

# Отображение информации в графическом виде

Николай Тоговеровский, 12 мая 2009

Есть множество способов отображения количественной информации. Простейший способ — предложение. Предложение «У Васи было 5 яблок и 4 вишни» позволяет с легкостью передать два числа. Однако если чисел становится больше, то предложение уже не так эффективно: у Васи было 5 яблок, 4 вишни, 8 орехов, десяток бананов, 35 конфет, 17 слив и 32 лимона. Когда есть необходимость передать больше одного-двух чисел, то лучше воспользоваться таблицей (→)

У Васи было:

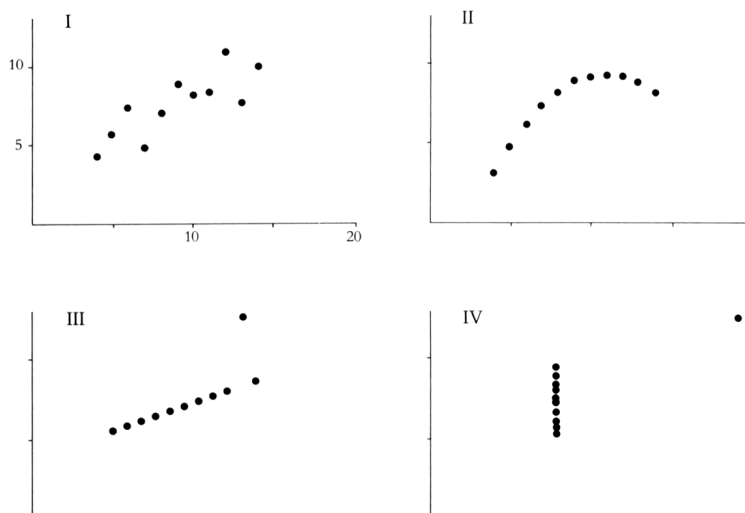
4 вишни  
5 яблок  
8 орехов  
10 бананов  
17 слив  
32 лимона  
35 конфет

Хотя графический способ отображения данных (то есть отображение с помощью графиков, диаграмм, схем, карт, чертежей и т. д.) позволяет показывать очень большие объемы информации, это не самое важное его достоинство. Замечательная особенность графики в том, что очень часто она позволяет *увидеть* суть данных. Рассмотрим квартет Анскомби<sup>1</sup>:

