

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Институт «Высшая школа журналистики и массовых коммуникаций»

На правах рукописи

РОСЛИКОВА Валерия Валерьевна

Способы визуализации big data в современной журналистике

Профиль магистратуры – «Медиадизайн»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Научный руководитель –
кандидат политических наук,
доцент Нигматуллина К. Р.

Вх. № _____ от _____
Секретарь _____

Санкт-Петербург
2017

Содержание

ВВЕДЕНИЕ	4
ГЛАВА 1. ФЕНОМЕН БОЛЬШИХ ДАННЫХ В СОВРЕМЕННОЙ ЖУРНАЛИСТИКЕ	10
1.1 К определению понятия big data.....	10
1.2 К определению «журналистики данных» (дата-журналистики).....	14
1.3 Мультимедийный проект как форма представления информации в журналистике данных	24
1.3.1 Отличительные свойства интернет-СМИ.....	24
1.3.2 Мультимедийность как свойство интернет-СМИ	27
1.3.3 Интерактивная инфографика как способ визуализации big data	32
1.3.4 Сторителлинг как способ визуализации в проектах big data.....	34
1.3.5 Элементы геймификации в проектах big data	39
ГЛАВА 2. СПОСОБЫ ВИЗУАЛИЗАЦИИ BIG DATA В СОВРЕМЕННОЙ ЖУРНАЛИСТИКЕ	48
2.1 Определение визуализации данных	48
2.2 Способы визуализации big data: определение и классификации.....	51
2.3 Контент-анализ визуализации big data изданий The New York Times, The Washington Post, The Guardian	55
2.3.1 Структура и тематические особенности материалов	58
2.4 Способы визуализации географического типа данных.....	62
2.4.1 Карты (map)	62
2.5 Способы визуализации дискретного типа данных	71
2.5.1 Деревья (tree map)	71
2.5.2 Солнечная диаграмма (sunburst diagram).....	73
2.5.3 Матричная диаграмма (dot matrix chart)	75
2.5.4 Параллельные наборы (parallel sets).....	77
2.5.5 Пиктограммы (pictogram chart).....	78
2.5.6 Хордовая диаграмма (chord diagram)	80
2.5.7 Диаграмма параллельных координат (parallel coordinates plot)	81
2.5.8 Сложная столбчатая диаграмма (stacked bar graphs).....	83
2. 6 Способы визуализации непрерывных временных данных	84
2.6.1 Поточковый график (streamgraph).....	84

2.6.2	Графики «открыто-высокого-низкого закрытия» (или диаграммы OHLC).....	85
2.6.3	Линейный график (line chart)	87
2.6.4	Диаграмма рассеяния (scatterplots).....	92
	ЗАКЛЮЧЕНИЕ	99
	СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	102
	ПРИЛОЖЕНИЯ	112

ВВЕДЕНИЕ

На современном этапе развития журналистика под влиянием технологий терпит глобальные, революционные изменения. Она находится в прямой зависимости от стремительно развивающихся информационных технологий (ИТ) и таких процессов, как конвергенция и цифровизация (дигитализация). В частности, с развитием ИТ зародилась и, опираясь на современные тенденции, трансформируется онлайн-журналистика.

С развитием Интернета в геометрической прогрессии начала увеличиваться публикуемая информация, а у читателя независимо от традиционных СМИ появилась возможность узнавать актуальную информацию в режиме реального времени. Появилась потребность в отсеивании некорректной информации и фильтрации источников — необходимость структурировать большие потоки информации в один материал.

Все это привело к поиску нового инструментария для создания текста — наглядного, простого для восприятия, интерактивного, требующего читательской рефлексии материала. Так появилась data journalism («журналистика данных»). Тенденция к использованию big data («больших данных») как профессионального направления в журналистике была зафиксирована в 2006 году в США, в 2013 — в России¹.

«Визуальная культура создает центрированные нации»² — писал Маршалл Маклюэн, и если в конце 90-х мир был объединен единой сетью, то сейчас — едиными визуальными формами. Изобилие этих форм и неподготовленность журналистов и дизайнеров к работе с большими

¹ Шилина М. Г., Левченко, В. Ю. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики [Электронный ресурс] // Журнал Факультета журналистики МГУ имени М.В. Ломоносова «МедиаСкоп», 2014. Режим доступа: <http://www.mediascope.ru/node/1486> (дата обращения: 14.05.17)

² Маклюэн М. Понимание Медиа: внешние расширения человека / перевод с английского В. Г. Николаева. — М.: Гиперборей; Кучково поле, 2007. — С. 464.

объемами данных затрудняет выбор правильной визуализации. Между тем, от этого решения зависит точность и правильность интерпретации информации, что для журналистского материала должно быть приоритетнее формы.

Актуальность работы связана с распространением исследуемого явления на мировом рынке, в том числе и в отечественной журналистике. Несмотря на то, что использование разных способов визуализации больших данных нашло свое отражение в таких СМИ, как BBC, The Guardian, Los-Angeles Times, Financial Times, РИА Новости, Slon.ru, Forbes и др., зависимость типа данных от способов их визуализации изучена мало. Это затрудняет выбор и приводит к малоэффективным и некорректным способам визуализации данных.

Степень разработанности темы. Зависимость типа информации от её графического представления была описана в трудах Э. Тафти в 1983 году в труде «Визуальное представление больших объёмов информации» (The Visual Display of Quantitative Information)³. В 1914 году в свет вышла книга инженера В. Бринтона «Графическое изображение фактов». По-настоящему прогрессивной в этой связи считается работа Джина Железны, в которой дизайнеру предложен алгоритм визуализации данных, напрямую зависящий от типа информации. Анализ проектов в области дата журнализма был проведен М.Г. Шилиной⁴ в 2017 году. Исследований, посвященных визуализации проектов журналистики данных и их зависимости от типа информации в области больших данных, мы не встречали.

Цель данного исследования — изучить зависимость между big data и способами их визуализации. Достижение цели планируется путем решения следующих научно-практических задач:

³ Tufte E.R. The Visual Display of Quantitative Information. 2nd edition. — Cheshire, CT: Graphics Press, 2001. — P. 199.

⁴ Шилина М. Г. Дата журнализм: особенности визуализации проектов. Медиа в современном мире. Молодые исследователи: материалы 16-й международной конференции студентов, магистрантов и аспирантов (15–17 марта 2017 года) / Под ред. А. С. Смоляровой; сост. А. Н. Марченко. — СПб.: С.-Петербург. гос. ун-т, 2017. С. 605.

— проанализировать существующие на сегодняшний день теории массовых коммуникаций на предмет используемых в них определений понятий «большие данные», «журналистика данных», «проект журналистики данных»;

— определить ключевые типы информации в больших данных, существенные для исследования визуализации big data в современной журналистике;

— проанализировать теории визуального представления информации на предмет используемых в них определений понятий «визуализация данных», «способы визуализации данных» и «интерактивная инфографика»;

— провести качественный и количественный анализ уже существующих проектов в области журналистики данных в таких зарубежных СМИ как The Guardian, The New York Times, The Washington Post.

Объектом исследования является журналистика данных в зарубежных СМИ.

Предмет исследования — способы визуализации big data в журналистских материалах.

Гипотеза нашего исследования звучит следующим образом — при выборе способа визуализации специалист руководствуется типом визуализируемых данных (дискретные, географические или непрерывные временные данные), темой проекта, а также его целью. Для проверки данной гипотезы был проведен формализованный анализ визуализаций big data ведущих интернет-СМИ Европы и США. Для достижения поставленной цели и решения необходимых задач в исследовании применялись **общенаучные методы** такие, как: анализ, индукция, дедукция, синтез. Для сбора эмпирического материала использовались: классический и контент-анализ.

Теоретическую базу данного исследования составляют работы по теории коммуникации Е.Л. Вартановой⁵, М.Г. Шилиной⁶, М.М. Лукиной⁷, Э. Тафти⁸, М. Маклюэна⁹, В. Майер-Шенбергера¹⁰, монографии и публикации, поднимающие вопросы графического метода представления информации Р. Арнхейма¹¹, Д. Желзязны¹², Н. Яу¹³, А.Ю. Зиновьева¹⁴, В.М. Розина¹⁵, В. Бринтона¹⁶; труды по социологии и социологическим методам исследования контента средств массовой информации А.В. Семёнова, М.В. Корсунской¹⁷, М.К. Горшкова и Ф.Э. Шереги¹⁸.

В качестве **эмпирической базы** было взято 53 проекта журналистики данных из исследуемых общественно-политических изданий (The Guardian, The New York Times, The Washington Post) за период с 2011 по 2016 год.

⁵ Вартанова Е. Л. Теория СМИ: Актуальные вопросы: сб. науч. статей. – М.: МедиаМир, 2009. С. 488.

⁶ Дата журнализм: особенности визуализации проектов. С. 521.

⁷ Лукина М.М. Интернет – СМИ: Теория и практика: Учеб. пособие для студентов вузов / М.М. Лукина. – М.: Аспект Пресс, 2010. С. 348.

⁸ Tufte E.R. Beautiful Evidence. – Cheshire, CT: Graphics Press, 2006. P. 213.; Idem. Envisioning Information. – Cheshire, CT: Graphics Press, 1990. P. 126; Idem. Visual Explanations: Images and Quantities, Evidence and Narrative. – Cheshire, CT: Graphics Press, 1997. P. 157.

⁹ Маклюэн Г. М. Галактика Гутенберга. Сотворение человека печатной культуры. — Перевод с английского и примечания: А. Юдин. — М., 2003. С 432.

¹⁰ Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / В. Майер-Шенбергер, К. Кукьер – Манн, Иванов и Фербер, 2014. С. 240.

¹¹ Арнхейм Р. Искусство и визуальное восприятие. М., Архитектура–С, 2007. С. 392.

¹² Желязны Д. Говори на языке диаграмм: Пособие по визуальным коммуникациям для руководителей / Пер. с англ. – М.: Институт комплексных стратегических исследований, 2004. С. 220.

¹³ Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми способами / Нейта Яу; пер. с англ. Светланы Кировой. – М.: Манн, Иванов и Фербер, 2013. С. 352.

¹⁴ Зиновьев А. Ю. Визуализация многомерных данных. Красноярск: Изд-во КГТУ, 2000. С. 168.

¹⁵ Розин В.М. Визуальная культура и восприятие: Как человек видит и понимает мир. — М.: Эдиториал УРСС, 1996. С. 224.

¹⁶ В. Бринтон. Графическое изображение фактов / Пер. с англ. С. Займовского. М.: Общероссийский общественный фонд «Общественное мнение», 2017. С. 348.

¹⁷ Семёнова А.В., Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения / Под ред. В.А. Мансурова. – М.: Институт социологии РАН, 2010. С. 324.

¹⁸ Горшков М.К., Шереги Ф.Э. Прикладная социология: методология и методы: Учебное пособие/ М.К. Горшков, Ф.Э. Шереги. — М.: Альфа-М: ИНФРА-М, 2009. С. 416.

Выборка за 5 лет позволит проследить тенденции в способах визуализации данных.

Практическая значимость заключается в упрощении способа выбора визуализации при работе с разными типами инфографики (статичной, интерактивной, motion-инфографики), а также повышении уровня визуальной культуры журналиста и дизайнера.

Структура данной работы обусловлена поставленными целями и задачами. **Первая глава** исследования посвящена изучению истории появления и развития журналистики данных, определению понятия больших данных, а также определению и описанию типов информации в больших данных. **Во второй главе** анализируются и описываются способы визуализации больших данных на основании результатов контент-анализа выборки проектов журналистики данных из изданий The Guardian, The New York Times, The Washington Post.

В библиографии представлен список исследований на данную тему на русском и английском языках, а также указаны интернет-источники.

Таблица с полными результатами контент-анализа выборки представлена в **приложении**.

Тезисы исследования прошли апробацию на следующих конференциях:

I Всероссийская студенческая научно-практическая конференция с международным участием «Традиционные и новые медиа: проблемы и перспективы взаимодействия» (2014 г., заочное участие с дальнейшей публикацией научных тезисов);

I Междисциплинарная научно-практическая конференция с международным участием «Сибирское медиапространство 2020» (2014 г., выступление с докладом с дальнейшей публикацией тезисов);

Международная научно-практическая конференция молодых исследователей «Язык, дискурс, (интер)культура в коммуникативном пространстве человека» (2015 г., выступление с докладом, диплом II степени в секции «Актуальные проблемы массовой коммуникации»);

Международный научный форум «Медиа в современном мире. 56-е Петербургские чтения» (2015 г., заочное участие с дальнейшей публикацией научных тезисов).

Положения, выносимые на защиту:

- дата-журналистика — это деятельность по анализу доступных баз данных, обработке полученной в ходе анализа информации и ее интерактивной визуализации с целью информирования читателя;
- проект журналистики данных — это мультимедийный проект, основанный на анализе баз данных и визуализированный с помощью интерактивного интерфейса;
- проект журналистики данных относится к сетевым жанрам, так как он обладает всеми функциями и свойствами присущими жанрам Интернет-СМИ. Одной из задач его является создание целостной картины какого-либо явления, события или предмета, представленного нелинейной системой воспроизведения. Все это дает нам право утверждать, что проект журналистики данных представляет собой мультимедийный текст;
- возможность прогнозирования и анализа современных явлений на основе больших данных позволяет нам утверждать, что журналистика данных (основа которой базы данных) является подвидом аналитической журналистики;
- интерактивная визуализация данных — способ графического представления информации, который позволяет читателю взаимодействовать с системой отображения информации и наблюдать ответную реакцию системы. Проекты журналистики данных почти всегда представлены интерактивной инфографикой;
- при выборе способа визуализации специалист руководствуется типом визуализируемых данных (дискретные, географические или непрерывные временные данные), темой проекта, а также его целью.

ГЛАВА 1. ФЕНОМЕН БОЛЬШИХ ДАННЫХ В СОВРЕМЕННОЙ ЖУРНАЛИСТИКЕ

1. 1 К определению понятия big data

На смену информационной эпохе, по мнению В. Майера-Шенбергера и А. Себранта, грядет эпоха больших данных¹⁹. Это выражается и в финансовых показателях аналитических рынков. Так, по итогам 2013 года коммерческие ЦОД (Центр обработки данных) заработали в России около 456,23 млн долларов, а их доля в объеме всей российской индустрии ИТ—услуг оценивается в 7 процентов²⁰. Отражением смены эпох являются данные International institute for analytics (Международного института аналитики). В 2014 году институт представил прогноз, согласно которому в 2015 году общемировой рынок больших данных достигнет 125 миллиардов долларов.

Несмотря на свою неоднозначность, термин big data имеет вполне конкретную дату рождения — 3 сентября 2008 года. В этот день вышел журнал Nature, создатели которого отвечали на вопрос: «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объемами данных?». Специальный выпуск журнала подводил черту в дискуссии о роли данных в науке.

Роль big data в академической науке обсуждают давно — английский астроном Томас Симпсон писал об обработке данных в труде «О преимуществах использования чисел в астрономических наблюдениях» в XVIII веке. Но только с появлением ЭВМ появилась возможность применять

¹⁹ Большие данные. Революция, которая изменит то, как мы живем... С. 110.; Себрант А. «Что такое Big Data и почему это страшно интересно» [Электронный ресурс] // Малый ШАД (канал на Youtube.com), 2014. Режим доступа: <https://youtu.be/zsUKYfXjpvo> (дата обращения: 14.05.17)

²⁰ Клесова, М. За какие профессии в ближайшее время будут хорошо платить [Электронный ресурс] // Интернет-издание «Рустория», 2014. Режим доступа: <http://sandbox.rustoria.ru/post/za-kakie-professii-v-blizhajshee-vremya-budut-mnogo-platit/> (дата обращения: 14.05.17)

компьютерные методы обработки данных во всех научных сферах. Любые оцифрованные данные могут стать объектом научной или любой другой публикации, а пользователь может самостоятельно проверить подлинность опубликованных в ней фактов. Процесс накопления данных бесконечен, и, осознавая масштабы вызванных этим фактом изменений, Клиффорд Линч предложил новое название для этой парадигмы — Большие Данные. Название было выбрано создателем по аналогии с метафорами «большой хлеб», «большая нефть», которые отражают переход от количества к качеству.

Исследователи McKinsey определяют большие данные как «наборы данных, размер которых выходит за пределы возможностей типичных программных средств для сбора, хранения, управления и анализа». Тем не менее, сами они считают это определение субъективным и недостаточно конкретным.

В книге «Большие данные: опровергая мифы, раскрывая возможности» (Big Data at Work: Dispelling the Myths, Uncovering the Opportunities) Том Дэвенпорт говорит о проблемах в определении больших данных и прогнозирует «относительно короткий срок службы для этого несчастного термина». Дэвенпорт предлагает свое определение: «Широкий спектр новых и массовых типов данных, появившихся за последнее десятилетие или около того»²¹. «Эти данные имеют настолько большие размеры, что при манипуляции и управлении ими возникают значительные материально—технические проблемы»²².

Колумнист журнала Forbes Гил Пресс (Gil Press) в статье с говорящим названием «12 определений больших данных: какое вы выберете?» (12 Big Data Definitions: What's Yours?) приводит 12 определений больших данных,

²¹ Press, G. 12 Big Data Definitions: What's Yours? [Электронный ресурс] / G. Press// Forbes, 2014. Режим доступа: <http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/> (дата обращения: 14.05.17)

²² Oxford Dictionaries - Dictionary, Thesaurus, & Grammar [Электронный ресурс]. Режим доступа: https://en.oxforddictionaries.com/definition/big_data (дата обращения: 14.05.17)

включая свое: «новые инструменты, помогающие нам найти соответствующие данные и проанализировать их последствия»²³.

В. Майер-Шенбергер и К. Кукьера объединяют все вышесказанное и раскрывают определение больших данных через характеризующие их признаки-этапы (количество, точность, причинность). Увеличение объёмов информации привело к тому, что «рассматриваемое количество уже не помещалось в памяти компьютера», и понадобились программы, которые смогли бы обработать эти массивы²⁴. Сегодня термин «большие данные» используется в значении «типы баз данных, в том числе общественно значимые компьютерные данные, которые могут храниться в общем доступе, беспрепятственно использоваться аудиторией».

Охарактеризовали big data через признаки и сотрудники компании Gartner, выделив 3+1 «V»: volume (объем), variety (разнообразие); velocity (скорость); value (польза). Под объемом данных мы понимаем их физическую величину, под скоростью — скорость обновления, обработки информации, а также получения результатов, а под разнообразием — возможность обрабатывать данные разных форматов. Последний признак (value) был добавлен позднее и означает то, что анализ big data должен приносить новые знания, а значит, пользу человеку, который с ними работает. Если данные удовлетворяют только нескольким из первых трех свойств, то они могут быть связаны с классом big data. Поэтому в настоящее время существуют следующие классы: Volume — Velocity, Volume — Variety, Velocity — Variety и Volume — Velocity — Variety.

Большие данные — это явление, которое может быть представлено бесконечным или неограниченным накоплением данных, объем которых экспоненциально растет. Эти данные могут быть представлены разными неструктурированными форматами. В конечном итоге, мы имеем набор

²³ 12 Big Data Definitions: What's Yours?

²⁴ Большие данные. Революция, которая изменит то, как мы живем... С. 110.

данных, который может быть слишком большим, слишком «сырым» или слишком неструктурированным для классических методов обработки. В контексте данной работы, нас интересует не объем, а методы использования big data в современной журналистике.

Тяжело поспорить с тем, что данных с каждым годом становится все больше и больше, но проблема заключается не в их количестве, а в неспособности справляться с такими объемами информации старыми методами. Парадоксально, но способность создавать данные сильнее, чем умение с ними работать. Л. Черняк объясняет это так: «Странно, математики столетиями разбираются с основными понятиями своей науки, такими, как число и системы счисления, привлекая к этому философов, а в нашем случае, данные и информация, отнюдь не тривиальные вещи, оставлены без внимания и отданы на откуп интуитивному восприятию. Вот и получилось, что все эти 65 лет невероятными темпами развивались собственно технологии работы с данными и почти не развивалась кибернетика и теория информации, оставшиеся на уровне 50-х годов, когда ламповые компьютеры использовались исключительно для расчетов»²⁵.

Big data представляет собой статистические данные, которые находятся в постоянном «движении». Такой подход к работе с данными открывает перед нами новую возможность — прогнозировать события и устанавливать связи между ними (корреляции), т.е. анализировать их. Результатом анализа может стать «неявное, скрытое значение информации»²⁶. Возможность прогнозирования и анализа современных явлений на основе больших данных позволяет нам утверждать, что журналистика данных (основа которой базы данных) является подвидом аналитической журналистики. А.А. Тертычный отмечает гибкость аналитических жанров и их постоянную модификацию: «отечественные СМИ выработали достаточно эффективную систему

²⁵ Черняк Л. Большие данные – новая теория и практика // Открытые системы. СУБД. 2011. № 10. С. 18–25.

²⁶ Большие данные. Революция, которая изменит то, как мы живем... С. 110.

аналитических жанров. Система эта не является чем-то раз и навсегда данным – она постоянно развивается, адаптируясь к тем задачам, которые встают перед аналитической журналистикой»²⁷.

Тенденция к использованию больших данных (баз данных) как профессионального направления в журналистике была зафиксирована в 2006 году в США, в 2013 – в России²⁸.

1.2 К определению «журналистики данных» (дата-журналистики)

В разных компаниях датчики, видеокамеры, интеллектуальные счетчики и другие подключенные устройства производят огромные объемы данных, которые добавляются к уже хранящейся на сервере информации. Журналист должен уметь разглядеть в этих огромных массивах данных полезную информацию. Исследование, проведенное в 2013 году по заказу компании Cisco, показало, что ИТ-специалисты и компании «с трудом извлекают пользу из поступающей информации»²⁹. Появление data journalism стало реакцией на проблемы и выходящие из них потребности.

Журналистика данных — достаточно новое явление, поэтому существует большое количество интерпретаций этого термина. Дата-журналистику определяют и как процесс (отсылка к традиционной журналистике), и как набор навыков, как жанр, как рассказ истории с помощью баз данных и как, непосредственно, сами структурированные данные (отсылка к исконному значению data).

Одним из первых определение «дата-журналистике» дал американский журналист Эдриан Головатый в 2006 году: «это структурированные,

²⁷ Тертычный, А.А. Жанры периодической печати [Электронный ресурс] / Учебное пособие. М.: Аспект Пресс, 2000. - Режим доступа: <http://www.evartist.narod.ru/text2/01.htm> (дата обращения: 14.05.17)

²⁸ Big Data, Open Data, Linked Data, метаданные в PR...

²⁹ Большие данные помогут наращиванию ценности ИТ — департаментов. Открытые системы, 2013, № 4. Точка доступа: <http://www.osp.ru/news/2013/0403/13018290/> (дата обращения: 14.05.17)

машиночитаемые данные, используемые вместе с традиционным текстом»³⁰. Данное определение не отражает современные журналистские реалии, поскольку журналистский материал не просто транслируется на компьютере, но и является частью интернет-пространства.

М.Н. Шерстюкова пишет о дата-журналистике как о новом жанре, «использующем для представления информации общественно доступные базы данных (отсюда и название), то есть статистические сводки, графики, списки, карты и многое другое»³¹. Эти данные могут касаться любой сферы общества: политики, образования, экономики, культуры, науки и других.

М.Г. Шилина, в свою очередь, дает широкое определение дата-журналистике, говоря, что это процесс создания конвергентного контента на основе использования массивов компьютерных и интернет-данных. Но журналистику данных можно и нужно понимать еще шире: как «набор специфических навыков для поиска, анализа, визуализации информации цифровых источников метаданных для формирования интерактивных форматов уникальной подачи авторского аналитического контента и эффективного взаимодействия с аудиторией; это формат актуального журнализма, формат медиатекста (медиаконтента), метод его создания, трансляции, потребления, который может быть использован как метаметод и метаоснова для иных жанров».

Журналистика данных подразумевает рассказ историй с помощью цифр, баз данных, с обязательной, понятной визуализацией. Принцип журналистики

³⁰ Шилина, М.Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакommunikации: к вопросу формирования теоретических исследовательских подходов [Электронный ресурс] / М.Г.Шилина // Журнал Факультета журналистики МГУ имени М.В. Ломоносова «Медиаскоп», 2013. - №1. – Режим доступа: <http://www.mediascope.ru/node/1263> (дата обращения: 14.05.17)

³¹ Шерстюкова, М.Н. Дата-журналистика как новое направление в системе средств массовой коммуникации [Электронный ресурс] / М.Н. Шерстюкова // Международный электронный научно-образовательный журнал «Медиа. Информация. Коммуникация». – 2012. -№1. – Режим доступа: <http://mic.org.ru/1-nomer-2012/65-1-shestukova-2> (дата обращения: 14.05.17)

данных — материал строится не вокруг новостного повода или явления, а вокруг статистики, цифр, сводок, отчетов и иной справочной информации³².

А.Л. Николов в статье «Новые направления в медиаиндустрии, или что такое дата-журналистика (data journalism)» выдвигает следующие требования к проектам дата-журналистики³³:

1. автор материала должен обработать большие массивы информации с помощью ЭВМ (для этого журналист должен владеть основами программирования);
2. в результате обработки мы имеем материал, не имеющий отношения к традиционной журналистике, – база данных, график, карта и т.д., который имеет интерактивное расширение;
3. итоговый продукт потребитель может использовать в соответствии со своими индивидуальными потребностями и интересами;
4. результатом проекта является знание, полученное потребителем самостоятельно (журналист, в данном случае, является проводником к этому знанию), это знание является общественно значимым и «распространяется с помощью традиционных СМИ или другим массовым общедоступным способом».

Журналистику данных исследователь определяет, как процесс создания интерактивного проекта. Такое определение, хоть и максимально приближенно к реальности, не является корректным. Журналистика данных и проект журналистики данных не являются синонимами. Они соотносятся как деятельность и продукт, созданный в результате этой деятельности.

³² Симакова, С.И. Журналистика данных как современное направление журналистики // С.И. Симакова // Научный журнал «Знак»: проблемное поле медиаобразования, 2013. - №1 (11). С. 52-56.

³³ Николов А.Л. Новые направления в медиаиндустрии, или что такое дата-журналистика (data journalism) [Электронный ресурс] / А.Л. Николов // Международный электронный научно-образовательный журнал «Медиа. Информация. Коммуникация», 2012. Режим доступа: <http://mic.org.ru/index.php/new/165-novye-napravleniya-v-mediaindustrii-ili-chto-takoe-data-zhurnalistika-data-journalism> Дата обращения: 14.05.17

На наш взгляд, разнообразие в интерпретации термина вызвано тем, что, во-первых, журналистика данных является новым направлением, во-вторых, каждый исследователь рассматривает дата-журналистику через призму своих личных, научных интересов. Мы же даем ей следующее определение:

Дата-журналистика – это деятельность по анализу доступных баз данных, обработке полученной в ходе анализа информации и ее интерактивной визуализации с целью информирования читателя. Итоговый продукт, создаваемый в ходе этой деятельности, – проект журналистики данных.

В свою очередь, проект журналистики данных – это мультимедийный проект, основанный на анализе баз данных и визуализированный с помощью интерактивного интерфейса.

1.3 Типы данных в проектах дата-журналистики

Общее, что объединяет все вышеперечисленные определения big data, это использование баз данных как источник информации. Максимально точное и юридически закрепленное определение этому термину мы нашли в Гражданском кодексе РФ. Часть четвертая, раздел VII «Права на результаты интеллектуальной деятельности и средства индивидуализации», глава 70 «Авторское право», 1.6. Статья 1260 «Переводы, иные производные произведения. Составные произведения»: «Базой данных является представленная в объективной форме совокупность самостоятельных материалов (статей, расчетов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины (ЭВМ)»³⁴.

