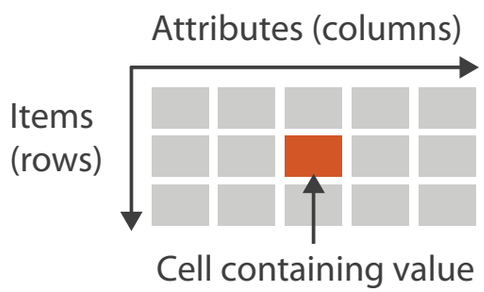
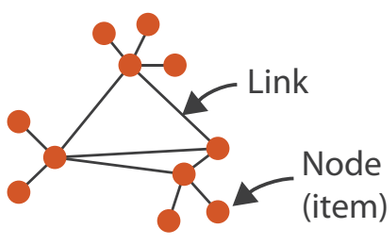
➔ Dataset Types

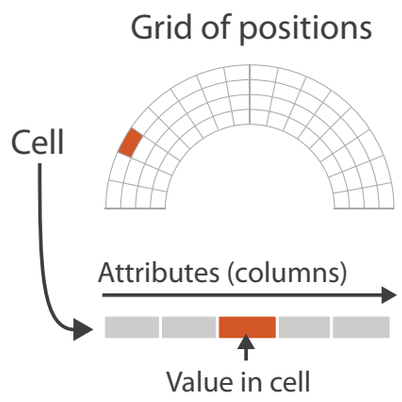
➔ Tables



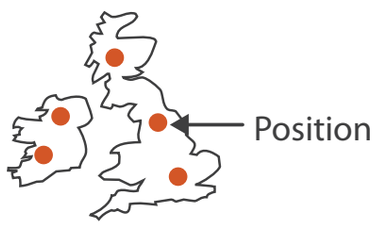
➔ Networks



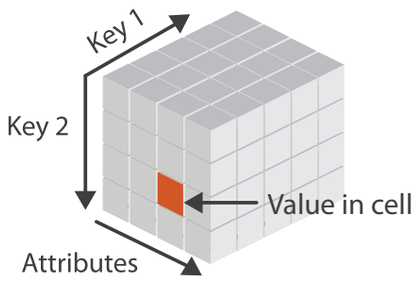
➔ Fields (Continuous)



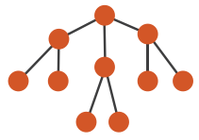
➔ Geometry (Spatial)



➔ *Multidimensional Table*

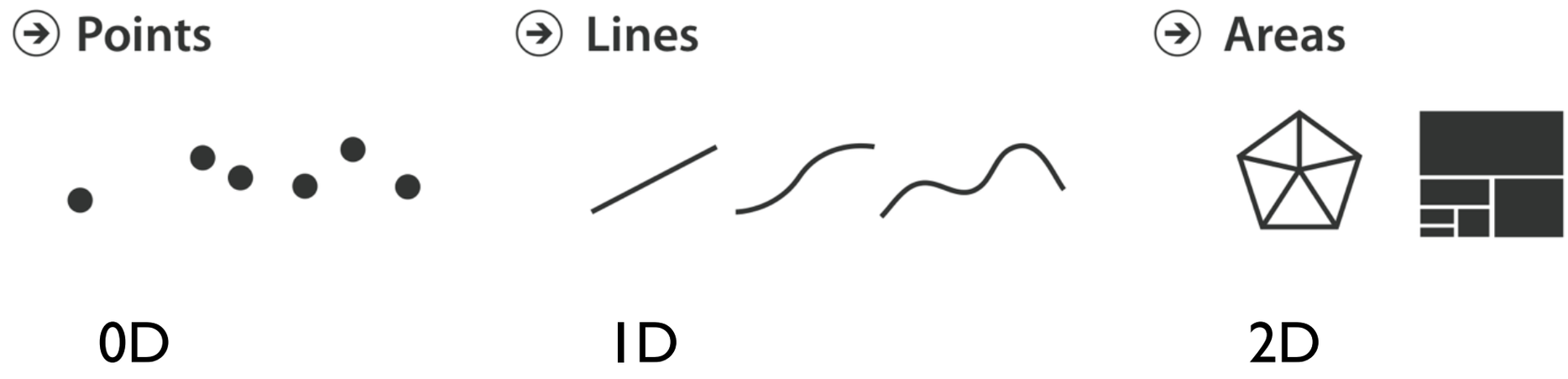


➔ *Trees*



Marques pour items

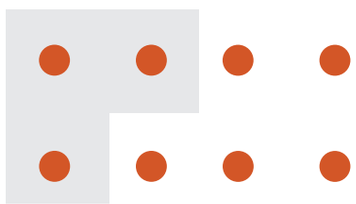
Elements géométriques basiques



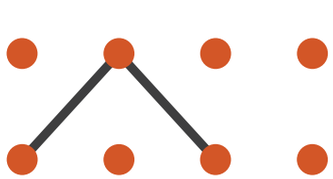
Marques 3D: Volume (rarement utilisé)

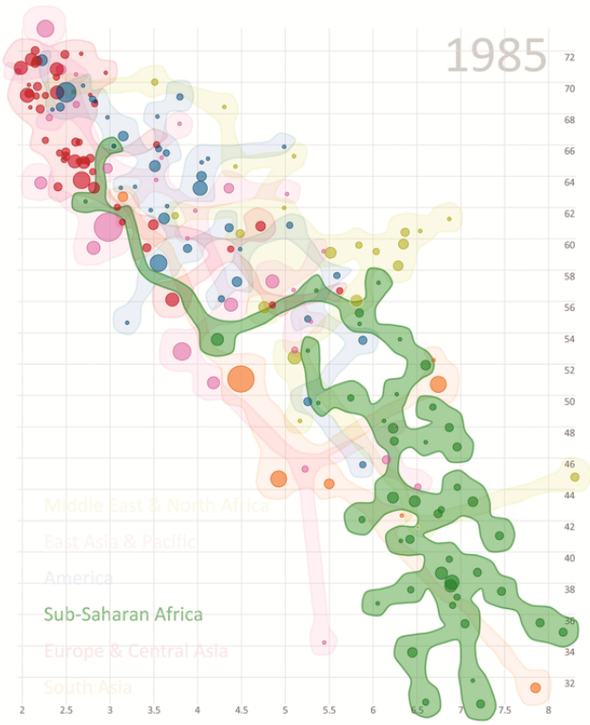
Marques pour liens

Containment



Connection

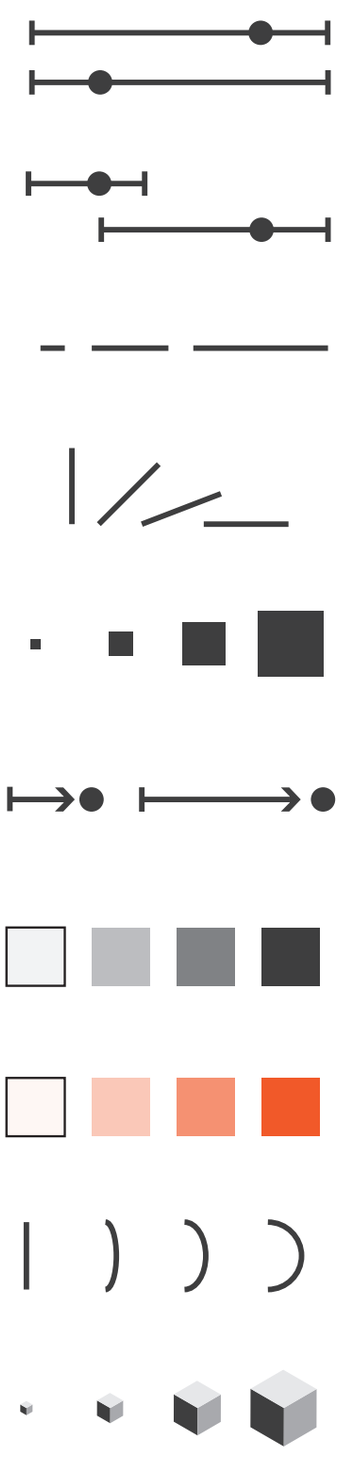




Channels: Expressiveness Types and Effectiveness Ranks

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

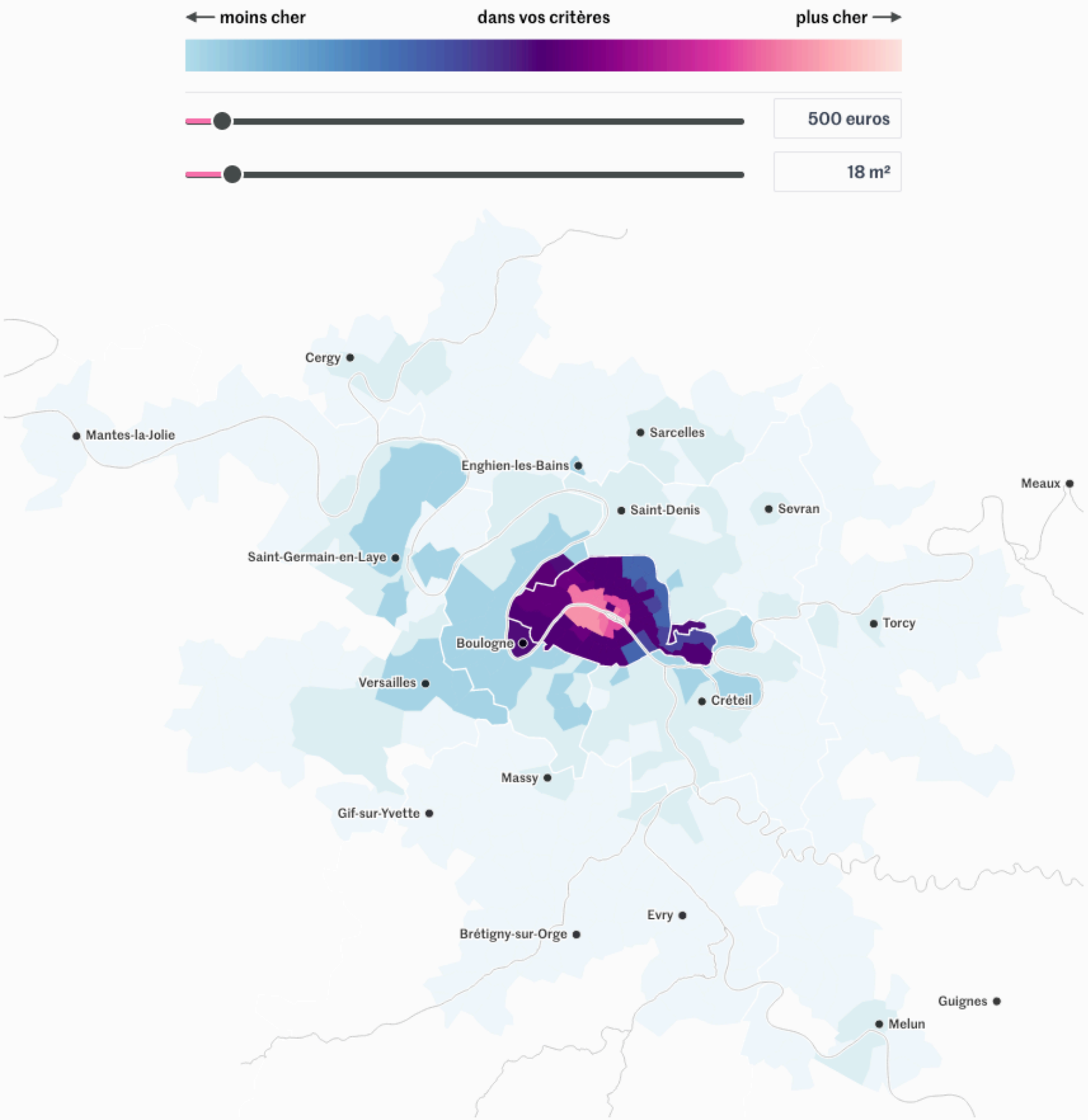
▼ Least

Same

Same

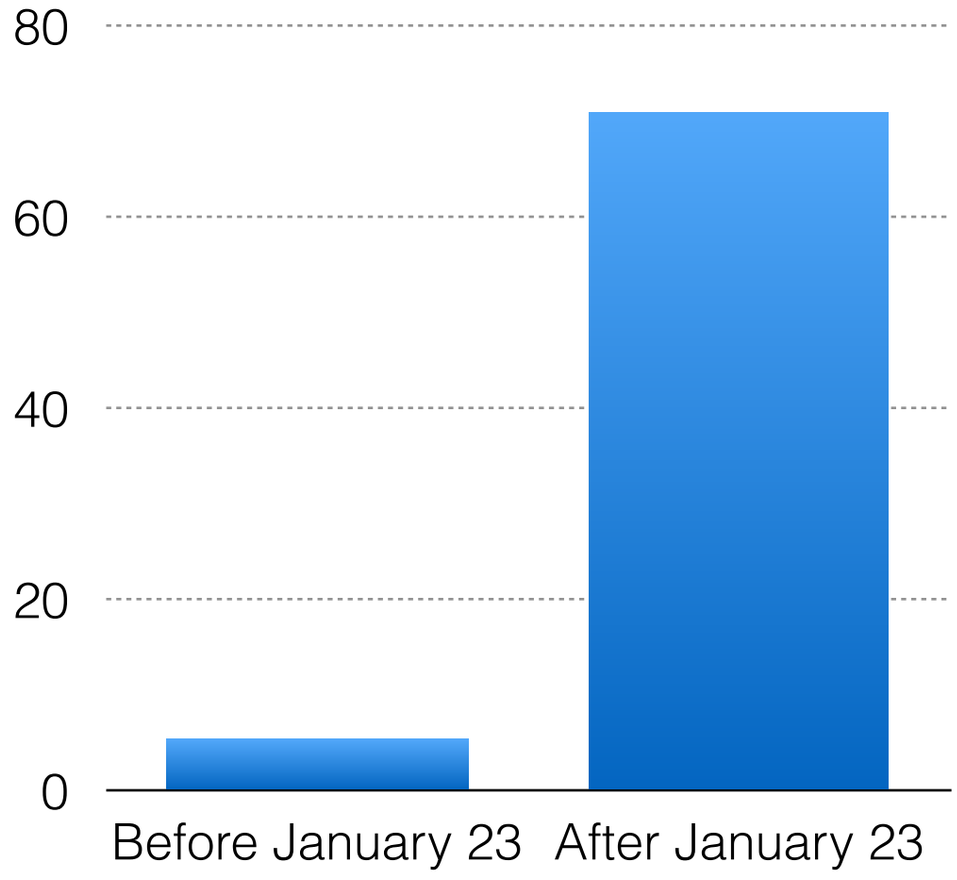
Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape


moins cher
dans vos critères
plus cher
500 euros
18 m²
Cergy
Mantes-la-Jolie
Sarcelles
Enghien-les-Bains
Meaux
Saint-Denis
Sevran
Saint-Germain-en-Laye
Torcy
Boulogne
Versailles
Créteil
Massy
Gif-sur-Yvette
Evry
Brétigny-sur-Orge
Guignes
Melun


STORM TEAM 5
SEASONAL SNOWFALL
Entire Season Average: 43.1"
Season to Date: 76.5"
5.5"
71.0"
BEFORE
JANUARY 23
AFTER
JANUARY 23

Que pensez-vous de ce bar-chart ?



Utiliser des canaux proportionnels aux données

Position

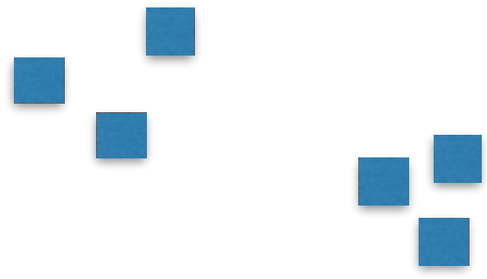
Variable visuelle la plus puissante

Tous types de données

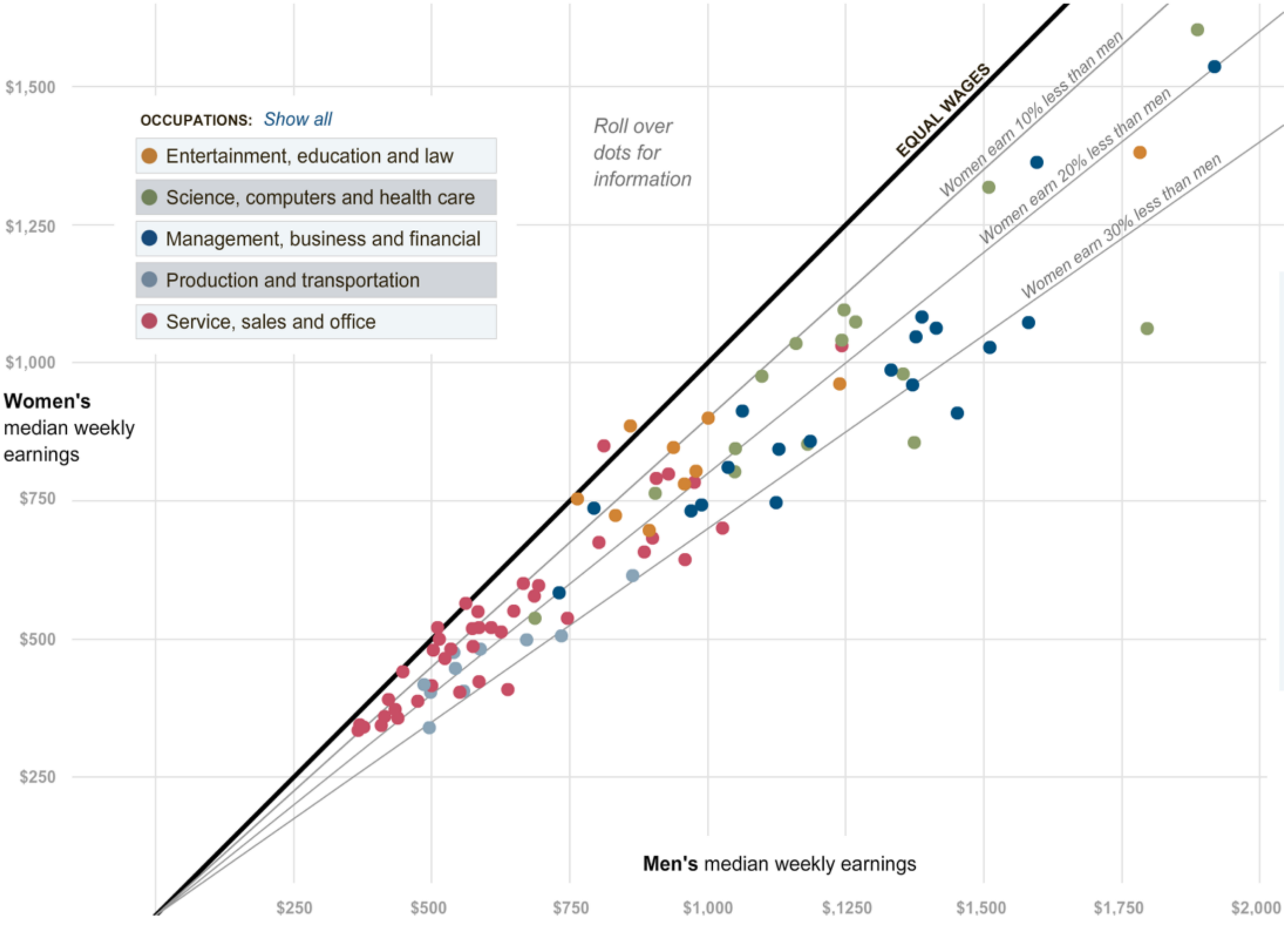
Problèmes:

parfois non disponible (données spatiales)

encombrement



Exemple: Nuage de points (scatterplot)



Longueur et taille

Bien pour 1D, OK pour 2D, mauvais pour 3D

Facile de voir si 1 est plus grand

Aligner les barres pour utiliser la position

Taille en 2D: buble-plot

Four Ways to Slice Obama's 2013 Budget Proposal

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

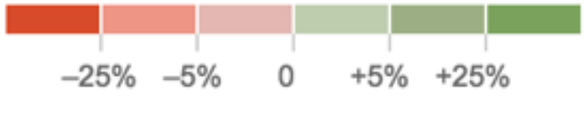
All Spending | Types of Spending | Changes | Department Totals

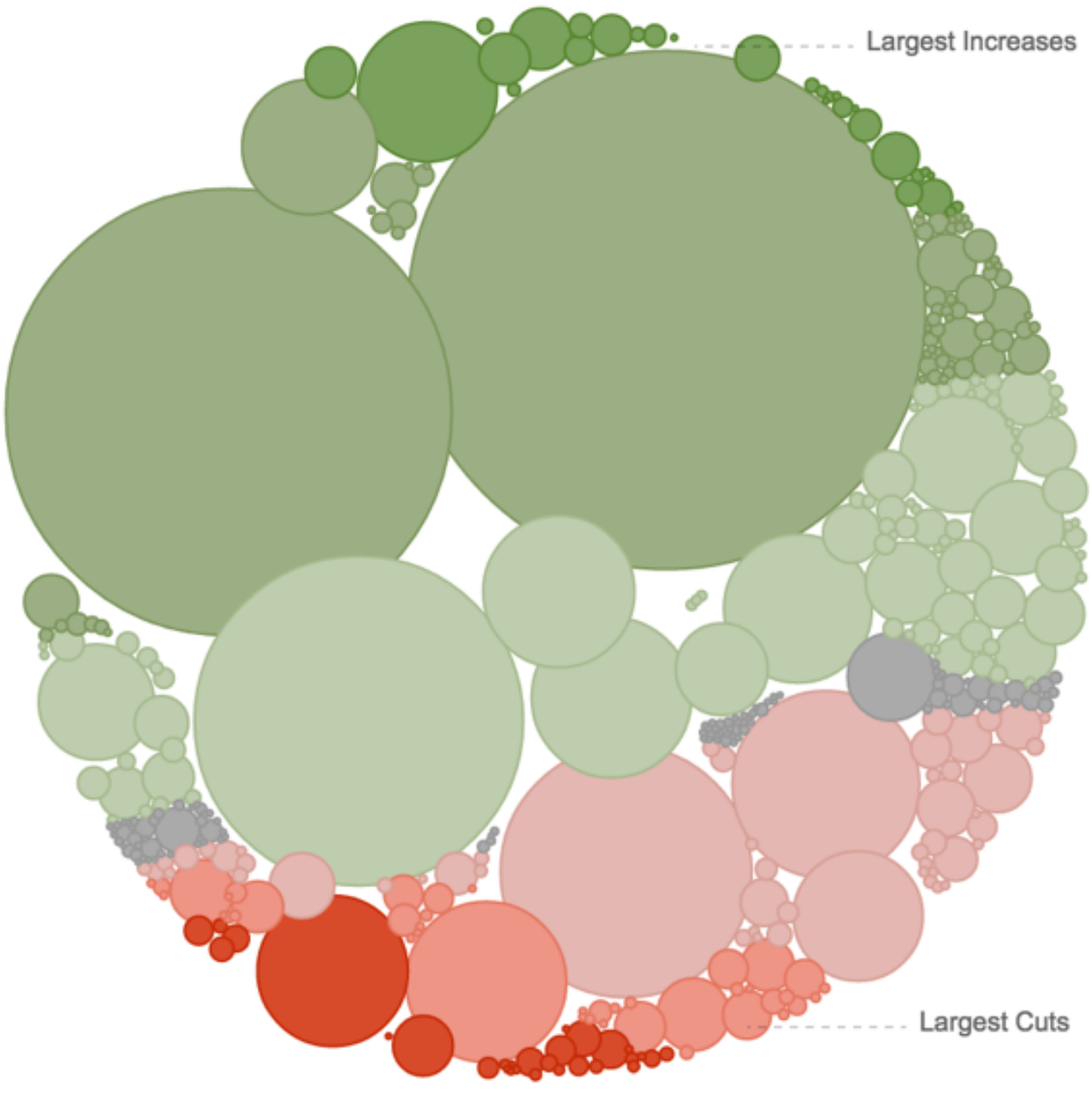
How $3.7 Trillion Is Spent

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.



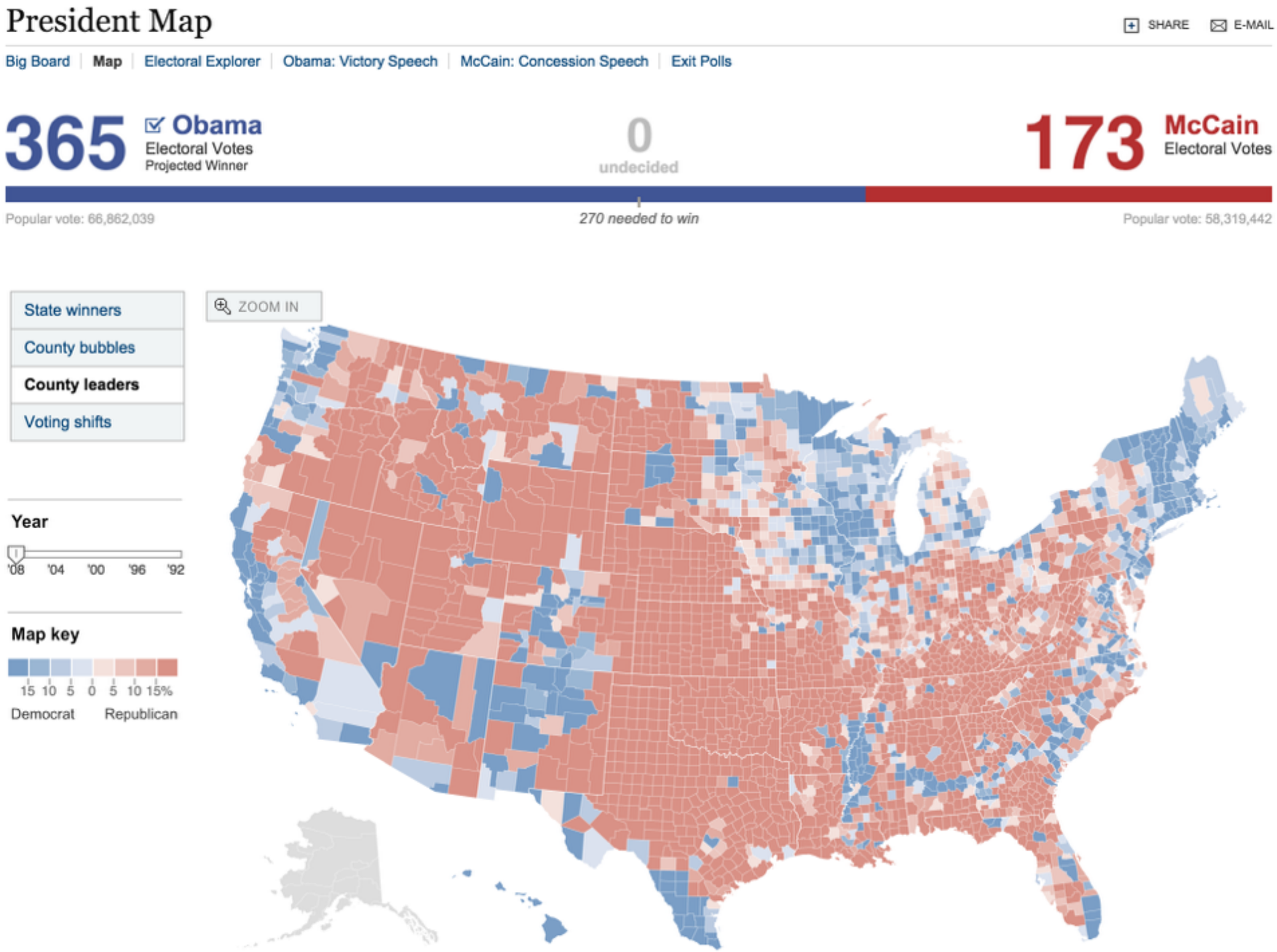


Luminosité/saturation

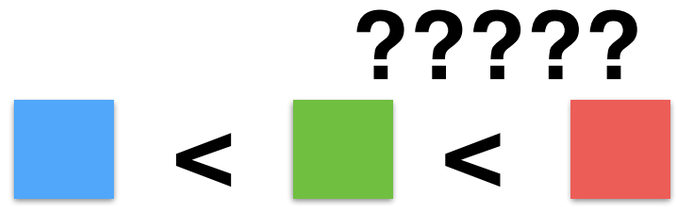
OK pour données quantitatives
quand longueur et taille utilisées

pas beaucoup de variations
reconnaissables

Exemple: echelle de saturation divergente



Couleur



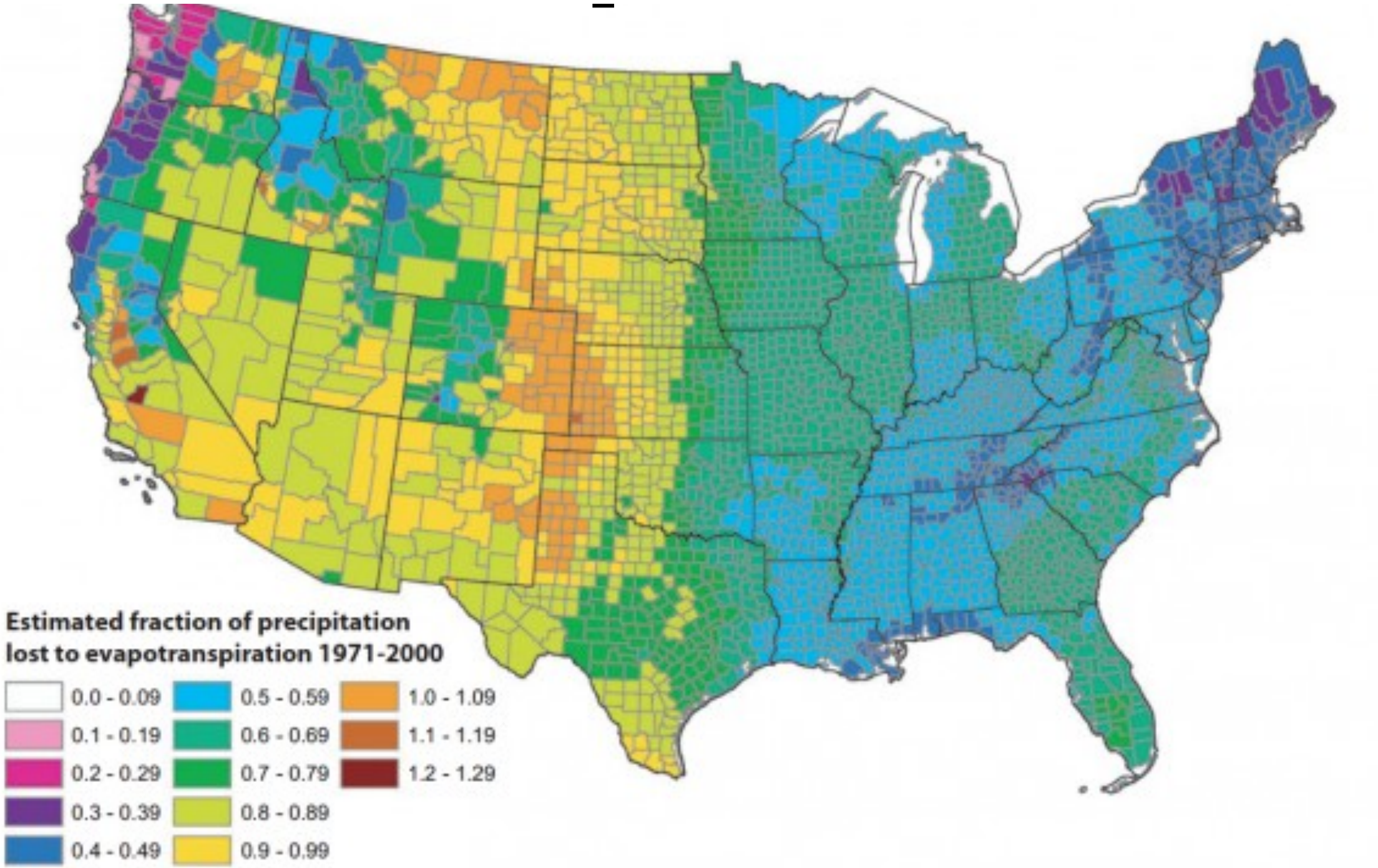
Bien pour variables qualitatives

7-10 classes max

Pas bon pour variables quantitatives

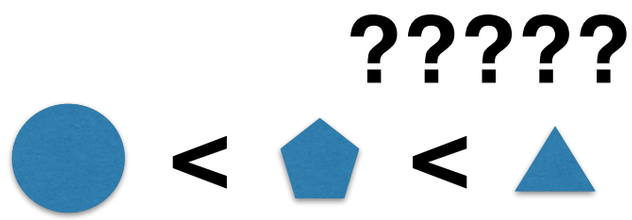
Attention, beaucoup de problèmes possibles

