

Practical Data Visualization

March 18, 2015

COMPSCI 216:
Everything Data

Angela Zoss

Data Visualization Coordinator
Data and Visualization Services



WHY VISUALIZE?

Preserve complexity

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Preserve complexity

Anscombe's Quartet

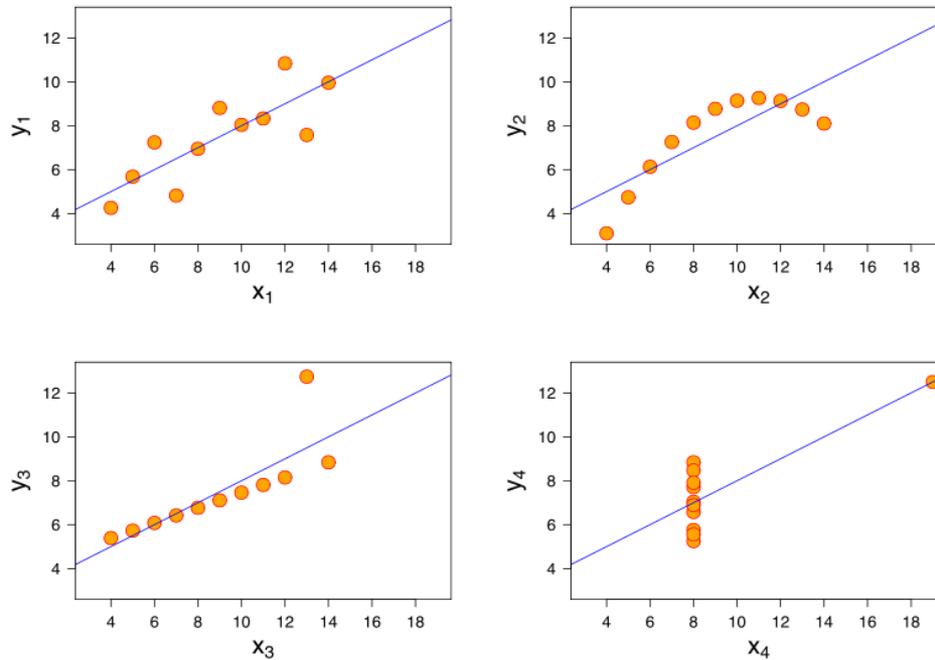
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Property	Value
Mean of x	9 (exact)
Variance of x	11 (exact)
Mean of y	7.50 (to 2 decimal places)
Variance of y	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y	0.816 (to 3 decimal places)
Linear regression line	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)

