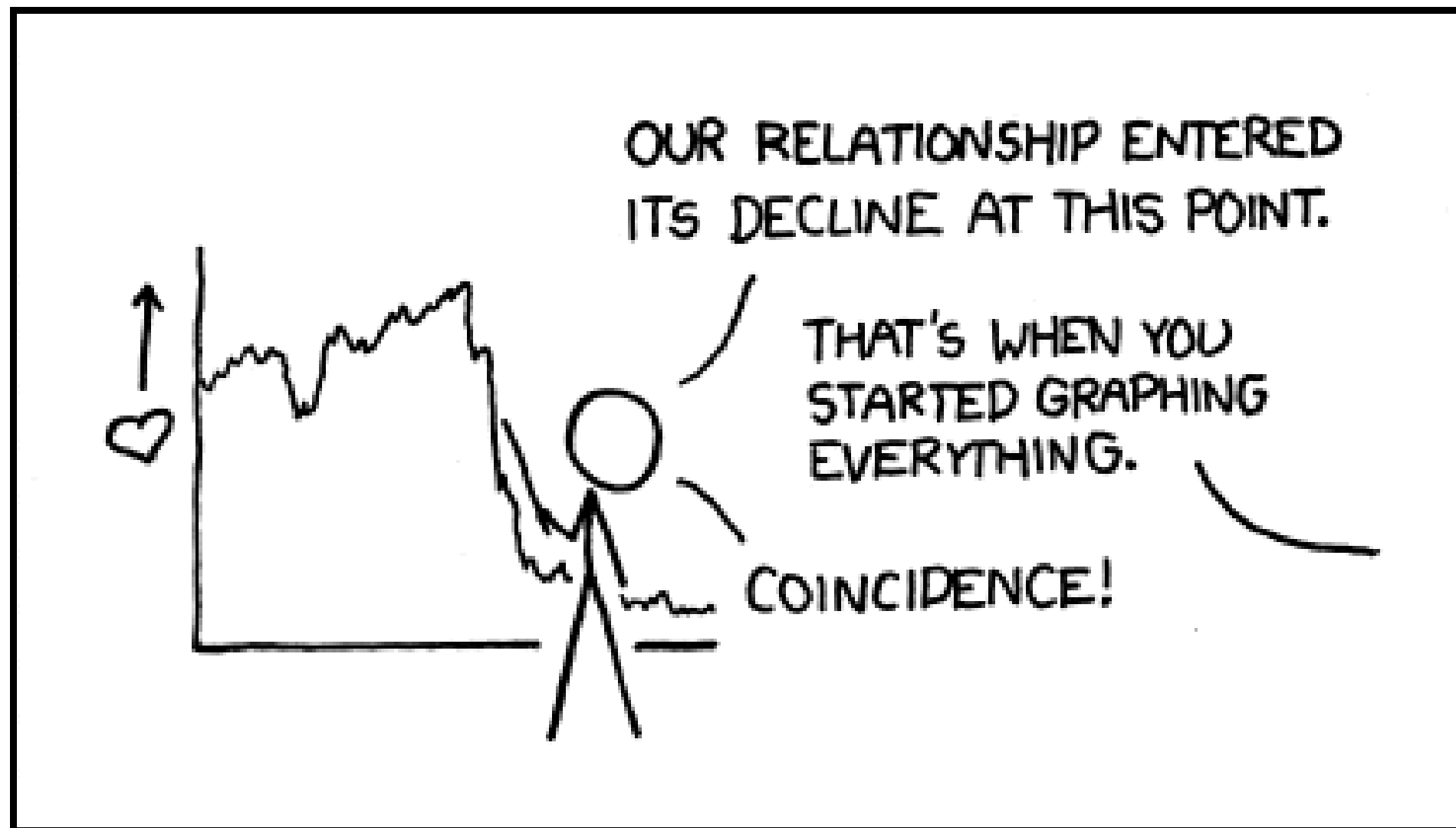


Data Visualization: Supplementary slides

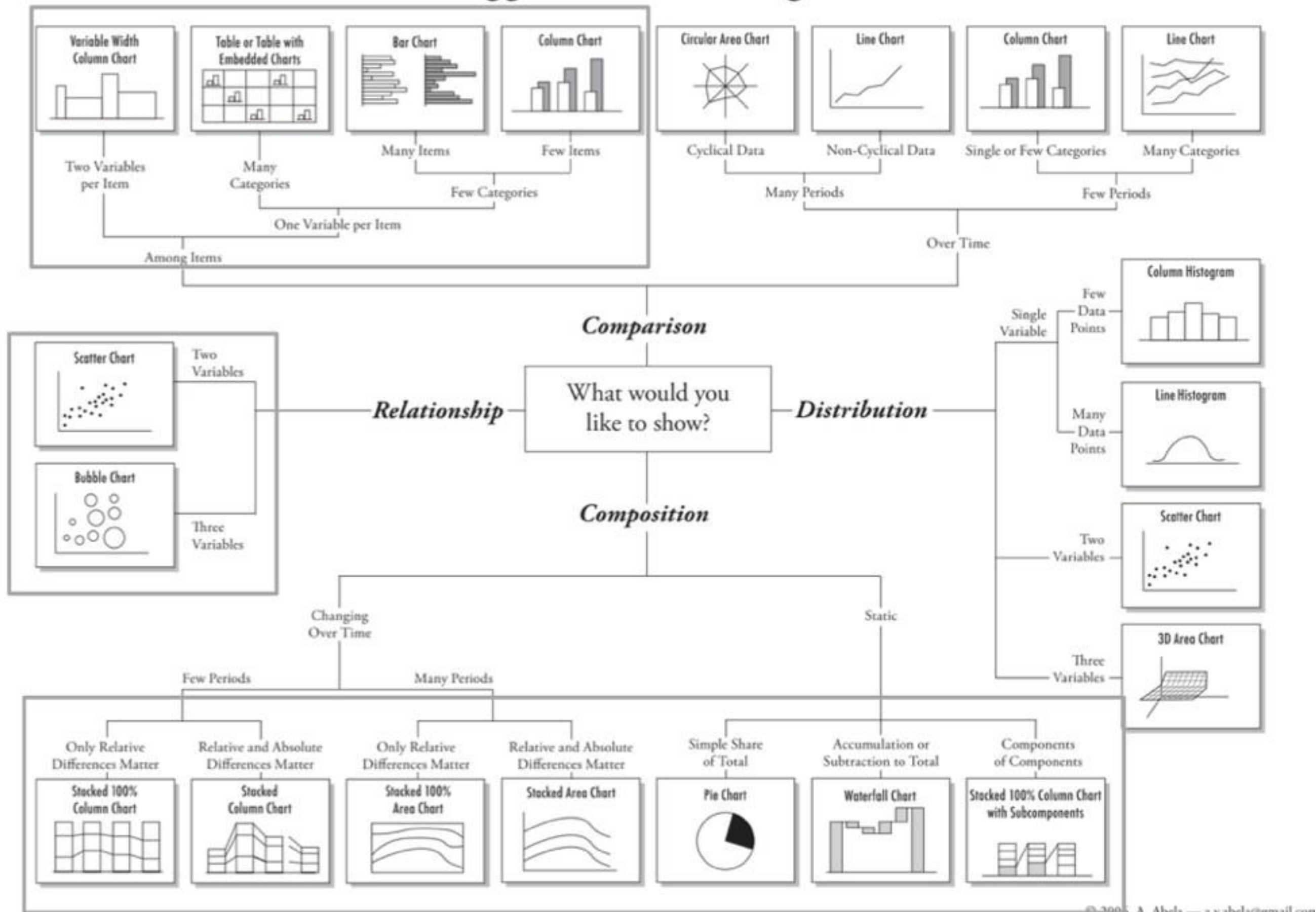
Sungkyu Jung

Statistical Graph Types



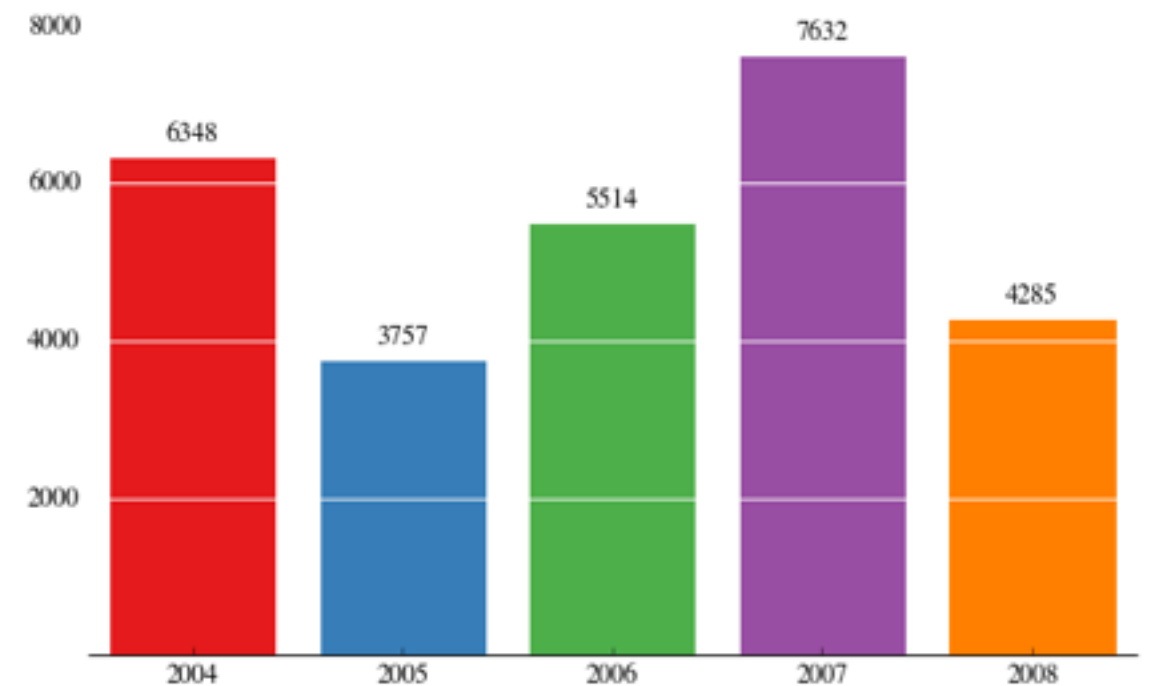
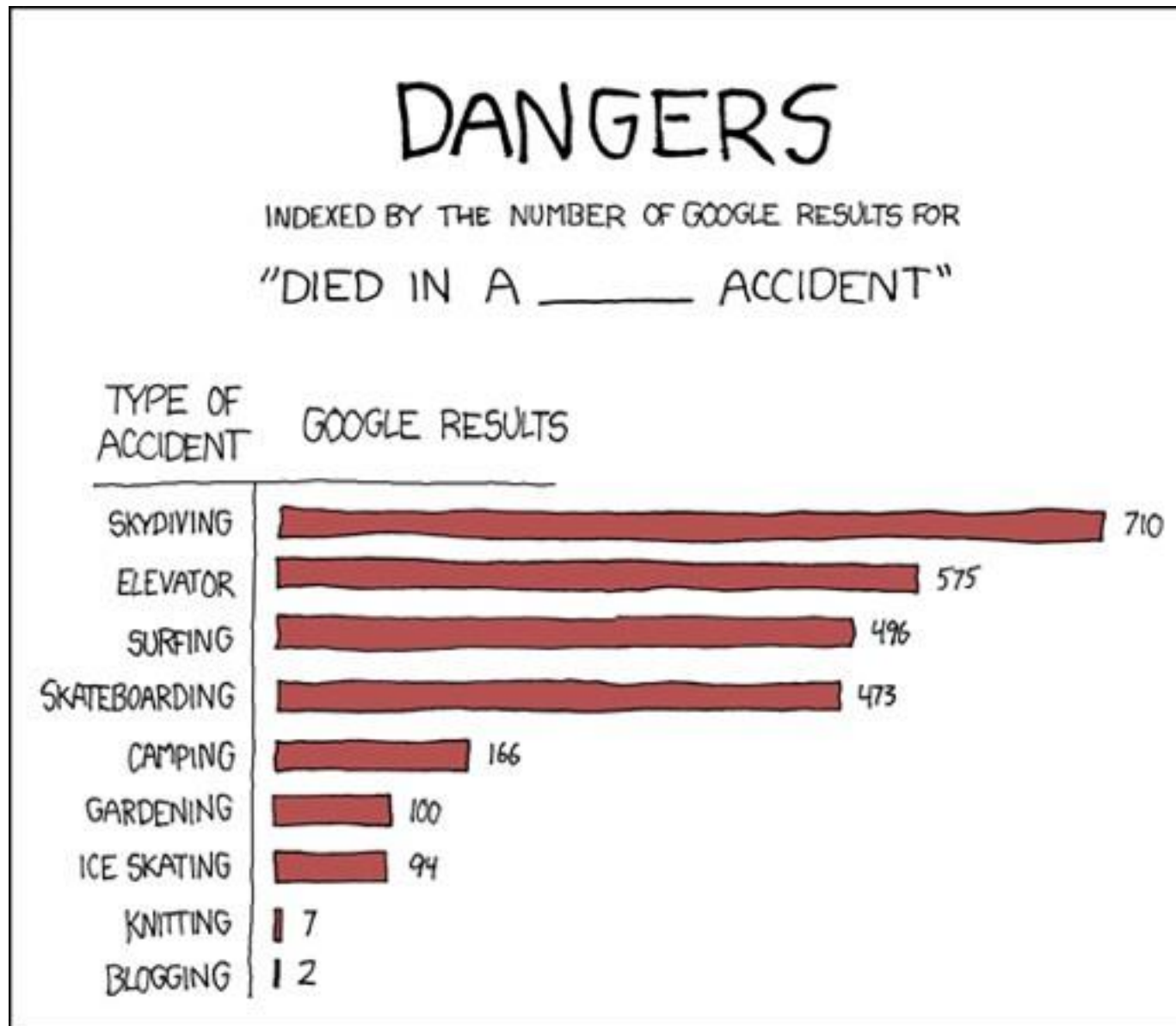
<http://xkcd.com/523/>

Chart Suggestions—A Thought-Starter



Comparisons

Bar Chart

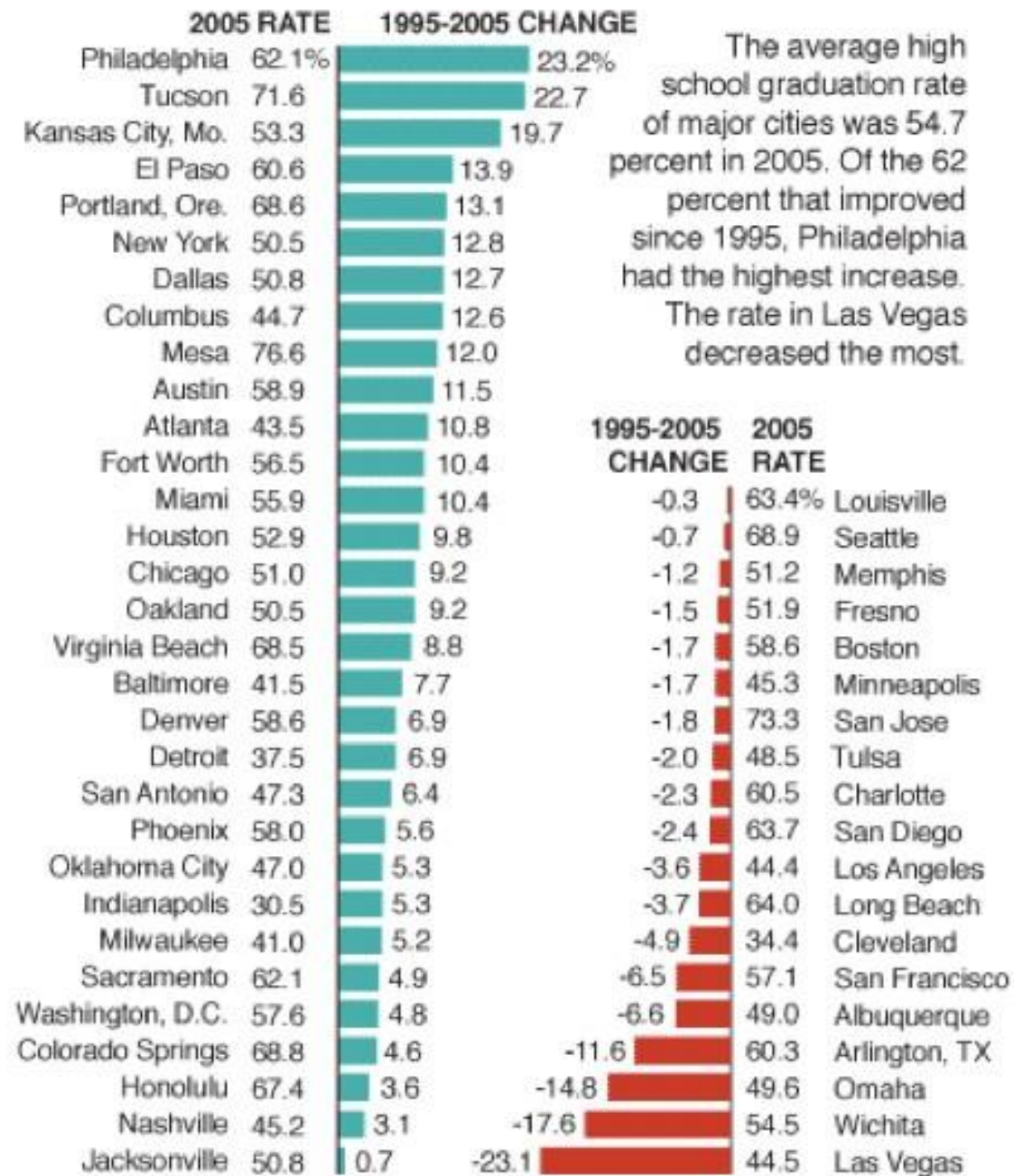


<http://nbviewer.ipython.org/gist/olgabot/5357268>

Direction

Graduation rates up in most cities

Graduation rate for principal school district of the largest cities



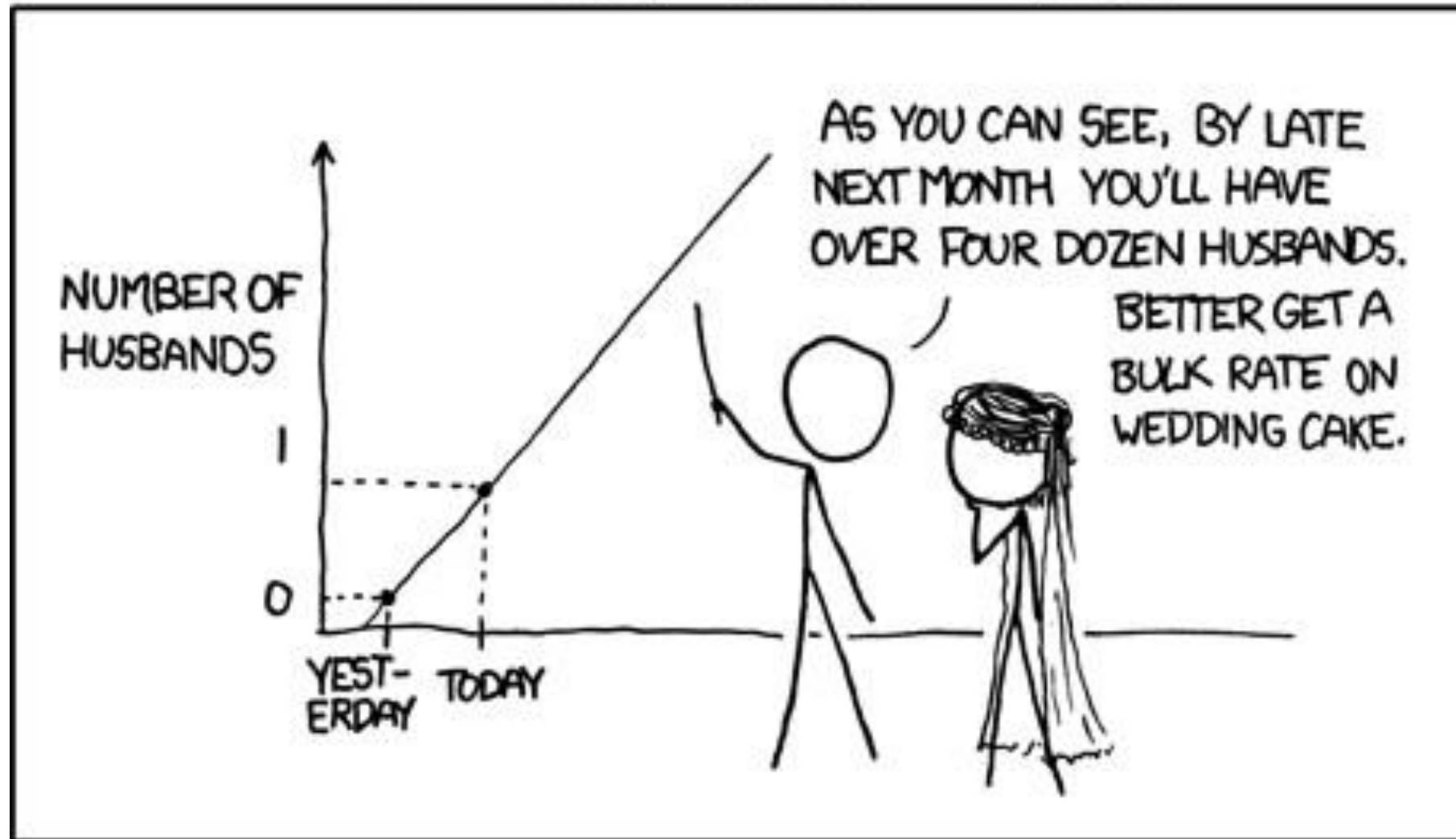
SOURCE: EPE Research Center

AP

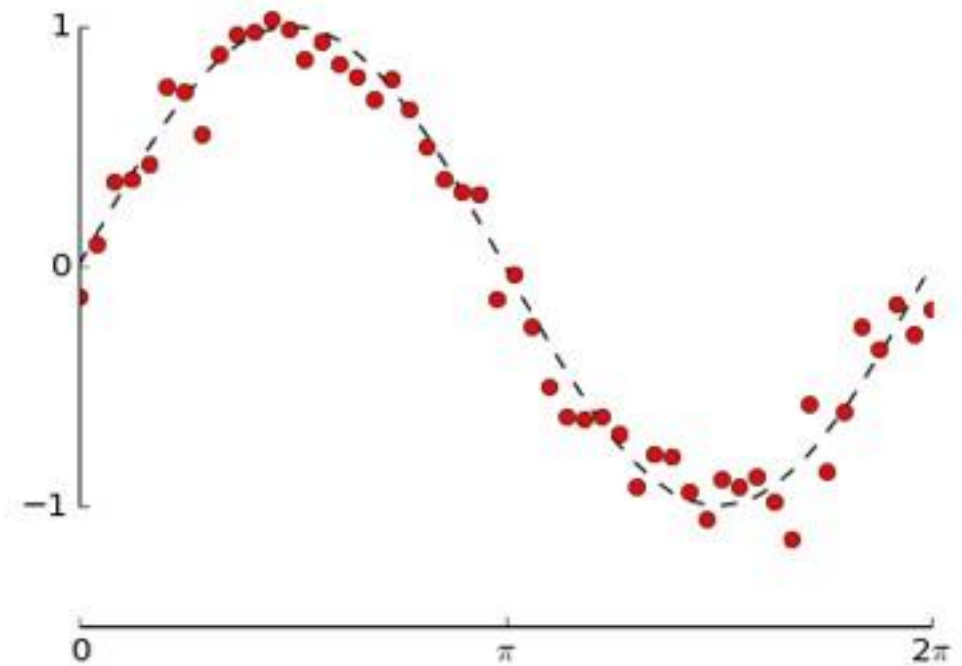
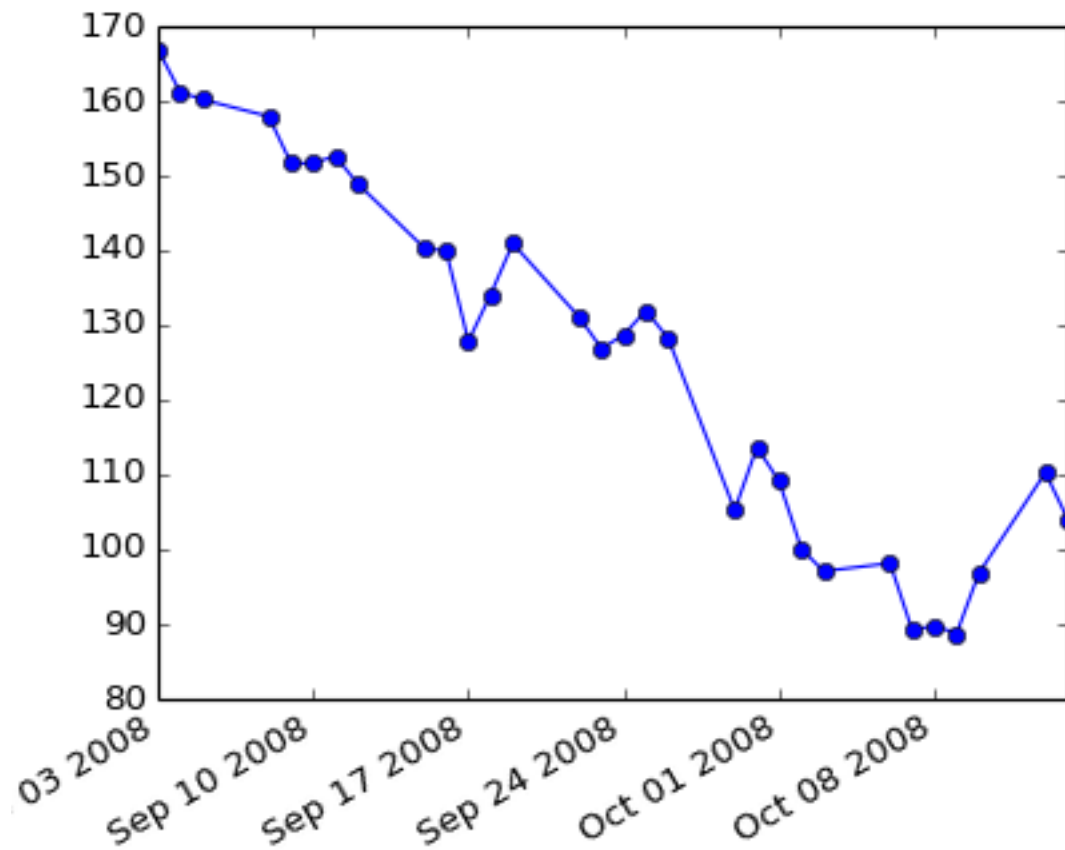
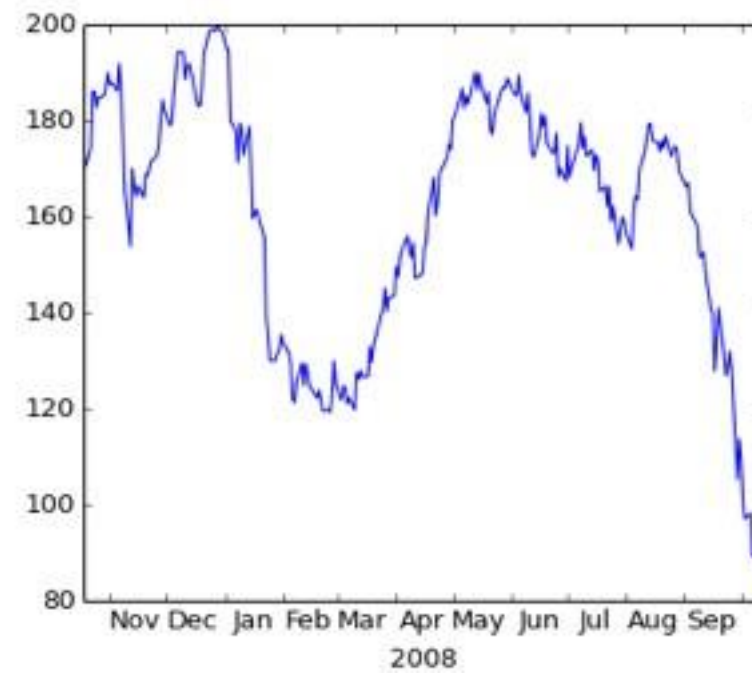
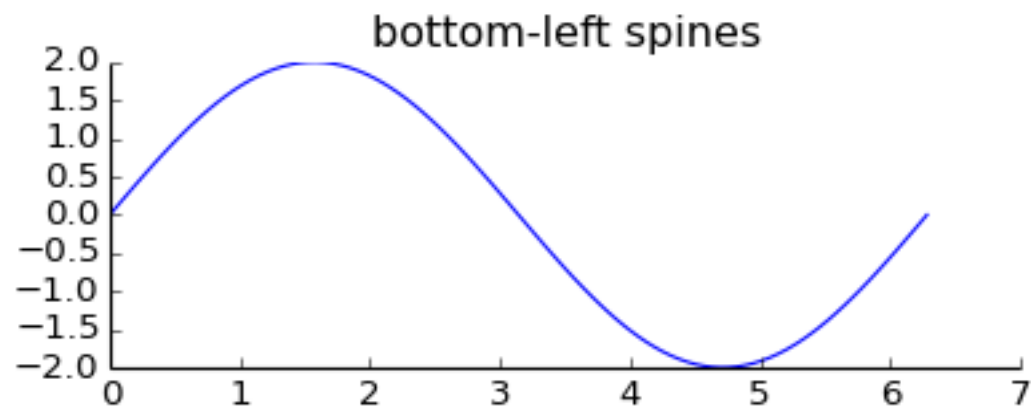
Nicolas Rapp

Trends Over Time

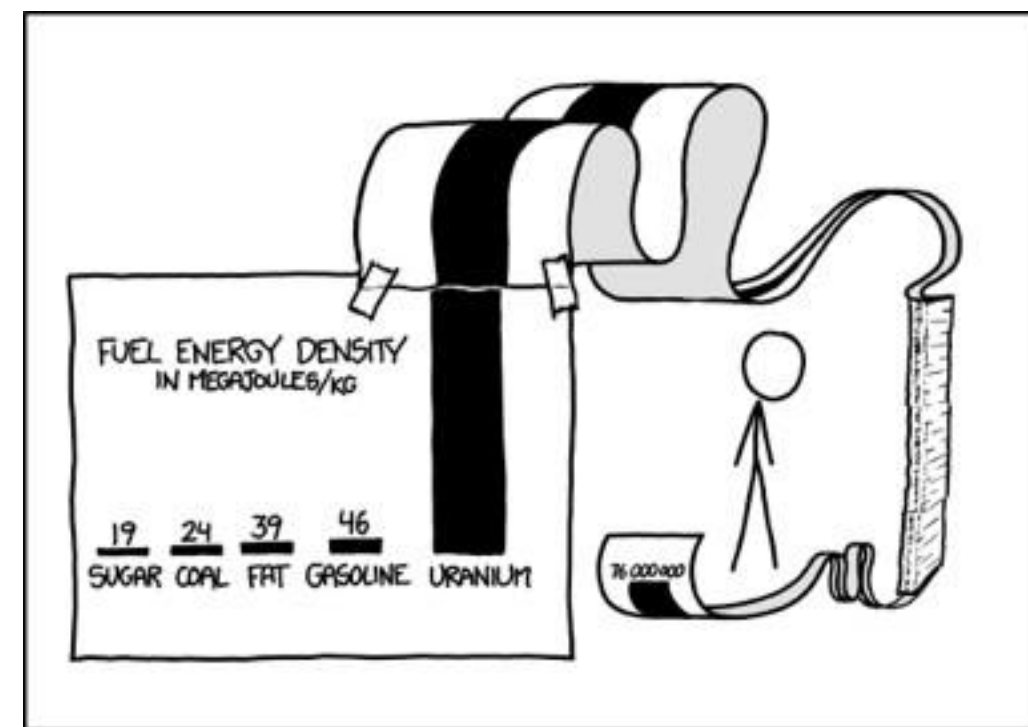
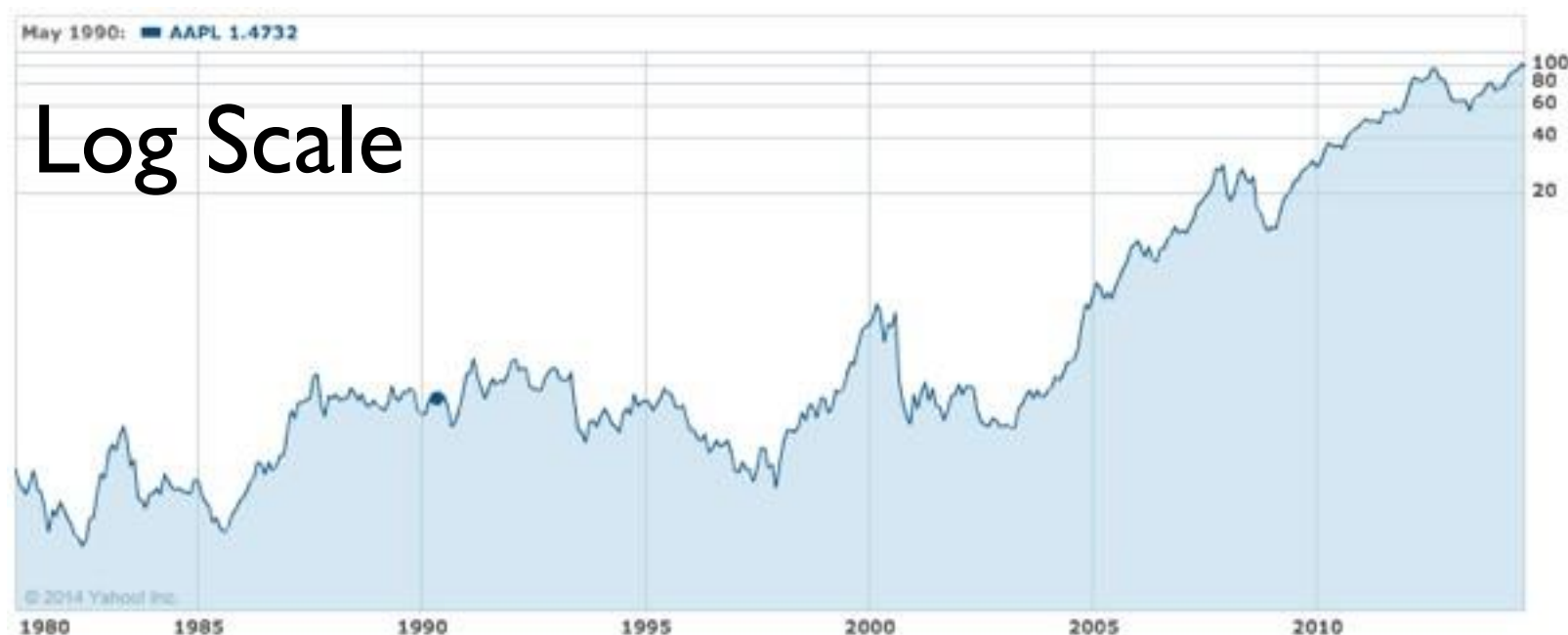
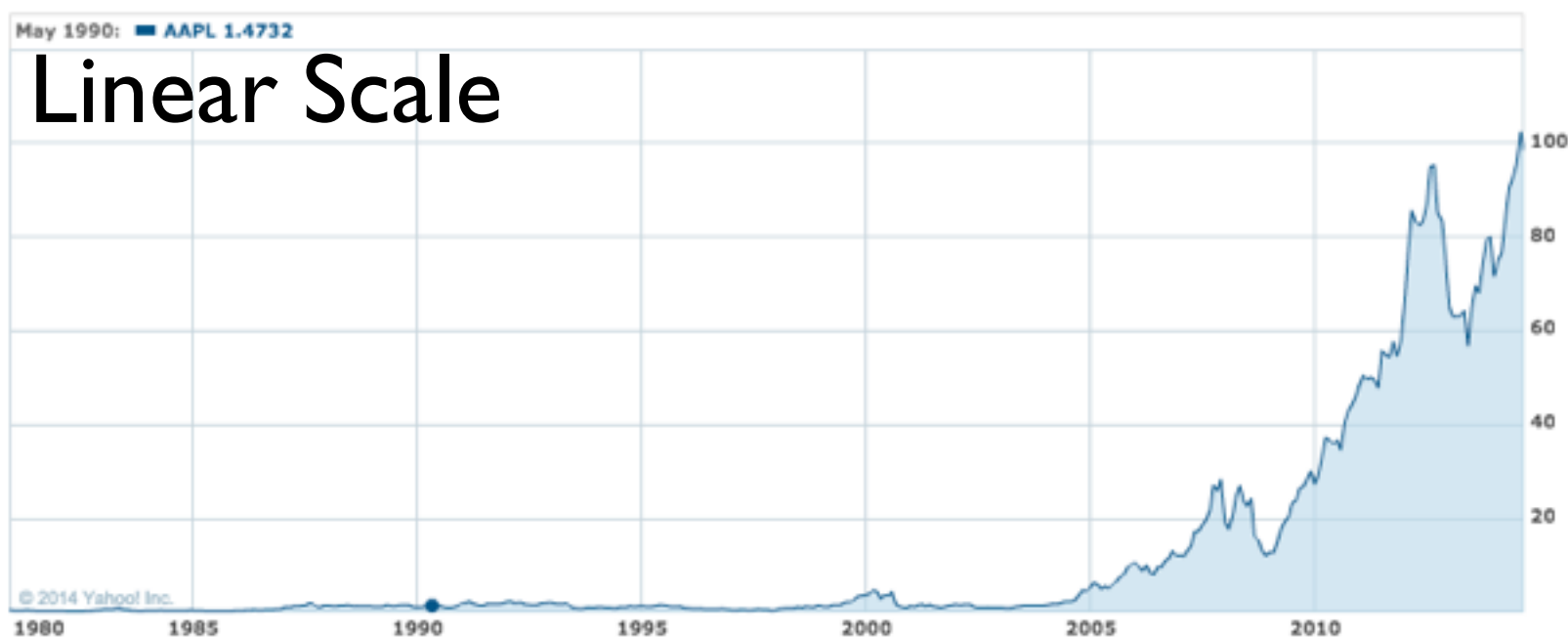
MY HOBBY: EXTRAPOLATING



