

Design Principles

Data Visualization SS23

Lecturer: Asil Çetin-Aufricht, MSc

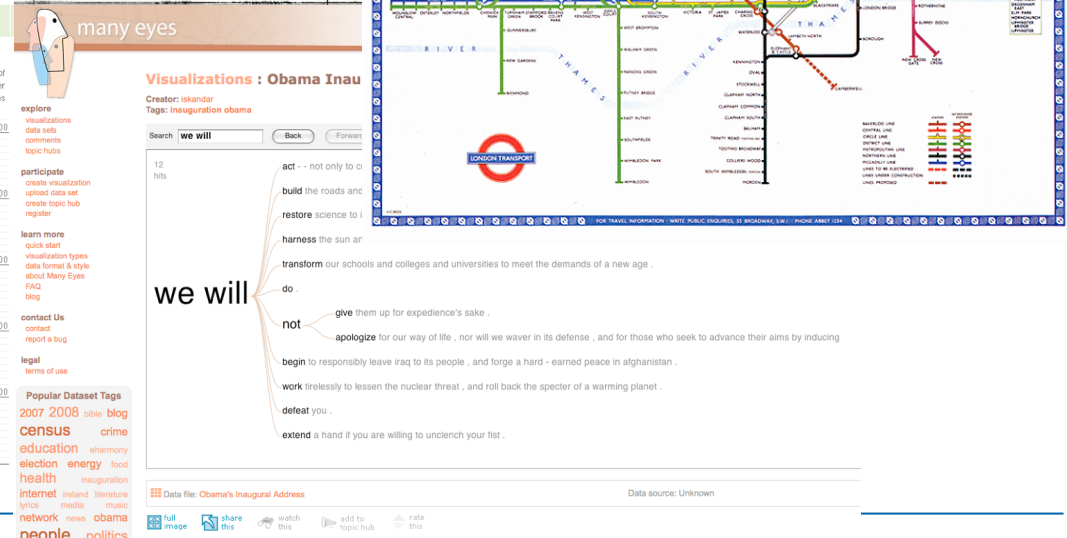
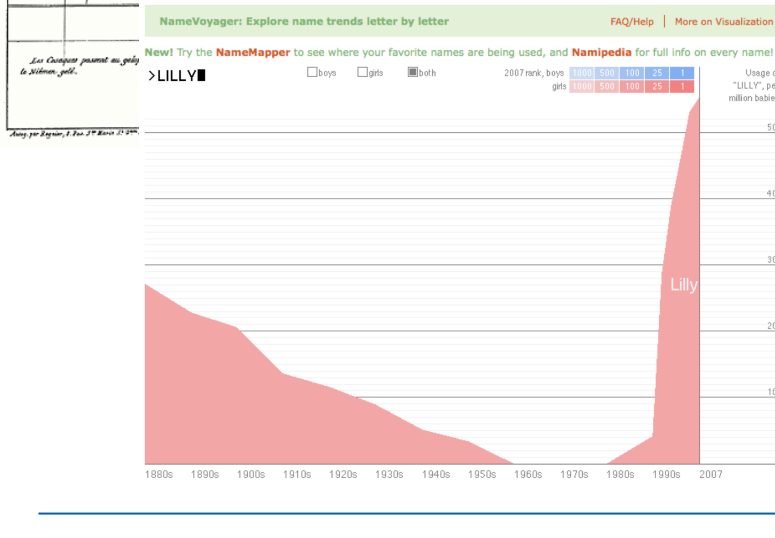
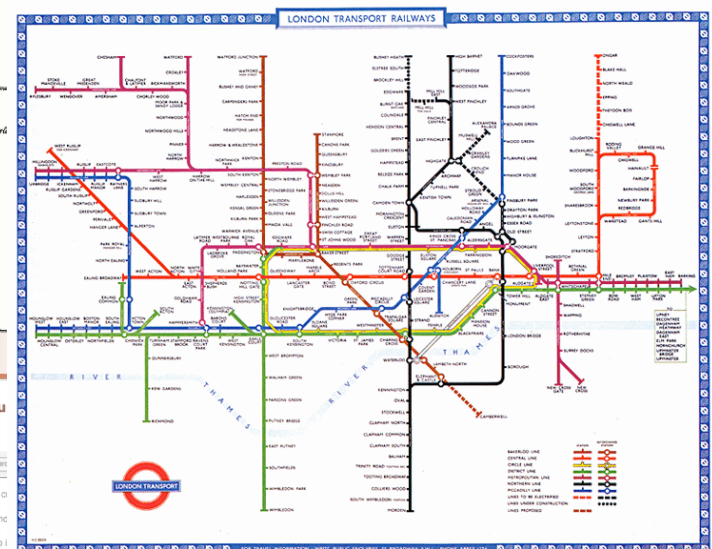
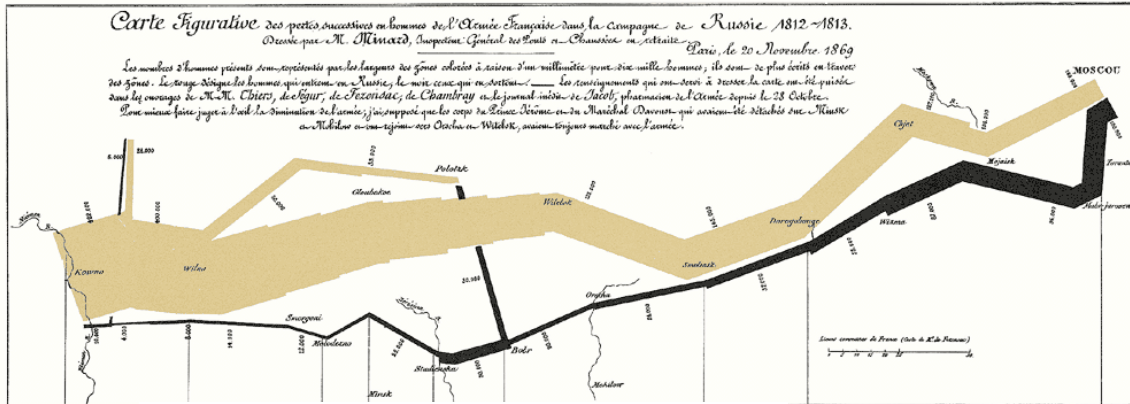


WHY do we create visualizations?



Visualization

To convey information through visual representations



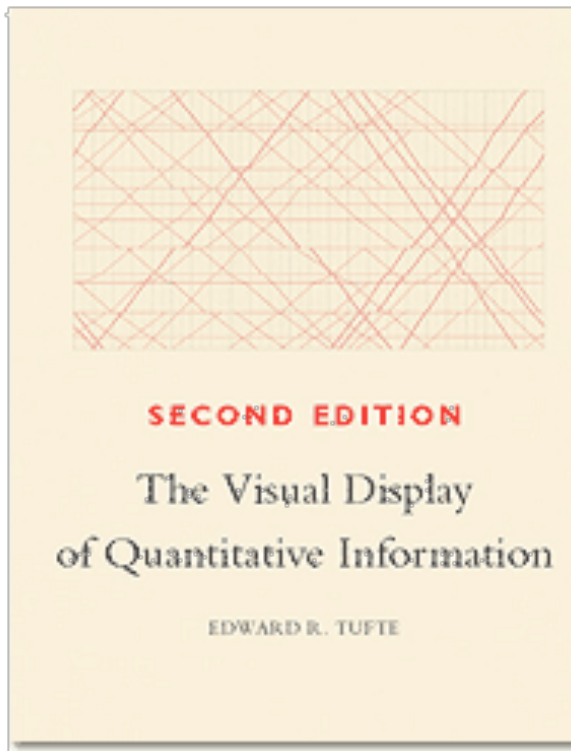
Design Excellence

“Well-designed presentations of interesting data are a matter of substance, of statistics, and of design.”

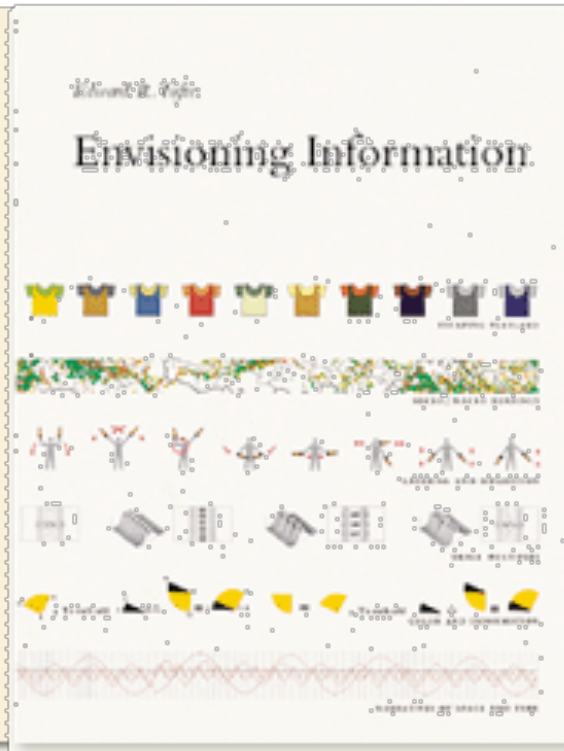
E. Tufte



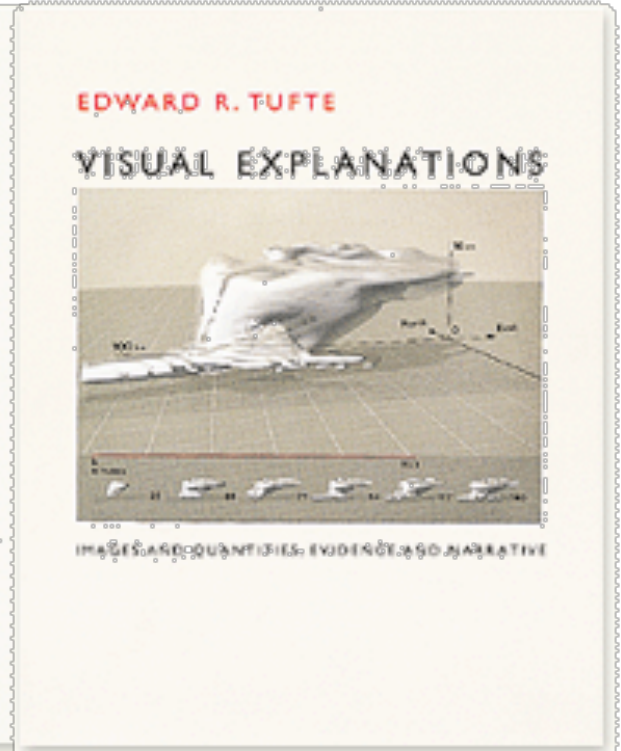
Edward Tufte



Tufte, VDQI, 2nd edition, 2001



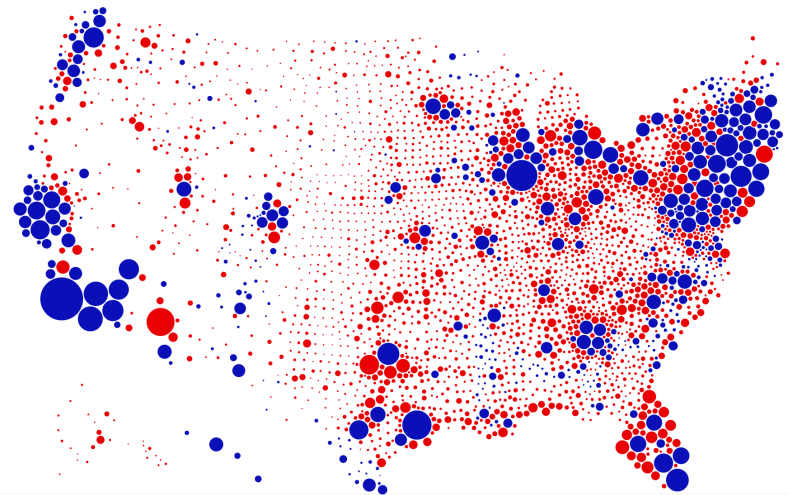
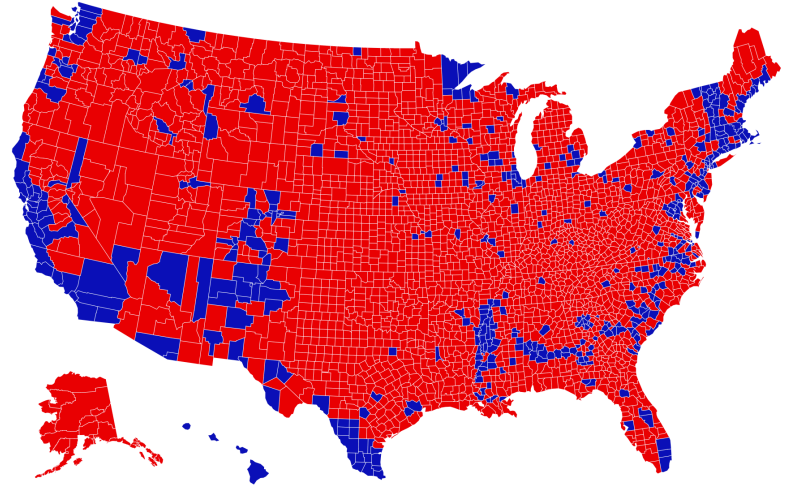
Tufte, Envisioning Information, 1990



Tufte, Visual explanations, 1997

Outline of Today

- **Graphical Integrity**
 - Visual Encoding
 - Lie Factor
 - Distortions
- **Design Principles**
 - Avoid Chartjunk
 - Maximize Data-Ink Ratio
 - Data Density
 - Tufte's Design Principles
 - Design Pyramid
- **Design Elements**
 - C.R.A.P.
 - Layering and Separation
 - Tufte's Graphical Excellence
 - Analysis Questions
- **Further Reading**



Surface area of US counties and their associated voting population from the 2016 U.S. presidential election. [Source: Karim Douieb.](#)

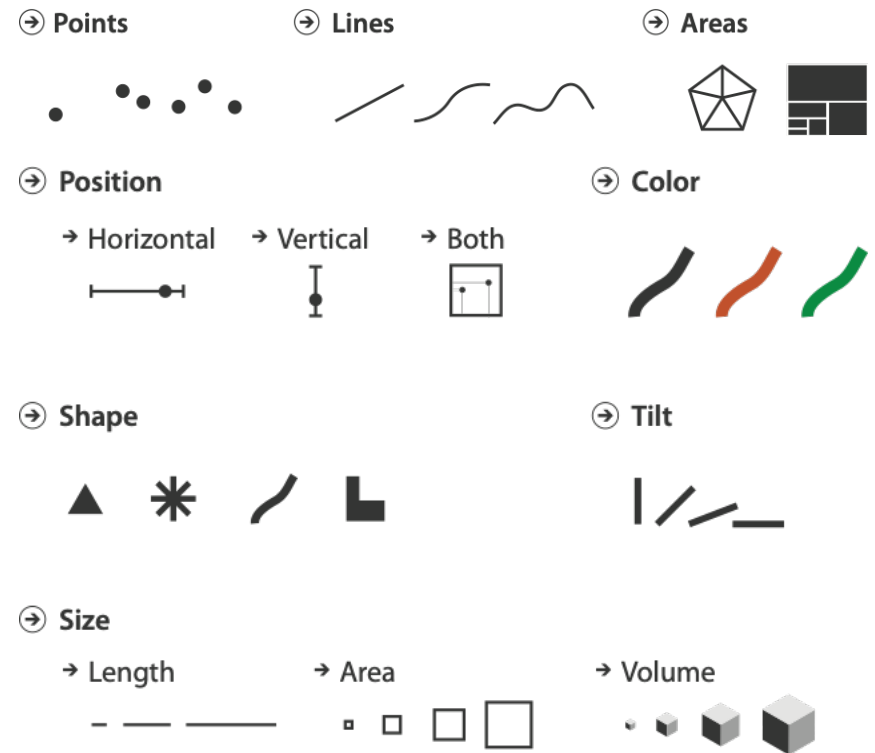
Graphical Integrity

Visual Encoding

= the way in which data is mapped into visual structures

Building blocks:

- Marks
- Channels



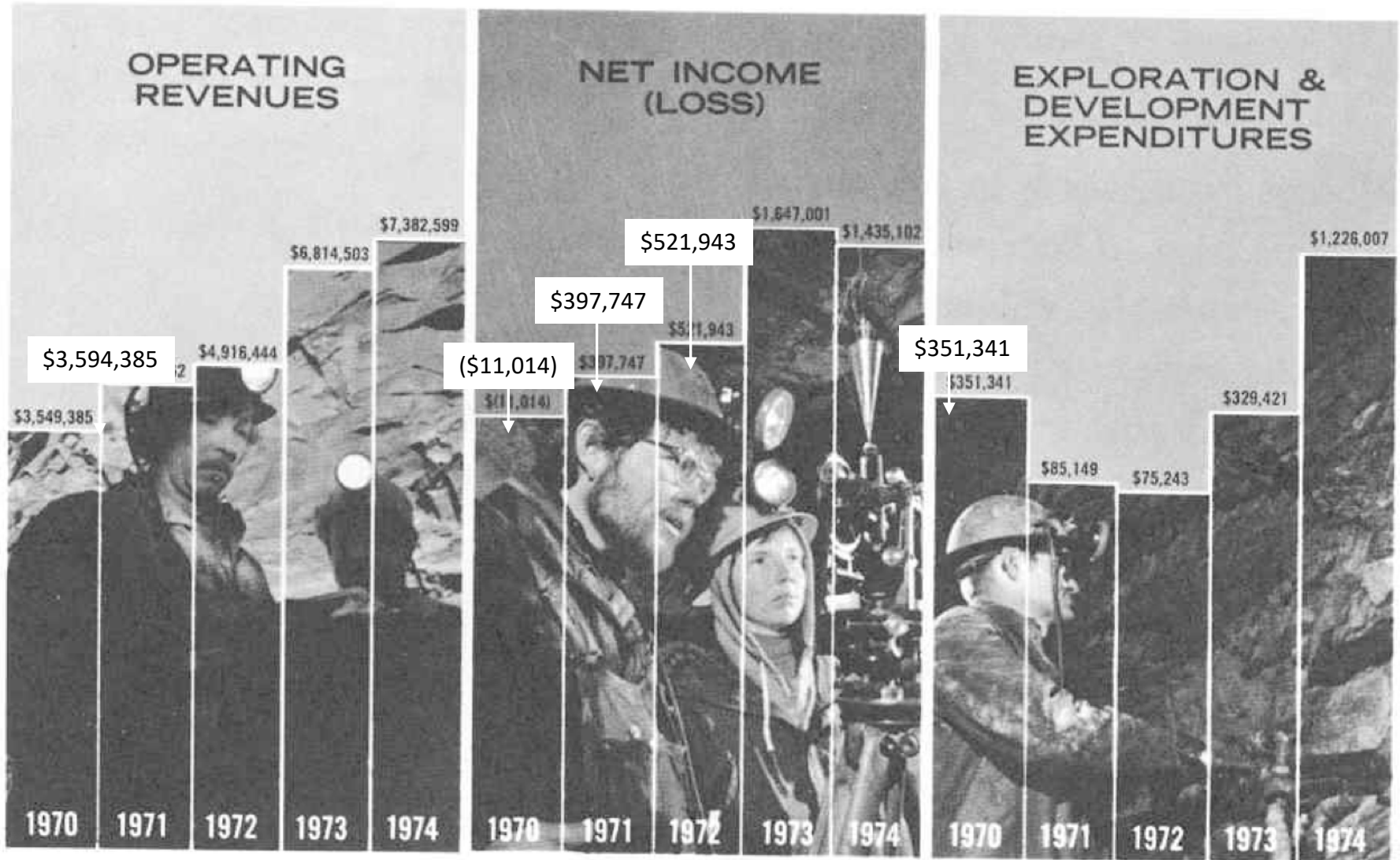
Munzner, Visualization Analysis and Design, 2014, p. 96

Lie factor

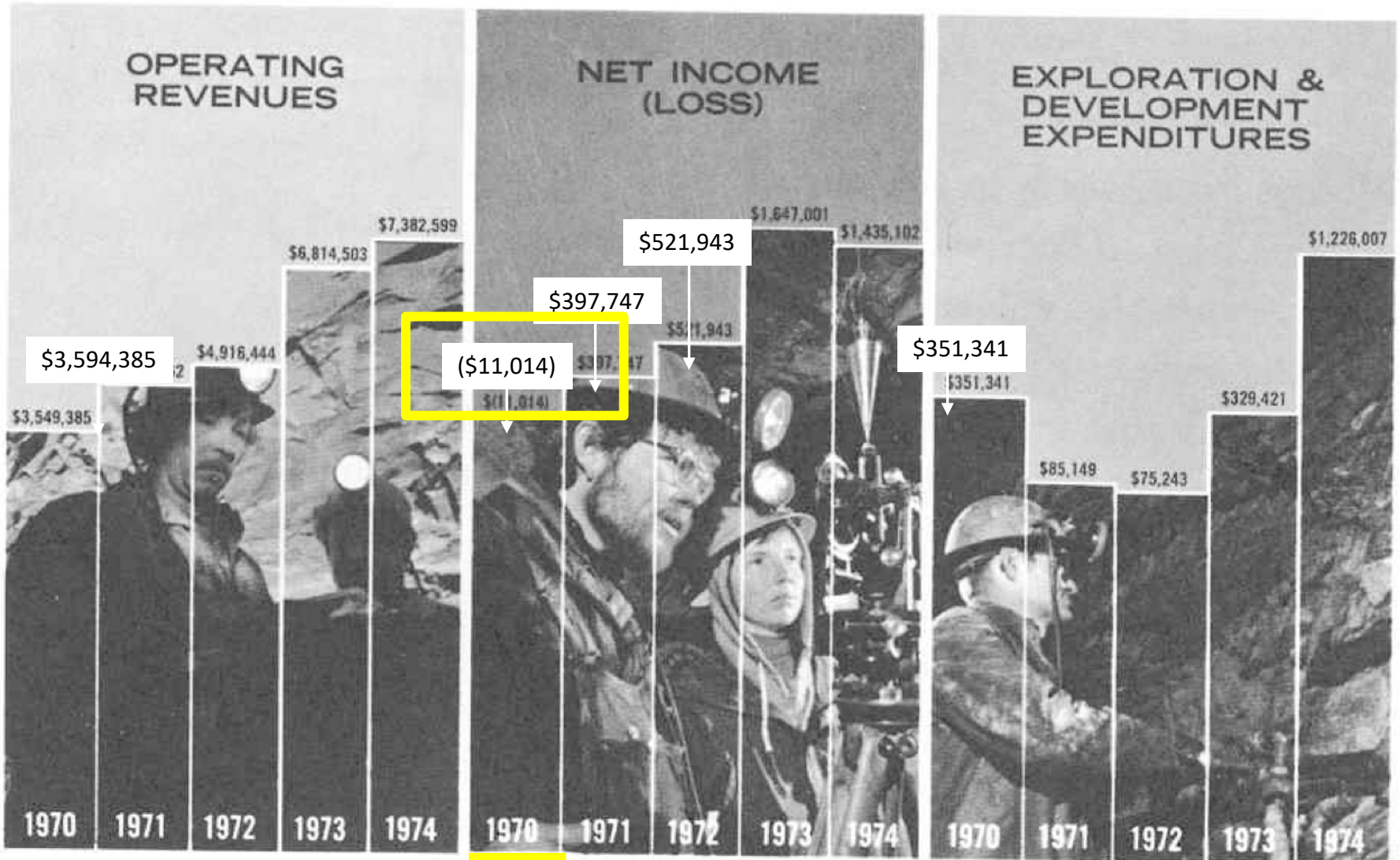


Photo by [Jametlene Reskp](#) on [Unsplash](#)

Lie Factor: Missing Scales



Lie Factor: Missing Scales



Baseline?

