

# Как рассказать историю с помощью данных

1

Подумайте обо всех популярных книгах по визуализации, которые существуют, — о тех, про которые вы постоянно слышите на лекциях или про которые читаете в блогах, а еще о тех, которые пришли вам на ум прямо сейчас, пока вы пробежали глазами по этим строчкам. Что у них всех общего? Они все рассказывают интересные истории. Вероятно, эти истории должны были убедить вас в чем-то. Или, возможно, подтолкнуть вас сделать что-то, познакомить с новой информацией или заставить пересмотреть свои предвзятые представления о реальности. Как бы то ни было, по-настоящему хорошая визуализация данных — неважно, большая или маленькая, созданная из любви к искусству или для презентации — помогает вам понять, о чем именно способны рассказать данные.

## Больше чем числа

---

Взглянем правде в глаза. Данные могут быть чем-то довольно досадным и раздражающим, если вы не знаете, что вы ищете, или не понимаете, есть ли в них нечто такое, что следует искать в первую очередь. Тогда они превращаются в грудку цифр и слов, которые не имеют иного смысла, кроме своего непосредственного значения. В этом и состоит великий смысл статистики и визуализации — в том, что они помогают увидеть, что именно стоит за всем этим. Помните: данные есть отражение реальной жизни. Это не просто грудка каких-то чисел. В этой грудке содержится множество историй. В ней есть и смысл, и правда, и красота. И, как и в реальной жизни, эти истории иногда бывают простыми и искренними, а иногда — сложными и иносказательными. Некоторые истории как будто взяты из учебника. Другие похожи на роман. И от вас — статистиков, программистов, дизайнеров и специалистов в обработке данных — зависит, как рассказать ту или иную историю.

В этом и состоял один из первых уроков, которые я усвоил, начав заниматься в магистратуре. Должен признаться, что до своего поступления туда я думал, что статистика — это чистый анализ, а данные — результат механического процесса сбора и обработки информации. В большинстве случаев так оно и есть. Ну, я же окончил колледж по специальности «электротехника», а потому неудивительно, что данные виделись мне именно в таком свете.

Не поймите меня неправильно. Я не хочу сказать этим ничего плохого. Просто со временем я убедился, что данные, будучи вещью объективной, очень часто имеют при этом и человеческое измерение.

В качестве примера давайте посмотрим еще раз на безработицу. Выдать средние значения по штатам нетрудно, но, как вы сами убедились, положение дел даже в пределах одного штата может сильно колебаться. В двух соседних районах ситуация может существенно различаться. Возможно, среди ваших знакомых есть человек, который потерял работу в последние несколько лет, и, как говорится, он для вас не просто статистическая единица, правильно? Числа представляют собой живых людей, так что вам следует подходить к данным именно так. Вам не нужно рассказывать историю каждого из людей. И все же есть тонкое, но от этого не менее важное различие между увеличением уровня безработицы на 5 процентов и несколькими сотнями тысяч людей, оставшихся не у дел. В первом случае сказанное воспринимается как довольно абстрактная цифра, а во втором данные имеют совсем иную смысловую нагрузку.

## Журналистика

В полной мере все вышесказанное до меня дошло во время стажировки в газете New York Times. Длилась она всего три летних месяца после второго курса магистратуры, но изменила мой подход к данным навсегда. Я не только научился создавать графику для новостей. Я научился преподносить данные как новость, а вместе с тем узнал многое о дизайне, организации и проверке фактов, а также об их поиске и изучении.

Однажды я посвятил день задаче проверки трех чисел из целого массива данных, а все потому, что когда отдел графики New York Times создает некую диаграмму, он должен удостовериться в точности сведений. Только убедившись в надежности полученной информации, мы переходили к созданию презентации. Именно такое внимание к деталям делает графику в этой газете столь хорошей.

Посмотрите на любую схему или диаграмму в New York Times. Данные в ней всегда представлены кратко, ясно и чрезвычайно красиво. Однако что это значит? Когда вы смотрите на инфографику, вы получаете возможность понять данные. Важные моменты или области сопровождаются комментариями; символы и цвета аккуратно объяснены в легенде или в примечаниях; а еще Times делает все для того, чтобы история, содержащаяся в данных, прочитывалась аудиторией легко. Это уже не просто диаграмма или схема. Это графика.

Графика, представленная на рис. 1.1, похожа на ту, которую вы можете найти в New York Times. На ней показано, как увеличивается вероятность того, что вы умрете в ближайший год, в зависимости от вашего возраста.



**Рис. 1.1.** Вероятность смерти с учетом вашего возраста

► Некоторые из лучших диаграмм и графиков New York Times можно посмотреть на странице <http://datafl.ws/nytimes>.

### ПРИМЕЧАНИЕ

Дополнительную информацию о том, как журналисты используют данные при подготовке репортажей о текущих событиях, можно получить из документального фильма Джеффа Макги «Журналистика в эпоху данных» (Geoff McGhee "Journalism in the Age of Data"). В нем вы увидите потрясающие интервью с некоторыми из лучших профессионалов в этой области.

В основе этого рисунка лежит простой линейный график. Однако элементы дизайна помогают лучше рассказать историю. Подписи и примечания поясняют контекст и помогают понять, почему эти данные интересны; а толщина линий и цвет фокусируют глаза на том, что важно.

Требования к дизайнеру графиков и диаграмм не сводятся только к визуализации статистических данных — от него требуется также дать понять, что именно выражает конкретный изображенный элемент.

## Искусство

New York Times — газета объективная. Она излагает вам данные и дает факты. В этом ей нет равных. На другом конце спектра находится визуализация, которая в меньшей степени связана с аналитикой и в большей — с погружением в эмоции. Именно этим занимаются Джонатан Харрис и Сеп Камвар в своей работе «Все у нас хорошо» (Jonathan Harris, Sep Kamvar, We Feel Fine), представленной на рис. 1.2.



Рис. 1.2. «Все у нас хорошо» (Джонатан Харрис, Сеп Камвар)

В их интерактивном творении собраны предложения и фразы из различных персональных блогов, и все они визуализированы как пузырьки, плывущие в пространстве. Каждый пузырек представляет собой эмоцию и окрашен в соответствующий цвет. В целом создается впечатление, как будто все эти люди несутся в космосе, но если вы посмотрите подольше, то заметите, как пузырьки начинают сбиваться вместе. Попробуйте отсортировать и категоризировать их, используя интерфейс, и вы увидите, как эти, казалось бы, произвольно разбросанные крошечные штучки объединяются. Кликните на какой-нибудь пузырек, чтобы узнать чью-то индивидуальную историю. Это так поэтично и одновременно так познавательно!

Можно привести еще много примеров. Скажем, работа Голана Левина «Мусорка» (Golan Levin, The Dumpster). В ней исследуются записи в блогах, в которых говорится о расставаниях со значимым человеком. Или «Sumedicina» Кима Азендорфа (Kim Asendorf), в которой рассказывается выдуманная история мужчины, убегающего от пораженной коррупцией организации, но не словами, а схемами и диаграммами. Или скульптуры Андреаса Николаса Фишера (Andreas Nicolas Fischer), изображающие экономический спад в Соединенных Штатах.

