

Основы анализа и визуализации данных для медиа

Лекции 5-6
16 сентября 2019

Рогович Татьяна
rogovich@gmail.com

Принципы хорошей визуализации. Восприятие визуальной информации

Частично базируется на курсе Роберты Синатры в CEU, Будапешт

Структура лекции

- Поговорим о принципах хороших визуализаций по Эдварду Тафти. Чего надо избегать, как визуализировать данные хорошо.
- Узнаем, какие есть психологические особенности восприятия визуальной информации и как они помогают нам делать визуализации лучше
- Поговорим про типичные ошибки самых популярных видов графиков (линейный график, столбчатая диаграмма и pie chart).
- Посмотрим на примеры неудачных визуализаций и обсудим, как они искажают данные
- Получим домашнее задание

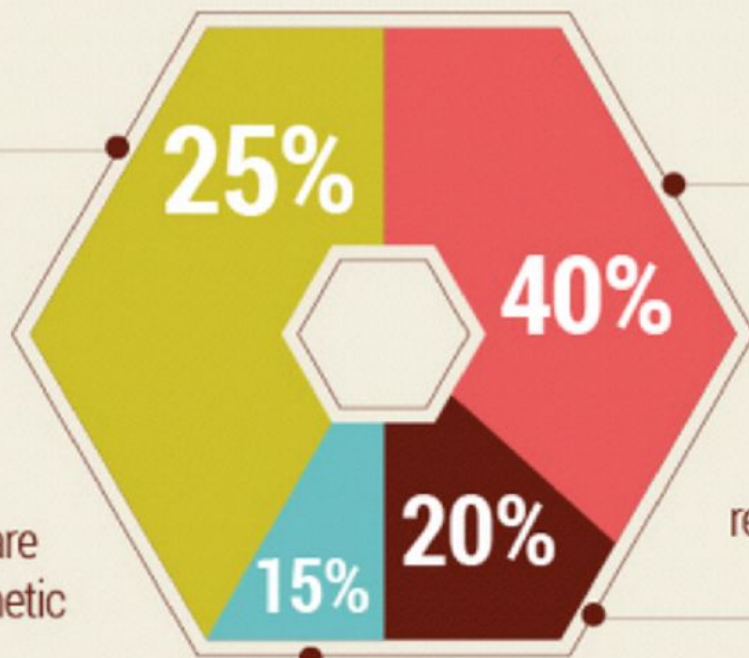
Оцениваем качество визуализации

В конце лекции мы обсудим домашнее задание. В нем нужно будет оценить плохую визуализацию и предложить, как ее улучшить.

А сегодня мы как раз поговорим о том, на что нужно обратить внимание.

Autistic children have communication disabilities.

Autistic children have intellectual disabilities.

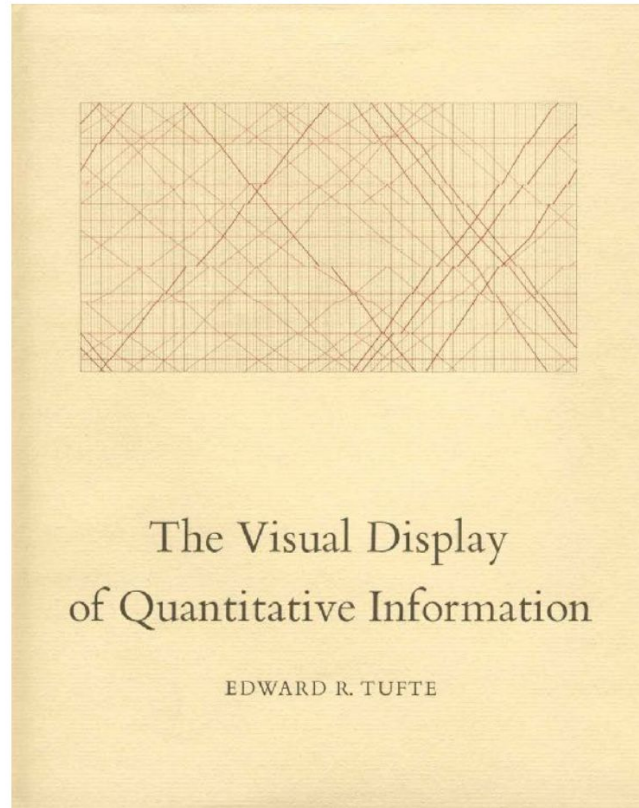


Autism cases are because of genetic disorders.

Autism recurrence rates in families.

Принципы хорошей визуализации по Э. Тафти

The Visual Display of Quantitative Information by *Edward R. Tufte*, Graphics Press, 2001



Graphical Excellence

VDQI Глава 1

Excellence in statistical graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

Э. Тафти

Graphical Excellence

- Это хорошая подача интересных данных - вопрос содержания, статистики и дизайна
- Это сложные идеи, которые переданы ясно, четко и эффективно
- Это то, что дает читателю наибольшее количество идей за наименьшее время с использованием наименьшего количества чернил и места
- ... и читатель узнает о данных только правду

Принципы дизайна Тафти

- Честность данных
- Фактор лжи
- Максимизация
соотношения данных
к чернилам
- Избегать мусора
в графиках

Data Integrity

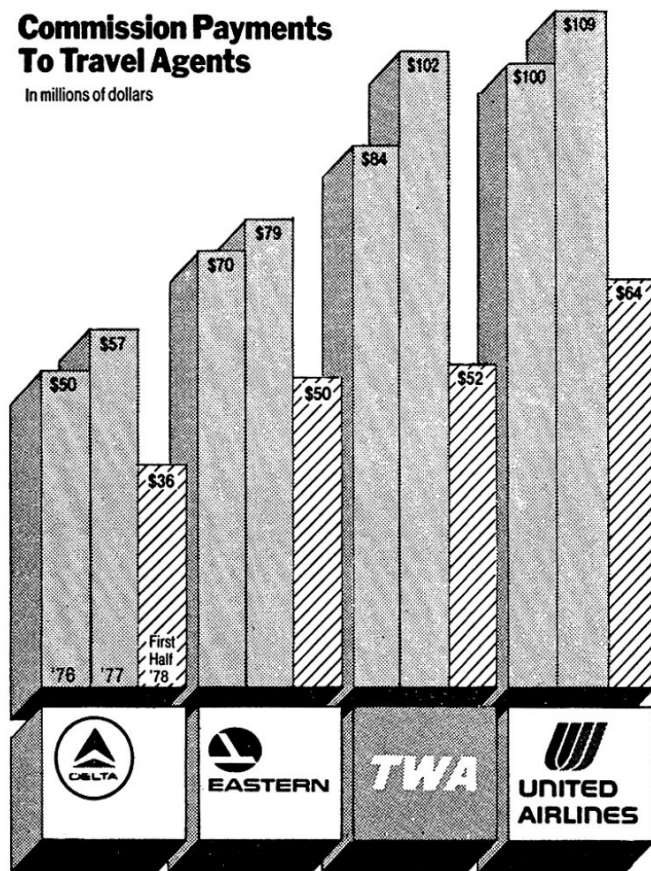
VDQI Глава 1

Честность данных

- Отсутствующие шкалы и лэйблы
- Отсутствующий контекст
- Искаженные шкалы
- Искажающий дизайн

Commission Payments To Travel Agents

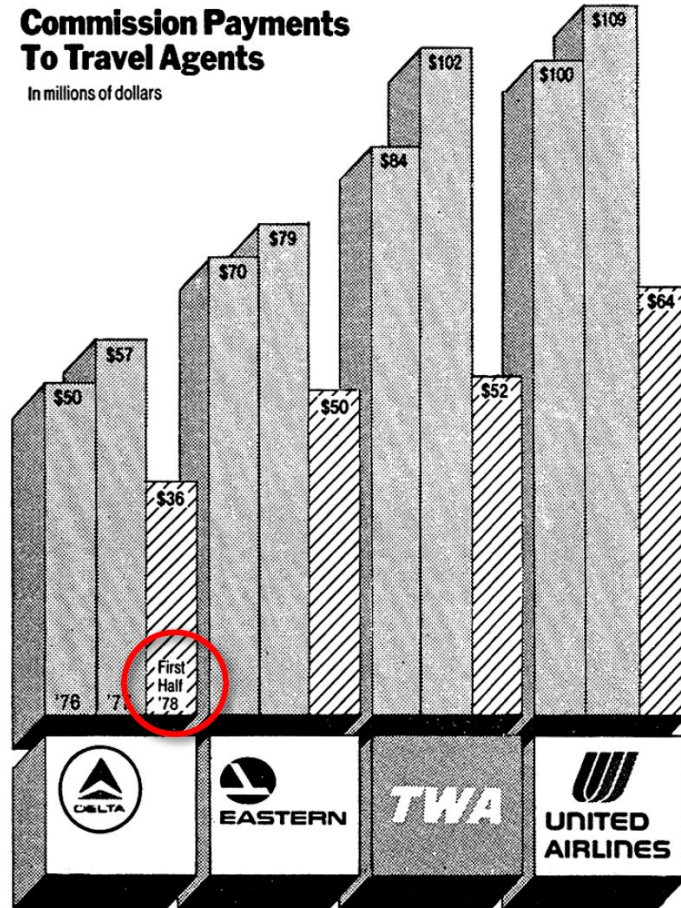
In millions of dollars



VDQI Example (p54)

Commission Payments To Travel Agents

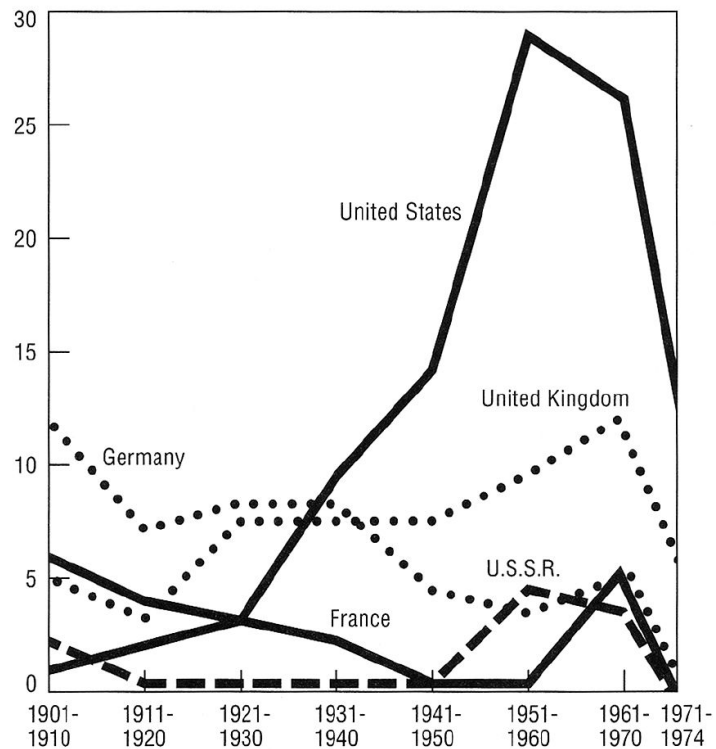
In millions of dollars



- Псевдо-снижение
- Полные годы сравниваются с полугодием

Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974

(Number of Prizes)

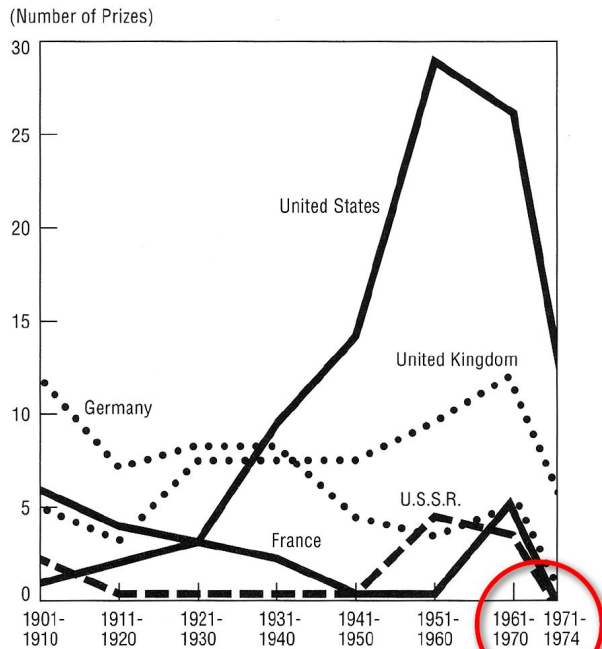


VDQI Example(p60)

- Найдите ошибку.

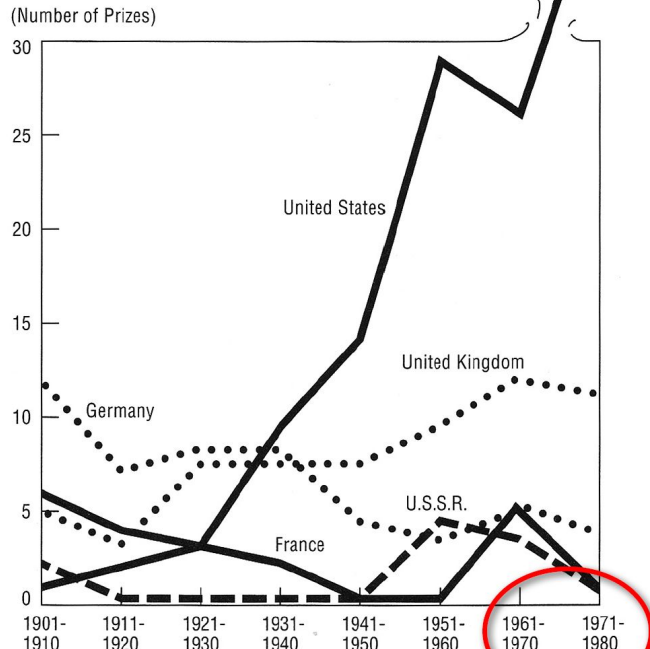
- Все периоды и отрезки должны быть равны.

**Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1974**



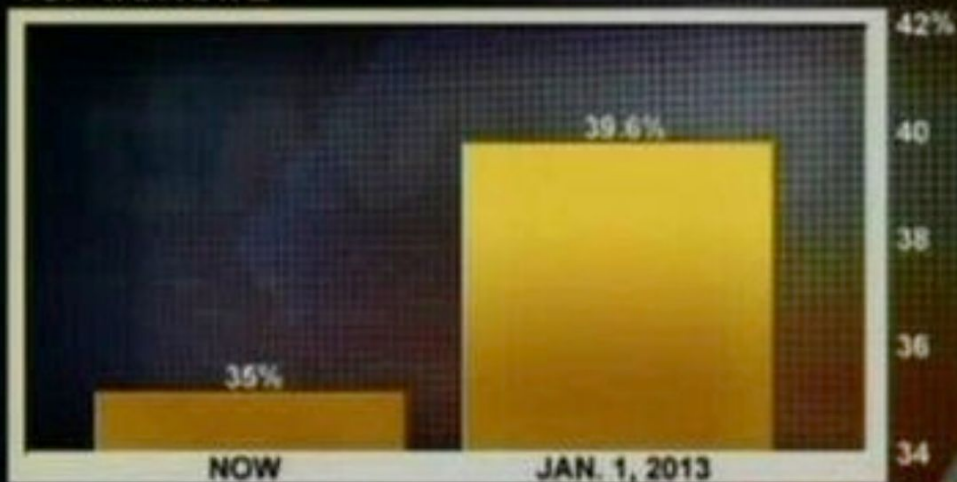
VDQI Example (p60)

**Nobel Prizes Awarded in Science,
for Selected Countries, 1901-1980**



IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

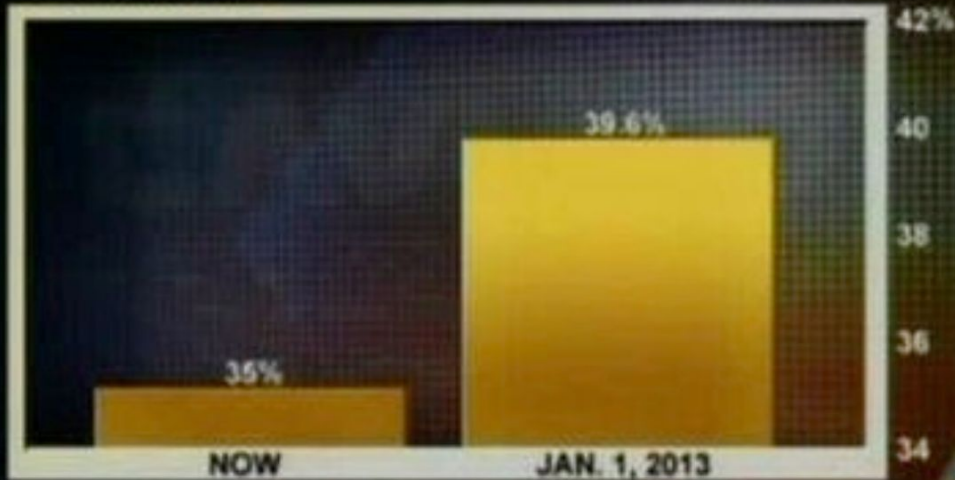
DOW 13008.68 ▲ 64.33

S&P 1379.32 ▲ 5.98

NASDAQ 2939.52 ▲ 6.32

IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



- Искажение шкалы (не с нуля)

8:01p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

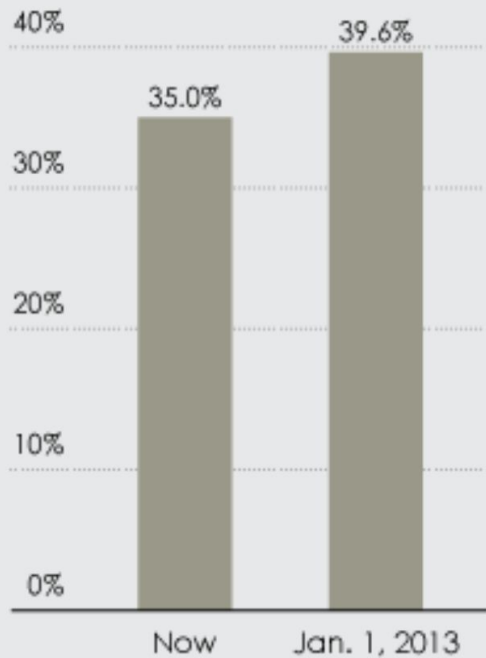
DOW 13008.68 ▲ 64.33

S&P 1379.32 ▲ 5.98

NASDAQ 2939.52 ▲ 6.32

If Bush tax cuts expire...

Top tax rate



- Правдивый тренд (разница не такая значительная)



- Искажение шкалы (не с нуля)

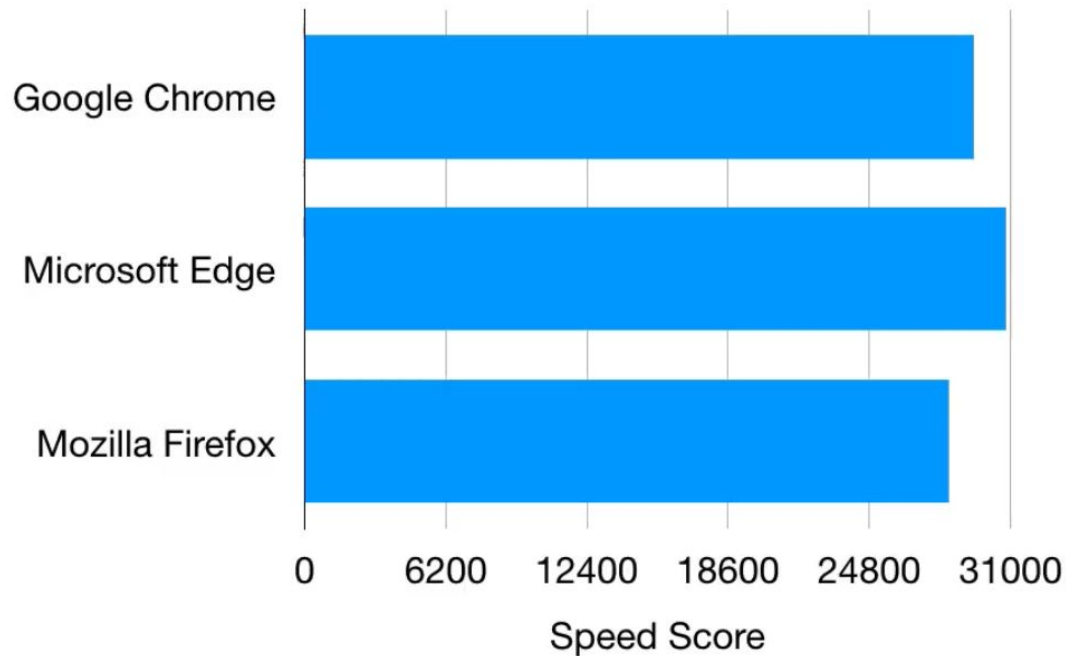
Welcome to Microsoft Edge

Faster than both Chrome and Firefox — according to Google's benchmark.

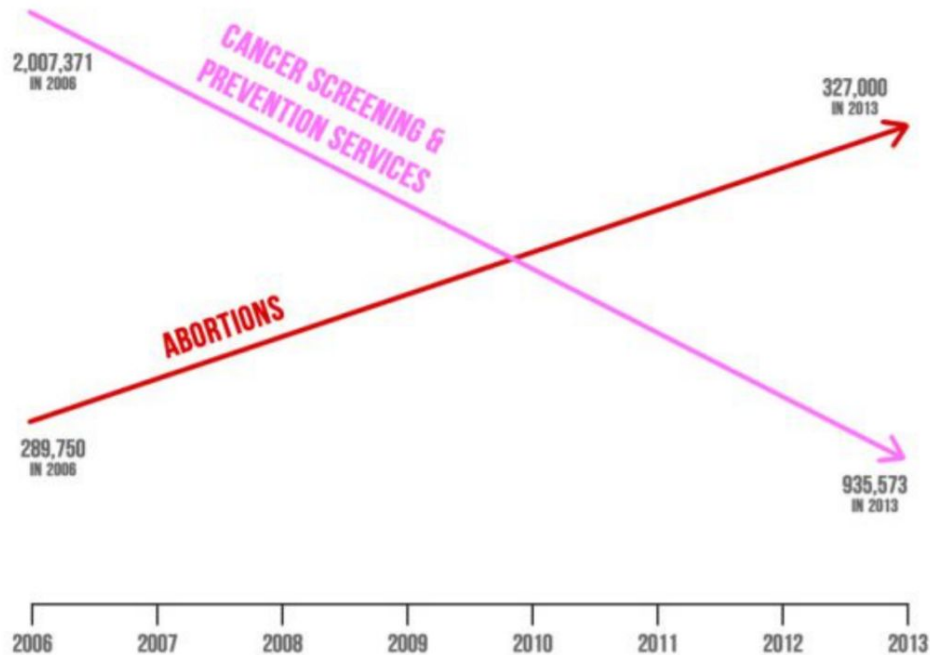


- Нарушение пропорций

- Если мы построим “честный график”, то правда, что разница такая значительная?



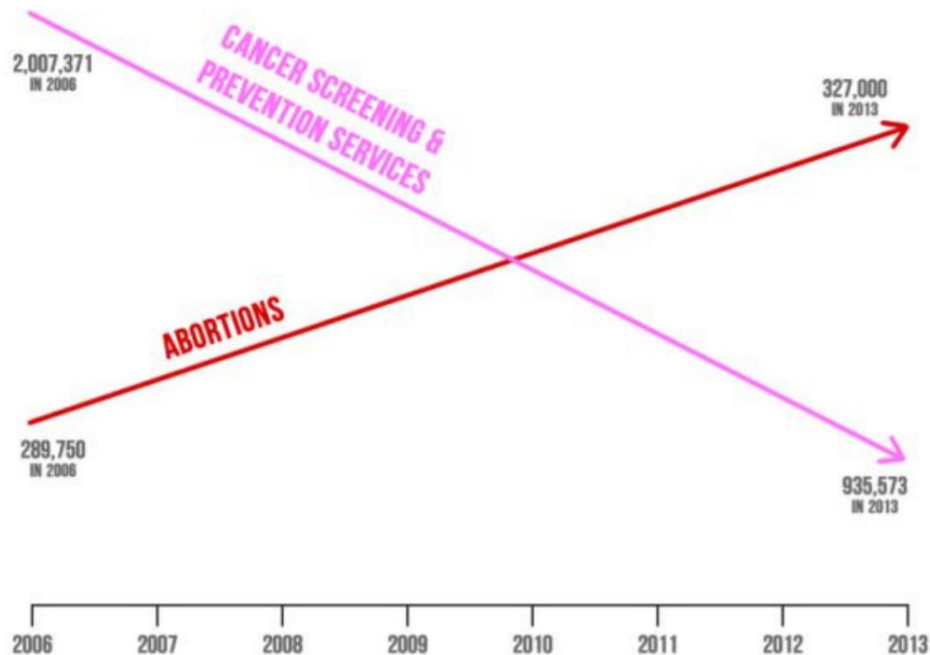
PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP — LIFE-SAVING PROCEDURES DOWN



SOURCE: AMERICANS UNITED FOR LIFE

Rep. Jason Chaffetz, Utah, 2015

PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP — LIFE-SAVING PROCEDURES DOWN

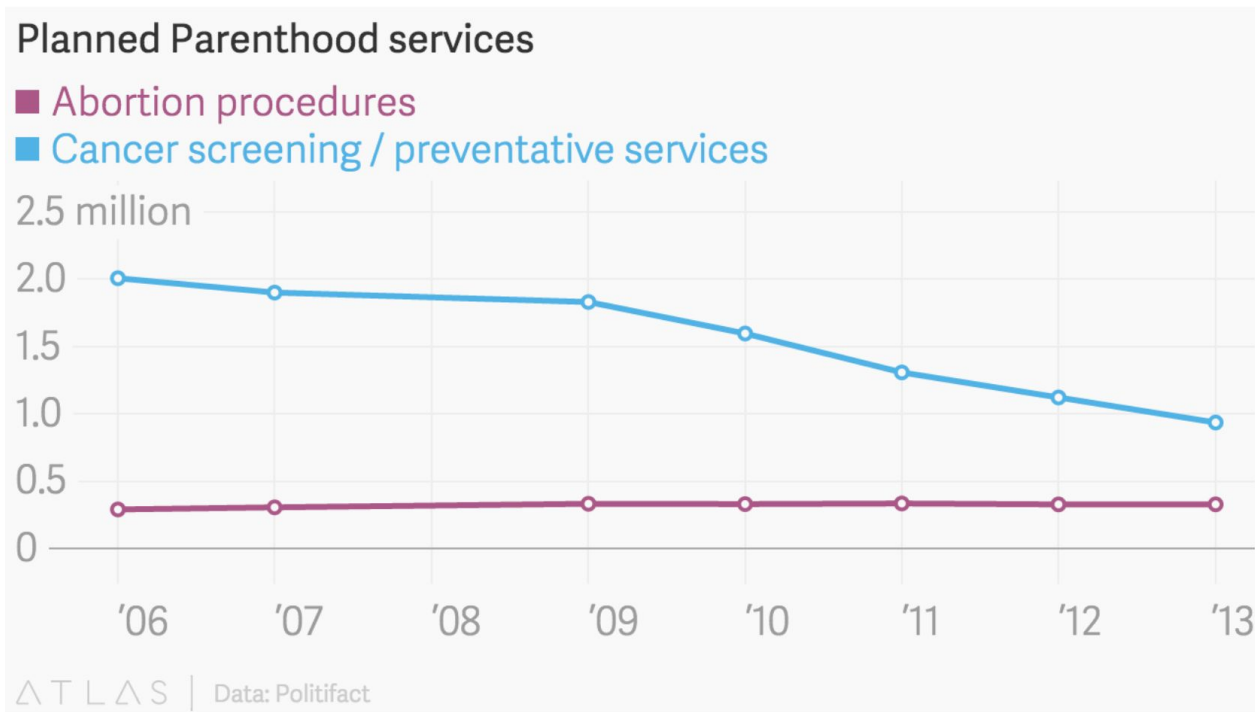


SOURCE: AMERICANS UNITED FOR LIFE

- Неудачное и непропорциональное совмещение шкал

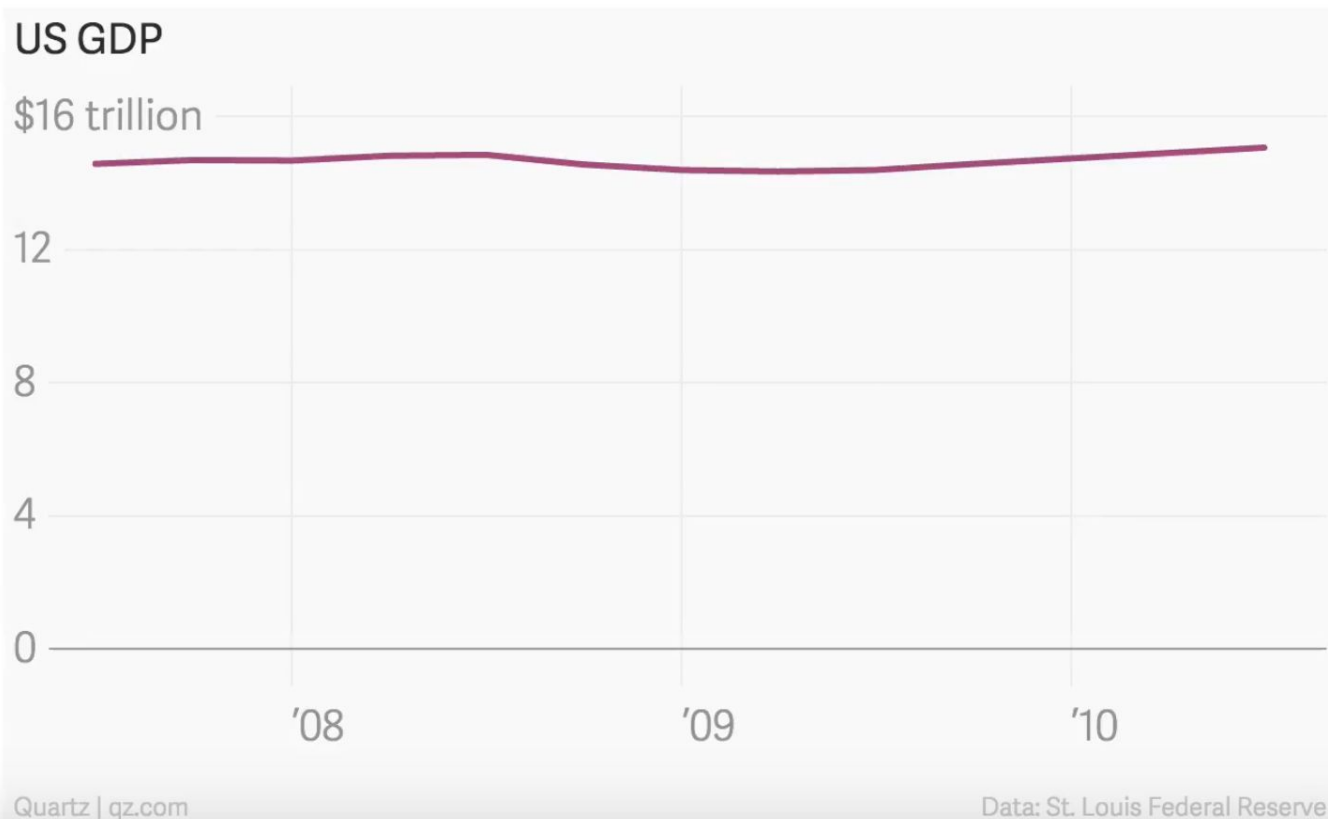
Rep. Jason Chaffetz, Utah, 2015

- И честный график.