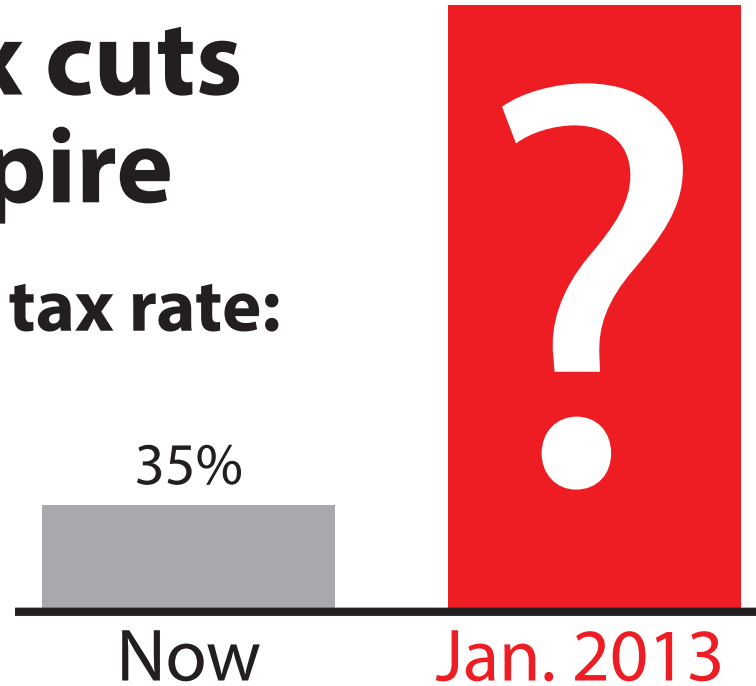
-\$4,200,000

Tuft, VDQI

Lie Factor

**If Bush
tax cuts
expire**

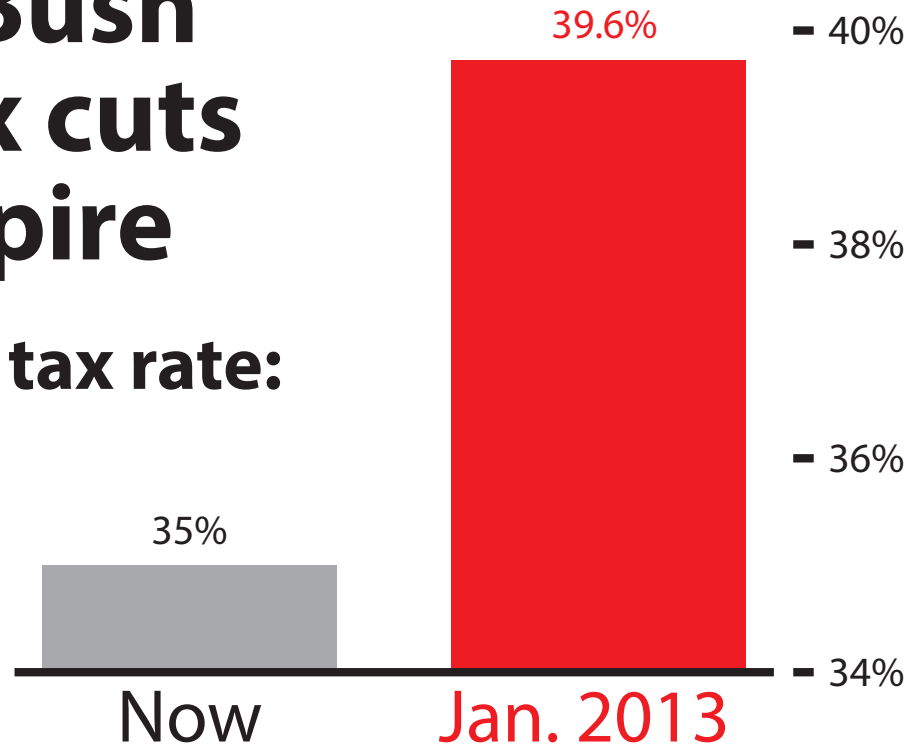
Top tax rate:



Lie Factor

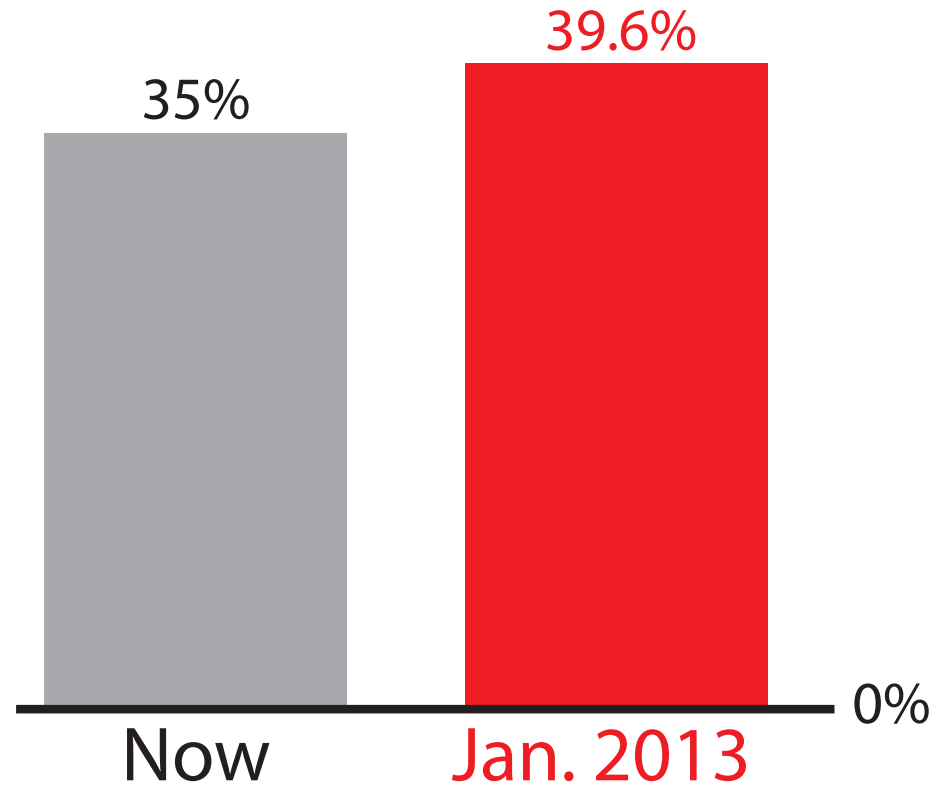
**If Bush
tax cuts
expire**

Top tax rate:

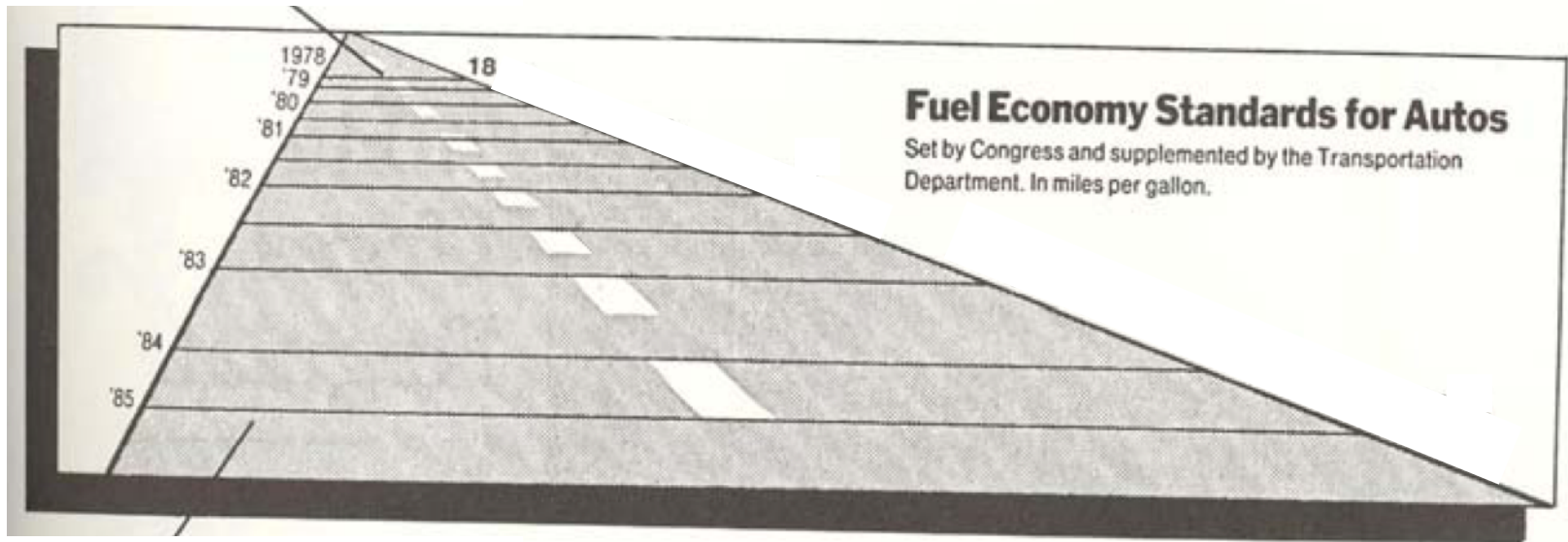


Lie Factor

**If Bush
tax cuts
expire
Top tax rate:**



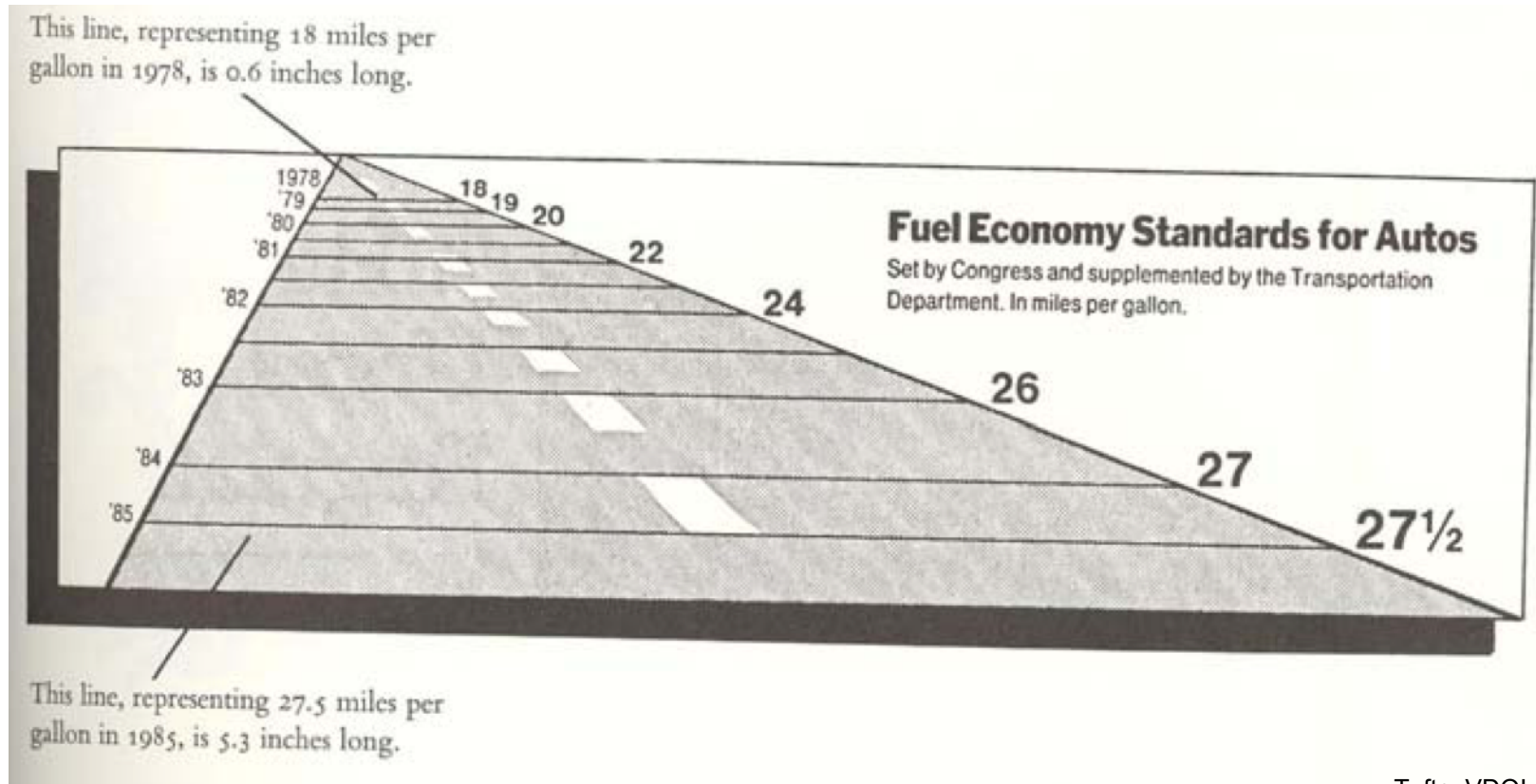
Lie Factor



Tufte, VDQI

Lie Factor

- (Size of effect in graphic)/ (size of effect in data)



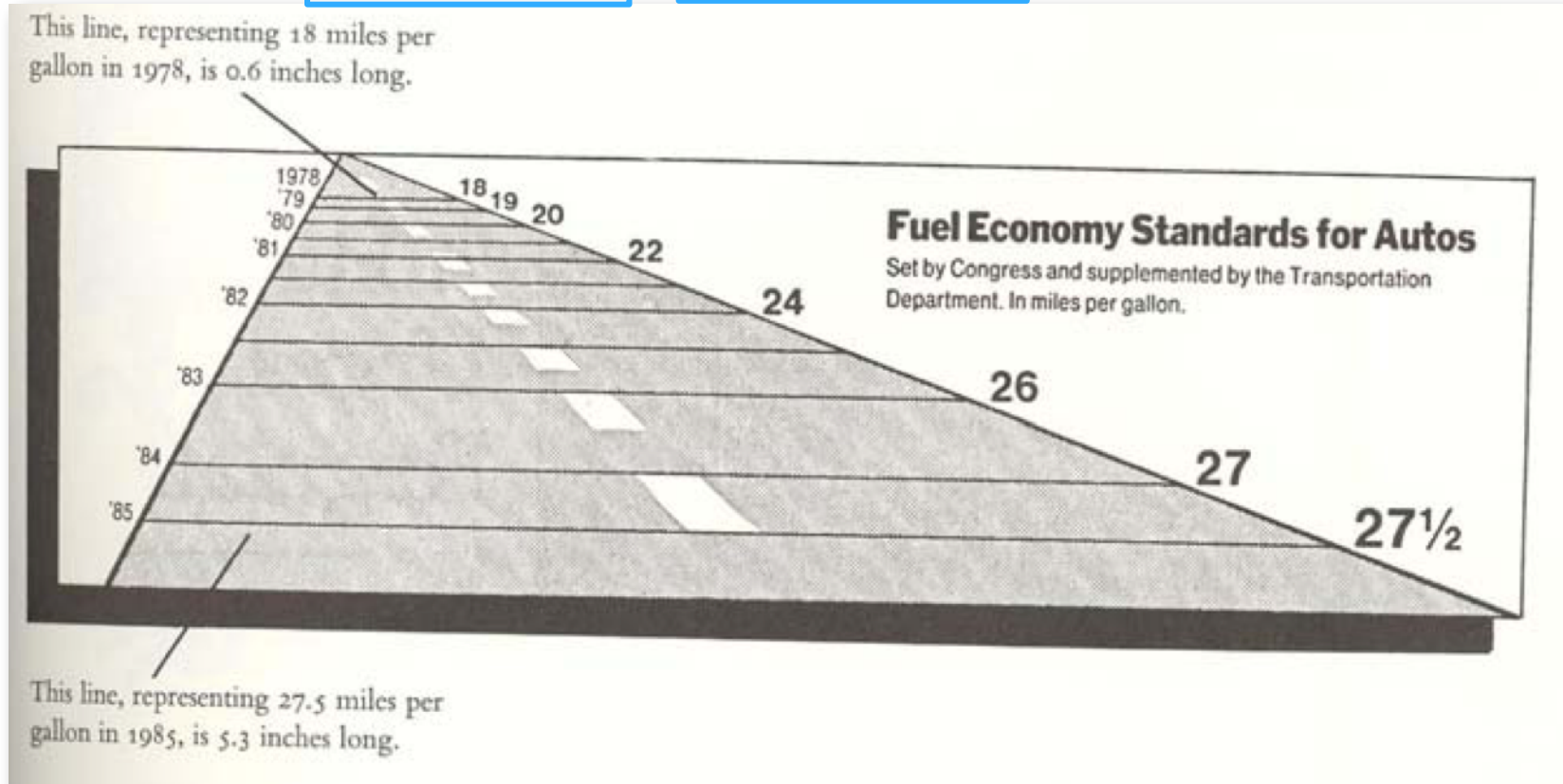
Tufte, VDQI

Lie Factor

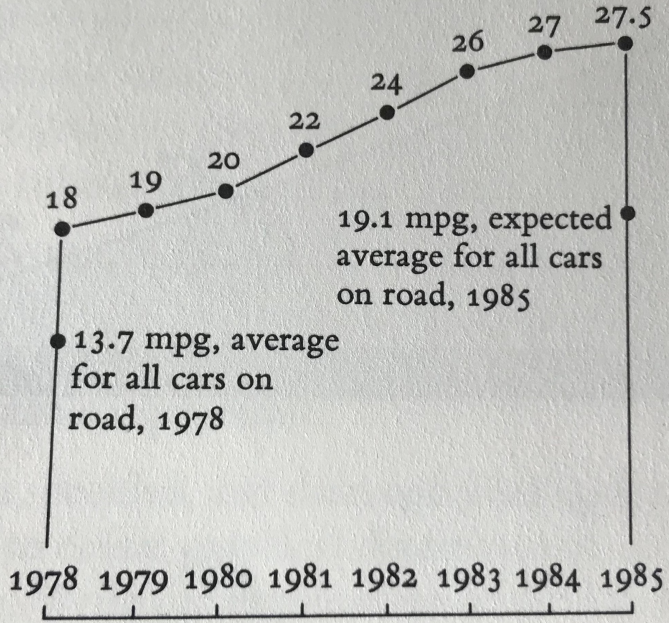
Size of effect
in graphic

$$\frac{5.3 - 0.6}{0.6} \bigg/ \frac{27.5 - 18}{18} = 14.8$$

Size of effect
in data



REQUIRED FUEL ECONOMY STANDARDS:
NEW CARS BUILT FROM 1978 TO 1985

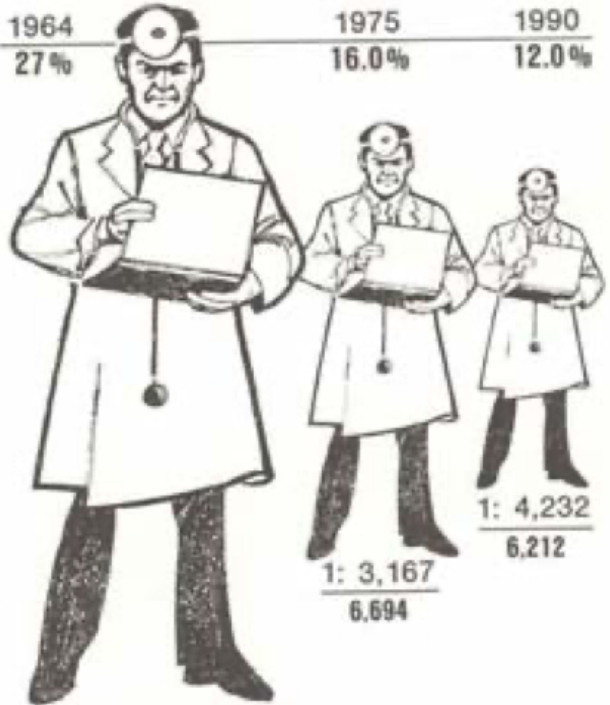


Lie Factor

THE SHRINKING FAMILY DOCTOR In California

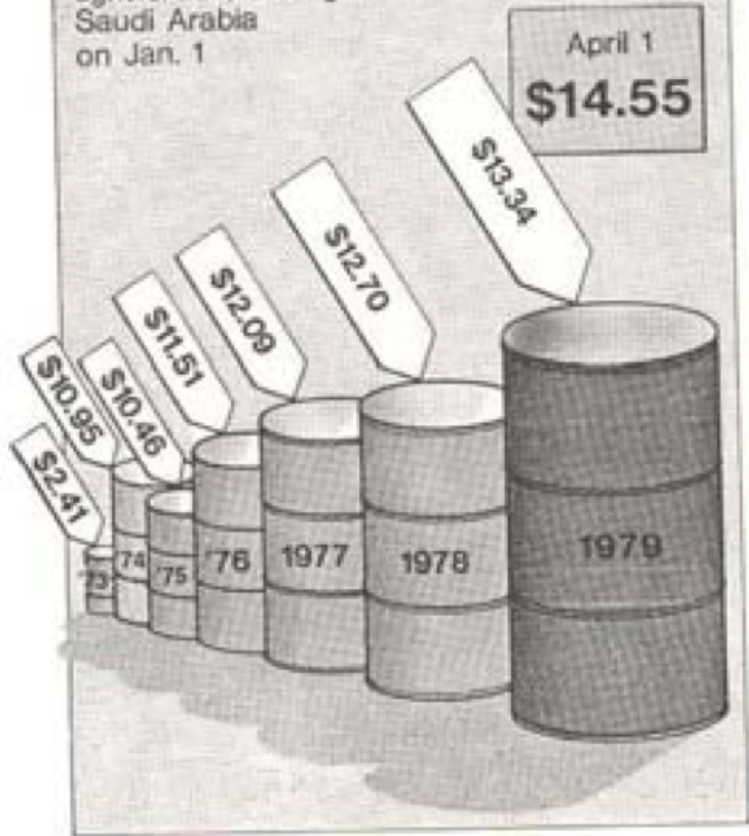
Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%

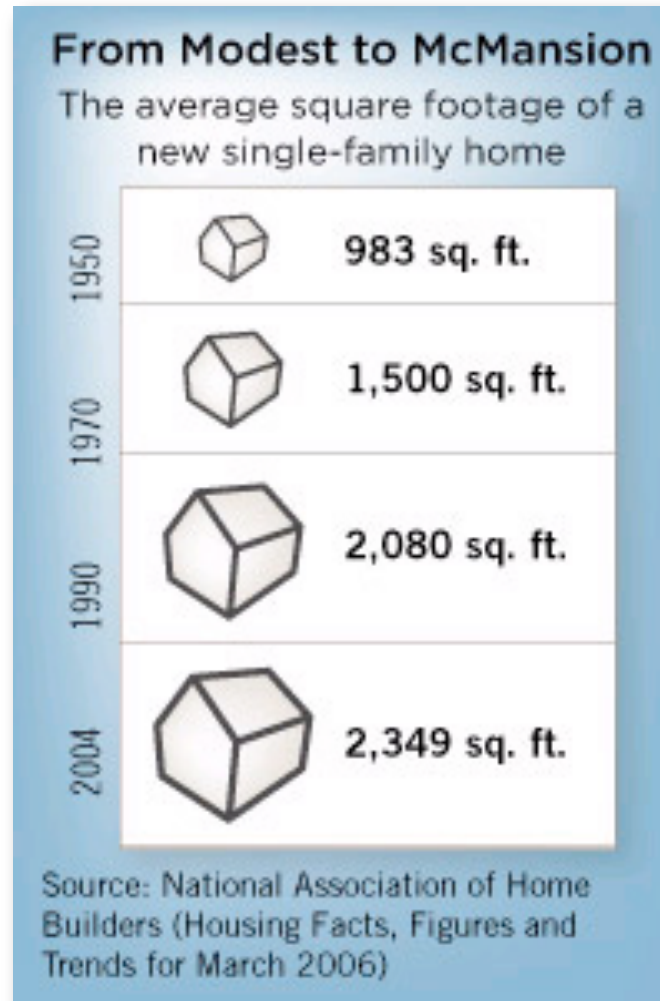


IN THE BARREL...