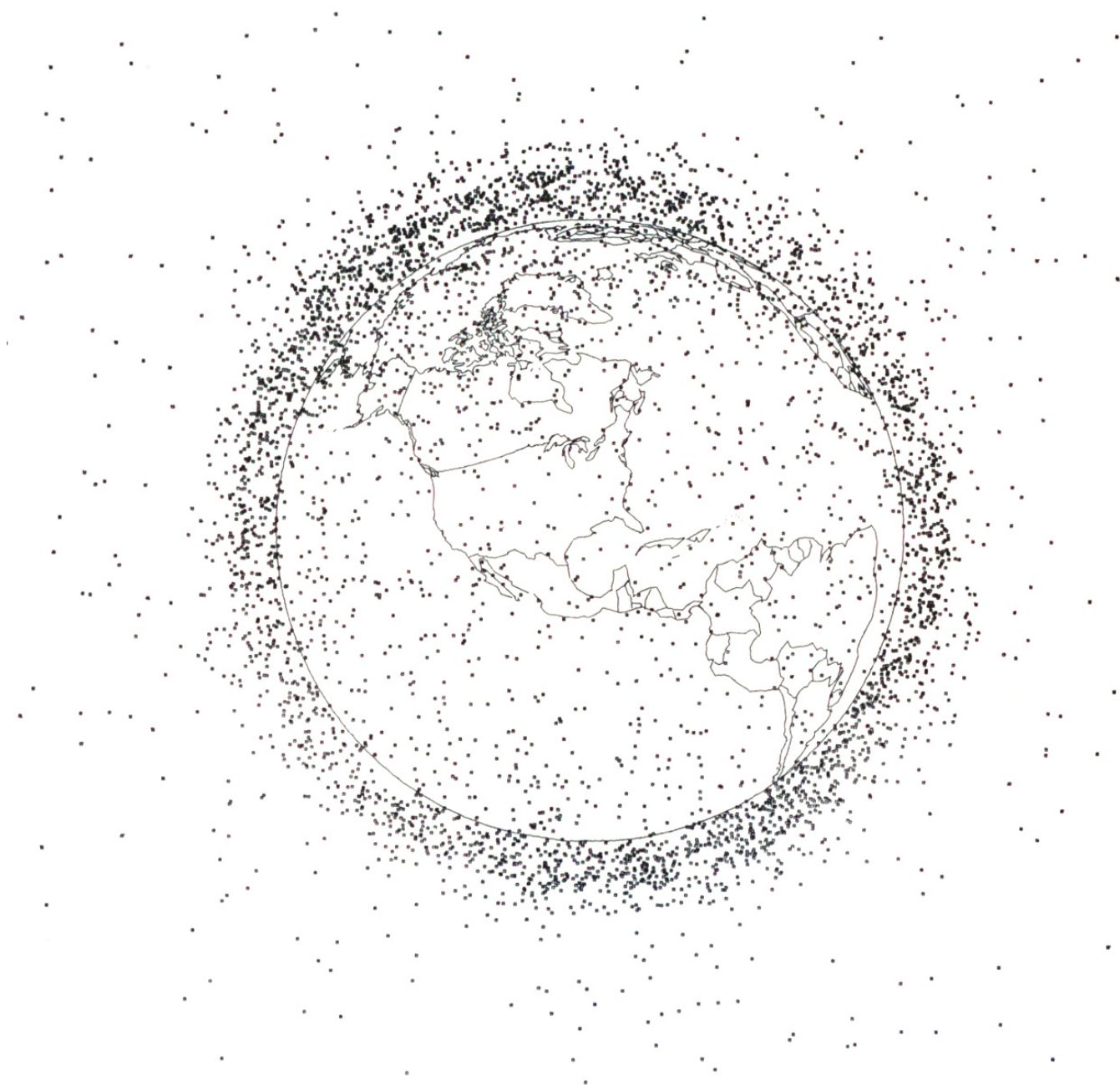
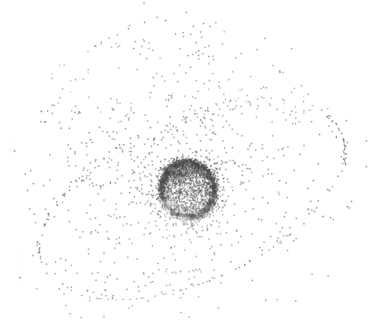
<sup>1</sup> [http://en.wikipedia.org/wiki/Anscombe's\\_quartet](http://en.wikipedia.org/wiki/Anscombe's_quartet)

| I    |       | II   |      | III  |       | IV   |       |
|------|-------|------|------|------|-------|------|-------|
| X    | Y     | X    | Y    | X    | Y     | X    | Y     |
| 10.0 | 8.04  | 10.0 | 9.14 | 10.0 | 7.46  | 8.0  | 6.58  |
| 8.0  | 6.95  | 8.0  | 8.14 | 8.0  | 6.77  | 8.0  | 5.76  |
| 13.0 | 7.58  | 13.0 | 8.74 | 13.0 | 12.74 | 8.0  | 7.71  |
| 9.0  | 8.81  | 9.0  | 8.77 | 9.0  | 7.11  | 8.0  | 8.84  |
| 11.0 | 8.33  | 11.0 | 9.26 | 11.0 | 7.81  | 8.0  | 8.47  |
| 14.0 | 9.96  | 14.0 | 8.10 | 14.0 | 8.84  | 8.0  | 7.04  |
| 6.0  | 7.24  | 6.0  | 6.13 | 6.0  | 6.08  | 8.0  | 5.25  |
| 4.0  | 4.26  | 4.0  | 3.10 | 4.0  | 5.39  | 19.0 | 12.50 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15  | 8.0  | 5.56  |
| 7.0  | 4.82  | 7.0  | 7.26 | 7.0  | 6.42  | 8.0  | 7.91  |
| 5.0  | 5.68  | 5.0  | 4.74 | 5.0  | 5.73  | 8.0  | 6.89  |