Основная идея заключается в том, что данные и визуализация не обязательно должны передавать лишь голые, холодные факты. Иногда то, что вам необходимо, это вовсе не аналитические прозрения. Порой бывает полезнее рассказать историю с эмоциональной точки зрения, чтобы подтолкнуть зрителей к размышлениям над имеющимися данными. Подумайте над этим. Не всегда кино должно быть документальным, и не всегда визуализация должна быть в виде традиционных графиков и диаграмм.

## Развлечение

Где-то между журналистикой и искусством визуализация нашла себе лазейку и пробилась в область развлечений. Если подумать о данных в более отвлеченном плане — выйти за рамки электронных таблиц и текстовых файлов с разделителями-запятыми и перенестись туда, где оцениваются фотографии и статусы, — заметить это будет нетрудно.

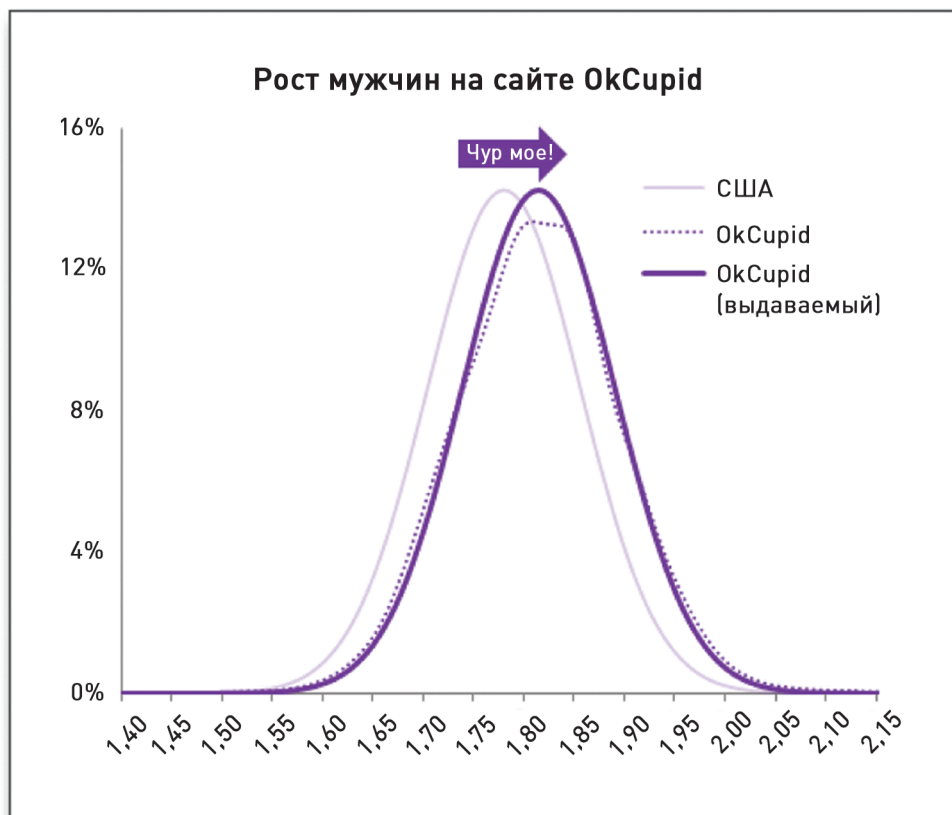
Facebook использовал обновление статусов для определения самого счастливого дня в году, а сайт онлайн-знакомств OkCupid использовал размещаемую на нем информацию для оценки того, в чем и как люди лгут, пытаясь сделать свое цифровое «я» лучше, чем есть на самом деле (рис. 1.3). Такой анализ данных имеет весьма отдаленную связь с развитием бизнеса, увеличением доходов и нахождением сбоев в системе. Но его результаты охватили Всемирную паутину как пожар — именно из-за своей развлекательной ценности. Эти данные раскрывают нам кое-что о нас самих и о нашем обществе.

Так, Facebook обнаружил, что самый счастливый день — это День благодарения, а OkCupid выяснил, что люди склонны к преувеличениям чаще всего тогда, когда говорят о своем росте, и любят приписать себе примерно 5 лишних сантиметров.

► Исследуйте человеческие эмоции в интерактивном режиме с живым онлайн-произведением Джонатана Харриса и Сепы Камвара на сайте <http://wefeelfine.org>.

► Еще большее количество примеров произведений искусства, созданных с использованием данных, вы найдете на сайте FlowingData: <http://dataf1.ws/art>.

► Зайдите на блог OkTrends, и вы узнаете еще много чего интересного об онлайн-свиданиях — например, что на самом деле нравится белым людям и как случайно не оплошать: <http://blog.okcupid.com>.



**Рис. 1.3.** Распределение значений роста мужчин на сайте OkCupid

## Побуждение

Конечно, задача историй не всегда состоит в том, чтобы информировать или развлекать людей. Иногда истории используются, чтобы донести мысль о необходимости срочно предпринять некие меры или чтобы подвигнуть людей на какие-то действия. Кто может забыть тот момент в «Неудобной правде», когда Ал Гор встает на ножничный подъемник, чтобы продемонстрировать, как поднимается уровень углекислого газа в атмосфере?

Однако, на мой взгляд, никто не сумел справиться с задачей побуждения людей к действию так хорошо, как это удалось Хансу Рослингу (Hans Rosling), профессору в области международного здравоохранения и директору фонда Garminder. Используя инструмент под названием «трендализер» (рис. 1.4), Рослинг создал анимированное изображение, показывающее изменения уровня бедности в разных странах. Выступление у него получилось просто изумительным, так что если вы его еще не видели, я горячо рекомендую вам это сделать.

