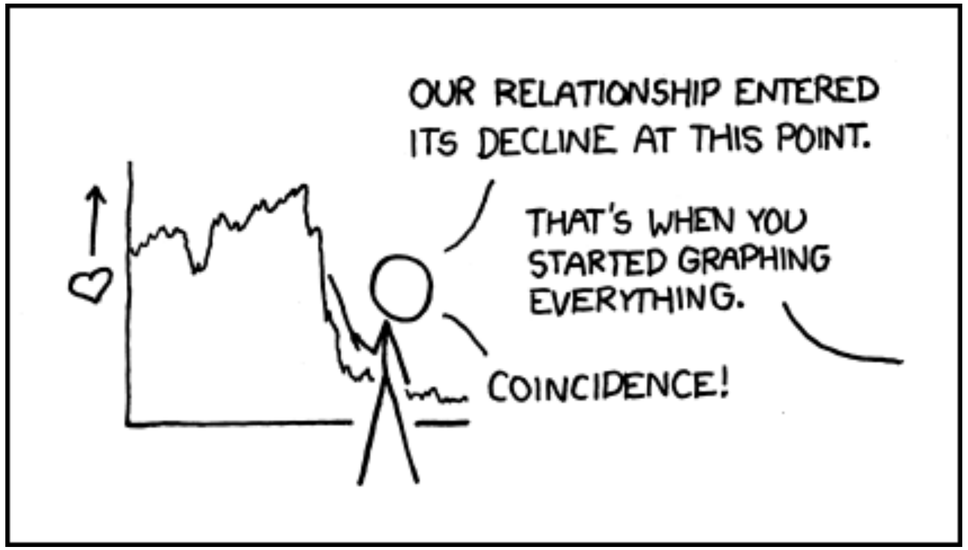
CS-5630 / CS-6630 Visualization for Data Science Design Guidelines

Alexander Lex
alex@sci.utah.edu

THE UNIVERSITY OF UTAH



[xkcd]

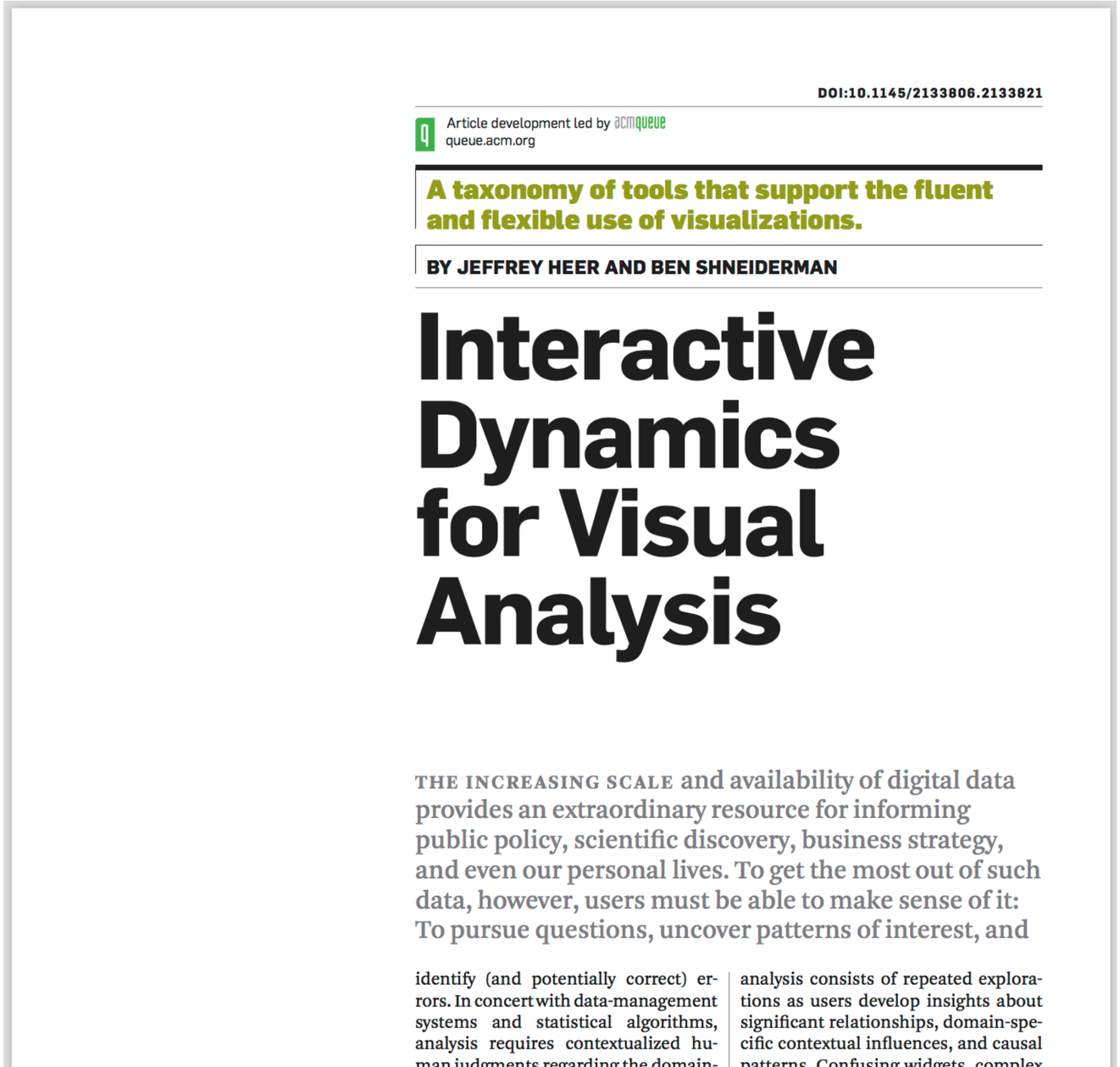
Next Week

Tuesday: D3 Layouts

Thursday: Interaction

Mandatory Reading

Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. https://doi.org/10.1145/2133806.2133821



DOI:10.1145/2133806.2133821

Article development led by acmqueue
queue.acm.org

A taxonomy of tools that support the fluent and flexible use of visualizations.

BY JEFFREY HEER AND BEN SHNEIDERMAN

Interactive Dynamics for Visual Analysis

THE INCREASING SCALE and availability of digital data provides an extraordinary resource for informing public policy, scientific discovery, business strategy, and even our personal lives. To get the most out of such data, however, users must be able to make sense of it: To pursue questions, uncover patterns of interest, and

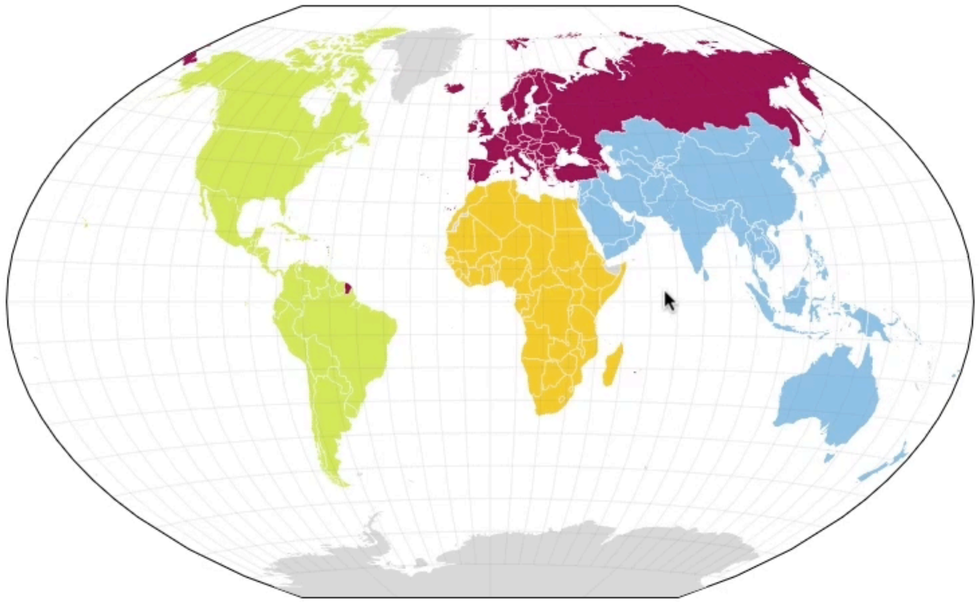
identify (and potentially correct) errors. In concert with data-management systems and statistical algorithms, analysis requires contextualized human judgments regarding the domain-

analysis consists of repeated explorations as users develop insights about significant relationships, domain-specific contextual influences, and causal

Next Homework

Gap Minder Inspired World Health Data

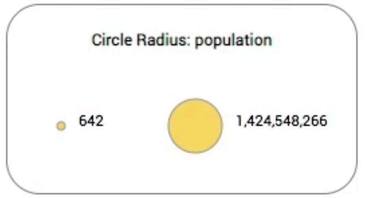
Name: Your Name; E-Mail: Your E-Mail; UID: Your UID



Singapore

Asia

Population: 5,935,053
GDP per capita: 86,473
Total fertility rate: 1.27
Child mortality (under age five): 2.424
Life expectancy: 84.37



Today's Reading

View This Article
FULL-TEXT (HTML)
FULL-TEXT (PDF)
IN DIGITAL EDITION
COMMENTS

Reader Tools
PRINT
TEXT SIZE
SHARE

PREV ISSUE | NEXT ISSUE
VIEW IN DIGITAL LIBRARY
DIGITAL EDITION FORMAT
VIEW IN PDF FORMAT

COVER STORY

XXV.4 July - August 2018
Page: 26
Digital Citation

THE GOOD, THE BAD, AND THE BIASED: FIVE WAYS VISUALIZATIONS CAN MISLEAD (AND HOW TO FIX THEM)

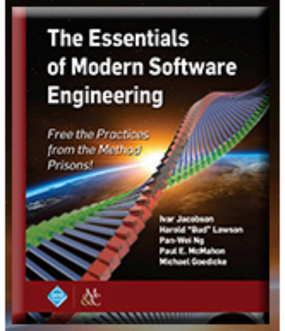
Authors:
Danielle Szafir

Data visualizations allow people to readily explore and communicate knowledge drawn from data. Visualization methods range from standard scatterplots and line graphs to intricate interactive systems for analyzing large data volumes at a glance. But how can we craft visualizations that effectively communicate the right information from our data? What aspects of data and design need to come together to develop accurate insights? The answer lies in the way we see the world: People use their visual and cognitive systems (i.e., our eyes and brain) to extract meaning from visualized data. However, flashy visualizations are not always optimized to help people see what matters. This article reviews common visualization practices that may inhibit effective analysis, why these designs are problematic, and how to avoid them. The discussion illustrates a need to better understand how visualizations can support flexible and accurate data analysis while mitigating potential sources of bias.

Insights

→ Visualizations allow people to readily analyze and communicate data. However, many common

Free the Practices from the Method Prisons!



The Essentials of Modern Software Engineering

Ivar Jacobson, et al

ISBN: 978-1-947487-24-6
DOI: 10.1145/3277669

books.acm.org
store.morganclaypool.com/acm

Design Guidelines

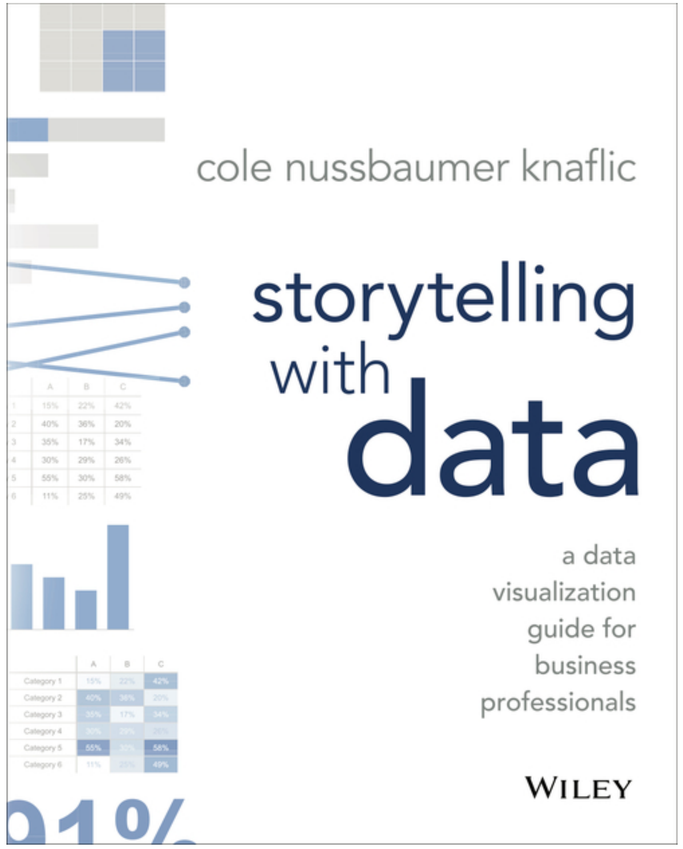
Rule #1: Use the Best Visual Channel Available for the Most Important Aspect of your Data

Rule #2: The visualization should show all of the data, and only the data

Book Recommendation

Great book with simple design guidelines

Not a “Visualization” book, but a “charting” book



Tufte's Integrity Principles

Show **data variation**, not design variation

Clear, detailed, and thorough **labeling** and **appropriate scales**

Size of the **graphic effect** should be **directly proportional to the numerical quantities** ("lie factor")

Scales

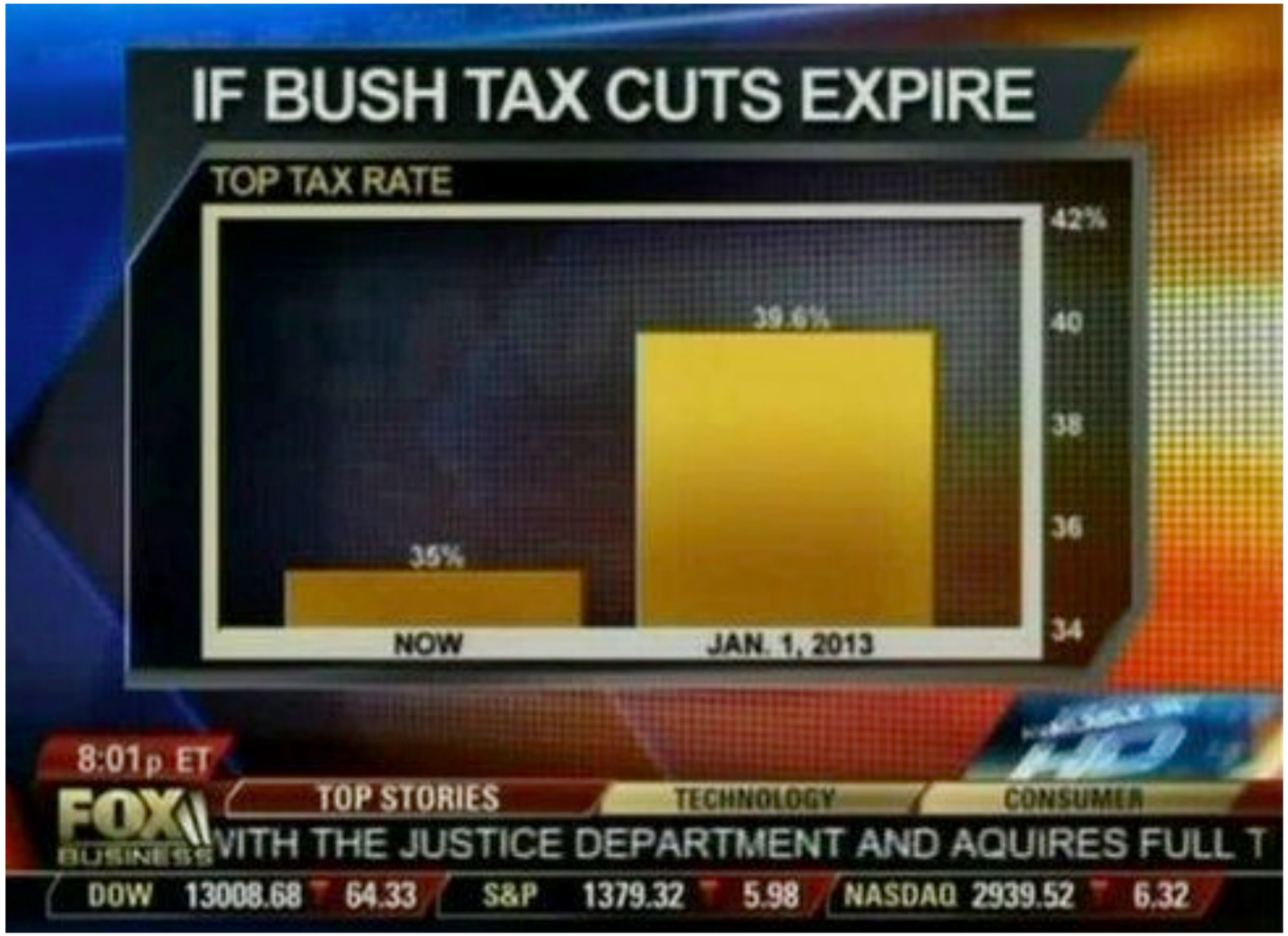
The Lie Factor

$$\frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$

Lie Factor - Graphical Integrity

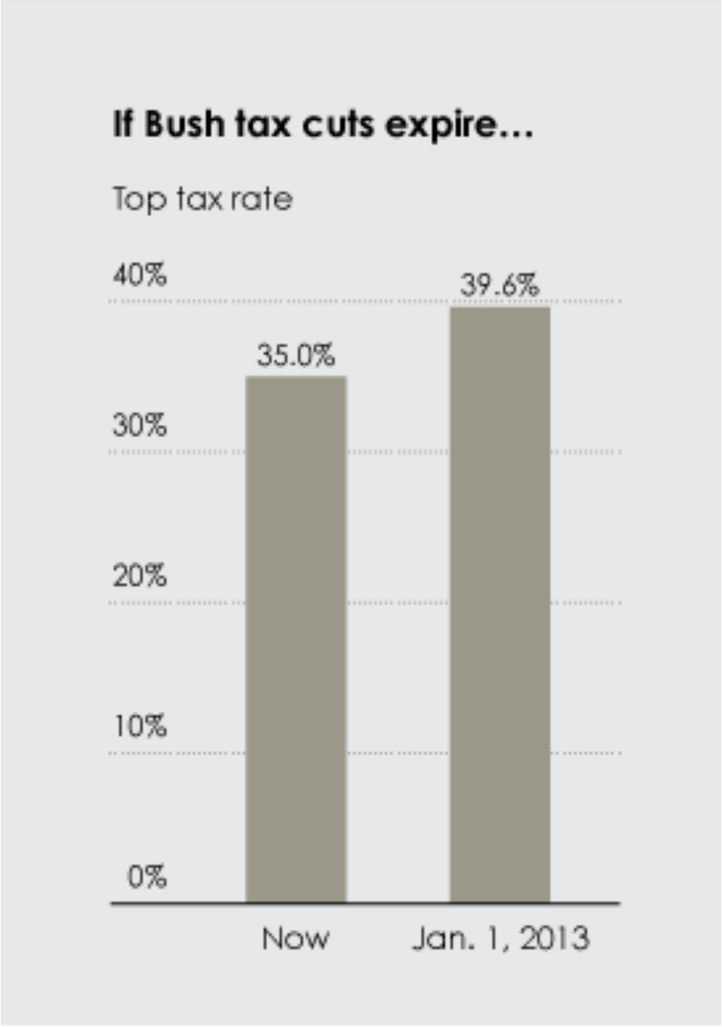
Magnitude in data must correspond to magnitude of mark

Effect in Data: factor 1.14
Effect in Graphic: factor 5
Lie Factor: 5/1.14 = 4.38



Flowing Data

Scale Distortions



Flowing Data

What’s wrong?

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

What’s wrong?

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

1
2
3

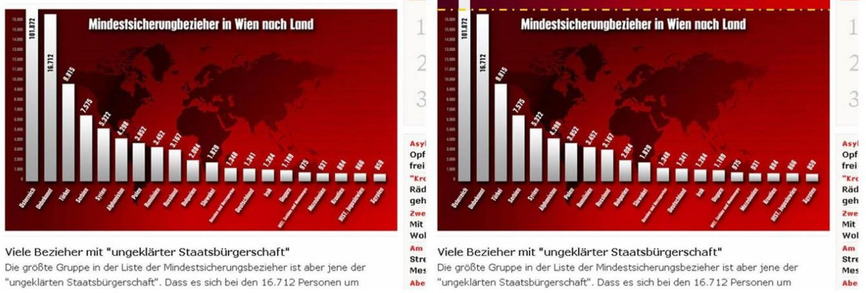
Asyl
Opf
frei
"Kr
Räd
geh
Zwe
Mit
Wol
Am
Stre
Mes
Abe

Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

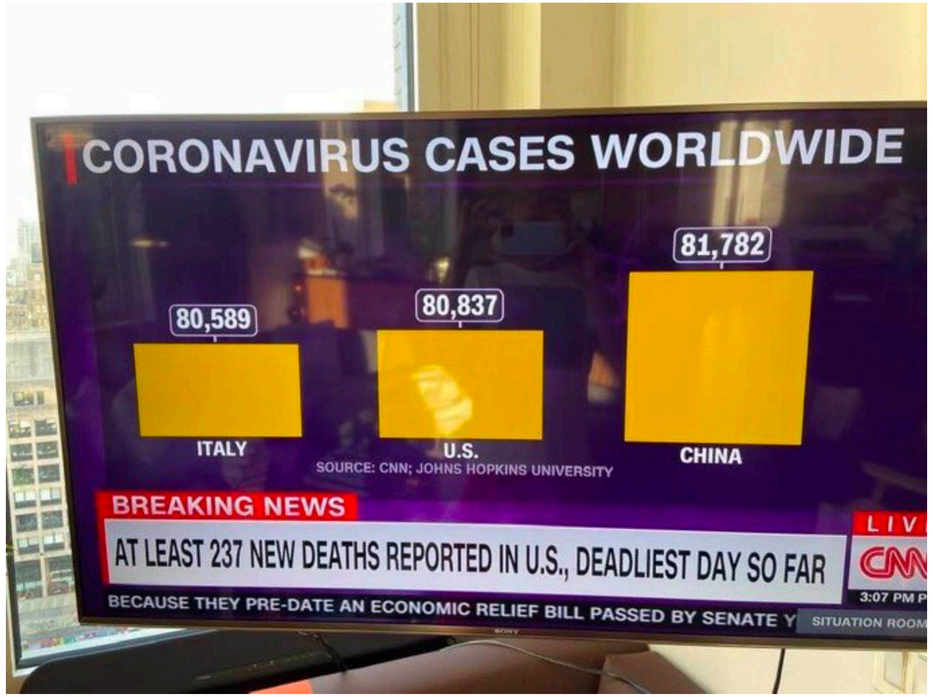
Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

What’s wrong?

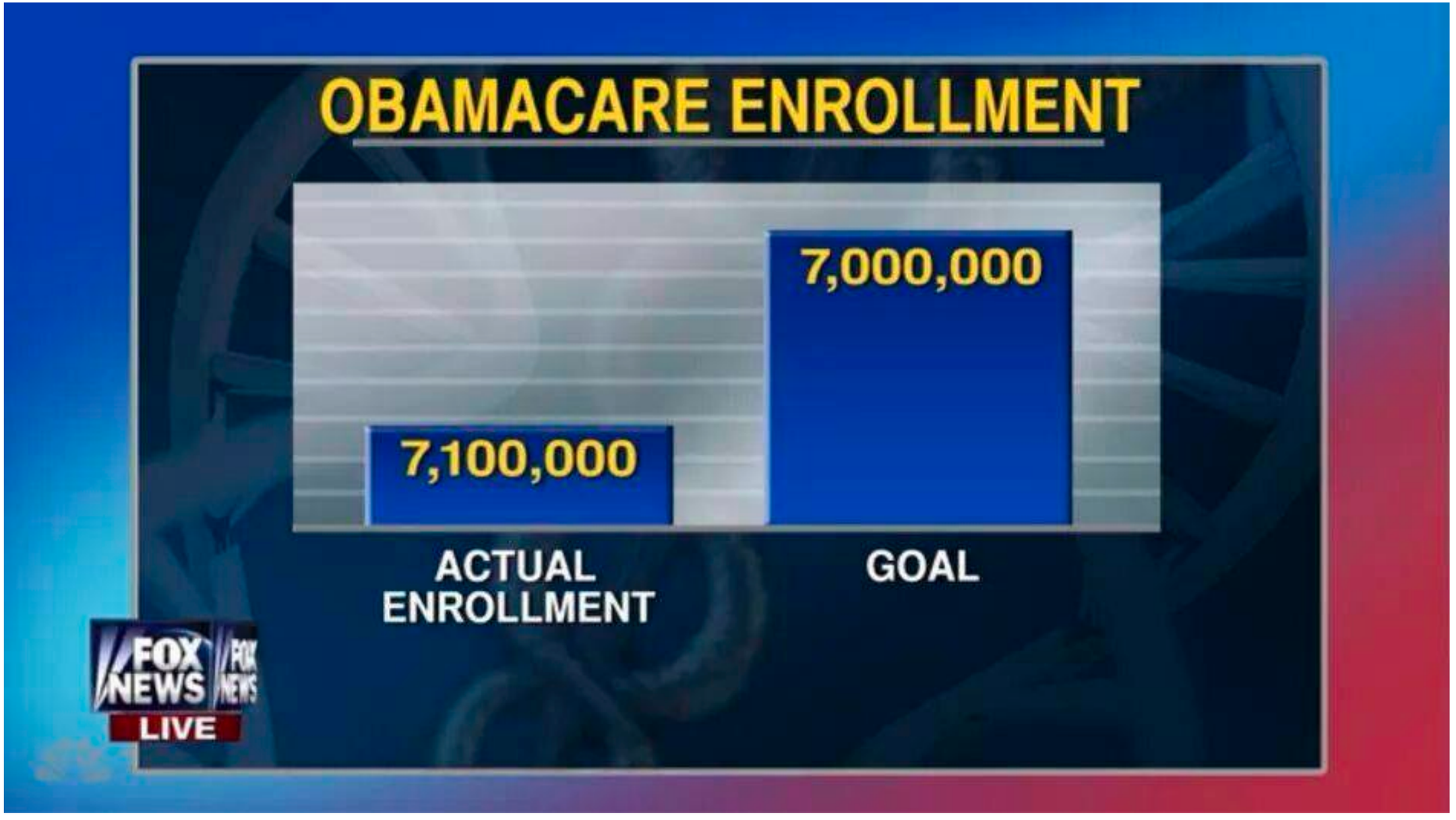
Grafik der Kronen-zeitung



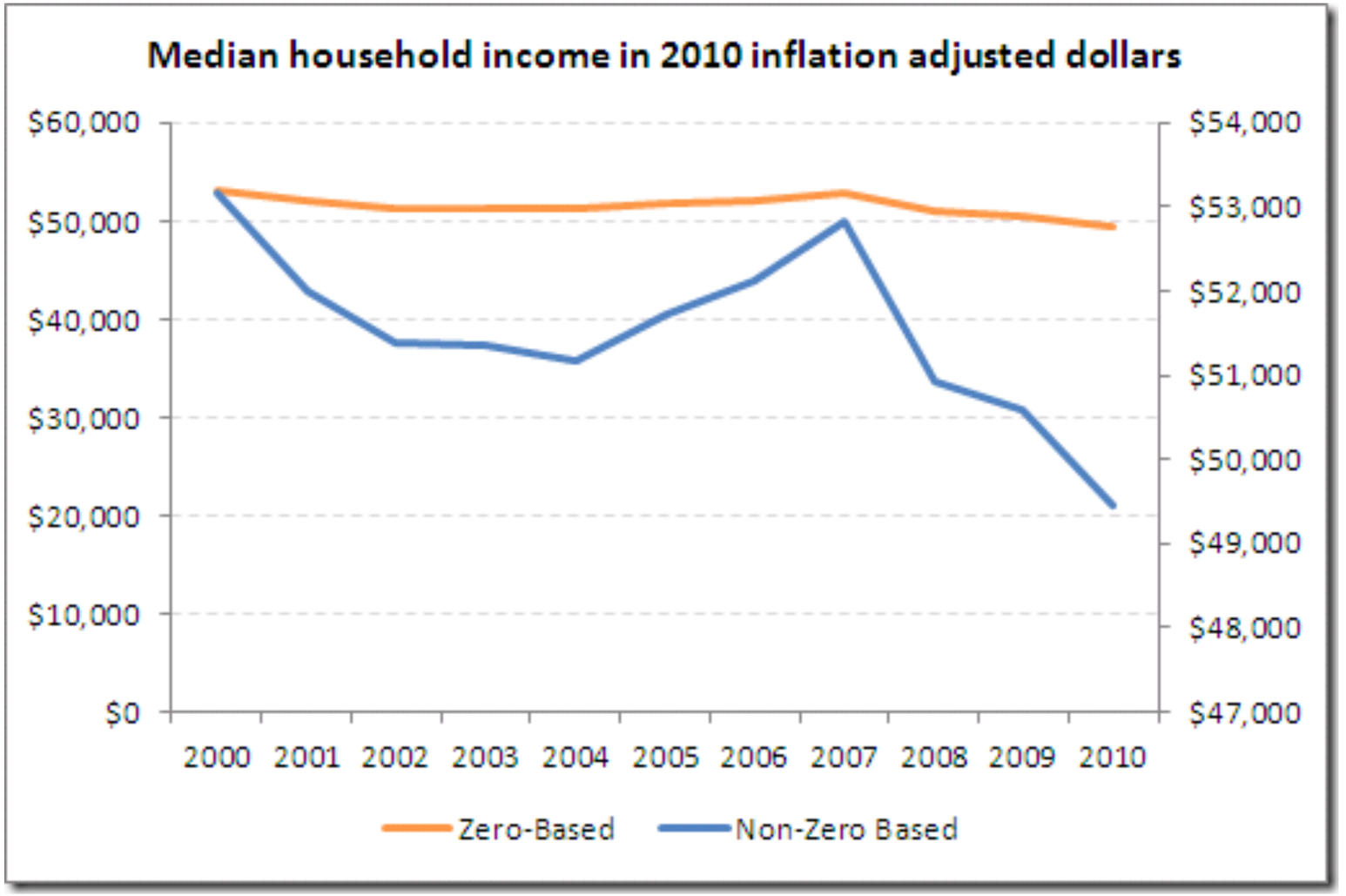
Grafik in echt



https://twitter.com/StatsbyLopez/status/1243564270970904581


OBAMACARE ENROLLMENT
7,100,000
7,000,000
ACTUAL ENROLLMENT
GOAL
FOX NEWS
LIVE

Start Scales at 0?