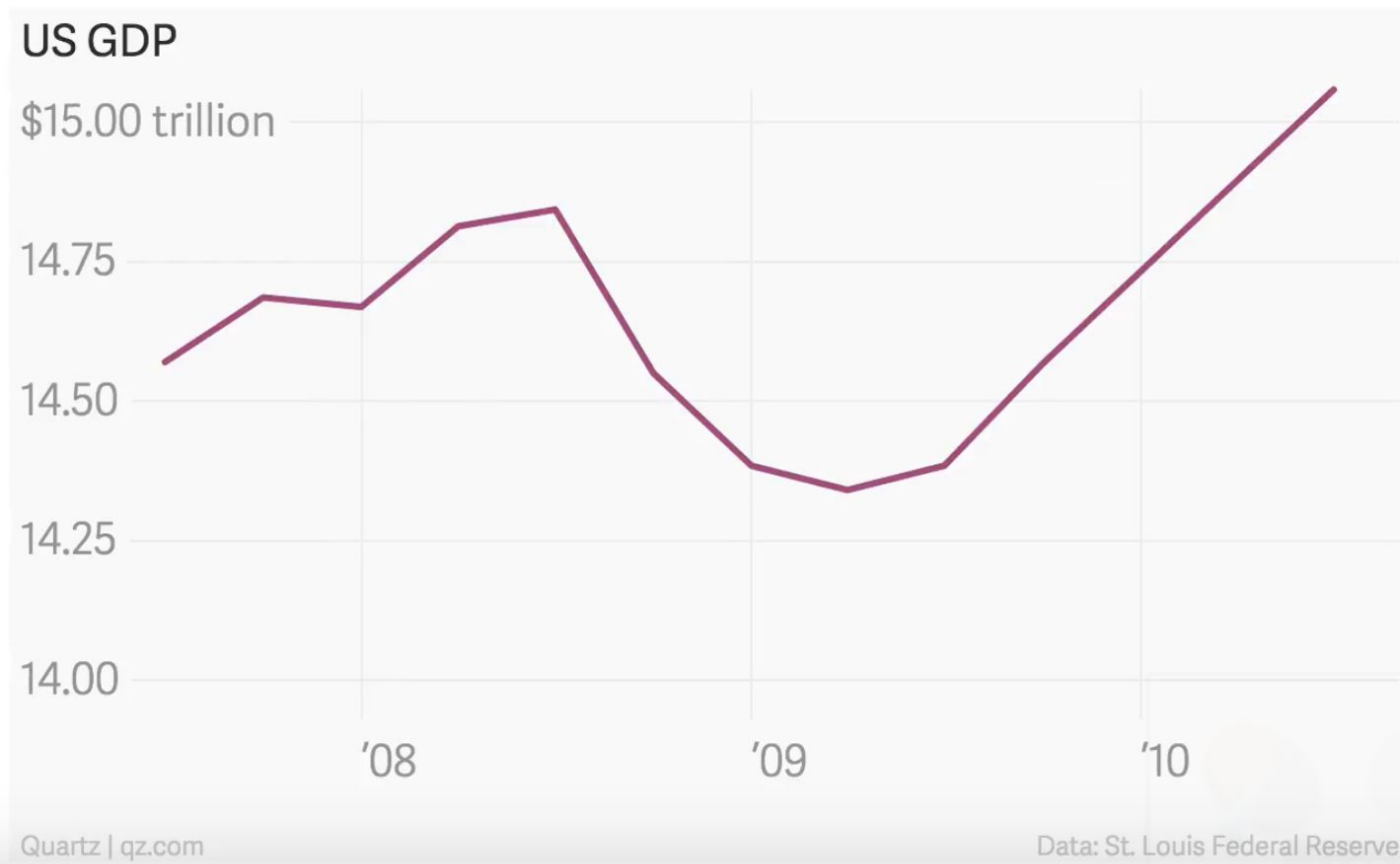


Politifact, 2015

- Слишком общая шкала (нужно приближать данные)



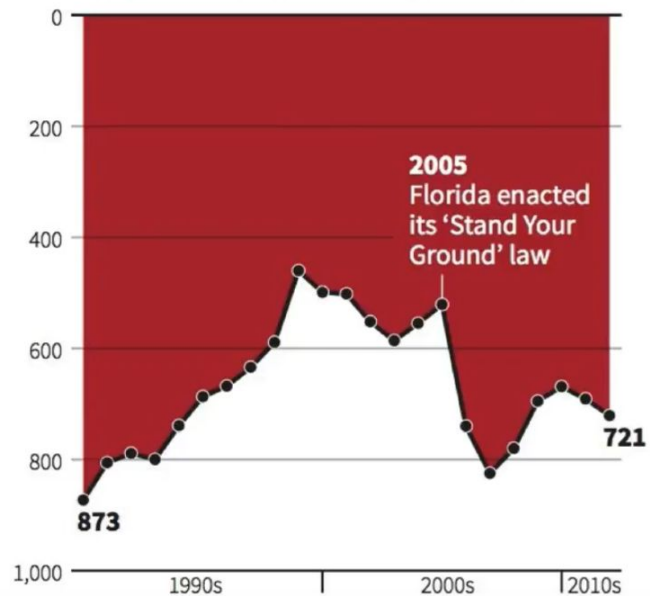
- На самом деле тренд другой!



- Не меняйте ориентацию шкал

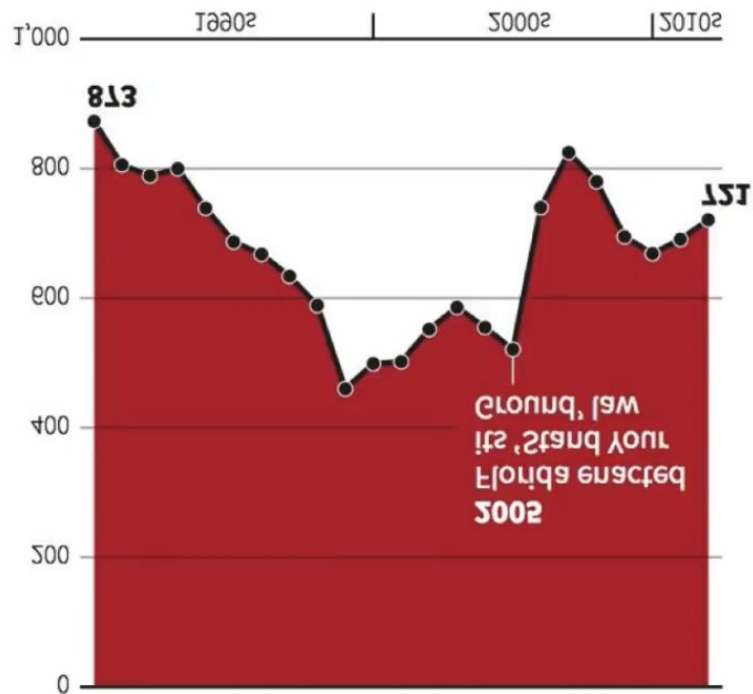
Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

SOURCE: Florida Department of Law Enforcement

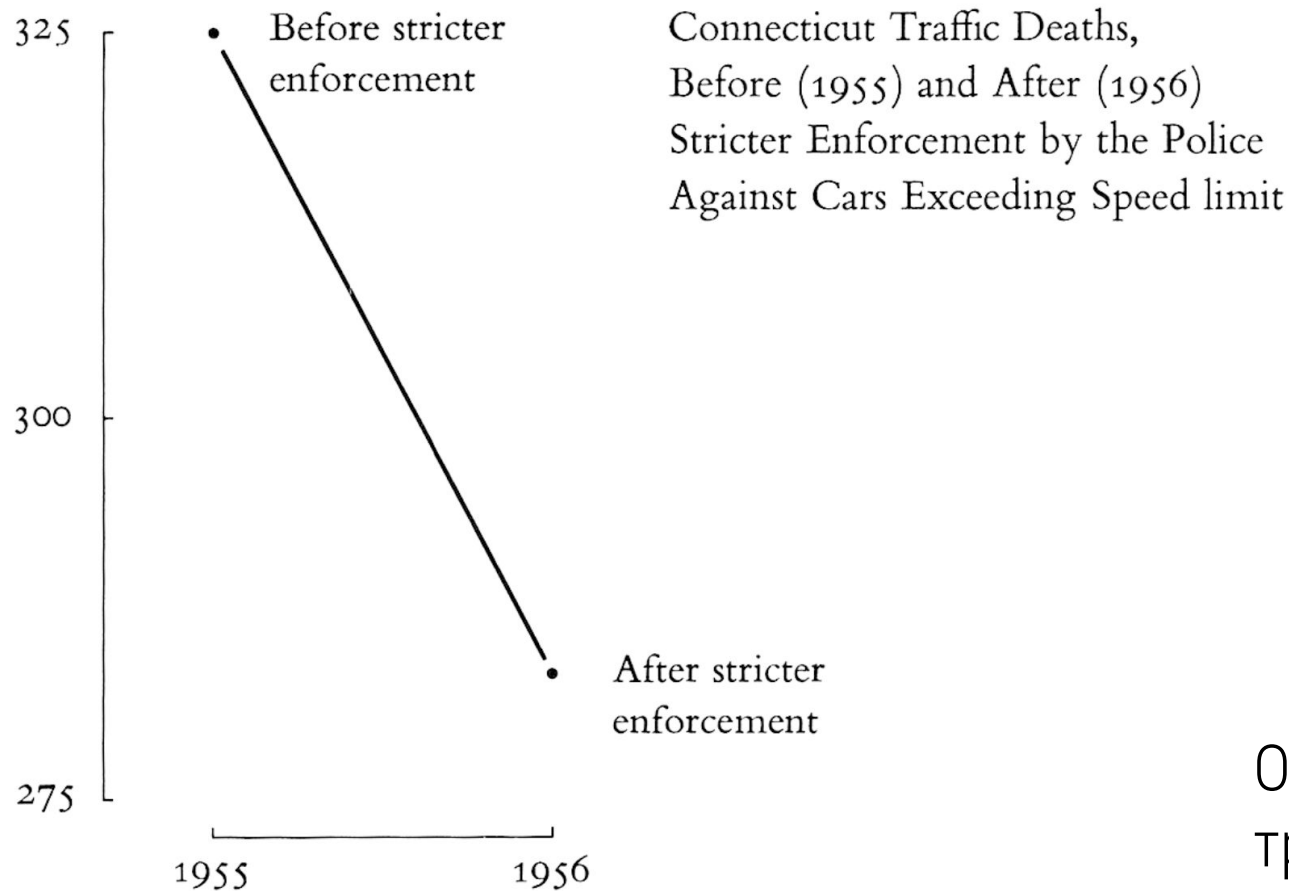


Number of murders committed using firearms

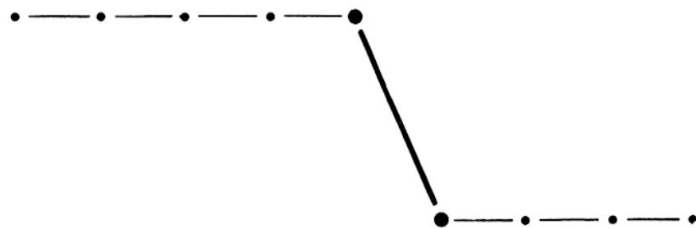
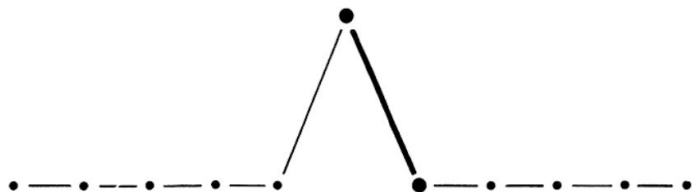
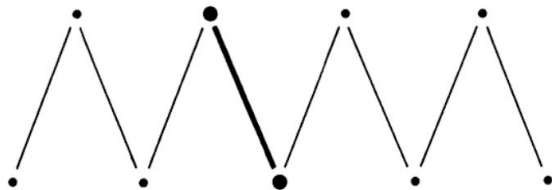
Gun deaths in Florida

Честность данных

- Отсутствующие шкалы и лэйблы
- Отсутствующий контекст
- Искаженные шкалы
- Искажающий дизайн



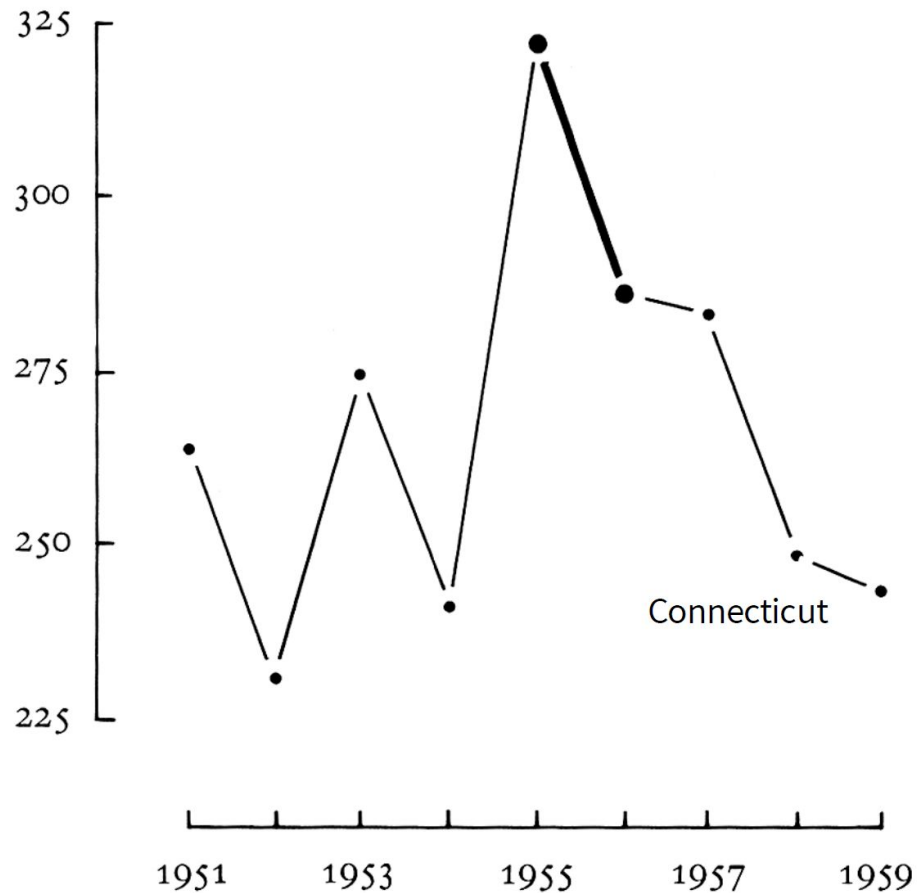
Опишите
тренд.



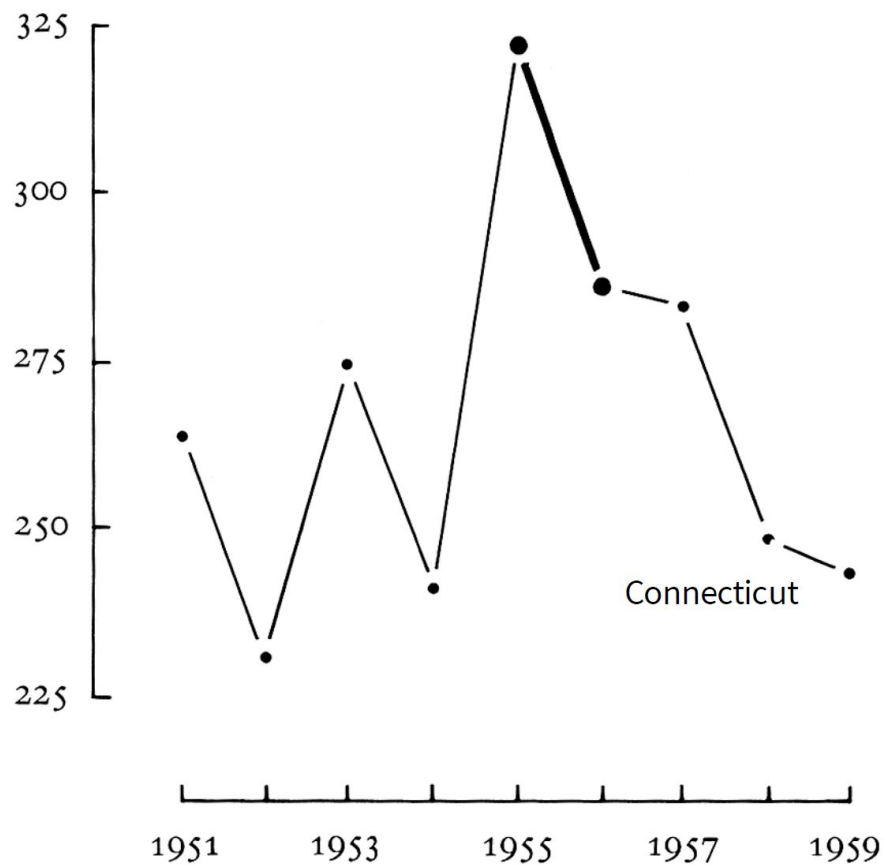
VDQI Example (p74)

Откуда взяли этот
сегмент?

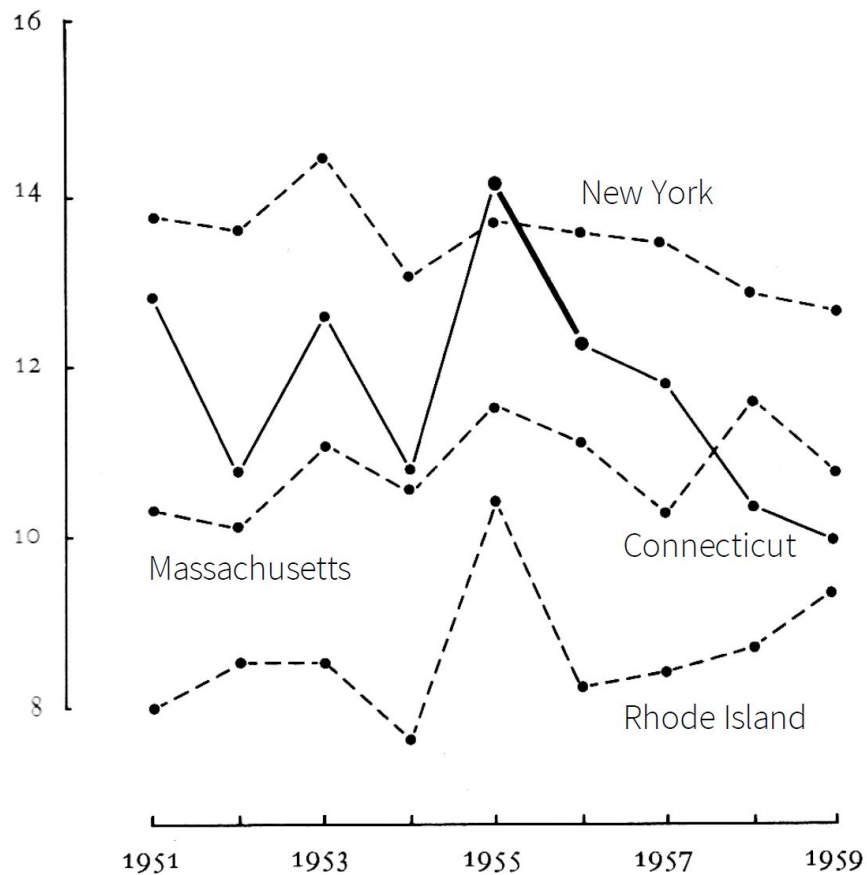
Откуда угодно!



Класс!
Появился контекст.
Опишите тренд снова.
Это много или мало?

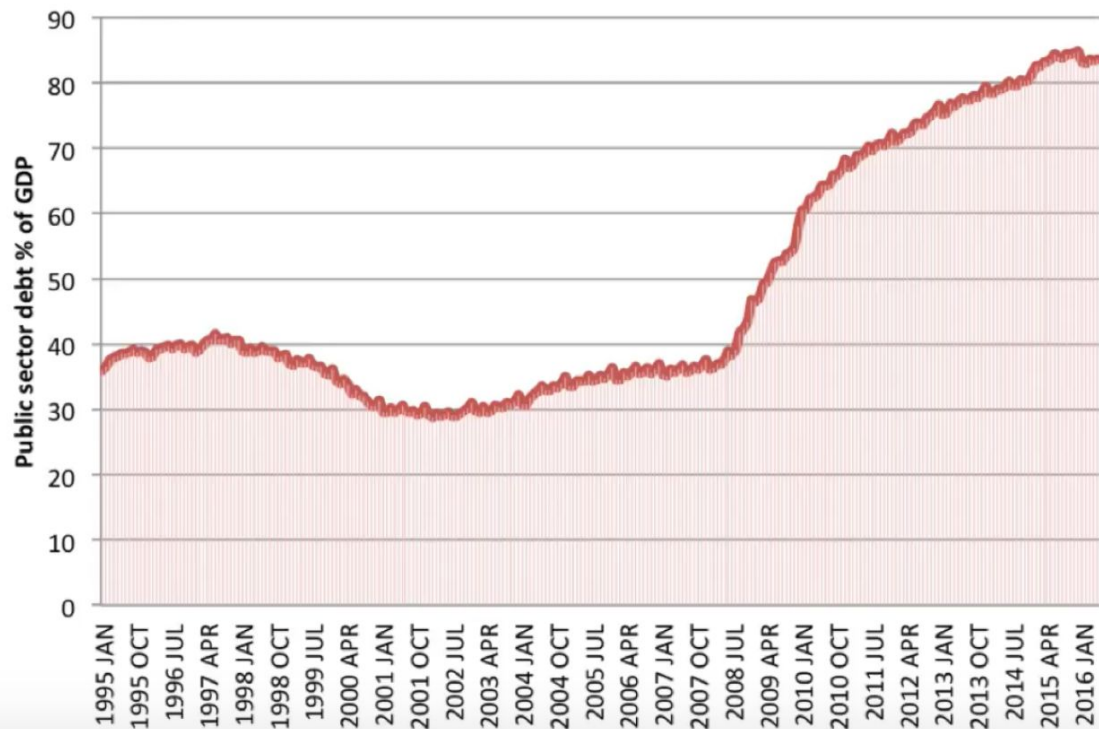


VDQI Example (p74)



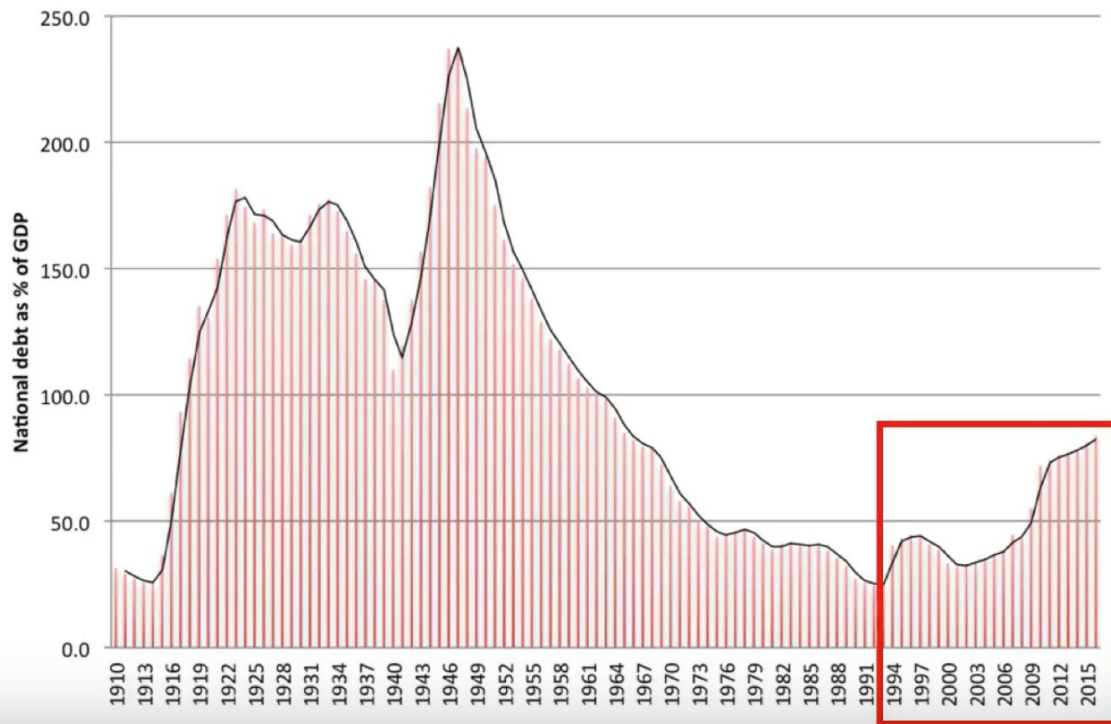
- Не упускайте контекст

UK national debt % GDP

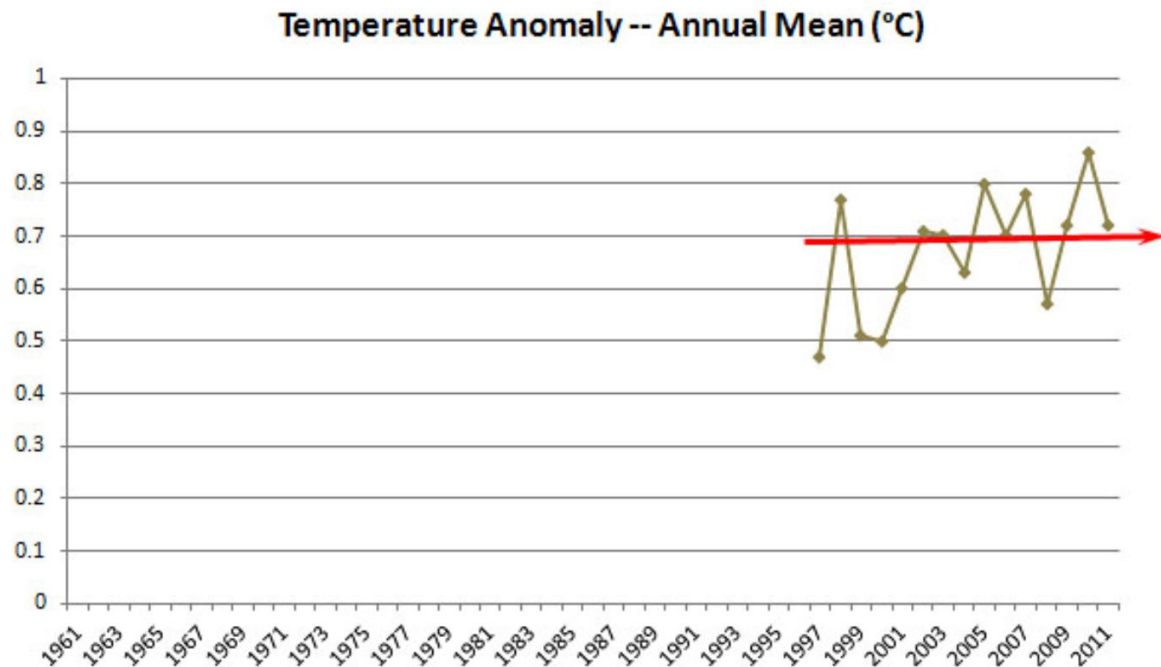


- Не упускайте контекст

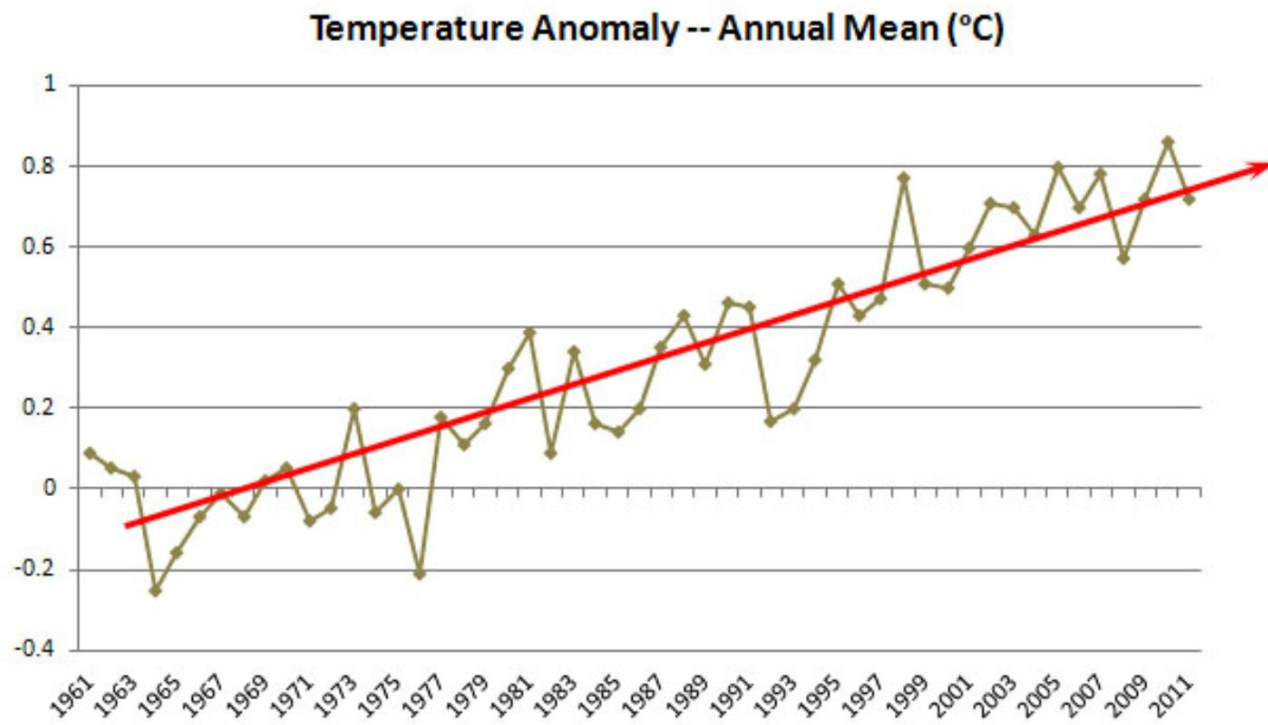
UK National debt



Существует ли глобальное потепление?