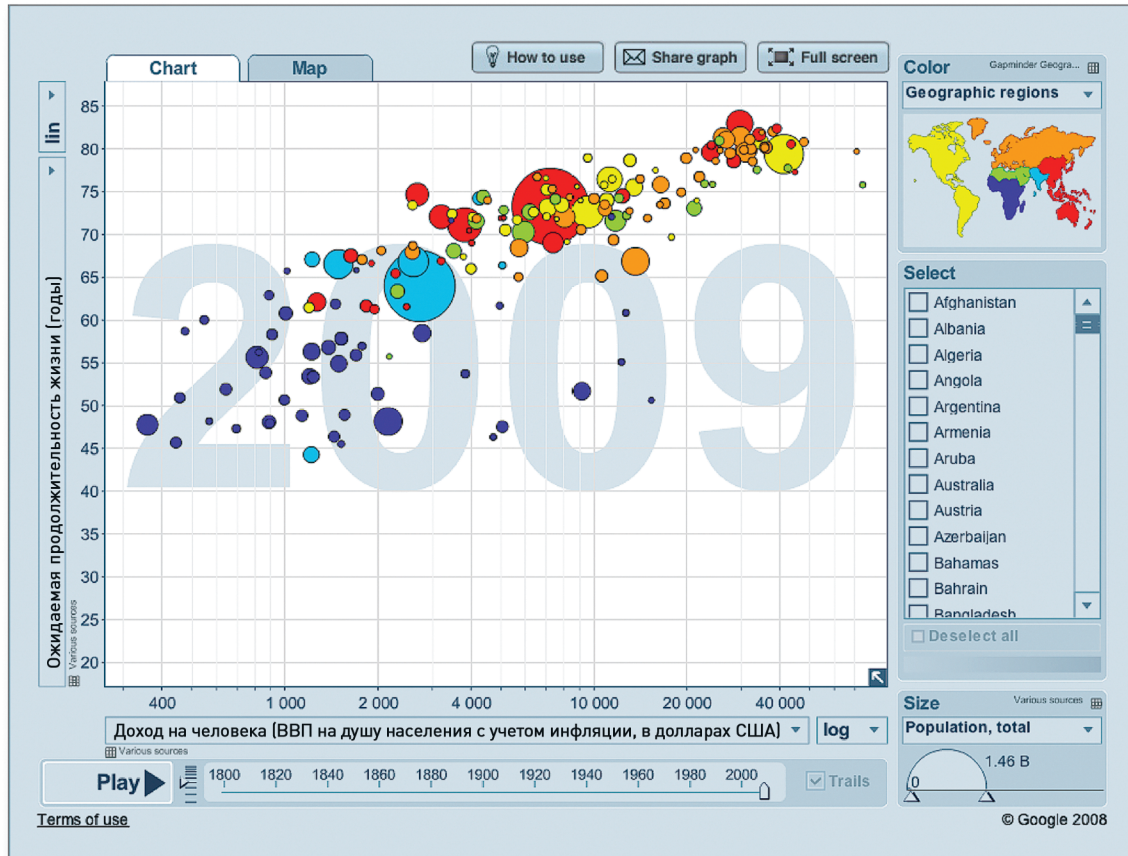


Рис. 1.4. Трендалайзер фонда Garminder

С точки зрения визуализации тут все довольно элементарно. Это просто динамическая диаграмма. Пузырьки обозначают различные государства и двигаются в соответствии с динамикой показателей бедности в конкретной стране на протяжении определенного года. Так в чем секрет популярности выступления Рослинга? В том, что он говорит с убеждением и волнением. Он рассказывает историю. Сколько раз вы видели презентации со схемами и диаграммами, от которых все зевают?! Но у Рослинга все иначе. Он извлек смысл, заложенный в данных, и в полной мере раскрыл его. А если к этому добавить еще и глотание шпаги в конце лекции, все становится на свои места — в зале не остается равнодушных. Посмотрев лекцию Рослинга, мне захотелось взять в руки эти данные и самому их рассмотреть. Мне не терпелось вникнуть в эту историю поглубже.

Позже я посмотрел другую презентацию фонда Garminder на ту же самую тему, с той же формой визуализации данных, но с другим докладчиком. Получилось отнюдь не так волнующе. Если говорить честно, я чуть не задремал. Эмоций было ноль. Данные не показались мне

► Посмотрите, как Ханс Рослинг вызывает восторг аудитории с помощью данных и потрясающих слайдов, на странице <http://dataf1.ws/hans>.

ни убедительными, ни увлекательными. Так что дело не только в данных — не они одни делают рассказ интересным. Важно то, как вы о них рассказываете и как их подаете. От этого зависит, запомнят ли их люди.

Когда все материалы собраны, вам необходимо запомнить: подходите к визуализации так, как будто вы собираетесь рассказать историю. Какого рода историю вы хотите поведать? Это будет репортаж? Или рассказ о жизни? Вы хотите убедить людей в необходимости каких-то действий?

Продумайте развитие персонажей. За каждой строчкой в массиве данных стоит история — точно так же, как у каждого героя в книге имеется свое прошлое, настоящее и будущее. Все эти числа находятся между собой в каких-то связях и отношениях. От вас требуется обнаружить их. Конечно, эксперты в сторителлинге, прежде чем засесть за написание текста, должны научиться составлять предложения.

## Что искать

Хорошо, истории. Договорились. А какого типа истории можно рассказать с помощью данных? Ну, все зависит от того, какими именно данными вы располагаете, но в общем случае, чему бы ни была посвящена ваша графика, вам следует всегда искать две вещи: паттерны и зависимости.

## Паттерны

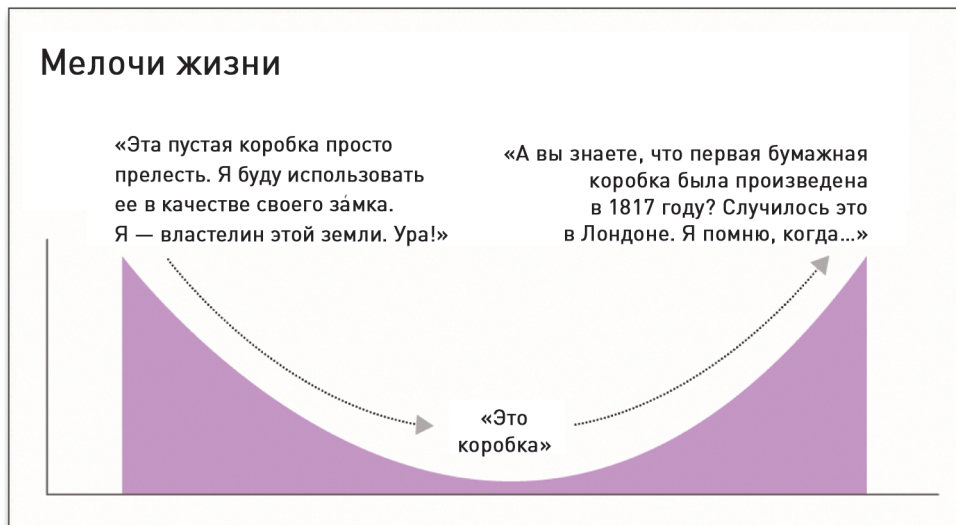


Рис. 1.5. Веселый взгляд на старение

Со временем все меняется. Вы становитесь старше, ваши волосы седеют, а зрение слабеет (рис. 1.5). Цены меняются. Логотипы меняются. Компании появляются. Компании исчезают. Иногда эти перемены случаются внезапно и без предупреждения. А иногда они происходят так медленно, что вы их даже не замечаете.

На какие бы данные вы ни смотрели, тенденция сама по себе может быть интересной, как может быть интересен любой процесс изменений. И здесь вы можете изучать паттерны во времени. Допустим, к примеру, что вы изучаете, как с ходом времени менялся курс акций. Он, конечно, то поднимается, то снижается, но на

