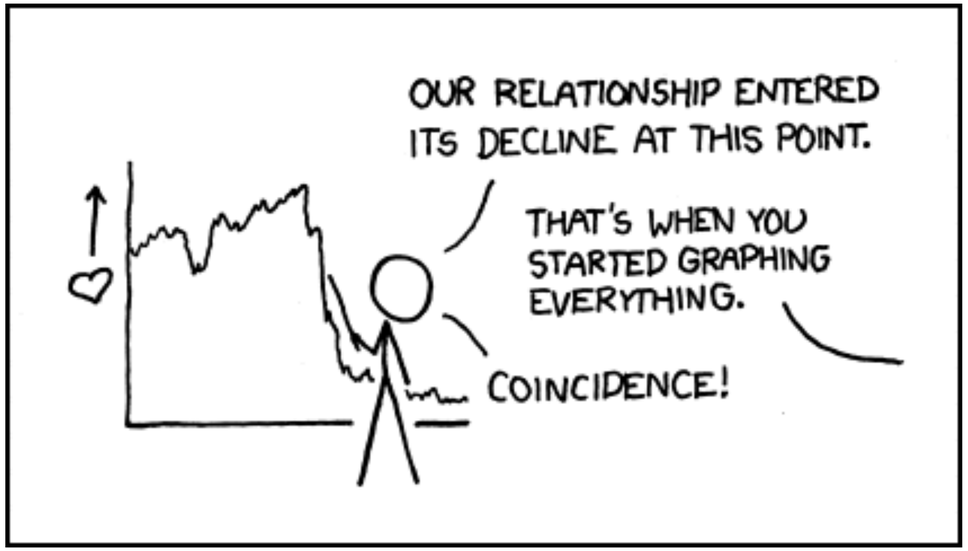
CS-5630 / CS-6630 Visualization Design Guidelines; Tasks

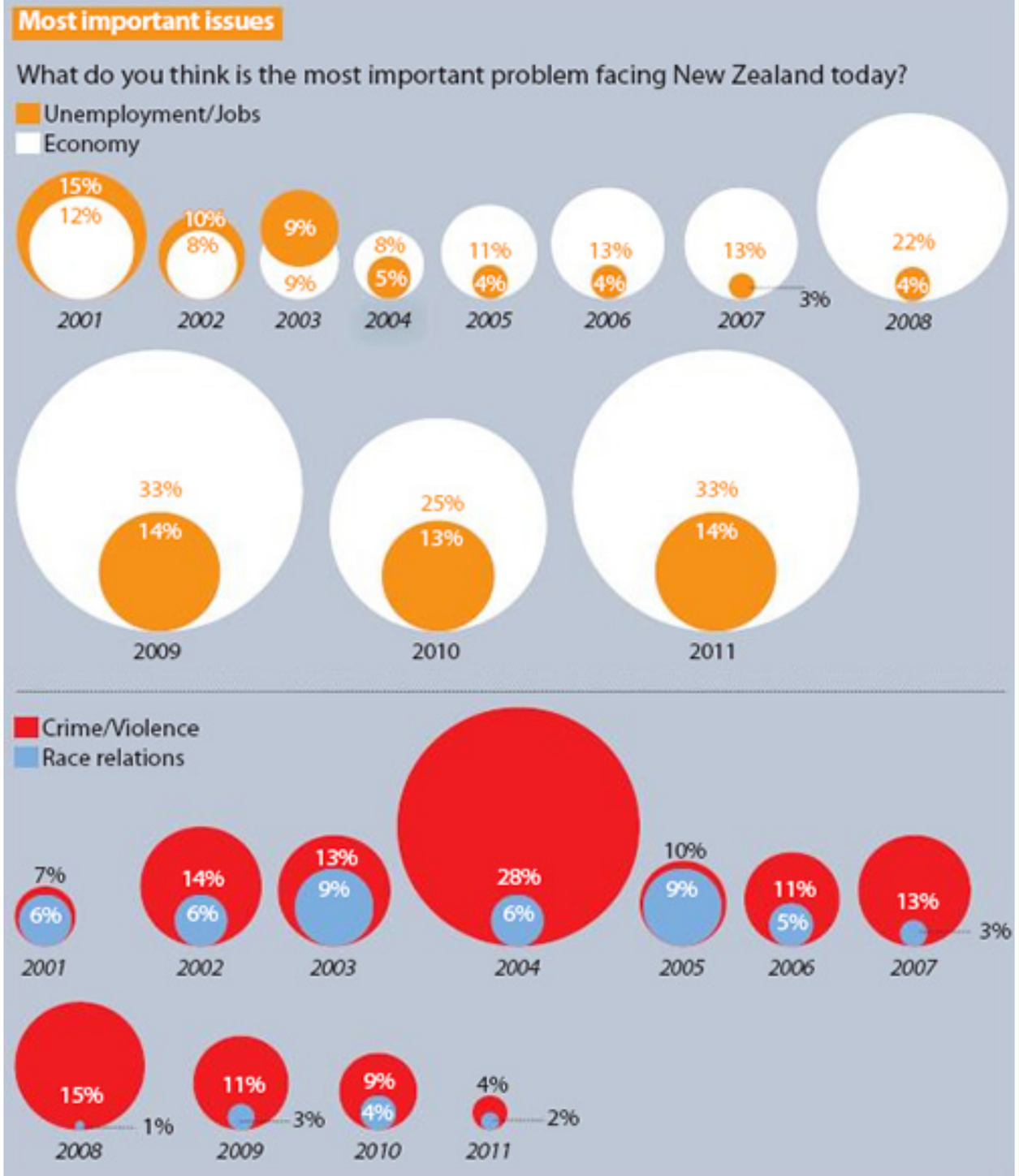
Alexander Lex
alex@sci.utah.edu

THE UNIVERSITY OF UTAH



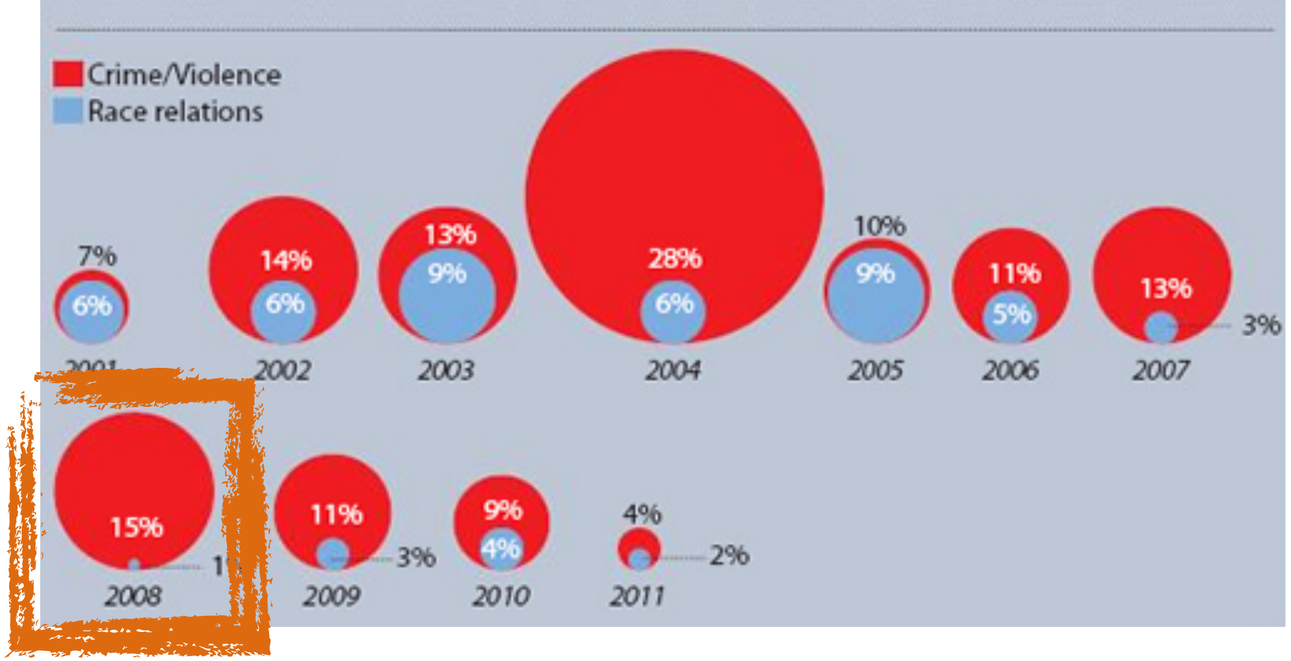
[xkcd]

Design Critique / Redesign



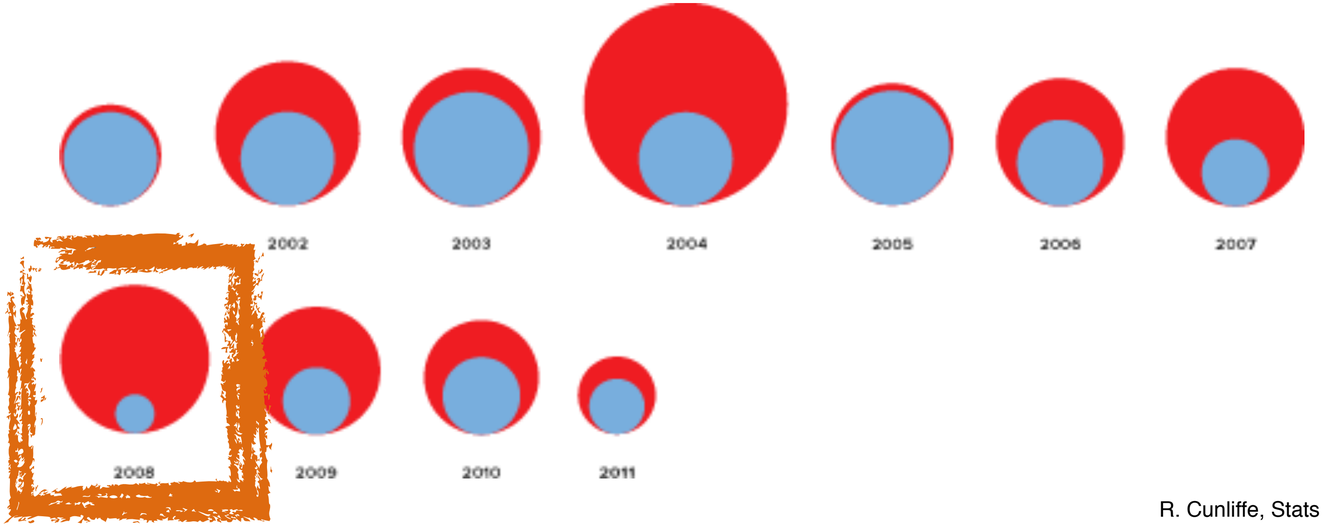
https://goo.gl/lHWp4x

Sunday Star Times, 2012

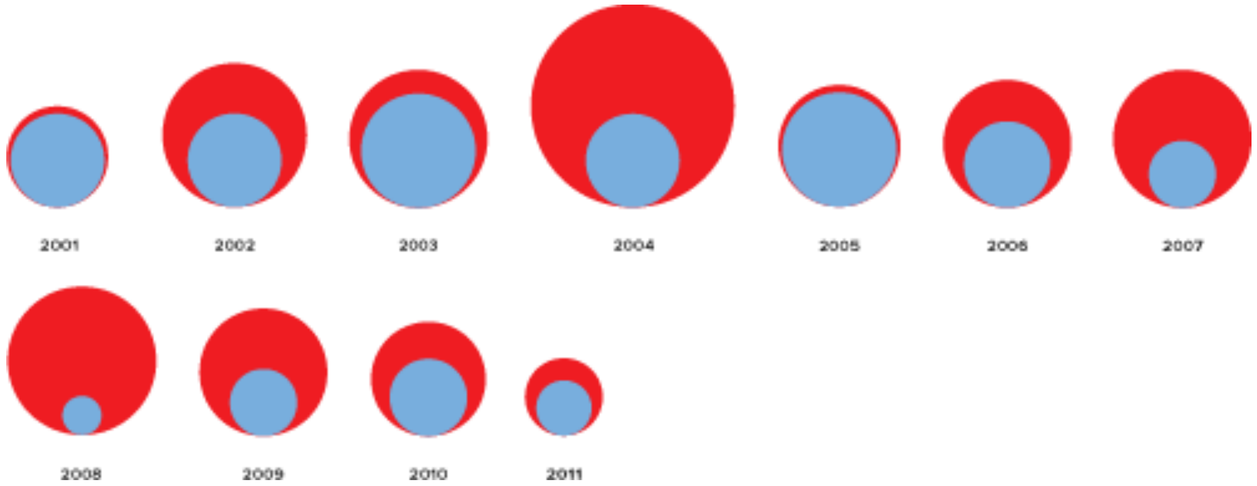


Quantity encoded by diameter, not area!

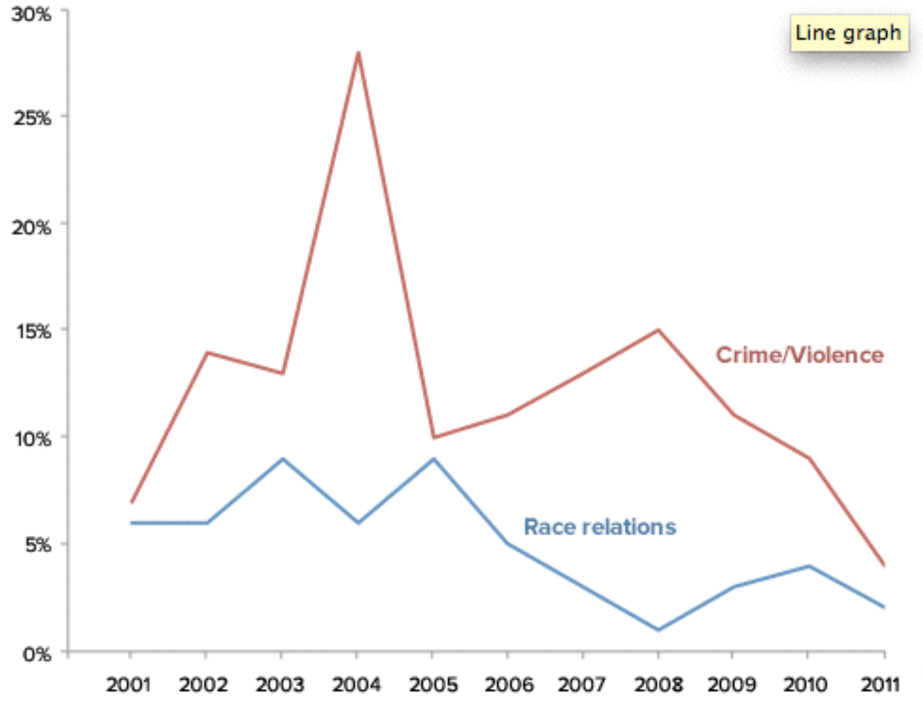
Fixing that:



R. Cunliffe, Stats Chat



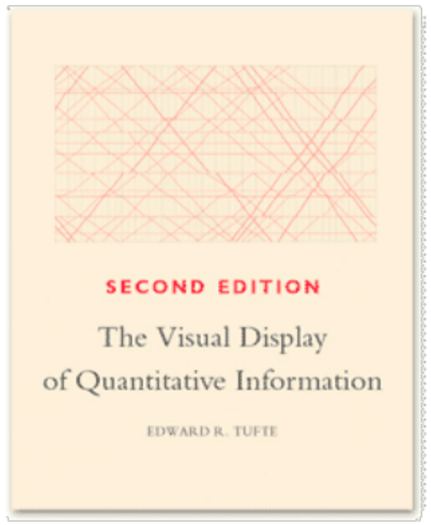
But is this visual encoding appropriate in the first place?

