

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Методическое пособие

Инфографика и визуализация данных

Екатерина Круглова

Экономический факультет имени М.В. Ломоносова

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Методическое пособие

Инфографика и визуализация данных

Зачем нужна визуализация данных? Правильный визуальный образ – это эффективная форма донесения Ваших идей. Хорошо продуманный и правильно подготовленный график, диаграмма или схема помогут донести идею до слушателя быстрее и точнее, чем текст или таблица с данными.

Инфографика – это использование визуальных средств для изображения того, что мы не можем увидеть или понять из информации, представленной в другой форме [1].

Визуализировать информацию нужно, чтобы:

- 1) видеть важные паттерны и взаимосвязи;
- 2) информация приобрела больше смысла;
- 3) поведать историю;
- 4) позволить читателю сфокусироваться на том, что реально важно [2].

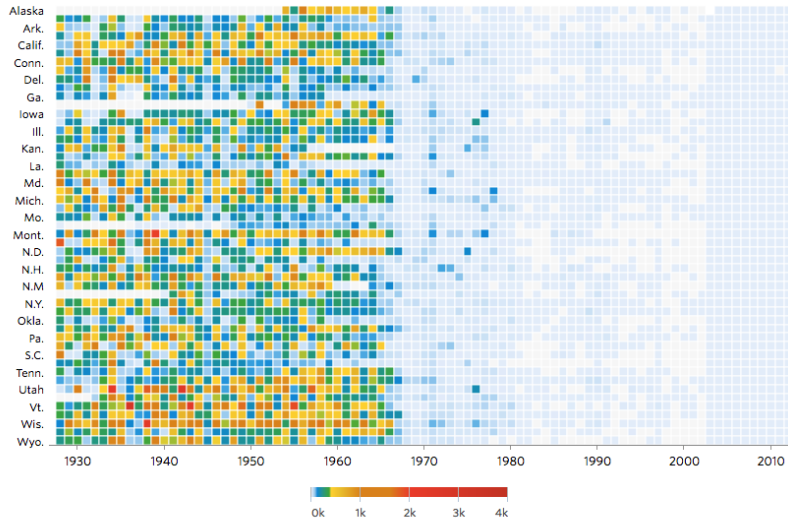
В чем разница между визуализацией и инфографикой? Нарисованные в графическом виде данные – это визуализация, а наличие интерпретации этих данных делает визуализацию инфографикой [1].

Рассмотрим различие на простом примере. На Рисунке1 представлена визуализация данных о заболеваемости корью в разных штатах США с 1928 по 2012 год. На Рисунке 2 уже инфографика на тех же данных от Wall Street Jordan.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 1. Заболеваемость корью в Американских штатах с 1928 по 2012 год (визуализация)

Measles

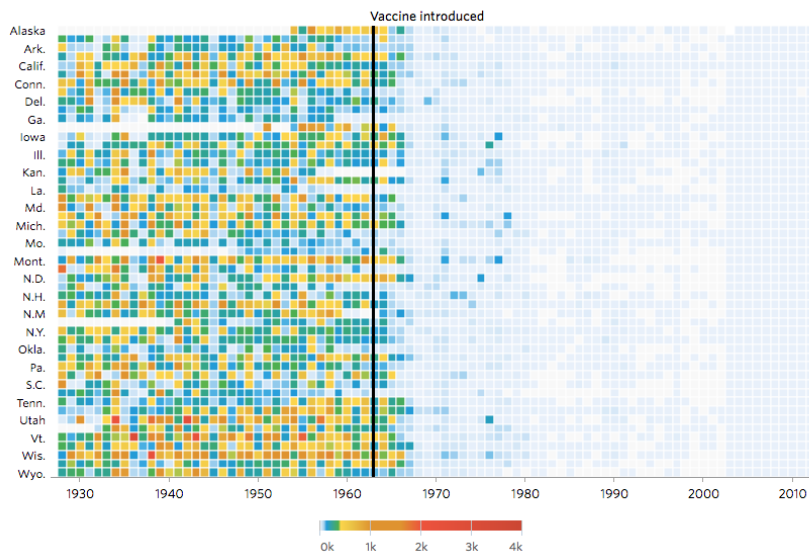


Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

Источник: сделано автором на основе материалов Wall Street Journal [4]

Рисунок 2. Внедрение вакцины позволило победить корь в США (инфографика)

Measles



Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Источник: [4]

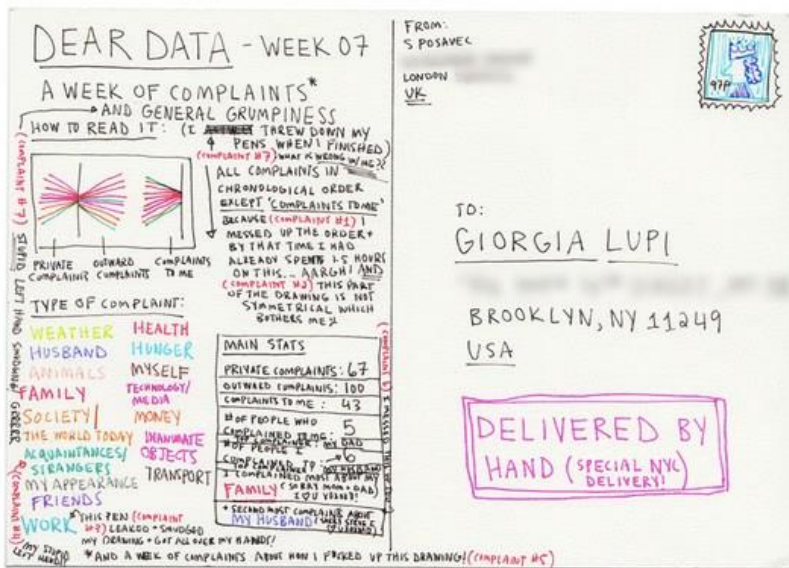
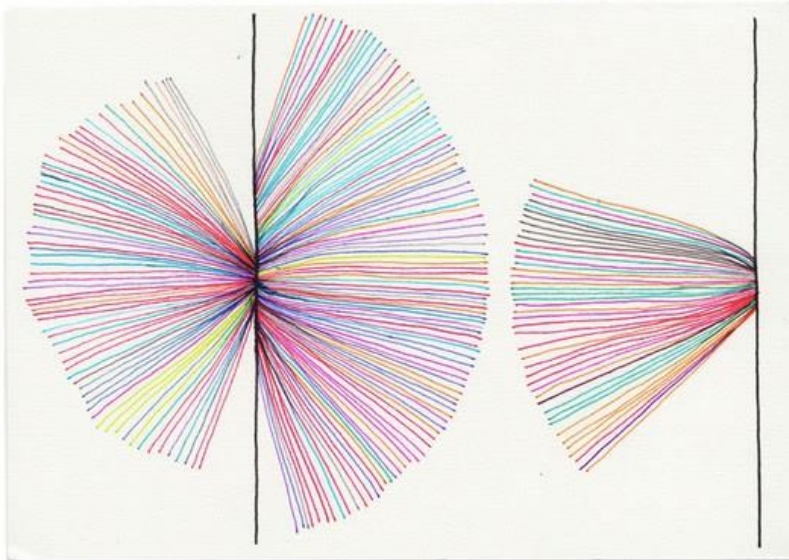
На Рисунке 1 в формате тепловой карты изображено количество случаев заболеваний корью в тысячах на 100 000 населения в каждом штате США. Единственный вывод, который можно сделать на основании представленных данных: заболеваемость корью снижается со временем. Рисунок 2 – это инфографика к статье «Борьба с инфекционными заболеваниями в 20-м веке: влияние вакцин», опубликованной в Walt Street Journal [4]. Вертикальная линия на графике обозначает год, когда была введена в медицинскую практику вакцина против кори. Эта линия позволяет читателю интерпретировать данные и сделать на основе этой инфографики вывод: введение вакцины за несколько лет значительно снизило заболеваемость корью в США. Кстати, эта работа удостоилась престижной премии Data Journalism Awards 2015 в номинации «Визуализация данных года (Крупное новостное издание)».

Таким образом, визуализация данных — это инструмент анализа, она помогает делать выводы. Инфографика — это готовые выводы, оформленные и структурированные показатели и закономерности [3].

Необходимо отметить еще одно направления работы с данными и их визуализацией – это Data art или искусство, основанное на данных. В этом направлении данные используются в качестве основы для создания визуализации, в которой эстетическая составляющая важнее, чем содержательная. Очень известный пример в этой области – проект [Dear Data](#), в рамках которого два дизайнера в области информации, Джоргия Лупи из Нью-Йорка (Giorgia Lupi) и Стефани Позавек из Лондона (Stefania Posavec), посылали друг другу по почте открытки с инфографикой, описывающей, как они провели эту неделю.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 3. Открытка «14 неделя» в рамках проекта Dear data от Стефани Позавек для Джоргии Лупе



Источник: [5]

Цель визуализации: сделать данные наглядными для читателя. Отсюда задачи визуализации связаны непосредственно с работой с данными: показать, сравнить, организовать или связать их [1].

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

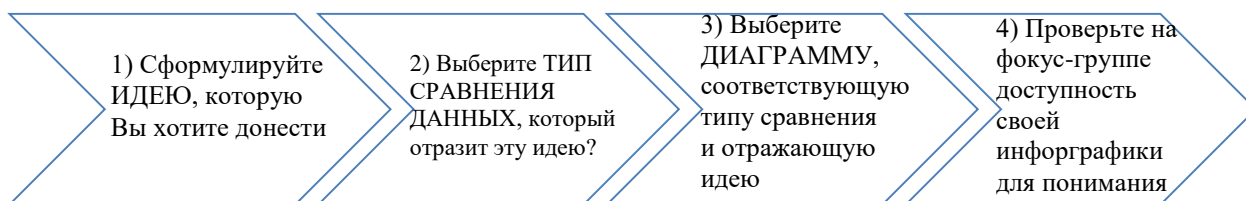
Цель инфографики: корректный вывод, который читатель делает из представленных данных. Если формулировать цель в терминах действия, то это показать то, что должно **стать очевидно** для читателя. Задачи инфографики связаны с читателем: привлечь его внимание, помочь ему понять или побудить его к чему-то [1].

Таблица 1. Цели и задачи инфографики и визуализации данных

	Визуализация	Инфографика
Цель	Сделать данные наглядными для читателя	Сделать идею очевидной для читателя, показать ему вывод
Задачи	Связаны с данными: 1) показать 2) сравнить 3) организовать 4) связать	Связаны с читателем: 1) привлечь внимание 2) помочь понять 3) побудить к чему-то

Источник: составлено автором на основе материалов [1].

Базовые шаги визуализации и создания инфографики



- 1) Прежде чем приступить к созданию инфографики, задайте себе вопрос: что должно стать очевидным читателю? Сформулируйте, что Вы хотите сказать, в чем основная идея, которую Вы хотите донести. Пусть основная мысль станет заголовком Вашего графика.
- 2) Определите, как именно Вы хотите сравнивать данные. Тип сравнения данных зависит от того, какую идею Вы хотите донести.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Существует 5 способов сравнения данных: покомпонентное, позиционное, временное, частотное и корреляционное [6].

3) Выберите тип диаграммы, который наилучшим образом позволит Вам отразить Вашу идею и сравнить Ваши данные. Тип диаграммы определяют не имеющиеся у Вас данные, а ИДЕЯ, которую Вы хотите через визуализацию донести до читателя.

4) Проверьте, насколько Ваша инфографика понятна другим людям. Разные люди по-разному воспринимают визуальные образы. Поэтому хорошо будет протестировать свои диаграммы на коллегах, чтобы удостовериться, что ваши материалы передают именно ту идею, которую Вы в них закладывали, и Ваши коллеги их хорошо понимают.

Рассмотрим подробнее типы сравнения данных [6]. К каждому типу сравнения будут предложены базовые типы диаграмм, позволяющие визуализировать это сравнение. Существует 5 основных видов диаграмм: линейный график, точечная диаграмма, гистограмма, линейчатая диаграмма и круговая диаграмма. Подробнее об этих и многих других видах диаграмм речь пойдет в другом разделе.

1. Покомпонентное сравнение

Данный тип сравнения показывает, как часть относится к целому. Другими словами, **процент от целого**.

Ключевые слова этого вида сравнения: доля, проценты.

Характерные фразы: За первый квартал на магазин А пришла наибольшая доля наших продаж. Почти половину рабочего времени наши сотрудники проводят на совещаниях.

Расходы на здравоохранение составляют X% бюджета на следующий год.

Подходящие диаграммы для визуализации: круговая диаграмма.

2. Позиционное сравнение

В случае позиционного сравнения объекты сравниваются между собой. Сравнивается **соотношение объектов**.

Ключевые слова: больше чем, меньше чем, равно, занимает X место.

Примеры: Военные расходы США в абсолютном выражении превышают расходы Российской Федерации на ВПК. МГУ занимает первое место в мире среди Высших учебных заведений по

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

числу запущенных спутников. Продолжительность жизни в России равна продолжительности жизни в Белизе, но меньше, чем в Бангладеш.

Подходящие диаграммы для визуализации: линейчатая диаграмма, гистограмма (больше подходит для временного сравнения)

3. Временное сравнение

Данное сравнение показывает нам, как **меняется показатель во времени**. Мы привыкли рассматривать время, как движение слева направо, а не сверху вниз или как-то еще. Поэтому слева располагаются более старые данные, а справа – более новые.

Ключевые слова: изменяется, растет, снижается, падает, колеблется, остается неизменным.

Примеры: Госдолг США постоянно растет. Продолжительность жизни в России увеличилась по сравнению с прошлым годом.

Подходящие диаграммы для визуализации: гистограмма, линейный график, линейный график с областями. Гистограмма хороша для небольшого количества наблюдений (до 7), а линейный график – если наблюдений много.

4. Частотное сравнение

Частотное сравнение характеризует структуру данных в количественном выражении: сколько объектов наблюдения соответствуют заданным критериям, какое **число объектов находится в интервале**. Для большей ясности приведем пример: частотное сравнение показывает, сколько студентов получили 5-ки за экзамен по математическому анализу, сколько 4-ки, сколько 3-ки и сколько не сдали экзамен. В результате основной идеей этого частотного сравнения будет то, что большинство студентов в этом году получили по математическому анализу 4-ки.

Этот тип сравнения актуален в двух ситуациях. Первая: чтобы обобщить сходные события на основе выборки наблюдений. Это особенно актуально для анализа риска, вероятности, возможностей. Подходит для визуализации азартных игр и других вероятностных событий. Вторая ситуация – это суммирование больших объемов информации, чтобы продемонстрировать значимые зависимости. Например, демографические данные. Этот тип сравнения часто используется при переписи населения, при анализе результатов голосования

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

и т.п. В данном случае есть два параметра сравнения, которые в графиках превращаются в две оси. Один параметр – это количество или доля значений/объектов, попадающих в интервал, второй – шкала распределения, те самые интервалы, между которыми распределяются значения. Оптимальное количество интервалов для анализа и визуализации: от 5 до 20. Лучше использовать интервалы одинакового размера. Однако возможны случаи, когда интервалы могут быть неравномерные. Это подходит для данных со смещенным распределением, например, если речь идет о распределении доходов среди населения.

Ключевые слова: в диапазоне от X до Y, большая часть соответствует значению X, распределение.

Примеры: Больше всего студентов отчисляется по результатам сессии после 1 курса.

Подходящие диаграммы для визуализации: гистограмма (если 5-7 интервалов значений) и линейный график (если больше, а также если не дискретные интервалы, а непрерывные значения).

5. Кореляционное сравнение

Данный тип сравнения показывает наличие или отсутствие связи между показателями. Есть ли **зависимость между переменными**.

Ключевые слова: связан с, возрастает при, снижается если, прямо пропорционален, имеет обратную зависимость от.

Примеры: Успешный результат экзамена прямо пропорционален количеству часов, затраченным на подготовку к нему.

Подходящие диаграммы для визуализации: точечная диаграмма (рассеивания), но также возможно использование двусторонней линейчатой диаграммы распределения.

Таблица 2. Типы сравнения и соответствующие им диаграммы

	Покомпонентное	Позиционное	Временное	Частотное	Корреляционное
Круговая					
Линейчатая диаграмма					
Гистограмма					
Линейный график					

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Точечная диаграмма					
-----------------------	--	--	--	--	--

Источник: [6]

Выше речь шла о содержательных способах сравнения данных. Рассмотрим теперь основные способы визуального сравнения данных [1]. Для количественного (точная разница) и качественного (больше/меньше, лучше/хуже, одинаковое/разное) существуют разные визуальные приемы.

Количественное сравнение данных:

1. Координаты

Сравниваемые объекты обладают числовыми характеристиками (одной или большим числом), которые наносятся на ось (оси) координат и на этом основании сравниваются между собой. Наиболее распространённый вариант – это сравнение по двум координатам в виде линейного графика или точечной диаграммы. Однако еще три измерения можно добавить с помощью площади, цвета (качественное измерение) и оттенка точки, о которых речь пойдет дальше. Нужно помнить, что чем больше параметров сравнивается одновременно, тем менее точным становится сопоставление для читателя, так как удерживать все параметры одновременно в сознании очень сложно.

2. Длина/Ширина

Показатели длины и ширины также используются для сравнения количественных характеристик объектов. Наиболее характерный пример использования этого способа: линейчатая диаграмма и гистограмма. При этом длина и ширина могут использоваться как по отдельности, так и вместе. Однако при совместном использовании этих параметров снижается точность сравнения.

3. Площадь

Еще один метод сравнения – площадной. Количественный параметр переводится в площадь фигуры, а затем площади сопоставляются между собой. К площадным сравнениям относятся такие типы диаграмм, как круговая диаграмма и трипам. Важно учитывать, что чем сложнее форма объекта, площади которых планируется сравнивать, тем ниже точность сравнения.

4. Угол

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

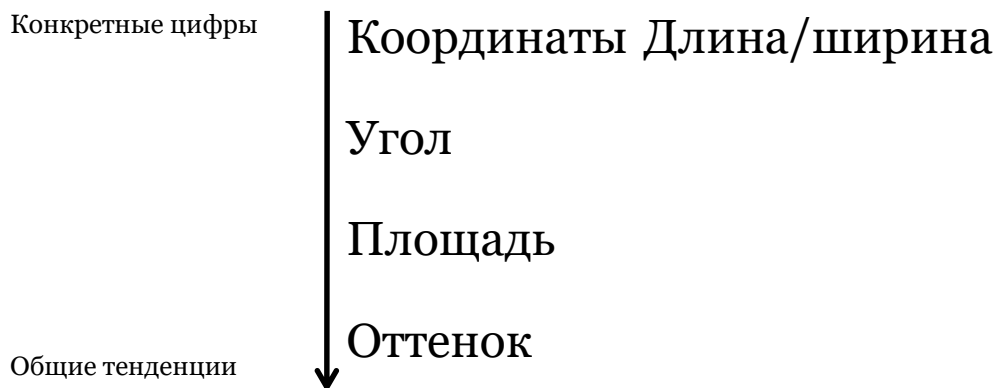
Угол наклона прямой также может быть использован для количественного сравнения. Подробнее об этом методе будет рассказано, когда речь пойдет об угловых диаграммах Эдварда Тафти.

