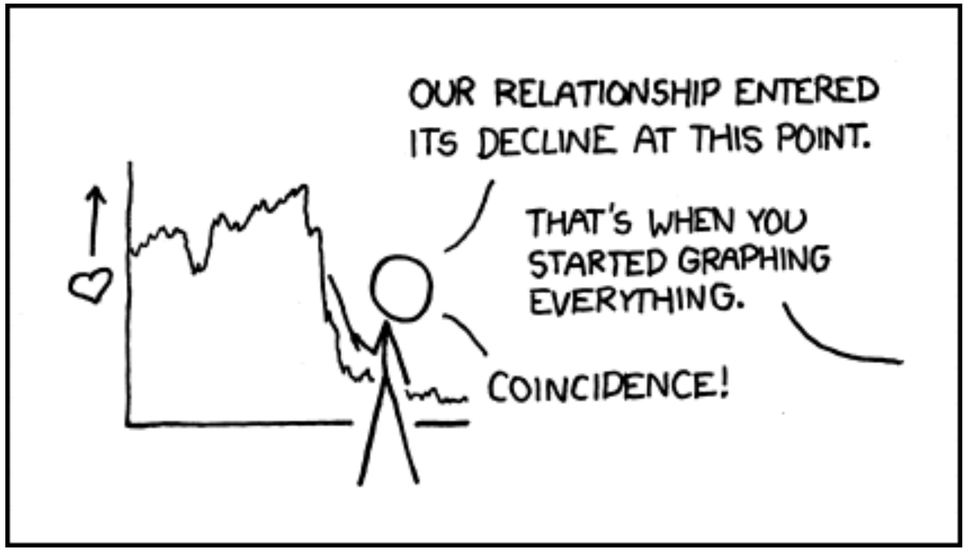
CS171 Visualization

Alexander Lex
alex@seas.harvard.edu

Design Guidelines
Tasks



[xkcd]

HARVARD
School of Engineering
and Applied Sciences

Next Week

Lecture 7: Homework 2 Design Studio

Lecture 8: Interaction
Guest Lecture, Jean-Daniel Fekete (INRIA)

Sections: D3 & JS: Data Structures, Layouts

Last Tuesday

The Visualization Alphabet: Marks and Channels

How can I visually represent two numbers, e.g.,

4 and 8

Marks & Channels

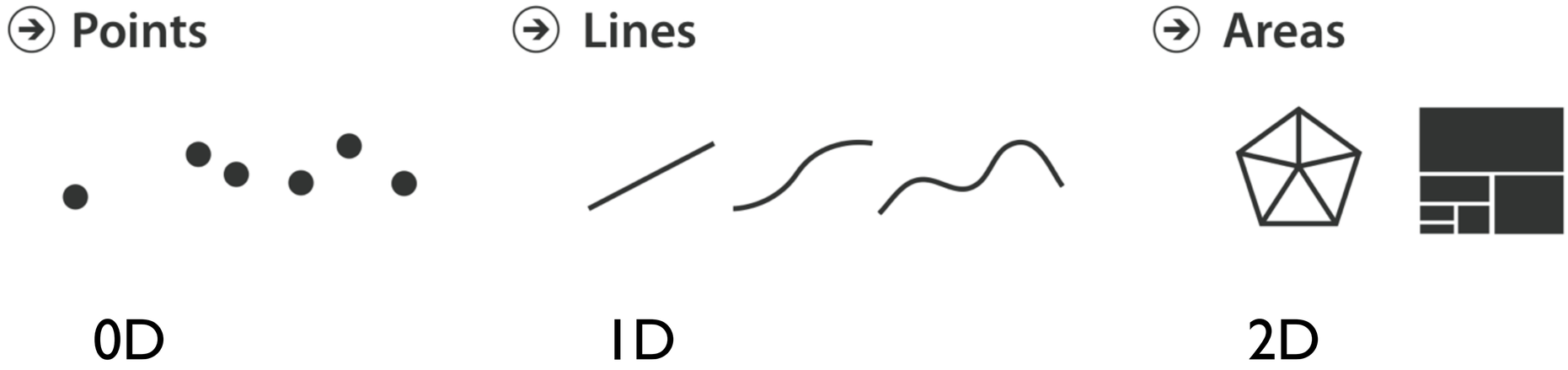
Marks: represent **items** or **links**

Channels: change **appearance** based on **attribute**

Channel = Visual Variable

Marks for Items

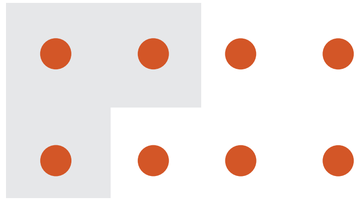
Basic geometric elements



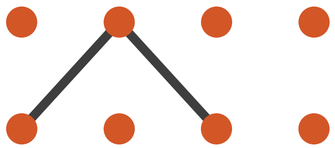
3D mark: Volume, but rarely used

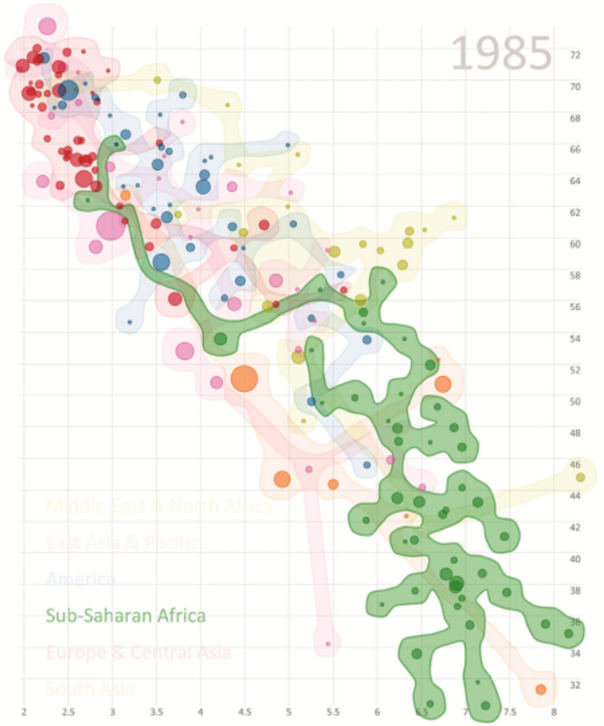
Marks for Links

Containment



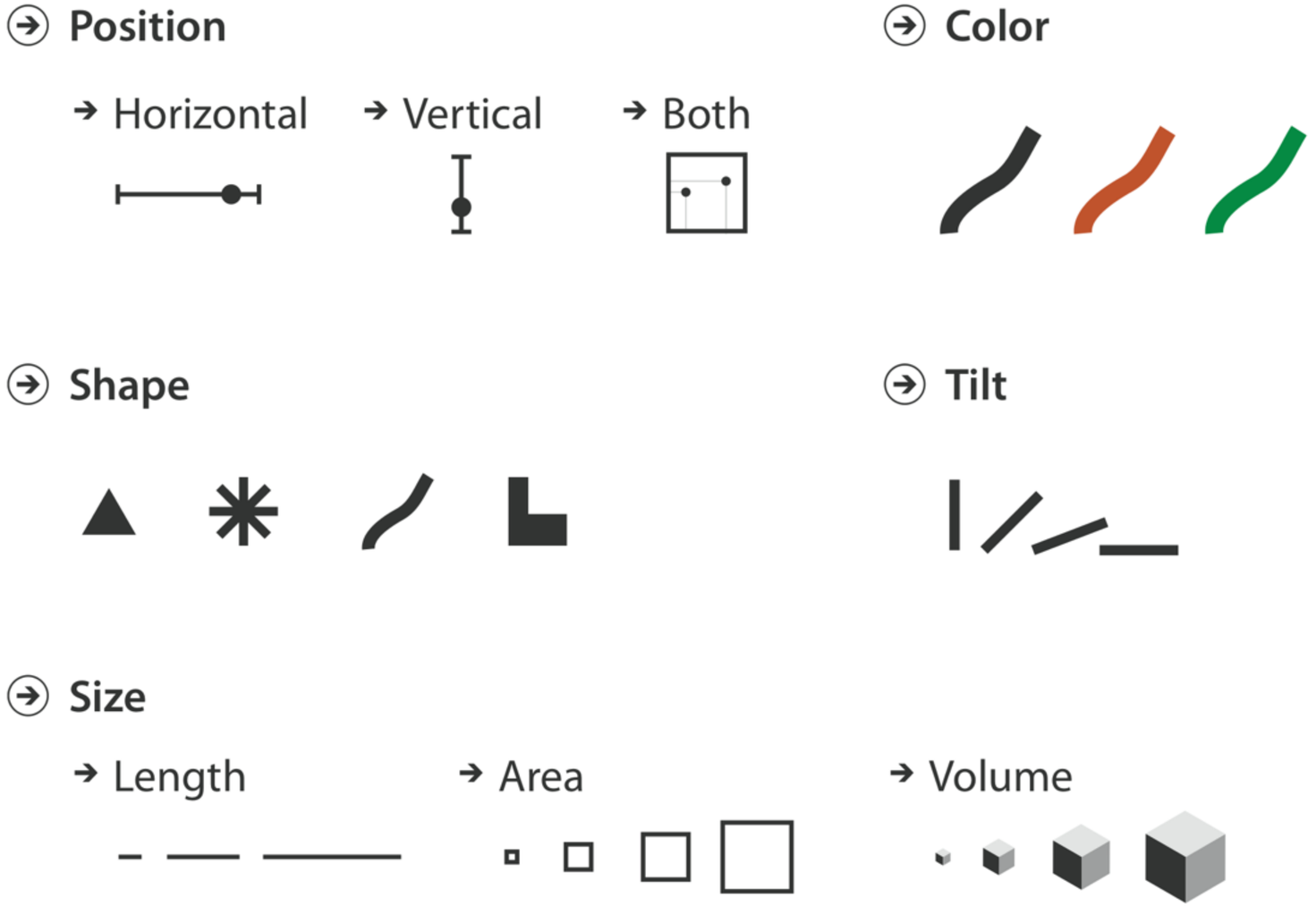
Connection





Channels (aka Visual Variables)

Control appearance proportional to or based on attributes



Types of Channels

Magnitude Channels

How much?

Position

Length

Saturation …

Identity Channels

What? Where?

Shape

Color (hue)

Spatial region …

Ordinal & Quantitative Data

Categorical Data

Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Most ▲ Effectiveness ▼ Least

Same (Color luminance, Color saturation)

Same (Curvature, Volume (3D size))

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

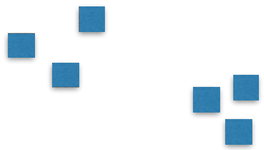
Position

Strongest visual variable

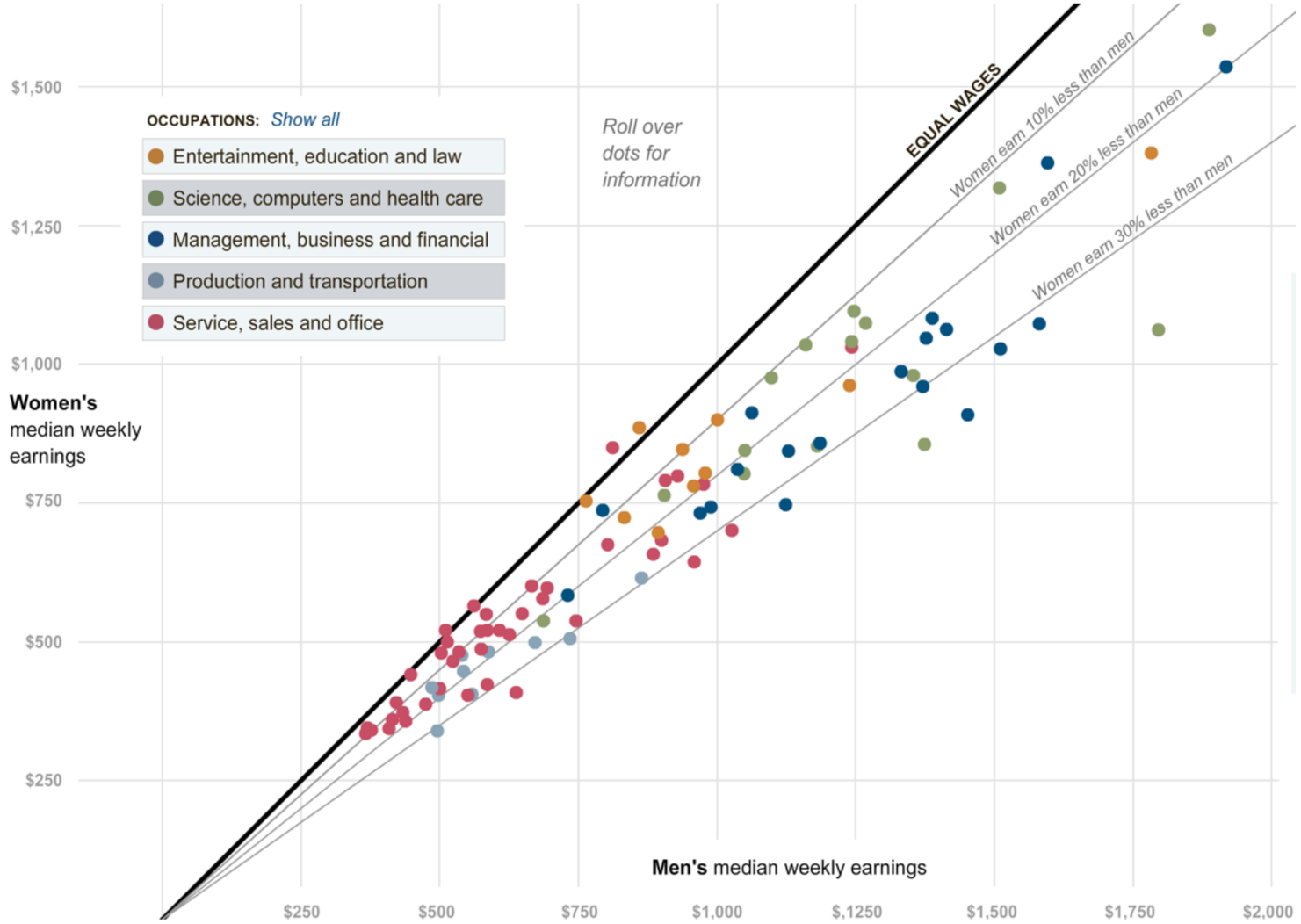
Suitable for all data types

Problems:

- Sometimes not available (spatial data)
- Cluttering



Example: Scatterplot



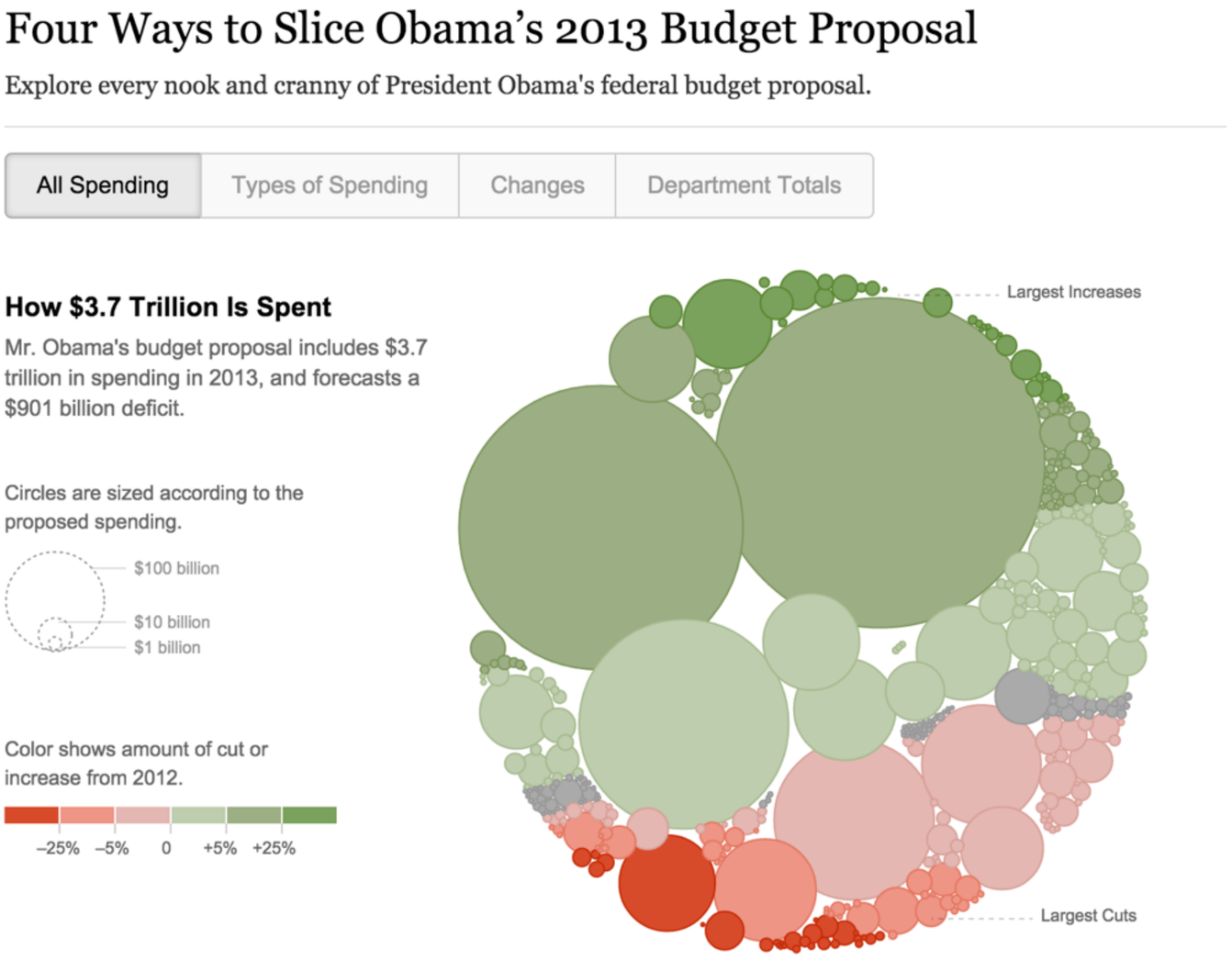
Length & Size

Good for 1D, OK for 2D, Bad for 3D

Easy to see whether one is bigger

Aligned bars use position redundantly

Example 2D Size: Bubbles



Value/Luminance/Saturation

OK for quantitative data when length & size are used.

Not very many shades recognizable

Selective: yes

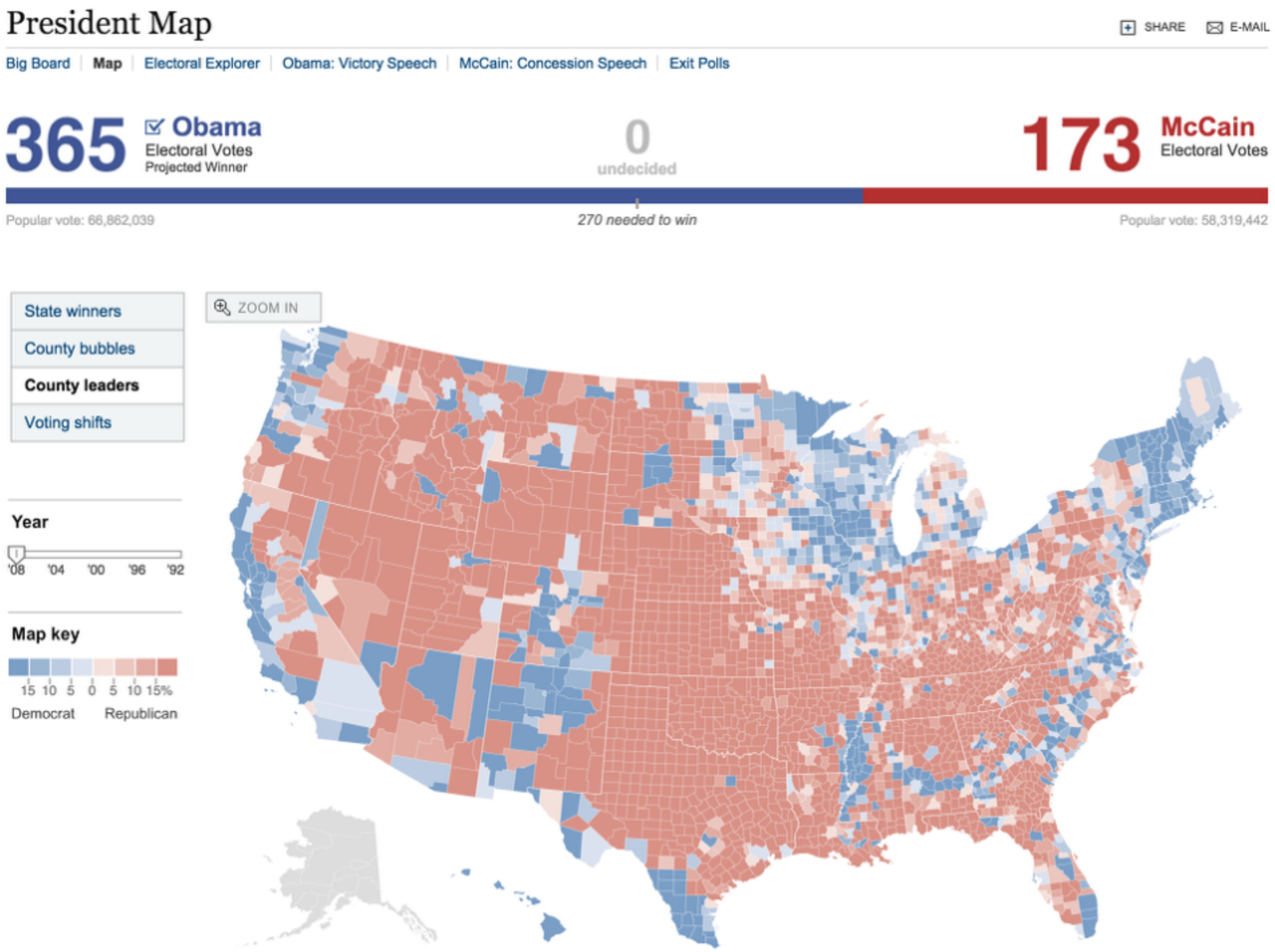
Associative: yes

Quantitative: somewhat (with problems)

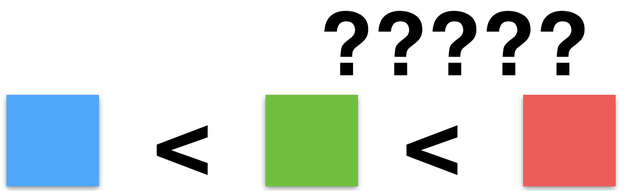
Order: yes

Length: limited

Example: Diverging Value-Scale



Color



Good for qualitative data (identity channel)

Limited number of classes/length (~7-10!)

Does not work for quantitative data!

Lots of pitfalls! Be careful!