A. Kriebel, VizWiz

Use a baseline that shows the data, not the zero-point.

Think about: what is a meaningful baseline?

E. Tufte

Scales at 0

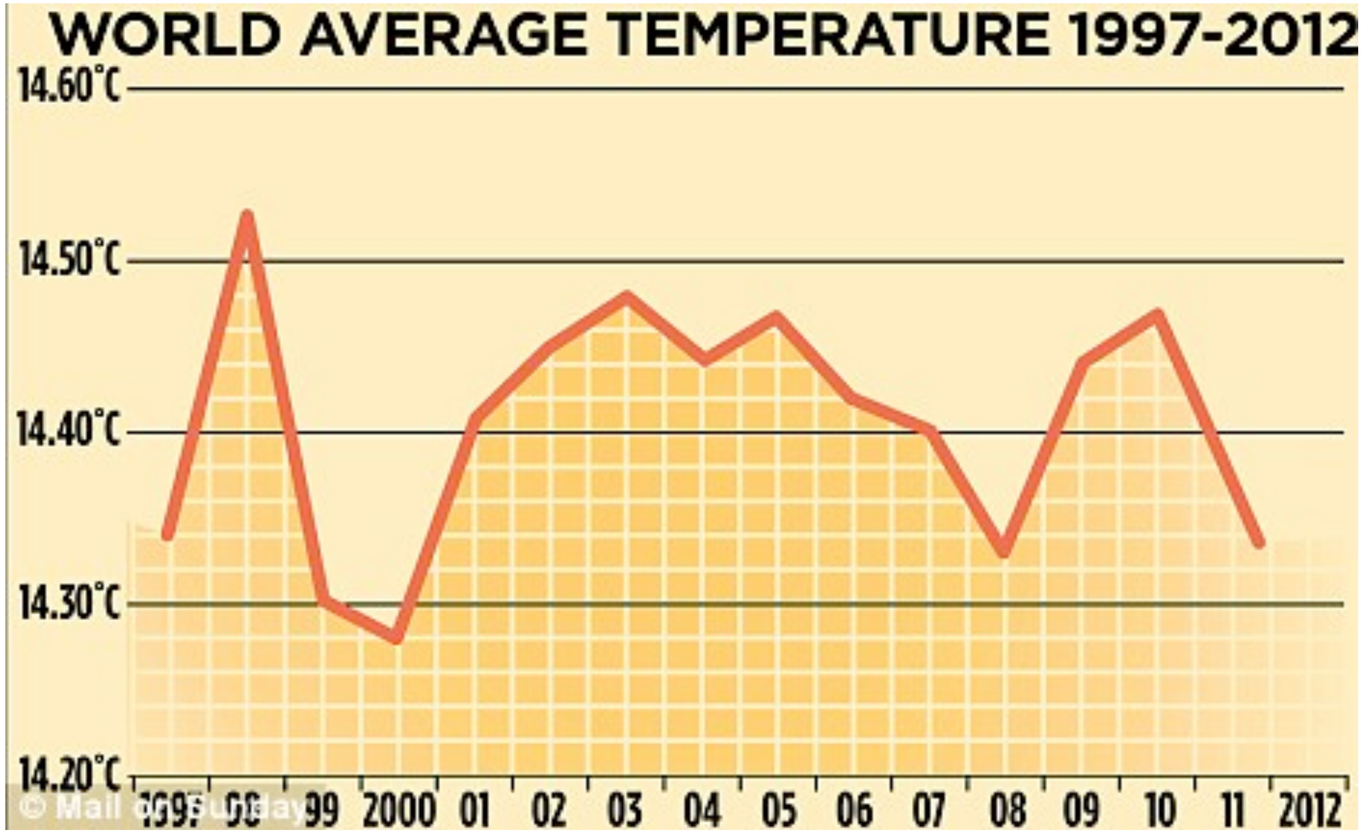
Framing

Vis can be used to lie

- just as language or statistics

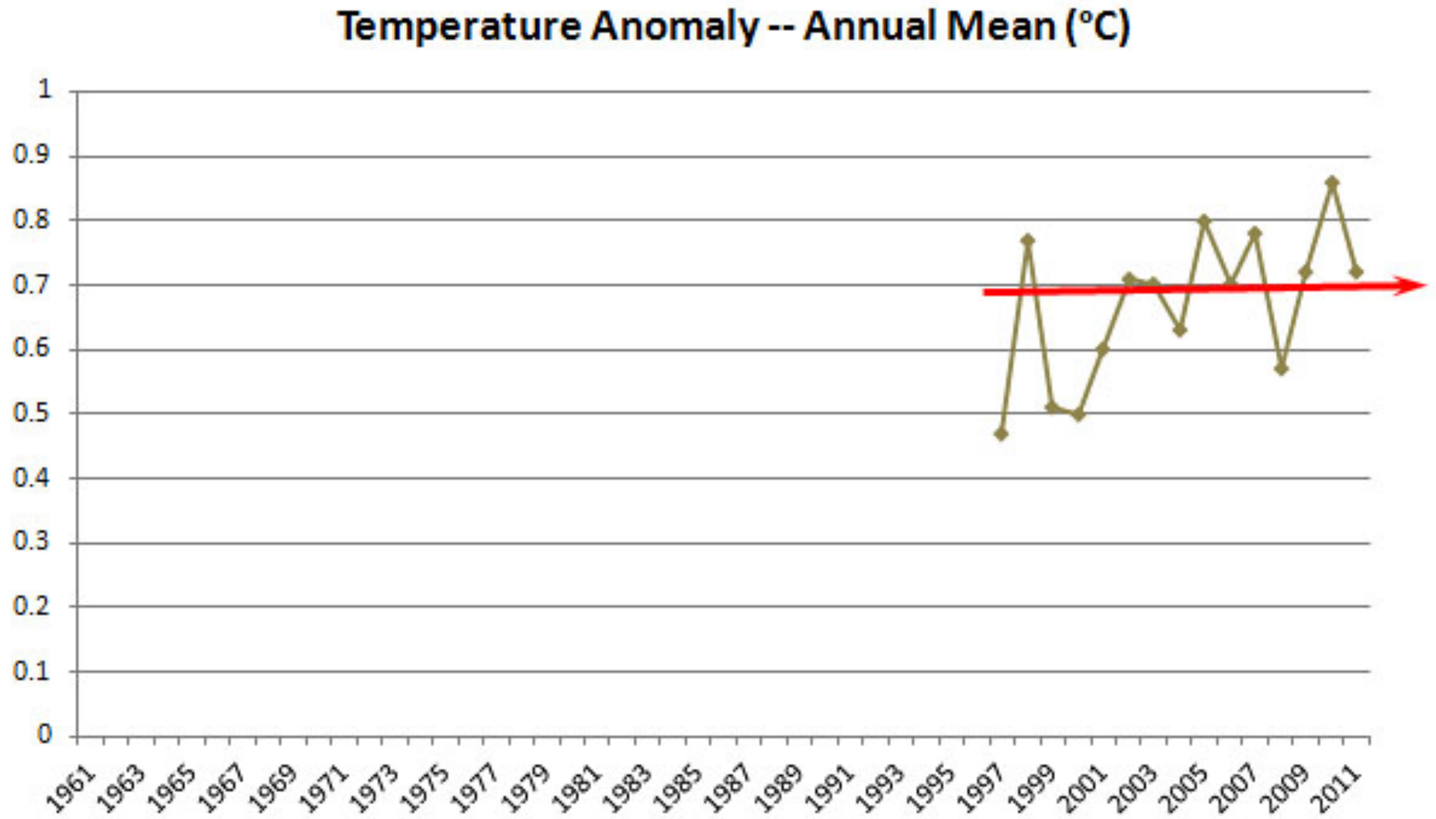
When showing something, make sure that you’re faithful to the data

Global Warming?



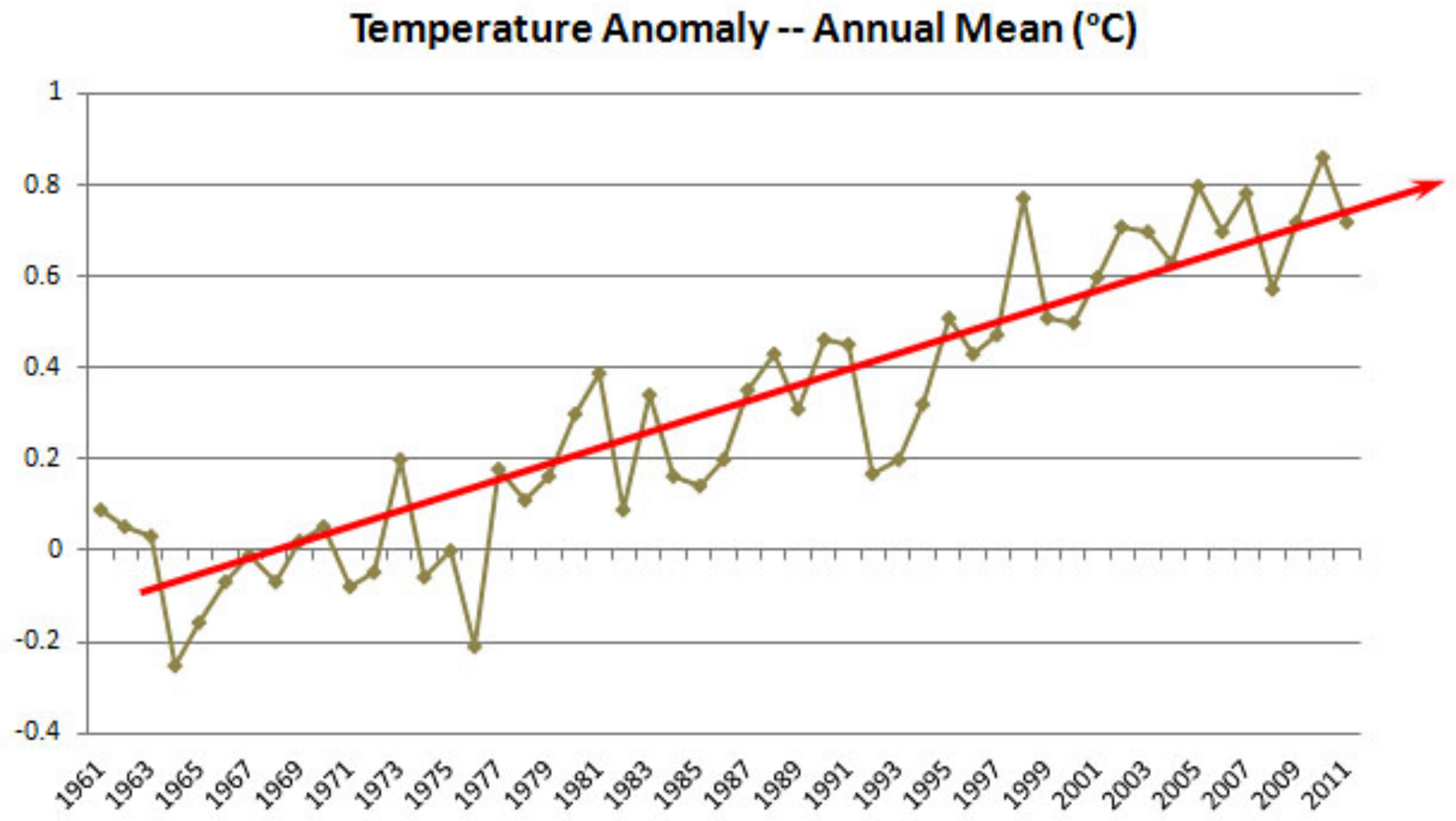
The Daily Mail, UK, Jan 2012

Global Warming?



Mother Jones

Global Warming - Frame the Data



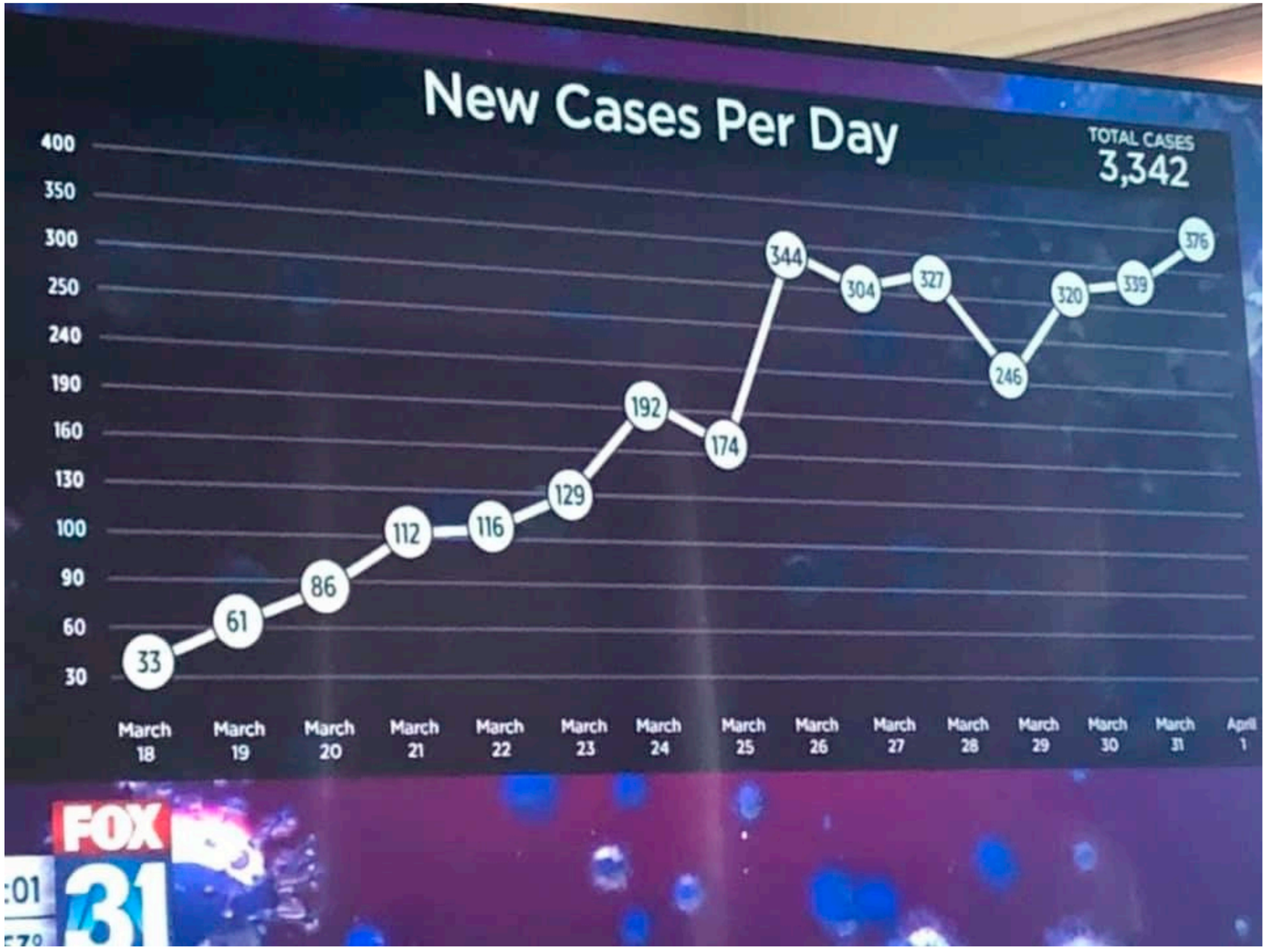
Also see: USA Temperature: can I sucker you?

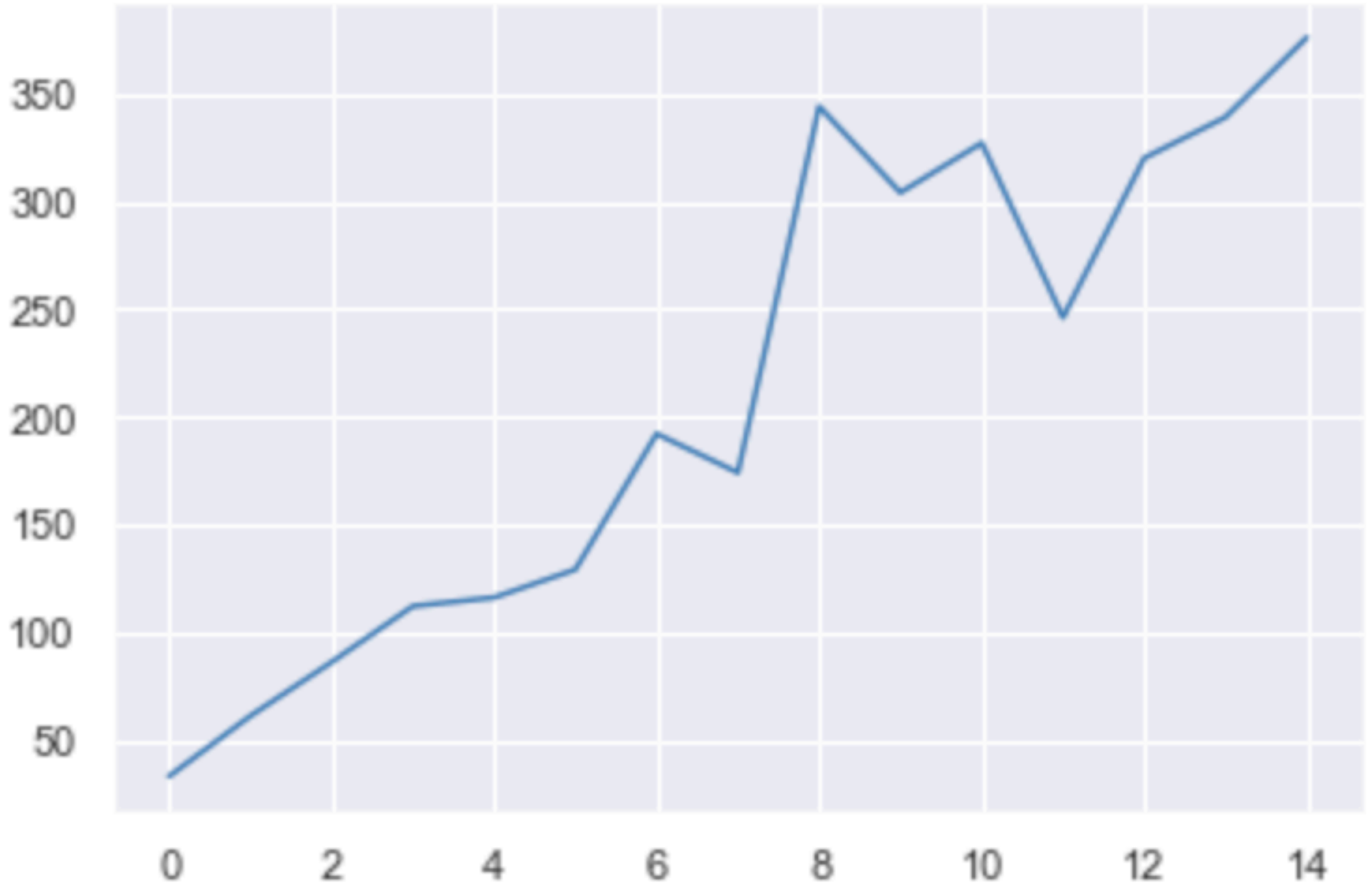
Mother Jones

What's wrong?

How 2012 Stacks Up
The Warmest Years on Record
Contiguous U.S.
55.3
54.31
54.32
54.14
53.93
53.81
55.5°
55.0°
54.5°
54.0°
53.5°
53.0°
1921
1999
1934
2006
1998
2012
Source: NOAA's National Climatic Data Center - State of the Climate National Overview
CLIMATE CENTRAL

Scale Distortions in Temporal Data


New Cases Per Day
TOTAL CASES
3,342
400
350
300
250
240
190
160
130
100
90
60
30
33
61
86
112
116
129
192
174
344
304
327
246
320
339
376
March 18
March 19
March 20
March 21
March 22
March 23
March 24
March 25
March 26
March 27
March 28
March 29
March 30
March 31
April 1
FOX
31


350
300
250
200
150
100
50
0
2
4
6
8
10
12
14

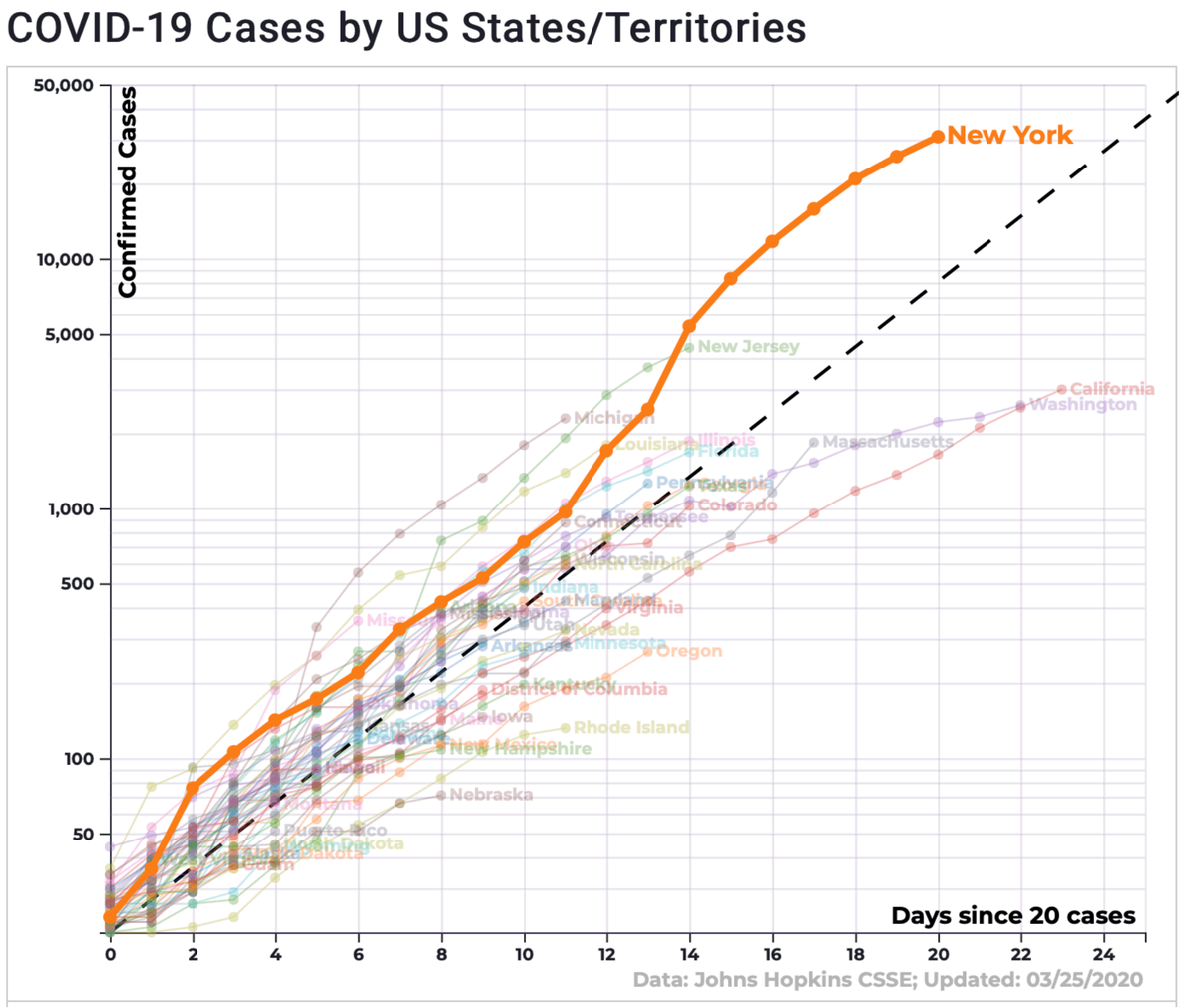
Log Scales

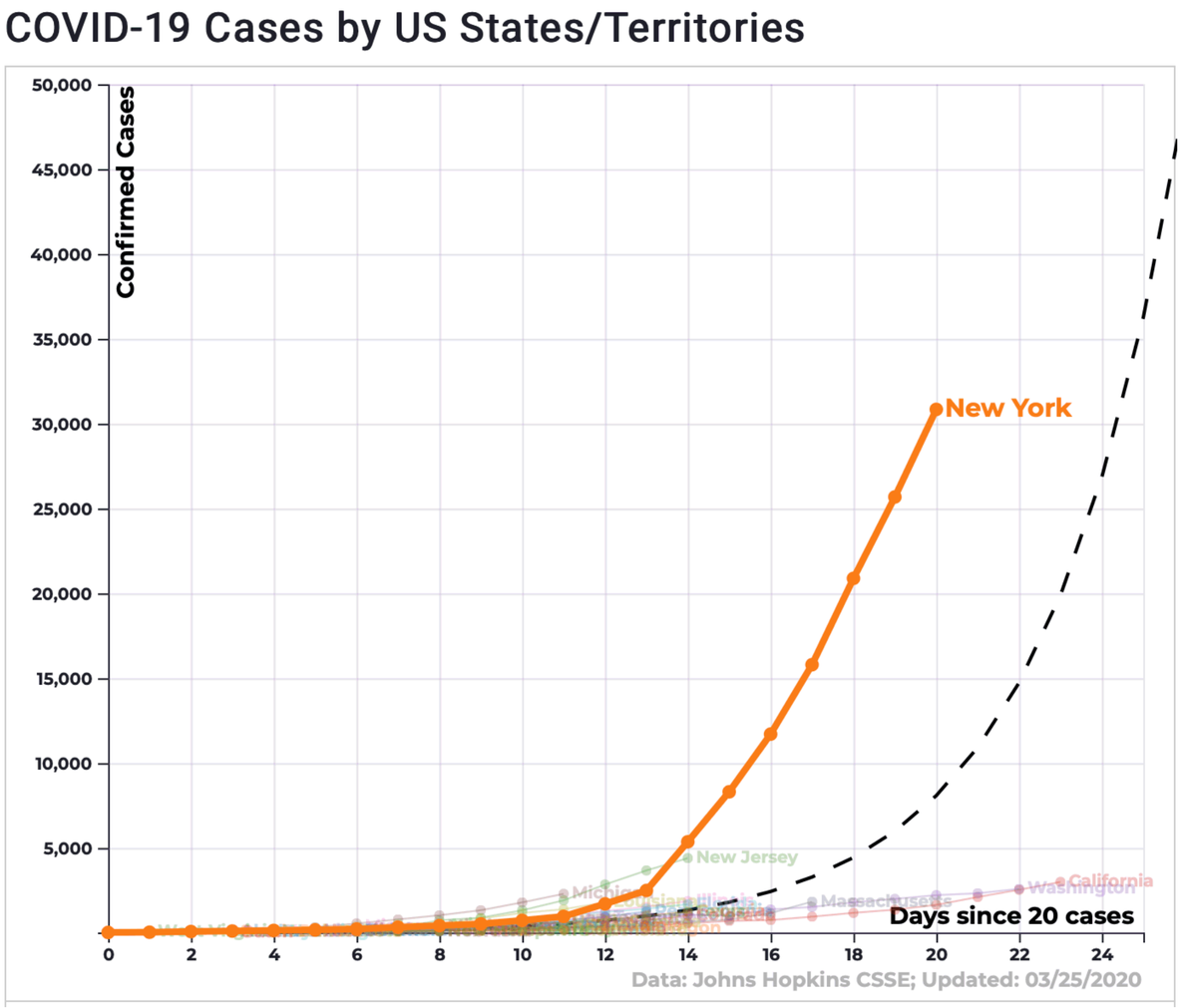
Use log scales if the underlying data warrants it

Typical use case: exponential growth curves

In practice: an expert tool

What are some interpretations?





