

Визуализация данных: Что лучше работает?



Эдвард Тафте, профессор Йельского университета хорошо известен своей новаторской работой в области информационного дизайна, исследовал картографию для упрощения изучения всех типов информации, а также визуализации данных, которая фокусируется на визуальных картографах для отображения данных.

Classic Readings:

Edward Tufte (Professor Emeritus at Yale University) has written four books on analytical design:

1983: The Visual Display of Quantitative Information

1990: Visual Explanations: Images and Quantities, Evidence and Narrative

1990: Envisioning Information

2006: Beautiful Evidence

HOME BOOKS ONE-DAY COURSE ET NOTEBOOKS SCULPTURE PRINTS POSTERS, GRAPH PAPER ABOUT ET CART

THE WORK OF EDWARD TUFT AND GRAPHICS PRESS GRAPHICS PRESS LLC P.O. BOX 430 CHESHIRE, CT 06410 800 822-2454

Edward Tufte is a statistician and artist, and Professor Emeritus of Political Science, Statistics, and Computer Science at Yale University. He wrote, designed, and self-published 4 classic books on data visualization. *The New York Times* described ET as the "Leonardo da Vinci of data," and *Bloomberg* as the "Galileo of graphics." He is now writing a book/film *The Thinking Eye* and constructing a 23-acre tree farm and sculpture park in northwest Connecticut, which will show his artworks and remain open space in perpetuity. He founded Graphics Press, ET Modern gallery/studio, and Hoggan Hill Farms LLC.

PRESENTING DATA AND INFORMATION: A ONE-DAY COURSE TAUGHT BY EDWARD TUFT

Topics covered in this [one-day course](#) include:

A new, widely-adopted method for presentations: meetings are smarter, more effective, 20% shorter.

Fundamental design strategies for all information displays: sentences, tables, diagrams, maps, charts, images, video, data visualizations, and randomized displays for making graphical statistical inferences.

New ideas on spectatorship, consuming reports. How to assess the credibility of a presentation and its presenter, how to detect cherry-picking, how to reason about alternative explanations.

Standards of comparison for workaday and for cutting edge visualizations. How to identify excellent information architectures and use them as models and comparison sets for your own work and for the work of your contractors. Monitoring the designs of others.

The future of information displays: 4K, 6K, 8K video maps moving in time.

Practical examples are from everywhere: science, social science, music, business, finance, sports, art, medicine, architecture, NASA, government reports.

Edward Tufte teaches the entire course.

Each student receives all four ET books on information design:



"One visionary day—the insights of this class lead to new levels of understanding both for creators and viewers of visual displays." *Wired*

"The 7 essentials to Visual of Data" *your company needs*



"An absolutely beautiful film. It picks up where *Helvetica* left off. Inge Druckrey's wonderful teaching is an inspiration." *Luke Coulson*, cinematographer of *Helvetica*. "A great story beautifully told." *Ken Cuthren* An ET MODERN film, 37 minutes, all for free click above.

ART: ET NOTEBOOKS

Hoggan Hill Farms artworks
Continuous ascent: structures of unknown significance
The "Trapezoidal" artworks made from steel and air

Arbitrarily Interpretative Explorer
Clever taking flight (200 Bromwichmont)
Magically finite (Spain Fall)
Floating Steel Caribbeion
Sloping sculptures
Blindfold blind engineering
Artful Feynman Diagrams

Philosophical Charmed Signs
Bookprints: 16 new prints
Rascal Science #1 (spacecraft)

Reckless Bottom #2 (Lunar Landing)
Territorial urbanism, original steel
Image data suite: our new website
Beautiful Evidence chronicles
Tong Bird of Paradise

Drawing Center show: ET at
Moose Quartz, building, 2011
Diggroot: "Whose home is it?"
Designing a museum garden
Ally Art

SCIENCE: ET NOTEBOOKS

Practical advice for medical patients
Sentences of the One
Developing data graphics
Skeptical theory and practice
Artful Feynman Diagrams