My rule:

- minimize color use for encoding data
- use for brushing

Selective: yes

Associative: yes

Quantitative: no

Order: no

Length: limited

Color: Bad Example

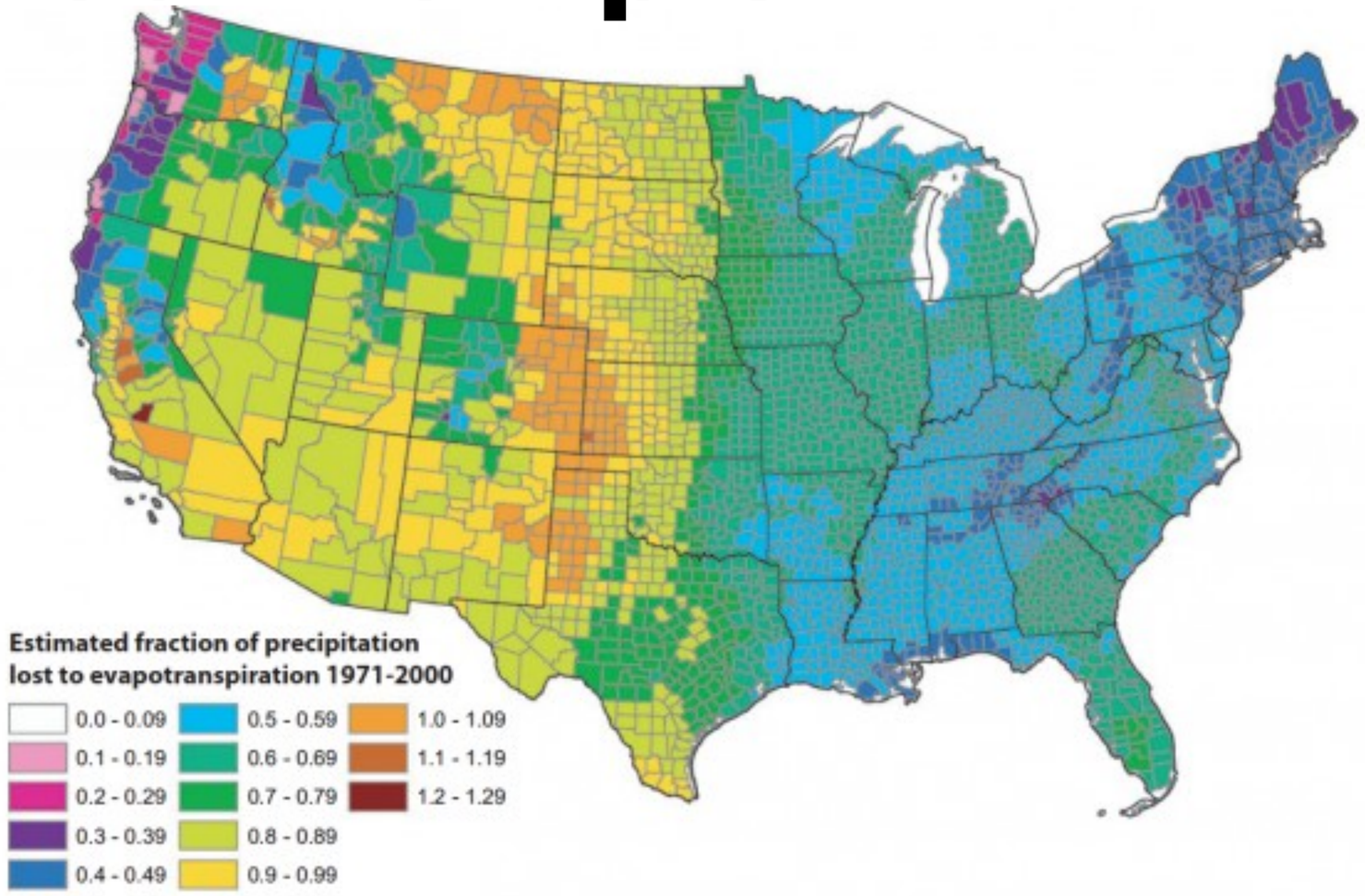
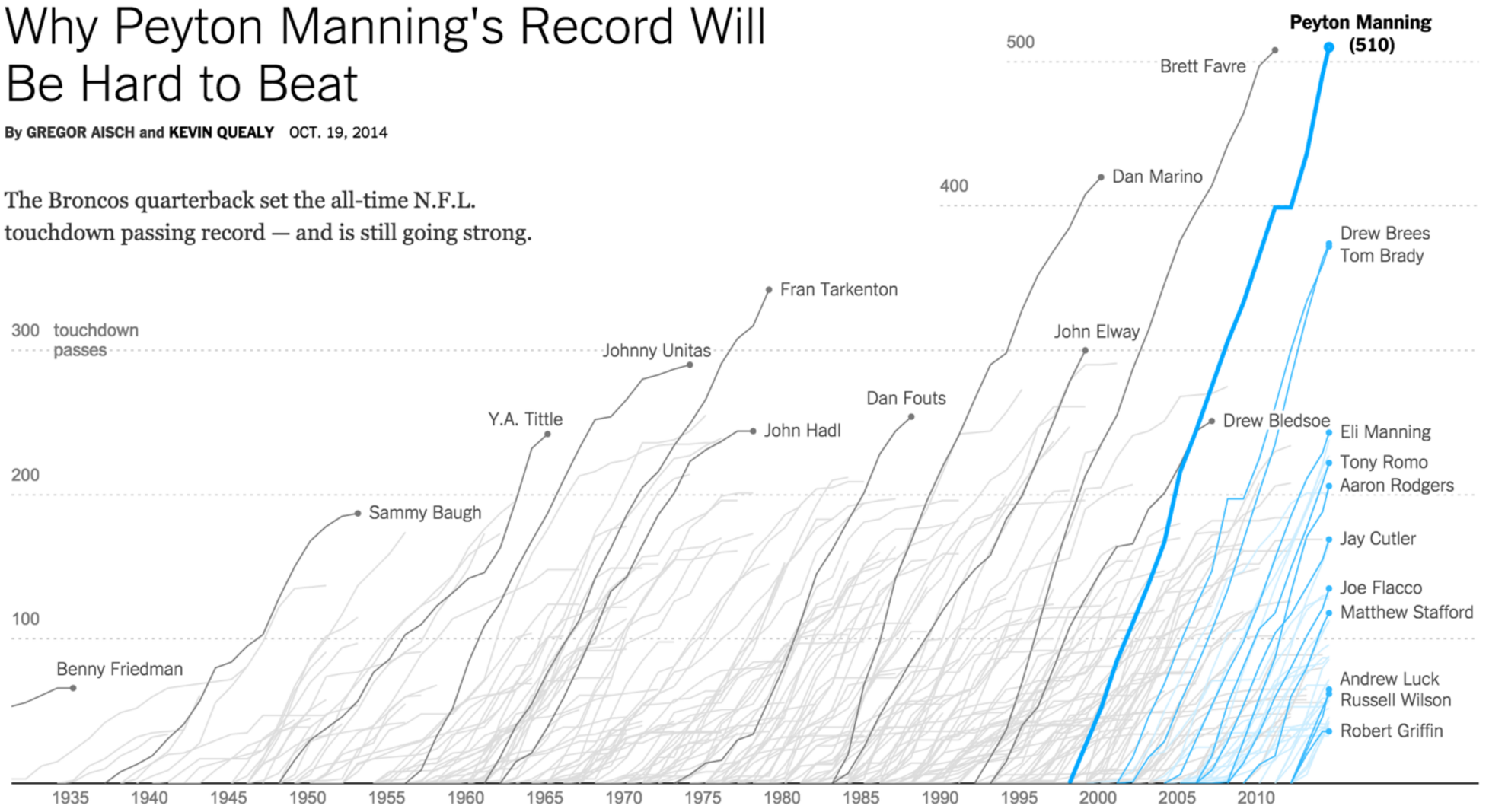


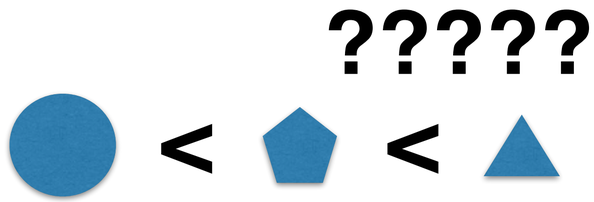
FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

Cliff Mass

Color: Good Example

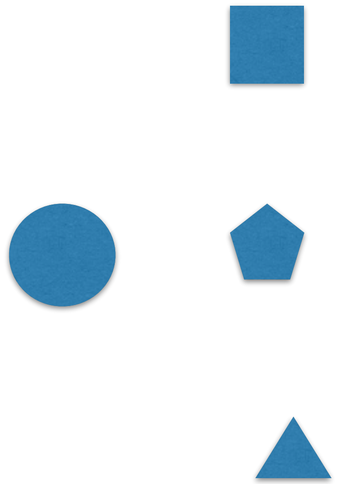


Shape



Great to recognize many classes.

No grouping, ordering.



Selective: yes

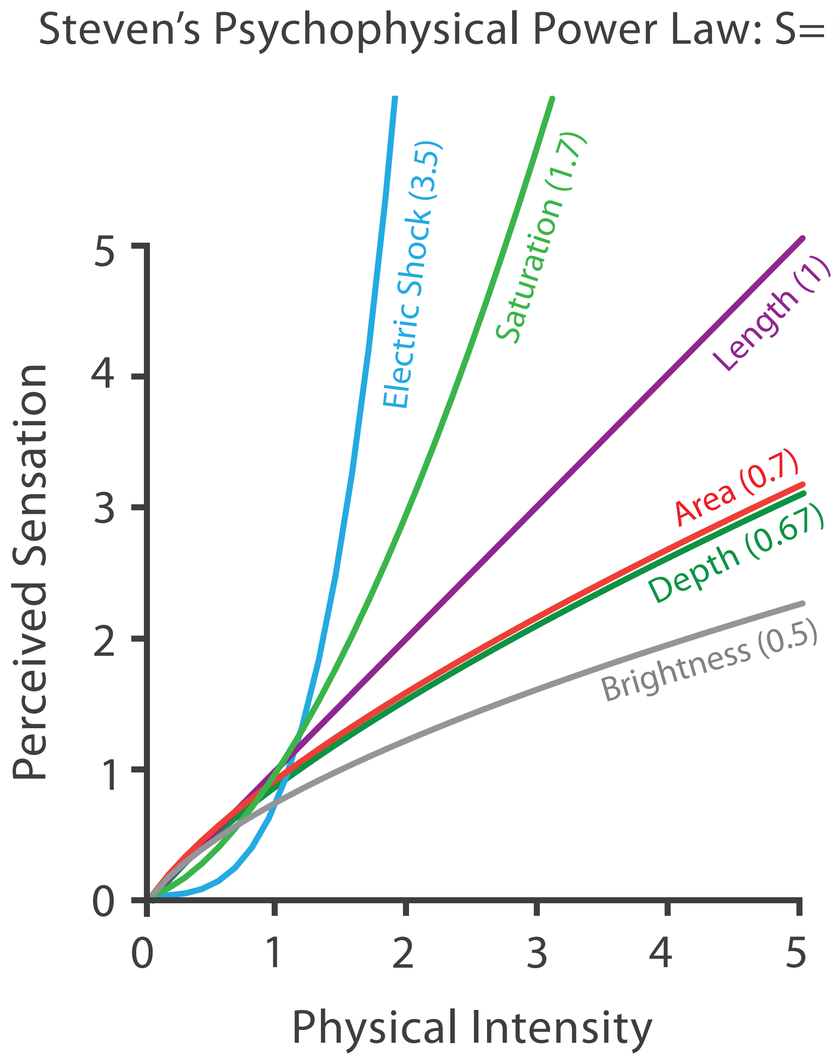
Associative: limited

Quantitative: no

Order: no

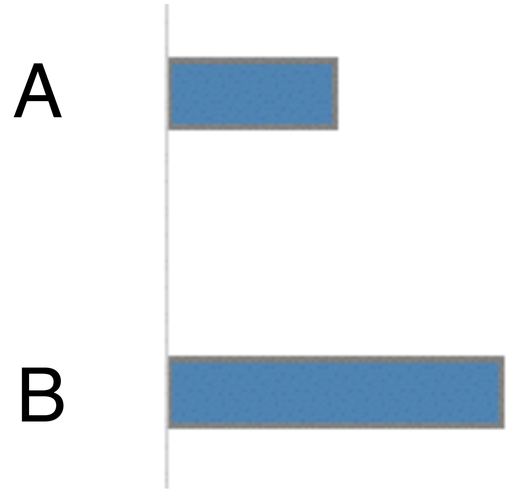
Length: vast

Why are quantitative channels different?



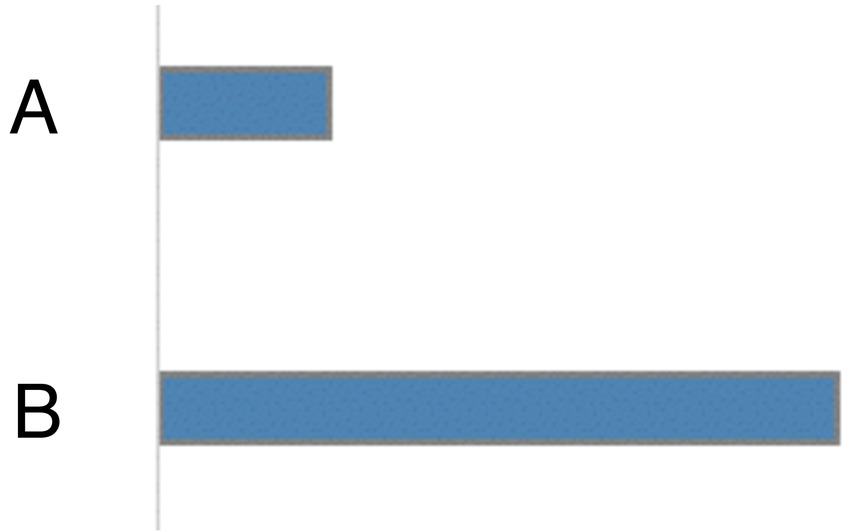
S = sensation
I = intensity

How much longer?



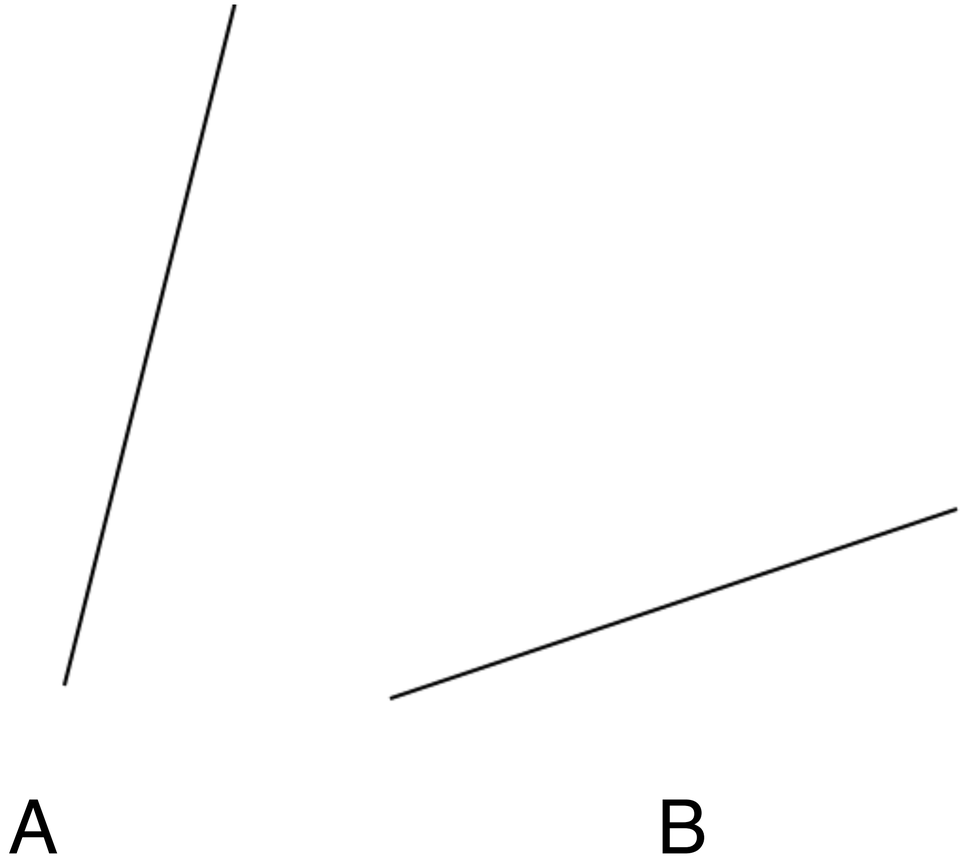
2x

How much longer?



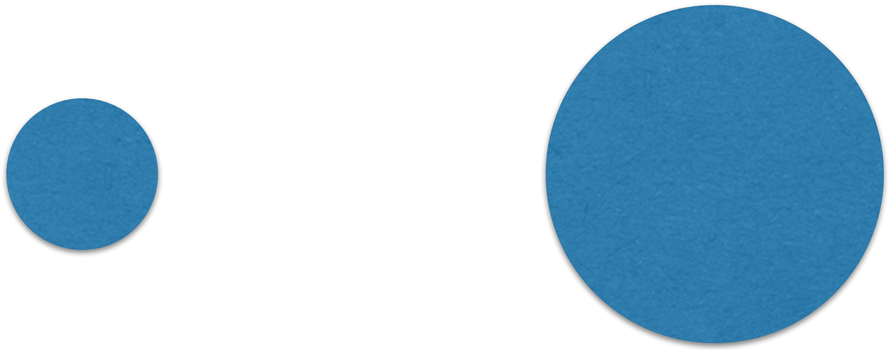
4x

How much steeper?



~4x

How much larger (area)?



A

B

5x

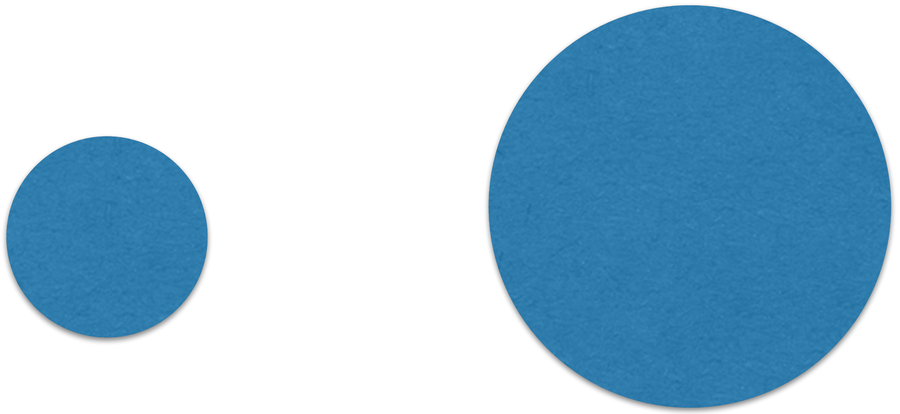
How much larger (area)?

A

B

3x

How much larger (diameter)?

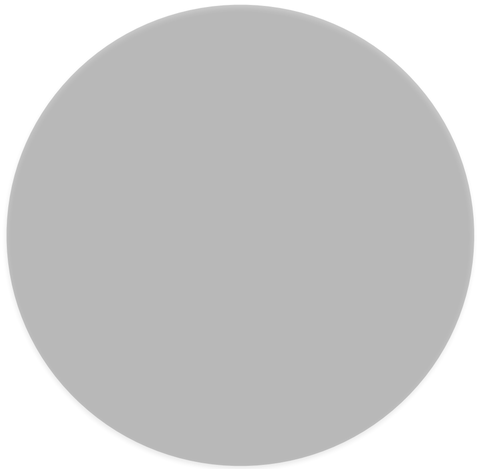


A

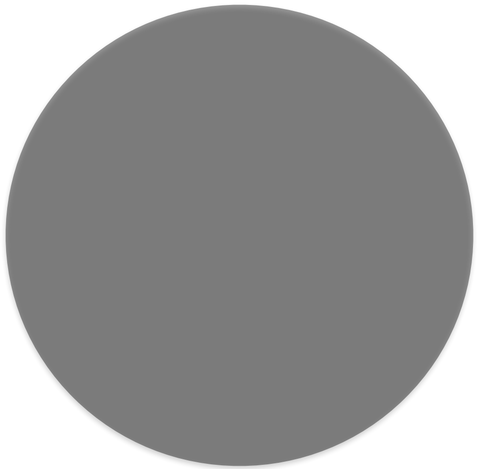
B

2x

How much darker?



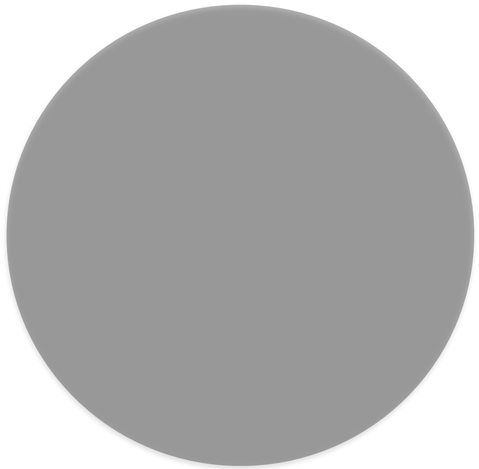
A



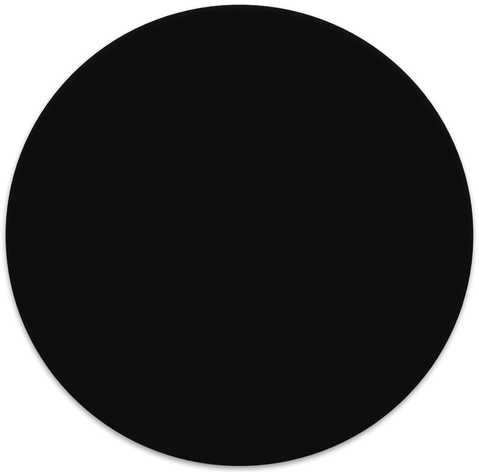
B

2x

How much darker?