5. Отенок

Сравнение количественных параметров с использованием оттенков применяется в тепловых картах. Разным значениям количественного параметра присваивается свой оттенок. Этот тип сравнения наименее точно отражает конкретные количественные значения.

Рисунок 4. Зависимость точности сравнения данных от способа визуализации

Точность сравнения снижается



Источник: [1]

Качественное сравнение данных:

1. Форма

Для обозначения различия качественных характеристик сравниваемых объектов могут использоваться формы или фигуры.

2. Цвет

Различие объектов может быть графически отражено с помощью цветового кодирования. Например, данные о рождаемости мальчиков обозначаются голубым, а девочек розовым и т.п.

Материалы к [вебинару](#) 22.01.2018 [«Инфографика и визуализация данных»](#)

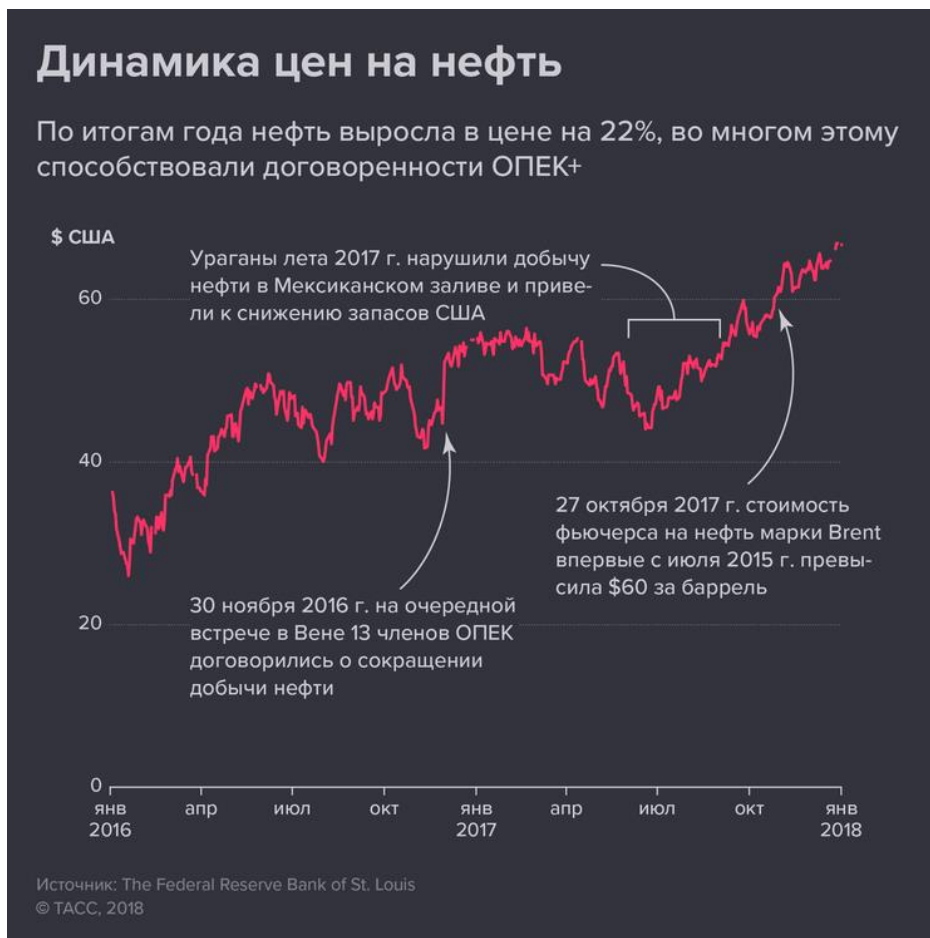
Базовые виды графиков

В классической литературе, посвященной визуальной коммуникации (например, «Говори на языке диаграмм» [6]) обычно выделяют 5 основных типов диаграмм, позволяющих визуализировать количественные данные: линейный график, линейчатая диаграмма, гистограмма, точечная диаграмма и круговая диаграмма. В обзоре ниже мы рассмотрим больше вариантов.

1) **Линейный график** показывает динамику показателей. Это самый простой и распространенный график. Подходит для временного и частотного сравнений. Может быть классического вида (как на Рисунке 5), плавной линией или ступенчатым. Ступенчатый график используется в случаях, когда показатель меняется резко и не регулярно (например, показатель кадрового обеспечения или цены на бензин в Белоруссии). Это единственный тип графика, у которого шкала может начинаться не от нуля. Однако необходимо указать значение, с которого начинается шкала в таком случае. Если Вы показываете временное сравнение и в нем есть прогнозные значения, выделяйте их пунктиром или другим цветом.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 5. Динамика цен на нефть

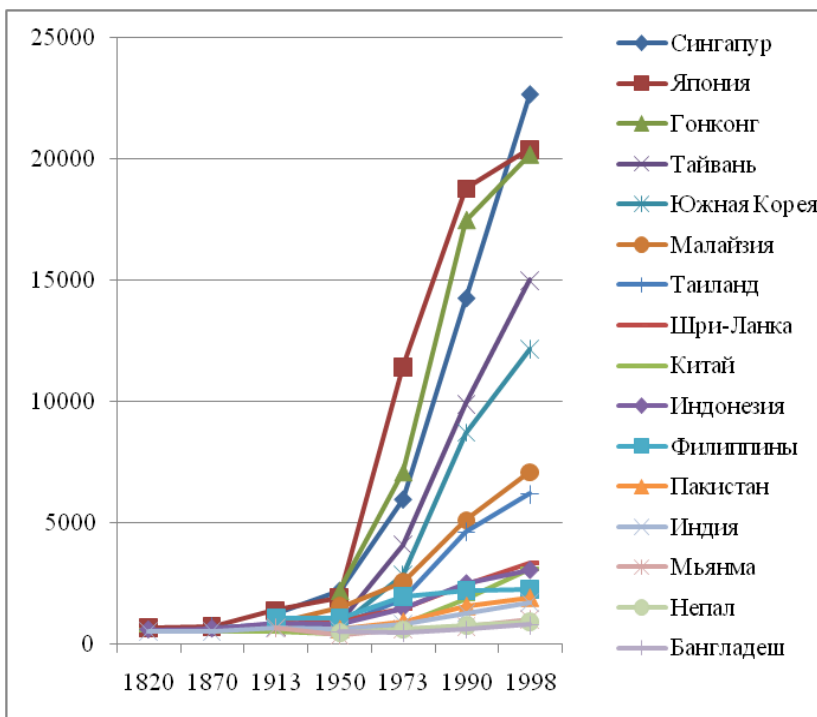


Источник: [7]

Не стоит располагать на линейном графике много наборов данных (линий), иначе получится «шум». Так можно делать только в ситуации, когда Вы хотите показать общую тенденцию. Спорный пример использования большого количества линий на одном линейном графике (Джин Желязны называет эту ситуацию «спагетти» [6]) приведен на Рисунке 6. Из этого графика можно сделать вывод о том, что страны имеют разные траектории развития. Однако траекторию конкретной страны оценить сложно. Если Вам необходимо показать много наборов данных, то, возможно, лучше разбить их на разные графики. Итоговое решение зависит от Вашей цели.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 6. Динамика ВВП на душу населения (долл.) стран Востока, перешедших на траекторию А, а также стран, находящихся на более низкой траектории Б



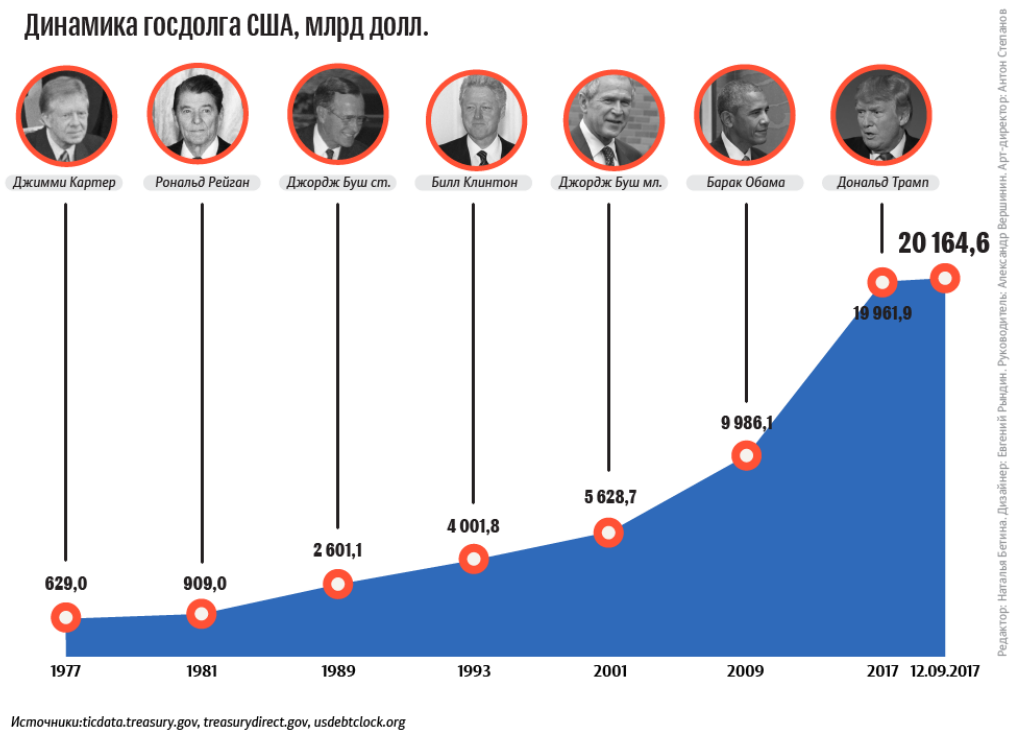
Источник данных: Maddison A. *The World Economy: A Millennial Perspective*. OECD, 2001

Источник: [8]

2) **Линейный график с областями** используется, когда важно показать общее суммарное значение. Детали здесь менее важны. Если это график с накоплением, то понятно только соотношение компонентов в целом, а оценить точный вклад каждого слоя сложно. Нельзя показывать слишком много слоев, не более 5. Если данных больше, выделите 4-5 основных, а остальные сгруппируйте в категорию «другое». Самый важный слой следует располагать непосредственно над осью абсцисс. Это связано с тем, что только нижняя категория имеет фиксированное основание. Это дает возможность точно определить объем нижней категории. Остальные категории не имеют фиксированных оснований, их объем может быть определен только приблизительно. Если нужно точно определить значение каждой категории, то рекомендуется привести отдельно линейный график с областями, отражающий суммарное значение, и отдельно графики каждой из входящих в сумму категорий.

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Рисунок 7. Динамика госдолга США, млрд. долл.

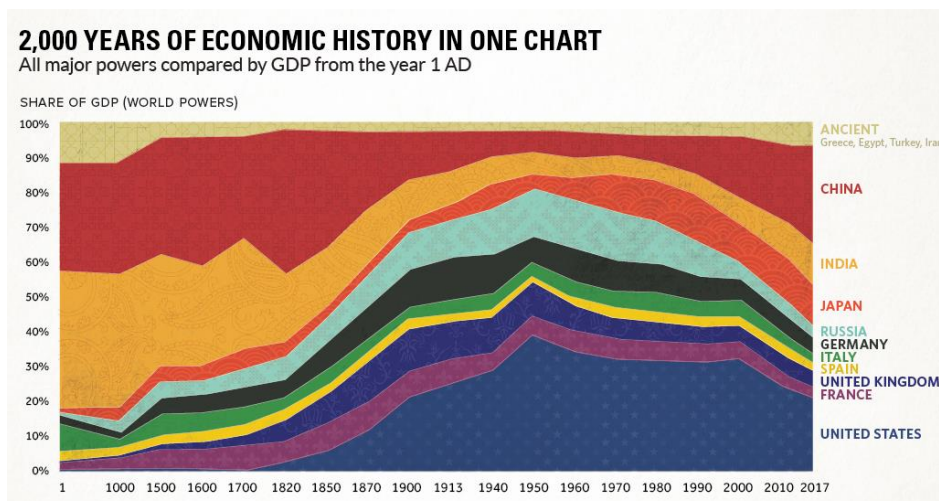


Источник: [28]

Линейный график с областями может отражать как абсолютные, так и относительные значения. Иными словами, он может быть нормирован. Но, так как нормированный график не отражает изменение абсолютных значений, то он искажает восприятие данных. В абсолютном выражении суммарное значение может сильно меняться во времени, а нормированный график этого не покажет. Он отразит только изменение в распределении вклада компонентов. Например, Рисунок 8: абсолютный ВВП в мире сильно увеличился. Объемы ВВП Индии в 1 году н.э. не сопоставимы с объемами ВВП США в начале XX века. Но, смотря на нормированный график и не имея значений абсолютного показателя мирового ВВП, можно сделать такой некорректный вывод, что они равны. Имейте это ввиду, если вы используете нормированные графики с областями, отражая временную динамику показателя.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 8. 2000 лет экономической истории в одном графике



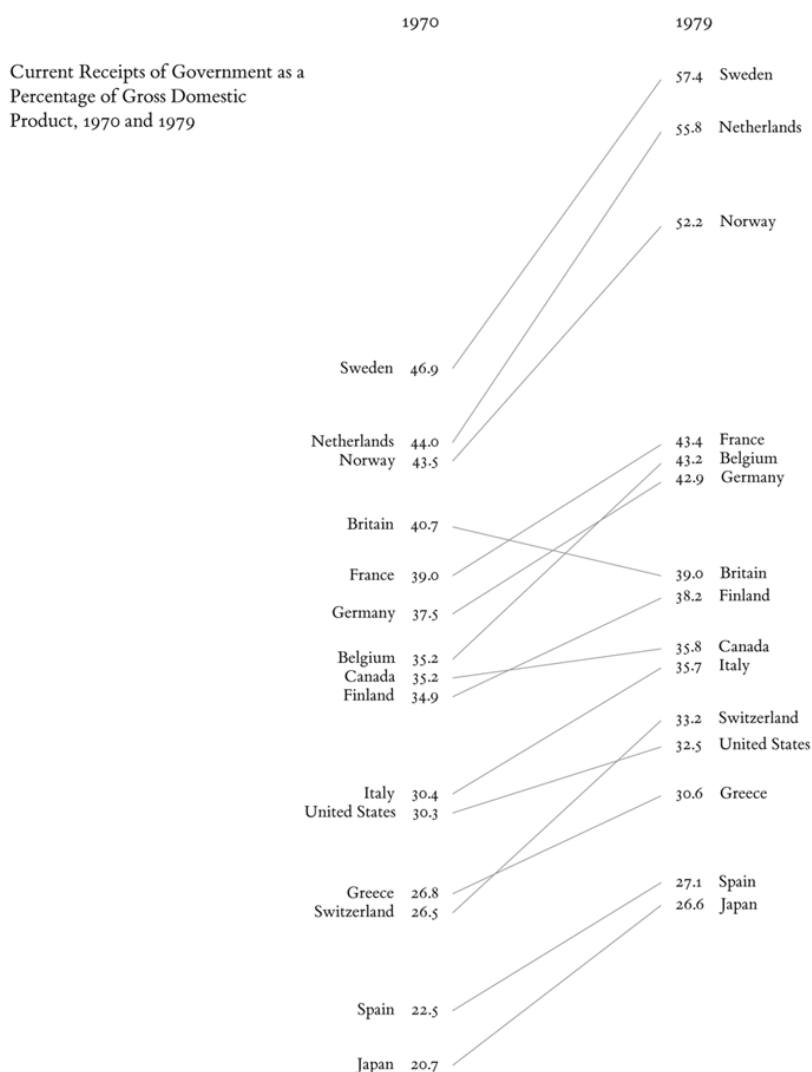
Источник: [9]

3) **График наклона** сравнивает между собой изменения разных объектов (разных рядов данных), ранжированных в определенном порядке или по определенной шкале, с течением времени [10]. Другими словами, как разные объекты меняются во втором периоде по сравнению с первым периодом абсолютно и относительно друг друга. Этот график придумал Эдвард Тафти, описав его в своей книге *The Visual Display of Quantitative Information* (1983) [12] – см. рисунок 9. В классическом виде у этого графика две идентичные параллельные оси. У этих осей одинаковые единицы измерения и одинаковая шкала, различаются только периоды. Он подходит для временного и позиционного сравнения во времени. Особенно выигрышно этот график демонстрирует изменения в рейтингах. Впоследствии график был доработан так, чтобы его можно было использовать для сопоставления нескольких периодов (см. Рисунок 10). Угловой график – образец минимализма. Его удобство заключается в том, что каждый раз название ряда данных (на Рисунке 9 – страны), находится рядом со значением показателя. Также мы сразу видим «рейтинг стран» в первом периоде, и как страны менялись местами во втором периоде. Затем, мы видим за счет угла наклона то, насколько интенсивными были изменения между периодами, и можем сопоставить темпы роста показателя. Сразу заметно, если динамика какого-то показателя принципиально отличается (на Рисунке 9 – это ситуация с Великобританией, у которой показатель снизился) [11]. Однако, если Вы хотите

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

использовать этот график для сравнения нескольких периодов, имейте в виду, что расстояние между осями каждого периода должно быть пропорционально длине этого периода. Другими словами, если у Вас между первым и вторым периодами прошло 10 лет, а между вторым и третьим – 5 лет, то расстояние между осями второго и третьего периодов должно быть в два раза меньше, чем между осями первого и второго, чтобы можно было сравнивать между собой темпы роста показателей на основе угла наклона графиков.