http://en.wikipedia.org/wiki/Anscombe%27s_quartet

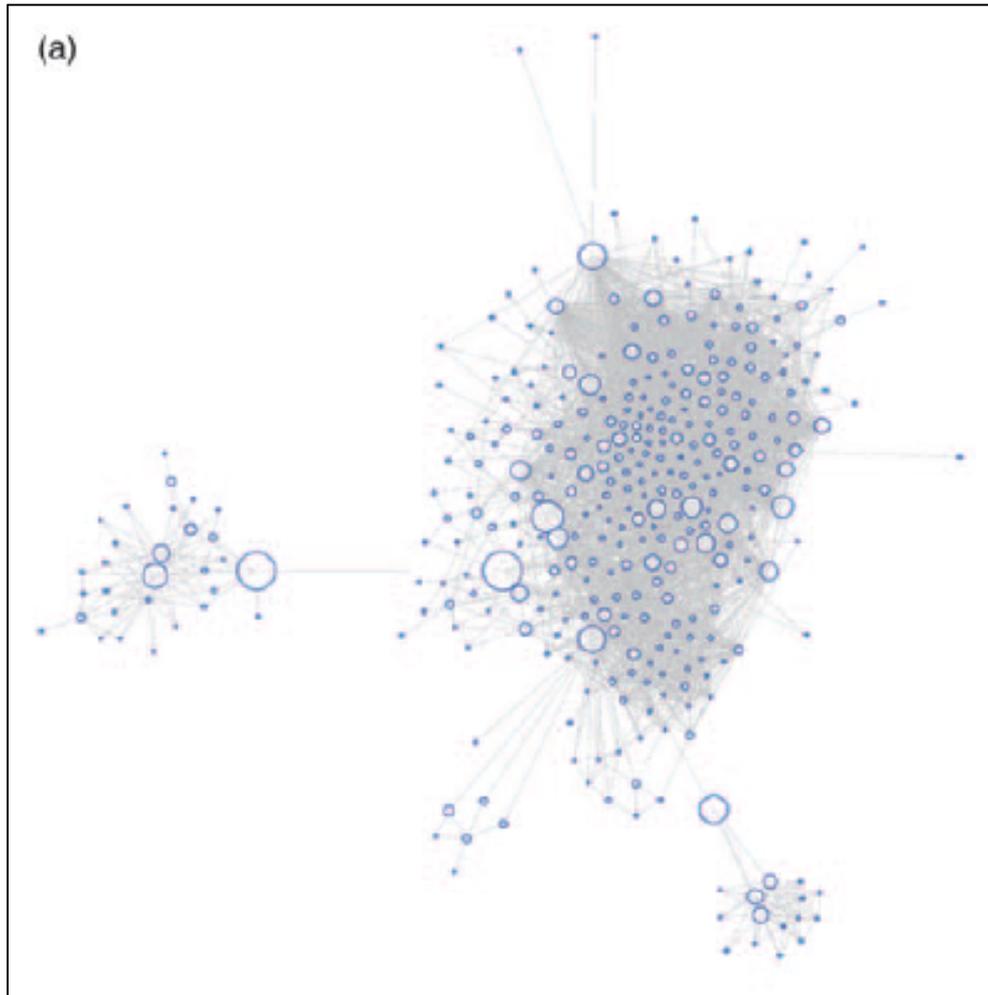
Preserve complexity

Anscombe's Quartet



http://en.wikipedia.org/wiki/Anscombe%27s_quartet

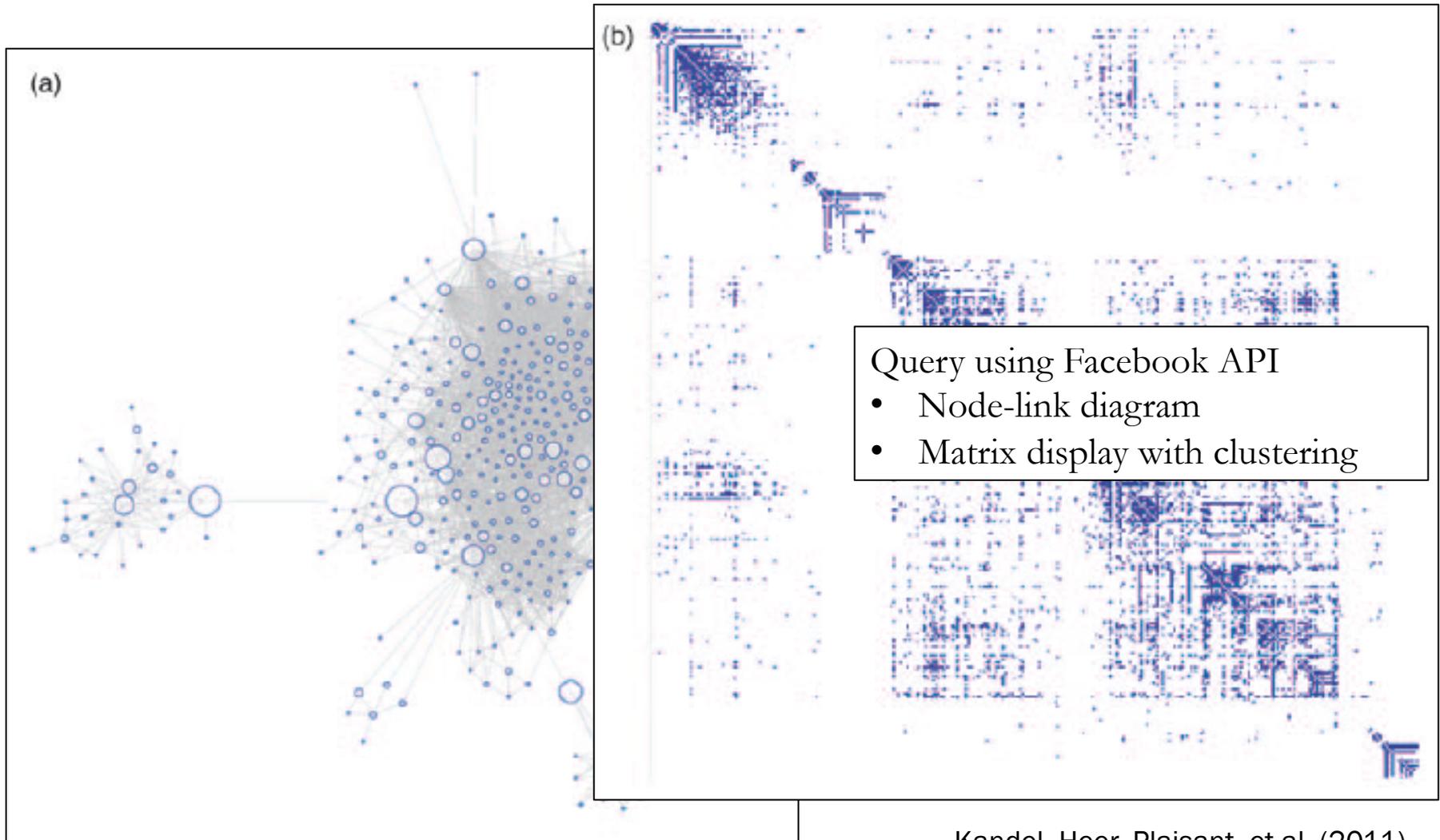
Evaluate data quality



Query using Facebook API

- Node-link diagram

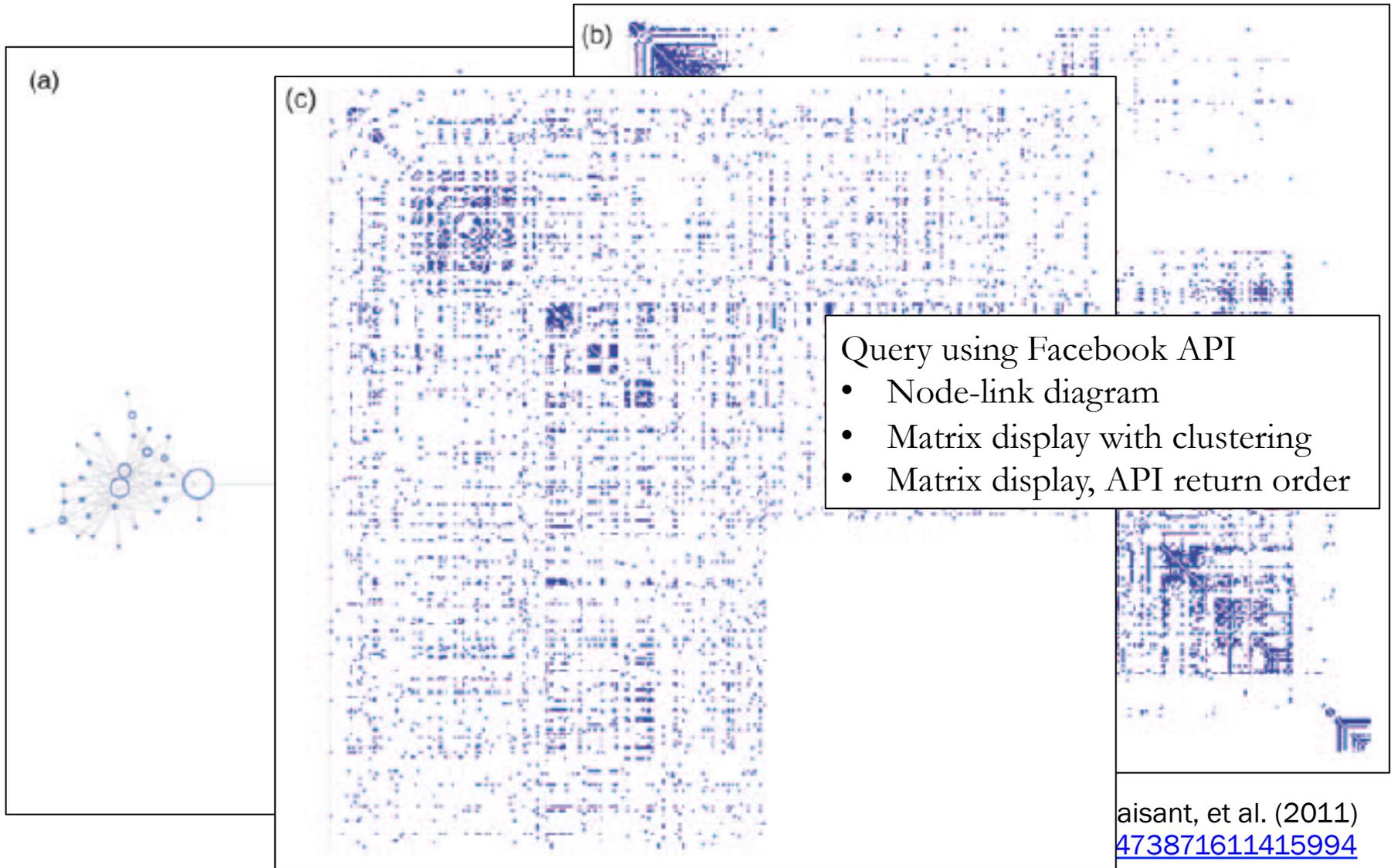
Evaluate data quality



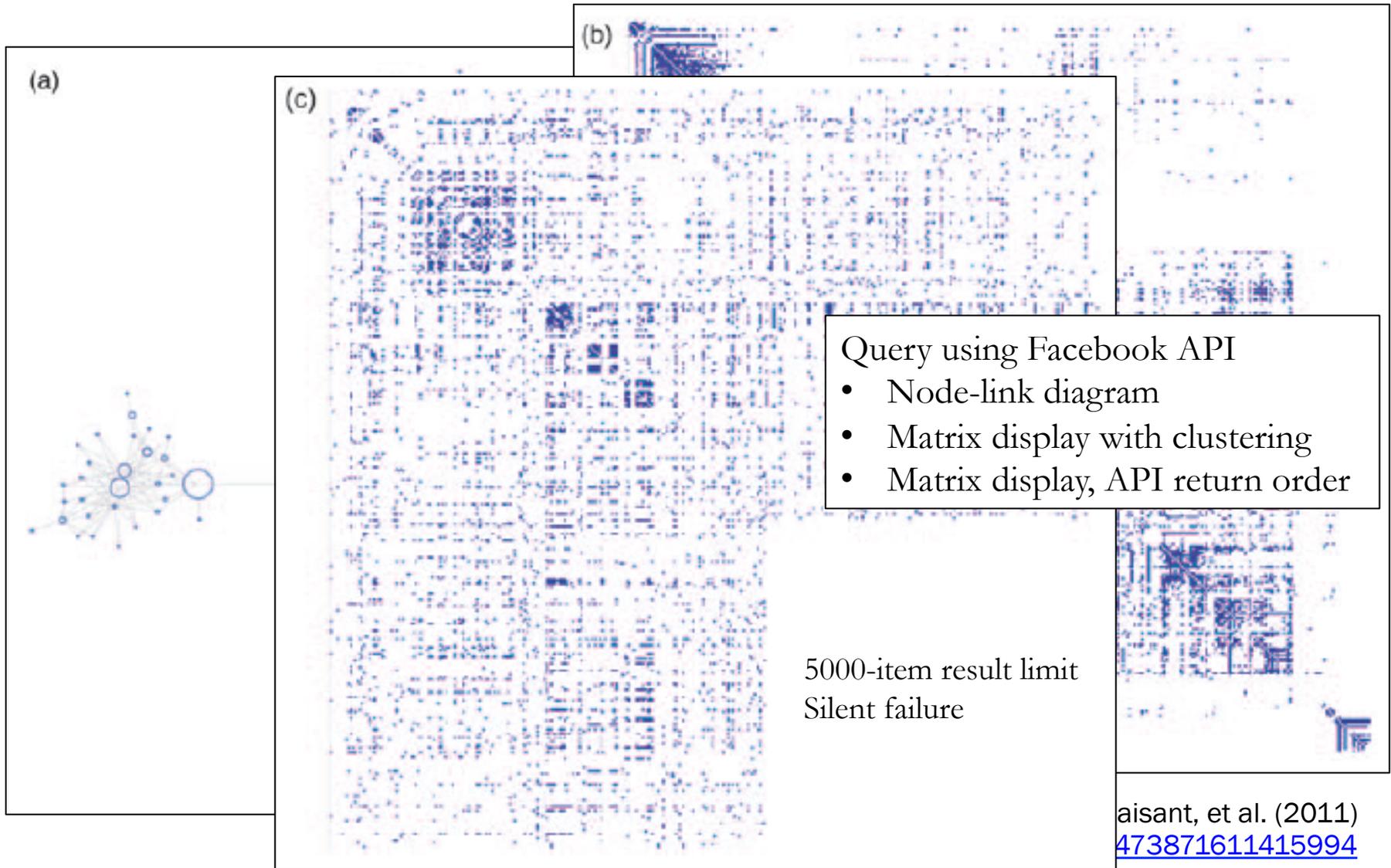
Kandel, Heer, Plaisant, et al. (2011)

<http://dx.doi.org/10.1177/1473871611415994>

Evaluate data quality



Evaluate data quality



Tell a story



Hans Rosling – The River of Myths

<http://www.youtube.com/watch?v=Owll-dwh-bk>

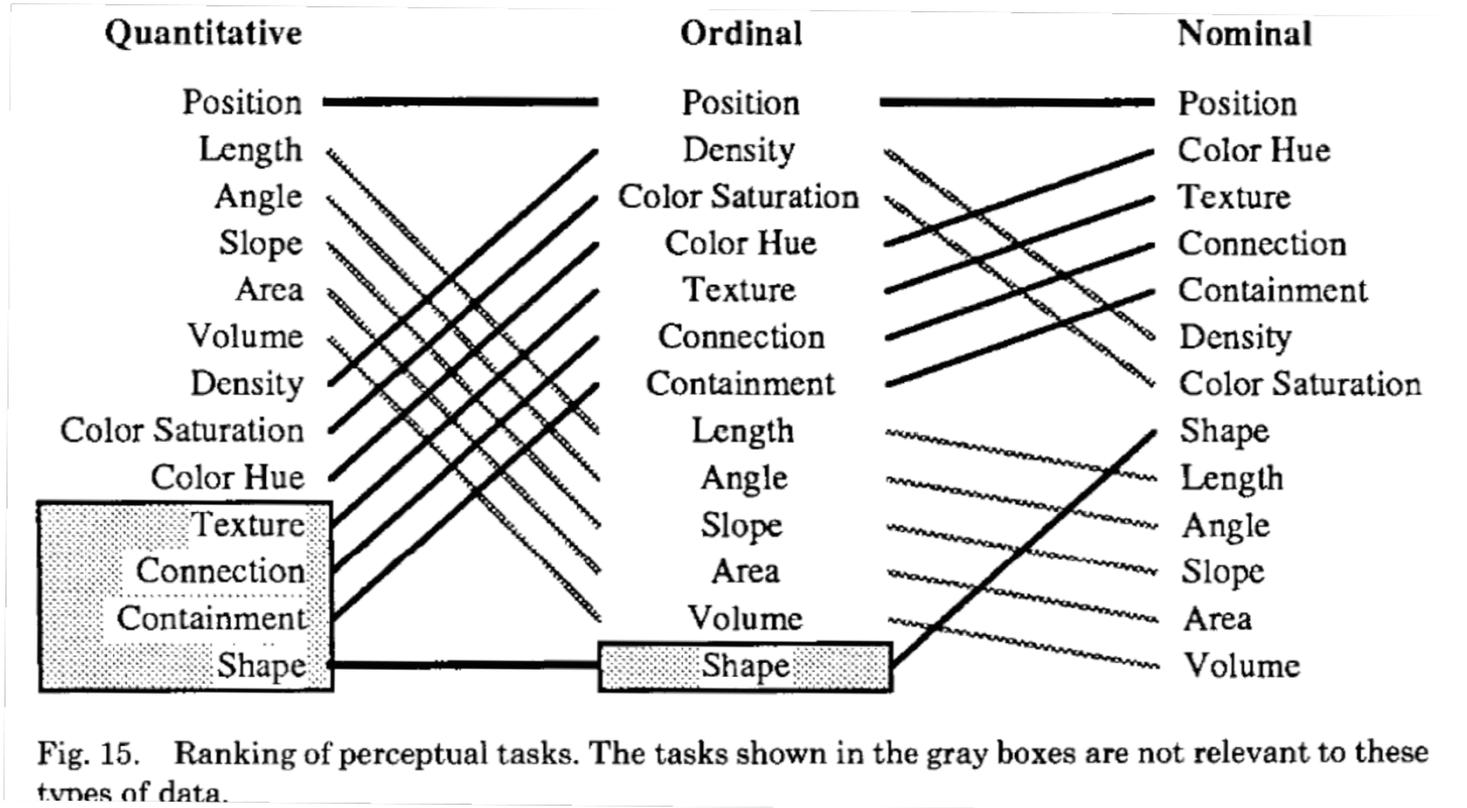
<http://www.gapminder.org/>

CREATING A VISUALIZATION

From Data to Graphic

- What data types are present in the data source?
Categorical? Numerical? Relational?