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



Lie Factor



<http://www.theoil drum.com/node/4645>

EU imports of crude oil

by partners 2019 (%)



Shares based on thousand tonnes
Imports from not specified countries excluded
Source : Eurostat

EU imports of natural gas

by partners 2019 (%)

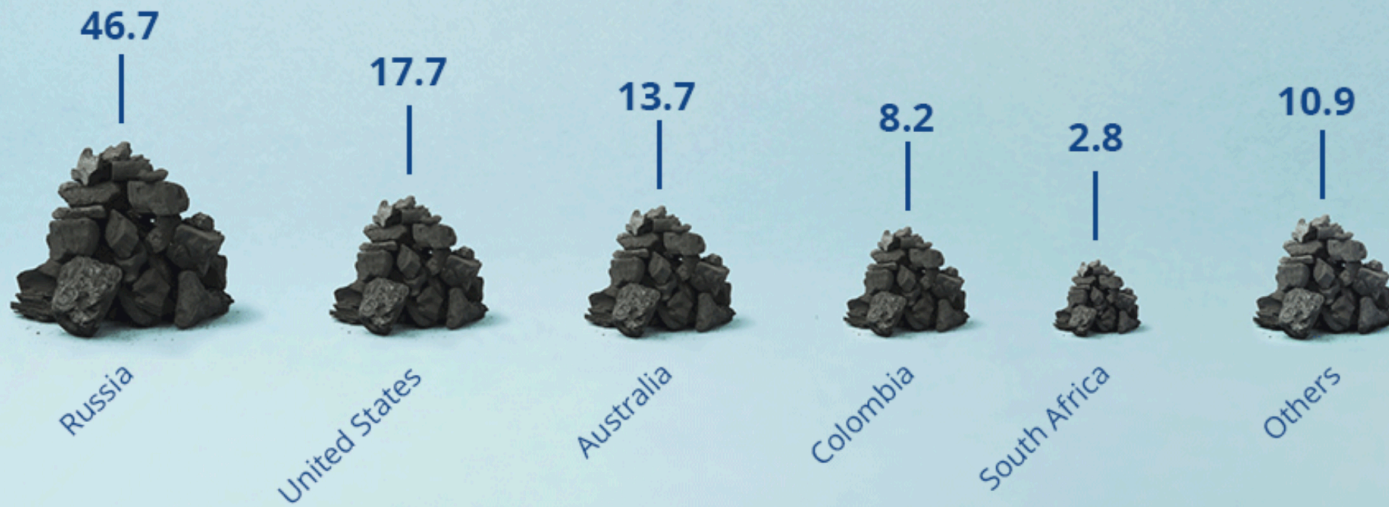


Shares based on million cubic metres
Imports from not specified countries excluded
Source : Eurostat

From:
<https://twitter.com/haschkat/status/1499754859175854088>

EU imports of solid fuel

by partners 2019 (%)

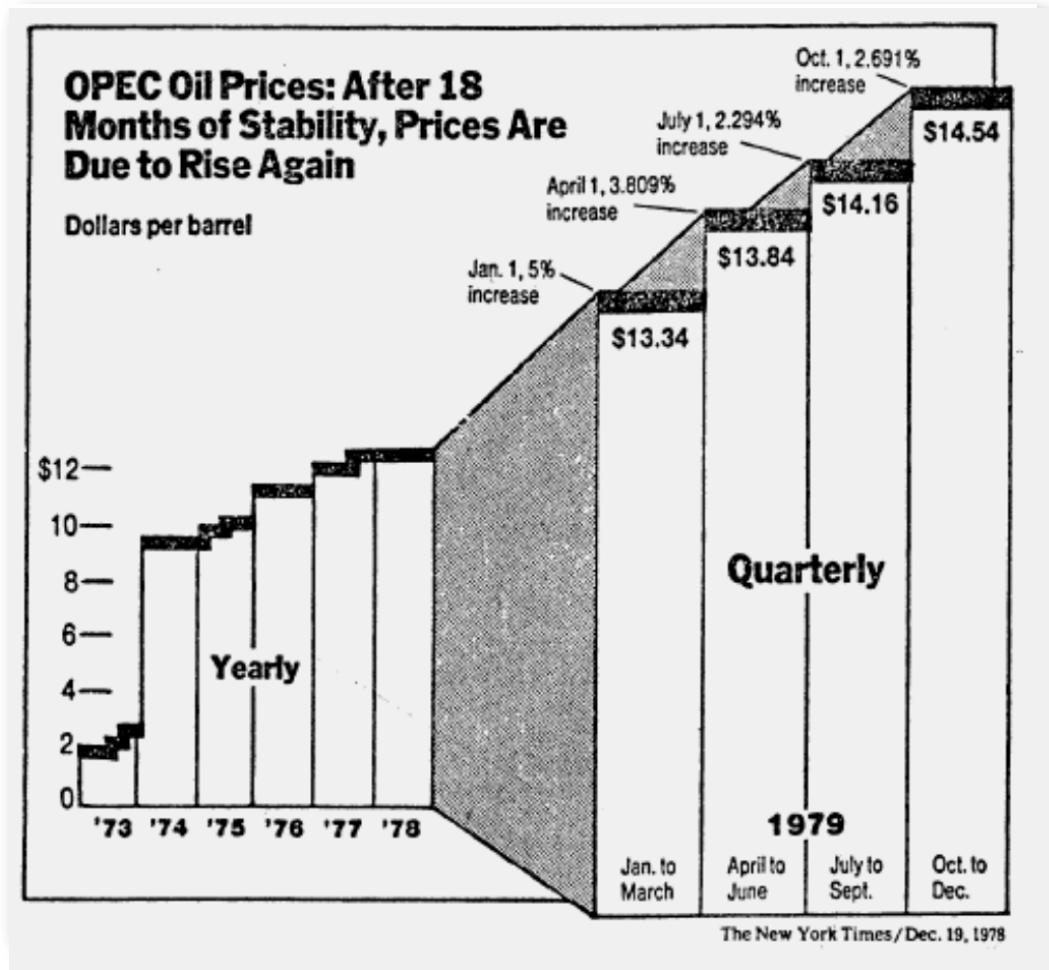


Shares based on thousand tonnes
Imports from not specified countries excluded
Source : Eurostat

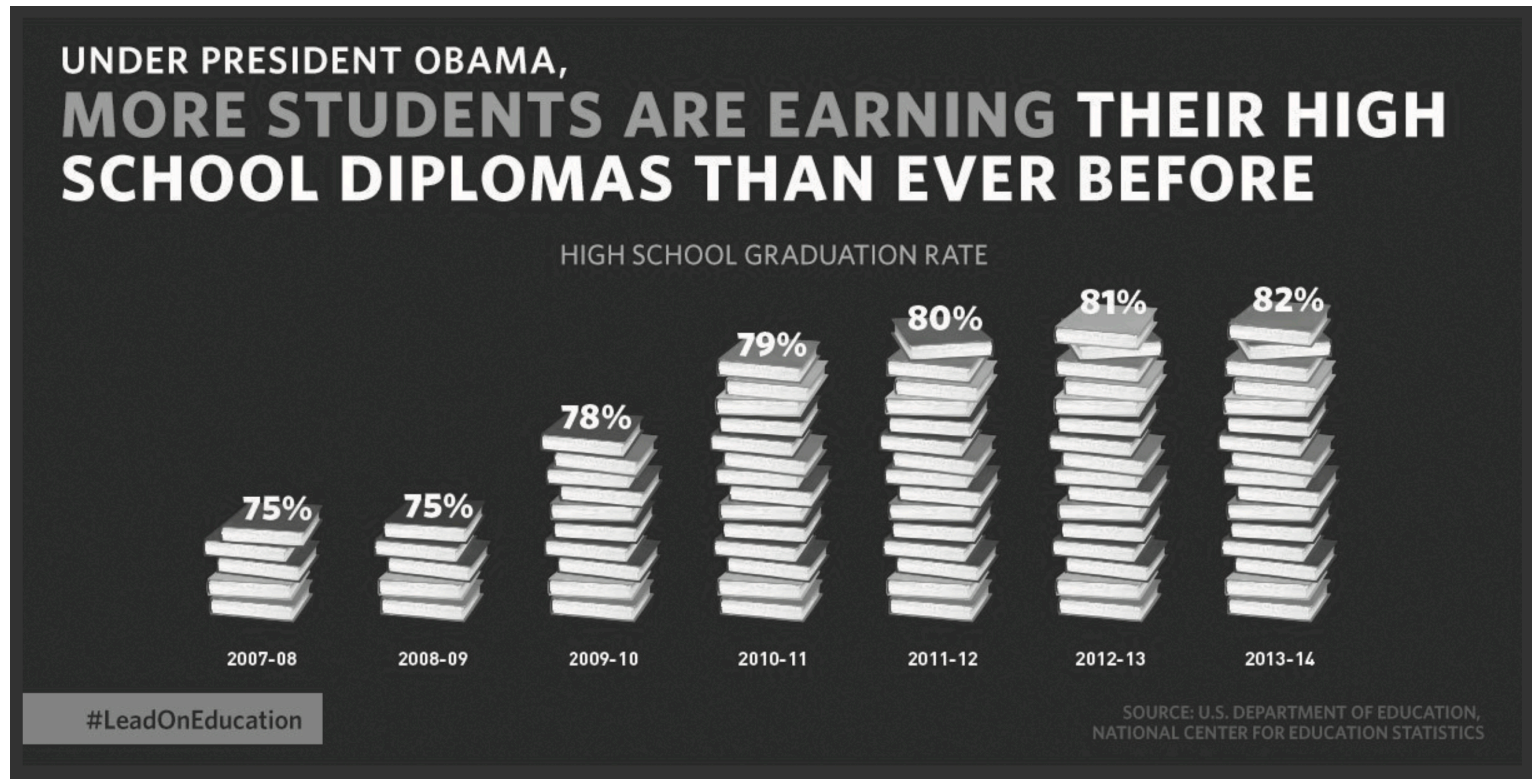
From: <https://twitter.com/haschkat/status/1499754859175854088>

Show data variation, not design variation

Design Distortions

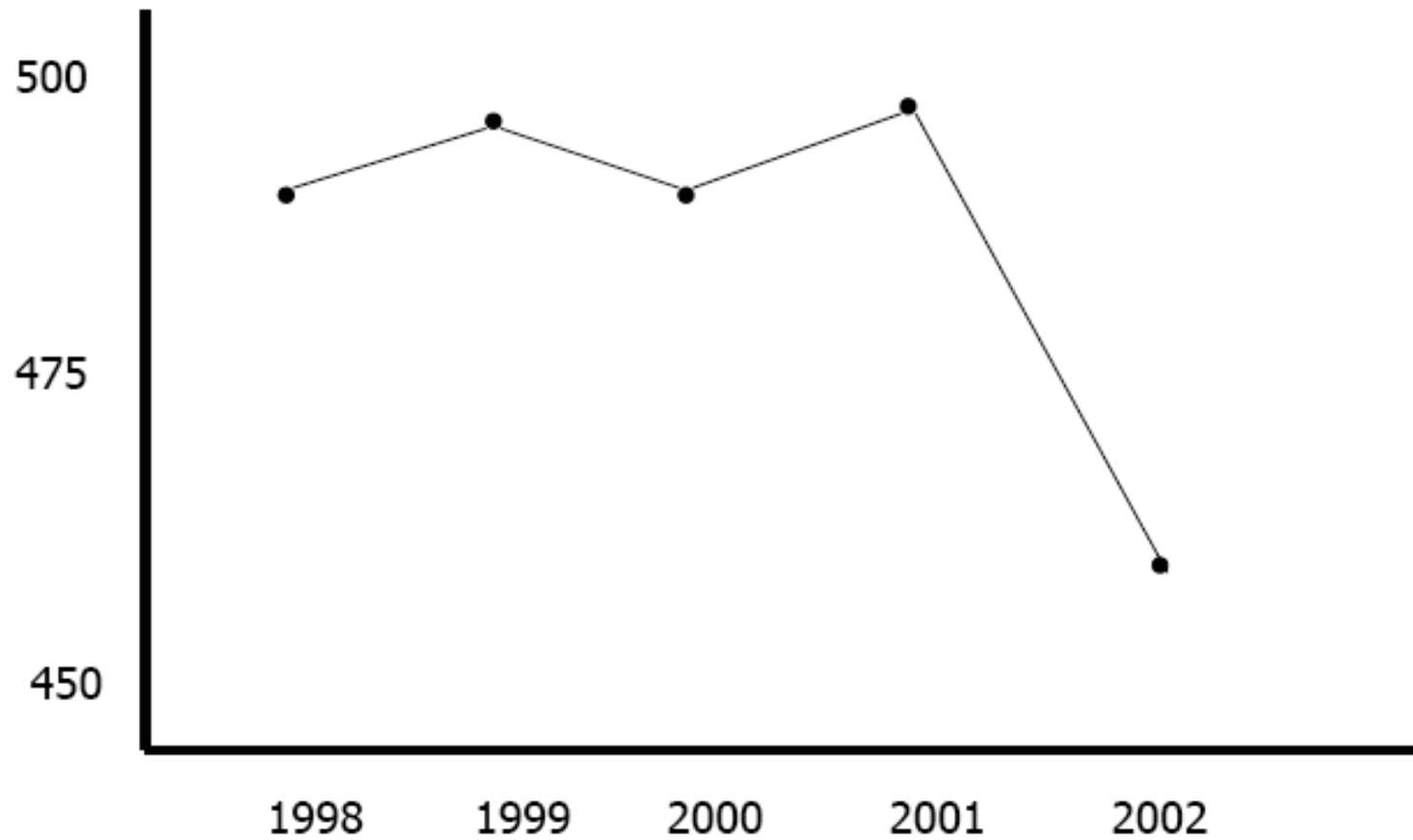


Design Distortions

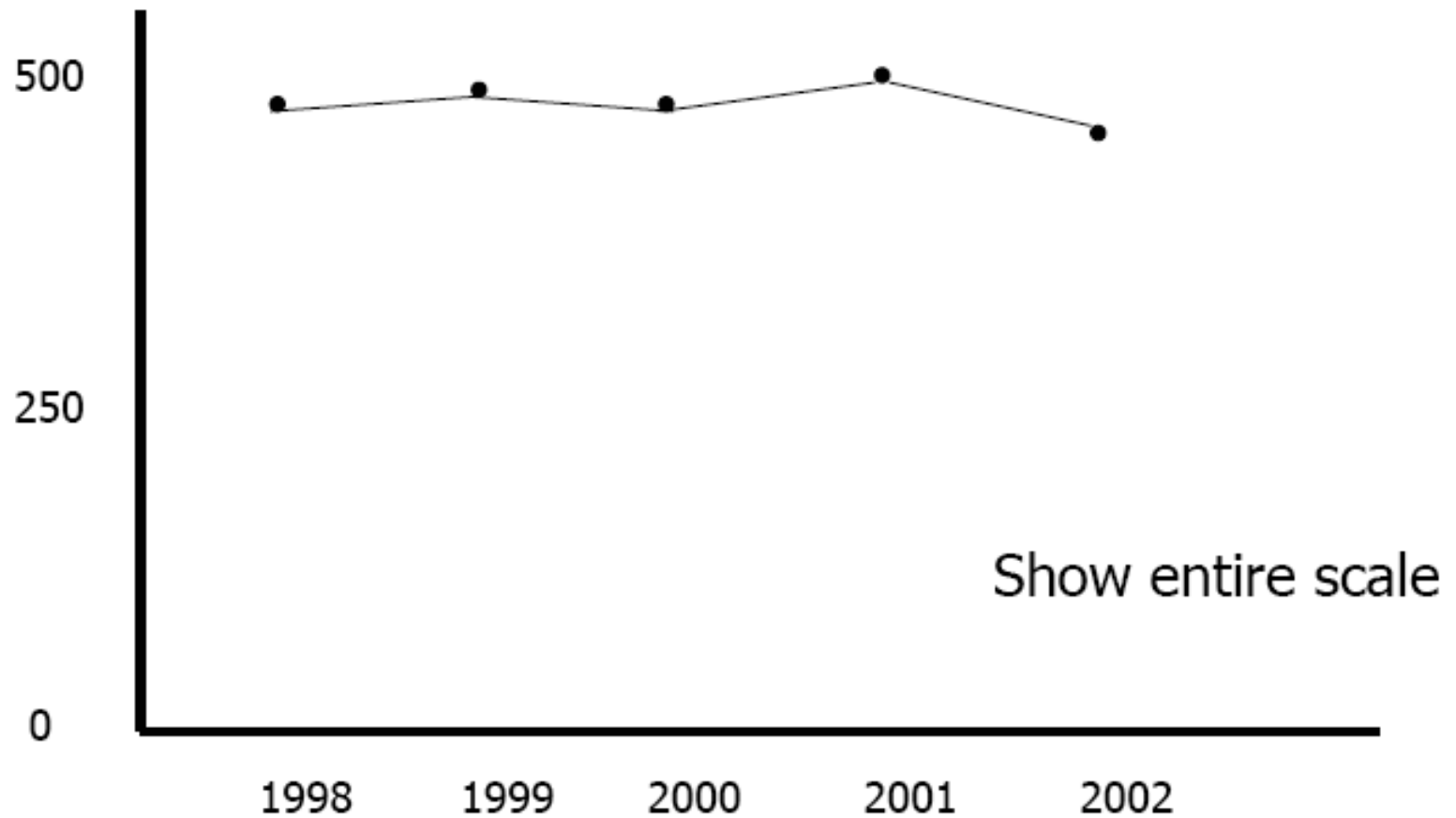


Alberto Cairo, HOW CHARTS LIE, 2019

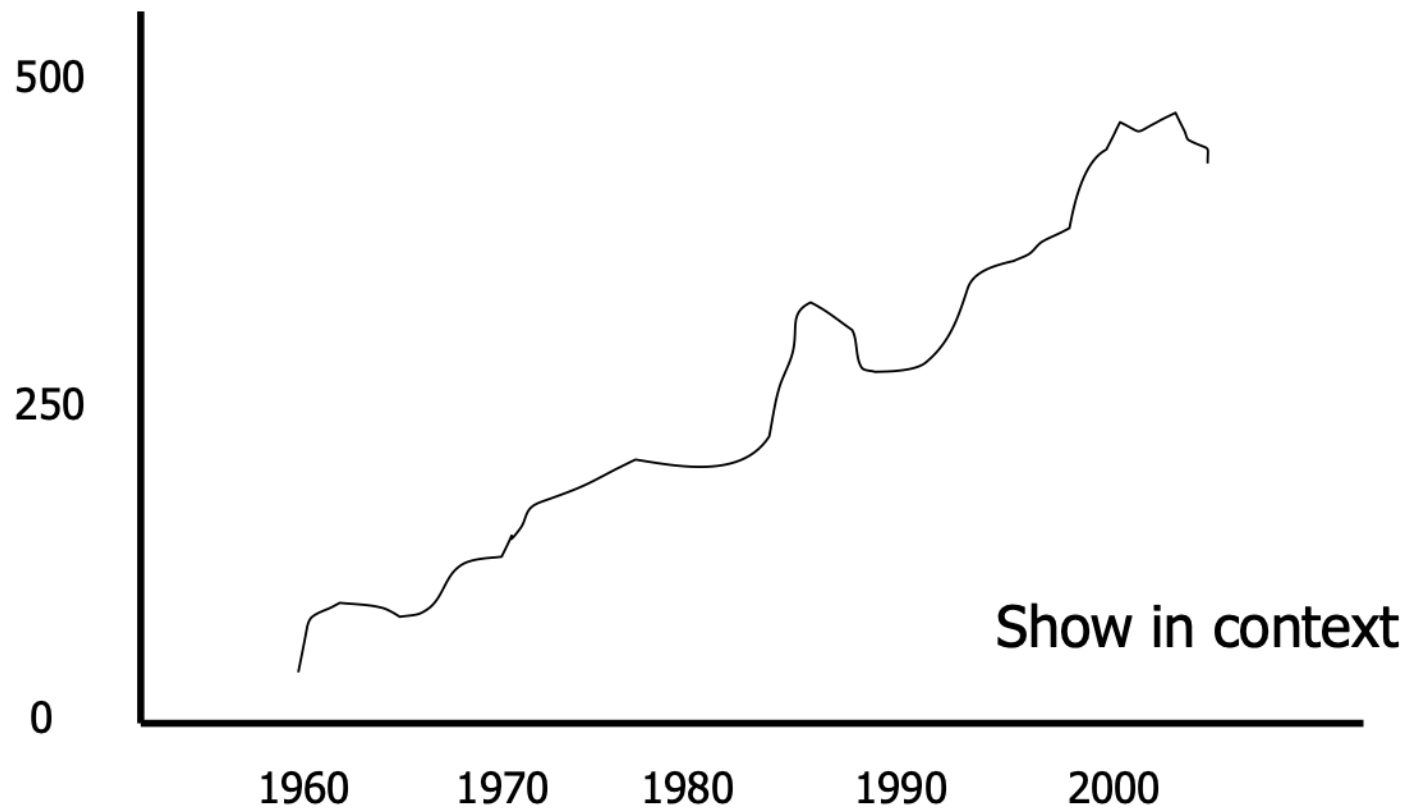
Scale Distortions



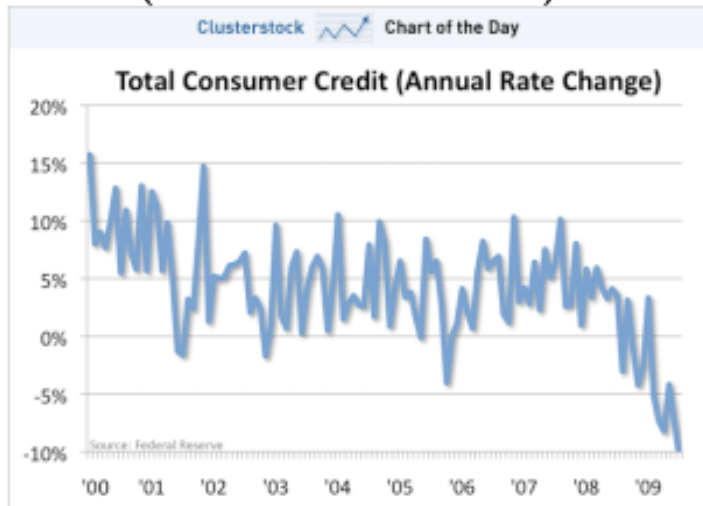
Scale Distortions



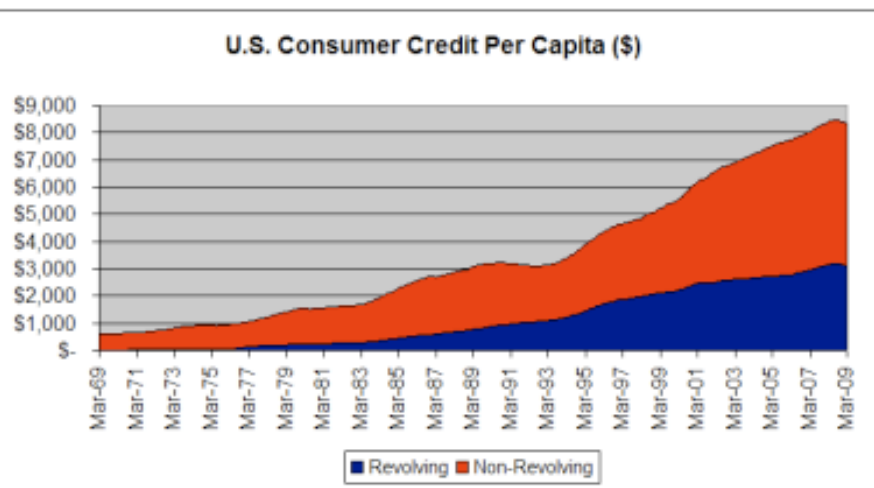
Scale Distortions



“the latest data shows total consumer credit collapsing at an accelerating rate”
(businessinsider.com)