A



B

3x

Other Factors Affecting Accuracy

Alignment

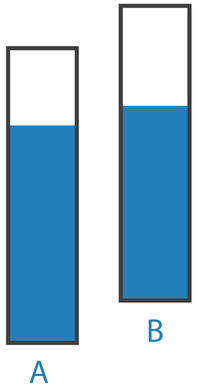
Distractors

Distance

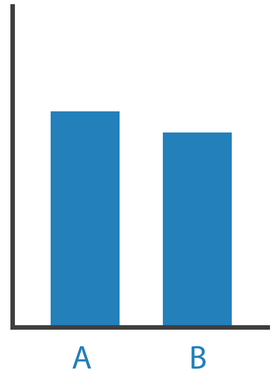
Common scale

…

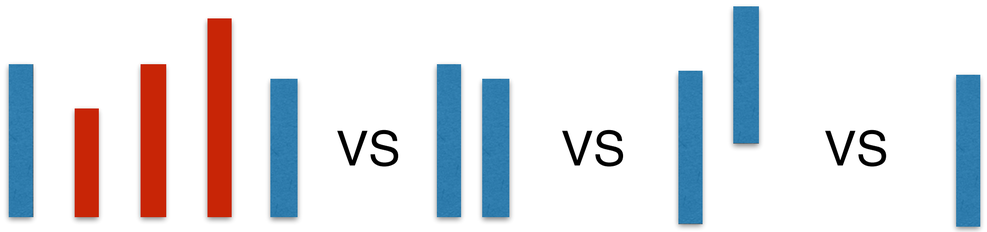
Unframed
Unaligned



Framed
Unaligned



Unframed
Aligned



Channels: Expressiveness Types and Effectiveness Ranks

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Most ▲ Effectiveness ▼ Least

Same (Color luminance, Color saturation)

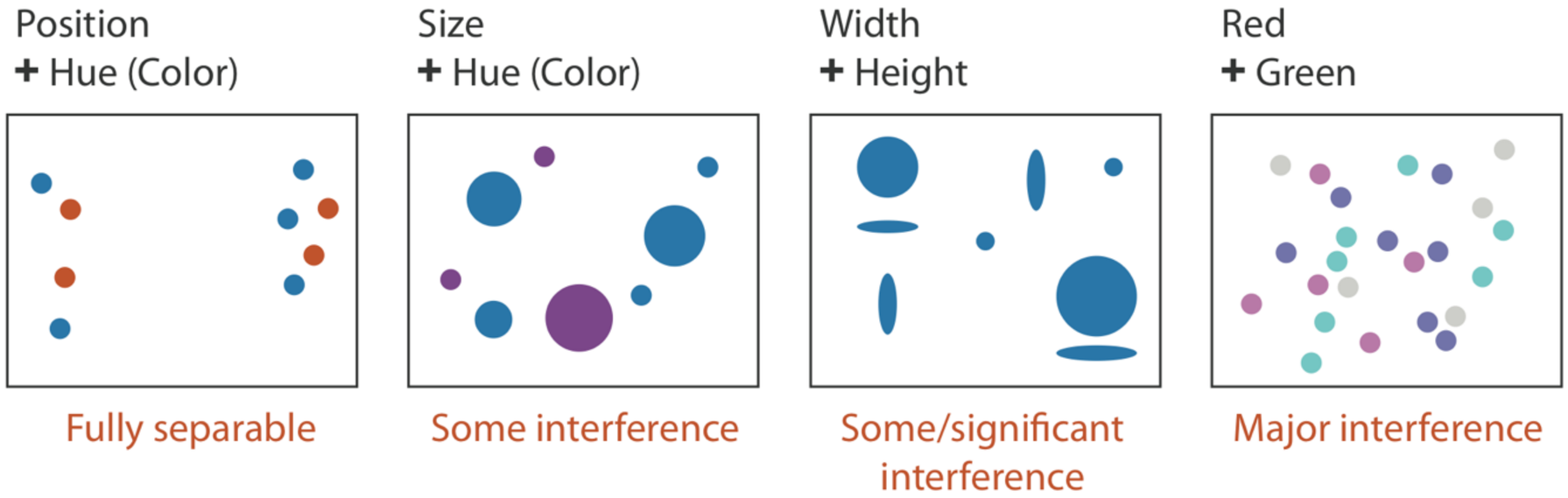
Same (Curvature, Volume (3D size))

Identity Channels: Categorical Attributes

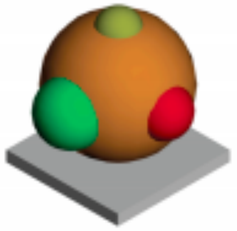
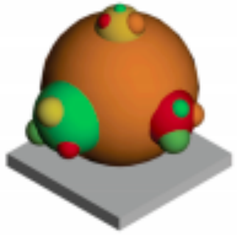
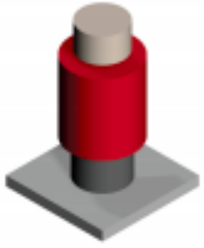
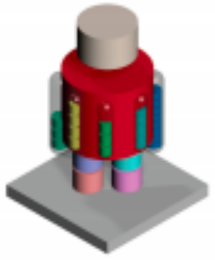
- Spatial region
- Color hue
- Motion
- Shape

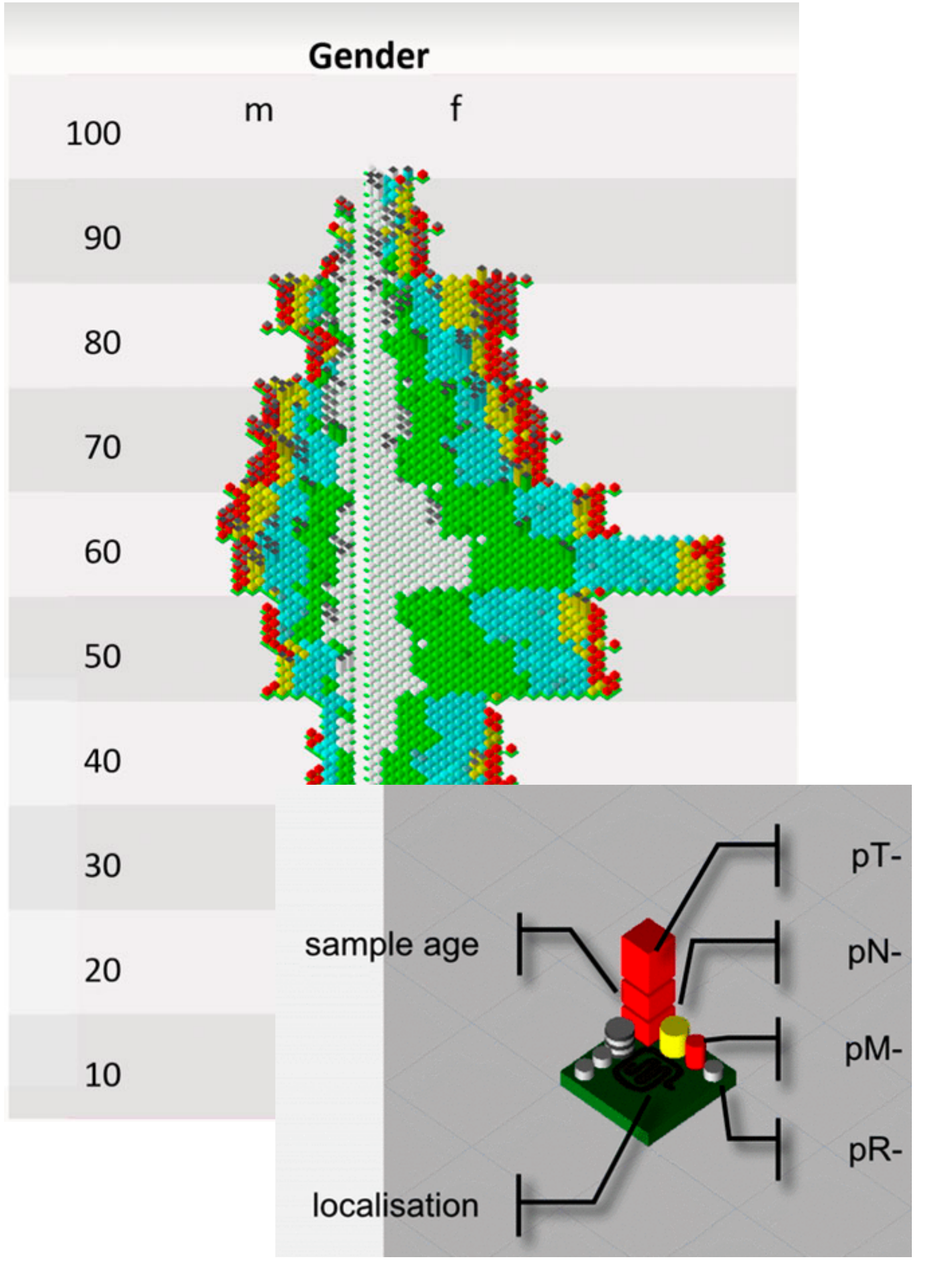
Separability of Attributes

Can we combine multiple visual variables?



Sins from the past…

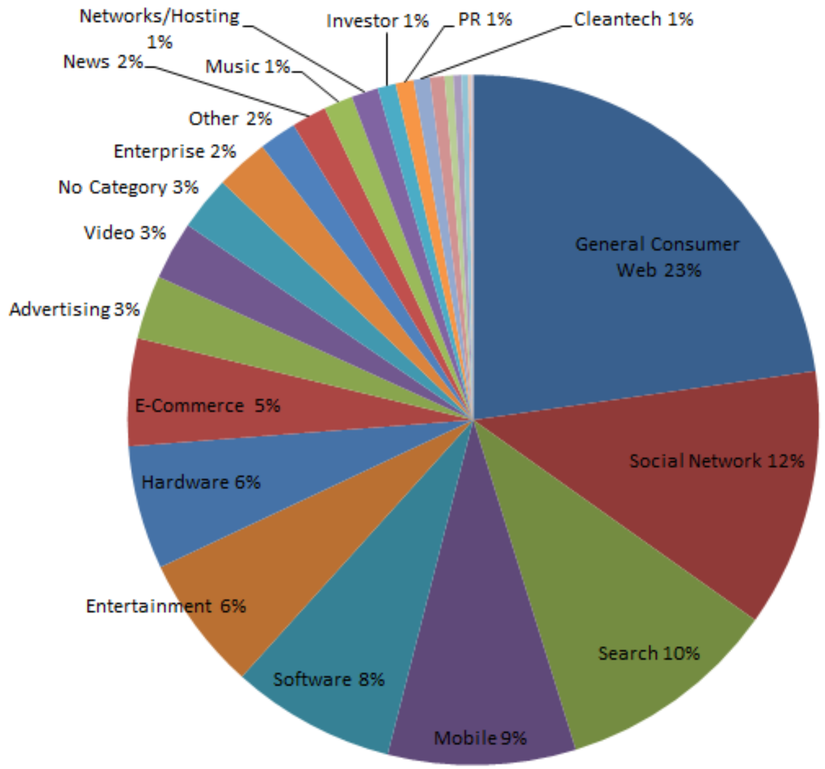
	Level 1	Level 2	Level 3
Cubic glyphs			
Crystal glyphs			
Human Glyphs			



[Mueller 09, Mueller 14]

Common Mistakes

Death to Pie Charts



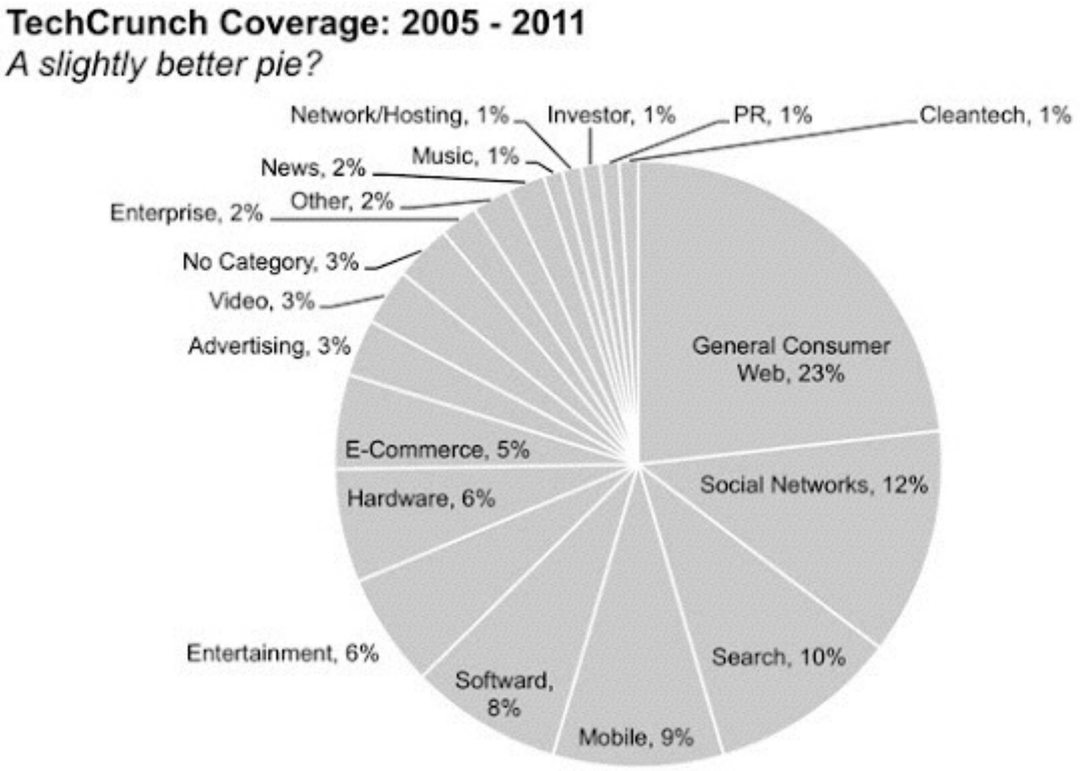
Share of coverage on TechCrunch

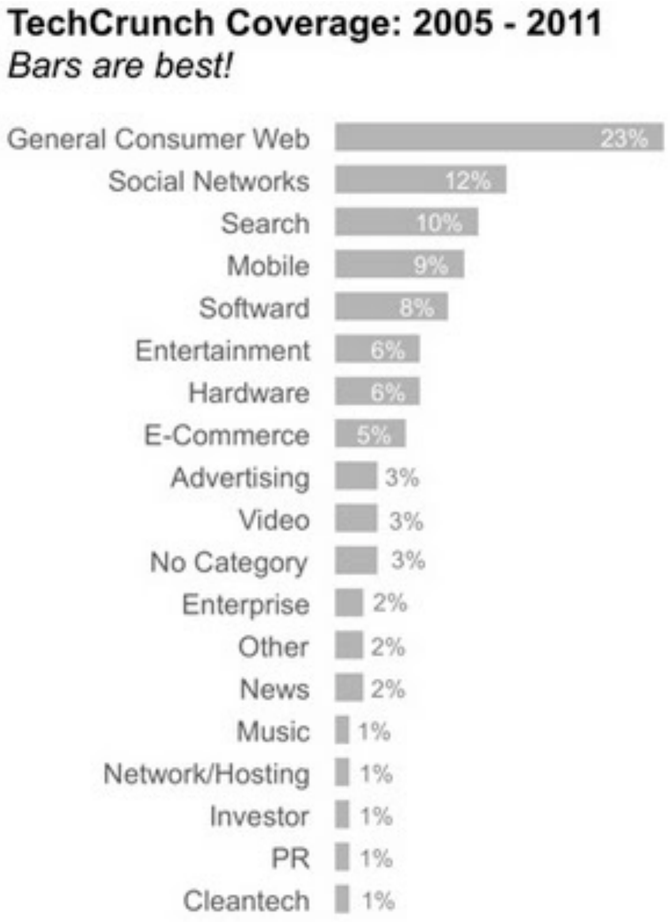
"I hate pie charts.
I mean, really hate them."

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

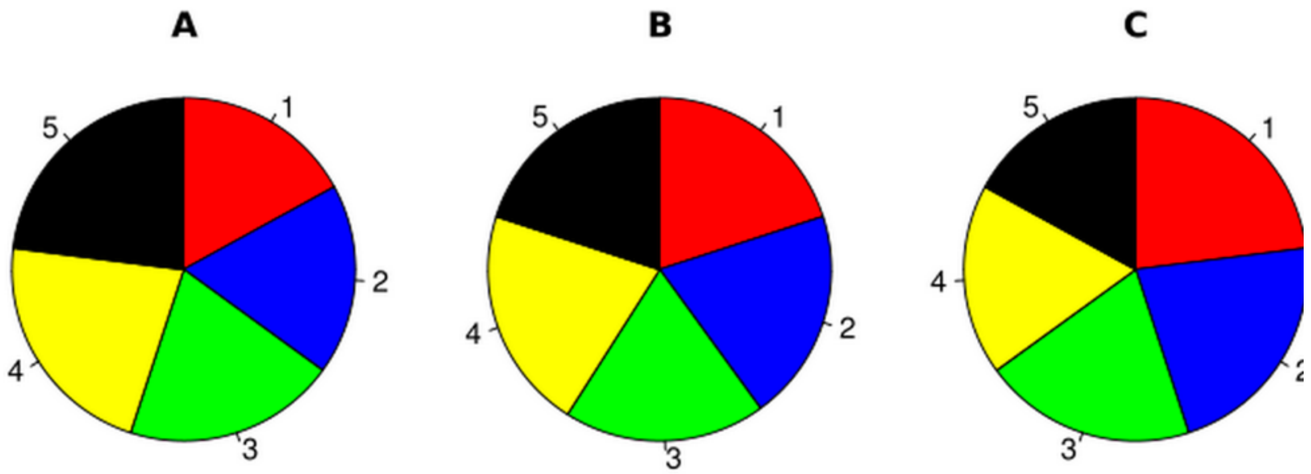
Cole Nussbaumer

Redesign

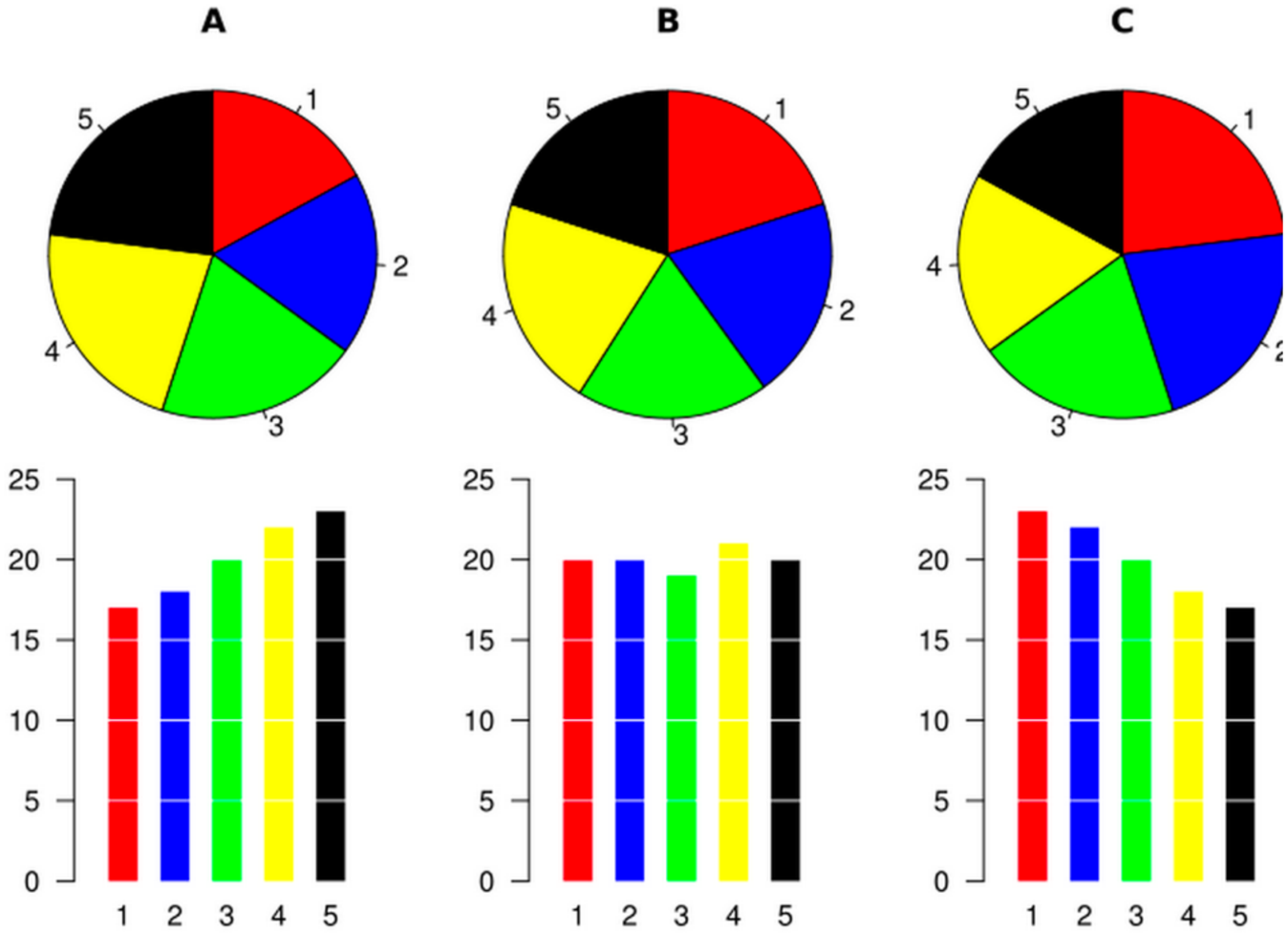

TechCrunch Coverage: 2005 - 2011
A slightly better pie?
Network/Hosting, 1%
Investor, 1%
PR, 1%
Cleantech, 1%
News, 2%
Music, 1%
Enterprise, 2%
Other, 2%
No Category, 3%
Video, 3%
Advertising, 3%
General Consumer Web, 23%
E-Commerce, 5%
Social Networks, 12%
Hardware, 6%
Entertainment, 6%
Softward, 8%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

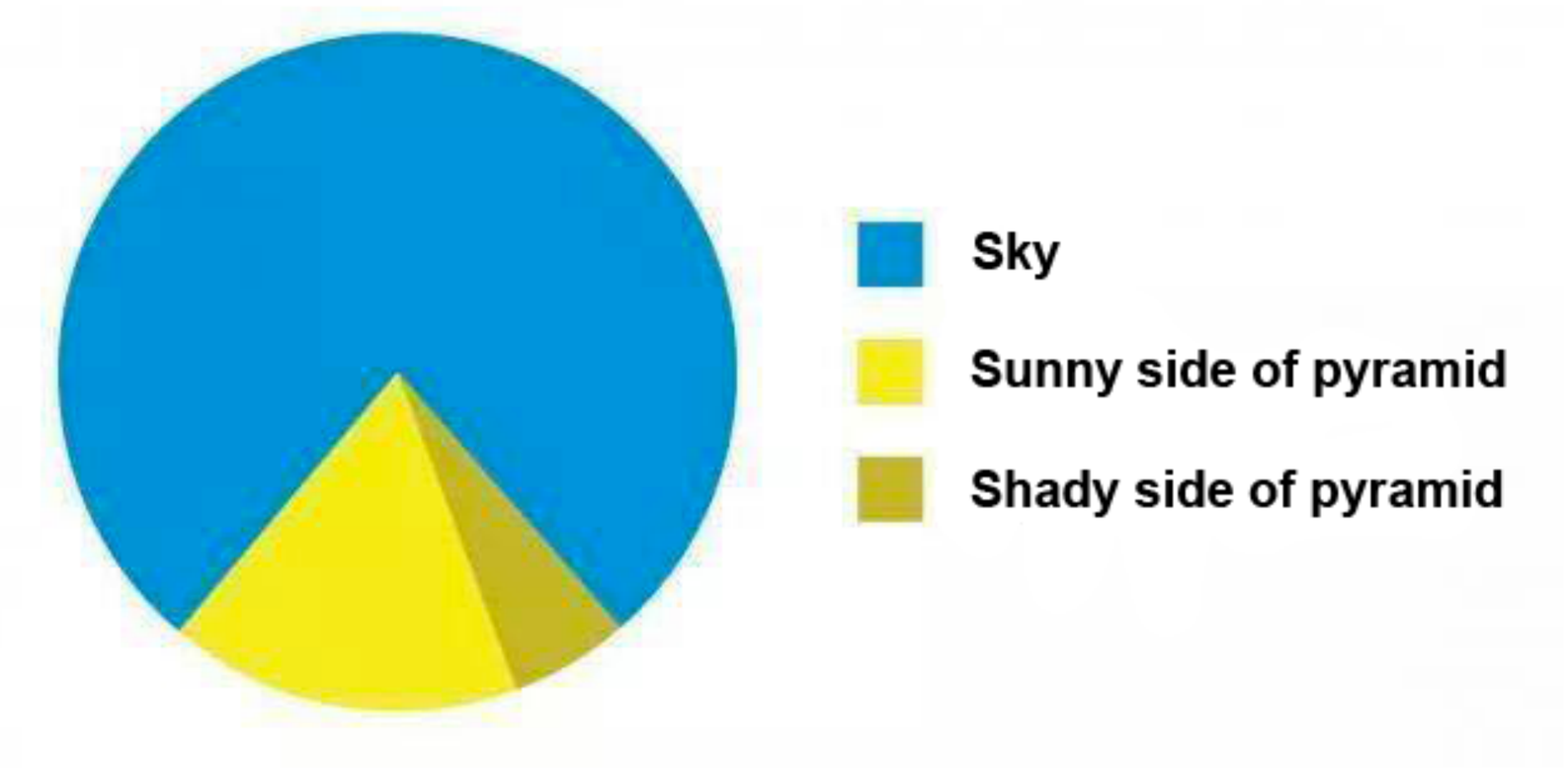
Can you spot the differences?