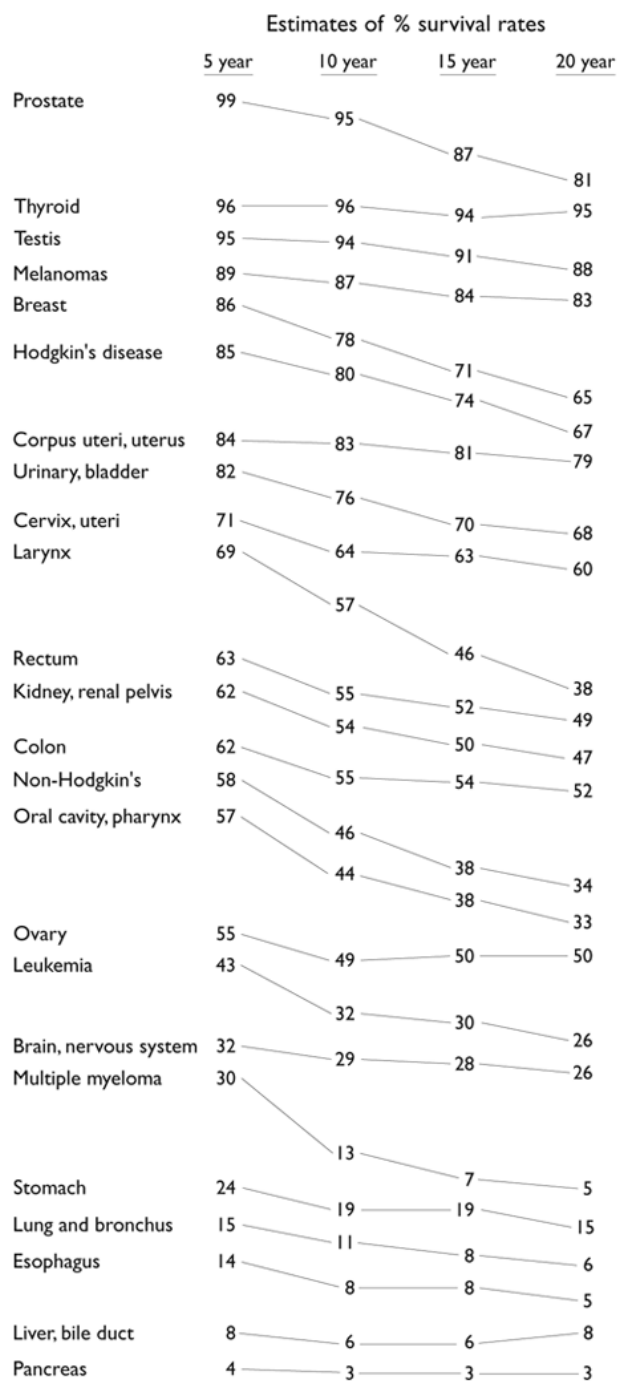
Рисунок 9. Доля налоговых поступлений в бюджет в ВВП в 1970 и 1979 годах.



Источник: [12]

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 10. Визуализация вероятности выживания больных при разных видах онкологических заболеваний на горизонте 5-10-15-20 лет



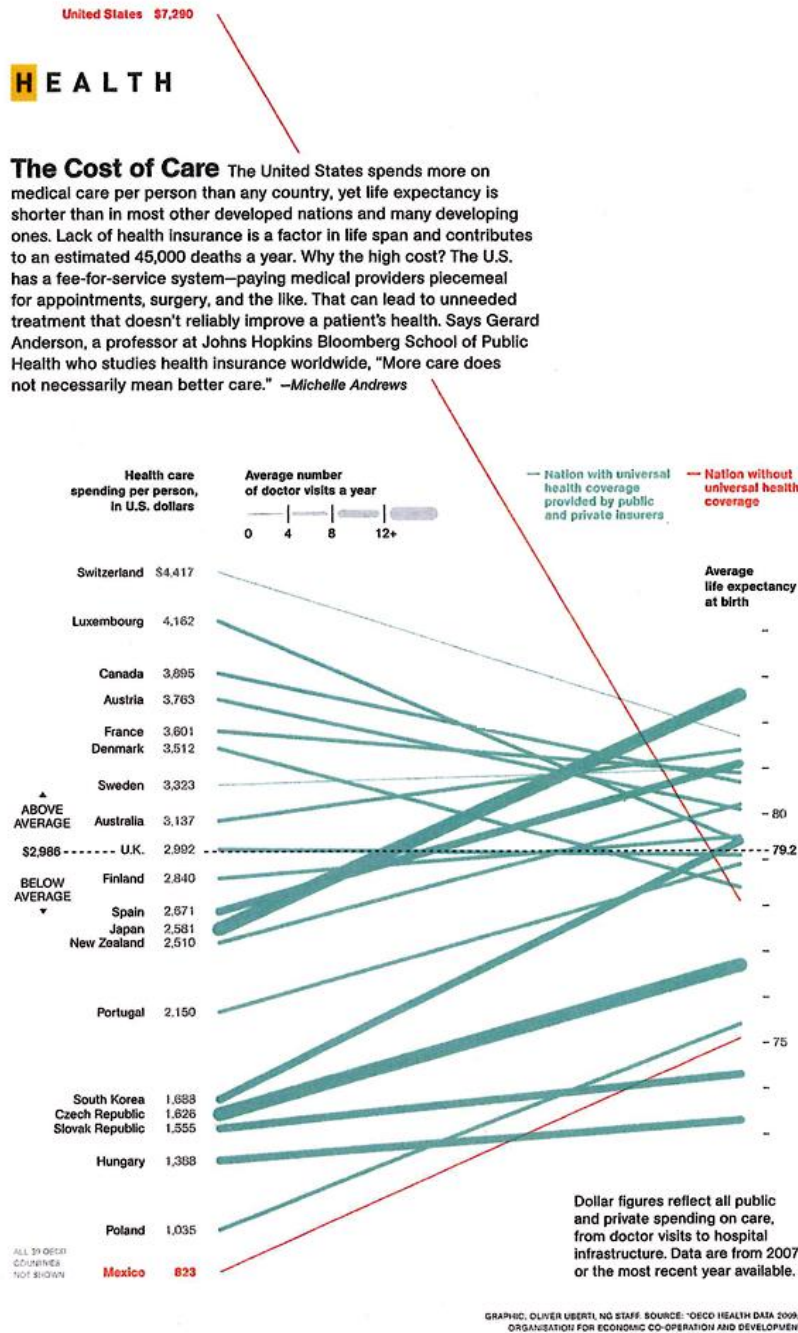
Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Источник: [13]

4) **График наклона с разными единицами измерения на осях.** В случае этого графика наклона две оси показывают разные показатели для одного и того же набора данных. По мнению авторов, использующих этот тип графика, он позволяет показать связь между двумя показателями у разных рядов данных, то есть подходит для корреляционного сравнения. Однако опыт использования данного графика показывает, что, во-первых, визуальным эффектом можно манипулировать за счет изменения масштаба одной из осей (как это сделано на Рисунке 11, чтобы драматизировать ситуацию в США). Во-вторых, этот тип графика менее достоверно показывает корреляцию, по сравнению с точечной диаграммой (диаграммой рассеивания). Так как график на Рисунке 10 вызвал множество замечаний, National Geographic в итоге удалили его со своего сайта, а дизайнер инфографики заменил его на точечную диаграмму (см. Рисунок 26).

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 11. Сопоставление расходов на здравоохранение в расчете на 1 жителя в государстве (в долл. США) и ожидаемой продолжительности жизни



Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

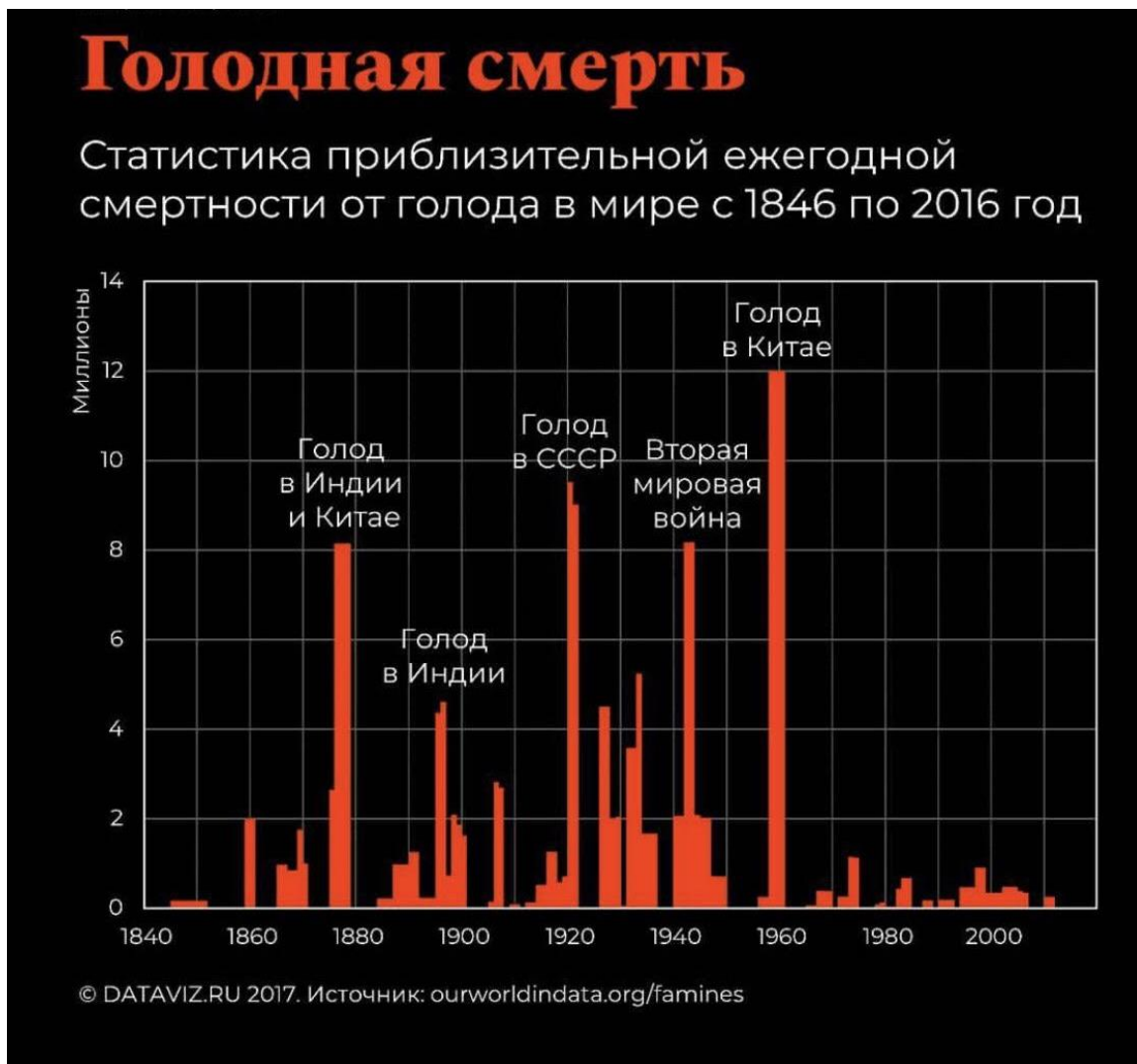
Источник: [14]

5) **Столбчатая диаграмма.** Эта диаграмма позволяет наиболее точно сравнить конкретные цифры. Хорошо подходит для позиционного, временного и частотного сравнений. Представляя данные в виде этой диаграммы, их можно сортировать разными способами: по времени, по возрастанию/убыванию, по алфавиту. Последний способ используется в таких случаях, как, например, данные о субъектах Российской Федерации, для простоты поиска нужного значения. Ограничением для использования столбчатой диаграммы является большая количественная разница между данными (на Рисунке 13 эту проблему попробовали решить, разместив на линейках Китая и Японии наклонную линию, показывающую разрыв линейки). В этом случае диаграмма становится не очень информативной. Столбчатая диаграмма бывает двух видов: вертикальная (гистограмма) и горизонтальная (линейчатая диаграмма).

Гистограмма традиционно используется для сравнения, особенно если это временная динамика. Также может использоваться и для позиционного сравнения. Если Вам нужно выделить какое-то значение на гистограмме, используйте контрастный цвет для соответствующей значению линейки. Также другим цветом или пунктирным контуром Вы можете выделить прогнозные значения при временном сравнении, если они у Вас есть.

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Рисунок 12. Голодная смерть



Источник: [29]

Линейчатая диаграмма чаще используется для позиционного сравнения, особенно если у категорий длинные названия, и они просто не помещаются при подписи значений на гистограмме. Этот тип диаграммы нельзя использовать для временного сравнения. Если Вы хотите выделить какое-то из значений, то можете расположить соответствующую ему линейку выше всех других, или акцентируйте эту линейку цветом, ближе всех к оси абсцисс. Мы настоятельно не рекомендуем Вам вставлять подписи данных между линейками. Это

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

визуально разбивает полотно графика, затрудняет читателю сравнение показателей и в целом плохо смотрится [1].

Рисунок 13. Крупнейшие иностранные держатели облигаций госдолга США, млрд. долл.



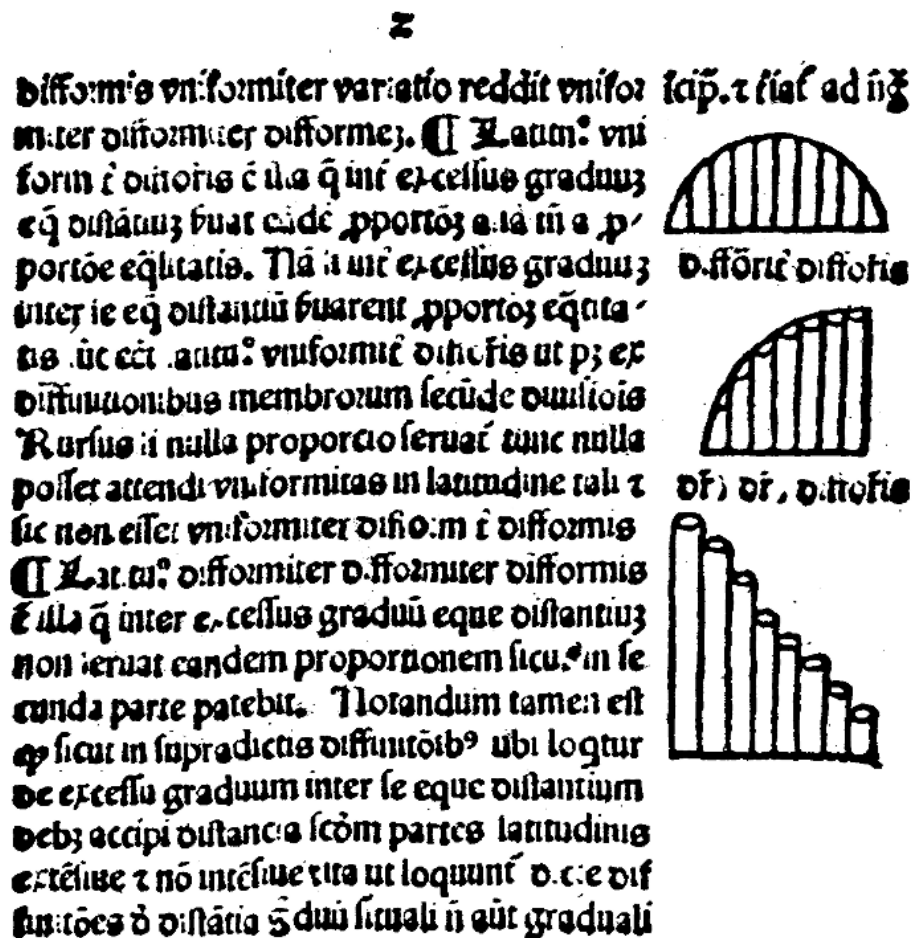
* Данные на июль 2017 года

Источник: [28]

Впервые столбчатая диаграмма встречается в трактате «Tractatus de latitudinibus formarum» (XIV век): на ней проиллюстрировано соотношение времени и скорости ускоряющегося объекта (см. Рисунок 14). Автором этого трактата является либо французский математик Николь Орезм (Nicole Oresme, 1320-1382), либо итальянский математик Якобус де Санто Мартино (Jacobus de Sancto Martino, XIV век).

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 14. Первая в истории гистограмма, XIV век

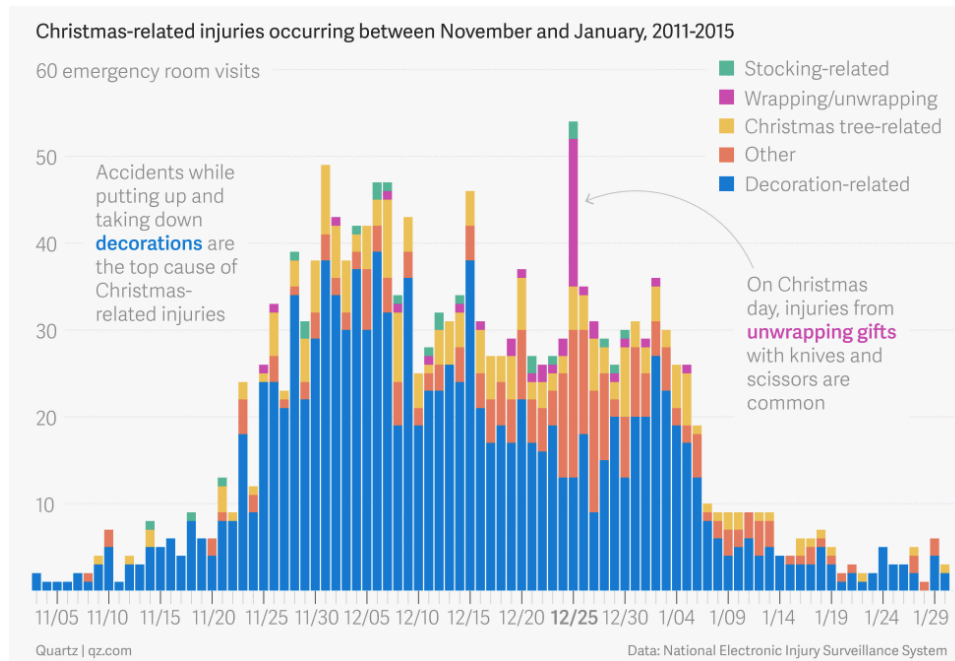


Источник: [35]

б) Стековая диаграмма (гистограмма с накоплением) позволяет проводить позиционное или временное сравнение, при этом сравнивая объект в целом, а также его структуру. Другими словами, она показывает изменение целого и вклад в это изменение отдельных компонентов. Не стоит использовать больше 5 компонентов для сравнения внутри целого на этой диаграмме. Если Вам важно отразить то, как меняется отдельный компонент, то следует разместить его у основания диаграммы (оси абсцисс) и выделить цветом. За счет наличия фиксированной нижней границы значение этой категории читателю будет оценить проще всего. Стековая диаграмма может быть представлена как в абсолютных значениях, так и нормирована на 100%.

