

# Добро пожаловать в визуализацию данных!

---

## **В этой главе**

- ✓ Что такое визуализация данных.
- ✓ Чем является и чем не является эта книга.
- ✓ Знание своей аудитории.
- ✓ Примеры из истории.
- ✓ Популярные современные инструменты для создания визуализаций.

Визуализация данных повсюду. Иногда она подкрадывается к вам в самых неожиданных местах. Играли ли вы последнее время в видеоигры? Наверняка там была какая-нибудь визуализация, показывающая оставшийся ресурс персонажа, ваш прогресс в достижении цели или представляющая собой карту с указанием всех посещенных вами мест.

Визуализация может быть не только интересной, но и очень важной для успеха отдельного человека, организации или общества в целом. В Англии XIX века Флоренс

Найтингейл перевернула мир медицины, применив визуализацию данных, которые собирала на протяжении десятилетий, когда работала в военных госпиталях. Она представляла свои идеи сильным мира сего, а они отвечали: «Мы так не делаем». Ей удалось что-то изменить, лишь показав свою визуализацию.

Хорошо продуманные визуализации спасают жизни и способны перевернуть целые отрасли. И они же могут сказать, какой ресурс остался у вашего маленького красного водопроводчика, прежде чем придется начать игру заново. :)

В главе 1 вы сначала получите общее представление о визуализации данных. Затем будет краткое введение, где сказано, что знание аудитории — главное, о чем следует помнить при создании любой визуализации. Далее мы рассмотрим несколько примеров визуализации данных в историческом контексте, а потом поговорим о некоторых самых популярных инструментах, используемых сегодня в этой сфере.

## 1.1. ЧТО ТАКОЕ ВИЗУАЛИЗАЦИЯ ДАННЫХ

Если вы взяли эту книгу в руки добровольно, велика вероятность, что вы уже имеете представление о визуализации данных. Или, возможно, находитесь в приемной у врача, и это единственное средство спасения от скуки (я бы хотела познакомиться с таким врачом!). Так или иначе, я дам такое определение: визуализация данных — это способ передачи числовой либо количественной информации в визуальной форме. Когда я рассказываю людям о том, чем занимаюсь, и упоминаю визуализацию данных, у них на лице проявляется недоумение. Поэтому я обычно упрощаю, говоря, что зарабатываю на жизнь созданием графиков и диаграмм.

В визуализации данных мы используем диаграммы, графики, карты, а иногда даже фотографии и иконографику, чтобы представить цифры визуально. Ведь не зря говорят, что лучше один раз увидеть, чем сто раз услышать, верно? В качестве носителя информации может выступать что угодно: печатная продукция, интерактивная веб-страница, статичная инфографика; доска с записями, сделанными маркерами, или трехмерные объекты из глины; и многое-многое другое. Уверена, вы даже можете станцевать интерпретационный танец, если захотите. Есть разные причины, по которым одна визуализация эффективнее другой. Четких предписаний, как ее создавать, нет — свобода действия здесь подобна бескрайнему небу. Но если вы будете держать мысль о своей аудитории в центре внимания и ставить ее во главу угла процесса, то обязательно добьетесь успеха.

## 1.2. ЧЕГО МОЖНО ЖДАТЬ ОТ ЭТОЙ КНИГИ

Я очень надеюсь, что эта книга станет для вас практическим, доступным и увлекательным руководством по созданию красивых и полезных визуализаций данных, которые наполняют повседневную жизнь энергией и информацией. Однако, прежде чем рассказать о том, что представляет собой книга, я хотела бы отметить, что она не является техническим руководством по установке каких-либо инструментов или изучению языка программирования. Мы не будем разбирать примеры кода, в конце каждой главы и части нет упражнений: пошаговые алгоритмы меня не привлекают. Эта книга — руководство по принципам хорошего дизайнера применительно к визуализации данных. Вы сможете использовать рассмотренные концепции на любом этапе любого проекта по визуализации, над которым работаете, независимо от инструментария, объема или характера среды.

Учитывая вышесказанное, эта книга будет вам полезна, если вы уже работаете над визуализацией данных, будь то личный проект для развлечения (в сообществе визуализации данных полно единомышленников-ботаников, которые с удовольствием тратят свое свободное время на создание графиков и диаграмм), или что-то для работы.