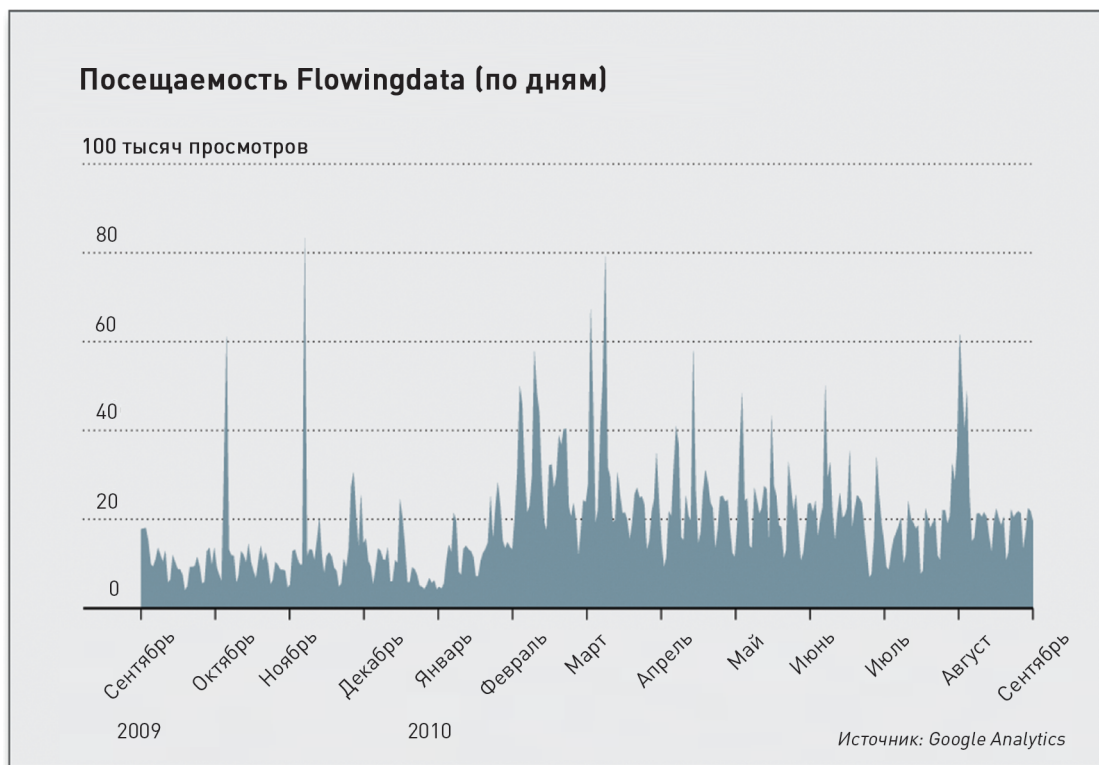


сколько пунктов он меняется за сутки? А за неделю? А за месяц? Существуют ли периоды, в которые курс поднимался сильнее, чем обычно? И если да, то почему это происходило? Имели ли место некие события, которые могли спровоцировать подобные изменения?

Как вы убедитесь сами, стоит начать с одного-единственного вопроса, взятого в качестве отправной точки, и тут же за ним потянутся следующие. Причем так обстоят дела не только с временными рядами, но и со всеми типами данных. Попробуйте подойти к вашему массиву данных в большей степени как исследователь, и тогда вы, скорее всего, получите в итоге более интересные ответы.

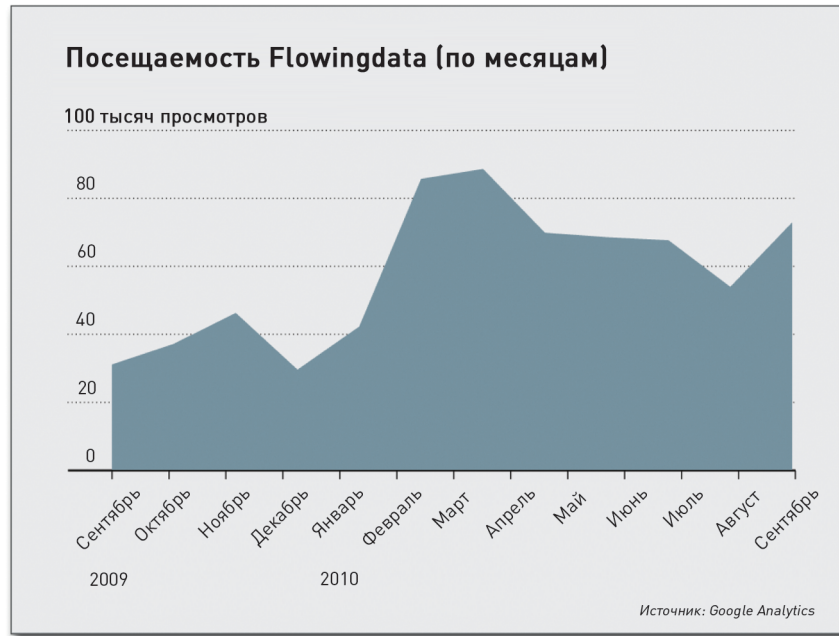
Данные типа временных рядов можно разделить по-разному. Иногда есть смысл представить значения по дням или по часам. В других случаях может оказаться, что полезнее взглянуть на данные по месяцам или по годам. В первом случае в графике, возможно, окажется больше шума, а во втором случае вы получите более обобщенный взгляд.

Те из вас, у кого есть сайт и кто пользуется той или иной компьютерной программой для анализа его посещаемости, могут в этом быстро убедиться. Если вы посмотрите на трафик на вашем сайте на базе посещаемости по дням, как это показано на рис. 1.6, то вы увидите, что он весьма бугристый. Колебания в нем довольно велики.



**Рис. 1.6.** Количество уникальных посетителей сайта FlowingData в день

А если посмотрите на график, выстроенный на базе посещаемости по месяцам (рис. 1.7), то увидите, что, хотя он охватывает тот же промежуток времени, точек на нем меньше и его линия более гладкая.



**Рис. 1.7.** Количество уникальных посетителей сайта FlowingData в в месяц

Я не хочу сказать, что один график лучше другого. На самом деле они могут дополнять друг друга. То, как вы проводите разбивку данных, зависит от того, какое количество деталей вам нужно (или не нужно).

Конечно, временные зависимости — это не единственное, что можно искать. Определенные паттерны можно найти также и в совокупностях, которые способны помочь вам сравнить некие группы людей или других объектов. Что вы чаще всего едите в течение недели? О чем обычно говорит президент в своем Послании Конгрессу? Какие штаты обычно голосуют за республиканцев? В данном случае полезно поискать паттерны по географическим регионам. И, как вы убедитесь, читая следующие главы, типы вопросов и данных могут быть разными, но ваш подход останется почти неизменным.

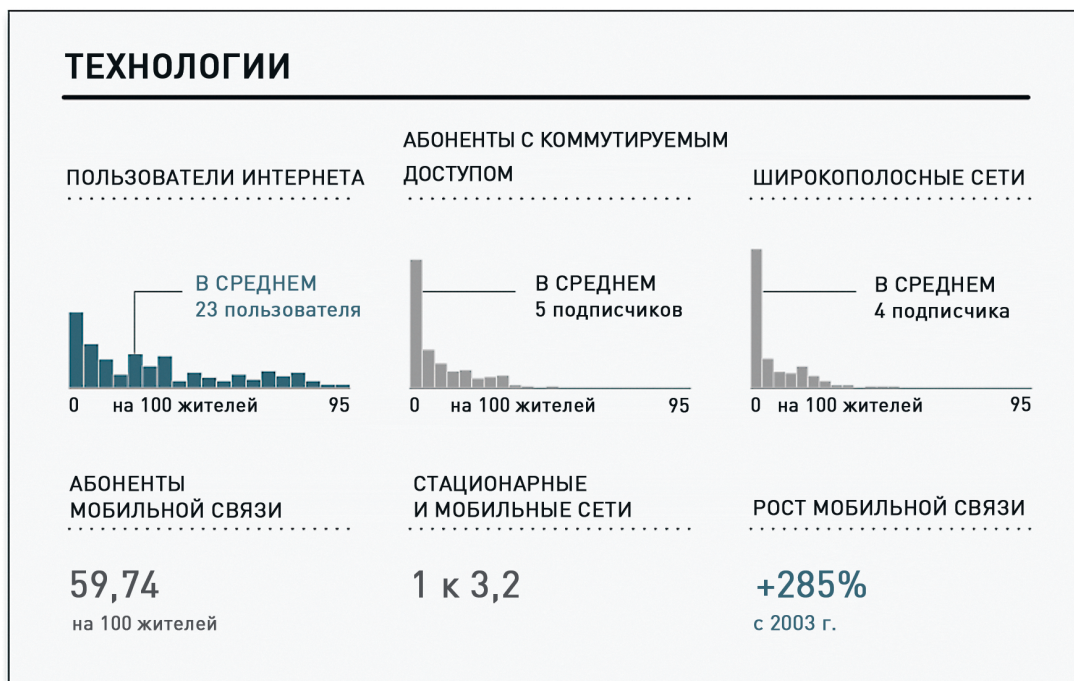
## Зависимости

Вам наверняка когда-нибудь попадались на глаза рисунки с целым набором графиков и диаграмм, расположенных как будто в совершенно случайном порядке? Я говорю о рисунках,