Cherish
Little design and construction, by Edward Tufte
Maps moving in time: a standard of excellence for data change
Beautiful Evidence chronicles
Advice for effective analytical reasoning

The shape of Saturn is blue
Megan: engineers' brilliant new graphics
Making better inferences from statistical graphics Edward Tufte
Image data suite: our new website
Cancer survival rates: tables, graphs
Feynman: Tube Principles
Statistics with ET

Tube and timeline design and typography
Destinations: decision displays
Field color charts: submit, test, dig
Feynman: "Nature cannot be fooled"
Bottom images in Goddard's test
Pioneer course plaque

Phone interface design
Diverse taking flight (200 Bromwichmont)
Biogeography for comparing gradients
Stoicgraph theory and practice
The meaning of "analysis"
Conrad: presentation techniques
Walking maps (and minimizing the usual mistakes)
London Underground maps
Lunar Landing



Целостность графики

- говорить правду о данных
- графика не должна цитировать данные вне контекста
- избегайте искажения в данных.
- представление чисел должно соответствовать ИСТИННЫМ пропорциям.
- маркировка должна быть четкой и подробной.
- для представления финансовых данных лучше всего использовать хорошо известные единицы измерения

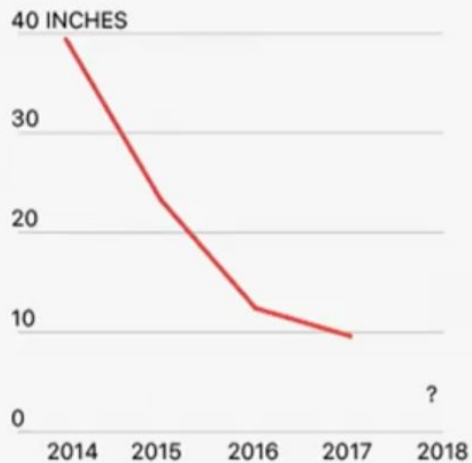
ДАННЫЕ ДОЛЖНЫ ОСНОВЫВАТЬСЯ
ТОЛЬКО НА ПРАВДЕ!



Снегопад

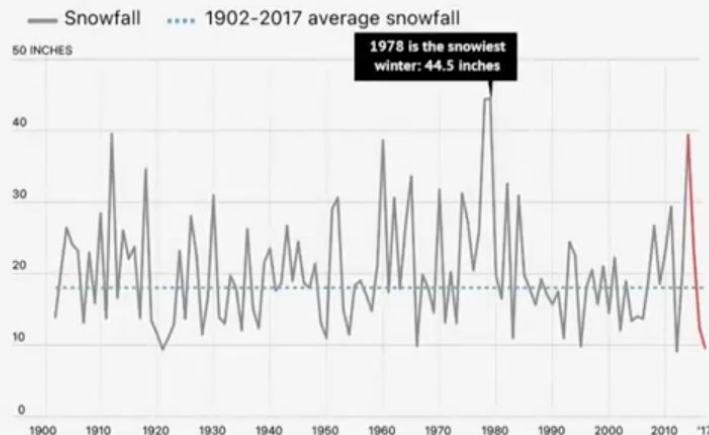
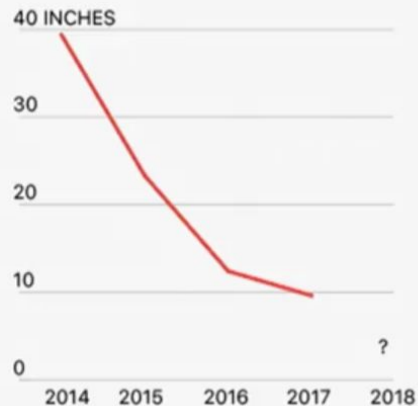
Context Is Essential