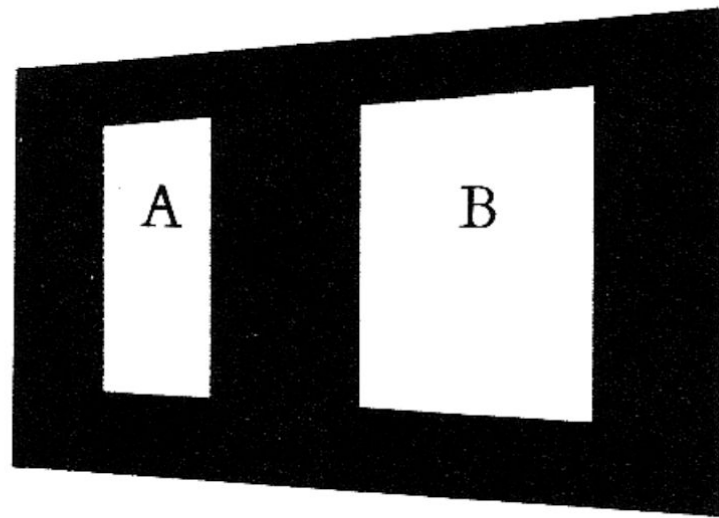
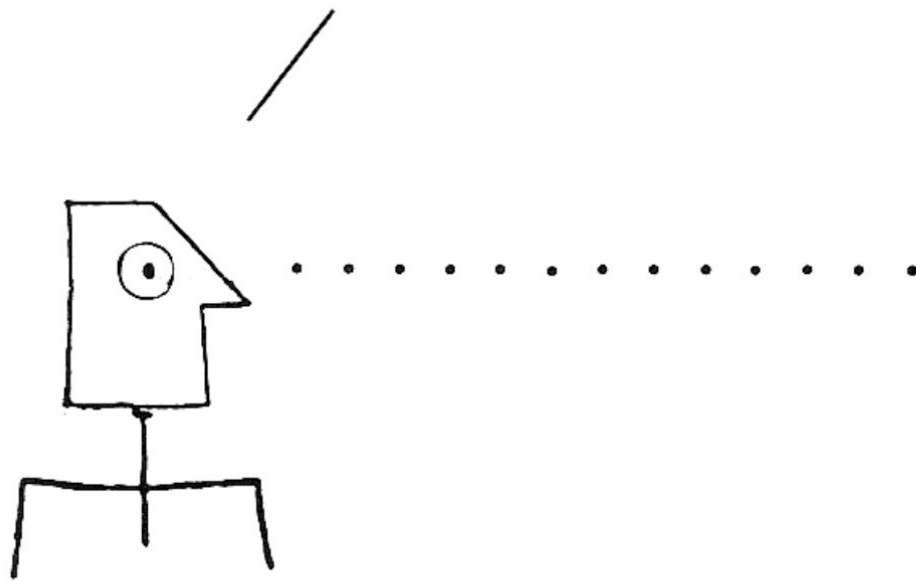
А так?



Честность данных

- Отсутствующие шкалы и лэйблы
- Отсутствующий контекст
- Искаженные шкалы
- Искажающий дизайн

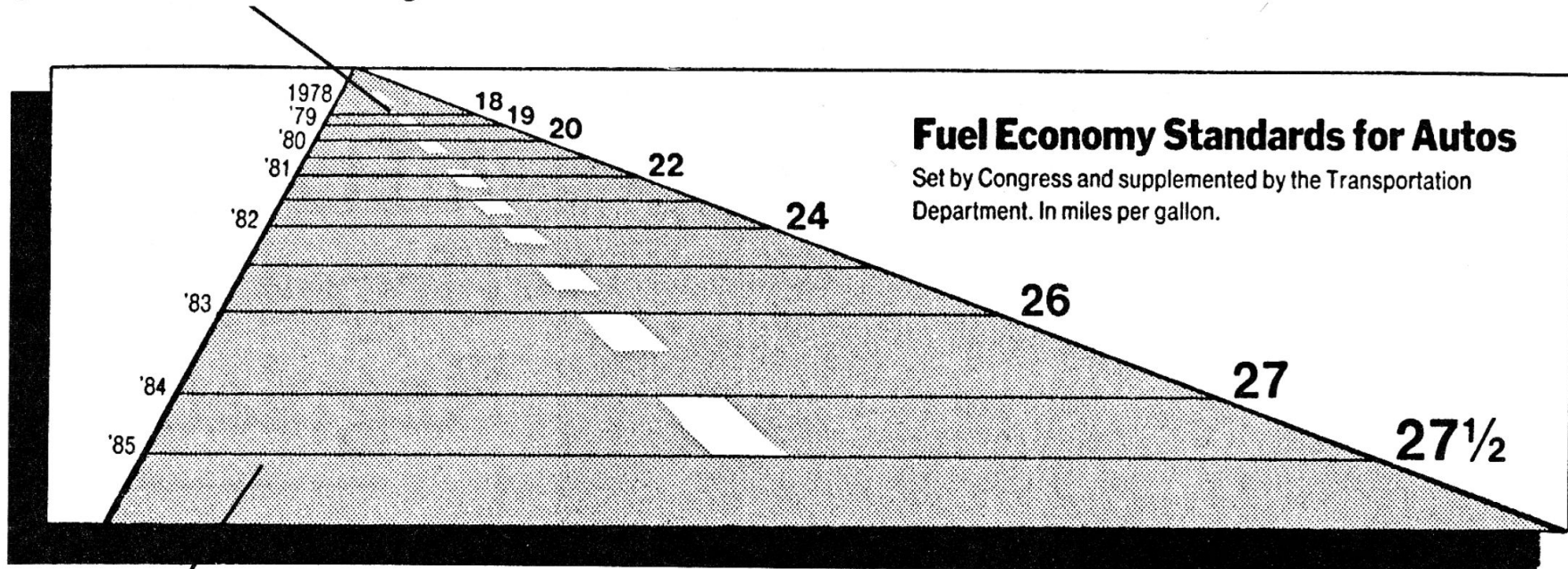
I think I see that area B
is 3.14 times bigger than
area A. Is that correct?



Искажение

- Физическое представление должно быть прямо пропорционально численным величинам
- Использовать ясное и детальное обозначение
- Отмечать важные моменты в данных

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



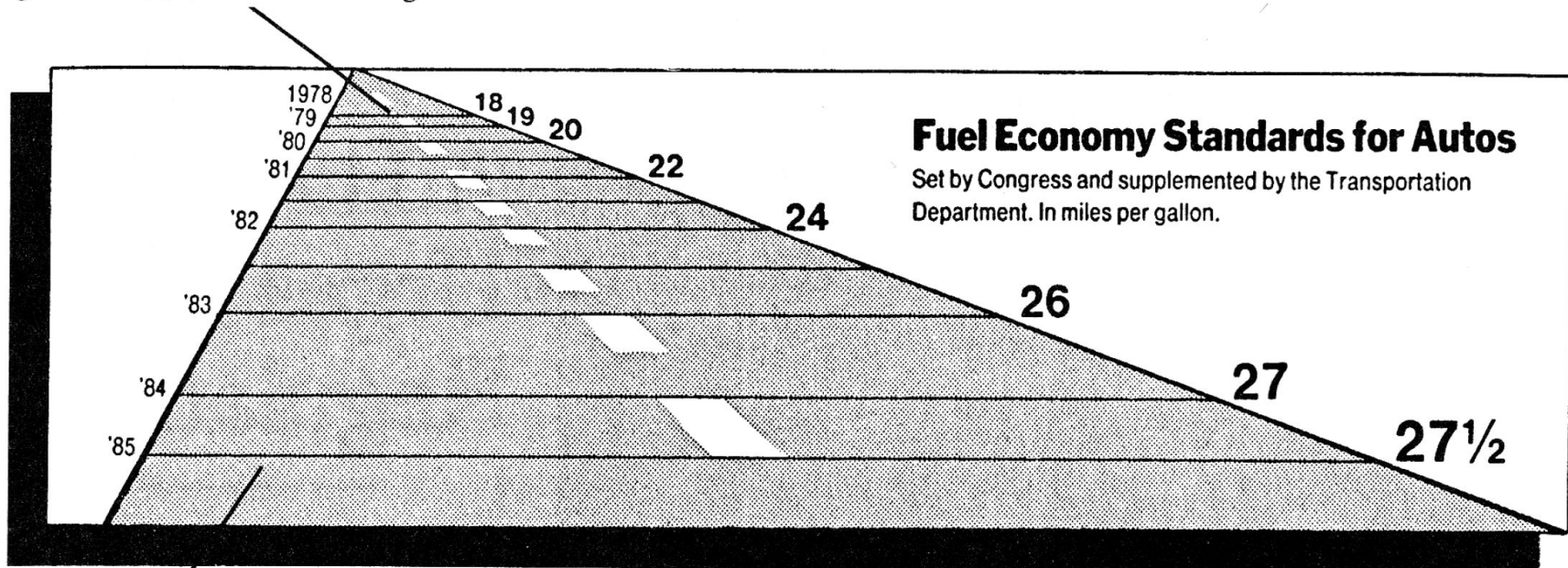
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

Фактор лжи

Фактор лжи =
размер эффекта, показанный на графике/
размер эффекта в данных

Фактор лжи должен быть ~ 1
 $1.05 > \text{Фактор лжи} > 0.95$

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



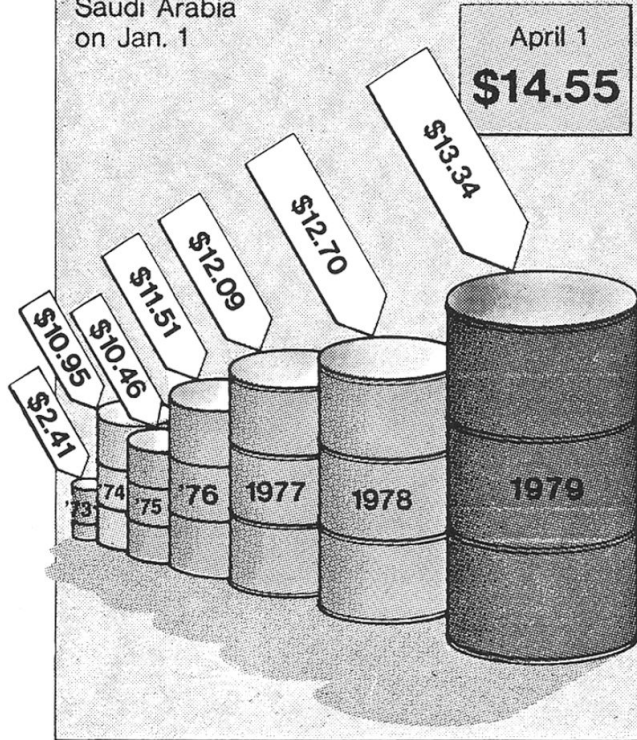
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

$$\frac{27.5 - 18.0}{18.0} = 0.53 \quad \frac{5.3 - 0.6}{0.6} = 7.83$$

$$\text{Фактор лжи} = \frac{7.83}{0.53} = 14.8$$

IN THE BARREL...

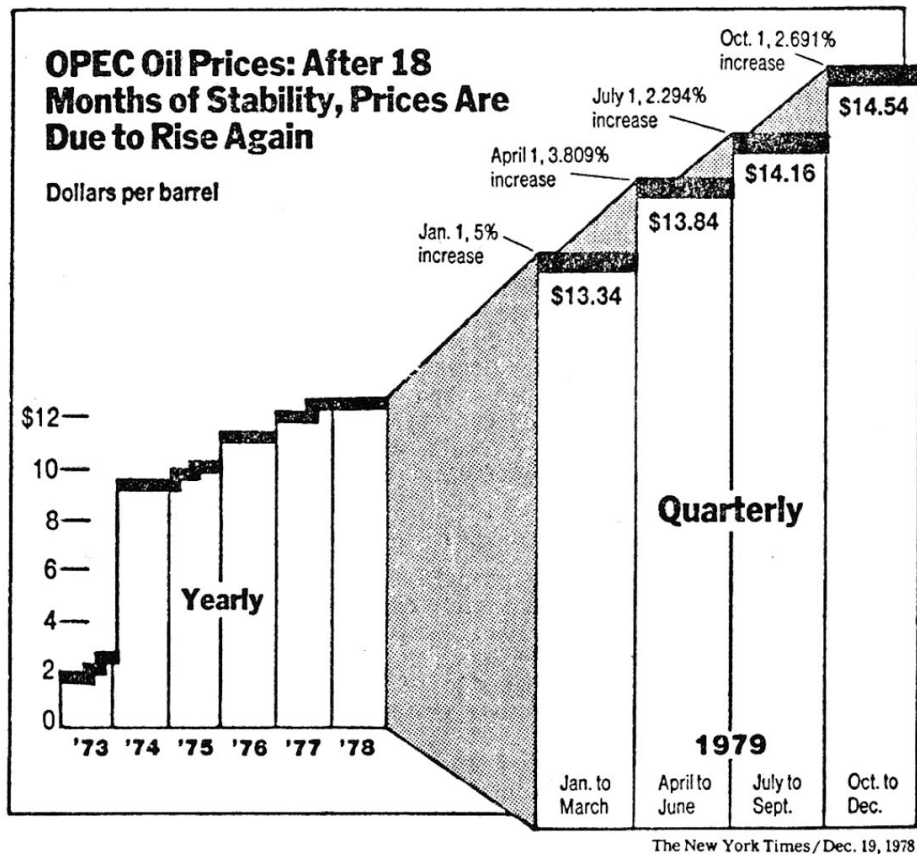
Price per bbl. of
light crude, leaving
Saudi Arabia
on Jan. 1



- Фактор лжи площади поверхности: 9.4
- Фактор лжи объема: 59.4
- Не используйте 3D

Честность данных

- Отсутствующие шкалы и лэйблы
- Отсутствующий контекст
- Искаженные шкалы
- Искажающий дизайн



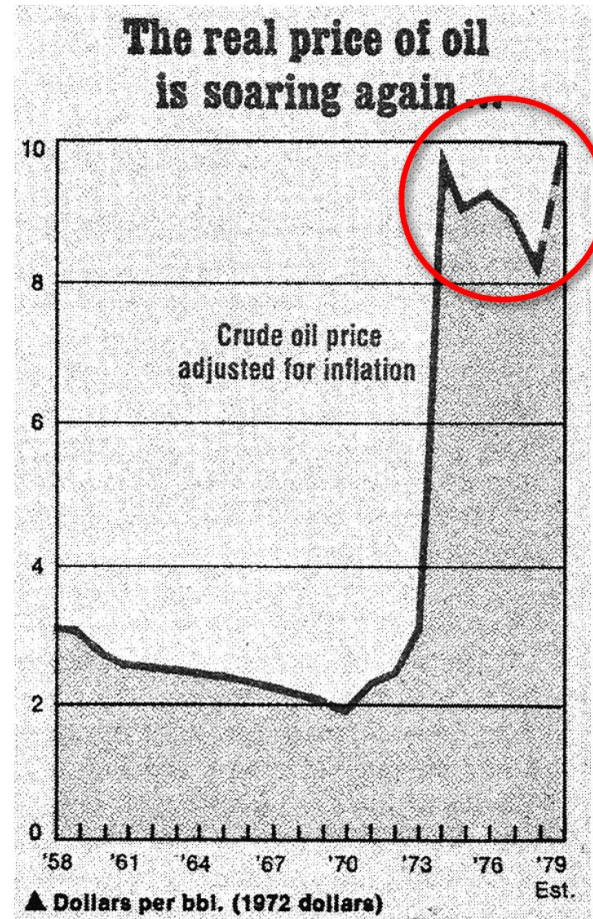
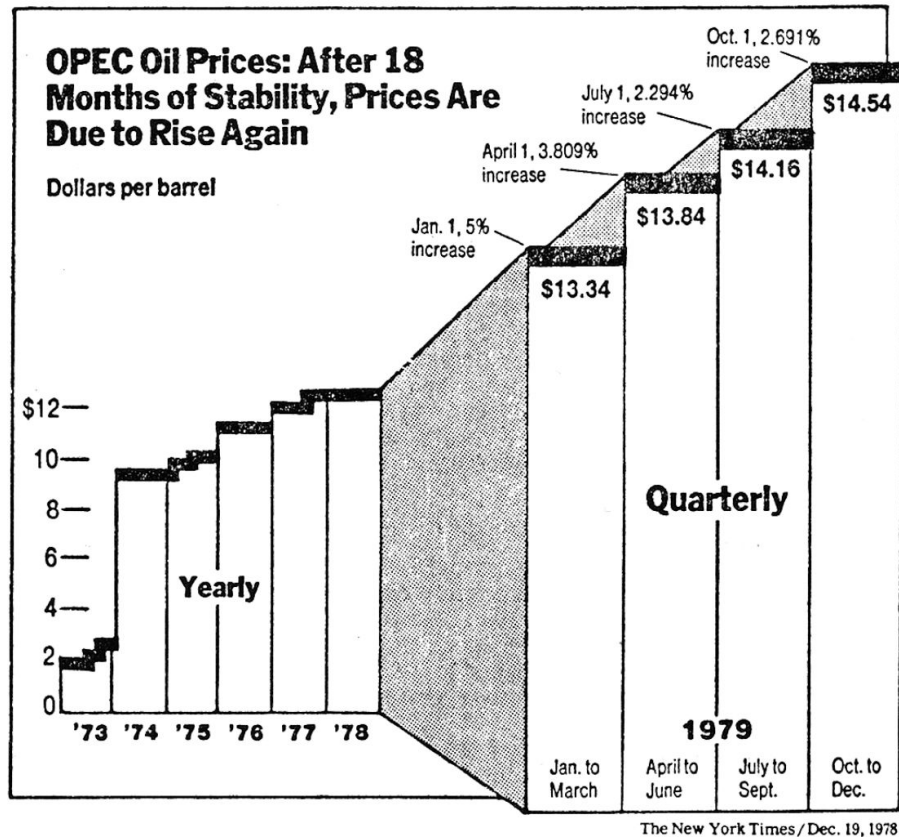
VDQI Example (p61)

- Вертикальная ось
С 1973 по 1978:
\$8.00 на дюйм

1979:
\$4.00 на дюйм
- Горизонтальная ось
С 1973 по 1978:
3.8 года на дюйм

1979:
0.57 лет на дюйм

Показывайте разницу в данных,
а не разницу в дизайне!



Comparative Annual Cost per Capita for care of Insane in
Pittsburgh City Homes and Pennsylvania State Hospitals.

\$147



South Mountain

\$172



Pittsburgh

\$198



Harrisburg

\$213



Norristown

\$214



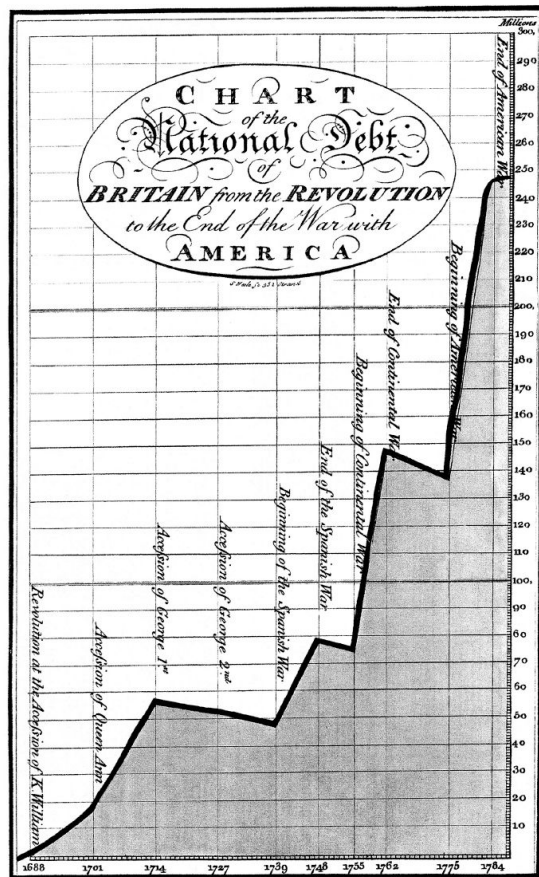
Warren

Pittsburgh Civic Commission, *Report on
Expenditures of the Department of Charities*
(Pittsburgh, 1911), p. 7.

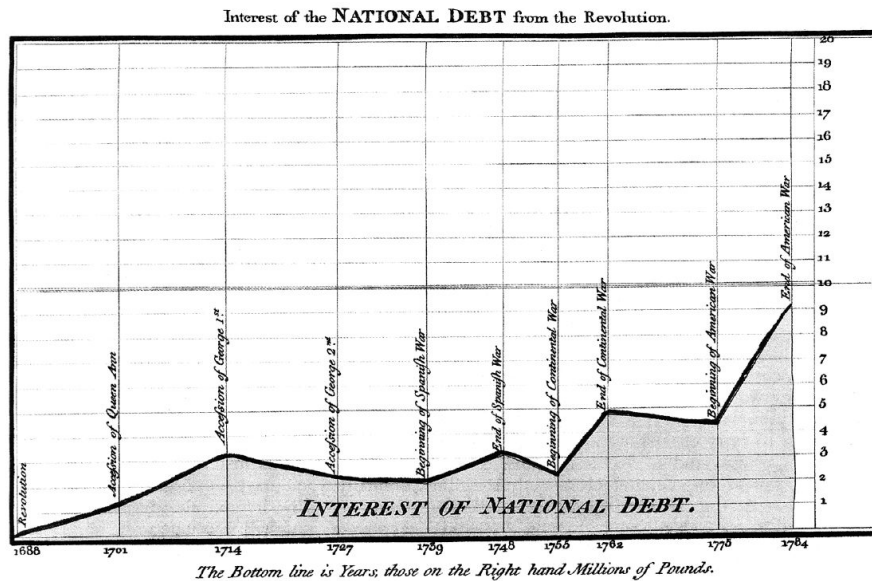
Честность данных

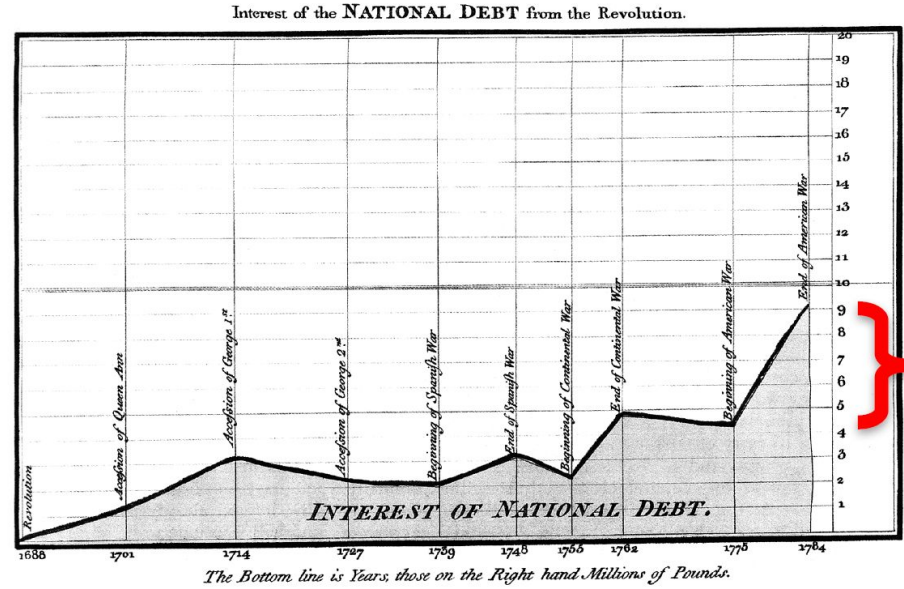
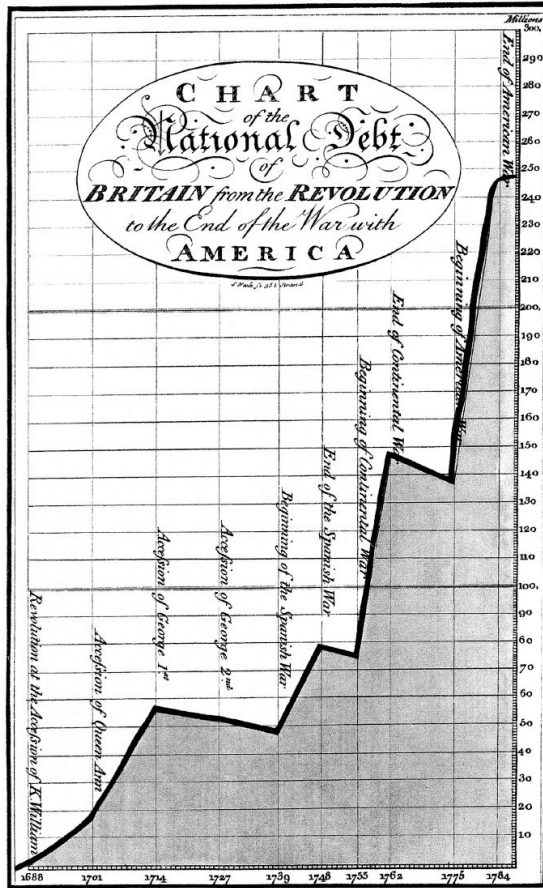
Принципы:

- Вертикальные графики в отличие от горизонтальных подчеркивают рост
- Горизонтальные графики в отличие от вертикальных подчеркивают время
- Разные уловки подчеркивают разные аспекты
- Абсолютные величины в отличие от пропорций не дают контекста



VDQI Example (p64)