которым очевидно чего-то не хватает — как будто графическому дизайнеру некогда было поразмыслить над данными, и он второпях придал им некую форму, лишь бы успеть к крайнему сроку. Очень часто то, чего так явно не хватает в подобных случаях, — и есть зависимость.

В статистике под зависимостью, как правило, подразумевают корреляцию и причинность. Множественные переменные могут быть переплетены между собой сложными связями. Шестая глава, «Визуализация зависимостей», посвящена именно этим понятиям и тому, как их визуализировать.

Однако и на более абстрактном плане, на котором вы не думаете об уравнениях и критериях проверки гипотез, вы можете разрабатывать диаграммы и графики для визуального сравнения и сопоставления величин и распределений. В качестве простого примера можно привести схему проникновения технологий в жизнь, взятую из «Доклада о мировом прогрессе» и представленную на рис. 1.8.



► «Доклад о мировом прогрессе» — это графический отчет, в котором сравниваются показатели прогресса во всем мире и для составления которого используются данные портала UNdata. Его полную версию вы можете найти на странице <http://dataf1.ws/12i>.

**Рис. 1.8.** Проникновение технологий в жизнь во всем мире

Столбцовые диаграммы показывают количество пользователей интернета, а также количество абонентов с коммутируемым доступом и широкополосных сетей на 100 жителей. Отметим, что разброс у интернет-пользователей (от 0 до 95 на 100 жителей) гораздо шире, чем у двух других групп.

Проще и быстрее всего — предоставить компьютерной программе, которая у вас установлена, самой решить, какой диапазон использовать для каждой столбцовой диаграммы. Однако все

столбцовые диаграммы здесь выстроены на базе одного и того же диапазона, несмотря на то что нет ни одной страны, у которой имелось бы по 95 абонентов с коммутируемым доступом или пользователей широкополосных сетей на 100 жителей. Но это дает вам возможность легко сравнивать распределения между группами.

А потому, когда вам приходится работать с множеством различных наборов данных, постарайтесь думать о них как о группе, а не как об отдельных, никак не связанных между собой массивах. Так вы сможете получить гораздо более интересные результаты.

## Сомнительные данные

Пытаясь разглядеть истории, скрывающиеся за вашими данными, стоит всегда подвергать сомнению то, что вы видите. Помните: то, что это — числа, еще не означает, что они верны.

Я должен признаться, что для меня проверка данных — самый нелюбимый этап создания графики. Конечно, когда некий человек, группа или организация предоставляет вам груды данных, то это должна быть их обязанность — проверить, все ли они достоверны. Но хорошие дизайнеры графики не принимают все на веру. Ведь и хорошие строители не берут для возведения фундамента дома абы какой цемент. Так что и вам для создания графики не следует брать абы какие данные.

Проверка и верификация данных — одна из важнейших, если не самая важная, часть графического дизайна.

По сути, вам необходимо просмотреть данные на предмет нахождения в них чего-то такого, что лишено смысла. Может, кто-то допустил ошибку при вводе и добавил лишний ноль или, наоборот, пропустил его. Может, во время сбора данных возникли некие проблемы со связью и часть информации дошла в искаженном виде. Во всех случаях, если что-то покажется вам подозрительным, вам необходимо будет свериться с первоисточником.

У человека, занимающегося сбором данных, обычно возникает некое чувство того, что именно можно ожидать. Если сбором данных занимались вы сами, тогда просто спросите себя: это нормально — что один штат дает 90 процентов чего бы там ни было, а все остальные, вместе взятые, уместаются в диапазоне 10–20 процентов от общего объема? Что там такое происходит?

Очень часто выясняется, что аномалия в ваших данных — всего лишь опечатка. Хотя бывает и такое, что она оказывается самым интересным фактом во всем массиве данных и становится двигателем вашей истории. Просто проверьте, с чем именно вы столкнулись в каждом конкретном случае.

## Дизайн

---

Когда уже все данные в порядке, вы можете приступать к визуализации. Над чем бы вы ни работали — будь то доклад, онлайн-инфографика или произведение data-арта, — вам следует

соблюдать несколько основных принципов. Каждый из них дает пространство для маневра, так что вам лучше воспринимать их, скорее, как некие ориентиры, а не как свод жестких правил, но если в графическом дизайне вы новичок, тогда вам именно с них и стоит начать.

## Объясните кодировки

Разработка любой диаграммы протекает примерно по одному и тому же сценарию. Вы собираете данные; вы кодируете данные с помощью кругов, столбцов и цветов; а затем предлагаете другим людям прочесть их. В этот момент читатели должны расшифровать ваши кодировки. Что означают все эти круги, столбцы и цвета?

В работе Уильяма Кливленда (William Cleveland) и Роберта Макгила (Robert McGill) вопросы кодировки рассмотрены в деталях. Некоторые типы кодировок работают лучше, чем другие. Но какой именно подход к кодировке вы выберете, будет не так уж и важно, если в конечном счете читатели не смогут понять, что именно выражают те или иные обозначения. Если они не смогут раскодировать ваше послание, значит, время, которое вы посвятили созданию диаграммы, окажется потраченным впустую.

Возможно, вы отмечали про себя, что в рисунках, находящихся где-то между data-артом и инфографикой, порой отсутствует контекст, а уж с произведениями data-арта это и вовсе обычная ситуация. Легенды и метки способны полностью разрушить впечатление от работы, но вы можете по крайней мере дать немного информации, описав самое необходимое в одном коротком абзаце. Это поможет другим оценить ваш труд.