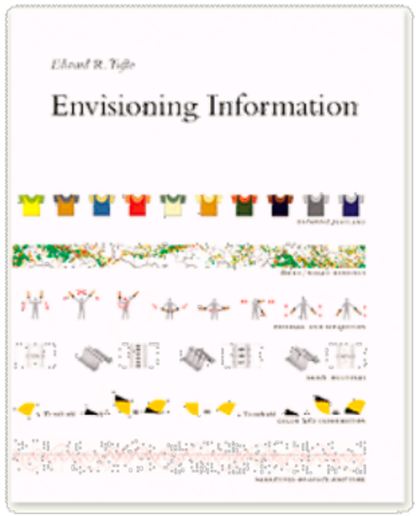


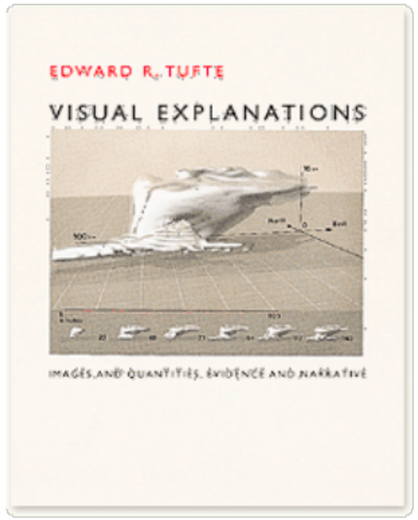
R. Cunliffe, Stats Chat

Design Guidelines

Edward Tufte

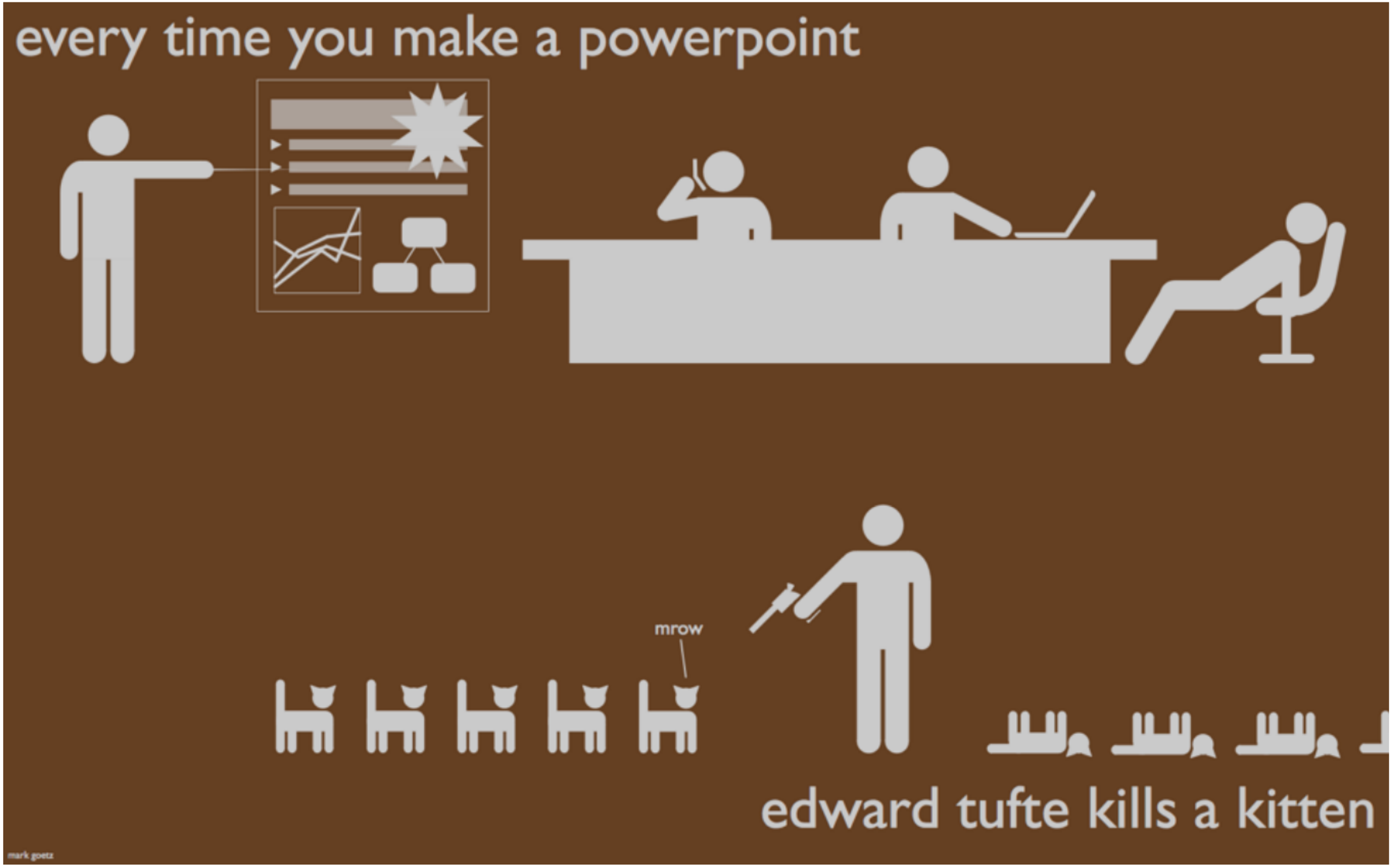






Design Excellence

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

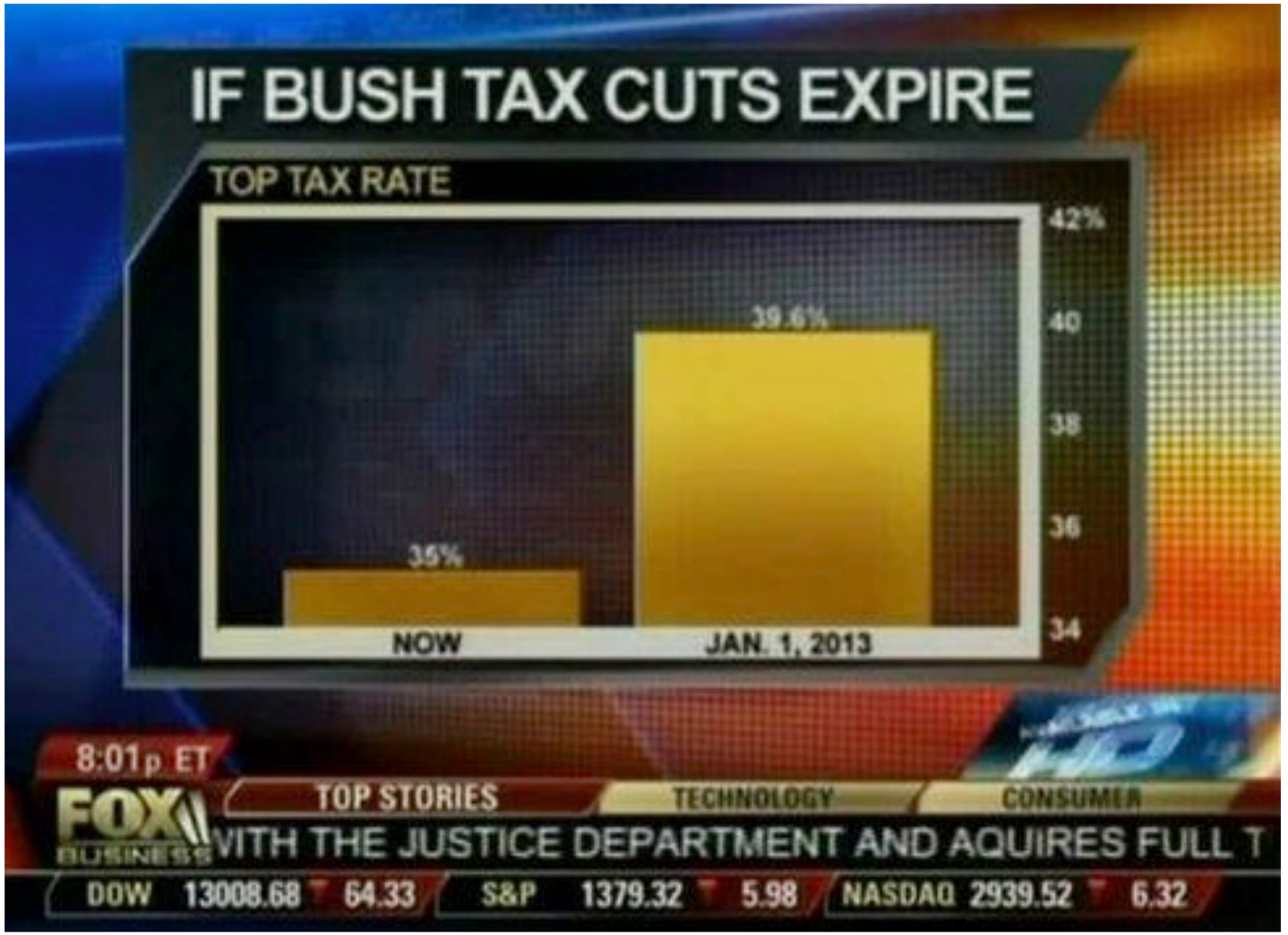

every time you make a powerpoint
mrow
edward tufte kills a kitten
mark goetz

Tufte's Lessons

Practice: graphical integrity and excellence

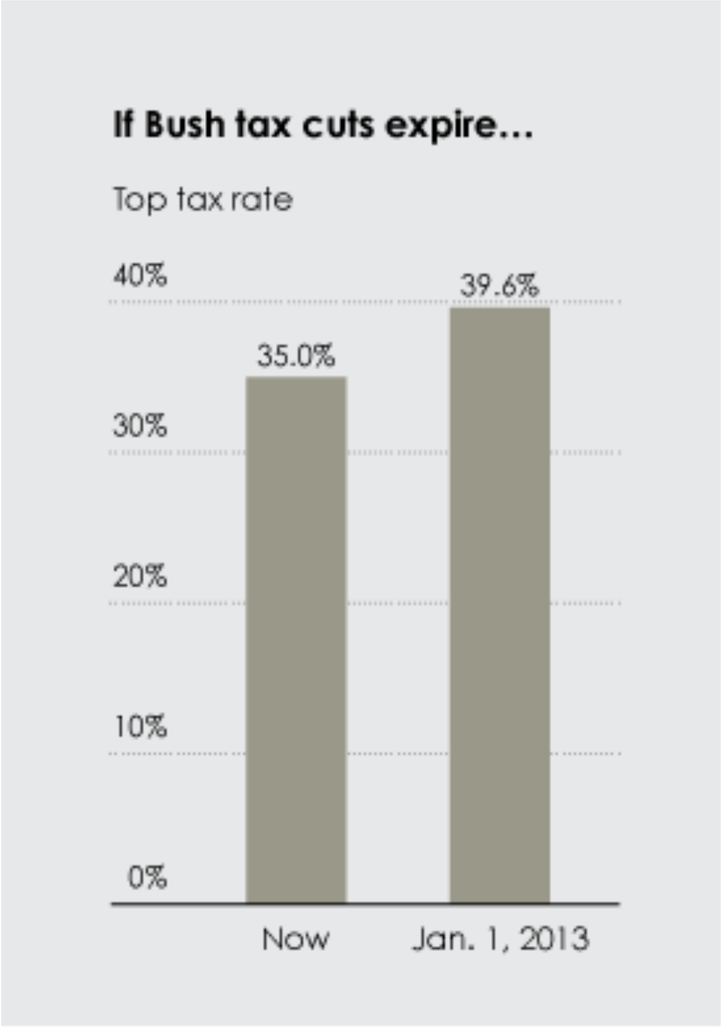
Theory: design principles for data graphics

Graphical Integrity



Flowing Data

Scale Distortions



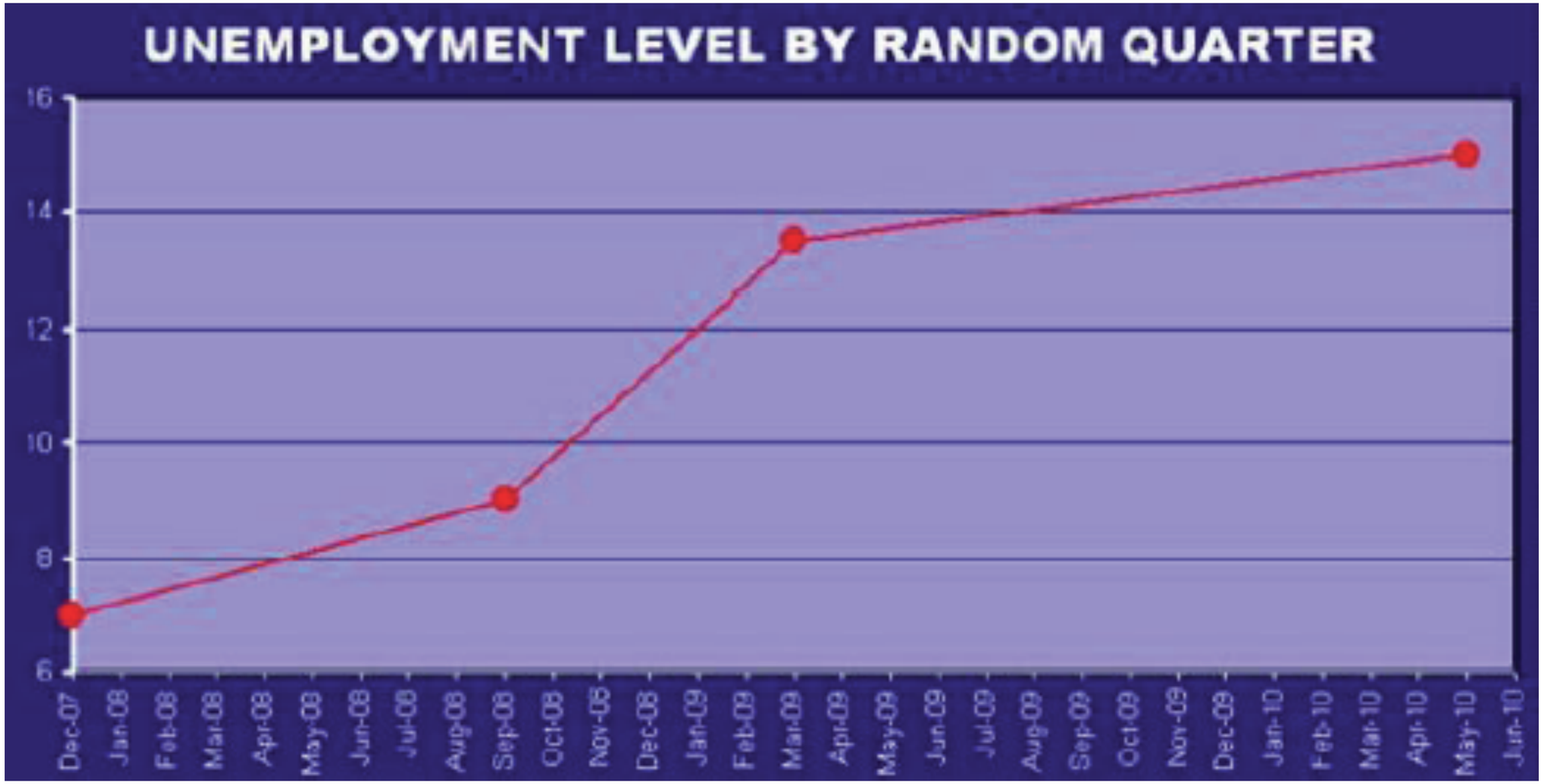
Flowing Data

What's wrong?

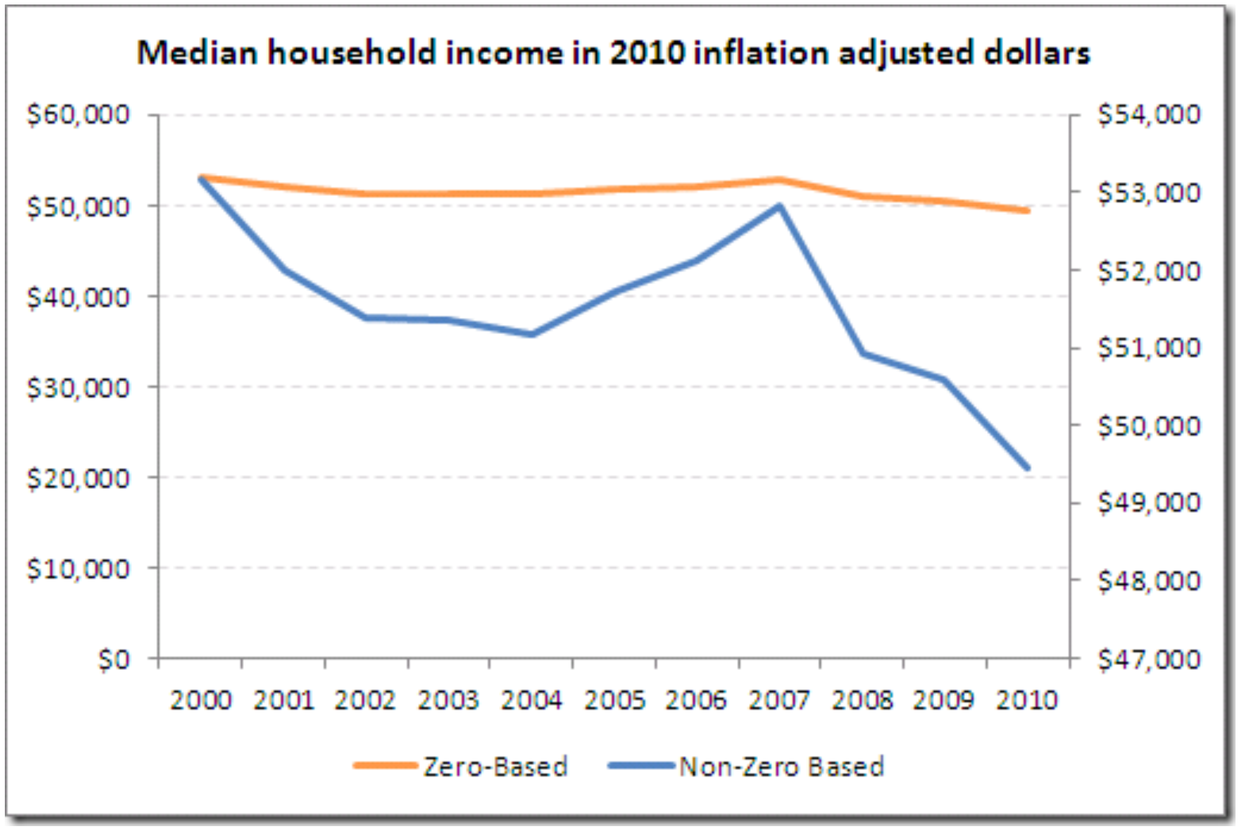

How 2012 Stacks Up
The Warmest Years on Record
Contiguous U.S.
55.3
54.31
54.32
54.14
53.93
53.81
55.5°
55.0°
54.5°
54.0°
53.5°
53.0°
1921
1999
1934
2006
1998
2012
Source: NOAA's National Climatic Data Center - State of the Climate National Overview
CLIMATE CENTRAL

Scale Distortions

Scale Distortions

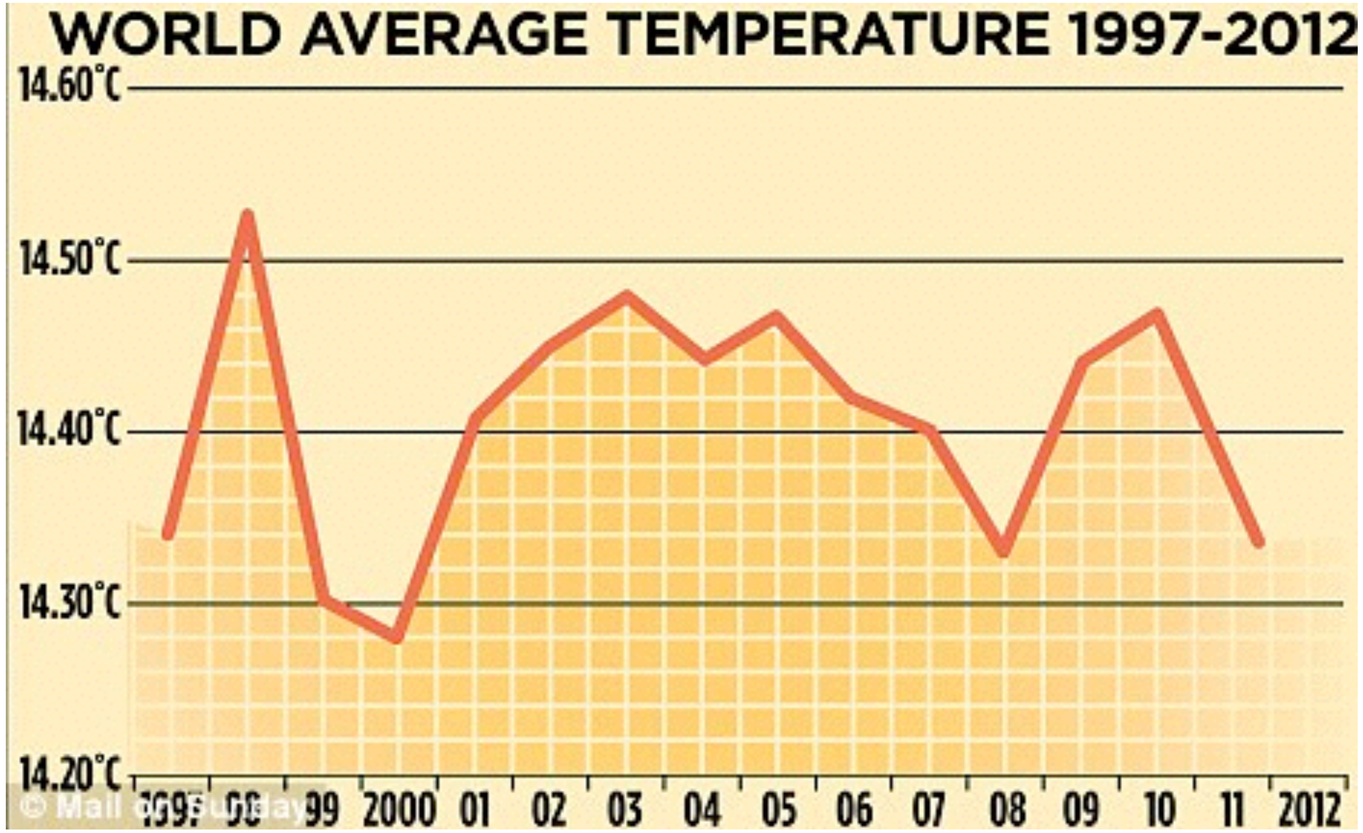


Start Scales at 0?



