Presentation of Results

Philip J. Cwynar

University of Pittsburgh

School of Information Sciences

Overview

- Introduction
- Exploratory Data Analysis
- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Data Visualization software
- Bad Examples
- Tufte's design principles for visualization of information
- In Class Videos

Introduction

What is the leading thought here?

“The purpose of computing is insight, not numbers”

Richard Hamming
(preface to his 1962 book on numerical methods)
http://en.wikipedia.org/wiki/Richard_Hamming

What is data visualization?

The basic objective of data visualization is to provide an efficient graphical display for summarizing and reasoning about quantitative information. 2011 SAGE Publications **Data Analysis, Exploratory**

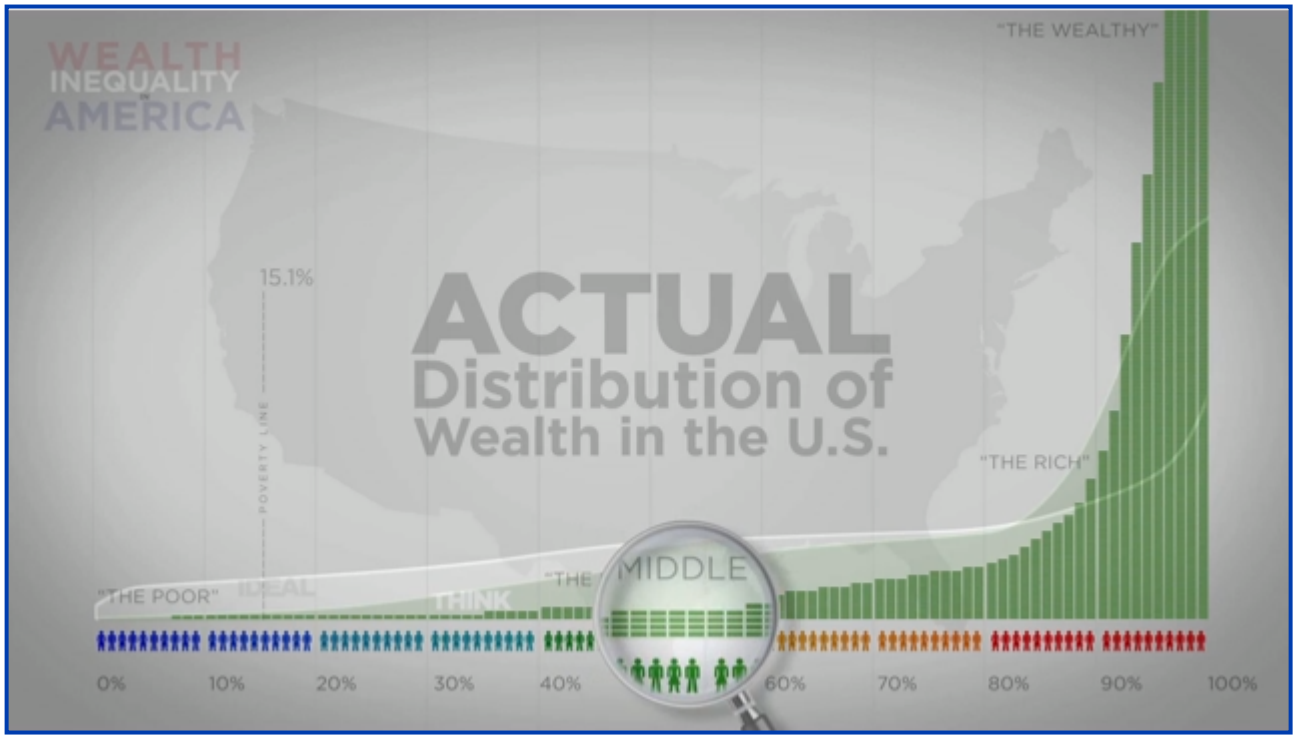
Data visualization is both an art and a science.

https://en.wikipedia.org/wiki/Data_visualization

Motivational videos: Great presentations of quantitative information

http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen.html (19'54")
http://www.ted.com/talks/lang/en/hans_rosling_on_global_population_growth.html (10'04")
http://www.ted.com/talks/hans_rosling_reveals_new_insights_on_poverty.html (19'01")
http://www.ted.com/talks/hans_rosling_the_good_news_of_the_decade.html (15'35")
http://www.ted.com/talks/hans_rosling_religions_and_babies.html (13'20")
http://www.ted.com/talks/hans_rosling_the_truth_about_hiv.html (10'0")
http://www.ted.com/talks/hans_rosling_asia_s_rise_how_and_when.html (15'51")

Motivational videos: Great presentations of quantitative information



http://www.youtube.com/watch?v=QPKKQnijnsM (6'24")

Motivational videos: Great presentations of quantitative information

Other cool presentations involving visualization of information:

http://www.ted.com/talks/lang/en/anders_ynnerman_visualizing_the_medical_data_explosion.html (16'37")
http://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization.html (18'17")

Exploratory Data Analysis

Exploratory Data Analysis

John W. Tukey, the definer of the phrase exploratory data analysis (EDA)

"'Exploratory data analysis' is an attitude, a state of flexibility, a willingness to look for those things that we believe are not there, as well as those we believe to be there."

2011 SAGE Publications **Data Analysis, Exploratory**

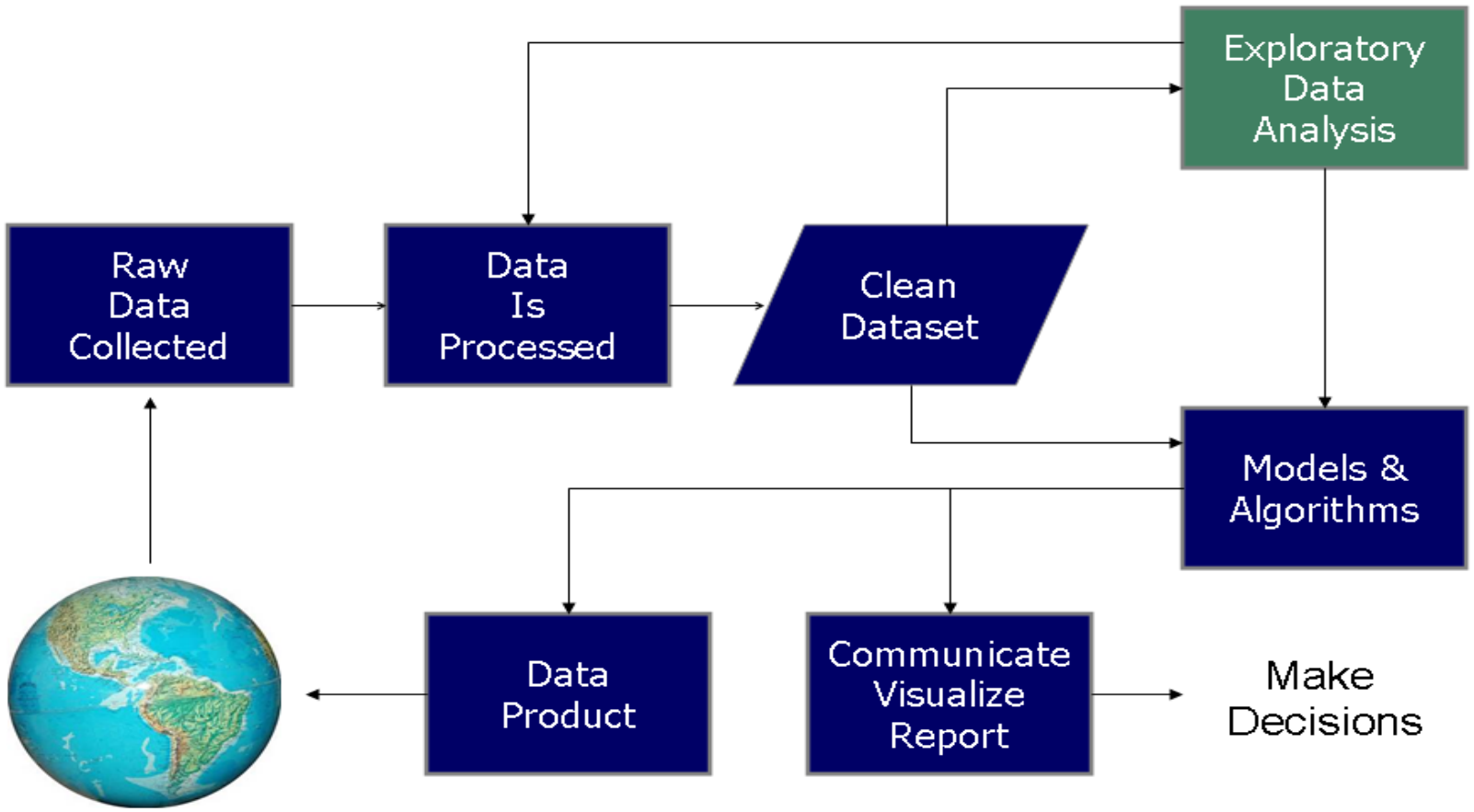
Exploratory Data Analysis

John W. Tukey

"I hope that I have shown that exploratory data analysis is actively incisive rather than passively descriptive, with real emphasis on the discovery of the unexpected." (p. lxii)

"Exploratory data analysis isolates patterns and features of the data and reveals these forcefully to the analyst." (Hoaglin, Mosteller, & Tukey,1983, p. 1)

Data Science Process



https://en.wikipedia.org/wiki/Exploratory_data_analysis

Descriptive Statistics

Basic statistics

Excel

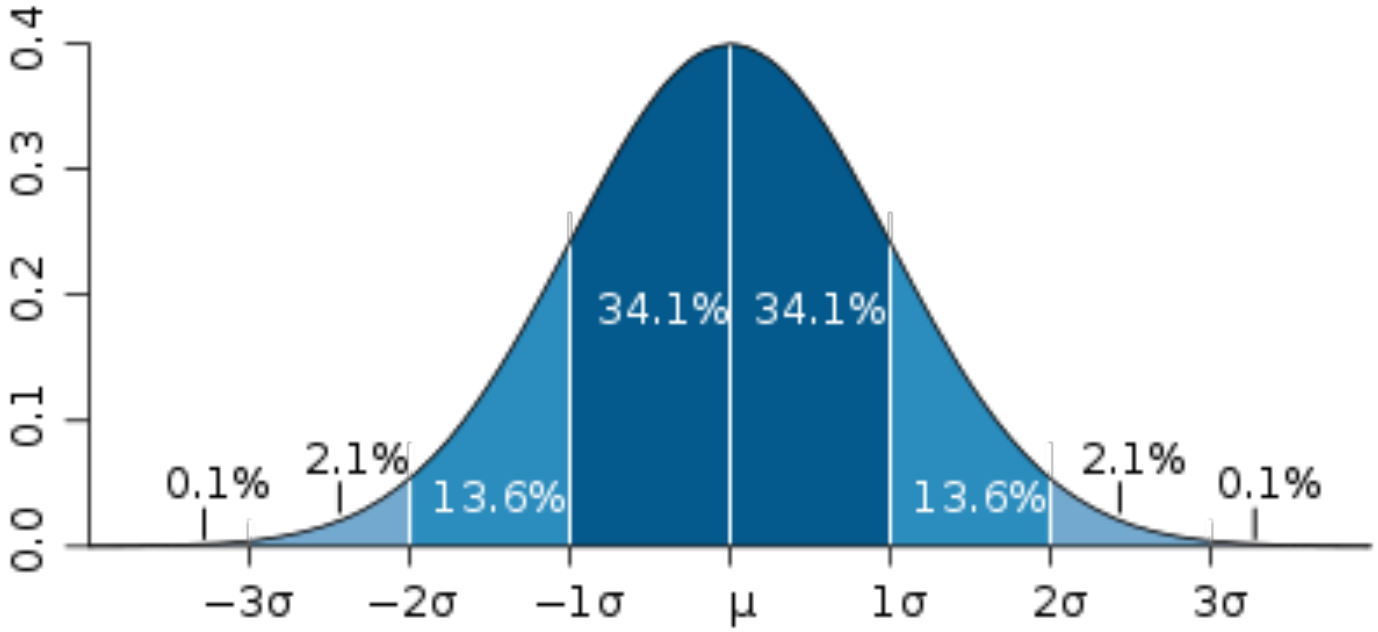
Apgra	
Mean	56.721 07647
Median	55.708 5
Mode	72
Standard Deviation	18.077 09676
Variance	32 6.78 14274
Kurtosis	-0.554450128
Skewness	0.0891 85832
Range	76.5
Minimum	18.75
Maximum	95.25
Sum	96 42.5 83
Count	17 0

GeNIe

	Mean	Variance	StdDev	Min	Max	Count
spend	10974.5	3.02507e+007	5500.07	4125	35863	170
apret	56.7211	326.781	18.0771	18.75	95.25	170
top10	38.4588	547.859	23.4064	8	98	170
rejr	30.6542	292.345	17.0981	0	84.067	170
tstsc	66.1642	48.6549	6.97531	48.125	87.5	170
pacc	43.1731	171.746	13.1052	8.964	76.253	170
strat	16.0865	16.0521	4.0065	7.2	29.2	170
salar	61357.6	9.60946e+007	9802.79	38640	87900	170

Probability distribution

Expresses the relative probabilities of different values taken by a random variable



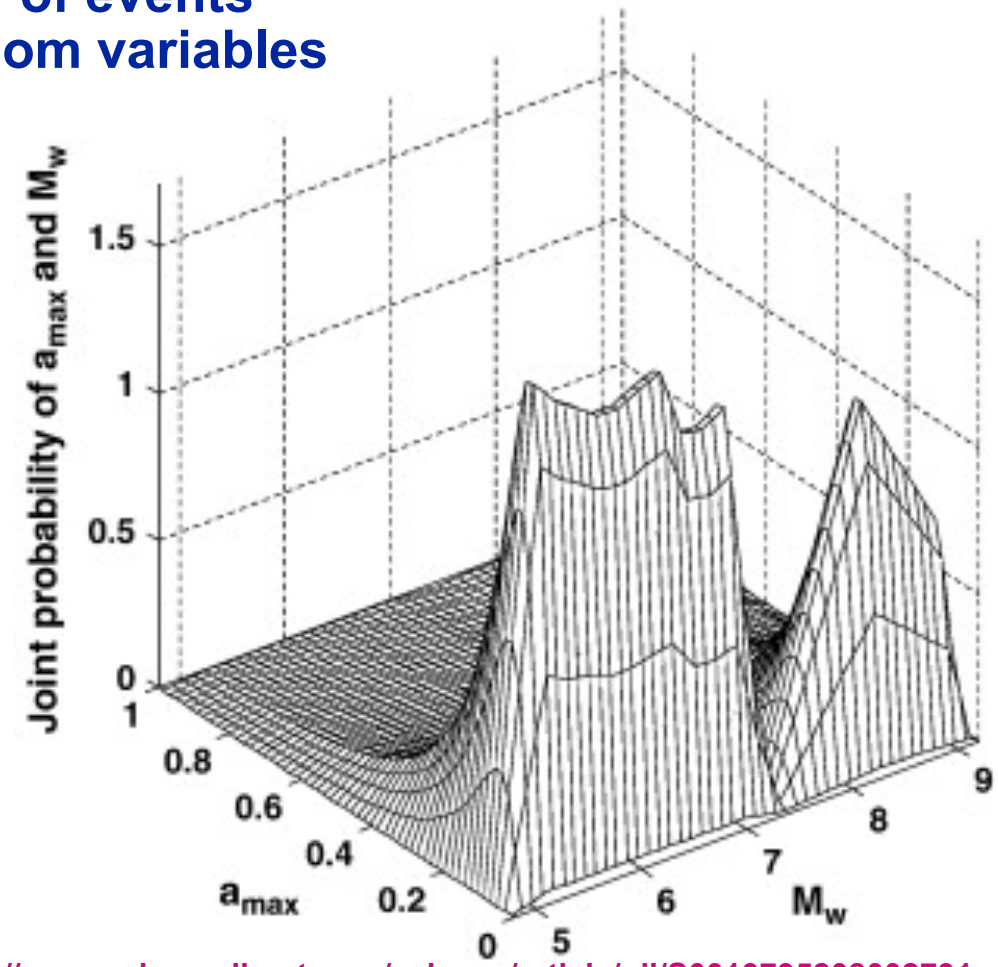
Source: http://en.wikipedia.org/wiki/Probability_distribution

e.g., grade distribution in a university course

Joint probability distribution

Expresses the probability of events defined over several random variables

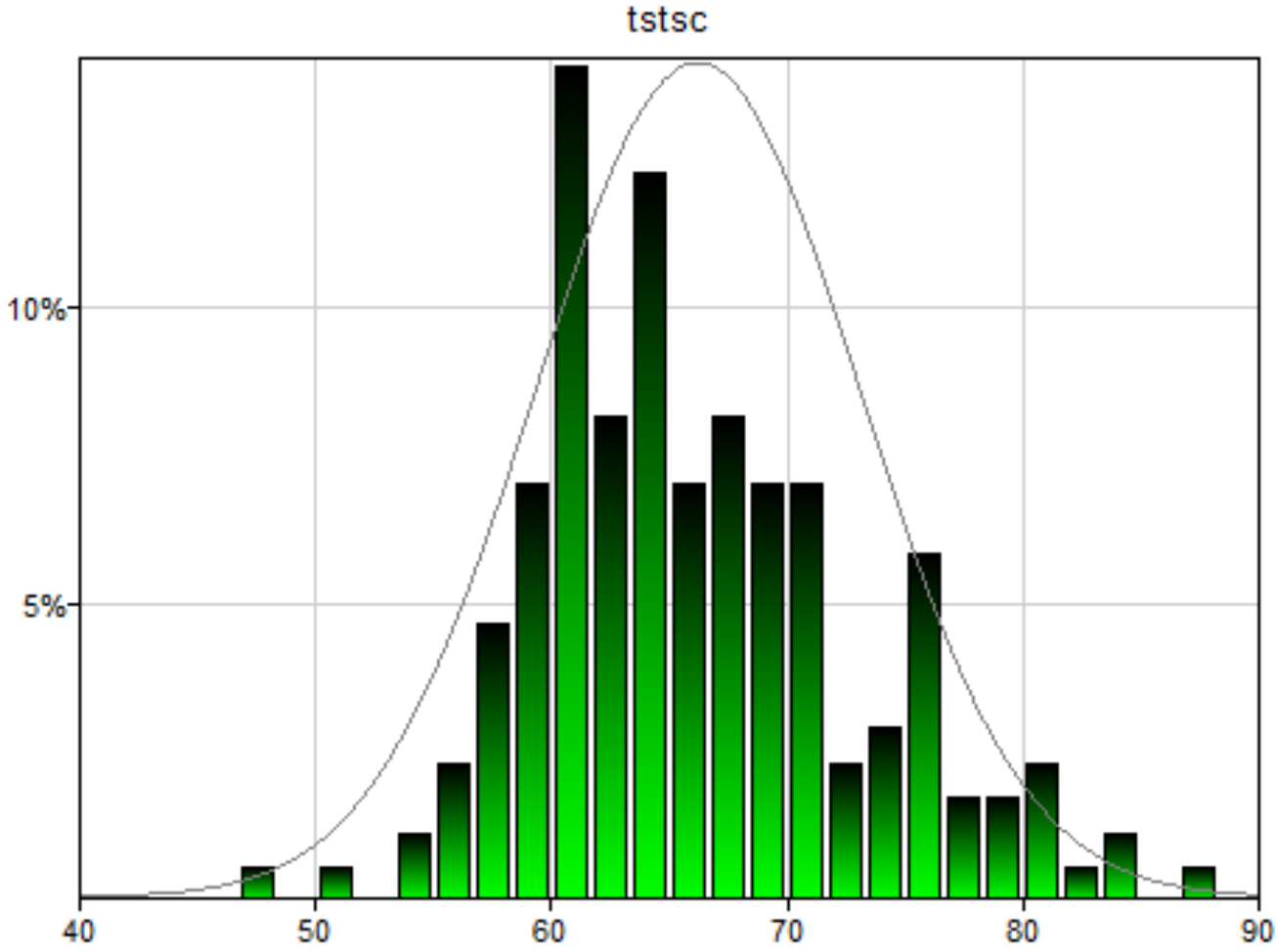
e.g., probability distribution over grades and the amount of work in a university course



Source: http://www.sciencedirect.com/science/article/pii/S0013795208002731

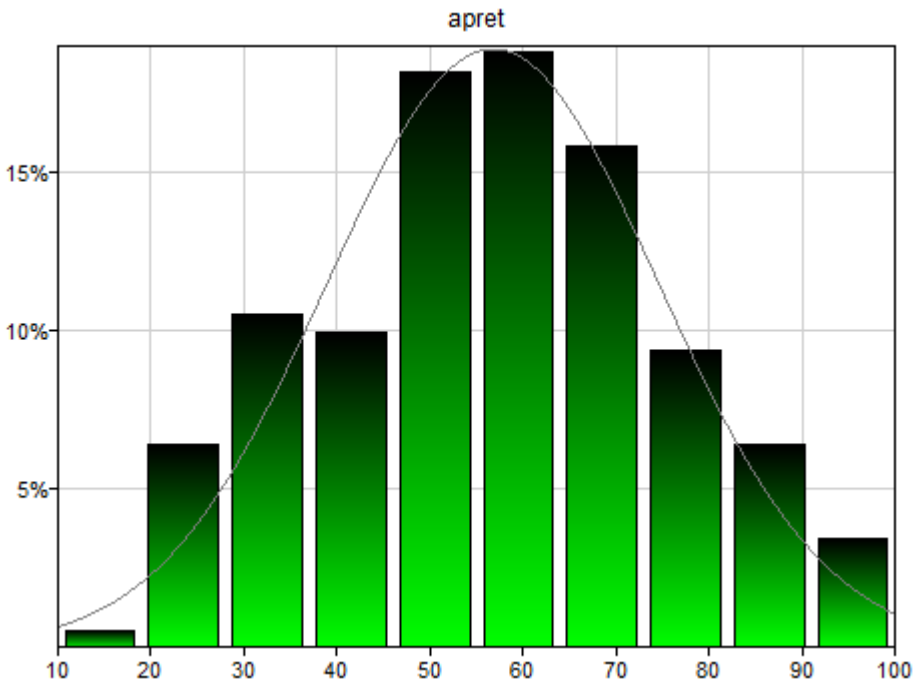
Histograms

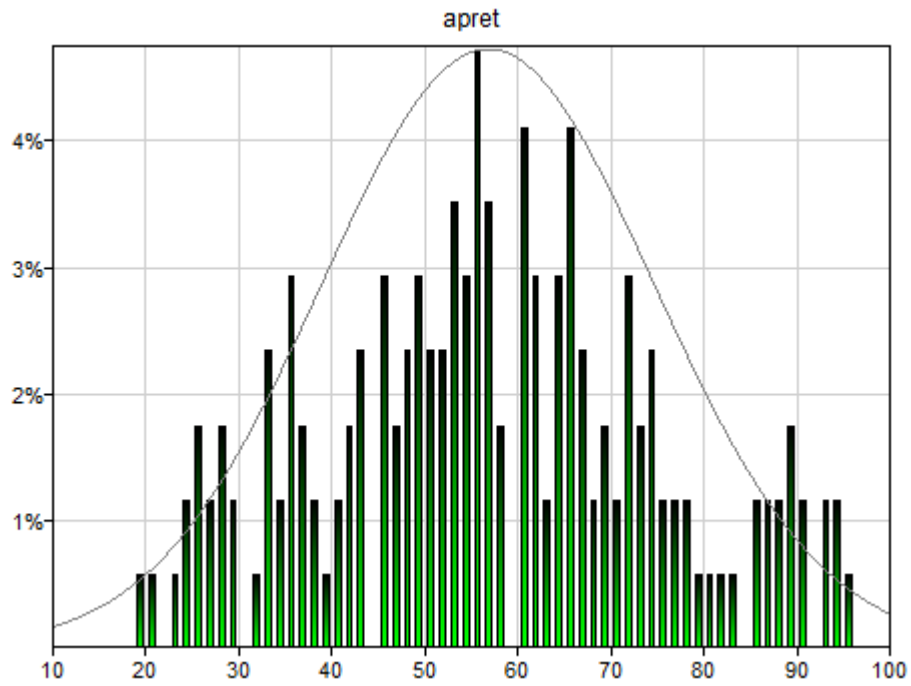
Values that a variable takes in a data set can be seen very nicely on plots called "histograms"



- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

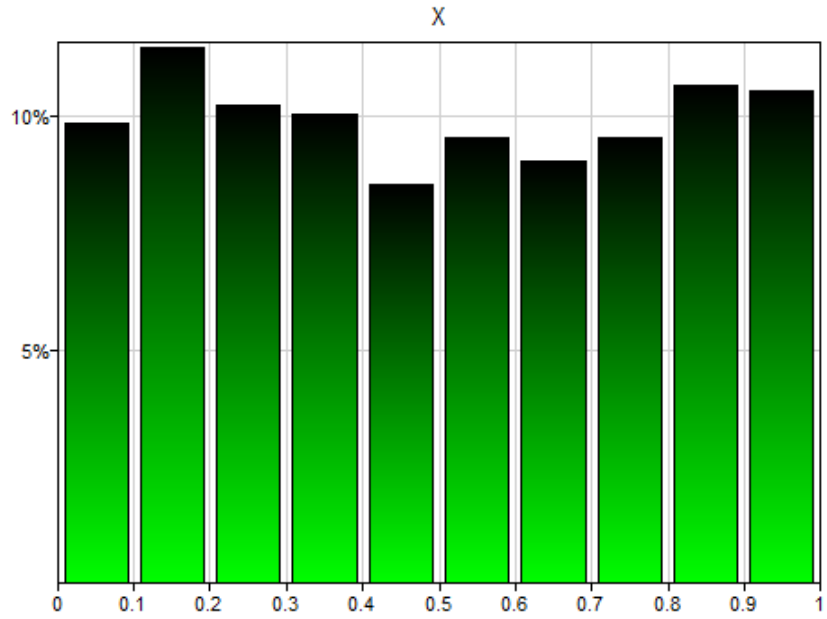
Histograms

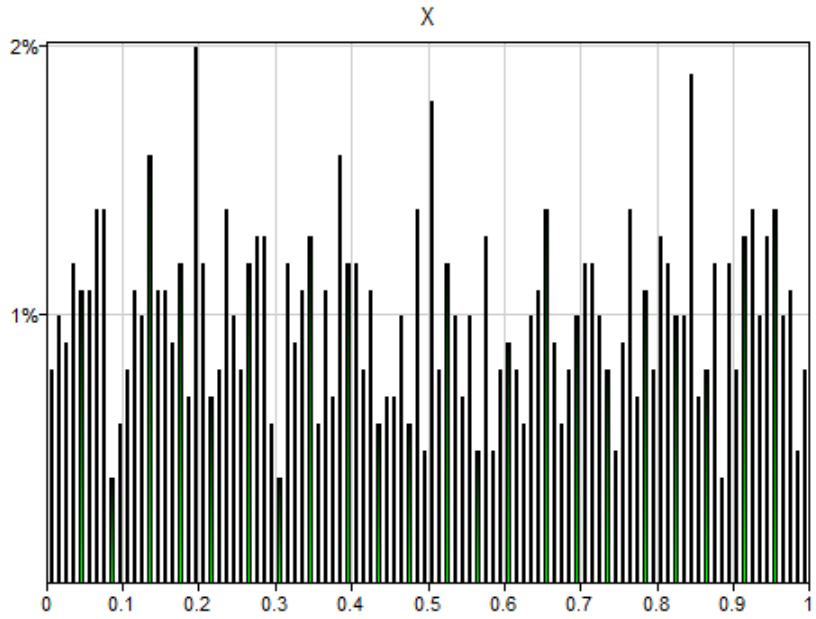




Bin size affects the form, good bin size is essentially an art: I'm not aware of any research on automatic selection of bins. I am aware of at least one computer program that does it right (see http://genie.sis.pitt.edu/).