A éviter


Estimated fraction of precipitation
lost to evapotranspiration 1971-2000
0.0 - 0.09
0.1 - 0.19
0.2 - 0.29
0.3 - 0.39
0.4 - 0.49
0.5 - 0.59
0.6 - 0.69
0.7 - 0.79
0.8 - 0.89
0.9 - 0.99
1.0 - 1.09
1.1 - 1.19
1.2 - 1.29

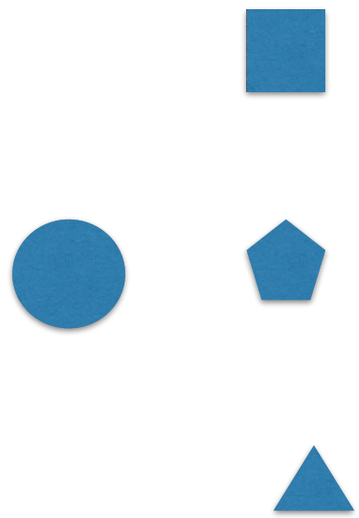
Forme

Très bien pour reconnaître de nombreuses classes

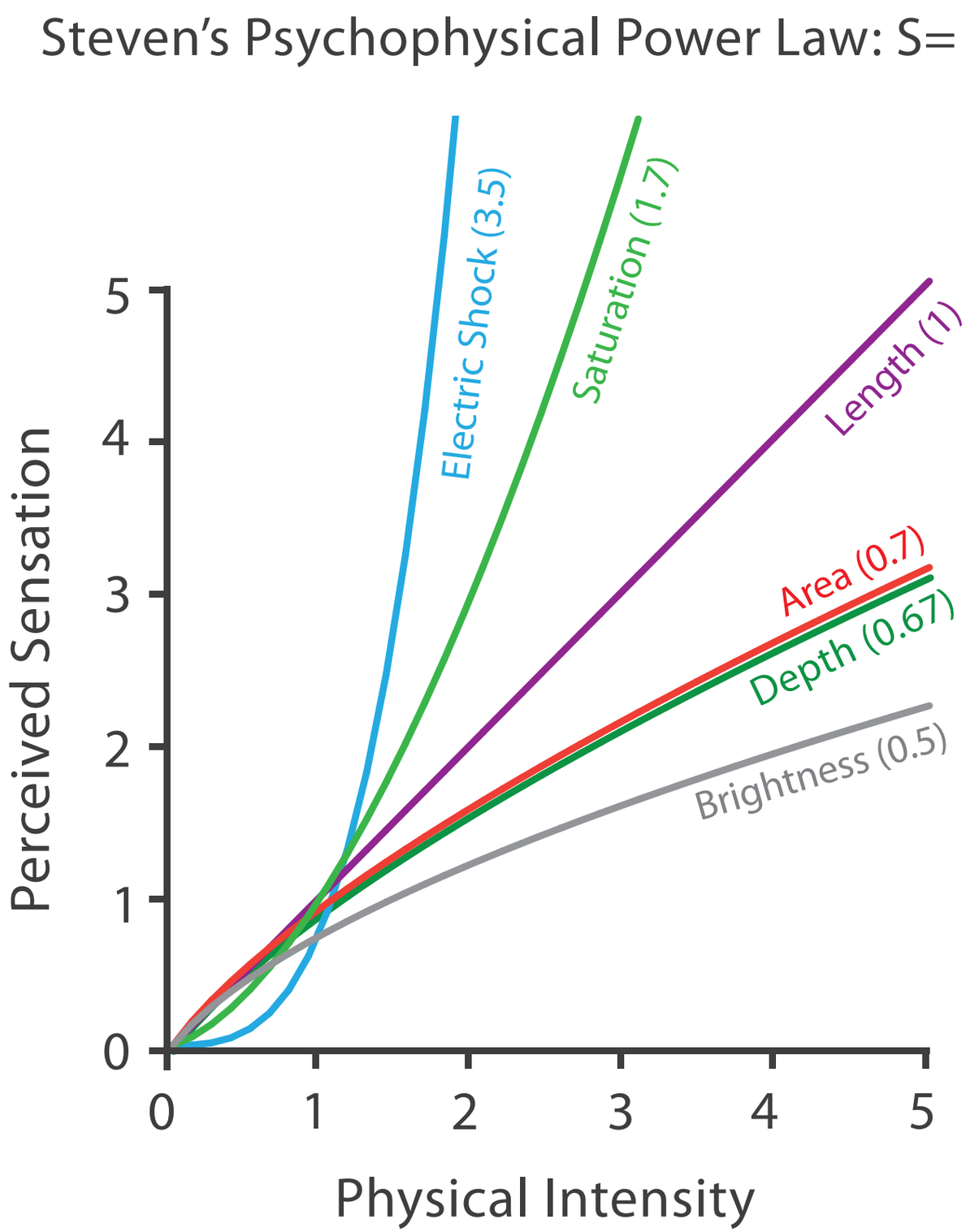


Pas de groupage

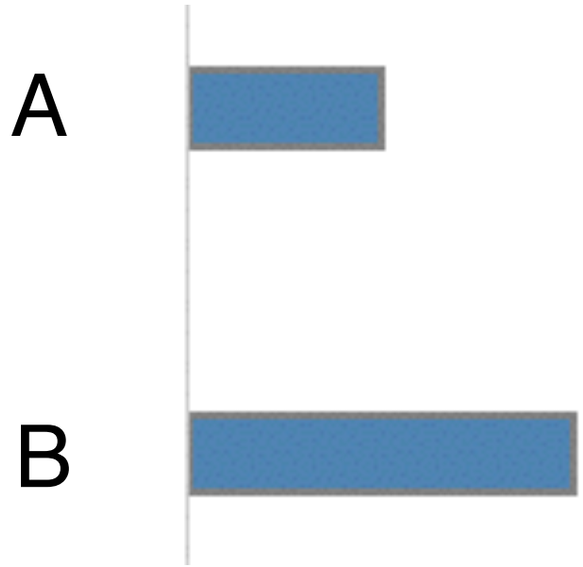
Pas d’ordre



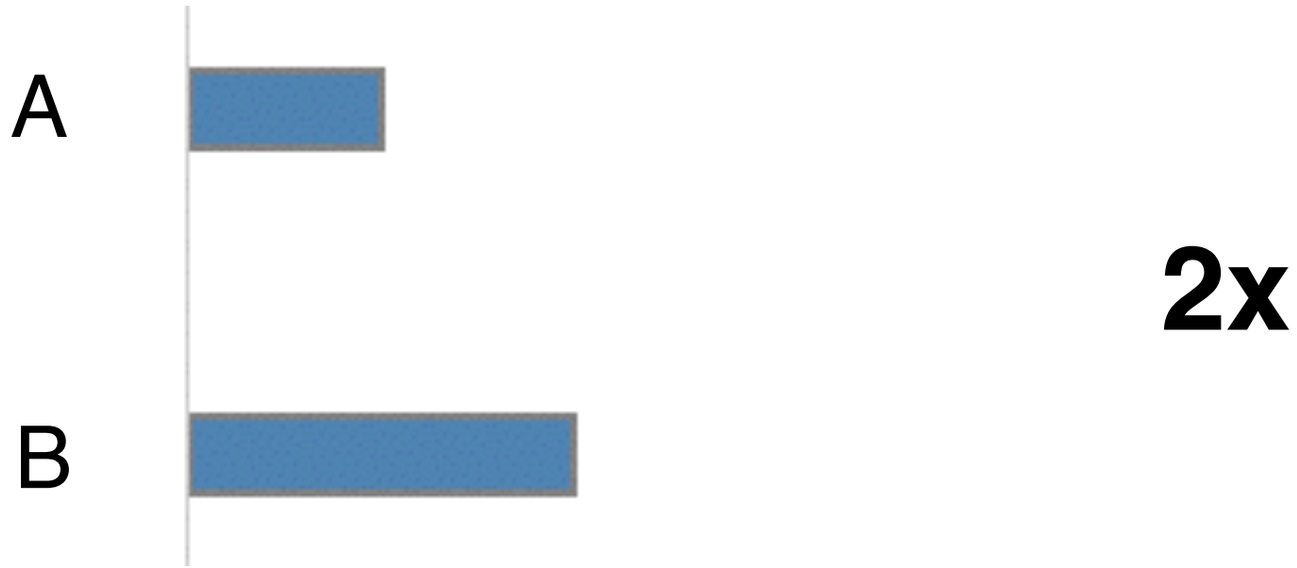
Les canaux ne sont pas égaux



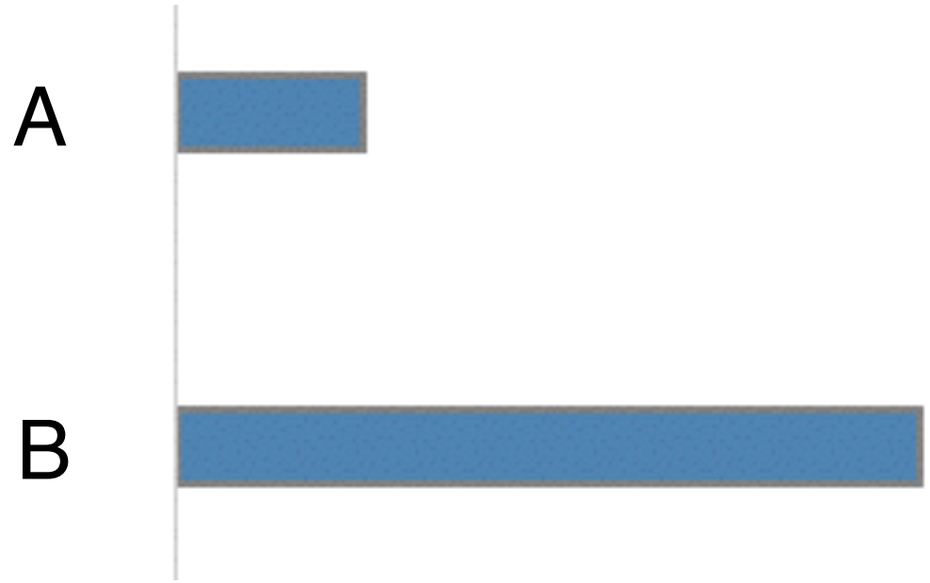
Combien de fois plus long ?



Combien de fois plus long ?

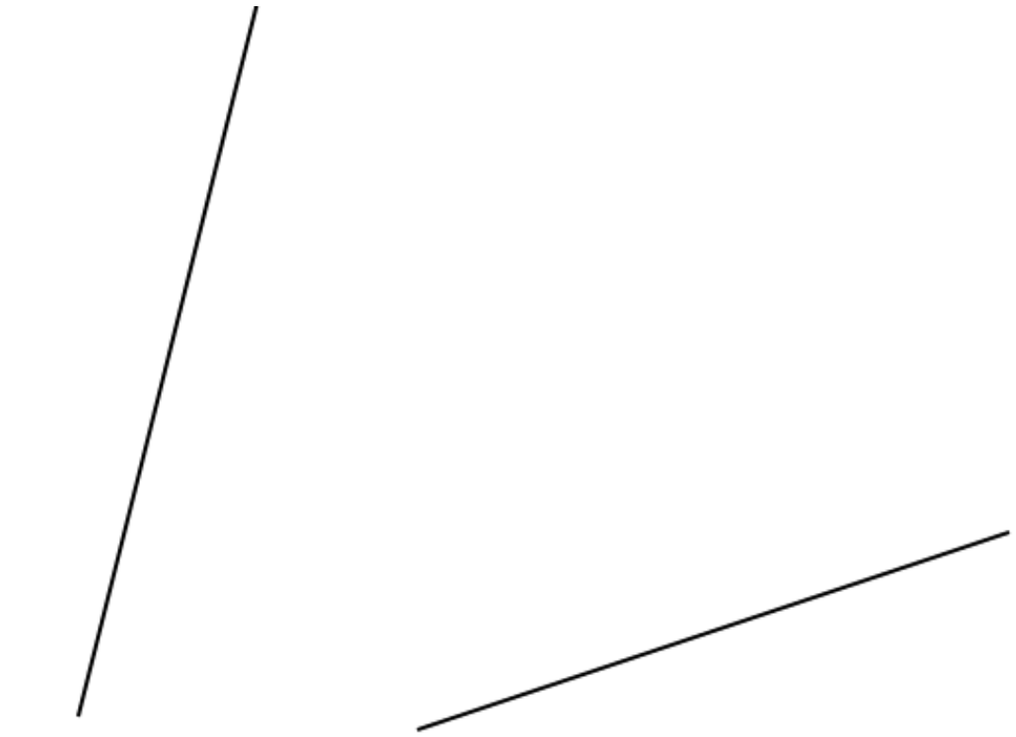


Combien de fois plus long ?



Combien de fois plus long ?

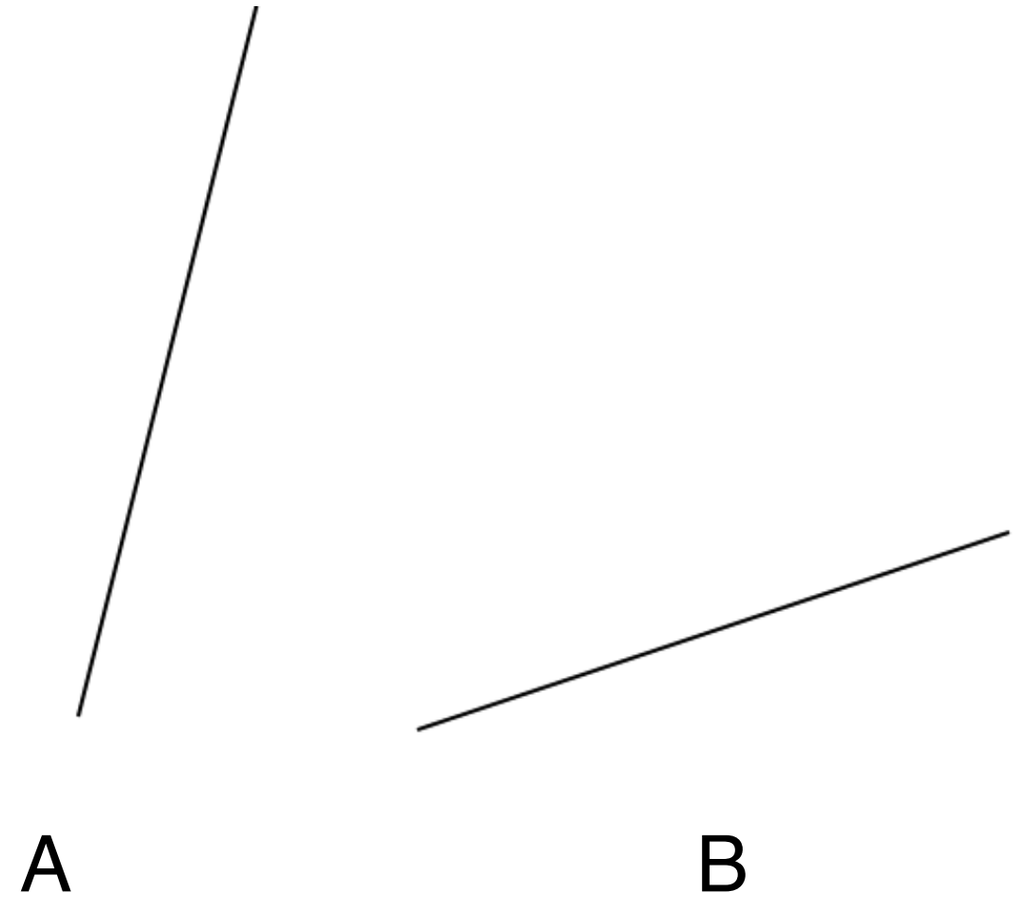
Combien de fois plus incliné ?



A

B

Combien de fois plus incliné ?

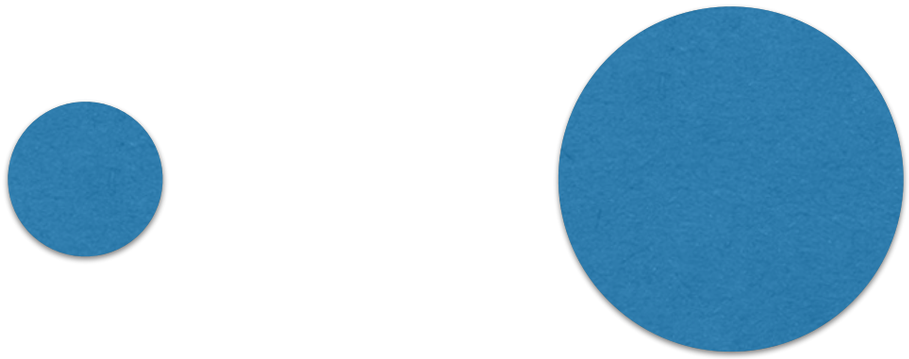


A

B

~4x

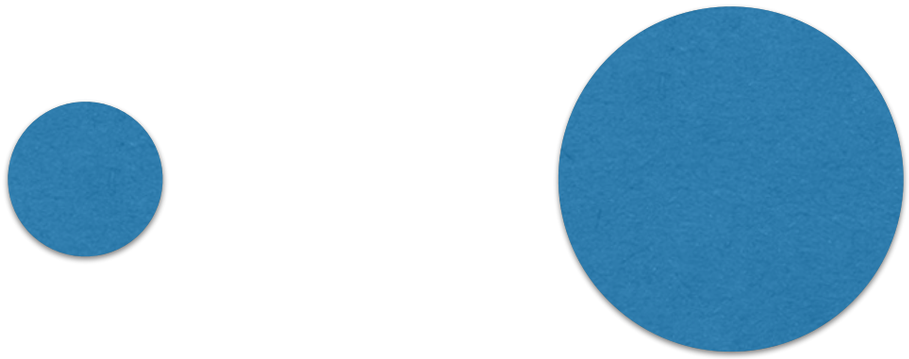
Combien de fois plus grand ? (aire)



A

B

Combien de fois plus grand ? (aire)



A

B

5x

Combien de fois plus grand ? (aire)

A

B

Combien de fois plus grand ? (aire)

A

B

3x

Combien de fois plus sombre ?

A

B

Combien de fois plus sombre ?

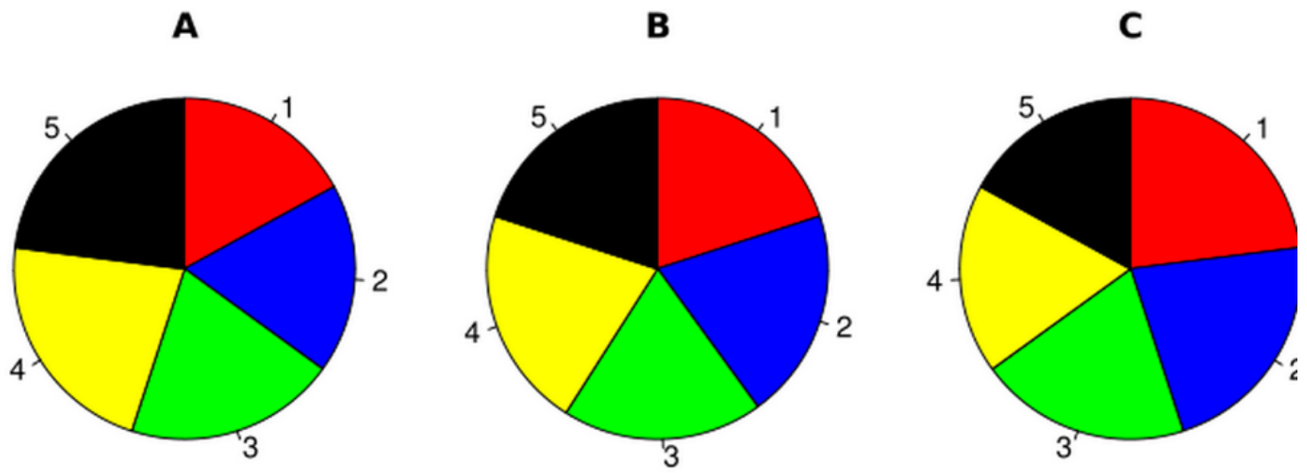
A

B

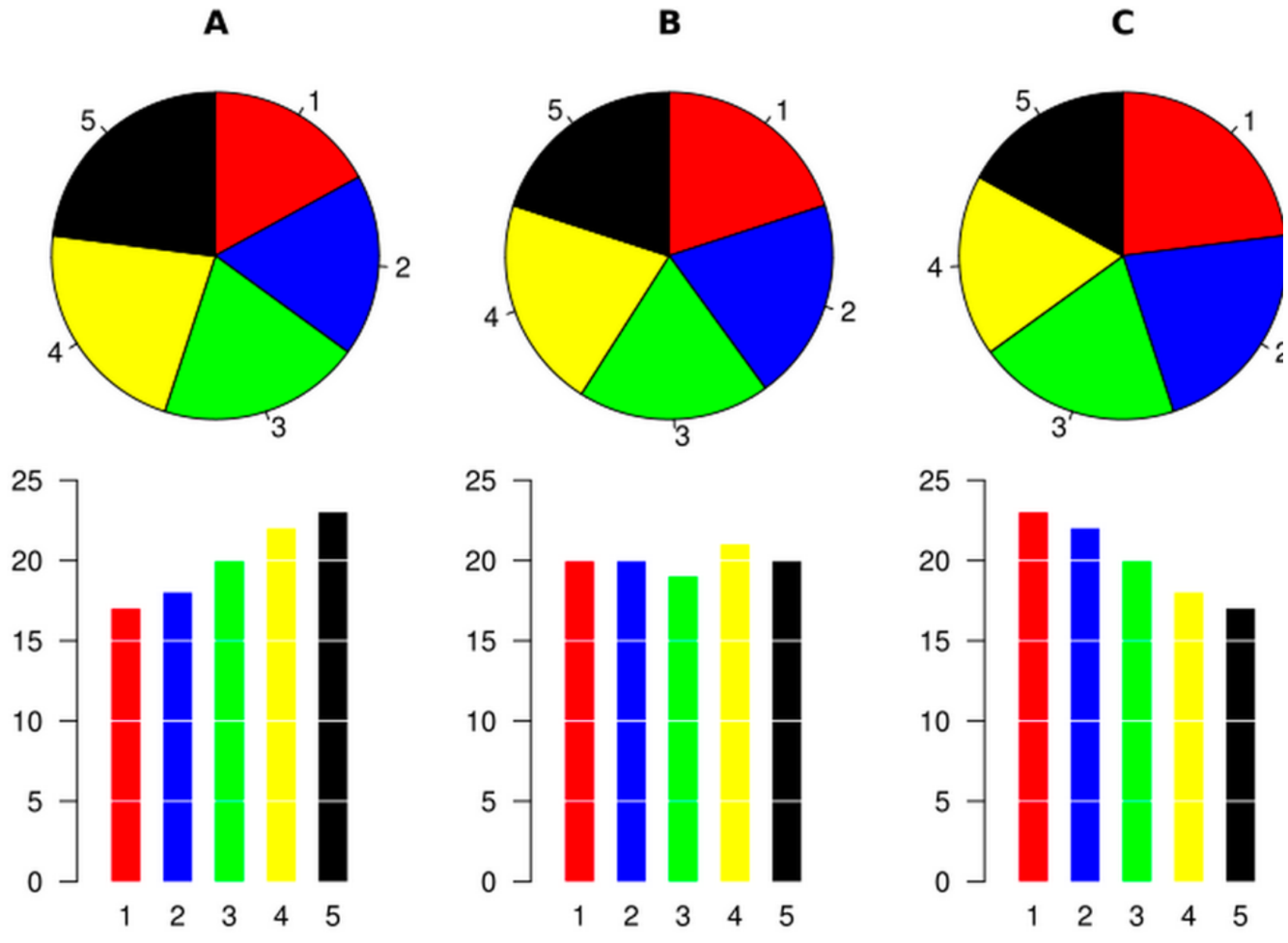
2x

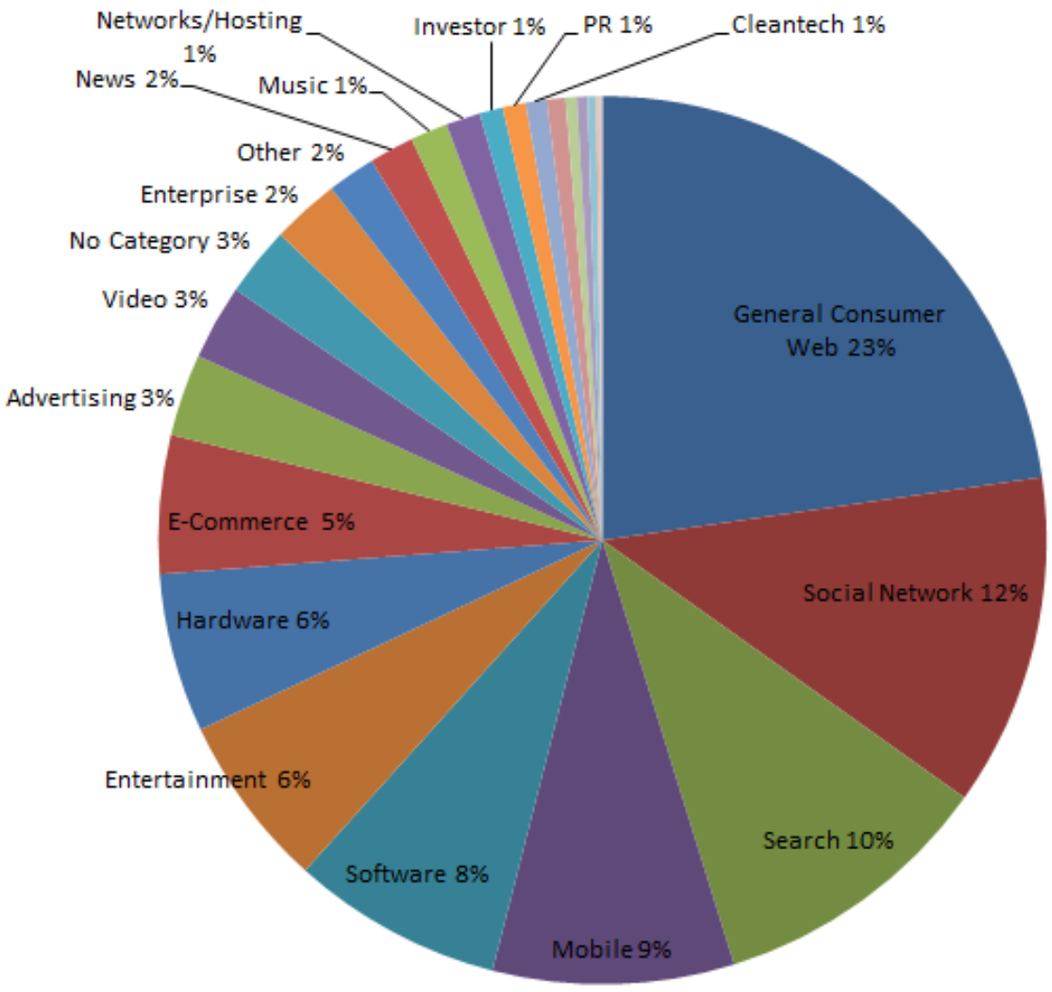
Erreurs communes

Voyez-vous les différences ?



Et maintenant ?





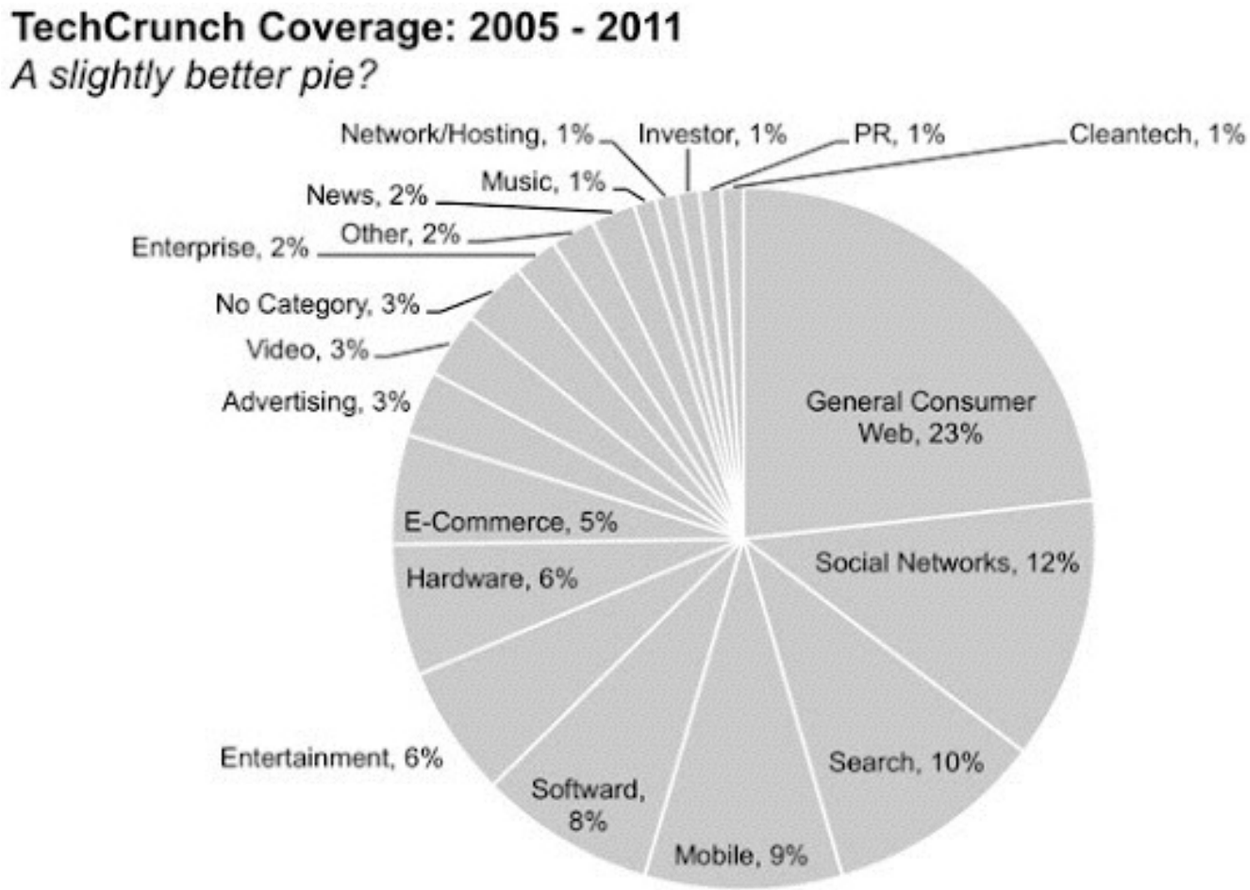
Share of coverage on TechCrunch

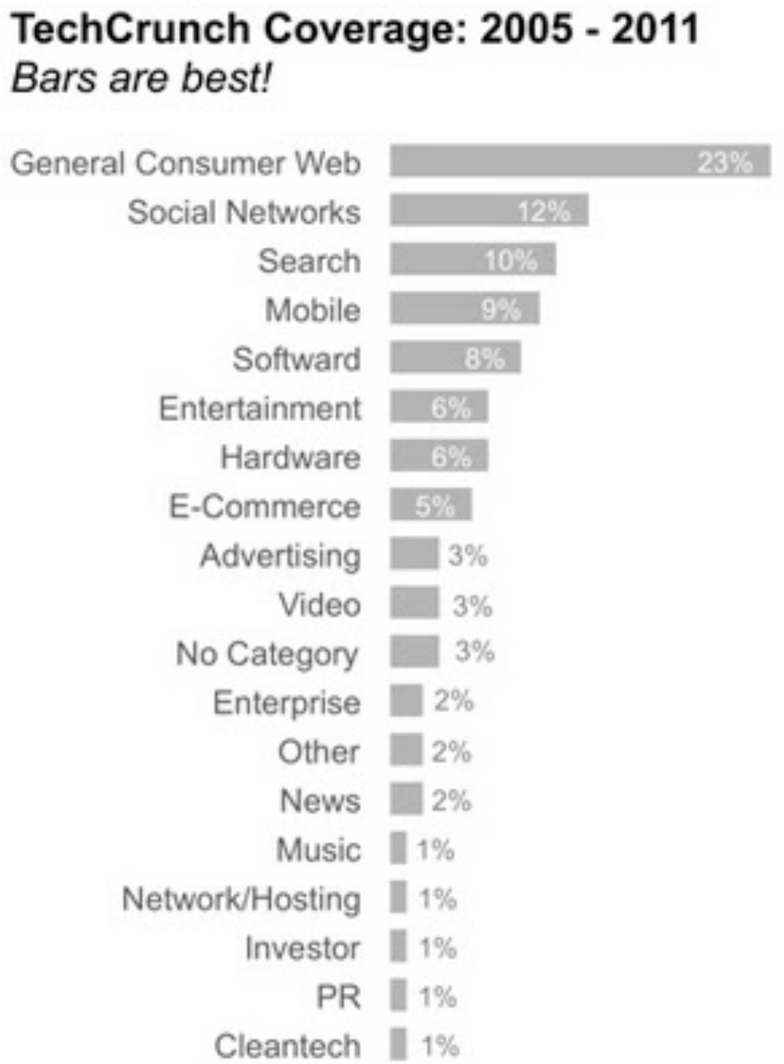
"I hate pie charts.
I mean, really hate them."