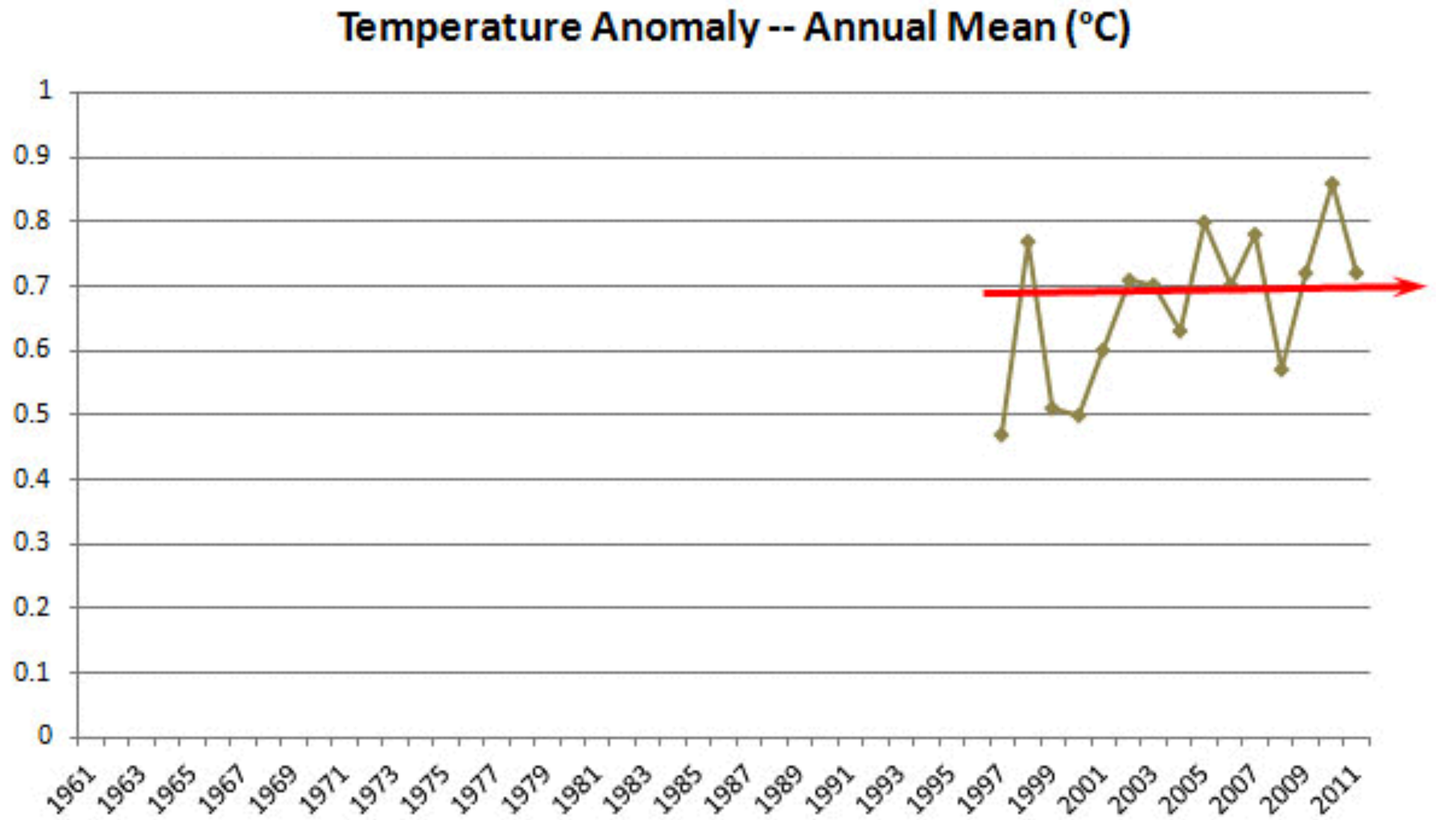
A. Kriebel, VizWiz

Global Warming?



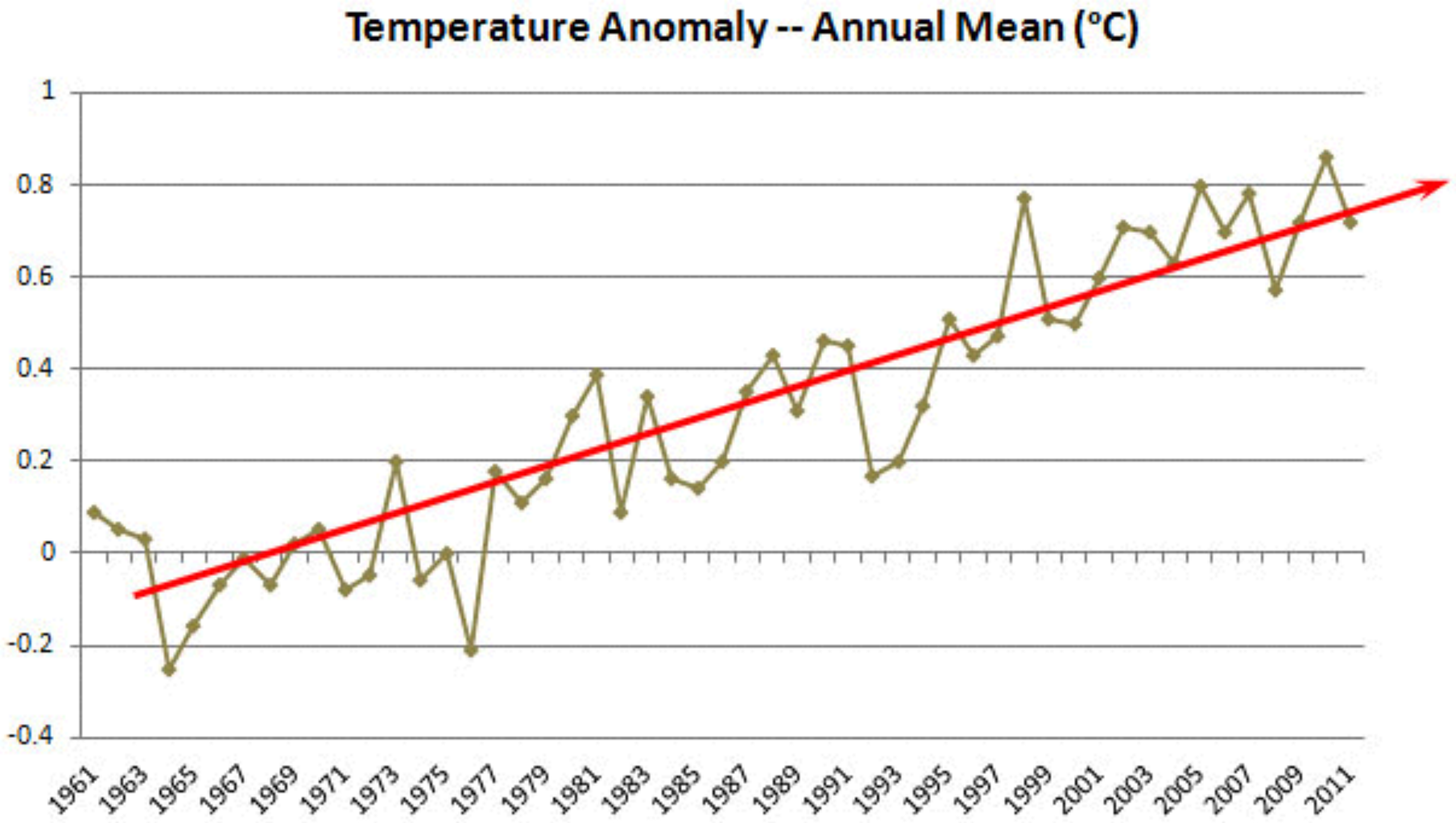
The Daily Mail, UK, Jan 2012

Global Warming?



Mother Jones

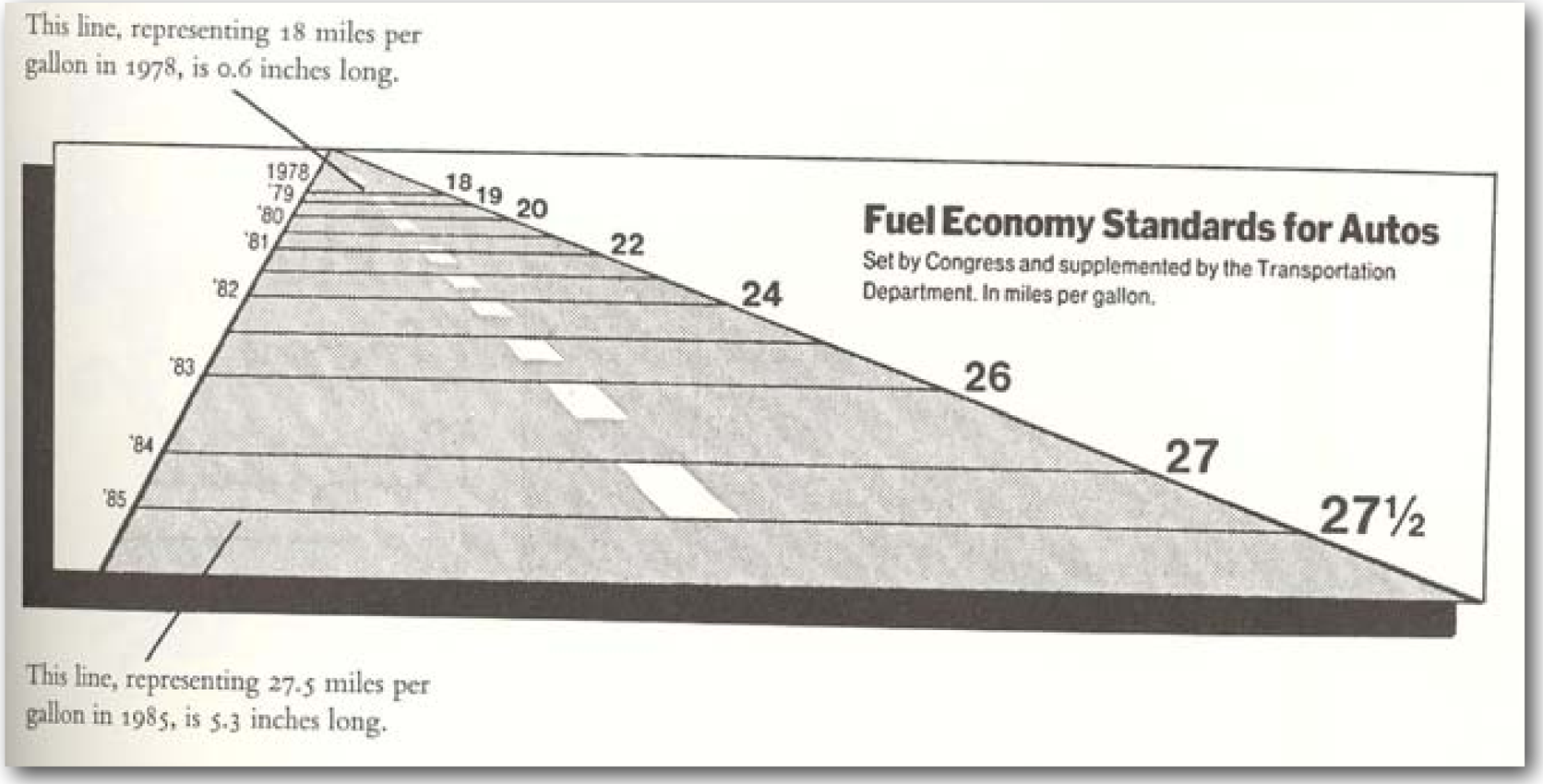
Global Warming - Frame the Data



Mother Jones

The Lie Factor

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

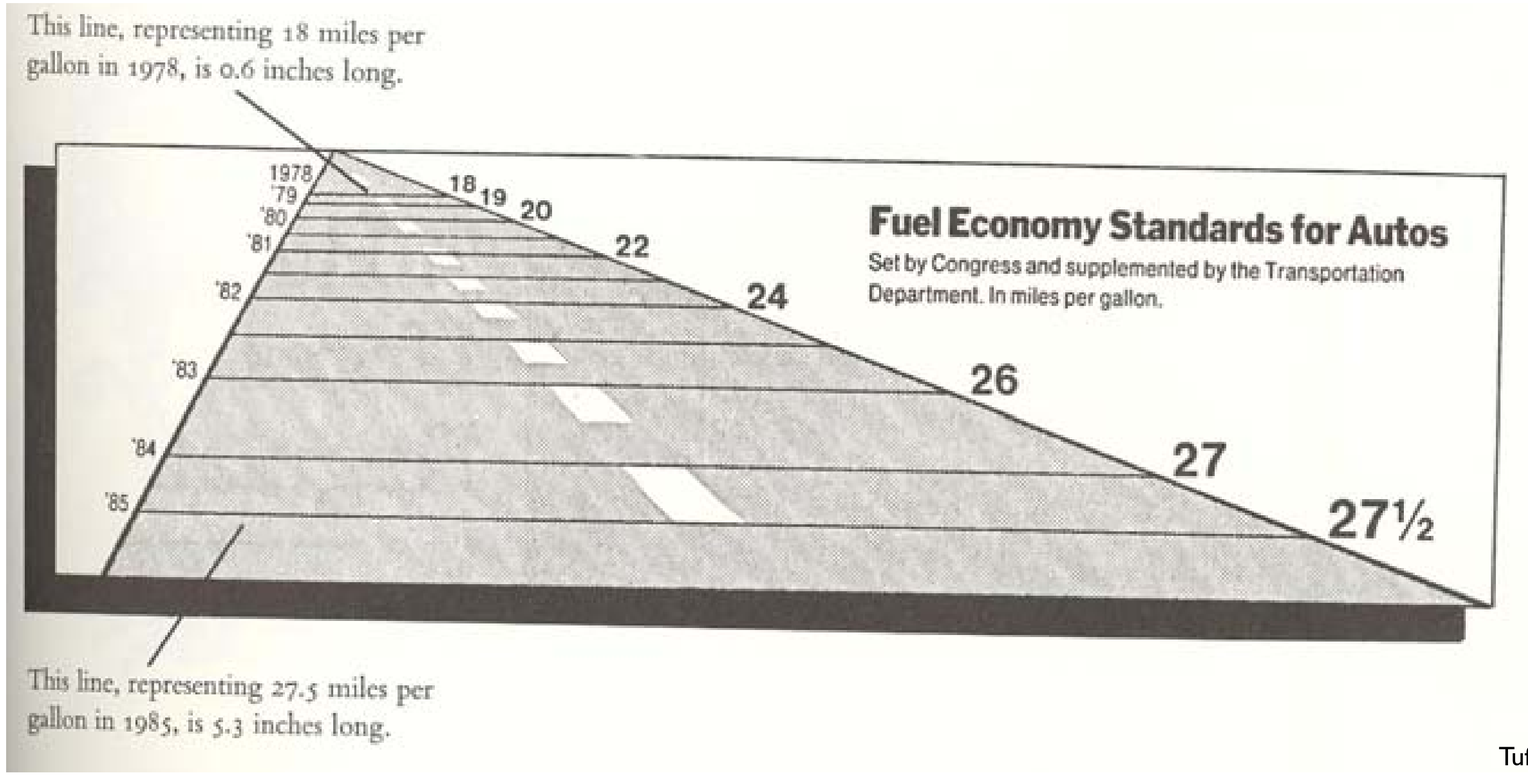


Tufte, VDQI

The Lie Factor

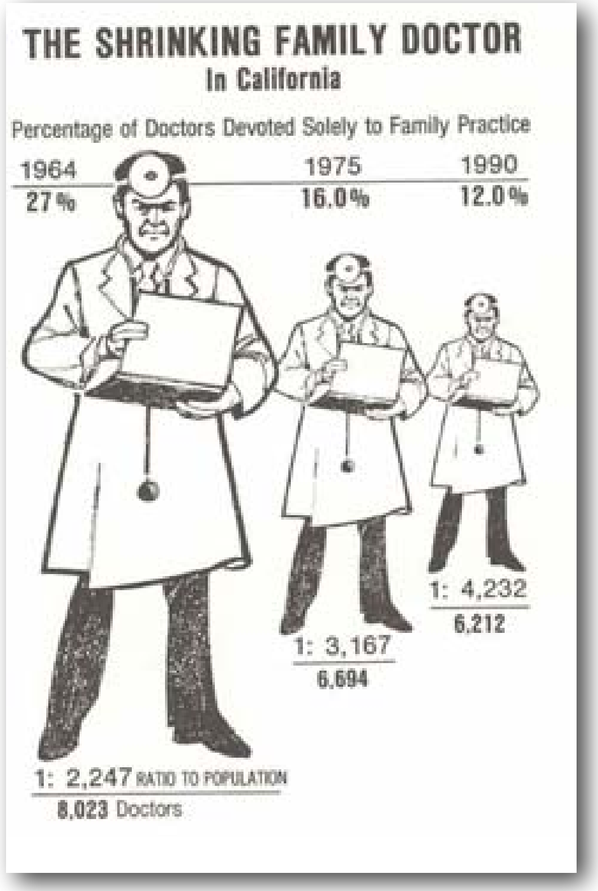
$$\frac{5.3 - 0.6}{0.6} \Big/ \frac{27.5 - 18}{18} = 14.8$$