SNOWFALL TRENDS IN ILLINOIS



Context Is Essential

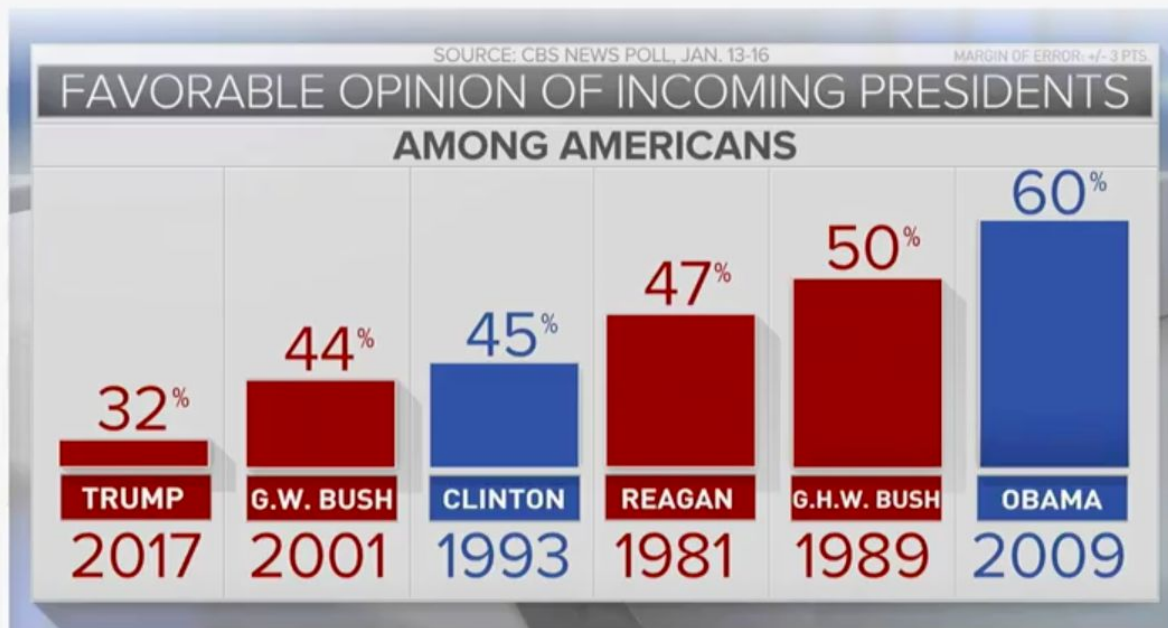
SNOWFALL TRENDS IN ILLINOIS



Когда мы видим график, где представлено больше данных за последние десятилетия или даже столетие, получается, что количество снега колеблется, и поэтому количество снега, которое мы ожидаем в 2018 году, не является точным.

How to Lie with Visualizations

No zero line

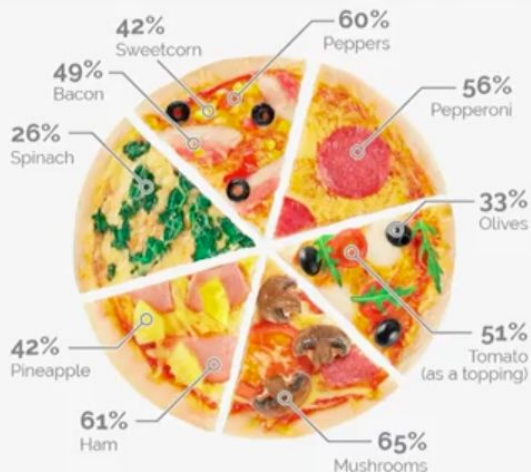


How to Lie with Visualizations

Doesn't add up to 100% and not in right proportion

Mushroom is the UK's most liked pizza topping

Generally speaking, which of the following toppings do you like on a pizza? Select as many as you like



Other items not depicted include: onions (52%), chicken (50%), beef (36%), chillies (31%), jalapeños (30%), pork (25%), tuna (22%), anchovies (18%), 2% of people say they only like Margherita pizzas

YouGov | yougov.com

February 26-28, 2017

Source: YouGov / Matthew Smith / <http://bit.ly/2lLzNFa/>



of domain owners have put moderate to high consideration to how much their domains are worth



of respondents have bought and sold domain names for a profit



of small business owners want to make time to enhance or update their online presence

Source: GoDaddy



Другой способ лгать - это манипулирование областью и размером. Тафте разработал концепцию под названием "Фактор лжи", которая относится к соотношению между размером эффекта, показанного на графике, и размером эффекта, показанного в данных.