Все четыре столбца с числами описываются одними и теми же элементарными статистическими свойствами (ср. значение  $X = 10$ ;  $Y = 7,5$ ; корреляция между  $X$  и  $Y$  равна  $0,816$ ; уравнение регрессии:  $Y = 3 + 0,5X$ ). Однако если представить эти данные в графическом виде, то мгновенно открывается их существенное различие:

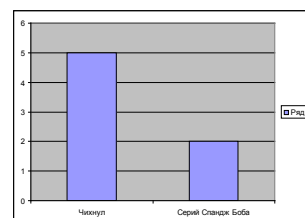


Существует огромное количество всевозможных способов отображения информации в графическом виде. Но в основе всех их лежит одно правило: *хорошая графика начинается с хороших данных*. Если у вас есть координаты более чем семи тысяч единиц различного космического мусора (остатков ракет, мертвых спутников, замороженные испражнения космонавтов, дробь, оставшейся от испытания противоспутниковых вооружений, 34-х ядерных реакторов и много чего еще), то такие данные, скорее всего, будут интересны, если, конечно, вы их хорошо покажете.



Иллюстрации Николаса Джонсона, Колорадо Спрингс, Колорадо.  
Точки не в масштабе с Землей.

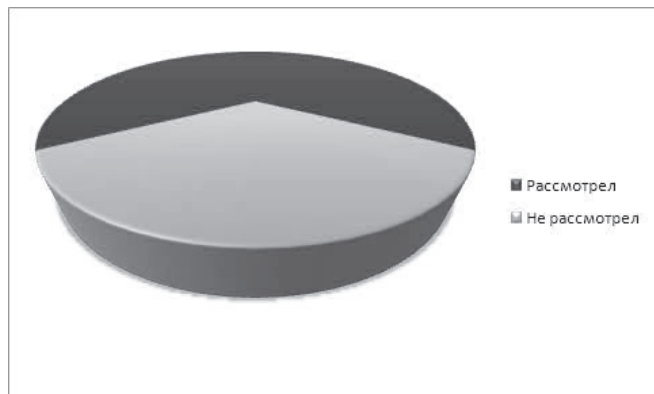
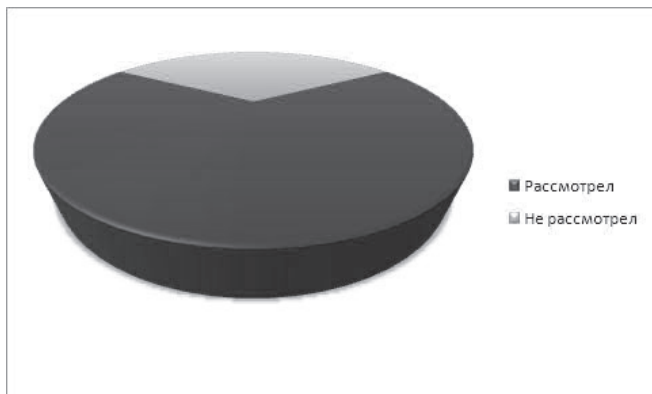
С другой стороны, если вы сосчитали сколько раз за день вы чихнули и хотите показать зависимость количества чихов от количества просмотренных за день серий Спандж Боба (→), то какие бы методы, приемы и технологии визуализации вы бы не использовали — результат будет прескверный, потому что *глупая теория ведёт к глупой графике*.



## Коэффициент брехни

При отображении информации в графическом виде могут возникать искажения, так как исходные цифры, обычно скрыты за точками и линиями (в отличие от таблиц, которые являются одним из лучших способов отображения точных значений).

К сожалению, такие искажения, иными словами *ложь*, очень распространены. Бывает, что искажения создаются намеренно (такая практика, например, широко распространена в маркетинге), тогда информация становится *дезинформацией*. Бывает, что авторы лгут не специально, просто не замечая своей лжи, тем более что такие офисные пакеты как Микрософт Эксель весьма способствуют этому. Рассмотрим небольшой пример. Пусть какой-то лектор разбирал со студентами их дипломные работы. Из 12 работ 8 он рассмотрел, а 4 нет. Такие «тонкие» данные, надо сказать, вообще не заслуживают графического представления, но наш лектор решил показать их «красивенько» и построил график в Экселе. Некоторые диаграммы Экселя могут отображать данные с чудовищными искажениями. К их числу относятся перспективные круговые диаграммы:



В зависимости от того какой стороной повернут пирожок<sup>1</sup> видимый эффект будет совершенно разным. На диаграмме слева кажется, что не рассмотренных работ очень мало, а справа наоборот, создается впечатление, что части почти равны.