Line Charts



Linear vs. Logarithmic Scale



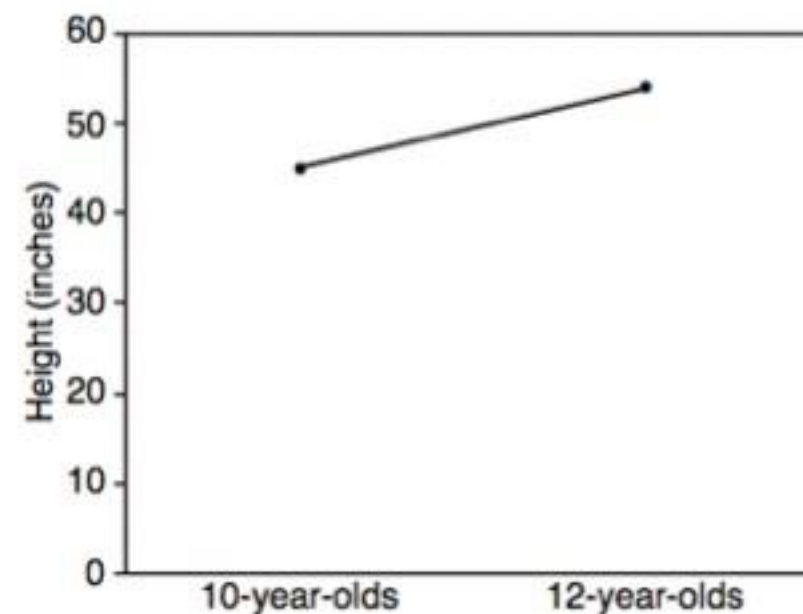
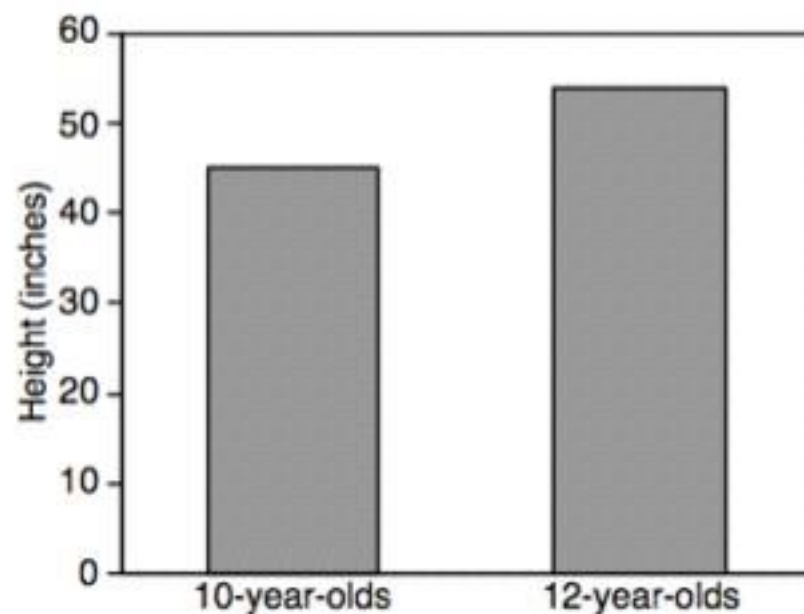
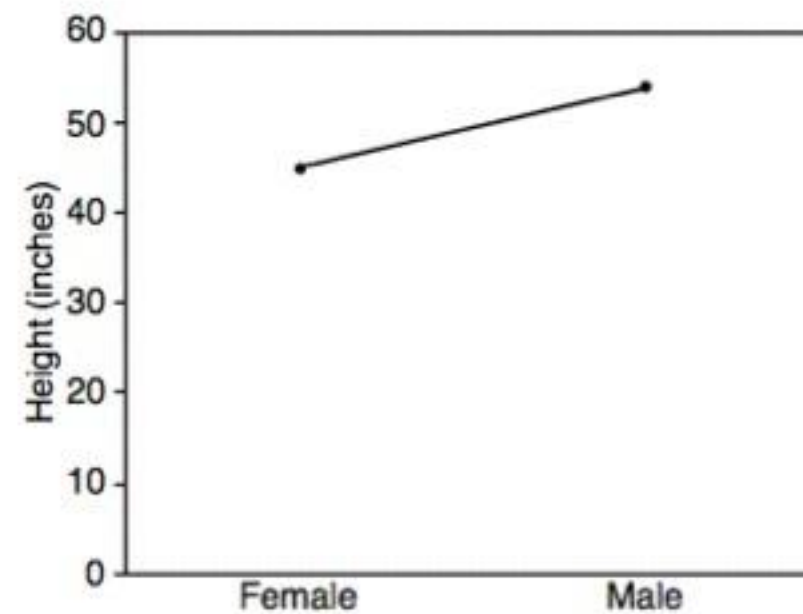
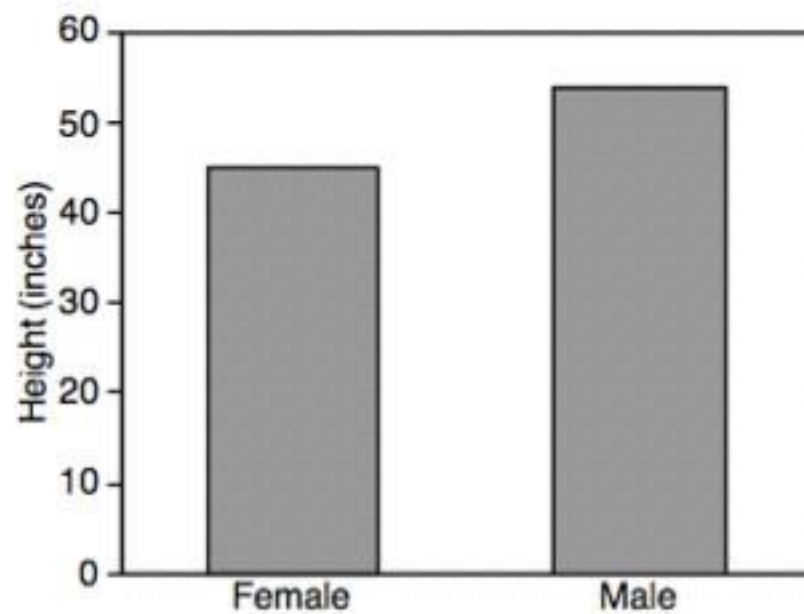
SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT *PROPERLY*.

<http://xkcd.com/1162/>

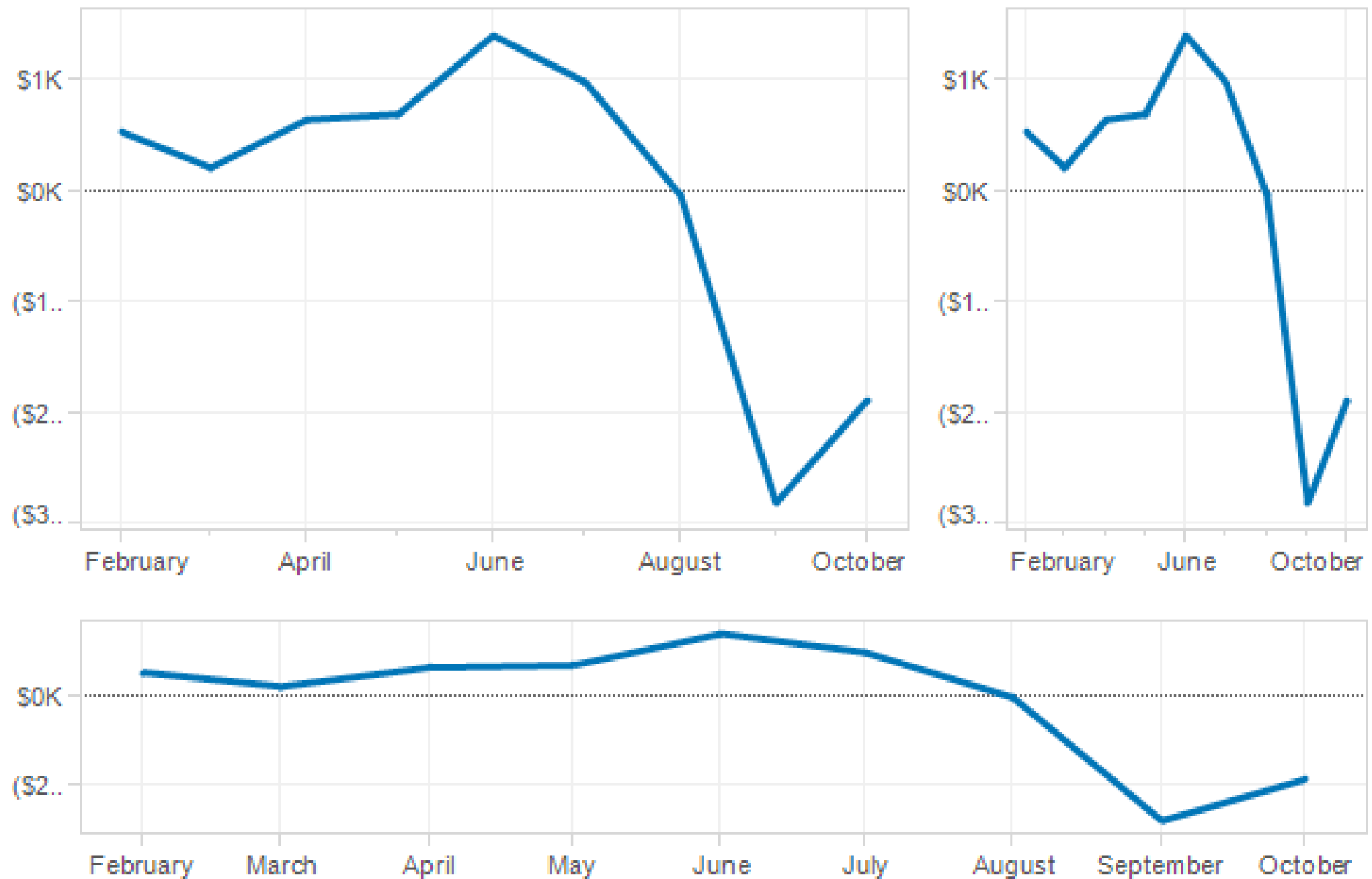
Apple Stock Price <http://finance.yahoo.com/echarts?s=AAPL>

Bars vs. Lines

Lines imply connections - do not use for categorical data

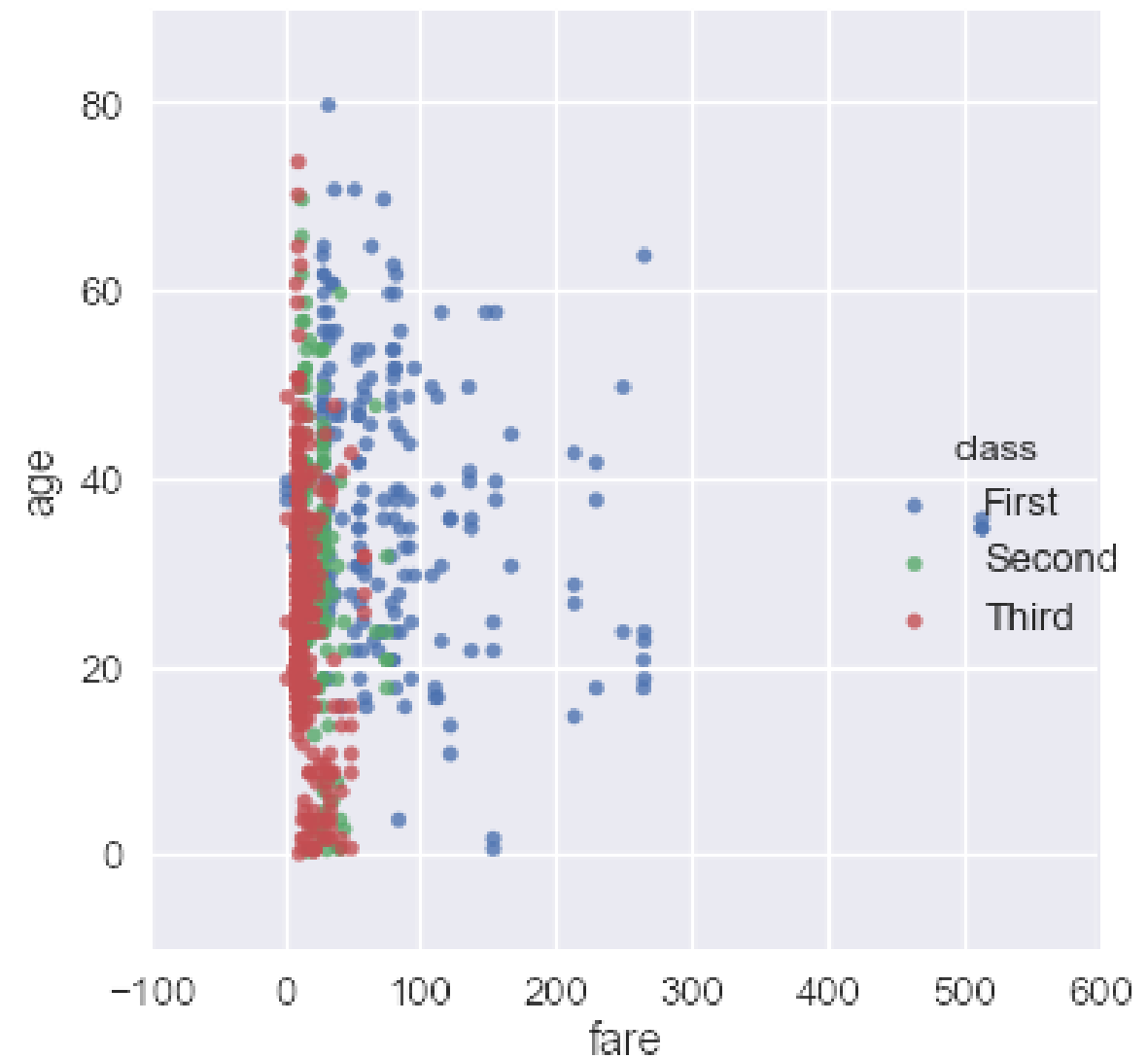
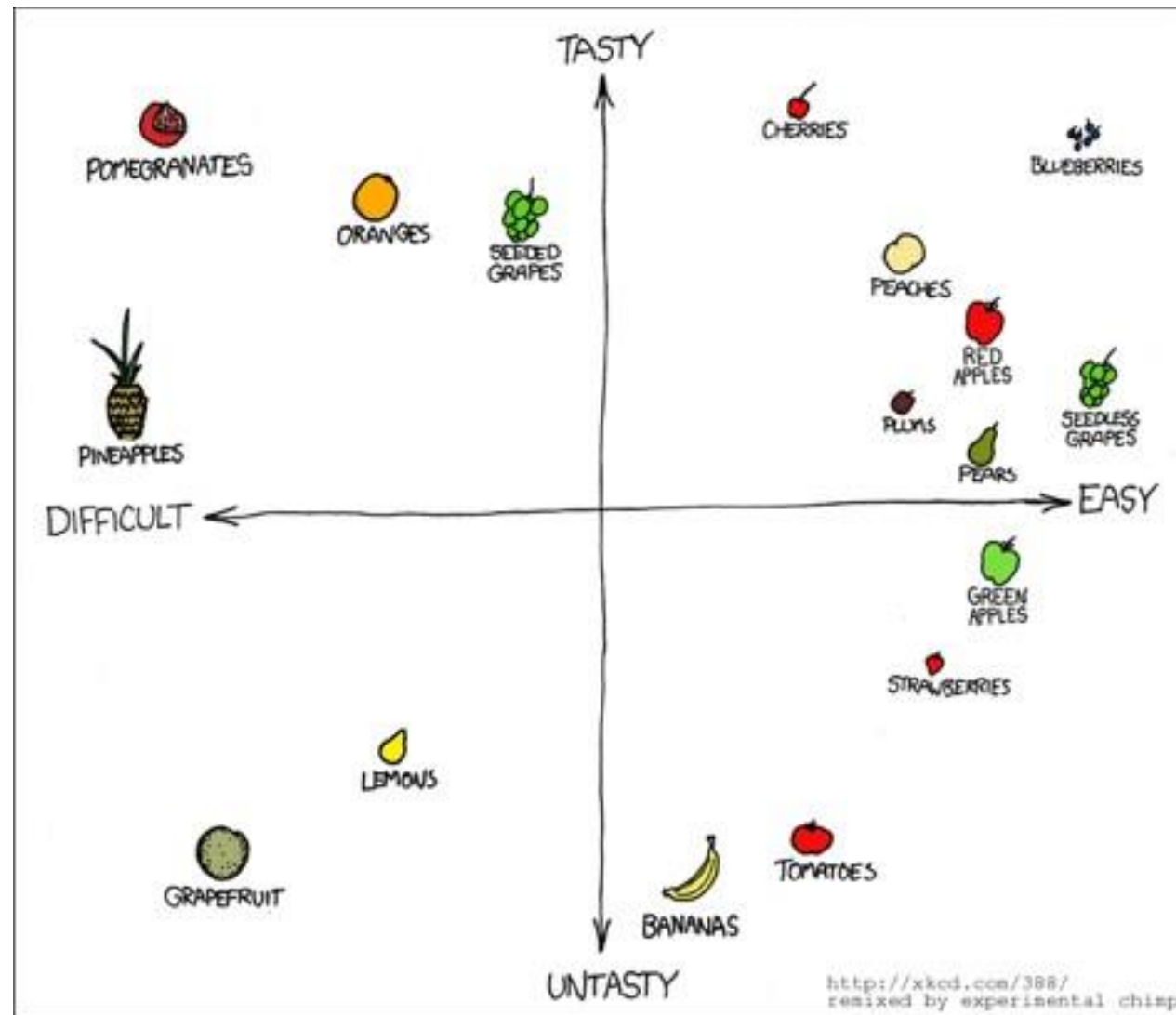


Aspect Ratios

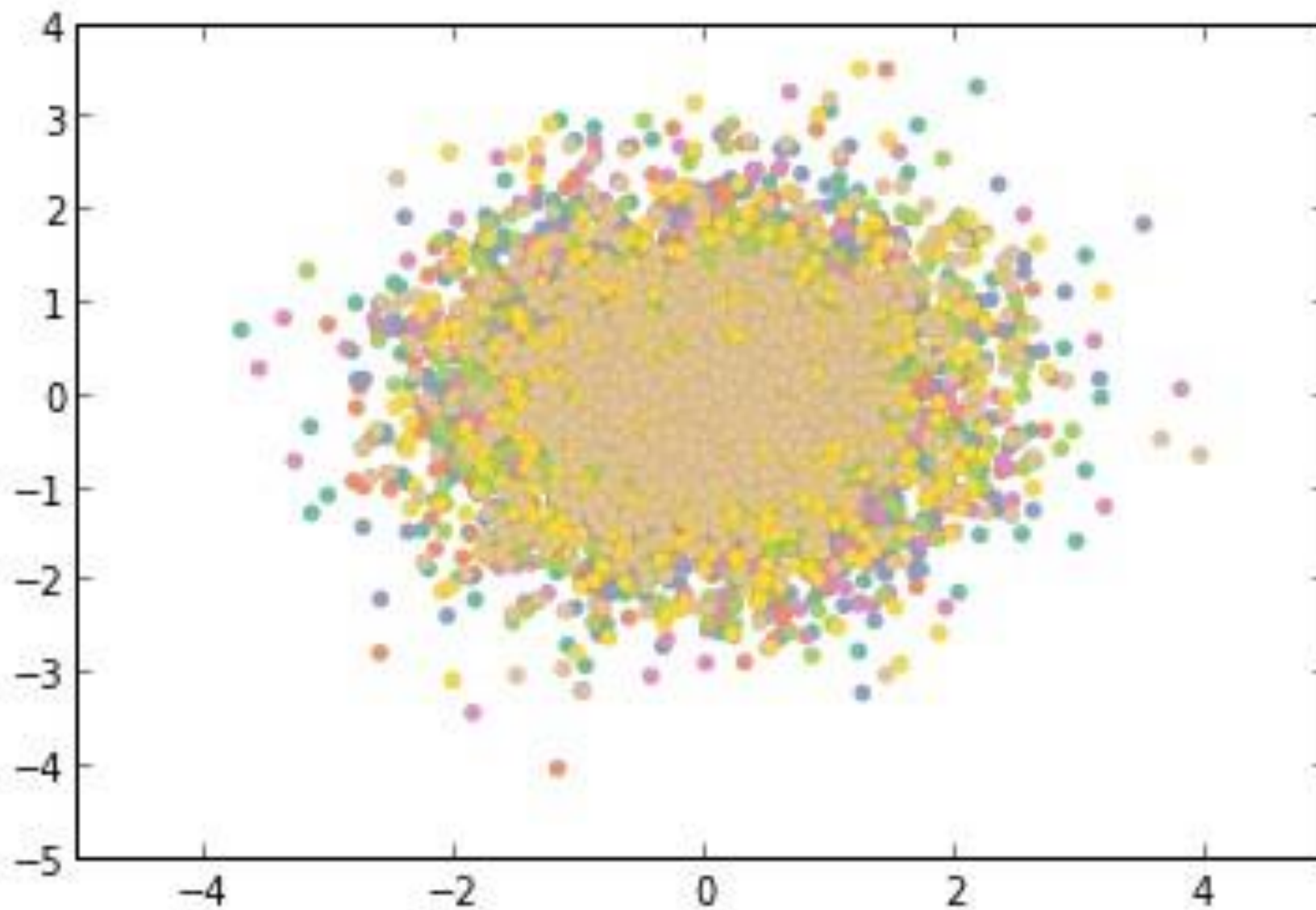


Correlations

Scatterplots

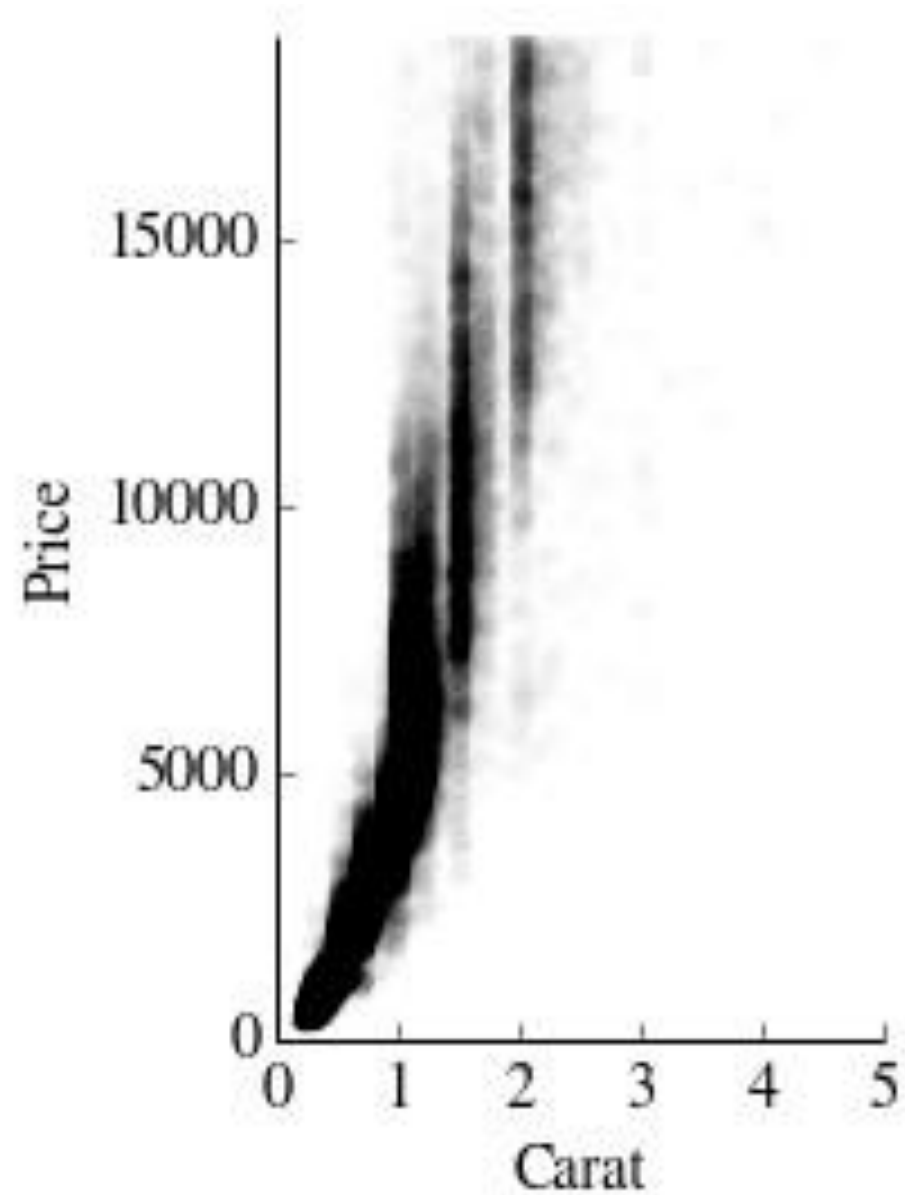
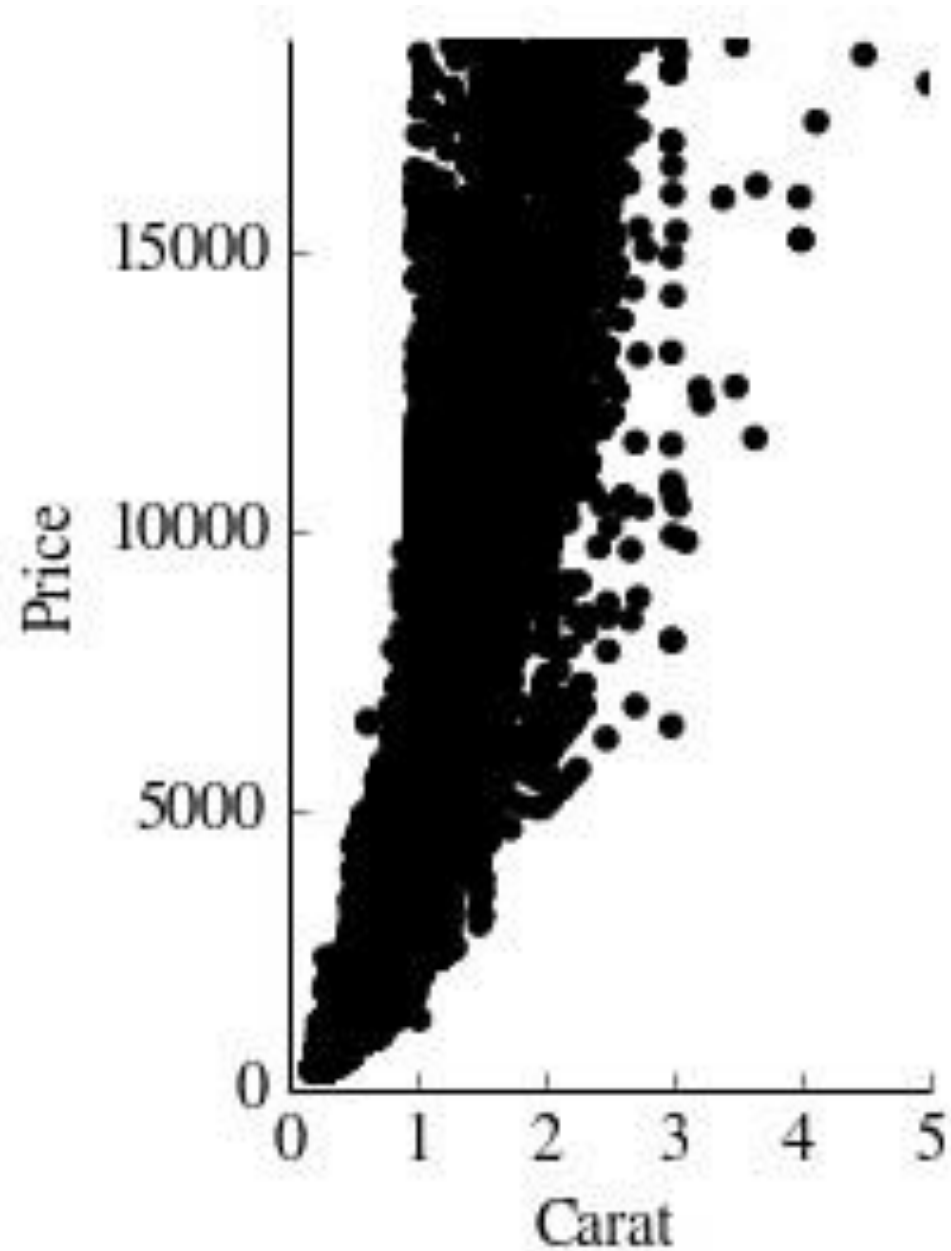


