

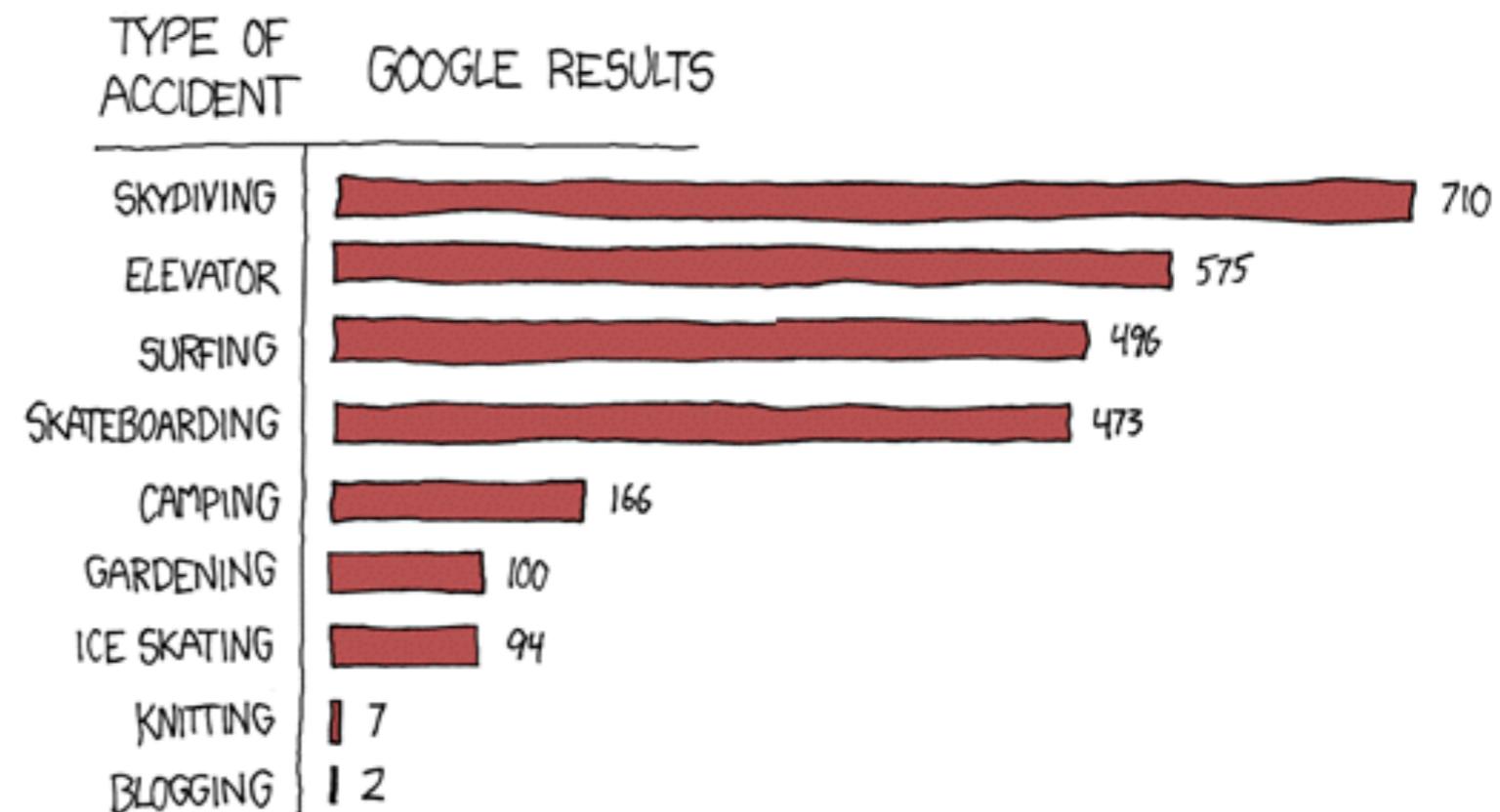
Choisir le type de
graph

Comparaisons

Bar Chart

DANGERS

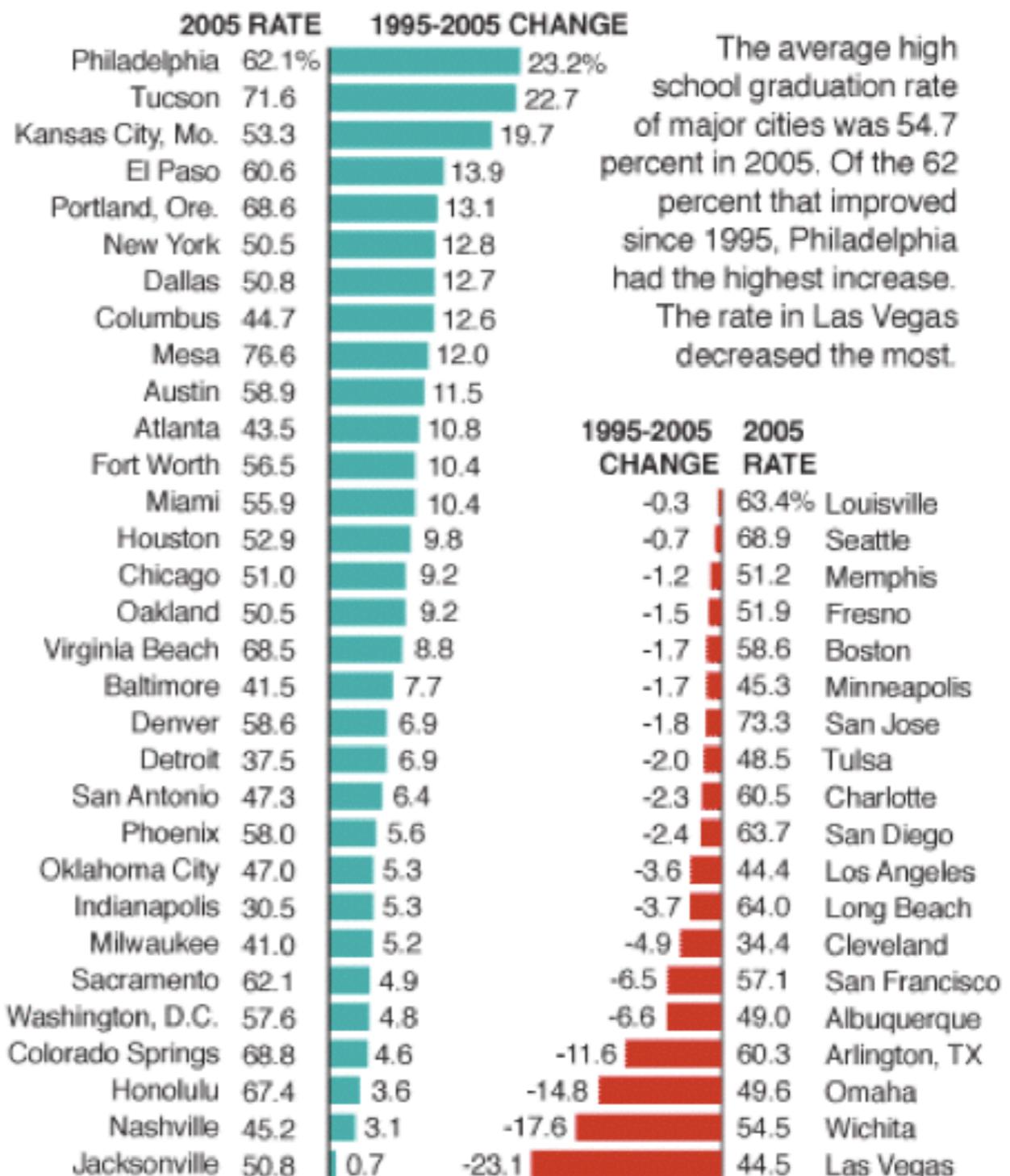
INDEXED BY THE NUMBER OF GOOGLE RESULTS FOR
"DIED IN A _____ ACCIDENT"



Direction

Graduation rates up in most cities

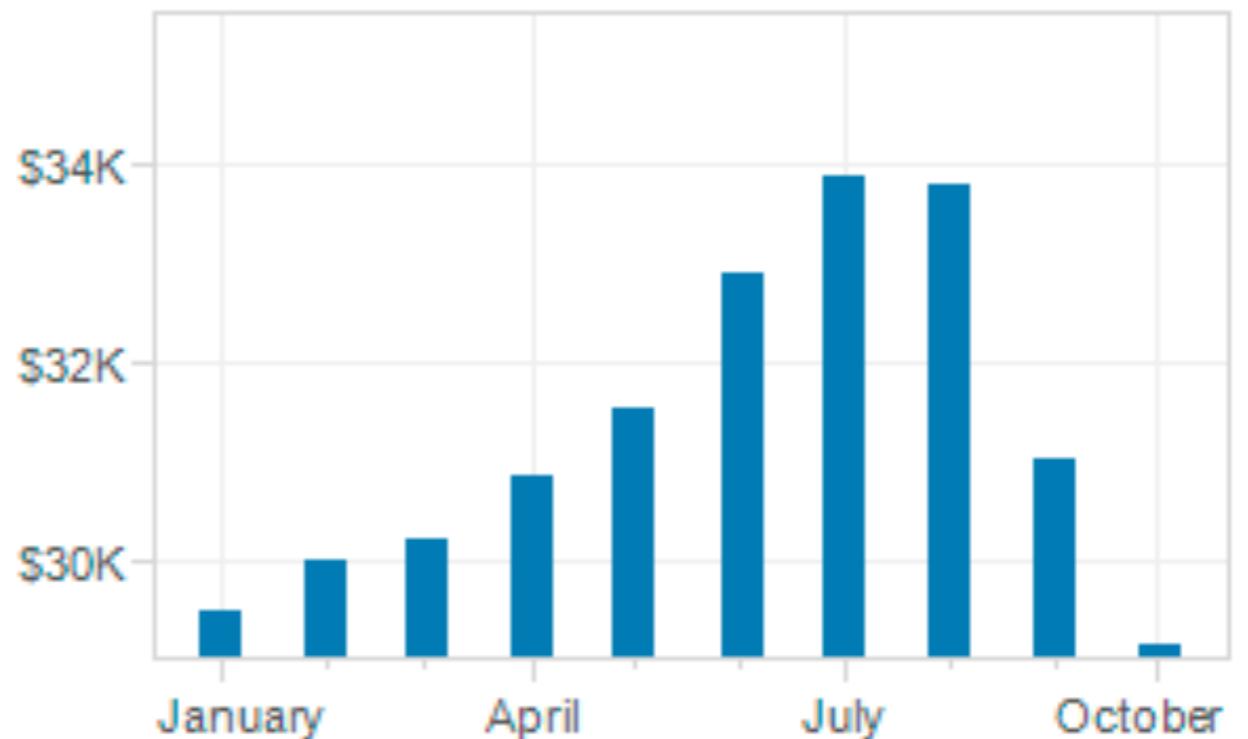
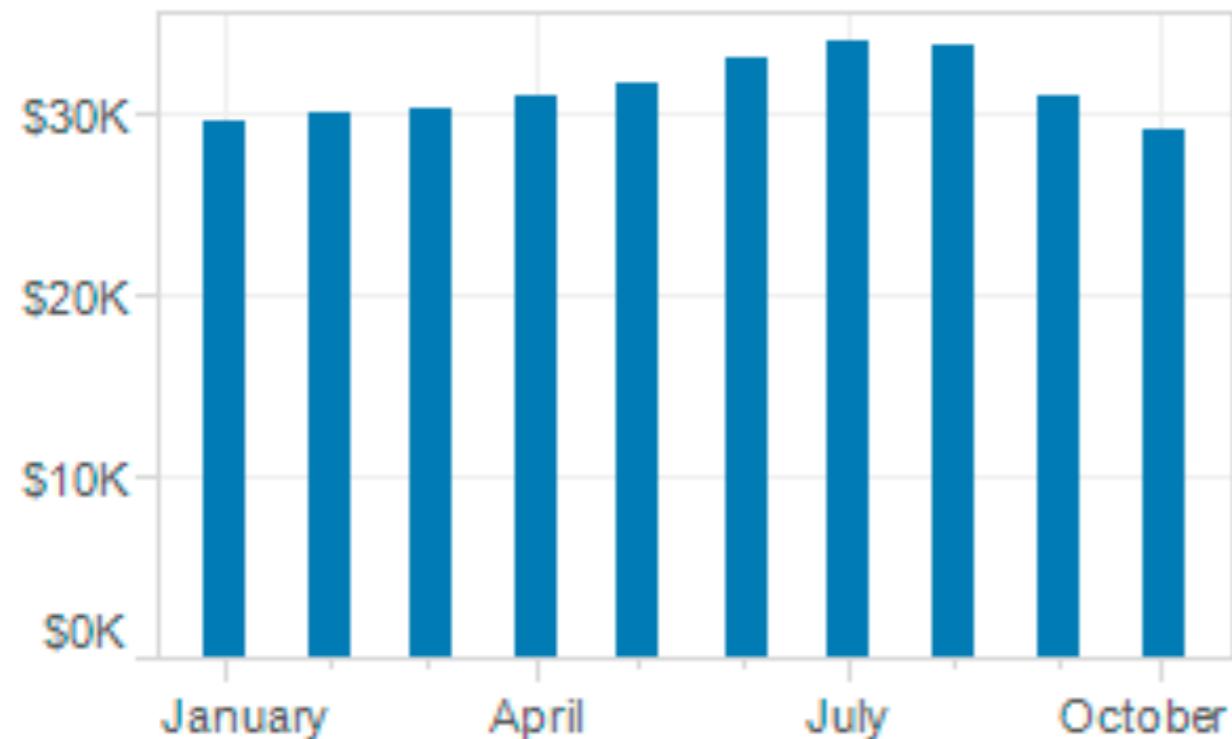
Graduation rate for principal school district of the largest cities