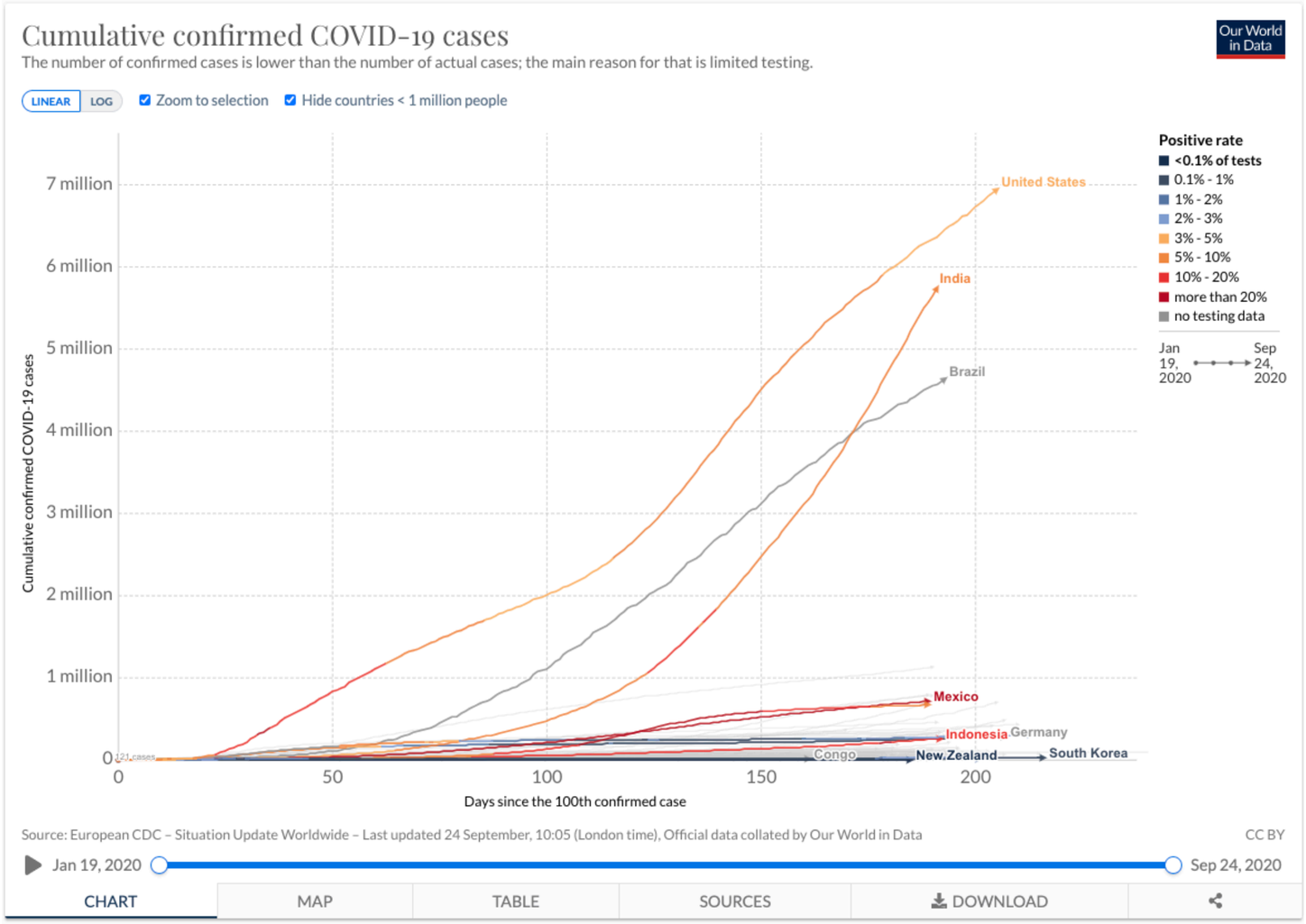
https://twitter.com/nothingelseis/status/1243203992848457733

Normalization

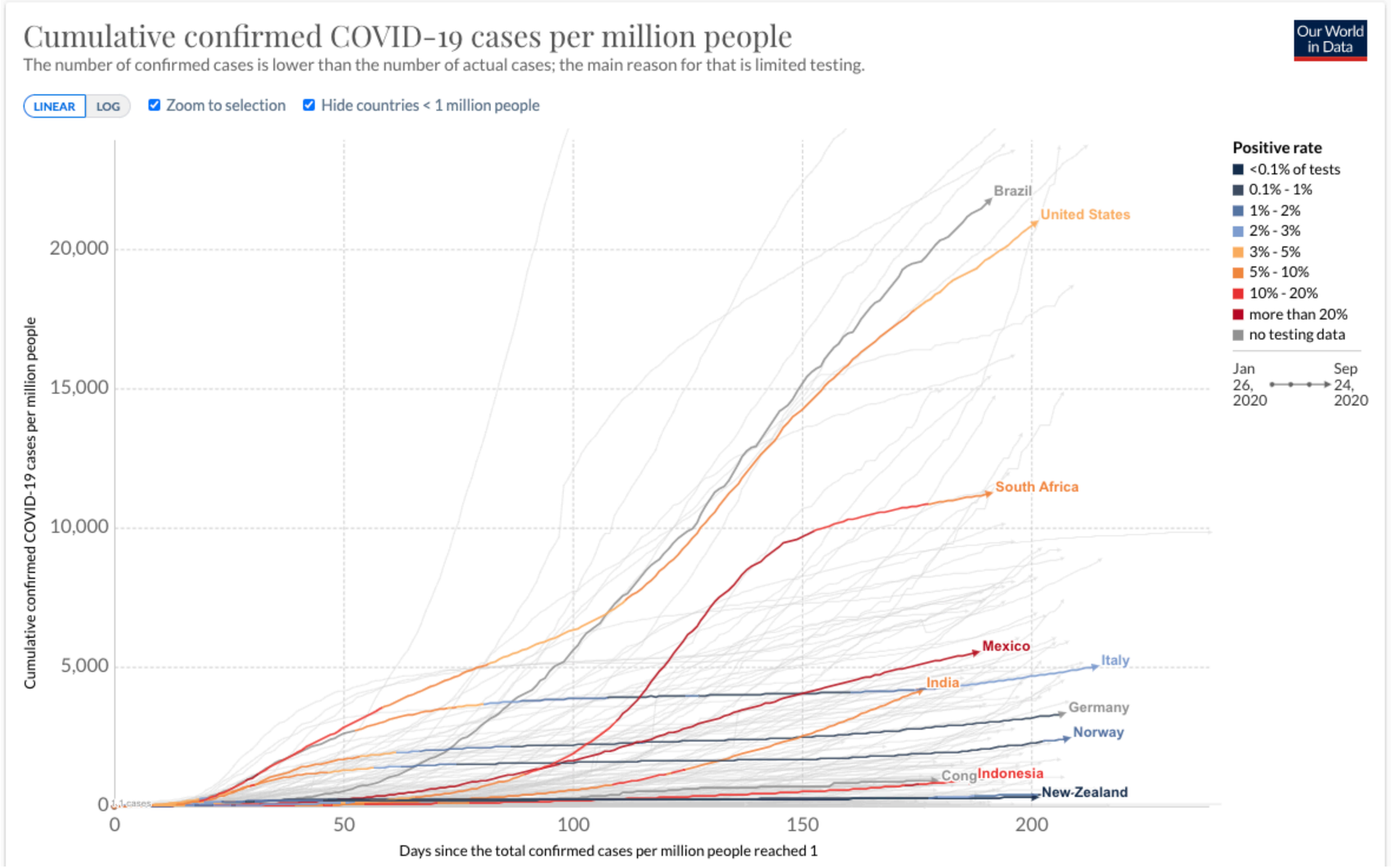
Comparing Apples to Apples

When we compare things that are different, we need to account for that difference. Normalize your data!

Cumulative Cases



Cumulative Cases Per Million



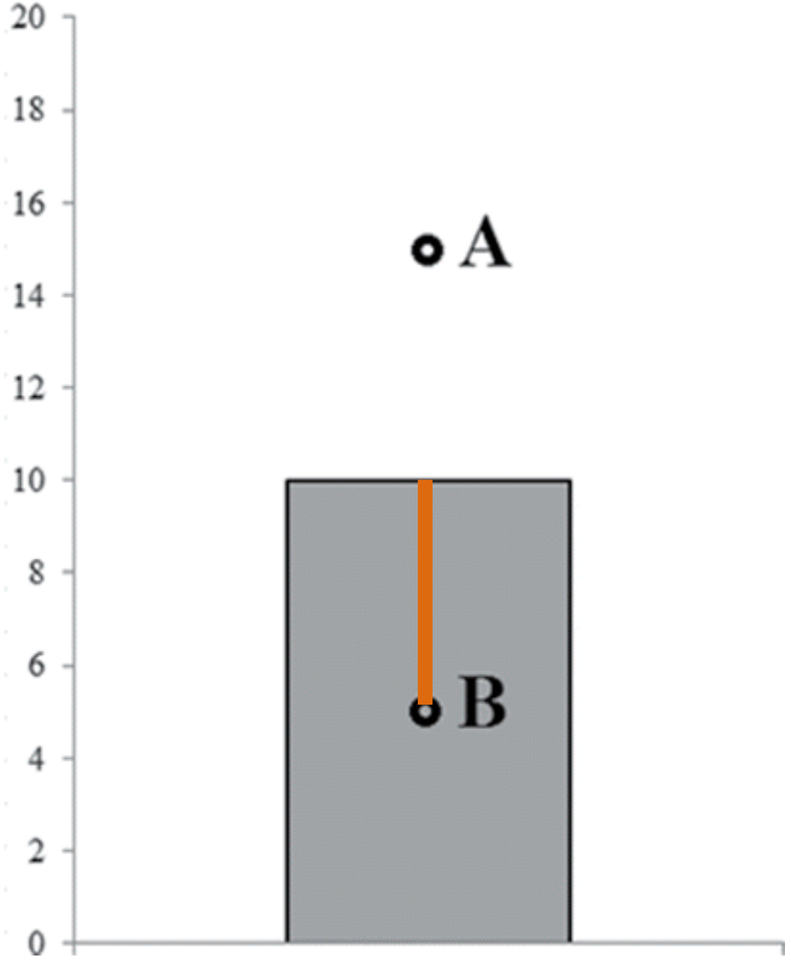
Different Perspectives

To get the full picture, you might look at more than one chart: https://ourworldindata.org/coronavirus

Distributions

Height of the Bar encodes mean of a distribution

Which value is more likely to belong to the distribution?
A or B?

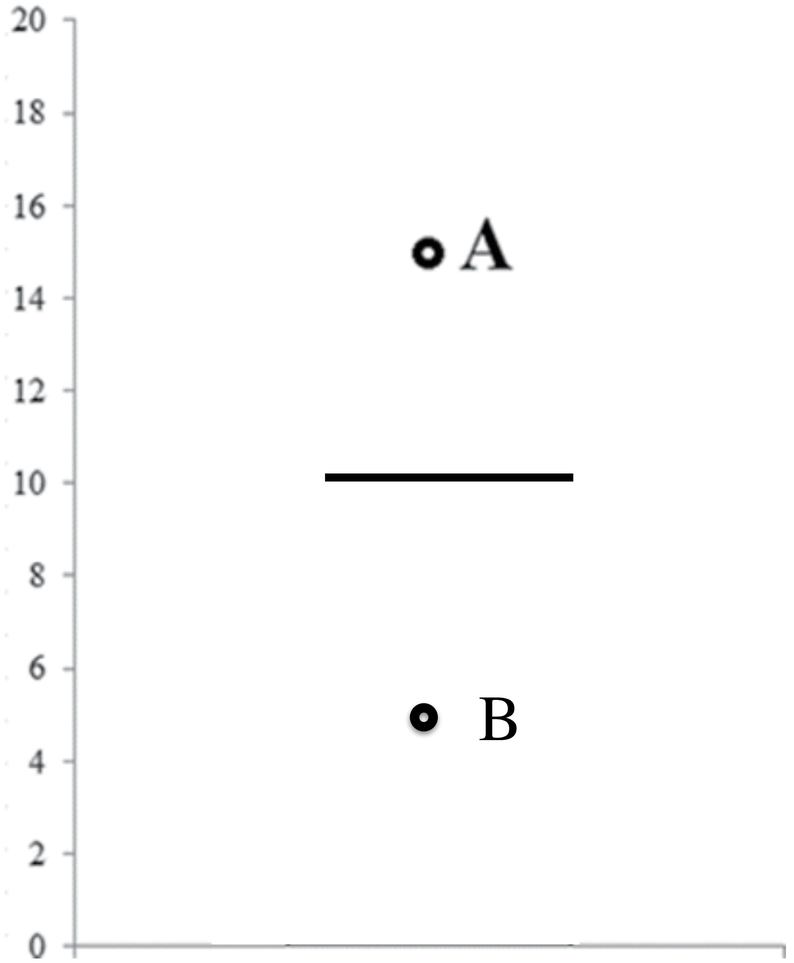


http://www.tandfonline.com/doi/full/10.1080/00031305.2016.1141706

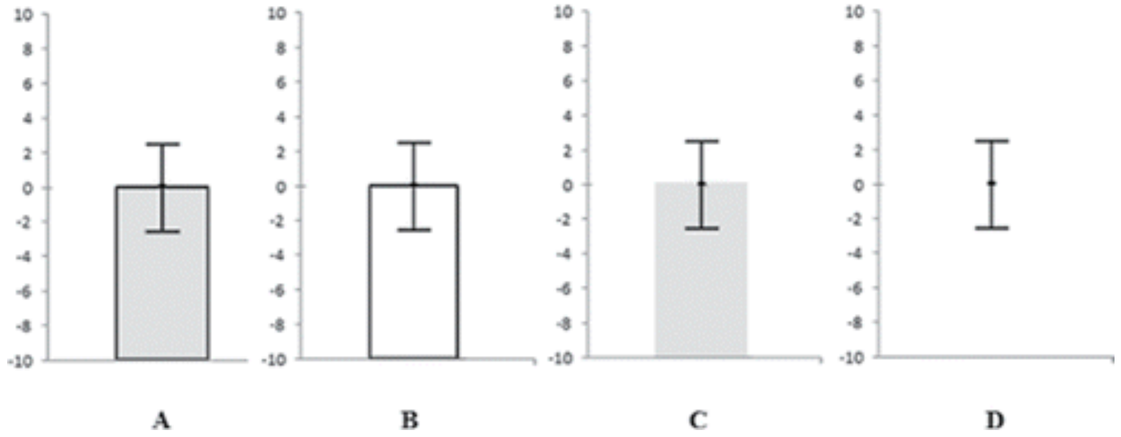
Biases

We can plot the data faithfully, but still perceive it wrongly!

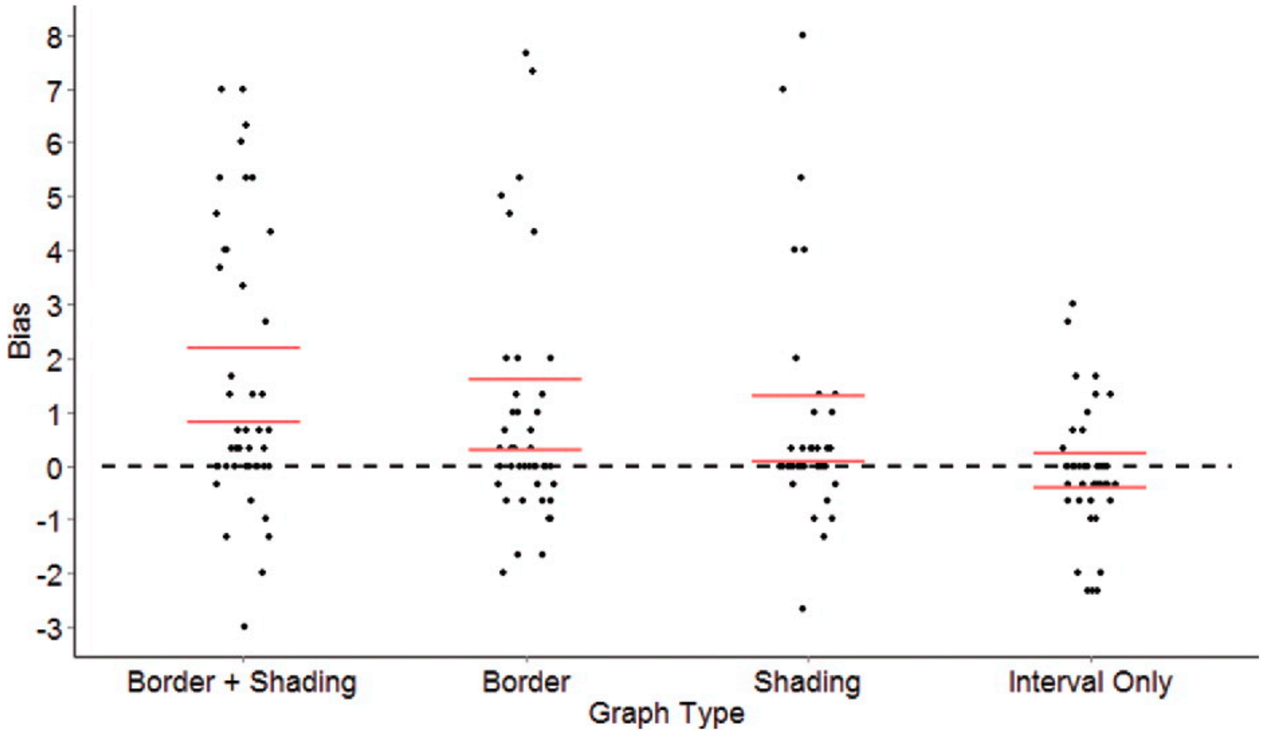
What about now?



Within the Bar Bias



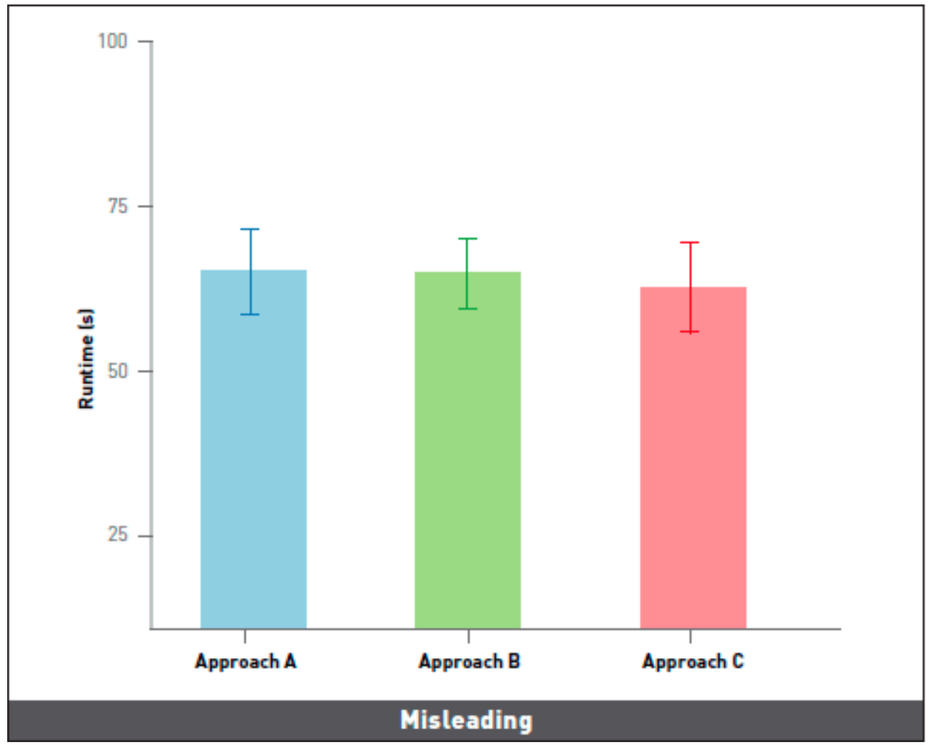
Experimental Conditions

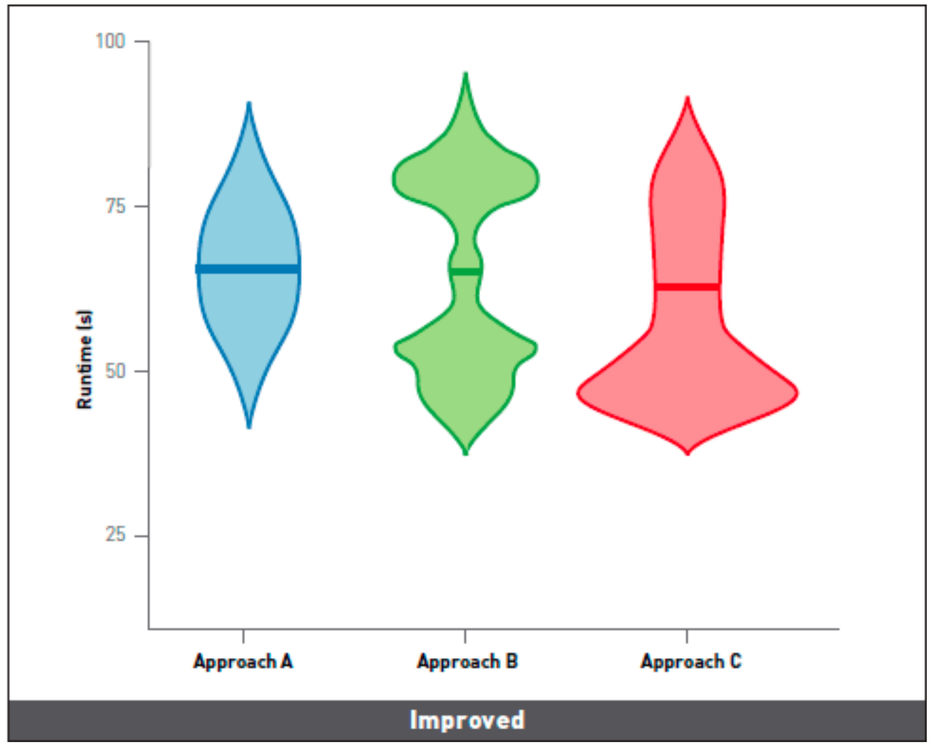


Results

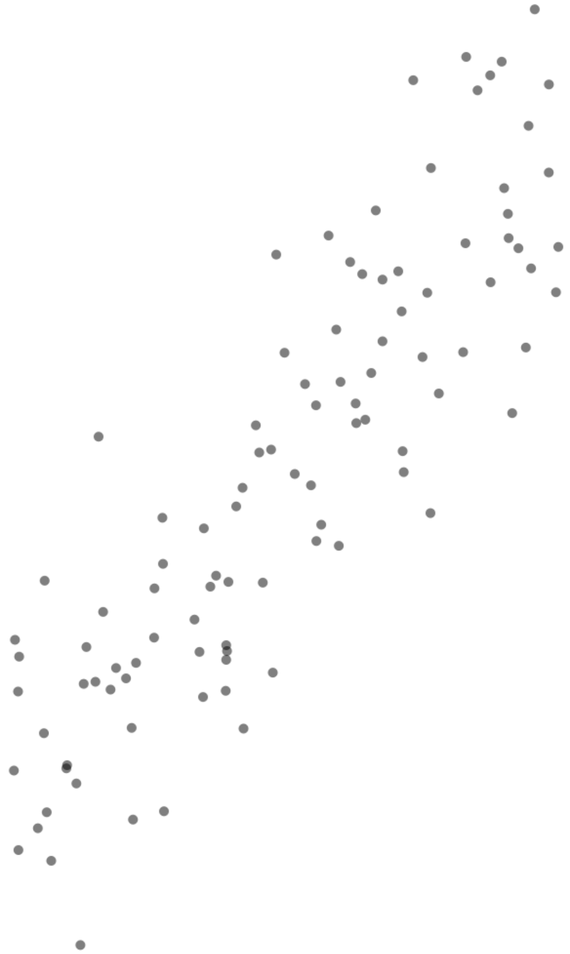
Christopher S. Pentoney & Dale E. Berger (2016) Confidence Intervals and the Within-the-Bar Bias, The American Statistician, 70:2, 215-220

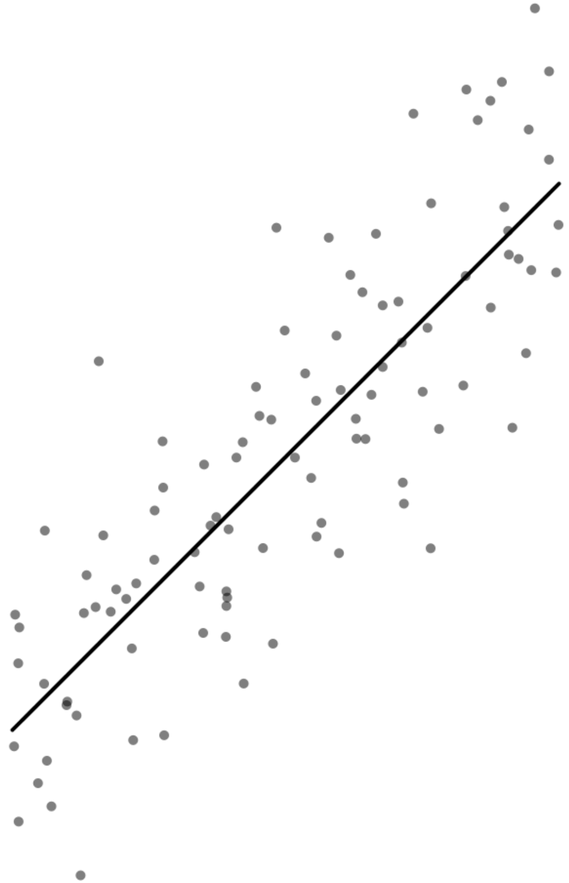
Careful when designing aggregated charts



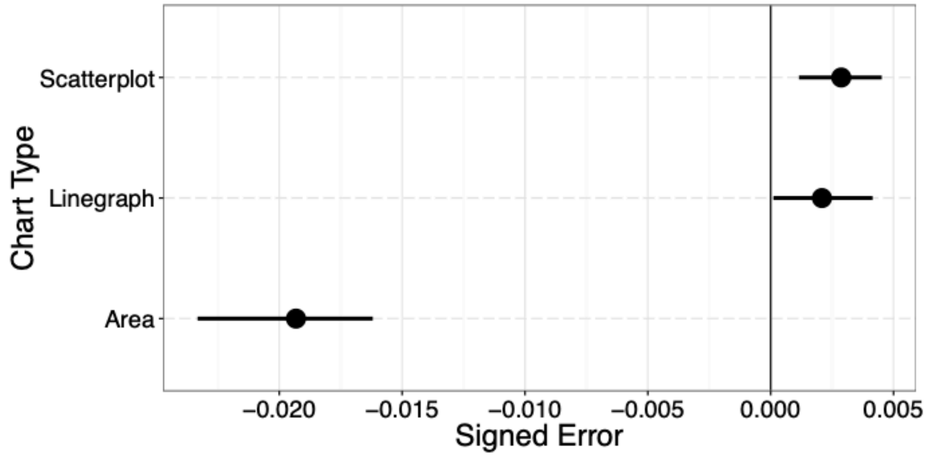


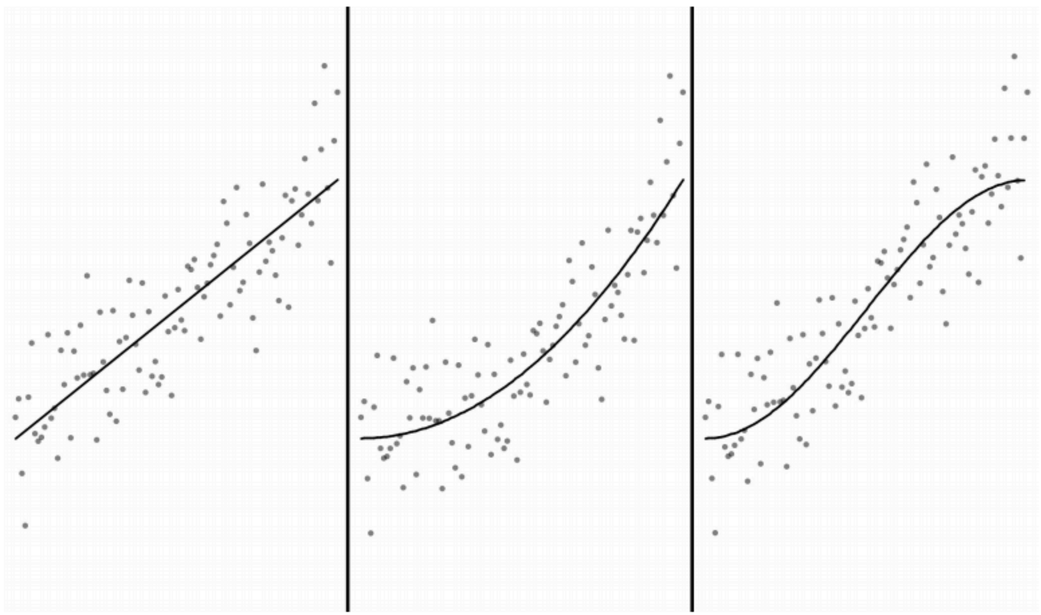
What’s the Trendline?