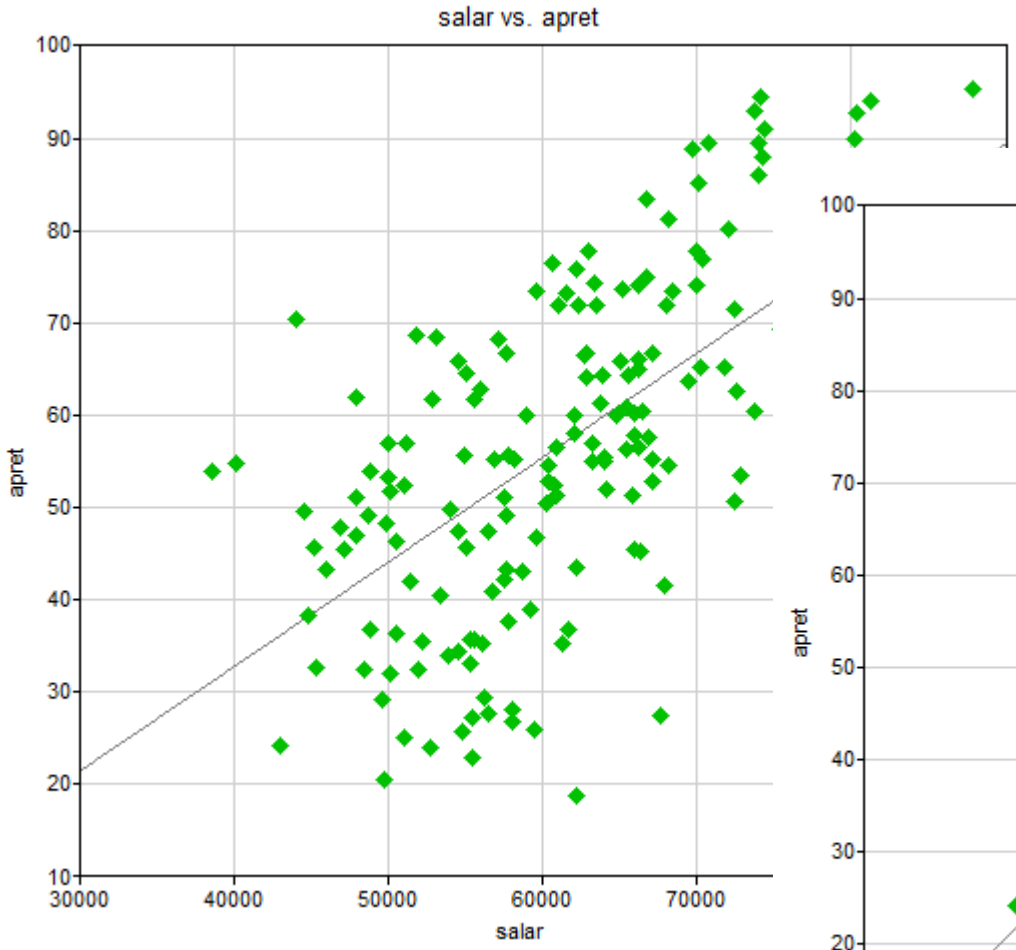
Histograms

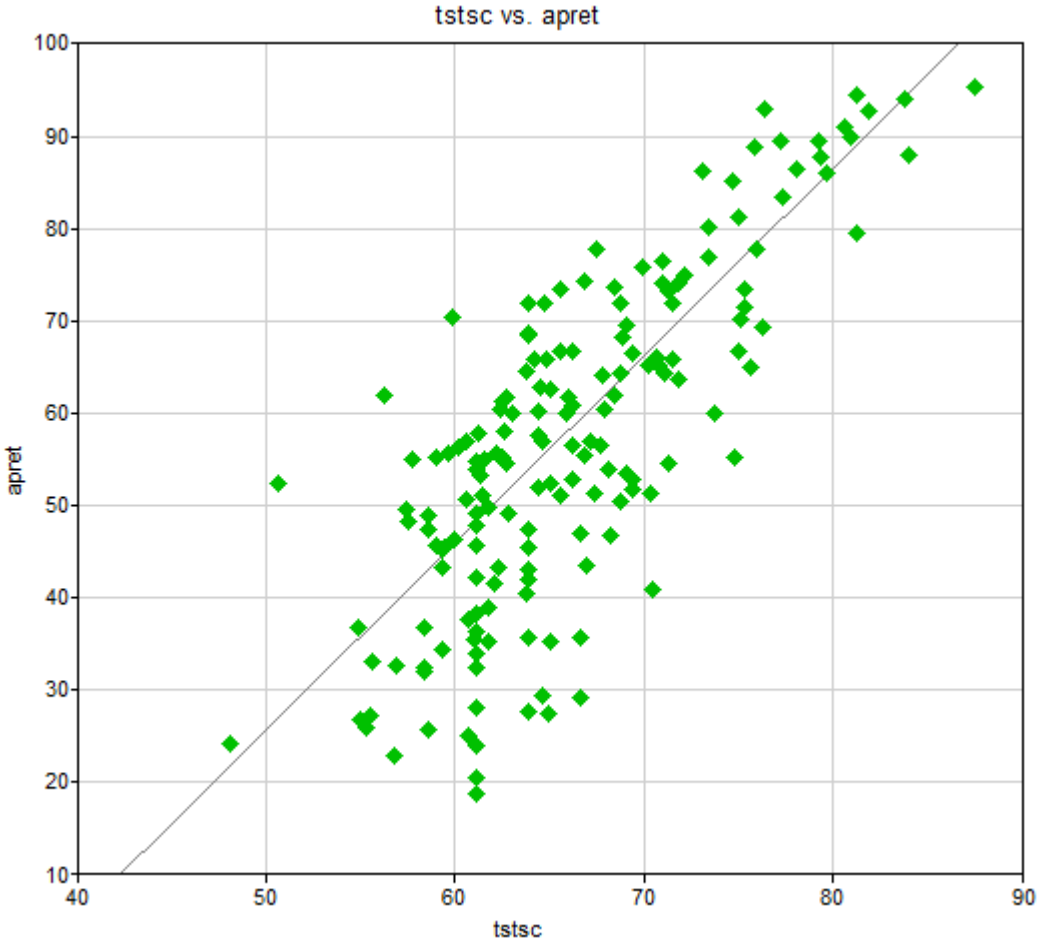




The effect of bin size is not that strong in case of some distributions (here: uniform distribution).

Scatter plots



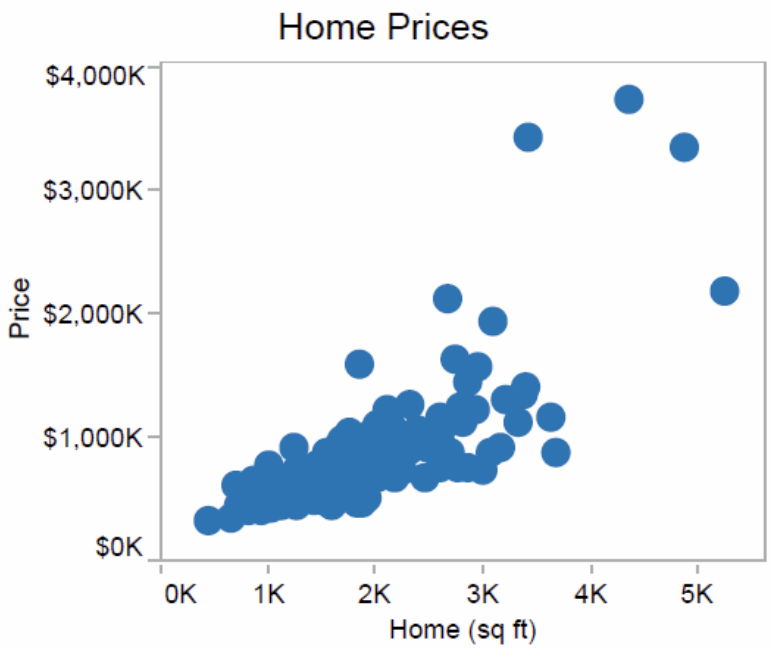


Plots of data known as *scatter plots* give an idea of the joint probability distribution between two variables.

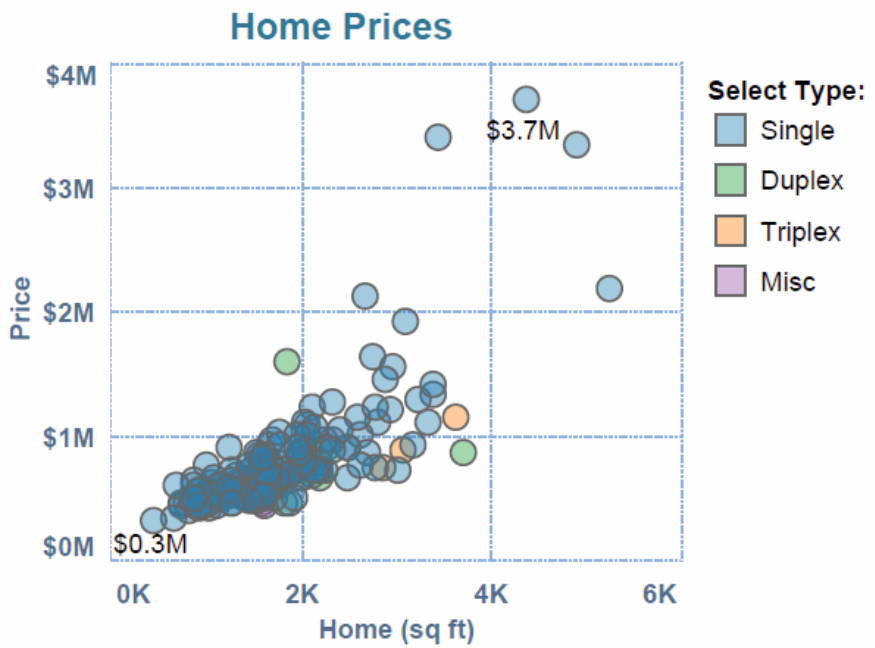
- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

Scatter plots

Good visualization



Great visualization



Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful," http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

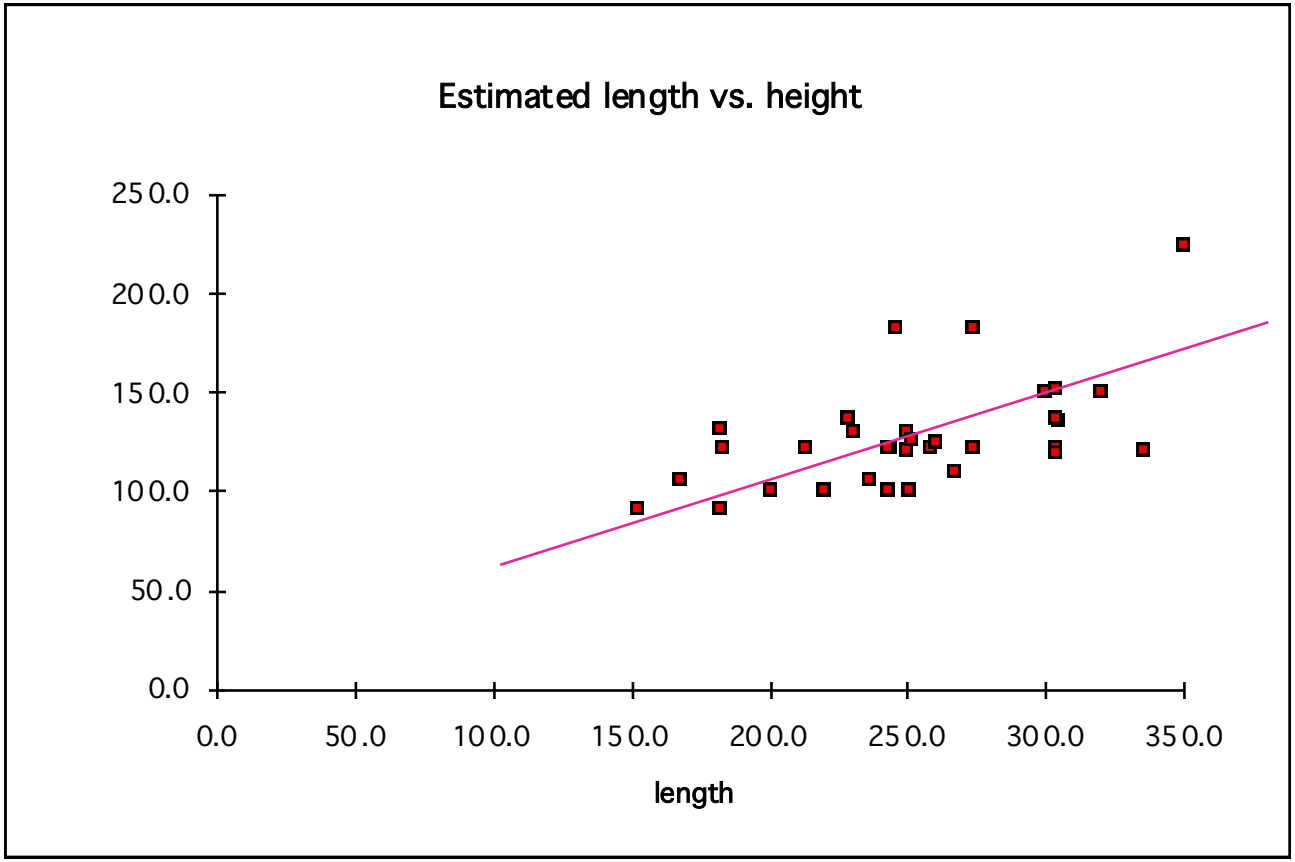
Correlation matrix

	spend	apret	top10	rejr	tstsc	pacc	strat	salar
spend	-							
apret	0.601231	-						
top10	0.675656	0.642464	-					
rejr	0.633544	0.514958	0.643163	-				
tstsc	0.71491	0.782183	0.798807	0.628601	-			
pacc	-0.23673	-0.302834	-0.207505	-0.0715207	-0.164223	-		
strat	-0.561755	-0.458311	-0.247857	-0.283617	-0.465226	0.131858	-	
salar	0.711838	0.635852	0.637648	0.606777	0.715472	-0.37524	-0.347673	-

A **correlation matrix** is used to investigate the dependence between multiple variables at the same time. The result is a table containing the **correlation coefficients** between each variable and the others.

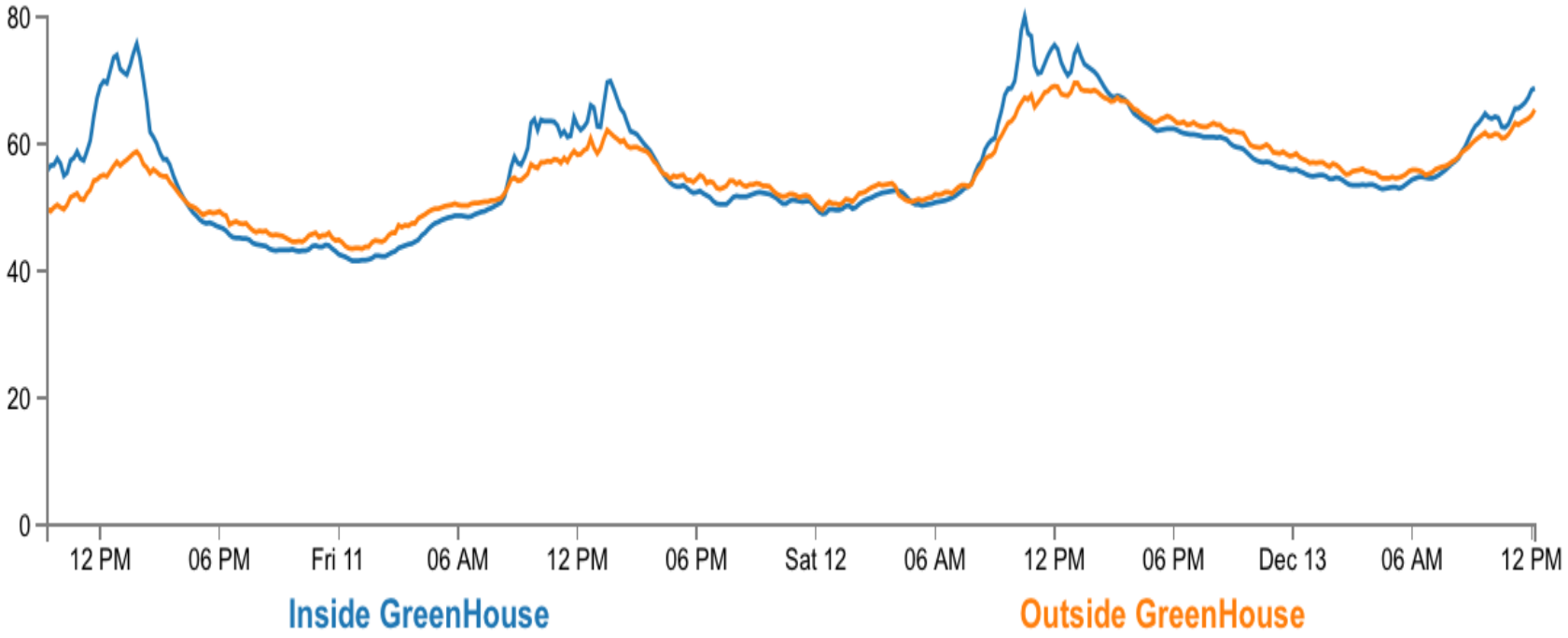
Linear regression

We fit a line to the data, the line equation is *y=a+bx*



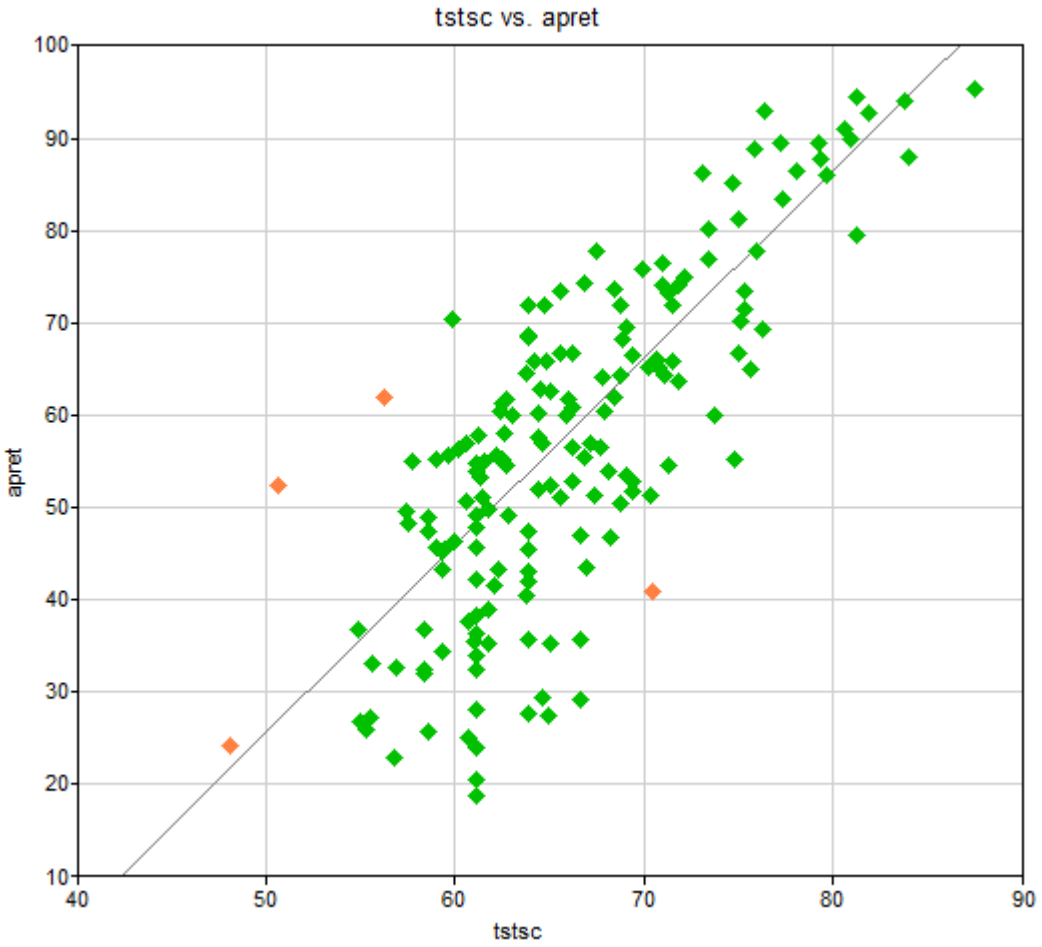
Time series

- **Measurements of variables that vary over time.**
- **This is often a matter of assumption: regular, static data also vary over time but we assume that they do not.**

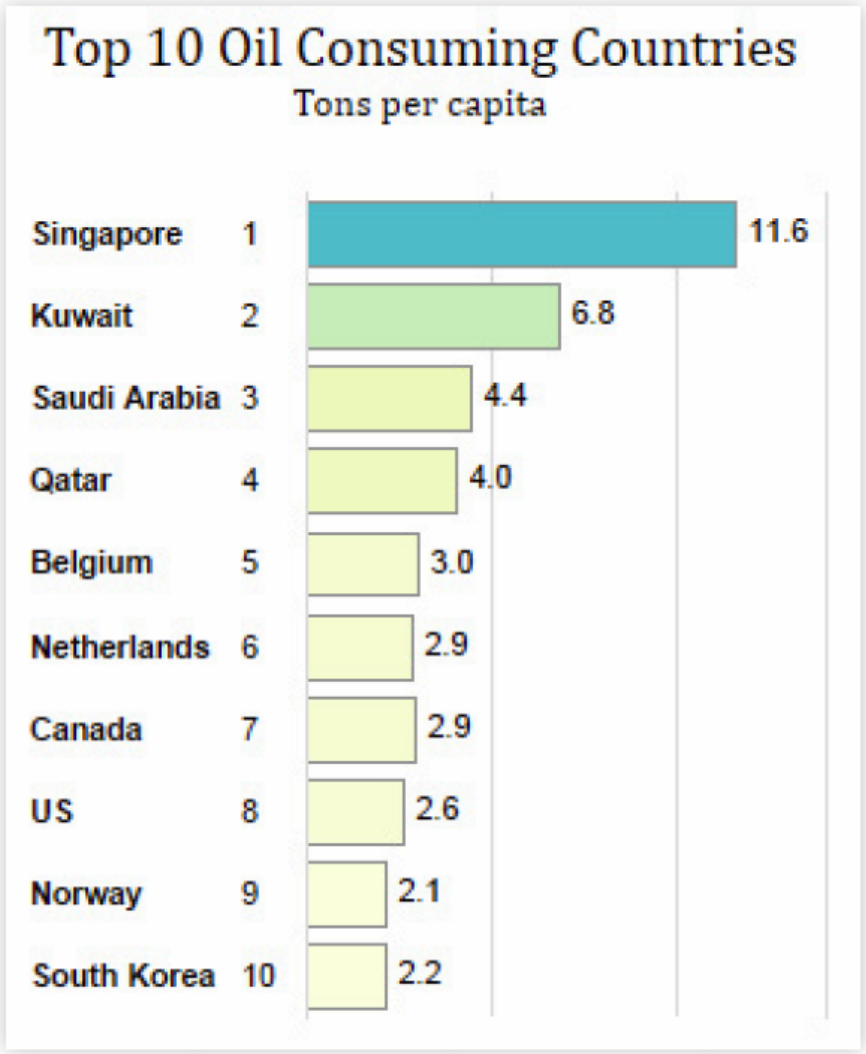


Outliers

- Values that come about because of errors in measurements, transcription, etc., or because of momentary failure in our assumptions.
- We remove them because they are potentially violating our assumptions.
- How to distinguish them? Typically done "manually." Visual inspection is usually very helpful.