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

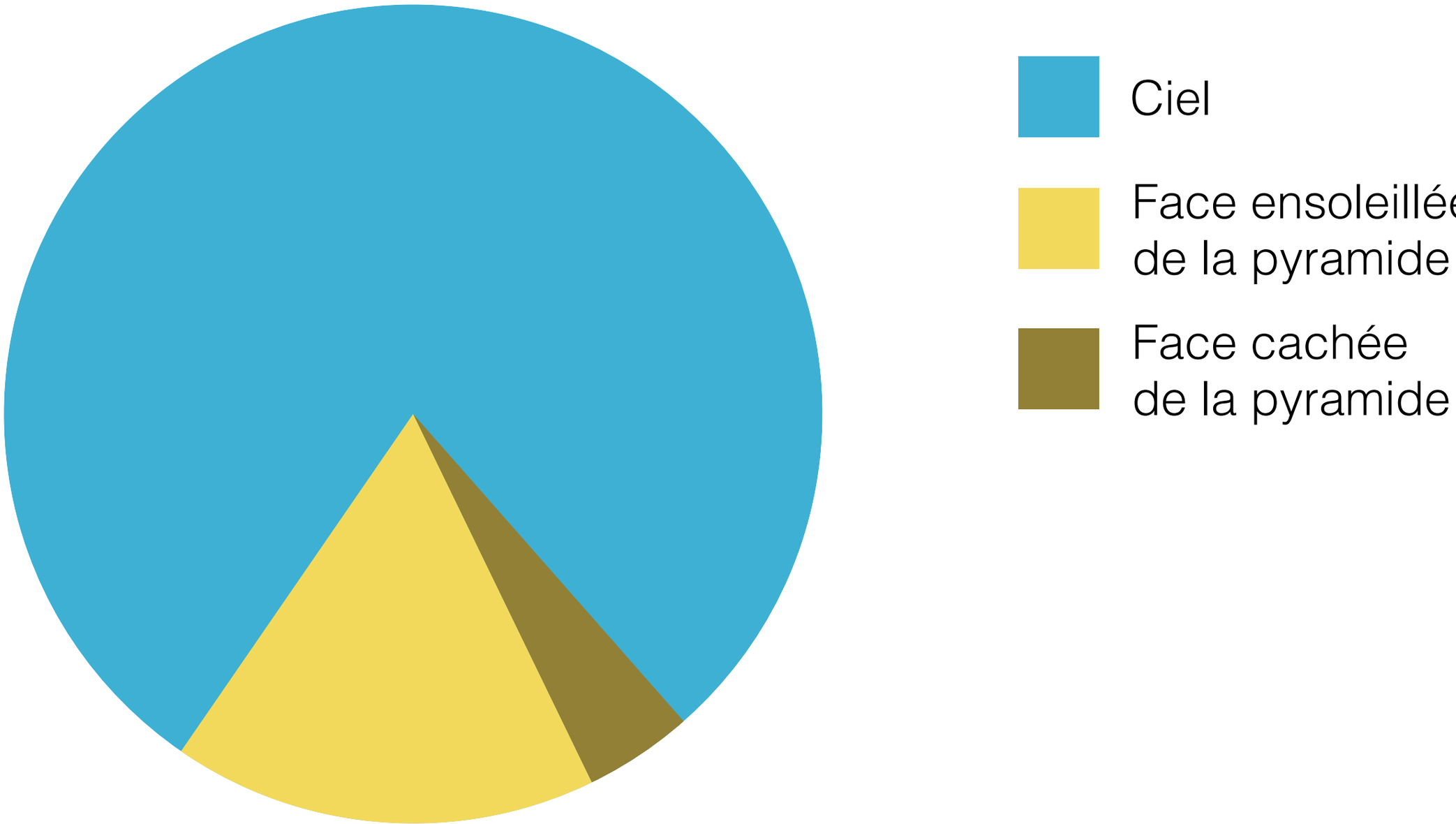
Cole Nussbaumer

Redesign

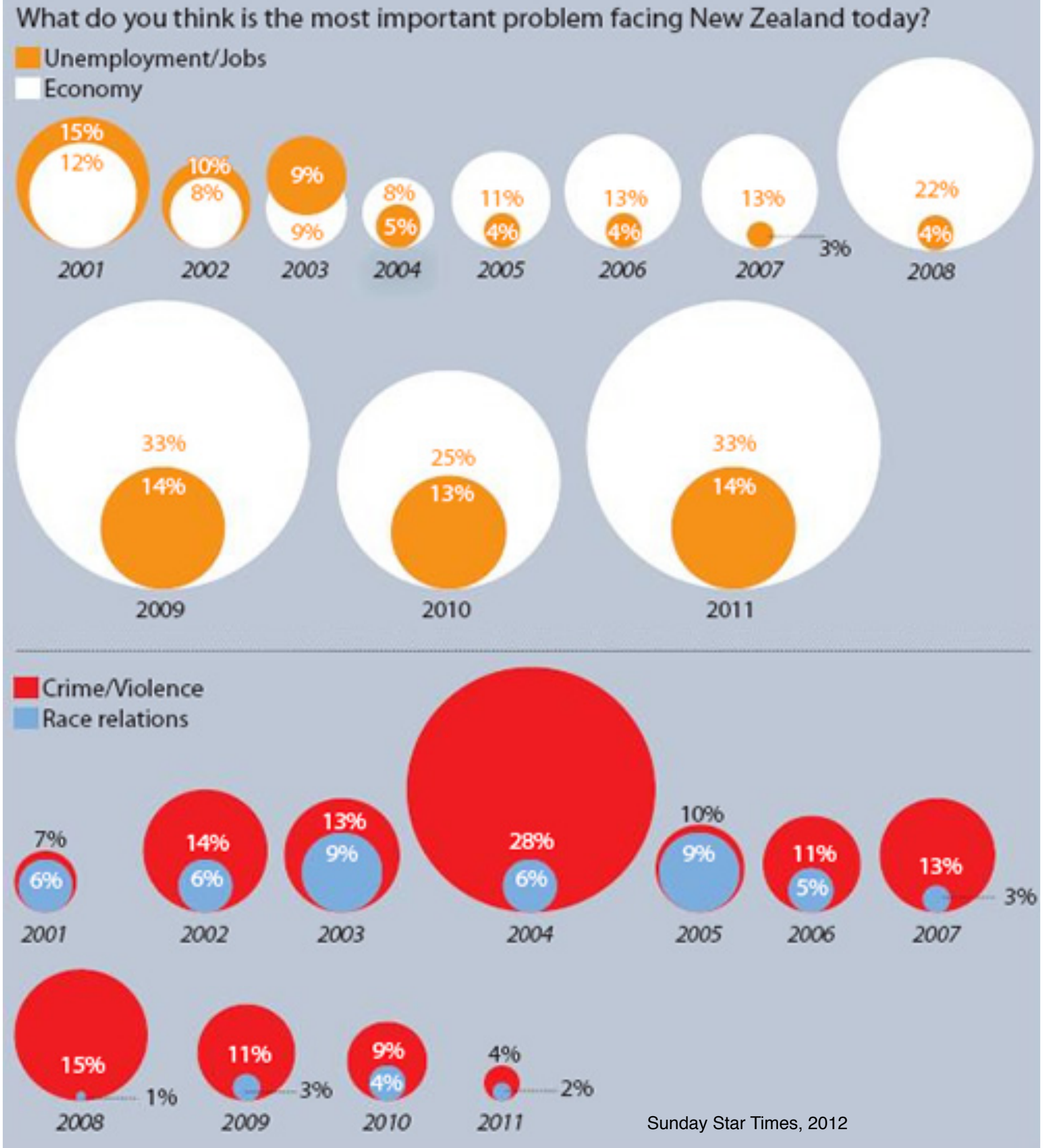

TechCrunch Coverage: 2005 - 2011
A slightly better pie?
Network/Hosting, 1%
Investor, 1%
PR, 1%
Cleantech, 1%
Music, 1%
News, 2%
Other, 2%
Enterprise, 2%
No Category, 3%
Video, 3%
Advertising, 3%
General Consumer Web, 23%
E-Commerce, 5%
Social Networks, 12%
Hardware, 6%
Entertainment, 6%
Softward, 8%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

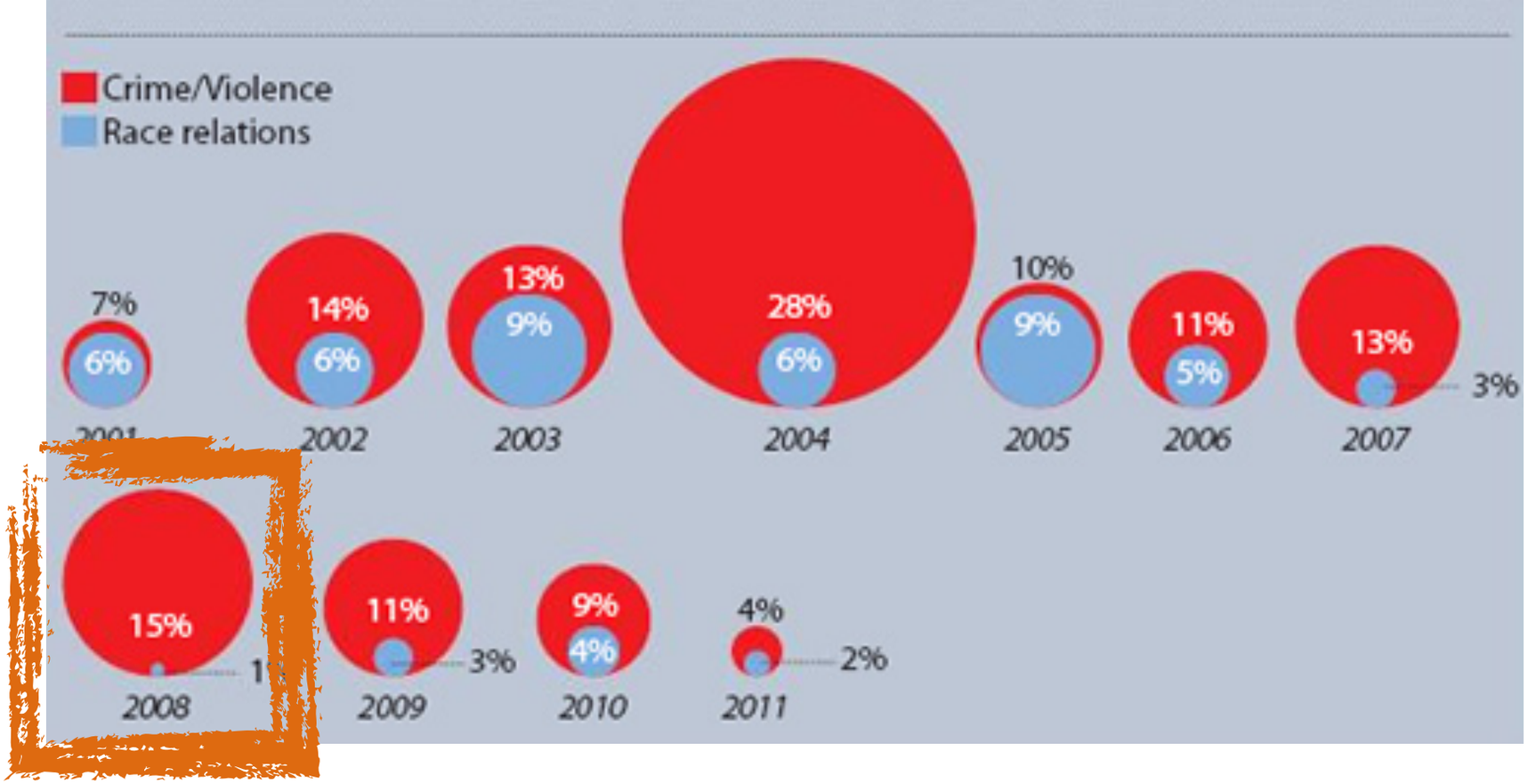
Le seul pie chart valable



Most important issues

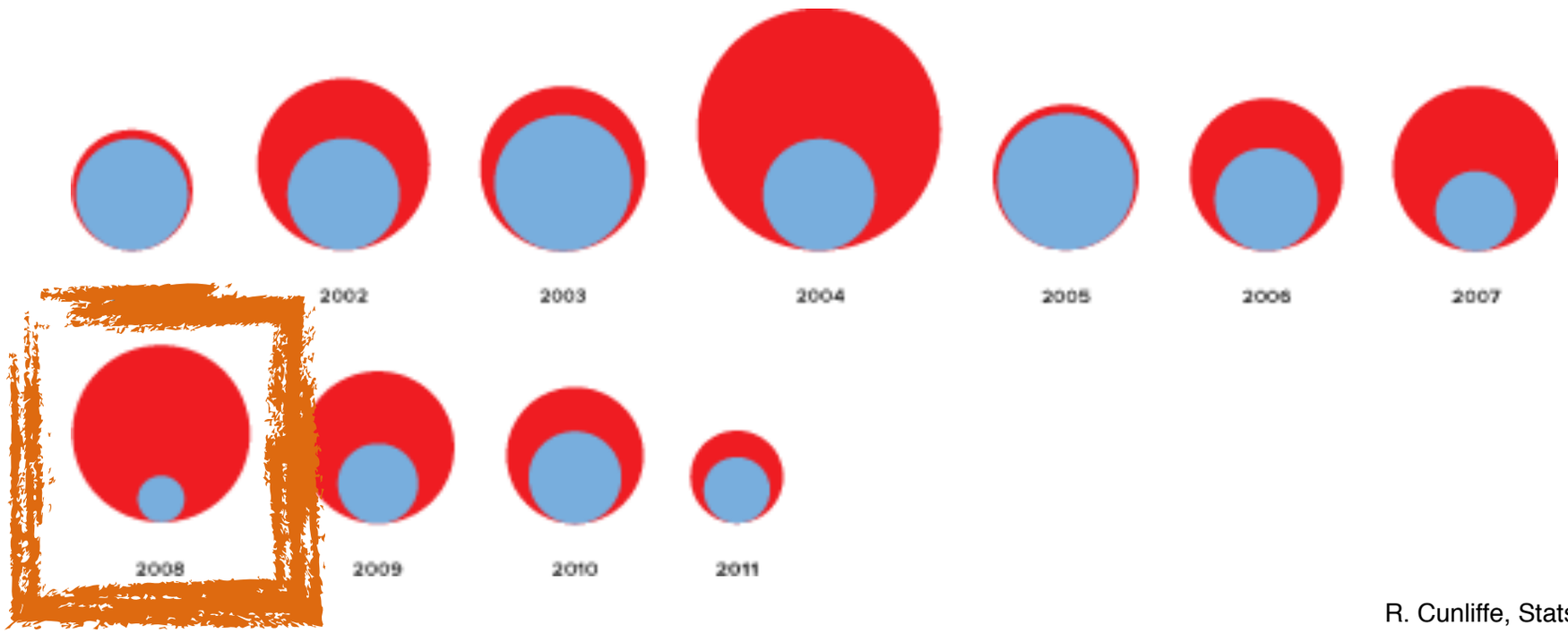


Sunday Star Times, 2012

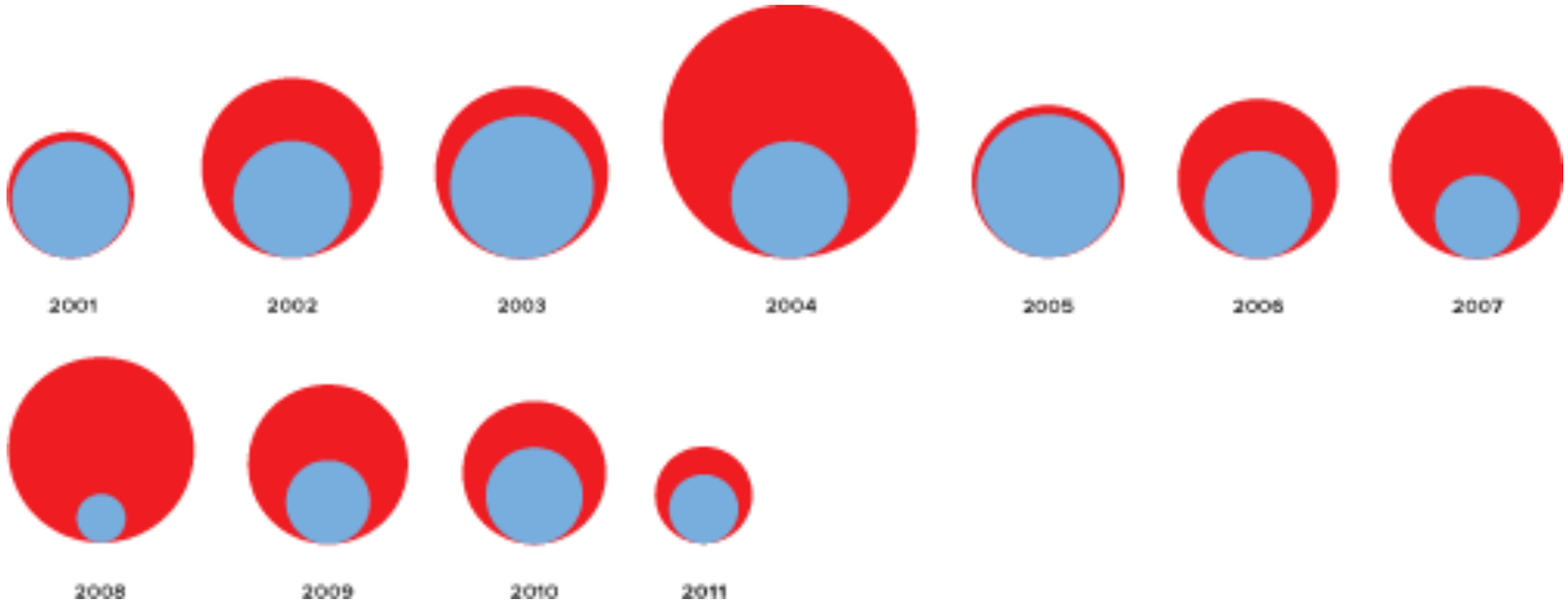


Quantity encoded by diameter, not area!

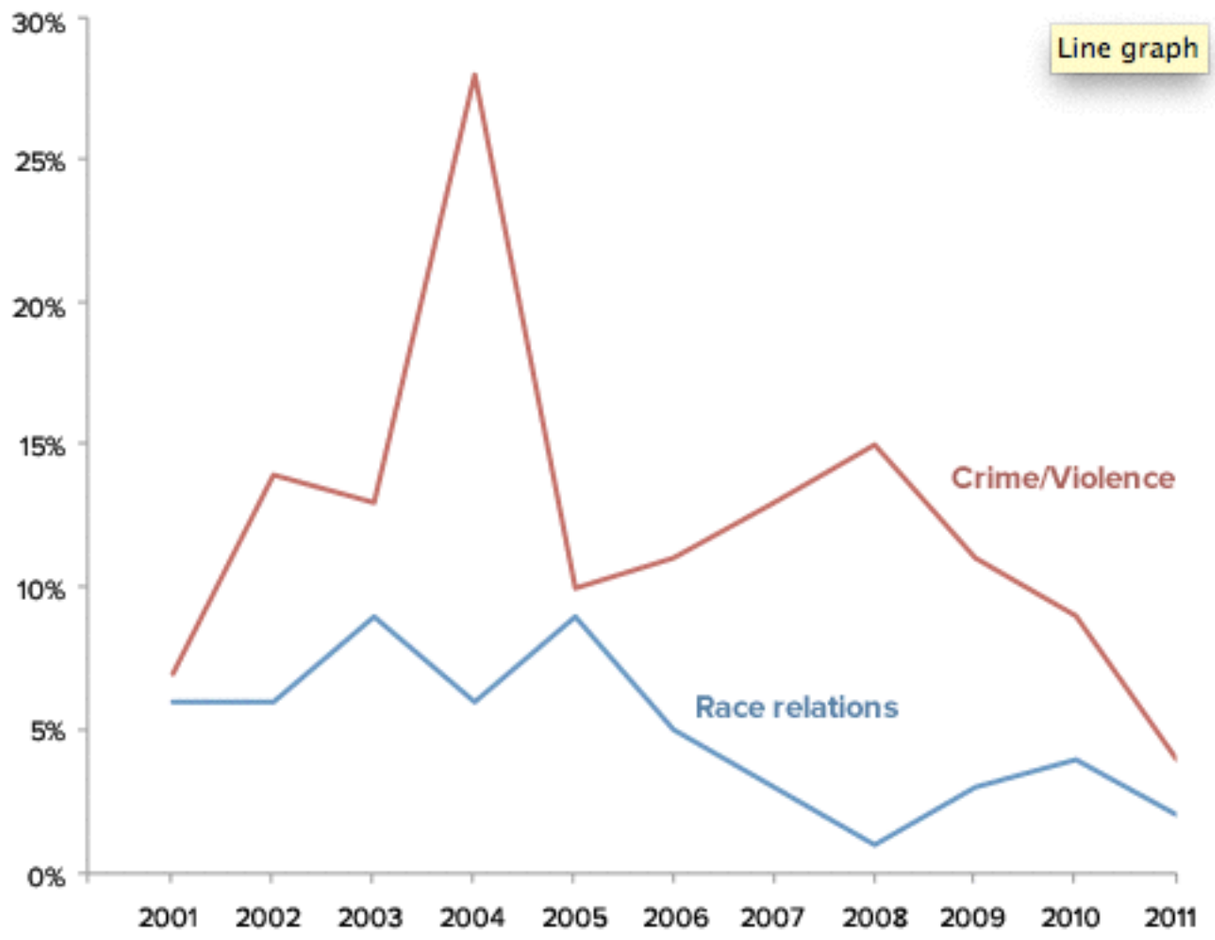
Fixing that:



R. Cunliffe, Stats Chat

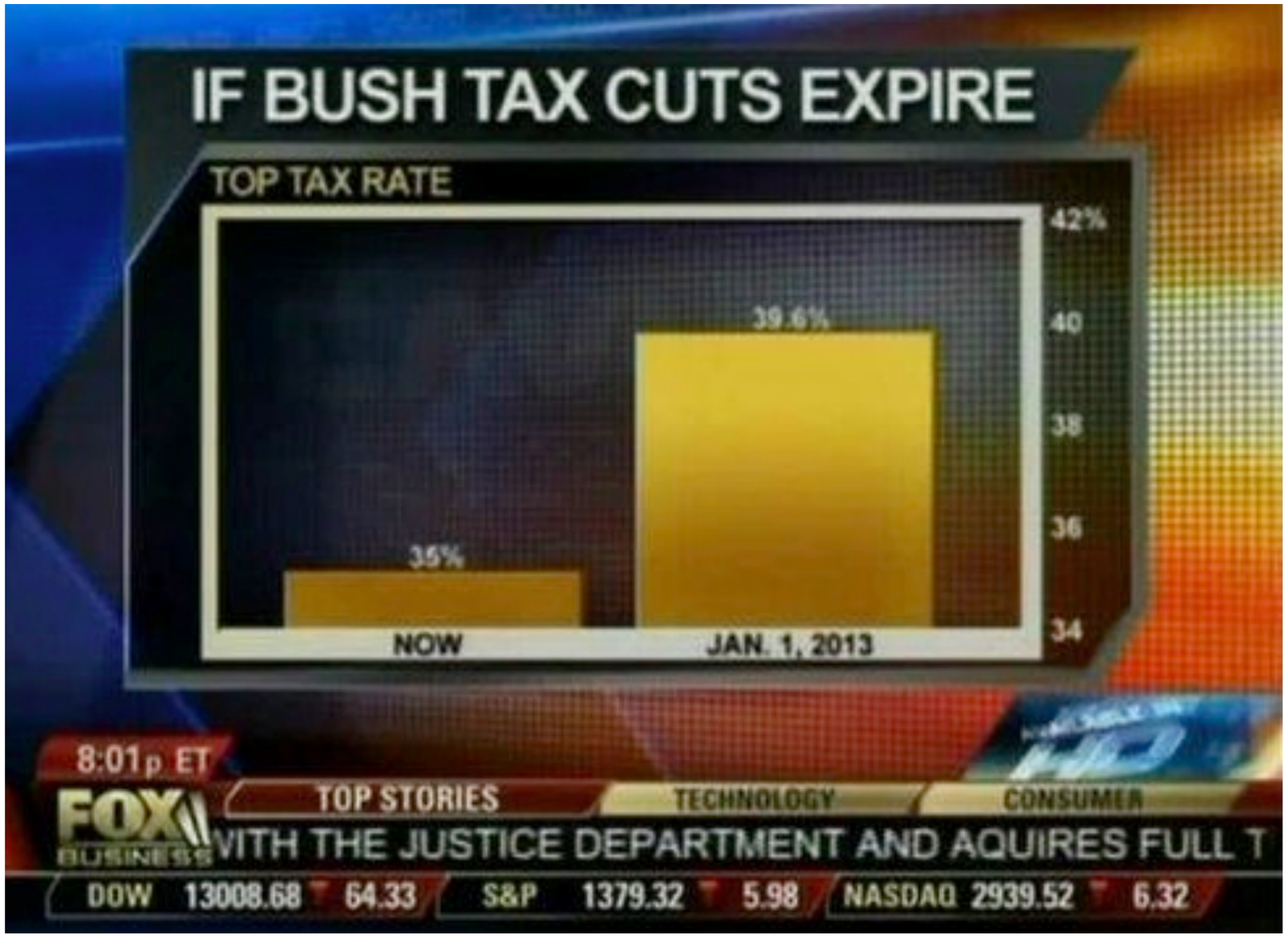


Mais le choix de l'encodage était-il approprié initialement ?



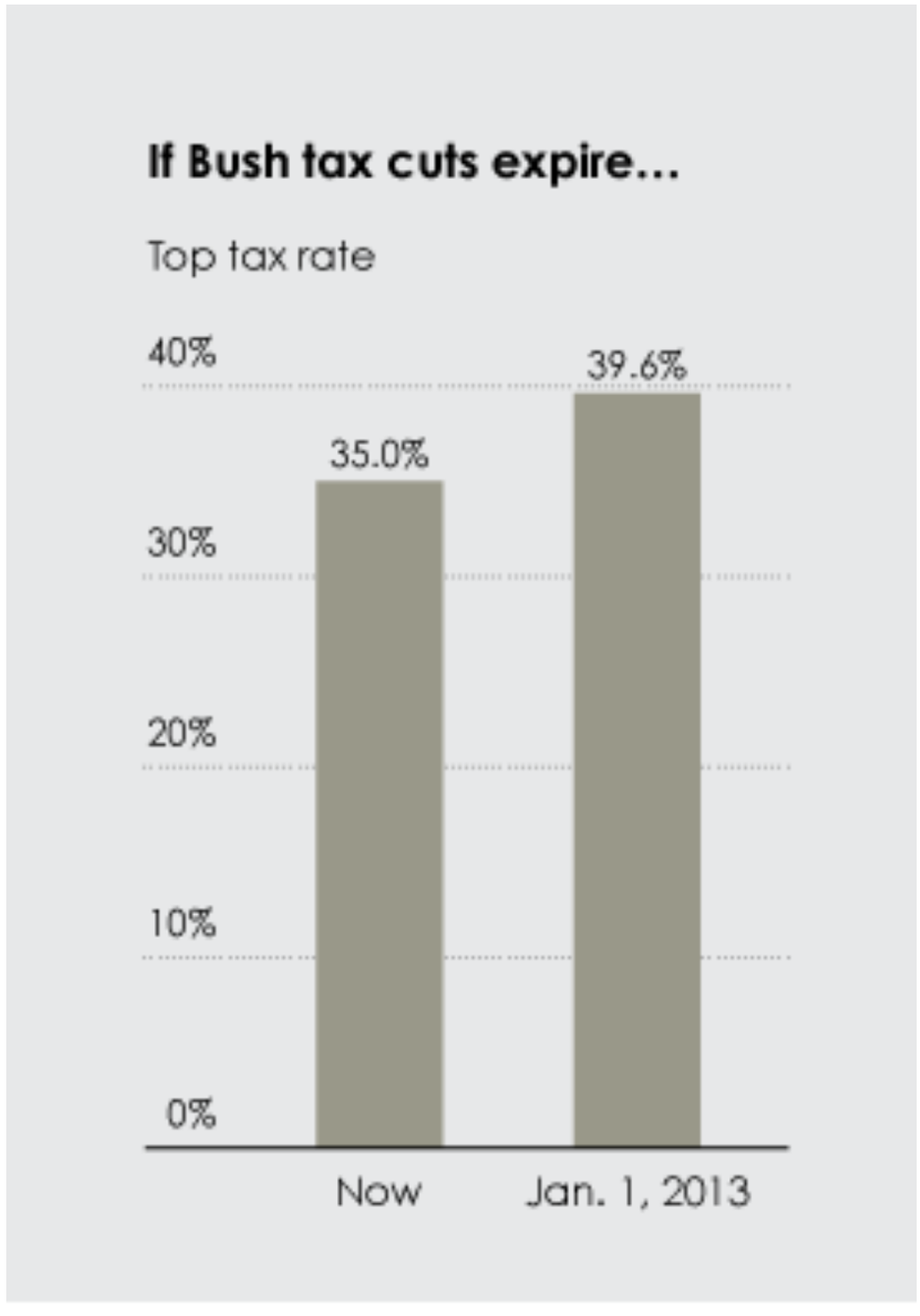
R. Cunliffe, Stats Chat

Graphical Integrity



Flowing Data

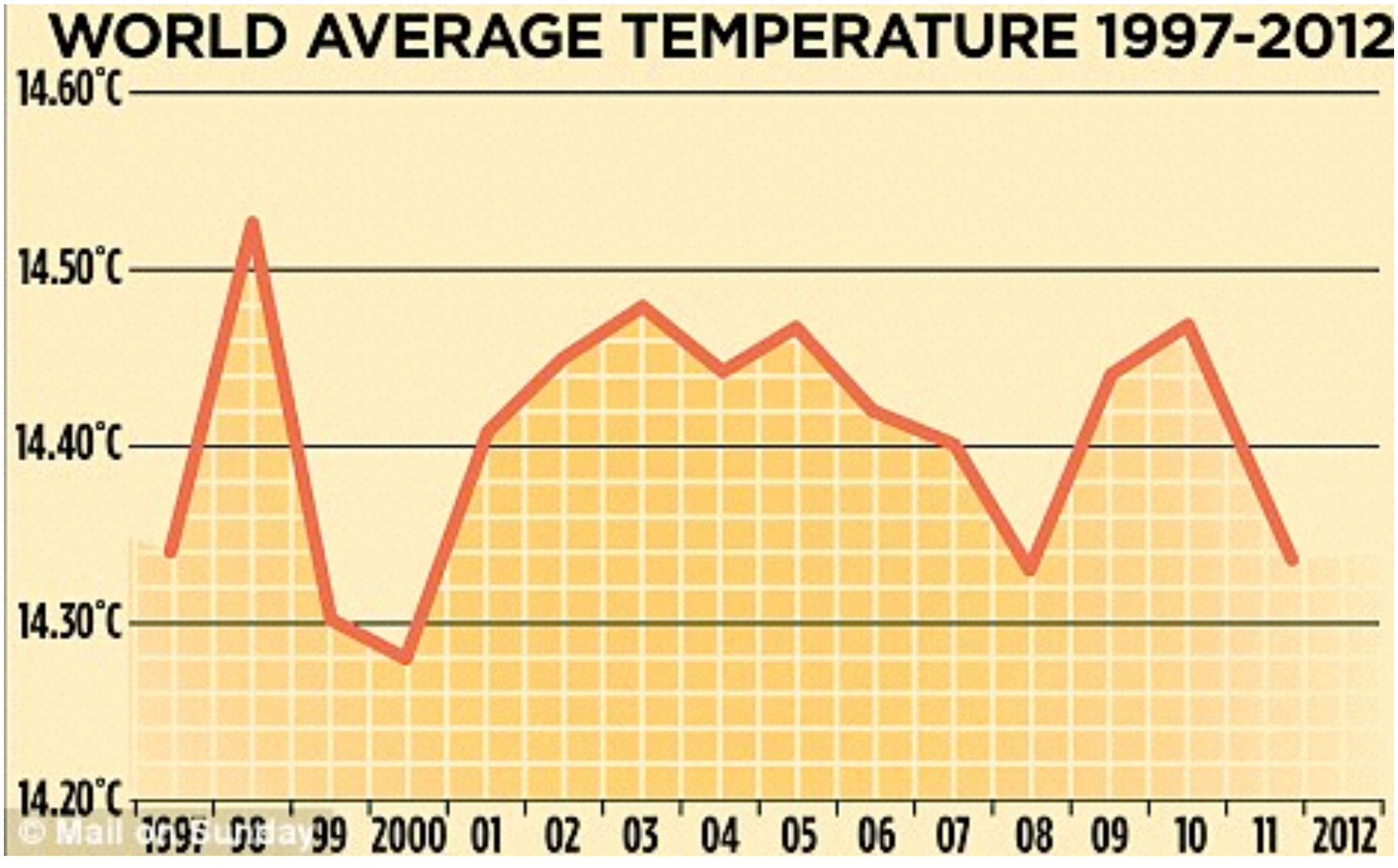
Scale Distortions



What's wrong?