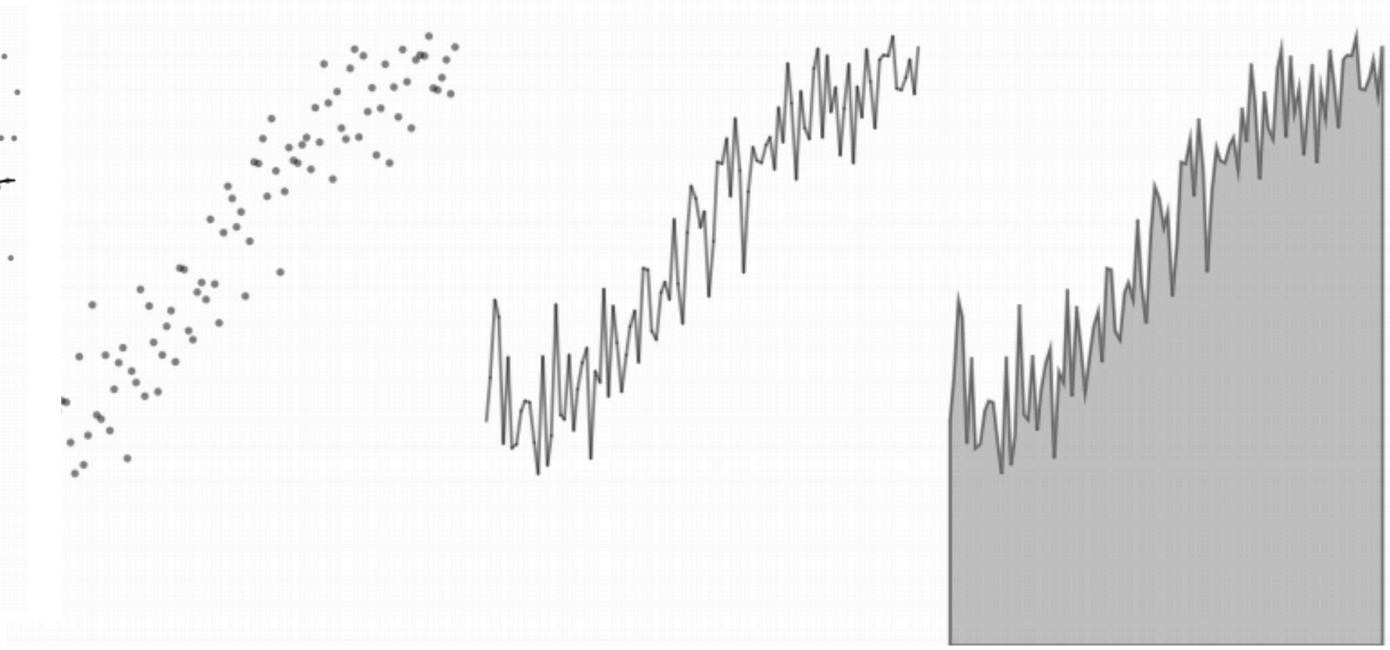




Regression by eye







We’re good at spotting trends

But the wrong vis technique can deceive us

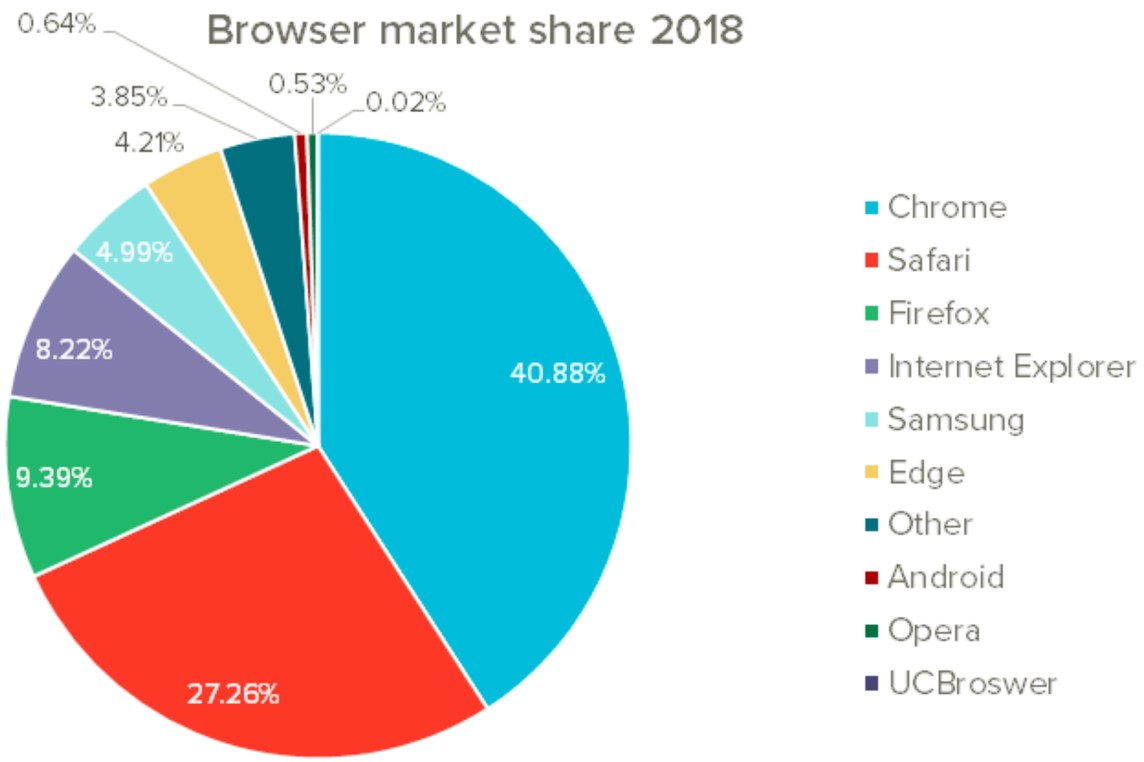
[Corell & Heer, 2017]

http://idl.cs.washington.edu/files/2017-RegressionByEye-CHI.pdf

Pie Charts

Why Pie Charts?

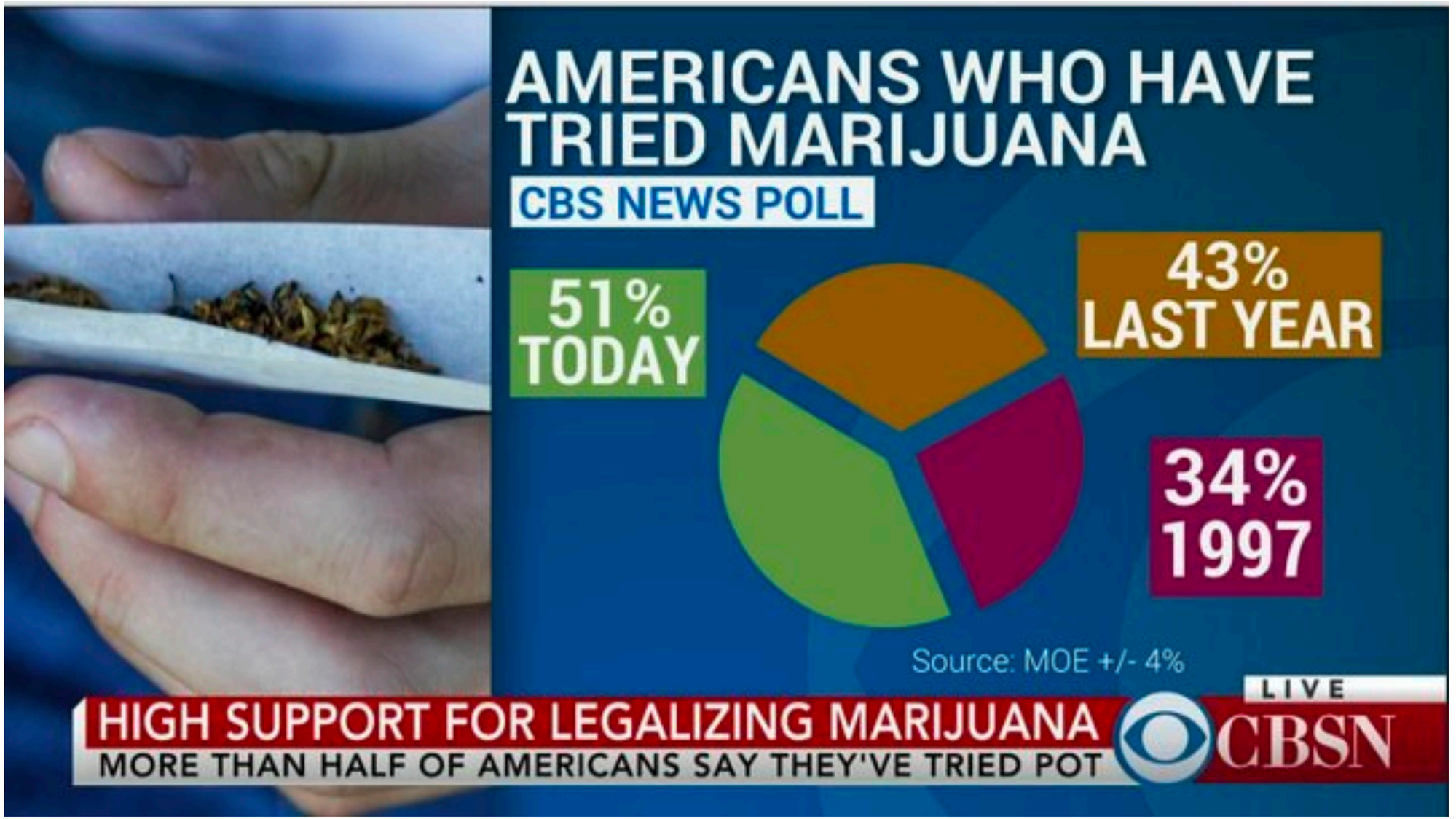
Show Part-of-Whole Relationships



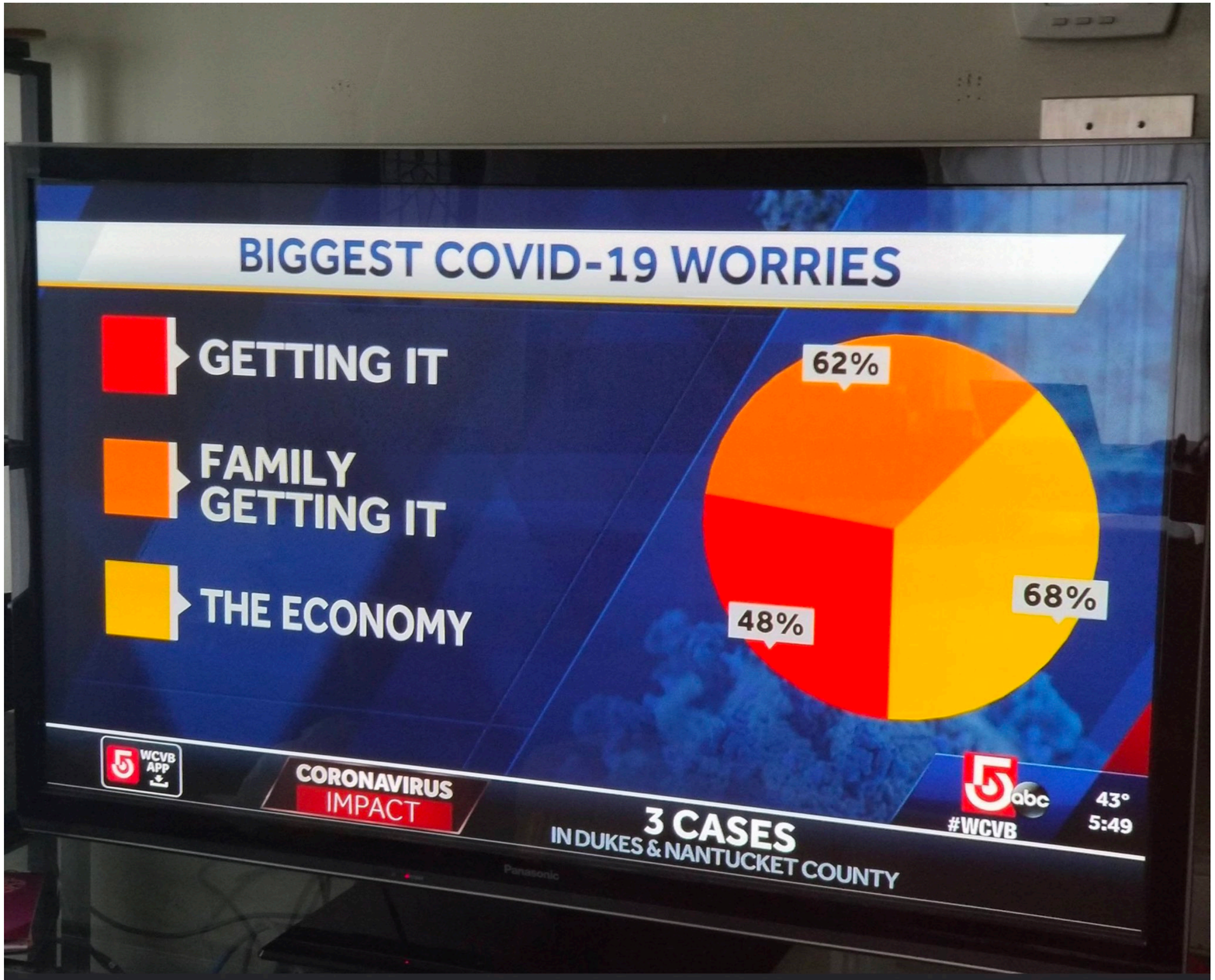
https://blog.uptrends.com/uptrends-research/browser-market-share-2018/

How can we make this better?

- Label the wedges directly, get rid of color scale
- Fewer segments: put more into “other”
- Make sure labels have contrast



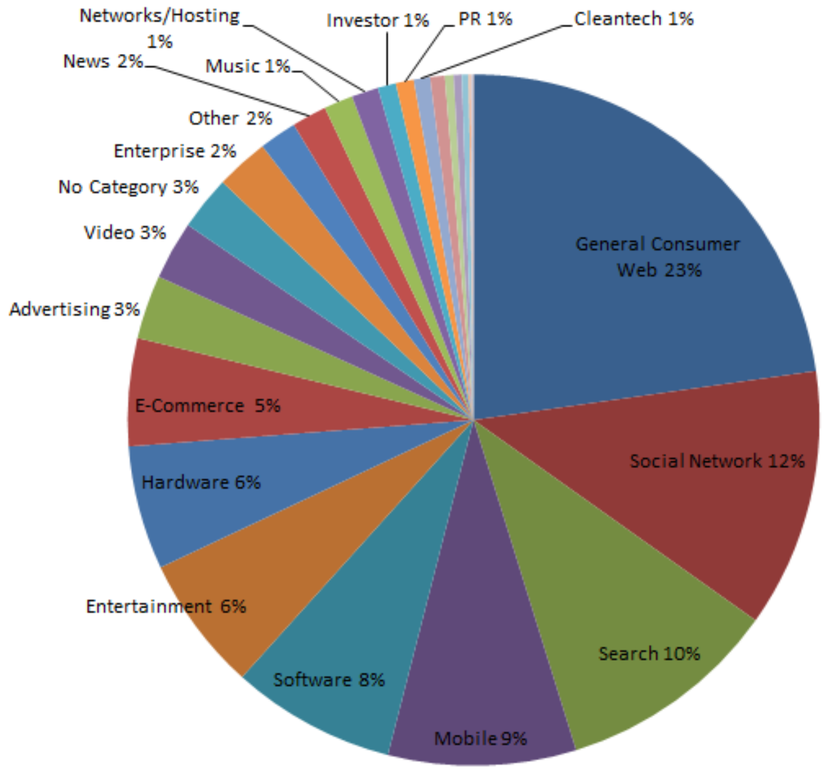
https://twitter.com/K_Graves/status/1118927857214873600



Posted in r/**funny** by u/**veganator**

reddit

Death to Pie Charts



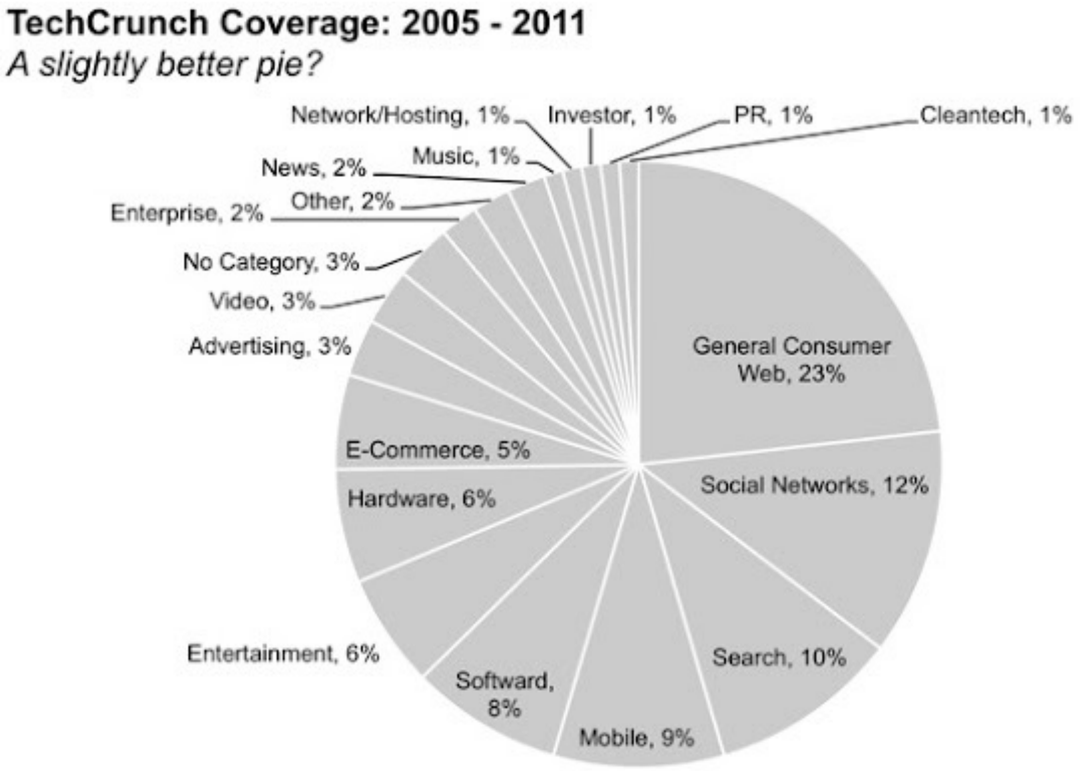
Share of coverage on TechCrunch

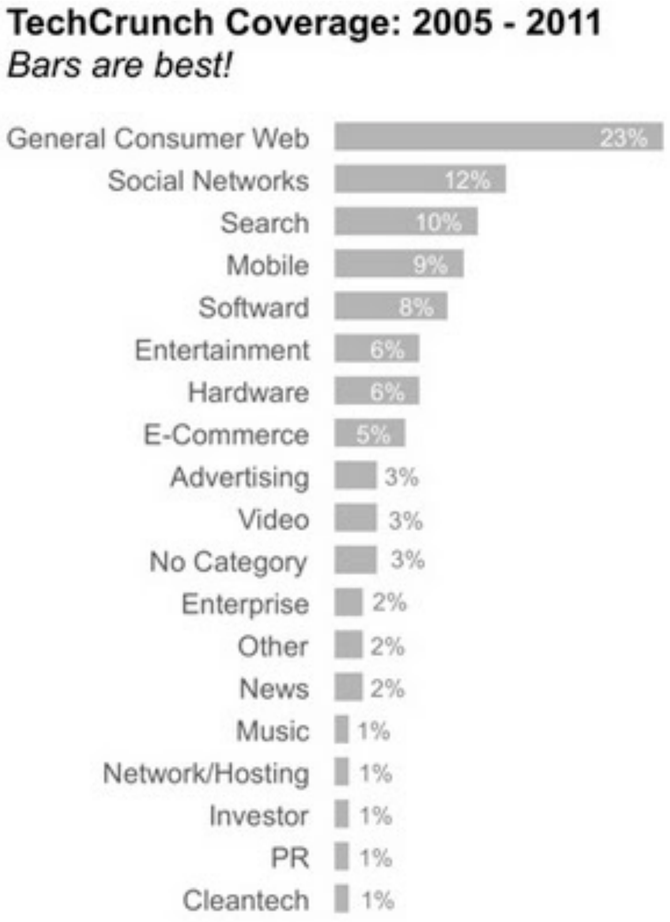
"I hate pie charts.
I mean, really hate them."

www.storytellingwithdata.com/2011/07/death-to-pie-charts.html

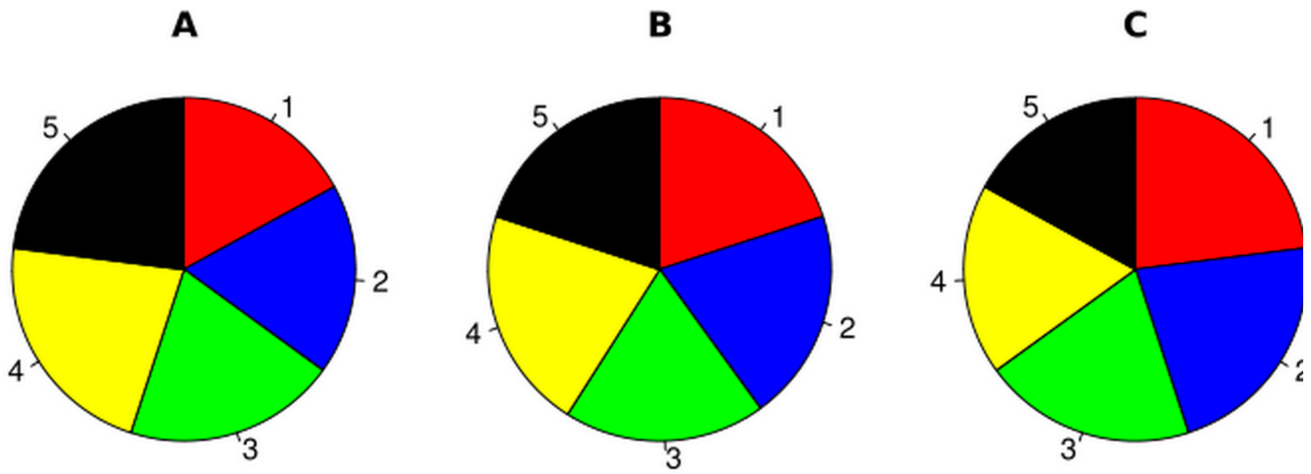
Cole Nussbaumer

Redesign

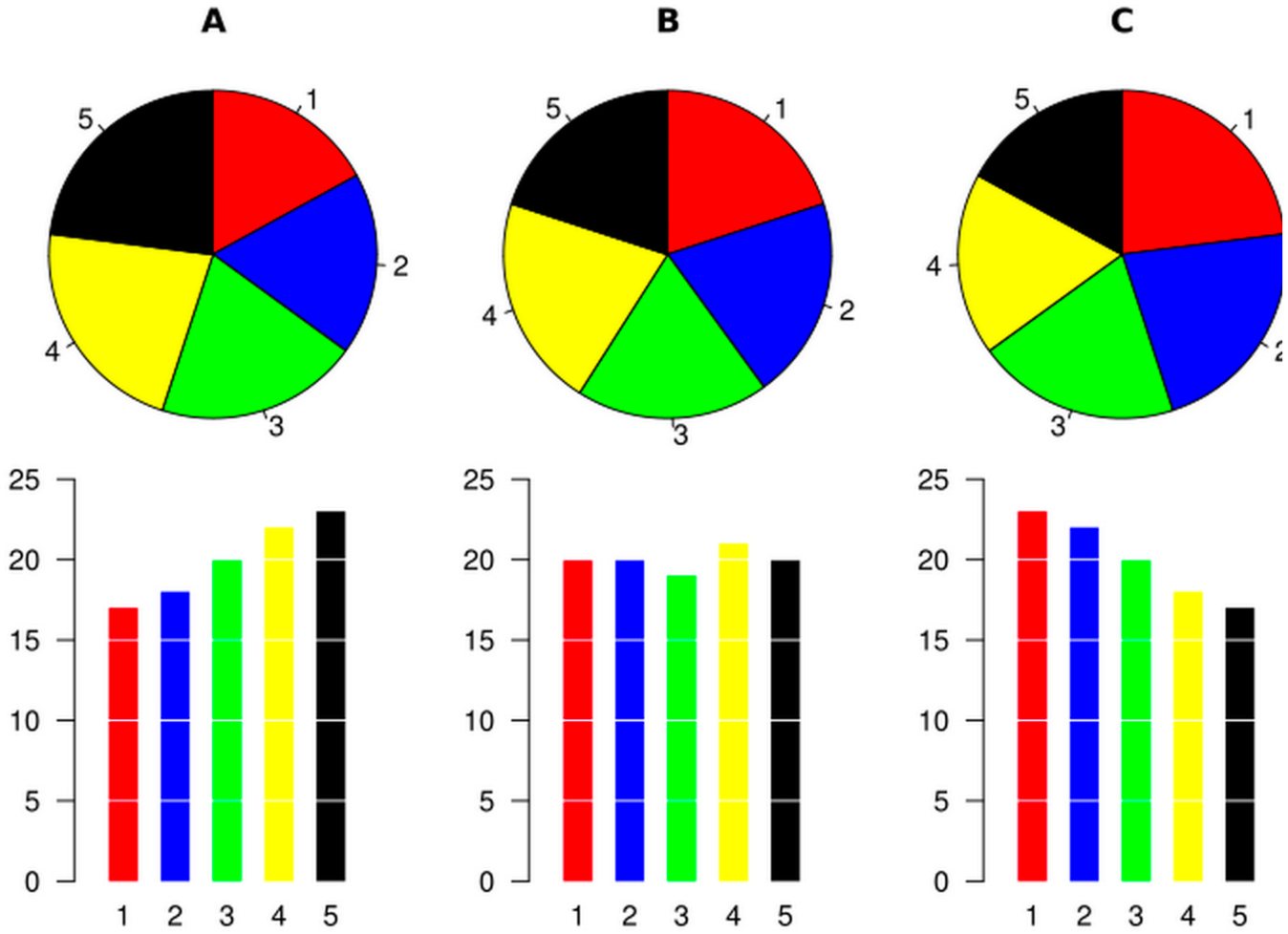

TechCrunch Coverage: 2005 - 2011
A slightly better pie?
Network/Hosting, 1%
Investor, 1%
PR, 1%
Cleantech, 1%
News, 2%
Music, 1%
Enterprise, 2%
Other, 2%
No Category, 3%
Video, 3%
Advertising, 3%
General Consumer Web, 23%
E-Commerce, 5%
Social Networks, 12%
Hardware, 6%
Entertainment, 6%
Softward, 8%
Search, 10%
Mobile, 9%