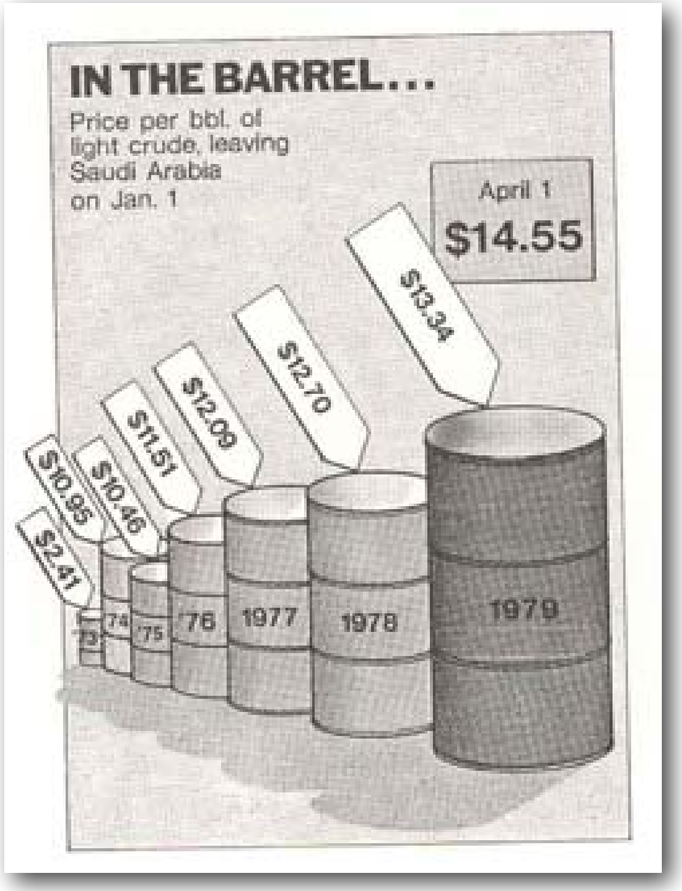
(Size of effect in graphic)/(size of effect in data)



Tufte, VDQI

The Lie Factor





Tufte, VDQI

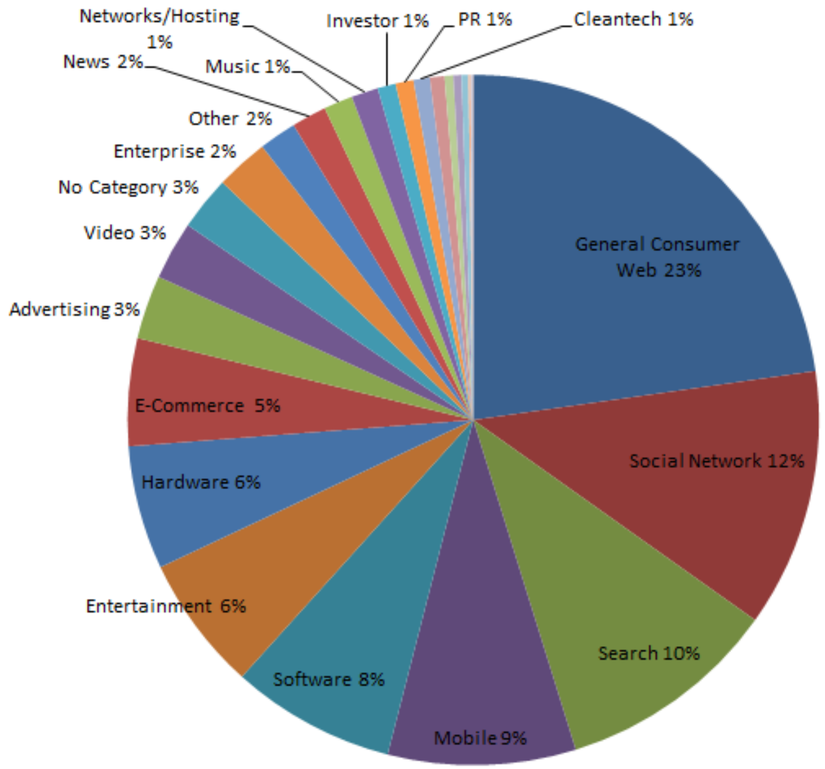
Tufte's Integrity Principles

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")

Death to Pie Charts



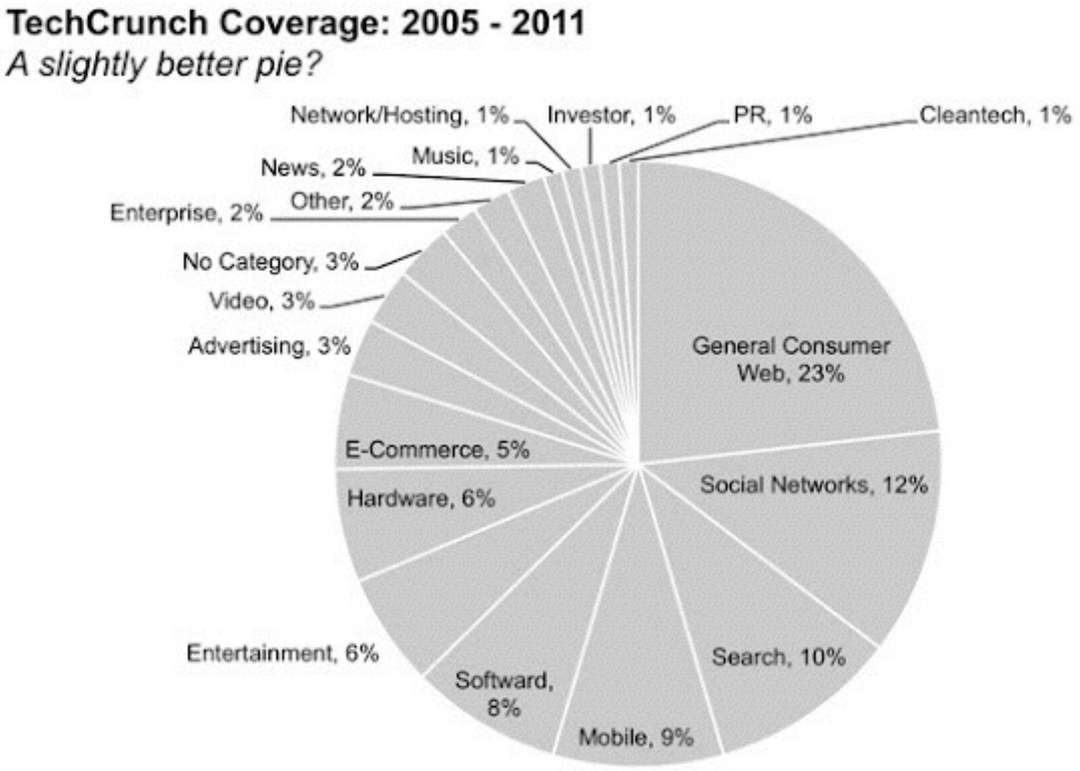
Share of coverage on TechCrunch

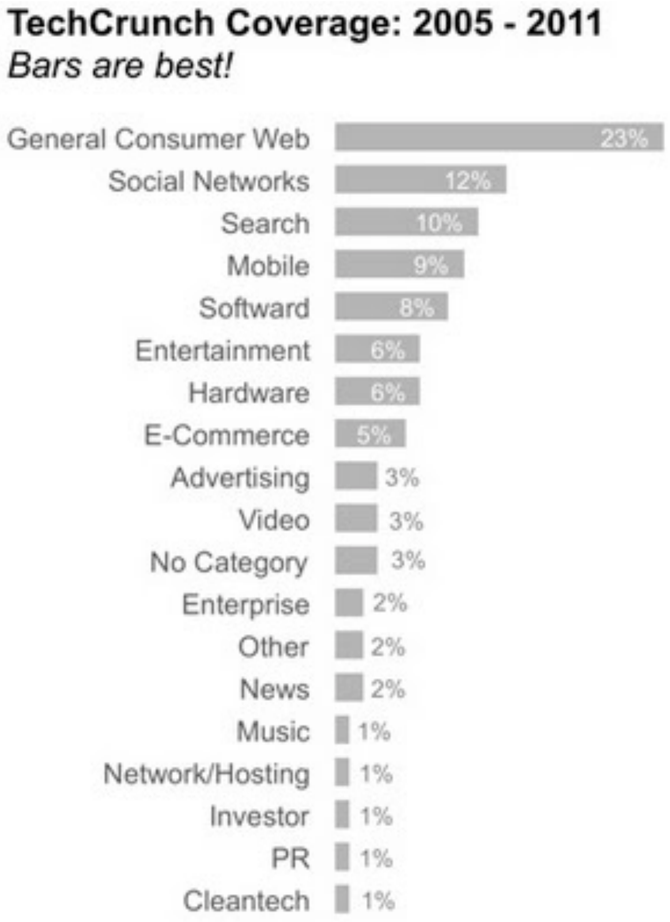
"I hate pie charts.
I mean, really hate them."

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

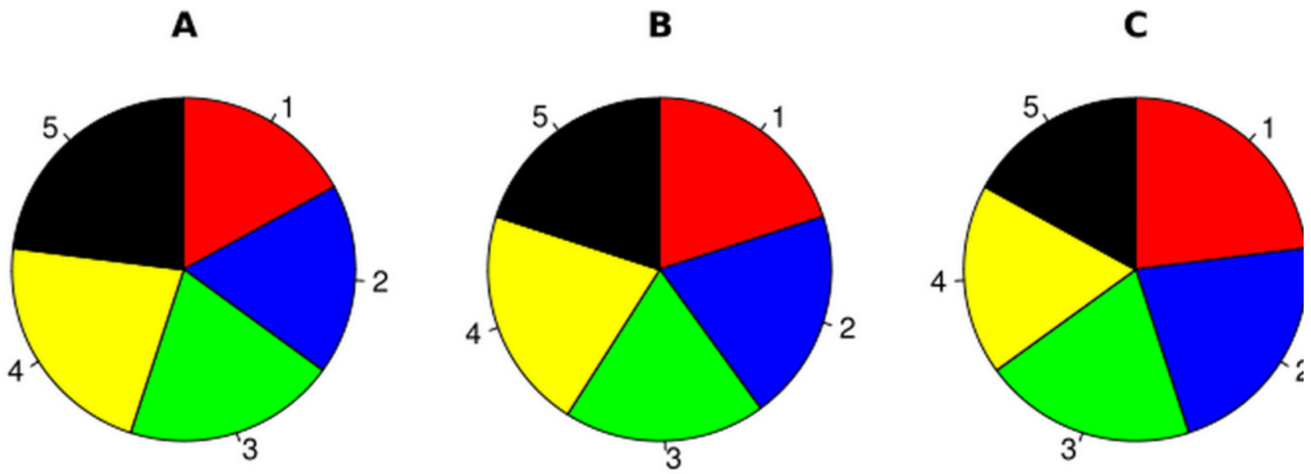
Cole Nussbaumer

Redesign

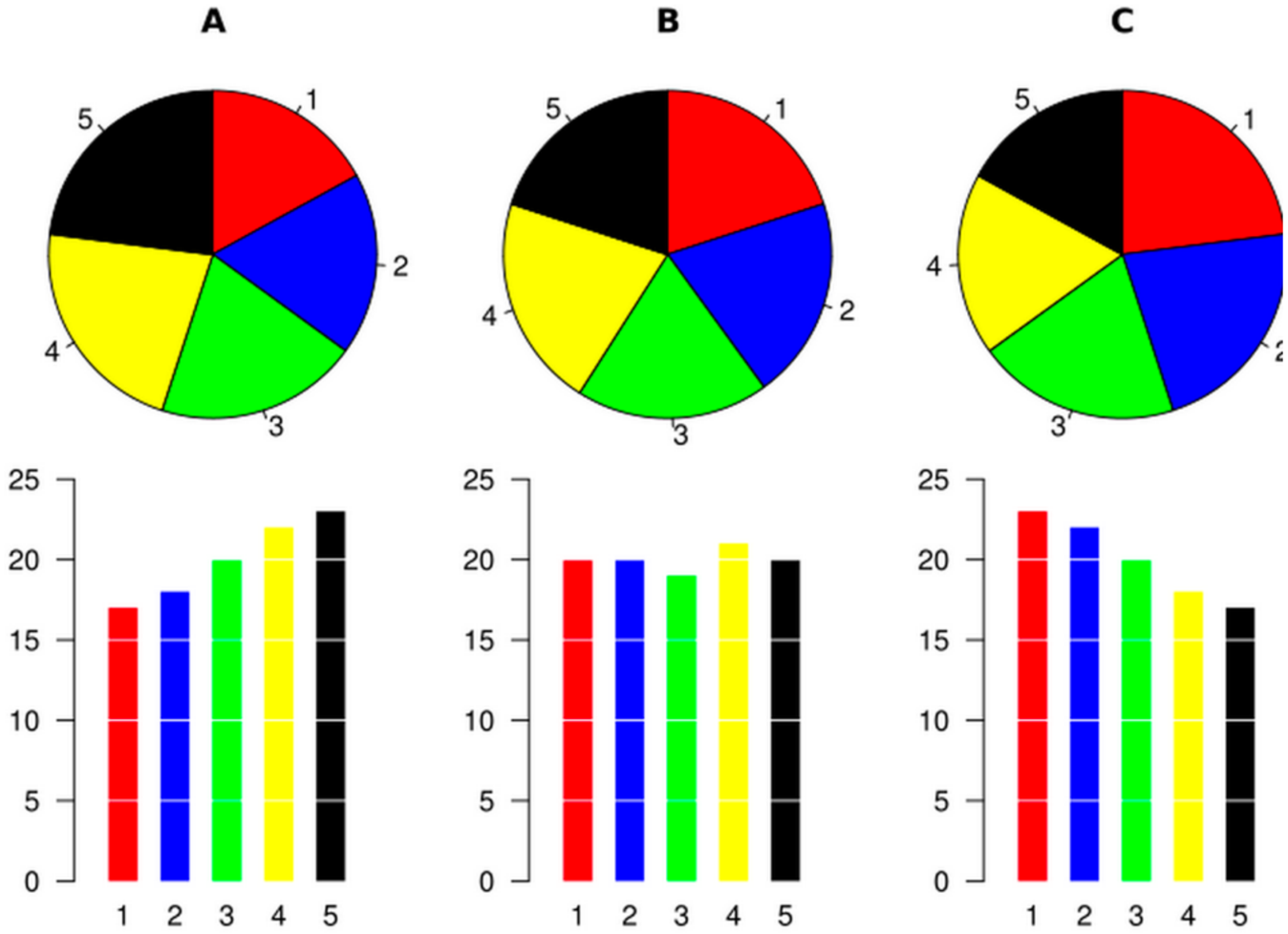

TechCrunch Coverage: 2005 - 2011
A slightly better pie?
Network/Hosting, 1%
Investor, 1%
PR, 1%
Cleantech, 1%
News, 2%
Music, 1%
Enterprise, 2%
Other, 2%
No Category, 3%
Video, 3%
Advertising, 3%
General Consumer Web, 23%
E-Commerce, 5%
Social Networks, 12%
Hardware, 6%
Entertainment, 6%
Softward, 8%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

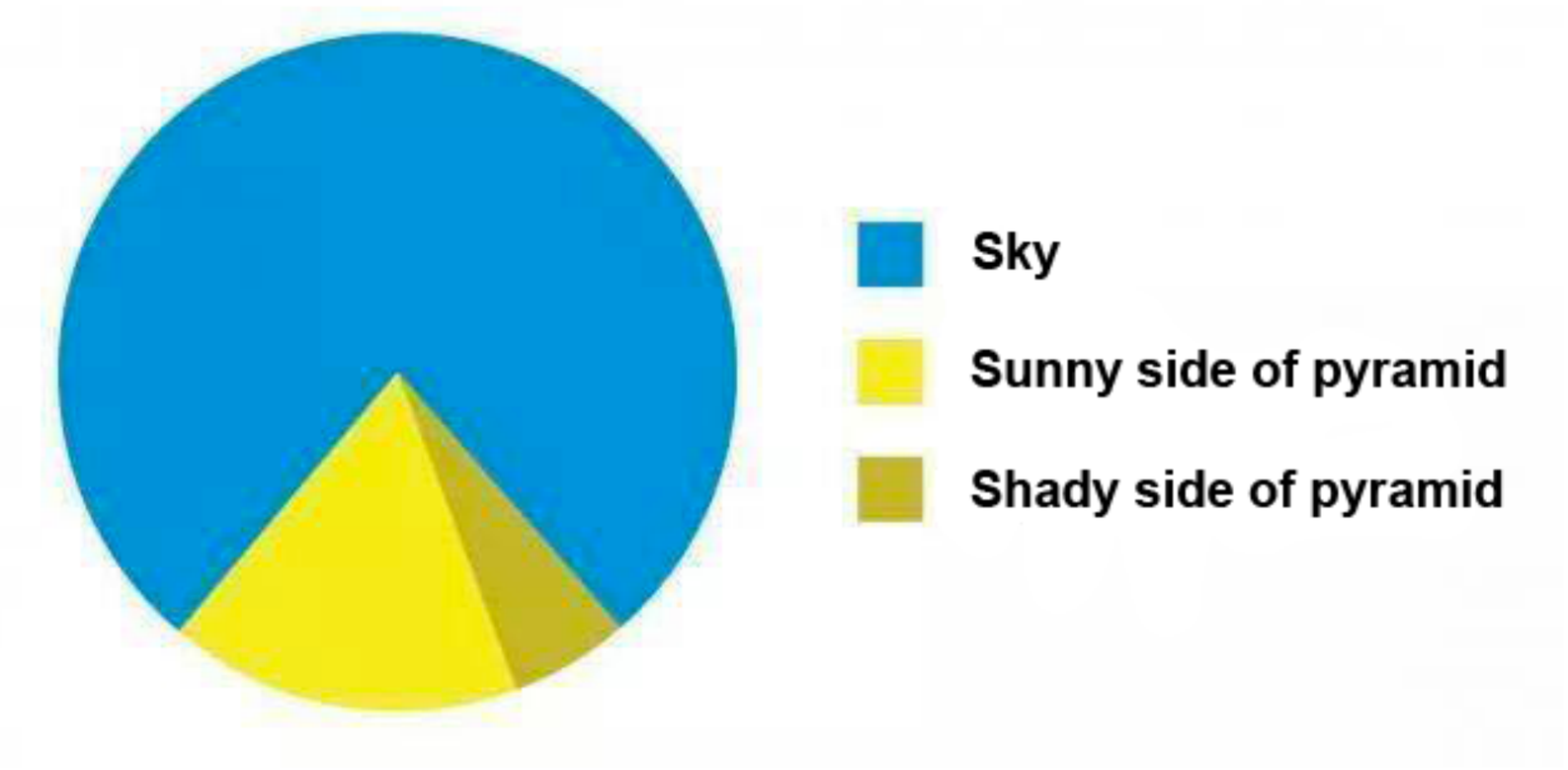
Can you spot the differences?