Scatterplots



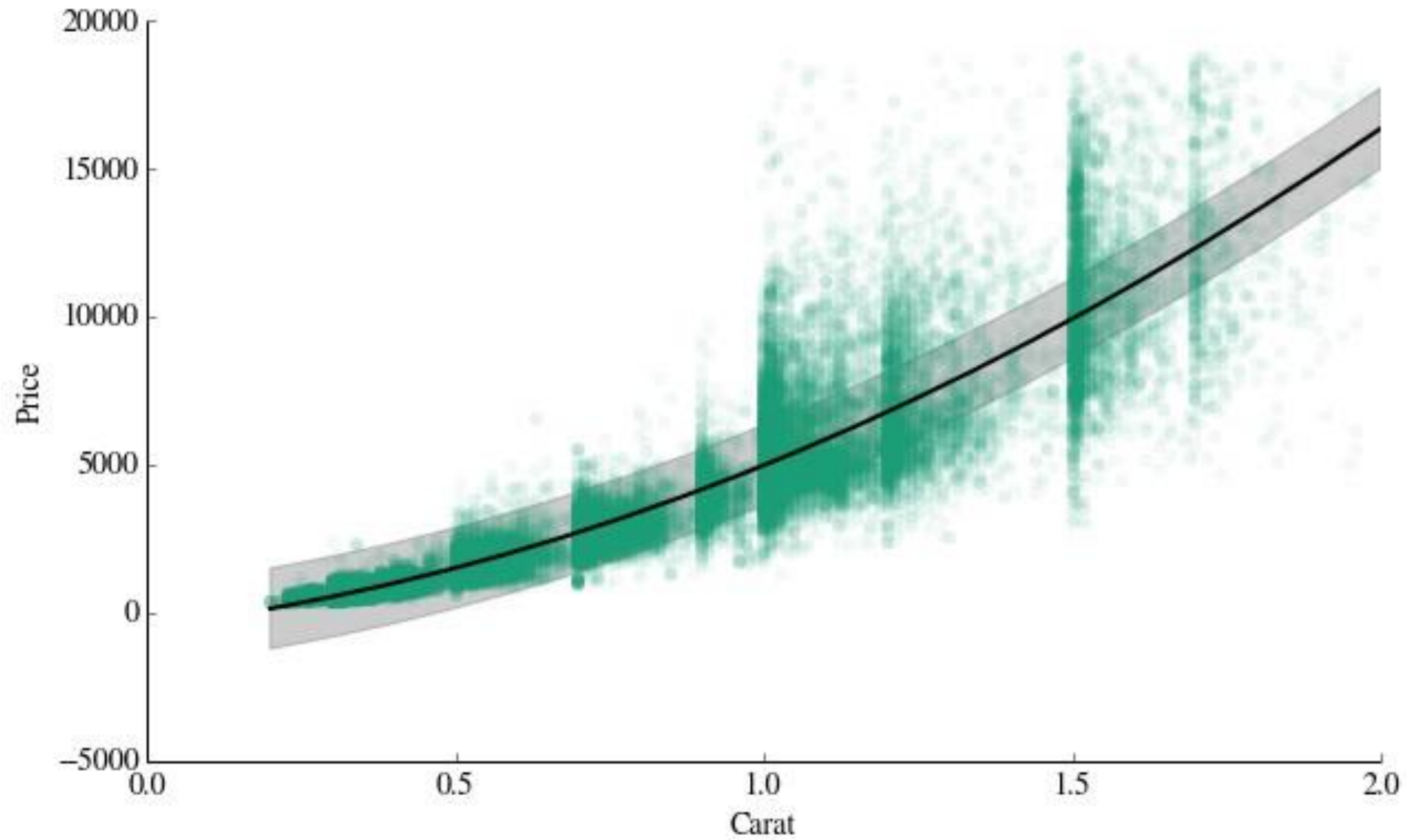
Light Grey Border

Overplotting

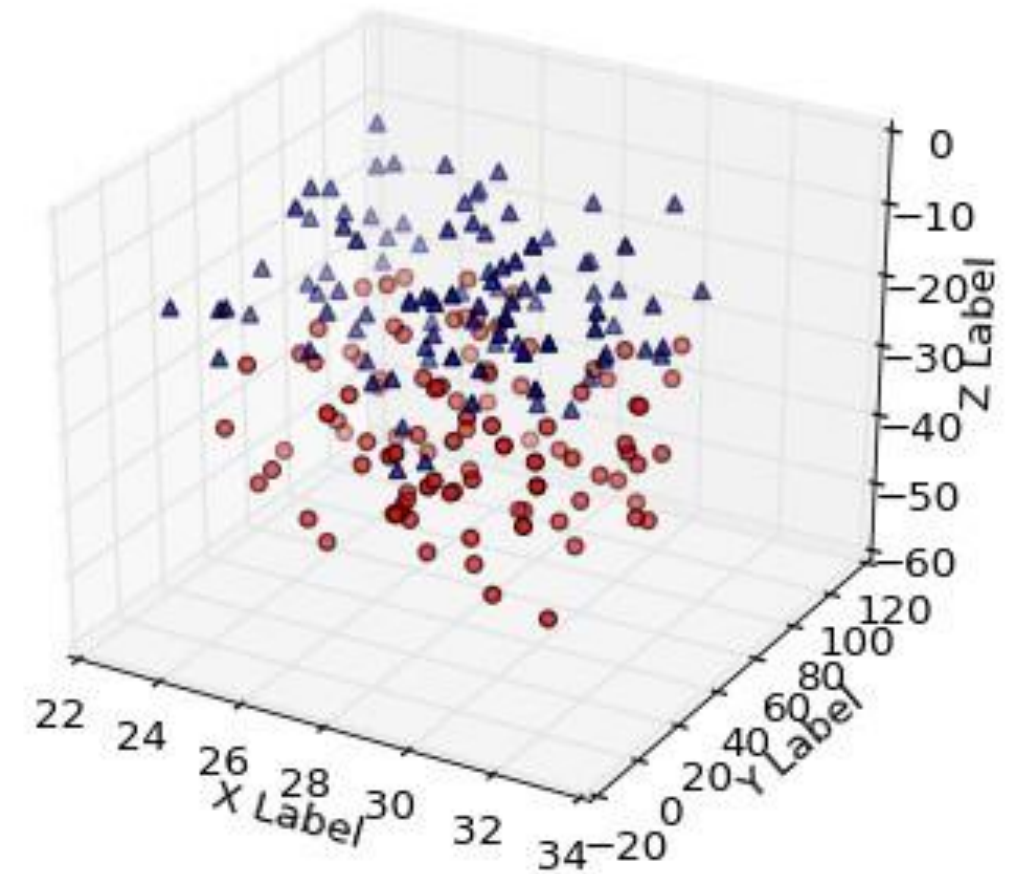
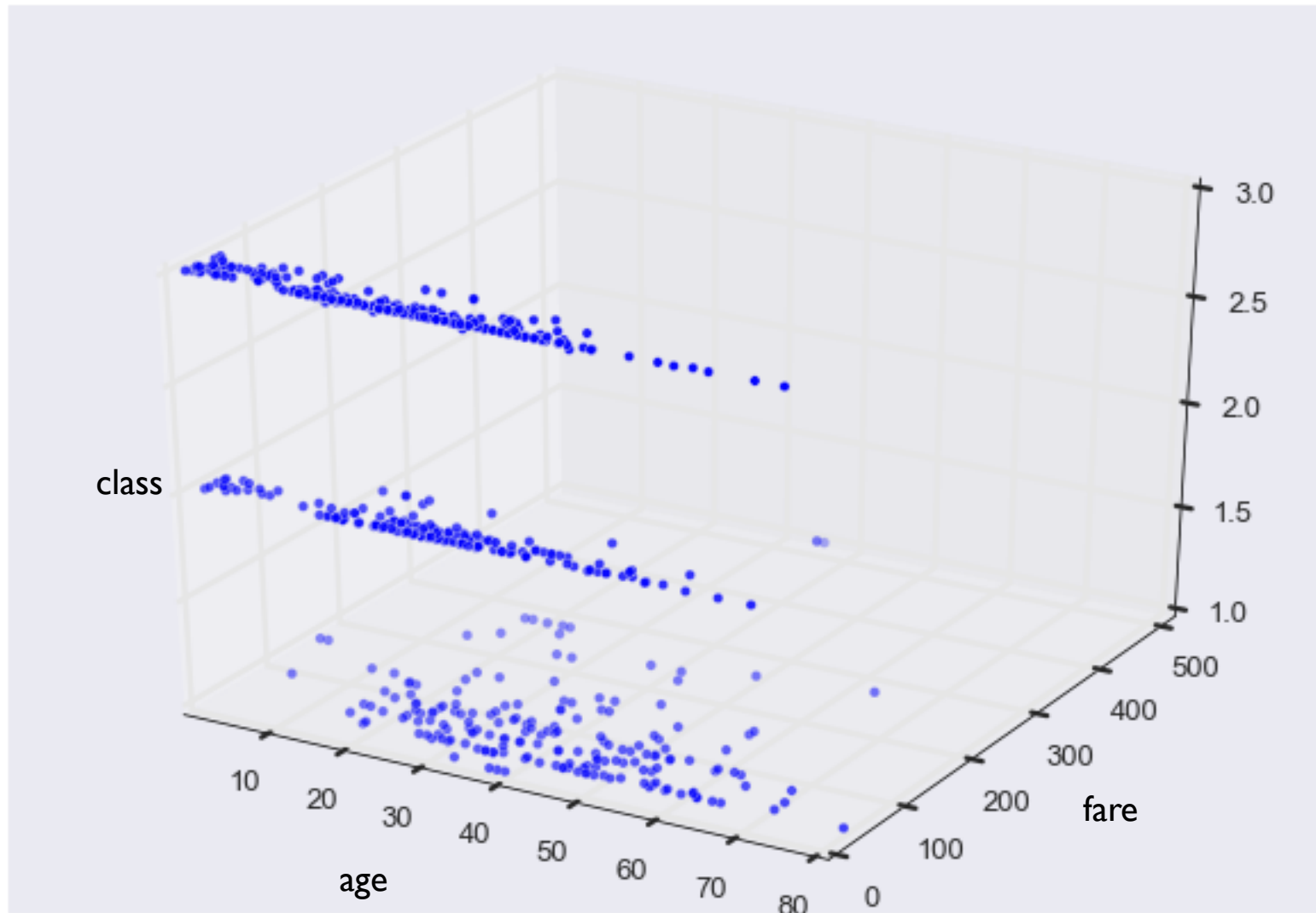


$\alpha = 1/100$

Trend Lines

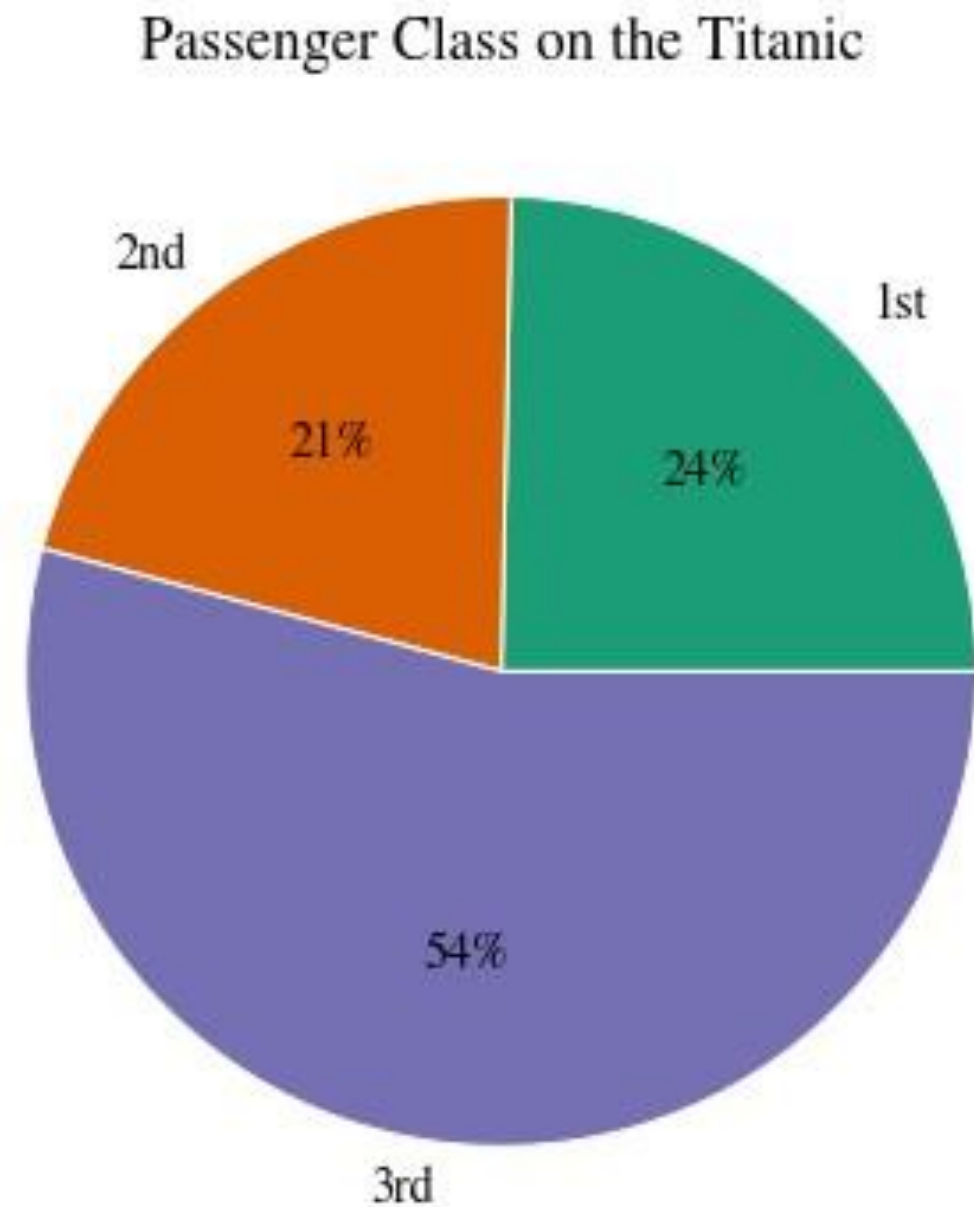
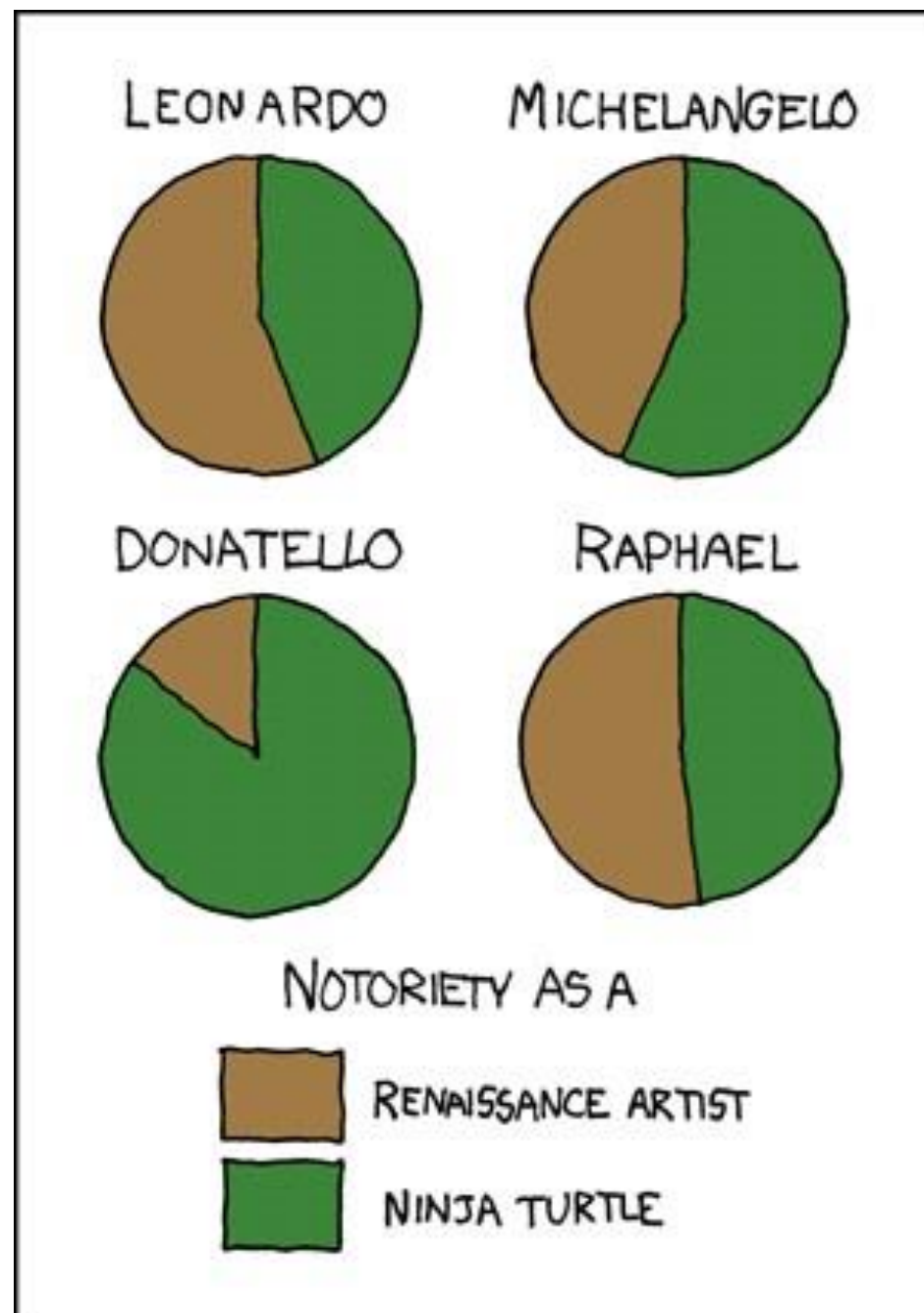


Don't



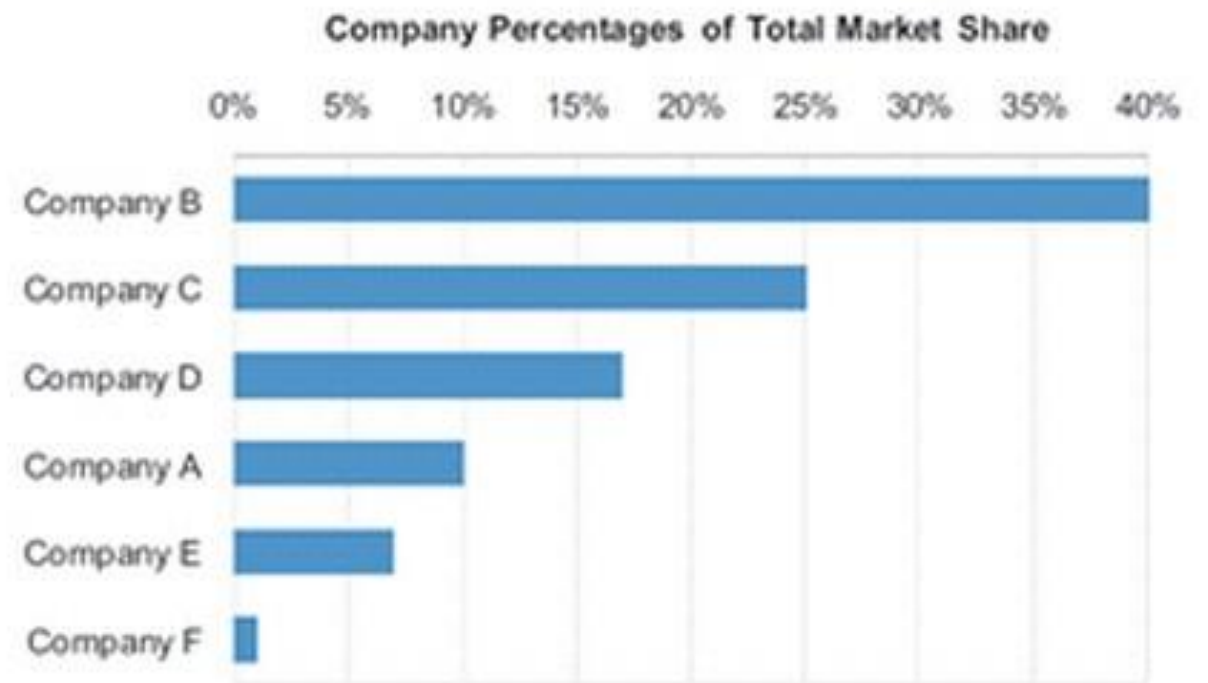
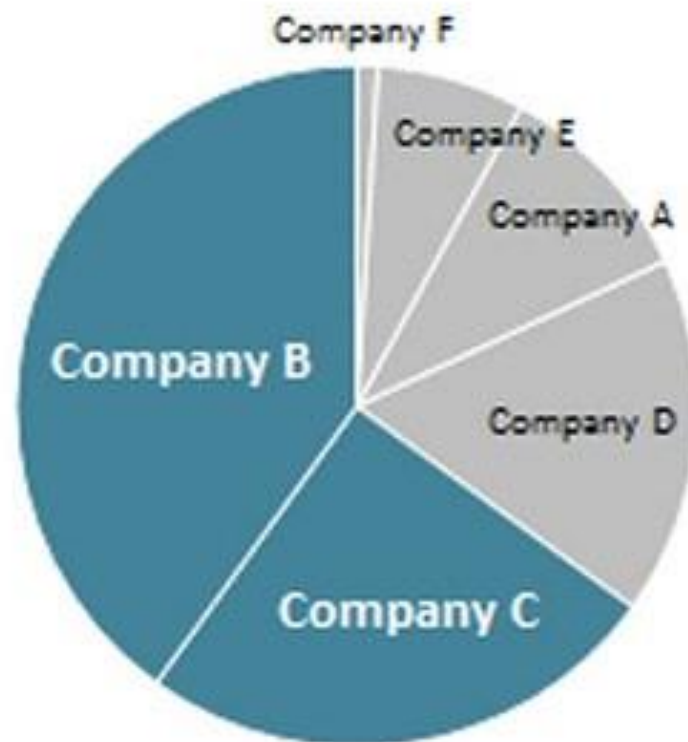
Compositions

Pie Charts

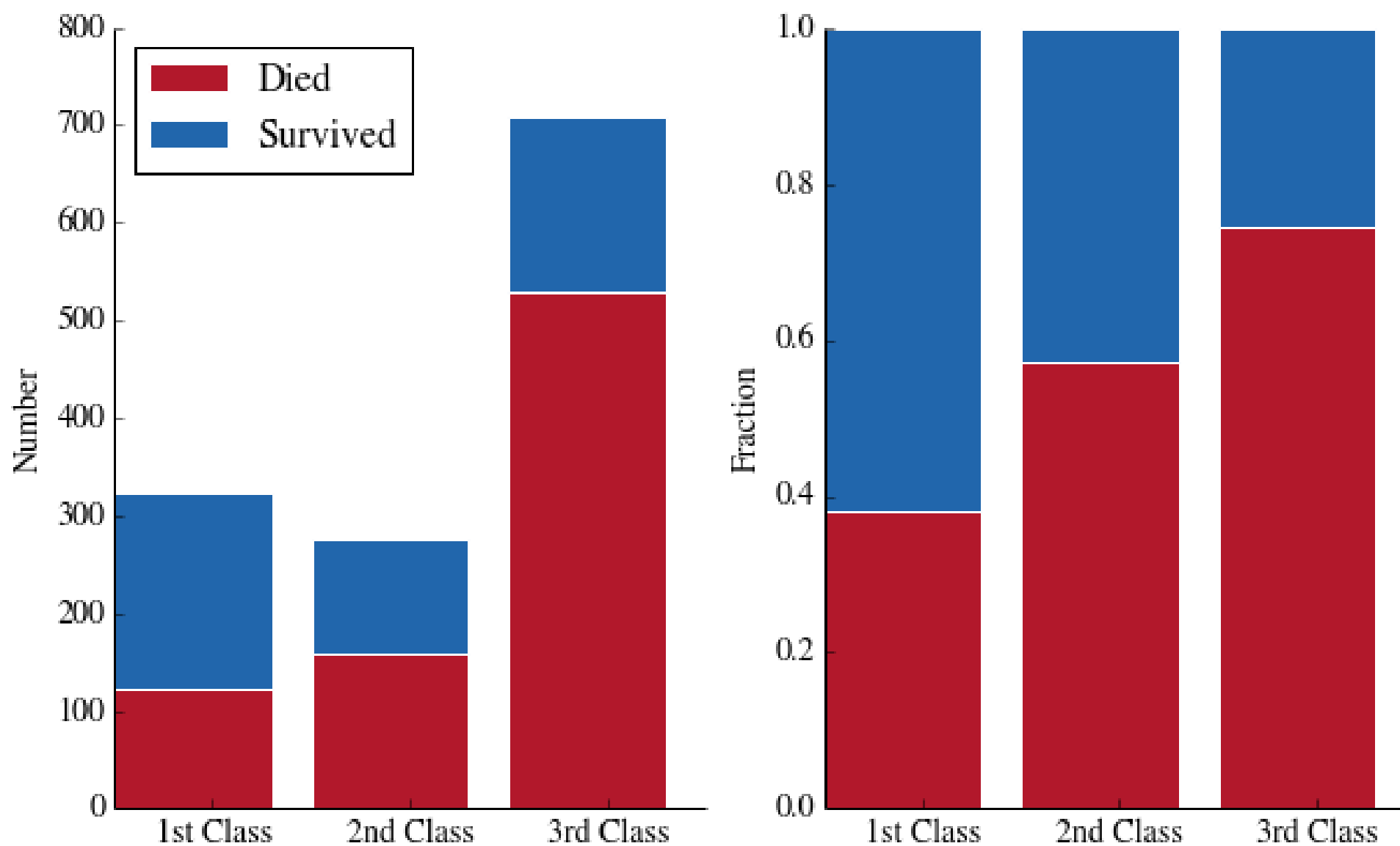


Pie vs. Bar Charts

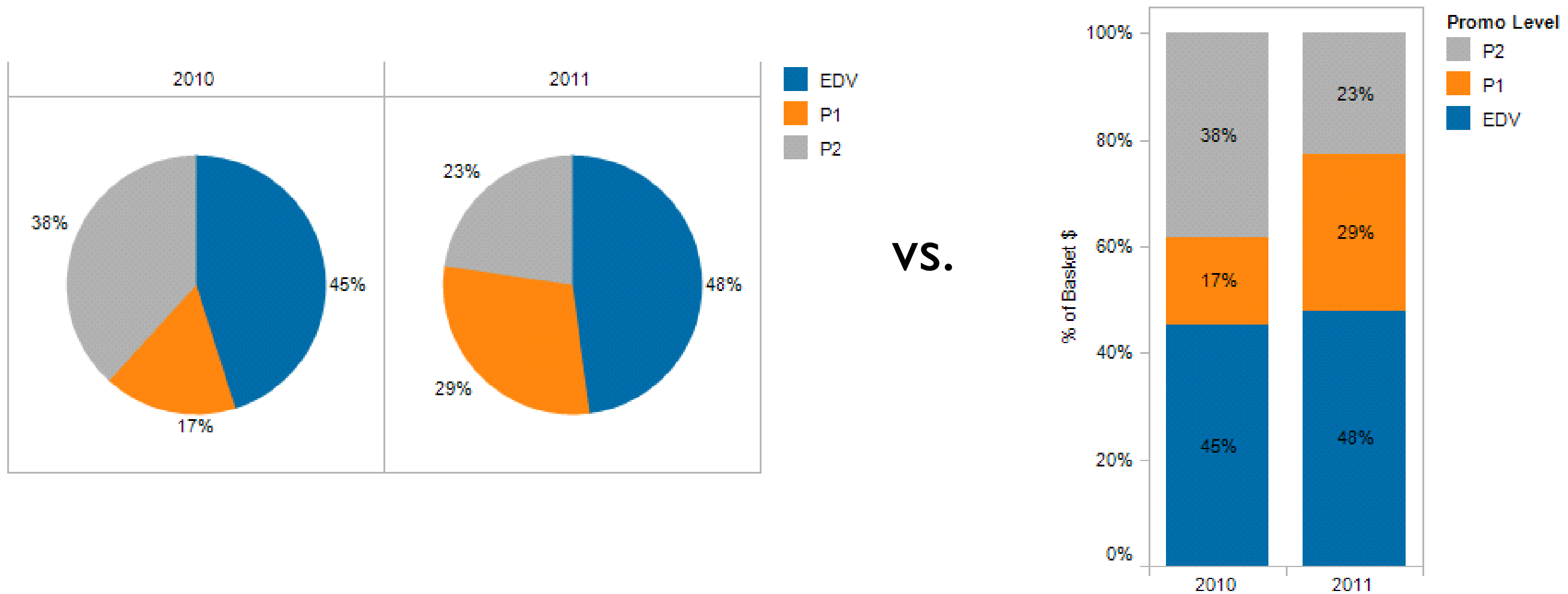
65% of the market is controlled by companies B and C