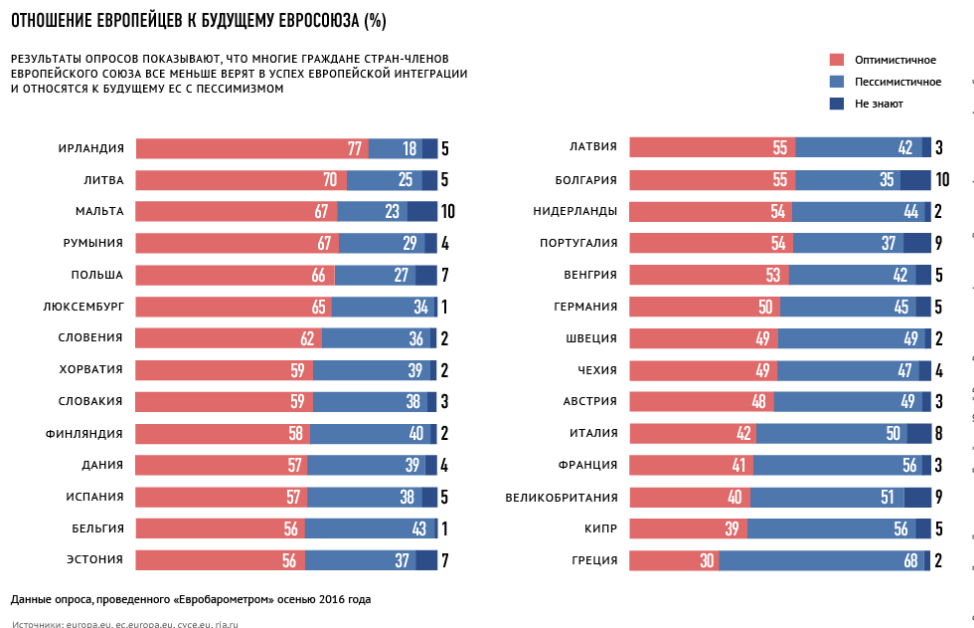
Материалы к вебинару 22.01.2018 «Инфографика и визуализация данных»

Рисунок 15. Наиболее распространенные травмы, связанные с Рождеством в США в 2011-2015 года



Источник: [36]

Рисунок 16. Отношение европейцев к будущему Евросоюза (%)



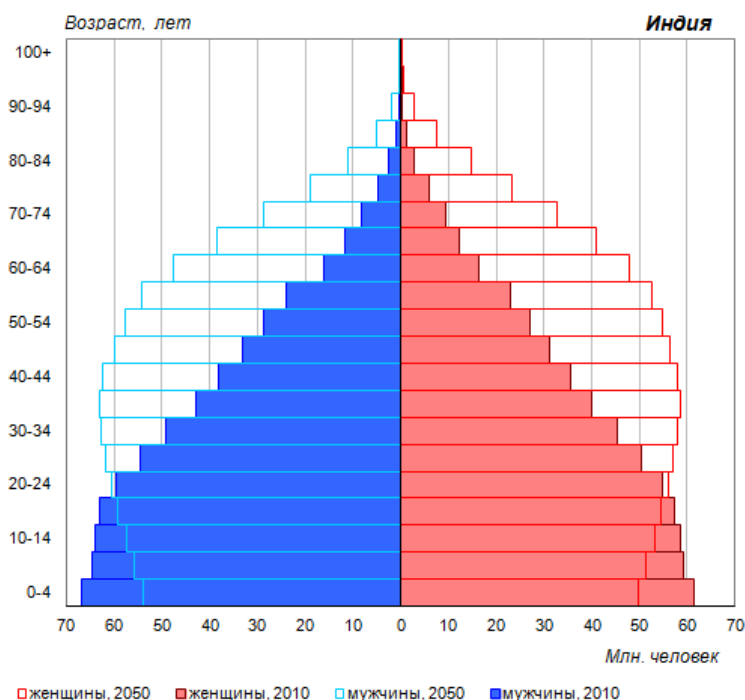
Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Источник: [30]

7) **Диаграмма распределения** (двусторонняя линейчатая) подходит для частотного сравнения. Она хорошо показывает, если имеется 2 два типа объектов, и они распределены по одинаковым интервалам: например, половозрастная структура населения. Она представляет собой две линейчатые диаграммы, зеркально отраженные относительно оси ординат. Диаграмма распределения позволяет Вам точно отразить количественную структуру двух рядов данных. Также она может использоваться для корреляционного сравнения, чтобы показать, есть ли зависимость между показателями. При таком подходе независимую переменную располагают слева по убыванию или возрастанию ее значений, а зависимая переменная справа при наличии зависимости должна будет зеркально отразить поведение независимой (прямо или обратно пропорционально, смотря какое направление зависимости). Для визуализации корреляционного сравнения диаграмму распределения следует использовать только, если у Вас не очень большое число наблюдений (не более 15). При большем числе значений мы рекомендуем для корреляционного сравнения использовать график рассеивания (точечную диаграмму).

Рисунок 17. Распределение населения Индии по полу и возрасту в 2010 и 2050 годах по среднему варианту прогноза ООН пересмотра 2010 года, миллионов человек

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)



Источник данных: United Nations, Department of Economic and Social Affairs, Population Division (2011). World Population Prospects: The 2010 Revision, CD-ROM Edition. POP/DB/WPP/Rev.2010/03/F02, F03

Источник: [32]

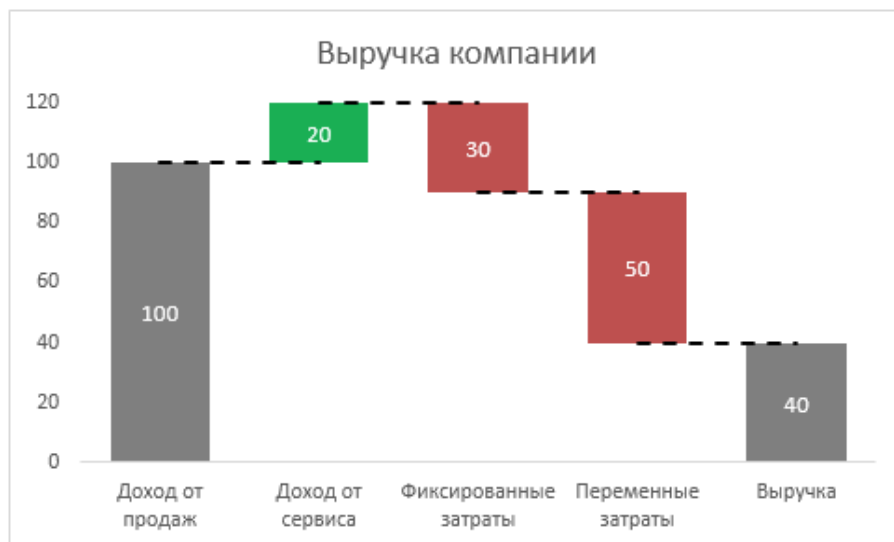
8) **Каскадная диаграмма** (диаграмма-водопад, диаграмма-мост) используется для наглядного понимания совокупного эффекта на изначальное значение показателя последовательного воздействия положительных и отрицательных значений. При этом изначальное и конечное значение представляет собой целую линейку (столбец), а промежуточные значения – плавающие линейки. Отправная точка (изначальное значение) – это всегда столбец, который начинается от нуля. Значения промежуточных положительных элементов отражаются в виде столбцов, которые начинаются в верхней точке предыдущего столбца и откладываются вверх. Отрицательные элементы начинаются в верхней точке предыдущего столбца и продлеваются вниз. Результирующее (конечное значение) представляет собой расстояние от верхней (для случая положительных воздействий) или нижней (для случая отрицательных воздействий) границы последнего промежуточного

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

элемента до нуля. Положительные и отрицательные промежуточные значения выделяют на диаграмме разными цветами. С помощью каскадной диаграммы можно отражать как статичные показатели (отчет о прибылях и убытках), так и изменяющиеся во времени (денежный поток). Положительные и отрицательные воздействия Вы можете показывать как вперемешку, так и группировать их, показывая сначала все воздействия одного направления, а затем все воздействия другого [15]. Эта диаграмма стала известна благодаря консалтинговой компании McKinsey, которая очень активно использует ее в своих презентациях. В целом этот тип диаграммы встречается исключительно в корпоративной сфере: отчетах о деятельности компании и т.п. Важно понимать, что каскадную диаграмму следует использовать только в тех случаях, когда на изначальное значение показателя имеются разнонаправленные воздействия, как положительные, так и отрицательные. Эту диаграмму часто некорректно используют для простого покомпонентного сравнения, показывая вклад отдельных элементов в итоговое значение. Для этих целей следует использовать круговую диаграмму или гистограмму с накоплением.

Каскадную диаграмму иногда называют диаграммой летающих кирпичей, а также диаграммой Марио, так как она напоминает кирпичи с бонусами из известной игры «Марио» для компьютерных приставок.

Рисунок 18. Выручка компании, визуализированная классической каскадной диаграммой



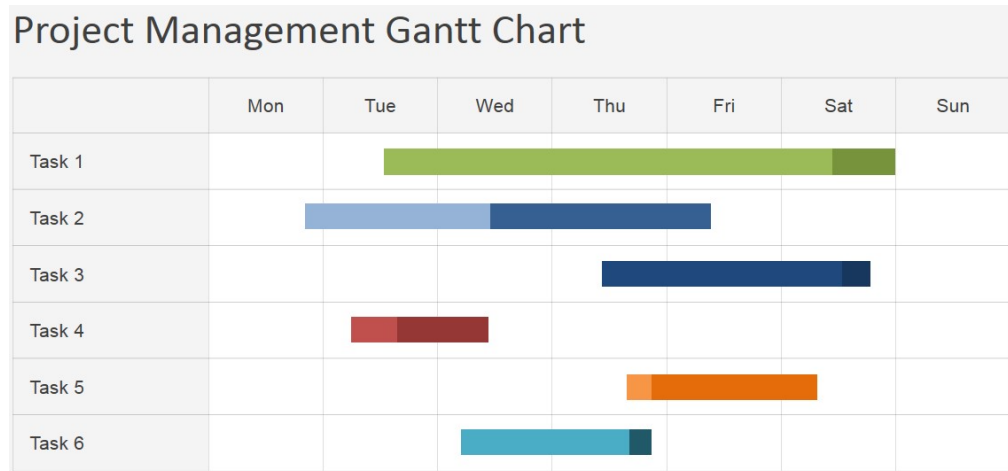
Материалы к [вебинару](#) 22.01.2018 [«Инфографика и визуализация данных»](#)

Источник: [37]

9) **Диаграмма Гантта** – это еще один частный случай линейчатой диаграммы. Она используется для визуализации графика работ, плана действий в рамках какого-то проекта. Диаграмма Гантта является одним из базовых элементов в процессах управления проектами. Она состоит из двух осей, одна из которых представляет собой временную шкалу (ось абсцисс), а вторая – перечень задач, требующих исполнения (ось ординат). Также в диаграмме Гантта используется такое понятие как «веха» для обозначения ключевых событий, контрольных точек, окончания определенных этапов и т.п. Введение в практику управления производством этой диаграммы стало одним из прорывов в управленческой мысли XX века. Но Вам нужно иметь в виду, что эта диаграмма плохо применима к сложным проектам с большим количеством задач, так как перестает быть наглядной.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

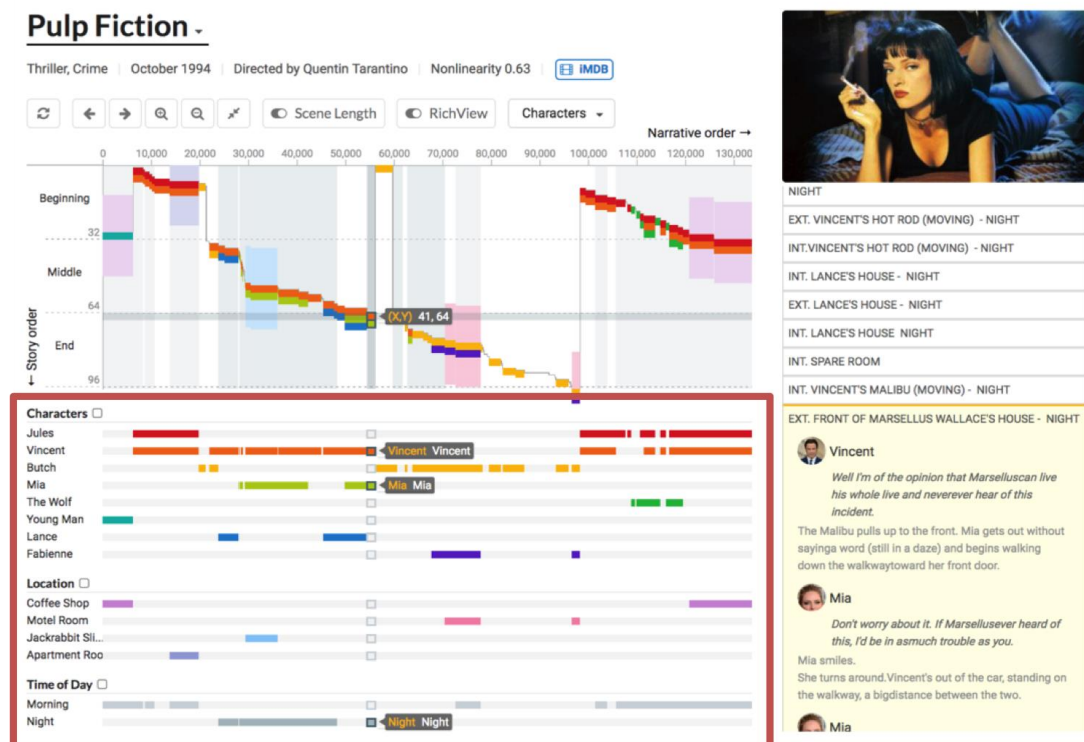
Рисунок 19. Классический вид диаграммы Гантта



Источник: [38]

На Рисунке 20 Диаграмма Гантта представлена в нижней секции (обведена красной рамкой) и визуализирует, когда и с кем вместе появляется герой в фильме «Криминальное чтиво».

Рисунок 20. Визуализация нелинейных историй: Криминальное чтиво



Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Источник: [39]

Диаграмма Гантта названа в честь Генри Гантта (Henry Laurence Gantt, 1861-1919), американского консультанта по управлению, ученика и последователя Фредерика Тейлора. Он предложил использовать диаграмму для иллюстрации и контроля выполнения задач на производстве. Однако до него этот метод управления проектами уже придумал польский экономист и специалист в области управления производством Кароль Адамецкий (Karol Adamiecki, 1866-1933) в 1896 году. Он назвал свою диаграмму «гармонограф», и даже написал отдельную статью, посвященную этому методу организации производства ([16]). В варианте Адамецкого временная ось шла по оси ординат, а каждый процесс представлял собой отдельную бумажную ленту, которую можно было вынуть, отредактировать или заменить. Кстати, помимо гармонографа Адаметский также придумал хронометраж – способ измерения продолжительности трудовых процессов [17 и 18].

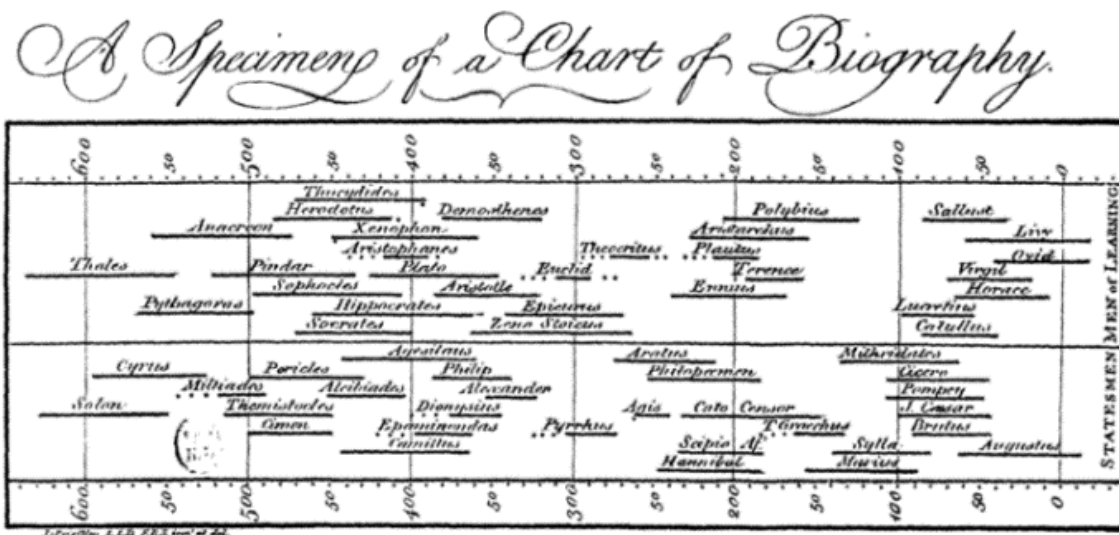
Рисунок 21. Портрет Кароля Адамецкого



Но на самом деле обоих этих титанов управленческой мысли обогнал Джозеф Присли (Joseph Priestley, 1733-1804), использовав прообраз диаграммы Гантта еще в 1765 году для своих лекций об истории и общей политике, которые впоследствии были опубликованы в виде сборника в 1788 году. Он использовал линейчатую диаграмму для отражения исторических событий, которые по природе своей являются дискретными: имеют начало и конец на временной шкале. Вместо перечня задач по оси ординат он откладывал имена исторических персонажей (около 2000 человек) [19].