Can you spot the differences?



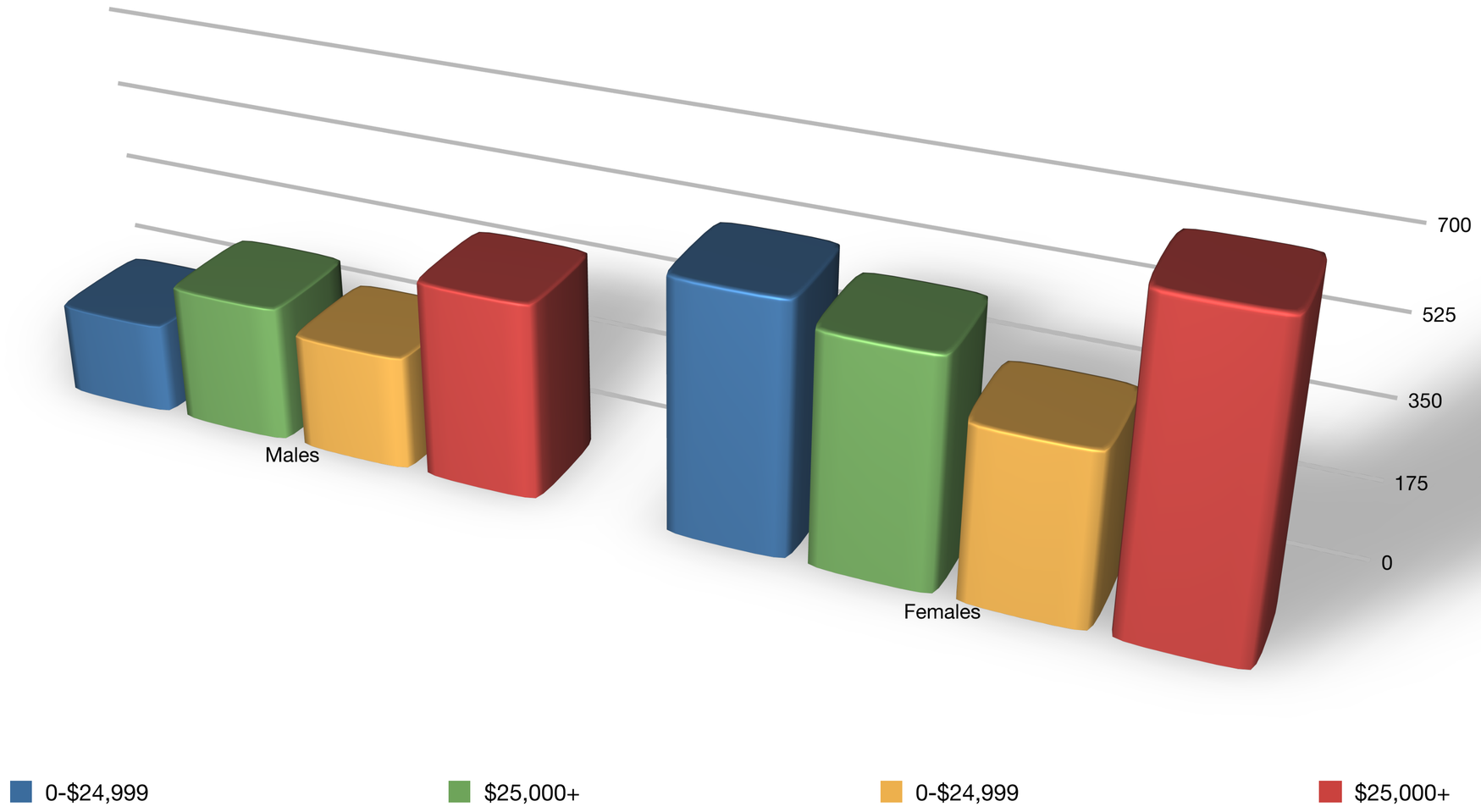
My favorite pie chart



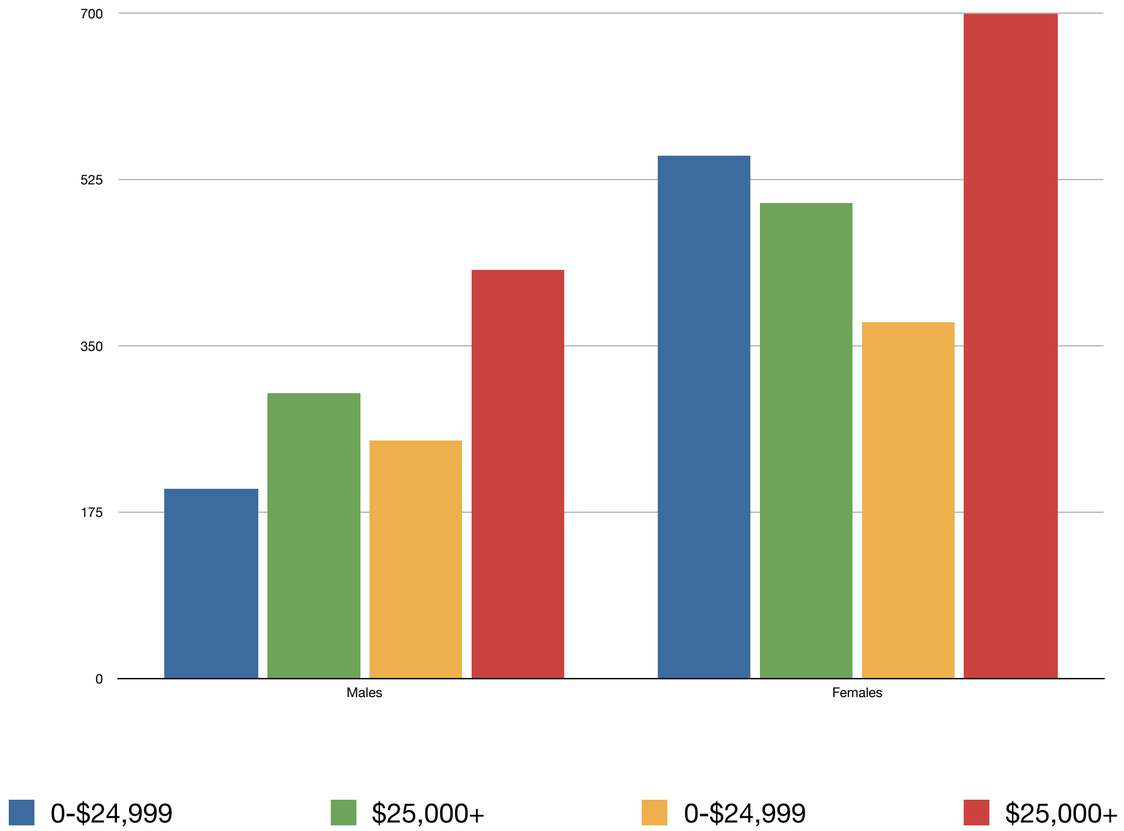
My second favorite pie chart

Visualization Design Principles

Maximize Data-Ink Ratio

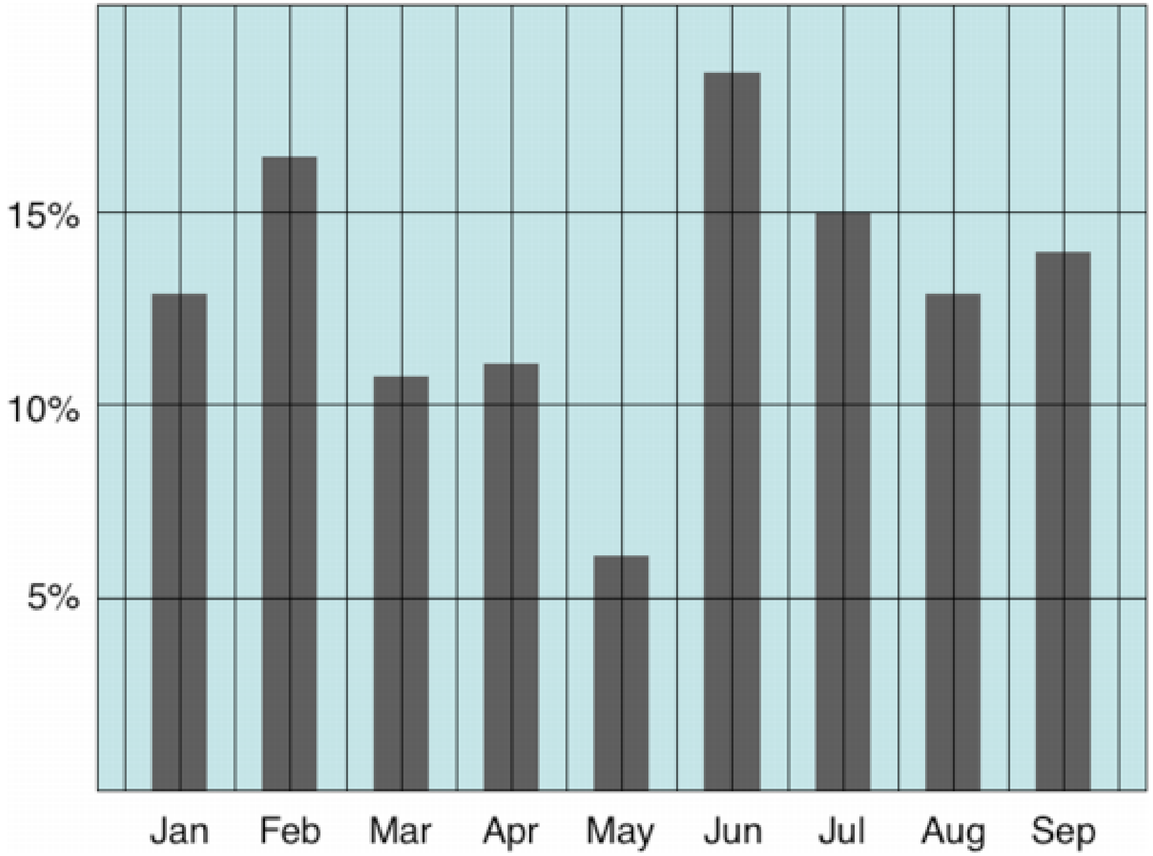


Maximize Data-Ink Ratio



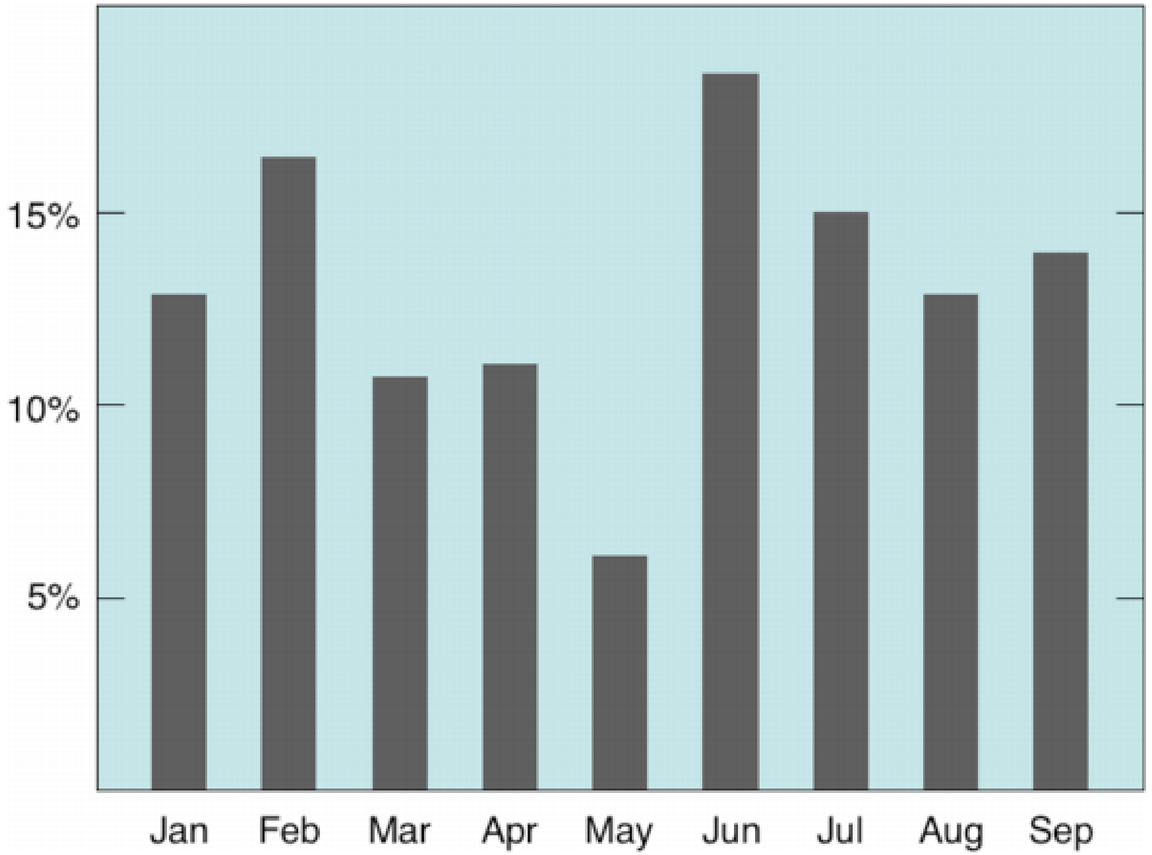
Avoid Chartjunk

Extraneous visual elements that distract from the message



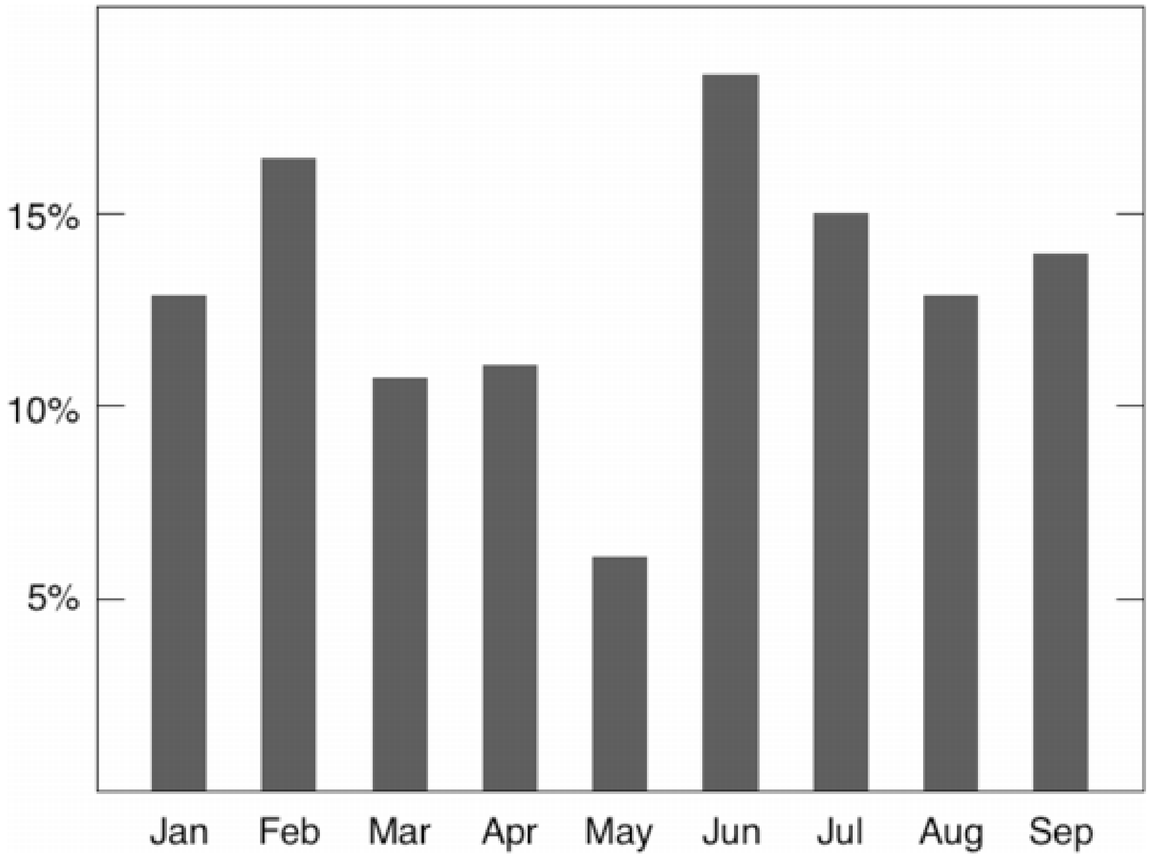
ongoing, Tim Brey

Avoid Chartjunk



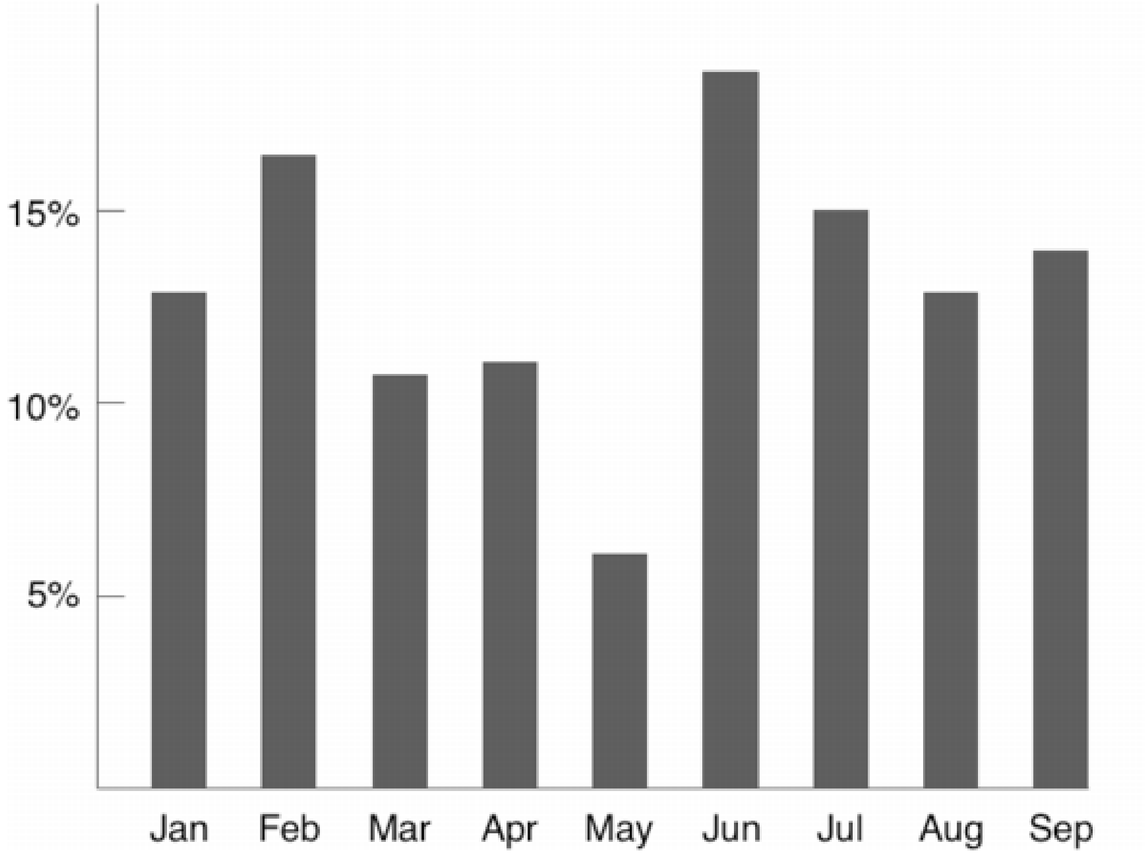
ongoing, Tim Brey

Avoid Chartjunk



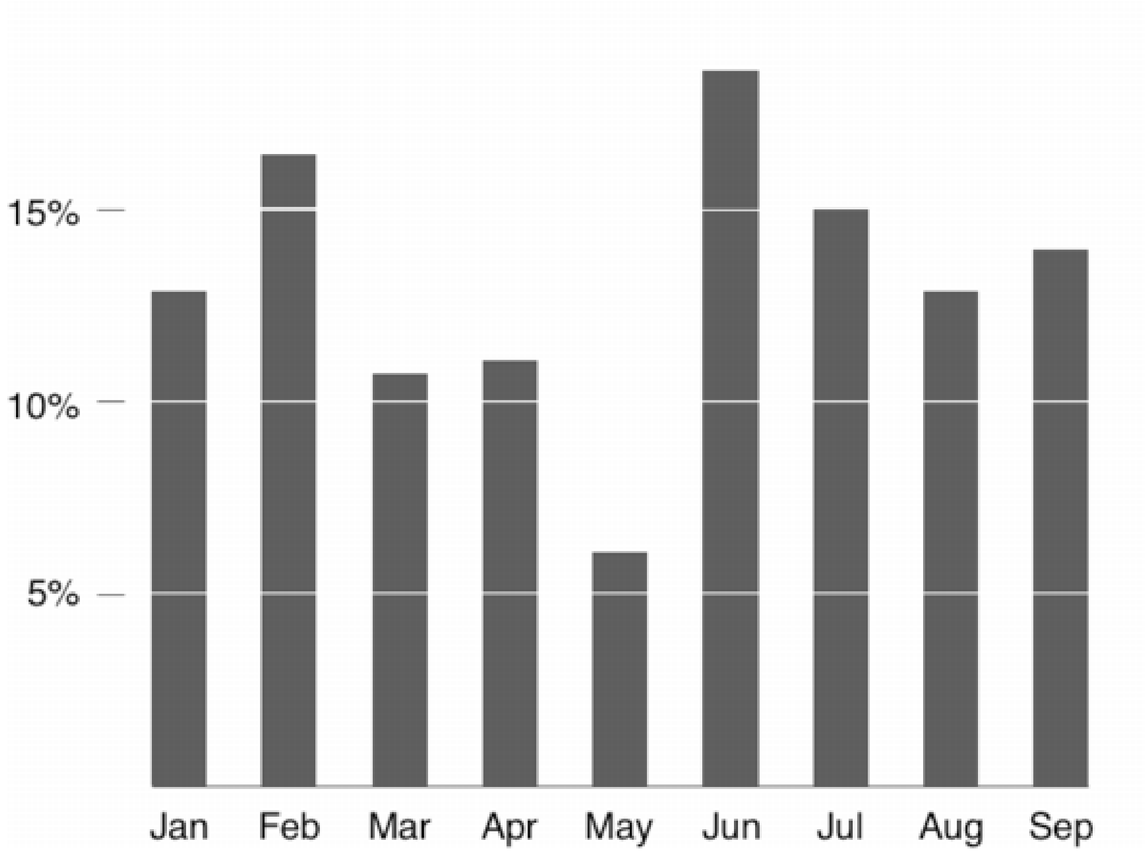
ongoing, Tim Brey

Avoid Chartjunk



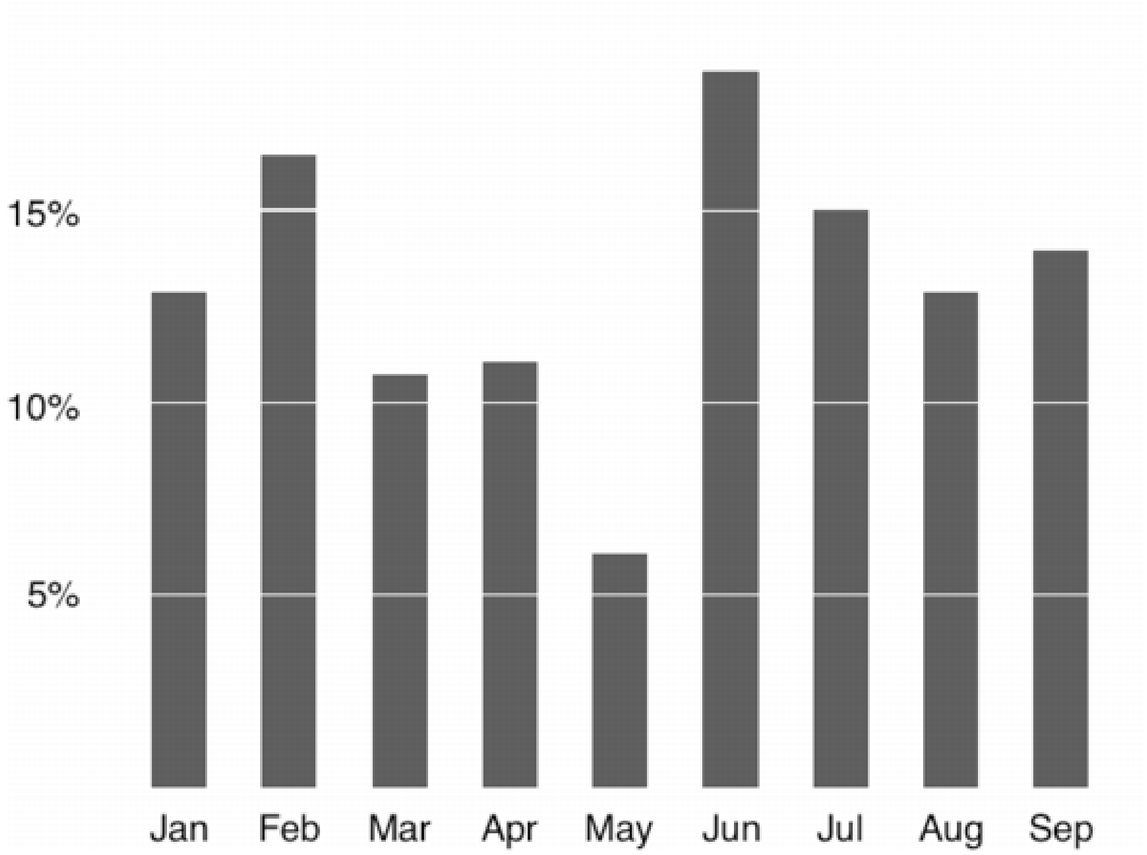
ongoing, Tim Brey

Avoid Chartjunk



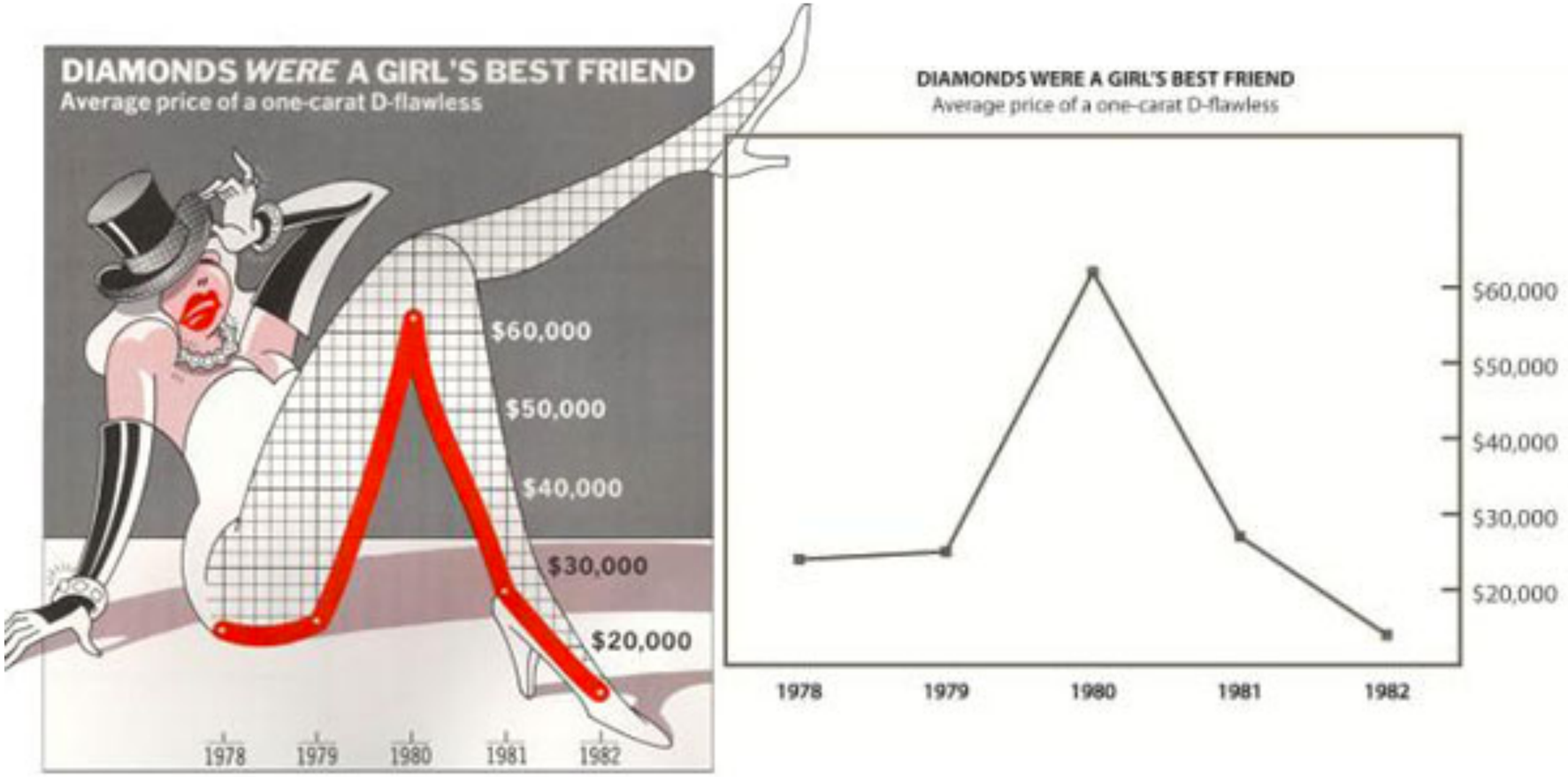
ongoing, Tim Brey

Avoid Chartjunk