Matching Data Types to Visual Elements



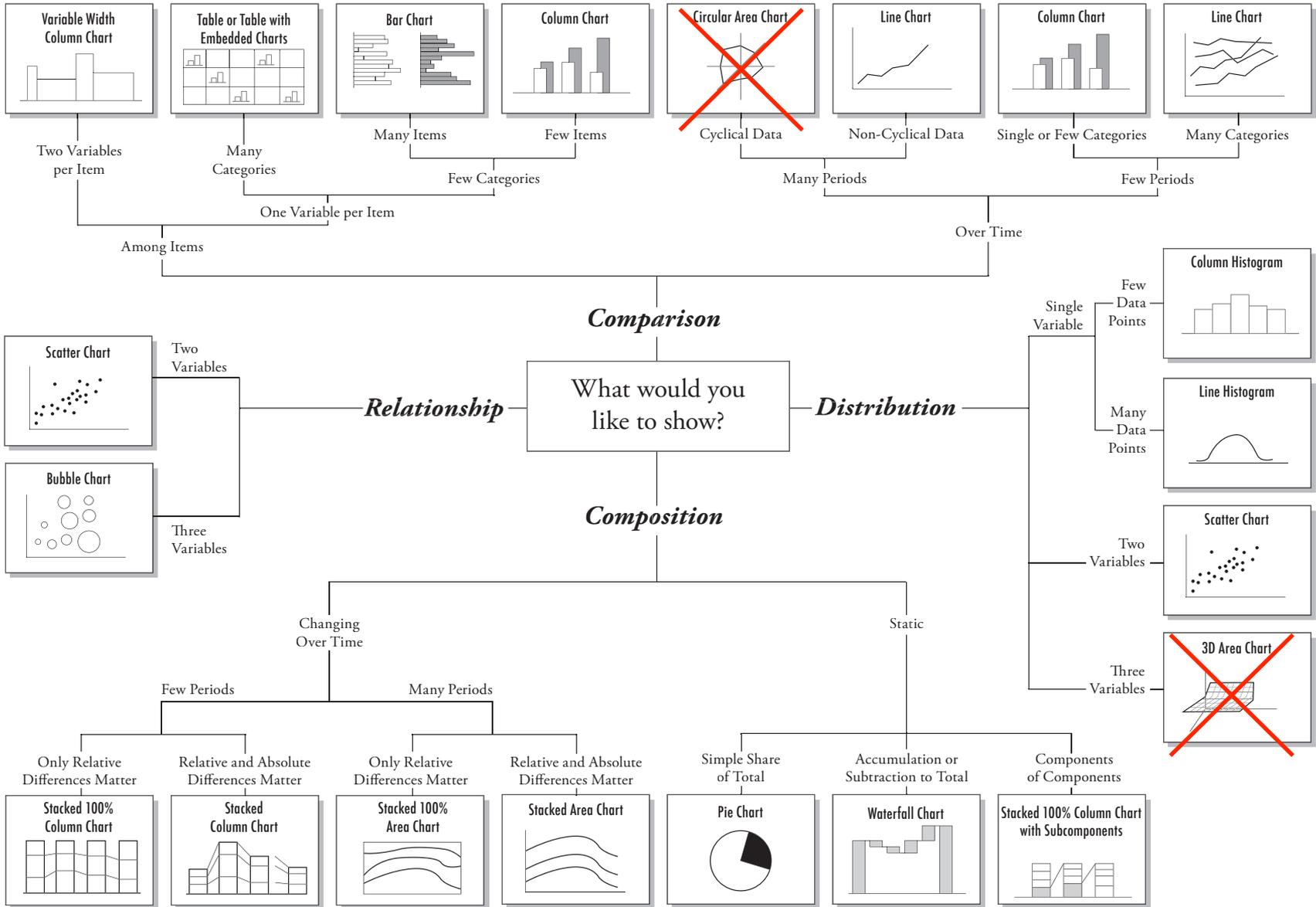
Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2), 110-141.

<http://dx.doi.org.proxy.lib.duke.edu/10.1145/22949.22950>

From Data to Graphic

- What data types are present in the data source?
- What type of analysis do you want to support?
Are you looking for correlations? Distributions?

Chart Suggestions—A Thought-Starter



From Data to Graphic

- What data types are present in the data source?
- What type of analysis do you want to support?
- What visualization type seems to be the best fit for the goal?

Do you want the visualization to be accessible for a broad audience? Flashy and engaging? Convincing?

POSITION IS EVERYTHING.

@moritz_stefaner

Basic tips

- Rotated text is harder to read
- People are very good at reading x/y position, bar length
- People are not as good at reading angles, areas
- Avoid overlap by filtering, aggregating, leaving space

<http://guides.library.duke.edu/topten>

COLOR IS DIFFICULT.

@moritz_stefaner

Basic tips

For categorical variables:

- People have trouble differentiating between more than 5-7 hues (colors)

For numerical variables:

- People have trouble differentiating between more than 5-7 shades
- Rainbow color gradients are very problematic

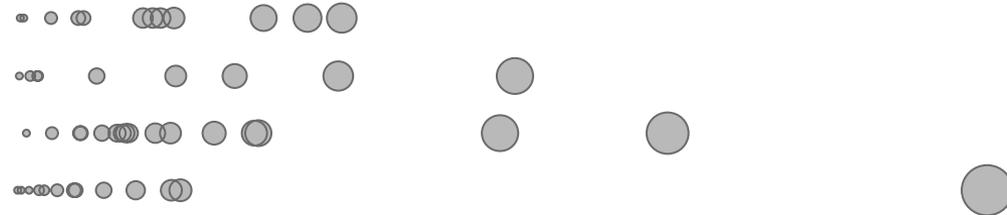
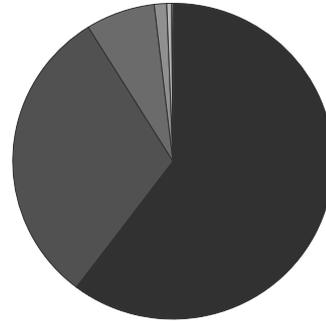
For highest contrast, only use color to highlight