Первая часть книги (главы 1–3) служит фундаментом, на котором строятся две другие части. Поэтому я рекомендую прочесть ее первой (в конце концов, она в самом начале). Во второй части (главы 4–7) мы углубимся в разные аспекты визуализации данных, чтобы понять, как все работает и почему это важно. Будет рассмотрена и практическая сторона, связанная с применением изложенных принципов дизайнера в реальной работе по визуализации. В третьей части (главы 8–9) подробно описан процесс создания визуализации данных от начала до конца, а затем даны рекомендации по устранению проблем и работе с проектами, которые идут не по плану.

Я постаралась написать эту книгу таким образом, чтобы вы получили из нее много полезного, независимо от того, на каком этапе пути в области визуализации находитесь.

## 1.3. ЗНАЙТЕ СВОЮ АУДИТОРИЮ

Самое главное, о чем нужно помнить при создании визуализации, — это о своей аудитории. Все дело в ней! Вы можете сделать очень красивую визуализацию, используя новейшие приемы и самые модные графики, опираясь на лучшие практики, но, если она не отвечает потребностям аудитории, вы зря потратили время. Хорошая новость заключается в том, что аудитория понимает историю, которую вы пытаетесь рассказать, и воспринимает сделанную работу. При этом повторюсь: сколько бы крови, пота и слез вы не пролили при работе над проектом, ничто не имеет значения, если ваша аудитория не понимает результат. Избавьте себя от лишних хлопот и поговорите с людьми как можно раньше, чтобы узнать, что им нужно. И общайтесь в течение всей разработки: важно убедиться, что вы

отвечаете на имеющиеся вопросы и не создаете чересчур сложные вещи, которые аудитория не сумеет быстро понять.

Если говорить подробнее, речь о знании, кто ваша аудитория и кто ею не является. Вы когда-нибудь пробовали пойти поесть всей семьей? Если да, то наверняка с содроганием вспоминаете, как долго решали, какой ресторан или бар почтить своим присутствием. Мама хочет итальянскую кухню, папа — «старое доброе мясо с картошкой», ваши братья и сестры — что-нибудь интересное вроде эфиопской или тайской кухни, а вы сами просто хотите поесть, черт побери. Независимо от сделанного выбора всем не угодишь. Даже если найдется ресторан-единорог, где подают все блюда мира, скорее всего, ни одно из них не будет приготовлено достаточно хорошо, чтобы обошлось без разочарований. Пытаясь угодить всем, в итоге вы не угодите никому.

Сейчас, когда я пишу эту книгу в свободное от работы время — честно говоря, в противном случае его большая часть была бы потрачена на просмотр сериалов и телешоу на моих любимых стриминговых сервисах, — необходимость знать свою аудиторию по-прежнему актуальна. Я думаю, что вы — аналитик или человек, работающий в смежной с аналитикой сфере, который обнаружил, что ему нужна визуализация, но нет дизайнерского ноу-хау для ее создания. Неважно, как долго и упорно я работаю, насколько впечатляющ мой слог и красивы приведенные примеры либо какое количество людей я уговариваю воспроизвести их великолепные визуализации — все это не имеет значения, если гипотетический аналитик ничему не научится. Всегда ставьте потребности аудитории выше собственных!

Теперь, вспомнив обо всем важном, мы совершим небольшое путешествие во времени, чтобы познакомиться с примерами из истории развития визуализации данных, а затем погрузимся в некоторые самые популярные инструменты, используемые сегодня для создания визуализации.

## **1.4. ПРИМЕРЫ ИЗ ИСТОРИИ**

Чтобы избавить вас от урока истории, предлагаю полистать метафорический фотоальбом с избранными новаторскими визуализациями. Ведь знакомство с историей той или иной дисциплины дает представление об имеющихся достижениях и разработках будущего. Многие находят в подобном экскурсе вдохновение для текущей работы.