³⁴ Гражданский кодекс Российской Федерации: в 4 ч.: по состоянию на 1 февр. 2010 г. – Москва: Кнорус, 2010. С. 540.

В проведенном контент-анализе, подробно описанном во 2 главе, мы выделили три типа информации, которые встречаются в визуализации big data в The Washington Post, The New York Times и The Guardian — дискретные, непрерывные временные и географические данные.

Дискретные данные. Набор данных называется дискретным, если значения, принадлежащие ему, различные и отдельные, и могут быть подсчитаны. Примеры дискретных данных: цвет глаз, или количество отжиманий; пациентов в хирургии, врачей и др.). В проекте The Guardian (Рисунок 1) «Права геев в США, по штатам» (Gay rights in the US, state by state)³⁵ дискретные данные представлены законами о правах лиц с нетрадиционной ориентацией. Если дискретные данные — это данные, ограниченные во времени и в пространстве, то непрерывные данные — это данные, которые не имеют временных и пространственных границ.

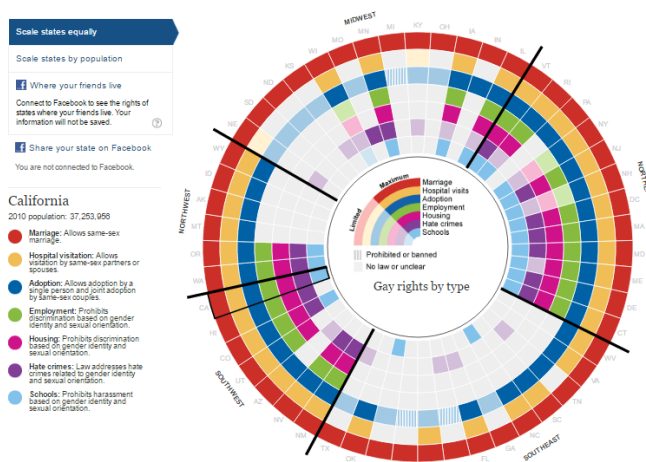


Рисунок 1 — Визуализация проекта Gay rights in the US, state by state

Непрерывные временные (темпоральные) данные. В широком смысле, «это данные, которые явно или неявно связаны с определенными датами или промежутками времени»³⁶. Под это определение могут попасть почти любые

³⁵ The Guardian [Электронный ресурс]. Режим доступа:

<https://www.theguardian.com/world/interactive/2012/may/08/gay-rights-united-states>

(дата обращения: 14.05.17)

³⁶ Туманов В.Е. Проектирование хранилищ данных для систем бизнес-аналитики: учебное пособие / В.Е. Туманов. — М.: Интернет-Университет Информационных Технологий: БИНОМ. Лаборатория знаний, 2010. С. 615.

данные и информация. У любой информации есть явная, или неявная зависимость от времени: например, время, когда системе стало известно, что данный факт существует. Существует вероятность, что факт изменится в какой-либо промежуток времени, поэтому его нельзя воспринимать как абсолютную истину.

К темпоральным базам данных относят данные, которые изменяются с течением времени. Можно выделить два вида данных для отображения времени: время фиксации определенного события или факта и время выполнения какого-либо действия или операции. Каждая запись представляет собой факт, который является истинным в определенный интервал или момент времени. Представление времени подобным образом (с точки зрения моделируемого мира) называется *valid time* («временем фиксации факта»). Представление подобным образом данных — интервальным. Например, в проекте The New York Times «Как год рождения влияет на политические взгляды?» (How Birth Year Influences Political Views)³⁷: факт про политические предпочтения респондента является истинным только в определенный интервал времени с 1937 по 1994 год (Рисунок 2).

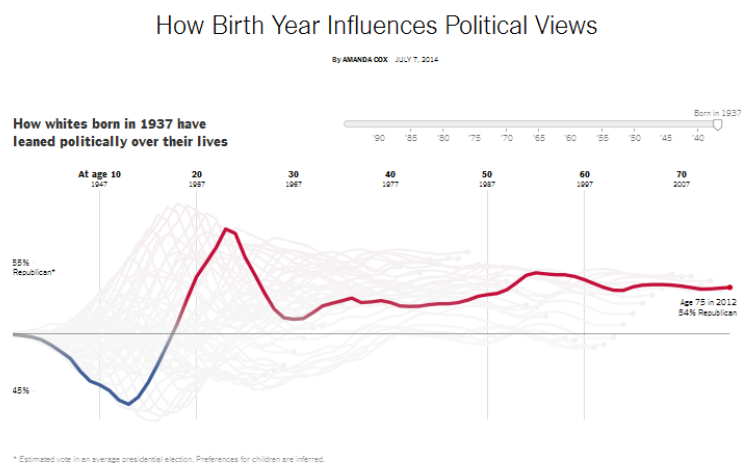


Рисунок 2 — Визуализация проекта How Birth Year Influences Political Views

³⁷ The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2014/07/08/upshot/how-the-year-you-were-born-influences-your-politics.html> (дата обращения: 15.04.17)

Еще один тип времени, который относится к темпоральным базам данных — это транзакционное время или время операции. Это промежуток времени, за который данные были внесены в базу данных, а потом удалены из нее. Помимо этого, можно рассматривать отдельный момент времени и те факты, которые были истинны в этот конкретный момент. Такой способ называется точечным представлением. Например, как в проекте The Washington Post «Убийства в округе» (Homicides in the District). Проект представляет собой базу с убийствами с 2000 по 2011 год, изучив которую пользователь может узнать количество убийств в отдельном районе Вашингтона, статус дела — открыто оно или закрыто, манеру убийства, мотивацию, а также возраст и расу преступника (Рисунок 3).

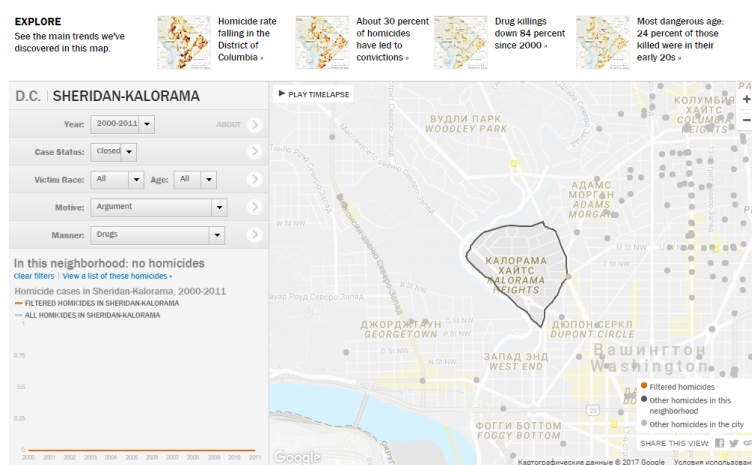


Рисунок 3 — Визуализация проекта Homicides in the District, точечная карта

В этом проекте речь идет о представлении времени с точки зрения пользователя, то есть о моделях, вокруг которых могут формулироваться запросы и их результаты. При использовании любого из этих представлений — точечного или интервального — истинность фактов не меняется, но в случае точечного представления мы получаем срез всех фактов на определенный момент времени, а для интервального представления нас волнует конкретный факт и периоды его истинности.

Во временных рядах принято выделять следующие компоненты, на которые могут составить основу для аналитики проекта³⁸:

1. Тренд — динамика, характеризующее общее развитие временного ряда;
2. Циклическая компонента — динамика, имеющая фазу возрастания и убывания, период которой занимает достаточно большой промежуток времени;
3. Сезонная компонента — это регулярные колебания уровней ряда в определенное время суток, недели, сезона и т. д. Связаны с сезонными явлениями (например, погодными условиями) и человеческими ритмами (например, с фазами бодрствования и сна);
4. Календарные эффекты — скачки временного ряда, связанные с некоторыми предсказуемыми календарными событиями (например, праздниками или выходными);
5. Аномальные явления (выбросы) — непредсказуемые скачки, приводящие к резким, но кратковременным отклонениям ряда от общей тенденции развития;
6. Структурные сдвиги — непредсказуемые скачки, приводящие к отклонениям ряда от общей тенденции развития, которые сказываются на всем его дальнейшем поведении;
7. Случайная компонента — беспорядочные движения достаточно большой частоты, связанные с влиянием большого количества неизвестных факторов.

Несмотря на то, что в случае с непрерывными временными данными измерения в теории могут производиться непрерывно, на практике фиксирование результатов измерений все равно происходит дискретным образом.

³⁸ Суслов В.И., Ибрагимов Н.М., Талышева Л.П., Цыплаков А.А. Эконометрия. Часть III Эконометрия - I: Анализ временных рядов — Учебное пособие — Новосибирск: Изд-во СО РАН, 2005. С. 744.

Географические данные. Геопространственные данные (географические данные) — «информация, которая идентифицирует географическое местоположение и свойства естественных или искусственно созданных объектов, а также их границ на земле. Эта информация может быть получена с помощью дистанционного зондирования, картографирования и различных видов съемок»³⁹. Географическое размещение объектов в пространстве визуализируется 2-х, 3-х или 4-х мерными координатами в географической системе координат. Объектом ГИС является цифровое представление целого или части реального объекта.

Гурьянова Л.В. выделяет четыре географических интегрированных компонента: местоположение, свойства и характеристики, пространственные отношения и время. Таким образом, в геоинформационной системе данные представлены двумя категориями: пространственные (местоположение), непространственные (атрибуты). Пространственные данные могут включать географические объекты, которые представлены точками, линиями или полигонами, то есть векторным способом. В реальном мире точки могут представлять собой деревья, дома или несколько объектов. Дугами являются те реальные объекты, которые можно рассматривать как линии: например, дорога, река, транспортная инфраструктура и др. Полигоны — это замкнутые области, которые представляют однородные участки. Полигонами обозначают типы почв, избирательные округа, земельные участки или контуры зданий. Способ представления пространственных объектов с помощью ячеек и сетки называется растровым.

Описательная информация пространственных объектов представлена атрибутами (свойствами) объектов. Атрибутивные данные могут включать идентификатор объекта, любую описательную информацию из баз данных,

³⁹ Гурьянова Л. В. Аппаратно-программные средства ГИС: Курс лекций. - Минск: БГУ, 2003. С. 140.

изображение и многое другое. Иногда атрибуты называют непространственной информацией.

Данные в геоинформационных системах хранятся в виде набора тематических слоев, объединенных географическим положением. Благодаря этому ГИС работает как с векторными, так и с растровыми моделями данных. Это способствует принятию эффективных решений задач, касающихся пространственной информации.

Про зависимость визуализации от типа информации писал директор по визуальным коммуникациям компании McKinsey Джин Железны⁴⁰. Сам процесс выбора способа визуализации он описывает следующим образом: сформулировать идею, определить тип сравнения данных и выбрать тип диаграммы.

Исследователь выделили следующие типы данных в зависимости от цели визуализации:

Отношения в данных — это знание о том, как данные зависят друг от друга, какая связь между ними. С помощью отношений в данных можно понять, есть ли зависимость между переменными в данных;

Распределение данных — это знание о том, как данные располагаются относительно чего-либо, и сколько объектов находится в определенных последовательных областях числовых значений;

Композиция данных — объединение данных с целью анализа общей картины, сравнения компонентов, составляющих процент от некоего целого;

Сравнение данных — объединение данных с целью сравнения некоторых показателей, выявление того, как объекты соотносятся друг с другом. Кроме того, это сравнение компонентов, изменяющихся с течением времени.

Впоследствии работа Джина Железны стала фундаментом для создания самых разных классификаций способов визуализации данных.

⁴⁰ Говори на языке диаграмм... С. 144.

1. 3 Мультимедийный проект как форма представления информации в журналистике данных

1. 3. 1 Отличительные свойства интернет-СМИ

Журналистика находится в прямой зависимости от стремительно развивающихся информационных технологий (ИТ) и таких процессов, как конвергенция и цифровизация (дигитализация). В частности, с развитием ИТ зародилась и, опираясь на современные тенденции, трансформируется онлайн-журналистика.

Для общего понимания процесса конвергенции обратимся к работам Е.Л. Вартановой. Конвергенция (от лат. *converge* — приближаюсь, схожусь) – термин, принятый в естественных и общественно-политических науках для обозначения процессов схождения, взаимоуподобления. В медиаэкономике используется для обозначения интеграции информационных и коммуникационных технологических платформ (компьютеров, телевизоров, телефонов), а также передаваемых ими содержательных (информационных) продуктов⁴¹.

Е.Л. Вартанова говорит о трех уровнях конвергенции — слиянии технологий, слиянии различных и разобщённых медиа и слиянии рынков. Примерно так же обозначает термин конвергенция американский ученый Итьеля де Сола Пула: «Стирание границ между медиа как средствами обоюдной коммуникации, такими как телефон, почта, телеграф как средствами массовой коммуникации, такими как пресса, радио и телевидение»⁴².

Конвергенция на технологическом уровне позволяет разными способами доставлять информацию потребителю: кабельными или

⁴¹ Теория СМИ... С. 244.

⁴² Качкаева, А.Г. Журналистика и конвергенция: почему и как традиционные СМИ превращаются в мультимедийные / под ред. А.Г. Качкаевой. – М.: 2010. С. 200.

телефонными сетями, беспроводной спутниковой связью. Цифровой формат контента позволяет распространять его в различных формах и на различных носителях. На втором уровне говорится о разных медиаплатформах, которые сейчас используют СМИ — никого не удивишь тем, что радиостанция имеет свое телевидение, или печатное издание занимается радиовещанием. На третьем уровне, по мнению автора, происходит зарождение нового интегрированного рынка, на котором «неразрывными связями скрепляются мультимедийные услуги, сетевое обслуживание, создание программных продуктов»⁴³.

Норвежские ученые Андерс Фагерйорд и Танья Сторсул, в свою очередь, выделили шесть интерпретаций понятия медиаконвергенция: конвергенция сетей, терминалов, жанров и форм, услуг, рынков и регулирования. Непосредственно проект журналистики данных является проектом, полученным в результате конвергенции разных жанров и форм. «Конвергенция жанров и форм предполагает, что в результате соединения различных медиаплатформ (например, печатных СМИ с телевизионными на базе интернет-порталов) жанры, ранее свойственные какой-либо одной медиаплатформе, проникают и ассимилируются с другими»⁴⁴.

Катализатором для структурных изменений, происходящих в СМИ и в медиаиндустрии стали изменения в технологической сфере. На ряду с конвергенцией этому способствовал процесс цифровизации контента. Цифровизация содержания (с англ. дигитализация) — перевод всех типов содержания СМИ в цифровой формат, который позволяет легко распространять информацию по любому каналу электронной коммуникации.

Использование различных форматов, медийных платформ и электронных каналов для передачи информации является отличительной

⁴³ Теория СМИ... С. 245.

⁴⁴ Журналистика и конвергенция... С. 13.

особенностью интернет-СМИ и проявляется в таких качествах как гипертекст, интерактивность и мультимедийность.

Впервые термин «гипертекст» был использован американским исследователем Тедом Нельсоном в 1960-ом году в значении «системы связей между отдельными документами с помощью встроенных в текст гиперссылок»⁴⁵. Гипертекст использовался в периодике и представлял собой ссылку на другой материал или издание («как сообщает газета...» или «как мы писали ранее»). Спустя пятьдесят лет значение термина не изменилось, однако, гипертекстуальность стала неотъемлемой частью интернет-пространства.

«Гипертекст — это представление информации как связанной (linked) сети гнезд (nodes), в которых читатели свободны прокладывать путь (navigate) нелинейным образом. Он допускает возможность множественности авторов, размывание функций автора и читателя, расширение работы с нечеткими границами и множественность путей чтения»⁴⁶. Благодаря гипертекстуальности как у создателя текста, так и у его потребителя есть возможность структурировать большой поток информации удобным для себя способом.

Гипертекст, который потенциально состоит из бесконечного множества текстов, объединенных системой встроенных гиперссылок, позволяет читать его не только горизонтально, как на бумаге, но и с помощью внутренних и внешних ссылок просматривать «вглубь». Использование гипертекста в журналистике придает материалам третье измерение — глубину⁴⁷.

Интерактивность в интернет-журналистике имеет большой спектр значений. В широком смысле под интерактивностью понимается

⁴⁵ Теория и практика... С. 56.

⁴⁶ Калмыков, А. А. Интерактивная гипертекстовая журналистика в системе отечественных СМИ. Научное издание / [Электронный ресурс] / А.А. Калмыков, под ред. В.С. Хелемендика. - М: Издательство ИПК работников ТВ и РВ, 2009. Режим доступа: <http://www.evartist.narod.ru/text19/159.html> (дата обращения: 15.05.17)

⁴⁷ Теория и практика...С. 58.

«взаимонаправленная коммуникация СМИ и пользователей»⁴⁸. По аналогии с гипертекстуальностью, есть ранние формы воплощения интерактивности в СМИ: например, письма в редакцию или звонки в студию, которые до сих пор являются популярными форматами теле- и радиопередач. Но с появлением интернет-СМИ коммуникационная функция интерактивности перешла на новый уровень — быстрого или даже simultанного реагирования пользователей на публикуемый контент.

В узком смысле, «интерактивное медиа — это медиа, которое читатель создает как бы самостоятельно на основе бесчисленного предложенного количества вариантов. В этом состоит принципиальное отличие языка мультимедиа от языка других сред — без вмешательства пользователя информация остается неполной». Существуют разные методы чтения мультимедийного продукта⁴⁹.

1. Линейная передача информации. Например, сеанс трансляции (потребитель может только остановить просмотр); сеанс чтения (можно прервать чтение, но вернуться к нему в любой момент);

2. Нелинейная передача информации. К примеру, навигация (перемещение по заранее запланированному маршруту); изучение (индивидуальное перемещение без заданного маршрута); виртуальный визит (пользователь полностью руководит процессом чтения материала).

1. 3. 2 Мультимедийность как свойство интернет-СМИ

Впервые термин «мультимедиа» был использован в 1965 году американским шоуменом Бобом Гольдштейном. Он применил его, чтобы описать поставленное им шоу, которое включало в себя световое оформление,

⁴⁸ Там же.

⁴⁹ Журналистика и конвергенция...С. 45.

кино и музыку⁵⁰. Сегодня же, говоря про мультимедийность, подразумевают способ представления информации с помощью различных медийных платформ: вербального текста, фотографии, аудио-, видео-, графики, анимации и других производных от них форм.

Несмотря на то, что традиционные медиа принято относить к мономедийным, разные форматы информации встречаются и в них. В печати – это не только вербальная информация, но и визуальная, а на радио и телевидении – не только звук и картинка, но и текст. Однако у создателя мультимедийного текста в Интернете есть возможность использовать все знаковые системы в одном материале.

М.Г. Шилина мультимедийность обозначает как «комплексное конвергентное представление информации в любом цифровом формате (условно-вербальном, условно-визуальном, условно-аудиальном и т.д.), которое позволяет создавать уникальный тип контента, применимый на любом канале медиакommunikации»⁵¹. Мы говорим о мультимедийности как о визуальной реализации идеи медиаконвергенции. Различные элементы различных языков (письменный текст, аудио – и видеоиллюстрации) могут взаимодействовать между собой разными способами⁵²:

- 1) методом дополнения (например, когда фотолента дополняет текст);
- 2) методом построения иерархии информации (менее важная информация представлена в другой форме);
- 3) методом навигации (другая форма информации в мультимедийном проекте упрощает навигацию по информационному пространству).

Онлайновые СМИ обладают как общими для всех СМИ свойствами, – производят уникальный социально значимый контент, имеют официальный

⁵⁰ Медиатренды [Электронный ресурс]: экспертный взгляд факультета журналистики МГУ имени М. В. Ломоносова на события в СМИ / – Электрон. журн. - факультет журналистики МГУ, 2011. - Режим доступа: http://www.mediascope.ru/files/MediaTrends_%2019.pdf (дата обращения: 14.05.17)

⁵¹ Data Journalism – дата-журналистика, журналистика метаданных... Режим доступа: <http://www.mediascope.ru/node/1263> (дата обращения: 14.05.17)

⁵² Журналистика и конвергенция...С. 56.

статус, имеют профессиональную редакцию, регулярно обновляются, ориентированы на массовую аудиторию (рассредоточенную и анонимную), используют традиционные журналистские жанры и рубрики — так и уникальными свойствами: гипертекстуальностью, интертекстуальностью и мультимедийностью. В интернет-СМИ изменилась подача и «упаковка» текстов — она стала мультимедийной, а требования к качеству информации остались прежними.

Уникальное свойство Интернета — передавать информацию с помощью разных медийных платформ — является предпосылкой для возникновения новых журналистских жанров, в которых для передачи информации используются разные знаковые системы.

Мультимедийный проект — журналистский материал, в котором тема раскрыта с помощью различных медийных платформ: текстовых и аудиовизуальных средств, единство которых создает объемную картину события, явления или предмета⁵³. При этом способ передачи информации — набор и конфигурация медийных платформ — может быть разная, как и возможность управления информацией со стороны читателя. М.М Лукина предлагает следующие технологические платформы мультимедийного текста:

- 1) текст в виде обычной html-страницы, в который инкорпорированы мультимедийные элементы;
- 2) интерактивная инфографика в основе которой лежит сюжет, пользователь получает возможность выбирать разные элементы и самостоятельно «передвигаться» по этому сюжету;
- 3) плеер, построенный на технологии flash или silverlight. Такой плеер может стать мультиэкраном, прототип которого мы можем увидеть на крупных новостных каналах и в Интернете.

В мультимедийном журналистском материале в разных сочетаниях могут быть использованы все семь медийных платформ — вербальный текст,

⁵³ Теория и практика... С. 160.

графика, инфографика, видео, аудио, анимация, фотография. А конфигурации этих платформ могут варьироваться от мономедийных (используется одна медийная платформа) до мультимедийных сочетаний (используется две или более медийных платформы), представленных в цельном журналистском произведении.

Давайте обратимся к основным уникальным жанрам интернет-СМИ, которые возникли благодаря конфигурации внутренних систем навигации и сочетанию медийных платформ⁵⁴. По приоритетности элементов внутри платформы их можно разделить на три группы — производные от статичной иллюстрации, от аудио- и видеозаписи.

1. Фотоленты, фоторепортажи, фотогалереи — повествование о событии в фотографиях. Это может быть, как последовательное расположение кадров друг за другом, так и хаотично оформленные фотографии по общей теме;
2. Слайд-шоу — автоматическая смена кадров прямо внутри иллюстрации;
3. Карикатура — отличается от традиционной карикатуры наличием анимации, звука и других интерактивных элементов;
4. Подкаст — звуковой файл, выложенный в Интернет, представляющий собой начитанный текст или фрагмент программы;
5. Видеосюжет — законченное по смыслу и структуре видеообщение;
6. Интерактивная видеоклонка, интерактивный видеосюжет — это видео, в тело которого встроены ссылки на другие мультимедийные элементы;
7. Мультискрипт (interactive video) — новый формат подачи информации, который упрощает навигацию по видеоматериалам большого объема. Все блоки синхронизированы. При просмотре какого-либо фрагмента видеозаписи в двух других блоках отражаются соответствующие фрагменты стенограммы и заголовок тематического блока.
8. Интерактивные игры (интерактивный инфотейнмент) — викторины, турниры flash-игры.

⁵⁴ Там же.

Конвергенция разных жанров и форм в онлайн-журналистике привела к созданию уникального формата журналистского материала — мультимедийного текста. Особый интерес в контексте нашей работы представляет такой интернет-жанр, как инфографика, которая аналитична и основана чаще всего на большом объеме данных. Совокупность данных свойств позволяет нам относить инфографику к журналистике данных.

«Что же отличает журналистику данных от остальной журналистики? Возможно, это новые возможности, которые открываются, когда вы объединяете традиционный “нюх на новости”, умение выведать все, что случилось, со способностью рассказать захватывающую и наглядную историю, с настоящим масштабом и разнообразием цифровой информации, которая ныне доступна»⁵⁵.

А.Г. Качкаева, на конференции, организованной РИА Новости («Как развивается «журналистика данных»») отметила, что работа с большими данными — это новый формат работы. Если до этого процесс работы над материалом можно было отобразить в следующей цепочке: поиск — проверка — публикация, то в дата-журналистике — это цепочка представлена в следующем виде: сбор — фильтрация — визуализации. Конечным продуктам журналистики данных является проект журналистики данных.

Проект журналистики данных относится к сетевым жанрам, так как он обладает всеми функциями и свойствами присущими жанрам Интернет-СМИ, в том числе и конвергентной передачей информации. Одной из задач проекта журналистики данных является создание целостной картины какого-либо явления, события или предмета, представленного нелинейной системой воспроизведения. Все это дает нам право утверждать, что проект журналистики данных представляет собой мультимедийный текст.

⁵⁵ Грэй Дж., Чемберс Л. Пособие по журналистике данных [Электронный ресурс] / Грэй Дж., Чемберс Л. // Сетевое издание «РИА Новости», 2013. Режим доступа: <http://ria.ru/files/book/site/index.html> (дата обращения: 15.05.17)

1. 3.3 Интерактивная инфографика как способ визуализации big data

Интерактивная визуализация данных — способ графического представления информации, который позволяет читателю взаимодействовать с системой отображения информации и наблюдать ответную реакцию системы. Данный способ используется для анализа изменений тенденций, взаимосвязей, соотношений и закономерностей в исследуемом наборе однородных параметров предметов или явлений. Данную технологию еще называют интерактивной инфографикой⁵⁶.

Интерактивная визуализация может использовать с помощью таких подходов, как масштабирование (увеличение и уменьшение масштаба), обзор и детализация, панорамирование, «рыбий глаз» и др. Шаги для интерактивной визуализации следующие⁵⁷:

1. Выбор. Интерактивный выбор объектов данных (группы объектов, части или всего набора данных в соответствии с интересом пользователя).

2. Связывание: полезно для связи информации между несколькими представлениями.

3. Фильтрация. Она помогает пользователям регулировать объем информации для отображения. Это уменьшает количество информации и фокусируется на информации, представляющей интерес.

4. Реорганизация или переназначение. Поскольку пространственная компоновка является наиболее важным визуальным отображением, переупорядочение пространственной структуры информации очень эффективно в плане получения различных сведений.

⁵⁶ Шиманский П. Интерактивная инфографика с анимациями CSS и SVG [Электронный ресурс] / П. Шиманский // Хабрахабр – 2013. Режим доступа: <http://habrahabr.ru/post/168681/> (дата обращения: 14.05.17)

⁵⁷ Khan M., Khan S.S. Data and Information Visualization Methods and Interactive Mechanisms: A Survey, International Journal of Computer Applications, 34(1), 2011. P. 1-14.

Д. Лэнкоу выделяет три вида инфографики⁵⁸: статическую, motion-инфографику и интерактивную.

1. Статическая инфографика — фиксированная информация, изображение статично. Взаимодействие пользователя с данными заключается в их просмотре и чтении. Лучше всего работает как повествование, но может быть и исследовательской.

2. Motion-инфографика — обычно фиксированная информация. Взаимодействие с пользователем состоит из просмотра, прослушивания и если есть выбор, то чтения данных. Изображение анимированное или перемещается. Лучше всего работает как повествование, почти никогда как исследование (за исключением использования в сочетании с интерактивным контентом).

3. Интерактивная инфографика. Информация может быть фиксированной, или динамичной. Взаимодействие с пользователем осуществляется с помощью щелчка, поиска определенных данных, активного формирования отображаемого содержимого и выбора информации, к которой осуществляется доступ и визуализация. Может быть и повествованием, и исследованием.

Интерактивная инфографика позволяет пользователям самостоятельно исследовать набор данных, отображая информацию при наведении указателем мыши, предоставляя заранее классифицированные данные, используя панорамирование и масштабирование. Часто интерактивная инфографика создается специальными отделами новостных организаций.