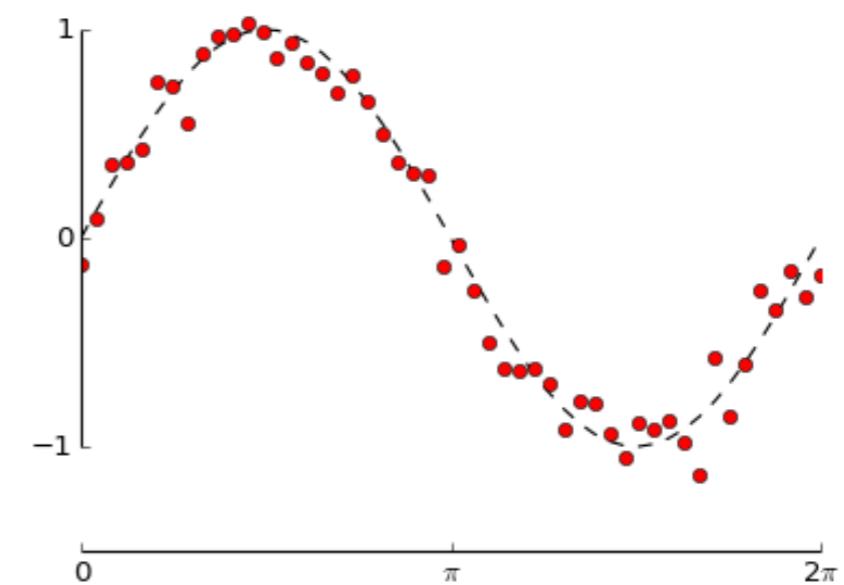
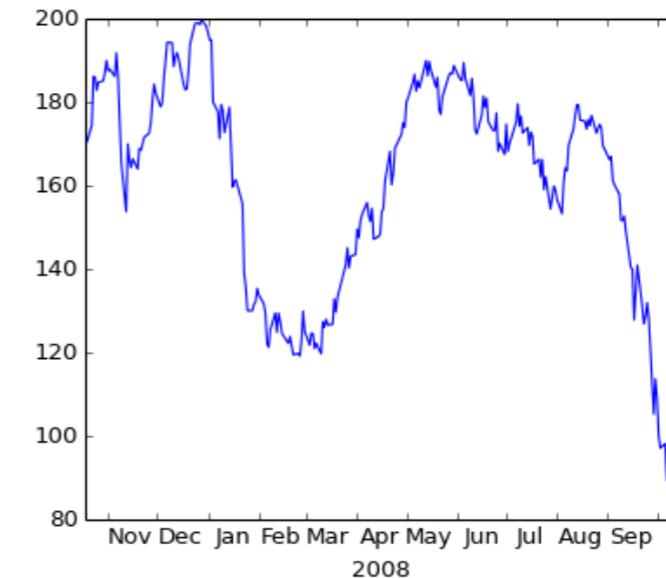
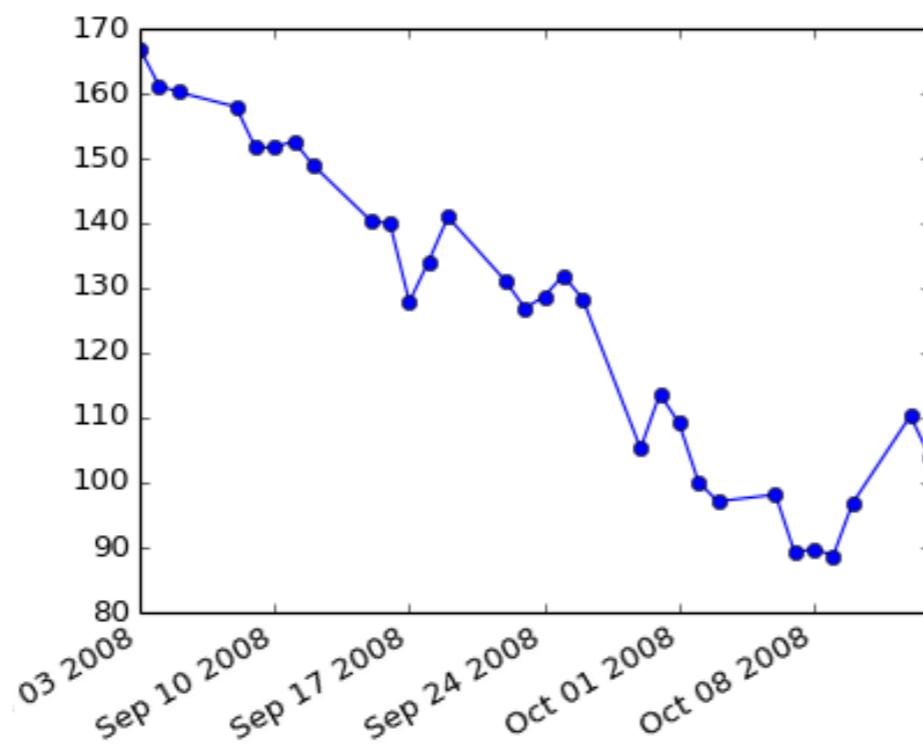
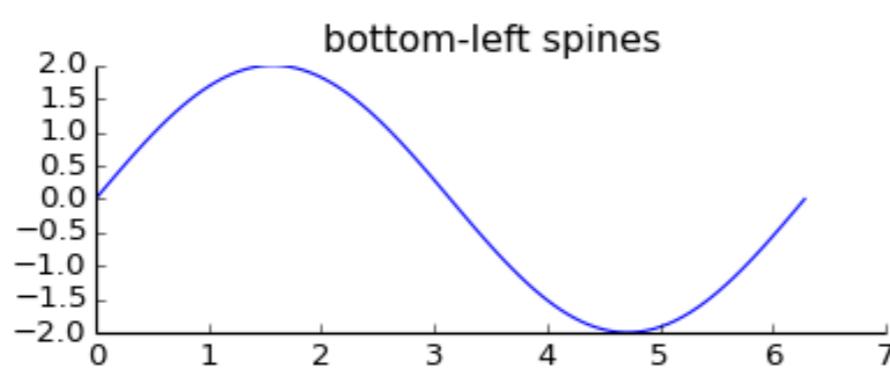
SOURCE: EPE Research Center

AP

attention à la baseline



Line Chart

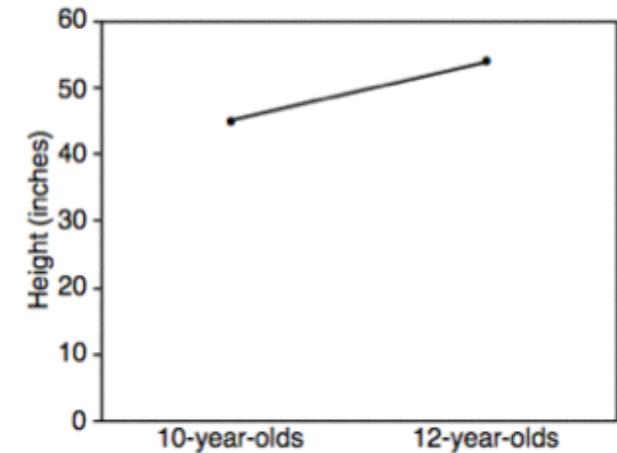
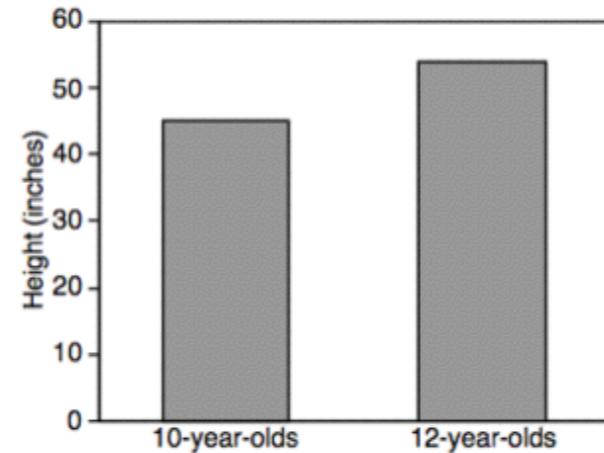
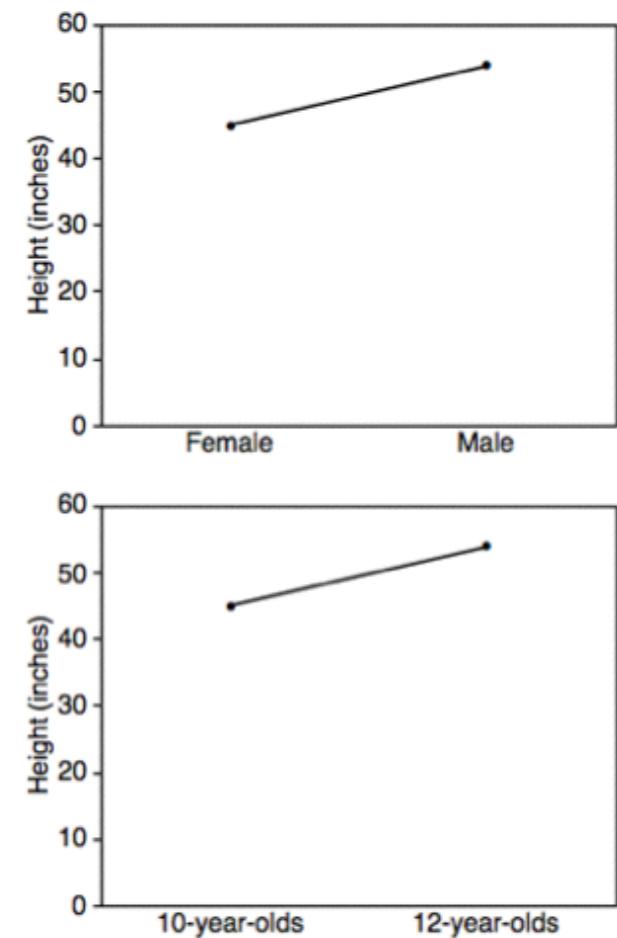
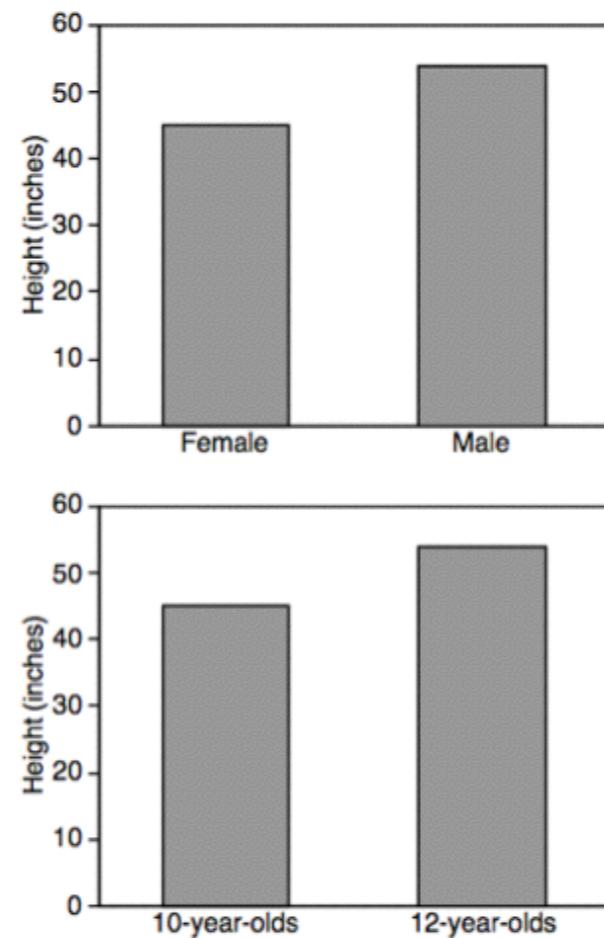


matplotlib gallery

Bars vs. Lines

Lines imply connections & sampling from continuous data.

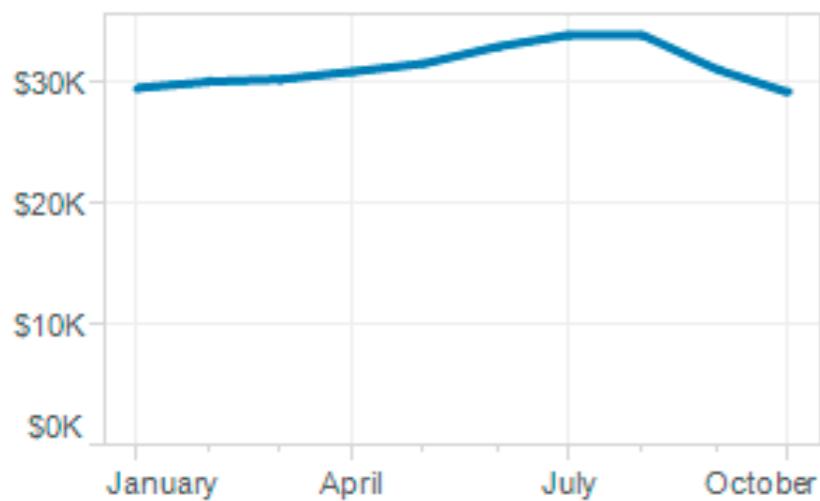
Do not use for categorical data.



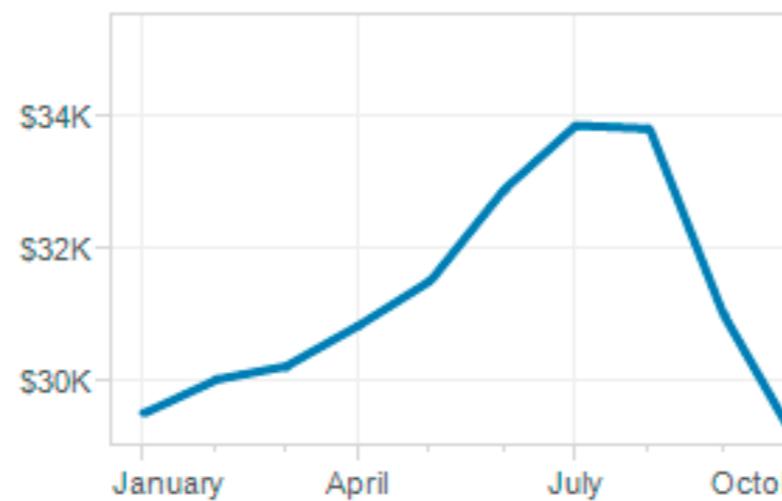
Zacks 1999

Baseline Problem (again)

