн-сервисы, poscode-инструменты, готовые библиотеки для Python, JavaScript и других языков программирования)<sup>23</sup>.

---

<sup>22</sup> Дата-журналистика в Центральной Азии: где копают глубже.// <https://newreporter.org/2019/04/18/data-zhurnalistika-v-centralnoj-azii-gde-kopayut-glubzhe/>

<sup>23</sup> [20] Буров В., Орлова Кс. Дата-журналистика и открытые данные в странах Центральной Азии. Скрытые возможности симбиоза данных, журналистики и технологий искусственного интеллекта для развития регионального потенциала. - [https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism\\_web.pdf](https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism_web.pdf)

## Кыргызстан

По темпам развития ЖД в регионе лидирует Кыргызстан. Уже несколько лет в стране действует ежегодная стипендиальная программа организации Internews, которая предусматривает обучение журналистов и выпуск дата-материалов с помощью менторов. В 2020 году заработали первые дата-отделы в трех изданиях — 24.kg, kaktus.media и Kloop.kg. Их материалы стали упоминаться в рейтингах Глобальной сети журналистских расследований (GIJN) и принимать участие в международном кросс-паблишинге. Другими словами, материалы местных изданий все чаще стали появляться в зарубежных медиа. Так, онлайн-медиа Kloop.kg расширило свой дата-отдел, включив туда группу программистов и сделав фокус на использовании машинного обучения. В настоящее время данный ресурс прекратил свою работу. Таким образом, появился проект «Лукошко», который использует нейросети, чтобы понять, как часто кыргызстанские политики мелькают по ТВ, или чтобы провести расследование «Кольца коррупции», которое стало возможным благодаря объединению базы госзакупок и зарегистрированных в Минюсте юридических лиц.

В качестве успешного примера ЖД можно привести статью «Дата 4.0. Про цены в Кыргызстане, и как они влияют на бедность населения», подготовленную журналистами Kloop.kg совместно с azattyk.org (опубликована 25 марта 2021 года). Статья сопровождается таблицей, отражающей повышение цен на основные продукты питания в 2020–2021 году. На основе сравнительного анализа цен за два года авторы выявили, что многие товары становятся в прямом смысле роскошью и на них уходит значительная доля бюджета граждан<sup>24</sup>. Кроме того, журналисты создали калькулятор, который считает, сколько дней в месяц придется работать, чтобы купить мясо. В тексте указана ссылка на тест, который может пройти любой желающий. В данном тесте на русском и кыргызском языках задан вопрос: «Сколько дней в месяц вам придется работать, чтобы купить мясо?» (Эсептеңиз: Эт сатып алуу үчүн сиз бир айда канча күн иштешиңиз керек?)

Первые проекты по ЖД в стране были реализованы в 2015 году при поддержке фондов “Сорос-Кыргызстан” (ФСК), Интерньюс, ПРООН, АУЦА и другие. В 2016 году Promotank Research Institute при поддержке Программы «Развитие СМИ» ФСК реализовал проект, целью которого стало развитие ЖД через работу с заинтересованными медиа-редакциями, повышение потенциала журналистов, редакторов и технических специалистов. В 2017 году проект уже включил в себя аналитический обзор и углубленное содействие в подготовке материалов с использованием современных инструментов обработки, анализа и визуализации данных. В данном аналитическом обзоре было отмечено существенное различие между:

1) собственной обработкой имеющихся данных, что может включать в себя манипуляции с первичными данными (несложные расчеты для выведения новых значений, сравнительные расчеты, очистка данных), а также анализ новых полученных смыслов и;

2) относительно простым изложением уже существующего анализа и вторичных данных, чаще всего лишь с дополнительной визуальной обработкой (подготовка графиков, иллюстраций). В последнем случае, хотя аудитория и получает возможность осмыслить данные, пользуясь визуальными подсказками, “добавочная стоимость” публикации ниже, чем в первом случае, где производится дополнительная и возможно уникальная ценность (при условии компетентной обработки)<sup>25</sup>.

Флагманом ЖД в Кыргызстане можно считать интернет-издание Kloop.kg. 10 мая 2021 года СМИ Кыргызстана сообщили, что объявлен шорт-лист претендентов на награды The Sigma Awards - 2021. Жюри отобрало 141 проект для участия в финале самого престижного международного конкурса по ЖД. В 2021 году две публикации Kloop.kg и одно совместное расследование издания с

<sup>24</sup> Источник: Kloop.kg, 25 марта 2021 года.

<sup>25</sup> Аналитический обзор. Есть ли журналистика данных в КР? – Бишкек, 2017. – С 3.

azattyk, Bellingcat (Нидерланды) и Центром по исследованию коррупции и организованной преступности (англ. *Organised Crime and Corruption Reporting Project, OCCRP*) попали в список финалистов этого конкурса, где были представлены работы ведущих изданий Америки, Европы, Азии и Африки.

Одна из представленных на конкурс Sigma Awards работ из Кыргызстана называется «Республика Матрастан»<sup>26</sup>. В первой части этого довольно объёмного расследования повествуется о семейном клане Матраимовых, которые осев в таможенной службе республики и годами покрывая семью Абдукадыр - контрабандистов из Китая, получали огромные прибыли. Семьи Матраимовых и Абдукадыр установили контроль над тремя важными таможенными терминалами, которые обеспечивают сверхприбыльный поток законного и не очень законного дохода. В расследовании развернуто показаны не только схемы доставки товаров из китайского Кашгара в Бишкек, а оттуда - в Казахстан, Россию и Узбекистан за последние 7-8 лет, но и процедуры незаконного растаможивания товаров.

Вторая часть расследования - «Политические игры могущественного кыргызского клана» - повествует об использовании кланом Матраимовых средств, полученных сомнительным путем, для завоевания в ходе парламентских выборов 2020 года политического и общественного влияния в Кыргызстане через партию «Мекеним Кыргызстан». Авторы разработали также интерактивную карту собственности Матраимовых, на которой отражены все объекты недвижимости, приобретенные семьей в разных уголках страны, с точным указанием владельцев, параметров, стоимости, вплоть до регистрационных номеров в кадастровой службе. Текст сопровождается также рисованными и фотоиллюстрациями.

«Я бы ее все равно убил» - ещё одно расследование Kloop, попавшее в список Sigma Awards, опубликовано 17 декабря 2020 года. Оно посвящено выявлению причин и мотивов убийства женщин в стране на почве семейного насилия. Ежегодно в Кыргызстане происходят десятки убийств, когда женщина умирает от рук супруга, сожителя, бывшего мужа или знакомого мужчины. В международной практике это называется фемцид. Авторы проанализировали статистику убийств и изучили более 54 тысяч криминальных пресс-релизов и новостей, чтобы воссоздать картину фемцида в стране. Сопоставив базы данных, они пришли к выводу, что большинство убийств были совершены с особой жестокостью. Визуализировав имеющуюся базу данных, авторы в процентном выражении показали характер травм в случаях фемцида. В частности, от множества травм погибло 42%, от множества ранений – 24%, от травм шеи от удушения – 15%, от ножевых ранений – 8%, от огнестрельных ранений – 4%, ожогов – 1%, неизвестно – 7% убитых.

В расследовании интернет-издания 24.kg «Три просчета властей, которые привели к вспышке COVID-19» (29 сентября 2020 года), доказано, что летнюю вспышку коронавируса в 2020 году можно было предотвратить или хотя бы смягчить ее последствия, если бы власти последовали советам ученых. Халатность государства привела к потере как минимум 1 тысячи 362 жизней в июне-июле — на пике эпидемии. Журналисты указали на три основные ошибки, допущенные властями:

Ошибка 1. Отсутствие информационной кампании о важности социального дистанцирования. Когда в мае постепенно стали снимать локдаун, государство перестало объяснять людям, как важно соблюдать социальную дистанцию в общественных местах и карантин при наличии симптомов. По данным проекта Our World in Data, уже в мае Кыргызстан стал одной из четырех стран в мире, включая Туркменистан, Беларусь и Йемен, где не было скоординированной информационной кампании для населения.

Данные Google о передвижении населения также говорят об отсутствии достаточной системы информирования граждан. С введением ЧП системы.

---

<sup>26</sup> См.: <https://kloop.kg/blog/2020/10/02/custom-empire/>

Google зафиксировали резкое падение передвижений по всей республике — от поездок на работу до отдыха в парках и торговых центрах. Именно поэтому случаев заражения в те дни было крайне мало. Но в середине мая, после снятия карантина, количество поездок и прогулок резко выросло, а вместе с ними (с учетом инкубационного периода) начали расти и случаи коронавируса. Источник: 24.kg, 29 сентября 2020 года.

Ошибка 2. Чрезвычайные меры вместо системного укрепления здравоохранения. Меры принимались поздно. Если бы выполнялись решения январского заседания Совета безопасности как следует, то многих вещей можно было избежать, но, видимо, их выполняли формально. Отдел данных 24.kg собрал базу данных решений Кабмина в период коронавируса. Из ее анализа следует, что именно в пик заболеваемости начали судорожно открывать новые медицинские отделения и стационары<sup>27</sup>.

Ошибка 3. Слабая система тестирования и отслеживания контактов. По информации Our World in Data, до 1 июля 2020 года тестирование в республике проводилось лишь для лиц с симптомами, которые находились в группе риска, например: прибыли из другой страны, являются ключевыми работниками или поступили в больницу. Эта политика снова вернулась с конца августа, когда вспышка заражения угасла.

Развитие ЖД в Кыргызстане тесно связано с тенденцией расширения доступа к данным. В этом большую роль играют государственные органы, которые проявляют все большую готовность делиться данными и признают право общества и СМИ на доступ к данным в ведении государства. Особенных трудностей с поиском данных в стране на сегодня нет, можно запросить и получить любую информацию. В некоторых редакциях постепенно появляется практика специализации журналистов по работе с данными, поскольку это начинает обеспечивать эксклюзивность материалов, обратную связь с аудиторией, а равно и повышает рейтинги издания.

В Кыргызстане действует портал открытых данных, где опубликовано 646 наборов данных в XLSX формате по разным тематикам. Данные доступны для скачивания без регистрации. В этом же формате публикуются данные таможенной статистики, кадастровой системы — списки национального имущества. Данные портала «Открытый бюджет» публикуются в основном в агрегированном виде в формате графиков и диаграмм и html-таблиц, которые можно собрать с сайта с помощью инструментов веб-скрейпинга. С помощью инструментов веб-скрейпинга, таких как Scraper API, Octoparse, DataOx, и др. Официальный портал госзакупок в стране имеет открытое API для разработчиков, а данные загружаются в компактном JSON формате. Официальные статистические данные Нацстаткома публикуются также в формате открытых данных, тоже для удобства в разных машиночитаемых форматах: JSON, XML, XLSX. В формате PDF публикуются базы данных юридических лиц, филиалов и представительств, отчеты об итогах Общереспубликанского тестирования и зачисления на грантовые и контрактные места вузов. В формате html-таблиц публикуются данные о товарных знаках, список банков и реестр лицензий. Данные можно собрать в табличный вид с помощью инструментов веб-<sup>28</sup>.

Несмотря на все эти достижения, в целом, ЖД в Кыргызстане находится на начальном этапе развития. Среди барьеров, ограничивающих развитие ЖД — трудоемкость и значительное время для подготовки качественного материала, отсутствие необходимых технических и аналитических навыков у журналистов и редакторов. Им необходимо регулярно совершенствовать навыки работы с графическими инструментами и другим программным обеспечением, позволяющим создавать инфографические медиапродукты<sup>29</sup>. При обработке первичных данных журналисты часто ограничиваются вычленением информации необходимой для данной конкретной статьи или

---

<sup>27</sup> Источник: 24.kg, 29 сентября 2020 года

<sup>28</sup> [https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism\\_web.pdf](https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism_web.pdf)

<sup>29</sup> Аналитический обзор. Есть ли журналистика данных в КР? — Бишкек, 2017

заметки. Лишь небольшой процент материалов содержит элементы интерактивности и предоставляет читателю возможность управлять отображением данных. Также имеет место ограниченное использование экспертного мнения. Большинство материалов, представляющих собой текст и столбчатый график, они основаны на открытых данных без использования дополнительных комментариев компетентных лиц.

Между тем, обучающие курсы по журналистике данных также впервые в регионе были введены в университетские программы в Кыргызстане. Первый такой курс был открыт в АУЦА (Американском Университете в Центральной Азии) в 2017 году. Факультет журналистики и массовых коммуникаций АУЦА провел и первый Центрально-Азиатский летний институт по журналистике данных для сотрудников СМИ стран региона при поддержке агентства ПРООН. В последующие годы проводились Хакатоны и форумы по журналистике данных. А в 2018 году в Кыргызстане была основана Школа Данных, силами которой были проведены различные тренинги по журналистике данных для преподавателей факультетов журналистики Кыргызстана. В настоящее время данная дисциплина кроме АУЦА преподается еще в международном университете Алатао и Университете Центральной Азии. В Кыргызстане также было опубликовано первое руководство по журналистике данных «Ищем сокровища в данных» А. Валеевой, А. Мамбетовой, Ч. Узакбаевой.

Потенциал журналистики данных в Кыргызстане недоиспользован и его раскрытие на должном уровне позволит укрепить качество общественного дискурса, улучшить информированность общественности по ключевым темам страновой повестки, и расширить вклад в улучшение качества принимаемых решений. Особенную важность усилиям по развитию дата журналистики придают продолжающееся улучшение доступа к большим массивам данных, в привязке к инициативам открытых данных, а также постоянное совершенствование технических платформ, сервисов и навыков их использования для обработки данных всеми заинтересованными сторонами, включая журналистов.

Анализ состояния журналистики данных в Кыргызстане, проведенный отмеченным уже выше Promotank Research Institute в 2017 году, показал, что большинство материалов (65.2% публикаций) можно охарактеризовать именно как графическую обработку и изложение вторичных данных. Такое распределение, где реже встречаются примеры дополнительной аналитической обработки данных может объясняться необходимостью более сложных навыков работы с данными в редакции, временных и материальных затрат на подготовку материала, уровнем доступности данных, позволяющим провести дополнительные манипуляции.

Тематический охват публикаций в выборке достаточно разнообразен и варьируется почти по всем основным сферам жизнедеятельности общества. При этом отмечается значительное преобладание материалов в рубрике “социальное”, широкой категории, включающей в себя такие подтемы, как жилищно-коммунальная инфраструктура и муниципальная сфера, преступность, и срез мнений по основным социально-политическим вопросам, волнующим общественность. Другой категорией с высокой частотой является финансово-экономическая сфера, что может быть обусловлено как высоким интересом аудитории, позволяющем поддерживать деятельность изданий, полностью специализирующихся в этом тематическом направлении, так и распространенностью первичных данных и источников данных. Из этой сферы были выделены публикации, связанные с государственным бюджетом и также встречающиеся с относительно высокой частотой. Как видно на представленном ниже графике, остальные тематические направления гораздо реже становятся основной для материалов в формате журналистики данных, что может быть связано с низким уровнем заинтересованности аудитории и доступностью данных.

Использование визуализаций данных является характерной особенностью большинства материалов, включенных в выборку. Основными группами визуализаций, наблюдаемыми в выборке являются публикации с преимущественным или центральным использованием: 1)

диаграмм и таблиц; 2) инфографики - графических символов, знаков и иконок; 3) картирования; 4) интерактивных визуализаций, со встроенной возможностью взаимодействия с представленными данными.

Крайне редко наблюдается комбинация видов визуализаций, но даже в таких случаях можно выделить вид визуализации, определяющий основную смысловую линию публикации. За исключением последней группы визуализаций, эти основные группы представлены равномерно. Проанализировав выборку материалов по визуализации, можно сделать вывод о том, что диаграммы и таблицы, традиционно представлявшие основной способ визуальной передачи смысла данных, получают свое продолжение в более комплексных форматах передачи смысла.

Исходя из анализа визуализаций можно сделать заключение, что в некоторых изданиях имеются творческие команды, состоящие из журналистов, которые занимаются анализом и обработкой данных, дизайнера или графического редактора, который подготавливает необходимые иллюстрации, и возможно веб-мастера (если это онлайн-издание). В других же изданиях можно полагать, что материалы подготовлены журналистами самостоятельно, зачастую без доступа к профессиональным техническим решениям (например, к платным сервисам и программам). Вследствие этого наблюдается большая разница между инфографикой, над которой работал профессиональный дизайнер, и теми материалами, где графики создавались самими журналистами: в первом случае это визуально привлекательная работа с использованием графиков, иконок, символики, карт и проч., во втором – простой линейный или столбчатый график, на основе функций бесплатных онлайн-сервисов.

Несмотря на то, что журналистика данных зачастую полагается на комплексные инструменты количественного анализа и графической обработки, в центре лучших журналистских произведений в этом формате всегда остаются живые люди, их истории и судьбы. Однако, в рассмотренной выборке материалов, «история героя», или, другими словами, человеческая история, используется в качестве основы или составного компонента для статьи лишь в 2% случаев, и отсутствует во всех остальных случаях. Возможное объяснение такой ограниченной представленности метода подачи материала с включением человеческой истории связано с тем, что данная выборка включала в себя большое количество новостных материалов, где информационный жанр не предусматривает расширенной публикации. В любом случае, недостаток внимания к человеческим историям свидетельствует о слабо освоенном, и упущенном потенциале для дополнительного сильного воздействия публикаций.

### **Казахстан**

ЖД в Казахстане также «набирает обороты». Тому способствует и стремление властей страны ко все большей открытости. Государственный портал «Открытые данные» утверждает, что с его помощью сегодня «общественность может легко находить, загружать и использовать наборы данных, которые создаются государственными органами». В 2018–2019 годы публикации прессы, подготовленные с помощью инструментария ЖД, в основном, представляли собой инфографику и повествование об актуальной социально-экономической проблеме.

Пандемия и корона-кризис основательно повлияли на развитие ЖД в стране. Стало популярным создавать интерактивные дэшборды, на которых отображаются различные показатели в режиме реального времени. На данный момент в режиме реального времени можно проследить количество вакцинированных граждан Казахстана. Карта вакцинации Казахстана доступна на сайте <https://www.coronavirus2020.kz/><sup>30</sup>

ЖД успешно практикуется в таких общенациональных газетах, как «Егемен Қазақстан», «Казахстанская правда», «Время», «Караван» и в онлайн изданиях [www.bnews.kz](http://www.bnews.kz), [tengrinews.kz](http://tengrinews.kz),

---

<sup>30</sup> <https://www.coronavirus2020.kz/>

nur.kz и др.<sup>31</sup>. Обзор материалов за 2019 год показал, что редакциями выше названных газет и онлайн порталов инфографика используется, но крайне редко. В то же время можно отметить, что казахстанские СМИ имеют тенденцию публиковать много материалов, часто довольно коротких, о последних цифрах, которые выходят из официальной национальной статистики. Принцип использования инфографики заключается в том, чтобы дополнить текстовый материал. Чаще всего инфографика используется под такими рубриками, как «Актуальные цифры», «Транспорт», «Социум» и т.д.

Так, в газете «Казахстанская правда», в отдельном блоке «Инфографика» (<https://www.kazpravda.kz/rubric/infografika>) за 2017–2018 годы можно найти 3, а за 2019 год - 4 статьи с инфографикой. Это - «Платежный расклад» (7 октября 2019 г.), «Социальное государство» (19 марта 2019 г.), «В поисках баланса» (31 января 2019 г.), «Год молодежи» (24 января 2019 г.). В корреспонденции «Когда начнут распределять места в детсады Нур-Султана» 9 июня 2020 года газета сообщает о том, что с 17 июня 2020 года планируется начать плановое распределение мест в дошкольные организации Нур-Султана. В электронной системе e.astana.kz или на сайте indigo24.kz публикуется информация об освободившихся местах в дошкольных организациях (частные сады и мини-центры). В 2020 году имелось 20 198 мест (в 2019 году освободилось 15 тыс. мест). Текст корреспонденции сопровождается таблицей с указанием наличия мест<sup>32</sup>. В «Егемен Казахстан» в 2021 году опубликованы две небольшие дата-заметки. В первой – «Куда чаще всего летают казахстанские самолеты?» - визуализированы данные состоянии международных рейсов из Казахстана в 12 странах мира<sup>33</sup>.

Во второй заметке визуализированы данные по строительству жилья в Казахстане за 2020 год<sup>34</sup>.

На интернет-платформе издания «Курсив» действует отдельный раздел «Тенденции и исследования», в котором опубликовано много историй и наборов данных по разным темам. Среди них можно отметить статью «Как стать звездой в Казахстане и сколько на этом можно заработать». В ходе подготовки материала автором проведен анализ данных столичного продюсерского центра KS Production. Проблемой для дата-журналиста в освещении подобных инфографик может стать отсутствие необходимых данных. Многие из компаний, фирм, те же самые продюсерские центры, ссылаясь на коммерческую тайну или чрезвычайную занятость, зачастую отказываются отвечать на официальные запросы журналистов.

На многих казахстанских новостных платформах, на сайтах государственных учреждений, банков и т.д. размещены также «персонализированные калькуляторы», призванные помочь людям принимать необходимые решения. Они могут касаться вопросов покупки автомобиля, дома (пример с сайта kursiv.kz), принятия решения об образовании или профессиональном жизненном пути, или тщательной проверки расходов, чтобы избежать долгов и так далее. Так, на портале Krisha.kz по вопросам недвижимости на основе открытых данных составляются полезные рейтинги городов Казахстана по ценам на недвижимость, продажам, покупкам жилья, безопасности районов и городов и т.д. Такие рейтинги, к примеру, помогли изданию «Курсив» создать инфографику на тему «Какие автомобили выбирают казахстанцы».

---

<sup>31</sup> Барлыбаева С.Х., Малтабаров М.А. Дата-журналистика в Казахстане как медиатренд. Хабаршы. Журналистика сериясы. №1 (55) 2020. <https://bulletin-journalism.kaznu.kz>

<sup>32</sup> <https://www.kazpravda.kz/news/obshchestvo/nazvana-data-nachala-raspredeleniya-mest-v-detsadi-nur-sultana/2020/06/09/>

<sup>33</sup> Қазақстандық ұшақтар қайда жиі ұшады? //egemen.kz/article/260475-qazaqstandyq-ushaqtar-qayda-dgii-ushady

<sup>34</sup> Қазақстандық ұшақтар қайда жиі ұшады? //egemen.kz/article/260475-qazaqstandyq-ushaqtar-qayda-dgii-ushady

Изучение деятельности казахстанских СМИ в ракурсе ЖД показал, что их работа условно организована на двух уровнях. На первом уровне происходит анализ с целью выведения смысла и структуры из бесконечного потока данных. И второй уровень - представление того, что хочет получить потребитель (читатель/зритель). То есть важно дать то, что важно и актуально на сегодня.

В этом смысле онлайн-издания более активны и выигрывают в инфографике. Для них характерно то, что в них размещаются как текстовые материалы, так и содержательная визуальная информация в виде графиков, таблиц и т.д.

К примеру, проект «Древо жизни» в интернет-издании «Власть» посвящен деятельности Национального управляющего холдинга «Байтерек»<sup>35</sup>. Холдинг создан в 2013 году как структура, которая призвана эффективно управлять работой финансовых институтов и институтов развития страны. Вся работа холдинга визуализирована в виде интерактивного дерева, на ветвях которого отражены данные по 5 основным направлениям деятельности: укрепление жилищной инфраструктуры, национальные чемпионы, освоение средств из национального фонда, оздоровление и поддержка проектов, привлечение инвестиций.

Несколько лет назад «Власть» в партнерстве с Kaspi составили первый в стране рейтинг казахстанских школ, ученики которых на протяжении последних семи лет показывали высокие результаты по математике. Этот рейтинг был подготовлен на основе открытых данных, и кроме, собственно, самого рейтинга в этом проекте была представлена полезная информация о том, чем математики занимаются в современных компаниях, например, в Kaspi или Beeline.

Factcheck.kz – один из самых значимых казахстанских ресурсов. Практически каждая публикация этого интернет-издания создана с помощью инструментов ЖД. Чиновники любят разбрасываться неправдоподобными цифрами, которые можно проверить за несколько часов, не покидая пределов редакции. Этим и занимаются фактчекеры издания. Например, на пленарном заседании Мажилиса РК 19 мая 2021 года депутат Ерлана Саиров в своем запросе на имя заместителя премьер-министра Романа Скляра предложил создать независимую государственную аудиторскую систему дорожного хозяйства местной сети автодорог, поскольку только 9,3% дорог страны могут воспринимать нагрузки более 10 тонн на ось, при этом 90,7% всех дорог не способны воспринимать даже минимальные нагрузки. Редакция по следам этого заявления провела собственное расследование на базе открытых данных по состоянию дорог и выявила, что они противоречат заявлению депутата<sup>36</sup>. Текст расследования сопровождается таблицами и инфографикой

Или история о том, как «Казахстан обеспечивает себя продовольствием на 80%»<sup>37</sup>. Такое заявление сделал министр национальной экономики Тимур Сулейменов. Редакция проверила заявление министра и доказала, что страна не обеспечивает себя продовольствием на 80 %. Изучив открытые данные Комитета статистики, журналисты построили диаграмму производства и импорта продуктов питания, диаграмму по показателям 8 крупнейших производителей и экспортеров зерновых и выявили, что министр манипулирует цифрами. Реальная ситуация оказалась ровно противоположной заявленному.

Еще один наглядный пример ЖД в Казахстане проект - Protenge.kz. Это аккаунт в социальной сети Instagram и Telegram, где публикуются результаты расследований, основанных на открытых данных. В описании изложено следующее «... следите вместе с нами на что тратят ВАШИ деньги» (орфография сохранена - *прим.авт.*). «Я люблю документы. Они никогда не подведут журналиста. Люди могут врать или отказываться от своих слов. А документы не врут никогда», - говорит Джамиля Маричева, один из лучших дата-журналистов Казахстана. Журналист также отмечает, что «в

---

<sup>35</sup> Древо жизни.// <https://vlast.kz/tree/>

<sup>36</sup> Макишева А. Ерлан Саиров о состоянии дорог страны и грузовиках.// [factcheck.kz/26/05/2021](https://factcheck.kz/26/05/2021)

<sup>37</sup> Клеменкова К. Сулейменов: Казахстан обеспечивает себя продовольствием // [factcheck.kz/19/09/2018](https://factcheck.kz/19/09/2018)

Казахстане много баз данных, они как Эверест, там кроется невероятная, качественная информация, которую никогда ни один чиновник не даст в ответ на запрос журналиста»<sup>38</sup>. Ниже приведем некоторые примеры проекта Protenge.kz которые были основаны на открытых данных<sup>39</sup>. Выводы экспертов: Казахстан имеет базовые условия для развития ЖД. В стране действует собственный национальный портал с открытыми данными. При этом данные на официальном портале открытых данных доступны только по регистрации для пользователей, имеющих специальный официальный идентификатор. Разработчикам, чтобы получить доступ к API (портал открытых данных, портал открытого бюджета, госзакупок) нужно запросить доступ - написать обращение в госорган, указав цель использования данных. После чего, в случае одобрения их заявки на данные, Департамент цифровизации и государственных услуг отправит им токен (ключ) для авторизации. Помимо того, что для работы с открытыми данными нужна авторизация. Статистические данные можно скачать свободно в агрегированном виде, в т.ч. в структурированном XLSX-формате<sup>40</sup>.

### Таджикистан

Первые проекты ЖД в Республике Таджикистан появились ещё в 2013 году, они представляли собой небольшие заметочного формата публикации с инфографикой. По мере освоения компетенций ЖД в рамках центральноазиатских тренингов в медиапространстве республики стали появляться проекты с использованием более сложных инструментов – интерактивных карт, дисковых и графических таблиц.

Журналисты из медиagrппы Asia-Plus визуализируют данные на основе различных исследований или отчётов. К примеру, в проблемной корреспонденции «Злостные неплательщики? Население Душанбе задолжало «Барки точик» 95 миллионов сомони» журналист Пайрав Чоршанбиев повествует о том, что жители Душанбе оплатили в 2020 году только 72,2% полученной электроэнергии<sup>41</sup>. Ежесуточное энергоснабжение города составляло 6,3 млн кВт.ч, что соответствует примерно 10% от общего энергоснабжения по республике. Технические потери электричества в сетях столицы, как и в 2019 году, оцениваются в 15,5%. Общая стоимость потребленной в столице электроэнергии в 2020 году составила 678 млн. сомони, а суммарный размер оплаченного электричества - 542 млн. сомони. То есть, потребители задолжали «Барки точик» 136 млн. сомони, что составляет около 20% в общем объеме потребленной энергии в Душанбе. Главными должниками называются бытовые потребители, на которых приходится почти 70% от совокупной задолженности по городу - около 95 млн. сомони. В мэрии отмечают, что население столицы получило электричество на сумму 341,5 млн., а оплатило только 246,7 млн. сомони (72,2%). В редакционной инфографике показано, сколько в среднем домашние приборы потребляют электроэнергии, и сколько надо будет платить за их работу. В статье «А он все растёт и растёт»<sup>42</sup>. Внешний долг Таджикистана за полгода вырос на 26 миллионов долларов» Asia-Plus поведала о структуре внешнего долга Таджикистана, общий размер которого составил 35,8% ВВП республики<sup>43</sup>. По словам первого заместителя министра финансов РТ Дж. Каримзода,

<sup>38</sup> <https://ru.internews.kz/2020/04/protenge-kz-dzhamili-marichevoj-kak-delat-data-zhurnalistiku-v-instagram/>

<sup>39</sup> [https://www.instagram.com/p/CQGnCVnHci/?utm\\_medium=copy\\_link](https://www.instagram.com/p/CQGnCVnHci/?utm_medium=copy_link) /  
[https://www.instagram.com/p/CFzRyQln2hd/?utm\\_medium=copy\\_link](https://www.instagram.com/p/CFzRyQln2hd/?utm_medium=copy_link)

<sup>40</sup> <https://www.gov.kz/memleket/entities/aqmola-selinograd/press/news/details/426804?lang=ru&ysclid=m3om850c85584213414>

<sup>41</sup> <https://www.asiaplustj.info/ru/news/tajikistan/society/20210111/naselenie-dushanbe-zadolzhalo-barki-tochik-95-millionov-somoni>

<sup>42</sup> <https://www.asiaplustj.info/ru/news/tajikistan/society/20210111/naselenie-dushanbe-zadolzhalo-barki-tochik-95-millionov-somoni>

<sup>43</sup> <https://www.asiaplustj.info/ru/news/tajikistan/economic/20170728/a-on-vse-rastet-i-rastet-vneshnii-dolg-tadzhikistana-za-polgoda-viros-na-26-millionov-dollarov>

обслуживание внешнего долга в первой половине года осуществлено в размере 84,56 млн. долларов. В выше приведенной редакционной инфографике указаны размеры внешнего долга страны по инвестиционным проектам, а также основные внешние кредиторы страны – китайский Эксимбанк, Всемирный банк, Азиатский банк развития, Исламский банк развития и Международный валютный фонд.

Серьёзные работы, построенные на открытых данных, опубликованы на сайте «Sputnik Таджикистан». В проекте «Скидки и льготы: сколько стоит проезд в Душанбе» проанализированы открытые данные по состоянию столичного транспорта<sup>44</sup>. Сегодня в Душанбе действуют 83 маршрутные линии транспорта. По ним можно добраться до любой локации в районах Фирдавси, Сино, Исмоили Сомони и Шохмансур. В городе насчитывается 49 микроавтобусных, 26 автобусных и 8 троллейбусных линий. Дешевле всего обойдется поездка в пределах столицы на троллейбусе. Ее стоимость оценивается в 1,00 сомони, а дороже - на маршрутке (1,60 сомони). Для детей до 5 лет проезд бесплатный. При этом в Душанбе действует ряд скидок при покупке проездного на 50 и более поездок. Инфографика к тексту представлена в цветном исполнении с использованием различных видов иконографии.

Много выступлений таджикских дата-журналистов опубликовано и на центральноазиатском портале «Sabar.Asia». К примеру, в аналитической публикации «Госбюджет Таджикистана: высокая налоговая зависимость» за 30 декабря 2020 года авторы, проанализировав открытые данные, пришли к выводу о том, что для обеспечения налоговых сборов, правительству необходимо снизить налоговую нагрузку на частный бизнес, который обеспечивает до 80% налоговых поступлений в бюджет. В противном случае предприниматели будут выводить свой капитал из страны при каждом удобном случае.

Анализ структуры налогов показывает, что налоговые поступления снижались во время кризисов и начали восстанавливаться в посткризисных периодах. Например, сборы налогов после глобального финансово-экономического кризиса 2008 года имели тенденцию к росту до 2014 года. Далее, вследствие экономического кризиса в 2014 году, в России, которая является главным торгово-экономическим партнером Таджикистана и основным рынком труда для таджикских мигрантов, объем налоговых доходов снова стал снижаться в течение двух лет. В посткризисный период сбор налогов снова несколько восстановился. Согласно официальным данным по исполнению бюджета за 10 лет (2010–2020 гг.), в среднем 68% от доходов приходится на налоговые поступления. Примерно 15% в среднем составляют различные кредиты и гранты, 9% – спецсредства бюджетных организаций, 7% – неналоговые поступления, 1,3% – свободные остатки предыдущего года. Текст статьи сопровождается тремя линейными графиками и дисковой диаграммой, на которых отражены соотношение налоговых поступлений к общим доходам госбюджета, источники госбюджета, соотношение видов налога к налоговым поступлениям в госбюджет и источники его пополнения в 2010–2020 годы. Как отмечает один из авторов публикации Бахранер Надимов, на обработку, очистку и составление инфографики времени было затрачено гораздо больше, чем на поиск открытых данных.

В развитие ЖД в Таджикистане большой вклад вносят и блоггеры. К примеру, Муллоражаб Юсуфзода на своем блоге дата-проекты в различных жанрах<sup>45</sup>. Так, в расширенной дата-заметке «Быстрый рост количества безработных в Таджикистане» он представил интерактивную инфографику с отражением темпов роста численности безработных по всем регионам страны с сентября 2019 года по сентябрь 2020 года<sup>46</sup>. Другой блоггер Алишер Чунайдов в своих дата-проектах

---

<sup>44</sup> <https://tj.sputniknews.ru/20210603/dushanbe-proezd-stoimost-2021-1039959178.html>

<sup>45</sup> <http://econewstj.blogspot.nl/>

<sup>46</sup> Афзоиши босуръати шумораи бекорон дар Тоҷикистон.// [http://mullorajab.blogspot.com/2016/07/blog-post\\_21.html](http://mullorajab.blogspot.com/2016/07/blog-post_21.html)

предпочитает освещать политические, экономические и социальные проблемы<sup>47</sup>. К примеру, в дата-корреспонденции «Может ли Бадахшан выйти из субсидий?» автор, анализируя государственное субсидирование Горно-Бадахшанской автономной области (ГБАО) за 2014–2016 годы приходит к выводу, что область, располагающая значительными природными ресурсами и туристским потенциалом, может стать самодостаточной<sup>48</sup>. Эти данные приведены ниже следующей столбичной таблице:

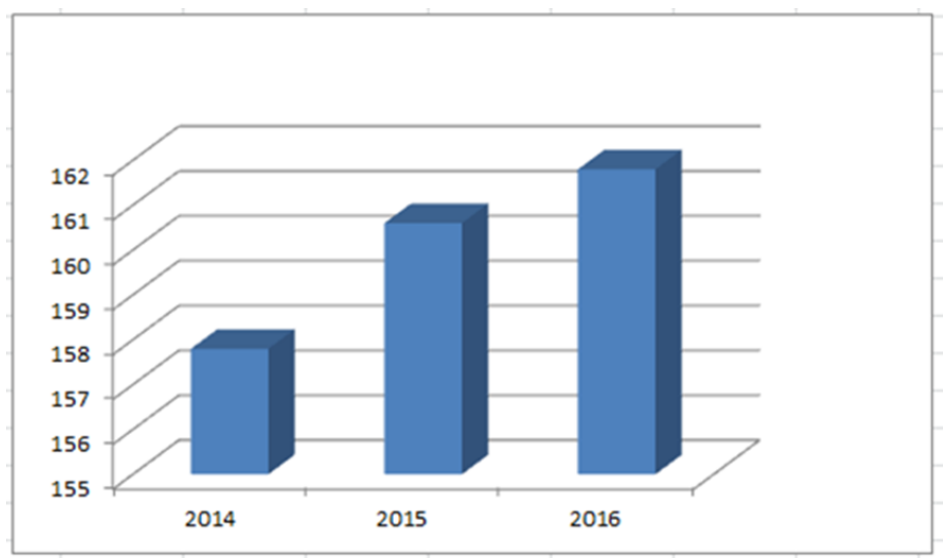


Рисунок № 3. Источник: Бадахшон тавони аз субвенсия баромаданро дорад?[https://openpamir4you.blogspot.com/2016/04/blog-post\\_47.html](https://openpamir4you.blogspot.com/2016/04/blog-post_47.html)

По данным Геологической службы страны, в горах Памира насчитывается более ста видов природных ресурсов, 57 из которых - месторождения драгоценных камней. Между тем, освободить от госдотаций Бадахшан может развитие энергетического сектора. Более того, продажа электроэнергии в соседние страны позволит принести в госказну значительные поступления. Компания «Памир Энерджи» более 10 лет развивает энергетический сектор региона, и впервые жители получили возможность пользоваться электроэнергией без ограничений. Публикацию включает интерактивная карта, на которой подробно указаны основные направления проекта трансграничной энергетической торговли между Таджикистаном и Афганистаном при содействии правительства США. Подробное описание целей и задач проекта к интерактивной карте представлено на английском языке.

Основные выводы: на сайтах госорганов можно найти данные о бюджете, национальную и таможенную статистику, которые публикуются агрегировано в открытом доступе в разных форматах (PDF, XLSX, DOC). В открытом доступе можно найти даже экологические данные в формате XLSX, собранные и опубликованные НКО 35 Living Asia. Однако большая часть данных являются

<sup>47</sup> <https://openpamir4you.blogspot.nl/>

<sup>48</sup> Бадахшон тавони аз субвенсия баромаданро дорад?// [https://openpamir4you.blogspot.com/2016/04/blog-post\\_47.html](https://openpamir4you.blogspot.com/2016/04/blog-post_47.html)

историческими, актуальные данные за последние 4 года не публиковались<sup>49</sup>. В прессе мало интересных материалов об экономике. Большинство журналистов не могут работать с данными, составлять таблицы, отчеты, проводить аналогии и т.п. Им не хватает знаний и опыта и навыков аналитического мышления для подготовки качественной экономической аналитики<sup>50</sup>. Есть трудности доступа к информации. Официальные государственные лица скрывают от журналистов важную информацию. Журналисты и эксперты сталкиваются со сложностями при обработке данных, так как данные самих правительственных органов зачастую серьезно расходятся между собой.

### **Узбекистан**

Первые обучающие тренинги по ЖД в Узбекистане были инициированы Представительством ЮНЕСКО в 2017 году. Тогда журналисты столицы и регионов впервые узнали об открытых международных и национальных базах данных, смогли на практике опробовать сервисы для обработки и визуализации больших массивов цифровой информации, научились искать истории в этих массивах. С тех пор журналисты заинтересованных редакций СМИ Узбекистана принимают активное участие в международных и региональных обучающих проектах по ЖД.

Безусловно, география дата-журналистских материалов в массмедиа Узбекистана еще не так обширна, однако практику таких интернет-изданий, как Gazeta.uz, Anxor.uz, Spot.uz, Podrobno.uz. можно рассматривать как важную основу для и развития ЖД в национальном медиапространстве.

Так, проблематика материалов с использованием атрибутов ЖД в Gazeta.uz затрагивает вопросы экономики, здравоохранения, социальные, в том числе занятость населения, включая и людей с инвалидностью, их социальную защиту в условиях пандемии.

К примеру, в статье «ВБ - о восстановлении экономики после кризиса COVID-19» (29 июня 2020 г.) Gazeta.uz анализирует ситуацию в экономике Узбекистана и её перспективы. По мнению автора, стране нужны быстрые «беспроигрышные» инвестиции, расширение программ социальной помощи самым уязвимым слоям населения и готовность к новой волне коронавируса. В статье использованы данные, собранные в ходе проведения исследования «Слушая граждан Узбекистана» (Listening to Citizens of Uzbekistan, L2CU). Данный проект осуществлялся с 2018 года при содействии Всемирного банка совместно с Центром «Стратегия развития», а также с Центром экономических исследований и реформ с целью оценки благополучия граждан и общественного мнения касательно реформ, осуществляемых правительством. Данные представлены в виде таблицы и диаграмм.

Gazeta.uz в рубриках «Колумнисты», «Экономика», «Коронавирус», «Мнения» систематически публикует статьи по актуальным вопросам экономики различных отраслей, в которых активно используются схемы, графики, диаграммы, гистограммы, таблицы. Gazeta.uz отличилась и тем, что в разгар пандемии редакция смогла провести опрос населения, результаты которого были представлены в публикации «Трудно, но о помощи пока не прошу». Всего в опросе приняли 328 читателей, прислав свои сообщения длиной от нескольких слов до короткого рассказа. Опрос не претендовал на репрезентативность, указанные в ответах участников обстоятельства не подвергались изучению, в отдельных случаях у респондентов были запрошены уточнения по телефону или по Telegram. Так, 71,6% участников опроса указали, что проживают в Ташкенте, 10,6%

---

<sup>49</sup> ] Буров В., Орлова Кс. Дата-журналистика и открытые данные в странах Центральной Азии. Скрытые возможности симбиоза данных, журналистики и технологий искусственного интеллекта для развития регионального потенциала. - [https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism\\_web.pdf](https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism_web.pdf)

<sup>50</sup> Ақобирова Ш. Проблемы освещения экономических проблем в современной таджикской прессе. – Душанбе: International Alert 2016.// <http://www.khoma.tj/khoma/wp-content/uploads/2018/09/Business-journalism-in-Tajikistan-RUS.pdf>

— в Ташкентской области, 4,5% — в Самаркандской, 2,7% — в Ферганской, 2,4% — в Навоийской, 1,8% — в Бухарской, 1,5% — в Каракалпакстане, 1,2% — в Андижанской, Сырдарьинской и Кашкадарьинской областях, менее 1% — в Наманганской и Сурхандарьинской областях.

Результаты опроса редакция проиллюстрировала конкретными историями участников. Так, *Евгения из Ташкента рассказывает*: «Мой бизнес — салон красоты. Пришлось закрыть. Двадцать человек без работы, без средств на жизнь. Если после 10 мая карантин продлят, буду закрывать фирму. На фирме имелись задолженности и до карантина, но при положительной динамике надеялась за год раскрутить и погасить долги. В основном, за аренду помещения. Теперь это кажется всё менее вероятным. Зарплату я всё равно должна начислять, а денег на счету уже нет. Карточка. Даже когда карантин снимут, клиентов будет мало и не все мастера выйдут, так как записи не будет полной. Так что останавливать молодой бизнес равно как убить его совсем». *Матлуба из Бухарской области сообщает*: «С внуками дома, без кормильца. Сын в России на карантине. Без работы. Семья с детьми: я, сноха и дети. Семь человек. Нуждаемся в материальной помощи». *Озода из Навоийской области*: «У нас с мужем четверо детей. ... Работаем на Навоийском центральном рынке, продаём женскую одежду. У нас нет сбережений, абсолютно. Все деньги вкладываем в оборот до последнего, чтобы покрыть кредиты и текущие расходы. Слава богу, живём с родителями мужа. Этот месяц дома прожили на их пенсию. Если бы не было пенсии, пришлось бы обращаться к органам соцподдержки»<sup>51</sup>.

В рубрике «Мнения» и «Общество» Gazeta.uz часто выступают эксперты из различных отраслей, которые оперируют открытыми базами данных. Как правило, эти выступления помимо фотоиллюстраций снабжены графиками, диаграммами и таблицами со статданными, а иногда — и картами. К таковым можно отнести публикацию «Где наши рабочие места?» Д. Юсупова, приуроченную к 3 декабря — Международному дню людей с инвалидностью<sup>52</sup>. В статье использованы данные ситуационного анализа ООН, Общества слепых и глухих РУз, Министерства занятости и трудовых отношений. Текст обогащен тремя столбичными диаграммами, иллюстрирующими доступ людей с инвалидностью и без инвалидности к трудоустройству, разницу в их среднемесячной зарплате, препятствия ВТЭК для трудоустройства людей с инвалидностью.

Автор, проанализировав огромный массив данных, пришел к выводу, что люди с инвалидностью в стране продолжают сталкиваться с разными формами дискриминации на рынке труда.

---

<sup>51</sup> Трудно, но о помощи пока не прошу, 27 апреля 2020 г.

<https://www.gazeta.uz/ru/2020/04/27/results/>

<sup>52</sup> Юсупов Д. Где наши рабочие места?// <https://www.gazeta.uz/ru/2019/12/03/workplace/>



Рисунок № 4. Доступ к трудоустройству и занятости  
 Источник: <https://www.gazeta.uz/ru/2019/12/03/workplace/>

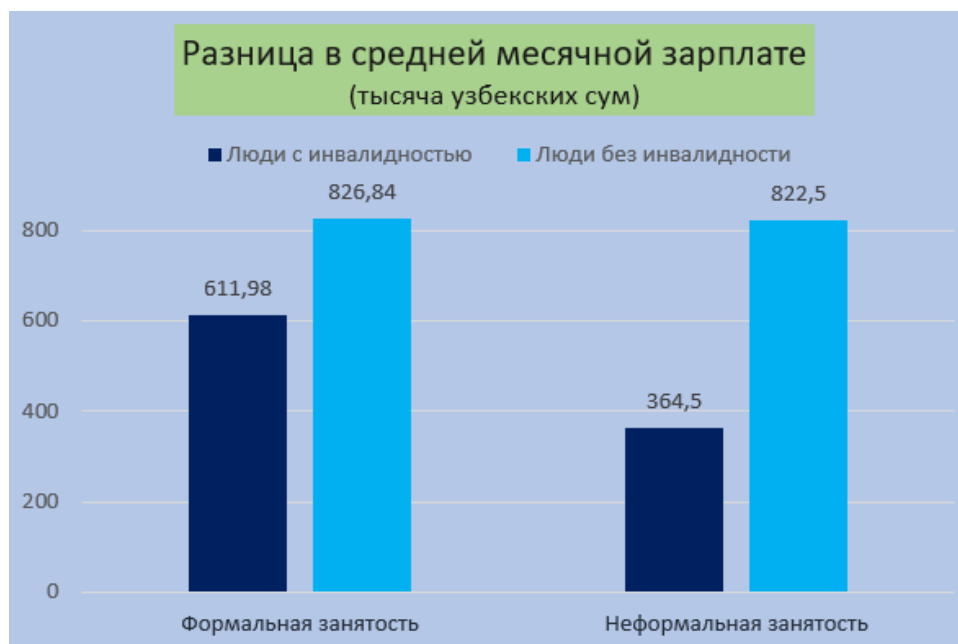


Рисунок № 5. Разница в средней месячной зарплате  
 Источник: <https://www.gazeta.uz/ru/2019/12/03/workplace/>



Рисунок № 6. Препятствия ВТЭК для трудоустройства  
Источник: <https://www.gazeta.uz/ru/2019/12/03/workplace/>

В Узбекистане официально зарегистрировано 346.084 человека с инвалидностью трудоспособного возраста. По данным ситуационного анализа ООН лишь около 7% людей с инвалидностью трудоустроены. Однако, в этих цифрах отображены только лица, которые смогли получить официальный статус инвалидности через врачебно-трудовые экспертные комиссии (ВТЭК), в то время как большинство остается невидимым. Данные исследования ООН показывают, что у людей с инвалидностью примерно в 4 раза меньше шансов трудоустроиться. Почти половина работают по неполному рабочему графику, в то время как на неполной занятости — лишь треть людей без инвалидности. В неформальном секторе заняты 40% людей с инвалидностью. В среднем зарплата людей с инвалидностью в формальном секторе намного меньше, чем у лиц без инвалидности. Еще хуже обстоят дела в неформальном секторе, где люди с инвалидностью зарабатывают более чем в два раза меньше, чем другие без инвалидности. Права людей с различными формами инвалидности на достойный труд нарушаются регулярно. Судьбы людей с инвалидностью в Узбекистане вершат всемогущие комиссии ВТЭК. Не зря их называют одновременно и врачебными, и трудовыми, потому что по закону они имеют право определять не только наличие физических или умственных нарушений, но и пригодность человека к работе. Заключение такой экспертизы содержит медико-экспертные, профессиональные и трудовые, а также социально-экспертные рекомендации, которые могут негативно влиять на возможности трудоустройства людей с инвалидностью.

Исследуя главные аспекты проблемы, автор заключает, что, не обеспечивая людей с инвалидностью достойной работой, мы подвергаем более высокому риску социальной и экономической уязвимости не только их, но и государство в целом.

Еще один значимый проект издания по ЖД – «COVID-19, женщины и их роль в восстановлении Узбекистана после пандемии» - посвящен рассмотрению вопросов гендерного равенства в обществе пост-пандемийного периода. В статье инфографический материал не представлен, но она оформлена рисуночными иллюстрациями<sup>53</sup>. Пандемия COVID-19 негативно повлияла на отрасли, в которых высока доля занятости женщин. Они преимущественно трудятся в неформальном секторе экономики, на сезонных работах и работах с частичной формой занятости, которые значительно пострадали из-за кризиса. Женщины широко представлены в индустрии питания и гостиничного

<sup>53</sup> <https://www.gazeta.uz/ru/2021/03/08/women/>

сервиса (52%), розничной торговле (51%), а также сферах искусства, развлечений и отдыха (45%). В этих сферах наблюдался наибольший уровень сокращения рабочих мест с начала пандемии.

Анализ данных OLX.uz, крупнейшего онлайн-ресурса вакансий в Узбекистане, показывает, что после снижения числа новых объявлений в апреле их количество так и не вернулось на прежний уровень к концу 2020 года. В период с апреля по декабрь секторами с наибольшим снижением вакансий по сравнению с годом ранее стали туризм (-51%); красота, фитнес и спорт (-43%), секретарская и административно-хозяйственная деятельность (-36%), культура и искусство (-30%), а также работа в барах и ресторанах (-17%). Эти отрасли традиционно создают рабочие места для женщин.

В 2020 году в Узбекистане также наблюдался рост домашнего насилия. Вместе с экономическими трудностями, вызванными пандемией, росло и число случаев физического, психологического, экономического и сексуального насилия над женщинами и девочками. По данным Министерства внутренних дел, в период с января по декабрь 2020 года было выдано около 14,8 тысячи охранных ордеров для обеспечения безопасности и интересов жертв гендерного насилия. В 82% из 13230 случаев насилие совершили мужья.

Первым исследованием интернет-издания Anhor.uz с использованием атрибутов ЖД можно считать статью «О трудовой миграции и безработице», опубликованную 2 ноября 2019 года<sup>54</sup>. В ней, помимо анализа данных портала Stat.uz, Министерства занятости и трудовых отношений РУз проанализированы информация Госдепа США за 1990-2016, 2019 гг. об иммиграционных визах (green card), выданных узбекистанцам, а также данные ФМС России. На основе сопоставления этих данных авторы приходят к выводу, что если в ближайшие 5–6 лет в Узбекистане не будет создано минимум 4–5 млн. рабочих мест, то начнется массовый отток из страны экономически активной части населения. Публикация вызвала широкую дискуссию среди общественности. А 22 ноября 2019 года Anhor.uz опубликовала опровержение Министерства занятости и трудовых отношений РУз на данную публикацию. В опровержении также приводится большой массив данных по состоянию занятости населения и рабочим местам. Тем самым был создан прецедент «битвы» открытых данных.

Полноценным дата-проектом Anhor.uz можно считать аналитическую статью «Что мешает развитию медиарынка в Узбекистане»<sup>55</sup>. В ней использованы открытые данные международных организаций International Media Service, KANTAR Research и национальной InstarAnalytics Узбекистан, которые позволили редакции представить следующие визуальные материалы для аудитории<sup>56</sup>. Проанализировав эти данные, автор приходит к выводу, что одним из главных сдерживающих факторов для национального рынка телевизионных медиа была и остается проблема наполнения эфира разнообразным контентом местного производства в необходимом объеме. Телевидение требует производства большого количества медийного продукта. А более 500 часов эфирного времени в сутки заполнить не так-то просто.

Другой формат дата-журналистских материалов в издании представлен в виде расширенной заметки с инфографикой. Так, в дата-заметке «Сколько заключенных трудилось в лагерях Узбекской ССР», наряду с текстом представлена инфографика - диаграмма, в которой отражены машинные подсчеты численности узников лагеря в УзССР в 1930–1943 годы<sup>57</sup>. А в дата-заметке «64% не знают об «этом»: история Марины»<sup>58</sup> инфографика отражает состояние использования контрацептивов в 2017 году в срезе всех вилоятов (областей) Узбекистана.

---

<sup>54</sup> <https://anhor.uz/society/20025-2/>

<sup>55</sup> <https://anhor.uz/vzglyad-iznutri/23708-2/>

<sup>56</sup> <https://anhor.uz/vzglyad-iznutri/23708-2/>

<sup>57</sup> <https://anhor.uz/history/skolyko-zaklyuchennih-trudilosy-v-lagere-uzbekskoy-ssr/>

<sup>58</sup> <https://anhor.uz/society/64-ne-znayut-ob-etom-istoriya-marini/>

В 2020 году Представительство ЮНЕСКО в Узбекистане в рамках проекта «Усиление потенциала узбекских СМИ для обеспечения соблюдения общественных интересов - Фаза II», финансируемого Правительством Великобритании, опубликовало издание «Дата-журналистика. Практический курс для начинающих. Инструментарий» журналистов и преподавателей журналистики на каракалпакском, русском и узбекском языках<sup>59</sup>.

Журналистику данных как отдельную дисциплину в Узбекистане сегодня преподают в Национальном университете Узбекистана (НУУЗ), в Узбекском госуниверситете мировых языков (УзГУМЯ), Каракалпакском госуниверситете, а также в недавно созданном университете журналистики и массовых коммуникаций (УЖМК). Кроме того, журналистику данных осваивают и журналисты-практики в рамках обучающих тренингов Центра переподготовки журналистов. Такие тренинги в республике проводятся при поддержке таких организаций, как ЮНЕСКО, ОБСЕ, Фонд К. Аденауэра и многих других.

Обобщая обзор ЖД в Узбекистане можно заключить, что открытые данные в стране публикуются на узбекском, русском и английском языках на национальном портале открытых данных [data.gov.uz](http://data.gov.uz). Открытые данные представлены в форматах CSV, XML, JSON. Для разработчиков есть также открытое API. На сайте [data.gov.uz](http://data.gov.uz) уже опубликовано 11.918 наборов данных. На портале «Открытый бюджет» также публикуются данные в машиночитаемом виде в разных форматах: XLSX, CSV, JSON. Официальная статистика страны тоже публикуется в формате открытых данных. Не публикуются файлы с данными судебных решений и госзакупок, однако устройство сайтов позволяет собрать записи о них с помощью инструментов веб-скрейпинга или специальных скриптов (роботов-парсеров)<sup>60</sup>. Вместе с тем, для динамичного развития ЖД в странах Центральной Азии необходимо формировать у журналистского корпуса региона новые технологические компетенции, повышать медиаграмотность сотрудников СМИ, смелее инициировать в практику СМИ журналистское расследование, а также в журналистских образовательных центрах открывать отдельные специализации бакалавриата и магистратуры по ЖД.

### **Туркменистан**

Журналистика данных как таковая в Туркменистане пока еще не обозначилась. В последние годы здесь проводятся обучающие тренинги для журналистов и преподавателей факультетов журналистики. Так, в марте 2021 года для журналистов и преподавателей были проведены тренинги в онлайн-формате по таким темам, как новые медиа, цифровая журналистика и медиаграмотность. Программы тренингов также включали в себя сторителлинг, визуализацию данных, особенности работы в социальных сетях, а также медиаграмотность и проверку фактов. Они стали площадкой не только для получения знаний, но и для знакомства и установления контактов между туркменскими и зарубежными медиаспециалистами, журналистами и преподавателями. Данные тренинги прошли в рамках проекта «Развитие новых медиа и цифровой журналистики в Центральной Азии» **Института** по освещению войны и мира (IWPR) при финансовой поддержке Правительства Великобритании.

На официальных сайтах госорганов Республики Туркменистан пока также не удалось найти опубликованных данных в любом из возможных форматов, в том числе неструктурированных (PDF, DOC). При этом агрегированные данные можно найти на сайтах международных баз данных: Knoema 36, Data.World 37.

---

<sup>59</sup> <https://ru.unesco.org/news/yunesko-opublikovala-instrumentariy-po-zhurnalistike-dannyh>

<sup>60</sup> [https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism\\_web.pdf](https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism_web.pdf)

### Основные проблемы развития журналистики данных в Центральной Азии

Оценивая практику и проблемы развития ЖД в Центральной Азии, эксперт А. Валеева, сооснователь "Школы Данных" в Кыргызстане, ментор по ЖД Internews Kyrgyzstan, заключила, что для региона Центральной Азии и географического контекста характерны неполная публикация данных, непостоянная и непрозрачная методология, что, в целом затрудняет, иногда и блокирует работу журналистов данных<sup>61</sup>.

В связи с пандемией, как считает эксперт, мы столкнулись с ситуацией, когда данные вроде бы есть, и их очень много, но, с другой стороны, ясно, что они неполные. Например, очень часто не было данных по смертности. Это заставило журналистов стать более изобретательными и найти другие точки входа, когда нет государственной статистики.

При работе с любыми данными журналисту, по мнению А. Валеевой, надо вырабатывать нюх, изначальную подозрительность, критическое мышление. Журналист должен проверять, откуда идут данные- не только от какого источника, но и как информацию собирали. Важно разбираться в методологии с критической точки зрения, в общем-то, это то, что журналист должен проделать до публикации. Он должен понять, что данные собраны так, что им можно доверять.

Для развития ЖД эксперт считает необходимым обеспечение **коллаборации** с другими сферами. Так, **дата-арт** – направление на стыке искусства и визуализации данных, которое преобразует и преобразует данные, где иногда получается что-то очень красивое. Данными может быть всё, что угодно: панно, мозаика, картина, даже платье, данные можно вшить даже в ткань, то есть очень много попыток очеловечить данные, сделать их более эмоциональными. **Data-humanism** – попытка прочувствовать человека за данными. Этот тренд только усилится в ближайшее время. **Life Logging** – это когда много-много чего в нашей жизни фиксируется, в том числе в телефоне: количество минут, которые мы проводим в приложениях, локации и пр., но можно это делать более осознанно. Можно куда-то записывать количество кружек кофе, к примеру, которые ты выпил, или показатели пробежек утром.

Люди любят персонифицированные данные, по своему городу, локации, где находятся, по стране, либо по тем праздникам, которые они отмечают. Людям всегда интересно посмотреть: «А что у меня на улице? А что у меня в городе происходит?» То есть если выбор – это данные по всей стране или детальные данные по городу, району, то, конечно, чем более близко данные относятся к читателю, тем они ему интереснее.

Между тем, большинство местных СМИ использует подход, основанный на новостях и событиях, с ограниченным количеством аналитики и данных в репортажах и расследованиях<sup>62</sup>. Более того, расследования с использованием ЖД еще не вошли в повседневную практику СМИ стран Центральной Азии. Языковой барьер также затрудняет взаимодействие с мировым сообществом дата-журналистов и продвижение своих навыков и знаний при помощи репрезентации на международных площадках. Хотя именно ЖД может способствовать решению вопросов обеспечения языковой инклюзии по различным направлениям (этнические меньшинства, образование, доступ к информации и т.д.) за счет своего универсального и точного визуального языка и опоры на реальные данные<sup>63</sup>.

В этой связи актуализируются и вопросы подготовки журналистских кадров для работы в сфере ЖД во всем регионе. Проект, инициированный Представительством ЮНЕСКО в Казахстане совместно с кафедрой печати и электронных СМИ КазНУ имени Аль-Фараби, в рамках которого было подготовлено и настоящее издание, направлен на повышение качества образования в сфере

---

<sup>61</sup> Что такое дата-журналистика и почему она нужна всем в Центральной Азии -

<https://newreporter.org/2021/08/31/chto-takoe-data-zhurnalistika-i-pochemu-ona-nuzhna-vsem-v-centralnoj-azii/?ysclid=m3pdpffewu101441639>

<sup>62</sup> <https://cabar.asia/ru/iwpr-central-asia-provel-treningi-dlya-zhurnalistov-i-prepodavatelej-iz-turkmenistana>

<sup>63</sup> 10 ответов о дата-журналистике, которые помогут пиарщику в подготовке контента для СМИ

ЖД в 19 университетах всего Центральноазиатского региона, где осуществляется подготовка профессиональных кадров для СМИ.

### 1.5. Ключевые инструменты журналистики данных

Рабочий процесс создания дата-журналистского материала можно разделить на следующие этапы:

1. Формулирование идеи и гипотезы будущей статьи - определение сути истории и потенциальной роли данных в этой истории.

2. Поиск данных — определение и получение нужных наборов данных для ответов на вопросы — раскрытия темы.

3. Обработка данных для подготовки их к анализу — например, исправление ошибок в наборах данных, переструктурирование данных, сведение разных наборов данных в один набор данных.

4. Анализ данных с помощью подходящих инструментов.

5. Создание истории: текст, визуализации, интерактивные элементы.

6. Публикация материала вместе с методикой расчета и наборами данных, которые использовались в анализе.

Для создания дата-журналистских материалов используется и соответствующий набор инструментов для работы с данными, которые постоянно обновляются. При этом следует помнить, что для создания такого материала предполагаются знания и компетенции в двух направлениях:

- 1) методология и статистика, поскольку используемые данные обычно являются результатом исследований. Выводы из кучи данных возможны только в том случае, если есть первоначальное понимание того, что на самом деле измеряется, значений и переменных, уровней измерения и границ погрешностей. Это позволяет сформулировать гипотезу, которая затем тестируется с использованием набора данных;

- 2) правильное определение интересных данных. Что представляет собой интересное место для поиска данных? Среди наиболее очевидных вариантов - базы данных правительства или организаций, (в том числе, международных), или веб-страницы этих организаций с данными. После того, как источник найден, следующий шаг - получение данных на ваш компьютер и логическая обработка цифр<sup>64</sup>.

Сегодня источники данных становятся все более массивными и, самое главное, разнообразными, что позволяет журналистам получать доступ к новым областям исследований. Администрации публикуют наборы открытых данных на интернет-портале. Все данные на портале доступны в форматах -.XLSX, .DOCX, .JSON и .XML. По словам Тима Бернерса-Ли, журналисты, которые знают свой CSV на RDF, могут создавать быстрые запросы MySQL для получения результатов на PHP или Python. Таким образом, они могут обнаружить скрытую историю в наборах данных, опубликованных правительствами, муниципалитетами, агентствами или любой их комбинацией, даже за пределами государственных границ.

#### Процесс сбора данных

Для сбора данных удобны следующие инструменты, облегчающие их сбор и извлечение (веб-скрейпинг) как из неструктурированных форматов файлов (PDF), так и из веб-страниц:

---

64 Модель учебной программы по журналистике: Сборник новых учебных планов, ЮНЕСКО, 2015.// [https://unesdoc.unesco.org/ark:/48223/pf0000221199\\_rus/PDF/221199rus.pdf.multi.page=51](https://unesdoc.unesco.org/ark:/48223/pf0000221199_rus/PDF/221199rus.pdf.multi.page=51)

- **Tabula** - приложение для извлечения таблиц из файлов PDF и преобразования их в машиночитаемый вид (CSV, XLSX). Однако Tabula работает только с текстовыми PDF-файлами, для работы со сканированными документами — их необходимо сначала распознать с помощью инструментов, в основе которых лежит технология искусственного интеллекта для распознавания изображений и текста (OCR)- [tabula.technology](http://tabula.technology).
- **Table Capture** — расширение для гугл-таблиц (Google Spreadsheets). Извлекает данные на веб-страницах из HTML-таблиц, копирует их в буфер обмена или экспортирует в форматах XLSX (Microsoft Excel), CSV. [georgemike.com](http://georgemike.com)
- **Instant Data Scraper** — автоматизированный инструмент поиска и извлечения данных с веб-страниц. Он использует искусственный интеллект для прогнозирования того, какие данные наиболее актуальны на HTML-странице, и позволяет сохранять их в Excel или CSV-файл (XLS, XLSX, CSV, JSON, XML) - [webrobots.io](http://webrobots.io)
- **Beautiful Soup** — это библиотека Python, разработанная для веб-скрейпинга. Библиотека предоставляет несколько простых методов для навигации, поиска и изменения дерева разбора веб-страниц: набор инструментов для изучения документа и извлечения необходимых данных. Beautiful Soup автоматически преобразует входящие документы в Unicode, а исходящие — в UTF-8 - [crummy.com](http://crummy.com).

#### Процесс подготовки данных к анализу

На данном этапе необходимы следующие инструменты, с помощью которых можно обработать и очистить данные, сделав их пригодными для дальнейшего анализа:

- **Open Refine** — инструмент с открытым исходным кодом для обработки и очистки данных - работает с разными форматами данных: CSV, XML, JSON, RDF/XML, XLSX, XLS и др. Инструмент поддерживает разные режимы работы с данными: ручная проверка и очистка, потоковая построчная обработка с использованием скриптов Python и GREL ([openrefine.org](http://openrefine.org)).
- **Power Query** позволяет работать с данным разных форматов: CSV, XLS(X), JSON, текстовыми файлами, базами данных и внешними API. Надстройка позволяет предобработать данные и загрузить их внутрь таблицы
- Excel в определенном формате и структуре.
- **Проект Natasha** — набор Python-библиотек для обработки текстов на естественном русском языке (Natural Language Processing). Библиотеки решают основные задачи NLP для русского языка: токенирование, сегментация предложения, вставка слова, теги морфологии, лемматизация, нормализация фразы, синтаксический разбор, NER-тегирование, извлечение фактов ([natasha.github.io](http://natasha.github.io)).

#### Процесс анализа данных

На этапе анализа данных **электронные таблицы** — самые популярные инструменты для ввода, хранения, анализа и визуализации данных.

**MS Excel и Google Spreadsheets** - самые популярные, доступные взаимозаменяемые по многим параметрам инструменты для работы с таблицами данных. Главное отличие MS Excel от гугл-таблиц заключается в том, что MS Excel является проприетарным (т.е. находящимся в частной собственности) программным обеспечением, требующим лицензию и плату за использование. Табличный редактор Google Spreadsheets полностью бесплатен в использовании и доступен в онлайн-режиме. Инструментом-аналогом со свободной лицензией является офисный пакет Libreoffice ([libreoffice.org](http://libreoffice.org)). Эти табличные редакторы позволяют подготовить данные к анализу и посчитать нужные показатели. Это могут быть:

- простые арифметические расчеты: сумма, среднее значение, доля и процент;

- расчеты базовых статистических показателей (медиана, мода и др.);
- поиск взаимосвязей в данных (регрессия), корреляций и трендов.

Табличные редакторы удобны для создания сводных таблиц, поиска и объединения данных с помощью функции ВПР (VLOOKUP) и проведения быстрого разведочного анализа данных с помощью встроенных возможностей для создания графиков и диаграмм.

Программа Microsoft Office Excel удобна для работы с электронными таблицами, для выполнения большинства базовых тестов. Вместо Excel есть также свободные пакеты офисных приложений OpenOffice и LibreOffice.

После обучения методам Excel основное внимание уделяется экспорту данных из базы данных веб-или PDF-формата в Excel, с несколькими заданиями по переводу, загрузке или очистке данных из Интернета. Обычно исходные данные, импортируемые в Excel, требуют отбора, чтобы начать анализ.

Простые статистические инструменты в программах электронных таблиц будут достаточными для расчета **средних, процентных различий или корреляций** (см. Глоссарий), в то время как для более продвинутого анализа, понадобится любой статистический пакет, доступный в университетах. Такие инструменты, как SPSS (статистический пакет для социальных наук) или Statistica, обычно используемые в университетах, более подходят, как и R-проект (бесплатный программный пакет с открытым исходным кодом для статистического анализа).

**Microsoft Power BI** — популярный инструмент анализа данных в области бизнес-аналитики, который объединяет программные пакеты: редактор запросов для загрузки и очистки данных, интерфейс для анализа данных, визуализация данных в отчетах. Помимо этого, инструмент позволяет анализировать данные с применением пользовательских моделей машинного обучения и мощных моделей из состава службы Microsoft Cognitive Services. Имеет бесплатный пакет для анализа данных- [powerbi.microsoft.com](http://powerbi.microsoft.com).

**NumPy** — это библиотека Python, добавляющая поддержку больших многомерных массивов и матриц, вместе с большой библиотекой высокоуровневых математических функций для операций с этими массивами ([github.com/numpy](http://github.com/numpy))

**Pandas** — это библиотека Python, которая является мощным инструментом для анализа данных. Пакет дает возможность строить сводные таблицы, выполнять группировки и фильтрацию данных — предоставляет удобный доступ к табличным данным, а при наличии пакета **matplotlib** дает возможность рисовать графики на полученных наборах данных- [github.com/pandas-dev](http://github.com/pandas-dev).

### Процесс визуализации данных

Данный процесс может включать в себя не только создание графиков и диаграмм, но и создание схем, технических иллюстраций и 3D-анимаций. Визуализация данных нужна для того, чтобы самым простым и наглядным образом представить количественные данные или результаты анализа данных для широкой аудитории. Основная задача визуализации - не поразить сложностью рисунков, а передать основную идею, сообщение или рассказать историю, сокрытую в данных. Иногда лучшая визуализация данных — ее отсутствие.

#### Наборы инструментов для выбора типа визуализаций:

- **DataViz Project** — сайт для подбора подходящей визуализации данных, где представ лены все актуальные виды графиков-[datavizproject.com](http://datavizproject.com).
- **Data Visualization Catalogue** - проект, разработанный для создания библиотеки различных типов визуализации информации. Есть перевод на русский язык - [datavizcatalogue.com](http://datavizcatalogue.com).
- **Visual Vocabulary** - список способов визуализации данных, с помощью которых в зависимости от категории взаимосвязи данных можно подобрать тип диаграммы или графика- [public.tableau.com](http://public.tableau.com).

### Книга «Fundamentals of Data Visualization».

В книге описывается наука визуализации данных. Например, в части «От данных к визуализации» описываются различные типы графиков и диаграмм, таких как гистограммы, диаграммы рассеяния или круговые диаграммы, и основные принципы их выбора в зависимости от структуры данных и целей визуализации<sup>65</sup>. Также в книге описываются такие технические моменты, как форматы файлов, используемые для хранения изображений и графиков, приводятся рекомендации по выбору инструмента для визуализации, а также объясняется, как поместить отдельные рисунки в контекст документа - [clauswilke.com](http://clauswilke.com)

**«Графики, которые убеждают всех».** В этой книге описаны универсальные принципы визуализации данных, которые помогут сделать из графиков и диаграмм сильный инструмент для донесения основной мысли вашей аудитории. Автор описывает основные этапы создания визуализации: подготовка данных, выбор графика, оформление, размерность шкалы, заголовки, подписи, аннотации и др.<sup>66</sup>

### Инструменты визуализации данных без программирования и специальных навыков:

- **Datawrapper** — веб-приложение для визуализации данных, доступное каждому. Помогает быстро создавать простые и корректные встраиваемые графики. Для использования Datawrapper не нужно иметь навыков программирования и дизайна. Поддерживает форматы данных: CSV, XLS(X) и другие скопированные текстовые форматы данных. Созданные визуализации вставляются на собственную веб-страницу в интерактивном режиме - [datawrapper.de](http://datawrapper.de) и [datawrapper.de/faq](http://datawrapper.de/faq) (FAQ)29.
- **Flourish** — простой инструмент для создания графиков, карт и интерактивных историй. Он является одним из самых популярных сервисов, с помощью которого можно рассказывать истории с помощью данных. Сервис создан настоящими редакторами и работает в онлайн-режиме. В основе сервиса — десятки готовых шаблонов для мультимедийных историй разных форматов, в которые можно добавить собственные данные. Поддерживаемые форматы данных: CSV, TSV, XLS(X), JSON и GeoJSON. Созданные визуализации вставляются на собственную веб-страницу в интерактивном режиме - [flourish.studio](http://flourish.studio) и [help.flourish.studio](http://help.flourish.studio) (блог).
- **Tableau** - пример более продвинутого инструмента для создания графиков на основе данных. Большинство данных также имеют географический компонент, поэтому добавление данных на карту (ГИС, географические информационные системы) с использованием сервиса Google Fusion Tables - еще один важный инструмент в журналистике данных.
- **Tableau Public** - бесплатная версия инструмента визуализации данных от Tableau, поддерживающая разные форматы данных, а также большой объем загружаемых данных. Позволяет создавать интерактивные графики, карты и аналитические отчеты (дашборды). Созданные визуализации вставляются на собственную веб-страницу в интерактивном режиме. Подходит как для создания простых графиков и диаграмм, не требующих много специальных навыков и знаний, так и для анализа данных и создания сложных визуализаций - [public.tableau.com](http://public.tableau.com) и [tableau.pro/bookofmarathon42](http://tableau.pro/bookofmarathon42) (обучение).
- **RAWGraphs** — веб-приложение с открытым исходным кодом для создания статических визуализаций данных, которые можно модифицировать и адаптировать под собственные

---

<sup>65</sup> Claus O. Wilke. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures, "O'Reilly Media, Inc.", 2019. - 390

<sup>66</sup> Александр Богачев. Графики, которые убеждают всех, 2019. 239 с.

задачи. Поддерживает форматы данных: CSV, TSV, SVG и другие скопированные текстовые форматы из электронных приложений- rawgraphs.io.

### Инструменты визуализации данных с помощью программирования

- **D3.js** — библиотека JavaScript для работы с данными. D3 помогает визуализировать данные с помощью HTML, SVG и CSS. Сочетает в себе мощные компоненты визуализации, основанные на данных. С помощью библиотеки D3JS можно визуализировать данные в разных типах графиков и диаграмм, представлять данные в анимированном формате, создавать интерактивные визуализации. Также D3 включает в себя инструменты для количественного анализа данных, поддерживает форматы данных: JSON, SVG, CSV, TSV, XML, HTML. Для использования необходимы навыки программирования на JavaScript - [d3js.org](https://d3js.org).
- **Observable** — инструмент визуализации данных, позволяющий выполнять исследовательский анализ данных, создавать собственные интерактивные визуализации на основе готовых шаблонов — тысяч блокнотов с кодом. Поддерживает разные форматы данных. Для использования необходимы навыки программирования на JavaScript - [observablehq.com](https://observablehq.com) и [observablehq.com/@anjana/visualizing-open-source](https://observablehq.com/@anjana/visualizing-open-source) (блог).
- **Matplotlib** — это библиотека Python для построения качественных двумерных графиков. Matplotlib является гибким, легко конфигурируемым пакетом, который вместе с NumPy, SciPy и IPython предоставляет возможности, подобные MATLAB - [github.com/matplotlib](https://github.com/matplotlib).
- **Seaborn** — библиотека для создания статистических графиков на Python. Она построена на основе библиотеки matplotlib и тесно интегрируется со структурами данных **pandas**. Seaborn помогает изучить и понять данные. Функции построения графиков работают с наборами данных и выполняют все необходимые преобразования для создания информативных графиков - [seaborn.pydata.org](https://seaborn.pydata.org).
- **Plotly** — библиотека визуализации данных на Python, с помощью которой достаточно легко строить интерактивные графики в Jupyter Notebook - [github.com/plotly](https://github.com/plotly).

## 1.6. Возможности искусственного интеллекта в ЖД

Искусственный интеллект (ИИ) стал ныне реальностью в деятельности СМИ. В силу этого важно понимать и уметь применять основные принципы работы систем на базе ИИ. В отчете информационного агентства Ассошиэйтед пресс (Associated Press, AP) выделены пять областей искусственного интеллекта, имеющих отношение к журналистике: машинное обучение, естественный язык — автоматизированная обработка и генерация текстов, речь — преобразование текста в речь и речи в текст, компьютерное зрение и распознавание изображений, производство видео<sup>67</sup>.

Технологии ИИ способны оказать широкое и глубокое влияние на то, как создается и потребляется журналистика. ИИ может освободить журналистов и помочь им автоматизировать работу с данными, и таким образом значительно сэкономить время для создания качественной публикации. Помимо этого, пока индустрия новостей борется за экономическую устойчивость, доверие и актуальность новостей для общества, ИИ может помочь общественности справиться с перегруженностью новостями и дезинформацией, удобным способом связывая ее с проверенным достоверным контентом.

---

<sup>67</sup> Artificial intelligence at The Associated Press - <https://www.ap.org/solutions/artificial-intelligence/>

История новостного бизнеса – это история данных. Способность собирать и публиковать важную для бизнеса информацию быстрее других была ключевым ценностным предложением с тех пор, как в 1734 году в Лондоне впервые был опубликован **Регистр Ллойда** (Lloyd's List), крупнейшее классификационное общество в мире. Такие информационные агентства, как Bloomberg и Thomson-Reuters, построили империи на своей способности предоставлять рыночные деловые данные читателям. Помимо этого, данные были важным фактором того, почему читатели обращались к новостным изданиям: результаты матчей, погода, результаты выборов, объявления о рождении и смерти, результаты опросов — все это стандартные элементы традиционной печатной газеты.

#### **Функции искусственного интеллекта:**

Автоматизированная проверка фактов - позволяет журналистам быстро проверять публичные заявления или утверждения. Умные компьютеры могут не только анализировать огромные объемы данных для проведения расследований, но и помогать находить источники и проверять факты, чтобы убедиться в их достоверности. Например, агентство Рейтер (Reuters) использует News Tracer для отслеживания последних новостей в социальных сетях и проверки достоверности сообщений в твиттере. Chequeabot используется организацией Chequeado в Аргентине. Это инструмент, который автоматически идентифицирует утверждения в СМИ и сопоставляет их с фактами.