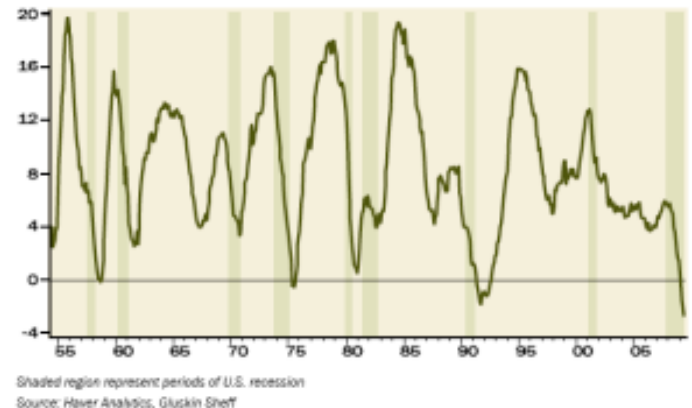
“Per capita is better than total”
(chartingtheeconomy.com)



“have to compare credit to something like disposable income.” (Rolfe Winkler, reuters blog)

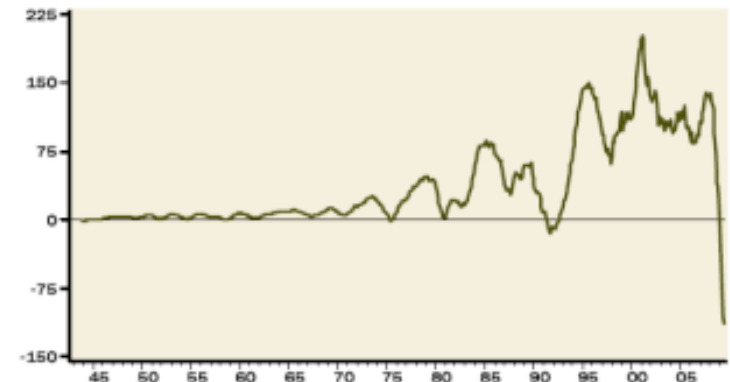
CHART 1: STEEPEST CONSUMER DELEVERAGING IN OVER 50 YEARS

United States: Consumer Credit Outstanding
(year-over-year percent change)



“on-going credit contraction– it’s astonishing”
(ritholtz.com – The Big Picture)

United States: Consumer Credit Outstanding
(year-over-year difference, US\$ bins)





Total Consumer Credit (Annual Rate Change)

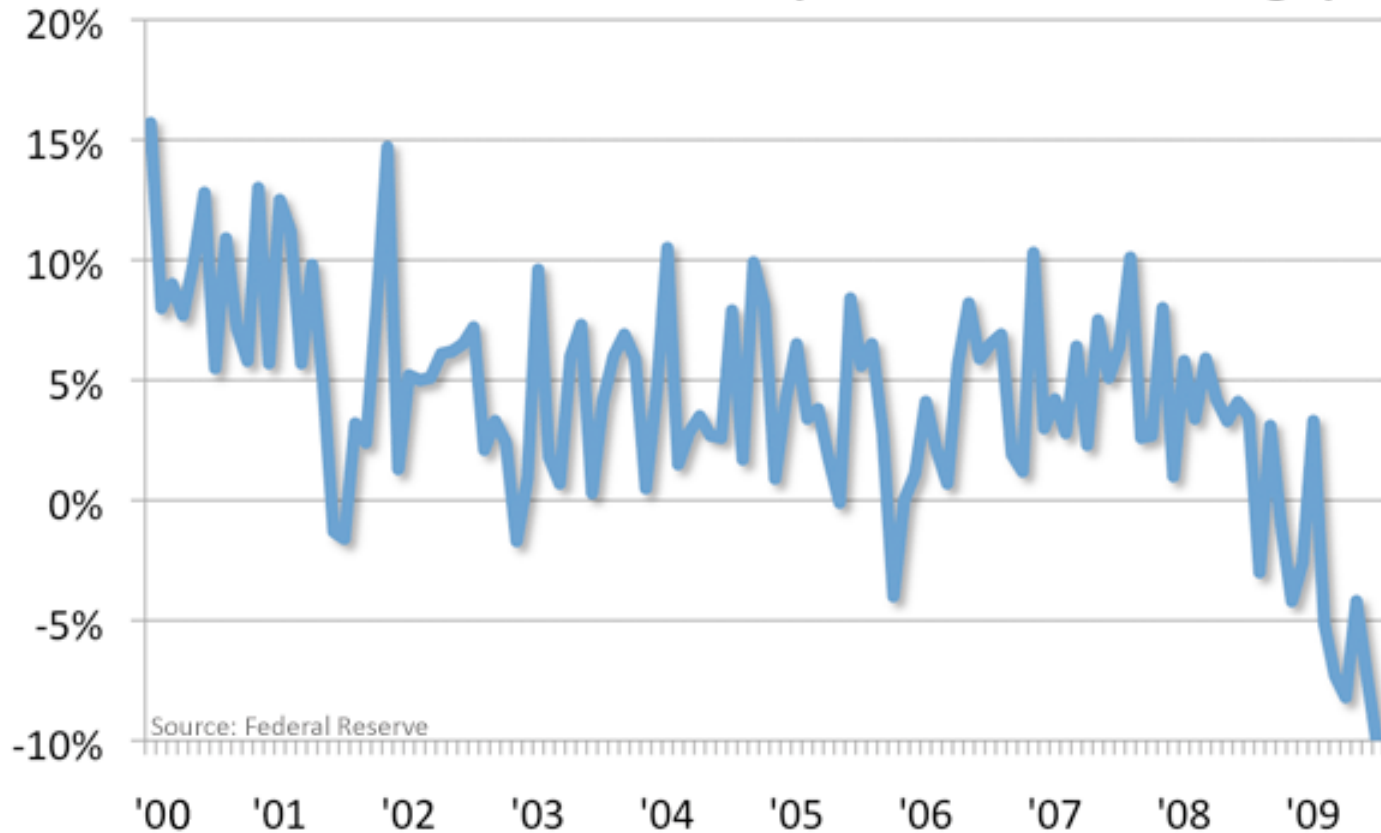
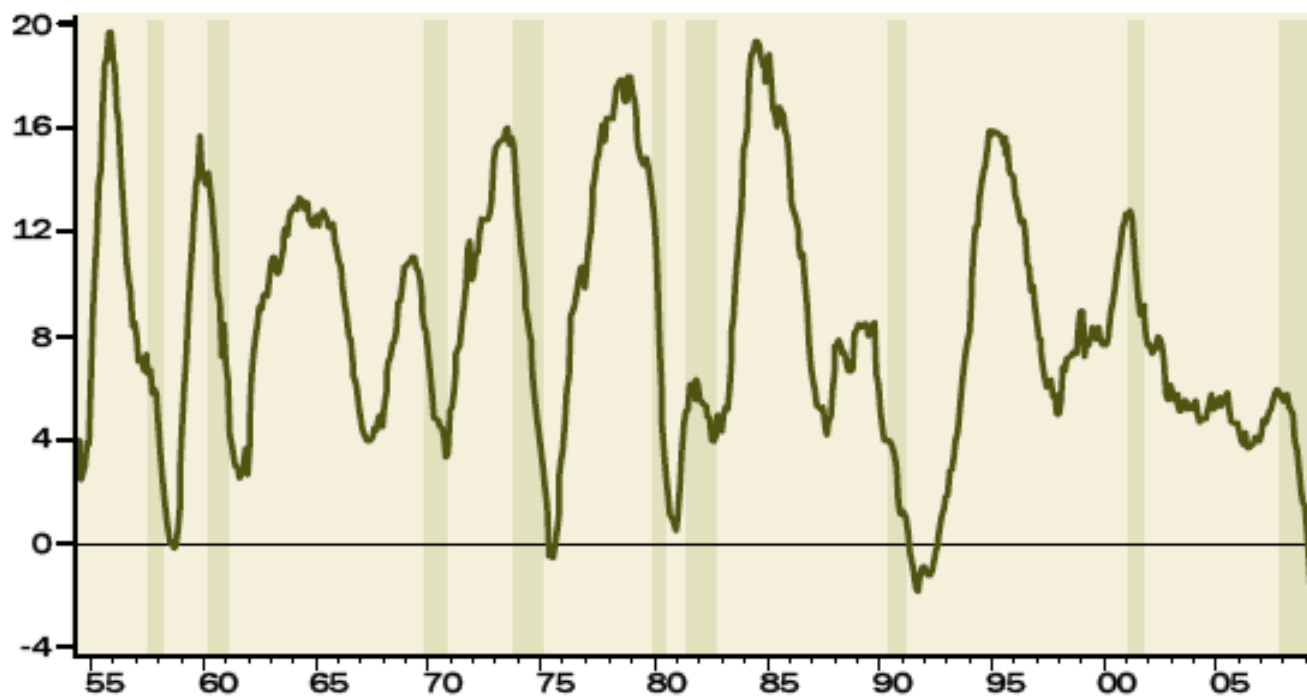


CHART 1: STEEPEST CONSUMER DELEVERAGING IN OVER 50 YEARS

United States: Consumer Credit Outstanding

(year-over-year percent change)

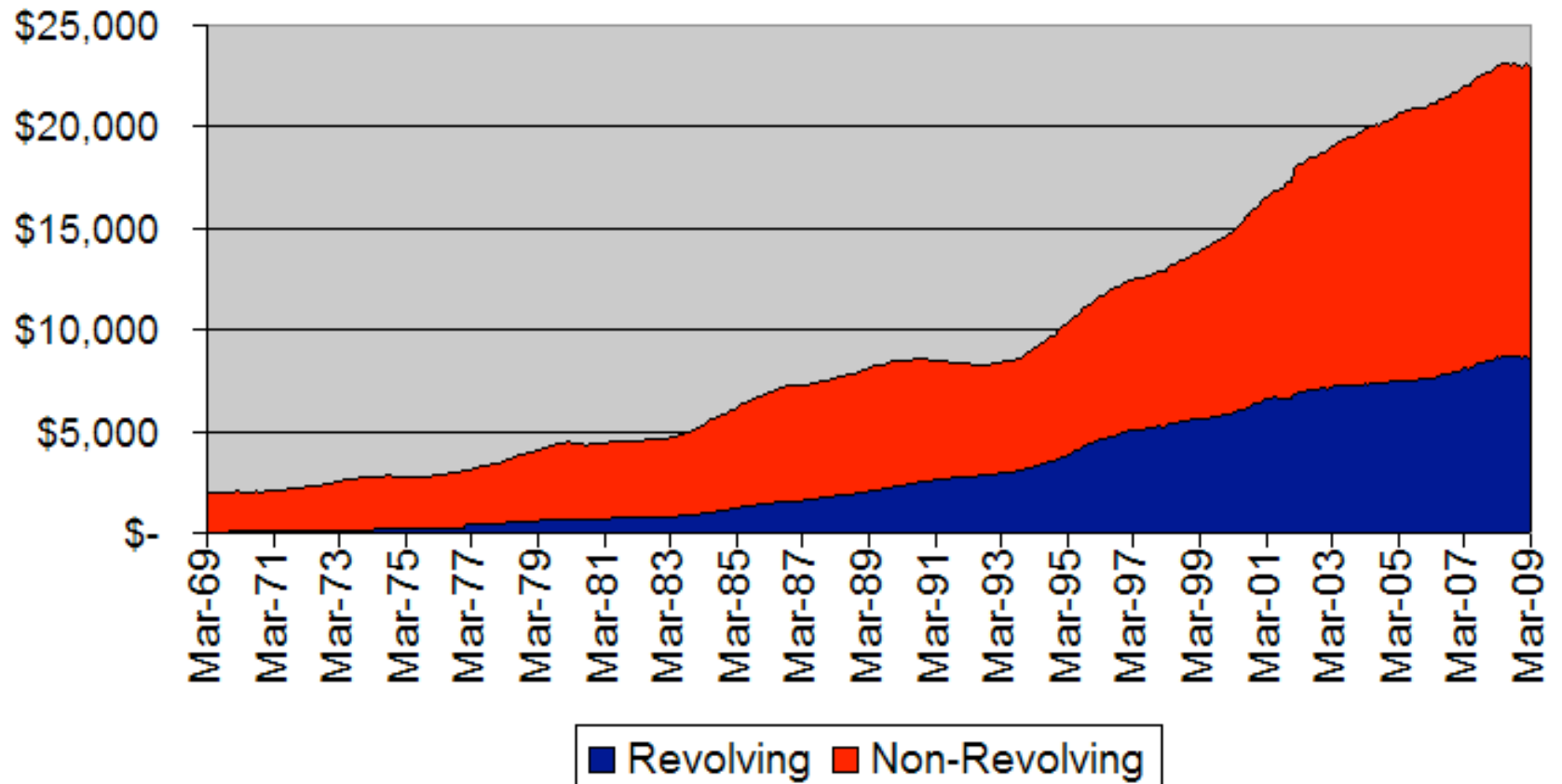


Shaded region represent periods of U.S. recession

Source: Haver Analytics, Gluskin Sheff

[Rolfe Winkler](#)

U.S. Consumer Credit Per Household (\$)



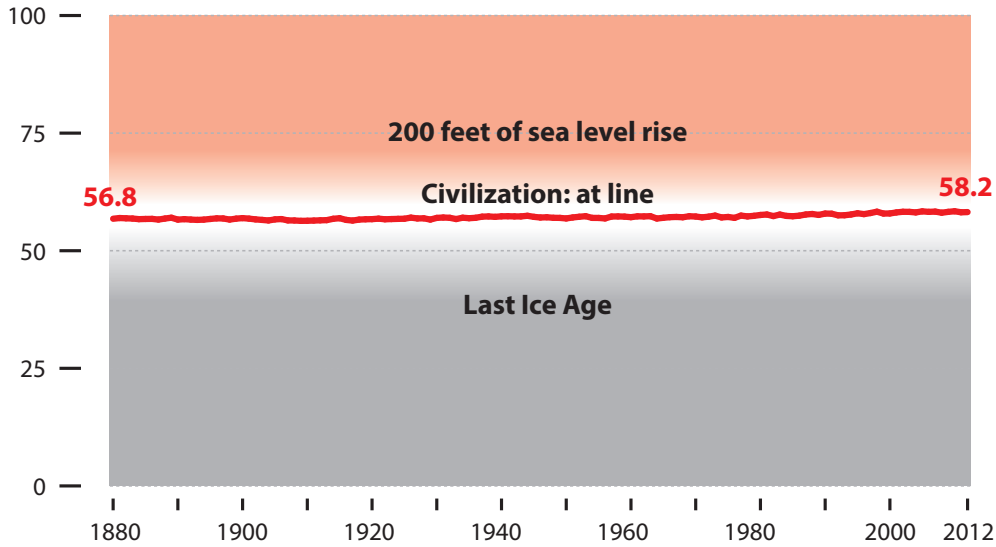
United States: Consumer Credit Outstanding

(year-over-year difference, US\$ blns)



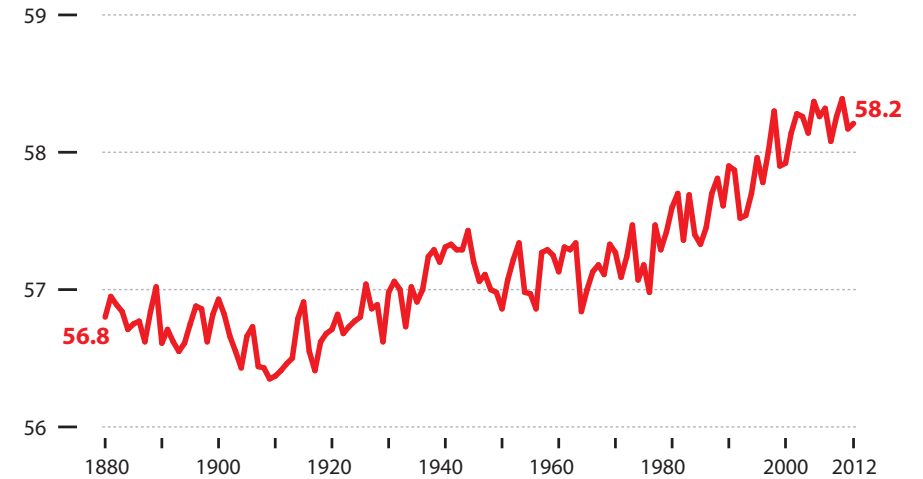
[The Big Picture](#)

Average annual global temperature in degrees Fahrenheit



0 °F = -17.8 °C
100 °F = 37.8 °C

Average annual global temperature in degrees Fahrenheit



Alberto Cairo, HOW CHARTS LIE, 2019

Summary: Graphical Integrity

- The **representation of numbers**, as physically measured on the surface of the graph itself, should be **directly proportional** to the numerical quantities represented.
- Clear, detailed and thorough **labelling** should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graph itself. Label important events in the data.
- Show **data variation**, not design variation
- In time-series displays of money, deflated and **standardized units of monetary measurement** are nearly always better than nominal units.
- The **number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data**
- Graphics must not quote data out of **context**

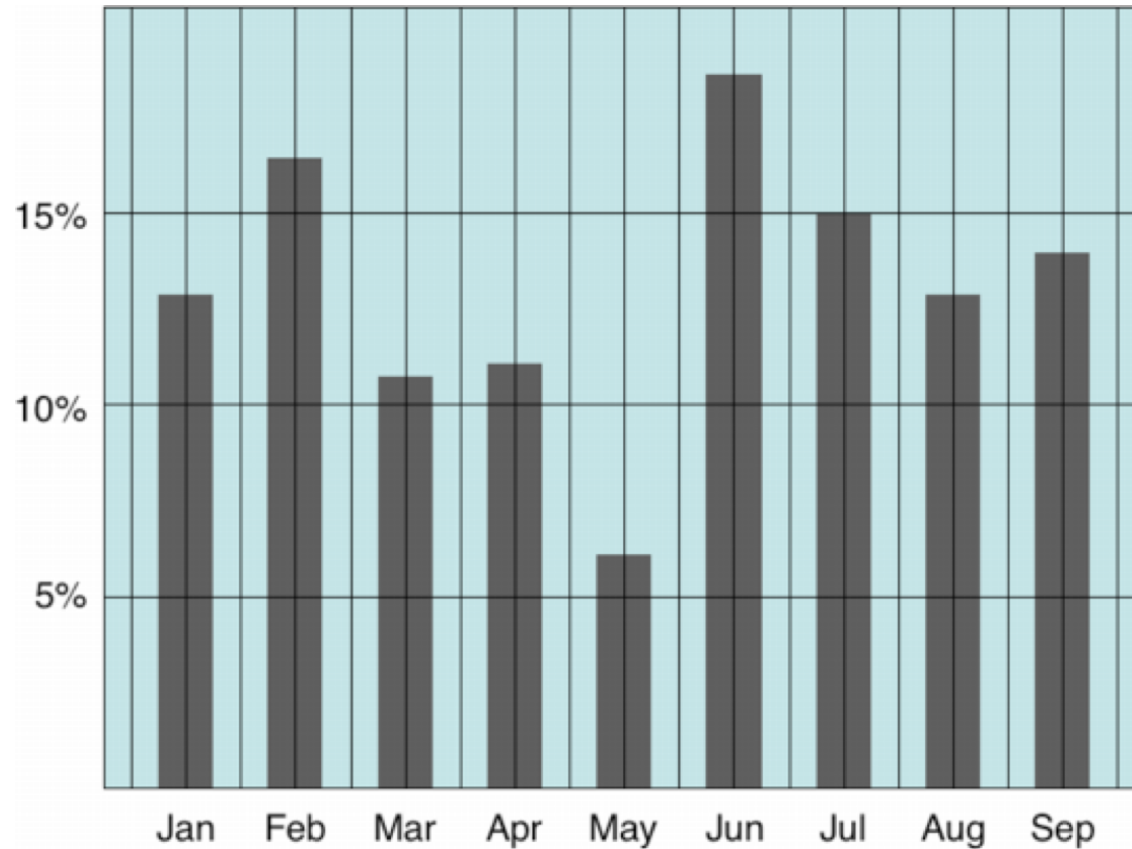
Tufte, VDQI

A top-down view of a wooden desk with various drawing tools. In the center is a spiral-bound notebook with a blank, light-colored page. To the left of the notebook are two pencils, one red and one yellow. To the right are two more pencils, one blue and one brown. Above the notebook is a black ruler and a compass. Below the notebook is a black pen. To the right of the notebook is a small glass containing a golden liquid. The background is a light-colored wooden surface with a visible grain.