<http://guides.library.duke.edu/topten>

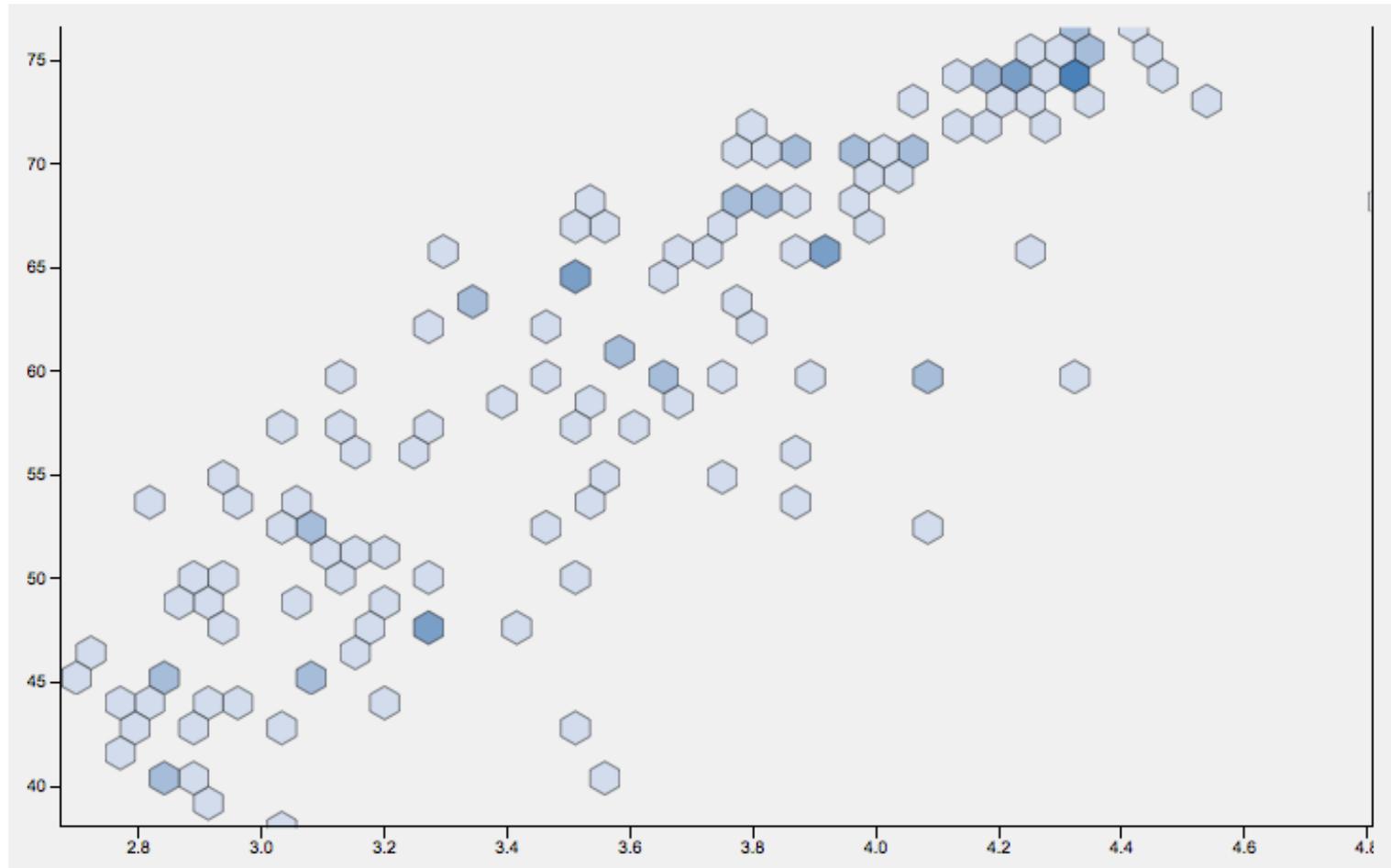
VISUALIZATION TYPES

http://guides.library.duke.edu/vis_types

Basic charts and graphs

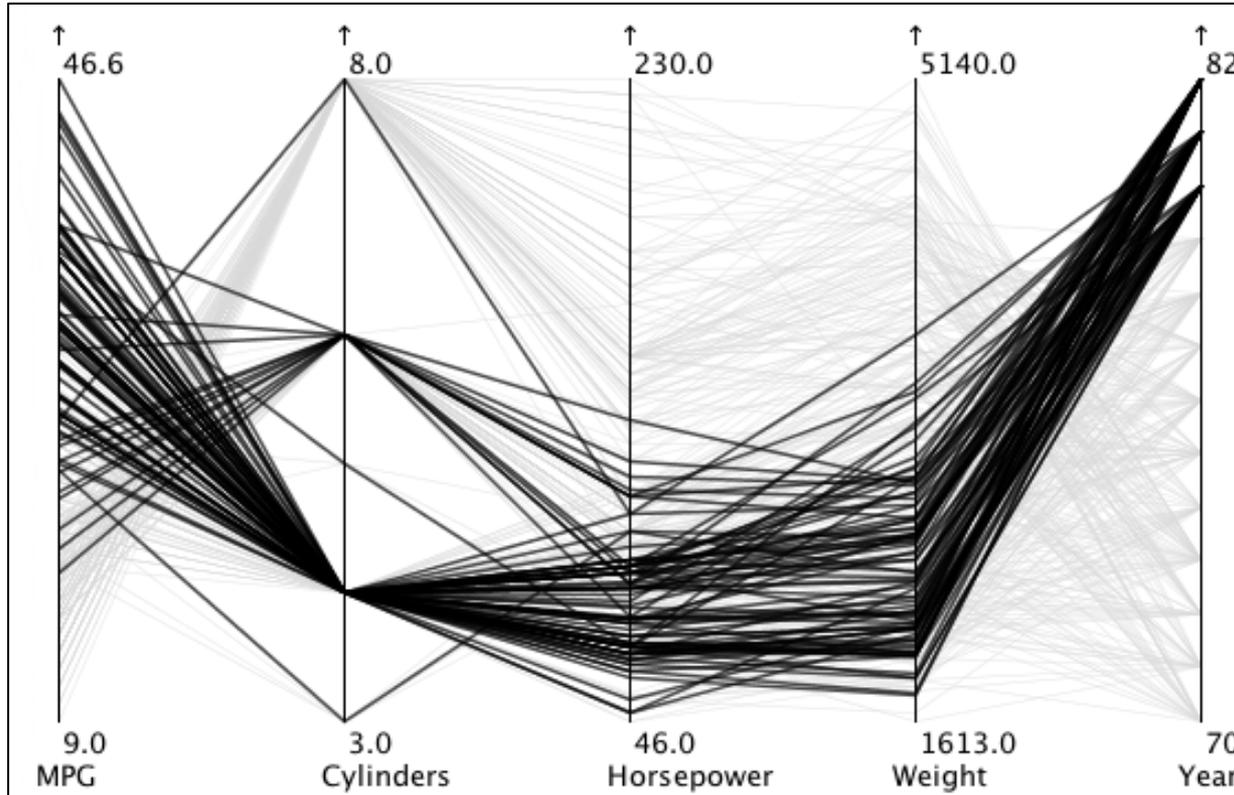


Binned Scatterplot



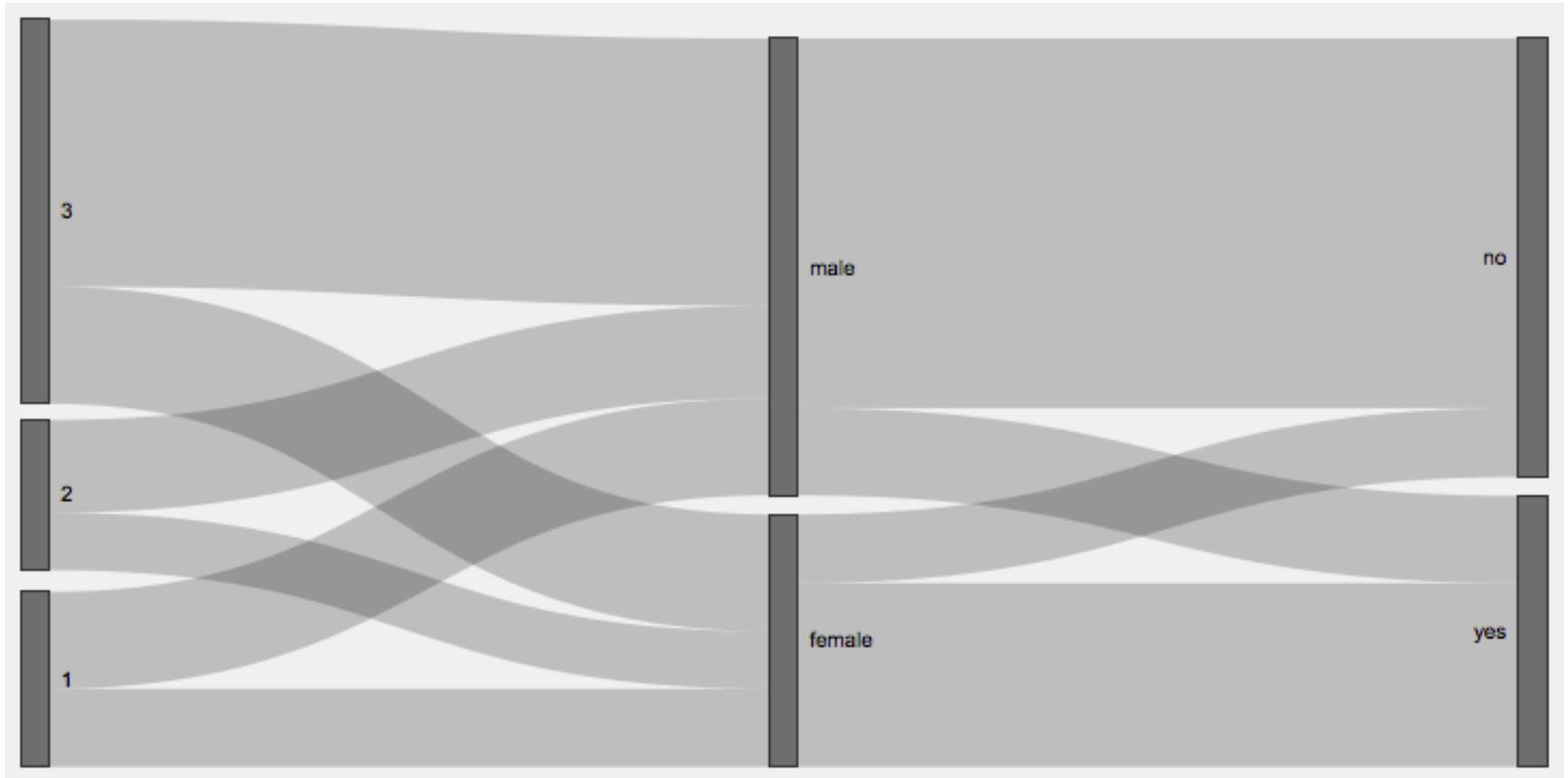
<http://raw.densitydesign.org/>

Parallel Coordinates



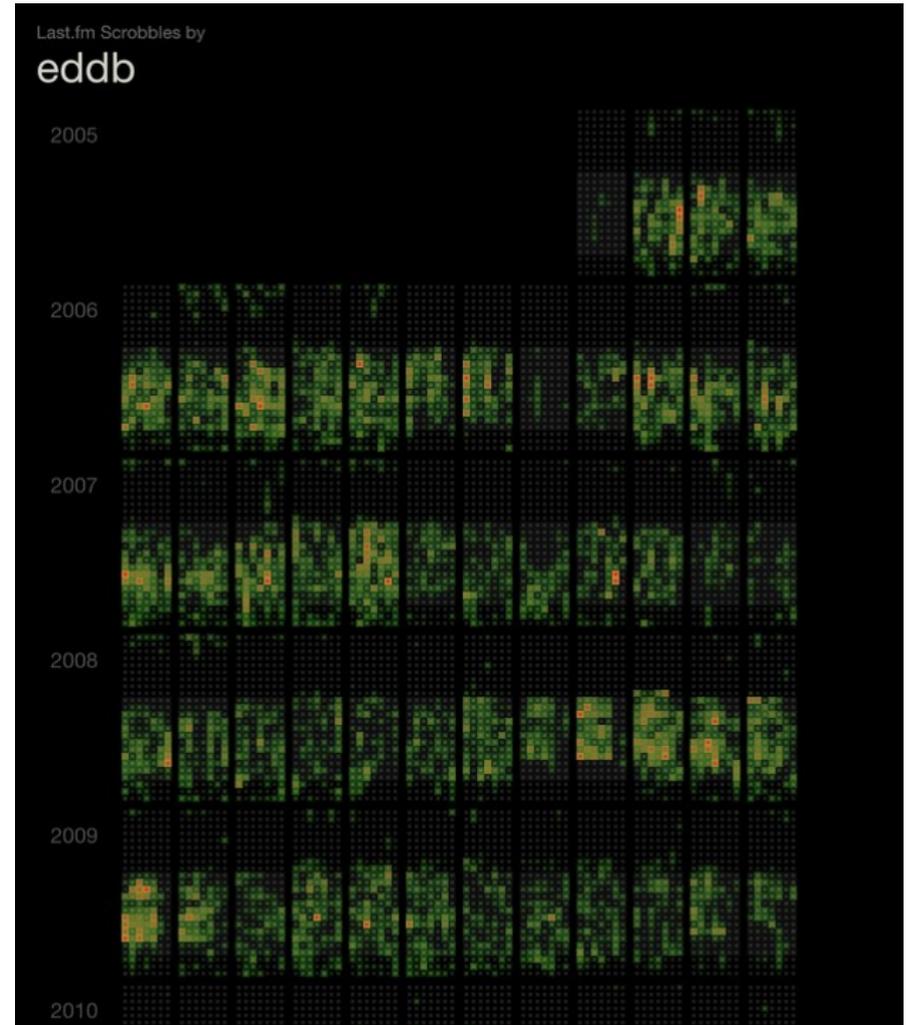
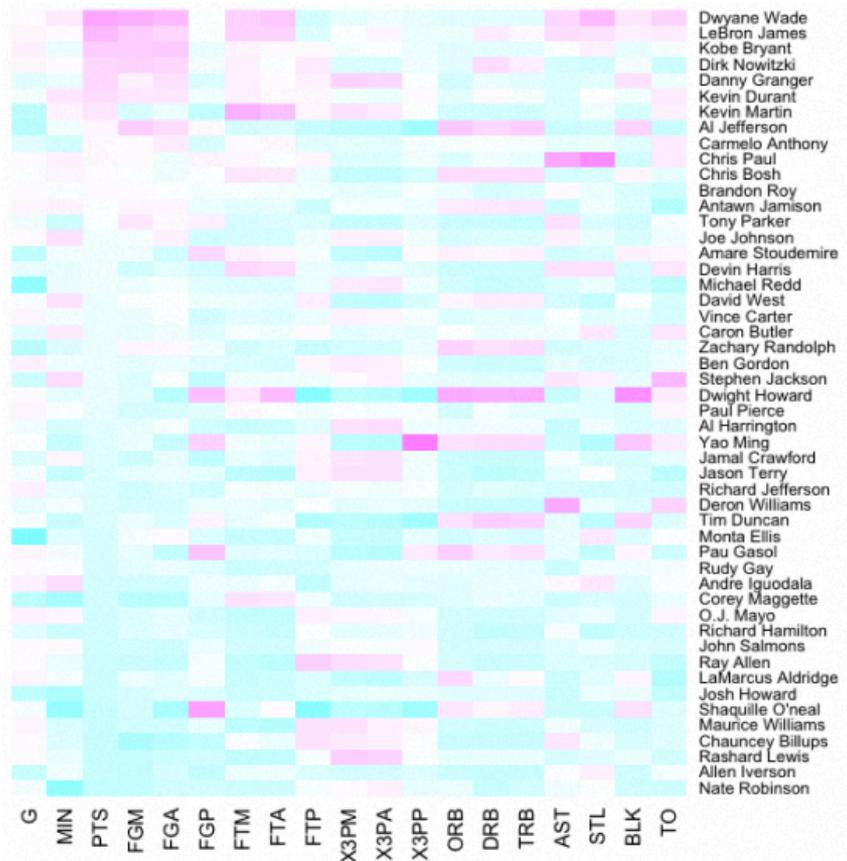
<http://eagereyes.org/techniques/parallel-coordinates>