Stacked Bar Chart

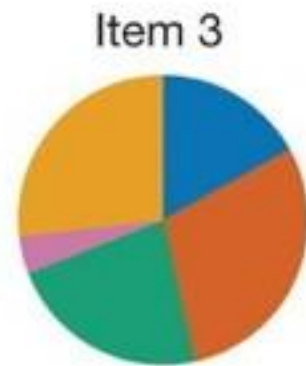


Stacked Bar Chart

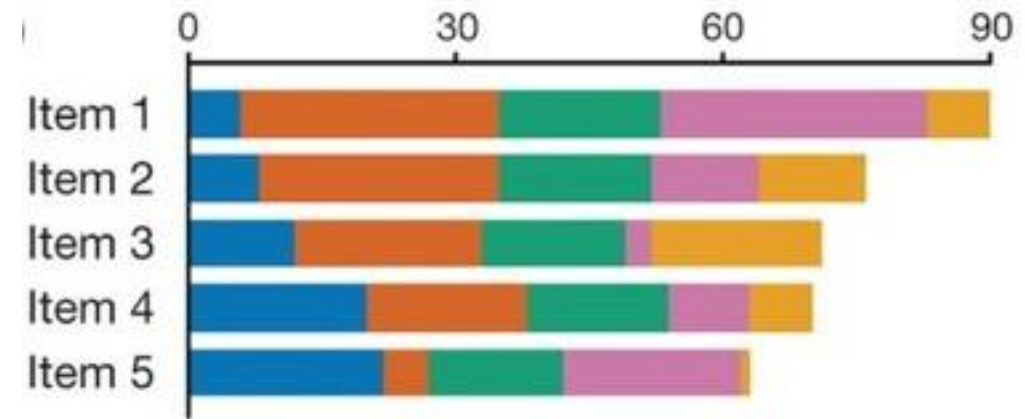


Comparison of bar chart types

Category 1 ●
Category 2 ●
Category 3 ●
Category 4 ●
Category 5 ●

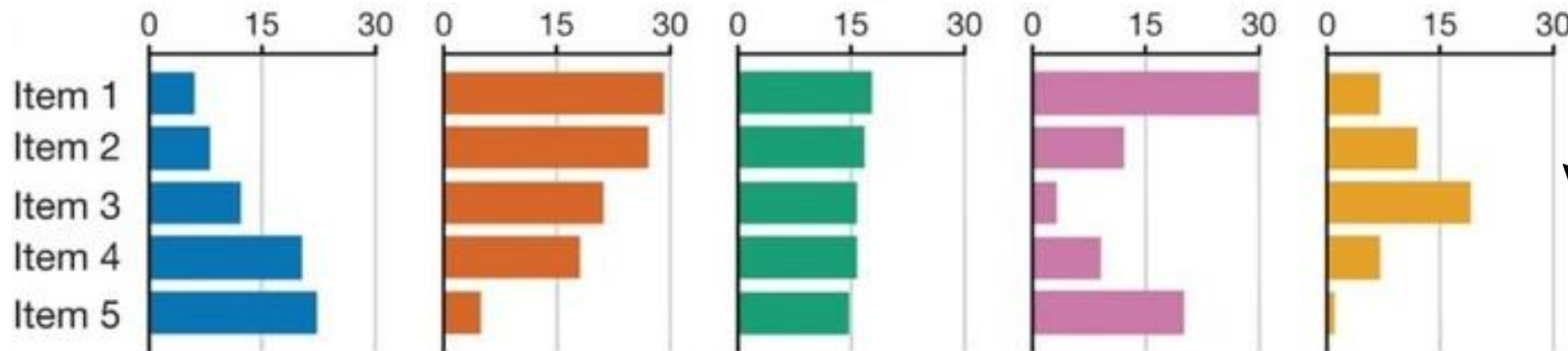


Pie Chart

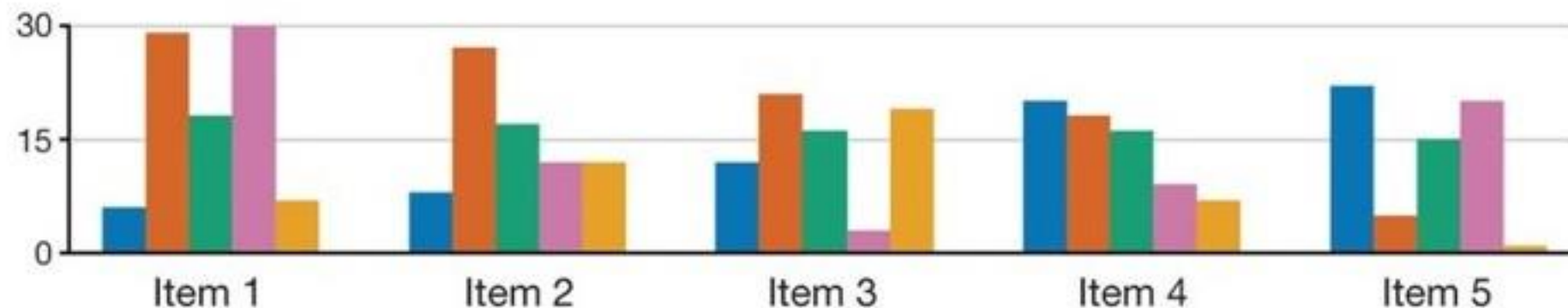


Stacked bar chart

Layered
Bar
Chart

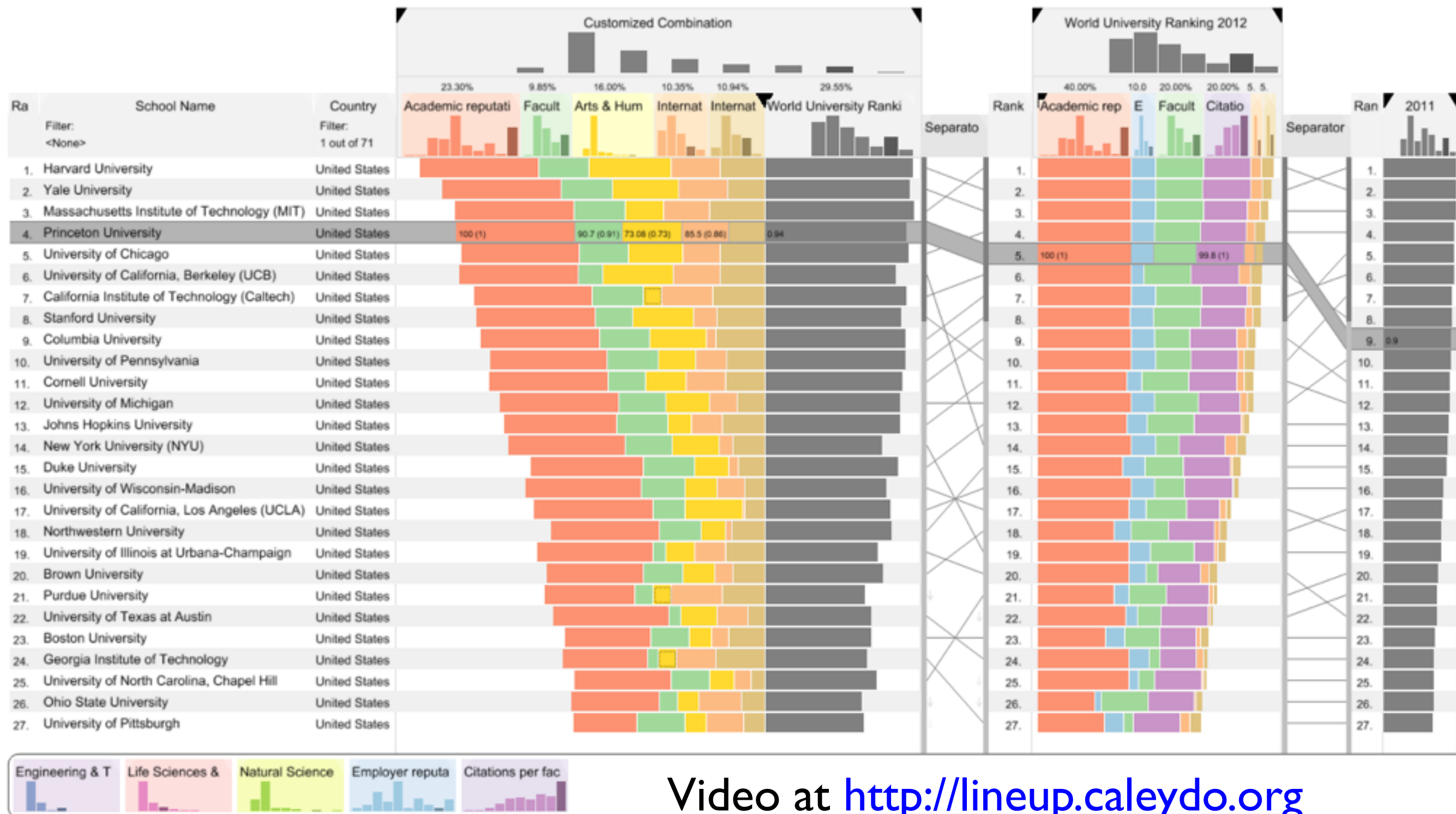


Grouped
Bar
Chart



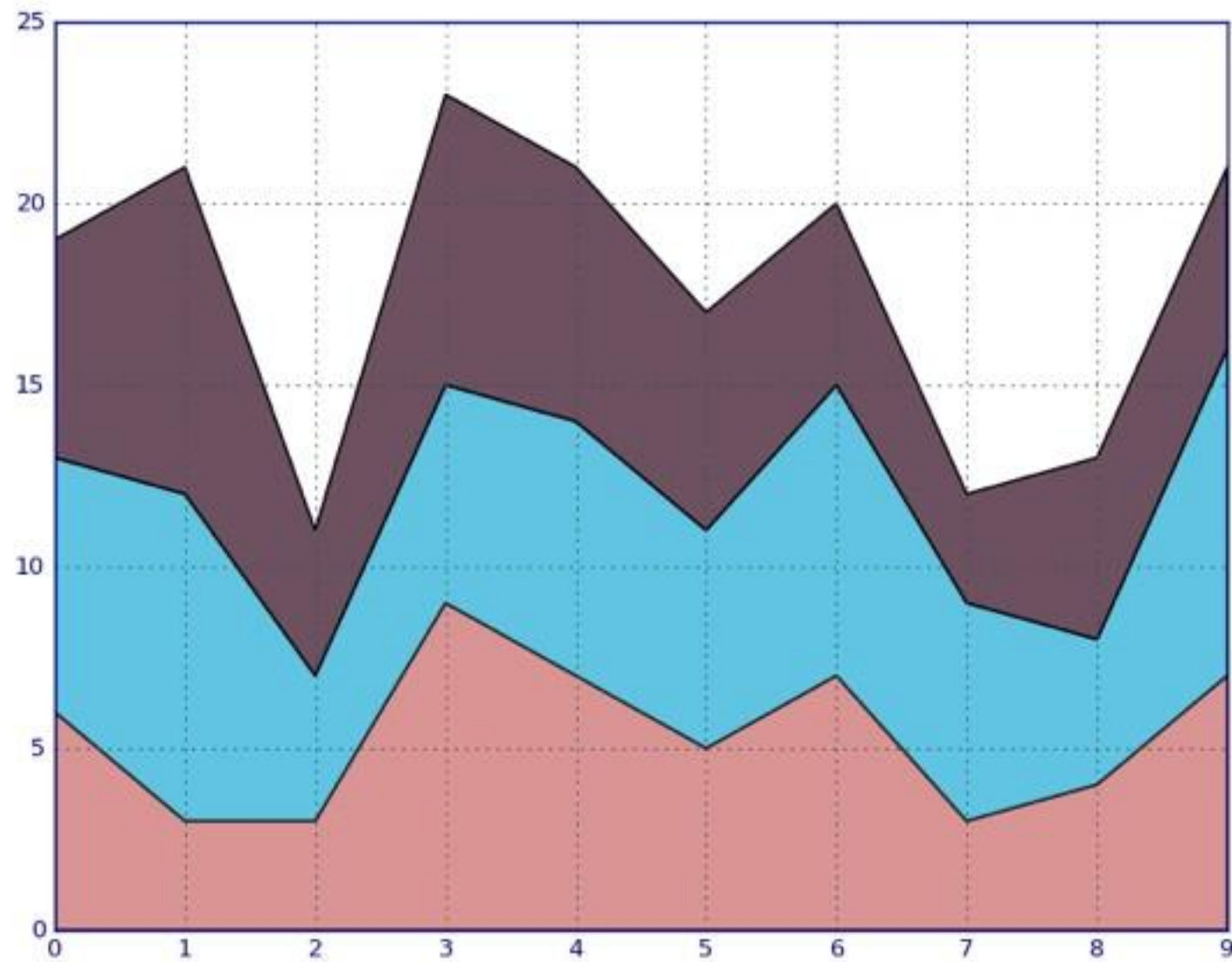
Small
Multiples

LineUp

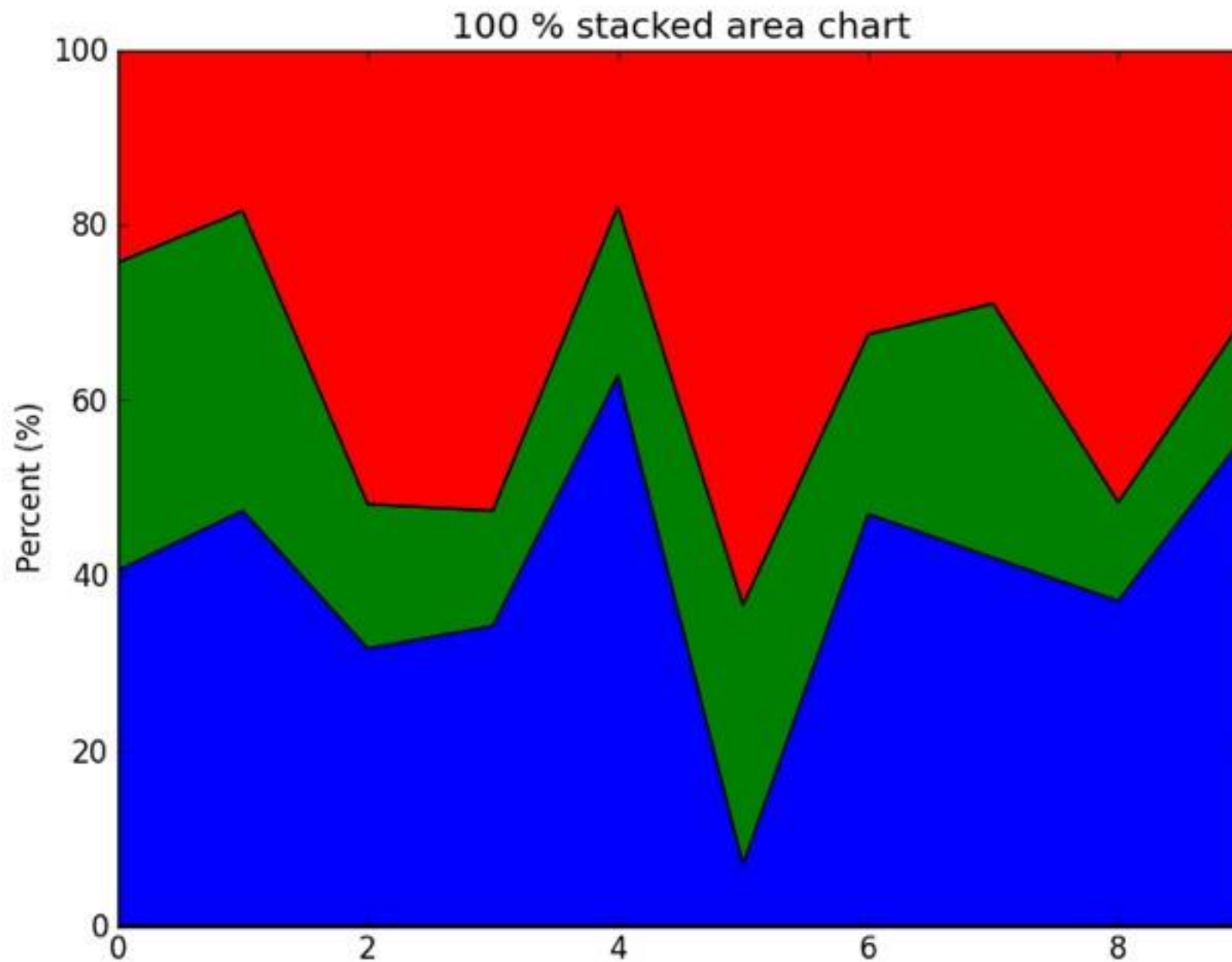


Video at <http://lineup.caleydo.org>

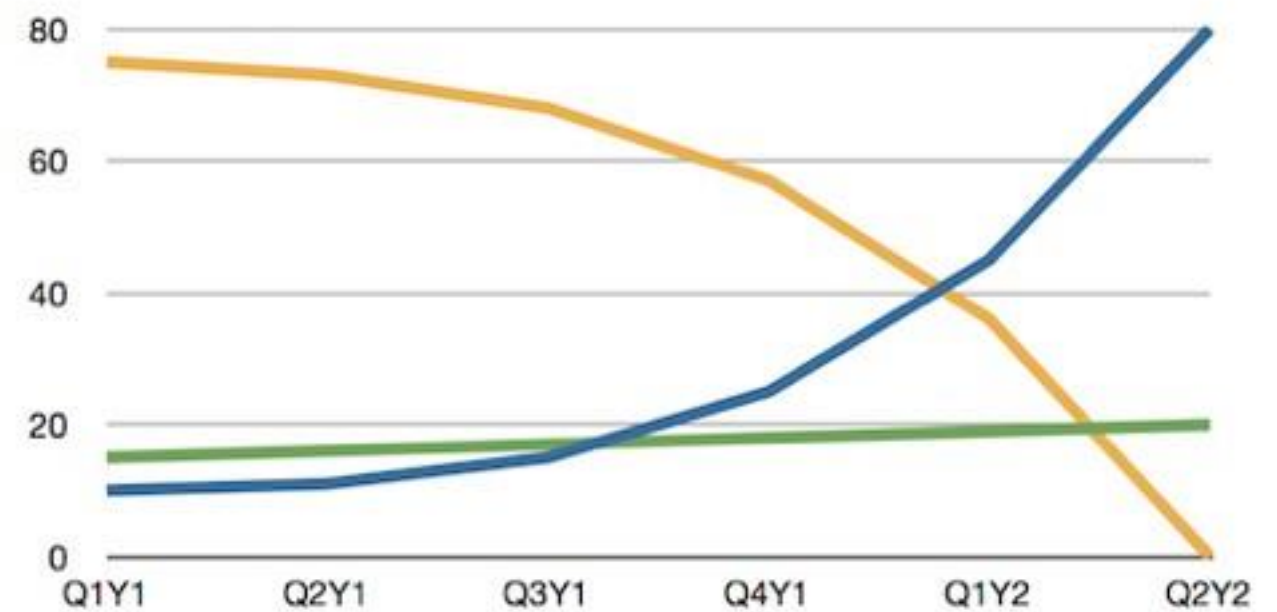
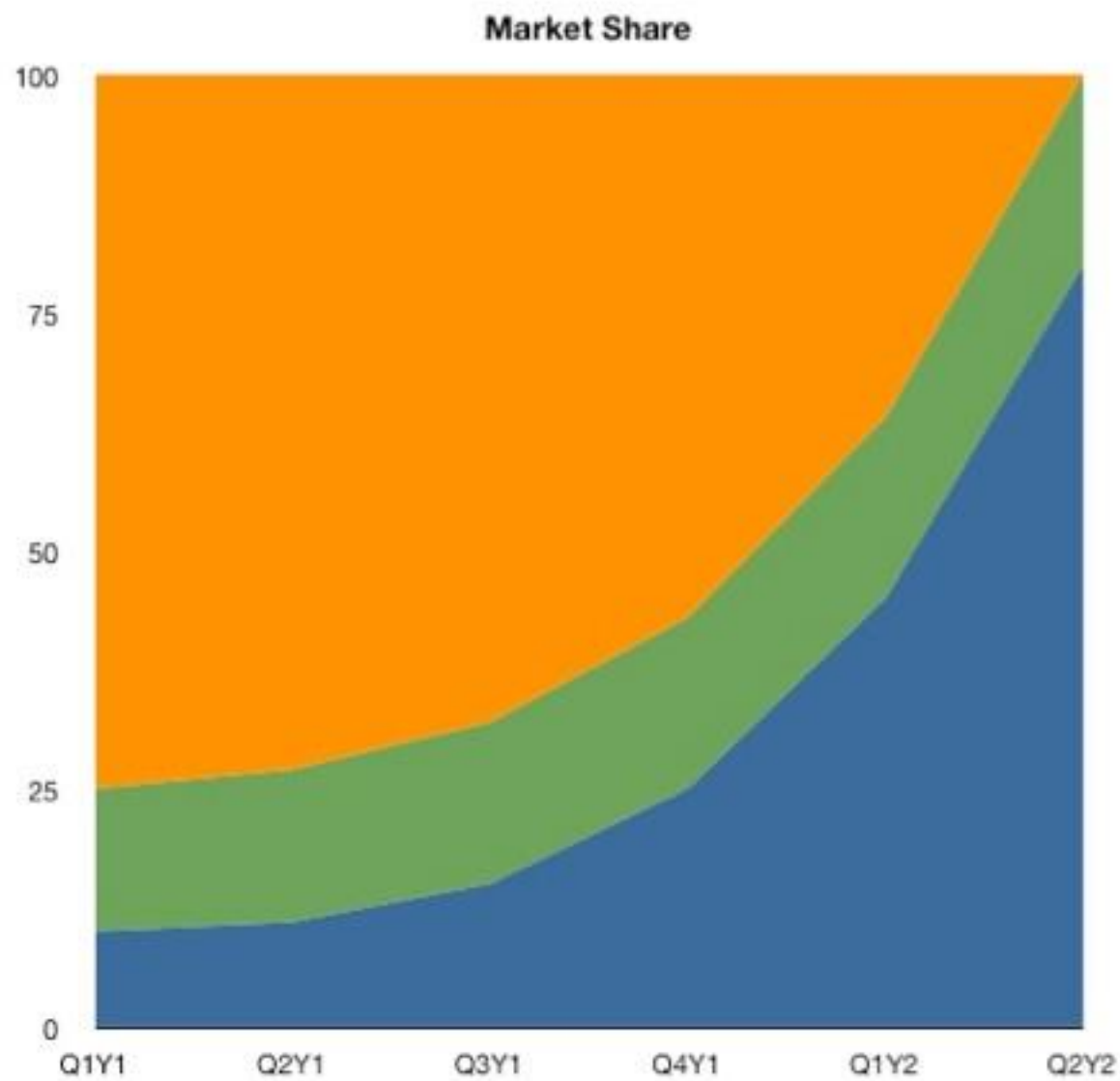
Stacked Area Chart



100% Stacked Area Chart



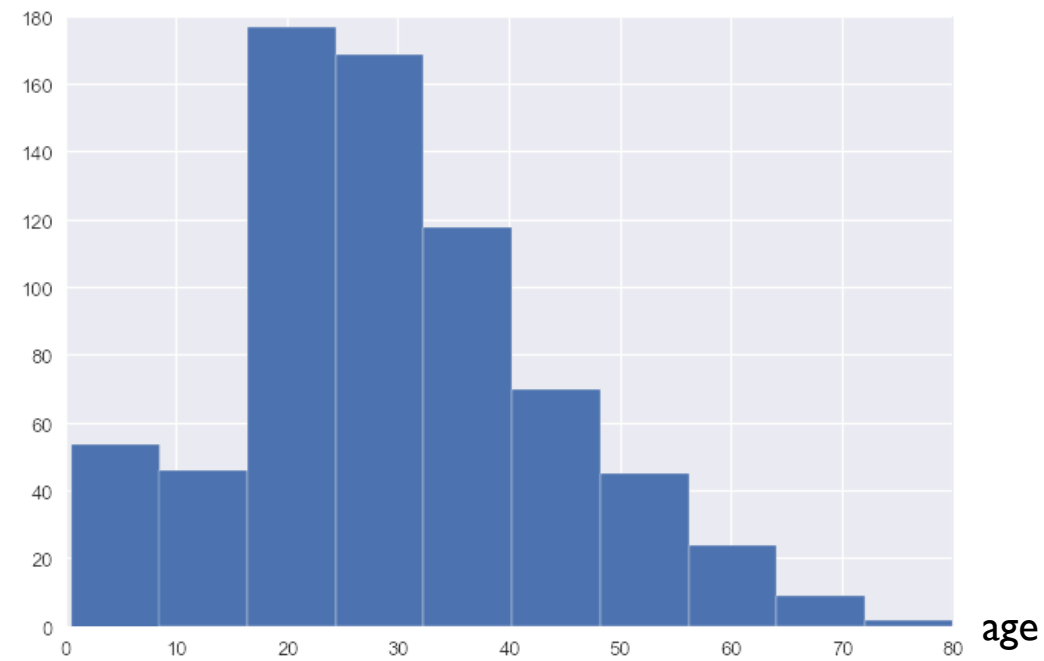
Stacked Area vs. Line Graphs



Distributions

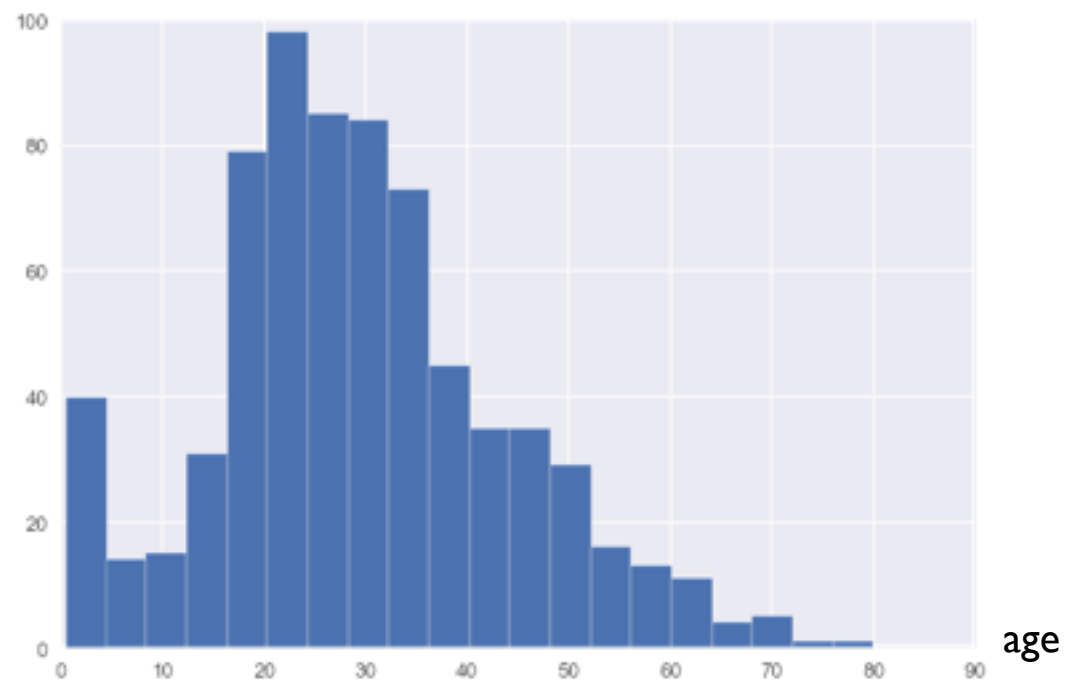
Histogram

passengers



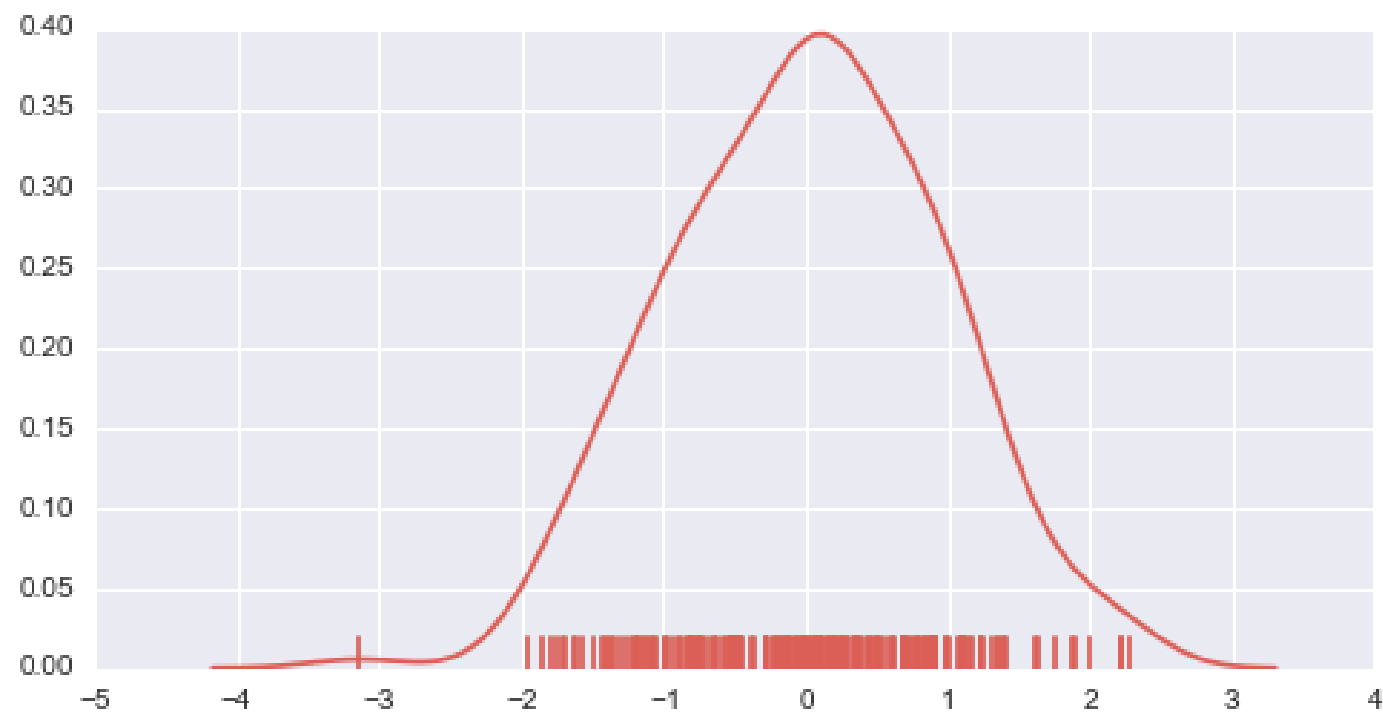
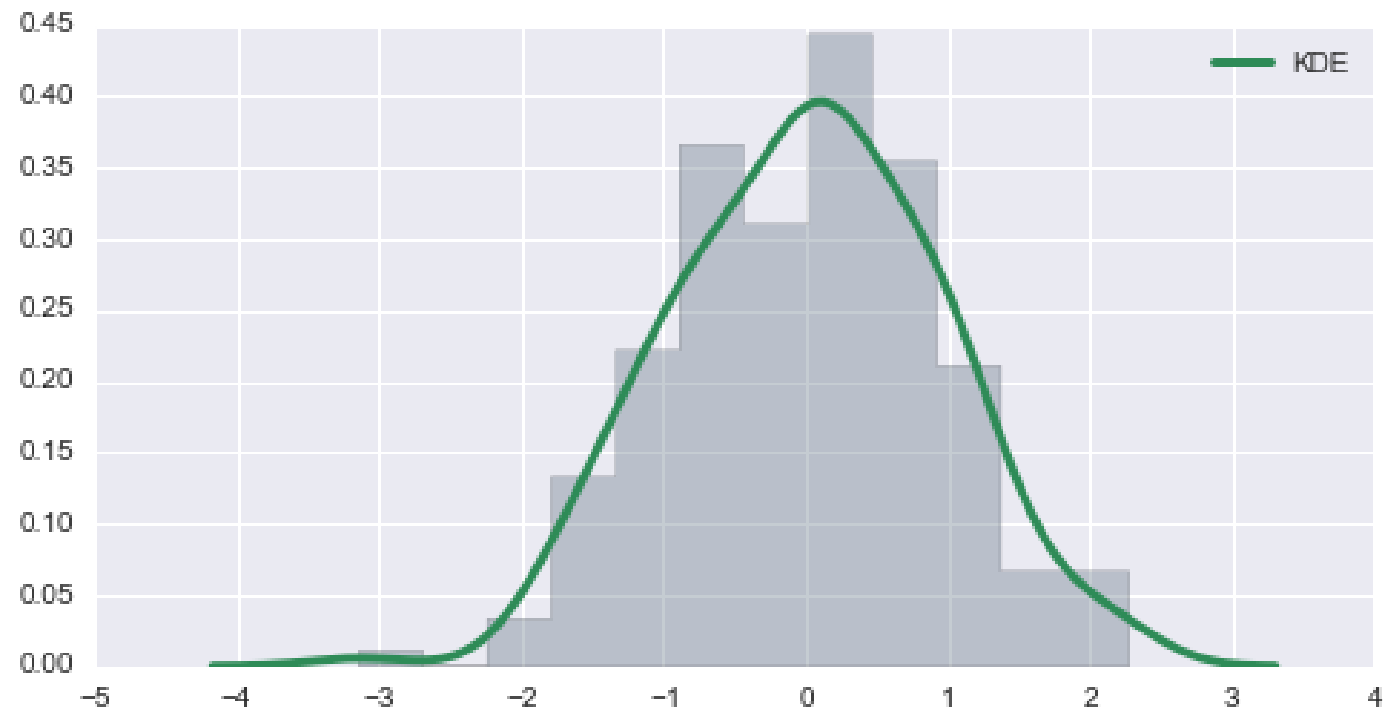
10 Bins

passengers

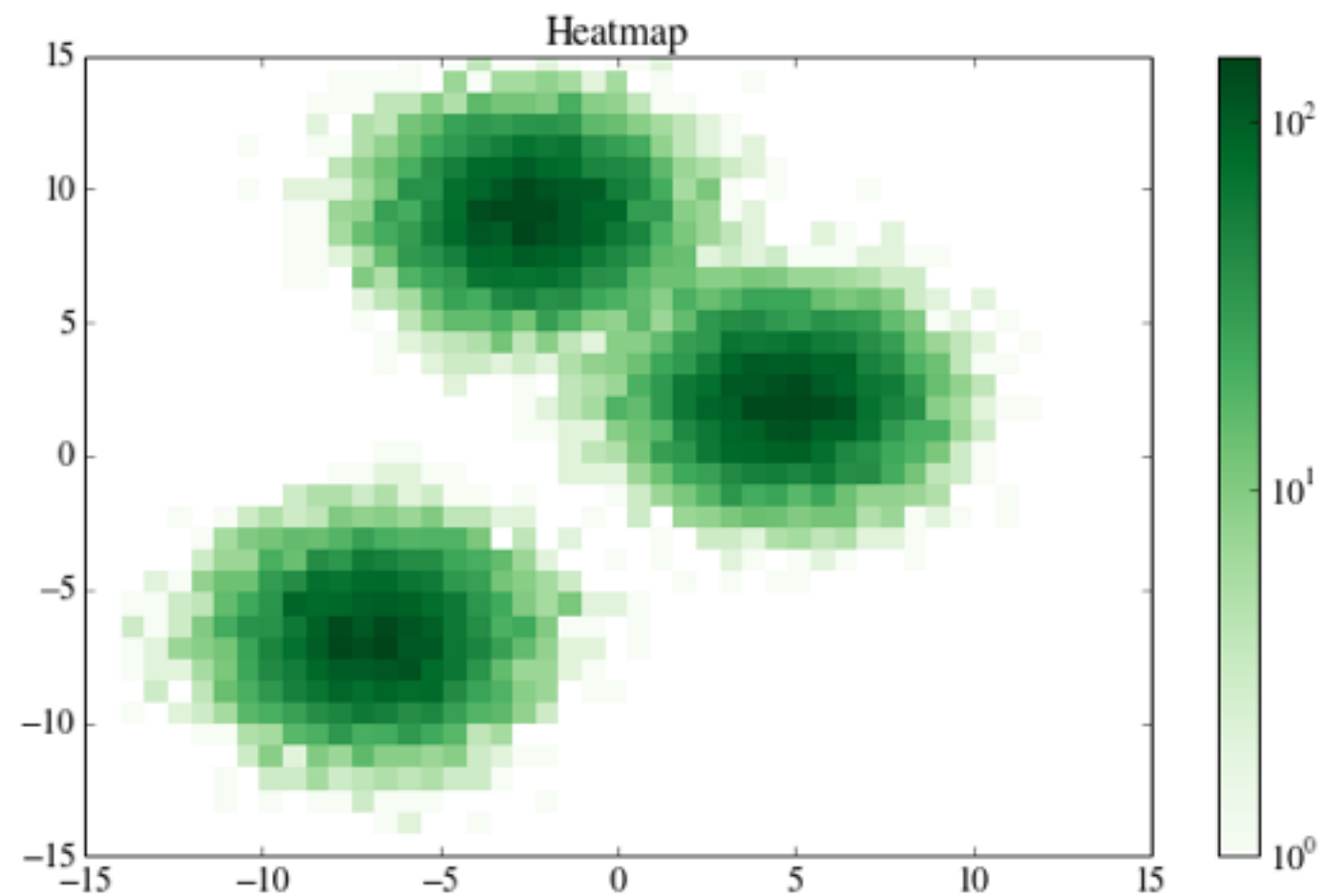
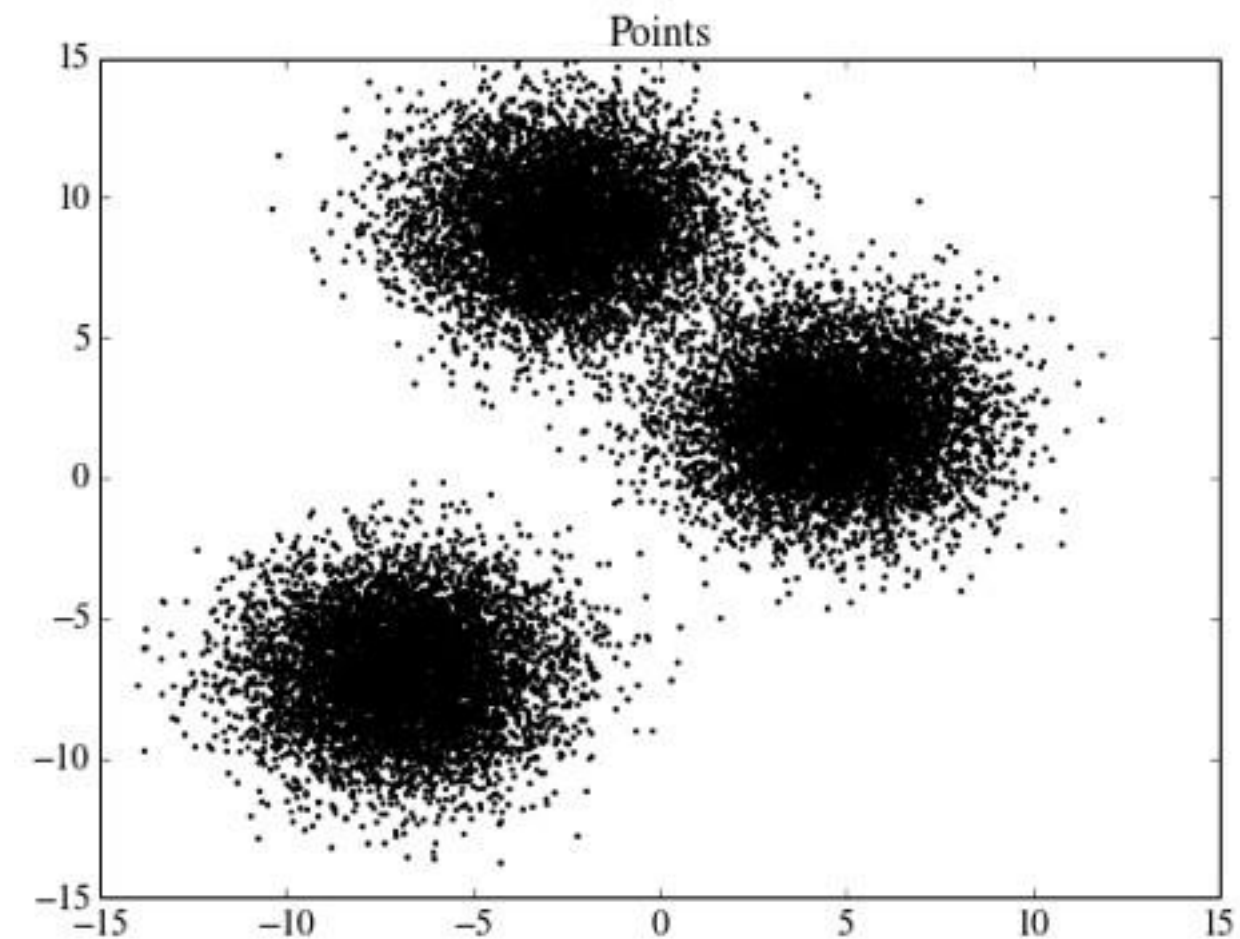
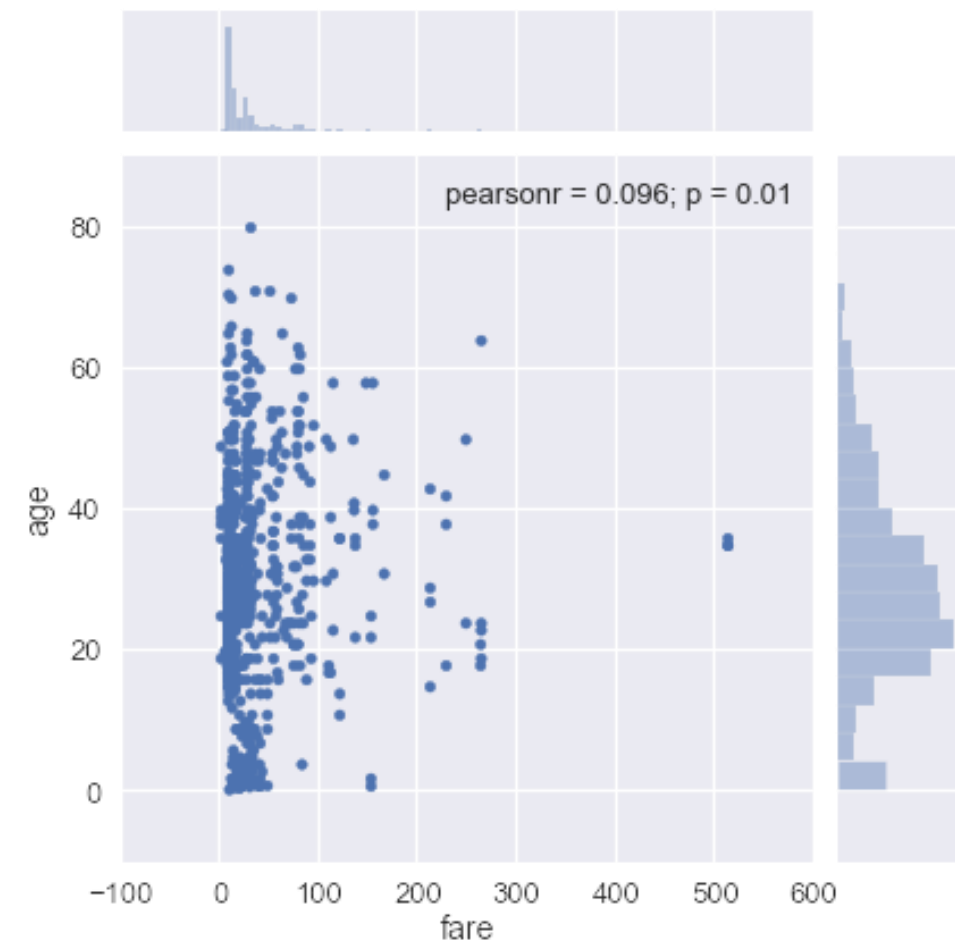