Design Principles

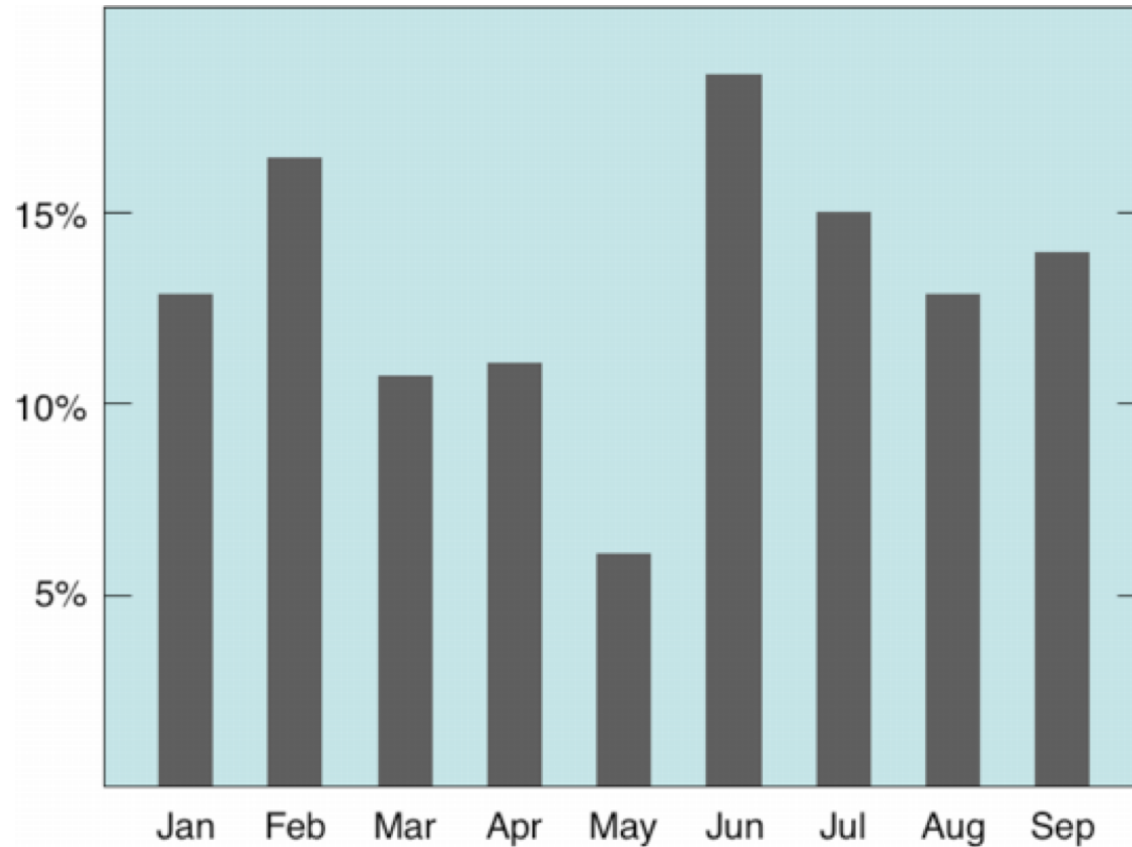
Avoid Chartjunk

Chartjunk: Extraneous visual elements that distract from the message



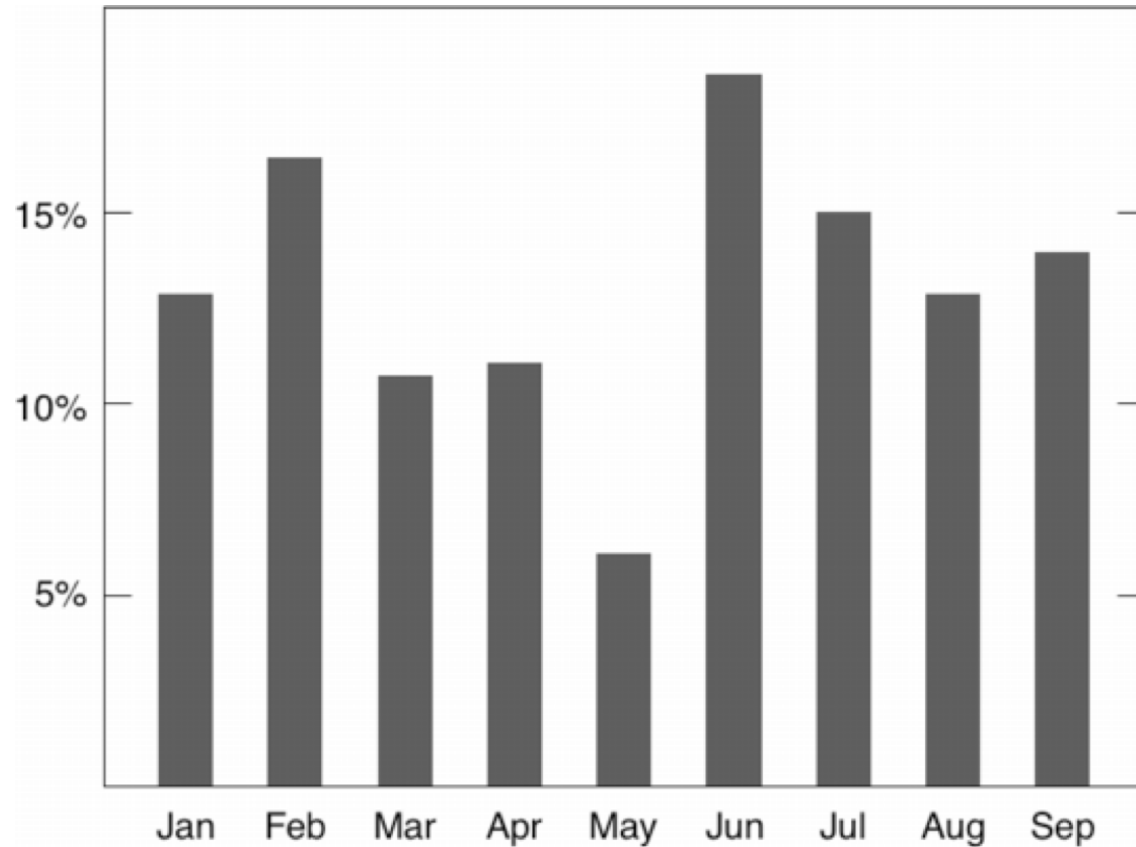
Avoid Chartjunk

Chartjunk: Extraneous visual elements that distract from the message



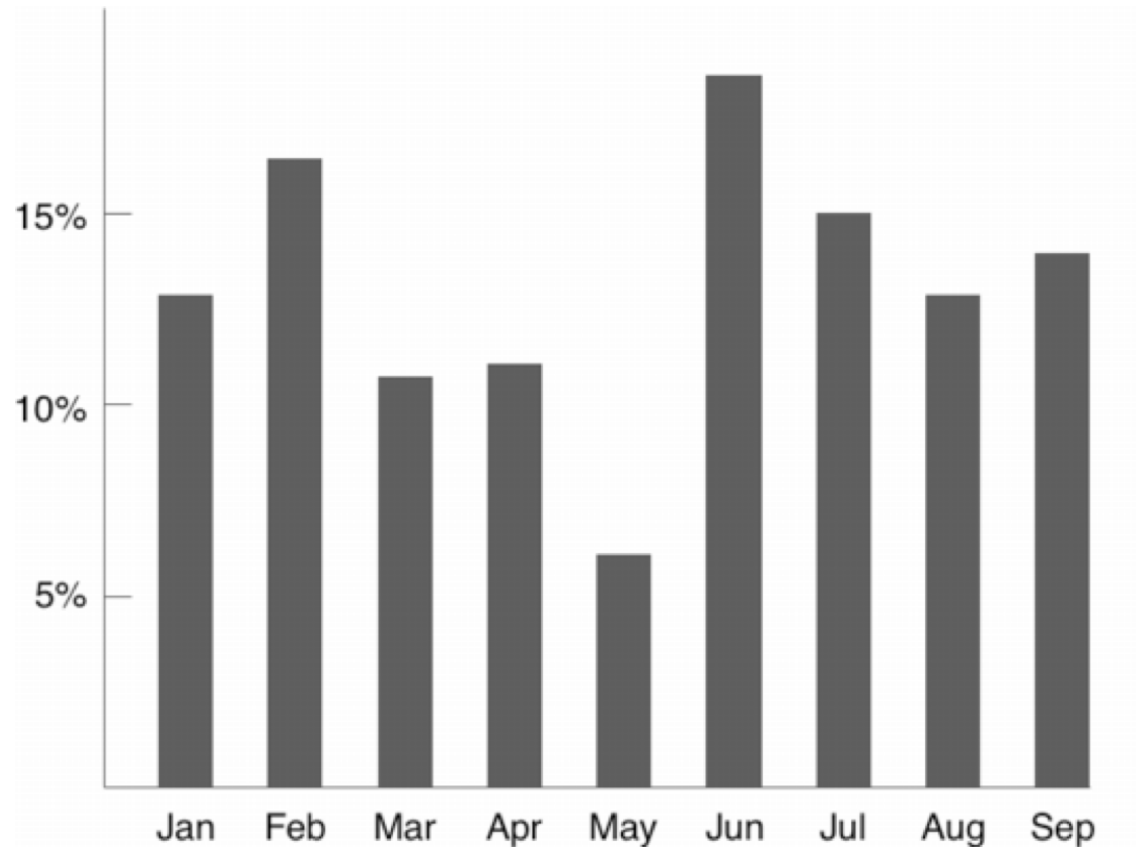
Avoid Chartjunk

Chartjunk: Extraneous visual elements that distract from the message



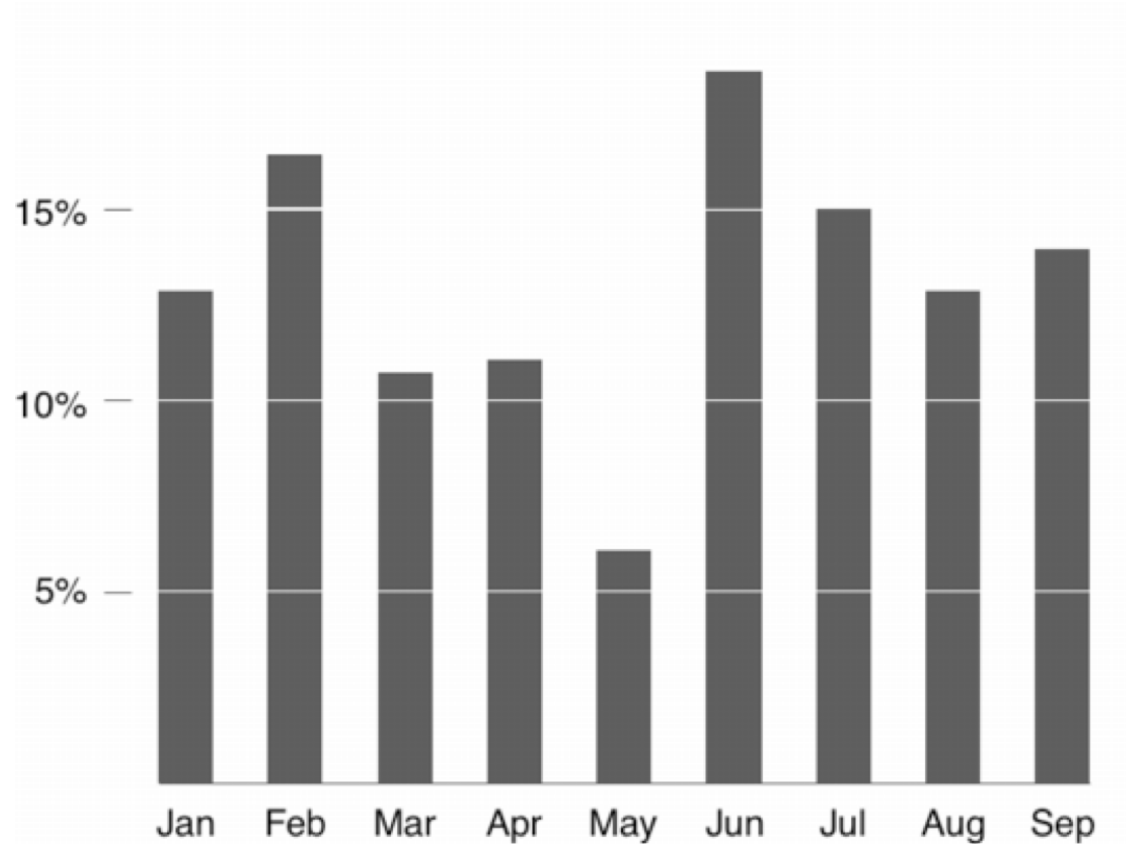
Avoid Chartjunk

Chartjunk: Extraneous visual elements that distract from the message



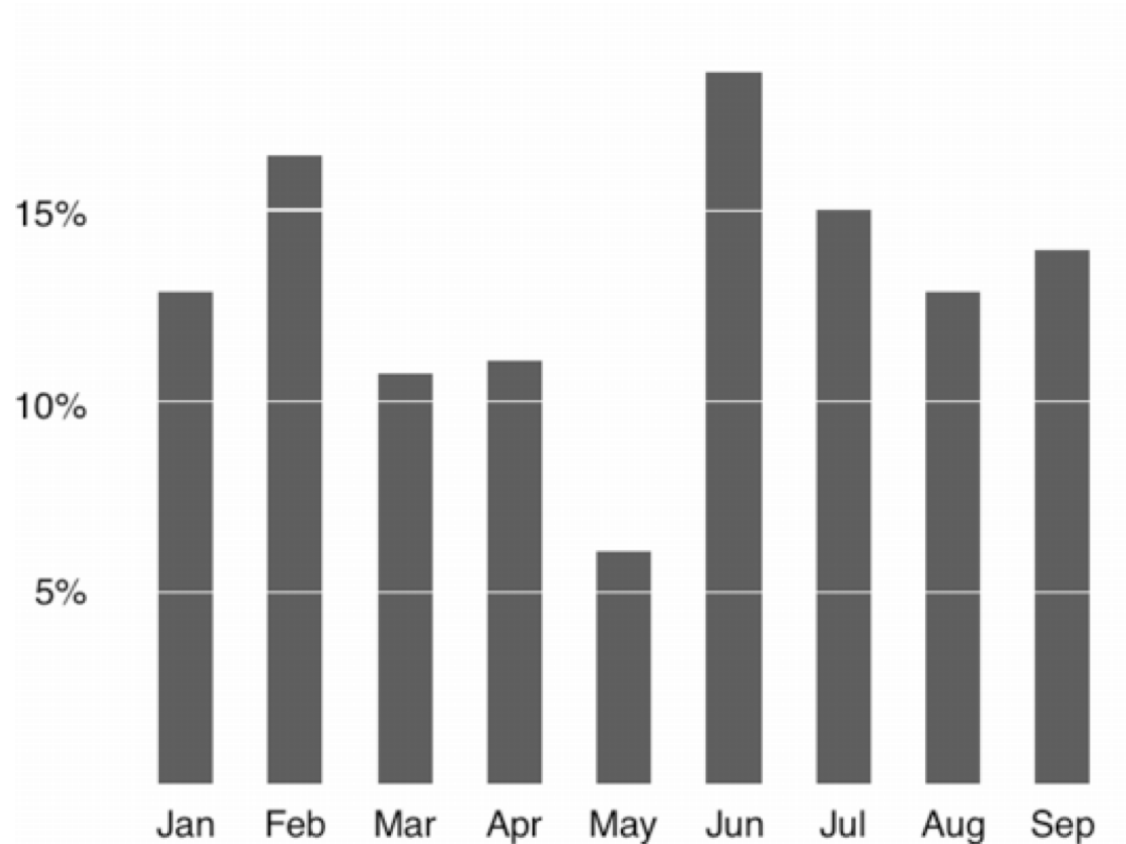
Avoid Chartjunk

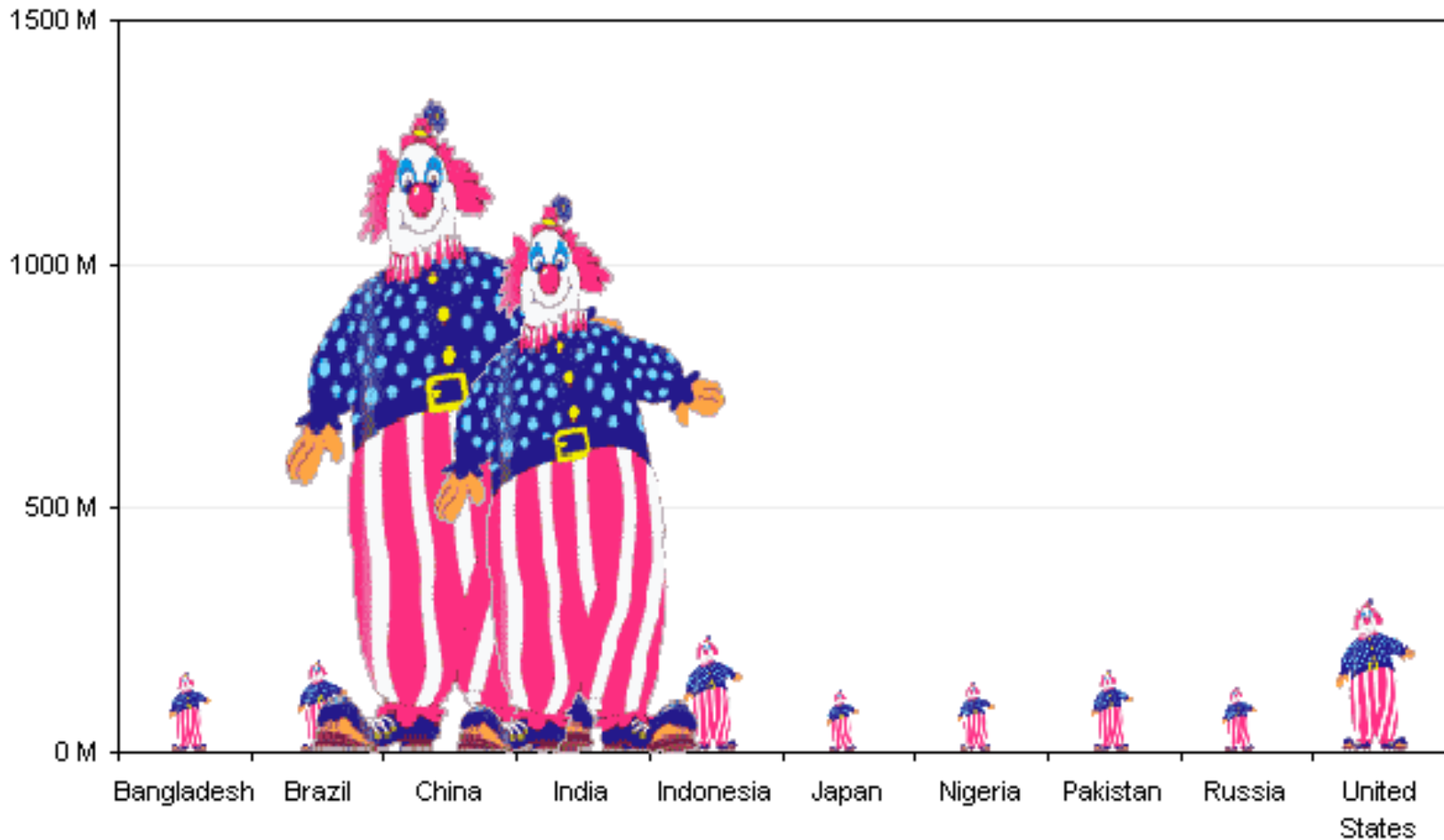
Chartjunk: Extraneous visual elements that distract from the message



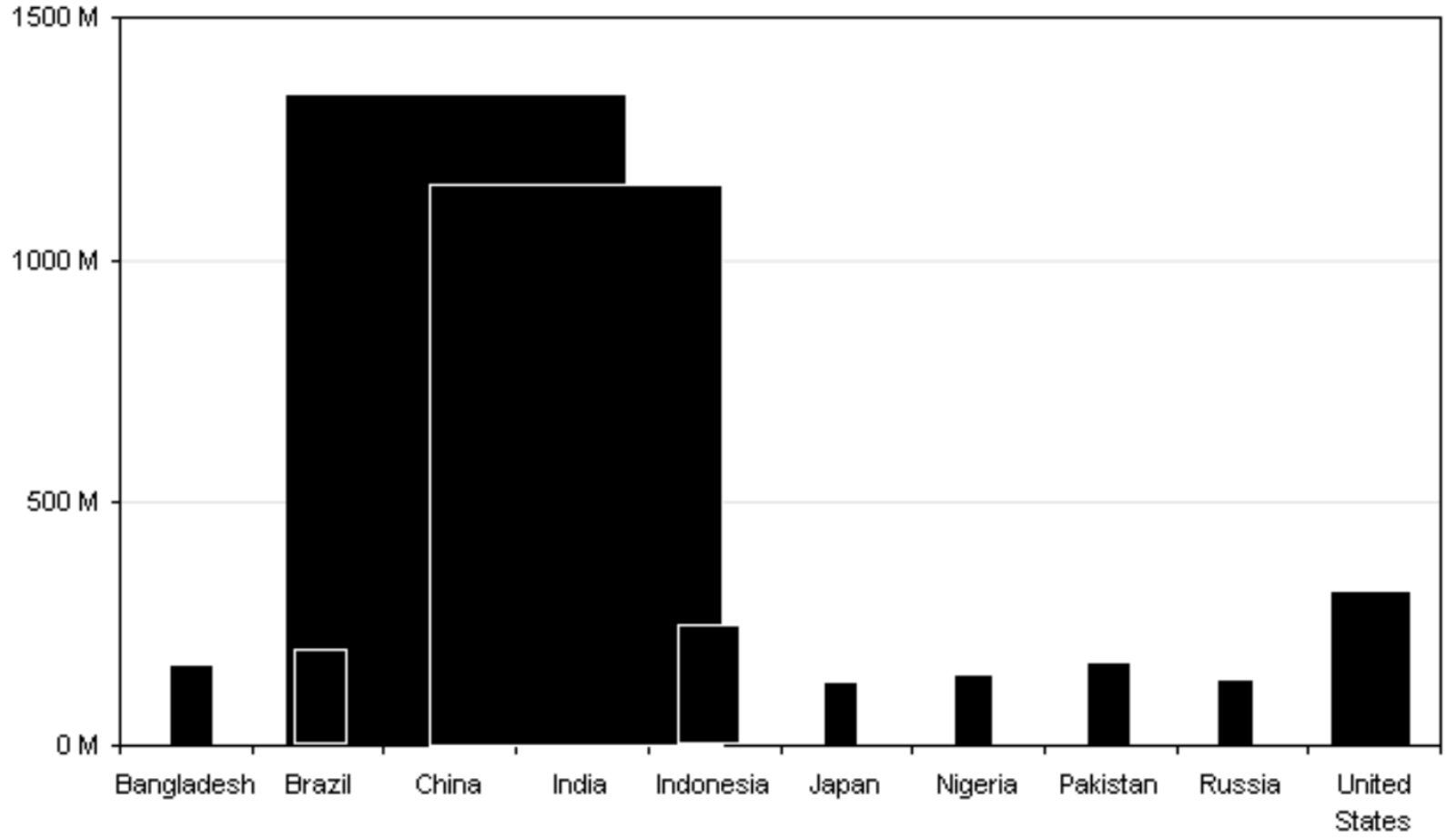
Avoid Chartjunk

Chartjunk: Extraneous visual elements that distract from the message



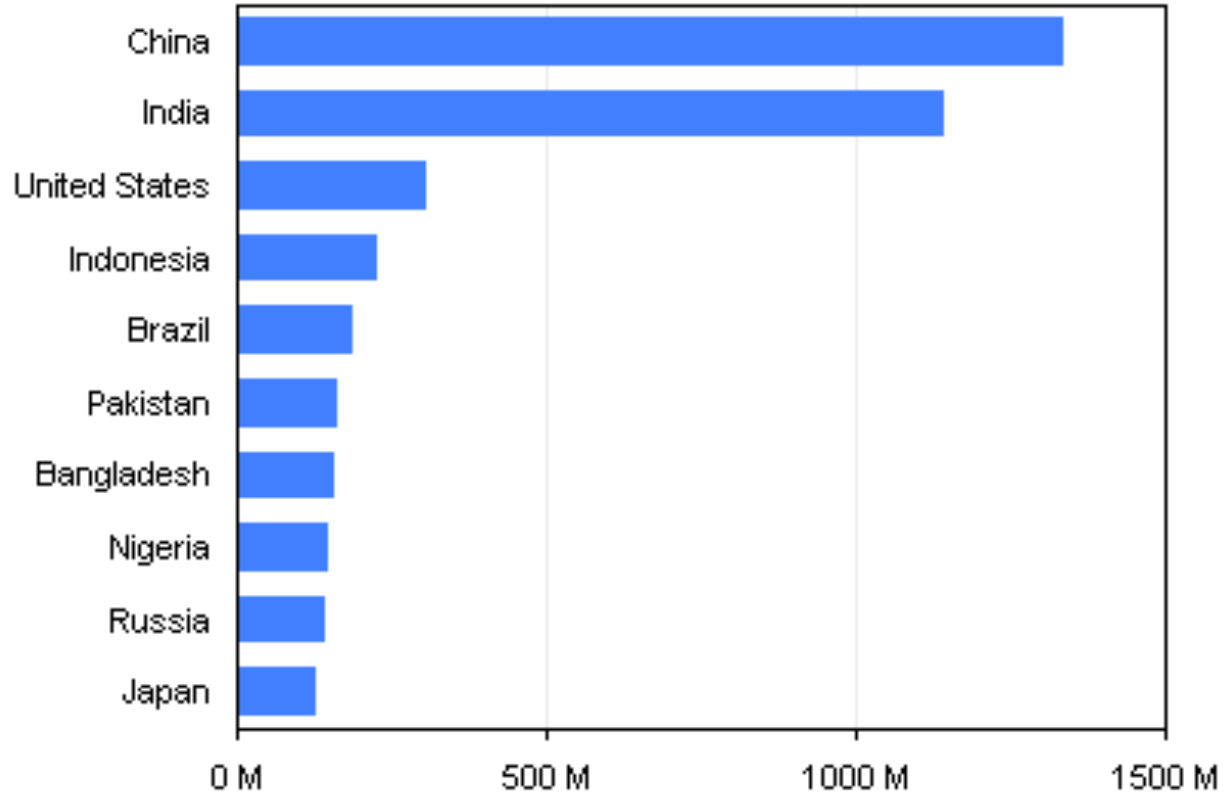


World Population in 2008



PTS Blog: <https://peltiertech.com/bad-bar-chart-practices-or-send-in-the-clowns/>

A better version...