Bar charts



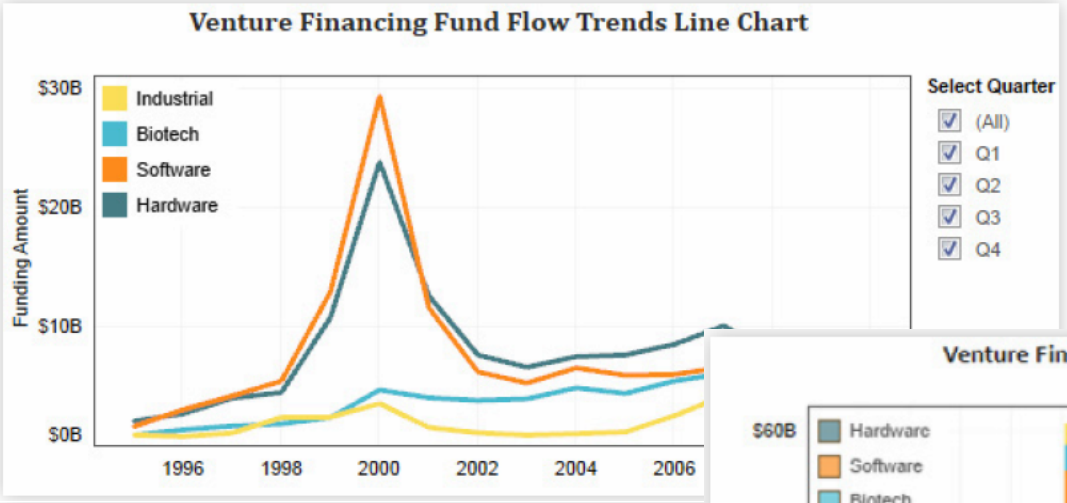
Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful,"
http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

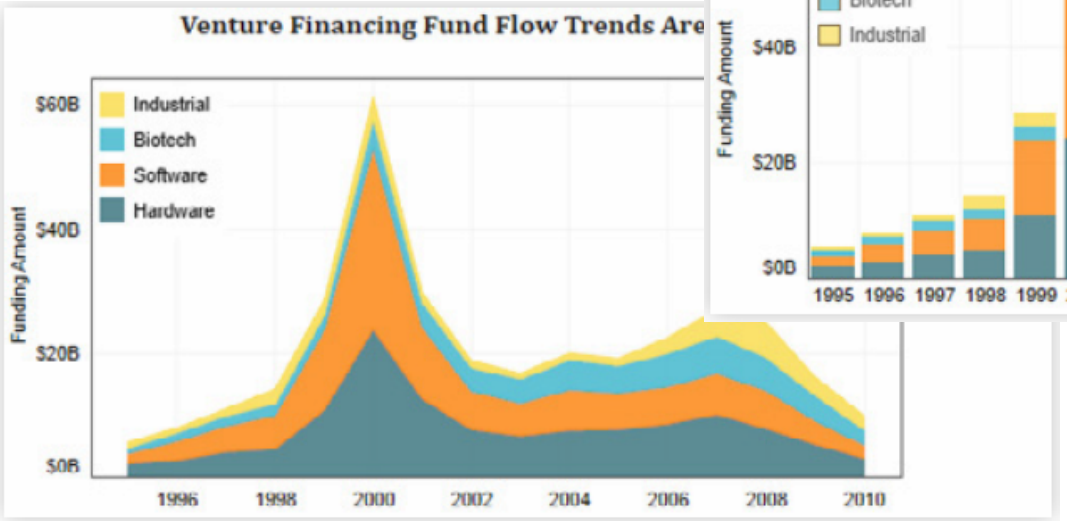
Trends and Temporal Data

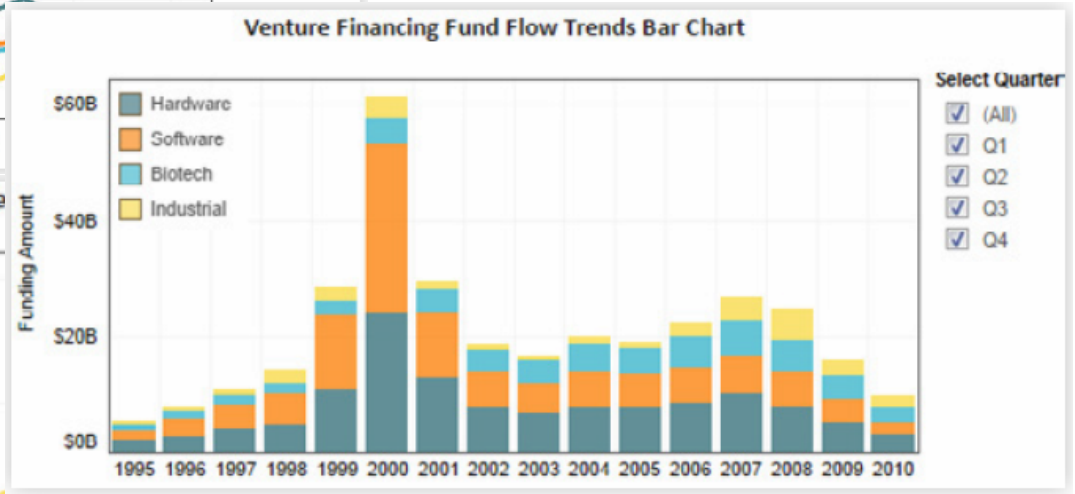
- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

Temporal data: Example

Various ways of showing the same information



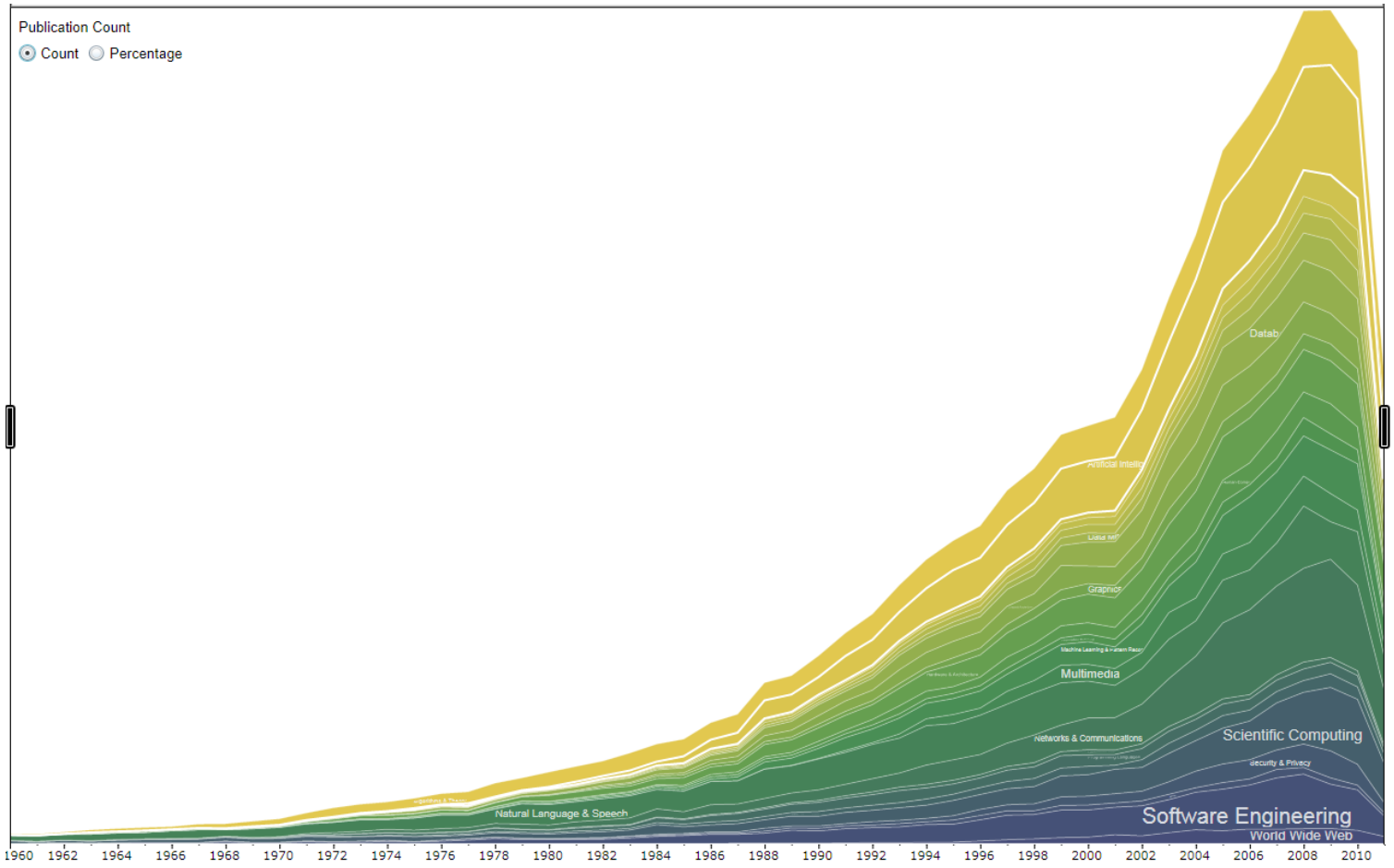




Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful,"
http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

- Descriptive statistics
- **Trends and temporal data**
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

Trends: Example

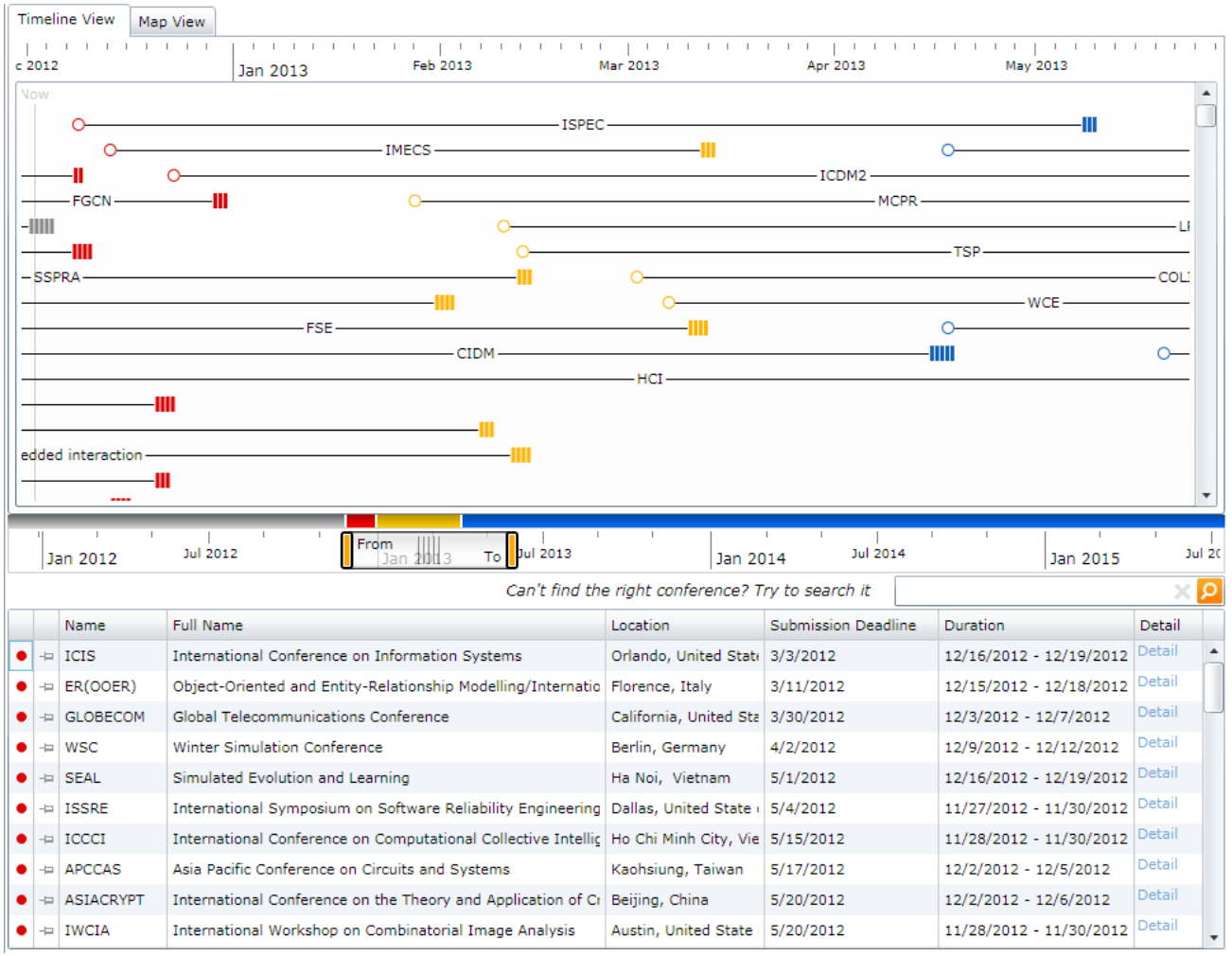


Microsoft Academic Search: Domain trends

http://academic.research.microsoft.com/

- Descriptive statistics
- **Trends and temporal data**
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

Temporal relationships: Example



Name	Full Name	Location	Submission Deadline	Duration	Detail
ICIS	International Conference on Information Systems	Orlando, United Stat	3/3/2012	12/16/2012 - 12/19/2012	Detail
ER(OOER)	Object-Oriented and Entity-Relationship Modelling/Internatio	Florence, Italy	3/11/2012	12/15/2012 - 12/18/2012	Detail
GLOBECOM	Global Telecommunications Conference	California, United Sta	3/30/2012	12/3/2012 - 12/7/2012	Detail
WSC	Winter Simulation Conference	Berlin, Germany	4/2/2012	12/9/2012 - 12/12/2012	Detail
SEAL	Simulated Evolution and Learning	Ha Noi, Vietnam	5/1/2012	12/16/2012 - 12/19/2012	Detail
ISSRE	International Symposium on Software Reliability Engineering	Dallas, United State	5/4/2012	11/27/2012 - 11/30/2012	Detail
ICCCI	International Conference on Computational Collective Intellig	Ho Chi Minh City, Vie	5/15/2012	11/28/2012 - 11/30/2012	Detail
APCCAS	Asia Pacific Conference on Circuits and Systems	Kaohsiung, Taiwan	5/17/2012	12/2/2012 - 12/5/2012	Detail
ASIACRYPT	International Conference on the Theory and Application of Cr	Beijing, China	5/20/2012	12/2/2012 - 12/6/2012	Detail
IWCIA	International Workshop on Combinatorial Image Analysis	Austin, United State	5/20/2012	11/28/2012 - 11/30/2012	Detail

Microsoft Academic Search: Scientific event calendar

http://academic.research.microsoft.com/

Graphical Representation of Relationships

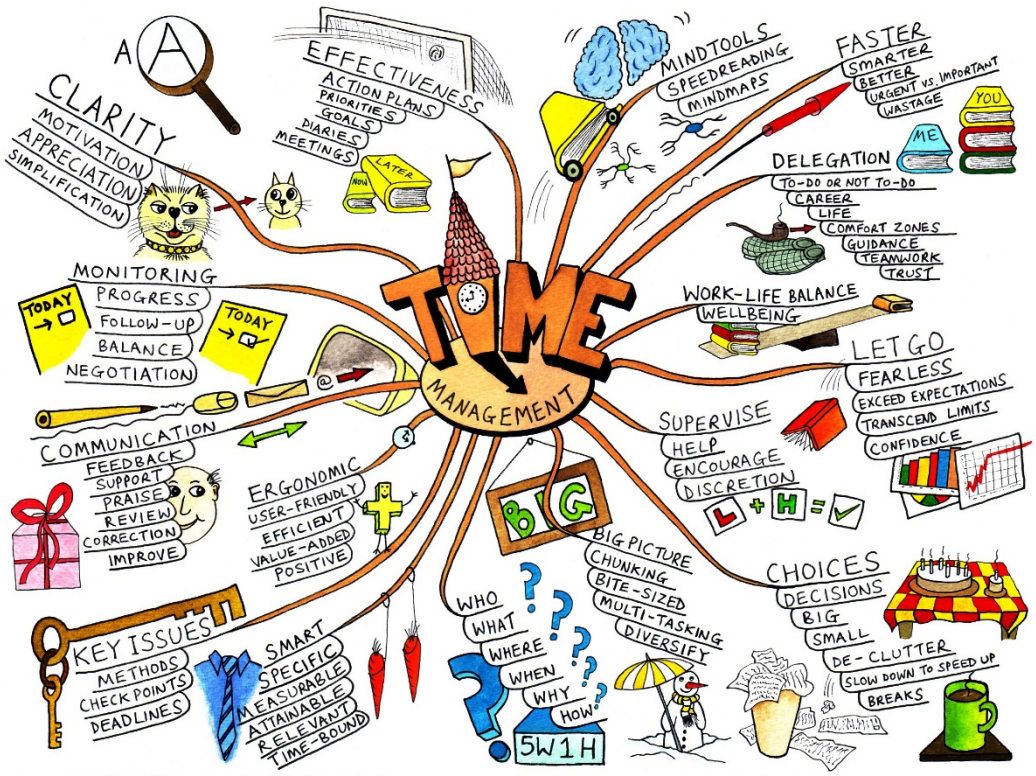
Mind maps

Diagrams used to visually outline information. Usually created around a single word or text, placed in the center, to which associated ideas, words and concepts are added.

Have usually the form of a snowflake: major categories radiate from a central node, minor categories radiate from the major categories, etc.

Categories represent ideas, concepts or other related items.

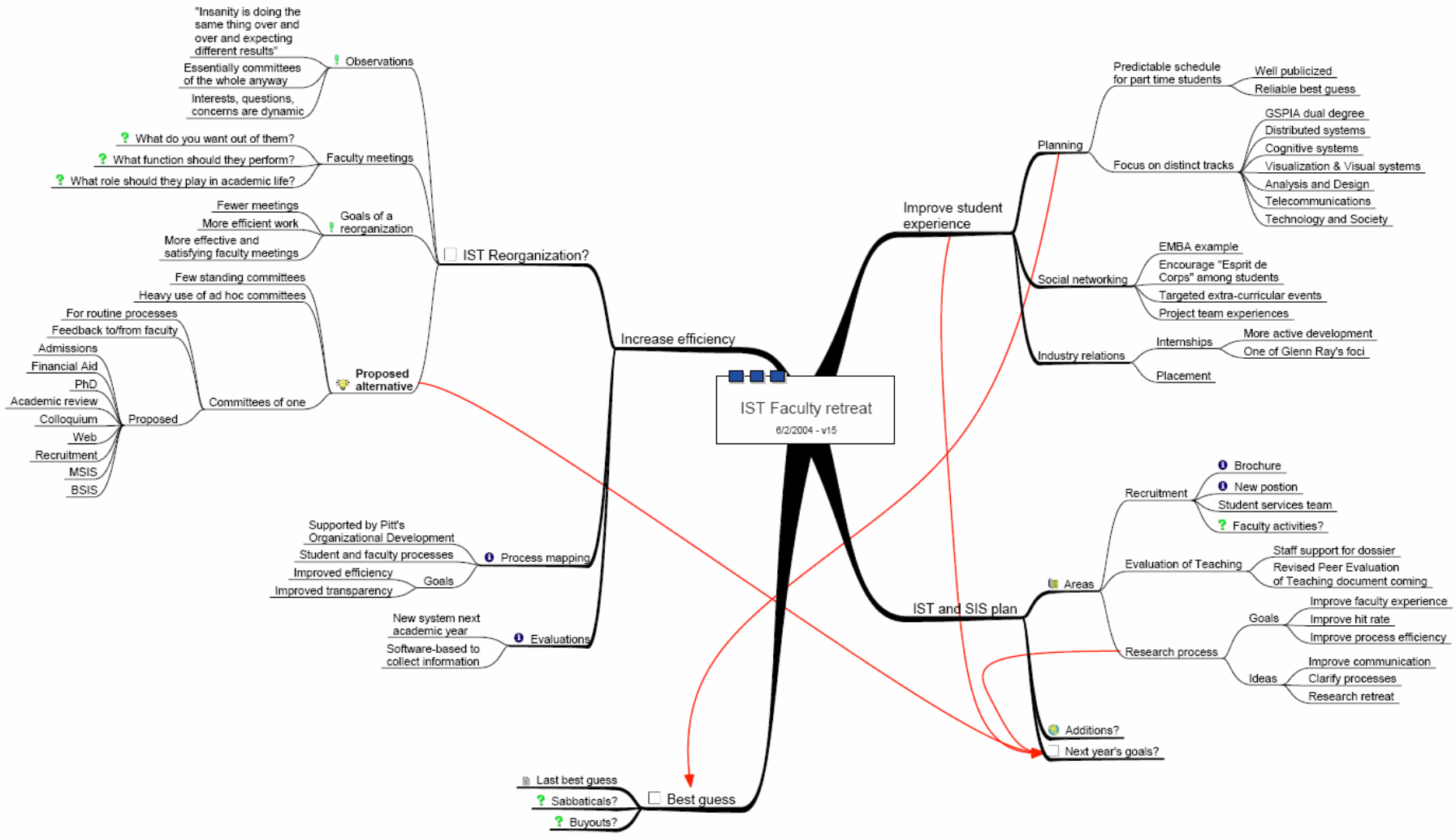
Date back to at least 3rd century AD
http://en.wikipedia.org/wiki/Mindmap



http://www.mindtools.com/pages/article/newISS_01.htm

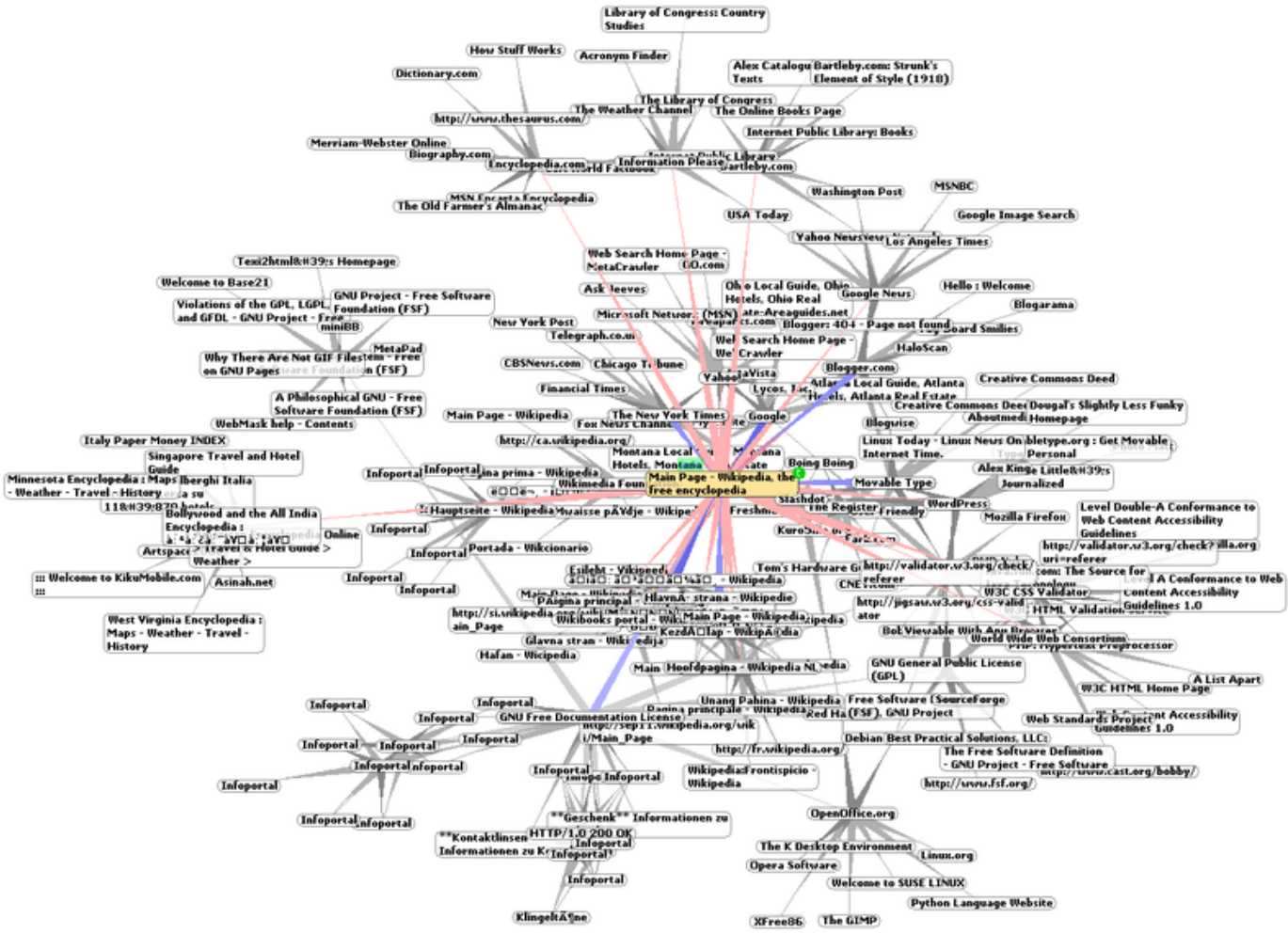
Mind maps: Example

- Descriptive statistics
- Trends and temporal data
- **Relationships**
- Spatial data
- Social networks
- Tufte's design principles