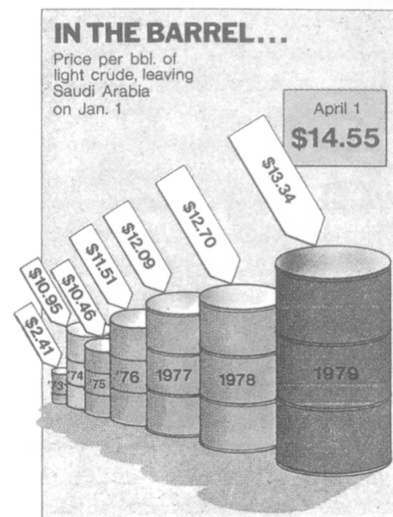
Для того чтобы понять насколько врет тот или иной график, диаграмма, и т. д. можно использовать коэффициент брехни.

$$\text{Коэффициент брехни} = \frac{\text{Смысл, показанный на графике}}{\text{Смысл, содержащийся в данных}}$$

Коэффициент брехни хорошей графики должен быть равен  $1 \pm 0,05$ . В правой круговой диаграмме примерно 40 000 (28%) пикселей показывают рассмотренные работы, а 100 000 (72%) — не рассмотренные. В исходных данных же не было рассмотрено 33,3% работ. Таким образом, коэффициент брехни равен  $71/33 = 2,15$ .

Когда с помощью объема пытаются показать изменение одной величины, почти всегда получается неправда. Если принять в расчёт площадь крышек бочек (справа), то коэффициент брехни будет 9,4, а если кто-то подумает, что величины показываются всем объемом, то исходный рост цены в 454 процента будет показан как рост в 27 000 процентов, а коэффициент брехни — 59,4!

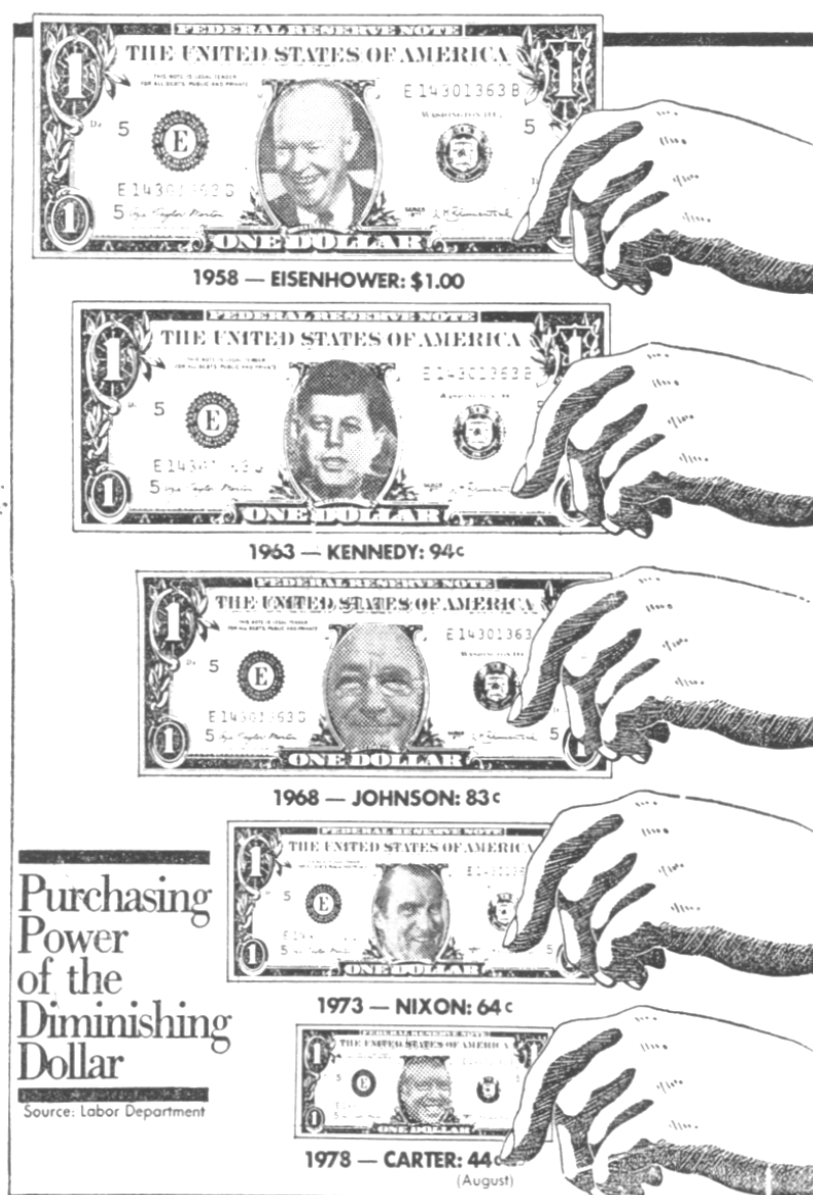
<sup>1</sup> По английский круговая диаграмма называется *pie chart* — пирожковая диаграмма.