Формула.

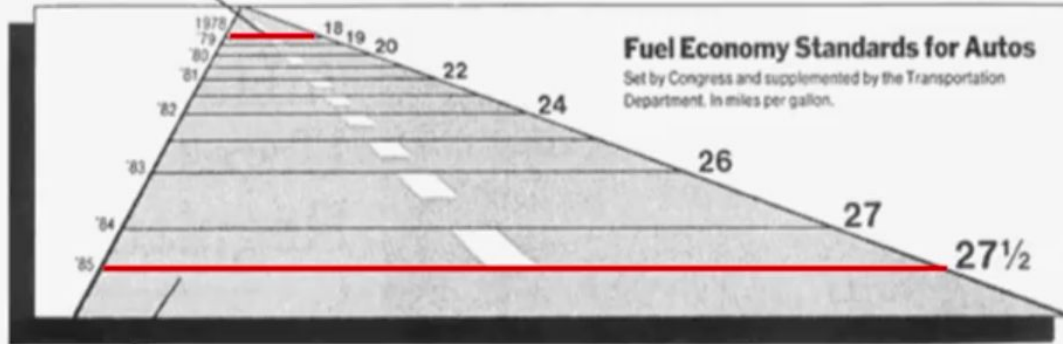
Представление чисел как физической меры на поверхности самого графика должно быть прямо пропорционально представляемой величине.

$$\textit{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

- This graphic (by *The New York Times*) tries to compare the mandated fuel economy standards for autos set by the US Department of Transportation.
- The line representing 18 miles per gallon in 1978, is 0.6 inches long. The line representing 27.5 miles per gallon in 1985, is 5.3 inches long.

$$\text{Lie Factor} = \frac{\frac{5.3 - 0.6}{0.6}}{\frac{27.5 - 18}{18}} = 14.8$$

Inaccurate visual area and numeric measure!

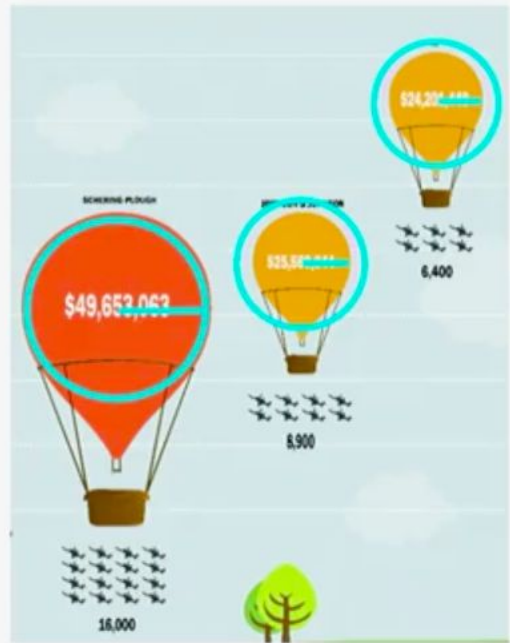


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

BALLOONING CEO SALARIES AND MASS LAYOFFS

When companies lay off workers, they are cutting costs, and it's one of the best ways to do that. But when CEO salaries are so high, it's hard to see how they can be cut. The chart below shows the relationship between CEO pay and layoffs. The chart shows that as CEO pay goes up, the number of layoffs goes down. This is a clear trend that is worth noting.

MOBI The Boss™



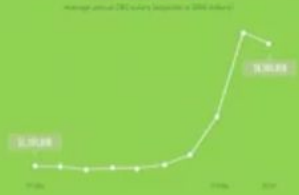
KEY



SLASHING JOBS PAYS



CEO SALARIES OVER TIME



CEO PAY AND UNEMPLOYMENT INSURANCE



Inaccurate visual area and numeric measure

ТЕСТ

Что такое фактор лжи?