Sankey/Alluvial Diagram



<http://raw.densitydesign.org/>

Heat Maps

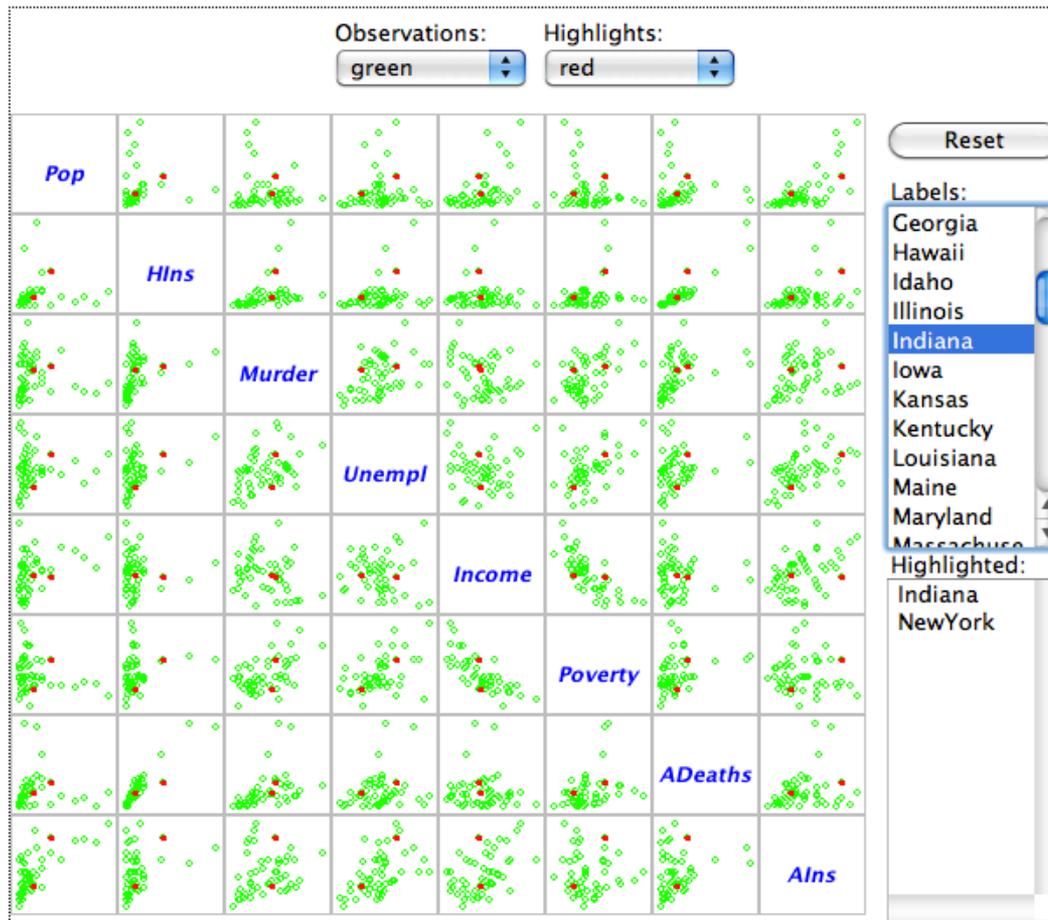


<http://flowingdata.com/2010/01/21/how-to-make-a-heatmap-a-quick-and-easy-solution/>

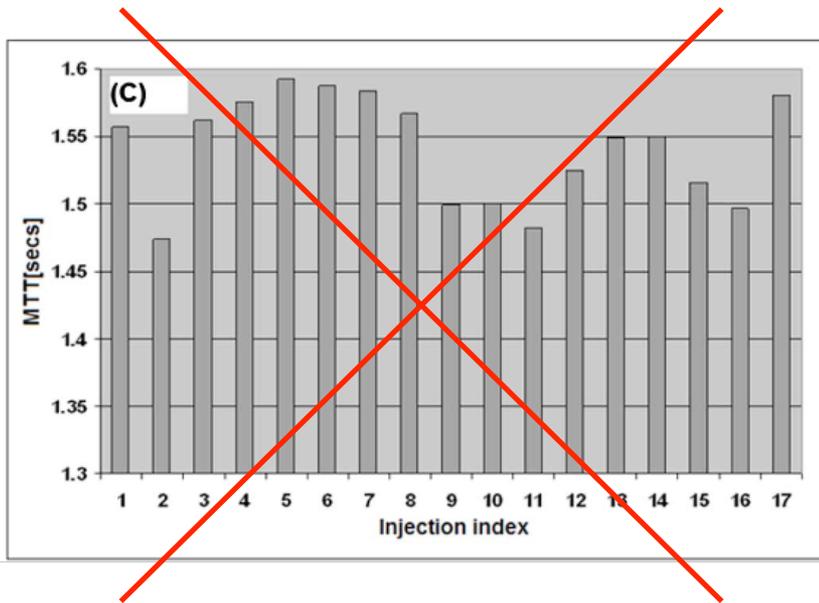
<http://flowingdata.com/2011/09/13/last-fm-scrobbles-as-calendar-heat-map/>

Pairs Plots

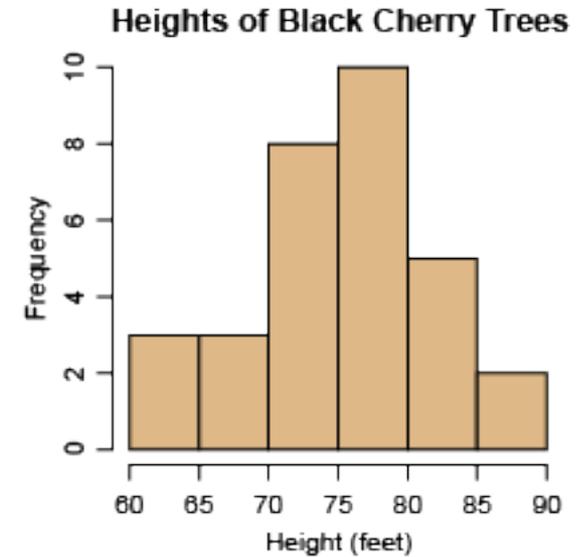
Dynamic Pairs Plot: <http://www.stat.sc.edu/~west/bradley/census.html>



Showing Distributions

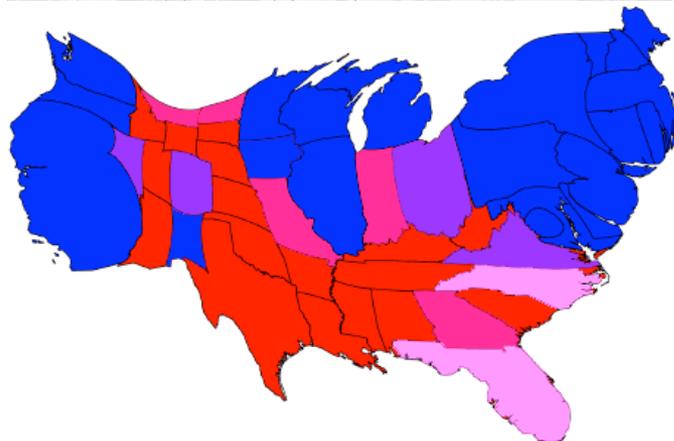
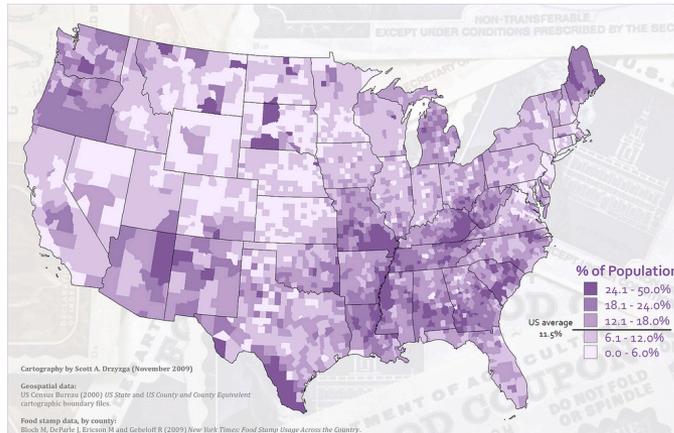