TechCrunch Coverage: 2005 - 2011
Bars are best!
General Consumer Web 23%
Social Networks 12%
Search 10%
Mobile 9%
Softward 8%
Entertainment 6%
Hardware 6%
E-Commerce 5%
Advertising 3%
Video 3%
No Category 3%
Enterprise 2%
Other 2%
News 2%
Music 1%
Network/Hosting 1%
Investor 1%
PR 1%
Cleantech 1%

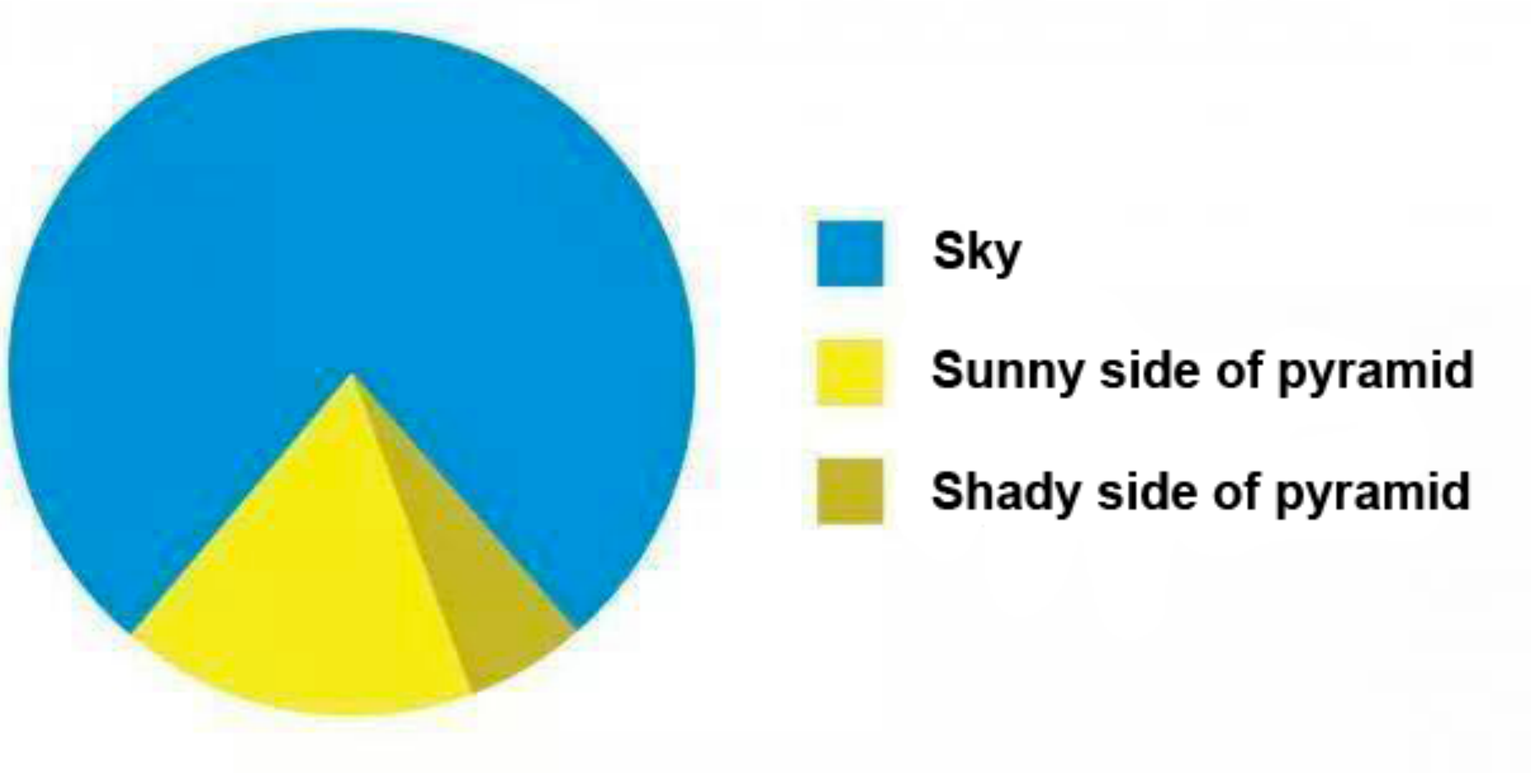
Can you spot the differences?



Can you spot the differences?

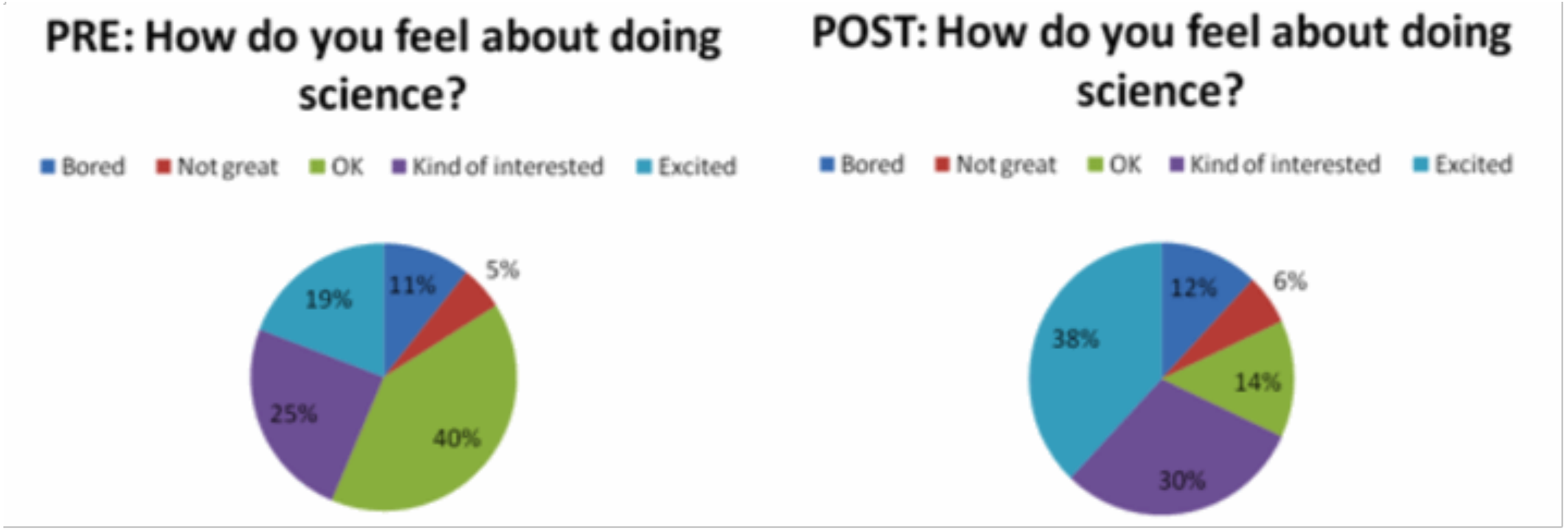


My favorite pie chart



My second favorite pie chart

So, what to use instead?

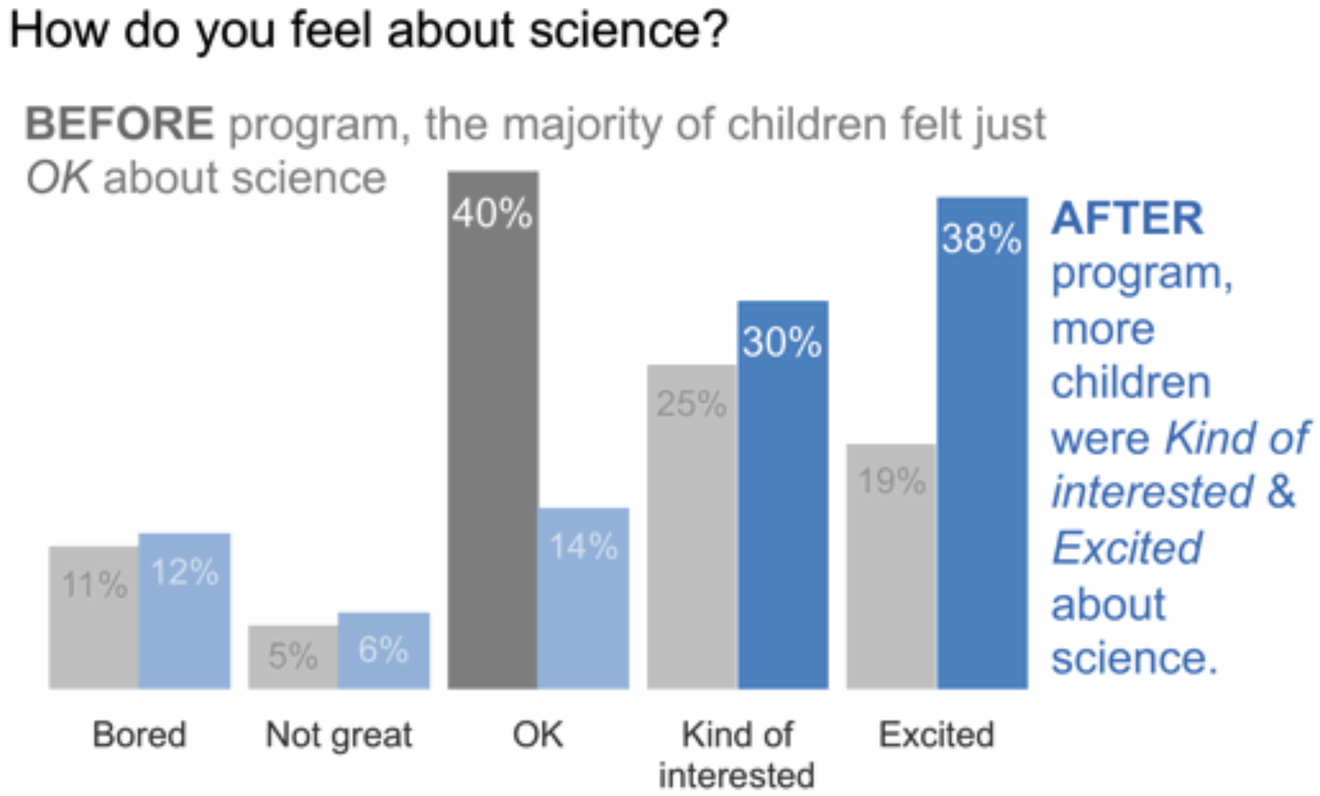


imagine you just completed a pilot summer learning program on science aimed at improving perceptions of the field among 2nd and 3rd grade elementary children

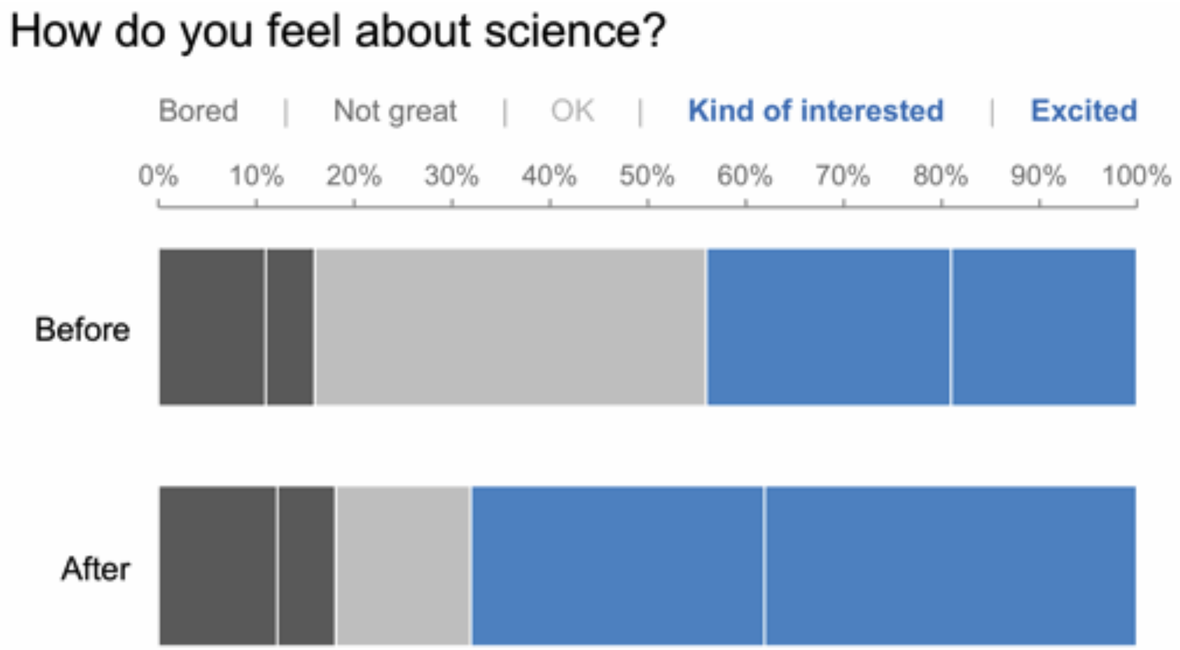
http://www.storytellingwithdata.com/blog/2014/06/alternatives-to-pies

Alternative #1: Show the Number(s) Directly

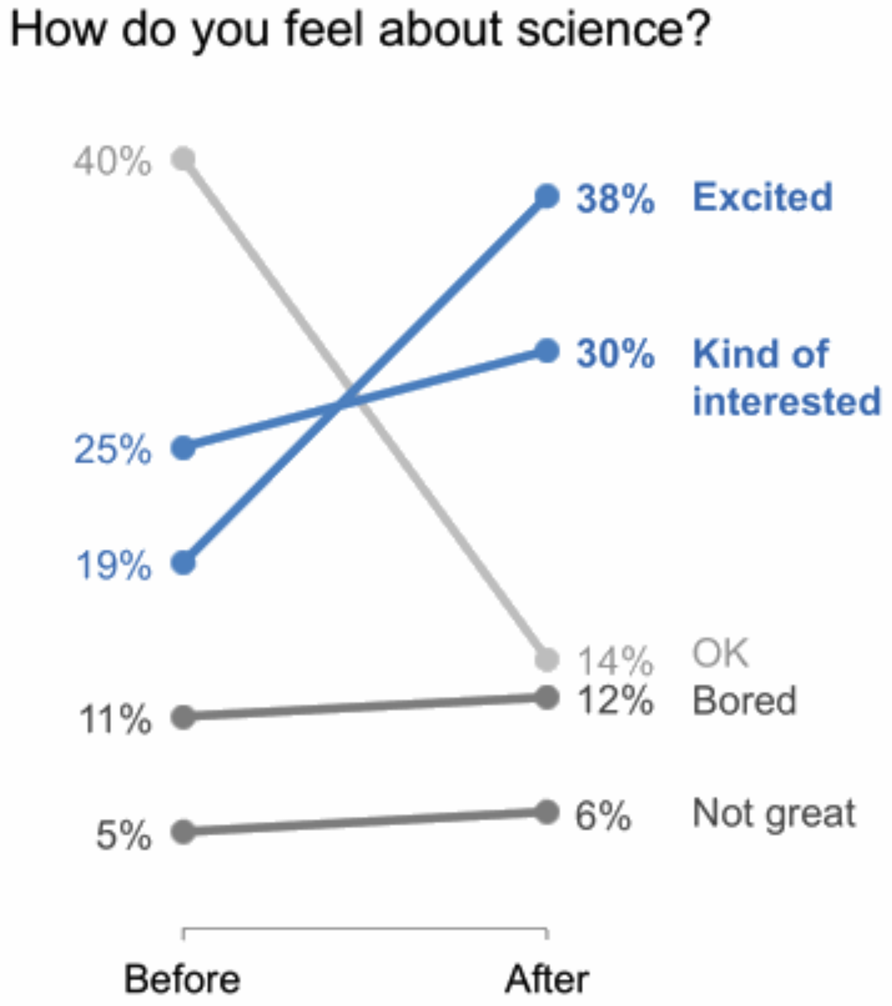
Alternative #2: Simple Bar Graph

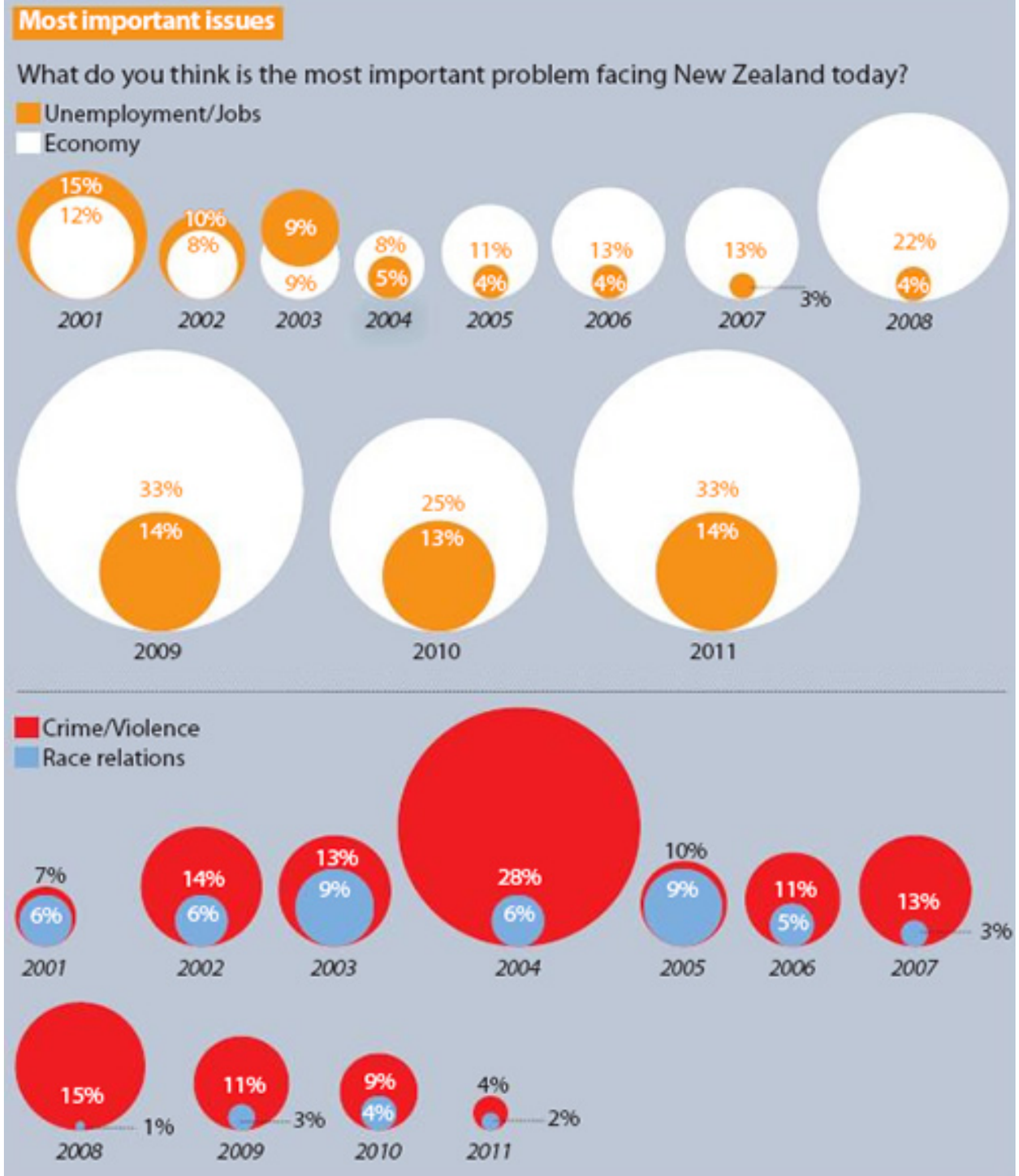


Alternative #3: 100% Stacked Horizontal Bar Graph



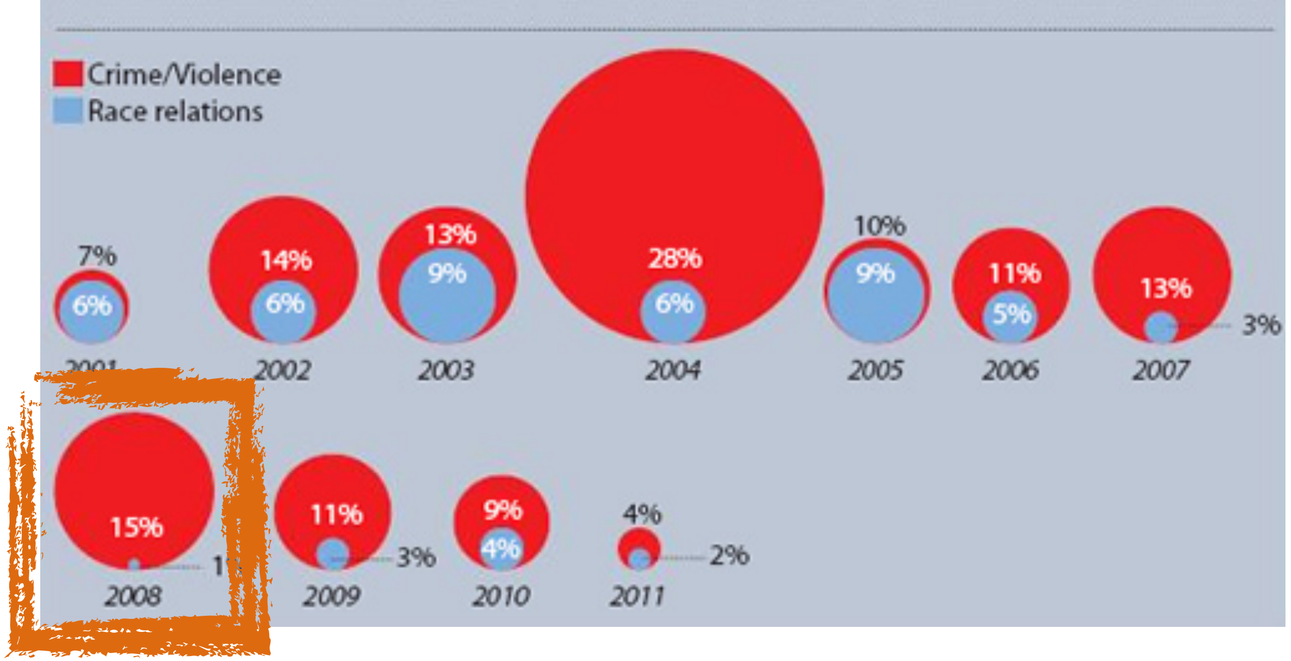
Alternative #4: Slopegraph





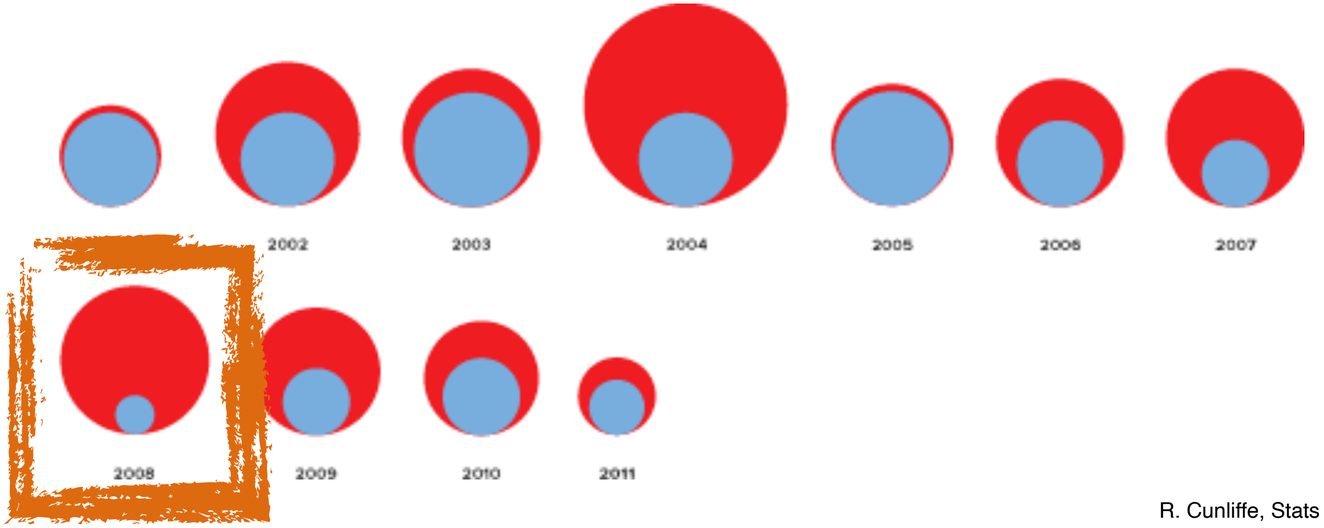
https://goo.gl/lHWp4x

Sunday Star Times, 2012

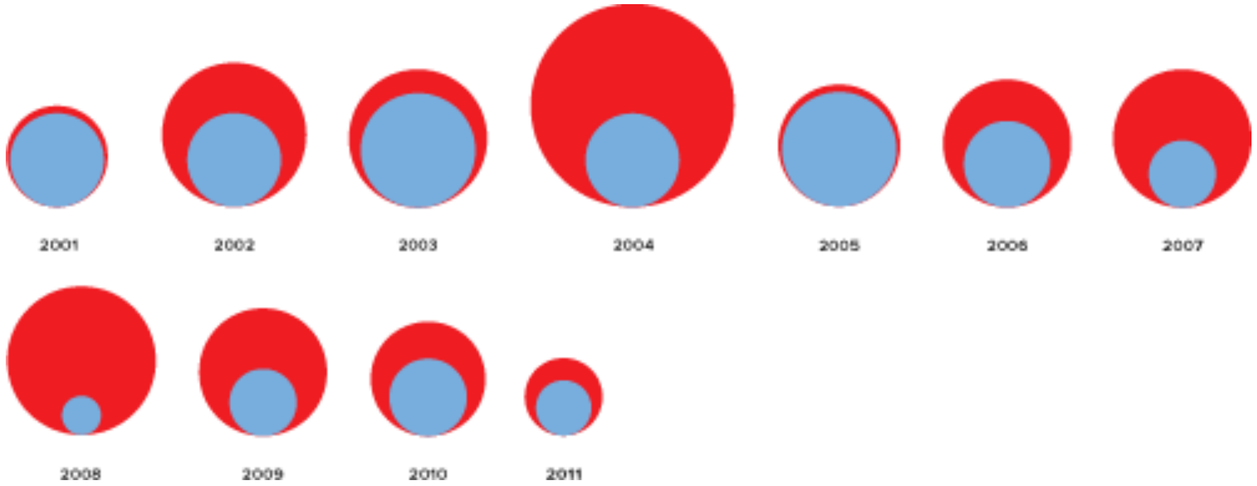


Quantity encoded by diameter, not area!