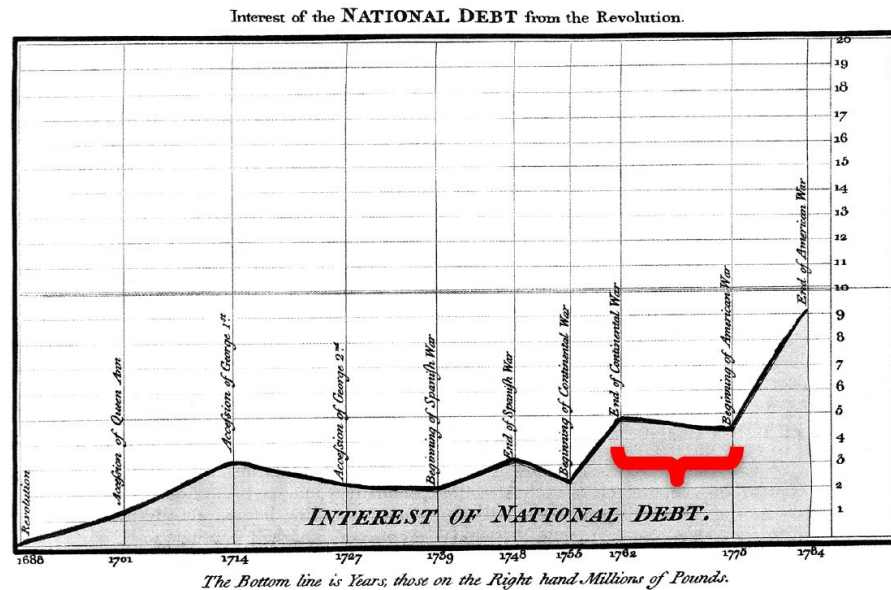
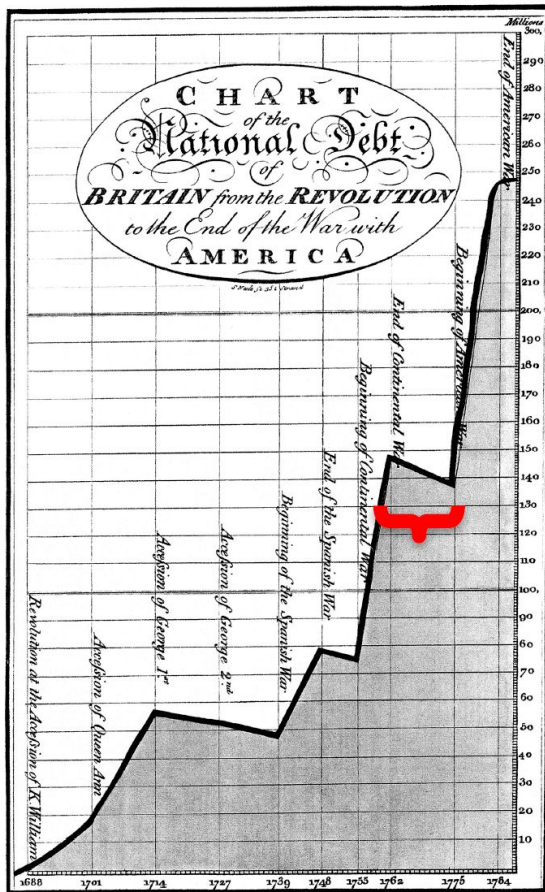




Честность данных

Принципы:

- Вертикальные графики в отличие от горизонтальных подчеркивают рост
- Горизонтальные графики в отличие от вертикальных подчеркивают время
- Разные уловки подчеркивают разные аспекты
- Абсолютные величины в отличие от пропорций не дают контекста



Честность данных

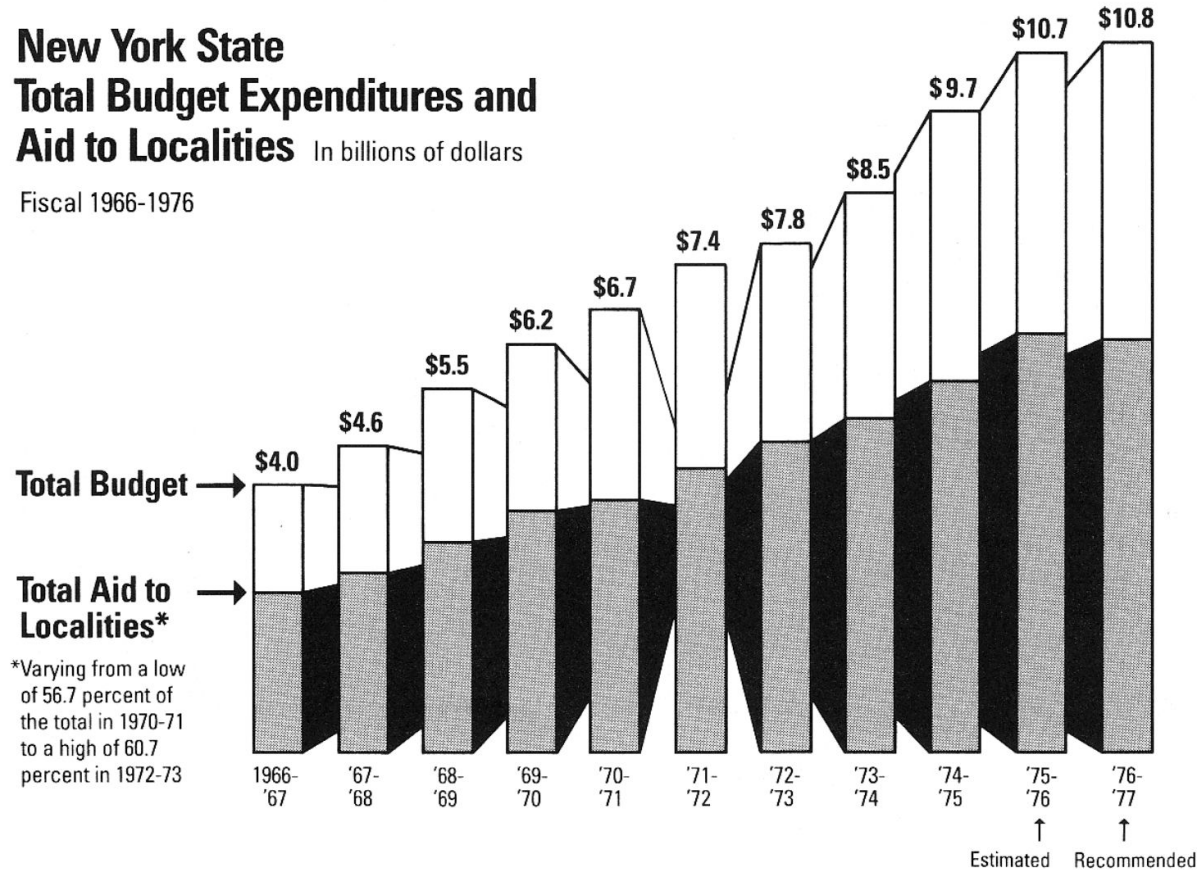
Принципы:

- Вертикальные графики в отличие от горизонтальных подчеркивают рост
- Горизонтальные графики в отличие от вертикальных подчеркивают время
- Разные уловки подчеркивают разные аспекты
- Абсолютные величины в отличие от пропорций не дают контекста

New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976



New York State Total Budget Expenditures and Aid to Localities

In billions of dollars

Fiscal 1966-1976

Без
поправки на
инфляцию

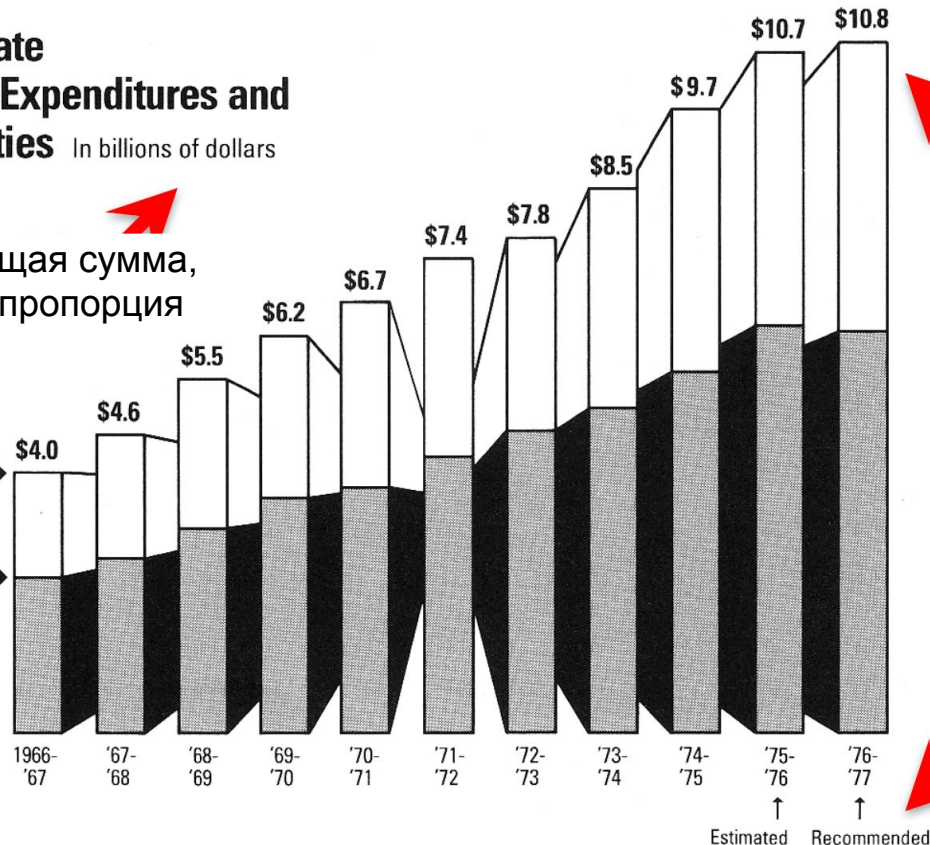
Общая сумма,
не пропорция

Из-за 3D новые
столбцы
возвышаются
над остальными

Total Budget

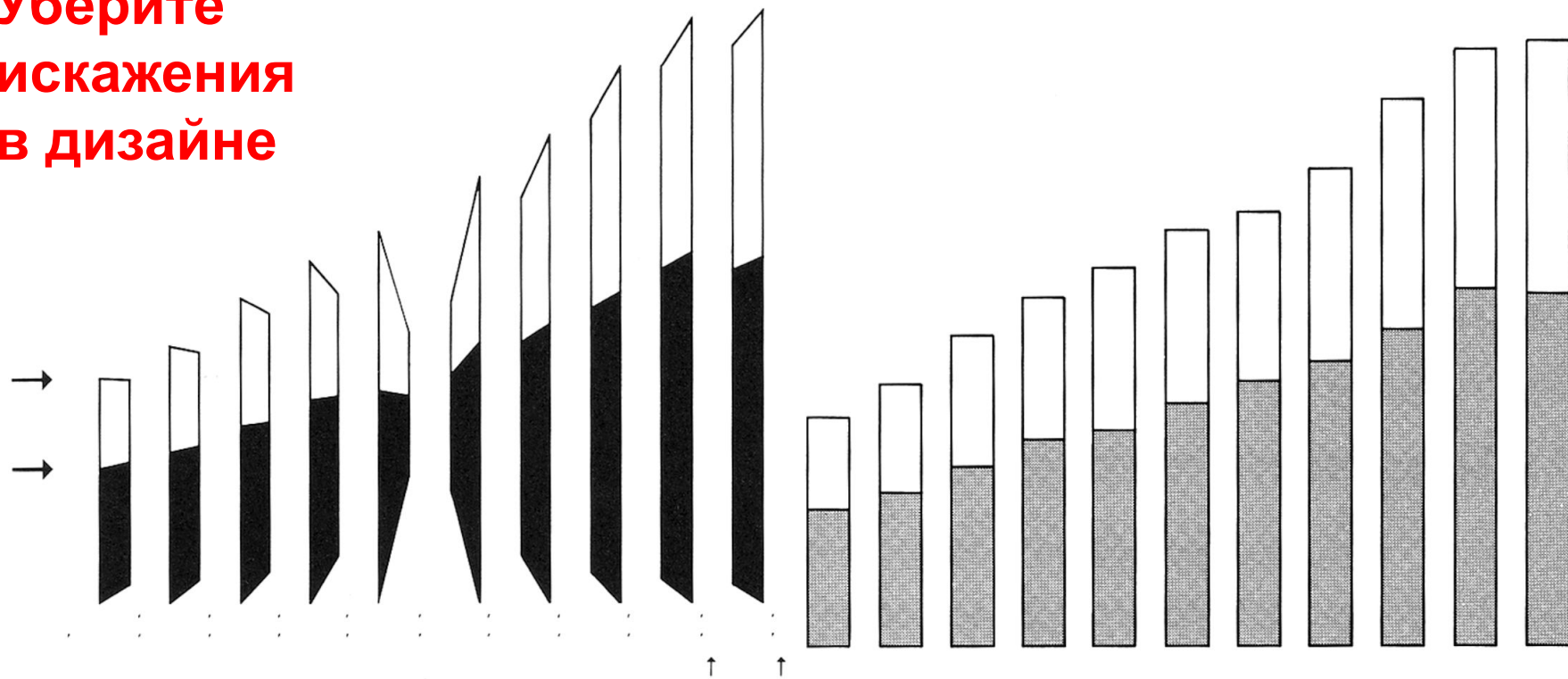
Total Aid to
Localities*

*Varying from a low
of 56.7 percent of
the total in 1970-71
to a high of 60.7
percent in 1972-73



Стрелки,
направленные
вверх,
подчеркивают
рост

Уберите искажения в дизайне

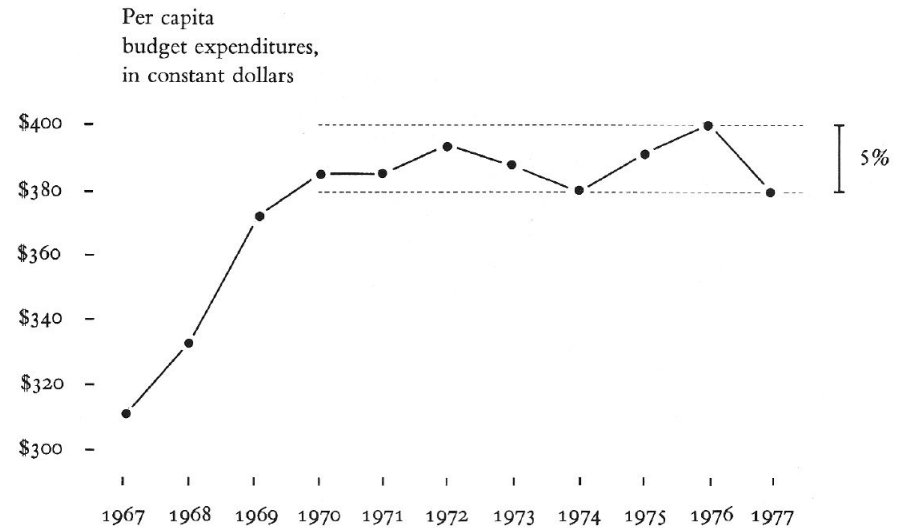
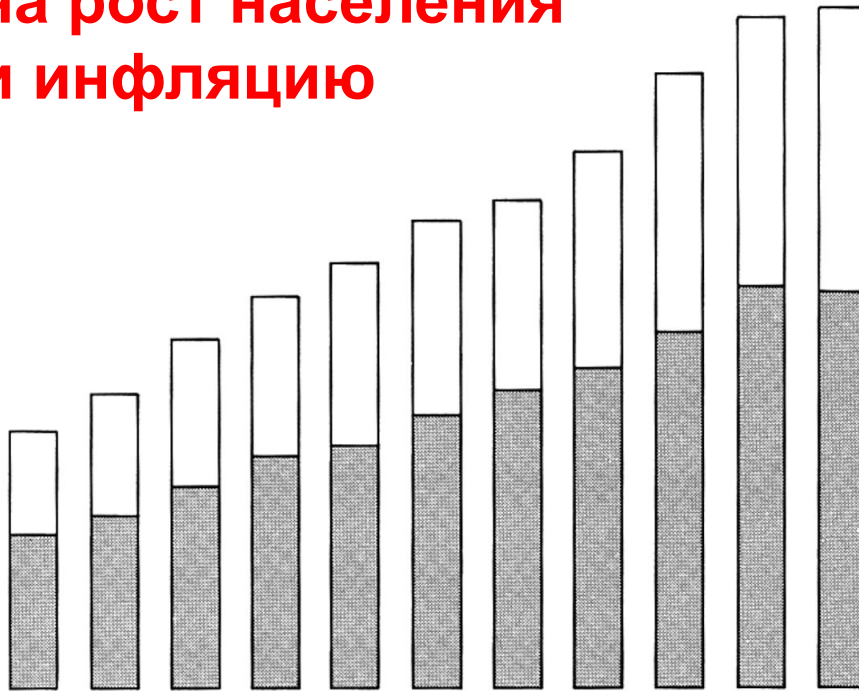


Честность данных

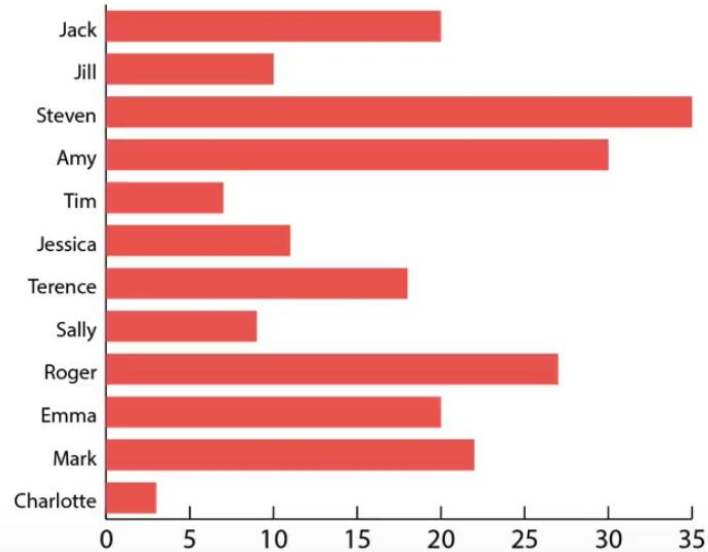
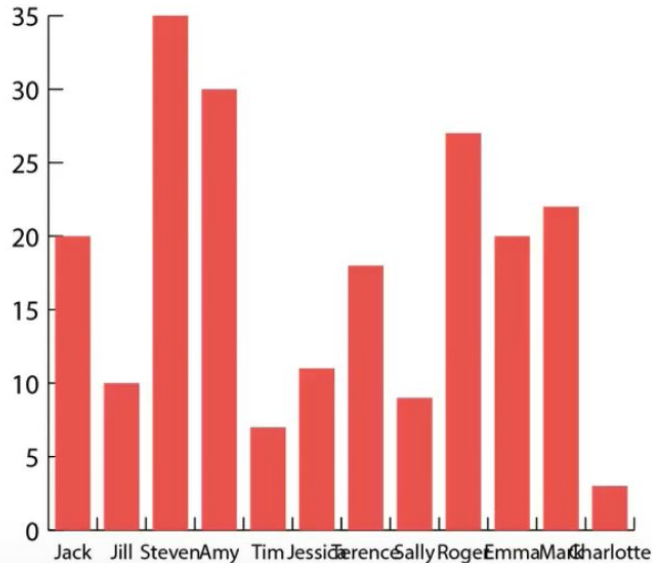
Принципы:

- Вертикальные графики в отличие от горизонтальных подчеркивают рост
- Горизонтальные графики в отличие от вертикальных подчеркивают время
- Разные уловки подчеркивают разные аспекты
- Абсолютные величины в отличие от пропорций не дают контекста

Делайте поправку на рост населения и инфляцию



Labelling can be problematic with vertical bars when there are a large number of bars. Therefore, it's better to use a (horizontal) bar chart



Lessons learned

- Избегайте акцентов, которые создаются вертикальными или горизонтальными графиками
- Избегайте уловок, которые излишне акцентируют данные
- Избегайте абсолютных данных, в отличие от пропорций
- Избегайте нескорректированных данных (например рубли, а не рубли поправкой на инфляцию)

Data Redesign

VDQI Глава 4

Прежде всего,
демонстрируйте
данные.

Данные–к–чернилам (Data–Ink Ratio)

Пропорция данных к чернилам =

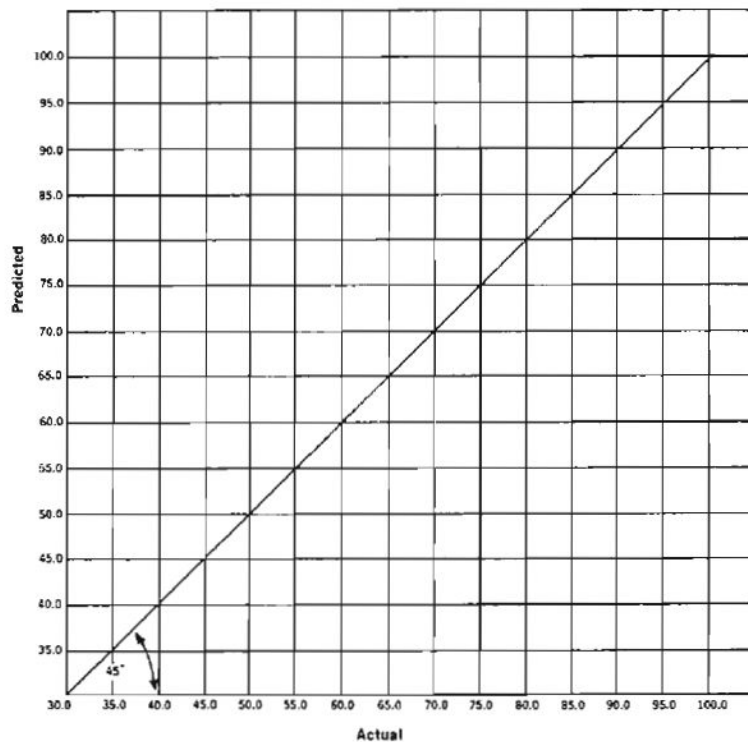
*данные/общее количество чернил,
использованных в графике =*

пропорция чернил в графике, потраченных на
неизбыточное отображение данных =

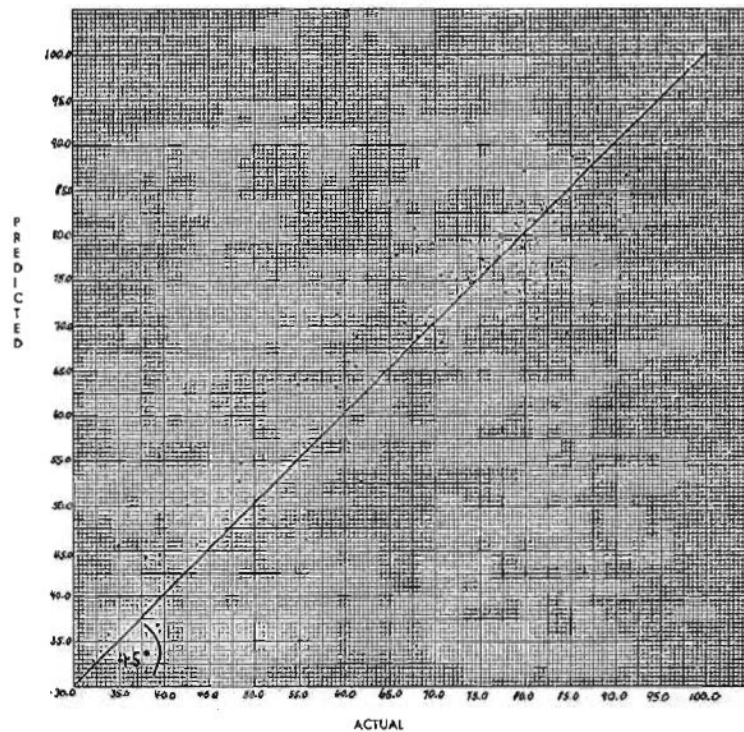
1.0 – пропорция графика, которую можно стереть без
потери данных

Данные-к-чернилам: низкая пропорция

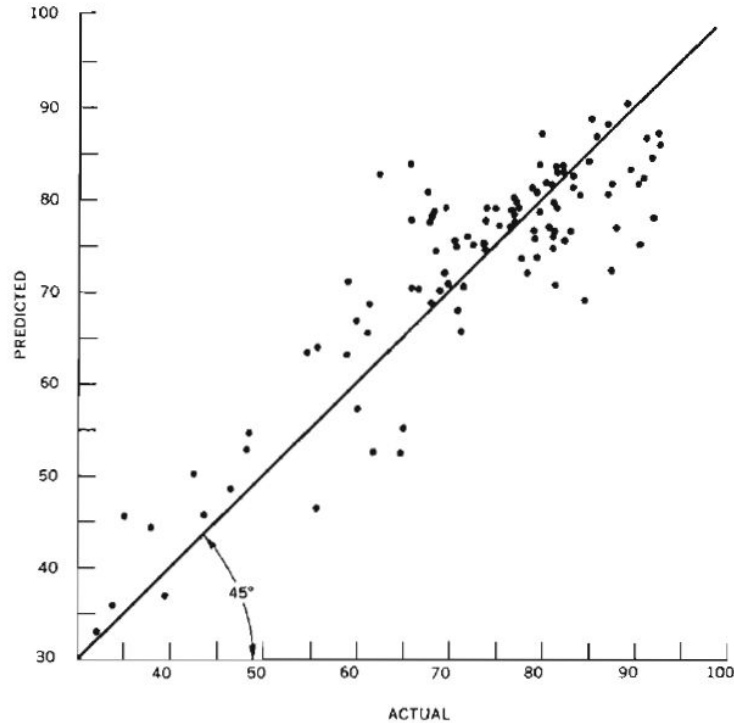
Figure 19.1 Relationship of Actual Rates of Registration to Predicted Rates
(104 cities, 1960)



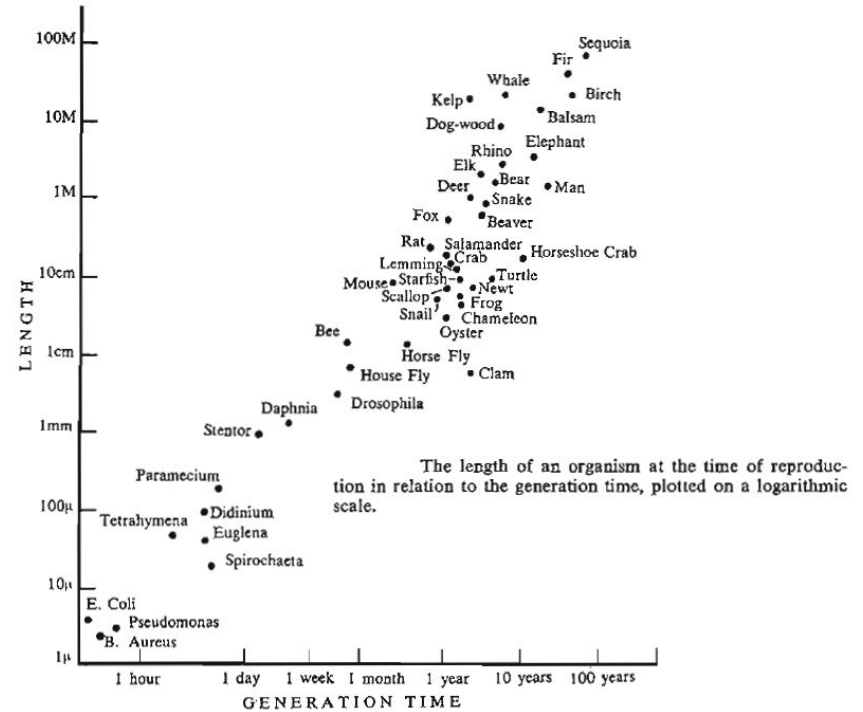
Relationship of Actual Rates of Registration to Predicted Rates
(104 cities 1960).



Данные-к-чернилам: высокая пропорция



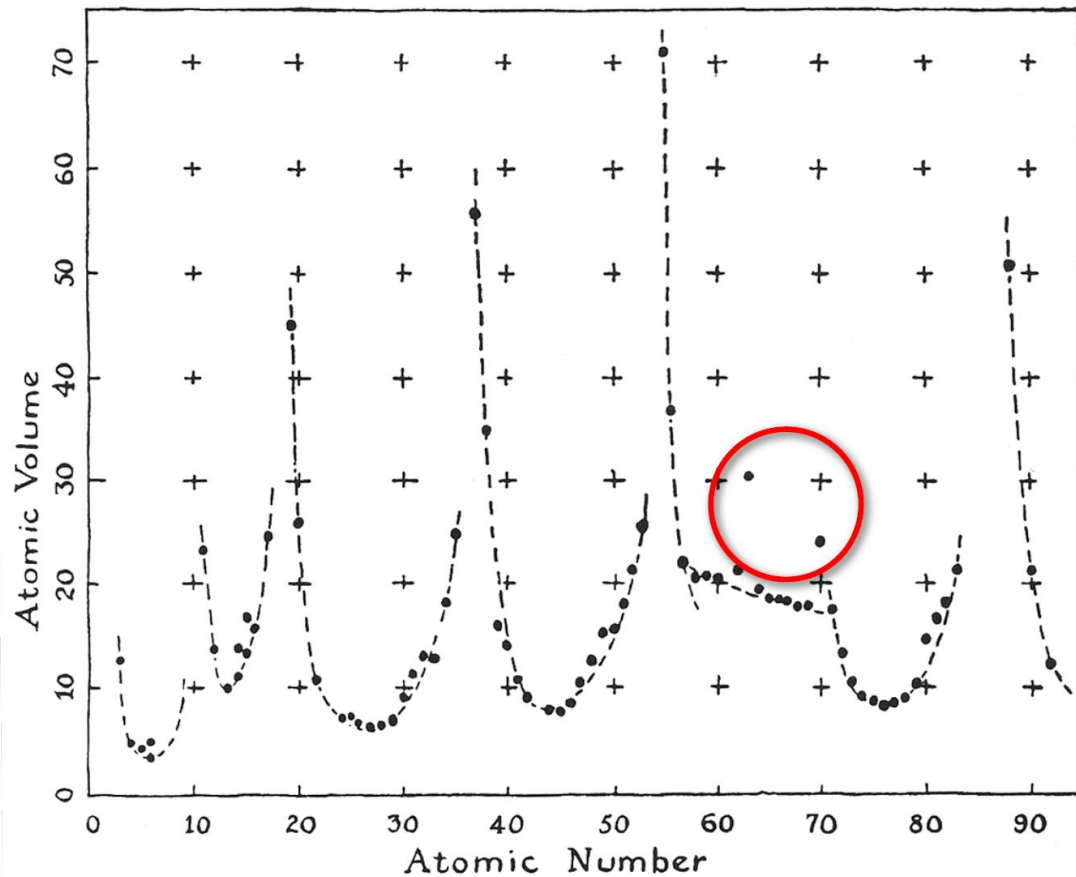
Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).