One-dimensional
scatter plot

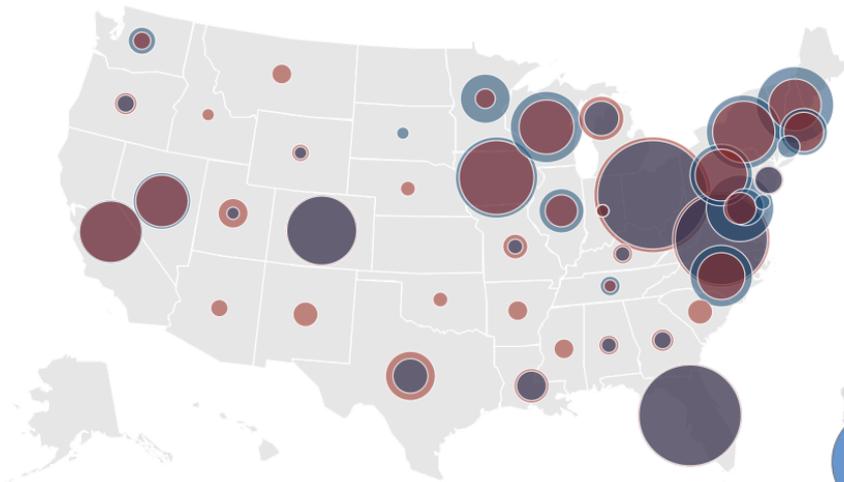


Histogram

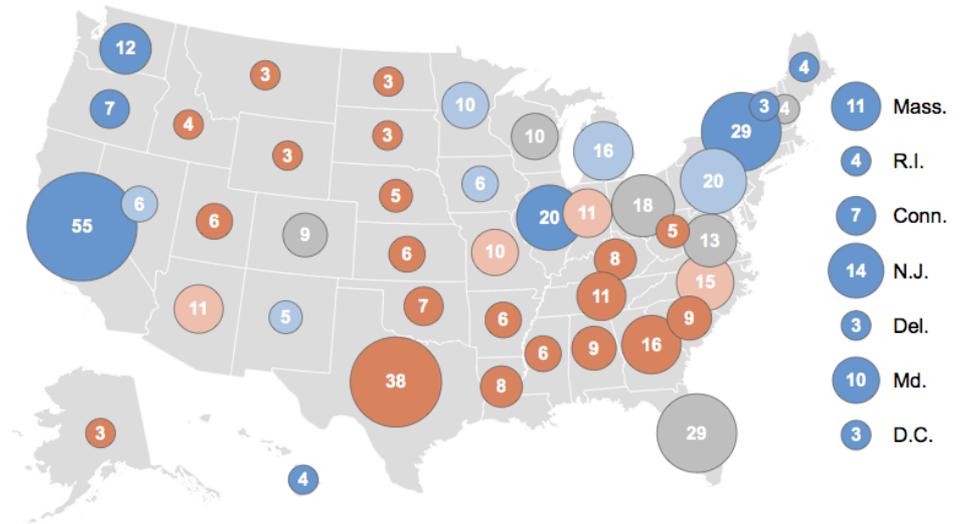
Showing Space



Proportional symbol

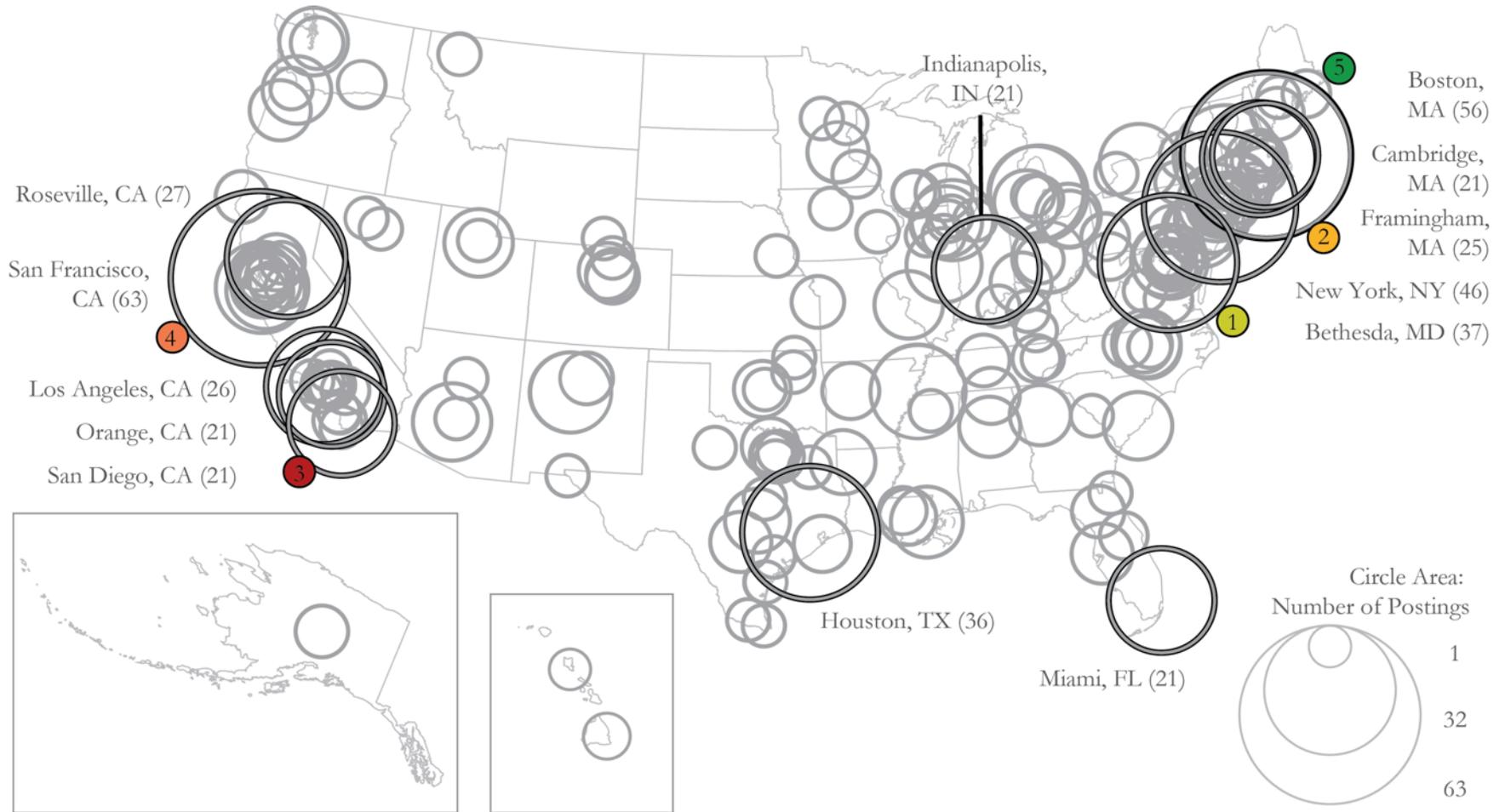


<http://wapo.st/2012-campaignvisits>

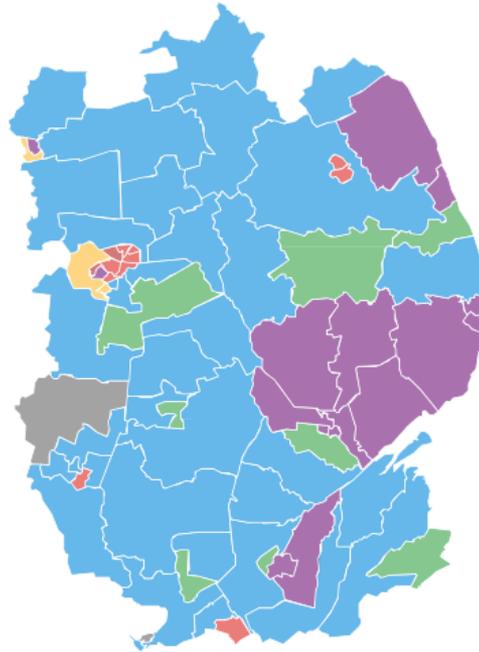


<http://ti.me/RQaRH9>

Proportional symbol



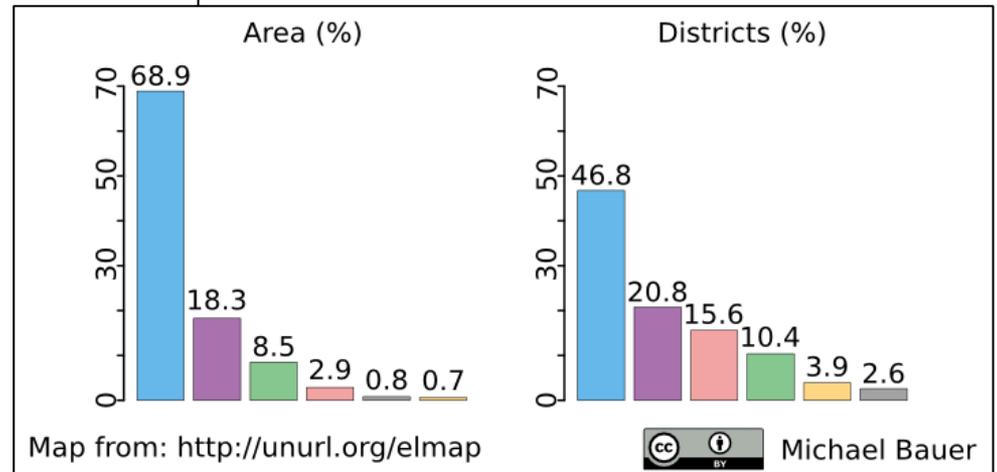
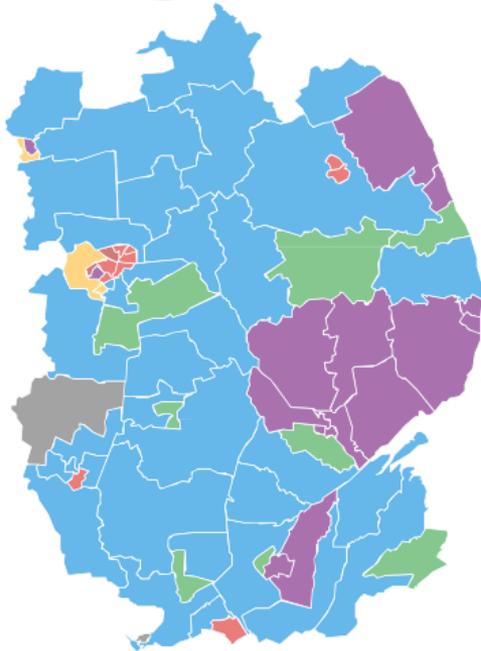
Choropleth



https://twitter.com/mihi_tr/status/330261204083810304/photo/1

Choropleth

Maps are great, except...



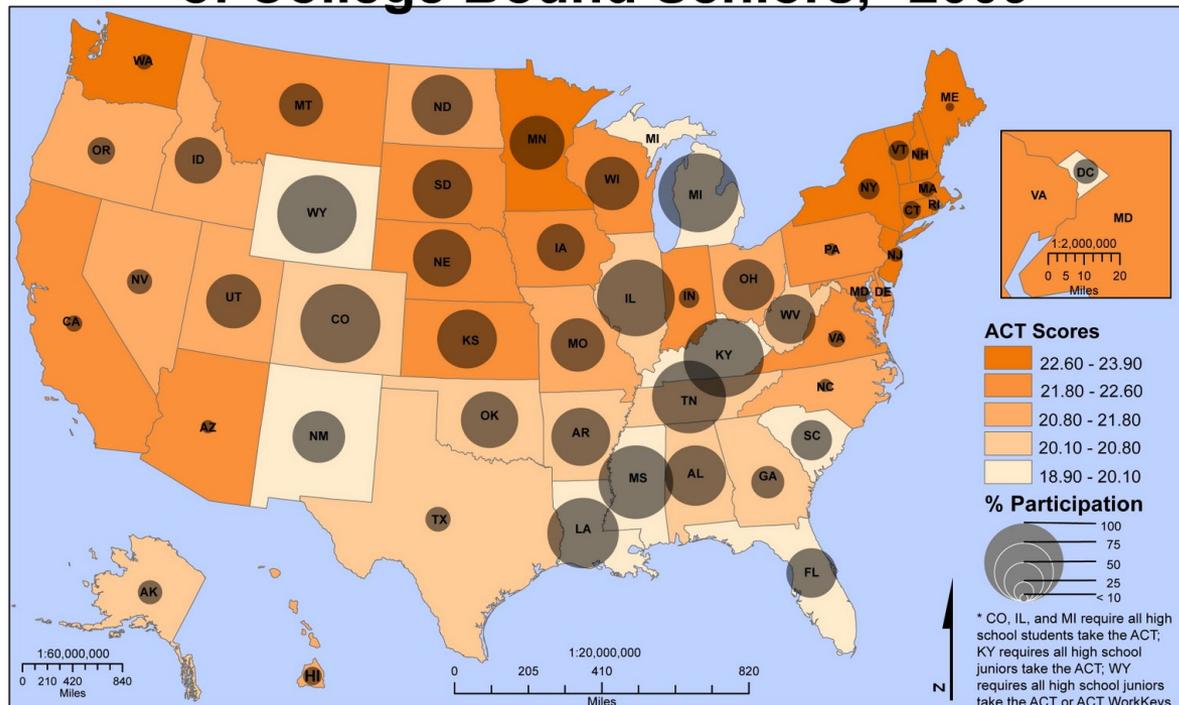
... they give a wrong idea about quantities.