«Технология интерактивной визуализации позволяет исследовать данные при помощи манипуляций с графическими изображениями, цветом, яркостью, размером, формой и передвижением визуальных объектов, представляющих аспекты анализируемых данных. Эти продукты

⁵⁸ Lankow, J. Infographics: The Power of Visual Storytelling / J. Lankow, J. Ritchie, R. Crooks — Wiley, 2012. P. 264.

предоставляют множество вариантов визуализации, которые выходят за рамки круговых диаграмм, линейных диаграмм и гистограмм, включая тепловые карты, древовидные схемы, географических карты, точечные диаграммы и других специальных визуальных объектов. Эти инструменты позволяют пользователям анализировать данные, взаимодействуя с визуальной репрезентацией данных»⁵⁹.

Интерактивная инфографика может быть представлена тремя формами: статическим или движущимся изображением и интерактивным интерфейсом. Проект журналистики данных почти всегда представлен интерактивной инфографикой с интерактивным интерфейсом (исходя из проведенного нами контент-анализа, 89% от общего числа проанализированных проектов интерактивны).

1.3.4 Сторителлинг как способ визуализации в проектах big data

Статическая визуализация давно используется для поддержки повествования с помощью диаграмм и графиков, встроенных в большой текст. Это формат текста, который рассказывает историю, обычно подтвержден данными или связанными с ними деталями. Подобный класс визуализаций пытается объединить повествование с интерактивной графикой.

Оксфордский словарь английского языка определяет «повествование» как «учет ряда событий, фактов и др., данных в нужном порядке и с установлением связей между ними»⁶⁰. Центральным понятием в этой связи является понятие «причинно-следственных событий». Такие истории часто имеют начало, середину и конец⁶¹. Стратегии сторителлинга могут различаться

⁵⁹ Электронный словарь Gartner IT Glossary [Электронный ресурс] // руководство к изучению терминов и определений из области информационных технологий, 2015. Режим доступа: <http://www.gartner.com/technology/home.jsp>. Дата обращения: 15.05.17.

⁶⁰ Oxford Dictionaries - Dictionary, Thesaurus, & Grammar [Электронный ресурс]. Режим доступа: <https://en.oxforddictionaries.com/definition/narrative> (дата обращения: 14.05.17)

⁶¹ Bordwell D. and Thompson K. Film Art: An Introduction. McGraw-Hill, 2003. P. 544.

в зависимости от медиа и жанра. Например, письменные рассказы, имеют доступ к одним описательным механикам (например, поток сознания), а истории, передаваемые через фильм — к другим (например, композиция с разделенным экраном).

Бланделл⁶² описывает повествовательные техники в журналистике: такие, как анонсирующий лид — начальную историю, которая может включать диалог между персонажами, своеобразный микромир в более крупном новостном сюжете; ореховый граф (nut graph) — абзац, прямо описывающий новостную ценность статьи. Такие приемы во многом уникальны для журналистики.

Хотя визуализацию данных часто сравнивают со сторителлингом, отношения между ними сформулированы нечетко. Одна из первых вещей, которую следует понять, — это различие между «данными» и «историей». Прежде чем выбрать определенный способ визуализации, данные проходят процедуру отбора, структурирования и сортировки, после чего становятся понятнее. Это открывает путь к созданию целостной истории. «Сторителлинг — это интерактивная форма коммуникации, в которой информация переводится в контекст, который люди могут понять, запомнить, обсудить и рассказать другим. В этом смысле есть два аспекта использования визуализации в качестве формы презентации в новостных материалах:

- 1) Исследовательский. Знакомство с данными и их анализ;
- 2) Коммуникативный. Для отображения и обсуждения визуализации с общественностью и создания истории, основанной на данных и фактах.

Джонатан Харрис, создатель We Feel Fine и Whale Hunt, считает себя в первую очередь рассказчиком, а во вторую - разработчиком визуализации: «Я думаю, что люди начали забывать, как сильны человеческие истории, пока они обменивались своими чувствами к фетишистскому увлечению данными,

⁶² W. E. Blundell. The Art & Craft of Feature Writing. Plume, 1988. P. 544.

сетями, шаблонами и информации в целом. Действительно, данные – это только часть истории. Человеческий материал — это главный материал, и данные должны его обогащать».

Однако само понятие «истории» Харрис раскрывает достаточно обобщенно: «Я объясняю “историю” довольно свободно. Для меня история может быть столь же мала, как жест или такой же большой, как жизнь. Но основные элементы истории, вероятно, можно суммировать с хрестоматийными Who / What / Where / When / Why / How»⁶³.

Американские исследователи Сегель и Хир разработали структуру, которая предлагает стратегии дизайна для нарративной визуализации в сторителлинге в СМИ:

1. Структура стакана Мартини: автор управляет вниманием пользователя только в начале, стартовой визуализации, и как только намеченный рассказ завершен, пользователь может свободно исследовать данные.

2. Интерактивное слайд-шоу: визуализация представлена в виде обычного слайд-шоу, где пользователь может взаимодействовать с определенными точками презентации, прежде чем переходить к следующему этапу.

3. Drill-Down Story: пользователь может взаимодействовать с определенными точками визуализации, чтобы выявить дополнительные детали и информацию. Тем не менее, для определения того, где будут происходить взаимодействия, по-прежнему требуется высокий уровень участия автора.

Рассказать историю попробовали в The Guardian, используя 6 интерактивных линейных графиков в проекте «Темная сторона комментариев

⁶³ Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. IEEE Transactions on Visualization and Computer Graphics, 16 (March), 2010. P. 1139–1148.

The Guardian» (The dark side of Guardian comments)⁶⁴. The Guardian заказало исследование 70 миллионов комментариев, оставленных на его сайте с 2006 года, и обнаружило, что из 10 самых раскритикованных писателей восемь — женщины, а двое мужчин — афроамериканцы. В основе проекта шесть линейных графиков, в которых показана зависимость негативных отзывов, количество статей в каждом из разделов, заблокированных пользователей и других категорий от гендерной и расовой принадлежности авторов. Например, в некоторых разделах издания гендерный разрыв больше, чем в других. В Sports, World News и Technology было наименьшее количество статей, написанных женщинами-писателями. Единственный раздел, в котором было значительно больше статей, написанных женщинами — это мода. В целом, отношение к авторам-женщинами в издании более критично, чем к мужчинам (Рисунок 4).

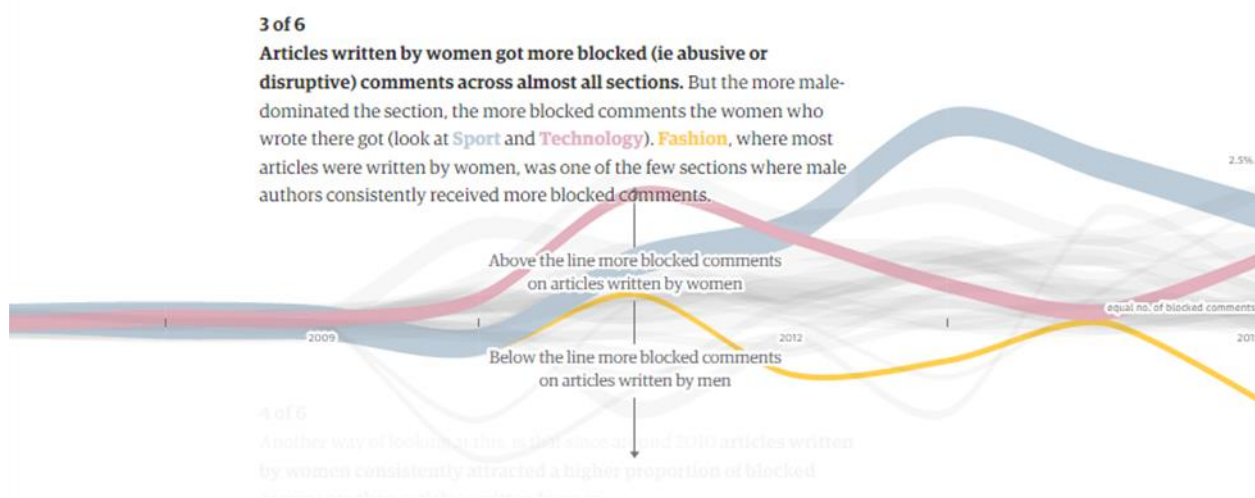


Рисунок 4 — Визуализация проекта The dark side of Guardian comments

Или проект In (International Lunar Network) и Guardian «В полете» (In flight)⁶⁵, в котором пользователю через визуализацию данных рассказывают 100-летнюю историю пассажирских авиаперевозок. Этот проект совмещает в

⁶⁴ The Guardian [Электронный ресурс]. Режим доступа:

<https://www.theguardian.com/technology/2016/apr/12/the-dark-side-of-guardian-comments>

(дата обращения: 15.05.17)

⁶⁵ The Guardian [Электронный ресурс]. Режим доступа:

<https://www.theguardian.com/world/ng-interactive/2014/aviation-100-years> (дата обращения:

15.05.17)

повествование с интерактивными диаграммами и архивными фотографиями (Рисунок 5).

В настоящее время наиболее сложные средства визуализации сосредоточены на исследовании и анализе данных. Приложения, такие, как электронные таблицы и средства визуализации, поддерживают массив процедур анализа и визуального кодирования, но помимо экспортирования изображений для презентации обычно обеспечивают небольшую поддержку для создания историй с результатами анализа. Сами визуализации могут включать в себя разнообразные носители, в том числе текст, изображения и видео, а также могут быть интерактивными, позволяя историям воздействовать одинаково и на читателя, и на автора.

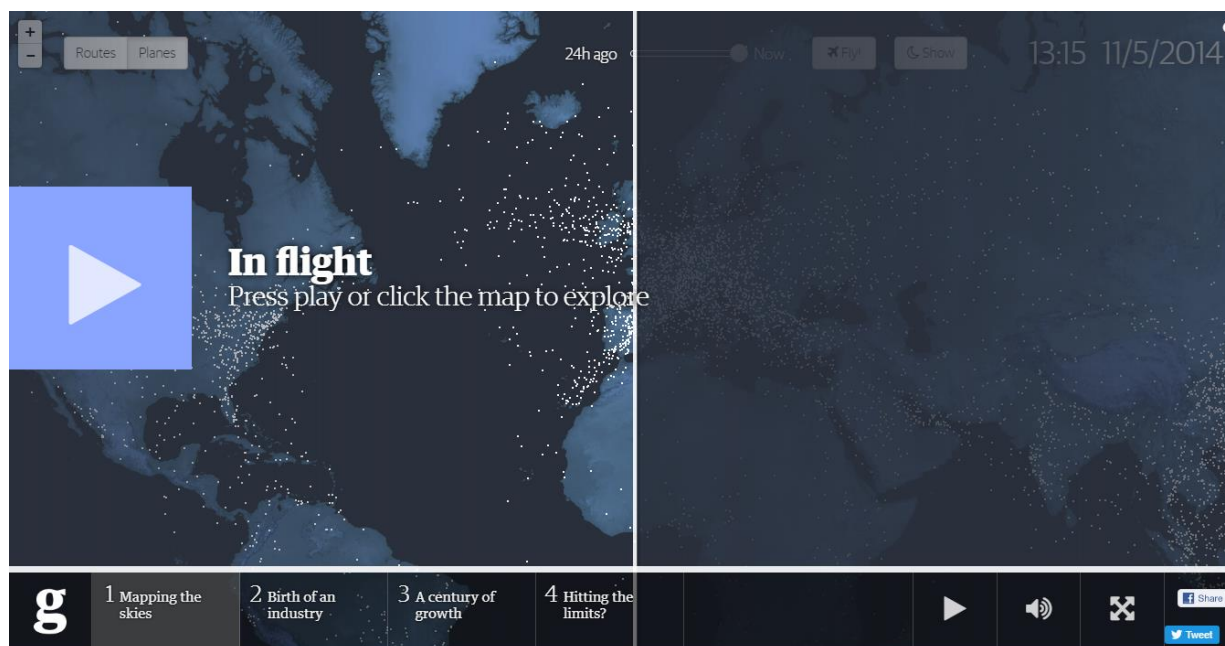


Рисунок 5 — Визуализация проекта In flight

Д. Мишель дает следующее определение «нарративу» — рассказ в игре, предоставляющий исходную информацию и контекст. Продолжающийся рассказ может мотивировать решение проблем, как это происходит в определенном жанре игры⁶⁶. Это создает эмоциональную близость, состояние, в котором пользователи чувствуют себя так или иначе связанными или

⁶⁶ Michele D. D. Game Design Narrative for Learning: Appropriating Adventure Game Design Narrative Devices and Techniques for the Design of Interactive Learning Environments. ETR&D, 54(3), 2006. P. 245.

идентифицирующими себя с персонажами внутри текущей системы. Некоторые исследователи относят повествование к динамике в геймификации.

1. 3. 5 Элементы геймификации в проектах big data

Себастьян Детердинг, Рилла Халед и др. вовремя СНГ в 2011 году предложили следующее определение геймификации — «использование элементов игрового дизайна в неигровых контекстах»⁶⁷. Гейб Зихерманн определяет геймификацию как «процесс игры и игровой механики для привлечения пользователей»⁶⁸. Объединяя эту информацию Майкл Ву, Джо Ли и Джессика Хаммер говорят об геймификации как об «использовании игровых атрибутов для управления игровым поведением игроков в неигровом контексте»⁶⁹ и «использовании механики и динамики в рамках игры для продвижения желаемого поведения»⁷⁰. Все эти дефиниции объединяет использование элементов с игровым фундаментом (game design) в существующей среде.

Определения геймификации в отраслевых компаниях мало чем отличаются от упомянутых выше, разве что тем, что ориентируются на цель использования. Badgeville⁷¹ рассматривает геймификацию как «современную стратегию бизнеса, которая использует проверенные методы социальных игр для измерения и влияния на поведение покупателя». Опять же, основное внимание в определении делается на трансплантацию игровой механики в

⁶⁷ Deterding S., Khaled R., Nacke L., Dixon D. Gamification: Toward a Definition [Электронный ресурс]. Точка доступа: <http://gamification-research.org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon.pdf> (дата обращения: 17.05.17)

⁶⁸ Zichermann, G. Gamification by Design. O'Reilly Media, 2011. P. 208.

⁶⁹ Wu M. What is gamification, really? [Электронный ресурс]. Точка доступа: <http://lithosphere.lithium.com/t5/Science-of-Socialblog/What-is-Gamification-Really-ба-р/30447> (дата обращения: 15.05.17)

⁷⁰ Lee J.J., Hammer, J. Gamification in Education: What, How, Why Bother? Academic Exchange Quarterly, 2011.

⁷¹ Badgeville. Gamification. Accessed on January 20, 2013. [Электронный ресурс]: <http://badgeville.com/main/gamification> (дата обращения: 15.05.17)

различные ситуации и среду, чтобы запретить изменение поведения пользователя. Bunchball⁷² больше фокусируется на конечном результате - изменении поведения: «Общая цель геймификации — это вовлечь людей в игру: делиться и взаимодействовать в какой-либо деятельности или сообществе, предлагая привлекательную, динамичную и устойчивую форму.

Рассматривая сходства между геймификацией и игровым дизайном, мы приходим к выводу, что определения для термина «геймификация» кажутся недостаточно глубокими. Было бы трудно найти определения, раскрывающие игровой дизайн. Поэтому М. Джейкобс рассуждает об геймификации⁷³ «как о процессе слияния преднамеренного расположения механики с существующей средой».

Элементы игры можно разделить на три группы: динамику, механику и компоненты⁷⁴. На следующем рисунке показана взаимосвязь между элементами игры (рисунок 5).



Рисунок 5 — Иерархия игровых элементов

Динамика игры. Под динамикой игры мы понимаем количество совершенных игроком действий за определенный отрезок времени. Динамика

⁷² Bunchball. Gamification 101: An Introduction to the Use of Game Dynamics to Influence Behavior, 2013. [Электронный ресурс]: <http://www.bunchball.com/sites/default/files/downloads/gamification101.pdf> (дата обращения: 15.05.17)

⁷³ Jacobs M., Gamification: Moving from 'Addition' to 'Creation', 2013.

⁷⁴ Werbach K., Hunter D. For the Win: How game thinking can revolutionize your business. Wharton Digital Press, 2012.

игр находится на вершине иерархии, она влияет на игровую систему на абстрактном уровне. Из-за влияния динамики на другие элементы в начале любого процесса проектирования ее необходимо правильно установить. Эффективное взаимодействие игрока с игровой механикой будет контролироваться посредством модификации динамики игры, как и взаимодействие пользователя с игровой системой⁷⁵. Всего существует пять игровых динамик: ограничения, эмоции, повествование, прогрессия и отношения.

Игровая механика — принцип взаимодействия игры с игроком. Например, сбор очков, индикаторы здоровья, повышение уровней и т. д. В новостной игре *Leo's Red Carpet Rampage*⁷⁶ от лондонской студии The Line есть три игровых механики: индикатор жизни героя, номер раунда и полученные за собранные награды очки. Это 8-битная браузерная игра, в которой персонаж актера пытается обогнать соперников на красной ковровой дорожке в погоне за Оскаром. Среди его соперников, к примеру, Мэтт Деймон, Майкл Фассбендер и Эдди Редмейн (Рисунок 6).



Рисунок 6 — Новостная игра *Leo's Red Carpet Rampage*

⁷⁵ Zichermann G., Cunningham, C. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. Sebastopol, Ca: O'Reilly, 2011.

⁷⁶ *Leo's Red Carpet Rampage* [Электронный ресурс]. Точка доступа: <http://redcarpetrampage.com/> (дата обращения: 15.05.17)

В отличие от других аспектов изучения геймификации, предметная игровая механика была подробно изучена исследователями игрового дизайна. В интернете есть множество описаний игровой механики.

Компоненты игры — это специфические формы игровой динамики и механики (достижения, значки, боссы, драки, коллекции, борьба, уровни, очки, виртуальные товары и др).

Технологии геймификации успешно используются в современной журналистике, в том числе и на телевидении. Впервые сыграть в игру телезрителям предложили в бразильском телешоу. В 2009 году в шоу «Fantástico» говорили об исчезнувшем несколько месяцев назад Белкиоре (Belchior), певце, снискавшем известность в 1970-ые. После эфира было решено дать ход этой истории, но уже в Интернете: на сайте канала «G1» было опубликовано небольшое информационное сообщение, спровоцировавшее отклик аудитории. Зрители присылали фотографии с Белкиором и сведения о том, где и когда в последний раз они видели певца. Команда сайта создала карту, на которой были отмечены последние местонахождения Белкиора с указанием даты. Не прошло и недели, как певца нашли, причем как раз в одном из тех мест, что указали зрители и посетители сайта — в небольшой деревне близ Сан-Грегорио-де-Поланко, в Уругвае. Это пример активного участия аудитории в журналистском процессе, проявившегося вполне естественно, без особого труда со стороны журналистов.

Существуют примеры успешного использования элементов геймификации и в проектах на основе big data. В процессе создания проекта «MP's Expenses» в 2009 году редакция The Guardian обратилась к читателям с просьбой проверить декларации расходов членов парламента на предмет их обоснованности. Читатели должны были с помощью специального приложения просмотреть сотни тысяч деклараций и указать в них те места, которые кажутся подозрительными и требуют проведения расследования. Результаты проекта оказались внушительными: около полутора миллиона фунтов стерлингов было возвращено в государственный бюджет. Из интервью

Саймона Роджерса, журналиста Guardian, блогу «Data Journalism», опубликованного в книге «Пособие по журналистике данных»: «Когда мы занимались проектом о расходах во второй раз, мы сделали его похожим на игру, в которой люди получали индивидуальные задания. То, что мы давали людям конкретные задачи, действительно помогло нам. Ведь когда вы просто даете людям горы информации и говорите им: “давайте, обработайте ее”, то вы ставите перед ними трудную и в какой-то мере неблагодарную задачу. И поэтому, я считаю, очень важно превратить эту работу в развлечение или игру»⁷⁷.

Какие же игровые механизмы представлены в «MPs Expenses»? Конкретное задание в игровой форме («challenge»). Огромный массив документов был разбит на небольшие фрагменты IT-командой «The Guardian». Для пользователей было создано веб-приложение, в котором нужно было выбрать варианты ответов на вопросы, а на некоторые вопросы дать короткий письменный ответ. Какого рода это документ? Что в нем подозрительного? Отвечая на подобные вопросы, читатели проверяли декларацию за декларацией, пока в конце концов не было проверено около половины всех опубликованных документов – более 220 тысяч деклараций. Следующий механизм - чувство принадлежности к комьюнити, сотрудничество («cooperation»). Читатели были заняты в одном большом краудсорсинговом проекте, чувствуя себя тесно связанными с журналистами издания. Последние активно поддерживали обратную связь с аудиторией («feedback»), публикуя периодические отчеты и интересные находки и вступая в диалог с пользователями.

1. 4 Журналист в эпоху big data

⁷⁷ Грэй Дж., Чемберс Л. Пособие по журналистике данных [Электронный ресурс] / Грэй Дж., Чемберс Л. // Сетевое издание «РИА Новости», 2013. Режим доступа: <http://ria.ru/files/book/site/index.html> (дата обращения: 15.05.17)

Некоторые исследователи говорят про журналистику данных как про специализацию. Она требует от журналиста владения новыми навыками – основами программирования, пониманием статистики и использования ее на практике, умением визуализировать информацию.

Журналистам придется, помимо литературы, осваивать сугубо технические дисциплины, в частности, работу с разными компьютерными программами и владение (хотя бы на базовом уровне) языком программирования. М.Н. Шерстюкова пишет: «журналист для успешной работы должен владеть двумя разными типами мышления: формально-логическим и ассоциативно-образным. При этом новые творческие возможности представления информации открываются на всех этапах работы журналиста»⁷⁸. Работа в сфере журналистики данных требует умения находить, анализировать и визуализировать полученные данные.

Конечно, если один журналист выполняет сверхзадачу и умеет все хорошо, что, естественно, бывает крайне редко, то это большой плюс для редакции. Но на практике, например, в РИА Новости над проектом работает целая творческая группа: дизайнер, программист и журналист. В этой триаде, естественно, главную роль играет журналист, которые является и идейным вдохновителем, и носителем информации, и шеф-редактором.

«Журналистика данных – это совмещение журналистом роли переводчика с экономического/ государственного/ цифрового языка в язык, понятный всем, и, одновременно, роли аналитика, представляющего материалы в предельно наглядной форме», - обобщает С.И. Симакова⁷⁹. И тут

⁷⁸ Шерстюкова, М.Н. Дата-журналистика как новое направление в системе средств массовой коммуникации [Электронный ресурс] / М.Н. Шерстюкова // Международный электронный научно-образовательный журнал «Медиа. Информация. Коммуникация». – 2012. - №1. – Режим доступа: <http://mic.org.ru/1-nomer-2012/65-1-shestukova-2> (дата обращения: 14.05.17)

⁷⁹ Симакова, С.И. Журналистика данных как современное направление журналистики // С.И. Симакова // Научный журнал «Знак»: проблемное поле медиаобразования, 2013. - №1 (11). С. 52-56.

возникает вопрос, а где должен получать молодой специалист все вышеперечисленные навыки?

С одной стороны, образовательные программы в университетах не всегда отвечают потребностям самих студентов и далеко не всегда формируют навыки, которые бы отвечали запросам работодателей. «Журналиста можно учить разным методам, которые уже есть в науке и исследовательской практике, замечать сенсационные факты там, где остальные видят просто информационный шум», – Виталий Лейбин. С другой стороны, сегодня появляется большое количество онлайн-курсов как иностранных, так и русских, как бесплатных, так и платных. Курсы по журналистике данных преподаются в некоторых зарубежных университетах, а для студентов факультетов журналистики проводятся конференции с участием профессиональных журналистов, в том числе и тех, которые работают с данными. Например, в 2013 году такую конференцию организовало РИА Новости (Как развивается «журналистика данных»).

Как раз на этой конференции Андрей Горянов (на то время главный редактор портала Slon.ru, а ныне генеральный продюсер «Русской службы BBC») представил два пути развития журналиста в области дата-журналистики. Первый путь – это то, о чем мы писали выше, – выполнять сверхзадачу и уметь все - программировать, визуализировать, находить и анализировать контент на основе базы данных, и второй путь – делать что-нибудь одно, но делать это намного лучше, чем другие: «Журналист должен блестяще овладеть своим стилем и этим покупать потребителя». Но в приоритете, естественно, стоит универсальный журналист. На практике журналисты Slon.ru умеют «чуть больше, чем просто писать», а дизайнеры понимают толк в коде. «Это не просто люди, которые создают красивую картинку - они могут сами написать простенький код. Они говорят с программистами на одном языке. Мне кажется, происходит дефрагментация – либо ты блестяще пишешь, либо осваиваешь технические данные», - Андрей Горянов.

Журналисту необходимо быть хорошим аналитиком, обладать достаточно развитым образным мышлением, быть уверенным пользователем технических инструментов и понимать возможности визуализации данных.

Работая с базой данных, журналист должен совмещать в себе три основных навыка – это программирование, отбор и анализ данных и визуализация проанализированной информации, то есть «должен сочетать в себе и «физика», и «лирика».

Конвергенция разных жанров и форм в онлайн-журналистике привела к созданию уникального формата журналистского материала – мультимедийного текста, который является формой для мультимедийного проекта и проекта журналистики данных. Мультимедийный проект – это журналистский текст, в котором тема раскрыта с помощью различных медийных платформ – текстовых и аудиовизуальных средств, единство которых создает объемную картину события, явления или предмета. Отсюда стоит сделать вывод, что в результате технологических процессов изменилась подача и «упаковка» журналистских материалов – она стала мультимедийной.

Проект журналистики данных относится к сетевым жанрам, так как он обладает всеми функциями и свойствами присущими жанрам Интернет-СМИ, в том числе и конвергентной передачей информации. Одной из задач проекта журналистики данных является создание целостной картины какого-либо явления, события или предмета, представленного нелинейной системой воспроизведения. Все это дает нам право утверждать, что проект журналистики данных относится к *аналитическому жанру* и представляет собой *мультимедийный текст*. В свою очередь, мультимедийный текст является одной из форм для реализации проекта журналистики данных. Мы даем следующее определение дата-журналистике:

Дата-журналистика – это деятельность по анализу доступных баз данных, обработке полученной в ходе анализа информации, и ее интерактивной визуализации.

Отсюда, проект журналистики данных – это мультимедийный проект, основанный на анализе баз данных и визуализированный с помощью интерактивного интерфейса.

ГЛАВА 2. СПОСОБЫ ВИЗУАЛИЗАЦИИ BIG DATA В СОВРЕМЕННОЙ ЖУРНАЛИСТИКЕ

2.1 Определение визуализации данных

Классическое определение визуализации «формирование визуальных образов, процесс интерпретации в визуальных терминах или введения в визуальную форму». С процессами цифровизации определение было модернизировано до «инструмент или метод для интерпретации данных, загружаемых в компьютер, и для создания изображений из сложных многомерных наборов данных». Цель визуализации данных — получить информацию путем сопоставления данных с графическими объектами.

Р.Б. Хабер и Д.А. Макнабб⁸⁰ еще в 1990 году дали следующее определение визуализации данных: использование компьютерной технологии визуализации в качестве инструмента для понимания данных, полученных путем моделирования или физического измерения. В этом понимании технология визуализации основана на интеграции более старых технологий, включая компьютерную графику, обработку изображений, компьютерное зрение, геометрическое моделирование с использованием автоматизированного проектирования, теорию приближения, психологию восприятия и исследования пользовательского интерфейса.