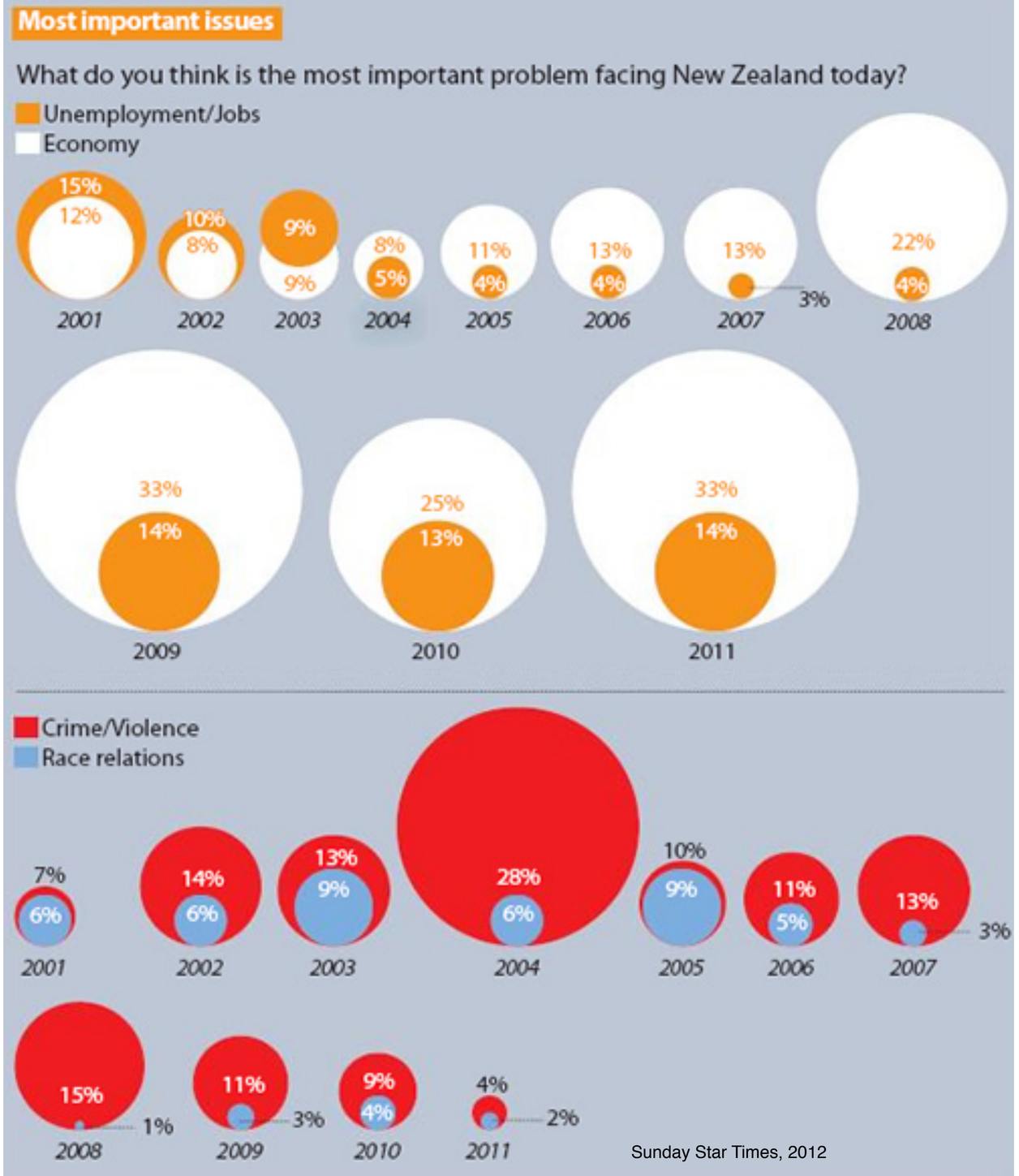
Can you spot the differences?



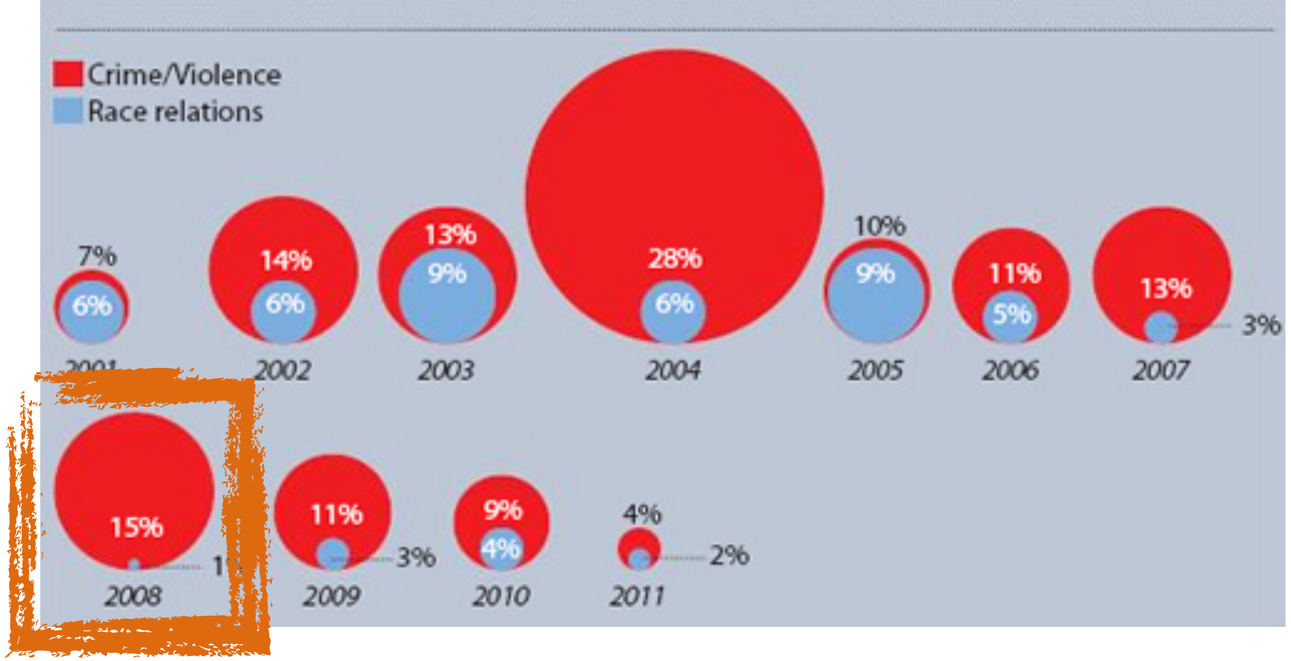
My favorite pie chart



My second favorite pie chart

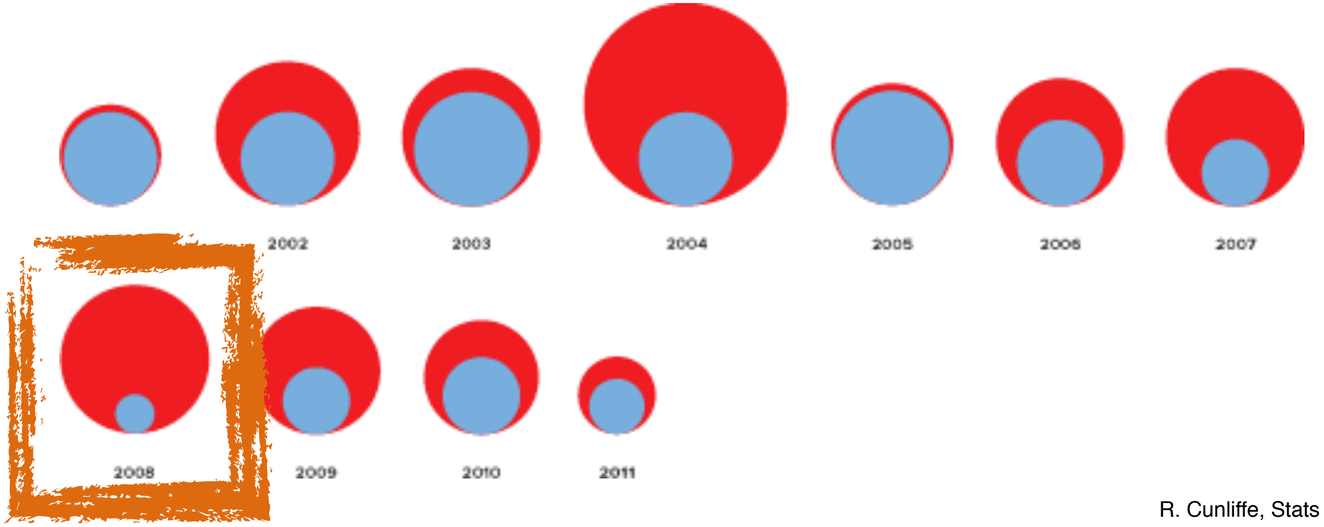


Sunday Star Times, 2012

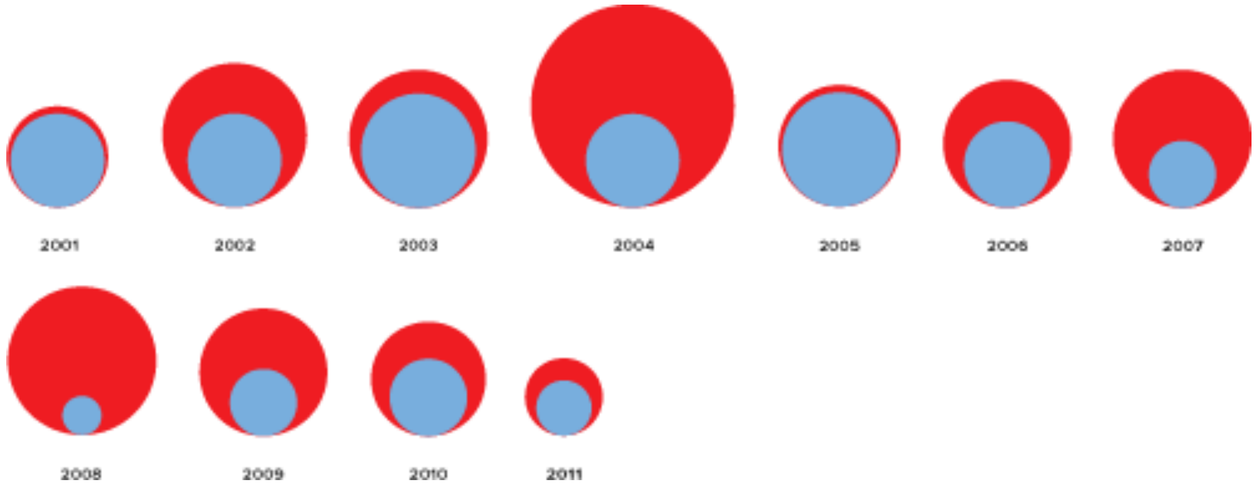


Quantity encoded by diameter, not area!

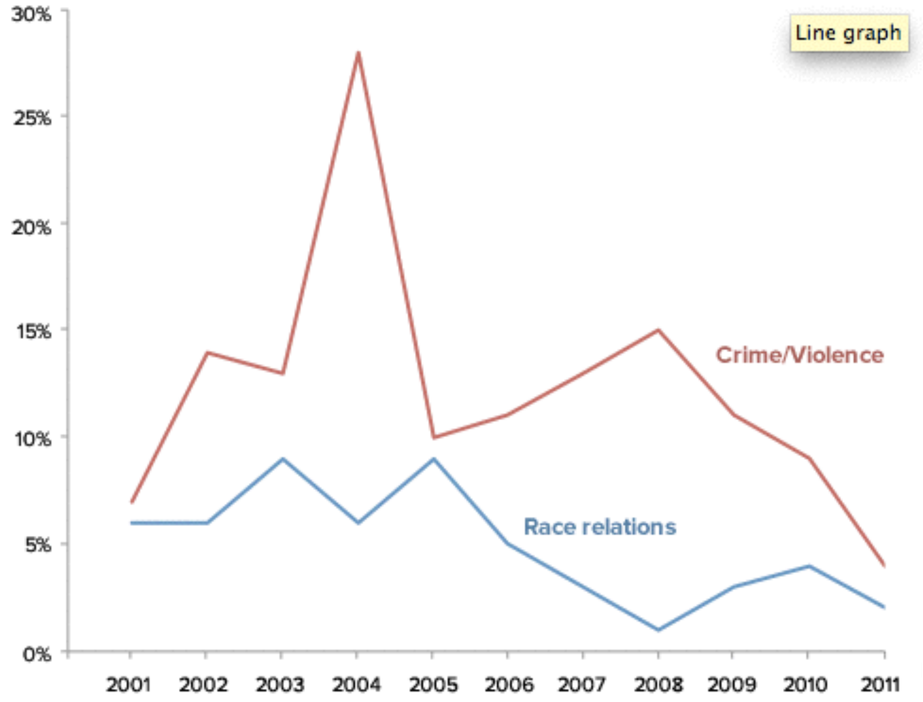
Fixing that:



R. Cunliffe, Stats Chat

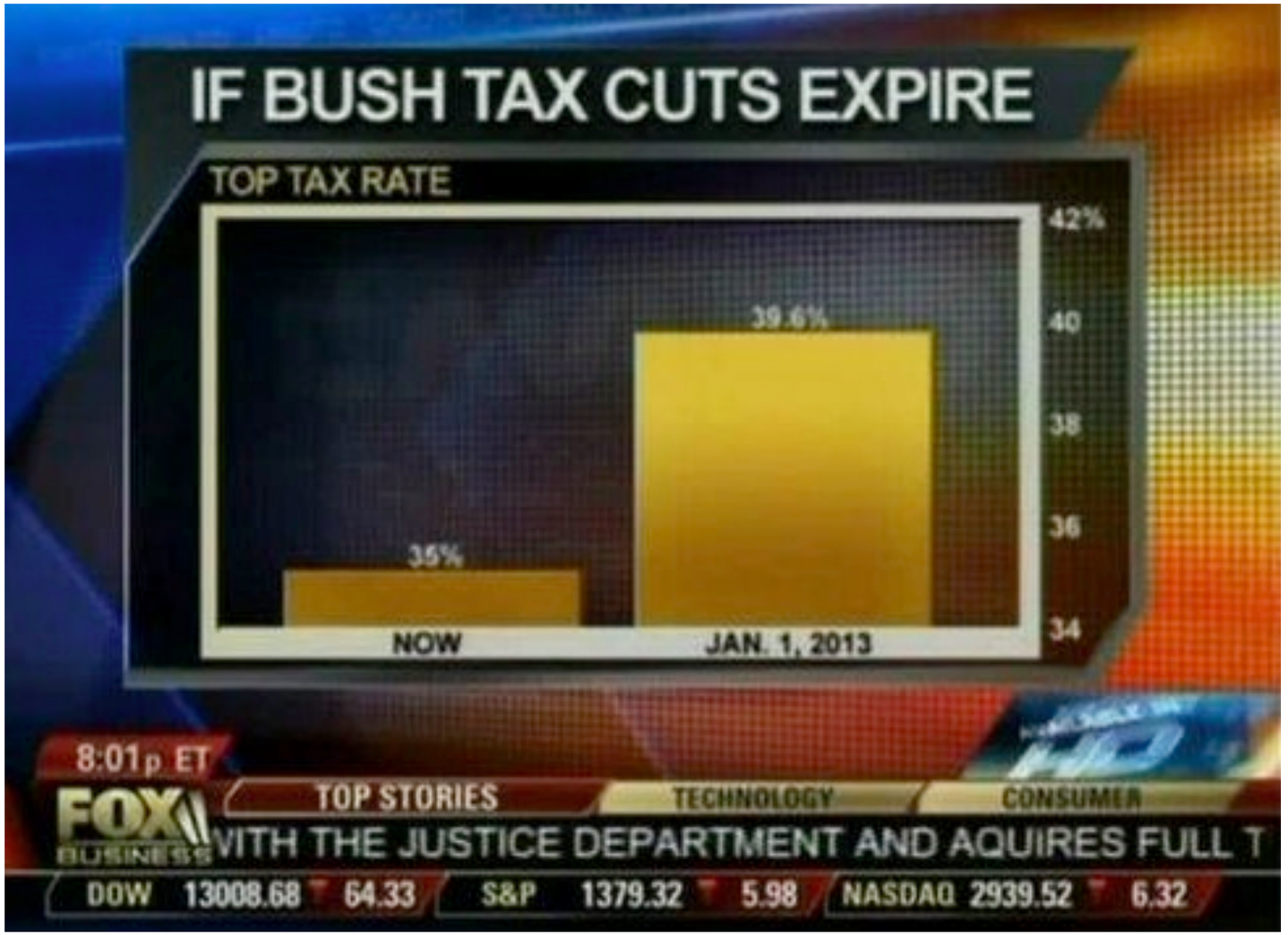


But is this visual encoding appropriate in the first place?



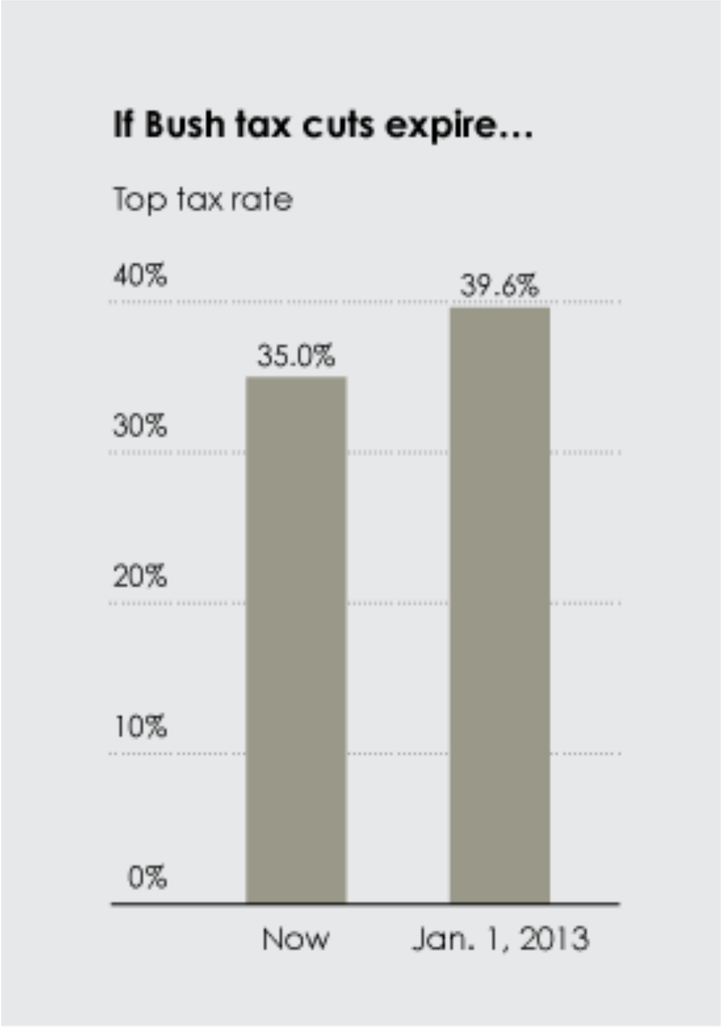
R. Cunliffe, Stats Chat

Graphical Integrity



Flowing Data

Scale Distortions



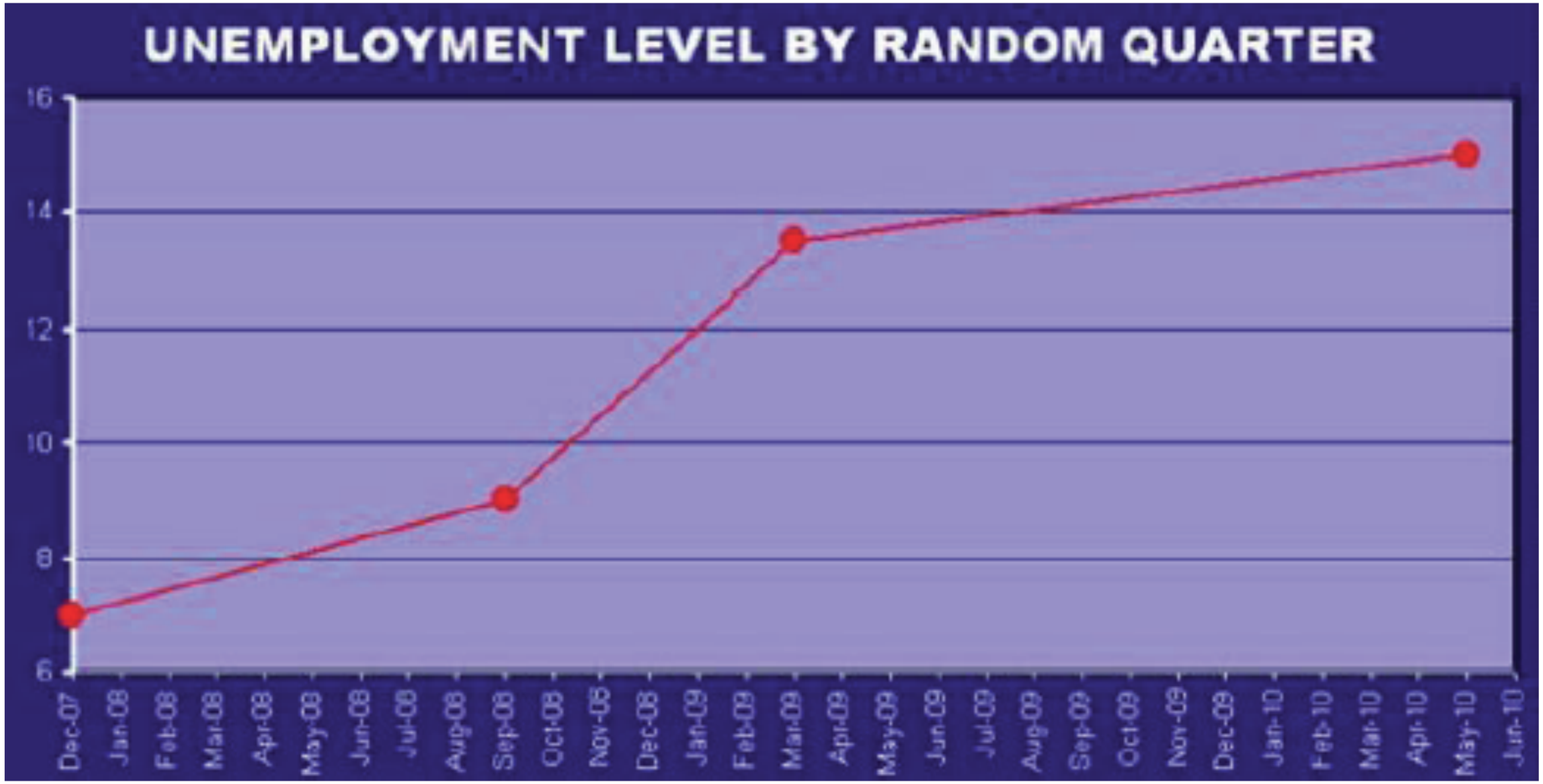
Flowing Data

What's wrong?