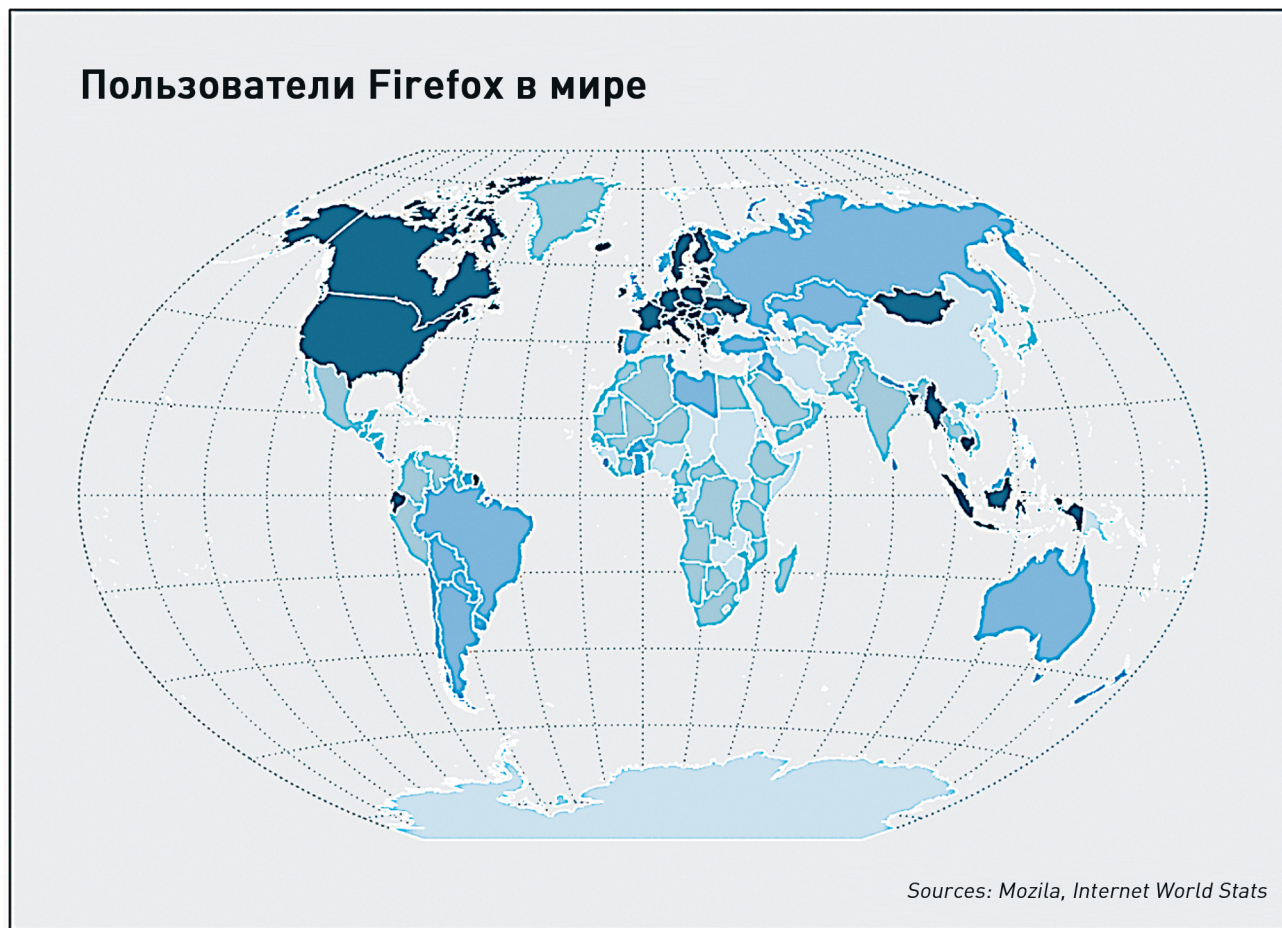
Однако с подобным отсутствием пояснений иногда можно столкнуться и в тех случаях, когда основная задача диаграмм и графиков как раз и состоит в представлении данных. В итоге читатели бывают разочарованы, а вам наверняка не хочется допускать этого. Происходит такое порой потому, что, работая с данными, вы с ними срастаетесь и знаете наизусть, что означает каждый элемент диаграммы. Читателям же все внове. Не зная контекста, с которым вы сами свыклись, анализируя данные, они оказываются словно слепыми — не могут увидеть то, что видите вы.

Итак, как вы можете убедиться, что читателям удастся расшифровать ваши кодировки? Просто объясните им значение каждого элемента с помощью подписей, легенд и ключей. Что именно выбрать — зависит от ситуации. Посмотрите, например, на карту мира на рис. 1.9, которая дает представление об использовании программы Firefox в различных странах земного шара.

Вы видите, что разные страны окрашены в разные оттенки голубого, но что означают эти оттенки? Темный цвет — это больше или меньше потребителей? И если больше, то больше какого именно количества? В том виде, в котором карта представлена на рисунке, она довольно-таки бесполезна. Но если вы снабдите ее легендой, такой, как на рис. 1.10, все сразу прояснится. В данном случае цветовая легенда выполняет и еще одну функцию: она представляет собой столбцовую диаграмму, демонстрирующую распространенность программы на основании количества пользователей.

### ПРИМЕЧАНИЕ

За дополнительной информацией о том, как люди кодируют и декодируют цвета и формы, обратитесь к публикации Кливленда и Макгила *Graphical Methods for Analyzing Data* («Восприятие графики и графические методы анализа данных»).



**Рис. 1.9.** Использование программы Firefox в мире, по странам

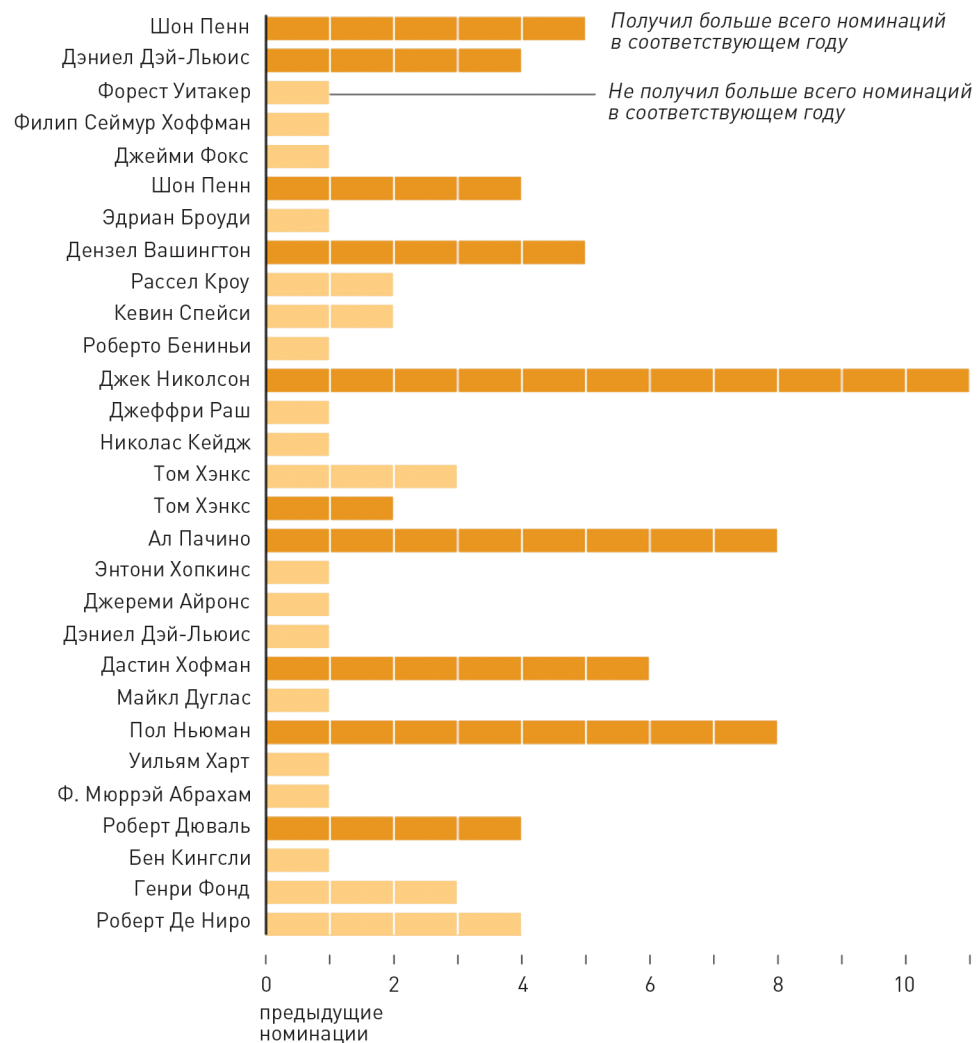


Если у вас есть достаточно места или слишком много категорий, вы можете вставить подписи непосредственно рядом с объектами вашей диаграммы, как это продемонстрировано на рис. 1.11. На данной диаграмме показано количество номинаций, которые тот или иной актер получил перед тем, как был удостоен премии «Оскар» за лучшую мужскую роль.