- **Роботы-репортеры (PP)** — самое перспективное направление для автоматизации в новостях. Отслеживая социальные сети и открытые базы данных, инструменты PP могут помочь в сборе новостей в цифровой среде. Вначале PP в СМИ использовались при подготовке спортивных и финансовых новостей. Газета Washington Post использует разработанную собственными силами технологию Heliograf для освещения спортивных событий и политических выборов. Газета Los Angeles Times стала одной из первых, где внедрила автоматизацию в редакции. В «Отчете об убийствах» газеты хранится база данных с информацией о каждом убийстве, о котором сообщил отдел полиции. Профиль каждой жертвы включает краткое автоматически создаваемое описание. Репортеры сами решают, какие истории заслуживают более подробного освещения.
- **Организация рабочего процесса.** К нему относятся: отслеживание срочных новостей, агрегация и систематизация новостей с использованием тегов и ссылок, модерация комментариев. Например, New York Times использует инструмент Perspective API, разработанный компанией Jigsaw (Alphabet), для модерации комментариев читателей.
- **Взаимодействие с аудиторией** – в этом направлении ИИ содействует прогнозированию трендовых тем. Здесь уже используются математические модели, чаще всего построенные с использованием машинного обучения и анализа временных рядов. Таким образом определяются темы статей и выявляются аномалии в данных. К примеру, в блоге «Продукты и технологии» газеты Financial Times опубликована запись об эксперименте, направленном на предсказание трендовых тем. Понимание предпочтений читателей Financial Times имеет решающее значение для улучшения пользовательского опыта и поддержки вовлеченности аудитории. Наличие точных индикаторов, показывающих, какая область становится все более важной, может дополнить работу журналистов, помогая им сосредоточиться на интересующих их темах. Другой пример: Би-Би-Си использовала боты для освещения референдума ЕС. Также она участвует в проектах с открытым исходным кодом и публикует все проекты со ссылками на документацию, на исходный код и информацию о том, как принять в них участие.
- **Анализ больших баз данных.** Агентство Рейтер разработало инструмент искусственного интеллекта Lynx Insight(см.: wired.co.uk), который помогает журналистам анализировать данные, предлагать идеи для статей, при этом не стремится заменить репортеров, а

дополняет их цифровым помощником по изучению данных и написанию текстов. В программе OCCRP «Распознавание образов преступлений» используется технология, которая анализирует большие базы данных документов, ищет в них экономические преступления с коррупционной составляющей и формирует связи (графы) между участниками.

- **Распознавание изображений (OCR)** — технология, позволяющая распознавать объекты, места, лица людей и даже чувства на изображениях. Так, газета New York Times использует инструмент API Recognition от Amazon для идентификации членов конгресса на фотографиях.

## 1.7. Открытость как условие развития журналистики данных: правовой аспект

### Понятие открытости журналистики данных

Журналистика данных предполагает работу журналиста с объективной и достоверной информацией, которая находится, прежде всего, в открытом доступе. В зависимости от темы, проблемы или аспекта ее исследования автор обращается к разным и по объему, и по характеру наборам данных. Так, если это область здравоохранения, то к разработке темы привлекаются открытые данные не только министерства здравоохранения, но и результаты опросов среди населения, врачей, другого медицинского персонала учреждений здравоохранения, их сравнительная статистика в разрезе регионов по стране, в целом.

Само понятие «открытость» несет в себе разные смысловые значения. Одни исследователи под открытостью понимают открытость творчества либо как главное свойство публичного мира. Другие видят в открытости рост доверия к власти со стороны общества и считает, что открытость, прозрачность, транспарентность обеспечивают обратную связь между ними<sup>68</sup>. В контексте ЖД можно выделить также несколько подходов к пониманию открытости. «Открытые данные – это общедоступные электронные информационные ресурсы, представленные в машиночитаемом виде и предназначенные для дальнейшего использования, повторной публикации в неизменном виде»<sup>69</sup>. Существует также обширное понятие открытости, характеризующееся использованием одних открытых данных в комбинации с другими наборами данных. Так, наряду с доступностью, повторным использованием, всеобщим участием, важным свойством является и интероперабельность, то есть «способность различных систем и организаций работать совместно... Возможность выделения компонентов и их «сборки» в единое целое – ключевая для построения больших и сложных систем»<sup>70</sup>. Другими словами, «открытый материал, содержащийся в них, можно беспрепятственно смешать с другим «открытым» материалом. «Такое взаимодействие – ключевой фактор, позволяющий достичь основного практического преимущества «открытости»: значительных улучшений в части комбинирования различных наборов данных, и, как **следствие, возможность разрабатывать больше качественных продуктов и услуг**».<sup>71</sup>

---

<sup>68</sup> Мэннинг Н., Парисон Н. Реформа государственного управления: международный опыт. М., 2003. - С. 80

<sup>69</sup> Современные технологии для оперативного доступа к информации: «Открытое правительство», электронные сервисы обращений, блог-платформа руководителей государственных органов и регионов // <http://info-dostup.kz/lesson10>

<sup>70</sup> Что такое открытые данные? // <https://opendatahandbook.org/guide/ru/what-is-open-data>; The Data Journalism Handbook. Edited by Jonathan Gray, Liliana Bounegru, and Lucy Chambers. (2012).238p.// [https://datajournalismhandbook.org/uploads/first\\_book/DataJournalismHandbook-2012.pdf](https://datajournalismhandbook.org/uploads/first_book/DataJournalismHandbook-2012.pdf)

<sup>71</sup> <https://do-business.kz/open-data?ysclid=m3pfrnebcv356990137>

## **Открытость и право**

Возможность широко использовать и распространять открытые данные вместе с тем не должна допускать дискриминации в смысле применения их в разных областях, в отношении людей или определенных социальных групп. «Главное, что когда мы открываем данные, нас интересуют обезличенные данные, то есть не содержащие информацию об отдельных людях. Использование правительственных данных порой также ограничено из соображений информационной безопасности».<sup>72</sup>

"Во всех сферах жизни технология значительно повысила прозрачность" <sup>73</sup>). Хелен Маргетс также отмечает, что Интернет и Всемирная паутина позволяют государству предоставлять гражданам информацию о своей деятельности, доступ к которой можно получить в любом месте и в любое время <sup>74</sup>. По мнению Бена Уорти, прозрачность может создать благоприятный круг повышения легитимности, демократического участия и доверия, что может стать прорывом, который в конечном итоге приведет к динамичным изменениям в правительстве, политике и даже "конфигурации" нации <sup>75</sup>.

Как уже отмечалось выше, в государствах Центральной Азии наблюдаются различные уровни ЖД. Если в Казахстане, Кыргызстане, и Узбекистане СМИ активно используют представленные в открытых источниках, официальных бюллетенях, отчетных докладах сведения и статистику, то в Таджикистане отдельные журналисты и группы исследователей занимаются сбором, анализом и распространением данных по определенным темам, а в Туркменистане продвижение знаний о ЖД только в зачаточном состоянии.

Прозрачность деятельности - важный инструмент доверия для любого демократического государства. Прозрачность характеризуется максимальной доступностью информации о деятельности органов управления и формированием жесткого процесса общественного влияния на сферу государственного управления.

## **Уровень открытости данных в Казахстане**

Прозрачность в современной журналистской медиапрактике возросла, следовательно растет спрос на полную прозрачность и открытый доступ к источникам информации, особенно в государственных учреждениях Казахстана.

Автор статьи «О доступности и гласности судебной системы» судья Б. Сегизова раскрыла значение принципа доступности в судопроизводстве для установления доверия общества к государству и его структурам. Она пишет «Принципы публичности, открытости, прозрачности гласности, доступности, или, называя все это одним объемным словом – транспарентности судопроизводства является одной из процессуальных гарантий справедливого и демократического правосудия в Казахстане»<sup>76</sup>. Автор также отмечает, что открытость правосудия – это оперативное информирование сторон и заинтересованных лиц. Представители СМИ имеют право знакомиться с содержанием судебных решений и иных документов посредством информационно-справочных систем либо иными способами.

В 2019 году на основе приказа министра информации и общественного развития (МИОР) РК в министерстве юстиции был принят документ «Об утверждении перечня данных, размещаемых на

---

<sup>72</sup> <https://opendatahandbook.org/guide/ru/what-is-open-data/>

<sup>73</sup> What Is Transparency? by R. E. Oliver, Trade Paperback, McGraw-Hill Companies, T.H.E.,

<sup>74</sup> Johannes W. Meijer. A Model for Internet Traffic Growth. Departamento de Ciencias Exactas e Ingeniería, Universidad Católica Boliviana. ACTA NOVA; Vol. 3, № 4, junio 2007. – P. 750-760

<sup>75</sup> The Impact of the Freedom of Information Act on Central Government in the UK, Robert Hazell, Ben Worthy, Mark Glover, Understanding Governance (TRG), 2010

<sup>76</sup> Сегизова Б. О доступности и гласности судебной системы // <https://medialaw.asia/node/9512>

интернет–портале открытых данных». В приказе были отмечено 39 наименований набора открытых данных и сроки размещения о них информации. В их числе подведомственные организации МИОР РК, статистические данные об организации, перечень аккредитованных субъектов частного предпринимательства в сфере информации и общественного развития, количество этнокультурных объединений Ассамблеи народа Казахстана, перечень выданных лицензий по деятельности теле - радиоканалов и др<sup>77</sup>.

Действующая законодательная база создает условия для эффективной и плодотворной деятельности СМИ в использовании различных данных из государственных и других источников. К примеру, в публикации «СМИ Казахстана запустили параллельный подсчет жертв COVID-19» указывается, что семь новостных проектов в Казахстане запустили ресурс Umytra.kz для публикации имен казахстанцев, умерших от коронавирусной инфекции: «В совместном заявлении редакций The Village Kazakhstan, Vlast.kz, «Медиазона Центральная Азия», Factcheck.kz, Радио «Азаттык», «За нами уже выехали» и «Тіл кеспек жоқ» отмечается, что ресурс Umytra.kz (в переводе с казахского «не забывай») запускается, чтобы граждане могли знать реальное положение дел, численность и имена жертв COVID-19»<sup>78</sup>.

Министерство культуры и информации Республики Казахстан ежегодно публикует Годовой отчет о состоянии сферы доступа к информации в Республике Казахстан, в котором представляется оценка соблюдения законодательства в области доступа к информации, включая статистику запросов и нарушения; анализ позиций Казахстана в международных рейтингах открытости; результаты деятельности в сфере международного сотрудничества и внедрения новых инициатив; рекомендации по улучшению доступа к информации, включая законодательные изменения и поддержку людей с инвалидностью. Отчет помогает укреплять принципы открытости, транспарентности и общественного участия<sup>79</sup>.

Казахстан занял 48-е место в мировом рейтинге открытых данных Open Data Inventory (ODIN), поднявшись на шесть позиций с 54-го места в 2020 году. В период 2021–2022 годов общий балл страны увеличился с 57 до 59 по охвату данных и с 65 до 68 по открытости. Рейтинг ODIN, разработанный международной организацией Open Data Watch, оценивает полноту и доступность официальной статистики, способствуя улучшению политики открытых данных и диалогу между национальными статистическими органами и пользователями. Первое место в рейтинге занял Сингапур, за ним следуют Дания и Финляндия<sup>80</sup>.

### **Кыргызстан: лидер открытости в регионе**

В мировых рейтингах открытости и свободы слова Кыргызстан занимает высокие позиции. Правовая база функционирования СМИ позволяет осуществлять дата-проекты самой различной сложности и масштаба. Так, Kloop.kg 15 июля 2020 года опубликовал аналитическую статью «Кольца коррупции» Е.Резниковой. Как кыргызские бизнесмены симитировали конкуренцию на 83 млн. сомов». Относительно цели публикации сказано: «Дата-отдел «Kloop»-а обнаружил, что в госзакупках регулярно участвуют компании, созданные одними и теми же людьми. Они имитируют конкуренцию и выигрывают тендеры без борьбы, ведь кроме них в торгах нет других участников.

---

<sup>77</sup> Об утверждении перечня открытых данных, размещаемых на интернет-портале открытых данных// <https://adilet.zan.kz/rus/docs/V1900019076>

<sup>78</sup> СМИ Казахстана запустили параллельный подсчет жертв COVID-19. // <https://fergana.media › news./13.06.2020/>

<sup>79</sup>

<https://www.gov.kz/memleket/entities/mam/documents/details/678995?lang=ru&ysclid=m3pyct5u9e744457577>

<sup>80</sup> <https://newtimes.kz/obshchestvo/164688-kazahstan-zanyal-48-mesto-v-rejtinge-otkrytyh-dannyh>

Закон запрещает заключать договоры с такими компаниями и называет это «конфликтом интересов»<sup>81</sup>. В рубрике «Тег: дата –журналистика» издание регулярно помещает расследования по актуальной проблематике. «Власти Бишкека ремонтируют больше дорог, чем тротуаров, почему это плохо», «О чем говорят фейки», «Одним постом: почему нельзя отрицать избыточную смертность в Кыргызстане?» и др.

В стране действуют более 200 традиционных СМИ. Издается большое количество газет на нескольких языках.<sup>82</sup> На начало 2020 года, по данным Единого государственного регистра статистических единиц, на территории республики насчитывалось 1,8 тысячи хозяйственных субъектов, зарегистрированных как СМИ<sup>83</sup>. Это даёт большую надежду на прорывное развитие ЖД.

### **Открытость как базовое условие развития ЖД в Узбекистане**

С принятием в стране в 2014 году закона «Об открытости деятельности органов государственной власти и управления» сформировалась правовая база реформирования информационной среды. И как результат этого – был создан портал открытых данных data.gov.uz. «Это создало благоприятные условия для деятельности журналистов, у них появилась практическая правовая основа для сбора цифр, статистических данных, которые ранее были не доступны для представителей СМИ»<sup>84</sup>. В Узбекистане утвержден и перечень открытых данных, подлежащих обязательной публикации госорганами. В 2021 году также вышел указ президента республики «О дополнительных мерах по обеспечению открытости деятельности государственных органов и организаций, а также эффективной реализации общественного контроля», который нацелен на обеспечение максимальной информационной открытости страны<sup>85</sup>.

Возможности ЖД активно используют и журналисты страны, специализирующиеся на экопроблематике. К примеру, элементы и инструментарий ЖД в своей практике активно применяет известный журналист Наргис Касимова, которая и на собственном сайте, и на аккаунте в соцсетях помещает инфографический и табличный материал по таким актуальным вопросам, как сруб деревьев, жестокое обращение с животными, нарушение экосистем столицы и вилоятов. Инфографику используют также журналисты Repost.uz., что наглядно отражено, например, в материале «Инфографика: как выступала сборная Узбекистана на Олимпийских играх с 1996 года»<sup>86</sup>.

Открытые данные удачно использованы интернет-изданием Podrobno.uz в статье «Реанимация в больнице в Загиате переполнена. Почти 20% пациентов в тяжелом и крайне тяжелом состоянии», опубликованной 21 июля 2021 года. Автор уточняет, что реанимация первой Зангиатинской специализированной многопрофильной больницы инфекционных заболеваний переполнена. А телеканалу «Узбекистан 24» заместитель директора больницы Хилола Алимова сказала: "У нас очень серьезная ситуация с COVID-19. Очень много обращающихся, много серьезных тяжелых случаев. Сейчас в нашей клинике лежит 2 тысячи больных, из которых 273 – находятся в реанимации в тяжелом состоянии. Среди них нет ни одного, кто получил вакцину. В целом, за весь

<sup>81</sup> Резникова Е. «Кольца коррупции». Как кыргызские бизнесмены симитировали конкуренцию на 83 млн сомов. (15 июля, 2020, 6 стр.) // <https://kloop.kg/blog/2020/07/15/kolca-korruptcii-goszakupki/>

<sup>82</sup> Исследование СМИ КР. (2013, 37 стр.) // [http://www.media.kg/wp-content/uploads/2013/10/issledovanie\\_media\\_rus2.pdf](http://www.media.kg/wp-content/uploads/2013/10/issledovanie_media_rus2.pdf)

<sup>83</sup> СМИ и книгоиздание: цифры и факты (06.11.2020, 2 стр)..Национальный статистический комитет Кыргызской Республики. // <http://www.stat.kg/ru/news/sredstva-massovoj-informacii-i-knigoizdanie-cifry-i-fakty/>

<sup>84</sup> Журналистика данных в Узбекистане. (14.05.2015, 4 стр.). // <https://data.gov.uz/ru/news/view/21>

<sup>85</sup> Утвержден перечень открытых данных, подлежащих обязательной публикации и. (2021, 17 июня, 11 стр.) // <https://www.gazeta.uz/ru/2021/06/17/publicity/>

<sup>86</sup> Инфографика: как выступала сборная Узбекистана на Олимпийских играх с 1996г. (22.07.2021, 3 стр.). // <https://repost.uz/tokyo2020/olympic-history>

прошедший месяц мы не наблюдали, чтобы люди, прошедшие вакцинацию, попадали в реанимацию»<sup>87</sup>.

### **Таджикистан: через тернии к открытости**

Исследователь Олег Панфилов представил новый ракурс оценки, говоря об уровне свободы СМИ в Таджикистане. Проблем у независимых таджикских масс-медиа, достаточно, пишет он, но в то же время «организация "Репортеры без границ" объявила, что по рейтингу состояния свободы слова Таджикистан занимает первое место среди стран СНГ, это вызвало в республике недоумение и усмешки.

Однако некоторые основания для подобных утверждений всё-таки есть, поскольку власти начинают осознавать, что держать под контролем прессу и не нужно, и опасно. У таджикской прессы есть хороший опыт, ведь до 1992 года выходило огромное количество газет, а с другой стороны, не стоит недооценивать внешнее влияние на Таджикистан, поскольку западные страны, оказывающие республике экономическую помощь и политическую поддержку, постоянно напоминают властям ... о необходимости приводить законодательство и деятельность чиновников в соответствие с международными стандартами»<sup>88</sup>.

Из национальных массмедиа информационное агентство Avesta в своих выступлениях открыто выражает свою, критическую по отношению к чиновникам точку зрения. Так, в публикации «СМИ Таджикистана: самоцензура, или критика должна быть конструктивной» от 24 февраля 2021 года отмечается: «Журналистика в Таджикистане переживает не самые лучшие времена, как в творческом, так и финансовом плане. Особенно это касается независимых СМИ. Хотя официально никакой цензуры в Таджикистане нет, но в последние годы среди журналистов укоренилась понятие самоцензура. Есть ли она (самоцензура) в журналистской среде или нет, отвечают люди, много лет посвятившие этой сфере». И далее приводятся мнения журналистов, руководителей газет, информационных коллективов.

Главный редактор одной из газет Сайфи Мизроб в ответе на заданный ему вопрос говорит: «Я еще раз, как главный редактор независимой газеты подчеркиваю, что никакой цензуры у нас нет. Если критика является конструктивной, с фактами и аргументами, то такая критика только приветствуется». А другой главный редактор - партийной газеты «Минбари халк» Бахтияр Хамдамов считает, что «причиной цензуры становится боязнь самого журналиста или главного редактора, независимо от того является ли СМИ государственным или независимым. Когда журналист пытается донести свои мысли во благо общества, почему он должен бояться? Если говорить о нашей газете, есть вещи, которые мы не можем публиковать. Почему? Потому, что тот, или иной материал вместо пользы, может только навредить обществу, поэтому каждая статья и заметка должна основываться на фактах, а не на домыслах, слухах и предположениях»<sup>89</sup>.

Независимый эксперт Ш.Хакимов, подытоживая мнения коллег по перу, отмечает, что, к сожалению, можно констатировать, что «самоцензура и корпоративные интересы социально-экономического характера и вероятность остаться без работы негативно влияют на качество и содержание работы журналистов, особенно работающих в независимых печатных изданиях и

---

<sup>87</sup> Реанимация в больнице в Загиате переполнена. Почти 20% пациентов - в тяжелом и крайне тяжелом состоянии. (21 июля 2021г, 3 стр) // <https://podrobno.uz/cat/obchestvo/reanimatsiya-v-bolnitse-v-zangiate-perepolnena>

<sup>88</sup> Гибельхауз Людвиг. Таджикистан: СМИ становятся все более независимыми. (11.06.2003, 3 стр.) // <https://www.dw.com/ru/%D1%82%>

<sup>89</sup> СМИ Таджикистана: самоцензура или критика должна быть конструктивной. 24.02.2021 // <http://avesta.tj/2021/02/24/smi-tadzhikistana-samotsenzura-ili-kritika>

других СМИ. В таких условиях ангажированные субъекты информационного пространства по известным причинам и дальше будут заниматься промыванием мозгов и манипуляцией общественного мнения. Независимые СМИ по мере своих возможностей, хотя чрезвычайно трудно, тем не менее, стараются последовательно удовлетворять потребности интеллектуалов и конструктивно мыслящих». Журналисты Таджикистана свои дата-проекты готовят посредством опросов фокус-групп, открытых заявлений в прессе, официальную статистику, а также на данных международных организаций.

#### **Проекты по журналистике данных в Таджикистане**

В Таджикистане по сей день в ВУЗах страны нет специального курса по дата журналистике. По информации преподавателей факультетов журналистики (ТНУ и РТСУ, 2 больших ВУЗа выпускающих журналистов) по своему усмотрению есть преподаватели по специальности журналистики могут включить пару часов в своем курсе о ЖД, однако в связи с тем, что нет учебного пособия, или учебной программы, то данный курс не обучается отдельно. Общественная организация «Хома» является одним из самых первых, кто начали внедрять дата журналистику через тренинги для таджикских журналистов. «Хома» по сей день проводит такие тренинги и на их сайте есть специальный раздел молодых журналистов, которые занимаются ЖД (см.: <https://www.khoma.tj/datajournalism/>)

#### **Туркменистан открывается миру**

До недавнего времени Туркменистан подвергался острой критике за информационную непрозрачность. Однако количество новостных сайтов и аудитория онлайн-СМИ Туркменистана значительно увеличилась с начала глобальной пандемии. Появились такие новостные сайты, как Turkmenistanlive.com, Nexttm.com, Alemjahan.net, Talyplar.com, Business.com.tm, Dashoguz.live, Ashgabat.in, Merw.news, Ahal.info, Balkantm.news, Jeyhun.news. По мнению авторов, это связано с тем, что у людей возникает повышенный интерес к тому, что происходит вокруг, потому что это касается их жизни, здоровья и благополучия»<sup>90</sup>. И это вселяет большую надежду на становление и развитие ЖД здесь.

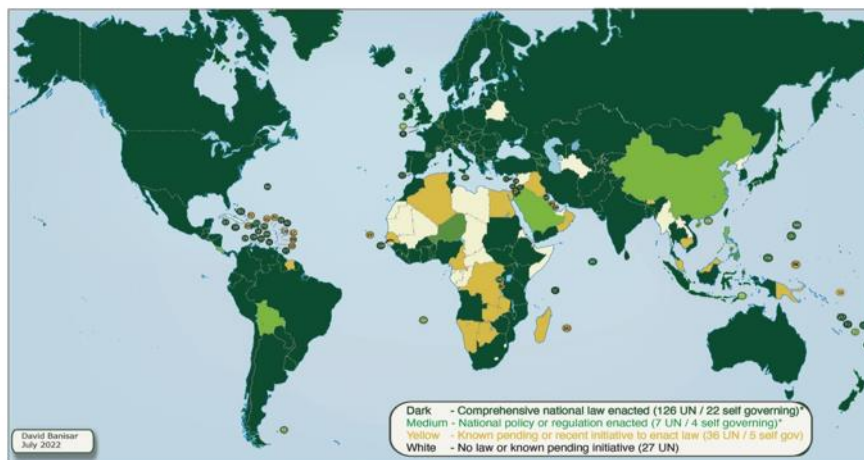
Так, на независимом информационном ресурсе «Хроника Туркменистана» правозащитной организации «Туркменская инициатива по правам человека» систематически публикуются материалы по важнейшим внешнеполитическим вопросам и актуальным событиям внутренней жизни страны. К примеру, в публикации «В выходные дни врачи обходят дома жителей Ашхабада и приглашают на вакцинацию» говорится: «Несмотря на выходные, связанные с празднованием Курбан байрамы, семейные врачи обходят дома жителей Ашхабада и приглашают их на вакцинацию. Об этом сообщают корреспонденты «Хроники Туркменистана». Сейчас вакцинируют людей в возрасте от 38 до 43 лет»<sup>91</sup>. У журналистов Туркменистана, которые обучаются на тренингах по ЖД, есть прекрасный шанс представлять аудитории дата-публикации о самоотверженной работе врачей и всей системы здравоохранении республике в борьбе с коронавирусом.

---

<sup>90</sup> Роль СМИ в обществе Туркменистана. 2стр. // <https://www.osce.org/files/f/documents/a/c/466695.pdf>

<sup>91</sup> В выходные дни врачи обходят дома жителей Ашхабада и приглашают на вакцинацию. —Электронный сайт издания «Хроника Туркменистана. 22 июля 2021. // <https://www.hronikatm.com> › Новости. ХТ

### National Right to Information Laws, Regulations and Initiatives 2022



#### Задания для самостоятельной работы:

1. Изучите СМИ вашей страны, в которых используются атрибуты журналистики данных. Охарактеризуйте жанры или формат дата-публикаций в них.
2. Изучите базы открытых данных в вашей стране. Определите характер информации, представленной наиболее широко и наименее узко в них. Проанализируйте сферы, которые они затрагивают, платные или бесплатные данные содержат, доступно ли скачивание.
3. Ознакомьтесь с материалами по журналистике данных на ведущих сайтах вашей страны. Охарактеризуйте их базовые признаки. Например: публикация..... по следующим признакам является дата-журналистским материалом: в нем использованы открытые данные; статистические данные представлены в сравнении; данные представлены в визуализированном формате.
4. Изучите базы данных крупнейших международных организаций - ООН, ЮНИСЕФ, ЮНФПА и др. Найдите дата-сеты, касающиеся Вашей страны. Выявите основные их характеристики.
5. Изучите гендерные аспекты публикаций журналистики данных в СМИ вашей страны. Определите локальные признаки гендерной проблематики, которая в них отражена. Выявите жанры данных публикаций и инструменты журналистики данных, которые в них использованы.
6. Изучите выступления СМИ вашей страны в срезе ЖД по вопросам социальной защиты уязвимых слоев населения, в том числе и проблем людей с инвалидностью. Выявите локальные аспекты данной проблематики.
7. Изучите материалы ЖД по вопросам национальных меньшинств в СМИ вашей страны. Определите актуальные аспекты данной проблематики.
8. Охарактеризуйте инструменты визуализации, которые использованы в материалах ЖД в СМИ вашей страны. Охарактеризуйте их влияние на качество публикаций в конкретных СМИ
9. Сравните разные периоды деятельности СМИ вашей страны, которые работают в направлении ЖД. Сравните типы визуализации, которые использовались ранее и используются сегодня.

#### Вопросы для самоконтроля

1. В чем заключается сущность и актуальность ЖД?
2. Как ЖД решает вопросы социальной инклюзии?
3. Как ЖД может содействовать обеспечению гендерного равенства в обществе?
4. Какими базовыми компетенциями должен обладать дата-журналист?

5. Каковы базовые различия между ЖД и БДЖ?
6. Охарактеризуйте основные этапы становления и развития ЖД в мировых СМИ.
7. Охарактеризуйте состояние ЖД в США, Германии, Франции, Великобритании, России, Китае.
8. Какие позитивные достижения в направлении ЖД есть у СМИ и журналистов Центральной Азии?
9. Какие барьеры и проблемы мешают ЖД активно развиваться в регионе Центральной Азии, в вашей стране?
10. Какие ключевые инструменты ЖД знаете?
11. Какие задачи решают инструменты ЖД на разных этапах подготовки дата-публикации?
12. Как и активно ли используются возможности ИИ в деятельности СМИ региона и вашей страны?
13. Как Вы оцениваете уровень открытости управления вашей страны? Какие факторы влияют на уровень открытости общества?
14. Какая взаимосвязь между открытостью СМИ и использованием открытых данных в сфере информации?
15. Определите источники и специфику использования данных для дата-публикаций в массмедиа вашей страны.

## Модуль 2. ПРИНЦИПЫ ИССЛЕДОВАНИЯ И СТАТИСТИКА

### 2.1. Стратегии исследований

Как это уже неоднократно подчеркивалось, журналистика данных по своей природе ближе к научным исследованиям, чем к чистой журналистике. Поэтому работа в этой области предполагает глубокие познания в различных методологиях по сбору данных и практические навыки в их статистическом анализе.

Сбор данных основан на двух видах методологий, или стратегиях исследований, - количественных и качественных. И выбор определенной методологии определяется тем, какой вид данных является доступным. Количественные стратегии исследования базируются на статистических выводах, которые выражаются в числах. Качественные стратегии в свою очередь основаны на нестатистических данных, которые собраны путем интервью, анализа всевозможных архивов и обзора литературы.

Количественная методология предполагает обработку большого количества однотипных данных, которые легко поддаются статистическому анализу. Такие данные обычно нацелены на выявление определенных трендов или тенденции распространения тех или иных явлений в отдельно взятой выборке. При таком анализе данных применяется дедуктивная логика, что предполагает выявление связей между различными явлениями, экстраполируя выводы из общих суждений к частным. Поэтому массовый опрос или контент-анализ публикаций стали наиболее распространенными методиками количественных исследований.

Качественная же методология предполагает всестороннее изучение одного или нескольких объектов, например, интервью участников того или иного события. После этого делается обобщение на всю группу, исходя из частных выводов, то есть - так называемая индуктивная логика. Однако из-за таких различий между количественными и качественными методологиями исследований, не стоит делать поспешных выводов, что в журналистике данных нет места качественным исследованиям. Многие результаты сбора качественных данных поддаются кодировке в количественные, и уже затем можно производить статистический анализ. К примеру, данные литературного обзора в частотный анализ лексики в публикациях на заданную тему и определенных авторов.

В журналистике данных особенно важно первоначальное определение объекта исследований, значений и переменных, а также уровня измерения в каждом конкретном случае и границ погрешности. Это позволяет сделать выводы из множества разрозненных данных и сформулировать гипотезу, которую можно уже проверить на наборе информации. И в зависимости от сформированной гипотезы, целей и задач исследования можно применить широкий спектр методов статистической обработки данных, таких как вычисление средних величин в основном массиве данных, установление частоты появления различных признаков в целом и в отдельно взятых группах, построение таблиц и гистограмм, установление коэффициентов тесноты связи между разрозненными признаками, наличие таких типов как кластеры или таксоны среди множества объектов и т.д.<sup>92</sup>

---

<sup>92</sup> <https://www.computerscience.org/resources/women-in-computer-science/>

## 2.2. Количественные методы исследования в журналистике

Количественные стратегии принято называть жесткими, потому что их сутью? является разработка и применение математических моделей, теорий и гипотез, относящихся к явлениям. В журналистике количественные исследования стимулировались интересом социологов к СМИ как социальному институту. За последние несколько лет, это направление журналистики, основанное на количественных данных, стало более заметным в профессии, поскольку оно сошло со все более повсеместной оцифровкой информации, как личной, так и общедоступной.

На этапе сбора данных, из-за их природы, осуществляется их измерение, установление наличия или отсутствия признаков, определенных задачами исследования или величины и частоты их проявления. В количественных стратегиях необходима апробация инструментов еще до начала полевого сбора данных, потому что он не может изменен на всем протяжении этого этапа.

Преимуществом количественных методов безусловно является охват большого количества объектов исследования. И их обобщение приводит к получению статистических показателей. По этим причинам результаты любой конкретной исследовательской деятельности должны быть помещены в контекст, в котором были построены выводы. В целом вместо того, чтобы изображать полную правду, исследования, статистика, в частности, раскрывает часть лишь изучаемого явления.

Задачами количественных исследований являются:

- Точные количественные описания объекта
- Сегментация, построение профайлов данных
- Отвечает на вопрос «Сколько?»
- Поиск закономерностей, проверка гипотез, прогнозирование

Таблица № 1. Работа с гипотезами количественных исследований

| Этап                     | Обоснование                                                                                                                                                                                               | Вопросы для формулировки                                                                                                   |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Формулировка гипотезы    | Качественный анализ помогает собрать подробную информацию по теме. Подходит, чтобы исследование вскрывало? Отражало? проблемы или возможности, о которых думают люди.                                     | Могут ли эти идеи стать гипотезами, которые будут доказаны посредством количественного исследования?                       |
| Подтверждение гипотезы   | Количественное исследование обеспечивает цифровые данные, к которым можно применить статистический анализ, чтобы подтвердить гипотезы. Полученные объективные факты позволяют принять правильное решение. | Имеет ли проблема реальную подоплеку или она возникла в результате чьего-то неверного восприятия?                          |
| Нахождение общих ответов | В количественном исследовании обычно принимают участие больше респондентов, чем в качественном, так как значительно легче проводить опросы с множественными                                               | Из-за чего возникает проблема? Имеют ли разрозненные проблемы единое решение? Какое объяснение является самым оптимальным? |

|  |                                                                                                                                               |  |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------|--|
|  | вариантами ответа, чем серии интервью или фокус-группы. Таким образом, это может помочь получить определенные ответы на широкий ряд вопросов. |  |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------|--|

### 2.3. Как кодировать данные качественных исследований?

В зависимости от видов источников и их степени формализованности, которые разнятся от неструктурированных записей интервью до таблиц, в качественных исследованиях используются всевозможные типы данных. Главная сложность заключается в их обобщении и кодировке. Конечно, анализ данных начинается уже на этапе первоначального сбора, что чаще всего происходит вручную. Затем происходит все более высокий уровень их обобщения, который выражается несколькими видами кодирования данных. Исследователь Л. Ньюман определил три основных вида кодирования: открытое, осевое и выборочное<sup>93</sup>. Эти три этапа можно представить в виде инверсивной пирамиды.

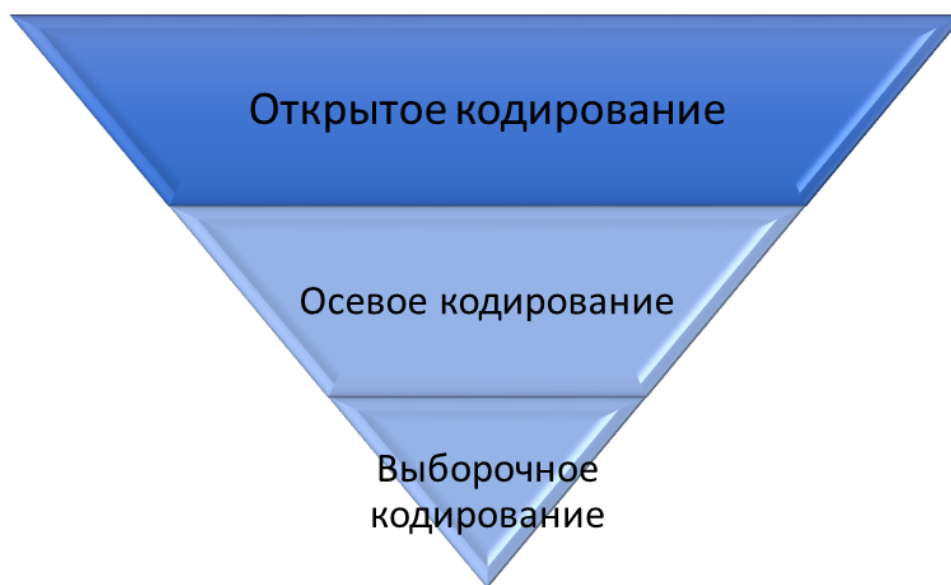


Рисунок № 7. Кодирование данных качественных исследований

Открытое кодирование предполагает маркировку тем в основном массиве при первом ознакомлении с данными. После первичного кодирования составляется общий список выявленных тем, что служит ориентацией и стимулом к дальнейшей конкретизации. Дальше выявленным темам присваиваются кодировочные обозначения, они могут быть в виде символов или цифр. Причем они могут обозначать разные величины, от отдельно взятых слов до больших описательных фрагментов.

Осевое кодирование – это следующий этап обобщения данных, при котором происходит повторный проход сквозь собранную информацию. Однако при этом происходит работа не непосредственно с первоначальными данными, а с обозначенными кодами. Целью осевого

<sup>93</sup> Л. Ньюман. Исследования СМИ: методология, подходы, методы. М: МГУ; 2011 - 236 с.

кодирования является выявление трендов или повторяющихся сочетаний между первоначальными кодами. Для оптимизации процесса анализа информации можно применить программы для анализа данных, например, тот же Excel. При осевом кодировании устанавливаются логическое отношение между центральной темой исследования и ее подтверждениями, разбросанными по разным отрывкам первичного текста.

Выборочное кодирование является заключительным этапом обработки качественных данных и последним проходом через массив информации. При выборочном кодировании определяются основные темы для анализа, их формулировки и выводов. Выборочно отыскиваются подтемы или эпизоды, которые наиболее ярко иллюстрируют основную тему, одновременно осуществляется сравнение или противопоставление разных отрывков текста.

## **2.4. Основной алгоритм исследований в журналистике данных**

В то время, когда источники становятся цифровыми, журналисты могут и должны быть ближе к этим источникам. Интернет открыл возможности, выходящие за рамки нашего нынешнего понимания. Журналистика данных – это только начало развития наших прошлых практик для адаптации к онлайн.

На сегодняшний день журналисты и общественность теперь гораздо лучше знакомы с данными, но и данные продолжают все быстрее распространяться: компьютерные технологии и автоматизация создали мир точных, записанных данных, которых раньше просто не существовало. Важно отметить, что сегодня компьютерные инструменты проще, ПК работают быстрее. Поэтому каждый день миллионы и миллионы строк данных поступают со всех уголков страны и мира.

Когда речь идет о проектах с использованием больших массивов данных, есть несколько ключевых моментов в алгоритме обработки информации, на которых мы остановимся подробнее:

1. Источники и тип данных
2. Понимание принципов основных API (application programming interfaces, или интерфейсов прикладного программирования).
3. Извлечение данных с сайтов при отсутствии доступа к структурированным данным, так называемый скрейпинг.
4. Основные инструменты для обработки данных.

## **2.5. Источники и тип данных**

Расследования в журналистике данных начинаются с новых доступных ресурсов, таких как программное обеспечение с открытым исходным кодом, публикации с открытым доступом и открытые данные, в то время как другие являются продуктами запросов на общедоступные записи или просочившихся материалов. Этот подход к журналистике основан на старых методах работы, в первую очередь на computer-assisted reporting (CAR). При этом очень важна оптимизация поиска информации.

Самым важным типом данных в ЖД является машиночитаемые данные (machine-readable data). Основная характеристика таких данных – это то, что они создаются для обработки специальными программами, а не для чтения пользователем. Структура таких данных относится к содержащейся в них информации, а не к способу ее отображения в конечном итоге. Примеры легко машиночитаемых форматов включают файлы CSV, XML, JSON и Excel, тогда как такие форматы, как документы Word, HTML-страницы и файлы PDF, больше ориентированы на визуальную компоновку

информации. Например, PDF — это язык, который напрямую общается с вашим принтером; он связан с положением линий и точек на странице, а не с различными символами.

Даже если на первый взгляд кажется, что многие базы данных почти невозможно найти, многие из них индексируются обычными поисковыми системами. Ниже приведены несколько распространенных советов, при поиске базы данных.

Одним из важных условий успешного поиска данных являются точные условия поиска. Нужно убедиться, что вписано как условия поиска, относящиеся к содержанию данных, которые вы пытаетесь найти, так и некоторую информацию о формате или источнике, в котором вы ожидаете, что они будут. Всем известно, что Google и другие поисковые системы позволяют выполнять поиск по типу файла. К примеру, можно выполнить поиск только по электронным таблицам, добавив к запросу «filetype: XLS filetype: CSV», по геоданным («filetype: shp») или выдержкам из базы данных («filetype: MDB, filetype: SQL, filetype: DB »). А добавление в поисковый запрос «filetype: pdf» поможет найти нужные документы в формате PDF.

Также возможно выполнить поиск только по одной части URL-адреса. К примеру, поиск в Google по такому запросу как «inurl: downloads filetype: xls» попытается найти все файлы Excel, в которых есть слово «downloads» (загрузки) в своем веб-адресе. Важно помнить, что если вы найдете одну загрузку, часто стоит просто проверить, какие еще результаты существуют для той же папки на веб-сервере. Равным образом можно ограничить свой поиск только результатами по одному доменному имени, например, выполнив поиск по запросу «site: agency.gov».

Еще одним популярным методом оптимизации поиска является поиск контента не напрямую, а локализация определенного места, где могут быть доступны большие объемы данных. Например, «site: agency.gov Directory Listing» может предоставить некоторые списки, созданные веб-сервером с легким доступом к необработанным файлам, в то время как «site: agency.gov Database Download» будет искать специально созданные списки<sup>94</sup>.

#### **Международные базы данных:**

1. Портал данных Мирового банка <https://data.worldbank.org>
2. Портал данных ООН <https://data.un.org>
3. Data.world <https://data.world>
4. Azure Open Datasets <https://azure.microsoft.com/en-us/services/open-datasets/>
5. Национальные базы данных разных стран:
6. Hubofdata.ru (Россия)
7. Data.gov (США)
8. Data.gov.uk (Великобритания)

#### **Национальные базы данных Казахстана:**

<https://data.egov.kz/> - Портал открытых данных электронного правительства Республики Казахстан.

<https://stat.gov.kz/> - Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан.

<https://zan.kz/> - База данных законодательства РК, кодексов и законов в последней редакции.

<http://www.elibrary.kz/> - Единая электронная библиотека Казахстана.

#### **Национальные базы данных Узбекистана:**

<https://data.gov.uz/> - Портал открытых данных Республики Узбекистан

<https://huquqiportal.uz/ru/legislation> - Национальная база данных законодательства Республики Узбекистан

---

<sup>94</sup> [https://datajournalismhandbook.org/uploads/first\\_book/DataJournalismHandbook-2012.pdf](https://datajournalismhandbook.org/uploads/first_book/DataJournalismHandbook-2012.pdf)

<https://lex.uz/ru/> - Национальная база данных законодательства

**Национальные базы данных Кыргызстана:**

<http://www.stat.kg/ru/> - Национальный статистический комитет Кыргызской Республики.

<https://e.srs.kg/ru/> - Информационный портал Государственной Регистрационной Службы.

<http://minjust.gov.kg/ru/content/824> - Электронная база данных юридических лиц, филиалов (представительств) и СМИ.

<http://www.kyrgyzstanvsl.org/> - Кыргызская Виртуальная научная библиотека (ВНБ).

**Национальные базы данных Таджикистана:**

<https://www.stat.tj/> - Агентство по статистике при Президенте Республики Таджикистан.

<https://www.stat.tj/ru/gender-database> - Динамические ряды гендерных индикаторов к Стратегии активизации роли женщин в РТ.

<http://ncz.tj/legislation> - База Законодательства Республики Таджикистан.

<https://www.namsb.tj/> - Национальная Ассоциация Малого и Среднего Бизнеса Республики Таджикистан.

**Национальные базы данных Туркменистана:**

<http://www.stat.gov.tm/ru/> - Государственный комитет Туркменистана по статистике.

<http://www.turkmenlegaldatabase.info/ru> - База данных законодательства Туркменистана.

## 2.6. Основные принципы API

Интерфейсы прикладного программирования (API) – это описание разных способов таких как набор классов, процедур, функций, структур или констант, которыми одна компьютерная программа может взаимодействовать с другой программой. Другими словами, API предоставляют платформу и среду для приложений, которые позволяют им общаться и понимать друг друга. Многие API имеют надлежащую документацию, которая также описывает характер и структуру ответа, который они отправляют, когда делается определенный запрос. Они также указывают необходимую информацию, которую запрашивающее приложение должно предоставить для успешного запроса к API<sup>95</sup>.

При определенных навыках можно создать собственные API, направленные на определенные задачи. Эти задачи зависят от конкретной базы данных, что является распространенной практикой в таких больших медиа-организациях как The Guardian или New York Times. Создание собственных API это существенный этап при освоении журналистики данных, но процесс их освоения может кропотливым и занимать много времени. Поэтому чаще всего целесообразней использовать уже созданные API или готовые шаблоны для их создания. Эти шаблоны имеют свою историю, разные требования и создают разные возможности для пользователей.

## 2.7. Скрейпинг

Скрейпинг (Web/ screen scraping) это процесс извлечения структурированного контента с обычной веб-страницы с помощью специальной утилиты или путем написания небольшого

---

<sup>95</sup> <https://sebwco.com/ru/ponimanie-osnov-raboty-api-i-rest-api-kratkoe-vvedenie/>

фрагмента кода. При этом журналистам необязательно писать код самим. Хорошим и весьма распространенным инструментом для скрейпинга является ScraperWiki, сайт, где журналисты могут обращаться к программистам за помощью в извлечении данных с веб-сайтов.

Investigative Dashboard содержит список готовых инструментов, которые могут помочь журналистам собирать, формировать и анализировать данные:

- Readability (<http://www.readability.com/>) – предназначен для извлечения текстов с вебстраниц.

- DownThemAll (<http://www.downthemall.net/>) – позволяет загружать сразу множество файлов одновременно, это помогает автоматизировать некоторые утомительные задачи.

- Расширение Chromescraper – помогает извлекать таблицы с веб-сайтов.

- Расширения разработчика FireBug (<http://getfirebug.com/>; для Firefox - то же самое уже включено в Chrome, Safari и IE) позволяет точно отслеживать структуру веб-сайта и обмен данными между вашим браузером и сервером.

- Кстати, важно помнить, что Википедия позволяет экспортировать данные в машиночитаемом формате. Можно использовать следующие ресурсы:

- <http://ru.wikipedia.org/w/api.php>

- Wikidata.org/Freebase.com, этот сервис принадлежит компании Google

«Викиданные (Wikidata) – это свободная, совместно наполняемая, многоязычная, вторичная база данных, в которой собрана структурированная информация с Википедии, Wikimedia Commons, а также других вики-проектов движения Викимедиа» (<https://www.wikidata.org/wiki/Wikidata:Introduction>), причем количество подключенных баз данных на сегодняшний день превышает 40 тысяч. Викиданные предоставляют большие возможности для дата-журналистов, предоставляя открытые данные с различных источников, причем система представления данных является независимой от языка. Связывание воедино данных позволяет выявить скрытые корреляции, привести к обнаружению новых фактов и закономерностей. Хранилище викиданных состоит в основном из элементов, каждый из которых имеет метку, описание и некоторое количество синонимов. Элементы имеют свой уникальный идентификатор. Каждый элемент сопровождается утверждением, который состоит из указания свойства элемента и его значения. Ссылки на сайты (Sitelinks) связывают каждый элемент с соответствующими ему статьями во всех вики-ресурсах. Особенностью викиданных является их актуальность, так как связывание данных происходит на основе информации, представленной в различных регулярно обновляемых базах данных. Викиданные позволяют выявить и установить множество связей между точками данных и взаимосвязей с другими открытыми наборами данных. Существует ряд способов обращения к викиданным с помощью встроенных средств, внешних инструментов например, таких как Wikidata Query и Reasonator, или программных интерфейсов Wikidata API.

- Dbpedia.org – один из наиболее гибких сервисов, но непрост в использовании для новичков.

**Почему иногда скрейпинг невозможен?** Важно помнить, что не всякая информация может быть пригодна для скрейпинга. Например, отформатированный HTML-код, который часто встречается на старых правительственных вебсайтах, в таком коде обычно встречается мало структурной информации для обработки. На самом сайте может отсутствовать возможность поиска по шаблону, что может привести к неполному списку данных. Также иногда администраторы сервера могут блокировать массовый доступ на сайт. В частности, такие системы аутентификации как коды CAPTCHA или платный доступ, которые должны предотвращать автоматический доступ, делают скрейпинг невозможным.

Существует еще такой аспект ограничений в сборе данных, как правовые барьеры. Есть ряд стран, в которых признаются права на использование базы данных, что может быть серьезной помехой для повторного использования информации, даже несмотря на то, что она опубликована в Интернете. Коммерческие организации и даже некоторые НПО, в отличие от свободно доступных государственных баз данных, могут рассматривать широкое извлечение информации с их сайтов как посягательство на коммерческую тайну и их интеллектуальную собственность. А скрейпинг социальных сетей может восприниматься как нарушение и проникновение в частную жизнь людей без их прямого согласия, тем самым нарушая законы о конфиденциальности данных или профессиональную этику журналиста. Впрочем, правовые и этические барьеры актуальны не только в журналистике данных, но и в традиционной журналистике тоже.

## 2.8. Основные инструменты для обработки данных

В журналистике данных необязательно иметь степень специалиста по статистике, чтобы эффективно обрабатывать данные. Как выразился Герд Гигеренцер, профессор Института Макса Планка, «лучшие инструменты не приведут к лучшей журналистике, если они не будут использоваться с пониманием».

Базовым инструментом любого аналитика, работающего с данными, является Excel, главный инструмент перед импортом данных в другие инструменты. Программа предоставляет возможности экономико-статистических расчетов, графические инструменты для визуализации данных и возможности макропрограммирования VBA (Visual Basic for Applications). Она создана для Microsoft Windows, и Mac OS, а также Android и iOS, чтобы подготавливать, обрабатывать и визуализировать однотипные данные. Большинство всей работы в дата-журналистике делается в Excel. Эта программа ответственна за основные операции по импорту и экспорту данных, их очистке, сортировке и структурированию.

К остальным популярным программам для обработки данных относятся:

- IBM SPSS – мощный пакет, способный справиться с абсолютным большинством статистических задач. Является платным, однако существует и бесплатная 14-дневная версия.
- StatSoft Statistica – главный конкурент SPSS на отечественном рынке. Также является коммерческим продуктом.
- R-commander – графический интерфейс для языка программирования R. Как и сам R, распространяется бесплатно.
- PSPP – бесплатный аналог SPSS со схожим интерфейсом.

### Базовые этапы работы с данными

В проектах с большим количеством данных необходимо иметь в виду, как минимум три ключевых понятия:

1. Запросы данных должны начинаться со списка вопросов, на которые вы хотите получить ответ.
2. Данные часто беспорядочные, и их необходимо “очистить”.
3. Данные могут иметь недокументированные особенности.

**Как очищать данные?** Одна из самых больших проблем в работе с базами данных в ЖД заключается в том, что вы часто будете использовать данные для анализа, которые были собраны для бюрократических целей. Проблема в том, что стандарты точности могут различаться для разных данных. Например, некоторые слова могут быть неправильно написаны или цифры могут перепутаться.

К сожалению, такие ошибки могут помешать журналистам обнаружить закономерности в базе данных. По этой причине первая большая работа, которую нужно предпринять при получении нового набора данных, это проверить, насколько данные находятся в беспорядке, а затем очистить их. Хороший быстрый способ найти беспорядок - создать частотные таблицы категориальных переменных, которые, как ожидается, будут иметь относительно небольшое количество различных значений. Например, при использовании Excel это можно сделать с помощью фильтров или сводных таблиц для каждой категории.

## 2.9. Статистика в журналистике

Часто возникают дискуссии о том, как долго существует сама «журналистика данных», потому что статистические данные используются в журналистике с самого ее зарождения. Если воспользоваться самым широким определением из числа стандартных определений, из которых следует, что дата-журналистика — это процесс сообщения фактов, в основе которого лежат структурированные данные, - то данный вид журналистики перед свое начало в XIX веке. Как отмечает Саймон Роджерс, автор книги «Священные факты», создатель и редактор блога «Data blog» в «The Guardian», первое применение журналистики данных в газете Guardian еще в 1821 году - предоставление таблицы школ в Манчестере (с указанием количества учащихся и затрат на школу) - помогло показать реальное количество учеников, получающих бесплатное образование, которое было намного выше официальных цифр. Данный пример показывает, что навыки, необходимые для сообщения данной новости, заключаются в использовании журналистами базы данных и статистики.

| <p>South end of Edgehill. In its progress it knocked down several workmen, one of whom was so violently affected in the back of the head, that for relief he had recourse to alcohol; and a girl in a garden had her arm suddenly lifted up by its effect, and felt it bound for some time after. The coachman of Mr. Huff was struck on the arm whilst on the box, but was merely stunned. A lady near Ilkington, who was sewing at the time, felt in the fingers in contact with the needle a sensation resembling that of a slight electric spark. The electric fluid entered the nose of Mrs. Clare, in Edge Vale, where its progress was not less alarming; to the inmates, thus destructive to the premises; and we have never heard of a more surprising escape, than that of the several individuals dispersed in a house, of which almost every room bears testimony of the ravages of the uncontrollable element. It appears probably, from an examination of the apertures which the fluid has made, and the direction in which the bricks, timber, &amp;c. have been forced, that, attracted by the iron railing in front of the house, it entered the wall on one side the door, where it has shattered the bricks, torn to pieces the wood and brick-work between the door-sill and the arch-way of the door; lifted the boards on the top; shook the fau-sight to pieces, burning part of the frame, and leaving a black soot on the paint-work; thence it passed up through the arch door, splitting the bricks and the stone at the bottom of the middle window, the glass of which was shattered to pieces, and the whole frame dislodged and forced into the house. Over the window it forced, in its way to the roof, a large hole, above which the soot appears exactly as if flames had issued from it. Its course appears next to have been towards the chimney: the eaves were shattered to pieces; the ridge stones displaced; many bricks and much cement torn from the wall; and the lead in many places forced up. It probably reached the rooms below by the chimneys. In the lower room the stoves, plaster and paper are in several places broken, and the hair, as if searching its way out, crept the gilding of the chimney glass, and scorching the top ornaments, but did not disturb the polished fire-iron, just below. Six squares of glass were driven out in this room. In the room above, on one corner, stood a bundle of rods, to which it made its way, perhaps from the chimney, between the railing and the wall, as it forced off the plaster, and shot a quantity of it against a chest of drawers, eleven or twelve feet distant, evidently with</p> | DAY SCHOOLS.— <i>Establishment</i> |       |       |           | Remarks.                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|-------|-------|-----------|-----------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Boys                               | Girls | Total | Ave. Exp. |                                                                                                           |
| Grammar School .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | 155                                | 155   | 310   | 1800      | Taught, clothed and boarded. And offertory money: do. do.                                                 |
| Blue Coat ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 80                                 | 80    | 160   | 2000      |                                                                                                           |
| Green Coat ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 50                                 | 50    | 100   | 250       |                                                                                                           |
| Collegiate Church ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 50                                 | 50    | 100   | 40        |                                                                                                           |
| Strangeways ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 10                                 | 10    | 20    | 100       | (Supposed)—Taught and clothed.                                                                            |
| St. Mary's ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 12                                 | 12    | 24    | 40        | Funds arising from Sacramental Offerings.                                                                 |
| St. John's ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 9                                  | 9     | 18    | 20        | (Supposed)—Expenses raised by voluntary Subscription.                                                     |
| St. Paul's ditto .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 30                                 | 30    | 60    | 250       | Taught, clothed and boarded, by voluntary Subscription.                                                   |
| Ladies' Jubilee .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 30                                 | 30    | 60    | 250       | (Supposed)—Taught and partly clothed. This School is supported by the benevolence of a single individual. |
| Back King-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | 21                                 | 21    | 42    | 40        | Voluntary Subscription, and Collections at Churches.                                                      |
| NATIONAL SCHOOLS, Granby-row .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 191                                | 119   | 310   | 600       |                                                                                                           |
| Bulton-street, Saltaire .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 300                                | 170   | 470   | 5         |                                                                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 851                                | 381   | 1232  | £5410     |                                                                                                           |
| <i>Disasters.</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                    |       |       |           |                                                                                                           |
| LANCASTERIAN SCHOOLS, Marshall-st.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 692                                | 225   | 917   | 400       | Voluntary Subscription.                                                                                   |
| UNITARIAN, Mosley-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 35                                 | 35    | 70    | 50        | Ditto ditto                                                                                               |
| CATHOLIC .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 198                                | 151   | 349   | 104       | Ditto ditto                                                                                               |
| SUNDAY SCHOOLS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 890                                | 381   | 1271  | £554      |                                                                                                           |
| Establishment .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 201                                | 205   | 406   |           |                                                                                                           |
| St. Ann's, Back King-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 50                                 | 50    | 100   |           |                                                                                                           |
| St. Mary's, Back South Parade .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 130                                | 110   | 240   |           |                                                                                                           |
| St. Paul's, Green-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 170                                | 163   | 333   |           |                                                                                                           |
| Turner-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 68                                 | 71    | 139   |           |                                                                                                           |
| Jenny-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 314                                | 281   | 595   |           |                                                                                                           |
| St. George's, St. George's .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 141                                | 112   | 253   |           |                                                                                                           |
| St. John's, St. John's-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 118                                | 163   | 281   |           |                                                                                                           |
| St. James's, St. James's-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 102                                | 198   | 300   |           |                                                                                                           |
| St. Michael's, Miller-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 234                                | 352   | 586   | £1023     |                                                                                                           |
| St. Peter's, Jackson's-row .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 120                                | 120   | 240   |           |                                                                                                           |
| St. Clement's and St. Luke's, Alport Town .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 90                                 | 90    | 180   |           |                                                                                                           |
| Beaumont-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 335                                | 1071  | 1406  |           |                                                                                                           |
| St. Stephen's, Bloom-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 181                                | 297   | 478   |           |                                                                                                           |
| Oldfield-road .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 120                                | 304   | 424   |           |                                                                                                           |
| Trinity, King's Head Yard .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 220                                | 300   | 520   |           |                                                                                                           |
| Hulme, Duke-street .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 165                                | 189   | 354   |           |                                                                                                           |
| All-Saints, Oxford road .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 195                                | 191   | 386   |           |                                                                                                           |
| Arwick .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 60                                 | 110   | 170   | 30        |                                                                                                           |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 3434                               | 4213  | 7647  | £10798    |                                                                                                           |

Рисунок № 8. Визуализация данных школ в Манчестере с указанием количества учащихся и затрат на школу в Guardian в 1821 г.<sup>96</sup>

<sup>96</sup> The Guardian News official website. The first Guardian data journalism: May 5, 1821 // <https://www.theguardian.com/news/datablog/2011/sep/26/data-journalism-guardian>

По мнению некоторых исследователей, значительная часть ошибок, которые допускают журналисты при работе со статистической информацией, — это не математические ошибки в расчетах, потому что действительно сложные формулы и операции им приходится использовать крайне редко. Более типичная причина — некорректная интерпретация имеющихся данных, недостаточно ясное понимание того, что именно скрывается за той или иной цифрой, откуда она взялась, о чем говорит. Математика, лежащая в основе корреляций и допустимых ошибок в исследованиях, безусловно, верна, по крайней мере, в большинстве случаев. Но если исследователи не будут искать корреляции, их результаты не имеют большого значения.

Статистика — это самая широко используемая область в математике в количественных исследованиях. Статистические методы распространены в социальных науках и экономике, и журналисты часто ссылаются на последние данные в своих статьях. Исследования с использованием статистических методов начинаются со сбора данных, основанных на гипотезе или теории. Обычно собирается большая выборка данных и проводится её валидация, а затем анализ. Для этой цели обычно используются программные пакеты, такие как SPSS и R. Причинно-следственные связи изучаются путем манипулирования факторами, которые влияют на интересующие явления, одновременно контролируя другие переменные, имеющие отношение к результатам эксперимента.

#### **Основные понятия, необходимые для понимания статистики**

Генеральная совокупность — группа объектов, которые вам интересны как исследователю.

Выборка — часть генеральной совокупности, доступная для исследования. Статистики стремятся к тому, чтобы результаты, полученные на выборках, были верны и для генеральной совокупности<sup>97</sup>.

Связанные выборки — ситуация, при которой любому объекту из первой выборки соответствует ровно один объект из второй. Можно сказать, что они образуют неразрывную пару (а в более сложных случаях — тройку, четверку и т. д.).

Наблюдение — измеренный объект.

Переменные — свойства объектов, которые поддаются измерению.

Значение переменной — степень выраженности того или иного свойства у конкретного объекта.

**Меры центральной тенденции** — они используются, чтобы отобразить наиболее типичные значения, присутствующие в вашей выборке.

Мода — наиболее часто встречающееся значение.

Медиана — середина упорядоченного ряда значений.

Среднее арифметическое — сумма значений, деленная на их количество.

**Меры изменчивости** — нужны, чтобы отобразить степень разброса значений относительно меры центральной тенденции.

Размах — разность между максимальным и минимальным значениями.

Дисперсия — сумма квадратов отклонений, деленная на их количество. Отклонение — это разность между средним арифметическим и конкретным значением. Дисперсии для генеральной совокупности и для выборки вычисляются по разным формулам.

Стандартное отклонение — корень из дисперсии.

#### **Виды статистических анализов:**

- Регрессионный анализ — это набор статистических методов, которые исследуют влияние одной или нескольких переменных. Это помогает предсказать их значения. Наиболее

---

<sup>97</sup> В. Савельев. Статистика и котики. М., 2017

распространенными методами регрессионного анализа являются линейная и логистическая регрессии. Линейная регрессия позволяет предсказать точное количественное значение некоторой переменной, представленной в метрической шкале. А логистическая регрессия может предсказать вероятность принадлежности объекта к тому или иному классу.

- Дискриминантный анализ – используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие группы. Он очень похож на логистическую регрессию, и отвечает на вопрос: по каким переменным я могу отнести конкретный объект в тот или иной класс.
- Факторный анализ – может значительно сократить число переменных и заменить их набором факторов. Более того, факторный анализ может являться предварительной процедурой перед проведением регрессионного анализа в случае, если ряд предикторов коррелирует между собой.
- Кластерный анализ – поможет вам разделить ваши данные на разные классы. При этом число классов может быть заранее неизвестным, либо вы точно знаете их количество. В первом случае, вам нужно применить метод иерархической кластеризации. Такая кластеризация последовательно объединяет объекты в группы, основываясь на расстоянии между ними. Если же вам неизвестно количество классов, необходим метод k-средних, который группирует ваши объекты вокруг так называемых центроидов.

#### **Советы по работе со статистическими данными<sup>98</sup>**

Лучший совет по работе с данными – это получать удовольствие. Данные могут на первый взгляд могут казаться непонятными, но не позволяйте этому стать преградой вашему журналистскому расследованию. Относитесь к цифрам как к чему-то, с чем можно поиграть и исследовать, и они часто с удивительной легкостью раскрывают секреты и истории. Так что обращайтесь с данными так же просто, как с другими доказательствами, без страха и предпочтений. В частности, думайте о работе со статистикой, как об упражнении в воображении.

В работе со статистическими данными очень важен креативный подход, чтобы строить приемлемые гипотезы, которые могут соответствовать данным и лучше объяснять их, а затем проверить их на большем количестве свидетельств. Нужно всегда задаваться вопросом «Какая еще теория могла бы это объяснить?».

Лучшие вопросы - старые: действительно ли это большое число? Откуда это? Вы уверены, что это важно то, что, по вашему мнению, имеет значение? Как правило, это просто побуждения обдумать данные, детали на грани, которые были сжаты при взгляде на одно число, реальные сложности, широкий спектр других возможных сравнений по времени, группе или географическому положению; контекст.

Если данные утверждают, что во время экономического спада потребление алкоголя возросло, вы можете предположить, что это потому, что все в депрессии. Если цифры, наоборот, утверждают, что потребление алкоголя прекратилось, вы можете предположить, что это потому, что все разорены. Другими словами, то, что говорят данные, не имеет значения для интерпретации, которую вы намерены применить к ним. Данных так много, что вы часто сможете найти подтверждение своим прежним убеждениям, если просто немного осмотритесь. Другими словами, журналистика данных, по крайней мере, для меня, не имеет большого значения, если вы не обладаете широкими взглядами. Это объективно настолько, насколько вы стремитесь к этому, а не потому, что основаны на числах.

Само исследование вопроса может стать интересной историей. Рассказ о том, как вы пытались что-либо выяснить, может стать отличной журналистикой, когда вы переходите от одного

---

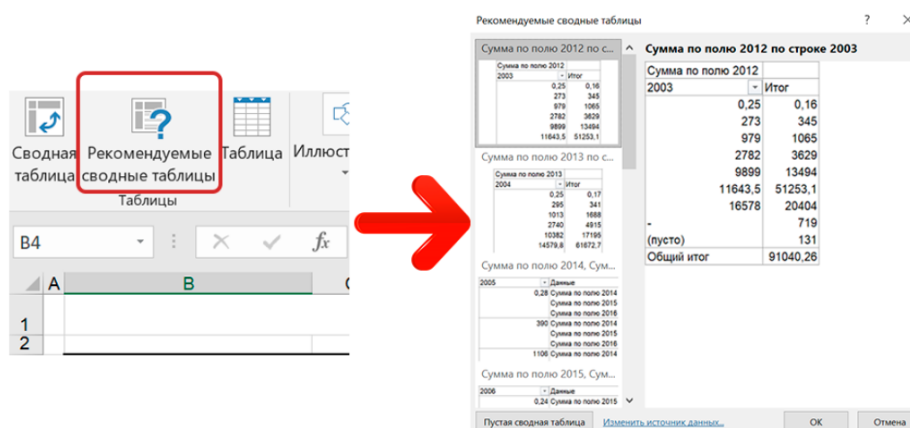
<sup>98</sup> «The Data Journalism Handbook» под редакцией Джонатана Грея



## Рекомендуемые сводные таблицы

С помощью этого приема можно в считанные секунды преобразовать миллион строк данных в краткий отчет. Для этого просто щелкните в любом месте таблицы данных. Перейдите на вкладку «Вставка» и выберите в меню «Рекомендуемые сводные таблицы».

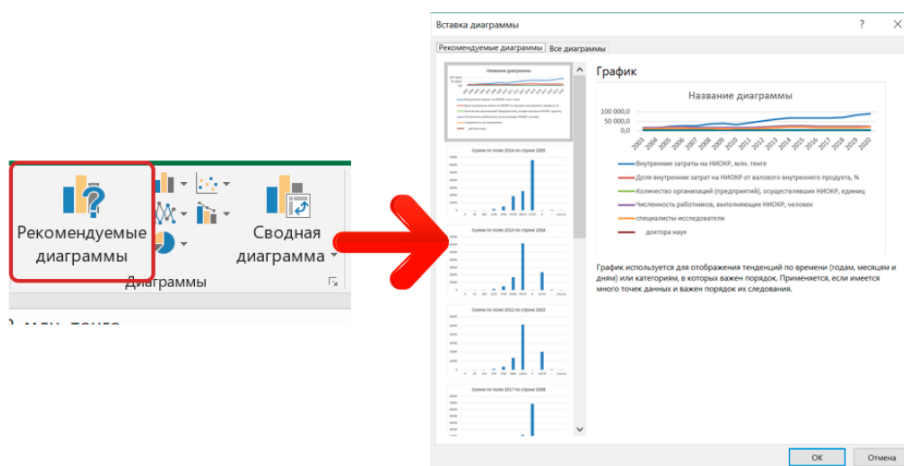
Таблица № 3. Рекомендуемые сводные таблицы



## Рекомендуемые диаграммы

Ваши данные можно всегда представить в виде диаграммы. Как и с таблицами, нужно щелкнуть на таблицу данных. Перейдите на вкладку «**Вставка**», выберите в меню «**Рекомендуемые Диаграммы**», и найдете самую подходящую.

#### Таблица № 4. Рекомендуемые Диаграммы



## Лист прогноза

Вы можете даже сделать прогноз на основе ваших данных. Выделите свою таблицу, перейдите на вкладку «**Данные**», нажмите кнопку «**Лист прогноза**». Вы можете изменить прогнозируемую дату окончания и другие параметры, такие как уровень достоверности ваших данных, изменить диапазон ваших данных, а затем нажать кнопку «**Создать**».

Для данной функции нужно, чтобы ваши данные были в определенном формате для правильной работы. Вам понадобится как минимум два столбца. Первый график времени. Это должно содержать даты, время или последовательность чисел, которые указывают временное изменение. Второй столбец – это ваши данные, то, что вы собираетесь прогнозировать в будущем.

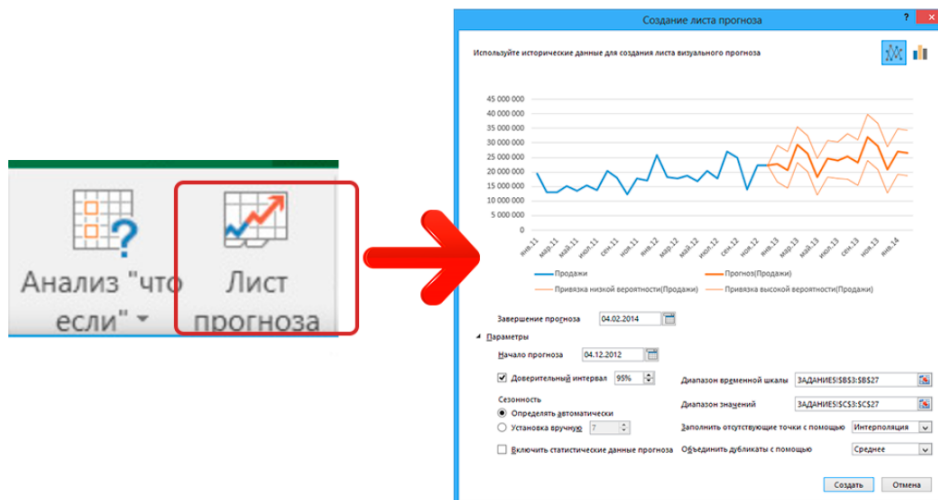


Таблица № 5. Лист прогноза

## Получить данные из интернета

Одним из самых удобных приемов для журналистов может быть получение данных в реальном времени с веб-страницы в Excel, например, давайте перейдем на веб-сайт X-Rates.com. Перейдите на страницу с таблицами курсов обмена долларов США.

Скопируйте ссылку, перейдите в Excel, перейдите в раздел «**Данные**», нажмите «**Получить данные из Интернета**», вставьте URL-адрес и нажмите «**ОК**». Теперь вы можете просмотреть таблицы, чтобы найти ту, которая выглядит так, как вам нужно, выберите и загрузите ее. Теперь перейдите в меню «**Обновить**» и перейдите в «**Свойства подключения**». Установите для него значение. Например, обновлять каждые 60 минут или в любое другое время, которое вам нужно — это обеспечивает автоматическое обновление данных. Нажмите «**ОК**».

## Ресурсы для самостоятельного изучения

Важно помнить, что будущее журналистики основана на больших данных, и поэтому журналистам необходимо хорошо разбираться в данных и их обработке. Раньше можно было проводить расследования просто интервьюируя экспертов, да и сейчас это действенный метод. Но когда речь идет о массивных исследованиях, обязательно детальное изучение данных и оснащение себя инструментами для их анализа и выбора того, что интересно. И держать это в перспективе, помогая людям, действительно видя, где все это сочетается и что происходит в стране.

Ниже представлен список рекомендованной литературы, которая поможет журналистам познакомиться с основами статистики и методологии:

1. У The Data Journalism Handbook, Edited by Jonathan Gray, 2012
2. Cohen S., Numbers in the Newsroom: Using Math and Statistics in News, 2013
3. Michael Blastland & Andrew Dilnot, The Tiger That Isn't: Seeing Through a World of Numbers, 2007
4. Савельев В. Статистика и котика , 2018
5. Бослаф С. Статистика для всех, 2007
6. Бредшоу П. Scraping for Journalists, 2-е издание, 2013
7. Богачев А. Графики, которые убеждают всех, 2015
8. Наследова А.Д. IBM SPSS Statistics 20 и AMOS: Профессиональный статистический анализ данных, 2012
9. Грей Д. Пособие по журналистике данных, 2013

## **2.10. Гендерные аспекты журналистики дезагрегированных данных**

Все методы журналистики данных — это прежде всего инструменты для расследования актуальных социальных проблем. Немало проектов в ЖД уделено гендерным проблемам в регионах, потому что именно объективные данные могут помочь в новом свете посмотреть на вечные, в общем-то, вопросы.

Так, в общественно-политической газете Казахстана «Время» была опубликована занимательная инфографика Владимира Кадырбаева «Жениться или разводиться?» (см. ниже). На были показаны данные за 2019 год, в каких регионах страны люди чаще вступают в брак, а где велико число разводов.

# ЖЕНИТЬСЯ или РАЗВОДИТЬСЯ?

Где в Казахстане чаще играют свадьбы, а где чаще расходятся

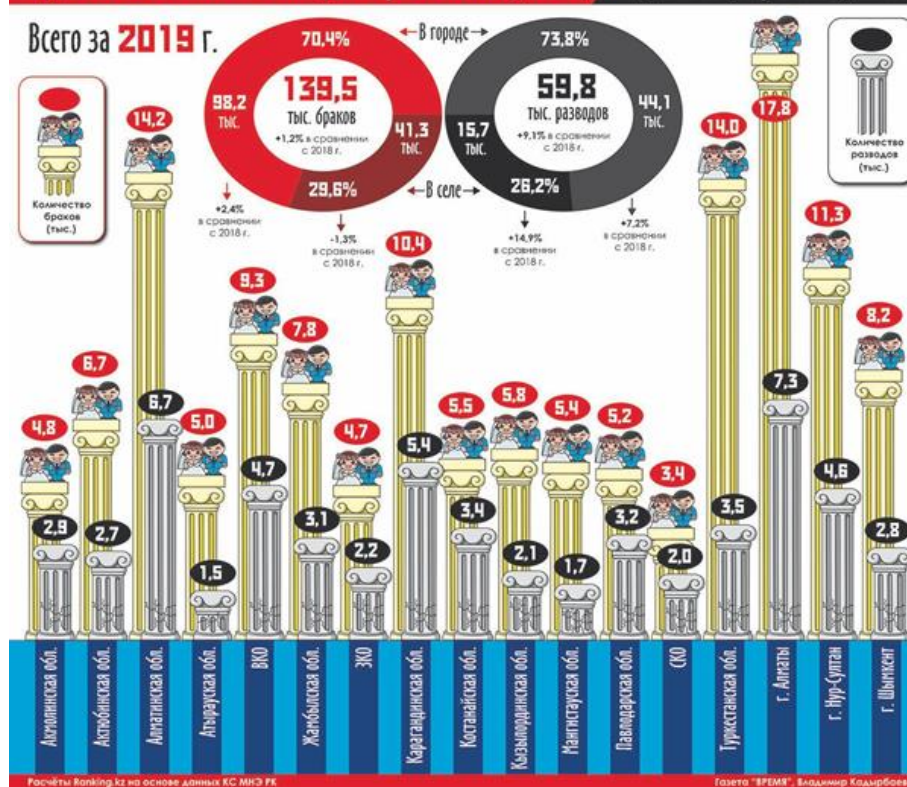


Рисунок № 9. Жениться или разводиться? Гендерные проблемы традиционной журналистики

Распределение гендерных ролей существует и в традиционной журналистике. Например, женщины-журналисты чаще всего освещают так называемые «мягкие» новостные темы, связанные с личной жизнью звезд или общебытовые темы, исходя из предположения, что они по природе своей лучше справляются с такими задачами чем их коллеги мужчины. Кроме того, женщины не могут продвигаться в высшие эшелоны журналистики, из-за существования «стеклянного потолка». Стекло́нный потоло́к – это метафора, которая описывает социальный барьер, препятствующий определенной демографической группе (например, женщинам) подняться выше определенного уровня в карьере. Считается, что стеклянные потолки так или иначе присутствуют почти во всех сферах для разных групп людей.

Хотя журналистика данных является важным инструментом для освящения проблем и обеспечения гендерного равенства, она сама не лишена проблем гендерного характера. Исследования показывают, что, как и в других сферах, в журналистике данных мужчины и женщины предстают перед разными проблемами. Результаты таких исследований показывают, что женщины-журналисты находят, как и больше возможностей, так и барьеры, которые характерны именно для журналистики данных. С одной стороны, некоторые женщины выбирают эту сферу, чтобы избежать среды, где доминируют мужчины журналисты, и гендерной сегрегации, что является факторами, ограничивающими их возможности для приема на работу и продвижения по службе в традиционной журналистике. С другой стороны, навыки журналистики данных имеют дискурсивный гендерный характер.

### **Вступление в карьеру**

В журналистике данных упор делается именно на различные навыки ИКТ и техническую компетентность. Журналисты во время сбора и обработки данных больше полагаются на свои навыки чем на социальные навыки или персональные связи. Поэтому ЖД сфера предоставляет больше равных возможностей, особенно для начинающих журналистов. Для профессионалов этой сферы «цифровой» капитал играет большую роль чем социальный, в него входят преимущественно навыки анализа и очистки больших баз данных и представления результатов общественности в доступной форме.

### **Навигация в сфере журналистики данных**

В сфере данных журналисты уже создают свои социальные ресурсы, по которым они могут делиться накопленным опытом. Сфера ИКТ изначально славилась тем, что компьютерные профессионалы делились своим опытом на форумах и чатах. Особенно, когда речь идет о конкретных технических проблемах.

Также, журналисты данных из разных стран всегда могут с большей легкостью объединяться в одни большие проекты. Например, в крупных международных проектах, таких как The Migrant Files, FarmSubsidy, LuxLeaks и Swiss Leaks, журналисты из разных стран объединили усилия для анализа огромных баз данных с целью выявления нарушений и несправедливостей. Журналистика данных предоставляет больше возможностей для работы с профессионалами из других сфер как программисты, разработчики или дизайнеры. И поскольку журналистика данных ориентирована на сотрудничество, прозрачность и разоблачение правонарушений в обществе, практики более осознают гендерную дискриминацию в своей сфере.

### **Стеклянный потолок в ИКТ сферах**

Несмотря на то, что предпринимаются попытки привлечения женщин в ИКТ сферу, она все равно продолжает ассоциироваться преимущественно с мужчинами. При слове программист, мы чаще представляем себе Марка Цукерберга, чем Дану Бойд (основатель и президент Data & Society Research Institute). Несмотря на высокий уровень предложений по работе, информатика остается областью, в которой доминируют мужчины. И возможно, так происходит из-за символической ассоциации между мужественностью и технологиями. Например, только 18% из всех, кто получает степень бакалавра по компьютерным наукам в Соединенных Штатах являются женщинами. И такие проблемы есть во многих странах. К сожалению, подобная ситуация может привести к тому, что такое гендерное соотношение появится и в журналистике данных, сфере, тесно связанной с ИКТ. Для предотвращения такой ситуации постепенно появляются специальные сообщества и ассоциации как Women Who Code, WebGrrls и Chicas Poderosas.

В заключение, можно сказать, что как новая сфера журналистики, ЖД все же дает больше гендерно-равных возможностей для будущих специалистов. И слово остается за их профессиональными навыками и чутьем.

### **Задания для самостоятельной работы:**

1. Посмотрите на ваши собственные исследования для статей и собранные данные. Определите, к какому виду исследований они относятся.
2. Объясните процесс кодирования данных качественных исследований, например взятых вами интервью или ответах при опросах населения, в качественные данные. Подумайте, как можно преобразовать данные, которые есть у вас на руках и использовать их в журналистике данных.
3. Найдите несколько примеров материалов и статей из сферы журналистики данных и попытайтесь определить, на каких гипотезах был построен первоначальный сбор данных.

4. Продумав рабочие гипотезы (см. пункт выше), продумайте алгоритм, который бы вы применили для сбора и обработки таких данных.

5. Найдите несколько статей из новостных порталов вашей страны и зарубежных информационных агентств похожей тематики. Определите источник данных, которые были в них использованы и выявите принципиальные различия в разных странах.

6. Найдите несколько примеров инфографики, которые иллюстрируют статьи на социальные темы, теперь посмотрите инфографики, которые используются в маркетинговых кампаниях. Определите есть между ними какие-либо принципиальные отличия.