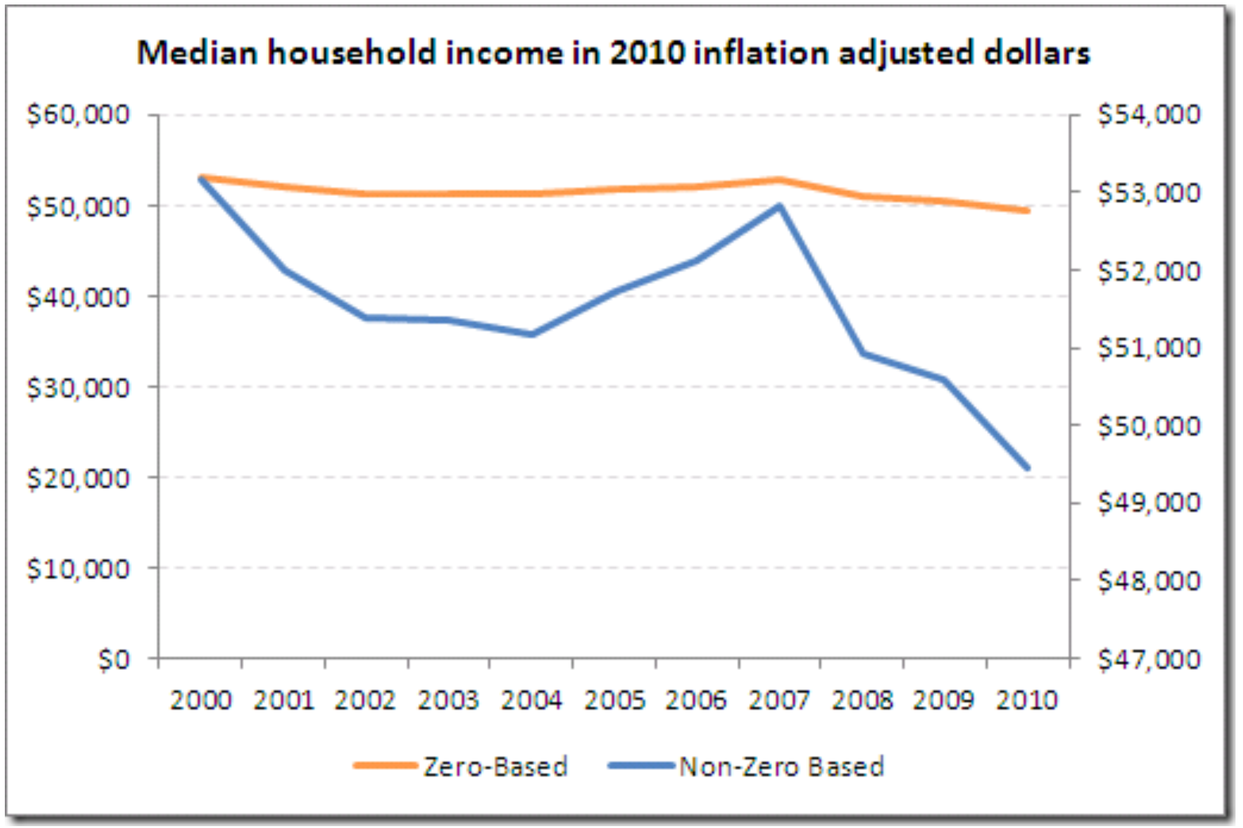

How 2012 Stacks Up
The Warmest Years on Record
Contiguous U.S.
55.3
54.31
54.32
54.14
53.93
53.81
55.5°
55.0°
54.5°
54.0°
53.5°
53.0°
1921
1999
1934
2006
1998
2012
Source: NOAA's National Climatic Data Center - State of the Climate National Overview
CLIMATE CENTRAL

Scale Distortions

Scale Distortions

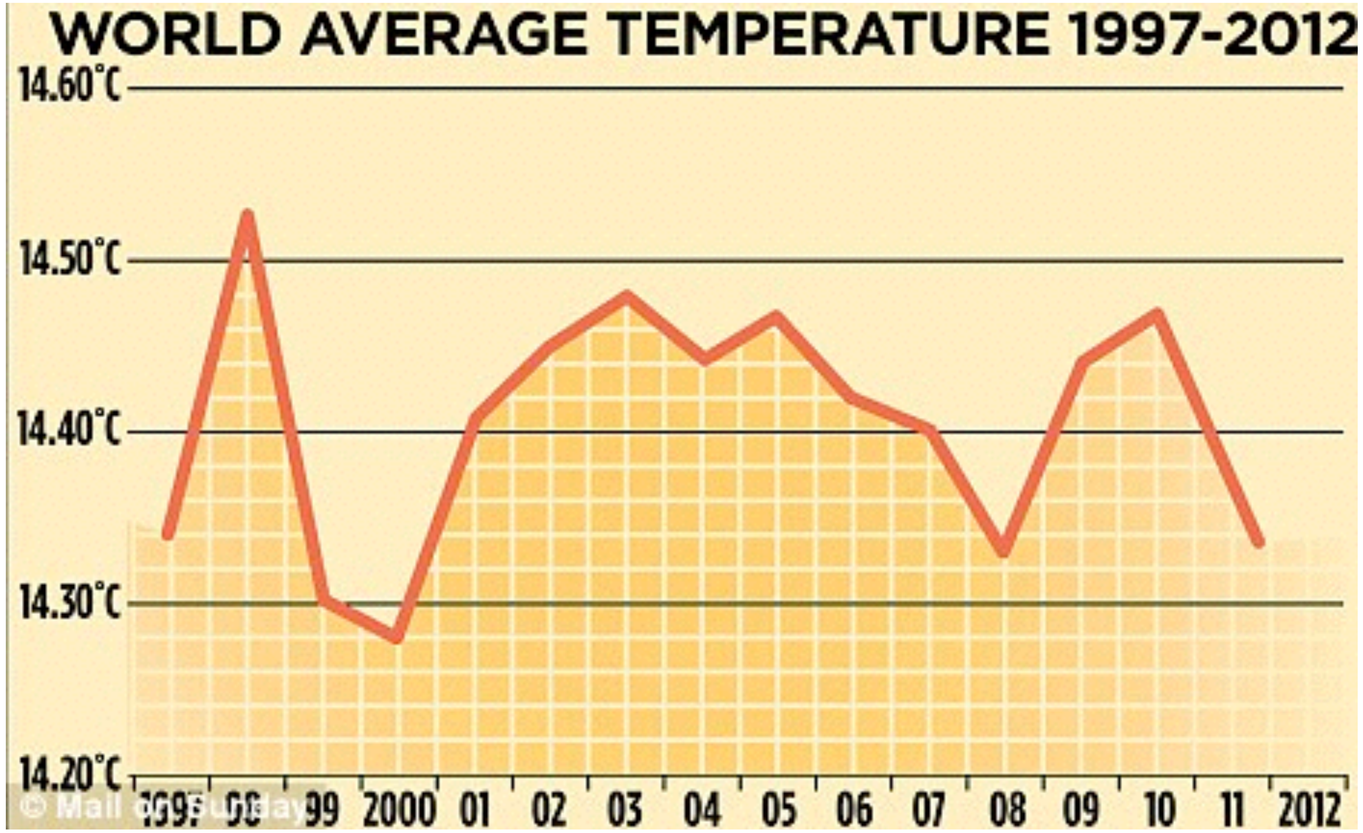


Start Scales at 0?



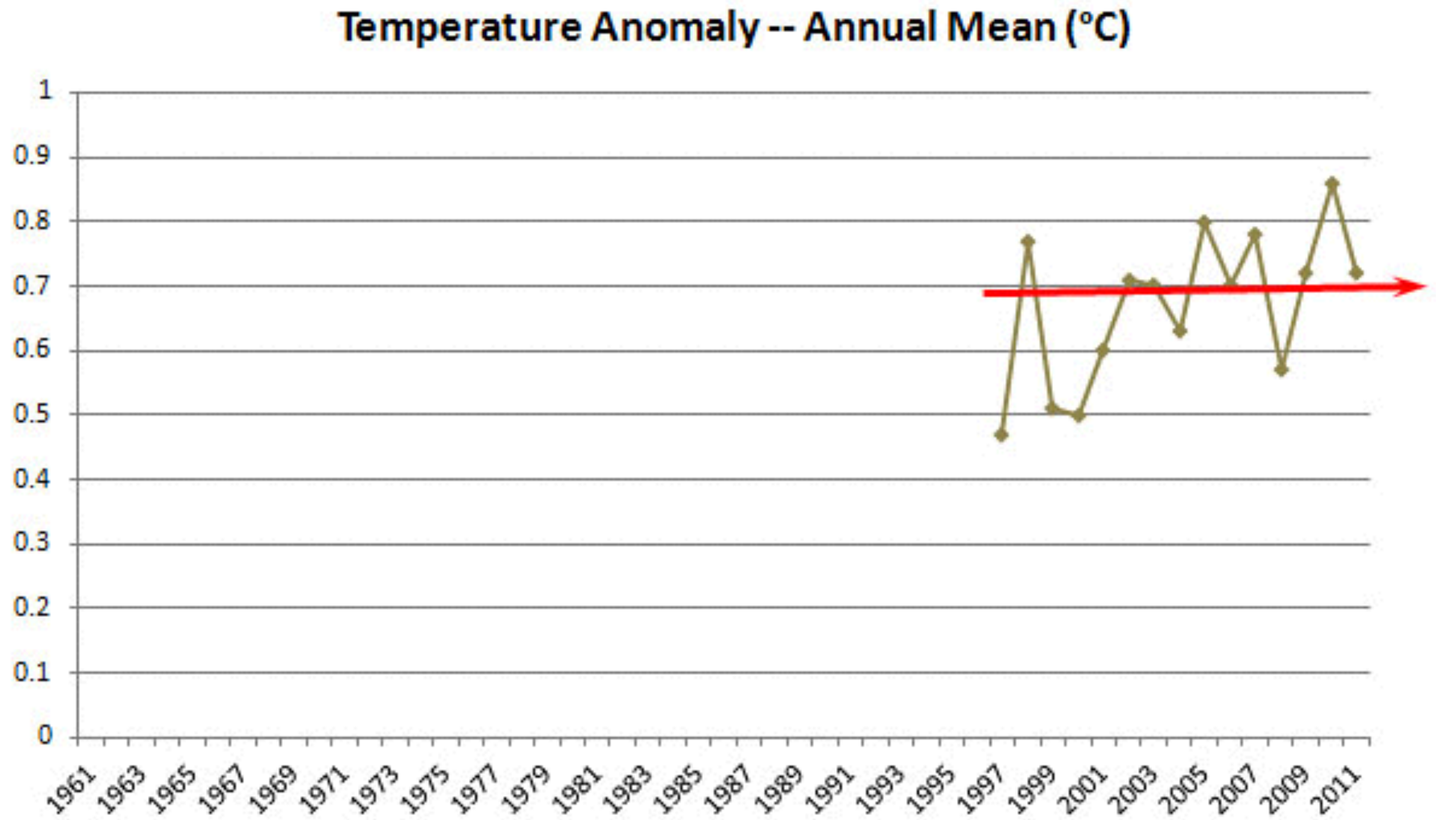
A. Kriebel, VizWiz

Global Warming?



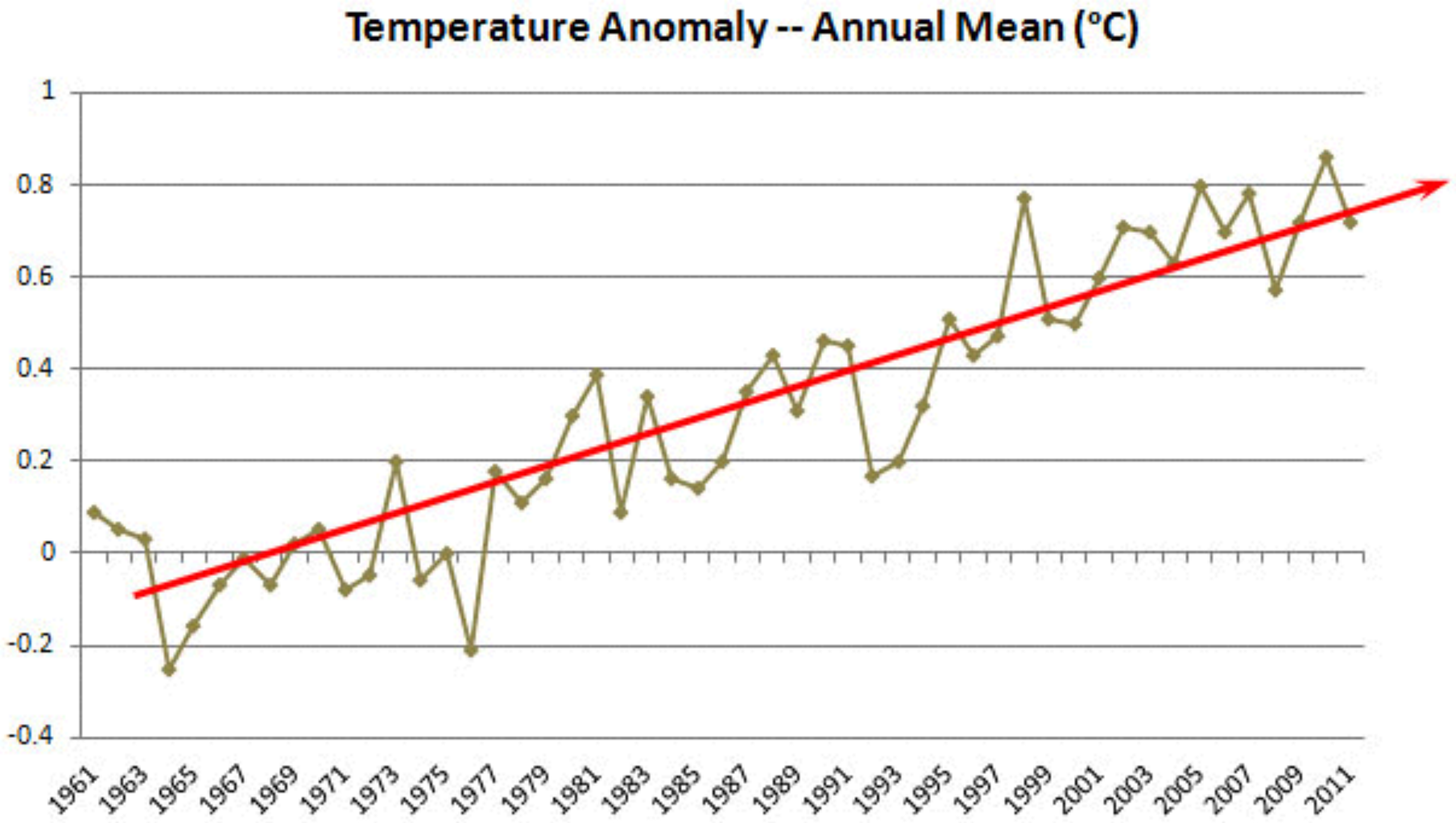
The Daily Mail, UK, Jan 2012

Global Warming?



Mother Jones

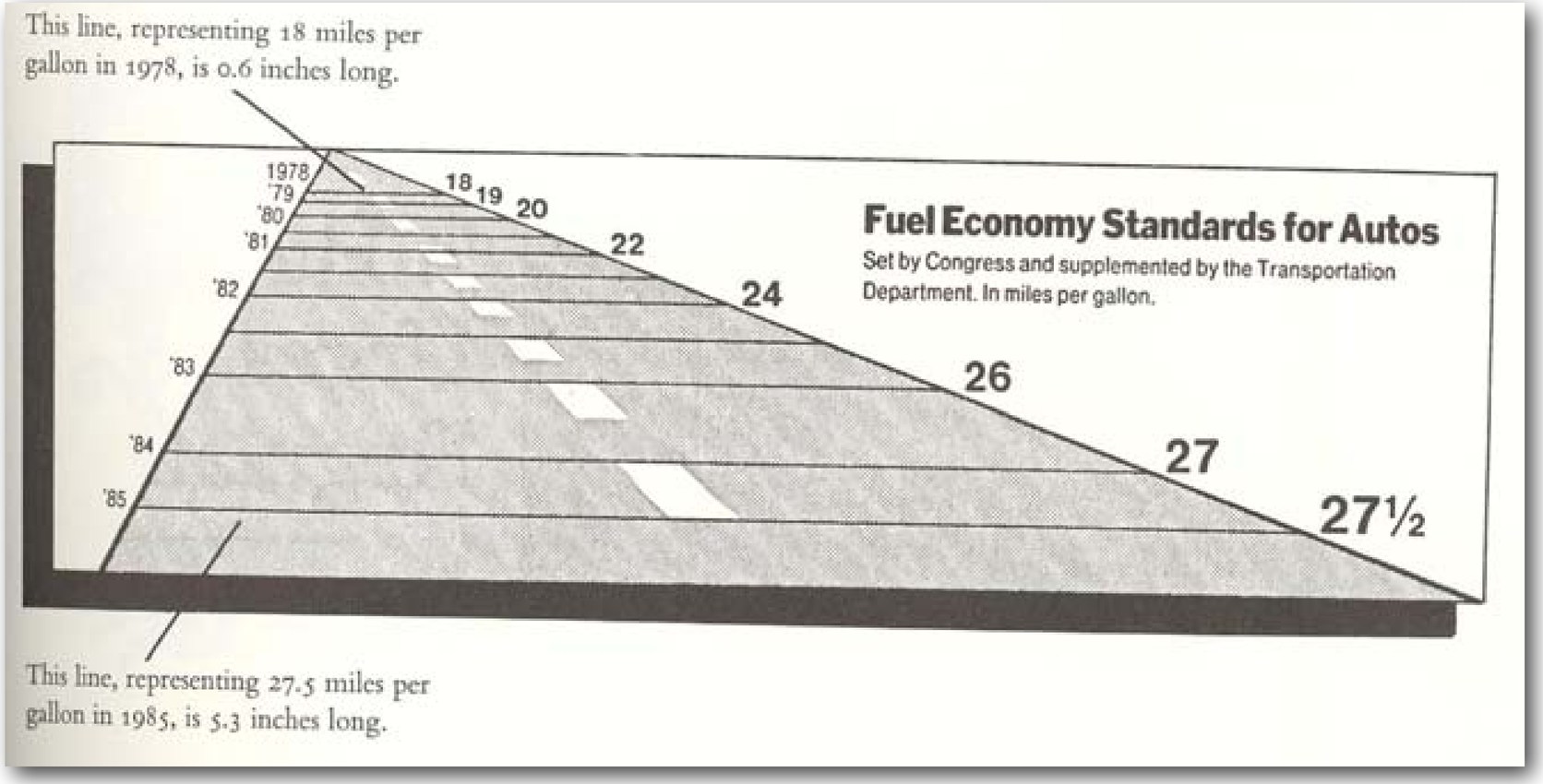
Global Warming - Frame the Data



Mother Jones

The Lie Factor

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

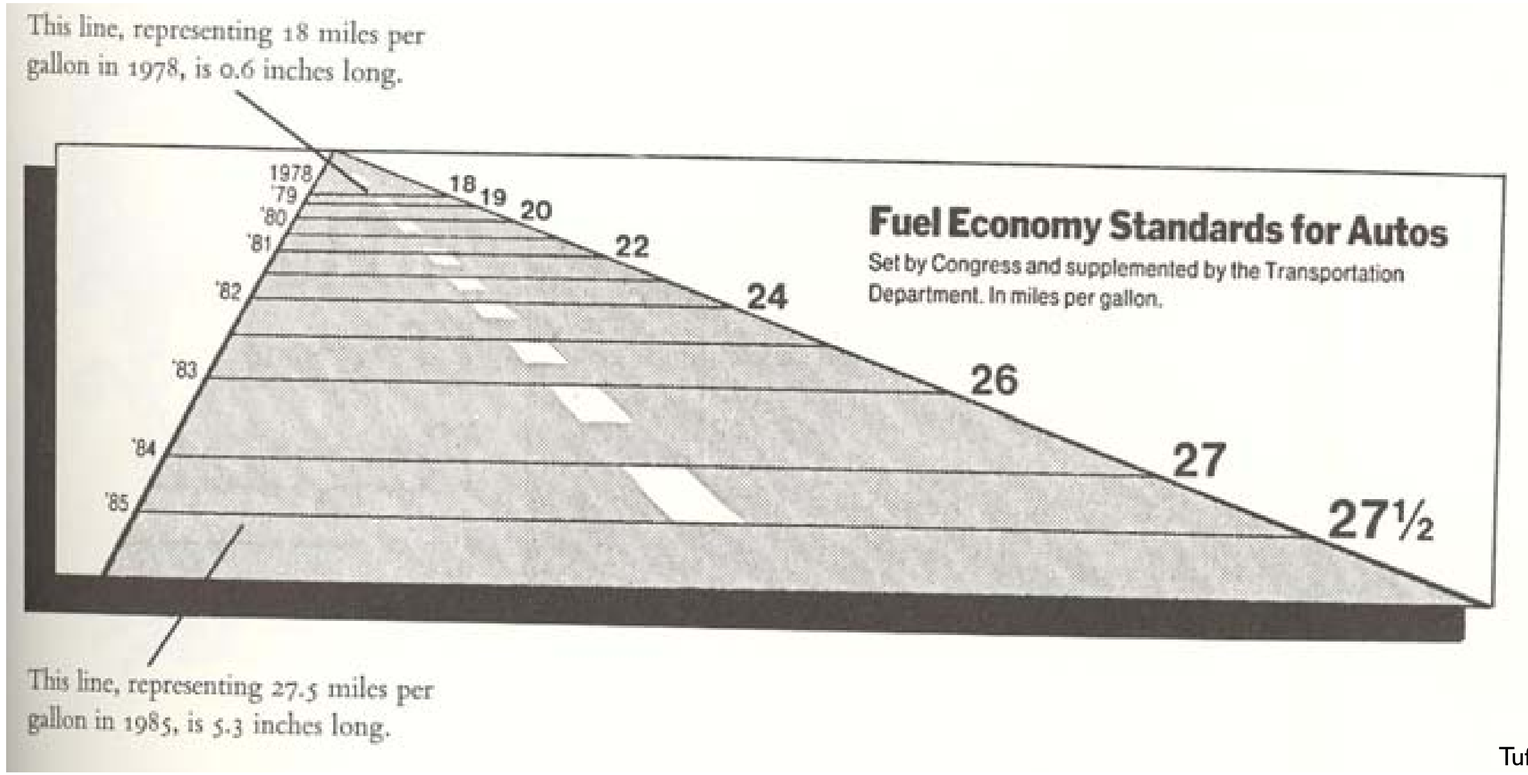


Tufte, VDQI

The Lie Factor

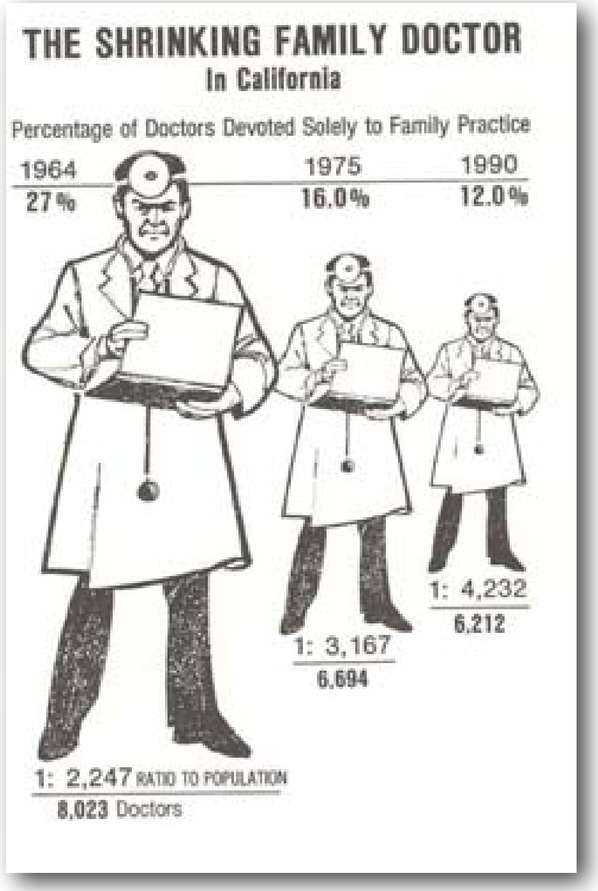
$$\frac{5.3 - 0.6}{0.6} \Big/ \frac{27.5 - 18}{18} = 14.8$$

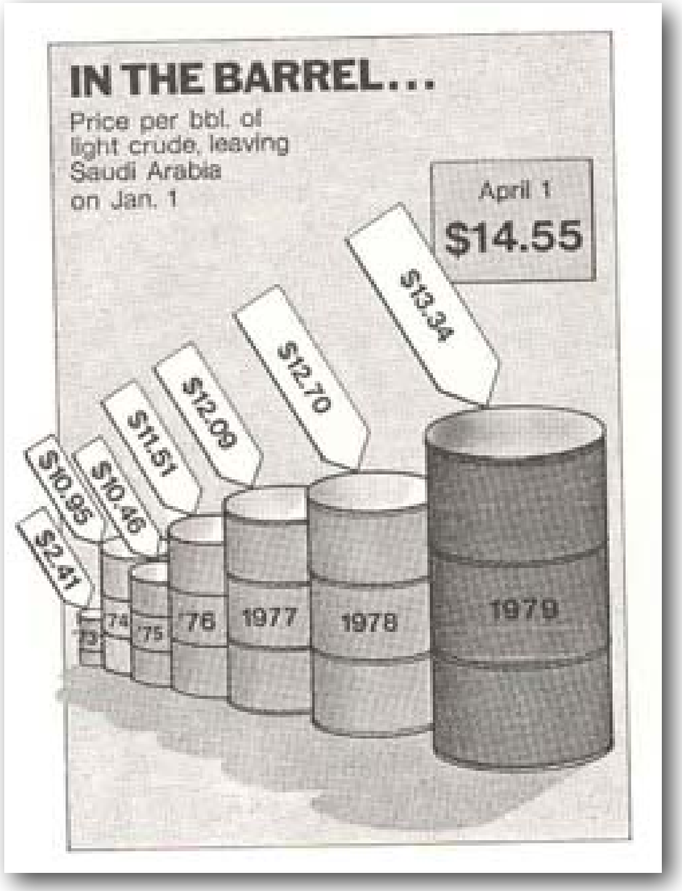
(Size of effect in graphic)/(size of effect in data)