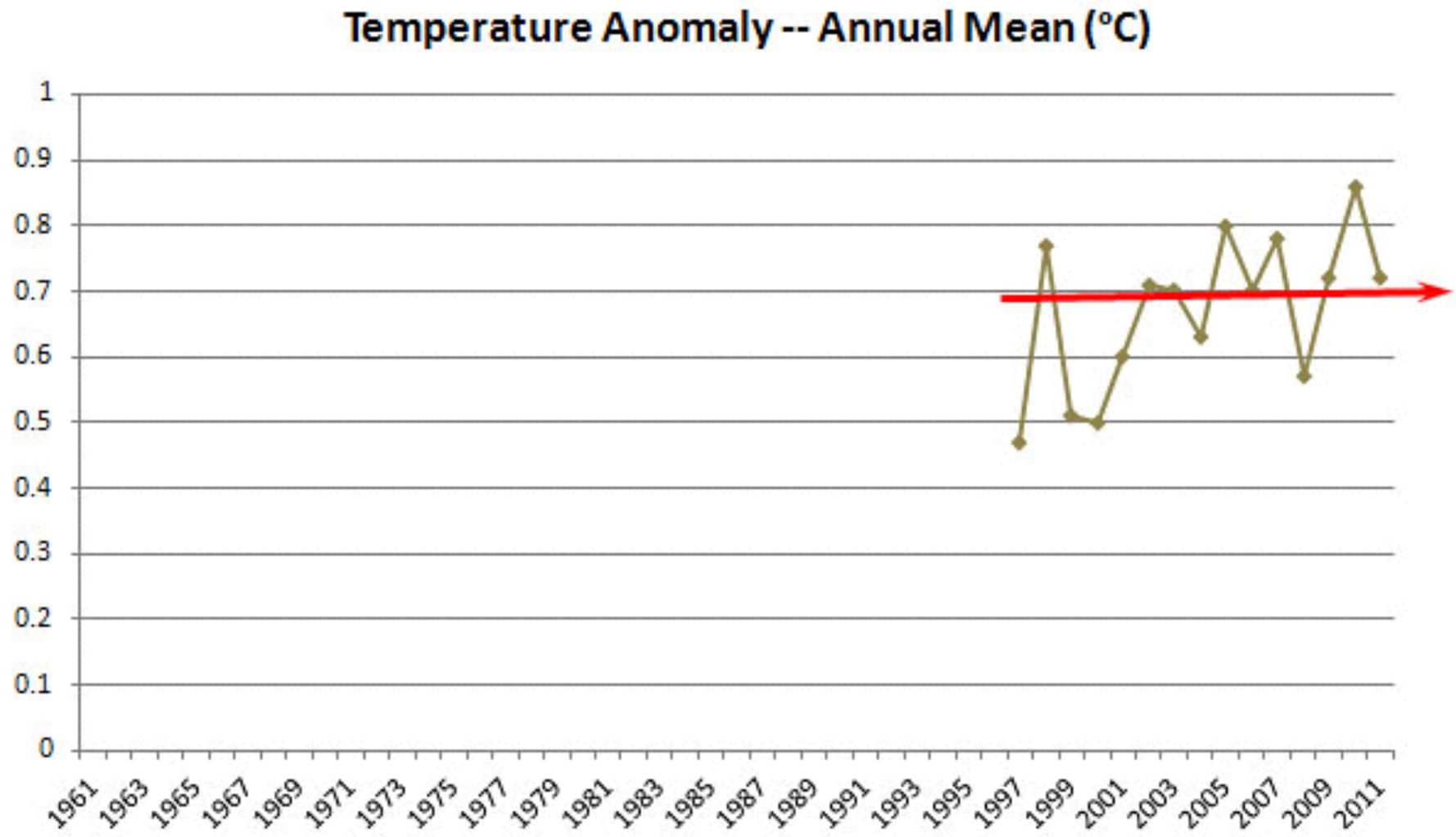

How 2012 Stacks Up
The Warmest Years on Record
Contiguous U.S.
55.3
55.5°
55.0°
54.5°
54.0°
53.5°
53.0°
54.31
54.32
54.14
53.93
53.81
1921
1999
1934
2006
1998
2012
Source: NOAA's National Climatic Data Center - State of the Climate National Overview
CLIMATE CENTRAL

Global Warming?



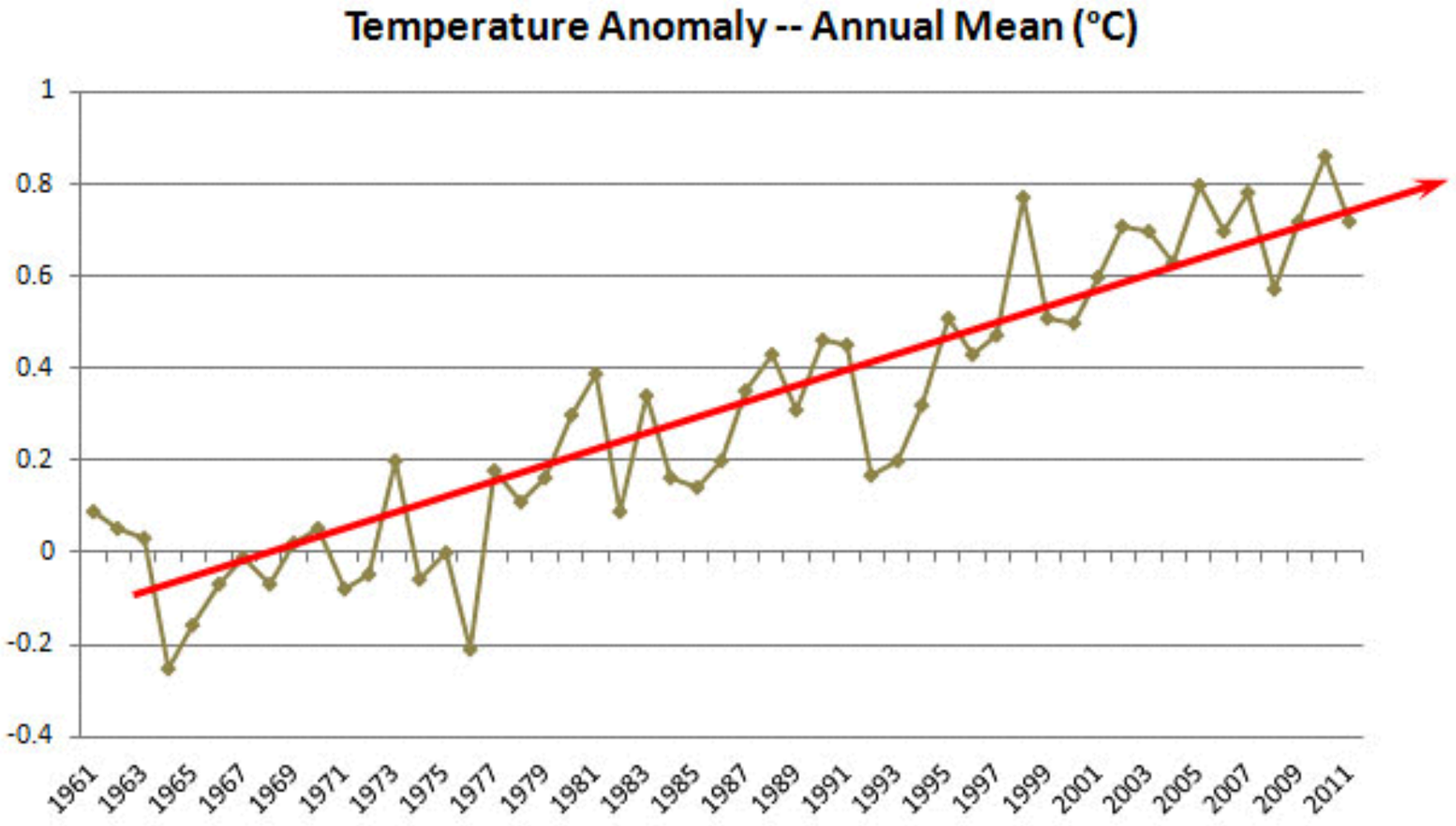
The Daily Mail, UK

Global Warming?



Mother Jones

Global Warming - Frame the Data



Mother Jones

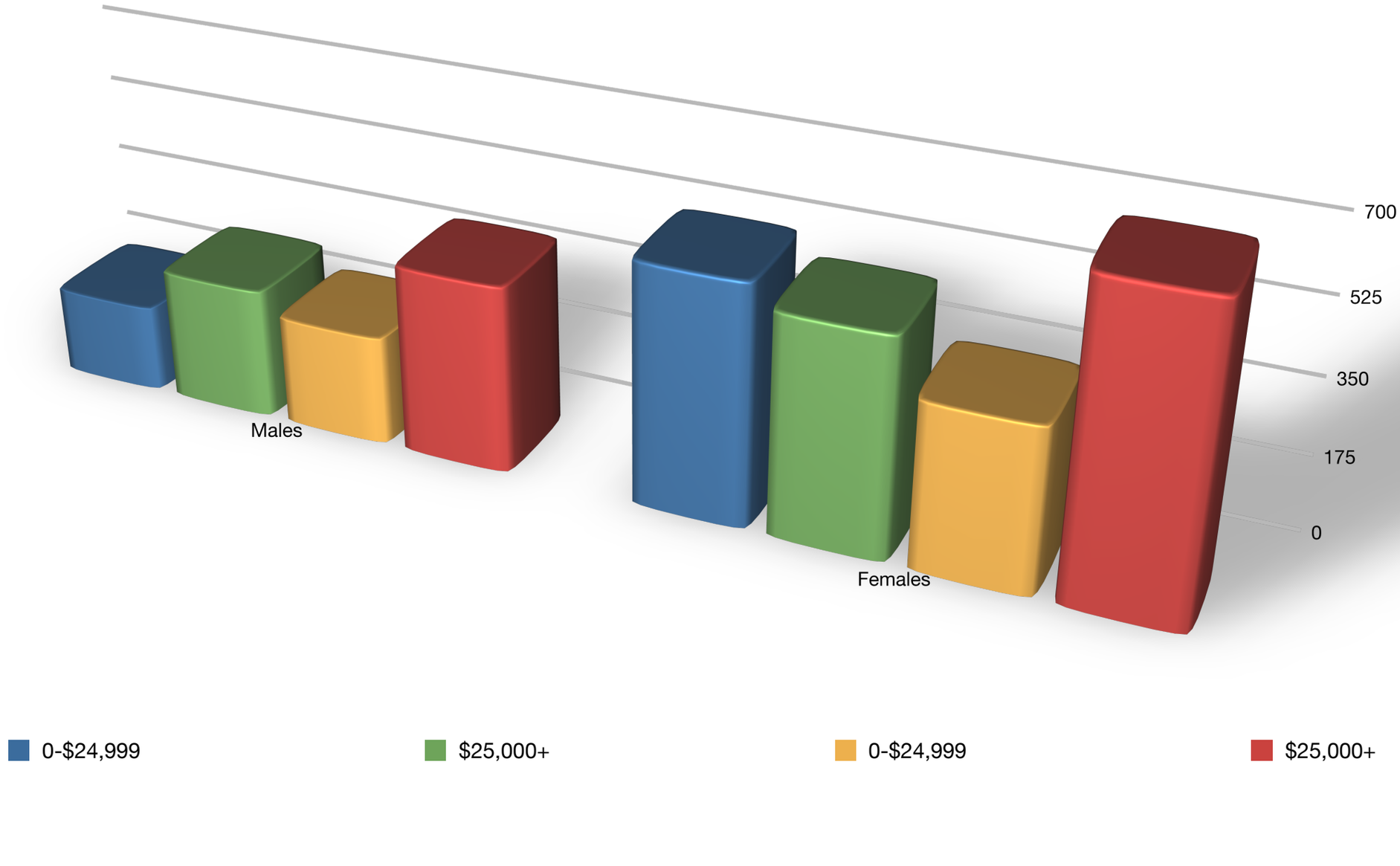
Tufte’s Integrity Principles

Show **data variation**, not design variation

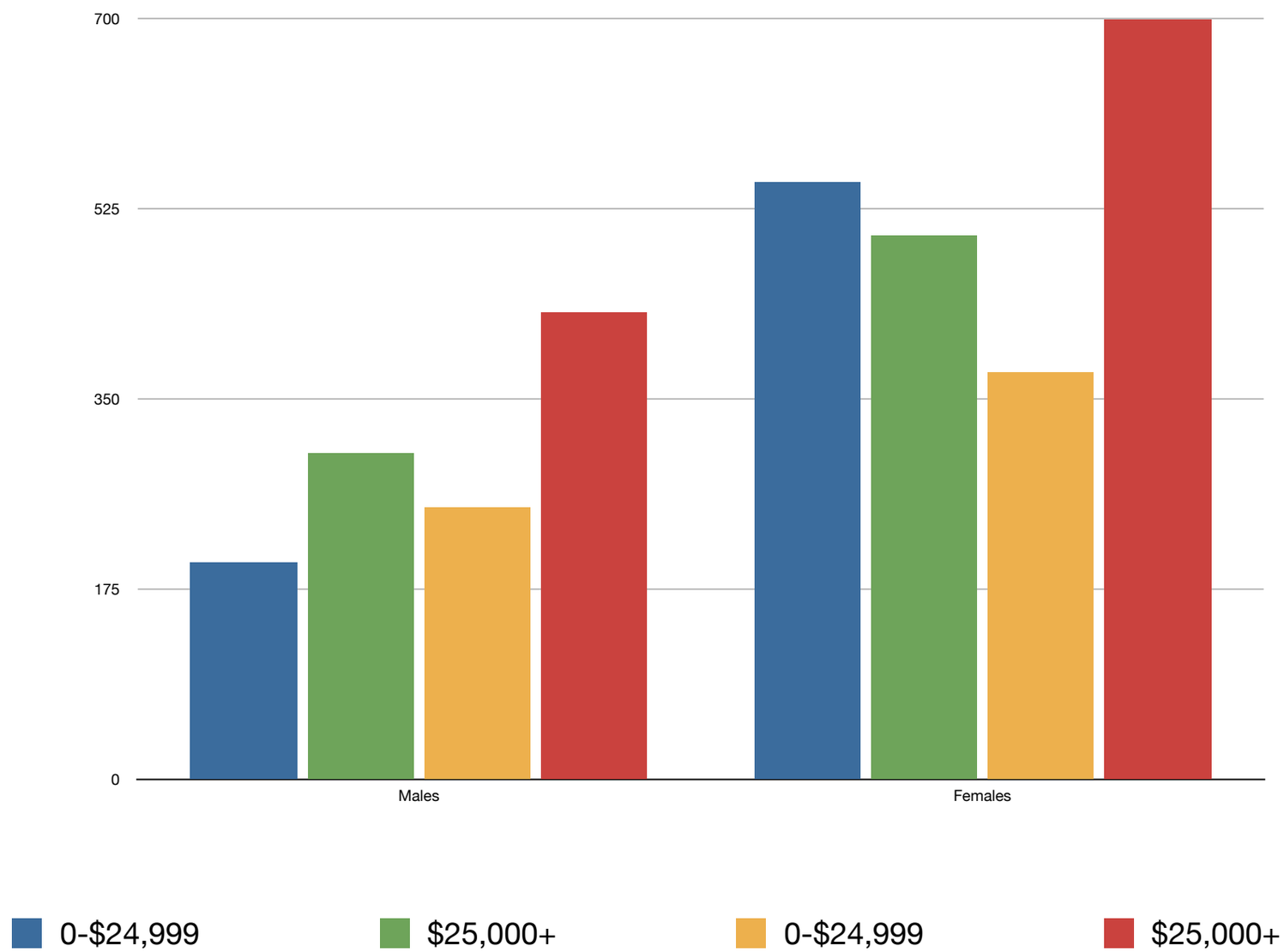
Clear, detailed, and thorough **labeling** and **appropriate scales**