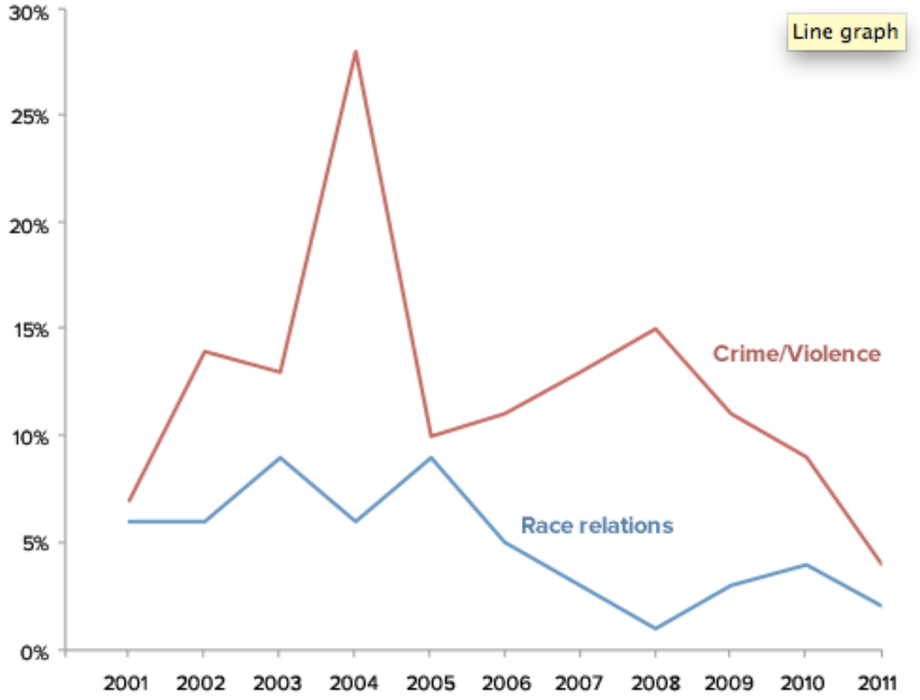
Fixing that:



R. Cunliffe, Stats Chat



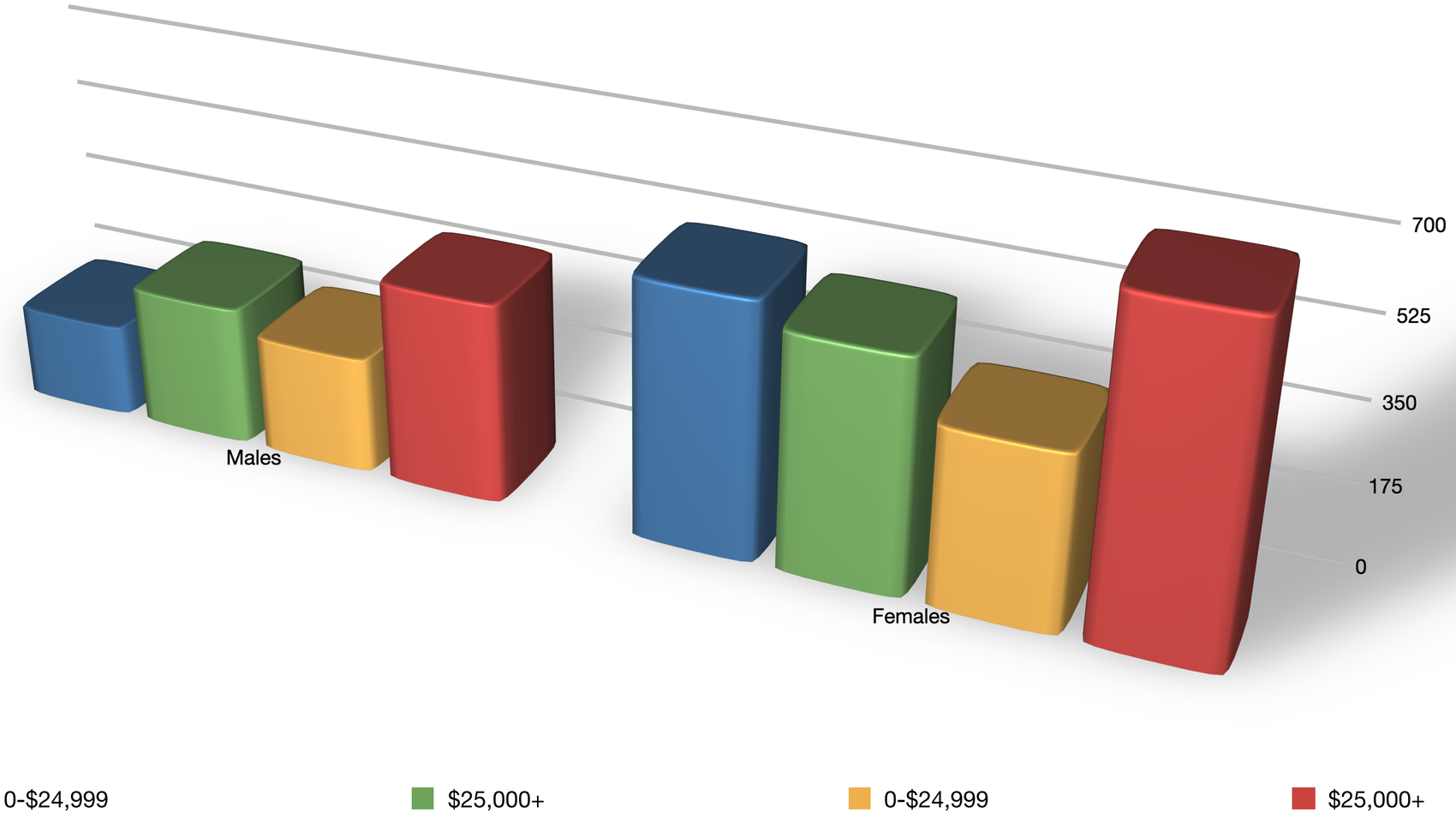
But is this visual encoding appropriate in the first place?