Tufte, VDQI

The Lie Factor





Tufte, VDQI

Tufte's Integrity Principles

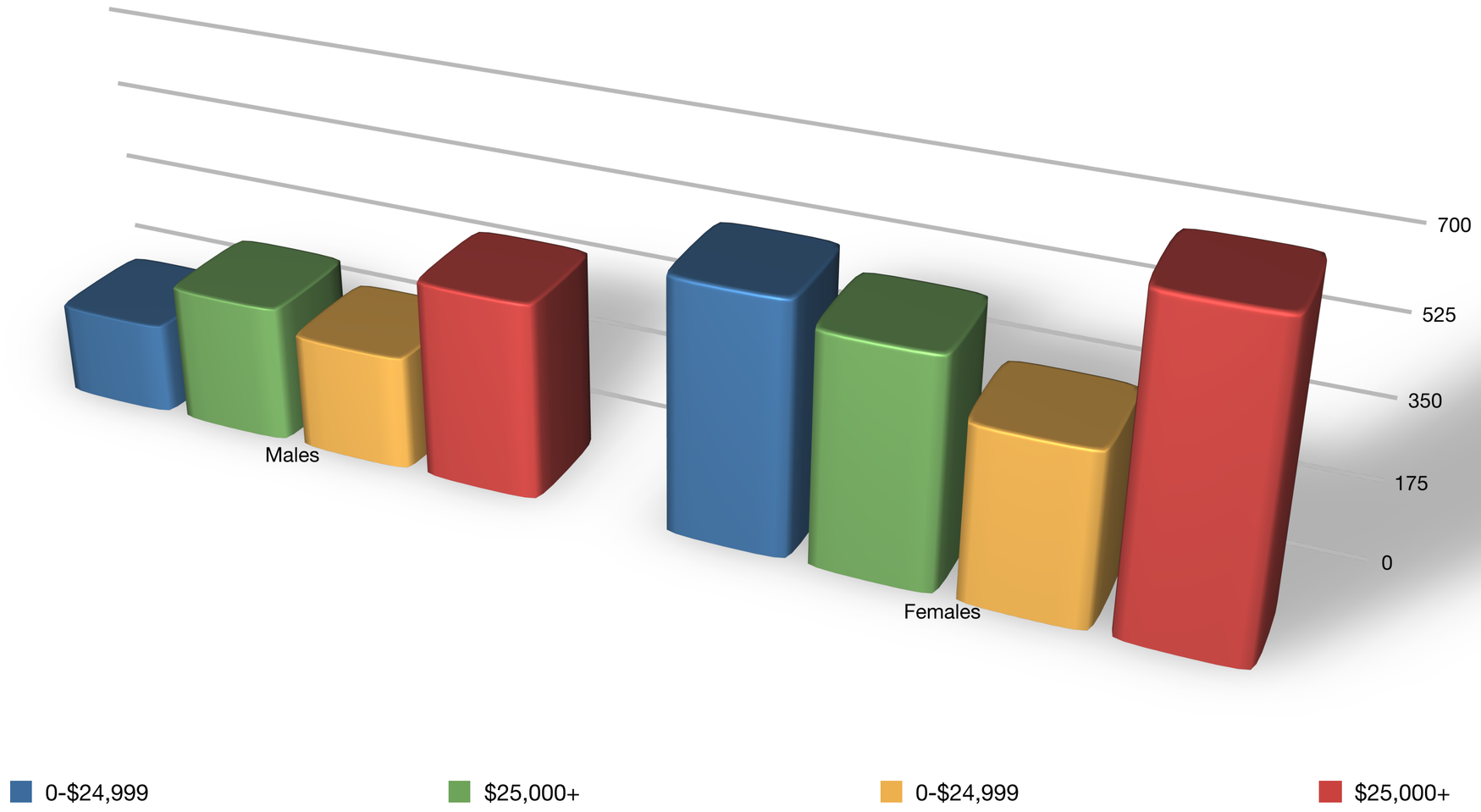
Show **data variation**, not design variation

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

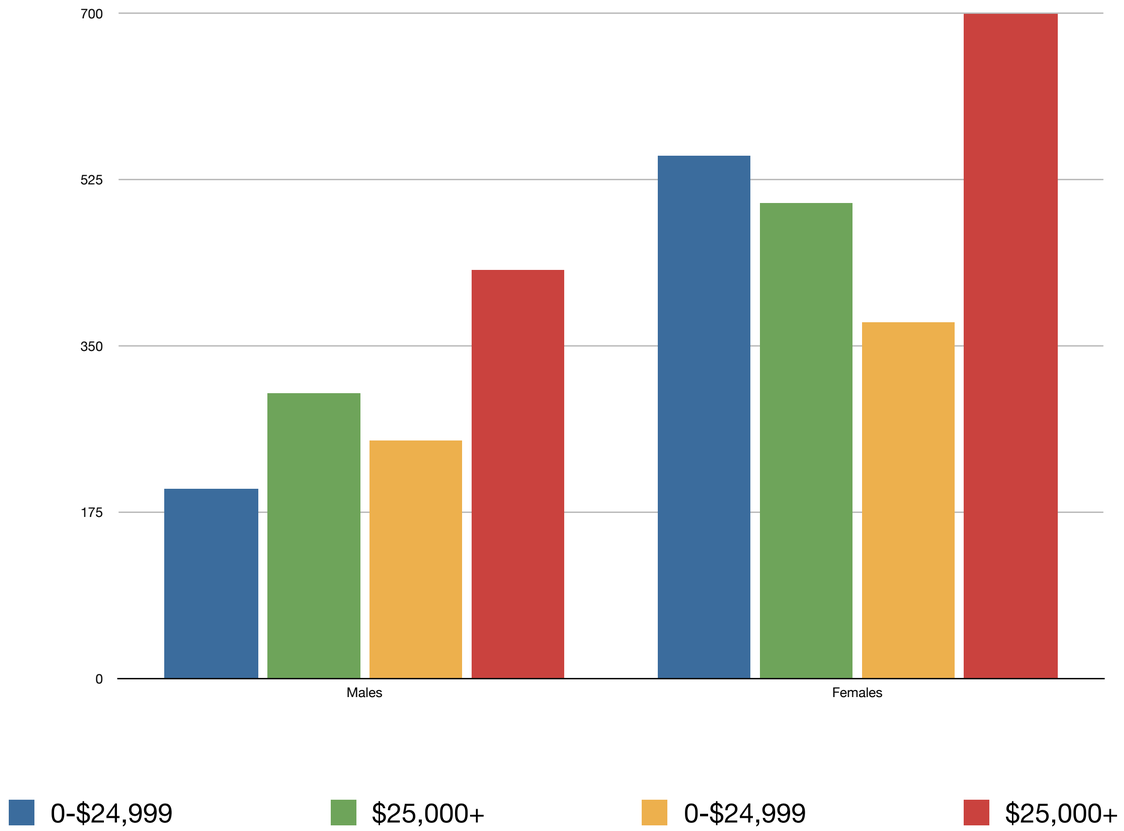
Size of the **graphic effect** should be **directly proportional to the numerical quantities** ("lie factor")