20 Bins

Density Plots



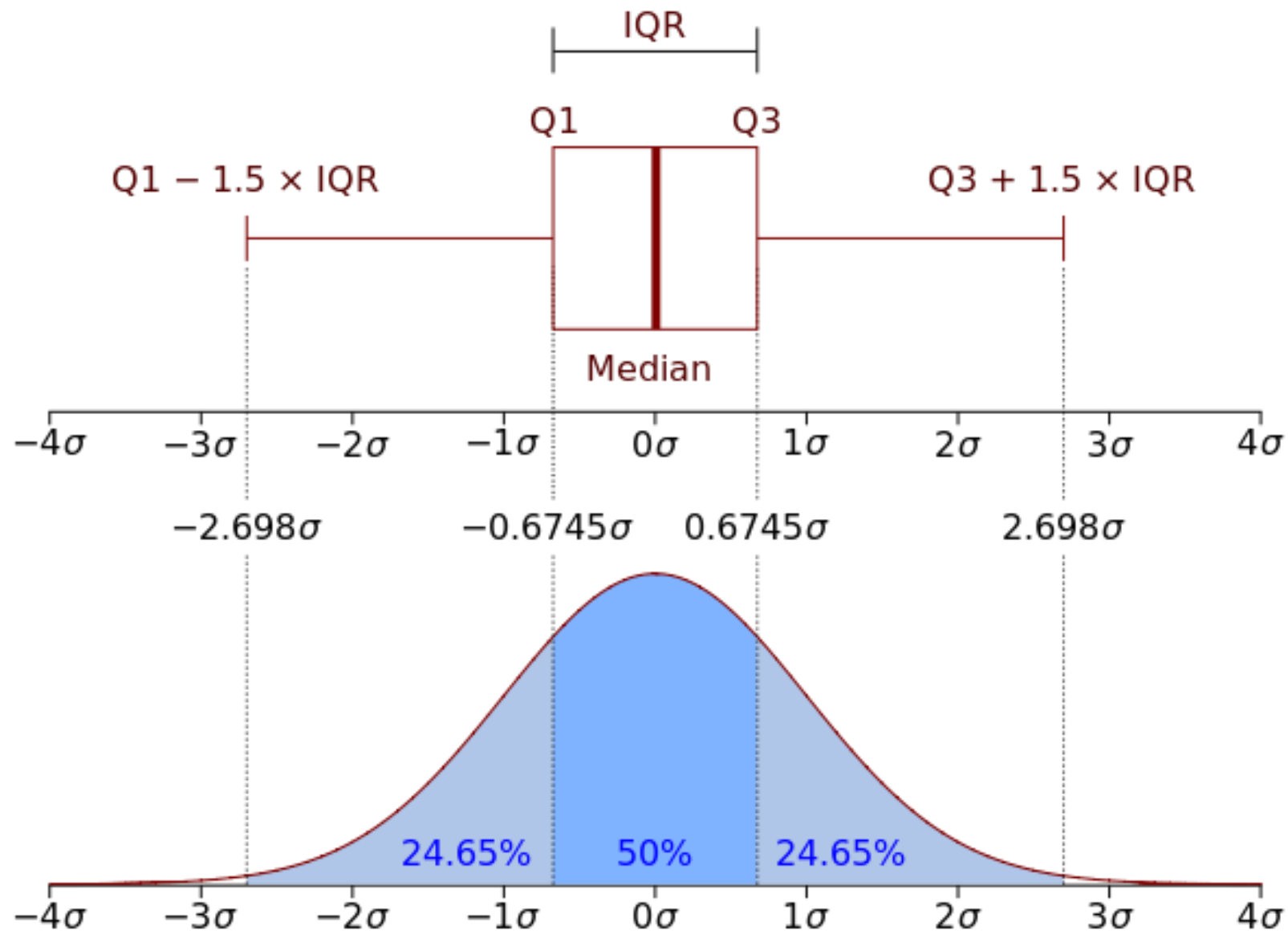
Heat Maps



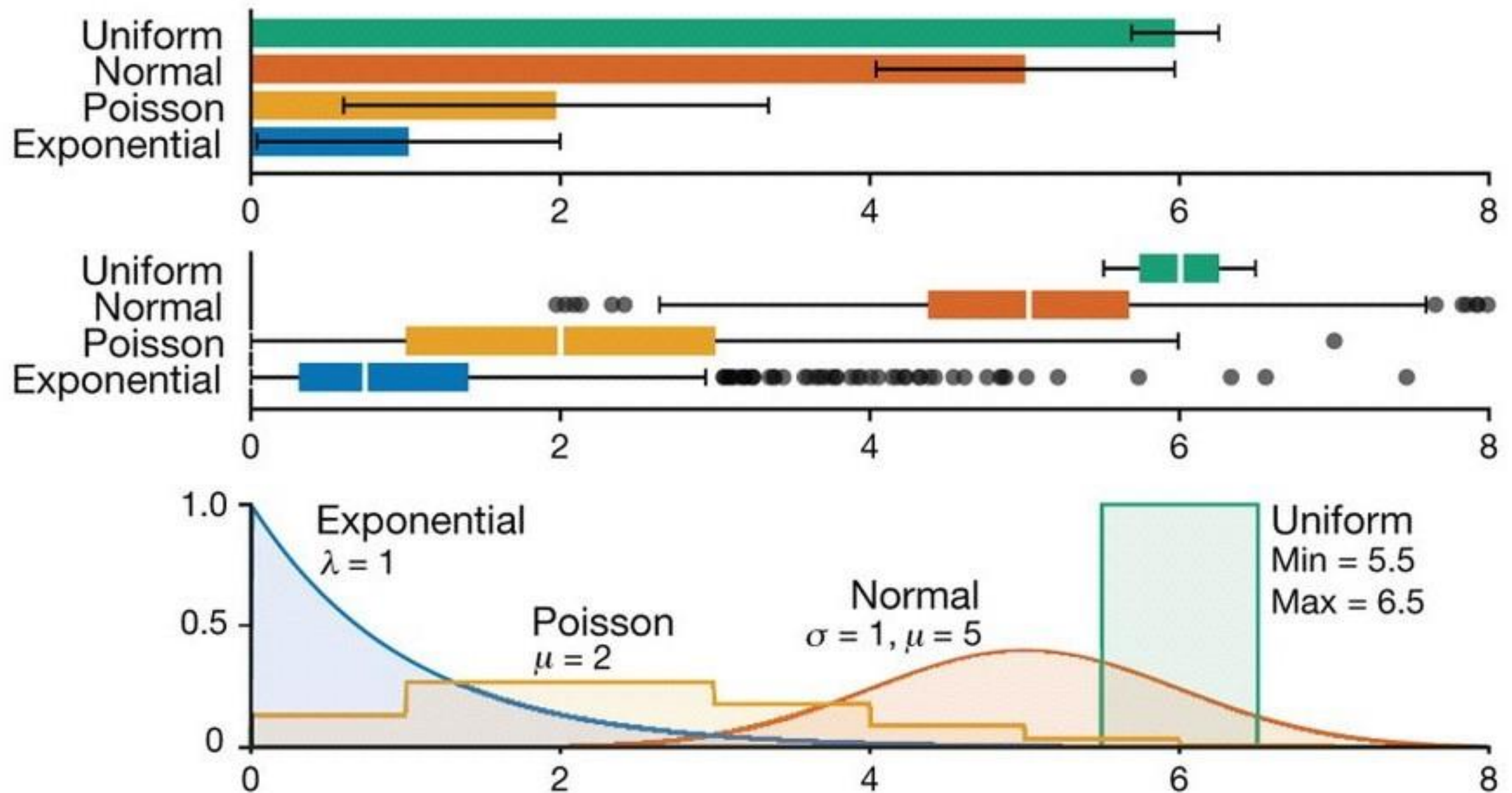
2D Density Plots

Box Plots

aka Box-and-Whisker Plot



Comparison



Spatial visualization

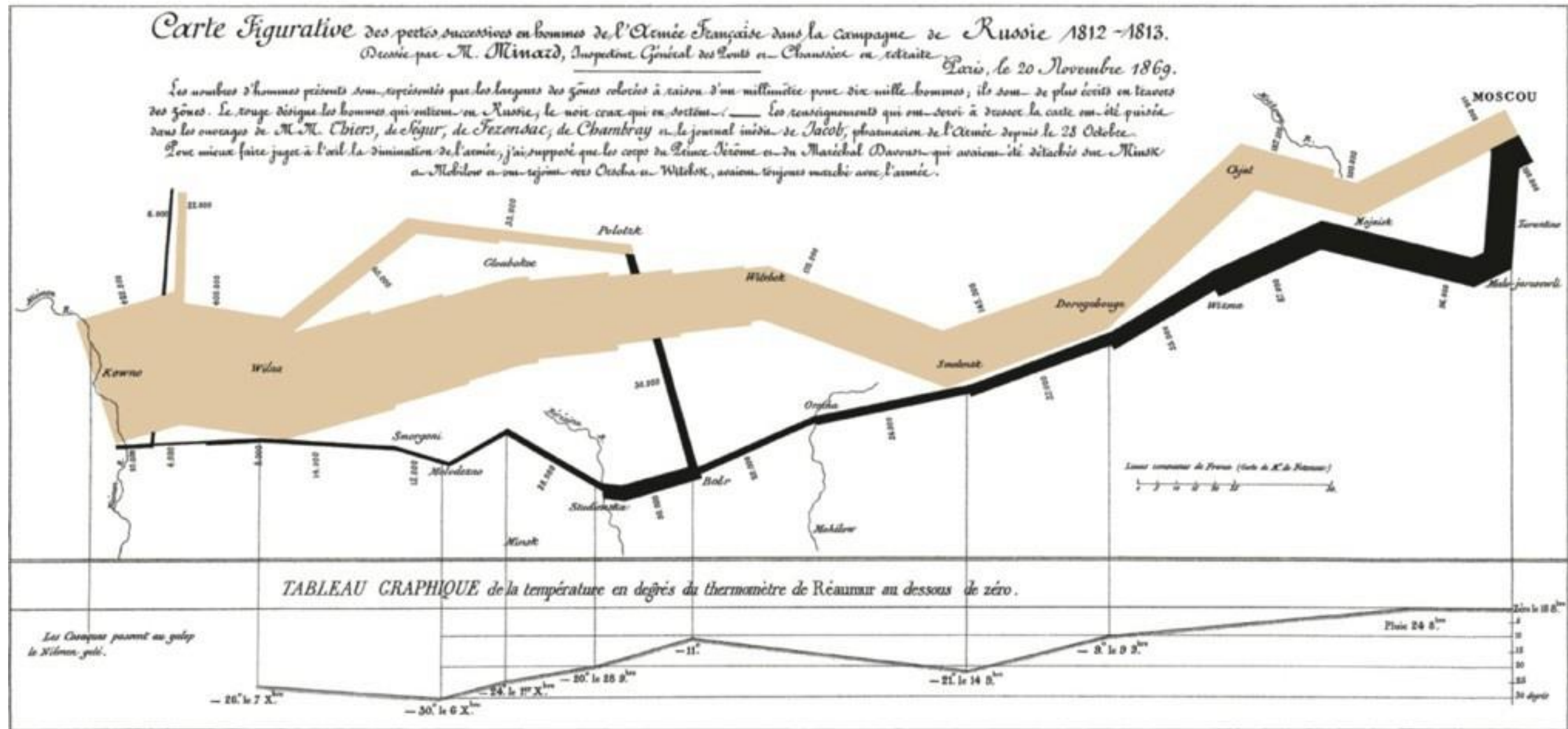
Minard's Map

Napoleon's March on Moscow

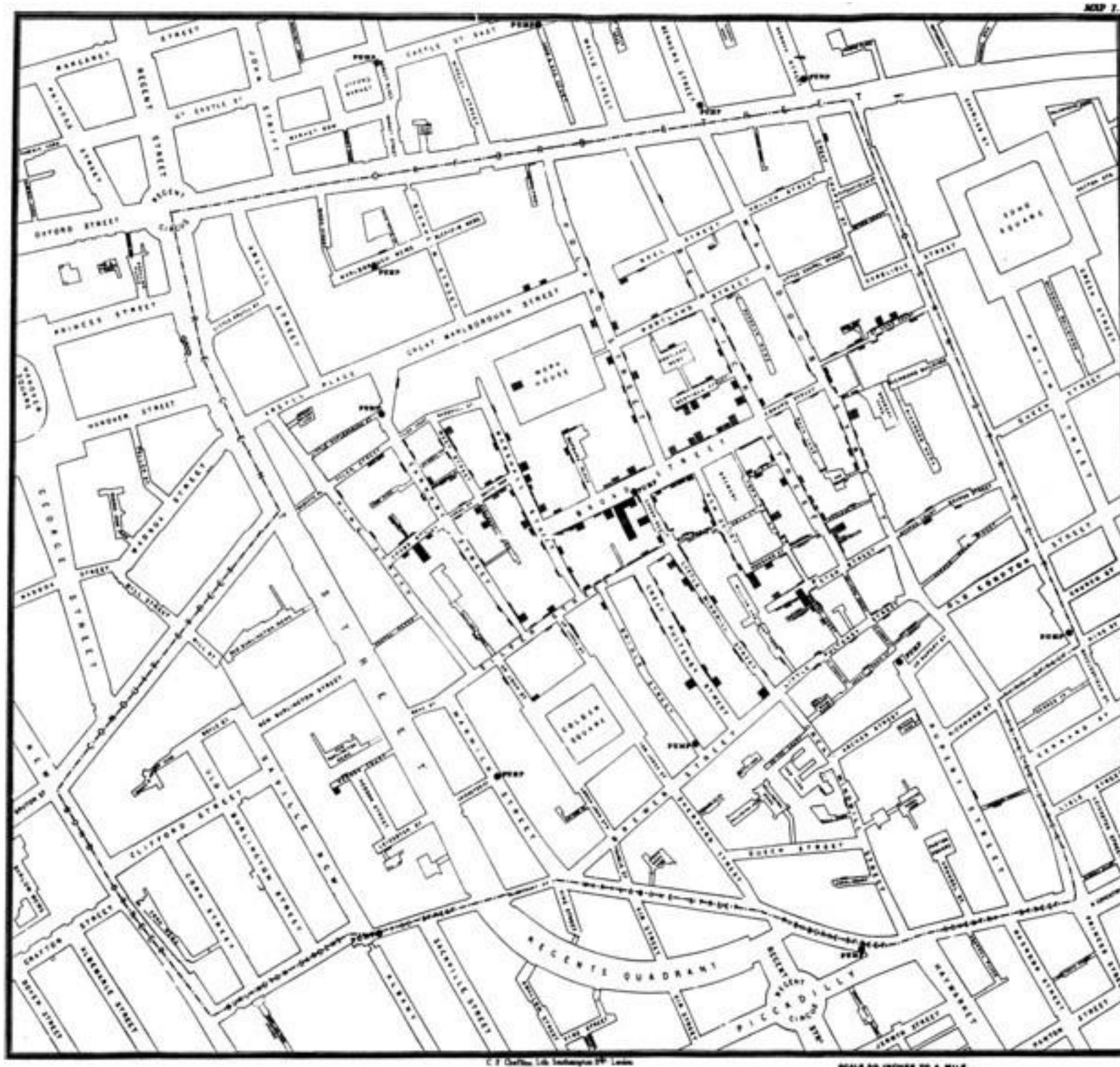
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Devisé par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Séjur, de Fozensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilew n'en rejoignent vers Orscha ou Wilkoff, avaient toujours marché avec l'armée.



Detect the expected – discover the unexpected



C. F. Duffin Ltd. London

SCALE 80 INCHES TO A MILE.

Wikimedia Commons



John Snow
(1813 – 1858)



Effective Visualizations

Not Effective...

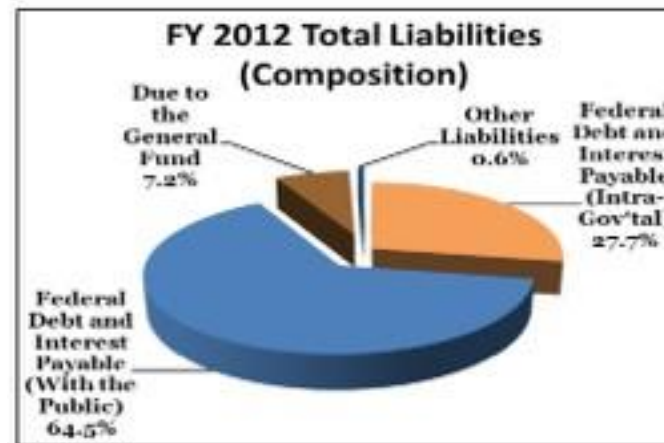
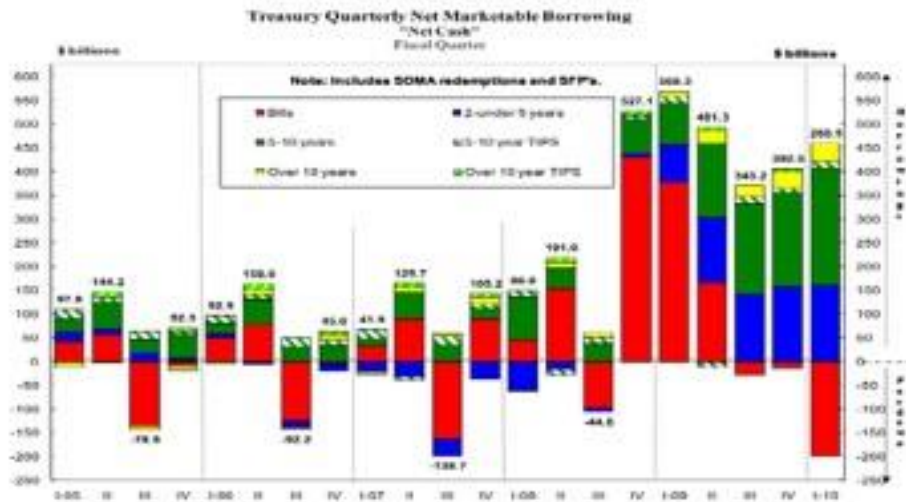


Figure 10

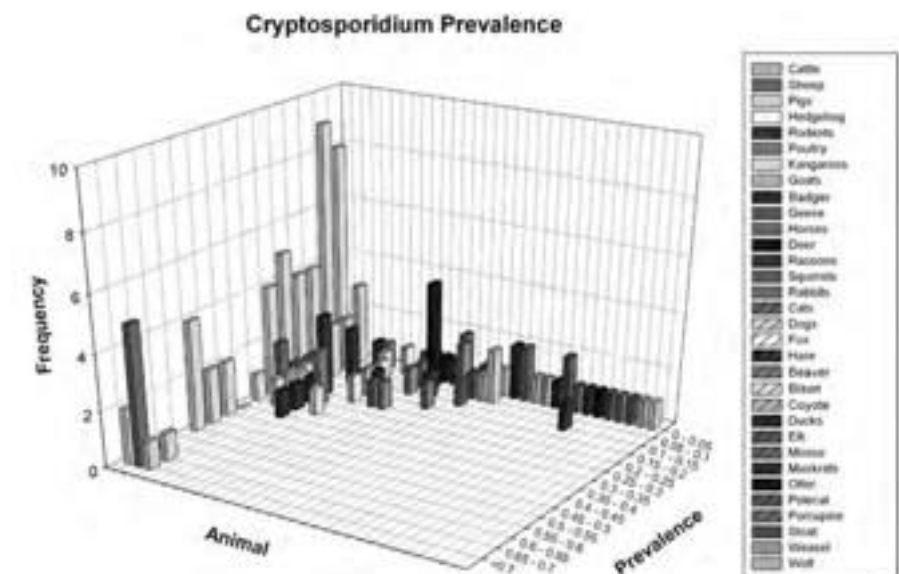
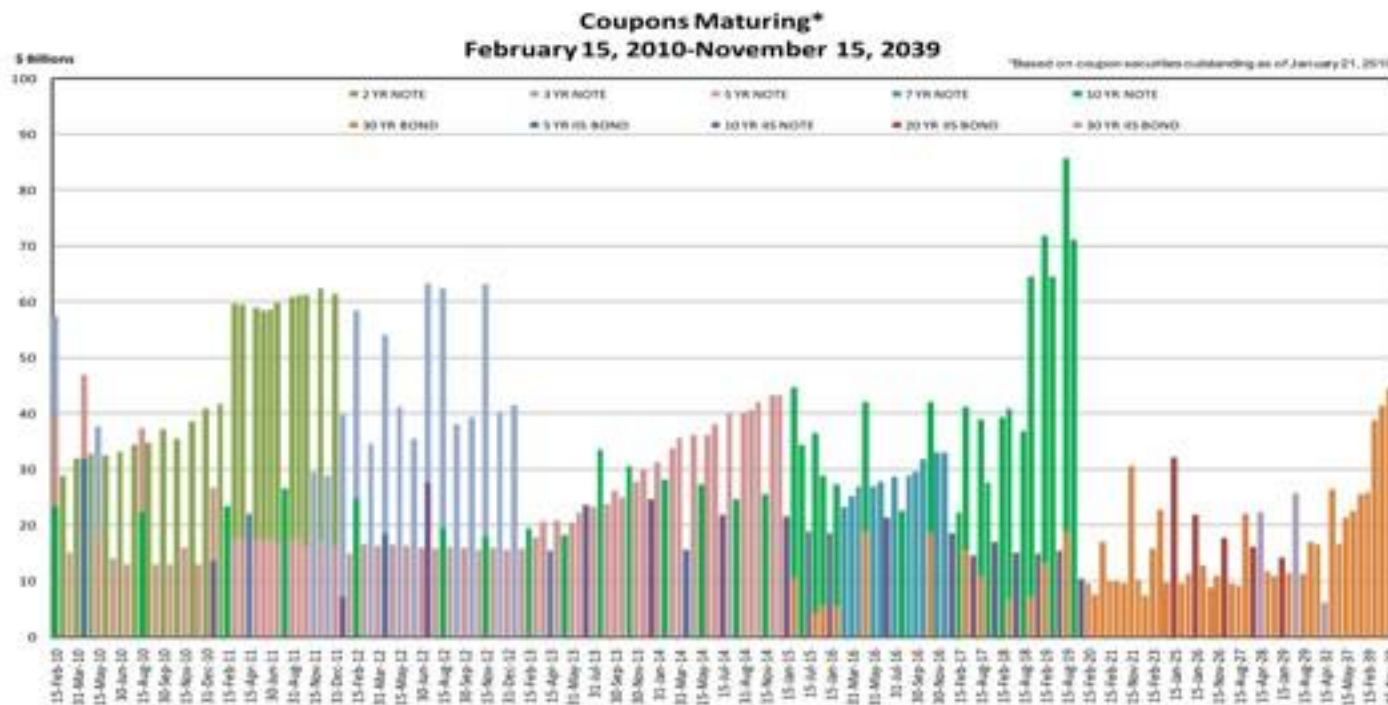
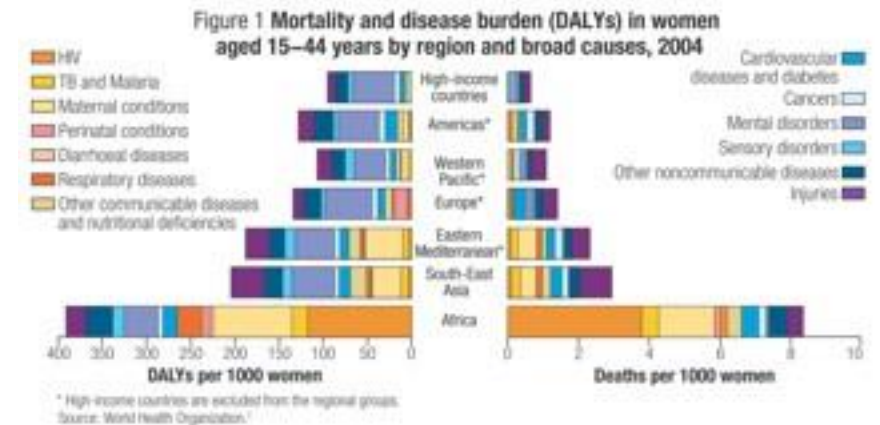
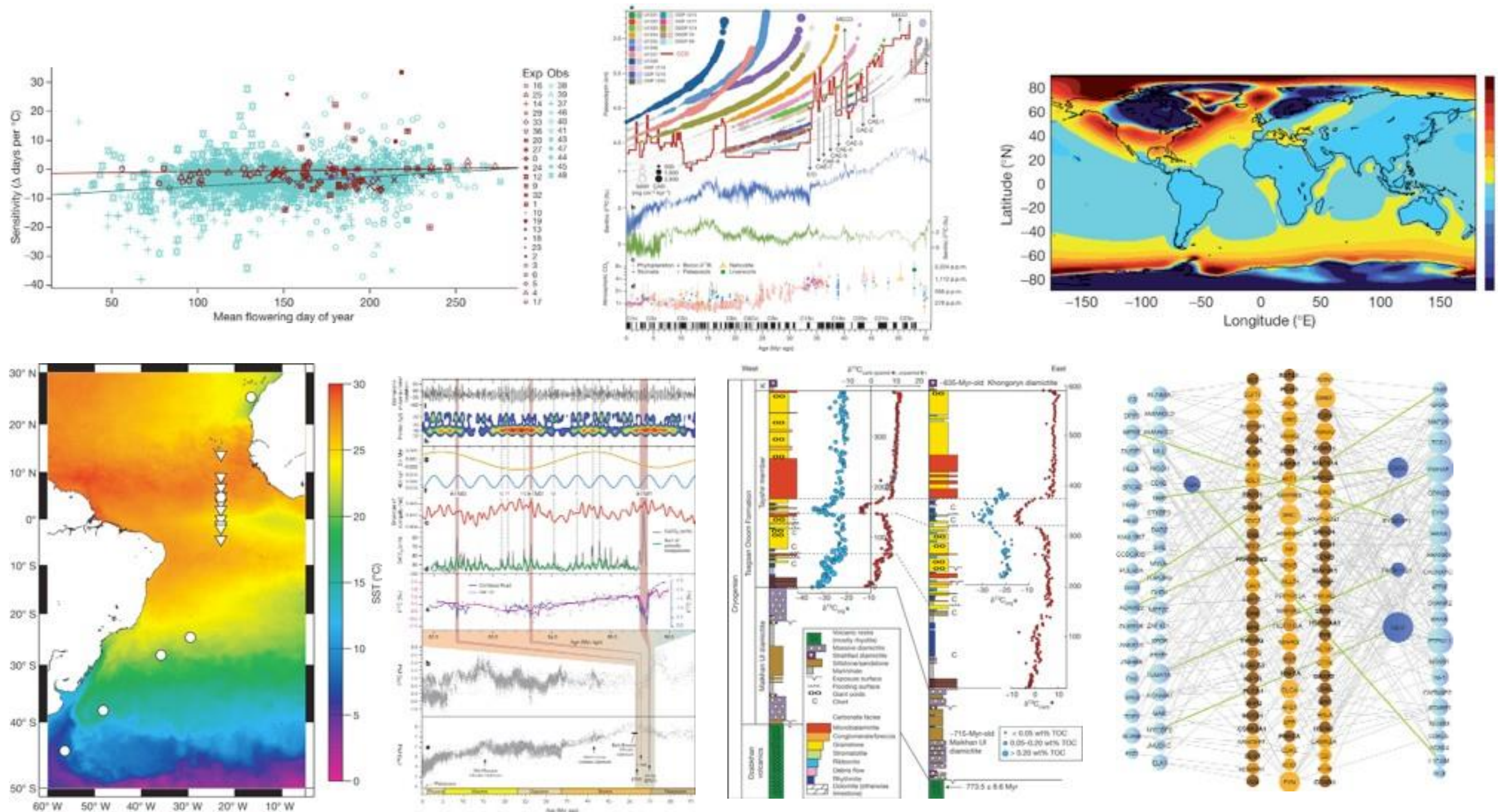
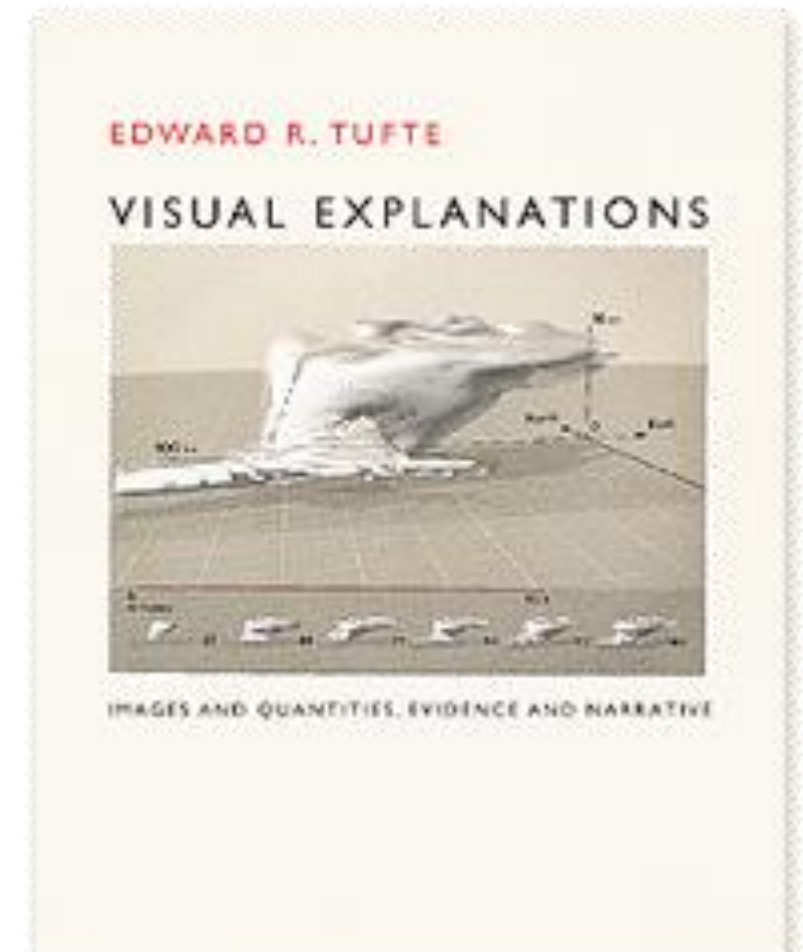
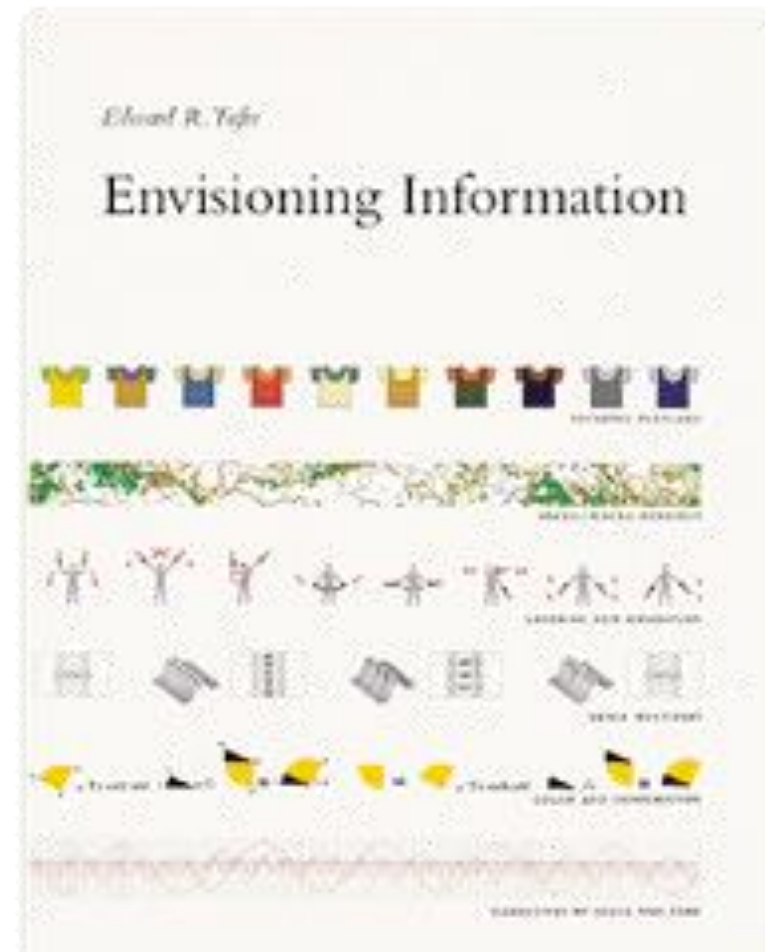
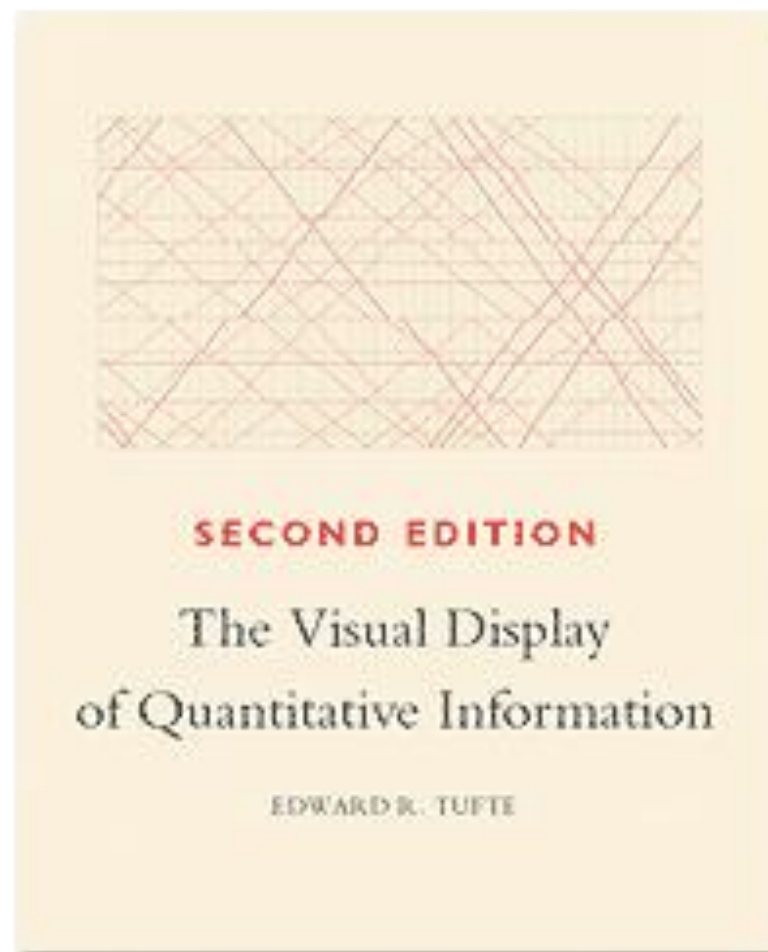


Figure 5.2 Mean prevalence rates of *Cryptosporidium* oocysts by animal species.

Also *not* effective...



Edward Tufte



Scale Distortions

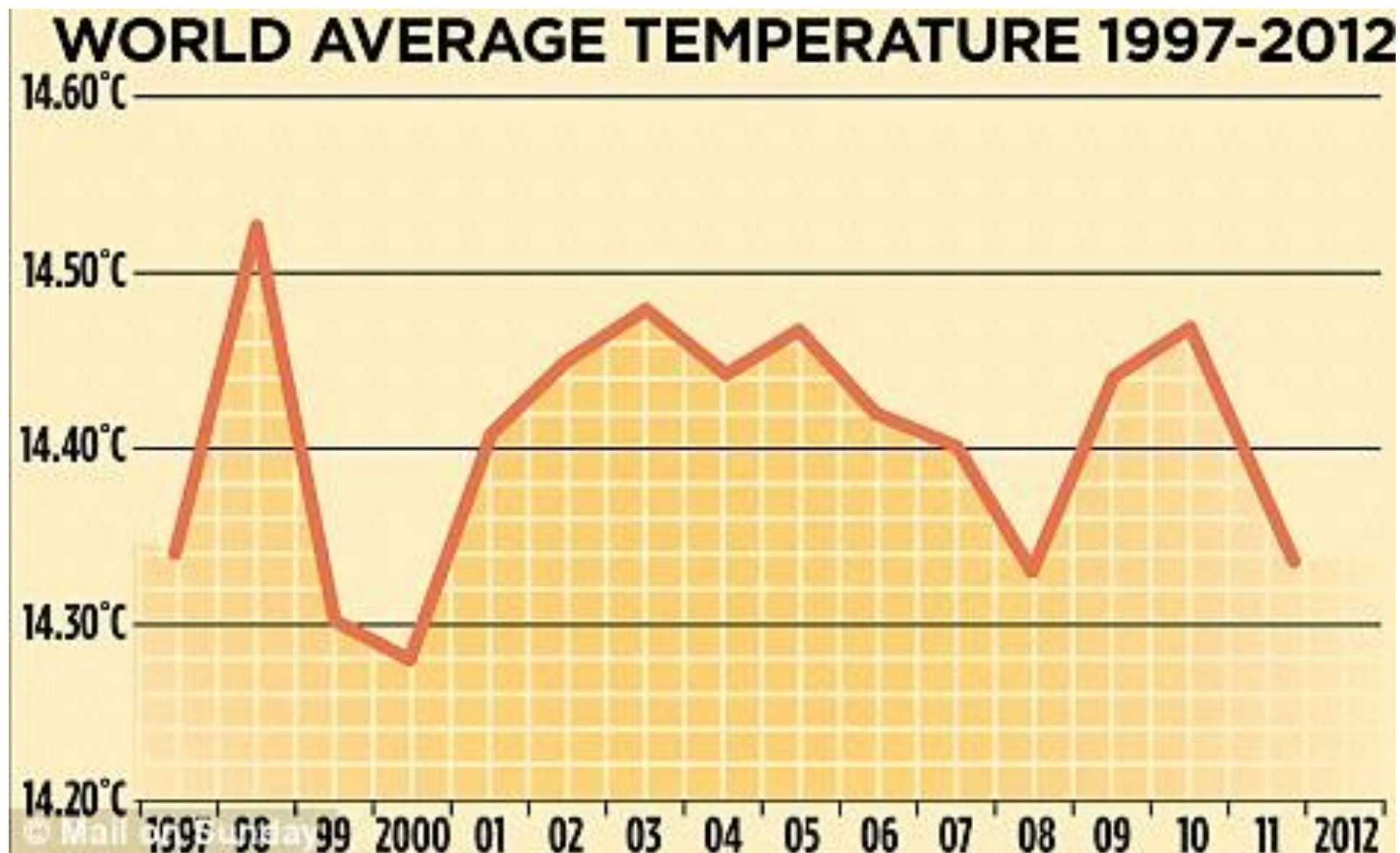
HOW 2012 STACKS UP

THE WARMEST YEARS ON RECORD
CONTIGUOUS U.S.