«Визуализация данных — способ представления многомерного распределения данных на двумерной плоскости, при котором качественно отражены основные закономерности, присущие исходному распределению — его кластерная структура»⁸¹. Одна из функций визуализации данных наглядно отобразить закономерности в big data. Именно в двумерной плоскости человек

⁸⁰ Haber R.B., McNabb D. A., Visualization Idioms: A Conceptual Model for Scientific Visualization Systems, in Visualization in Scientific Computing, 1990.

⁸¹ Зиновьев А. Ю. Визуализация многомерных данных. Красноярск: Изд-во КГТУ, 2000. С. 168.

естественно воспринимает геометрические формы, а отношения между структурами наглядны.

Коммуникация между создателем визуализации и ее потребителем находится на плоскости — одномерной, двумерной, трехмерной. Главная задача, по мнению Э. Тафти⁸², в представлении информации уйти от плоскостей в многомерную реальность. В 2015 года исследователи зарубежной и отечественной журналистики начали активно обсуждать перспективы *immersive journalism* («журналистики погружения»), когда пользователь с помощью технологий 3D и виртуальной реальности может присутствовать на месте события из новостных сводок. С помощью 3D-очков и 3D мониторов создается виртуальный мир с участниками события, местом действия, телесными и звуковыми ощущениями – «эффект присутствия». Так в 2012 году американская журналистка Нонни Де Ла Пенья с помощью технологий виртуальной реальности создала серию репортажей про сирийских беженцев. А в 2015 году *The New York Times* отправили своим подписчикам *Google Cardboard* – картонный шлем со встроенными виртуальными очками, которые можно собрать самостоятельно и использовать для просмотра фильмов, созданных в виртуальных мирах. В перспективе пользователь сможет иметь доступ к виртуальной базе данных, по которой сможет путешествовать как по «чертогам разума» персонажа известного британского сериала.

Изображения, полученные в результате анализа данных, должны быть доступны для пользователя за счет разнообразных средств: границ, пропорций, масштаба, цвета и т.д. От этого зависит эффективность коммуникации между пользователем и создателем проекта. Х. Чернофф⁸³ описал основные принципы комбинации визуальных средств представления информации: принцип акцента на основных смысловых элементах, принцип

⁸² Tufte E.R. *Envisioning Information*. – Cheshire, CT: Graphics Press, 1990. P. 126.

⁸³ Chernoff H. *Journal of the American Statistical Association*, No. 68, 1973. P. 361-368.

лаконичности, обобщения и унификации, автономности, структурности, стадийности, принцип автономности, принцип использования привычных ассоциаций и стереотипов. Кроме этого, средство визуализации быть надежным и иметь скорость, которая устроит пользователя, принимающего на основе этих данных решения.

Визуализация — это инструмент, который знакомит нас уже с конечным результатом анализа баз данных, позволяет нам управлять процессом этого знакомства и даже выбрать нужное направление в исследовании данных. Может выступать инструментом для показа данных и побуждать зрителя задуматься о сути, а не о способе; побуждать зрителя сравнивать фрагменты данных, объединяя их в единое целое. Визуализация данных следует четким целям: описанию, исследованию, упорядочиванию или украшению. А график является самым простым и популярным способом визуализации данных.

Графический образ и вспомогательные элементы являются составной частью графика. Графический образ данных – это результат визуализации. В статистике под графическим образом понимают «множество точек, фигур и линий, которые иллюстрируют статистические данные». Вспомогательные элементы графика включают в себя его название, оси координат, шкалы, числовые сетки и данные, которые уточняют или дополняют изображаемые показатели. Представление данных в виде графических образов позволяет пользователю увидеть изучаемое явление или процесс изнутри, облегчая понимание данных и обеспечивая поддержку для принятия решений.

Цель визуализации — предоставить пользователю возможность легко извлекать информационное содержимое данных. Связь, которая не очевидна из самих данных, становится видимой с помощью визуализации. Данные могут быть проанализированы с помощью визуального представления. Кроме того, визуализация упрощает коммуникацию. Даже неспециалисты должны получить сообщение, которое им передают, а интерактивное управление процессом визуализации должно дать возможность понять явления как можно скорее.

2.2 Способы визуализации big data: определение и классификации

Способы визуализации данных направлены на то, чтобы представлять данные пользователю таким образом, чтобы они точно передавали информацию и требовали минимальных усилий для понимания. Хорошая визуализация данных может облегчить эффективный анализ больших объёмов данных и облегчить их понимание. Поэтому разными теоретиками и практиками в области визуализации осуществлялись попытки классифицировать ее способы в зависимости от цели визуализации и типа информации.

Авторы блога компании DevExpress выделяют несколько типов визуализации⁸⁴:

1. Визуальное представление количественной информации в схематической форме (гистограммы и спектрограммы, линейные и круговые диаграммы, таблицы и точечные графики);
2. Данные, преобразованные в форму, которая усиливает восприятие и анализ информации (диаграмма Эйлера, карта или полярный график, таймлайн, график с параллельными осями);
3. Концептуальная визуализация (диаграмма Ганта, граф с минимальным путем, концептуальные карты и т. д.);
4. Стратегическая визуализация — различные данные о работе организации (диаграммы производительности, жизненного цикла и графики структур организации);
5. Комбинированная визуализация — несколько сложных графиков, объединенных в одну схему (карта погоды);
6. Метафорическая визуализация — графически организованная структурная информация с помощью деревьев, карт и пирамид.

⁸⁴ Зачем и как использовать визуализацию данных? [Электронный ресурс] / П. Шиманский // Хабрахабр, 2014. Режим доступа: <https://habrahabr.ru/company/devexpress/blog/240325/> (дата обращения: 15.05.2017)

В 1996 году вышел справочник по визуализации данных Роберта Харриса⁸⁵, в котором была опубликована авторская классификация визуализации данных (Рисунок 7). Исследователь предлагает определить количественные ли данные будут лежать в основе вашего проекта, и на какой вопрос должен ответить ваша аудитория, обратившись к визуализации. В зависимости от этого, вы подбираете категорию, которая подходит для вашего графика — карта, диаграмма, таблица и др.

Major Categories of Charts					
Key features	Graph (plot)	Map	Diagram	Table	Other Charts
Quantitative	Quantitative patterns and complexities	Spatial and directional relationships	Nonquantitative interrelationships	Precision of information and ease of reference	Differs depending on specific chart
Primary Function	Examples are shown below				
Shows patterns and/or relationships of quantitative data at point in time	Area graph Bar graph Circle graph Column graph Line graph Nomograph Polar graph Radar graph Scatter graph Tilted graph	Contour map Demographic map Distorted map Elevation map Prism map Shaded map Smooth statistical map Weather map		Analytical table Bidirectional table General table Quantitative table Reference table Spreadsheet	Icon companion display Pie chart Proportional chart Ranking chart Unit chart Venn diagram
	Shows patterns and/or relationships of quantitative data over time	Area graph Candlestick chart Cash flow graph Column graph Control chart Index graph Line graph Run chart		Analytical table General table Quantitative table Reference table Spreadsheet	Comparative chart Icon companion display Point & figure chart Proportional chart Star chart Unit chart
Shows how/where things are distributed or located	100% graph Border plot Box plot Histogram Pareto graph Population pyramid Probability graph Quantis graph	Block map Blot or patch map Dot density map Geological map Pin map Profile map Topographic map Weather map	Block diagram Network diagram Voronoi diagram	Analytical table Frequency table General table Percent table Quantitative table Reference table Spreadsheet	Business matrix Conceptual chart Floor plan
	Relates time and activities		CPM chart PERT chart Time line	Calendar Time table	Explanatory chart Gantt chart Loading chart Milestone chart Scheduling chart
Shows how nonquantitative things are organized, arranged, interconnected or interested		Network map Ray map Road map Strip map Thematic map	Block diagram Cause & effect diag. Conceptual diagram Flow chart Network diagram Organization chart Relational diagram Venn diagram	Conceptual chart General table Pictorial table Reference table	Conceptual chart Distribution channel chart Exploded diagram Gantt & milestone charts Minimum spanning Process chart Structure diagram
	Shows how nonquantitative things proceed		Flow map Weather map	Conceptual diagram Decision chart Flow chart PERT & CPM charts Process chart Tree chart	Analytical table General table Reference table Spreadsheet
Shows how nonquantitative things evolve or work			Conceptual diagram Flow chart Process chart	Analysis table General table Pictorial table Reference table	Conceptual chart Cross-section Exploded diagram Illustration chart
Shows how to do things			Calculation chart "How to" diagram Procedural diagram		Calculation chart How-to chart Illustration chart Procedural instruction

Рисунок 7 — Способы визуализации данных в книге Information Graphics: A Comprehensive Illustrated Reference

В 2006 году Э. Эбела, вдохновленный Д. Желязны, создал схему Chart Suggestions — A Thought-Starter, упрощающую выбор визуализации данных для дизайнера⁸⁶. В первую очередь, он предлагает ответить на вопрос: What would you like to show? («Что бы вы хотели отобразить?»), если адаптировать данные вопрос на русский язык, то он звучит, как «Какова цель вашей визуализации?» (Рисунок 8). В определении целей Эбела ориентировался на

⁸⁵ Robert L. Harris. Information Graphics: A Comprehensive Illustrated Reference. Publisher: Oxford University Press; 1 edition, 2000.

⁸⁶The Extreme Presentation(tm) Method [Электронный ресурс]. Точка доступа: http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html (дата обращения: 15.05.17)

классификацию Д. Желязны⁸⁷: отношения данных, распределение данных, композицию данных или сравнение данных. Данные цели отображают самые разные аспекты анализа информации: с помощью отношений можно выявить наличие или отсутствие зависимостей между переменными; распределение данных отвечает за количество объектов, попадающих в определенные области числовых значения; композиция данных — объединяет данные с целью сравнения компонентов, составляющих процент от некоего целого. Сравнение данных — это объединение данных, с целью сравнения некоторых показателей, выявление того, как объекты соотносятся друг с другом. В дальнейшем каждый из случаев рассматривается в отдельности в зависимости от количества и значения переменных.

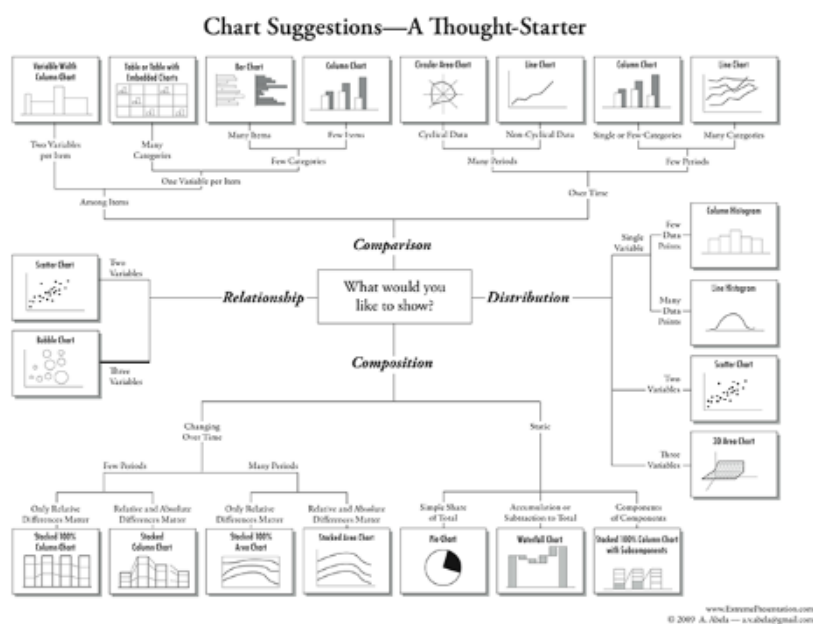


Рисунок 8 — Способы визуализации данных Эндрю Эбела

В 2013 году классификацию Э. Эбелы модернизировал Ж. Камоес⁸⁸ (Рисунок 9). Он выделяет шесть типов категорий, на которые можно ориентироваться при выборе способа визуализации данных: сравнение,

⁸⁷ Желязны Д. Говори на языке диаграмм: Пособие по визуальным коммуникациям для руководителей / Пер. с англ. – М.: Институт комплексных стратегических исследований, 2004. С. 220.

⁸⁸ A classification of chart types [Электронный ресурс]. Точка доступа: <http://excelcharts.com/classification-chart-types/> (дата обращения: 15.05.17)

композиция, отношения, распределение, эволюция (временные модели) и профилирование (сравнение шаблонов).

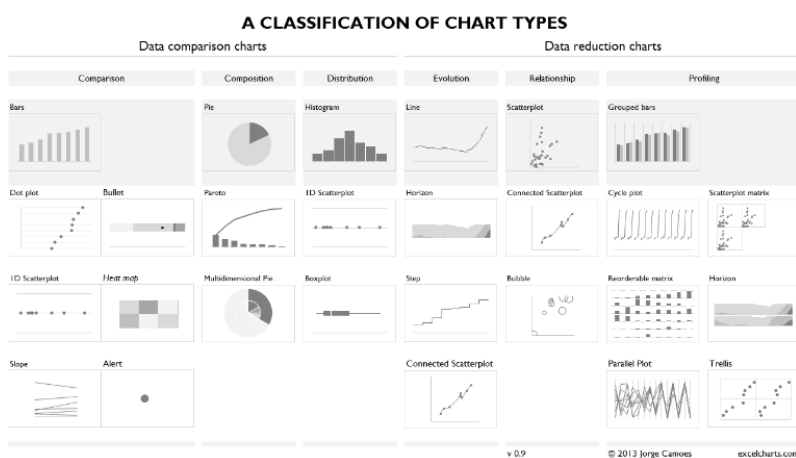


Рисунок 9 — Способы визуализации данных Жорже Камоеса

Свой «Визуальный словарь» создали и в международной британской газете Financial Times в ноябре 2016 года⁸⁹. По сравнению с предыдущими примерами, классификация в данной схеме более конкретизирована и проработана: отклонения, корреляции, рейтинги, распределение данных, изменение данных с течением времени, сравнение их размеров, поток данных и др.



Рисунок 10 — Способы визуализации данных Financial Times

⁸⁹ Visual vocabulary [Электронный ресурс]. Точка доступа: <https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf> (дата обращения: 15.05.17)

Последовательные попытки систематизировать большое количество способов для визуализации данных говорит о наличии проблем в этой области, которые требуют решения. Мы сформулировали следующую гипотезу исследования: при выборе способа визуализации специалист руководствуется типом визуализируемых данных (дискретные, географические или непрерывные временные данные), темой проекта, а также его целью. Для проверки данной гипотезы мы провели формализованный анализ визуализаций big data ведущих интернет-СМИ Европы и США.

2.3 Контент-анализ визуализации big data изданий The New York Times, The Washington Post, The Guardian

Мы визуализируем информацию для того, чтобы удовлетворить основную потребность — рассказать историю. Это одна из самых примитивных форм коммуникации, известных человеку, которая имеет свое происхождение от пещерных рисунков. С течением времени появились новые способы визуализации информации. Однако мы редко задумываемся о том, почему линейные графики, диаграммы, гистограммы более эффективны, чем таблицы, текст или числа. Мы не всегда можем обосновать правильный выбор визуализации, потому что не понимаем ее зависимость от внешних факторов.

Основной метод исследования, применявшийся в работе, — контент-анализ. Контент-анализ или анализ содержания — процедура качественного анализа всех видов документов (вербальных, визуальных и пр.) с последующей квантификацией (количественной обработкой) полученных данных. Таким образом, контент — анализ⁹⁰ относится к качественно — количественному (формализованному) типу анализа документов.

⁹⁰Горшков М.К., Шереги Ф.Э. Прикладная социология: методология и методы: Учебное пособие/ М.К. Горшков, Ф.Э. Шереги. — М.: Альфа-М, 2009. С. 416.

В качестве эмпирической базы для контент-анализа было взято 53 проекта журналистики данных из общественно-политических изданий (The Guardian, The New York Times, The Washington Post) за период с 2011 по 2016 год. Из них 6 материалов The Guardian — 15%, 21 материал The New York Times — 39%, 24 материала The Washington Post — 46%. Это все опубликованные материалы изданий за указанный промежуток времени с визуализацией большого массива данных (Рисунок 11).

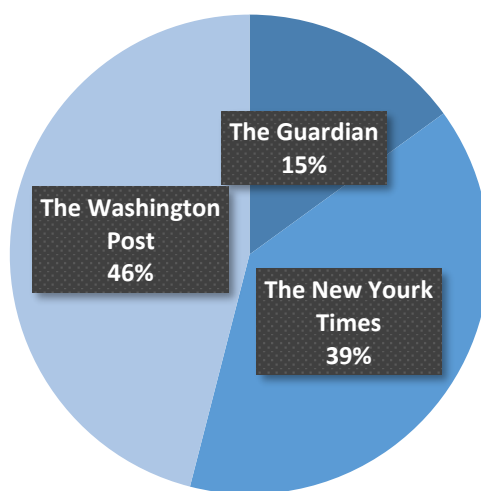


Рисунок 11 — Общее число проектов в контент-анализе

Data Journalism Awards — одна из престижных международных премий в области информационной журналистики, которая с 2012 года ежегодно выделяет выдающиеся работы: от лучших проектов на основе анализа данных до лучших работ в области их визуализации. Работы сетевых изданий The Guardian, The New York Times, The Washington Post неоднократно попадали в шорт-лист конкурса.

Основным инструментом формализованного анализа выступает кодификатор или бланк кодировки. В кодификаторе (Таблица 2.1) отмечается два критерия, раскрытых в предыдущей главе научной работы — тип информации и интерактивность, а также способы визуализации и тематика текстов, которые будут подробно описаны в этой главе. Базу данных исследования дополняют название и дата публикации, а также название издания.

Таблица 2.1 — Кодификатор для контент-анализа

№	Признак, градация признака	Код
1	Тематика издания: Национальные новости Мировые новости Технологии Бизнес Экономика Политика Спорт Наука Здоровье	1 2 3 4 5 6 7 8 9
2	Способ визуализации: Упаковка кругов Солнечная диаграмма Пузырьковая диаграмма Матричная диаграмма Карта Линейный график График «открыто-высокого-низкого-закрытия» Хордовая диаграмма Поточковый график Диаграмма параллельных координат Сложная столбчатая диаграмма Диаграмма рассеяния Интерактивная инфографика Пиктограмма Параллельные наборы	10 11 12 13 14 15 16 17 18 19 20 22 23 24 25
3	Тип информации: Дискретные данные Географические данные Непрерывные временные данные	26 27 28
4	Интерактивность Присутствует	0

	Отсутствует	1
--	-------------	---

2.3.1 Структура и тематические особенности материалов

The Guardian. Ежедневная британская газета *The Guardian* имеют столетнюю историю. А онлайн-версия газеты — theguardian.com — является одной из самых посещаемых сетевых версий британских газет. В 2012 году она стала третьим среди самых читаемых сайтов в мире. Вместе с посетителями сайта газеты число читателей *The Guardian* достигает 9 миллионов.

В марте 2009 года издание запустило проект *Guardian Datablog*, в котором изначально планировалось публиковать для читателей полные наборы данных, которые использовались в материалах онлайн-газеты. Сейчас *Datablog* стал образовательной площадкой, через которую можно знакомиться с данными мирового правительства, разными решениями в области визуализации данных художников-графиков *Guardian* и других проектов.

Первый проекты в области визуализации *big data* стали появляться на сайте в 2011 году и продолжают публиковаться на сайте в количестве от 1 до 3 проектов в год (Рисунок 12).

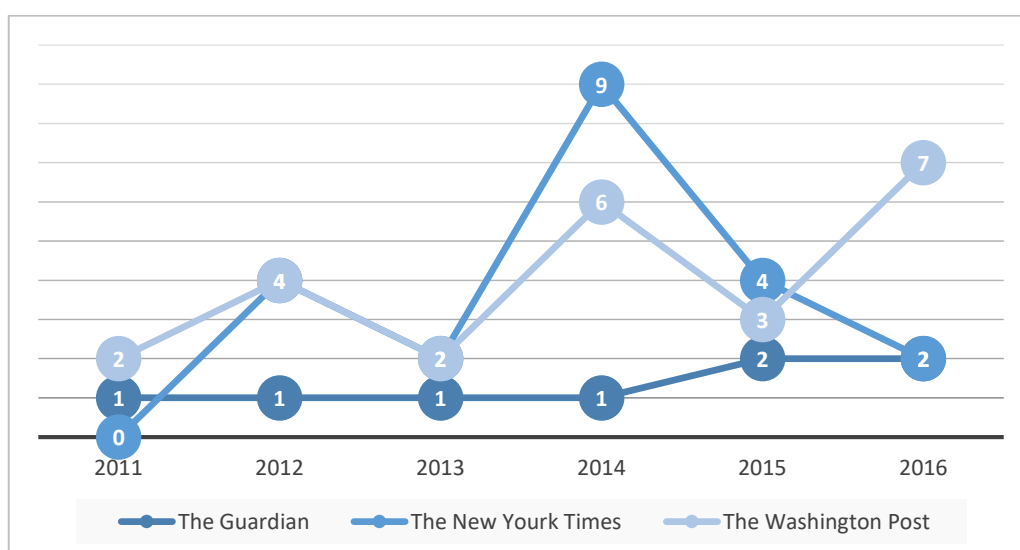


Рисунок 12 — График публикаций проектов на основе *big data*

Половина проанализированных материалов издания относятся к разделу *national news* — 50%, аналитике международных событий и явлений — 40% и 10% — развитию технологий. Визуализация больших данных The Guardian отличается проработкой и разнообразием способов визуализации — все проекты представлены разными графиками. Но при этом 60% проектов основаны на дискретном типе информации, что говорит о приоритетности для издания этапа визуализации исходных баз данных (Рисунок 13).



Рисунок 13 — Разнообразие тем в проектах The Guardian

Материалы, за редким исключением, имеют стандартную структуру: заголовок, имя автора, лид, основная визуализация, статья по теме с дополнительными графиками. Количество графиков варьируется от 1 до 10, когда общее количество уникальных способов визуализации в рамках одного проекта не превышает 3.

The New York Times — ежедневная газета, издаваемая в Нью-Йорке с 18 сентября 1851 года, первоначально создавалась как региональное издание. Однако сейчас это одна из крупнейших и многотиражных газет в США, и влиятельная газета в мире. Сайт TNY появился в 1996 году, а сейчас имеет посещаемость в десятки миллионов человек в месяц.

43% проанализированных материалов издания относятся к разделу *national news*, аналитике политических событий посвящено 19%, по 5% науке и экономике и по 14% спорту и бизнесу. Материалы The New York Times

отличается частотой публикаций: в среднем, издание публикует по 4 проекта в год, а 2014 год стал самым плодотворным — было опубликовано 9 проектов. Но при этом 52% проектов основаны на непрерывных временных данных, что говорит о внимательном отношении к способу обработки и подачи данных пользователю (Рисунок 14).

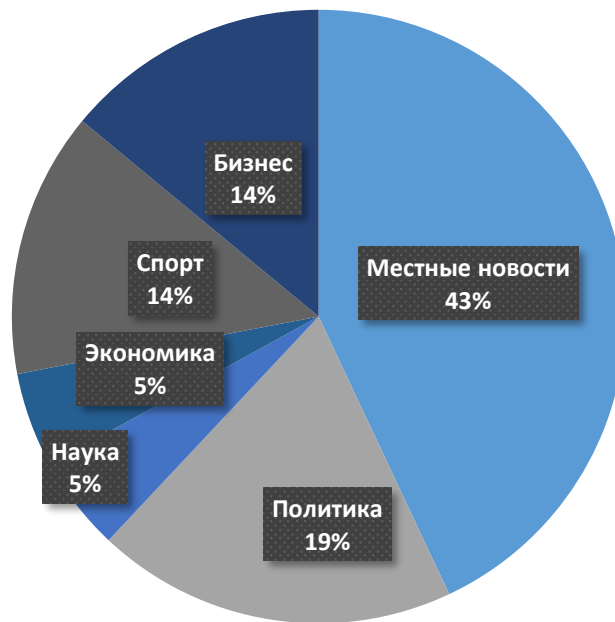


Рисунок 14 — Разнообразие тем в проектах The New York Times

Материалы строятся по следующей структуре — заголовок, дата публикации и имя автора работы, основной график с комментариями к нему, и множество дополнительных графиков, раскрывающих подробно основную. Помимо этого, обязательной частью является сноска с источниками данных, в которых пишут не только про источники, но и про погрешности в работе с данными, что минимизирует вероятность ввести читателя в заблуждение. Общее количество уникальных способов визуализации в рамках одного проекта не превышает 3, а общее количество графиков варьируется в среднем от 1 до 7. Исключение: проекты *State Gun Laws Enacted in the Year After Newtown* (109 графиков), *How the Recession Reshaped the Economy, in 255 Charts* (255 графиков), *How Fan Loyalty Changed During the World Cup* (33 графика); *The Race Gap in America's Police Departments* (116 графиков).

The Washington Post. Ежедневная газета была основана в 1877 году в Вашингтоне, где и выходит по настоящий момент. Создавая проекты с визуализацией больших данных, газета акцентирует внимание на национальных новостях (37%) и политике (20%), уделяет внимание международным (13%) и спортивным новостям (8%), а также здоровью (17%) и бизнесу (4%) (Рисунок 15).

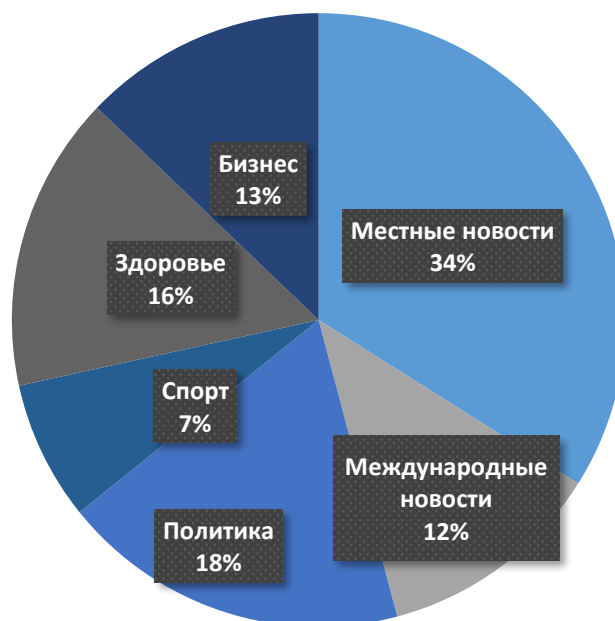


Рисунок 15 — Разнообразие тем в проектах The Washington Post

Проекты The Washington Post публикуют от 2 до 4 проектов в год, а в 2014 и в 2016 годах было опубликовано по 6 проектов. В визуализации данных в равных долях используют такие типы информации, как географические данные, непрерывные временные данные и дискретные данные. Структура материалов не отличается от The New York Times.

Общее количество уникальных способов визуализации в рамках одного проекта не превышает 3 (за исключением материала NFL draft pick quality for your team, в котором данные визуализированы с помощью 7 уникальных типов графиков), а количество графиков варьируется в среднем от 1 до 7. Исключение: проекты Building the U.S. tax code, break by break (12 графиков), Washington: A world apart (650 графиков); The number of 'mass shootings' in the U.S. depends on how you count (354 графика).

На основе анализа также была определена зависимость способа визуализации данных от тематики проекта. Местные и политические новости чаще всего были проанализированы с помощью карт (41% и 45%) и линейных графиков (18% и 33%). Способы визуализации мировых событий и новостей бизнеса разнообразны, но чаще всего использовались карты (50%). Тема здоровья чаще всего раскрывается с помощью диаграммы рассеяния (50%). Мы делаем вывод, что несмотря на большое количество существующих способов визуализации, дизайнеры чаще всего используют классические способы визуализации данных, которые знакомы и понятны читателю.

