

Lecture 2: Introduction To Data Visualization

Instructor: Saravanan Thirumuruganathan

Outline

- ① Data Mining Terminology
- ② Basics of Visualization
 - Graph integrity
 - 2D visualization
 - Basics of higher dimensional visualization

- Enrollment done for registered students
- Auditing students - send me your email id
- “Search for Teammates” enabled

- Instant Student Feedback
- Accessible via Smart Phone, Tablet, Laptop
- No login needed from Student's end
- Use a consistent name throughout the semester

In-Class Quizzes

- URL: <http://m.socrative.com/>
- Room Name: **4f2bb99e**

Misc Announcements

- Slides for Lecture 1 updated
- Change Office hour timings?
- Installation of Scientific Python

Other Relevant Online Classes

- Machine Learning, Stanford:
<https://www.coursera.org/course/ml>
- Mining of Massive Datasets, Stanford:
<https://www.coursera.org/course/mmds>
- Statistical Learning, Stanford: <https://class.stanford.edu/courses/HumanitiesandScience/StatLearning/Winter2015/about>

Data Mining Terminology

Data Matrix

Table 1.1. Extract from the Iris dataset

	Sepal length	Sepal width	Petal length	Petal width	Class
	X_1	X_2	X_3	X_4	X_5
\mathbf{x}_1	5.9	3.0	4.2	1.5	Iris-versicolor
\mathbf{x}_2	6.9	3.1	4.9	1.5	Iris-versicolor
\mathbf{x}_3	6.6	2.9	4.6	1.3	Iris-versicolor
\mathbf{x}_4	4.6	3.2	1.4	0.2	Iris-setosa
\mathbf{x}_5	6.0	2.2	4.0	1.0	Iris-versicolor
\mathbf{x}_6	4.7	3.2	1.3	0.2	Iris-setosa
\mathbf{x}_7	6.5	3.0	5.8	2.2	Iris-virginica
\mathbf{x}_8	5.8	2.7	5.1	1.9	Iris-virginica
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
\mathbf{x}_{149}	7.7	3.8	6.7	2.2	Iris-virginica
\mathbf{x}_{150}	5.1	3.4	1.5	0.2	Iris-setosa

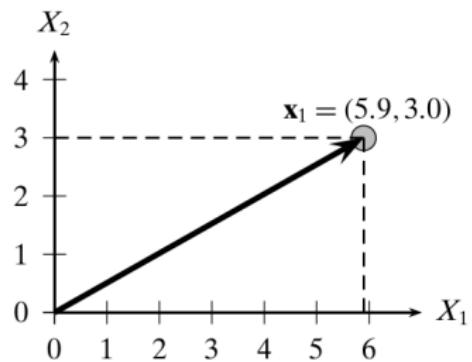
Data Matrix

$$\mathbf{D} = \left(\begin{array}{c|ccccc} & X_1 & X_2 & \cdots & X_d \\ \hline \mathbf{x}_1 & x_{11} & x_{12} & \cdots & x_{1d} \\ \mathbf{x}_2 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n & x_{n1} & x_{n2} & \cdots & x_{nd} \end{array} \right)$$

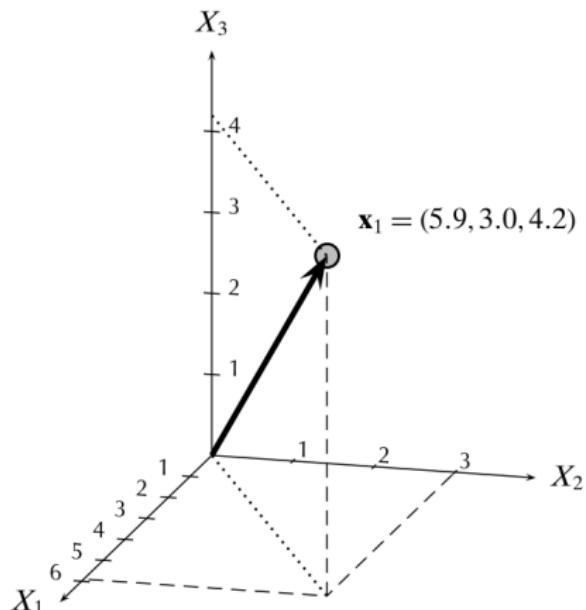
Data Matrix

- n rows and d columns
- Row \Rightarrow Tuple/Entities
- Column \Rightarrow attribute/feature
- Special column called **Class**
- x_i : i -th row, X_j : j -th column
- Row \Rightarrow entities, instances, examples, records, transactions, objects, points, feature-vectors, tuples
- Column \Rightarrow attributes, properties, features, dimensions, variables, fields
- $n \Rightarrow$ size, $d \Rightarrow$ dimensionality of data

Geometric View



(a)

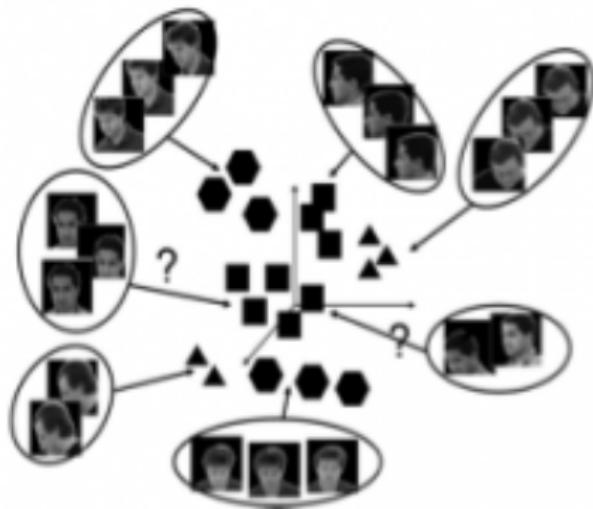


(b)

Figure 1.1. Row \mathbf{x}_1 as a point and vector in (a) \mathbb{R}^2 and (b) \mathbb{R}^3 .

Implications

- Each photo in the universe is some point in high dimension
- Each book (written or in future) are some point in high dimension



Data Types

Ben Shneiderman, 1996:¹

- 1D (sequences)
- Temporal
- 2D (maps)
- 3D (shaped)
- nD (relational)
- Trees (hierarchical)
- Networks (graphs)
- Others (text)

¹The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization [Shneiderman, 96]

Semantics vs. Types

- Data Semantics: real-world meaning
 - e.g., company name, day of the month, person height, etc.
- Data Type: Interpretation in terms of scales of measurements
 - e.g., quantity or category, sensible mathematical operations etc.

Data Types

- Nominal (Categorical) (N)

Are = or \neq to other values

Apples, Oranges, Bananas,...



- Ordinal (O)

Obey a $<$ relationship

Small, medium, large



- Quantitative (Q)

Can do arithmetic on them

10 inches, 23 inches, etc.



On the theory of scales and measurements [S. Stevens, 46]

Data Types

- Q - Interval (location of zero arbitrary)
 - Dates: Jan 19; Location: (Lat, Long)
 - Like a geometric point. Cannot compare directly.
 - Only differences (i.e., intervals) can be compared
- Q - Ratio (zero fixed)
 - Measurements: Length, Mass, Temp, ...
 - Origin is meaningful, can measure ratios & proportions
 - Like a geometric vector, origin is meaningful

Data Types

- N - Nominal (labels)
 - Operations: $=, \neq$
- O - Ordinal (ordered)
 - Operations: $=, \neq, >, <$
- Q - Interval (location of zero arbitrary)
 - Operations: $=, \neq, >, <, +, -$
- Q - Ratio (zero fixed)
 - Operations: $=, \neq, >, <, +, -, \times, \div$

Quiz!

What is the data type of:

- Gender:

Quiz!

What is the data type of:

- Gender: Categorical/Nominal
- Age:

Quiz!

What is the data type of:

- Gender: Categorical/Nominal
- Age: Ordinal
- Height:

Quiz!

What is the data type of:

- Gender: Categorical/Nominal
- Age: Ordinal
- Height: Quantitative - Ratio
- Date:

Quiz!

What is the data type of:

- Gender: Categorical/Nominal
- Age: Ordinal
- Height: Quantitative - Ratio
- Date: Quantitative - Interval

Data Dimensions

- Univariate (1D)
- Bivariate (2D)
- Trivariate (3D)
- Multivariate (nD)

Introduction To Data Visualization

Visualization Goals

- **Presentation**

- Known facts about data
- Task: Communicate results

- **Exploration**

- Data without hypothesis
- Task: Generate hypothesis

- **Confirmation**

- Hypothesis is given
- Task: Verify / falsify hypothesis

Visualization Goals

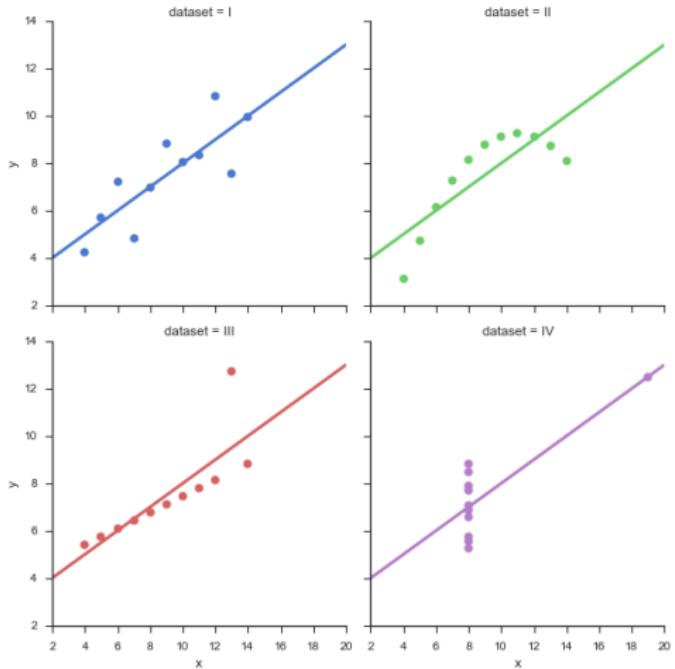
“The greatest value of a picture is when it forces us to notice what we never expected to see.”

-John Tukey (1915 - 2000)



Anscombe's Quartet

Same mean, variance, correlation, and linear regression line

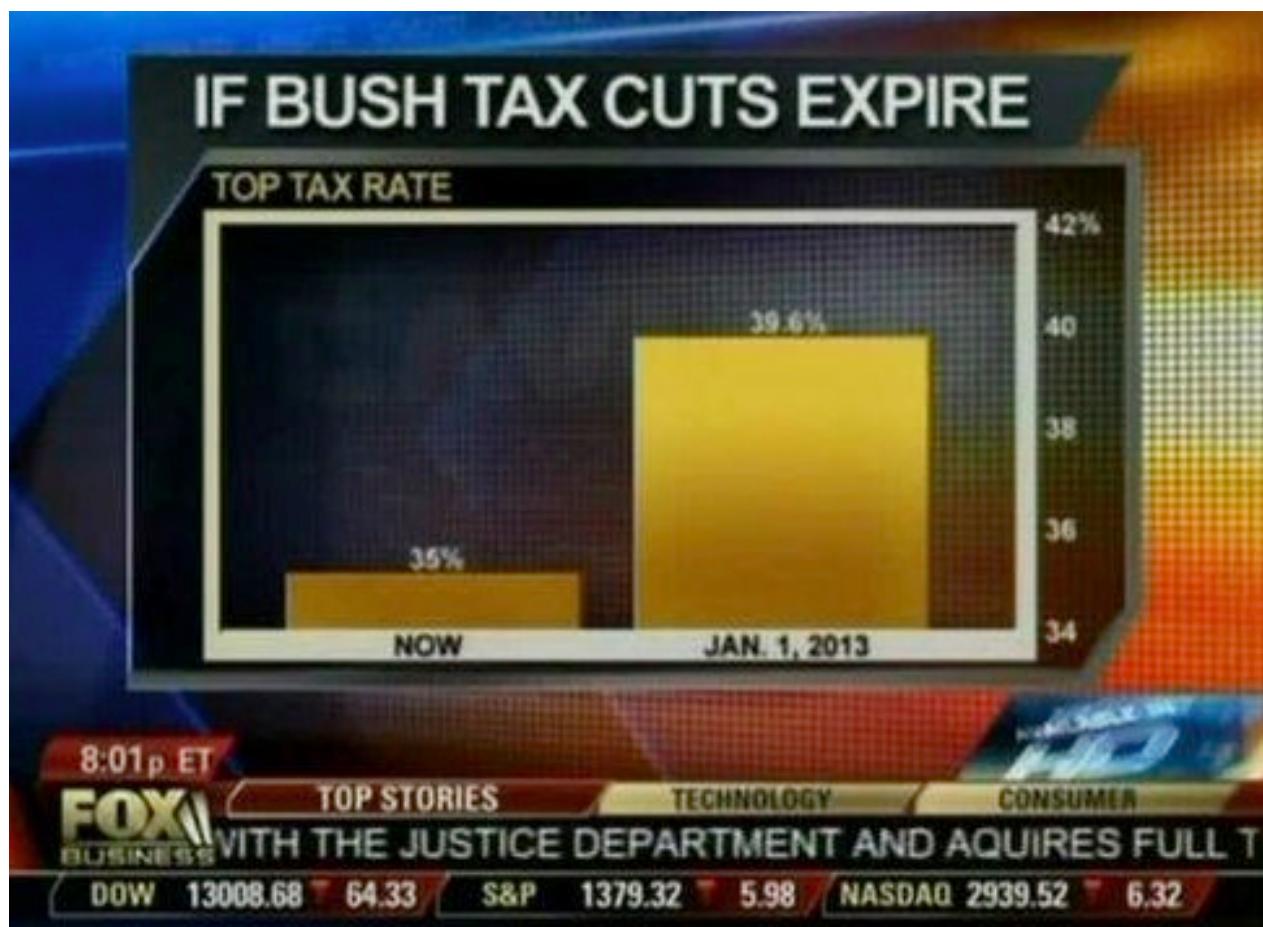


Graphical Integrity

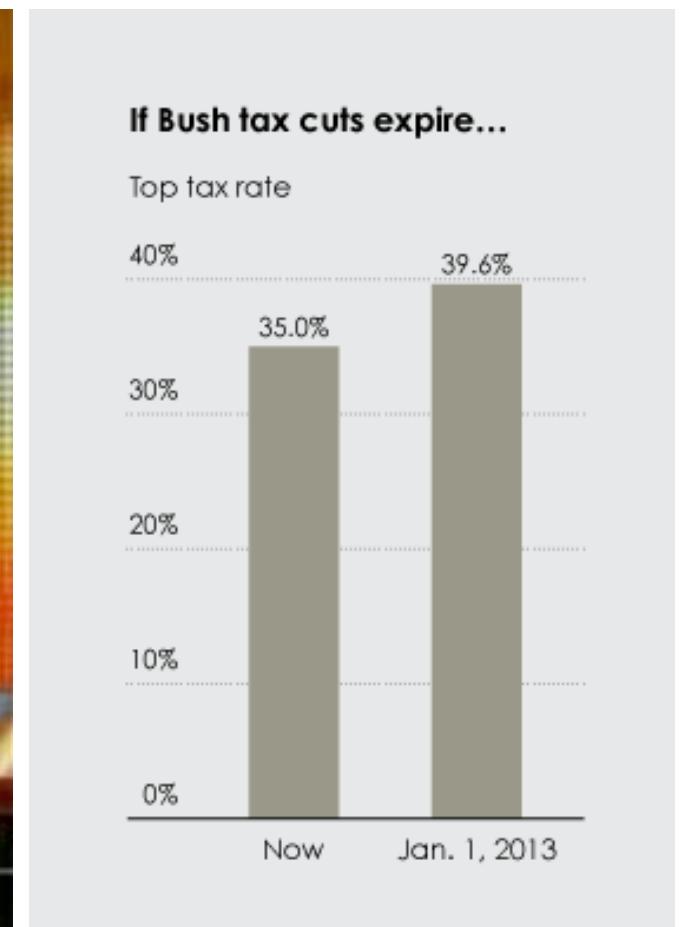
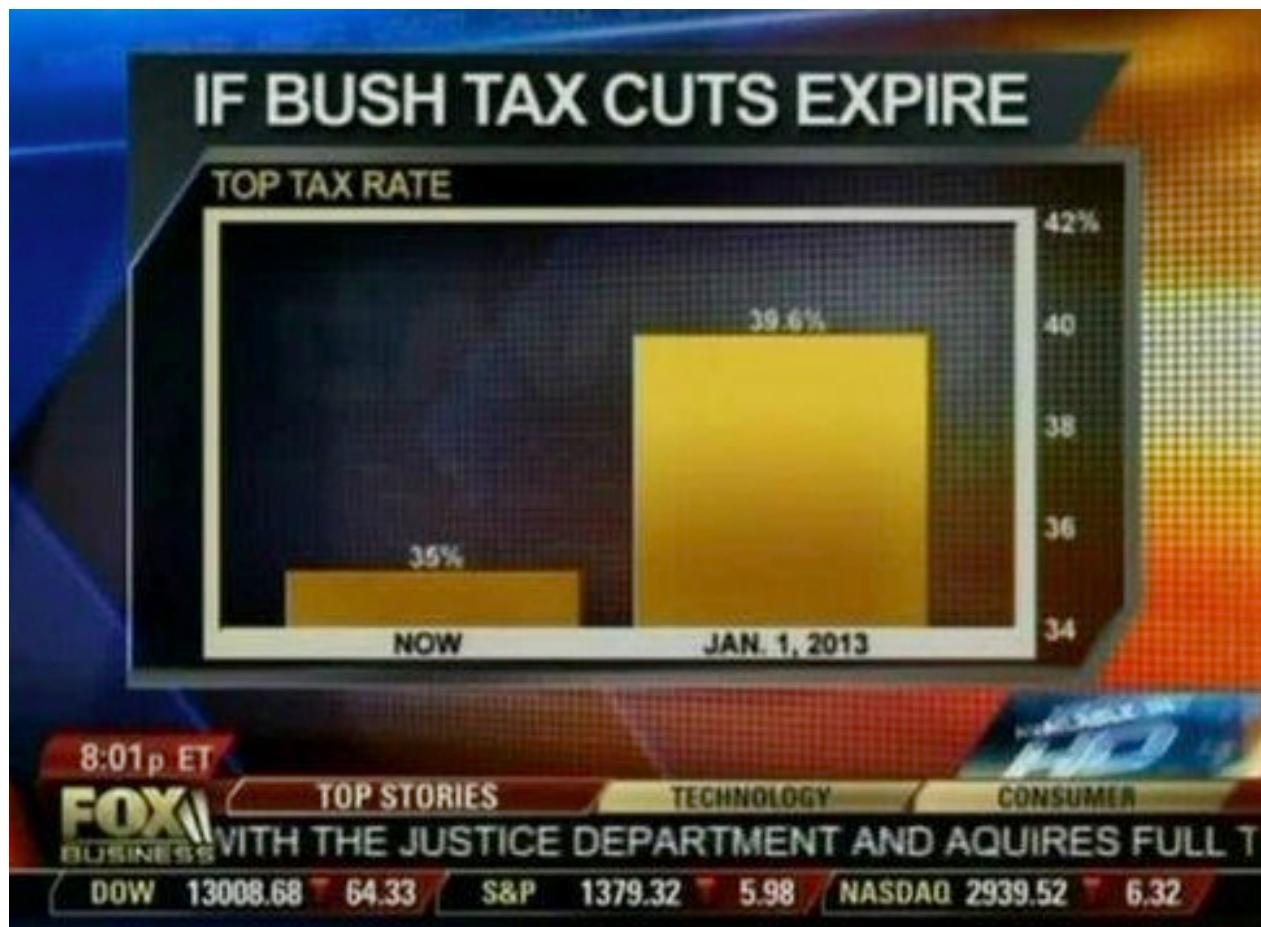
“There are three kinds of lies: lies, damned lies, and statistics.”

- attributed to Benjamin Disraeli in 19th Century

Graphical Integrity



Scale Distortions



Scale Distortions

How 2012 STACKS UP

THE WARMEST YEARS ON RECORD
CONTIGUOUS U.S.

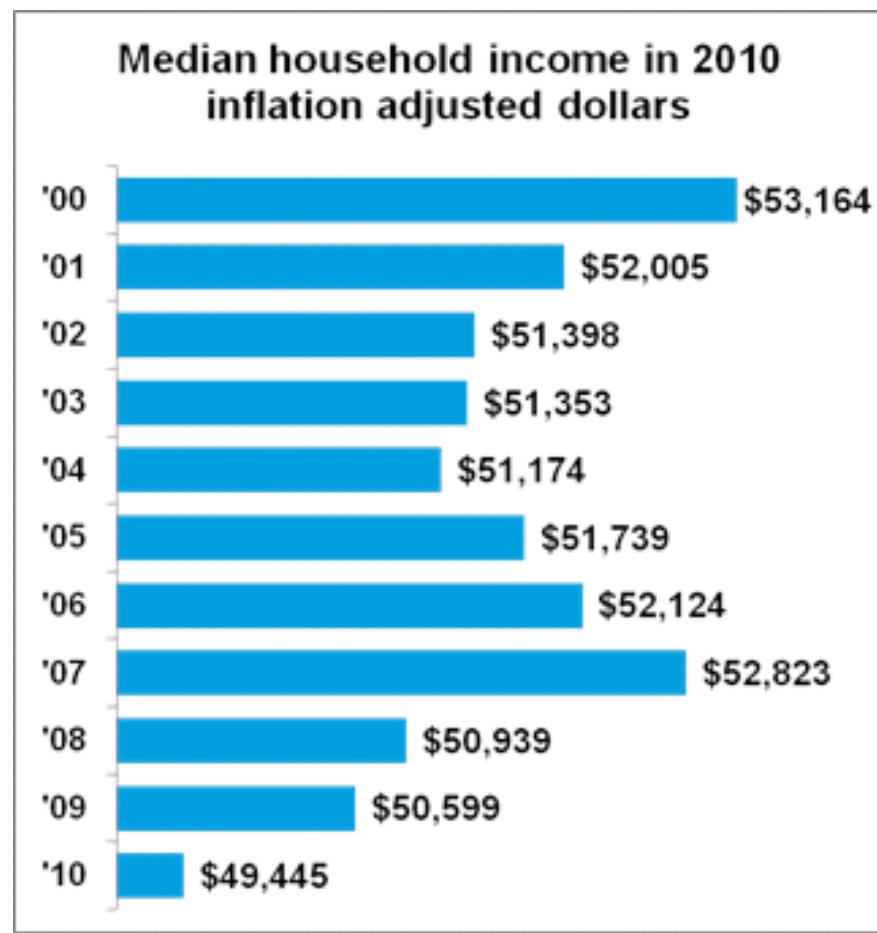


Source: NOAA's National Climatic Data Center - State of the Climate National Overview

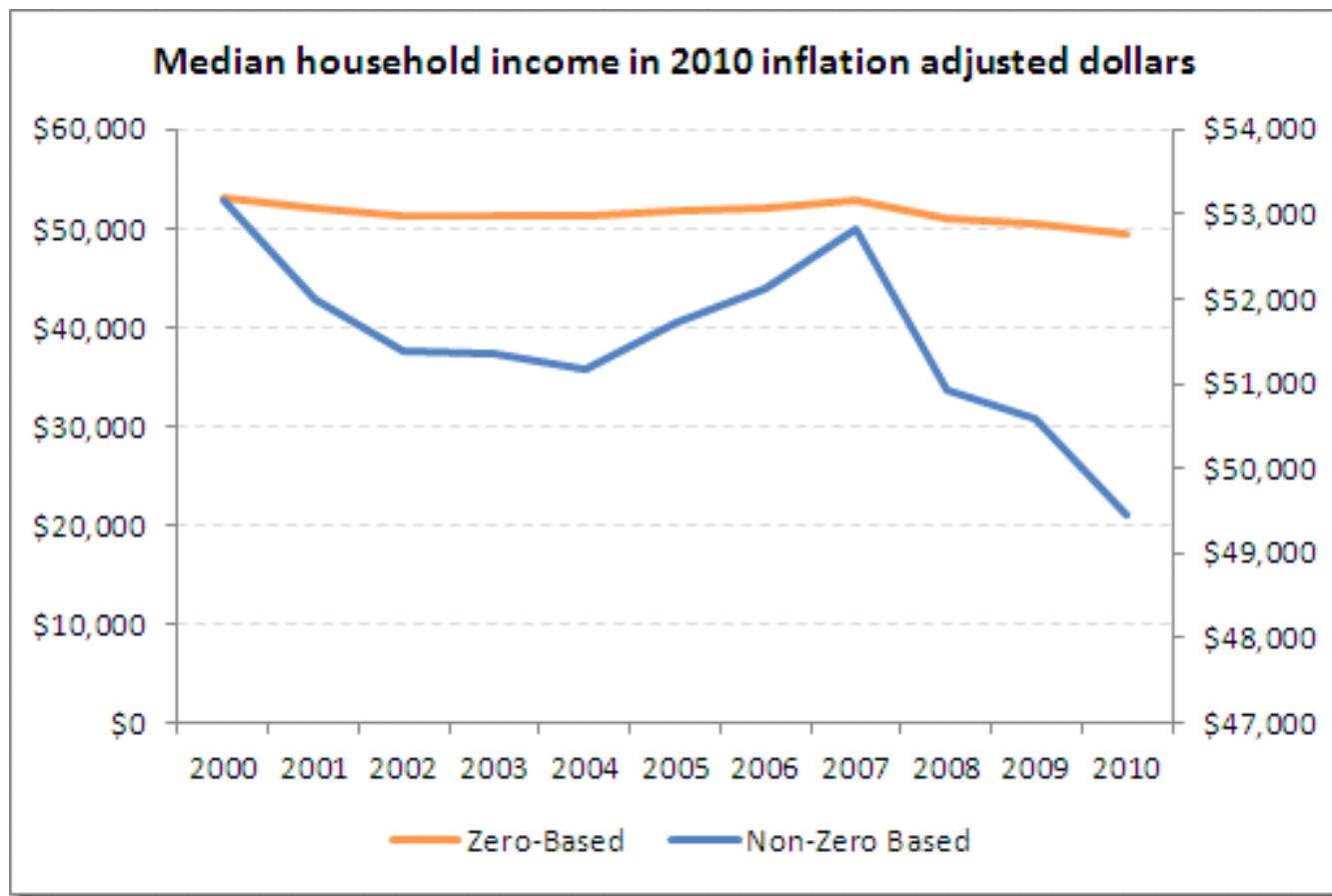
CLIMATE  CENTRAL

Scale Distortions

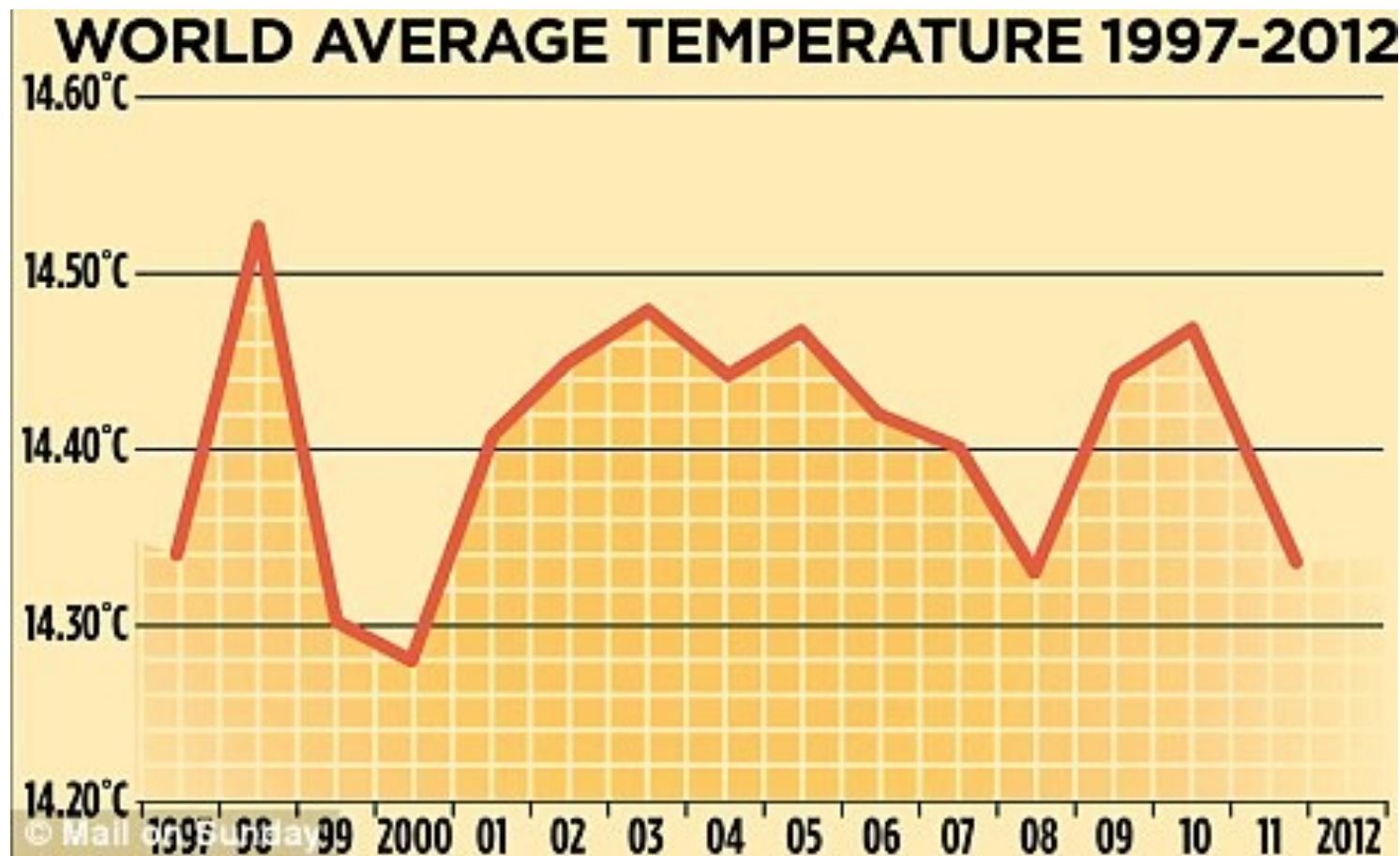
Always start your bar graphs at zero!



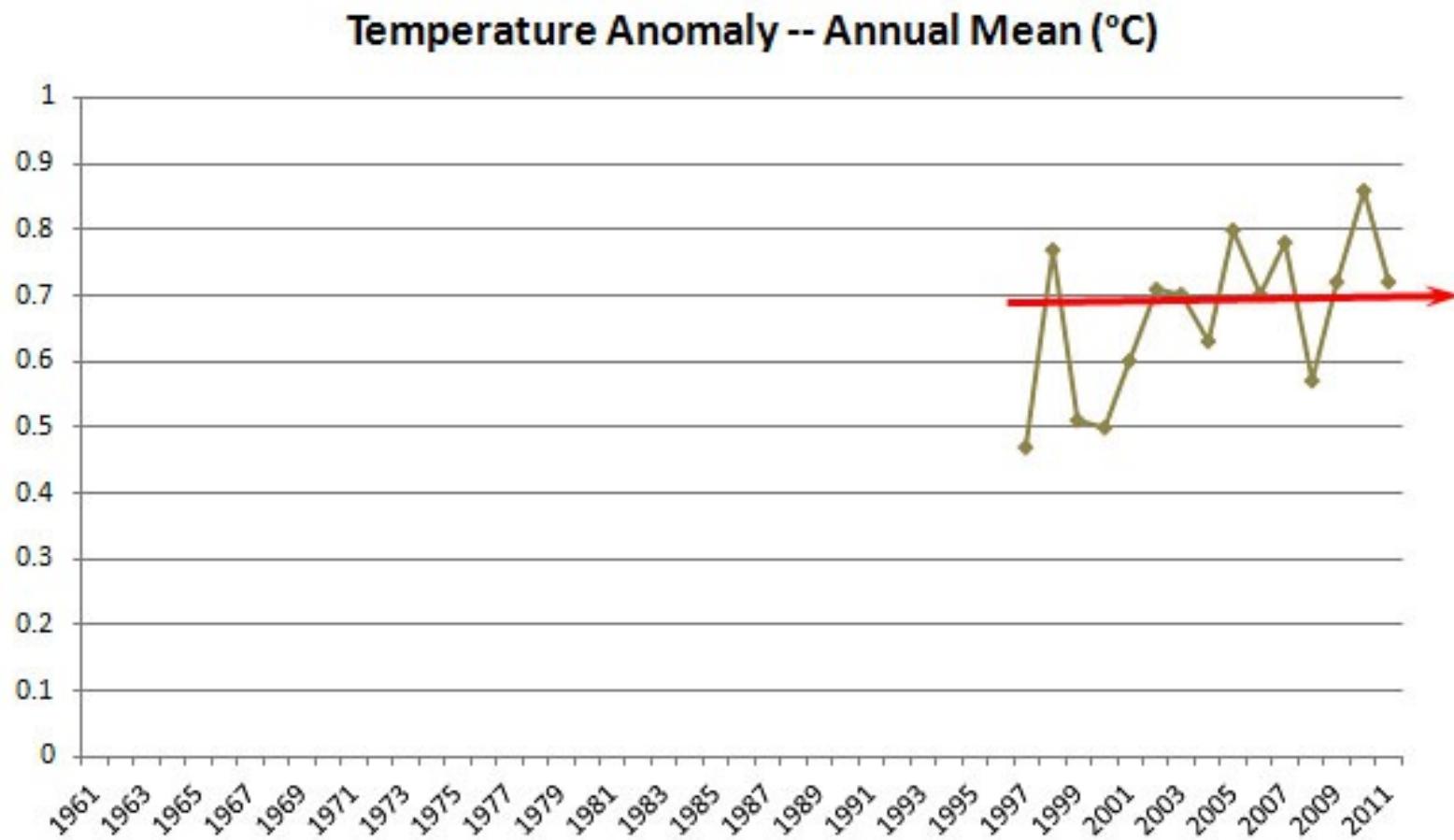
Scale Distortions



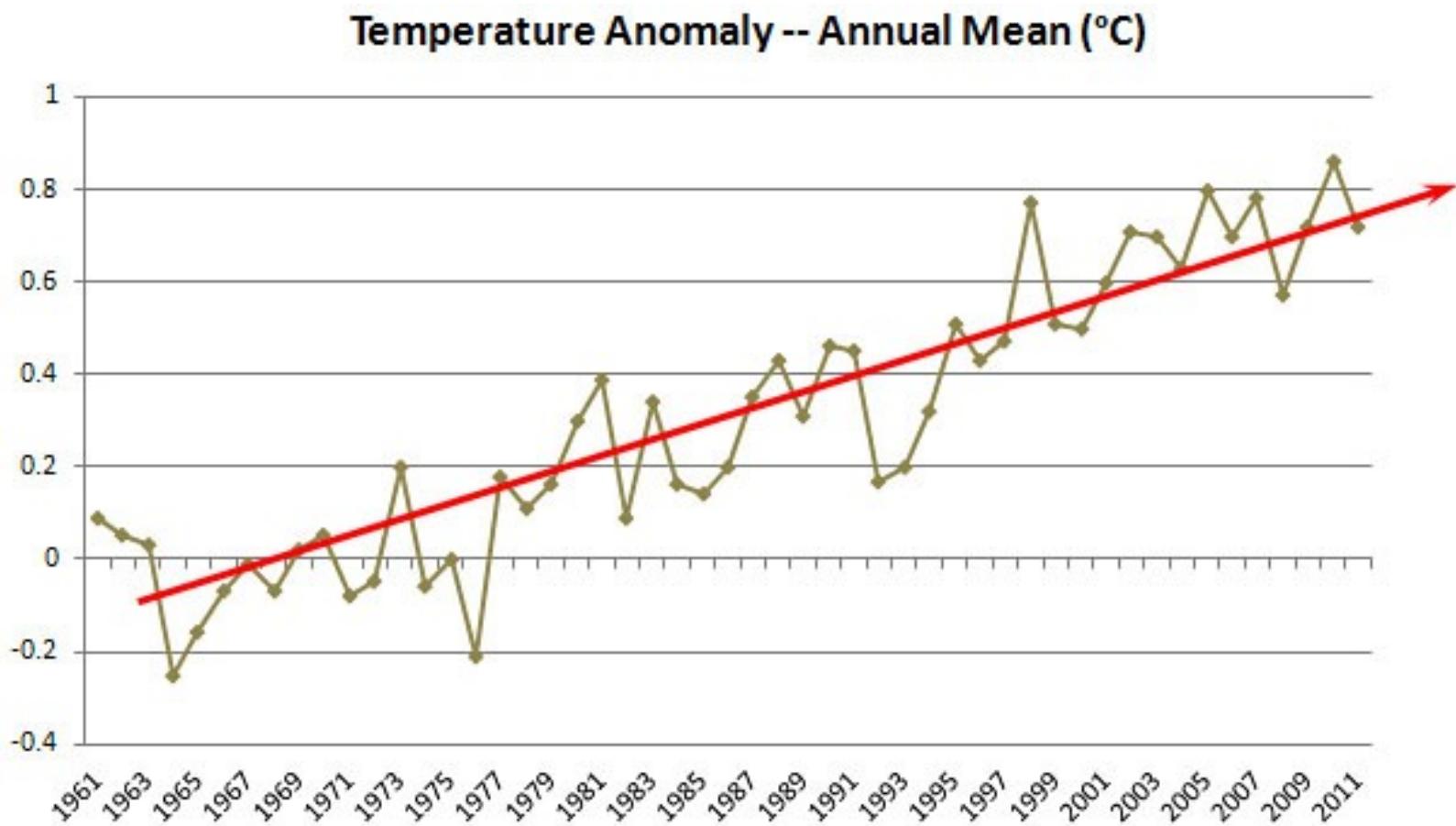
Global Warming?