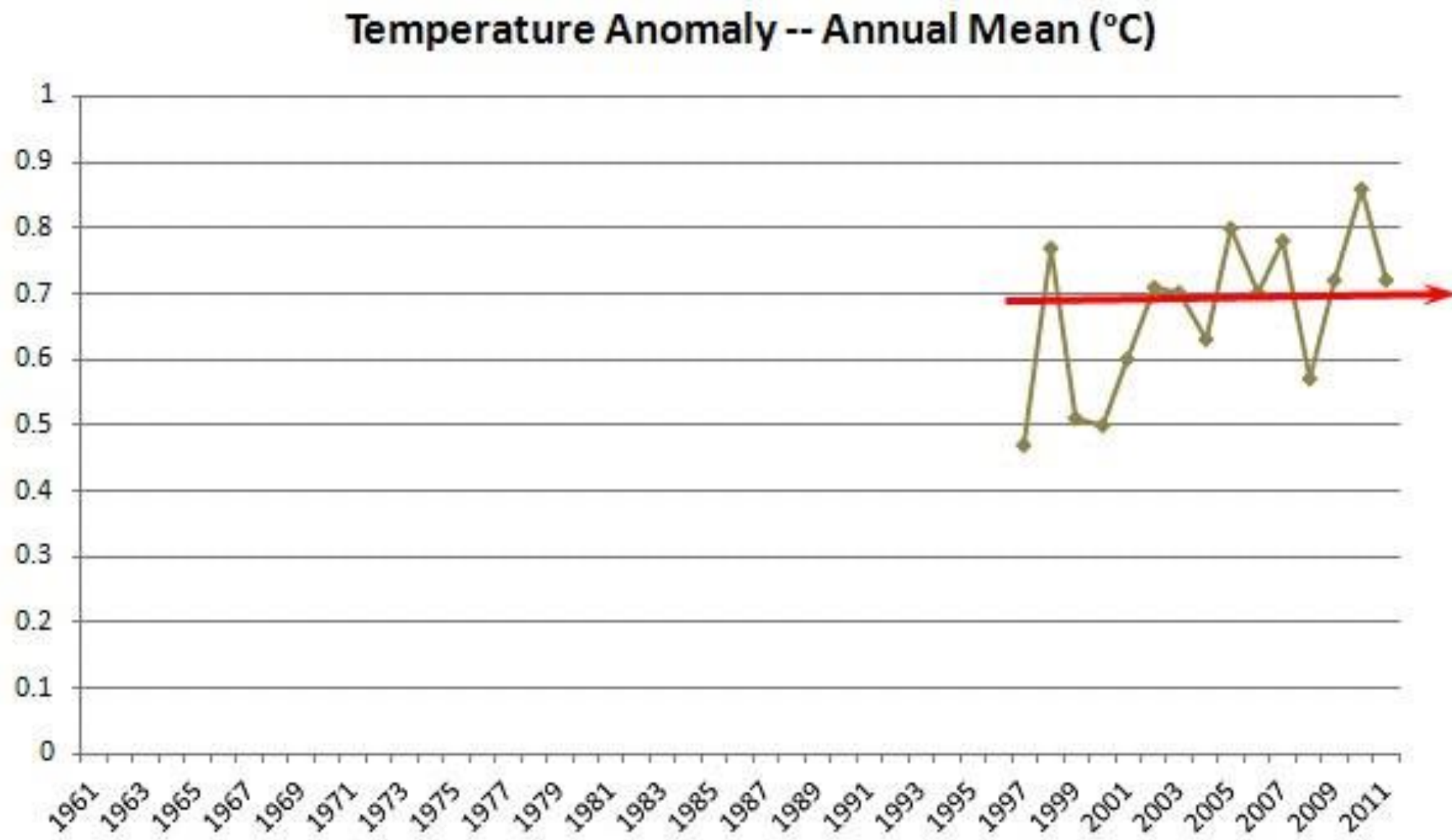


Source: NOAA's National Climatic Data Center - State of the Climate National Overview

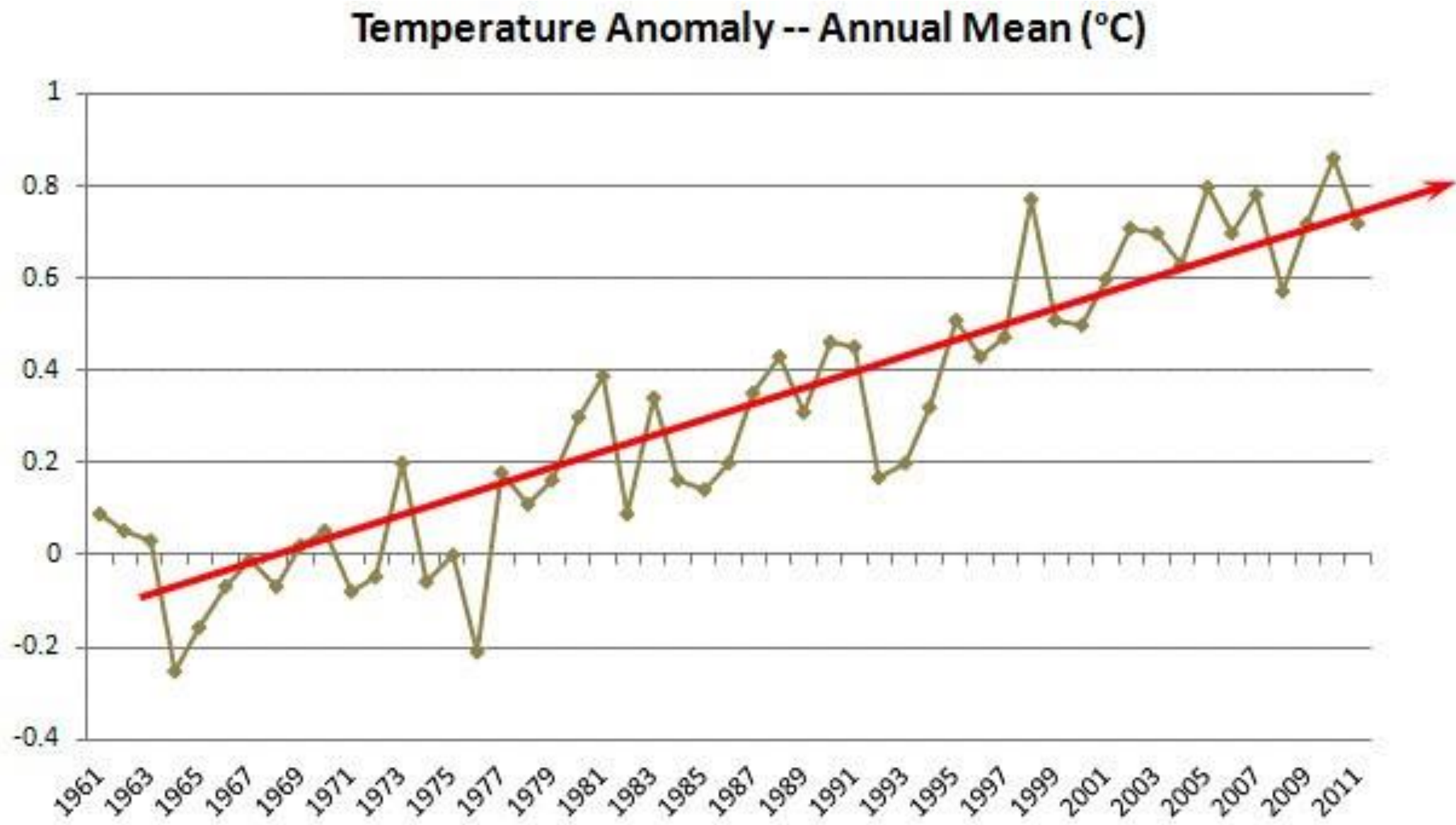
Global Warming?



Global Warming?

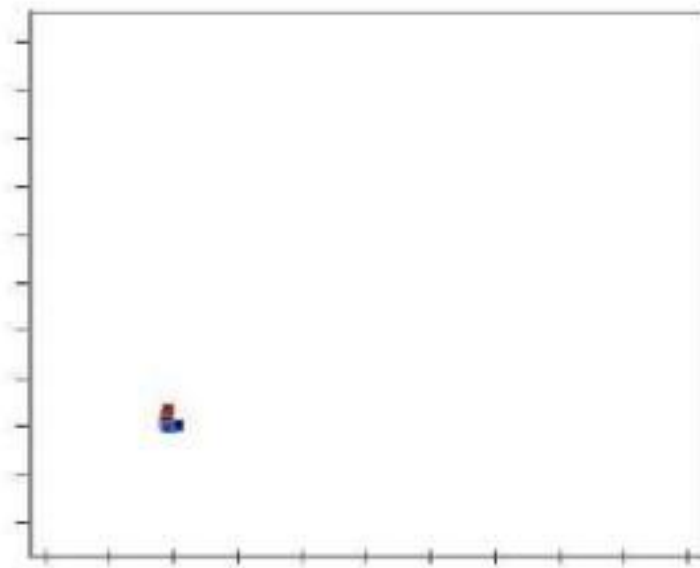


Global Warming!

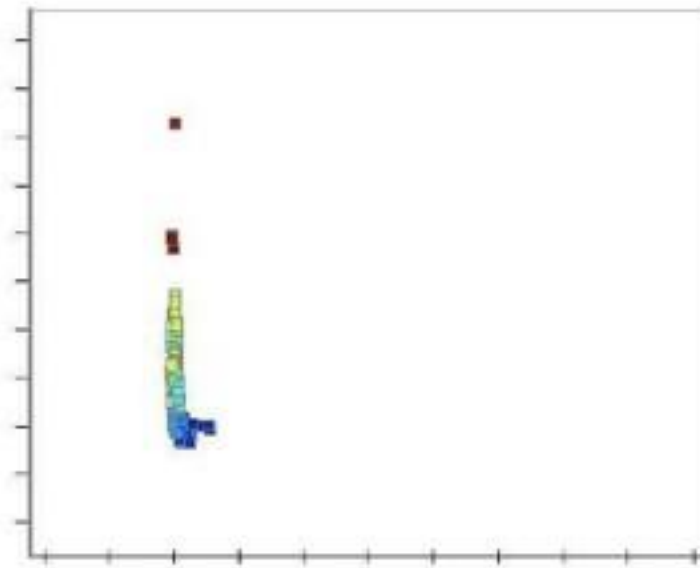


Lying with Scales

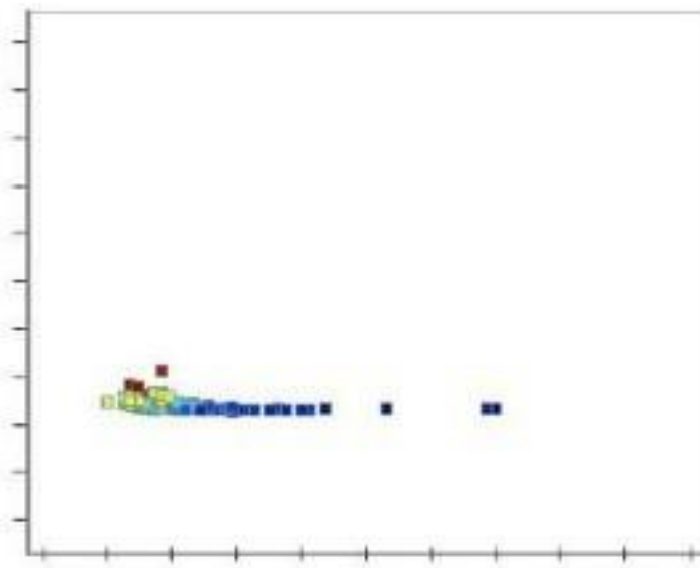
Same data - different scales



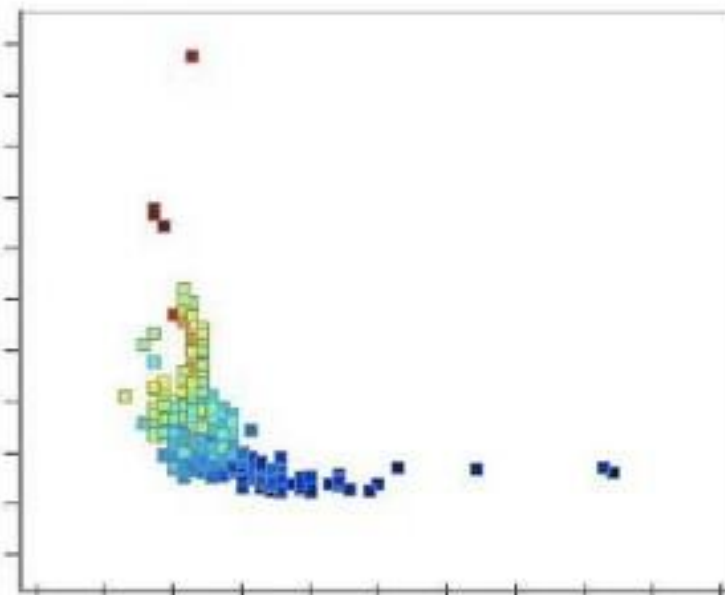
Uniform scale in both x and y



Larger scale in y



Larger scale in x



Larger scale in x and y

Scales are critical!

What are your bounds – upper and lower?

What scale works?

Linear? Log? Clipping? Breaks?

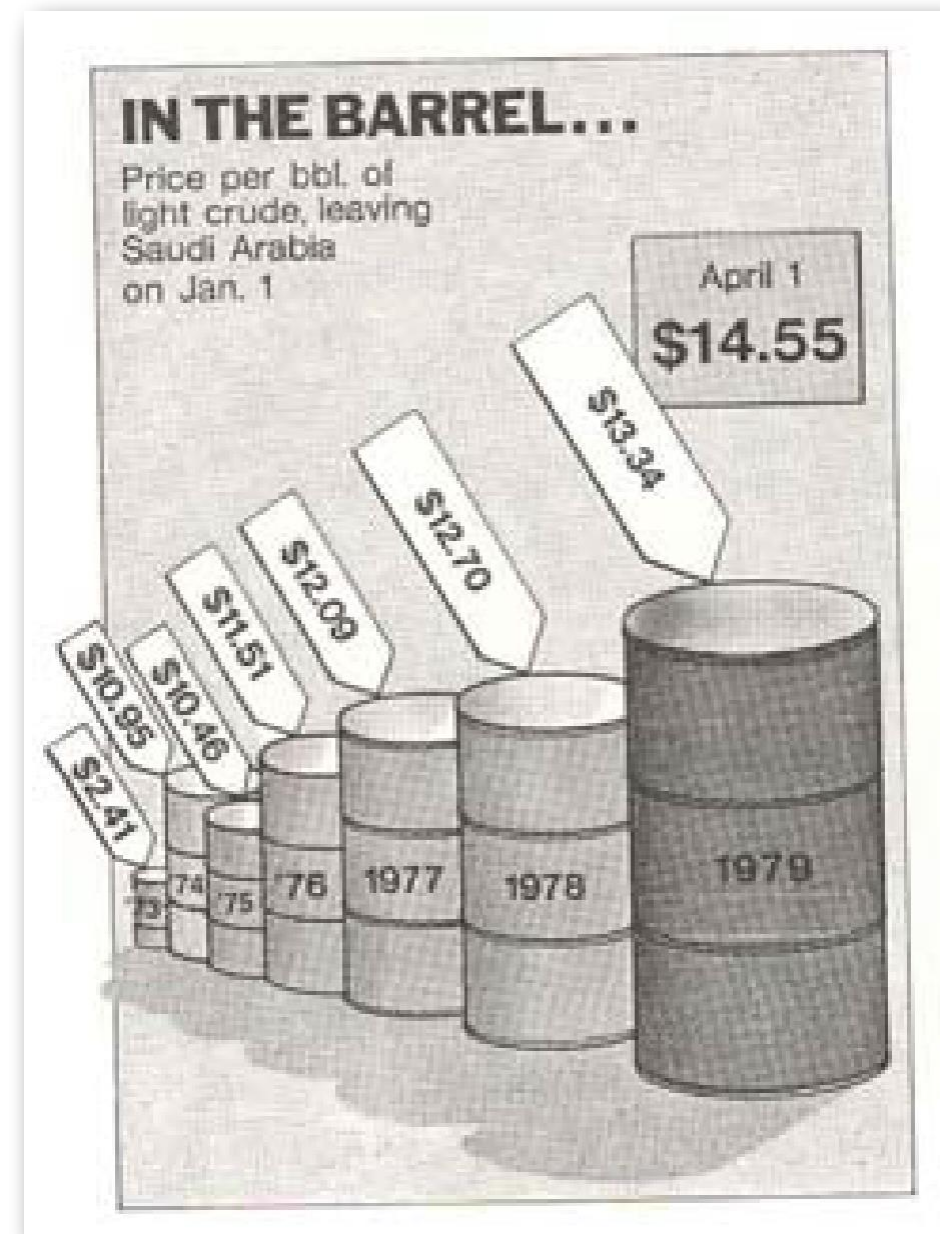
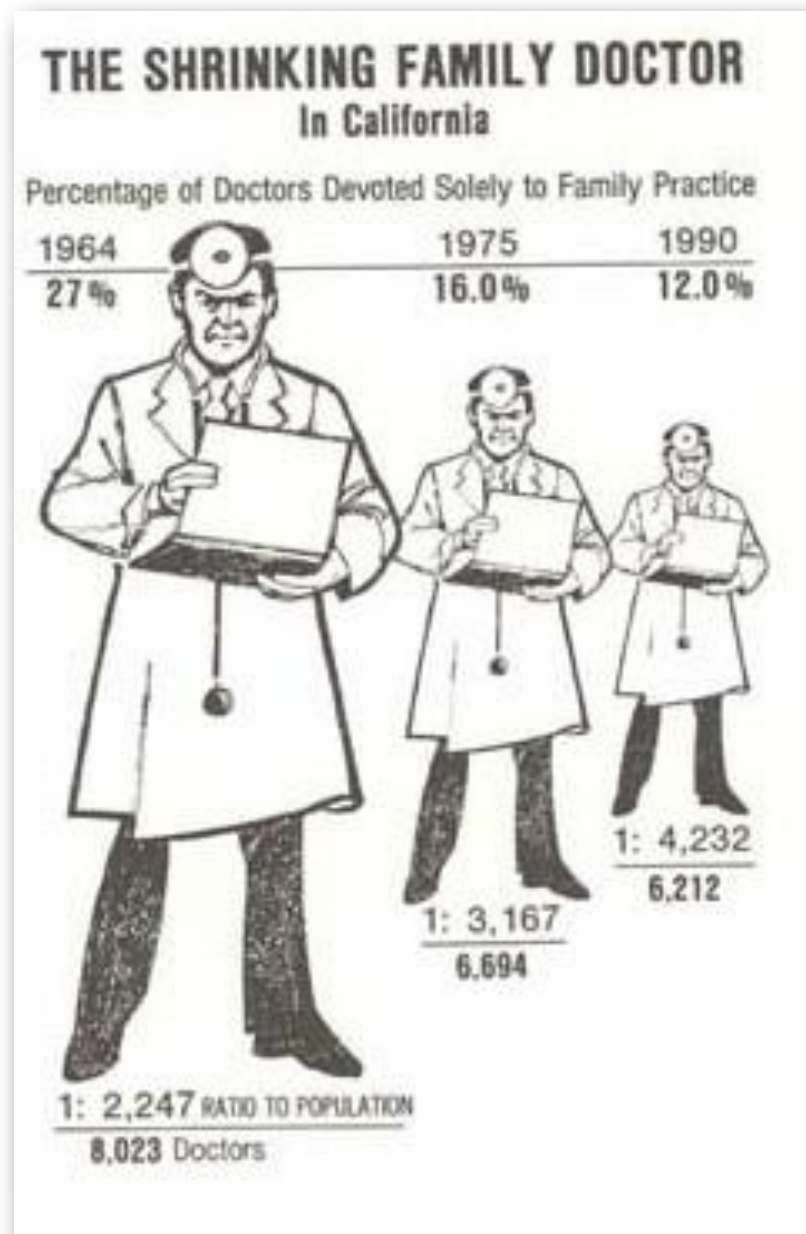
Relative or absolute values?

How can you make things comparable?

The Lie Factor

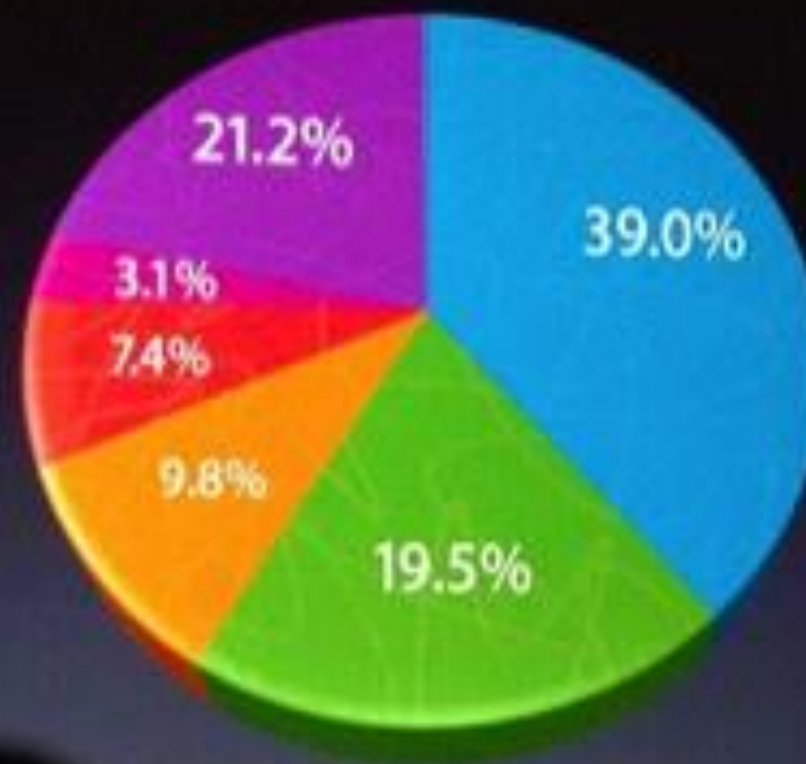
Size of effect shown in graphic

Size of effect in data



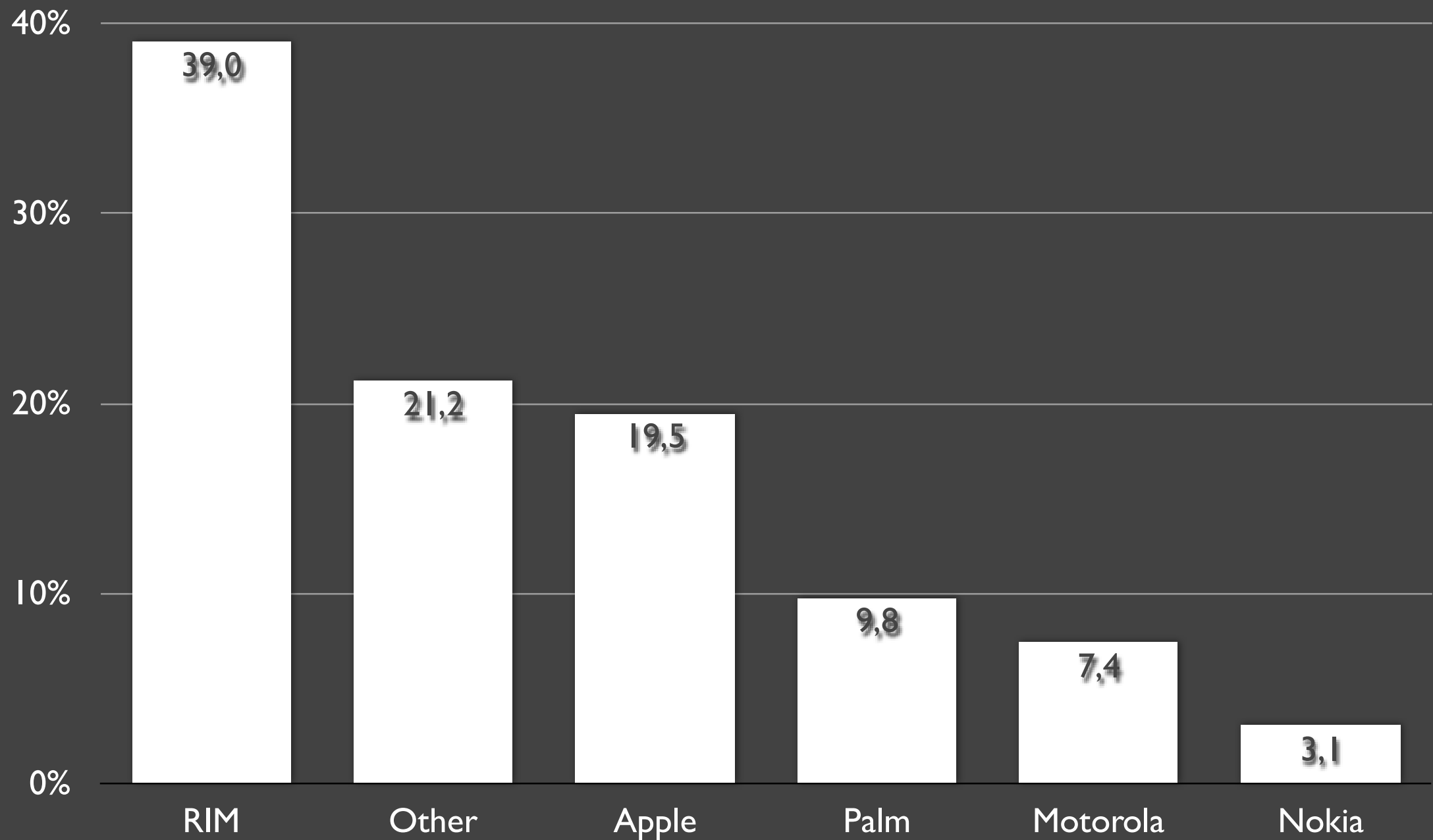
U.S. SmartPhone Marketshare

- RIM
- Apple
- Palm
- Motorola
- Nokia
- Other



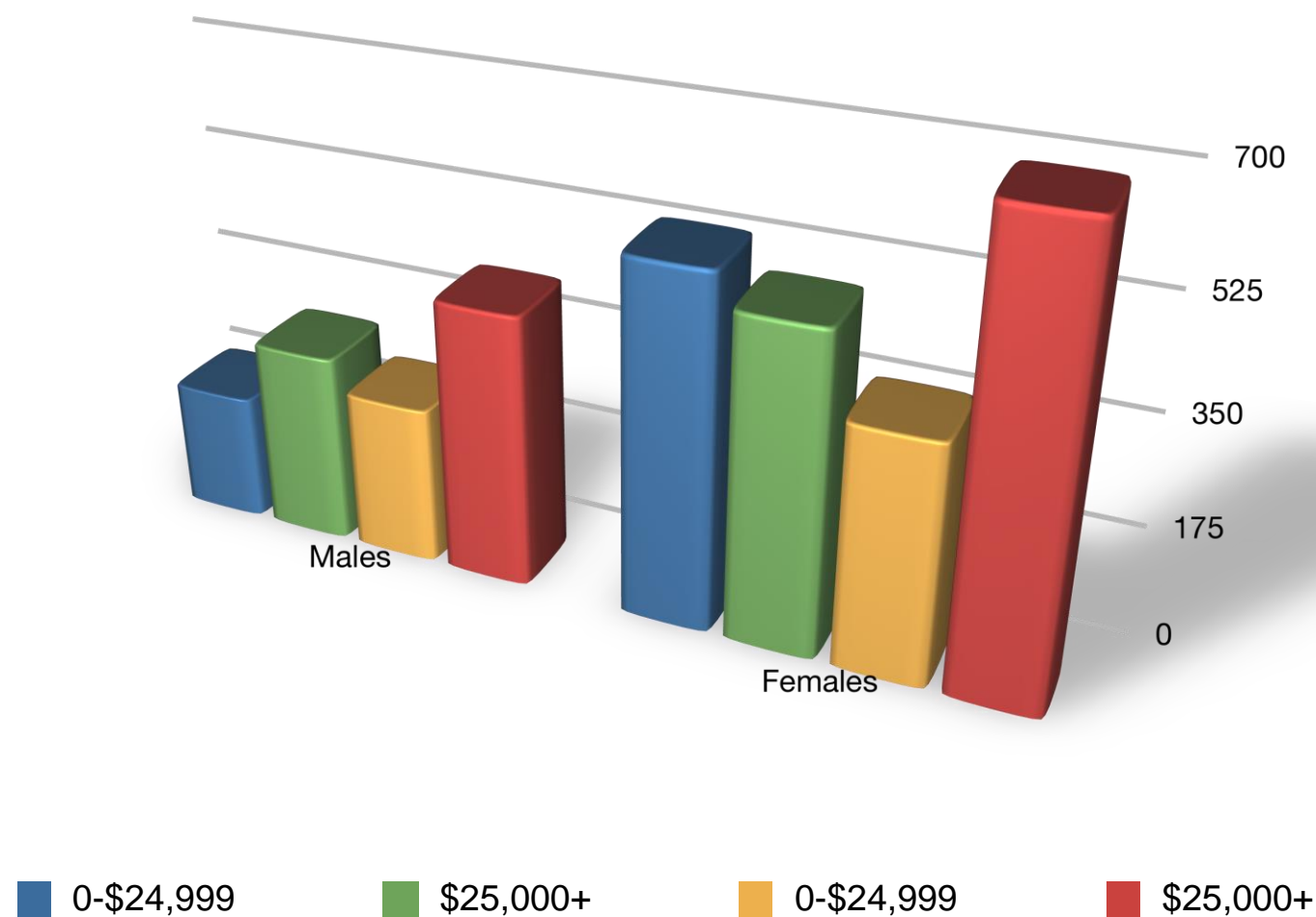
Ernst & Gartner for

U.S. SmartPhone Marketshare



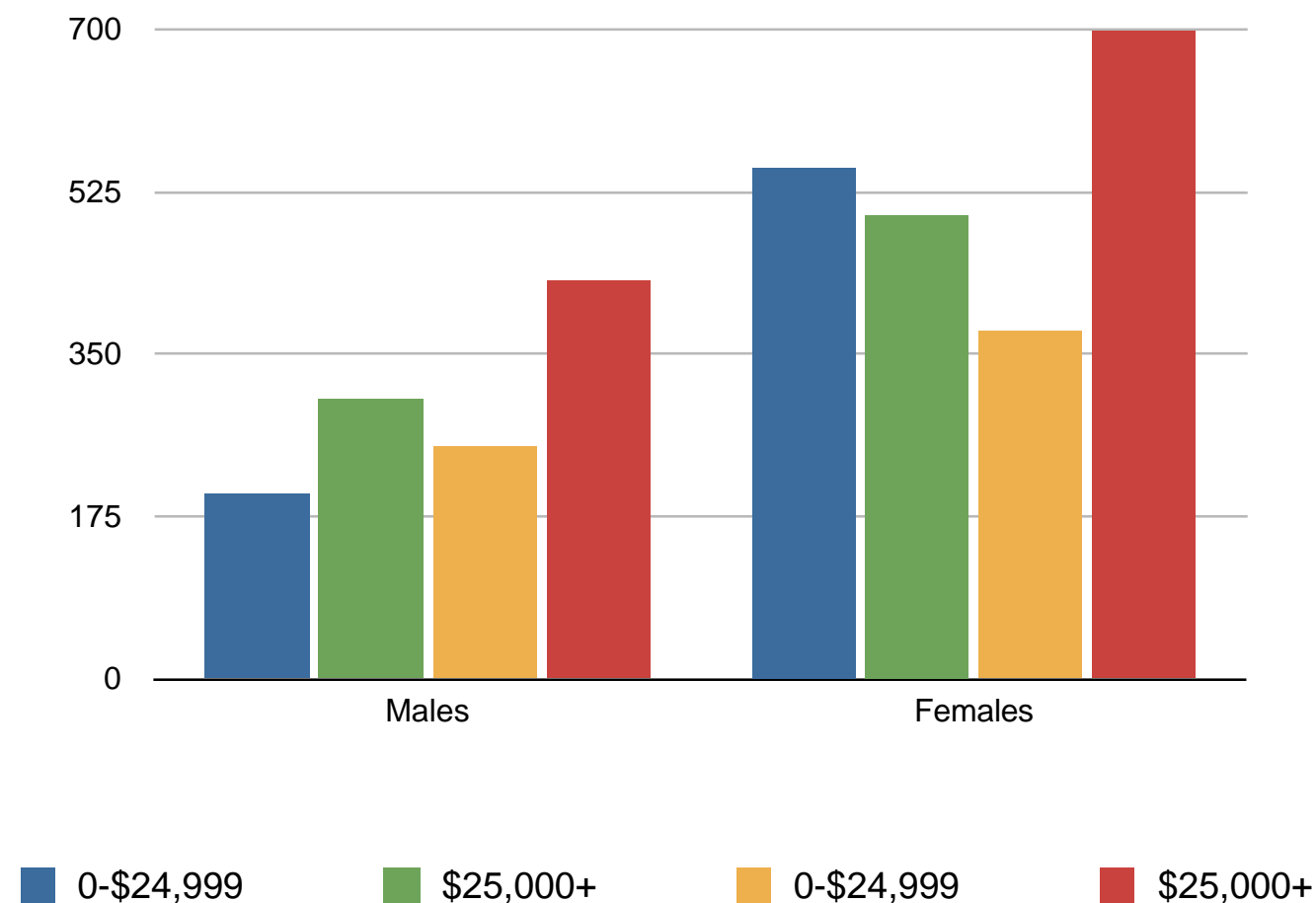
Maximize Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