Редизайн графики

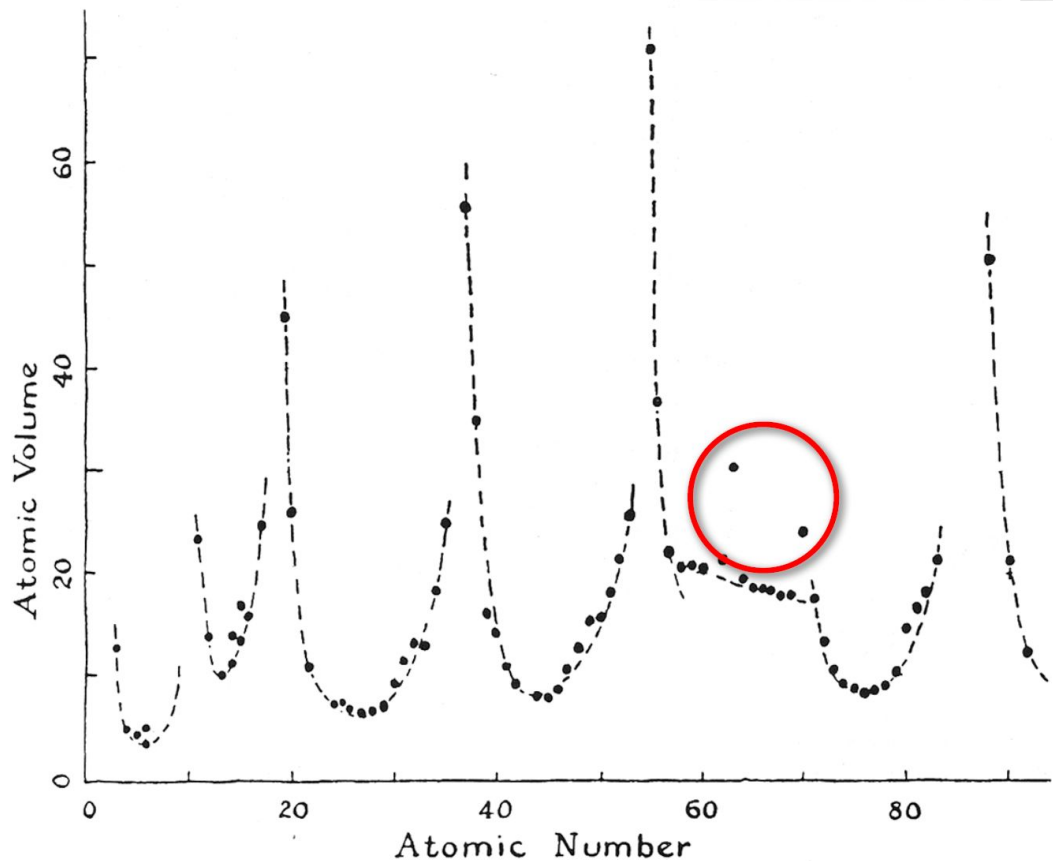
- Максимизируйте соотношение данных-к-чернилам в разумных пределах.
- Убирайте чернила-без-данных, в разумных пределах.
Немного орнамента, маркировка осей и т.д. вполне допустимы.
- Убирайте избыточные чернила-данные, в разумных пределах.
Некоторая избыточность полезна.
- Проверяйте и правьте.

**data-ink:
low**



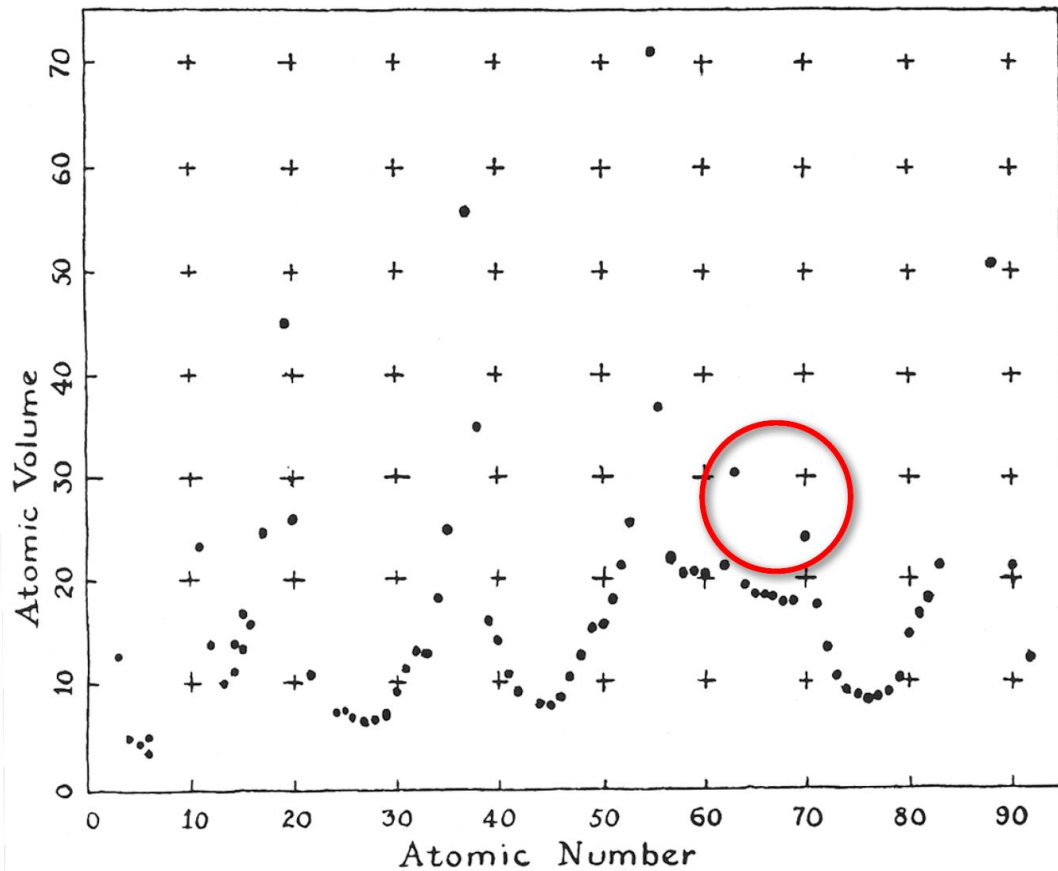
VDOI Example (p102)

Убрать
шкалу?



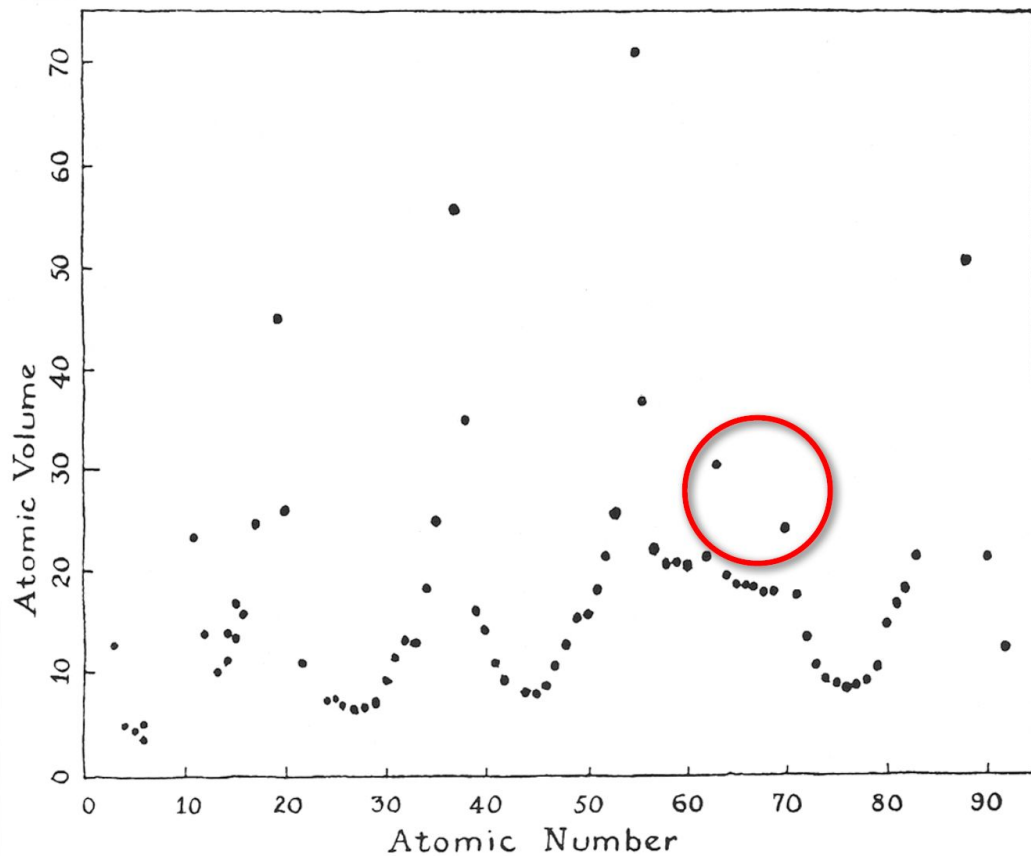
VDOI Example (p102)

Убрать
линии?



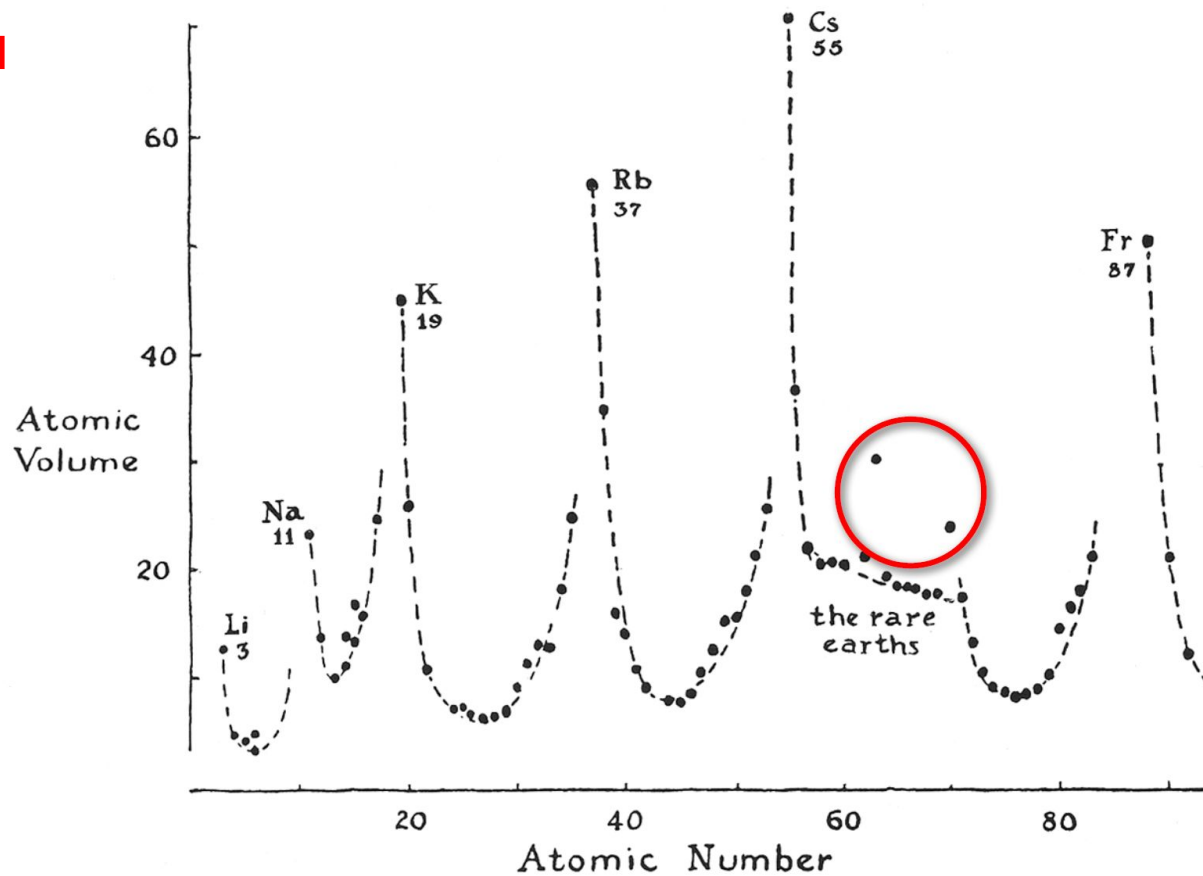
VDQI Example (p102)

Убрать
и то,
и другое?

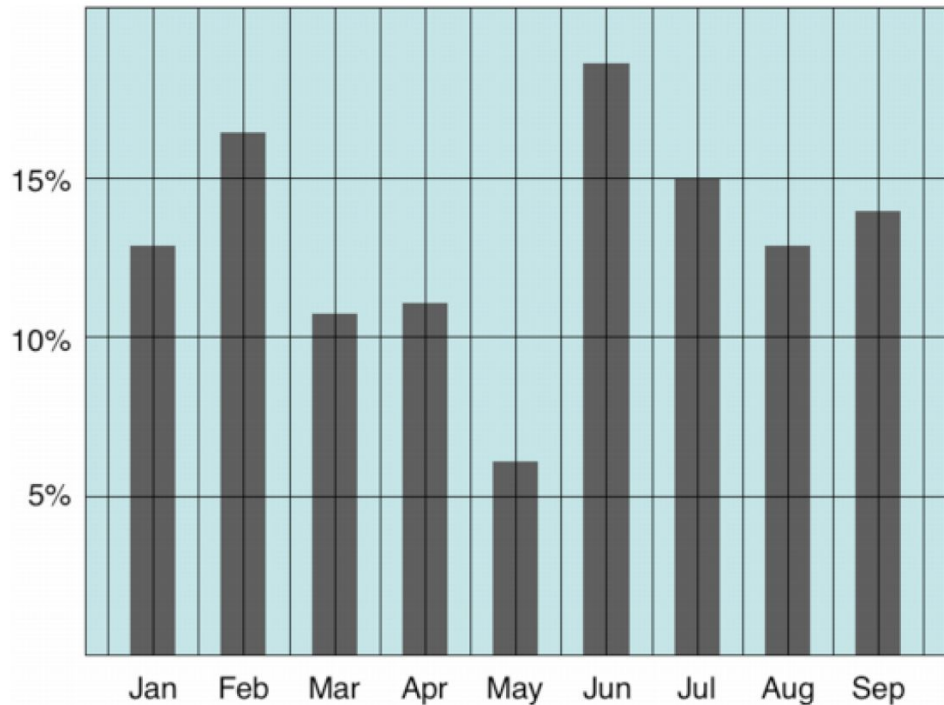


VDQI Example (p102)

Итоговая версия



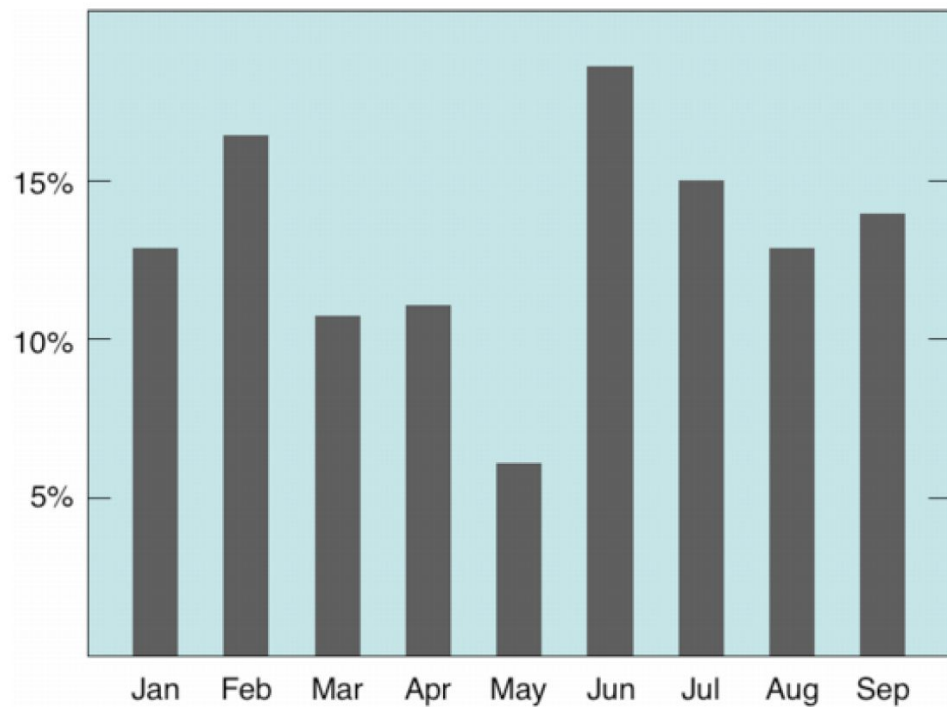
Избегайте мусора в графике



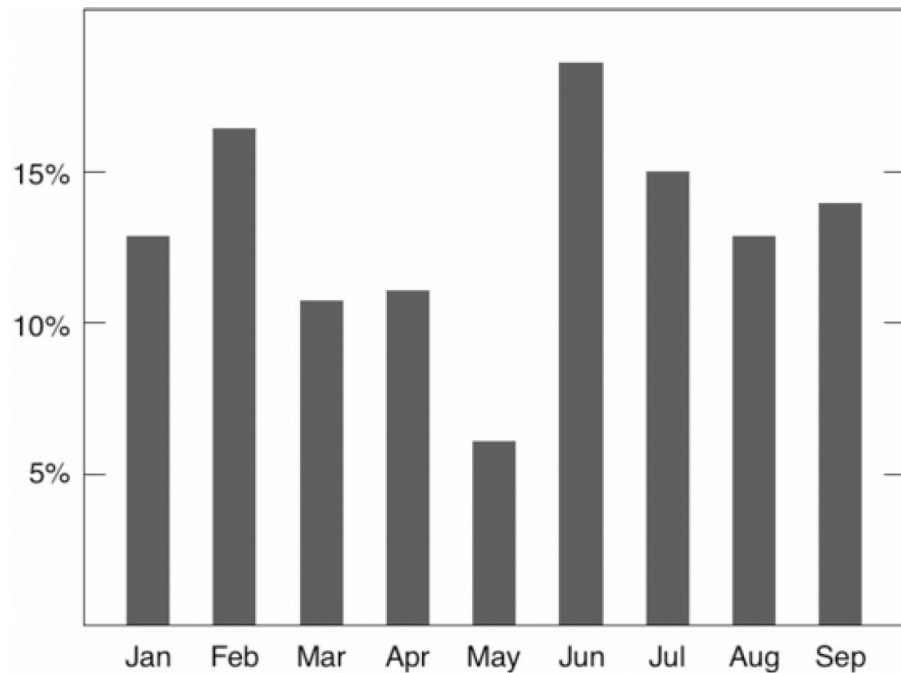
Посторонние визуальные
элементы, которые
отвлекают от послания

Упражнение: давайте
упростим график, но
сохраним информацию.

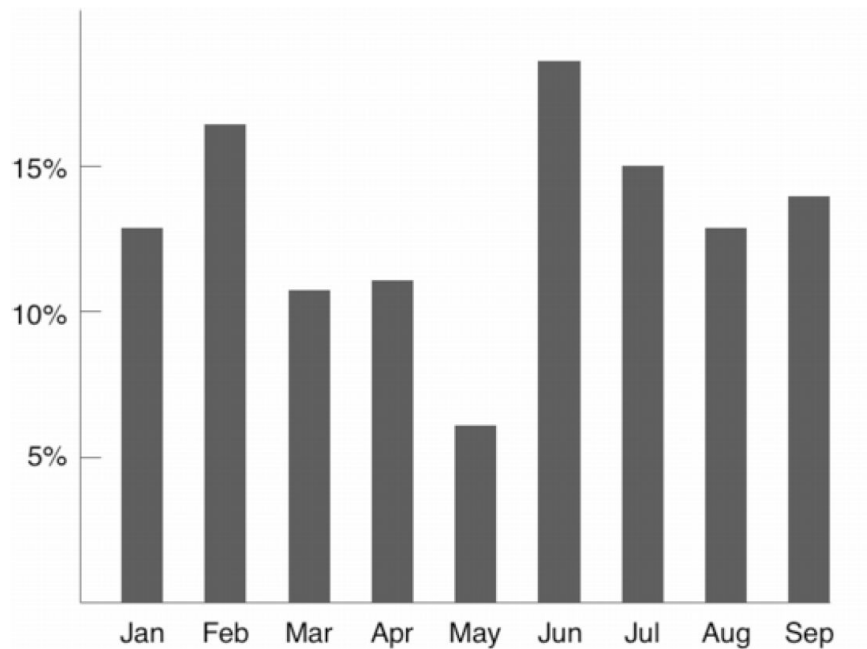
Избегайте мусора в графике



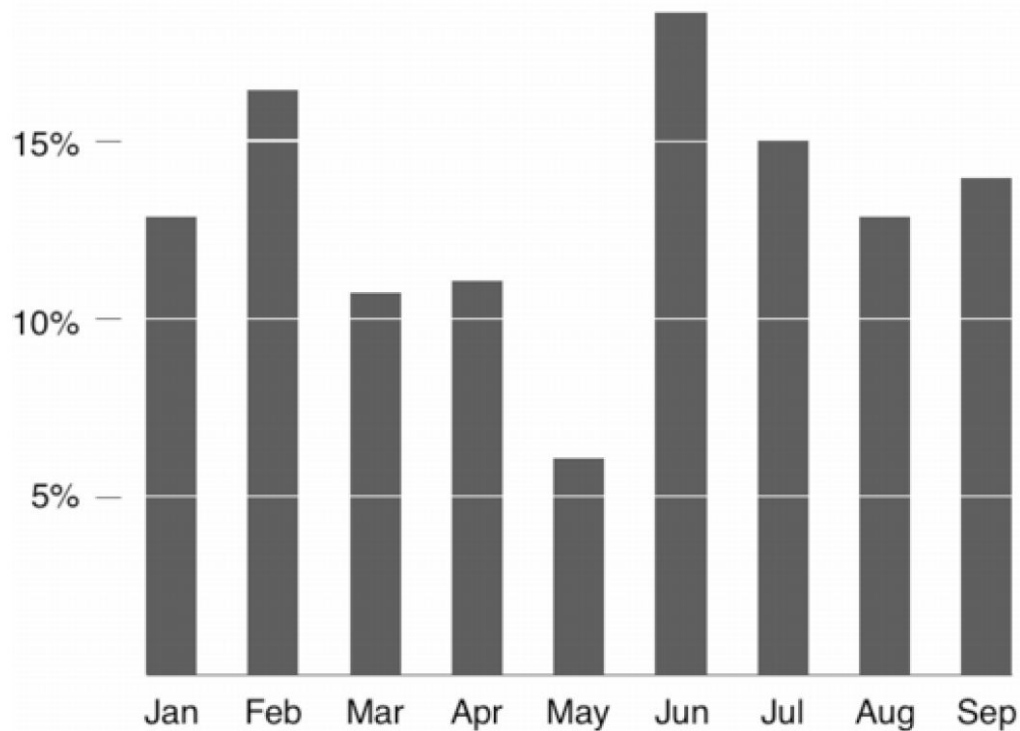
Избегайте мусора в графике



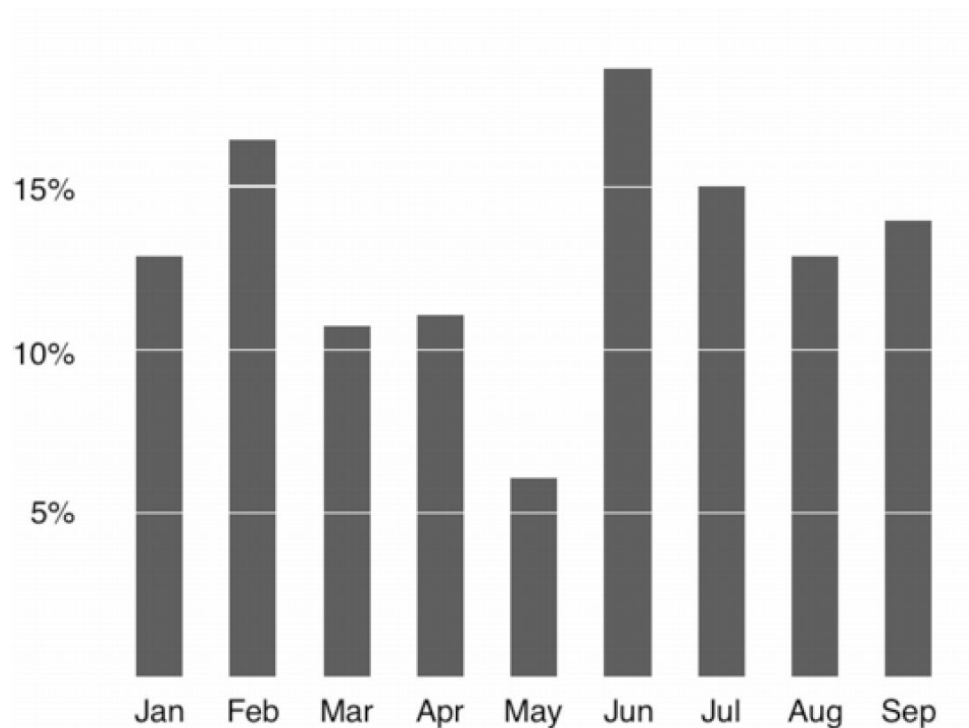
Избегайте мусора в графике



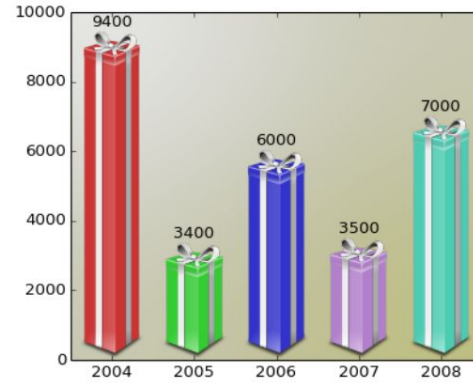
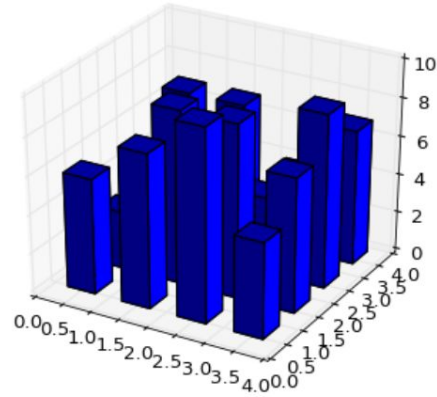
Избегайте мусора в графике



Избегайте мусора в графике

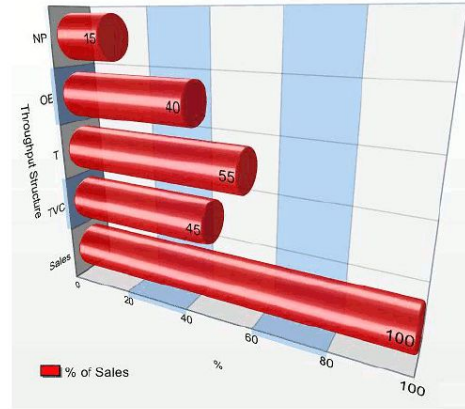
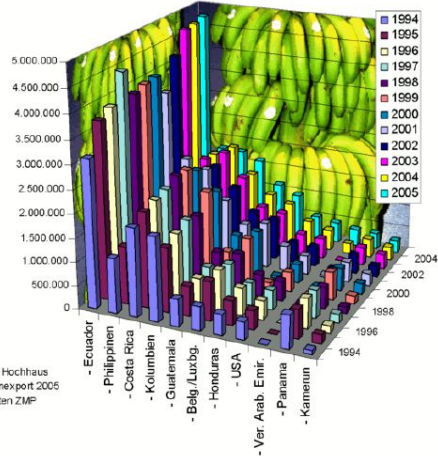


Удалили все лишние
элементы/заливки без
ущерба информативности.