Global Warming?



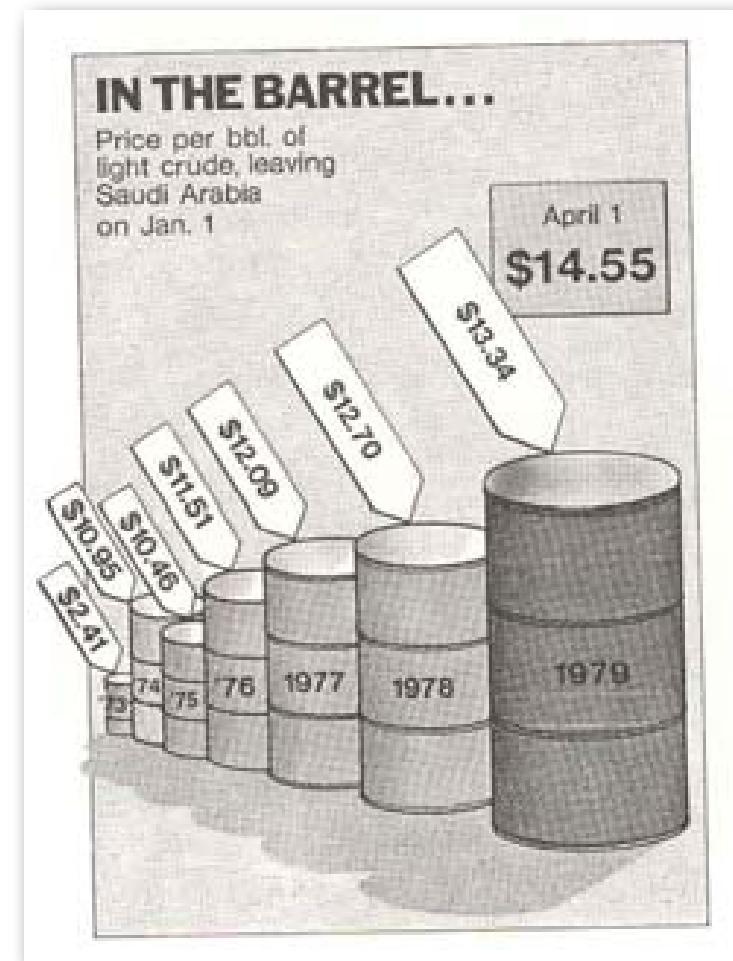
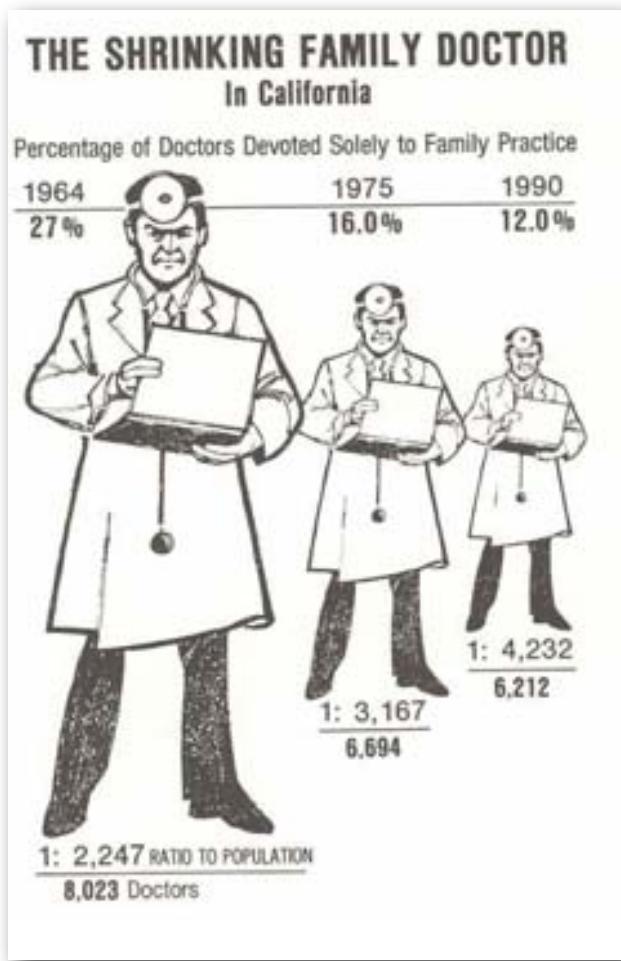
Global Warming!



The Lie Factor

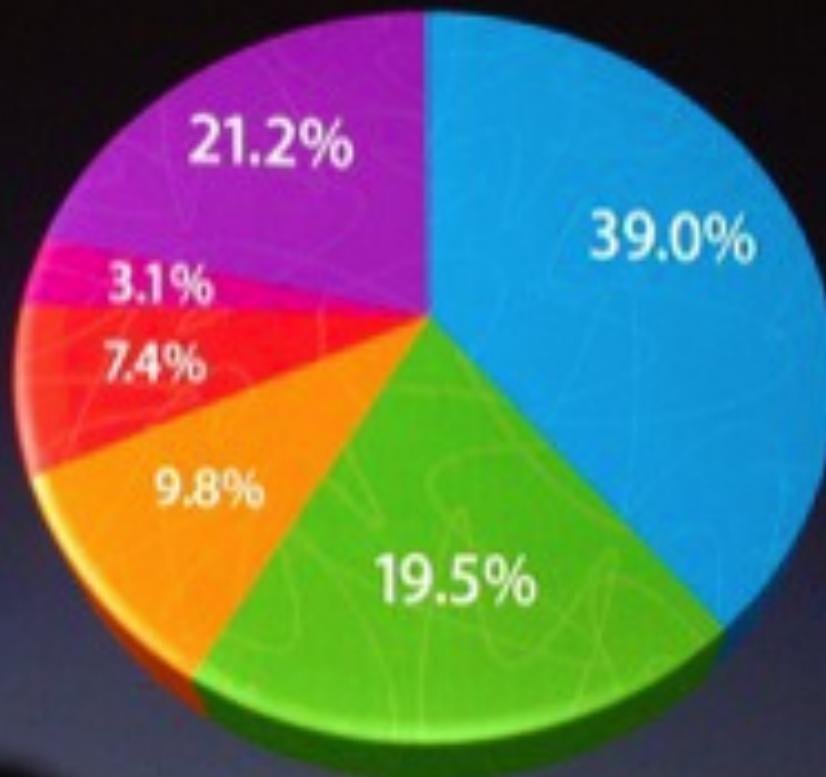
Size of effect shown in graphic

Size of effect in data

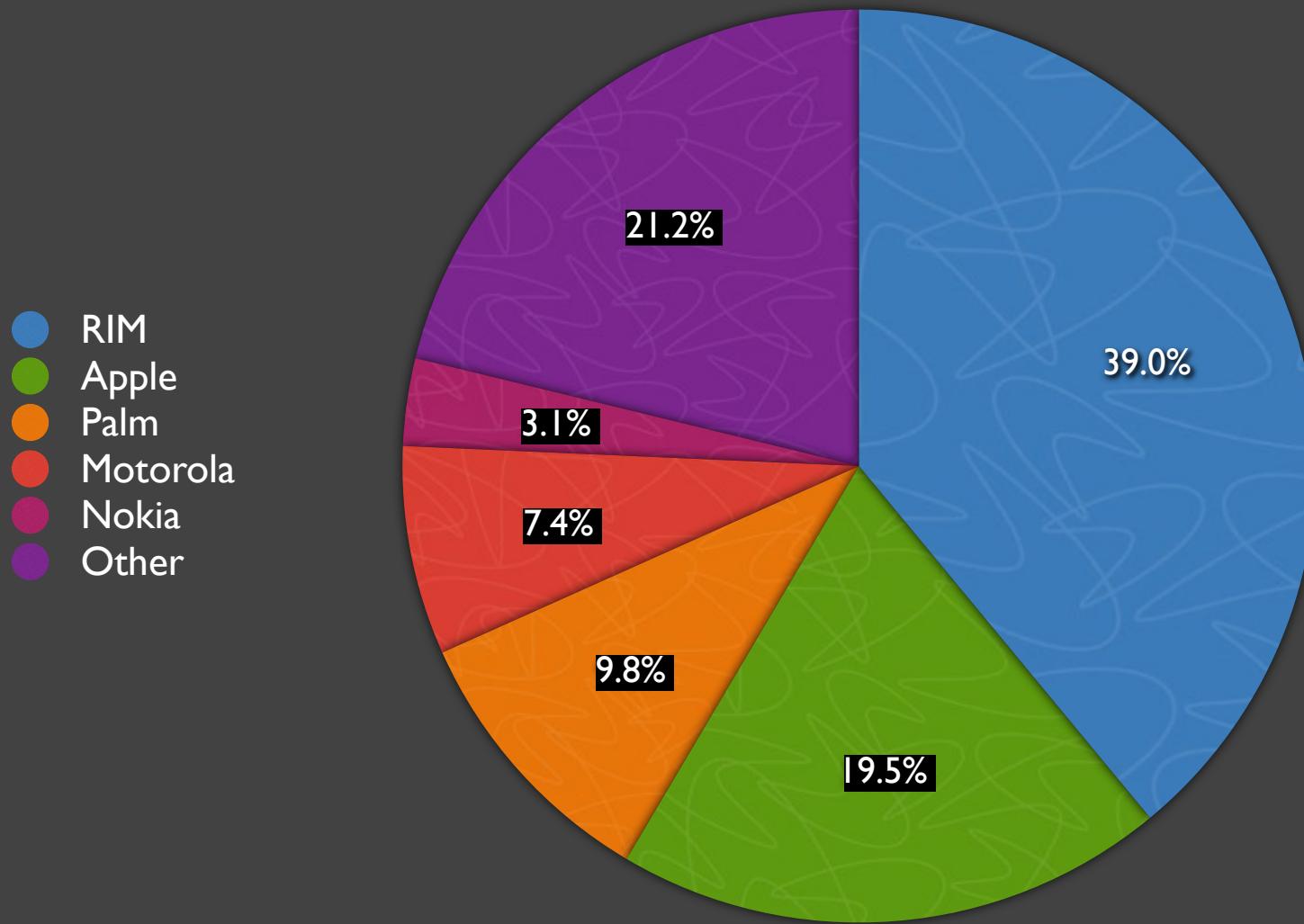


U.S. SmartPhone Marketshare

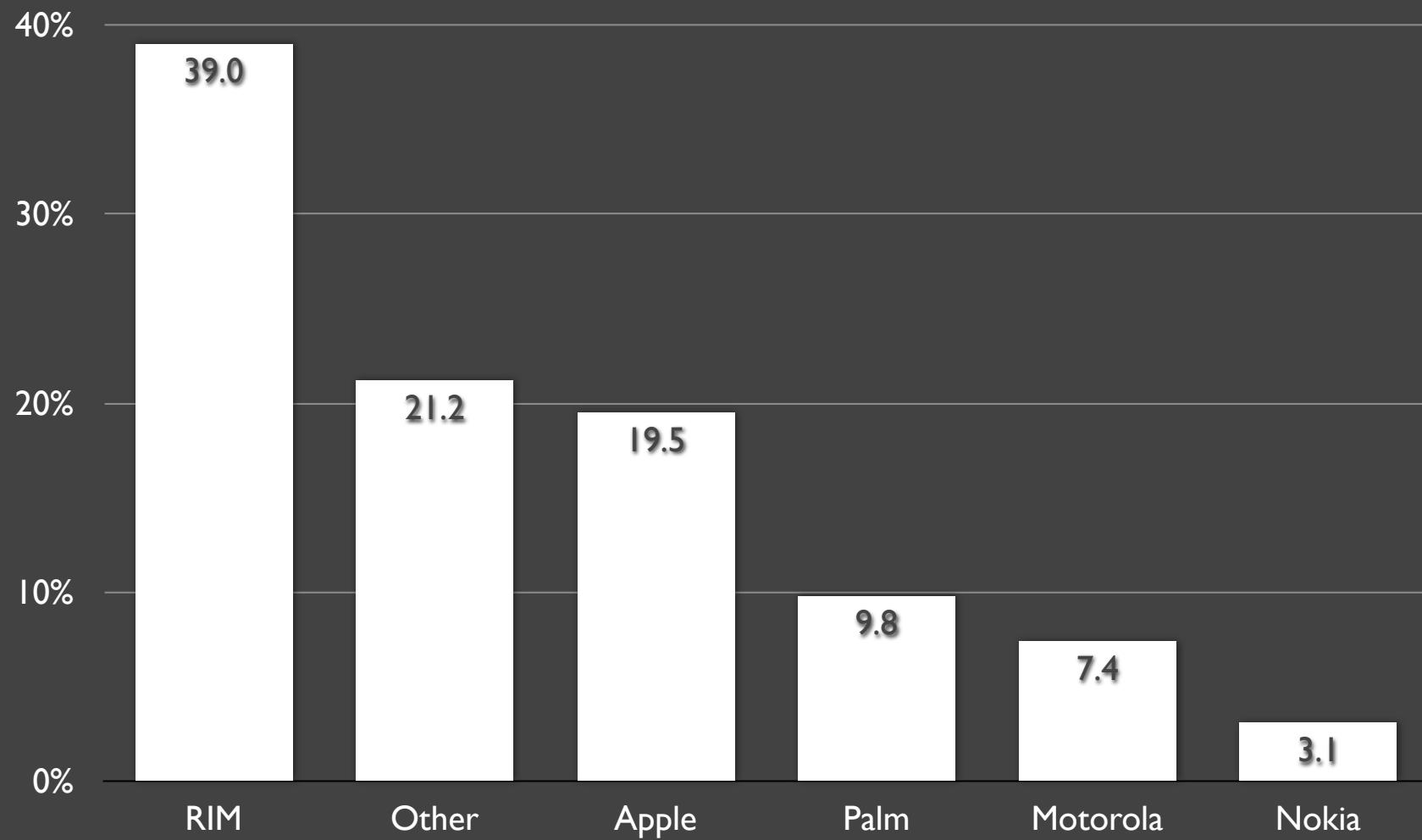
- RIM
- Apple
- Palm
- Motorola
- Nokia
- Other



U.S. SmartPhone Marketshare



U.S. SmartPhone Marketshare



Labelling Chart Axes



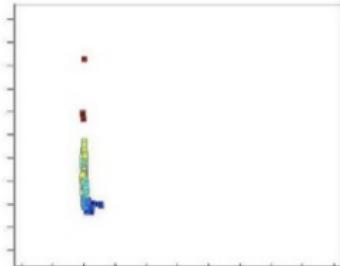
²<http://xkcd.com/833/>

Lying with Scales

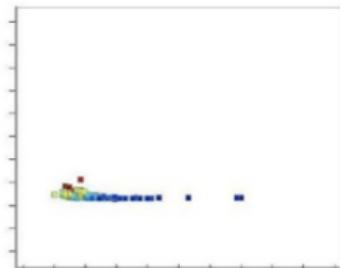
Same data - different scales



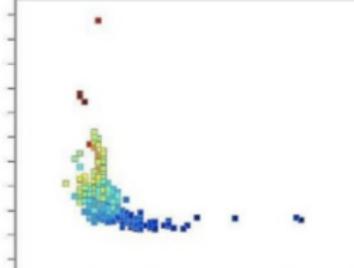
Uniform scale in both x and y



Larger scale in y



Larger scale in x



Larger scale in x and y

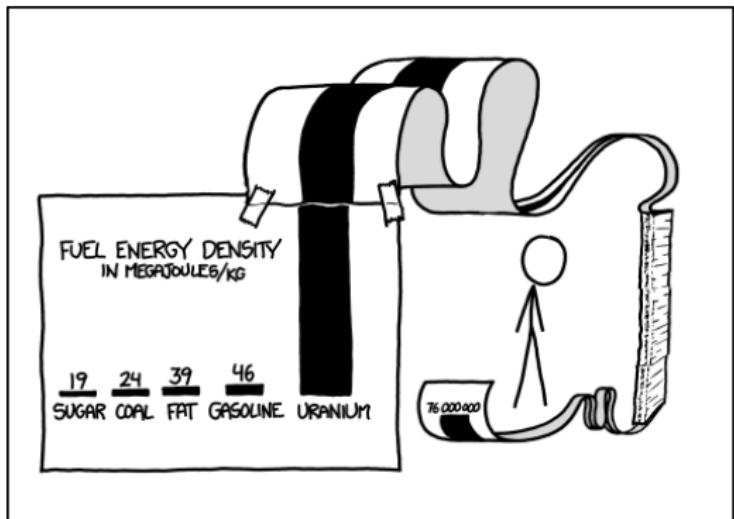
3

³Ward, Grinstein, Keim, 2011

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?

Log Scale



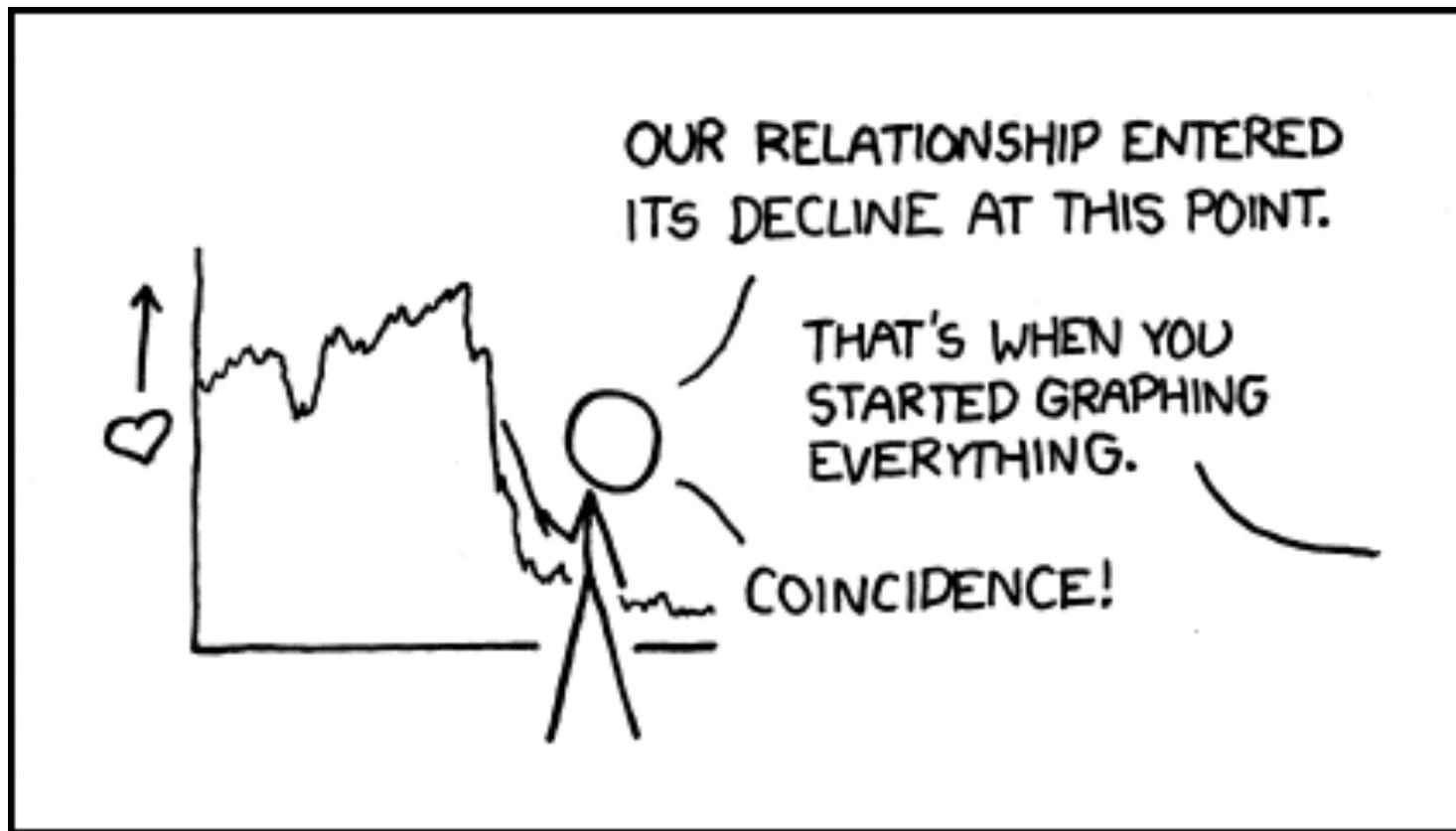
SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT PROPERLY.

4

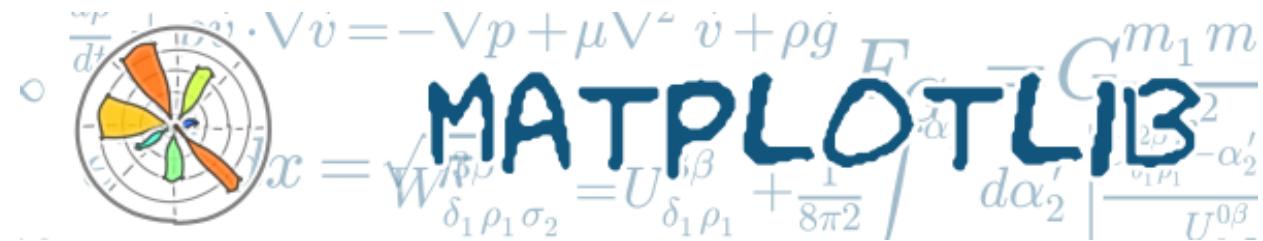
⁴<http://xkcd.com/1162/>

Graph Types (2D and nD)

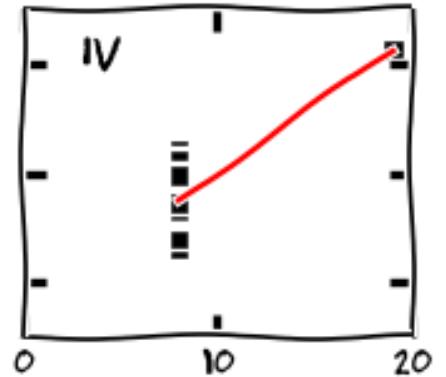
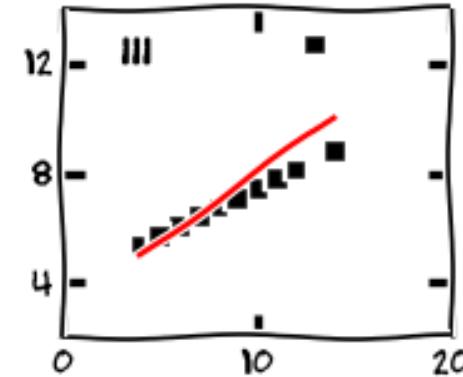
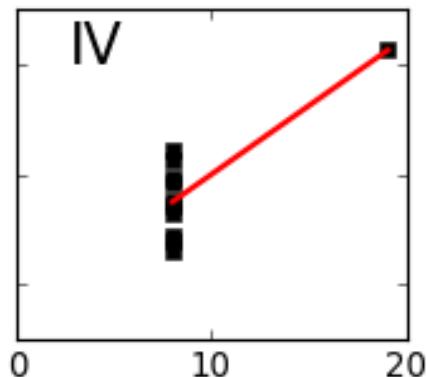
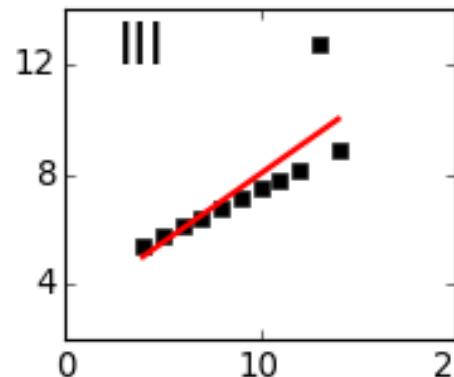
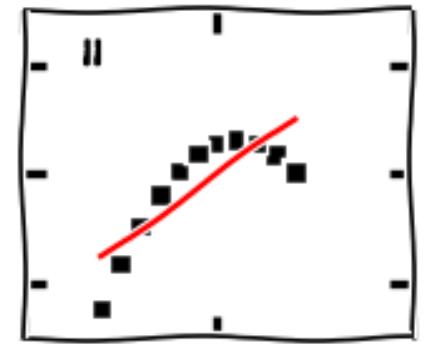
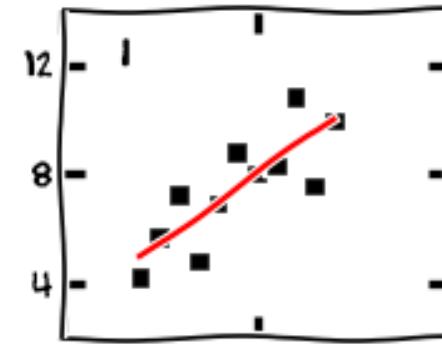
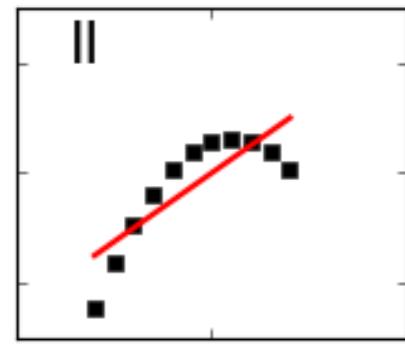
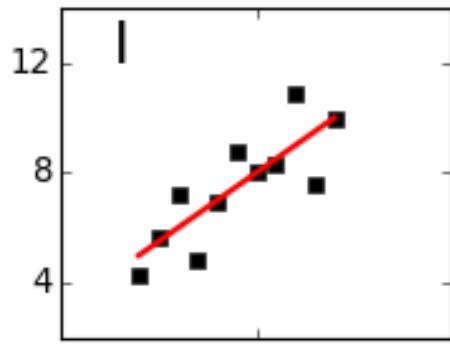
Statistical Graph Types



Side Note

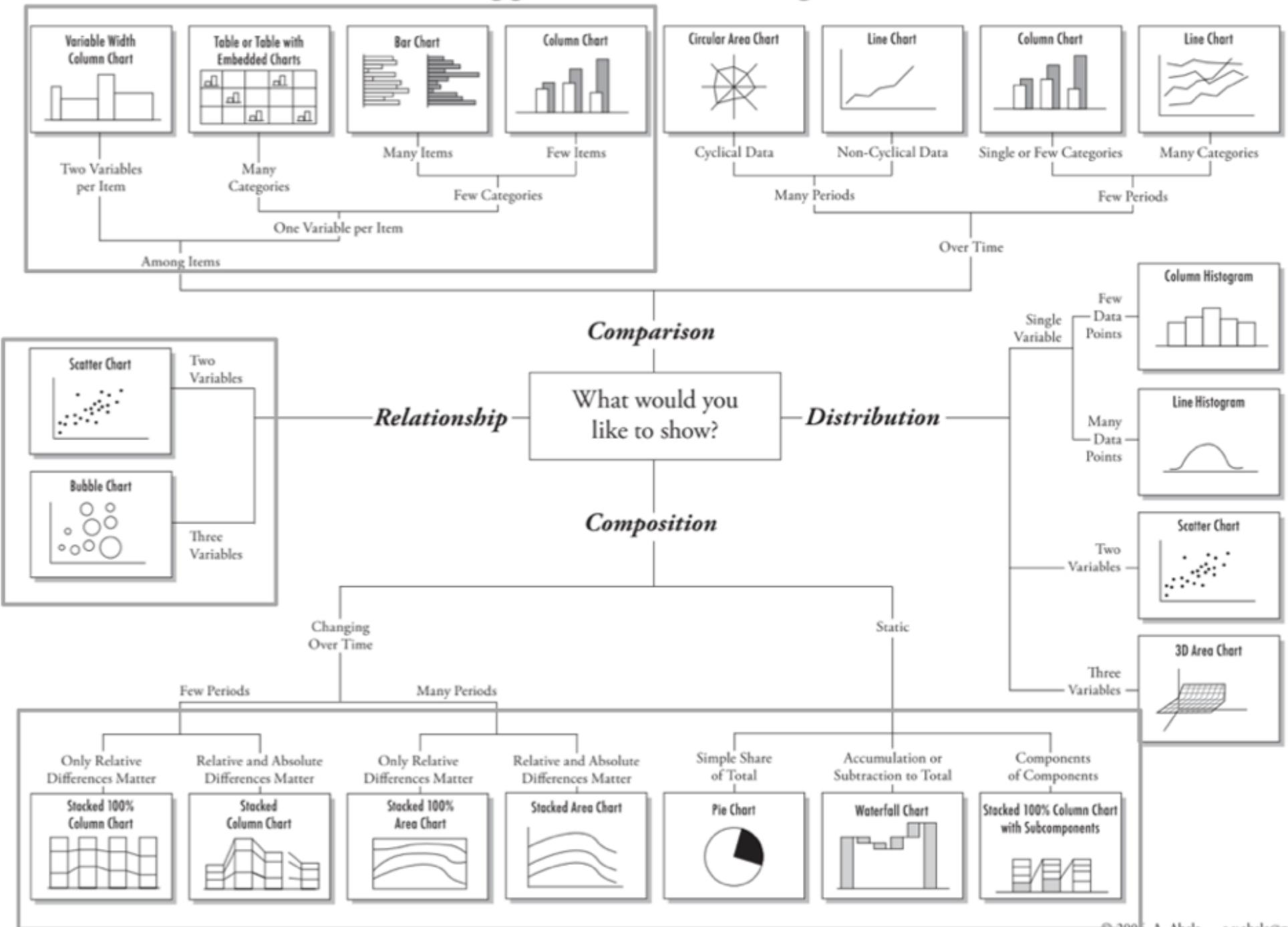


XKCD-ify your plot



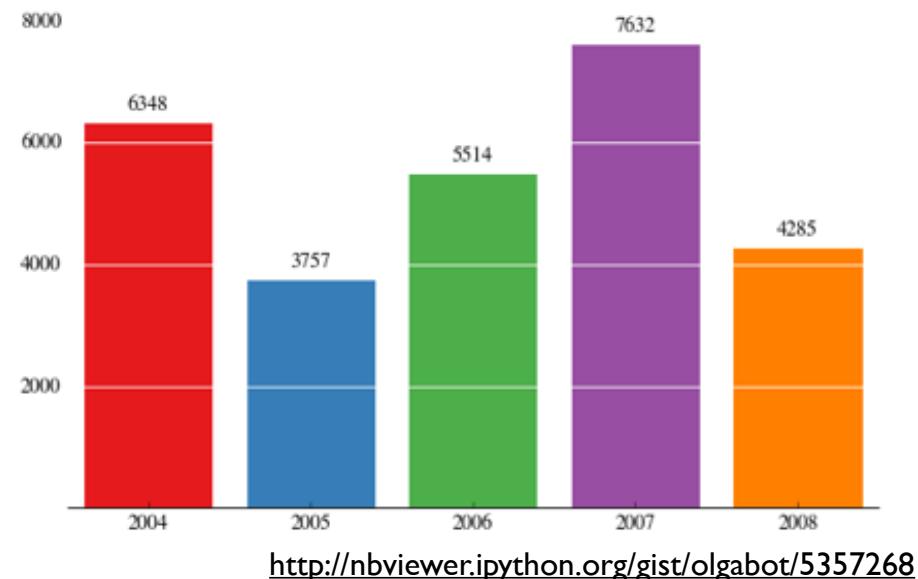
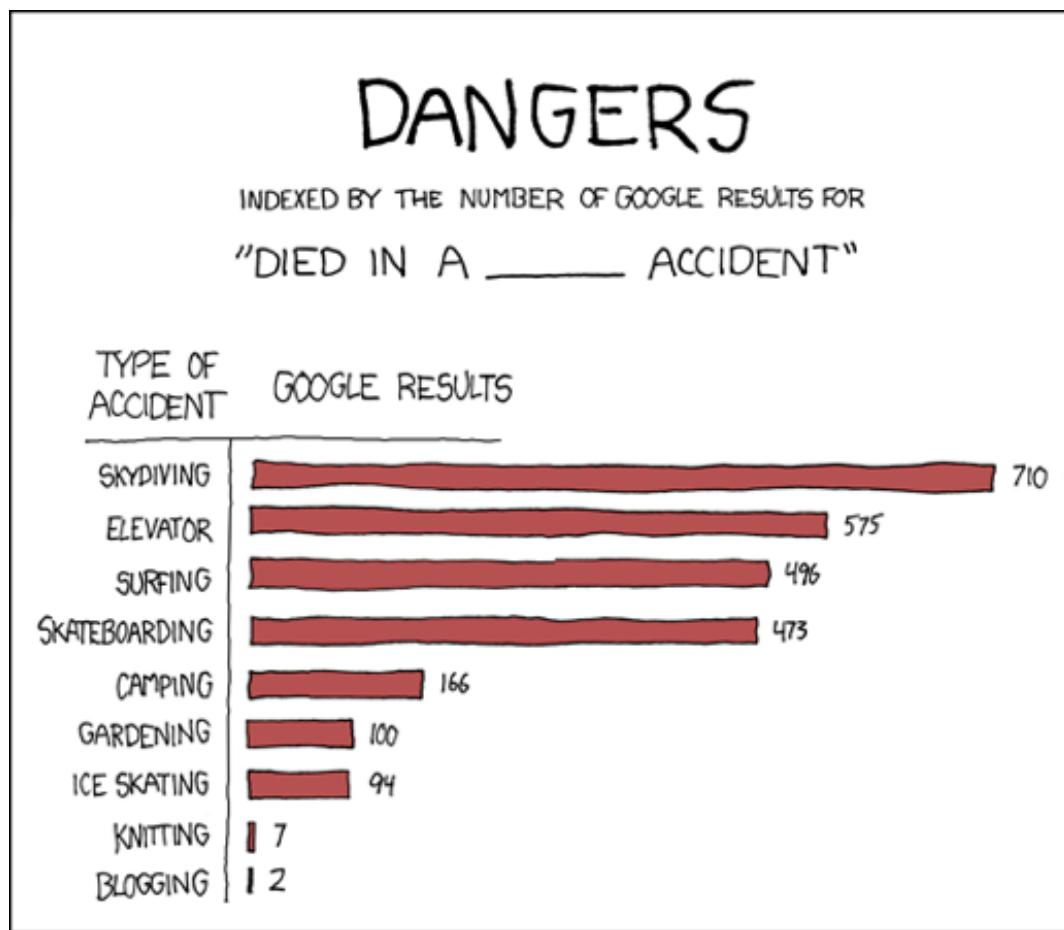
<http://matplotlib.org/xkcd>