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

Maximize Data-Ink Ratio

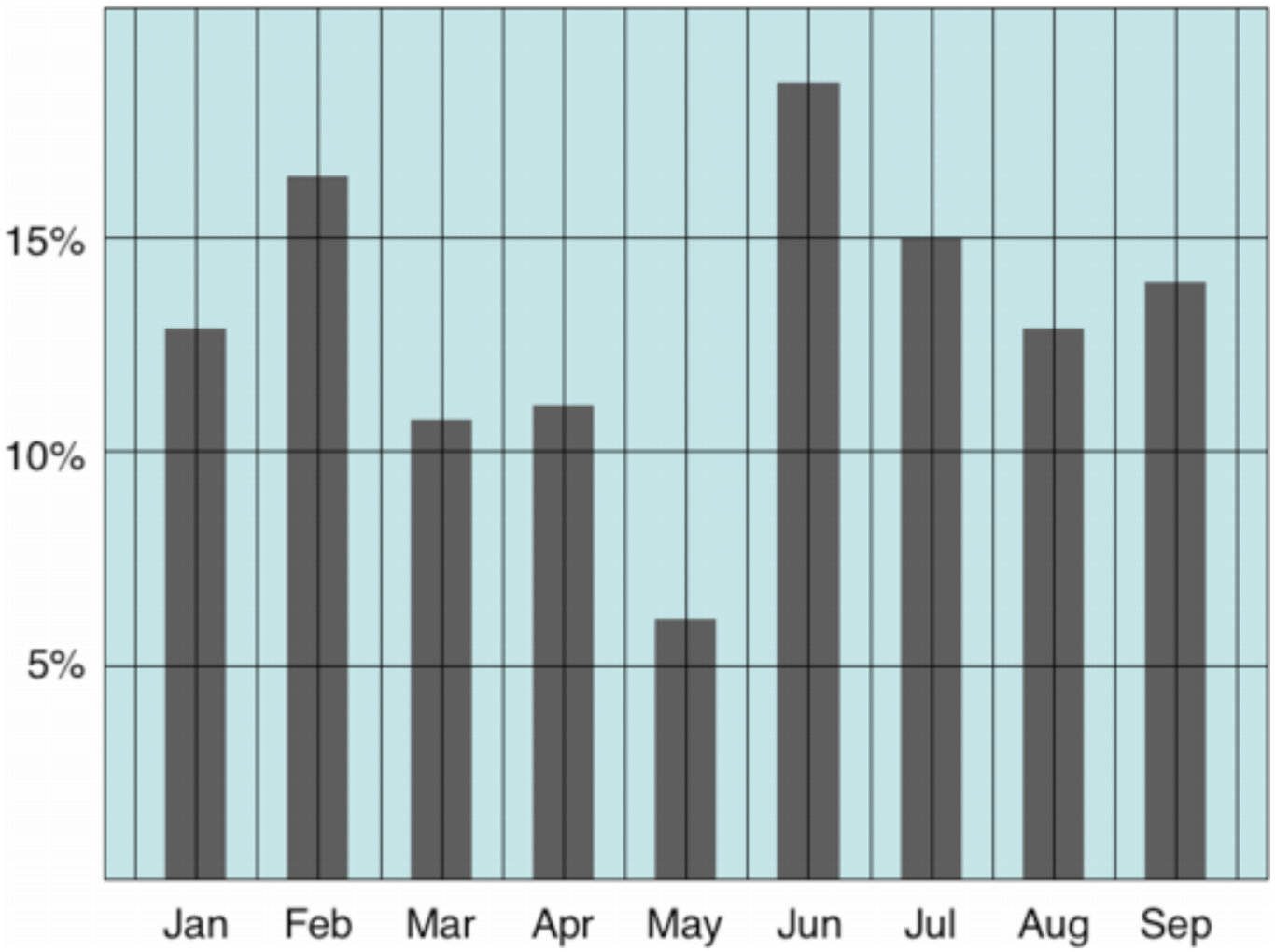


Maximize Data-Ink Ratio



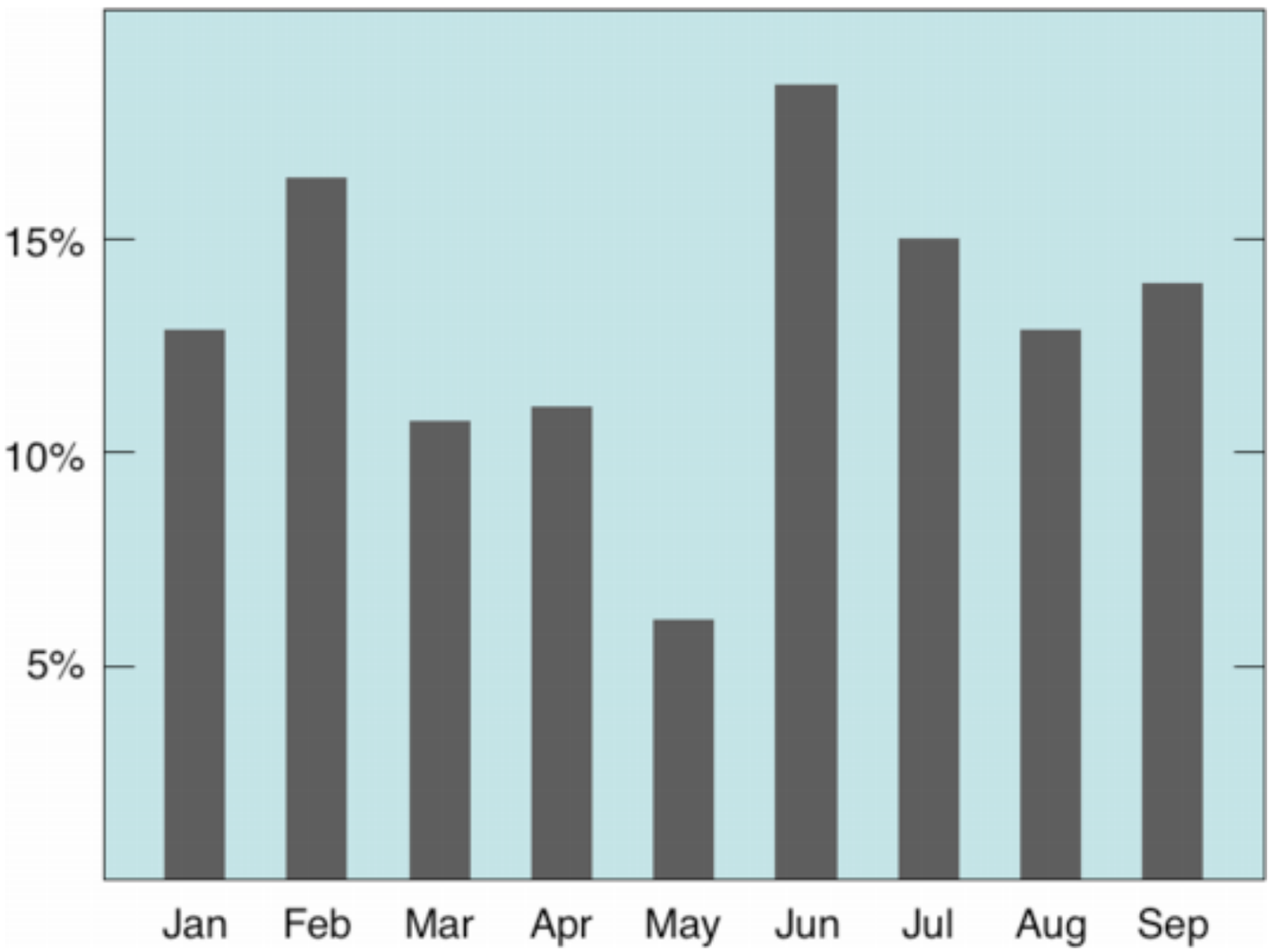
Avoid Chartjunk

Extraneous visual elements that distract from the message



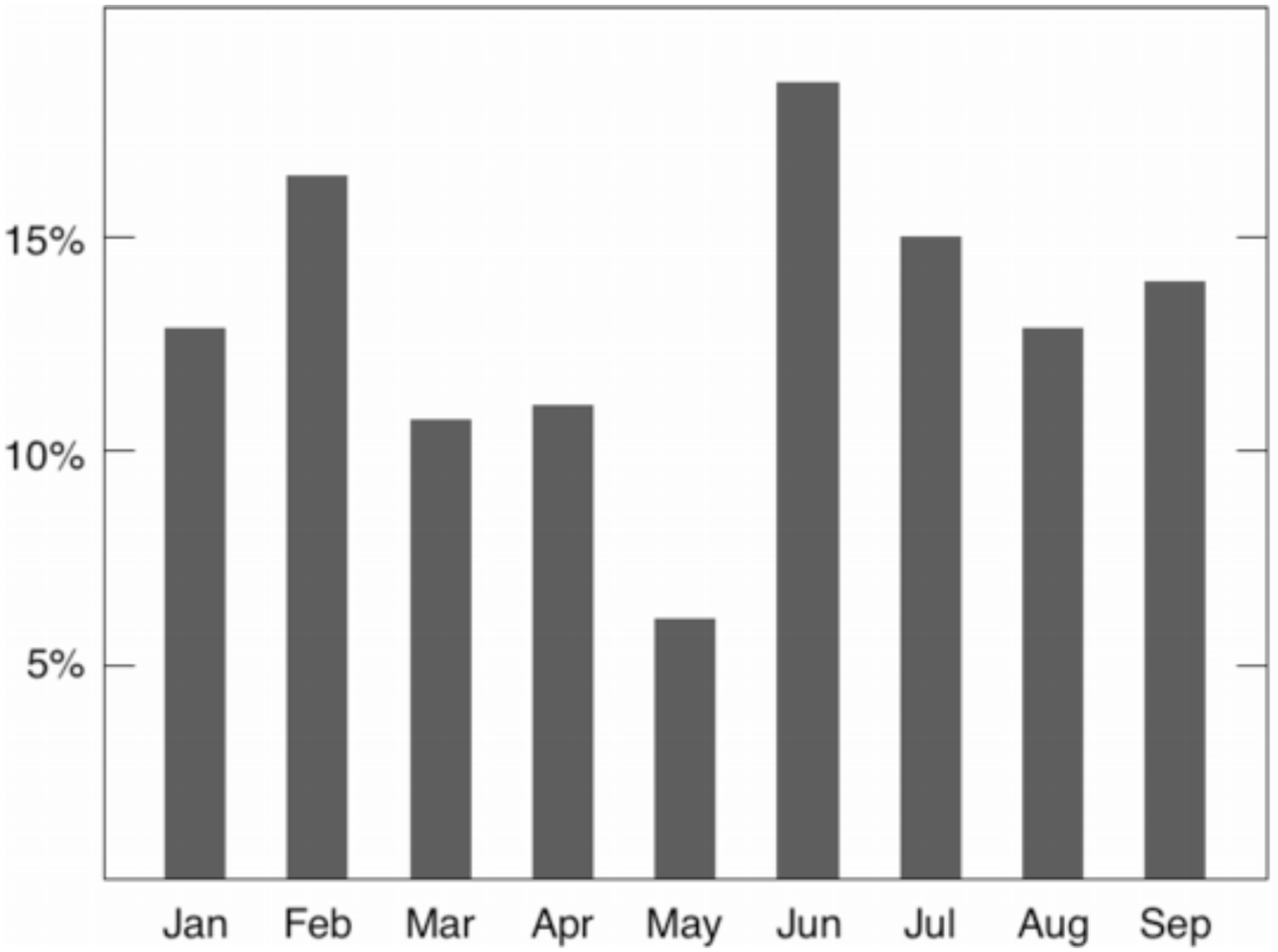
ongoing, Tim Brey

Avoid Chartjunk



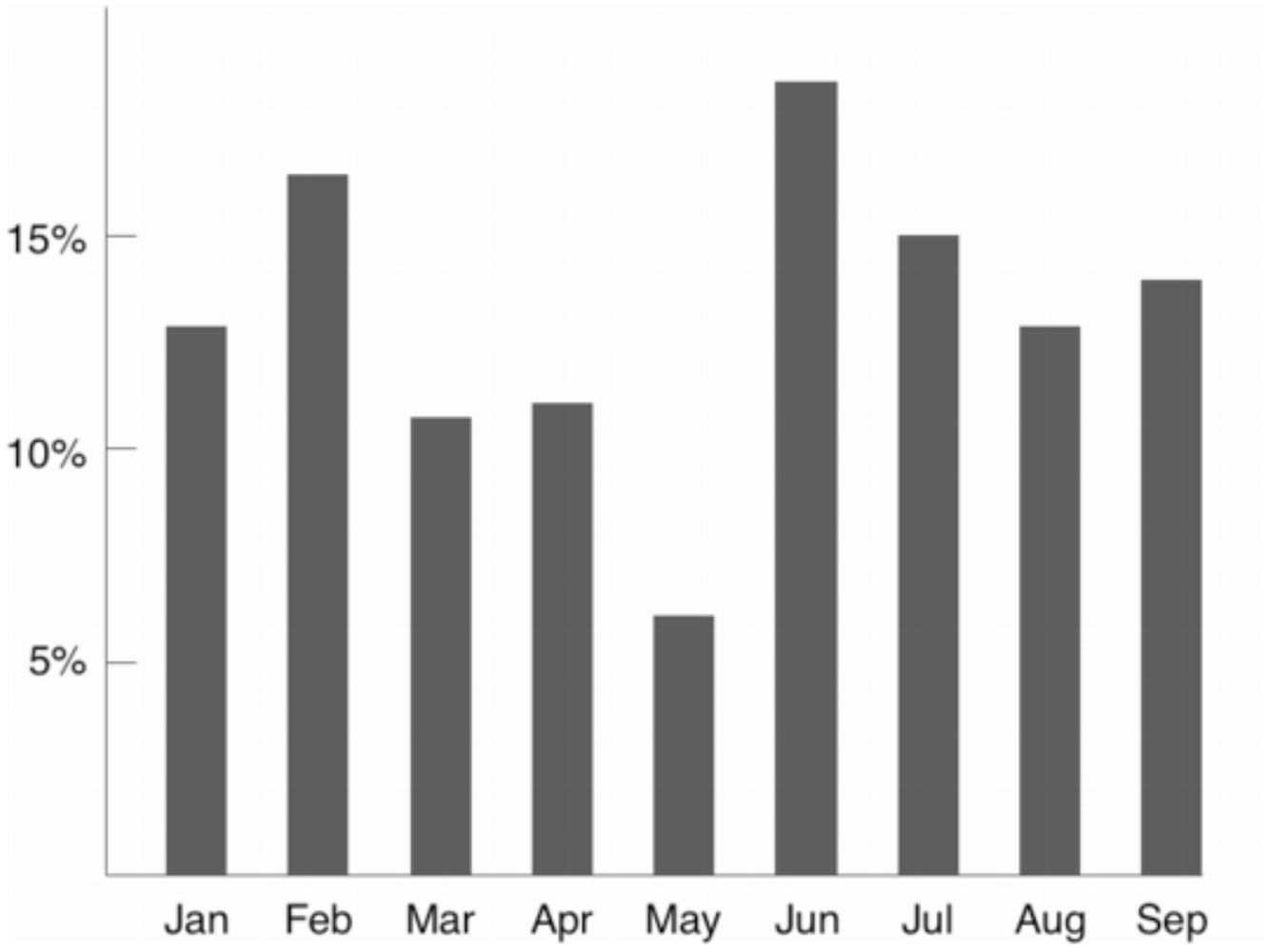
ongoing, Tim Brey

Avoid Chartjunk



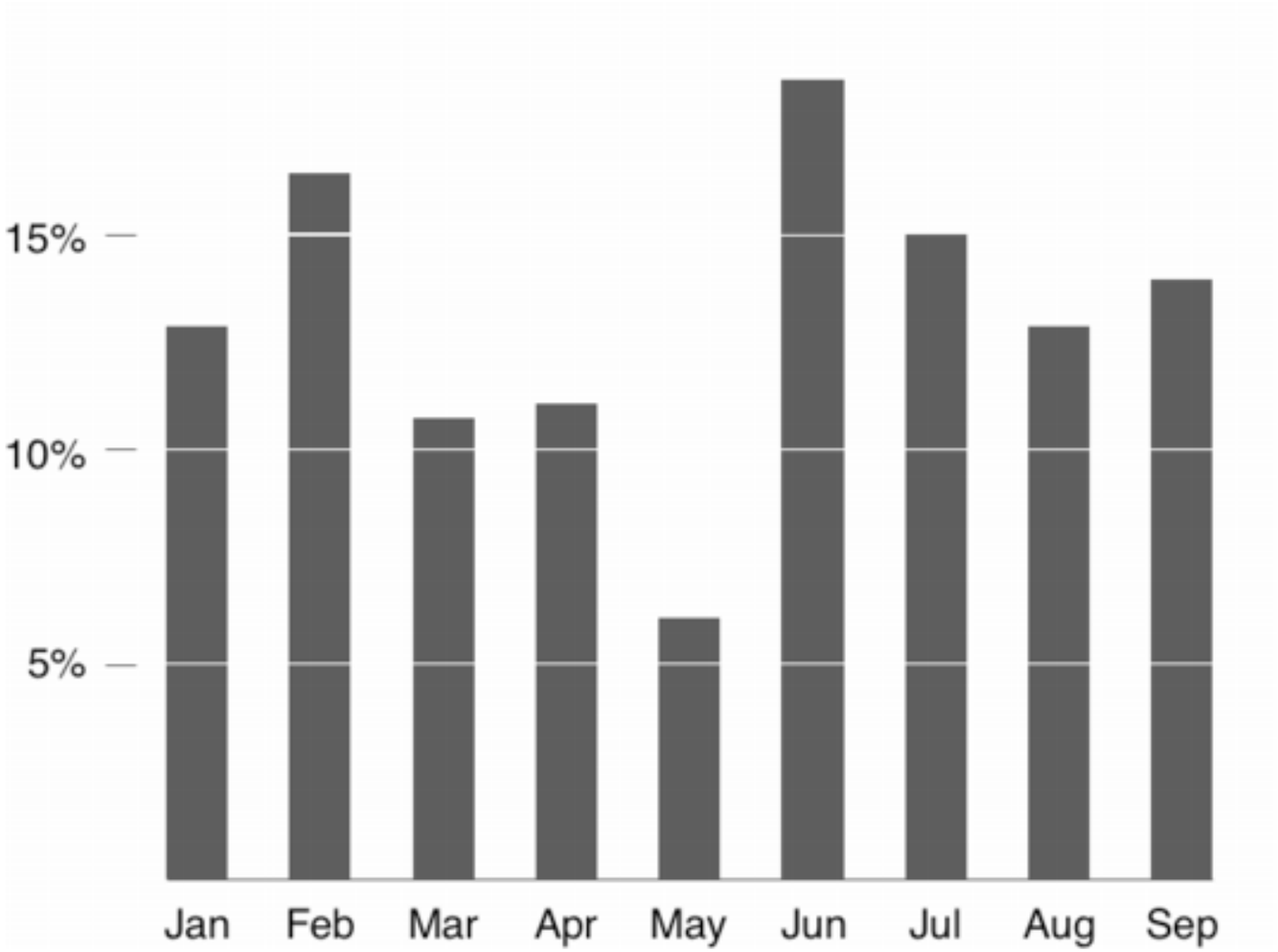
ongoing, Tim Brey

Avoid Chartjunk



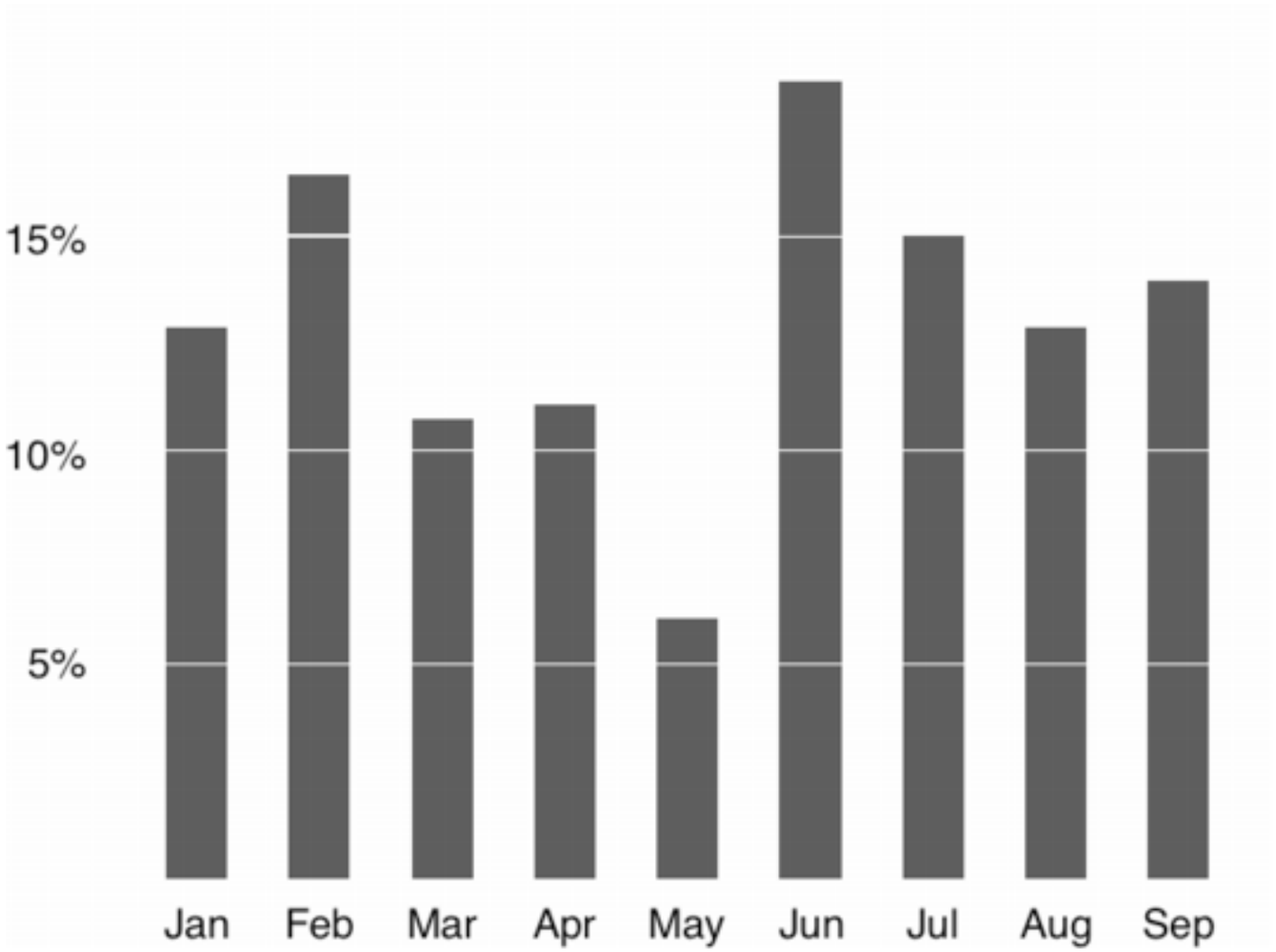
ongoing, Tim Brey

Avoid Chartjunk



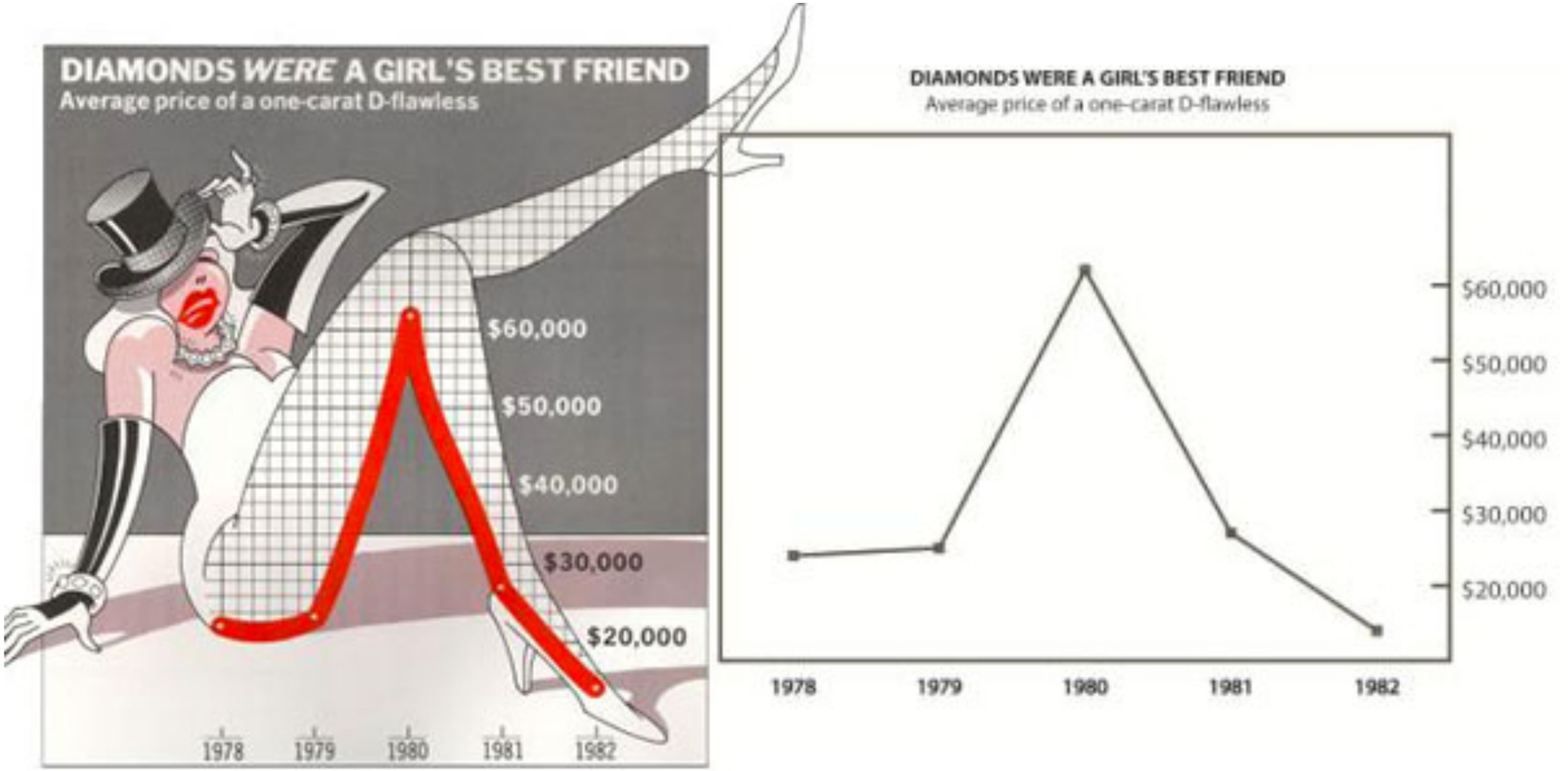
ongoing, Tim Brey

Avoid Chartjunk



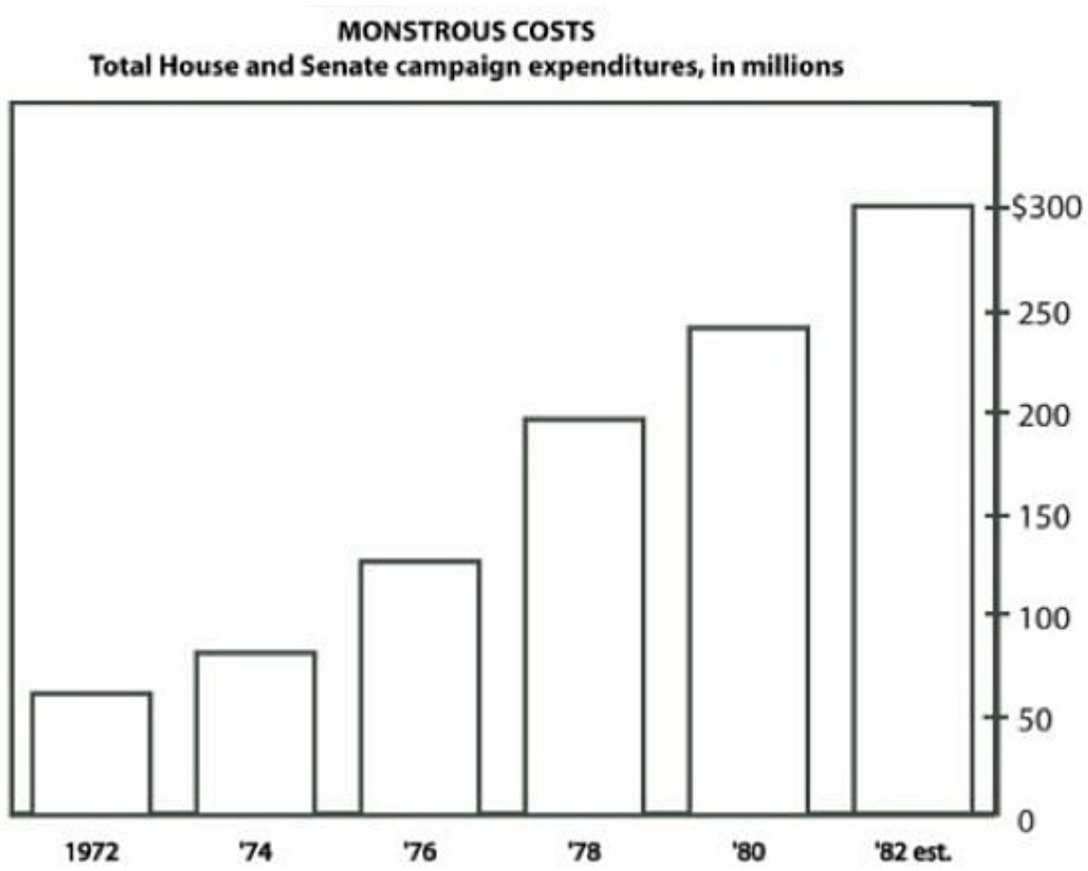
ongoing, Tim Brey

Which is better?



[Bateman et al. 2010]

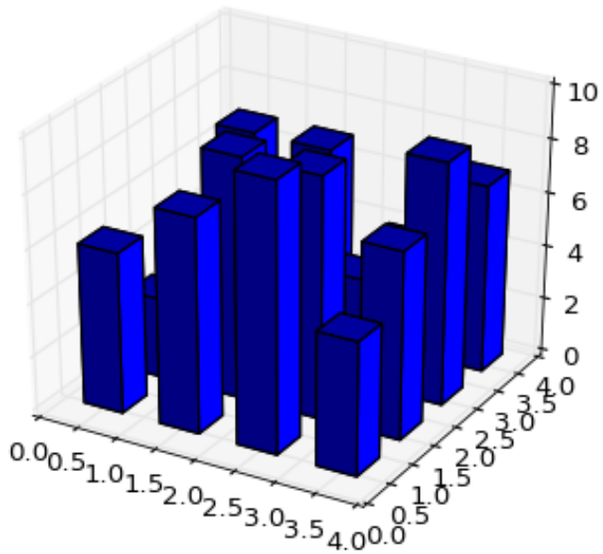
Which is better?

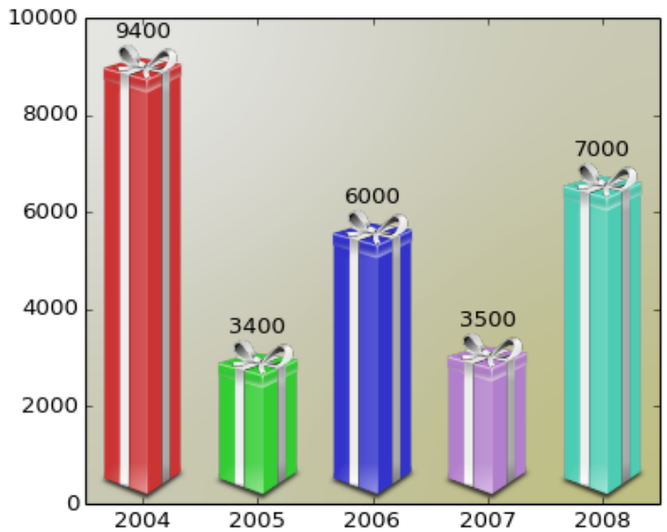


[Bateman et al. 2010]

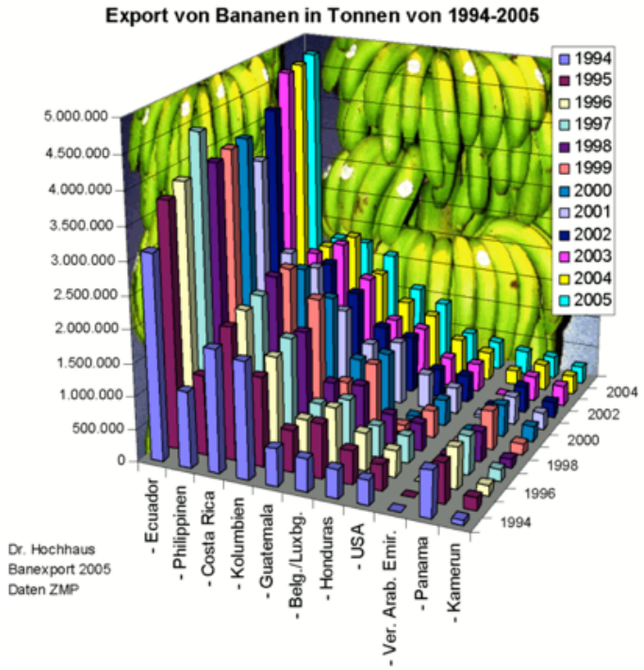
https://eagereyes.org/criticism/chart-junk-considered-useful-after-all

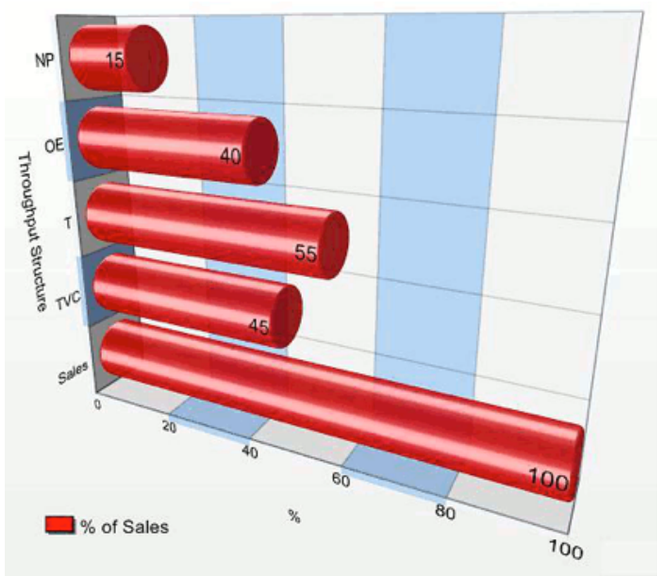
Don’t





matplotlib gallery



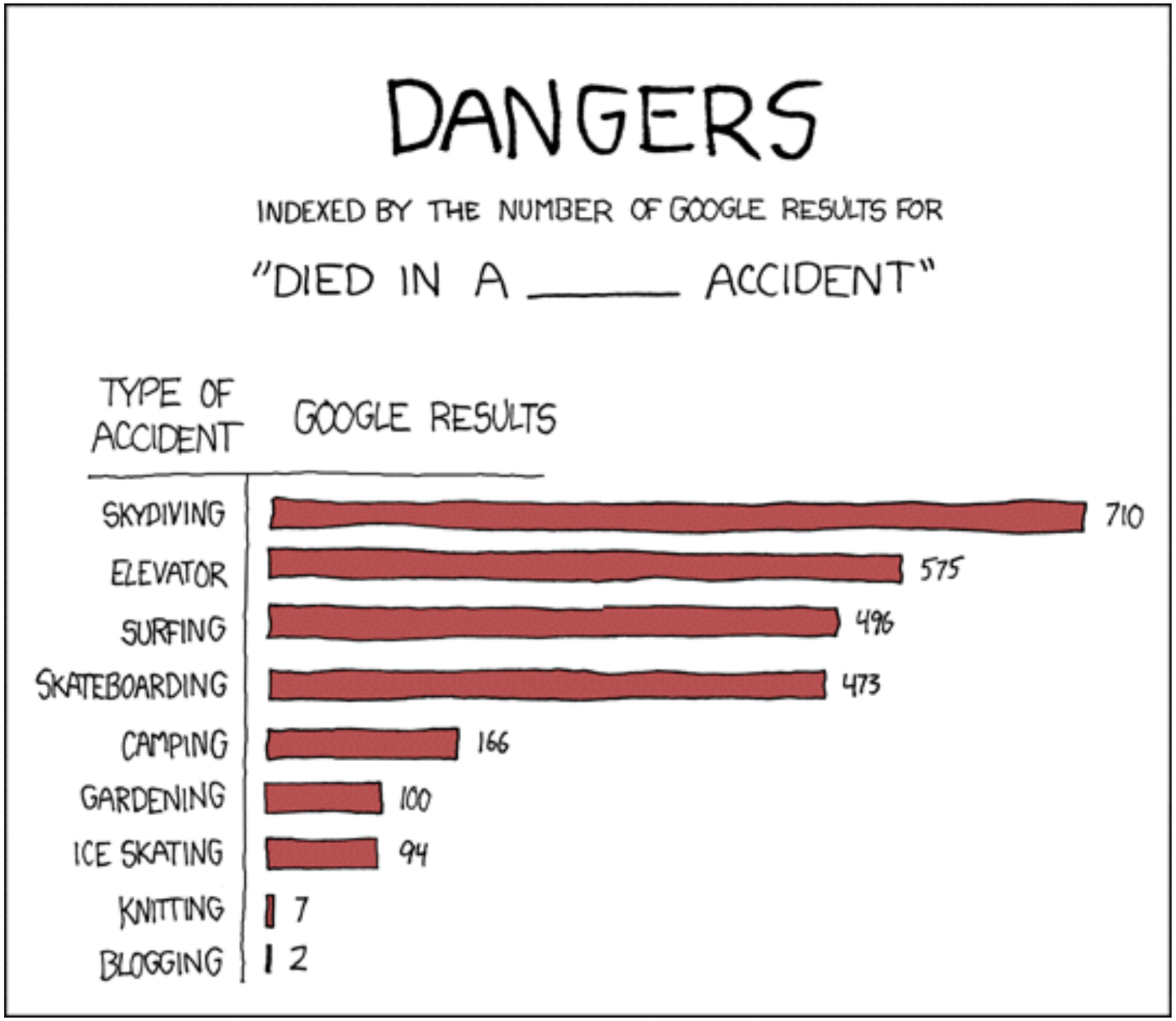


Excel Charts Blog

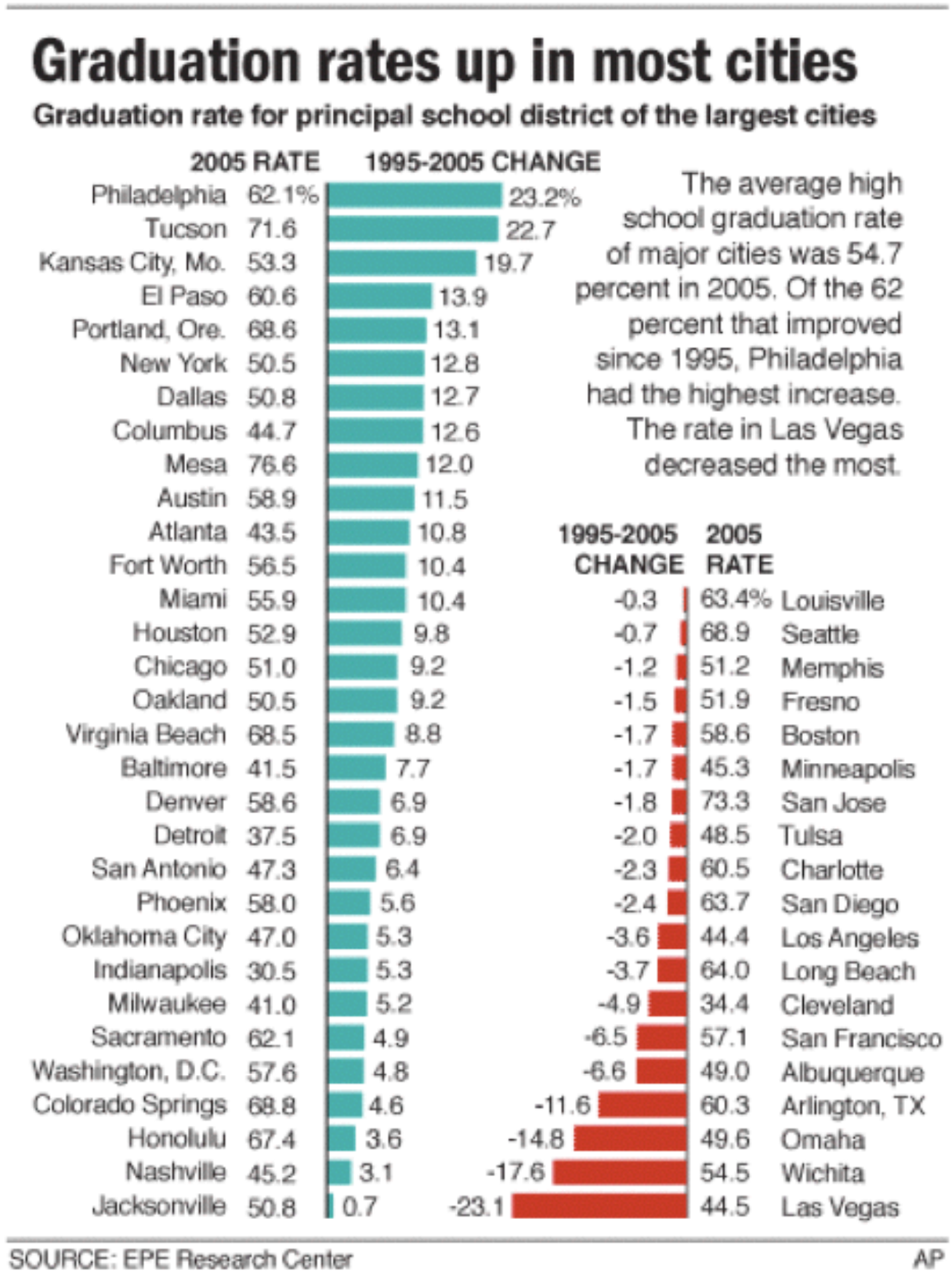
Choisir le type de graph

Comparaisons

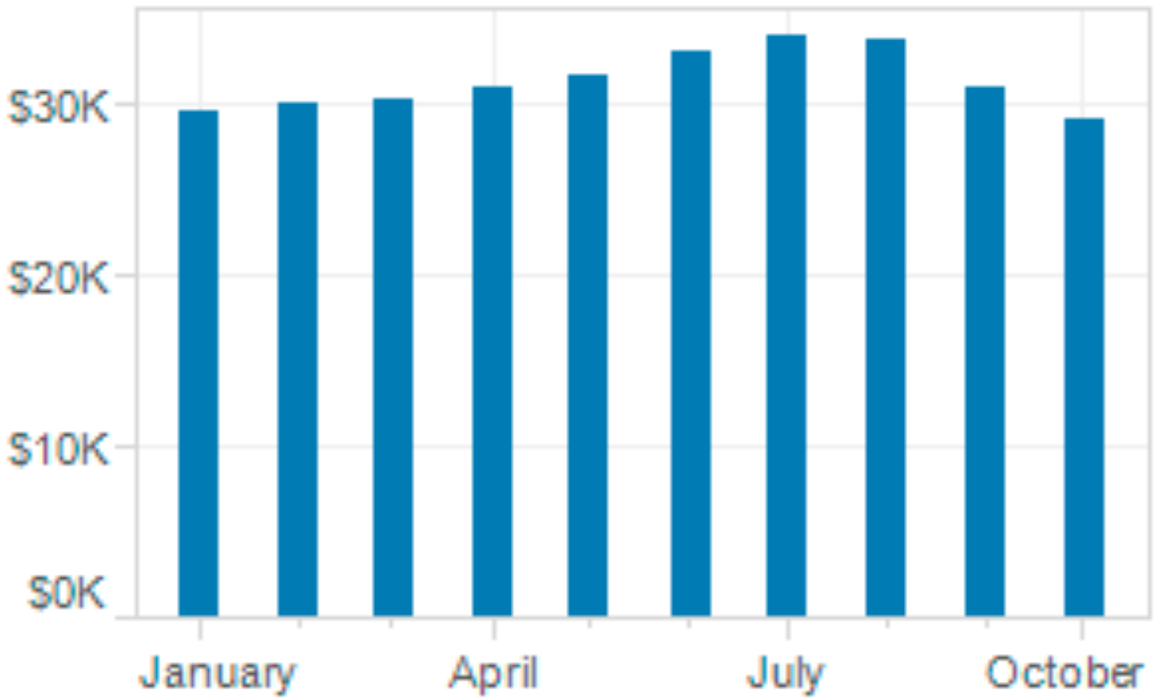
Bar Chart

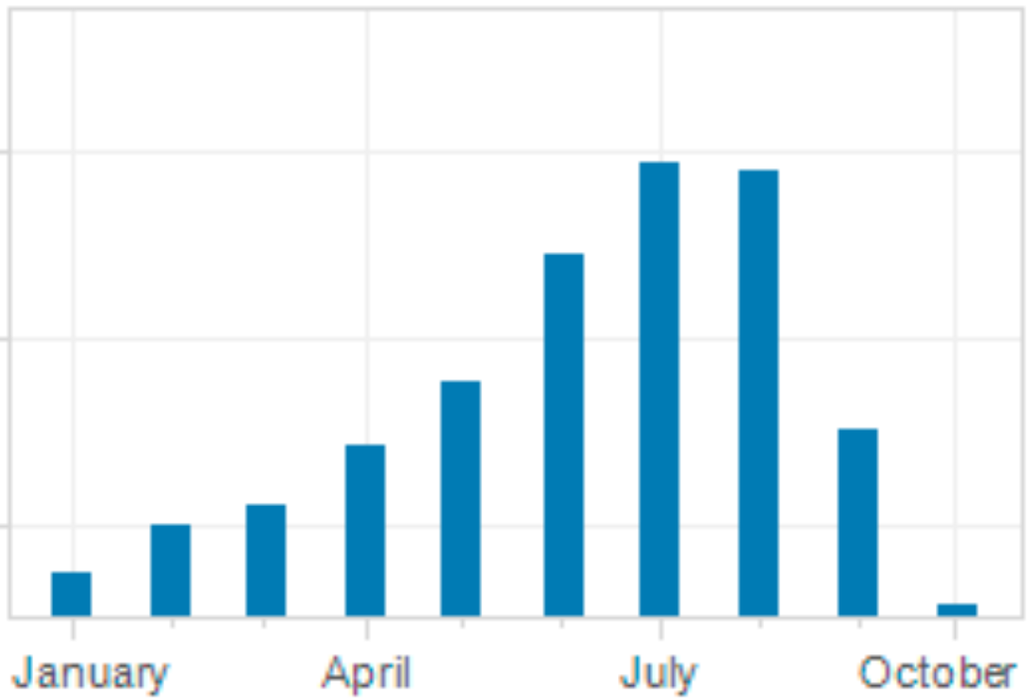

DANGERS
INDEXED BY THE NUMBER OF GOOGLE RESULTS FOR
"DIED IN A ____ ACCIDENT"
TYPE OF ACCIDENT
GOOGLE RESULTS
SKYDIVING
710
ELEVATOR
575
SURFING
496
SKATEBOARDING
473
CAMPING
166
GARDENING
100
ICE SKATING
94
KNITTING
7
BLOGGING
2

Direction


Graduation rates up in most cities
Graduation rate for principal school district of the largest cities
2005 RATE
1995-2005 CHANGE
Philadelphia 62.1% 23.2%
Tucson 71.6 22.7
Kansas City, Mo. 53.3 19.7
El Paso 60.6 13.9
Portland, Ore. 68.6 13.1
New York 50.5 12.8
Dallas 50.8 12.7
Columbus 44.7 12.6
Mesa 76.6 12.0
Austin 58.9 11.5
Atlanta 43.5 10.8
Fort Worth 56.5 10.4
Miami 55.9 10.4
Houston 52.9 9.8
Chicago 51.0 9.2
Oakland 50.5 9.2
Virginia Beach 68.5 8.8
Baltimore 41.5 7.7
Denver 58.6 6.9
Detroit 37.5 6.9
San Antonio 47.3 6.4
Phoenix 58.0 5.6
Oklahoma City 47.0 5.3
Indianapolis 30.5 5.3
Milwaukee 41.0 5.2
Sacramento 62.1 4.9
Washington, D.C. 57.6 4.8
Colorado Springs 68.8 4.6
Honolulu 67.4 3.6
Nashville 45.2 3.1
Jacksonville 50.8 0.7
The average high school graduation rate of major cities was 54.7 percent in 2005. Of the 62 percent that improved since 1995, Philadelphia had the highest increase. The rate in Las Vegas decreased the most.
1995-2005 CHANGE
2005 RATE
-0.3 63.4% Louisville
-0.7 68.9 Seattle
-1.2 51.2 Memphis
-1.5 51.9 Fresno
-1.7 58.6 Boston
-1.7 45.3 Minneapolis
-1.8 73.3 San Jose
-2.0 48.5 Tulsa
-2.3 60.5 Charlotte
-2.4 63.7 San Diego
-3.6 44.4 Los Angeles
-3.7 64.0 Long Beach
-4.9 34.4 Cleveland
-6.5 57.1 San Francisco
-6.6 49.0 Albuquerque
-11.6 60.3 Arlington, TX
-14.8 49.6 Omaha
-17.6 54.5 Wichita
-23.1 44.5 Las Vegas
SOURCE: EPE Research Center
AP

