

## Искусство графического представления данных

Наиболее простыми и эффективными способами представления статистических данных являются графические изображения.<sup>1</sup> Хороший рисунок позволяет сразу выявить основные закономерности, скрытые в массиве информации. Для улучшения анализа данных необходимы ясные и точные таблицы и графики. Излишние украшения и вычурность лишь мешают. В последние годы широкое распространение электронных таблиц и графических пакетов привело к интенсивному использованию рисунков для иллюстрации статистических данных. Несмотря на то что графические изображения довольно часто приносят пользу, злоупотребление графикой создает впечатление, что единственной целью статистики является наукообразный обман.

Вероятно, одним из наиболее известных пропагандистов правильного представления данных с помощью графических средств является Эдвард Р. Тафти. Ранее я уже излагал его идеи в заметках [Принцип Эдварда Тафти минимизации количества элементов диаграммы](#) и [Как с помощью диаграммы приукрасить действительность? или о факторе лжи Эдварда Тафти](#).

Идеальная диаграмма должна обладать следующими основными свойствами (по Тафти):

- Иллюстрировать данные.
- Концентрировать внимание на существе графического изображения, а не на способе его создания.
- Предотвращать искажения.
- Облегчать сравнение данных.
- Быть наглядной.
- Быть тесно связанной со статистическими и словесными описаниями изображения.

Тафти так же сформулировал пять принципов графического представления данных:

- Графическое представление данных должно отражать суть дела, статистические свойства данных и быть хорошо продуманным.
- Графическое представление данных должно просто, ясно и эффективно представлять сложные идеи.
- Графическое представление данных должно порождать у наблюдателя наибольшее количество идей за минимальный промежуток времени.
- Графическое представление данных всегда многомерно.
- Графическое представление данных должно отображать истинный смысл данных.

Существует несколько способов оценки качества графического представления данных. Одним из них является вычисление информативности рисунка на основе доли чернил, затраченных на иллюстрацию данных.

Информативность рисунка равна доле чернил, затраченных на иллюстрацию данных, в общем объеме чернил, затраченных на весь рисунок. Цель этого показателя — не допустить излишеств.

$$\text{Информативность} = \frac{\text{Объем чернил, затраченных на иллюстрацию данных}}{\text{Общий объем затраченных чернил}}$$

Цель этого отношения — максимизировать долю чернил, затраченных на иллюстрацию данных. Не следует перегружать рисунок элементами, не имеющими отношения к делу. Например, это часто относится к линиям сетки на графике. Такие элементы называются графическим хламом. Графический хлам — это декоративные украшения, не имеющие отношения к данным или являющиеся их излишними уточнениями. Графический хлам часто представляет собой самостоятельное графическое изображение, фокусирующее внимание на способе его создания, а не на данных.

При создании рисунка нельзя искажать данные. График считается верным, если он полностью соответствует исходным данным. Количество искажений, которые вносятся графиком, называются фактором лжи. Фактор лжи — это отношение величины эффекта, изображенного на графике, к величине эффекта исходной выборки.

---

<sup>1</sup> Используются материалы книги Левин и др. Статистика для менеджеров. — М.: Вильямс, 2004. — с. 141–145

Любое изменение, показанное на графике, должно соответствовать изменениям, существующим в исходных данных. Довольно часто при построении графика этим правилом пренебрегают. Это приводит к искажениям и несоответствиям между графическим изображением и данными. Чтобы лучше разобраться с этими принципами, рассмотрим несколько примеров, которые нарушают правила построения графических изображений.

На рис. 1 представлена иллюстрация к статье в журнале The Time, посвященной возрастающему экспорту австралийского вина в США. На ней, в частности, показан бокал, символизирующий 6,77 млн. галлонов вина, поставленного из Австралии в США в 1997 году. Обратите внимание на то, что объем вина в этом бокале должен мене чем в два раза превышать объем вина в предыдущем бокале, соответствующем 3,67 млн. галлонов, а тот, в свою очередь, должен содержать в три с половиной раза больше вина, чем первый бокал. На самом деле этот не так. Отчасти это объясняется тем, что иллюстраторы использовали трехмерный рисунок вместо двухмерного. Такие иллюстрации могут привлекать внимание, но все же правильнее было бы использовать двухмерную диаграмму или график.