A. Значение, подсчитывающее количество неточностей графика.

B. Значение отношения, которое некорректно отображает эффект данных из-за неточного размера графика.

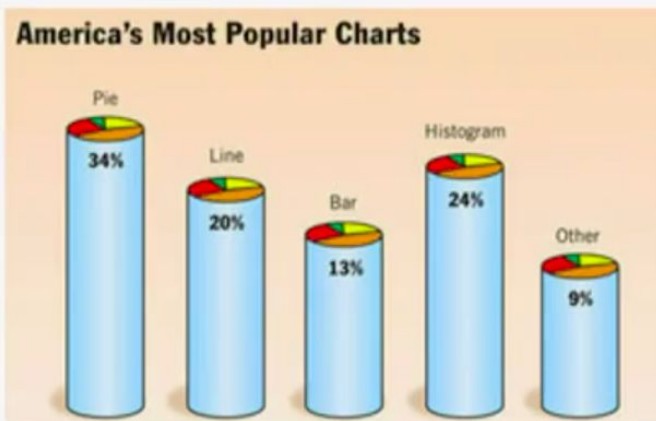
C. Значение коэффициента, отображающего эффект данных в связи с размером графика

D. Значение, подсчитывающее количество ложных данных.

Graphical Integrity (2)

Show data variation, not design variation.

- Induce the viewer to think about the substance rather than about the data.



Source: The Onion / <http://bit.ly/2GUQUOu>



Source: Junk Charts / <http://bit.ly/2UfPohw>

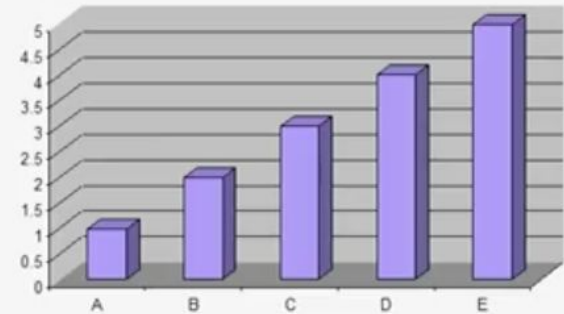
Тафте также просит дизайнера не путать вариации дизайна.

Chartjunk

- Chartjunk refers to anything in a chart that does not represent data as not just unnecessary, but harmful.

Common types:

- **Vibrating chartjunk** – which is cross-hatching or other patterns that distract the mind from the information being presented
- **Grids** – over-busy grid lines and excess ticks
- **Self-promoting graphics** (“The Duck chart”) – where color schemes and patterns are introduced for artistic appeal rather than information content







Chart Junk (графический мусор), относится ко всему, что представлено на графике кроме данных.





Соотношение чернильных данных (ink)