This is the map of the Lincolnshire election results. It shows a clear dominance of the conservative (blue) party candidates. Looking closely at the number of districts won shows a different picture though: The problem: Some districts have more area than others.

And don't make users do "visual math."

<http://eagereyes.org/criticism/visual-math-wrong>

Mean ACT Scores and Participation of College Bound Seniors,* 2009

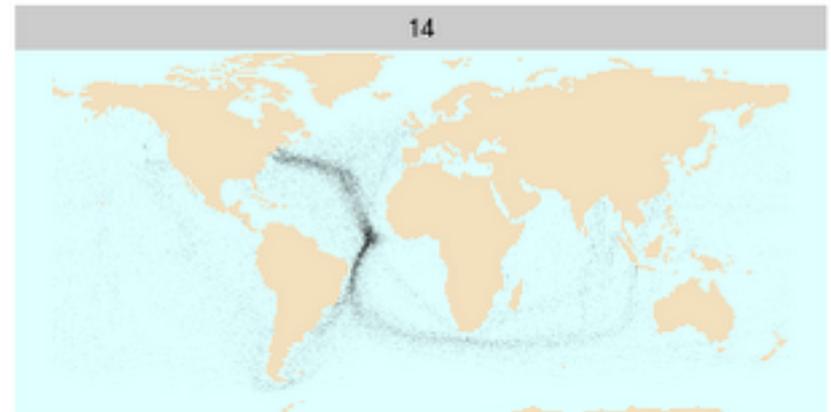
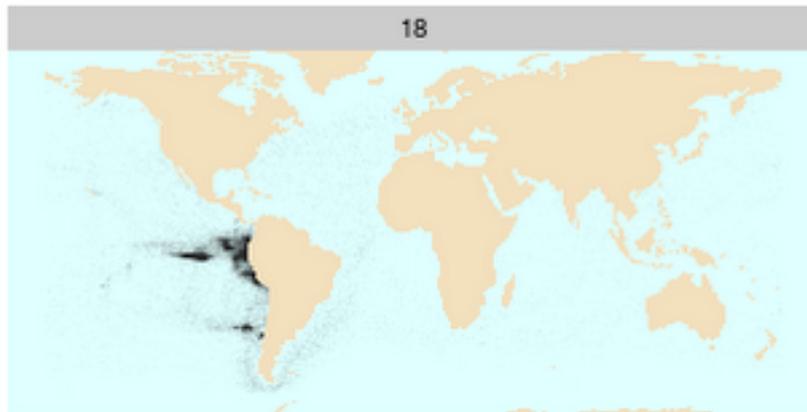
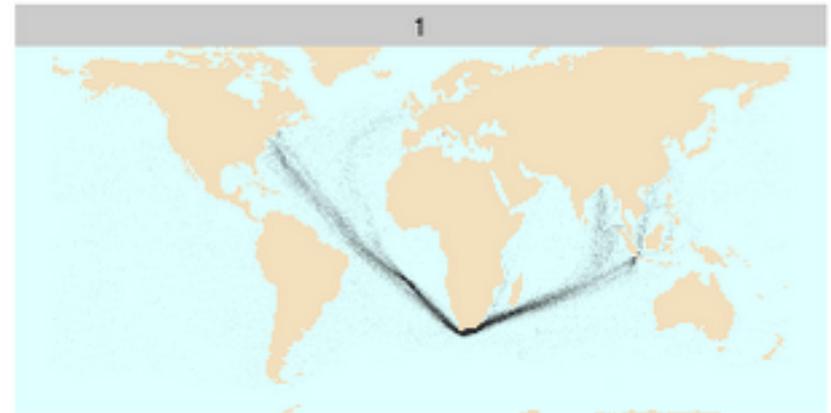
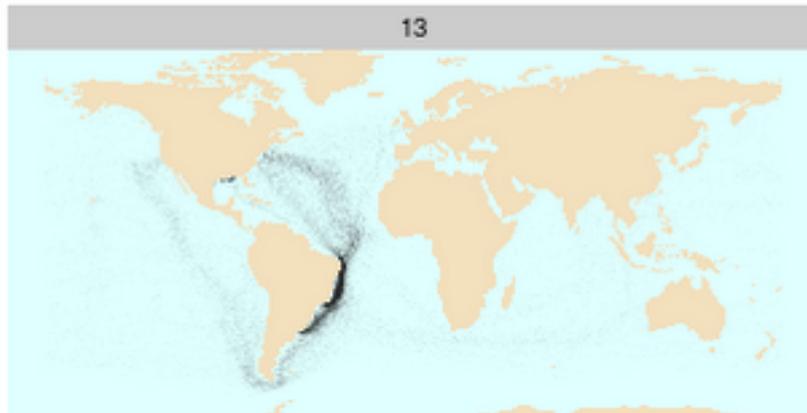


Data Source: 2009 Average ACT Scores by State
<http://www.act.org/news/data/09/states.html>
Map Projection: Albers Equal Area Conic

Map Created by Lisa Capps, 2010

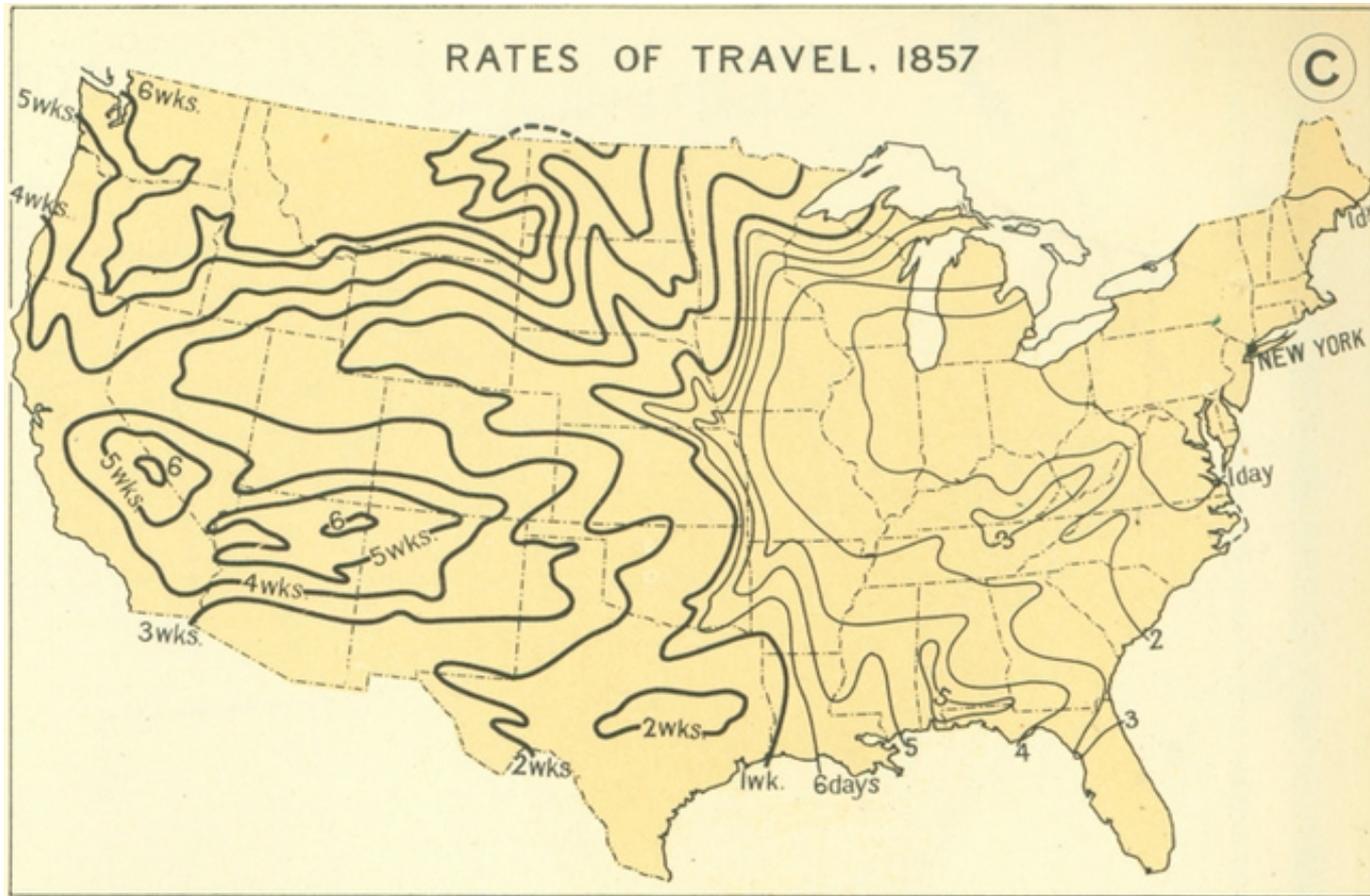
<http://enb105-2012s-rw.blogspot.com/2012/02/lab-two-mapping-exercice.html>

Common Routes Based on Ship Log Data



<http://bit.ly/1i3PSQh>

Atlas of the Historical Geography of the United States (1932)



<http://bit.ly/1qv0Lvo>

Possible tools for mapping

- ArcGIS
- QGIS
- Tableau Public
- CartoDB
- Google Fusion Tables
- Google Earth
- GeoCommons
- JavaScript
 - D3
<http://d3js.org/>
 - Leaflet
<http://leafletjs.com/>
 - Kartograph
<http://kartograph.org/>
 - Polymaps
<http://polymaps.org/>
 - Google Maps API
[https://
developers.google.com/
maps/documentation/
javascript/](https://developers.google.com/maps/documentation/javascript/)
- Very basic:
 - Google Spreadsheets
 - BatchGeo
<http://batchgeo.com/>
 - OpenHeatMap
[http://
www.openheatmap.com/](http://
www.openheatmap.com/)