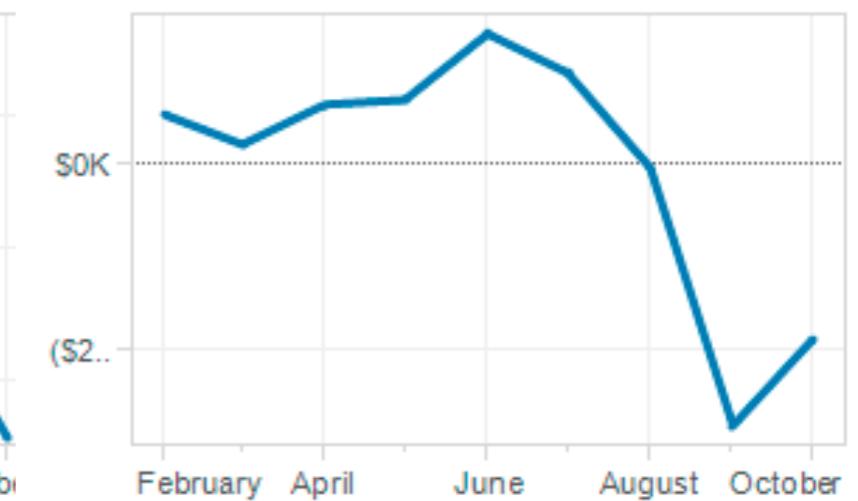
True Baseline



Clipped Baseline



Plotting Change

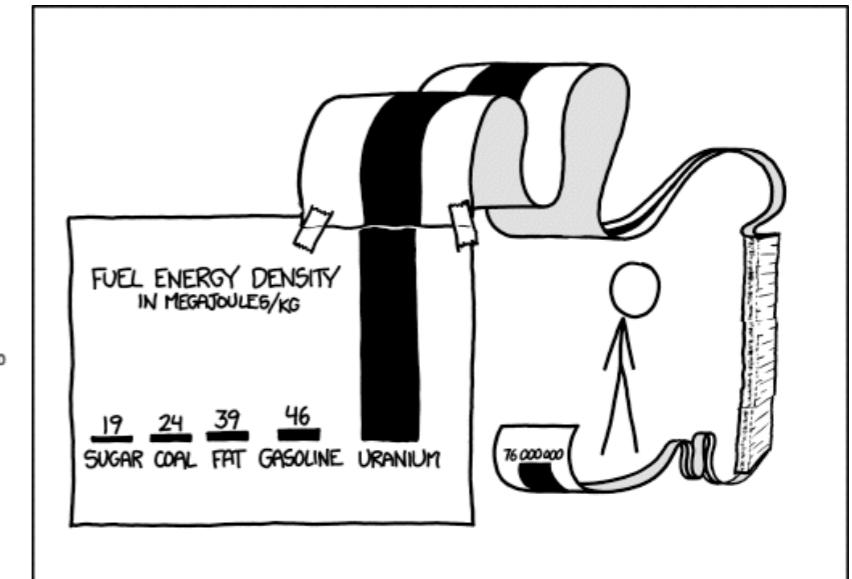


Linear vs. Logarithmic Scale



Apple Stock Price

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

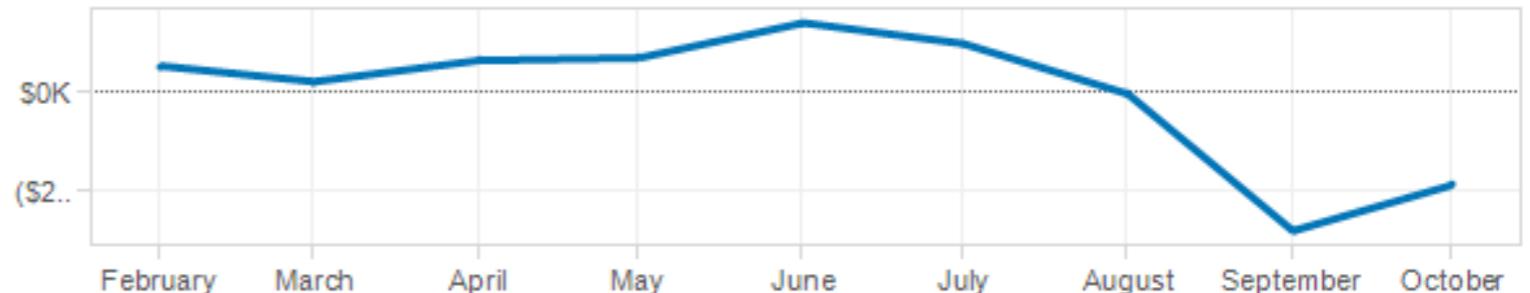
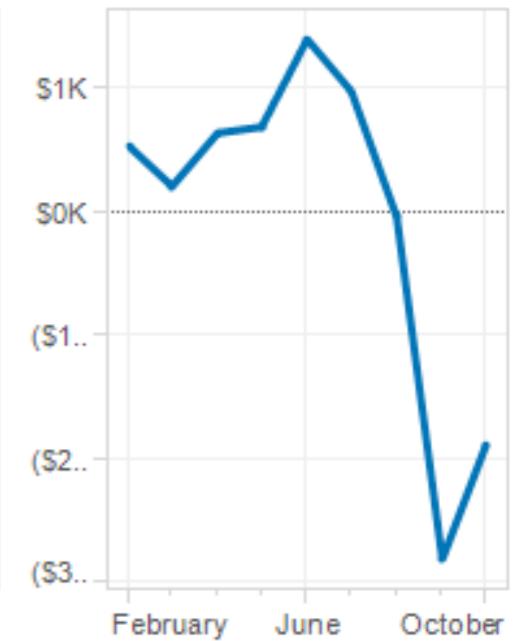
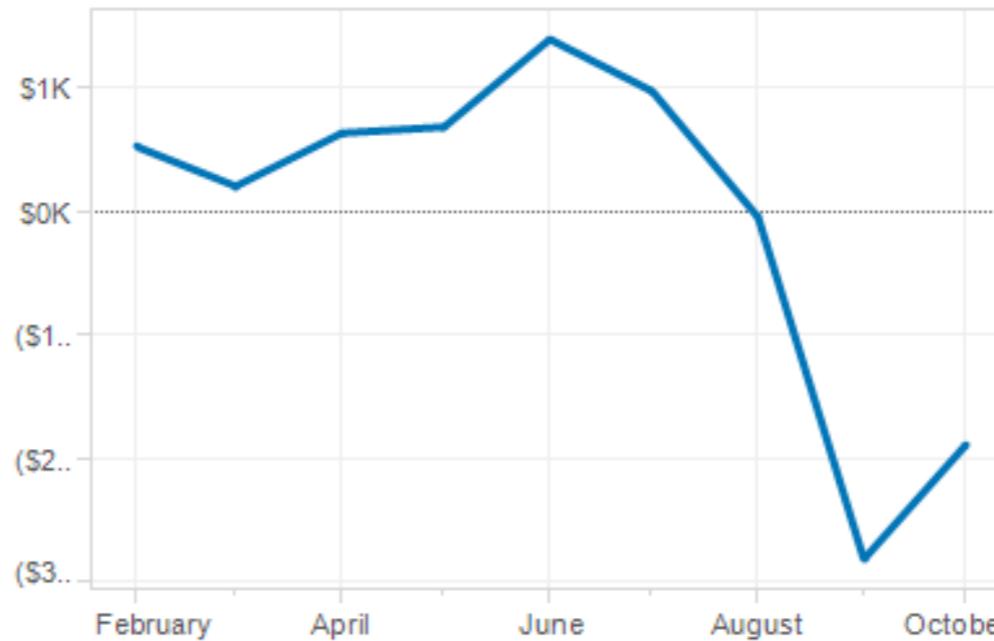


SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT PROPERLY.

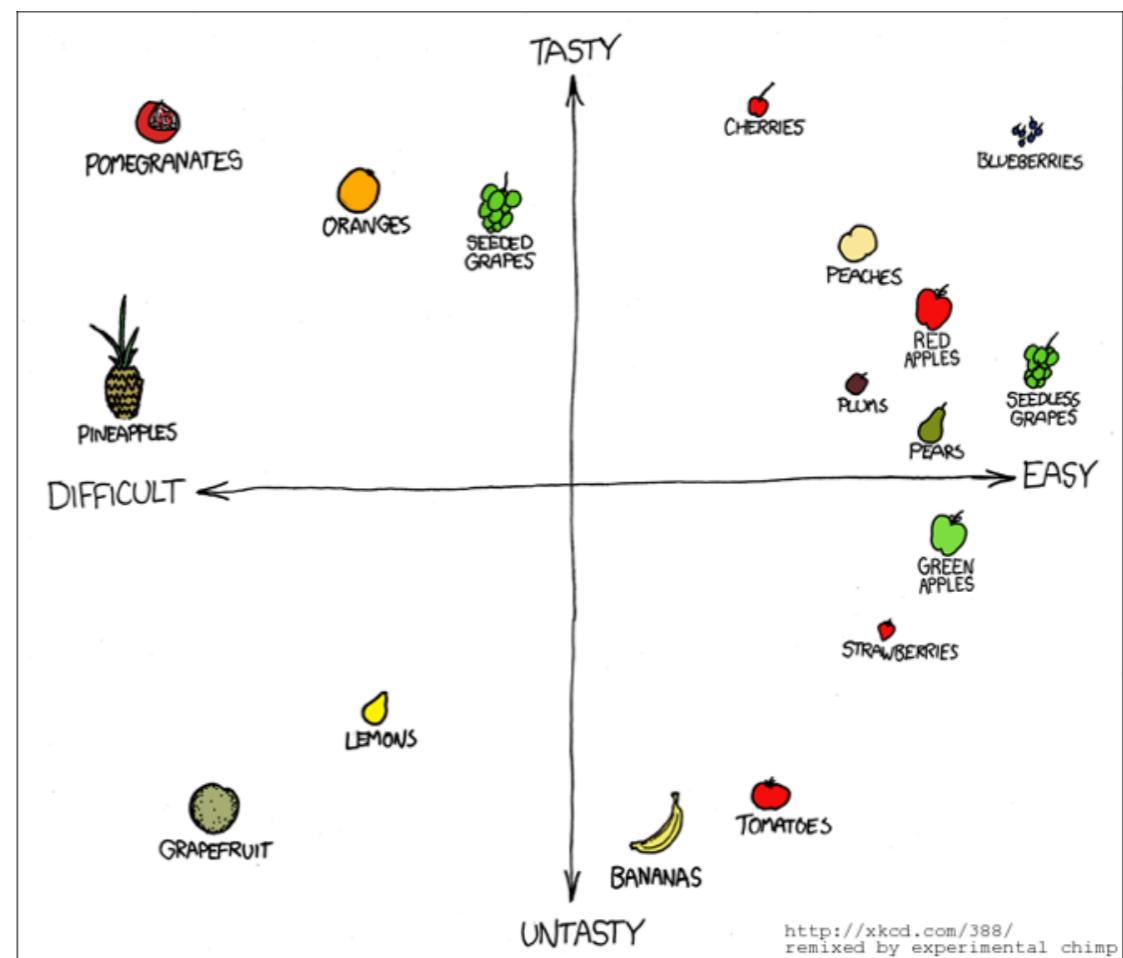
<http://xkcd.com/1162/>

Aspect Ratios

Rule of Thumb:
Banking to 45°
(average line
slope: 45°)

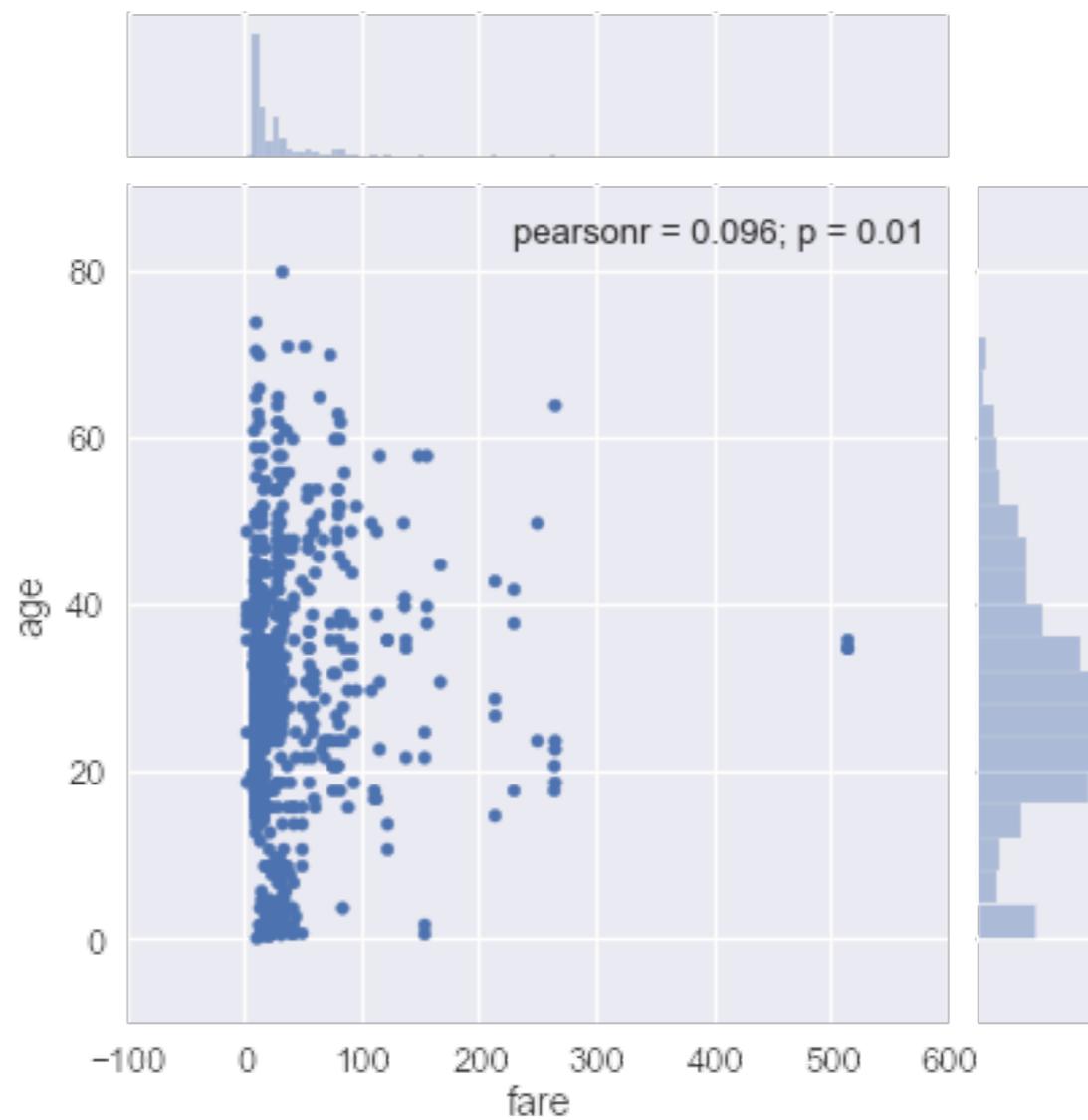


Correlations



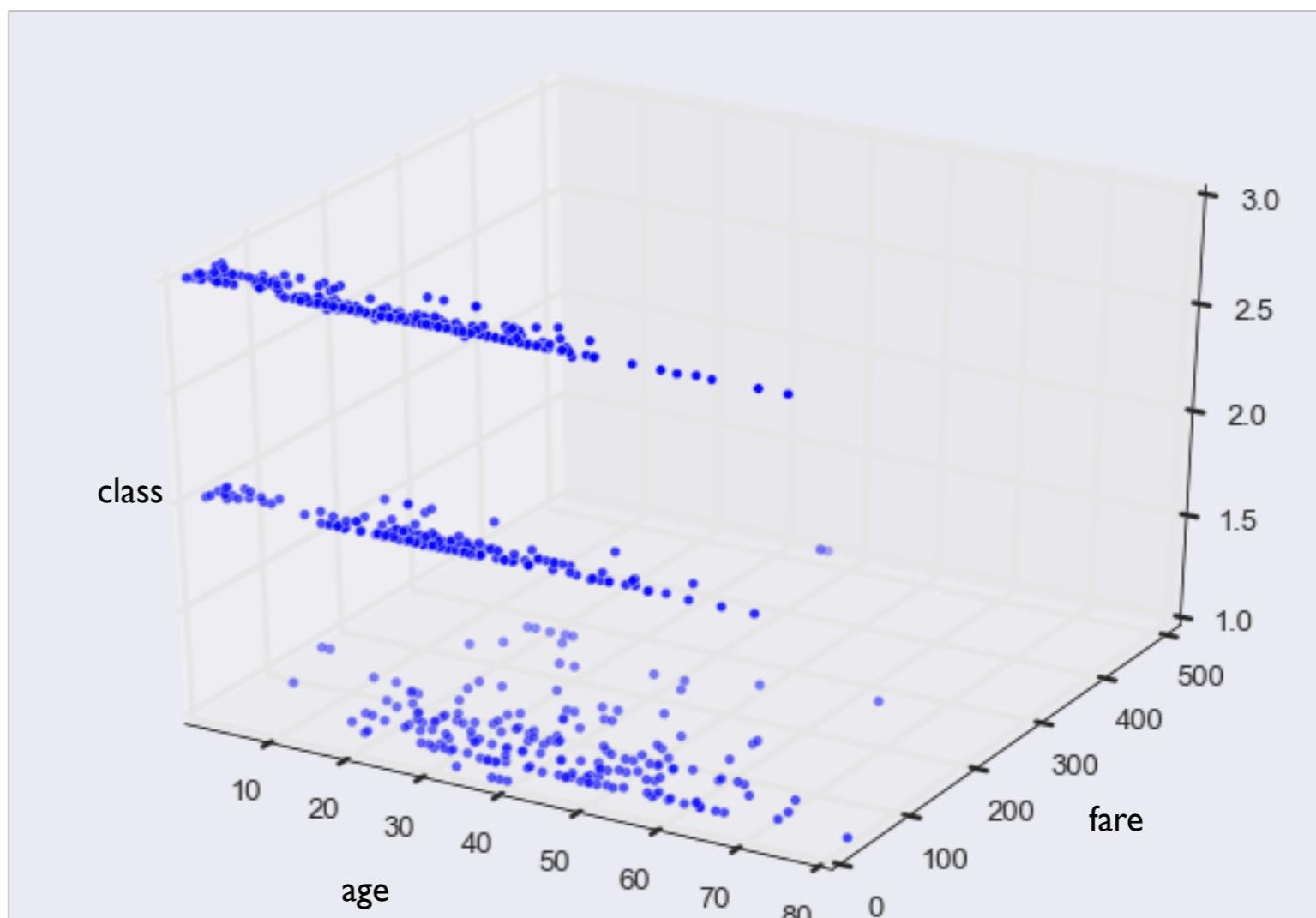
age	fare
22.0	7.25
38.0	71.2833
26.0	7.925
35.0	53.1
35.0	8.05
	8.4583
54.0	51.8625
2.0	21.075
27.0	11.1333
14.0	30.0708
4.0	16.7
58.0	26.55
20.0	8.05
39.0	31.275
14.0	7.8542
55.0	16.0
2.0	29.125
	13.0
31.0	18.0
	7.225
35.0	26.0
34.0	13.0
15.0	8.0292