7. Найдите национальные базы данных вашей страны. Определите их наполненность и частоту обновления информации.

8. Изучите традиционные статьи, посвященные гендерным проблемам в вашей стране. Подумайте какой бы подход можно было бы использовать, если бы вы задействовали в них методы журналистики данных.

9. Возьмите любой набор данных будь то из старины Википедии или национальной базы данных, откройте Excel и скопируйте данные в таблицы программы. Выделите данные и щелкните значок в правом нижнем углу «Быстрый Анализ».

10. Попробуйте проделать тоже самое, но только с функцией «Анализ данных» и «Лист прогноза» в Excel. Проанализируйте ваши результаты.

#### **Вопросы для самоконтроля:**

1. Какие есть основные виды стратегии исследований?
2. Как формулируются и подтверждаются гипотезы?
3. Что такое машиночитаемые данные?
4. Что такое интерфейсы прикладного программирования?
5. Для чего нужен процесс скрейпинга?
6. Какие инструменты можно использовать для скрейпинга?
7. В каких случаях невозможно сделать скрейпинг?
8. Какие существуют инструменты для обработки данных?
9. Как можно очистить данные?
10. Какие есть основные виды статистического анализа?
11. Какие есть инструменты в Excel для обработки данных?

### Модуль 3. СБОР ИНФОРМАЦИИ ИЗ ОТКРЫТЫХ ИСТОЧНИКОВ

В 2013 году утечки Сноудена помогли публично подтвердить, что транснациональный аппарат слежки за государствами и технологическими компаниями является фактом, а не фантазией. Эти утечки показали, как граждане стали узнаваемыми с помощью практики больших данных, показывая темную сторону знакомых устройств, приложений и платформ для обработки данных<sup>100</sup>. В Соединенных Штатах запуск специализированного издания Нейта Сильвера по журналистике данных FiveThirtyEight в 2014 году был встречен негативной реакцией за его чрезмерную уверенность в конкретных видах количественных методов и презрение к “журналистике мнений”<sup>101</sup>. В то время как Джон Стюарт из Daily Show провозгласил Сильвера “господом и богом алгоритма” за успешное предсказание результатов выборов 2012 года, статистические методы, которые он отстаивал, подверглись дальнейшей критике и оспариванию после избрания Дональда Трампа в 2016 году. Эти выборы наряду с голосованием по Брекситу в Соединенном Королевстве и ростом популистских правых лидеров по всему миру, как говорили, соответствовали моменту “постправды”<sup>102</sup>, характеризующемуся повсеместной потерей веры в государственные институты, экспертные знания и посредничество в общественной и политической жизни онлайн-платформ, которые сделали их пользователей уязвимыми для таргетинга, манипуляций и дезинформации<sup>103</sup>.

Воспринимается ли так называемый момент “постправды” как свидетельство неудачи или в качестве призыва к действию ясно одно: данные больше нельзя воспринимать как должное, как и журналистику данных. Данные не просто дают нейтральное и прямое представление о мире, но скорее связаны с политикой и культурой, деньгами и властью. Институты и инфраструктуры, лежащие в основе сбора данных—от обследований до статистики, науки о климате до платформ социальных сетей—были поставлены под сомнение. По мере того как пандемия COVID-19 продолжала распространяться по всему миру, цифры, графики и рейтинги стали широко распространяемыми, тематизированными, политизированными и деполитизированными, о чем свидетельствовали ежедневно циркулирующие эпидемиологические диаграммы, на которые ссылается повсеместная стратегия общественного здравоохранения по “сглаживанию кривой”. В то же время широко сообщалось о хрупкости и предварительном характере таких данных, при этом высказывались опасения по поводу недостаточной отчетности, непредоставления отчетности и классификации случаев, а также растущее осознание социальных и политических последствий различных видов данных из источников—от данных больниц до оценок исследований, самоотчетов и транзакционных данных из приложений отслеживания. Пандемия расширила осведомленность не только об использовании, но и о критической отчетности по цифрам и данным<sup>104</sup>.

---

<sup>100</sup> Gray, J., & Bounegru, L. (2019). What a difference a dataset makes? Data journalism and/as data activism. In J. Evans, S. Ruane, & H. Southall (Eds.), *Data in society: Challenging statistics in an age of globalisation* (pp. 365–374). The Policy Press. <https://doi.org/10.5281/zenodo.1415450>

<sup>101</sup> Byers, D. (2014, March 19). Knives out for Nate Silver. Politico. <https://www.politico.com/blogs/media/2014/03/knives-out-for-nate-silver-185394.html>

<sup>102</sup> Davies, W. (2016, August 24). The age of post-truth politics. The New York Times. <https://www.nytimes.com/2016/08/24/opinion/campaign-stops/the-age-of-post-truth-politics.htm>

<sup>103</sup> Для критического взгляда на этот термин см. Jasanoff, S., & Simmet, H. R. (2017). No funeral bells: Public reason in a “post-truth” age. *Social Studies of Science*, 47(5), 751–770.

<sup>104</sup> Bounegru, L. and J. Gray (eds.), *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press, 2021

Журналистику данных можно рассматривать не только с точки зрения того, как представлены вещи, но и с точки зрения того, как она организует отношения—так, чтобы речь шла не только о создании историй данных (посредством сбора, анализа, визуализации и изложения данных), но и о том, кого и что эти истории объединяют (включая аудиторию, источники, методы, учреждения и платформы социальных сетей). Таким образом, мы можем спросить, как недавно выразился Нортье Маррес: “Каковы методы, материалы, методы и механизмы, которые мы курируем, чтобы создать пространства, в которых проблемы можно решать по-другому?”<sup>105</sup>.

### 3.1. Платформизация трансграничной журналистики расследований

Платформизация трансграничной журналистики расследований является растущим явлением, поддерживаемым тем же технопозитивизмом, что и нынешняя тенденция платформизации общества <sup>106</sup>. Платформы для размещения данных для трансграничных расследований начали приобретать известность примерно в 2010 году в контексте проведения расследований с утечкой данных. Возможно, наиболее заметным примером масштабного журналистского сотрудничества на платформе являются т.н. “Панамские документы”, удостоенные Пулитцеровской премии. Для того, чтобы организовать сбор данных и отчетность для 500 журналистов, участвующих в расследовании “Панамских документов”, Международный консорциум журналистов-расследователей (МКЖР) разработал платформу под названием Global I-Hub<sup>107</sup>.

Райл (2017) описывает платформу как “специально разработанную технологию..., используемую для опроса и распространения информации, объединения журналистов в онлайн-отделе новостей и обеспечения того, чтобы журналисты работали как одна глобальная команда”. Это называется “виртуальный офис МКЖР... Facebook для журналистов” как редакционным, так и исследовательским персоналом МКЖР<sup>108</sup>.

Предполагается, что данные и трансграничные расследования идеально сочетаются и расширяют возможности независимого журналистского сотрудничества<sup>109</sup>. Есть последствия того, что очень мало участников управляют такими платформами, и большое количество журналистов зависит от них в сфере трансграничной журналистики. Один из них можно понять как то, что в

---

<sup>105</sup> Вопрос, который Нортье Маррес задала в своем пленарном выступлении на конференции EASST 2018 в Ланкастере: <https://twitter.com/jwyg/status/1023200997668204544>

<sup>106</sup> Dijck, J. van, Poell, T., & Waal, M. de. (2018). The platform society: Public values in a connective world. Oxford University Press

<sup>107</sup> Wilson-Chapman, A. (2017, August 29). Panama Papers a “notable security success.” ICIJ. <https://www.icij.org/blog/2017/08/panama-papers-notable-security-success>

<sup>108</sup> Hare, K. (2016, April 4). How ICIJ got hundreds of journalists to collaborate on the Panama Papers. Poynter. <https://www.poynter.org/reporting-editing/2016/how-icij-got-hundreds-of-journalists-to-collaborate-on-the-panama-papers>; Raab, B. (2016, April 8). Behind the Panama Papers: A Q&A with International Consortium of Investigative Journalists director Gerard Ryle. Ford Foundation. <https://www.fordfoundation.org/ideas/equal-change-blog/posts/behindthe-panama-papers-a-qa-with-international-consortium-of-investigativejournalists-director-gerard-ryle/>

<sup>109</sup> Coronel, S. (2016, June 20). Coronel: A golden age of global muckraking at hand. Global Investigative Journalism Network. <https://gijn.org/2016/06/20/a-golden-age-of-global-muckraking/>; Houston, B. (2016, April 14). Panama papers showcase power of a global movement. Global Investigative Journalism Network. <https://gijn.org/2016/04/13/panama-papers-showcase-power-of-a-global-movement>

ландшафте “больших технологий” было названо “гиперсовременной формой феодализма”, основанной на владении данными<sup>110</sup>.

Доступ к таким платформам по многим веским причинам обеспечивается многими уровнями безопасности, и не каждый журналист может получить доступ. Основной вопрос заключается в том, кто решает, кого включать и исключать, и каковы правила, регулирующие эти решения, а также любые трения и конфликты, которые могут возникнуть в результате этого. Участие в таких платформах, как правило, регулируется базовым соглашением о неразглашении или партнерским соглашением, в котором подробно перечисляются обязанности журналиста или средства массовой информации, получающего доступ, обычно с недостаточной ссылкой на их права. Такие системы и схемы их управления разработаны не с учетом принципов совместного владения, а скорее как структуры, находящиеся в централизованном владении, с наблюдением за действиями пользователей и контролем за нарушениями соглашений в качестве встроенных функций<sup>111</sup>.

Принятие этой модели в журналистских расследованиях, как и в остальной части “экономики совместного использования”, чревато возникновением прекариата в сфере журналистских расследований. Об этом риске свидетельствуют самоописания некоторых организаций, работающих на этих платформах. Например, OCCRP описывает себя как “AirBnb или Uber журналистов”, которые хотят проводить “большие трансграничные расследования”<sup>112</sup>.

Бизнес - модель крупного сетевого посредника в расследованиях сегодня можно рассматривать как подобие цифровой платформы с гигантской экономикой. Доступ к платформе может быть отозван в любое время, управление не открыто для обсуждения, встроено наблюдение за деятельностью пользователей, и “деньги лучше не учитывать”<sup>113</sup>.

### 3.2. Настройка оповещений о данных

Чтобы создавать горячие истории, журналисты данных должны иметь доступ к самым свежим, новейшим выпускам данных, поступающим от наиболее заслуживающих доверия организаций. Чтобы быть в курсе того, какие наборы данных где публикуются, журналисты данных подписываются на системы оповещения, которые отправляют им уведомления каждый раз, когда публикуются потенциально важные данные. Эти системы оповещения часто отправляют уведомления через RSS-каналы или по электронной почте. Также можно настроить пользовательское приложение, такое как DataStringer, для отправки push-уведомлений при внесении существенных изменений или обновлений в исходные базы данных<sup>114</sup>.

После того как вы подпишетесь на оповещения о данных и сформируете четкое представление о расписании выпуска данных, вы можете начать планировать выпуски данных заранее. Например, если вы занимаетесь журналистикой данных в нише бизнес-аналитики и знаете, что особенно интересный квартальный отчет должен быть выпущен через неделю, вы можете

---

<sup>110</sup> Morozov, E. (2016, April 24). Tech titans are busy privatising our data. The Guardian. <https://www.theguardian.com/commentisfree/2016/apr/24/the-new-feudalism-silicon-valley-overlords-advertising-necessary-evil>

<sup>111</sup> Căndeă, Ș (2021), Data Feudalism: How Platforms Shape Cross-border Investigative Networks, in Bounegru, L. and J. Gray (eds.), The Data Journalism Handbook: Towards a Critical Data Practice. Amsterdam: Amsterdam University Press, 2021 doi 10.5117/9789462989511\_ch41

<sup>112</sup> OCCRP, 2017

<sup>113</sup> Lillian Pierson, Data Journalism: Collecting Data for Your Story, <https://www.dummies.com/programming/big-data/data-science/data-journalism-collecting-data-story/>

<sup>114</sup> Lillian Pierson, Data Journalism: Collecting Data for Your Story, <https://www.dummies.com/programming/big-data/data-science/data-journalism-collecting-data-story/>

использовать время, которое у вас есть до его выпуска, чтобы сформулировать план того, как вы будете анализировать данные, когда они станут доступны.

Первым шагом любой истории данных является поиск набора данных для анализа. Для научных авторов одним из естественных источников является раздел результатов любой статьи, которая, по вашему мнению, рассказывает убедительную историю. Многие ученые публикуют свои неаналитические данные на платформах открытого доступа, таких как Dryad и GitHub, - практика, которая позволяет другим, будь то другие ученые или журналисты, исследовать и использовать опубликованные результаты. И даже данные, которые не передаются по каналам открытого доступа, часто доступны по запросу<sup>115</sup>.

В любом случае, выбор в пользу использования результатов одного конкретного исследования в истории данных требует тщательной проверки; рассмотрите учетные данные авторов и изучите раздел "Методы", прежде чем погрузиться в него.

Приянка Ранвал, научный писатель и репортер данных в Climate Central, отмечает, что процесс поиска набора данных может зависеть от задания. В некоторых случаях у вас может возникнуть вопрос (скажем, "Сколько американцев прошли тестирование на COVID-19?"), и вы можете найти конкретный набор данных, который отвечает на этот вопрос. В других случаях можно наткнуться на интересный набор данных (скажем, Глобальный индекс безопасности здравоохранения) и попытаться сформулировать вопрос на его основе.

При изучении потенциального набора данных для использования в проекте подумайте, рассказывают ли данные убедительную историю. Есть ли очевидные тенденции или интересные отклонения? Захотят ли читатели изучить цифру или они предпочтут сразу перейти к выводу? Например, история, объясняющая обзор горячих точек биоразнообразия, может быть полезна с помощью карты или диаграммы, показывающей, где расположены эти места обитания по всему миру и как им угрожают люди. Напротив, сосредоточение внимания на числовых результатах различных испытаний в рассказе о тестировании нового медицинского лечения может отвлечь читателей от понимания качественных выводов о том, чего, по-видимому, достигло лечение до сих пор, и о необходимых шагах, которые предстоит предпринять.

Помимо этих вопросов, представляющих ценность для читателя, рассмотрите логистические проблемы. Разрешается ли загрузить эти данные? Были ли они выпущены по лицензиям Creative Commons? Что представляют собой все метки данных? Понимаете ли вы методы исследования, предостережения и последствия, или вам нужно будет обратиться за разъяснениями к ученому или сотруднику пресс-службы?

Помимо самих научных работ, существует множество общедоступных и удобных для журналистов источников данных. Вот несколько из них:

- Глобальная обсерватория здравоохранения Всемирной организации здравоохранения, хранилище международных данных по широкому спектру показателей здоровья.
- Центры по контролю и профилактике заболеваний (CDC), центральный источник медицинской информации в США, включая данные и информационные бюллетени по вопросам, начиная от случаев гриппа и заканчивая предотвращением лесных пожаров.
- Национальное управление океанических и атмосферных исследований: Национальные центры экологической информации (NOAA: NCEI), центральный источник данных о погоде и стихийных бедствиях в Америке.
- Climate Central, некоммерческая организация по исследованию климата, которая обслуживает местных журналистов и метеорологов в рамках своей программы по вопросам климата.

---

<sup>115</sup> Betsy Ladyzhets, 2020, Interrogating Data: A Science Writer's Guide to Data Journalism, <https://www.theopennotebook.com/2020/07/28/interrogating-data-a-science-writers-guide-to-data-journalism/>

- Кокрейновские обзоры, хранилище медицинских доказательств. (Члены Национальной ассоциации научных писателей получают бесплатный доступ к этому ресурсу.)
- Глобальный информационный фонд по биоразнообразию (GBIF), платформа биоразнообразия с открытым доступом, на которой размещено более 1 миллиона записей о встречаемости видов как от учреждений, так и от платформ гражданской науки.
- Красный список Международного союза охраны природы (МСОП), данные о видах, находящихся под угрозой исчезновения; в Красном списке есть интерфейс прикладного программирования (или API), который по сути является платформой программирования, которую исследователи могут использовать для массовой загрузки огромных объемов данных. Журналисты могут подать заявку на получение ключа API для использования интерфейса.
- Данные во множественном числе, коллекция “полезных/любопытных наборов данных”, собранных редактором данных новостей BuzzFeed Джереми Сингером-Вайном. Сингер-Вайн рассылает дополнения к коллекции в бесплатной еженедельной рассылке.
- Информация прекрасна, публикация, посвященная визуализации данных, сделала все наборы данных, лежащие в основе ее визуализации, доступными в свободном доступе. Эти наборы данных очищаются и обновляются по мере необходимости, что облегчает их изучение начинающим журналистам данных.
- Поиск по наборам данных Google позволяет пользователям искать данные по любой теме с помощью легко настраиваемых фильтров для форматов наборов данных и прав на использование.
- Tabula, инструмент для превращения PDF-файлов в файлы данных. DocumentCloud, аналогичный инструмент, также может похвастаться хранилищем общедоступных документов с открытым исходным кодом, которые прошли через этот процесс.
- Закон о свободе информации (FOIA) запрашивает материалы расследований, которые требуют, чтобы журналисты запрашивали информацию у государственных учреждений. Справочник по журналистике данных включает в себя учебник по FOIA, подготовленный журналистом-расследователем Джорджем Падейски.

Британское бюро журналистских расследований собирает данные из различных источников, собранные в ходе своих расследований, в базе данных под названием Drone Warfare<sup>116</sup>.

Базы данных также могут быть созданы в целях дальнейшей журналистики, в качестве инструмента для содействия исследованиям. Международный консорциум журналистов-расследователей создал и поддерживает базу данных об утечках в оффшорах, которая собирает данные из "Панамских документов" (Panama Papers), "Райских документов" (Paradise Papers) и других расследований.

Аналогичным образом, Проект по освещению организованной преступности и коррупции (OCCRP) поддерживает и обновляет данные OCCRP, что позволяет зрителям просматривать более 19 миллионов общедоступных записей<sup>117</sup>.

В обоих случаях основным пользователем инструментов не является предполагалось, что это будет средний читатель, но вместо этого журналисты и исследователи предполагали провести дальнейшие исследования любой информации, найденной с помощью этих инструментов.

Bellingcat - это нидерландский сайт журналистских расследований, специализирующийся на проверке фактов и разведанных с открытым исходным кодом (OSINT). Он был основан британским журналистом и бывшим блогером Элиотом Хиггинсом в июле 2014 года. Bellingcat публикует результаты как профессиональных, так и гражданских журналистских расследований в зонах

<sup>116</sup> Drone Warfare — The Bureau of Investigative Journalism (en-GB) (thebureauinvestigates.com)

<sup>117</sup> <https://data.occrp.org/>

военных действий, нарушений прав человека и преступного мира. Участники сайта также публикуют руководства по своим методам, а также тематические исследования<sup>118</sup>.

В целях содействия сотрудничеству Международный Консорциум журналистов-расследователей (ICIJ) (МКЖР) предоставляет коммуникационную платформу под названием Global Hub, основанную на компонентах программного обеспечения с открытым исходным кодом<sup>119</sup>.

Его пользователи описали его как "частный Facebook" и позволяют осуществлять такой же прямой обмен информацией, который происходит в физическом отделе новостей.

Репортеры присоединяются к группам, которые следят за конкретными темами—странами, спортом, искусством, судебными разбирательствами или любой другой интересующей темой. В рамках этих групп они могут публиковать сообщения на еще более конкретные темы, такие как политик, которого они нашли в данных или конкретную транзакцию, которую они изучают. Именно здесь происходит большая часть дискуссий, когда журналисты перепроверяют информацию и делятся заметками и интересными документами<sup>120</sup>.

МКЖР потребовалось несколько проектов, чтобы журналисты освоились с iHub. Чтобы облегчить им доступ к платформе и решить технические вопросы, региональные координаторы МКЖР предлагают поддержку. Это является ключом к обеспечению соответствия журналистов требуемым стандартам безопасности

### 3.3. База данных оффшорных утечек

Международный Консорциумом журналистов-расследователей (МКЖР) расследует, кто стоит за более чем 785 000 оффшорными компаниями, фондами и трастами из "Панамских документов", утечками из офшоров, утечками с Багамских островов и изучением "Райских бумаг"<sup>121</sup>. МКЖР опубликовал свое первое издание базы данных оффшорных утечек в 2013 году с использованием графических баз данных, чтобы читатели могли изучить связи между должностными лицами и более чем 100 000 оффшорных организаций. На момент написания этой статьи это число выросло до более чем 785 000 оффшорных компаний, в том числе в результате последующих утечек, таких как Панамские и Райские документы. МКЖР впервые попытался использовать графические базы данных со швейцарскими утечками, но именно с Панамскими документами графические базы данных начали играть ключевую роль на этапе исследований и отчетности. Для изучения 11,5 миллионов сложных финансовых и юридических записи объемом 2,6 терабайта данных были непростой задачей. Используя инструменты сетевого графика, такие как Neo4J и Linkurious, МКЖР смог позволить партнерам быстро исследовать связи между людьми и оффшорными организациями<sup>122</sup>.

Для проекта "Панамские документы" графические базы данных и другие специальные технологии, такие как Knowledge Center и Global I-Hub, были подключены журналисты из почти 80 стран работают на 25 языках через глобальный виртуальный отдел новостей<sup>123</sup>. По состоянию на

---

<sup>118</sup> Wikipedia. Bellinacat - Wikipedia

<sup>119</sup> <https://www.icij.org/blog/2014/07/icij-build-global-i-hub-new-secure-collaboration-tool/>

<sup>120</sup> Díaz-Struck, E. et.al. (2021) *Infrastructuring Collaborations Around the Panama and Paradise Papers*, in Bounegru, L. and J. Gray (eds.), *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press, 2021 doi 10.5117/9789462989511\_ch41

<sup>121</sup> <https://offshoreleaks.icij.org/>

<sup>122</sup> Díaz-Struck, E. et.al. (2021) *Infrastructuring Collaborations Around the Panama and Paradise Papers*, in Bounegru, L. and J. Gray (eds.), *The Data Journalism Handbook: Towards a Critical Data Practice*. Amsterdam: Amsterdam University Press, 2021 doi 10.5117/9789462989511\_ch41

<sup>123</sup> <https://ru.wikipedia.org/wiki/Ibid.>

апрель 2019 года правительства по всему миру выплатили более 1,2 миллиарда долларов США в виде штрафов и неуплаченных налогов в результате расследования Панамских документов<sup>124</sup>.

### **Документы Пандоры**

Около 35 нынешних и бывших лидеров и более 300 государственных чиновников фигурируют в файлах оффшорных компаний, получивших название "Документы Пандоры". Документы раскрывают владельцев некоторых из 95 000 оффшорных фирм, стоявших за этими покупками. Документы Пандоры - это утечка почти 12 миллионов документов и файлов, раскрывающих тайные богатства и сделки мировых лидеров, политиков и миллиардеров. Данные были получены Международным консорциумом журналистов-расследователей в Вашингтоне, округ Колумбия, и привели к одному из крупнейших в истории глобальных расследований. Более 600 журналистов из 117 стран изучили скрытые состояния некоторых из самых влиятельных людей на планете.

Большинство файлов бывают в формате Pdf, изображения, электронного письма, счета-фактуры и тому подобное, которых нелегко найти. Инженеры МКЖР используя такие технологии, как Apache Tika (для извлечения метаданных и текста), Apache Solr (для создания поисковых систем) или Tesseract (для преобразования изображений в текст), создали программное обеспечение с открытым исходным кодом под названием Extract с единственной целью-превратить эти документы в доступный для поиска, машиночитаемый контент.

## **3.4. Открытые данные в Кыргызстане**

По Закону «О доступе к информации, находящейся в ведении государственных органов и органов местного самоуправления КР» основными принципами свободы доступа к информации являются ее общедоступность, объективность, своевременность, открытость и достоверность. В ноябре 2017 года Кыргызстан стал 75 страной, вступившей в OGP (Open government partnership) – «Открытое правительство». Его основными направлениями являются открытые данные, прозрачные госзакупки, доступные госуслуги и прозрачность горнодобывающей отрасли. В рамках реализации страной инициативы 16 ноября 2018 года решением правительства утверждена дорожная карта – Национальный план действий (НПД), куда вошли 19 инициатив.

В январе 2018 года в Кыргызстане запустили проект Госкомитета информационных технологий и связи «Открытые данные». Он предполагает размещение на специальном портале информации госорганов, которую сейчас не получить иначе как по запросу. «Открытые данные» - это 18-месячный проект Госкомитета информационных технологий и связи. Его запустили еще в ноябре 2017 года, но полностью свою деятельность он начал в январе 2018 года. Завершился он в июне-июле 2019 года.

Главной задачей проекта было создание государственного портала открытых данных и обеспечение механизмов доступа населения к ним. Это уникальный ресурс для широкого круга пользователей: бизнеса, СМИ, научной среды, студентов, НПО, гражданского общества и исследовательских компаний. Любой гражданин получит к данным всех госорганов бесплатный доступ на едином сайте. Данные, о которых идет речь - это обезличенная информация, которая не носит персонального характера<sup>125</sup>.

---

<sup>124</sup> GitHub - ICIJ/extract: A cross-platform command line tool for parallelised content extraction and analysis

<sup>125</sup> Vesti.kg, 2018. <https://vesti.kg/politika/item/57290-dannye-vsekh-gosorganov-kyrgyzstana-obnaroduyut-do-2023-goda.html>

В пилотной фазе были вовлечены 12 госорганов КР. Это Министерство образования, Министерство экономики, Министерство финансов, Национальный статистический комитет, Государственная регистрационная служба, Налоговая служба, Фонд обязательного медицинского страхования, Агентство по привлечению и защите инвестиций, Министерство здравоохранения, Госкомитет промышленности, энергетики и недропользования, Министерство юстиции и Таможенная служба. Портал объединил 215 наборов данных от 12 пилотных госведомств Кыргызстана, теперь этой информацией могут свободно воспользоваться все желающие.

Согласно исследованию ГКИТиС, сейчас в КР наиболее востребованы данные по торговле, рынку труда, образованию. Востребованность в такой информации особенно характерна для бизнес-структур. Теперь данные государственных ведомств открыты, их наборы будут пополняться, а информация доступна также в машиночитаемом формате.

Государственная регистрационная служба КР оцифровала данные госсектора, и эти госорганы имеют готовые базы данных в машиночитаемом формате и могут ими поделиться. К середине 2019 года данные этих госорганов разместили на сайте.

**Данный проект был реализован** под зонтиком большого пятилетнего проекта Digital CASA Kyrgyzstan, частью которого является «Открытые данные». Он предполагает полную цифровую трансформацию Кыргызстана до 2023 года, обеспечение Интернетом всех отдаленных регионов страны для получения доступа к открытым данным на уровне органов местного самоуправления и региональных представительств.

Важна и готовность раскрывать свои данные. Открытые данные и работа с ними – новый формат для Кыргызстана. В Кыргызстане привыкли работать с госорганами как владельцами данных в формате вопрос-ответ, в рамках проекта «Открытые данные» ставится цель переформатировать сознание чиновников, приучив его к открытости и подотчетности перед населением.

Большой пласт работы посвящен обучению пользователей в лице СМИ, бизнеса, НПО, госорганов и исследователей в целях обеспечения востребованности открытых данных. Дело в том, что представитель бизнеса часто сталкивается с проблемой поиска бесплатной информации. Нацстатком делится большой категорией данных на сайте, но многие говорят, что их недостаточно для аналитики, они слишком обобщенные. А чтобы получить детальные, нужно платить.

В 2018 году в КР была принята Концепция Цифровой Трансформации “Цифровой Кыргызстан” - 2019-2023 «Санарип Кыргызстан»<sup>126</sup>. Стратегические цели национальной стратегии «Санарип Кыргызстан» - это создание новых возможностей для населения через развитие цифровых навыков, предоставление качественных цифровых услуг, повышение эффективности, результативности, открытости, прозрачности, подотчетности и борьбы с коррупцией, обеспечение экономического рывка через цифровую трансформацию частного сектора и усиление международного партнерства.

Стратегия включает в себя 60 проектов. В их числе создание центра оцифровки документов, внедрение системы «Санарип аймак», цифровизация системы Государственного агентства архитектуры и строительства, внедрение системы электронного документооборота, развитие систем «Э-Кызмат», «Тундук» и многое другое.

14 сентября 2021 года указом президента КР Садыра Жапарова в Кыргызстане появилось Государственное агентство по защите персональных данных - в целях оптимизации системы управления и реализации Закона «Об информации персонального характера». Агентство в соответствии с данным законом будет обеспечивать защиту персональных данных, прав субъектов персональных данных, регистрацию держателей (обладателей) массива персональных данных, ведение Реестра держателей массивов персональных данных, а также новосозданный орган будет осуществлять другие задачи и функции.

---

<sup>126</sup> Санарип Кыргызстан - Концепция (ict.gov.kg)

«Как показывает практика, на сегодня процесс цифровизации в стране испытывает большие трудности. Об этом свидетельствуют жалобы и недовольства со стороны граждан из-за бюрократии в получении государственных услуг. Именно в целях улучшения ситуации начата работа над новой концепцией «Цифровая экономика Кыргызской Республики — 2021-2023 годы», направленной на создание клиентоориентированной среды. Нельзя забывать о базовых и инфраструктурных проектах», – подчеркнул Глава кабинета министров КР Улукбек Марипов на совещание по вопросам цифровизации 11 мая 2021 года.

В ходе презентации концепции «Цифровая экономика Кыргызской Республики - 2021-2023 годы» отмечено, что документ включает в себя основные направления, такие как цифровое образование и развитие человеческого капитала в области ИКТ, создание устойчивой цифровой инфраструктуры, построение «Электронного правительства e-Gov», создание условий по развитию цифровой экономики.

В мае 2021 года в Кыргызстане ОО «МувГрин» запустило открытую базу экологических данных. Отмечается, что данные можно использовать для проведения анализа, подготовки научных исследований и журналистских материалов, косвенно или прямо касающихся вопросов охраны окружающей среды в Кыргызстане.

База экологических данных включает в себя девять компонентов о:

- качестве воздуха;
- биоразнообразии;
- энергетике;
- изменении климата;
- земельных ресурсах;
- транспорте;
- водных ресурсах;
- сельском хозяйстве;
- отходах.

Кроме того, каждый компонент имеет свои параметры, которые содержат актуальную информацию с указанием источника данных. В базе также есть ссылка на полезные публикации, такие как национальные доклады о состоянии окружающей среды в КР, экологический обзор КР и другие. Она доступна [по ссылке](#).

«Данные являются общедоступными, будут обновляться постоянно и могут быть использованы в исследованиях по охране окружающей среды, также для создания профессиональных материалов журналистами и исследователями с обязательным упоминанием источника данных и ссылкой на ОО «МувГрин», — отметили в организации<sup>127</sup>.

---

<sup>127</sup> 24.kg, 2021. <https://24.kg/obschestvo/194322/>

## Модуль 4. ВВЕДЕНИЕ В ЭЛЕКТРОННЫЕ ТАБЛИЦЫ

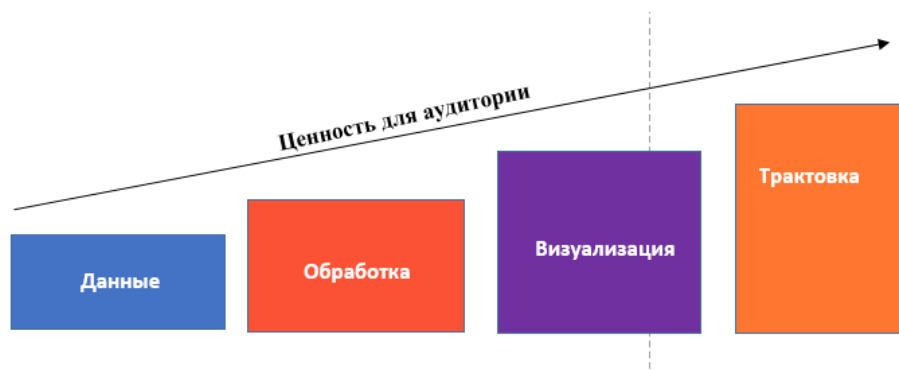
Целью модуля является освоение учащимися методов обработки, анализа и визуализации статистических данных с применением MS Excel и Python Jupyter Notebook, с целью конечной ее публикации и оформления в виде журналистского материала.

На текущий момент журналистика данных является одной из наиболее быстро развивающихся трендов на территории СНГ. Это обусловлено несколькими факторами, самым значимым из которых является распространение блогов и соцсетей, вследствие чего информационная монополия СМИ разрушилась, и их аудитория стала сокращаться. Специализация на журналистике данных, требующей особых навыков извлечения информации и зачастую существенных временных затрат, позволяет журналистам становиться посредниками между базами данных и широкой аудиторией. Иными словами, начала образовываться новая ниша, в которой работа журналистов востребована.

Текст в журналистике данных является вспомогательным, а не основным инструментом. Эта особенность существенно меняет и форму подачи материалов. Информация становится сложнее сделать эмоциональной, но и одновременно она куда более наглядна.

Предметная область журналистики данных включает в себя спектр работ над журналистским материалом, в основе которого лежит:

1. анализ большого объема данных;
2. поиск изначальных данных;
3. очистка массива данных и его последующая аналитика;
4. трактовка и визуальное представление для читательской аудитории



Журналистика данных служит связующим звеном для двух различных направлений работы — технической (которая предполагает проведение работ по сбору и аналитике большого объема данных) и журналистской (которая предполагает исследование темы, а также трактовку и подготовку концепции для наилучшего представления выводов исследования).

Понятие журналистики данных во многом отражает новый виток развития работы с данными, о чем заявляют и эксперты: если раньше основной работой журналиста было именно заполучить еще неопубликованные данные, то сейчас, когда в Интернете появляется все больше данных в открытом виде, все большую ценность получает именно умение обработки данных и поиска в них собственно журналистских историй.

Наконец, понятие журналистики данных связано также с появлением целого спектра доступных инструментов для самостоятельной обработки данных и с появлением огромного объема данных в открытом виде.

#### **4.1. Сложности в развитии журналистики данных как отдельной дисциплины**

1) Аналитический аспект, а именно ложные корреляции является одной из наиболее распространенных проблем в дата журналистике. В корне данной проблемы лежит создание несуществующих зависимостей между различными данными с похожей корреляцией по времени, вследствие чего, их трактовка является как минимум ошибочной, а как максимум - фальсификацией информации.

2) Малое количество открытых для каждого инструментов для обработки и визуализации данных — большая часть представленных на текущий момент решений ориентирована именно на разработчиков, имеющих знания в области языков программирования, либо требует хотя бы базового представления о языках программирования. Таким образом, при работе с данными, журналисты могут совершать целый ряд как намеренных преступлений против честного представления имеющейся информации, так и неумышленных, основанных на неумении работать и анализировать данные.

3) Юридический аспект вопроса. Так, например, в пользовательском соглашении API (инструментов для доступа разработчиков к данным) статистических порталов Республики Казахстан указано, что данные ресурс «не гарантирует достоверность, точность, полноту и своевременность Данных, получаемых с помощью Сервиса». В этой связи, юридическая ценность размещаемых на портале данных сводится к нулю и требует дополнительных юридических консультаций в случае отсылки к полученным данным в результатах журналистского исследования.

#### **4.2. Перспективы журналистики данных**

Одной из главных причин актуальности работы с данными именно сейчас является все больший доступ к данным для каждого. Если раньше именно поиск и сбор информации был основной целью журналиста, то сейчас важнее становится обработка уже имеющейся информации, уже имеющихся данных (как полученных из открытых источников, так и полученных в результате журналистской работы — запросов в государственные структуры, либо поиска и публикации не известных ранее данных).

Предметная область журналистики данных связывает два разрозненных ранее направления: техническую сторону (статистику и аналитику) с журналистской работой (исследование, трактовка и представление). При этом, по мнению журналистского сообщества, журналистику данных нельзя считать заменой традиционной журналистики, но, напротив, еще одним способом расширения сферы ее действия и работы, а также инструментария. Используя данные, дата-журналисты сегодня: во-первых — получают уникальные сюжеты, не прибегая к новостным темам;

во-вторых — получают уникальный инструментарий по контролю тех организаций и государственных структур, которые следуют правилам открытого правительства и публикуют данные о своей деятельности в открытом доступе;

в-третьих — влияют на формирование информационной культуры общества и активной гражданской культуры общества как системы ценностных ориентаций политической системы государства.

Для региональной прессы журналистика данных дает совершенно новые возможности для получения сюжетов, связанных с ситуацией в конкретном городе или районе. Используя возможности по аналитике и трактовке больших объемов данных по региональному обустройству, бюджетированию и ряду других вопросов, становится возможным получить еще больше

инструментов для контроля, изучения и информировании жителей об актуальных проблемах в регионе.

#### 4.3. Поиск и сбор информации. Теоретический аспект алгоритма поиска релевантной информации

Журналистика данных основана на том, что журналисту требуется провести работы по поиску, обработке, анализу и представлению данных. Данная последовательность работ стала уже сформировавшимся подходом для работы над данными в медиа. Данный алгоритм работы можно представить в виде схемы:

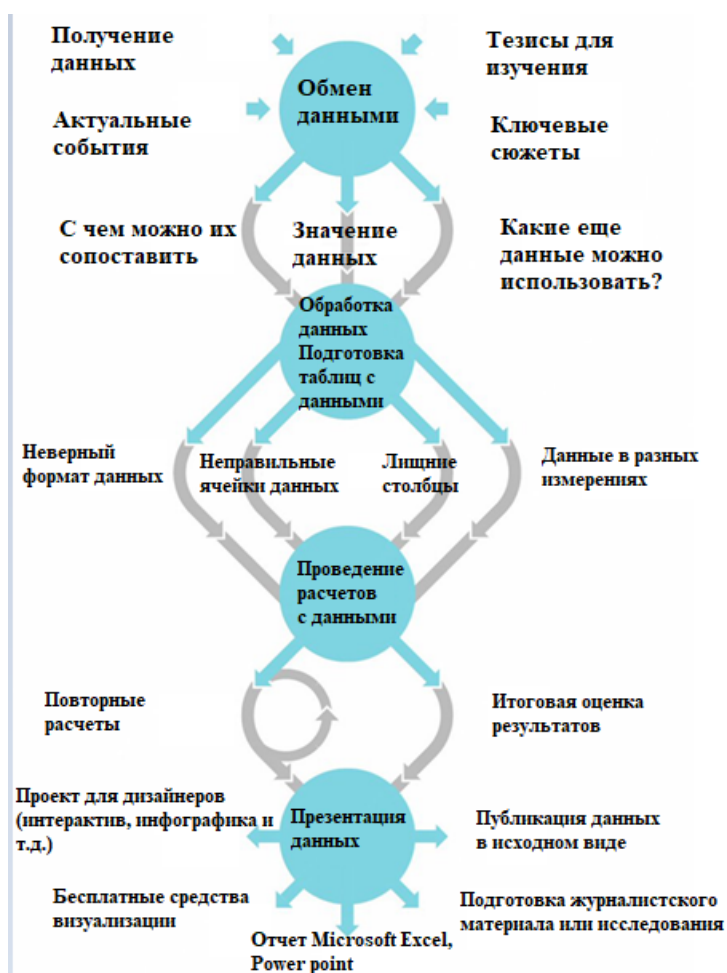


Рисунок № 10. Алгоритм работы анализа и обработки данных

1.1. Таким образом, на первом этапе сотрудники редакции занимаются поиском данных: из множества источников они получают полноценную базу актуальных данных по определенному вопросу (включая данные из новостных сводок, поступающие из государственных органов, журналистских исследований и других источников). Затем, идет анализ полученного массива:

помимо общей трактовки, исследуется, с какими еще пакетами данных возможно использовать найденную информацию, какие корреляции и зависимости возможно проследить в рамках исследуемой темы.

На следующем этапе, происходит обработка и очистка данных: помимо рутинной работы по удалению склеенных (сдвоенных) табличных ячеек и очистки базы от «мусорных» элементов, встречающихся фактически в любом крупном массиве данных, на данном этапе требуется полная дешифровка имеющихся таблиц. В силу того, что многие значения при табличном представлении данных оказываются зашифрованы (скажем, регионы России — соответствующим региональным кодом, поликлиники — своим уникальным идентификатором, и так далее), журналисту совместно со специалистом по обработке данных (или без него) приходится восстанавливать всю логическую цепочку и расшифровывать значение каждого столбца и каждой строки.

При слиянии таблиц, полученных из разных источников, потребуется также привести к единому виду все имеющиеся кодировки (скажем, если страны в одной из таблиц закодированы двухбуквенным кодом, то к точно такому же виду нужно будет привести и все другие таблицы, чтоб объединить данные в них).

При работе с исходными данными в форматах, отличных от машиночитаемых (скажем, в PDF) журналисту приходится либо вручную переносить данные в таблицу, либо использовать специальное ПО для распознавания текста (что в случае с таблицами может приводить к целому ряду сбоя и ошибок).

Наконец, по итогам технической обработки данных и понимания того, какие выводы из исследования журналист хочет донести до аудитории, возникает вопрос представления данных. На этом этапе возможно как оформить материал в виде привычной журналистской статьи, так и воспользоваться инструментами визуализации: от интерактивных форматов до публикации данных через Microsoft Excel, Google Fusion Tables.

При поиске данных для журналистского исследования, существует целый спектр возможностей для нахождения подходящих массивов информации:

1. Поиск данных через стандартные поисковые машины в Интернете (Google, Yandex, Yahoo и другие). Помимо стандартного поиска данных по основной теме, существует возможность найти базы данных, используя более точечные поисковые запросы, учитывающие особенности работы поисковых машин. В частности, через Google с использованием специальных поисковых команд возможно искать данные на определенном портале (на котором хранится большой массив открытых данных), можно искать в сети только файлы определенного типа (скажем, только представленные в виде таблиц Excel).

1.2. Все большее распространение концепции открытых данных и открытого правительства позволяет найти достаточно большой объем данных на официальных сайтах государственных органов. В случае с Республикой Казахстан, вопрос о необходимости постоянного улучшения доступа граждан к открытым данным, а равно и максимально оперативная реакция на ответы по запросам граждан на основе FOI (Freedom of Information — закона о свободе доступа к информации) позволяет значительно продвинуться в сфере предоставления данных в том числе и для журналистов. На данный момент вопрос улучшения работы системы открытых данных в Казахстане все еще является открытым.

На сегодняшний день Правительство Республики Казахстан уделяет внимание открытости и доступности данных, явным доказательством тому служит 29-е место по индексу е-участия (e-Participation), которое страна заняла в рейтинге ООН в 2020 году, что является наилучшим результатом среди стран СНГ. Обновляемый раз в два года индекс EGDl рассчитывается на основе трех показателей каждой из стран: развитие электронных услуг, развитие человеческого капитала и развитие телекоммуникационной инфраструктуры на протяжении двух лет, предшествующих обновлению рейтинга.

В государственной программе «Информационный Казахстан - 2020» были закреплены целевые индикаторы по е-участию и использованию сервисов на базе Портала «Открытых данных»: количество активно используемых приложений, базирующихся на сервисах «открытых данных», в 2017 году - не менее 50, в 2020 году – 65. Интернет-отрасль проявила высокий интерес к Открытым данным. ОЮЛ «Ассоциация Казахстанского интернет бизнеса и мобильной коммерции» готово активно сотрудничать, развивать проекты и создавать приложения на основе «Открытых данных».

Однако говорить о значительных подвижках в данной сфере еще рано: на текущий момент существует целая масса примеров так называемых «плохих данных», которые публикуются в открытом виде на государственных порталах в Республике Казахстан.

Анализируя источники открытых данных Республики Казахстан, можно выделить целые массивы данных, которые выложены на официальных порталах в нечитаемом виде: наборы текста в одну строку, представление данных в устаревших версиях ПО, представление информации в виде графиков и диаграмм, без доступа к исходным данным, и так далее. Вследствие слабого контроля за качеством выкладываемых открытых данных, многие государственные структуры Республики Казахстан имеют возможность под видом открытых данных выкладывать в Интернете, по сути, нечитабельные документы, которые невозможно проанализировать.

Центральными ресурсами, на которых можно найти большие открытые базы данных от официальных структур, являются Бюро национальной статистики, Открытые данные egov, проект Open Data Зерде. Представленные наборы данных посвящены самым разным вопросам: от результатов ЕНТ по областям до данных правительства о динамике смертности и рождаемости, или информации Национального Банка Республики Казахстан о суммарных размерах вкладов юридических и физических лиц в казахстанских банках. Используя представленные ресурсы, можно найти целый ряд документов и баз данных, которые журналисту можно использовать для достаточно комплексного анализа выбранных вопросов.

1.3. Доступ к данным через API (интерфейс программирования приложений). Используя навыки программирования (или воспользовавшись услугами специалистов), становится возможным получить доступ к данным, которые не выложены в Интернете в удобном для доступа и скачивания виде. Скажем, если государственное ведомство предоставляет отчетность в формате интерактивного графика, то с использованием технологий парсинга, то есть захвата данных становится возможным получить полноценный, структурированный набор данных там, где их использование третьими лицами изначально не предполагалось.

Необходимость в очистке данных (процессе поиска и корректировки неверных или неточных элементов базы данных, который нужен для последующей работы с базой данных) возникает при работе практически с любыми данными, полученными из открытых источников. В виду того, что в каждой организации выставляются свои требования к оформлению данных и к контролю за качеством вносимых в базы данных сведений, журналисту в любом случае придется изучить полученную базу данных, а в случае ее слияния с другими — еще и привести к единому форматированию. Так, в разных таблицах, или даже в пределах одной базы данных часто встречаются примеры, когда одни и те же графы оказываются закодированы по-разному. Журналисту приходится приводить к единому формату самые разные столбцы: от кодировок стран (которые можно закодировать двумя буквами, можно цифрами, можно полными названиями стран и так далее) до самой простой графы — «пол», которую можно обозначить либо цифрами, либо словами, либо сокращениями. Другая потенциальная проблема — использование синонимов или неточных определений, которые также усложняют верную обработку данных (скажем, использование в рамках одного столбца кодировок «студент», «учащийся», «школьник»). В этом случае требуется либо привести все названия к единому виду, либо упростить базу данных в этом пункте, чтобы избежать неточностей при дальнейшем представлении.

Полноценный функционал по очистке больших массивов данных доступен в более комплексном ПО, с использованием языков программирования: Python, R и другие решения для работы с данными дают пользователю практически неограниченные возможности по очистке данных, в том числе и для их последующей визуализации. Несмотря на то, что основным способом представления большого объема данных в журналистике служит визуализация (в форматах от инфографии до интерактивных веб-приложений), во многих случаях важнее оказывается трактовка полученных данных. Будь это аналитический материал или пояснительная записка к визуализации, часто именно журналистская трактовка и аналитика полученных результатов позволяет раскрыть картину происходящего. Таким образом, многие журналисты признают, что в ряде случаев, разумнее пренебречь полноценной визуализацией, особенно в случаях, когда для ее качественной разработки недостаточно ресурсов. Вместо этого возможно опубликовать вместе с журналистским материалом либо «сырые» данные, либо представить выводы в более традиционном формате, со статичными графиками и аналитикой.

Многие западные издания отмечают, что, несмотря на множество бесплатных, недорогих или условно-бесплатных инструментов для визуализации (Overview Project, Tableau, Dipity, Google Fusion Tables), все они не могут сравниться с теми возможностями, которые получает журналист при работе с отдельной дизайн-командой. Основная проблема бесплатных инструментов заключается в том, что большая их часть заточена под всего один или несколько стандартных вариантов визуализации и не может дать необходимой гибкости в работе для создания качественного визуального проекта, заточенного под требования конкретного журналистского материала. В многочисленных подборках лучших инструментов для визуализации, большую часть представляют дополнительные библиотеки или надстройки для профессионалов JavaScript, HTML5, Ruby и других языков программирования, которыми активно пользуются крупнейшие медиа и организации по всему миру, от New York Times до ЮНЕСКО. Учитывая наличие тысяч примеров визуализации с D3 (крупная библиотека с шаблонами для визуализации на базе JavaScript), использовать этот инструмент в журналистике становится еще удобнее. Во многом именно появление бесплатных инструментов визуализации спровоцировало в журналистских кругах дискуссии о примерах «плохой визуализации» — использовании неточных, примитивных и (или) шаблонных вариантов для изображения информации в журналистских материалах. Журналисты сходятся во мнении, что для поиска наилучшего решения по визуализации (а также принятия решения о том, необходима ли она в принципе), следует создать множество черновых, вспомогательных графиков, и лишь после жесткого их отбора выбрать наилучший способ представления информации. Таким образом, на выходе, в журналистском материале не должно быть неинформативной визуализации или непроработанных графиков (без указания точной «легенды», подписей к осям и так далее).

Таким образом, несмотря на то, что бесплатные инструменты для визуализации иногда используются также и крупными СМИ, в большинстве случаев, приоритет отдается именно внутренним разработкам, которые осуществляют внешние или штатные команды дизайнеров и разработчиков. На основе полученных после обработки объема данных результатов, они также могут принимать решение как о необходимости визуализации в принципе, так и о наилучших способах ее реализации.

#### **4.4. Подготовка материала для анализа данных**

В настоящем пособии используются открытые и доступные на сегодняшний день данные. Как было указано ранее, основными источниками статистических данных по Республике Казахстан являются:

- Бюро национальной статистики, <https://stat.gov.kz/>

- Открытые данные egov, <https://data.egov.kz/>
- Проект Open Data Зерде, <https://zerde.gov.kz/activity/center-competencies/open-data/>
- Евразийская экономическая статистика,

[http://www.eurasiancommission.org/ru/act/integr\\_i\\_makroec/dep\\_stat/union\\_stat/Pages/default.aspx](http://www.eurasiancommission.org/ru/act/integr_i_makroec/dep_stat/union_stat/Pages/default.aspx)

При выборе основного источника данных мы будем руководствоваться двумя интересующими нас параметрами: предметной отрасли статистических данных и ее релевантности. Основной отраслью в нашем случае будет являться наука и инновационная деятельность Республики Казахстан. Это связано с тем, что в настоящий момент Казахастанский Национальный Университет им. аль-Фараби, являясь одним из наиболее передовых и технически развитых университетов Казахстана, ведет исследовательскую и научную деятельность, развивая новые отрасли науки, в частности, информационные технологии. Соответственно, имея достаточно обширную научную базу и исследовательский опыт, тема науки и образования в Казахстане является одной из приоритетных для университета.

Вторым параметром является релевантность. К настоящему моменту одним из наиболее существенных факторов, влияющих на темпы развития дата-журналистики как отдельной и самостоятельной дисциплины в Казахстане, является проблема открытых данных. Несмотря на то, что ведется статистический учет, выпускаются отчеты с данными, правительство активно участвует в развитии этой отрасли, пока относительно рано говорить о полной интеграции дата-журналистики в информационно-техническую среду Казахстана.

Тем не менее прогресс в данном направлении имеет положительную тенденцию, так как с каждым годом доступ к данным улучшается, все больше организаций, как частных, так и государственных, публикуют отчеты своей деятельности в открытый доступ. Улучшается качество оформления статистических данных, становится значительно меньше ошибок и неточностей.

Принимая во внимание вышесказанное, и учитывая, что дата-журналистика в Республике Казахстан - это достаточно молодая, и, как, следствие, развивающаяся предметная область, рассмотрим параметр релевантности данных.

Бюро Национальной статистики – это крупнейший и наиболее структурированный и содержательный портал, с большим объемом статистических данных. Каждая отрасль имеет свой раздел, где, в свою очередь, данные делятся на экспресс-данные, бюллетени, основные показатели, которые мы и будем использовать, а также динамические таблицы.

Соответственно, для анализа и визуализации данных нами будут использованы следующие статистические данные:

- 1) Объем инновационной продукции (товаров, услуг),  
<https://stat.gov.kz/api/getFile/?docId=ESTAT098609>
- 2) Основные показатели инновационной деятельности,  
<https://stat.gov.kz/api/getFile/?docId=ESTAT100456>
- 3) Инновационная активность предприятий,  
<https://stat.gov.kz/api/getFile/?docId=ESTAT098608>
- 4) Основные показатели развития и состояние науки,  
<https://stat.gov.kz/api/getFile/?docId=ESTAT100458>
- 5) Число организаций, выполняющих исследования и разработки,  
<https://stat.gov.kz/api/getFile/?docId=ESTAT095997>
- 6) Численность магистрантов и докторантов,  
<https://stat.gov.kz/api/getFile/?docId=ESTAT294850>  
<https://stat.gov.kz/api/getFile/?docId=ESTAT294849>
- 7) Количество высших учебных заведений Республики Казахстан по городам,  
[https://enic-kazakhstan.edu.kz/ru/reference\\_information/universities](https://enic-kazakhstan.edu.kz/ru/reference_information/universities)

Все данные были отобраны и структурированы по релевантности и точности данных. Так как сфера образования и науки является одной из самых значимых отраслей, оказывающей непосредственное влияние на научно-технические, экономические и социальные показатели страны, в данном пособии мы рассмотрим основные показатели и темпы развития, а также общие тенденции к положительному росту в данной отрасли.

### **Лабораторная работа № 1. “Дата журналистика - основные определения и понятия”**

#### **Вопросы:**

1. Что такое дата-журналистика?
2. Обосновать все преимущества дата журналистики над классическим журналистским расследованием.

#### **Задание:**

3. Дать определение дата-журналистики, описать краткую историю появления данного понятия.

4. Дать собственную качественную оценку дата журналистике как принципиально нового способа обработки и трактовки информации.

### **Лабораторная работа № 2. “Алгоритм работы дата-журналиста”**

#### **Вопросы:**

1. Что такое итоговая публикация?
2. Что такое алгоритм?

#### **Задание:**

Создайте и опишите подробную схему алгоритма работы журналистики данных. Найдите и опишите примеры работ, использующих дата-журналистику как основу.

### **Самостоятельная работа студента. “Дата-журналистика в Казахстане”**

#### **Вопросы:**

1. Приведите положительные, и отрицательные факторы, влияющие на журналистику данных в Казахстане на сегодняшний день.

#### **Задание:**

2. Перечислите источники статистических данных по Казахстану. Опишите их основные достоинства и недостатки
3. Приведите данные по пяти различным отраслям, используя ресурс [stat.gov.kz](http://stat.gov.kz). Проведите краткий устный анализ ваших данных.

#### **Вопросы для самоконтроля:**

1. Приведите основные достоинства и недостатки дата журналистики как предметной области.
2. Опишите алгоритм работы дата журналиста от постановки задачи до итоговой публикации.
3. Приведите основные источники статистических данных Республики Казахстан.
4. Опишите основные факторы, препятствующие развитию дата-журналистики в Казахстане.
5. Какое событие в публицистике можно считать первым опытом применения дата анализа в журналистике?
6. Какие существуют варианты визуализации данных?
7. В чем основное отличие дата журналистики от классического журналистского исследования?
8. Какие критерии являются основными при отборе информации для дата анализа?
9. Что такое очистка данных?

10. Какой этап дата-анализа является последним и наиболее ценным для аудитории при публикации?

## **Модуль 5. ЭЛЕКТРОННЫЕ ТАБЛИЦЫ: ОБРАБОТКА, АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ТАБЛИЧНЫХ ДАННЫХ**

### **5.1. Анализ данных в Microsoft Excel**

Следующим этапом дата анализа в журналистике данных является подбор инструментария. В данном пособии в качестве инструментов дата журналиста будут разобраны два решения: Microsoft Excel и язык программирования Python, а именно веб-оболочка Jupyter Notebook.

При анализе работы с данными, можно определить два ключевых направления для представления больших объемов данных:

- **интерактивные форматы визуализации** (краудсорсинг форматы, интерактивная инфографика);
- **статичное представление данных** (визуализация в виде инфографики, публикация текстовой журналистской истории, представление в табличном виде адаптированных данных, публикация просто «сырых» или прошедших первичную обработку данных для самостоятельного анализа аудиторией). Именно это представление будет разобрано в нижеследующих главах.

В данной главе будет разобран процесс сортировки, фильтрации и анализа данных в Microsoft Excel.

### **5.2. Обзор программы Microsoft Excel**

**Microsoft Excel** — это мощный инструмент, который справится с большинством задач, полезных для журналиста, которому необходимо анализировать данные, чтобы обнаруживать интересные закономерности. Эти задачи включают:

1. Сортировка
2. Фильтрация
3. Использование математических и текстовых функций
4. Сводные таблицы

Excel будет обрабатывать большие объемы данных, организованных в виде таблиц со строками и столбцами. В столбцах (которые помечены A, B, C...) перечислены переменные (например, имя, возраст, количество преступлений и т. Д.). Как правило, первая строка содержит имена переменных. Остальные строки предназначены для отдельных записей или анализируемых дел. Каждая ячейка (например, A1) содержит фрагмент данных.

Как уже отмечалось выше, в качестве основы для дата-анализа мы будем использовать данные из открытых источников, а именно данные о высшем образовании и основных показателях состояния и развития науки в Республике Казахстан.

Excel предоставляет команды, функции и инструменты, которые облегчают выполнение ваших задач анализа данных. Вы можете избежать многих трудоемких и (или) сложных расчетов, используя Excel. В этом руководстве вы получите представление о том, как выполнять анализ данных с помощью Excel. Вы поймете на соответствующих примерах, шаг за шагом использование команд Excel и снимки экрана на каждом этапе.

#### **Диапазоны и таблицы**

Данные, которые у вас есть, могут быть в диапазоне или в таблице. Определенные операции с данными могут выполняться независимо от того, находятся ли данные в диапазоне или в таблице.

Однако существуют определенные операции, которые более эффективны, когда данные находятся в таблицах, а не в диапазонах. Есть также операции, которые предназначены исключительно для таблиц.

### Условное форматирование

Excel предоставляет вам команды условного форматирования, которые позволяют раскрашивать ячейки или шрифт, рядом со значениями в ячейках отображаются символы на основе предварительно определенных критериев. Это помогает визуализировать выдающиеся ценности. Вы поймете различные команды для условного форматирования ячеек.

### Сортировка и фильтрация

Во время подготовки анализа данных и / или для отображения определенных важных данных вам может потребоваться отсортировать и / или отфильтровать ваши данные. Вы можете сделать то же самое с простыми в использовании опциями сортировки и фильтрации, которые есть в Excel.

Таблица — это прямоугольный диапазон структурированных данных. Ключевые особенности:

1. Каждая строка в таблице соответствует отдельной записи данных.
2. Каждый столбец содержит определенный фрагмент информации.
3. Верхняя строка описывает информацию, содержащуюся в каждом столбце, и называется строкой заголовка.

Каждая запись в верхней строке называется заголовком столбца.

|    | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T |
|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 2  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 3  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 4  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 5  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 6  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 7  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 8  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 9  |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 10 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 11 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 12 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 13 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 14 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
| 15 |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Таблица № 7. Основные показатели состояния и развития науки РК

Вы можете создавать и использовать таблицу Excel для удобного управления и анализа данных. Кроме того, с помощью таблиц Excel вы получаете встроенные функции фильтрации, сортировки и затенения строк, которые облегчают ваши действия по составлению отчетов. Кроме того, Excel реагирует на действия, выполняемые на столе, интеллектуально. Например, у вас есть формула в столбце или вы создали диаграмму на основе данных в таблице. Когда вы добавляете больше данных в таблицу (т.е. больше строк), Excel расширяет формулу для новых данных, и диаграмма расширяется автоматически.

### Разница между таблицами и диапазонами

Таблица — это более структурированный способ работы с данными, чем диапазон. Вы можете конвертировать диапазон в таблицу, и Excel автоматически предоставляет:

1. Имя таблицы
2. Имена заголовков столбцов

3. Форматирование данных (цвет ячейки и цвет шрифта) для лучшей визуализации Таблицы предоставляют дополнительные функции, которые недоступны для диапазонов:

1. Excel предоставляет на ленте табличные инструменты от свойств до стилей.  
2. Excel автоматически предоставляет кнопку «Фильтр» в заголовке каждого столбца, чтобы отсортировать данные или отфильтровать таблицу так, чтобы отображались только те строки, которые соответствуют заданным критериям.

3. Если у вас есть несколько строк в таблице, и вы прокручиваете лист вниз, чтобы строка заголовка исчезла, буквы столбца на рабочем листе заменяются заголовками таблицы.

4. Когда вы помещаете формулу в любую ячейку в столбце таблицы, она распространяется на все ячейки в этом столбце.

5. Вы можете использовать имена таблиц и заголовков столбцов в формулах, не используя ссылки на ячейки или создавая имена диапазонов.

6. Вы можете расширить размер таблицы, добавив больше строк или столбцов, щелкнув и перетаскивая небольшой треугольный элемент управления в нижнем правом углу правой ячейки.

7. Вы можете создавать и использовать слайсеры для таблицы для фильтрации данных. Excel предоставляет на ленте табличные инструменты от свойств до стилей. Excel автоматически предоставляет кнопку «Фильтр» в заголовке каждого столбца, чтобы отсортировать данные или отфильтровать таблицу так, чтобы отображались только те строки, которые соответствуют заданным критериям.

8. Если у вас есть несколько строк в таблице, и вы прокручиваете лист вниз, чтобы строка заголовка исчезла, буквы столбца на рабочем листе заменяются заголовками таблицы.

9. Когда вы помещаете формулу в любую ячейку в столбце таблицы, она распространяется на все ячейки в этом столбце.

### 5.3. Сортировка

Одна из самых полезных возможностей Excel - сортировать данные в более понятном порядке. Слишком часто нам выдаются списки в алфавитном порядке, что полезно только для поиска определенной записи в длинном списке. В журналистике нас обычно больше интересуют крайности: наибольшее, наименьшее, наибольшее, наименьшее, лучшее, наихудшее. Рассмотрим данные:

|                                                                     | 2003     | 2004     | 2005     | 2006     | 2007     | 2008     | 2009     | 2010     | 2011     | 2012     | 2013     | 2014     | 2015     | 2016     | 2017     | 2018     | 2019     | 2020     |
|---------------------------------------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Внутренние затраты на НИОИР, млн. руб.                              | 11 643,5 | 14 579,8 | 21 527,4 | 24 799,9 | 26 825,5 | 34 761,6 | 38 968,7 | 33 466,8 | 43 351,6 | 51 253,1 | 61 672,7 | 66 847,6 | 69 302,8 | 66 605,1 | 68 884,2 | 72 224,6 | 82 333,1 | 89 028,7 |
| Доля внутренних затрат на НИОИР от валового внутреннего продукта, % | 0,25     | 0,25     | 0,28     | 0,24     | 0,25     | 0,22     | 0,23     | 0,15     | 0,15     | 0,16     | 0,17     | 0,17     | 0,17     | 0,14     | 0,13     | 0,12     | 0,12     | 0,13     |
| Количество организаций (предприятий), осуществивших НИОИР, единиц   | 273      | 295      | 390      | 437      | 438      | 421      | 414      | 424      | 412      | 345      | 343      | 392      | 390      | 383      | 386      | 384      | 388      | 396      |
| Численность работников, выполняющих НИОИР, человек                  | 16 570   | 16 715   | 18 912   | 19 563   | 17 774   | 16 304   | 15 793   | 17 021   | 18 003   | 20 404   | 23 712   | 25 793   | 24 735   | 22 965   | 22 081   | 22 379   | 21 843   | 22 665   |
| в том числе:                                                        |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| специальности-исследователи наук                                    | 9 890    | 10 382   | 11 910   | 12 404   | 11 524   | 10 780   | 10 095   | 10 870   | 11 488   | 13 494   | 17 195   | 18 930   | 18 454   | 17 421   | 17 056   | 17 454   | 17 124   | 18 228   |
| иные                                                                |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| инженеры наук                                                       | 970      | 1 013    | 1 026    | 1 157    | 1 166    | 1 193    | 1 135    | 1 341    | 1 063    | 1 688    | 2 006    | 1 821    | 1 820    | 1 818    | 1 740    | 1 709    | 1 883    | 1 920    |
| доктора наук                                                        | -        | -        | -        | -        | -        | -        | -        | -        | 2 488    | 2 739    | 4 029    | 3 96     | 3 499    | 4 097    | 3 54     | 3 36     | 3 177    | 4 000    |
| доктора философии PhD                                               | -        | -        | -        | -        | -        | -        | -        | 60       | 59       | 391      | 218      | 330      | 435      | 466      | 529      | 496      | 5 205    | 3 778    |
| исследователи наук                                                  | 2 780    | 2 740    | 3 016    | 3 147    | 3 068    | 2 861    | 2 734    | 3 012    | 3 008    | 3 676    | 4 913    | 5 254    | 5 119    | 4 726    | 4 661    | 4 560    | 4 240    | 4 529    |

Таблица № 8. Сортировка данных

В MS Excel есть два метода сортировки. Первый метод быстрый и может использоваться для сортировки по одной переменной. Поместите курсор в столбец, по которому вы хотите отсортировать, а затем щелкните значок A – Z:

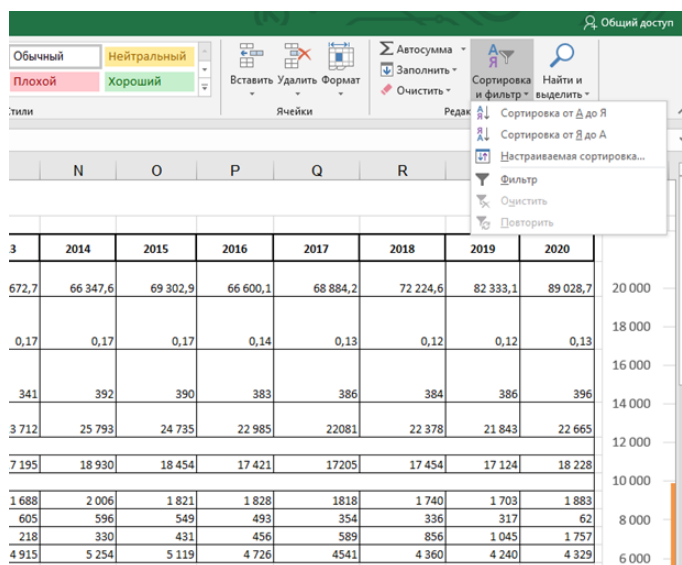


Таблица № 9. Меню сортировки и фильтрации

Сортируйте в любом порядке. Но будьте осторожны! Поместите курсор в столбец, но НЕ выбирайте букву столбца (в данном случае D), а затем выполняйте сортировку.

## 5.4. Фильтрация

Иногда вам нужно изучить только отдельные записи из большого набора данных. Для этого вы можете использовать инструмент Excel "Фильтр". На панели инструментов перейдите на вкладку «Данные», затем нажмите «Фильтр». Вверху каждого столбца появятся маленькие кнопки:

|                                                                     | 2003     | 2004     | 2005     |
|---------------------------------------------------------------------|----------|----------|----------|
| Внутренние затраты на НИОКР, млн. тенге                             | 11 643,5 | 14 579,8 | 21 527,4 |
| Доля внутренних затрат на НИОКР от валового внутреннего продукта, % | 0,25     | 0,25     | 0,28     |
| Количество организаций (предприятий), осуществлявших НИОКР, единиц  | 273      | 295      | 390      |
| Численность работников, выполняющих НИОКР, человек                  | 16 578   | 16 715   | 18 912   |

Таблица № 10. Итоговый результат сортировки

Предположим, мы хотим видеть только записи «Количество организаций, осуществляющих НИОКР». Нажмите кнопку в столбце «Область», снимите флажок «Выбрать все», а затем выберите из списка «Количество организаций, осуществляющих НИОКР», например:

|    |                                                                     |          |          |
|----|---------------------------------------------------------------------|----------|----------|
| 1  |                                                                     |          |          |
| 2  |                                                                     |          |          |
| 3  |                                                                     | 2003     | 2004     |
| А↓ | Сортировка от А до Я                                                | 11 643,5 | 14 579,8 |
| Я↓ | Сортировка от Я до А                                                |          |          |
|    | Сортировка по цвету                                                 |          |          |
| ✖  | Удалить фильтр из столбца "(Столбец В)"                             | 0,25     | 0,25     |
|    | Фильтр по цвету                                                     |          |          |
|    | Текстовые фильтры                                                   |          |          |
|    | Поиск                                                               |          |          |
| 2  | (Выделить все)                                                      | 16 578   | 16 715   |
|    | <input checked="" type="checkbox"/> Внутренние затраты на НИОКР, мл |          |          |
|    | <input checked="" type="checkbox"/> доктора наук                    | 9 899    | 10 382   |
|    | <input checked="" type="checkbox"/> доктора по профилю              |          |          |
|    | <input checked="" type="checkbox"/> доктора философии PhD           | 979      | 1 013    |
|    | <input checked="" type="checkbox"/> Доля внутренних затрат на НИОКР |          |          |
|    | <input checked="" type="checkbox"/> из них:                         | -        |          |
|    | <input checked="" type="checkbox"/> кандидаты наук                  | 2 782    | 2 740    |
|    | OK                                                                  |          |          |
|    | Отмена                                                              |          |          |

Таблица № 11. Алгоритм фильтрации данных

3

Сортировка от А до Я

Сортировка от Я до А

Сортировка по цвету

Удалить фильтр из столбца "(Столбец В)"

Фильтр по цвету

Текстовые фильтры

Поиск

☐ доктора по профилю
 ☐ доктора философии PhD
 ☐ Доля внутренних затрат на НИОКР
 ☐ из них:
 ☐ кандидаты наук
 ☒ **Количество организаций (предпри**
☐ специалисты-исследователи
 ☐ Численность работников, выполни

OK

Отмена

| 2003     | 2004     |
|----------|----------|
| 11 643,5 | 14 579,8 |
| 0,25     | 0,25     |
| 273      | 295      |
| 16 578   | 16 715   |
| 9 899    | 10 382   |
| 979      | 1 013    |
| -        | -        |
| 2 782    | 2 740    |

Таблица № 12. Алгоритм фильтрации данных

Вот результат:

| Основные показатели состояния и развития науки                     |      |      |      |      |      |      |      |      |      |      |  |
|--------------------------------------------------------------------|------|------|------|------|------|------|------|------|------|------|--|
|                                                                    | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |  |
| Количество организаций (предприятий), осуществлявших НИОКР, единиц | 273  | 295  | 390  | 437  | 438  | 421  | 414  | 424  | 412  | 345  |  |

Таблица № 13. Результат фильтрации данных

Обратите внимание, теперь вы видите только строки в данной области. Остальные все еще там, но скрыты. Возможны более сложные фильтры. Например, предположим, что вы хотите видеть только записи, в которых «Численность работников, выполняющих НИОКР» больше или равно 2000, за 2020 год. Вы начинаете с фильтрации численности работников следующим образом:

|    |                                                    |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|----|----------------------------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 7  | Численность работников, выполняющих НИОКР, человек | 18 578 | 18 715 | 18 912 | 19 560 | 17 734 | 18 504 | 15 791 | 17 021 | 18 703 | 20 404 | 23 712 | 25 791 | 24 735 | 22 985 | 22 081 | 22 378 | 23 843 | 22 665 |
| 8  | из них:                                            |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 9  | специалисты-исследователи                          | 9 899  | 10 362 | 11 910 | 13 406 | 11 524 | 10 780 | 10 090 | 10 870 | 11 488 | 13 454 | 17 193 | 18 930 | 18 454 | 17 421 | 17 005 | 17 434 | 17 124 | 18 228 |
| 10 | из них:                                            |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
| 11 | доктора наук                                       | 879    | 1 015  | 1 108  | 1 117  | 1 188  | 1 190  | 1 241  | 1 401  | 1 091  | 1 600  | 2 306  | 1 821  | 1 828  | 1 818  | 1 740  | 1 703  | 1 883  | 1 883  |
| 12 | доктора по профессии                               | -      | -      | -      | -      | -      | -      | -      | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  | 1 000  |
| 13 | доктора философии PhD                              | -      | -      | -      | -      | -      | -      | -      | 121    | 121    | 121    | 121    | 121    | 121    | 121    | 121    | 121    | 121    | 121    |
| 14 | кандидаты наук                                     | 1 782  | 2 742  | 3 004  | 3 147  | 3 058  | 2 880  | 2 734  | 3 011  | 3 398  | 4 804  | 5 187  | 5 130  | 4 736  | 4 541  | 4 365  | 4 240  | 4 329  | 4 329  |

Таблица № 14. Выделение численности работников

| 2015     | 2016     | 2017     | 2018     | 2019     | 2020     |
|----------|----------|----------|----------|----------|----------|
| 69 302,9 | 66 600,1 | 68 884,2 | 72 224,6 | 82 333,1 | 89 028,7 |
| 0,17     | 0,14     | 0,13     | 0,12     | 0,12     | 0,13     |
| 390      | 383      | 386      | 384      | 386      | 396      |
| 24 735   | 22 985   | 22 081   | 22 378   | 21 843   | 22 665   |
| 18 454   | 17 421   | 17 205   | 17 454   | 17 124   | 18 228   |
| 1 821    | 1 828    | 1 818    | 1 740    | 1 703    | 1 883    |
| 549      | 493      | 354      | 336      | 317      | 62       |
| 431      | 456      | 589      | 856      | 1 045    | 1 757    |
| 5 119    | 4 726    | 4 541    | 4 360    | 4 240    | 4 329    |

Таблица № 15. Последовательность действий при сложной фильтрации

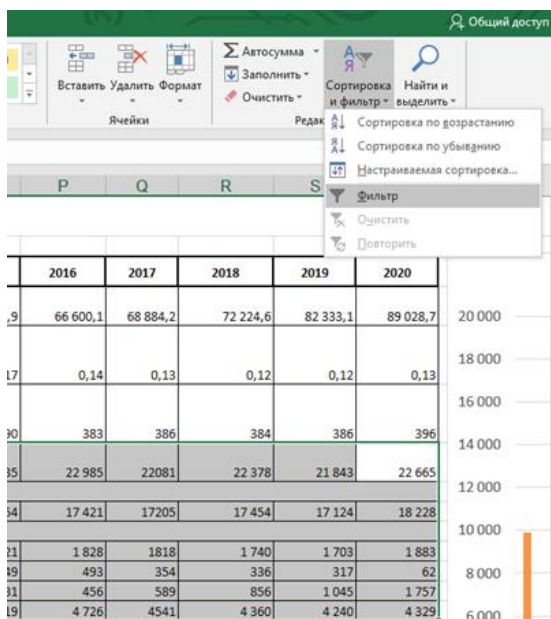


Таблица № 16. Последовательность действий при сложной фильтрации

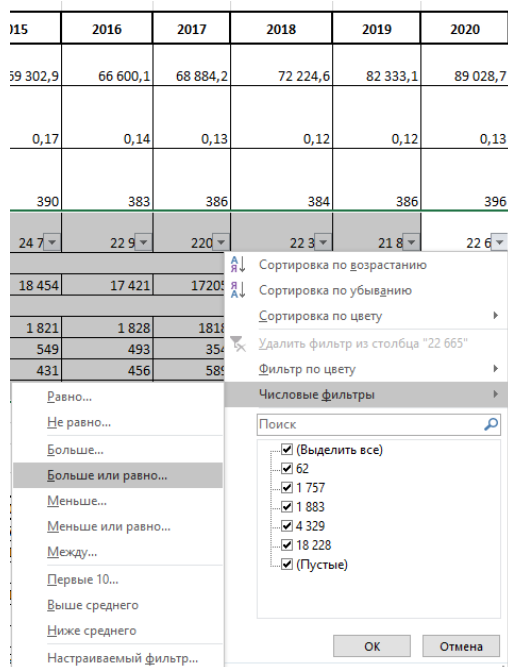


Таблица № 17. Меню сложных фильтров

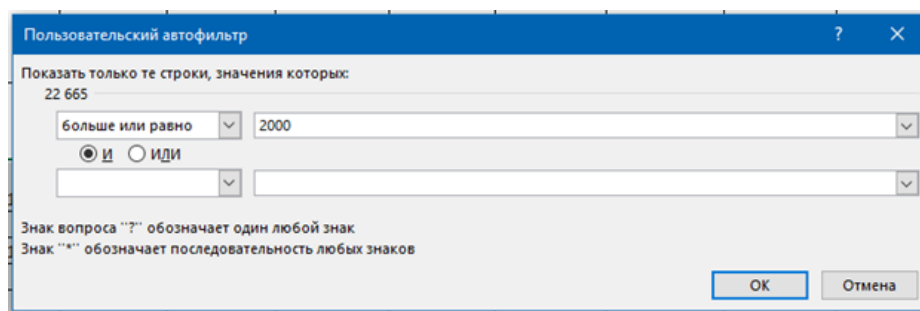


Таблица № 18. Пользовательский автофильтр

Итоговый результат:

|    |                                                    |        |
|----|----------------------------------------------------|--------|
| 7  | Численность работников, выполняющих НИОКР, человек | 22 665 |
| 9  | специалисты-исследователи                          | 18 228 |
| 14 | кандидаты наук                                     | 4 329  |
| 15 |                                                    |        |

Таблица № 19. Итоговый результат фильтрации

## 5.5. Функции

Excel имеет множество встроенных функций, полезных для выполнения математических вычислений и работы с датами и текстом. Например, предположим, что мы хотим подсчитать суммарные затраты на НИОКР в млн. тенге в промежутке с 2003 по 2020 годы. Для этого мы должны перейти к правой части столбца U, пропустить строку и затем ввести эту формулу в ячейку U:4 = СУММ (C4:T4). Знак равенства (=) необходим для всех функций. Двоеточие (:) означает «все числа от ячейки C4 до ячейки T4». После того, как вы нажмете Enter, результат будет следующим:

| Q        | R        | S        | T        | U         |
|----------|----------|----------|----------|-----------|
|          |          |          |          |           |
| 2017     | 2018     | 2019     | 2020     |           |
| 68 884,2 | 72 224,6 | 82 333,1 | 89 028,7 | 877601,87 |
| 0,13     | 0,12     | 0,12     | 0,13     |           |

Таблица № 20. Подсчет суммарных затрат на НИОКР, в млн. тенге.

Было бы очень утомительно повторять это вычисление в каждой из строк данных. К счастью, в Excel есть способ быстро скопировать формулу в столбец ячеек. Для этого осторожно переместите курсор (обычно это большой жирный белый крест) к точке в правом нижнем углу ячейки, содержащей формулу. Когда он окажется в нужном месте, курсор изменится на маленький черный крестик. В этот момент вы можете дважды щелкнуть, и формула скопирует столбец, пока не достигнет пустой ячейки в столбце слева. Это будет результат:

| R        | S        | T        | U         |
|----------|----------|----------|-----------|
|          |          |          |           |
|          |          |          |           |
| 2018     | 2019     | 2020     |           |
| 72 224,6 | 82 333,1 | 89 028,7 | 877601,87 |
| 0,12     | 0,12     | 0,13     | 3,29      |
| 384      | 386      | 396      | 6907      |
| 22 378   | 21 843   | 22 665   | 363259    |

Таблица № 21. Копирование формулы на другие ячейки

Вот еще несколько полезных функций Excel, которые можно использовать аналогичным образом:

(Вы можете складывать, вычитать, умножать или делить, используя символы + - \* и/)

= СРЕДНЕЕ - вычисляет среднее арифметическое столбца или строки чисел.