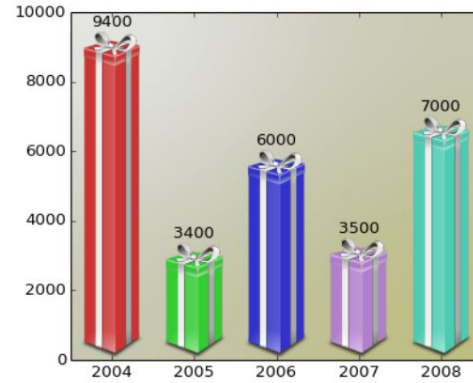
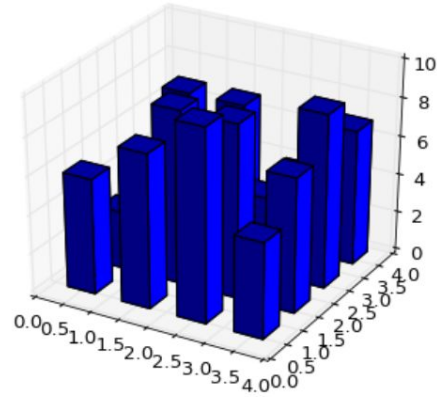


matplotlib gallery

Export von Bananen in Tonnen von 1994-2005

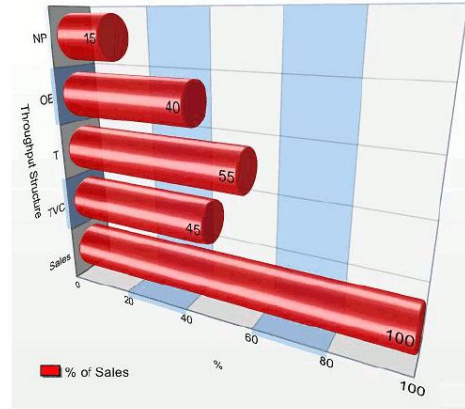
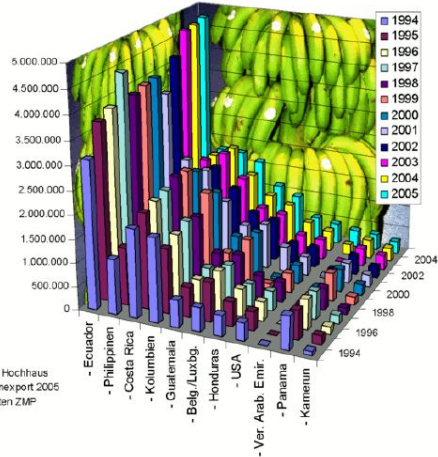


Excel Charts Blog



matplotlib gallery

Export von Bananen in Tonnen von 1994-2005



Excel Charts Blog

Her.

Особенности визуального восприятия

Сколько букв В
на следующем
слайде?

E L O H A N
V I B D H T C H A E
P B T A Q F C N J Y G R A S Q D
M A B Q A C U A K P G A B X O Q Z A G
A I D V D U G C E Y A V A S S Z A
R F S A C T A C A F O J G Q A
A Z A P W X A A F A B F Z

Сколько букв В
на следующем
слайде?

E L V I O H A N
P B D H T C H P A E
T A Q F C N J Y G R A S Q D
M A B Q A U A K P A V B X O Q Z A G
A I D E V A D U G A C E Y A O J G A
R F S C T A C A F O J G Q
A Z A P W X A A F A B F Z

Преаттентивная обработка

- Независимо от сознательного контроля
Вы заметите, хотите вы этого или нет
- Информация обрабатывается без необходимости фокусироваться
Видно даже боковым зрением
- Схоже с фильтром иконической памяти
Привлекает внимание только к тому, что важно

Память

Долговременная память

- Длится годы или всю жизнь
- Быстро доступна, сложно хранить

Кратковременная память

- Длится несколько секунд или минуту
- Ограниченный объем хранения (5-7 элементов)
- Сознательная, сфокусированная, внимательная обработка

Преаттентивные характеристики

Цвет

- Оттенок, интенсивность, и т.д.

Форма

- Длина, форма и т.д.

Позиция

- Местоположение, глубина

Движение

- Мерцание, дрожание, и т.д.

Преаттентивные характеристики

- Аккуратно настройте данные под преаттентивные атрибуты
Благоразумно используйте самые сильные
- НЕ ОТВЛЕКАЙТЕ от данных
Не злоупотребляйте этими атрибутами!
- Помните о кратковременной памяти
Слишком много элементов будет сбивать с толку



Цвет

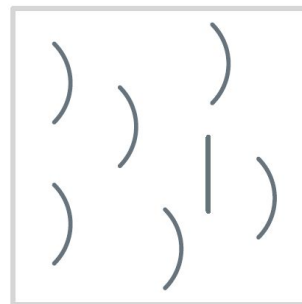
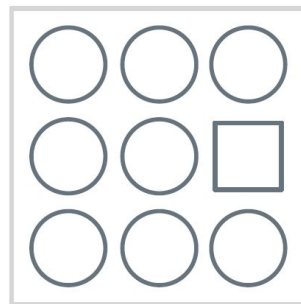
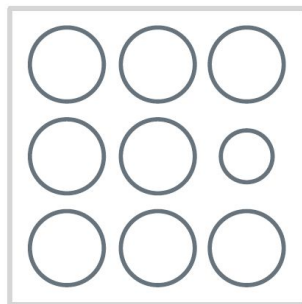
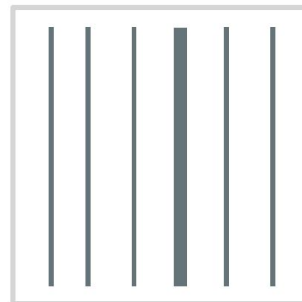
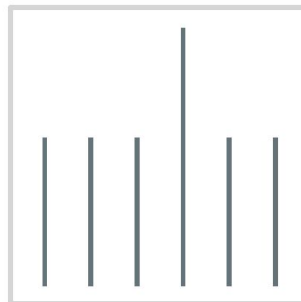
- Отдельные цвета должны обозначать отдельные данные
- Одинаковые цвета должны обозначать одинаковые данные
- Цвет и тематика не должны вызывать когнитивного диссонанса
- Никогда не используйте больше цветов, чем может храниться в кратковременной памяти (5-7)
- Помните о дальтонизме!

Ресурсы о цвете

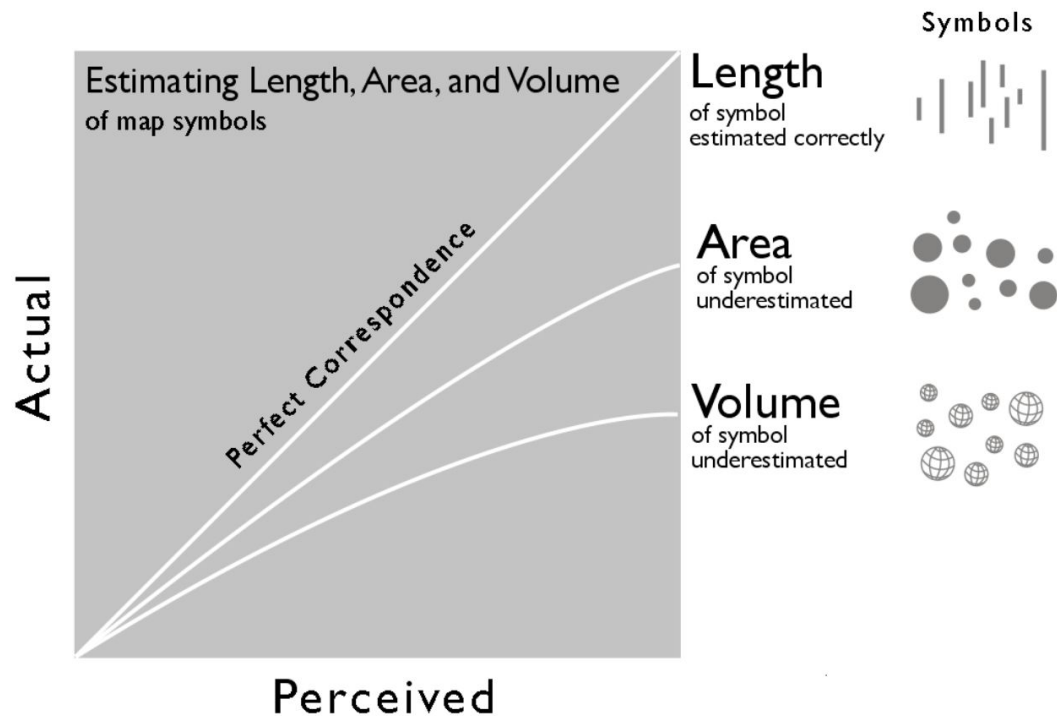
- "A Field Guide to Digital Color" by Maureen C. Stone, 2003.
- Выбор цвета для визуализации данных
<http://www.b-eye-network.com/newsletters/ben/2235>
- Советы по цветам для картографии
<http://colorbrewer2.org/>
- Симулятор дальтоника
<http://www.vischeck.com/>

Внешний вид

- Ориентация
- Длина
- Ширина
- Размер
- Форма
- Кривизна

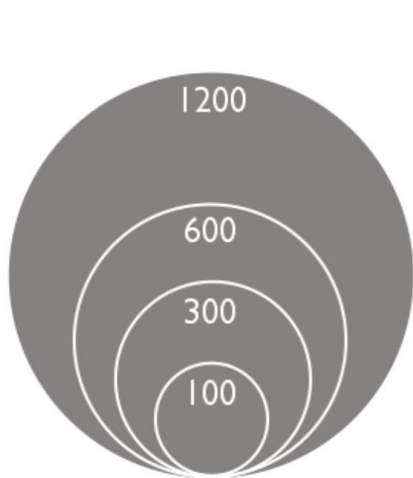


Внешний вид: форма и размер

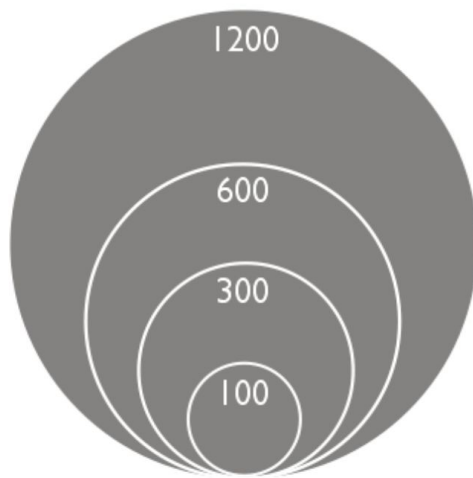


<http://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/>

Внешний вид: форма и размер



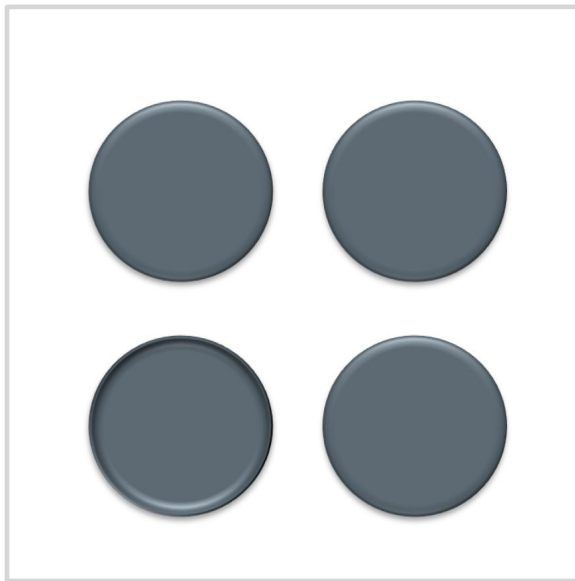
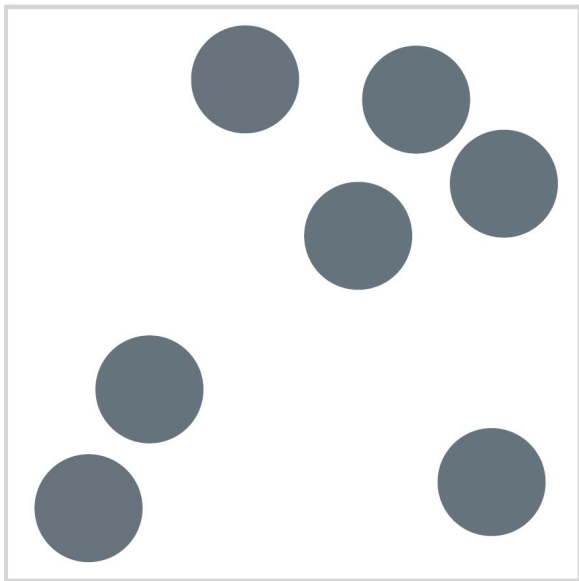
Absolute Scaling



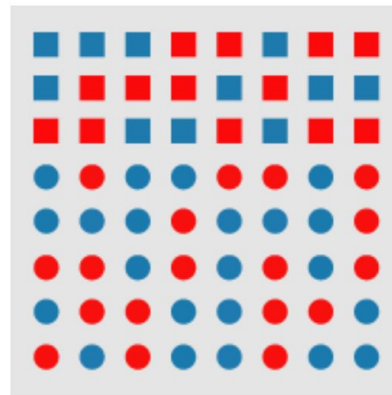
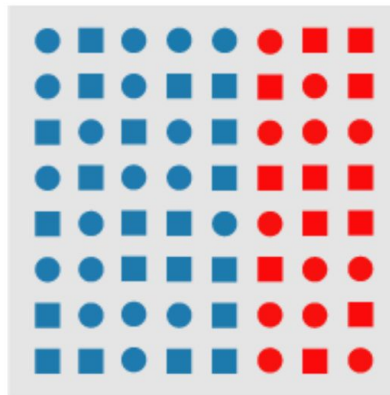
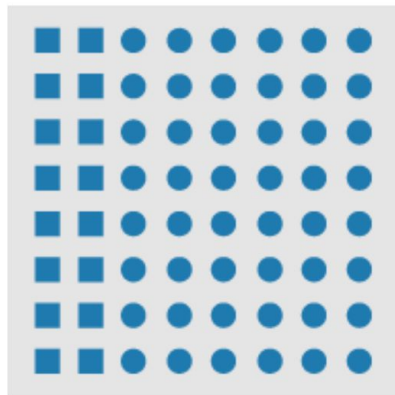
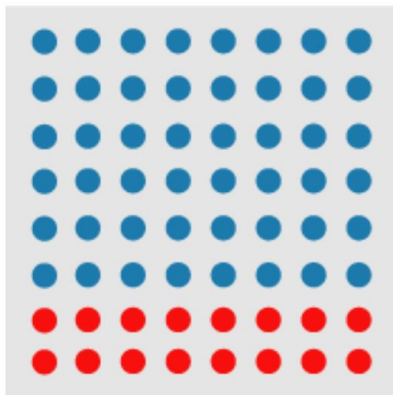
Apparent Scaling
(Flannery's Compensation)

<http://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/>

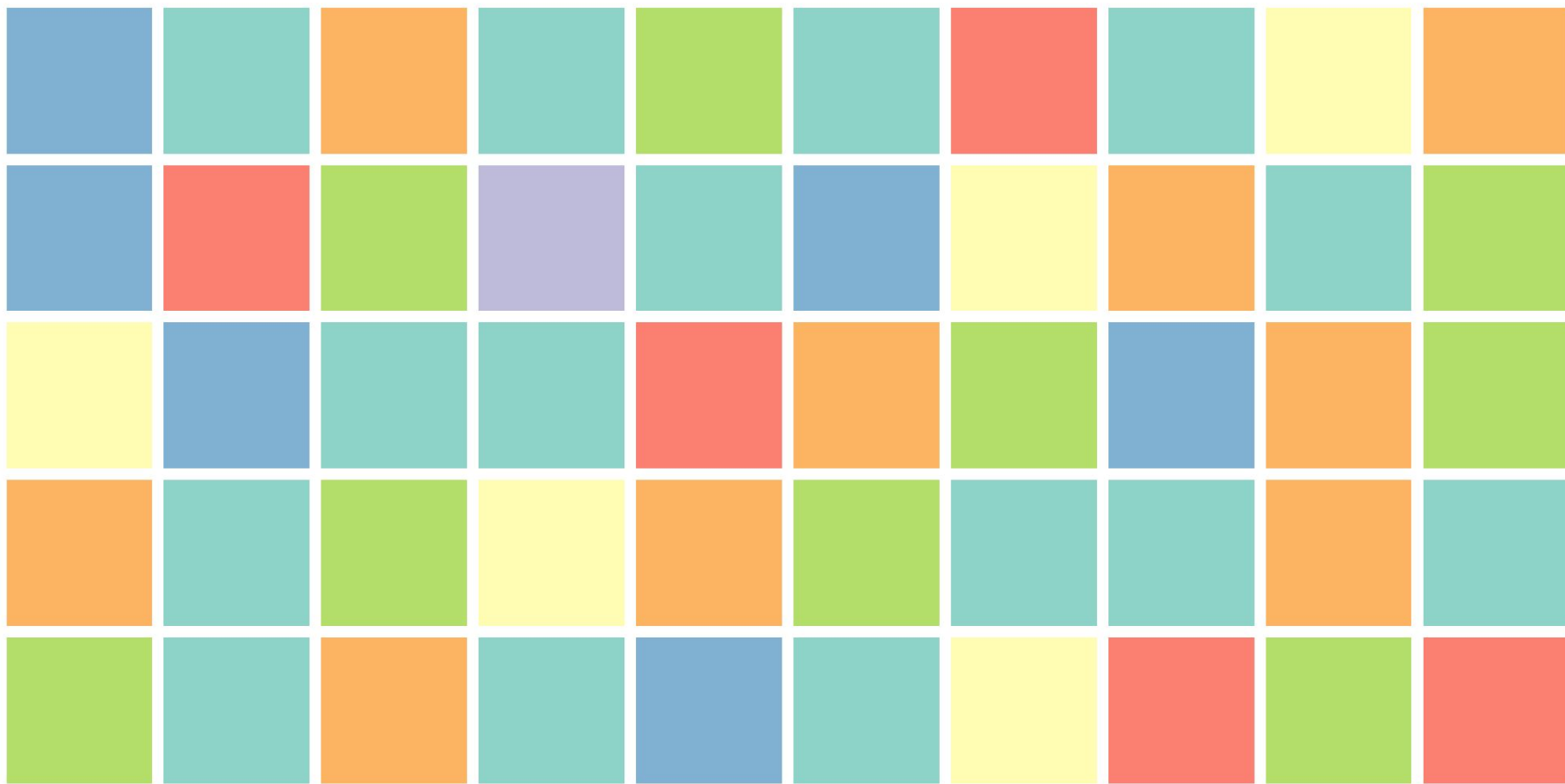
Положение: 2D и глубина



Положение: группировка



Найдите
уникальный
цвет



<http://steveharoz.com/research/attention/>



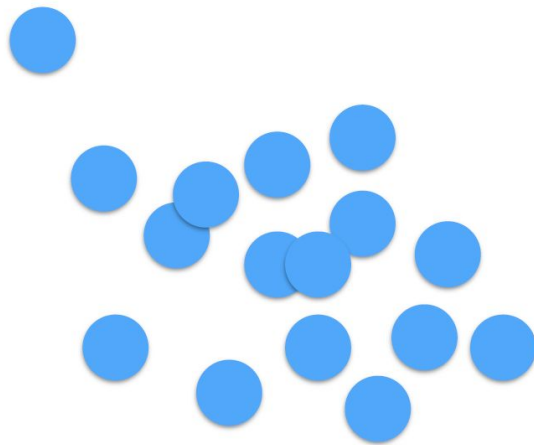
Группировка

- Группировка может увеличить емкость 4154224174 или (415) 422-4174
- Размер групп должен быть небольшим
- Группировка также может ускорить обработку информации

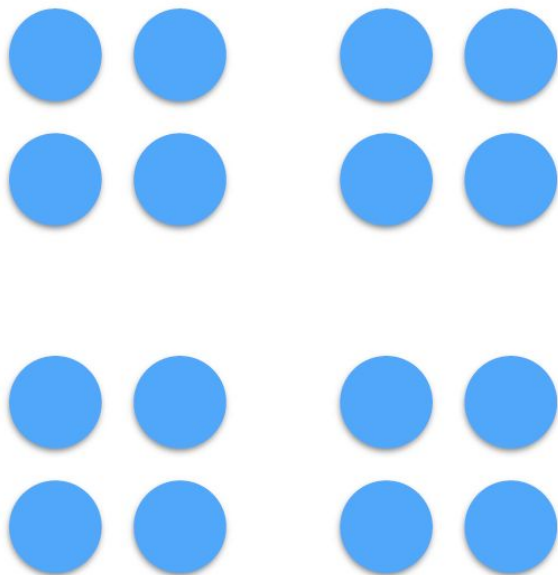
Группировка

- Улучшает возможность найти выброс
- Особенно важна если краткосрочная память загружена (уже 7 цветов)
- Работает и для других преаттентивных характеристик (например, движение)

СКОЛЬКО ТОЧЕК?



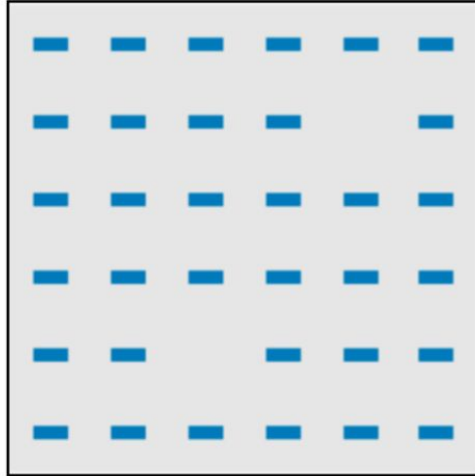
СКОЛЬКО ТОЧЕК?



Движение

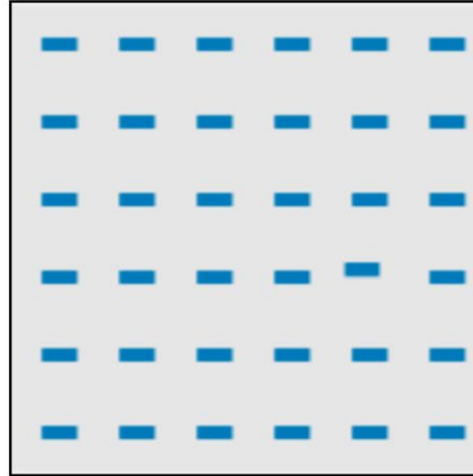
- *Две характеристики*
Мерцание (исчезновение и появление вновь)
Передвижение (изменение позиции)
- Один из самых эффективных способов привлечь внимание
- Больше всего злоупотребляется в маркетинге

<https://www.csc.ncsu.edu/faculty/healey/PP/>



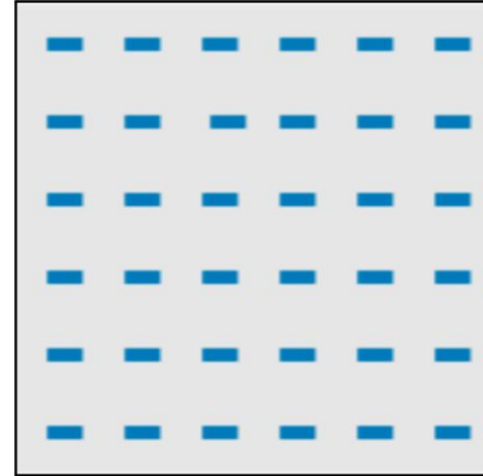
flicker

Gebb et al. 55; Mowbray & Gebhard 55; Brown 65; Júlész 71; Huber & Healey 2005



direction of motion

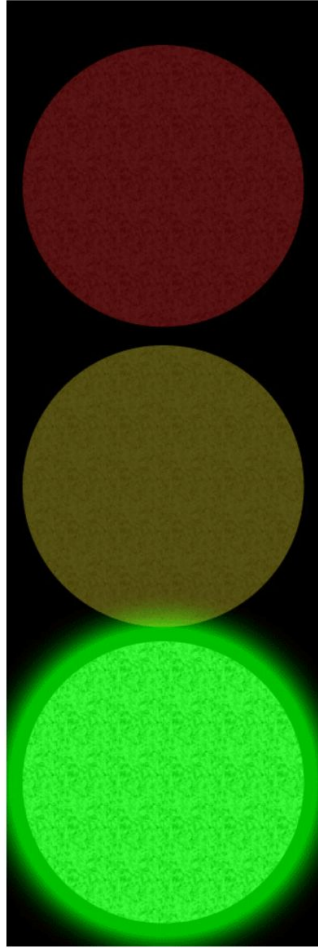
Nakayama & Silverman 86; Driver & McLeod 92; Huber & Healey 2005



velocity of motion

Tynan & Sekuler 82; Nakayama & Silverman 86; Driver & McLeod 92; Hohnsbein & Mateff 98; Huber & Healey 2005

Пример
из повседневной
жизни?



Память

Кратковременная память

- Внимание/фокус доставляет информацию из сенсорной памяти в кратковременную память
- Длится от нескольких секунд до минуты
- Ограниченная емкость
Минимум: 5 элементов
В среднем: 7 элементов
Максимум: 9 элементов

Практические индикаторы

- Можно использовать пять разных характеристик
- Не стоит использовать больше 7, чтобы быть понятным
Не более 7 разных цветов или форм!
- Характеристики складываются
3 формы, 4 цвета = 7 характеристик
- Как только фокус потерян, информация теряется
Отвлечение внимания обходится дорого

Упражнение

Сейчас вы увидите
9 фотографий,
по паре секунд на каждую.

Ваше задание - запомнить
все эти фотографии.



















Упражнение

Сейчас мы снова увидим фотографии. Если вы видели их, то хлопните в ладоши.

Если нет - не делайте ничего.





Нет!





Нет!





Нет!



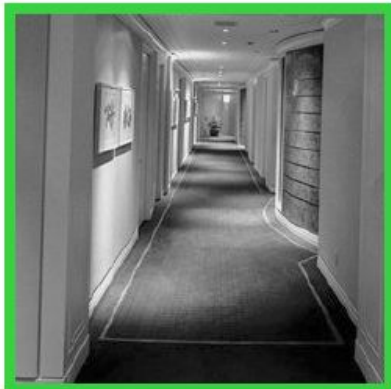


Да!





Нет!



Упражнение: обман памяти

- Наша краткосрочная память не очень хорошо справляется с хранением более 5-7 деталей.
- Поэтому фотографии с одинаковым ракурсом и тематикой так легко спутать.
- Нужно приложить усилие (и время), чтобы запомнить детали.

39 studies about human perception in 30 minutes

by Kennedy Elliott, visual journalist at Washington Post

<https://medium.com/@kennelliott/39-studies-about-human-perception-in-30-minutes-4728f9e31a73>

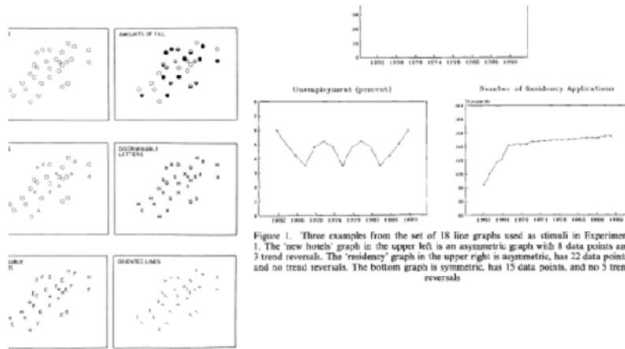


Figure 1. Three examples from the set of 18 line graphs used as stimuli in Experiment 1. The 'new home' graph in the upper left is an asymmetric graph with 8 data points and 3 trend reversals. The 'residency' graph in the upper right is asymmetric, has 22 data points, and no trend reversals. The bottom graph is symmetric, has 15 data points, and no 3 trend reversals.

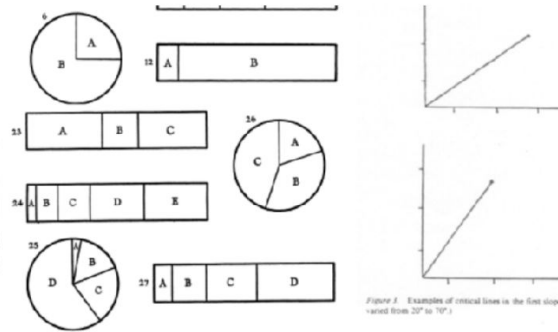
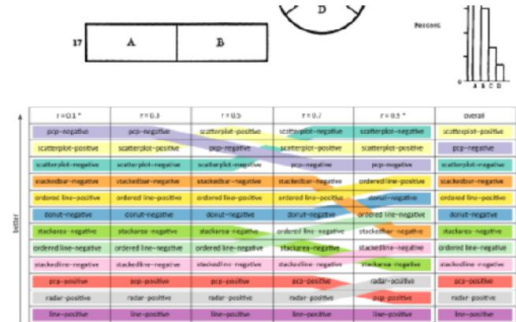


Figure 2. Examples of critical lines in the first slope study. (Angles varied from 20° to 70°.)