Chart Suggestions—A Thought-Starter



Comparisons

Bar Chart

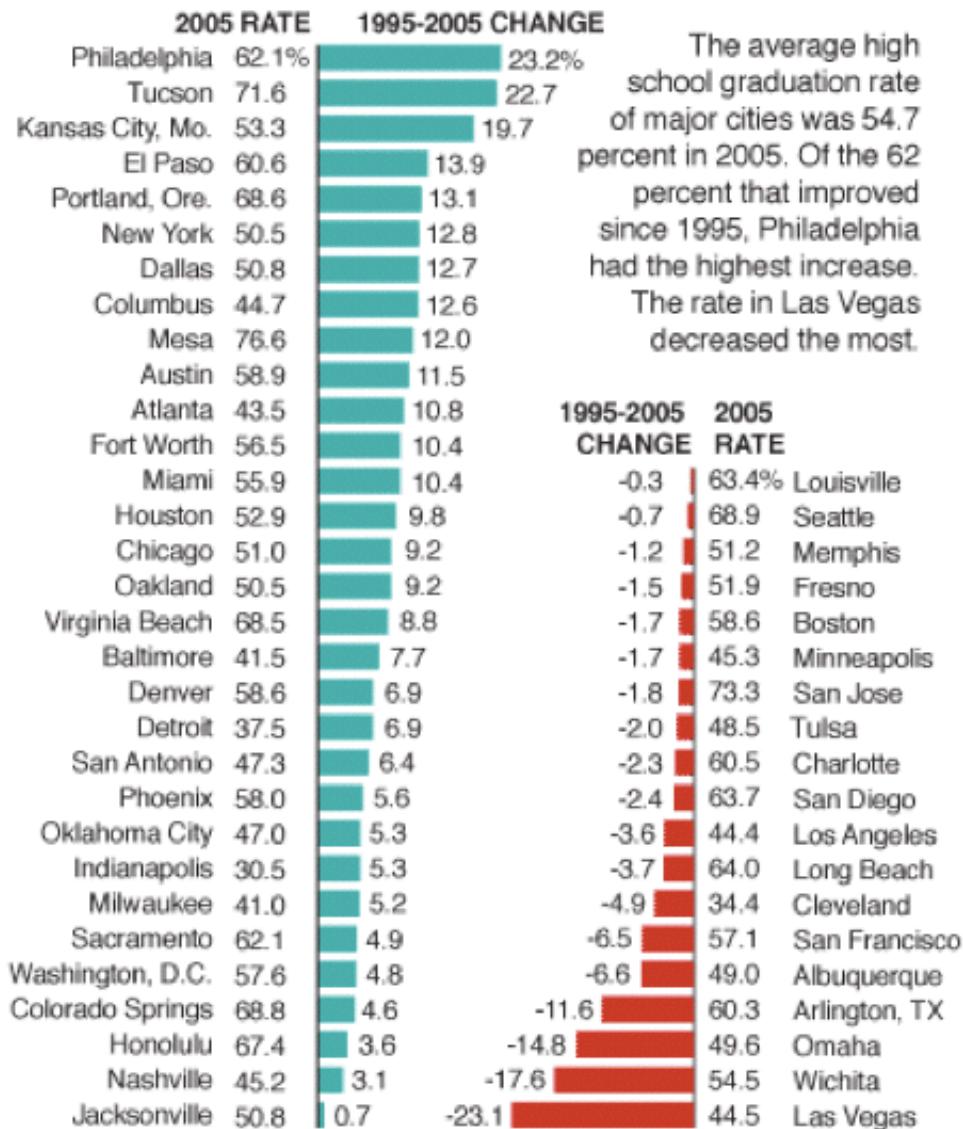


<http://nbviewer.ipython.org/gist/olgabot/5357268>

Direction

Graduation rates up in most cities

Graduation rate for principal school district of the largest cities



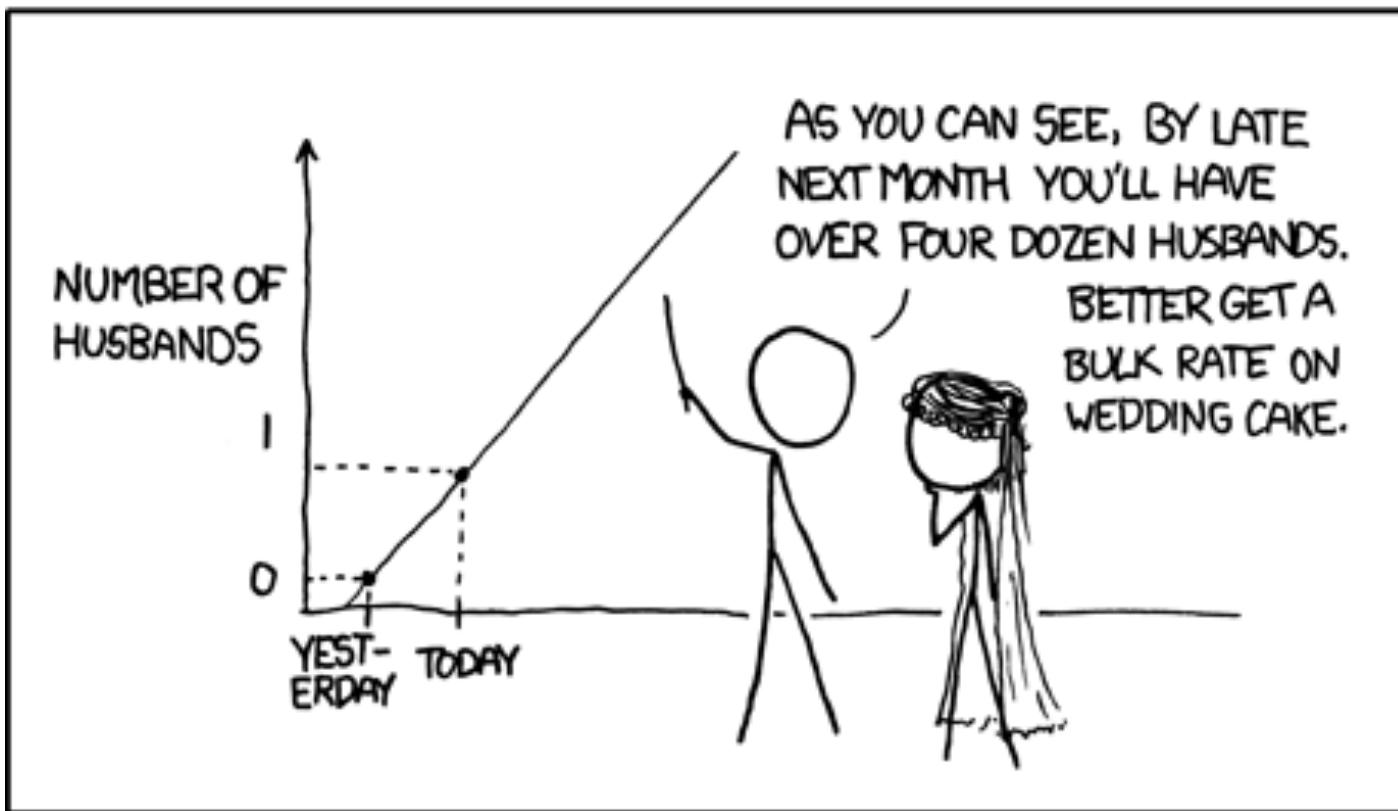
SOURCE: EPE Research Center

AP

Nicolas Rapp

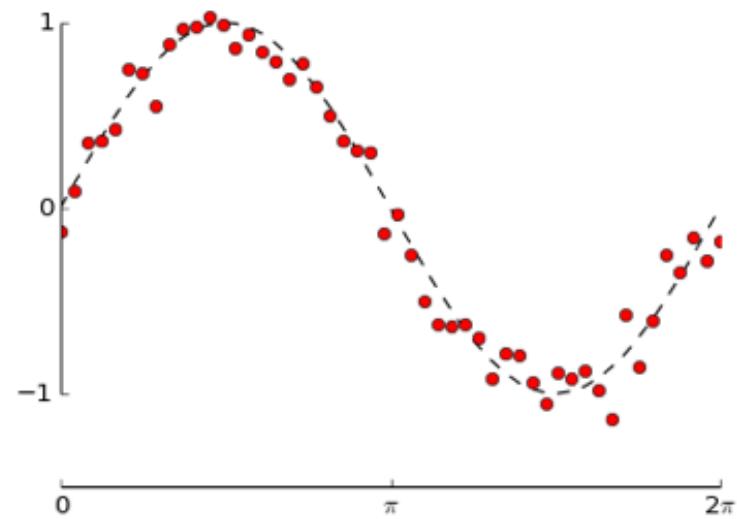
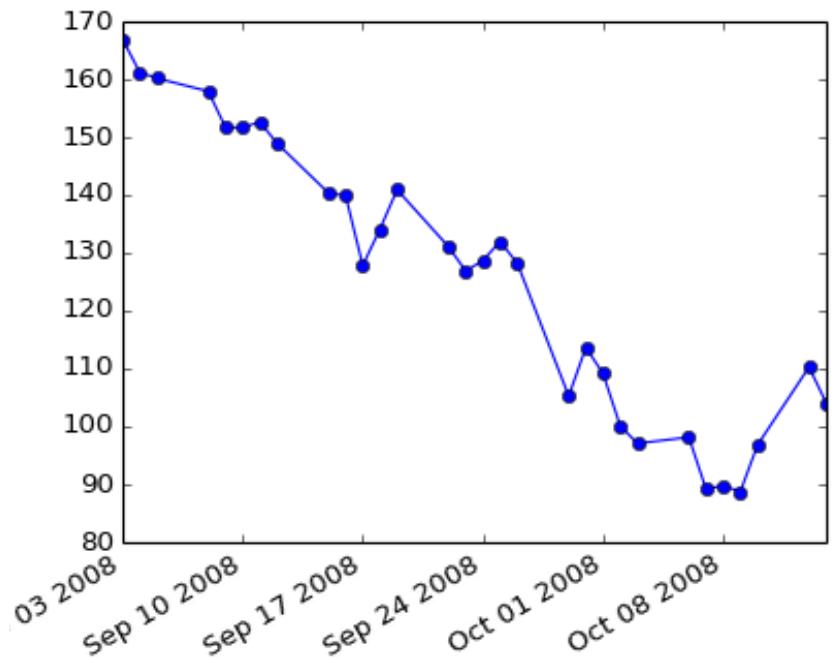
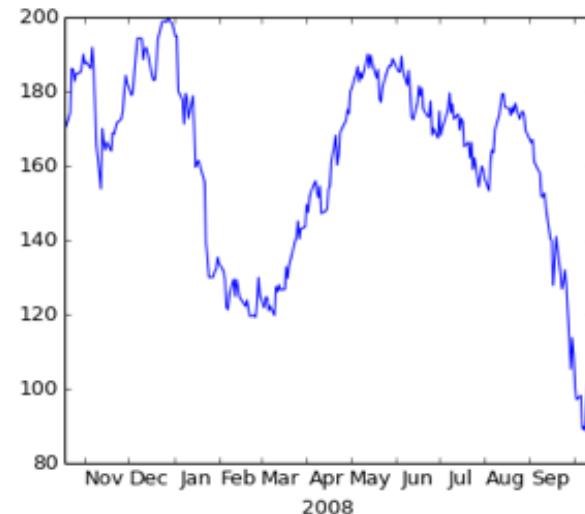
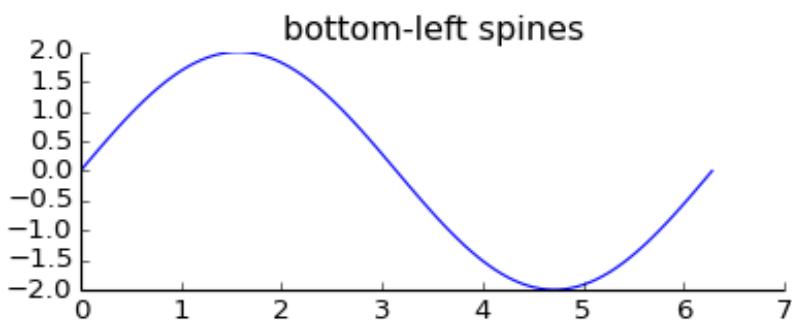
Trends Over Time

MY HOBBY: EXTRAPOLATING



<http://xkcd.com/605/>

Line Charts



Linear vs. Logarithmic Scale

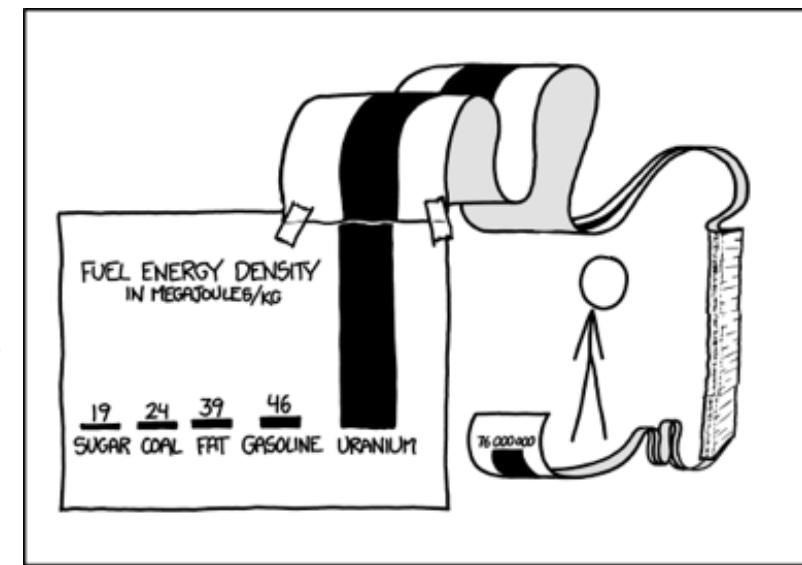
May 1990: AAPL 1.4732

Linear Scale

May 1990: AAPL 1.4732

Log Scale

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

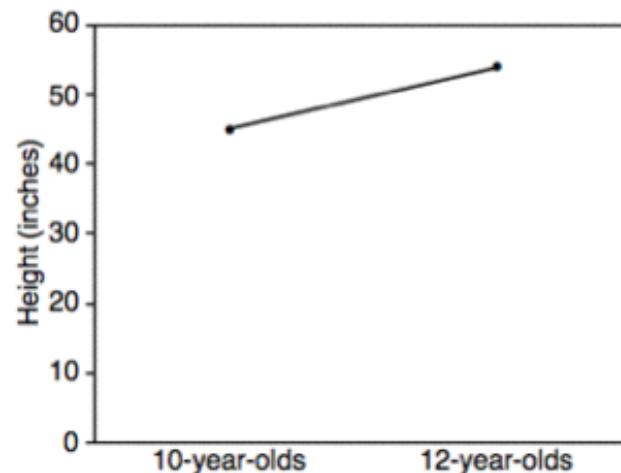
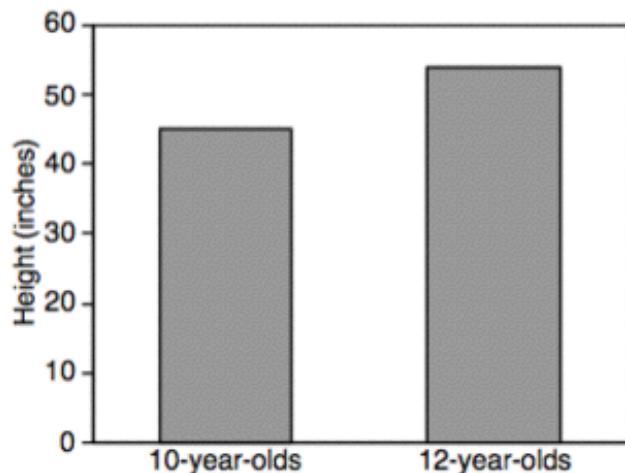
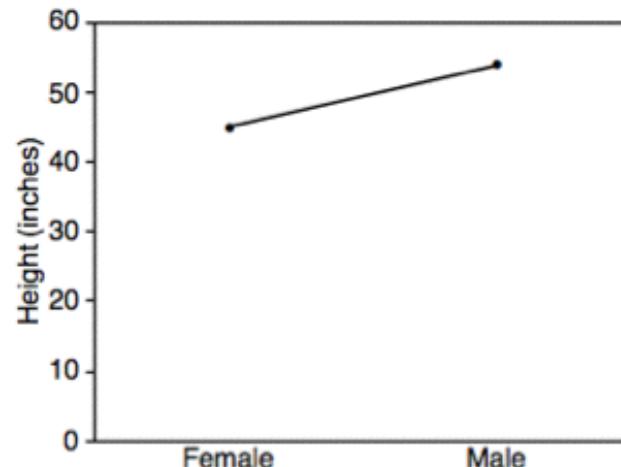
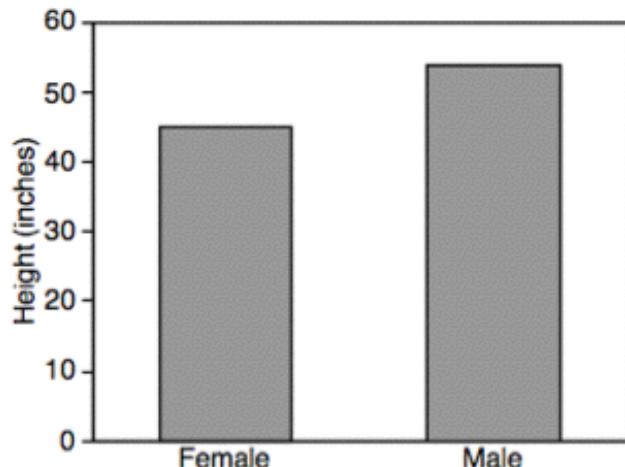


SCIENCE TIP: LOG SCALES ARE FOR QUITTERS WHO CAN'T FIND ENOUGH PAPER TO MAKE THEIR POINT PROPERLY.

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

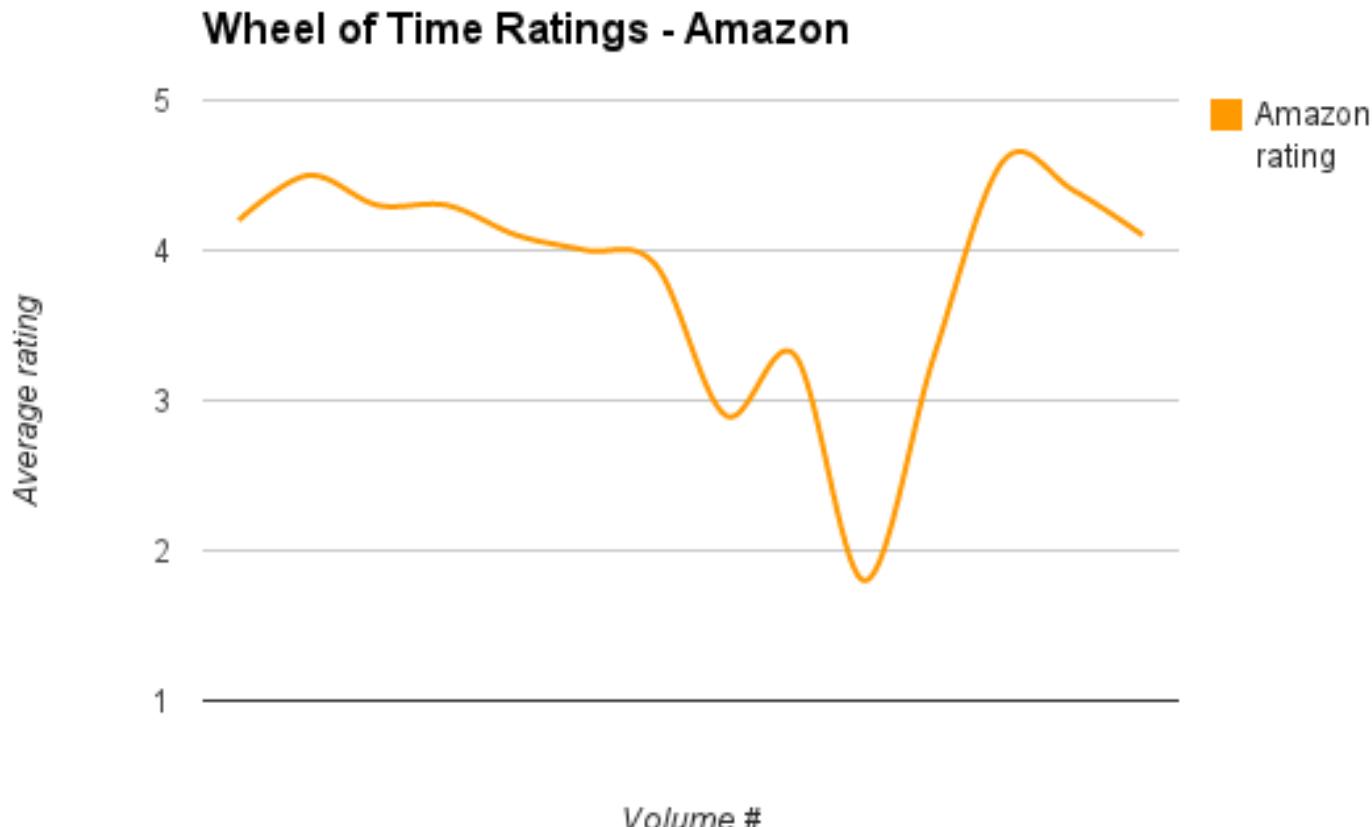
Bars vs. Lines

Lines imply connections - do not use for categorical data



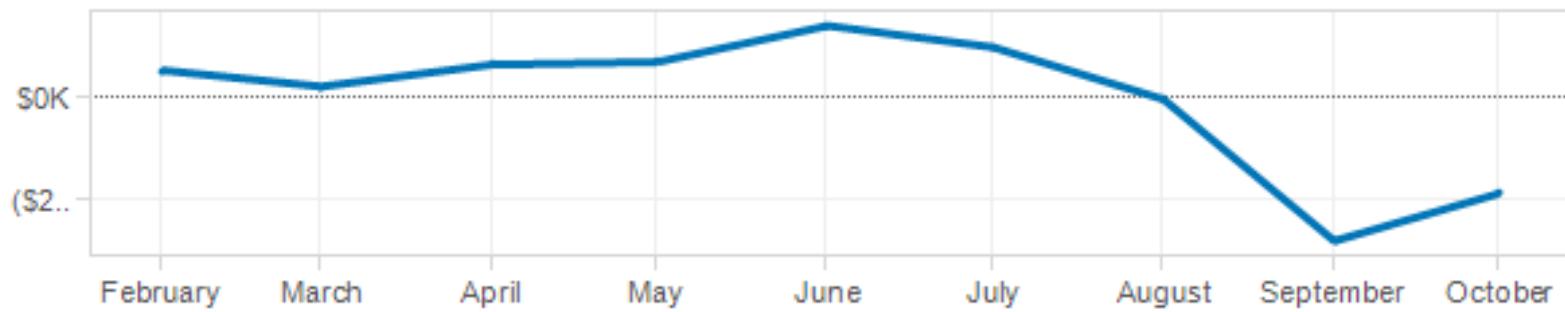
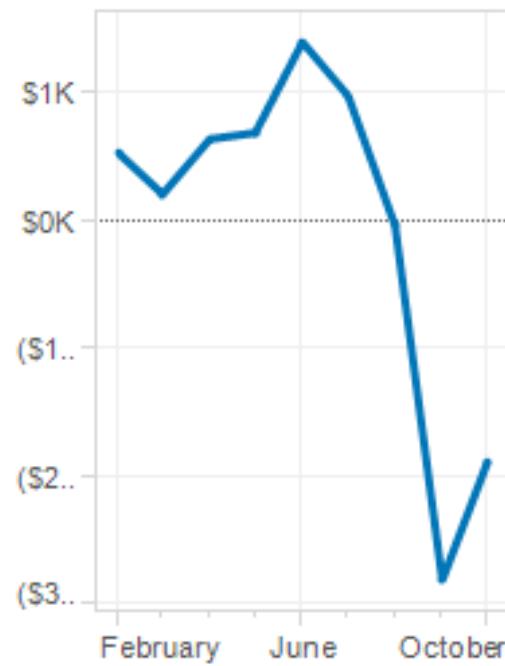
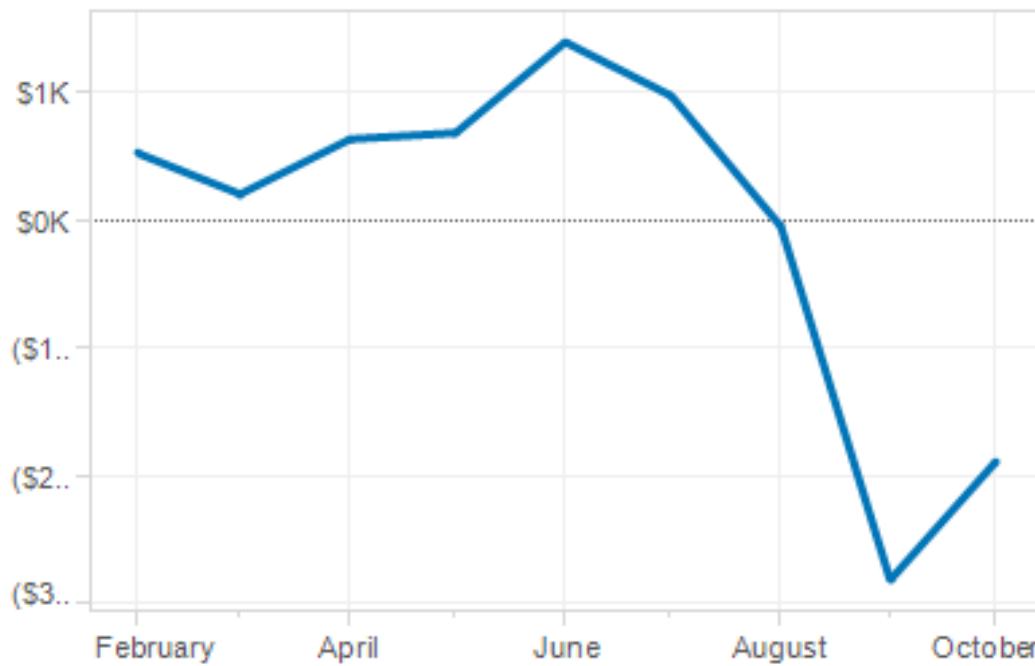
Don't

Use bar charts to compare book ratings



“Visualizing The Wheel of Time: Reader Sentiment for an Epic Fantasy Series”, J. Siddle, Sept 2013

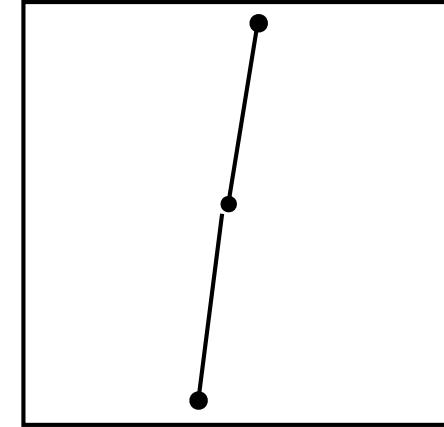
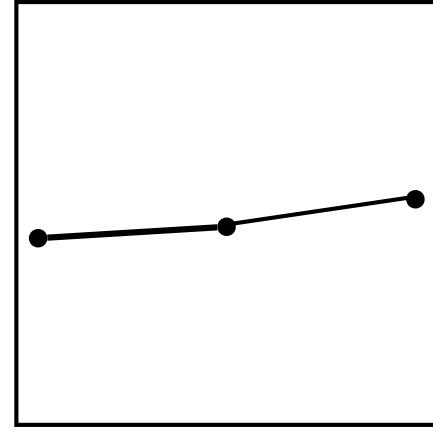
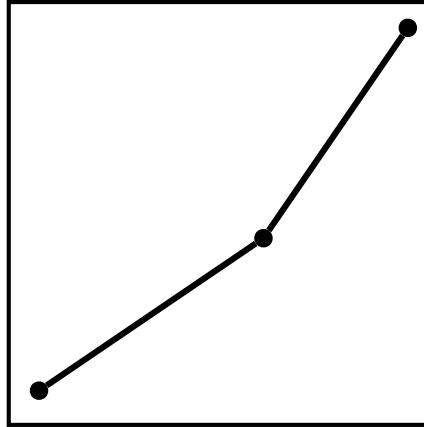
Aspect Ratios



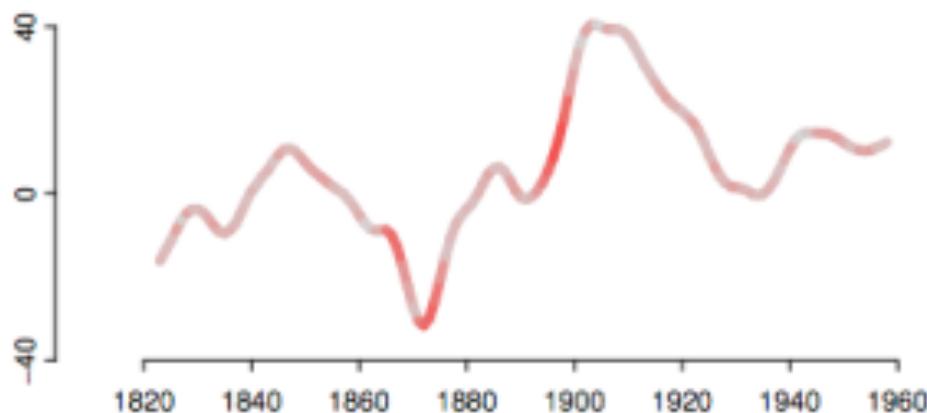
Banking to 45°

Two line segments are maximally discriminable when
their average absolute angle is 45°

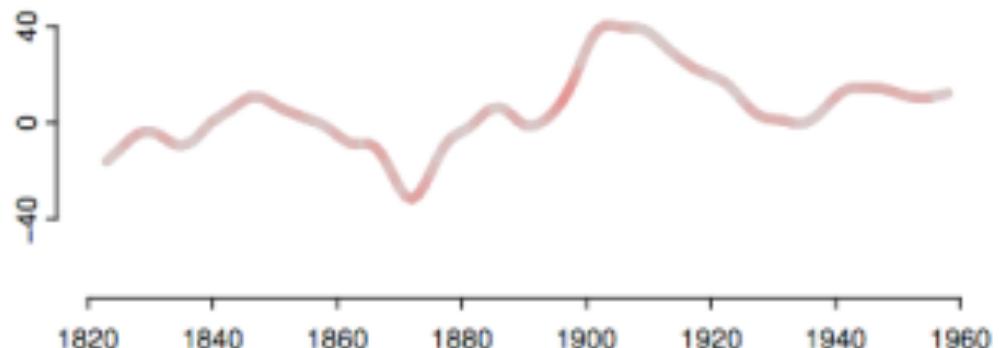
W. Cleveland



Banking to 45°

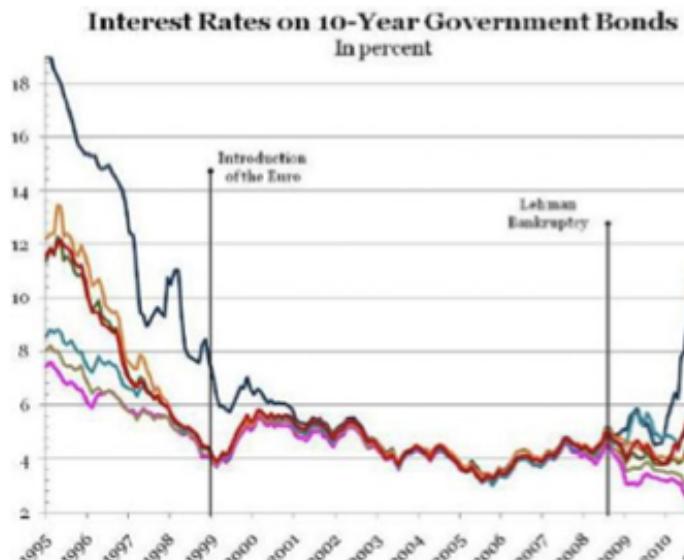


Error Prone

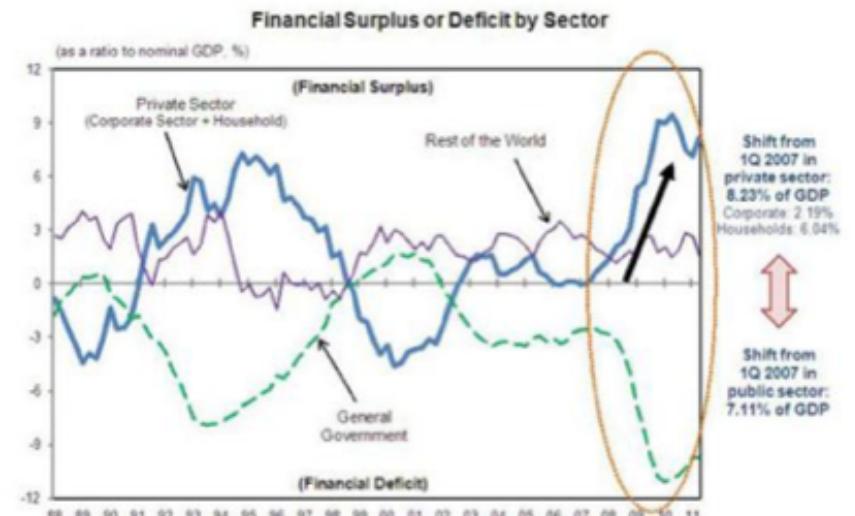


Optimal Aspect Ratio

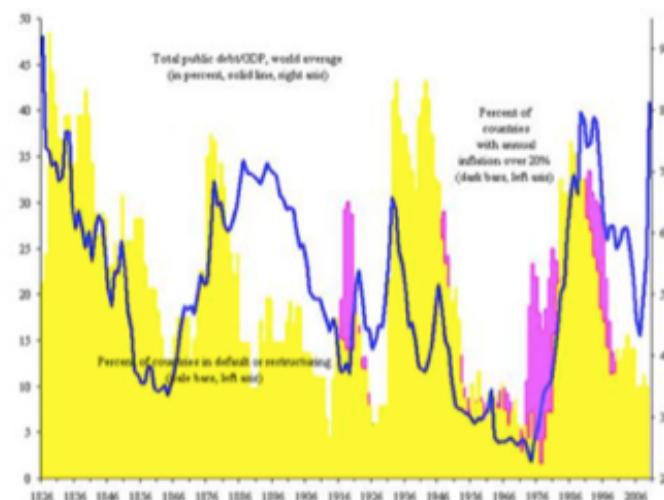
Don't



UK in Balance Sheet Recession: UK Private Sector Increased Savings Massively after the Bubble

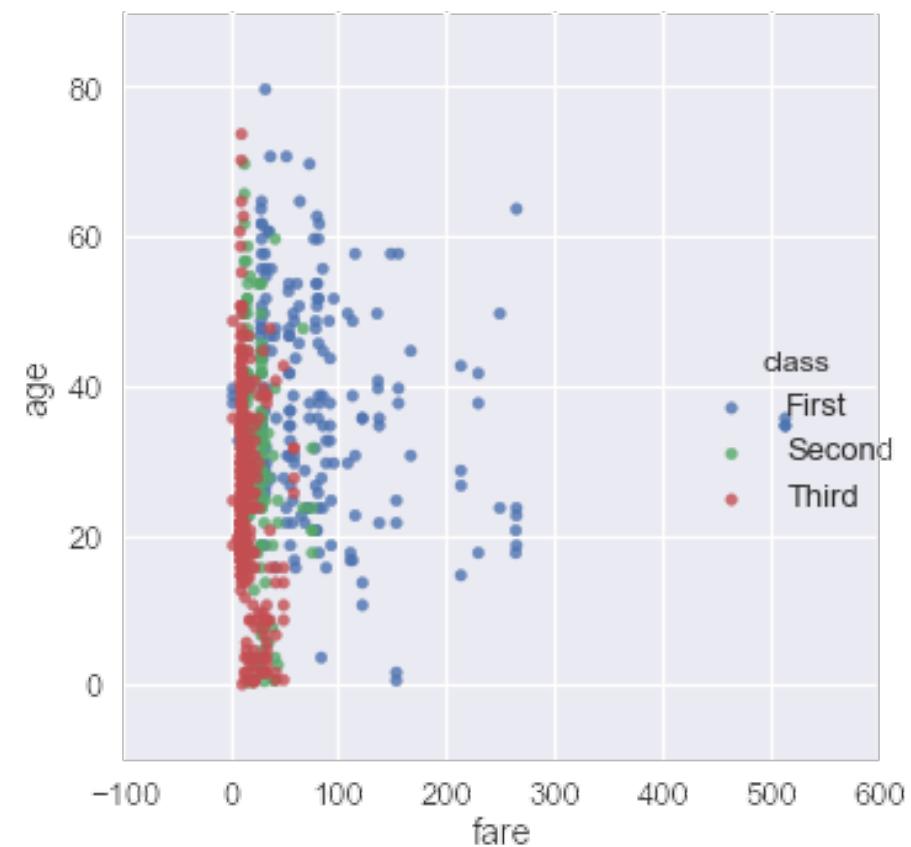
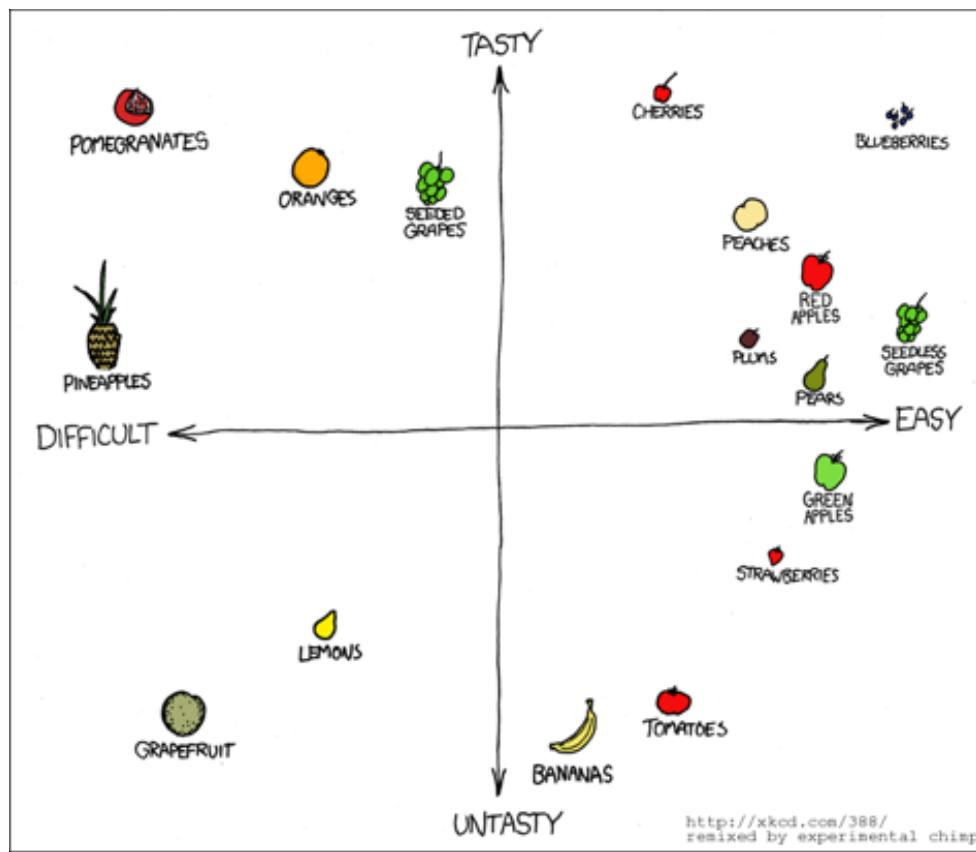


Note: For the latest figures, 4 quarter averages ending with 2Q/11 are used.
Source: Office for National Statistics, UK