= МЕДИАНА - находит среднее значение столбца или строки чисел

= COUNT - показывает, сколько элементов находится в столбце или строке

= MAX - сообщает вам наибольшее значение в столбце или строке

= MIN - сообщает вам наименьшее значение в столбце или строке

Есть также множество текстовых функций, которые могут соединять и разделять текстовые строки.

Другие текстовые функции включают:

= ПОИСК - найдет начало желаемой строки текста в более крупной строке.

= LEN - это покажет вам, сколько символов в текстовой строке.

= ВЛЕВО - будет извлекаться указанное вами количество символов, начиная с левого.

= ВПРАВО - это будет извлекать символы, начиная с правого.

= MID - это запустит извлечение с того места, где вы укажете, что он должен начинаться, и получит столько символов, сколько вы укажете.

Вы также можете выполнять арифметические операции с датами, например, вычислять количество дней или лет между двумя датами или часы, минуты и / или секунды между двумя датами.

## 5.6. Горячие клавиши и другие полезные приемы

Чтобы выделить блок данных, поместите курсор где-нибудь в середине ваших данных, а затем одновременно нажмите Ctrl-Shift-звездочка (\*). (Работает только в Windows):

|                                                                     | 2003     | 2004     | 2005     | 2006     | 2007     | 2008     | 2009     | 2010     | 2011     | 2012     | 2013     | 2014     | 2015     | 2016     | 2017     | 2018     | 2019     | 2020     |
|---------------------------------------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Внутренние затраты на НИОКР, млн. тенге                             | 11 643,5 | 14 579,8 | 21 527,4 | 29 790,0 | 29 855,5 | 34 762,0 | 38 989,7 | 33 486,0 | 49 331,0 | 52 253,5 | 61 672,7 | 66 347,0 | 69 302,9 | 66 005,5 | 68 884,2 | 72 224,0 | 82 333,5 | 89 528,7 |
| Доля внутренних затрат на НИОКР от валового внутреннего продукта, % | 0,25     | 0,25     | 0,28     | 0,34     | 0,31     | 0,32     | 0,33     | 0,33     | 0,35     | 0,35     | 0,37     | 0,37     | 0,37     | 0,34     | 0,33     | 0,33     | 0,33     | 0,35     |
| Количество организаций (предприятий), осуществлявших НИОКР, единиц  | 273      | 295      | 380      | 437      | 438      | 433      | 434      | 424      | 412      | 345      | 340      | 381      | 380      | 381      | 380      | 384      | 388      | 386      |
| Полнота данных                                                      | 10 876   | 18 724   | 18 842   | 19 583   | 17 774   | 18 304   | 15 795   | 17 051   | 18 885   | 20 404   | 23 712   | 23 780   | 24 735   | 22 885   | 23 081   | 22 878   | 23 845   | 22 880   |
| Специальность исследователей                                        | 9 899    | 10 362   | 11 703   | 12 424   | 11 524   | 10 780   | 10 093   | 10 870   | 11 089   | 10 494   | 17 192   | 18 930   | 18 454   | 17 423   | 17 029   | 17 454   | 17 134   | 18 138   |
| из них:                                                             |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| дoктoры наук                                                        | 976      | 1 023    | 1 109    | 1 117    | 1 090    | 1 092    | 1 090    | 1 042    | 1 060    | 1 060    | 1 060    | 1 060    | 1 060    | 1 060    | 1 060    | 1 060    | 1 060    | 1 060    |
| дoктoры по профилю                                                  |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| дoктoры философии PhD                                               |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |          |
| кандидаты наук                                                      | 2 782    | 2 740    | 3 058    | 3 147    | 3 208    | 2 862    | 2 738    | 3 012    | 3 288    | 3 428    | 4 812    | 5 024    | 5 118    | 4 728    | 4 942    | 4 865    | 5 245    | 4 528    |

Таблица № 22. Таблица состояния и развития науки РК

Чтобы проверить четыре угла выделенного фрагмента данных, выделите данные (используя только что упомянутый трюк), затем одновременно нажмите control-period (.). И повторить это четыре раза. Каждый раз, когда вы нажимаете клавиши, они переходят в другой угол данных. Это полезно для того, чтобы убедиться, что вы выделили весь фрагмент (и ничего постороннего) перед сортировкой или копированием.

|                                                                     | 2003     | 2004     |
|---------------------------------------------------------------------|----------|----------|
| Внутренние затраты на НИОКР, млн. тенге                             | 11 643,5 | 14 579,8 |
| Доля внутренних затрат на НИОКР от валового внутреннего продукта, % | 0,25     | 0,25     |
| Количество организаций (предприятий), осуществлявших НИОКР, единиц  | 273      | 295      |

Таблица № 23. Пример работы горячей клавиши «Домой»

Закрепить панели — это полезный способ сохранить заголовки вверху страницы, чтобы вы могли видеть их при прокрутке вниз большого набора данных. Поместите курсор в ячейку чуть ниже строки, которую вы хотите зафиксировать (обычно это будет где-то около A2 или A3, в зависимости от того, с какой строки начинаются ваши данные). Затем перейдите в меню «Просмотр» и выберите «Закрепить панели». Он предоставит вам несколько вариантов, например, заморозить верхнюю строку или заморозить один, или несколько столбцов слева, а также верхнюю строку.

Чтобы вернуться к верхней части экрана данных, одновременно нажмите клавиши управления и «Домой». Вы попадете в A1.

|   |    |                                                                     |          |
|---|----|---------------------------------------------------------------------|----------|
|   | C3 |                                                                     | 2003     |
|   | A  | B                                                                   | C        |
| 1 |    |                                                                     |          |
| 2 |    |                                                                     |          |
| 3 |    |                                                                     | 2003     |
| 4 |    | Внутренние затраты на НИОКР, млн. тенге                             | 11 643,5 |
| 5 |    | Доля внутренних затрат на НИОКР от валового внутреннего продукта, % | 0,25     |
| 6 |    | Количество организаций (предприятий), осуществлявших НИОКР, единиц  | 273      |
| 7 |    | Численность работников, выполняющих НИОКР, человек                  | 16 578   |

Таблица № 24. Выделение всего столбца

Чтобы выделить весь столбец, поместите курсор на первую ячейку данных, затем нажмите Ctrl-Shift и клавишу со стрелкой вниз. Он будет выделяться полностью вниз, пока не попадет в пустую строку. Это работает так же для строк, за исключением того, что вы нажимаете Ctrl-Shift и клавишу со стрелкой вправо.

Специальная вставка позволяет избавиться от формул, чтобы остались только ответы. Однако хорошенько подумайте, хотите ли вы это сделать. Иногда оставление формул там дает вам своего рода «бумажный след» того, что вы сделали, или позволяет другим увидеть, что вы сделали. Чтобы использовать Специальную вставку, выделите строки или столбцы, содержащие формулы. Скопируйте данные, используя Ctrl-C или кнопку копирования. Затем поместите курсор в то место, куда вы хотите вставить данные, и щелкните правой кнопкой мыши, затем выберите «Специальная вставка». Появится небольшое окно: в разделе Paste сверху выберите «Значения» и нажмите ОК.

|    |    |                                                                     |          |        |          |
|----|----|---------------------------------------------------------------------|----------|--------|----------|
|    | C2 |                                                                     |          |        |          |
|    | A  | B                                                                   | C        | D      | F        |
| 1  |    |                                                                     |          |        |          |
| 2  |    |                                                                     |          |        |          |
| 3  |    |                                                                     | 2003     |        | 2006     |
| 4  |    | Внутренние затраты на НИОКР, млн. тенге                             | 11 643,5 |        | 24 799,5 |
| 5  |    | Доля внутренних затрат на НИОКР от валового внутреннего продукта, % | 0,25     |        | 0,24     |
| 6  |    | Количество организаций (предприятий), осуществлявших НИОКР, единиц  | 273      |        | 437      |
| 7  |    | Численность работников, выполняющих НИОКР, человек                  | 16 578   | 16 715 | 18 912   |
| 8  |    | из них:                                                             |          |        |          |
| 9  |    | специалисты-исследователи                                           | 9 899    | 10 382 | 11 910   |
| 10 |    | из них:                                                             |          |        |          |
| 11 |    | доктора наук                                                        | 979      | 1 013  | 1 106    |
| 12 |    | доктора по профилю                                                  | -        | -      | -        |
| 13 |    | доктора философии PhD                                               | -        | -      | -        |
| 14 |    | кандидаты наук                                                      | 2 782    | 2 740  | 3 018    |

Таблица № 25. Функция «Скрыть» столбцы

Скрыть столбцы: вы можете скрыть столбцы, чтобы убрать их с дороги или избежать их печати, выделив нужные столбцы (щелкните букву сверху столбца) и щелкните правой кнопкой мыши, чтобы выбрать «Скрыть столбцы». Столбцы исчезнут. Чтобы вернуть их, выделите два столбца по обе стороны от отсутствующих, щелкните правой кнопкой мыши и выберите «Показать столбцы».

|    | A | B                                                                   | D        | E        |
|----|---|---------------------------------------------------------------------|----------|----------|
| 1  |   |                                                                     |          |          |
| 2  |   |                                                                     |          |          |
| 3  |   |                                                                     | 2004     | 2005     |
| 4  |   | Внутренние затраты на НИОКР, млн. тенге                             | 14 579,8 | 21 527,4 |
| 5  |   | Доля внутренних затрат на НИОКР от валового внутреннего продукта, % | 0,25     | 0,28     |
| 6  |   | Количество организаций (предприятий), осуществлявших НИОКР, единиц  | 295      | 390      |
| 7  |   | Численность работников, выполняющих НИОКР, человек                  | 16 715   | 18 912   |
| 8  |   | из них:                                                             |          |          |
| 9  |   | специалисты-исследователи                                           | 10 382   | 11 910   |
| 10 |   | из них:                                                             |          |          |
| 11 |   | доктора наук                                                        | 1 013    | 1 106    |
| 12 |   | доктора по профилю                                                  | -        | -        |
| 13 |   | доктора философии PhD                                               |          |          |
| 14 |   | кандидаты наук                                                      | 2 740    | 3 018    |

Таблица № 26. Результат работы функции

|   | A | B                                                                   | D      | E      | F        |
|---|---|---------------------------------------------------------------------|--------|--------|----------|
| 1 |   |                                                                     |        |        |          |
| 2 |   |                                                                     |        |        |          |
| 3 |   |                                                                     |        |        | 2006     |
| 4 |   | Внутренние затраты на НИОКР, млн. тенге                             |        |        | 24 799,9 |
| 5 |   | Доля внутренних затрат на НИОКР от валового внутреннего продукта, % |        |        | 0,24     |
| 6 |   | Количество организаций (предприятий), осуществлявших НИОКР, единиц  |        |        | 437      |
| 7 |   | Численность работников, выполняющих НИОКР, человек                  | 16 715 | 18 912 | 19 563   |

Таблица № 27. Функция «Показать» скрытые столбцы

Добавление столбцов или строк: чтобы добавить новый столбец в середину ваших данных, щелкните букву столбца непосредственно СПРАВА от того места, где вы хотите создать новый столбец. Убедитесь, что выделен весь столбец (если вы щелкните букву над столбцом, Excel автоматически выделит весь столбец). Затем щелкните правой кнопкой мыши и выберите «вставить». Чтобы добавить новую строку, щелкните номер строки НИЖЕ, где вы хотите создать новую. Убедитесь, что выделена вся строка, затем щелкните правой кнопкой мыши и выберите «вставить».

Рабочие листы: вы можете переключаться между различными рабочими листами в книге Excel, используя вкладки в нижнем левом углу. У них есть имена по умолчанию: «Лист 1» или «Лист 2» и т. Д. Чтобы изменить имя, нажмите на нем дважды, и оно станет черным - затем вы можете начать вводить новое имя. Чтобы добавить новый рабочий лист, вы увидите небольшую кнопку справа от вкладок, чтобы добавить рабочий лист. Вы можете перемещать листы, чтобы они отображались в другом порядке.

|                                                                                                                               |          |          |           |           |
|-------------------------------------------------------------------------------------------------------------------------------|----------|----------|-----------|-----------|
| 19                                                                                                                            | 2003     | 2004     | 2005      | 2006      |
| 20                                                                                                                            | 65 020,4 | 74 718,5 | 120 408,4 | 156 039,9 |
| 21                                                                                                                            | 1,1      | 9,8      | 6 278,1   | 31 947,9  |
| 22                                                                                                                            | 6 493,5  | 8 076,5  | 1 249,8   | 6 450,4   |
| 23                                                                                                                            | 501,4    | 897,8    | 6 571,9   | 10 370,2  |
| 24                                                                                                                            | 0,0      | 9,2      | 334,9     | 1 940,1   |
| 25                                                                                                                            | 1 827,7  | 391,4    | 1 181,0   | 2 327,9   |
| 26                                                                                                                            | 848,1    | 378,0    | 1 135,0   | 1 575,1   |
| 27                                                                                                                            | 34 798,0 | 37 483,8 | 56 239,0  | 59 871,8  |
| 28                                                                                                                            | 203,3    | 89,7     | 184,7     | 861,6     |
| 29                                                                                                                            | 0,0      | 4,6      | 3,8       | 2,0       |
| 30                                                                                                                            | 1 126,2  | 3 005,2  | 8 576,2   | 6 639,6   |
| 31                                                                                                                            | 8 765,3  | 7 691,4  | 10 147,8  | 1 249,7   |
| <div> <div>рус</div> <div>+</div> <div>Добавить</div> </div> <div> <div>Готово</div> <div>Лист 1</div> <div>Лист</div> </div> |          |          |           |           |

Таблица № 28. Добавление новых листов

## 5.7. Визуализация данных в Excel

Вы можете отображать свои отчеты анализа данных в Excel несколькими способами. Однако если результаты анализа данных можно визуализировать в виде диаграмм, которые выделяют заметные точки в данных, ваша аудитория может быстро понять, что вы хотите спроецировать в данные. Это также оставляет хорошее влияние на ваш стиль презентации. В этой главе вы узнаете, как использовать диаграммы Excel и функции форматирования Excel на диаграммах, которые позволяют с акцентом представлять результаты анализа данных. Далее, на основе полученных данных, будет дана итоговая визуализация статистических данных по Республике Казахстан в сфере науки и инноваций и приведена их трактовка.

## 5.8. Визуализация данных с помощью диаграмм

В Excel диаграммы используются для графического представления любого набора данных. Диаграмма — это визуальное представление данных, в котором данные представлены такими символами, как столбцы на линейчатой диаграмме или линии на линейной диаграмме. Excel предоставляет вам множество типов диаграмм, и вы можете выбрать тот, который подходит вашим данным, или вы можете использовать опцию Excel Recommended Charts, чтобы просмотреть диаграммы, настроенные для ваших данных, и выбрать один из них.

В этой главе вы познакомитесь с различными методами, которые вы можете использовать с диаграммами Excel, чтобы более эффективно освещать результаты анализа данных. Процесс визуализации можно разделить на несколько основных категорий:

1. Создание комбинационных диаграмм
2. Различающие серии и оси категорий
3. Элементы диаграммы и стили диаграммы
4. Метки данных
5. Быстрый макет
6. Диаграмма

Ганта

Инструментарий для визуализации данных в MS Excel представляет собой достаточно мощное средство для визуальной интерпретации информации. Он включает в себя возможность создания смежных таблиц для сопоставления различных данных, создание различных видов диаграмм, инфографик и чартов.

Условное форматирование. Первый способ визуализировать данные – это раскрасить выбранные ячейки, следуя определенной закономерности или правилу. Условное форматирование позволяет подобное.

|                        | 2003     | 2004     | 2005      | 2006      | 2007      | 2008      | 2009     | 2010      | 2011      | 2012      | 2013      | 2014      | 2015      | 2016      | 2017      | 2018        | 2019        | 2020        |
|------------------------|----------|----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|-------------|-------------|
| Республика Казахстан   | 65 626,4 | 24 718,5 | 120 408,4 | 154 039,9 | 152 108,8 | 113 118,1 | 82 187,4 | 142 168,4 | 215 962,7 | 379 085,6 | 519 262,1 | 580 286,0 | 677 196,7 | 845 774,7 | 884 714,9 | 1 084 987,4 | 1 111 586,1 | 1 715 588,2 |
| Алматы                 | 0,1      | 9,8      | 6 176,1   | 11 547,5  | 40 275,9  | 4 124,9   | 2 070,0  | 6 959,4   | 9 812,5   | 19 912,1  | 18 261,7  | 19 911,0  | 13 217,2  | 21 544,0  | 15 712,9  | 25 644,8    | 17 790,0    | 56 584,5    |
| Актюбинская            | 6 493,1  | 8 976,5  | 1 149,8   | 6 402,4   | 11 080,0  | 14 076,9  | 4 428,1  | 9 791,1   | 10 880,9  | 6 542,4   | 8 101,6   | 4 404,4   | 1 818,0   | 19 680,2  | 19 442,0  | 44 299,9    | 51 421,7    | 59 526,1    |
| Авлятинская            | 301,4    | 897,8    | 6 171,9   | 10 170,2  | 6 577,4   | 104,0     | 426,1    | 321,0     | 5 408,1   | 13 288,0  | 13 151,8  | 14 408,9  | 15 699,2  | 12 600,0  | 12 624,0  | 18 101,2    | 20 443,4    | 59 171,0    |
| Атырауская             | 0,0      | 9,2      | 134,8     | 1 940,1   | 186,9     | 1 196,4   | 15,1     | 126,1     | 1 628,1   | 4 772,2   | 18 076,1  | 18 455,1  | 7 506,1   | 7 419,4   | 3 748,0   | 8 619,8     | 7 136,1     | 482 420,1   |
| Восточно-Казахстанская | 1 817,7  | 291,4    | 1 181,0   | 2 161,9   | 2 341,0   | 291,4     | 728,7    | 24 904,9  | 4 199,1   | 9 028,1   | 9 996,1   | 1 134,7   | 8 407,1   | 18 111,2  | 23 288,7  | 24 713,4    | 21 471,0    |             |
| Жамбылская             | 848,1    | 1 176,0  | 1 115,0   | 1 575,1   | 2 241,4   | 9 186,1   | 2 190,4  | 723,8     | 11 271,8  | 19 181,2  | 19 617,4  | 25 293,1  | 23 143,1  | 34 104,1  | 30 854,7  | 66 781,2    | 77 091,2    | 67 430,7    |
| Карагандинская         | 14 798,0 | 17 481,8 | 54 219,0  | 19 871,8  | 17 886,5  | 18 471,8  | 14 412,4 | 14 817,7  | 14 388,8  | 30 891,1  | 53 711,2  | 21 576,1  | 18 442,0  | 21 127,2  | 31 048,0  | 54 178,0    | 74 007,0    | 145 720,4   |
| Костанайская           | 201,1    | 85,7     | 184,7     | 881,4     | 1 144,4   | 4 084,0   | 1 846,1  | 1 966,4   | 12 451,0  | 29 701,7  | 37 128,9  | 57 813,1  | 47 252,4  | 45 195,9  | 61 101,0  | 114 114,9   | 111 088,1   | 849 112,4   |
| Кызылординская         | 0,0      | 4,4      | 3,0       | 1,0       | 6,0       | 10,0      | 46,7     | 2 341,0   | 1 440,0   | 4 441,7   | 4 761,2   | 4 902,1   | 4 201,0   | 5 101,8   | 4 401,0   | 14 421,2    | 19 101,0    |             |
| Магистральная          | 1 126,2  | 1 900,2  | 8 576,1   | 6 409,8   | 7 051,0   | 5 901,0   | 113,0    | 213,1     | 418,4     | 1 409,4   | 1 144,4   | 1 144,4   | 1 144,4   | 1 144,4   | 1 144,4   | 1 144,4     | 1 144,4     | 1 144,4     |
| Павлодарская           | 8 763,1  | 7 891,4  | 10 147,8  | 1 149,1   | 1 001,7   | 10 062,2  | 10 420,1 | 72 192,4  | 72 179,0  | 17 420,0  | 81 076,0  | 1 818,0   | 9 501,0   | 177 881,0 | 250 011,0 | 44 101,0    | 44 101,0    | 96 984,4    |
| Северо-Казахстанская   | 279,7    | 2 215,9  | 1 430,4   | 2 789,4   | 919,1     | 1 128,4   | 2 791,0  | 2 880,4   | 1 449,5   | 4 098,1   | 14 028,0  | 18 505,4  | 11 751,8  | 11 111,8  | 13 804,9  | 9 196,5     | 8 601,1     | 24 066,0    |
| Семейская              | 109,0    | 101,0    | 4 177,1   | 1 841,0   | 407,1     | 701,8     | 1 844,4  | 2 149,1   | 10 120,2  | 14 018,1  | 19 414,0  | 9 612,0   | 10 101,0  | 18 111,0  | 13 140,0  | 13 140,0    | 13 140,0    | 14 847,4    |
| Восточно-Казахстанская | 4 144,1  | 4 144,1  | 11 411,9  | 13 747,1  | 29 012,4  | 5 118,2   | 6 919,1  | 13 854,1  | 10 950,1  | 19 101,1  | 109 178,9 | 17 778,9  | 13 420,9  | 54 299,5  | 80 472,0  | 174 068,0   | 211 418,8   | 128 147,0   |
| г. Нур-Султан          | 0,0      | 0,0      | 2 779,0   | 2 886,8   | 13,0      | 13,2      | 11,1     | 72,2      | 1 818,8   | 4 787,0   | 119 823,4 | 125 507,0 | 112 219,8 | 120 119,8 | 149 277,5 | 112 184,2   | 119 468,7   | 67 114,0    |
| г. Алматы              | 3 146,1  | 4 291,4  | 7 627,1   | 7 877,1   | 7 616,1   | 11 085,4  | 7 711,1  | 13 088,4  | 10 801,4  | 12 576,1  | 12 576,1  | 22 088,1  | 18 876,9  | 17 189,1  | 29 181,1  | 30 216,4    | 48 948,4    | 54 911,2    |
| г. Шымкент             | 493,1    | 9 811,1  | 781,1     | 1 841,4   | 2 107,1   | 748,1     | 781,1    | 3 101,1   | 9 248,1   | 8 576,1   | 13 107,1  | 20 546,1  | 10 916,1  | 19 489,1  | 112 091,4 | 101 478,4   | 138 048,4   | 120 148,4   |

Таблица № 29. Объем инновационной продукции

Для примера, выделим данные за все годы, нажмем «Формат» – «Условное форматирование». Доступный нам функционал это правила отбора и выделения ячеек, непосредственно само выделение, возможное в виде гистограмм, цветовых шкал, наборов знаков, таких как стрелки, кружки, крестики и т.д. Попробуем использовать гистограмму и цветовую шкалу.

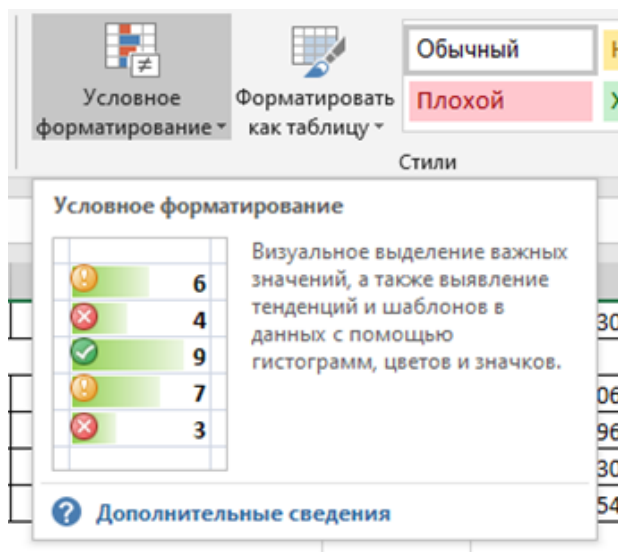


Таблица № 30. Условное форматирование

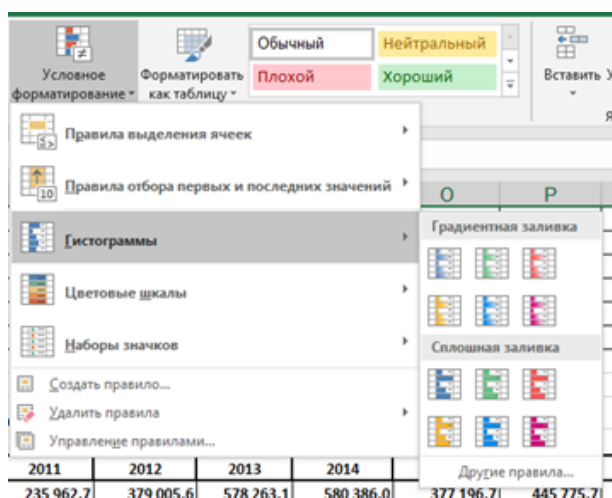


Таблица № 31. Меню условного форматирования «Гистограммы»

|                        | 2002     | 2003     | 2004      | 2005      | 2006      | 2007      | 2008     | 2009      | 2010      | 2011      | 2012      | 2013      | 2014      | 2015      | 2016      | 2017        | 2018        | 2019        | 2020     |
|------------------------|----------|----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|-------------|-------------|----------|
| Республика Казахстан   | 65 430,6 | 74 728,5 | 520 408,4 | 156 039,9 | 152 508,6 | 111 513,3 | 82 587,8 | 142 566,8 | 295 962,7 | 379 805,6 | 579 283,3 | 580 386,0 | 377 396,7 | 445 775,7 | 844 736,9 | 1 064 067,4 | 1 113 566,5 | 1 735 506,3 |          |
| Алматы                 | 5,1      | 9,0      | 6 278,1   | 31 947,9  | 41 235,6  | 4 316,9   | 3 039,0  | 6 939,6   | 9 822,5   | 19 902,3  | 18 301,6  | 13 217,2  | 21 036,2  | 15 721,9  | 25 664,4  | 17 793,0    | 56 366,5    |             |          |
| Актюбинская            | 6 493,5  | 8 076,5  | 1 249,8   | 4 450,4   | 11 680,0  | 14 056,9  | 4 428,3  | 9 792,3   | 16 880,9  | 6 542,4   | 8 300,6   | 4 434,4   | 1 838,8   | 16 660,2  | 44 299,5  | 51 421,7    | 59 526,1    |             |          |
| Авлатовская            | 502,4    | 897,8    | 6 571,9   | 10 370,3  | 8 577,4   | 334,0     | 425,5    | 512,0     | 5 498,1   | 13 388,0  | 13 151,8  | 16 606,9  | 15 699,2  | 12 630,0  | 12 434,2  | 18 351,2    | 20 443,6    | 59 571,0    |          |
| Алматы                 | 0,0      | 9,2      | 334,9     | 1 940,1   | 186,9     | 1 598,4   | 55,3     | 126,1     | 2 828,1   | 4 772,2   | 36 078,2  | 18 655,3  | 7 906,1   | 7 429,4   | 5 768,0   | 8 818,8     | 7 516,3     | 402 426,3   |          |
| Западно-Казахстанская  | 1 827,7  | 391,4    | 1 181,0   | 2 327,9   | 2 343,8   | 390,6     | 728,7    | 2 190,4   | 723,8     | 11 251,8  | 19 281,2  | 19 637,4  | 25 250,3  | 18 383,7  | 34 154,1  | 50 854,7    | 66 782,3    | 77 292,5    | 67 430,7 |
| Жамбылская             | 848,1    | 578,0    | 1 135,0   | 1 575,1   | 2 241,6   | 6 286,3   | 2 190,4  | 723,8     | 11 251,8  | 19 281,2  | 19 637,4  | 25 250,3  | 18 383,7  | 34 154,1  | 50 854,7  | 66 782,3    | 77 292,5    | 67 430,7    |          |
| Карагандинская         | 34 796,0 | 37 483,8 | 56 239,0  | 59 871,8  | 37 886,5  | 16 473,8  | 14 612,4 | 14 897,7  | 34 388,8  | 30 891,5  | 50 731,2  | 21 578,1  | 18 443,3  | 31 327,2  | 32 048,0  | 54 778,0    | 74 007,0    | 145 720,6   |          |
| Костанайская           | 203,3    | 85,7     | 184,7     | 861,6     | 1 344,8   | 4 064,0   | 1 848,3  | 1 966,9   | 12 453,0  | 29 789,7  | 35 728,9  | 37 633,9  | 47 252,6  | 45 095,9  | 10 502,4  | 124 014,9   | 211 088,3   | 149 012,4   |          |
| Кокшетауская           | 0,0      | 4,6      | 3,6       | 2,0       | 8,0       | 30,3      | 66,7     | 2 281,3   | 3 640,0   | 8 641,7   | 4 763,2   | 6 990,2   | 6 291,2   | 3 505,8   | 6 401,7   | 16 425,2    | 19 925,7    |             |          |
| Мангистауская          | 1 126,2  | 3 005,2  | 8 576,2   | 6 639,6   | 7 351,0   | 5 002,9   | 133,0    | 233,1     | 618,6     | 3 609,0   | 1 395,4   | 1 546,6   | 1 234,6   | 506,4     | 294,9     | 651,0       | 7 971,3     | 5 317,2     |          |
| Павлодарская           | 8 765,3  | 7 651,4  | 10 147,8  | 1 249,7   | 2 003,7   | 10 080,2  | 35 420,1 | 72 562,4  | 71 279,0  | 97 620,0  | 83 368,0  | 83 670,6  | 1 838,3   | 9 320,0   | 177 881,5 | 250 032,0   | 44 503,7    | 96 984,4    |          |
| Северо-Казахстанская   | 279,7    | 2 235,9  | 1 430,4   | 2 789,4   | 955,9     | 3 228,6   | 2 751,0  | 2 080,6   | 3 469,5   | 6 098,3   | 16 038,0  | 16 500,4  | 11 753,8  | 11 112,8  | 13 804,9  | 9 296,5     | 8 632,1     | 28 266,0    |          |
| Туркестанская          | 389,8    | 750,3    | 4 277,2   | 3 881,0   | 607,3     | 701,3     | 3 640,4  | 1 509,7   | 10 105,2  | 34 028,1  | 19 870,4  | 9 612,4   | 10 535,0  | 14 727,1  | 13 140,0  | 13 375,7    | 13 797,5    | 14 847,6    |          |
| Восточно-Казахстанская | 4 544,3  | 4 188,2  | 11 411,9  | 13 747,5  | 26 025,4  | 5 118,2   | 6 939,5  | 13 854,5  | 31 592,5  | 99 112,1  | 209 176,9 | 17 776,9  | 13 420,9  | 54 296,5  | 80 472,0  | 134 068,8   | 223 618,8   | 158 747,0   |          |
| Город-Сатпау           | 0,0      | 0,0      | 2 779,0   | 2 866,8   | 11,0      | 13,2      | 31,1     | 72,2      | 1 818,4   | 4 787,0   | 125 923,4 | 125 567,0 | 111 239,6 | 125 159,3 | 149 270,5 | 112 146,2   | 129 468,7   | 87 114,0    |          |
| Г.А.Ялы                | 5 146,1  | 4 093,6  | 7 827,3   | 7 877,5   | 7 836,9   | 11 080,8  | 7 711,3  | 12 686,8  | 10 602,4  | 12 578,3  | 12 564,9  | 21 588,8  | 38 876,9  | 17 186,3  | 36 183,3  | 30 228,4    | 48 948,4    | 56 491,2    |          |
| Г.Самарканд            | 495,9    | 5 813,3  | 780,4     | 1 661,4   | 2 337,7   | 748,1     | 782,1    | 3 350,2   | 5 268,8   | 8 570,6   | 13 307,0  | 35 540,8  | 50 900,6  | 19 689,0  | 112 091,6 | 101 678,0   | 136 084,8   | 250 588,3   |          |

Таблица № 32. Условное форматирование «Гистограмма»

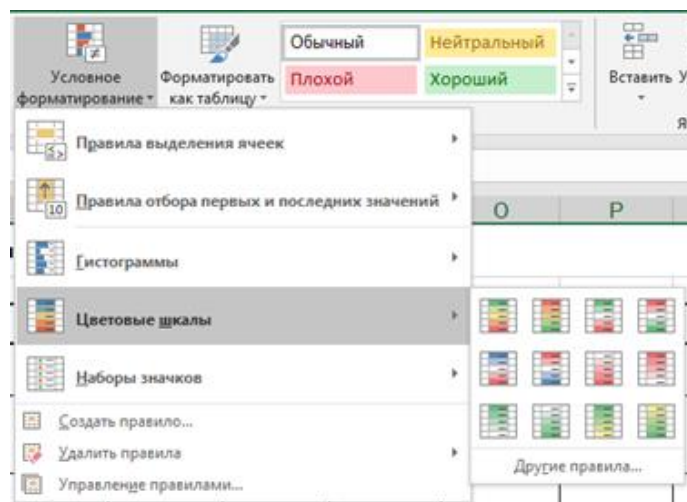


Таблица № 33. Меню условного форматирования «Цветовые шкалы»

|                      | 2001     | 2004     | 2005      | 2006      | 2007      | 2008      | 2009     | 2010      | 2011      | 2012      | 2013      | 2014      | 2015      | 2016      | 2017      | 2018        | 2019        | 2020        |
|----------------------|----------|----------|-----------|-----------|-----------|-----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|-------------|-------------|
| Республика Казахстан | 85 620,4 | 74 718,5 | 120 408,4 | 150 639,9 | 152 500,6 | 111 511,1 | 82 187,4 | 142 186,8 | 235 963,7 | 579 005,6 | 579 263,1 | 580 086,0 | 177 936,7 | 445 775,7 | 844 734,9 | 1 064 067,4 | 1 113 566,5 | 1 715 500,1 |
| Алматы               | 3,1      | 3,0      | 6 276,1   | 11 947,9  | 41 210,6  | 4 318,9   | 1 039,0  | 6 959,6   | 9 822,5   | 19 902,1  | 18 205,7  | 11 802,6  | 13 217,2  | 22 036,2  | 19 723,9  | 25 644,6    | 17 793,0    | 56 366,5    |
| Актюбинская          | 6 493,5  | 6 176,5  | 1 349,8   | 6 450,4   | 11 680,0  | 14 194,9  | 4 420,1  | 9 751,1   | 10 880,5  | 6 542,4   | 8 300,4   | 4 454,4   | 1 838,8   | 10 660,2  | 19 442,0  | 44 299,9    | 51 421,7    | 59 026,1    |
| Амурская             | 182,4    | 89,9     | 4 471,9   | 10 170,1  | 9 177,4   | 184,0     | 825,5    | 521,0     | 1 486,1   | 11 264,0  | 13 151,6  | 16 458,4  | 19 499,2  | 13 640,0  | 12 424,2  | 18 151,2    | 20 443,6    | 59 571,0    |
| Атырауская           | 6,0      | 3,2      | 316,9     | 1 546,1   | 1 88,9    | 1 159,4   | 35,2     | 126,1     | 1 428,8   | 4 772,0   | 10 076,2  | 18 455,1  | 7 504,1   | 7 424,4   | 9 768,0   | 8 819,8     | 7 536,3     | 402 420,3   |
| Байель-Кавказская    | 1 827,7  | 391,4    | 1 183,0   | 2 327,9   | 2 343,8   | 390,6     | 728,7    | 24 804,9  | 4 899,8   | 9 009,5   | 5 996,5   | 3 334,7   | 3 407,1   | 18 123,1  | 23 398,7  | 24 713,4    | 21 671,3    |             |
| Белгородская         | 848,1    | 378,0    | 1 135,0   | 1 575,1   | 2 343,8   | 6 288,1   | 2 190,4  | 722,8     | 11 251,8  | 19 181,2  | 19 627,4  | 25 250,1  | 13 163,7  | 34 104,0  | 50 854,7  | 66 782,3    | 77 092,5    | 67 430,7    |
| Белгородская         | 14 798,0 | 17 483,8 | 16 230,0  | 19 871,8  | 17 888,5  | 16 471,8  | 14 412,4 | 14 897,7  | 14 188,8  | 80 893,5  | 53 713,2  | 21 578,1  | 18 442,5  | 31 127,2  | 12 048,0  | 54 778,0    | 74 007,0    | 145 720,6   |
| Белгородская         | 251,1    | 85,7     | 184,7     | 861,4     | 1 144,6   | 4 084,0   | 1 848,3  | 1 966,9   | 12 411,0  | 29 769,7  | 35 728,9  | 57 613,9  | 47 252,6  | 45 095,9  | 91 502,6  | 124 014,9   | 211 088,3   | 349 012,4   |
| Белгородская         | 5,0      | 4,6      | 9,8       | 6,8       | 4,0       | 4,0       | 4,0      | 4,0       | 4,0       | 4,0       | 4,0       | 4,0       | 4,0       | 4,0       | 4,0       | 4,0         | 4,0         | 4,0         |
| Белгородская         | 1 124,2  | 3 395,2  | 6 576,1   | 6 419,8   | 7 111,3   | 5 202,5   | 1 111,3  | 1 111,3   | 1 111,3   | 1 111,3   | 1 111,3   | 1 111,3   | 1 111,3   | 1 111,3   | 1 111,3   | 1 111,3     | 1 111,3     | 1 111,3     |
| Белгородская         | 6 765,7  | 7 481,4  | 10 147,4  | 1 349,7   | 2 902,7   | 88 064,2  | 10 430,1 | 72 582,4  | 79 279,0  | 97 620,0  | 83 368,0  | 83 970,4  | 1 858,8   | 9 520,0   | 177 880,0 | 290 032,0   | 44 503,7    | 96 984,4    |
| Белгородская         | 276,7    | 2 135,9  | 1 430,4   | 2 789,4   | 959,9     | 2 123,6   | 2 751,0  | 2 880,4   | 1 489,9   | 6 068,3   | 14 028,0  | 16 500,4  | 11 751,8  | 11 112,8  | 11 804,9  | 9 196,1     | 8 652,1     | 26 066,0    |
| Белгородская         | 189,8    | 250,0    | 4 277,2   | 1 861,0   | 807,9     | 703,8     | 1 444,4  | 1 506,7   | 10 105,2  | 14 016,1  | 19 870,4  | 9 612,6   | 10 535,0  | 14 727,1  | 13 140,0  | 13 175,7    | 13 797,5    | 14 847,6    |
| Белгородская         | 4 144,3  | 4 388,2  | 11 411,9  | 13 747,5  | 20 013,4  | 5 113,2   | 6 919,5  | 13 454,3  | 33 192,1  | 89 132,1  | 109 378,9 | 97 778,9  | 11 430,9  | 54 295,5  | 80 472,0  | 174 068,8   | 223 618,8   | 116 747,0   |
| Белгородская         | 10       | 10       | 2 779,0   | 2 884,9   | 18,0      | 18,0      | 18,0     | 18,0      | 18,0      | 18,0      | 18,0      | 18,0      | 18,0      | 18,0      | 18,0      | 18,0        | 18,0        | 18,0        |
| Белгородская         | 5 148,1  | 4 988,4  | 7 827,1   | 7 877,5   | 7 036,9   | 11 085,4  | 7 711,5  | 12 486,4  | 10 401,4  | 12 576,1  | 12 584,9  | 22 084,6  | 18 876,6  | 17 188,1  | 26 183,3  | 37 228,4    | 48 148,4    | 56 491,2    |
| Белгородская         | 490,8    | 5 813,3  | 786,4     | 1 941,4   | 2 381,7   | 148,4     | 762,1    | 3 350,2   | 5 268,8   | 8 570,4   | 13 307,0  | 15 140,4  | 10 930,6  | 19 885,0  | 112 093,6 | 101 978,0   | 136 084,8   | 150 588,3   |

Таблица № 34. Условное форматирование «Цветовые шкалы»

| 2018        | 2019        | 2020        |
|-------------|-------------|-------------|
| 1 064 067,4 | 1 113 566,5 | 1 715 500,1 |
| 25 644,6    | 17 793,0    | 56 366,5    |
| 44 299,9    | 51 421,7    | 59 026,1    |
| 18 351,2    | 20 443,6    | 59 571,0    |
| 8 819,8     | 7 536,3     | 402 420,3   |
| 23 398,7    | 24 713,4    | 21 671,3    |
| 66 782,3    | 77 092,5    | 67 430,7    |
| 54 778,0    | 74 007,0    | 145 720,6   |
| 124 014,9   | 211 088,3   | 349 012,4   |
| 6 401,7     | 16 425,2    | 19 925,7    |
| 651,0       | 7 971,3     | 5 317,2     |
| 250 032,0   | 44 503,7    | 96 984,4    |
| 9 396,5     | 8 652,1     | 26 066,0    |
| 13 375,7    | 13 797,5    | 14 847,6    |
| 174 068,8   | 223 618,8   | 116 747,0   |
| 112 146,2   | 129 468,7   | 67 314,0    |
| 30 228,4    | 48 948,4    | 56 491,2    |
| 101 678,0   | 136 084,8   | 150 588,3   |

| 2018        | 2019        | 2020        |
|-------------|-------------|-------------|
| 1 064 067,4 | 1 113 566,5 | 1 715 500,1 |
| 25 644,6    | 17 793,0    | 56 366,5    |
| 44 299,9    | 51 421,7    | 59 026,1    |
| 18 351,2    | 20 443,6    | 59 571,0    |
| 8 819,8     | 7 536,3     | 402 420,3   |
| 23 398,7    | 24 713,4    | 21 671,3    |
| 66 782,3    | 77 092,5    | 67 430,7    |
| 54 778,0    | 74 007,0    | 145 720,6   |
| 124 014,9   | 211 088,3   | 349 012,4   |
| 6 401,7     | 16 425,2    | 19 925,7    |
| 651,0       | 7 971,3     | 5 317,2     |
| 250 032,0   | 44 503,7    | 96 984,4    |
| 9 396,5     | 8 652,1     | 26 066,0    |
| 13 375,7    | 13 797,5    | 14 847,6    |
| 174 068,8   | 223 618,8   | 116 747,0   |
| 112 146,2   | 129 468,7   | 67 314,0    |
| 30 228,4    | 48 948,4    | 56 491,2    |
| 101 678,0   | 136 084,8   | 150 588,3   |

Таблица № 35. Выделение столбца

Таблица № 36. Условное форматирование

Условное форматирование можно применять и к отдельным столбцам. Применим данный способ к нашим данным, а именно к объему инновационной продукции. Выясним, какая область является лидером по объему выпущенной продукции в каждом году. Для этого необходимо сперва выделить интересующий вас столбец, а после в настройках форматирования выбрать цветовую шкалу от красного до зеленого.

### Мини-диаграммы

Еще один способ быстро оценить большое количество данных – построить мини-диаграммы в ячейке рядом с данными. Сделать это помогает функция SPARKLINE. Чтобы воспользоваться ей, нужно прописать в ячейке рядом с данными =SPARKLINE(), а в скобках указать диапазон данных. Например: =SPARKLINE(B3:F3). Тогда в этой ячейке появится мини-график, иллюстрирующий колебания в данных. Также спарклайны можно вызвать через соответствующий пункт меню в разделе «Вставка»:

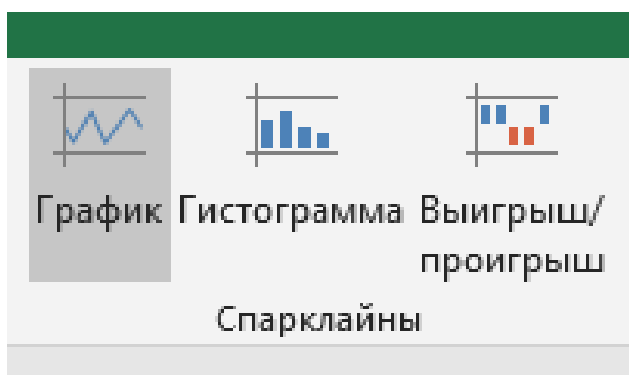


Рисунок № 11. Меню спарклайнов

Можно задавать дополнительные параметры для этой функции: например, цвет и толщину линии.

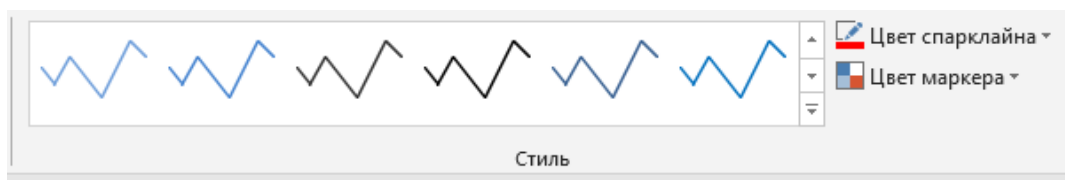


Рисунок № 12. Меню стилей сарклайнов

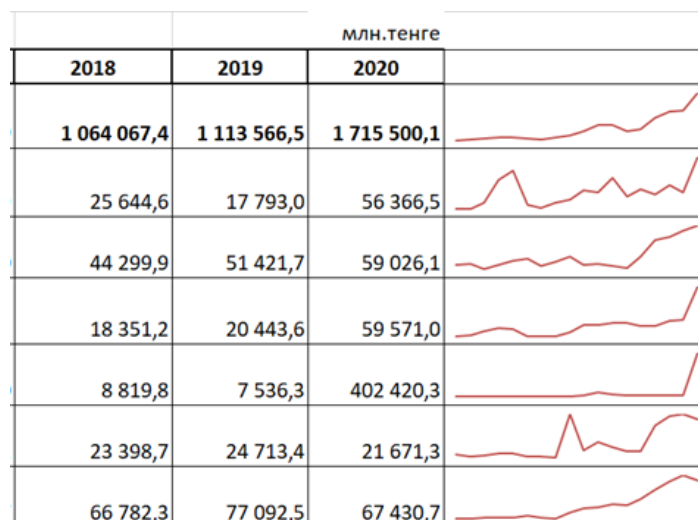


Рисунок № 13. Визуализация данных спарклайнами

### Редактор диаграмм

После того, как мы проанализировали наши данные, можем переходить к их графическому представлению в виде публикации. Визуализацию в виде диаграмм можно производить как в Microsoft Excel, там и с помощью Google Spreadsheets. Последняя является удобным средством для визуализации в тех случаях, когда вам надо показать интерактивные графики в онлайн среде, или во время конференций с общим доступом. Оба вариант полностью совместимы между собой, и в большинстве случаев переход между двумя решениями является плавным и интуитивно понятным. Мы же будем использовать стандартный инструментарий для создания диаграмм MS Excel. Чтобы построить любую диаграмму, надо выделить таблицу с данными, нажать в главном меню «Вставка» – «Диаграмма». Появится меню редактора диаграмм. В нем можно выбирать тип визуализации, менять стиль диаграммы, задавать заголовок, подзаголовок, названия осей диаграммы, работать с легендой и сеткой. Когда диаграмма готова, ее можно выгрузить в одном из форматов:

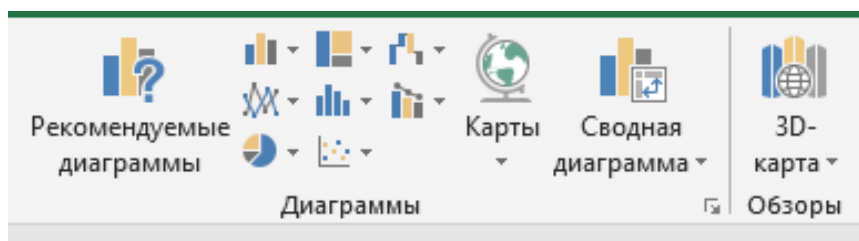


Рисунок № 14. Меню диаграмм

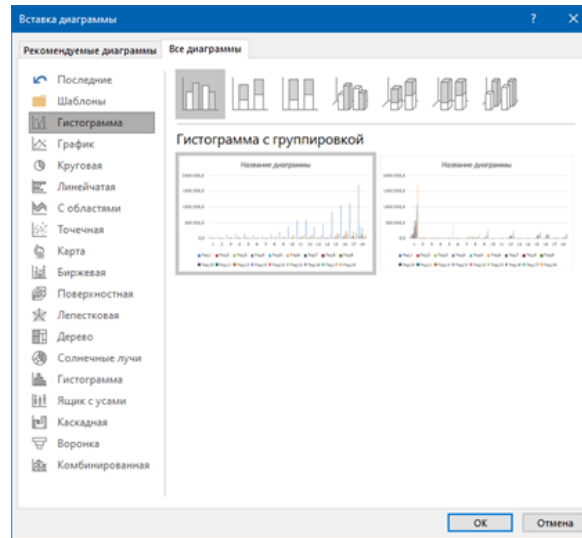


Рисунок № 15. Окно вставки диаграмм



Рисунок № 16. Итоговая диаграмма

## 5.9. Итоговая визуализация статистических данных в Microsoft Excel

В данной главе будет рассмотрена итоговая визуализация в пакете Microsoft Excel в качестве примера, однако аналогичные результаты были также получены в среде Jupyter Notebook.

Все данные были взяты из открытых источников, впоследствии были очищены и немного доработаны для удобства их дальнейшего использования. Все данные в виде .xls таблиц также прилагаются к данному учебному пособию.

Основой данного анализа являются открытые данные о состоянии и развитии науки в Республике Казахстан в период с 2003 по 2020 годы, Объем инновационной продукции по областям и по Республике в целом, а также количество предприятий, осуществляющие научную деятельность. В дополнении к этому для анализа и трактовки полученных результатов были использованы статистические данные количества образованного населения, так или иначе участвующих в научно-исследовательской деятельности, а именно магистрантов, докторантов, докторов и кандидатов наук.



Рисунок № 17. Диаграмма количества специалистов-исследователей по профилю

Анализируя полученные итоговые данные, в целом можно наблюдать положительную статистику научной деятельности в Республике Казахстан. С увеличением объема инвестиций, что является ярким показателем заинтересованности Государства в ее дальнейшем развитии и улучшении, как следствие, положительную динамику роста можно наблюдать в сфере образования. В данном конкретном случае нас интересует количество работников, непосредственно занимающихся научно-исследовательской деятельностью.



Рисунок № 18. Количество организаций, осуществляющих НИОКР, единиц

Как видно из статистических данных, за прошедшее десятилетие стремительное развитие получили практически все области Республики Казахстан, увеличилось количество обучающихся магистрантов и докторантов, как следствие, увеличилось число научных сотрудников и исследователей. Все эти показатели являются одними из ключевых в дальнейшем развитии Казахстанской экономики и инновационной деятельности.



Рисунок № 19. Численность работников, выполняющих НИОКР, человек

В стране можно наблюдать достаточно стабильную положительную динамику роста после девальвации тенге 2015 года. На конец 2020 года в количество предприятий, осуществляющих научно-исследовательскую деятельность, увеличилось в среднем на 3.8% по сравнению с прошлым годом, показывая тенденцию к дальнейшему росту.



Рисунок № 20. Рост количества предприятий РК, осуществляющих научную деятельность

## 5.10. Объем инновационной продукции

Анализируя основные показатели инновационной деятельности Республики Казахстан, основным и наиболее значимым параметром является рост объема инновационной продукции.

Инновационная продукция – это продукция, включающая в себя такие товары, работы и услуги, которые в течение последних трех лет подвергались различным технологическим изменениям. Состав инновационной продукции определяется в соответствии с типами технологических инноваций, и охватывает как изделия новые, то есть вновь внедренные, подвергшиеся значительным технологическим изменениям, так и уже усовершенствованных. Исходя из этого, определение новой продукции соответствует понятию радикальных продуктовых инноваций, значительно технологически измененной, то есть инкрементальной инновации, а усовершенствованной продукции – процессной инновации.



Рисунок № 21. Объем инновационной продукции по годам

В настоящее время в Республике Казахстан инновационная продукция представлена как в виде усовершенствованной продукции, так и в виде новой, с небольшим перевесом в сторону первой.



Рисунок № 22. Сравнительная диаграмма роста объема инновационной продукции РК в 2003 и 2020 годах

Связано это прежде всего с изначально обширной материальной и научно-технической базой в приоритетных для Казахстана отраслях производства и промышленности, доставшейся молодой Республике после развала Советского Союза.



Рисунок № 23. Пример итоговой визуализации статистических данных при помощи Python и Adobe Illustrator с применением гео-карты

Этими отраслями, как видно по анализу объема инновационной продукции, проведенного с помощью языка программирования Python в среде Jupyter Notebook, который мы рассмотрим в следующей главе, являются: нефтегазовый сектор (Атырауская область), угольная, металлургическая промышленность (Карагандинская и Восточно-Казахстанская области), сельскохозяйственная промышленность (Костанайская, Туркестанская и Восточно-Казахстанская области). Во всех приведенных областях наблюдается активное внедрение инновационной продукции, применительно к соответствующей отрасли экономики.

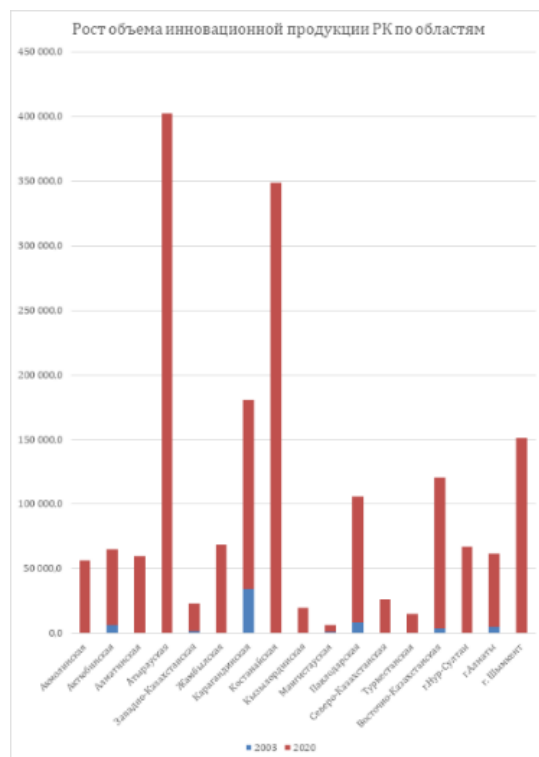


Рисунок №24. Сравнительная диаграмма роста объема инновационной продукции РК по областям с 2003 по 2020 годы

Также значимым показателем в данной статистике является научная и образовательная деятельность. На данной карте ее обозначения не совсем очевидны, однако по статистической таблице и приведенных ниже данных с разделением по областям показывают, что в процессе роста объема инновационной продукции решающее значение имеет рост образовательной и научной отрасли. Так, по сравнению с 2003 годом, в связи с развитием образования, цифровизации, роста в стране доли образованного населения, к 2020 году мы можем наблюдать результат всех этих действий: значительно возросла роль каждой отдельно взятой области Республики, увеличено количество учебных заведений, научных предприятий в городах областного и республиканского значения.



Рисунок № 25. Сравнительная диаграмма роста объема инновационной продукции РК по областям в 2003 году



Рисунок № 26. Сравнительная диаграмма роста объема инновационной продукции РК по областям в 2020 году

### Лабораторная работа

#### Знакомство с Microsoft Excel

Вопросы:

1. Какие виды операций над данными присутствуют в Microsoft Excel?
2. Что такое таблицы Excel?

Задание:

1. Знакомство с Microsoft Excel. Загрузите выбранные вами ранее таблицы, полученные с ресурса stat.gov.kz. Проведите простые операции над данными, примените формулы.
2. Сделайте краткий анализ данных ваших таблиц в MS Excel.
3. Примените функции “Сумма”, “Среднее Значение”, “Процентное соотношение” к вашим данным.
4. Примените инструменты форматирования и сортировки к вашим данным по критериям на ваше усмотрение.

### **Самостоятельная работа студента**

#### **Визуализация данных в Excel**

Вопросы:

1. Какие виды визуализации поддерживает MS Excel?
2. Что такое условное форматирование?
3. Что такое фильтрация данных?
4. Что такое диаграмма MS Excel?

Задание:

Визуализируйте данные вашей таблицы, применив следующие методы визуализации:

- а) Условное форматирование
- б) Спарклайны
- в) Гистограммы и графики
- г) Круговую диаграмму

Дайте трактовку визуализации, проведите итоговый анализ ваших данных

#### **Вопросы для самоконтроля:**

1. Какие виды операций над данными присутствуют в MS Excel?
2. Что такое сводные таблицы?
3. Дайте определение таблицам в Microsoft Excel.
4. Какие операции с данными позволяет осуществлять Microsoft Excel?
5. Какие виды визуализации присутствуют в Microsoft Excel?
6. Что такое Спарклайн?
7. Какие диаграммы доступны в MS Excel?
8. Что такое условное форматирование?
9. Что такое инновационная продукция?
10. Как осуществляется фильтрация данных в Excel?

## Модуль 6. ОБРАБОТКА И АНАЛИЗ ДАННЫХ ПРОГРАММНЫМИ МЕТОДАМИ

### 6.1 Обзор программных средств для обработки и анализа данных

На текущий момент инструментарий для журналистики данных представляет собой обширный набор различных программ, онлайн платформ и библиотек для сбора данных, их анализа и визуализации. Также к инструментам журналистики данных можно отнести OSINT технологии (open-source intelligence), позволяющие собирать, анализировать информацию из Интернет-пространства. Данные о человеке, компании могут быть получены из бесплатных публичных источников. OSINT охватывает данные, которые можно искать на разных видах носителей. Кроме текста, это фото-, видеоконтент, материалы из вебинаров, публичных собраний, конференций. Ключевое преимущество в том, что скорость, сбор и анализ информации происходит быстро, и это не зависит от типа запроса.

Определив понятие OSINT, рассмотрим основные платформы, использующие данную технологию. На текущий момент существует множество приложений, среди которых можно выделить Maltego, DarkOwl Vision, CobWebs, SpiderFoot, Google Dorks и другие. Все они имеют общую основу и, как правило, используются в сфере кибербезопасности, журналистике данных, разведке данных, однако методы и цели их применения могут быть различными. Так, начинающим дата-журналистам целесообразно будет использовать Google Dorks. Метод использует простой синтаксис Google для фильтрации данных поисковой системы, в котором сайты автоматически индексируются во время сканирования ботами Google, и, соответственно, пользователи могут детально проверить сервер и найти нужную информацию. Для детализированного и обширного поиска подходит SpiderFoot, решениями для анализа угроз и платформами для веб-исследований являются DarkOwl Vision и CobWebs.

Однако, помимо перечисленных преимуществ, OSINT платформы также обладают рядом недостатков, по причине которых методика работы с ними не была описана как основная в данном пособии. На примере платформы Maltego, являющейся одной из наиболее популярных платформ для разведки данных, выделим основные минусы работы с платформами данного типа. К ним, в первую очередь, относится тот факт, что большинство OSINT платформ используют стандартный алгоритм поиска, анализируя данные из сети Интернет, то есть обладают одним и тем же набором недостатков, а именно: невозможностью трактовки данных, в некоторых случаях неправильной фильтрацией и интерпретацией данных, отсутствием структурированной и «подогнанной» под алгоритмы поиска информации. Большинство OSINT платформ являются проприетарными, и, как следствие, платными платформами. Еще одним недостатком проприетарных OSINT платформ является отсутствие контроля со стороны стороннего специалиста, невозможность подбирать и настраивать инструментарий под свои задачи, модифицируя работу алгоритма поиска данных. OSINT платформы являются одними из наиболее востребованных и актуальных платформ в сфере кибербезопасности, разведке данных и криминалистике. Также они применимы в сфере журналистики данных, однако данное направление не является основным для этих платформ.

В разделе 6.2 рассмотрим OSINT платформу на примере Maltego, одной из наиболее популярных проприетарных программ для разведки данных. Далее более подробно изучим применение языка программирования Python для обработки и анализа данных.

## 6.2 Программное обеспечение Maltego как пример проприетарной платформы

Maltego — это программное обеспечение для поиска, построения и анализа связей между различными объектами и субъектами информации, итогом работы которого является формирование графа на основе анализа полученных связей. Данное ПО применяется в онлайн-расследованиях для автоматизации процесса и поиска связей между блоками информации, которые размещены в различных источниках сети интернет. Основными особенностями Maltego являются возможность визуализации данных, взятых из публичных или открытых источников, возможность глубокого анализа, а также автоматическое построение взаимосвязей между обнаруженными данными.

Maltego – это комплексный инструмент графического анализа ссылок, который предлагает интеллектуальный анализ данных и сбор информации в реальном времени, а также представление этой информации в виде графа на основе узлов, что позволяет легко идентифицировать шаблоны и множественные связи между данными. С помощью Maltego можно легко извлекать данные из разрозненных источников, автоматически объединять совпадающую информацию в диаграмму и визуально отображать ее для изучения вашего ландшафта данных. Maltego предлагает возможность легко подключать данные и функции из различных источников с помощью преобразований. Через Transform Hub можно подключать данные от более чем 30 партнеров по данным, различных общедоступных источников (OSINT), а также свои собственные данные. Maltego не является бесплатным программным обеспечением, но поставляется в нескольких редакциях, включая бесплатную имеющуюся в открытом доступе версию под названием Maltego Community Edition, который широко используется многими профессионалами в области информационной безопасности. Maltego может использоваться на стадии предварительного сбора информации и проведения работ, связанных с безопасностью и защитой данных. Визуальное отображение информации, мощная поисковая система и наличие собственной базы данных и алгоритмов поиска на основе искусственного интеллекта являются основными возможностями данного программного обеспечения. Ниже представлена рабочая среда Maltego (рис. 41).

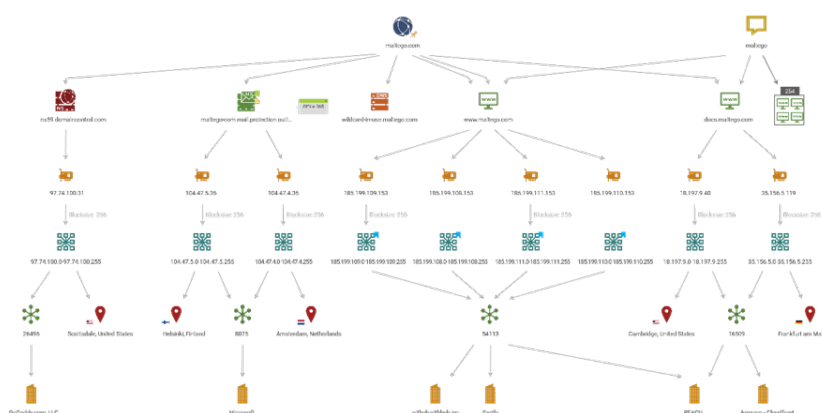


Рисунок 41. Рабочая среда Maltego

На рисунке 42 показан процесс установки и выполнения программного обеспечения Maltego в Windows. Для использования Community Edition Maltego CE необходима регистрация на веб-платформе Maltego (<https://www.maltego.com/blog/beginners-guide-to-maltego-setting-up-maltego-community-edition-ce/>).

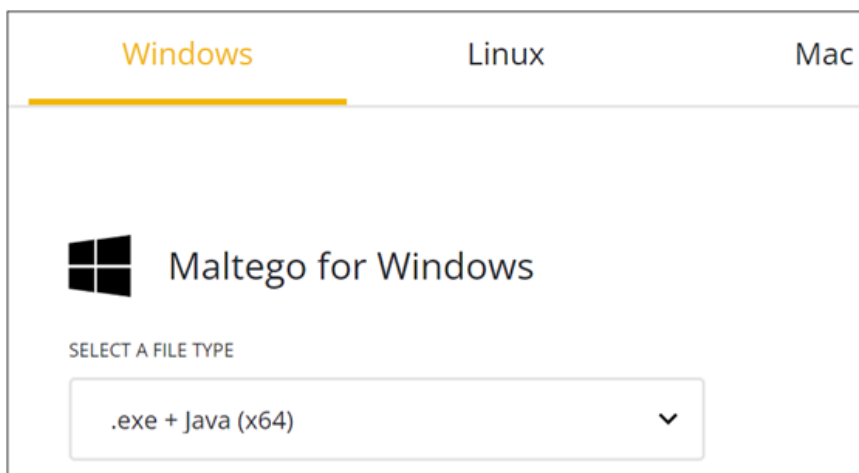


Рисунок 42. Официальный загрузочный сайт Maltego для Windows

После завершения установки программа Maltego Desktop Client будет доступна на панели задач и на рабочем столе. При первом запуске Maltego Desktop Client откроется домашняя страница, включающая стартовую страницу слева и меню редактирования Transform Hub справа. Transform Hub позволяет устанавливать дополнения Transforms, предоставляемые сторонними поставщиками, а также дополнительные инструменты Transforms, предоставляемые Maltego Technologies и Paterva. Каждый из пакетов в Transform Hub называется элементом меню редактирования (рис. 43, 44).

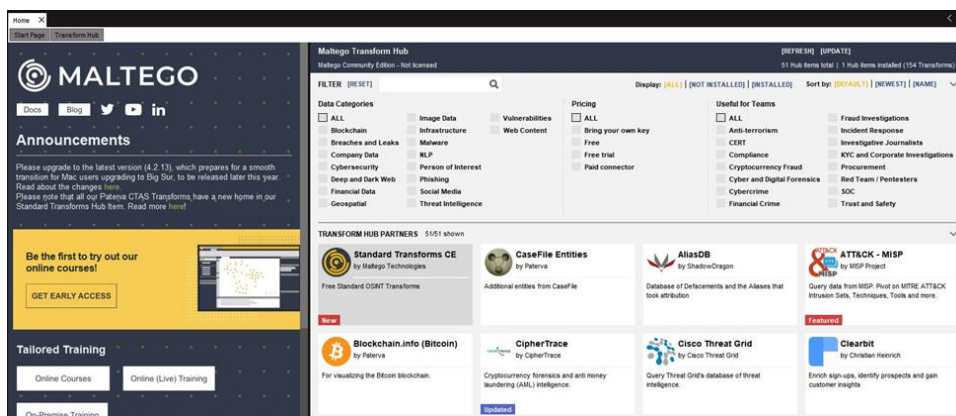


Рисунок 43. Домашняя страница Maltego

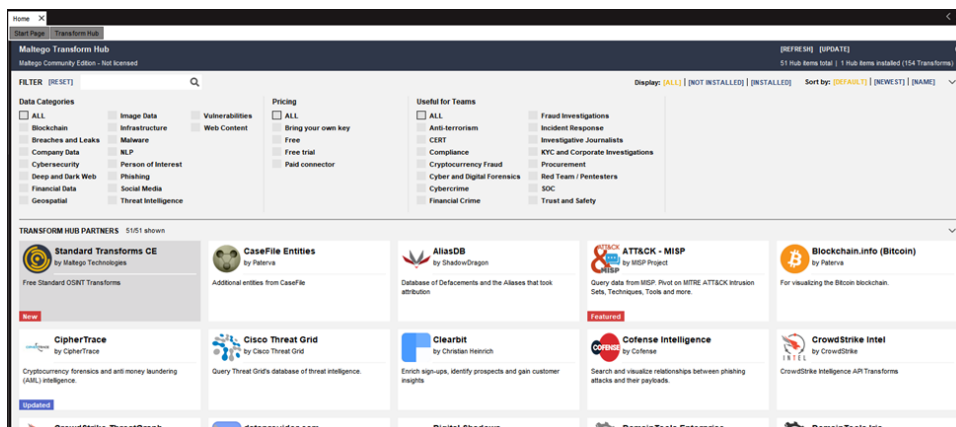


Рисунок 44. Меню редактирования Transform Hub

### 6.3. Python и Jupyter Notebook как инструменты дата - журналиста

Данный модуль посвящен более гибким подходам для журналистики данных с применением языка программирования Python. По аналогии с MS Excel, будут рассмотрены такие аспекты обработки и анализа данных, как фильтрация, сортировка, очистка датасетов, их структуризация и конечная визуализация.

Ссылки на используемые библиотеки и программное обеспечение:

1) Дистрибутив языка программирования и среда разработки Anaconda, <https://www.anaconda.com/products/individual>

2) Дистрибутив Python для Windows WinPython, с предустановленными библиотеками и средой разработки Jupyter Notebook для дата анализа,

<https://winpython.github.io/>

3) Библиотека pandas для загрузки и взаимодействия с данными, <https://pandas.pydata.org/>

4) Библиотека matplotlib для визуализации, построения двумерной и трёхмерной графики, <https://matplotlib.org/>

5) Библиотека seaborn для создания статистических графиков на Python, <https://seaborn.pydata.org/>

6) Официальный сайт языка программирования python, <https://www.python.org/>

**Знакомство с Python и Jupyter Notebook.** Прежде чем мы начнём изучать Python, рассмотрим небольшой пример с загрузкой данных в Jupyter Notebook и их просмотром с помощью библиотеки pandas. В качестве источника информации будет использован сайт википедии, а именно раздел «Высшие учебные заведения Республики Казахстан». Зайдем на страницу википедии и найдем там таблицу с информацией об учебных заведениях:

```
import pandas as pd
import seaborn as sns
%config InlineBackend.figure_format = 'svg'
%matplotlib inline
```

Рисунок № 27. Данные об учебных заведениях РК

Для начала импортируем несколько библиотек. Скорее всего у вас уже установлена `pandas` (это библиотека для работы с датафреймами), но может не быть библиотеки `seaborn`. Её надо установить, например, набрав в командной строке `pip3 install seaborn` или `conda install seaborn`.

Мы хотим программно вытащить эту информацию со страницы и немножко её обработать.

```
In [13]: tables = pd.read_html("https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%88%D1%81%D0%B8%D0%BA_%D0%B2%D1%8B")  
         table = tables[0]  
  
         table  
         < ...
```

Рисунок № 28. Обработка информации об учебных заведениях РК

Итоговый результат:

B [13]:

```

tables = pd.read_html("https://ru.wikipedia.org/wiki/%D0%A1%D0%BD%D0%B8%D1%81%D0%B8%D0%BA_%D0%A0%D0%B8%D1%81")
table = tables[0]
table

```

Out[13]:

|   | Unnamed: 0                                        | Unnamed: 1 | Unnamed: 2               | Unnamed: 3                                        | Unnamed: 4                                        |
|---|---------------------------------------------------|------------|--------------------------|---------------------------------------------------|---------------------------------------------------|
| 0 | Евразийский национальный университет имени Л. ... | Нур-Султан | 1996                     | Ректор: доктор исторических наук Ерлан Баттаев... | NaN                                               |
| 1 | Казахская национальная академия искусств имени... | Алма-Ата   | 1976                     | Ректор: кандидат исторических наук, доцент Ады... | Преемник Алма-Атинского театрально-художествен... |
| 2 | Казахская национальная консерватория имени Кур... | Алма-Ата   | 1944                     | Ректор: народная артистка Казахстана, профesso... | NaN                                               |
| 3 | Казахская национальная академия хореографии       | Нур-Султан | 2015                     | Ректор: народная артистка России, лауреат Госу... | Первое высшее учебное заведение в Центральной ... |
| 4 | Казахский национальный аграрный университет       | Алма-Ата   | 1929 (АЭВН/1930)(KazCGU) | Ректор: доктор экономических наук, профессор Т... | Образован слиянием бывшего Казахского сельског... |
| 5 | Казахский национальный медицинский университет... | Алма-Ата   | 1931                     | Ректор: доктор медицинских наук, профессор Айт... | NaN                                               |
| 6 | Казахский национальный педагогический универси... | Алма-Ата   | 1926                     | Ректор: доктор педагогических наук, профессор...  | Преемник Казахского педагогического института ... |
| 7 | Казахский национальный технический университет... | Алма-Ата   | 1934                     | Ректор: доктор экономических наук, профессор Б... | Преемник Казахского политехнического института... |
| 8 | Казахский национальный университет имени аль-Ф... | Алма-Ата   | 1934                     | Ректор: доктор технических наук, профессор Гал... | NaN                                               |
| 9 | Казахский национальный университет искусств       | Нур-Султан | 1996                     | Ректор: народная артистка Казахстана, профesso... | NaN                                               |

B [ ]:

Рисунок № 29. Импорт таблицы в Jupyter Notebook

План работы. В данной части мы разберемся с фильтрацией, сортировкой и очисткой датафреймов перед непосредственной работе с ними в среде Jupyter Notebook. Займемся декодированием текста с помощью библиотеки `ftfy` и определением языка через библиотеку `google_trans_new`. Блокнот Jupyter и наша шпаргалка по `pandas` прилагаются.

Общий процесс работы в первой части в свою очередь можно разделить на 8 этапов:

- ## 1. Установка pandas в Jupyter Notebook

2. Объекты Series и DataFrame
3. Импорт файлов и просмотр содержимого
4. Операции над строками и столбцами
5. Сортировка
6. Очистка данных.
7. Фильтрация и поиск
8. Сохранение датафрейма

**Установка Pandas.** Установка библиотеки Pandas в Jupyter Notebook является достаточно простым действием. Мы используем Python 3 и оболочку Jupyter (ранее известный как IPython Notebook). Чтобы их установить, проще всего скачать пакет Anaconda. Обратите внимание: вам нужна версия с Python 3.x. В качестве альтернативы можно использовать пакет WinPython, с предустановленными заранее библиотеками и пакетами. После установки Anaconda у вас появится: в меню «Пуск» пункт «Anaconda», в нём подпункт «IPython Notebook».

После запуска IPython Notebook у вас откроется окно браузера (и лучше пусть это будет не Internet Explorer, под ним IPython Notebook работает с некоторым скриптом, а иногда не работает) со списком файлов. В нём надо выбрать New → Python 3. Откроется новая вкладка браузера, в ней будет запущен пустой notebook, состоящий из отдельных ячеек (cells). В ячейку с кодом можно вписать код и нажать Shift+Enter — он выполнится и вам покажут тут же результат.

Далее, непосредственно в среде разработки импортируем саму библиотеку:

```
In [ ]: import pandas as pd
```

Рисунок № 30. Импорт в библиотеку

**Объекты Series и Data Frame.** Объект Series – одномерный массив, состоящий из элементов и их индексов:

```
In [15]: import pandas as pd

series_example = pd.Series(['серия', 100, 'объект', 'ряд'])
series_example

Out[15]: 0    серия
         1     100
         2  объект
         3     ряд
         dtype: object
```

Рисунок № 31. Объект Series

Здесь:

0, 1, 2, 3 – индексы. a, 48, d, hello115 – элементы.

В свою очередь объект DataFrame – многомерный массив, как таблица в Excel.

```

In [16]: df = pd.DataFrame({
        'telephone number': [123456789, 123679, 15645445, 564645546],
        'name': ['Алимжан', 'Нурлан', 'Айгерим', 'Томирис'],
    })
df

```

Out[16]:

|   | telephone number | name    |
|---|------------------|---------|
| 0 | 123456789        | Алимжан |
| 1 | 123679           | Нурлан  |
| 2 | 15645445         | Айгерим |
| 3 | 564645546        | Томирис |

Рисунок № 32. Объект DataFrame

Здесь:  
 telephone number, name – названия столбцов с соответствующими значениями 123456789 ..., Алимжан ....

## 6.5. Импорт файлов и просмотр содержимого

Импортируем дата-сет основных показателей развития науки РК в Jupyter:

```

In [35]: df = pd.read_excel(r'D:/excel/dataset.xls')
print(df)

```

Рисунок 33. Импорт дата-сет в Jupyter

Аналогично есть команды для импорта других файлов:  
 read\_csv() – импорт csv-файла.  
 read\_sql() – импорт базы данных SQL.  
 read\_json() – импорт json-файла.  
 read\_html() – создает из HTML-таблицы список объектов DataFrame.

```

Unnamed: 0      Основные показатели состояния и развития науки Unnamed: 2  \
0      NaN      NaN      NaN      NaN
1      NaN      NaN      NaN      2003
2      NaN      Внутренние затраты на НИОКР, млн. тенге      11643.5
3      NaN      Доля внутренних затрат на НИОКР от валового вн...      0.25
4      NaN      Количество организаций (предприятий), осуществ...      273
5      NaN      Численность работников, выполняющих НИОКР, чел...      16578
6      NaN      из них:      NaN
7      NaN      специалисты-исследователи      9899
8      NaN      из них:      NaN
9      NaN      доктора наук      979
10     NaN      доктора по профилю      -
11     NaN      доктора философии PhD      NaN
12     NaN      кандидаты наук      2782

```

Рисунок № 34. Импорт Excel-таблицы «Основные показатели состояния и развития науки РК»

Отображаются все 12 строк, так как количество данных в нашем файле больше ориентируется по столбцам.

Посмотрим первые пять строчек датафрейма с помощью `df.head(5)`:

```

В [38]: df.head(5)

```

Out[38]:

|   | Unnamed: 0 | Основные показатели состояния и развития науки    | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 | Unnamed: 8 | Unnamed: 9 | Unnamed: 10 |
|---|------------|---------------------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| 0 | NaN        | NaN                                               | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN         |
| 1 | NaN        | NaN                                               | 2003       | 2004       | 2005       | 2006       | 2007       | 2008       | 2009       | 2010       | NaN         |
| 2 | NaN        | Внутренние затраты на НИОКР, млн. тенге           | 11643.5    | 14579.8    | 21527.4    | 24799.9    | 26835.5    | 34761.6    | 38988.7    | 33466.8    | 432         |
| 3 | NaN        | Доля внутренних затрат на НИОКР от валового вн... | 0.25       | 0.25       | 0.28       | 0.24       | 0.21       | 0.22       | 0.23       | 0.15       | NaN         |
| 4 | NaN        | Количество организаций (предприятий), осуществ... | 273        | 295        | 390        | 437        | 438        | 421        | 414        | 424        | NaN         |

Рисунок № 35. Отображение первых пяти строчек датафрейма с помощью `df.head(5)`

Последние 5 строчек:

```

В [39]: df.tail(5)

```

Out[39]:

|    | Unnamed: 0 | Основные показатели состояния и развития науки | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 | Unnamed: 8 | Unnamed: 9 | Unnamed: 10 |
|----|------------|------------------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| 8  | NaN        | из них:                                        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN         |
| 9  | NaN        | доктора наук                                   | 979        | 1013       | 1106       | 1157       | 1166       | 1191       | 1338       | 1341       | NaN         |
| 10 | NaN        | доктора по профилю                             | -          | -          | -          | -          | -          | -          | -          | -          | 14          |
| 11 | NaN        | доктора философии PhD                          | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | 68         | 59         | NaN         |
| 12 | NaN        | кандидаты наук                                 | 2782       | 2740       | 3018       | 3147       | 3058       | 2861       | 2734       | 3012       | 3012        |

Рисунок 36. Отображение последних пяти строчек датафрейма с помощью `df.tail(5)`

И пять случайных строчек:

`In [40]: df.sample(5)`

`Out[40]:`

|    | Unnamed: 0 | Основные показатели состояния и развития науки    | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 | Unnamed: 8 | Unnamed: 9 | Unnamed: 10 |
|----|------------|---------------------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|
| 11 | NaN        | доктора философии PhD                             | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | 68         | 59         |             |
| 4  | NaN        | Количество организаций (предприятий), осуществ... | 273        | 295        | 390        | 437        | 438        | 421        | 414        | 424        |             |
| 12 | NaN        | кандидаты наук                                    | 2782       | 2740       | 3018       | 3147       | 3058       | 2861       | 2734       | 3012       |             |
| 6  | NaN        | из них:                                           | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        |             |
| 5  | NaN        | Численность работников, выполняющих НИОКР, чел... | 16578      | 16715      | 18912      | 19563      | 17774      | 16304      | 15793      | 17021      |             |

Рисунок 37. Отображение случайных пяти строчек датафрейма с помощью `df.sample(5)`

Узнаем количество строчек и столбцов:

`In [41]: df.shape`

`Out[41]: (13, 20)`

Рисунок № 38. Количество строчек и столбцов

Здесь:

13 – количество строчек.