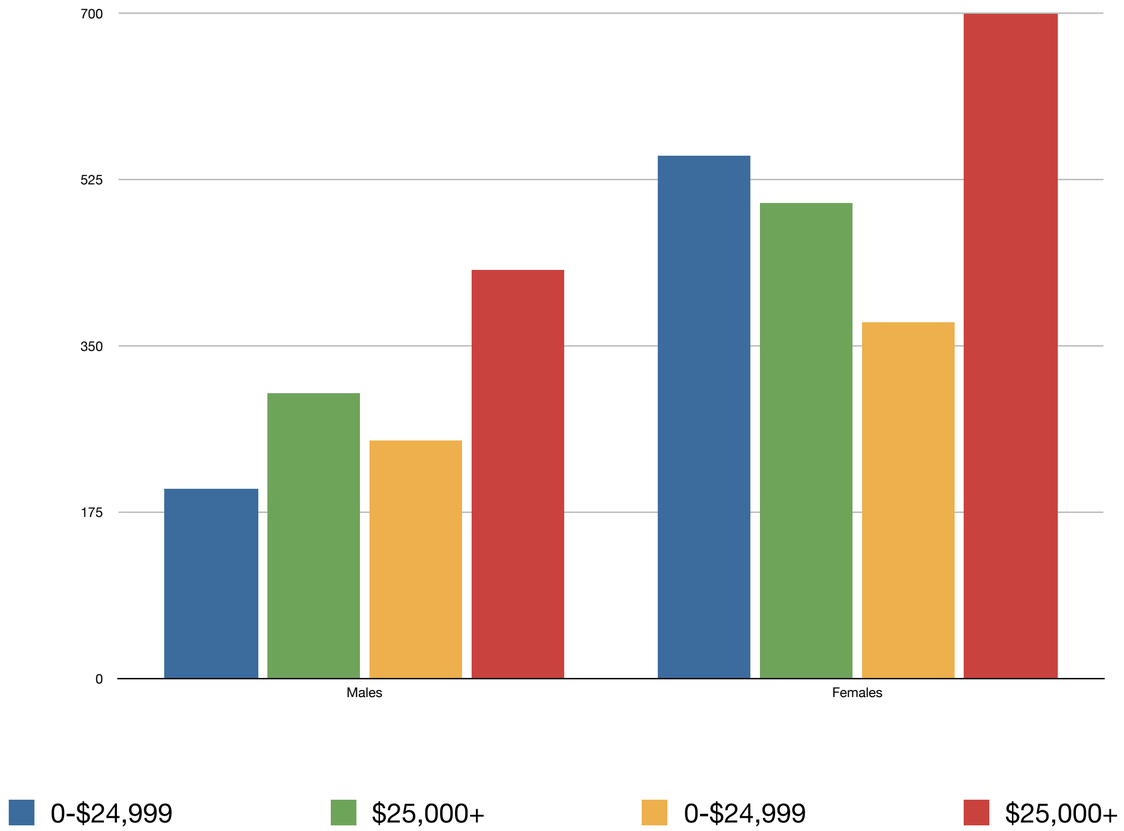
R. Cunliffe, Stats Chat

Clean vs Embellished

Maximize Data-Ink Ratio

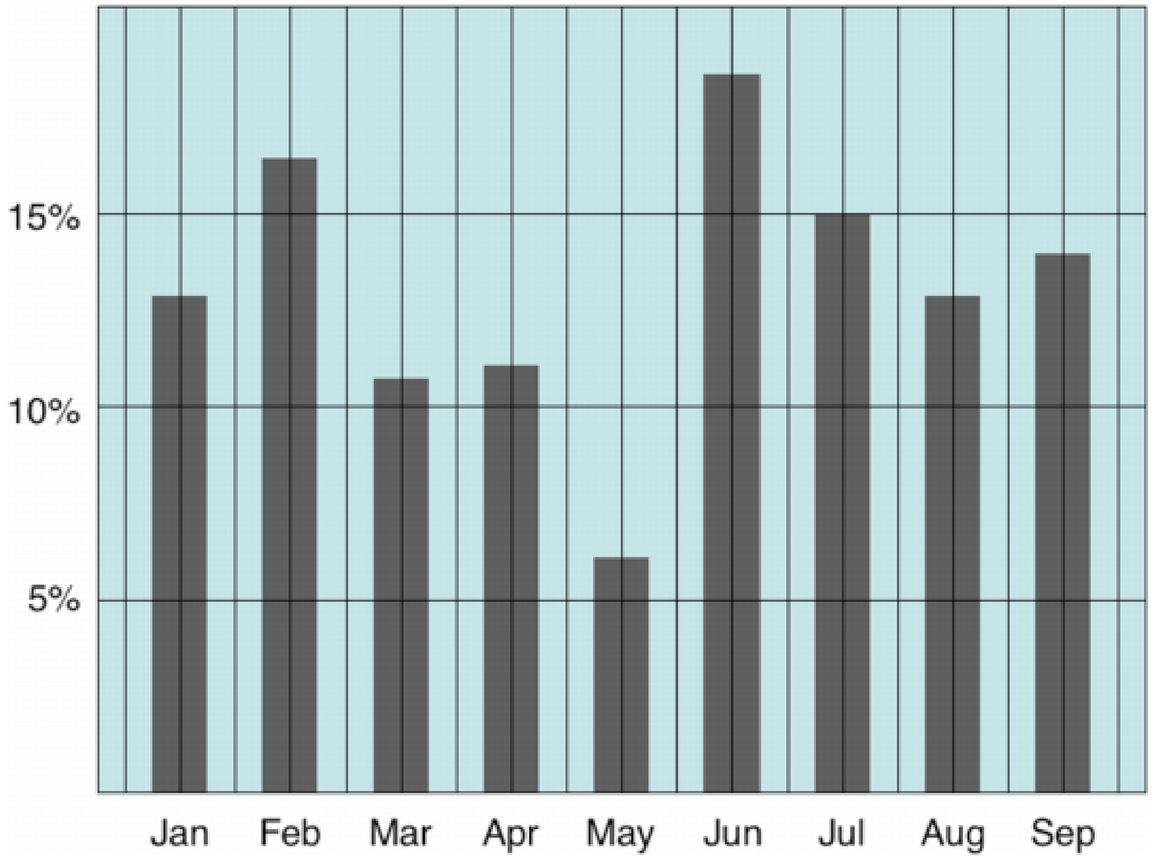


Maximize Data-Ink Ratio



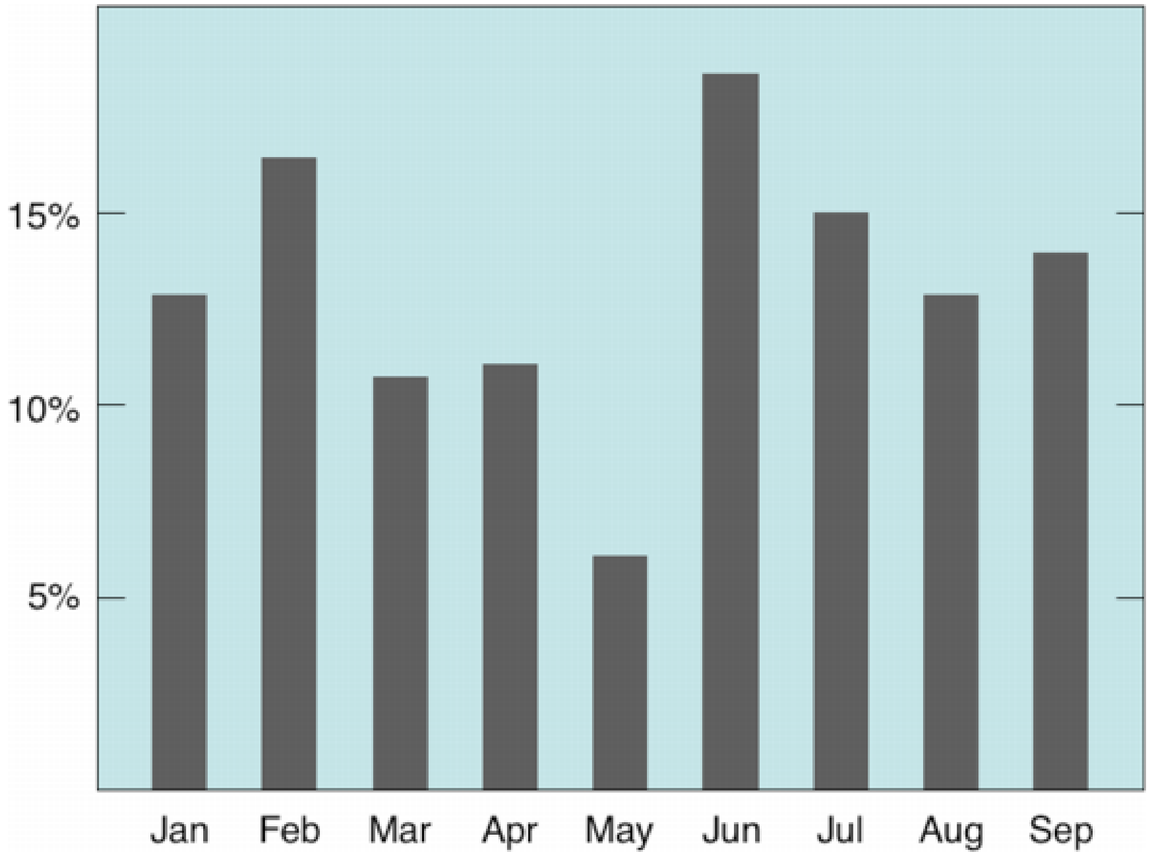
Avoid Chart Junk

Extraneous visual elements that distract from the message



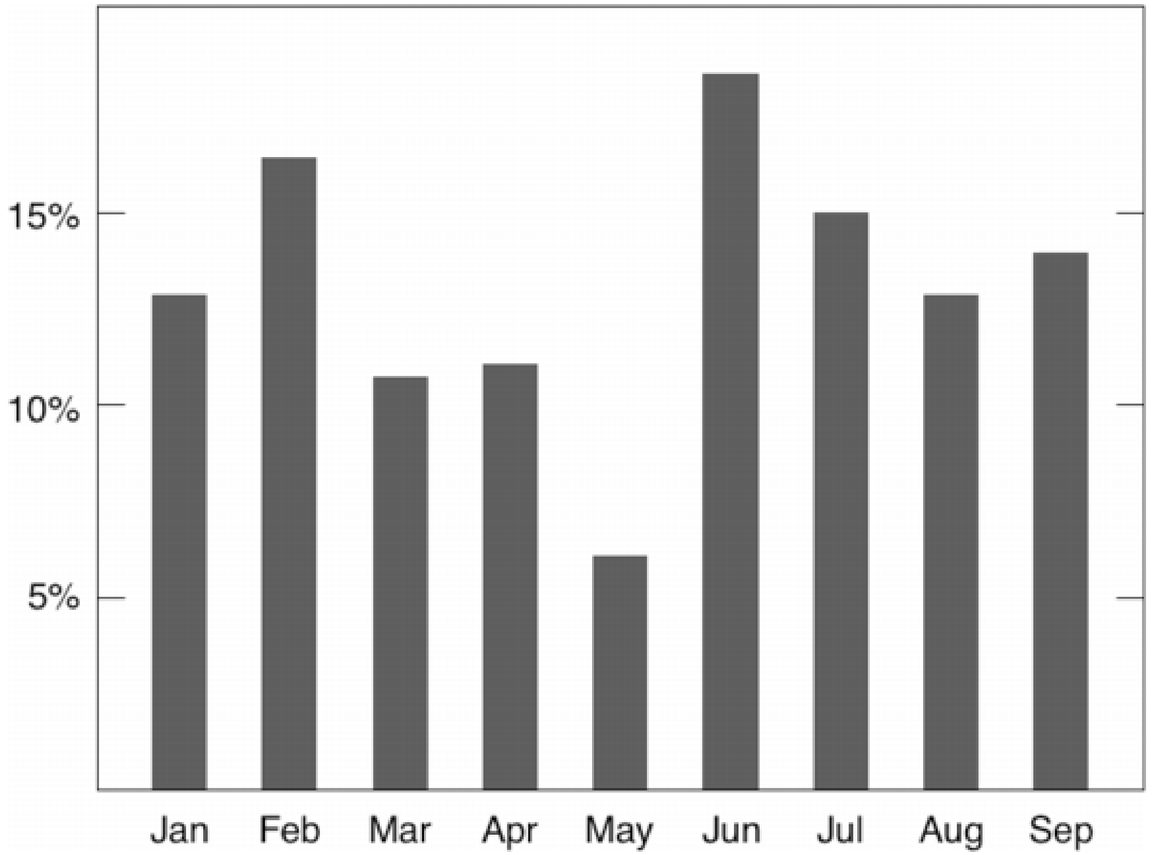
ongoing, Tim Brey

Avoid Chart Junk



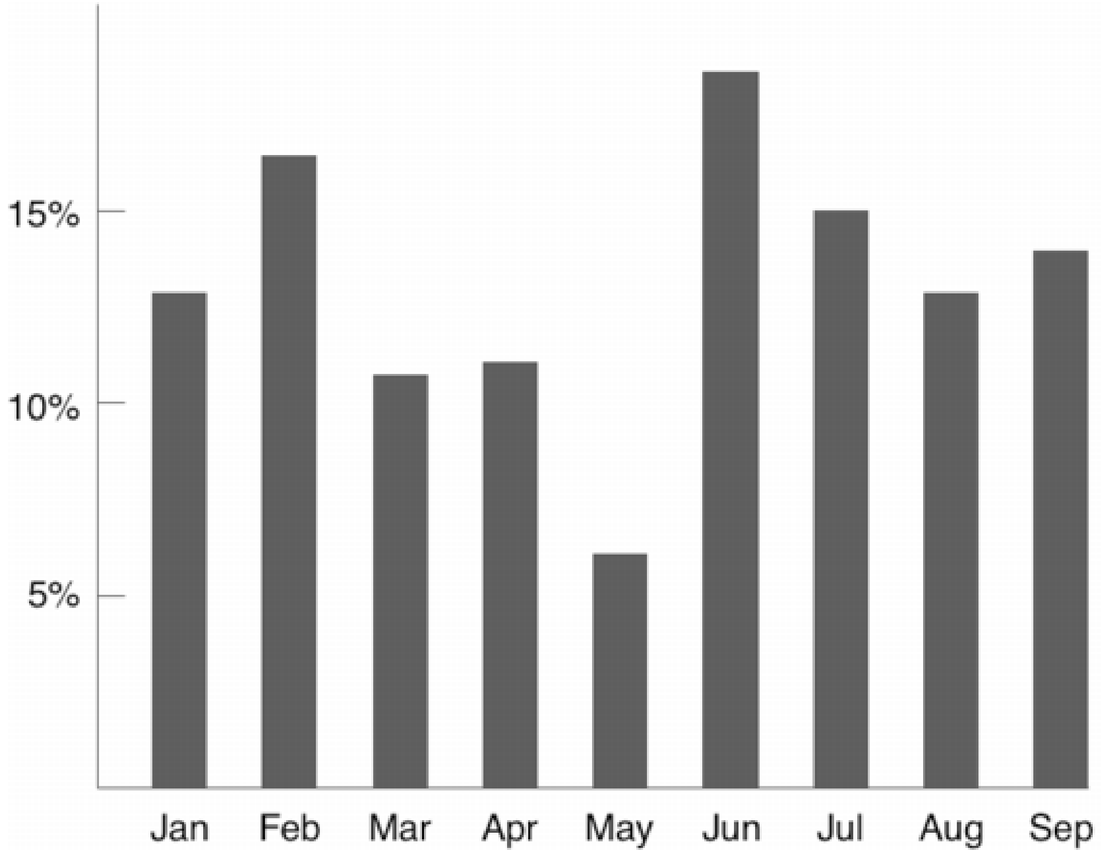
ongoing, Tim Brey

Avoid Chart Junk



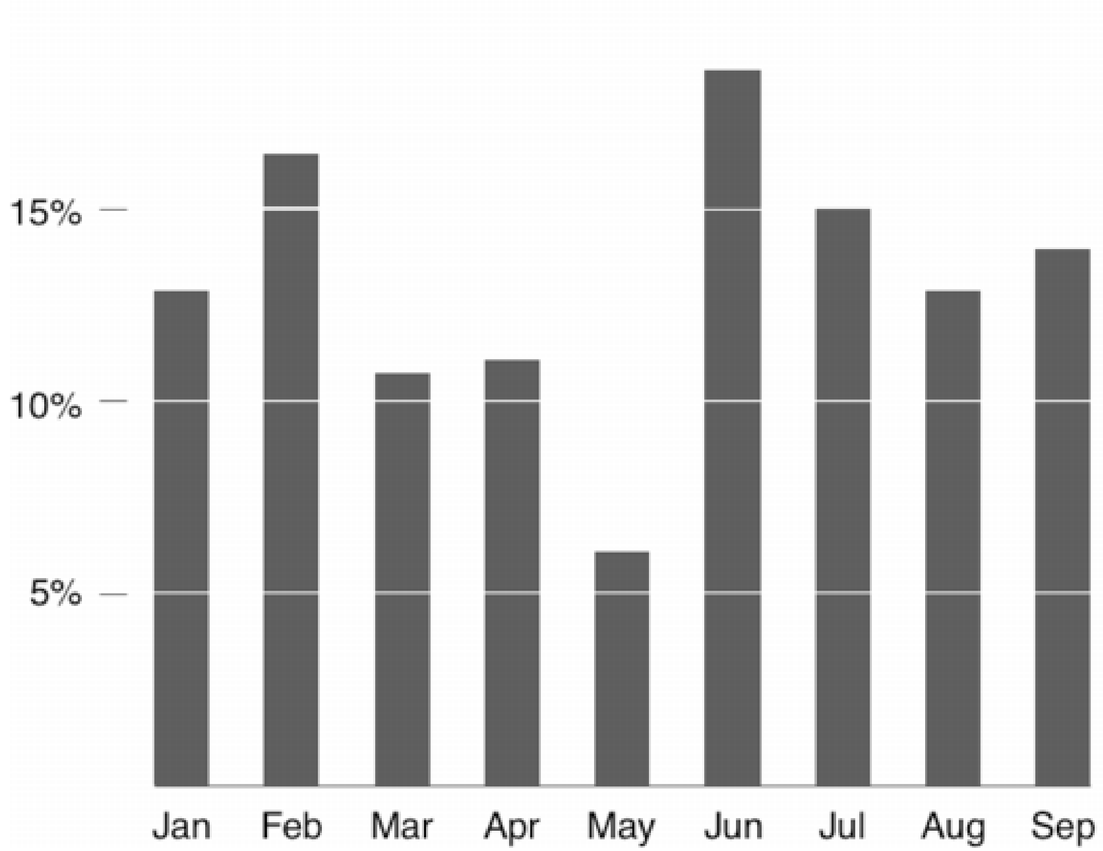
ongoing, Tim Brey

Avoid Chart Junk



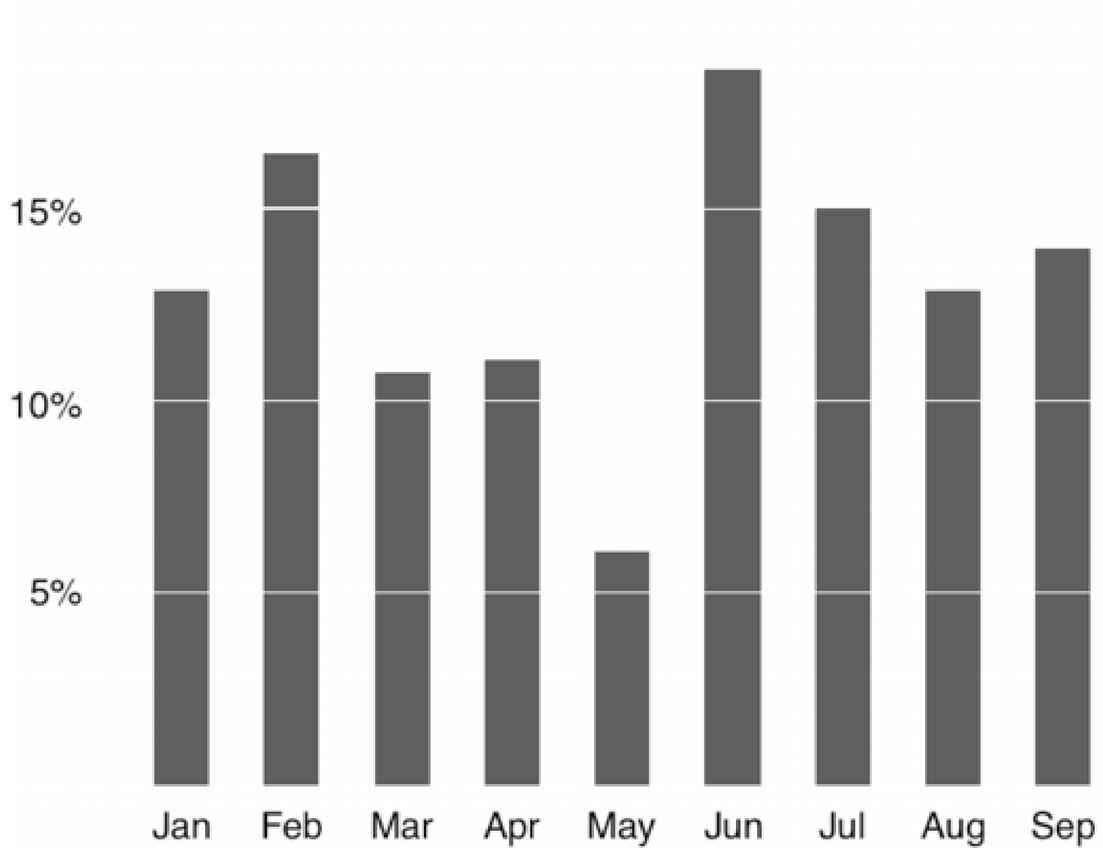
ongoing, Tim Brey

Avoid Chart Junk



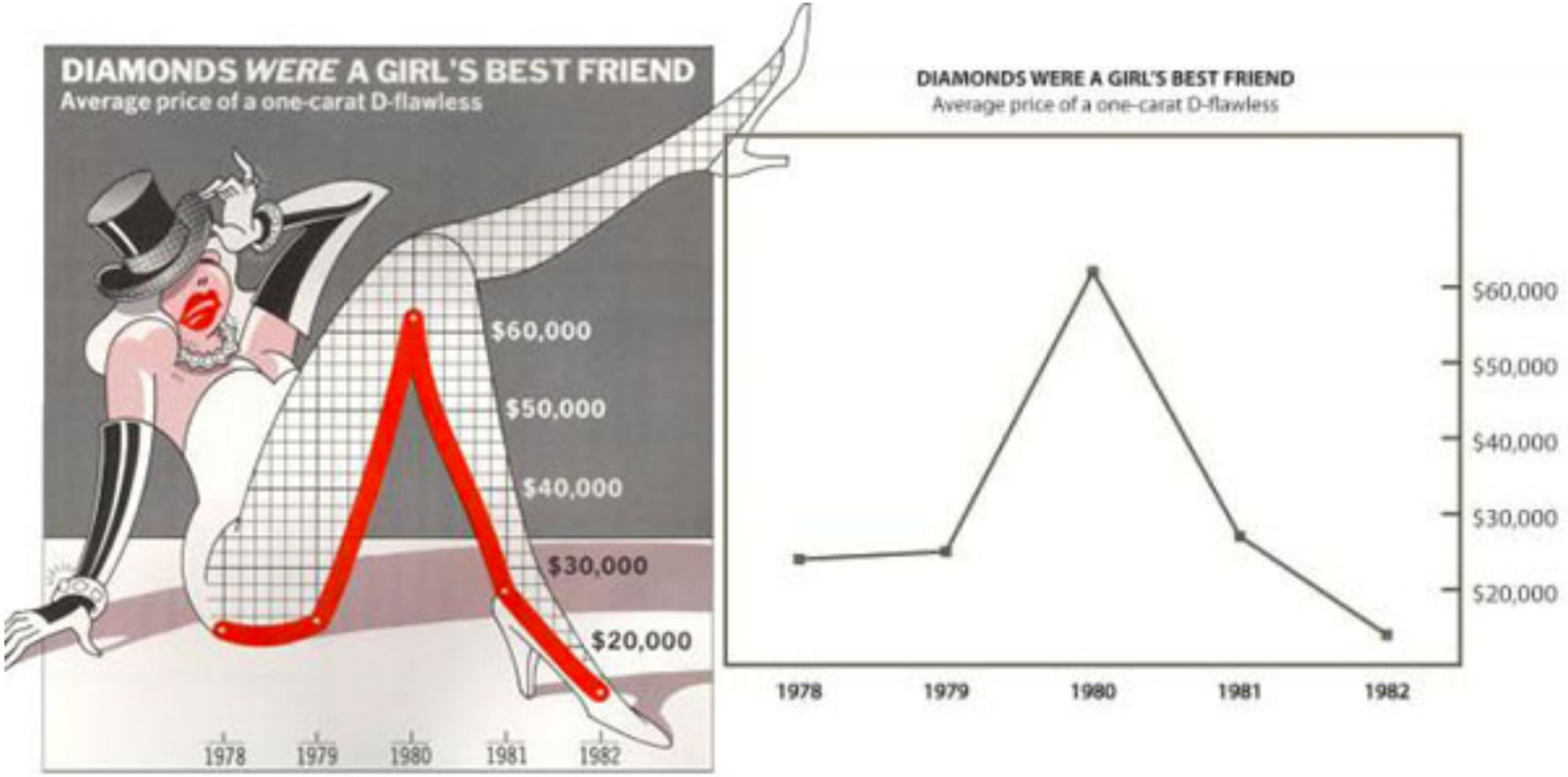
ongoing, Tim Brey

Avoid Chart Junk



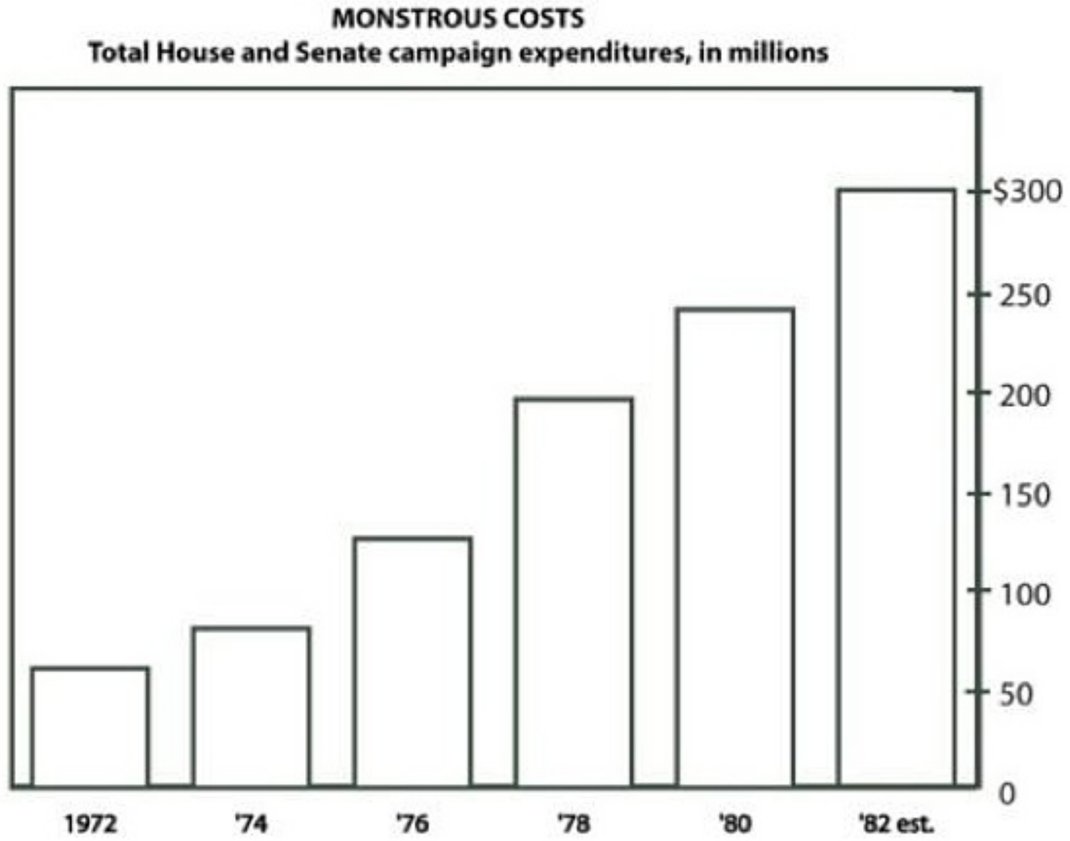
ongoing, Tim Brey

Which is better?



[Bateman et al. 2010]

Which is better?



[Bateman et al. 2010]

https://eagereyes.org/criticism/chart-junk-considered-useful-after-all

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.

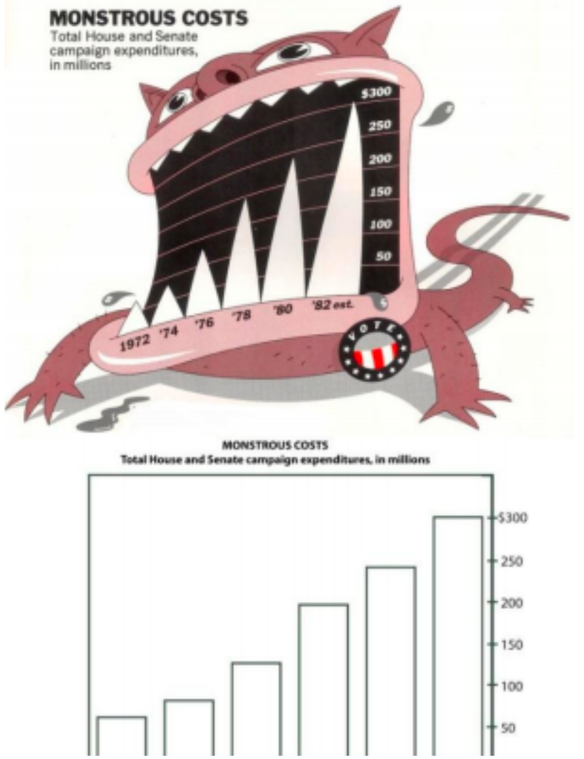
Author Keywords
Charts, information visualization, imagery, memorability.

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

General Terms
Design, Human Factors

INTRODUCTION
Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk,* decorations and other kinds of non-essential imagery, to a chart can make interpretation more difficult and can distract readers from the data [22]. This *minimalist* perspective advocates *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).



EXPERIMENTAL RESULTS

1. No difference for **interpretation accuracy**
2. No difference in **recall accuracy after a five-minute gap**
3. Significantly **better recall for Holmes charts** of both the chart topic and the details (categories and trend) **after long-term gap** (2-3 weeks).
4. Participants **saw value messages** in the Holmes charts significantly more often than in the plain charts.
5. Participants found the Holmes charts **more attractive, most enjoyed** them, and found that they were **easiest and fastest to remember**.

Use Chart Junk? It depends!

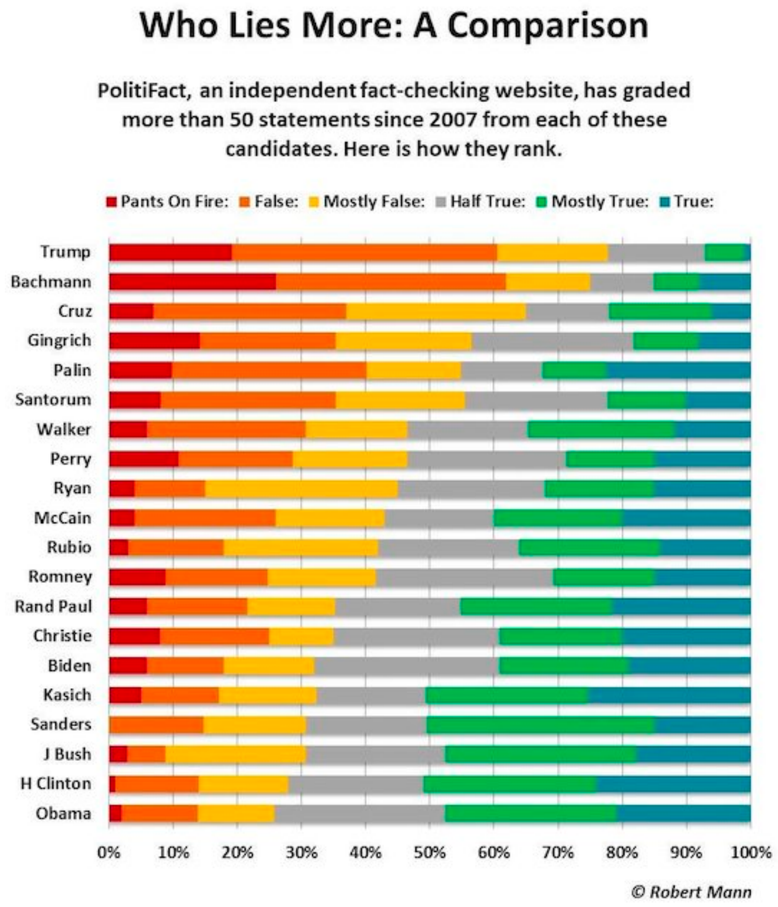
PROS

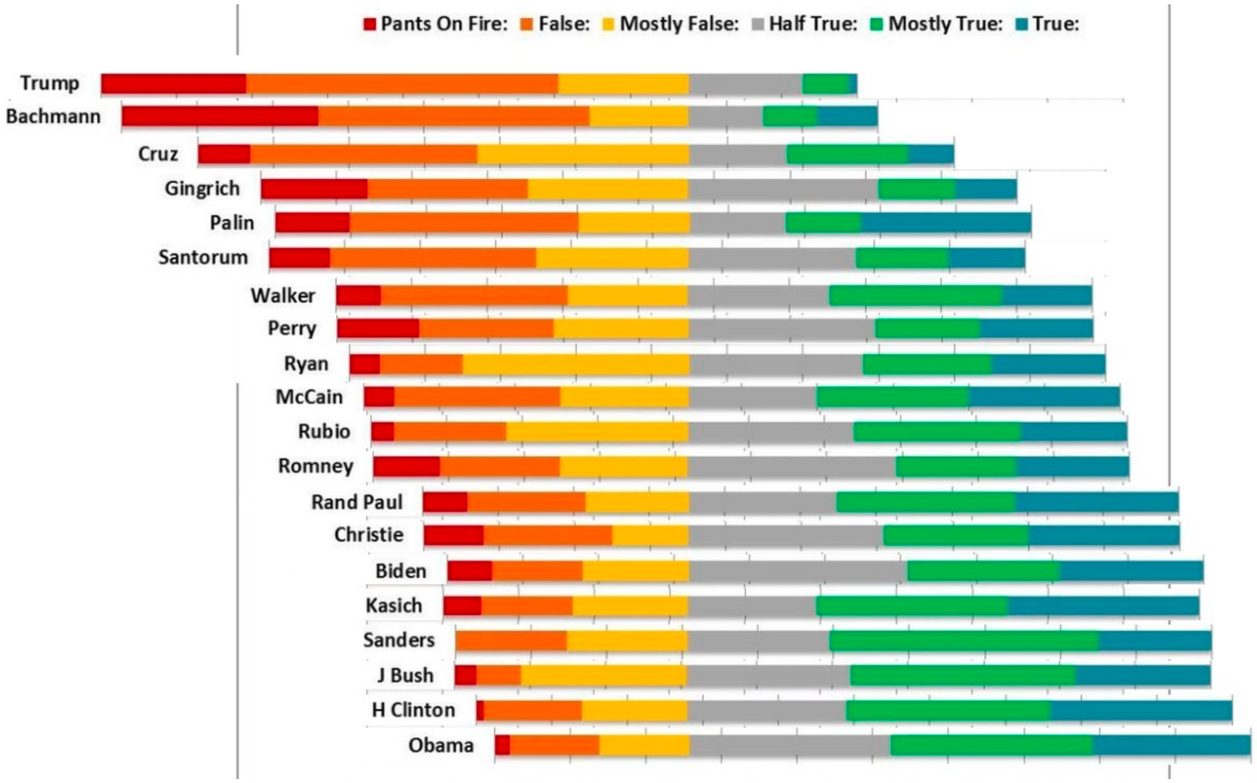
- persuasion
- memorability
- engagement

CONS

- biased analysis
- trustworthiness
- interpretability
- space efficiency
- effort

Alignment Matters





http://www.visualisingdata.com/2016/08/little-visualisation-design-part-21/

https://twitter.com/infowetrust/status/760521739092627457

3D

No Unjustified 3D

Depth judgment is bad

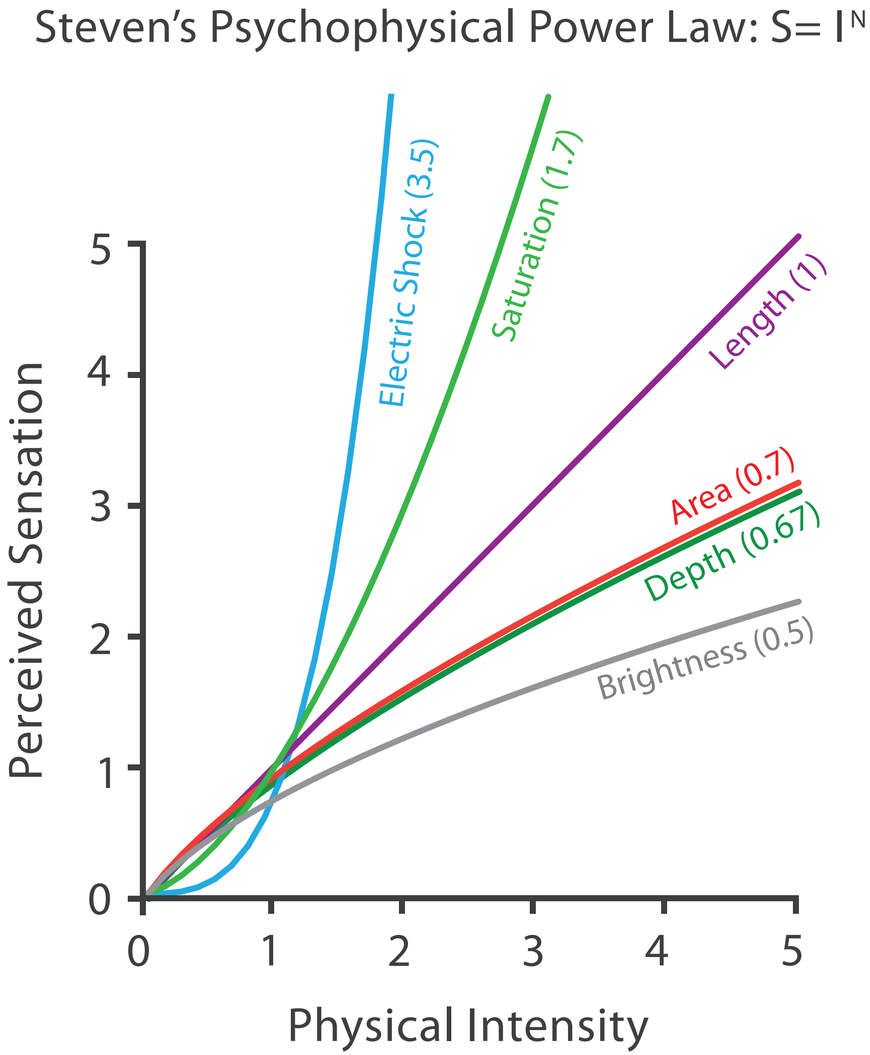
N = 0.67 Sensation=Intensity^N

Occlusion

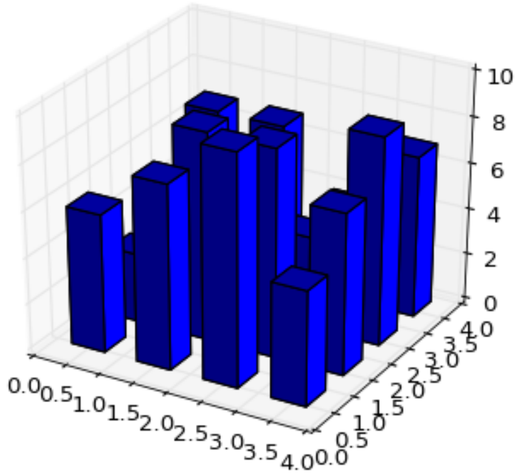
Perspective Distortion

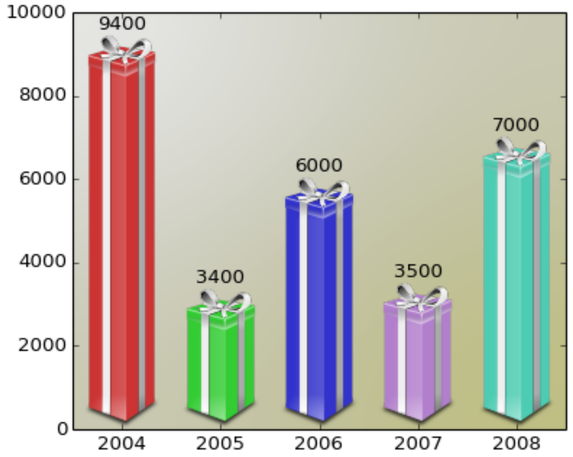
Color: Lighting / Shadows / Shading

Tilted Text illegible

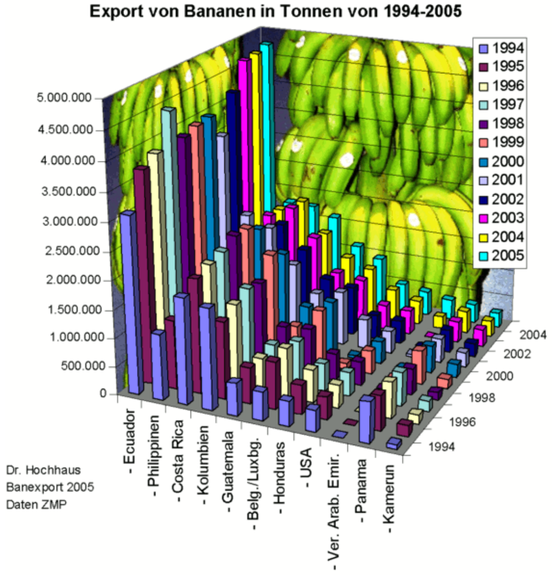


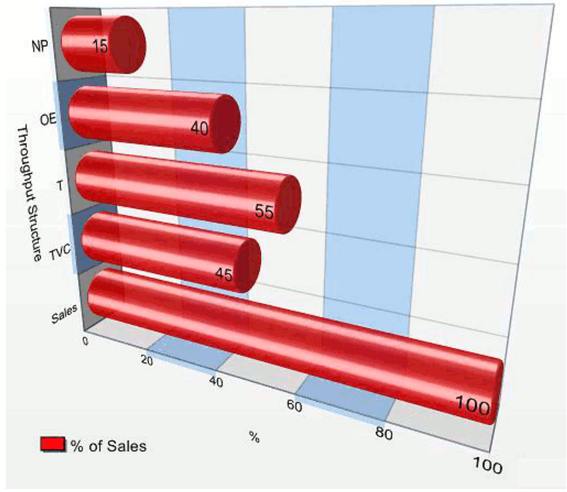
Don’t





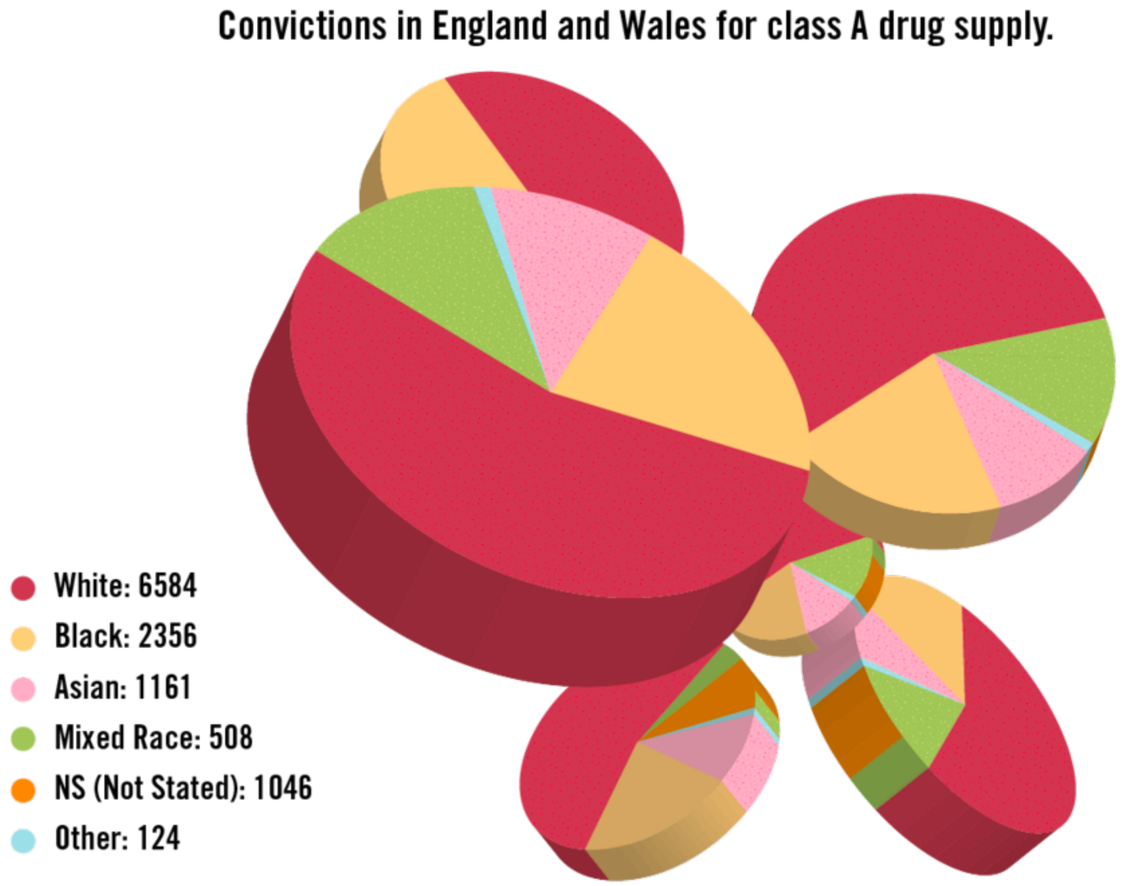
matplotlib gallery





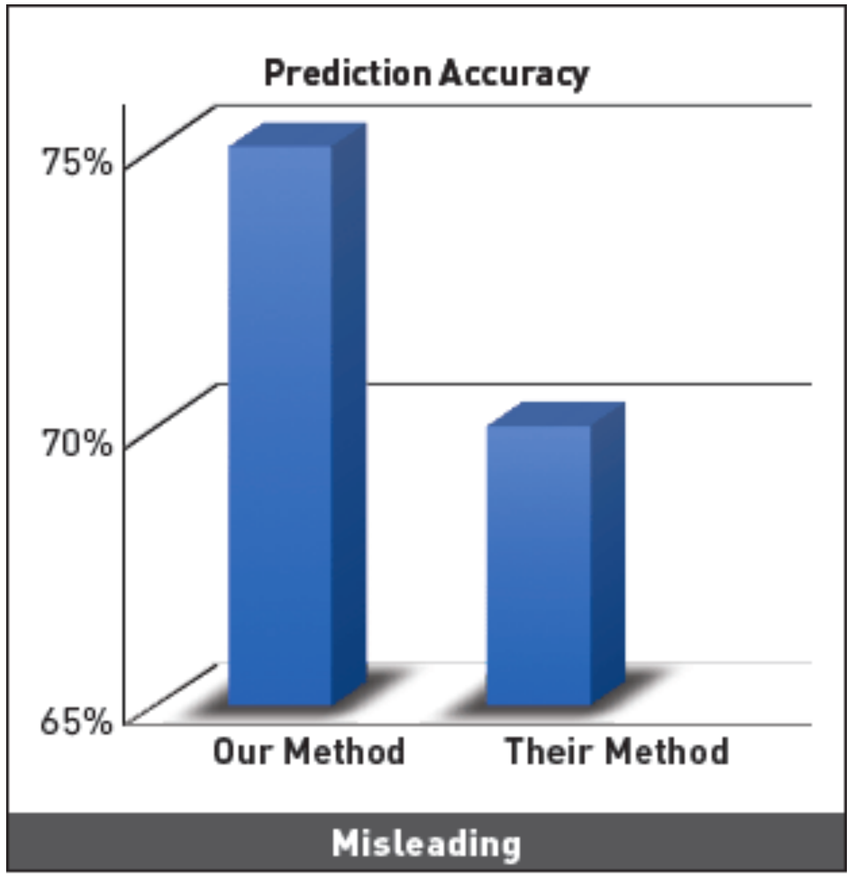
Excel Charts Blog

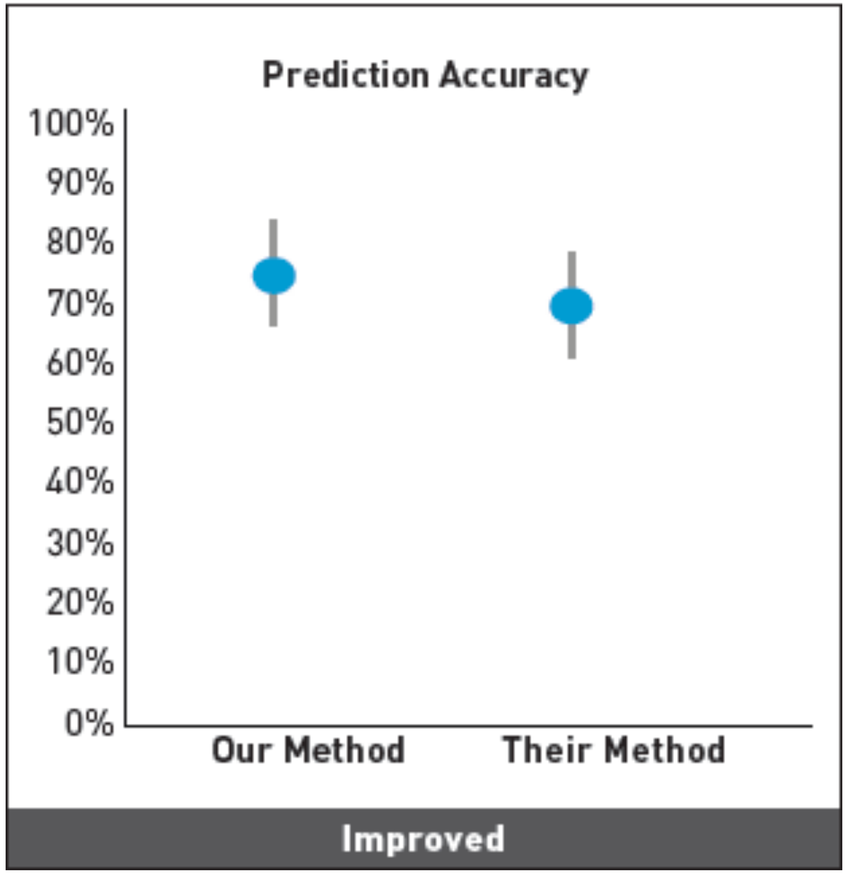
Don’t



https://www.vice.com/en_uk/read/foi-uk-drug-conviction-ethnicity-282

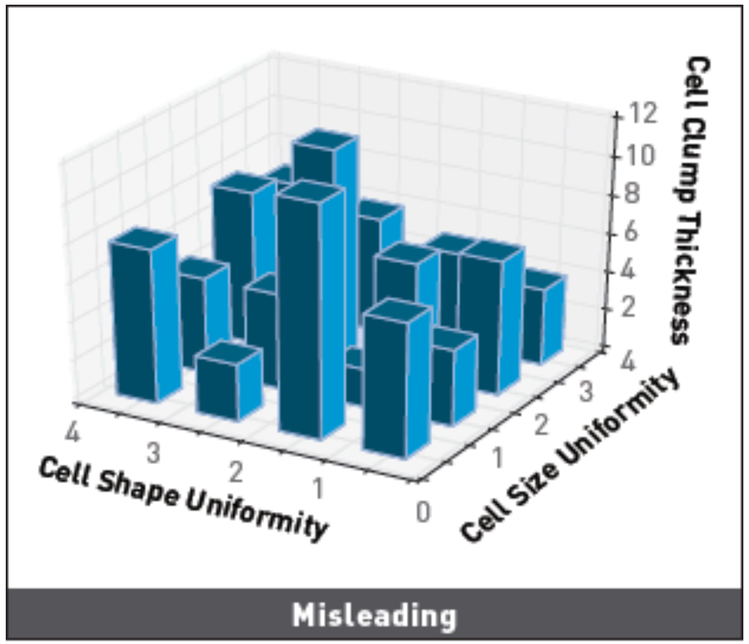
3D Design Alternatives

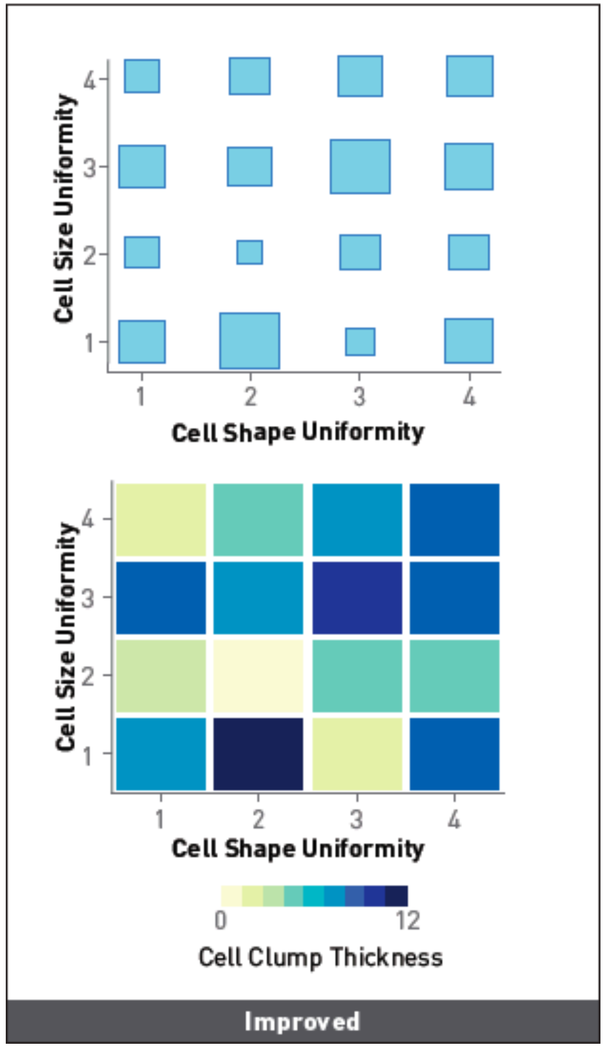




http://interactions.acm.org/archive/view/july-august-2018/the-good-the-bad-and-the-biased