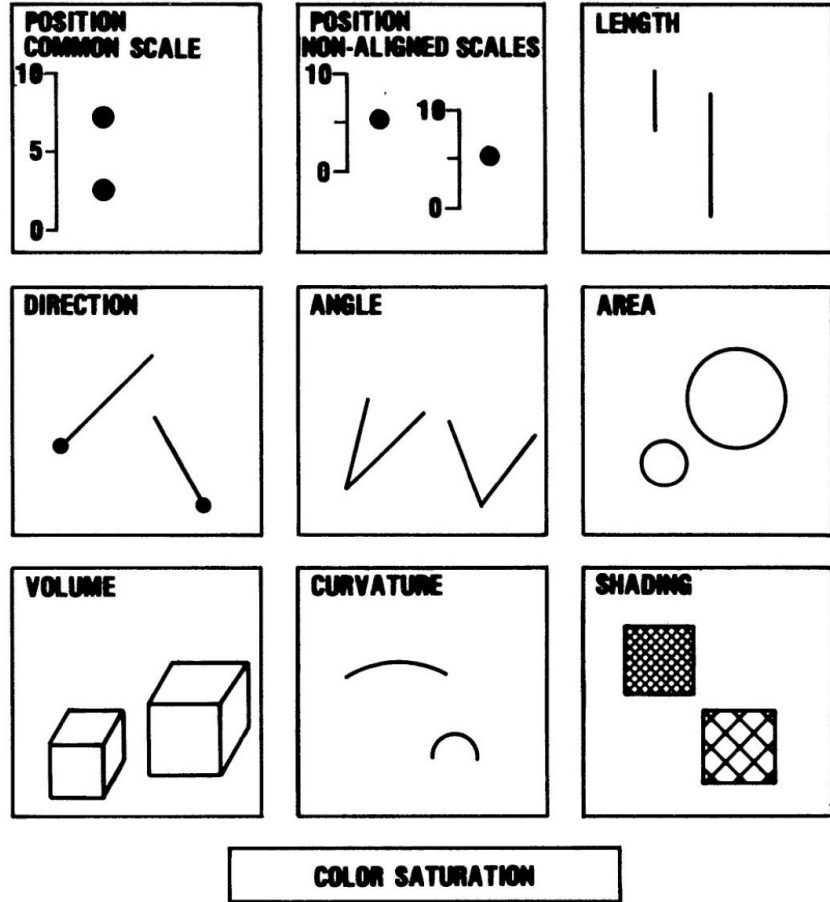


39 studies about human perception in 30 minutes

These are my speaker notes from a talk I gave at OpenVis in April 2016. Originally this talk was supposed to be called "Everything we know about how humans perceive graphics," which is... at a minimum, pretty pretentious. I scaled back a bit.

foundations

In 1984, William Cleveland and Robert McGill (1) published a study that can only be described as the archetypical seminal study for information visualization.



In several experiments, Cleveland and McGill identified "elementary perceptual tasks": the most basic tasks they believe viewers perform when evaluating a visualization.

foundations

However, this study was conducted three decades ago. How relevant is it today?

Fortunately for us, we have some clue. Heer and Bostock (2) revisited some of Cleveland and McGill's old experiments in a study they did as more of a proof of concept for using Mechanical Turk users in scientific studies.

Their results were very similar to those of Cleveland and McGill, at least with the perceptual tasks they tested for.

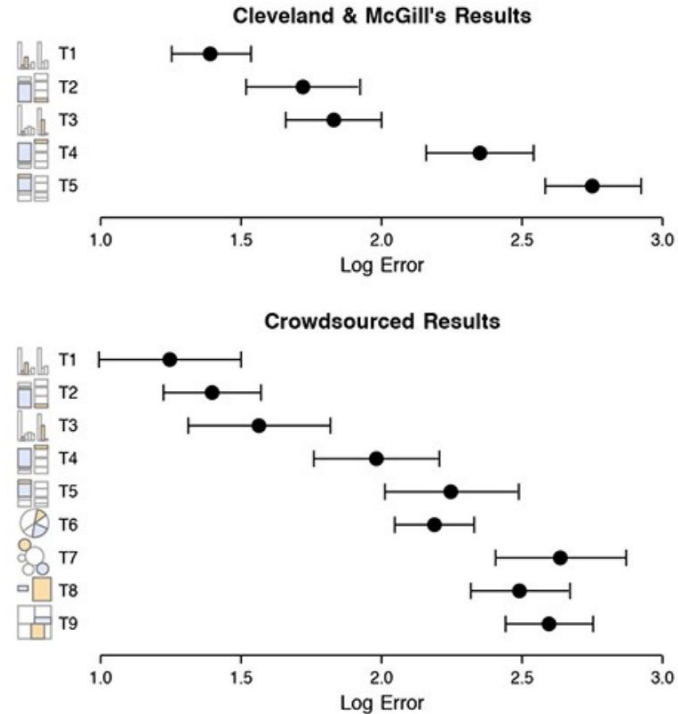


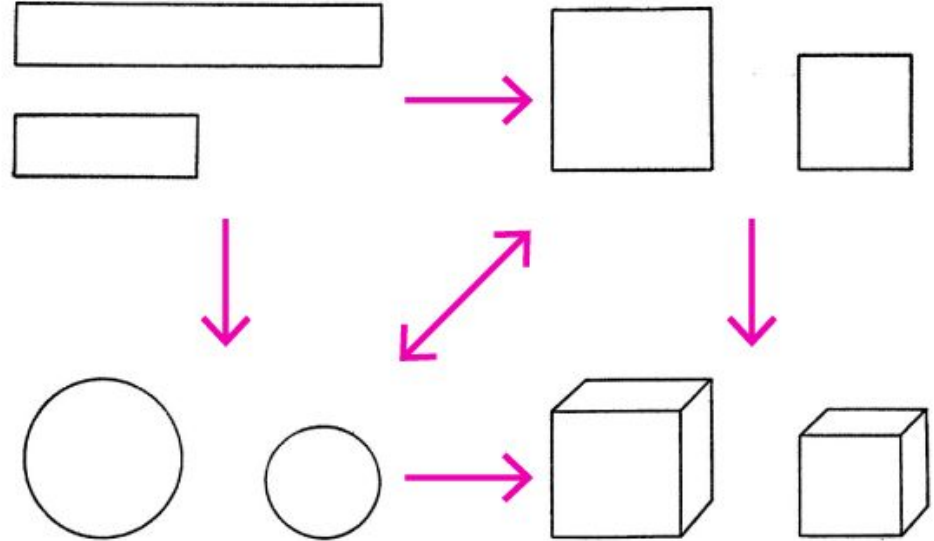
Figure 4: Proportional judgment results (Exp. 1A & B). Top: Cleveland & McGill's [7] lab study. Bottom: MTurk studies. Error bars indicate 95% confidence intervals.

Heer and Bostock's experiments with Mechanical Turk participants yielded similar results to those of Cleveland and McGill.

basic shapes

Croxton (6) found over eight decades ago that bars were more effective in communicating comparative values than either circles, squares or cubes. Circles and squares were about as effective as the other. Cubes were undoubtedly the worst. More on 3d later.

BARS, SQUARES, CIRCLES, AND CUBES SHOWING 50 TO 100 RELATIONSHIP



bars and pies for proportions

Much is said about the relative merits of bars and circles for showing proportions.

All five of these studies legitimize the use of pie charts when conveying proportions and some even show their superiority over bar charts.

5] *Circles and Bars for Representing Component Parts* 123

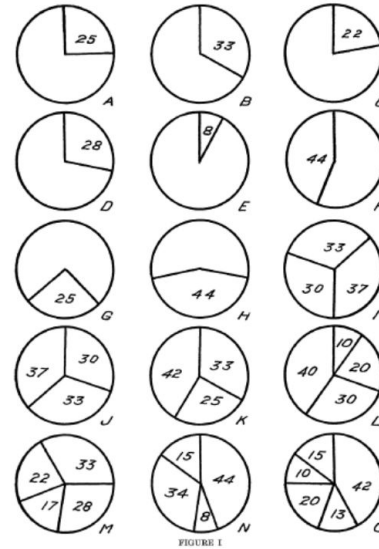


FIGURE I

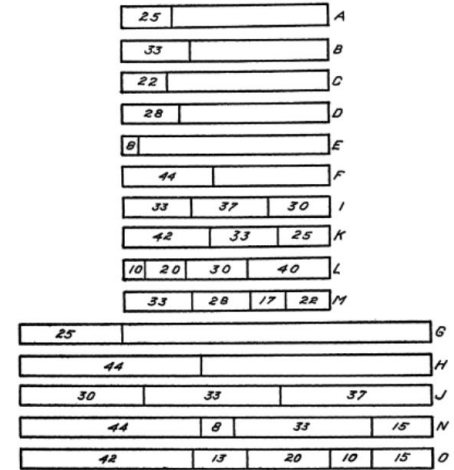


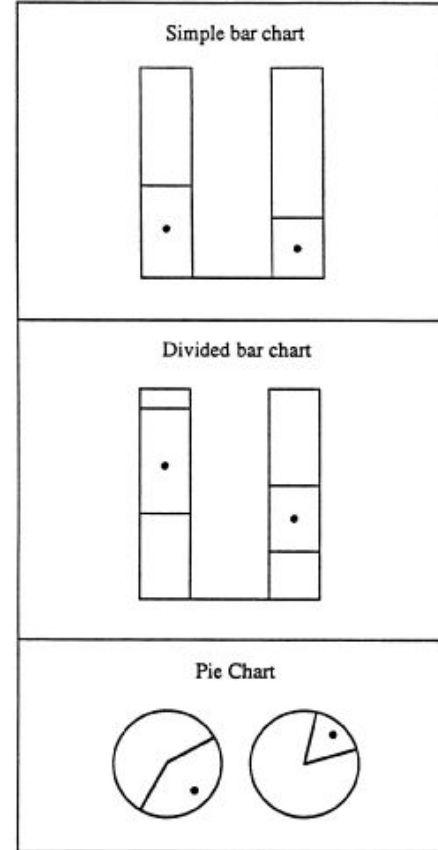
FIGURE II

Eells gave a psychology class two worksheets — a series of pie charts and bar charts — and asked them to estimate the proportion of each segment to the whole.

bars and pies for proportions

Simkin and Hastie (9) had participants make proportional judgments and segment-to-segment (comparison) judgments. And found that for segment-to-segment judgments, simple bar charts worked best, followed by divided bar charts and then pie charts.

segment-to-segment

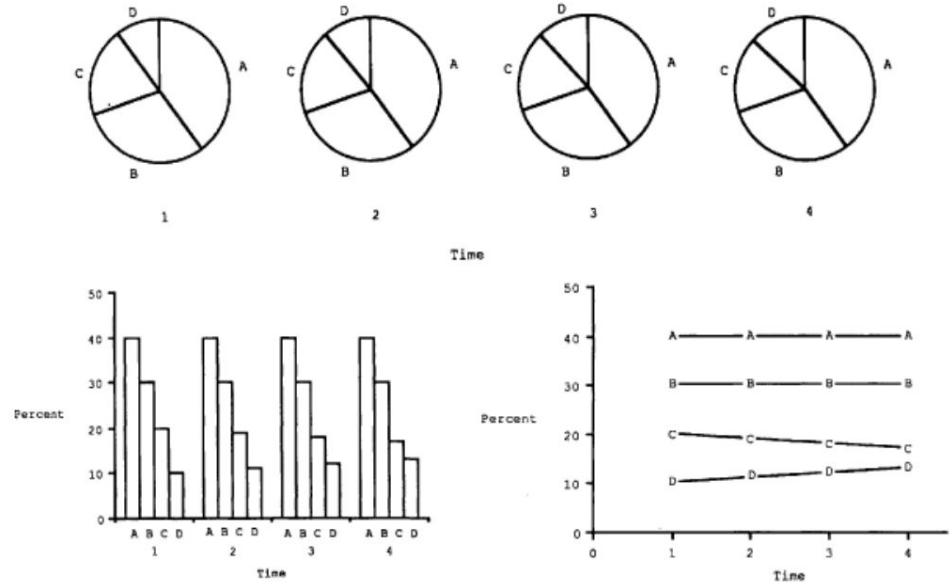


proportional

bars and pies and lines

They tested participants' perception of change and proportion among bar, pie and line charts.

Pie charts obviously failed at communicating change efficiently, but they found that bar charts had similar success to line charts

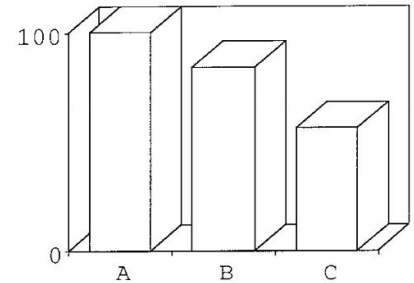
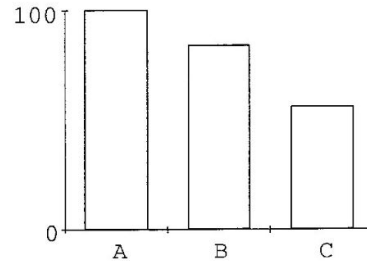
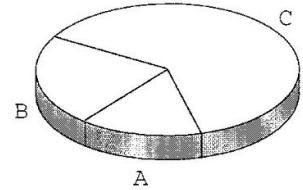
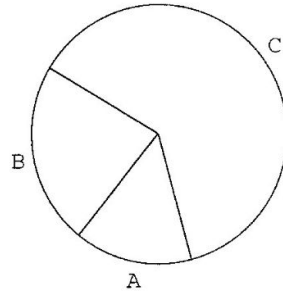


In their first experiment in the study, Hollands and Spence presented participants with these three graphs to evaluate how accurately they could evaluate change.

3d

Siegrist (17) finds that among bar charts, 2d is not superior to 3d, but 3d charts take slightly longer to process. With pies, 2d is better, and the perspective angle makes a big difference in how accurately the slices are evaluated, most likely because some of the slices are more obscured than others.

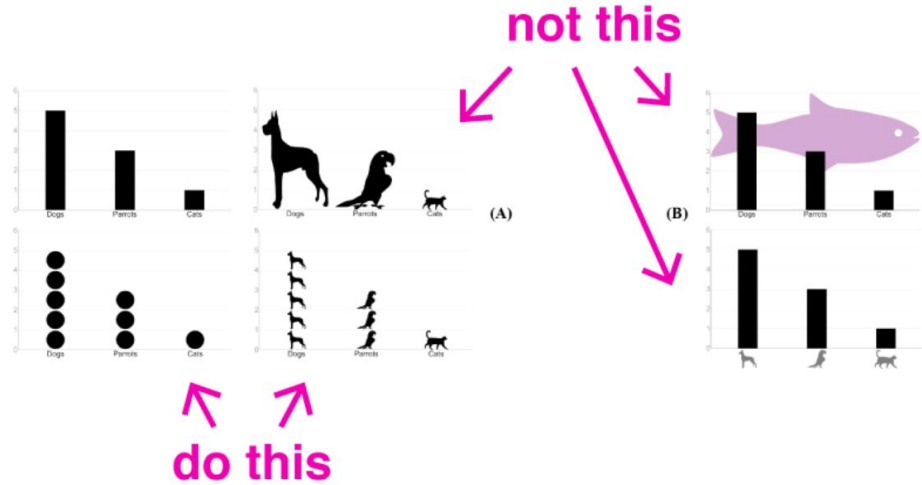
Levy and coauthors (18) acknowledge that 3d graphics, while “glitzy” and “sexy,” do not convey any additional information and force the reader to “deal with redundant and extraneous cues.”



Siegrist observed how well participants could evaluate size differences among 2d and 3d variants of pie and bar charts. Participants were about as accurate with 3d bar charts as the 2d versions, but were less successful with 3d pie charts.

pictographs and drawings

Haroz, Kosara and Franconeri (31) experimented with using pictographs instead of generic shapes to represent data in simple charts. They found that using discrete shapes, whether they were generic circles or pictographs, helped people remember the data better than a single bar. Using pictographs as replacement for text on the axis led to more errors.



Haroz, Kosara and Franconeri found that replacing generic stacked shapes with stacked pictographs did not result in a decrease in accuracy for participants.

colors

Finally, in another one of my favorite studies, Lin and coauthors (40) created an algorithm to identify “semantically resonant colors.” For example, if I want to talk about oceans, I use the color blue. If I want to talk about love, I use pink or red.

The algorithm searched Google Images and assigned a particular color to a keyword. They crowdsourced how well this algorithm did with Mechanical Turk participants.

Car Color (4.7)	A	T	Food (3.6)	A	T
Red	Red	Red	Sour Cream	Blue	Gray
Silver	Gray	Gray	Blue Cheese Dressing	Orange	Blue
Black	Gray	Gray	Porterhouse Steak	Brown	Brown
Green	Green	Green	Iceberg Lettuce	Green	Light Green
Brown	Brown	Brown	Onions (Raw)	Orange	Light Green
Blue	Blue	Blue	Potato (Baked)	Light Green	Light Green
			Tomato	Red	Red
Features (3.2)	A	T	Activities (2.7)	A	T
Speed	Red	Red	Sleeping	Brown	Blue
Reliability	Blue	Blue	Working	Orange	Red
Comfort	Brown	Light Blue	Leisure	Blue	Light Green
Safety	Orange	Green	Eating	Pink	Green
Efficiency	Green	Light Green	Driving	Red	Gray

Figure 4: Category sets ordered by mean colorability rating. Ratings are shown in parentheses. Column A contains colors selected by our algorithm; Column T contains the Turk-er-chosen colors with highest overlap.

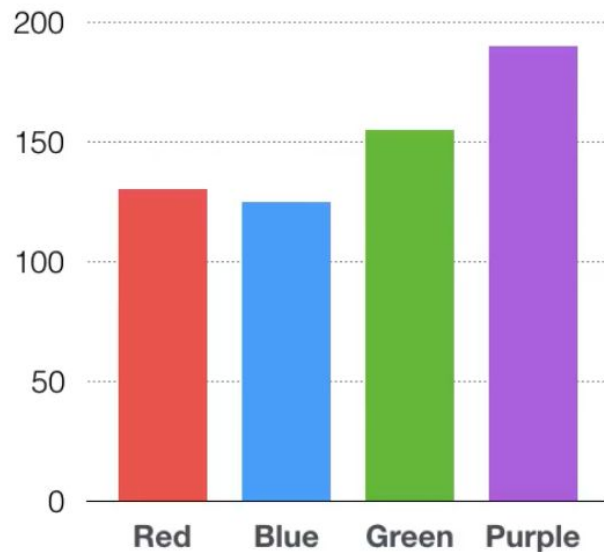
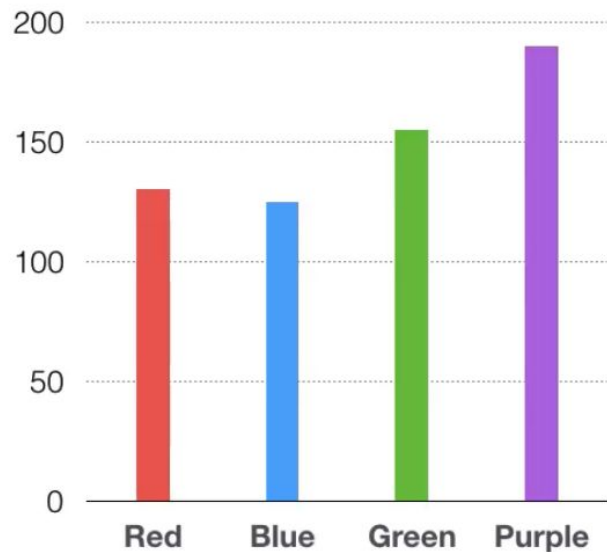
In the first experiment of the study, Lin and coauthors tested how well their algorithm assigned colors to particular keywords by comparing their results to assignments made by Mechanical Turk participants.

Лучшие практики

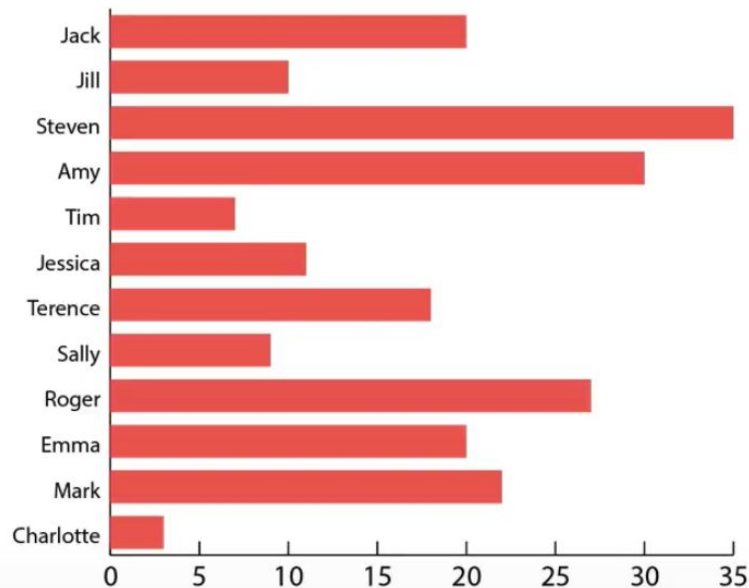
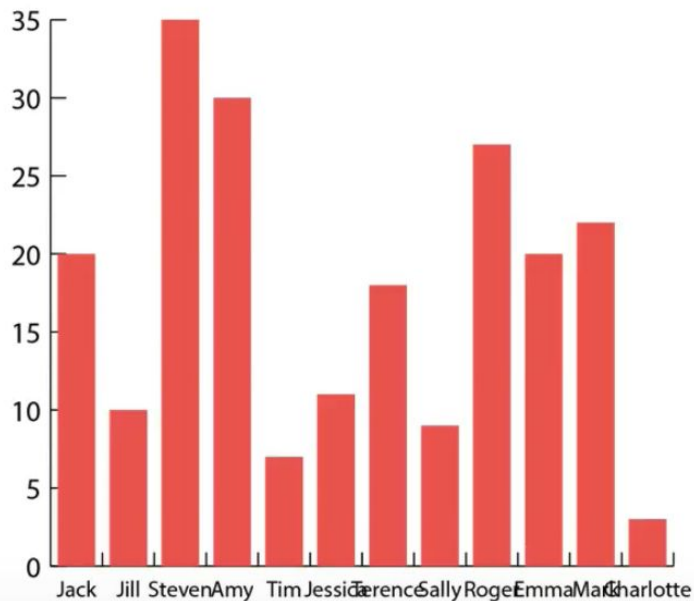
<https://www.youtube.com/watch?v=yntLglYUHXI&feature=youtu.be>

Лучшие
практики:
столбчатая
диаграмм

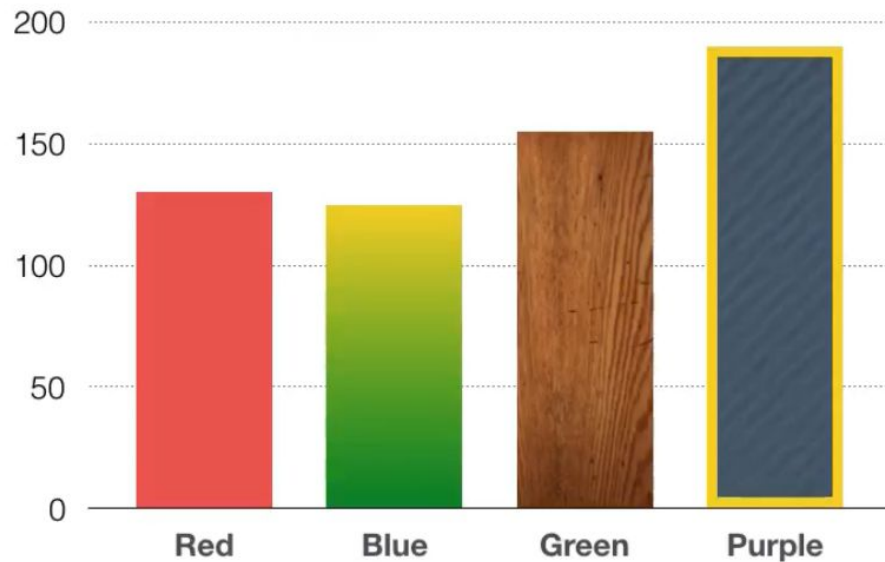
Не рисуйте столбцы слишком тонкими, потому что это смещает акцент на пустое пространство



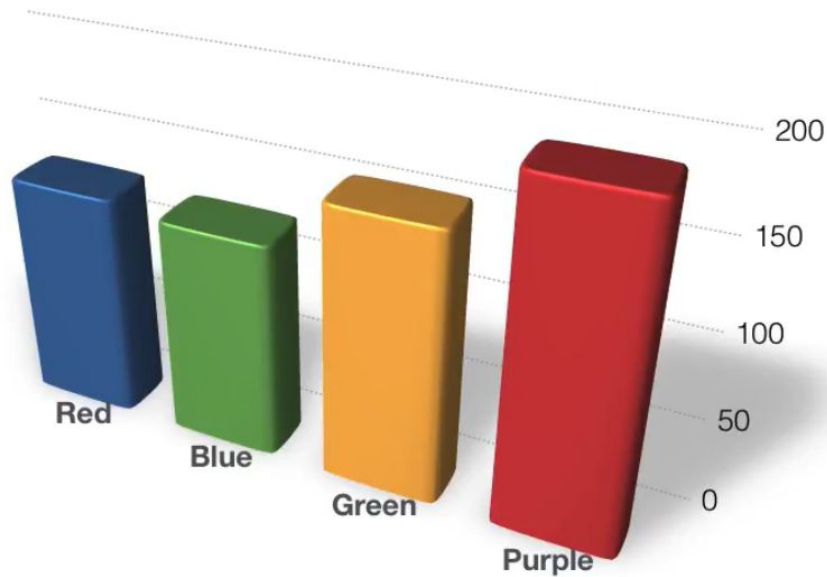
При большом количестве столбцов, размещение лейблов может стать проблематичным. Поэтому лучше использовать горизонтальную диаграмму



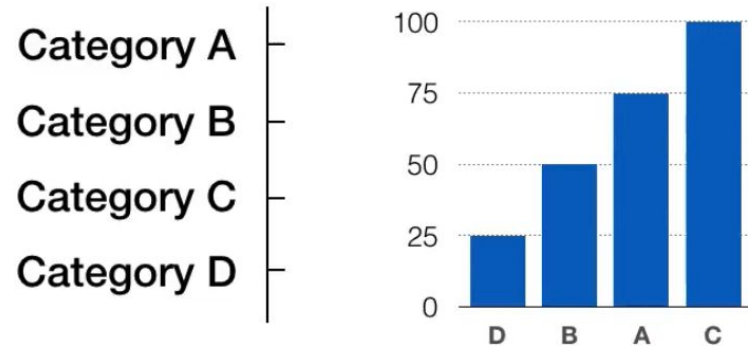
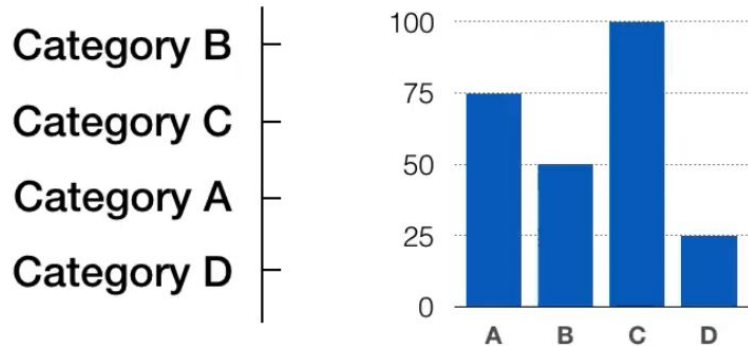
Не используйте радикально разные стили столбцов



**Не используйте 3D столбчатые диаграммы,
т.к. сложно считывать и сравнивать значения**

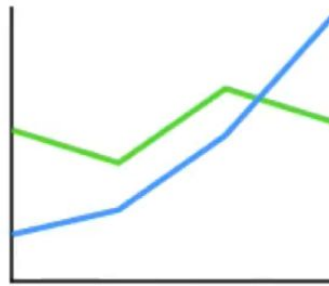
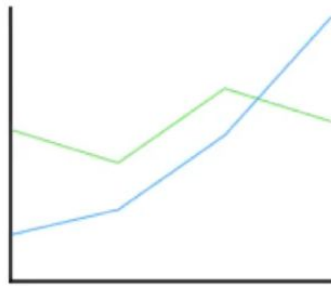


Логически сортируйте категории
(например по алфавиту или по величине)

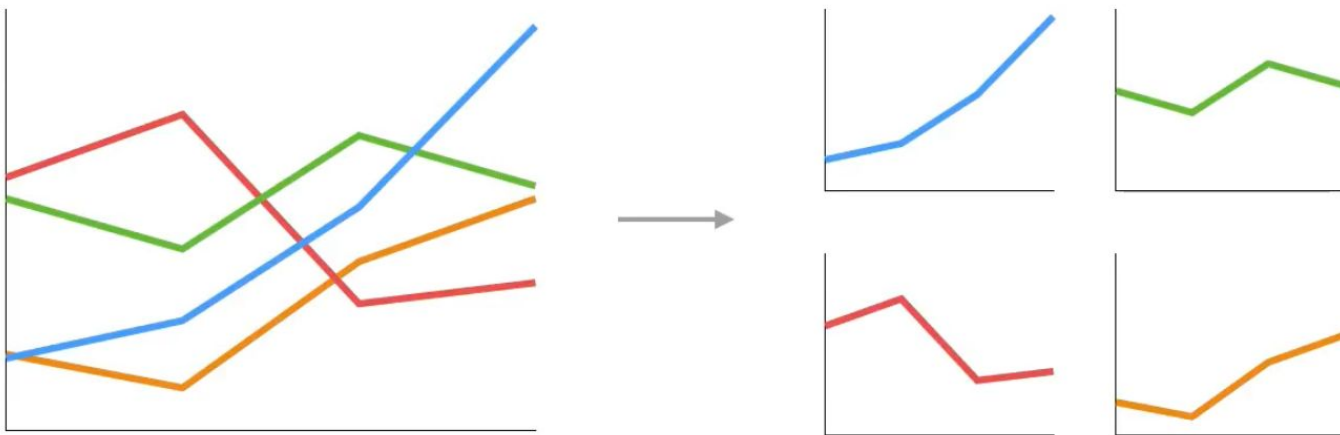


Лучшие
практики:
линейный
график

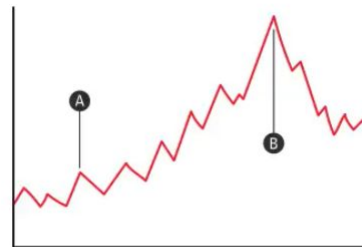
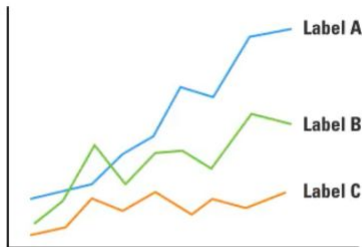
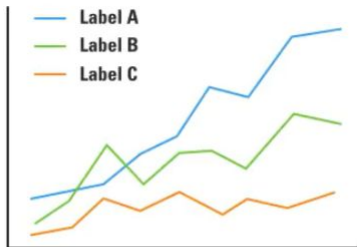
Старайтесь избегать слишком тонких или слишком толстых линий



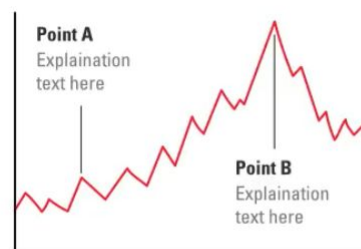
Никаких спагетти:
Если есть больше 3-4 линий, то разбейте график
на несколько маленьких



Постарайтесь наносить лейблы прямо у соответствующих графиков

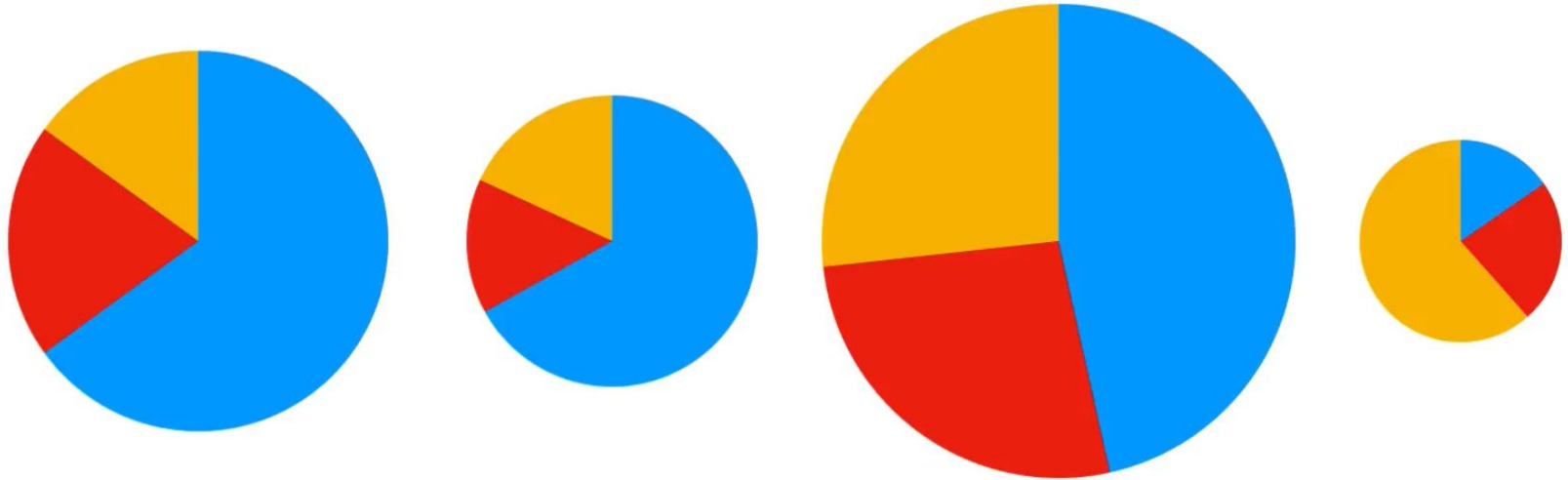


- A Point A** The explanation text goes here
- B Point B** The explanation text goes here



Лучшие
практики:
круговая
диаграмма

- Не нарушайте пропорции кругов. Задача их сравнения невыполнима.