### **1.4.1. Предыстория визуализации данных**

Несмотря на то что визуализация данных появилась не так давно, как искусство, она имеет такую же долгую историю, как, например, астрономия. Действительно, рисунки в пещере Ласко на юге Франции — это древняя наскальная живопись, относящаяся к эпохе палеолита. По сведениям Мартина Свитмана и Алистера Кумбса, опубликованным в их статье о древних знаниях о предварении равноденствий

(*Athens Journal of History*, 2018), некоторые из этих изображений могли использоваться для шифрования даты «встречи с метеорным потоком Таурид». На рис. 1.1 показана часть сцены из зала «Шахта», где слева изображен носорог и за ним — человек с птичьей головой. Выше нарисована птица, которая может быть уткой или гусем. Справа — бизон или вепрь, а на противоположной стене (на рисунке не показана) — лошадь. Свитман и Кумбс утверждают, что это не просто животные, а созвездия. Так, носорог соответствует созвездию Тельца, бизон/вепрь — Козерогу, утка/гусь — Весам, а лошадь — созвездию Льва.



**Рис. 1.1.** Сцена из зала «Шахта» в пещере Ласко с изображениями: носорога — слева; птицы, которая может быть уткой или гусем, — посередине; бизона — справа. Кумбс и Свитмен утверждают, что это могут быть изображения созвездий, указывающих на определенные даты (Алистер Кумбс. Использовано с разрешения)

Увидеть визуализацию количественных данных можно в *кипу*, или «говорящих узлах» (рис. 1.2). Их делали инки, жившие на территории современного Перу, начиная примерно с 2600 года до нашей эры. «Говорящие узлы» использовались для ведения записей обо всем, начиная с переписи населения и заканчивая налогами.

### 1.4.2. Карты

Самыми ранними *задокументированными* визуализациями, несомненно, являются карты, например Туринская папирусная карта, созданная в Египте и датированная 1150-ми годами до нашей эры (рис. 1.3). На карте длиной около 2,7 м, нарисованной Аменнахтом, изображено место, где находится золото и камень для создания статуй фараона Рамзеса IV.



**Рис. 1.2.** Узлы кипу, которые инки использовали для фиксирования количественных данных, от переписи населения до налогов (Pi3.124, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, Wikimedia Commons)



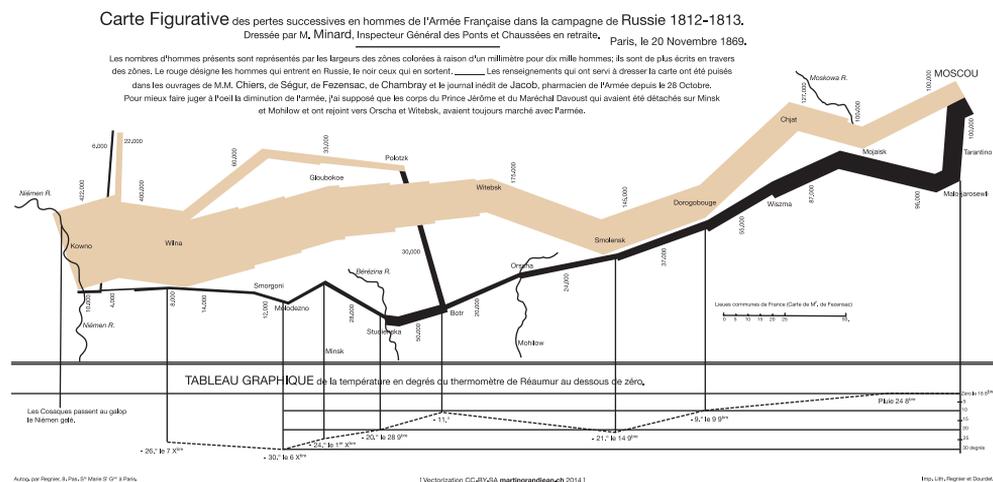
**Рис. 1.3.** Туринская папирусная карта, нарисованная Аменнахтом (между 1156 и 1150 годами до н. э.), на которой изображены дороги, пересекающие восточную часть египетской пустыни, Хаммамет (общественное достояние, Wikimedia Commons)

В течение следующих нескольких тысяч лет самые заметные успехи в области визуализации данных были достигнуты в картографии. Однако в X или XI веке мы снова встречаем свидетельства того, как люди фиксировали астрономическую информацию, в том числе о движении планет. К 1500-м годам картографы, мореплаватели и астрономы (среди прочих) восточной и западной цивилизации регулярно использовали приборы и инструменты для точных измерений географических и астрономических координат, а также других физических величин.