2. 4 Способы визуализации географического типа данных

В результате контент-анализа было определено, что 23% от общего числа проектов основаны на географических данных. Чаще всего карту для визуализации больших данных использовали в *The Washington Post* (38%). В *The Guardian* — 30%, а в *The New York Times* — 29%. Несмотря на наименьший показатель, редакция *The New York Times* уделяет большое внимание работе с этим типом графиков — в отделе графики у них есть подраздел, который занимается исключительно маппингом.

2. 4. 1 Карты (map)

Термин «картографический дизайн», который обозначает «формирование (конструирование) эстетического облика карты, атласа, любого другого картографического произведения в соответствии с его функциональным назначением, тематикой, современными художественными принципами и техническими возможностями», стал упоминаться в литературе по картографии в конце XX века. По факту, картографический дизайн появился вместе с появлением первых карт, однако в разные периоды развития искусства и картографии менялась и роль дизайна.

Сегодня для картографического дизайна важен синтез инженерно-технических и естественнонаучных знаний и гуманитарных дисциплин. Используя карты в data журналистике, мы имеем дело с широким слоем потребителей, неподготовленных к техническому типу визуализации, поэтому должны опираться на художественное мышление пользователя. С появлением мультимедийных средств, в картографическом изображении изменилась содержательная сущность, условные обозначения, способы отображения, возможность использования картографической информации, информативность и особенность восприятия пользователями содержания карты⁹¹.

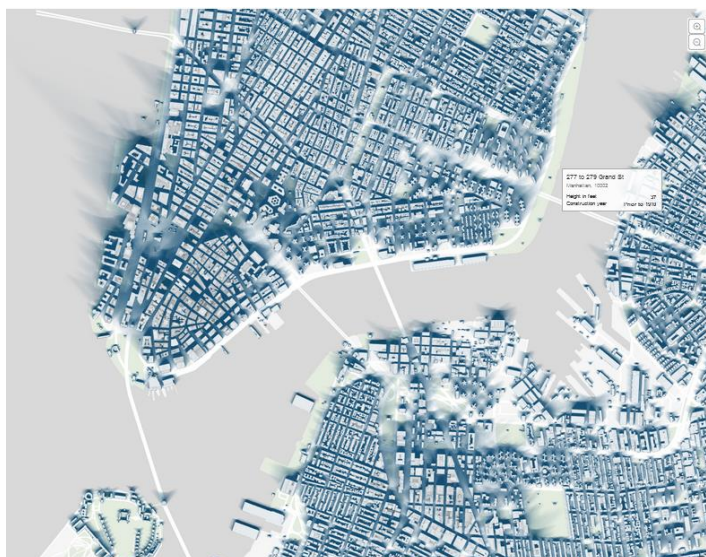


Рисунок 16 — Визуализация проекта

Mapping the Shadows of New York City: Every Building, Every Block

Проект The New York Times «Картирование теней Нью-Йорка: каждое здание, каждый блок» (Mapping the Shadows of New York City: Every Building, Every Block)⁹². В основу этой визуализации легли результаты исследования (Рисунок 16), проведенные Школой инженерии в Нью-Йоркском

⁹¹ Сизикова Т. С. Тенденция развития картографического дизайна в геоинформационном обеспечении территории. Новосибирск: СГГА, 2012.

⁹² The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2016/12/21/upshot/Mapping-the-Shadows-of-New-York-City.html> (дата обращения: 15.04.17)

университете. Ученые использовали метод трассировки лучей, когда имитируются эффекты, которые луч света может оказывать на здание и его окрестности. Тени, брошенные зданиями, влияют на ощущение и поток города, а отсутствие солнечного света может повлиять на такие аспекты, как покупка или аренда квартиры. В Нью-Йорке, где много высоких зданий, наличие данной проблемы очевидно. Обратившись к этому материалу, пользователь может узнать год постройки каждого дома в городе, его высоту, адрес, продолжительность нахождения участка рядом с домом в тени в зависимости от времени года. Для интерактивной визуализации используются спутниковые карты (Рисунок 17).



Рисунок 17 — Визуализация проекта
Mapping the Shadows of New York City: Every Building, Every Block

Точечные карты (Dot Map) — это способ обнаружения и распределения данных по географическому региону путем размещения по нему точек одинакового размера. Существует два типа Dot Map: один — к — одному (одна точка представляет один объект) и один — ко — многим (одна точка представляет конкретный блок объектов). Одним из преимуществ отображения статистической информации, таким образом, является то, что помимо графического отображения общего числа, расположение точек так же показывает, как эти предметы распределены.

Точечные карты идеально подходят для наблюдения за тем, как объекты распределяются по географическому региону и могут выявлять закономерности, когда точки группируются на карте, например, как в проекте The Washington Post «Убийства в округе» (Homicides in the District)⁹³. Он визуализирован с помощью точечной карты. Проект представляет собой базу с убийствами с 2000 по 2011 год, изучив которую пользователь может узнать количество убийств в отдельном районе Вашингтона, статус дела — открыто оно или закрыто, манеру убийства, мотивацию, а также возраст и расу преступника. Помимо этого, можно посмотреть статистику убийств по интерактивному линейному графику – их количество и год совершения. Посмотрев на преступления в своем районе, можно ознакомиться с общими тенденциями города (Рисунок 18).

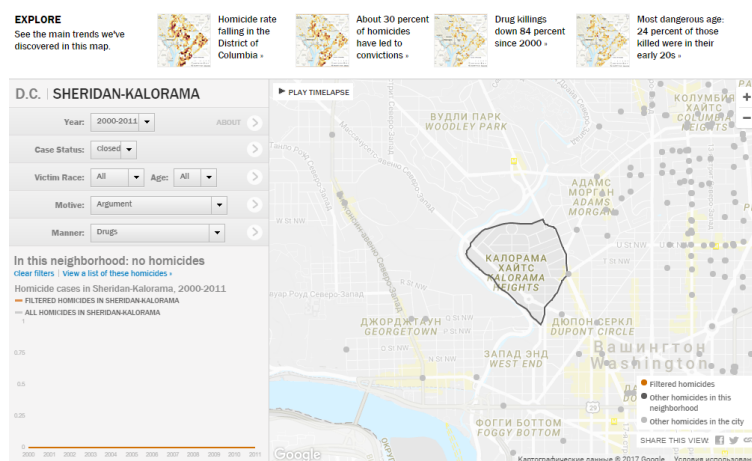


Рисунок 18 — Визуализация проекта Homicides in the District, точечная карта Пузырьковая карта (*Bubble Map*). На этой карте данных круги отображаются в обозначенной географической области, площадь области которой пропорциональна ее значению в наборе данных. Этот тип карты хорош для сравнения пропорций по географическим регионам. Но необходимо учитывать, что на bubble maps слишком большие «пузырьки» могут перекрывать другие «пузырьки» и области на карте.

⁹³ The Washington Post [Электронный ресурс]. Точка доступа: <http://apps.washingtonpost.com/investigative/homicides/#0:0:0:0:0:0:0> (дата обращения: 15.05.17)

Проект The Washington Post «Картирование того, как в США производят электричество» (Mapping how the United States generates its electricity)⁹⁴. Интерактивная карта показывает то, как в США производят электричество. Цветовая кодировка указывает на способ получения электричества. Красный – сжигание нефтяных продуктов, пурпурный – атомная энергия. Размер точки зависит от количества вырабатываемого электричества. По каждому штату приводится общая статистика с января по май 2015 года (Рисунок 19).

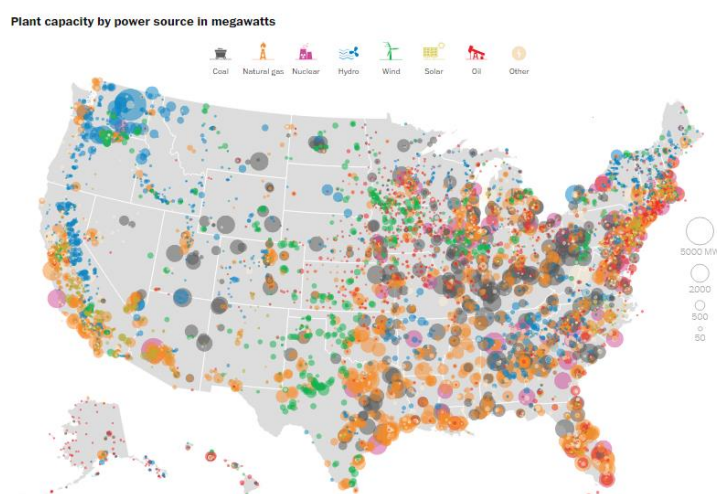


Рисунок 19 — Визуализация проекта Homicides in the District, пузырьковая карта *Хороплет (choropleth-map)* — это карта, на которой для различных областей штриховкой, точками различной густоты или окраской различной степени насыщенности показывается интенсивность какого-либо показателя. Областями в этом случае обычно являются объекты административно-территориального деления. Недостатком использования цвета является то, что вы не можете точно прочитать или сравнить значения с карты. Другая проблема заключается в том, что большие регионы кажутся более подчеркнутыми, чем более мелкие, и это влияет на восприятие зрителем затененных значений.

⁹⁴ The Washington Post [Электронный ресурс]. Точка доступа: https://www.washingtonpost.com/graphics/national/powerplants/?utm_term=.7c3e788cf7f7 (Дата обращения: 15.05.17)

Проект The Washington Post «Все более разнообразная Америка» (The increasingly diverse United States of America)⁹⁵. Расовое и этническое разнообразие общества сильно различается по всей Америке, но быстрая смена захватывает и наименее этнически богатые области. Хороплет имеет характерный для себя цветовой кодификатор с окраской разной степени насыщенности: желтый — малое разнообразие этнических групп, но много изменений в 2016 году; ярко-желтый — штаты не разнообразны, но испытывают наибольшую скорость изменений; зеленый — мало разнообразия в 2000 году и мало изменений в 2016; синий — большое разнообразие в 2000 году и мало изменений в 2016; серый — много разнообразия в 2000 году и много изменений сейчас (Рисунок 20).

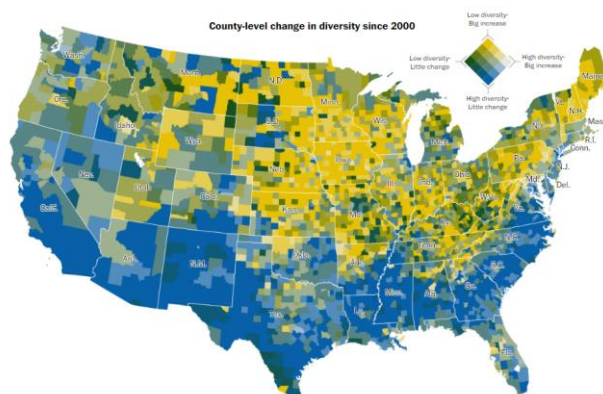


Рисунок 20 — Визуализация проекта

The increasingly diverse United States of America, хороплет

Проект The New York Times «Шанс поднять свой уровень дохода зависит от местоположения» (In Climbing Income Ladder, Location Matters)⁹⁶. В исследовании, проведенном учеными из Гарварда и Калифорнийского университета в Беркли, данные показывают пространственные различия в отношении шансов выхода из бедности ребенка в зависимости от округа, в

⁹⁵ The Washington Post [Электронный ресурс]. Точка доступа: <https://www.washingtonpost.com/graphics/national/how-diverse-is-america/> (Дата обращения: 15.05.17)

⁹⁶ The New York Times [Электронный ресурс]. Точка доступа: http://www.nytimes.com/2013/07/22/business/in-climbing-income-ladder-location-matters.html?pagewanted=all&_r=0 (Дата обращения: 15.05.17)

котором он проживает. Нью-Йорк Таймс сообщает: «Показатели взлетают по доходной лестнице реже на юго-востоке и на промышленном Среднем Западе, а также показывают, что разногласия особенно низкие в Атланте, Шарлотте, Мемфисе, Роли, Индианаполисе, Цинциннати и Колумбусе. Одни из самых высоких показателей наблюдаются на северо-востоке, Великих равнинах и Западе, в том числе в Нью-Йорке, Бостоне, Солт-Лейк-Сити, Питтсбурге, Сиэтле и больших полосах Калифорнии и Миннесоты (Рисунок 21).

In Climbing Income Ladder, Location Matters

A study finds the odds of rising to another income level are notably low in certain cities, like Atlanta and Charlotte, and much higher in New York and Boston.

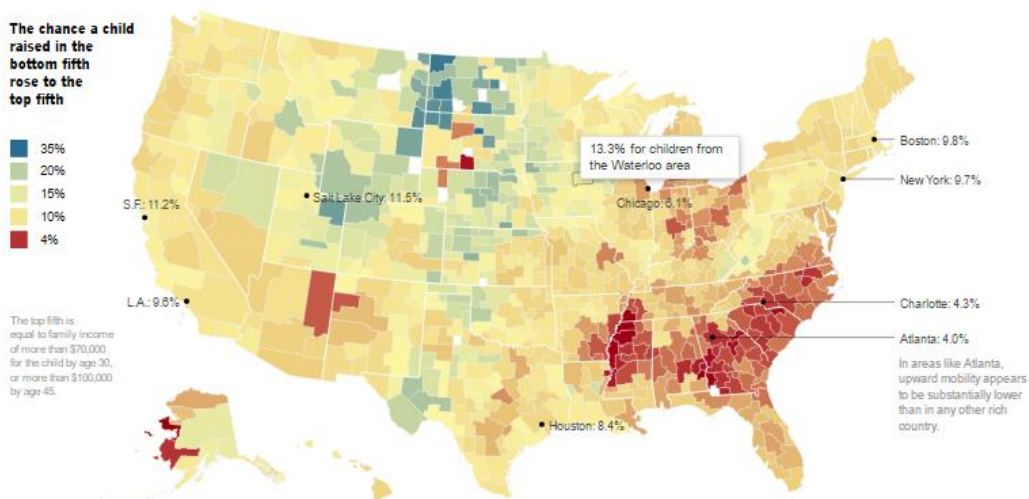


Рисунок 21 — Визуализация проекта

In Climbing Income Ladder, Location Matters, хороплет

Проект The Guardian «Смертельный кризис: отображение распространения эпидемии передозировки наркотиков в Америке» (A deadly crisis: mapping the spread of America's drug overdose epidemic)⁹⁷. Передозировка наркотиками убивает больше американцев, чем автокатастрофы или оружие — и эксперты говорят, что кризис еще не достиг своего пика. Данные показывают, как увеличилось количество смертей от передозировки

⁹⁷ The Guardian [Электронный ресурс]. Точка доступа: URL: <https://www.theguardian.com/society/ng-interactive/2016/may/25/opioid-epidemic-overdose-deaths-map> (дата обращения: 15.05.17)

наркотиков с 1999 года по 2014 год, а местная проблема превратилась в национальную эпидемию (Рисунок 22).

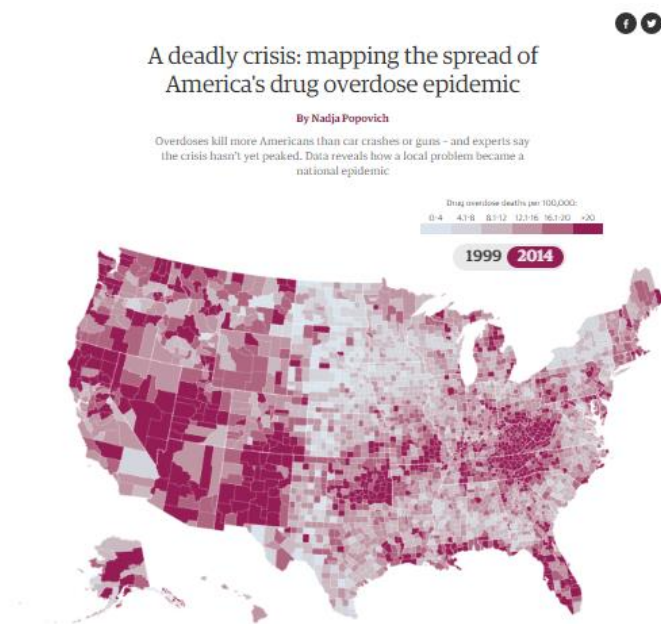


Рисунок 22 — Визуализация проекта

A deadly crisis: mapping the spread of America's drug overdose epidemic, хороплет

Карты связей (connection map) рисуются путем соединения точек, расположенных на карте прямыми или изогнутыми линиями. Карты связей подходят для отображения географических связей, а также их можно использовать для отображения маршрутов карты через одну цепочку ссылок. Карты связей также могут быть полезны при выявлении пространственных шаблонов через распределение соединений или за счет того, насколько концентрированные соединения находятся на карте.

Проект The Washington Post «Шесть карт, которые показывают анатомию американских обширных инфраструктур» (Six maps that show the anatomy of America's vast infrastructure)⁹⁸. В этом материале показаны масштабные карты важнейших американских инфраструктур — электрических линий, мест выработки энергии, мостов, труб и карты

⁹⁸ The Washington Post [Электронный ресурс]. Точка доступа: URL: <https://www.washingtonpost.com/graphics/national/maps-of-american-infrastructure/> (дата обращения: 15.05.17)

воздушных, морских и железнодорожных пути. Они были созданы на основе данных OpenStreetMap и различных правительственных ресурсов. Карта связей включает в себя элементы bubble map. Например, на карте с аэропортами большие точки — крупные аэропорты, маленькие точки — маленькие. Помимо этого, на графике отображено направление полетов (Рисунок 23).

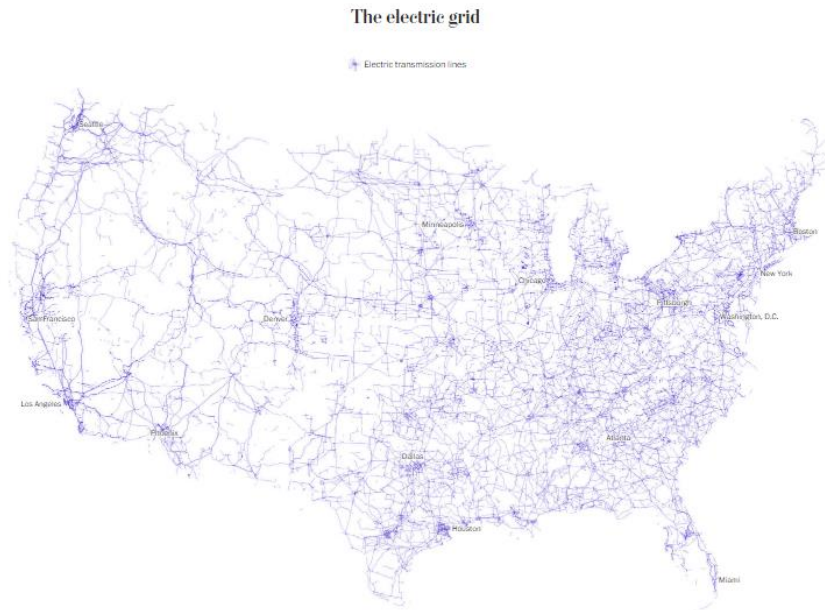


Рисунок 23 — Визуализация проекта
A deadly crisis: mapping the spread of America's drug overdose epidemic,
карта связей электрических линий

Карты потока (flow maps) географически показывают перемещение информации или объектов из одного места в другое и их количество. Карты потока берутся из точки происхождения и разветвляются свои «поточные линии». Стрелки могут использоваться, чтобы показать направление, или если движение является входящим или исходящим. Без стрелок может использоваться для обозначения движения, идущего назад и вперед. Объединение / объединение потоков линий и избежание пересечений может помочь уменьшить визуальный беспорядок на карте.

Проект The Guardian «В полете» (In flight)⁹⁹. IIn (International Lunar Network) и Guardian исследовали 100-летнюю историю пассажирских авиаперевозок, а для интерактивной карты потока использовали данные авиарейса FlightStats в реальном времени. На карте показаны все текущие полеты в воздухе прямо сейчас (Рисунок 24).

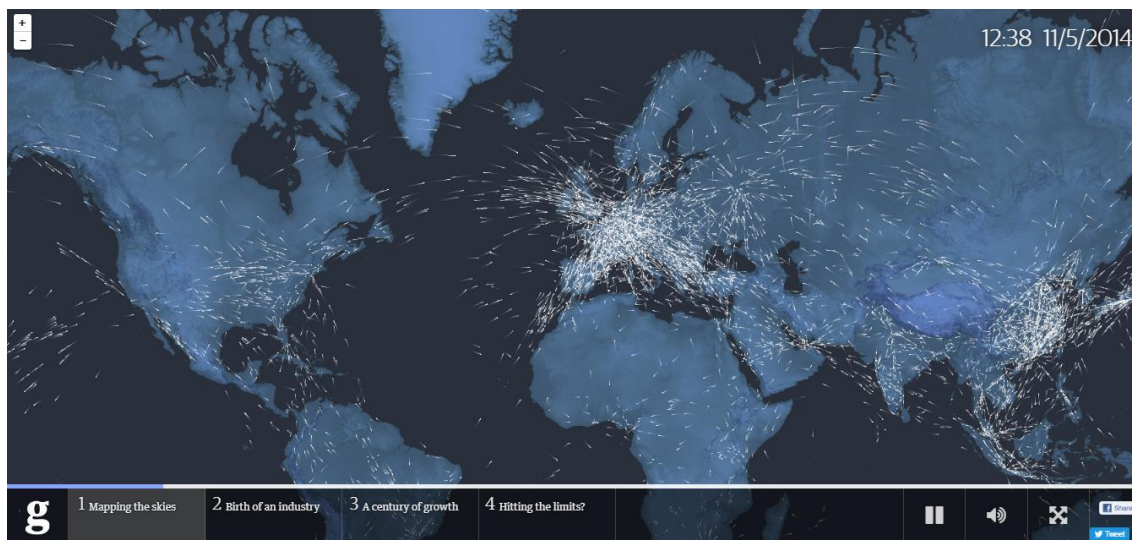


Рисунок 24 — Визуализация проекта In flight, карта потока

2. 5 Способы визуализации дискретного типа данных

В результате контент-анализа было определено, что 36% от общего числа проектов основаны на дискретном типе данных. Чаще всего, большие массивы данных дискретного типа использовали в The Washington Post (19%). В The Guardian — 9%, а в The New York Times — 8%.

2. 5. 1 Деревья (tree map)

Изначально tree maps были разработаны Беном Шнейдерманом для визуализации огромного каталога файлов на компьютере. Важное качество

⁹⁹ The Guardian [Электронный ресурс]. Точка доступа: URL: <https://www.theguardian.com/world/ng-interactive/2014/aviation-100-years> (дата обращения: 15.05.17)

графика в его компактности — он не занимал много места на экране. Этот способ основан на пространственной визуализации иерархических данных. И существует строгое требование к объектам данных, которые должны быть связаны иерархически. График представляет собой прямоугольник корня, разделенный на группы и представленный меньшими прямоугольниками, которые соответствуют объектам данных из набора. Treemaps удобны для отображения иерархии и хороши при сравнении пропорций между категориями через их размер. Визуализация может отображать только два фактора данных. Первый — это коэффициент, используемый для вычисления объема формы. Второй — цвет, используемый для группировки фигур.

Упаковка кругов (Circle Packing). Этот метод является прямой альтернативой treemap, помимо того, что в качестве примитивной формы он использует круги, которые также могут быть включены в круги с более высокого уровня иерархии. Контейнер внутри каждого круга представляет собой уровень иерархии: каждая ветвь дерева представлена в виде круга, а его ответвления представлены как окружности внутри него. Область каждого круга также может использоваться для представления дополнительного произвольного значения, такого, как количество или размер файла. Цвет также может использоваться для присвоения категорий или представления другой переменной с помощью разных оттенков.

Круговая упаковка — это вариант treemap, который использует круги вместо прямоугольников. Контейнер внутри каждого круга представляет собой уровень иерархии: каждая ветвь дерева представлена в виде круга, а его ответвления представлены как окружности внутри него. Область каждого круга также может использоваться для представления дополнительного произвольного значения, такого, как количество или размер файла. Цвет также может использоваться для присвоения категорий или представления другой переменной с помощью разных оттенков.

Проект The Guardian «Как распространяются слухи о массовых беспорядках в Twitter» (How riot rumours spread on Twitter)¹⁰⁰. Визуализация основана на анализе 2,6 млн твитов. Нужно выбрать слух, чтобы увидеть, насколько необоснованные претензии распространяются в твиттере, прежде чем их успевают подтвердить или опровергнуть в СМИ. Зеленые круги – это поддержка слуха пользователем (чем темнее зеленый, тем больше поддержка). Отрицание слуха — красный. Сомнение в слухе — желтый. Комментирование — серый. Диаметр круга отражает влияние поста на формирование мнения — чем больше круг, тем больше влияние (Рисунок 25).

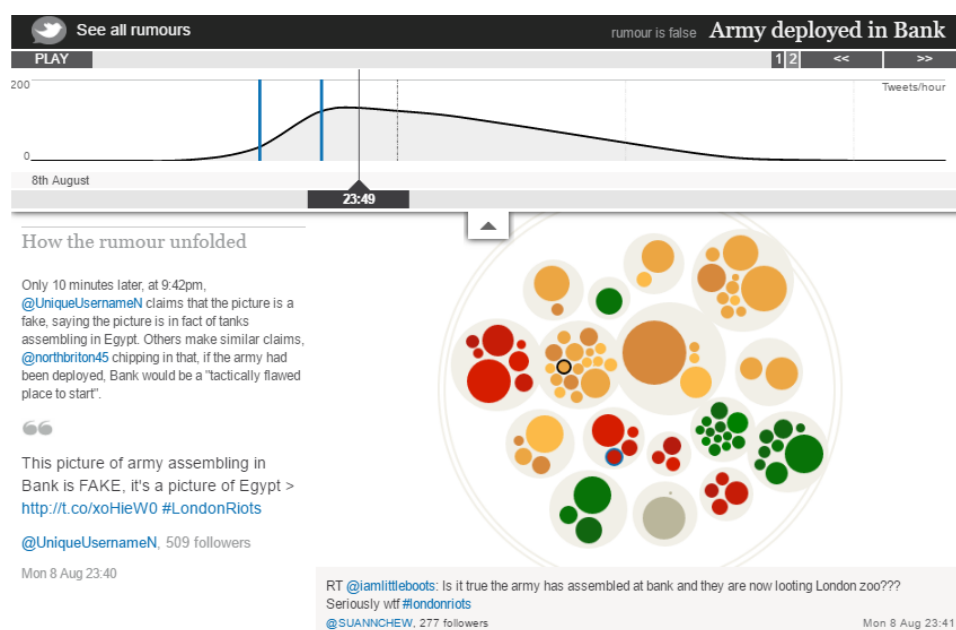


Рисунок 25 — Визуализация проекта How riot rumours spread on Twitter, упаковка кругов

2.5.2 Солнечная диаграмма (sunburst diagram)

Этот тип визуализации показывает иерархию через серию колец, которые разделяются для каждого узла категории. Каждое кольцо соответствует уровню в иерархии: с центральным кругом, представляющим корневой узел, и иерархией, которая направлена к внешним кольцам.

¹⁰⁰ The Guardian [Электронный ресурс]. Точка доступа: URL: <https://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter> (дата обращения: 15.05.17)

Кольца разрезаны и разделены на основе их иерархической связи с исходным срезом. Угол каждого среза либо делится в соответствии с исходным узлом, либо может быть пропорционален значению. Цвет может использоваться для выделения иерархических групп или определенных категорий, как в проекте The Guardian «Права геев в США, по штатам» (Gay rights in the US, state by state).¹⁰¹

Законы о правах в Америке защищают или ограничивают права лиц нетрадиционной ориентации по целому ряду вопросов, включая брак, усыновление, посещение больниц, проживание, издевательства в школе и др. Обработка вопросов прав данной социальной группы варьируется в зависимости от штата, и отслеживает тенденции по регионам. Государство организовано по регионам, поэтому легко увидеть, где находятся штаты страны. Сегментированные концентрические круги составляют основу интерактивного пространства, где каждый круг является проблемой, а каждый сегмент является её состоянием (Рисунок 26).