Тайм, 9 апреля 1979, с. 57

Площадь — тоже опасный дизайн для отображения одной величины. Если бы площадь долларов честно показывала изменение ценности, то доллар 1978 года должен был бы быть почти в два раза больше.

Вашингтон пост, 25 октября 1978,  
с. 1



## Коэффициент чернил-данных.

Хорошая информационная графика привлекает внимание пользователя к данным, а не к чему-либо еще. Иными словами, каждая капля чернил (или каждый пиксель, если графика отображается на экране компьютера) должна использоваться для отображения данных. Достичь этого не всегда возможно, но к этому надо постоянно стремиться. Коэффициент чернил-данных показывает сколько чернил в какой-либо графике используется для отображения данных, а сколько для других, второстепенных целей.

$$\text{Коэффициент чернил-данных} = \frac{\text{Чернила, отображающие данные}}{\text{Все чернила графика}}$$

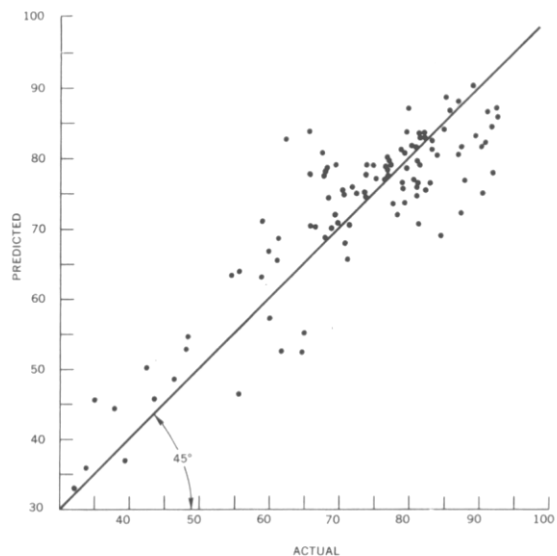
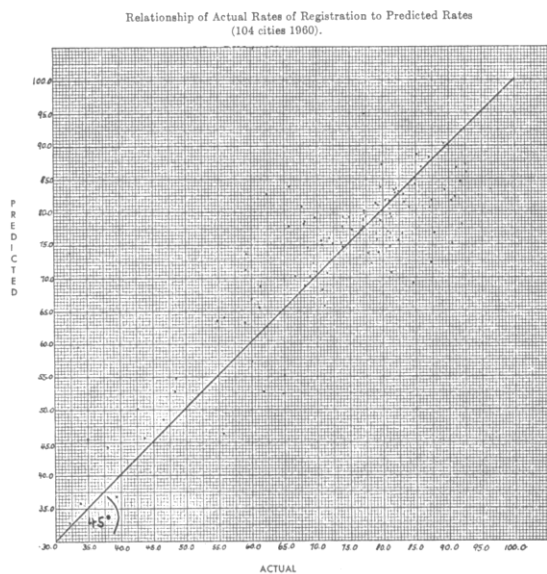
Одна из важнейших задач информационного дизайнера состоит в максимизации коэффициента чернил-данных.