- Descriptive statistics
- Trends and temporal data
- **Relationships**
- Spatial data
- Social networks
- Tufte's design principles

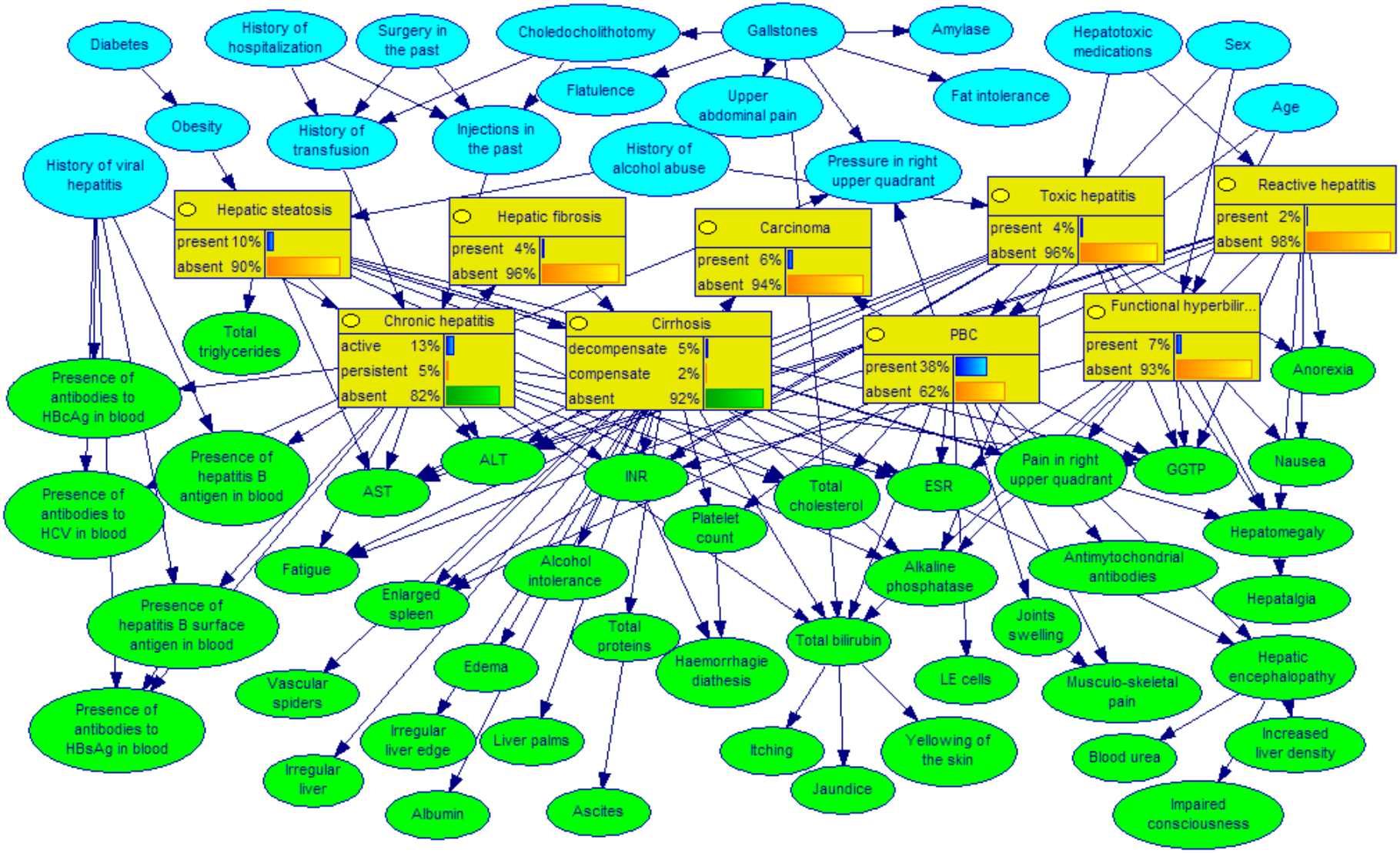
World Wide Web: Example



A data visualization of Wikipedia as part of the World Wide Web, demonstrating hyperlinks

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

Relationships among variables (causal relationships)



Spatial Data

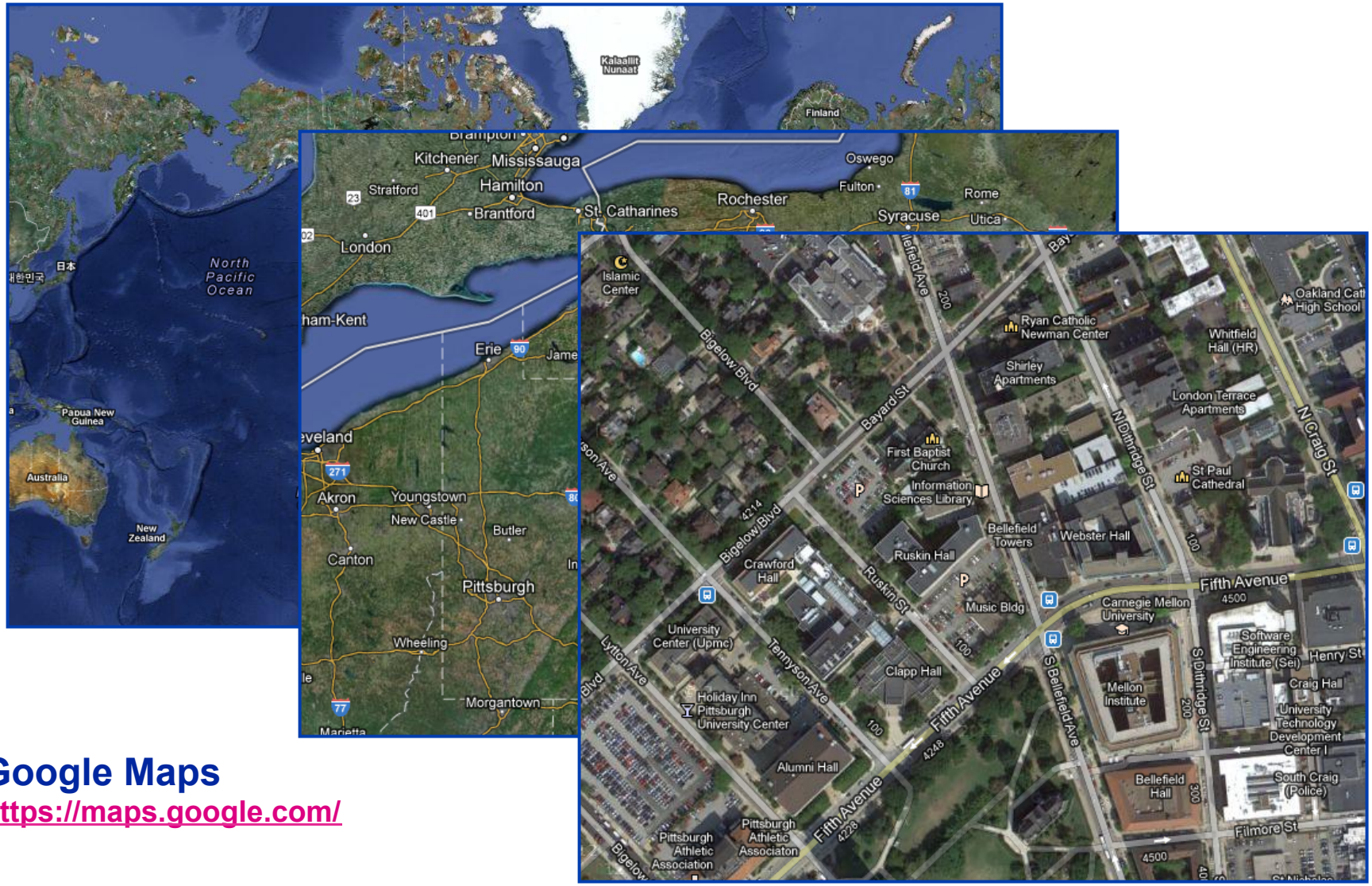
- Descriptive statistics
- Trends and temporal data
- Relationships
- **Spatial data**
- Social networks
- Tufte's design principles

Spatial data: Example

Google Maps

https://maps.google.com/

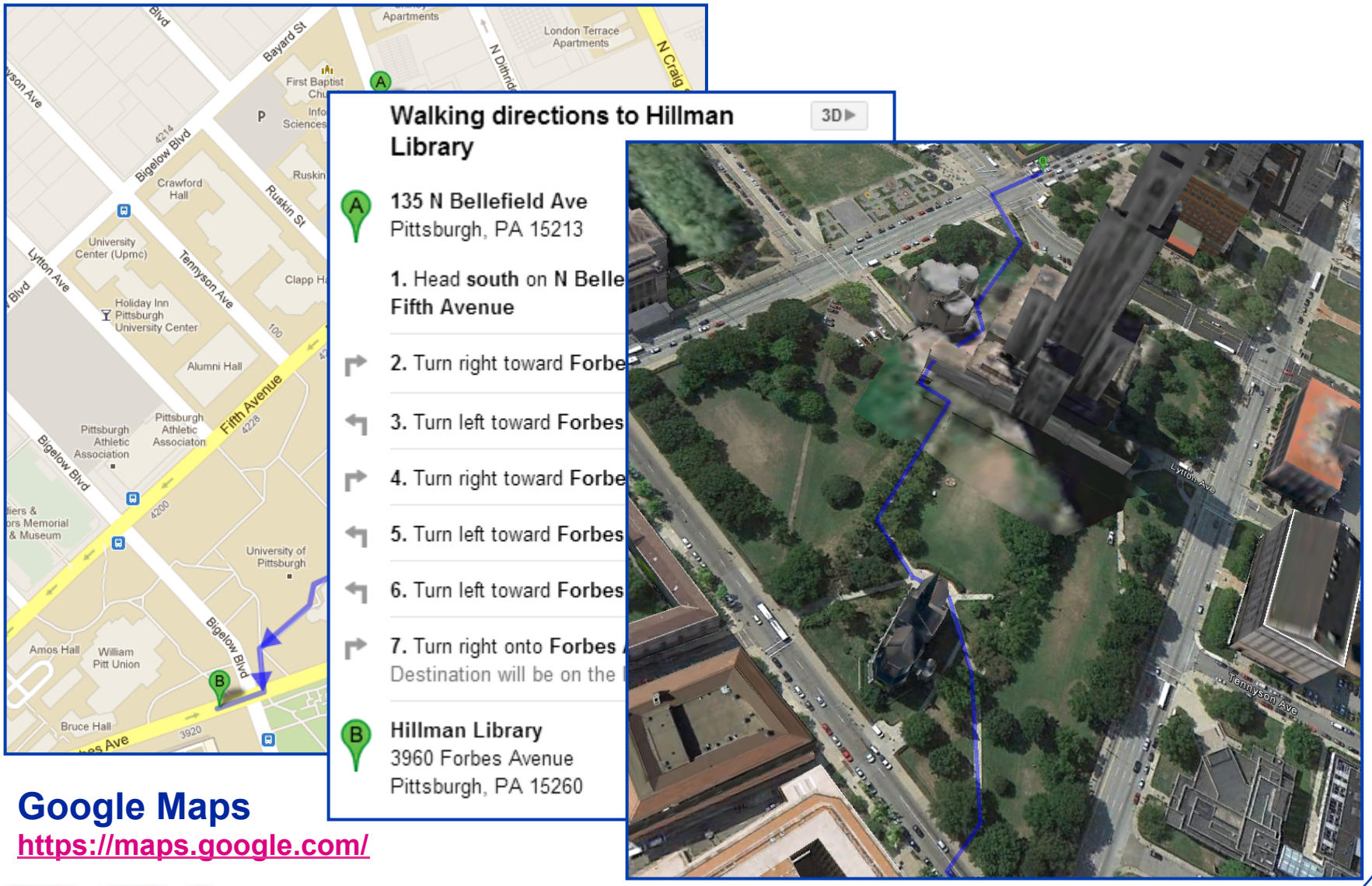
Spatial data: Example



Google Maps

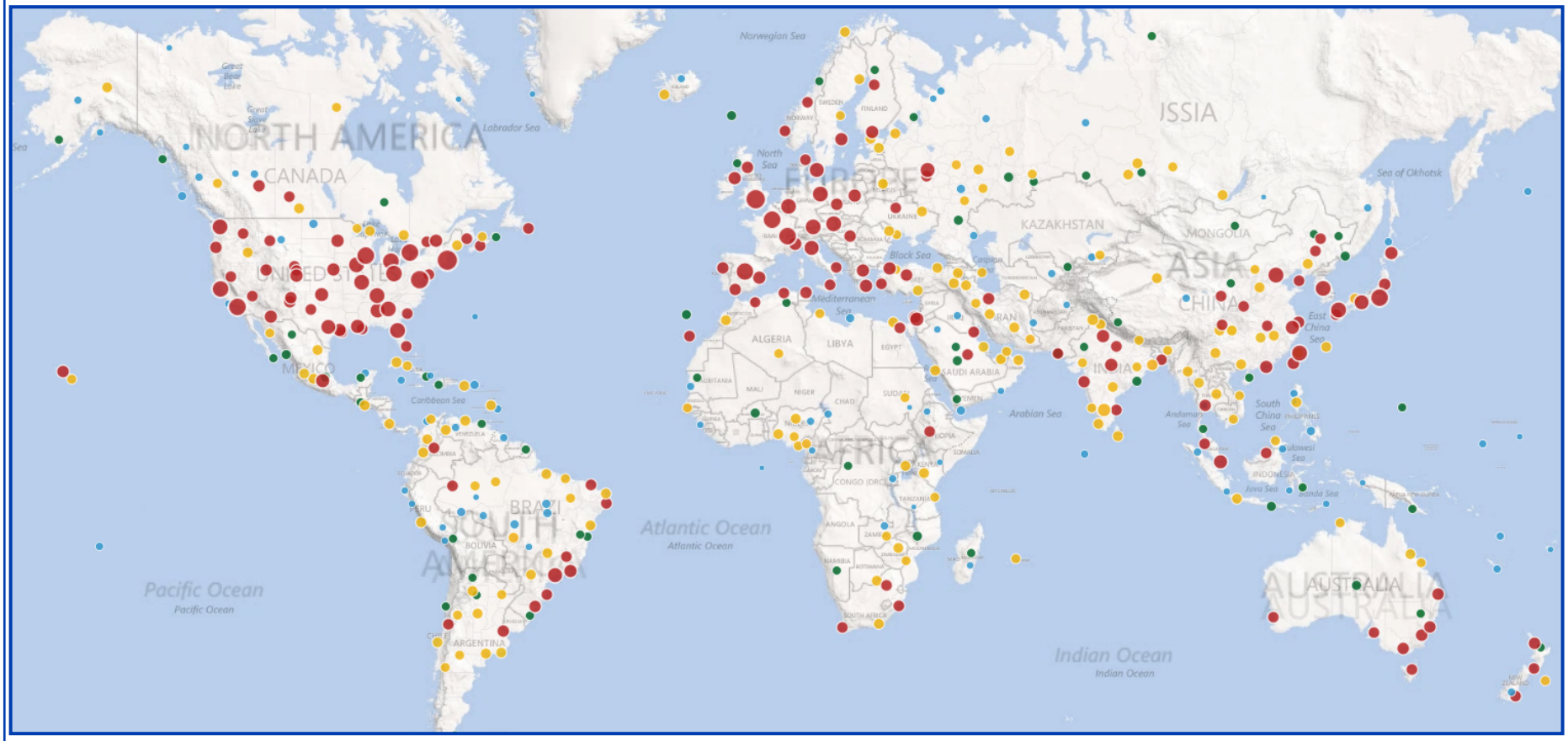
https://maps.google.com/

Spatial data: Example



Google Maps
https://maps.google.com/

Spatial relationships: Example

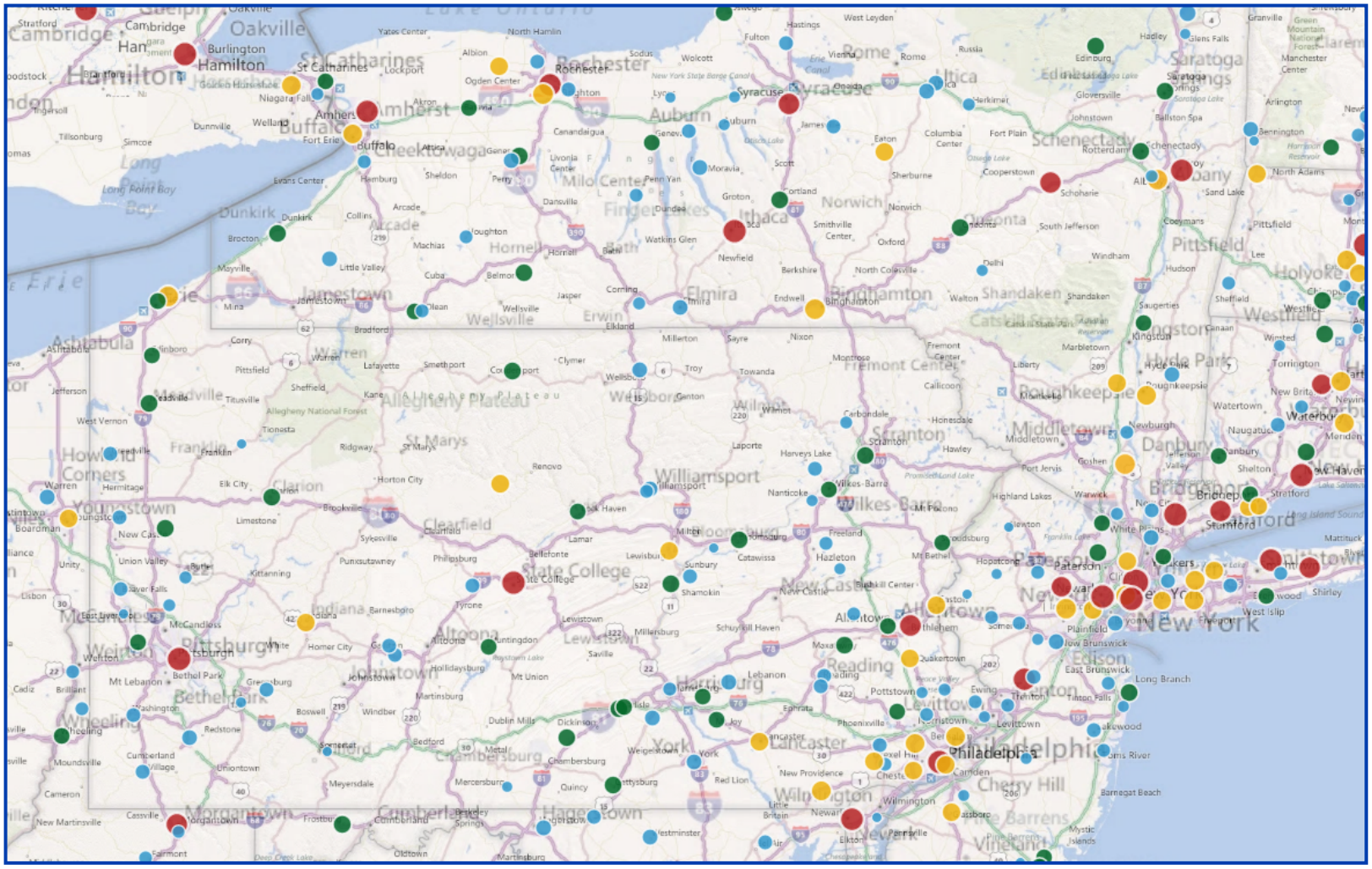


Microsoft Academic Search: Academic Map

http://academic.research.microsoft.com/AcademicMap

- Descriptive statistics
- Trends and temporal data
- Relationships
- **Spatial data**
- Social networks
- Tufte's design principles

Spatial relationships: Example

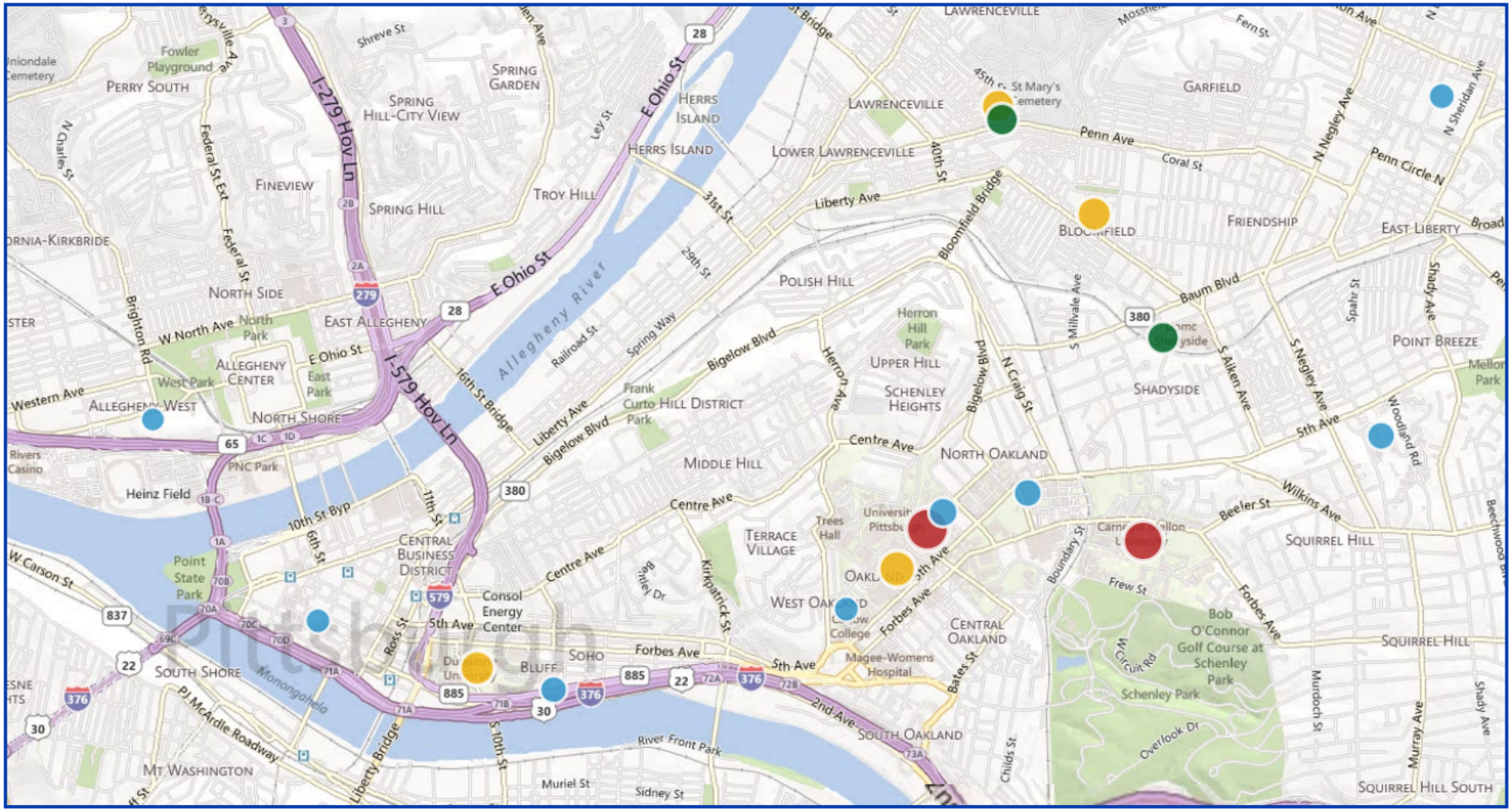


Microsoft Academic Search: Academic Map

http://academic.research.microsoft.com/AcademicMap

- Descriptive statistics
- Trends and temporal data
- Relationships
- **Spatial data**
- Social networks
- Tufte's design principles