The size of the Mobile market

www.iStrategy2009.com

i Strategy

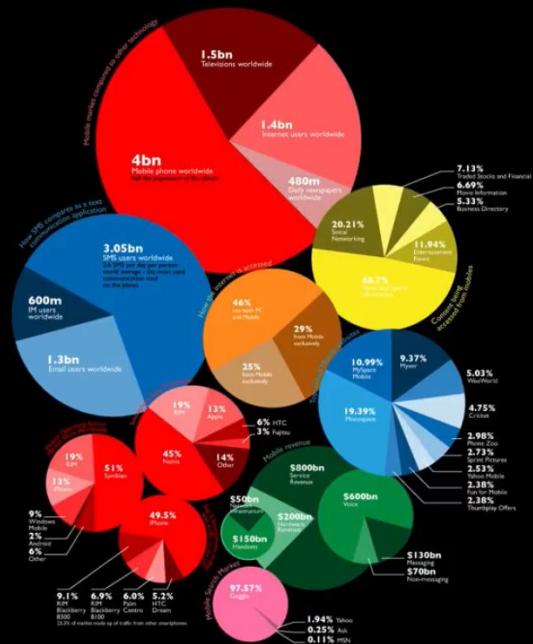
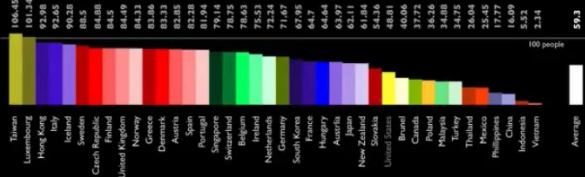


Chart shows the number of mobiles for every 100 people within that country, for example in United States there are 48.81 mobiles for every 100 people



2012 PRESIDENTIAL RUN

GOP CANDIDATES



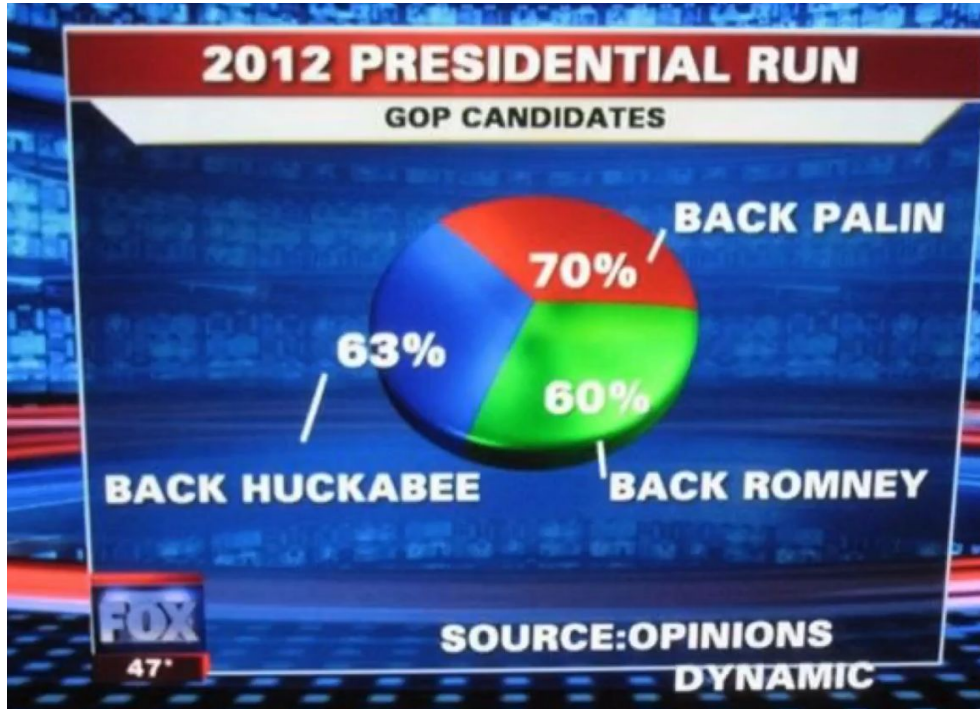
FOX

47°

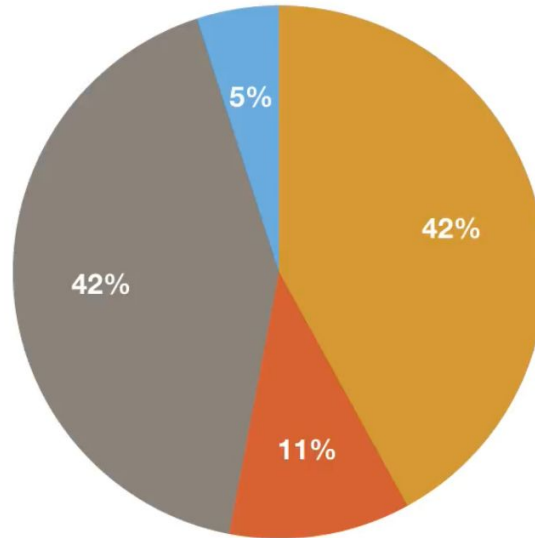
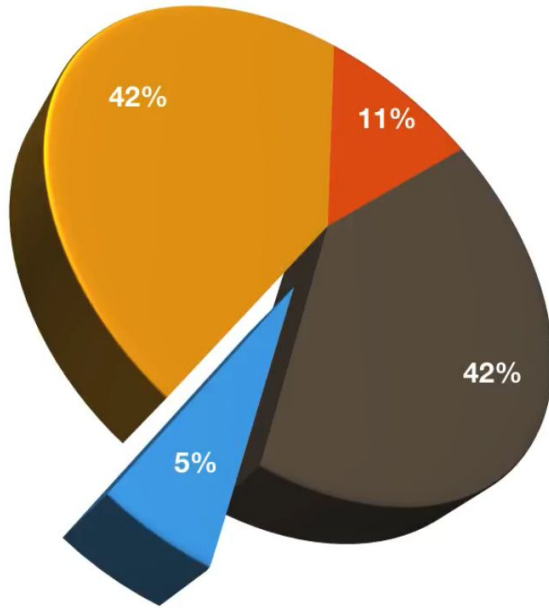
SOURCE: OPINIONS

DYNAMIC

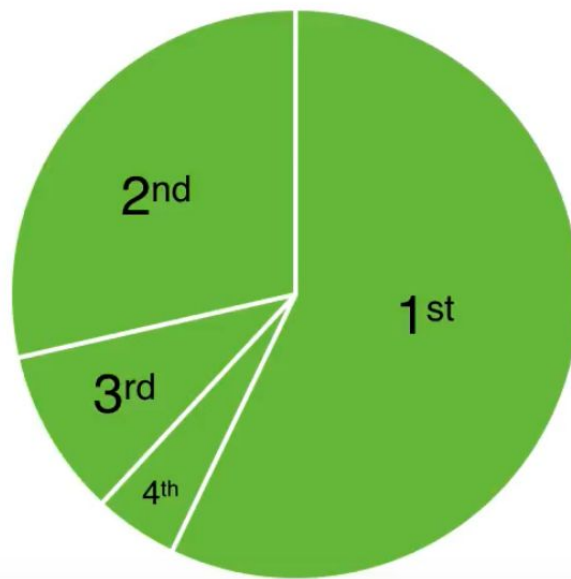
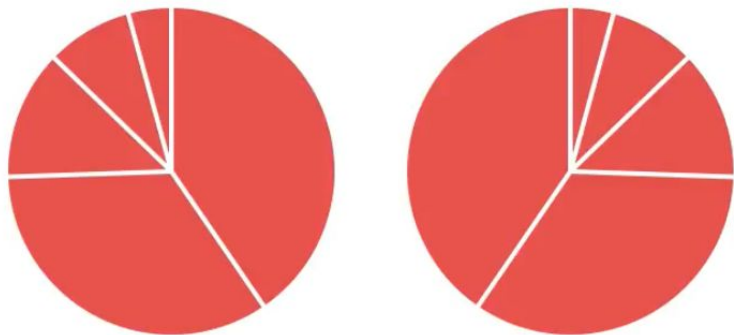
- Круговая диаграмма = 100%!
(не в этом случае)



- 3D особенно сильно искажает пропорции секторов.



Постарайтесь избегать сортировки сегментов от самого маленького к самому большому (и наоборот)



Домашнее задание

Оценка Презентации

Всего 10 баллов

6 баллов: 1 бал за каждый обоснованный пункт почему это плохая визуализация (всего шесть пунктов)

+ 0.5 доп. балла за два дополнительных пункта (всего 8).

2.5 балла: Вы предлагаете свою визуализацию этих данных, которая может быть нарисована в любом графическом редакторе/на листке от руки/на доске в классе. Главное воспроизвести паттерн данных, не обязательно сделать ее 100% точной.

+0.5 доп. балла, если вы вдруг сильно заморочились с альтернативной визуализацией.

Из этих 2,5 баллов могут быть вычтены баллы за допущенные ошибки: выбран неправильный тип визуализации для конкретного типа данных, использованно слишком много цветов, неправильный масштаб шкалы измерений и т.д. - **0.5 балла за каждую ошибку.**

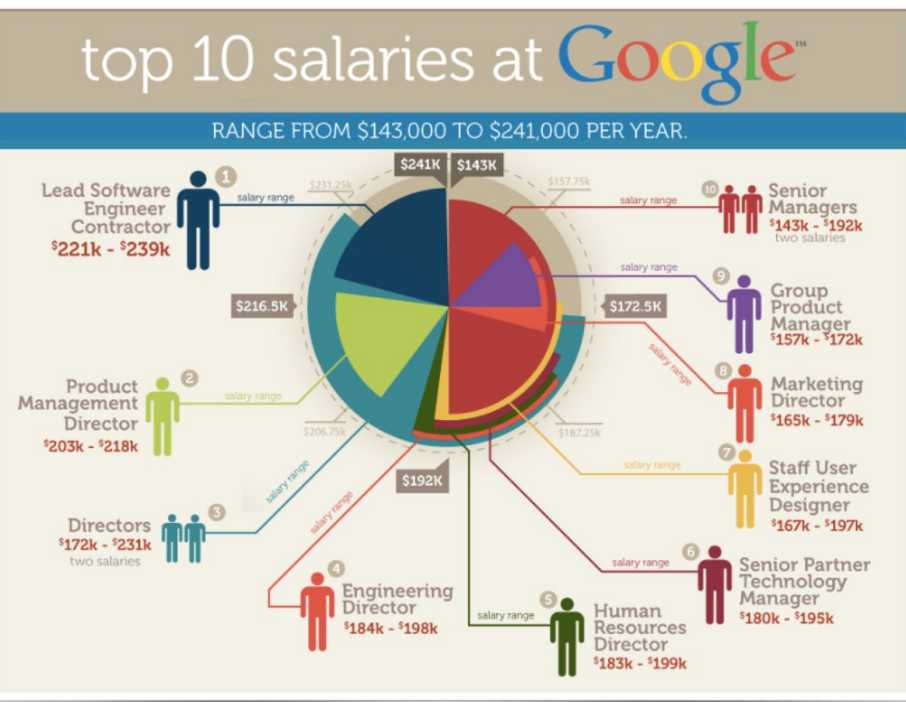
1.5 балла: презентация. Качество вашей презентации (уверенность, ответы на вопросы преподавателя и т.д.).

В группе должны презентовать оба человека: и слабые пункты, и ре-визуализацию. Преподаватель оставляет за собой право остановить презентующего и попросить коллегу продолжить. Оценка за презентацию общая на группу. Если презентует один человек, то максимальная оценка **0.5 балла.**

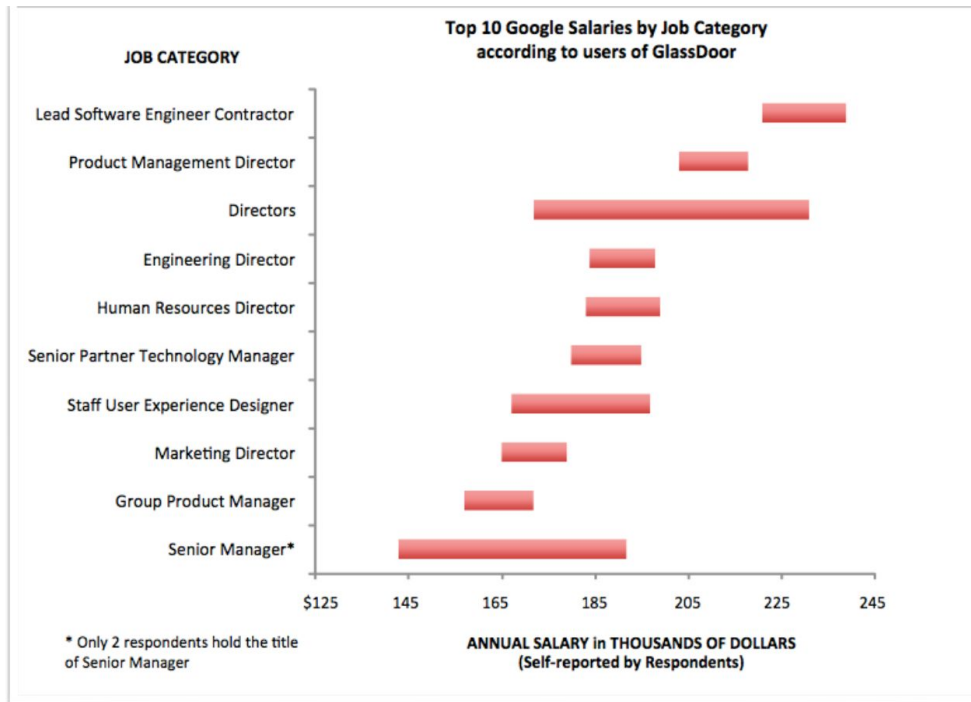
Источник: <http://viz.wtf/>

Или любая достаточно плохая визуализация

Какая визуализация лучше?



<http://bit.ly/top10salariesatgoogle>



<http://junkcharts.typepad.com/.a/6a00d8341e992c53ef014e8c55224e970d-pi>

KNOW YOUR POOP

0.9 KG

IS THE AVERAGE
AMOUNT A HUMAN
POOPS IN A DAY



1%
PROTEIN

4%
SALTS

75%
WATER

8%
INDIGESTIBLE
FIBERS

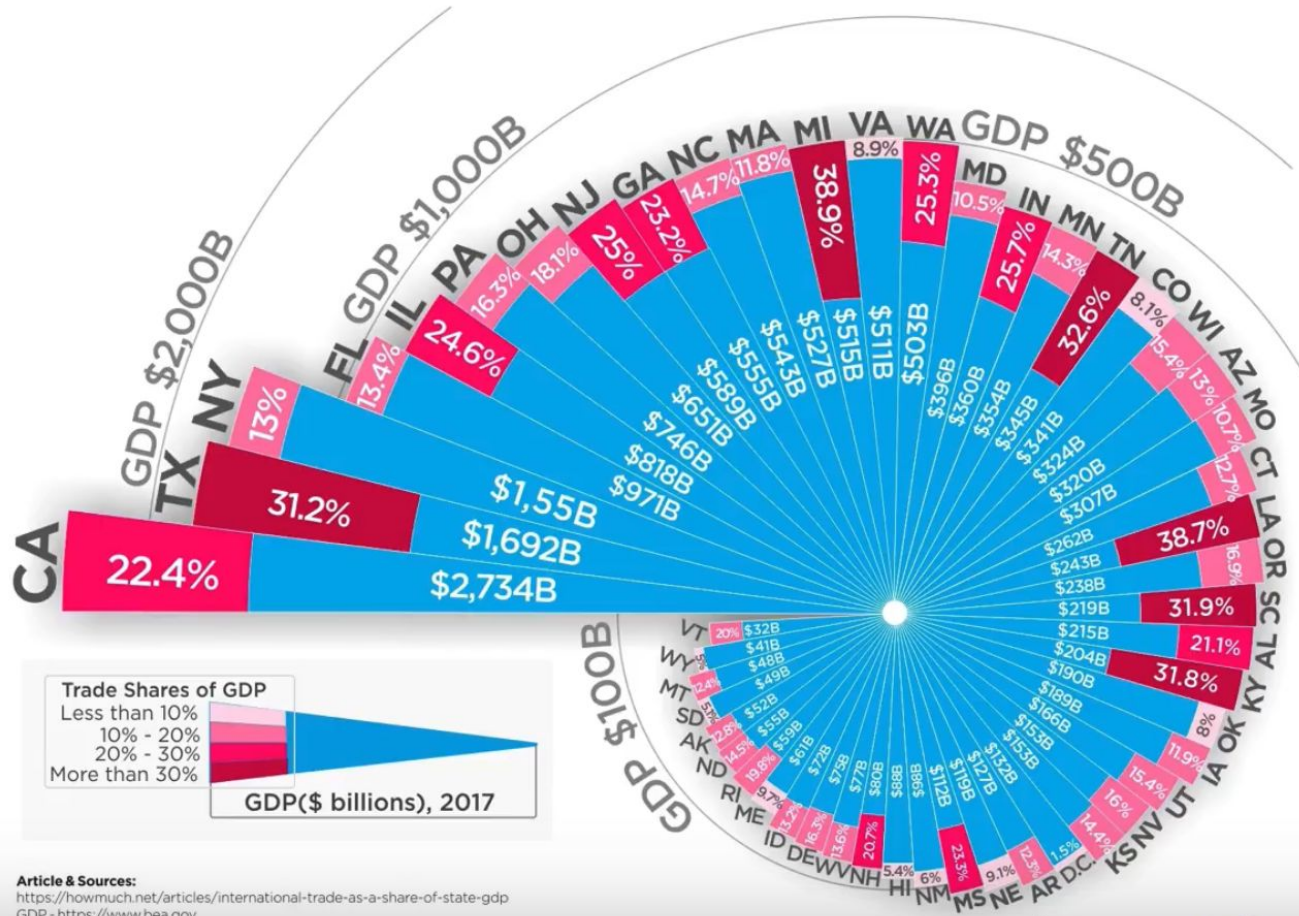
8%
DEAD BACTERIA

4%
FATS

POOP COLOR CHART



International Trade as a Share of State GDP (2017)



Article & Sources:

<https://howmuch.net/articles/international-trade-as-a-share-of-state-gdp>

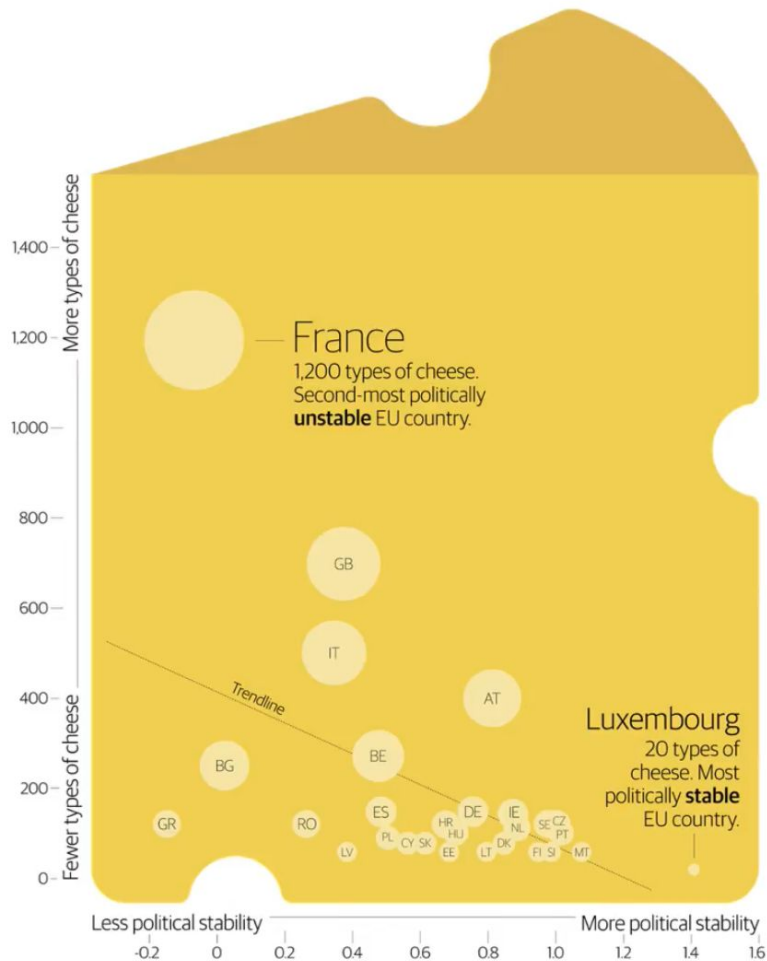
GDP - <https://www.bea.gov>

Trade - <https://www.census.gov>

Via: <https://www.aei.org>

Longer team name = more wins???





Source:
The EU cheese stability index,
POLITICO

DATA VISUALIZATION CHECKLIST Project with Stephanie Evergreen

Stephanie Evergreen and I knew some time ago that evaluators and social scientists had a thirst for better graphs, a clear understanding of why better graphs were necessary, but they lacked efficient guidance on how, exactly, to make a graph better. Introducing the Data Visualization Checklist.

Read more: annkemery.com/dataviz-checklist and stephanieevergreen.com/dataviz-checklist

Download the checklist: annkemery.com/wp-content/uploads/2014/05/DataVizChecklist_May2014.pdf

Who's Talking about the Checklist

Arrangement

Improper arrangement of graph elements can confuse readers at best and mislead viewer at worst. Thoughtful arrangement makes a data visualization easier for a viewer to interpret.

Proportions are accurate

A viewer should be able to take a ruler to measure the length or area of the graph and find that it matches the relationship in the underlying data.

2 1 0 n/a

Data are intentionally ordered

Data should be displayed in an order that makes logical sense to the viewer. Data may be ordered by frequency counts (e.g., from greatest to least for nominal categories), by groupings or bins (e.g., histograms), by time period (e.g., line charts), alphabetically, etc.

2 1 0 n/a

Axis intervals are equidistant

The spaces between axis intervals should be the same unit, even if every axis interval isn't labeled.

2 1 0 n/a

Graph is two-dimensional

Avoid three-dimensional displays, bevels, and other distortions.

2 1 0 n/a

Display is free from decoration

Graph is free from clipart or other illustrations used solely for decoration. Some graphics, like icons, can support interpretation.

2 1 0 n/a

https://depictdatastudio.com/wp-content/uploads/2014/05/DataVizChecklist_May2014.pdf

MISLEADING DATA VISUALISATION AND BEST CHARTING PRACTICES

▶ 🔊 1:16 / 20:33

📺 ⚙️ HD YouTube 🔍

<https://youtu.be/yntLglYUHXI>



David McCandless turns complex data sets (like worldwide military spending, media buzz, Facebook status updates) into beautiful, simple diagrams that tease out unseen patterns and connections. Good design, he suggests, is the best way to navigate information glut -- and it may just change the way we see the world.

https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization

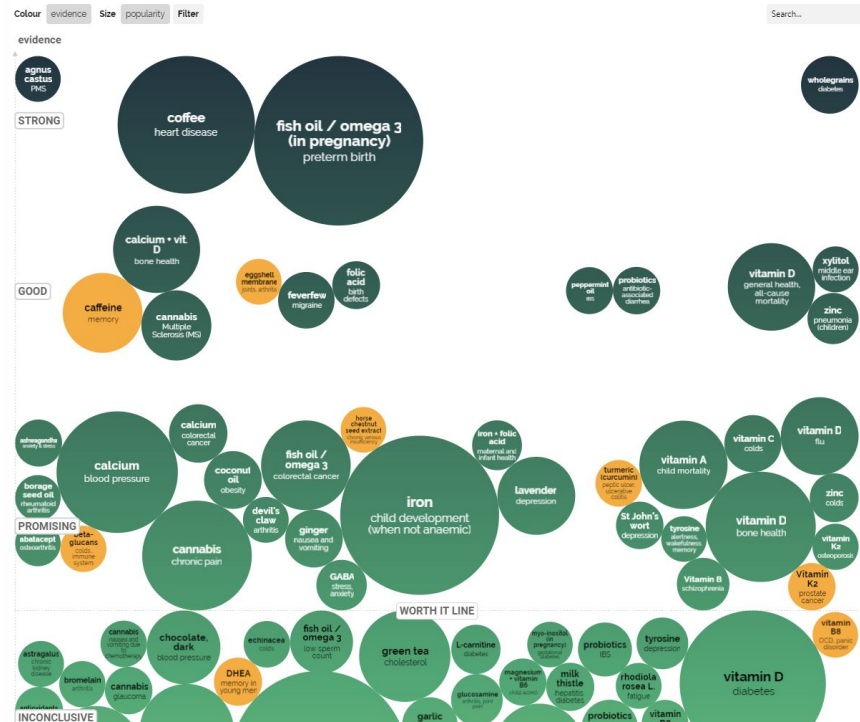
И хорошая
визуализация,
чтобы
закончить на
позитивной
ноте :)

Snake Oil Supplements?

Scientific evidence for popular health supplements

Showing tangible human health benefits when taken orally by an adult with a healthy diet

Last updated: 24 April 2019



<https://www.informationisbeautiful.net/visualizations/snake-oil-scientific-evidence-for-nutritional-supplements-vizsweet/>