Для этого следует использовать два правила стирания:

*Обоснованно стирайте чернила, которые не отображают данные.*

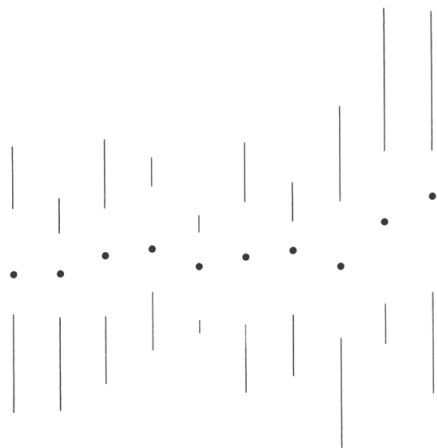
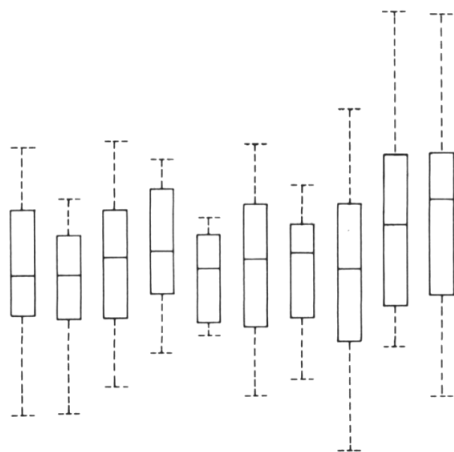
В графике слева яркая сетка скрывает данные, если большую её часть удалить, как на графике справа, то увидеть данные становится гораздо проще.



Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

*Обоснованно стирайте лишние чернила.*

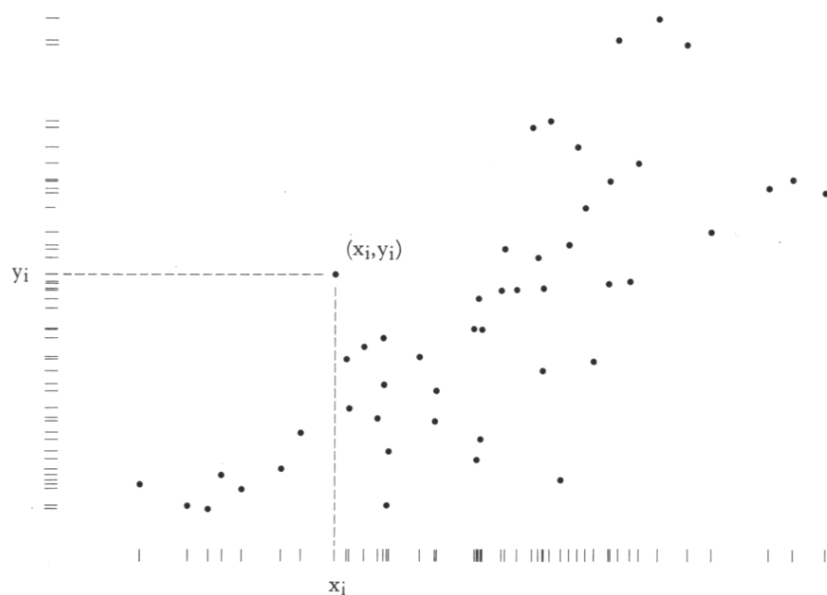
В примере ниже стандартный график, показывающий тренды, очищен от лишних чернил без какой-либо потери функциональности.



Оси в графиках являются вспомогательным элементом. Для того чтобы немного поднять коэффициент чернил-данных ненужные части осей можно стереть.



Оставшееся после стирания лишних чернил место можно заполнить чем-нибудь полезным — информацией. В графике ниже оси заменены небольшими штрихами, которые позволяют увидеть как распределяются точки в графике.



Показывайте данные.

Максимизируйте коэффициент чернил-данных.