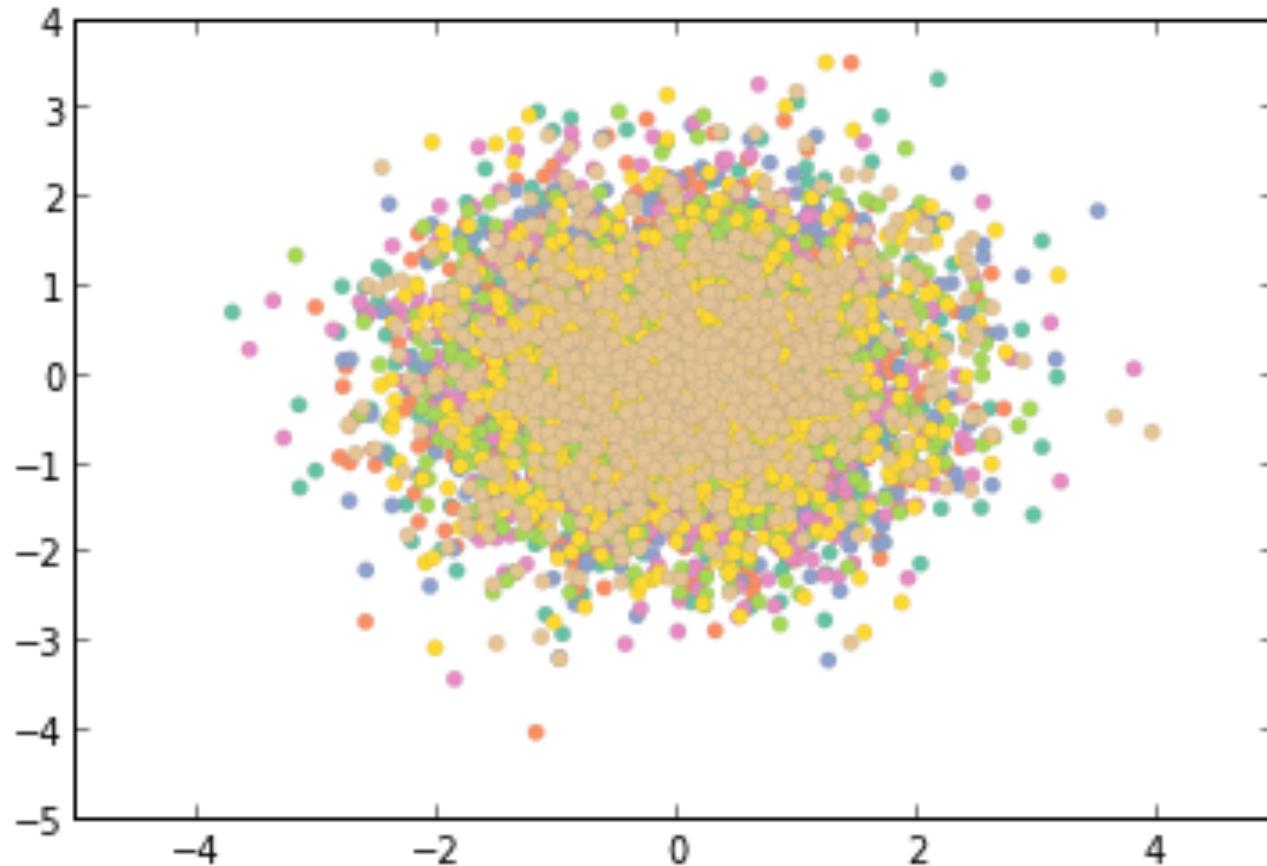


Correlations

Scatterplots

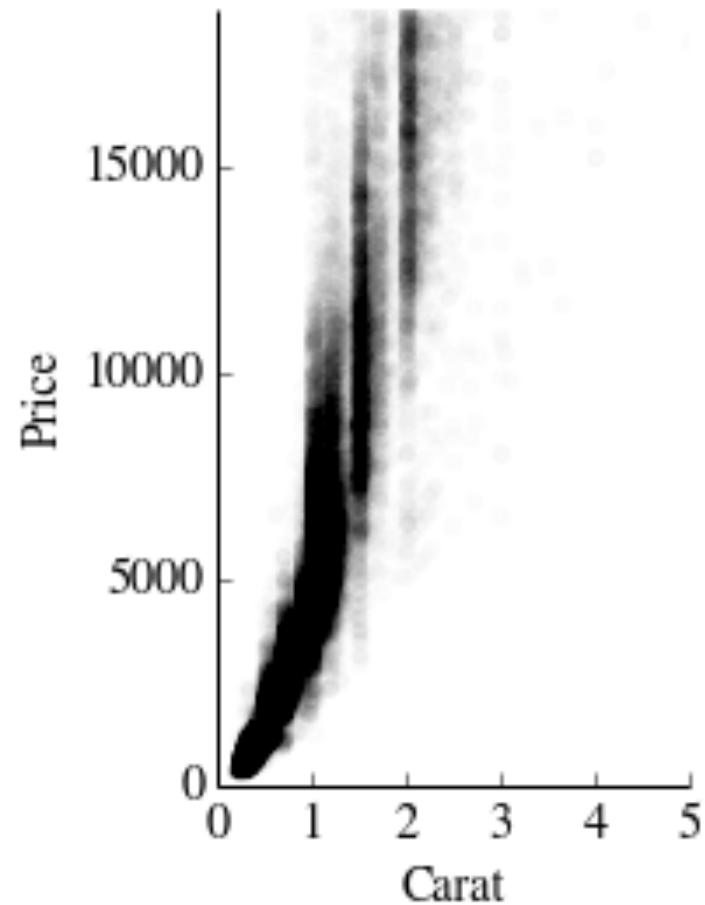
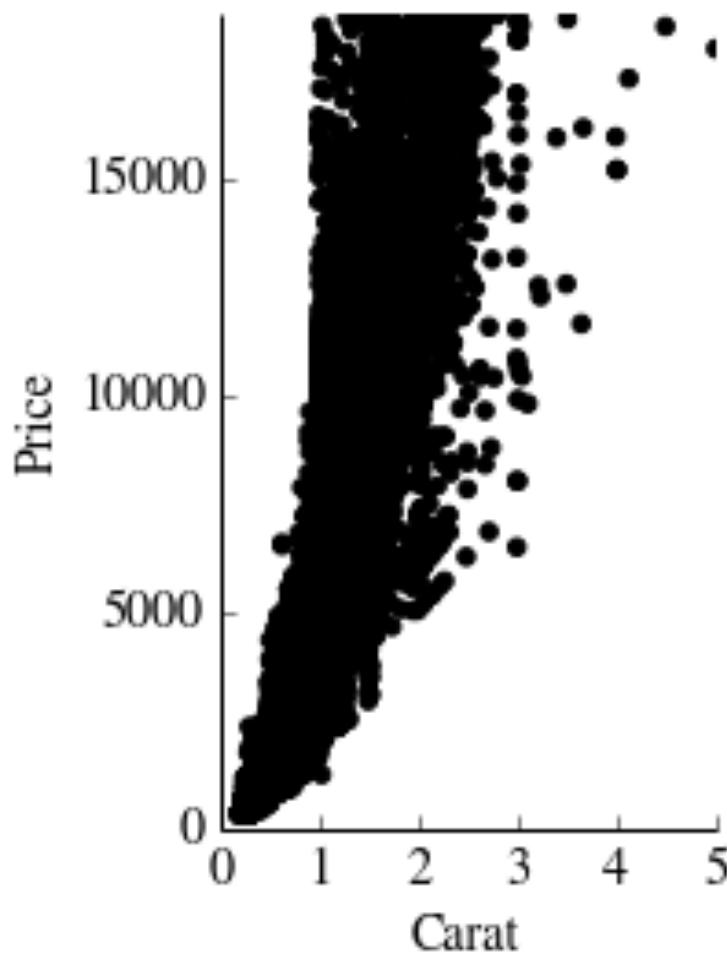


Scatterplots



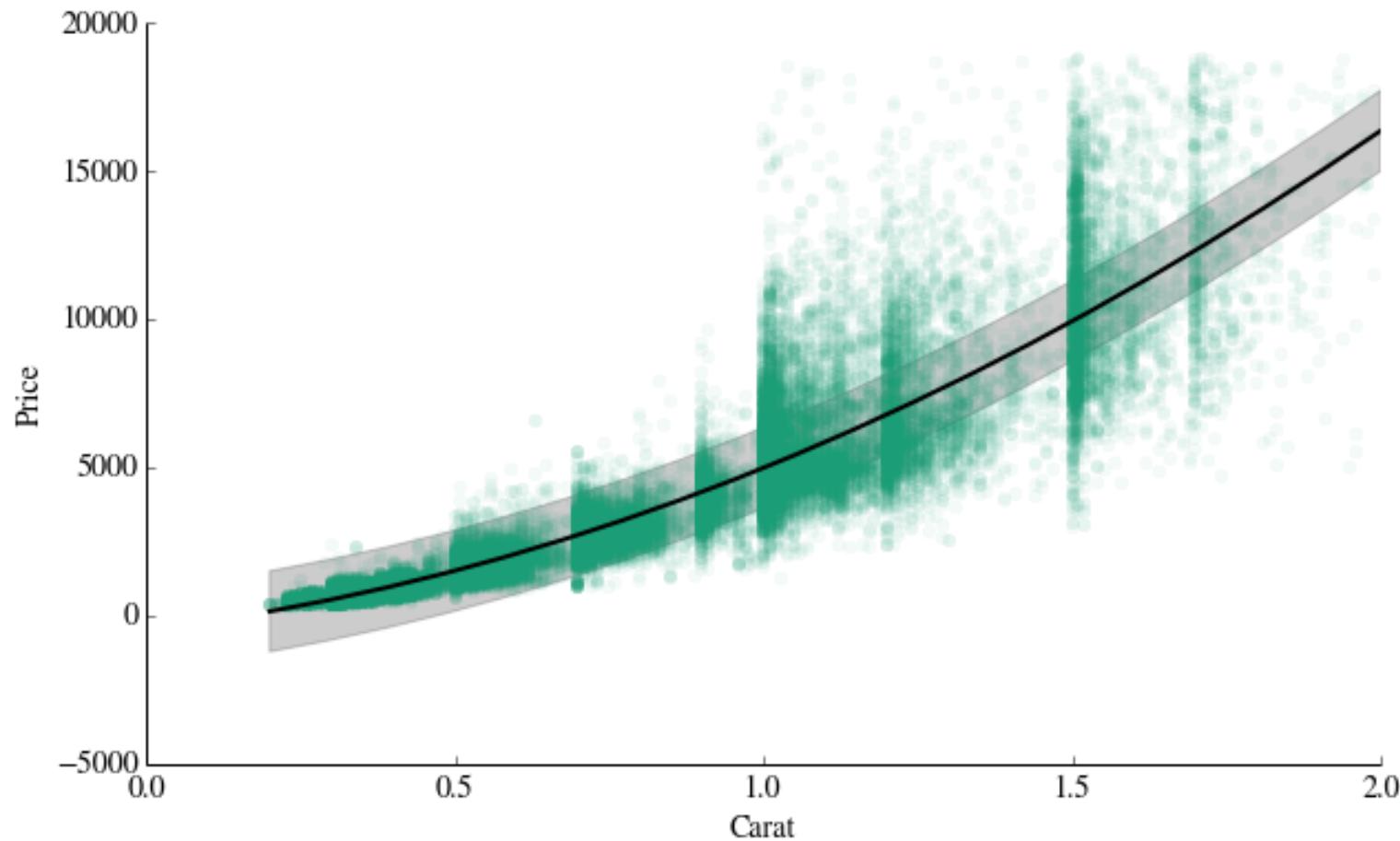
Light Grey Border

Overplotting

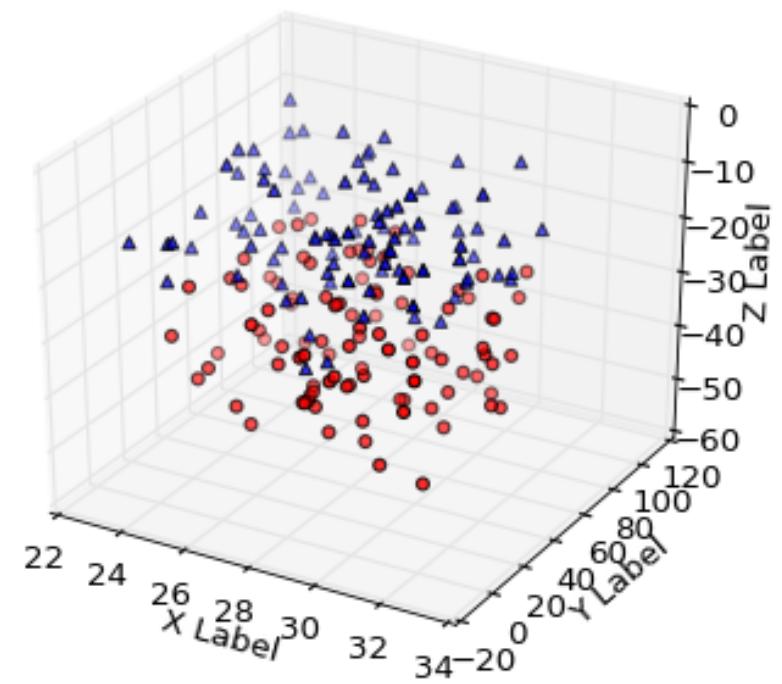
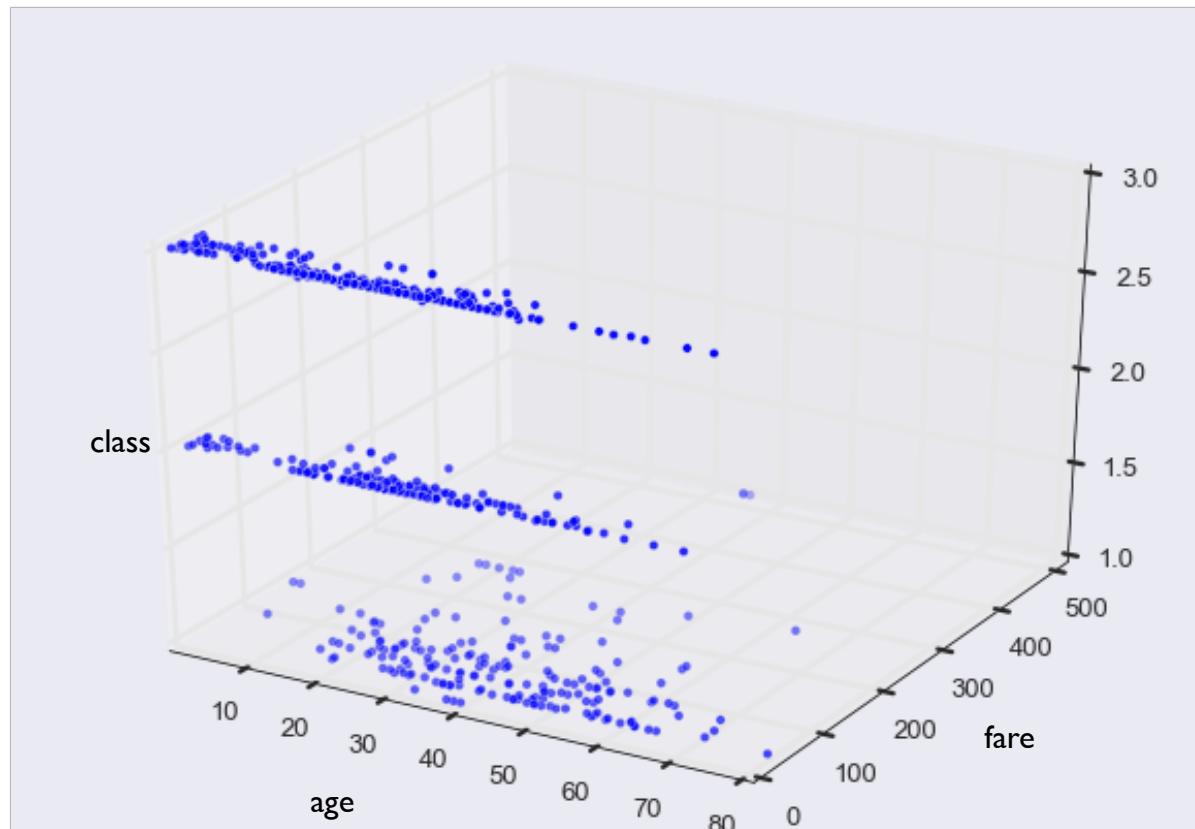


$\text{alpha} = 1/100$

Trend Lines

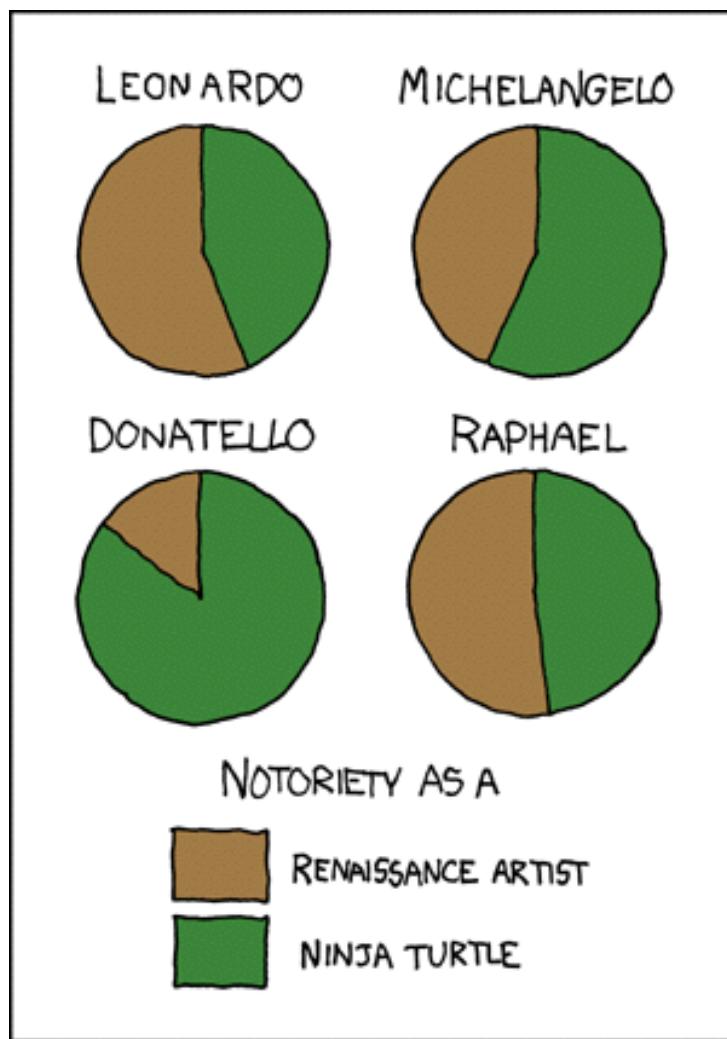


Don't

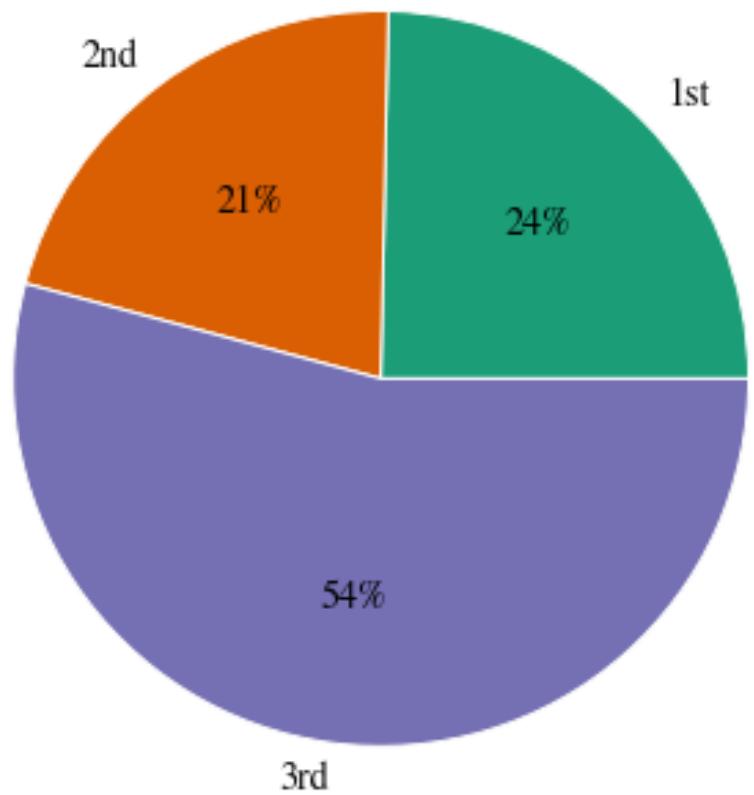


Compositions

Pie Charts

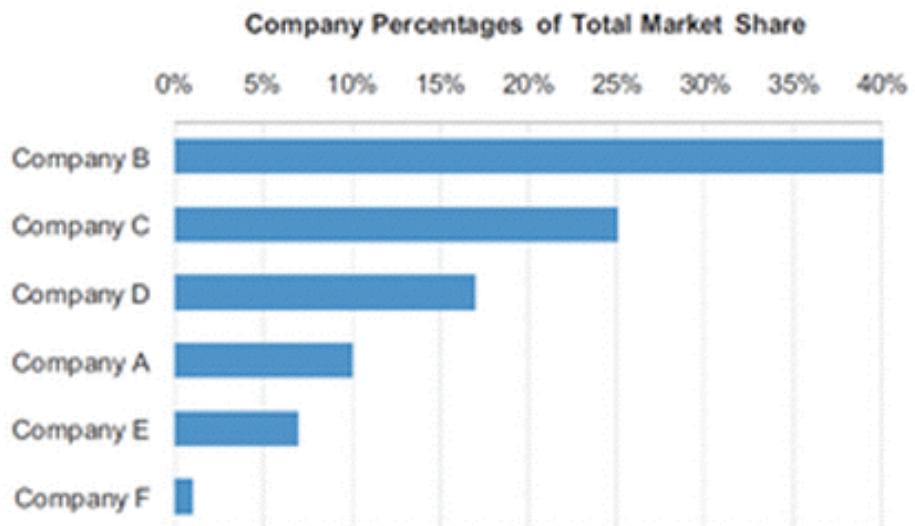
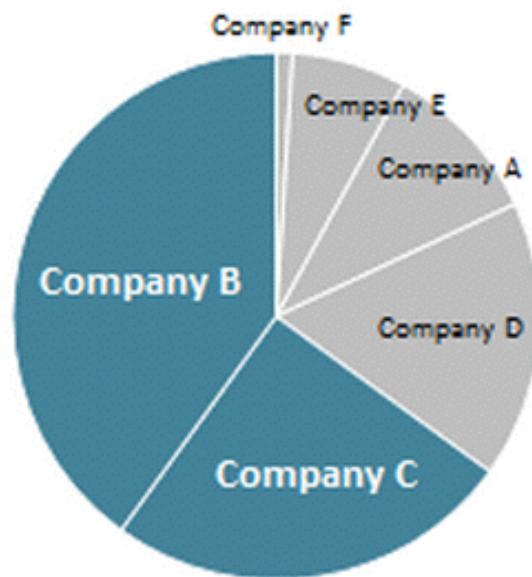


Passenger Class on the Titanic



Pie vs. Bar Charts

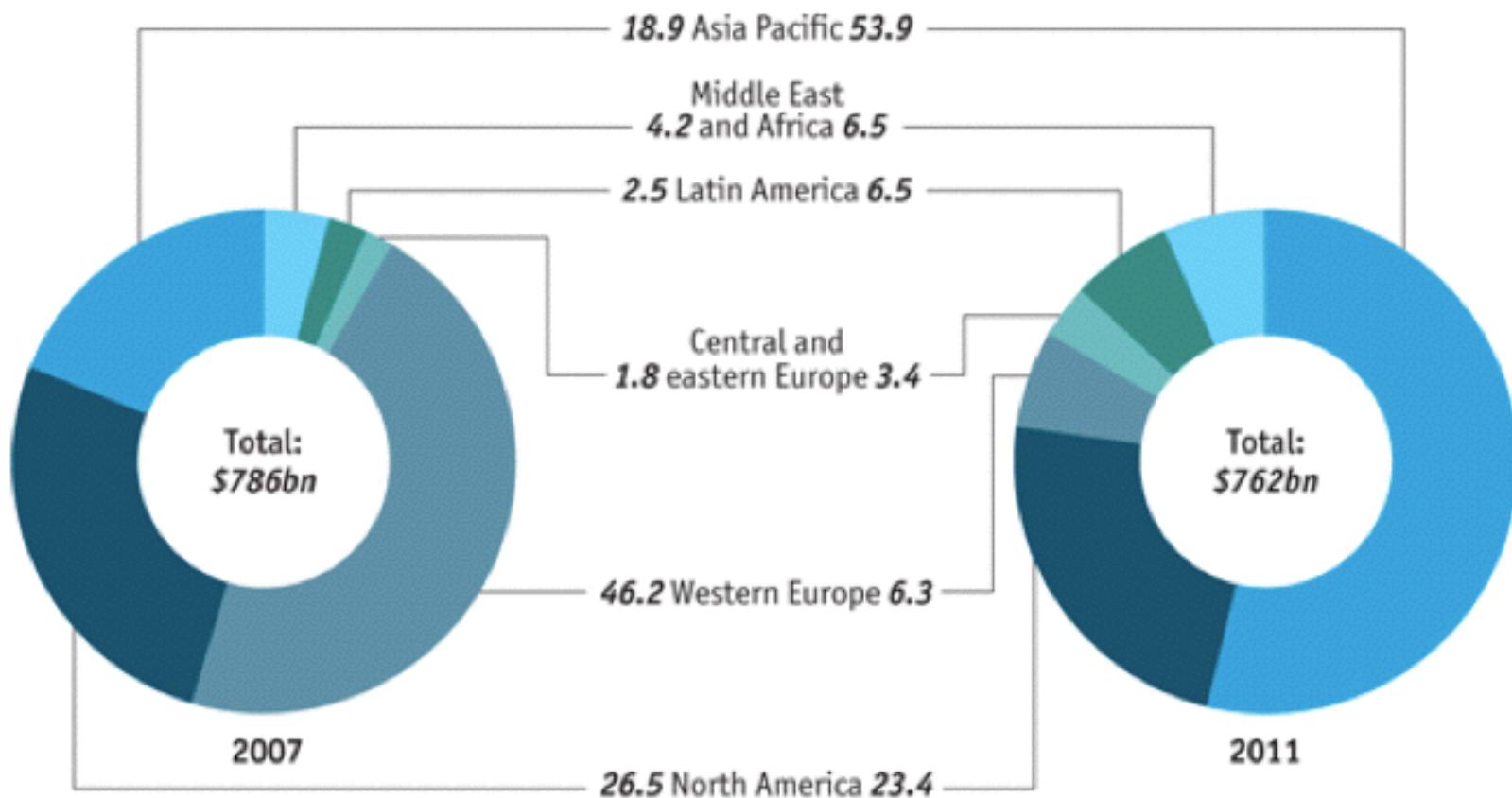
65% of the market is controlled by companies B and C



Donut Chart

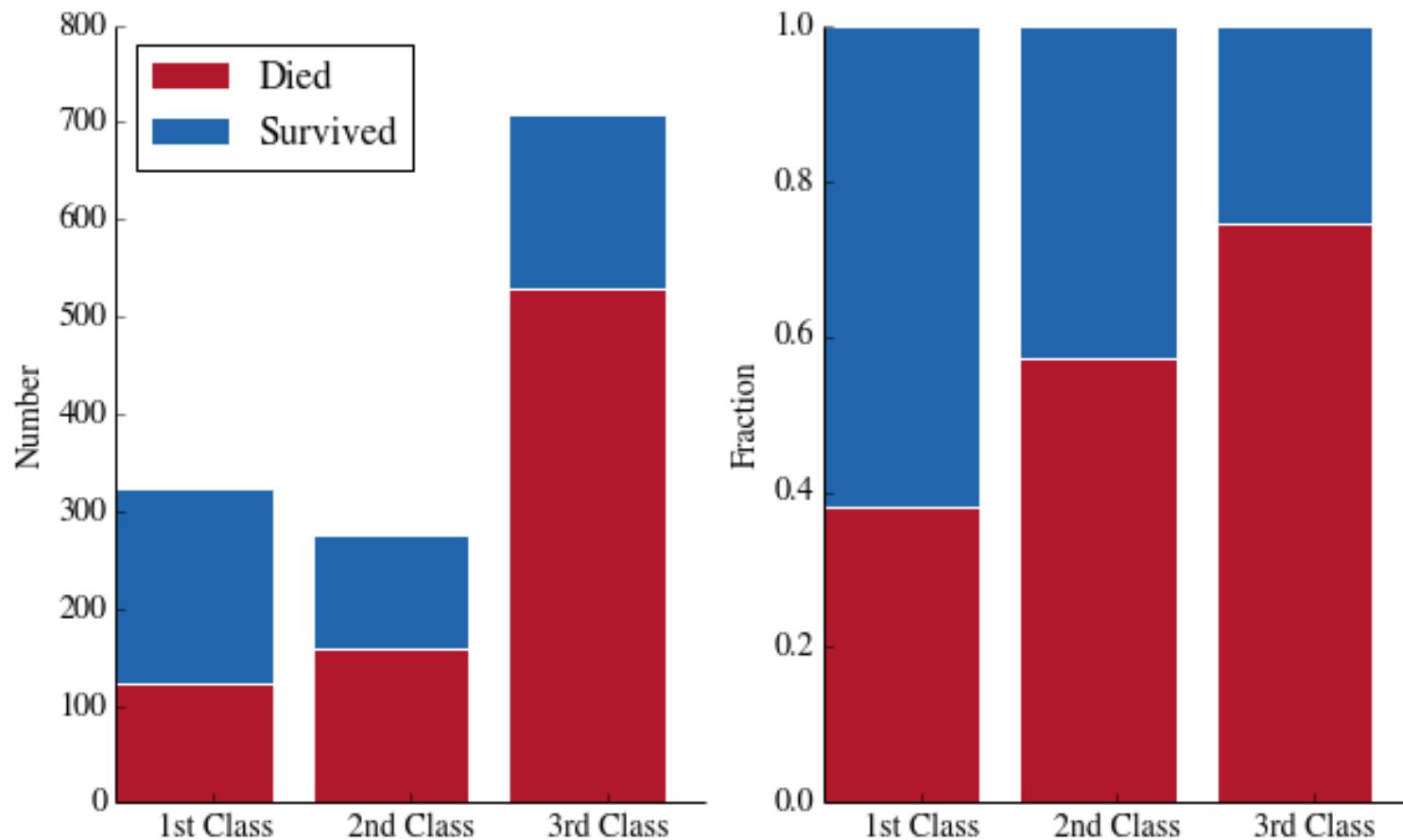
Pre-tax profits of the 1,000 largest banks

By tier-one capital and domicile, % of total

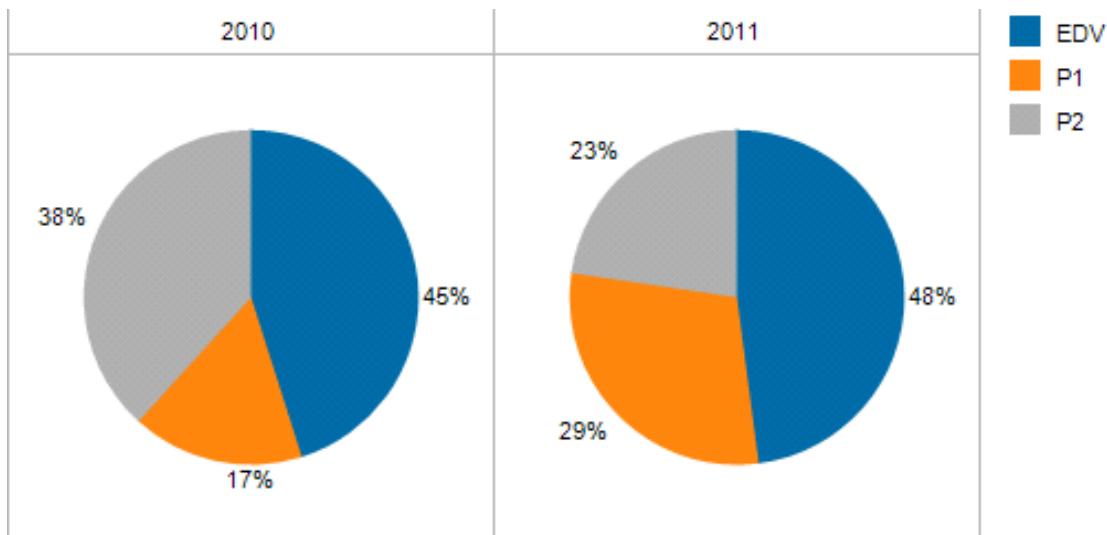


Source: *The Banker Top 1000*

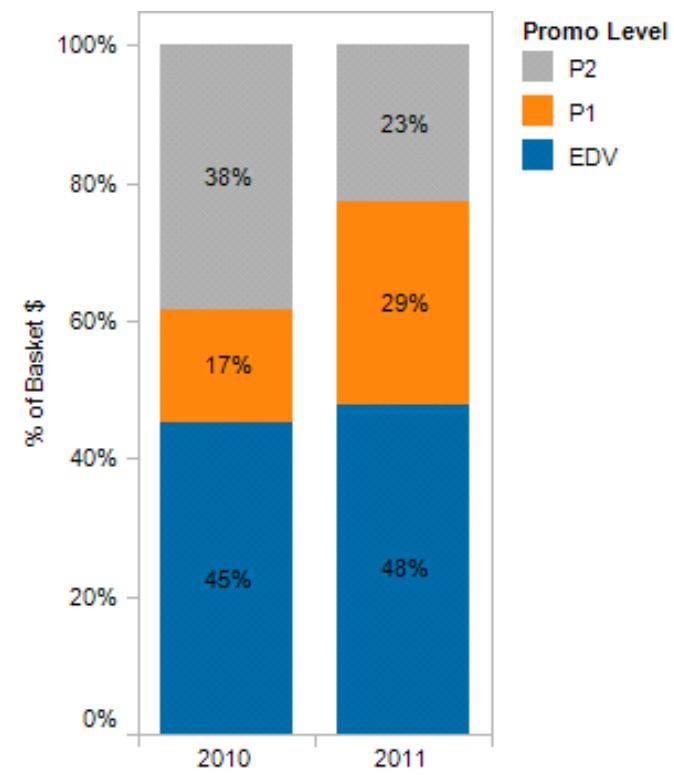
Stacked Bar Chart



Stacked Bar Chart



VS.