Материалы к вебинару 22.01.2018 «Инфографика и визуализация данных»

Рисунок 22. «График биографий» Джозефа Присли

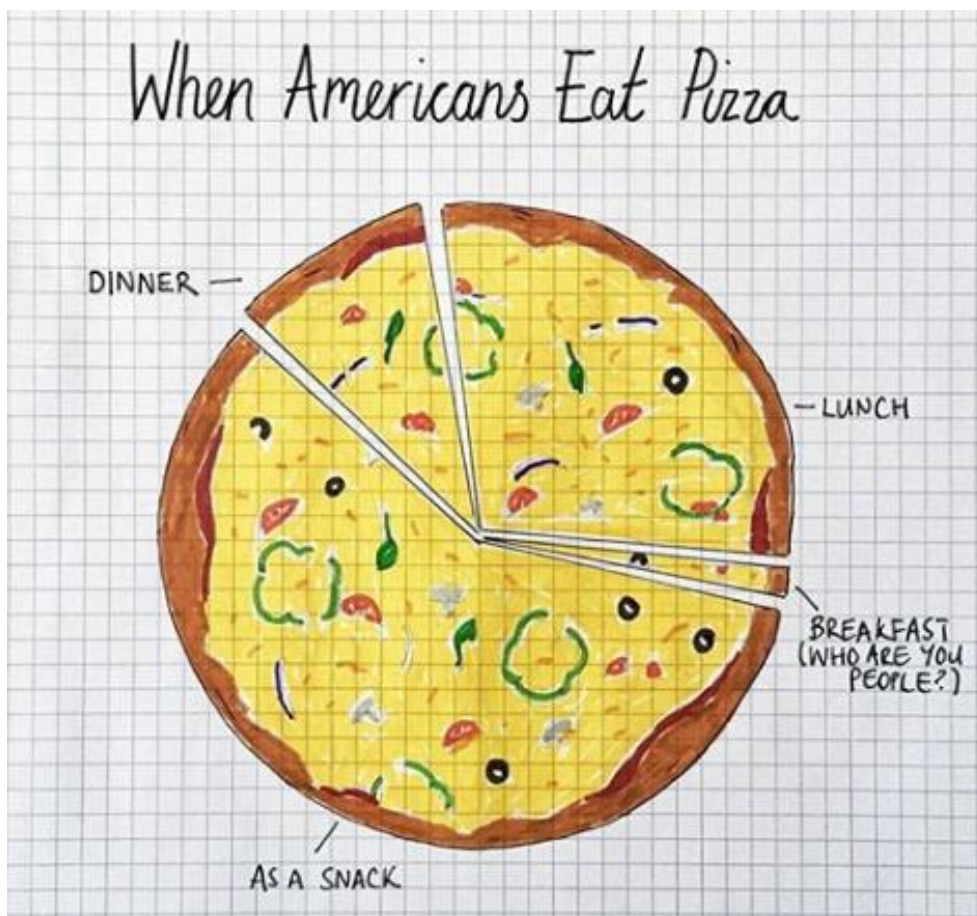


Источник: [19]

10) **Круговая диаграмма** (пайчарт) показывает пропорции и проценты между категориями, путем деления окружности на пропорциональные отрезки. Каждая длина дуги представляет собой долю каждой категории, в то время как полный круг представляет собой общую сумму всех данных, равный 100%. Круговая диаграмма идеально иллюстрирует идеи пропорционального распределения данных. Форма круга лучше всего подходит для создания впечатления целого, а сама диаграмма – чтобы показать, сколько изучаемый объект занимает от целого (в %). На ней наглядно видно, кому достается «большая часть пирога». Если Вам необходимо привлечь внимание к какому-то сегменту, то расположите его на линии 12 часов и выделите контрастным цветом. В остальных случаях располагайте показатели по убыванию и помните, что взгляд читателя обычно движется по часовой стрелке. При этом не забывайте, что сравнение долей здесь не очень точное. Это может быть недостатком, если Вам необходимо продемонстрировать точное сравнение. К тому же круговая диаграмма занимает много места из-за своего размера и необходимости отдельной легенды. Более компактно покомпонентное сравнение можно с помощью нормированной гистограммы с накоплением (стековой диаграммы) [1 и 33]. Также круговую диаграмму иногда критикуют за то, что она привлекает слишком много внимания читателя к сопоставлению всех долей, акцентируя

Материалы к [вебинару](#) 22.01.2018 [«Инфографика и визуализация данных»](#)

внимание на площади секторов. В итоге внимание читателя отвлекается от основной идеи. Частично эту проблему позволяет решить кольцевая диаграмма. Рисунок 23. Когда американцы едят пиццу?



Источник данных: US department of Agriculture, 2014

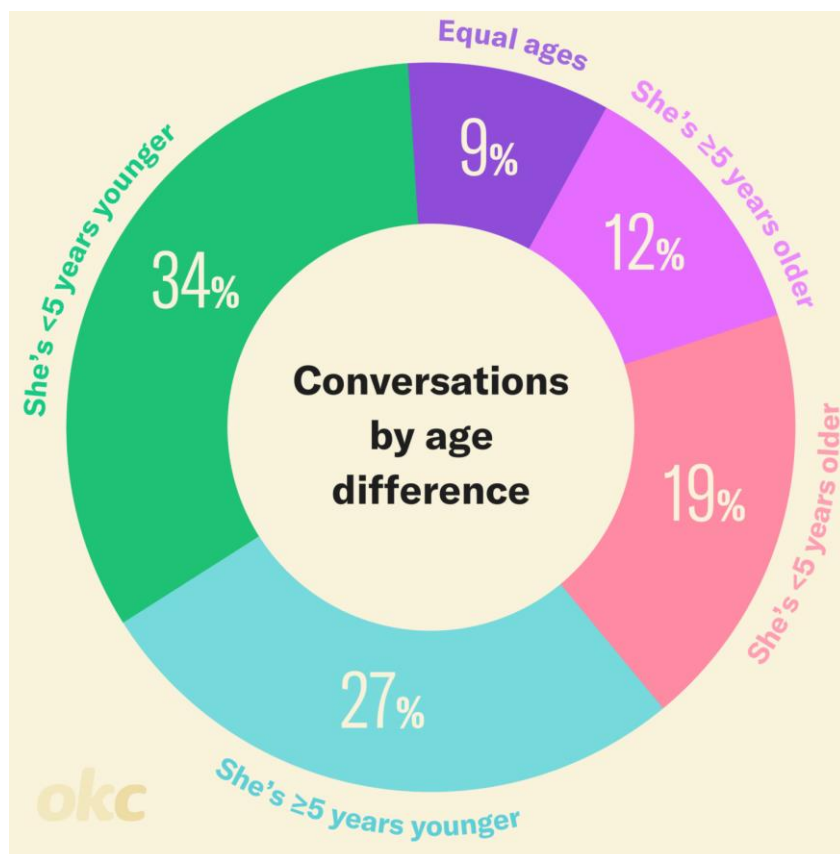
Источник: [40]

Не следует использовать круговую диаграмму, если: 1) больше 6 значений (если значений больше, используйте Тримап или выберите 5 наиболее важных частей, а остальные объедините в категорию «прочие»); 2) сильный разброс значений; 3) значения близки, а нам важна точность сравнения; 4) есть динамика показателей или сравнение структур двух объектов – в этом случае используйте стековую диаграмму, разбитую на секторы.

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

11) **Кольцевая диаграмма** (пончиковая) на самом деле является круговой диаграммой, с вырезанной в середине площадью. В кольцевой диаграмме меньший акцент делается на площади секторов, и внимание читателя переключается на чтение длины дуг. К тому же кольцевая диаграмма занимает меньше места, так как пространство внутри может использоваться для размещения легенды или другой имеющей отношение к графику информации [34].

Рисунок 24. Более половины «успешных»* разговоров на сайте знакомств происходят между более взрослым мужчиной и более молодой девушкой



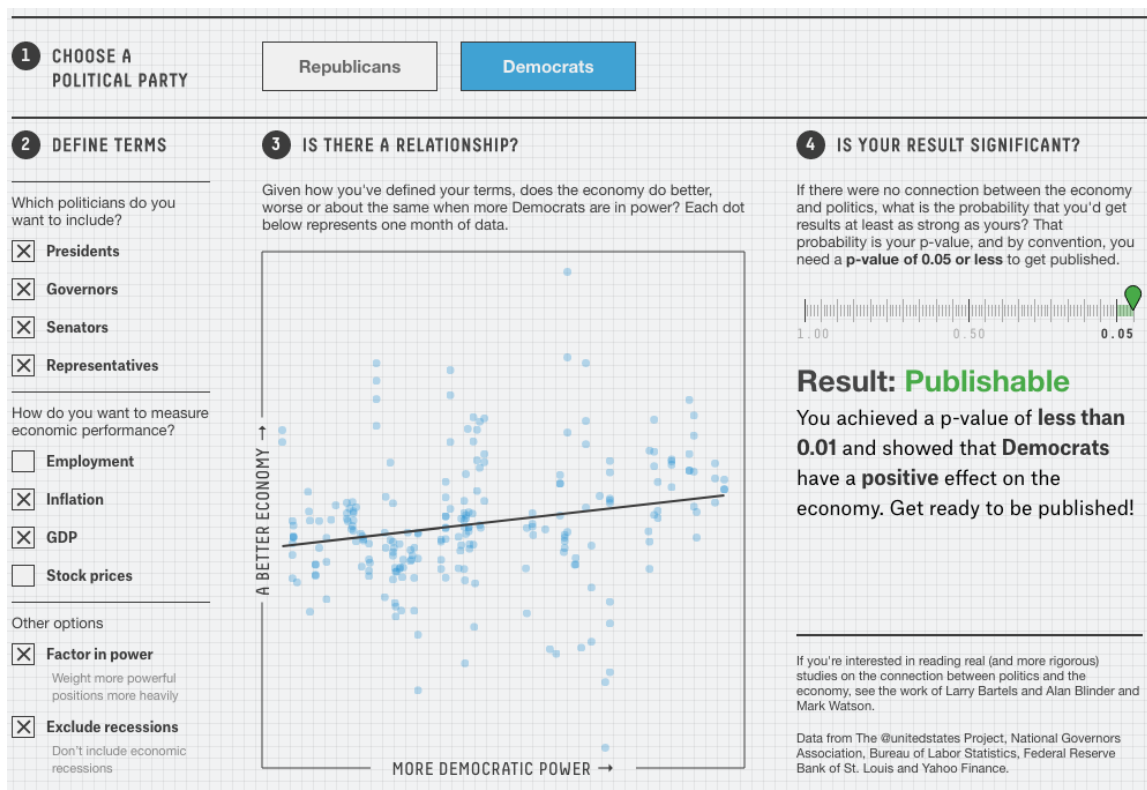
Источник: [41]

* Под «успешным» подразумевается диалог в форме сообщений на сайте знакомств, в котором было не менее 4-ех сообщений от каждого участника и произошел обмен контактами.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

12) **Диаграмма рассеивания (точечная диаграмма)** показывает зависимость между двумя показателями, отражаемыми на осях координат. Этот тип диаграммы используется для корреляционного сравнения. Точечная диаграмма позволяет увидеть закономерности, тренды и выпадающие значения (или их отсутствие). На точечной диаграмме можно выделять тренды и проверять ожидаемые зависимости. Одна из сложностей использования этой диаграммы связана с подписью значений: при их большом количестве трудно ставить подписи данных (название ряда и значение точки). Если Вам необходимо продемонстрировать только тренд, наличие и направление зависимости, то точки можно не подписывать. Если обозначения точек важны, Вы можете обозначить их буквами или цифрами, а затем расшифровать их значение в легенде. При наличии общего тренда и каких-то выбросов из общей закономерности, мы советуем Вам подписать эти выбросы, так как они в первую очередь привлекут внимание читателя. Нас всегда интересует то, что отличается, что не соответствует правилу и нашим ожиданиям.

Рисунок 25. Связь между качеством экономических показателей США и властью демократов



Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

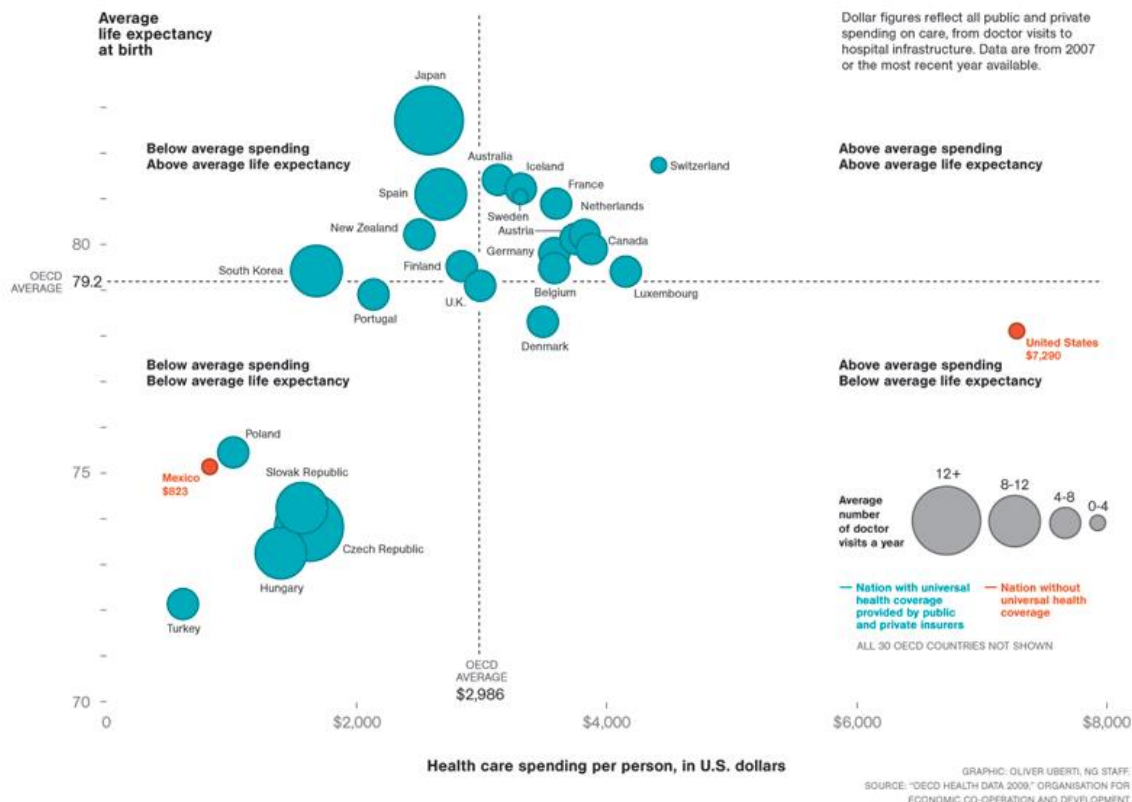
Источник: [42]

13) **Пузырьковая диаграмма** представляет собой смешение точечной и площадной диаграммы. На ней данные отражаются в трех измерениях. Каждое значение изображается в виде диска (пузыря), где два его параметра соответствуют координатам по осям, а третье изображается с помощью площади диска. На самом деле в такой диаграмме Вы можете отразить до 6 параметров. Дополнительные три измерения можно получить через различия в форме, цвете и оттенке пузыря. Однако форма и цвет отражают качественные параметры, и только оттенок добавляет еще один количественный.

Проблема пузырьковых диаграмм заключается в том, что масштабирование пузырей нелинейное. Это создает трудности при сравнении показателей между собой, особенно, если между ними сильный разброс в значениях. Из-за этого мы не рекомендуем Вам использовать пузырьковую диаграмму там, где требуется высокая точность сравнения данных. Также трудности у Вас может вызвать изображение на пузырьковой диаграмме наблюдений, у которых третий параметр имеет отрицательное значение. Вы можете выйти из этой сложной ситуации, использовав другой цвет для пузыря, соответствующего отрицательному значению. Также Вы можете поместить в центр пузыря (диска) особый значок, который в легенде опишете как индикатор отрицательного значения показателя.

Материалы к вебинару 22.01.2018 «Инфографика и визуализация данных»

Рисунок 26: Альтернативная визуализация расходов на здравоохранение в расчете на 1 жителя в стране (в долл. США) и ожидаемой продолжительности жизни



Источник: [11]

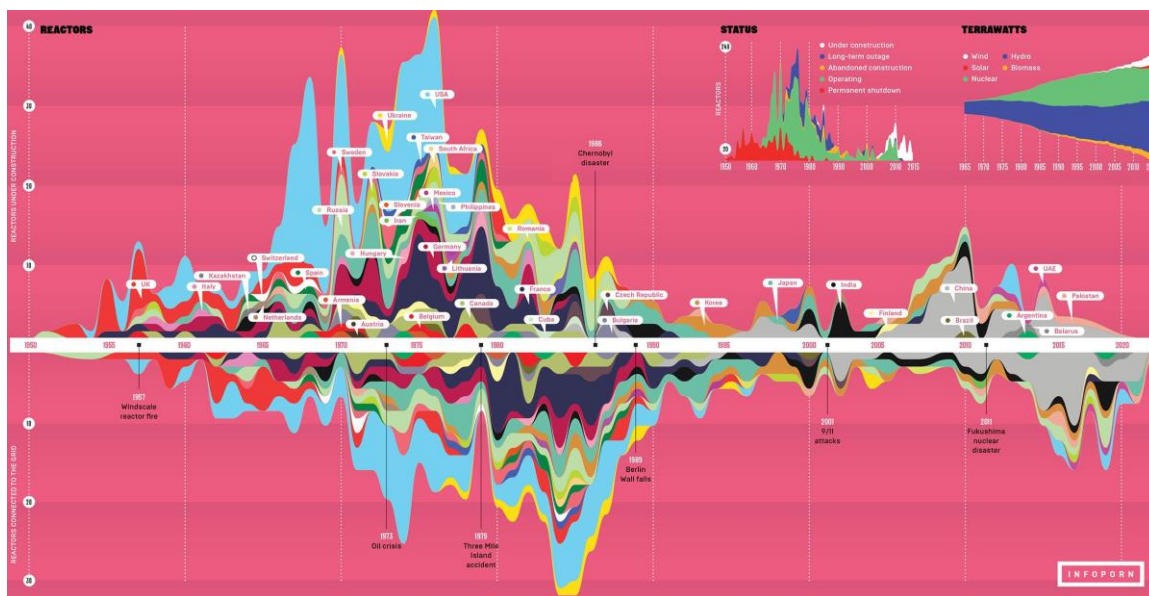
14) **Диаграмма потока** визуализирует распределение одномерного явления в зависимости от значения независимого фактора, влияющего на это явление. Чаще всего при использовании потоковой диаграммы внешним фактором является время. Эта диаграмма – вариация на тему линейного графика с областями в сочетании с площадной диаграммой. В данном случае значения откладываются не относительно оси абсцисс, а относительно центра потока, а ось значений показателя отсутствует. Диаграмма отражает изменение значений данных в зависимости от изменения независимого параметра (например, времени) в виде потока, напоминающего реку. Ширина потока на каждом конкретном участке зависит от значений показателей. При этом в потоковом графике также могут отражаться значения с накоплением, то есть разные категории в рамках одного потока. Они могут выделяться, например, разными

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

цветами. Однако мы не советуем Вам визуализировать больше 4-5 категорий данных в одном потоке.