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used}}$$

Data-Ink Ratio

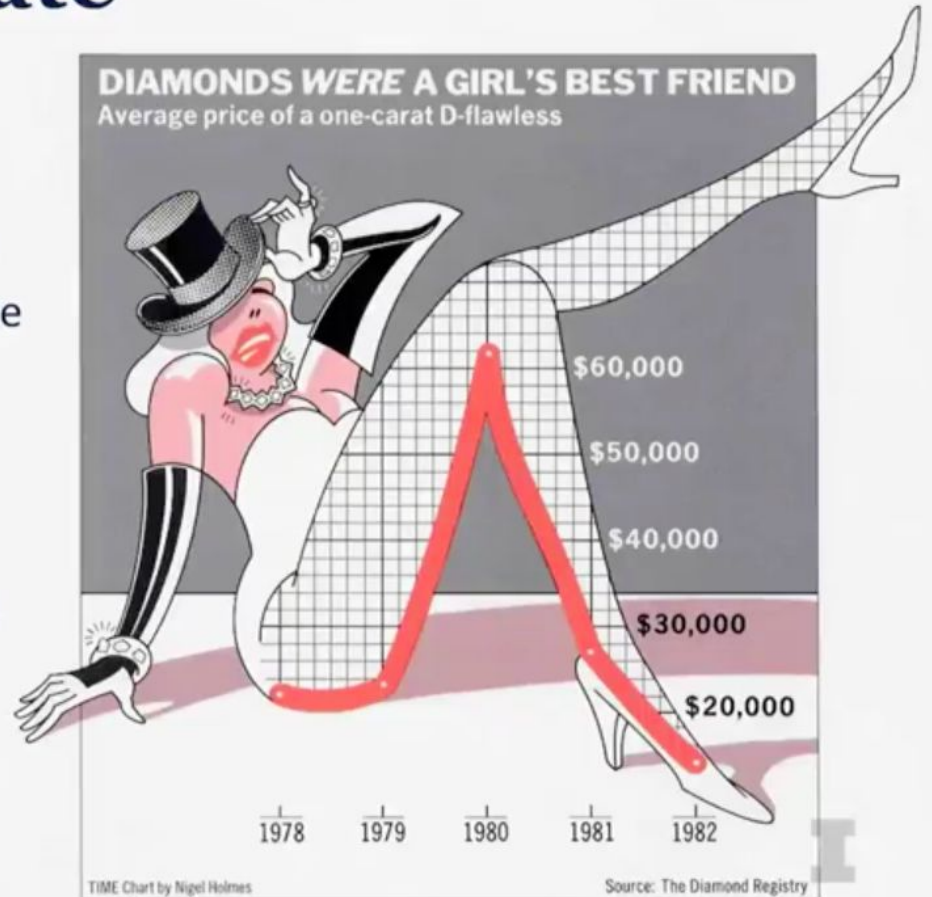
- It is an electroencephalogram – a graph that records the electrical activity from the brain.
- This graph would have a data-ink ratio of 1.



Использование декоративных элементов без цели может привести к искажению данных и неправильной интерпретации данных пользователем.

The Chartjunk Debate

- Holmes: “From ankle to knee, the price changes from lowest to highest over time. By combining image with graph, it made viewers remember the spike shape of the graph more easily.”
- Tufte: “Everything counts, but nothing matters...The data-thin (and thus uncontextual) chart mixes up changes in the value of money with changes in diamond prices...”



Chartjunk Тафта рассматривает как графический мусор

The Chartjunk Debate

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT

Guidelines for designing information charts often state that the presentation should reduce 'chart junk' – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords

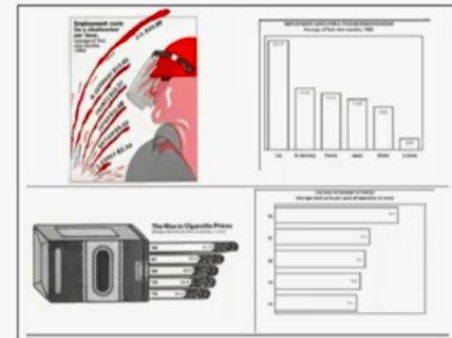
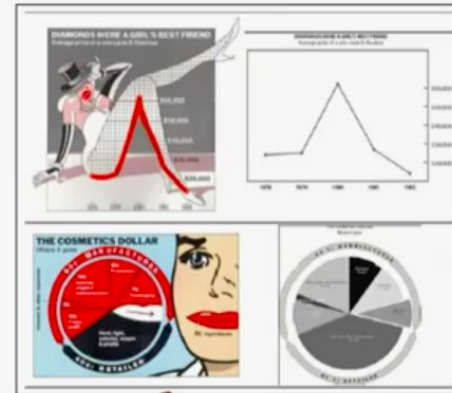
Charts, information visualization, imagery, memorability.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