World Population in 2008

Chartjunk

Any ink on your chart that does not convey data is considered junk (Tufte).

Is chart junk always bad?

How about:

Chart junk is any element of a chart that does not contribute to clarifying the intended message.

Stephen Few for more context:

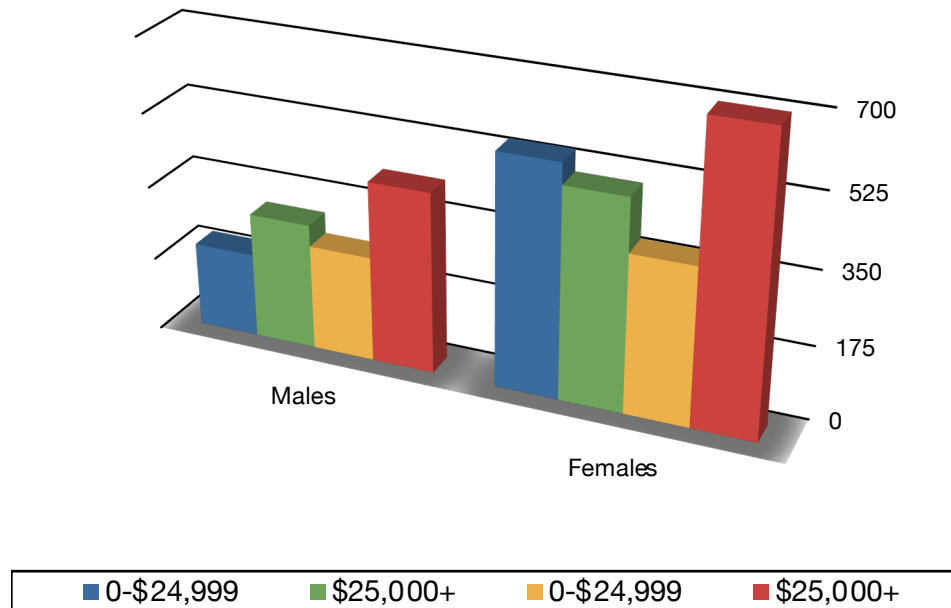
http://www.perceptualedge.com/articles/visual_business_intelligence/sometimes_we_must_raise_our_voices.pdf

Is chartjunk always bad?

- Memorability
- Engagement
- Support cognitive processes through repetition?

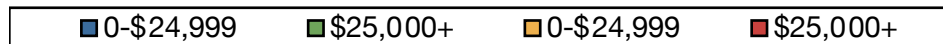
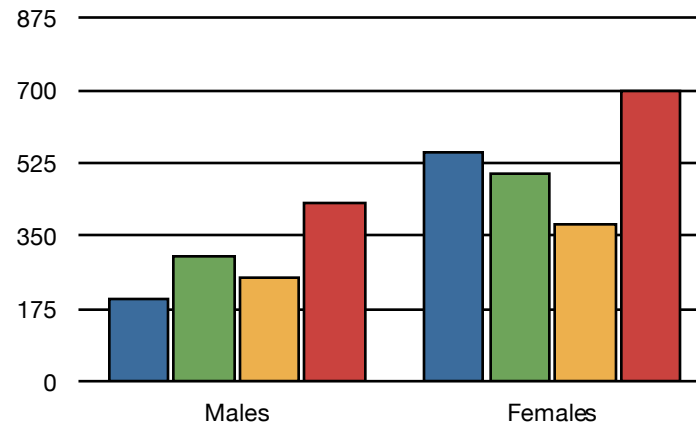
Maximize Data-Ink Ratio

- Data-ink = the ink used to show data
- Data-ink ratio = data-ink / total ink used



Maximize Data-Ink Ratio

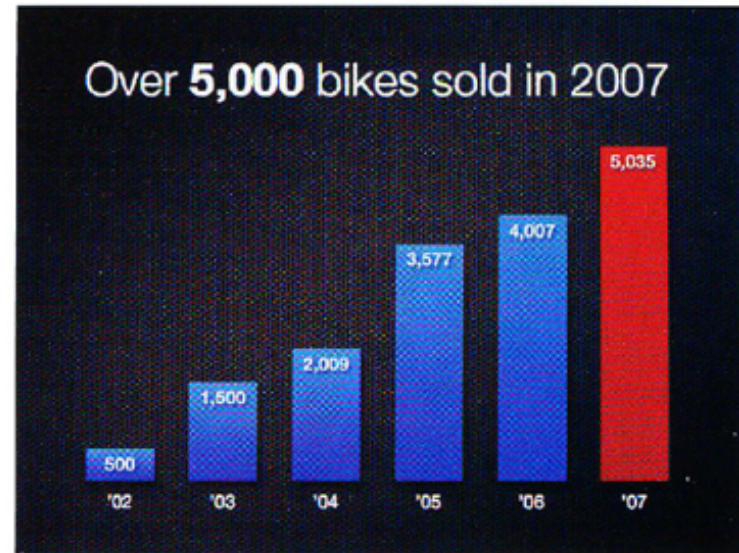
- Data-ink = the ink used to show data
- Data-ink ratio = data-ink / total ink used



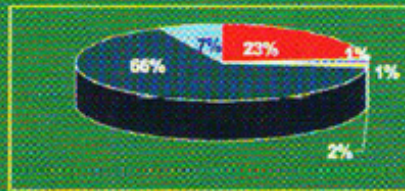
Before



After



Arable land in organic production

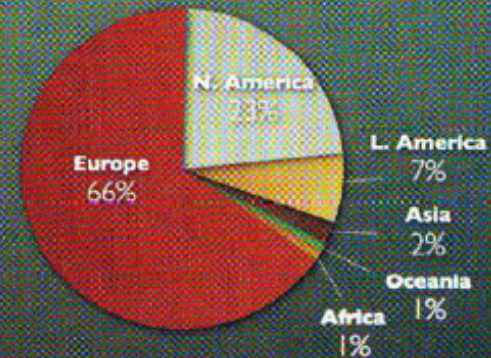


- N. America
- Oceania
- Africa
- Asia
- Europe
- L. America

Source: SOEL-FIBL Survey '07

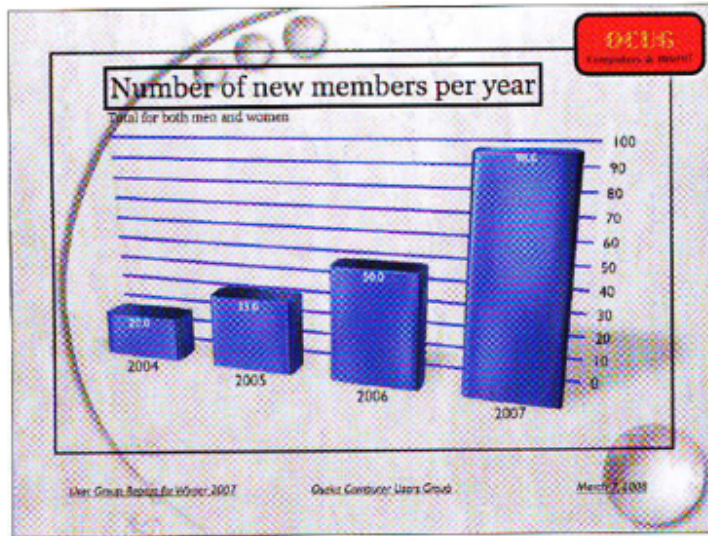


Arable land in organic production

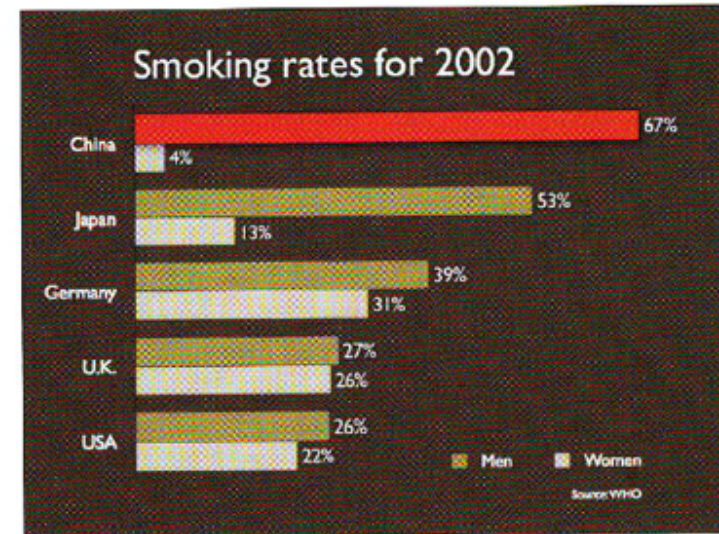
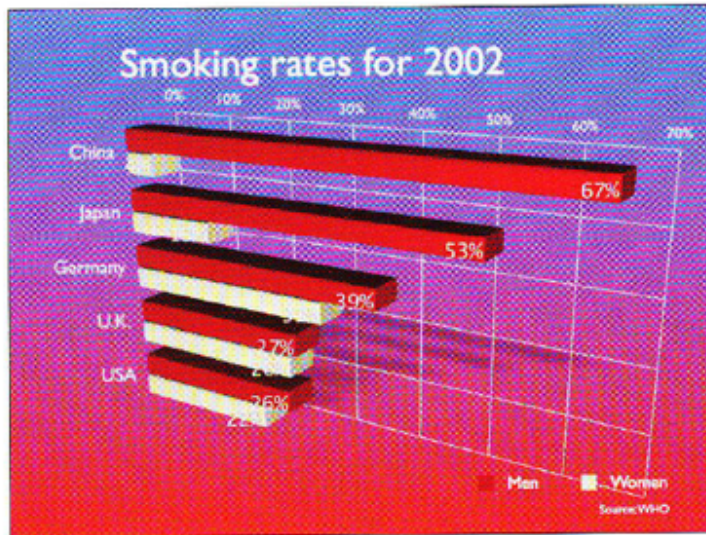
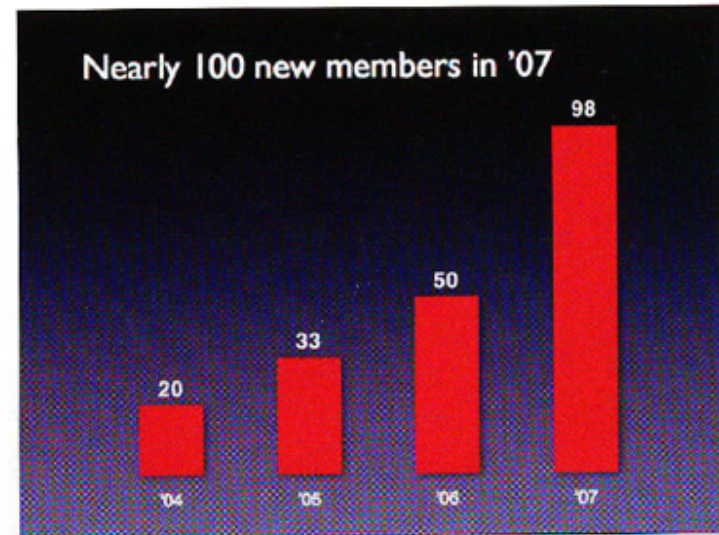


Source: SOEL-FIBL Survey '07

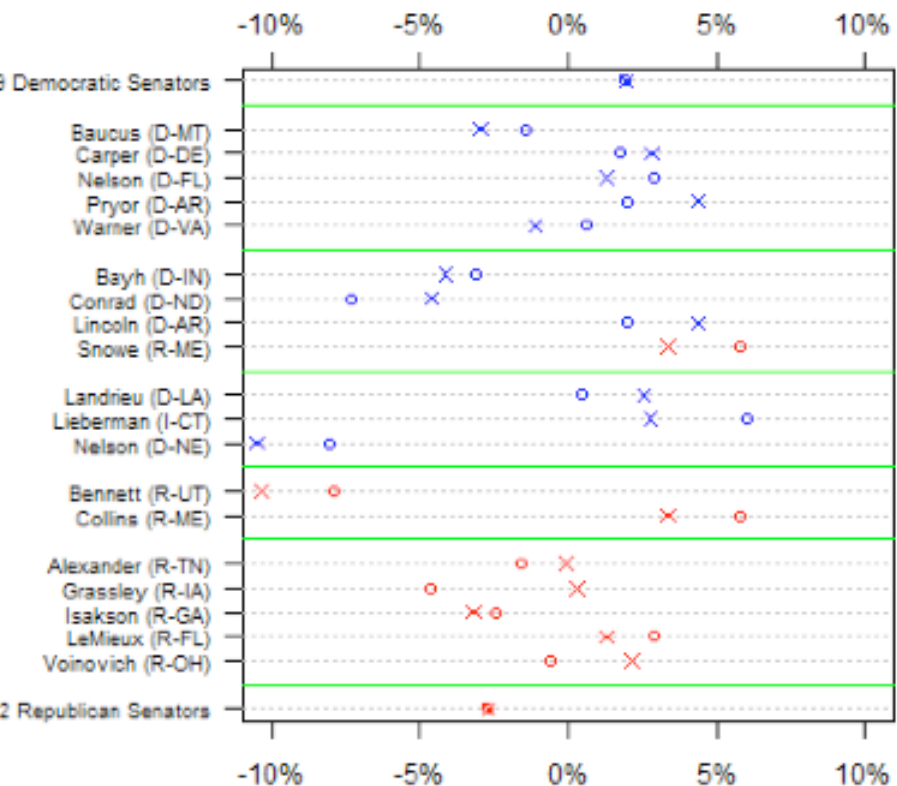
Before



After



Senators and State support for health care



o: 2000 avg. is 67%
 x: 2004 avg. is 73%

SENATORS AND STATE SUPPORT FOR HEALTH CARE

49 Democrats who fully support the bill

SUPPORTS WITH RESERVATIONS

- Baucus MONTANA
- Nelson FLORIDA
- Carper DELAWARE
- Warner VIRGINIA
- Conrad NORTH DAKOTA
- Pryor ARKANSAS

ON THE FENCE

- Bayh INDIANA
- Lincoln ARKANSAS
- Landriau LOUISIANA
- Nelson NEBRASKA

- Lieberman (Independent) CONNECTICUT

- Snowe MAINE

OPPOSES IN PRACTICE

- Collins MAINE
- Bennett UTAH
- Isakson GEORGIA
- Grassley IOWA
- LeMieux FLORIDA
- Alexander TENNESSEE
- Voinovich OHIO

32 Republicans who fully oppose the bill

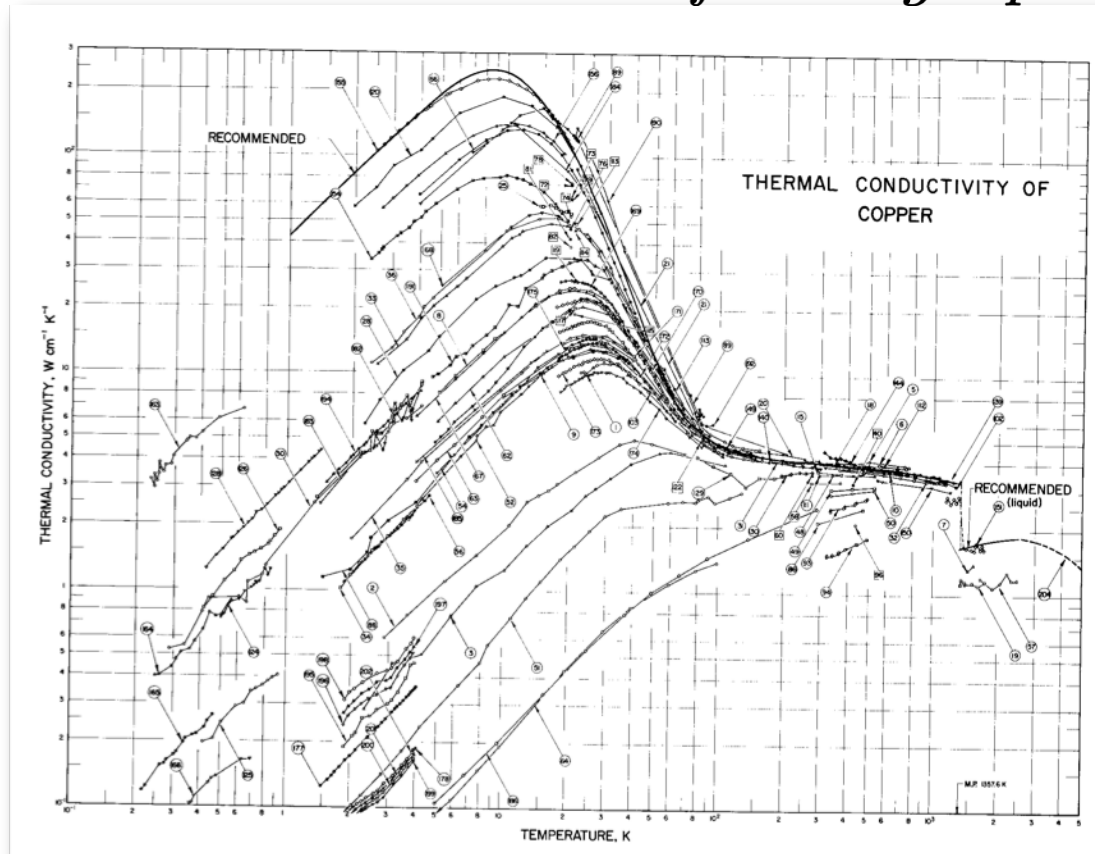


X = percentage of state voters who supported increased federal spending on health care in the 2004 Annenberg survey; percentages are relative to the national average.

Andrew Gelman, Nov. 2009

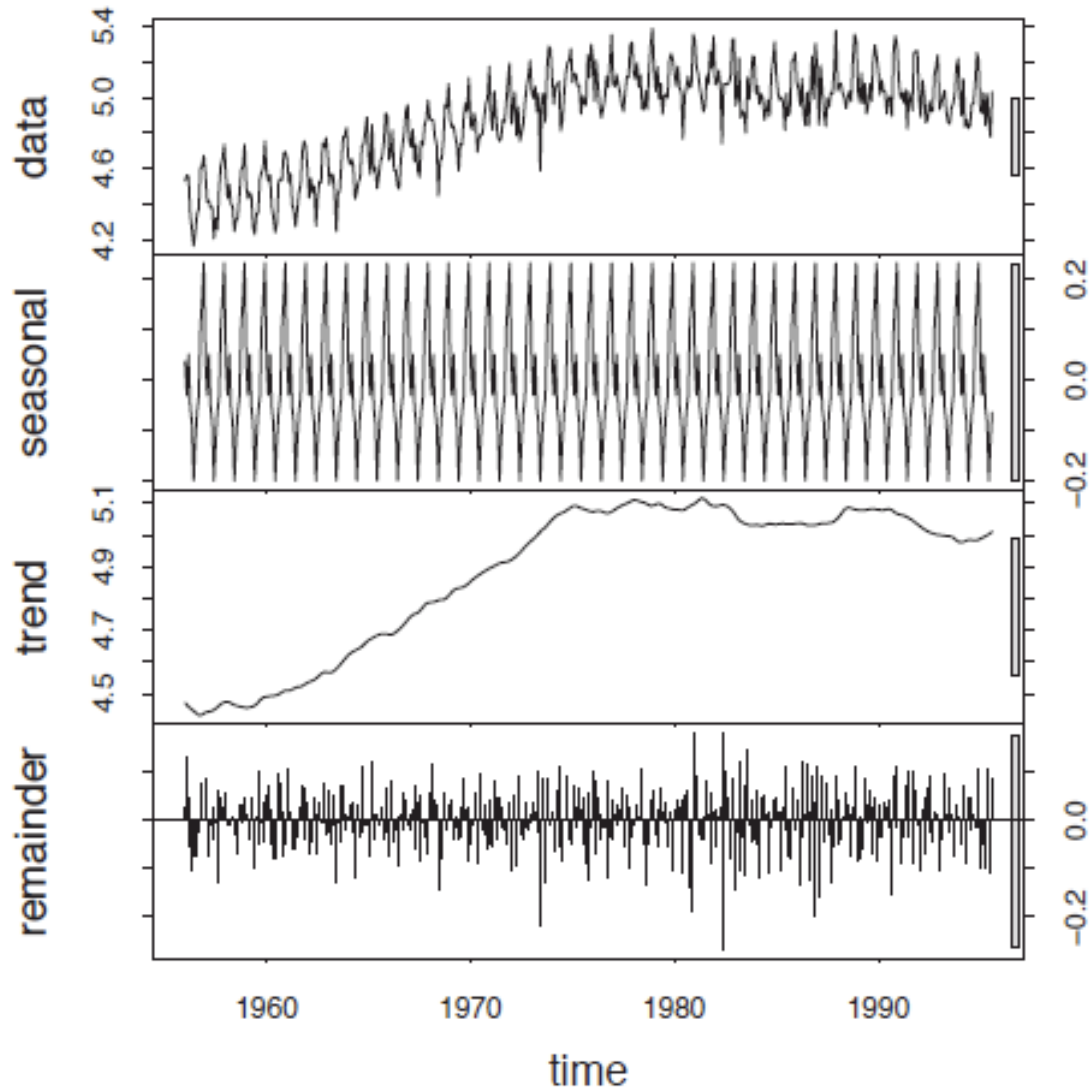
Data Density

$$\text{data density} = \frac{\text{number of entries in data array}}{\text{area of data graphic}}$$