Рис. 1. Импорт вина из Австралии

Кроме того, если на рисунке не указано начало координат, набор данных оказывается искаженным. В качестве примера обратимся к рис. 2, приведенному в той же статье. У этого рисунка есть несколько недостатков. Во-первых, на оси X не отложено ни одного значения. Поскольку рис. 2 представляет собой график, изображающий рост площадей, занятых виноградниками в винной промышленности с течением времени, на оси X следовало бы отметить годы. Вместо этого, годы (в правильном порядке) указаны рядом с объемами площадей. Во-вторых, точки изображены неверно. Это более серьезная ошибка. Точка, соответствующая 135 326 акрам в 1949-1950 г., изображена выше, чем точка, соответствующая 150 300 акрам в 1969-1970 г! Кроме того, разница между 1979-1980 и 1997-1998 г. должна в три раза превышать разницу между 1979-1980 и 1969-1970 г., а на рисунке эта пропорция нарушена. В-третьих, интервалы времени на оси X изображены неверно. Точка, соответствующая 1979-1980 г., намного ближе к точке, соответствующей 1989-1990 г., чем к точке, изображающей площадь виноградников в 1969-1970 г.



Рис. 2. Площадь виноградников

В журналах и газетах часто печатают рисунки, содержащие излишнюю информацию. Например, рис. 3 иллюстрирует рынок газированных безалкогольных напитков в 1999 году. Хотя в целом рисунок правильно отображает разницу между долями рынка, занятыми разными напитками, он содержит массу ненужных деталей. Количество чернил, затраченных на изображение пены, рвущейся из бутылок, намного превышает разумную величину. Гораздо лучше было бы изобразить эти данные в виде круговой диаграммы.

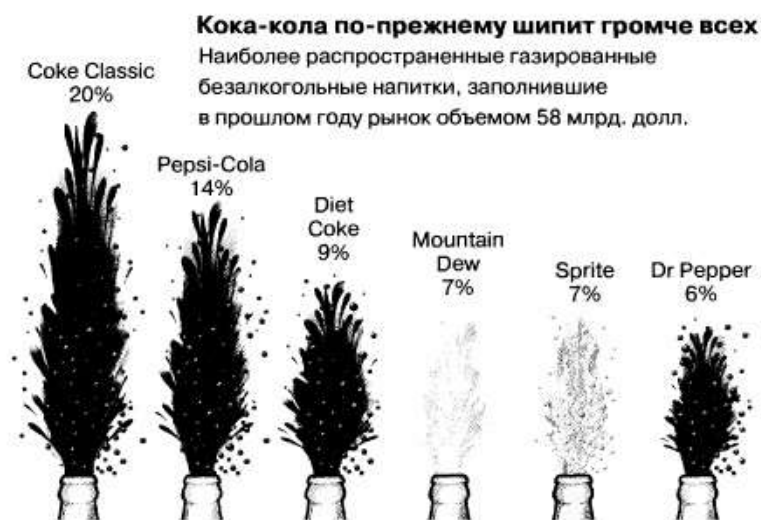


Рис. 3. Доли рынка, занятого газированными безалкогольными напитками в 1999 году

Подведем итоги. Каждый из нас поглощает огромный объем информации из газет и журналов. Поскольку большая ее часть представляет собой ненужный хлам, необходимо научиться отбрасывать лишнее. Следует также помнить, что иногда графики захламляются по невежеству, а иногда — чтобы ввести читателей в заблуждение. Следовательно, очень важно сохранять скептический настрой.

Как указывает Тафти, первое, что приходит в голову людям, рассматривающим иллюстрации статистических данных, — «неправда». Слишком часто графики искажают реальные данные, затрудняя читателям поиск истины. При выборе способа изображения статистических данных — таблиц или рисунков — возникает много этических проблем. Необходимо одинаково честно изображать как хорошие, так и плохие результаты. Делая устный или письменный доклад, необходимо излагать информацию честно, объективно и нейтрально. Следует различать неудачную и нечестную презентацию. Критерий, с помощью которого это можно сделать, — намерения докладчика. Очень часто искажения и излишества при иллюстрации статистических данных возникают в результате невежества. Однако бывает, что под этим скрывается стремление обмануть читателей и слушателей.