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

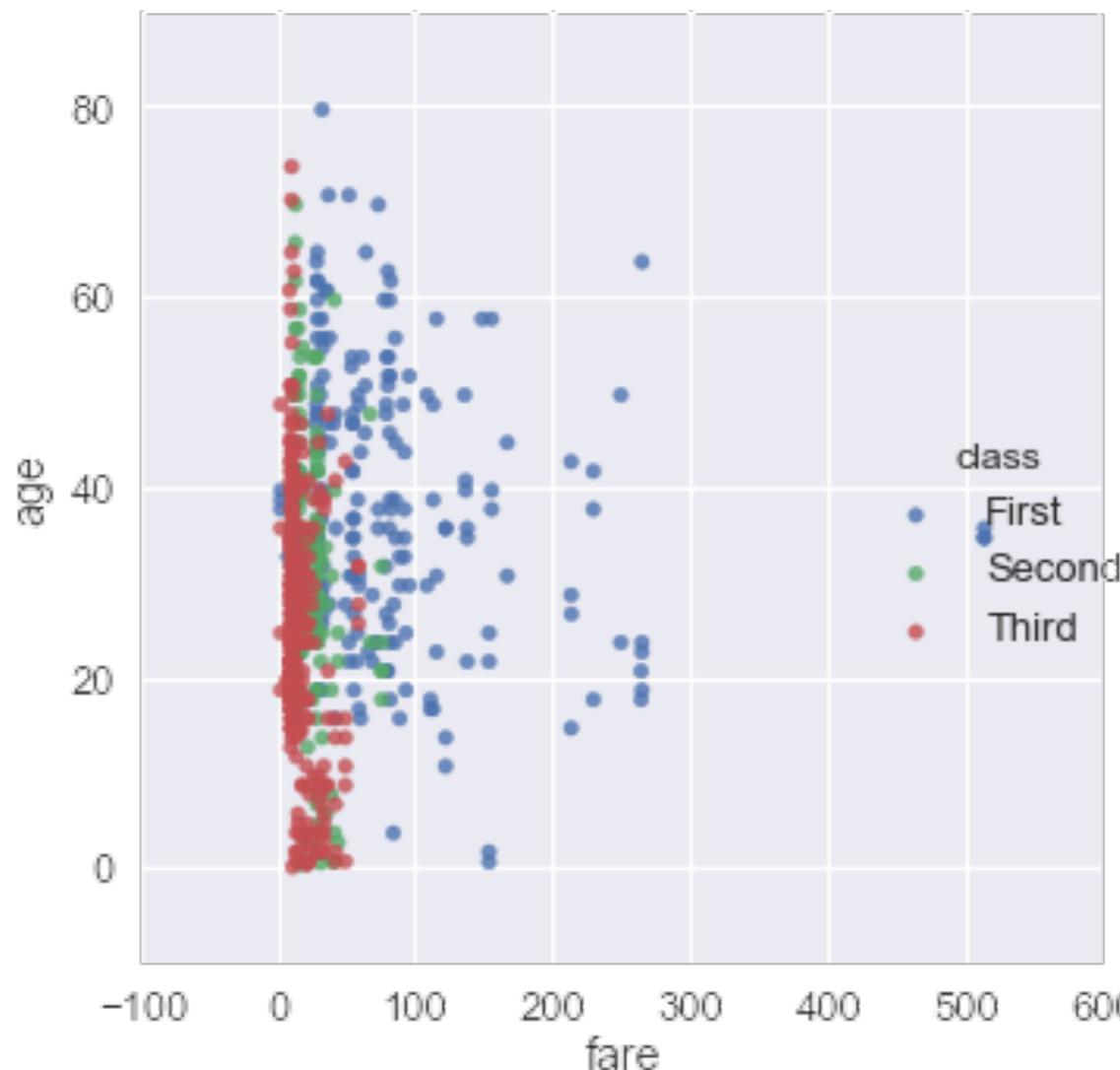
Trivariate Data



Do NOT use 3D scatterplots!

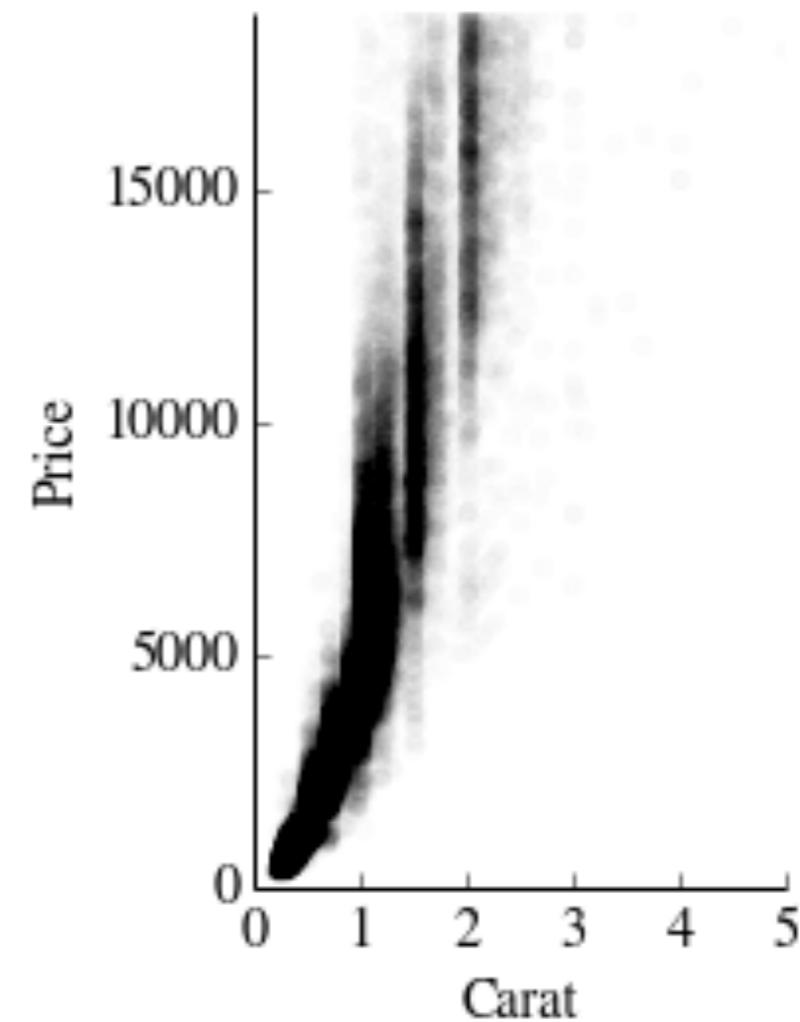
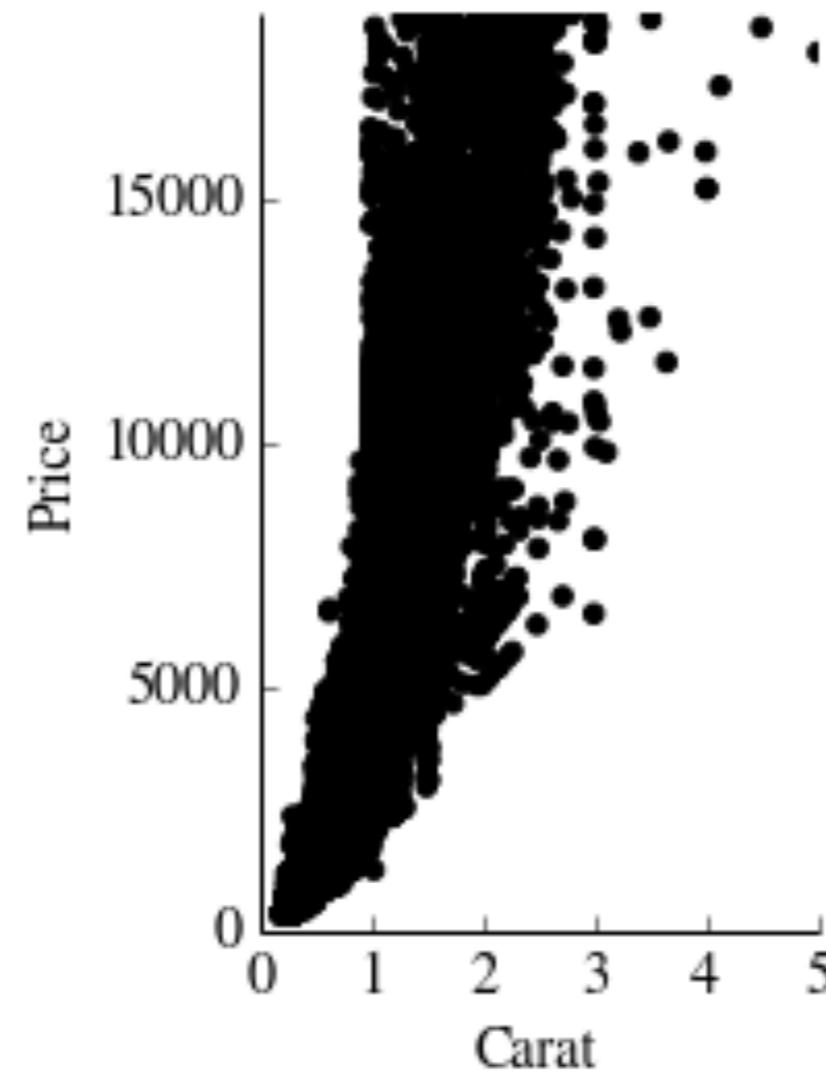
age	fare	class
22.0	7.25	Third
38.0	71.2833	First
26.0	7.925	Third
35.0	53.1	First
35.0	8.05	Third
	8.4583	Third
54.0	51.8625	First
2.0	21.075	Third
27.0	11.1333	Third
14.0	30.0708	Second
4.0	16.7	Third
58.0	26.55	First
20.0	8.05	Third
39.0	31.275	Third
14.0	7.8542	Third
55.0	16.0	Second
2.0	29.125	Third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

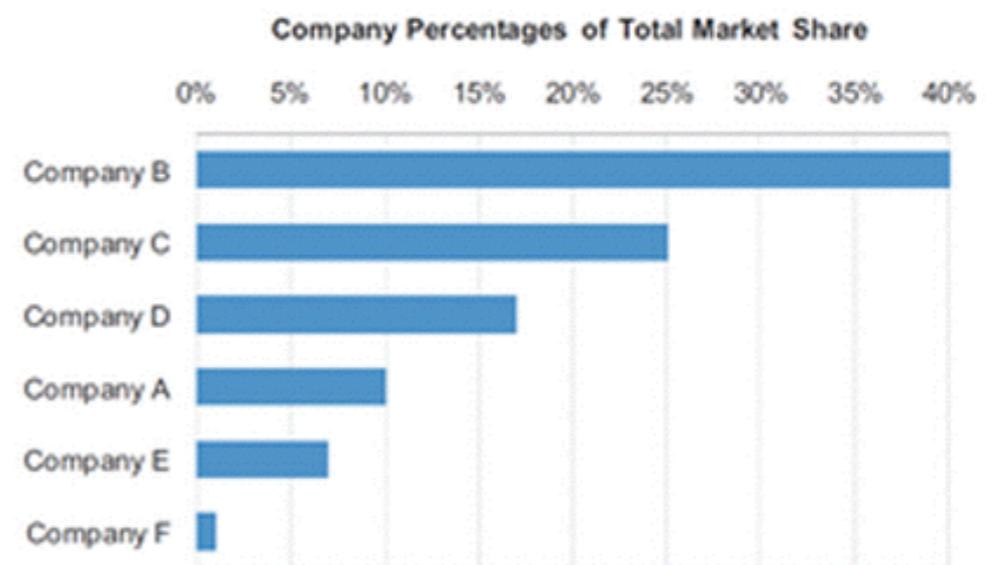
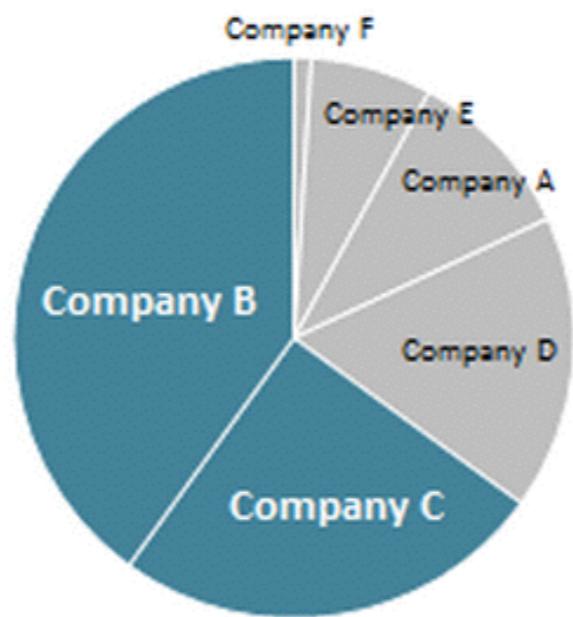