Visualization Design Principles

Maximize Data-Ink Ratio

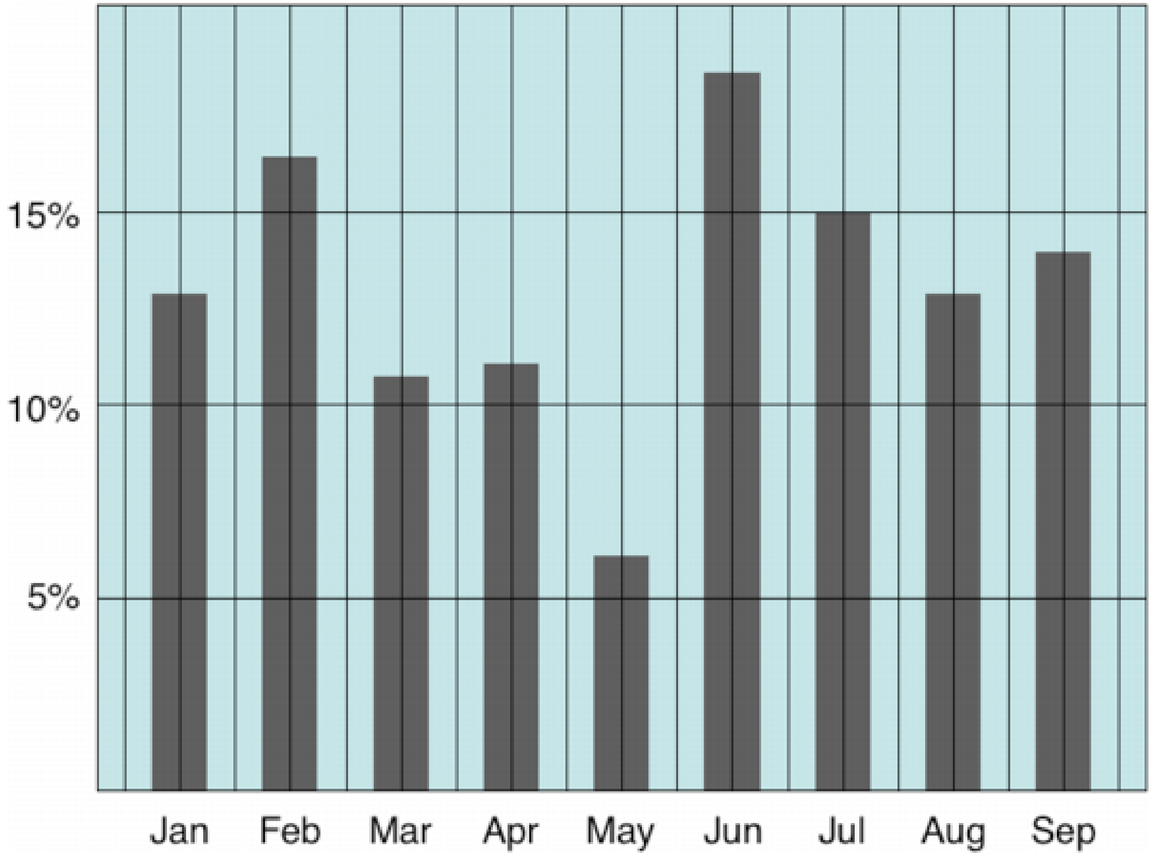


Maximize Data-Ink Ratio



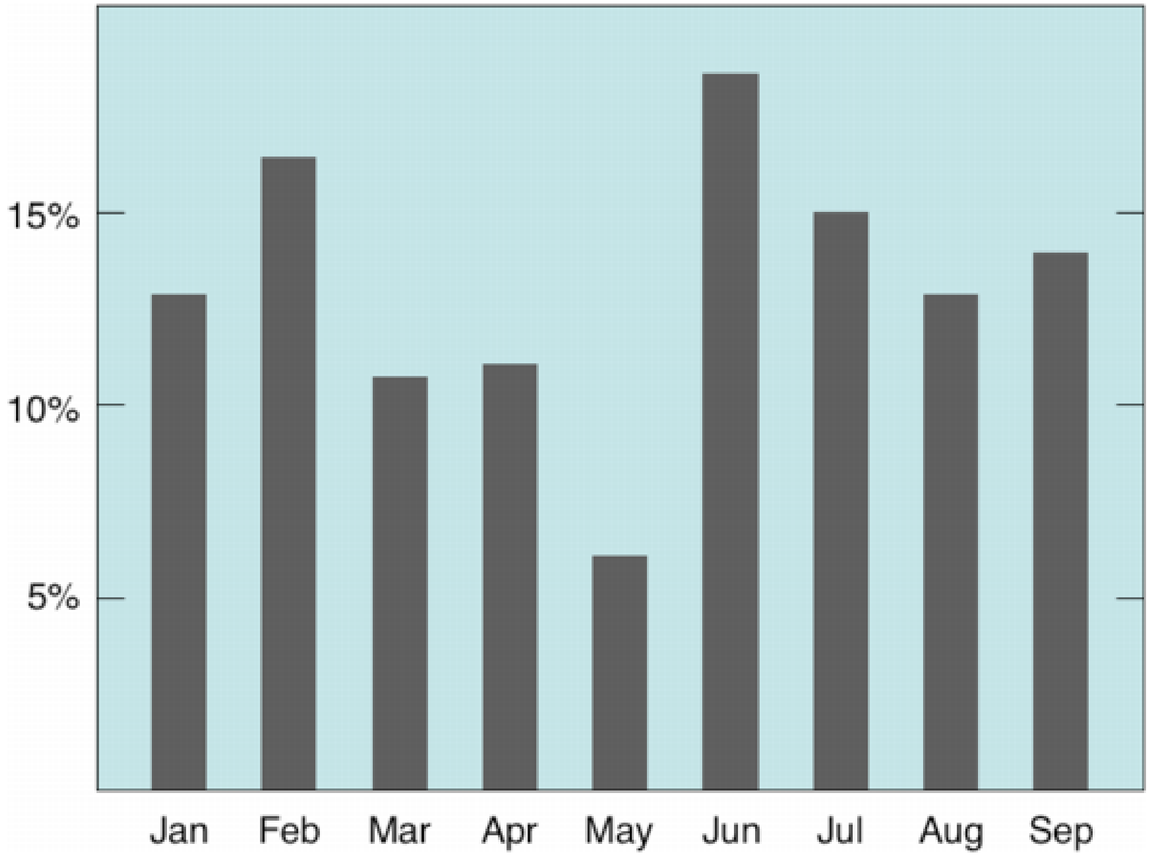
Avoid Chartjunk

Extraneous visual elements that distract from the message



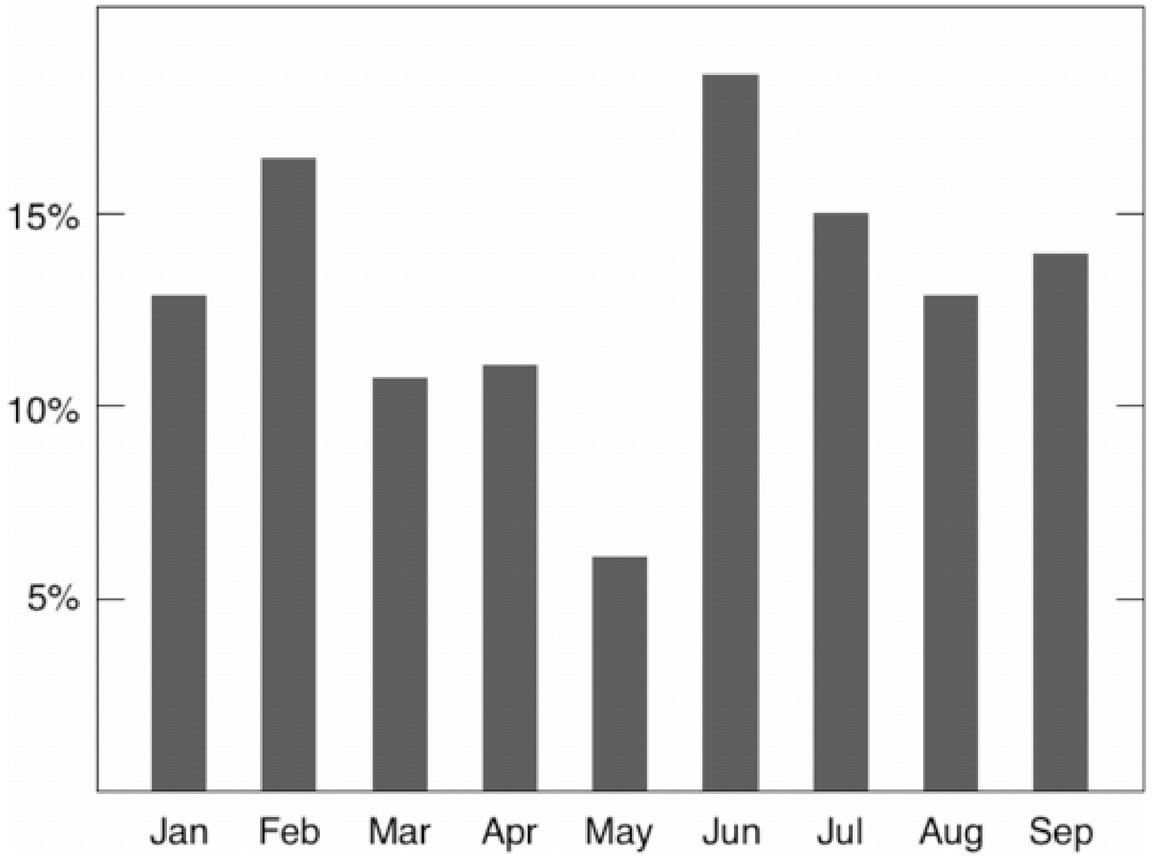
ongoing, Tim Brey

Avoid Chartjunk



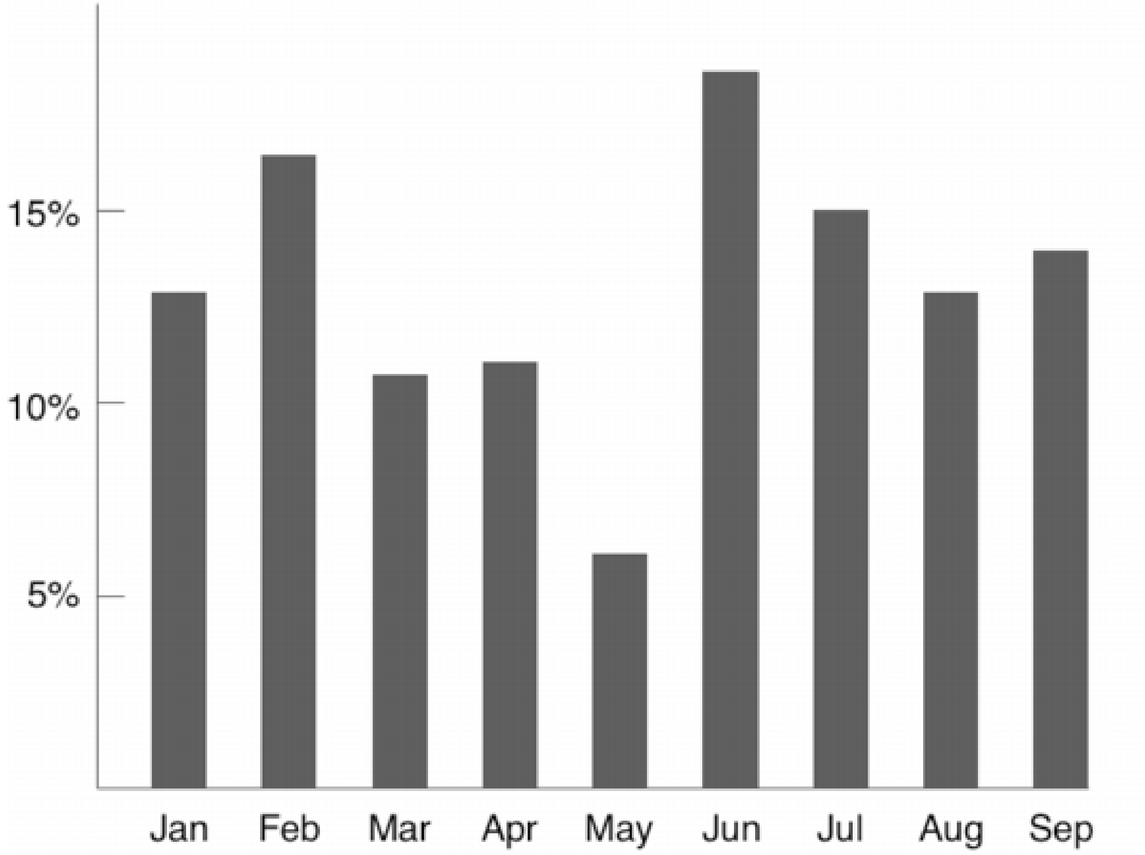
ongoing, Tim Brey

Avoid Chartjunk



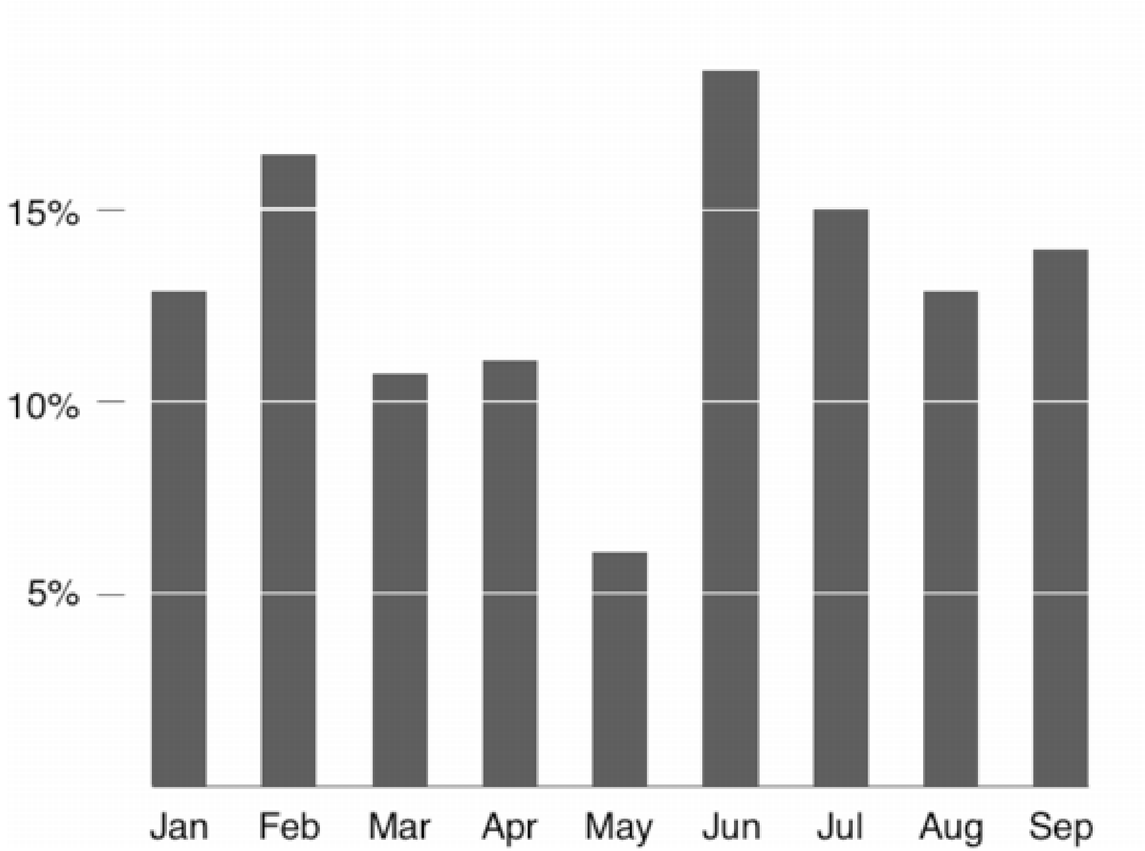
ongoing, Tim Brey

Avoid Chartjunk



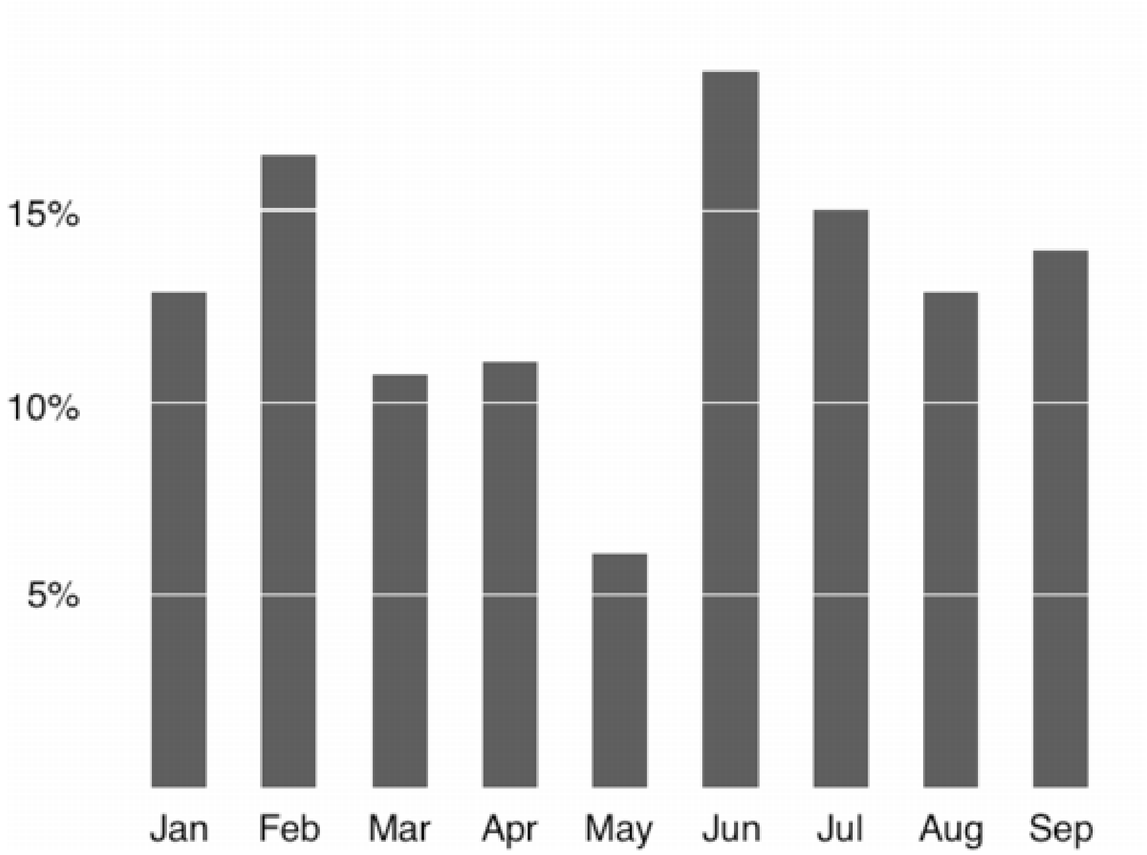
ongoing, Tim Brey

Avoid Chartjunk



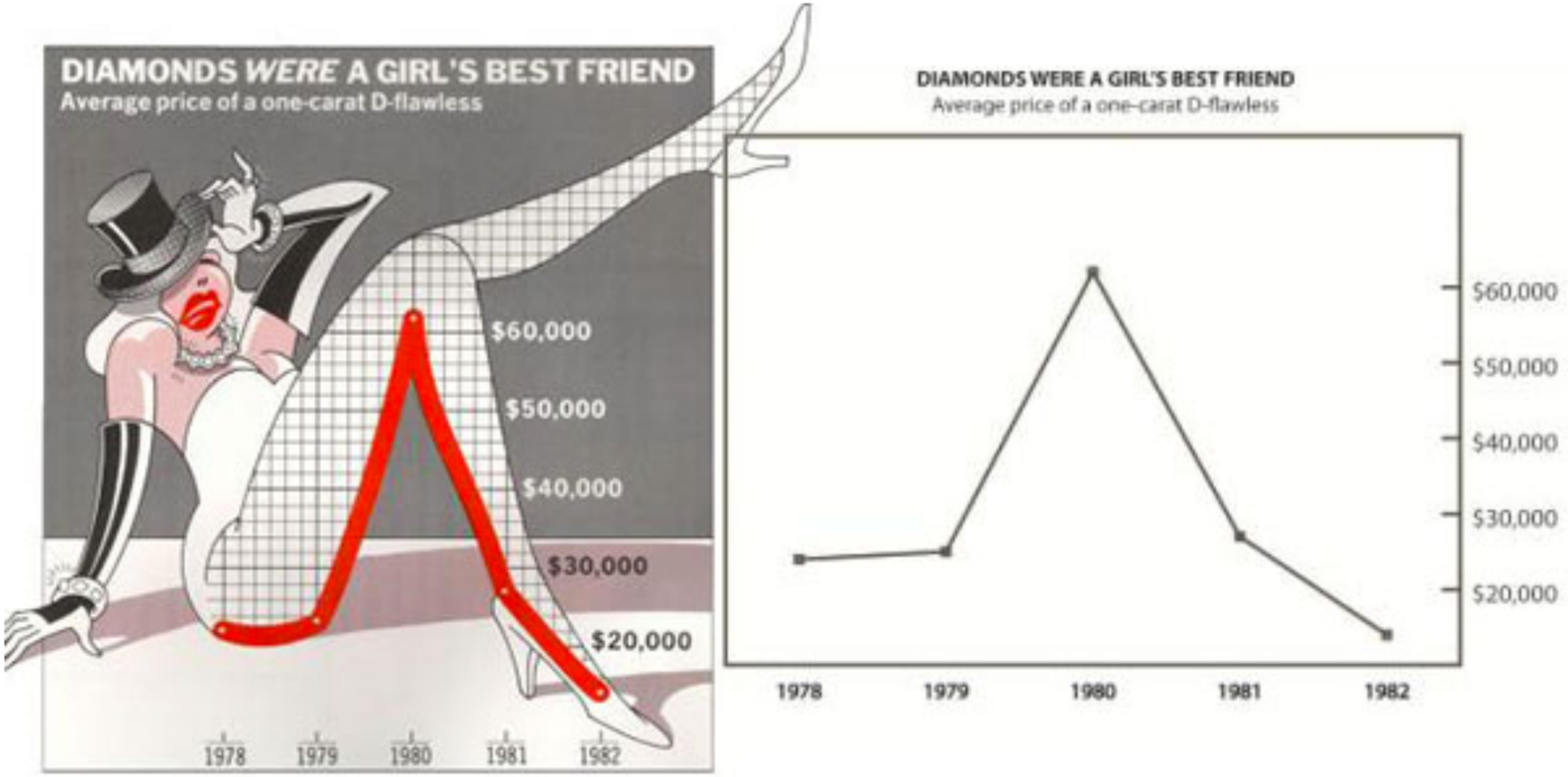
ongoing, Tim Brey

Avoid Chartjunk



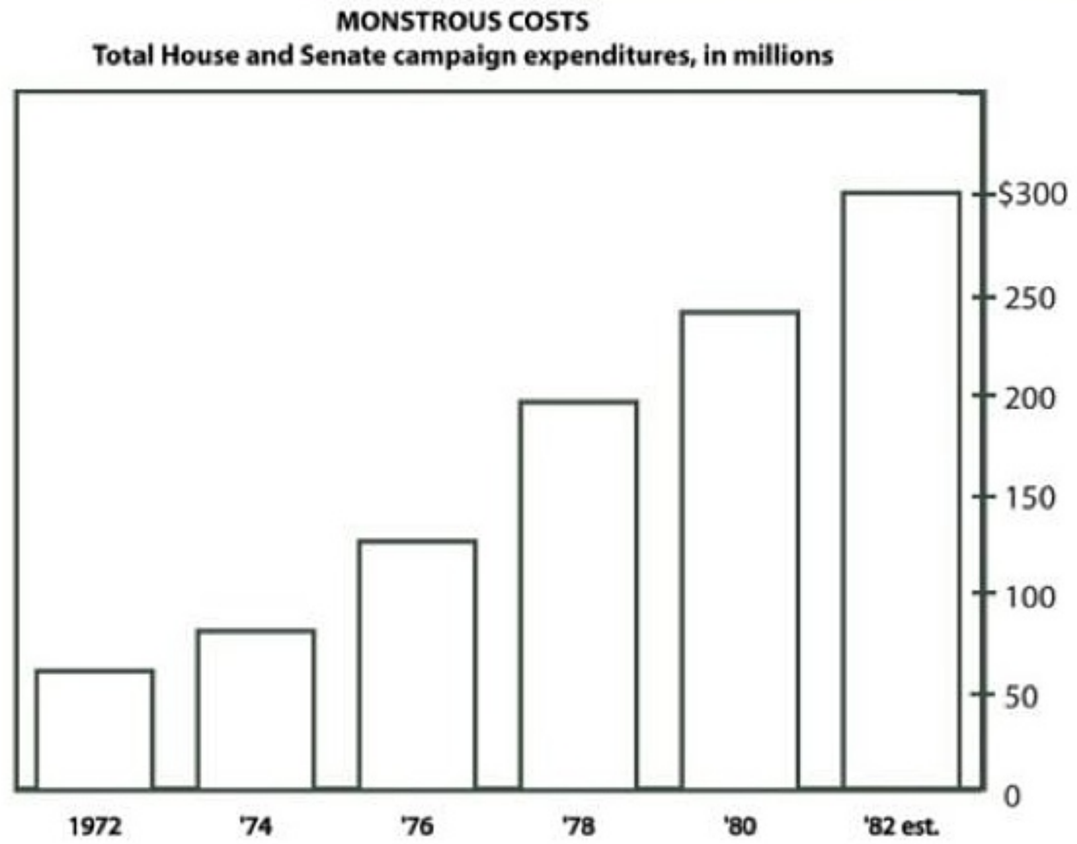
ongoing, Tim Brey

Which is better?



[Bateman et al. 2010]

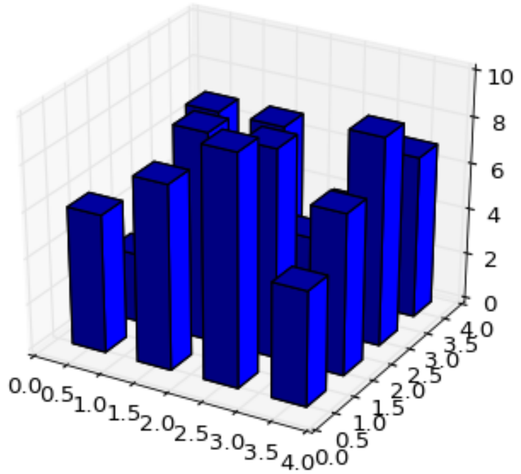
Which is better?

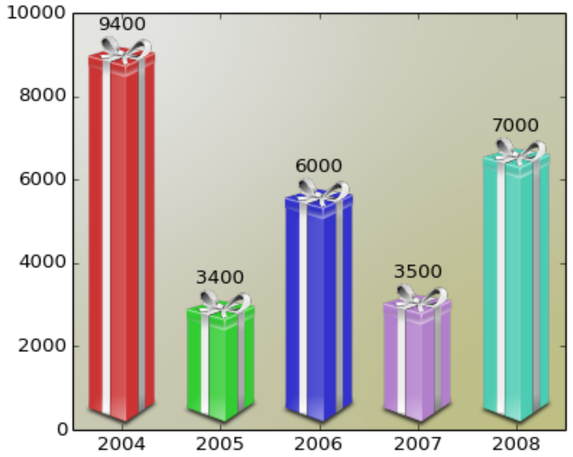


[Bateman et al. 2010]

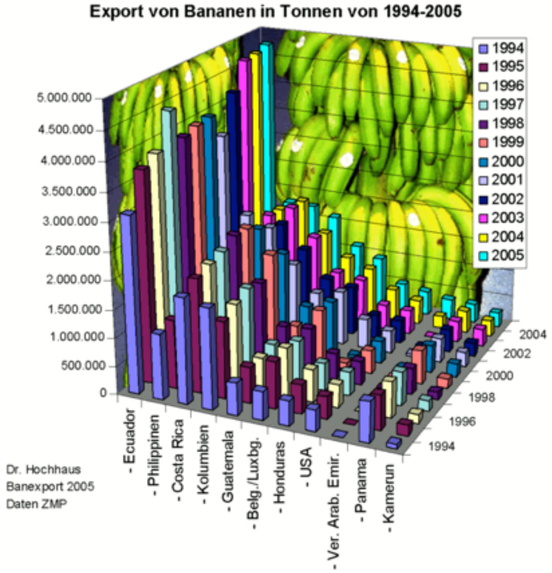
https://eagereyes.org/criticism/chart-junk-considered-useful-after-all

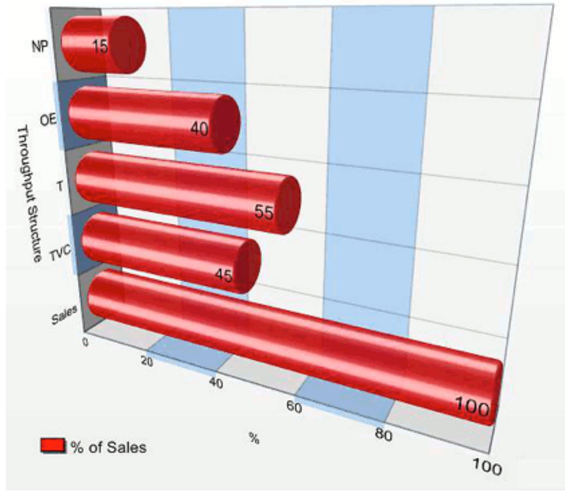
Don’t





matplotlib gallery





Excel Charts Blog

Design Critique

Design Critique

Four Ways to Slice Obama's 2013 Budget Proposal

Explore every nook and cranny of President Obama's federal budget proposal.

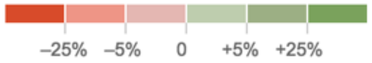
All Spending | Types of Spending | Changes | Department Totals

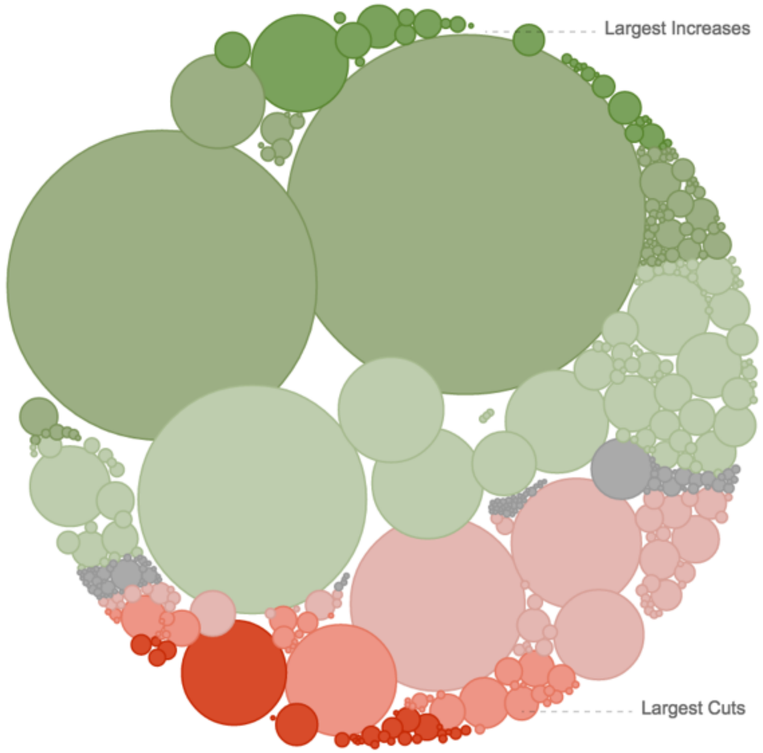
How $3.7 Trillion Is Spent

Mr. Obama's budget proposal includes $3.7 trillion in spending in 2013, and forecasts a $901 billion deficit.

Circles are sized according to the proposed spending.

Color shows amount of cut or increase from 2012.





http://goo.gl/DA67PG

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

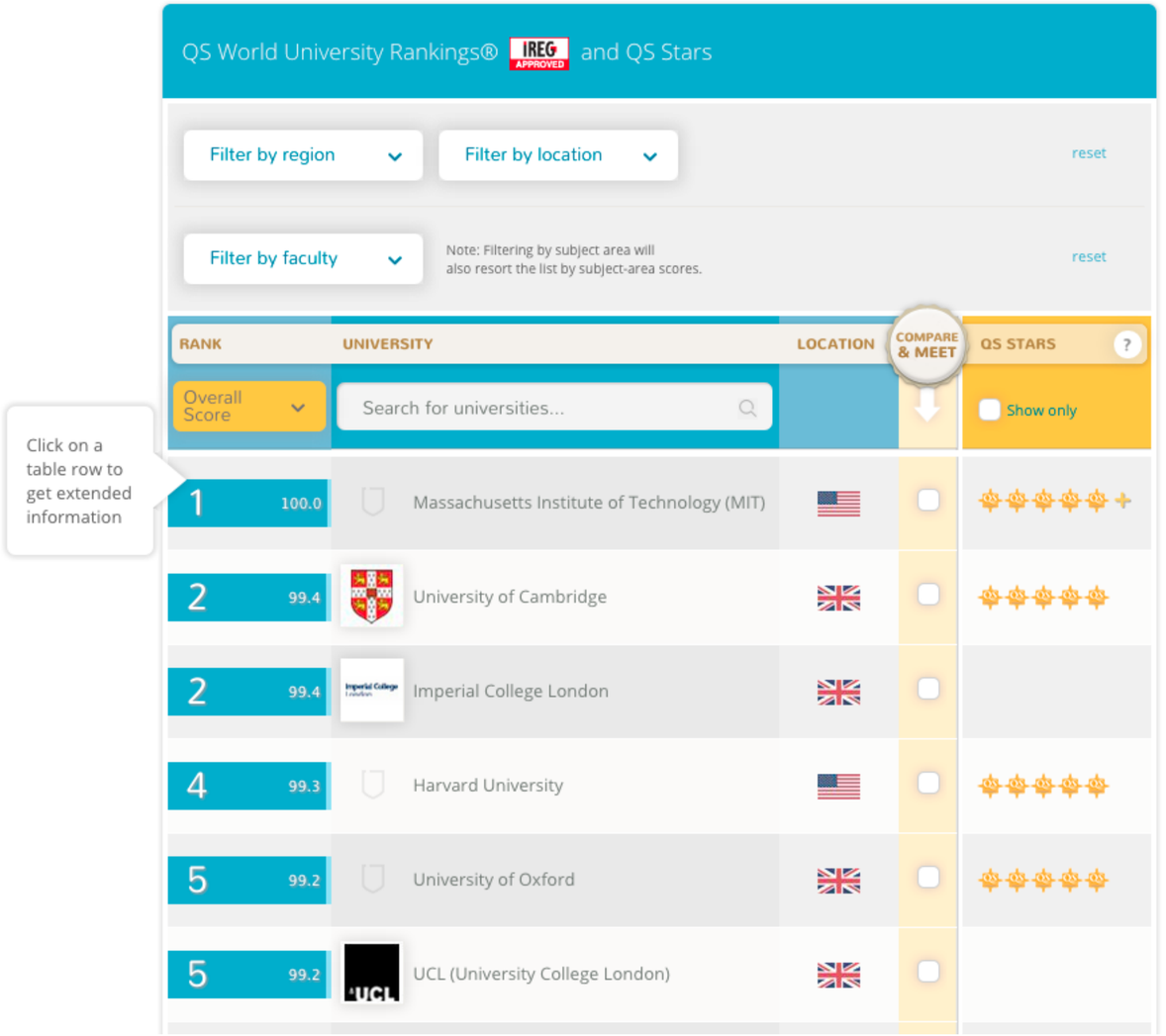
Find good universities with a high faculty student ratio.