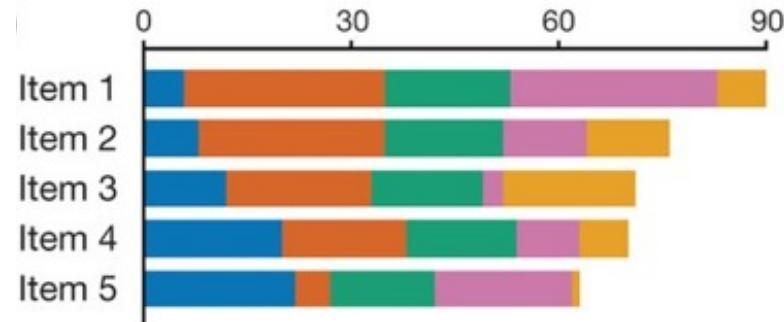


Comparison of bar chart types

- Category 1
- Category 2
- Category 3
- Category 4
- Category 5

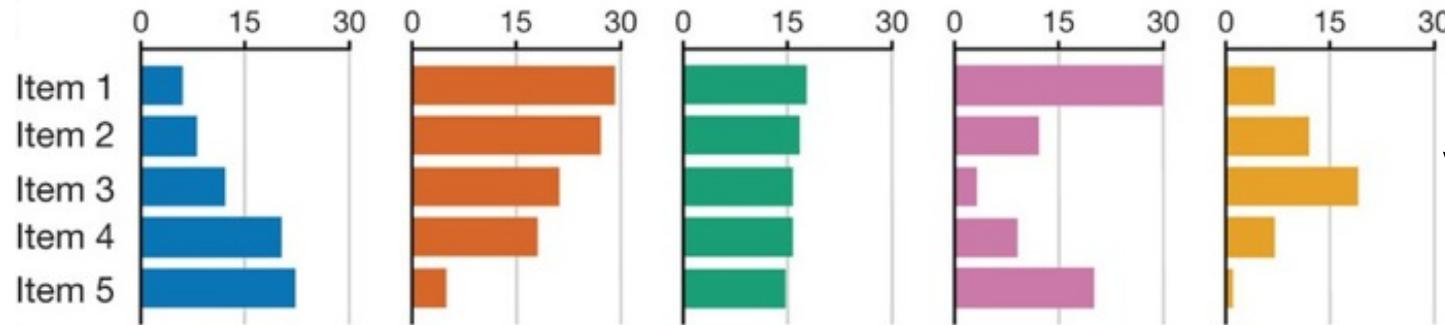


Pie Chart



Stacked bar chart

Layered
Bar
Chart



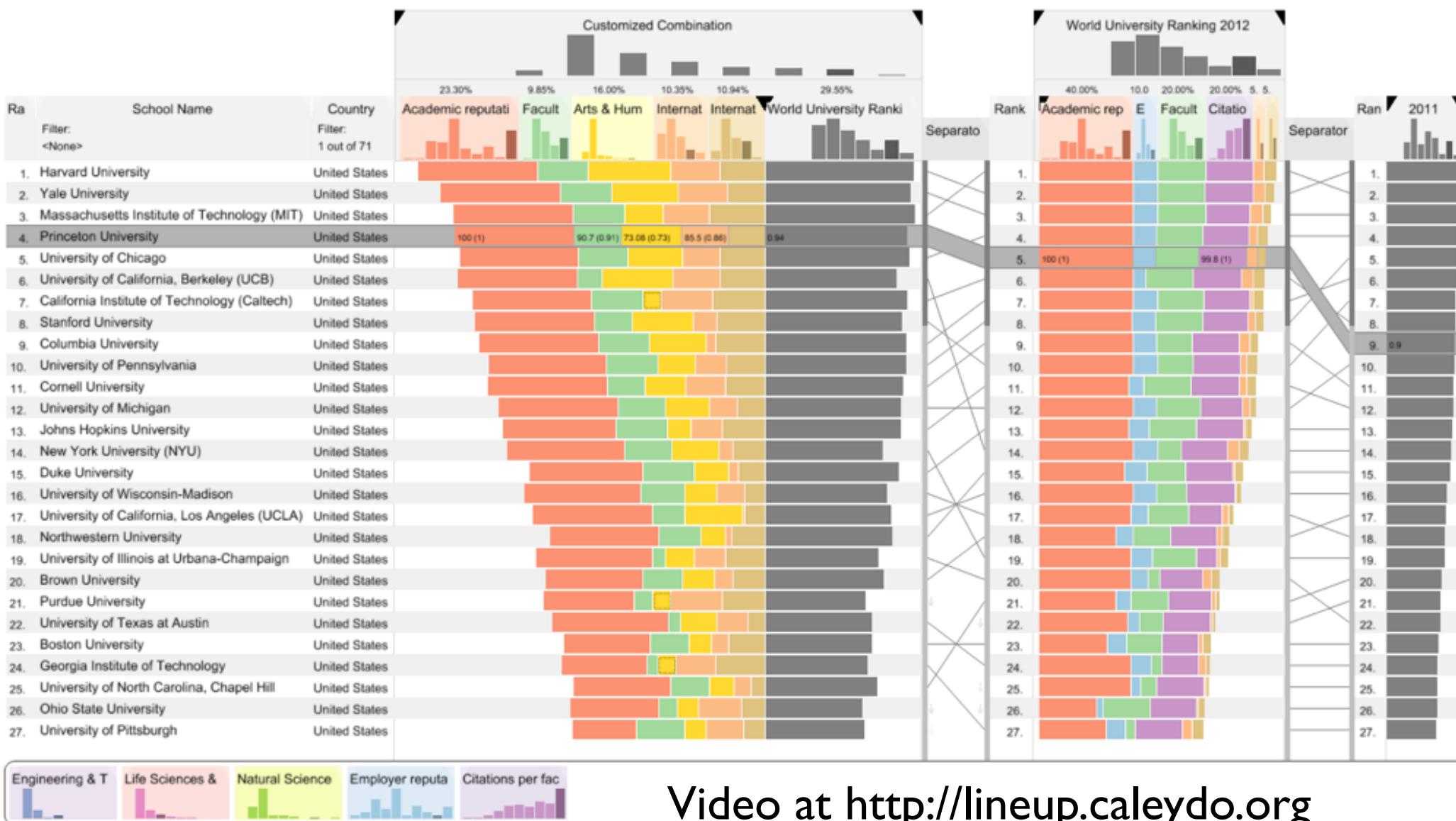
Small
Multiples

Grouped
Bar
Chart

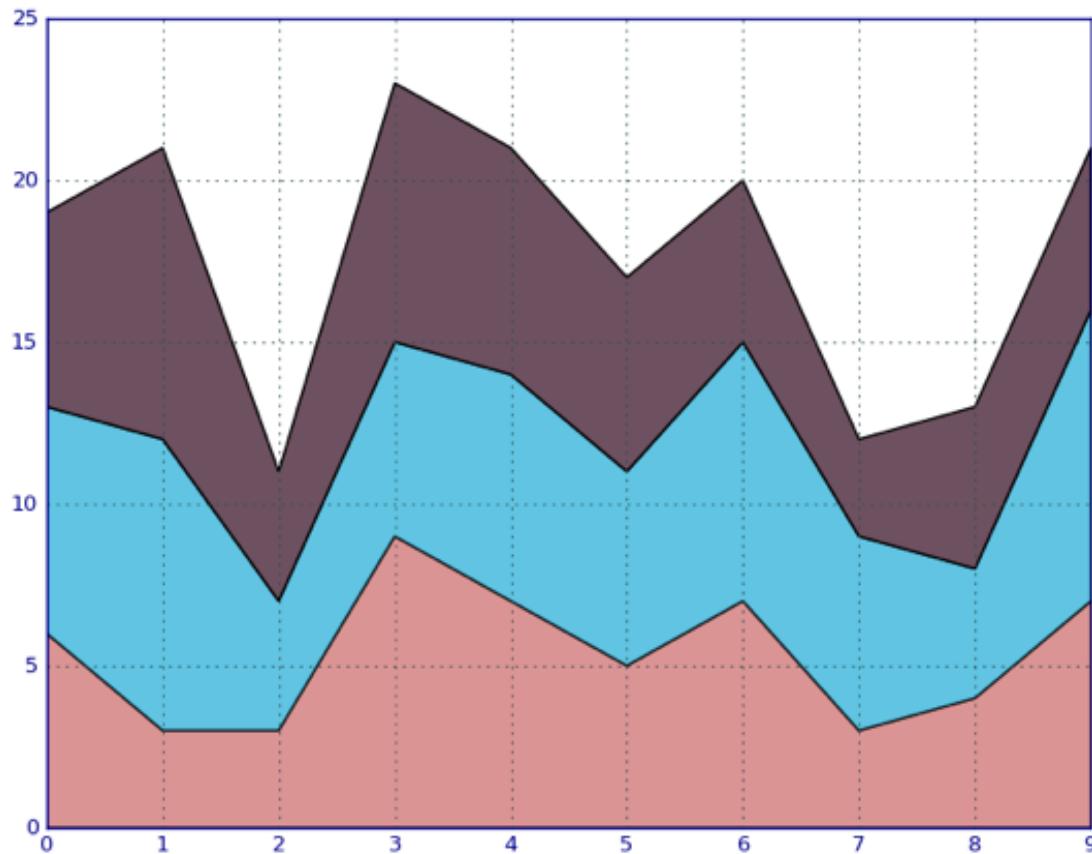


Small
Multiples

LineUp

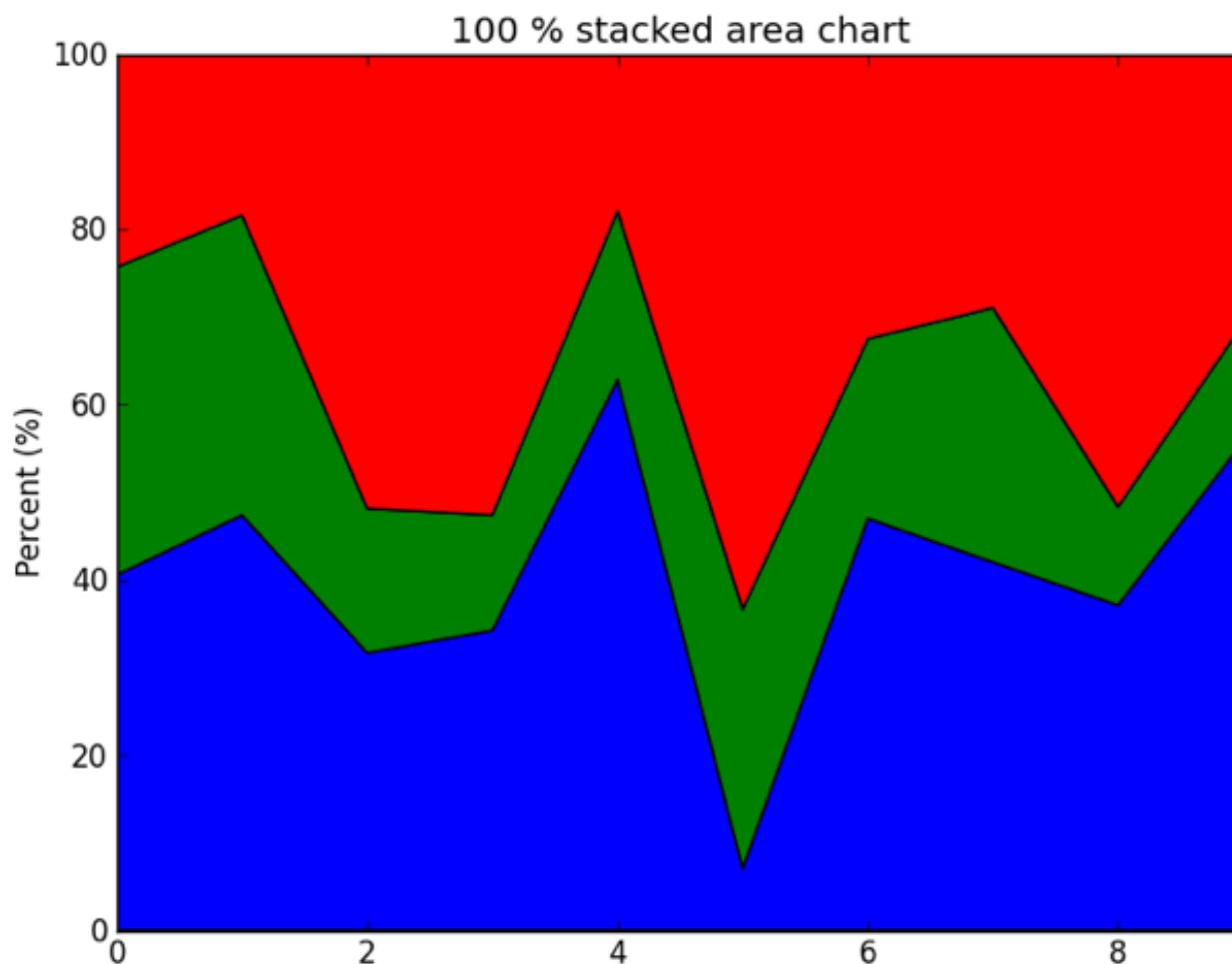


Stacked Area Chart



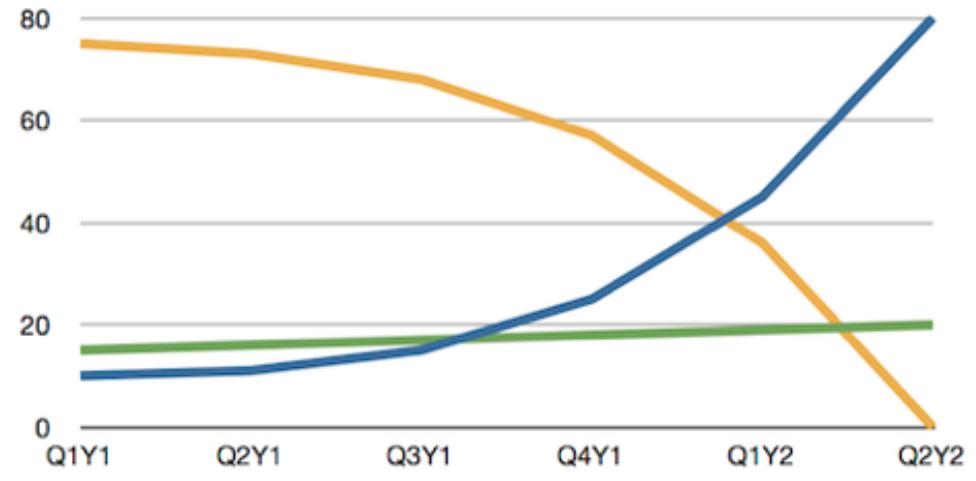
<http://stackoverflow.com/questions/2225995/how-can-i-create-stacked-line-graph-with-matplotlib>

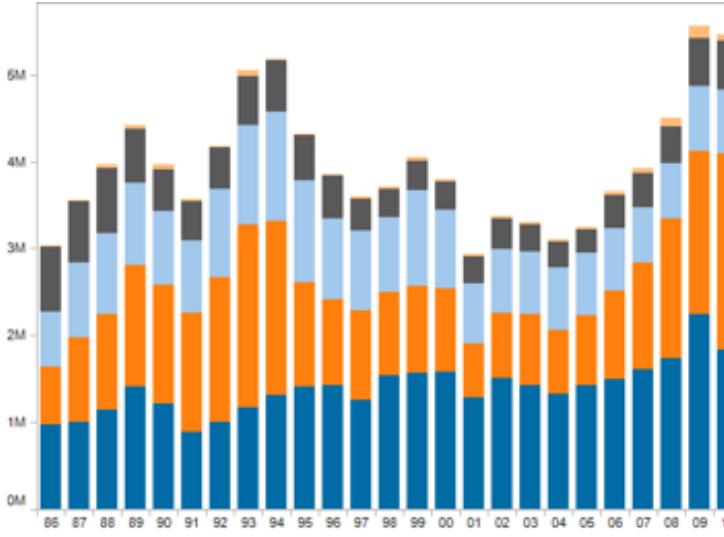
100% Stacked Area Chart



<http://stackoverflow.com/questions/16875546/create-a-100-stacked-area-chart-with-matplotlib>

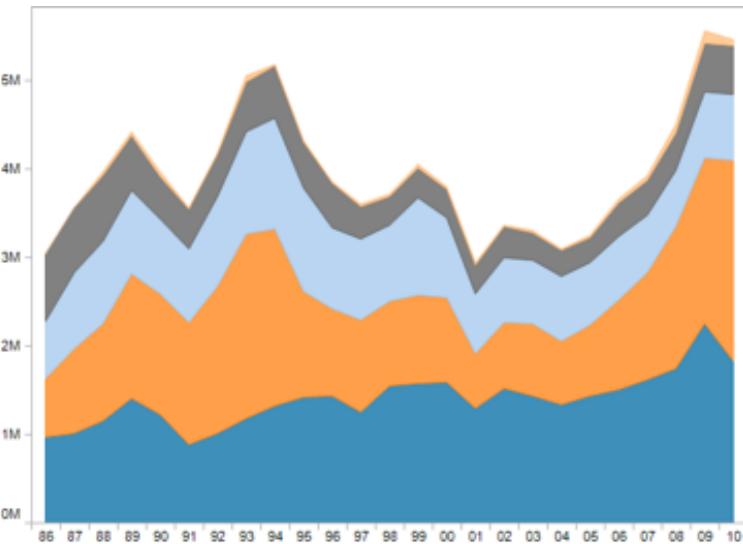
Stacked Area vs. Line Graphs





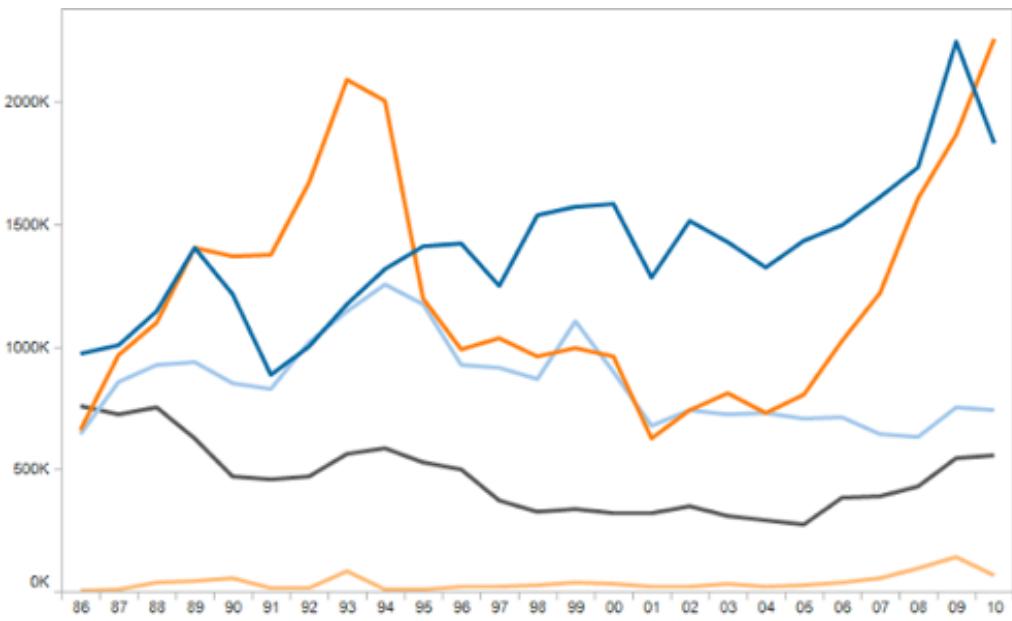
Weapon

- Misc
- Revolvers
- Shotguns
- Pistols
- Rifles



Weapon

- Misc
- Revolvers
- Shotguns
- Pistols
- Rifles

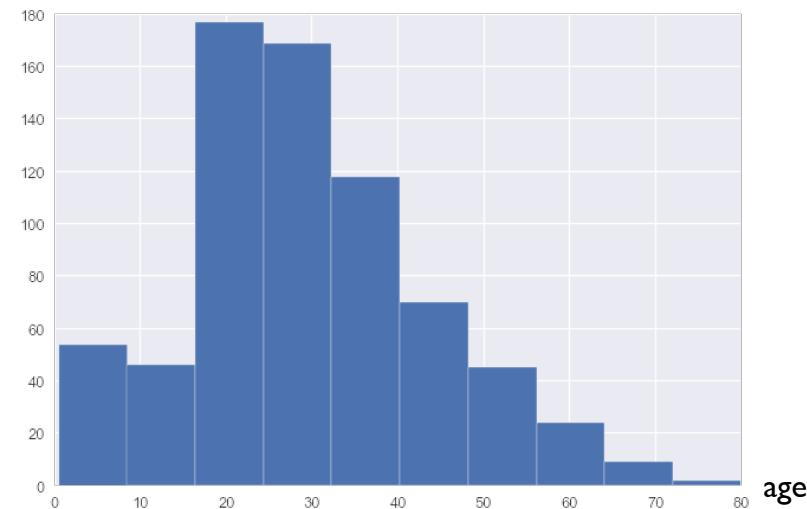


0K
500K
1000K
1500K
2000K

Distributions

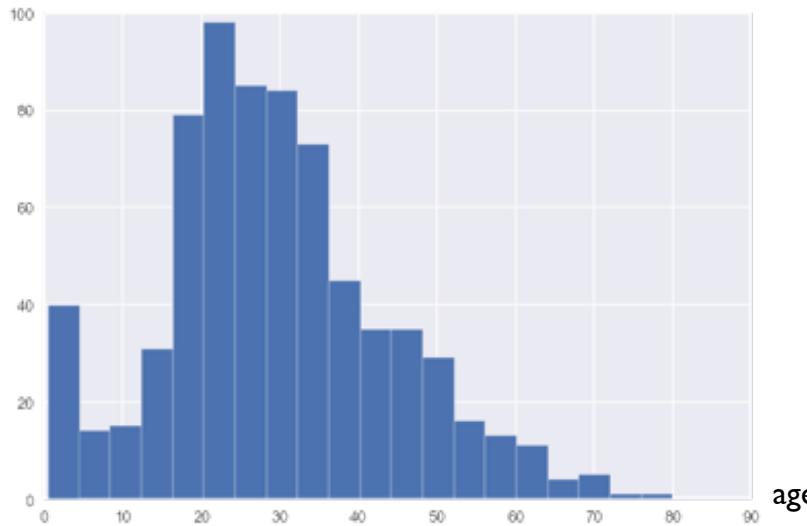
Histogram

passengers



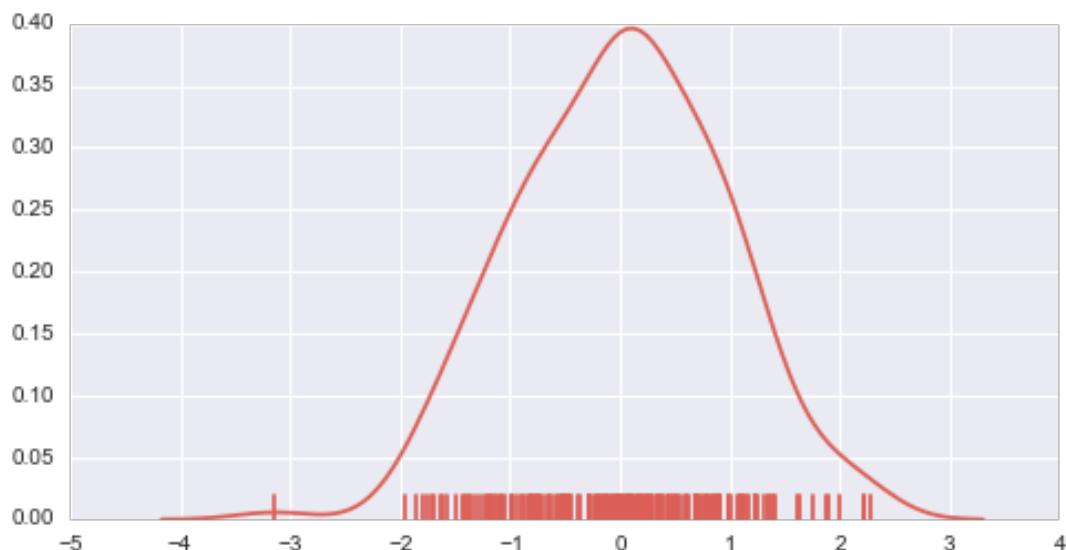
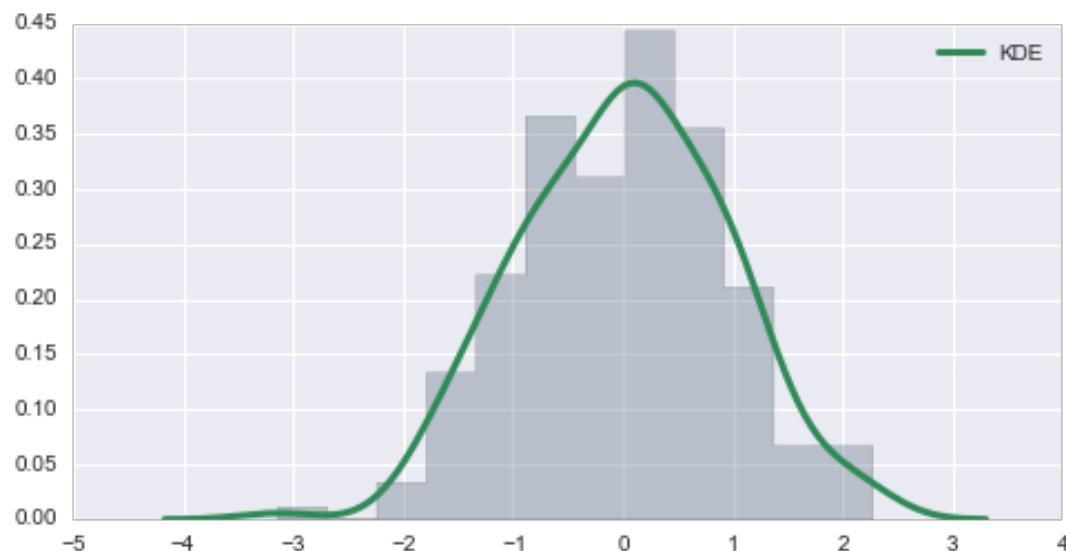
10 Bins

passengers

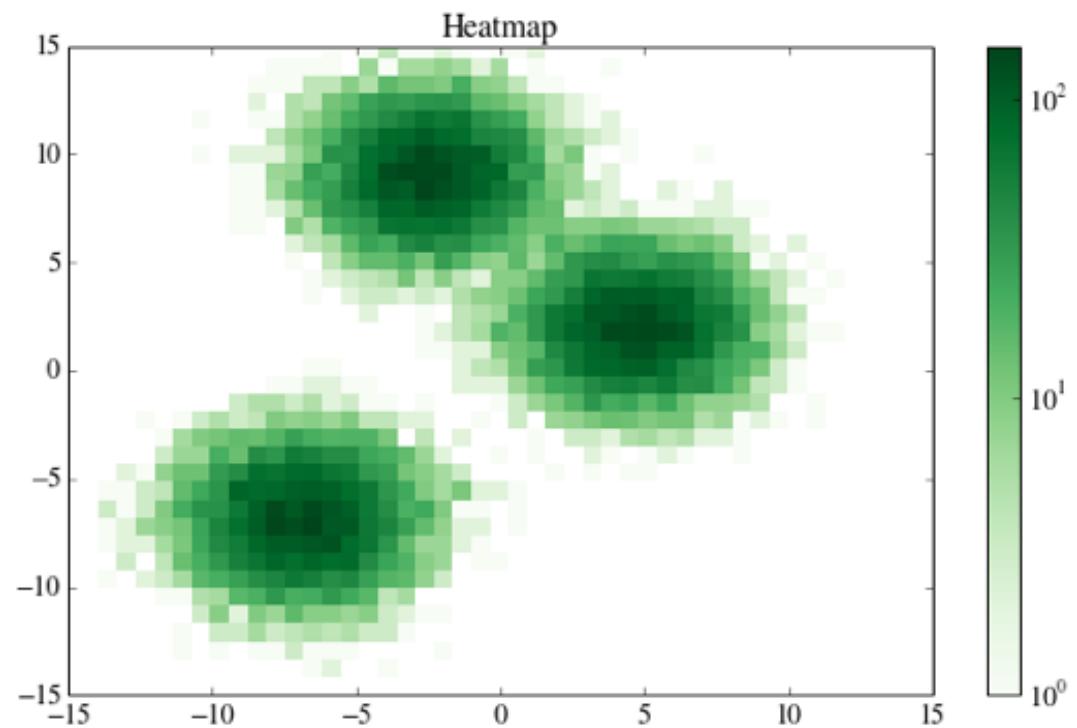
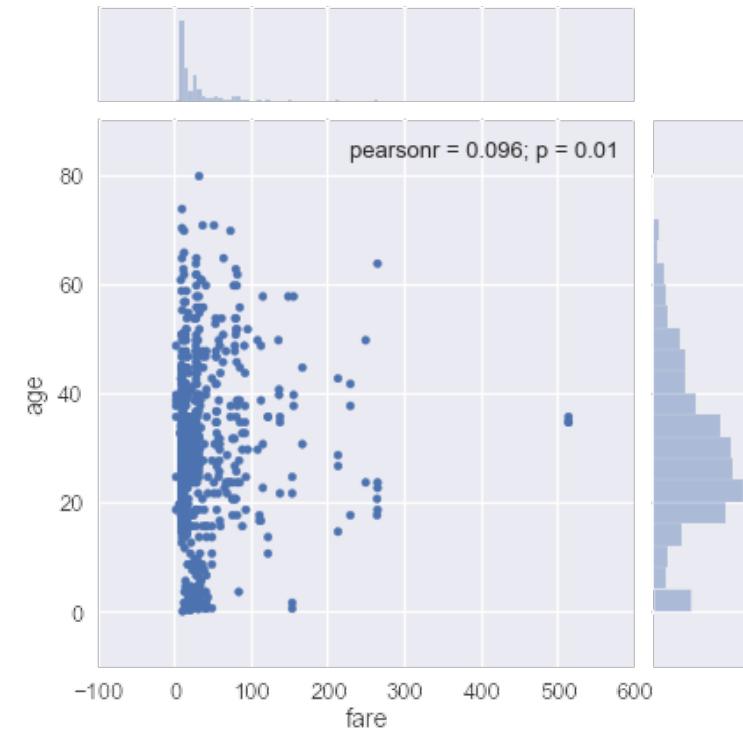
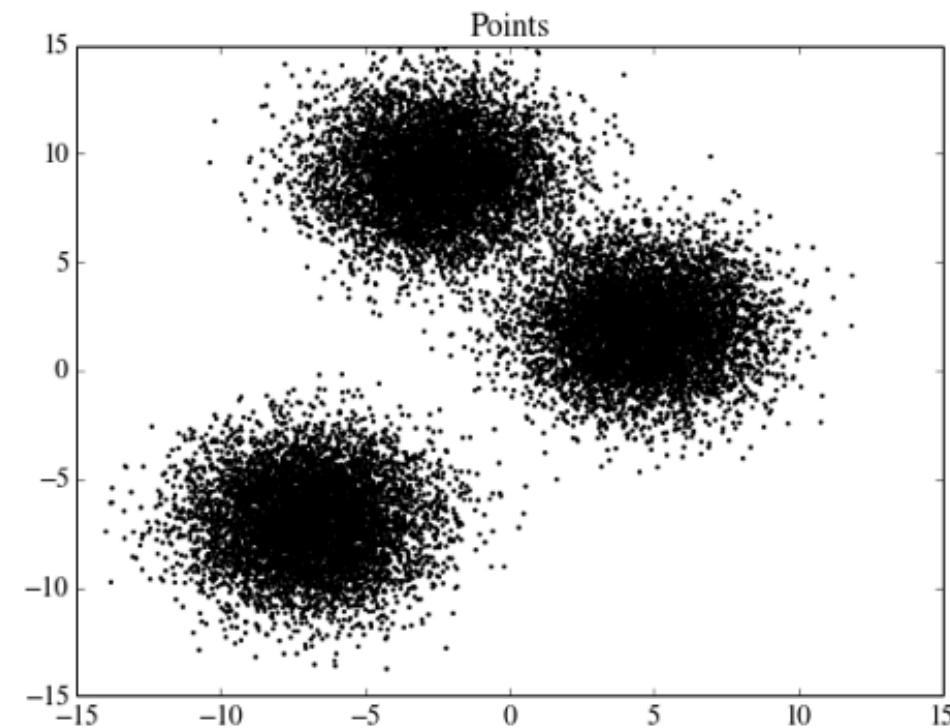


20 Bins

Density Plots



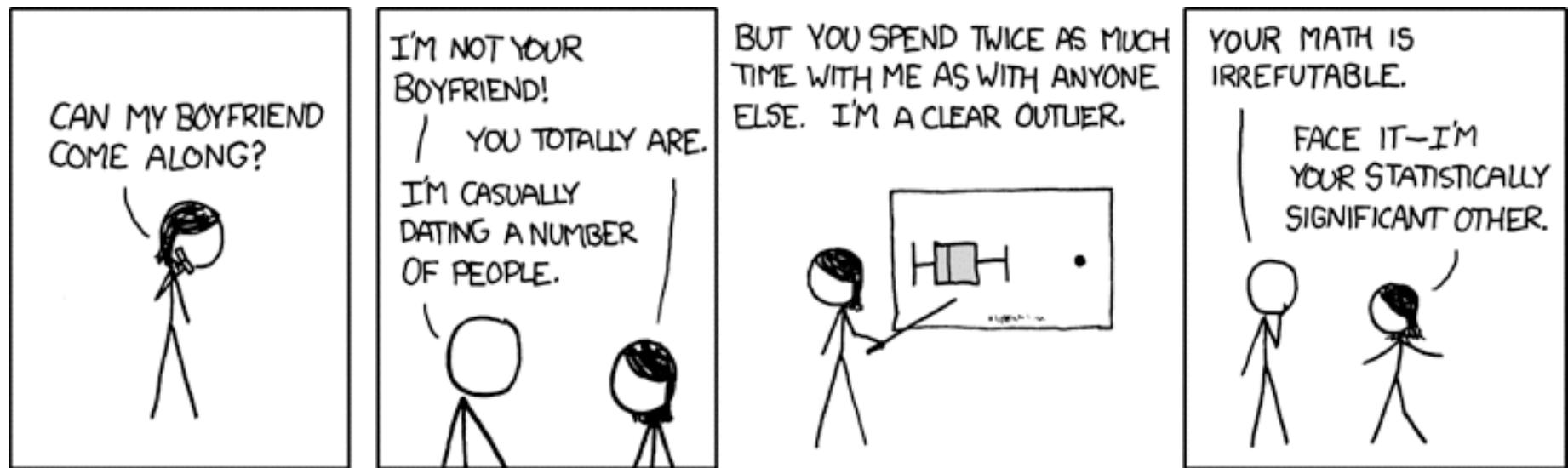
Heat Maps



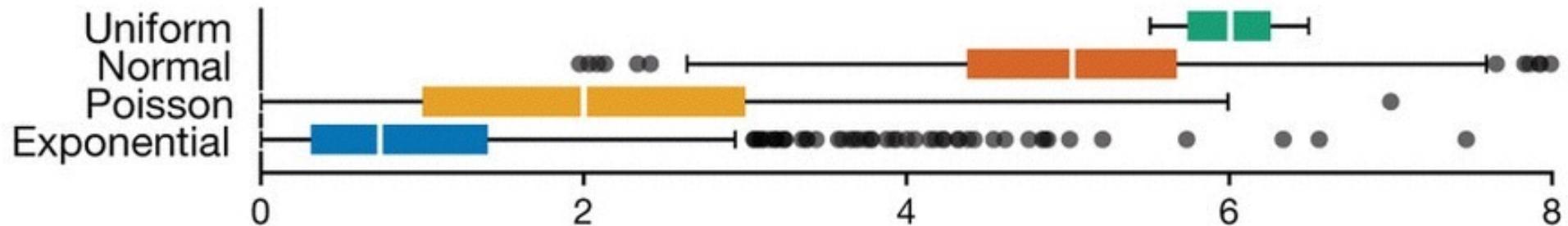
2D Density Plots

Box Plots

aka Box-and-Whisker Plot

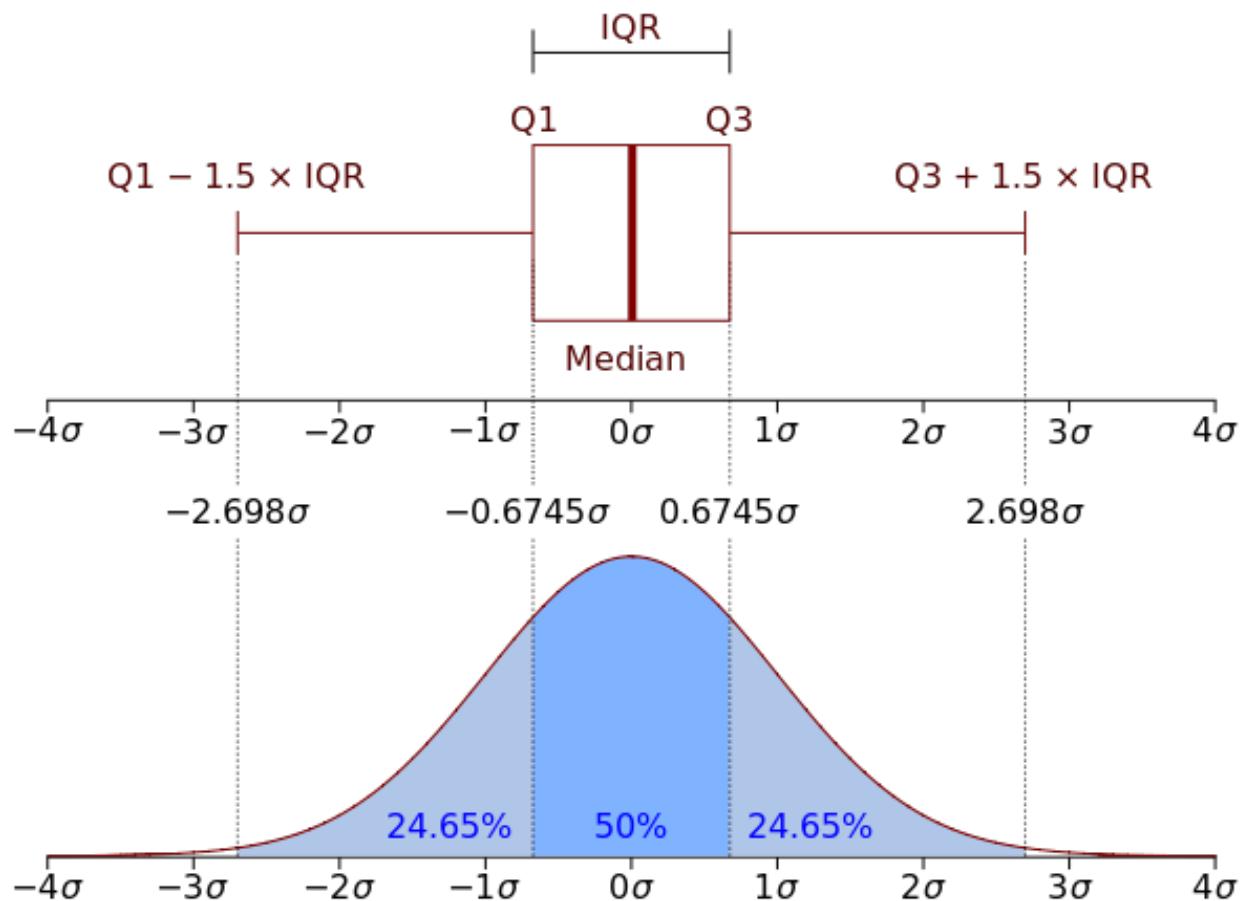


<http://xkcd.com/539/>



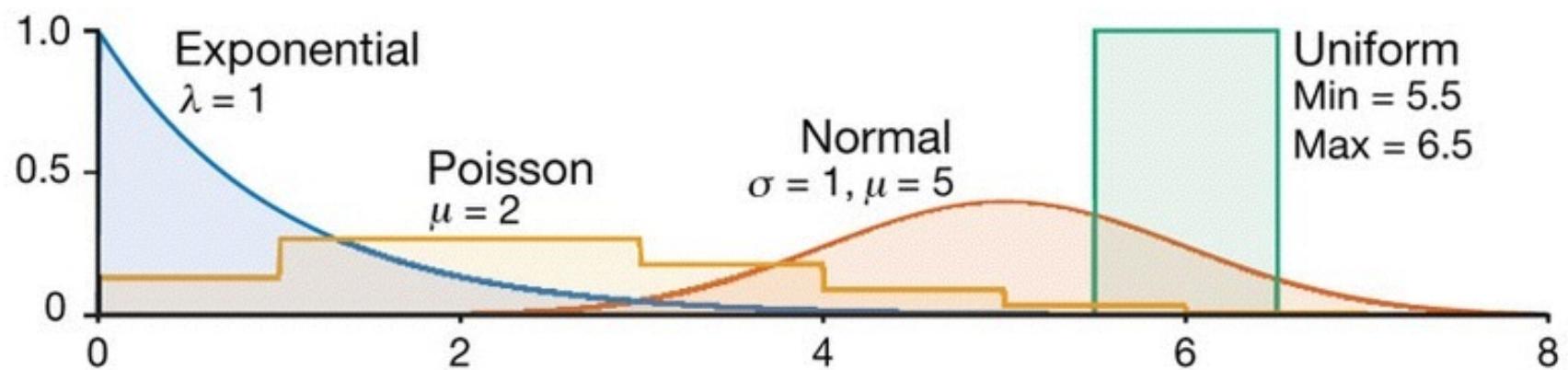
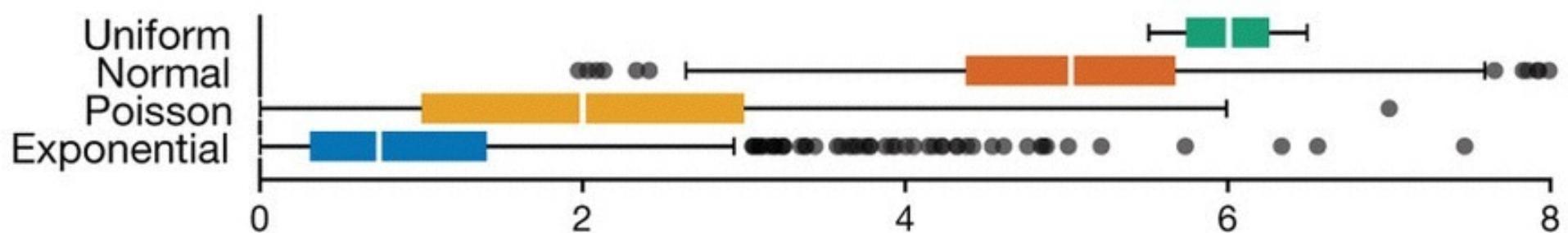
Box Plots