**Рис. 1.10.** Легенда к карте использования Firefox

### Актеры, получившие «Оскар» за лучшую мужскую роль, и количество их номинаций

Можно ли сказать, что у актеров, имеющих большее количество номинаций на получение премии, также и больше шансов на победу? Только 10 из последних 29 актеров-лауреатов в данной категории номинировались большее количество раз, чем другие претенденты



Source: Wikipedia

Рис. 1.11. Непосредственное обозначение объектов

По интернету в свое время разгуливала теория о том, что статуэтку, как правило, выигрывает тот актер, который в соответствующий год получил больше всего номинаций среди представителей своей когорты. Как следует из подписей, оскароносцы, отмеченные темно-оранжевым, действительно получили больше всего номинаций, а светло-оранжевым отмечены те, кто не получил больше всего номинаций, однако тем не менее был удостоен статуэтки в соответствующем году.

Как видите, у вас есть множество вариантов действия. Все эти детали не требуют больших усилий в использовании, они вроде мелкие, но разница от их применения может быть огромной для восприятия вашей работы.

## Дайте осям названия

Вы уже поняли, почему необходимо всегда давать пояснения к кодировке. Точно так же вам следует всегда давать названия осям в ваших графиках и диаграммах. Без подписей и объяснений оси будут исполнять лишь декоративную роль. Дайте им названия, чтобы читатели знали, какая именно шкала применяется в каждом конкретном случае. Логарифмическая, дифференциальная, экспоненциальная или вы измеряете все в расчете на 100 сливных бачков унитаза? Лично я, когда не вижу у осей названий, всегда по умолчанию принимаю, что речь идет именно о последнем.

Чтобы продемонстрировать свою мысль, я вернусь к конкурсу, который организовал пару лет назад на сайте FlowingData. Я разместил на страничке изображение, представленное на рис. 1.12, и попросил читателей дать самые интересные названия осям.

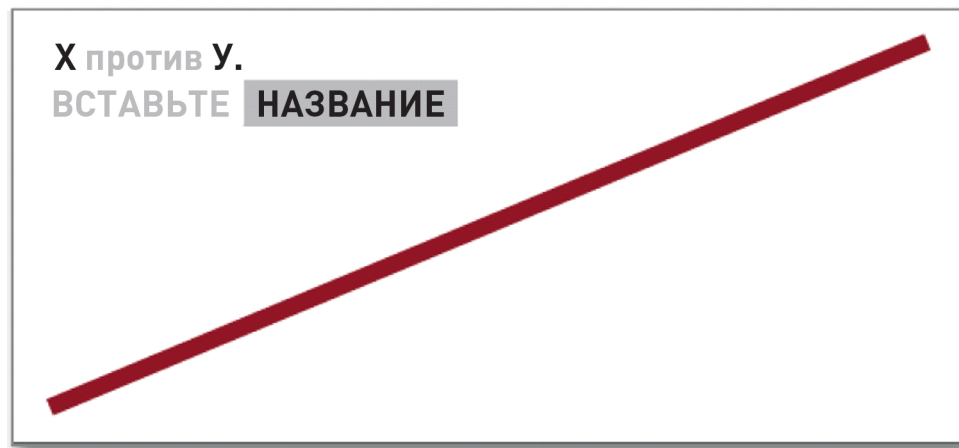


Рис. 1.12. Вставьте название

Для одного и того же графика я получил около 60 различных вариантов названий. Часть из них представлена на рис. 1.13.