$\alpha = 1/100$

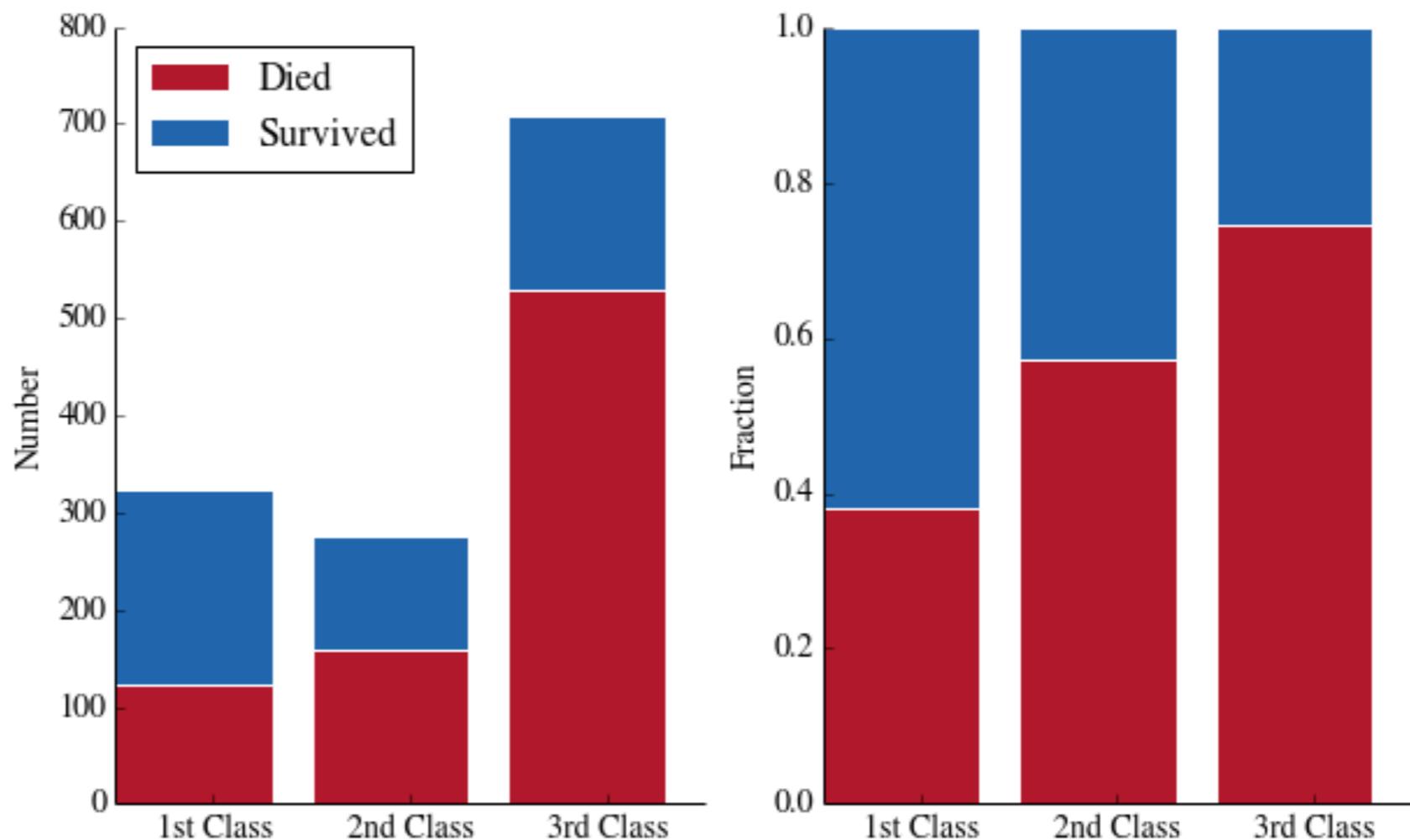
Compositions

Pie vs. Bar Charts

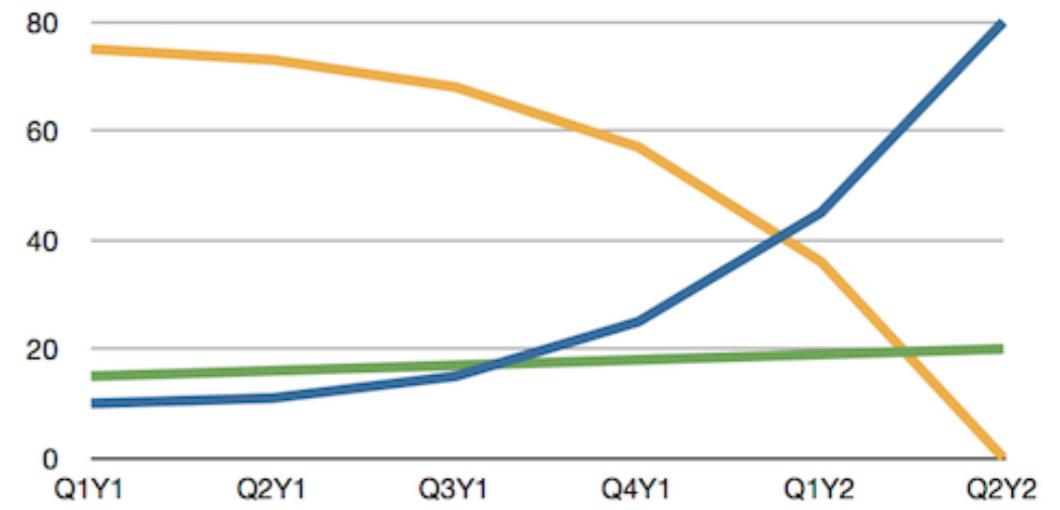
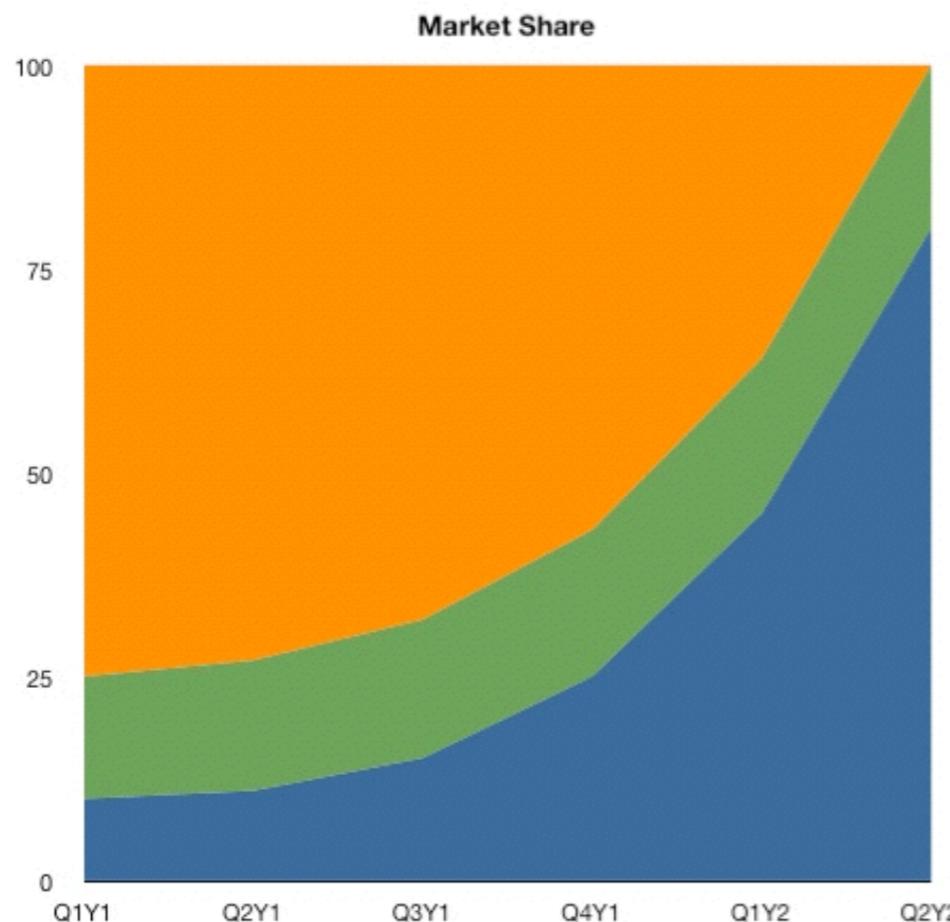
65% of the market is controlled by companies B and C



Stacked Bar Chart



Stacked Area vs. Line Graphs



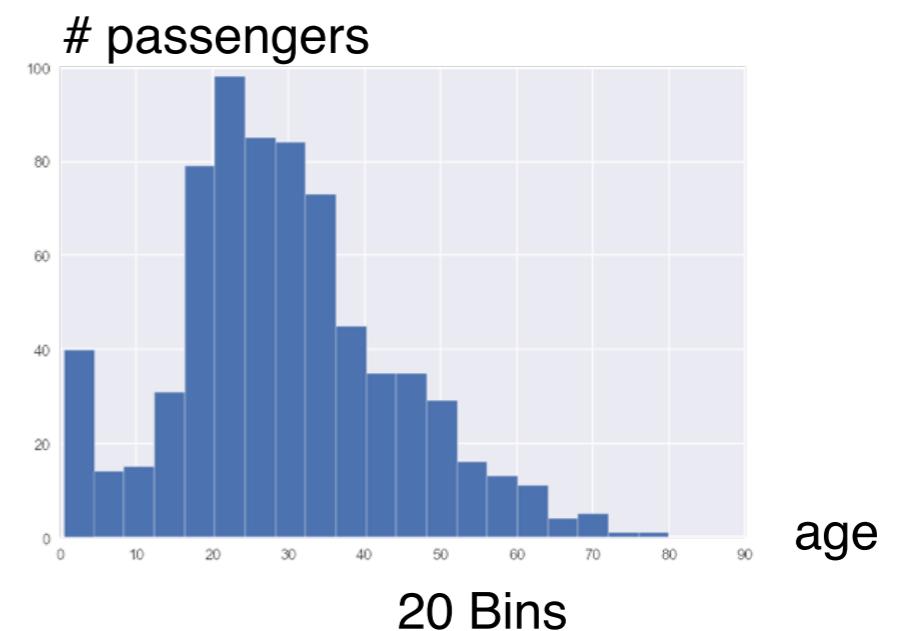
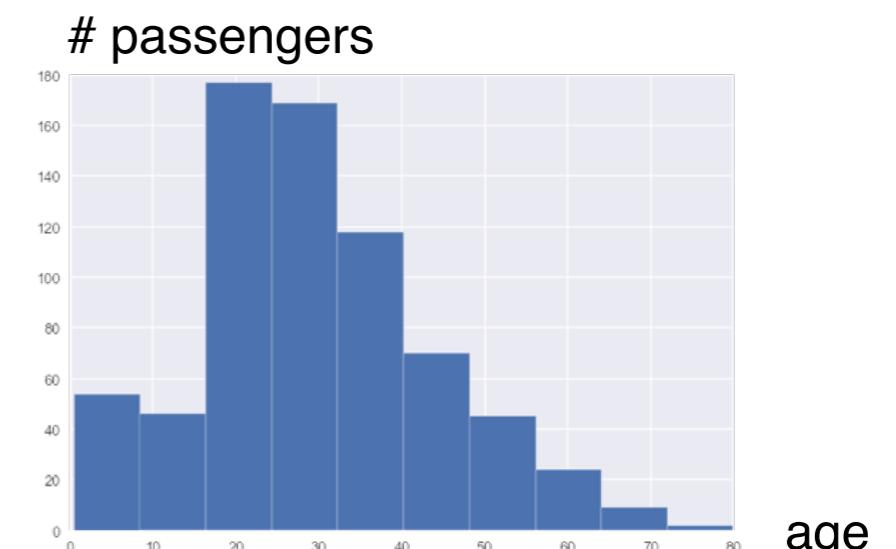
Distributions

Histogram

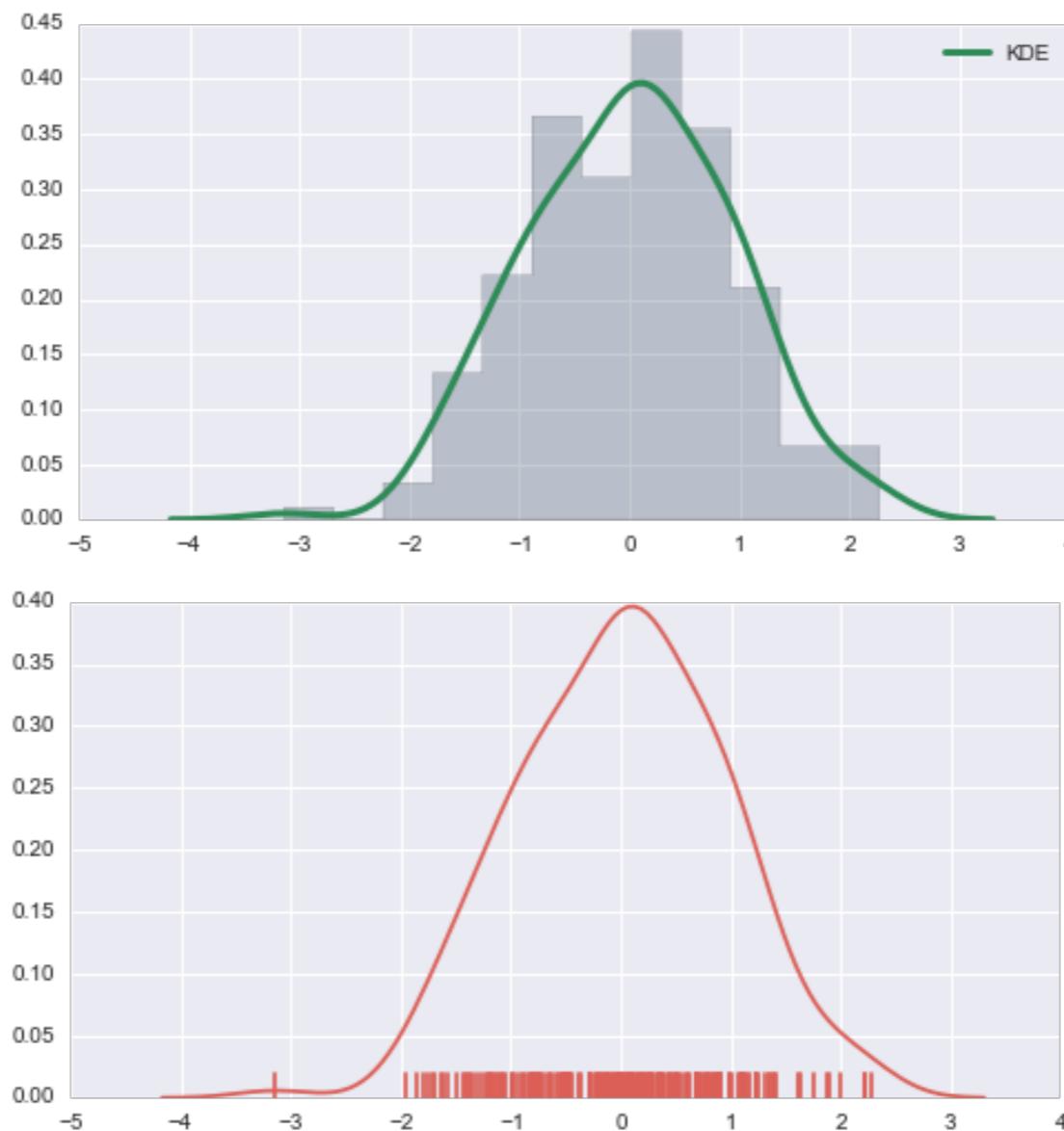
#bins hard to predict

make interactive!