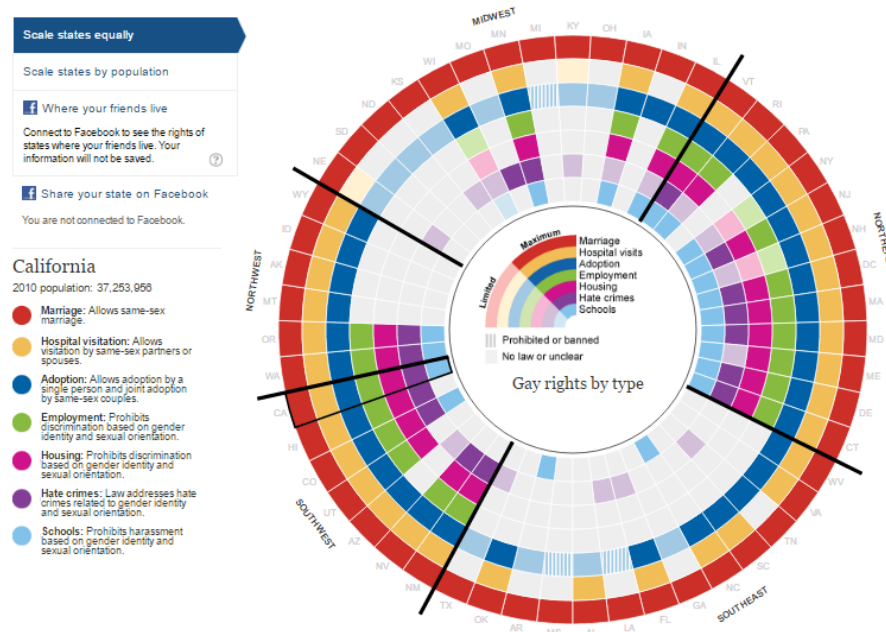


Рисунок 26 — Визуализация проекта Gay rights in the US, state by state, упаковка кругов

¹⁰¹ The Guardian [Электронный ресурс]. Точка доступа: URL: <https://www.theguardian.com/world/interactive/2012/may/08/gay-rights-united-states> (дата обращения: 15.05.17)

2.5.3 Матричная диаграмма (Dot Matrix Chart)

Точечные матричные диаграммы отображают дискретные данные в единицах точек, каждые из которых окрашены для представления определенной категории и сгруппированы в матрице. Они используются, чтобы дать краткий обзор распределения и пропорций каждой категории в наборе данных, а также сравнить распределение и пропорцию по другим наборам данных, чтобы обнаружить шаблоны.

В проекте The Guardian «Понаблюдайте за тем, как распространяются вспышки кори, когда дети были вакцинированы, а когда нет» (Watch how the measles outbreak spreads when kids get vaccinated — and when they don't).¹⁰² В основе визуализации лежат данные предоставленные Министерством здравоохранения США. Матричная диаграмма (симулятор) моделирует ситуацию заражения 10 разных сообществ с разными уровнями вакцинации: вакцинированные, восприимчивые к заболеванию, инфицированные, находящиеся в контакте с инфицированными и вакцинированные, но все равно восприимчивые к инфекции кори. Дети в каждой из групп находятся в тесном контакте, и они смешиваются случайным образом — как, например, на игровой площадке. Важно то, что у каждого из них есть равные шансы войти в контакт с кем-то зараженным корью. Если запускать симуляцию долгое время, то только очень высоко вакцинированные группы из их сообщества могут последовательно блокировать инфекцию кори (Рисунок 27).

¹⁰² The Guardian [Электронный ресурс]. Точка доступа: <https://www.theguardian.com/society/ng-interactive/2015/feb/05/-sp-watch-how-measles-outbreak-spreads-when-kids-get-vaccinated> (дата обращения: 15.05.17)

Watch how the measles outbreak spreads when kids get vaccinated - and when they don't

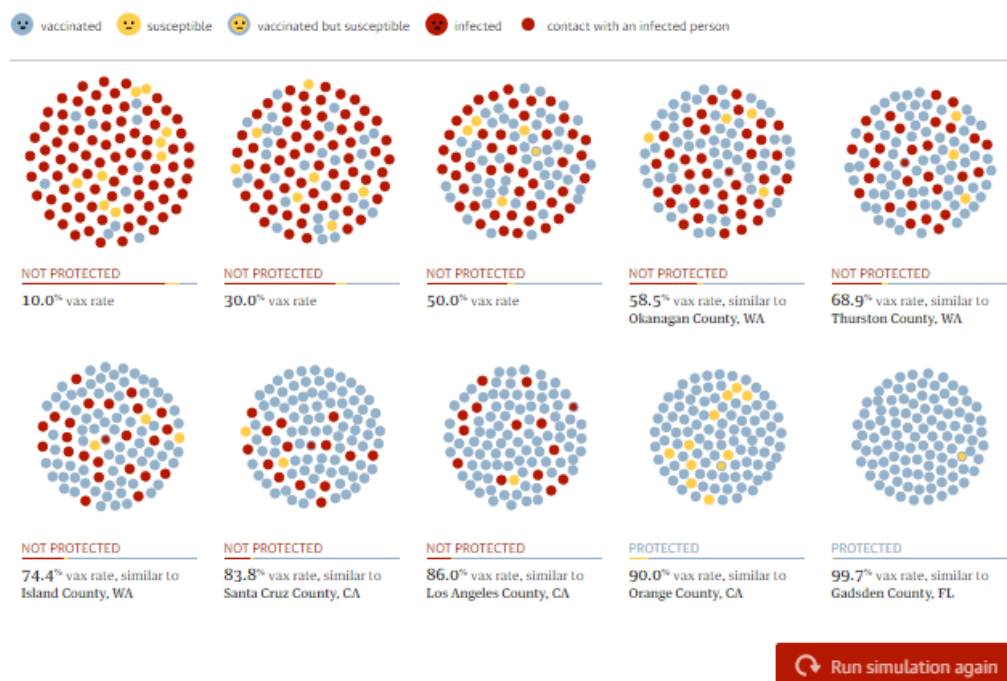


Рисунок 27 — Визуализация проекта

Watch how the measles outbreak spreads when kids get vaccinated — and when they don't, матричная диаграмма

Годом ранее подобный симулятор был опубликован в The Washington Post под названием «Эбола распространяется медленнее, но убивает больше, чем другие болезни» (Ebola spreads slower, kills more than other diseases)¹⁰³. С помощью матричной диаграммы создатели визуализации симулируют распространение инфекции на 100 человек от одного зараженного разными болезнями: Эбола, пневмония, ветряная оспа, корь, свинка, краснуха, коклюш, грипп, дифтерия, корь. По сравнению с другими инфекционными заболеваниями, Эбола распространяется медленно и затрагивает относительно малое количество людей, но по данным Всемирной организации здравоохранения в 70% случаев исход болезни летальный (Рисунок 28).

¹⁰³The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/health/how-ebola-spreads/> (дата обращения: 15.05.17)

Ebola spreads slower, kills more than other diseases

This simulation shows how quickly 10 diseases, from more fatal to less fatal, could spread from one person to 100 unvaccinated people.

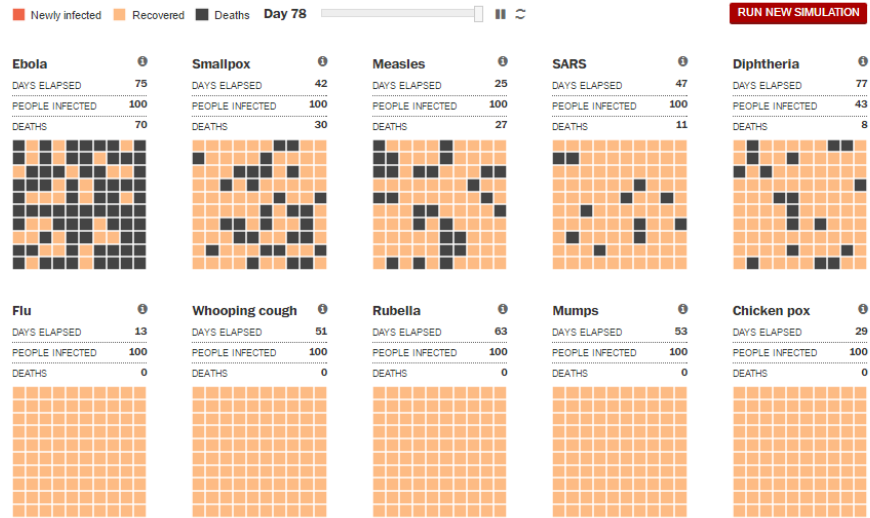


Рисунок 28 — Визуализация проекта

Ebola spreads slower, kills more than other diseases, матричная диаграмма

2.5.4 Параллельные наборы (parallel sets)

В параллельной диаграмме набора каждая линия отражает определенную часть общего набора данных. Ширина каждой линии и путь ее прохождения определяются пропорциональной долей от общего набора. Каждому пути потока можно присвоить свой цвет, чтобы показать и сравнить отношения между разными группами данных.

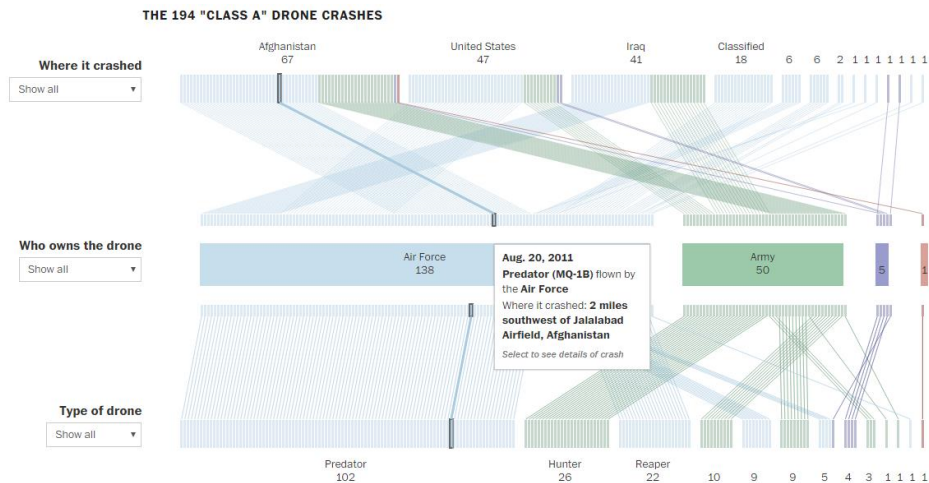


Рисунок 29 — Визуализация проекта Fallen from the skies, параллельные наборы

В проекте The Guardian «Павшие с небес» (Fallen from the skies).¹⁰⁴ Визуализация основана на данных о 194 падений военных американских дронов, каждое из которых нанесло государству ущерб как минимум в 2 миллиона долларов. По каждому случаю падения предоставлена информация о дате и месте падения, а также классе и принадлежности к роду войск дрона (Рисунок 29).

2.5.5 Пиктограммы (pictogram chart)

Для визуальной привлекательности графика используются значки, которые представляют тему или категорию данных, например, данные о населении будут использовать значки людей. Каждый значок может представлять одну единицу или любое число, или единицы (например, каждый значок представляет 10). Наборы данных сравниваются бок о бок в обоих столбцах или рядах значков, чтобы сравнить каждую категорию друг с другом.

Использование значков может иногда помочь преодолеть различия в языке, культуре и образовании. Значки также могут давать более репрезентативное представление данных (например, если у нас данные про 8 домов, то в диаграмме отображаются 8 иконок с домами. Чаще всего пиктограммы встречаются как способ визуализации малых наборов данных, но в некоторых случаях графики являются хорошим дополнением и к большому массиву данных.

¹⁰⁴ The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/national/drone-crashes/database/> (дата обращения: 15.05.17)



Рисунок 30 — Визуализация проекта An eye for an eye?, пиктограмма

Например, как в проекте The Washington Post под названием «Глаз за глаз» (An eye for an eye?)¹⁰⁵, который основан на данных о смертных приговорах за 2085 убийств в США с 1977 по 2014 год. В левой части пиктограммы представлены казненные убийцы, а в правой части — количество убитых этими казненными преступниками (Рисунок 30). С помощью цветового кодификатора можно определить расу преступников и жертв. Вторая часть проекта — это гистограммы на основе тех же данных, в которой данные распределяются по следующим категориям: возрасту, методу казни, гендерной принадлежности и региону преступления (Рисунок 31).

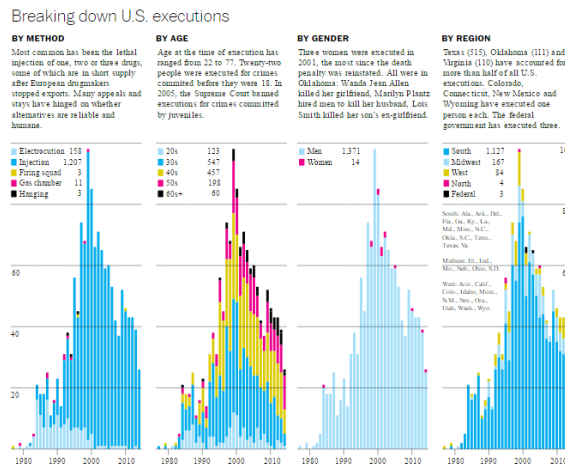


Рисунок 31 — Визуализация проекта An eye for an eye?, гистограмма

¹⁰⁵ The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/outlook/death-penalty/> (дата обращения: 15.05.17)

Или пиктограмма The Washington Post под названием «Число жертв в Палестино-Израильском конфликте» (The death toll in the Gaza-Israel conflict)¹⁰⁶, который основан на данных о жертвах Палестино-Израильского конфликта с 8 июля по 6 августа 2014 года. Люди распределены на гражданских и военных с левой и с правой стороны и на тех, чей статус неизвестен. При наведении на иконку появляется информация про возраст и дату смерти человека (Рисунок 32).



Рисунок 32 — Визуализация проекта
The death toll in the Gaza-Israel conflict, пиктограмма

2.5.6 Хордовая диаграмма (chord diagram)

Этот тип диаграммы визуализирует взаимосвязи между объектами. Связи между объектами используются для того, чтобы показать, что их объединяет. Это делает диаграмму идеальной для сравнения сходств в наборе данных или между различными группами данных.

Узлы в графике расположены вокруг круга, причем отношения между точками связаны друг с другом либо с помощью дуг, либо с помощью кривых

¹⁰⁶ The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/world/gaza-counter/> (дата обращения: 15.05.17)

Безье. Значения присваиваются каждому соединению, которое выглядит пропорционально размеру каждой дуги. Цвет можно использовать для группировки данных по различным категориям, что помогает в сравнении и идентификации каждой из групп.

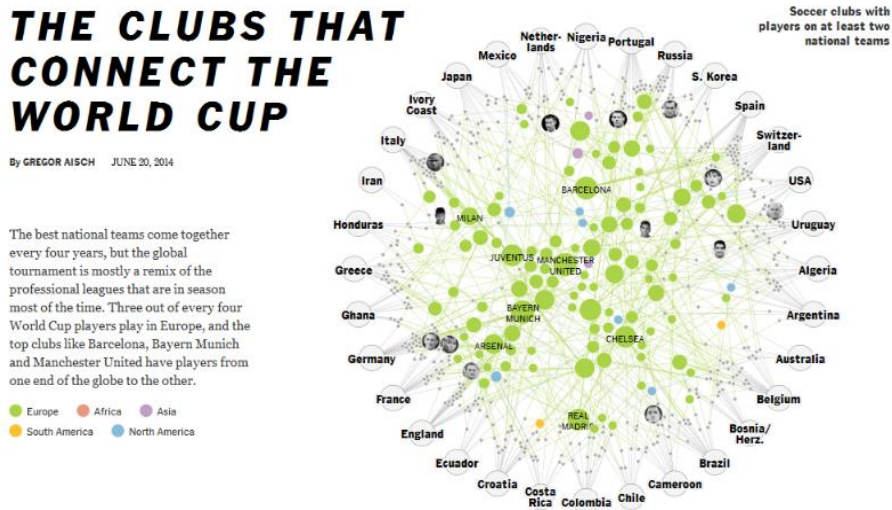


Рисунок 33 — Визуализация проекта Clubs that connect World Cup, не-ленточная хордовая диаграмма

Не-ленточная хордовая диаграмма представляет собой усеченную версию хордовой — с указанием только узлов и линий соединения. Это обеспечивает акцент на соединениях внутри данных. Проект The New York Times «Клубы, объединенные кубком мира» (Clubs that connect World Cup)¹⁰⁷. Диаграмма показывает интернациональность команд — как сборная команда каждой страны может быть связаны с клубами других стран. Цветовая кодировка по материкам – Европа, Азия, Африка, Северная и Южная Америки (Рисунок 33).

2.5.7 Диаграмма параллельных координат (parallel coordinates plot)

Этот вид визуализации используется для построения многомерных дискретных данных. Диаграмма параллельных координат отлично подходит для одновременного сравнения многих переменных и выявления взаимосвязей

¹⁰⁷ URL: <https://www.nytimes.com/interactive/2014/06/20/sports/worldcup/how-world-cup-players-are-connected.html>

между ними. В данном типе визуализации каждой переменной присваивается своя ось, а все оси расположены параллельно друг другу. Каждая ось может иметь разный масштаб в зависимости от переменной. Значения отображаются в виде ряда линий, соединенных по каждой оси. Это означает, что каждая строка представляет собой набор точек, расположенных на каждой оси, которые были соединены вместе. Изменение порядка осей может помочь в обнаружении закономерностей или корреляций между переменными.

Проект The New York Times «Как лояльность болельщиков менялась во время Кубка мира» (How Fan Loyalty Changed During the World Cup)¹⁰⁸. С помощью диаграммы параллельных координат были проанализированы данные социальной сети Facebook, а это активность около 300 млн. болельщиков по всему миру. Создатели анализировали аудиторию стран участниц кубка мира в Facebook, оценивая, за какие сборные болели в той или иной стране на каждом этапе чемпионата. Большинство болельщиков поддерживало сборную своей страны, а в случае вылета команды, в основном болели за сборную Германии (Рисунок 34).

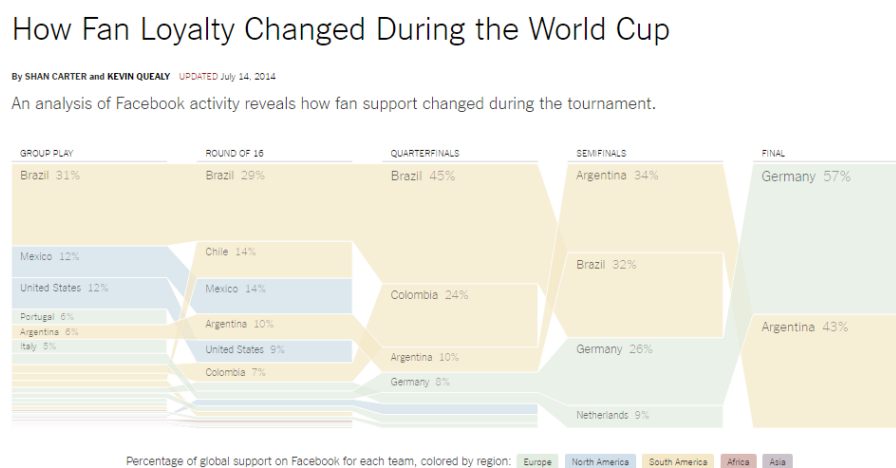


Рисунок 34 — Визуализация проекта
How Fan Loyalty Changed During the World Cup, диаграмма параллельных
координат

¹⁰⁸ The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2014/06/20/sports/worldcup/how-world-cup-players-are-connected.html> (Дата обращения: 15.05.17)

2.5.8 Сложная столбчатая диаграмма (Stacked Bar Graphs)

Классический вариант столбчатой диаграммы использует горизонтальные или вертикальные столбцы, чтобы показать дискретное числовое сравнение по категориям. Одна ось диаграммы показывает, какие конкретные категории сравниваются, а другая ось представляет собой шкалу дискретных значений. Сложная столбчатая диаграмма показывает, как более крупная категория делится на мелкие категории и какова взаимосвязь каждой этой части от общей суммы. Одним из основных недостатков этой визуализации является то, что большое количество сегментов затрудняет понимание графика. Кроме того, сложно сравнивать каждый сегмент друг с другом, поскольку они не выровнены по общей базовой линии.

Проект The New York Times «Расовый разрыв в полицейских подразделениях Америки» (The Race Gap in America's Police Departments)¹⁰⁹. По данным Бюро статистики юстиции, в сотнях полицейских департаментов по всей стране доля европеоидной расы в правоохранительных органах на 30 процентов выше, чем в округах, которым они служат. Согласно данным опроса 2007 года, меньшинства составляют четверть полицейских сил. В сложной столбчатой диаграмме перечислены местные полицейские отделы из 17 столичных районов, отсортированные таким образом, чтобы наверху находились отделы с наибольшим процентным соотношением белых офицеров и белых жителей. График отображает расовое соотношение сотрудников полиции и населения округа, с которым она работает (Рисунок 35).

¹⁰⁹ The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2014/09/03/us/the-race-gap-in-americas-police-departments.html> (Дата обращения: 15.05.17)

Baltimore

Baltimore's police department has a lower percentage of blacks than the population it serves. But in contrast to other cities that have been wracked by tension and protests over police confrontations with black men, the city's mayor, its police commissioner, the state's attorney are all black, giving a somewhat different tenor to clashes between the power structure and its critics.

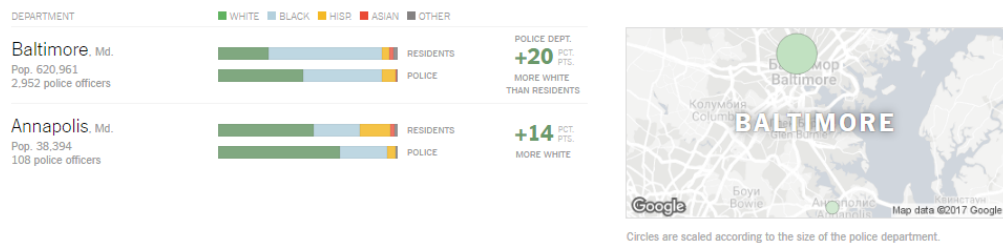


Рисунок 35 — Визуализация проекта

The Race Gap in America's Police Departments, сложная столбчатая диаграмма

2. 6 Способы визуализации непрерывных временных данных

В результате контент-анализа, было определено, что 57% от общего числа проектов основаны на непрерывных временных данных. Чаще всего большие массивы данных данного типа использовали в The New York Times (28%). В The Guardian — 4%, а в The Washington Post — 25%.

2.6.1 Потокный график (streamgraph)

Потокный график — это тип визуализации с накоплением, который перемещается вокруг центральной оси, что приводит к плавной и органичной форме (особенно при работе с big data). Этот метод показывает тенденции для разных наборов событий, количества его появлений, его относительных показателей и т. д. Таким образом, может быть множество подобных событий, показанных на временной шкале на изображении.

Метод имеет две цели: показать множество отдельных временных рядов, а также передать их сумму. Так как высоты отдельных слоев суммируются с высотой общего графика, можно одновременно достичь обеих целей. В то же время это предполагает определенные компромиссы. Между слоями не может

быть пробелов, так как это исказит их сумму. Из-за отсутствия промежутков между слоями изменения среднего слоя неизбежно вызовут вихри во всех других окружающих слоях, искажения, которые не имеют никакого отношения к базовым данным тех затронутых временных рядов.

Проект The New York Times «Откуда мы вышли, и куда мы пришли» (Where We Came From and Where We Went, State by State)¹¹⁰. Этот потоковый график воспроизводит миграцию внутри страны с начала прошлого века. В данной визуализации ленты окрашены в цвета по регионам, а жители, наоборот переехавшие в этот регион из другого места, размещены на серых лентах внизу. Многослойность графика позволяет создать полную картину миграции и репатриации выбранного округа. Главное удобство этого графика в том, что вы можете видеть изменения в населении округа на протяжении многих лет (Рисунок 36).

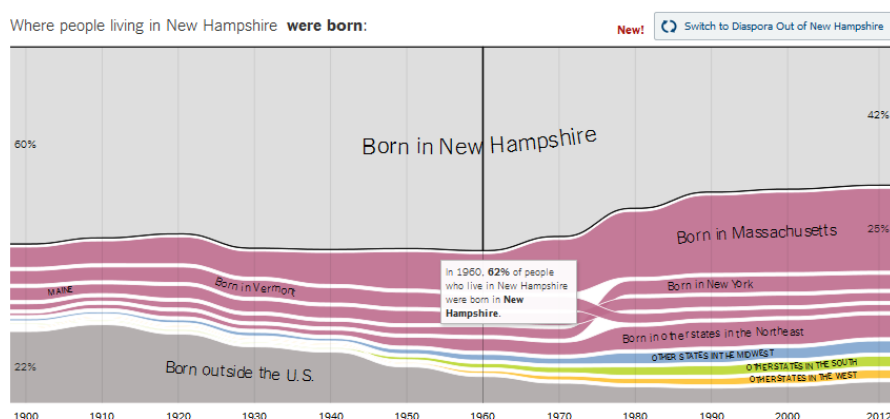


Рисунок 36 — Визуализация проекта Where We Came From and Where We Went, State by State, потоковой график

2.6.2 Графики «открыто-высокого-низкого закрытия» (или диаграммы OHLC)

Эти графики чаще всего используются для визуализации изменения цен валют, акций, ценных бумаг, товаров, валют во времени. Они полезны для

¹¹⁰ The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2014/09/03/us/the-race-gap-in-americas-police-departments.html> (дата обращения: 15.05.17)

прогноза будущих изменений цен через созданные графики. Каждый отдельный график «открытого-высокого-низкого закрытия» представляет собой два диапазона по разным осям: самые высокие цены и самые низкие цены, а также цена открытия и закрытия в одном временном промежутке (например, за неделю). Высокие и низкие ценовые диапазоны представлены длиной вертикальной линией. Цены открытия и закрытия представлены вертикальным позиционированием галочек, которые появляются слева (представляющие цену открытия), и справа (представляющие цену закрытия) сторон высоко — низкой вертикальной линии.

Несмотря на, казалось бы, узкую направленность этого способа визуализации, необычное применение ему нашли в проекте The New York Times «Законы об оружии принятые в течение года после Ньютауна» (State Gun Laws Enacted in the Year After Newtown)¹¹¹. Вместо ценовой категории создатели проекта использовали в качестве единицы — законы об оружии. По оси икс — дни принятия закона, по оси игрек — этапы принятия решения по закону (подписание, внесение изменений, принятие). Закон подписывается, когда линия достигает вершины, где зеленая галочка обозначает более свободные ограничения в пользовании оружием (график открытия), а оранжевый цвет — жесткие ограничения в пользовании оружием (график закрытия). Все законы группируются на следующие тематические подгруппы: разрешение на оружие, разрешение на его публичный перенос, штурмовое оружие, недействительный федеральный закон, доступ к оружию, потерянное или украденное огнестрельное оружие и др. (Рисунок 37).