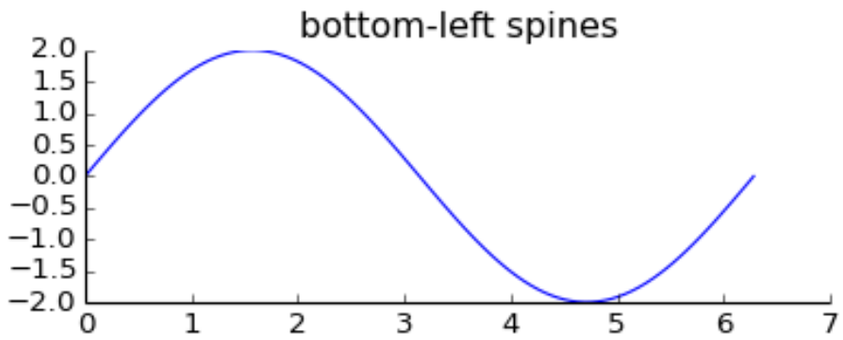
attention à la baseline

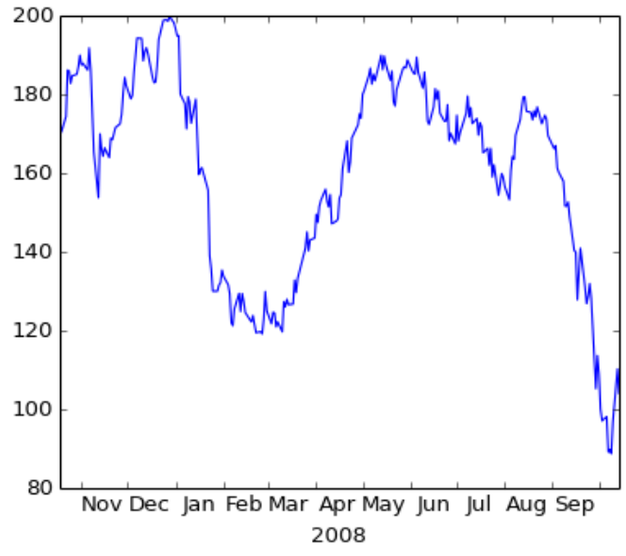


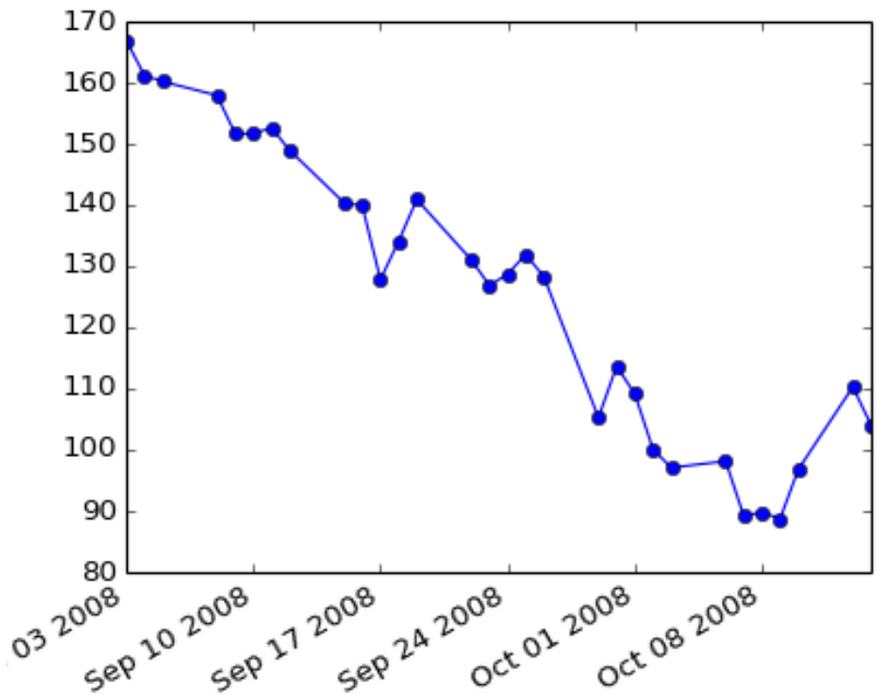


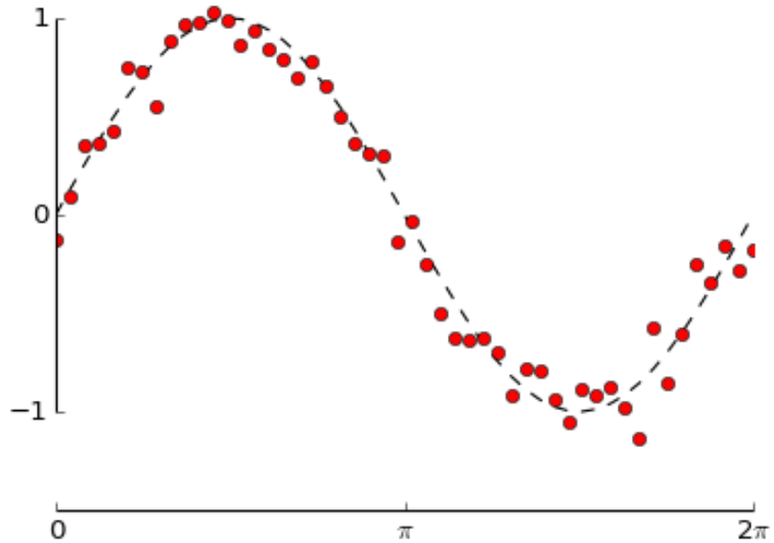
https://eagereyes.org/basics/ba

Line Chart







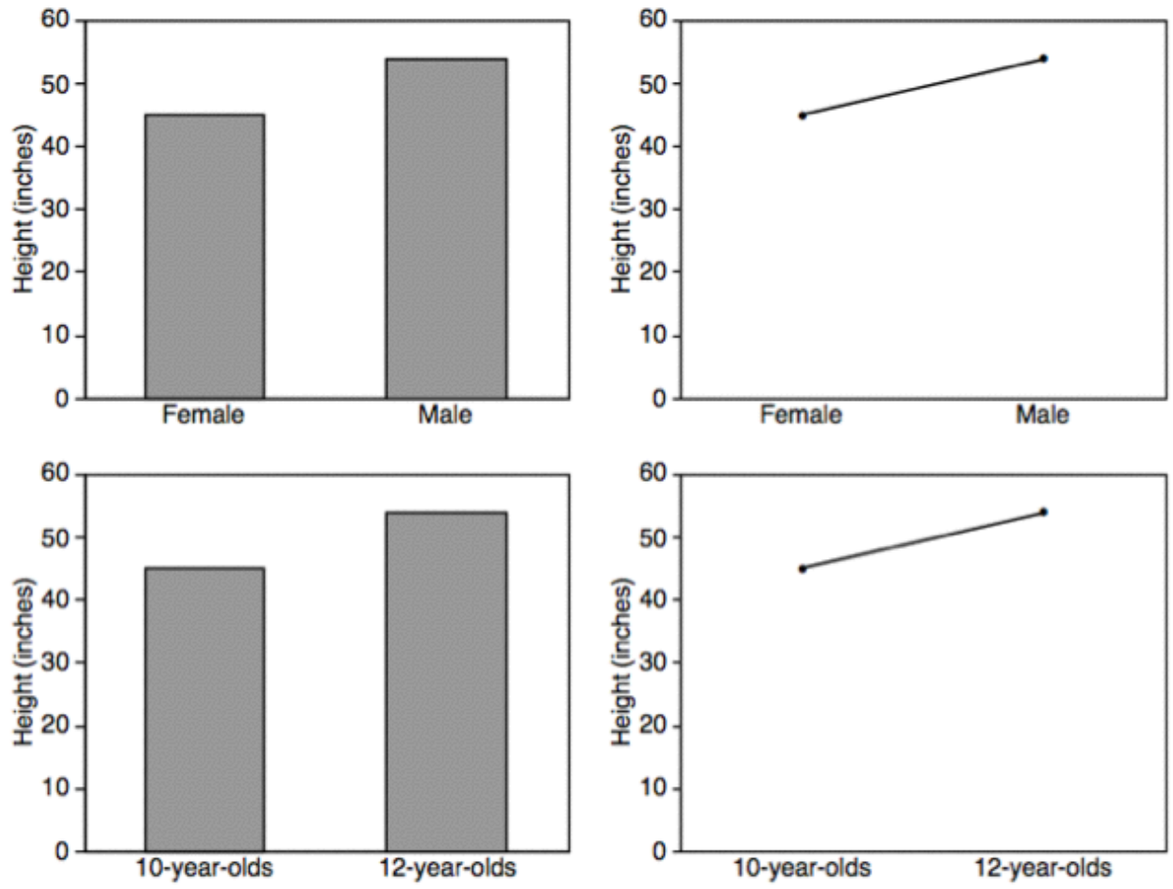


matplotlib gallery

Bars vs. Lines

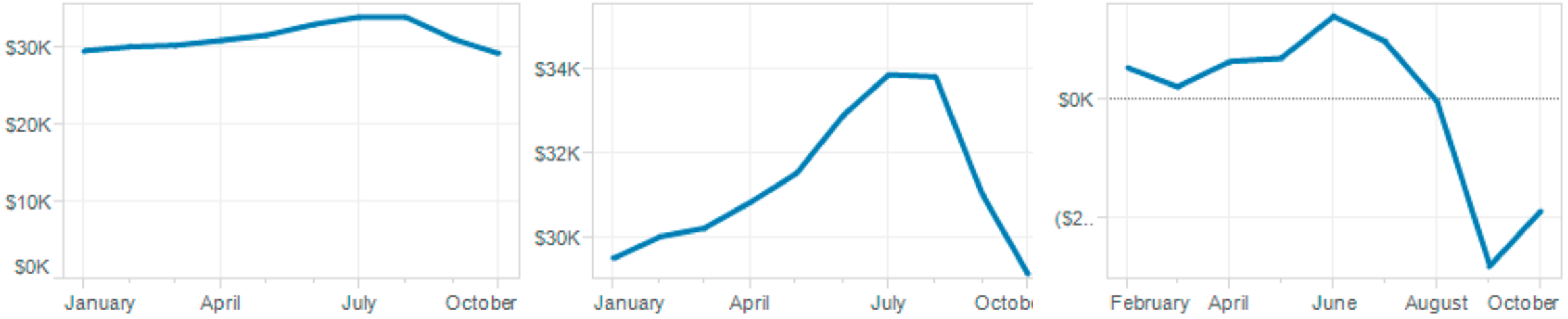
Lines imply connections & sampling from continuous data.

Do not use for categorical data.



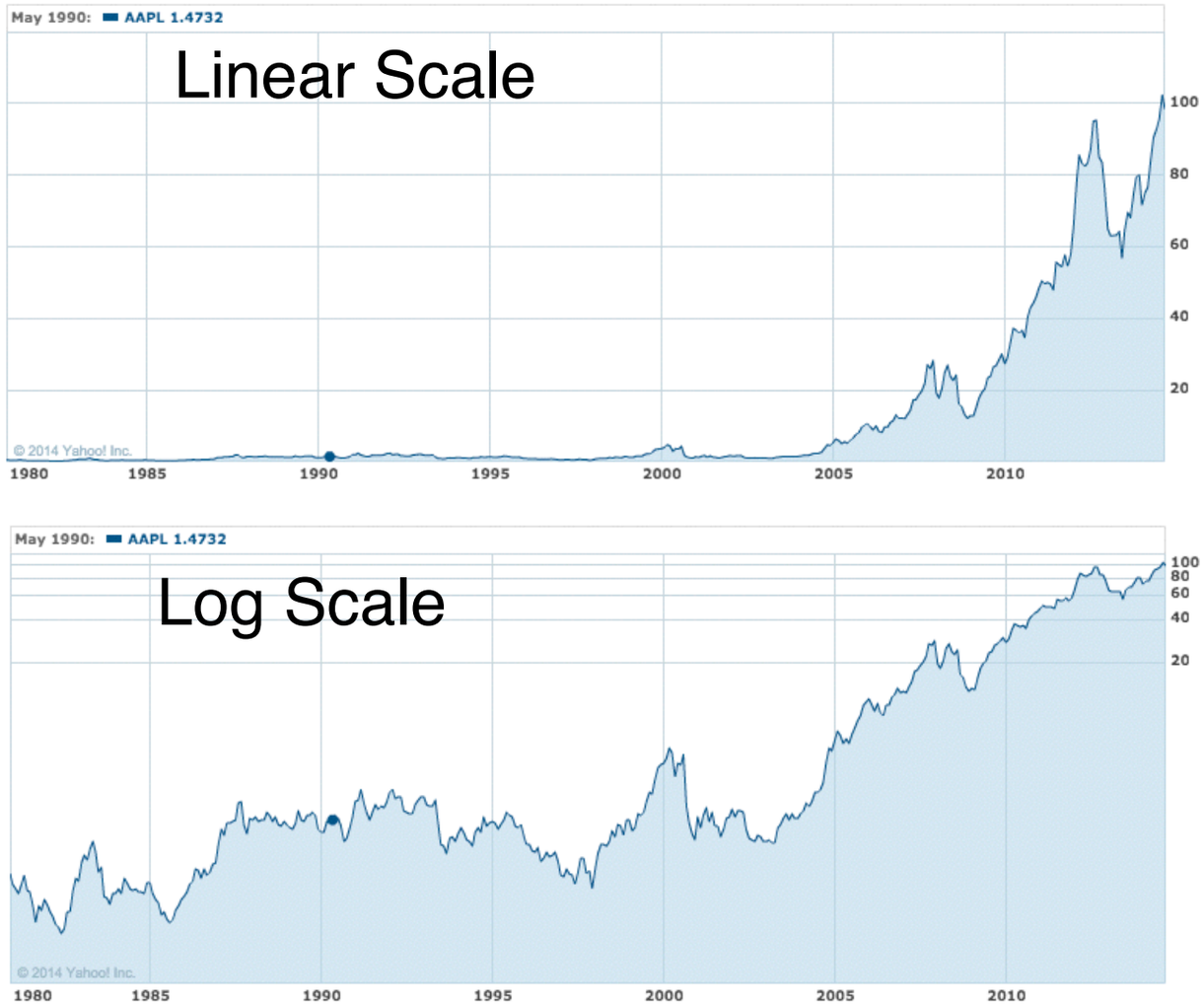
Zacks 1999

Baseline Problem (again)



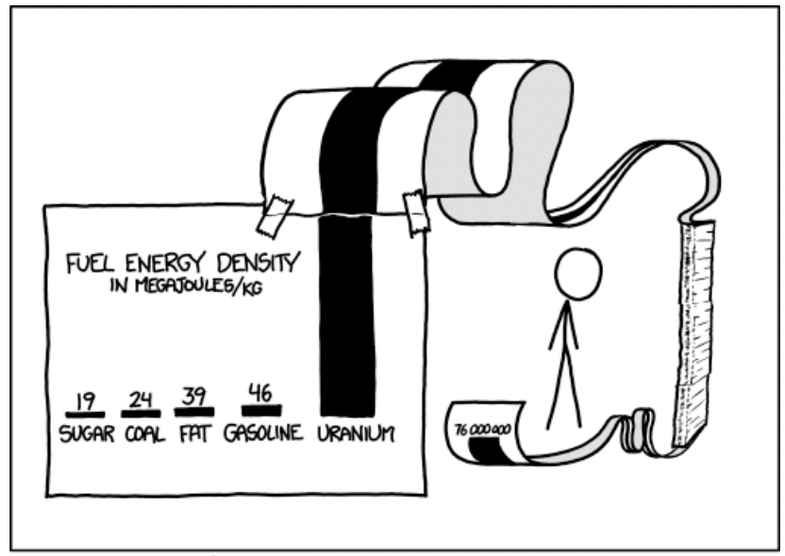
https://eagereyes.org/basics/baselines

Linear vs. Logarithmic Scale



Apple Stock Price

http://finance.yahoo.com/echarts?s=AAPL



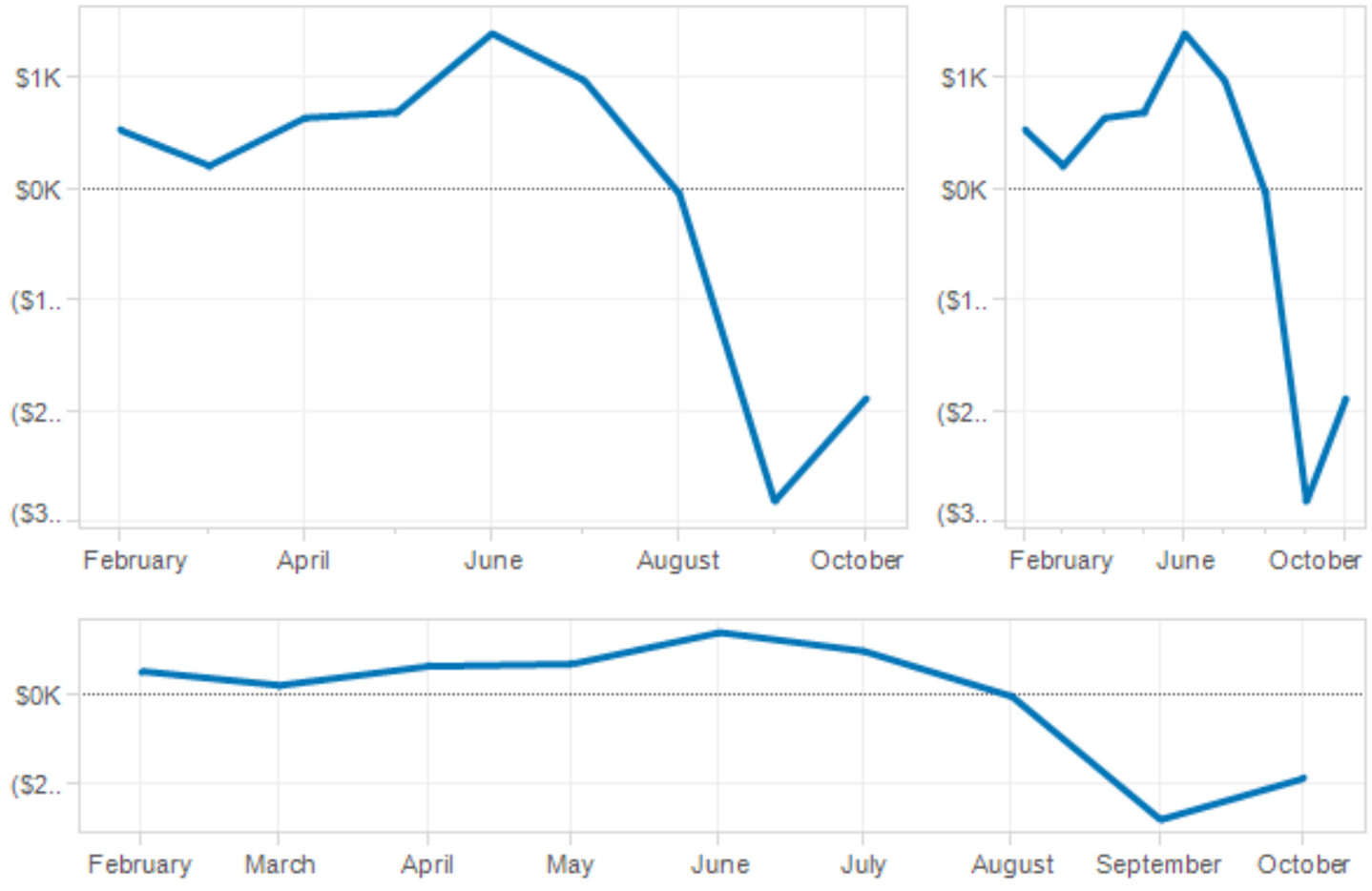
http://xkcd.com/1162/

Aspect Ratios

Rule of Thumb:

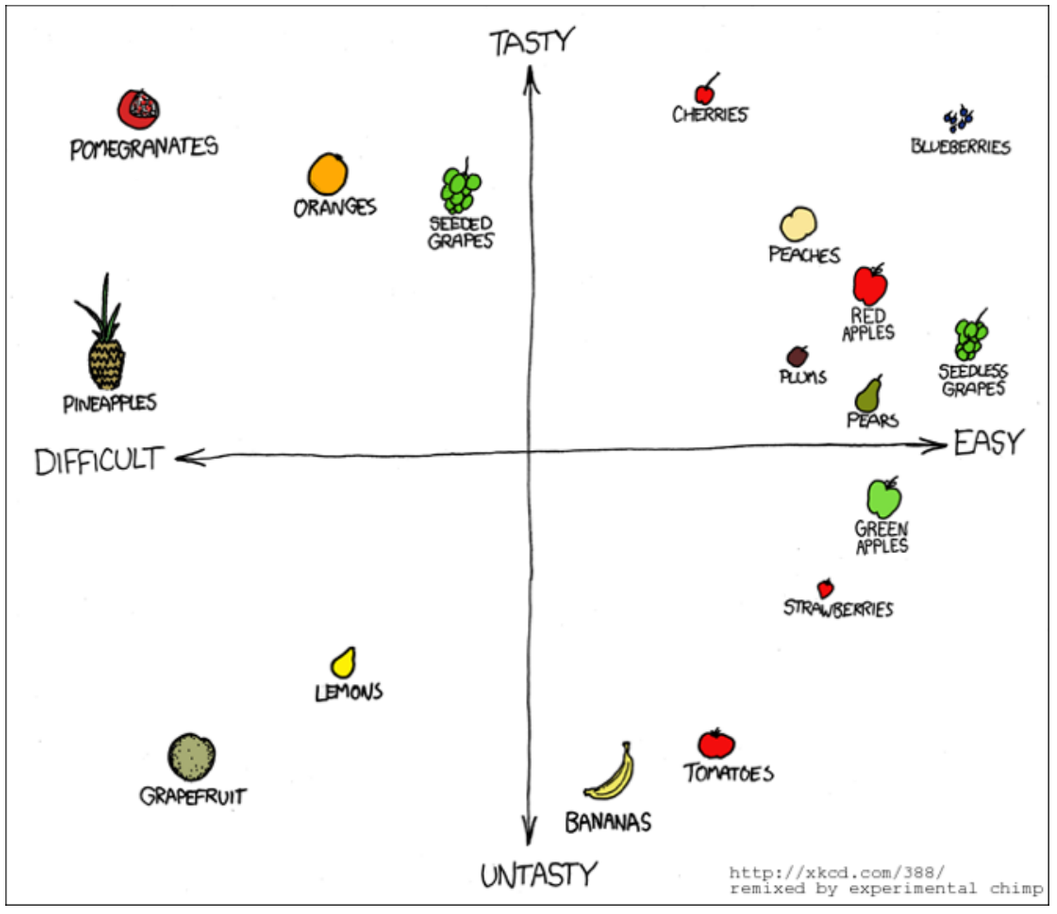
Banking to 45°

(average line slope: 45°)



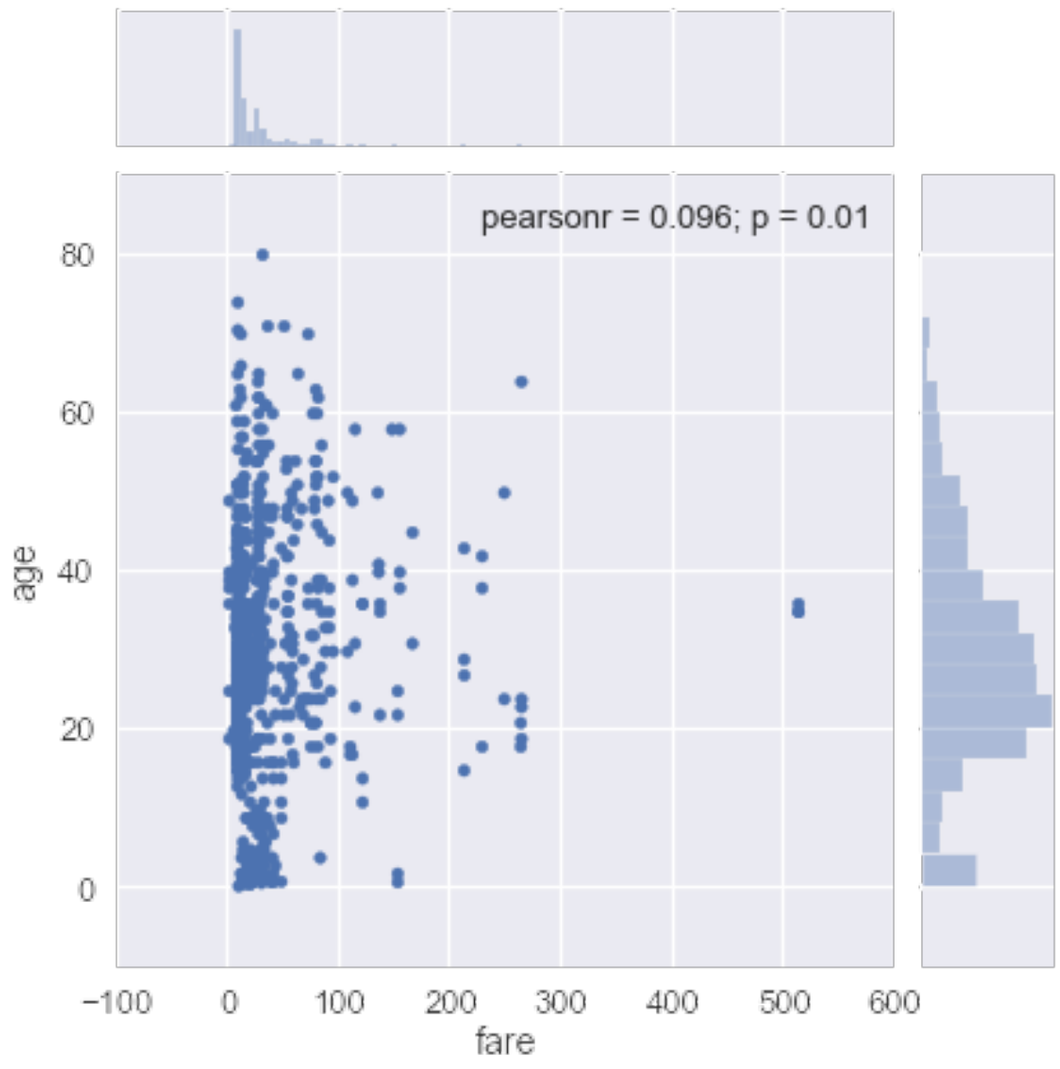
eagereyes.org

Correlations



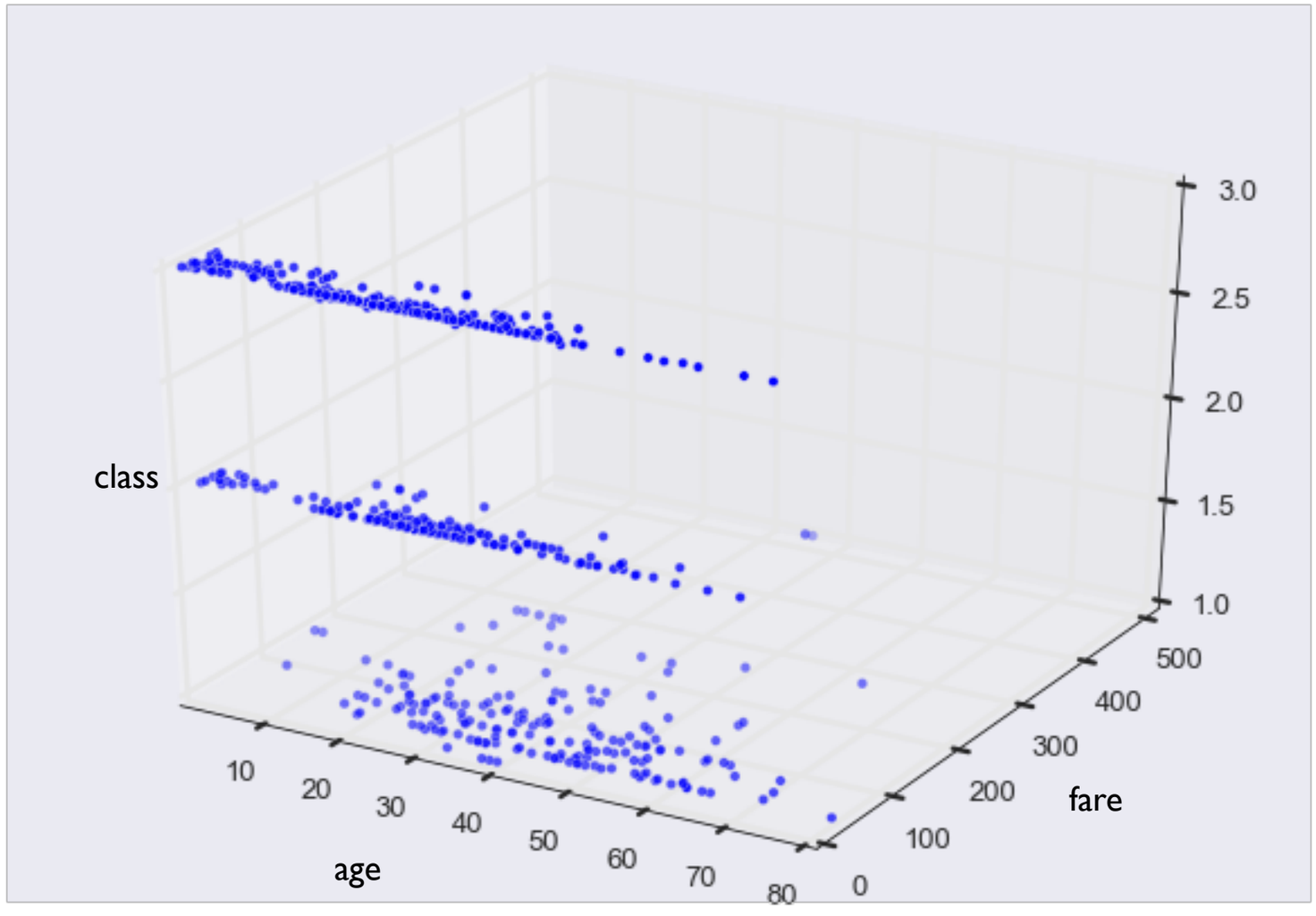
age	fare
22.0	7.25
38.0	71.2833
26.0	7.925
35.0	53.1
35.0	8.05
	8.4583
54.0	51.8625
2.0	21.075
27.0	11.1333
14.0	30.0708
4.0	16.7
58.0	26.55
20.0	8.05
39.0	31.275
14.0	7.8542
55.0	16.0
2.0	29.125
	13.0
31.0	18.0
	7.225
35.0	26.0
34.0	13.0
15.0	8.0292

Scatterplots



age	fare	class
22.0	7.25	Third
38.0	71.2833	First
26.0	7.925	Third
35.0	53.1	First
35.0	8.05	Third
	8.4583	Third
54.0	51.8625	First
2.0	21.075	Third
27.0	11.1333	Third
14.0	30.0708	Second
4.0	16.7	Third
58.0	26.55	First
20.0	8.05	Third
39.0	31.275	Third
14.0	7.8542	Third
55.0	16.0	Second
2.0	29.125	Third
	13.0	Second
31.0	18.0	Third
	7.225	Third
35.0	26.0	Second
34.0	13.0	Second
15.0	8.0292	Third

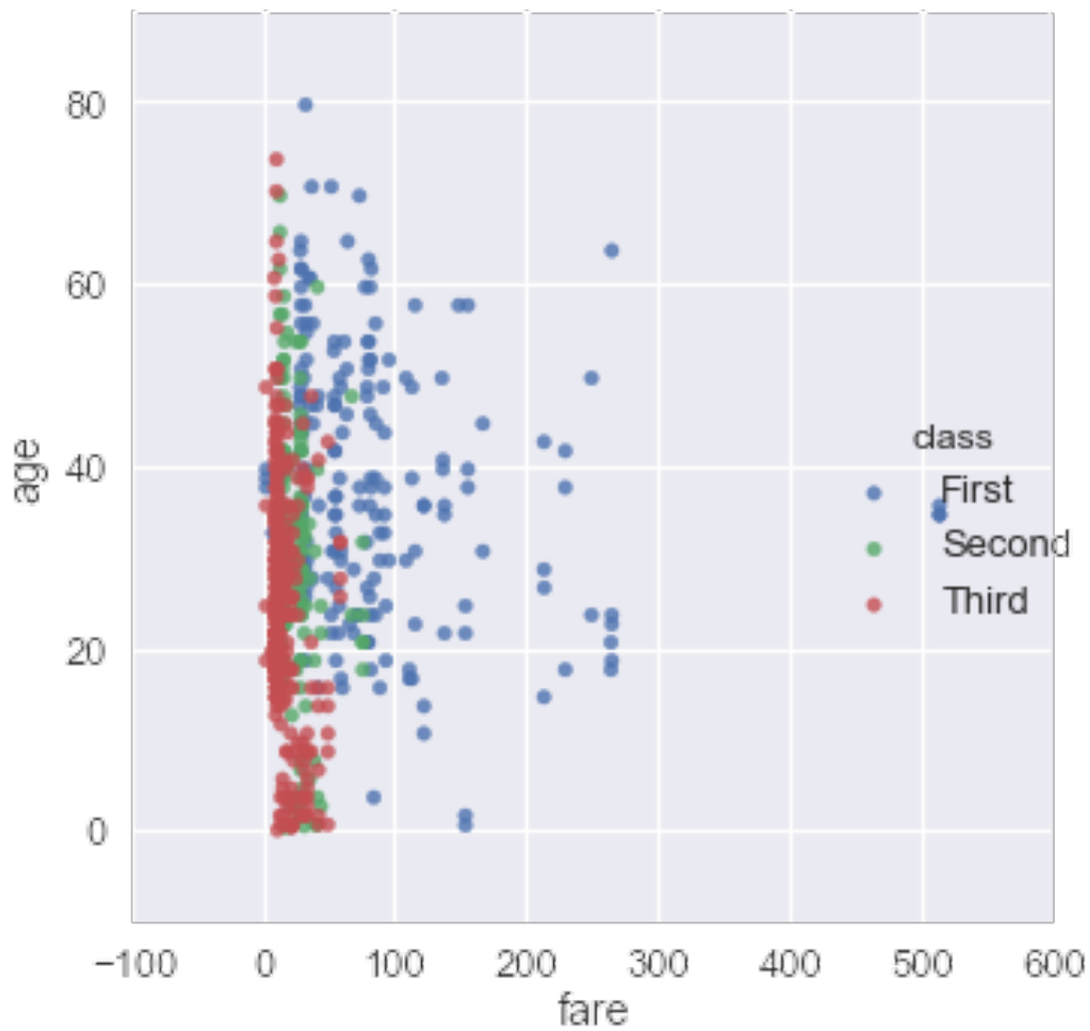
Trivariate Data



Do NOT use 3D scatterplots!