Spatial relationships: Example

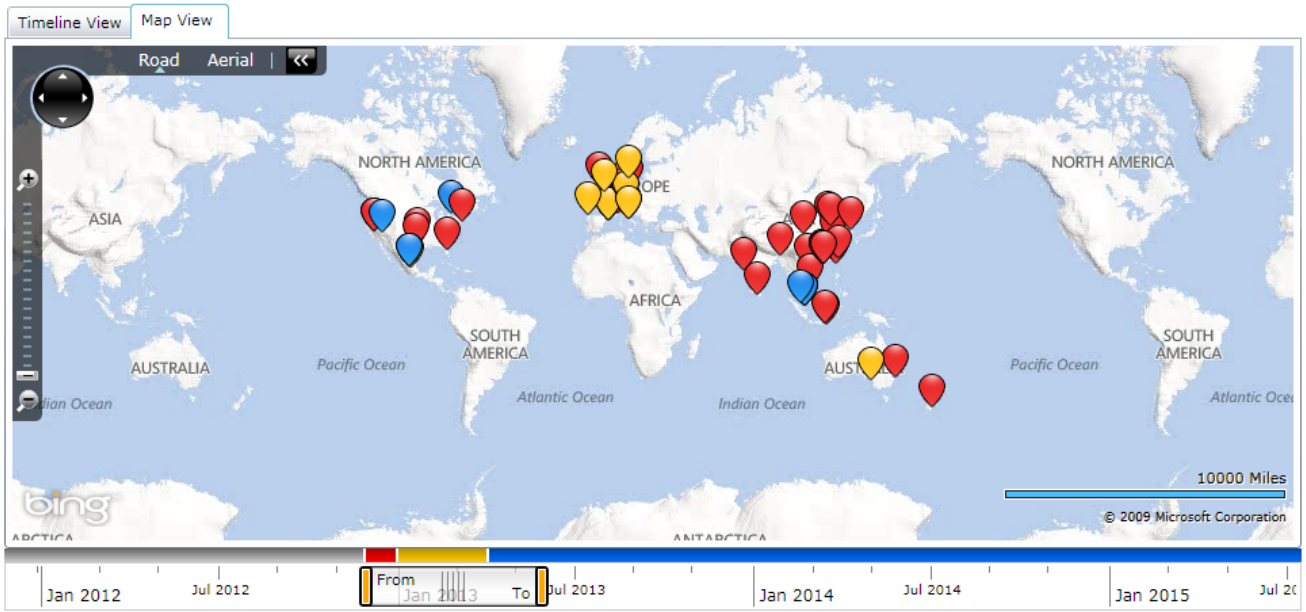


Microsoft Academic Search: Academic Map

http://academic.research.microsoft.com/AcademicMap

Spatial relationships: Example

- Descriptive statistics
- Trends and temporal data
- Relationships
- **Spatial data**
- Social networks
- Tufte's design principles



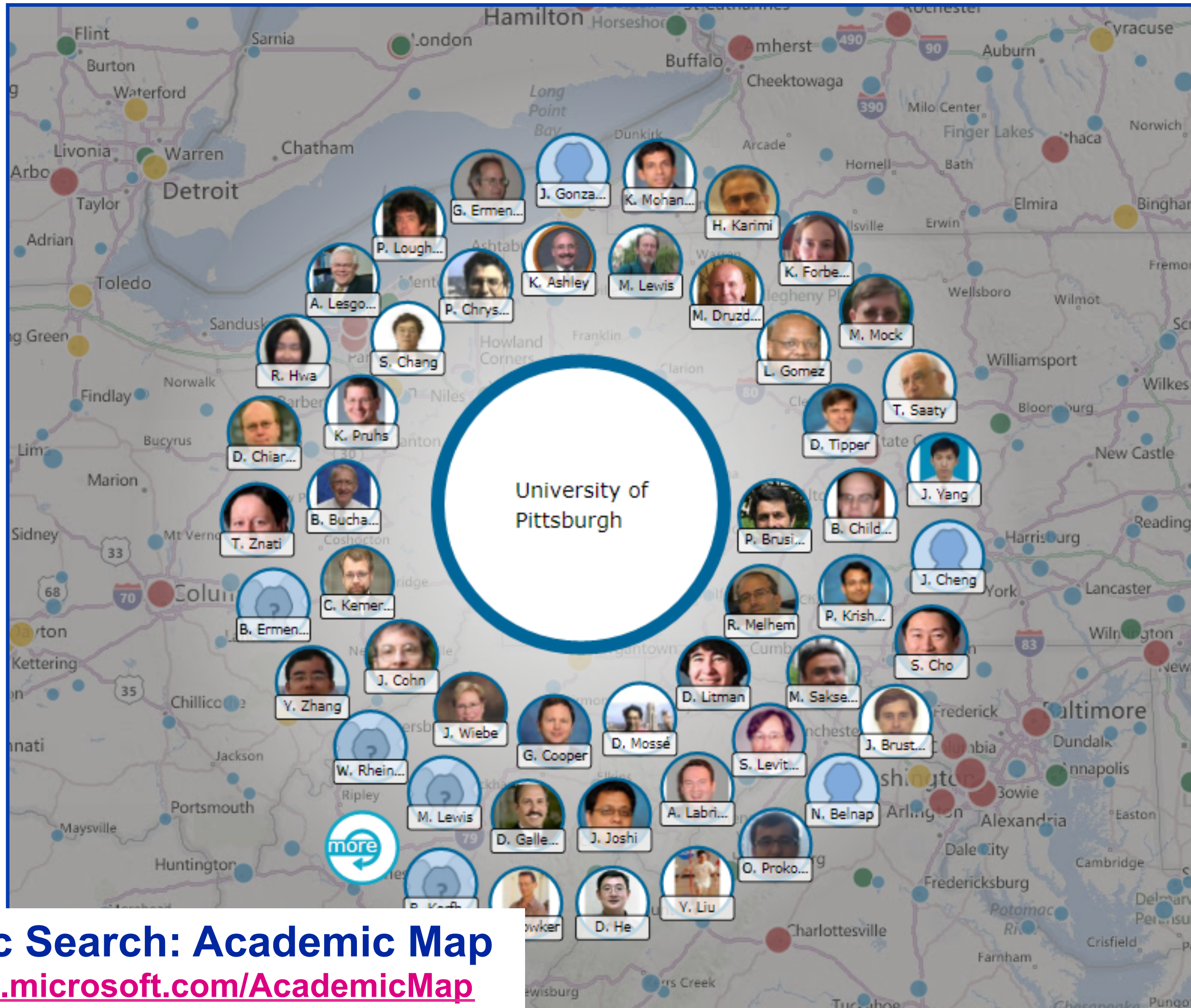
Can't find the right conference? Try to search it

Name	Full Name	Location	Submission Deadline	Duration	Detail
ICIS	International Conference on Information Systems	Orlando, United State	3/3/2012	12/16/2012 - 12/19/2012	Detail
ER(OOER)	Object-Oriented and Entity-Relationship Modelling/Internatio	Florence, Italy	3/11/2012	12/15/2012 - 12/18/2012	Detail
GLOBECOM	Global Telecommunications Conference	California, United Sta	3/30/2012	12/3/2012 - 12/7/2012	Detail
WSC	Winter Simulation Conference	Berlin, Germany	4/2/2012	12/9/2012 - 12/12/2012	Detail
SEAL	Simulated Evolution and Learning	Ha Noi, Vietnam	5/1/2012	12/16/2012 - 12/19/2012	Detail
ISSRE	International Symposium on Software Reliability Engineering	Dallas, United State	5/4/2012	11/27/2012 - 11/30/2012	Detail
ICCCI	International Conference on Computational Collective Intellig	Ho Chi Minh City, Vie	5/15/2012	11/28/2012 - 11/30/2012	Detail
APCCAS	Asia Pacific Conference on Circuits and Systems	Kaohsiung, Taiwan	5/17/2012	12/2/2012 - 12/5/2012	Detail
ASIACRYPT	International Conference on the Theory and Application of Cr	Beijing, China	5/20/2012	12/2/2012 - 12/6/2012	Detail
IWCIA	International Workshop on Combinatorial Image Analysis	Austin, United State	5/20/2012	11/28/2012 - 11/30/2012	Detail

Microsoft Academic Search: Scientific event calendar

http://academic.research.microsoft.com/

Spatial relationships: Example

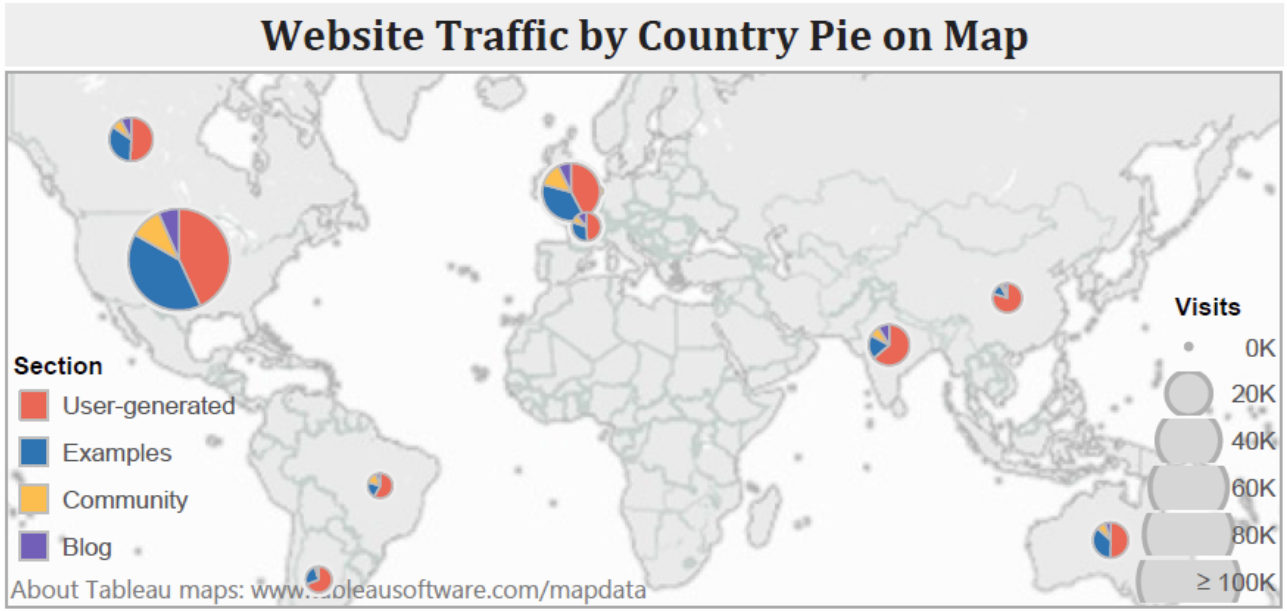


Microsoft Academic Search: Academic Map

http://academic.research.microsoft.com/AcademicMap

Decision Systems Laboratory

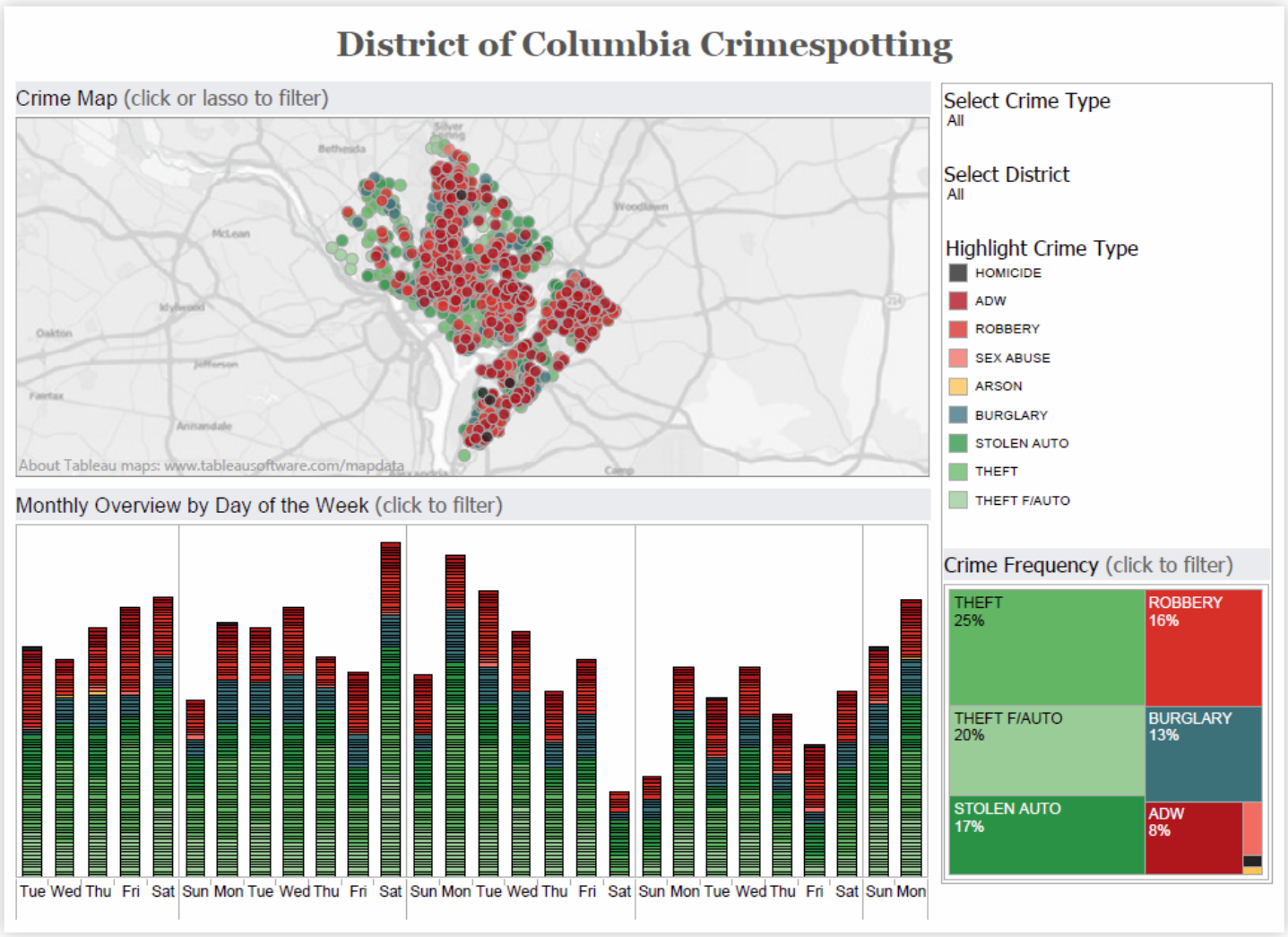
Spatial relationships: Example



	Blog	Community	Examples	User-generated
United States	10K	16K	62K	67K
United Kingdom	2K	4K	12K	14K
Canada	1K	1K	6K	9K
India	1K	1K	3K	10K
Australia	1K	1K	4K	6K
China	0K	0K	1K	7K
France	1K	1K	2K	4K

Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful," http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

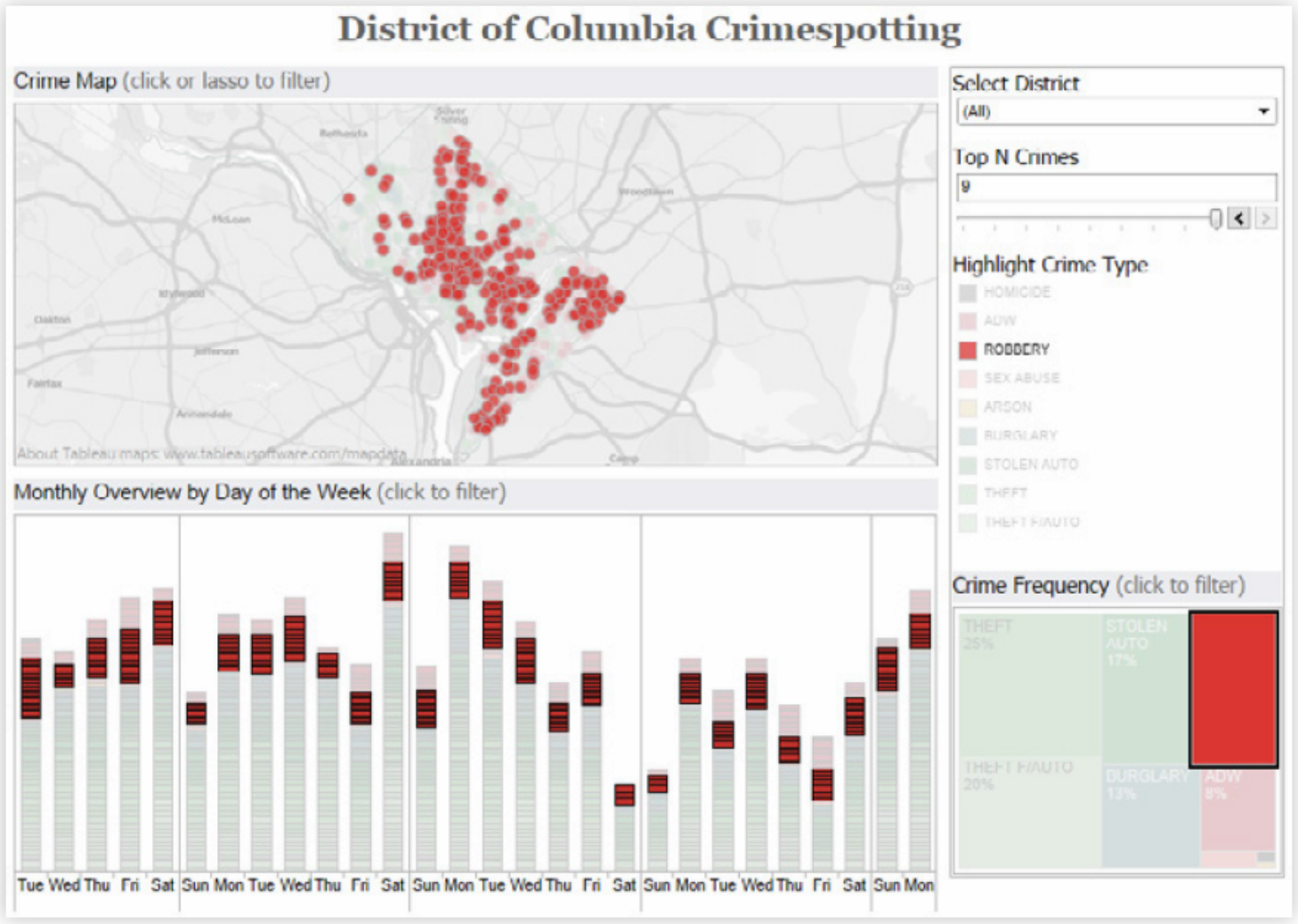
Spatial relationships: Example



Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful,"
http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

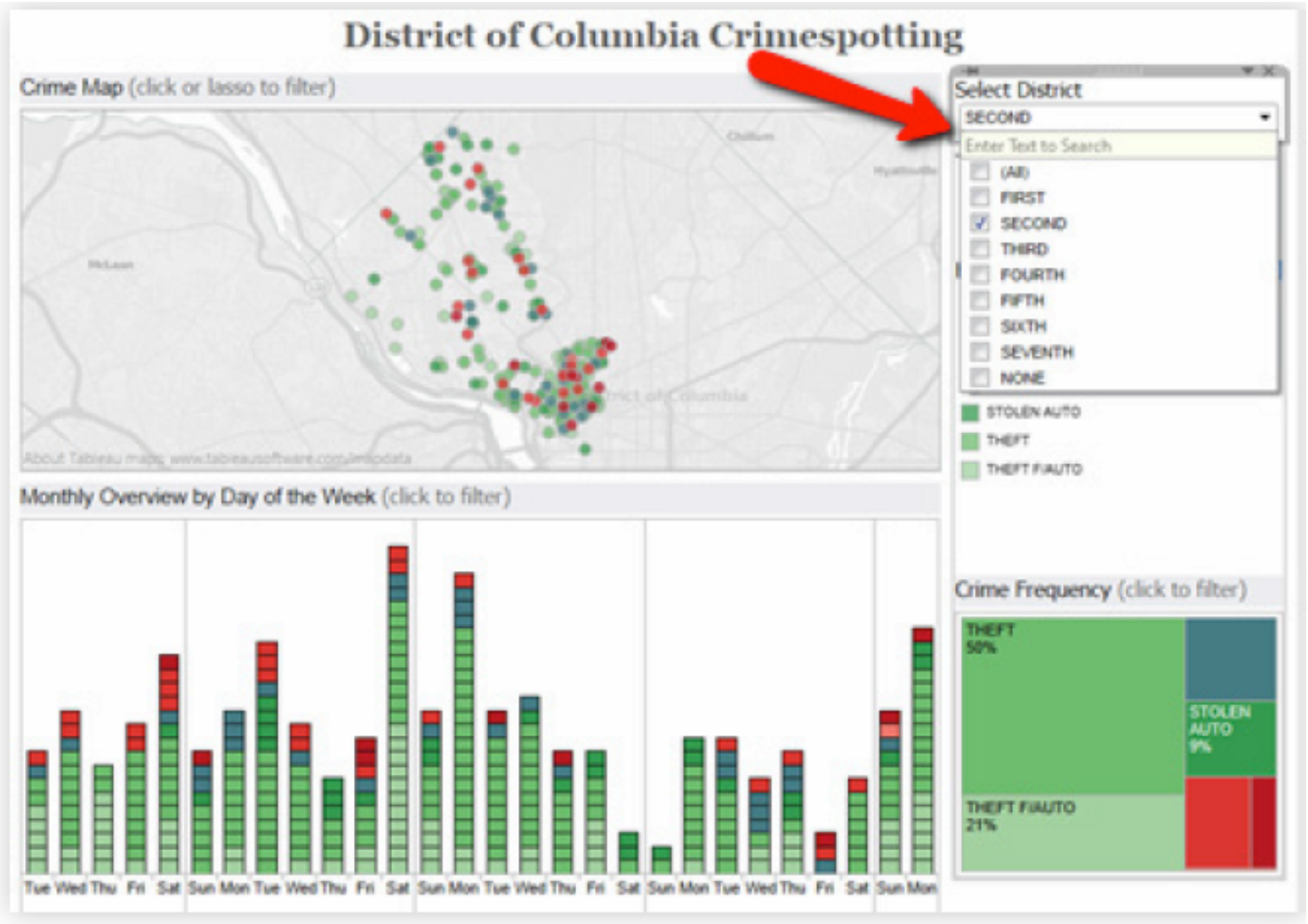
- Descriptive statistics
- Trends and temporal data
- Relationships
- **Spatial data**
- Social networks
- Tufte's design principles

Spatial relationships: Example (highlighting)