¹¹¹ The New York Times [Электронный ресурс]. Точка доступа: <http://www.nytimes.com/interactive/2013/12/10/us/state-gun-laws-enacted-in-the-year-since-newtown.html> (дата обращения: 15.05.17)

State Gun Laws Enacted in the Year After Newtown

About 1,500 state gun bills were introduced in the year after the Newtown massacre. 178 passed at least one chamber of a state legislature. 109 have become law.

In the 12 months after the mass shooting at Sandy Hook Elementary School in Newtown, Conn., almost every state enacted at least one new gun law, according to a database compiled by the Law Center to Prevent Gun Violence. Nearly two-thirds of the new laws ease restrictions and expand the rights of gun owners. Most of those bills were approved in states controlled by Republicans. Those who support stricter regulations won some victories — mostly in states where the legislature and governorship are controlled by Democrats — to increase restrictions on gun use and ownership. Select categories from the table below to see all gun bills that passed at least one chamber of a state legislature.

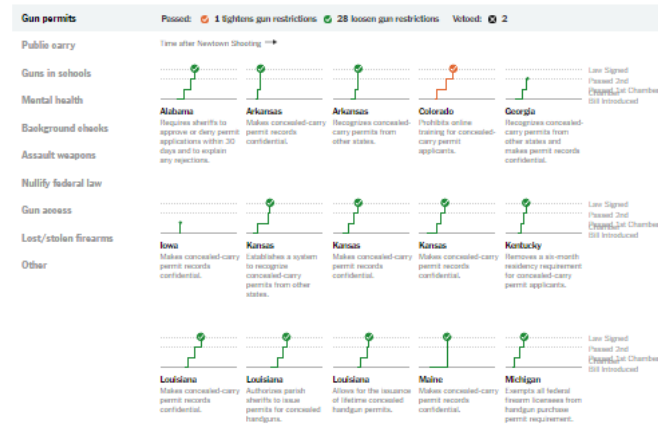
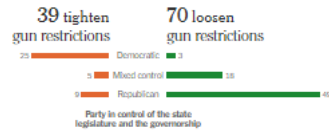


Рисунок 37 — Визуализация проекта State Gun Laws Enacted in the Year After Newtown, график «открытого-высокого-низкого закрытия»

2.6.3 Линейный график (line chart)

Самые разные виды линейных графиков широко представлены в проектах big data. Они используются для отображения количественного значения в определенном временном интервале. Чаще всего используется для отображения тенденций или отношений (в случае, если сгруппированы с другими линиями). Стандартность графика не мешает специалистам в области визуализации находить неординарные темы для визуализации. Например, как в проекте The New York Times «Как год рождения влияет на политические взгляды?» (How Birth Year Influences Political Views)¹¹².

¹¹² The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2014/07/08/upshot/how-the-year-you-were-born-influences-your-politics.html> (дата обращения: 15.05.17)

Новая модель президентских выборов предполагает, что рейтинг одобрения президента Обамы, который в настоящее время находится на низком уровне, будет информировать не только о выборах 2016 года, но и о выборах в 2076 году. Исследователи из Колумбийского университета и американская компания Catalist использовали сотни тысяч ответов на опрос и новое статистическое программное обеспечение для оценки того, как политические предпочтения людей изменяются на разных этапах их жизни.

Модель предполагает, что поколения избирателей выбирают свою команду, демократов или республиканцев, исходя из накопленного ими жизненного опыта — «подсчета» событий. Используя рейтинг президента Гэллапа в качестве доверенного лица для этих событий, Яир Гица, главный ученый в Каталисте, и Эндрю Гельман, политолог и статистик из Колумбийского университета, смогли оценить, в каком возрасте формируются политические предпочтения респондентов.

Изменяя возраст рождения на таймлайне, мы можем наблюдать изменения на линейном графике. Республиканцам или демократам отдали предпочтения люди, рожденные в конкретный год (через процент проголосовавших) на выборах в 2012 году. Например, 53% людей, родившихся со мной в один год, проголосовали за демократов (Рисунок 38).

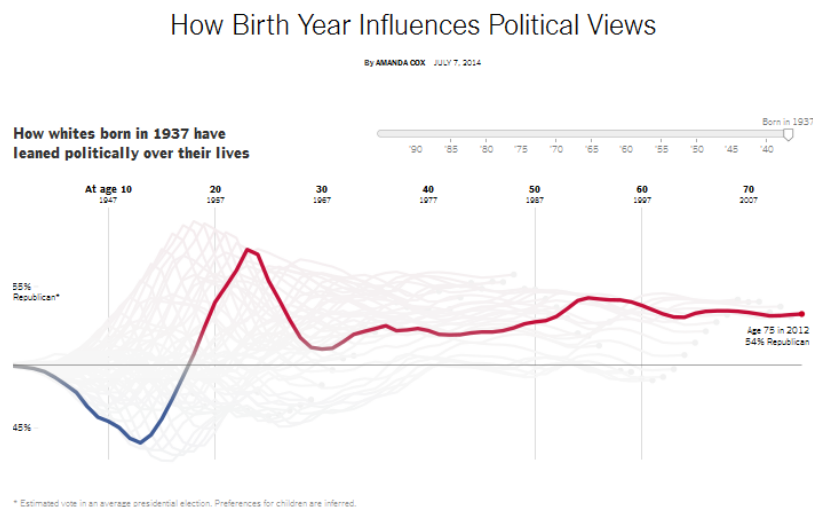


Рисунок 38 — Визуализация проекта How Birth Year Influences Political Views, линейный график

Еще один проект The New York Times «Как изменилась налоговая нагрузка?» (How the Tax Burden Has Changed)¹¹³. Большинство американцев платили меньше налогов в 2010 году, чем люди с теми же доходами, скорректированными с учетом инфляции, выплаченные в 1980 году из — за сокращения федеральных подоходных налогов. Тем не менее, при более низких уровнях доходов большая часть сэкономленных средств была скомпенсирована ростом федеральных налогов на заработную плату, государственных налогов с продаж и местных налогов на недвижимость. Около половины домохозяйств, зарабатывающих менее 25 тыс. долларов, вообще ничего не сохранили. Майк Босток, Мэтью Эриксон и Роберт Гебелов из New York Times исследовали изменение налоговых ставок с 1980 по 2010 год для различных уровней доходов (Рисунок 39).

How the Tax Burden Has Changed

Most Americans paid less in taxes in 2010 than people with the same inflation-adjusted incomes paid in 1980, because of cuts in federal income taxes. At lower income levels, however, much of the savings was offset by increases in federal payroll taxes, state sales taxes and local property taxes. About half of households making less than \$25,000 saved nothing at all. [About the Data](#) | [Related Article](#)

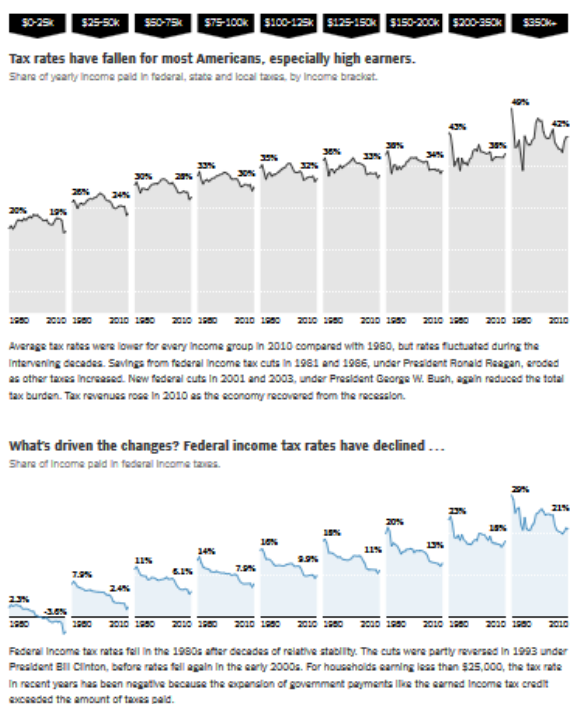


Рисунок 39 — Визуализация проекта How the Tax Burden Has Changed, линейный график

¹¹³ The New York Times [Электронный ресурс]. Точка доступа:

<http://www.nytimes.com/interactive/2012/11/30/us/tax-burden.html> (дата обращения: 15.05.17)

Проект The Washington Post «Почему растут показатели смертности белых женщин в американских деревнях?» (Why death rates for white women in rural America are spiking?)¹¹⁴. Данный проект представляет собой интерактивный линейный график. Пользователь может выбрать определенную категорию женщин: по возрасту, гендерному признаку, расе и месту жительства, и узнать то, как увеличивалась смертность той или иной социальной страты с 1994 по 2014 год (Рисунок 40). Создатели проекта работают с пропорциями во времени — анализируются непрерывные данные Центра по контролю и заболеваемости и профилактики смертности. Проект дополнен столбчатой диаграммой, в основе которого лежат данные об изменении среднего числа смертей на сто тысяч человек с начала века (четыре категории — мужчины, женщины, передозировка или самоубийство).

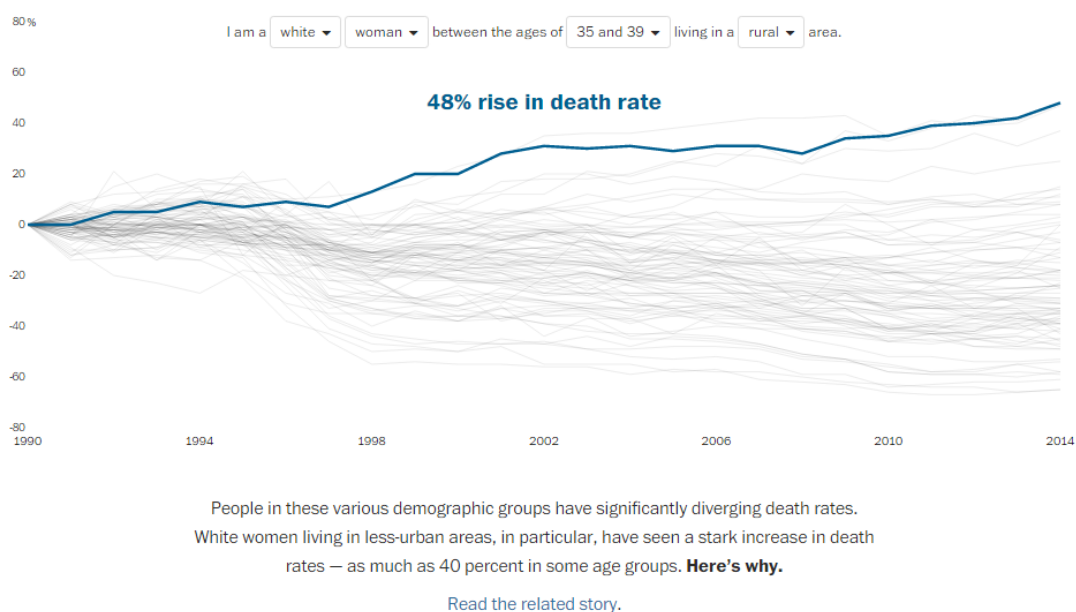


Рисунок 40 — Визуализация проекта Rising death rates for white women, линейный график

¹¹⁴ The Washington Post [Электронный ресурс]. Точка доступа: <https://www.washingtonpost.com/graphics/national/white-death/> (дата обращения: 15.05.17)

Проект The New York Times «Как рецессия изменила экономику?» (How the Recession Reshaped the Economy, in 255 Charts)¹¹⁵. Через пять лет было восстановлено девять миллионов рабочих мест, которые потеряла во время Великой рецессии экономика США. Но не все отрасли восстановились одинаково (Рисунок 41).

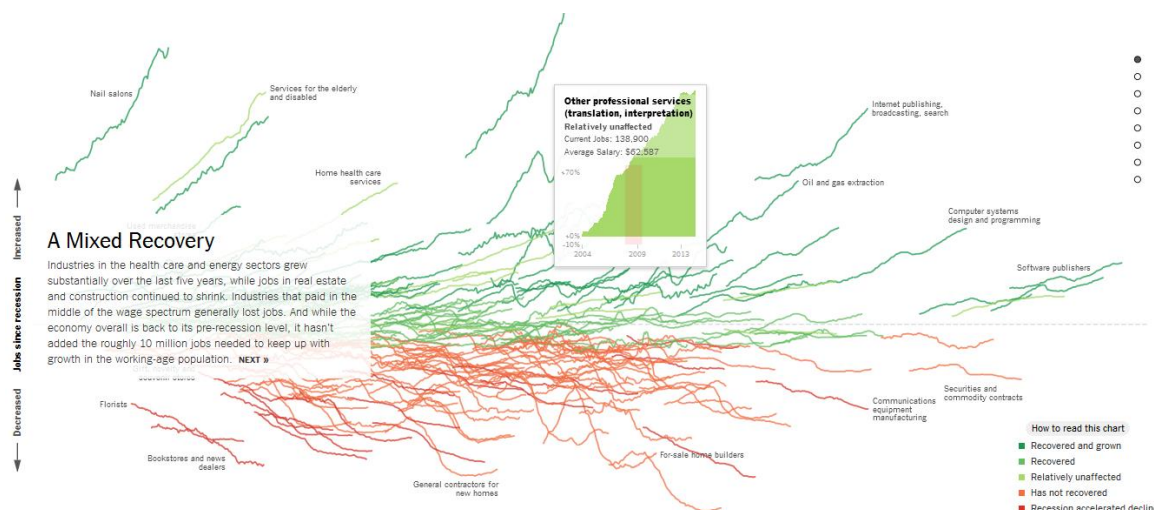


Рисунок 41 — Визуализация проекта How the Recession Reshaped the Economy, in 255 Charts, линейный график

В основе проекта два типа графиков – интерактивный линейный график, на основе дискретной информации, и вспомогательный график с областями на основе непрерывных временных данных. График в верхней части страницы включает линейные диаграммы для каждой отрасли, которые показывают изменения в занятости за последнее десятилетие. Отдельные линии помещаются по оси X (горизонтально) на основе средней заработной платы рабочих, выплачиваемой в этой отрасли. Другие линии размещаются по оси Y (вертикально) на основе процентного изменения занятости с начала рецессии в декабре 2007 года. Цветовой кодификатор означает состояние сферы деятельности: от восстановленной и улучшенной до все еще находящейся в упадке.

¹¹⁵ The New York Times [Электронный ресурс]. Точка доступа: <https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html> (дата обращения: 15.05.17)

Профессиональные области, которые были отображены в графике: Business, Buying & Selling, Construction, Eating & Drinking, Energy & Resources, Entertainment, Finance, Health, Home, Insurance, Manufacturing, Media, Other Services, Private Services, Professional Services, Real Estate, Research, Retail, Social Services and Organizations, Sports, Transportation, Travel (Рисунок 42).

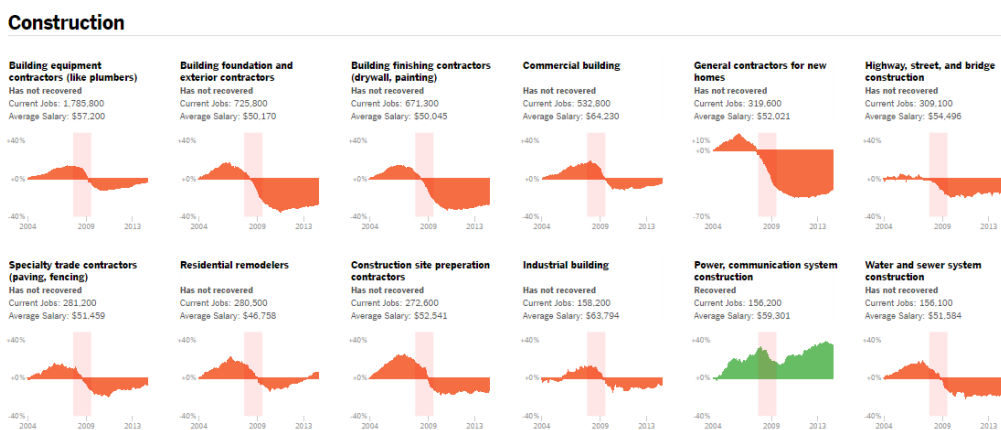


Рисунок 42 — Визуализация проекта How the Recession Reshaped the Economy, in 255 Charts, график с областями

Данные поступают из программы статистики занятости «Бюро статистики труда», которая предоставляет ежемесячные отчеты на основе опросов предприятий в сотнях отраслей.

2.6.4 Диаграмма рассеяния (scatterplots)

Диаграммы рассеяния используют набор точек, размещенных на оси координат, чтобы отображать значения двух переменных. Отображая переменную по каждой оси, график определяет, существует ли связь или корреляция между этими двумя переменными. Внутри графика могут располагаться линии или кривые, чтобы помочь пользователю в анализе данных, их можно нарисовать как можно ближе ко всем точкам, а также показать, как будут выглядеть все сгруппированные точки в одной линии — линии тренда.

Проект The Washington Post «Как долго мы будем жить и насколько хорошо?» (How long will we live — and how well?)¹¹⁶. В основе проекта лежит интерактивная диаграмма рассеяния. Каждая точка на диаграмме обозначает страну, а цветом показаны материки, к которым относятся эти страны. Нажав на точку, пользователь может узнать, как за 20 лет изменилась продолжительность жизни и возраст, в котором житель страны еще не сталкивался с серьезными заболеваниями. Точки распределены по двум осям — возрасту человека и проценту здоровой жизни. Создатели проекта визуализировали исследование Института показателей и оценки состояния здоровья «Ожидаемая продолжительность здоровой жизни для 187 стран с 1990 — 2010 годы: систематический анализ для исследования глобального бремени болезней» (Рисунок 43).

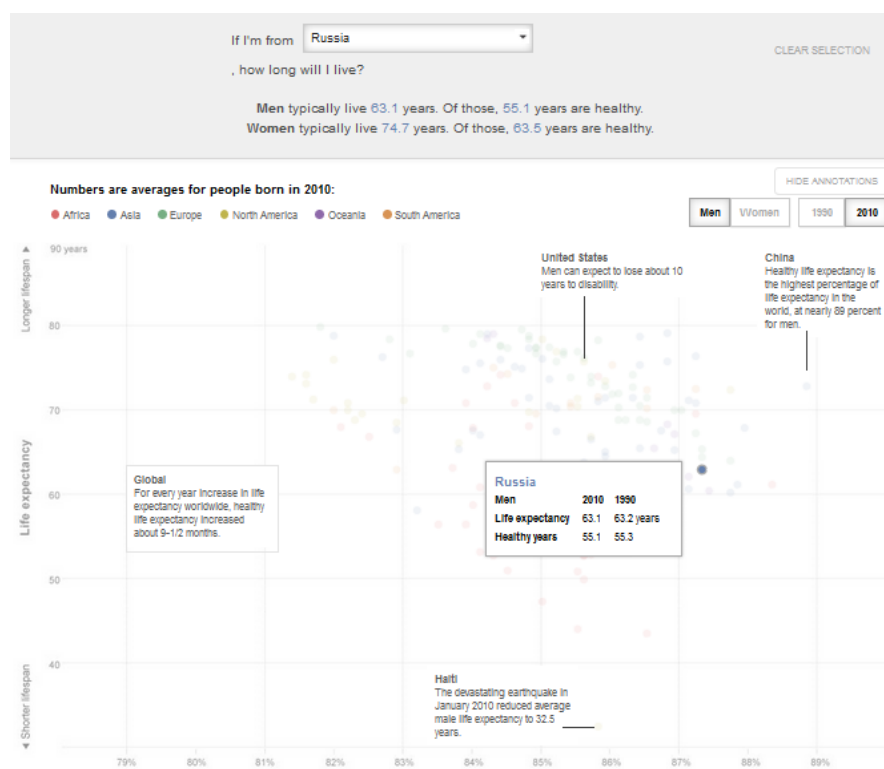


Рисунок 43 — Визуализация проекта How long will we live — and how well?,
диаграмма рассеяния

¹¹⁶ The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/health/healthy-life-expectancy/> (дата обращения: 15.05.17)

Проект The Washington Post «Глобальный взгляд на сердечные факторы риска» (A global look at cardiac risk factors)¹¹⁷. В проекте представлена интерактивная диаграмма рассеяния с данными с 1980 по 2008 год. Мужчины и женщины в возрасте от 18,5 до 35 лет и данные об индексе массы тела. Категории в основе графика: нормальный вес, перевес и ожирение. График показывает, как уменьшалась или увеличилось количество людей с избытком веса с 1980 по 2008 год (Рисунок 44).



Рисунок 44 — Визуализация проекта A global look at cardiac risk factors, диаграмма рассеяния

В результате контент-анализа, было определено, что 45% от общего числа проектов основаны на географических данных, 32% — на дискретных и 23% на непрерывных временных данных. Способы визуализации напрямую

¹¹⁷ The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/health/weight-of-the-world-bmi/> (дата обращения: 15.05.17)

зависят от типа информации в базах данных, которые используются для визуализации (Рисунок 45).

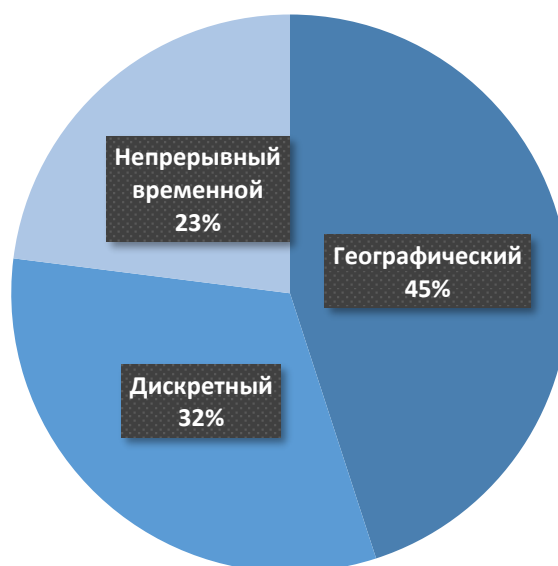


Рисунок 45 — Типы данных в проанализированных проектах big data

Выработанные критерии и индикаторы показали свою эффективность в решении практической задачи по определению зависимости типа информации в проекте от способа его визуализации. Изучив проекты журналистики данных и систематизировав их по типу информации, мы пришли к выводу, что способы визуализации находятся в прямой зависимости от данных в их основе (Таблица 2.2). Помимо этого, в некоторых случаях мы обнаружили зависимость способа визуализации от темы и цели проекта. Например, для визуализации данных в проектах «Глаз за глаз» (An eye for an eye?) или «Число жертв в Палестино-Израильском конфликте» (The death toll in the Gaza-Israel conflict) про жертв вооруженного конфликта и несправедливой судебной системы используется пиктограмма. Матричную диаграмму выбрали сразу два издания, чтобы показать, как распространяется болезнь на разные группы населения¹¹⁸.

¹¹⁸ The Washington Post [Электронный ресурс]. Точка доступа: <http://www.washingtonpost.com/wp-srv/special/health/how-ebola-spreads/> (дата обращения: 15.05.17); The Guardian [Электронный ресурс]. Точка доступа: <https://www.theguardian.com/society/ng-interactive/2015/feb/05/-sp-watch-how-measles-outbreak-spreads-when-kids-get-vaccinated> (дата обращения: 15.05.17)

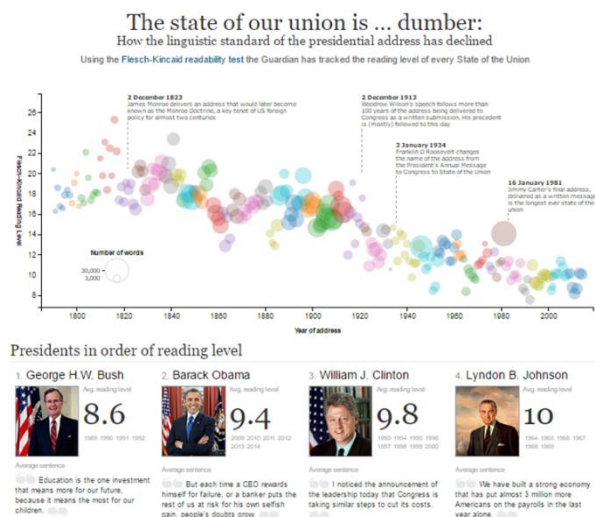


Рисунок 46 — Визуализация данных с помощью пузырьковой диаграммы

Для визуализации данных в проектах 2016 года «История в словах президента» (History through the president’s words) и 2013 года «Обращение “О положении страны” ... глупеет: как уменьшается лингвистический стандарт президентского послания» (The state of our union is ... dumber: how the linguistic standard of the presidential address has declined)¹¹⁹ используются пузырьковые диаграммы. В обоих проектах анализируется такой дискретный тип информации как слова, сказанные разными президентами в разный временной период.

Таблица 2.2 — Классификация способов визуализации по типу данных

№	Тип данных	Способ визуализации
1	Дискретные данные	Деревья (упаковка кругов) Солнечная диаграмма Матричная диаграмма Параллельные наборы Пиктограммы Хордовая диаграмма Диаграмма параллельных координат Сложная столбчатая диаграмма

¹¹⁹ The Washington Post [Электронный ресурс]. Точка доступа: <https://www.washingtonpost.com/graphics/politics/2016-sotu/language/> (дата обращения: 15.05.17); The Guardian [Электронный ресурс]. Точка доступа: <https://www.theguardian.com/world/interactive/2013/feb/12/state-of-the-union-reading-level> (дата обращения: 15.05.17)

2	Географические данные	Точечные карты Пузырьковая карта Хороплет Карты связей Карты потока
3	Непрерывные временные данные	Диаграмма рассеяния Потоковый график График «открыто-высокого-низкого-закрытия» Линейный график

Мы выделили некоторые тенденции, характерные для визуализации проектов на основе big data в современной журналистике.

Во-первых, широко распространено использование географического типа информации для визуализации данных. Несмотря на сложность и многоаспектность визуализации с этим типом данных, у проанализированных изданий карты неизменно оставались в приоритете с 2011 по 2016 год. На данном этапе активно развиваются и осваиваются онлайн-ресурсы для работы с геоинформационными данными, которые позволяют просто загрузить файлы с расширением XLSX или CSV в программу и получить эффектную визуализацию. Использование программного кода помогает обогатить ее дополнительной информацией и графиками.

Во-вторых, нами зафиксировано частое использование классических способов визуализации данных, таких, как карта или линейный график. Эта тенденция является следствием первой: зависимость географического типа информации от такого способа визуализации, как карта. Линейный графики и карты — наиболее привычные пользователю графики, которые легко и удобно считываются и чаще всего не требуют дополнительных усилий для понимания.

В-третьих, идентичная структура проектов в области дата-журналистики. Проводя данное исследование, мы не планировали исследовать жанровые характеристики в проектах журналистики данных. Тем не менее, мы

обратили внимание на характерную для всех проектов структуру: заголовок, дата публикации, имя автора, основной график и статья с второстепенными графиками (или с составными графиками основного), которая раскрывает тему проекта. В The Guardian обязательной частью проекта является сноска с источниками данных, в которых пишут не только про источники, но и про погрешности в измерениях данных в работе.

И в-четвертых, ориентация проектов на локальные новости. Во многом это связано с ориентацией на местную аудиторию исследуемых изданий.

ЗАКЛЮЧЕНИЕ

Роль больших данных в академической науке обсуждали еще в XVIII веке давно, но только с появлением компьютеров появилась возможность применять компьютерные методы обработки данных во всех научных сферах. Любые оцифрованные данные могут стать объектом научной или любой другой публикации, а пользователь может самостоятельно проверить подлинность опубликованных в ней фактов. Процесс накопления данных бесконечен, и, осознавая масштабы вызванных этим фактом изменений, Клиффорд Линч предложил новое название для этой парадигмы — Big Data. Эти данные могут быть выражены разными неструктурированными форматами. В конечном итоге, мы имеем набор данных, который может быть слишком большим, слишком «сырым» или слишком неструктурированным для классических методов обработки.