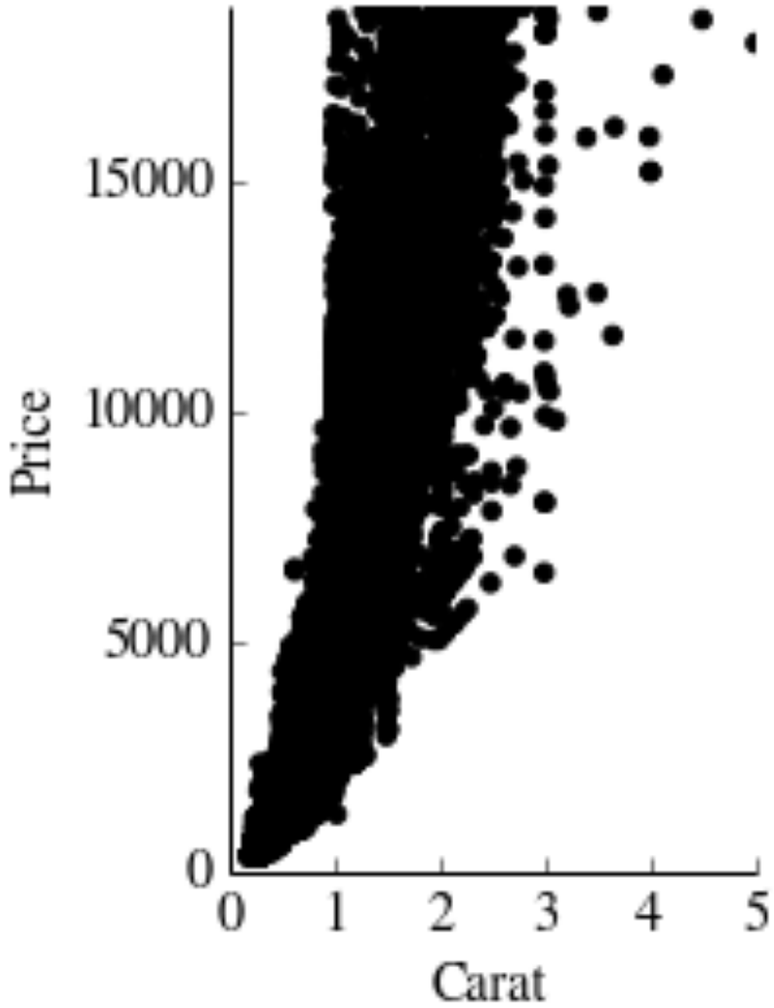
age	fare	class
22.0	7.25	Third
38.0	71.2833	First
26.0	7.925	Third
35.0	53.1	First
35.0	8.05	Third
	8.4583	Third
54.0	51.8625	First
2.0	21.075	Third
27.0	11.1333	Third
14.0	30.0708	Second
4.0	16.7	Third
58.0	26.55	First
20.0	8.05	Third
39.0	31.275	Third
14.0	7.8542	Third
55.0	16.0	Second
2.0	29.125	Third
	13.0	Second
31.0	18.0	Third
	7.225	Third
35.0	26.0	Second
34.0	13.0	Second
15.0	8.0292	Third

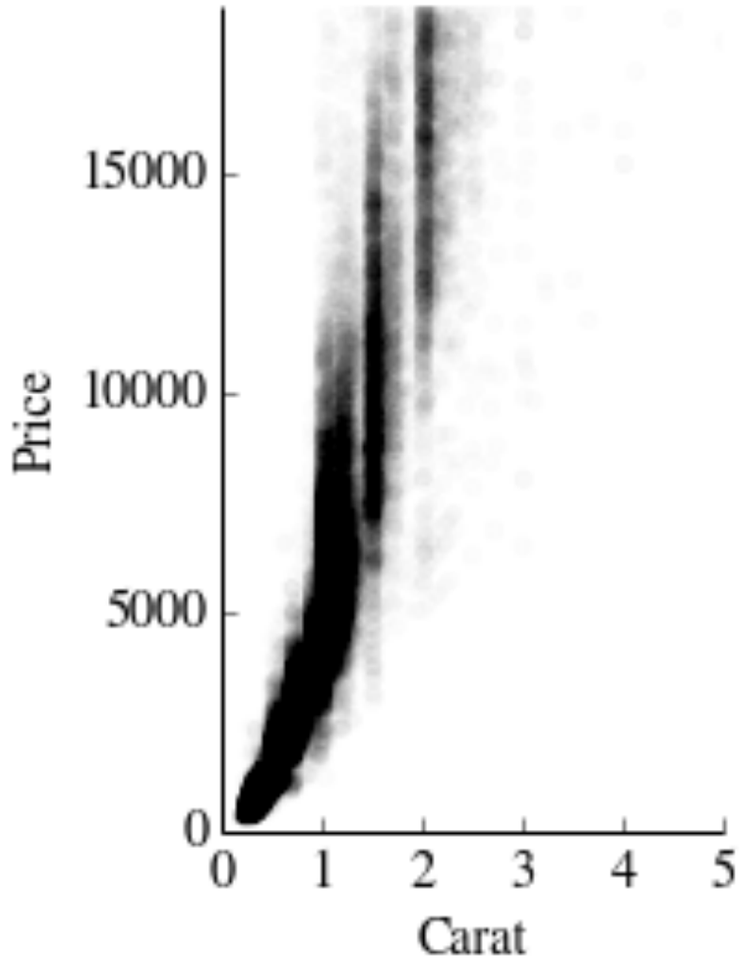
Trivariate Data



Map the third dimension to some other visual attribute

Overplotting

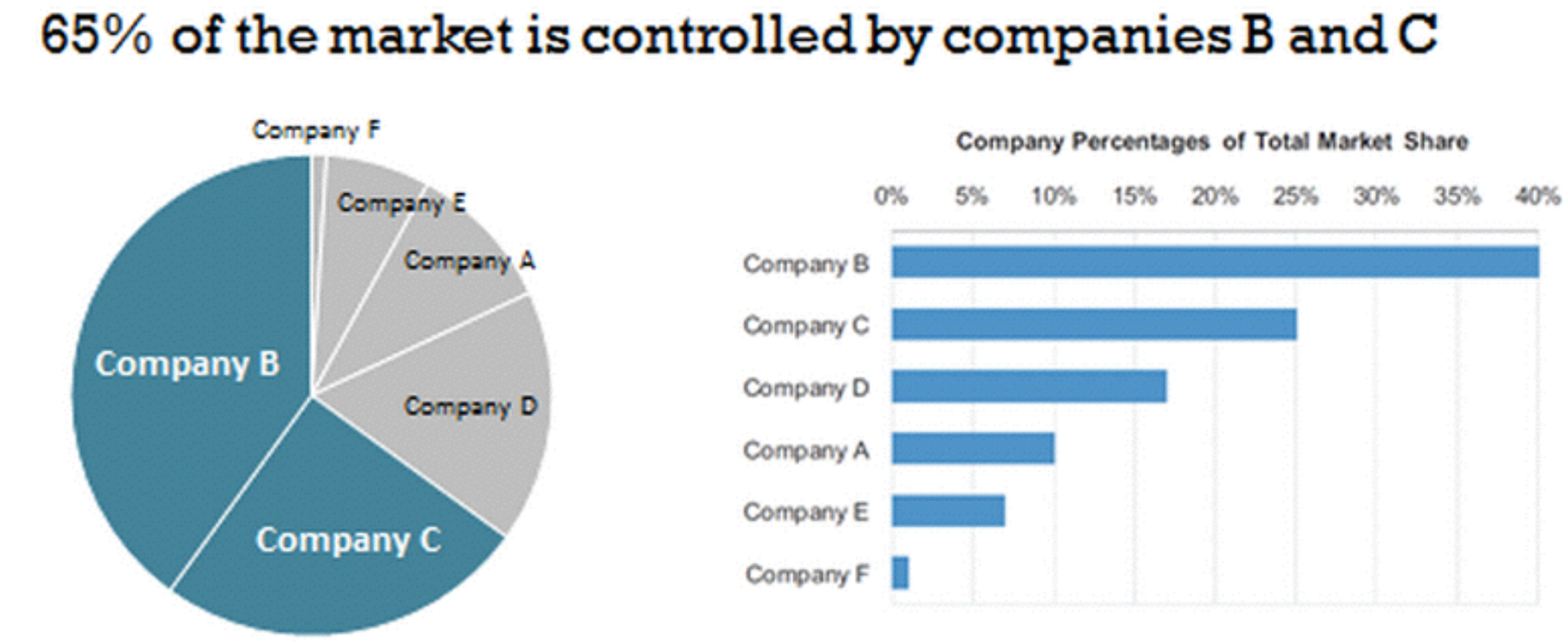




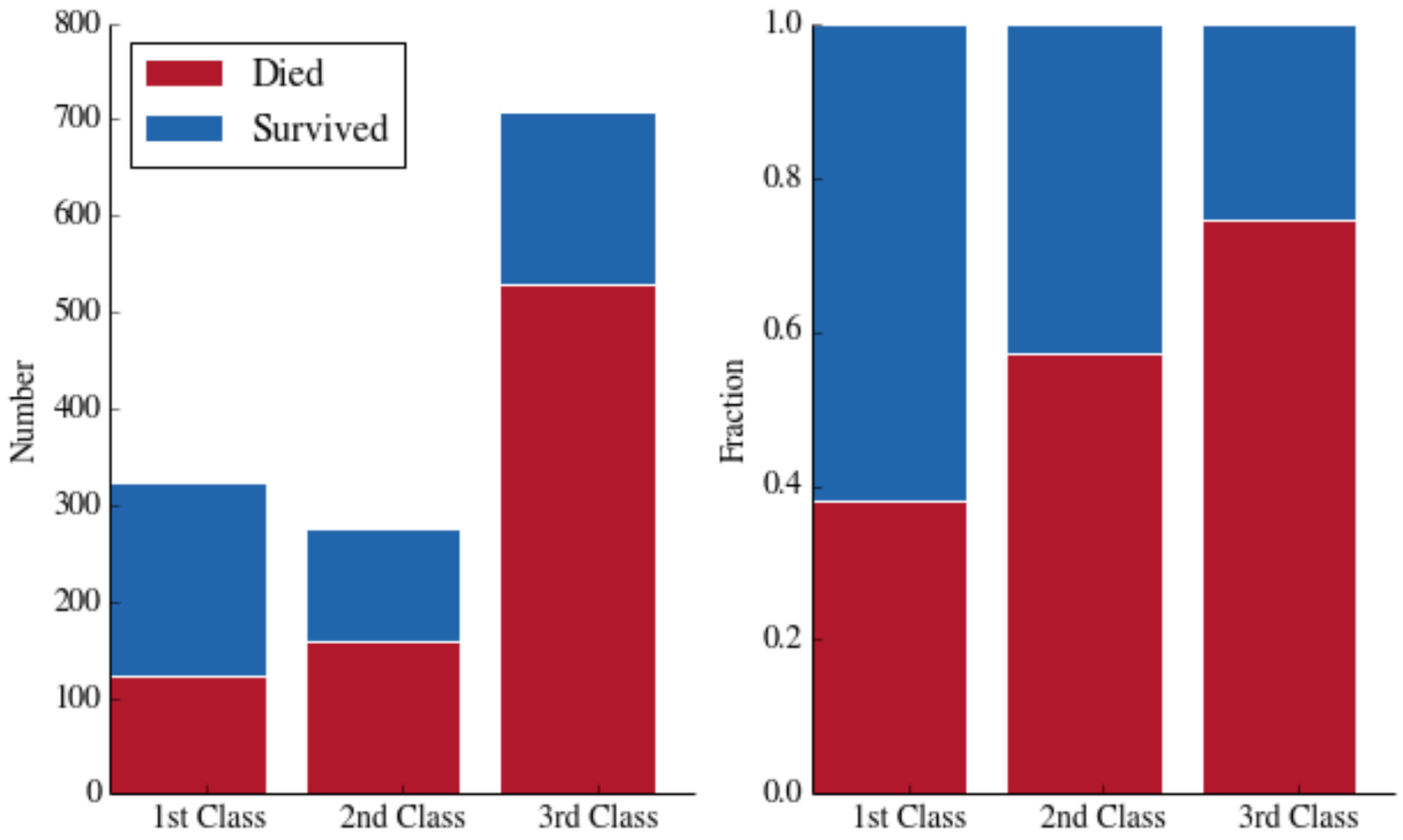
alpha = 1/100

Compositions

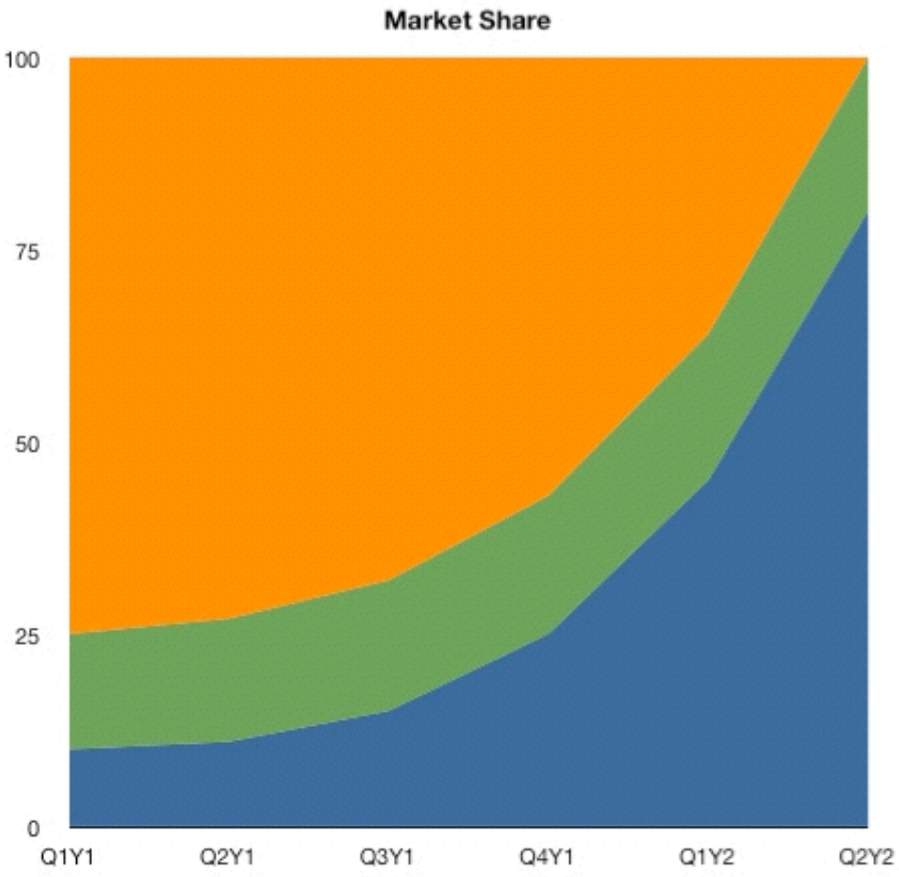
Pie vs. Bar Charts

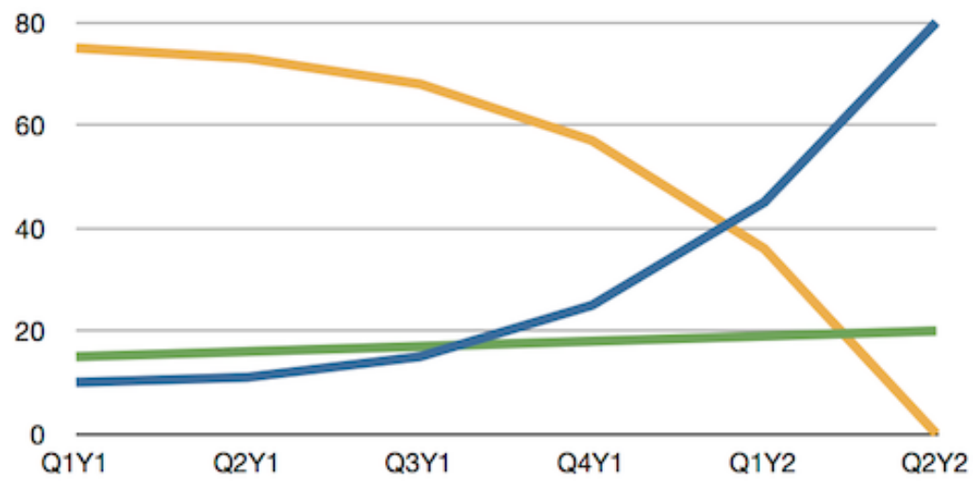


Stacked Bar Chart



Stacked Area vs. Line Graphs





leancrew.com &
Practically Efficient

Distributions

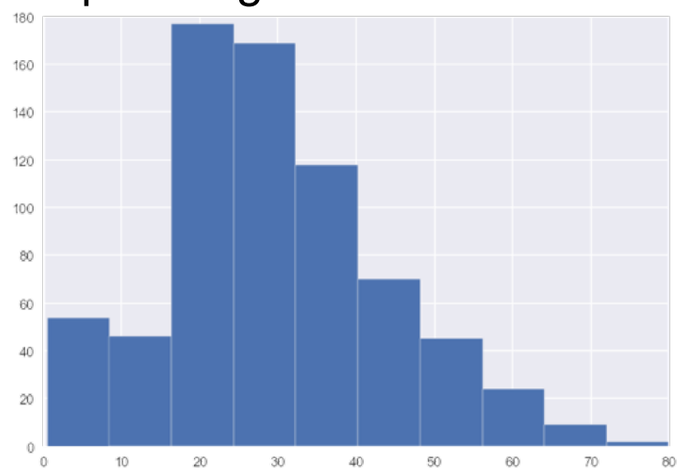
Histogram

#bins hard to predict

make interactive!

rule of thumb: #bins = sqrt(n)

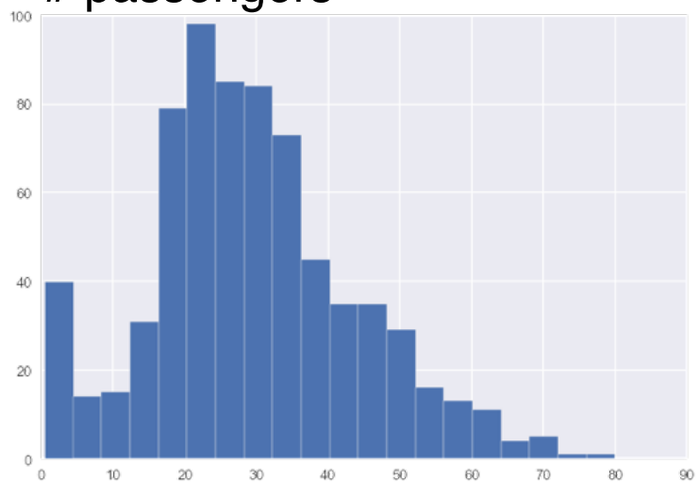
passengers



age

10 Bins

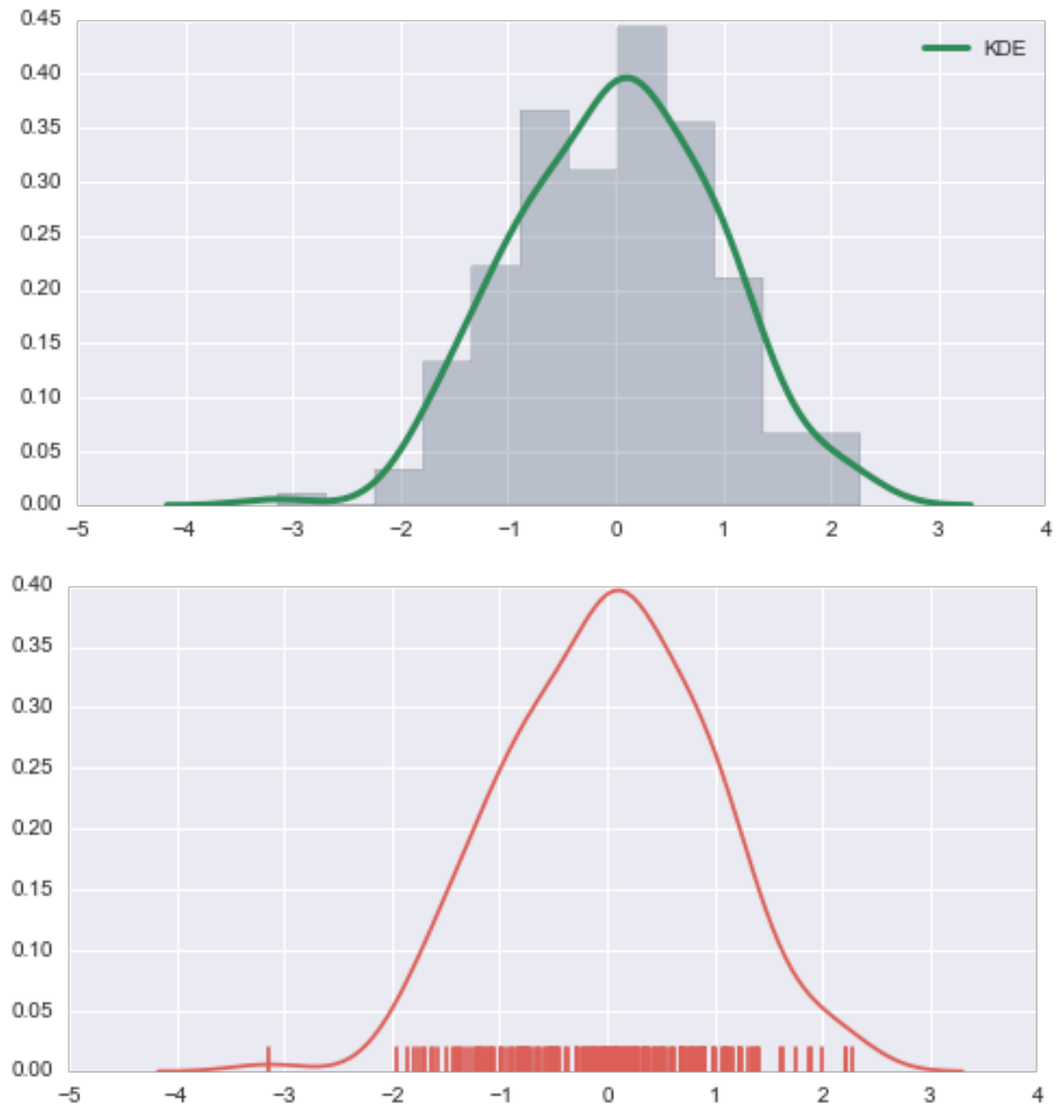
passengers



age

20 Bins

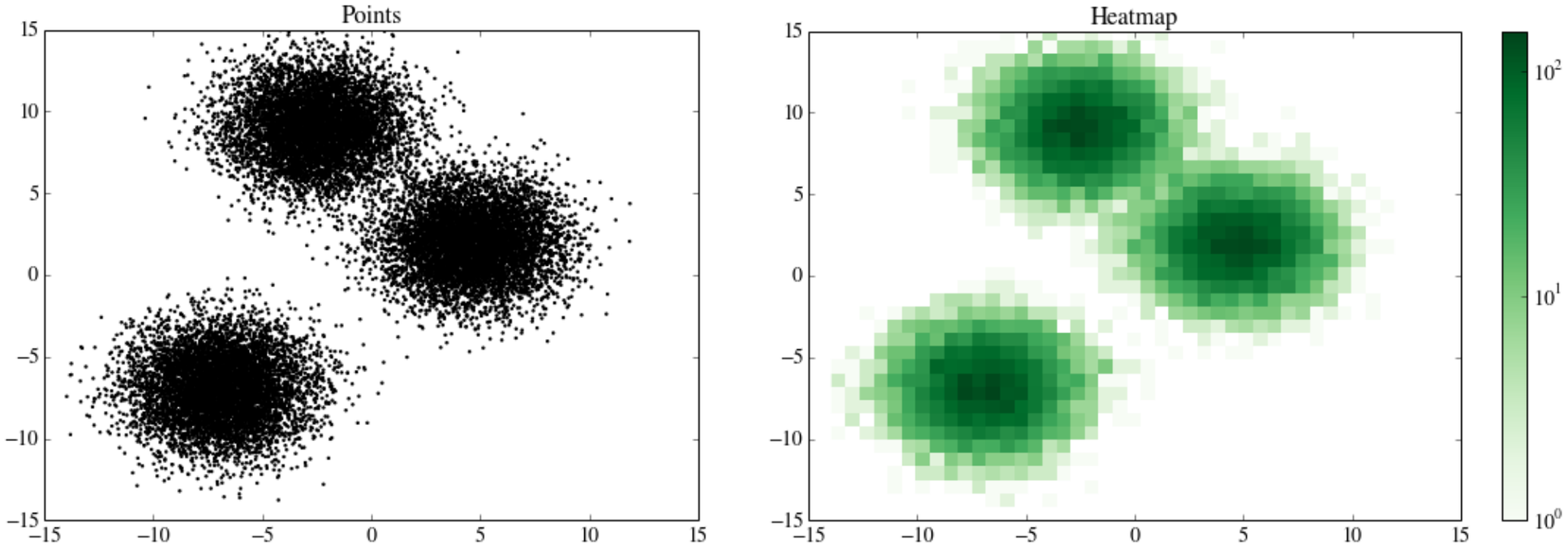
Density Plots



http://web.stanford.edu/~mwaskom/software/seaborn/tutorial/plotting_distributions.html