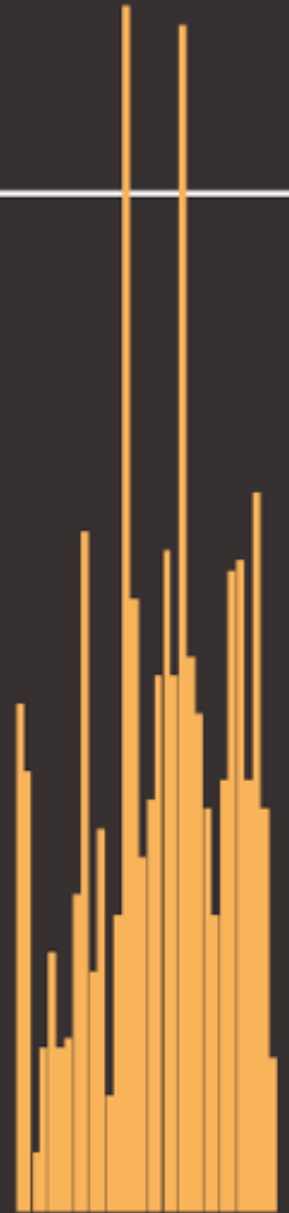
Data Density: Use Decomposition

Beer sales

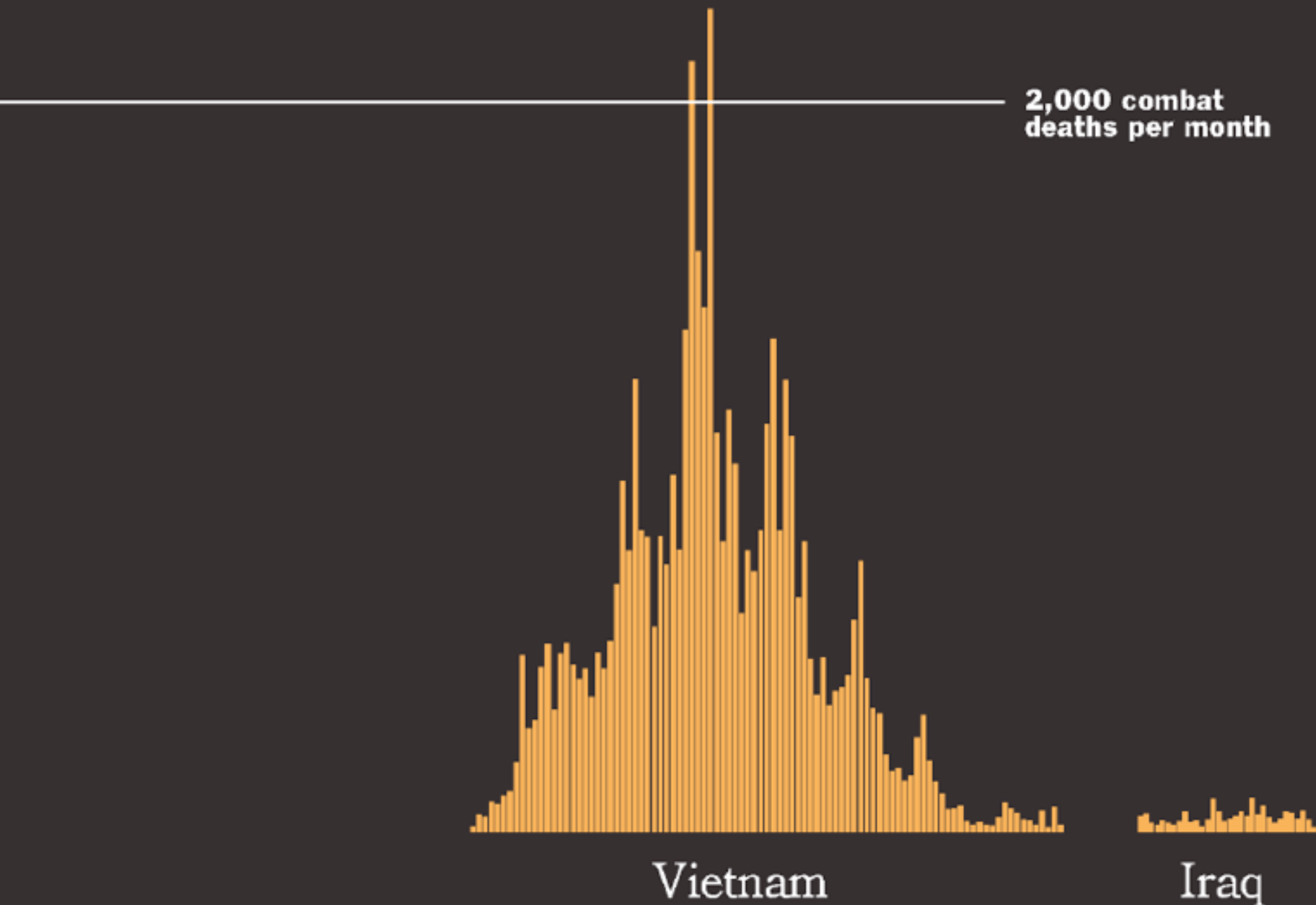


Data Density: Show Context

100 combat
deaths per month



Iraq



2,000 combat deaths per month

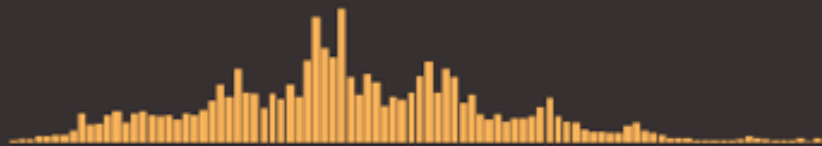
Vietnam

Iraq

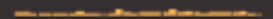
20,000 combat
deaths per month



World War II



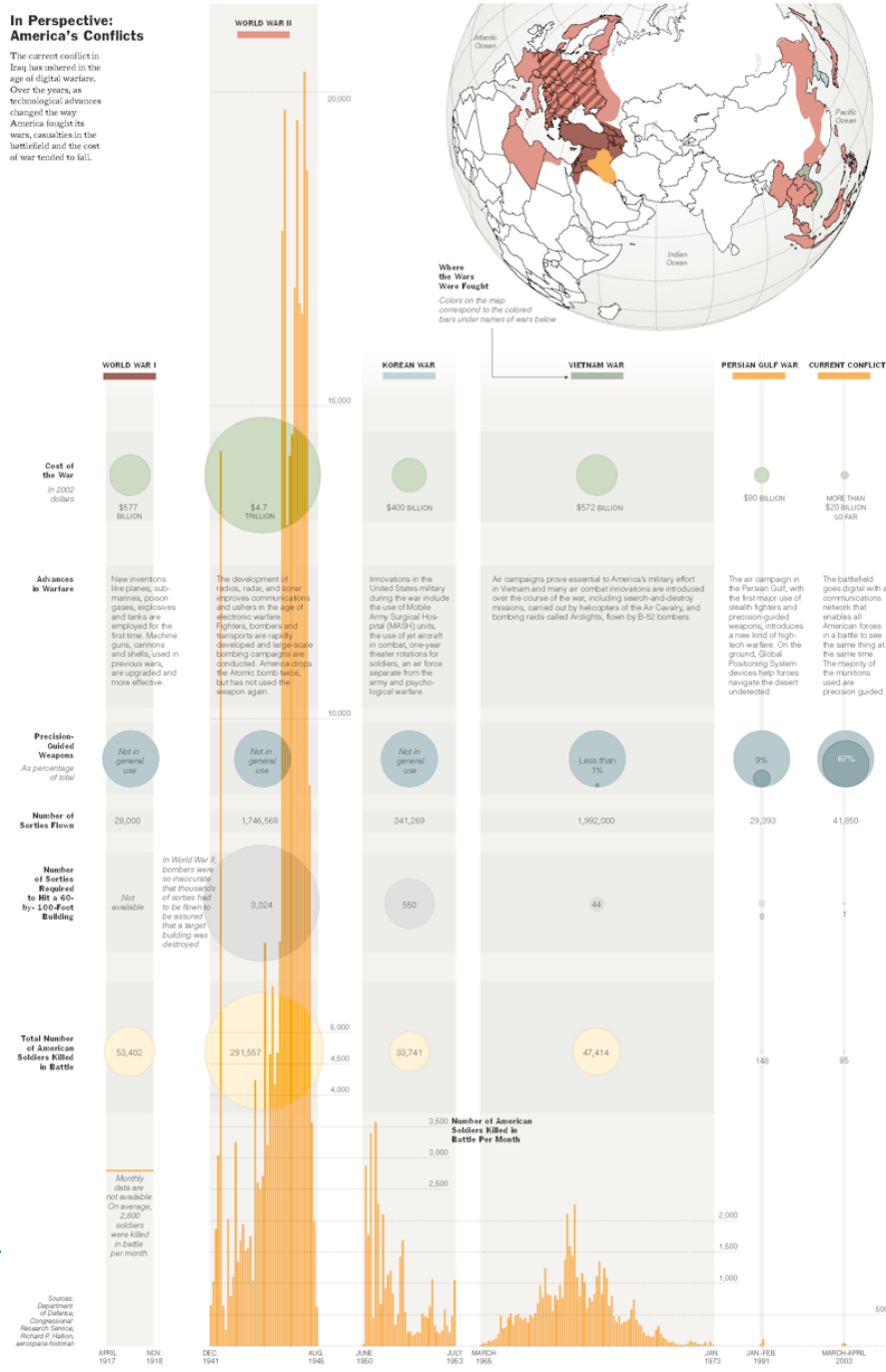
Vietnam



Iraq

In Perspective: America's Conflicts

The current conflict in Iraq has ushered in the age of digital warfare. Over the years, as technological advances changed the way America fought its wars, casualties in the battlefield and the cost of war tended to fall.



Fertility rate in European countries, 1990–2016

Compared to the replacement rate: 2.1 children per woman

Note: not all these countries belong to the European Union



Alberto Cairo, HOW CHARTS LIE

MARKETS US **Europe** Asia Currencies Crypto

↓ DAX 15,236.66 -0.093% -14.20	↑ FTSE 100 7,089.25 +0.16% +11.21	↓ CAC 40 6,579.02 -0.32% -21.17	↓ IBEX 35 8,961.40 -0.016% -1.40	↓ STOXX 50 4,089.77 -0.21% -8.57
--	---	---	--	--

.DJI · INDEX

Dow Jones Industrial Average

+ Follow Share

34,754.94 ↑ 14.99% +4,531.05 YTD

8 Oct, 08:39:55 UTC-4 · INDEXDJX · Disclaimer

1 D 5 D 1 M 6 M YTD 1 Y 5 Y MAX



 Dow Jones Industri...	34,754.94	+4,531.05	↑ 14.99%	
 Nasdaq Composite	14,654.02	+1,955.57	↑ 15.40%	X
 KOSPI	2,956.30	+11.85	↑ 0.40%	X
+ Add comparison		Clear all		

Index	
PREVIOUS CLOSE	34,416.99
DAY RANGE	34,509.72 - 34,975.19
YEAR RANGE	26,143.77 - 35,631.19

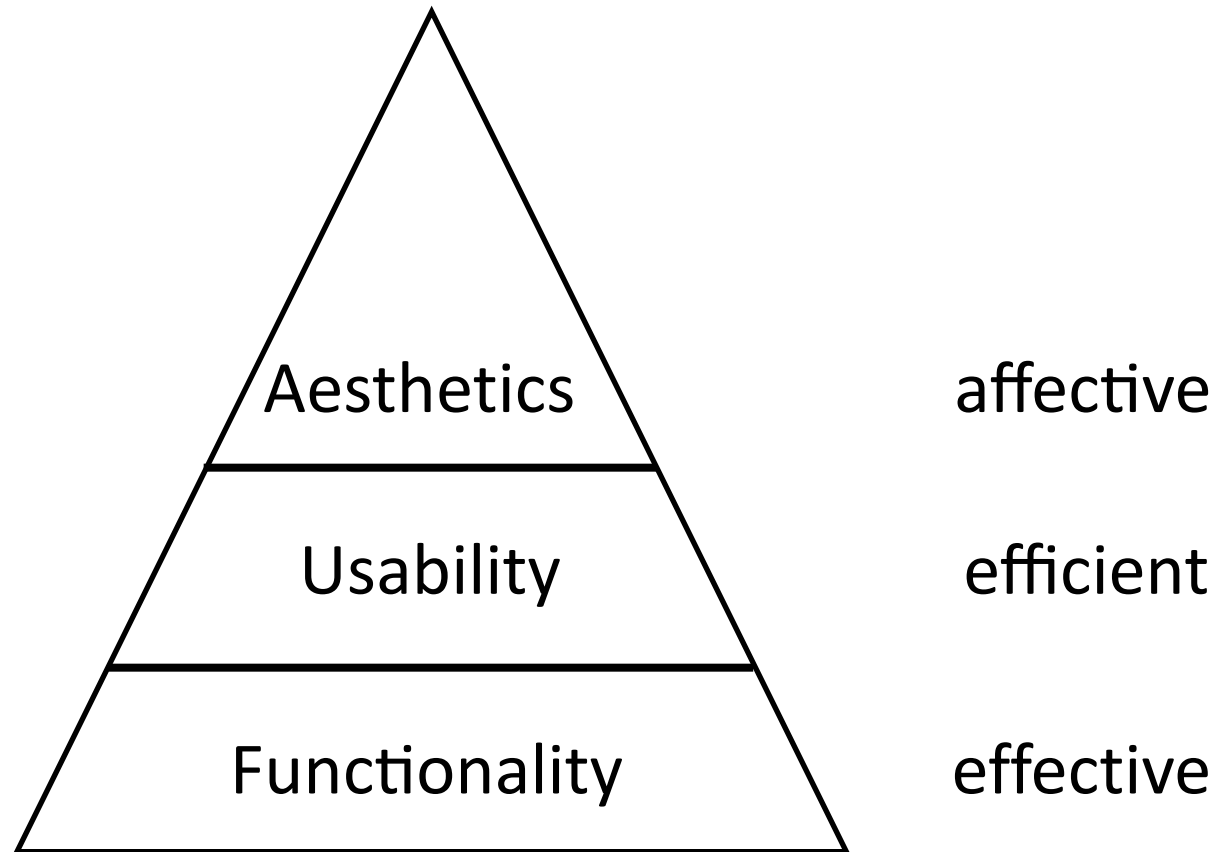
About ^

The Dow Jones Industrial Average, Dow Jones, or simply the Dow, is a price-weighted measurement stock market index of 30 prominent companies listed on stock exchanges in the United States. Although the DJIA is one of the oldest and the most commonly followed equity indices, many professionals consider the Dow to be an inadequate representation of the overall U.S. stock market compared to broader market indices such as the S&P 500 or Russell 3000 Index. The DJIA includes only 30 large companies and is a price-weighted index, unlike later stock indices which use market capitalization. Furthermore, the DJIA does not use a weighted arithmetic mean. The value of the index is the sum of the stock prices of the companies included in the index, divided by a factor which is currently approximately 0.152. The factor is changed whenever a constituent company undergoes a stock split so that the value of the index is unaffected by the stock split. First calculated on

Tufte's Design Principles

- Above all else show the data
- Maximize data-ink ratio
- Erase non-data ink
- Erase redundant data ink
- Revise and edit

Design Pyramid



Subjective Dimensions

- **Aesthetics**
 - Attractive things are perceived as more useful than unattractive ones
- **Style**
 - Communicates brand, process, who the designer is
- **Playfulness**
 - Encourages experimentation and exploration
- **Vividness**
 - Can make a visualization more memorable

Design Elements

CRAP

Contrast

Repetition

Alignment

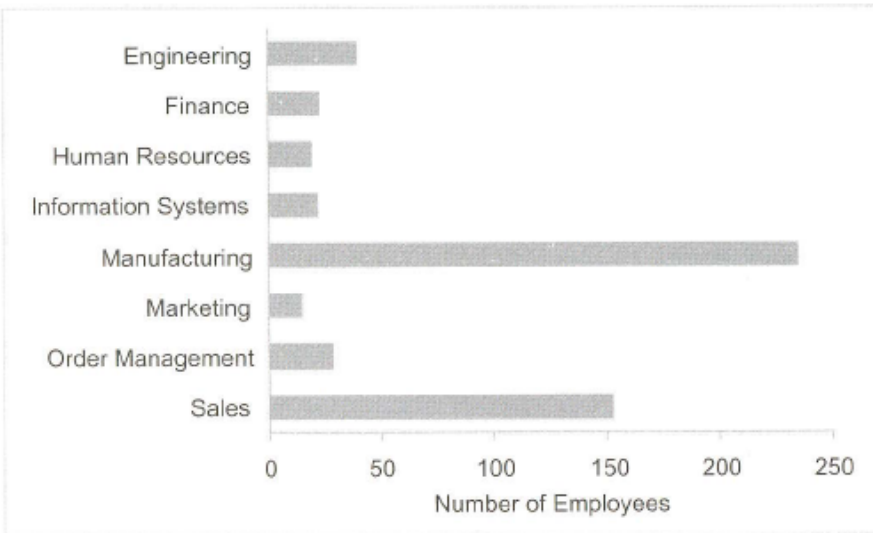
Proximity

Peter's cake metaphor ties in nicely with Galls Law

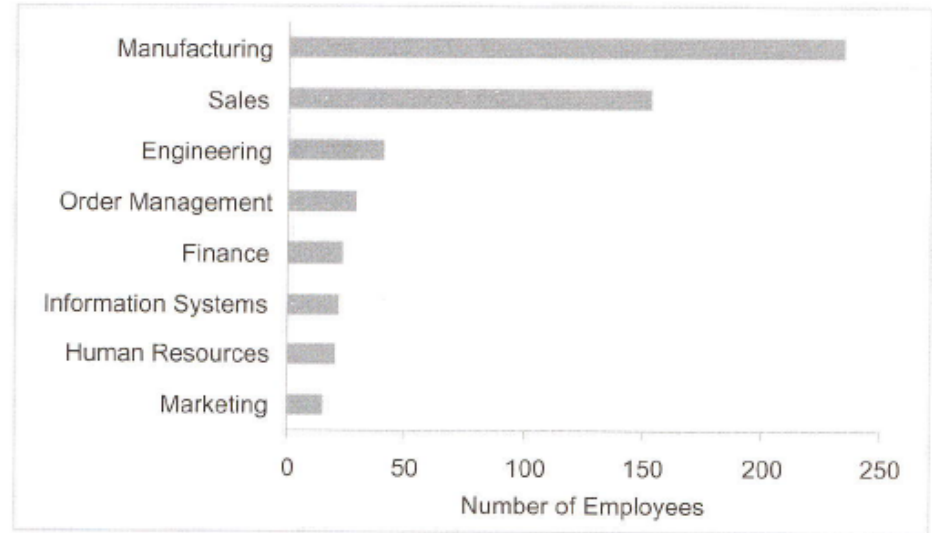
A complex system that works is invariably found to have evolved from a simple system that worked. The inverse proposition also appears to be true: A complex system designed from scratch never works and cannot be made to work. You have to start over, beginning with a working simple system.

Alignment

Before



After



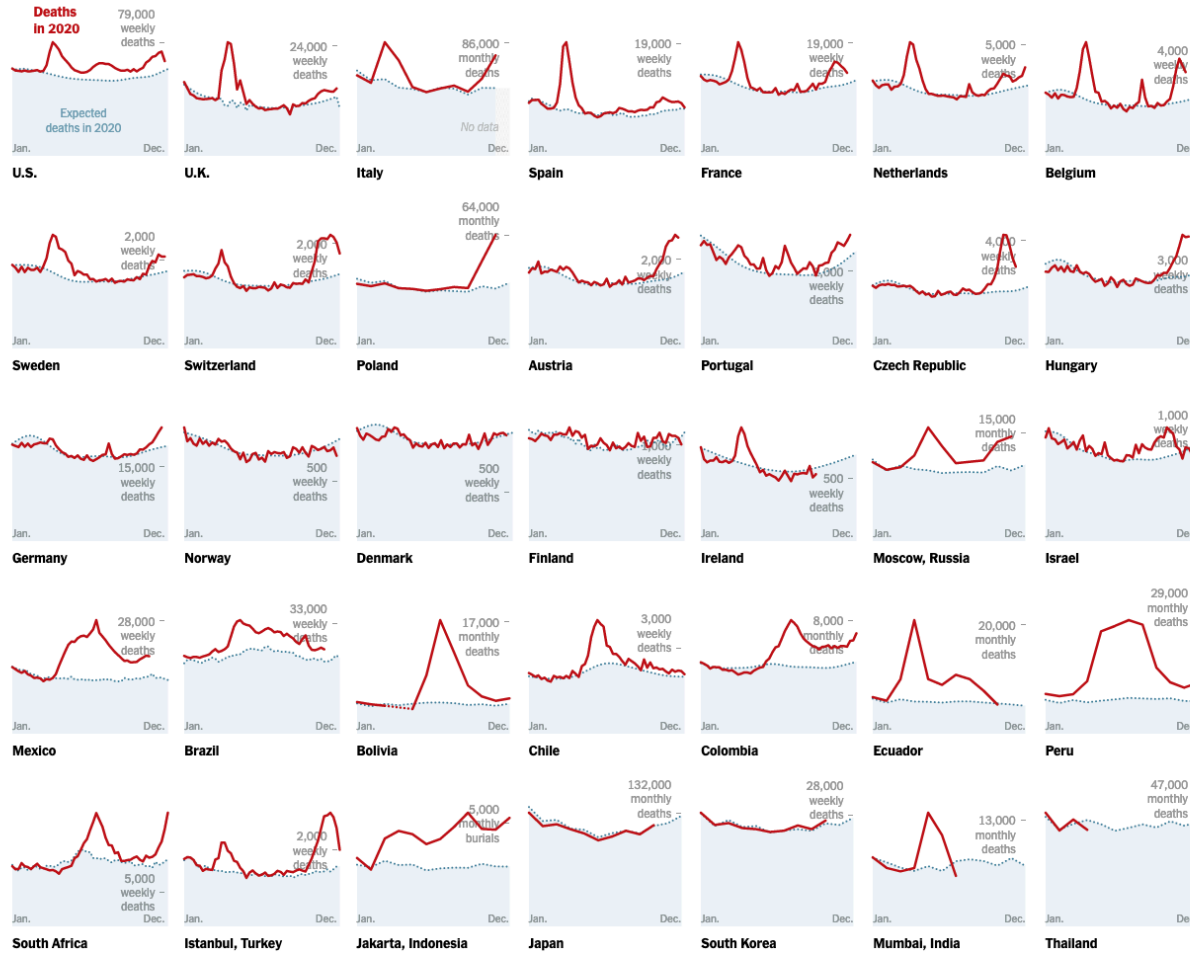
S. Few, Show me the numbers

Repetition



The Pandemic's Hidden Toll: Half a Million Deaths

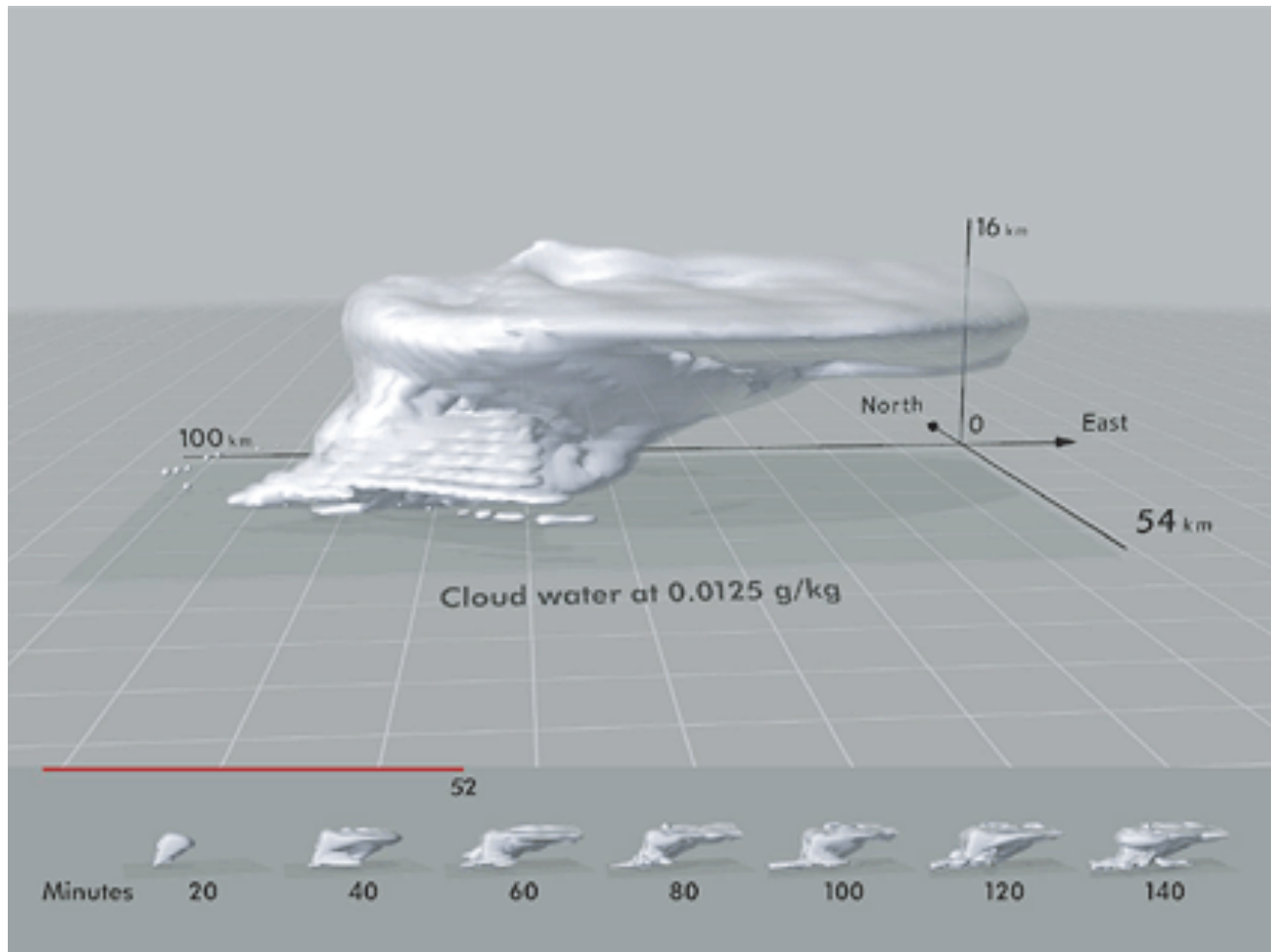
By Jin Wu, Allison McCann, Josh Katz, Eilan Peltier and Karan Deep Singh Updated Feb. 9, 2021