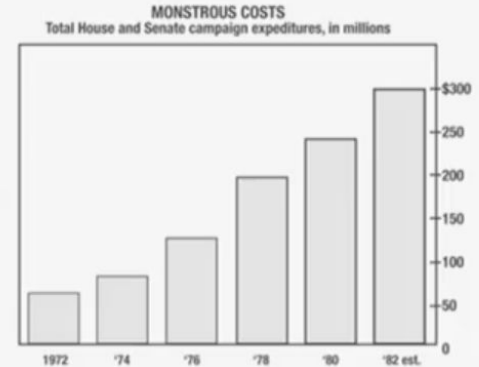
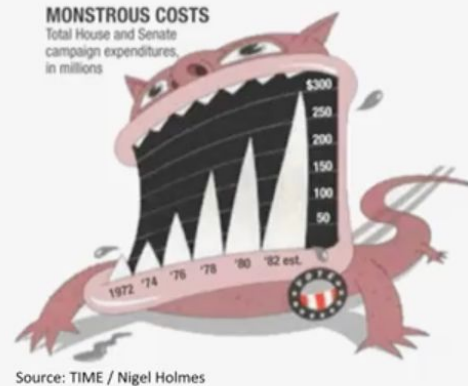
data-ink – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).



Criticism

- Readers do not find highly embellished charts difficult or easy to comprehend, but they do remember those charts better and find them more attractive and more enjoyable to look at than their plainer alternatives. (Bateman et al., 2010)
- Borkin et al. (2013) also found that visually distinctive infographic style charts were more memorable than more standard statistical charts.



Читатели считают, что **chartjunk** не влияет на понимание читателя, но на самом деле читатели запоминают, что приукрашенные **chartjunk** смотрятся интересней, чем те же обычные графики.

Data-Ink Ratio Debate

- Anderson et al. (2011) asked participants to choose the box plot with the largest range from a set
- Six representations
- Measured cognitive load from EEG brain waves
- The simplest (c) box plot is the hardest to interpret

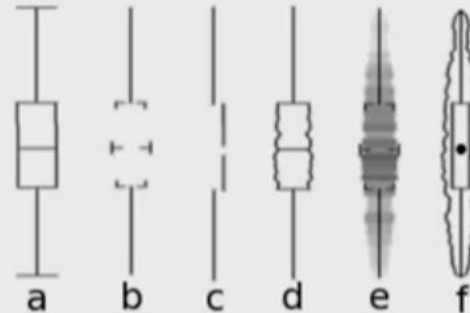
A User Study of Visualization Effectiveness Using EEG and Cognitive Load

E. W. Anderson¹, K. C. Potter¹, L. E. Matzen², J. F. Shepherd², G. A. Preston³, and C. T. Silva¹

¹SCI Institute, University of Utah, USA

²Sandia National Laboratories, USA

³Utah State Hospital, USA





Мы должны помнить о пяти принципах построения графиков данных.

- максимизируйте соотношение "чернил" в пределах разумного
- стирайте не чернил в пределах разумного
- стирайте избыточные чернила в пределах разумного
- пересматривайте и редактируйте

Тафте предлагает сделать большие наборы данных согласованными, *представив множество данных наименьшим образом.* Другими словами, *использовать наименьшее количество чернил в самом маленьком пространстве.*

Small Multiple

Uses a single design repeated several times within the eyespan, each example showing a different value of the independent variable(s)

Allows easy comparison



Published July 19, 2012

FACEBOOK TWITTER GOOGLE+ EMAIL SHARE

Drought's Footprint

More than half of the country was under moderate to extreme drought in June, the largest area of the contiguous United States affected by such dryness in nearly 60 years. Nearly 1,300 counties across 29 states have been declared federal disaster areas. Areas under moderate to extreme drought in June of each year are shown in orange below. Related Article »



Source: The New York Times / Haeyoun Park and Kevin Quealy / <https://gox>



Основные типы маркировки включают в себя:

- заголовок диаграммы
- заголовок осей
- этикетки и шкалу
- некоторые примечания, которые отмечают специфику внутри диаграммы

Совершенство достигается не тогда, когда больше нечего добавить, а там, где больше нечего убрать.