aka Box-and-Whisker Plot



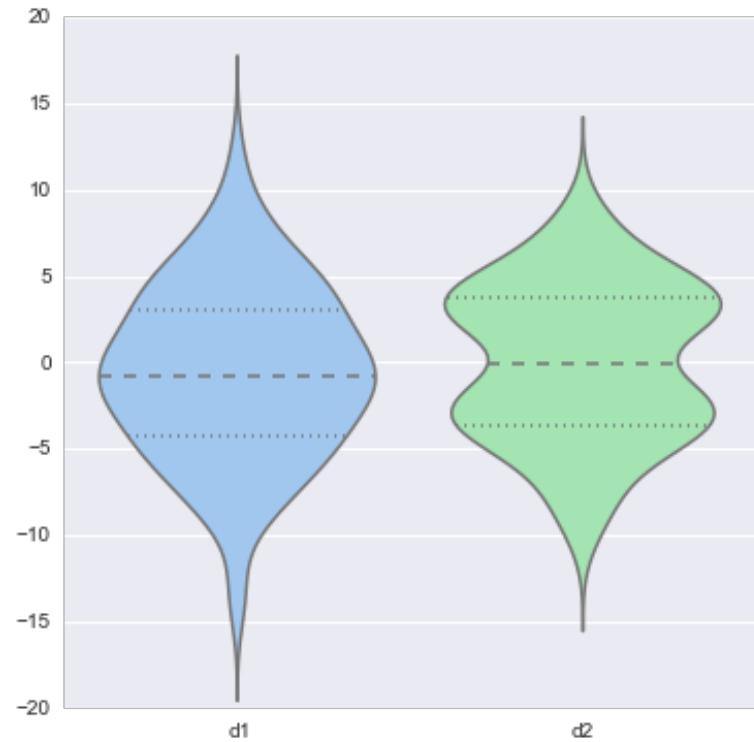
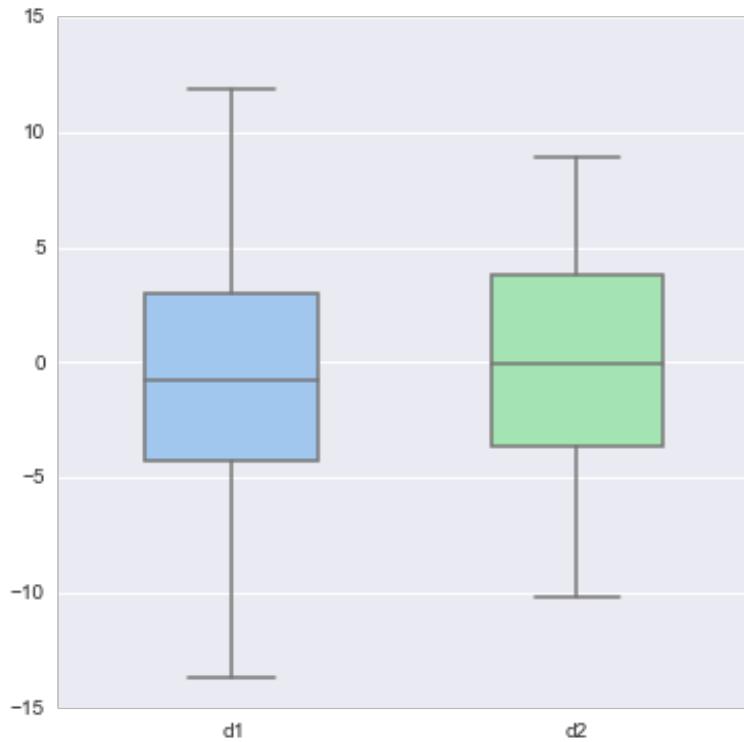
Wikipedia

Comparison



Violin Plot

= Box Plot + Probability Density Function

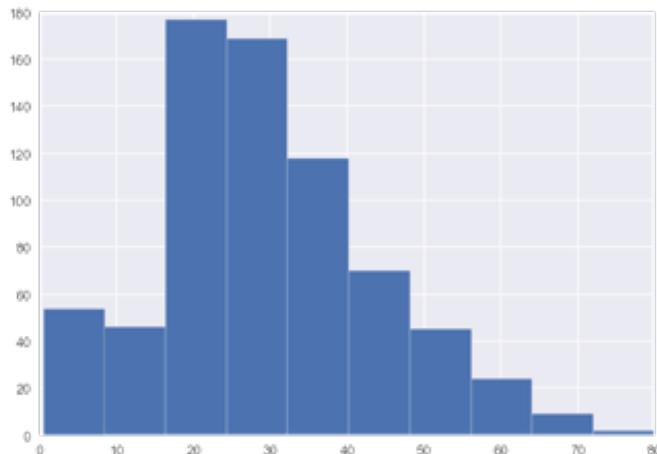


Multi-Dimensional Data Visualization

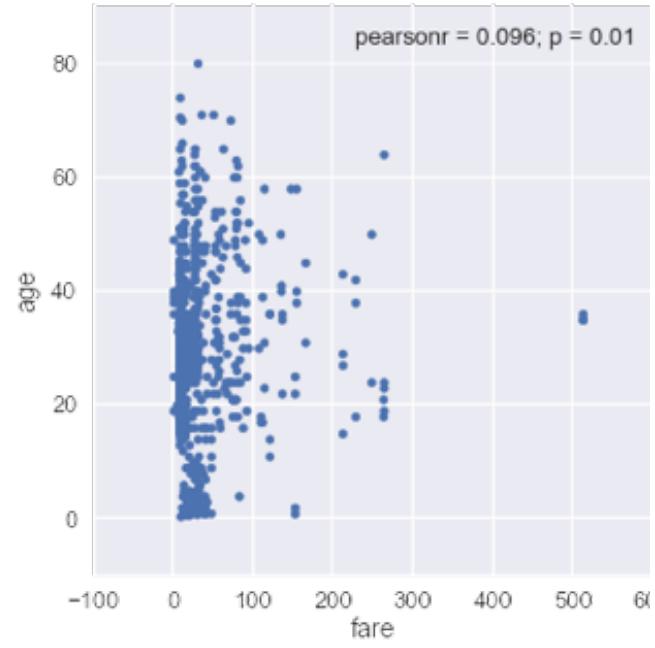
survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	3	male	22.0	1	0	7.25	S	Third	man	True		Southampton	no	False
1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
1	3	female	26.0	0	0	7.925	S	Third	woman	False		Southampton	yes	True
1	1	female	35.0	1	0	53.1	S	First	woman	False	C	Southampton	yes	False
0	3	male	35.0	0	0	8.05	S	Third	man	True		Southampton	no	True
0	3	male		0	0	8.4583	Q	Third	man	True		Queenstown	no	True
0	1	male	54.0	0	0	51.8625	S	First	man	True	E	Southampton	no	True
0	3	male	2.0	3	1	21.075	S	Third	child	False		Southampton	no	False
1	3	female	27.0	0	2	11.1333	S	Third	woman	False		Southampton	yes	False
1	2	female	14.0	1	0	30.0708	C	Second	child	False		Cherbourg	yes	False
1	3	female	4.0	1	1	16.7	S	Third	child	False	G	Southampton	yes	False
1	1	female	58.0	0	0	26.55	S	First	woman	False	C	Southampton	yes	True
0	3	male	20.0	0	0	8.05	S	Third	man	True		Southampton	no	True
0	3	male	39.0	1	5	31.275	S	Third	man	True		Southampton	no	False
0	3	female	14.0	0	0	7.8542	S	Third	child	False		Southampton	no	True
1	2	female	55.0	0	0	16.0	S	Second	woman	False		Southampton	yes	True
0	3	male	2.0	4	1	29.125	Q	Third	child	False		Queenstown	no	False
1	2	male		0	0	13.0	S	Second	man	True		Southampton	yes	True
0	3	female	31.0	1	0	18.0	S	Third	woman	False		Southampton	no	False
1	3	female		0	0	7.225	C	Third	woman	False		Cherbourg	yes	True
0	2	male	35.0	0	0	26.0	S	Second	man	True		Southampton	no	True
1	2	male	34.0	0	0	13.0	S	Second	man	True	D	Southampton	yes	True
1	3	female	15.0	0	0	8.0292	Q	Third	child	False		Queenstown	yes	True

Example:Titanic Dataset

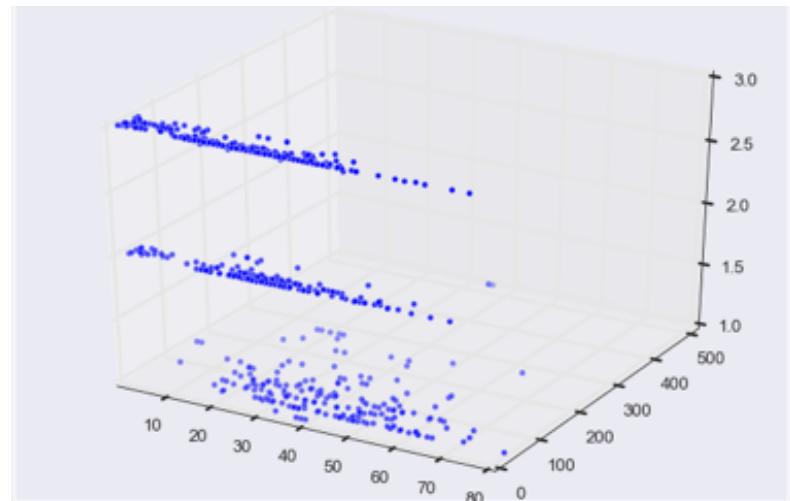
ID



2D



4D? ←



3D

What is “high” dimensional?

How many dimensions (attributes)?

- ~50 – tractable with “just” vis
- ~1,000 – need analytical methods

How many items?

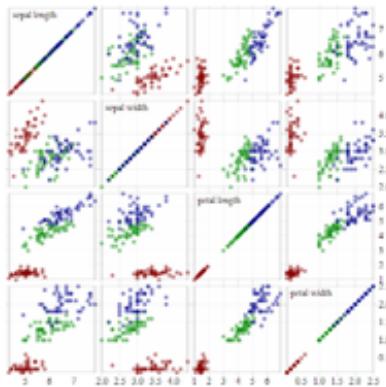
- ~ 1,000 – “just” vis is fine
- >> 10,000 – need analytical methods

Homogeneity

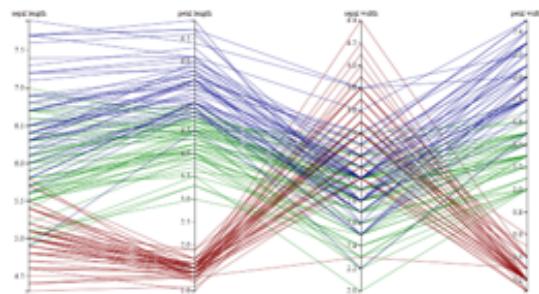
Same data type?

Same scales?

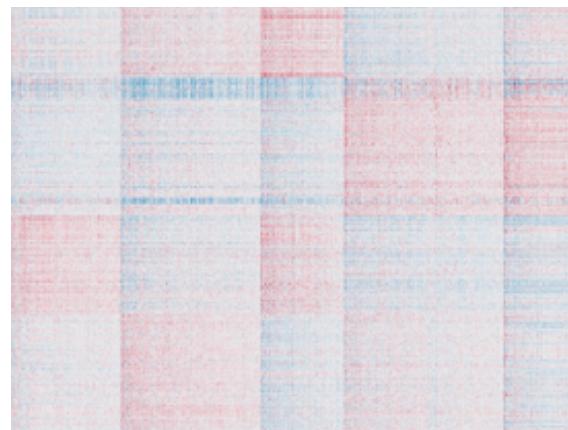
Analytic Component



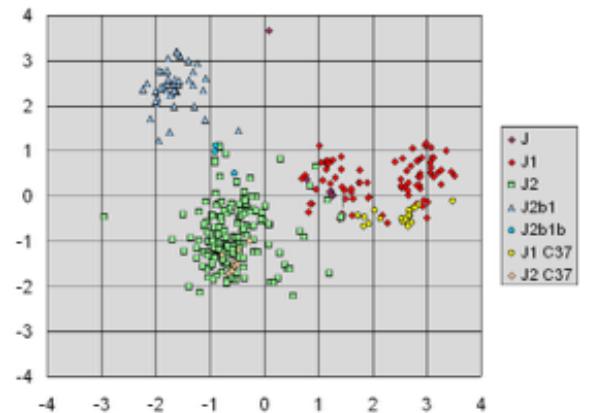
Scatterplot Matrices



Parallel Coordinates



Pixel-based Visualizations /
Heat Maps



Dimensionality
Reduction
(e.g., PCA)



no / little analytics

strong analytics
component

Heat Map

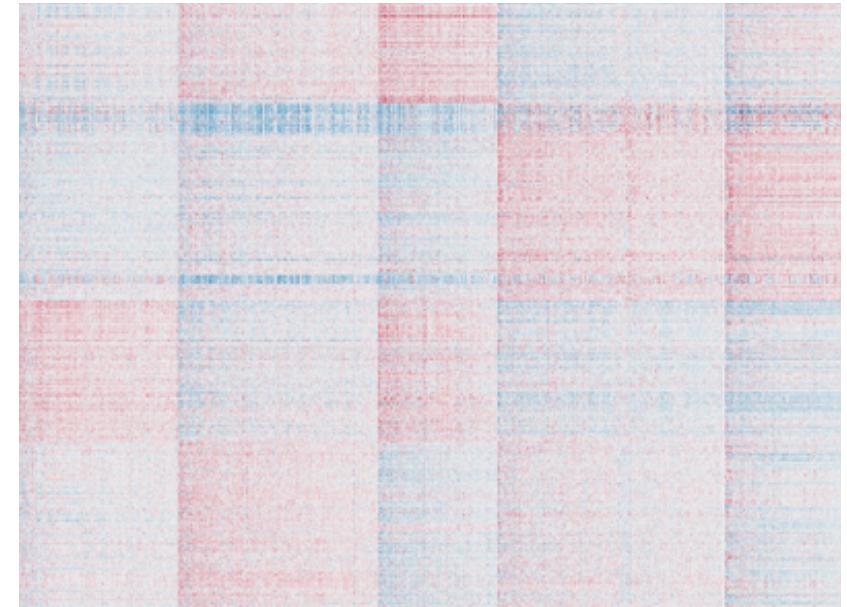
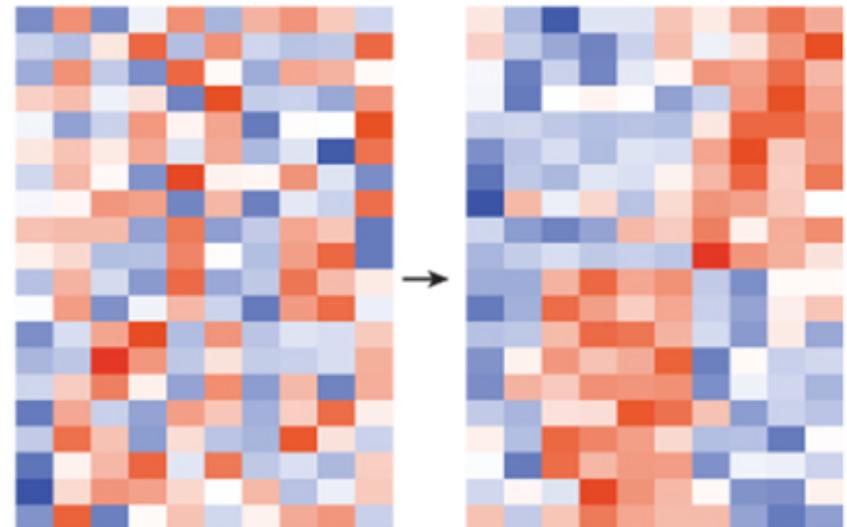
Each cell is a “pixel”, value encoded using color

Meaning derived from ordering

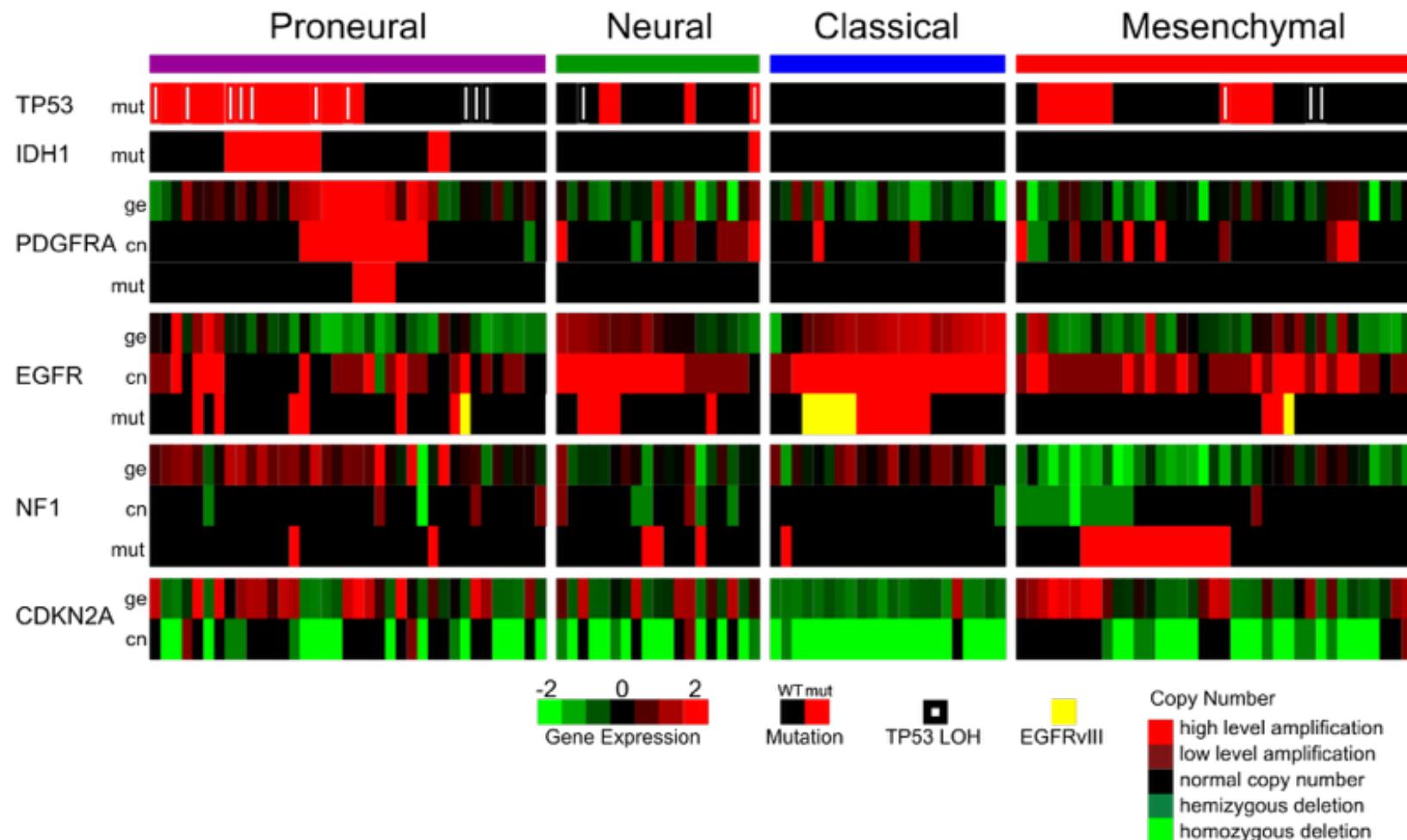
If no ordering inherent,
clustering is used

Scalable – 1 px per item

Good for homogeneous data



Heterogeneous Data?

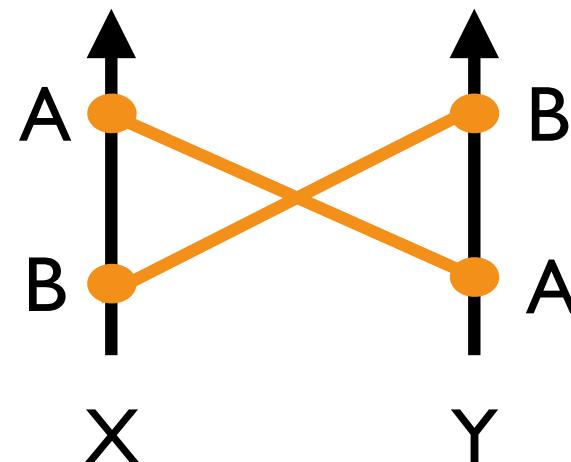
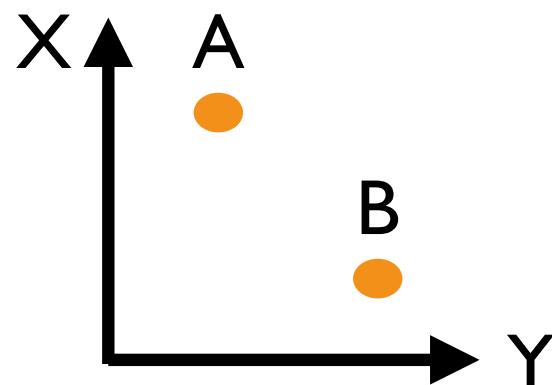


Parallel Coordinates (PC)

Inselberg 1985

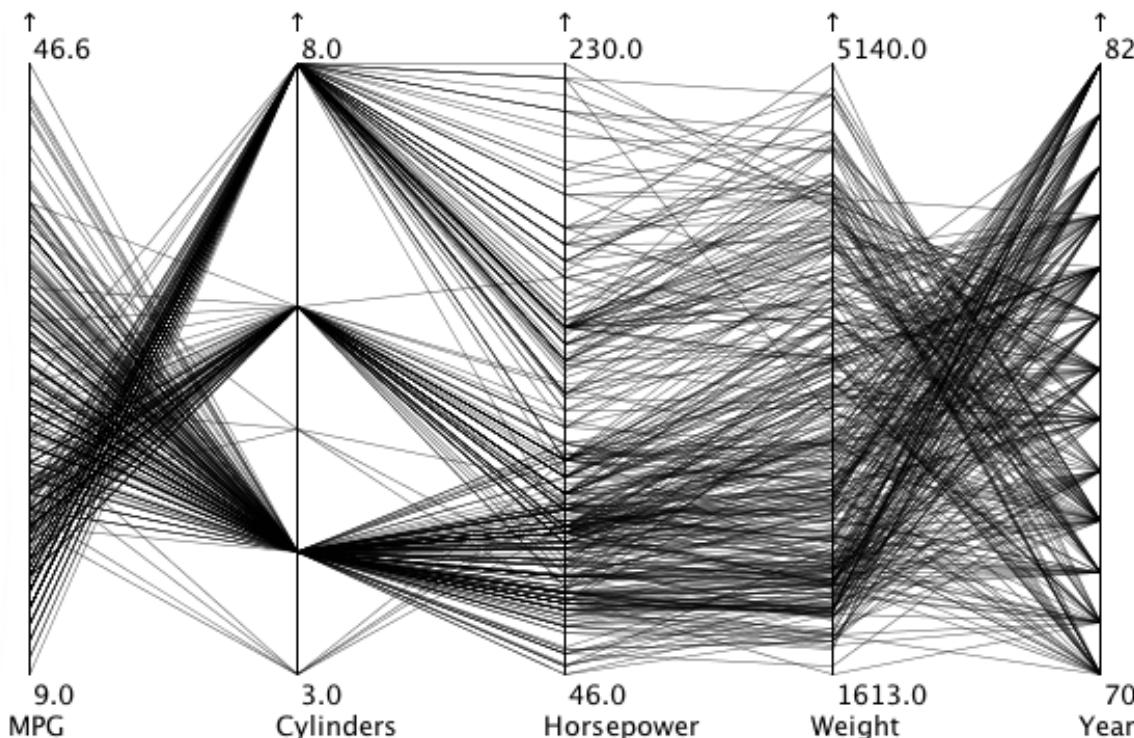
Axes represent attributes

Lines connecting axes represent items



Example: Cars Dataset

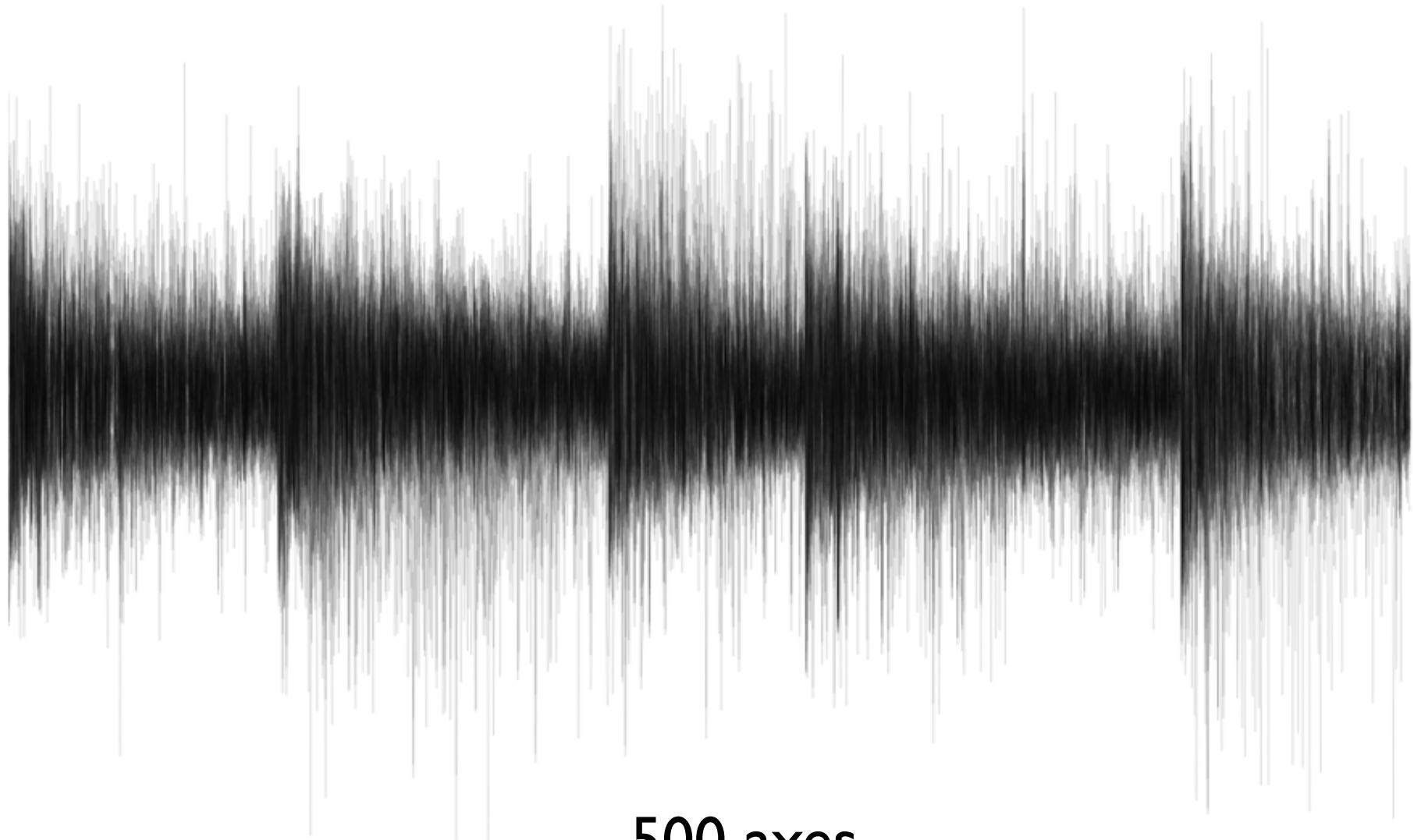
1	MPG	Cylinders	Horsepower	Weight	Acceleration	Year	Origin
2	18	8	130	3504	12	70	USA
3	15	8	165	3693	11.5	70	USA
4	18	8	150	3436	11	70	USA
5	16	8	150	3433	12	70	USA
6	17	8	140	3449	10.5	70	USA
7	15	8	198	4341	10	70	USA
8	14	8	220	4354	9	70	USA
9	14	8	215	4312	8.5	70	USA
10	14	8	225	4425	10	70	USA
11	15	8	190	3850	8.5	70	USA
12	15	8	170	3563	10	70	USA
13	14	8	160	3609	8	70	USA
14	15	8	150	3761	9.5	70	USA
15	14	8	225	3086	10	70	USA
16	24	4	95	2372	15	70	Europe
17	22	6	95	2833	15.5	70	USA
18	18	6	97	2774	15.5	70	USA
19	21	6	85	2587	16	70	USA



Limitations of PC?

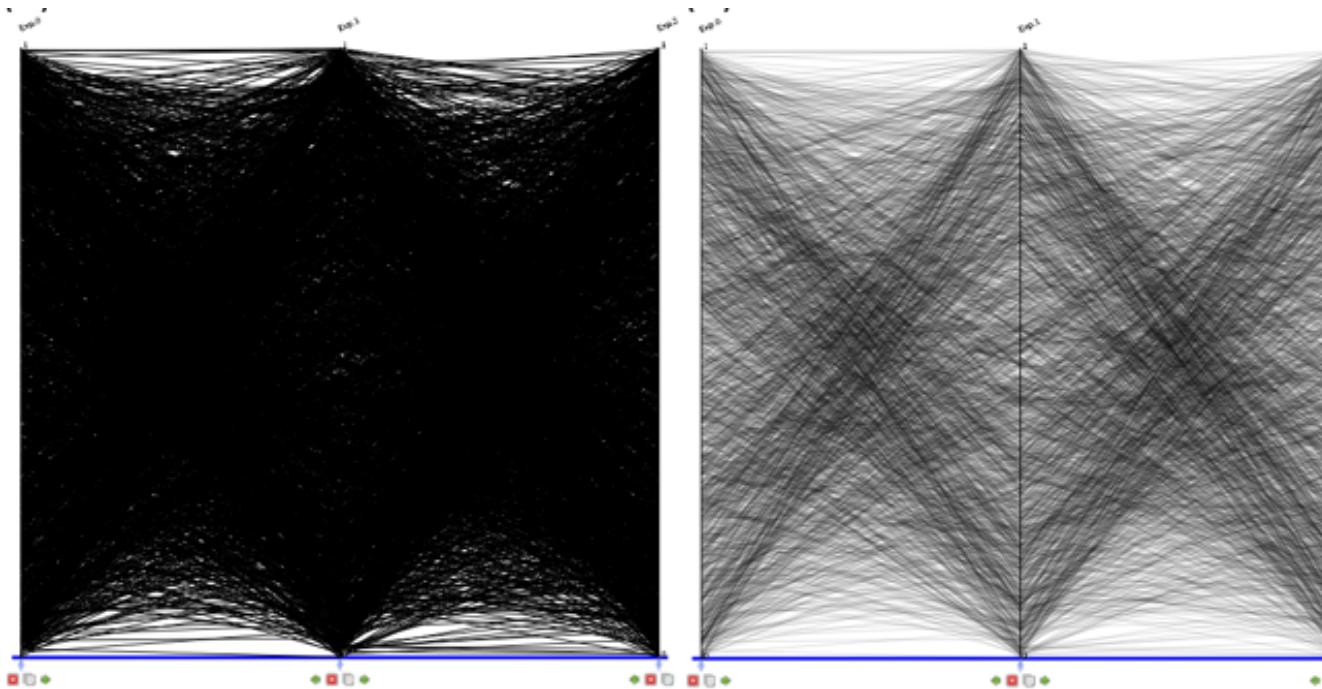
PC Limitations

Scalability to Many Dimensions



PC Limitations

Scalability to Many Items



Solutions:

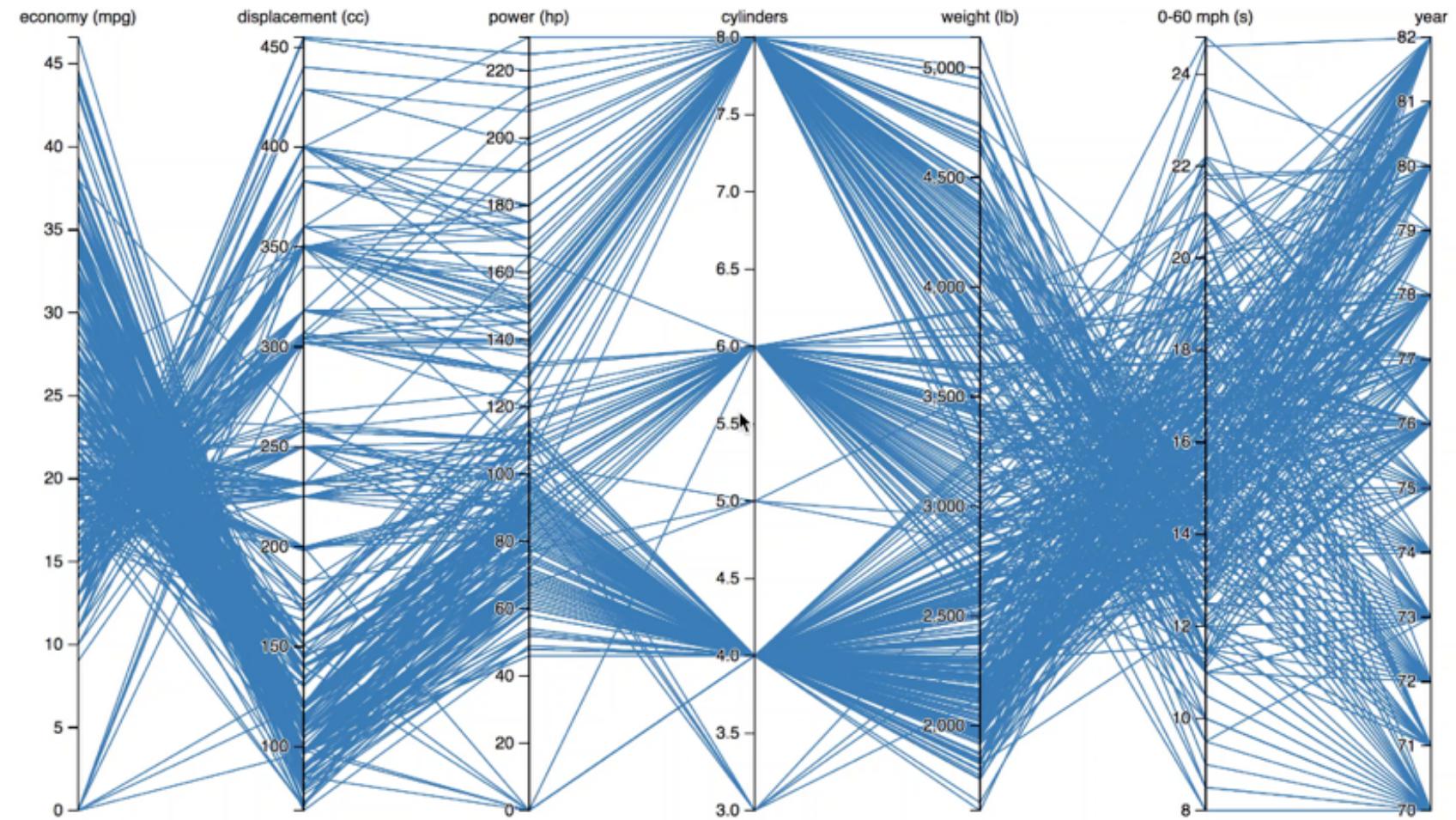
Transparency

Bundling, Clustering

Sampling

PC Limitations

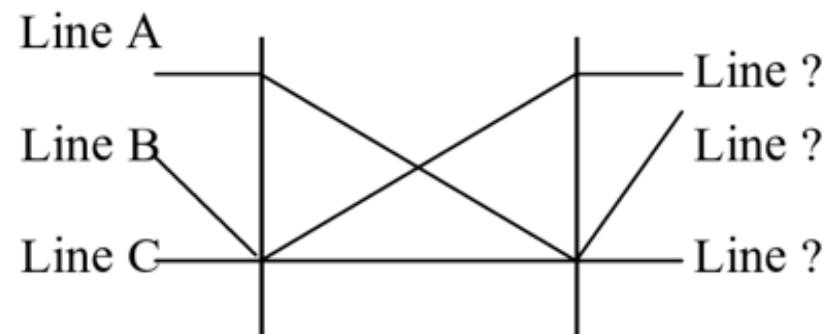
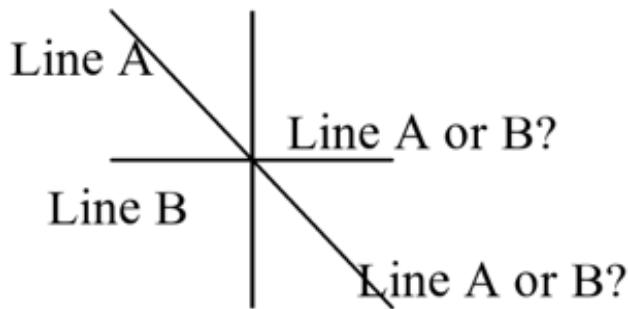
Correlations only between adjacent axes