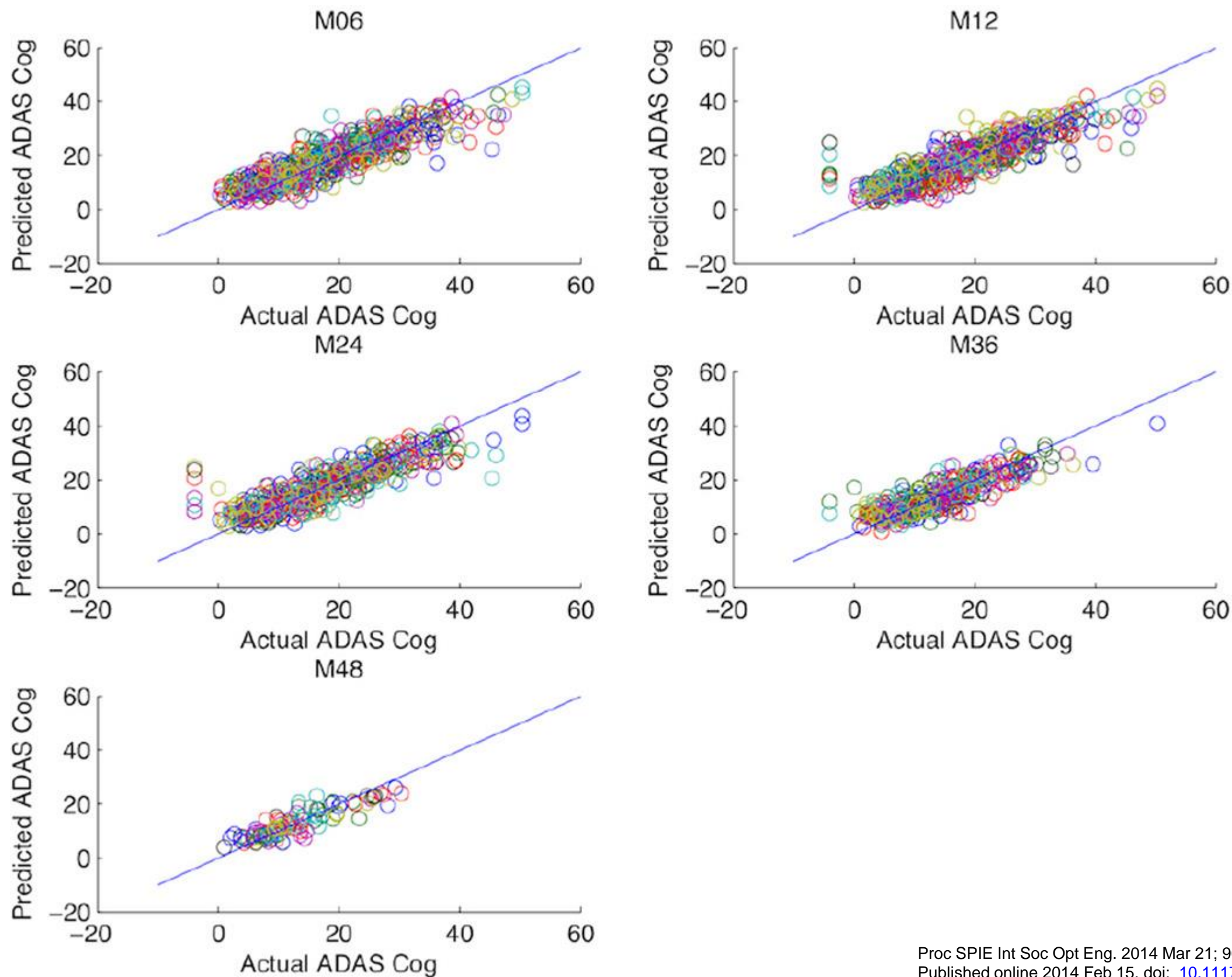


Maximize Data-Ink Ratio

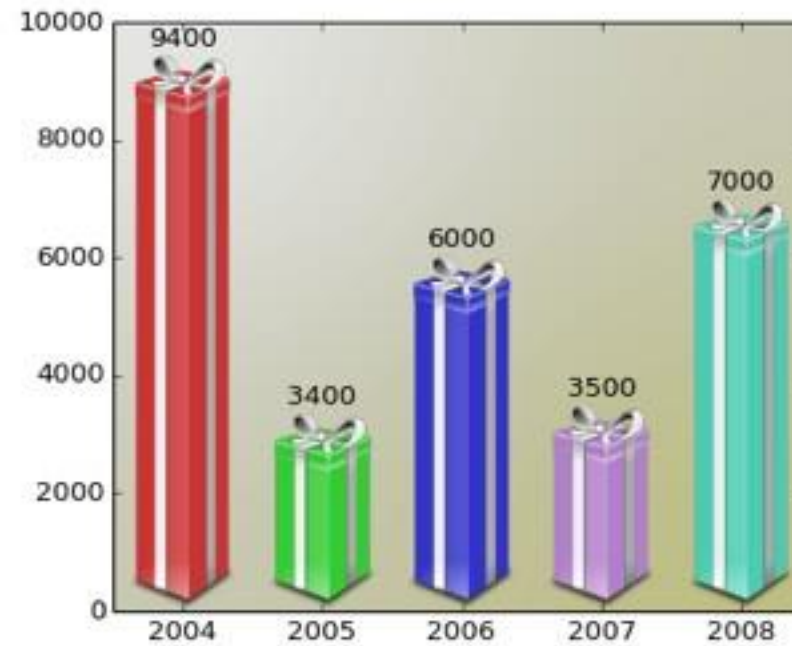
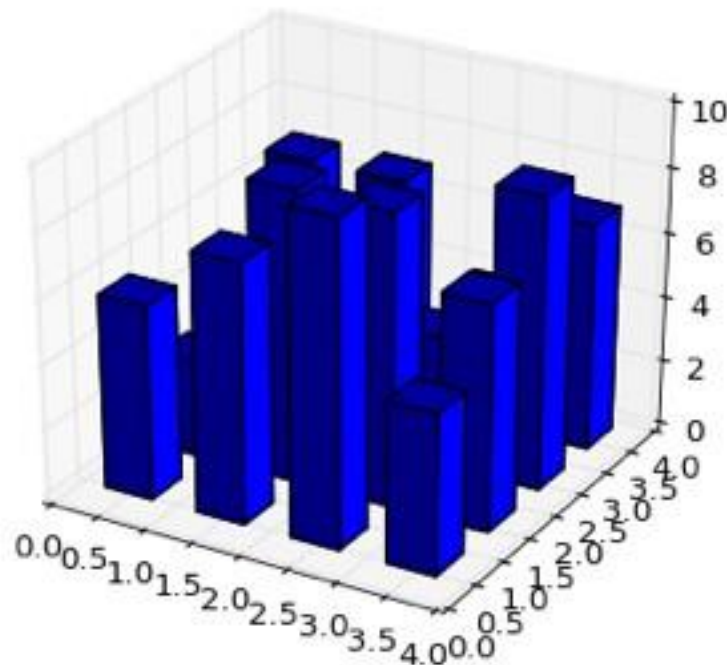
$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$



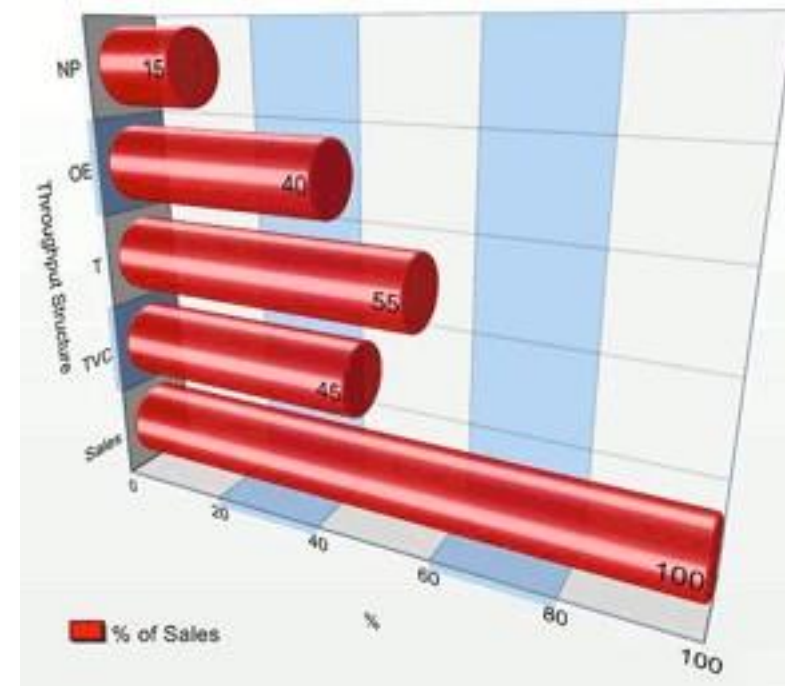
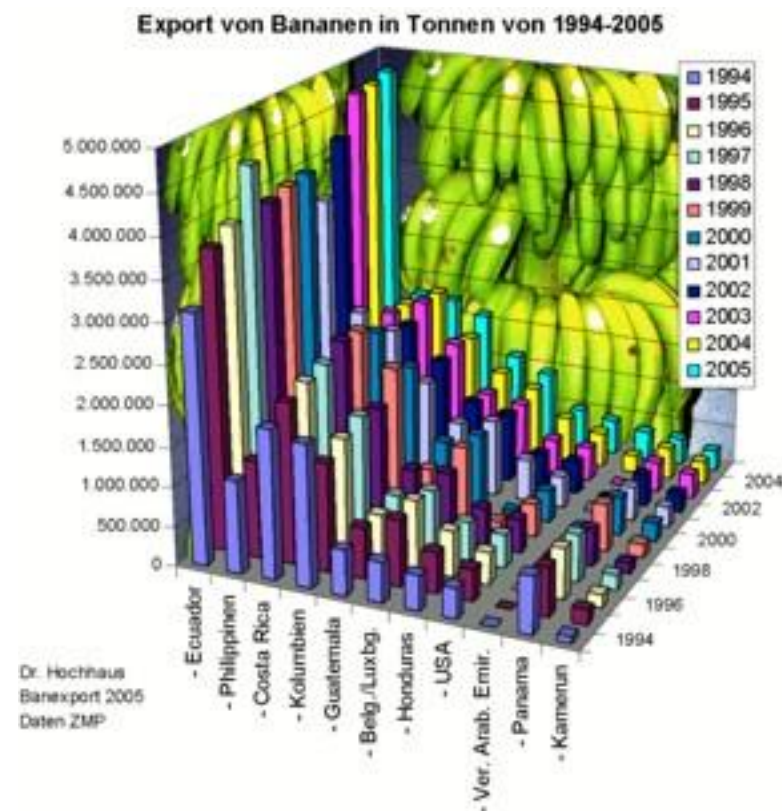
What do colors mean?



Don't Use 3D

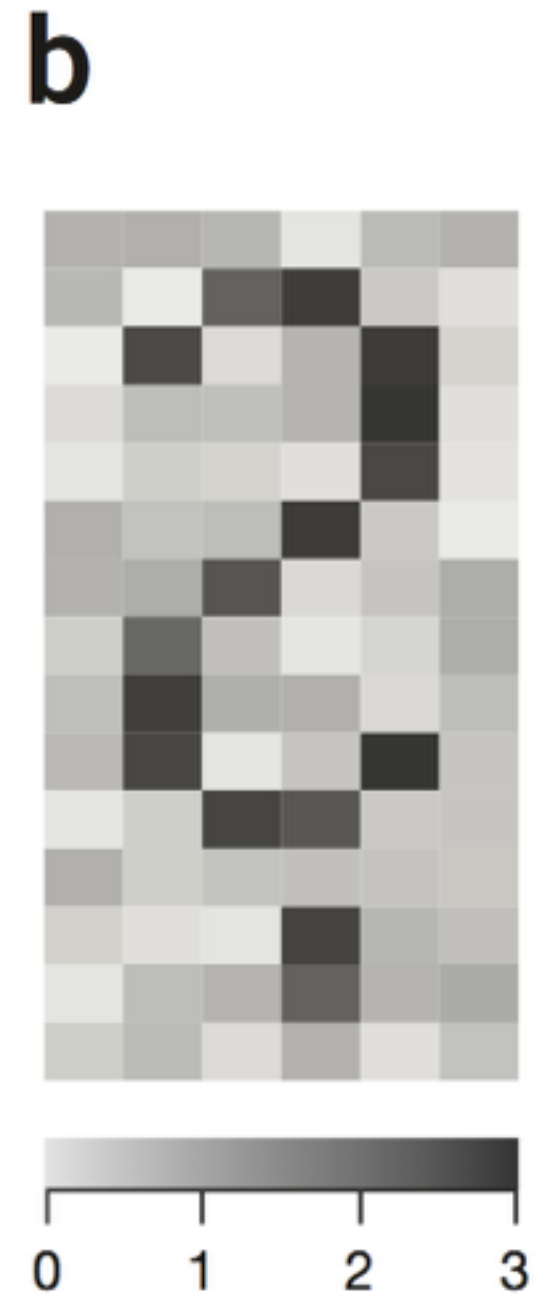
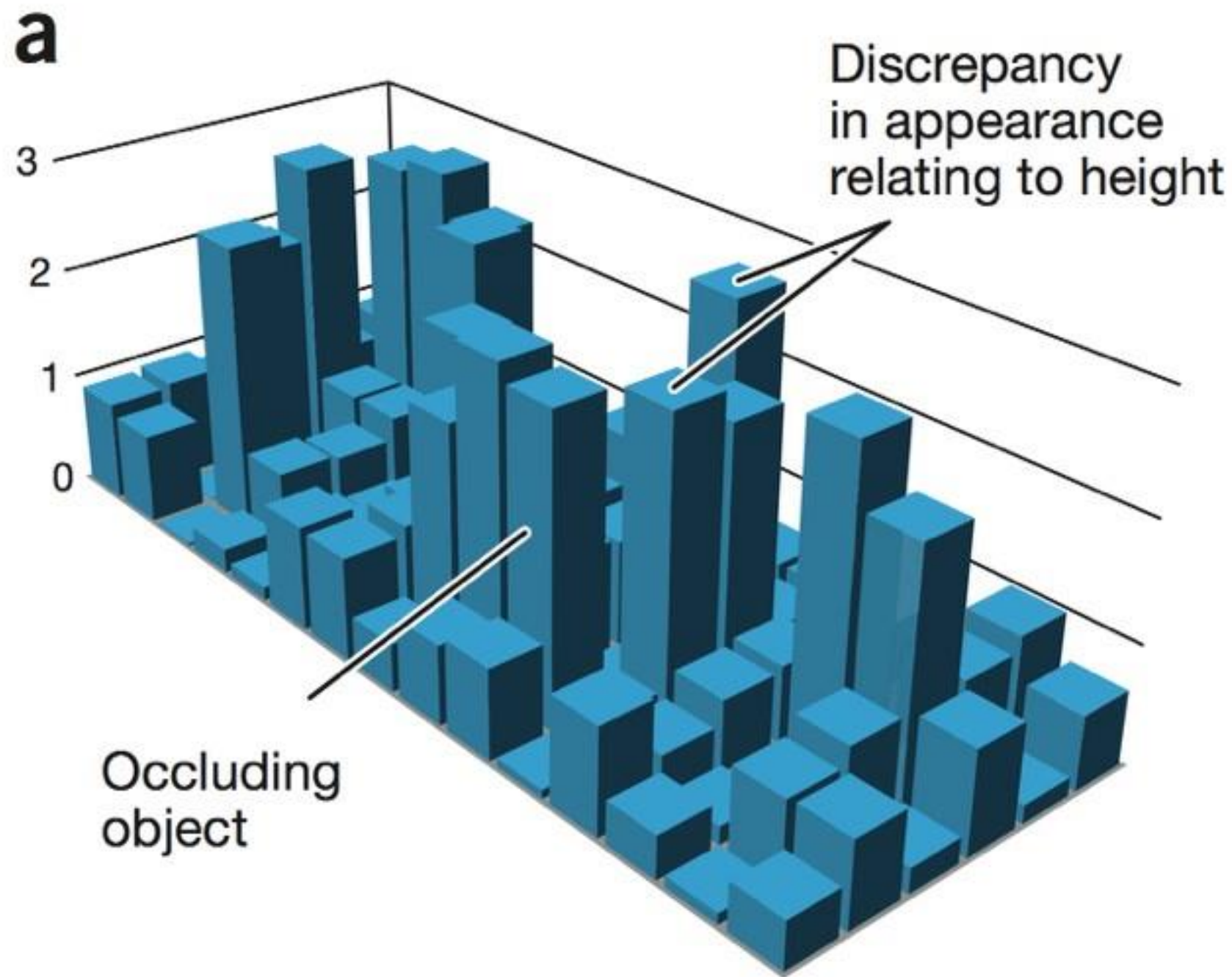


matplotlib gallery



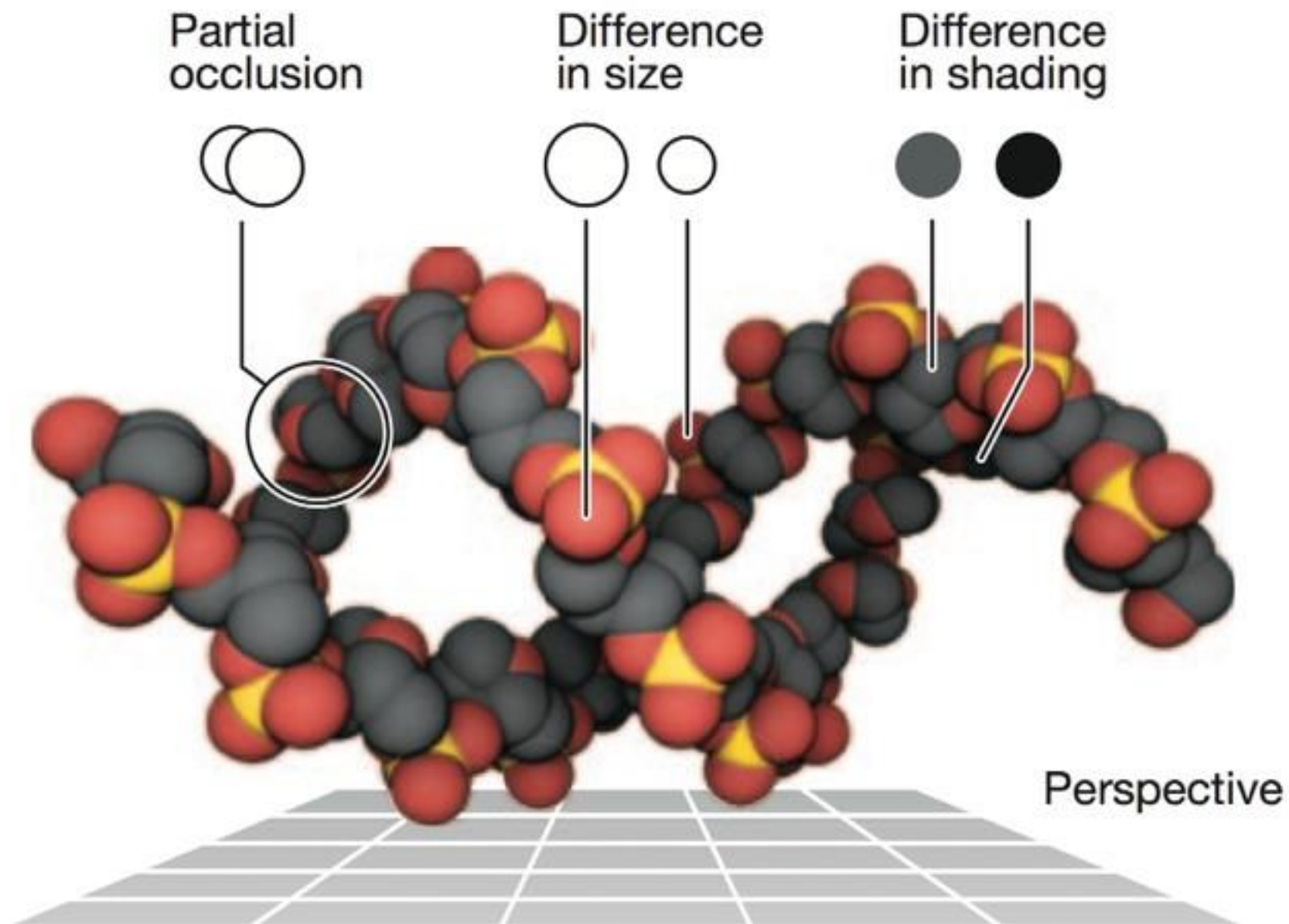
Excel Charts Blog

3D Bar Plot



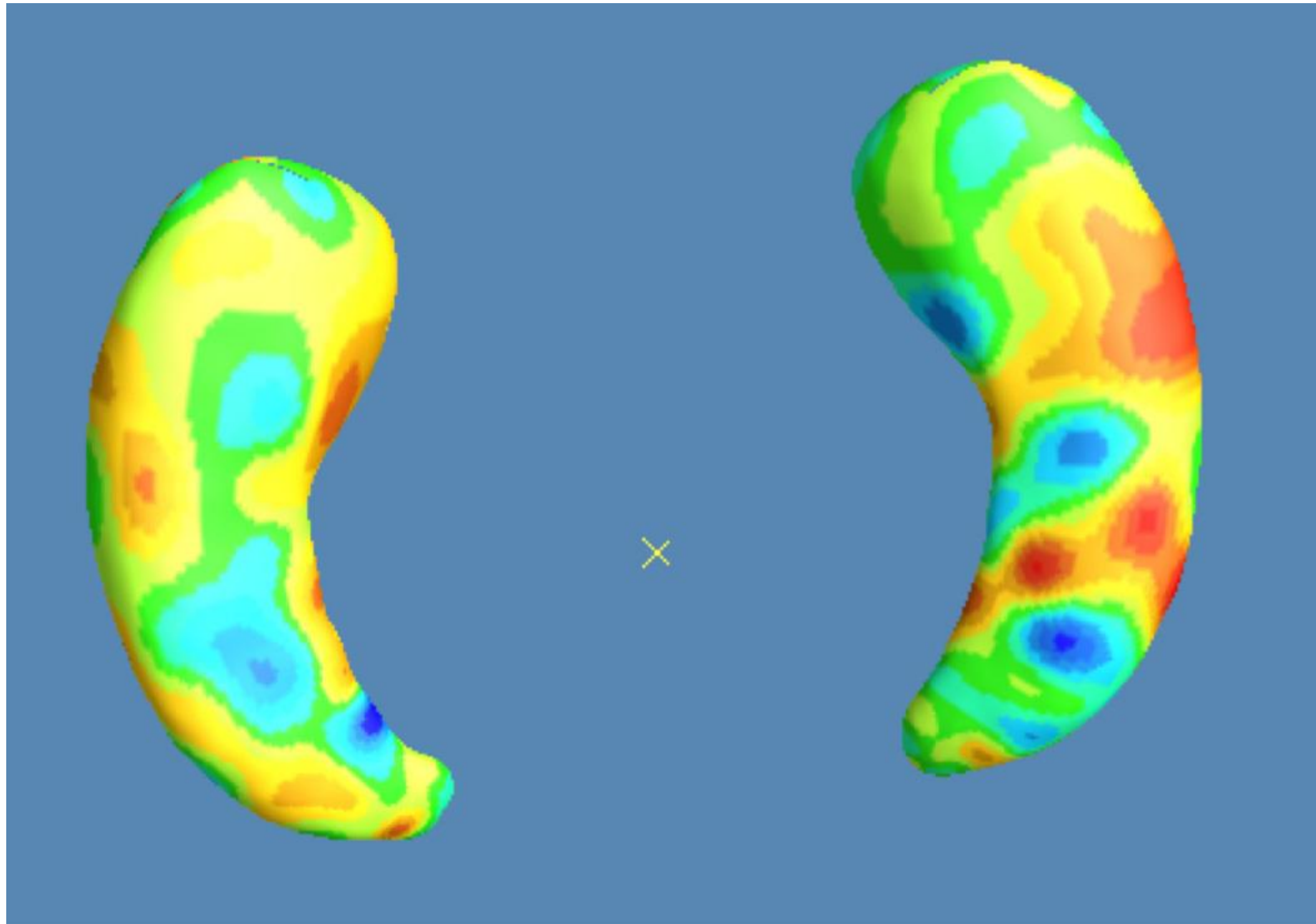
Which one is the tallest bar?
What is the pattern in the data?

Exception: 3D Phenomena



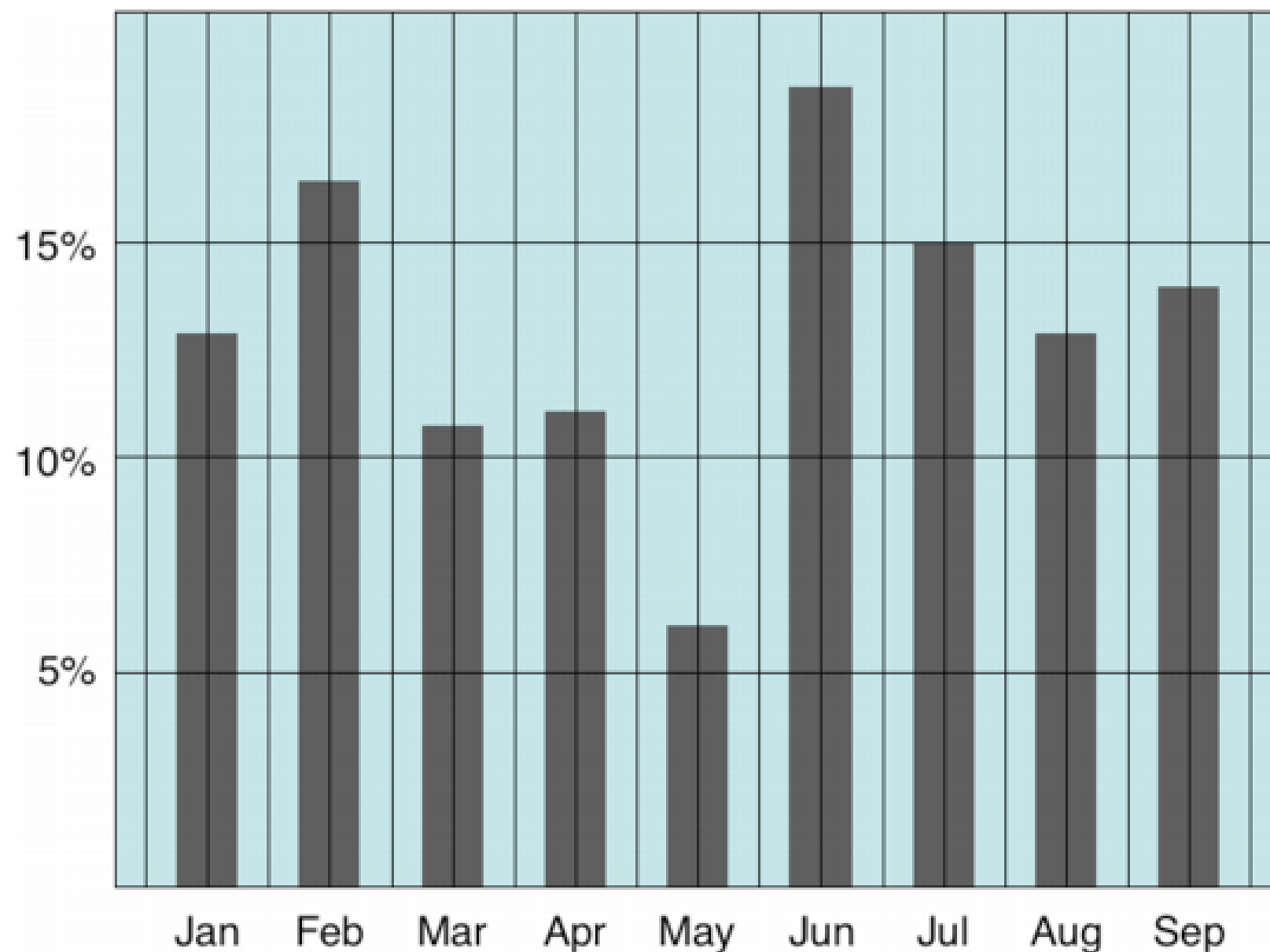
Depth cues enable us to perceive 2D images as 3D objects