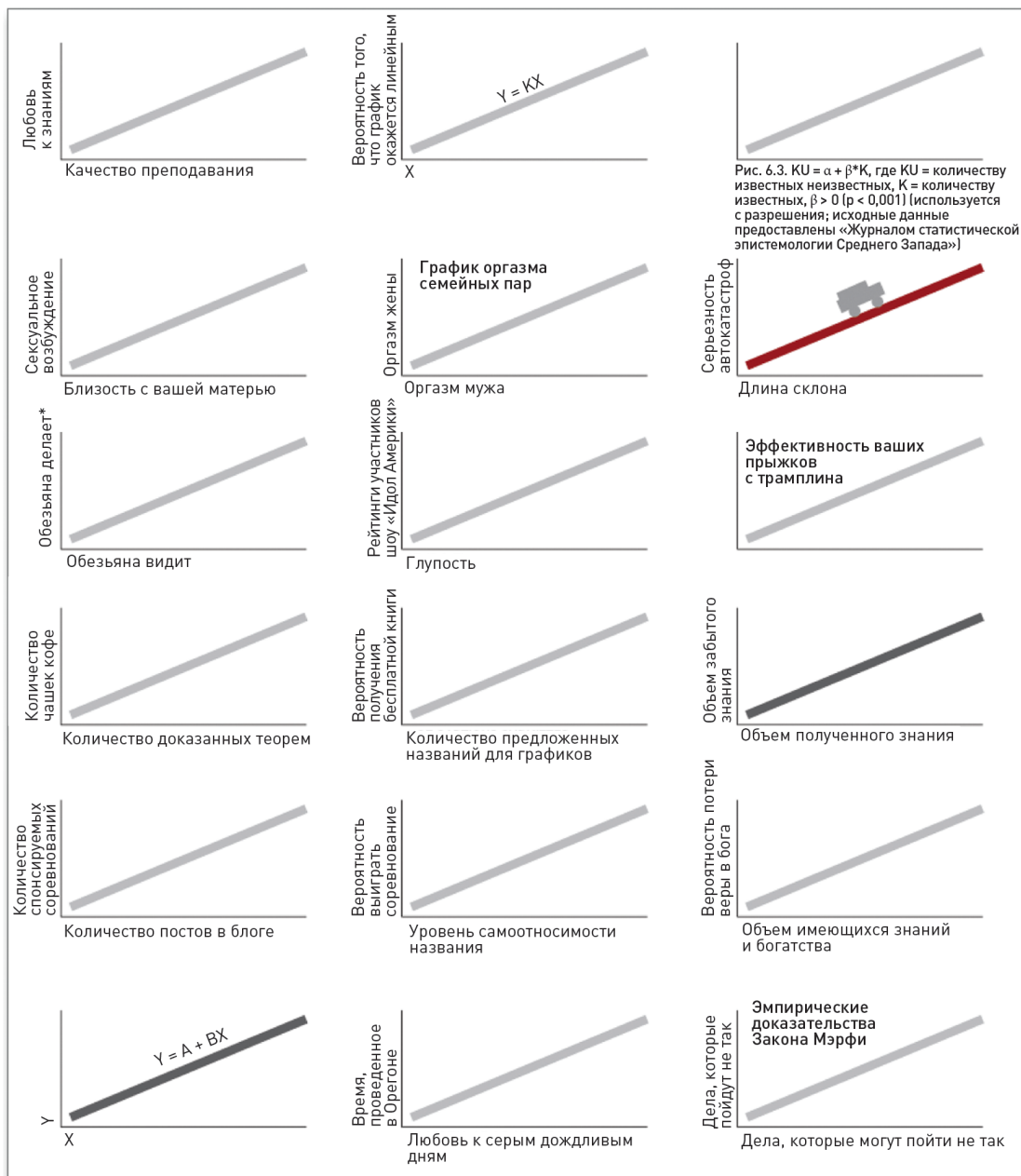


Рис. 6.3.  $KU = \alpha + \beta * K$ , где KU = количеству известных неизвестных, K = количеству известных,  $\beta > 0$  ( $p < 0,001$ ) (используется с разрешения; исходные данные предоставлены «Журналом статистической эпистемологии Среднего Запада»)

\* «Обезьяна видит — обезьяна делает» — ямайская поговорка, описывающая процесс копирования чужого поведения без понимания его смысла. Прим. пер.

Рис. 1.13. Некоторые из результатов, полученных во время соревнования по придумыванию названий на FlowingData

Как вы можете убедиться, притом что график везде один, в результате простой смены названий осей получаются совершенно разные истории. Конечно, все это делалось забавы ради. А теперь представьте себе, что данный график должен был рассказать о чем-то очень серьезном. Без подписей и названий ваша диаграмма просто бессмысленна.

## Держите геометрию под контролем

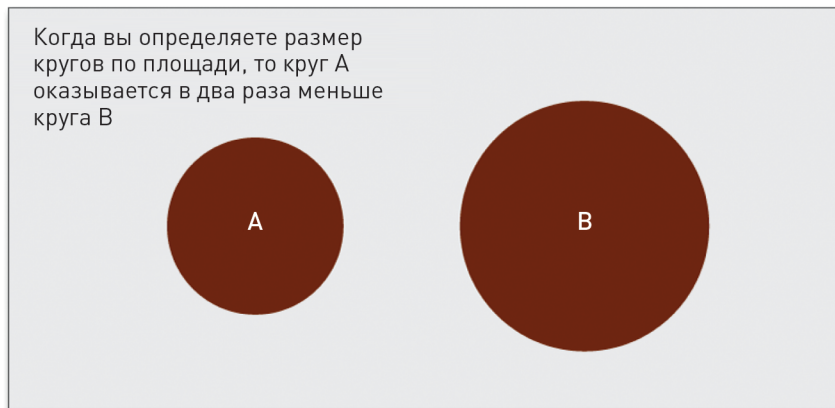
Когда вы создаете диаграмму, вы используете геометрические фигуры. В столбцовой диаграмме вы используете прямоугольники. При этом величины вы выражаете высотой прямоугольников. В точечной диаграмме индикатором величины служит местоположение. То же самое и с графическим представлением временных рядов. В круговых диаграммах для выражения величины используется угол, а сумма величин всегда равняется 100 процентам (рис. 1.14). Это несложно, но будьте начеку, потому что так же несложно все перепутать. Если вы не проявите достаточного внимания, вы легко можете допустить ошибку, и когда вы облажаетесь, люди, особенно в Сети, не побоятся вам об этом сказать прямым текстом.



**Рис. 1.14.** Правильное и неправильное составление круговых диаграмм

Другая распространенная ошибка обычно происходит тогда, когда дизайнеры используют для обозначения величин двумерные формы, но измеряют их так, как если бы они использовали только одно измерение. Прямоугольники в столбцовой диаграмме двумерные, но в качестве индикатора используется только их высота. Ширина не означает ничего. Однако, создавая пузырьковую диаграмму, для обозначения величин следует использовать площадь кругов. Начинающие дизайнеры вместо этого часто применяют радиус или диаметр, и соотношение получается в корне неверным.

На рис. 1.15 показаны два круга, размер которых задавался по площади. Это правильный подход.



**Рис. 1.15.** Правильный подход при определении размеров кругов в диаграммах

На рис. 1.16 представлена пара кругов, чьи размеры определялись по диаметру. Первый круг диаметром в два раза меньше второго, но по площади он меньше в четыре раза.



**Рис. 1.16.** Неправильный подход при определении размеров кругов в диаграммах

То же самое происходит и с прямоугольниками, например в тримапах. В качестве индикатора величины там используется площадь, а не высота или ширина.

## Укажите источники

Казалось бы, не нужно говорить о необходимости указывать источники данных — но уж слишком много людей забывают это сделать. Откуда взялись все эти данные? Если вы посмотрите на диаграммы в газетах, вы увидите, что источник в них всегда указан — обычно мелким

шрифтом где-нибудь внизу. Вы должны поступать так же. Иначе читатели не будут знать, насколько достоверными являются ваши графики и диаграммы.

Как еще они могут убедиться в том, что данные не были просто выдуманы? Конечно, вы бы такое никогда не сделали, но не все об этом знают. Если вы укажете в своей работе источник, читатели не только будут относиться к ней с бóльшим уважением, но они смогут также самостоятельно проверить и проанализировать данные.

Включение источника данных способствует также обогащению контекста. Вполне очевидно, что опрос, проведенный на региональной ярмарке, будет интерпретирован иначе, чем тот, который осуществлялся во время переписи населения страны.

## Учитывайте аудиторию

И наконец, всегда помните о вашей аудитории и о цели вашей работы. Например, диаграмма, создаваемая для слайда презентации, должна быть простой. Вы, конечно, можете включить в нее массу деталей, но их увидят только те, кто сидит в первом ряду. А вот если вы делаете постер, который люди будут пристально разглядывать и изучать, тогда вы можете включить в диаграмму гораздо больше подробностей.

Вы готовите деловой отчет? Тогда не пытайтесь создать самое красивое произведение data-арта, которое когда-либо видел мир. Вместо этого постарайтесь разработать четкую и ясную диаграмму, в которой все строго по делу. Графика нужна вам как подспорье в процессе анализа? В таком случае результат визуализации данных увидите только вы и, следовательно, не стоит особо тратить время на оформление и аннотации. А может, ваша диаграмма или график предназначается для печати в средствах массовой информации? Тогда вы должны стремиться к простоте и пояснять все мысли, восприятие которых может оказаться затруднительным для аудитории.

## Закругляясь

---

Короче говоря, начните с вопроса, осмотрите данные критическим взглядом и определитесь с целью визуального средства, а также узнайте, для кого оно предназначено. Это поможет вам разработать понятную диаграмму или график, которые будут заслуживать того, чтобы люди потратили время на их рассмотрение. И неважно, какого именно типа графикой вы воспользуетесь.

Как это делать, вы научитесь в следующих главах. Вы поймете, как подавать и визуализировать данные. Вы узнаете, как разрабатывать диаграммы с самого начального этапа и до конца. А затем вы примените полученные знания к вашему собственному массиву данных. Вы определитесь с тем, какую историю вы хотите рассказать, и создадите соответствующий графический объект.