Данную диаграмму хорошо использовать для анализа больших массивов данных и их изменений во времени (или под воздействием другого фактора). Эта диаграмма выигрышно смотрится приятно с точки зрения эстетического восприятия. Но мы не рекомендуем Вам использовать ее, если Вам необходима высокая точность визуализации. Это связано с тем, что у данного графика отсутствует шкала значений и оценить точное значение ширины потока читателя практически невозможно. При использовании в одном потоке нескольких категорий те из них, которые имеют маленький вес, станут попросту незаметны на фоне потоков других категорий. Таким образом, использовать потоковую диаграмму нужно для ситуаций, когда Вам нужно отразить общие изменения данных без необходимости углубляться в детали и общие значения [20].

Рисунок 27. Сокращение использования ядерных реакторов



Источник: [43]

15) **Диаграмма Санкей** – подвид потоковой диаграммы, в которой поток разделяется на отдельные ручки-стрелки, каждая из которых пропорциональна количеству попадающих в нее элементов. Эта диаграмма отражает количественное соотношение рукавов потока между

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

собой. Она позволяет оценить, какое направление делает наибольший вклад в общее явление. Чем шире поток, тем больше его вклад. При этом потоки могут разделяться или наоборот объединяться на разных этапах.

Рисунок 28. Перспективы либерализации визового режима для Украины

Перспективы либерализации визового режима

Что ждут украинцы
от отмены виз, %

Куда украинцы
чаще всего ездят, %

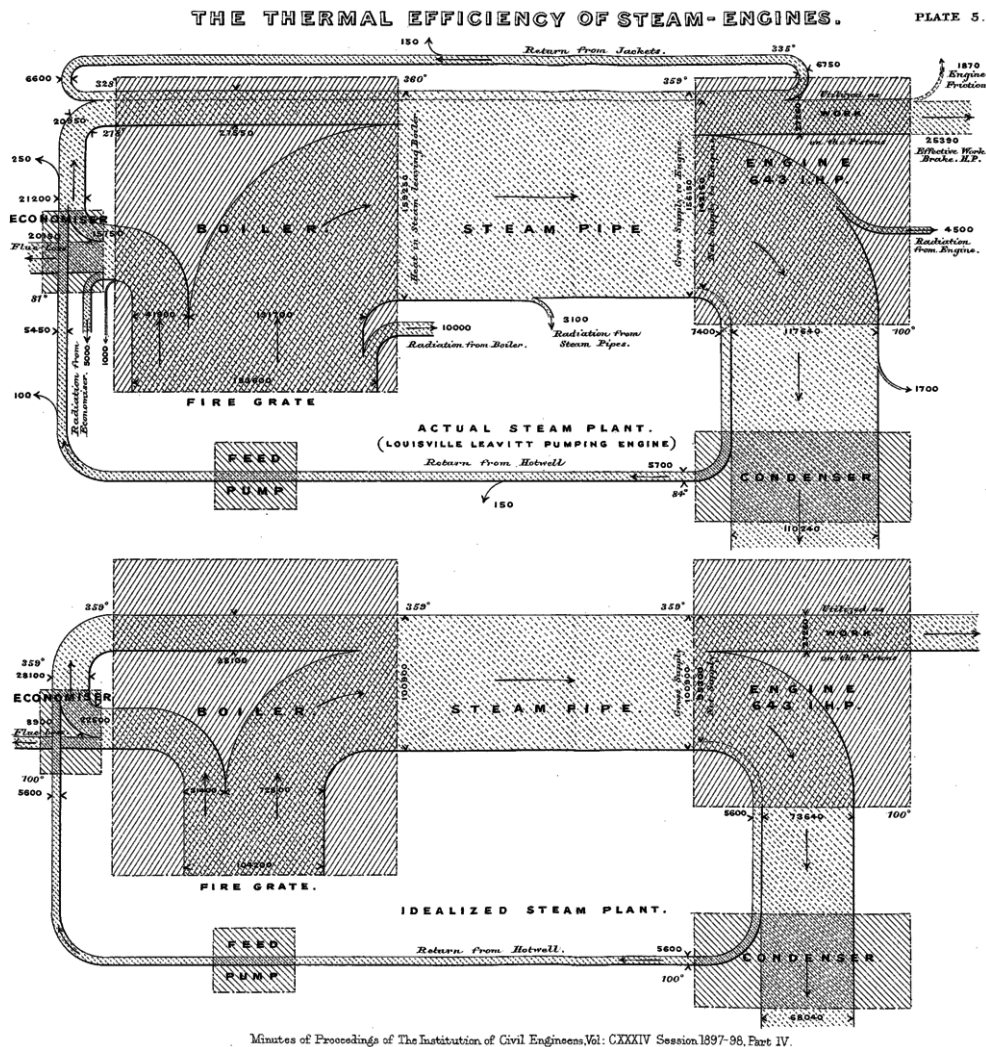


Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Источник: [31]

Название диаграмма получила по имени ирландского капитана Мэтью Санкея (Matthew Henry Phineas Riall Sankey, 1853 - 1926), который в 1898 году изобразил с ее помощью эффективность использования энергии парового двигателя [21].

Рисунок 29: Температурная эффективность парового двигателя



Источник: [21]

16) **Карта потока** (карта-схема) визуализирует пространственное распределение одномерных явлений, то есть географическое движение объектов из одного места в другое и количество этих объектов. Чаще всего эта диаграмма используется для изображения миграции

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

населения и животных, а также потоков товаров между разными географическими рынками. В своей основе эта диаграмма имеет потоковую диаграмму, которую накладывают на карту. Как и в случае обычной потоковой диаграммы, толщина карты потока отражает количество перемещаемых объектов [22]. Эта диаграмма позволит Вам наглядно показать различия в интенсивности потока по мере его перемещения или сравнить потоки, связывающие разные географические точки.

Рисунок 30. Мировая торговля ресурсами в 2015 году

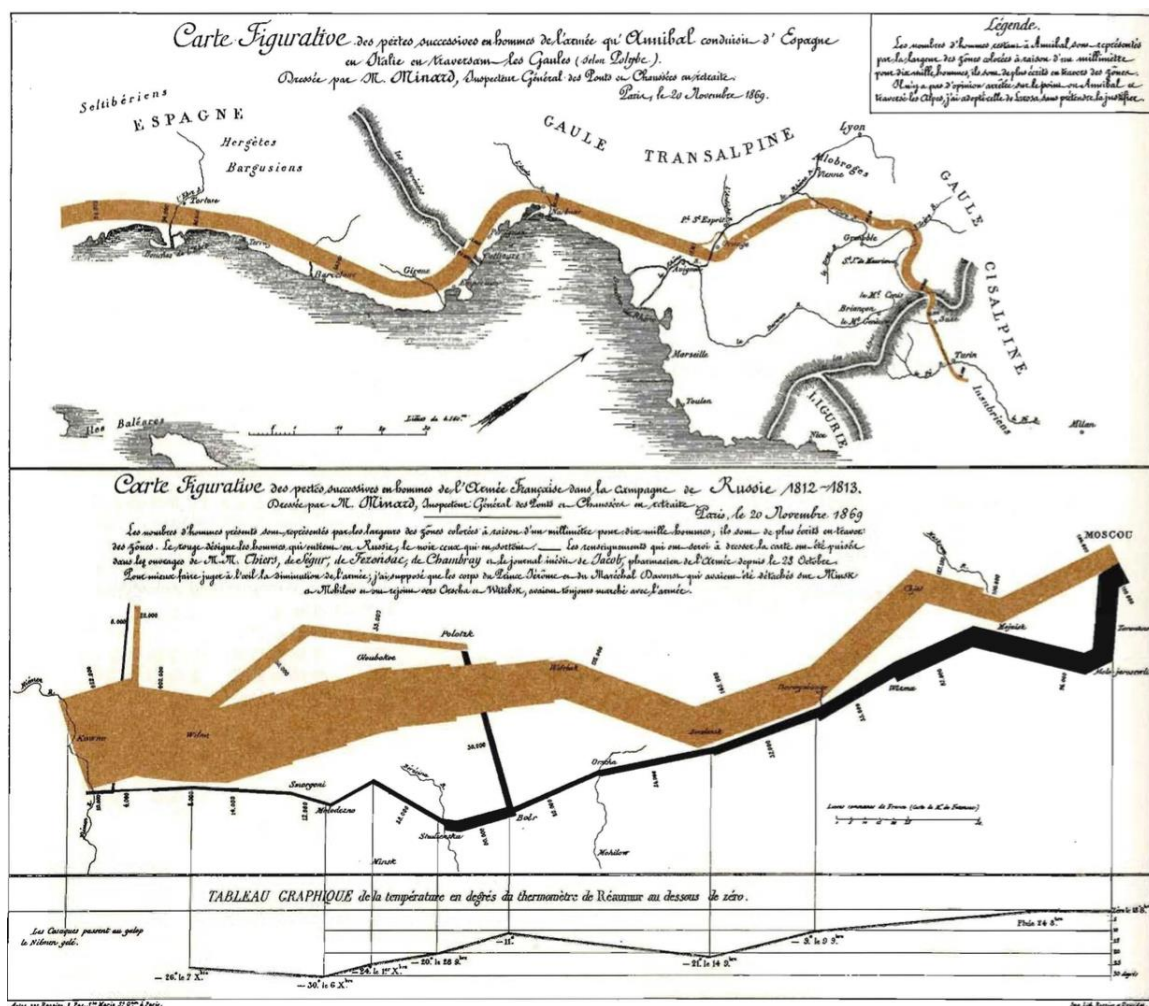


Источник: [44]

Самые первые потоковые карты иллюстрировали интенсивность пассажиропотоков на железных дорогах в Ирландии [23]. Но одной из самых известных потоковых карт в истории является карта-схема русского похода Наполеона, созданная французским инженером Шарлем Минаром (Charles-Joseph Minard, 1781-1870).

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 31. Графическая карта последовательных потерь во французской армии в период Русской Кампании 1812-13 (по сравнению с потерями Ганнибала во время 2-й Пунической войны)



Источник: [24]

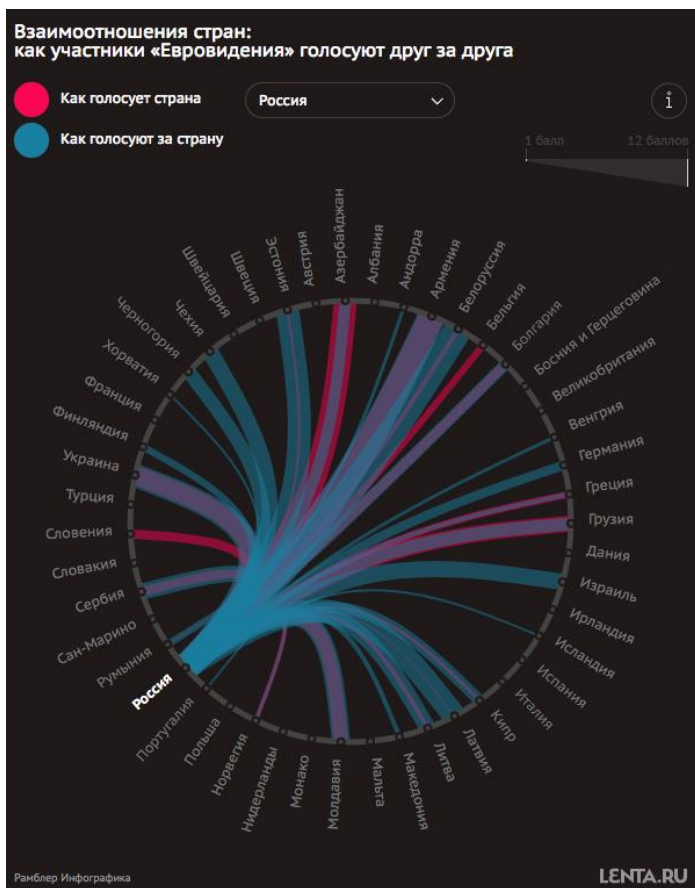
Студия инфографики ТАСС сделала интересный проект, связанный с потоковой диаграммой Минара: [«Когда пришёл Наполеон»](#) [25]. Инфографы ТАСС решили проверить точность данных Минара и показать, где именно пролегал маршрут Великой армии Наполеона. Реконструкцию маршрута они показали на современной карте мира.

17) **Хордовая диаграмма** иллюстрирует взаимосвязи между наблюдениями следующим образом: данные располагаются радиально по кругу, а связь между значениями изображается

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

с помощью хорд (отрезков, соединяющих две точки на окружности; хорды на самом деле являются кривыми Безье). Эти связи между точками показывают, что у них есть нечто общее. Толщина хорд отражает количественные характеристики этой общности. Хордовые диаграммы подходят для сравнения сходства в наборе данных или между различными группами данных. Однако точность такого сравнения довольно низкая. Информативна эта диаграмма только в случае, если на ней отражать небольшое количество значений и связь между ними. При большом количестве данных Вы не сможете на ней толком ничего понять [26 и 27]. Однако она может очень красиво выглядеть, поэтому ее любят использовать в искусстве, основанном на данных датаарт.

Рисунок 32. Как участники голосуют друг за друга: средние показатели голосования на «Евровидении» за последние 20 лет

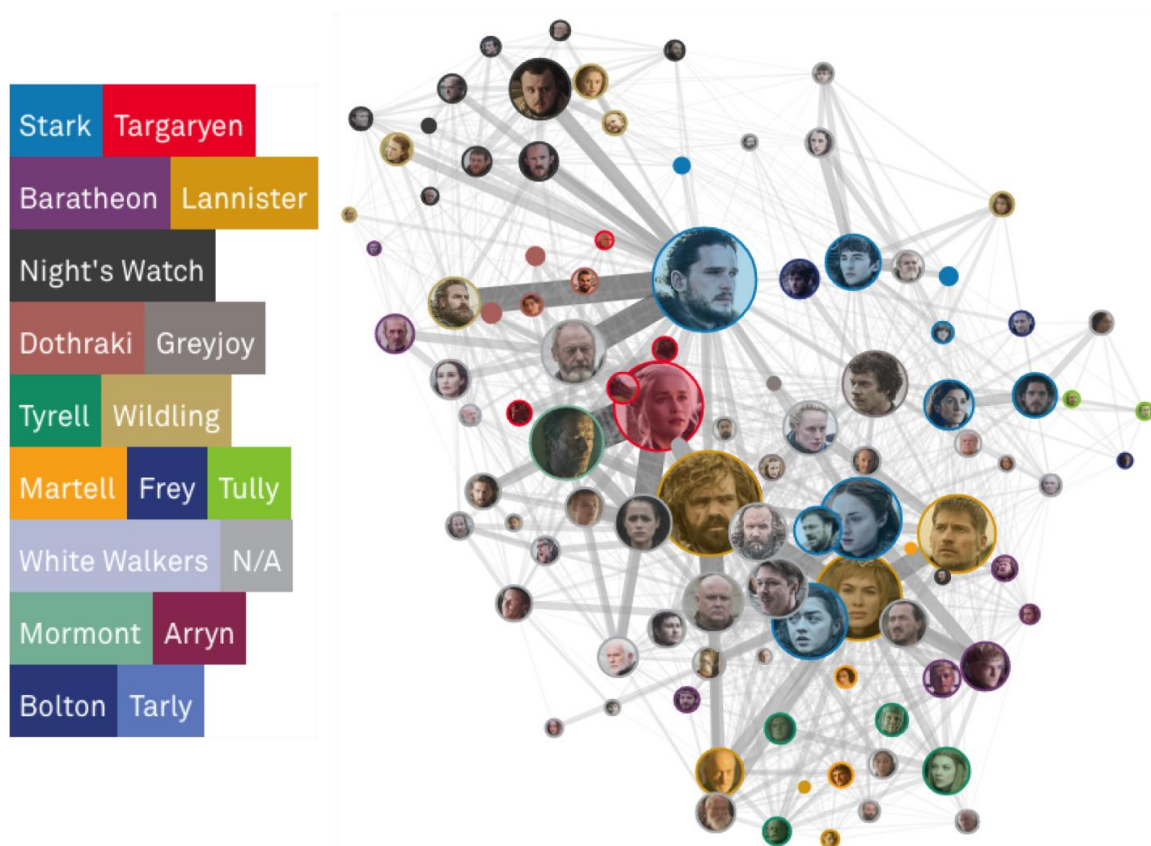


Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Источник: [45]

18) **Граф** (сетевая диаграмма) показывает, как наблюдения (которые именуются узлы или вершины) связаны между собой. Эти связи отражены с помощью ребер, которые соединяют вершины. Вершины обычно изображаются в виде точек, но также могут использоваться иконки. Разные вершины могут быть разного размера (например, в зависимости от числа связей других элементов с этой вершиной). Соединительные линии могут быть разной толщины или цвета, чтобы отразить параметры связи. Граф позволяет на основе анализа связей проанализировать структуру и выявить кластеры вершин. Графы позволяют анализировать небольшое количество взаимосвязей. По мере роста числа вершин и связей граф превращается в клубок ниток.

Рисунок 33. Социальная сеть «Игра престолов»

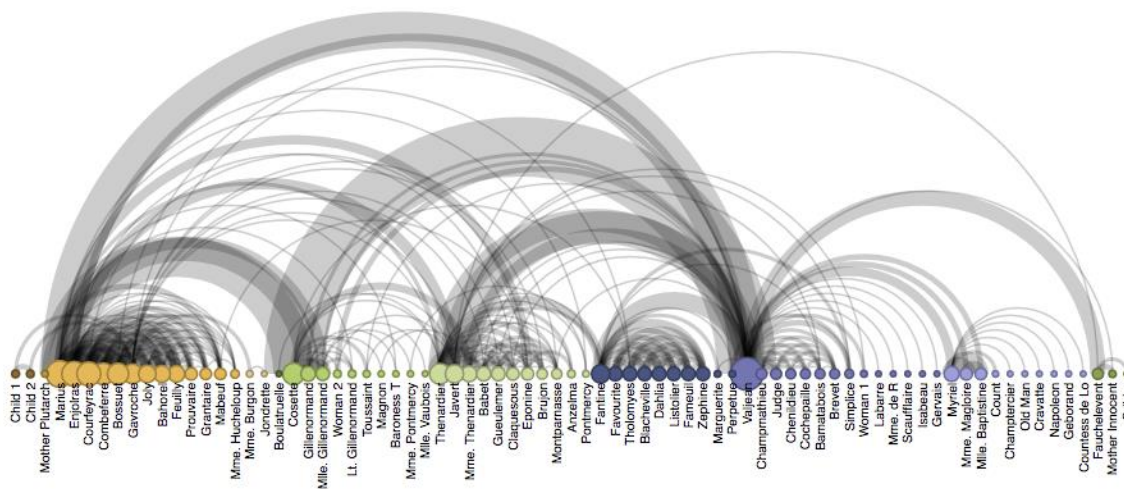


Источник: [46]

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

19) **Арочная диаграмма** – это одномерный граф (сетевая диаграмма), у которого вершины расположены вдоль одной оси, а ребра, соединяющие вершины имеют форму арок. Она позволяет изобразить граф на плоскости, отразив при этом корректно связи вершин. Другими словами, данная диаграмма используется для изображения процессов или объектов, имеющих связи между собой. У данной диаграммы очень низкая точность при большом количестве данных.