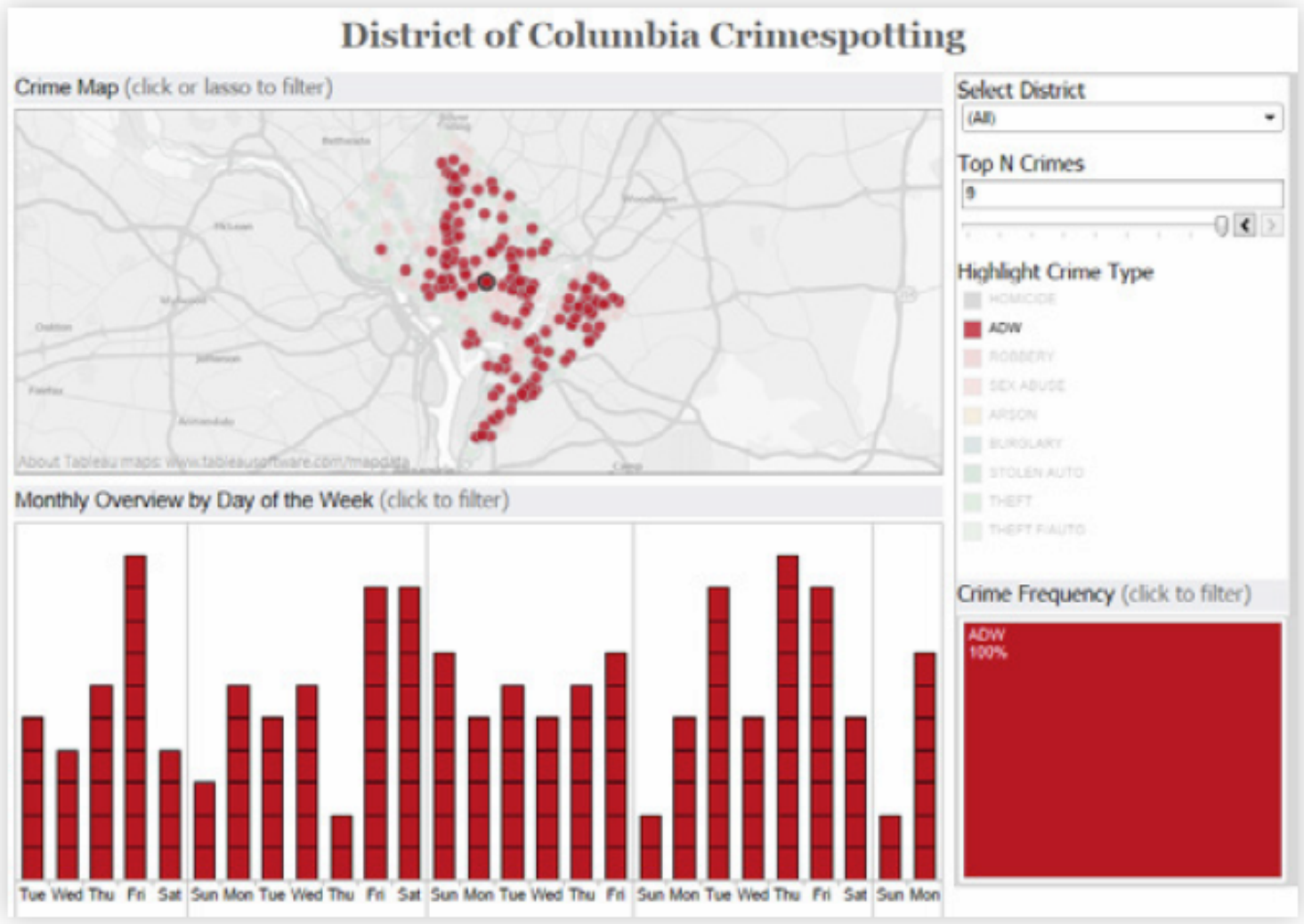
Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful," http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

Spatial relationships: Example (filtering)



Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful,"
http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

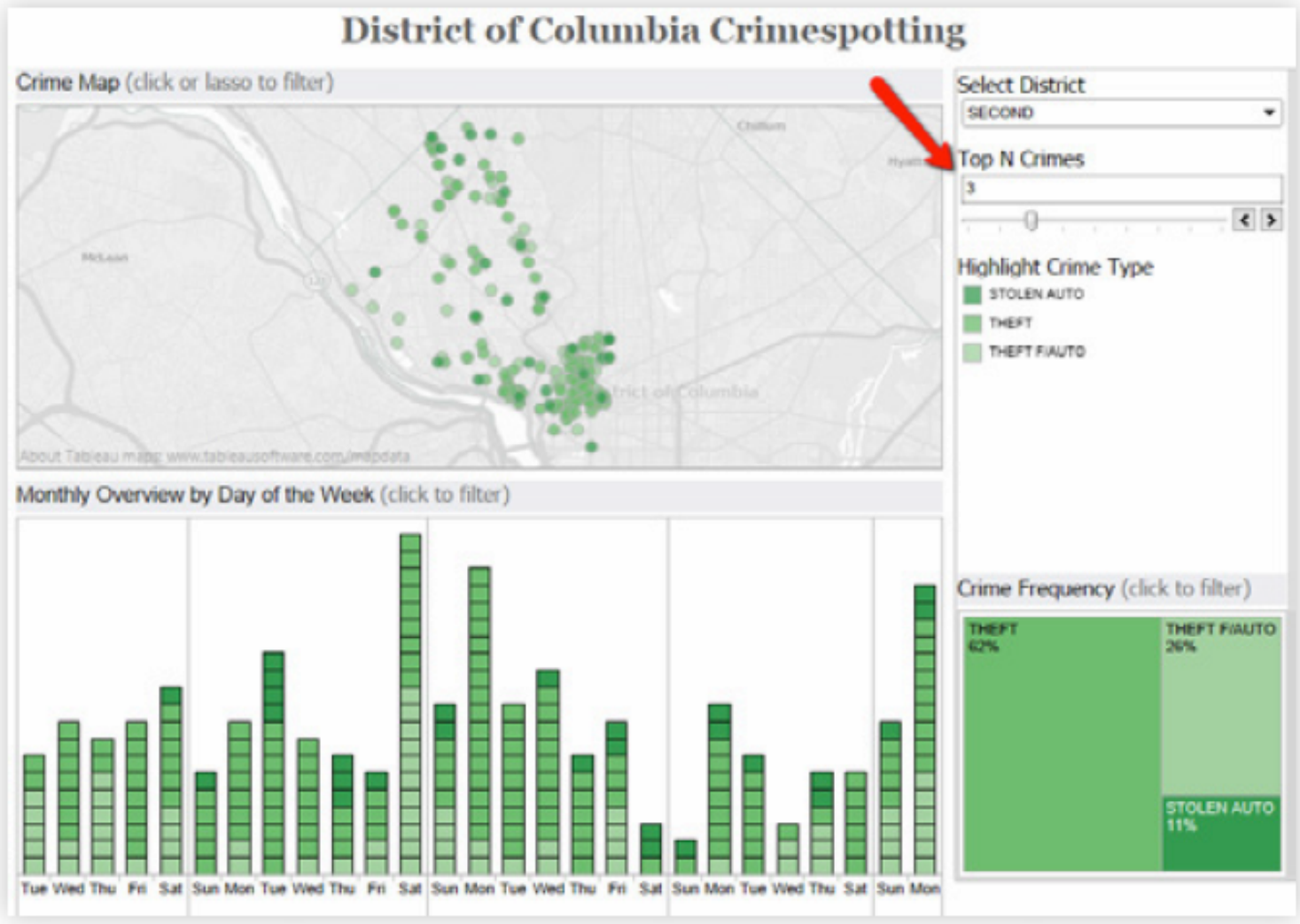
Spatial relationships: Example (filtering)



Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful," http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

- Descriptive statistics
- Trends and temporal data
- Relationships
- **Spatial data**
- Social networks
- Tufte's design principles

Spatial relationships: Example (filtering)

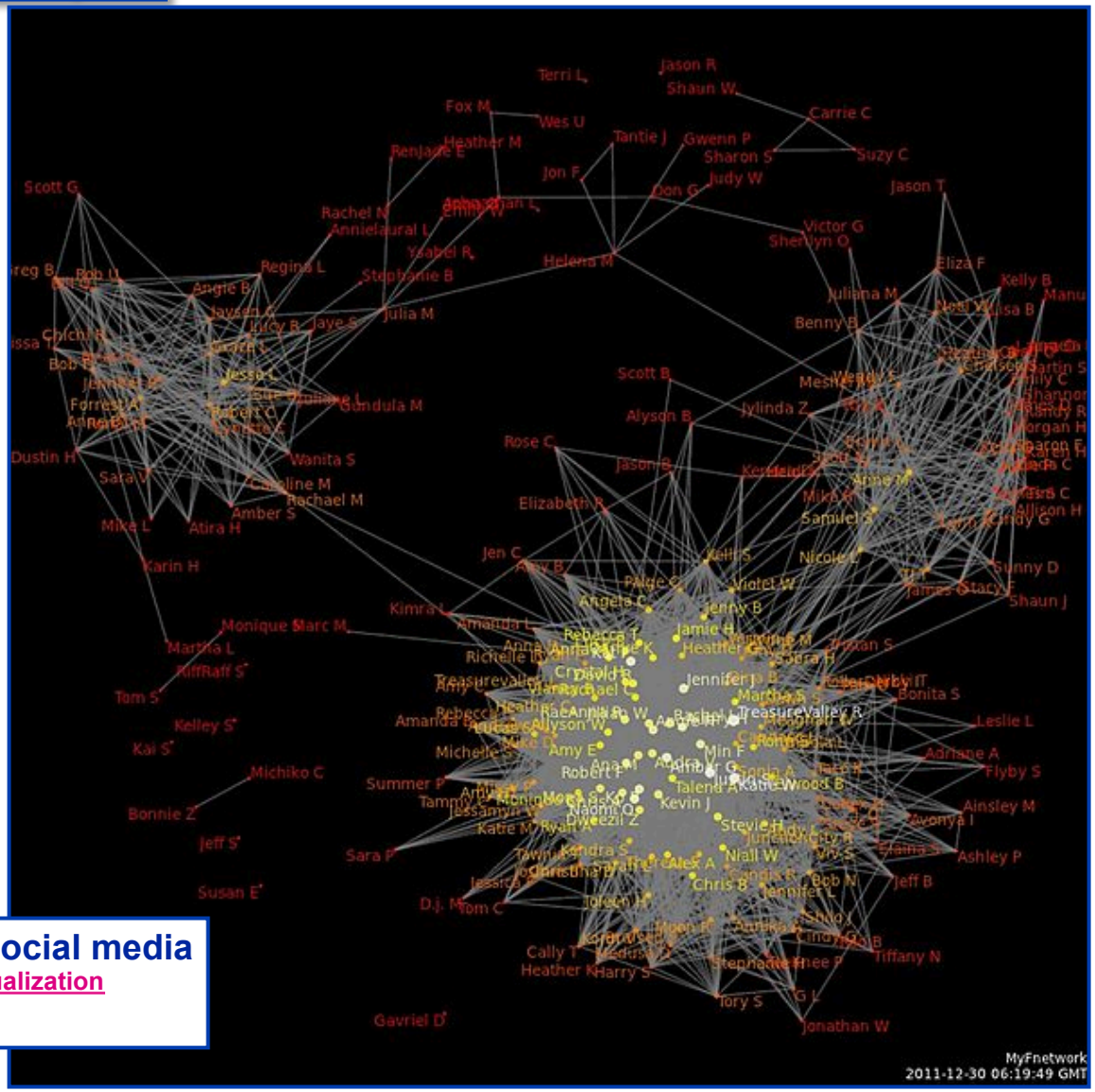


Source: Tableau White Paper "Visual Analysis Best Practices Simple Techniques for Making Every Data Visualization Useful and Beautiful," http://www.tableau.com/learn/whitepapers/tableau-visual-guidebook

Visualization of Social Networks

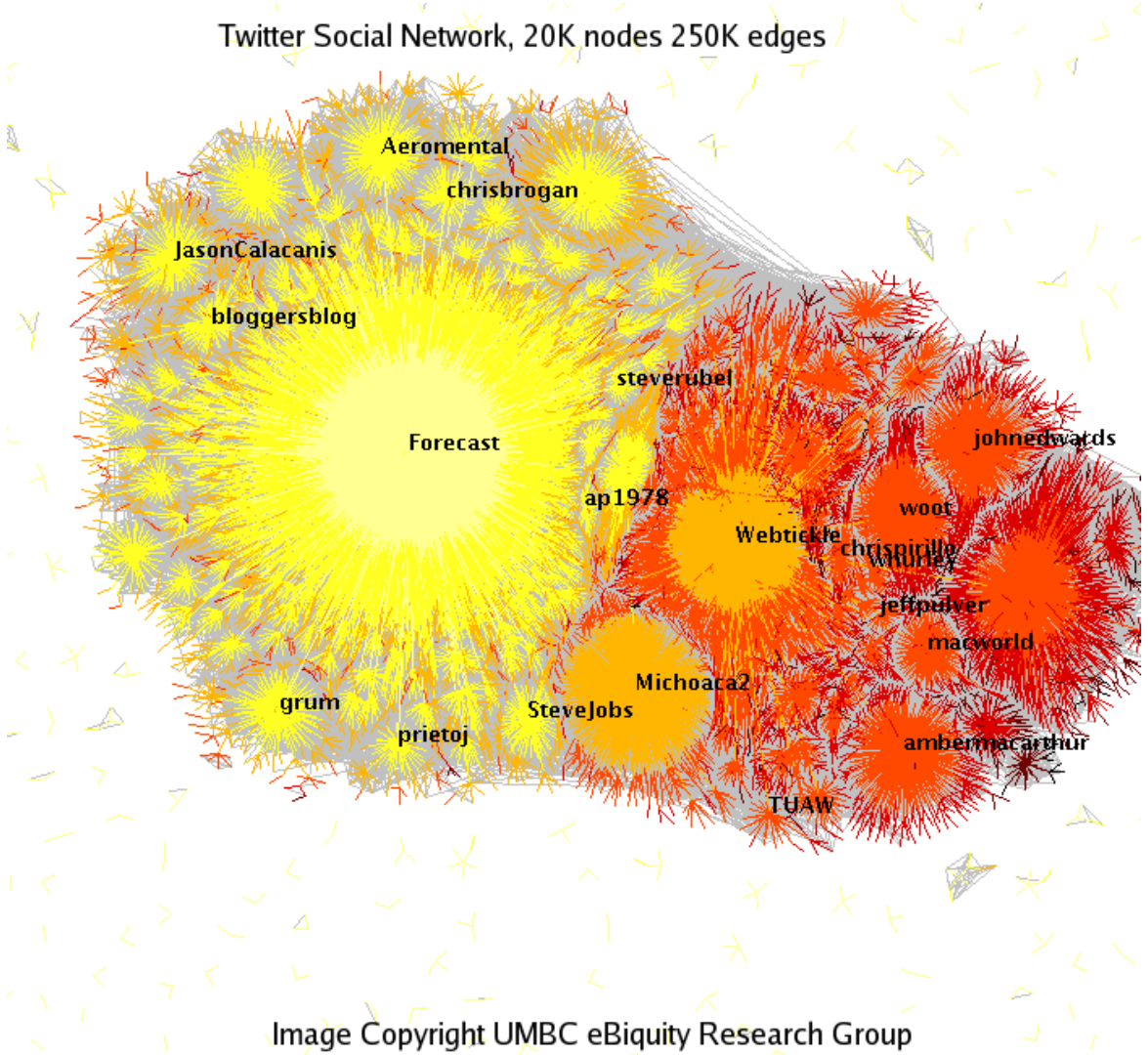
Social media: Example

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- **Social networks**
- Tufte's design principles



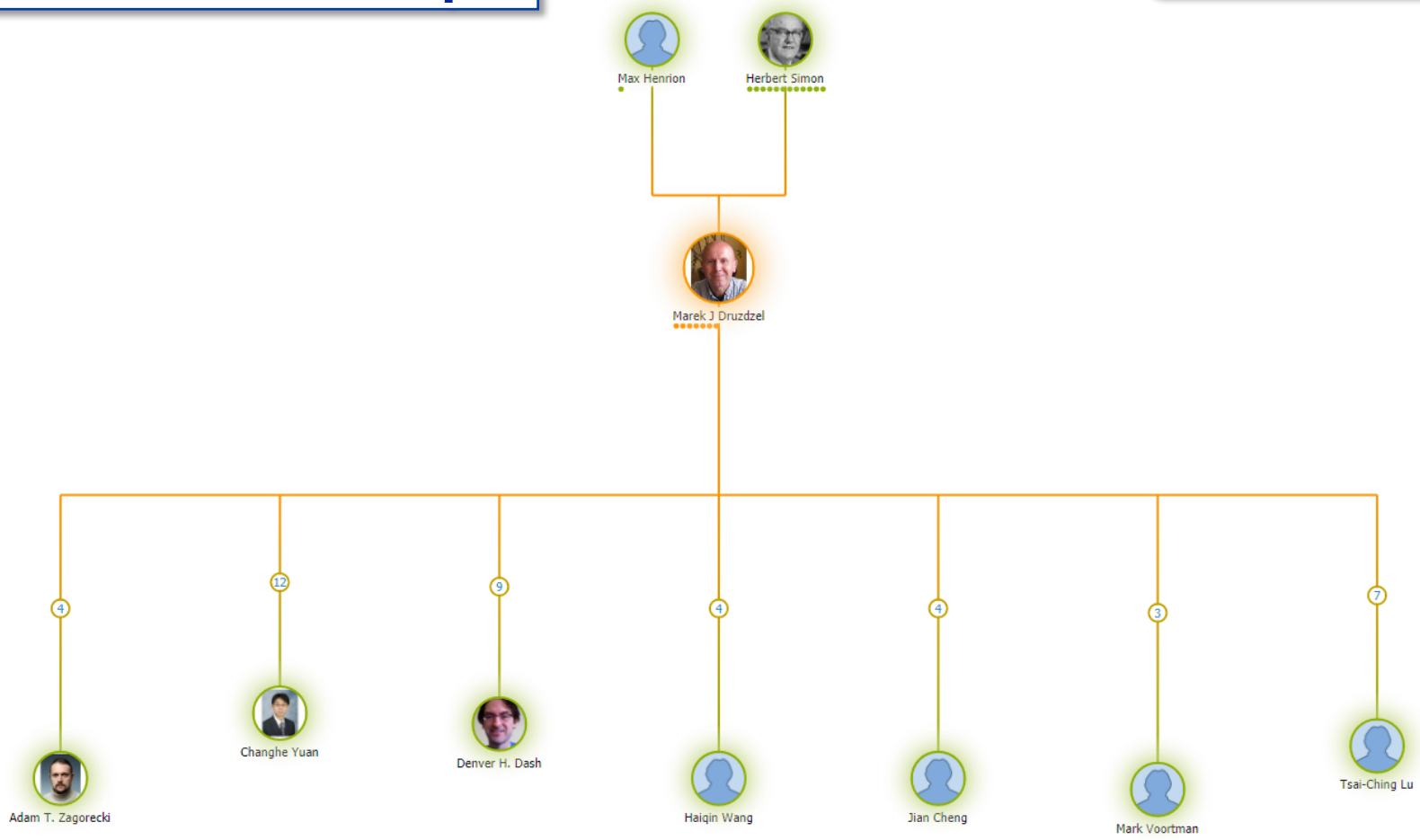
A data visualization from social media
http://en.wikipedia.org/wiki/Data_visualization

Social media: Example



- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- **Social networks**
- Tufte's design principles

Social media: Example

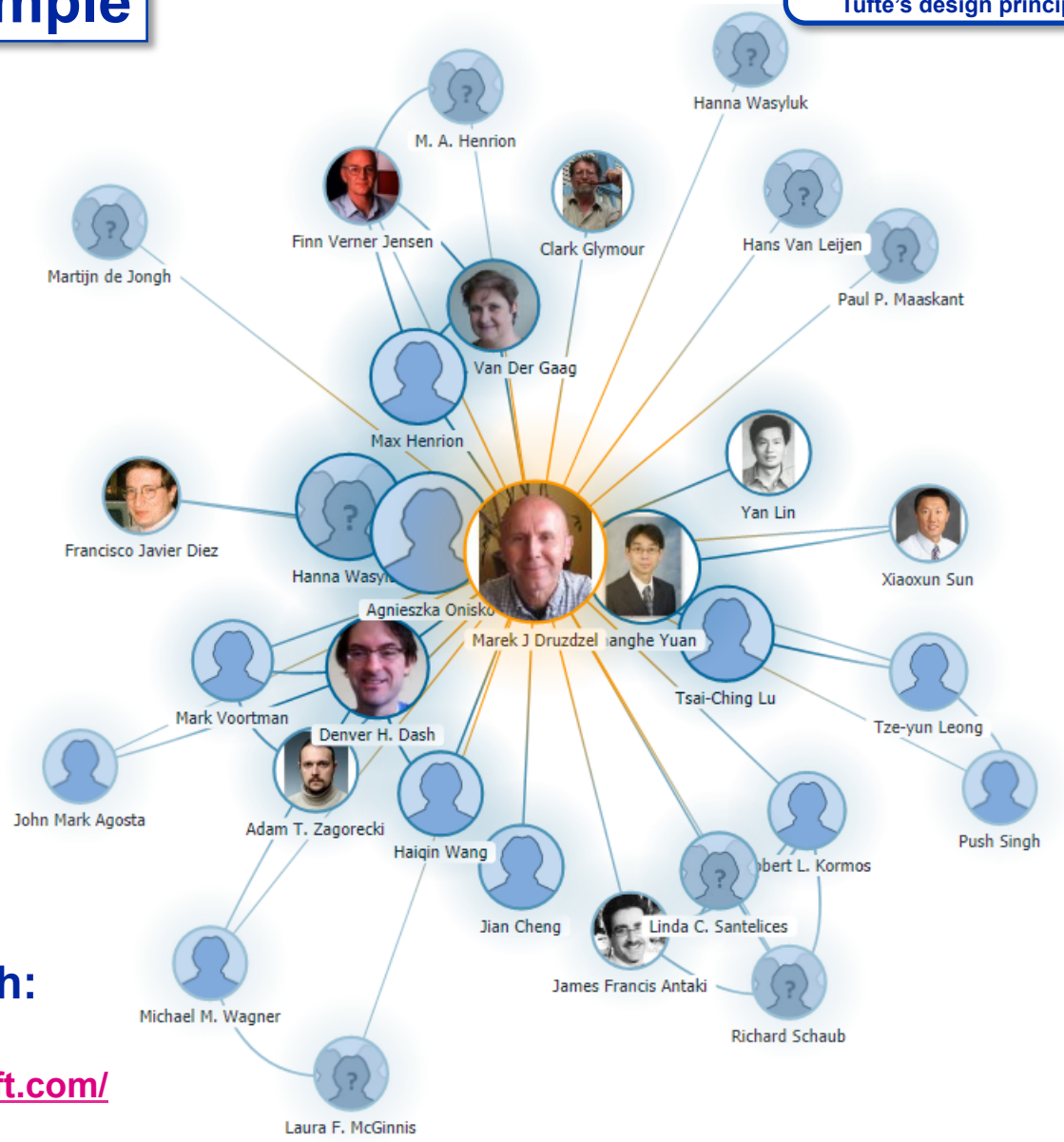


Microsoft Academic Search: Genealogy graphs

http://academic.research.microsoft.com/

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- **Social networks**
- Tufte's design principles