### 1.4.3. Эпоха раннего модерна

Согласно интерактивной хронологии инфографики Р. Дж. Эндрюса, к 1600-м годам человечество вошло в эпоху культовой инфографики. Это восхитительная визуализация данных, поэтому я настоятельно рекомендую вам взглянуть на нее (<http://history.infowetrust.com>).

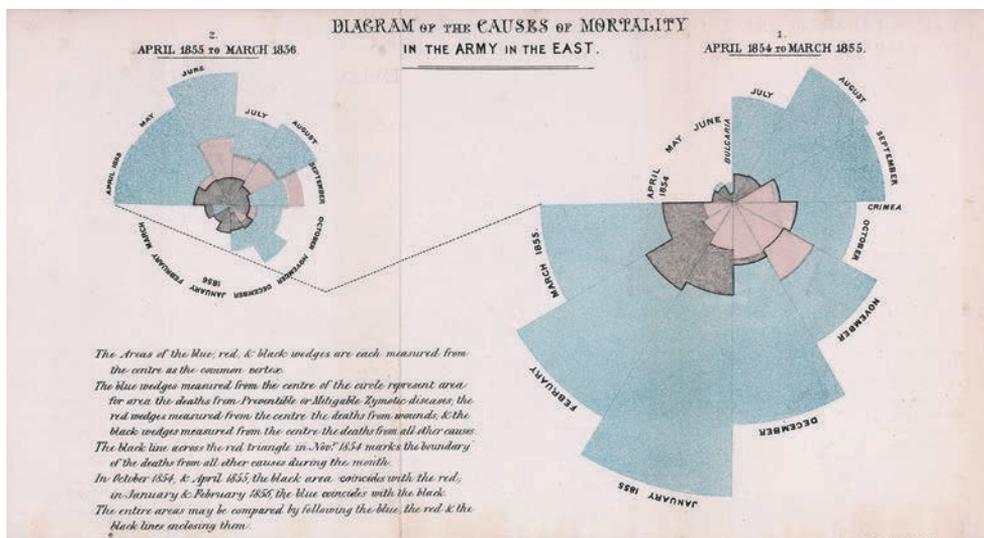
Один из моих любимых примеров в хронологии Эндрюса — Русская кампания Наполеона 1812 года — показан на рис. 1.4. Всего на одной странице искусно отобраны шесть видов данных: численность войск; пройденное расстояние; температурный режим, в котором военным приходилось действовать; географические координаты; направления движения и временные отрезки. По мере совершенствования навыков в области визуализации данных вы сможете по достоинству оценить этот впечатляющий результат визуализации!



**Рис. 1.4.** Самая известная работа топографа Шарля Минара — карта Русской кампании Наполеона 1812 года, где на одной странице отобраны шесть видов данных (Martin Grandjean, CC BY-SA 3.0, общественное достояние, Wikimedia Commons)

### 1.4.4. Флоренс Найтингейл

Возможно, вы слышали о Флоренс Найтингейл — английской аристократке и основательнице современного сестринского дела, которая собирала и записывала данные, опираясь на собственный опыт работы медицинской сестрой во время Крымской войны 1854–1856 годов. Используя данные, полученные в военном госпитале Скутари (Ускюдар, Турция), она создавала визуализацию за визуализацией, чтобы лоббировать улучшение условий лечения. На рис. 1.5 представлена одна из самых известных сравнительных полярных диаграмм, которую придумала эта женщина и которая сегодня известна под названием «Роза Найтингейл». На диаграммах хорошо видно, что основной причиной смерти солдат были не раны (показаны красным цветом), а поддающиеся профилактике болезни (показаны синим цветом). В наши дни подобные визуализации могут посчитать спорными, поскольку, как мы узнаем далее, человеческий глаз плохо интерпретирует углы. Тем не менее, любите вы их или нет, а использование этих углов способствовало большим переменам.