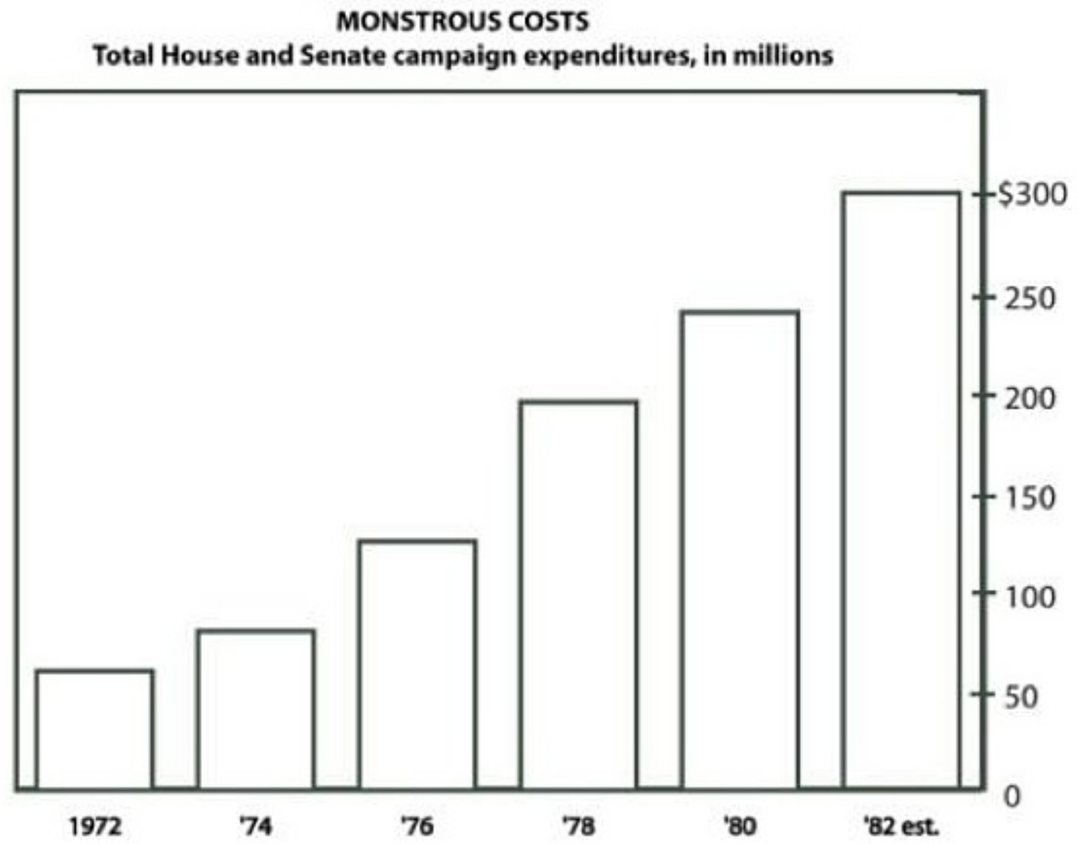
ongoing, Tim Brey

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.

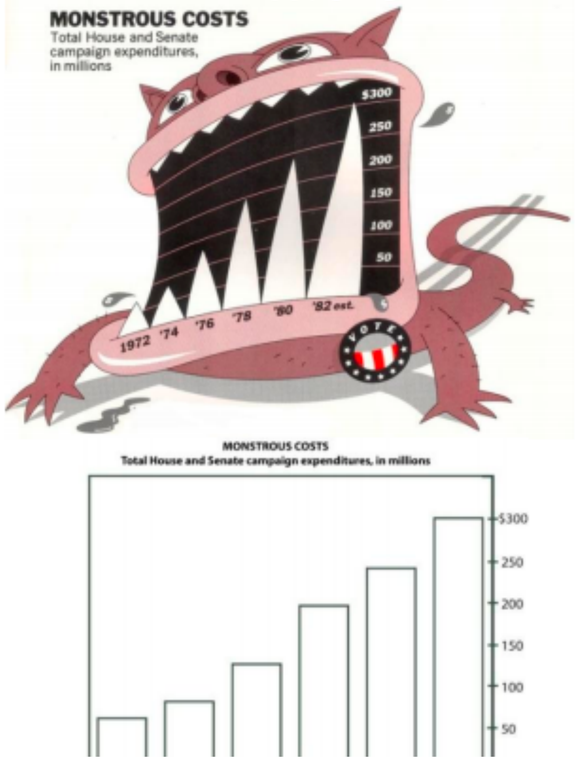
Author Keywords
Charts, information visualization, imagery, memorability.

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 significant difference between plain and image charts for interactive **interpretation accuracy**
2. No significant 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!

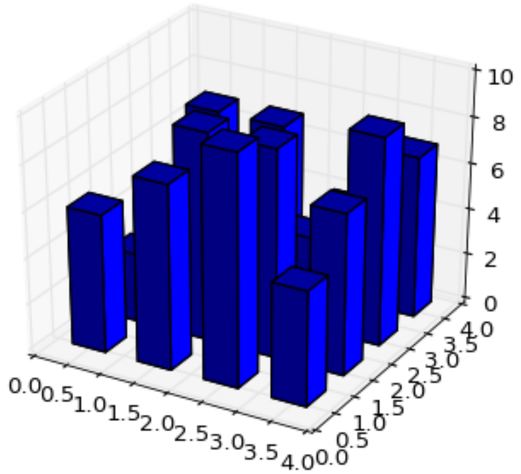
PROS

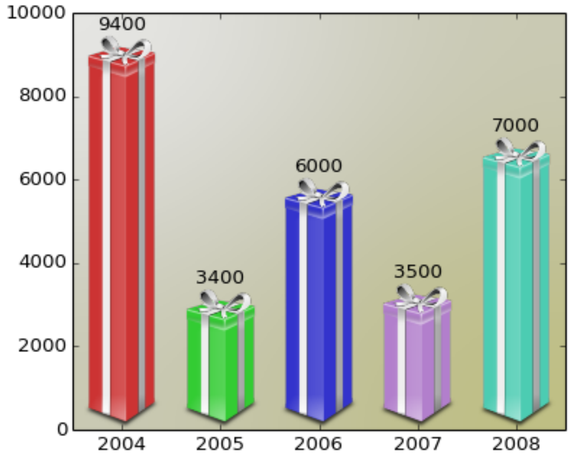
- persuasion
- memorability
- engagement

CONS

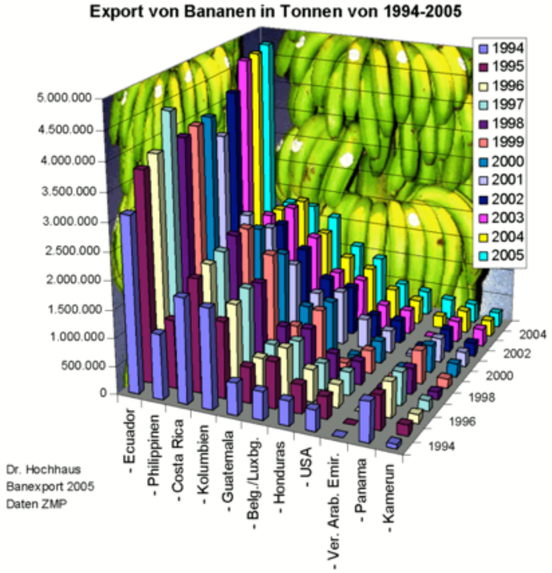
- unbiased analysis
- trustworthiness
- interpretability
- space efficiency

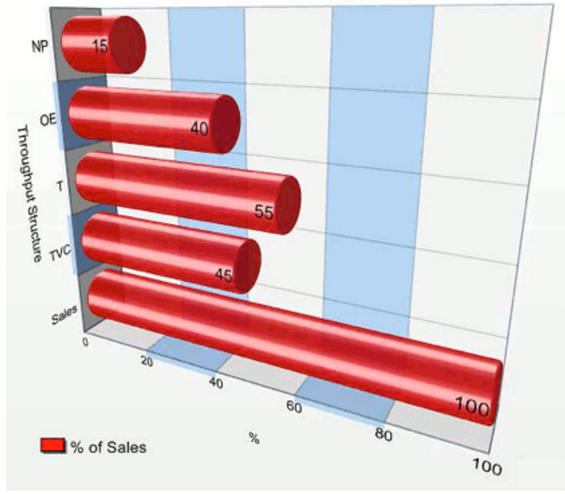
Don’t





matplotlib gallery





Excel Charts Blog

Tasks

Why are we using Visualization?