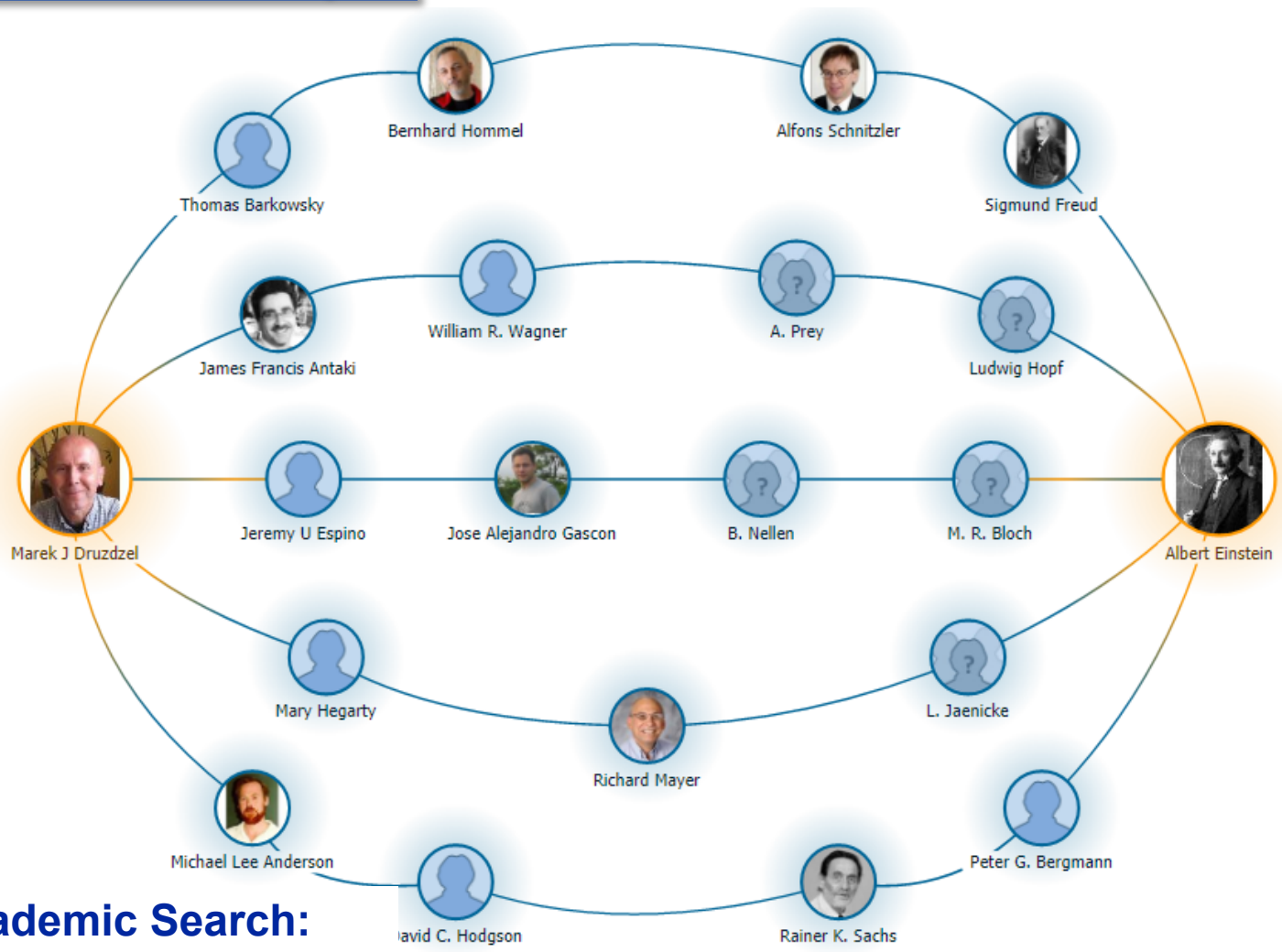
Social media: Example



Microsoft Academic Search: Co-author graphs

http://academic.research.microsoft.com/

Social media: Example

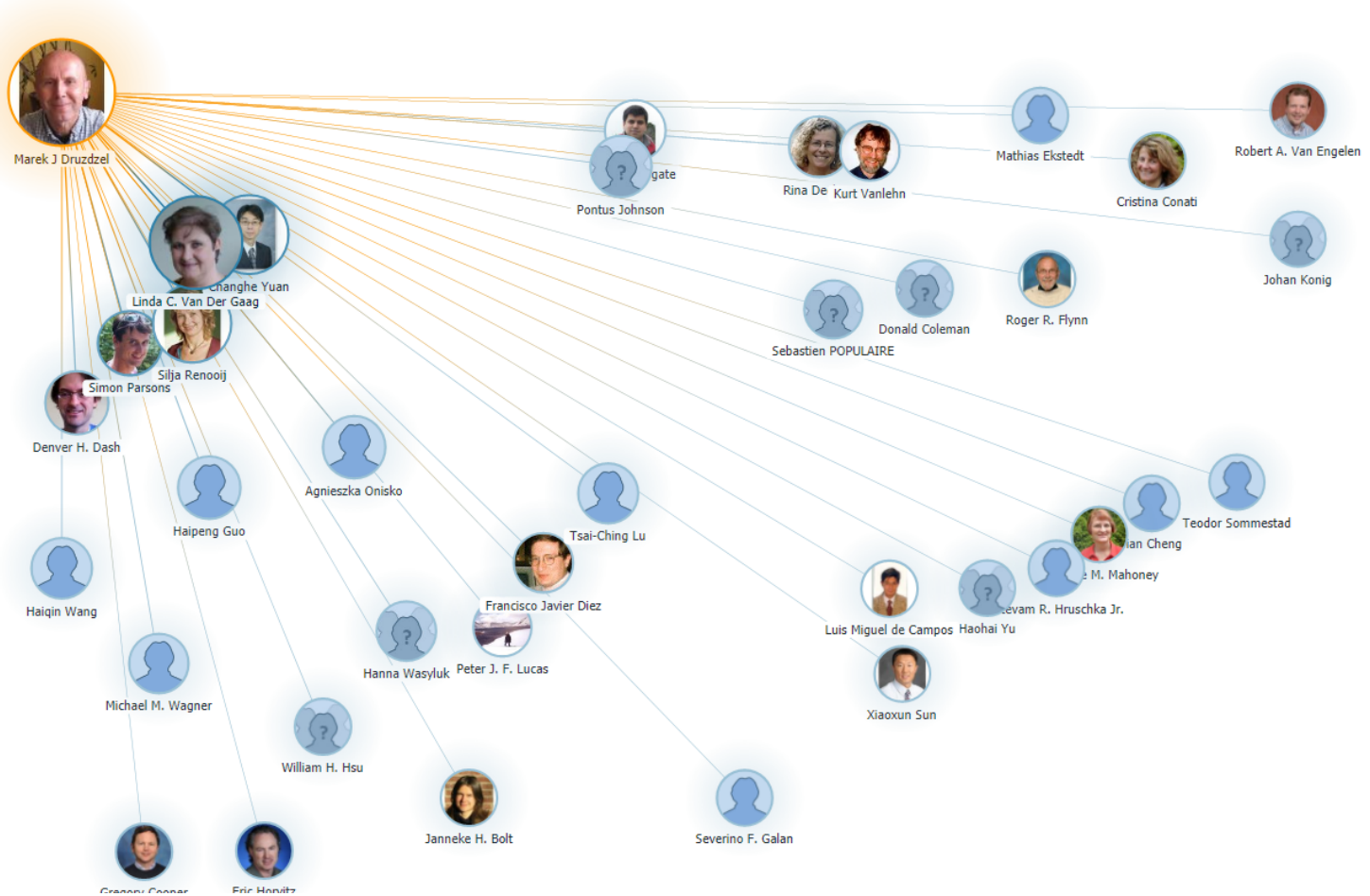


Microsoft Academic Search: Co-author paths

http://academic.research.microsoft.com/

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- **Social networks**
- Tufte's design principles

Social media: Example



Microsoft Academic Search: Citation graphs

http://academic.research.microsoft.com/

Popular Data Visualization Software

http://www.tableau.com/

Put together an Academy Award-winning professor, a brilliant computer scientist at the world's most prestigious university, and a savvy business leader with a passion for data. Add in one of the most challenging problems in software: making databases and spreadsheets understandable to ordinary people.

Data-Driven Documents

https://D3js.org

D3.js is a JavaScript library for manipulating documents based on data. **D3** helps you bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

Download the latest version (3.5.16) here:

d3.zip

Bad Examples of Data Visualization

Tufte's "Duck"

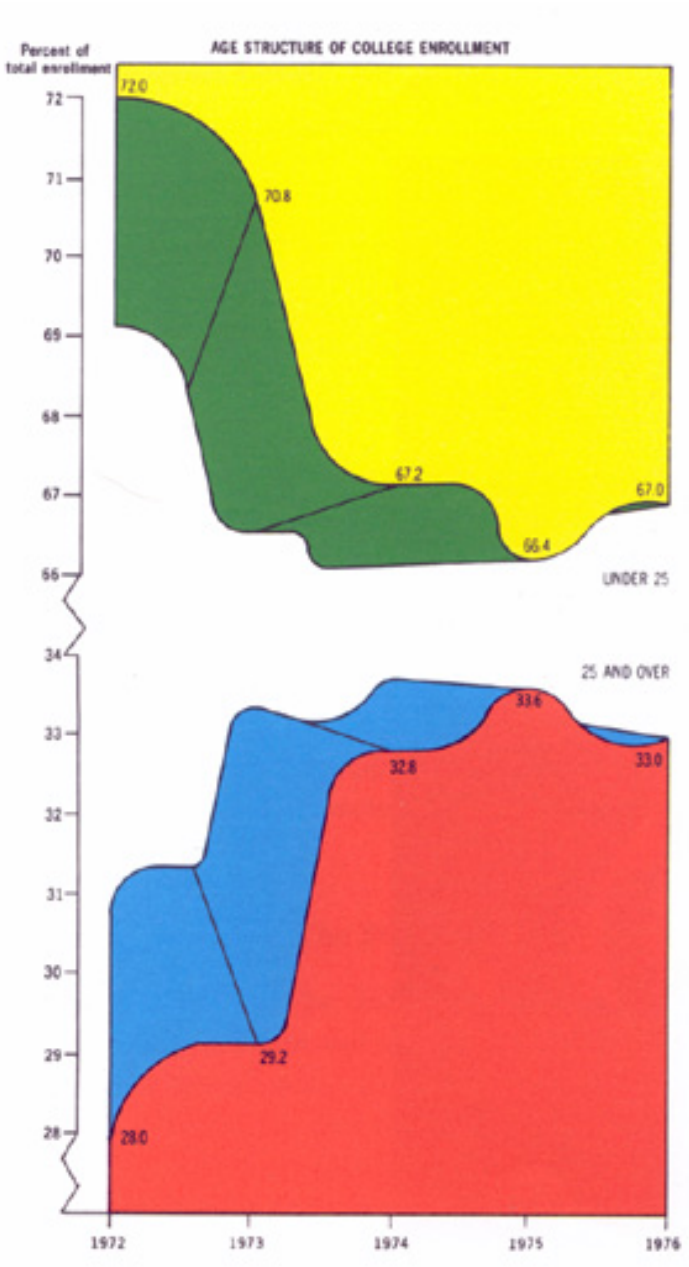
Self-Promoting Graphics: The Duck

When a graphic is taken over by decorative forms or computer debris, when the data measures and structures become Design Elements, when the overall design purveys Graphical Style rather than quantitative information, then that graphic may be called a *duck* in honor of the duck-form store, "Big Duck." For this building the whole structure is itself decoration, just as in the duck data graphic.

Graphs can also be terrible

The addition of a fake perspective to the data structure clutters many graphics.

This variety of **chartjunk**, now at high fashion in the world of Boutique Data Graphics, abounds in corporate annual reports, the phony statistical studies presented in advertisements, the mass media, and the more muddled sorts of social science research.



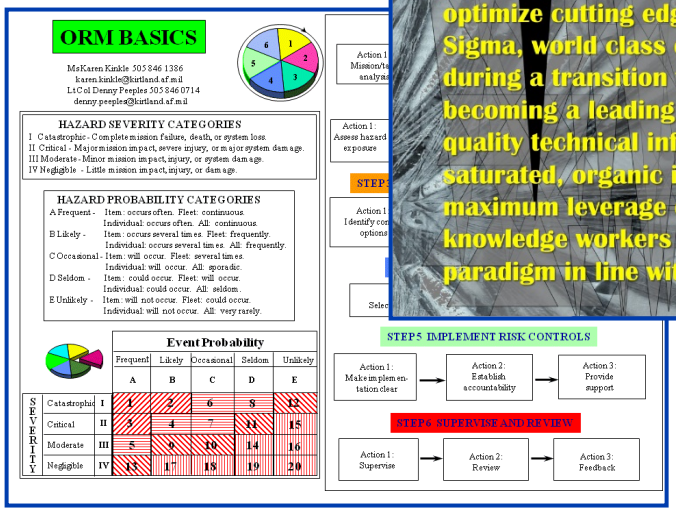
PowerPoint slides can be terrible as well ☺

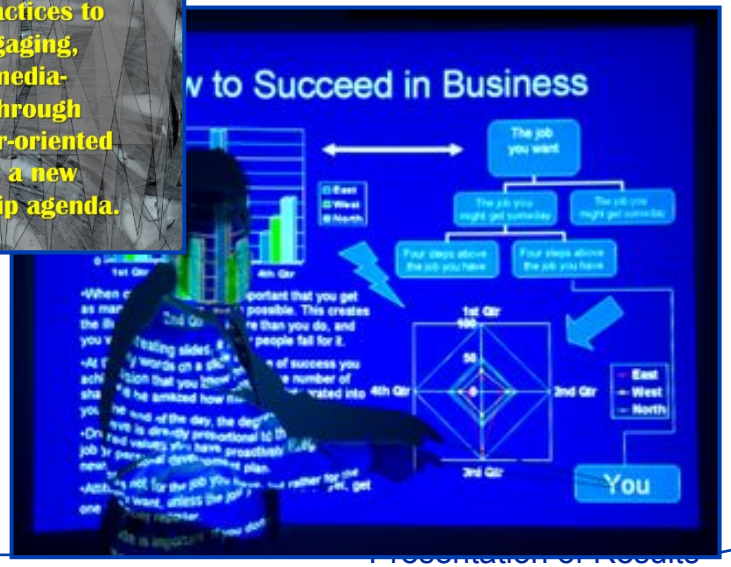
Lorem *ipsum* DOLOR **sit**

- Mauris eget dui. Phasellus mollis dolor quis libero. Aliquam ipsum massa, fermentum eget, tincidunt molestie, dictum id, purus.
- Nullam sed sem. Aliquam dictum felis ac ante. Pellentesque aliquam. Nulla in leo. Pellentesque tempor elit a nisl. Maecenas sagittis purus nec purus. In vulputate tempus erat.
- Integer viverra quam sed metus. Aliquam tincidunt sollicitudin orci. Quisque dignissim, dolor eget ornare ornare, ante enim gravida nisl, eget mattis ipsum orci nec est. Aliquam tempus suscipit lorem. Suspendisse auctor, enim vitae eleifend tempus, sapien urna cursus elit, et bibendum nibh purus at libero.
- Etiam leo. Integer ut arcu.
- Pellentesque sed dui at ligula posuere auctor. Morbi quam nibh, tristique non, malesuada eu, laoreet id, libero. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Integer vel nisl. Nullam nisi. Donec et nisl. Cras quis neque scelerisque ligula malesuada tempus. Integer non libero et nibh adipiscing tempus. In convallis. Aliquam a lorem in elit condimentum bibendum.

…l. Phasellus nisl odio, auctor et, luctus sed, luctus sit amet, nisl. Etiam …sit amet nunc. Maecenas vehicula iaculis quam. In hac habitasse platea …la ante ligula, rutrum vel, pharetra in, laoreet id, ipsum.
…na purus, mollis et, feugiat id, sollicitudin eu, orci. Vestibulum tempor. …us. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Vivamus

The revitalization project is an experiment in using Media Wiki and Wiki technology to foster collaborative best-of-breed forward thinking backend technologies in order to optimize cutting edge, revolutionary, Six Sigma, world class centers of excellence during a transition from legacy practices to becoming a leading provider of engaging, quality technical information in a media-saturated, organic infrastructure through maximum leverage of our customer-oriented knowledge workers as we architect a new paradigm in line with our leadership agenda.

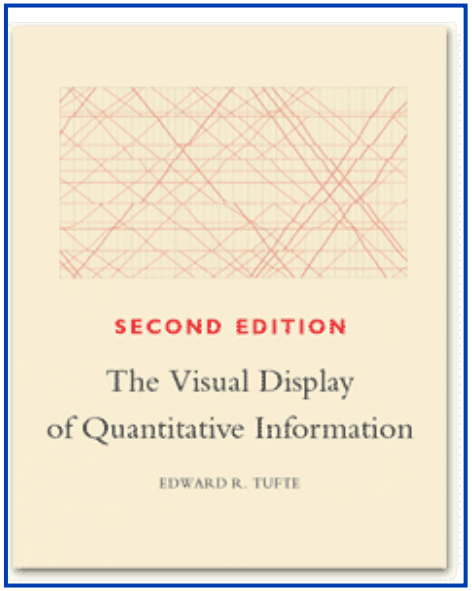


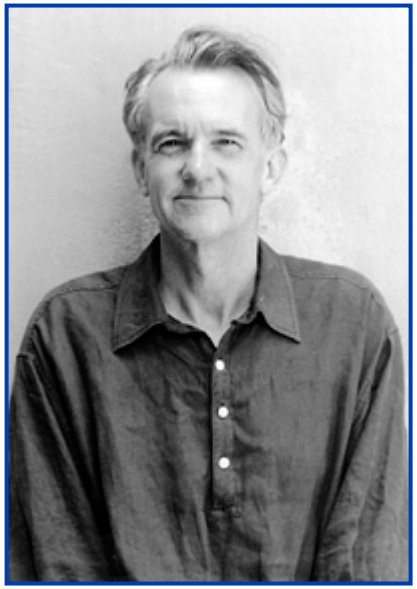


DSL
Decision Systems Laboratory

Design Principles for Visualization of Information

After Edward R. Tufte. *The visual display of quantitative information*. Graphics Press, 1983





Exemplary graphical displays

Tufte encourages the use of data-rich illustrations that present all available data.

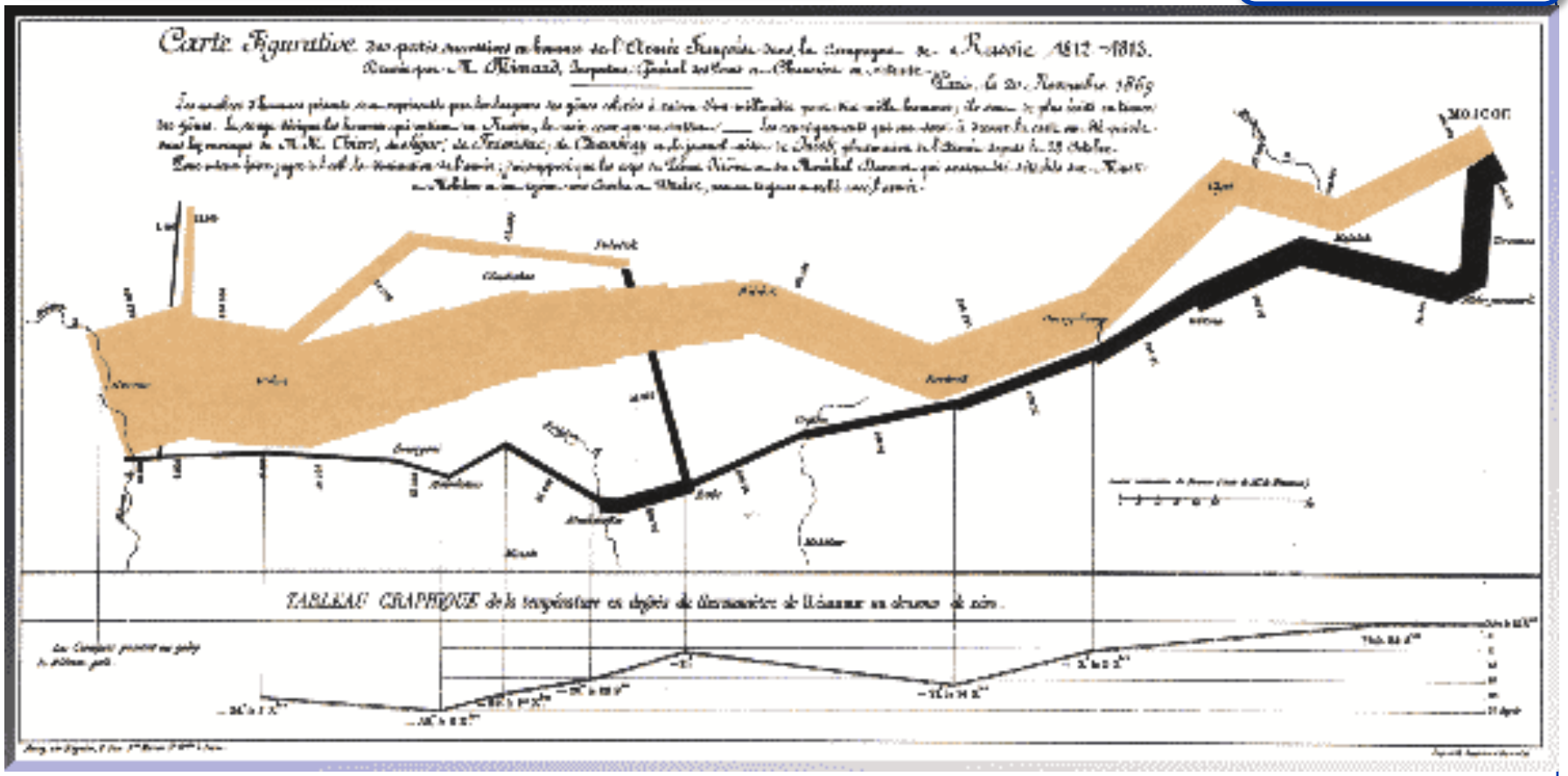
When such illustrations are examined closely, every data point has a value, but when they are looked at more generally, only trends and patterns can be observed.

Tufte suggests these macro/micro readings be presented in the space of an eye-span, in the high resolution format of the printed page, and at the unhurried pace of the viewer's leisure.

https://en.wikipedia.org/wiki/Edward_Tufte

Exemplary graphical displays

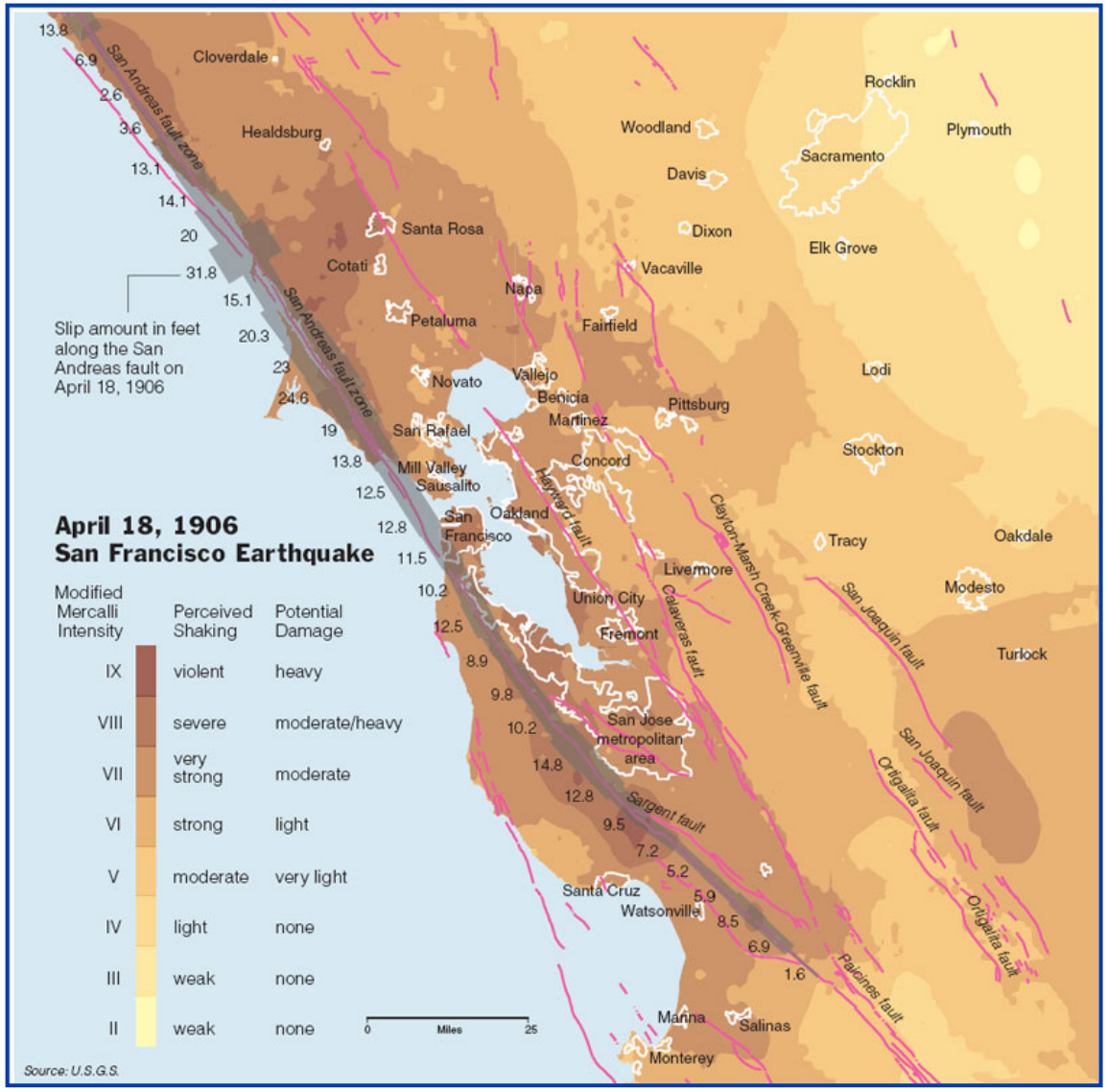
- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- **Tufte's design principles**



Charles Minard's map of Napoleon's disastrous Russian campaign of 1812. The graphic is notable for its representation in two dimensions of six types of data: *the number of Napoleon's troops; distance; temperature; the latitude and longitude; direction of travel; and location relative to specific dates.*[2]