**Рис. 1.5.** Самая известная пара полярных диаграмм «Роза Найтингейл», отображающих причины смерти солдат во время Крымской войны: поддающиеся профилактике болезни — синий цвет, боевые раны — красный (Florence Nightingale, Wikimedia Commons)

Вооружившись своими данными и визуализациями, Найтингейл стала одним из первых пропагандистов мытья рук и санитарии в больницах, задолго до появления микробной теории болезней. Она рассылала информацию и самостоятельно

изданные собственные труды всем сильным мира сего и специалистам — врачам, военным, членам Палаты общин и Палаты лордов, даже королеве Виктории. И так стала движущей силой масштабных перемен, последствия которых мы ощущаем до сих пор.

### 1.4.5. Поздний современный период

Время не стоит на месте, мы тоже — перемещаемся в XX и XXI века. Мне, человеку, зарабатывающему на жизнь в основном созданием интерактивных визуализаций, не дает покоя тот факт, что до 1980-х годов при желании сделать визуализацию данных приходилось доставать (в лучшем случае) ручку и бумагу. Электронные таблицы стали использоваться только после того, как публике представили VisiCalc<sup>1</sup>: в 1979 году — на Apple II, в 1981 году — на IBM PC (да, великое разделение корнями уходит в далекое прошлое). Даже сегодня это самый любимый способ превращения данных в идеи: финансового работника по-прежнему невозможно представить без старой доброй электронной таблицы, разве что вырвать ее из холодных пальцев после смерти.

Раз уж мы затронули тему электронных таблиц, давайте немного поговорим об инструментах для создания визуализации данных.

## 1.5. ИНСТРУМЕНТЫ ДЛЯ ВИЗУАЛИЗАЦИИ ДАННЫХ

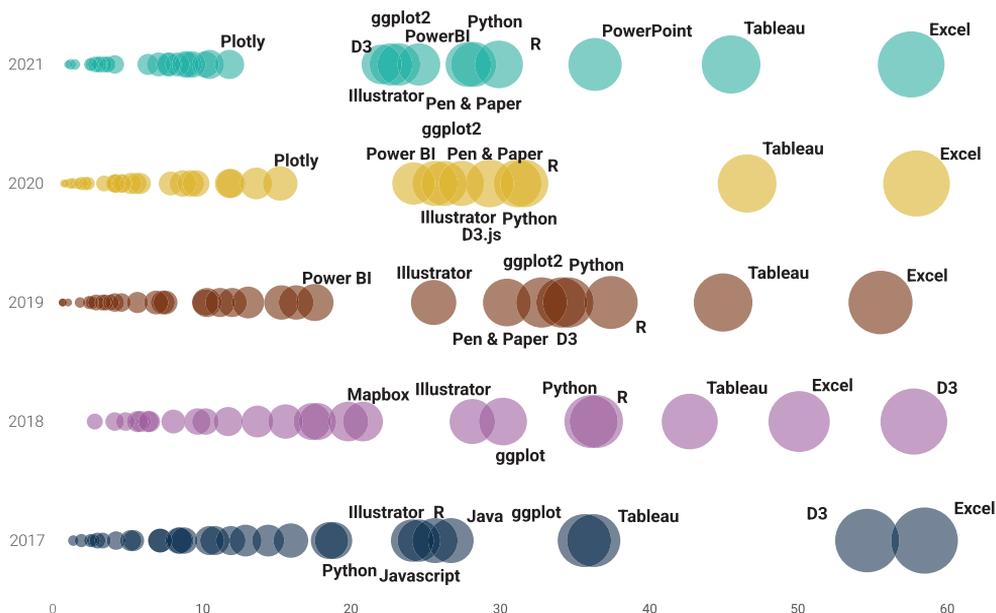
Как мы уже говорили в разделе 1.1, инструментами для визуализации данных может быть практически все, что дает наглядный результат. Хотя лепка из глины и интерпретационный танец тоже подходят, в книге мы сосредоточимся на двухмерных и, как правило, вычислительных средствах.

Общество визуализации данных (Data Visualization Society, DVS) ежегодно проводит исследование состояния отрасли. Один из постоянно задаваемых вопросов касается инструментов, которые респонденты используют для создания визуализаций. На рис. 1.6 показано их процентное соотношение. Каждый инструмент представлен в виде кружка определенного цвета, а процент использующих его респондентов — последовательностью кружков, диаметр которых увеличивается слева направо. Важно отметить, что изменения в рейтинге инструментов из года в год могут быть связаны не только с ростом или падением популярности, но и с типом респондентов, участвовавших в опросе, а также с наличием конкретного инструмента в списке инструментов на выбор.