20 – количество столбцов.

Определив основные команды загрузки и начального взаимодействия с таблицами с Jupyter Notebook, следует рассмотреть такое понятие, как очистка данных и приведение к удобному для анализа виду.

Источники данных в различных системах часто находятся в разрозненном виде и в различных состояниях, вследствие чего возникает необходимость в их подготовке к анализу. Преобразования выполняются автоматически (в соответствии с набором правил) либо вручную (в интерактивном режиме).

Как видно из скриншотов, нам необходимо удалить столбы с показателями NaN (Not a Number), так как для статистического анализа эти столбцы не представляют значимости, а также переименовать столбцы с неверно отображенными обозначениями (Unnamed).

Удаление столбцов можно реализовать несколькими командами, рассмотрим каждую из них:

`In [7]: df = df.drop('Unnamed: 0', 1)`

```
B [8]: df = df.drop(df.columns[[1]], axis=1, inplace=True)
```

Рисунок № 39. Удаление столбцов

В первом случае мы задаем команду на удаление столбца посредством определения его имени. Цифра 1 в конце команды обозначает осьY.

Вторая команда отличается тем, что названия столбцов записаны в python в виде списка. В результате, для удаления соответствующего столбца мы указываем только на его порядок. Цифра 1 в конце команды также обозначает осьY.

Итоговый результат работы команд идентичен:

Out[40]:

|    | Основные показатели состояния и развития науки    | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 | Unnamed: 8 | Unnamed: 9 |
|----|---------------------------------------------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 11 | доктора философии PhD                             | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | 68         | 59         |
| 4  | Количество организаций (предприятий), осуществ... | 273        | 295        | 390        | 437        | 438        | 421        | 414        | 424        |
| 12 | кандидаты наук                                    | 2782       | 2740       | 3018       | 3147       | 3058       | 2861       | 2734       | 3012       |
| 6  | из них:                                           | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        | NaN        |
| 5  | Численность работников, выполняющих НИОКР, чел... | 16578      | 16715      | 18912      | 19563      | 17774      | 16304      | 15793      | 17021      |

Рисунок № 40. Удаление лишних столбцов в Jupyter

В данном случае мы избавились от лишнего второго столбца. Рассмотрим теперь процесс переименования. Для переименования нескольких столбцов наиболее быстрым и удобным вариантом является их перечисление в виде словаря, с использованием команды `df.rename`, в нашем случае – столбцов:

```
B [12]: df.rename(columns={'Unnamed: 2': '2003', 'Unnamed: 3': '2004'}, inplace=True)
print(df)
```

Рисунок 41. Процесс переименования столбцов

Результат будет выглядеть так:

| Основные показатели состояния и развития науки |                                                   | 2003  | 2004  | Unnamed: 4 | Unnamed: 5 |
|------------------------------------------------|---------------------------------------------------|-------|-------|------------|------------|
| 11                                             | доктора философии PhD                             | NaN   | NaN   | NaN        | NaN        |
| 4                                              | Количество организаций (предприятий), осуществ... | 273   | 295   | 390        | 437        |
| 12                                             | кандидаты наук                                    | 2782  | 2740  | 3018       | 3147       |
| 6                                              | из них:                                           | NaN   | NaN   | NaN        | NaN        |
| 5                                              | Численность работников, выполняющих НИОКР, чел... | 16578 | 16715 | 18912      | 19563      |

Рисунок № 42. Результат переименования столбцов

И проделывая ту же операцию на других столбцах, получаем итоговый вариант:

| Основные показатели состояния и развития науки |                                                   | 2003  | 2004  | 2005  | 2006  |
|------------------------------------------------|---------------------------------------------------|-------|-------|-------|-------|
| 11                                             | доктора философии PhD                             | NaN   | NaN   | NaN   | NaN   |
| 4                                              | Количество организаций (предприятий), осуществ... | 273   | 295   | 390   | 437   |
| 12                                             | кандидаты наук                                    | 2782  | 2740  | 3018  | 3147  |
| 6                                              | из них:                                           | NaN   | NaN   | NaN   | NaN   |
| 5                                              | Численность работников, выполняющих НИОКР, чел... | 16578 | 16715 | 18912 | 19563 |

Рисунок № 43. Итоговая чистая таблица

## 6.4. Основные команды для работы в Jupyter Notebook

Ниже приведен список основных команд для работы с данными в среде Jupyter Notebook. Данный набор команд является основным и наиболее всеобъемлющим при работе с анализом данных в Jupyter. Так как основной библиотекой python для дата анализа является pandas, то перед использованием нижеследующего набора команд рекомендуется ознакомиться с документацией данной библиотеки:

Pandas 1.3.3. Documentation, <https://pandas.pydata.org/docs/>

```

# Импорт данных
df = pd.read_csv('file') - импорт файла
df = pd.read_csv('file', sep=",") - импорт файла с разделителем «запятая»

# Экспорт данных
save_file.to_csv('file.csv', encoding='utf8') - сохранение датафрейма в csv-файл с кодировкой utf8

# Вывод на экран
df - вывести весь датафрейм
df.info - информация о датафрейме
df.head(5) - первые пять строчек
df.tail(5) - последние пять строчек
df.sample(5) - случайные пять строчек
df.shape - количество строк и столбцов
df.dtypes - типы данных в столбцах
df.columns - названия столбцов
df['col1'] - значения столбца col1
df['col1'][0:20] - срез столбца col1
df[['col1', 'col2']] - значения нескольких столбцов
df.loc['index_name'] - строки и/или столбцы по нечисловому значению индекса
df.iloc[0] - строки и/или столбцы по числовому индексу
df.iloc[2,4] - значение ячейки в третьей строчке и пятом столбце

# Объединение датафреймов
df1 = pd.concat([df1, df2], axis=0) - к df1 добавляем строчки df2
df1 = pd.concat([df1, df2], axis=1) - к df1 добавляем столбцы df2
df1 = df1.append(df2) - к df1 добавляем строчки df2

# Копирование датафрейма
df2 = df1.copy - создает глубокую копию датафрейма

```

Рисунок № 44. Основные команды Jupyter Notebook для дата-анализа(1)

```

# Добавление строчек и столбцов
new_row = {'col1': 'Hello', 'col2': 123}
df = df.append(new_row, ignore_index=True) - добавляет в конец df строчку new_row
df.loc[2] = ['Hello', 123] - добавляет строчку на третье место с начала фрейма
df['new_column'] = 'abcd' - создает столбец со значениями abcd

# Удаление строчек и столбцов
df.drop(26954, 0) - удаляет строчку номер 26954
df.drop('col', 1) - удаляет столбец col

# Переименование столбцов
df.rename(columns={'col1': 'new_col1'}) - меняет имя столбца с col1 на new_col1

# Изменение значения в ячейке
df['col'] = df['col'].replace(to_replace=2020, value=2021) - меняем все значения с 2020 на 2021
df['col'] = df['col'].mask(df['col'] == 2001, 2021) - аналогично, меняем все значения с 2020 на 2021

# Сортировка
df.sort_values('col', ascending=False) - сортировка в обратном алфавитном порядке
df.sort_values(['col1', 'col2']) - сортировка по двум столбцам. col1 - в приоритете, так как первый

# Изменение типов
df['col'] = df['col'].astype(str) - меняет тип переменных на str
df['col'] = df['col'].astype(float) - на float
df['col'] = df['col'].astype('int32') - на int32
df['col'] = df['col'].astype('int64') - на int64
df['col'] = pd.to_numeric(df['col1']) - на int64

# Удаление и замена NaN-значений
df.dropna() - удаляет строчки с отсутствующими значениями
df.dropna(axis=1) - удаляет столбцы с отсутствующими значениями
df.fillna('abcd') - меняет NaN на abcd

```

Рисунок № 45. Основные команды Jupyter Notebook для дата-анализа(2)

```

# Удаление лишних пробелов
df['col'] = df['col'].map(str.strip) - удаляет пробелы слева и справа

# Обработка дат
df['col'] = pd.to_datetime(df['col']) - меняет тип на datetime64[ns]
df['col'].dt.day - вытаскивает из столбца col только значение дня
df['col'].dt.month - только месяц
df['col'].dt.year - только год

# Фильтрация и поиск
df['col'].str.startswith('a') - ищет совпадение по первому символу строки, не поддерживает regex
df['col'].str.endswith('a') - ищет совпадение по последнему символу строки, не поддерживает regex
df['col'].str.match('a') - определяет, начинается ли каждая строка с шаблона
df['col'].str.findall('a') - возвращает совпадения шаблонов
df['col'].str.contains('a') - в результате поиска возвращает булево значение
df['col'].str.extractall('a') - вернет столбец с поисковым шаблоном
df.loc[df['col'].isin(['a', 'b'])] - ищет совпадения в столбцах

# Статистические данные
df.describe() - статистическая сводка по численным значениям

# Подсчет количества повторов
df['col'].value_counts() - показывает сколько раз значения повторяются в столбце
df['col_result'] = df.groupby('col')['col'].transform('count') - создает столбец с количеством повторов значений

```

Рисунок № 46. Основные команды Jupyter Notebook для дата-анализа(3)

## 6.7. Курс по программированию на Python

Такие операции над данными, как сортировка, фильтрация и форматирование также применимы в среде Jupyter, однако в отличие от Microsoft Excel, представляют собой набор вручную вводимых команд и, как следствие, являются более продвинутым и сложным инструментом в освоении. Отсюда следует вывод, что окончательная и итоговая визуализация с применением Jupyter Notebook может быть реализована при наличии у обучающихся достаточных компетенций и знаний в области программирования, а именно языка программирования Python.

Вследствие чего, дабы не нагружать материал данного учебного пособия, в качестве учебного материала для даты журналиста возможно использование курса Высшей школы экономики (Москва) «Программирование на языке Python для сбора и анализа данных». В данном курсе приводится подробное изложение основных моментов при обучении учащихся программированию, и рассматриваются как базовые команды для взаимодействия с языком, такими как арифметические операции (сложение, вычитание, умножение, деление и т.д.), списки, циклы и функции, так и более углубленное изучение основ Jupyter Notebook и библиотек Pandas и matplotlib.

Ссылка на данный курс: <http://math-info.hse.ru/s15/m>

Весь материал подается в понятной и удобной форме, и по его итогу обучающийся будет способен применять навыки и знания для анализа и трактовки данных, с последующей их визуализацией.

## 6.8. Визуализация данных в Jupyter Notebook

Поскольку основной задачей данного руководства является ознакомление с наиболее популярными и актуальными на сегодняшний день инструментами для работы с данными, раздел визуализации данных в Jupyter Notebook будет являться неким кратким обзором функционала и возможностей среды разработки с итоговыми примерами. Данная глава является заключительной стадией визуализации данных. Соответственно, для более углубленного понимания сути работы дата-журналиста, следует внести кое-какие пояснения к этому разделу.

Первый, и наиболее значимый пункт – отличие готовых решений для работы с данными и использование языков программирования и соответствующих сред разработки. В большинстве случаев объем и количество данных на практике не являются большими значениями, соответственно анализ и визуализацию вполне можно выполнить, используя готовые решения. Применение языков программирования оправдано в том случае, когда у нас имеются большие массивы данных, плохо структурированных, или нуждающихся в предварительной обработке. Также использование сред разработки позволяет загружать как полноценные xls или csv файлы, так и загружать таблицы и данные напрямую с веб-сайтов посредством ссылки. Так, анализ данных в реальном времени, с необходимостью регулярно изменять визуальную презентацию в связи с обновлением данных, предпочтительнее использовать языки программирования.

| Объем инновационной продукции (товаров, услуг) |         |         |         |          |          |          |         |          |          |          |          |          |          |          |          |           |           |           |             |  |
|------------------------------------------------|---------|---------|---------|----------|----------|----------|---------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-------------|--|
|                                                | 2003    | 2004    | 2005    | 2006     | 2007     | 2008     | 2009    | 2010     | 2011     | 2012     | 2013     | 2014     | 2015     | 2016     | 2017     | 2018      | 2019      | 2020      | нарастающим |  |
| Республика Казахстан                           | 65020,4 | 74718,2 | 82048,4 | 106030,9 | 152100,0 | 111531,1 | 82587,4 | 142366,4 | 235862,7 | 379005,4 | 578202,1 | 586386,0 | 377186,7 | 445775,7 | 644754,0 | 1094607,4 | 1120565,1 | 1715506,1 |             |  |
| Акимовская                                     | 1,1     | 0,0     | 0,278,1 | 11947,9  | 41230,0  | 4300,0   | 1030,0  | 0000,0   | 9822,5   | 19902,1  | 18305,7  | 30801,0  | 13217,2  | 21016,2  | 15701,0  | 26444,0   | 17791,0   | 16360,1   |             |  |
| Актюбинская                                    | 0401,0  | 8070,5  | 12490,0 | 0400,4   | 11600,0  | 14050,0  | 4430,3  | 9703,3   | 16880,0  | 0540,4   | 0300,0   | 4454,4   | 1838,0   | 10002,2  | 39402,0  | 48299,0   | 51421,7   | 59000,1   |             |  |
| Алматинская                                    | 501,0   | 897,0   | 0571,0  | 10330,2  | 057,4    | 336,0    | 410,0   | 021,0    | 5490,1   | 13200,0  | 13151,0  | 30000,0  | 15000,2  | 12030,0  | 12042,0  | 30351,2   | 30443,0   | 59571,0   |             |  |
| Атырауская                                     | 0,0     | 0,2     | 234,0   | 1940,1   | 180,0    | 3500,4   | 00,0    | 00,0     | 1820,1   | 4770,2   | 38000,2  | 1805,3   | 7500,1   | 7810,4   | 5700,0   | 8810,0    | 7510,1    | 40240,1   |             |  |
| Западно-Казахстанская                          | 1827,7  | 391,0   | 1181,0  | 2327,0   | 2340,0   | 390,0    | 720,7   | -        | 24804,0  | 4300,2   | 0000,0   | 5900,5   | 3310,7   | 3407,1   | 18022,1  | 20300,7   | 24713,4   | 21071,3   |             |  |
| Карагандинская                                 | 848,1   | 370,0   | 1135,0  | 1570,1   | 2340,0   | 0300,1   | 2100,4  | 03,0     | 11251,0  | 19100,2  | 19037,4  | 26250,3  | 23103,7  | 34104,1  | 50004,7  | 10702,3   | 77002,1   | 07430,7   |             |  |
| Костанайская                                   | 14790,0 | 3780,0  | 50210,0 | 50871,0  | 37000,0  | 10471,0  | 10112,4 | 14007,7  | 14380,0  | 30000,1  | 53701,7  | 20570,1  | 18442,0  | 31337,2  | 32000,0  | 54770,0   | 78007,0   | 105730,0  |             |  |
| Кокшетауская                                   | 00,0    | 0,0     | 0,0     | 0,0      | 100,0    | 000,0    | 1000,0  | 1000,0   | 1000,0   | 1000,0   | 1000,0   | 1000,0   | 1000,0   | 1000,0   | 1000,0   | 1000,0    | 1000,0    | 1000,0    |             |  |
| Кызылординская                                 | 1100,0  | 3000,0  | 0570,0  | 0000,0   | 1300,0   | 000,0    | 130,0   | 00,0     | 000,0    | 000,0    | 000,0    | 1000,0   | 1000,0   | 1000,0   | 1000,0   | 1000,0    | 1000,0    | 1000,0    |             |  |
| Ленинградская                                  | 0700,1  | 7001,0  | 10147,0 | 1240,7   | 2000,0   | 10000,2  | 20420,1 | 72000,0  | 72700,0  | 07000,0  | 03300,0  | 00000,0  | 1800,3   | 00000,0  | 17700,1  | 200032,0  | 40003,0   | 00000,0   |             |  |
| Северо-Казахстанская                           | 270,7   | 2330,0  | 1430,0  | 2700,4   | 000,0    | 2300,0   | 2701,0  | 2000,0   | 1400,0   | 0000,0   | 10000,0  | 10000,0  | 11701,0  | 11112,0  | 13000,0  | 00000,0   | 0000,0    | 20000,0   |             |  |
| Семейская                                      | 100,0   | 350,1   | 0277,0  | 3000,0   | 000,0    | 700,0    | 3040,1  | 1000,7   | 10100,2  | 14000,1  | 10000,1  | 00100,0  | 10000,0  | 14777,1  | 13000,0  | 13375,7   | 13707,1   | 14000,0   |             |  |
| Восточно-Казахстанская                         | 4140,2  | 4100,2  | 11411,0 | 11707,2  | 20000,0  | 5100,2   | 0000,0  | 11000,0  | 10000,0  | 00000,0  | 00000,0  | 00000,0  | 00000,0  | 00000,0  | 00000,0  | 00000,0   | 00000,0   | 00000,0   |             |  |
| Нур-Султан                                     | 0,0     | 0,0     | 0,0     | 0,0      | 0,0      | 0,0      | 0,0     | 0,0      | 0,0      | 0,0      | 0,0      | 0,0      | 0,0      | 0,0      | 0,0      | 0,0       | 0,0       | 0,0       |             |  |
| Алматы                                         | 5100,1  | 4000,0  | 7007,0  | 7007,0   | 7000,0   | 11000,0  | 7711,0  | 12000,0  | 10001,0  | 12500,0  | 12500,0  | 20000,0  | 10000,0  | 17100,1  | 20000,0  | 30300,0   | 40000,0   | 50000,0   |             |  |
| г.Шымкент                                      | 400,0   | 5011,3  | 700,0   | 1000,0   | 2100,0   | 700,0    | 701,1   | 3000,0   | 5700,0   | 0500,0   | 13107,0  | 20540,0  | 00000,0  | 10000,0  | 11200,0  | 101070,0  | 10000,0   | 00000,0   |             |  |

Таблицы № 37. Объем инновационной продукции (товаров, услуг),  
<https://stat.gov.kz/api/getFile/?docId=ESTAT098609>

Рассмотрим пример продвинутой визуализации данных в Jupyter. В качестве источника будем использовать данные Таблицы 5.

На *рисунке 60* можно видеть график с точечными значениями. На данном графике можно визуализировать процент вклада той или иной области в общие показатели роста инновационной деятельности в сфере образования. Цветовой градиент является визуализацией зависимости показателей отдельно взятой области и ее процент от общего числа по РК. Так, ось X является осью инновационной продукции в сфере промышленности, тогда как ось Y – образовательной и научной. По итогам данной визуализации наглядно можно сделать вывод о том, что Атырауская, Восточно-Казахстанская и Костанайская области являются лидерами в сфере промышленной и сельскохозяйственной инновационной деятельности, тогда как такие города, как Нур-Султан, Алматы, а также Карагандинская и Восточно-Казахстанская области – лидерами в сфере образовательной и научной деятельности.

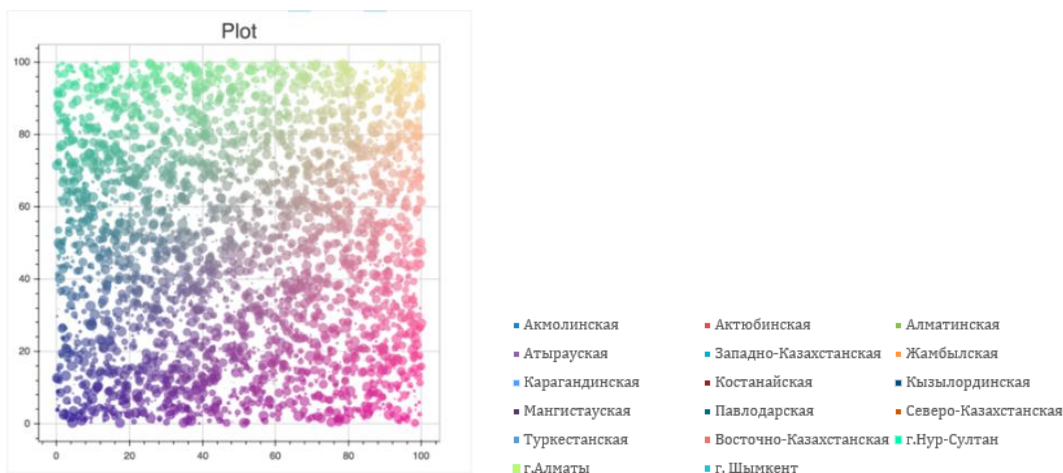


Рисунок № 47. Рост объема инновационной продукции в сфере образования  
за 2020 год в процентном соотношении.

На *рисунке 61* мы проанализировали показатели роста объема инновационной продукции, используя все ту же таблицу 5(?), по Республике Казахстан в промежутке с 2003 по 2020 год. В настоящее время в Республике Казахстан инновационная продукция представлена как в виде

усовершенствованной продукции, так и в виде новой, с небольшим перевесом в сторону первой. Связано это прежде всего с изначально обширной материальной и научно-технической базой в приоритетных для Казахстана отраслях производства и промышленности, доставшейся молодой республике после распада Советского Союза. С 2003 года по настоящее время мы можем видеть стремительный рост инновационной продукции, применительной к соответствующей отрасли экономики.

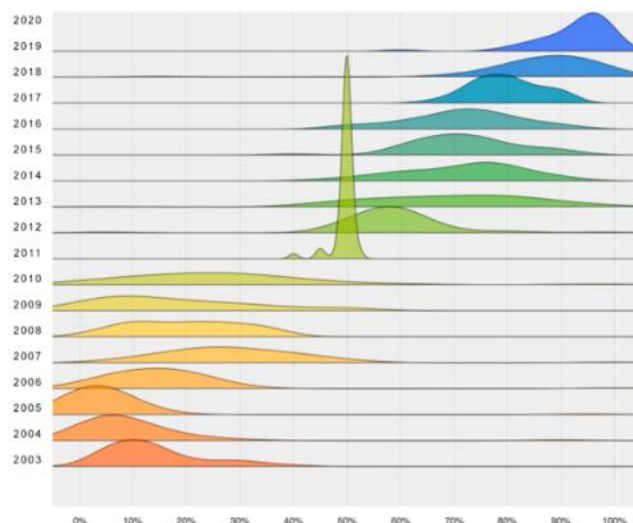


Рисунок № 48. Визуализация показателя роста объема инновационной продукции по всем областям РК с 2003 по 2020 год

Как видно из графика, наиболее резкий и существенный скачок был осуществлен в период 2011-2012 года. Связано это прежде всего со вступлением Казахстана в Таможенный Союз Евразийского Экономического Союза в 2010 году. В связи с увеличением количества товаров, обмена опытом и знаниями, а также внедрением новых технологий в различные отрасли экономики, это позволило Республике Казахстан совершить уверенный скачок в развитии инновационной продукции.

### Лабораторная работа

#### Знакомство с Python и начало работы с Jupyter Notebook

Вопросы:

1. Что такое язык программирования?
2. Что такое среда разработки?
3. Какие существуют типы данных?
4. Для чего необходимо программирование в журналистике данных?

Задание:

5. Дайте определение языку программирования Python и его особенностям для дата-журналистики
6. Установите Python и соответствующую среду разработки на ваш компьютер. Для использования Jupyter Notebook вы можете установить:
7. а) Дистрибутив языка программирования и среду разработки Anaconda.

8. б) более быстрый и простой способ - скачать пакет WinPython для Windows, с заранее установленным дистрибутивом python, средой разработки Jupyter, библиотеками pandas, matplotlib и др.

9. Запустите Jupyter Notebook на своем ПК, запустите простую программу (например, калькулятор)

**Вопросы для самоконтроля:**

1. Что такое Jupyter Notebook?
2. Опишите, в каких случаях целесообразно использовать языки программирования для визуализации, а в каких готовые решения(Microsoft Excel).
3. Для чего нужна библиотека pandas?
4. Опишите способы установки и использования Jupyter Notebook для Windows? Какие программы необходимы для загрузки/установки,
5. Как осуществляется очистка данных в Jupyter Notebook?
6. Какой общий функционал имеют MS Excel и программирование на Python в Jupyter Notebook?
7. Что необходимо для запуска Jupyter Notebook?
8. Какие команды Jupyter Notebook позволяют загружать файлы в среду разработки?
9. Какие форматы загружаемых файлов поддерживаются Jupyter Notebook для анализа данных?
10. Какие библиотеки python являются необходимыми для работы с данными?

## Модуль 7. ВИЗУАЛИЗАЦИЯ В ЖУРНАЛИСТИКЕ

### 7.1. Создание графики с помощью веб-инструментов

Визуализация данных позволяет проиллюстрировать анализ гигантских данных в более компактном виде. Сегодня многие новостные сюжеты основаны на визуализации данных, а повествование с использованием данных стало популярным жанром в журналистике. Благодаря своему эмоциональному воздействию и когнитивной эффективности сторителлинг стал неотъемлемой частью журналистики. Более того, большинство современных читателей в большинстве, являются – визуалами.

Нас окружают сплошные данные, они везде и всюду. Данные производятся различными гаджетами, которые есть у каждого человека. Данные создаются из всего, что мы делаем, говорим, и видим. Иногда достаточно трудно принять решение, проанализировав все данные. Для получения необходимых данных может потребоваться много времени, порой даже годы. Здесь на помощь приходит визуализация данных, которая не только экономит время, но и помогает человеческому мозгу легко понять эти данные.

Визуализация данных – это графическое представление данных с помощью визуальных элементов, таких как диаграммы, графики, карты и т.д. Она позволяет по-новому понять и увидеть тенденции, закономерности и отклонения в данных. Она позволяет быстро принимать решения. Она обеспечивает большую скорость и меньшее количество ошибок при анализе данных.

### Plotting the Brexit conundrum

When two-dimensional scatters aren't sufficient, turn to the ternary chart



James Fransham

Follow



Sep 26, 2019 · 5 min read

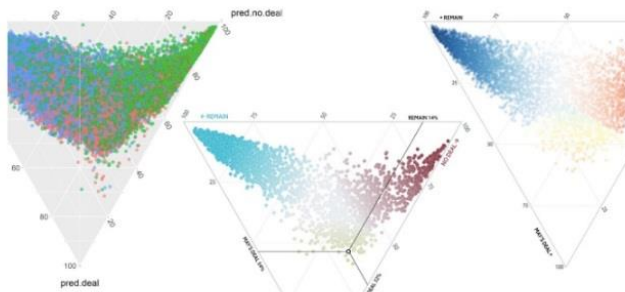


Рисунок № 49. Пример визуализации данных

<https://medium.economist.com/plotting-the-brexit-conundrum-fb8e804474fa>

Визуализация данных – это самый мощный и эффективный способ запоминания решений. Известные типы визуализации данных – это: гистограмма, круговая диаграмма, линейный график, вставка, диаграмма рассеяния, радарная или паутинная диаграмма, точечная карта, карта плотности, диаграммы воронки и многие другие.

Часто имеет смысл выбрать инструмент, который охватывает как работу с данными, так и их визуализацию. Поскольку во время разделение этих задач в разных инструментах журналисту придется очень часто импортировать и экспортировать данные. Поэтому ниже предлагаем список некоторых инструментов для визуализации и обработки данных одновременно:

### **Электронные таблицы Google Docs, Excel. Табличные диаграммы Google**

Хотите что-то простое – например, столбиковую или линейную диаграмму, или круговую диаграмму? Вы обнаружите, что в электронных таблицах Google (которые вы создаете в разделе «Документы» своего аккаунта Google) можно создавать довольно красивые диаграммы – в том числе анимированные пузырьки, используемые в книге Ганса Рослинга «Gapminder».

На конференции TED 2006 года Ганс Рослинг представил проект «Невероятная статистика». Была проделана огромная работа с драматическим развитием сюжета. Активная ссылка на выступление Рослинга можно посмотреть здесь - <https://youtu.be/eXtigm0WHjw>

Excel может превращать абстрактные ряды и столбцы чисел в привлекательные, информативные графики и диаграммы. Более того, Excel поддерживает более 30 типов различных двух и трехмерных диаграмм. В отличие от API диаграмм, вам не нужно беспокоиться о коде – это очень похоже на создание диаграммы в Excel: вы выделяете данные и нажимаете на виджет диаграммы. Возможности настройки также стоит изучить – вы можете менять цвета, заголовки и масштабы. Они довольно нейтральны по дизайну, что полезно для небольших диаграмм. Линейные графики также имеют несколько хороших опций, включая опции аннотаций.

Рекомендация: изучите настройку диаграмм – вы можете создать собственную цветовую палитру.

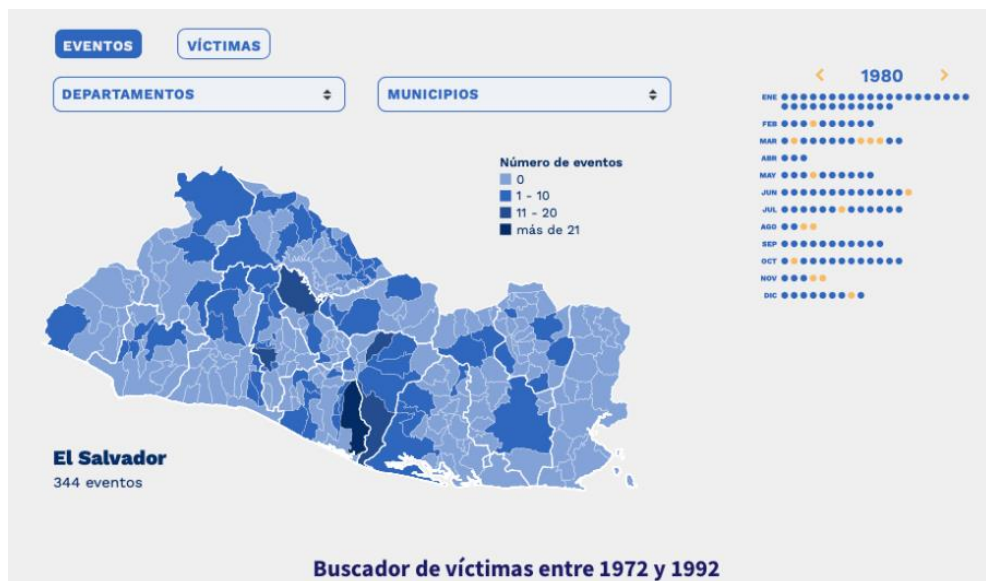


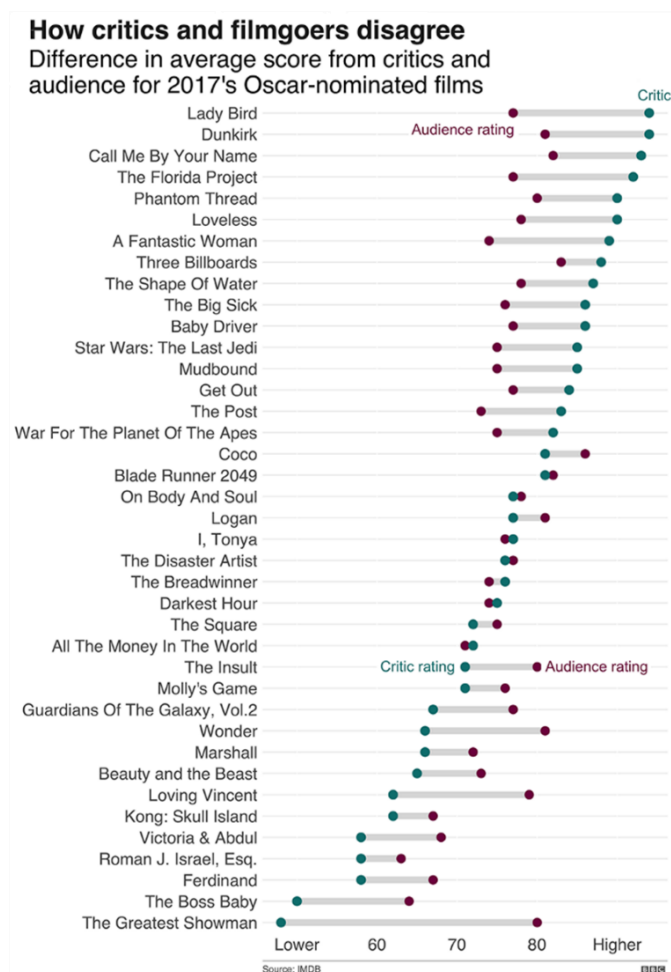
Рисунок № 50. Применение систематизированных данных в CSV-файлы  
<https://rostrosyvocesdelconflicto.elsalvador.com>

В вышеуказанном примере команда из пяти человек изучила четыре различные печатные газеты с 1972 по 1992 год. Найденную информацию о нарушениях прав человека и возможных преступлениях против человечности и о военных преступлениях агрегировали в электронные таблицы Google. Фотограф El Diario de Hoy сделал снимок высокого разрешения каждого элемента памятника. После этого с помощью библиотеки OpenCV применили пороговую обработку к каждому изображению, а затем использовали инструмент Tesseract для распознавания имен. Затем, используя библиотеку языка программирования Python под названием Pillow, каждая фотография

была вырезана, а результаты систематизированы в CSV-файлы, которые были объединены для хранения.

Одним из наглядных примеров в журналистике данных является практика телекомпании Би-би-си (BBC). За последние несколько лет журналисты команды BBC Visual and Data Journalism значительно изменили способ создания графиков для публикации на сайте BBC News. Мы нашли практическую статью BBC о том, как и почему их команда использовала пакет ggplot2 из R для создания готовых к производству графиков, задокументировали процесс и код, тем самым поделились личным практическим опытом. Оставим тут ссылку для более детального изучения - <https://medium.com/bbc-visual-and-data-journalism/how-the-bbc-visual-and-data-journalism-team-works-with-graphics-in-r-ed0b35693535>

Отметим, что журналисты BBC News по визуальной журналистике используют R для сложного и воспроизводимого анализа данных и создания прототипов. К примеру, они использовали R для извлечения, обработки, очистки и изучения данных из сотен электронных таблиц о том, достигаются ли целевые показатели национальной системы здравоохранения (NHS), для проекта NHS tracker, который удостоился награды. В 2017 году команда проанализировала более 8 миллионов сделок с жилой недвижимостью в Англии и Уэльсе для проекта по изучению изменения цен на жилье в реальном выражении, который также получил награду Королевского статистического общества. Ниже пример графики, который был построен благодаря R:



### Что делает пакет 'bbplot'?

Как говорится в статье, пакет был разработан для того, чтобы справиться со всеми повторяющимися препятствиями, упрощая рабочий процесс добавления объектов, которые должны быть во всех графиках. Когда команда BBC начала работать в R, каждый раз при создании графика приходилось настраивать каждый отдельный элемент, чтобы перейти от стандартного стиля ggplot к собственному стилю BBC. Также они отмечают то, что было тесное сотрудничество с дизайнерами из отдела визуальной журналистики и журналистики данных. Таким образом, команда собрала все решения в одном месте, чтобы обеспечить последовательность и сделать все как можно более легко воспроизводимым в bbplot. По словам создателей, основная идея заключалась в том, что `bbc_style()`, функция, которую создала команда BBC для изменения внешнего вида ggplot2 по умолчанию на собственный стиль BBC, должна помочь другим на 90%, оставляя возможность вносить любые дополнительные изменения в график. Более того, команда BBC собрал cookbook, своего рода справочник маленьких советов и хитростей, где можно найти все знания команды о ggplot2.

Однако, команда BBC напоминает, что R не решает все потребности журналистов в построении диаграмм. R не используют для интерактивной графики. Для этого лучше подходит библиотека D3 Javascript. Но для статических графиков R и ggplot2 очень полезны и удобны.

## 7.2. Системы статистического программирования: R (r-project.org) или Pandas (pandas.pydata.org)

Журналист должен обладать определенным багажом знаний в технологиях, чтобы не только более свободно и самостоятельно использовать эти технологии, но, и чтобы общаться с техническими специалистами на одном языке. Базовые познания в языках программирования становятся нормой для журналистов во всем мире. Знание HTML и умение программировать с помощью простейшего языка программирования «R» — это новая степень свободы журналиста и его независимости от других.

(<https://unesdoc.unesco.org/ark:/48223/pf0000374922/PDF/374922rus.pdf.multi.page=58> )

R — это язык и среда для статистических вычислений и графики. Это проект GNU, который похож на язык и среду S, разработанную в Bell Laboratories (ранее AT&T, сейчас Lucent Technologies) Джоном Чемберсом и его коллегами.

R предоставляет широкий спектр статистических (линейное и нелинейное моделирование, классические статистические тесты, анализ временных рядов, классификация, кластеризация и т. д.) и графических методов, а также обладает высокой расширяемостью. R предоставляет путь с открытым исходным кодом для участия в этой деятельности. Одной из сильных сторон R является простота создания хорошо оформленных графиков, пригодных для публикации.

R включает в себя:

- эффективные средства обработки и хранения данных;
- набор операторов для вычислений с массивами, в частности с матрицами;
- большой, последовательный, интегрированный набор промежуточных инструментов для анализа данных;
- графические средства для анализа и отображения данных на экране или на бумаге, а также хорошо разработанный, простой и эффективный язык программирования, который включает в себя

условные выражения, циклы, определяемые пользователем рекурсивные функции и средства ввода и вывода.

R - это очень мощный язык статистического программирования, популярный среди журналистов работающих с данными. Он содержит множество удивительных пакетов для работы с данными и позволяет создать воспроизводимый процесс анализа данных.



Пакеты R для обработки данных

В интересной статье «How and why we made our first audio data visualization» на Medium журналисты делятся опытом создания аудиовизуализации. Основная идея аудиовизуализации заключается в том, что если слушатель закроет глаза, то сможет слушать, как разворачивается история. Как и в случае с визуализацией данных, где каждая визуальная картинка сопоставляется с точкой данных, визуализация аудиоданных должна быть сопоставлена с отдельным аудиофайлом.

Чтобы завершить весь интерактив, они создали шаблонное React-приложение (create-react-app) и начали создавать соответствующие компоненты, используя ReactStrap, React-Player и Nivo-Calendar. React был использован для интерактивных элементов в течение восьми месяцев и показал преимущества этой среды разработки компонентов многократного использования. Одним из таких компонентов является D3.js компонент календаря Nivo.rocks, который был использован журналистами для быстрой визуализации данных.

Pandas – программная библиотека на языке Python для обработки и анализа данных. Pandas предоставляет быстрые, гибкие и выразительные структуры данных, разработанные для того, чтобы сделать работу с «реляционными» или «маркированными» данными простой и интуитивно понятной. Его цель – стать фундаментальным высокоуровневым строительным блоком для практического анализа реальных данных на Python. Кроме того, у него есть более широкая цель – стать самым мощным и гибким инструментом анализа и манипулирования данными с открытым исходным кодом, доступным на любом языке.

Возможности Pandas:

- Быстрый и эффективный объект DataFrame для манипулирования данными с интегрированной индексацией.

- Инструменты для чтения и записи данных между структурами данных в памяти и различными форматами: CSV и текстовые файлы, Microsoft Excel, базы данных SQL и быстрый формат HDF5.

- Интеллектуальное выравнивание данных и интегрированная обработка отсутствующих данных: автоматическое выравнивание на основе меток при вычислениях и легкое преобразование беспорядочных данных в упорядоченную форму.

- Гибкое изменение формы и поворот наборов данных.

- Интеллектуальная нарезка на основе меток, фантазийное индексирование и подмножество больших наборов данных.

- Колонки можно вставлять и удалять из структур данных для изменения размера;

- Агрегирование или преобразование данных с помощью мощного механизма группировки, позволяющего выполнять операции разделения-применения-комбинирования наборов данных.

- Высокопроизводительное слияние и объединение наборов данных.

- Иерархическое индексирование по осям обеспечивает интуитивно понятный способ работы с высокоразмерными данными в низкоразмерной структуре данных.

- Функциональность временных рядов: создание диапазона дат и преобразование частоты, статистика скользящего окна, сдвиг и запаздывание дат. Даже создание специфических для домена временных смещений и объединение временных рядов без потери данных.

- Высокая оптимизация производительности, с критическими путями кода, написанными на Cython или C.

- Python с pandas используется в самых разных академических и коммерческих областях, включая финансы, неврологию, экономику, статистику, рекламу, веб-аналитику и многое другое.

### 7.3. Географические информационные системы (ГИС): Quantum GIS, ArcGIS, GRASS

**Quantum GIS (QGIS)** – это географическая информационная система с открытым исходным кодом, позволяющая управлять геоданными, отображать, редактировать и анализировать их, а также создавать макеты карт. Благодаря интеграции с GRASS, эта ГИС обладает мощным аналитическим функционалом. QGIS работает в Linux, Unix, Mac OSX и Windows, поддерживает множество векторных, растровых форматов, а также различные базы данных. ([https://live.osgeo.org/archive/8.0/ru/overview/qgis\\_overview.html](https://live.osgeo.org/archive/8.0/ru/overview/qgis_overview.html) )

Quantum GIS может получать доступ к векторным данным, хранящимся в широком разнообразии форматов, включая файловые (например, ESRI Shape Files, KML, GML), базы геоданных (например, PostgreSQL/PostGIS, ODBC, ESRI Personal GeoDatabase, SQLite), и сетевые протоколы (OPeNDAP, GeoJSON); растровые данные в одном из более чем 40 форматов, поддерживаемых базовой растровой библиотекой GDAL (включая NetCDF, HDF5, GeoTIFF, GRIB и JPEG-2000); и сервисы визуализации и доступа к данным Open Geospatial Consortium (Web Map и Web Feature Services [WMS и WFS, соответственно]). В зависимости от конфигурации хост-системы, QGIS также может выступать в качестве альтернативного графического интерфейса пользователя для большой коллекции функций обработки геопространственных данных GRASS GIS. QGIS включает в себя архитектуру "подключаемых модулей", в которой можно разрабатывать и использовать расширения основной функциональности приложения; в настоящее время подключаемые модули включают поддержку интеграции GPS, взаимодействие с серверами данных OpenStreetMap и инструменты преобразования данных. (<https://old.dataone.org/software-tools/quantum-gis-qgis> )

**GRASS (Geographic Resources Analysis Support System)** – полнофункциональная геоинформационная система, предоставляющая мощные механизмы для обработки растровых и векторных геоданных в единой интегрированной программной среде. GRASS включает средства для пространственного моделирования, визуализации растровых и векторных данных, управления геоданными и их анализа, для обработки данных дистанционного зондирования Земли и аэрофотоснимков. В её состав входят продвинутый инструментарий временного анализа для обработки временных рядов и API для быстрой разработки на Python. GRASS была оптимизирована для высокой производительности и анализа больших объёмов пространственных данных. Помимо этого, в GRASS возможно создание сложной презентационной графики и макетов карт. GRASS переведена на множество языков и поддерживает большое количество форматов данных. GRASS используется по всему миру в академических институтах и коммерческих организациях, а также в правительственных учреждениях и компаниях по экологическому консалтингу. Эта ГИС работает на различных аппаратных платформах и является свободным программным обеспечением с открытым исходным кодом, распространяемым под лицензией GPL. ([https://gis-lab.info/docs/osgeo/ru/overview/grass\\_overview.html](https://gis-lab.info/docs/osgeo/ru/overview/grass_overview.html))

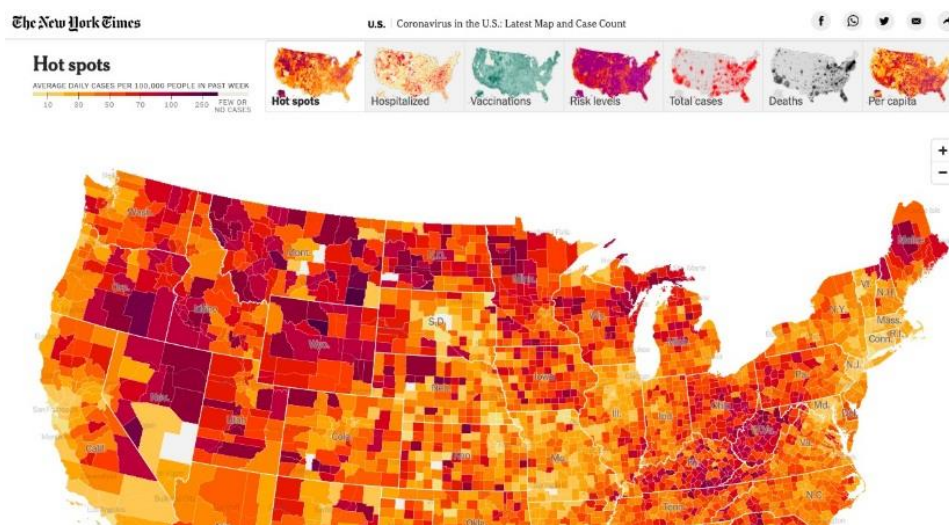


Рисунок № 51. Применение в журналистике данных Google Drive, Google Documents, Microsoft Excel, Microsoft Access и R. <https://www.nytimes.com/interactive/2021/us/covid-cases.html>

Данные часто попадают на тысячи страниц сайта [nytimes.com](https://www.nytimes.com), где с помощью JavaScript (D3 и Svelte) динамически генерируется текст и визуализации, которые объясняют последнее состояние вспышки в каждом штате и округе, в домах престарелых, колледжах, тюрьмах и т. д. Для управления данными и проведения анализа также использовались Google Drive, Google Documents, Microsoft Excel, Microsoft Access и R.

**Pont Map.** Точечные карты Pont Map очень удобны для отображения данных с широким географическим распределением. К примеру, некоторые компании имеют широкий спектр бизнеса. Если компания хочет просмотреть данные каждого сайта (конкретного местоположения) в определенной области, это будет сложнее осуществить с помощью общих карт, да и точность будет невысокой. Тогда для точного и быстрого позиционирования можно использовать карту точек.



**Follow maps.** Карты потоков часто используются для визуализации данных о потоках "начало - место назначения". Источником и пунктом назначения могут быть точки или поверхности. Данные о взаимодействии между точкой отправления и точкой назначения обычно выражаются линией, которая соединяет геометрический центр тяжести единицы пространства. Ширина или цвет линии указывает на значение направления потока между местом возникновения и местом назначения. Каждое пространственное место может быть как местом возникновения, так и местом назначения.

В статье «Top 10 Map Types in Data Visualization» вы можете найти различные виды карт для визуализации данных (<https://towardsdatascience.com/top-10-map-types-in-data-visualization-b3a80898ea70>)

Если нужно просто показать обработанные данные, можно выбрать Excel. В случае, если у вас есть все виды данных, но вы не разбираетесь в моделировании данных, программировании, очистке данных или даже оптимизации SQL, тогда нужно использовать простые инструменты визуализации данных, такие как FineReport или Tableau.

**Библиотеки визуализации, например d3.js ([mbostock.github.com/d3](https://github.com/d3)), Prefuse ([prefuse.org](https://prefuse.org)), Flare ([flare.prefuse.org](https://flare.prefuse.org))**

D3.js (или просто D3 – Data – Driven Documents) – это библиотека JavaScript, использующая цифровые данные для создания и контроля динамических и интерактивных графических элементов, которые могут отображаться в веб-браузере. Кроме всего прочего, D3 как инструмент визуализации данных поддерживает технологии, предусмотренные стандартом W3C – SVG (Scalable Vector Graphics – масштабируемая векторная графика), JavaScript, HTML5 и CSS3. Вот отличная галерея примеров использования этой библиотеки.

*Основные принципы работы библиотеки D3:*

- «бесшовная» интеграция с самыми современными технологиями вроде HTML5 и CSS3;
- использование стандартной, универсальной модели DOM, которая не только работает с любым языком, но и позволяет представить любой документ (в частном случае, представление набора данных) в удобном для конечного пользователя формате;

- прозрачность разработки – функции библиотеки не скрыты внутри неких классов (как это происходило с некоторыми ее предшественниками), а могут использоваться в любом месте документа при условии обращения к библиотеке;
- контроль над представлением данных – как в рамках программирования, так и относительно конечного результата визуализации.

**Prefuse** – это инструмент на базе Java для создания интерактивных приложений визуализации информации. Prefuse поддерживает богатый набор функций для моделирования данных, визуализации и взаимодействия. Он предоставляет оптимизированные структуры данных для таблиц, графов и деревьев, множество методов компоновки и визуального кодирования, а также поддержку анимации, динамических запросов, интегрированного поиска и подключения к базе данных. Prefuse написан на языке Java с использованием библиотеки Java 2D graphics library и легко интегрируется в приложения Java Swing или веб-апплеты. Сложная в использовании и освоении журналистами.

**Flare** – это библиотека ActionScript для создания визуализаций, запускаемых в Adobe Flash Player. Набор инструментов поддерживает управление данными, визуальное кодирование, анимацию и методы взаимодействия - от базовых диаграмм и графиков до сложной интерактивной графики. Более того, Flare имеет модульную конструкцию, которая позволяет разработчикам создавать индивидуальные методы визуализации без необходимости изобретать велосипед. Сложная в использовании и освоении журналистами.

#### **Инструмент для обработки данных: Google Refine**

**OpenRefine (ранее Google Refine)** – это мощный инструмент для работы с беспорядочными данными: их очистки, преобразования из одного формата в другой, расширения с помощью веб-сервисов и внешних данных. (<https://openrefine.org>)

OpenRefine всегда сохраняет конфиденциальность ваших данных на вашем собственном компьютере, пока вы не захотите поделиться ими или сотрудничать. Ваши личные данные никогда не покидают ваш компьютер, если вы сами этого не захотите. (Он работает, запуская небольшой сервер на вашем компьютере, а вы используете веб-браузер для взаимодействия с ним). OpenRefine доступен на более чем 15 языках.

**Основная задача программы** — переработка и очистка данных. Звучит довольно абстрактно, до тех пор, пока ее не попробуешь в деле. Среди ее основных возможностей можно перечислить:

- Возможность быстро получить представление о больших массивах информации
- Перерабатывать данные в нужный формат, делать базовые расчеты
- Фильтровать и объединять данные
- Находить ошибки и неожиданности — например, слишком большие цифры, слова вместо чисел, пустые значения
- Автоматически находить потенциальные опечатки и несоответствия в названиях, позволяя приводить записи к единому виду (кластеризация текстовых записей)
- Отправлять http-запросы к любому api, к примеру, для геокодирования адресов
- Сверять (reconcile) данные с открытыми базами данных, (в первую очередь Freebase) для уточнения названия и загрузки комплиментарных данных (см. видео №3 в конце поста)
- Создавать «макросы», автоматизируя обработку данных
- Большое преимущество OR – его простота. Программа не сложнее, чем, к примеру, excel (<https://medium.com/data-journalism/open-refine-a2a7f595934d>)

**Программное обеспечение для визуализации без программирования: например, ManyEyes, Tableau Public ([tableausoftware.com/products/public](http://tableausoftware.com/products/public))**

В последние годы в журналистике развернулась оживленная дискуссия о том, как рассказывать истории. Под влиянием новых технологий и растущей конвергенции СМИ появились новые формы и гибридные жанры (например, аудиослайдшоу, геймифицированные интерактивные материалы, моушн-графика или VR-фрагменты), которые ученые и практики часто объединяют под расплывчатым термином "мультимедийное повествование", "журналистика длинных форм" или "повествовательная журналистика онлайн". Эти новые продукты выходят далеко за рамки традиционных текстовых жанров, таких как новости, очерк или мнение. Они пересекают границы изображений, текстов и цифр, фактов и вымысла, дистанции и погружения; они смешивают письмо и рисунок, рассказ и показ, повествование и исследование; они сочетают объективность и субъективность, грамотность и устность. Таким образом, они стоят в традициях повествовательной журналистики, также называемой литературной журналистикой, которая стремится найти частную историю за публичной историей. Одной из таких новых форм, получивших огромный импульс после появления журналистики данных, является визуализация данных.

В настоящее время мы являемся свидетелями все более широкого использования визуализации данных в журналистике, поскольку данные становятся видимыми и общедоступными благодаря их визуализации. Журналисты и дизайнеры используют не только стандартные виды визуализации данных, такие как гистограммы или карты, но и создают новые формы, разработанные специально для того, чтобы рассказать историю наиболее понятным и увлекательным способом. Здесь мы снова сталкиваемся с термином *storytelling*: рассказ с помощью данных.

Мы должны помнить, главная цель визуализации – упростить и ускорить восприятие информации. Выбранный инструмент, формат и тип графика должны этому способствовать, и облегчать задачу. В зависимости от поставленной задачи, визуальный материал можно разрабатывать при помощи програм-мирования или обойтись без нее. В этом подразделе мы рассмотрим программные обеспечения для визуализации без метода применения программирования.

**Tableau** – это решение для бизнес-аналитики, которое позволяет визуализировать данные. Пользователи могут создавать и совместно использовать интерактивную приборную панель, которая отображает тенденции, отклонения и плотность данных в виде графиков и диаграмм. Для получения и обработки данных Tableau может подключаться к файлам, реляционным базам данных и источникам Big Data. Программа допускает смешивание данных и совместную работу в режиме реального времени, что делает ее очень необычной. Его используют для визуального анализа данных корпорации, академические исследователи и многие правительственные организации. В Магическом квадранте Gartner для платформ бизнес-аналитики и аналитики она также является лидером.

Tableau считается одним из лучших инструментов бизнес-аналитики и визуализации данных, и с момента своего появления он неоднократно возглавлял хит-парады. Самой важной особенностью этого приложения является то, что оно значительно упрощает для пользователей организацию, управление, визуализацию и интерпретацию данных. Данные могут быть настолько сложными и загадочными, насколько мы можем себе представить, и извлечение смысла из них требует правильных инструментов. Такие инструменты позволяют нам глубоко проникнуть в данные, чтобы выявить тенденции и получить ценные сведения. Tableau предоставляет нам группу инструментов, с помощью которых мы можем попытаться обнаружить данные, визуализировать их и поделиться информацией на глубоком уровне.

Функциональность Tableau, позволяющая легко перетаскивать данные, является интригующей особенностью, которая помогла ему стать любимым BI-приложением. Для работы с Tableau не нужно иметь техническое образование или знать много кодировок. Чтобы создавать информационные панели и оценивать отчеты, вы можете просто освоить этот инструмент, зная и освоив его возможности и функции, основанные на пользовательском интерфейсе.

### Google Fusion Tables

Этот онлайн-инструмент для работы с базами данных и картами позволяет быстро создать карты, где требуется увеличение масштаба. Вы получаете все высокое разрешение карт Google, но он может открывать большое количество данных - например, 100 Мб CSV. При первом использовании таблицы Fusion могут показаться немного сложными, но вы не пожалеете. Инструмент можно использовать для создания карт, подобных приведенной выше карте Ирака, а также пограничных карт, подобных этой карте бездомных.

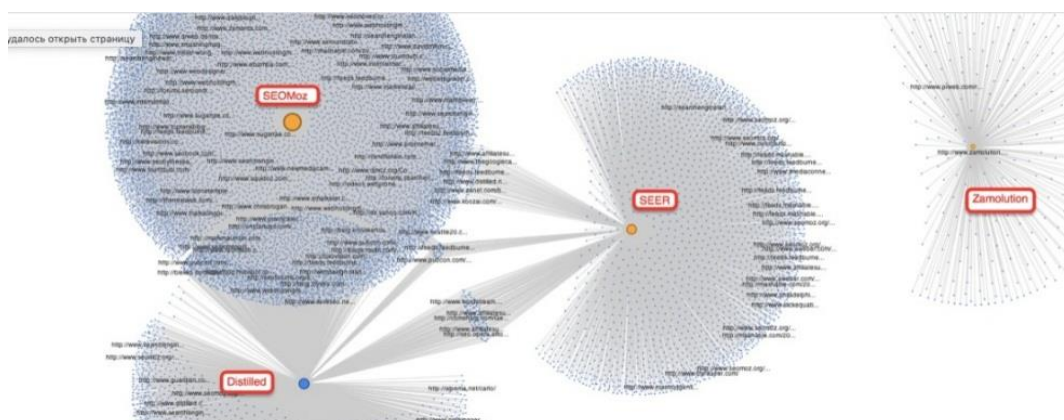


Рисунок № 52. Применение Fusion Tables в визуализации данных

<https://www.seerinteractive.com/blog/visualize-your-backlinks-with-google-fusion-tables/>

Помимо хранения огромного количества данных, Fusion Tables может визуализировать связи между вашими данными в виде простой интерактивной карты. Эта функция Network Graph настраивается за несколько секунд и может иметь множество различных применений.

Основным преимуществом является гибкость - вы можете загрузить, например, файл kml с границами регионов, а затем объединить его с таблицей данных. Он также получает новый пользовательский интерфейс, что должно упростить его использование. Не нужно быть программистом, чтобы создать такую карту - этот инструмент Fusion layers позволяет объединить различные карты или создать параметры поиска и фильтрации, которые затем можно вставить в блог или на сайт.

Рекомендация: используйте shapescare для преобразования официальных файлов shp в таблицы Fusion. Также остерегайтесь слишком сложных карт - Fusion не справляется с более чем миллионом точек в одной ячейке.

**Grafana** — это программное обеспечение для визуализации с открытым исходным кодом, позволяющее пользователям создавать динамические приборные панели и другие визуализации. Она поддерживает смешанные источники данных, аннотации и настраиваемые функции оповещения, а также может быть расширена с помощью сотен доступных плагинов. Это делает его одним из самых мощных инструментов визуализации.

Функции экспорта позволяют разработчикам делиться снимками приборных панелей, а также приглашать к сотрудничеству других пользователей. Grafana поддерживает более 50 источников данных с помощью плагинов. Ее можно скачать бесплатно, а облачная версия стоит \$49 в месяц. (Существует также очень ограниченная бесплатная версия для хостинга). В загружаемой версии также доступны планы поддержки, чего не предлагают многие другие инструменты с открытым исходным кодом.

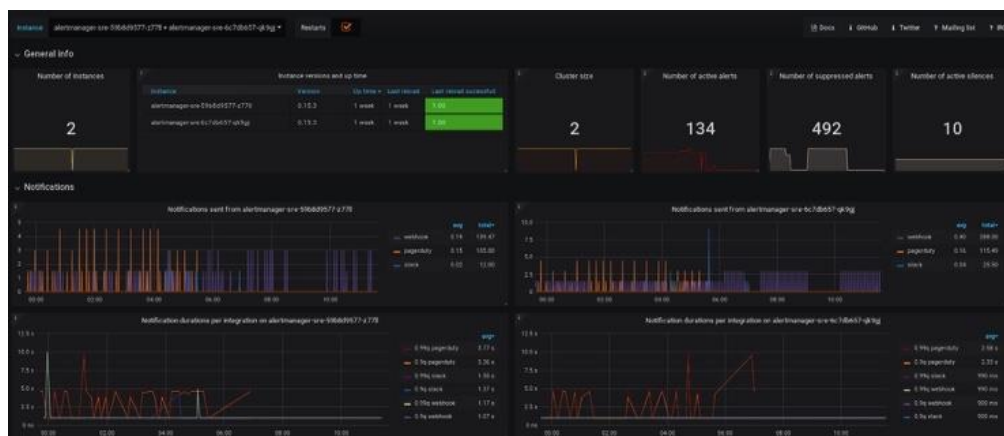


Рисунок № 53. Применение Grafana для визуализации источников данных с помощью плагинов

#### Плюсы Grafana

1. Открытый исходный код, доступны бесплатные и платные варианты.
2. Большой выбор источников данных.
3. Разнообразие типов диаграмм.
4. Простота создания динамических информационных панелей.
5. Возможность работы со смешанными источниками данных.

#### Минусы Grafana

1. Сложный для создания простых визуализаций.
2. Не предлагает столько возможностей по настройке визуализации, как другие инструменты.
3. Не лучший вариант для создания изображений визуализации.
4. Невозможно встраивать приборные панели в веб-сайты, хотя это возможно для отдельных панелей.

#### Datawrapper

Datawrapper был создан специально для добавления графиков и карт в новостные материалы. Созданные графики и карты являются интерактивными и предназначены для вставки на новостные сайты. Однако источники данных ограничены, и основным методом является копирование и вставка данных в инструмент.

После импорта данных диаграммы создаются одним щелчком мыши. Типы визуализации включают столбчатые, линейные и столбиковые диаграммы, диаграммы выборов, диаграммы областей, диаграммы рассеяния, хороплеты, карты символов, карты локаторов и другие. Готовые визуализации напоминают те, что можно увидеть на таких сайтах, как New York Times или Boston Globe. На самом деле, их диаграммы используют такие издания, как Mother Jones, Fortune и The Times. Бесплатный тарифный план идеально подходит для встраивания графики на небольших

сайтах с ограниченным трафиком, но платные тарифные планы стоят дороже - от 39 долларов в месяц.

#### Плюсы Datawrapper:

1. Специально разработан для визуализации данных в новостных отделах.
2. Бесплатный тарифный план хорошо подходит для небольших сайтов.
3. Инструмент включает встроенную проверку на дальтонизм.

#### Минусы Datawrapper:

1. Ограниченные источники данных.
2. Платные тарифные планы стоят дороже.

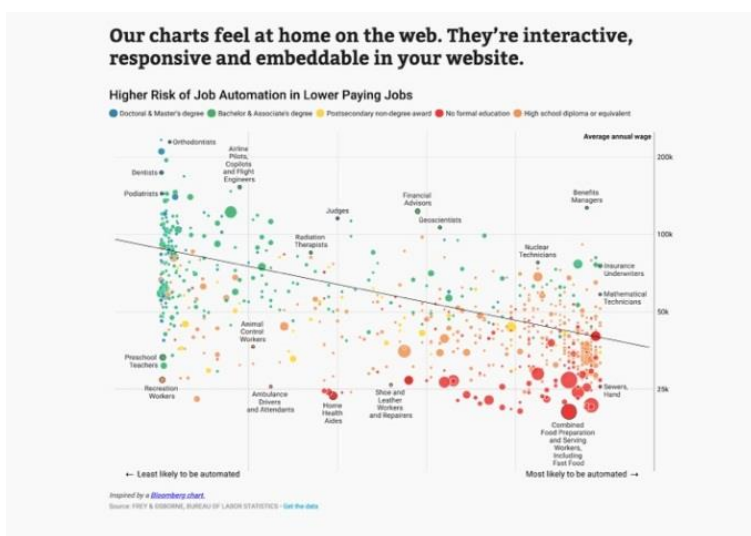


Рисунок № 54. Применение Datawrapper для создания интерактивных график и карт

#### Color brewer

Не являясь, строго говоря, инструментом визуализации, Color Brewer - изначально созданный при федеральном финансировании и разработанный в штате Пенсильвания - действительно предназначен для выбора цветов карты, и с ним стоит провести некоторое время, если вы планируете сделать много других. Вы можете выбрать базовый цвет и получить коды для всей палитры.

Приведенные выше инструменты не являются единственными, это лишь те, которые используются чаще всего. Существует множество других, в том числе:

**Chartsbin** – инструмент для создания кликабельных карт мира.

**iCharts** – специализируется на небольших виджетах диаграмм.

**Geocommons** – обмен данными и данными о границах для создания глобальных и локальных карт.

**Piktochart.com** – предоставляет шаблоны для текстовых/числовых визуализаций, которых сейчас очень много.

Визуализация данных является незаменимым инструментом при подготовке новостей. Они передают сюжеты, то, как люди изучают данные, и уточняют информацию. Однако не все визуализации данных достигают одной и той же цели. Некоторые из них отлично подходят для

сравнения двух переменных, другие показывают более широкую картину или указывают на тенденции.

Вопрос об инструментах – не самый простой. Каждый доступный инструмент визуализации данных имеет свои плюсы и минусы. Визуализация и работа с данными должны быть простыми и менее затратными. Сложные алгоритмы изменения параметров визуализации может иногда занять несколько часов, что не позволит журналистам экспериментировать так много, сколько хотелось бы. Это не обязательно означает, что вам не нужно учиться пользоваться инструментом. Но как только вы его освоили, он должен быть действительно эффективным.

Визуализация данных может быть поразительно красивой и привлекать внимание – ценная социальная валюта для обмена информацией и привлечения читателей, - но и использовать мощное когнитивное преимущество: половина человеческого мозга посвящена обработке визуальной информации. Когда вы представляете пользователю информационную графику, вы обращаетесь к нему через самый высокоскоростной путь сознания. Хорошо продуманная визуализация данных может произвести на зрителей мгновенное и глубокое впечатление. Несмотря на эстетическую привлекательность, она менее эмоциональна и больше направлена на то, чтобы констатировать факты. То есть, благодаря визуализации данных можно повествовать, опираясь на факты, а не на фанатизм, драматизм и романтизм истории.

#### 7.4. Работа с инструментами для более сложных графических объектов и карт

Научная визуализация, информационная визуализация и визуальная аналитика часто рассматриваются как три основные ветви визуализации. Мы в свою очередь рассматриваем информационную визуализацию.

Информационная визуализация — это изучение интерактивных визуальных представлений абстрактных данных для улучшения человеческого познания. Абстрактные данные включают как цифровые, так и нецифровые данные, такие как географическая информация и текст. Такие графики, как гистограммы, графики тенденций, диаграммы потоков и древовидные диаграммы, относятся к визуализации информации, и дизайн этих графиков преобразует абстрактные понятия в визуальную информацию.

**Dygraphs** — это библиотека диаграмм JavaScript с открытым исходным кодом для работы с огромными массивами данных. Она быстрая, гибкая и хорошо настраиваемая, работает во всех основных браузерах (включая IE8) и имеет активное сообщество.

Dygraphs определил для себя нишу использования и не будет идеальным решением для всех ваших потребностей. Но чаще всего он будет максимально полезен если вы работаете с большими наборами данных.

##### *Особенности*

- Работа с огромными наборами данных: dygraphs строит графики из миллионов точек, не загромождая их.
- Интерактивность из коробки: масштабирование, панорамирование и наведение курсора мыши включены по умолчанию.
- Сильная поддержка полос ошибок / доверительных интервалов.
- Высокая настраиваемость: используя опции и пользовательские обратные вызовы, вы можете заставить dygraphs делать практически все, что угодно.



Рисунок № 55. Применение Dygraphs для интерактивной графики

Данная графика является интерактивной: можно навести курсор мыши, чтобы выделить отдельные значения или щелкнуть и перетащить мышью, чтобы изменить масштаб. При двойном щелчке масштаб уменьшается. При перетаскивании с нажатой клавишей Shift происходит панорамирование. Вы можете изменить число и нажать Enter, чтобы настроить период усреднения.

**Vega** — это грамматика визуализации, декларативный язык для создания, сохранения и обмена интерактивными проектами визуализации. С помощью Vega можно описать внешний вид и интерактивное поведение визуализации в формате JSON и создать веб-представления с помощью Canvas или SVG.

Vega предоставляет базовые строительные блоки для широкого спектра проектов визуализации: загрузка и преобразование данных, масштабы, проекции карты, оси, легенды и графические метки, такие как прямоугольники, линии, символы черчения и т. д. Методы взаимодействия могут быть заданы с помощью реактивных сигналов, которые динамически изменяют визуализацию в ответ на потоки входных событий. Спецификация Vega определяет интерактивную визуализацию в формате JSON. Спецификации анализируются программой выполнения JavaScript Vega для создания статических изображений или интерактивных веб-визуализаций. Vega обеспечивает удобное представление для вычислительной генерации визуализаций и может служить основой для новых API и инструментов визуального анализа.

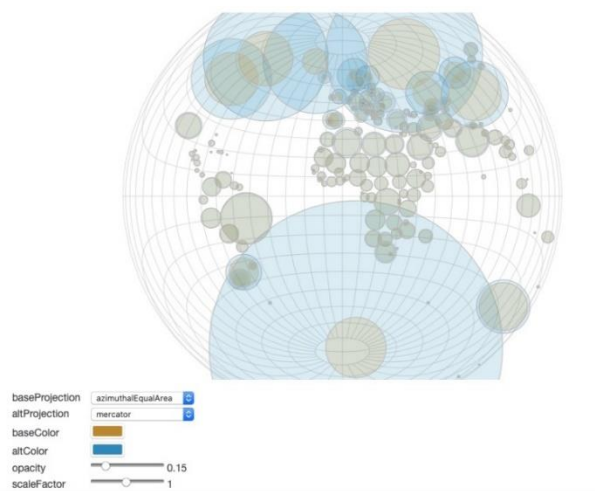


Рисунок № 56. Применение Vega для генерации визуализаций данных

Этот пример визуализирует разницу в прогнозируемой площади стран при использовании двух различных картографических проекций. Каждая страна абстрагируется в круг, размер которого соответствует площади этой страны в определенной картографической проекции. Круги стран располагаются на основе положения центроида страны в основной проекции. Затем накладывается второй набор окружностей с проецируемыми площадями, основанными на вторичной проекции. Этот пример демонстрирует функции языка выражения Vega geoArea и geoCentroid.

**PyGraphistry** — это библиотека Python для быстрой загрузки, формирования, встраивания и изучения больших графиков с помощью визуального анализатора графиков Graphistry с GPU-ускорением. Graphistry привносит человеческий интерфейс в эпоху больших и сложных данных. Она автоматически преобразует ваши данные в интерактивные, визуальные карты исследований, созданные для нужд аналитиков. Быстро выявляйте взаимосвязи между событиями и сущностями без написания запросов или перебора данных. Используйте все свои данные, не заботясь о масштабе, и поворачивайте их на лету, чтобы проследить, куда бы ни привело вас расследование. Идеально подходит для любых задач - от расследований в области безопасности, мошенничества и ИТ до 3600 представлений о клиентах и цепочках поставок.

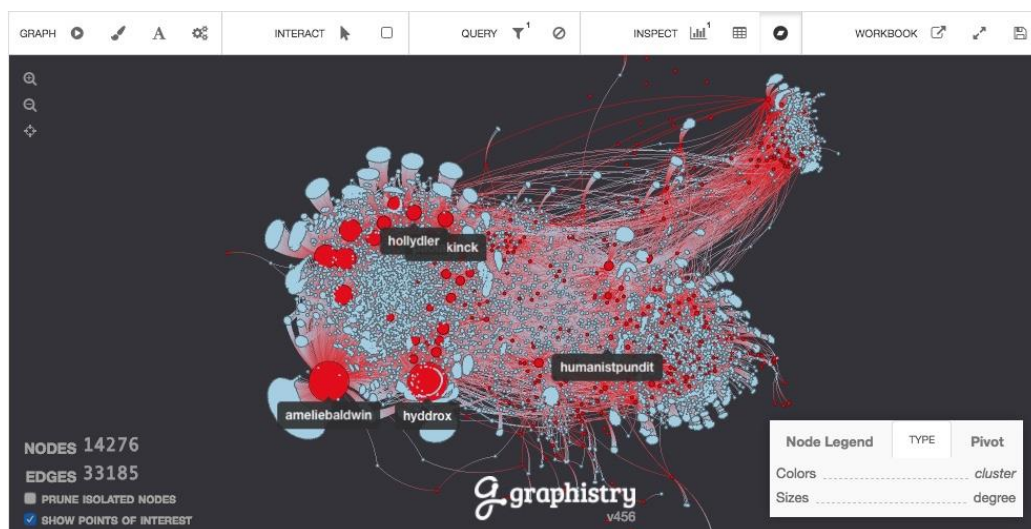


Рисунок № 57. Применение Graphistry для визуализации графиков

#### Бесплатные ресурсы по визуализации:

1. <https://datavizcatalogue.com>
2. Бесплатный курс от Табло - <https://www.tableau.com/learn/training/20213>
3. Каталог визуализации данных - <https://datavizcatalogue.com/RU/>
4. Курсы по OpenRefine - [https://youtu.be/B70J\\_H\\_zAWM](https://youtu.be/B70J_H_zAWM)

#### Задания для самостоятельной работы:

1. Создайте таблицу данных в Excel для построения диаграммы.
2. Найдите примеры журналистов стран СНГ, где были использованы инструменты визуализации данных.
3. Проанализируйте дата блог The Guardian datablog и определите часто используемые виды визуализации данных.
4. Преобразуйте свою таблицу данных в графический формат и создайте карту на Fusion Tables.

**Вопросы для самоконтроля:**

1. Что такое визуализация данных?
2. Позволяет ли визуализация данных быстро принимать решения и запоминать решения?
3. Почему журналисту желательно выбрать инструмент, который охватывает как работу с данными, так и их визуализацию?
4. Можно ли создавать анимированные пузырьки в электронных таблицах Google?
5. Чем отличаются табличные диаграммы Google от API диаграмм?
6. Каковы возможности R и Pandas?
7. Что такое Quantum GIS?
8. Каковы возможности GRASS (Geographic Resources Analysis Support System)?
9. Как называется библиотека JavaScript, использующая цифровые данные для создания и контроля динамических и интерактивных графических элементов?
10. Опишите основные задачи работы программы OpenRefine.
11. Нужно ли иметь журналисту техническое образование для работы на Tableau?
12. Расскажите об особенностях работы на Google Fusion Tables.
13. Можно ли в Google Fusion использовать более миллиона точек в одной ячейке?
14. Применяется ли визуализация данных при подготовке новостей?
15. Опишите возможности, в том числе плюсы и минусы Datawrapper.
16. Какая из библиотек JavaScript работает во всех основных браузерах (включая IE8)?
17. Расскажите о спецификации Vega.

## Модуль 8. РОЛЬ МЕДИЙНОЙ И ИНФОРМАЦИОННОЙ ГРАМОТНОСТИ (МИГ) В БОРЬБЕ С НЕКОРРЕКТНОЙ ИНФОРМАЦИЕЙ И ДЕЗИНФОРМАЦИЕЙ

### 8.1. Влияние некорректной информации и дезинформации на современное общество

Стремительное развитие информационных технологий за первые два десятилетия XXI века изменило привычные способы потребления информации. Печатные СМИ, радио и телевидение — главные источники или медиа распространения информации XX века — уже утратили или утрачивают свою актуальность. Сегодня современные общества в большей своей части получают информацию из Интернета, а именно из новостных сайтов или социальных сетей.

Интернет, в отличие от телевидения, радио или печатных СМИ, обладает следующим преимуществом: во-первых, это низкие издержки по созданию и распространению информации. Во-вторых, как отмечает американский исследователь интернет-технологий Клей Ширли, интернет — это та среда, в которой ее пользователи могут не только потреблять контент, но также выступать в роли его создателей: «Интернет — это первое средство коммуникации в истории, которое основана на поддержании групп и дискуссии одновременно. Телефон позволяет нам поддерживать связь один на один, телевизор, радио, журналы и книги позволяют осуществлять общение в одностороннем порядке. Интернет предоставляет нам возможности общения множества со множеством. <...> Это одно из огромных достижений», — говорил Ширли в своем выступлении на TED в 2009 году<sup>128</sup>. Ну и в-третьих, Интернет расширил форматы подачи информации: сегодня для современных новостных статей и заметки (текст), подкасты на Google или Apple Podcasts (аудио), видеоролики на YouTube или кэпш (caption) - видео в социальных сетях (видео).

Интернет действительно совершил революцию в межличностной и массовой коммуникациях в XXI веке, однако в то же время это породило определенные проблемы. Первая — децентрализованный и бесконтрольный Интернет стал причиной информационного переизбытка, который, по словам журналиста и писателя Николая Кононова, приносит усталость для современного человека<sup>129</sup>. Вторая и куда более серьезная проблема — огромное количество недостоверной и ложной информации (дезинформации) в Интернете.

«В Сети огромное количество площадок (сайты, блоги, соцсети, почта, форумы, чаты, СМИ, реклама и т. д.), которые предлагают информацию, часто не подтвержденную доказательствами, аргументами, фактами». — пишет исследовательница Елена Горина. Распространение дезинформации, по ее словам, происходит при переадресации сведений не при помощи ссылки-перепоста, а в форме вольного пересказа, в котором пересказчик чрезмерно свободно поступает с фактами. Это напоминает устно передаваемые в межличностной коммуникации сплетни, слухи — так называемое «сарафанное радио», когда часть информации трансформируется, искажается, а сведения не доходят до получателя в первичном виде, теряют свою суть, меняют значение, приобретают новые оттенки<sup>130</sup>.

---

<sup>128</sup> Ширки К. Как социальные сети создают историю // TED, июнь 2009. Точный адрес: [https://www.ted.com/talks/clay\\_shirky\\_how\\_cellphones\\_twitter\\_facebook\\_can\\_make\\_history?language=ru](https://www.ted.com/talks/clay_shirky_how_cellphones_twitter_facebook_can_make_history?language=ru)

<sup>129</sup> Кононов Н. В. Автор, ножницы, бумага. - Москва.: Издательство Манн, Иванов и Фербер, 2021

<sup>130</sup> Горина Е. В. Коммуникативные технологии манипуляции в СМИ и вопросы информационной безопасности. Учебное пособие. - Екатеринбург.: Издательство Уральского университета, 2016. [https://elar.urfu.ru/bitstream/10995/42384/1/978-5-7996-1807-0\\_2016.pdf](https://elar.urfu.ru/bitstream/10995/42384/1/978-5-7996-1807-0_2016.pdf)

При изучении дезинформации следует определиться с ее понятием. Ложная информация или дезинформация — это сообщение или данные, которые содержат в себе заведомо ложные, искаженные или недостоверные сведения. Создание и участие в распространении дезинформации может быть целенаправленным или ненамеренным, а в этом могут быть уличены не только человек, СМИ, но и государство. Весной 2003 года США и их союзники начали военную операцию против режима Саддама Хусейна в Ираке: администрация тогдашнего президента Джорджа Буша-младшего объясняло это вторжение тем, что правительство Хусейна разрабатывает оружие массового поражения, а также сотрудничает с террористической организацией «Аль-Каида». В феврале 2003 года, за месяц до начала вторжения, Колин Пауэлл, занимавший на тот момент пост госсекретаря США, выступил на заседании Совета Безопасности ООН с обоснованием для начала операции. На заседание Пауэлл принес пробирку с белым порошком и заявил, что в ней содержится образец производимого в Ираке оружия массового уничтожения. Позже выяснилось, что госсекретарь был дезинформирован Центральным разведывательным управлением (ЦРУ) США, а пробирки оказались фальшивыми.