Exception: 3D Phenomena



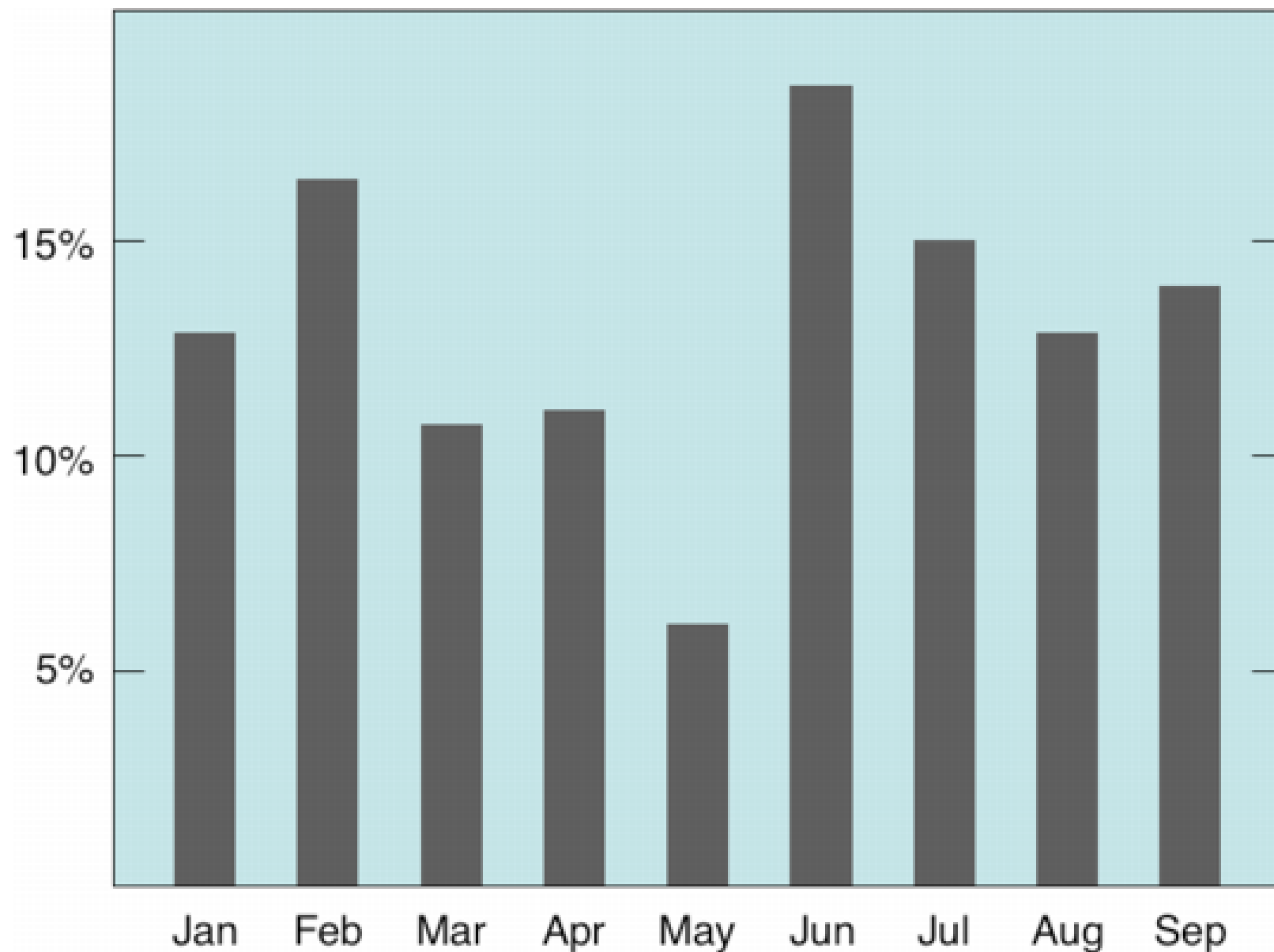
Avoid Chart Junk

Extraneous visual elements that distract from the message



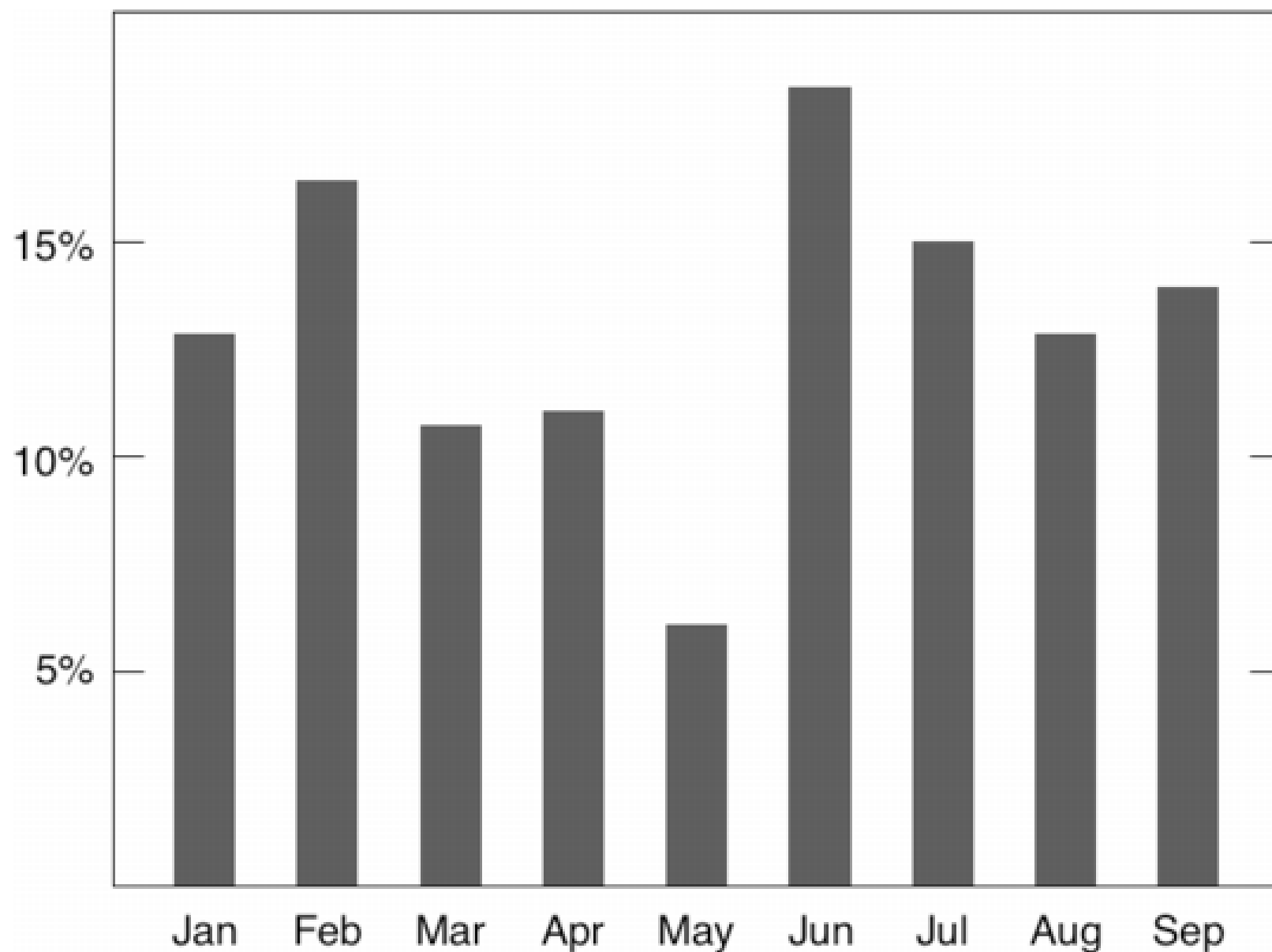
Avoid Chart Junk

Extraneous visual elements that distract from the message



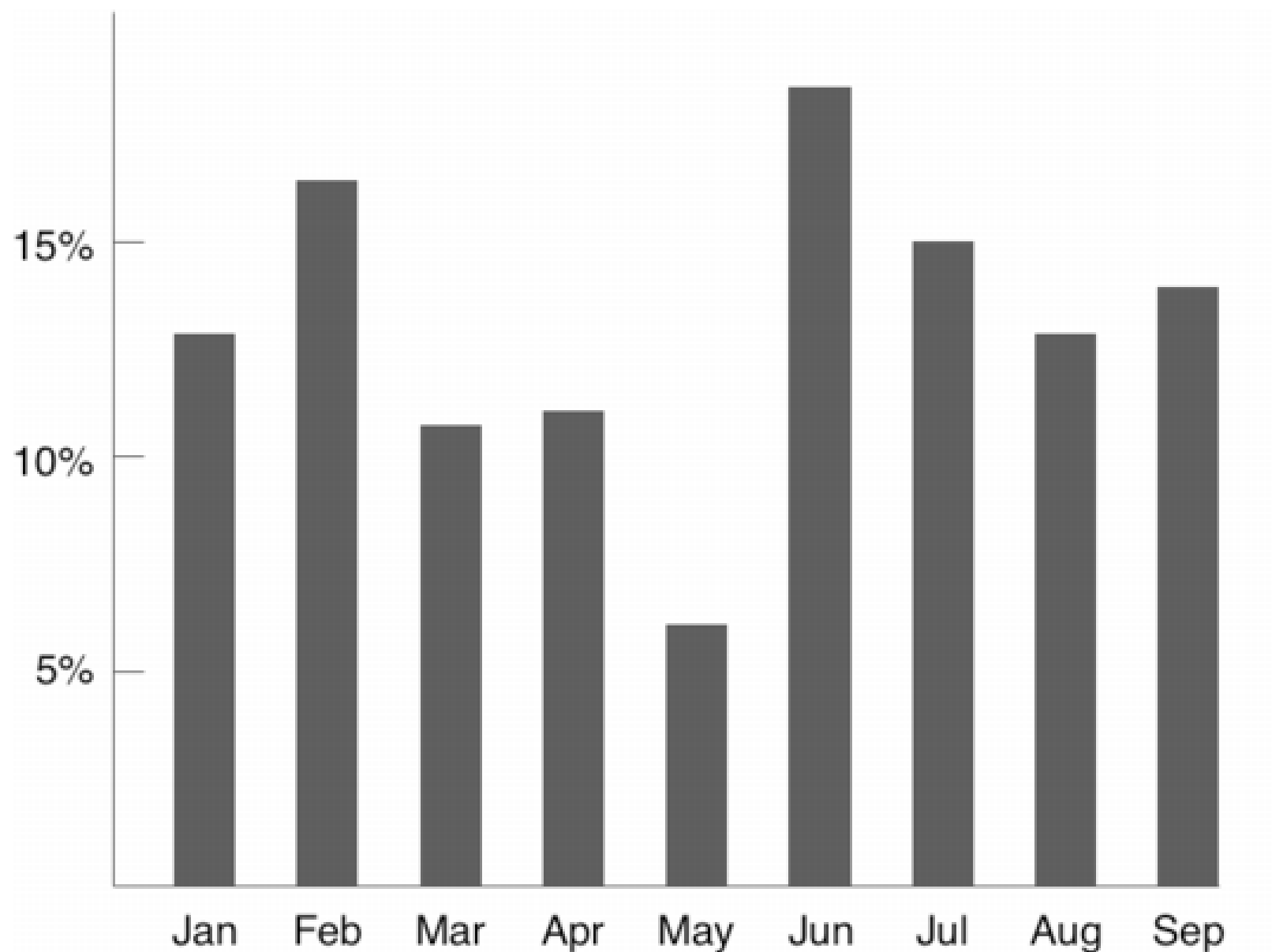
Avoid Chart Junk

Extraneous visual elements that distract from the message



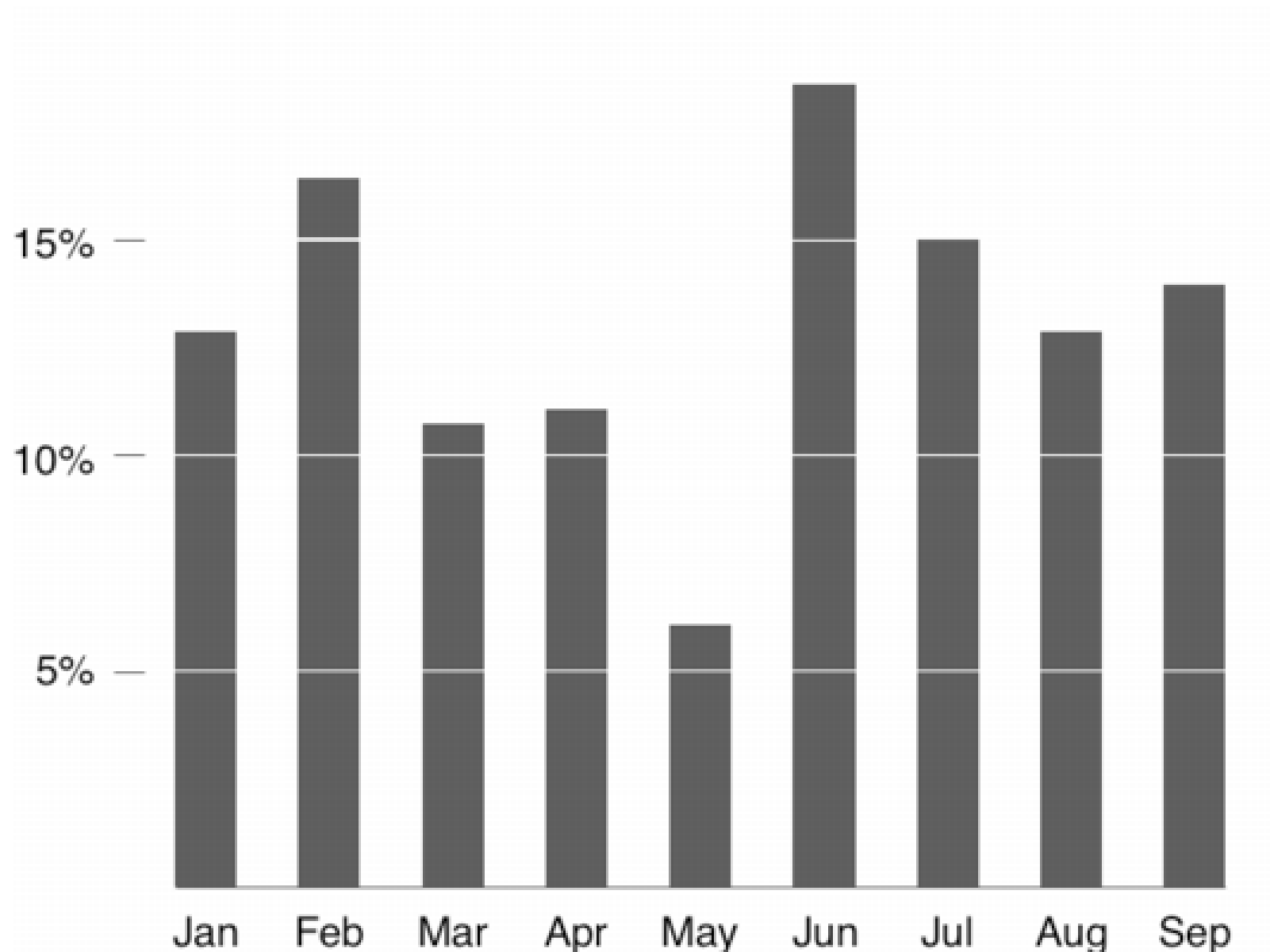
Avoid Chart Junk

Extraneous visual elements that distract from the message



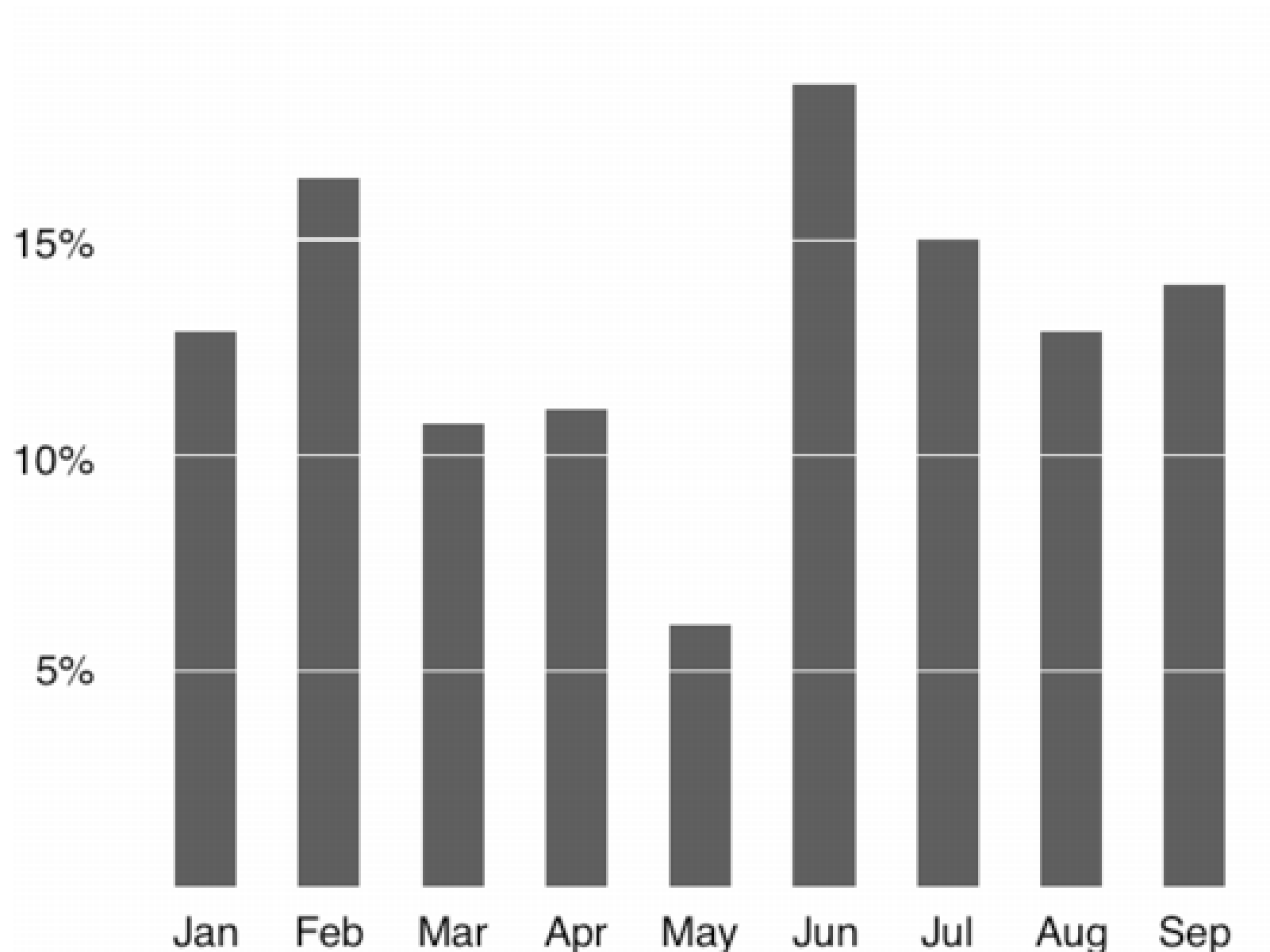
Avoid Chart Junk

Extraneous visual elements that distract from the message



Avoid Chart Junk

Extraneous visual elements that distract from the message



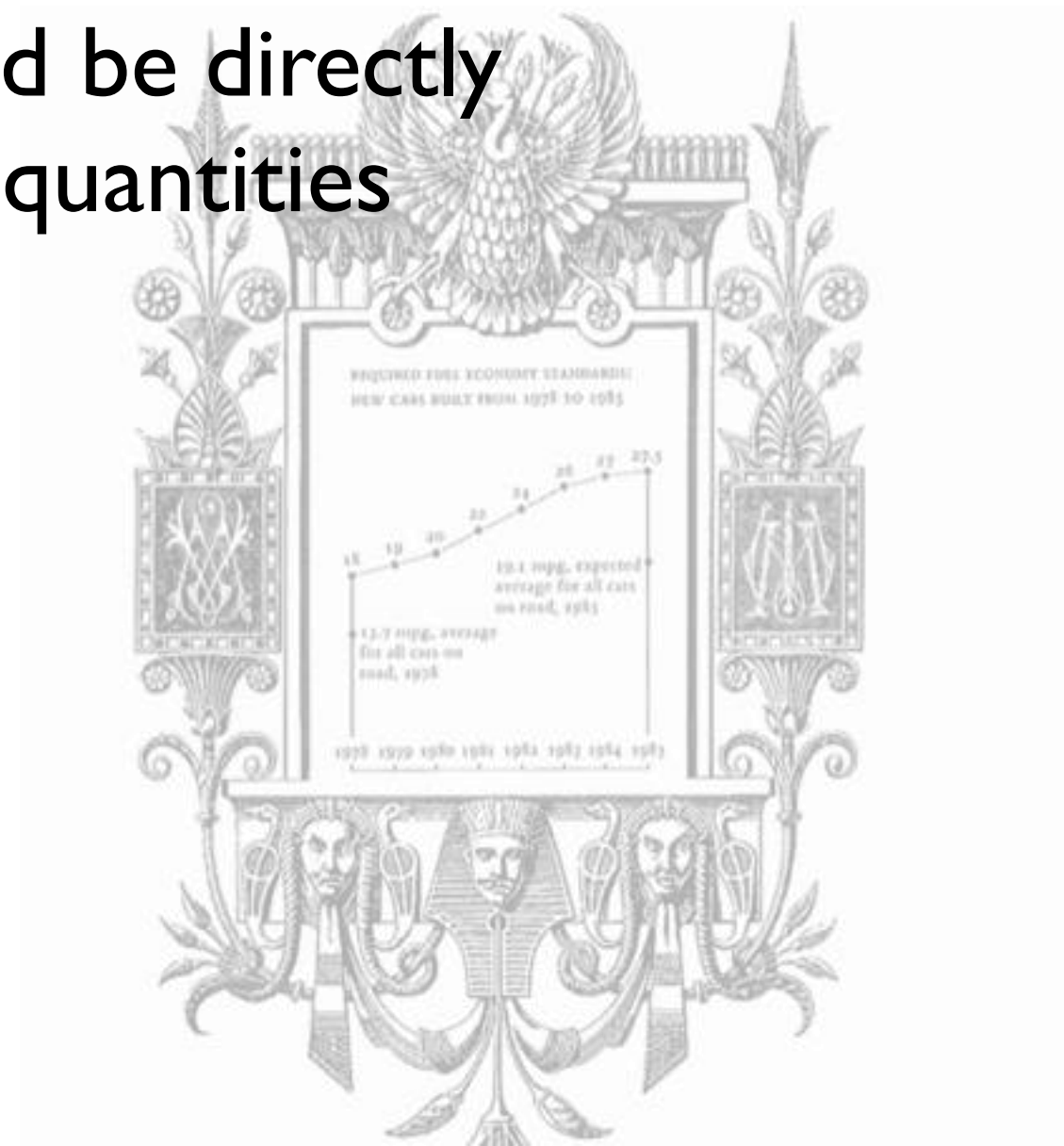
Tufte's Design Principles

Clear, detailed, and thorough labeling and appropriate scales

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

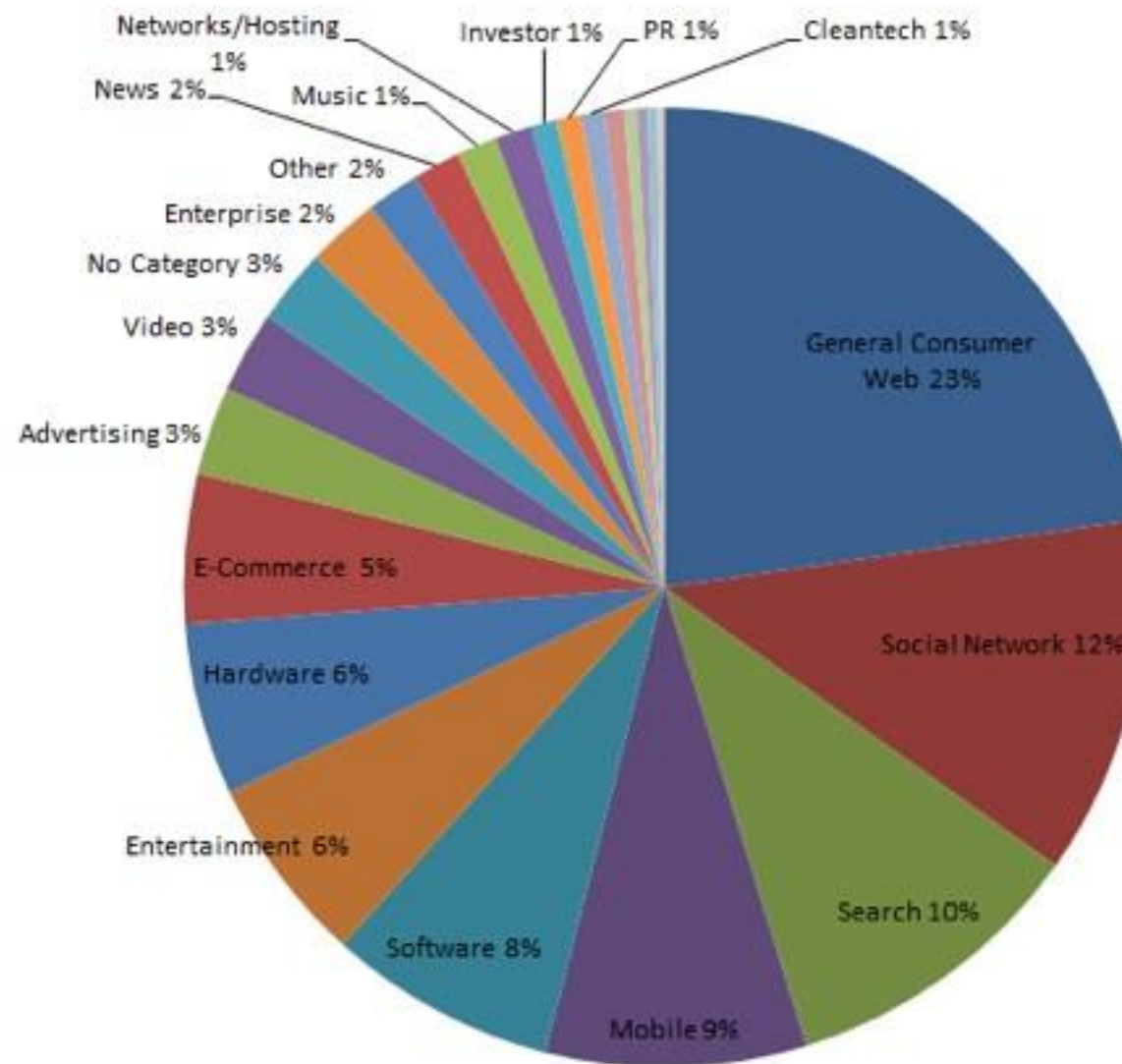
Maximize data-ink ratio

Avoid chart junk



Visualization Critique

Death to Pie Charts

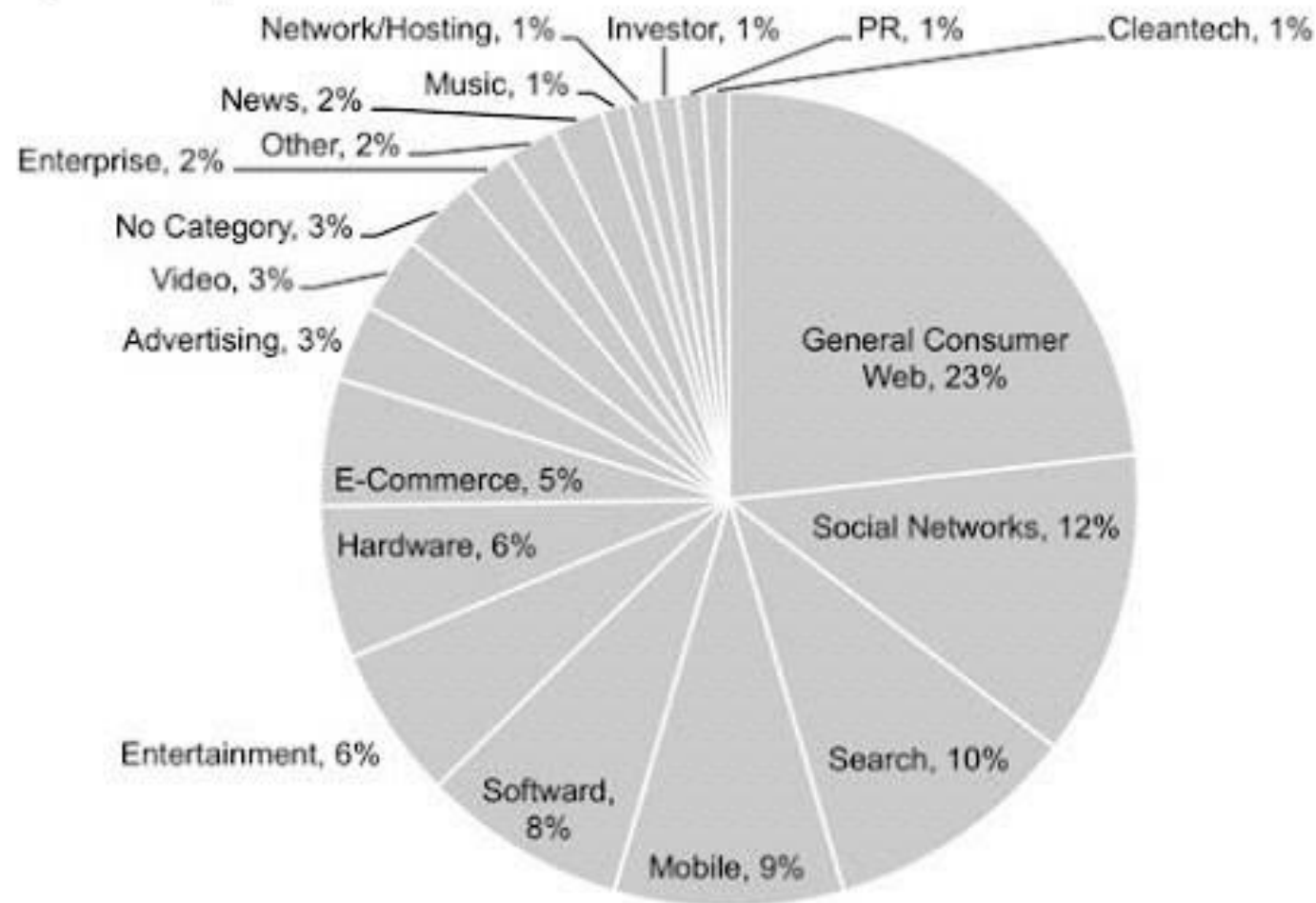


“I hate pie charts.
I mean, really hate them.”

Redesign

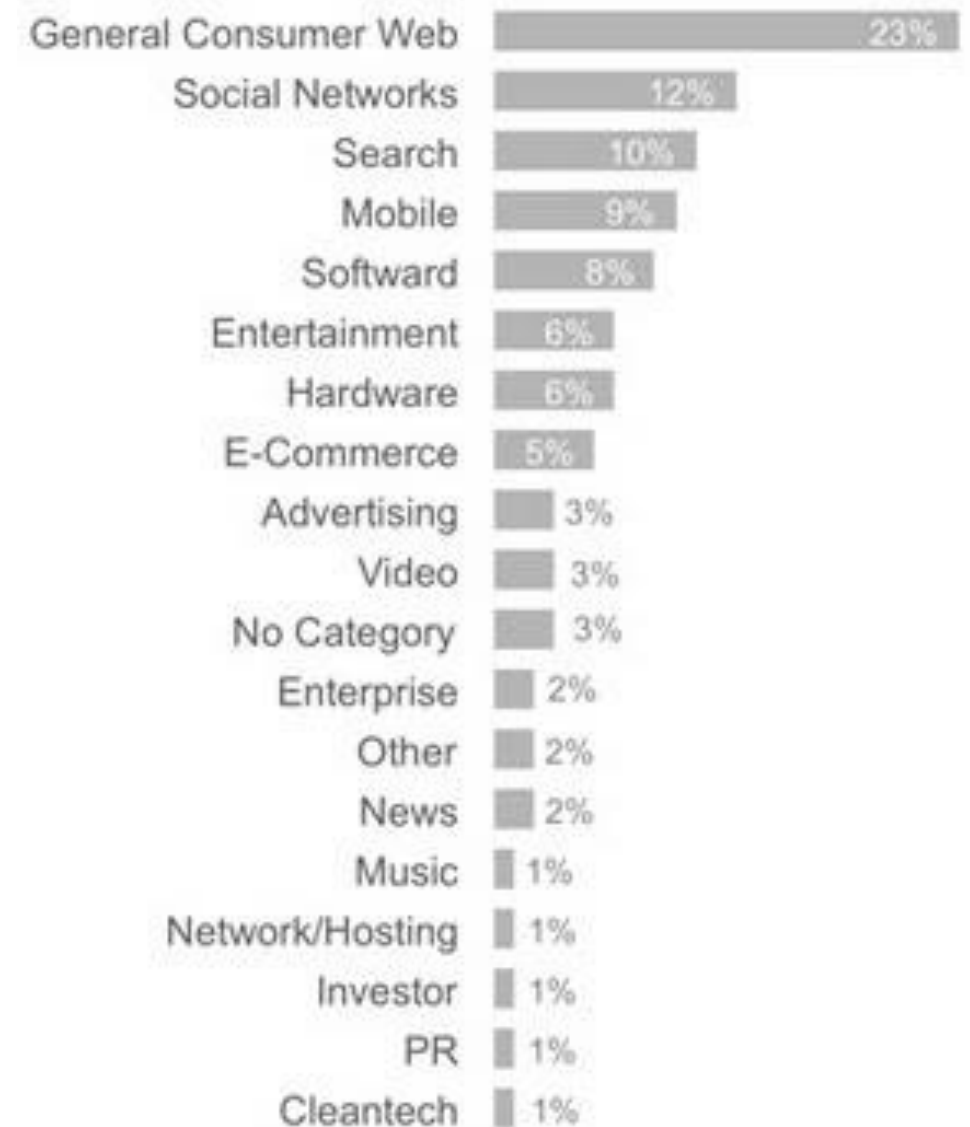
TechCrunch Coverage: 2005 - 2011

A slightly better pie?

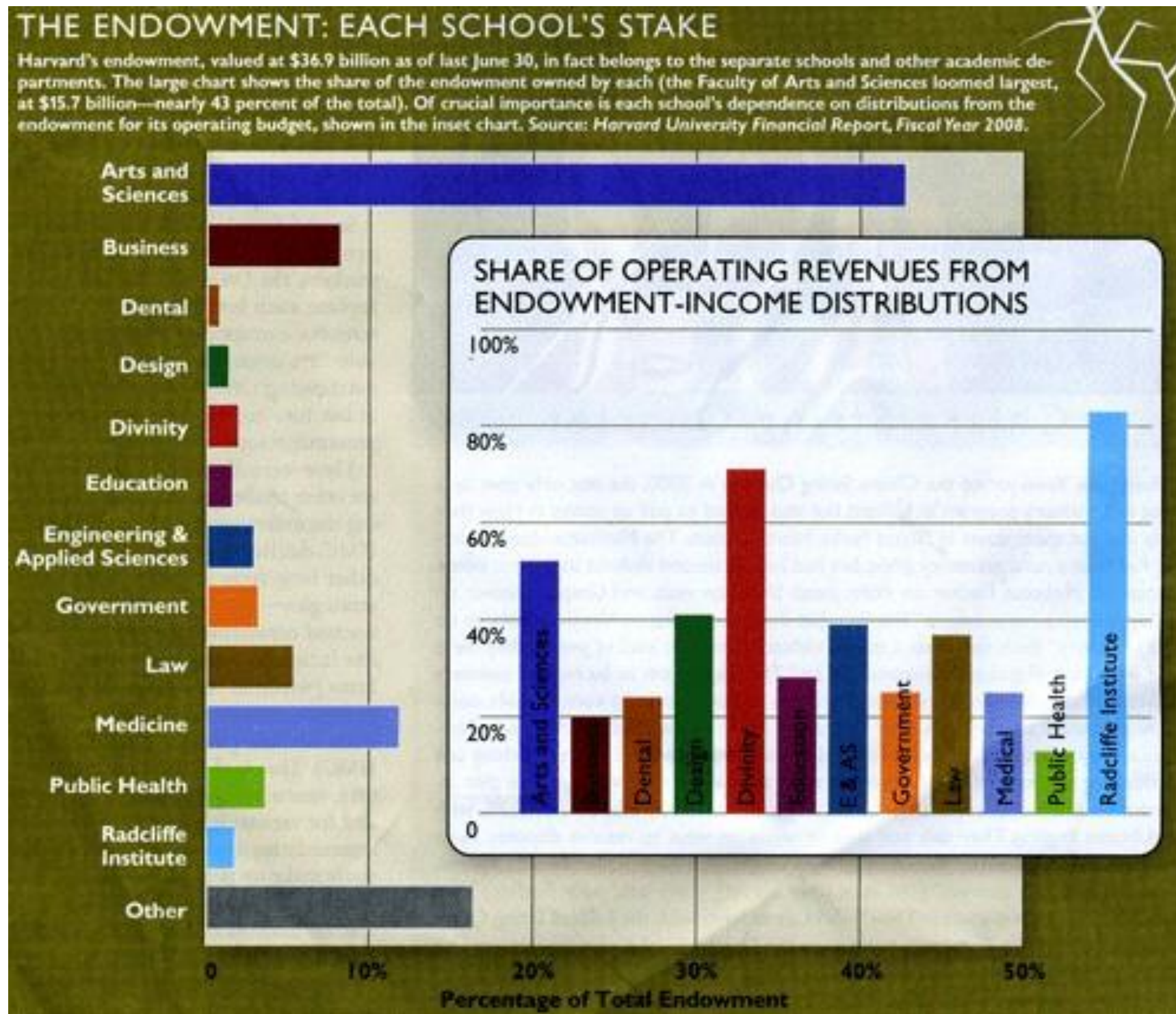


TechCrunch Coverage: 2005 - 2011

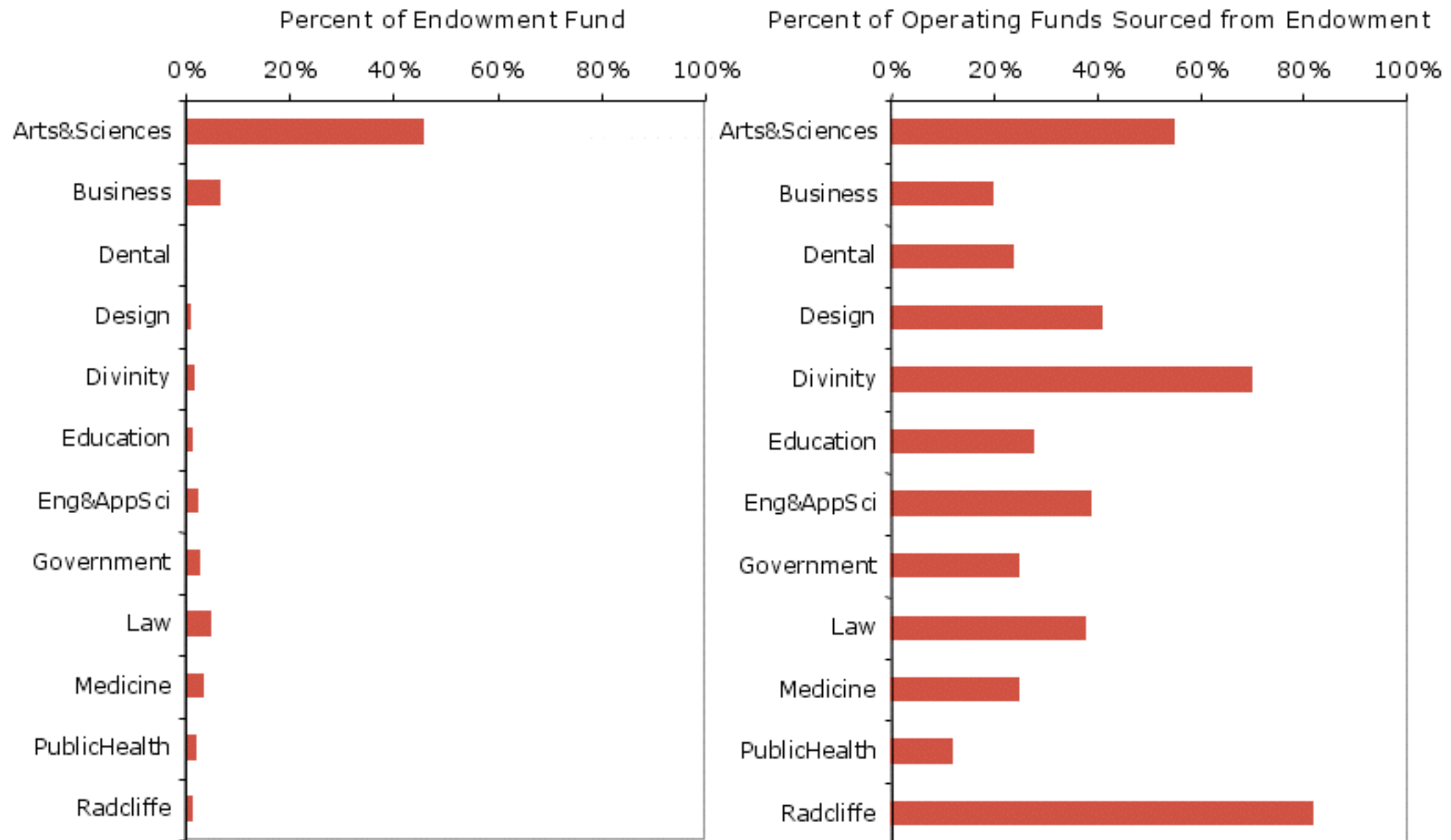
Bars are best!



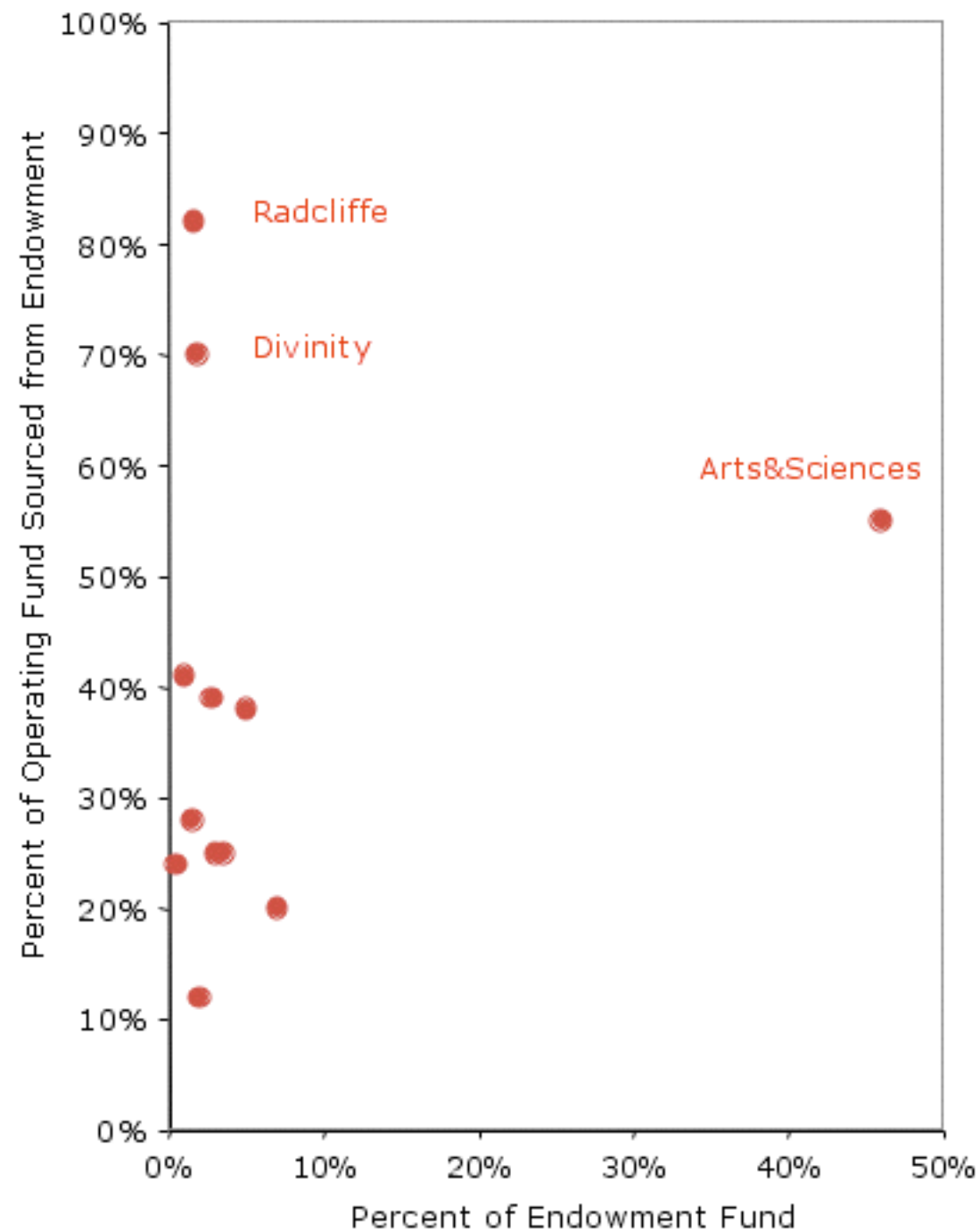
Harvard's Troubles



Redesign I



Redesign II



Redesign III