Solution: Let user change order

PC Limitations

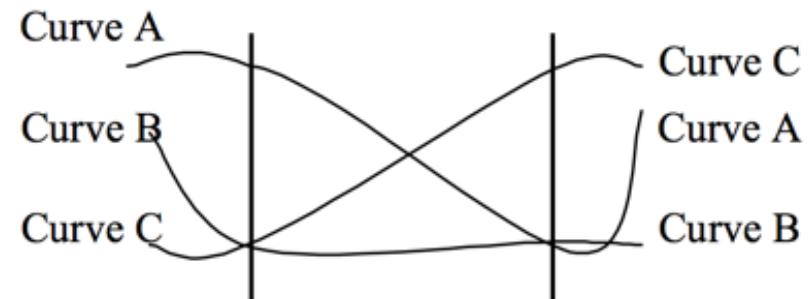
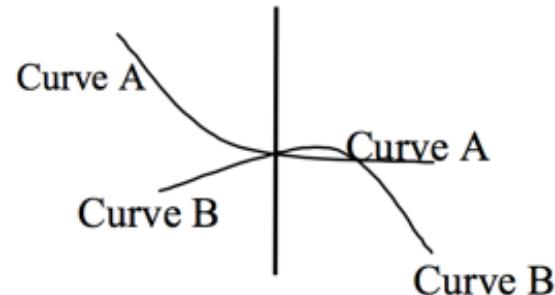
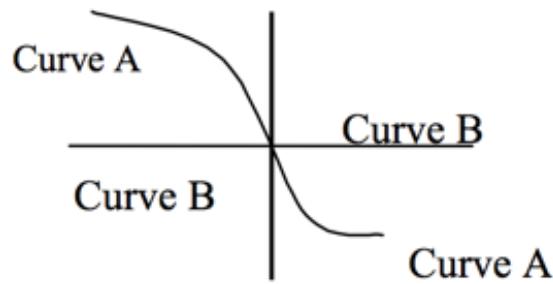
Ambiguity



Solutions:

Interactive highlighting

Curves

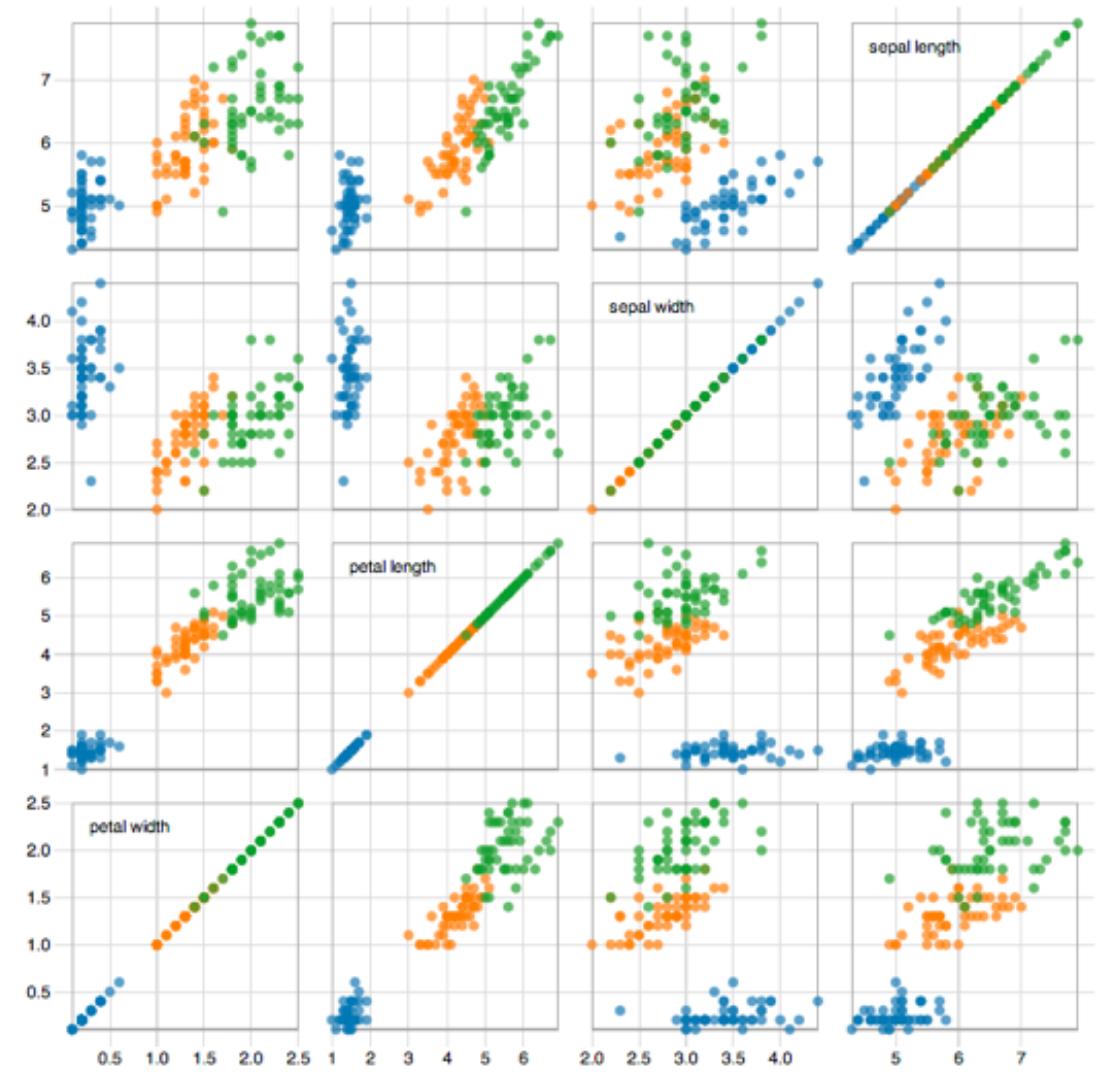


Scatterplot Matrix (SPLOM)

N dimensions

N^2 scatterplots

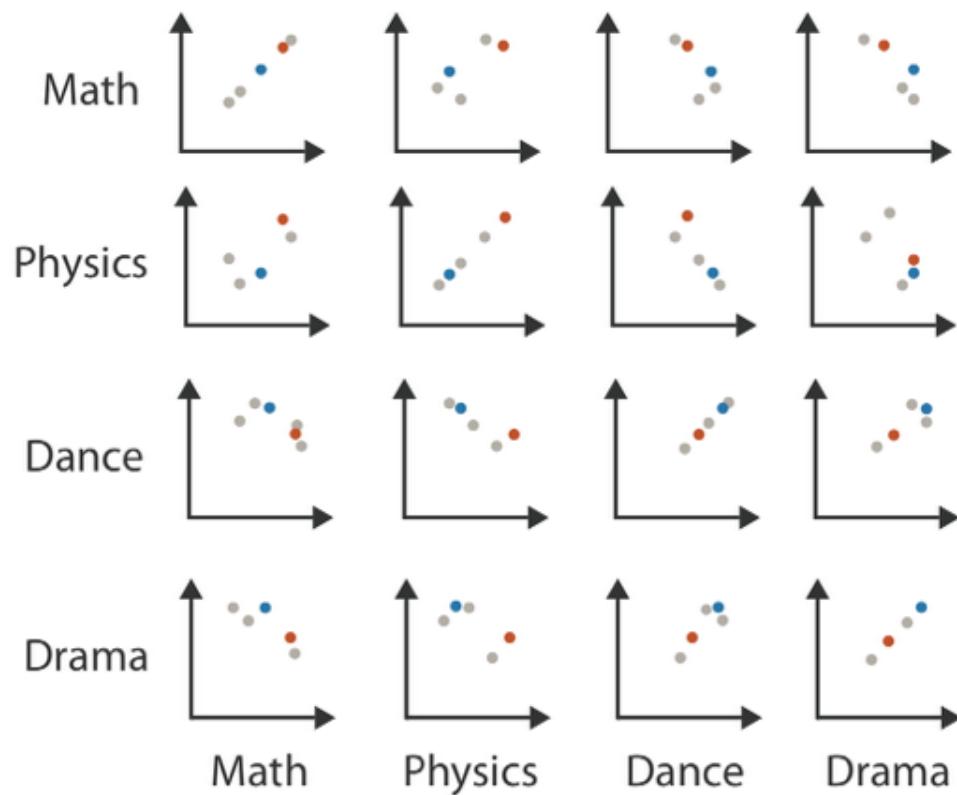
Limited scalability
(~20 dims,
~500-1k items)



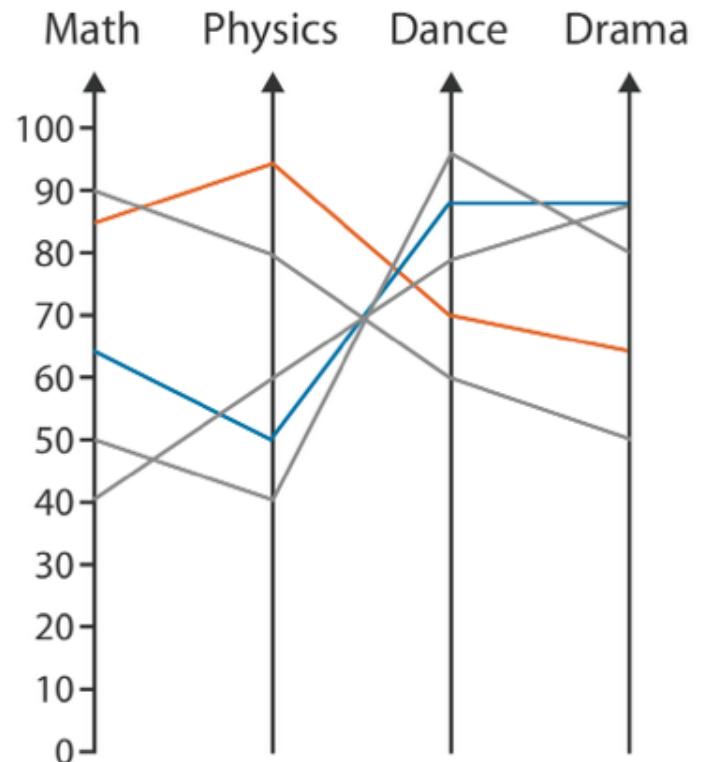
Table

	Math	Physics	Dance	Drama
	85	95	70	65
	90	80	60	50
	65	50	90	90
	50	40	95	80
	40	60	80	90

Scatterplot Matrix



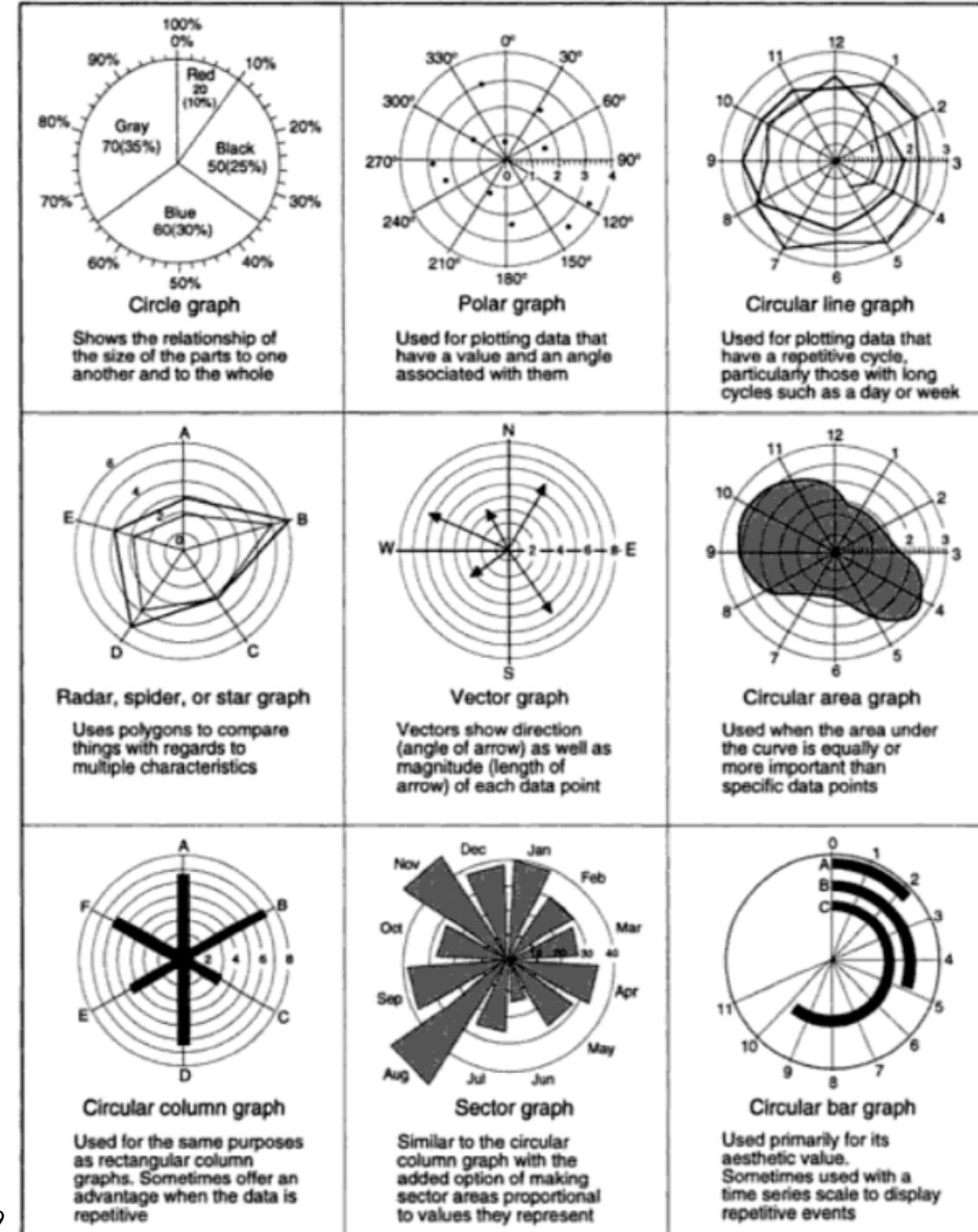
Parallel Coordinates



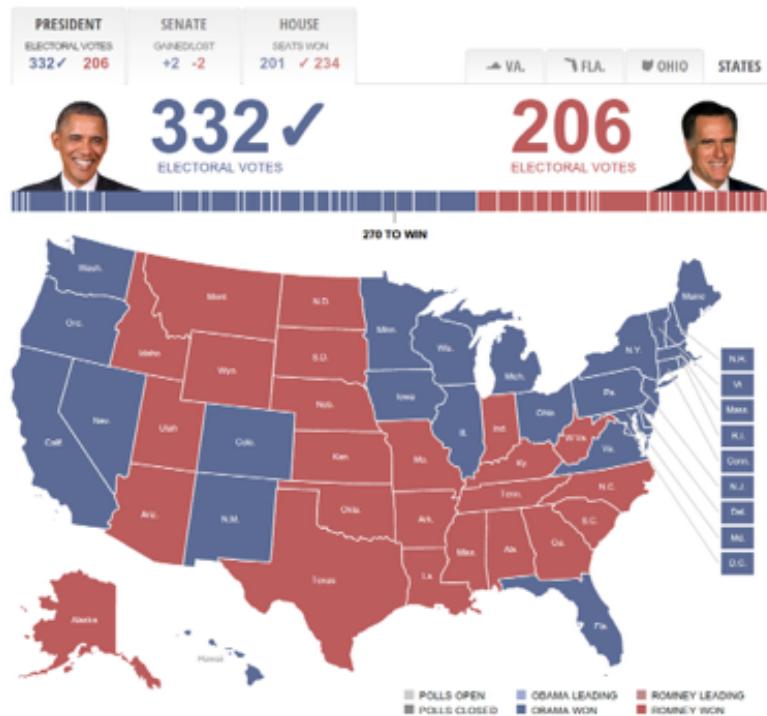
Radial Axis Techniques

Similar to parallel coordinates

Axes radiate from a common origin



Map Visualization



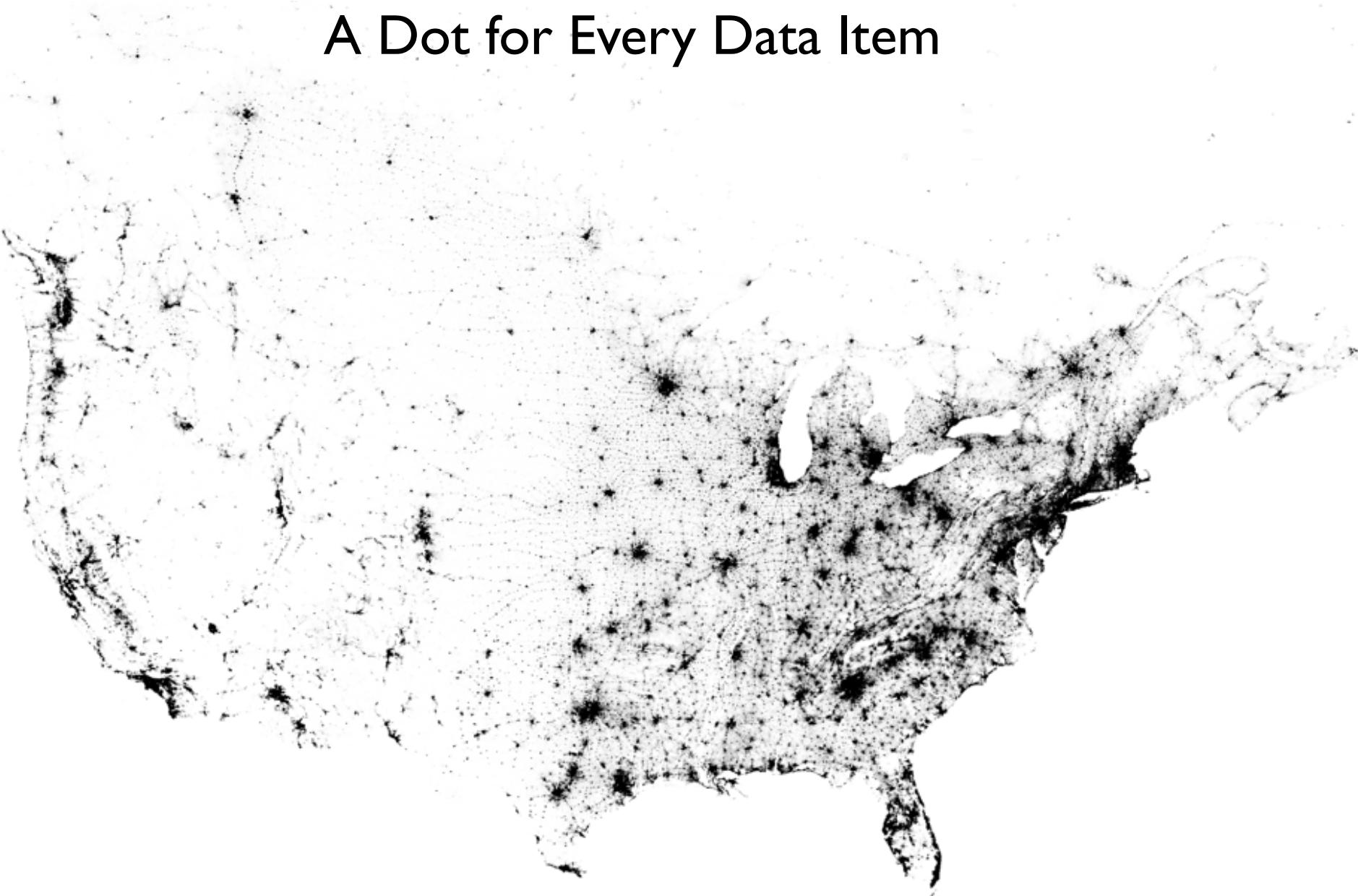
Visual Variables for Spatial Data

	Size	Shape	Brightness	Color	Orientation	Spacing	Perspective height	Arrangement
Point								
Linear								
Areal								

Slocum 1999

Dot Map

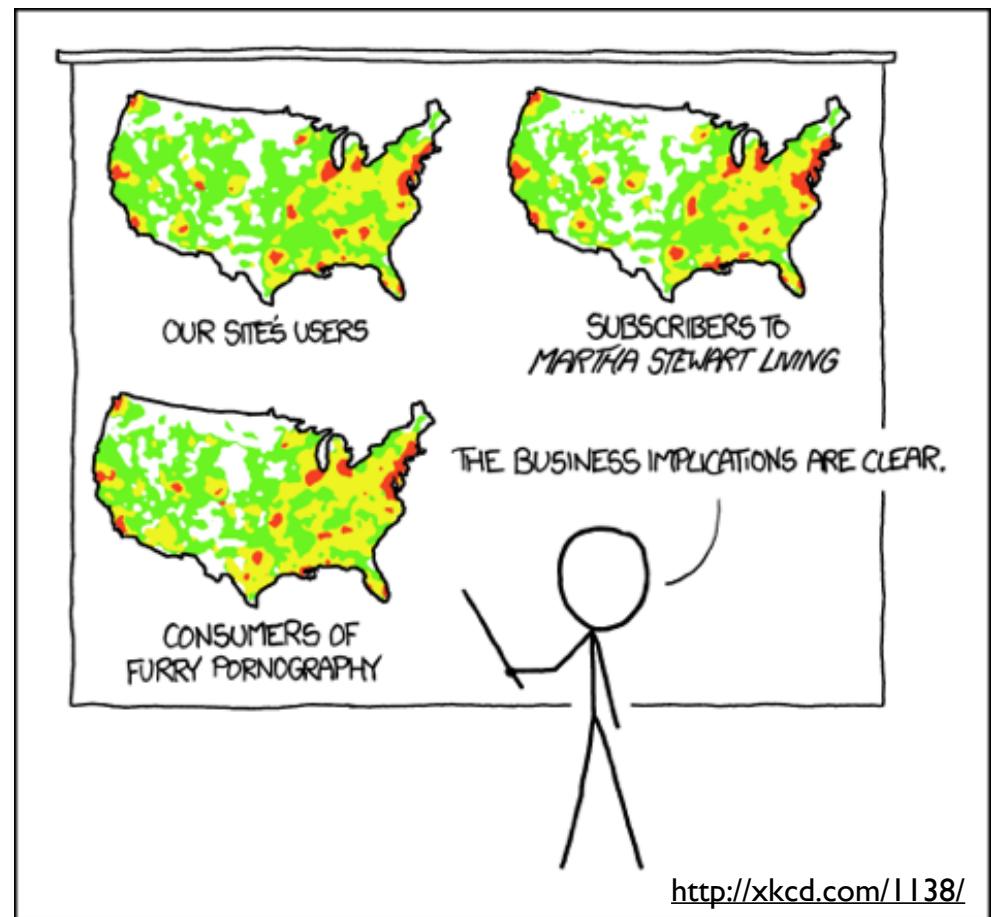
A Dot for Every Data Item



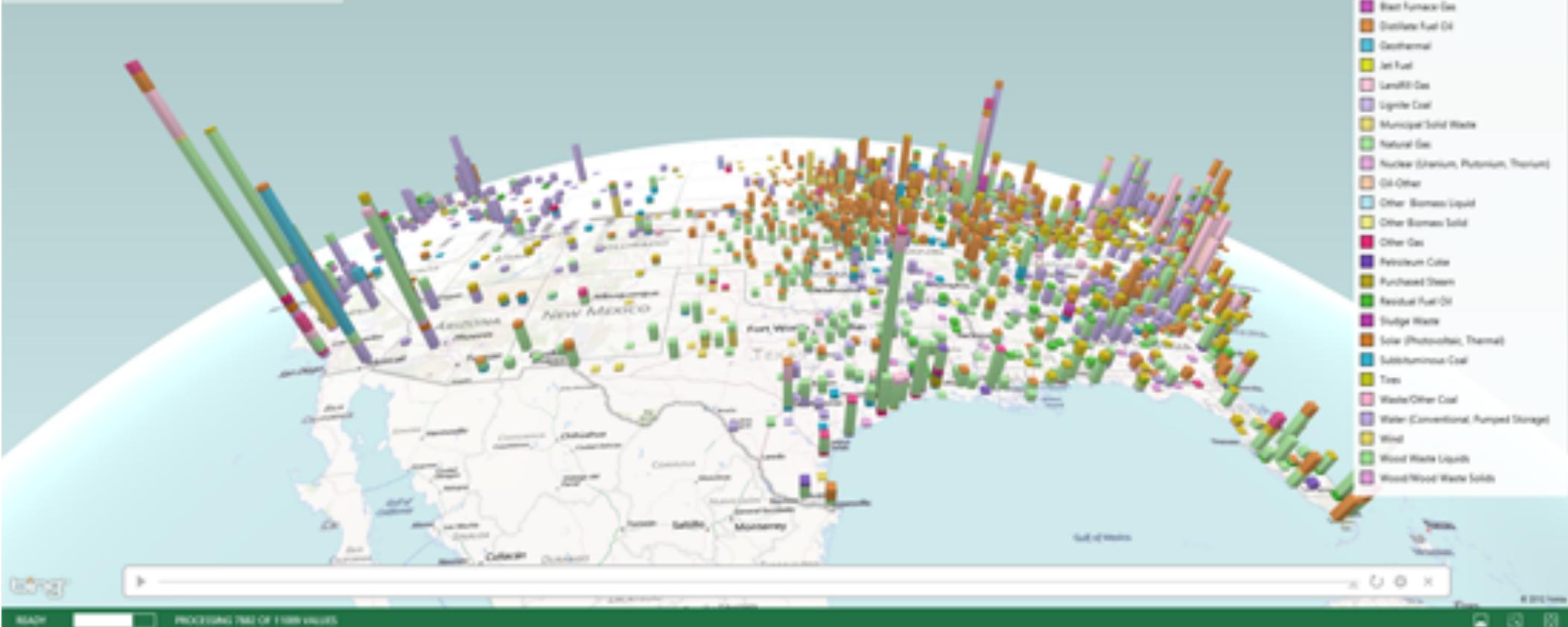


<http://www.cairco.org/news/those-scary-dots-population-america>

Maps can lie, too!



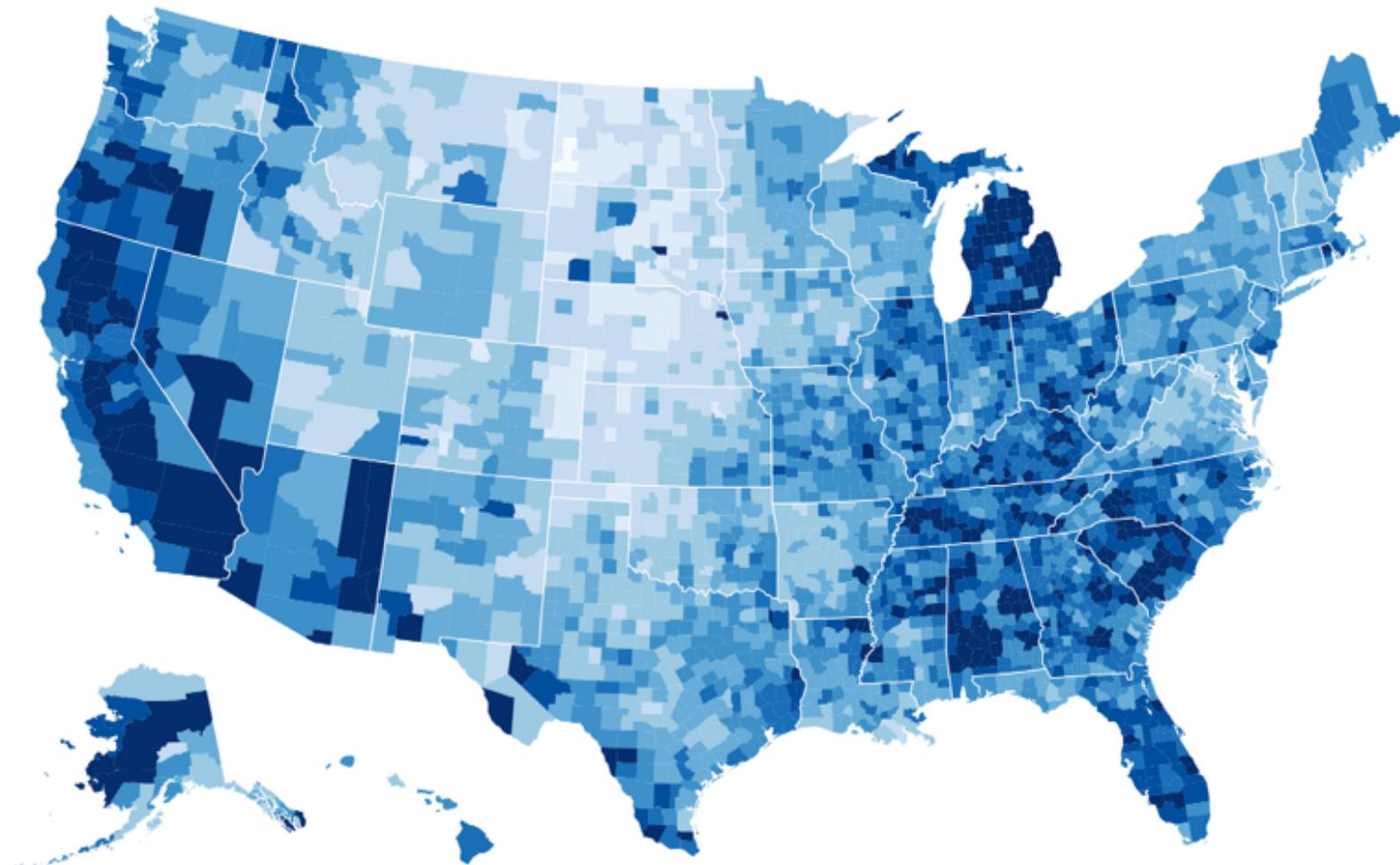
PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS



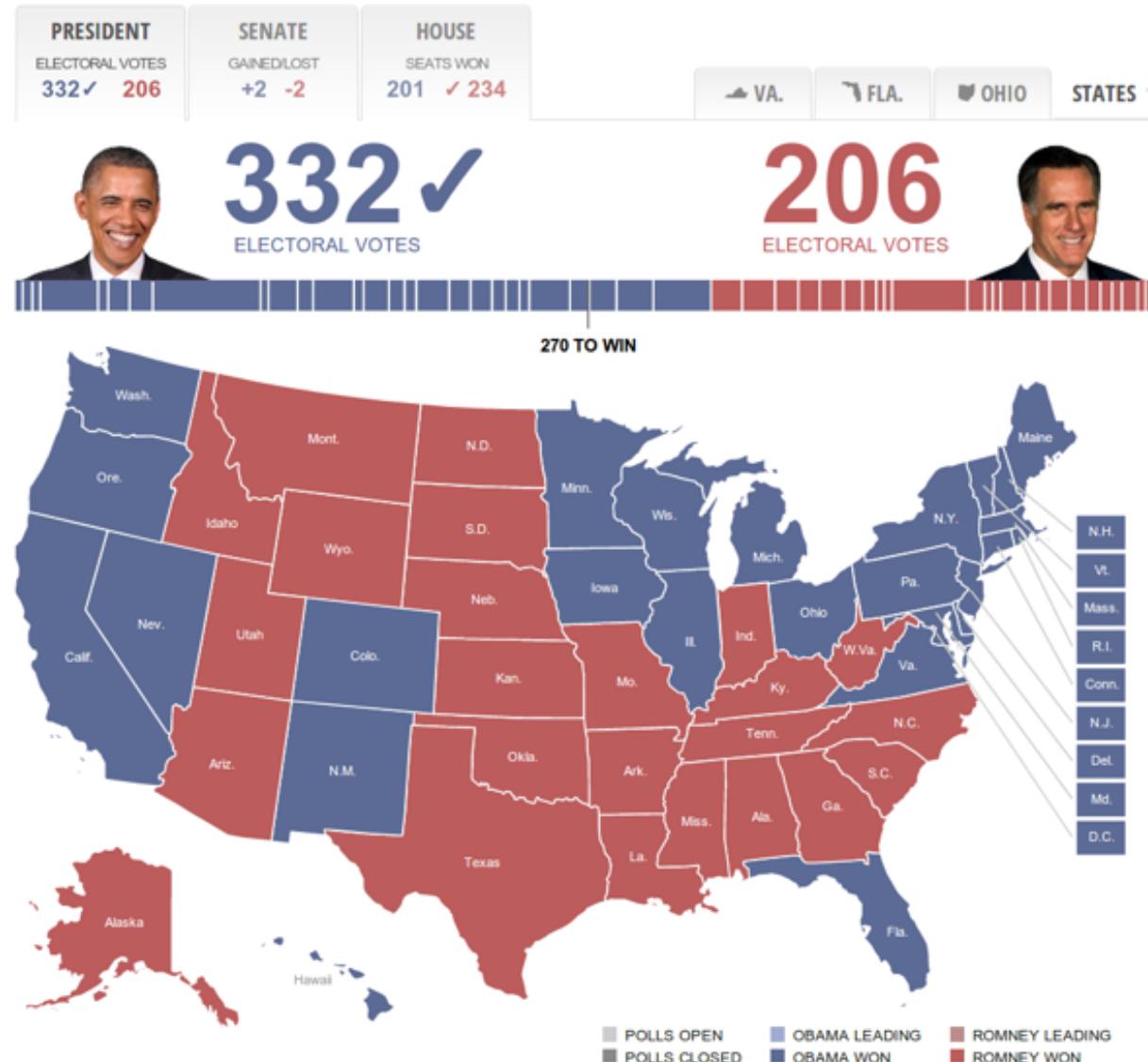
Microsoft GeoFlow – Part of Excel 2013

Choropleth Map

Attribute uniformly distributed in region



Misleading Coropleth Map



Better Version by NYT

In a Decisive Victory, Obama Reshapes the Electoral Map

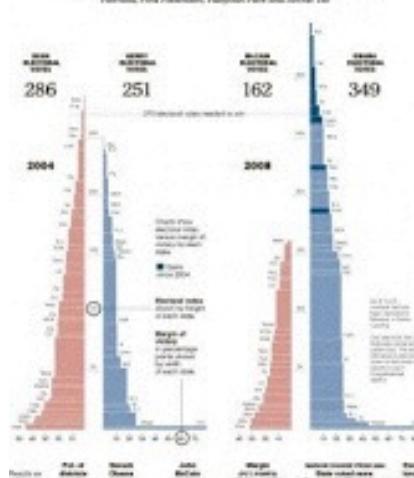
Barack Obama's historic win, with at least 349 electoral votes to John McCain's 162, can be attributed to his victories in several high-population states like Florida, Michigan and Ohio, that George W. Bush won handily in 2004.

The struggling economy, especially in manufacturing states, and high numbers of new immigrants helped tip key areas from red to blue.

Even where Mr. McCain beat Mr. Obama, he won by a thinner margin, as most of the swing states went to Mr. Obama and some less-volatile states went to Mr. McCain.

— moving toward the Democratic Party.