Будет неверно полагать, что дезинформация — это побочный продукт развития интернет-технологий. Распространение ложной информации существовало и в XX веке, однако разница в том, кто мог участвовать в этом процессе. Исследователи Александр Федоров, Анастасия Левицкая и Андрей Новиков пишут, что возможностями для массового распространения неправдивой информации через медиа (пресса, телевидение, радио) в прошлом столетии обладали государственные структуры и частные организации, а население могло распространять фальшивую информацию только через слухи. «С массовым приходом интернета ложные слухи стали передаваться не только из уст в уста, а через социальные сети и желтые сайты. При этом нередко фальшивая информация в интернете распространяется быстрее, чем достоверная»<sup>131</sup>.

Начавшаяся в 2020 году пандемия коронавируса показала, насколько стремительно может распространяться дезинформация и слухи о вирусе с помощью Глобальной Сети. В интернете, социальных сетях, мессенджерах и видеохостингах начали появляться материалы и сообщения о том, что коронавирус — это заговор мирового правительства, а последовавшая за ней вакцинация породила очередную волну бездоказательных домыслов и сторонников этих теорий. Свой вклад в распространение дезинформации внесли и известные личности: например, 45-й президент США Дональд Трамп считал, что в лечении коронавирусной инфекции помогает противомаларийный препарат гидроксихлорохин, а президент Беларуси Александр Лукашенко советовал лечиться от коронавируса в банях и саунах, а также игрой в хоккей<sup>132</sup>.

Заслуженный режиссер из России Никита Михалков посвятил вакцинам от коронавируса целый фильм, в котором обвинил американского миллиардера и филантропа Билла Гейтса в «чипировании» людей под предлогом прививки<sup>133</sup>. В других теориях заговора американского бизнесмена уже обвинили в создании коронавируса, а разработанная им «чип-вакцина» будет применяться в борьбе с «неудобными людьми»<sup>134</sup>. Согласно еще одной теории, вакцина коронавируса вызывает аутизм у детей.

---

<sup>131</sup> Дезинформация // Википедия, последняя дата редакция статьи 09.09.2021.

<https://en.wikipedia.org/wiki/Disinformation>

<sup>132</sup> Злобин А. Пробирка ценой \$7 трлн. Трамп назвал худшую ошибку в истории США. // Форбс, 20.09.2018. <https://www.forbes.ru/biznes/367071-probirka-cenoy-v-7-trln-tramp-nazval-hudshuyu-oshibku-v-istorii-ssha>

<sup>133</sup> Федоров А. В., Левицкая А. А., Новиков А. С. Коронавирус как источник медийных манипуляций. // Crede Experto, 2020. <https://cyberleninka.ru/article/n/koronavirus-kak-istochnik-mediynyh-manipulyatsiy/viewer>

<sup>134</sup> Твиттер удалил ретвит Трампа о существовании лекарства от коронавируса // РБК, 29.06.2020. <https://www.rbc.ru/society/29/07/2020/5f20b8fb9a79470416a1dafa>

У теорий заговоров появились свои последователи, которые не верят в официальные и научные публикации о коронавирусе и необходимости прививок. В 2021 году во многих городах мира, включая такие страны Европы, как Франция и Греция, прошли митинги противников вакцинации. Часть протестов закончились столкновениями с полицией.

Всемирная организация здравоохранения и ООН считают, что инфодемия о коронавирусе — поток ложных слухов, непроверенной и намеренно искаженной информации, продвижение недоказанного лечения, — несет куда более серьезный вред, нежели сам вирус. Главный санитарный врач США Вивек Мурти заявил, что дезинформация о коронавирусе стало причиной того, что люди отказывались носить маски, игнорировали проверенные методы лечения и вакцинацию, что, в свою очередь, «привело к смертям и заболеваниям, которые можно было предотвратить»<sup>135</sup>. В Иране более 700 людей умерли от отравления алкоголем — погибшие употребляли ядовитый метанол как лекарство от коронавируса<sup>136</sup>. А в Великобритании сторонники конспирологической теории о том, что коронавирус распространяется через телевышки 5G, начали сжигать телекоммуникационные станции<sup>137</sup>.

Специалисты в области когнитивистики отмечают, что главная проблема дезинформации заключается в том, что она продолжает влиять на суждения и выводы даже после выхода опровержения. Феномен был исследован учеными университета Мичигана, которые распространили дезинформацию о пожаре на складе среди студентов, но после предоставили им опровержения. В итоге исследователи пришли к выводу что человек не может просто стереть из памяти первичную информацию, даже если опровержение будет представлено сразу же за дезинформацией. Исследовательница Диана Богданова пишет, что при разоблачении фейков «людям легче принять ложную информацию за истину, так как она кажется им знакомой»<sup>138</sup>.

## **8.2. Как страны Центральной Азии борются с распространением ложной и некорректной информации на законодательном уровне**

В четырех из пяти стран Центральной Азии (Казахстан, Кыргызстан, Узбекистан и Таджикистан) действуют законы, предусматривающие ответственность за распространение заведомо ложной или недостоверной информации. В Казахстане, Узбекистане и Таджикистане за распространение ложной информации предусмотрены схожие наказания: это штраф, общественные работы, исправительные работы, ограничение или лишение свободы.

### **Казахстан**

В Уголовном Кодексе Казахстана прописана уголовная ответственность за распространение ложной информации, она регулируется статьей 274. Распространителей ложной информации может ждать штраф, общественные работы, либо лишение свободы до 1 года. Если распространение ложной информации имело место при режиме чрезвычайного положения, то

---

<sup>135</sup> БесогонTV. Снятый с эфира выпуск БесогонTV выпуск «У кого в кармане государство?» // YouTube, БесогонTV от 01.05.2020. Точный адрес видео: <https://www.youtube.com/watch?v=MEtaaNNUVCI>

<sup>136</sup> Телеканал Дождь. Вакцина Гейтса. Откуда взялась теория заговора, за выпуск о которой программу Михалкова сняли с эфир. // YouTube, Телеканал Дождь от 04.05.2020. Точный адрес видео: <https://www.youtube.com/watch?v=OA5yxXfUIkU>

<sup>137</sup> Бешлей О. Чипирование, аутизм и заговоры. Что такое инфодемия и как дезинформация вредит людям. // Настоящее Время, 25.07.2021. <https://www.currenttime.tv/a/chto-takoe-infodemia-i-kak-ona-vredit-ludyam/31374201.html>

<sup>138</sup> В Иране более 700 человек умерли из-за употребления метанола как "лекарства" от COVID-19 // Интерфакс, 27.04.2020. <https://www.interfax.ru/world/706332>

нарушителей могут ограничить или лишить свободы сроком от 3 до 7 лет. При этом ответственность по настоящей статье несут не только авторы, но также распространители ложной информации<sup>139</sup>.

В 2019 году бывший президент Казахстана Нурсултан Назарбаев предложил ужесточить наказание за клевету. Министр информации и социального развития страны Даурен Абаев позже дал объяснение предложению экс-президента: по словам министра, «Назарбаева беспокоит рост анонимных источников, фейковых аккаунтов и страниц, участвующих в распространении неправды», а жесткие наказания за дезинформацию — это общемировая тенденция<sup>140</sup>. К моменту выхода этой публикации поправки, ужесточающую ответственность за дезинформацию, еще не приняты.

### **Кыргызстан**

23 августа 2021 года президент Кыргызстана Садыр Жапаров подписал закон «О защите от недостоверной (ложной) информации», который вступил в силу в тот же день<sup>141</sup>.

Закон был разработан депутатами 6 созыва парламента Гульшат Асылбаевой и Айнурой Осмоновой в 2020 году, еще до прихода Жапарова к власти после революции 4 октября 2021 года. Рассмотрение законопроекта в парламенте сопровождалось митингами и протестами со стороны гражданского общества. Предшественник Жапарова, Сооронбай Жээнбеков, на фоне недовольства в обществе не стал подписывать законопроект и отправил его на доработку. О законопроекте не вспоминали до лета 2021 года.

В принятом законе Асылбаевой и Осмоновой прописывается деанонимизация пользователей интернета, а также блокировка сайтов без решения суда, если там будет опубликована ложная информация.

С критикой закона о фейковой информации выступают юристы и правозащитники. По их мнению, этот закон противоречит нормам Конституции, нарушает права человека, а также предоставляет возможность для коррумпированных чиновников и людям близким к власти использовать закон в преследовании журналистов и граждан, критикующих власть и проводящих журналистские расследования. Общественный фонд «Институт Медиа Полиси» отмечает, что закон был принят с нарушениями установленных регламентов и без учета возражений бывшего главы государства.

Стоит отметить, что и самого Садыра Жапарова уличали в распространении недостоверной информации: в апреле 2021 года Фейсбук удалил его пост в социальной сети и Инстаграме о лечении коронавируса с помощью отвара из иссык-кульского корня или аконита. Как объяснял Президент КР, настойка аконита помогла ему выздороветь от болезни, когда он был в заключении, хотя сам аконит является ядовитым растением, а его применение, предупреждают ряд медиков и экспертов, может быть опасным для жизни и привести к смерти<sup>142</sup>.

---

<sup>139</sup> В Великобритании жгут вышки 5G из-за их якобы связи с распространением коронавируса. // Коммерсантъ, 07.04.2020. <https://www.kommersant.ru/doc/4316217>

<sup>140</sup> Богданова Д. Социальные сети как источник глобального риска. Как его нейтрализовать. // Школьные технологии, 2020. <https://cyberleninka.ru/article/n/sotsialnye-seti-kak-istochnik-globalnogo-riska-kak-ego-neytralizovat>

<sup>141</sup> Тилеухан Ж. COVID-19 и фейки | В каких случаях грозят штрафы и уголовная ответственность? // Factcheck.kz, 17.03.2020. <https://factcheck.kz/health/covid-19-i-fejki-v-kakix-sluchayah-grozyat-shtrafy-i-ugolovnaya-otvetstvennost/>

<sup>142</sup> Мамырханова М. За распространение фейков могут усилить ответственность в Казахстане. // Курсив, 23.08.2019. <https://kursiv.kz/news/obschestvo/2019-08/za-rasprostranenie-feykov-mogut-usilit-otvetstvennost-v-kazakhstan>

### **Узбекистан**

Весной 2020 года в Узбекистане ввели уголовную ответственность за распространение недостоверной информации о коронавирусе — вступивший в силу закон предусматривает нарушителям штрафы, общественные работы, ограничения свободы. Крайняя мера наказания — лишение свободы сроком до 3 лет<sup>143</sup>.

Однако уже в декабре парламент страны принял дополнительные поправки в закон про распространение дезинформации. Согласно изменениям, нарушителям могут грозить штрафы в размере 2 - 85 тысяч долларов, общественные работы, исправительные работы, а также лишение свободы. Максимальное наказание предусмотренное законом — ограничение или лишение свободы сроком до 5 лет<sup>144</sup>.

### **Таджикистан**

Летом 2020 года парламент Таджикистана ужесточил наказания за распространение ложной информации об эпидемиях и коронавирусе. Соответствующие поправки внесли в Административный и Уголовный кодексы страны.

В подписанном президентом страны Эмомали Рахмоном законе предусмотрены такие наказания как штрафы в размере 56-1125 долларов, либо административный арест до 15 суток<sup>145</sup>.

## **8.3. Медийная и информационная грамотность как инструмент просвещения в эпоху постправды**

Страны Центральной Азии — не единственные государства, чьи парламенты приняли санкционные законы против дезинформации на фоне распространения пандемии и инфодемии. Аналогичные законы были одобрены в Венгрии, Румынии, России, Сингапуре и Таиланде. В Сингапуре власти страны приняли самые жесткие санкции: максимальное наказание за распространение фейков предусматривает лишения свободы до 10 лет или свыше миллиона долларов штрафа<sup>146</sup>.

Джули Посетти, директор по глобальным исследованиям Международного центра для журналистов (ICFJ), считает, что запретительные законы «не доказали свою эффективность в борьбе с фейками», но отдельные правительства государств уже используют их в подавлении инакомыслия, преследовании профессиональной журналистики и контроле над публичными обсуждениями, представляющими общественно значимый интерес<sup>147</sup>.

Ключевая угроза дезинформации для общества — это обесценивание научных трудов и подрыв авторитетных институтов, и рост недоверия к ним. По мнению ученых Йельского

---

<sup>143</sup> Розиев И. Президент Жапаров подписал скандальный закон о фейках. Что это значит? // Kloop.kg, 23.08.2021. <https://kloop.kg/blog/2021/08/23/prezident-zhaparov-podpisal-skandalnyj-zakon-o-fejkah-chto-eto-znachit/>

<sup>144</sup> Facebook удалил публикацию Садыра Жапарова о лечении коронавируса аконитом. // Радио Азаттык, 19.04.2021. <https://rus.azattyk.org/a/facebook-udalil-post-sadyra-zhaparova-o-lechenii-koronavirusa-issyk-kulskim-kornem/31211192.html>

<sup>145</sup> В Таджикистане запретили распространять ложную информацию об эпидемиях. // Интерфакс-Азербайджан, 08.07.2020. <http://interfax.az/view/807102>

<sup>146</sup> Bothwell E. Fake news laws may 'catch on' during coronavirus. // The World University Rankings, 06.04.2020. <https://www.timeshighereducation.com/news/fake-news-laws-may-catch-during-coronavirus>

<sup>147</sup> Bothwell E. Fake news laws may 'catch on' during coronavirus. // The World University Rankings, 06.04.2020. <https://www.timeshighereducation.com/news/fake-news-laws-may-catch-during-coronavirus>

университета, в борьбе с дезинформацией нужно использовать не запретительные законы, а укреплять уважение к фактам, восстанавливать доверие общества к авторитетным источникам и институтам<sup>148</sup>. Исследователи разработали дорожную карту действий, которая состоит из таких пунктов и предложений как:

- Обучение пользователей интернета навыкам критического мышления и анализа потребляемой информации (общество должно быть осведомлено о том, как распространяется информация в сегодняшнем мире, о вреде дезинформации, как ее распознать и т.д.);
- Борьба с кликбейтом в сети (к примеру, обязывать пользователей знакомиться с материалами, прежде чем они смогут поделиться ими на своей ленте);
- Аккредитация производителей (журналистов, блогеров) и распространителей контента (новостные агрегаторы, блог платформы и т.д.);
- Запрет на использование ботов или введение идентификации пользователей интернета. Принятие пользовательских соглашений, запрещающих пользователям выдавать себя за других лиц и предусматривающих ответственность за такие нарушения;
- Финансирование местной, расследовательской журналистики;
- Изменение алгоритмов поисковых систем, где приоритет в показах отдается проверенной и точной информации, а не просмотрам или количеству кликов;

Как видно из рекомендаций ученых Йельского университета, борьба с дезинформацией состоит из комплекса мероприятий, в которые вовлечено государство, образовательные институты, технологические компании, СМИ и так далее. Реализация всех вышеперечисленных мер зависит от уровня мотивации заинтересованных лиц, их приверженности этических норм и принципов. Если технологические компании или новостные медиа могут быть не мотивированы в продвижении точной и достоверной информации, отдавая предпочтение сенсационному или провокационному контенту как способу привлечь внимание как можно большего количества пользователей, то обучение пользователей навыкам критического мышления, умению распознавать манипуляцию, дезинформацию и пропаганду, можно назвать одним из базовых и институциональных решений стоящей проблемы.

Медиаграмотность — а именно умение критически анализировать и оценивать потребляемую или распространяемую информацию — определенно важнейший навык современного человека в эпоху постправды. Европейский проект MediaCoach считает, что медиаграмотность — это «предпосылка для полноценного гражданства в информационном обществе», поэтому развитие данного навыка должно охватывать людей всех возрастов, родителей, педагогов и медиа-специалистов<sup>149</sup>.

Международная организация ЮНЕСКО предлагает рассматривать навык медиаграмотности в совокупности с информационной грамотностью, тем самым объединяя их в общий термин «медийно-информационной грамотности» или МИГ<sup>150</sup>. Основу МИГ составляют такие понятия как:

- Понимание роли и функции медиа в демократическом обществе;
- Понимание условий, при которых медиа могут выполнять свои функции;
- Взаимодействие с медиа для самовыражения и участия в демократических процессах;
- Критическая оценка информации;
- Соблюдение этики при использовании информации и так далее;

---

<sup>148</sup> Fighting Fake News // Yale Law School, 07.03.2017.

[https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting\\_fake\\_news\\_-\\_workshop\\_report.pdf](https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf)

<sup>149</sup> Асанбаева Г. Основные понятия и принципы МИГ // Учебное пособие по медиаграмотности, 23.10.2018.

<https://newreporter.org/uchebnoe-posobie-po-mediagramotnosti/>

<sup>150</sup> Wilson C. et all. Медийная и информационная грамотность: программа обучения педагогов. - Париж: UNESCO, 2012. [https://unesdoc.unesco.org/ark:/48223/pf0000192971\\_rus](https://unesdoc.unesco.org/ark:/48223/pf0000192971_rus)

Согласно Индексу медиаграмотности для европейских стран, способность общества противостоять фейкам, манипуляции и пропаганде зависят от таких факторов как состояние свободы и уровень образования<sup>151</sup>. В данном рейтинге первые пять мест заняли Финляндия, Дания, Эстония, Швеция и Ирландия — страны с высоким уровнем свободы слова. Все эти государства входят в топ-20 стран Всемирного индекса свободы слова.

| № | Страна    | Занимаемое место |
|---|-----------|------------------|
| 1 | Финляндия | 2                |
| 2 | Дания     | 4                |
| 3 | Эстония   | 15               |
| 4 | Швеция    | 3                |
| 5 | Ирландия  | 12               |

Таблица № 38. Всемирный Индекс Свободы Слова за 2021 год

Всемирный рейтинг уровня свободы слова составляется неправительственной организацией "Репортеры без границ". В рейтинге участвуют 180 стран мира. С другой стороны страны Центральной Азии (Казахстан, Кыргызстан, Узбекистан, Таджикистан и Туркменистан). 4 из них — Казахстан, Узбекистан, Таджикистан и Туркменистан — не входят даже в первые 100 государств этого рейтинга.

| № | Страна       | Занимаемое место |
|---|--------------|------------------|
| 1 | Кыргызстан   | 79               |
| 2 | Казахстан    | 155              |
| 3 | Узбекистан   | 157              |
| 4 | Таджикистан  | 162              |
| 5 | Туркменистан | 178              |

Таблица № 39. Всемирный Индекс Свободы Слова за 2021 год

Для измерения качества образования Индекс медиаграмотности ориентируется на рейтинг школьного образования PISA. Из всех стран Центральной Азии, в последнем рейтинге PISA в 2018 году участие принял только Казахстан, заняв 62 место из 79 участвовавших стран. Кыргызстан в рейтинге PISA участвовал только в 2009 году.

Странам Центральной Азии необходимо инвестировать в образование граждан, а также создание необходимых условий для развития независимой журналистики. Принятие запретительных законов на фоне пандемии коронавируса показало, что это не решает проблему с распространением дезинформации, однако усугубляет ее. Региональный консультант по медиаграмотности Internews Гульнар Асанбаева отмечает, что «несвободная пресса лишает граждан полной и достоверной информации»: «Мы не сможем по-настоящему участвовать в жизни своих сообществ, принимать верные жизненно важные решения, останемся под влиянием

<sup>151</sup> Асанбаева Г. Факторы влияния на уровень медиаграмотности в странах Центральной Азии // Медийно-цифровая грамотность в цифровом мире: как научить учителей. - ЮНЕСКО, 2021. [https://iite.unesco.org/wp-content/uploads/2021/05/MIL\\_Teachers\\_book\\_2021.pdf](https://iite.unesco.org/wp-content/uploads/2021/05/MIL_Teachers_book_2021.pdf)

политической пропаганды, недостоверной информации и уязвимыми перед любой дезинформацией»<sup>152</sup>.

#### **Задания для самостоятельной работы:**

1. Найдите примеры дезинформации, в создании и распространении которой участвовали государства или государственные СМИ. Какую цель или цели преследовались данными фейками?
2. Найдите примеры, где журналистов привлекали к ответственности за якобы распространение ложной информации. В чем заключается особенность этих дел? Как эти дела комментируют сами журналисты?
3. Найдите и перечислите фейки, которые распространялись в период выборов Президента США в 2016 году.
4. Найдите и перечислите 3 примера с фейковой информацией, в распространении которых участвовали знаменитости или лидеры мнений.
5. Назовите аргументы, по которым цензура — это плохое (или эффективное) решение в борьбе с фейками.

#### **Вопросы для самоконтроля:**

1. Назовите преимущества интернета перед другими каналами коммуникации (печатная пресса, ТВ, радио)?
2. Почему интернет стал площадкой, где распространение фейковой информации стало обыденным явлением?
3. Назовите определение фейка.
4. В чем заключается опасность фейков для общества?
5. Какие страны Центральной Азии приняли законы, предусматривающие ответственность за создание и распространение фейковой информации?
6. Назовите причины, по которым наказания за распространение фейков — это не самый эффективный метод в борьбе с дезинформацией.
7. Перечислите рекомендации по борьбе с фейковой информацией в интернете, предложенные учеными Йельского университета
8. Что такое медиаграмотность? Как медиаграмотность помогает в борьбе с фейками?
9. Какие факторы, согласно Индексу медиаграмотности, влияют на способности обществ противостоять фейкам, манипуляции и пропаганде?
10. В чем заключаются негативные последствия несвободной прессы для государства?

### **8.4. Журналистика, фейковые новости и дезинформация**

Интернет: новые возможности для журналистики, источник фейков и дезинформации

Переход журналистики в цифровое пространство породил понятие авторской журналистики — то есть формата, в котором лицом редакции выступает известный журналист или же известная личность. Трендсеттером этого формата стал спортивный журналист из России Юрий Дудь, который в 2017 году запустил одноименный канал на YouTube. Далее за Дудем свои каналы начали открывать блогеры Илья Варламов, Татьяна Мингалимова (Нежный редактор), Антон Лядов (The

---

<sup>152</sup> Как медиаграмотность способствует свободе слова? Мы спросили у шести экспертов из Казахстана, Таджикистана и Узбекистана // Stan Radar, 07.05.2021. <https://stanradar.com/news/full/44584-kak-mediagramotnost-sposobstvuet-svobode-slova-my-sprosili-u-shesti-ekspertov-iz-kazahstana-tadzhikistana-i-uzbekistana.html>

Люди), журналисты Павел Лобков (Моральный Камертон), Вадим Борейко (Гиперборей), Леонид Парфенов (Parfenon).

В 2019 году бывший тележурналист, медиаменеджер и основатель проекта «Редакция» Алексей Пивоваров ввел понятие ютуб-журналистики<sup>153</sup>. Развитие этого жанра, по словам Пивоварова, произошло благодаря двум факторам: государственному контролю и цензуре над телевидением в России, что отразилось на качестве контента, а также возможности зарабатывать достаточного количества денег за счет рекламы, необходимого для жизнеобеспечения редакций и развития.

Развитие интернет-технологий в период второго десятилетия (2010 - 2020 годы) принесло изменения и новые возможности в работе профессиональных редакций, среди которых можно выделить:

- Прямые трансляции при освещении общественно-политических событий (этот формат использовали редакции «Клооп» и «Азаттык» при освещении революции 4 октября 2020 года, приведшей к смене власти в Кыргызстане);

- Запись и выпуск подкастов (практикует редакция латвийского интернет-издания «Медуза», ориентированного на русскоязычную аудиторию в России и странах СНГ. Журналисты записывают подкасты на общественно-политические и другие темы).

- Нативная реклама (один из относительно новых способов заработка для интернет-изданий).
- Анализ данных и появление жанра дата-журналистики. (это стало возможным после того, как правительства стран начали выкладывать открытые данные в сеть, которые можно анализировать).

- Сбор пожертвований (в 2021 году Минюст России внес издание «Медуза» в список иностранных агентов, из-за чего редакция потеряла существенную часть дохода от рекламодателей и оказалась на грани закрытия. В результате редакция решила собирать пожертвования или донаты со своих читателей, а позже запустила мерчандайзинг в партнерстве с российскими дизайнерами).

Благодаря цифровым источникам значительно расширился охват журналистики, социальных сетей и взаимодействия с общественностью. Однако вместо того, чтобы использовать цифровые инструменты для информирования людей и активизации гражданской дискуссии, некоторые люди используют социальные и цифровые платформы для обмана, введения в заблуждение или нанесения вреда другим путем создания или распространения поддельных новостей и дезинформации.

Проблема дезинформации и сенсационных новостей — это не новомодный тренд. Например, война США с Испанией в конце XIX века за освобождение испанской колонии Куба произошла благодаря желтой прессе. Влиятельные на тот момент издатели Уильям Херст и Джордж Пулитцер не стеснялись публиковать недостоверные материалы (например, про обыск американской туристки, которую испанские таможенники раздели догола - эта новость вышла в газете «Нью Йорк Джорнал» и была полной выдумкой журналиста). Провокационные материалы побудили общественный запрос среди американцев на военную интервенцию в Кубе<sup>154</sup>.

Американский писатель Марк Твен, будучи журналистом, выдумывал сенсационные истории. Ирония заключается в том, что Твен в итоге сам стал жертвой дезинформации в 1897 году, когда

---

<sup>153</sup> TedxTalks. "YouTube-журналистика: есть ли она на самом деле? | Alexey Pivovarov | TEDxDvortsovayaSquare" YouTube. n.p., 5.11.2019. Web. 15.09.2021.  
URL: <https://www.youtube.com/watch?v=5I5k53rr3h0>

<sup>154</sup> Ефимова М. «Обеспечьте зарисовки, я обеспечу войну» // Радио Свобода. 06.05.2018.  
URL: <https://www.svoboda.org/a/29147182.html>

прочитал заметку в газете о своей смерти. Писатель, к слову, тогда отшутился на страницах газеты «Нью Йорк Джорнал» фразой, позже ставшей крылатой: «Слухи о моей смерти несколько преувеличены»<sup>155</sup>. Советский Союз на протяжении своего существования также намеренно скрывал информацию или дезинформировал общественность о крупных техногенных катастрофах и происшествиях: авария на Чернобыльской АЭС в 1986 году, взрыв баллистической ракеты на космодроме Байконур в 1960-м, массовая давка футбольных болельщиков на стадионе Лужники в Москве в 1982-м<sup>156</sup>.

Широкое применение термина фейкньюс начало использоваться после выборов президента США в 2016 году и победы Дональда Трампа над его оппоненткой Хиллари Клинтон. Волна компроматов, недостоверной информации и политической рекламы из сомнительных источников, сопровождавшая предвыборную кампанию, побудило ученых и исследователей заговорить о необходимости изучения деструктивной роли дезинформации<sup>157</sup>. Критики Трампа, в частности, утверждали, что 45-й президент США был избран благодаря фейковой информации и пропаганде, распространяемой в социальной сети Фейсбук. Дело дошло до того, что Конгресс США вызвал главу социальной сети Марка Цукерберга к себе на допрос<sup>158</sup>.

Пандемия коронавируса, начавшаяся в 2020 году, породила кучу домыслов и неправдивой информации о болезни. В Сети стали распространяться слухи и теории заговоров: например, об искусственном происхождении вируса, или же распространении вируса с помощью телевышек 5G. Ложная информация привела к трагедии в Иране, где сотни жителей погибли и тысячи отравились от употребления метанола как лекарства от коронавирусной инфекции. Фейк, что алкоголь помогает при лечении и профилактике коронавируса, распространялся в соцсетях<sup>159</sup>.

Слухи и домыслы не обошли стороной и разработанную в короткие сроки вакцину против коронавируса. Одна из распространенных теорий про вакцину утверждает, что под ее предлогом проводится массовая чипизация людей (эту теорию в частности поддерживал известный российский режиссер Никита Михалков). Другая — что вакцина вызывает аутизм у детей. В 2021 году в Греции, Франции, США и других стран мира прошли массовые митинги противников вакцинации и антипрививочников, часть протестов закончилась столкновениями демонстрантов с органами правопорядка<sup>160</sup>. И это несмотря на то, что ученым удалось доказать эффективность вакцин от тяжелого течения болезни, госпитализации и смерти<sup>161</sup>.

В 2020 году генеральный секретарь ООН Антонио Гуттериш со ссылкой на анализ одной из популярных соцсетей (Гуттериш не уточняет, что это за ресурс) заявил, что 40 процентов постов про коронавирус были созданы ботами — автоматическими программами, замаскированными под

---

<sup>155</sup> Твен, Марк // ИА «ТАСС». дата публикации неизвестна. URL: <https://tass.ru/encyclopedia/person/tven-mark>

<sup>156</sup> Трагедии, которые в СССР скрывали от народа // Новые Известия. 3.07.2017. URL: <https://newizv.ru/article/general/03-07-2017/tragedii-kotorye-v-sssr-skryvali-ot-naroda-5391ca91-7a8e-4f06-865c-acbd58ae2209>

<sup>157</sup> Качкаева А. Г., Шомова С. А., Душакова И. С. Фейк: от академических дискуссий к практическим решениям // Коммуникации. Медиа. Дизайн, 6(2), 5-32, 2021. URL: <https://cmd-journal.hse.ru/article/view/12600>

<sup>158</sup> Цукерберг: Фейсбук не виноват в том, что Трамп стал президентом // Би-Би-Си, 11.11.2016. URL: <https://www.bbc.com/russian/news-37946959>

<sup>159</sup> Манаева Н. В Иране 300 человек погибли от метанола, используя его как лекарство от коронавируса // Знак, 28.03.2020. URL: [https://www.znak.com/2020-03-28/v\\_irane\\_300\\_chelovek\\_pogibli\\_ot\\_metanola\\_ispolzuya\\_ego\\_kak\\_lekarstvo\\_ot\\_koronavirusa](https://www.znak.com/2020-03-28/v_irane_300_chelovek_pogibli_ot_metanola_ispolzuya_ego_kak_lekarstvo_ot_koronavirusa)

<sup>160</sup> Настоящее Время. Что такое инфодемия и как она вредит людям // Радио Азаттык. 26. 07.2021. URL: <https://rus.azattyk.org/a/chto-takoe-infodemiya/31377009.html>

<sup>161</sup> Работают ли вакцины против дельты? Это главный вопрос, который задают себе ученые. Ответ не самый приятный // Медуза. 13.08.2021. URL: <https://meduza.io/feature/2021/08/13/rabotayut-li-vaktsiny-protiv-delta-eto-glavnyy-vopros-kotoryy-zadayut-sebe-uchenye-otvet-ne-samyy-priyatnyy>

определенных людей. «Откровенная ложь распространяется в интернете ужасающими темпами», — заявил генсек ООН<sup>162</sup>.

Всемирная Организация Здравоохранения обеспокоена распространением фейков про коронавирус и вакцинацию и назвала происходящее инфодемией. «Ложные слухи приводят к поляризации общественного мнения по вопросам, касающимся COVID-19; провоцируют риторику ненависти; повышают риск конфликтов, насилия и нарушений прав человека и ставят под угрозу долгосрочные перспективы укрепления демократии, прав человека и социальной сплоченности», — говорится в заявлении организации. Опасность дезинформация, по мнению ученых Йельского университета, заключается в том, что фейки обесценивают научных труды исследователей, подрывают авторитет институтов и порождают недоверие к экспертному мнению в глазах общества<sup>163</sup>.

Почему пользователи интернета участвуют в распространении фейков? Интернет-ресурс для студентов Turnitin определяет следующие причины популярности фейков:

- **Фейки распространяются через знакомых.** Когда информация передается от человека к человеку, это не только обеспечивает скорость, а еще и порождает доверие.
- **Фейки укрепляют наши предрассудки.** Пользователи социальных сетей общаются с людьми, разделяющими их точку зрения. Алгоритмы соцсетей подстраиваются под интересы каждого пользователя — таким образом он или она оказывается запертым(ой) в мыльном пузыре единого мнения.
- **Сенсационность и простота.** Фейки апеллируют к эмоциям, что побуждает пользователей распространять сенсационную публикацию, не проверяя. Еще одна причина распространения фальшивых новостей — изменение модели потребления информации в интернете. Сегодня люди предпочитают узнавать о новостях не на сайте конкретного интернет-издания, а с помощью новостных агрегаторов<sup>164</sup>. К таковым относятся сервисы «Яндекс.Новости» и «Google News», «Новости Mail.ru», «Rambler Новости» и другие.

## 8.5. Виды и типы фейковой информации

В научной литературе ученых, исследующих природу фейков, существует множество определений и классификаций для неправдивой информации, такие как:

**Некорректная информация (misinformation):** данный термин в основном используется в англоязычной литературе, означает непреднамеренное участие (или же по незнанию) пользователя или человека в распространении ложной или неточной информации<sup>165</sup>.

---

<sup>162</sup> Генсек ООН: 40% постов о COVID-19 были генерированы ботами // Новости ООН.

4.05.2020. URL: <https://news.un.org/ru/story/2020/05/1377312>

<sup>163</sup> Fighting Fake News // Yale Law School,

07.03.2017. [https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting\\_fake\\_news\\_-\\_workshop\\_report.pdf](https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf)

<sup>164</sup> 4 причины привлекательности фейков // Turnitin. 27.08.2020. URL: <https://www.turnitin.com/ru/blog/4-prichiny-privlekatelnosti-feykov>

<sup>165</sup> Аналитики выяснили, откуда россияне чаще всего узнают новости // 360 TV.

18.09.2018. URL: <https://360tv.ru/news/nauka/analitiki-vyjasnili-otkuda-rossijane>

**Дезинформация:** намеренное и умышленное распространение ложной информации (классический пример — пропаганда)<sup>166</sup>.

**Сатира или фейк, созданный в юмористических целях**<sup>167</sup>

К такого рода публикациям, например, можно отнести новости сатирического проекта ИА «Панорама» («Time включил Александра Лукашенко в список влиятельных людей мира», «Песков сообщил, что во время самоизоляции Путин будет учиться играть на арфе»)<sup>168</sup>. Кыргызский аналог «Панорамы» — это пародийный твиттер-аккаунт Qloop («Сооронбай Жээнбеков попросил у ГКНБ на допросе факты его виновности»)<sup>169</sup>.

**Слухи, сплетни, городские легенды**<sup>170</sup>

Существуют также классификация фейков, касающаяся только журналистской сферы или фейк-ньюс. Такие определения, к примеру, упоминаются в работе Красовской Н., Гуляева А. и Юлины Г. (при классификации используется такой параметр как правдивость), а также автора книги «EVERYDATA: дезинформация, скрытая в небольших данных, которые вы потребляете каждый день» Джона Джонса<sup>171</sup>.

**Новости или публикации, ложные от начала и до конца.** Известным примером в журналистике является история бывшего журналиста The New Republic Стивена Гласса. В 1998 году Гласс выпустил в журнале историю про 15-летнего хакера, который взломал сеть IT-компании Jukt Micronics, а позже был нанят этой же компанией в качестве консультанта по безопасности. Позже внутреннее расследование The New Republic выявило, что Гласс полностью выдумал эту историю от начала и до конца<sup>172</sup>.

**Частично правдивые новости или публикации.** В 2018 году репортер немецкого издания Der Spiegel Клаас Релотиус был разоблачен своим коллегой в манипулировании и подмене фактов во время работы над материалом про отряды гражданской самообороны в США. Начатое расследование выяснило, что как минимум в 14 материалах Релотиуса есть выдуманные факты и персонажи. Клаас Релотиус был звездой Der Spiegel — его материалы перепечатывали

---

<sup>166</sup> Архангельская И. Б., Архангельская А. С. Фейк-ньюс в цифровую и доцифровую эпохи // Знак: проблемное поле медиаобразования, 2020. URL: <https://cyberleninka.ru/article/n/feyk-nyus-v-dotsifrovuyu-i-tsifrovuyu-epohi/viewer>

<sup>167</sup> Time включил Александра Лукашенко в список самых влиятельных людей мира // ИА «Панорама». 2021. URL: <https://panorama.pub/news/time-vklyuchil-aleksandra-lukashenko-v-spisok-samyx-vliyatelnyx-lyudej-mira> (дата обращения: 16.09.2021)

<sup>168</sup> Песков сообщил, что во время самоизоляции Путин будет учиться играть на арфе. // ИА «Панорама». 2021. URL: <https://panorama.pub/news/peskov-soobshhil-cto-vo-vremya-samoizolyaczii-putin-budet-uchitsya-igrat-na-arfe> (дата обращения: 16.09.2021)

<sup>169</sup> Qloop. «Сооронбай Жээнбеков попросил у ГКНБ на допросе факты его виновности» // [публикация в Твиттере]. <https://twitter.com/qloopnews/status/1439200715344334849> (дата обращения: 18.09.2021)

<sup>170</sup> Красовская Н. Р., Гуляев А. А., Юлина Г. Н. Фейковые новости как феномен современности // Власть. 2019. URL: <https://cyberleninka.ru/article/n/fejkovyie-novosti-kak-fenomen-sovremennosti/viewer>

<sup>171</sup> Виды фейков. // Banana Watch. Три точки. дата публикации неизвестна. URL: [https://banana.watch/falsehood\\_of\\_fakes](https://banana.watch/falsehood_of_fakes)

<sup>172</sup> Афера Стивена Гласса: 16 лет спустя. The New Republic вспоминает самый громкий скандал в своей истории. // Медуза. 12.11.2014. <https://meduza.io/news/2014/11/12/afeta-stivena-glassa-16-let-spustya>

авторитетные немецкие СМИ, а за некоторые из них журналист удостоивался журналистских премии<sup>173</sup>.

**Новости и публикации, искажающие суть реального события.** В эту категорию входят фразы, цитаты, выдернутые из контекста или изложенные не полностью, а выборочно. Такие факты встречаются в кликбейтных заголовках: то есть заголовок частично или полностью не соответствует содержанию материала.

**Предвзятая новость. Неточные данные или неверное использование данных.** Профессор университета Аризоны Стив Доиг в интервью исследовательскому центру American Press Institute приводит историю, когда одно из американских изданий (Доиг не раскрыл названия издания) опубликовало материал с данными про детскую смертность в своем городе<sup>174</sup>.

В дата-материале говорилось, что в одном из неблагополучных районов этого наблюдается самый высокий показатель смертности среди детей в городе. Однако после публикации этого материала выяснилось, что журналисты ошиблись в анализе данных: в районе, который заинтересовал их, располагается крупный учебный госпиталь, где лечили больных детей со всего штата.

Сегодня фейковые новости создаются не только из-за ошибок журналиста, но и с целью продвижения определенных интересов<sup>175</sup>. К таковым (если речь идет о намеренном распространении фейков, то есть дезинформации) можно отнести:

- Нанесение удара по политическим или экономическим противникам (например, в преддверии выборов);
- Манипулирование общественным мнением и формирование определенной картины мира;
- Привлечение аудитории на сайт;
- Коммерческое мошенничество.

## 8.6. Борьба с фейками: от медиаграмотности до лишения свободы

Информационная угроза, исходящая от фейков, потребовала согласованной политики, в которую включились практически все интернет-гиганты вроде Фейсбук, Google, Майкрософт, Твиттер и YouTube. Например, Фейсбук (сейчас - компания Метэ) опубликовал отчет о первом квартале 2020 года, где сообщается: в этот период компания удалила 1,7 млрд фейковых аккаунтов.

Сколько из них «специализировалось» на коронавирусной тематике, неизвестно. Но есть статистика обнаруженных фейковых сообщений: если в марте в соцсети Марка Цукерберга было зафиксировано 40 млн постов с недостоверным содержанием, то в апреле их число возросло до 50 млн.

• Ответом на распространение фейков про коронавирус в Сети стало принятие (или ужесточение) отдельными странами запретительных законов. Законы, предусматривающие наказание вплоть до лишения свободы, были приняты и действуют в Азербайджане, Румынии,

---

<sup>173</sup> Лучший репортер Германии годами фабриковал свои статьи для Der Spiegel. // Esquire. 21.12.2018.

Точный адрес: <https://esquire.ru/articles/75842-luchshiy-reporter-germanii-godami-fabrikoval-svoi-stati-dlya-der-spiegel/#part0>

<sup>174</sup> Samantha Sunne. "The challenges and possible pitfalls of data journalism, and how you can you avoid them" AmericanPress Institute. n.p., 9.03.2016. Web. 20.09.2021.

URL: <https://www.americanpressinstitute.org/publications/reports/strategy-studies/challenges-data-journalism/>

<sup>175</sup> Jamie Wiseman. "Rush to pass 'fake news' laws during Covid-19 intensifying global media freedom challenges" International Press Institute. n.p., 22.10.2020. Web. 20.09.2021. <https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/>

Сингапуре, Боливии, Венгрии, Вьетнаме. Ответственность за публикацию и рассылку фейковой информации в свои Административный и Уголовный кодексы в 2020 году ввели Узбекистан и Таджикистан<sup>176</sup>.

- Критики подобных мер (среди них есть международная правозащитная организация Human Rights Watch) считают, что санкционные законы создают возможности для правительств в подавлении инакомыслия и преследовании критиков власти<sup>177</sup>.

С другой стороны, проблема фейков объединила усилия неправительственных организации, профессиональных журналистов и IT-корпораций. Логичным шагом в противодействии недостоверному контенту стало развитие фактчекинговых сервисов и инициатив<sup>178</sup>.

В 2015 году Институт Пойнтера, занимающийся исследованиями журналистики, создал Международную сеть фактчекеров (International Fact-Checking Network), которая отслеживает форматы и тенденции развития фейк-ньюс, занимается обучением.

Китайская платформа Вейбо (Weibo) в 2018 году предоставила возможность для «заслуживающим доверия СМИ» и правительству Китая помечать фейковые сообщения на платформе специальным уведомлением. Какая информация является фейком, а какая — нет, решают госорганы Китая и аккредитованные правительством СМИ<sup>179</sup>.

Фейсбук, с другой стороны, в своем Пользовательском соглашении (раздел Нормы сообщества) обещает своим пользователям не удалять фейковую информацию, однако ограничивать ее распространение<sup>180</sup>. Однако в 2021 году корпорация удалила сеть аккаунтов, распространяющих фейки о вакцинах AstraZeneca и Pfizer<sup>181</sup>. Фейсбук также удалил пост президента Кыргызстана Садыра Жапарова о том, что коронавирус лечится настойкой ядовитого аконита<sup>182</sup>.

В 2020 году российская блог-платформа Яндекс.Дзен запустила пилотный проект проверки фактов в партнерстве с НИУ ВШЭ, а также журналистами из ТАСС, Интерфакс, Ведомости и The Bell. Фактчекерам на проверку отправляют: публикации, которые быстро набирают популярность на платформе; публикации, которые получают много жалоб пользователей на недостоверную информацию<sup>183</sup>.

В сентябре 2018 году российский телеканал «Дождь» запустил передачу Fake News, в которой развенчивает фейки и пропаганду на федеральных каналах.

---

<sup>176</sup> Jamie Wiseman. "Rush to pass 'fake news' laws during Covid-19 intensifying global media freedom challenges" International Press Institute. n.p., 22.10.2020. Web. 20.09.2021. <https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/>

<sup>177</sup> Unknown. "Singapore: 'Fake News' Law Curtails Speech" Human Rights Watch. n.p., 13.01.2021. Web. 20.09.2021. URL: <https://www.hrw.org/news/2021/01/13/singapore-fake-news-law-curtails-speech>

<sup>178</sup> Фейсбук. Фальшивые новости // Нормы сообщества Facebook. дата публикации неизвестна. URL: <https://transparency.fb.com/ru-ru/policies/community-standards/false-news/>

<sup>179</sup> Качкаева А.Г., Шомова С.А., Душакова И. С. Фейк: от академических дискуссий к практическим решениям / Коммуникации. Медиа. Дизайн, Том 6, №2, 2021

<sup>180</sup> Фейсбук. Фальшивые новости // Нормы сообщества Facebook. дата публикации неизвестна. URL: <https://transparency.fb.com/ru-ru/policies/community-standards/false-news/>

<sup>181</sup> Facebook удалил сеть аккаунтов из России, распространяющих фейки о западных вакцинах // Би-Би-Си. 11.08.2021. URL: <https://www.bbc.com/russian/news-58171489>

<sup>182</sup> Facebook удалил посты Жапарова о лечении COVID-19 настойкой ядовитого корня. В аппарате президента заявили, что они сделали это сами // Настоящее  
Время. 19.04.2021. URL: <https://www.currenttime.tv/a/kyrgyzstan-akonit-zhaparov/31211747.html>

<sup>183</sup> Назаров А. «Фейковые» новости в современном мире - оружие массового психоза? // Центр-1. 17.01.2018. URL: <https://centre1.com/kyrgyzstan/fejkovye-novosti-v-sovremennom-mire-oruzhie-massovogo-psihoza/>

## 8.7. Борьба с фейками в Кыргызстане

Кыргызстан, как бывшая часть СССР, зачастую попадает под влияние российской пропаганды – центра и наследницы Советского Союза. Нынешняя негласная стратегия борьбы против дезинформации в стране заключается в раскрытии недостоверной информации и обучению журналистов, а данная работа финансируется с поддержки международных доноров и неправительственных организаций.

К примеру, в 2020 году при поддержке программы Интерньюс в Кыргызстане журналисты изданий [Factcheck.kg](http://Factcheck.kg), «ПолитКлиника» и [T-Media.kg](http://T-Media.kg) объединились в фактчекинг-группу, которая занималась изучением и разоблачением фейков про коронавирус. В рамках проекта фактчекеры опубликовали 115 статей и видеоматериалов на кыргызском и русском языках. Их контент состоит не только из классического фактчекинга – верификации информации на достоверность, но также включает жанры медиакритики и журналистских расследований.

### **Комментарии фактчекеров-участников программы Интерньюс**

Главный редактор [Factcheck.kg](http://Factcheck.kg) Болот Темиров:

— Паника и страх – расчет делается именно на эти чувства читателей, когда придумываются фейки. Особенно «востребованной» оказалась тема коронавируса. Проснулись и сторонники различных теорий заговора. Активизировались создатели геополитических фейков. В условиях пандемии количество и скорость распространения фейковых новостей и слухов многократно увеличились.

Но вместе с этим повысилась и скорость наших разоблачений. Некоторые из них моментально попадали в обсуждение на различных форумах и активно комментировались в социальных сетях, не только в Кыргызстане, но и в других государствах. Посещения нашего сайта через поисковые запросы увеличились в разы. Это говорит об актуальности фактчека в мире.

### **Главный редактор издания «ПолитКлиника» Дильбар Алимова:**

— До пандемии мы лишь изредка обращались к жанру фактчекинга, да и выбирали легкие инфоповоды вроде баек о наличии в бананах ВИЧ. С такими объемами и сложностью тематики столкнулись впервые и очень рады, что приняли решение помочь коллегам из [Factcheck.kg](http://Factcheck.kg) справиться с возросшим из-за коронавируса потоком фейковых новостей.

Это оказалось очень полезным для нас. Мы и раньше занимались журналистскими расследованиями, а фактчекинг является очень важной частью этого направления. А теперь наши расследования станут более качественными. Но проверкой информации как таковой мы будем заниматься отдельно, на постоянной основе, для чего планируем создать в редакции отдел фактчекинга. За это время мы увидели, что такие проверочные материалы всегда находятся в топе самых читаемых статей нашего сайта.

### **Выводы**

Президентские выборы в США в 2016 году и пандемия коронавируса в 2020 году показали, насколько деструктивной и серьезной проблемой является фейковая информация. Фейки и дезинформация опасны тем, что из-за них принимаются неправильные решения, происходит поляризация обществ, рост недоверия к авторитетному мнению. В крайних случаях фейковая информация может стать причиной гибели людей.

В современных реалиях фейки распространяются с помощью социальных медиа, поэтому логичным шагом IT-корпораций стало внедрение механизмов по борьбе и предупреждению фейков. Методы против фейков могут быть разными: от ограничения в распространении до

блокировки аккаунтов до привлечения экспертов-фактчекеров для проверки сомнительной информации или данных.

Фейки про коронавирус стали поводом для введения запретительных законов в ряде государств мира, некоторые из которых предусматривают жесткие санкции вплоть до лишения свободы или огромных штрафов. Подобные законы, по сути, развязывают руки авторитарным правительствам, которые под предлогом защиты от фейковой информации могут использовать данные законы для подавления критики и преследования оппозиции.

В то же время проблема фейков стала толчком для развития фактчекинга и развития профессиональных сообществ. Это не стало исключением и для Кыргызстана и развития таких проектов как Factcheck.kg, «Политклиника» и так далее. Финансирование инициатив по фактчекингу в Кыргызстане на 2021 год пока идет за счет средств международных доноров и неправительственных организаций.

#### **Задания для самостоятельной работы:**

1. Перечислите фейки про коронавирус и вакцинацию, с которыми вы сталкивались в 2020 и 2021 годах. От кого вы получили данную информацию? Поверили ли вы в данный фейк?

2. Изучите материалы, посвященные выводу американских войск из Афганистана. Затем посмотрите отрывок из выпуска российского блогера и урбаниста Ильи Варламова на YouTube по этой теме (выпуск Чё Происходит #80, начало с 18:26). Какую ошибку допустил Варламов?

3. Посмотрите один из выпусков передачи Fake News на YouTube на выбор. Перечислите методы и приемы, используемые журналистами телеканала «Дождь» при разоблачении фейков и пропаганды.

4. 1 марта 2021 года российский дизайнер Артемий Лебедев выпустил фейк про немецкое гражданство Юлии Навальной, жены оппозиционера и основателя Фонда по Борьбе с Коррупцией Алексея Навального. К какому виду / типу фейка относится данная информация?

5. Найдите 2-3 примера фейковой информации, созданной в развлекательных целях

6. Найдите 2-3 примера фейковой информации, созданной для привлечения аудитории на сайт.

7. Найдите 2-3 примера фейковой информации, созданной в целях манипулирования общественным мнением и формировании определенной картины мира

8. Посмотрите передачу РЕН.ТВ на выбор, посвященную теориям заговора. Определите приемы и техники, используемые тележурналистами для привлечения внимания зрителей

#### **Вопросы:**

1. Какие возможности открылись для журналистики с развитием интернет-технологий?

2. Назовите причины, по которым фейки пользуются популярностью.

3. В чем разница между некорректной информацией, дезинформацией и сатирой?

4. Почему, по мнению Всемирной организации здравоохранения, фейки представляют серьезную опасность для общества?

5. Перечислите меры, принимаемые государствами, IT-корпорациями, неправительственными организациями в борьбе с фейками в интернете.

6. Перечислите цели, которые преследует намеренное распространение дезинформации.

7. Что такое фактчекинг?

8. Назовите базовую рекомендацию, которая эффективна в борьбе с фейками.

## 8.8. Проверка социальных сетей: оценка источников и визуального контента

### Вопросы:

- Верификация источников в соцсетях
- Верификация фотографий
- Верификация видеоматериалов

### Верификация источников в соцсетях

Проверка фактов — один из важнейших компонентов работы медиаспециалиста. Представляем подборку простых инструментов, с помощью которых каждый может провести факт-чекинг той или иной информации. В первой части материала — читатель найдет онлайн-сервисы для работы с социальными сетями. Какие-то ресурсы вам наверняка известны, что-то из рекомендаций окажется не подходящим для конкретной ситуации. Однако, чем больше инструментов для проверки данных вы будете использовать, тем более достоверным будет окончательный вариант журналистского материала.

Для проверки аккаунтов в соцсетях специалисты применяют ряд инструментов, таких как: Facebook Graph Search;

- LinkedIn;
- Twitter Advanced Search;
- Topsy — поиск пользователей, твитов, видео, фото, ссылок в Twitter;
- Webmii — отслеживает онлайн-данные о пользователях — опубликованные на разных ресурсах фотографии, Википедия, новости (Yahoo), профили в социальных сетях.

Для того, чтобы получить информацию о пользователе социальной сети, нужно начать с поиска в социальных сетях. Это логично, однако многие игнорируют этот шаг. Логично воспользоваться специальными поисковыми настройками твиттера, а также попробовать возможности крупнейшей профессиональной сети LinkedIn.

Например, для того, чтобы узнать, является ли твиттер-аккаунт настоящим или нет, применяется сервис Bot or Not. Он анализирует данные указанного твиттер-аккаунта — количество твитов, ретвитов, друзей и так далее. По этим данным вы сможете составить картину того, является ли пользователь реальным или это — программа.

Контактная информация. Эти сервисы, большей частью, предназначены для поиска информации о пользователях на английском языке. Однако функционал инструментов разный, и даже немногие полученные данные могут оказаться полезными. Facebook Graph Search, LinkedIn, Twitter Advanced Search, Pipl.com — поиск по имени, электронному адресу, телефонному номеру.

Spokeo.com — аналогичный предыдущему сервис.

SearchSystems.net — глобальный сервис, где выложено множество публичных данных, с функцией поиска людей.

Rapportive — плагин для Gmail-почты. Покажет вам полную информацию о ваших контактах, включая местонахождение, фото, аккаунты в социальных сетях.

AnyWho — поиск людей по имени, адресу, телефонному номеру.

GeoSocial Footprint — поиск информации по твиттер-аккаунту (местонахождение, информация профиля). Работает не для всех аккаунтов, к сожалению.

Numberway — каталог справочников телефонных номеров, находящихся в открытом доступе, в той или иной стране.

### **Фактор «синей галочки»**

Это весьма полезно знать журналистам. Конечно, значок подлинности не самое главное, но все же – одно из социальных доказательств. Галочка рядом с названием аккаунта или сообщества однозначно повышает доверие пользователей. Если речь о селебрити или популярном блогере, то наличие синего значка – это также отстройка от многочисленных фейковых профилей.

Синий значок в Instagram показывает, что профиль действительно принадлежит публичной личности, знаменитости или международному бренду.

Ниже мы рассмотрим некоторые примеры требований соцсетей для получения такой синей галочки (Instagram, VK, FB, Twitter, OK, YouTube, TikTok и Telegram). Это даст определенное понимание надежности данного параметра оценки источников в соцмедиа.

### **Instagram**

Требования к аккаунту:

Представляет реального человека или зарегистрированный бизнес. Есть фотография, текст в био и хотя бы один пост. В профиле нет ссылок, призывающих добавиться в других соцсетях. Это уникальный, то есть единственный аккаунт, который представляет человека или компанию. Известность личности или бренда подтверждается популярностью среди аудитории, несколькими независимыми источниками. Рекламный контент – не в счет.

Автоматически галочку получить нельзя, сначала нужно отправить запрос на верификацию. Для этого нужно указать свое реальное имя и приложить необходимые документы.

### **«ВКонтакте»**

До ноября 2019 VK выдавал галочки только страницам известных лиц или организаций. Из-за этого верификацию воспринимали «в качестве медали, которую дают за особые достижения», сообщает команда «ВКонтакте». Теперь процедуру упростили. Шансы есть и у малоизвестных музыкантов, и у локального бизнеса.

Никаких особых требований нет:

- Контент на странице уникальный, регулярно обновляется.
- Не менее 100 подписчиков, а страница создана больше 3 месяцев назад.
- Есть фото и описание, страница публичная / группа открытая.
- Настроена двухфакторная аутентификация.
- Заявка на верификацию раньше подавалась через службу поддержки, теперь для этого есть соответствующие формы в настройках личного аккаунта / сообщества. Надо просто рассказать о себе и своей деятельности, а также почему вы нуждаетесь в галочке. Плюс необходимо приложить документы, удостоверяющие личность / подтверждающие принадлежность к компании. Также следует добавить ссылки на упоминания в СМИ, верифицированные аккаунты в других социальных медиа.

Как и Instagram, «ВКонтакте» ничего не обещает и оставляет за собой право отказать в прохождении верификации. Но через некоторое время можно подать заявку снова.

### **Facebook**

В Facebook тоже есть синие галочки, они также показывают подлинность страниц и профилей. Пройти верификацию имеют право только публичные личности, СМИ и бренды.

Facebook предъявляет следующие требования к аккаунту:

- Есть фото обложки и профиля.
- Название страницы – без слоганов, ошибок, общих слов.
- Есть опубликованные материалы.

- Включена кнопка «Подписаться» (у профилей).
- Чтобы отправить запрос на верификацию профиля или страницы, нужно перейти на специальную форму. Затем приложить документы, удостоверяющие личность, дать ссылки, иллюстрирующие общественный интерес, а также объяснить, зачем вообще вам нужно получать галочку Facebook.

Дополнительное условие для представителей правоохранительных органов – включение двухфакторной аутентификации. Отправлять сразу несколько запросов бесполезно. Если заявку отклонят, можно запросить верификацию повторно через 30 дней.

### **Twitter**

Синяя галочка в Twitter означает, что аккаунт представляет большой интерес для общества, а его подлинность подтверждена. Сам владелец верифицированного профиля получает доступ к некоторым дополнительным настройкам: например, может отказаться от групповых личных сообщений. Пока что временно возможность верификации аккаунта Twitter приостановлена.

Однако помимо «синей галочки» существует много других факторов, с которыми сталкивается журналист.

### **YouTube**

Условия прохождения верификации:

- Канал должен принадлежать автору, исполнителю, общественному деятелю или бренду.
- Наличие 100 тыс. подписчиков.
- Профиль – общедоступный, активный, заполненный.
- Есть еще дополнительные факторы – например, признание за пределами YouTube, известность в интернете. Также администрация платформы может запросить больше информации или документацию.

### **TikTok**

Строгих требований к верификации нет. Нужны подлинность, уникальность, активность и отсутствие нарушений условий и принципов сообщества. Однако страница центра поддержки, посвященная процедуре верификации, возвращает ошибку 404. Судя по всему, никаких заявок соцсеть не принимает, и нужно ждать благословления от сотрудников / алгоритмов. Скорее всего, как и везде, играет роль популярность самого автора и контента, который он делает.

### **Telegram**

«Телеграм» верифицирует только каналы / ботов крупных компаний («Яндекс», «Альфа Банк») и публичных людей (Павел Дуров, Эммануэль Макрон), которые всем известны, имеют галочки в других соцмедиа. Также синие значки есть у официальных каналов и ботов самого Telegram. Для широких масс верификации у Telegram нет, и писать по этому поводу в саппорт не имеет смысла. Но на продвижение в Telegram наличие галочки гарантированно никак не влияет.

## **8.9. Как искать информацию о пользователях в социальных сетях?**

Сегодня социальные сети служат журналисту не только источником инфоповодов, но и позволяют проверить данные и найти дополнительные сведения, чтобы установить связи, сопоставить данные и пр. Далее рассмотрим инструменты, которые помогают журналисту найти информацию в различных соцмедиа.

### **«ВКонтакте»**

220vk – онлайн-сервис для наблюдения за активностью пользователя во ВКонтакте: приложение показывает исходящие лайки, скрытых друзей, проведенное в сети время и т.д.

Сервис VK Paranoid Tools позволяет посмотреть, кому пользователь ставит лайки, какие темы комментирует.

SnRadar помогает искать фотографии пользователей «ВКонтакте» по геолокации. Установив на карте точку, определив радиус и задав время, можно узнать, кто в зоне поиска выкладывал фотографии в открытый доступ. Например, с места ЧП.

### **Facebook**

Несколько лет назад Facebook ограничил возможность широкого поиска по социальной сети, например, лайков и комментариев. Защита данных была усилена в том числе после скандала с участием компании Cambridge Analytica, которая использовала алгоритмы соцсети для политической рекламы.

Тем не менее внутренний поиск соцсети позволяет находить некоторые данные.

Поиск человека: указать имя и фамилию человека. Применить разные фильтры соцсети: город и страна, общие друзья, образование, место работы.

Поиск по номеру телефона: пользователи Facebook могут указывать номер телефона в открытом доступе. Для поиска достаточно указать номер в строке поиска. По нему можно найти человека или установить его связь с какой-нибудь организацией.

Искать посты, фотографии, фильтровать выдачу, по ключевым словам, и датам позволяет сервис Graph.tips.

Инструмент Lookup-ID помогает установить цифровой ID пользователя, если он заменен словами.

Сервис Social Searcher позволяет искать пользователей и посты с упоминанием человека, компании или продукта сразу в нескольких сетях: Facebook, Instagram, LinkedIn, Vimeo и других. Например, если вбить в строку поиска «Иван Иванов», то выборка покажет последние посты всех зарегистрированных Иванов Ивановых сразу в нескольких соцсетях, а также их упоминания другими пользователями.

### **Instagram**

Одним из самых удобных инструментов поиска в Instagram является внутренний поиск: по хештегу и геолокации. В приложении искать удобнее, чем в версии соцсети для ПК, однако эта проблема решается с помощью сторонних сервисов, например: Gramfind.

Соцсеть также позволяет находить аккаунт по номеру телефона при условии, что номер «привязан» к странице или указан в ее контактах. Для поиска нужно зайти в личный профиль, нажать на символ в верхнем правом углу, выбрать раздел «Интересные люди», а затем – «Контакты». Во вкладке появится список всех пользователей, которые так или иначе связаны с телефонной книгой.

Расширение для браузера Google Chrome Helper Tools for Instagram v2 помогает выявлять связи между пользователями Instagram: например, установить, с кем человек взаимодействовал в соцсети. Как пользоваться расширением, можно узнать в видеоинструкции на YouTube.

### **YouTube**

При загрузке видео YouTube позволяет проверить, был ли ролик уже опубликован и кем.

Сервис YouTube GitHub помогает искать видео по локации, периоду, ключевым словам, заданному радиусу. В поиске необходимо указать несколько параметров: город (или другой

населенный пункт), временной интервал, радиус (например, 100 км), ключевое слово (если оно есть).

Результаты поиска сортируются по дате публикации.

Поиск по геометке сработает в том случае, если автор указал в ролике геометку. Сортировка, по ключевым словам, понадобится, чтобы сузить результаты выдачи.

## **8.10. Верификация фотографий**

### **Как проверить подлинность снимка или документа?**

Самый простой способ проверить фотографию – загрузить ее в поисковики Google Image или Яндекс-картинки, чтобы узнать, где она ранее публиковалась.

Проследить путь картинки в Интернете также позволяет сервис TinEye.

Удаленные с сайтов и из соцсетей снимки можно найти в веб-архиве Cached Pages.

Сервис FotoForensics показывает, было ли изображение изменено в фотошопе.

Внутренняя информация о файле или метаданные позволяют уточнить происхождение фотографии или документа: когда и где был создан документ или сделан снимок, на какое устройство, кем и т.п. Исследовать метаданные можно с помощью сервиса Jeffrey's Image Metadata Viewer.

На что стоит обратить внимание при проверке снимка?

Авторство снимка, место съемки и аппаратура: изучить метаданные, загрузить снимок в поисковики и узнать, публиковался ли он ранее. «Авторский стиль»: оценить, похожи ли снимки на серию, обратить внимание на свет, резкость деталей. Размер и качество фотографии: насторожиться, если снимки нестандартного размера или очень хорошего качества.

### **Как журналисту искать и применять данные из открытых источников?**

#### **Кейс «Технологии и ресурсы предвыборной кампании в Республике Коми»**

Использованные сервисы: поиск в соцсетях, метаданные документа.

Редакция получила флешку с документами, в том числе концепцией избирательной кампании кандидата в Госсовет Республики Коми. В документе подробно описывались мероприятия для избирателей: чаепития в детсадах и школах, встречи кандидата с директорами и коллективами учреждений, тезисы выступлений и бюджет.

Чтобы проверить подлинность информации и понять, действительно ли в предвыборной кампании использовался административный ресурс, редакция провела расследование. Журналисты нашли фотографии с чаепитий и встреч в соцсетях. Метаданные документа показали, что концепция создавалась в администрации региона.

Теперь рассмотрим способы и инструменты для поиска данных в некоторых наиболее популярных социальных сетях.

#### **Facebook**

- Искать посты, фотографии, фильтровать выдачу по ключевым словам и датам позволяет сервис Graph.tips
- Инструмент Lookup ID помогает установить цифровой ID пользователя, если он заменен словами.
- Сервис Social Searcher позволяет искать пользователей и посты с упоминанием человека, компании или продукта сразу в нескольких сетях: Facebook, Instagram, LinkedIn, Vimeo и других.

Например, если вбить в строку поиска «Иван Иванов», то выборка покажет последние посты всех зарегистрированных Иванов Ивановых сразу в нескольких соцсетях, а также их упоминания другими пользователями.

### **Instagram**

Одним из самых удобных инструментов поиска в Instagram является внутренний поиск: по хештегу и геолокации. В приложении искать удобнее, чем в версии соцсети для ПК, однако эта проблема решается с помощью сторонних сервисов, например: Gramfind.

Соцсеть также позволяет находить аккаунт по номеру телефона при условии, что номер «привязан» к странице или указан в ее контактах. Для поиска нужно зайти в личный профиль, нажать на символ в верхнем правом углу, выбрать раздел «Интересные люди», а затем — «Контакты». Во вкладке появится список всех пользователей, которые так или иначе связаны с телефонной книгой.

Расширение для браузера Google Chrome Helper Tools for Instagram v2 помогает выявлять связи между пользователями Instagram: например, установить, с кем человек взаимодействовал в соцсети.

### **YouTube**

При загрузке видео YouTube позволяет проверить, был ли ролик уже опубликован и кем.

Сервис YouTube GitHub помогает искать видео по локации, периоду, ключевым словам, заданному радиусу. В поиске необходимо указать несколько параметров: город (или другой населенный пункт), временной интервал, радиус (например, 100 км), ключевое слово (если оно есть). Результаты поиска сортируются по дате публикации. Поиск по геометке сработает в том случае, если автор указал в ролике геометку. Сортировка по ключевым словам понадобится, чтобы сузить результаты выдачи.

## **8.11. Верификация видеоматериалов**

В работе журналиста часто возникает потребность проверки видео на его действительное отношение к тому или иному факту. Сегодня источников видеоматериалов в соцсетях огромное количество, и существует значительная вероятность того, что информация не соответствует реальности.

К примеру, вам необходимо проверить видео из Facebook, источник которого утверждает, что это съемка урагана в Уругвае. Вы, разумеется, не можете слепо брать данное утверждение на веру и вам лучше проверить что же на самом деле запечатлено на видео. Ваши действия будут следующие:

1. Сделайте скриншот этого видеоролика - <https://www.ntv.ru/novosti/1622222/> (см. рис. 1);
2. Откройте Google Images;
3. Загрузите скриншот в Google Images;
4. Посмотрите результат поисковой выдачи: там могут быть различные источники с историями об урагане в Уругвае, использующие фото, совпадающее с вашим скриншотом (либо же нет).

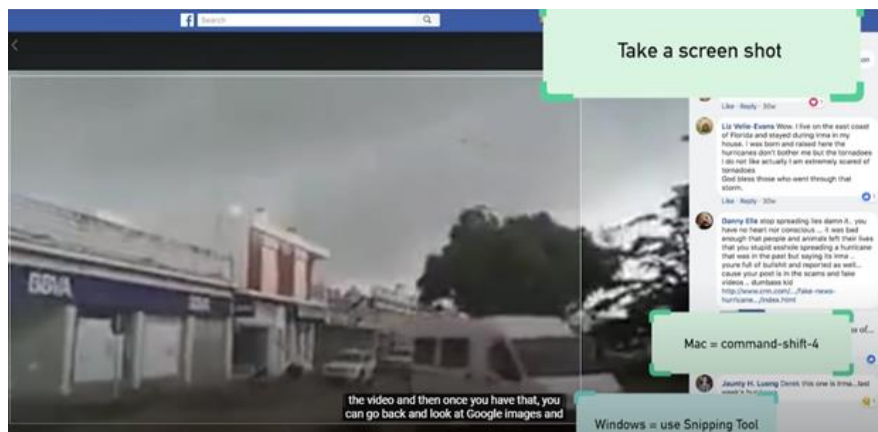


Рисунок № 58. История об урагане в Уругвае

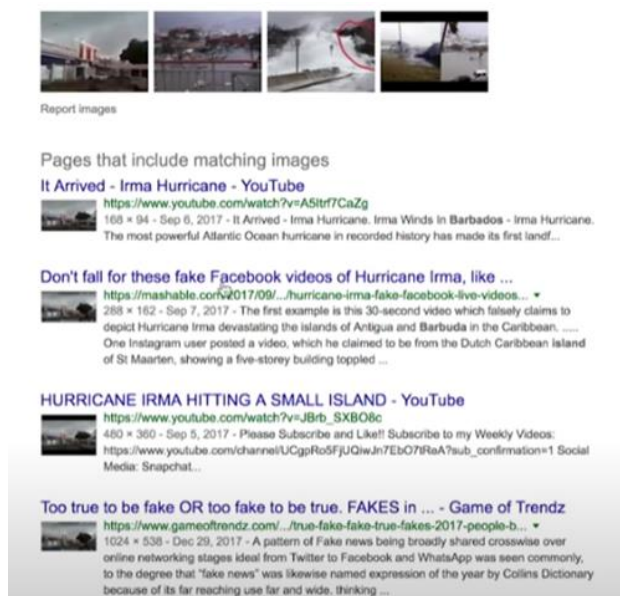


Рисунок № 59. Результаты поиска в Google Images

### Выводы:

1. Верифицировать информацию в соцсетях необходимо по причине обилия сомнительных источников.
2. Верификация фотографий, видео и аккаунтов осуществляется весьма просто.
3. Социальные сети заинтересованы в повышении уровня качества и достоверности контента, хотя это и сложно в условиях огромного количества аккаунтов, владельцы которых сами становятся жертвой низкой медиаграмотности.
4. В соцсетях вирусность контента нередко является поводом для подозрений в сомнительности контента и источников.

### Вопросы:

1. Каков общий принцип верификации информации в соцсетях?
2. Перечислите основные сервисы, применяемые для верификации источников в соцсетях.

3. На что стоит обратить внимание при проверке снимка?
4. Приведите алгоритм верификации видеоматериалов.
5. Какой сервис загрузке видео в YouTube позволяет проверить, был ли ролик уже опубликован и кем?
6. Позволяет ли Instagram находить аккаунт по номеру телефона при условии, что номер «привязан» к странице или указан в ее контактах?
7. С какой целью социальные сети дают «синюю галочку» некоторым аккаунтам?
8. Каковы условия прохождения верификации канала в YouTube?
9. Каким сервисом вы бы воспользовались для проверки наличия PhotoShop-эффектов в фотографии?
10. Что дает внутренняя информация о файле или метаданные фотографии или документа или как можно ее получить?
11. Какие полезные функции обеспечивает сервис Social Searcher?

**Задания:**

1. Проверьте любую новость на достоверность, используя описанной в данной статье метод верификации фотографий.
2. Проверьте любую фотографию на наличие PhotoShop-манипуляций, используя сервис FotoForensics.
3. Найдите посты в соцсетях с упоминанием любого человека на ваш выбор, используя сервис Social Searcher.

**Список источников:**

1. Журналистика данных: инновация или пустая трата времени и сил? <https://test.ru/2018/05/30/data-journalism-innovation-or-a-waste-of-time/>
2. Чурбакова Е. А. Журналистика баз данных как медиатренд // Медиа-исследования, 2017. № 4–2. С. 217–221.
3. Шевченко В. Э. Визуальный контент как тенденция современной журналистики // Медиаскоп. 2014. №4. С. 6-8.
4. Иванов А. Д. Роботизированная журналистика как следующий этап развития дата-журналистики // Журналистика в системе альтернативных источников информации: сб. материалов науч. конф, Н. Новгород, 2017. С. 113–119.
5. Методы сбора информации: в журналистике и репутинге <https://reputation.moscow/journal/metody-sbora-informatsii/>

## Модуль 9. МЕТОДИКА ПОДГОТОВКИ ДАТА-ПУБЛИКАЦИИ ДЛЯ МАССМЕДИА

В предыдущих модулях мы останавливались на том, как находить данные и как их правильно обрабатывать. И последним шагом является правильная презентация ваших данных. Существует множество различных способов написания статей в журналистике данных – от публикации необработанных наборов данных с сопровождающей историей до создания красивых визуализаций и интерактивных веб-приложений.

Журналисты данных могут помочь сопоставить, синтезировать и представить разнообразные и часто сложные источники информации таким образом, чтобы дать их аудитории реальное представление о сложных проблемах. Они могут дать читателям ясную и понятную перспективу с помощью интерактивной графики и прямых ссылок к первоисточникам.

### 9.1. SMART -Планирование материала

Первый шаг в написании любой статьи – это безусловно планирование её главного фокуса, цели и правильный выбор презентации. На практике журналисты могут применить SMART целеполагание, которое широко распространено в менеджменте, проектном управлении и даже при создании контент маркетинга. Но в отличие от этих сфер, значение этой аббревиатуры в журналистике данных несколько иное.



Рисунок № 60. SMART целеполагание

- **Specific** (конкретность) – определите какие конкретные задачи стоят перед вашей статьей. Будет ли это сугубо информационное сообщение? Статья на бытовую тему или материал развлекательного характера?

- **Measurable** (измеримость) – насколько правильно вы измерили свои данные и поддаются ли они простой и понятной презентации.

- **Attainable** (достижимость) – все ли данные, которые вы хотите представить в своей статье достижимы? Не выйдет ли ваша статья слишком однобокой, потому что вы показали только, те данные, которые смогли получить?

- **Relevant** (актуальность) – журналистика данных, это сфера, которая находится в постоянном развитии вместе со все более развивающимися технологиями. Убедитесь, что ваш способ подачи материала будь то инфографика или интерактивные графики не потеряли своей актуальности.

- **Time-bound** (ограничение во времени) – к сожалению, даже журналистика данных не свободна от редакционных дедлайнов. Какой бы вы способ презентации вы не выбрали, удостоверьтесь, что у вас хватает достаточно времени, чтобы обработать данные и правильно их

оформить. То, что материал можно превратить в сложное красивое новостное приложение, не означает, что вам нужно его создавать. Баланс приоритетов имеет решающее значение.

Каждый из компонентов в SMART планировании поможет вам правильно распределить свои ресурсы и время, а также прояснить главные задачи во время написания вашей статьи.

### **Ориентируйтесь на свою аудиторию**

Во время сбора больших данных и дальнейшей статистической обработки, легко упустить самую главную часть вашей работе – ваших читателей. Вам нужно твердое понимание своей аудитории. Кого ваша статья действительно заинтересует? Предназначена ли она для широкой аудитории или только для экспертов? Это всё - разные категории читателей, которым интересны неодинаково сформулированные и по-разному поданные темы, с разными примерами и формулировками. Без понимания своей аудитории вы рискуете создать слишком обобщенный контент.

Несмотря на свое серьезное название, журналистика данных не всегда обязана быть серьезной или перегруженной информацией. Наоборот, она может подавать сухие статистические цифры в простой и занимательной форме. Например, использование карикатур в качестве персонажей может сделать ее более занимательной. В газете “Время” в мае 2020 года было проиллюстрировано сколько продуктов в месяц съедает среднестатистический казахстанец (см. рис. ниже). Очевидным преимуществом стоит назвать то, что каждая иллюстрация сопровождается аналитическим материалом, что помогает лучше отразить содержание истории. Кроме обычной инфографики в газете «Время» присутствуют целые информационные блоки, сделанные в виде «комиксов». Автором большого количества материалов инфографики, некоторые из которых сделаны в мультяшном виде, является Владимир Кадырбаев.



Рисунок № 61. В. Кадырбаев. Комплексный обед<sup>184</sup>

Наконец, что очень характерно именно для журналистики данных, в проектах с очень сложными техническими решениями, вам будет необходима помощь программистов-разработчиков. В таких случаях особенно важно помнить об интересах ваших читателей. Даже лучший программист, предоставленный самому себе, может предпочесть технически интересную работу вместо того, чтобы создавать более легкий для восприятия контент. Также, как и журналистам, программистам нужны редакторы и креативные наставники.

## 9.2. Визуализация

Без приемов визуализации журналистика данных, несмотря на всю свою важность и высокую технологичность, не была бы настолько интересной для читателей. Перегруженный цифрами и остальными данными текст может быть трудным даже для экспертов, не говоря уже о широкой аудитории. Брайан Суда, один из авторов учебника «Пособие по журналистике данных», заключает,

<sup>184</sup> В. Кадырбаев. Комплексный обед.

<https://time.kz/articles/Infografika/2020/05/06/kompleksnyj-obod>

что «некоторые из самых известных графиков и диаграмм появились вследствие необходимости лучше объяснять малопонятные таблицы с данными».

Удачно сделанная визуализация не только сразу привлекает внимание читателей, она также обладает мощным когнитивным преимуществом: половина человеческого мозга посвящена обработке визуальной информации. Хорошо спроектированная визуализация данных может дать зрителям мгновенное и глубокое впечатление и упорядочить сложные статистические данные, чтобы сразу перейти к сути поднимаемой проблемы.

Но в отличие от других визуальных средств массовой информации, таких как фотография или видео, визуализация данных все же предоставляет больше информации и измеримых фактов. Не менее эстетически привлекательные, различного рода инфографики бывают менее эмоционально заряжены. Хорошим примером является инфографика на сайте Национального статистического комитета Кыргызской Республики «В Кыргызстане сокращается число ранних браков»<sup>185</sup>.

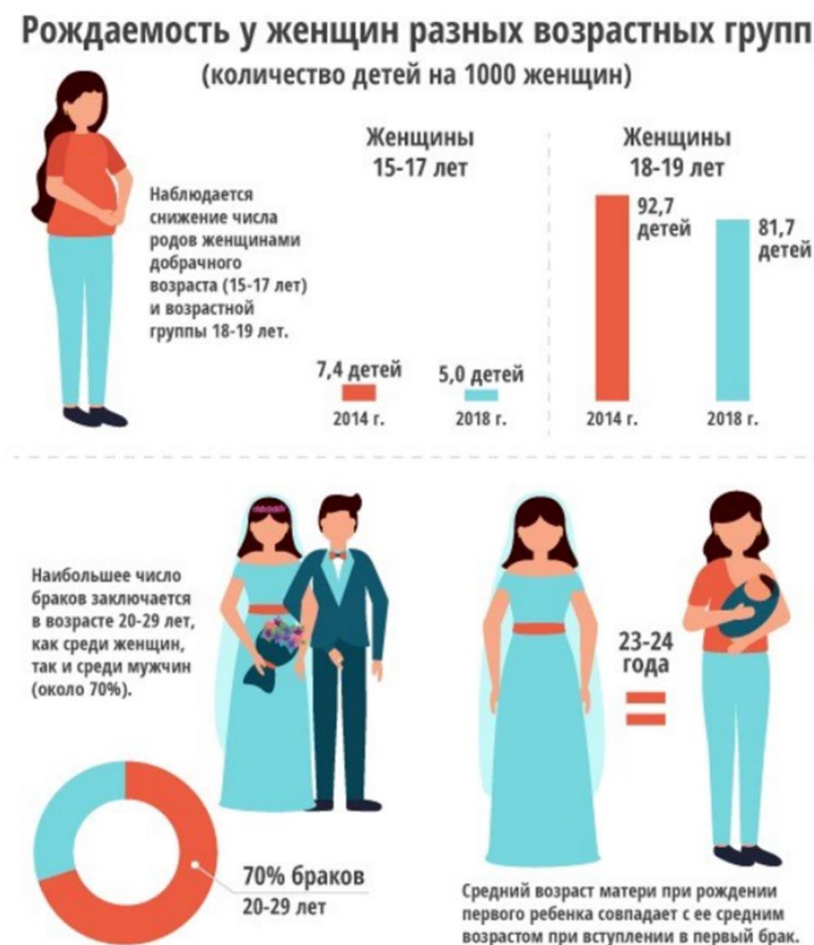


Рисунок №62. В Кыргызстане сокращается число ранних браков

<sup>185</sup> <http://www.stat.kg/ru/news/infografika-8-marta-2020-mezhdunarodnyj-zhenskij-den/>

### **Советы при создании инфографики**

Инфографика часто становится кульминацией вашей кропотливой работы по сбору данных. И ее создание может занять не меньше времени. Ниже представлены несколько советов перед началом работы<sup>186</sup>.