rule of thumb: #bins = \sqrt{n}

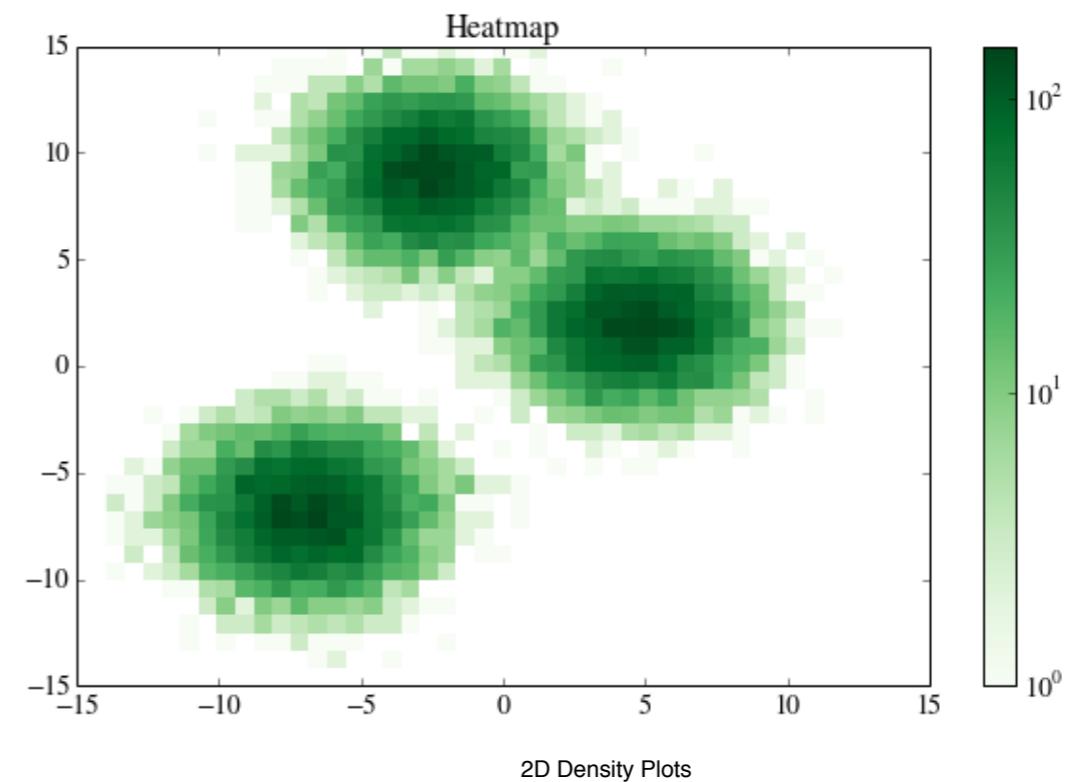
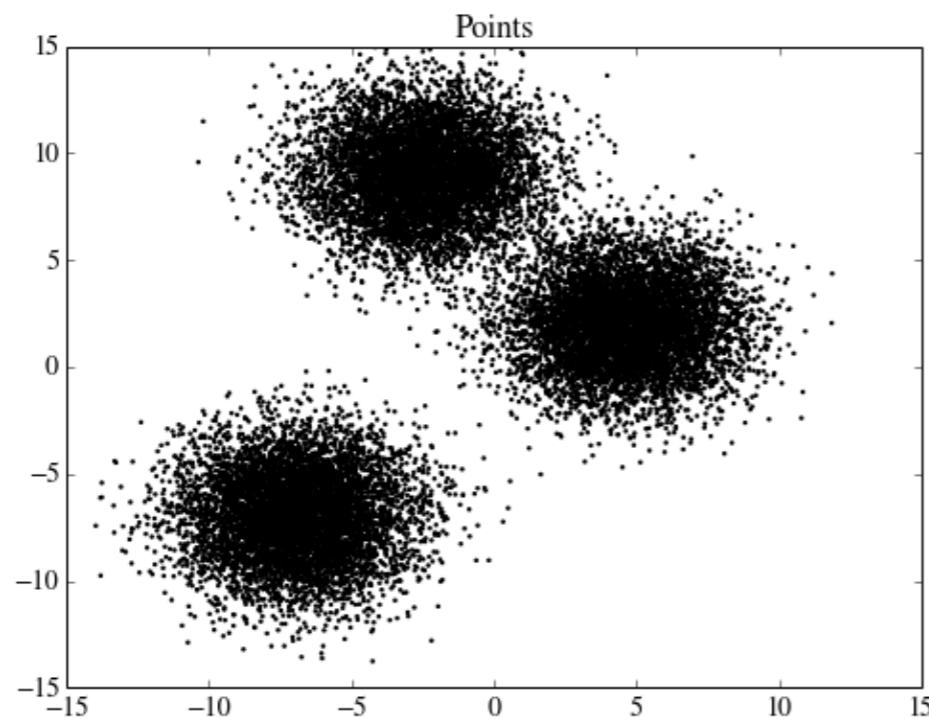


Density Plots



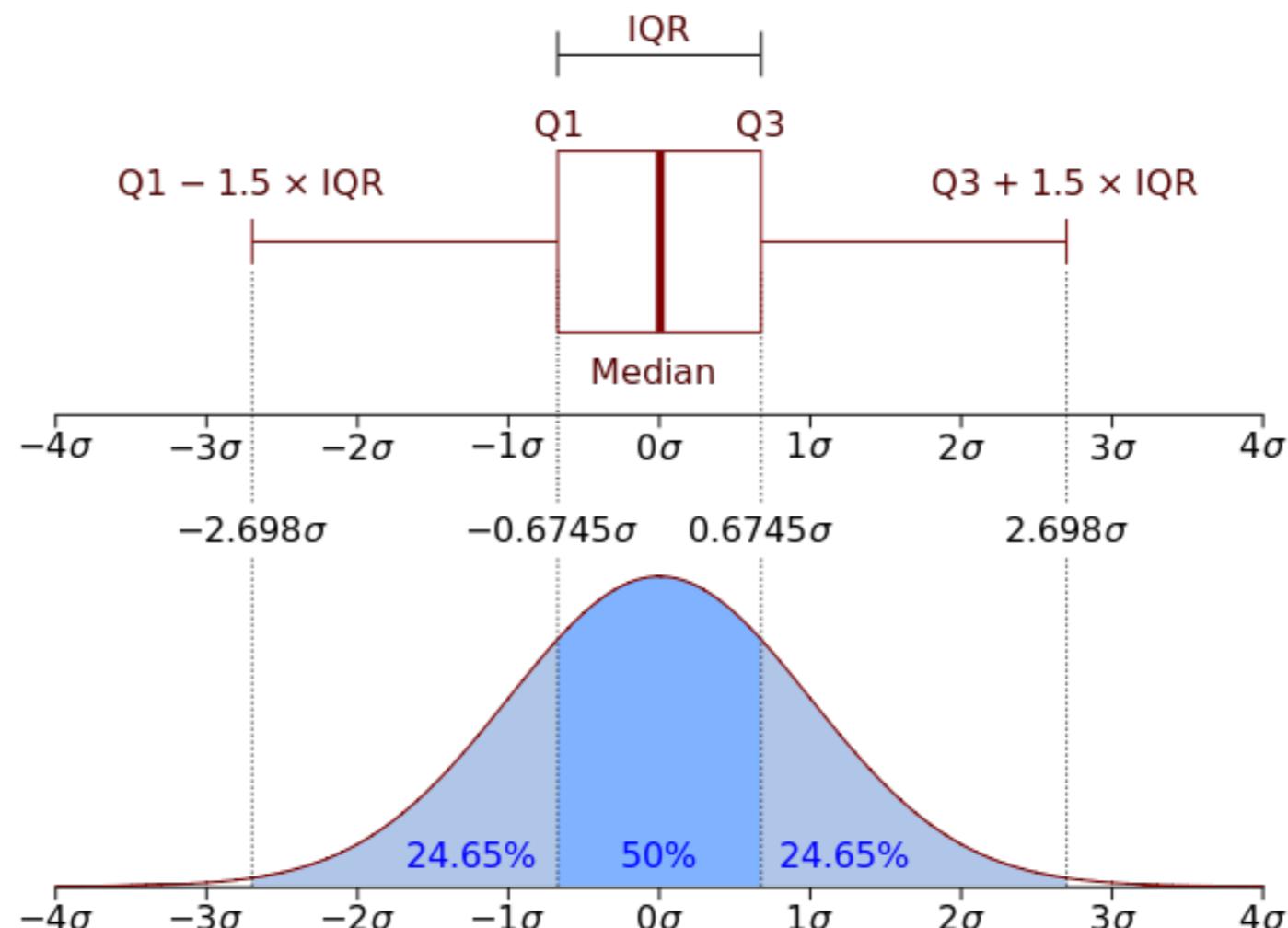
Heat Maps

binning of scatterplots



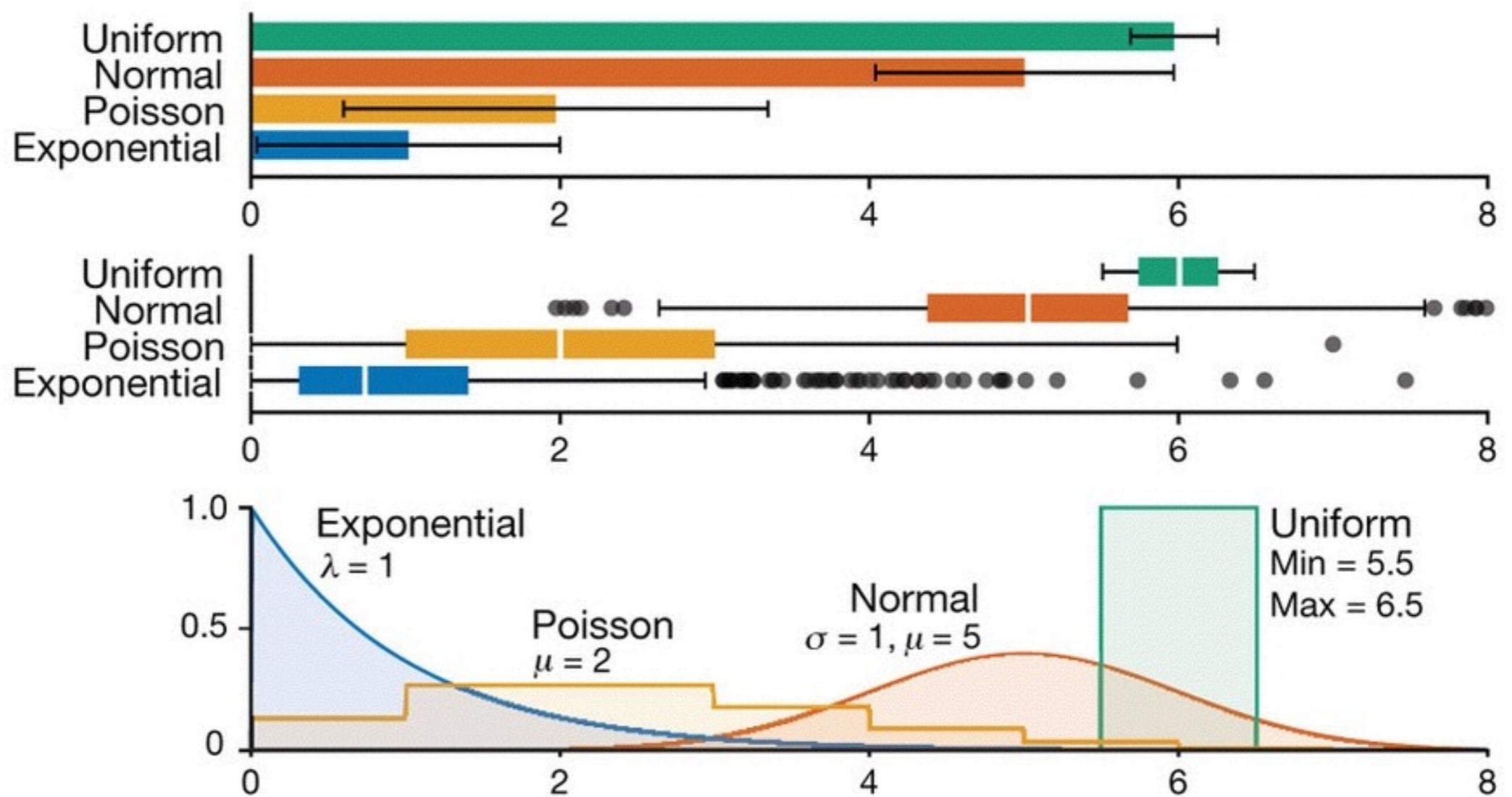
Box Plots

aka Box-and-Whisker Plot



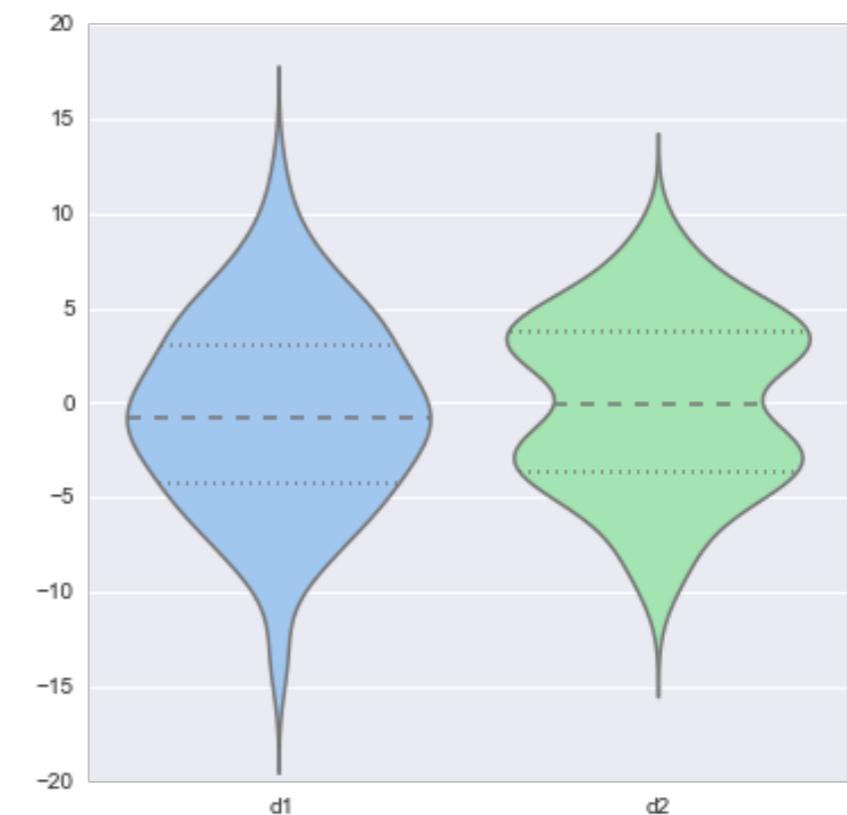
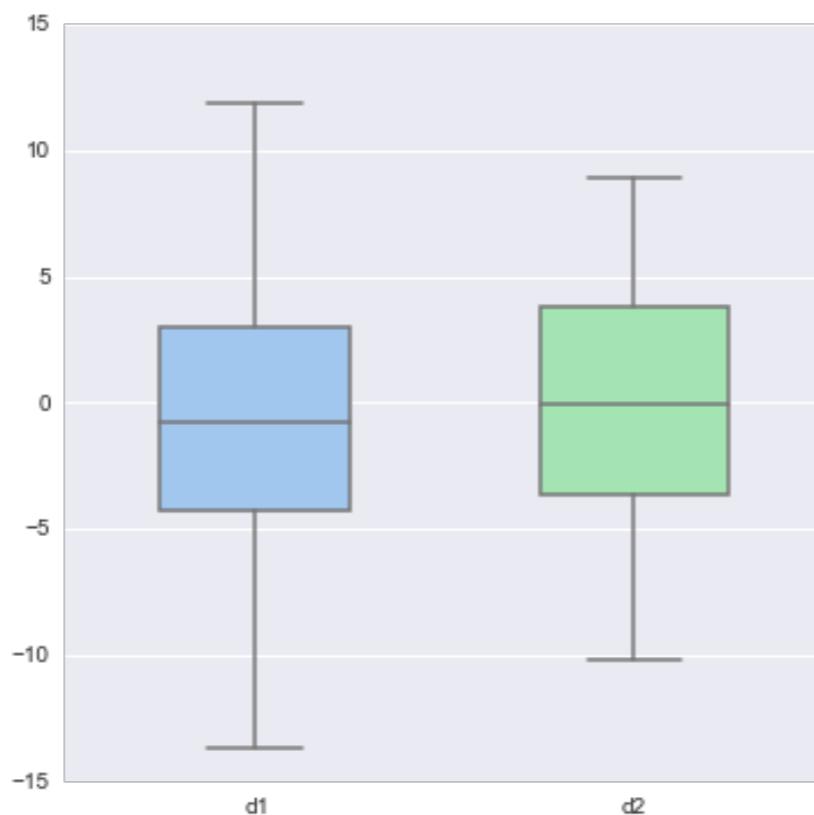
Wikipedia