Рисунок 34. Совместное появление в главах героев романа «Отверженные» Виктора Гюго

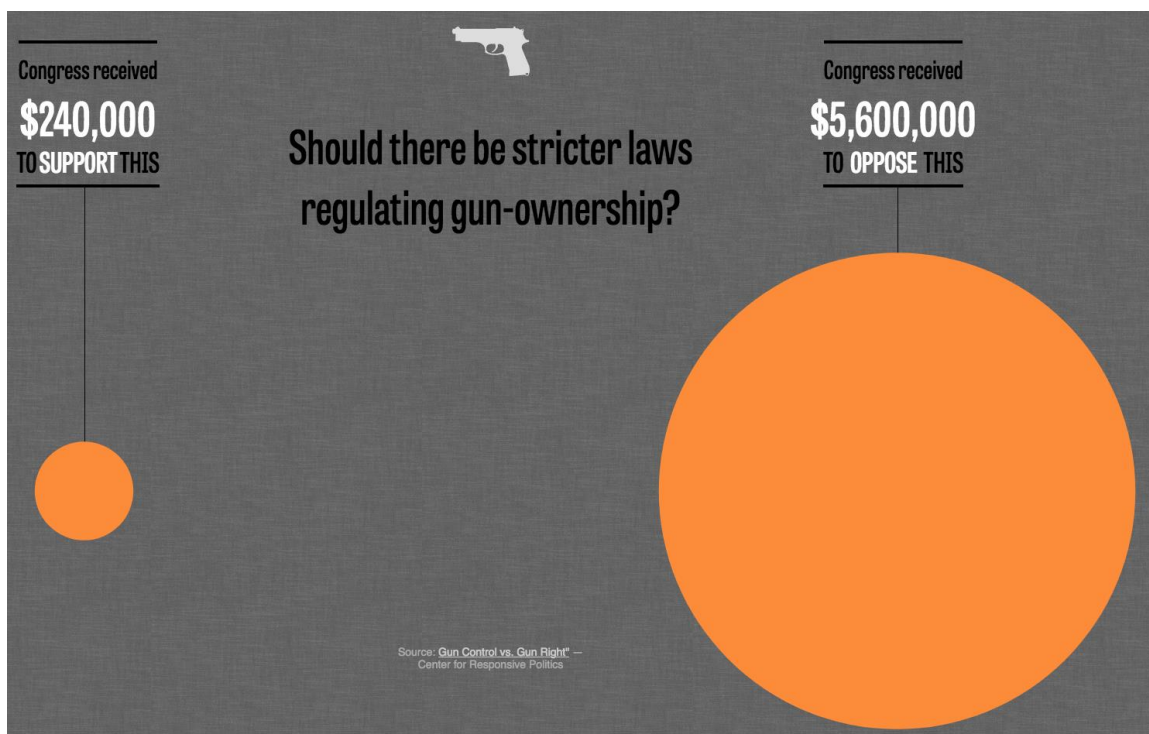


Источник: [47]

20) **Диаграмма соотношения площадей** используется для сравнения значений и визуального отображения пропорций этих значений. Она позволяет быстро сориентировать читателя в том, как соотносятся данные без использования шкал. Чаще всего для площадного сопоставления используются круги или квадраты, но на самом деле может использоваться абсолютно любая фигура. Только Вам нужно иметь в виду, что чем сложнее форма, тем больше трудностей у читателя вызывает сопоставление размеров. Важно также понимать, что сопоставляется площадь фигур, а не длина/ширина какой-то из сторон. Это довольно распространенная ошибка при интерпретации площадных диаграмм. Недостаток данной диаграммы заключается в том, что точно оценить значение показателя или разницу между двумя показателями на основе площадных диаграмм невозможно. Вы не можете использовать эту диаграмму в аналитических целях, только в иллюстративных [48].

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Рисунок 35. Должны ли быть ужесточены законы, регулирующие владение оружием



Источник: [49]

21) **Диаграмма Тримап** – альтернатива круговой диаграмме, позволяет помимо размера каждой категории отразить иерархичную структуру данных. На диаграмме тримап каждая категория отображается в виде прямоугольника и ее площадь соответствует ее вкладу в целое. Тримап подходит для визуализации покомпонентного сравнения. Особенно актуально использовать этот тип диаграммы в ситуации, если категорий много, и, соответственно, визуализировать их с помощью круговой диаграммы нельзя. Также только благодаря тримап

Материалы к [вебинару](#) 22.01.2018 [«Инфографика и визуализация данных»](#)

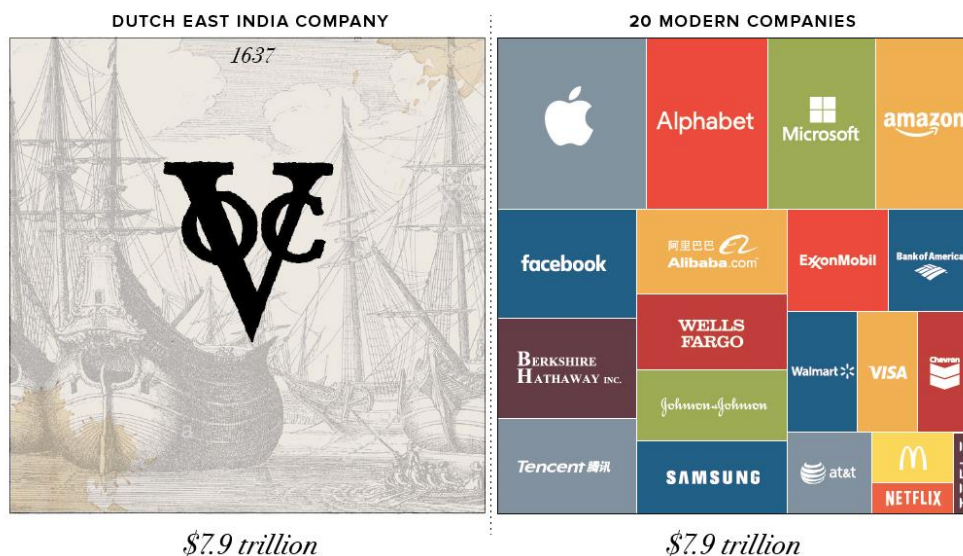
можно отразить многоуровневую структуру, когда внутри категорий есть подкатегории. В результате тримап позволяет компактно отразить структуру и сравнить между собой категории и подкатегории на основе площади соответствующих прямоугольных сегментов. Впервые трипам применил американский информатик Бен Шнейдерман (Ben Shneiderman, 1947) для того, чтобы проиллюстрировать иерархичную структуру папок на компьютере более компактно, чем с помощью диаграммы в виде дерева [50].

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Рисунок 36. Самые высоко оцениваемые компании в истории

THE MOST VALUABLE COMPANIES OF ALL-TIME

Modern juggernauts like Apple and Amazon don't even come close

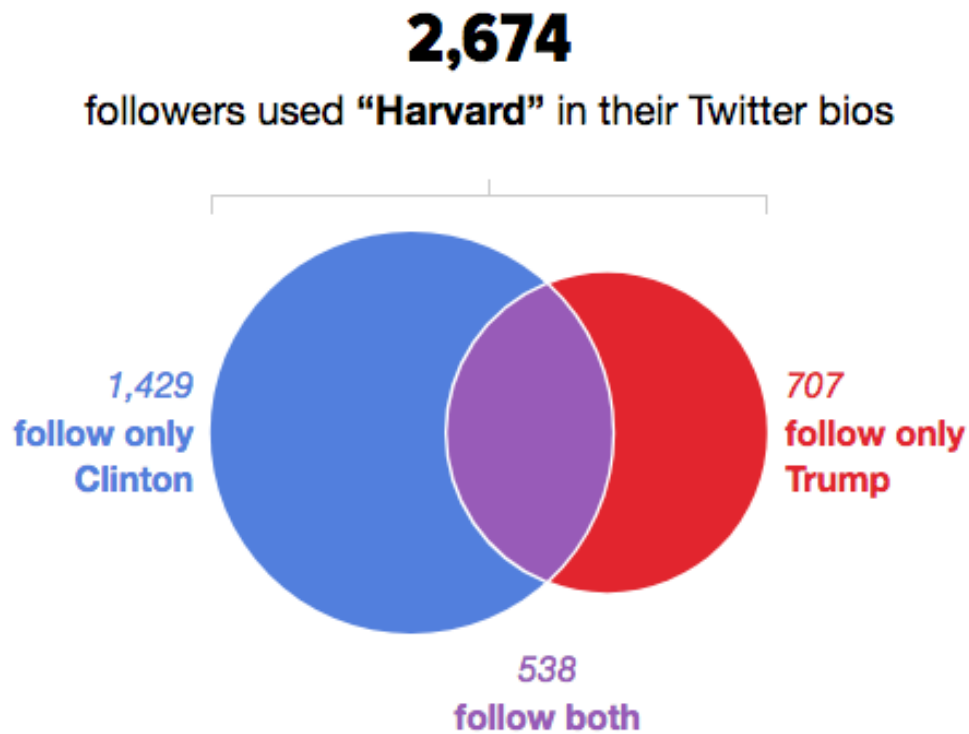


Источник: [51]

22) **Диаграмма Венна-Эйлера** иллюстрирует связи между набором множеств. Каждый набор, как правило, представлены в виде круга. Содержимое в каждом наборе представляет собой совокупность объектов или объектов, которые все имеют что-то общее. Когда наборы пересекаются, это известно как область пересечения. Объекты, попадающие в область пересечения обладают признаками как одного, так и второго недобора данных.

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Рисунок 37. Среди поклонников Клинтон больше пользователей твиттер, связанных с Гарвардом

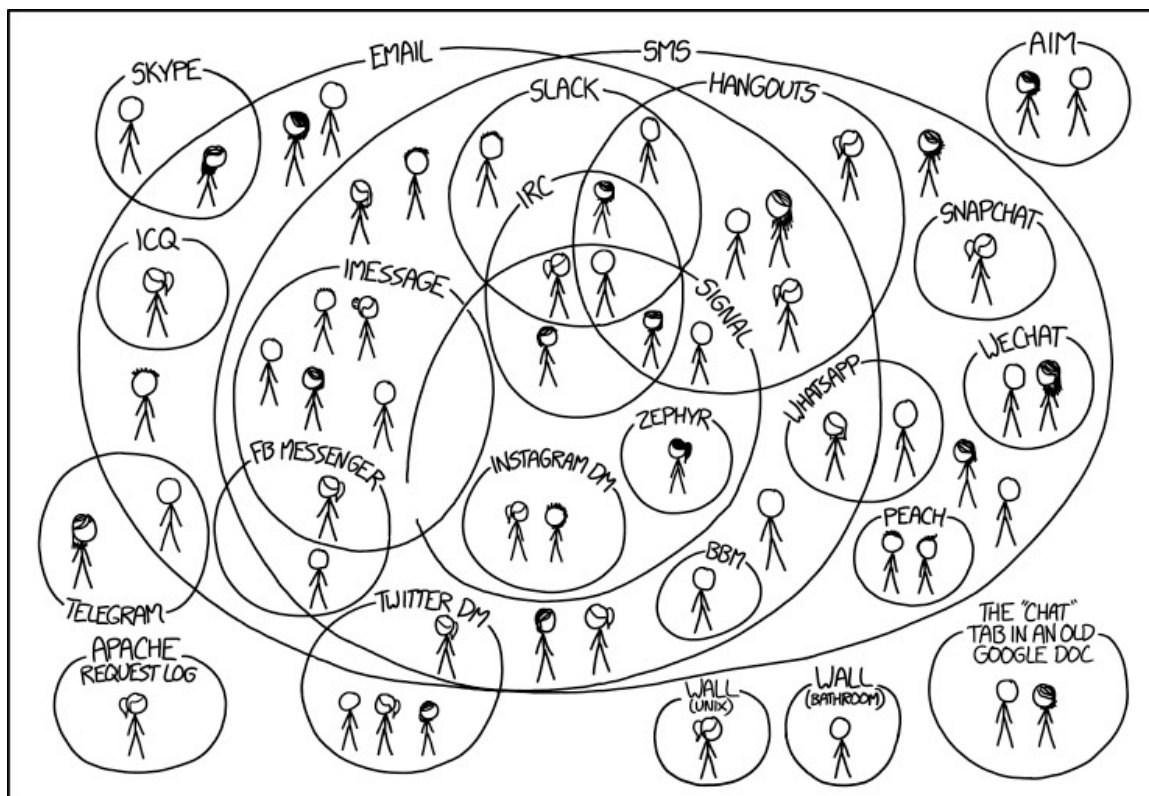


Источник: [52]

На Рисунке 38 представлена диаграмма Венна, нарисованная известным автором комиксов и инфографики в стиле комиксов xkcd. На этой шуточной диаграмме показаны пользователями с одной или несколькими системами обмена сообщений. Идея автора заключается в том, что обладание большим количеством способов связаться с людьми не обязательно облегчает вам общение с ними.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Рисунок 38. Система чатов



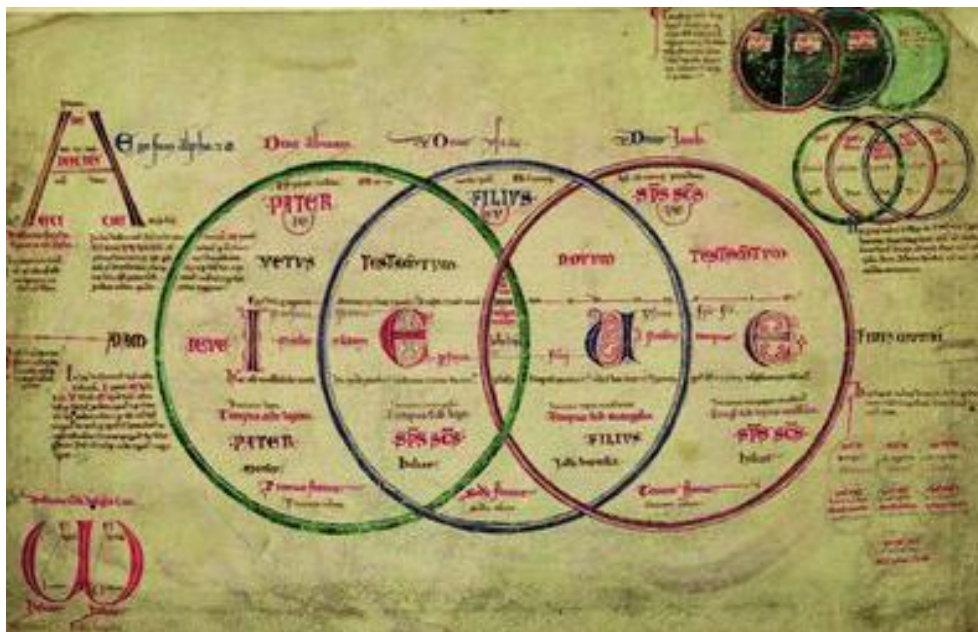
I HAVE A HARD TIME KEEPING TRACK OF WHICH CONTACTS USE WHICH CHAT SYSTEMS.

Источник: [53]

Впервые диаграмма Вена-Эйлера встречается в работе «Книга фигур» средневекового итальянского богослова и мистика Иоахима Флорского (Giacchino da Fiore, 1132-1202).

Материалы к [вебинару](#) 22.01.2018 [«Инфографика и визуализация данных»](#)

Рисунок 39. Тринитарные круги



Источник: [54]

Основные правила хорошей инфографики

Начнем с нескольких правил, которые приводит в своей книге Джин Желязны.

1) Главное из этих правил: **лучше меньше**. Во-первых, подготовка хорошей инфографики требует больших усилий. Так что следует задуматься, насколько окупятся Ваши усилия, какое оптимальное количество диаграмм донесет до читателя Вашу идею. Во-вторых, чем больше диаграмм используется в тексте или презентации, тем хуже их запоминают. Самые известные работы в мире инфографики запомнились по одному ключевому изображению.

2) **Лучше упростить**. Не пытайтесь вместить в диаграмму все данные, которыми Вы располагаете. Задайте вопрос: что именно надо передать в этой диаграмме? какие данные достаточны для того, чтобы передать Вашу основную идею? Отбросьте все остальные данные. Если они очень интересные, то добавьте их где-то на другом уровне (меньшим шрифтом или в стороне от основных данных) для самых любопытных читателей. Не подтверждайте один и тот же тезис двумя разными наборами данных: читатель не идиот и понимает чаще всего с

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

одного раза. Используйте принцип достаточного минимума, так как перегруженная данными инфографика может запутать читателя, привести его к неверному выводу или просто не запомниться.

3) **Лучше показать больше.** Иногда подробности важны. В этой ситуации часто всю информацию по одной теме стараются впихнуть в один слайд (если мы говорим о презентации и раздаточном материале). Этот подход очень популярен в консалтинговых компаниях. Но, либо это нужно делать очень талантливо, чтобы получилась «история» из данных, которую читателю захочется изучить, а не плюнуть на эту кучу наваленной информации и перевернуть страницу, либо разнести идеи на разные слайды/страницы. Прочитируем Джина Желязны: «Презентация пяти идей на одном слайде и одной идеи на пяти слайдах занимает одинаковое время» [Желязны Дж., 2016, стр. 194]. Так что, если вы не можете отказаться от подробностей, то учитывайте все условия, чтобы принять корректное решение по форме визуализации. Будут ли Ваши материалы читать или Вы будете их рассказывать с презентацией на проекторе и т.п. Разнесение информации на несколько слайдов как минимум позволит Вам использовать читаемый кегель, вместо 7ого, которым Вы набивали все идеи на один слайд.

4) **Лучше сделать по-другому.** Если Вы все сделали по правилам, а Ваша фокус-группа все равно не понимает основной идеи, то сделайте по-другому, прислушайтесь к мнению коллег. Либо измените диаграмму, либо саму идею и способ сопоставления данных.