- **Семь раз отмерь, перед тем как начать работу**

В журналистике данных невозможно спрятать неполные или приблизительные данные за витиеватыми формулировками. Инфографика приглашает читателей к пристальному изучению, и поэтому вам будет тщательно необходимо перепроверять цифры, и оставлять ссылки на источники.

- **Имейте в виду два типа читателей**

Визуализация, будь то автономная интерактивная функция или статическая картинка, сопровождающая статью, должна удовлетворять два разных типа читателей. Она должна быть легкой для понимания при беглом просмотре, и в то же достаточно детальной и информативной, для тех, кто хочет проблему поглубже.

- **Не нужно перегружать инфографику**

Определитесь с главной идеей, которую вы хотите донести для своей аудитории, и сосредоточьтесь только на ней. Избыточные детали, не имеющие к проблеме прямого отношения, могут только запутать читателей. Если ваша основная цель не заключается в прозрачности отчетности, большинство данных, которые вы собрали, не имеют большого значения.

- **Текст в инфографике должен быть ясным и понятным**

Любая статья должна быть ясной по мыслям, по композиции и по языку. Тот же принцип относится и к визуализации данных. Если ваша инфографика не состоит из сплошных цифр и картинок, убедитесь, что в тексте нет незнакомых для читателей терминов и профессионального жаргона. Пояснения терминов особенно важны, если иллюстрируется сложный технический процесс.

- **Встречайтесь с графическим дизайнером как можно раньше и чаще**

Начинайте обсуждать будущую инфографику как можно раньше, чтобы знать какой тип данных для вас будет наиболее важен. Еще у дизайнеров, может быть, нестандартный и интересный подход к презентации ваших данных.

- **Разные графики для разных историй**

Визуальная журналистика стала неотъемлемой частью журналистики данных. И наряду с фотографиями, видеоматериалами и карикатурами, к ней относятся таблицы, диаграммы и интерактивные карты. И хотя есть много различных видов инфографики, все они состоят из базовых элементов: иконок, графиков, простых картинок, таблиц и заголовков.

## **9.3. Виды инфографики**

### **Хронологическая или историческая инфографика**

Возможно, наиболее распространенное использование визуализации данных — иллюстрирование того, как разные значения меняются с течением времени. Инфографика с временной шкалой как нельзя лучше подходит для того, чтобы показывать историю развития чего-либо и общий обзор событий, с выделением важных дат. Для этих целей как нельзя лучше подходят линейные графики. Направление линий на графике работает как хорошая метафора для данных:

---

<sup>186</sup> Суда Б. Дизайн для данных // Пособие по журналистике данных / Под ред. Д. Грея. URL: <https://unotices.com/book.php?id=125825>

наклон вверх указывает, где значения увеличились, а наклон вниз указывает, где значения уменьшились. Например, журналисты информационного портала halva.tj сделали интересную инфографику о населении Таджикистана и об изменении численности за последние годы<sup>187</sup>.

### Сопоставительная инфографика

Инфографика является самым наглядным инструментом для презентации сравнения двух и более дискретных значений. Как правило, это диаграмма, обобщающая основную информацию о сравниваемых объектах, которая призвана помочь другим людям лучше понять эти объекты и, как следствие, составить свое мнение о них.

Конечно, иногда и простая таблица может проиллюстрировать сравнение между значениями, однако использование разных цветов, иконок и пояснительных надписей, делает сухие цифры визуально более привлекательными. В сопроводительном тексте можно написать собственное заключение, чтобы подвести читателей к определённому выводу. Или ограничиться наглядной демонстрацией фактов и позволить аудитории сделать заключение самим. Например, в честь олимпийских игр в Токио в 2021 году редакция Repost совместно с 7SABER создали инфографику, в которой показывается, как сборная Узбекистана выступала на всех Олимпийских играх с 1996 по 2016 год<sup>188</sup>. На ней очень наглядно продемонстрированы достижения атлетов на разных олимпиадах, иллюстрация ниже показывает лишь часть инфографики. Инфографика: как выступала сборная Узбекистана на Олимпийских играх с 1996 года.

### Карты или географическая инфографика

Карты тоже могут быть использованы в журналистике данных для отражения географических, временных или количественных показателей. Карта может отображать различные затраты в отрасли по всей стране или миру, места, откуда поступают те или иные материалы, политические настроения в разных регионах и другую интересную информацию. В статье «Какую часть зарплаты отдают на коммуналку казахстанцы и наши соседи по СНГ» журналисты использовали карты наряду с таблицами, чтобы сравнить стоимость коммунальных услуг<sup>189</sup>.



Рисунок № 63. Карта количественных показателей

<sup>187</sup> [https://halva.tj/articles/faktcheking/tadzhikistan\\_v\\_chislakh\\_infografika\\_halva\\_tj/](https://halva.tj/articles/faktcheking/tadzhikistan_v_chislakh_infografika_halva_tj/)

<sup>188</sup> <https://repost.uz/tokyo2020/olympic-history>

<sup>189</sup> <https://informburo.kz/stati/kakuyu-cast-zarplaty-otdayut-na-kommunalku-kazaxstancy-i-nasi-sosedi-po-sng>

Одним из самых удобных инструментов для создания быстрых и детализированных карт является GoogleFusion. Этот ресурс помогает даже объединять несколько имеющихся карт в одну и добавлять функции поиска и фильтрации. Основным преимуществом является гибкость – вы можете загрузить, например, KML-файл с региональными границами, а затем объединить его с таблицей данных.

### Отображение иерархии

Иерархическую структуру данных обычно принято показывать через древовидную диаграмму как генеалогическое древо или выстраивать по принципу пирамиды, как пирамиду потребностей Маслоу. Такие диаграммы - очень простые и их можно нарисовать даже в программе PowerPoint.

Однако информацию можно проиллюстрировать и посредством древовидной карты (treemap). Несмотря на название, визуально она немного отличается от одноименной диаграммы. На ней количественным значениям разных категорий присваиваются разные размеры области квадрата. Площадь данного квадрата представляет собой количество, которое оно представляет, как само по себе, так и как совокупность его содержимого. Это компактный и интуитивный визуальный прием, чтобы показывать составные части данных и сравнения относительно друг друга. Древовидная карта была придумана в 1991 году Беном Шнайдерманом как метод презентации обширных каталогов файлов на компьютере так, чтобы они не занимали слишком много места на экране<sup>190</sup>. Например, на сайте Викисклад (Wikimedia Commons) при помощи древовидной карты показаны средние значения экспорта Туркменистана за 2017 год, на основе данных медиа-лаборатории Массачусетского технологического института и центра международного развития Гарвардского университета<sup>191</sup>.



Рисунок № 64. Древовидная карта экспорта Туркменистана в 2017 году

<sup>190</sup> [8] [https://datavizcatalogue.com/RU/metody/drevovidnaja\\_karta.html](https://datavizcatalogue.com/RU/metody/drevovidnaja_karta.html)

<sup>191</sup> [https://commons.wikimedia.org/wiki/File:Turkmenistan\\_Exports\\_Treemap\\_2017.svg](https://commons.wikimedia.org/wiki/File:Turkmenistan_Exports_Treemap_2017.svg)

### **Процессы и пошаговые инструкции**

Если в вашей статье вам нужно объяснить работу сложных систем или дать пошаговые инструкции, то инфографика - это самый наглядный способ продемонстрировать это. Для этого нужно разбить процесс на отдельные этапы, дать им понятное название и подобрать или нарисовать подходящие иконки для каждого раздела. Каждый этап можно снабдить коротким сопровождающим текстом. На сайте Академии государственного управления при Президенте Республики Казахстан с помощью простых пиктограмм объясняются курсы переподготовки<sup>192</sup>. А на сайте государственной программы «Цифровой Казахстан» объясняется, как покупать товары в интернет-магазинах, в виде круговой диаграммы с пояснительными надписями<sup>193</sup>.

### **Анимированная и интерактивная инфографика**

Преимуществом интернет-изданий является то, что визуализация необязательно должна быть представлена в виде статичных изображений. Хотя разработка может занять дополнительное время, анимированные элементы делают изображения более интересными и быстро привлекают внимание читателей. Однако следует помнить о том, чтобы движущиеся элементы не слишком отвлекали внимание от информационного содержания вашей инфографики.

Интерактивная инфографика еще более сложная в исполнении и требует высокой квалификации, но также позволяет использовать больше возможностей веб-дизайна и мультимедиа. Можно разработать инфографику таким образом, чтобы читатели могли сами управлять анимацией и переходить в разные информационные разделы.

**Бывает ли неоправданная визуализация?** Несмотря на то, что визуализация набирает все большую популярность, встречаются так же примеры неудачных работ. В конце концов, инфографика бывает хороша ровно настолько, насколько хороши данные, на которых она построена.

Еще один аспект неоправданной визуализации – это, когда графики и прочие данные бывают опубликованы без должного контекста. Разумеется, иногда хорошая картинка может стоить тысячи слов, но без описания ситуации и проблемы, она не несет для читателей ценности или рискует остаться непонятой. Как писал автор «Основных принципов визуализации данных» А. Кайро, «Создание привлекательной визуализации крайне важно, но стоит подумать о том, что будет после того, как вы привлечете внимание своих читателей? Журналист должен сделать визуализацию с полезными данными. И проблема, с которой я сталкиваюсь сегодня во многих визуализациях, заключается в том, что это просто красиво, но на самом деле не несет никакого смысла»<sup>194</sup>.

Ниже приведены несколько примеров, когда возможно стоит предпочесть другие методы презентации информации:

#### **Только голые факты**

Иногда голые факты и данные не могут сами по себе объяснить всю важность проблемы. Разумеется, диаграммы и свободная статистика могут быть полезными иллюстрирующими материалами, но непосредственное повествование о реальных последствиях проблемы и интервью с экспертами может быть более действенным для читателя.

#### **Данные не укладываются в четкую структуру**

Возможно, у вас нет большого разнообразия в данных, нет четкой тенденции или невозможно сделать четкие выводы из них. В таких случаях данные, внесенные в графики Excel или аналогичное приложение, могут оказаться слишком зашумленными со слишком большими

---

<sup>192</sup> <https://akm.apa.kz/infografik/>

<sup>193</sup> <https://ecampusontario.pressbooks.pub/communicationatwork/chapter/11-2-conflict-resolution-strategies/>

<sup>194</sup> Cairo, A. The main principles of data visualization

колебаниями, или же наоборот показывать плоский тренд. И вам стоит либо продолжать собирать данные или применить другой вид статистического анализа.

### **Выбранный метод не самый эффективный**

Иногда географические данные не являются значимыми в вашей подборке данных, в таких случаях не рекомендуется применять карты или другие географические графики. Они могут только отвлекать внимание читателей от других данных, таких как изменение во времени или демонстрация сходства между несмежными областями.

### **Иногда подходит и простая таблица**

Если у вас относительно мало точек данных, но есть информация, которая может быть полезна некоторым из ваших читателей, подумайте о том, чтобы просто представить контент в простой форме таблиц. Фактически, таблицы могут быть очень эффективным и элегантным макетом для основной информации. К тому же при их оформлении всегда можно использовать интересные цветовые решения и другие графические элементы<sup>195</sup>.

### **Материалы для самостоятельного изучения**

В основе журналистики данных всегда будет ваша статья, неважно представлена она в виде графики или написана простым текстом. Поэтому особенно важно продолжать развиваться как автор. Ниже представлены краткий список книг и других ресурсов для самостоятельного изучения.

#### **Книги:**

1. «Пиши, сокращай» Ильяхов Максим, Сарычева Людмила
2. «Бессистемные советы» Александр Амзин
3. «Как новые медиа изменили журналистику» Сергей Паранько, Наталья Лосева
4. «Копирайтинг. Как не съесть собаку. Создаем тексты, которые продают» Дмитрий Кот
5. «Пишите – не пишите» Евгения Пельтек
6. «On Writing Well: The Classic Guide to Writing Nonfiction» William Zinsser

#### **Статьи:**

1. «Как написать статью?» Лариса Павловна Нода <https://newreporter.org/2018/12/12/kak-napisat-statyu/>

2. «Как создать текст для печати» <http://www.oshibok-net.ru/for-journalists/journalistic-text/jur-text-pechat/>

«Data-журналистика. Пример того как использовать данные для репортажей» <https://xn----7sbafuabraerjymxvsmn8f.xn--p1ai/data-zhurnalistika-primer-togo-kak-ispolzovat-dannye-dlya-reportazhej/> Бесплатные онлайн курсы

1. Научные тексты — учимся писать на русском и английском <https://www.coursera.org/learn/naychnie-teksti>
2. Коммьюнити-журналистика: цифровые и социальные медиа <https://www.futurelearn.com/courses/community-journalism>
3. Английский для журналистики <https://www.coursera.org/learn/journalism?action=enroll>

---

<sup>195</sup> The Data Journalism Handbook

Модель учебной программы по журналистике: Сборник новых учебных планов, ЮНЕСКО, 2015  
[https://unesdoc.unesco.org/ark:/48223/pf0000221199\\_rus/PDF/221199rus.pdf.multi.page=49](https://unesdoc.unesco.org/ark:/48223/pf0000221199_rus/PDF/221199rus.pdf.multi.page=49)

#### Список использованных ресурсов

1. В. Кадырбаев. Комплексный обед.  
<https://time.kz/articles/Infografika/2020/05/06/kompleksnyj-obed>
2. Суда Б. Дизайн для данных // Пособие по журналистике данных / Под ред. Д. Грея. URL:  
<https://unotices.com/book.php?id=125825>
3. <http://www.stat.kg/ru/news/infografika-8-marta-2020-mezhdunarodnyj-zhenskij-den/>
4. The Data Journalism Handbook
5. [https://halva.tj/articles/faktcheking/tadzhikistan\\_v\\_chislakh\\_infografika\\_halva\\_tj/](https://halva.tj/articles/faktcheking/tadzhikistan_v_chislakh_infografika_halva_tj/)
6. <https://repost.uz/tokyo2020/olympic-history>
7. [https://datavizcatalogue.com/RU/metody/drevovidnaja\\_karta.html](https://datavizcatalogue.com/RU/metody/drevovidnaja_karta.html)
8. [https://commons.wikimedia.org/wiki/File:Turkmenistan\\_Exports\\_Treemap\\_2017.svg](https://commons.wikimedia.org/wiki/File:Turkmenistan_Exports_Treemap_2017.svg)
9. <https://informburo.kz/stati/kakuyu-cast-zarplaty-otdayut-na-kommunalku-kazaxstancy-i-nasi-sosedi-po-sng>
10. <https://akm.apa.kz/infografik/>
11. <https://digitalkz.kz/o-programme/>

## Модуль 10. СТОРИТЕЛЛИНГ КАК ТРЕНД В ЖУРНАЛИСТИКЕ ДАННЫХ

### 10.1. Типы контента в новых медиа

Роберт Макки и Томас Джерас в своей книге «Сториномика» рассказывают, как в США потребители буквально восстали против назойливой традиционной рекламы и, пользуясь новыми технологиями, стали избегать контакта с ней. К примеру, в 2006 году производители видеорекордеров (DVR) заявили о новой функции своего продукта, которая позволяла пропускать 30-секундные рекламные видеоролики. Поставщики кабельного телевидения стали предлагать сервис «видео по запросу» (VOD), благодаря которому потребитель мог смотреть видео без рекламы. Сегодня мы имеем ситуацию, когда огромное количество пользователей устанавливает блокировщики рекламы и наслаждается видеоконтентом в YouTube без рекламы.

Действительно, традиционная реклама раздражает многих. Она крайне редко бывает действительно интересной и креативной. Ее воспринимают как грубое вмешательство в личное пространство. В наше время потребительские привычки сформировались таким образом, что пользователи мессенджеров, например, испытывают негативные эмоции, когда звонит телефон. Точно так же они относятся и к рекламным вставкам.

Исследование, проведенное Национальной ассоциацией рекламодателей и аналитической фирмой Forrester Research, показало, что маркетологи с тревогой и пессимизмом наблюдали за антирекламными нововведениями, как в 2006 году, так и сегодня. Однако на этом работа маркетологов, разумеется, не прекращается. Выход есть всегда и он заключается в смене традиционной рекламной парадигмы на сторителлинговую. Что это значит? Это значит, что людям следует давать то, ради чего они избегают рекламы – добротный контент. Но как добавить рекламную информацию в этот контент? Ответ: надо рассказывать истории и органично вплетать продукт в них. Как это делают голливудские киностудии, занимаясь продакт-плейсментом. Или как это делают некоторые редакции СМИ, создавая нативную рекламу для своих рекламодателей.

О видах контента можно говорить очень долго, и здесь мы можем лишь показать примерную типологию контента в Интернете из того, что мы наблюдаем ежедневно. Разновидностей одного только текстового контента большое множество, к примеру, встречаются такие его варианты:

- Адверториал (рекламная статья)
- Имиджевая статья (PR-статья)
- Обзор
- Опрос
- Эксперимент
- Развлечение: юмор, короткая история
- Новости
- Рейтинг
- Инструкция
- Памятка
- Лайфхак
- Развенчивание мифов
- Проблемная статья
- Обучающая, или полезная статья
- Кейс-стади
- Подборки (идей, инструментов, рецептов, источников)
- Прогнозы и аналитика

- Чеклист
- Познавательная статья
- Очерк
- Репортаж
- Интервью
- Рассказ

Однако современный потребитель интернет-контента не хочет просто читать текст. Особенно, если это длинный текст. Поэтому появился такой интересный формат, как мультимедийный лонгрид. С его помощью длинный текст обрел новую жизнь в Сети. На самом деле люди с удовольствием читают большие статьи, если они содержат реально интересную историю, мастерски написаны и эффективно оформлены. Об этом как раз и пишут эксперты по лонгридам. Но к этому формату мы еще вернемся.

Что касается сторителлинга в целом, то, по нашим, наблюдениям, в современном интернет-пространстве можно выделить следующие форматы:

- Лонгрид/мультимедийный лонгрид
- Блог (ЯндексДзен, Blogspot, Facebook и пр.); галереи в ЯндексДзен
- Telegra.ph (Telegram), а также посты-галереи в Telegram
- Instagram Stories, карусель в Инстаграме
- Tik-Tok (видео)
- Чат-бот
- Компьютерная игра
- Мобильное приложение
- Email-рассылка
- Публичное выступление
- Видео-сторителлинг: видеоролик для YouTube/Instagram/Tik-Tok/видеокаст/видеопрезентация/документальный фильм/мультфильм/кинофильм
- Аудио-сторителлинг: аудиоролик/подкаст/аудиостатья/аудиокнига
- Маркетинг-кит (Marketing Kit)
- Специальный отчет (Special Report)

## 10.2. Мультимедийный лонгрид и его место в дата-сторителлинге

Сравнительно недавно в английском языке появился новый термин лонгрид (англ. long – длинный, read – читать, т.е. длительное чтение). Согласно своей этимологии, термин лонгрид означает любой длинный текст. Вместе с тем, на основе данного понятия, возникло такое интересное интернет-явление как мультимедийный лонгрид, который являет собой особый медиаформат и жанр публикаций в Интернете. “Мультимедийный” в данном случае означает наличие в публикации различных медийных платформ: таких как вербальный текст, фотографии, аудио-, видео, графики, анимации и прочих производных от них форм. Еще одной особенностью лонгрида является то, что он оформляется как отдельная веб-страница со своим доменом, часто второго уровня (иногда первого). Это позволяет избавить читателя от многих отвлекающих от чтения элементов и сосредоточить внимание на истории. Лонгрид является наследником традиционной лонгформ-журналистики (англ. long form - длинная форма) или нарративной журналистики. С точки зрения задач работы с аудиторией лонгриды можно условно разделить на две крупные категории: некоммерческие и коммерческие. Коммерческие лонгриды относятся к

инструментам маркетинга, а некоммерческие – информационным, обучающим, пропагандистским и пр.

Достаточно часто в международной практике, особенно в западных СМИ, встречается термин “мультимедийная история” (от англ. story), а также “фичер” (англ. Feature - очерк). Эти понятия применяют для обозначения объемных публикаций в Интернете, к которым относятся и лонгриды. Пример термин feature пришел в онлайн-среду из традиционных бумажных СМИ, особенно глянцевого.

Первым в истории опубликованным примером жанра мультимедийного лонгрида считается публикация портала The New York Times "Snowfall" (<http://www.nytimes.com/projects/2012/snowfall/index.html>), увидевшая свет в 2012 г. Он рассказывает весьма увлекательную историю о приключениях группы горнолыжников в Каскадных горах США, которую прочитали десятки миллионов интернет-пользователей.

С тех пор данный жанр получил большое развитие и распространение на просторах Сети, став устойчивым трендом в онлайн-медиа, к которым относятся как онлайн-СМИ, так и корпоративные блоги, а также персональные блоги.

Другими яркими примерами лонгридов (в том понимании, которое мы закладываем в данной работе) в мировых англоязычных онлайн-СМИ являются: “A Game of Shark and Minnow” (The New York Times) - <http://www.nytimes.com/newsgraphics/2013/10/27/south-china-sea/index.html>, “Out in the Great Alone” (Espin.com) - [http://www.espn.com/espn/feature/story/\\_/id/9175394/out-great-alone](http://www.espn.com/espn/feature/story/_/id/9175394/out-great-alone). В русскоязычных, в частности, российских онлайн-СМИ также активно ведется работа в данном направлении и их творения заслуживают не меньшего внимания. В качестве примеров можно привести следующие: “Конец советов. 20 лет” (Lenta.ru) - <https://age.lenta.ru/1993>, “Реальный Готэм-сити” (Модный Петербург) - <http://modny.spb.ru/articles/i-love-ny>, “Чернобыль. Земля отчуждения” (Kommersant.ru) - <https://www.kommersant.ru/projects/chernobyl>. Казахстанские СМИ тоже внедряют мультимедийный формат публикаций. Так онлайн-издание Vlast.kz на регулярной основе размещает в одном из своих разделов подобные посты, поднимающие их авторами злободневные темы: “Идеализм как выбор” (Vlast.kz) - <https://vlast.kz/avtory/33418-idealizm-kak-vybor.html>, “Маленькие люди на грани нервного срыва” (Vlast.kz) - <https://vlast.kz/avtory/33143-malenkie-ludi-na-grani-nervnogo-sryva.html>, “Как алматинцы начали бояться автобусов” (Vlast.kz) - <https://vlast.kz/gorod/30533-kak-almatincy-nacali-boatsa-avtobusov.html>, “Переломит ли гибель Дениса Тена ход истории Казахстана?” (Vlast.kz) - <https://vlast.kz/avtory/28776-perelomit-li-gibel-denisa-tena-hod-istorii-kazahstana.html>.

#### **Типы лонгридов:**

- Репортажный
- Рекламный
- Портретный
- Биографический
- Инструктивный
- Реконструктивный
- Путевой

### **10.3. Особенности дата-сторителлинга**

Мультимедийный лонгрид, безусловно, является серьезным трендом. Однако, требуется высокое журналистское мастерство, чтобы создать достойный мультимедийный лонгрид, который прочтут многие пользователи и станут делиться им, придавая ему вирусный эффект. Кроме того, поисковые системы «любят» этот формат, поскольку считают его более полезным и интересным

для аудитории, интересам которой они служат. По этой причине мультимедийные лонгриды часто называют “evergreen content”, т.к. такие материалы годами сохраняют свои позиции в органической поисковой выдаче. Однако следует рассмотреть и особенности дата-журналистских материалов и понять, насколько мультимедийный лонгрид как формат можно адаптировать к задачам дата-журналистики.

Приведем типичный пример дата-сторителлинга без признаков лонгрида. По сути это одностраничный сайт с контентом: <http://worldfoodclock.com/>

Как мы видим, дата-сторителлинг преимущественно реализуется на основе визуализации данных. Текста в привычном статейном понимании при этом используется мало, либо вообще обходятся без него. Приведем еще несколько успешных примеров от известных брендов. Заодно рассмотрим, какую пользу способен приносить дата-сторителлинг обществу.

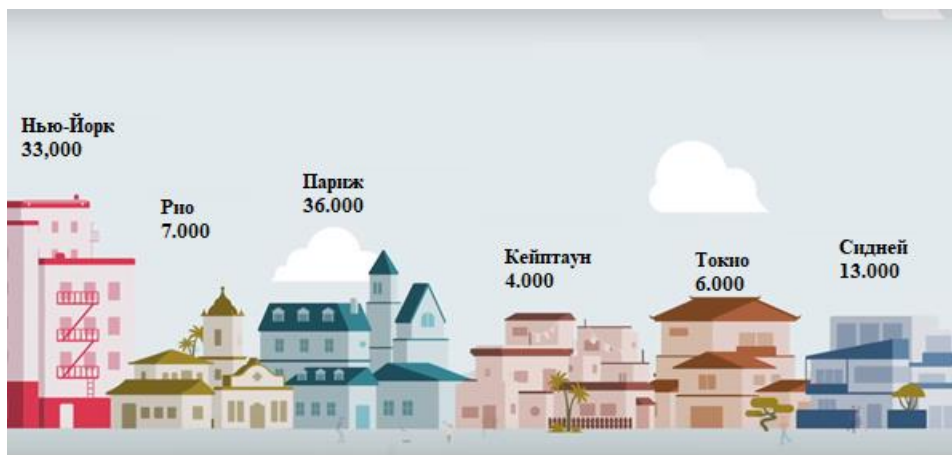


Рисунок № 65. Успешный пример визуализации данных

#### **AirBnB.**

Сторителлинг — это основа маркетинга Airbnb. В маркетинговых сообщениях особое внимание уделяется комьюнити и традициям местного гостеприимства. Это направлено на то, чтобы пользователи сервиса получали уникальный ценный опыт во время своих поездок. Кроме того, Airbnb отличается креативным использованием данных о пользователях — сервис превращает их в истории. Например, в преддверии 2015 года компания подготовила видео, в котором рассказала о своей истории. Также в ролике сообщалось, что около 550 тысяч путешественников провели канун Нового года в арендованном с помощью сервиса жилье в 20 тысячах городов мира. А всего пятью годами ранее сервисом пользовались около 2 000 человек. Самым популярным городом в канун Нового года по версии Airbnb стал Нью-Йорк: там с помощью сервиса сняли жилье около 47 тысяч путешественников. Источник: <https://bakunin.com/data-storytelling/>

#### **SPOTIFY.**

Spotify постоянно собирает данные о том, какие песни, плейлисты и исполнителей выбирают 30 миллионов пользователей. Эту информацию объединяют с данными о местонахождении и демографии слушателей. Так получается оригинальный контент для блога Spotify Insights. Так, в мае 2017 года Spotify представил исследование о том, как и какую музыку слушают американские студенты колледжей и университетов. Исследование было представлено в виде интерактивного лендинга. На сайте сообщалось, где чаще всего слушали музыку, какая это была музыка и т.д. (Источник: <https://bakunin.com/data-storytelling/>).

## **GOOGLE.**

Классический пример сторителлинга у Google — это видео «Год в поиске». Корпорация выпускает такие видео ежегодно и собирает в них наиболее популярные поисковые запросы (которые в каком-то смысле отражают состояние и настроение всей нации). Так, в 2016 году Google подготовила двухминутный ролик о главных поисковых запросах, а также показал ключевые моменты года — как радостные, так и трагические. В ходе тестирования зрители в большинстве своем оставили положительные отзывы, и видео попало в 1% лучших рекламных объявлений за год. Как Google удастся вызвать такие сильные эмоции у зрителей? Во-первых, освещаются события, которые каким-то образом затронули каждого, а во-вторых, используются данные, позволяющие определить, какие темы и события наиболее актуальны для аудитории (См.: <https://bakunin.com/data-storytelling/> )

**ZILLOW.** Базирующийся в США онлайн-рынок недвижимости Zillow располагает данными о более чем 110 миллионах домов. Эта информация включает не только стоимость и метраж, но и, например, аэрофотоснимки. Компания использует эти данные для создания контента. В блоге компании есть как стандартные материалы, например, о том, где миллениалам найти доступное жилье, так и нестандартные. Так, в 2016 году Zillow в преддверии Хэллоуина составил рейтинг из 20 лучших городов для выпрашивания сладостей (trick or treat) — традиционной для праздника забавы. Рейтинг был основан на данных о ценах домов, их близости друг к другу, а также об уровне преступности за последние 10 лет. Рейтинг сопровождался инфографикой. Лидерами рейтинга стали Филадельфия, Сан-Хосе, Сан-Франциско, Милуоки и Лос-Анджелес (Источник: <https://bakunin.com/data-storytelling/> )

## **HINGE**

Hinge — это приложение для знакомств для тех, кто «выше» Tinder. Hinge собирает данные о пользователях, чтобы рассказывать им истории о них же. По статистике Hinge, 81% пользователей никогда не заводили долгосрочных отношений при помощи приложений для знакомств. Эта идея легла в основу центральной истории Hinge: The Dating Apocalypse («Апокалипсис для свиданий»). История оформлена в виде ролика, в котором молодой человек бродит по заброшенному парку аттракционов, символизирующему приложения для знакомств (в частности, Tinder), пока наконец не находит выход и перед ним не возникает перспектива построить долгосрочные отношения (при помощи Hinge). По заявлениям Hinge, приложение поощряет людей «избегать игр и находить что-то реальное» и открывает для них мир возможностей. Ключевое сообщение бренда сформулировано следующим образом: «Приложения для знакомств превратились в игру, и с каждым свайпом мы отдаляемся от реальных отношений, которых так хотим. Мы создали кое-что получше». (Источник: <https://bakunin.com/data-storytelling/> )

## **HUGGIES.**

Главным конкурентом Huggies в Канаде была компания Pampers (лидер рынка и поставщик подгузников по контракту во все больницы страны). Чтобы отвоевать долю рынка, Huggies должны были убедить рожениц выбрать именно их марку еще до приезда в роддом. Причина должна была иметь, с одной стороны, мощный эмоциональный, а с другой стороны — материальный компонент. Решение проблемы нашли в названии марки: историю решили строить вокруг объятий (huggies можно перевести с английского как «объятия» или даже, точнее, «обнимашки»: от глагола to hug — «обнимать»). Основная идея, которую стали транслировать Huggies, — ребенка нельзя оставлять без объятий. Компания привела данные более чем 600 исследований, которые доказали, что объятия «укрепляют иммунную систему детей, предотвращают болезни и улучшают развитие мозга». Чтобы выполнить свою миссию — не оставлять ни одного ребенка без объятий — Huggies

развили деятельность по двум направлениям: объяснение матерям важности объятий с ребенком (контакта кожа к коже) и контроль за тем, чтобы в канадских больницах не было недостатка в волонтерах — «обнимателях» детей. Кампания позволила Huggies увеличить продажи в 2016 году на 30%. При этом уровень вовлеченности клиентов был в три раза выше, чем в среднем по отрасли.

Источник: <https://bakunin.com/data-storytelling/>

#### **IBM.**

Каждые шесть часов в Австралии от меланомы умирает один человек. Эта информация побудила компанию IBM задействовать искусственный интеллект, чтобы «перехитрить меланому» и обеспечить раннее выявление этого вида рака.

Был создан ИИ, получивший имя Watson. Он диагностировал меланому на 31% точнее, чем это можно было сделать невооруженным глазом. Кампанию запустили прямо на пляже Бонди-Бич в Сиднее в самый разгар сезона. Участниками эксперимента стали обычные австралийцы. Их приглашали посмотреть в зеркало, в то время как ИИ анализировал их пол, возраст и то, каким солнцезащитным кремом пользуется человек. На площадке проведения эксперимента дежурил медик, если ИИ выявлял какие-то риски или нарушения, человека отправляли к специалисту. За один выходной было осмотрено 800 человек. 22% из них были направлены к медикам. Так IBM показала себя не только как высокотехнологичный бренд, но и как бренд, заботящийся о здоровье пользователей.

Источник: <https://bakunin.com/data-storytelling/>

MALTESERS. В 2016 году британский бренд Maltesers решил осветить в рекламе проблему отношения в обществе к людям с инвалидностью. Выявив тот факт, что 80% людей с ограниченными возможностями чувствуют себя недостаточно представленными на телевидении и в средствах массовой информации, Maltesers создал серию рекламных роликов, вдохновленных реальными историями людей с ограниченными возможностями. При этом акцент был сделан на неловкие ситуации, которые случаются в жизни любого человека. Например, в одном из роликов девушка с ДЦП рассказывает о неловкой ситуации со своим бойфрендом. Инклюзивная реклама помогла Maltesers увеличить продажи на 8,1%. Кроме того, она стала самым популярным роликом на YouTube-канале бренда. После кампании 57% потребителей заявили, что Maltesers изменили их отношение к инвалидности.

Источник: <https://bakunin.com/data-storytelling/>

Ссылка на видео: <https://www.youtube.com/watch?v=YgUqmKQ9Lrg>

**WHIRLPOOL.** Бренд бытовой техники Whirlpool обнаружил неожиданную причину, по которой 4000 американских учащихся каждый день уходят из школы. Выяснилось, что их семьи не могли позволить себе регулярно стирать одежду (речь идет о бедных слоях населения). Компания запустила программу Care Counts, в рамках которой в школах устанавливали стиральные машинки и сушилки.

После того как детям из бедных семей предоставили доступ к стиральным машинам, посещаемость улучшилась на 90%. Также ребята стали активнее проявлять себя в классе.

Источник: <https://bakunin.com/data-storytelling/>

**REFINERY29.** Популярный американский сайт лайфстайл-тематики Refinery29 обнаружил, что 67% американок — дамы plus size, но в медиапространстве их представленность составляет всего 2%. Refinery29 объединился с популярным фотостоком Getty Images, чтобы создать новую коллекцию изображений, соответствующую реальной ситуации. Коллекция была бесплатна. Аудиторию призывали распространять ее в соцсетях с тегом #SeeThe67. Только глубоко поняв свою аудиторию, бренд может выступить как ее защитник, сформулировать наиболее подходящие ценностные сообщения и вовлечь людей в свою историю.

Источник: <https://bakunin.com/data-storytelling/>

#### 10.4. Примеры лонгридов в журналистике данных

Теперь рассмотрим некоторые примеры того, как длинные тексты применяются для дата-сторителлинга. Бренд Tinkoff периодически публикует интересные истории на различные актуальные темы. Пример: <https://clck.ru/XtAR6>

По сравнению с полноформатными мультимедийными лонгридами данный пример больше принадлежит к категории просто статей. Однако если его подтянуть по формату до уровня современного мультимедийного лонгрида, то материал может получиться еще интереснее. В качестве иллюстрации возможности интеграции лонгрида в дата-историю данный пример является показательным. Вопрос имплементации мультимедийного лонгрида как современного медийного формата мы позиционируем в данной публикации как перспективный и имеющий основания для дальнейших исследований.

##### Вопросы:

1. Как вы понимаете суть и предназначение сторителлинга в современном медиапространстве?
2. Какие факторы обуславливают актуальность сторителлинга как формы коммуникации в современной культуре потребления интернет-контента?
3. В чем особенность мультимедийного лонгрида как особого формата в современном Интернете?
4. Какие форматы сторителлинга вы можете перечислить?
5. Сторителлинг-кейс от Whirlpool: в чем заключалась идея?
6. Какой ключевой месседж вы поняли из видеоистории от бренда Maltesers?
7. Для чего компания Spotify собирает данные о своих слушателях?
8. Расскажите о своем видении развития дата-сторителлинга и обоснуйте свою позицию с точки зрения современных привычек потребления контента у аудитории.
9. Как повлияло на развитие сторителлинга в Интернете отношение пользователей (аудитории) к рекламе?

##### Задания:

1. Возьмите любой небольшой новостной материал на новостном сайте и создайте на его основе небольшую историю, которую читать было бы интереснее, чем просто обычную новость.
2. Расскажите или напишите небольшую историю о своем соседе по подъезду. Выделите характерные свойства его личности, его явную ключевую мотивацию в ежедневной жизни, его взгляды и привычки и пр. Каким должен быть небольшой рассказ о нем, чтобы заинтересовать читателей?
3. Проанализируйте кейс Hinge: The Dating Apocalypse. В чем коммуникационная задача бренда в данном случае? Почему и чем эта история должна заинтересовать аудиторию?

##### Литература

1. Колесниченко А.В. Long Texts (Longreads) in the Modern Russian Press, электронный научный журнал “Медиаскоп”, выпуск №1, 2015.
2. The Revival of Long Reading: A New Multimodal Narrative Format Milena Tsvekova, Postmodernism problems, Volume 8, Number 2, 2018.

3. Чигаев Д.П. Лонгрид как разновидность креолизованного текста // Медиаскоп. 2017. Вып. 1.
4. Колесниченко Александр Васильевич, Long Texts (Longreads) in the Modern Russian Press (<http://mediascope7.mediascope.ru/?q=node/1691>).
5. Климко А.Ю. ПОРТРЕТНЫЙ ОЧЕРК В ФОРМАТЕ ЛОНГРИДА.
6. Распопова С.С. АВТОР МУЛЬТИМЕДИЙНОГО ТЕКСТА // Вестник Челябинского государственного университета. 2012. № 13 (267). Филология. Искусствоведение. Вып. 65. С. 100–102.
7. Бозрикова С.А., Татару Л.В. НАРРАТИВНАЯ ЖУРНАЛИСТИКА В АМЕРИКЕ И В РОССИИ.

## Модуль 11. ЭТИЧЕСКИЕ АСПЕКТЫ ЖУРНАЛИСТИКИ ДАННЫХ

### 11.1. Специфика этики в журналистике данных

Согласно Брэдшоу (2013), этические соображения в журналистике данных существенно не отличаются от таковых в любой другой области журналистики. В принципе, журналистика данных сталкивается с теми же этическими проблемами, что и традиционная журналистика, например, в отношении угрозы конфиденциальности и неправдивого представления<sup>196</sup>. Однако информационные технологии, используемые в такого рода журналистской практике, создали дополнительные проблемы, связанные с доступом к информационным материалам, использованием данных при анализе, прозрачностью публикации результатов и открытостью для общественности, обратной связью и исправлением ошибок в рассказе.

Журналист может задавать неуместные вопросы о данных, анализировать неверные данные или пропускать важные данные в своих репортажах. Проблемы могут возникнуть, если отсутствуют данные, методы анализа недостаточны, выводы не обоснованы или общая картина проблемы недостаточно широка. Однако ошибки должны быть устранены, если сотрудники редакции изучают и обсуждают исходные данные и методы работы, а также если используемые данные и анализ становятся прозрачными в статье для общественности.

С этической точки зрения одним из важных моментов в журналистике данных является отношение журналистов к своей собственной работе и к тому, как они должны вести себя в своей профессии и выполнять свою роль сторожевого пса.

Данные - это удобный инструмент для "выяснения правды", который является одним из важнейших принципов работы журналистики<sup>197</sup>. На самом деле формат информации, основанной на данных, открывает новые перспективы и позволяет журналистам взглянуть на сюжетные темы с высоты птичьего полета. Информация, созданная таким образом, открывает журналистам "новые горизонты" в их работе. Подход к журналистике данных позволяет журналистам создавать новые типы объектов и видеть более широкие закономерности охватываемых тем. Используя методы журналистики данных, журналисты могут выявить связи и причины, лежащие в основе, казалось бы, отдельных проблем, которые не раскрываются только отдельными данными. Анализ данных как таковой также перенаправляет более глубокий сбор информации на проблемы, которые вызвали результаты анализа. Зависимость журналистов от властей как источника информации уменьшается за счет использования общедоступных данных, в то время как их собственный опыт в этой теме возрастает. Общественные учреждения сферы услуг теряют свое прежнее исключительное право на интерпретацию материалов данных и свою монополию на производство статистических данных и отчетов о своей собственной деятельности.

Технические характеристики цифровых данных могут вызвать ряд этических проблем. Этические проблемы в журналистике данных в основном связаны с подотчетностью данных и наборов данных. При изучении наборов данных важно обращать внимание на происхождение, достаточность и нехватку данных. Журналисты должны задавать данным те же вопросы, что и любому другому источнику, и оценивать этику данных с нескольких точек зрения: были ли найдены

---

<sup>196</sup> Bradshaw, Paul (2013). Ethics in data journalism: accuracy. Online journalism blog. [Blog] September 13, 2013. Available at <https://onlinejournalismblog.com/2013/09/13/ethics-indata-journalism-accuracy/> {Accessed 25 November, 2017}

<sup>197</sup> Kuutti, H (2016). Ethics of data journalism Four ethical phases in the working process, a paper presented at Nordic Data Journalism Conference NODA 2016, 21 April 2016 in Helsinki, Finland

все данные, необходимые для статьи, насколько обоснован и достаточен метод анализа, насколько обоснованы выводы, достаточно ли широка общая картина и точен ли контекст. Брэдшоу подчеркивает, что журналисты должны задавать следующие вопросы:

- В чем заключается личная заинтересованность лица, предоставляющего данные?
- Как была собрана эта информация, и что или кто (или когда или где) может отсутствовать в ней?
- Как были сформулированы вопросы и какие вопросы использовались для предварительной обработки данных?

- Есть ли второй независимый источник той же информации или другая ее интерпретация?

- Какова допустимая погрешность?

- Есть ли у журналиста знания, чтобы задавать правильные вопросы всем этим источникам

Лафлер (2014) советует проверить целостность данных с помощью следующих шагов: прочитайте (оригинальную) документацию, проверьте, сколько записей у вас должно быть, сверьте подсчеты и итоги с отчетами и найдите возможные недостающие данные, дубликаты и внутренние проблемы<sup>198</sup>.

## 11.2. Этические коллизии в работе с «грязными» данными

“Грязные” данные относятся к нескольким ошибкам в данных и широко используются среди журналистов данных. Проблемы с гними обычно возникают в процессе хранения, передачи и регистрации и приводят к исчезновению или смешиванию связи между полями данных и наборами данных. Кроме того, ошибки при сборе и передаче данных могут вызывать проблемы. Манипулировать данными на различных этапах работы легко, поскольку те, кто регистрирует, хранит и передает данные, являются людьми.

Связи между полями данных и базами данных могут исчезнуть или перепутаться, или переменные, используемые в анализе, устарели.

Вебстер (2014) предупреждает, что если данные не останутся в исходном состоянии, они могут быть изменены и, следовательно, содержать ошибки программирования. “Грязь” может быть в самих данных или в больших наборах данных. Данные являются грязными, если они неверны, хранятся в другом формате или устарели. Грязь в наборах данных относится к базам данных, в которых могут отсутствовать соответствующие данные или целые поля данных, данные хранятся в неправильных полях или одни и те же данные хранятся более одного раза. Хорошей отправной точкой является то, что все данные являются грязными<sup>199</sup>.

Соответственно, последний вопрос заключается в том, насколько на самом деле грязные данные. Причины “грязных” данных обычно кроются в различных мотивах использования баз данных. Если власти обычно хранят данные для своих собственных различных целей сбора информации, то журналисты просматривают весь набор данных для поиска новостей или подтверждения своей гипотезы о сюжете.

Запросы открытых данных отличаются от традиционных запросов информации. Запросы данных относятся к крупным информационным объектам, и цель состоит в том, чтобы повторно использовать данные. Национальные политические стратегии направлены на то, чтобы открыть

---

198 LaFleur, Jennifer (2014). Data journalism tips from data driven ethics. Conference at Markkula Center for Applied Ethics [Blog] 1 January, 2014. Available at <https://www.scu.edu/ethics/focus-areas/journalism-ethics/resources/data-journalism-tips-/> [Accessed 20 November, 2017]

199 Webster, Maryjo (2014). Data State of Mind. IRE Journal, Winter 2014:11.

данные, которые будут использоваться для различных целей без специального регулирования и, следовательно, для развития информационной отрасли с помощью новых приложений.

Однако открытые данные не являются решением проблемы доступа к данным, поскольку это всего лишь часть всех общедоступных данных, опубликованная добровольно. Даже если он не содержит четкого измерения связей с общественностью (PR), власти владельца проверили и очистили его. Кроме того, публикация открытых данных может быть способом скрыть более важные с журналистской точки зрения данные.

Общедоступные данные содержат разнообразную информацию, и власти обычно не располагают точным пониманием ее общего содержания. Поэтому они могут неохотно выпускать ее.

Обычно органы власти только собирают данные без какой-либо конкретной ответственности за их анализ.

Запросы данных могут вызвать опасения у властей по поводу возможных “вредных” результатов, выявленных “непредсказуемыми” анализами неизвестных журналистов.

Согласно финскому исследованию (Kuutti 2011), было выявлено несколько проблем, связанных с поведением государственных учреждений в ответ на запросы данных. Даже если информация носит публичный характер, власти могут категорически избегать??? массовый выпуск данных. Основная проблема в доступе к данным, по-видимому, заключается в электронном формате информации и полноте запросов, которые считаются более сложными задачами, чем отдельные запросы на информацию или документы. Компьютерные системы государственных учреждений не предназначены для обслуживания свободы доступа к информации. Тесты доступа, проведенные финским исследованием, выявили ряд недостатков в администрировании документов и внедрении информационных систем. Отклоняя запросы о предоставлении данных, власти ссылаются на негибкость информационных систем. Во многих информационных запросах на предоставление данных государственное учреждение выполняет “специальные задания” за вознаграждение для подготовки и редактирования материалов, содержащих как общедоступные, так и конфиденциальные данные.

Согласно нескольким государственным учреждениям, невозможно отделить конфиденциальные элементы от общедоступных элементов, или разделение может быть проведено только вручную, т. е. трудоемким и дорогостоящим способом. Чтобы быть в безопасности в неоднозначных ситуациях, государственные учреждения предпочитают не предоставлять информацию. Даже несмотря на то, что это может быть оспорено, любой отказ в передаче публичных материалов как таковой является юридическим толкованием и правом соответствующего государственного учреждения, в то время как предоставление конфиденциальных материалов при запросе может привести к уголовному преследованию.

С другой стороны, доступ к данным может привести к их неограниченному использованию и побудить журналистов воспользоваться преимуществами поспешного и небрежного использования данных без учета ограничений данных в статье. Кроме того, такого рода данные, вызванные слепотой, могут привести к ограничению сбора информации только для материалов баз данных и тех (менее важных) сюжетных идей, где имеются данные.

### **11.3. Доступ к данным в ракурсе этики**

Этические проблемы, связанные с доступом к данным, также имеют свои последствия при анализе данных. По словам Брэдшоу (2013), вероятно, основным этическим соображением в журналистике данных является необходимость быть точным и обеспечивать надлежащий контекст

историям, которые рассказывают журналисты. Это должно влиять на то, как журналисты анализируют данные, сообщают об историях данных или публикуют данные.

Точность, пожалуй, является главной заботой журналистов, работающих с любой формой данных. Хьюстон (2004) напоминает журналистам быть точными и справедливыми, а не видеть только то, что они надеются увидеть. Это возможно, поскольку журналистика данных работает с числами, которые могут быть использованы в качестве резервных копий собственных интерпретаций. Единственная хорошая история - это та, которая точно резюмирует то, что вы нашли.

Предположения и выводы могут быть оспорены. Журналисты, использующие данные, должны быть осторожны, чтобы не делать ложных сравнений<sup>200</sup>. Например, Лефлер (2014) советует журналистам искать возможные альтернативные результаты и спрашивать, чем еще можно объяснить их выводы, а также проверять, собраны ли все необходимые данные. Ответственность журналистов заключается в том, чтобы убедиться, что данные, используемые в статье, действительно говорят о том, что они должны делать, и, соответственно, не говорят о том, чего они не могут сделать<sup>201</sup>. Каир (2014,) указывает, что даже если выбранная переменная имеет отношение к истории, она может быть не лучшей альтернативой для рассказа истории и, следовательно, нуждается в изменении<sup>202</sup>. Скептицизм также необходим, когда журналисты сами собирают данные. Зависящий от того, как используются результаты, сбор данных, например, с помощью краудсорсинга, может потребовать проверки необычных моделей поведения, которые могут свидетельствовать о том, что человеческие источники играют в систему. Также полезен второй механизм проверки, при котором некоторые материалы проверяются другими пользователями или даже требуют представления некоторых доказательств вместе с данными.

О'Нейл (2013) критикует журналистов за то, что они думают, что все поддается измерению и что выбранные ими переменные являются жизнеспособными для выявления конкретных проблем. В отличие от этого, журналисты не обязательно технически ориентированы, чтобы обнаружить основное понятие в данных и их анализе. Они не знают, как преобразовать важные вопросы журналистики в математические модели<sup>203</sup>.

Для аналогового журналиста очень сложно быть математически точным и технически подкованным, чтобы получить то, что он хотят. Как говорит Отьено (2013), "каждый алгоритм, как бы он ни был написан, содержит человеческое и, следовательно, редакционное суждение. Те решения, принятые о том, какие данные включать и исключать, добавляют уровень перспективы к предоставленной информации"<sup>204</sup>.

Цифры, диаграммы и карты обладают авторитетом, которого часто не хватает другим видам информации, и все же они в равной степени подвержены манипуляциям. Журналистам необходимо быть осторожными как в отношении достоверности, которую они придают числовым и графическим источникам, так и в том, как они представляют свои собственные истории в цифрах и графически.

---

<sup>200</sup> Houston, Brant (2004). Computer assisted reporting. A practical guide. Bedford / St. Martin's

<sup>201</sup> LaFleur, Jennifer (2014). Data journalism tips from data driven ethics. Conference at Markkula Center for Applied Ethics [Blog] 1 January, 2014. Available at <https://www.scu.edu/ethics/focus-areas/journalism-ethics/resources/data-journalism-tips/> [Accessed 20 November, 2017]

<sup>202</sup> Cairo, Alberto (2014). Ethical infographics. IRE Journal, Spring 2014: 25-27.

<sup>203</sup> O'Neil, Cathy (2013). On Being a Data Skeptic. O'Reilly Media Available at <http://www.oreilly.com/data/free/being-a-data-skeptic.csp> [Accessed 25 November, 2017]

<sup>204</sup> Otieno, Samuel (2013). Numbers don't lie: my experience with data journalism. Data Journalism Fellowship Blog [Blog] Data Dredger. Available at <http://internewsenya.org/fellowshipblog/numbers-dont-lie-my-experience-with-datajournalism/#sthash.dkVisHqo.dpbs> [Accessed 25 November, 2017]

Поскольку все большая доля источников журналистов связана с данными, математика становится такой же важной, как грамотность. Случай путаницы “процентного увеличения” с “увеличением на процентный пункт” должен быть таким же постыдным, как неправильное написание чьего-либо имени. Журналисты должны быть настолько же конкретны в формулировках, касающихся данных, насколько они конкретны в описании событий и людей, где нет места расплывчатости или путанице.

Журналистика данных также требует тщательного обдумывания предубеждений, встроенных в оригинал данных и их интерпретацию. Врожденные бессознательные предубеждения могут влиять на принятие решений, искажать интерпретацию и создавать “белые пятна” в сборе, анализе данных и их восприятии общественностью. Журналисты должны осознать свои собственные предубеждения, например, взяв на себя обязательство выполнять этическую работу, зная свои собственные ограничения и ограничения данных.

Однако лихорадочная журналистская деятельность может испортить анализ данных, который требует внимания, как указывает Сегнини (2015). По его словам, неправильное использование статистики может это происходит легко, и риски возрастают, когда журналисты пытаются установить причинно-следственную связь на основе больших или необъяснимых наборов данных<sup>205</sup>.

Данные, используемые в статье, должны иметь четкую связь с результатами. Данные могут быть ошибочными, вводящими в заблуждение, вредными, неудобными или агрессивными. Представление данных как форма журналистики требует, чтобы журналисты подвергали данные журналистскому процессу<sup>206</sup>.

Даже если данные точны, их связь с реальностью может быть недостаточной, и они могут неадекватно соотноситься с реальностью, которую они должны описывать. Это может быть серьезной проблемой в историях об опросах общественного мнения, проведенных с использованием недостаточных выборок или неправильных вопросов. Сила, которой обладают журналисты как профессионалы, - это внимание. Журналисты привлекают внимание к чему-то, и это внимание имеет хорошие и плохие последствия. Решения, которые принимают журналисты, часто касаются того, что происходит, когда внимание сосредоточено на этой вещи<sup>[25]</sup>.

Даже анонимизированные данные могут раскрывать личную информацию. Журналистам также может потребоваться проявлять осторожность в отношении защиты своих источников в том, как они публикуют утечку данных. Метаданные, хранящиеся в файлах, касающиеся даты и местоположения доступа, используемых компьютеров и учетных записей, а также других данных, также могут идентифицировать источник.

Сондерман (2013) отмечает, что журналисты не должны предполагать, что данные по своей сути являются точными, справедливыми и объективными. Кроме того, они не должны ошибочно принимать свой доступ к данным или свое право на их публикацию как законные основания для этого. При публикации журналистам следует критически относиться к общественному благу и потенциальному вреду, контексту, окружающему данные, и их значимости для других репортажей.

Журналисты, занимающиеся опросами, всегда должны запрашивать доступ к исходным данным, включая все заданные в его ходе вопросы. Если эта информация не будет предоставлена, то истинность утверждений не будет установлена, и журналисту необходимо принять решение не

---

<sup>205</sup> Segnini, Giannina (2015). Investigating with databases: Verifying data quality. Craig Silverman (ed.): Verification handbook. [online] European Journalism Center. Available at <https://gijn.org/2015/05/14/investigating-with-databases-verifying-data-quality/> [Accessed 25 November, 2017]

<sup>206</sup> Sonderman, Jeff (2013). Programmers explain how to turn data into journalism & why that matters. Poynter institute. Available at <https://www.poynter.org/news/programmers-explainhow-turn-data-journalism-why-matters> [Accessed 25 November, 2017]

публиковать его итоги. Однако, если опрос уже находится в открытом доступе, этическое решение тогда касается того, должен ли журналист сообщать об отсутствии в публикации более подробной информации или данных, а также о критике в отношении используемых методов.

В осторожной истории с данными вопрос заключается в том, насколько журналисты могут растянуть данные анализ, если данные не являются чистыми на сто процентов. Загрязнение данных является относительным, и определенное количество загрязненных данных не обязательно загрязняет весь набор данных. Однако журналисты должны учитывать даже небольшие угрозы ошибок, вызванные грязными данными в результатах, и объяснять их общественности в статье. Одним из аспектов, поддерживающих поспешную публикацию в подобной ситуации, может быть важность публикации даже приблизительных результатов.

Прогнозы - это один из видов данных, в котором конфликт между принципом точности и минимизация вреда выходят на первый план. Публикация прогнозов может быть самореализующейся или самоуничтожающейся, а также опасной для публикации. Журналисты должны быть особенно осторожны, когда другие прогнозы не указывают на то же самое, как это может произойти в опросах о намерениях голосования на выборах.

#### **11.4. При представлении данных ключевым является контекст**

Одни только измеренные результаты ничего не говорят о том, являются ли эти цифры выше или ниже, чем они должны быть, растут или падают, являются ли они лучшими или худшими в регионе, стране или мире. Представление их в историческом контексте, по лицам или по датам помогает сделать цифры более значимыми. Персонализация, однако, может создавать свои собственные проблемы. Если журналисты рассказывают пользователям, как события влияют на них, они также должны иметь представление о более широкой картине.

Одним из этических вопросов является уровень детализации, на котором журналист фактически обязан рассказывать историю. Совокупная и менее личная информация может дать более четкое представление о более широких тенденциях, в то время как случайные проверки достоверности данных могут привести к появлению историй о недостатках в такой общедоступной информации.

Добавление контекста является жизненно важной частью процесса журналистики данных. Журналисты должны приводить абсолютные цифры в контексте численности местного населения, исторических закономерностей и даже различий в демографии. Более того, они должны проверять тенденции в отношении изменений границ или методов сбора и классификации данных.

Журналисты должны придерживаться этического принципа прозрачности при атрибуции источников и ссылках на полные данные, где это возможно, с исключениями, подробно описанными как выше, так и в разделе о защите источников ниже.

Когда история завершена, журналистам нужно вернуться назад и проверить факты построчно. Всякий раз, когда это возможно, факты должны быть прослежены до любых или обоих данных и документов. Согласно Хьюстону (2004), журналистам необходимо играть роль “адвоката дьявола” в работе с данными и выдвигать аргументы против любого первоначального вывода. Хьюстон утверждает, что журналисты должны делиться основными моментами собственных выводов во время своих интервью, а также выслушивать и учитывать любую критику. Как он резюмирует, “гораздо лучше выяснить, что вы ошибаетесь, до того, как опубликуете историю, чем после того, как ее увидит публика”. Кроме того, журналистам необходимо постоянно думать о “скрытой переменной”, относящейся к другим факторам, которые могут привести к тому, что данные будут выглядеть так, как их видит журналист.

### **11.5. Этическая ответственность при работе с необработанными данными и их точность**

Еще одна этическая проблема связана с размещением необработанной базы данных публичных записей в Интернете. Если журналисты не занимаются каким-либо анализом или какой-либо дополнительной работой с данными, этот вид издательской деятельности просвещает общественность. Чтобы журналисты могли ответственно публиковать данные, Сондерман (2013) предлагает журналистам спросить самих себя о необходимости что-то публиковать, о причинах отказа от публикации и о том, как лучше всего опубликовать.

При рассмотрении вопроса о том, следует ли публиковать, журналисты должны иметь четкое представление о том, чего они пытаются достичь, публикуя данные: какое влияние они намерены оказать, действительно ли данные представляют ценность для читателей и связаны ли они с другими элементами отчетности. Журналисты отбрасывают данные, когда не могут придумать более веской причины, чем "потому что мы можем" или "потому что мы думаем, что это будет выглядеть круто".

Учитывая причину, по которой что-то не публикуется, журналистам следует потратить некоторое время на размышления о вероятных проблемах, которые могут возникнуть в результате публикации определенного набора данных. Вопросы о пострадавших от публикации особенно важны, если набор данных включает информацию о конкретных лицах.

Еще одна проблема связана с точностью данных. Даже если данные поступают из государственного источника, есть вероятность, что они неточные. Любая база данных на самом деле не является базой данных, а набором данных, содержащих последнюю известную записанную информацию, которая может быть устаревшей, неточно записанной или неточно предоставленной или не содержать полную информацию. Следовательно, представление данных в истории может вводить в заблуждение.

Третья проблема связана с тем, имеют ли данные отношение к вашей истории. Журналисты должны спросить себя, добавили ли они достаточно контекста, представляют ли они данные и как читатель должен их интерпретировать.

Наконец, журналисты должны решить, как представить данные таким образом, чтобы максимизировать выгоды и свести к минимуму вред. Вопросы заключаются в том, какие аспекты данных действительно важны, а какие журналист может ограничить или отредактировать. Журналистам, пишущим статьи, часто приходится решать, использовать ли цитату дословно или перефразировать ее. То же самое относится и к представлению данных. Журналисты могут манипулировать необработанными исходными данными для повышения ясности, контекста или других принципов. Например, в США Бюро переписи населения, собирая, анализируя и отображая всевозможные персональные данные в сводных таблицах или картах, никогда не идентифицирует ни одного человека.

В эпоху мгновенной информации журналисты несут этическое обязательство представлять данные четко и в контексте. Поскольку простые диаграммы и инфографика предлагают только моментальный снимок данных, соответствующая справочная информация и любые руководства по интерпретации информации имеют важное значение. Это правда, что визуализация истории вызывает свои собственные этические проблемы. Согласно Каиру (2014), визуализация данных является аргументом и доказательством точности информации в истории, поскольку тогда люди менее скептически относятся к новой информации. Визуализация создает у общественности впечатление, что информация в фоновом режиме является достоверной. Чтобы избежать подобных этических проблем, визуализация и данные должны иметь четкую связь друг с другом. Кроме того, журналисты должны знать, какой вид визуализации наиболее подходит для каких видов

презентаций. Неэтичная визуализация может вызвать путаницу среди общественности и привести к ложным интерпретациям. Легко заманить публику прочесть историю с хорошим примером визуализации, но не ценой ясности.

Брэдшоу (2013) предполагает, что визуализация данных должна быть четкой. Например, диаграммы и таблицы, как правило, должны иметь нулевую базовую линию, чтобы избежать искажения изменений как более серьезных, чем они есть на самом деле. Журналистам следует выбирать временные рамки для представления долгосрочных тенденций, а не искажать результаты, начиная или выбирая с рекордно низкого или высокого уровня. Portelance (2014) критикует визуализацию данных за то, что она вводит в заблуждение недостаточными точками данных и усеченным графиком, из-за которого в некоррелированных данных появляются поверхностные тенденции<sup>207</sup>.

Как источники, так и журналисты могут манипулировать визуальными представлениями данных. Базовые показатели, которые не начинаются с нуля, могут быть особенно обманчивыми. Линейные графики, начинающиеся с самой низкой или самой высокой точки, могут предполагать гораздо большее падение или подъем, чем долгосрочная реальность. Использование 3D-эффектов может активно исказить пропорции в диаграмме и добавлять к ней бессмысленный шум. Круговая диаграмма, удаляющаяся вдаль, страдает от той же проблемы, что и все, что находится на расстоянии: срез, который находится дальше, меньше, чем тот же срез на переднем плане.

При публикации истории данных важно помнить, что данные - это всего лишь объект для анализа, и для завершения истории требуются дополнительные источники. Во многих ситуациях дополнительные источники могут выявить лазейки в анализе и показать в истории важные моменты сравнения с реальностью. Однако данные, собранные от общественности, могут привести к ложным результатам из-за возможности сфабриковать ответы и проблемы с идентификацией респондентов.

Например, Лефлер (2014) утверждает, что анализ данных - это только начало журналистского процесса. Репортеры должны спросить себя, согласуется ли их анализ с выводами. Они должны публиковать подробную методологию, касающуюся данных и журналистского процесса. Кроме того, важно привлекать отзывы и исправления общественности<sup>208</sup>.

Выводы о результатах должны открыто обсуждаться среди коллег в редакциях, журналистам рекомендуется искать сторонних экспертов для оценки результатов. Получение ответа на данные перед публикацией является жизненно важным шагом в проверке их точности.

Таким образом, журналисты не могут отказаться от своих этических обязанностей в отношении грязных данных и должны сделать все, чтобы очистить их перед анализом. В некоторых незначительных нерешенных проблемах, связанных с грязными данными, может быть приемлемо предоставить общественности возможность с осторожностью относиться к результатам анализа. Недопустимо, чтобы журналист пытался решить редакционные проблемы с грязными данными, просто публикуя данные и информируя людей о личных потребностях каждого, чтобы решить, насколько достоверной может быть история.

Ожидается, что в будущем частота применения журналистики данных значительно возрастет. Следовательно, ожидается, что ее этические проблемы станут более распространенными в

---

<sup>207</sup> Portelance, Eric (2014). Misleading with statistics. How journalists make arguments with distorted data. Medium.com [Blog] Available at <https://medium.com/i-data/misleading-withstatistics-c63780efa928> [Accessed 10 November, 2017]

<sup>208</sup> LaFleur, Jennifer (2014). Data journalism tips from data driven ethics. Conference at Markkula Center for Applied Ethics [Blog] 1 January, 2014. Available at <https://www.scu.edu/ethics/focus-areas/journalism-ethics/resources/data-journalism-tips-/> [Accessed 20 November, 2017]

отношении журналистской практики, и поэтому им необходимо уделять особое внимание, чтобы их избегать.

Этические проблемы в журналистике данных могут быть более заметными, чем в традиционной журналистике. Поэтому доверие общественности к журналистике данных и журналистам нуждается в большей прозрачности журналистской деятельности.

#### **Контрольные вопросы:**

1. Является ли зависимость журналистики данных от самих данных также ее слабостью?
2. Какие ограничения существуют в доступности данных, их надежности, качества и возможности манипулирования ими?
3. Согласно финскому исследованию (Kuutti 2011), какие проблемы были выявлены, связанные с поведением государственных учреждений в ответ на запросы данных?
4. По словам Брэдшоу (2013), что является самым основным этическим соображением в журналистике данных?
5. За что О'Нейл (2013) критикует журналистов?
6. Согласно Хьюстону (2004), журналистам необходимо играть роль адвоката дьявола в работе с данными. Что это означает?
7. Какая этическая ответственность существует при работе с необработанными данными?
8. Согласно Брэдшоу (2013) каким должна быть визуализация данных?
9. За что Portelance (2014) критикует визуализацию данных?
10. Что при важно помнить публикации истории данных?

#### **Задание:**

1. Что, по-вашему, ответственная журналистика данных? Напишите эссе (минимум 500 слов)
2. Этические проблемы в журналистике данных могут быть более заметными, чем в традиционной журналистике. Объясните почему (минимум 300 слов).
3. Доверие общественности к журналистике данных и журналистам нуждается в большей прозрачности журналистской деятельности. Объясните, почему это так? (минимум 500 слов)
4. Получение ответа на данные перед публикацией является жизненно важным шагом в проверке их точности. Объясните почему это так? (минимум 300 слов)
5. При публикации истории данных важно помнить, что данные - это всего лишь объект для анализа, и для завершения истории требуются дополнительные источники. Объясните, почему? (минимум 300 слов)

#### **Список источников:**

1. <https://texterra.ru/blog/galochku-vam-v-akkaunt-kak-proyti-verifikatsiyu-v-sotssetyakh.html>
2. <https://denezhnye-rucheyki.ru/verifikatsiya-akkaunta/>
3. <https://www.youtube.com/watch?v=7eKG9RuqUE4>
4. <https://anri.org.ru/2020/05/29/instrumenty-i-prilozheniya-dlja-zhurnalista-v-socsetjah-poisk-po-geolokacii-svjazi-proverka-foto-i-video/>

## **ЗАКЛЮЧЕНИЕ**

Настоящий курс по журналистике данных является первой попыткой регионального сотрудничества преподавателей журналистики 19 университетов Центральноазиатского региона под эгидой ЮНЕСКО. Собранные и представленные в нём материалы теоретического и

практического характера подтверждают, что журналистика данных выступает важным инструментом обеспечения открытости деятельности органов государственной власти и управления, вовлечения в процессы дальнейшей демократизации общества всех слоев населения, включая национальные меньшинства и уязвимые группы. Такие усилия преподавателей факультетов журналистики университетов региона нацелены на подготовку журналистов и других специалистов высшей квалификации в информационной сфере, способных обеспечить посредством дата-проектов и расследований в странах Центральной Азии максимальный уровень прозрачности деятельности органов власти и их подотчетности перед гражданами.

Журналистика данных выступает одним из весомых факторов обеспечения гендерного равенства. Она позволяет посредством данных представлять аудитории многомерную картину действительности, оставляя за аудиторией свободу интерактивного взаимодействия с контентом. В силу этого дата-журналисты призваны осваивать навыки работы с открытыми данными, а также с их программированием, статистическим анализом и дизайном. Дата-журналистам необходимо знать о том, как устроены основные инструменты дата-аналитики, как проектировать разного рода исследования, как собирать и обрабатывать данные, какие механизмы и подходы можно применить к очистке и оценке их достоверности. Кроме того, дата-журналист в силу специфики своей работы призван следить за обновлением самих инструментов анализа данных, а значит, непрерывно повышать и собственный уровень медиаграмотности и медикультуры.

Важным достижением СМИ и журналистов Центральноазиатского региона является понимание необходимости активного развития журналистика данных как на локальном, национальном, так и на региональном уровне. Для этого необходимо расширять рамки использования дата-аналитики и расследований, шире представлять публикации, подготовленные на основе совместных проектов. Важно расширять спектр дата-публикаций на языках народов региона, активнее представлять материалы по гендерно маркированным темам, по вопросам расширения участия в общественной жизни национальных меньшинств, уязвимых слоев населения. А самое главное, - совместными усилиями представлять в глобальном информационном пространстве многообразную и богатую жизнь народов региона, устремленных жить в мире, безопасности и большом уважении ко всем языкам и культурам.

По итогам совместной работы команды преподавателей журналистики университетов региона Центральной Азии над курсом по журналистике данных разработаны нижеследующие рекомендации.

## **РЕКОМЕНДАЦИИ:**

### **Для ВУЗов:**

- Обучение преподавателей дата-журналистике.
- Разработка учебных планов, учебных программ, учебно-методических пособий, руководств и книг по дата-журналистике.
- Обеспечить техническую инфраструктуру для обучения по дата-журналистике
- Улучшить сотрудничество с практикующими дата-журналистами и с сектором
- Открыть дата-лабы на факультетах журналистики.
- Находить редакции для студентов для прохождения профессиональной практики по дата-журналистике.
- Поддержка студенческих проектов по дата-журналистике
- Поддержка студенческих дипломных работ по дата-журналистике
- Проведение хакатонов, форумов, летних школ по дата-журналистике
- Проведение научных исследований по дата-журналистике
- Разработка видео-уроков и онлайн-курсов по дата-журналистике
- Поддерживать сотрудничество факультетов журналистики региона в формате конференций, форумов, фестивалей и медиа-лабов по дата-журналистике
- Поддерживать инициативу по приглашению ведущих профессоров и преподавателей региона для чтения лекций и обмена опытом по дата-журналистике в заинтересованных ВУЗах

### **Для органов государственной власти:**

- Обеспечить доступ к данным
- Обучение сотрудников государственных структур работе с данными
- Своевременно обновлять базы-данных
- Принимать решения на основе данных и материалов дата-журналистики
- Поддержка конструктивного диалога между государственными дата-базами, дата-центрами, дата-журналистами и факультетами журналистики с целью организации стажировки исследователей и практики студентов
- Поддержка и оказание практической помощи в реализации дата-журналистских инициатив в реализации приоритетных государственных программ и стратегий
- Поддержка организации и проведения национальных и региональных конкурсов по номинации “ Самая лучшая дата-журналистская публикация”.

### **Для медиа сектора**

- Обучение журналистов дата-журналистике
- Открытие дата-отделов в редакциях СМИ
- Повышение квалификации дата-журналистов
- Создать техническую инфраструктуру для работы дата-журналистов
- Проведение совместных дата-журналистских расследований с коллегами из других стран, а также с международными организациями по дата-журналистике
- Сотрудничество с факультетами журналистики
- Принимать студентов факультетов журналистики на практику
- Участие редакций СМИ в разработке требований к дата-журналистским компетенциям выпускников бакалавриата и магистратуры факультетов журналистики
- Активное участие журналистов и редакций СМИ в национальных, региональных и международных конкурсах по дата- журналистике
- Поощрение и поддержка дата-журналистов-расследователей

## ГЛОССАРИЙ

### Главы № 1-3.

**Агрегирование информации** – процесс сбора, обработки и представления информации в окончательном виде; преобразование детализированной информации в пакеты(агрегаты) данных, что позволяет анализировать состояние определенной отрасли жизнедеятельности человека.

**База данных (Database)** – это особым образом организованные и хранимые в электронном виде данные. Особым образом организованные означает, что данные организованы неким конкретным способом, способным облегчить их поиск и доступ к ним для одного или нескольких приложений. Также такая организация данных предусматривает наличие минимальной избыточности данных. Базы данных являются одной из разновидностей информационных технологий, а также формой хранения данных.

**Большие данные** – огромный объем структурированной или неструктурированной(бессистемной) информации, обрабатываемые программными инструментами.

**Визуализация** – общее обозначение приёмов и методов представления числовой информации в удобном для восприятия, наблюдения и анализа виде – рисунков, таблиц, диаграмм, графиков, схем, карт и т.д.

**Данные** - в широком понимании данные представляют собой факты, текст, графики, картинки, звуки, аналоговые или цифровые видео-сегменты. Данные могут быть получены в результате измерений, экспериментов, арифметических и логических операций. Данные должны быть представлены в форме, пригодной для хранения, передачи и обработки. Иными словами, данные — это необработанный материал, предоставляемый поставщиками данных и используемый потребителями для формирования информации на основе данных.

**Деагрегирование информации** – разукрупнение, детализация информации; процесс, обратный агрегированию.

**Журналистика данных** - направление в журналистике, в основе которого лежит сбор, анализ и обработка данных с целью создания медиаматериалов, представления актуальной и социально значимой информации в визуализированном виде

**Инфографика** – графический способ подачи информации, данных и знания; форма графического и коммуникационного дизайна, целью которой является донесение сложной информации до аудитории быстрым и понятным образом. Инфографика способна не только организовать большие объёмы информации, но и более наглядно показать соотношение предметов и фактов во времени и пространстве, а также продемонстрировать тенденции.

**Корреляция** – от позднелатинского *corelatio* – соотношение, взаимная связь, взаимозависимость, соотношение предметов и понятий. В математической статистике – вероятностная или статистическая зависимость, которая возникает тогда, когда зависимость одного из признаков от другого осложняется наличием ряда случайных факторов.

**Метаданные (Metadata)** — это данные о данных. В состав метаданных могут входить: каталоги, справочники, реестры. Метаданные содержат сведения о составе данных, содержании, статусе, происхождении, местонахождении, качестве, форматах и формах представления, условиях доступа, приобретения и использования, авторских, имущественных и смежных с ними правах на данные и др. Метаданные имеют важное значение в управлении хранилищем данных. Метаданные, применяемые при управлении хранилищем, содержат информацию, необходимую для его настройки и использования. Различают бизнес-метаданные и оперативные метаданные.

**МИГ** – Медийная и информационная грамотность

**Расследовательская журналистика** – вид журналистики, характеризующий планомерное и долговременное исследование предмета публикации. Она характеризуется тремя важными особенностями: 1) оно посвящено актуальной проблеме; 2) проводится по собственной инициативе журналиста и 3) как правило, предает публичной огласке информацию, которую преднамеренно скрывали определенные социальные силы общества.

**Фактоид** – недостоверное, ложное утверждение, которое облекается в форму достоверного и выдается за достоверное сообщение, факт.

**Образ художественный** - категория, обозначающая особый способ освоения и преобразования действительности, явление творчески воссозданное в художественном или журналистском произведении.

**Открытые данные** — это общедоступная информация, размещенная в интернете для неоднократного, свободного и бесплатного использования в виде машиночитаемых систематизированных данных, в формате, позволяющем их автоматизированную обработку.

**Поликодовость текста** – смешение и взаимодействие различных кодов, «упаковывается» по определенным принципам, известным адресату (кто её принимает) и адресанту (кто ее отправляет). Разновидность кодов зависит от канала СМИ, жанра сообщения. В поликодовом тексте интегрируются разнородные компоненты - визуальные, аудиовизуальные, вербальные и другие – в единое смысловое пространство. В традиционных СМИ поликодовость означала интегрирование вербального текста с фото, картинкой, рисунком. В интернет-СМИ поликодовость предполагает сложное текстовое образование, в котором вербальные и невербальные элементы образуют единое визуальное, структурное и функциональное целое, нацеленное на комплексное воздействие на аудиторию.

**«Перфомансность» поведения объекта коммуникации** – в процессе коммуникации исполнение определенной роли, демонстрация умений, поддержание определенного образа.

**Тренд** - цикличные изменения чего-либо, например времен года или числового ряда в математике. В широком смысле – основная тенденция изменения, направление развития чего-либо, например: в одежде, технологиях или дизайне. В ЖД тренд означает актуальные темы и проблемы, поставленные на повестку дня.

**Фемицид** - убийство женщины, совершаемое, как правило, мужчиной на почве ненависти к женщине, гендерной дискриминации, и/или в результате гендерного насилия при попустительстве государства. Фемицид возникает из идеи превосходства мужчины над женщиной — например, убежденности в наличии права собственности на нее. Зачастую фемицид возникает, когда мужчина заставляет женщину соответствовать стереотипным гендерным ролям - женщина должна выглядеть и вести себя так, как считает мужчина.

#### **Глава № 4.**

**Атрибуция** - установление первоисточника информации, авторства.

**Аудитория** - понятие, которое трактуется по-разному в свете функционального значения. В медиаметрии этот термин подразумевает некое множество потенциальных потребителей медиапродукта, рассматриваемое в заданных границах.

**Визуализация** - Представление данных в графической форме. В отличие от термина “инфографика”, который может относиться к любым формам графического отображения информации (в том числе, например, инструкций), визуализация подразумевает, в первую очередь, отображение численных данных. При работе с данными она используется двояко. Во-первых, при исследовании данных строятся черновые графики, позволяющие быстро заметить закономерности и потенциальные взаимосвязи. Во-вторых, представление результатов анализа данных зачастую требует визуального сопровождения, как наиболее эффективного способа продемонстрировать некоторые наблюдения.

**Данные** - Представление информации в формализованном виде, пригодном для передачи, интерпретации или обработки.

**Дата-журналист** - Журналист, специализирующийся на журналистике данных.

**Дата-экспедиция** — это неформальное общедоступное образовательное онлайн-мероприятие, строящееся на принципах проектного обучения и взаимопомощи. Под “проектным” имеется в виду, что итогом участия в идеале должно стать создание индивидуального или коллективного проекта.

**Журналистика данных** — направление в журналистике, в основе которого лежит обработка данных и их использование для создания журналистского материала. Данные могут служить как инструментом для раскрытия определенной журналистской истории, так и ее источником. Развитие журналистики данных связано с развитием технологий, позволяющих хранить и обрабатывать большие объемы данных, и движением в сторону большей открытости информации.

**Машиночитаемые данные** — это данные в формате, который может обрабатываться компьютером. Машиночитаемые данные должны быть структурированными. Машиночитаемые данные можно разделить на две группы: человеко-читаемые данные, которые размечены, чтобы они также могли быть прочитаны машинами (например, микроформаты, RDFa, HTML) и файл данных форматы, предназначенные в основном для машинной обработки (CSV, RDF, XML, JSON). Эти форматы машиночитаемы только в том случае, если содержащиеся в них данные формально структурированы.

**Открытые данные** – это одна из форм представления информации в машиночитаемом виде, когда информация представлена в формате, пригодном для загрузки в системы баз данных и может быть интегрирована с другими массивами данных или быть объектом анализа.

**Цифровая история** — раздел цифровых гуманитарных наук, изучающий использование компьютерных технологий и цифровых медиа для исторического анализа, исследований и визуализации данных. Два основных применения данного направления — знакомство интернет-аудитории с цифровыми архивами, интерактивными картами, хрониками событий и создание новых исследовательских инструментов для учёных-историков, помогающих им в развитии истории как науки.