Comparison



Violin Plot

= Box Plot + Probability Density Function



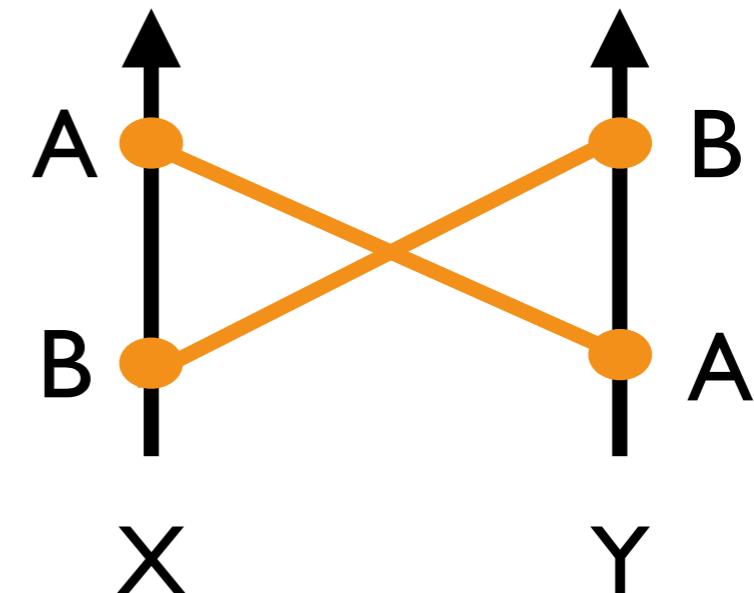
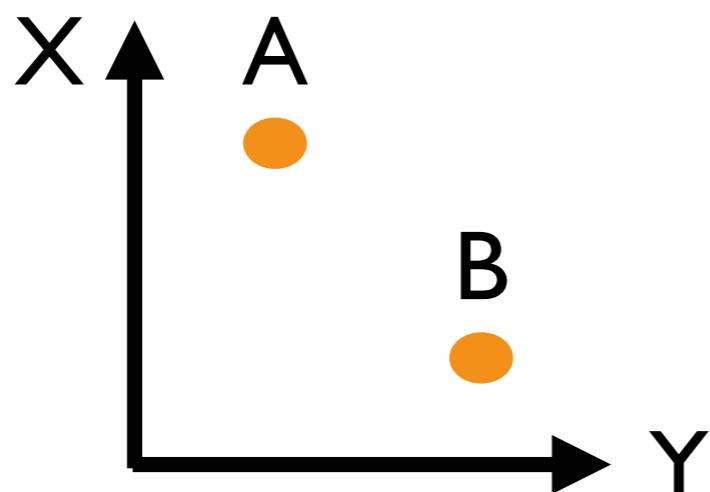
Highdimensional Data

Parallel Coordinates (PC)

Inselberg 1985

Axes represent attributes

Lines connecting axes represent items



Parallel Coordinates

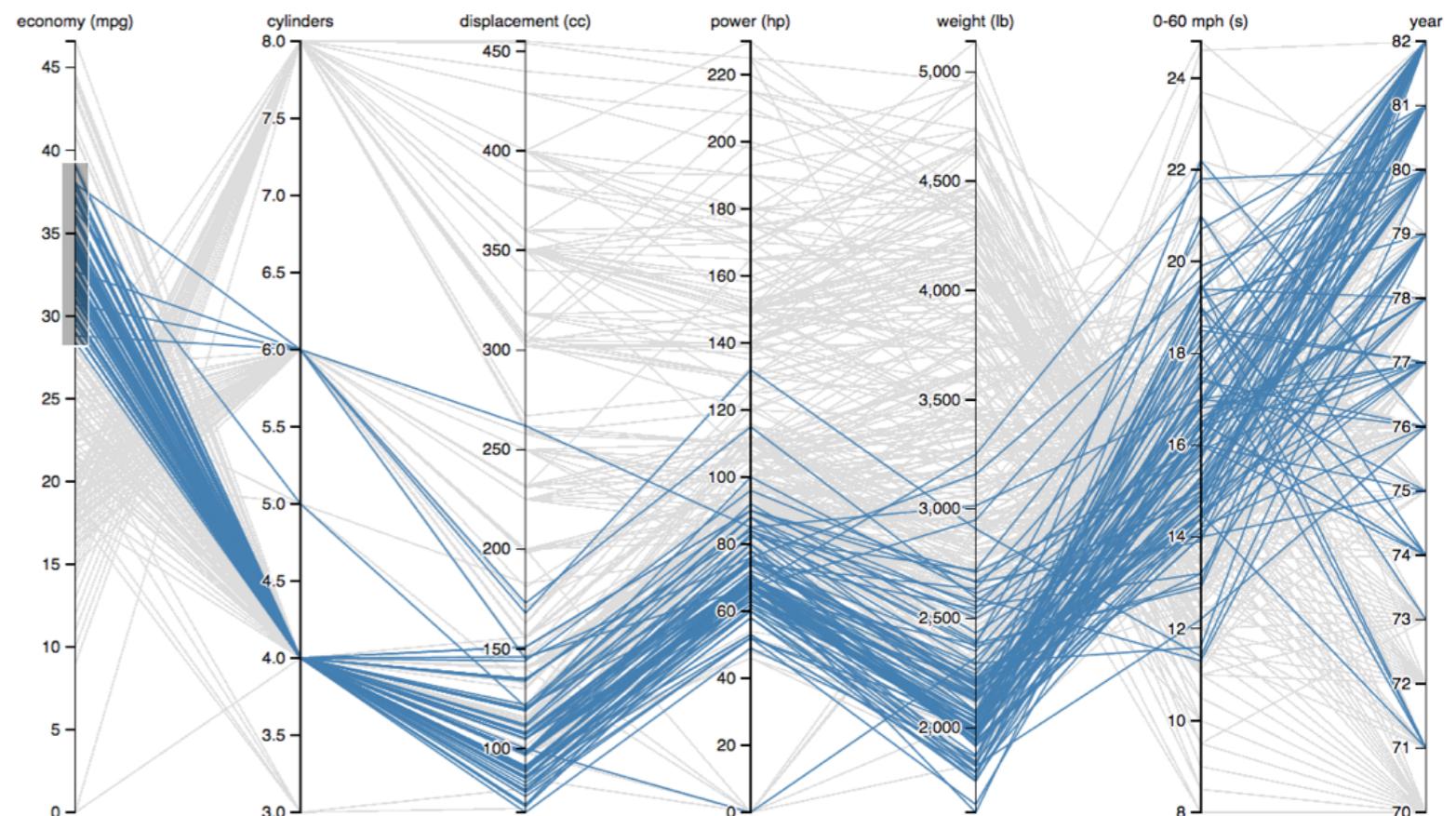
Each axis represents dimension

Lines connecting axis represent records

Suitable for

all tabular data types

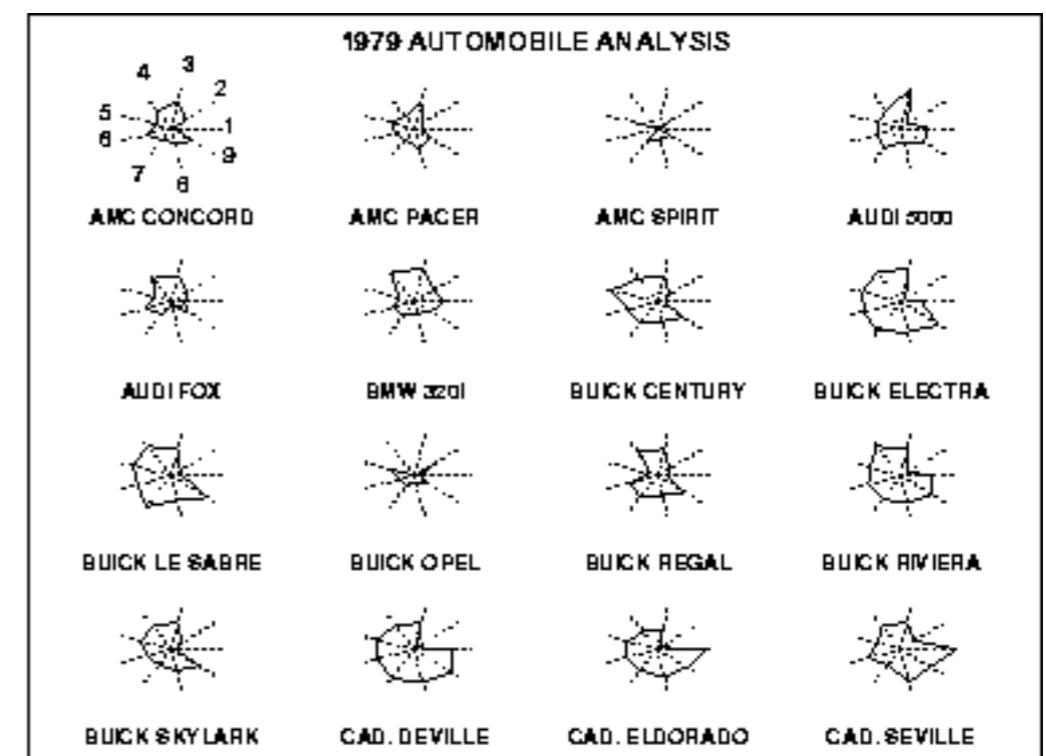
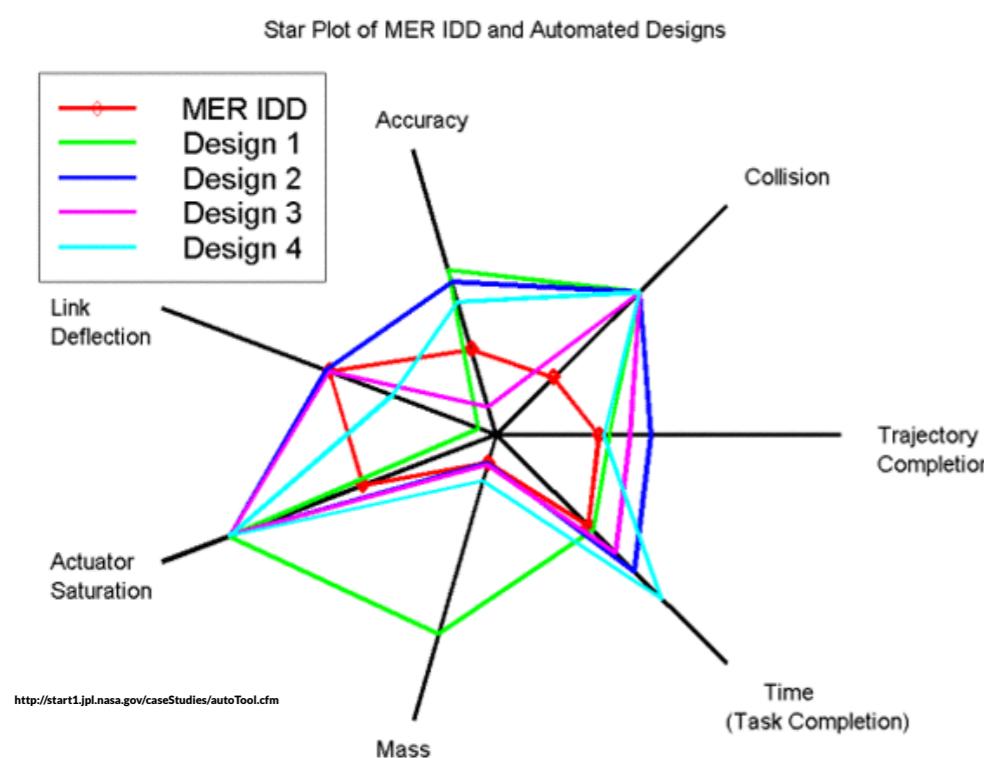
heterogeneous data



Star Plot

[Coekin1969]

Similar to parallel coordinates
Radiate from a common origin



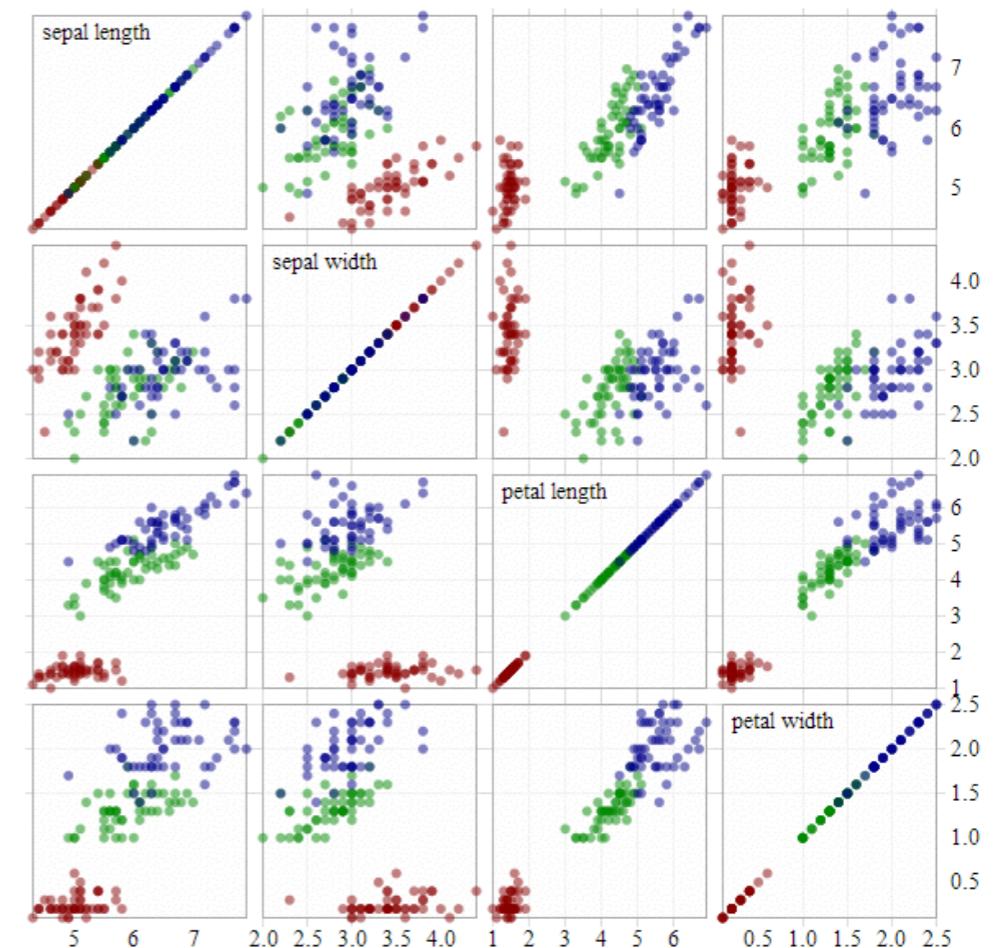
<http://blocks.org/kevinschaul/raw/8833989/>