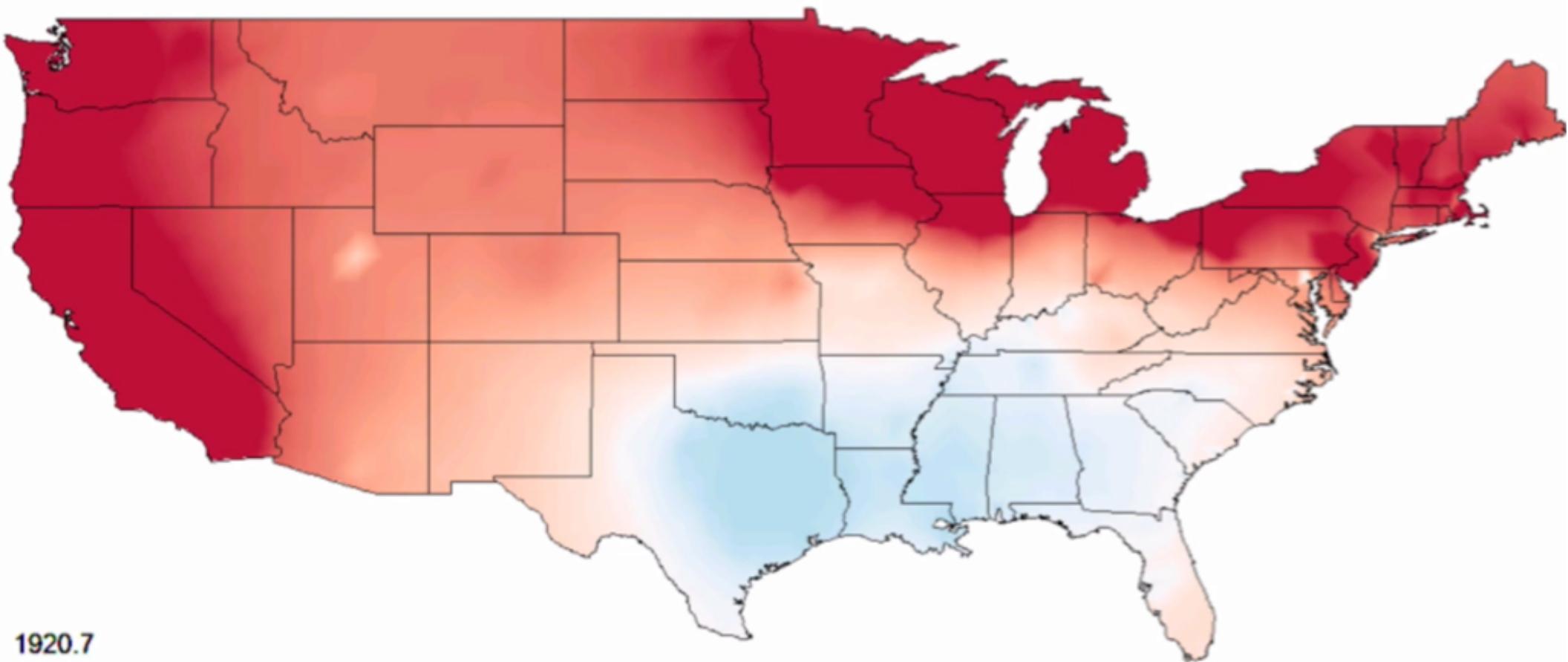
By Eric Alterman, Joe Borgna, Badri Chaganti, Marlene Fracica, Howard Fineman, Fred Glaeser, Hartley Park and Jennifer Tsai



States won by Obama		2008	
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	100%	1,001,004,375	+1.0%
North Dakota	100%	100,174,400	+1.0%
Wyoming	100%	1,001,004,375	+1.0%
Montana	100%	1,001,004,375	+1.0%
Idaho	10		

Isarithmic Map

Color coding continuous phenomena

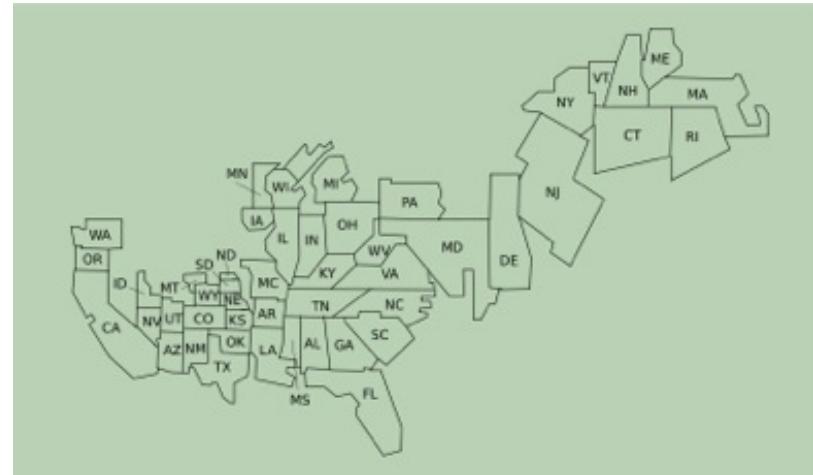


Cartograms

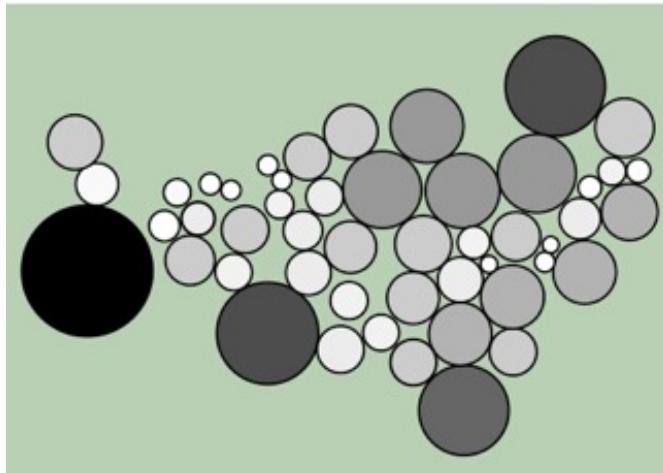
Size of region scaled to attribute value



Noncontinuous cartogram



Noncontiguous cartogram

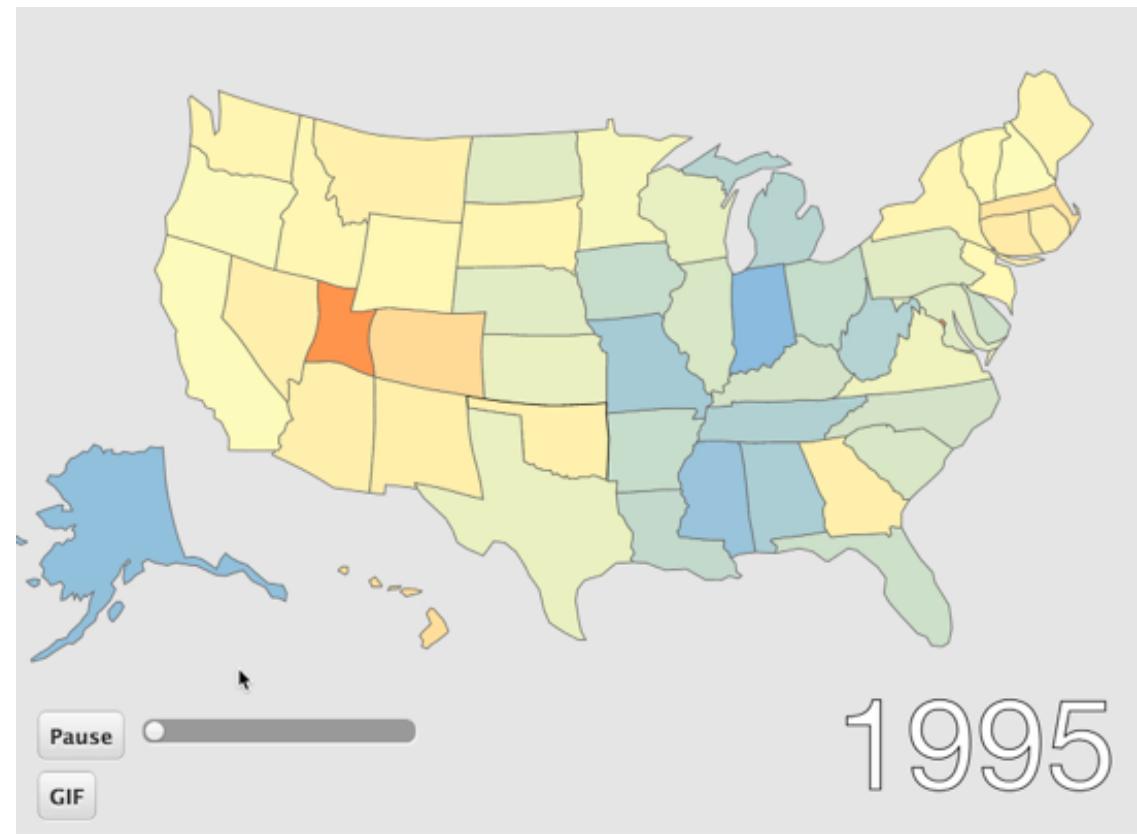
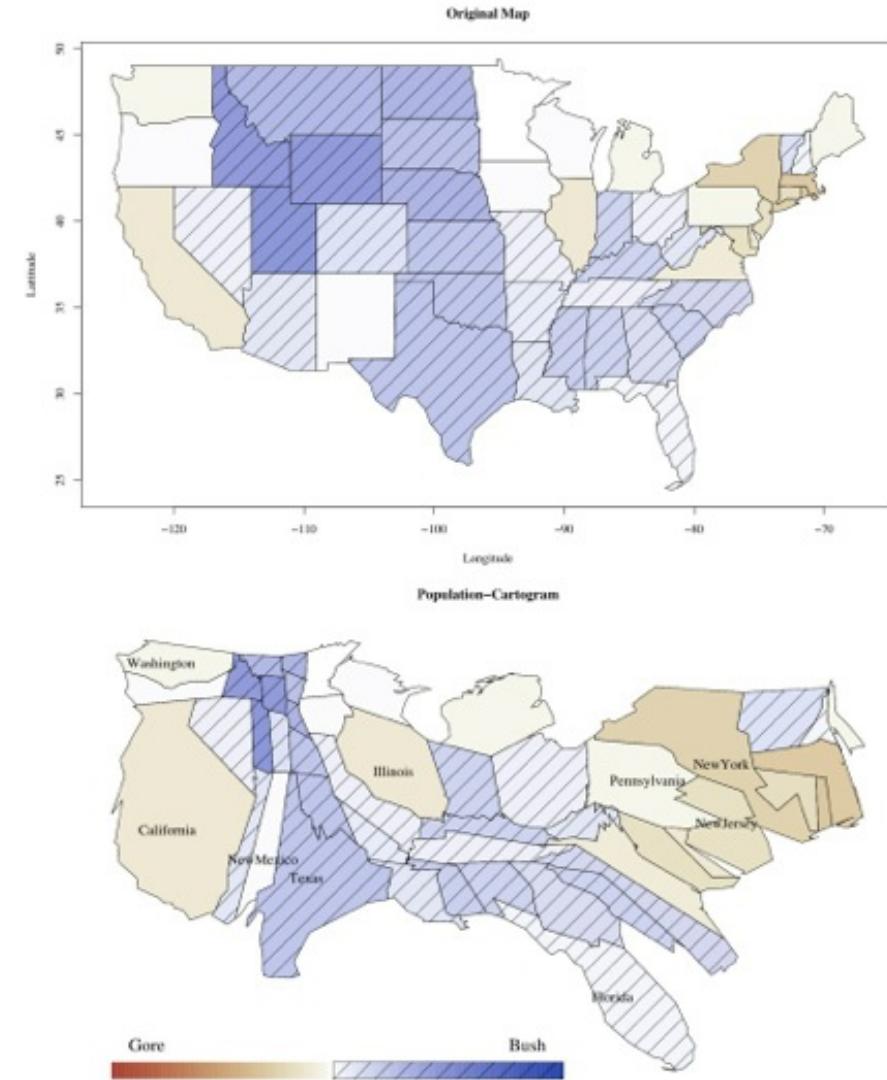


Circular cartogram



Continuous cartogram

Continuous Cartogram Example



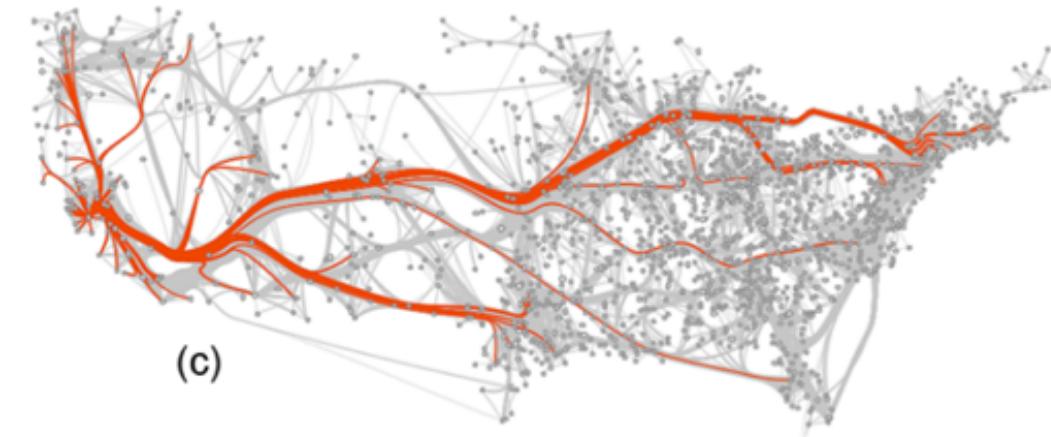
Edges Augmented Onto Map



(a)



(b)



(c)

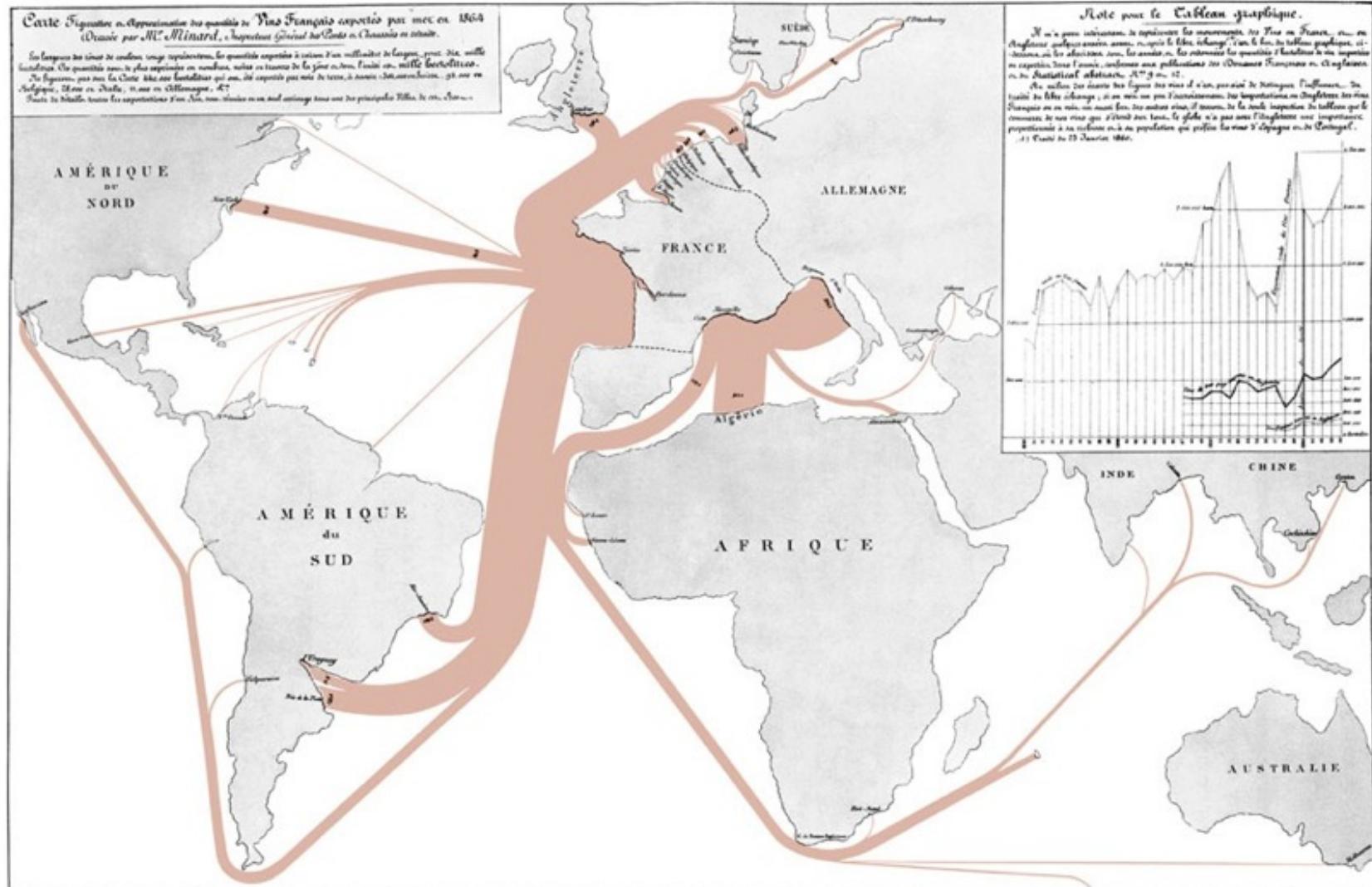


(d)

Holten 2009

Force Directed Edge Bundles
US migration graph (1715 nodes, 9780 edges)

Flow Maps



Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845-1869*, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

Text Visualization

\$59,413,405,476,974

The outstanding United States public debt as of Sept 15, 2014.

Tag Cloud / Word Cloud / Wordle

Change word size/color by frequency

Institutional freshman facilities learning established
article Henry Technology Aid well MIT 19 institution North culture
February Information external See body including Financial Robert 31 Handbook Admissions
House Libraries 2001 Search Ivy January 2007 Washington Medical 2003 many
Degrees Sciences Science Largest Yale Alumni Sports league expanded
Ranking John River Game Museums Main David Arts York United High William
Professional ms 42 Museum teams Wikipedia New Arts hockey Princeton Brown
Foundation dated Library October Association September Center General Teaching
Identifiers Study w School ha Graduate Century Universities ISBN class Top Office
South system links News Mens Institute Cambridge First Times Harvards 2009 million
Great among Large 18 NCAA colleges 14 also academic First 1 August 28 5 State potentially
Admission President Higher Research 11 2012 England men Rankings Carnegie
Faculty Program 85 Early 9 Education Retrieved M Crimson States Allston 2004 Stadium
cheating containing campus Students Hall Years 3 2010 Women Report Making
several Undergraduate 16 1636 Radcliffe 4 Public became National
two staff Business Team won Court Penn statements endowment Member
Columbia athletic Home Medicline references Cornell 2006 Rivalry throughout
Texas 2005 expansion Canadian 1926 1988 Americas Quincy
Western Curriculum

Enter a URL below, or paste some text

http://en.wikipedia.org/wiki/Harvard_University

Go!

Spiral: Archimedean Rectangular

Scale: $\log n$ \sqrt{n} n

Font: Impact

orientations from 0 ° to 0 °

Number of words: 250

One word per line

Download: [SVG](#) | [PNG](#)

President Bush, January 29, 2002

President Obama, January 25, 2011

<http://www.jasondavies.com/wordcloud/>

Spark Clouds

Lee 2010

Convey trends between multiple tag clouds over time

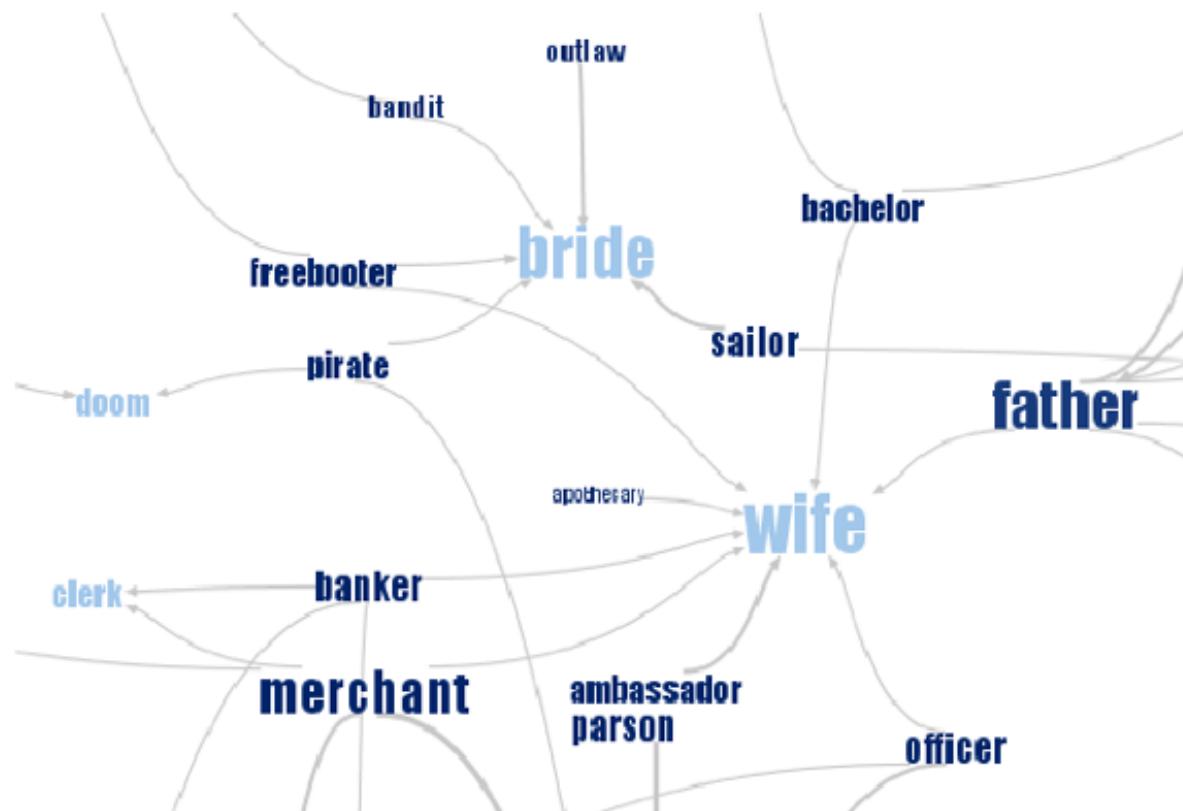


Example: Phrase Net

van Ham et al. 2009

Visual overviews of unstructured text

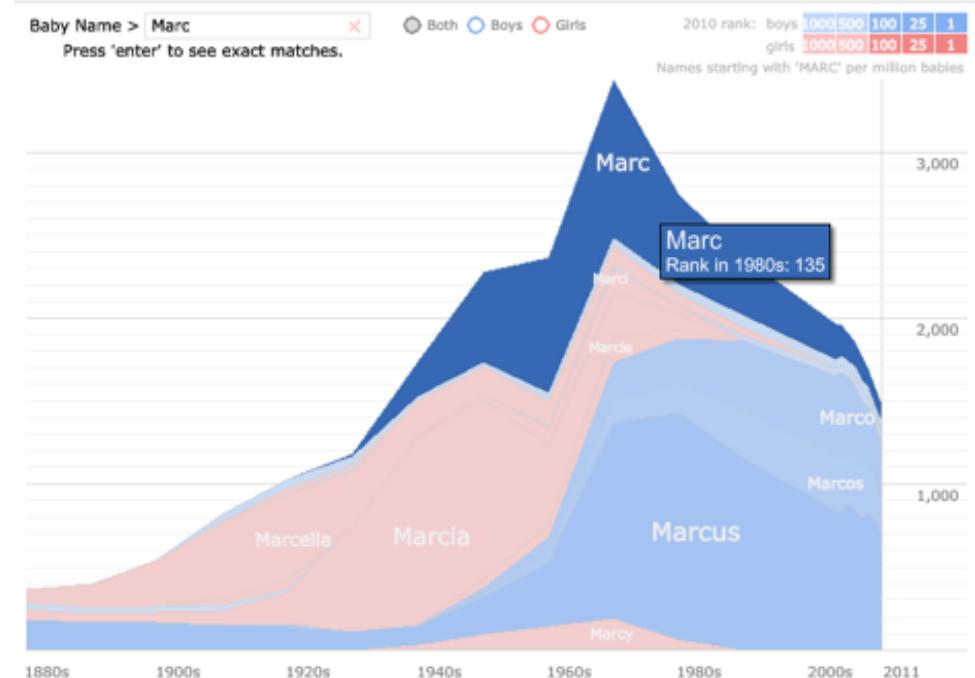
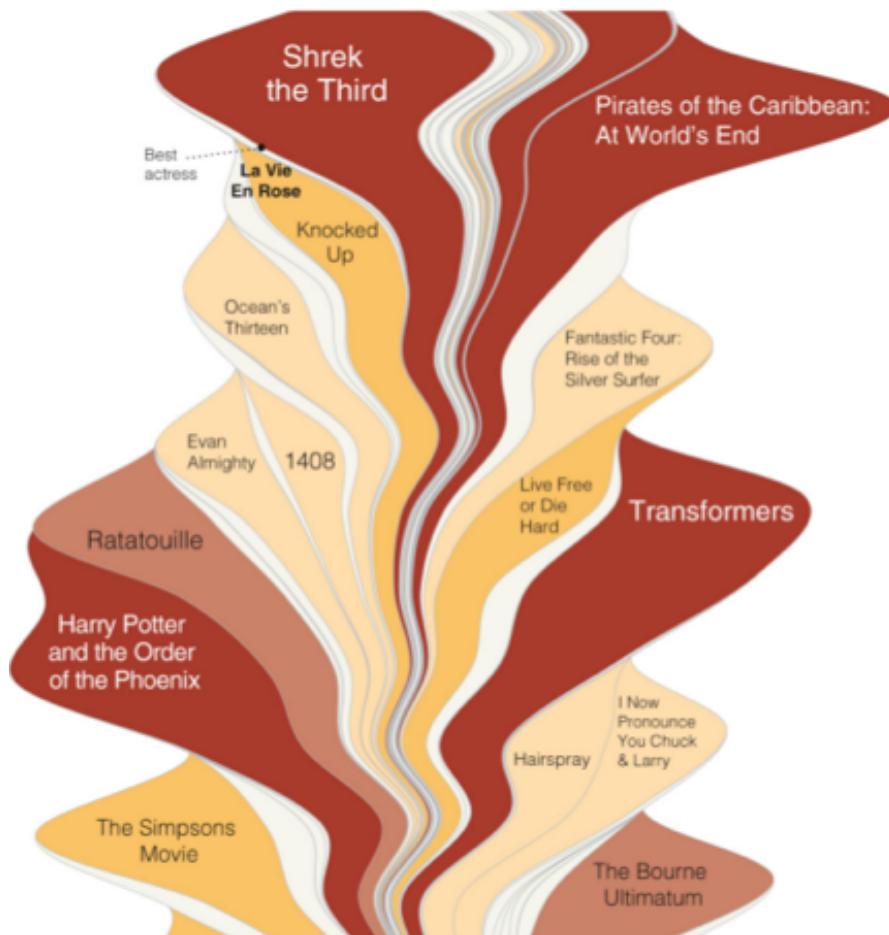
Graph; nodes = words; edges = linked by user specified relation



Relations in 18th and 19th century novels

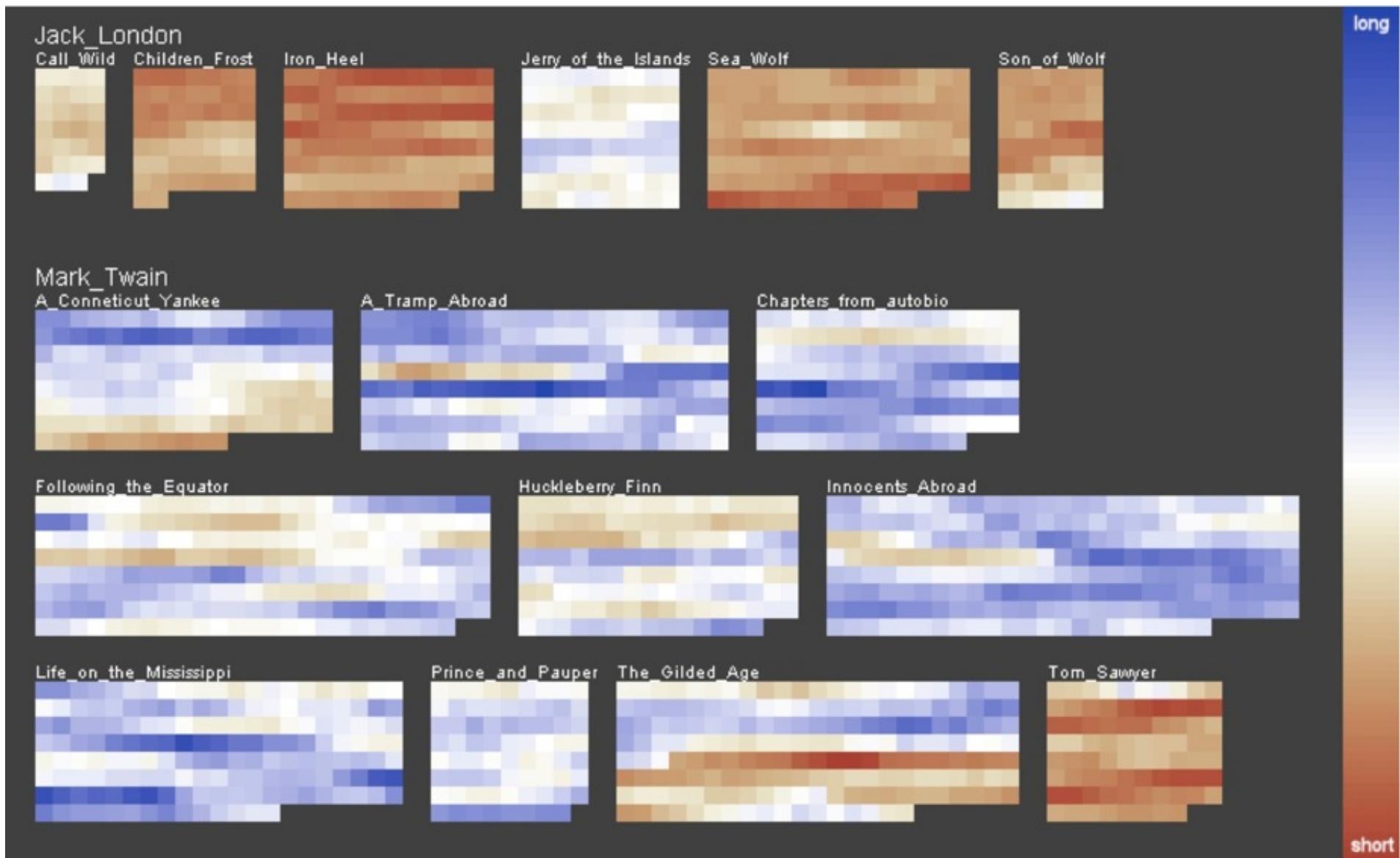
Theme River / Stream Graph

Thematic changes over time. Height = frequency



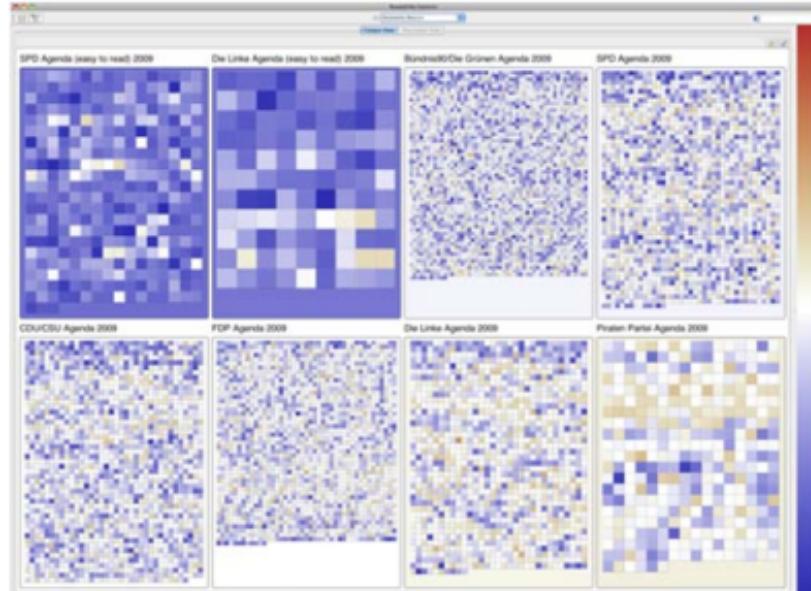
Baby Name Voyager
<http://www.babynamewizard.com/>
Wattenberg 2005

Example: Literature Fingerprinting

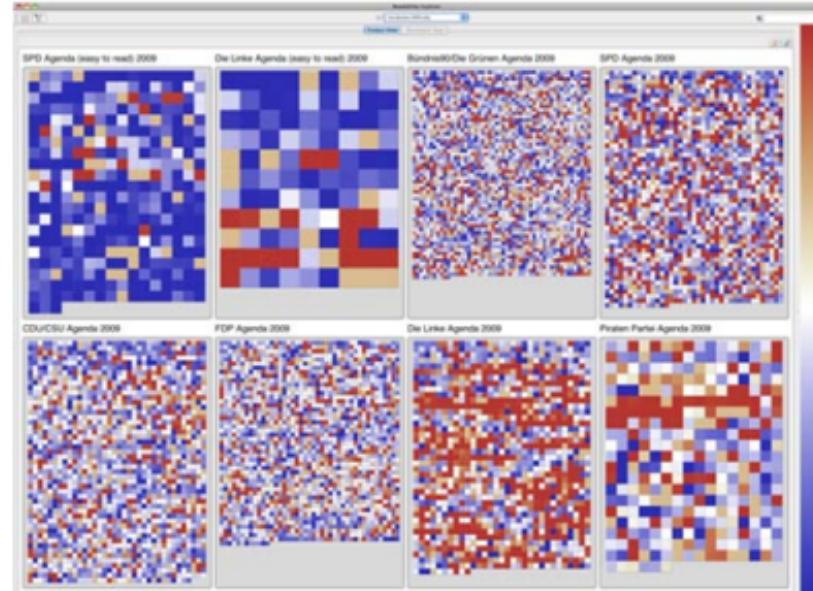


Visualize several text measures to discriminate between authors.
Pixel = text block, Group = book, color = average sentence length

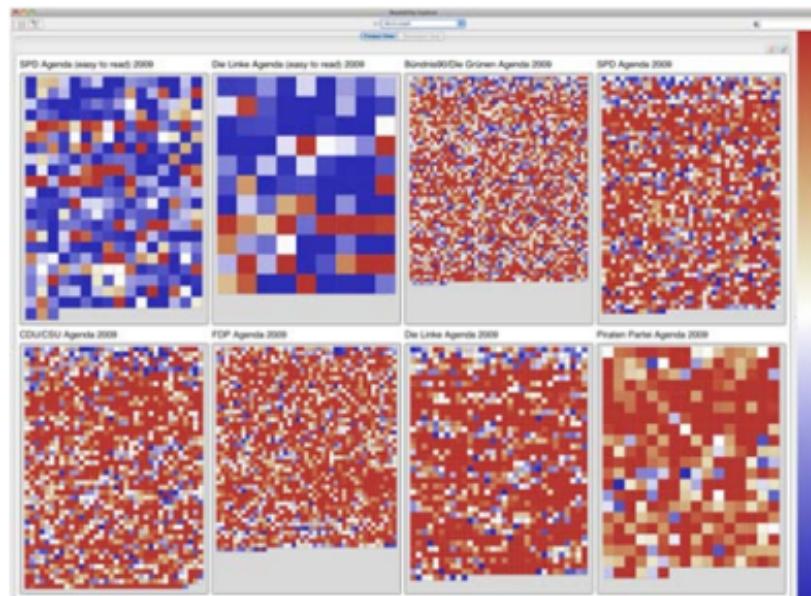
Example: Readability Analysis



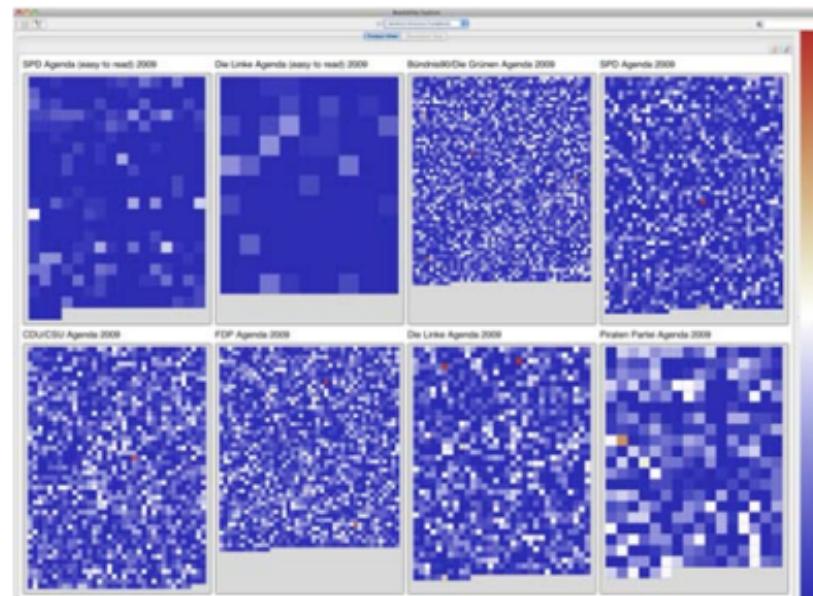
(a) Average Readability Score



(b) Feature: Vocabulary Difficulty



(c) Feature: Word Length



(d) Feature: Sentence Structure Complexity

Summary

Major Concepts:

- Data mining Terminology
- Visualization basics
- Graphical Integrity
- Graph Types for 2D and nD

Slide Material References

- Slides from Harvard CS 109 (2013 and 2014)