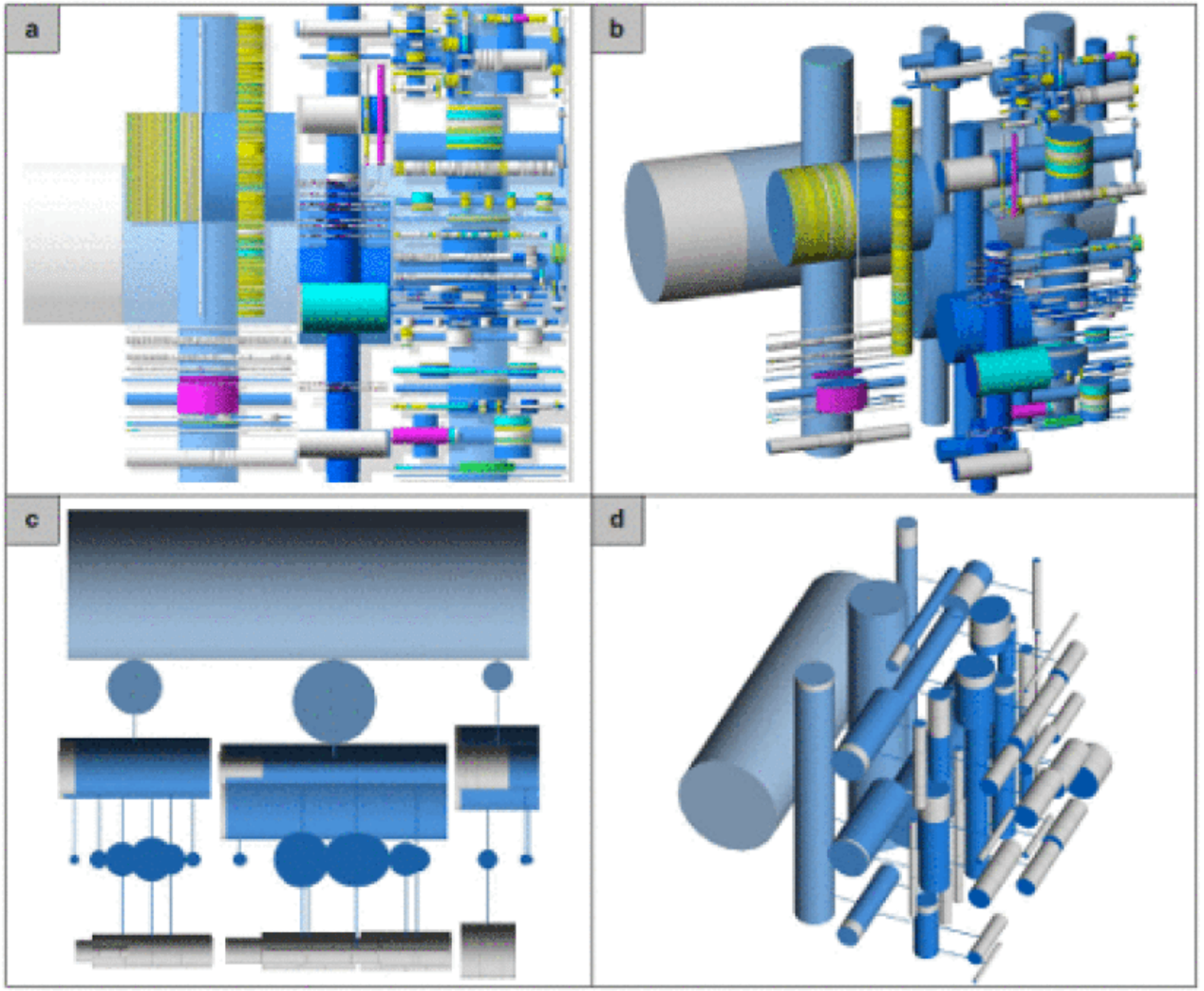
3D Design Alternatives





http://interactions.acm.org/archive/view/july-august-2018/the-good-the-bad-and-the-biased

Example: Hierarchy Visualization

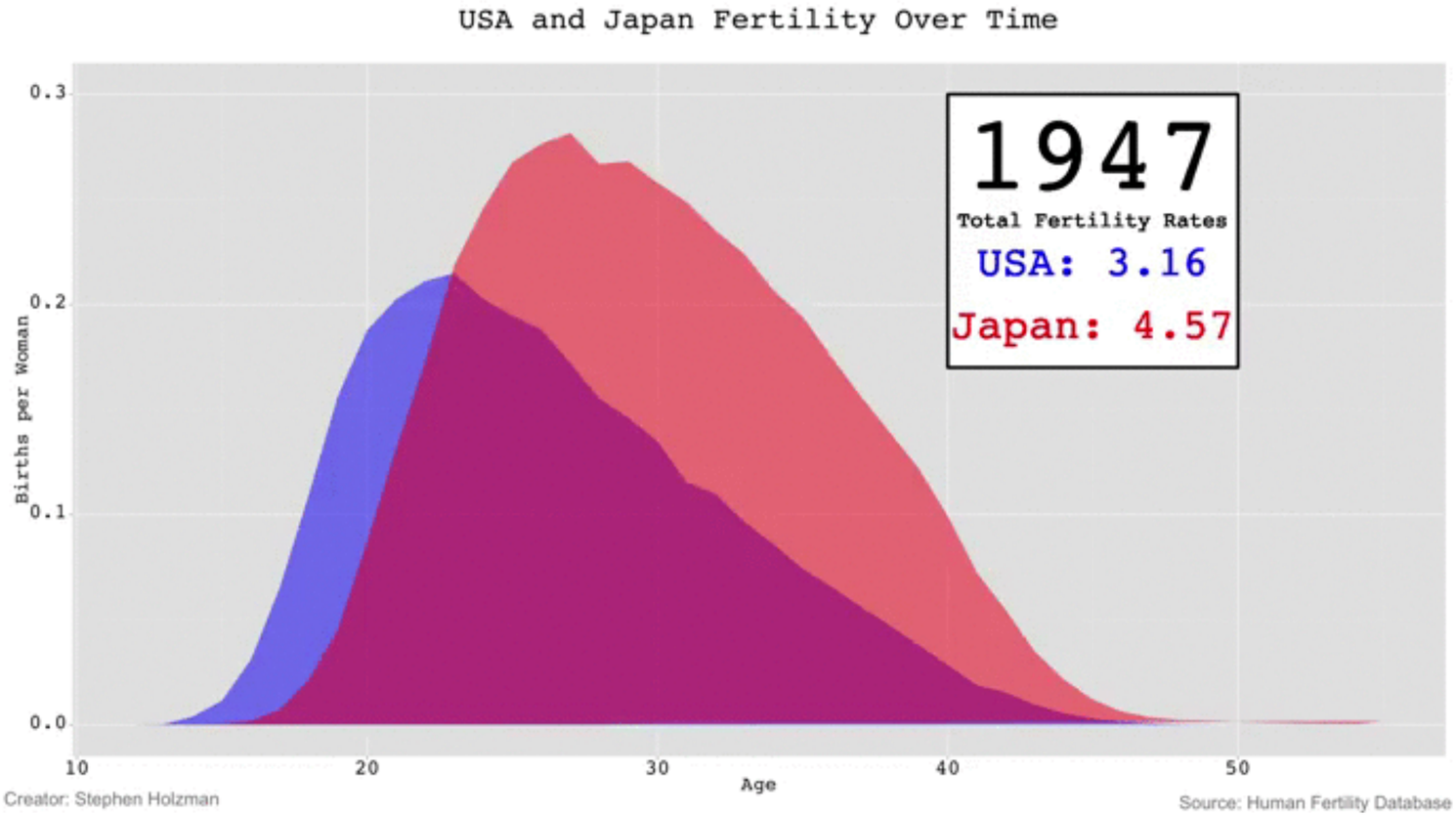


[F. van Ham ; J.J. van Wijk, 2002]

More data than fits one chart:
Animation, Multiple Views

Eyes Beat Memory

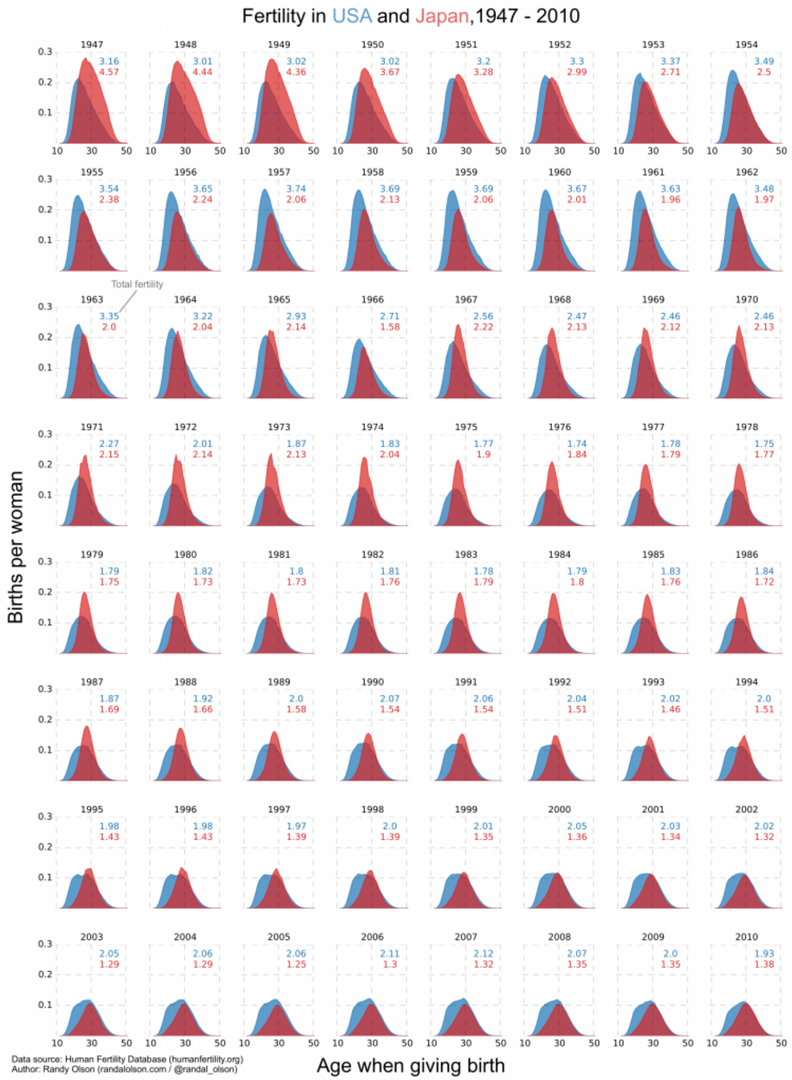
Don’t make people memorize: Show them



http://www.randalolson.com/2015/08/23/small-multiples-vs-animated-gifs-for-showing-changes-in-fertility-rates-over-time/

What can we do differently?

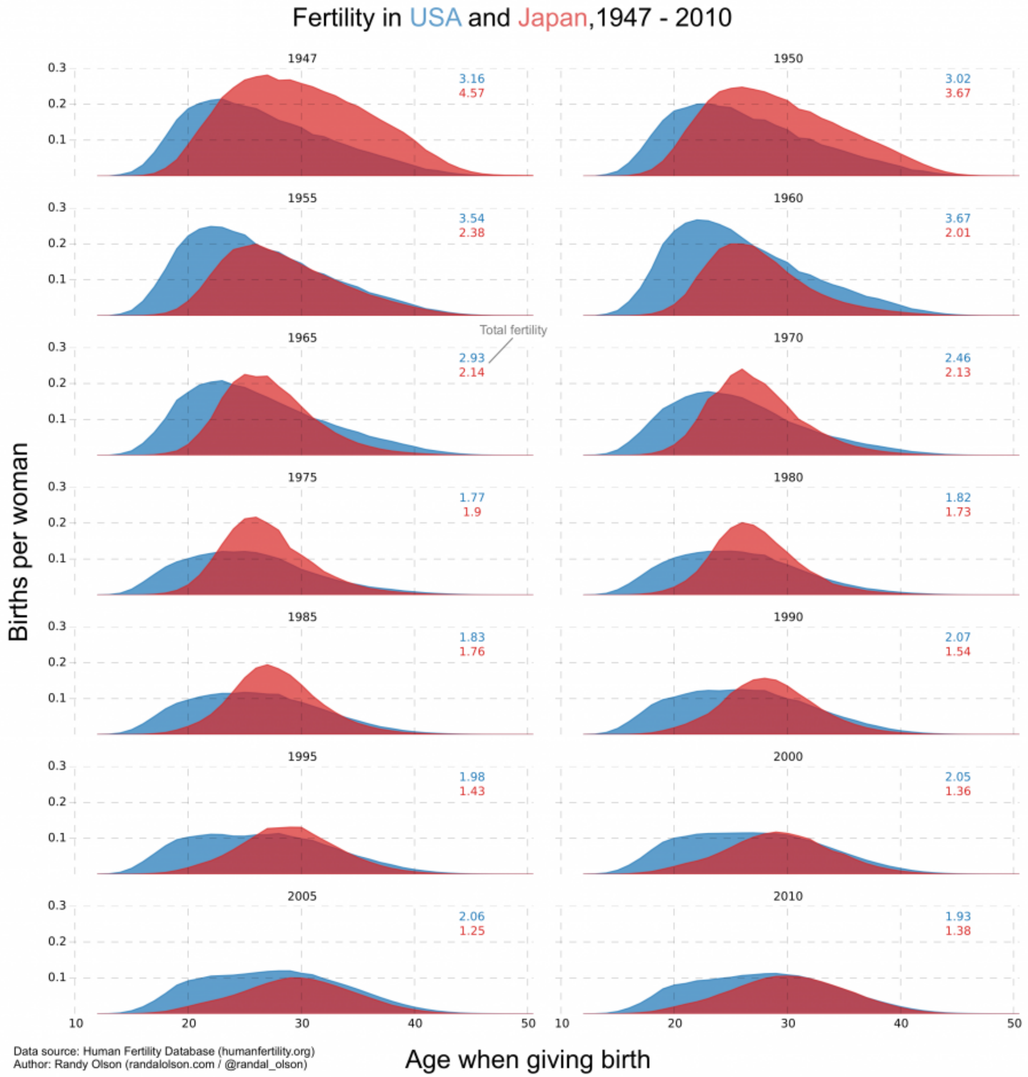
Eyes Beat Memory: Small Multiples



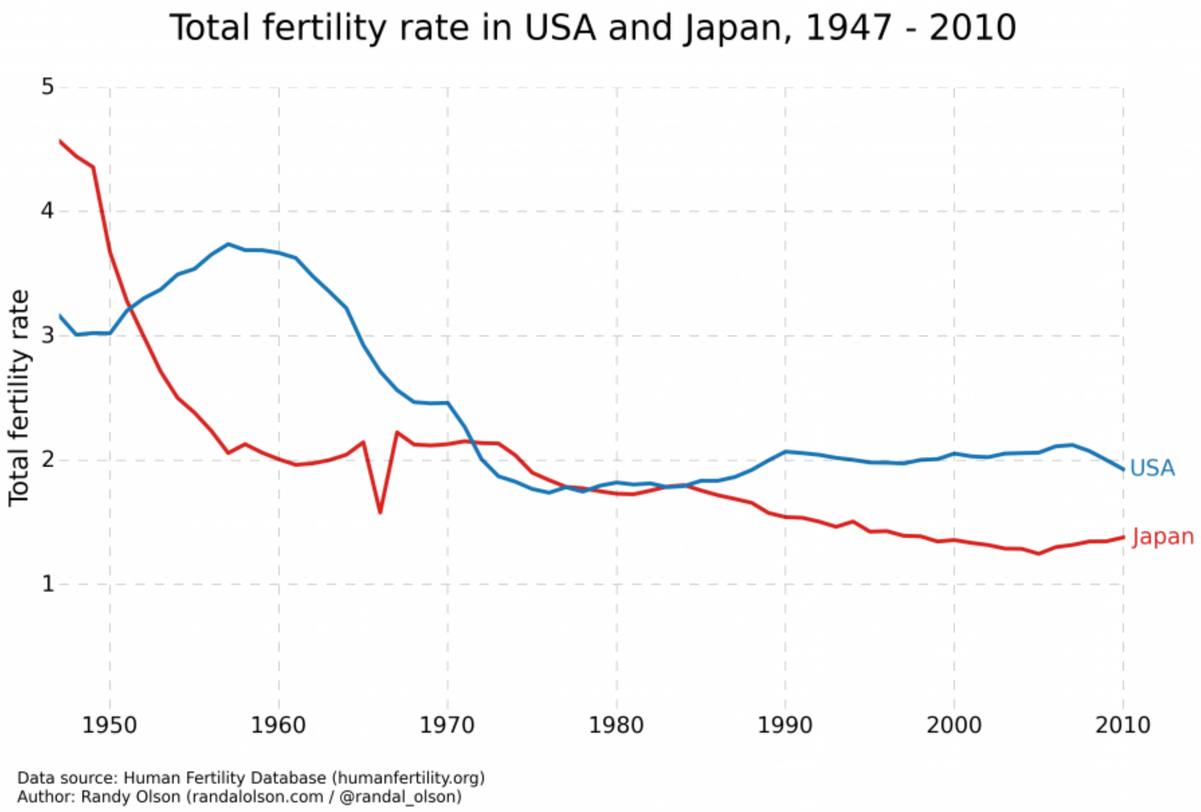
A lot of charts

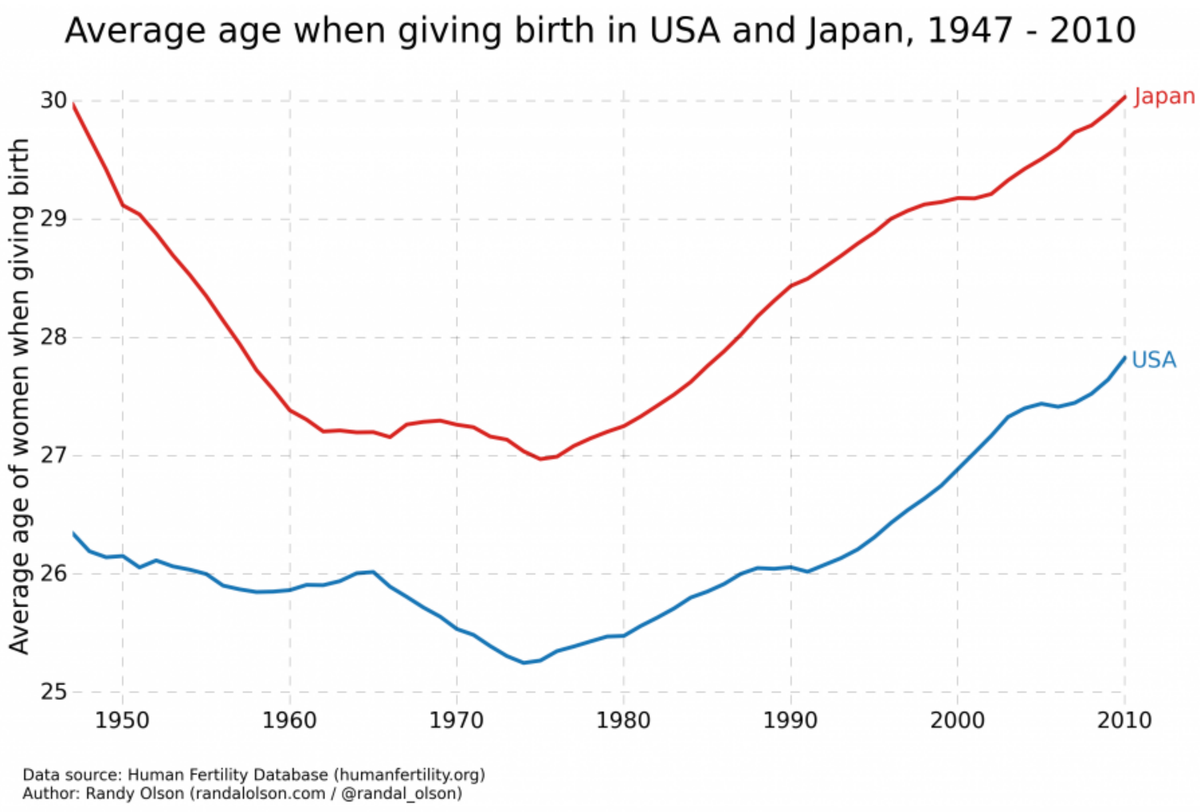
Do we need all of them?

Eyes Beat Memory: Small Multiples

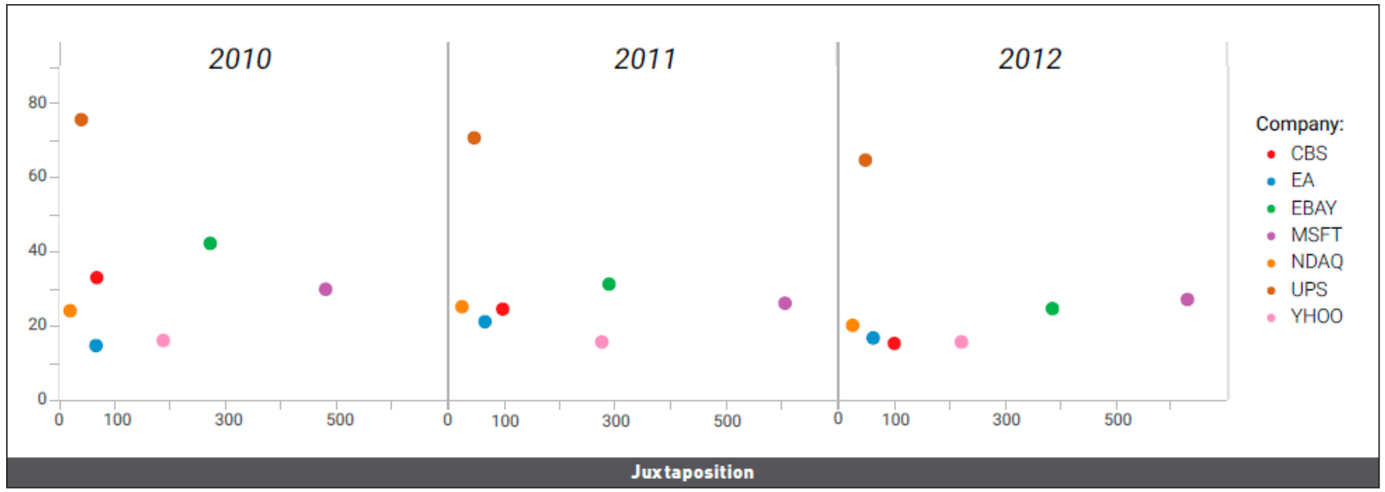


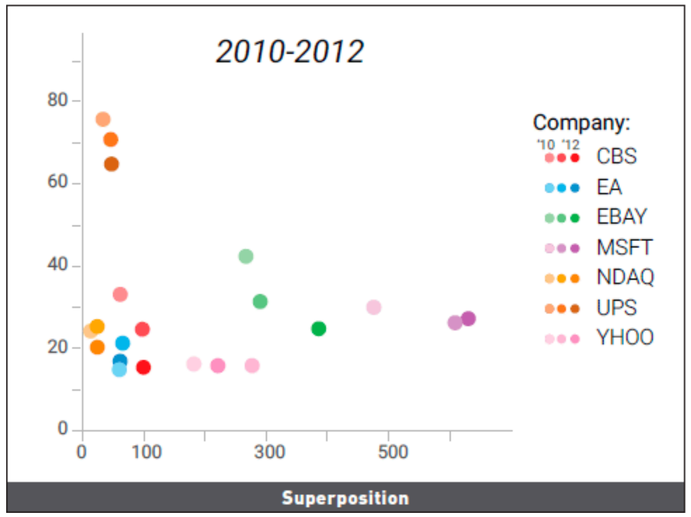
Simplify!

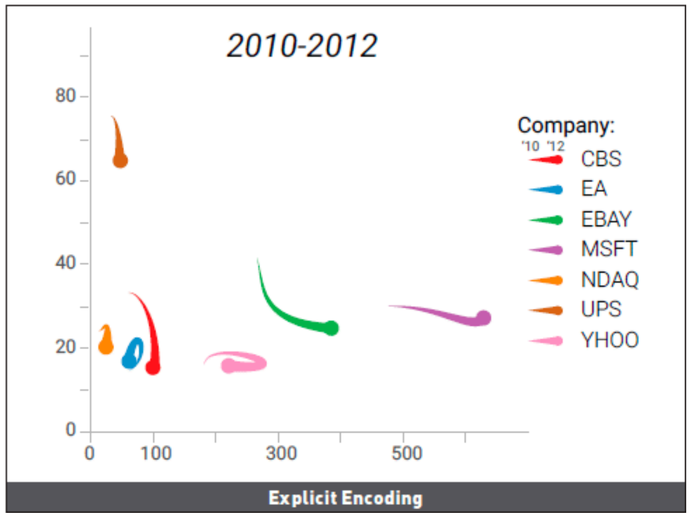

Total fertility rate in USA and Japan, 1947 - 2010
Total fertility rate
5
4
3
2
1
1950
1960
1970
1980
1990
2000
2010
USA
Japan
Data source: Human Fertility Database (humanfertility.org)
Author: Randy Olson (randalolson.com / @randal_olson)


Average age when giving birth in USA and Japan, 1947 - 2010
Average age of women when giving birth
30
29
28
27
26
25
1950
1960
1970
1980
1990
2000
2010
Japan
USA
Data source: Human Fertility Database (humanfertility.org)
Author: Randy Olson (randalolson.com / @randal_olson)

Small Multiple Design Alternatives







http://interactions.acm.org/archive/view/july-august-2018/the-good-the-bad-and-the-biased