Стирайте чернила, не отображающие данные.

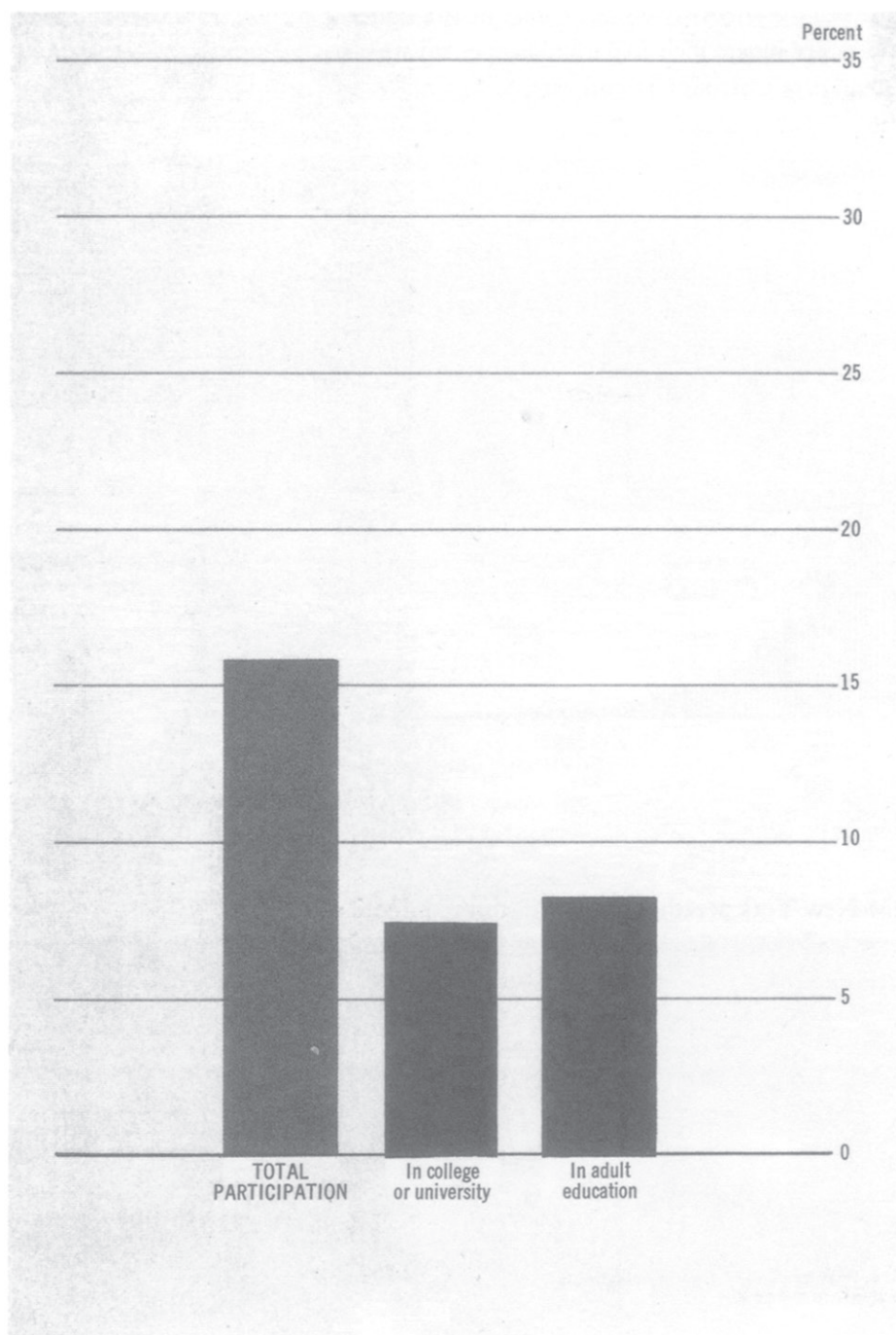
Стирайте лишние чернила.