Domain and Abstract Tasks

Infinite numbers of domain tasks

Can be broken down into simpler abstract tasks

We know how to address the abstract tasks!

Identify task - data combination: solutions probably exist

Tasks

Analyze

- high-level choices
- consume vs produce

Search

- find a known/unknown item

Query

- find out about characteristics of item
- by itself or relative to others

Example 1

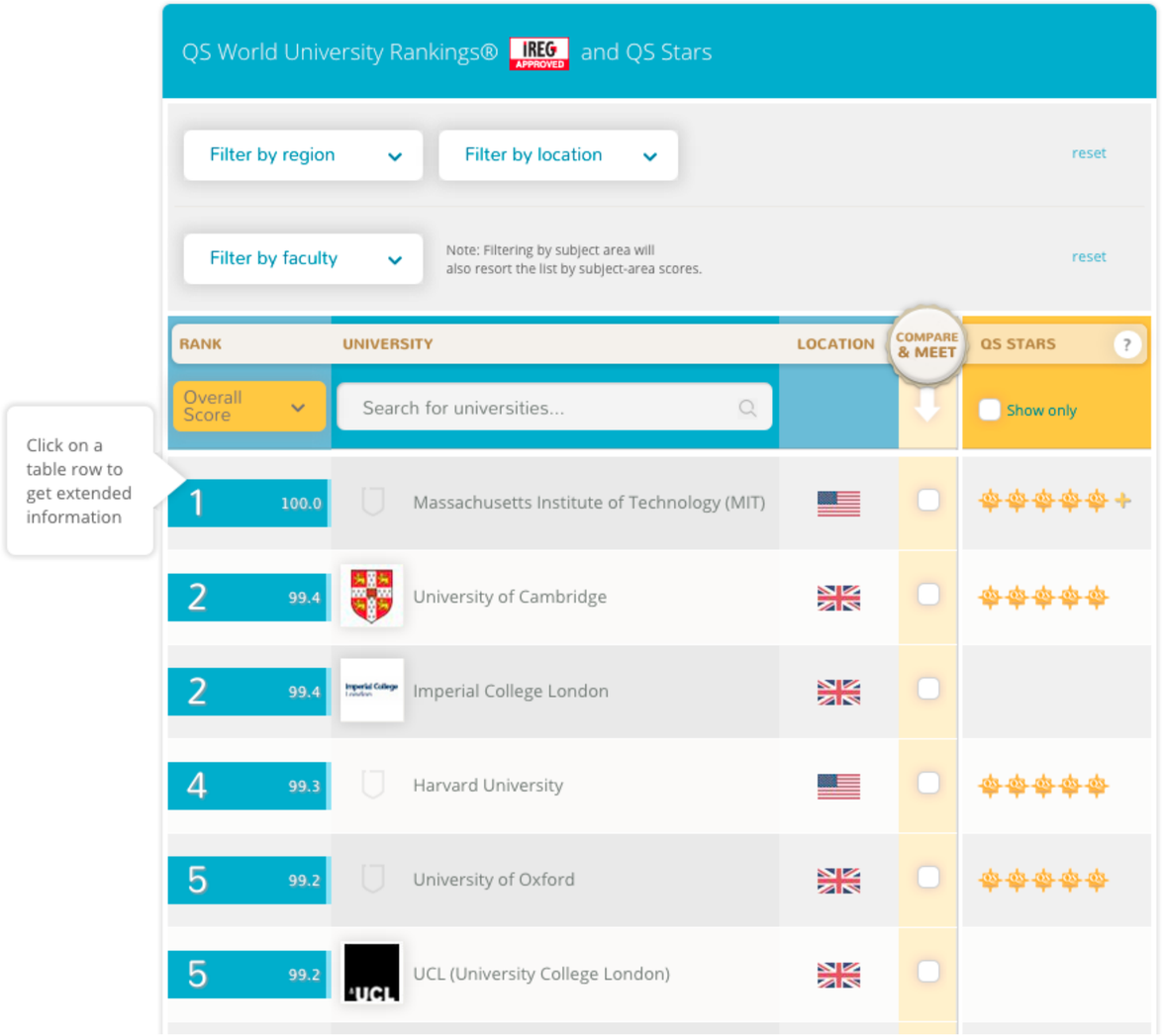
Find good universities with a high faculty student ratio.

Identify high-ranked universities

In this subset: **compare** universities & **identify** high faculty student ratio

OR

Derive a ranking with a high weight for faculty student ratio

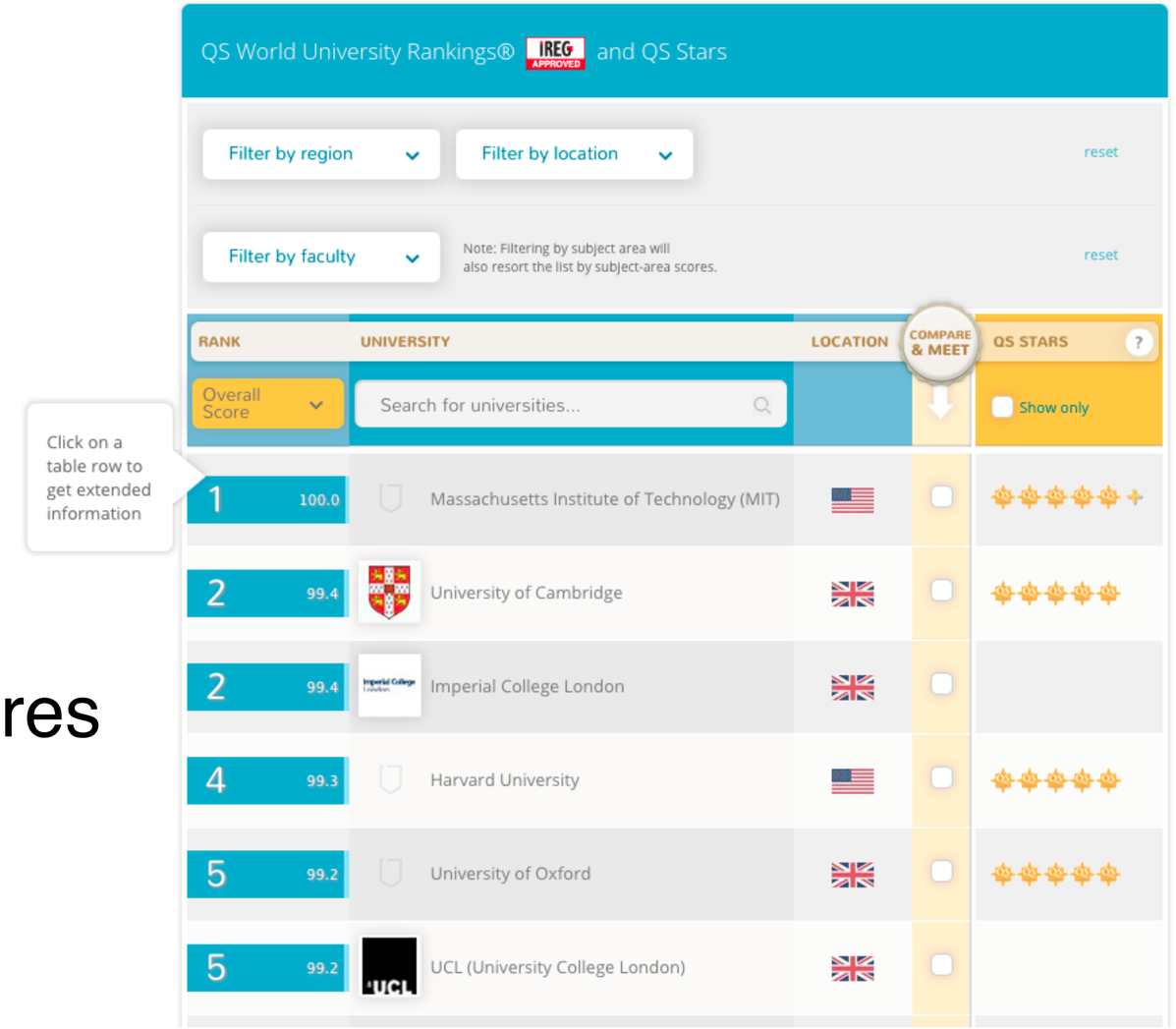


Example 2

Contrast Harvard's reputation scores with MIT's

Match up Harvard with Yale

First, **find** Harvard and Yale, then **compare** their (two) reputation scores

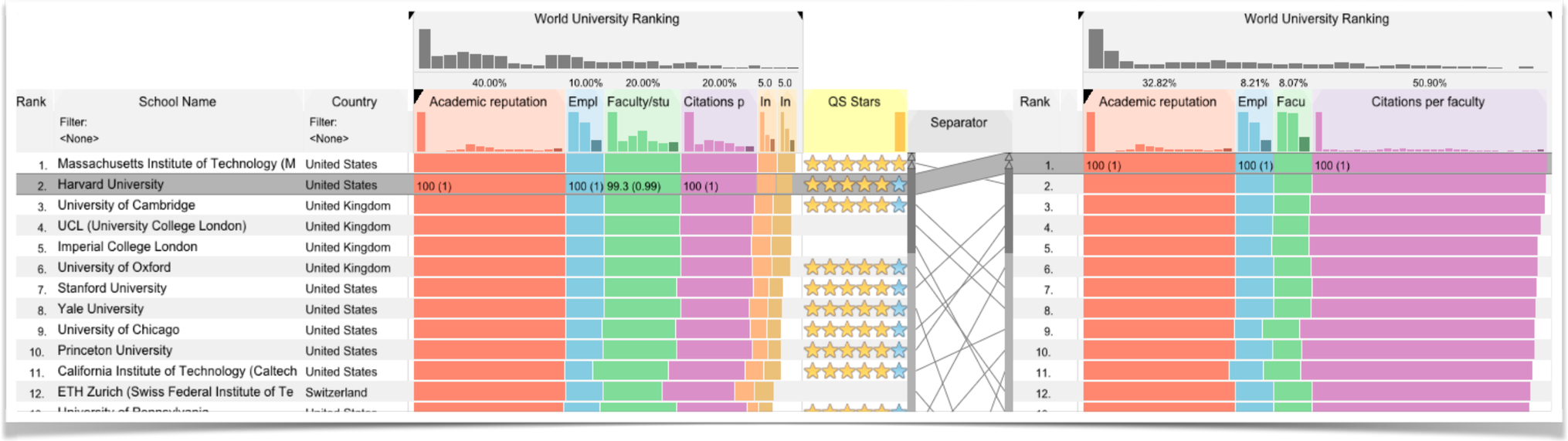


Example 3

Find a combination of weights and parameters where Harvard is better than MIT

Produce a new dataset by **deriving** from the input parameters

Result


World University Ranking
40.00%
10.00%
20.00%
20.00%
5.0
5.0
Rank
School Name
Country
Filter:
<None>
Filter:
<None>
Academic reputation
Empl
Faculty/stu
Citations p
In
In
QS Stars
Separator
Rank
World University Ranking
32.82%
8.21%
8.07%
50.90%
Academic reputation
Empl
Facu
Citations per faculty
1. Massachusetts Institute of Technology (M United States
2. Harvard University United States 100 (1) 100 (1) 99.3 (0.99) 100 (1)
3. University of Cambridge United Kingdom
4. UCL (University College London) United Kingdom
5. Imperial College London United Kingdom
6. University of Oxford United Kingdom
7. Stanford University United States
8. Yale University United States
9. University of Chicago United States
10. Princeton University United States
11. California Institute of Technology (Caltech United States
12. ETH Zurich (Swiss Federal Institute of Te Switzerland
1. 100 (1) 100 (1) 100 (1)
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.