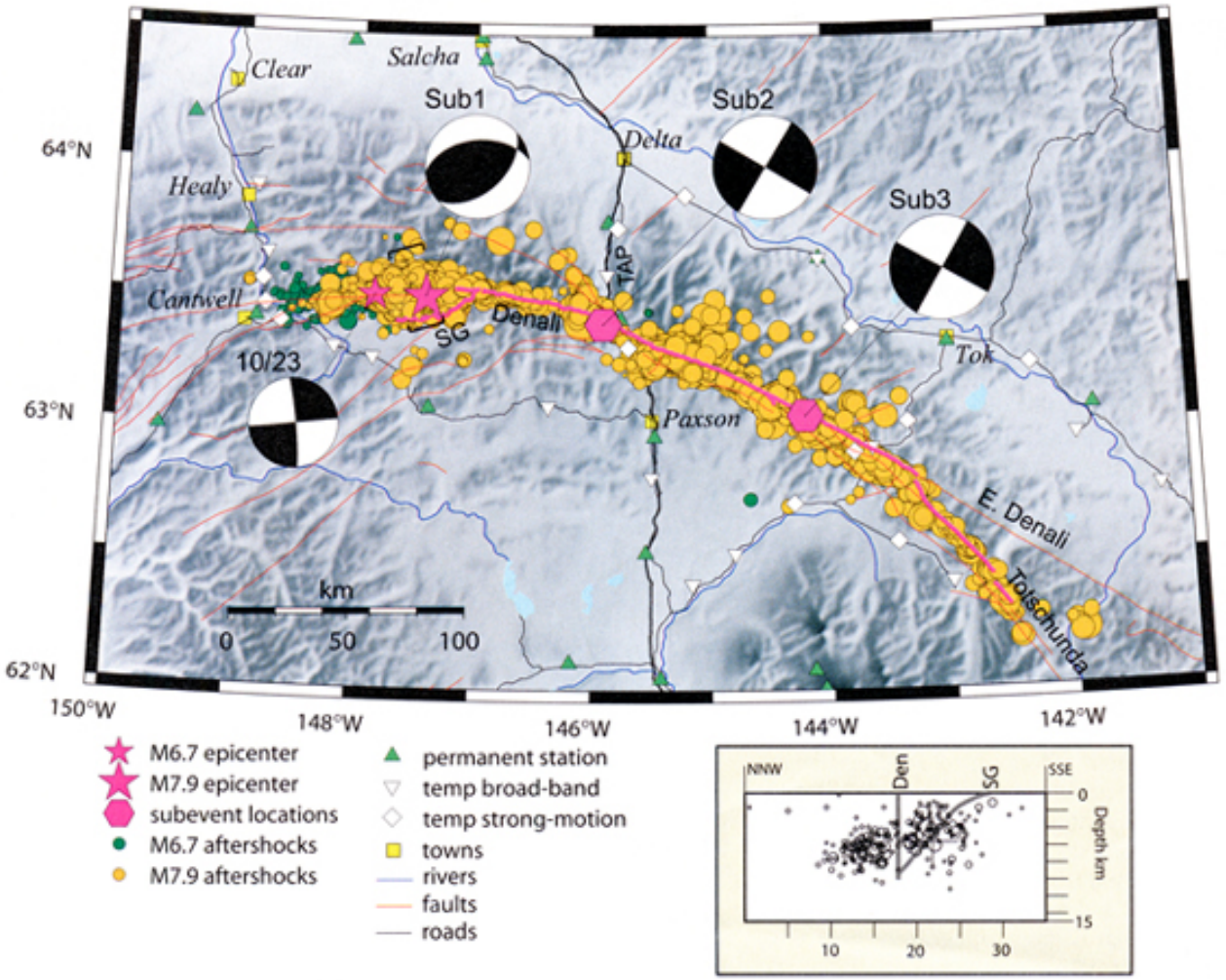
Exemplary graphical displays

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles



Exemplary graphical displays

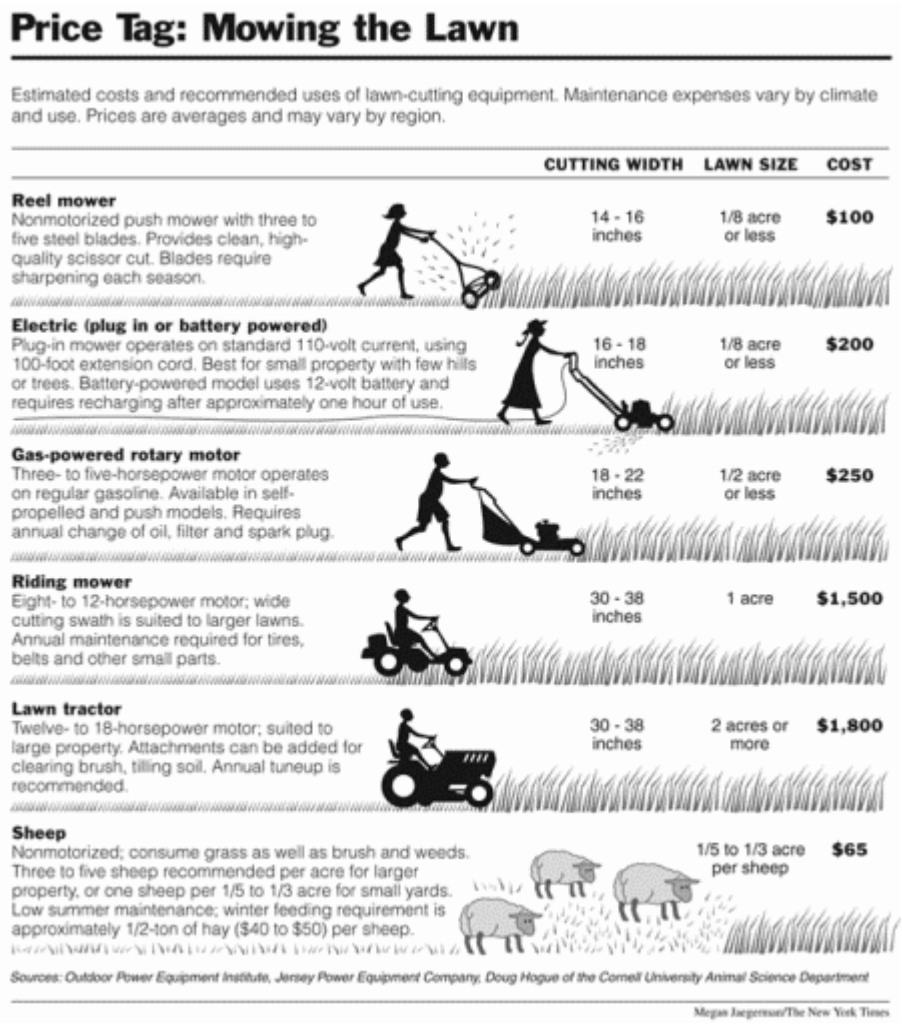
Fig. 2. Locations of principal earthquakes and aftershocks. Stars show the hypocenters of the 23 October M_w 6.7 and 3 November M_w 7.9 earthquakes, with double-difference relocated aftershocks shown in green and orange, respectively. Focal mechanisms show the first motion solution for the M_w 6.7 earthquake and the 3 subevents (sub1 to -3) determined for the M_w 7.9 earthquake. Mapped surface rupture shown as heavy magenta line; red lines indicate other faults. The inset cross section shows schematic faults and $M_L \geq 2.5$ aftershocks in the bracketed zone across the Susitna Glacier (SG) thrust, inferred to splay off the Denali (Den) fault. Cross, mainshock.



- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

Exemplary graphical displays

"At the heart of quantitative reasoning is a single question: Compared to what? *Small multiple* designs, multivariate and data bountiful, answer directly by visually enforcing comparisons of changes, of the differences among objects, of the scope of alternatives. For a wide range of problems in data presentation, small multiples are the best design solution."



Price Tag: Mowing the Lawn

Estimated costs and recommended uses of lawn-cutting equipment. Maintenance expenses vary by climate and use. Prices are averages and may vary by region.

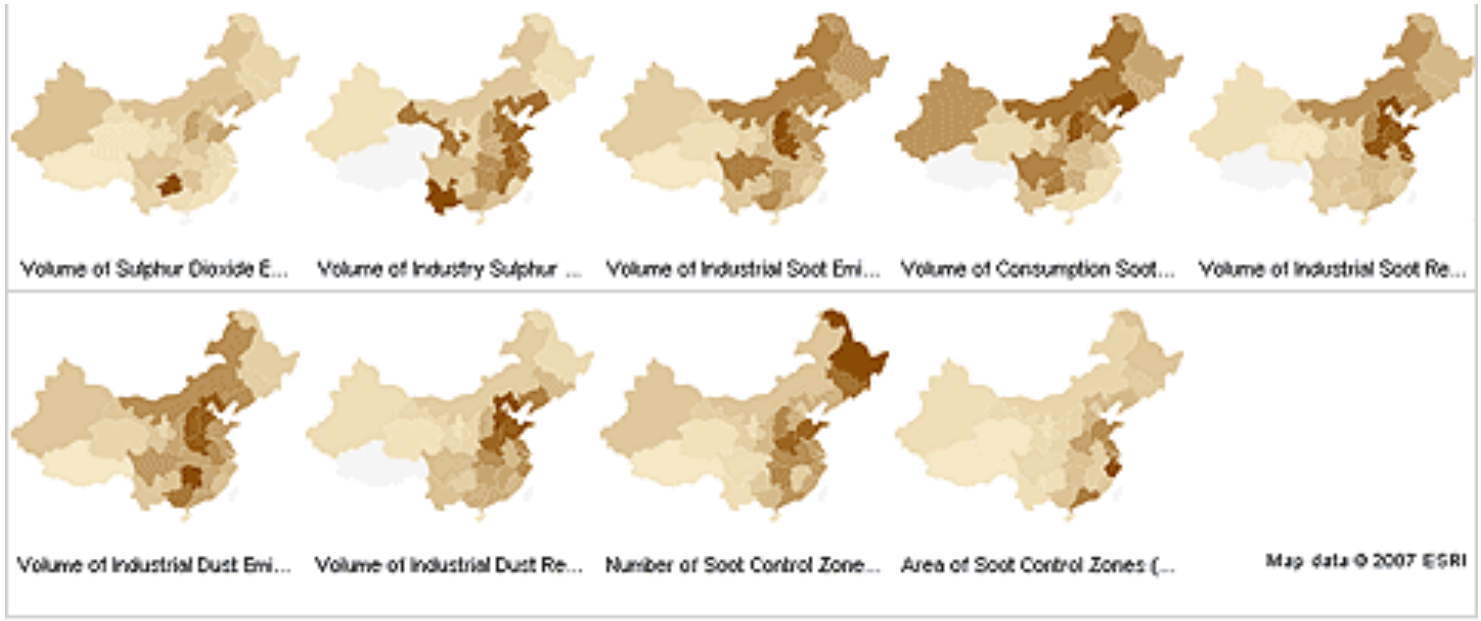
	CUTTING WIDTH	LAWN SIZE	COST
Reel mower Nonmotorized push mower with three to five steel blades. Provides clean, high-quality scissor cut. Blades require sharpening each season.	14 - 16 inches	1/8 acre or less	$100
Electric (plug in or battery powered) Plug-in mower operates on standard 110-volt current, using 100-foot extension cord. Best for small property with few hills or trees. Battery-powered model uses 12-volt battery and requires recharging after approximately one hour of use.	16 - 18 inches	1/8 acre or less	$200
Gas-powered rotary motor Three- to five-horsepower motor operates on regular gasoline. Available in self-propelled and push models. Requires annual change of oil, filter and spark plug.	18 - 22 inches	1/2 acre or less	$250
Riding mower Eight- to 12-horsepower motor; wide cutting swath is suited to larger lawns. Annual maintenance required for tires, belts and other small parts.	30 - 38 inches	1 acre	$1,500
Lawn tractor Twelve- to 18-horsepower motor; suited to large property. Attachments can be added for clearing brush, tilling soil. Annual tuneup is recommended.	30 - 38 inches	2 acres or more	$1,800
Sheep Nonmotorized; consume grass as well as brush and weeds. Three to five sheep recommended per acre for larger property, or one sheep per 1/5 to 1/3 acre for small yards. Low summer maintenance; winter feeding requirement is approximately 1/2-ton of hay ($40 to $50) per sheep.		1/5 to 1/3 acre per sheep	$65

Sources: Outdoor Power Equipment Institute, Jersey Power Equipment Company, Doug Hogue of the Cornell University Animal Science Department

Megan Jaegerman/The New York Times

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- **Tufte's design principles**

Exemplary graphical displays

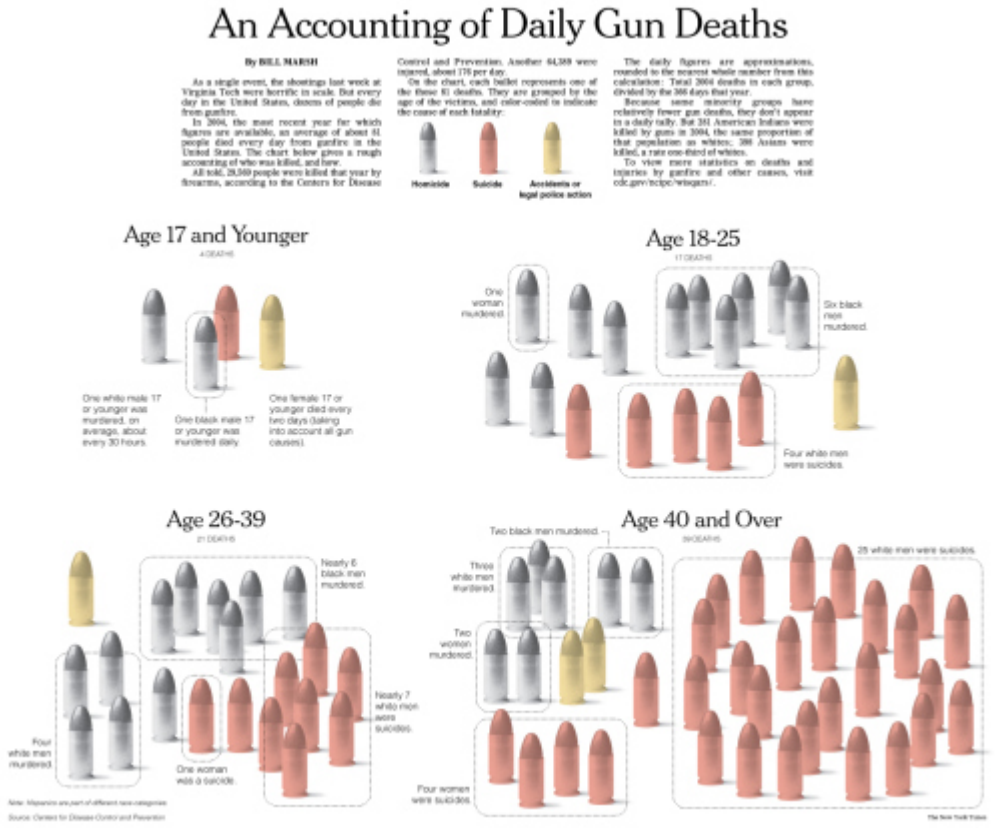


Small-multiples view: map of China showing several dimensions of air pollution.

"I do believe that there are some universal cognitive tasks that are deep and profound – indeed, so deep and profound that it is worthwhile to understand them in order to design our displays in accord with those tasks."

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles

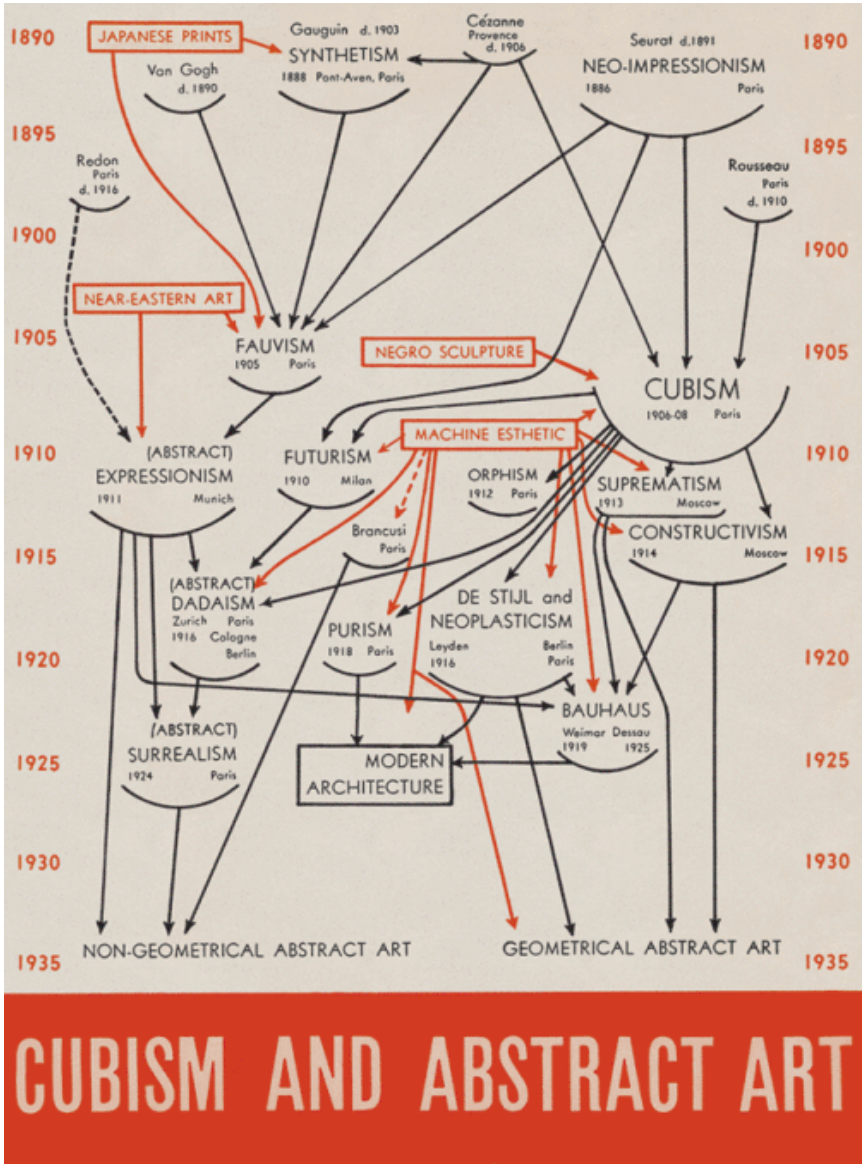
Exemplary graphical displays



"It's not that PowerPoint brought the Columbia down, but the method of presentation broke up the argument into tiny fragments, and it's intensely hierarchical-no sentences, just little phrases."

Exemplary graphical displays

- Descriptive statistics
- Trends and temporal data
- Relationships
- Spatial data
- Social networks
- Tufte's design principles



Graphical displays should

- Show the data.
- Induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else.
- Avoid distorting what the data have to say.
- Present many numbers in a small space.
- Make large data sets coherent.
- Encourage the eye to compare different pieces of data.
- Reveal the data at several levels of detail, from a broad overview to the fine structure.
- Serve a reasonably clear purpose: description, exploration, tabulation, or decoration.
- Be closely integrated with the statistical and verbal descriptions of a data set.

Principles of graphical excellence

- Graphical excellence is the well designed presentation of interesting data – a matter of substance, of statistics, and of design.
- Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency.
- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- Graphical excellence is nearly always multivariate.
- And graphical excellence requires telling truth about the data.

Graphical integrity

Graphical integrity is more likely to result if these six principles are followed:

- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.
- Show data variation, not design variation.
- In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
- The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
- Graphics must not quote data out of context.

Data-ink and graphical redesign

Five principles in the theory of data graphics produce substantial changes in graphical design. The principles apply to many graphics and yield a series of design options through cycles of graphical revision and editing.

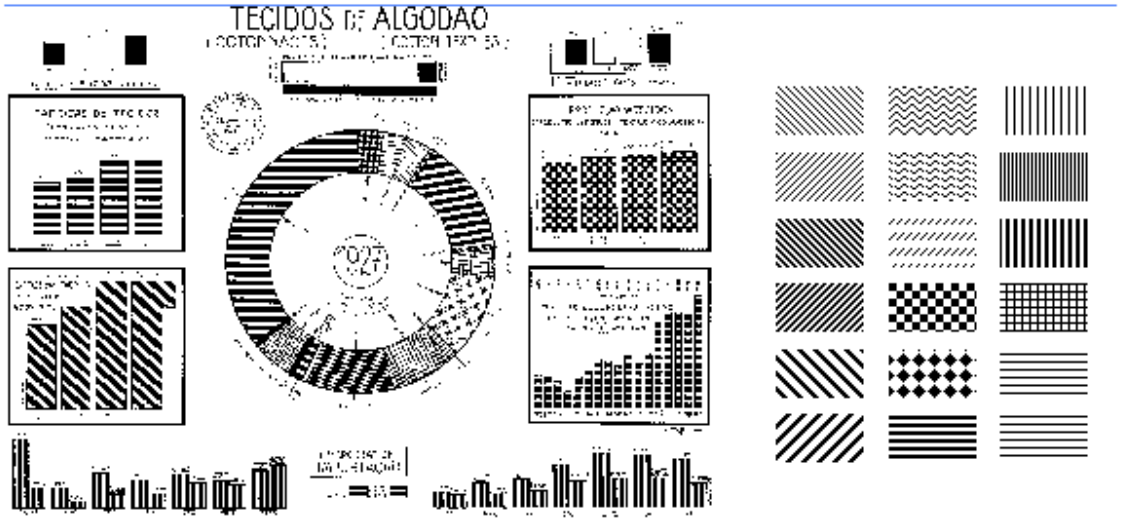
1. **Above all else show the data.**
2. **Maximize the data-ink ratio.**

   It's wrong to distort the data measures—the ink locating values of numbers—in order to make an editorial comment or fit a decorative scheme.
3. **Erase non-data-ink.**
4. **Erase redundant data-ink.**
5. **Revise and edit.**

"Chartjunk:" vibrations, grids, and ducks

- "Chartjunk" generates no information, no sense of discovery, no wonder, and no substance.
- Forgo "chartjunk," including Moire vibrations, the grid, and the duck.

Moiré Vibration Chartjunk



From Tufte, *The Visual Display of Quantitative Information*

Data-ink maximization and graphical design

Tufte proposes several improvements on standard graphics and argues that

1. They are necessarily better within the principles of the theory.
2. They get better as they get revised and improved and the theory shows the direction of improvements
3. They are more intelligent and we should not underestimate the intelligence of the audience.
4. The new designs may appear odd, but this is probably because we have not seen them before.

Malfunctioning graphical elements

Mobilize every graphical element, perhaps several times over, to show the data.

Data density and small multiples

Well designed small multiples are

- inevitably comparative
- deftly multivariate
- shrunken, high-density graphics
- usually based on a large data matrix
- drawn almost entirely with data-ink
- efficient in interpretation
- often narrative in content, showing shifts in the relationship between variables as the index variable changes (thereby revealing interaction or multiplicative effects).

Small multiples reflect much of the theory of data graphics:

- For non-data-ink, less is more.
- For data-ink, less is a bore.

Aesthetics and technique in data graphical design

Graphical elegance is often found in simplicity of design and complexity of data.

Attractive displays of statistical information:

- have a properly chosen format and design
- use words, numbers, and drawing together
- reflect a balance, a proportion, a sense of relevant scale
- display an accessible complexity of detail
- often have a narrative quality, a story to tell about the data
- are drawn in a professional manner, with the technical details of production done with care
- avoid content-free decoration, including chartjunk

Data graphics are paragraphs about data and should be treated as such.

Conclusion

- Design is choice.
- The theory of the visual display of quantitative information consists of principles that generate design options and that guide choices among options. The principles should not be applied rigidly or in a peevish spirit; they are not logically or mathematically certain; and it is better to violate any principle than to place graceless or inelegant marks on paper. Most principles of design should be greeted with some skepticism, for word authority can dominate our vision, and we may come to see only through the lenses of word authority rather than with our own eyes.
- What is to be sought in designs for the display of information is the clear portrayal of complexity. Not the complication of the simple; rather the task of the designer is to give visual access to the subtle and the difficult –that is, the revelation of the complex.

Concluding remarks

- The purpose of data analytics is insight
- Think very carefully about how to present the insight that you have gained from data analysis
- Do it well – if you don't you will have produced perhaps great results but in poor packaging and poor "user interface" – essentially useless

In class Videos

- The Art of Data Visualization | Off Book | PBS Digital Studios
  - https://youtu.be/AdSZJzb-aX8
- Edward Tufte: Beautiful Evidence (Highlights)
  - http://youtu.be/Th_1azZA2OY