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

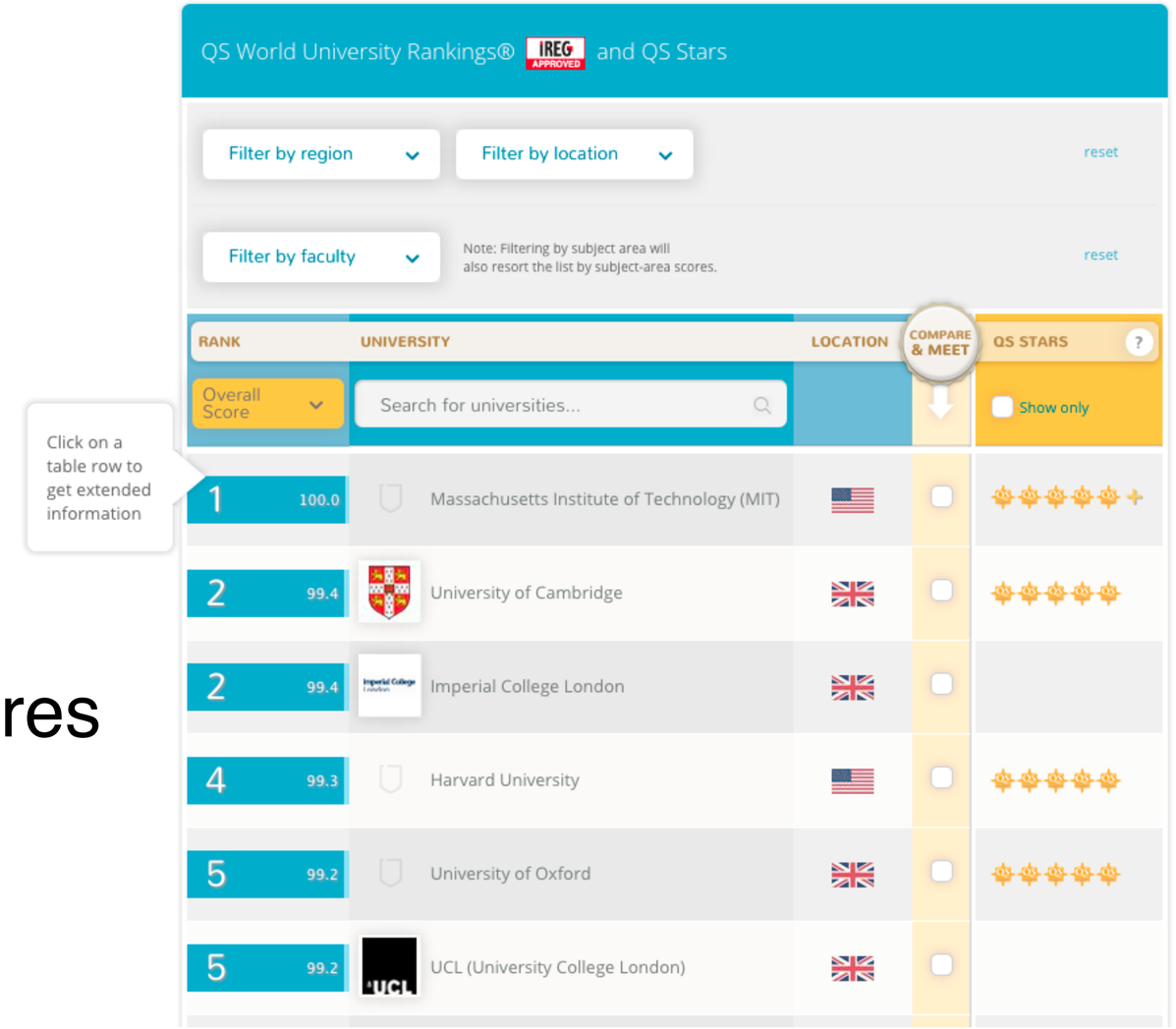


Example 2

Contrast Harvard's reputation scores with MIT's

Match up Harvard with Yale

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

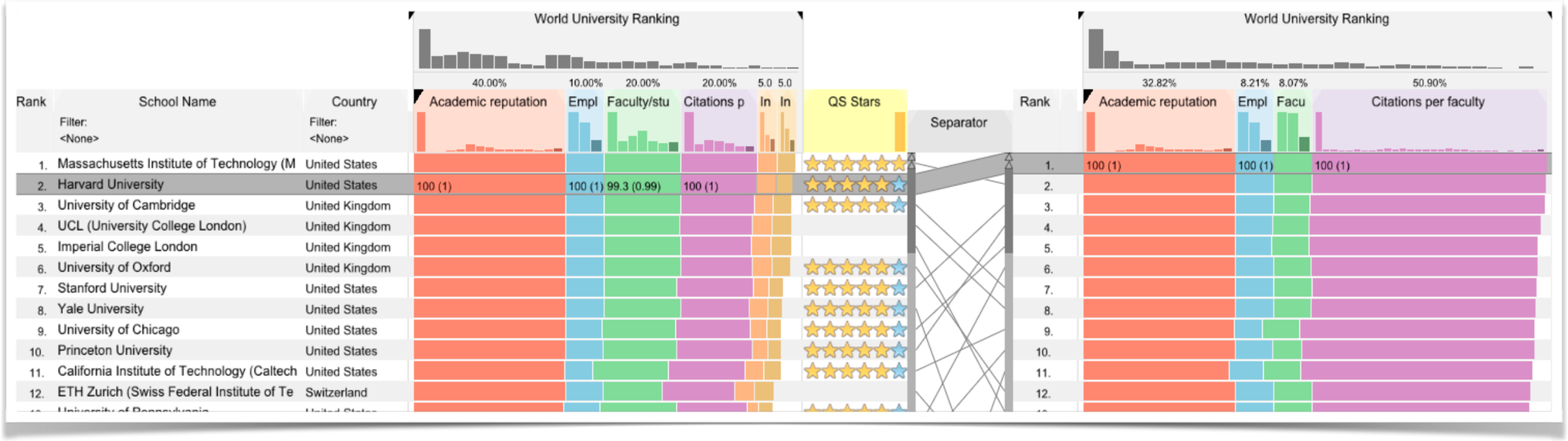


Example 3

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

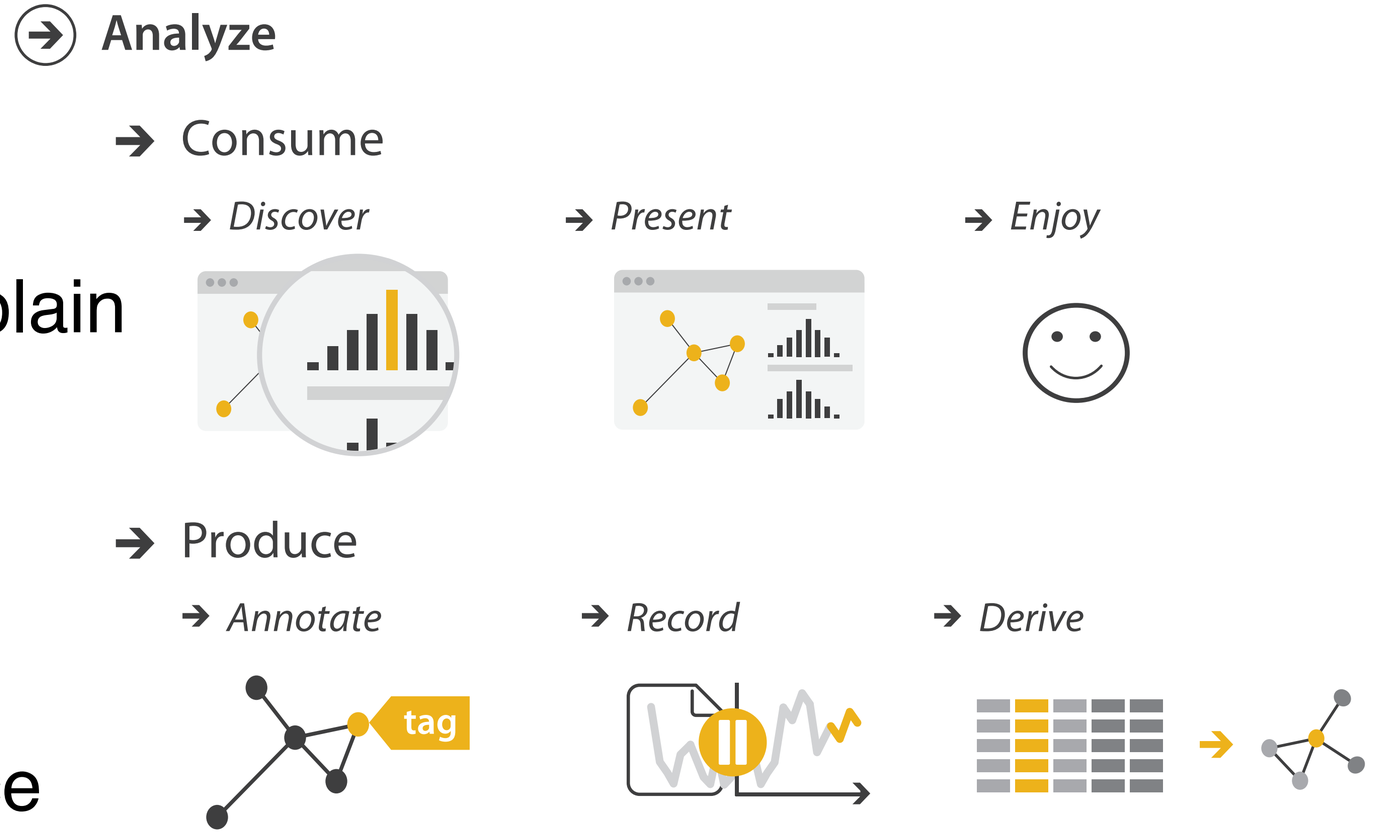
classic split: explore vs explain

enjoy: casual, social

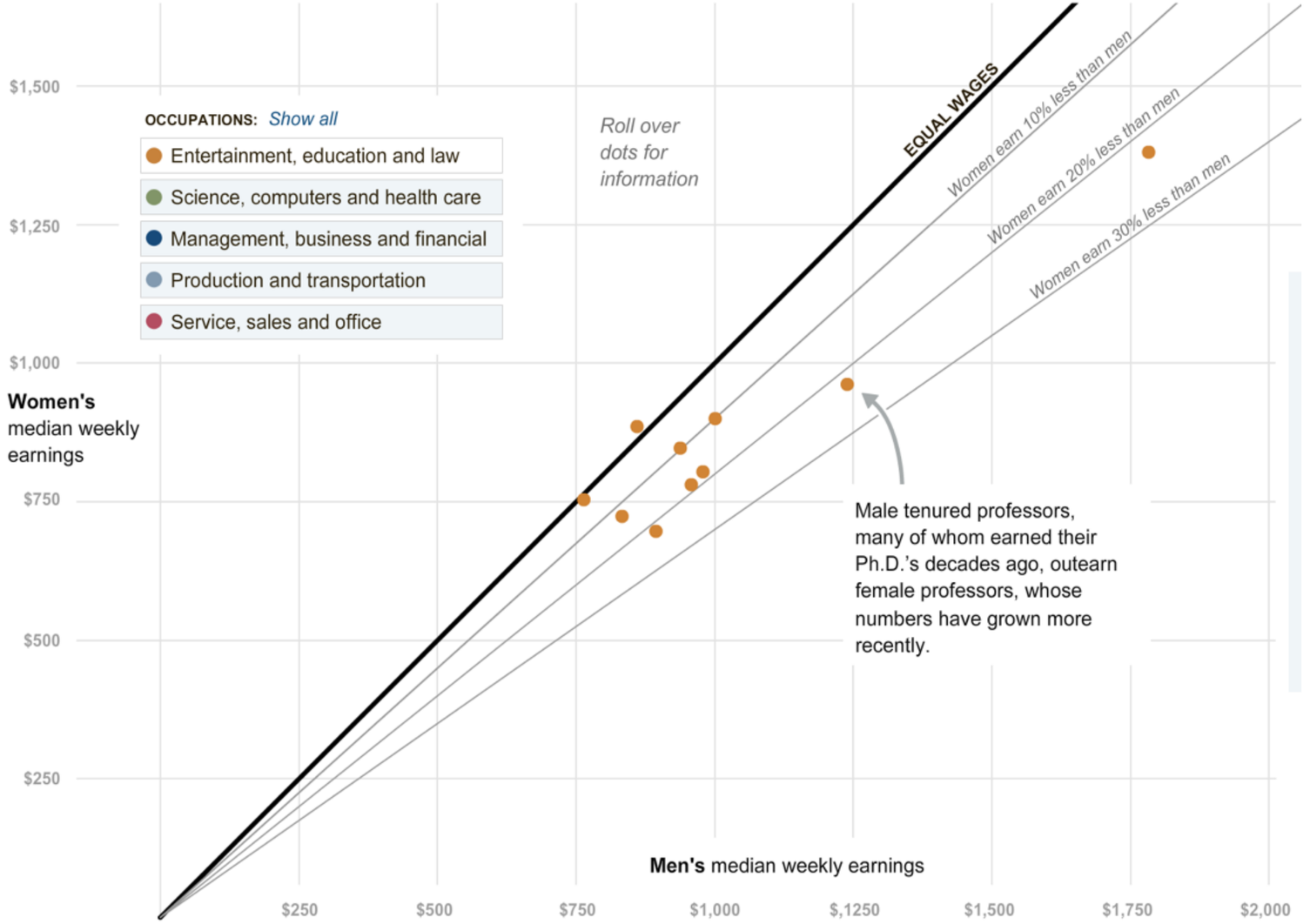
Produce

Annotate, record

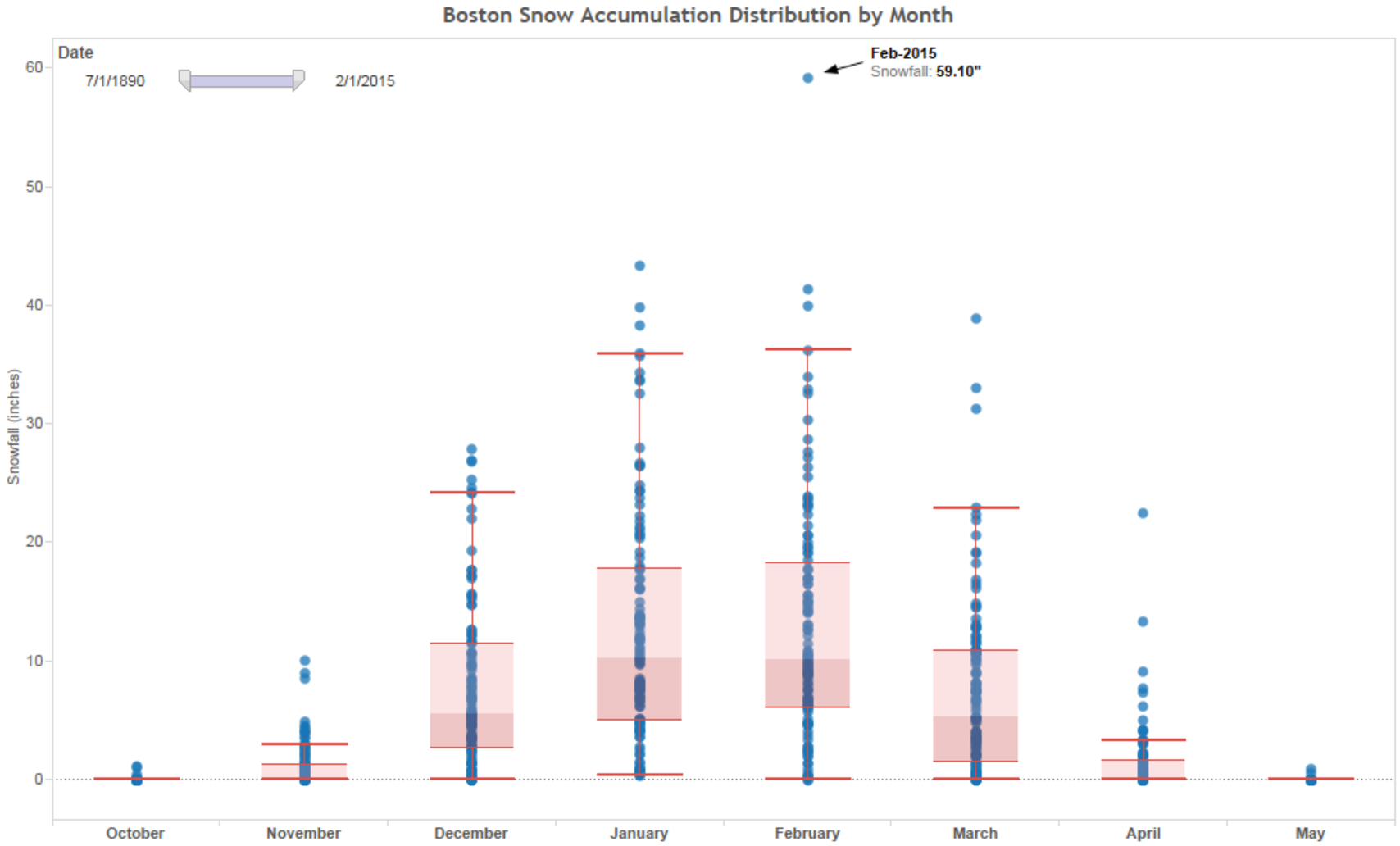
Derive: crucial design choice



Example: Annotate

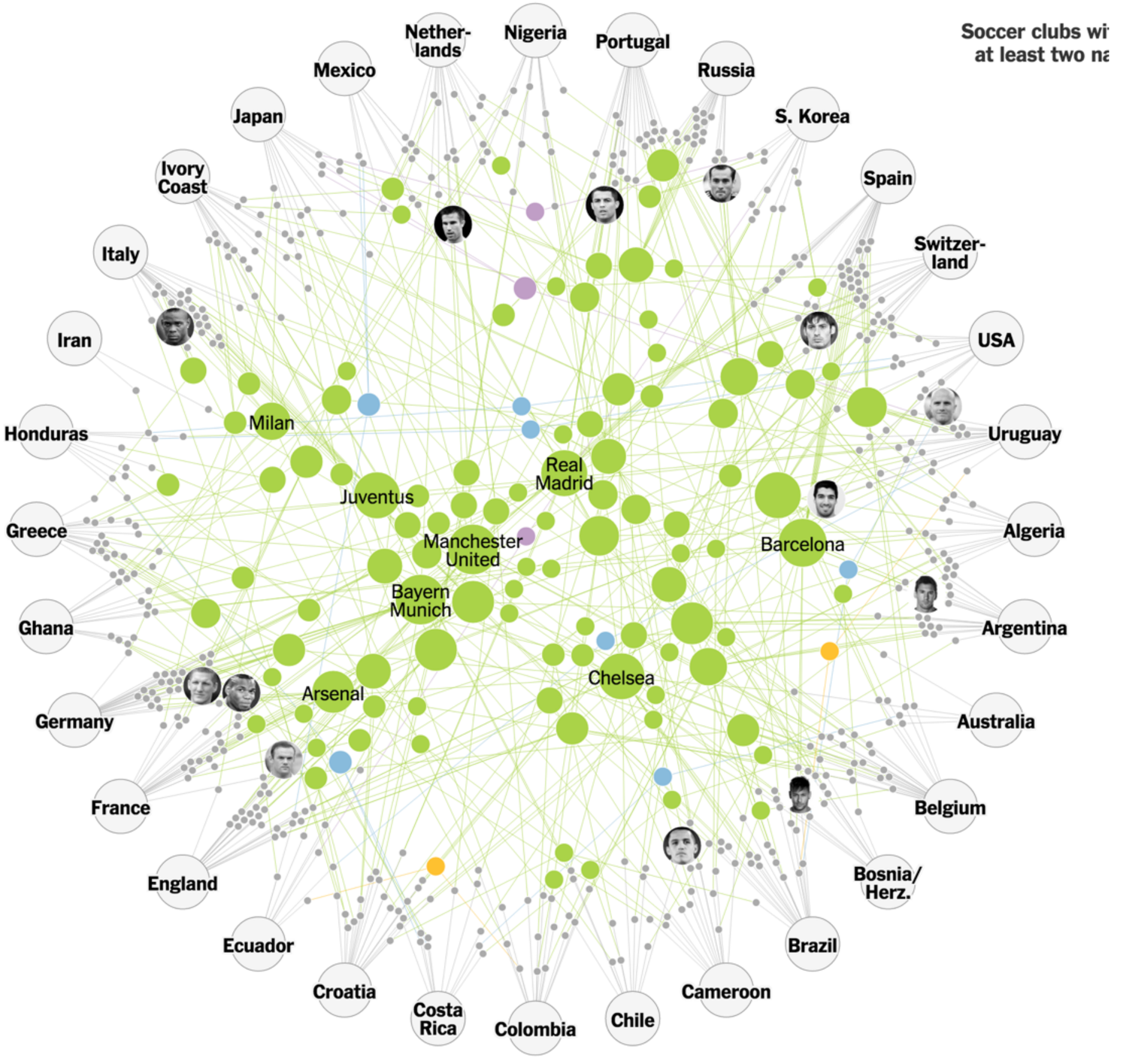


Example: Derive



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?

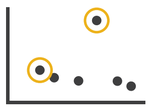
one, some, all

Search

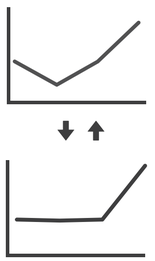
	Target known	Target unknown
Location known	*Lookup*	*Browse*
Location unknown	*Locate*	*Explore*

Query

→ Identify

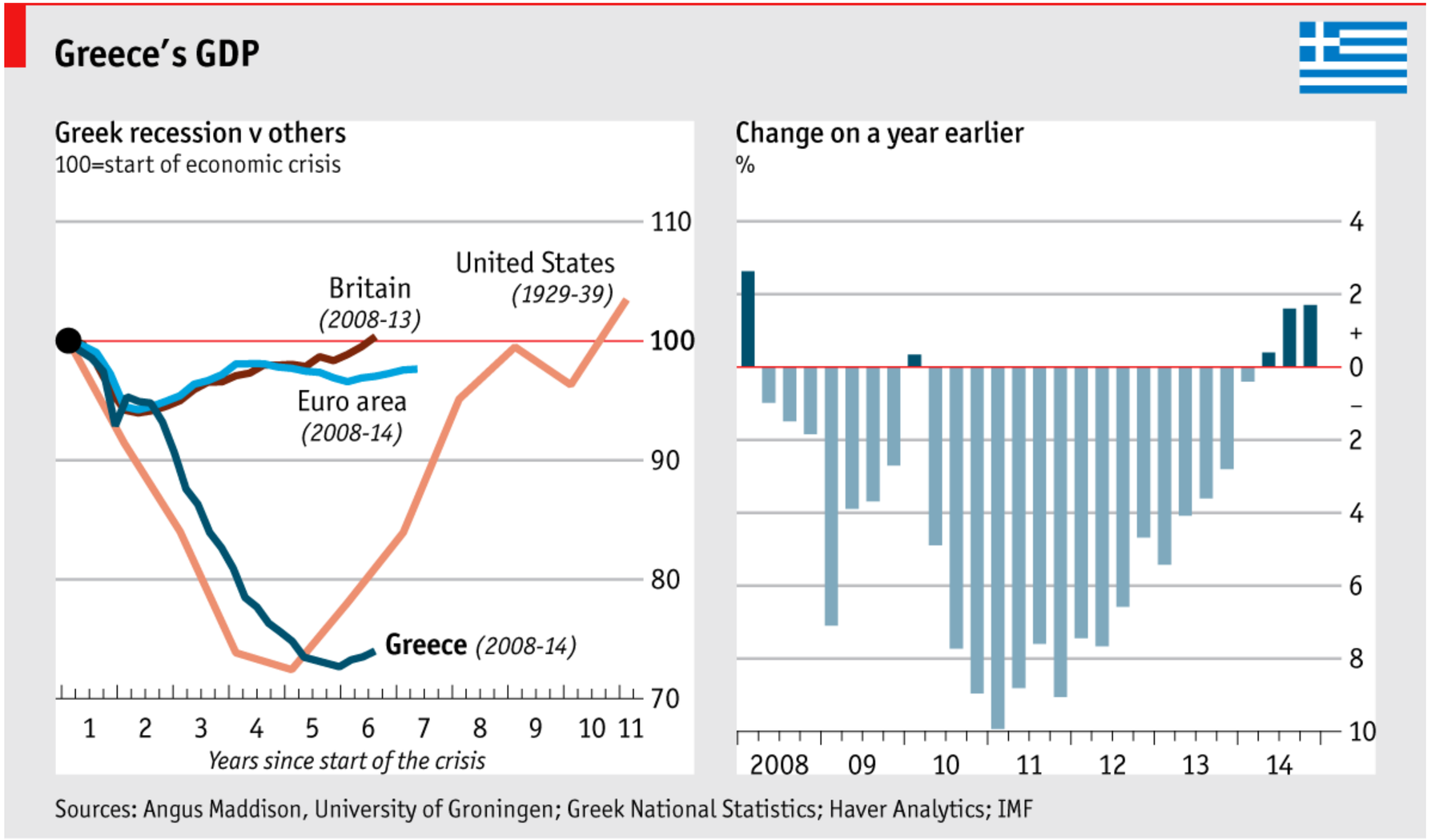


→ Compare

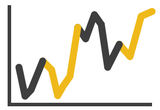
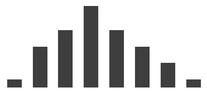
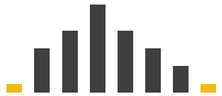
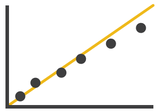
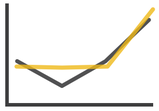
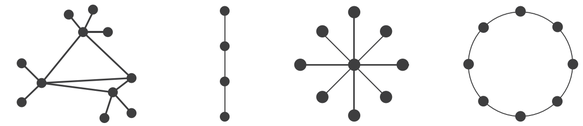
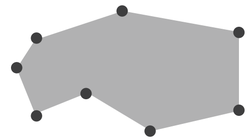


→ Summarize

Example Compare (& Derive)



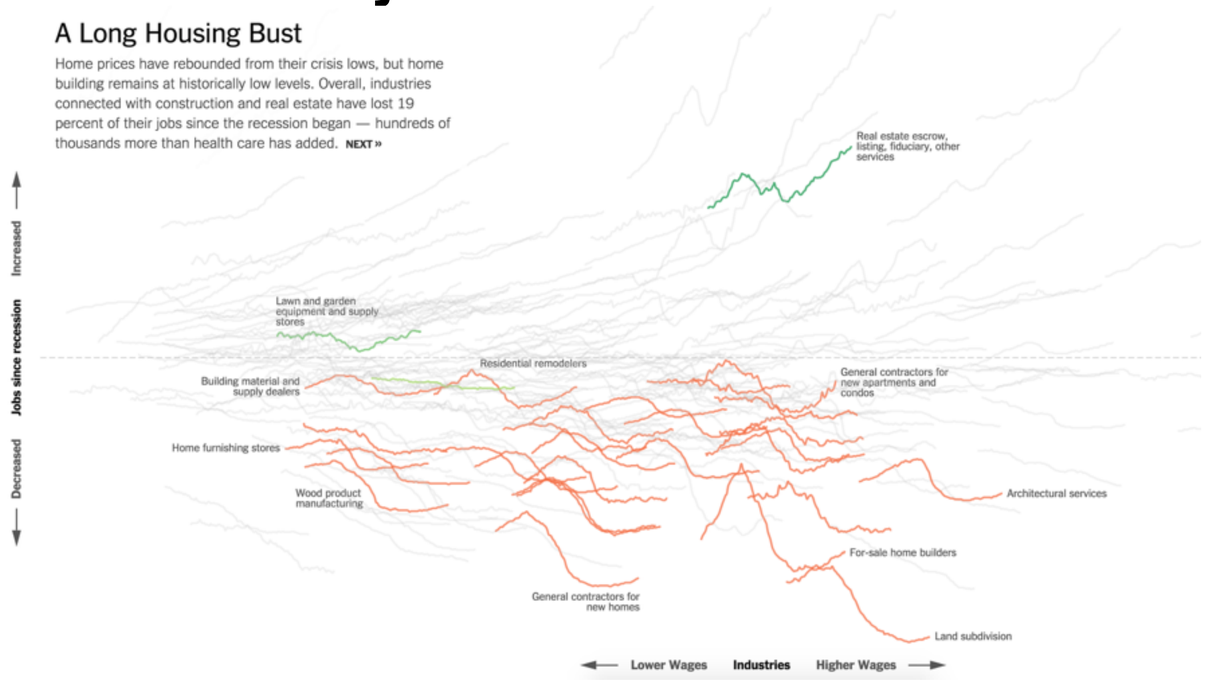
Why: Targets

- 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

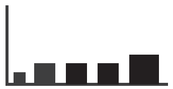


How? A Preview

Encode

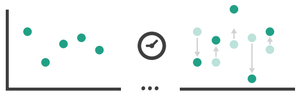
- **Arrange**
  - Express
  - Separate
  - Order
  - Align
  - Use



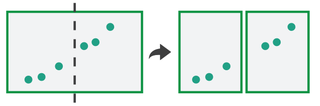

Manipulate

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



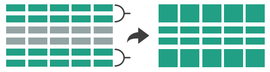
Facet

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


Reduce

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


Next time: Evaluation