**Компьютеризированная отчетность** - описывает использование компьютеров для сбора и анализа данных, необходимых для написания новостей.

## **Глава № 5.**

**Диаграмма** – графический способ представления числовых значений. Программа Excel поддерживает работу с разными диаграммами: гистограммы, круговые, пузырьковые, графики и др.

**Имя** – идентификатор, который предоставляет возможность ссылаться на какой-либо объект (ячейку, диапазон, формулу и т. д.).

**Константа** – постоянное (невычисляемое) значение. Константой может быть, например, фиксированное число либо какой-либо текст.

**Контекстное меню** – меню, содержащее список команд, которые предназначены для работы с конкретным объектом.

**Массив** – набор ячеек либо значений, используемых как единое целое. Другими словами, массив представляет собой группу элементов одного типа, объединенных под одним именем.

**Настройка** – изменение текущих параметров работы программы штатными средствами.

**Операнд** – элемент вычисления (константа, функция либо ссылка).

**Оператор** – знак либо символ, который определяет тип вычисления в формуле над операндами. В Excel используются следующие виды операторов: математические, текстовые, операторы сравнения и операторы ссылок.

**Панель инструментов** – панель, состоящая из кнопок и иных элементов управления, используемых для активизации команд и перехода в разные режимы работы.

**Пользовательский интерфейс** – средство взаимодействия пользователя с программой. Интерфейс включает в себя меню, панели инструментов, и др. В Excel применяется стандартный пользовательский интерфейс, используемый в приложениях MS Office.

**Примечание** – вспомогательная информация произвольного характера, относящаяся к определенной ячейке и хранящаяся независимо от содержимого этой ячейки.

**Рабочая книга** – файл, который создается, редактируется и сохраняется средствами Excel. Основной структурной единицей рабочей книги является рабочий лист (см. ниже).

**Рабочий лист** – основной элемент рабочей книги, предназначенный для ввода, редактирования и хранения данных, а также для выполнения расчетов. По умолчанию в состав рабочей книги включено три рабочих листа. Основной структурной единицей рабочего листа является ячейка.

**Ссылка** – адрес ячейки или диапазона ячеек.

**Форматирование** – изменение оформления ячейки (ее «внешнего вида» – границ, заливки, и др.) либо изменение представления данных, содержащихся в ячейке. Параметры форматирования ячейки независимы от ее содержимого, и наоборот.

**Формула** – специальный инструмент Excel, предназначенный для расчетов, вычислений и анализа данных. Формула может включать в себя константу, оператор, ссылку, имя ячейки (диапазона) и функцию.

**Арифметический оператор** – оператор, предназначенный для выполнения арифметических действий и выдающий в качестве результата числовое значение;

**Оператор сравнения** – данный оператор используется для сравнения данных и выдает в качестве результата логическое значение ИСТИНА или ЛОЖЬ;

**Текстовый оператор** – оператор, используемый для объединения данных.

**Функция** – готовая штатная формула Excel. Каждая функция может включать в себя константу, оператор, ссылку, имя ячейки (диапазона) и формулу (см. выше).

**Пользовательская функция** – это функция, написанная пользователем на языке VBA.

**Электронная таблица** – интерактивная программа, состоящая из набора строк и столбцов, которые выводятся на экран в отдельном окне.

**Ячейка** – наименьшая (элементарная) часть рабочего листа (электронной таблицы), предназначенная для ввода и хранения информации. Каждая ячейка может содержать текст, число или формулу.

**Адрес** – это месторасположение (координаты) ячейки; адрес состоит из буквы (номера) столбца и номера строки, на пересечении которых расположена данная ячейка;

**Ссылка** – указание на адрес ячейки. Ссылки могут быть абсолютными (т. е. не изменяющимися при перемещении и копировании ячейки), относительными (эти ссылки изменяются при перемещении и копировании ячейки) и смешанными.

**Внешняя ссылка** – это ссылка на ячейку, расположенную в другой рабочей книге.

**Круговой график** – диаграмма, показывающая частотное распределение категориальной переменной. Круг делится на сегменты, площадь каждого пропорциональна частоте категории, к которой он относится.

**Линейный график** – линейный график, который упорядочивает данные по реальной линии. Также называется точечным графиком.

**Гистограмма** – диаграмма, представляющая распределение частот значений переменной (или частот значений на каждом из интервалов, на которые разбита выбранная область изменения переменной). Огибающая гистограммы показывает форму функции плотности распределения.

**Минимальное значение** – наименьшее значение в выборке.

## Главы № 6-11.

**Арифметическое среднее** - мера положения, полученная делением суммы значений переменной по наблюдениям на число слагаемых, часто называемая просто средним.

**Алгоритм** - упорядоченный набор действий (операций, процедур), которые приводят к достижению заранее поставленной цели. Например, алгоритм Евклида указывает, как найти наибольший общий делитель (НОД) двух натуральных чисел  $a$  и  $b$ .

**База данных (Database)**. Для данных необходим особый способ хранения и обработки, чтобы они могли трансформироваться в информацию и далее использоваться для каких-либо полезных выводов. Базы данных обычно содержат совокупности записей данных или файлов, таких как последовательность производственных действий, транзакции, каталоги продуктов, запасы, профили клиентов и т.д. Данные обновляются, расширяются и удаляются по мере добавления новой информации. Данные организованы в строки, столбцы, таблицы, которые индексируются, чтобы упростить поиск необходимой информации.

**Python** — язык программирования, доступный с 1994 года, популярный среди исследователей, занимающихся наукой о данных. Python отличается простотой использования среди новичков и большой мощностью при использовании опытными пользователями, особенно когда используются преимущества специализированных библиотек, таких как библиотеки, предназначенные для машинного обучения и генерации графиков.

**Переменная** - в языках программирования - именованная часть памяти, в которую могут помещаться разные значения переменной. Причем в каждый момент времени переменная имеет единственное значение. В процессе выполнения программы значение переменной может изменяться. Тип переменных определяется типом данных, которые они представляют.

**Данные** — поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи или обработки информационными системами и их оператором.

**Тип данных** - характеристика набора данных, которая определяет: диапазон возможных значений данных из набора; допустимые операции, которые можно выполнять над этими значениями; способ хранения этих значений в памяти.

**Дата фреймы (data frame)** - прямоугольная двумерная таблица с данными (Excel spreadsheet, SQL-таблица). По своей сути дата фреймы — это стандартный способ хранения данных в формате “наблюдения/переменные”: строки соответствуют наблюдениям, а столбцы - изучаемым переменным.

**Библиотека** в программировании — сборник подпрограмм или объектов, используемых для разработки программного обеспечения (ПО).

**Модуль** - Термин Python, обозначающий коллекции функций и методов. В общем случае представляет собой файл с расширением.py

**Пакет** - Термин Python, набор модулей. В простом случае представляет собой папку с файлами модулей.

**Интерфейс** — это компонент языка программирования, предназначенный для обеспечения взаимодействия между его компонентами.

**Pandas** — библиотека Python для манипулирования данными, популярная среди исследователей данных.

**Matplotlib** — библиотека на языке программирования Python для визуализации данных двумерной графикой. Получаемые изображения могут быть использованы в качестве иллюстраций в публикациях.

**Веб-скрейпинг** — это технология получения веб-данных путем извлечения их со страниц веб-ресурсов. Веб-скрейпинг может быть сделан вручную пользователем компьютера, однако термин

обычно относится к автоматизированным процессам, реализованным с помощью кода, который выполняет GET-запросы на целевой сайт.

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Барлыбаева С.Х., Малтабаров М.А. Дата-журналистика в Казахстане как медиатренд. Хабаршы. Журналистика сериясы. №1 (55) 2020.
2. Бозрикова С.А., Татару Л.В. Нарративная журналистика в Америке и в России
3. Вайгенд А. Big Data. Вся технология в одной книге. Пер. с англ. С. Богданова.- М.: Эксмо, 2018.
4. Иванов А. Д. Роботизированная журналистика как следующий этап развития дата-журналистики // Журналистика в системе альтернативных источников информации: сб. материалов науч. конф, Н. Новгород, 2017. С. 113–119
5. Климко А.Ю. Портретный очерк в формате лонгрида
6. Кононов Н. В. Автор, ножницы, бумага. - Москва.- Издательство Манн, Иванов и Фербер, 2021.
7. МИГ в журналистике, ЮНЕСКО, Ташкент, 2019
8. Мэннинг Н. Парисон Н. Реформа государственного управления: международный опыт. М., 2003.
9. Распопова С.С. Автор мультимедийного текста // Вестник Челябинского государственного университета. 2012. № 13 (267). Филология. Искусствоведение. Вып. 65. С. 100–102.
10. Руководство по дата-журналистике «Ищем сокровища в данных», Анастасия Валеева Алтынай Мамбетова, Чолпон Узакбаева, Бишкек, 2021
11. Чигаев Д.П. Лонгрид как разновидность креолизованного текста // Медиаскоп. 2017. Вып. 1.
12. Чурбакова Е. А. Журналистика баз данных как медиатренд // Медиа-исследования, 2017. № 4–2. С. 217–221.
13. Шилина М. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов. // Медиаскоп, 2013, №1/[www.mediascope.ru/](http://www.mediascope.ru/)20.07.2021.
14. Шевченко В. Э. Визуальный контент как тенденция современной журналистики // Медиаскоп. 2014. №4. С. 6-8.

## Электронные источники на русском языке:

1. Акобиров Ш. Проблемы освещения экономических проблем в современной таджикской прессе. – Душанбе: International Alert 2016.// <http://www.khoma.tj/khoma/wp-content/uploads/2018/09/Business-journalism-in-Tajikistan-RUS.pdf>
2. Алма-Атинская Декларация о содействии развитию независимых и плюралистических средств информации в Азии. (1992,5стр.) // [https://www.un.org/ru/documents/decl\\_conv/declarations/pdf/alma-ata](https://www.un.org/ru/documents/decl_conv/declarations/pdf/alma-ata).
3. Analiticheskij-obzor.-ZHurnalistika-dannyh-v-KR.pdf (soros.kg)
4. Асанбаева Г. Основные понятия и принципы МИГ // Учебное пособие по медиаграмотности, 23.10.2018. Точный адрес: <https://newreporter.org/uchebnoe-posobie-po-mediagramotnosti/>
5. Асанбаева Г. Факторы влияния на уровень медиаграмотности в странах Центральной Азии // Медийно-цифровая грамотность в цифровом мире: как научить учителей. - ЮНЕСКО, 2021.

Ссылка на электронный ресурс: [https://iite.unesco.org/wp-content/uploads/2021/05/MIL\\_Teachers\\_book\\_2021.pdf](https://iite.unesco.org/wp-content/uploads/2021/05/MIL_Teachers_book_2021.pdf)

6. Архангельская И. Б., Архангельская А. С. Фейк-ньюс в цифровую и доцифровую эпохи // Знак: проблемное поле медиаобразования, 2020. URL: <https://cyberleninka.ru/article/n/feyk-nyus-v-dotsifrovuyu-i-tsifrovuyu-epohi/viewer>

Архивы дата-журналистика | KLOOP.KG - Новости Кыргызстана

7. Бешлей О. Чипирование, аутизм и заговоры. Что такое инфодемия и как дезинформация вредит людям. // Настоящее Время, 25.07.2021. Точный адрес: <https://www.currenttime.tv/a/chto-takoe-infodemia-i-kak-ona-vredit-ludyam/31374201.html>

8. Богданова Д. Социальные сети как источник глобального риска. Как его нейтрализовать. // Школьные технологии, 2020. Точный адрес статьи: <https://cyberleninka.ru/article/n/sotsialnye-seti-kak-istochnik-globalnogo-riska-kak-ego-neytralizovat>

9. Борис Лозовский Журналистика и средства массовой информации: краткий словарь [https://elar.urfu.ru/bitstream/10995/48851/1/Lozovsky\\_2007.pdf](https://elar.urfu.ru/bitstream/10995/48851/1/Lozovsky_2007.pdf)

10. Буров В., Орлова Кс. Дата-журналистика и открытые данные в странах Центральной Азии. Скрытые возможности симбиоза данных, журналистики и технологий искусственного интеллекта для развития регионального потенциала. - [https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism\\_web.pdf](https://www.infoculture.ru/wp-content/uploads/2021/06/Opendata-journalism_web.pdf).

11. В Великобритании жгут вышки 5G из-за их якобы связи с распространением коронавируса. // Коммерсантъ, 07.04.2020. Точный адрес статьи: <https://www.kommersant.ru/doc/4316217>

12. В Таджикистане запретили распространять ложную информацию об эпидемиях. // Интерфакс-Азербайджан, 08.07.2020. Точный адрес статьи: <http://interfax.az/view/807102>

13. Гайсина, 2019. Дата-журналистика в Центральной Азии: где копают глубже | Новый репортер ([newreporter.org](http://newreporter.org))

14. Гибельхауз Л. Таджикистан: СМИ становятся все более независимыми. Электронный сайт «НЕМЕЦКАЯ ВОЛНА» рубрика: ПОЛИТИКА. МИР (11.06.2003, 2р.) // <https://www.dw.com/ru/%D1%82%>

15. Горина Е. В. Коммуникативные технологии манипуляции в СМИ и вопросы информационной безопасности. Учебное пособие. - Екатеринбург.: Издательство Уральского университета, 2016. Точный адрес учебника: [https://elar.urfu.ru/bitstream/10995/42384/1/978-5-7996-1807-0\\_2016.pdf](https://elar.urfu.ru/bitstream/10995/42384/1/978-5-7996-1807-0_2016.pdf)

16. Дата-журналистика: «Получается, что журналист не нужен»./ <https://polit.ru/article/2020/09/22/begin/>.

17. Дата-журналистика в Центральной Азии: где копают глубже.// <https://newreporter.org/2019/04/18/data-zhurnalistika-v-centralnoj-azii-gde-kopayut-glubzhe/>.

18. Дезинформация // Википедия, последняя дата редакция статьи 09.09.2021. Точный адрес: <https://en.wikipedia.org/wiki/Disinformation>

19. Журналистика данных: инновация или пустая трата времени и сил? <https://test.ru/2018/05/30/data-journalism-innovation-or-a-waste-of-time/>

20. Журналистика, «фейковые новости» и дезинформация: Руководство для академической и профессиональной подготовки журналистов <https://unesdoc.unesco.org/ark:/48223/pf0000371785/PDF/371785rus.pdf.multi.page=127>

21. За распространение фейков в Узбекистане будут штрафовать и ограничивать свободу. // Международное агентство новостей «Фергана», 16.12.20. Точный адрес статьи: <https://fergana.agency/news/121493/>

22. Злобин А. Пробирка ценой \$7 трлн. Трамп назвал худшую ошибку в истории США. // Форбс, 20.09.2018. Точный адрес статьи: <https://www.forbes.ru/biznes/367071-probirka-cenoy-v-7-trln-tramp-nazval-hudshuyu-oshibku-v-istorii-ssha>

23. Как медиаграмотность способствует свободе слова? Мы спросили у шести экспертов из Казахстана, Таджикистана и Узбекистана // Stan Radar, 07.05.2021. Точный адрес статьи: <https://stanradar.com/news/full/44584-kak-mediagramotnost-sposobstvuet-svobode-slova-my-sprosil-u-shesti-ekspertov-iz-kazahstana-tadzhikistana-i-uzbekistana.html>

24. «Как написать статью?» Лариса Павловна Нода <https://newreporter.org/2018/12/12/kak-napisat-statyu/>

25. «Как создать текст для печати» <http://www.oshibok-net.ru/for-journalists/journalistic-text/jur-text-pechat/>

26. Качкаева А. Г., Шомова С. А., Душакова И. С. Фейк: от академических дискуссий к практическим решениям // Коммуникации. Медиа. Дизайн, 6(2), 5-32, 2021. URL: <https://cmd-journal.hse.ru/article/view/12600>

27. Красовская Н. Р., Гуляев А. А., Юлина Г. Н. Фейковые новости как феномен современности // Власть. 2019. URL: <https://cyberleninka.ru/article/n/fejkovye-novosti-kak-fenomen-sovremennosti/viewer>

28. Колесниченко Александр Васильевич, Long Texts (Longreads) in the Modern Russian Press электронный научный журнал «Медиаскоп», выпуск №1, 2015. (<http://mediascope7.mediascope.ru/?q=node/1691>)

29. Коммьюнити-журналистика: цифровые и социальные медиа <https://www.futurelearn.com/courses/community-journalism>

30. Макарова Н.Я. Журналистика данных в системе профессиональных компетенций журналиста // Знак: проблемное поле mediaobrazovaniya. 2020. № 4 (38).

31. Манаева Н. В Иране более 700 человек умерли из-за употребления метанола как "лекарства" от COVID-19. // Интерфакс, 27.04.2020. Точный адрес статьи: <https://www.interfax.ru/world/706332>

32. Мамырханова М. За распространение фейков могут усилить ответственность в Казахстане. // Курсив, 23.08.2019. Точный адрес статьи: <https://kursiv.kz/news/obschestvo/2019-08/za-rasprostranenie-fejkov-mogut-usilit-otvetstvennost-v-kazakhstane>

33. Методы сбора информации: в журналистике и репутинге <https://reputation.moscow/journal/metody-sbora-informatsii/>

34. МИГ в журналистике, ЮНЕСКО, Ташкент, 2019 <https://unesdoc.unesco.org/ark:/48223/pf0000374922/PDF/374922rus.pdf.multi.page=51>

35. Модель учебной программы по журналистике: Сборник новых учебных планов, ЮНЕСКО, 2015 [https://unesdoc.unesco.org/ark:/48223/pf0000221199\\_rus/PDF/221199rus.pdf.multi.page=49](https://unesdoc.unesco.org/ark:/48223/pf0000221199_rus/PDF/221199rus.pdf.multi.page=49)

36. Назаров А. «Фейковые» новости в современном мире - оружие массового психоза? // Центр-1. 17.01.2018. URL: <https://centre1.com/kyrgyzstan/fejkovye-novosti-v-sovremennom-mire-oruzhie-massovogo-psihoza/>

37. Нисневич Ю.А. Открытость, прозрачность и подотчетность государственного управления. (2010, 20с.). // <https://www.hse.ru/data/2010/11/09/1223230150/pdf>.

38. Об утверждении перечня открытых данных, размещаемых на интернет-портале открытых данных // <https://adilet.zan.kz/rus/docs/V1900019076>.

39. Розиев И. Президент Жапаров подписал скандальный закон о фейках. Что это значит? // Kloop.kg, 23.08.2021. Точный адрес статьи: <https://kloop.kg/blog/2021/08/23/prezident-zhaparov-podpisal-skandalnyj-zakon-o-fejkah-chto-eto-znachit/>

40. Руководство по дата-журналистике «Ищем сокровища в данных», Анастасия Валеева Алтынай Мамбетова, Чолпон Узакбаева, Бишкек, 2021

41. Сегизова Б. О доступности и гласности судебной системы // <https://medialaw.asia/node/9512>.

42. Смолин А. Уотергейтский скандал как школа транспарентности. Сайт Российского агентства правовой и судебной информации (РАПСИ). [//http://rapsinews.ru/international\\_publication/20110813/](http://rapsinews.ru/international_publication/20110813/)

43. Современные технологии для оперативного доступа к информации: «Открытое правительство», электронные сервисы обращений, блог- платформа руководителей государственных органов и регионов [// http://info-dostup.kz/lesson10](http://info-dostup.kz/lesson10).

44. Суда Б. Дизайн для данных // Пособие по журналистике данных / Под ред. Д. Грея. URL: <https://unotices.com/book.php?id=125825>

45. Телеканал Дождь. Вакцина Гейтса. Откуда взялась теория заговора, за выпуск о которой программу Михалкова сняли с эфир. // YouTube, Телеканал Дождь от 04.05.2020. Точный адрес видео: <https://www.youtube.com/watch?v=OA5yxXfUIkU>

46. Твиттер удалил ретвит Трампа о существовании лекарства от коронавируса. // РБК, 29.06.2020. Точный адрес статьи: <https://www.rbc.ru/society/29/07/2020/5f20b8fb9a79470416a1dafa>

47. Тилеухан Ж. COVID-19 и фейки | В каких случаях грозят штрафы и уголовная ответственность? // Factcheck.kz, 17.03.2020. Точный адрес статьи: <https://factcheck.kz/health/covid-19-i-fejki-v-kakix-sluchayah-grozyat-shtrafy-i-ugolovnaya-otvetstvennost/>

48. Что такое открытые данные? [//https://opendatahandbook.org/guide/ru/what-is-open-data](https://opendatahandbook.org/guide/ru/what-is-open-data); The Data Journalism Handbook. Edited by Jonathan Gray, Liliana Bounegru, and Lucy Chambers. (2012). 238p. [// https://datajournalismhandbook.org/uploads/first\\_book/](https://datajournalismhandbook.org/uploads/first_book/)

49. Федоров А. В., Левицкая А. А., Новиков А. С. Коронавирус как источник медийных манипуляций. // Crede Experto, 2020. Точный адрес статьи: <https://cyberleninka.ru/article/n/koronavirus-kak-istochnik-mediynyh-manipulyatsiy/viewer>

50. Учебное пособие по дата-журналистике | Новый репортер (newreporter.org)

51. Promotank, 2017. Аналитический обзор. Есть ли журналистика данных в КР?, Analiticheskij-obzor.-Zhurnalistika-dannyh-v-KR.pdf (soros.kg)

52. Facebook удалил публикацию Садыра Жапарова о лечении коронавируса аконитом. // Радио Азаттык, 19.04.2021. Точный адрес статьи: <https://rus.azattyk.org/a/facebook-udalil-post-sadyra-zhaparova-o-lechenii-koronavirusa-issyk-kulskim-kornem/31211192.html>

53. Wilson C. et all. Медийная и информационная грамотность: программа обучения педагогов. - Париж: UNESCO, 2012. Ссылка на электронный ресурс: [https://unesdoc.unesco.org/ark:/48223/pf0000192971\\_rus](https://unesdoc.unesco.org/ark:/48223/pf0000192971_rus)

### **Электронные источники на английском языке:**

1. Bradshaw, Paul (2016). Data journalism. ONA Ethics. The Ethics & Excellence in Journalism Foundation. Online News Association [Blog] Available at <https://ethics.journalists.org/topics/data-journalism/> [Accessed 26 November, 2017]

2. Bothwell E. Fake news laws may 'catch on' during coronavirus. // The World University Rankings, 06.04.2020. Точный адрес статьи: <https://www.timeshighereducation.com/news/fake-news-laws-may-catch-during-coronavirus>

3. Buzenberg, W. E. (2015, July 6). Anatomy of a global investigation: Collaborative, data-driven, without borders. Shorenstein Center. <https://shorensteincenter.org/anatomy-of-a-global-investigation-william-buzenberg/>

4. Byers, D. (2014, March 19). Knives out for Nate Silver. Politico. <https://www.politico.com/blogs/media/2014/03/knives-out-for-nate-silver-185394.html>

5. Cairo, Alberto (2014). Ethical infographics. IRE Journal, Spring 2014: 25-27.

6. Coronel, S. (2016, June 20). Coronel: A golden age of global muckraking at hand. Global Investigative Journalism Network. <https://gijn.org/2016/06/20/a-golden-age-of-global-muckraking/>
7. Data Journalism: Beginner Excel <https://mjwebster.github.io/DataJ/BeginnerExcel.html>
8. Davies, W. (2016, August 24). The age of post-truth politics. The New York Times. <https://www.nytimes.com/2016/08/24/opinion/campaign-stops/the-age-of-post-truth-politics.htm>
9. Dijck, J. van, Poell, T., & Waal, M. de. (2018). The platform society: Public values in a connective world. Oxford University Press
10. Kuutti, Heikki (2011). Julkisuusjournalismi. Jyväskylän yliopisto [Access to information in journalism] University of Jyväskylä
11. Kuutti, H (2016). Ethics of data journalism Four ethical phases in the working process, a paper presented at Nordic Data Journalism Conference NODA 2016, 21 April 2016 in Helsinki, Finland
12. LaFleur, Jennifer (2014). Data journalism tips from data driven ethics. Conference at Markkula Center for Applied Ethics {Blog} 1 January, 2014. Available at <https://www.scu.edu/ethics/focus-areas/journalism-ethics/resources/data-journalism-tips-/> [Accessed 20 November, 2017]
13. Morozov, E. (2016, April 24). Tech titans are busy privatising our data. The Guardian. <https://www.theguardian.com/commentisfree/2016/apr/24/the-new-feudalism-silicon-valley-overlords-advertising-necessary-evil>
14. O'Neil, Cathy (2013). On Being a Data Skeptic. O'Reilly Media Available at <http://www.oreilly.com/data/free/being-a-data-skeptic.csp> [Accessed 25 November, 2017]
15. Otieno, Samuel (2013). Numbers don't lie: my experience with data journalism. Data Journalism Fellowship Blog [Blog] Data Dredger. Available at <http://internnewskenya.org/fellowshipblog/numbers-dont-lie-my-experience-with-datajournalism/#sthash.dkVisHqo.dpbs> [Accessed 25 November, 2017]
16. Portelance, Eric (2014). Misleading with statistics. How journalists make arguments with distorted data. Medium.com [Blog] Available at <https://medium.com/i-data/misleading-with-statistics-c63780efa928> [Accessed 10 November, 2017]
17. Raab, B. (2016, April 8). Behind the Panama Papers: A Q&A with International Consortium of Investigative Journalists director Gerard Ryle. Ford Foundation. <https://www.fordfoundation.org/ideas/equal-change-blog/posts/behind-the-panama-papers-a-qa-with-international-consortium-of-investigative-journalists-director-gerard-ryle/>
18. Rahman, Z. and Wehrmeyer, S. (2021) Searchable Databases as a Journalistic Product, in a book "The Data Journalism Handbook: Towards A Critical Data Practice" edited by Bounegru, L. and Gray, J. Amsterdam University Press. <https://www.jstor.org/stable/j.ctv1qr6smr.31>
19. Samantha Sunne. "The challenges and possible pitfalls of data journalism, and how you can you avoid them" American Press Institute. n.p., 9.03.2016. Web. 20.09.2021. URL: <https://www.americanpressinstitute.org/publications/reports/strategy-studies/challenges-data-journalism/>
20. Segnini, Giannina (2015). Investigating with databases: Verifying data quality. Craig Silverman (ed.): Verification handbook. [online] European Journalism Center. Available at <https://gijn.org/2015/05/14/investigating-with-databases-verifying-data-quality/> [Accessed 25 November, 2017]
21. Sonderman, Jeff (2013). Programmers explain how to turn data into journalism & why that matters. Poynter institute. Available at <https://www.poynter.org/news/programmers-explain-how-turn-data-journalism-why-matters> [Accessed 25 November, 2017]
22. The Data Journalism Handbook -1. European Journalism Centre. (2012, 191p.)// <https://s3-eu-central-1.amazonaws.com/datajournalismcom/handbooks/The-Data-Journalism-Handbook-1.pdf>.
23. Vallance- Jones Fred, David McKie. The Data Journalist Getting. The Story- Oxford. 2017, 320p.//<https://services.su.nu/galway.ie/bookstore/book/4526/>

24.Vervoska M. Data journalism tips: Eastern Europe, central Asia and Caucasus edition 01.09.2020, 7p. // <https://ijnet.org/en/story/data-journalism-tips-eastern-europe-central-asia-and-caucasus-edition>.

25.Visualization with Matplotlib <https://jakevdp.github.io/PythonDataScienceHandbook/04.00-introduction-to-matplotlib.html>

26.Wilson-Chapman, A. (2017, August 29). Panama Papers a “notable security success.” ICIJ. <https://www.icij.org/blog/2017/08/panama-papers-notable-security-success>

## ПОЛЕЗНЫЕ РЕСУРСЫ

### Приложение № 1 Международные источники данных

Мировые открытые данные:

| п/п | Название ресурса                                                  | Ссылка                                                                      | Характеристика                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----|-------------------------------------------------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1   | OpenCorporates                                                    | <a href="https://opencorporates.com/">https://opencorporates.com/</a>       | Крупнейшая открытая база данных компаний в мире.<br>Корпоративная прозрачность является важнейшим требованием для более справедливого общества. Компания предлагает данные, к которым можно получить доступ, проанализировать и запросить, когда и как это необходимо. Данные, которые нужны миру.<br>Компания стремится быть основным источником данных о компаниях во всем мире, оказывая положительное влияние на общество с помощью принципов OpenCorporates. Данные бесплатно доступны для общественности через наш веб-сайт, и бесплатно предоставляет доступ к информационным продуктам журналистам, неправительственным организациям и ученым для проведения общественных исследований.          |
| 2   | <u>Открытые базы данных иностранных государств (nalog.gov.ru)</u> | <a href="https://www.nalog.gov.ru/rn77/">https://www.nalog.gov.ru/rn77/</a> | Федеральная налоговая служба (ФНС России) является федеральным органом исполнительной власти, осуществляющим функции по контролю и надзору за соблюдением законодательства о налогах и сборах, за правильностью исчисления, полнотой и своевременностью внесения в соответствующий бюджет налогов, сборов и страховых взносов, в случаях, предусмотренных законодательством Российской Федерации, за правильностью исчисления, полнотой и своевременностью внесения в соответствующий бюджет иных обязательных платежей, за производством и оборотом табачной продукции, за применением контрольно-кассовой техники, а также функции органа валютного контроля в пределах компетенции налоговых органов. |

|   |                                                                                                                           |                                                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---|---------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 | Global Open Data Index - ( <a href="http://index.okfn.org/">okfn.org</a> )                                                | <a href="http://index.okfn.org/">http://index.okfn.org/</a>                                                                     | Глобальный индекс открытых данных (GODI) - ежегодный глобальный ориентир для публикации открытых правительственных данных, проводимый Сетью открытых знаний. Наш краудсорсинговый опрос измеряет открытость государственных данных в соответствии с определением Open.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 4 | <a href="https://data.worldbank.org/">World Bank Open Data   Data</a>                                                     | <a href="https://data.worldbank.org/">https://data.worldbank.org/</a>                                                           | Во Всемирном банке Группа данных о развитии координирует статистическую работу и сбор данных и ведет ряд макро-, финансовых и отраслевых баз данных. Тесно сотрудничая с регионами Банка и мировой практикой, группа руководствуется профессиональными стандартами при сборе, обобщении и распространении данных, чтобы гарантировать, что все пользователи данных могут быть уверены в качестве и целостности полученных данных.                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 5 | <a href="https://data.un.org/">UNdata</a>                                                                                 | <a href="https://data.un.org/">https://data.un.org/</a>                                                                         | UNdata - это веб-сервис передачи данных для глобального сообщества пользователей. Это делает международные статистические базы данных легко доступными для пользователей через единую точку входа. Пользователи могут осуществлять поиск и загружать различные статистические ресурсы, составленные статистической системой Организации Объединенных Наций (ООН) и другими международными учреждениями. Многочисленные базы данных или таблицы, известные под общим названием "datamarts", содержат более 60 миллионов точек данных и охватывают широкий спектр статистических тем, включая сельское хозяйство, преступность, связь, помощь в целях развития, образование, энергетику, окружающую среду, финансы, гендер, здравоохранение, рынок труда, производство, национальные счета, население и миграцию, науку и технологии, туризм, транспорт и торговля. |
| 6 | <a href="https://unstats.un.org/unsd/mbs/app/DataSearchTable.aspx">MBS Monthly Bulletin of Statistics Online (un.org)</a> | <a href="https://unstats.un.org/unsd/mbs/app/DataSearchTable.aspx">https://unstats.un.org/unsd/mbs/app/DataSearchTable.aspx</a> | Ежемесячная статистика предназначена для того, чтобы дать исчерпывающую характеристику нынешних социальных и экономических условий в районе Тайваня, Китайская Республика, и предоставить последние данные, которые включают все важные статистические серии, опубликованные в стране.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 7 | <a href="https://comtrade.un.org/">UN Comtrade   International Trade Statistics Database</a>                              | <a href="https://comtrade.un.org/">https://comtrade.un.org/</a>                                                                 | База данных Comtrade Организации Объединенных Наций объединяет подробную глобальную годовую и ежемесячную статистику торговли по продуктам и торговым партнерам для использования правительствами, академическими кругами,                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

|    |                                                                                                          |                                                                                                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                          |                                                                                                 | исследовательскими институтами и предприятиями. Данные, собранные Статистическим отделом Организации Объединенных Наций, охватывают примерно 200 стран и представляют более 99% мировой торговли товарами. Информация может быть извлечена в различных форматах, включая инструменты разработчика API для интеграции в корпоративные приложения и рабочие процессы. Подписчики получают доступ к дополнительным функциональным возможностям для повышения эффективности и специфичности.                                                                                                 |
| 8  | <a href="https://hdr.undp.org/">Human Development Data Center   Human Development Reports (undp.org)</a> | <a href="https://hdr.undp.org/">https://hdr.undp.org/</a>                                       | <p>ЧТО ТАКОЕ ЧЕЛОВЕЧЕСКОЕ РАЗВИТИЕ?</p> <p>В 1990 году в первом докладе о человеческом развитии был представлен новый подход к повышению благосостояния людей. Человеческое развитие – или подход к человеческому развитию – заключается в расширении богатства человеческой жизни, а не просто в богатстве экономики, в которой живут люди. Это подход, который ориентирован на людей, их возможности и выбор.</p>                                                                                                                                                                      |
| 9  | <a href="https://data.undp.org/">Home - UNDP Covid-19 Data Futures Platform</a>                          | <a href="https://data.undp.org/">https://data.undp.org/</a>                                     | <p>Платформа Data Futures - это открытый ресурс, который объединяет “лучшие” данные и преобразует многомерные данные в действенные, основанные на фактических данных идеи, способствующие системному мышлению и совместному созданию политических инициатив. Он охватывает широкий круг социально-экономических вопросов с целью оказания поддержки правительствам, специалистам по развитию и другим лицам, принимающим решения, в анализе потенциальных последствий различных комбинаций политических решений в разных странах и регионах, прежде чем будут привлечены инвестиции.</p> |
| 10 | <a href="https://www.who.int/data">Data at WHO</a>                                                       | <a href="https://www.who.int/data">https://www.who.int/data</a>                                 | <p>Забываясь о благополучии всех людей и руководствуясь наукой, Всемирная организация здравоохранения возглавляет и поддерживает глобальные усилия, направленные на то, чтобы предоставить всем и везде равные шансы вести здоровый образ жизни.</p>                                                                                                                                                                                                                                                                                                                                     |
| 11 | <a href="https://www.transparency.org">2020 - CPI - Transparency.org</a>                                 | <a href="https://www.transparency.org/en/cpi/2020">https://www.transparency.org/en/cpi/2020</a> | <p>Transparency International - это глобальное движение, работающее в более чем 100 странах, чтобы покончить с несправедливостью коррупции. Мы фокусируемся на вопросах, оказывающих наибольшее влияние на жизнь людей, и привлекаем сильных мира сего к ответственности ради общего блага. Посредством нашей адвокации,</p>                                                                                                                                                                                                                                                             |

|    |                                                                                           |                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                           |                                                                                                | кампаний и исследований мы работаем над разоблачением систем и сетей, которые позволяют процветать коррупции, требуя большей прозрачности и неподкупности во всех сферах общественной жизни.                                                                                                                                                                                                                                                                                                                          |
| 12 | <a href="#"><u>Data   UN Women Data Hub</u></a>                                           | <a href="https://data.unwomen.org/"><u>https://data.unwomen.org/</u></a>                       | <p>Центр данных Women Count стал возможен благодаря партнерам программы Women Count, в число которых входят правительства Австралии, Ирландии, Франции, Мексики, Швеции, Соединенных Штатов и Соединенного Королевства, а также Фонд Билла и Мелинды Гейтс, Фонд Alibaba и Elizabeth Arden.</p> <p>Центр данных Women Count представляет информационные панели и данные по гендерным показателям ЦУР, данные по ключевым тематическим областям Структуры «ООН-женщины» и последние ресурсы, связанные с COVID-19.</p> |
| 13 | <a href="#"><u>UNICEF DATA - Child Statistics</u></a>                                     | <a href="https://data.unicef.org"><u>https://data.unicef.org</u></a>                           | Информационная группа ЮНИСЕФ является глобальным центром сбора данных о детях. Она руководит сбором, проверкой, анализом, использованием и распространением наиболее достоверных со статистической точки зрения, сопоставимых на международном уровне данных о положении детей и женщин во всем мире. Команда поддерживает качество, целостность и организованность этих данных и делает их доступными как глобальное общественное благо на data.unicef.org веб-сайт.                                                 |
| 14 | <a href="#"><u>Situations (unhcr.org)</u></a>                                             | <a href="https://data.unhcr.org/en/situations"><u>https://data.unhcr.org/en/situations</u></a> | The Operational Data Portal (ODP) was created in 2011 to enable UNHCR's institutional responsibility to provide an information and data sharing platform to facilitate coordination of refugee emergencies. This was achieved using independent 'situation views' covering major emergencies such as the Syria situation or the Central African Republic emergency, among others.                                                                                                                                     |
| 15 | <a href="#"><u>Центр исследования коррупции и организованной преступности – OCCRP</u></a> | <a href="https://www.occrp.org/ru/"><u>https://www.occrp.org/ru/</u></a>                       | OCCRP поддерживает и развивает всемирную сеть журналистов-расследователей и публикует расследовательские материалы, тем самым информируя общественность о преступности и коррупции и привлекая власти к ответу.                                                                                                                                                                                                                                                                                                       |
| 16 | <a href="#"><u>Global Investigative Journalism Network (giin.org)</u></a>                 | <a href="https://giin.org/"><u>https://giin.org/</u></a>                                       | Глобальная сеть журналистских расследований (GIJN) - международная ассоциация некоммерческих организаций, которые поддерживают, продвигают и производят журналистские расследования. GIJN проводит                                                                                                                                                                                                                                                                                                                    |

|    |                                                                                                             |                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                             |                                                                                                             | конференции, тренинги, предоставляет ресурсы и справочную службу, а также поощряет создание аналогичных некоммерческих групп.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 17 | <a href="https://www.bellingcat.com/">bellingcat - the home of online investigations</a>                    | <a href="https://www.bellingcat.com/">https://www.bellingcat.com/</a>                                       | Bellingcat – это независимый международный коллектив исследователей, дознавателей и гражданских журналистов, использующий открытые источники и расследования в социальных сетях для изучения различных тем - от мексиканских наркобаронов и преступлений против человечности до отслеживания применения химического оружия и конфликтов по всему миру. Имея сотрудников и спонсоров в более чем 20 странах мира, мы работаем в уникальной области, где сочетаются передовые технологии, судебно-медицинские исследования, журналистика, расследования, прозрачность и подотчетность.                                                                                                                                                           |
| 18 | <a href="https://offshoreleaks.icij.org/">https://offshoreleaks.icij.org/</a>                               | <a href="https://offshoreleaks.icij.org/">https://offshoreleaks.icij.org/</a>                               | База данных снимает секретность, которая скрывает компании и трасты, зарегистрированные в налоговых убежищах, и разоблачает людей, стоящих за ними. Это включает в себя, если таковые имеются, имена реальных владельцев этих непрозрачных структур. В общей сложности интерактивное приложение раскрывает более 750 000 имен людей и компаний, стоящих за секретными оффшорными структурами. Они получены из просочившихся записей, а не из стандартизированного корпоративного реестра, поэтому могут быть дубликаты, в том числе в одной и той же утечке. В некоторых случаях компании перечислены в качестве акционеров другой компании или траста, что часто помогает скрыть людей из плоти и крови, стоящих за оффшорными организациями. |
| 19 | <a href="https://www.icij.org/investigations/paradise">https://www.icij.org/investigations/paradise</a>     | <a href="https://www.icij.org/">https://www.icij.org/</a>                                                   | Международный консорциум журналистов-расследователей рассказывает истории, которые пробиваются сквозь шум, показывая, как на самом деле устроен мир, вызывая позитивные изменения.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 20 | Github                                                                                                      | <a href="https://github.com/">https://github.com/</a>                                                       | Полноценная платформа для разработчиков для создания, масштабирования и доставки защищенного программного обеспечения.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 21 | <a href="https://www.google.com/sheets/about/">https://www.google.com/sheets/about/</a>                     | <a href="https://www.google.com/sheets/about/">https://www.google.com/sheets/about/</a>                     | Создавайте онлайн-таблицы и работайте над ними совместно с другими людьми на любых устройствах. Изменения отображаются в режиме реального времени.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 22 | <a href="https://medium.economist.com/plotting-the-brex">https://medium.economist.com/plotting-the-brex</a> | <a href="https://medium.economist.com/plotting-the-brex">https://medium.economist.com/plotting-the-brex</a> | подавляющее большинство методов визуализации данных имеют дело с данными, представленными в двух пространственных измерениях. Возьмите                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |

|    |                                                                         |                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|----|-------------------------------------------------------------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | <a href="#"><u>brexit-conundrum-fb8e804474fa</u></a>                    | <a href="#"><u>conundrum-fb8e804474fa</u></a>                            | линейную диаграмму, столбчатую диаграмму или диаграмму с разбросом по x-y. Каждый из них берет две переменные — ВВП и время, например, — и отображает их в двумерном пространстве. Но в инструментарию есть диаграмма, которая ломает эту модель: трехмерный разброс, называемый “троичным” графиком. В последние годы эти сюжеты в основном ограничивались специализированными публикациями (примеры можно увидеть здесь, здесь и здесь). Это история о том, как мы использовали троичную диаграмму для отображения головоломки Brexit для статьи для The Economist. |
| 23 | <a href="#"><u>https://rostrosyvocesdelconflicto.elsalvador.com</u></a> | <a href="#"><u>https://rostrosyvocesdelconflicto.elsalvador.com/</u></a> | События, которые произошли в Сальвадоре в период с 1972 по 1992 год, кажутся далекими. Этот сайт стремится придать этим событиям лицо и голос. Серьезные нарушения прав человека, военные преступления и преступления против человечности, совершенные теми и другими, могут остаться смутными воспоминаниями в памяти тех, кто пережил эти события, но это фрагменты памяти, которые все еще отмечают их жизнь и жизнь целой страны.                                                                                                                                 |

## Приложение № 2 Открытые данные - Центральная Азия

### Модуль 3: Ресурсы открытых данных в Кыргызстане

|  |                                                   |                                       |                                                                                                                                                                                                                                                             |
|--|---------------------------------------------------|---------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | Портал открытых данных Кыргызской Республики      | <a href="#"><u>data.gov.kg</u></a>    | Портал открытых данных Кыргызской Республики — это один из ключевых инструментов реализации государственной политики в области открытых данных, которому отводится роль системообразующего элемента, ядра экосистемы открытых данных Кыргызской Республики. |
|  | Реестр Министерства Юстиции Кыргызской Республики | <a href="#"><u>minjust.gov.kg</u></a> | Электронная база данных юридических лиц, филиалов (представительств) и СМИ Кыргызской Республики                                                                                                                                                            |
|  | Национальная интегрированная кадастровая система  | <a href="#"><u>cadastre.kg</u></a>    | Информация по недвижимому имуществу (земельные участки, строения), технические параметры недвижимости, информация по правам собственности на единицу                                                                                                        |

|  |                                                        |            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--|--------------------------------------------------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | Кыргызской Республики                                  |            | недвижимого имущества, информация по рынку недвижимости Кыргызстана и т.д.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|  | Министерство цифрового развития Кыргызской Республики  | ict.gov.kg | Министерство цифрового развития образовано Указом Президента Кыргызской Республики «О Кабинете Министров Кыргызской Республики» от 5 мая 2021 года № 114 в результате слияния Государственной службы цифрового развития при Правительстве КР и Государственной регистрационной службы при Министерстве юстиции КР. МЦР призвано вырабатывать и осуществлять поддержку в реализации государственной политики прежде всего в области цифровизации и электронного управления.                                                                                                                                                             |
|  | Государственная Налоговая Служба Кыргызской Республики | Salyk.kg   | Государственная налоговая служба при Правительстве Кыргызской Республики является государственным органом исполнительной власти, осуществляющим сбор налогов, страховых взносов и других платежей в бюджет и контроль за полнотой и своевременностью их уплаты, государственную регистрацию физических лиц в качестве индивидуальных предпринимателей, а также государственное регулирование и контроль за импортом, производством и/или оборотом этилового спирта и алкогольной и спиртосодержащей продукции.                                                                                                                         |
|  | Открытый бюджет                                        | okmot.kg   | Министерством финансов КР запущен Интернет портал "Прозрачный бюджет" (далее Портал) , который представляет собой автоматизированную систему предоставления данных по доходам и расходам республиканского и местных бюджетов в сети Интернет. Портал не публикует данные, связанные с предоставлением информации, доступ к которой ограничен в соответствии с законодательством КР. Посредством Портала реализуется принцип прозрачности бюджетных данных для широкого круга заинтересованных лиц, связь бюджетных средств с результатами деятельности государственных органов и органов местного самоуправления. Данный ресурс делает |

|  |                                                                                               |                |                                                                                                                                                                                                                                                                                                                                 |
|--|-----------------------------------------------------------------------------------------------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  |                                                                                               |                | доступной информации для всех категорий пользователей сети Интернет.                                                                                                                                                                                                                                                            |
|  | Судебные акты и заседания                                                                     | Act.sot.kg     | Сайт судебных актов по гражданским (административным) делам – по налоговым спорам, связанным с администрированием НДС.                                                                                                                                                                                                          |
|  | Официальный Портал Государственных Закупок Кыргызской Республики                              | Zakupki.gov.kg | Данные обо всех государственных закупках в единой информационной системе Веб-Портал Государственных Закупок Кыргызской Республики.                                                                                                                                                                                              |
|  | Национальный статистический комитет Кыргызской Республики - Статистика Кыргызстана            | stat.kg        | Национальный статистический комитет Кыргызской Республики (далее Нацстатком) является государственным органом, осуществляющим государственную статистическую деятельность на принципах профессиональной независимости и самостоятельности и координирующим деятельность в области учета и статистики на всей территории страны. |
|  | Портал открытых данных Министерства Образования Кыргызской Республики                         | edu.gov.kg     | Портал государственного органа исполнительной власти Киргизстана, осуществлявший функции по выработке государственной политики и нормативно-правовому регулированию в сфере образования, научной, научно-технической и инновационной деятельности                                                                               |
|  | Счетная палата Кыргызской Республики                                                          | esep.kg        | Портал государственного органа, который осуществляет контрольную, экспертно-аналитическую, информационную и иные виды деятельности; осуществляет внешний государственный финансовый контроль в сфере бюджетных правоотношений                                                                                                   |
|  | Государственная Таможенная Служба при Министерстве Экономики и Финансов Кыргызской Республики | customs.kg     | Государственная таможенная служба при Правительстве Кыргызской Республики является государственным органом исполнительной власти, осуществляющим непосредственное руководство таможенным делом в Кыргызской Республике и обладающим статусом вооруженного правоохранительного органа.                                           |
|  | Государственный портал электронных                                                            | tunduk.kg      | "Государственный портал электронных услуг является государственной информационной                                                                                                                                                                                                                                               |

|  |                             |             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--|-----------------------------|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | услуг Кыргызской Республики |             | системой, обеспечивающей предоставление государственных и муниципальных услуг в электронной форме и доступ заявителей к сведениям о государственных и муниципальных услугах. На Госпортале предоставляются государственные и муниципальные услуги, а также сервисы такие как проверка штрафов, проверка готовности общегражданского паспорта, водительских удостоверений, получение справки об отсутствии судимости и других справок и документов, которые реализованы в Системе межведомственного электронного взаимодействия «Түндүк»." |
|  | Школа Данных в Кыргызстане  | opendata.kg | Образовательный ресурс по дата-грамотности, часть международной сети School of Data. Целью организации является развитие площадки для нового поколения специалистов в Кыргызстане, готовых работать с данными и учить этому других.                                                                                                                                                                                                                                                                                                       |

## ПОЛЕЗНЫЕ РЕСУРСЫ

### Модуль1:

|  |                        |                                                                                                          |                                                                                                                                                                                |
|--|------------------------|----------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | Google Docs            | <a href="https://docs.google.com/">https://docs.google.com/</a>                                          | Сервис работы с документами и таблицами от Google.                                                                                                                             |
|  | Wordle                 | <a href="http://www.wordle.net">http://www.wordle.net</a>                                                | Wordle - сервис по созданию «карт текста».                                                                                                                                     |
|  | IBM ManyEyes           | <a href="http://958.ibm.com/software/data/cognos/manyeyes">958.ibm.com/software/data/cognos/manyeyes</a> | Бесплатный онлайн-инструмент визуализации данных IBM                                                                                                                           |
|  | Visual Complexity      | <a href="http://www.visualcomplexity.com/vc/">http://www.visualcomplexity.com/vc/</a>                    | Основная цель проекта - использовать критическое понимание различных методов визуализации в ряде дисциплин, таких разных, как биология, социальные сети или всемирная паутина. |
|  | Flowing Data           | <a href="http://flowingdata.com/">http://flowingdata.com/</a>                                            | Блог — сочетание работ и руководств по визуализации — является бесплатным ресурсом для всех.                                                                                   |
|  | Well-formed Data       | <a href="http://well-formed-data.net/">http://well-formed-data.net/</a>                                  | Блог визуализации данных                                                                                                                                                       |
|  | Information Aesthetics | <a href="http://infosthetics.com/">http://infosthetics.com/</a>                                          | Блог визуализации данных                                                                                                                                                       |

|  |                        |                                                                                                         |                                                                                                                                                              |
|--|------------------------|---------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | Good Magazine          | <a href="http://www.good.is/">http://www.good.is/</a>                                                   | Сайт некоммерческого идейного лидера и аналитического центра социальных действий, помогая социально ответственным организациям и отдельным лицам             |
|  | Simple Complexity      | <a href="http://simplecomplexity.net/">http://simplecomplexity.net/</a>                                 | Превратите данные и сложную аналитику в простые идеи и действия HR.                                                                                          |
|  | Infografistas          | <a href="http://infografistas.blogspot.com/">http://infografistas.blogspot.com/</a>                     | Испаноязычный ресурс по инфографике                                                                                                                          |
|  | Cool Infographics      | <a href="http://www.coolinfographics.com/">http://www.coolinfographics.com/</a>                         | Сайт для всего, что связано с визуализацией данных, инфографикой, презентациями, информационными панелями и визуальными коммуникациями.                      |
|  | Datenjournalist        | <a href="http://www.datenjournalist.de/">http://www.datenjournalist.de/</a>                             | Немецкоязычный ресурс по дата журналистике.                                                                                                                  |
|  | Cloud                  | <a href="http://www.documentcloud.org/home">http://www.documentcloud.org/home</a>                       | Универсальная платформа для документов: загрузка, организация, анализ, аннотация, поиск и вставка.                                                           |
|  | Google Code Playground | <a href="http://code.google.com/apis/ajax/playground/">http://code.google.com/apis/ajax/playground/</a> | Ресурс, который помогает изучить данные и инструменты Google.                                                                                                |
|  | Zemanta                | <a href="http://www.zemanta.com/">http://www.zemanta.com/</a>                                           | Zemanta — первая в мире многоканальная платформа для взаимодействия со спросом, созданная для взаимодействия и повышения производительности.                 |
|  | Open Calais            | <a href="http://www.opencalais.com/">http://www.opencalais.com/</a>                                     | Служба Thomson Reuters, которая автоматически извлекает семантическую информацию из веб-страниц в формате, пригодном для использования в семантической сети. |
|  | API Playground         | <a href="http://apiplayground.org/">http://apiplayground.org/</a>                                       | Предоставляет простой и удобный интерфейс написания и тестирования javascript-кода для восьми сервисов Google, использующих API.                             |
|  | Gapminder Desktop      | <a href="http://www.gapminder.org/desktop/">http://www.gapminder.org/desktop/</a>                       | Сервис для анимированной статистики.                                                                                                                         |
|  | Yahoo Pipes            | <a href="http://pipes.yahoo.com/pipes/">http://pipes.yahoo.com/pipes/</a>                               | Веб-приложение от Yahoo!, предоставляющее графический пользовательский интерфейс для создания приложений, объединяющих                                       |

|  |         |                                                                                           |                                                                                   |
|--|---------|-------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
|  |         |                                                                                           | ленты новостей, веб-страницы и другие сервисы.                                    |
|  | Tableau | <a href="http://www.tableausoftware.com/public">http://www.tableausoftware.com/public</a> | Программное обеспечения для интерактивной визуализации данных и бизнес-аналитики. |

#### Таджикистан:

| <b>Законодательство и нормативно-правовые документы:</b> |                                                               |                                                                                                                                                                                                                                                                           |                                                                                               |
|----------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|
|                                                          | Сайт Президента Республики Таджикистан                        | <a href="http://president.tj/">http://president.tj/</a>                                                                                                                                                                                                                   | законодательная база, новости, все послания президента, состав аппарата (четырёхязычный сайт) |
|                                                          | Сайт Национального Центра законодательства при Президенте     | <a href="http://www.mmk.tj/">http://www.mmk.tj/</a>                                                                                                                                                                                                                       | вся законодательная база, расположенная по годам на таджикском и частично русском языках.     |
|                                                          | Сайт Парламента Республики Таджикистан                        | <a href="http://www.parlament.tj/">http://www.parlament.tj/</a>                                                                                                                                                                                                           | законодательные инициативы, информация о депутатах и аппарате                                 |
| <b>Сайты местных органов исполнительной власти:</b>      |                                                               |                                                                                                                                                                                                                                                                           |                                                                                               |
|                                                          | Сайт Местного исполнительного органа гос. власти в г. Душанбе | <a href="http://www.dushanbe.tj/">http://www.dushanbe.tj/</a>                                                                                                                                                                                                             | законодательные инициативы в г. Душанбе                                                       |
|                                                          | Сайт хукумата Хатлонской области                              | <a href="http://www.khatlon.tj/">http://www.khatlon.tj/</a>                                                                                                                                                                                                               | законодательные инициативы в Хатлонской области                                               |
| <b>Публичные финансы:</b>                                |                                                               |                                                                                                                                                                                                                                                                           |                                                                                               |
|                                                          | Сайт Министерства финансов Таджикистана                       | <a href="http://www.minfin.tj/">http://www.minfin.tj/</a>                                                                                                                                                                                                                 | Официальная информация от Министерства финансов Таджикистана                                  |
|                                                          | Сайт Налогового комитета Таджикистана                         | <a href="http://andoz.tj">http://andoz.tj</a>                                                                                                                                                                                                                             | Официальная информация от Налогового комитета Таджикистана                                    |
|                                                          | Сайт Национального Банка Таджикистана                         | <a href="http://www.nbt.tj/tj/">http://www.nbt.tj/tj/</a>                                                                                                                                                                                                                 | Официальная информация от Национального Банка Таджикистана                                    |
| <b>Сведения о предприятиях:</b>                          |                                                               |                                                                                                                                                                                                                                                                           |                                                                                               |
|                                                          | Реестр юридических/физических лиц-предпринимателей:           | <a href="http://andoz.tj/index.php/ru/elektronnye-uslugi/informatsiya-pogсударstvennomu-reestru/reestr-yuridicheskikh-lits">http://andoz.tj/index.php/ru/elektronnye-uslugi/informatsiya-pogсударstvennomu-reestru/reestr-yuridicheskikh-lits</a>                         |                                                                                               |
|                                                          | Реестр физических лиц - индивидуальных предпринимателей       | <a href="http://andoz.tj/index.php/ru/elektronnye-uslugi/informatsiya-pogсударstvennomu-reestru/reestr-individualnykh-predprinimatelej">http://andoz.tj/index.php/ru/elektronnye-uslugi/informatsiya-pogсударstvennomu-reestru/reestr-individualnykh-predprinimatelej</a> |                                                                                               |
|                                                          | Налоговый комитет РТ:                                         |                                                                                                                                                                                                                                                                           |                                                                                               |
|                                                          | Раздел «Безответственные налогоплательщики»                   | <a href="http://andoz.tj/index.php/ru/bezotvetstvennye">http://andoz.tj/index.php/ru/bezotvetstvennye</a>                                                                                                                                                                 |                                                                                               |

|                                                                                                           |                                                                                                             |                                                                                                                                                                                                                                                                                                                          |
|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Список налогоплательщиков, исключенных из списка безответственных                                         | <a href="http://andoz.tj/index.php/ru/bezotvetstvennyye">http://andoz.tj/index.php/ru/bezotvetstvennyye</a> |                                                                                                                                                                                                                                                                                                                          |
| Единый реестр лицензий и разрешений Госкомитета по инвестициям и управлению государственным имуществом РТ | <a href="http://www.ijozat.tj/">http://www.ijozat.tj/</a>                                                   |                                                                                                                                                                                                                                                                                                                          |
| <b>Статистика экономики и торговли:</b>                                                                   |                                                                                                             |                                                                                                                                                                                                                                                                                                                          |
| Сайт Агентства по статистике при Президенте РТ                                                            | <a href="http://www.stat.tj/ru/">http://www.stat.tj/ru/</a>                                                 | Агентство по статистике при Президенте Республики Таджикистан является центральным статистическим органом страны.                                                                                                                                                                                                        |
| Сайт Минэкономразвития и торговли РТ                                                                      | <a href="http://www.medt.tj/index.php/tj/">http://www.medt.tj/index.php/tj/</a>                             | Официальный сайт Министерства Экономического Развития и Торговли Республики Таджикистан                                                                                                                                                                                                                                  |
| Сайт Таможенной службы                                                                                    | <a href="http://www.customs.tj/">http://www.customs.tj/</a>                                                 | Официальный сайт со статистикой внешней торговли Республики Таджикистан                                                                                                                                                                                                                                                  |
| Сайт МИД РТ                                                                                               | <a href="http://mfa.tj/tj/">http://mfa.tj/tj/</a>                                                           | Официальный сайт Министерства иностранных дел Республики Таджикистан со статистикой иностранных инвестиций                                                                                                                                                                                                               |
| Сайт Минэнергопрома                                                                                       | <a href="http://www.minenergo.prom.tj">http://www.minenergo.prom.tj</a>                                     | Официальный сайт Министерства энергетики и промышленности Республики Таджикистан                                                                                                                                                                                                                                         |
| <b>Государственные закупки:</b>                                                                           |                                                                                                             |                                                                                                                                                                                                                                                                                                                          |
| Сайт профильного Агентства по госзакупкам                                                                 | <a href="http://zakupki.gov.tj">http://zakupki.gov.tj</a>                                                   | Агентство по государственным закупкам товаров, работ и услуг при Правительстве Республики Таджикистан                                                                                                                                                                                                                    |
| Сайт Госкомитета по инвестициям и управлению государственным имуществом                                   | <a href="http://www.gki.tj/">http://www.gki.tj/</a>                                                         | Официальный сайт, на котором размещается информация по проводимым тендерам, аукционам и конкурсам в рамках полномочий данного министерства:<br><a href="http://www.muzoyada-kvd.tj/index.php?category=1&amp;material_s=2&amp;lang=toj">http://www.muzoyada-kvd.tj/index.php?category=1&amp;material_s=2&amp;lang=toj</a> |
| Тендеры по государственным                                                                                | <a href="http://www.gki.tj/ru/state-tender/">http://www.gki.tj/ru/state-tender/</a>                         | Государственный комитет по инвестициям и управлению                                                                                                                                                                                                                                                                      |

|                                                                  |                                                                                                                                                                                                                                         |                                                                                                                                                                  |
|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| инвестиционным проектам Республики Таджикистан                   |                                                                                                                                                                                                                                         | государственным имуществом Республики Таджикистан                                                                                                                |
| Объявления о госзакупках в Душанбе                               | <a href="http://www.dushanbe.tj/ru/news/">http://www.dushanbe.tj/ru/news/</a>                                                                                                                                                           | Исполнительный орган государственной власти города Душанбе                                                                                                       |
| Сайт Управления по инвестициям и госимуществу Согдийской области | <a href="http://www.sugdinvest.tj/index.php?option=com_content&amp;view=category&amp;id=63&amp;Itemid=154&amp;lang=ru">http://www.sugdinvest.tj/index.php?option=com_content&amp;view=category&amp;id=63&amp;Itemid=154&amp;lang=ru</a> |                                                                                                                                                                  |
| Генплан                                                          | <a href="http://www.dushanbe.tj/ru/masttplan/graphics/">http://www.dushanbe.tj/ru/masttplan/graphics/</a>                                                                                                                               | Генеральный план города Душанбе                                                                                                                                  |
| Патентно-информационный центр Минэкономразвития и торговл        | <a href="http://www.ncpi.tj/">http://www.ncpi.tj/</a>                                                                                                                                                                                   | Официальный сайт Государственного учреждения «Национальный патентно-информационный центр» Министерства экономического развития и торговли Республики Таджикистан |
| <b>Правоохранительная система:</b>                               |                                                                                                                                                                                                                                         |                                                                                                                                                                  |
| Сайт МВД РТ                                                      | <a href="http://www.vkd.tj/">http://www.vkd.tj/</a>                                                                                                                                                                                     | Использование информации официального сайта Министерства внутренних дел Республики Таджикистан разрешается с указанием на ссылку сайта                           |
| Контроль и борьба с коррупцией РТ                                | <a href="http://www.akn.tj">www.akn.tj</a>                                                                                                                                                                                              | Официальный сайт агентства государственного финансового контроля и борьбы с коррупцией                                                                           |
| Сайт Генпрокуратуры РТ                                           | <a href="http://www.prokuratura.tj/new/ru/structura/generalprok.html">http://www.prokuratura.tj/new/ru/structura/generalprok.html</a>                                                                                                   | Генеральная Прокуратура Республики Таджикистан                                                                                                                   |
| КХФ Таджикистан                                                  | <a href="http://www.khf.tj">http://www.khf.tj</a>                                                                                                                                                                                       | Официальный сайт комитета по чрезвычайным ситуациям и гражданской обороне                                                                                        |
| Сайт Совета Юстиции РТ                                           | <a href="http://shuroiadliya.tj/">http://shuroiadliya.tj/</a>                                                                                                                                                                           | На сайте также находятся ссылки на сайты 28 судов, (в т.ч. экономических) городов и районов республики                                                           |
| Сайт омбудсмена                                                  | <a href="http://www.ombudsman.tj">http://www.ombudsman.tj</a>                                                                                                                                                                           | Официальный сайт Министерства Внутренних Дел Республики Таджикистан                                                                                              |
|                                                                  |                                                                                                                                                                                                                                         | Официальный сайт омбудсмена Республики Таджикистан в т.ч.                                                                                                        |

|  |  |                                                              |
|--|--|--------------------------------------------------------------|
|  |  | информация об общественных приемных омбудсмена в регионах РТ |
|--|--|--------------------------------------------------------------|

### Приложение № 3 Инструменты по журналистике данных

|                                     |                                                                                                                                                               |                                                                                                                                                                                                                                            |
|-------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Awesome Data Journalism             | <a href="https://github.com/infoculture/awesome-datajournalism">https://github.com/infoculture/awesome-datajournalism</a>                                     | Открытый исходный код, открытые данные и просто открытый репозиторий журналистики данных, чтобы изучить и понять практическую журналистику данных.                                                                                         |
| Essential DDJ resources             | <a href="https://ijnet.org/en/story/essential-data-journalism-tips-and-resources">https://ijnet.org/en/story/essential-data-journalism-tips-and-resources</a> | Важные советы и ресурсы по журналистике данных                                                                                                                                                                                             |
| Data Journalism                     | <a href="https://datajournalism.com/">https://datajournalism.com/</a>                                                                                         | Обширный ресурс по дата журналистике                                                                                                                                                                                                       |
| Планета Excel                       | <a href="https://www.planetaexcel.ru/">https://www.planetaexcel.ru/</a>                                                                                       | Приемы, трюки и видео уроки по работе в Microsoft Excel с формулами, таблицами, макросами, диаграммами.                                                                                                                                    |
| Онлайн курсы по журналистике данных | <a href="https://mdq.kz/ru/education/online/data-zhurnalistika/">https://mdq.kz/ru/education/online/data-zhurnalistika/</a>                                   | Корпоративный Фонд “Медиа дамыту қоры” создан по инициативе ведущих государственных средств массовой информации Казахстана с целью содействовать повышению конкурентоспособности отечественных средств массовой информации и медиа отрасли |
| Онлайн курсы по журналистике данных | <a href="https://www.coursera.org/learn/socialnaya-set">https://www.coursera.org/learn/socialnaya-set</a>                                                     | «Ловцы человеков» или социальные сети в медиа, бизнесе, рекрутинге и образовании                                                                                                                                                           |
| Visualization for Data Journalism   | <a href="https://www.coursera.org/learn/visualization-for-data-journalism">https://www.coursera.org/learn/visualization-for-data-journalism</a>               | Цель онлайн курса — научиться думать о визуальном представлении данных, о том, как и почему это работает, и как это сделать правильно.                                                                                                     |

|                                                                      |                                                                                                                                                                         |                                                                                                                                                                                                                                                         |
|----------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| BSc Computer Science University of London                            | <a href="https://www.coursera.org/degrees/bachelor-of-science-computer-science-london">https://www.coursera.org/degrees/bachelor-of-science-computer-science-london</a> | Востребованные навыки работы с компьютером и отточите свои инновации и креативность, получив диплом Лондонского университета.                                                                                                                           |
| Master of Science in Data Science University of Colorado Boulder     | <a href="https://www.coursera.org/degrees/master-of-science-data-science-boulder">https://www.coursera.org/degrees/master-of-science-data-science-boulder</a>           | Теория, методы, инструменты и языки программирования — навыки, которые позволят работать на различных должностях специалистов по данным                                                                                                                 |
| Online Master of Computer Science Arizona State University           | <a href="https://www.coursera.org/degrees/master-of-computer-science-asu">https://www.coursera.org/degrees/master-of-computer-science-asu</a>                           | Программа получения степени магистра компьютерных наук (MCS) Университета штата Аризона обеспечивает высококачественное обучение в сочетании с реальным опытом посредством прикладных проектов.                                                         |
| MBA in Business Analytics                                            | <a href="https://www.coursera.org/degrees/mba-business-analytics-jgu">https://www.coursera.org/degrees/mba-business-analytics-jgu</a>                                   | Программа по набору аналитических навыков, необходимых для продвижения по карьерной лестнице в современной бизнес-среде, ориентированной на данные и аналитику                                                                                          |
| Maestría en Inteligencia Analítica de Datos Universidad de los Andes | <a href="https://www.coursera.org/degrees/maestria-analitica-de-datos-uniandes">https://www.coursera.org/degrees/maestria-analitica-de-datos-uniandes</a>               | Онлайн-программа Analytical Data Intelligence (MIAD) готовит профессионалов, которые выделяются своими техническими и передовыми навыками на стыке трех областей знаний: математическое моделирование, информационные технологии и управление бизнесом. |
| Специализация Наука о данных                                         | <a href="https://www.coursera.org/specializations/jhu-data-science">https://www.coursera.org/specializations/jhu-data-science</a>                                       | Введение в науку о данных из десяти курсов, разработанное и преподаваемое ведущими профессорами.                                                                                                                                                        |

|                                                  |                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                    |
|--------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Deep Learning Specialization                     | <a href="https://www.coursera.org/specializations/deep-learning">https://www.coursera.org/specializations/deep-learning</a>                           | Специализация по глубокому обучению — это базовая программа, которая поможет вам понять возможности, проблемы и последствия глубокого обучения и подготовит вас к участию в разработке передовых технологий искусственного интеллекта.                                                                                                                                                             |
| Специализация Data Science: Foundations using R  | <a href="https://www.coursera.org/specializations/data-science-foundations-r">https://www.coursera.org/specializations/data-science-foundations-r</a> | Эта специализация охватывает основные инструменты и методы обработки данных, включая получение, очистку и исследование данных, программирование в R и проведение воспроизводимых исследований. Учащиеся, завершившие эту специализацию, будут готовы пройти специализацию «Наука о данных: статистика и машинное обучение», в которой они создают продукт данных, используя данные реального мира. |
| Специализация Прикладная наука о данных с Python | <a href="https://www.coursera.org/specializations/data-science-python">https://www.coursera.org/specializations/data-science-python</a>               | Курсы Мичиганского университета знакомят учащихся с наукой о данных с помощью языка программирования Python.                                                                                                                                                                                                                                                                                       |
| Специализация Навыки Excel для бизнеса           | <a href="https://www.coursera.org/specializations/excel">https://www.coursera.org/specializations/excel</a>                                           | Эта специализация предназначена для всех, кто стремится развить один из самых важных и фундаментальных цифровых навыков на сегодняшний день. Программное обеспечение для работы с электронными таблицами остается одним из самых распространенных программных продуктов, используемых на рабочих местах по всему миру.                                                                             |

|                                                                     |                                                                                                                         |                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Набор инструментальных средств для специалистов по обработке данных | <a href="https://www.coursera.org/learn/data-scientists-tools">https://www.coursera.org/learn/data-scientists-tools</a> | Курс дает обзор данных, вопросов и инструментов, с которыми работают аналитики данных и специалисты по данным.                                                                                                                                                                                           |
| Специализация от Excel до MySQL: способы анализа бизнес-данных      | <a href="https://www.coursera.org/specializations/excel-mysql">https://www.coursera.org/specializations/excel-mysql</a> | В этой специализации вы научитесь формулировать бизнес-задачи как вопросы данных. Вы будете использовать мощные инструменты и методы, такие как Excel, Tableau и MySQL, для анализа данных, создания прогнозов и моделей, разработки визуализаций и обмена идеями.                                       |
| Программирование на языке R                                         | <a href="https://www.coursera.org/learn/r-programming">https://www.coursera.org/learn/r-programming</a>                 | Курс охватывает практические вопросы статистических вычислений, включая программирование в R, чтение данных в R, доступ к пакетам R, написание функций R, отладку, профилирование кода R, а также организацию и комментирование кода R. Темы статистического анализа данных предоставят рабочие примеры. |

### Приложение № 3 Инструменты по журналистике данных

#### ПОЛЕЗНЫЕ РЕСУРСЫ

##### Модуль1:

1. Google Docs — <https://docs.google.com/> - Сервис работы с документами и таблицами от Google.
2. Wordle — <http://www.wordle.net> Wordle - сервис по созданию «карт текста».
3. IBM ManyEyes — <http://www->
4. 958.ibm.com/software/data/cognos/manyeyes
5. Visual Complexity, <http://www.visualcomplexity.com/vc/>
6. Flowing Data, <http://flowingdata.com/>
7. Well-formed Data, <http://well-formed-data.net/>
8. Information Aesthetics, <http://infosthetics.com/>
9. Good Magazine, <http://www.good.is/>
10. University of Amsterdam, <http://www.digitalmethods.net/>
11. Simple Complexity, <http://simplecomplexity.net/>
12. A Beautiful WWW, <http://abeautifulwww.com/>

- 13..Infografistas, <http://infografistas.blogspot.com/>
- 14.Visual Editors, <http://www.coolinfographics.com/>
- 15.Cool Infographics, <http://www.coolinfographics.com/>
- 16.Datenjournalist. German blog on ddj. <http://www.datenjournalist.de/>
- 17.Cloud, <http://www.documentcloud.org/home>
- 18.Google Code Playground: Helps to explore Google data and
- 19.tools. <http://code.google.com/apis/ajax/playground/>
- 20.Zemanta, <http://www.zemanta.com/>
- 21.Open Calais, <http://www.opencalais.com/>
- 22.API Playground - helping journalists to
- 23.understand API Data. <http://apiplayground.org/>
- 24.Gapminder Desktop, <http://www.gapminder.org/desktop/>
- 25.Yahoo Pipes, <http://pipes.yahoo.com/pipes/>
- 26.Tableau, <http://www.tableausoftware.com/public>

## **Модуль 2:**

1. Awesome Data Journalism <https://github.com/infoculture/awesome-datajournalism>
2. Essential DDJ resources <https://ijnet.org/en/story/essential-data-journalism-tips-and-resources>
3. Data Journalism <https://datajournalism.com/>
4. - Планета Excel <https://www.planetaexcel.ru/>
5. Онлайн курсы по журналистике данных
6. Онлайн курс "Data-журналистика" <https://mdq.kz/ru/education/online/data-zhurnalistika/>
7. «Ловцы человеков» или социальные сети в медиа, бизнесе, рекрутинге и образовании  
<https://www.coursera.org/learn/sotsialnaya-set>
8. Visualization for Data Journalism <https://www.coursera.org/learn/visualization-for-data-journalism>

## **Полезные ресурсы по визуализации:**

1. <https://unesdoc.unesco.org/ark:/48223/pf0000374922/PDF/374922rus.pdf.multi.page=58>
2. [https://live.osgeo.org/archive/8.0/ru/overview/qgis\\_overview.html](https://live.osgeo.org/archive/8.0/ru/overview/qgis_overview.html)
3. <https://old.dataone.org/software-tools/quantum-gis-qgis>
4. <https://www.esri.com/en-us/arcgis/products/arcgis-online/overview>
5. [https://gis-lab.info/docs/osgeo/ru/overview/grass\\_overview.html](https://gis-lab.info/docs/osgeo/ru/overview/grass_overview.html)
6. <https://www.nytimes.com/interactive/2021/us/covid-cases.html>
7. <https://openrefine.org>
8. <https://public.tableau.com/s/>
9. <https://www.toptal.com/designers/data-visualization/data-visualization-tools>
10. <https://support.google.com/fusiontables/answer/184641?hl=en>
11. <https://www.seerinteractive.com/blog/visualize-your-backlinks-with-google-fusion-tables/>
12. <https://www.ibm.com/software/data/cognos/manyeyes/>
13. <http://www.icharts.net>
14. <http://geocommons.com> <https://piktochart.com>
15. <https://github.com>
16. <https://dygraphs.com>
17. <https://vega.github.io/vega/>

#### Приложение № 4 Рекомендуемые курсы MOOK по дата журналистике

1. <https://www.coursera.org/degrees/master-of-data-science-hse> - Master of Data Science HSE University
2. <https://www.coursera.org/degrees/bachelor-of-science-computer-science-london> - BSc Computer Science University of London
3. <https://www.coursera.org/degrees/master-of-science-data-science-boulder> - Master of Science in Data Science University of Colorado Boulder
4. <https://www.coursera.org/degrees/master-of-computer-science-asu> - Online Master of Computer Science Arizona State University
5. <https://www.coursera.org/degrees/master-of-applied-data-science-umich> - Master of Applied Data Science University of Michigan
6. <https://www.coursera.org/degrees/mba-business-analytics-jgu> - MBA in Business Analytics
7. Jindal Global University
8. <https://www.coursera.org/degrees/master-of-data-network-analytics-hse> - Master of Data and Network Analytics HSE University
9. <https://www.coursera.org/degrees/maestria-analitica-de-datos-uniandes> - Maestría en Inteligencia Analítica de Datos Universidad de los Andes
10. <https://www.coursera.org/specializations/jhu-data-science> - Специализация Наука о данных
11. <https://www.coursera.org/specializations/deep-learning> - Deep Learning Specialization
12. <https://www.coursera.org/specializations/data-science-foundations-r> - Специализация Data Science: Foundations using R - <https://www.coursera.org/learn/python-data> Структуры данных в Python
13. <https://www.coursera.org/specializations/data-science-python> - Специализация Прикладная наука о данных с Python
14. <https://www.coursera.org/specializations/excel> - Специализация Навыки Excel для бизнеса
15. <https://www.coursera.org/learn/data-scientists-tools> - Набор инструментальных средств для специалистов по обработке данных
16. <https://www.coursera.org/specializations/excel-mysql> - Специализация от Excel до MySQL: способы анализа бизнес-данных

17. <https://www.coursera.org/learn/r-programming> - Программирование на языке R



**unesco**

МПРК - Международная  
программа развития  
коммуникации

Кафедра ЮНЕСКО по  
журналистике и  
коммуникации,  
Казахский  
национальный  
университет  
им.аль-Фараби,  
Алматы, Казахстан

|    |                                                                                                             |                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    |                                                                                                             |                                                                                                             | конференции, тренинги, предоставляет ресурсы и справочную службу, а также поощряет создание аналогичных некоммерческих групп.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 17 | <a href="https://www.bellingcat.com/">bellingcat - the home of online investigations</a>                    | <a href="https://www.bellingcat.com/">https://www.bellingcat.com/</a>                                       | Bellingcat – это независимый международный коллектив исследователей, дознавателей и гражданских журналистов, использующий открытые источники и расследования в социальных сетях для изучения различных тем - от мексиканских наркобаронов и преступлений против человечности до отслеживания применения химического оружия и конфликтов по всему миру. Имея сотрудников и спонсоров в более чем 20 странах мира, мы работаем в уникальной области, где сочетаются передовые технологии, судебно-медицинские исследования, журналистика, расследования, прозрачность и подотчетность.                                                                                                                                                           |
| 18 | <a href="https://offshoreleaks.icij.org/">https://offshoreleaks.icij.org/</a>                               | <a href="https://offshoreleaks.icij.org/">https://offshoreleaks.icij.org/</a>                               | База данных снимает секретность, которая скрывает компании и трасты, зарегистрированные в налоговых убежищах, и разоблачает людей, стоящих за ними. Это включает в себя, если таковые имеются, имена реальных владельцев этих непрозрачных структур. В общей сложности интерактивное приложение раскрывает более 750 000 имен людей и компаний, стоящих за секретными оффшорными структурами. Они получены из просочившихся записей, а не из стандартизированного корпоративного реестра, поэтому могут быть дубликаты, в том числе в одной и той же утечке. В некоторых случаях компании перечислены в качестве акционеров другой компании или траста, что часто помогает скрыть людей из плоти и крови, стоящих за оффшорными организациями. |
| 19 | <a href="https://www.icij.org/investigations/paradise">https://www.icij.org/investigations/paradise</a>     | <a href="https://www.icij.org/">https://www.icij.org/</a>                                                   | Международный консорциум журналистов-расследователей рассказывает истории, которые пробиваются сквозь шум, показывая, как на самом деле устроен мир, вызывая позитивные изменения.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 20 | Github                                                                                                      | <a href="https://github.com/">https://github.com/</a>                                                       | Полноценная платформа для разработчиков для создания, масштабирования и доставки защищенного программного обеспечения.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 21 | <a href="https://www.google.com/sheets/about/">https://www.google.com/sheets/about/</a>                     | <a href="https://www.google.com/sheets/about/">https://www.google.com/sheets/about/</a>                     | Создавайте онлайн-таблицы и работайте над ними совместно с другими людьми на любых устройствах. Изменения отображаются в режиме реального времени.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 22 | <a href="https://medium.economist.com/plotting-the-brex">https://medium.economist.com/plotting-the-brex</a> | <a href="https://medium.economist.com/plotting-the-brex">https://medium.economist.com/plotting-the-brex</a> | Поддавляющее большинство методов визуализации данных имеют дело с данными, представленными в двух пространственных измерениях. Возьмите                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