## Плотность данных

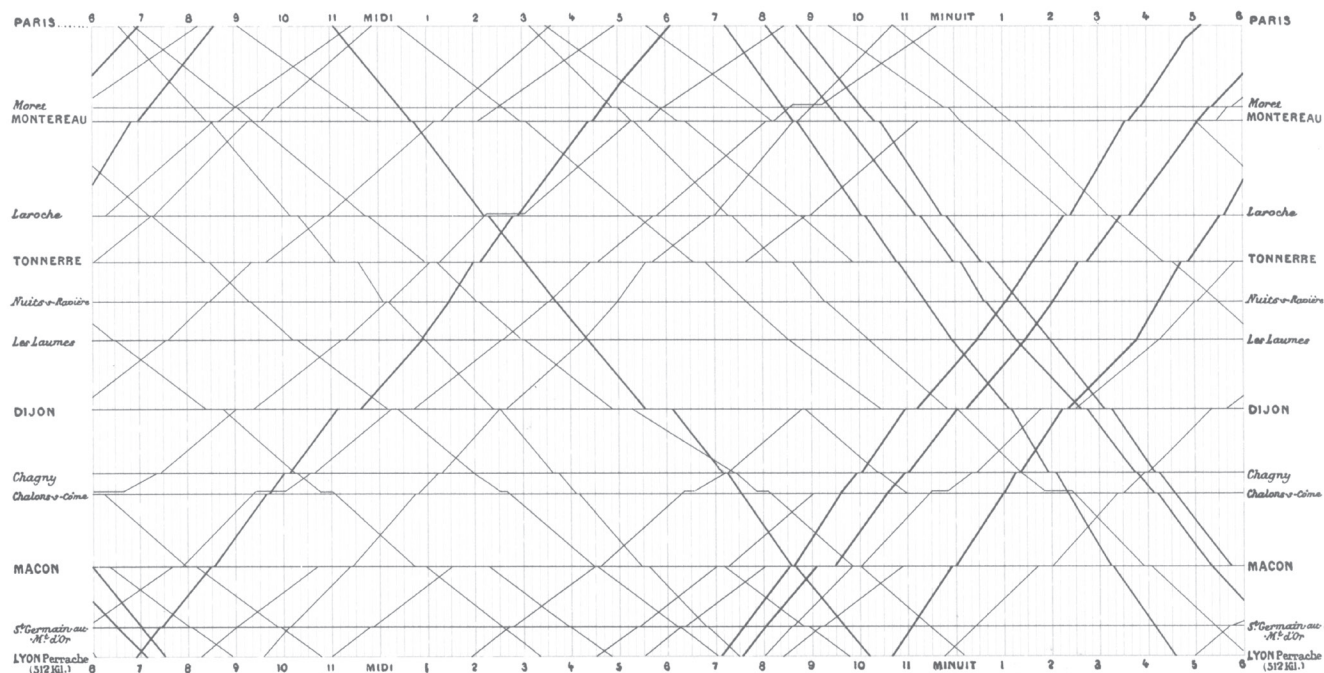
Человеческий глаз способен различать огромное количество отдельных элементов. Рассмотреть 100 точек в одном квадратном сантиметре — не проблема для человеческого глаза. Количество точек, расположенных на какой-то площади, то есть плотность данных нужно стараться максимизировать.

$$\text{Плотность данных} = \frac{\text{Количество записей в матрице с данными}}{\text{Общая площадь графика}}$$

К сожалению, многие графики обладают весьма низкой информационной плотностью, показывая всего несколько элементов на значительной площади. График справа отображает матрицу из всего 4 записей (два числа и две подписи; третий столбик — просто сумма двух первых) и имеет коэффициент плотности данных всего 0,02 числа на кв. сантиметр, что весьма мало.



Управление делами президента, департамент управления и бюджета, *социальные показатели, 1973* (Вашингтон, округ Колумбия, 1973), с. 86.



Плотная в плане информации графика всегда интересна. Её интересно изучать, рассматривать. Она позволяет увидеть как всю картину в целом, так и рассмотреть отдельные детали.

Этьен-Жюль Маре, расписание движения поездов, 1885.

## Обратная связь

Пишите по адресу [kolan@ksoftware.ru](mailto:kolan@ksoftware.ru).

Читайте ещё на сайте [www.ksoftware.ru/wiki](http://www.ksoftware.ru/wiki).