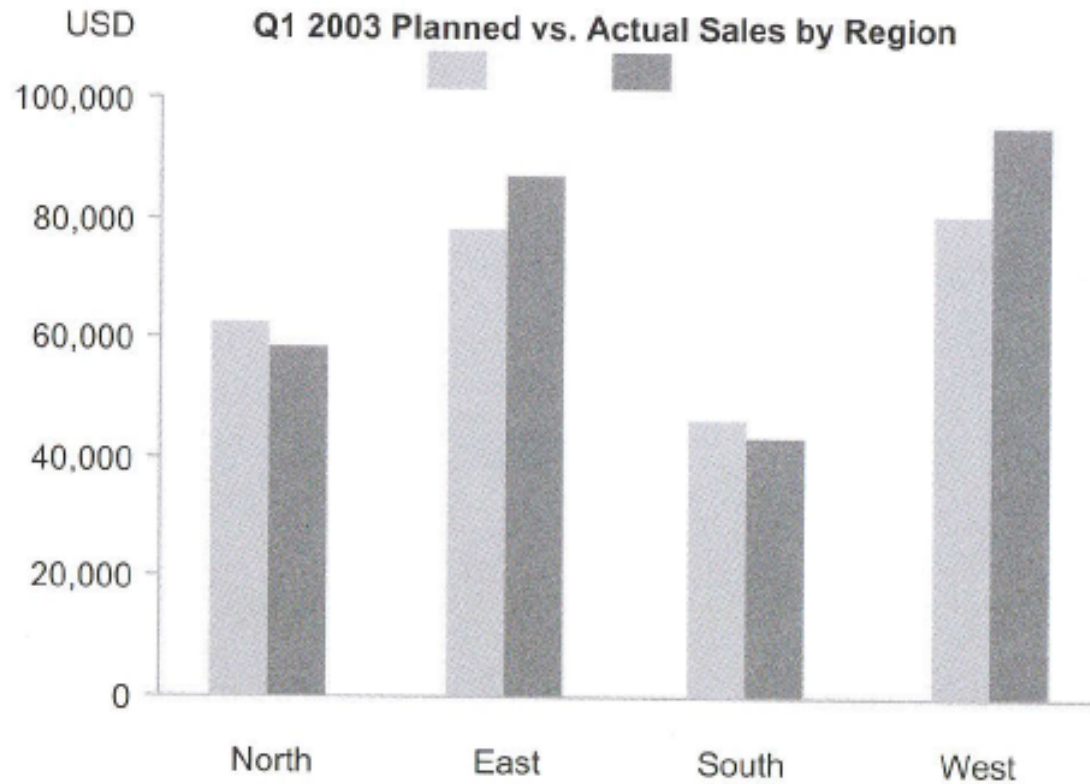
Note: Bolivia's Civil Registry recorded almost no deaths in April due to the closure of government offices during a lockdown. Officials said least some of the deaths that occurred in April could have been registered in later months.

NYT, retrieved October 2021: <https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html>

Small Multiples

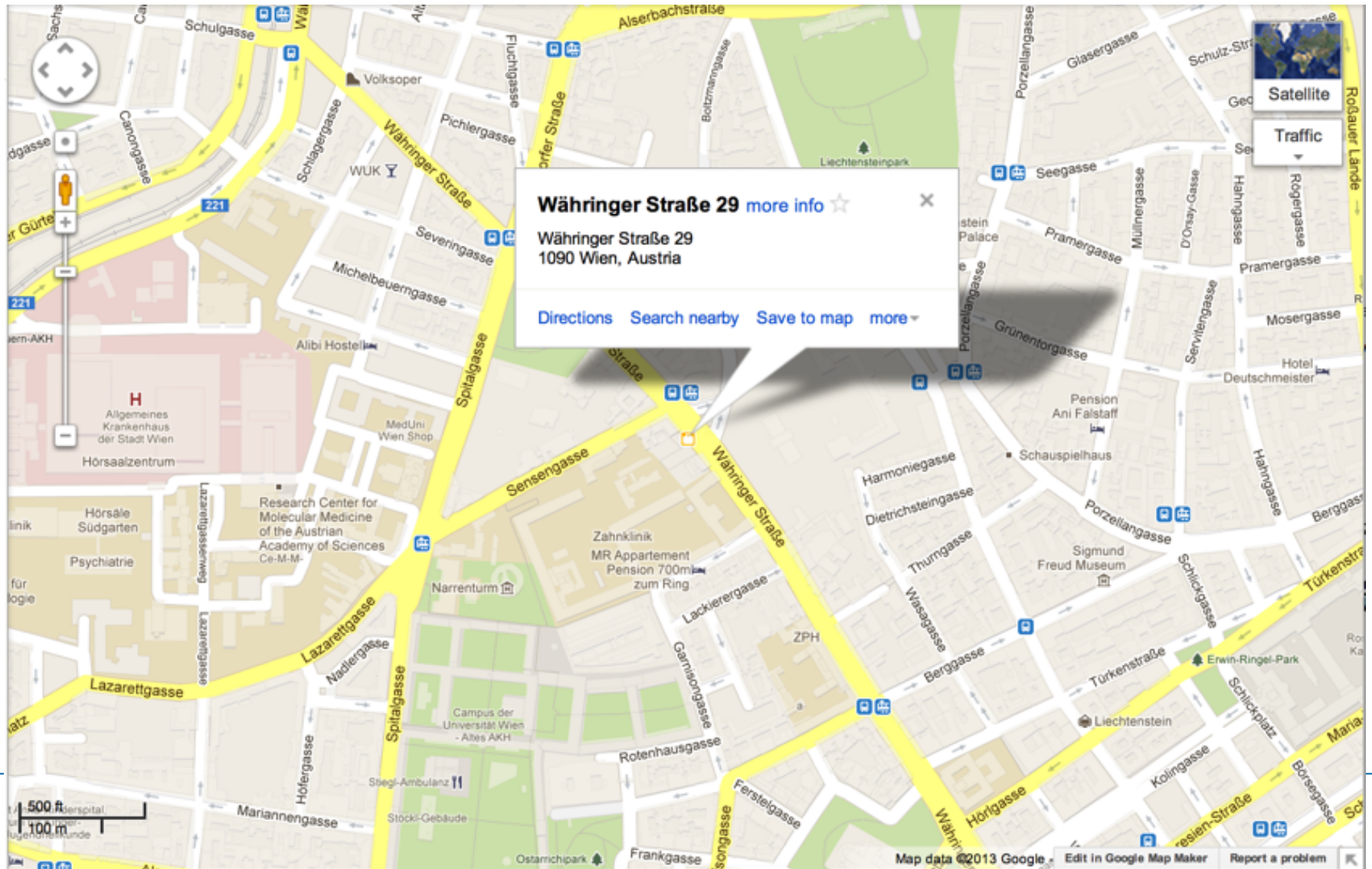


Proximity



S. Few, Show me the numbers

Layering and Separation



Layering and Separation

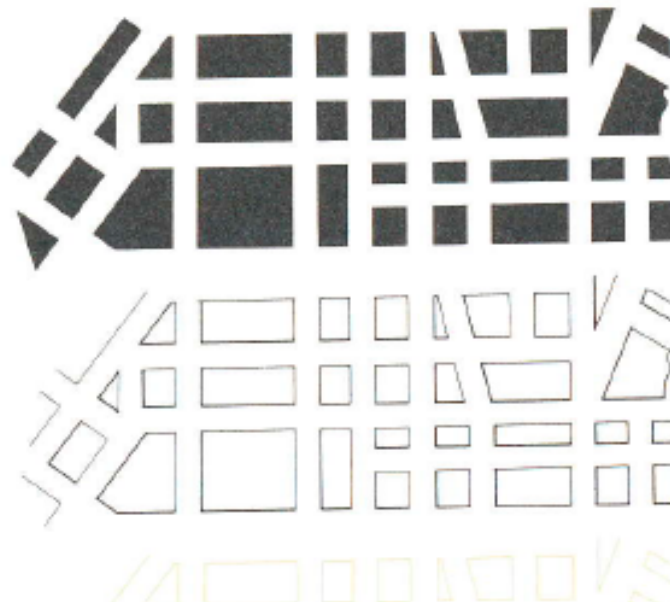
Train No.	3701	3301	3801	3542	3765
New York	12:10	1:30	3:45	7:30	4:33
Newark, N. J.	1:43	10:30	5:21	8:50	11:45
North Elizabeth	6:45
Elizabeth	3:33	2:05	7:05
Peekskill	5:34	6:40	7:20	8:50
Ediison, N. J.	4:45	5:20	4:40	2:10	11:05
Princeton, N. J.	1:30	3:30	7:30

Layering and Separation

New York	12:10	1:30	3:45	7:30	4:33
Newark, N. J.	1:43	10:30	5:21	8:50	11:45
North Elizabeth	6:45
Elizabeth	3:33	2:05	7:05
Peekskill	5:34	6:40	7:20	8:50
Ediison, N. J.	4:45	5:20	4:40	2:10	11:05
Princeton, N. J.	1:30	3:30	7:30
Train No.	3701	3301	3801	3542	3765

Negative Space

CROSSWIND TAKEOFF
THE SLIP
CROSSWIND LANDINGS
SHORT FIELD TAKE OFF & LANDING
SOFT FIELD TAKE OFF & LANDING
FORCED LANDING
720 POWER TURNS



Negative Space



Negative Space Logos

Tufte's Graphical Excellence

- Interesting data
 - Complex ideas, multivariate data
- Clear, precise, concise presentation
 - Data-ink ratio
- Accurate communication
 - Lie factor

Graphical displays should...

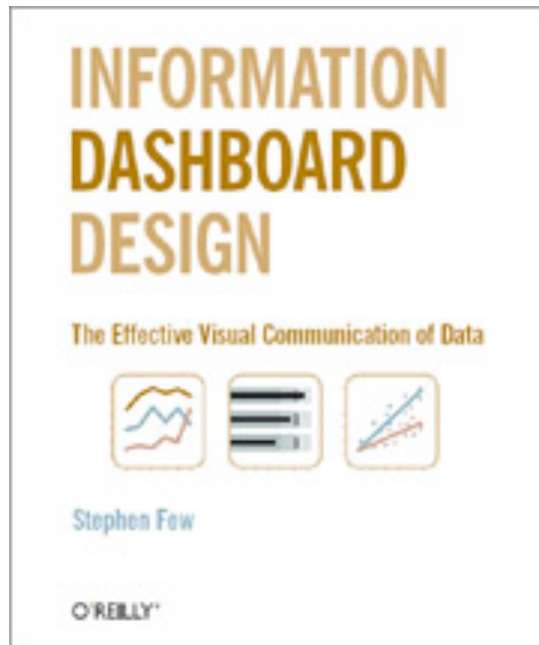
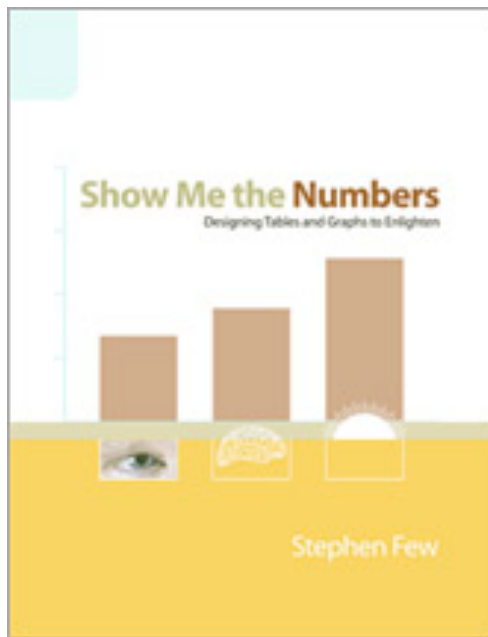
- Show the data
- Induce the viewer to think about the substance, rather than about methodology, graphic design, [or] the technology of graphic productions...
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail
- Serve a reasonably clear purpose
- Be closely integrated with the statistical and verbal descriptions

Analysis Questions

- Who is the intended audience?
 - What information does this visualization represent?
 - How many data dimensions does it encode?
 - List several tasks, comparisons or evaluations it enables
 - What principles of excellence best describe why it is good / bad?
 - Can you suggest any improvements?
 - Why do you like / dislike this visualization?
-

Further Reading

Stephen Few



Few is Applied Tufte

Relationship	Value-Encoding Object			
	Points	Lines	Points & Lines	Bars
<i>Nominal Comparison</i>	When there is a need to narrow the quantitative scale, and in so doing, remove zero from its base	Avoid	Avoid	Either horizontal or vertical bars
<i>Time Series</i>	Avoid	Categorical subdivisions on X axis, quantitative values on Y axis; emphasis on overall pattern	Categorical subdivisions on X axis, quantitative values on Y axis; mutual emphasis on overall pattern and individual values	Categorical subdivisions on X axis, quantitative values on Y axis; emphasis on individual values
<i>Ranking</i>	When there is a need to narrow the quantitative scale, and in so doing, remove zero from its base	Avoid	Avoid	Horizontal bars are preferable, with values sorted in descending order.
<i>Part-to-Whole</i>	Avoid	Avoid	Avoid	Either horizontal or vertical bars
<i>Deviation</i>	Avoid	Especially useful when combined with time series	Useful when combined with time series and when a slight emphasis on individual values is desired	Either horizontal or vertical bars, except when combined with time series, which requires vertical bars
<i>Distribution Single</i>	Avoid	Known as a <i>frequency polygon</i> ; emphasis on overall pattern	Avoid	Known as a <i>histogram</i> ; emphasis primarily on individual values
<i>Multiple</i>	Use to mark the median in a box plot	Avoid	Avoid	Use in the form of range bars in box plots
<i>Correlation</i>	Known as a <i>scatter plot</i>	Avoid	In this case the line is a trend line, not a line that connects the points.	Either horizontal or vertical bars; can be structured either as a <i>correlation bar graph</i> or a <i>paired bar graph</i>

S. Few, Show Me the Numbers, p. 87

Paper

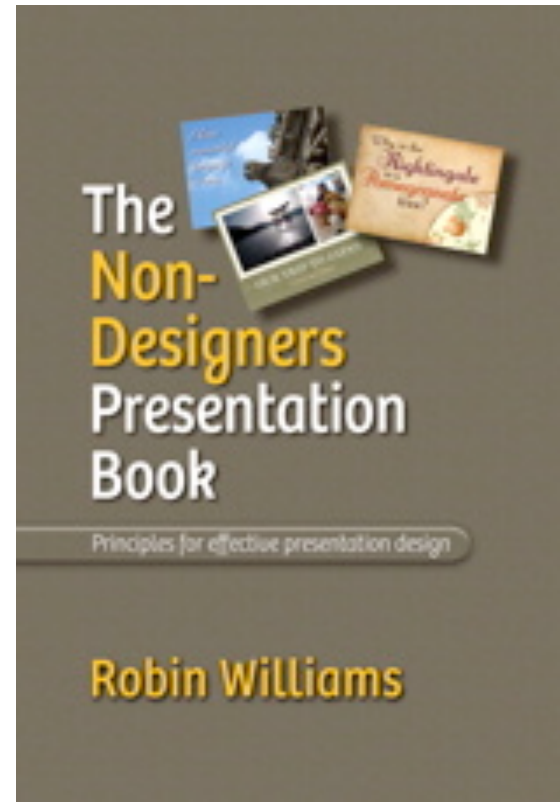
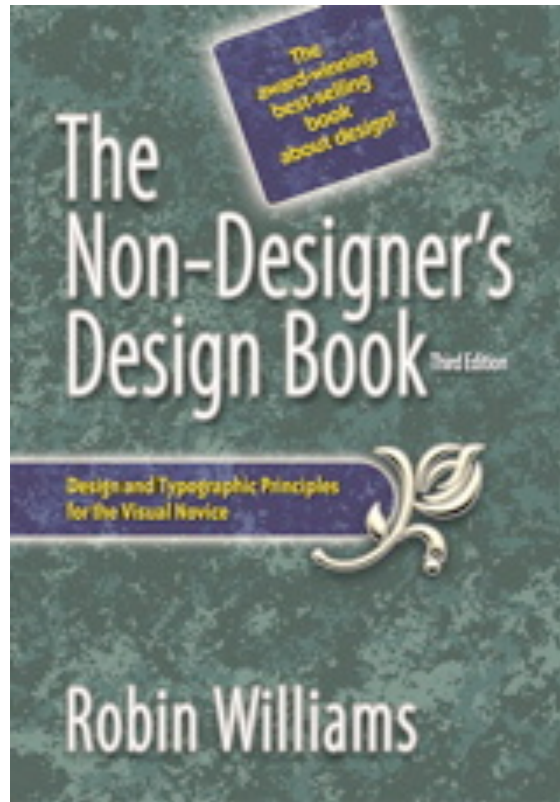
The Science of Visual Data Communication: What Works

Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman⁵

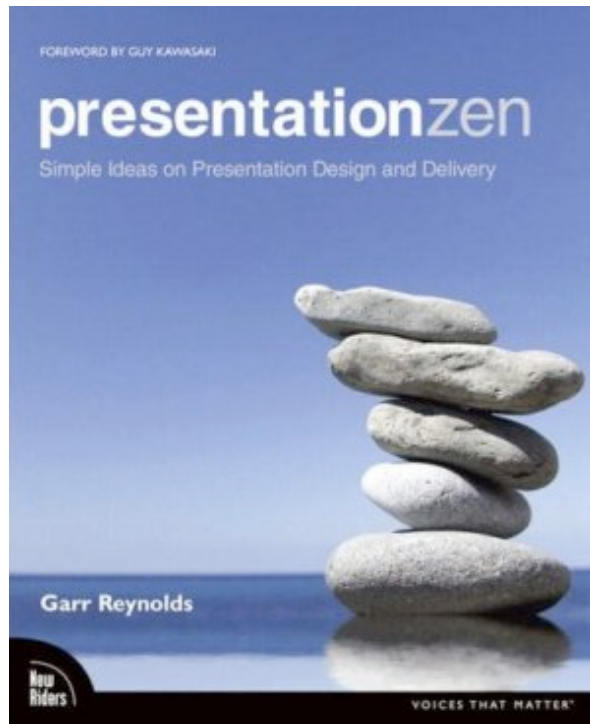
Abstract

Effectively designed data visualizations allow viewers to use their powerful visual systems to understand patterns in data across science, education, health, and public policy. But ineffectively designed visualizations can cause confusion, misunderstanding, or even distrust—especially among viewers with low graphical literacy. We review research-backed guidelines for creating effective and intuitive visualizations oriented toward communicating data to students, coworkers, and the general public. We describe how the visual system can quickly extract broad statistics from a display, whereas poorly designed displays can lead to misperceptions and illusions. Extracting global statistics is fast, but comparing between subsets of values is slow. Effective graphics avoid taxing working memory, guide attention, and respect familiar conventions. Data visualizations can play a critical role in teaching and communication, provided that designers tailor those visualizations to their audience.

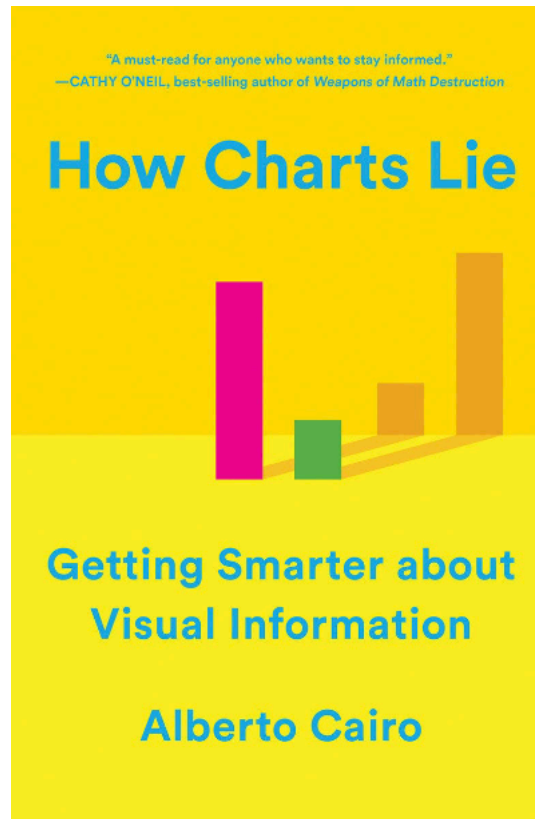
Robin Williams



Garr Reynolds



Alberto Cairo



2019

From: <https://albertocairo.com/>