Heat Maps

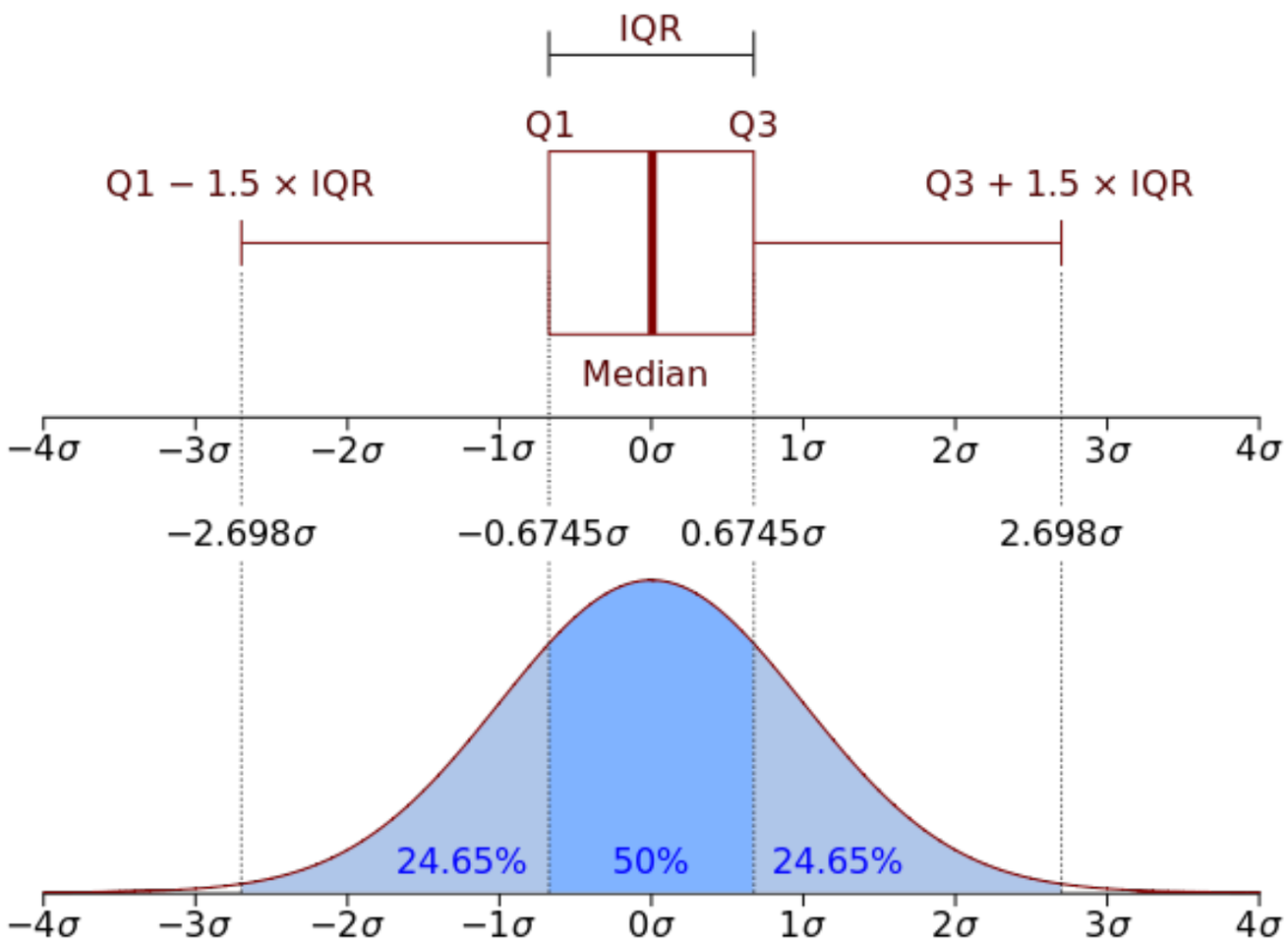
binning of scatterplots



2D Density Plots

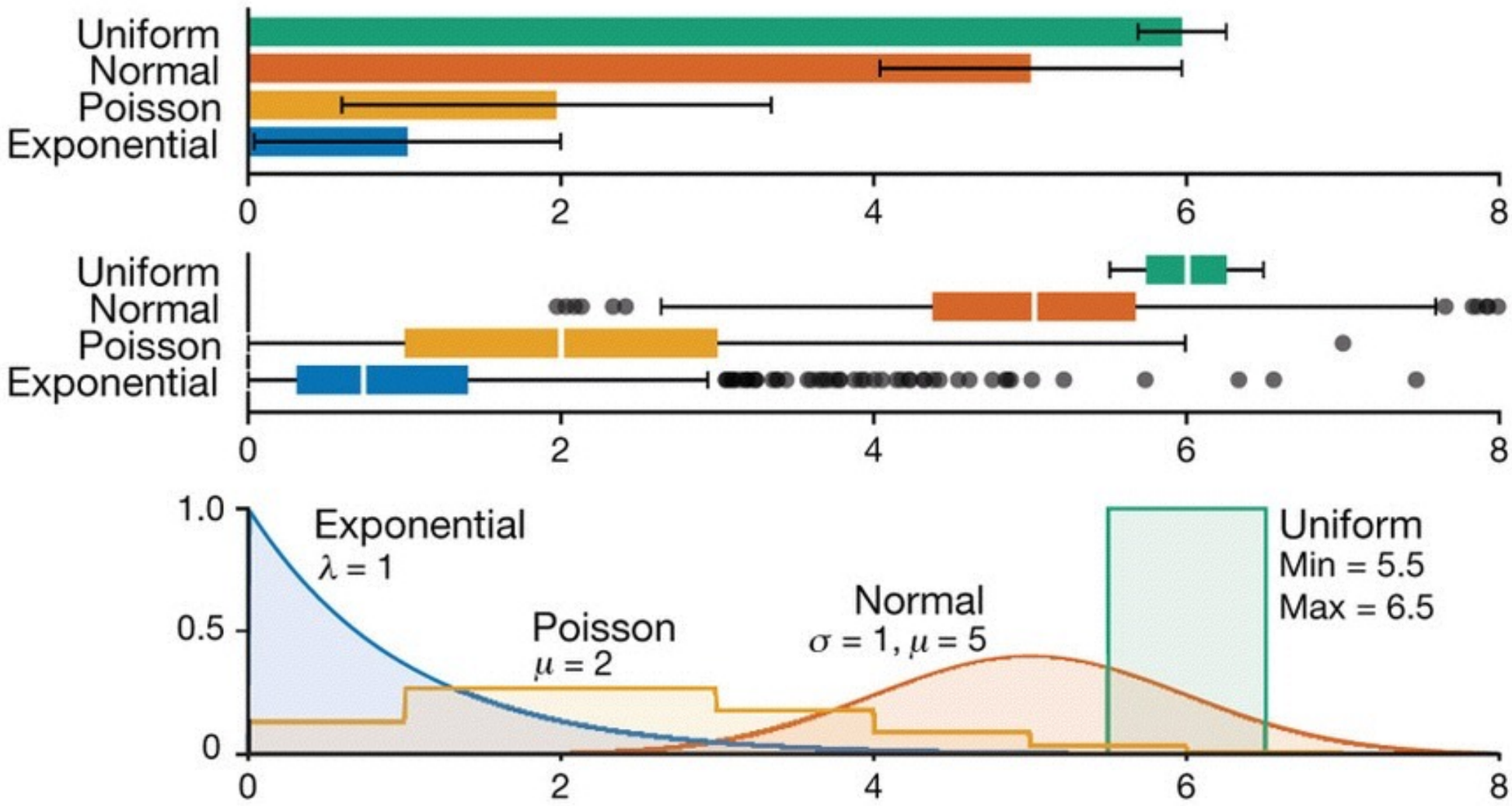
Box Plots

aka Box-and-Whisker Plot



Wikipedia

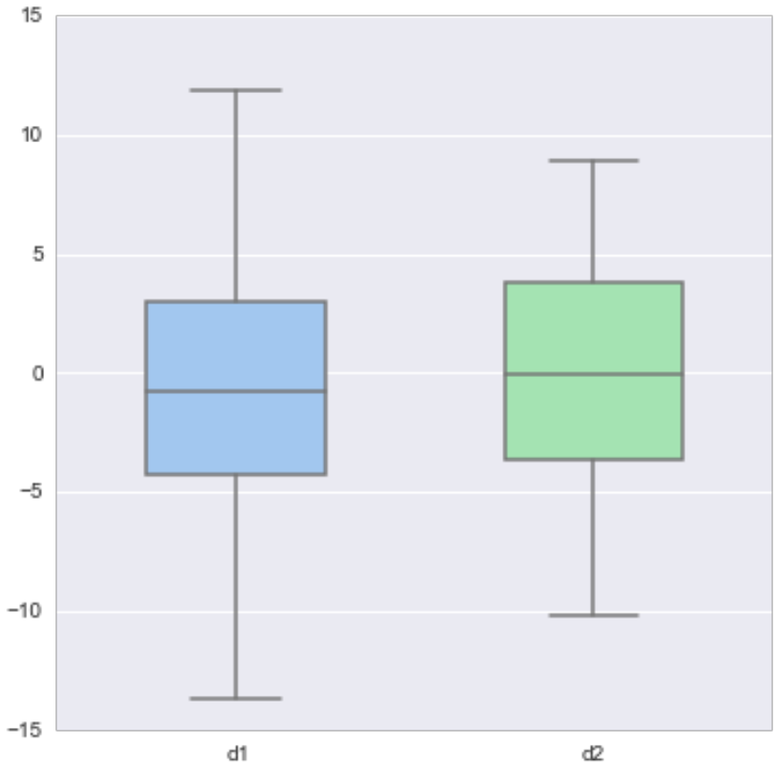
Comparison

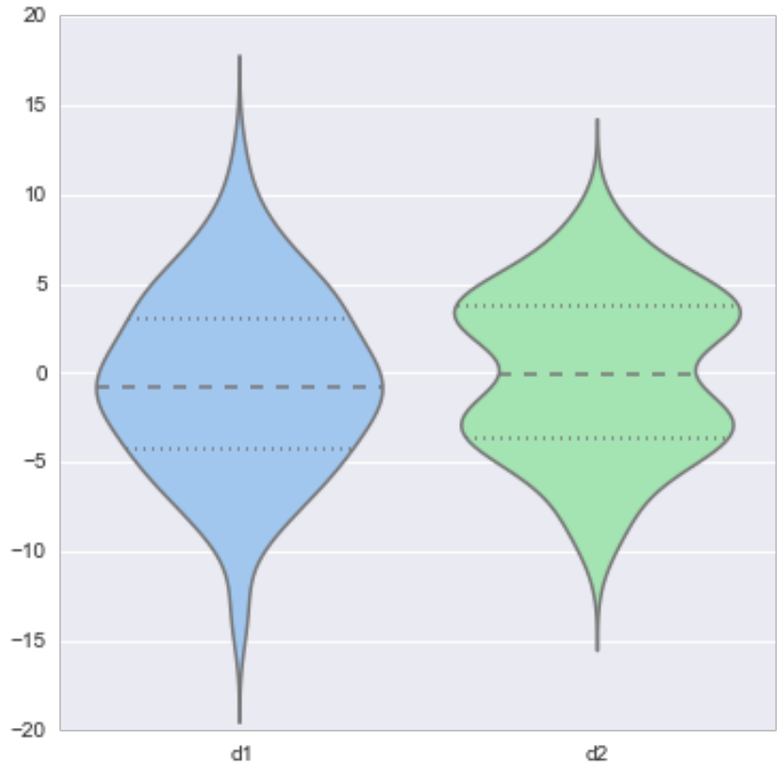


Streit & Gehlenborg, PoV, Nature Methods, 2014

Violin Plot

= Box Plot + Probability Density Function





http://web.stanford.edu/~mwaskom/software/seaborn/tutorial/

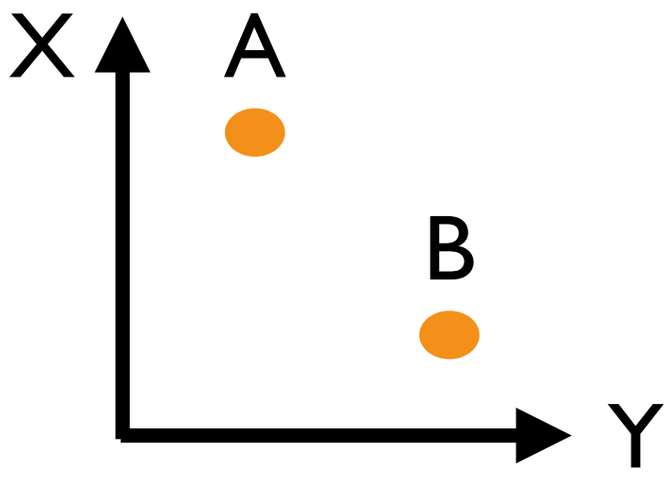
Highdimensional Data

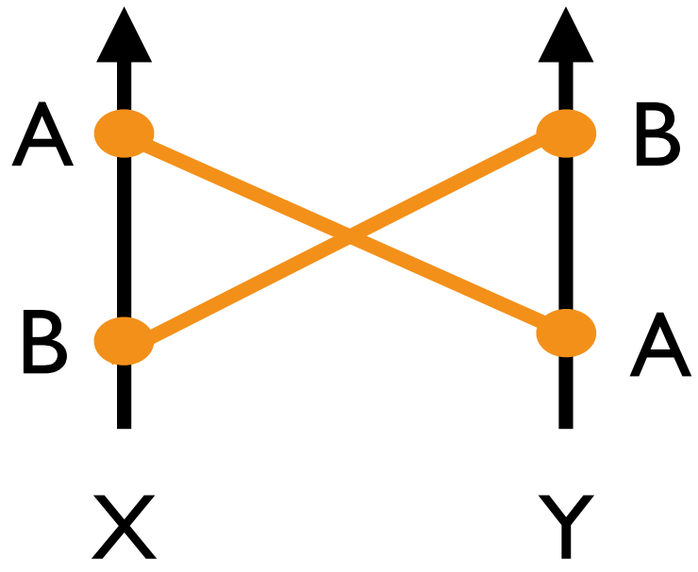
Parallel Coordinates (PC)

Inselberg 1985

Axes represent attributes

Lines connecting axes represent items





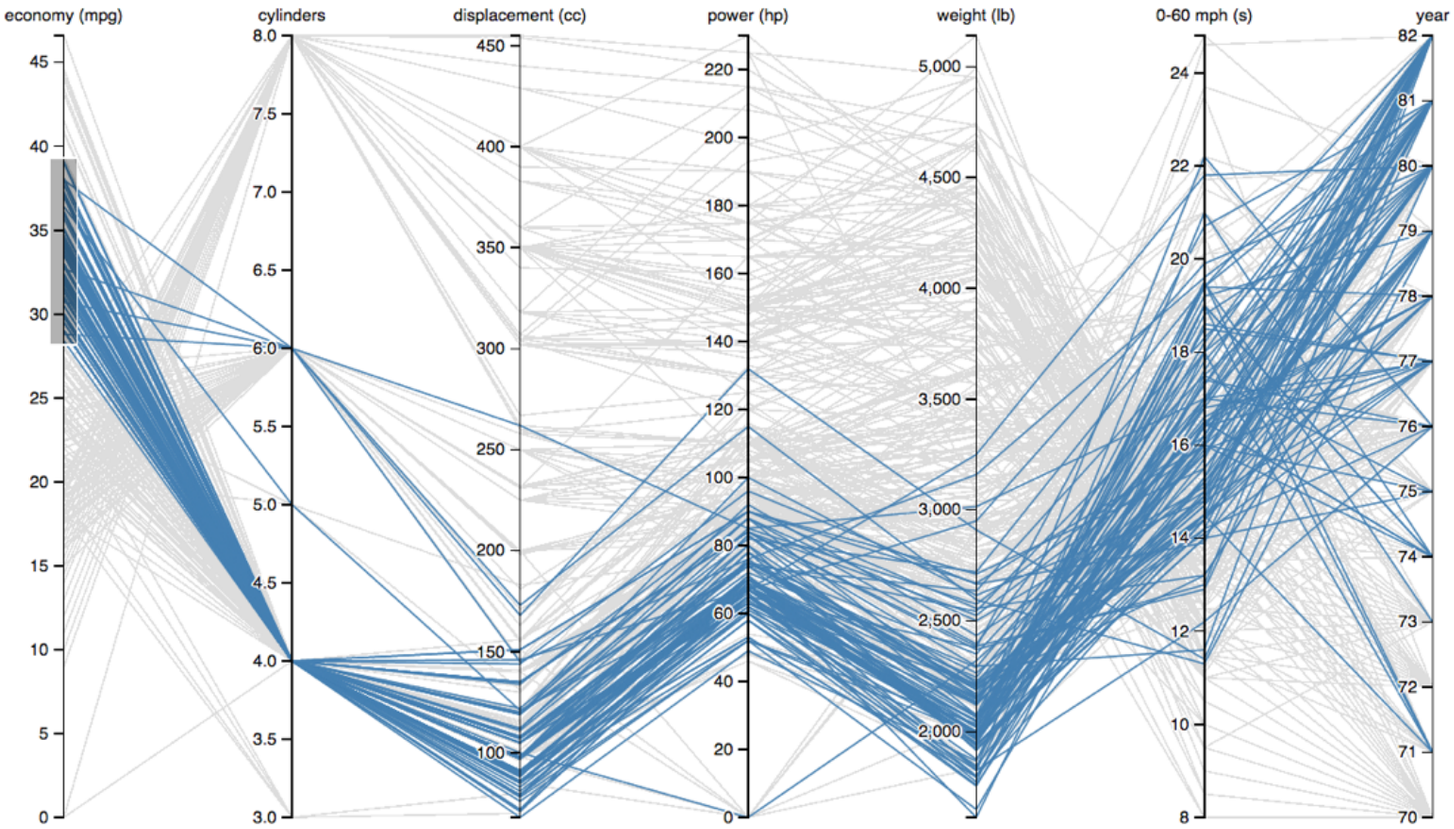
Parallel Coordinates

Each axis represents dimension

Lines connecting axis represent records

Suitable for

- all tabular data types
- heterogeneous data



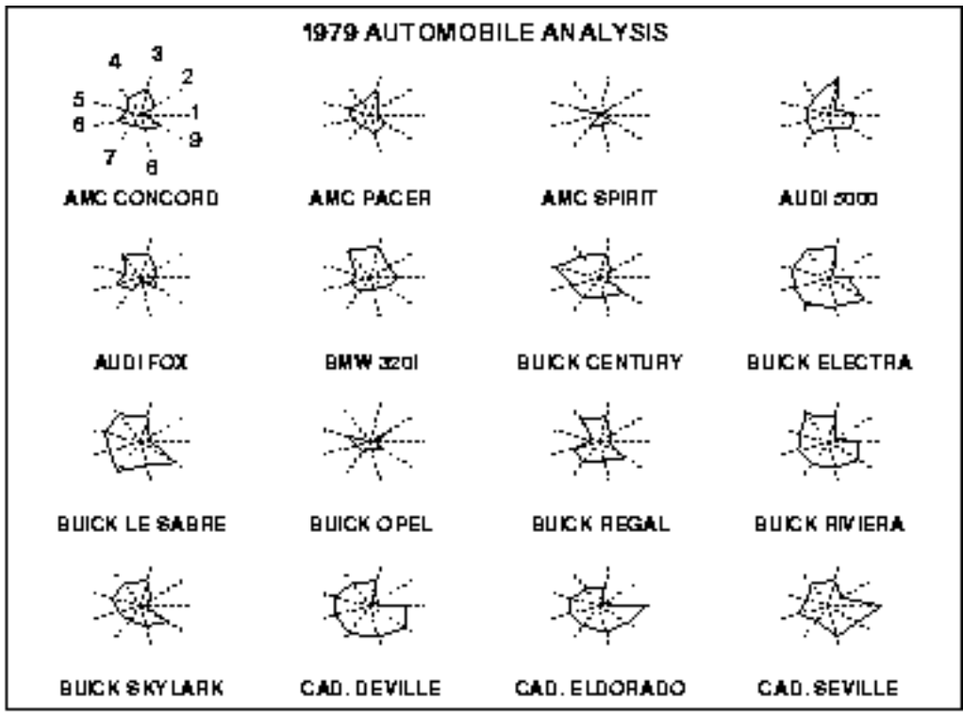
Star Plot

[Coekin1969]

Similar to parallel coordinates

Radiate from a common origin

http://start1.jpl.nasa.gov/caseStudies/autoTool.cfm



http://www.itl.nist.gov/div898/handbook/eda/section3/starplot.htm

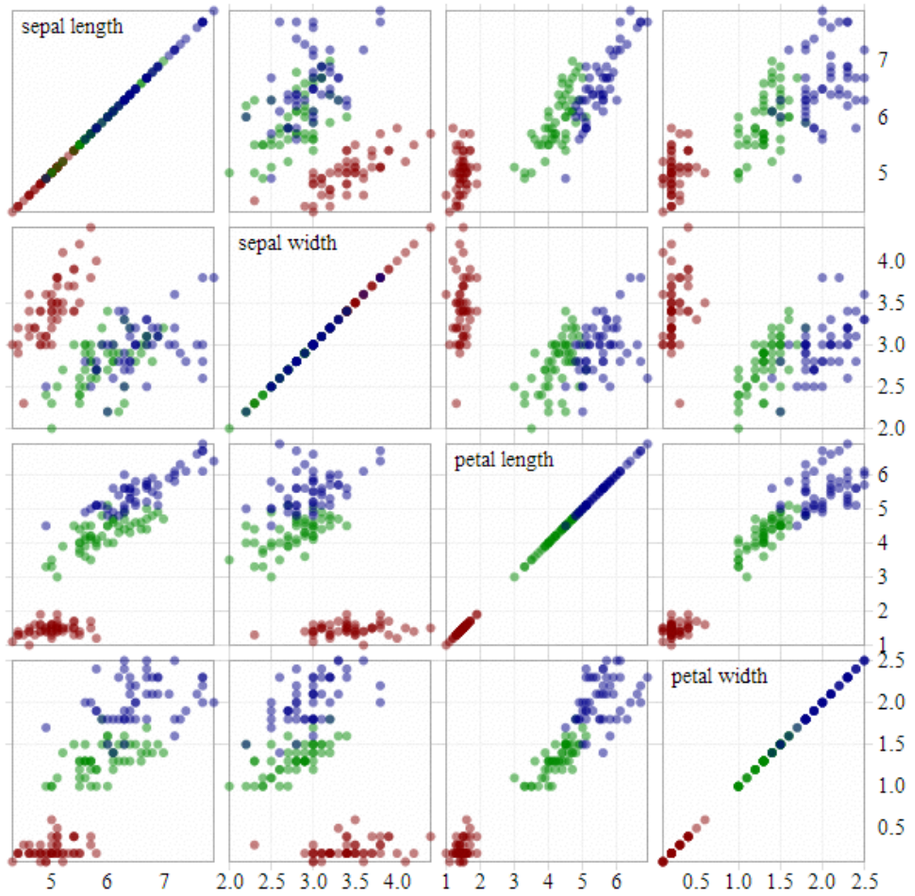
http://bl.ocks.org/kevinschaul/raw/8833989/

Scatterplot Matrices (SPLOM)

Matrix of size d*d

Each row/column is one dimension

Each cell plots a scatterplot of two dimensions



Pixel Based Displays

Each cell is a “pixel”, value encoded in color / value

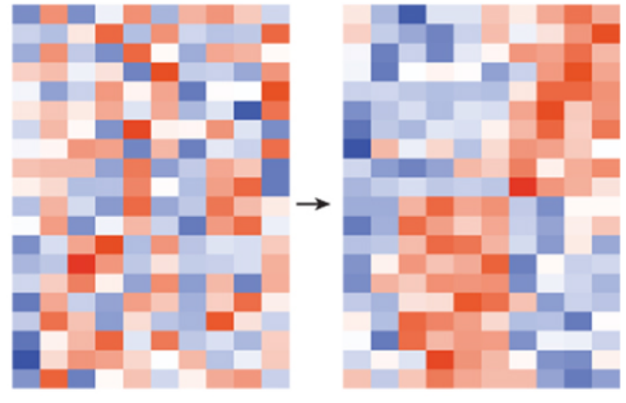
Meaning derived from ordering

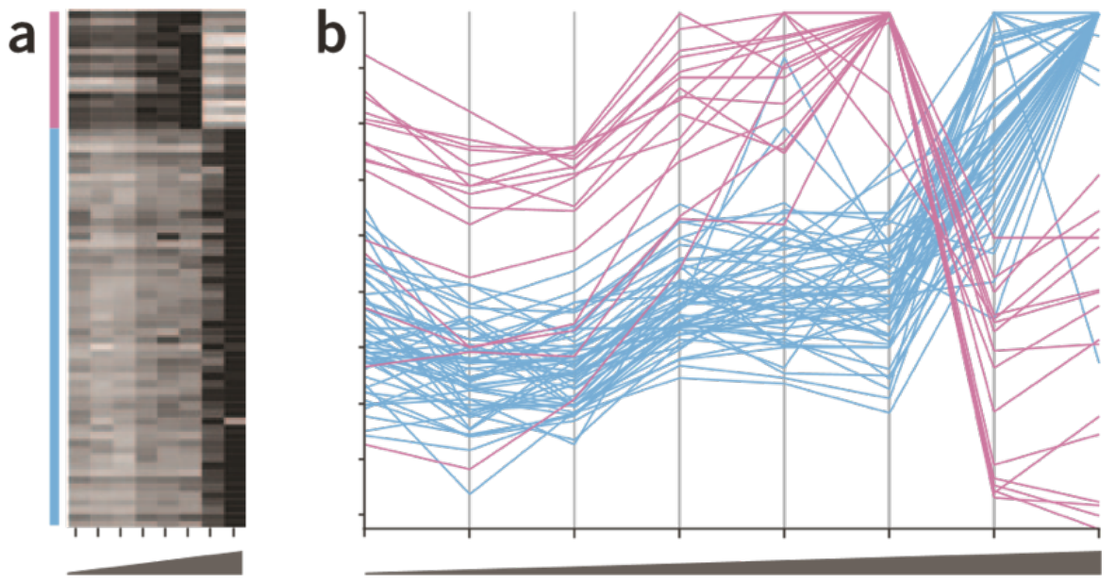
If no ordering inherent, clustering is used

Scalable – 1 px per item

Good for homogeneous data

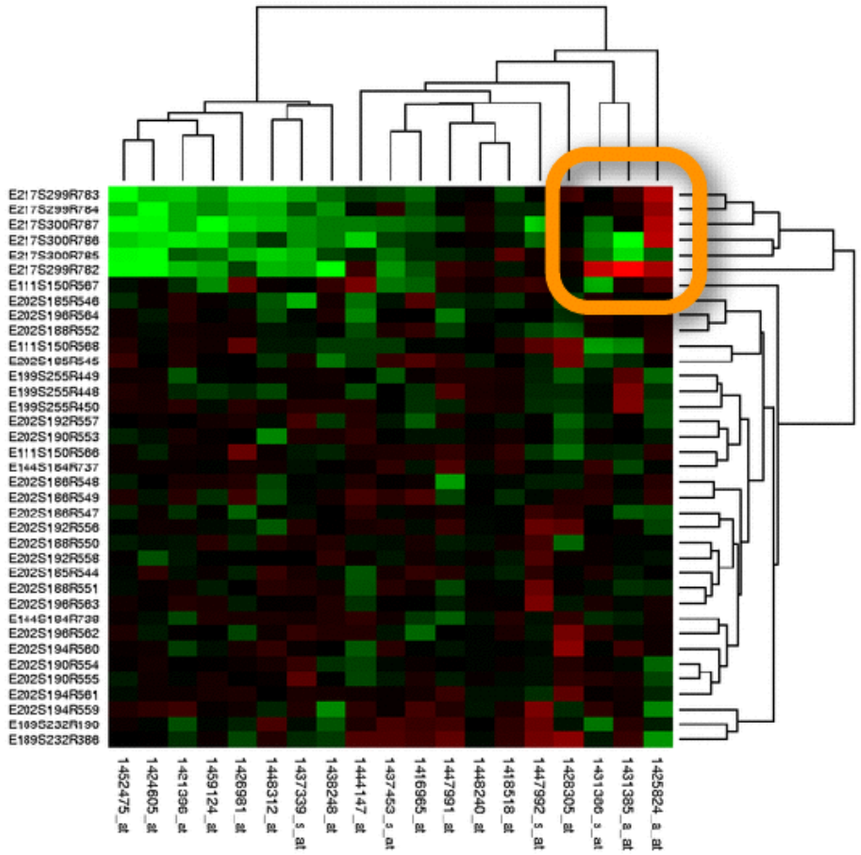
same scale & type



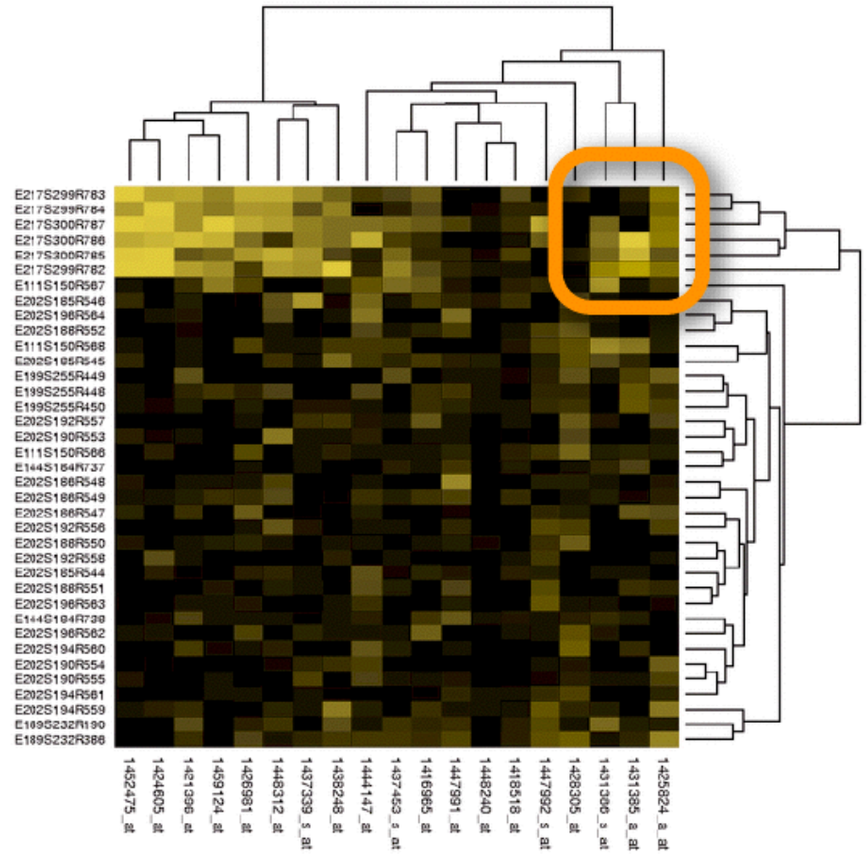


[Gehlenborg & Wong 2012]

Bad Color Mapping

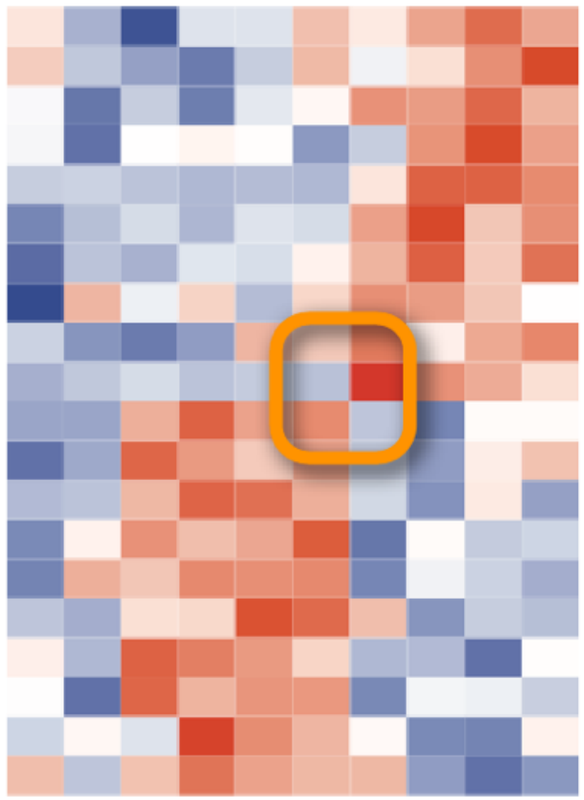


Normal Vision

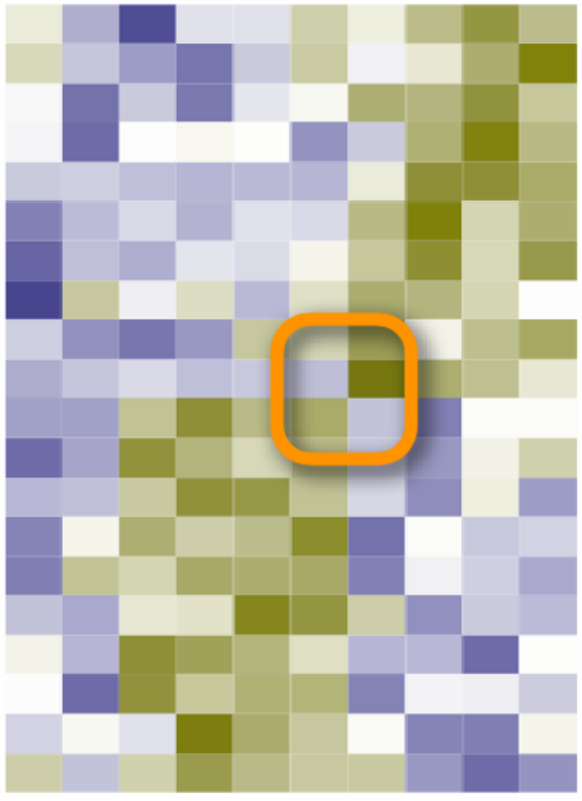


Deuteranope Vision
("Red-Green Blindness")

Good Color Mapping

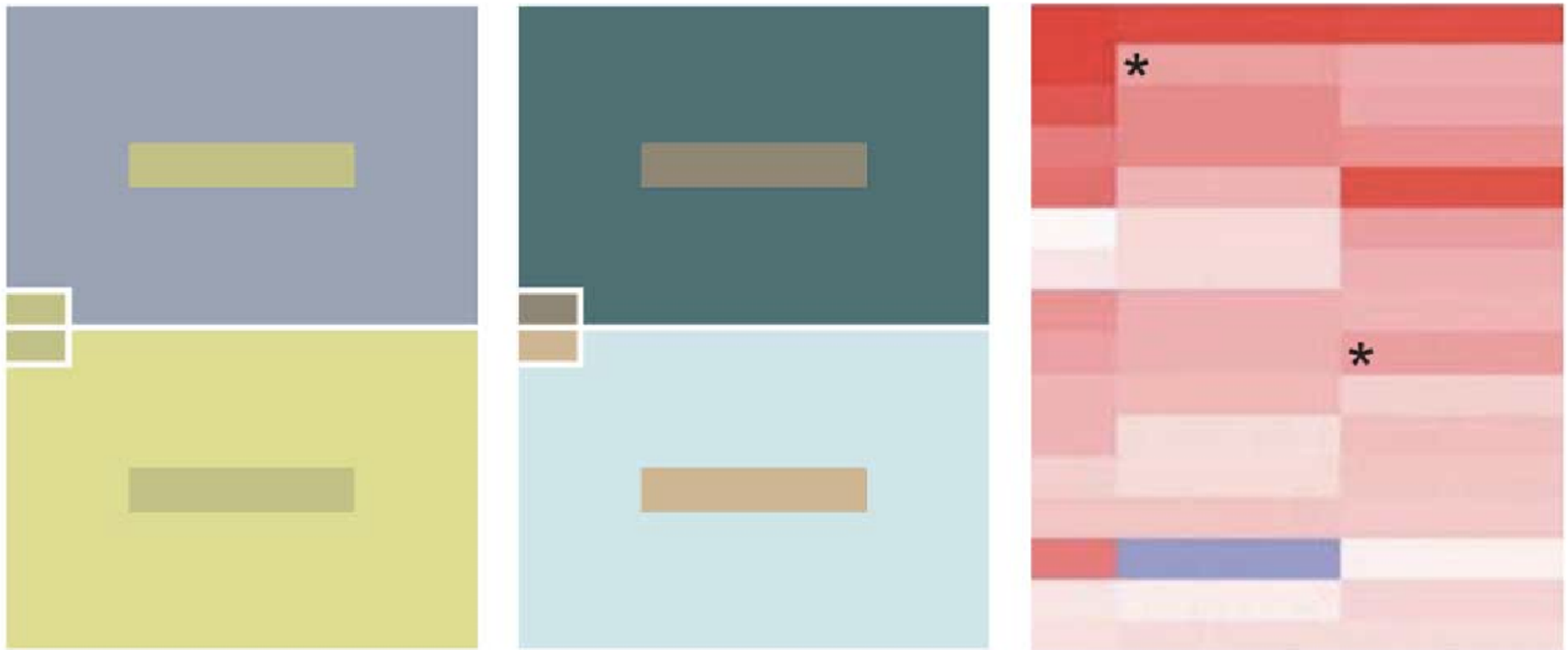


Normal Vision



Deuteranope Vision
("Red-Green Blindness")

Color is relative!



THE GRAPHIC CONTINUUM

The Graphic Continuum shows several ways that data can be illustrated individually or combined to show relationships. Use of various shapes, chart types, and colors can help identify patterns, tell stories, and reveal relationships between different sets and types of data. Bar charts, or histograms, for example, can illustrate a distribution of data over time, but they also can show categorical or geographic differences. Scatterplots can illustrate data from a single instance or for a period, but they also can be used to identify a distribution around a mean.

This set of charts does not constitute an exhaustive list, nor do the connections represent every possible pathway for linking data and ideas. Instead, the Graphic Continuum identifies some presentation methods, and it illustrates some of the connections that can bind different representations together. The six groups do not define all possibilities: Many other useful, overlapping data types and visualization techniques are possible.

This chart can guide graphic choices, but your imagination can lead the way to other effective ways to present data.

DISTRIBUTION

Graphical representations of the distribution of data

TIME

Track changes over time

COMPARING CATEGORIES

Compare values across categories

GEOSPATIAL

Relate data to its geography

PART-TO-WHOLE

Visualizations that relate the part of a variable to its total

RELATIONSHIP

Illustrate correlations or relationships between variables