---

<sup>1</sup> VisiCalc — первая электронная таблица для персональных компьютеров. Считается, что именно VisiCalc превратил микрокомпьютеры из инструмента развлечения и хобби в серьезный инструмент ведения делового оборота. — *Примеч. ред.*

Ничего страшного, если вы не знакомы с инструментами, перечисленными на рис. 1.6. В этом разделе мы рассмотрим некоторые из них. А чтобы вы могли хотя бы в общих чертах понять, что представляет собой этот список, я разбила все инструменты по типам (табл. 1.1).



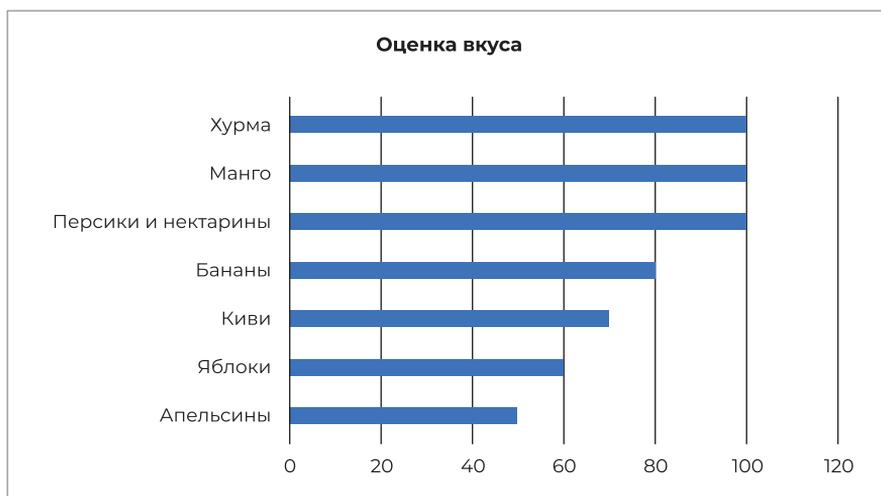
**Рис. 1.6.** Самые популярные инструменты по годам (в процентах от количества опрошенных) согласно ежегодному опросу Общества визуализации данных. Обратите внимание, что появление и рост популярности некоторых инструментов могут быть обусловлены их добавлением в список инструментов на выбор для участников опроса, а не реальным ростом частоты использования (Общество визуализации данных. Использовано с разрешения)

**Таблица 1.1.** Разбивка по типам инструментов, представленных на рис. 1.6

Тип инструмента	Популярные инструменты на рис. 1.6
Электронная таблица	Excel
Инструменты анализа	Tableau, Power BI
Код	D3 (он же D3.js, или d3, или d3.js), Python, R, ggplot, ggplot2, Plotly, Java
Программное обеспечение для дизайна	Illustrator
Другое	Ручка и бумага, Mapbox, PowerPoint

### 1.5.1. Электронные таблицы

Как следует из рис. 1.6, Excel от Microsoft — самый популярный инструмент для визуализации данных (за исключением 2018 года, но и тогда он потерял лишь одну позицию, став вторым). Популярность Excel вполне объяснима: это программа, к которой в той или иной степени имеют доступ почти все специалисты по визуализации данных, а значит, у нее очень низкий барьер входа. Практически ничего не нужно делать, чтобы добавить или ввести некоторые данные. Затем два щелчка кнопкой мыши — и график построен. Полученный результат, представленный на рис. 1.7, может быть не очень красивым, но, тем не менее, это наглядный график.



**Рис. 1.7.** Гистограмма, созданная в Excel с использованием настроек по умолчанию

Хотите более активно участвовать в создании диаграмм? В этом вам помогут программы для работы с электронными таблицами, в том числе выпускаемые не только Microsoft. С их помощью можно вести сложные расчеты для создания графиков, менять заголовки и оси, осуществлять форматирование и создавать легенды, использовать большие и малые данные. Не зря у финансистов такая крепкая хватка: они — Макгайверы<sup>1</sup>, а электронные таблицы — швейцарский армейский нож для анализа данных.

<sup>1</sup> Отсылка к американскому приключенческому сериалу «Секретный агент Макгайвер».