Scatterplot Matrices (SPLOM)

Matrix of size $d \times d$

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

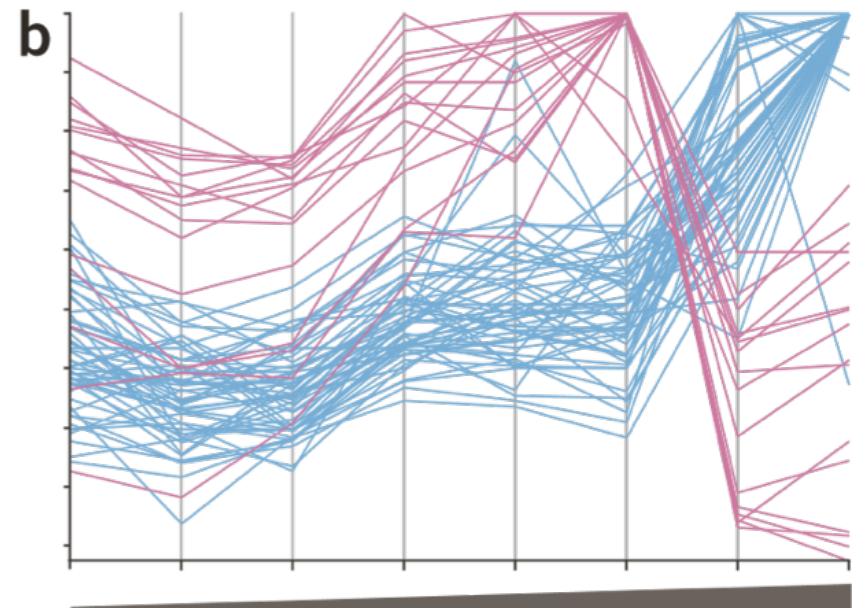
Meaning derived from ordering

If no ordering inherent, clustering is used

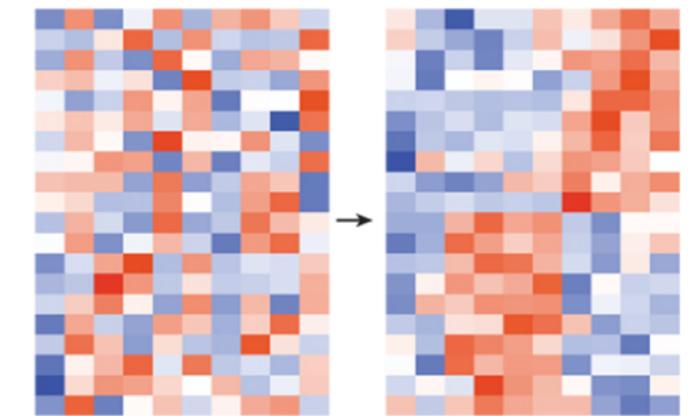
Scalable – 1 px per item

Good for homogeneous data

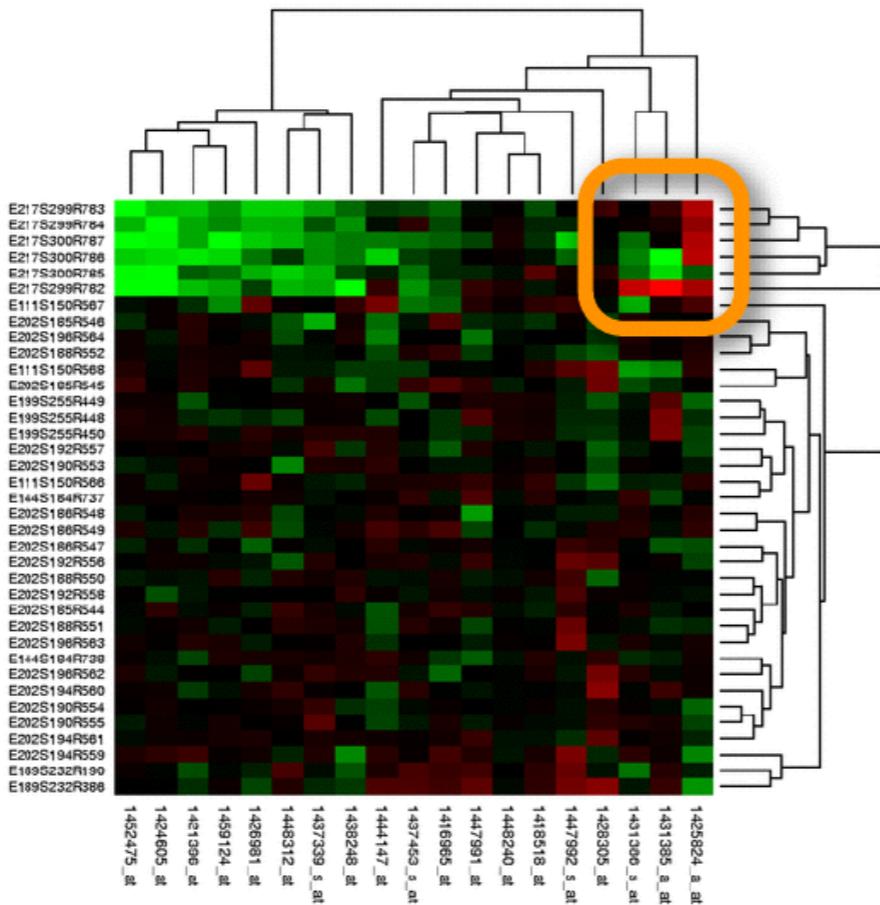
same scale & type



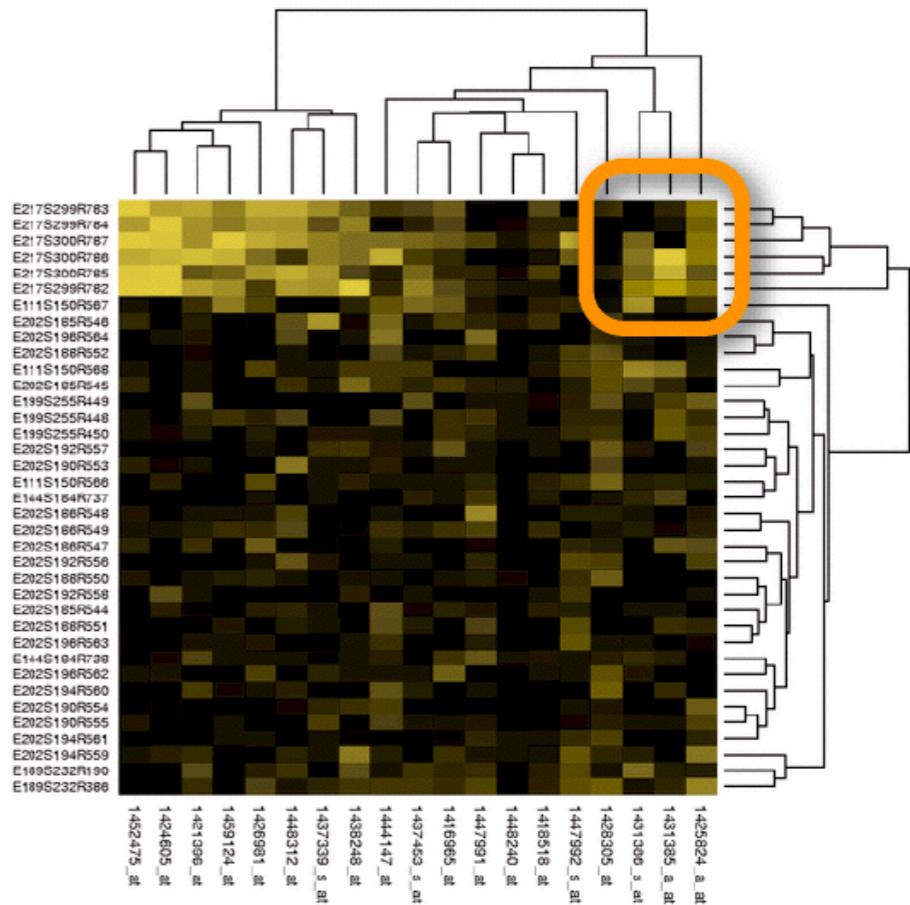
[Gehlenborg & Wong 2012]



Bad Color Mapping

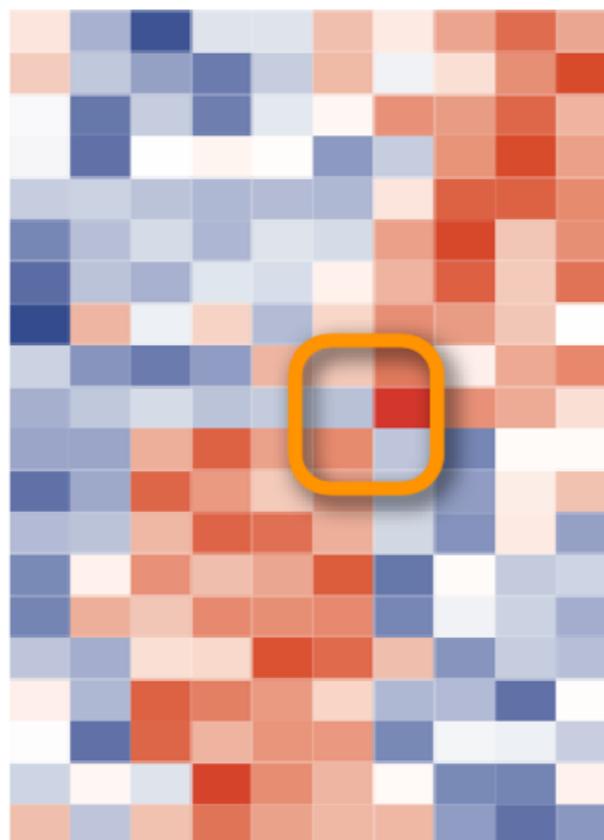


Normal Vision

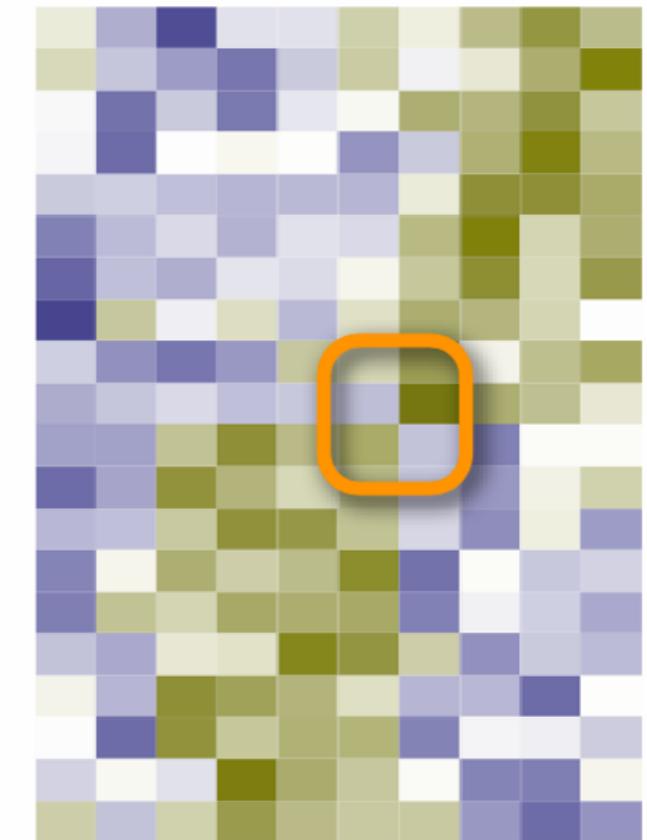


Deuteranope Vision
("Red-Green Blindness")

Good Color Mapping

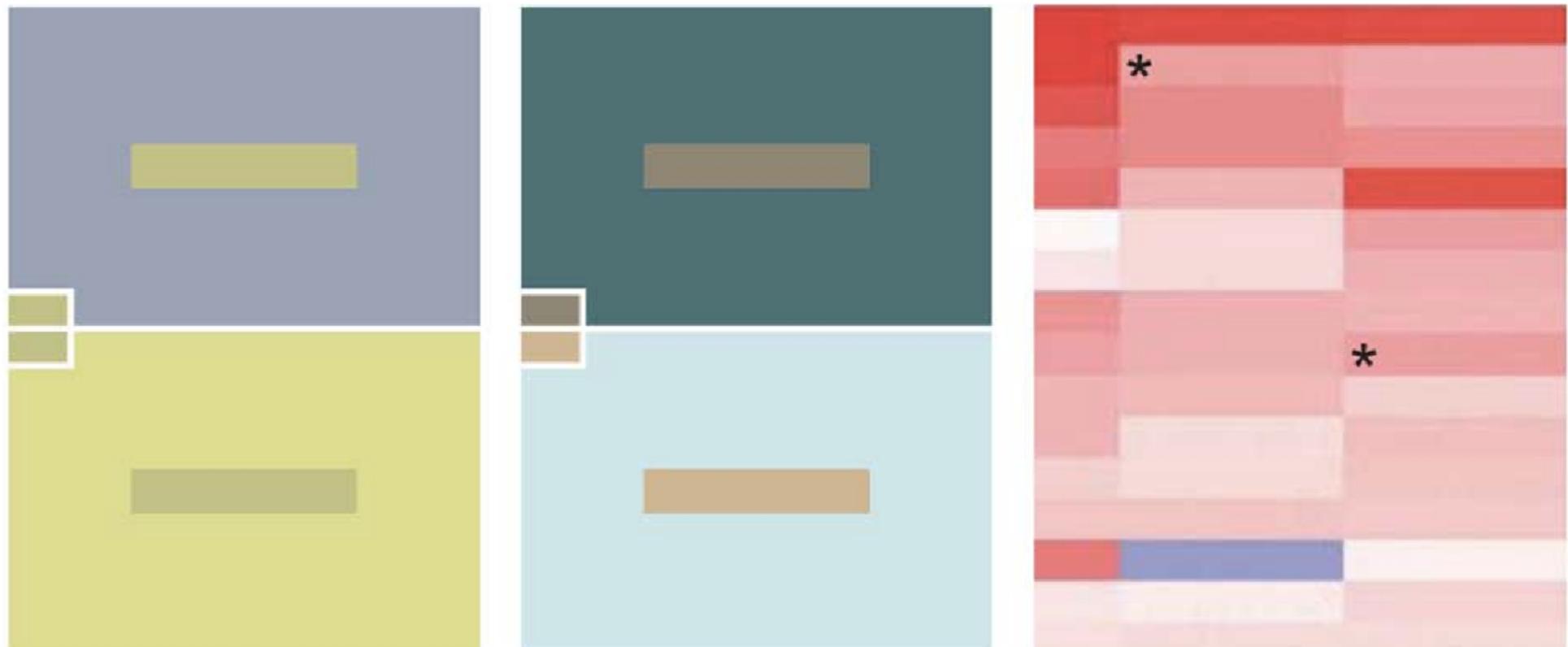


Normal Vision



Deutanope 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.

COMPARING CATEGORIES

Compare values across categories.



PART-TO-WHOLE

Visualizations that relate the part of a variable to its total.



GEOSPATIAL

Relate data to its geography.



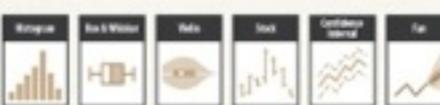
RELATIONSHIP

Illustrate correlations or relationships between variables.



DISTRIBUTION

Illustrate representations of the distribution of data.



TIME

Track changes over time.