В разных компаниях датчики, видеокамеры, интеллектуальные счетчики и другие подключенные устройства производят огромные объемы данных, которые добавляются к уже хранящейся на сервере информации. Журналист нужно уметь разглядеть в этих огромных массивах данных полезную информацию. Появление data journalism стало реакцией на проблемы и выходящие из них потребности.

Дата-журналистика — это деятельность по анализу доступных баз данных, обработке полученной в ходе анализа информации, и ее интерактивной визуализации с целью информирования читателя. Отсюда, проект журналистики данных — это мультимедийный проект, основанный на анализе баз данных и визуализированный с помощью интерактивного интерфейса.

Визуализация — это инструмент, который знакомит нас уже с конечным результатом анализа баз данных, позволяет нам управлять процессом этого знакомства и даже выбрать нужное направление в исследовании данных.

Графический образ и вспомогательные элементы являются составной частью графика. Графический образ данных — это результат визуализации. Представление данных в виде графических образов позволяет пользователю увидеть изучаемое явление или процесс изнутри, облегчая понимание данных и обеспечивая поддержку для принятия решений.

Последовательные попытки систематизировать большое количество способов для визуализации данных говорит о наличии проблем в этой области, которые требуют решения. Мы сформулировали следующую гипотезу исследования: при выборе способа визуализации специалист руководствуется типом визуализируемых данных (дискретные, географические или непрерывные временные данные), темой проекта, а также его целью. Для проверки данной гипотезы мы провели формализованный анализ визуализаций big data ведущих интернет-СМИ Европы и США.

Анализ показал, что при выборе способа визуализации специалист может руководствоваться следующей классификацией:

1. **Географический** тип данных (точечные и пузырьковые карты, карты-хороплеты, карты связей и потока);
2. **Дискретный** тип данных (солнечные, матричные, хордовые, сложные столбчатые диаграммы, а также параллельные наборы, пиктограммы, диаграмм параллельных координат);
3. **Непрерывно временной** тип данных (линейный и потоковые графики. диаграмма рассеяния, график «открыто-высокого-низкого-закрытия»).

В качестве эмпирической базы для контент-анализа было взято 53 проекта журналистики данных из изданий (The Guardian, The New York Times, The Washington Post) за период с 2011 по 2016 год. Из них 6 материалов The Guardian — 15%, 21 материал The New York Times — 39%, 24 материала The Washington Post — 46%. Это все опубликованные материалы изданий за указанный промежуток времени с визуализацией большого массива данных.

Изучив проекты журналистики данных и систематизировав их по типу информации, мы пришли к выводу, что способы визуализации находятся в

прямой зависимости от данных в их основе. Помимо этого, частично подтвердилась и зависимость способа визуализации от темы и цели проекта.

Мы выделили некоторые тенденции, характерные для визуализации проектов на основе big data в современной журналистике. Во-первых, широко распространено использование географического типа информации для визуализации данных. Во-вторых, нами зафиксировано частое использование классических способов визуализации данных, таких, как карта или линейный график. Эта тенденция является следствием первой: зависимость географического типа информации от такого способа визуализации, как карта. Линейный график и карты — наиболее привычные пользователю способы визуализации данных, которые легко и удобно считываются и чаще всего не требуют дополнительных усилий для понимания. В-третьих, идентичная структура проектов в области дата-журналистики. И в-четвертых, ориентация проектов на локальные новости. Во многом это связано с ориентацией на местную аудиторию исследуемых изданий.

На секции «Стратегии развития новых медиа – 2025» Международного форума «Медиа в современном мире» Е. Л. Вартановой поднимался вопрос о необходимости прогнозирования развития современных медиа. Ведь, если верить концепции нового типа общества, технологии напрямую влияют на него. Big data — это тот новый инструмент работы для журналиста, источник информации и эффективной, и интересной визуализации. Генеративный дизайн никогда не станет заменой для информационного, также, как и электронная книга никогда не станет заменой для печатной. Поэтому журналистам и дизайнерам необходимо думать наперед о тех реалиях, которые их ожидают.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Аппаратно-программные средства ГИС: Курс лекций. - Минск: БГУ, 2003. С. 140.
2. Арнхейм Р. Искусство и визуальное восприятие. М., Архитектура–С, 2007. С. 392.
3. Бринтон В. Графическое изображение фактов / Пер. с англ. С. Займовского. М.: Общероссийский общественный фонд «Общественное мнение», 2017. С. 348.
4. Вартанова Е. Л. Теория СМИ: Актуальные вопросы: сб. науч. статей. – М.: МедиаМир, 2009. С. 488.
5. Геоинформатика. Толковый словарь основных терминов / под ред. А.М. Берлянта и А.В. Кошкарева. -М.: ГИС-Ассоциация, 1999. С. 204.
6. Голомбински, К., Хаген, Р. Добавь воздуха! Основы визуального дизайна для графики, веб и мультимедиа / К. Голомбински, Р. Хаген - СПб.: Питер, 2013. С. 275.
7. Горшков М.К., Шереги Ф.Э. Прикладная социология: методология и методы: Учебное пособие/ М.К. Горшков, Ф.Э. Шереги. — М.: Альфа-М: ИНФРА-М, 2009. С. 416.
8. Гражданский кодекс Российской Федерации: в 4 ч.: по состоянию на 1 февр. 2010 г. Москва: Кнорус, 2010. С. 540.
9. Ефремова, Т.Ф. Новый толково-словообразовательный словарь русского языка / под ред. Т. Ф. Ефремовой. Дрофа. Русский язык, 2000.
10. Желязны Д. Говори на языке диаграмм: Пособие по визуальным коммуникациям для руководителей / Пер. с англ. – М.: Институт комплексных стратегических исследований, 2004. С. 220.
11. Захарова, О.А. Интерактивное повествование и мультимедиа в системе профессионального обучения и повышения квалификации /

- О.А. Захарова// Журнал «Мир науки, культуры, образования» - Выпуск№ 1 (38) - 2013.
12. Зиновьев А. Ю. Визуализация многомерных данных. Красноярск: Изд-во КГТУ, 2000. С. 168.
 13. Йоргенсен М. В., Филлипс Л. Дж. Дискурс-анализ. Теория и метод / пер. с англ. — 2-е изд., испр. — Х.: Изд-во «Гуманитарный центр», 2008. С. 352.
 14. Качкаева А.Г. Журналистика и конвергенция: почему и как традиционные СМИ превращаются в мультимедийные / под ред. А.Г. Качкаевой. – М.: 2010. С. 200.
 15. Крам Р. Инфографика. Визуальное представление данных. – СПб.: Питер, 2014.
 16. Лукина М.М. Интернет – СМИ: Теория и практика: Учеб. пособие для студентов вузов / М.М. Лукина. – М.: Аспект Пресс, 2010. С. 348.
 17. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / В. Майер-Шенбергер, К. Кукьер – Манн, Иванов и Фербер, 2014. С. 240.
 18. Маклюэн Г. М. Галактика Гутенберга. Сотворение человека печатной культуры. — Перевод с английского и примечания: А. Юдин. — М., 2003. С 432.
 19. Неганова Л. М., Шевелева Ю. Г., Замедлина Е. А. Статистика. – А-Приор, 2011. С. 144.
 20. Пашинян И. А. Контент — анализ как метод исследования: возможности и ограничения. Журнал Научная периодика: проблемы и решения, Выпуск № 3 (9), май—июнь, 2012. С. 13-18.
 21. Розин В.М. Визуальная культура и восприятие: Как человек видит и понимает мир. — М.: Эдиториал УРСС, 1996. С. 224.
 22. Семёнова А.В., Корсунская М.В. Контент-анализ СМИ: проблемы и опыт применения / Под ред. В.А. Мансурова. – М.: Институт социологии РАН, 2010. С. 324.

23. Сизикова Т. С. Тенденция развития картографического дизайна в геоинформационном обеспечении территории/Т. С. Сизикова, Е. В. Комиссарова // Интерэкспо ГЕО-Сибирь-2012. - Новосибирск: СГГА, 2012. Т.2. С.13-18.
24. Симакова, С.И. Журналистика данных как современное направление журналистики // С.И. Симакова // Научный журнал «Знак»: проблемное поле медиаобразования. – 2013. - №1 (11). С. 52-56.
25. Суслов В.И., Ибрагимов Н.М., Талышева Л.П., Цыплаков А.А. Эконометрия. Часть III Эконометрия — I: Анализ временных рядов — Учебное пособие — Новосибирск: Изд — во СО РАН, 2005. С. 744.
26. Тоффлер Э. Третья волна. – М.: ООО "Фирма "Издательство АСТ", 1999. С. 784.
27. Туманов В.Е. Проектирование хранилищ данных для систем бизнес-аналитики: учебное пособие / В.Е. Туманов. — М.: Интернет-Университет Информационных Технологий: БИНОМ. Лаборатория знаний, 2010. С. 615.
28. Черняк Л. Большие данные – новая теория и практика // Открытые системы. СУБД. 2011. № 10. С. 18–25.
29. Шилина А. Г. Дата журнализм: особенности визуализации проектов. Медиа в современном мире. Молодые исследователи: материалы 16-й международной конференции студентов, магистрантов и аспирантов (15–17 марта 2017 года) / Под ред. А. С. Смоляровой; сост. А. Н. Марченко. — СПб.: С.-Петерб. гос. ун-т, 2017. С. 605.
30. Шилина, М.Г. Текстогенные трансформации инфосферы. Методологический эскиз становления Интернета: Монография. — РИЦ «Северо-Восток» Москва, 2012. С. 736.

31. Яу Н. Искусство визуализации в бизнесе. Как представить сложную информацию простыми способами / Нейта Яу; пер. с англ. Светланы Кировой. – М.: Манн, Иванов и Фербер, 2013. С. 352.
32. Blundell W. E. The Art & Craft of Feature Writing. Plume, 1988. P. 259
33. Bordwell D., Thompson K. Film Art: An Introduction. McGraw-Hill, 2003. P. 544.
34. Chernoff H. Journal of the American Statistical Association, No. 68, 1973. P. 361-368.
35. Edward Segel and Jeffrey Heer. Narrative visualization: Telling stories with data. IEEE Transactions on Visualization and Computer Graphics, 16 (March), 2010. P. 1139–1148.
36. Jacobs M., Gamification: Moving from 'Addition' to 'Creation', 2013.
37. Khan M., Khan S.S. Data and Information Visualization Methods and Interactive Mechanisms: A Survey, International Journal of Computer Applications, 34(1), 2011. P. 1-14.
38. Lankow J. Infographics: The Power of Visual Storytelling / J. Lankow, J. Ritchie, R. Crooks — Wiley, 2012. P. 264.
39. Lee J.J., Hammer, J. Gamification in Education: What, How, Why Bother? Academic Exchange Quarterly, 2011.
40. Michele D. D. Game Design Narrative for Learning: Appropriating Adventure Game Design Narrative Devices and Techniques for the Design of Interactive Learning Environments. ETR&D, 54(3), 2006. P. 245.
41. Robert L. Harris. Information Graphics: A Comprehensive Illustrated Reference. Publisher: Oxford University Press; 1 edition, 2000.
42. Tufte E.R. Envisioning Information. – Cheshire, CT: Graphics Press, 1990. P. 126.
43. Tufte E.R. Envisioning Information. – Cheshire, CT: Graphics Press, 1990. – 126 p.

44. Tufte E.R. The Visual Display of Quantitative Information. 2nd edition. – Cheshire, CT: Graphics Press, 2001. – 199 p.
45. Tufte E.R. Visual Explanations: Images and Quantities, Evidence and Narrative. – Cheshire, CT: Graphics Press, 1997. – 157 p
46. Werbach K., Hunter D. For the Win: How game thinking can revolutionize your business. Wharton Digital Press, 2012.
47. Zichermann G., Cunningham, C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. Sebastopol, Ca: O'Reilly, 2011.
48. Zichermann, G. Gamification by Design. O'Reilly Media, 2011. P. 208.
49. Вартанова, Е.Л. Медиэкономика зарубежных стран. Учебное пособие [Электронный ресурс] / Е.Л. Вартанова - М.: Аспект Пресс, 2003. Точка доступа: <http://evartist.narod.ru/text11/42.htm> (дата обращения: 14.05.17)
50. Васильев В. Р., Волобой А. Г., Вьюкова Н. И., Галактионов В. А. Контекстная визуализация пространственных данных [Электронный ресурс] // Препринты ИПМ им. М. В. Келдыша. 2004. № 56. 23 с. Точка доступа: http://www.keldysh.ru/pages/cgraph/articles/dep20/vis_pp.pdf (дата обращения: 14.05.17)
51. Гражданский кодекс Российской Федерации (ГК РФ) // Информационно-правовой портал «Гарант» [Электронный ресурс]. Точка доступа: <http://base.garant.ru/10164072/> (дата обращения: 14.05.17)
52. Грэй Дж., Чемберс Л. Пособие по журналистике данных [Электронный ресурс] / Грэй Дж., Чемберс Л. // Сетевое издание «РИА Новости», 2013. Точка доступа: http://ria.ru/files/book/_site/index.html (дата обращения: 14.05.17)

53. Журнал Инфографика // Infogra.ru Точка доступа: <http://infogra.ru/infographicsmag> (дата обращения: 14.05.17)
54. Зачем и как использовать визуализацию данных? [Электронный ресурс] / П. Шиманский // Хабрахабр, 2014. Точка доступа: <https://habrahabr.ru/company/devexpress/blog/240325/>
<http://infogra.ru/infographicsmag> (дата обращения: 14.05.17)
55. Калмыков, А. А. Интерактивная гипертекстовая журналистика в системе отечественных СМИ. Научное издание / [Электронный ресурс] / А.А. Калмыков, под ред. В.С. Хелемендика. - М : Издательство ИПК работников ТВ и РВ, 2009. Точка доступа: <http://www.eartist.narod.ru/text19/159.html> (дата обращения: 14.05.17)
56. Клесова, М. За какие профессии в ближайшее время будут хорошо платить [Электронный ресурс] // Интернет-издание «Рустория» – 2014. Точка доступа: <http://sandbox.rustoria.ru/post/za-kakie-professii-v-blizhajshee-vremya-budut-mnogo-platit/> (дата обращения: 14.05.17)
57. Мастицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R, 2014 [Электронный ресурс] Точка доступа: <http://r-analytics.blogspot.com> (дата обращения: 14.05.17)
58. Медиаатренды [Электронный ресурс]: экспертный взгляд факультета журналистики МГУ имени М. В. Ломоносова на события в СМИ / – Электрон. журн. - факультет журналистики МГУ, 2011. Точка доступа: [http://www.mediascope.ru/files/MediaTrends %2019.pdf](http://www.mediascope.ru/files/MediaTrends_%2019.pdf). (дата обращения: 14.05.17)
59. Научный понедельник: как развивается «журналистика данных» [Электронный ресурс] // Сетевое издание «РИА Новости» (канал на Youtube.com), 2013. Точка доступа: <http://www.youtube.com/watch?v=-jCrE9qT5hY> (дата обращения: 14.05.17)

60. Николов, А.Л. Новые направления в медиаиндустрии, или что такое дата-журналистика (data journalism) [Электронный ресурс] / А.Л. Николов // Международный электронный научно-образовательный журнал «Медиа. Информация. Коммуникация», 2012. Точка доступа: <http://mic.org.ru/index.php/new/165-novye-napravleniya-v-mediaindustrii-ili-chto-takoe-data-zhurnalistika-data-journalism> (дата обращения: 14.05.17)
61. Рачинский, С. Журналистика баз данных: модный тренд или будущее профессии? [Электронный ресурс] / С. Рачинский // Блог журналиста Сергея Рачинского «Институт новых медиа (Киев)», 2011. Точка доступа: <http://newmedia.in.ua/articles/business/zhurnalistika-baz-dannyx-modnyj-trend-ili-budushhee-professii.html> (дата обращения: 14.05.17)
62. Социология [Электронный ресурс]: энциклопедия. - Минск: Книжный Дом, 2003. Точка доступа: <http://voluntary.ru/dictionary/568/page/65> (дата обращения: 14.05.17)
63. Тепловая карта как способ презентации данных исследований [Электронный ресурс]. Точка доступа: <http://edwvb.blogspot.ru/2013/10/teplovaja-karta-kak-sposob-prezentacii-dannykh-issledovaniij.html> (дата обращения: 14.05.17)
64. Тертычный, А.А. Жанры периодической печати [Электронный ресурс] / Учебное пособие. М.: Аспект Пресс, 2000. Режим доступа: <http://www.eartist.narod.ru/text2/01.htm> (дата обращения: 14.05.17)
65. Шевченко, В. Э. Теоретические основы визуальной коммуникации [Электронный ресурс] // Научные ведомости БелГУ. Серия: Гуманитарные науки, 2013. Точка доступа: <http://cyberleninka.ru/article/n/teoreticheskie-osnovy-vizualnoy-kommunikatsii> (дата обращения: 14.05.17)
66. Шевченко, В. Э. Теоретические основы визуальной коммуникации: тенденции технологий и форм передачи информации [Электронный

- ресурс] // Вестник Челябинского государственного университета., 2015. Режим доступа: <http://cyberleninka.ru/article/n/vizualnye-kommunikatsii-tendentsii-form-i-tehnologiy-peredachi-informatsii> (дата обращения: 14.05.17)
67. Шерстюкова, М.Н. Дата-журналистика как новое направление в системе средств массовой коммуникации [Электронный ресурс] / М.Н. Шерстюкова // Международный электронный научно-образовательный журнал «Медиа. Информация. Коммуникация». – 2012. Точка доступа: <http://mic.org.ru/1-nomer-2012/65-1-shestukova-2> (дата обращения: 14.05.17)
68. Шилина, М. Г. Медиа XXI века как объект исследования (Несколько тезисов об инновационных подходах к изучению массовых коммуникаций) [Электронный ресурс]. Точка доступа: http://www.mediaalmanah.ru/files/59/2013_6_7_shilina.pdf (дата обращения: 14.05.17)
69. Шилина, М. Г., Левченко, В. Ю. Big Data, Open Data, Linked Data, метаданные в PR: актуальные модели трансформации теории и практики. [Электронный ресурс] // Журнал Факультета журналистики МГУ имени М.В. Ломоносова «Медиаскоп», 2014. Режим доступа: <http://www.mediascope.ru/node/1486> (дата обращения: 14.05.17)
70. Шилина, М.Г. Data Journalism – дата-журналистика, журналистика метаданных – в структуре медиакоммуникации: к вопросу формирования теоретических исследовательских подходов [Электронный ресурс] / М.Г.Шилина // Журнал Факультета журналистики МГУ имени М.В. Ломоносова «Медиаскоп». – 2013. - №1. Точка доступа: <http://www.mediascope.ru/node/1263> (дата обращения: 14.05.17)
71. Шилина, М.Г. Гипертекст интернета как формат интеграции коммуникации: дата-журналистика [Электронный ресурс] // Вестник

- ТвГУ. Серия: Филология (4). С. 192-199. Режим доступа: http://tjournal8.net/journals_n/1403272055.pdf (дата обращения: 14.05.17)
72. Шиманский, П. Интерактивная инфографика с анимациями CSS и SVG [Электронный ресурс] / П. Шиманский // Хабрахабр, 2013. Точка доступа: <http://habrahabr.ru/post/168681/> (дата обращения: 14.05.17)
73. Электронный словарь Gartner IT Glossary [Электронный ресурс] // руководство к изучению терминов и определений из области информационных технологий, 2015. Точка доступа: <http://www.gartner.com/technology/home.jsp> (дата обращения: 14.05.17)
74. A classification of chart types [Электронный ресурс]. Точка доступа: <http://excelcharts.com/classification-chart-types/> (дата обращения: 15.05.17)
75. Deterding S., Khaled R., Nacke L., Dixon D. Gamification: Toward a Definition [Электронный ресурс]. Точка доступа: <http://gamification-research.org/wp-content/uploads/2011/04/02> (дата обращения: 17.05.17)
76. Oxford Dictionaries - Dictionary, Thesaurus, & Grammar [Электронный ресурс]. Режим доступа: https://en.oxforddictionaries.com/definition/big_data (дата обращения: 14.05.17)
77. Press, G. 12 Big Data Definitions: What's Yours? [Электронный ресурс] / G. Press// Forbes, 2014. Режим доступа: <http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/> (дата обращения: 14.05.17)
78. The Extreme Presentation(tm) Method [Электронный ресурс]. Точка доступа: http://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html (дата обращения: 15.05.17)

79. Visual vocabulary [Электронный ресурс]. Точка доступа: <https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf> (дата обращения: 15.05.17)
80. Wu M. What is gamification, really? [Электронный ресурс]. Точка доступа: <http://lithosphere.lithium.com/t5/Science-of-Socialblog/What-is-Gamification-Really/ba-p/30447> (дата обращения: 15.05.17)

ПРИЛОЖЕНИЯ

Приложение 1. Контент-анализа выборки визуализации проектов на основе big data изданий The Guardian, The New York Times, The Washington Post

Тематика издания:

- 1 Национальные новости
- 2 Мировые новости
- 3 Технологии
- 4 Бизнес
- 5 Экономика
- 6 Политика
- 7 Спорт
- 8 Наука
- 9 Здоровье

Способ визуализации:

- 10 Упаковка кругов
- 11 Солнечная диаграмма
- 12 Пузырьковая диаграмма
- 13 Матричная диаграмма
- 14 Карта
- 15 Линейный график
- 16 График «открыто-высокого-низкого-закрытия»
- 17 Хордовая диаграмма
- 18 Поточковый график

19 Диаграмма параллельных

- координат
- 20 Сложная столбчатая диаграмма
- 21 Диаграмма рассеяния
- 22 Интерактивная инфографика
- 23 Пиктограмма
- 24 Параллельные наборы

Тип информации:

- 25 Дискретные данные
- 26 Географические данные
- 27 Непрерывные временные данные

Интерактивность:

- 0 – отсутствует
- 1 - присутствует

	Название проекта	Издание	Дата публикации	Тема	Способ визуализации	Тип информации	Количество типов графиков	Интерактивность	Всего графиков
1	Mapping the Spread of Drought Across the U.S.	The New York Times	09.04. 2015	1	14	27, 28	3	1	3
2	How Has the Drought Affected California's Water Use?	The New York Times	02.06. 2015	1	14	27, 28	1	1	3
3	Mapping the Shadows of New York City: Every Building, Every Block	The New York Times	21.12.2016	1	14	27, 28	2	1	3
4	How Much Warmer Was Your City in 2015?	The New York Times	19.02.2016	1	15	28	2	1	2
5	Kepler's Tally of Planets	The New York Times	23.07.2015	8	23	28	1	1	1
6	The Race Gap in America's Police Departments	The New York Times	08.04.2015	1	20	26	2	1	116

7	How Fan Loyalty Changed During the World Cup	The New York Times	07.12.2014	7	19	26	1	1	33
8	The Most Detailed maps You'll See From the Midterm Elections	The New York Times	06.11.2014	6	14	26	1	1	7
9	Why Peyton Manning's Record Will Be Hard to Beat	The New York Times	19.10.2014	7	15	28	1	1	2
10	Home Prices in 20 Cities	The New York Times	26.08.2014	4	15	28	2	1	2
11	Where We Came From and Where We Went, State by State	The New York Times	13.08.2014	1	18	28	1	1	2
12	How Birth Year Influences Political Views	The New York Times	07.07.2014	6	15	28	1	1	1
13	Clubs that connect World Cup	The New York Times	20.06.2014	7	17	26	1	1	1
14	How the Recession Reshaped the Economy, in 255 Charts	The New York Times	06.06.2014	5	15	28	2	1	255
15	Can You Live on the Minimum Wage?	The New York Times	08.02.2014	1	23	28	1	1	1
16	State Gun Laws Enacted in the Year After Newtown	The New York Times	10.12.2013	1	16	28	1	1	109
17	In Climbing Income Ladder, Location Matters	The New York Times	22.07.2013	4	14	28, 27	3	1	3
18	How the N.R.A. Rates Lawmakers	The New York Times	19.12.2012	6	14	27, 28	2	1	2
19	How the Tax Burden Has Changed	The New York Times	30.11.2012	1	15	28	1	1	7
20	Over the Decades, How States Have Shifted	The New York Times	07.11.2012	6	15	28	1	1	1
21	What Percent Are You?	The New York Times	14.01.2012	4	14	27	1	1	1
22	The death toll in the Gaza-Israel conflict	The Washington Post	7.08.2014	2	24	26	1	1	1

23	Fallen from the skies	The Washington Post	20.06.2014	1	25	26	3	1	3
24	Interactive: What is the safest time to drive?	The Guardian	27.10.2013	2	23	28	2	1	2
25	A global look at cardiac risk factors	The Washington Post	25.06.2011	9	22	28	1	1	1
26	In flight: see the planes in the sky right now	The Guardian	22.01.2014	2	14	27	1	1	1
27	Ebola spreads slower, kills more than other diseases	The Washington Post	14.10.2014	9	13	26	1	1	1
28	Rising death rates for white women	The Washington Post	09.04.2016	1	15	28	2	1	5
29	The electoral college misrepresents every state, but not as much as you may think	The Washington Post	06.12.2016	6	20	28	2	0	2
30	Six 14s that show the anatomy of America's vast infrastructure	The Washington Post	01.12.2016	1	14	27	1	0	6
31	The increasingly diverse United States of America	The Washington Post	25.11.2016	1	14	28	1	0	
32	How Trump redrew the electoral map, from sea to shining sea	The Washington Post	09.11.2016	6	14	26	2	0	4
33	A deadly crisis: mapping the spread of America's drug overdose epidemic	The Guardian	12.05.2016	1	14	27, 26	1	1	1
34	NFL draft pick quality for your team	The Washington Post	21.04.2016	7	14	26	7	1	7
35	The dark side of Guardian comments	The Guardian	12.04.2016	3	15	28	1	1	6
36	History through the president's words	The Washington Post	12.01.2016	6	12	26	1	1	7
37	The number of 'mass shootings' in the U.S. depends on how you count	The Washington Post	04.12.2015	1	24	26	1	1	354

38	Mapping how the United States generates its electricity	The Washington Post	31.07.2015	2	14	28		1	2
39	The counted. People killed by police in the USE	The Guardian	01.06.2015	1	14	26	3	1	3
40	Where carbon emissions are greatest	The Washington Post	31.03.2015	2	14	28	2	1	2
41	Watch how the measles outbreak spreads when kids get vaccinated – and when they don't	The Guardian	05.02.2015	1	13	26	1	1	10
42	Too much of a good thing	The Washington Post	26.11.2014	9	15	28	2	1	3
43	An eye for an eye?	The Washington Post	24.07.2014	1	24	26	2	0	5
44	Across the nation, income rise hints at recovery	The Washington Post	18.07.2014	1	15	28	1	1	1
45	Washington: A world apart	The Washington Post	09.11.2013	1	14	27, 28	2	1	650
46	Mapping where English is not the language at home	The Washington Post	20.08.2013	6	14	27, 28	1	1	1
47	How long will we live — and how well?	The Washington Post	13.12.2012	9	22	28	1	1	1
48	Homicides in the District	The Washington Post	19.10.2012	1	14	27, 26	2	1	2
49	Building the U.S. tax code, break by break	The Washington Post	18.09.2012	6	15	28	2	1	12
50	Are you over the hill for Olympic sports?	The Washington Post	30.07.2012	7	20	26	2	0	2
51	Gay rights in the US, state by state	The Guardian	08.05.2012	2	11	26	1	1	43
52	How riot rumours spread on Twitter	The Guardian	07.12.2011	1	10	26	1	1	7
53	A gadget's life: From gee-whiz to junk	The Washington Post	10.01.2011	4	12	28	1	1	5