High-level actions: Analyze

Consume

discover vs present

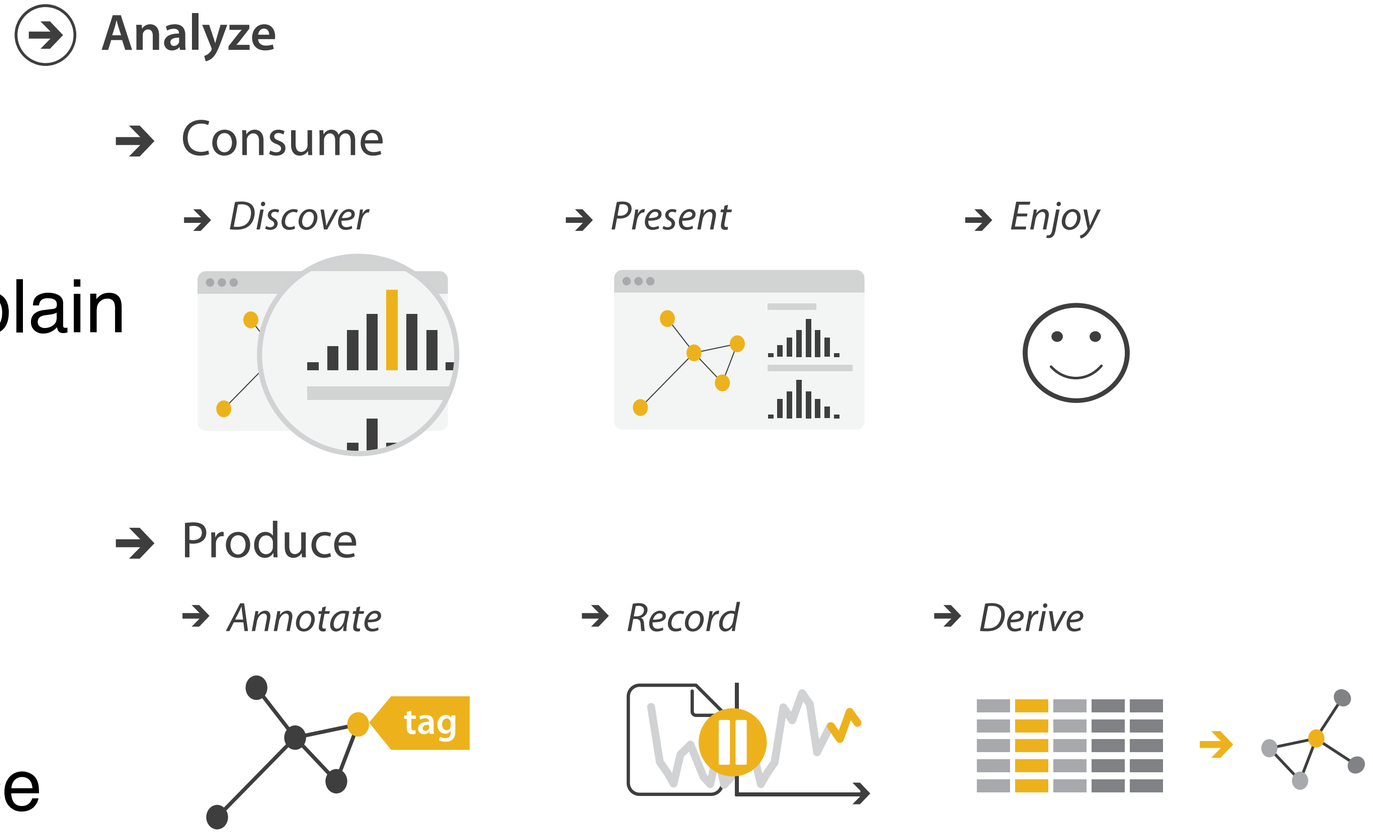
classic split: explore vs explain

enjoy: casual, social

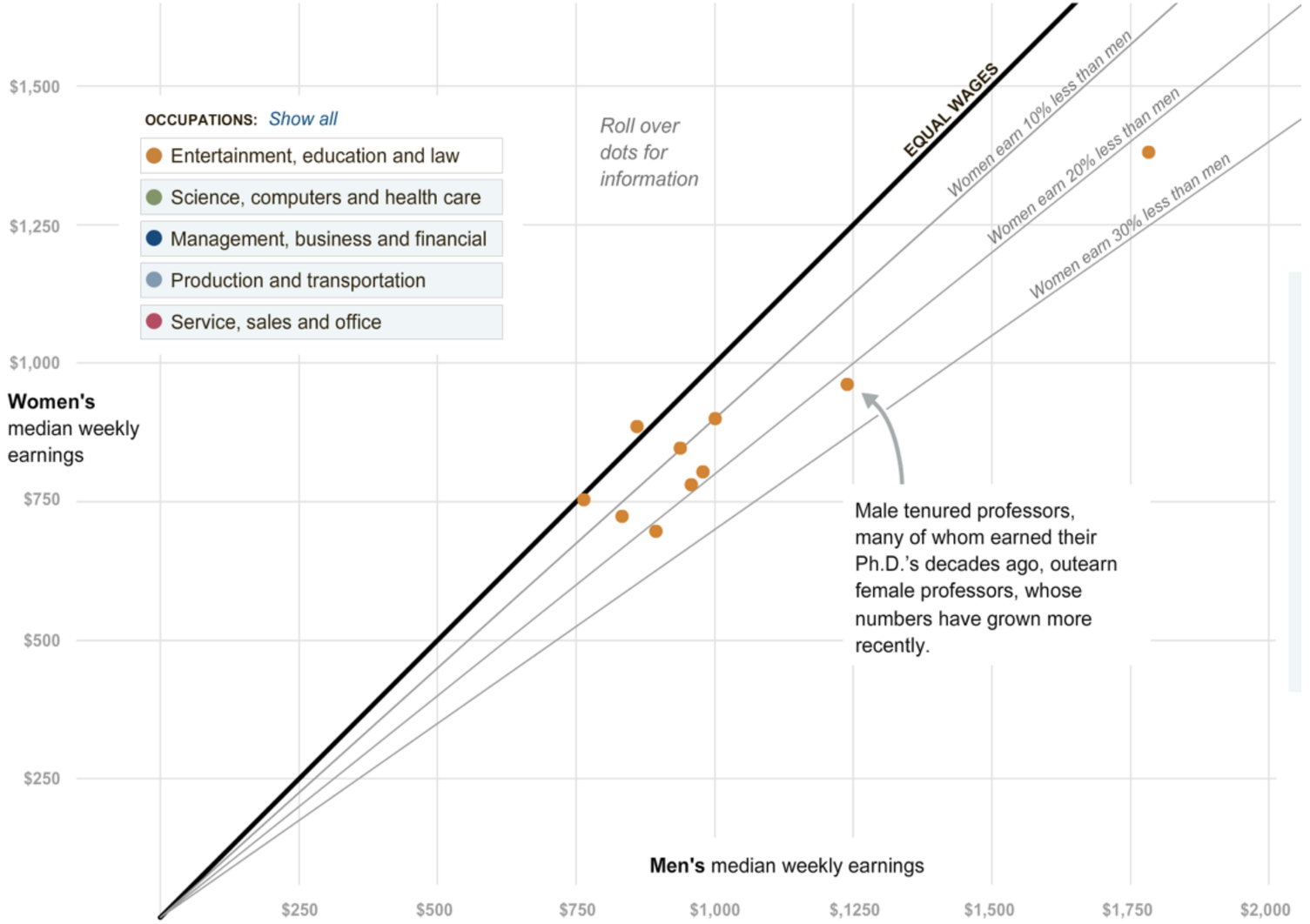
Produce

Annotate, record

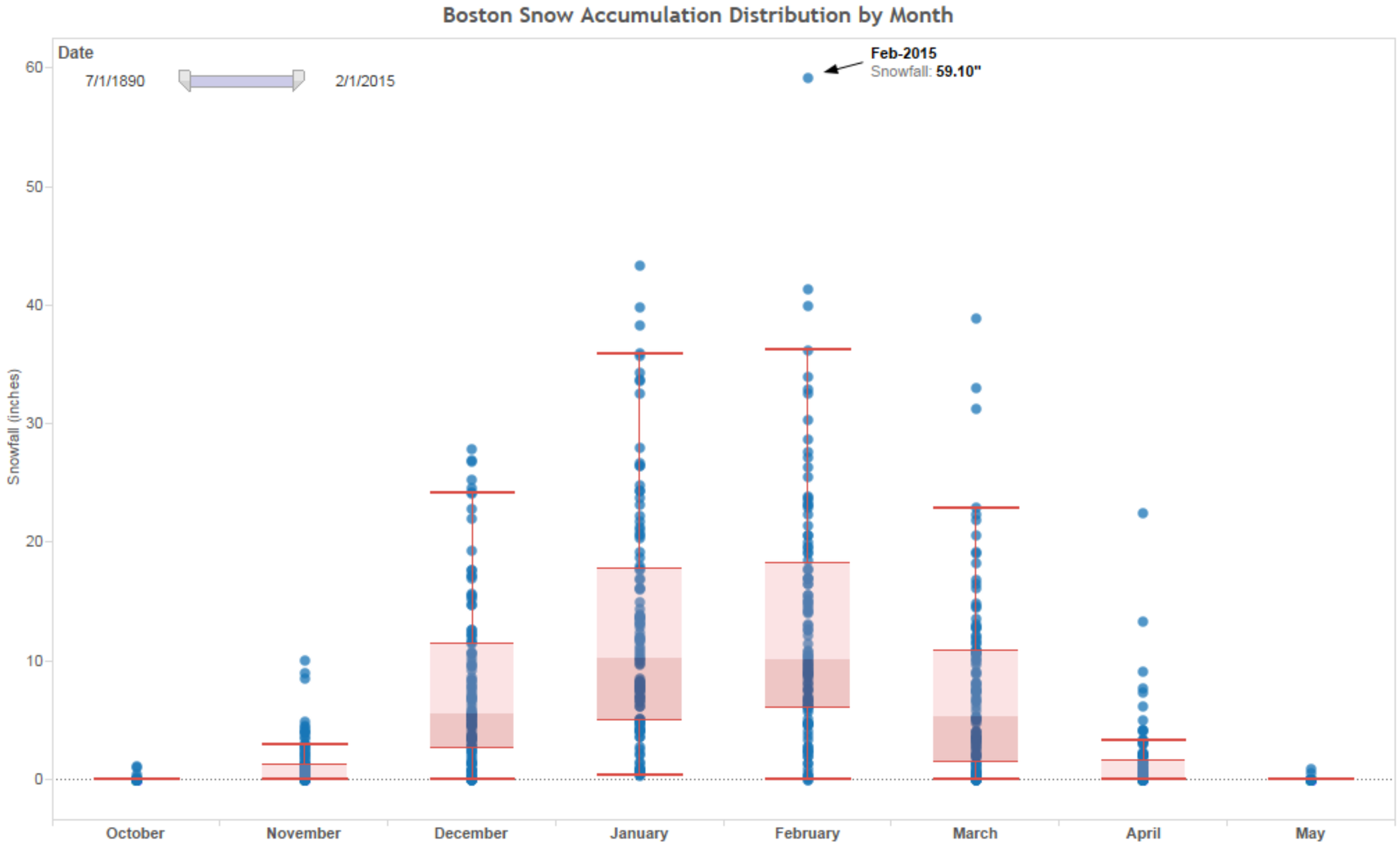
Derive: crucial design choice



Example: Annotate

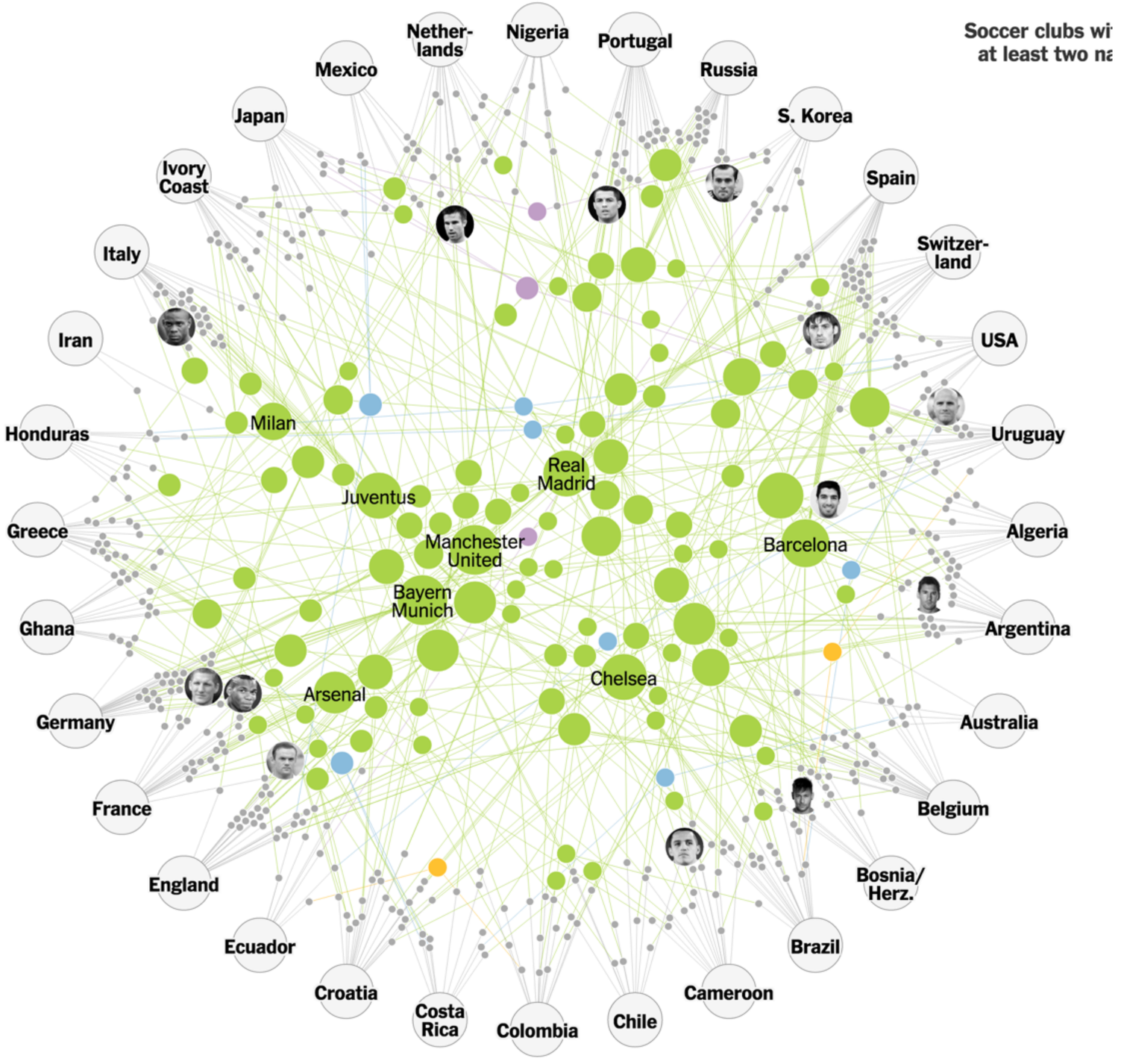


Example: Derive



Example: Derive

	Country	Club	Club Continent
Ronaldo	Portugal	Real Madrid	Europe
Lahm	Germany	Bayern München	Europe
Robben	Netherlands	Bayern München	Europe
Khedira	Germany	Real Madrid	Europe
Phogba	Italy	Juventus	Europe
Messi	Argentina	Barcelona	Europe


Soccer clubs wi
at least two na
Nether-
lands
Nigeria
Portugal
Mexico
Russia
Japan
S. Korea
Ivory
Coast
Spain
Italy
Switzer-
land
Iran
USA
Honduras
Milan
Uruguay
Real
Madrid
Juventus
Greece
Algeria
Manchester
United
Barcelona
Bayern
Munich
Ghana
Argentina
Chelsea
Arsenal
Germany
Australia
France
Belgium
England
Bosnia/
Herz.
Ecuador
Brazil
Croatia
Cameroon
Costa
Rica
Colombia
Chile

Actions: Mid-level search, low-level query

what does user know?

target, location

how much of the data matters?

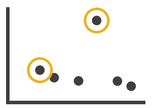
one, some, all

Search

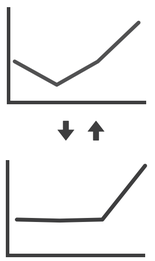
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

→ Identify

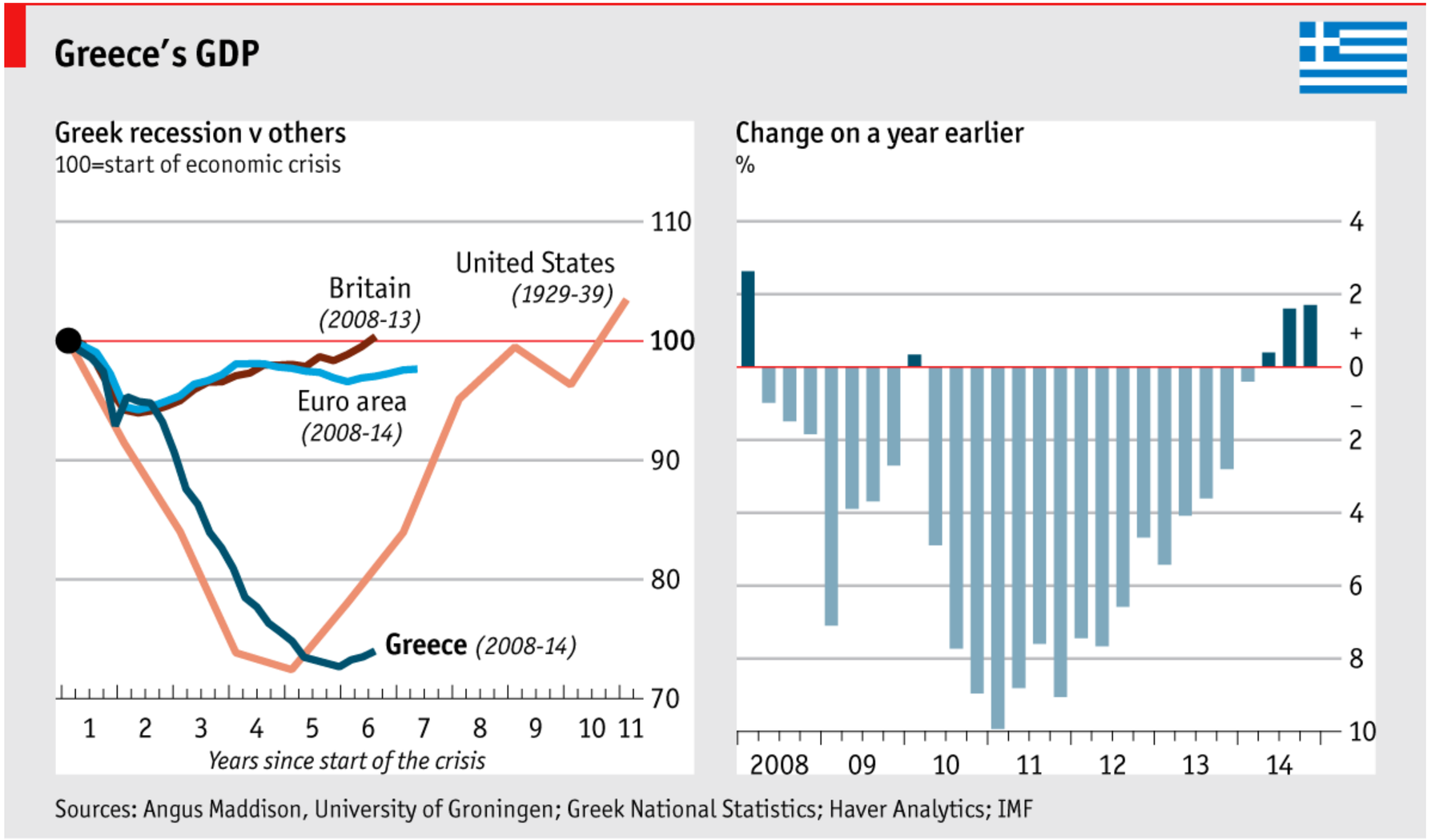


→ Compare

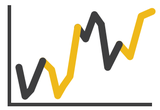
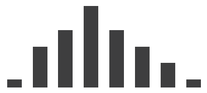
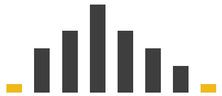
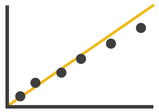
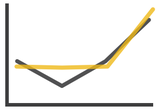
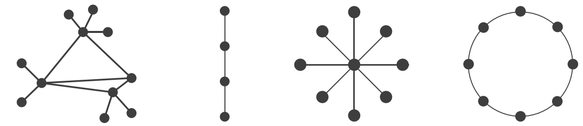
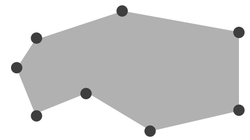


→ Summarize

Example Compare (& Derive)



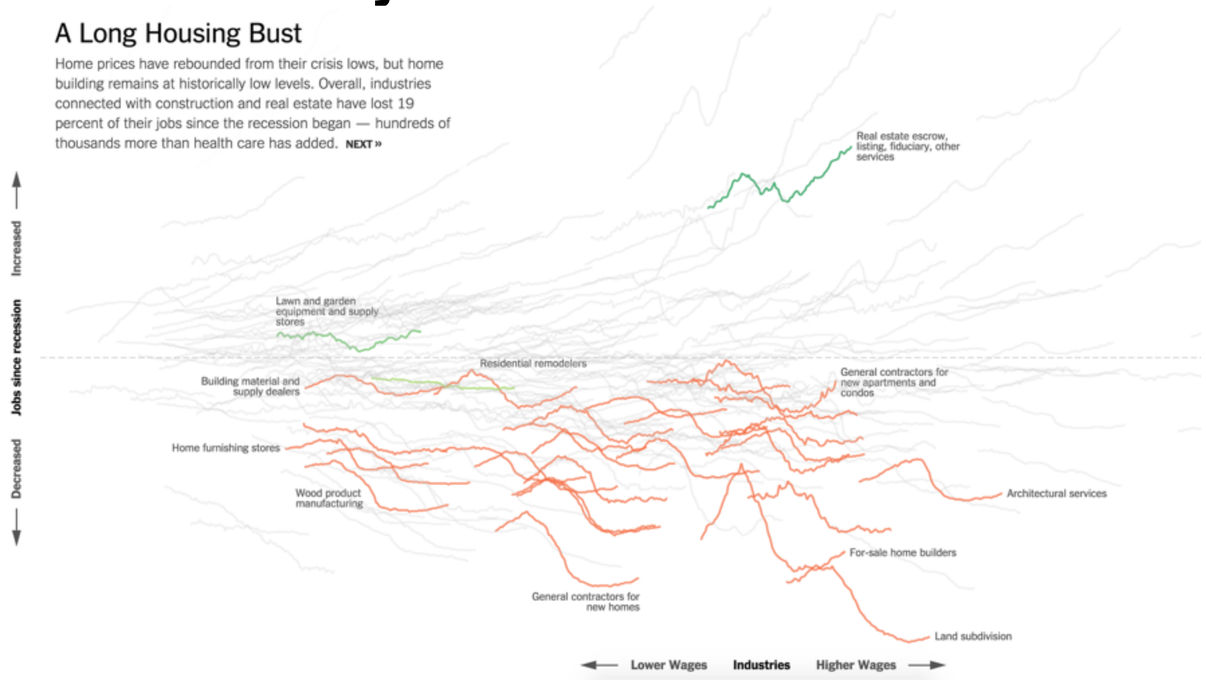
Why: Targets

- ALL DATA
  - Trends
    
  - Outliers
    
  - Features
    
- ATTRIBUTES
  - One
    - *Distribution*
      
    - *Extremes*
      
  - Many
    - *Dependency*
      
    - *Correlation*
      
    - *Similarity*
      
- NETWORK DATA
  - Topology
    
  - *Paths*
    
- SPATIAL DATA
  - Shape
    

Examples

Trends: How did the job market develop since the recession overall?

Outliers: Looking at real estate related jobs

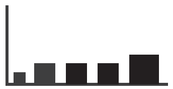


How? A Preview

Encode

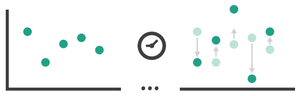
- **Arrange**
  - Express
  - Separate
  - Order
  - Align
  - Use



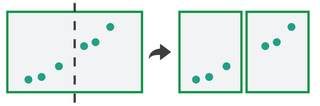

Manipulate

- **Change**
- **Select**
- **Navigate**



Facet

- **Juxtapose**
- **Partition**
- **Superimpose**


Reduce

- **Filter**
- **Aggregate**
- **Embed**