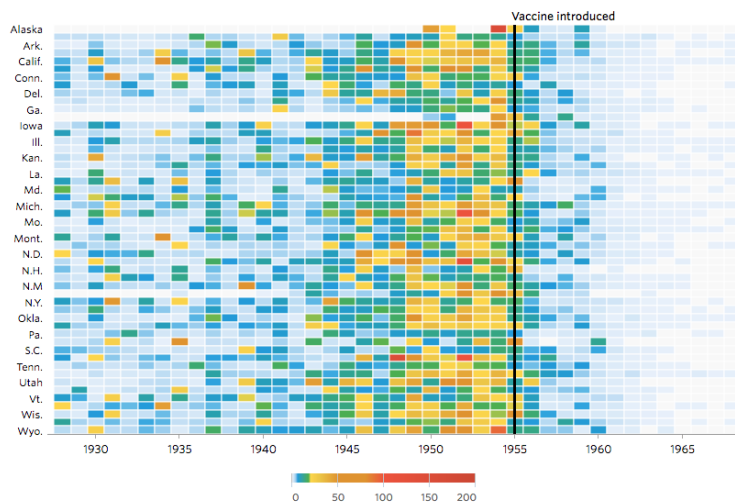
5) **Лучше проявить изобретательность** (а иногда наоборот). Бесспорно, оригинальное решение в подаче визуальной информации сделает вашу идею запоминающейся. Например, в самом начале мы уже рассматривали инфографику из статьи «Борьба с инфекционными заболеваниями в 20-м веке: влияние вакцин», опубликованной в Walt Street Journal [55]. Авторы решили использовать оригинальный способ подачи данных с помощью тепловой карты, хотя вполне могли бы обойтись линейным графиком, о чем в своей критике к инфографике WSJ пишет Рэнди Олсон [56]. На Рисунках 40 и 41 представлены оба варианта визуализации. Именно оригинальная подача данных сделала эту инфографику победителем престижной награды и известной в мире инфографики. Но иногда Вашу идею можно наиболее

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

эффективно передать с помощью простого текстового сообщения. В погоне за оригинальностью не усложните себе жизнь напрасно.

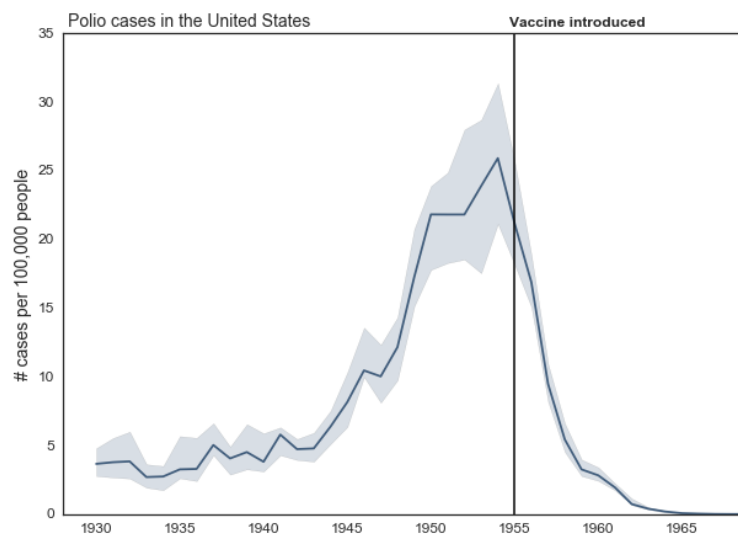
Рисунок 40: Число случаев заболеваний полиомиелитом на 100 000 человек населения в США с 1928 по 1968 год (тепловая карта)

Polio



Источник: [55]

Рисунок 41: Число случаев заболеваний полиомиелитом на 100 000 человек населения в США с 1928 по 1968 год (линейный график)



Data source: Project TYCHO (tycho.pitt.edu) | Author: Randy Olson (randalolson.com / @randal_olson)

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Источник: [56]

Рассмотрим другие правила хорошей инфографики [1].

6) **Заголовок должен все рассказать.** Сформулируйте свою основную идею и вынесите ее в заголовок. Пусть в нем будет отражено все, что вы хотите показать, а диаграмма просто подтвердит ваши слова. Не заставляйте читателя самого анализировать и делать собственные выводы, если вы хотите быть уверены, что он поймет именно то, что Вы хотели ему донести.

7) **Легенда к графику должна быть.** Легенда – это интрига, которая привлекает к себе внимание читателя. Она должна быть, иначе без нее читатель не поймет, что именно Вы изобразили. Она должна быть понятной по той же самой причине. Лучше всего ставить легенду под названием, но перед самой диаграммой, чтобы читатель сначала изучил, как именно ему читать Вашу диаграмму, а потом уже перешел к изучению данных. Однако легенда может находиться и в других местах, что Вы можете видеть на многих примерах, представленных в данном тексте. Иногда легенду можно внести в название диаграммы, выделив ключевые слова заголовка цвета, которые соответствуют данным на графике.

8) **Шкала играет роль.** В зависимости от шкалы одни и те же данные могут выглядеть стабильно или колебаться и производить совершенно другое впечатление. Для обозначения количественных величин можно использовать шкалу, а можно отказаться от нее и подписать значения. Лучше не делать и того, и того сразу, так как получится слишком много информации. Если важную роль играют значения всех категорий, то следует подписать их на линейках (при учете, что значений не много). Если важно одно значение на фоне остальных, то можно использовать шкалу. При желании можно дополнительно подписать важное значение на выделенной цветом линейке.

Нельзя отрезать низ шкалы (значения должны идти от нуля). Исключения возможны только для линейного графика, но тогда надо обязательно отметить, что шкала начинается со значения, отличного от нуля.

В графиках с временным сравнением временная шкала должна идти слева направо: более старые данные в левой части диаграммы, а более новые данные – в правой. Именно так большинство людей воспринимают течение времени.

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

9) **Качество инфографики определяется качеством данных.** Поэтому для хорошей инфографики у Вас должны быть релевантные, достоверные и полные данные. При это как абсолютные, так и относительные данные не всегда дают нам правильное понимание картины.

10) **Округляйте цифры и повышайте порядки.** 38118093727725 тысяч рублей составляет страховая сумма по договорам страхования ответственности перевозчика за причинение вреда жизни, здоровью и имуществу пассажиров, действовавшим на конец 2016 года согласно статистическим показателям, которые публикует Банк России о страховании, за 2016 год [57]. Вряд ли кто-то с ходу может разобраться, что это за цифра. Для того, чтобы облегчить жизнь читателю, необходимо делать следующее. Во-первых, порядки следует отделять пробелами: 38 118 093 727 725 тысяч рублей. Во-вторых, порядок представляемой читателю суммы следует повысить до оптимального для чтения: 38 118 триллионов рублей, то есть 38 квадрильонов рублей. Между прочим, эта сумма составляет 443,0088010834% от ВВП Российской Федерации в 2016 году. Но для удобства читателя, мы округлим значение процентов до 443%.

11) **Форма имеет значение.** Алексей Новичков на своем курсе «Основы инфографики и презентации данных» (совместно с Надеждой Андриановой) [1] любит повторять: «хочешь сделать хорошо – нарисуй круг». **Круг** – это наиболее универсальная форма для привлечения внимания и создания приятного эстетического впечатления у читателя. Однако в зависимости от ситуации, использование элементов различных форм может помочь Вам добиться необходимого эффекта. **Треугольник** – форма, требующая внимания. Ассоциируется с острым и опасным: зуб, шип. **Стрелочки** заставляют глаза читателя бегать по указанным направлениям. Если стрелочки использованы удачно, они помогут читателю изучить визуальную историю целиком, совершить увлекательное путешествие по вашим данным. **Фигура человека** вызывает эмпатию у читающего, он начинает себя ассоциировать с героем Вашей инфографики. Добавление к этой фигуре элементов, связанные с темой инфографики, делает ее еще более запоминающейся.

Работая с формами, необходимо быть осторожным при использовании трехмерных диаграмм. Они часто искажают восприятие данных. Это хорошо показано на Рисунке 42.

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

Рисунок 42: Число недобросовестных рекламных роликов у ведущих сотовых операторов



Источник: [58]

На первый взгляд видно, что оператор Зеленого цвета – самый недобросовестный рекламодатель: какой огромный зеленый сектор! Однако, если посмотреть на диаграмму внимательно и обратить внимание на цифры, то мы с удивлением обнаружим, что число недобросовестных роликов у оператора Красного цвета прилично опережает показатель Зеленого оператора. Трехмерный график визуально искажает доли.

Материалы к [вебинару](#) 22.01.2018 [«Инфографика и визуализация данных»](#)

Полезные ссылки

Сервисы для создания инфографики:

<https://infogram.com/>

<https://public.tableau.com/>

<https://www.canva.com/>

<https://venngage.com/>

<https://caco.com/>

<https://creatly.com/>

<https://gephi.org/>

Виды графиков:

<https://datavizcatalogue.com/index.html>

Источники вдохновения:

<https://www.informationisbeautifulawards.com/showcase?type=awards>

<https://ria.ru/infografika/>

<http://tass.ru/infographics>

<http://infographics.tass.ru/>

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

Библиография

1. Андрианова Н., Новичков А., «Основы инфографики и презентации данных», доступно по ссылке: http://design.hse.ru/dop/programs/1/program_instances/42
2. McCandless D., «Красота визуализации данных», выступление на TED, доступно по ссылке: https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization?language=ru
3. Беляев К., Личный блог, 2015, доступно по ссылке <http://kirillbelyaev.com/all/main-difference-between-visualization-and-infographic/>
4. Friedman D., DeBold T., The Wall Street Journal, 11 февраля 2015, доступно по ссылке <http://graphics.wsj.com/infectious-diseases-and-vaccines/>
5. Dear Data, доступно по ссылке: <http://www.dear-data.com/>
6. Желязны Дж., Говори на языке диаграмм: пособие по визуальным коммуникациям, — М.: Манн? Иванов и Фербер 2016
7. Динамика цен на нефть в 2017 г. Досье, ТАСС, доступно по ссылке: <http://tass.ru/info/4860528>
8. А.А. Аузан и соавторы, «Культурные факторы модернизации», Москва, 2011
9. <http://www.visualcapitalist.com/wp-content/uploads/2017/09/share-of-gdp-share.jpg>
10. Tufte E., Slopegraphs for comparing gradients: Slopegraph theory and practice, доступно по ссылке: https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0003nk&topic_id=1
11. Park Ch., Edward Tufte’s “Slopegraphs”, 2011, доступно по адресу <http://charliepark.org/slopegraphs/>
12. Tufte Ed., The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press; 1983; p. 158

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

13. Tufte Ed., Beautiful Evidence, 2006. P.174, 176.
14. <http://www.oliveruberti.com/infographics/ewpc0ylkniugxuulgm8o0zy8f4jvft>
15. Rasiel E.M., The McKinsey Way, p. 117-118
16. Adamiecki K., Harmonograf (Harmonograful), Przegląd Organizacji, Nr. 4, Warszawa, trad. Autorilor, 1931
17. Bart J. Debicki "Forgotten contributions to scientific management: work and ideas of Karol Adamiecki", Journal of Management History, Vol. 21 Issue: 1, pp.40-67, 2015, доступно по ссылке: <https://doi.org/10.1108/JMH-09-2013-0043>
18. Marsh, Ed. R., The Harmonogram of Karol Adamiecki, The Academy of Management Journal, Vol. 18, No. 26 1975
19. Friendly M., A brief History of Data Visualization, Handbook of Computational Statistics on Data Visualization 2006, доступно по ссылке: <https://www.researchgate.net/publication/226400313>
20. Stream graph, The Data Visualisation Catalogue, доступно по ссылке: https://datavizcatalogue.com/methods/stream_graph.html
21. Sankey diagram, The Data Visualisation Catalogue, доступно по ссылке: https://en.wikipedia.org/wiki/Sankey_diagram
22. Slocum T.A., Thematic cartography and visualization. Prentice Hall. New Jersey. 1999
23. Dent B.D., Cartography: Thematic map design. McGraw-Hill. New York. 1999
24. Carte figurative des pertes successives en hommes de l'armée française dans la Campagne de Russie 1812-13 (comparées à celle d'Hannibal durant la 2ème Guerre Punique) / Minard. Régnier & Dourdet, 1896, Bibliothèque nationale de France, département Cartes et plans, GE DON-4182, доступно по ссылке: <http://gallica.bnf.fr/ark:/12148/btv1b52504201x/f1.item>

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

25. Когда пришел наполеон, ТАСС, доступно по ссылке: <https://1812.tass.ru/>

26. Chord Diagram, Wikipedia – The Free Encyclopedia, доступно по ссылке:

https://en.wikipedia.org/wiki/Chord_diagram

27. Chord diagram, The Data Visualisation Catalogue, доступно по ссылке:

https://datavizcatalogue.com/methods/chord_diagram.html

28. Структура госдолга США, Инфографика РИАНовости, доступно по ссылке:

<https://ria.ru/infografika/20170912/1504466667.html>

29. Dataviz.ru, доступно по ссылке: https://www.instagram.com/p/BeIp_O1DHwI/

30. История и перспективы Европейского союза, Инфографика РИАНовости, доступно по

ссылке: <https://ria.ru/infografika/20170207/1487330737.html>

31. Безвизовый режим между Евросоюзом и Украиной, Инфографика РИАНовости, доступно

по ссылке: <https://ria.ru/infografika/20170609/1496134072.html>

32. Щербакова Е., 38% населения мира переживают третью фазу демографического перехода, на которой рождаемость приближается к уровню простого воспроизводства. Пример Индии, Демоскоп weekly № 481 – 482, октябрь 2011, доступно по ссылке:

<http://www.demoscope.ru/weekly/2011/0481/barom06.php>

33. Pie chart, The Data Visualisation Catalogue, доступно по ссылке:

https://datavizcatalogue.com/methods/pie_chart.html

34. Donut Chart, The Data Visualisation Catalogue, доступно по ссылке:

https://datavizcatalogue.com/methods/donut_chart.html

35. Taschow U., Nicole Oresme und der Frühling der Moderne, доступно по ссылке: <http://nicole-oresme.com/seiten/oresme-biography.html>

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

36. Collins K., America's most common Christmas-related injuries, Quartz, доступно по ссылке:
<https://www.informationisbeautifulawards.com/showcase/1844-america-s-most-common-christmas-related-injuries-in-charts>

37. Диаграмма водопад (waterfall chart) в Excel, Exceltip, доступно по ссылке:
<http://exceltip.ru/%D0%B4%D0%B8%D0%B0%D0%B3%D1%80%D0%B0%D0%BC%D0%BC%D0%B0-%D0%B2%D0%BE%D0%B4%D0%BE%D0%BF%D0%B0%D0%B4-waterfall-chart-%D0%B2-excel/>

38. Project Work Scheduling using Gantt Chart, SlideModel, доступно по ссылке:
<https://slidemodel.com/templates/project-management-gantt-chart-powerpoint-template/project-work-scheduling-using-gantt-chart/>

39. Kim N.W., et al., Visualizing Nonlinear Narratives with Story Curves,
<http://storycurve.namwkim.org/>

40. Chalabi M., доступно по ссылке: <https://www.instagram.com/p/BQS0KNRA2yK/>

41. OkCupid, Undressed: What's the Deal With the Age Gap in Relationships?, доступно по ссылке: <https://theblog.okcupid.com/undressed-whats-the-deal-with-the-age-gap-in-relationships-3143a2ca5178>

42. Aschwanden Chr., King R., Hack Your Way To Scientific Glory, доступно по ссылке:
<https://fivethirtyeight.com/features/science-isnt-broken/#part1>

43. Pellegrini V., NUCLEAR SLOWDOWN – Wired UK, доступно по ссылке:
<https://www.behance.net/gallery/52825037/NUCLEAR-SLOWDOWN-Wired-UK>

44. Chatham House, <https://resourcetrade.earth/data>

45. Золотарев Д., Логинова Е., Никонец С., Шоу-голос. Кто за кого голосует на «Евровидении»: интерактивная инфографика «Ленты.ру», доступно по ссылке:
<https://lenta.ru/articles/2014/05/10/eurovision/>

Материалы к [вебинару 22.01.2018 «Инфографика и визуализация данных»](#)

46. Petronio L., Data Visualization Designer, Communications Department in External Affairs, Northeastern University, The 'Game of Thrones' Social Network, доступно по ссылке:
<https://news.northeastern.edu/2017/08/the-game-of-thrones-social-network/>
47. Bostock M., Arc diagrams, доступно по ссылке: <http://mbostock.github.io/protovis/ex/arc.html>
48. Proportional Area Charts, The Data Visualisation Catalogue, доступно по ссылке:
https://datavizcatalogue.com/methods/area_chart.html
49. Schu T., Money Wins Elections, доступно по ссылке: <http://letsfreecongress.org/>
50. Treemap, The Data Visualisation Catalogue, доступно по ссылке:
<https://datavizcatalogue.com/methods/treemap.html>
51. Desjaerdins J., The Most Valuable Companies of All-Time, доступно по ссылке:
<http://www.visualcapitalist.com/most-valuable-companies-all-time/>
52. Mathias Chr., Storey S., Hooper A., Huff Post, We the Tweeples, The Huffington Post, доступно по ссылке: <http://data.huffingtonpost.com/2016/we-the-tweeps/Harvard>
53. Munroe R., XKCD, доступно по ссылке: <https://xkcd.com/1810/>
54. Флорский И., «Книга фигур», XIII в. (Библиотека колледжа Корпус Кристи, Оксфорд. Ms. 255A. Fol. 7v)
55. Friedman D., DeBold T., The Wall Street Journal, 11 февраля 2015, доступно по ссылке
<http://graphics.wsj.com/infectious-diseases-and-vaccines/>
56. Olson R., Revisiting the vaccine visualisations, Significance, март 2016 доступно по ссылке
<https://www.statslife.org.uk/health-medicine/2717-revisiting-the-vaccine-visualizations>
57. Статистические показатели и информация об отдельных субъектах страхового дела, Центральный Банк Российской Федерации, доступно по ссылке:
http://www.cbr.ru/finmarket/supervision/sv_insurance/

Материалы к [вебинару](#) 22.01.2018 «Инфографика и визуализация данных»

58. Назван худший оператор России, Akket.com, доступно по ссылке:

<https://akket.com/raznoe/68089-nazvan-hudshij-operator-rossii-kotoryj-postoyanno-durit-svoih-abonentov.html>