See also:

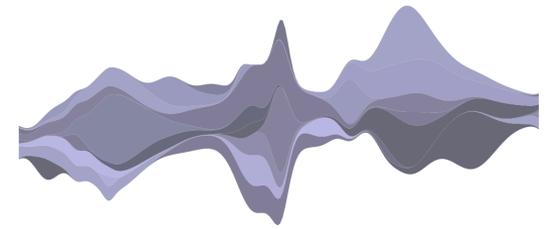
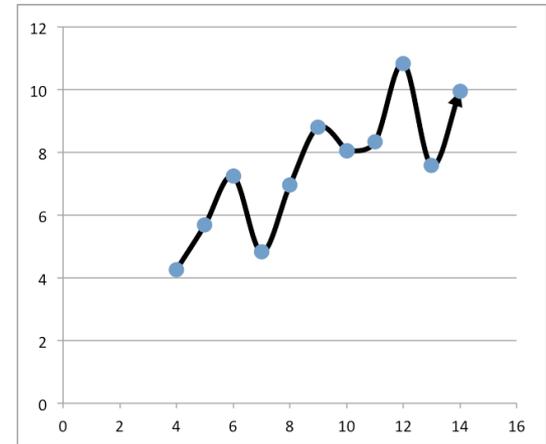
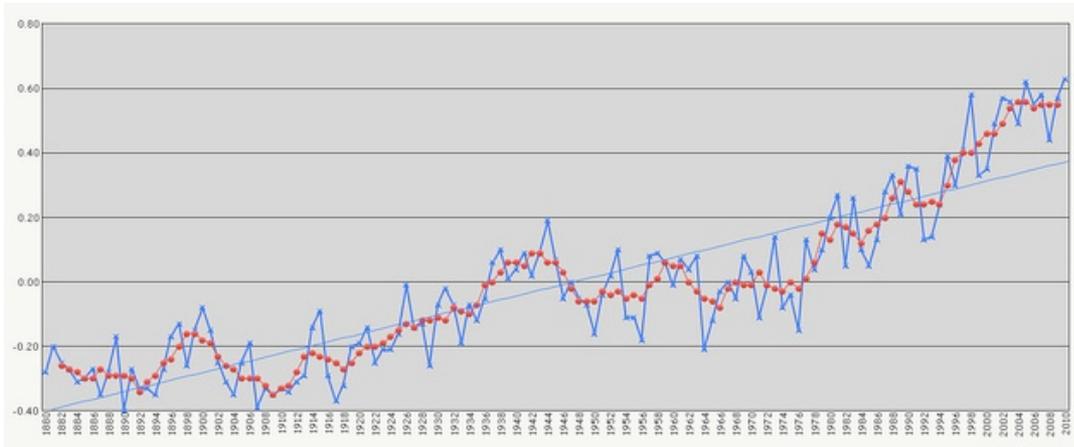
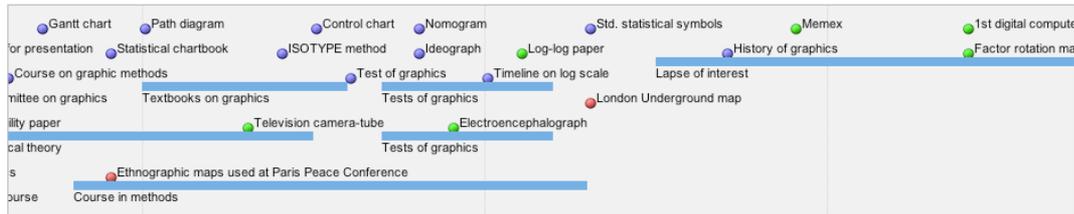
<http://library.duke.edu/data/gis>

<https://github.com/veltman/learninglunches/tree/master/maps>

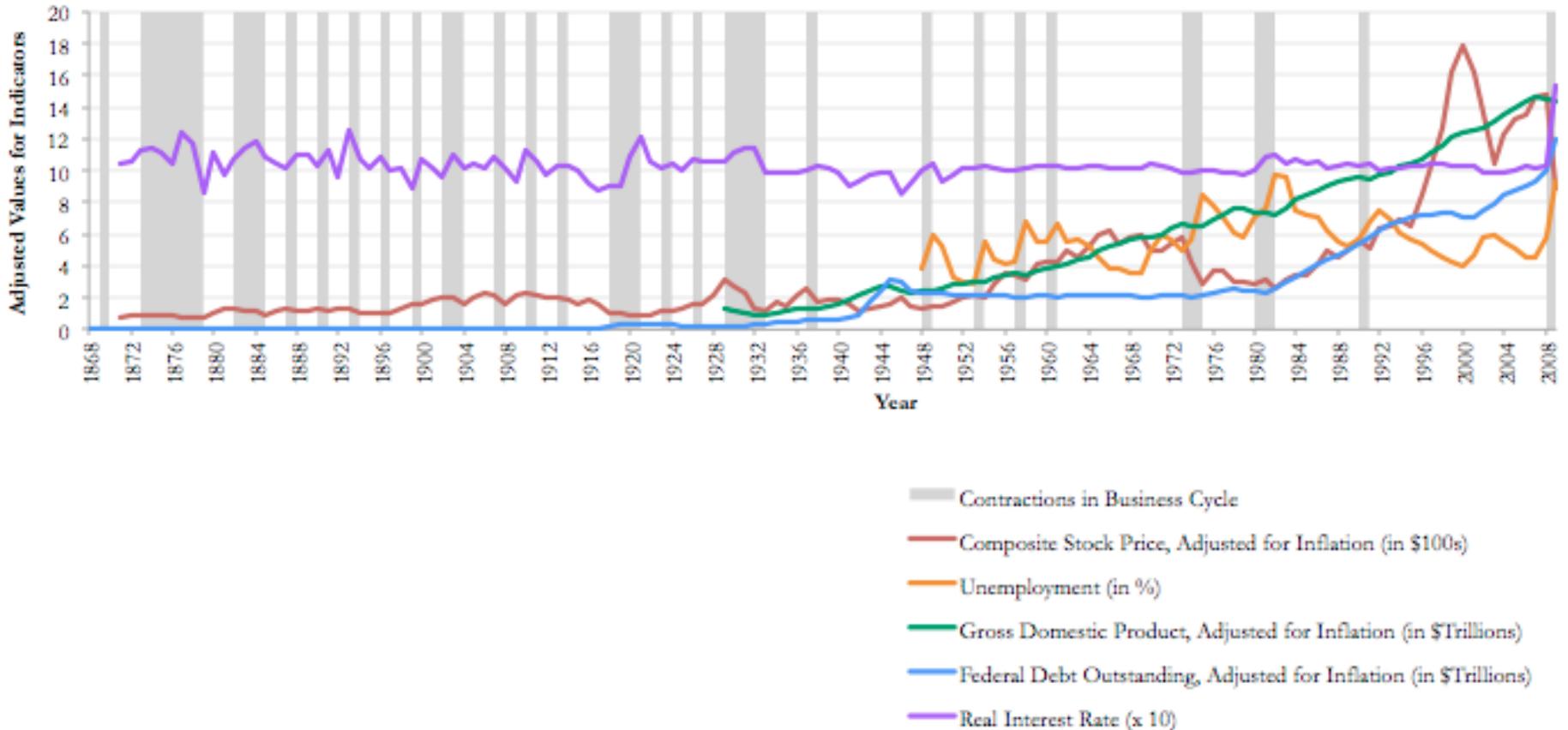
For congress data in Tableau

- <http://www.tableau.com/public/blog/2012/10/congressional-districts-1696>
- http://onlinehelp.tableau.com/current/pro/online/en-us/help.htm#maps_geographicroles.html

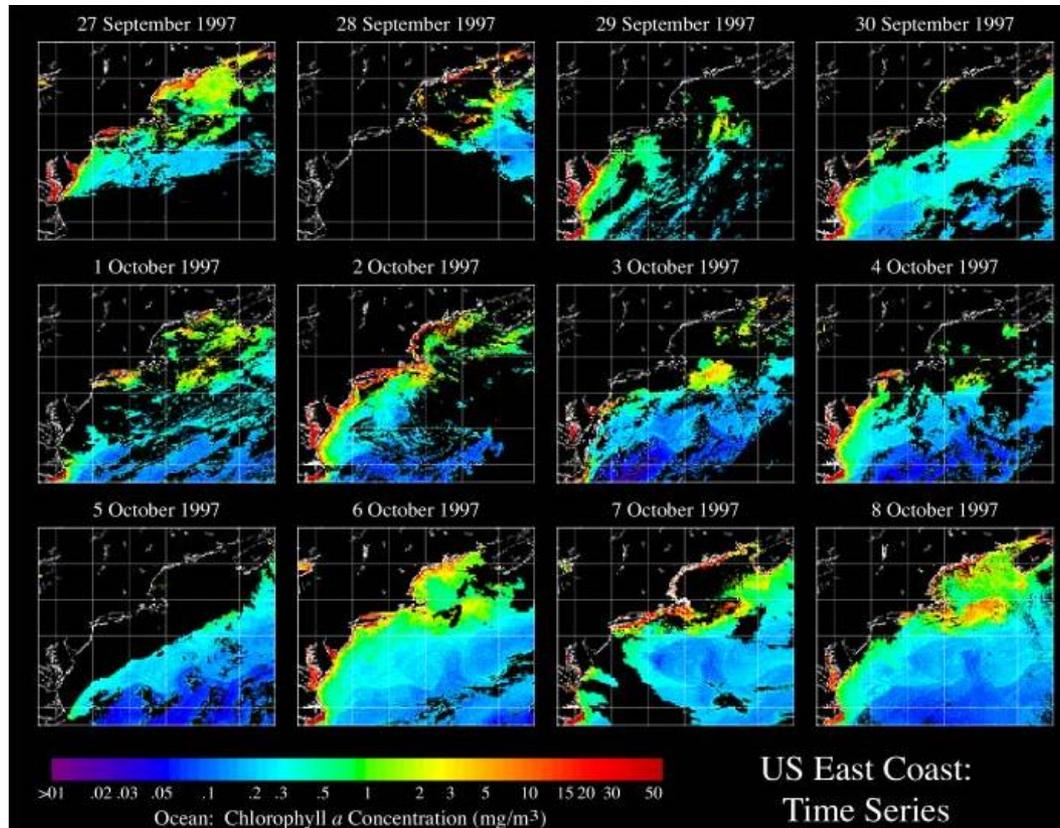
Showing Time



Economic indicators over time

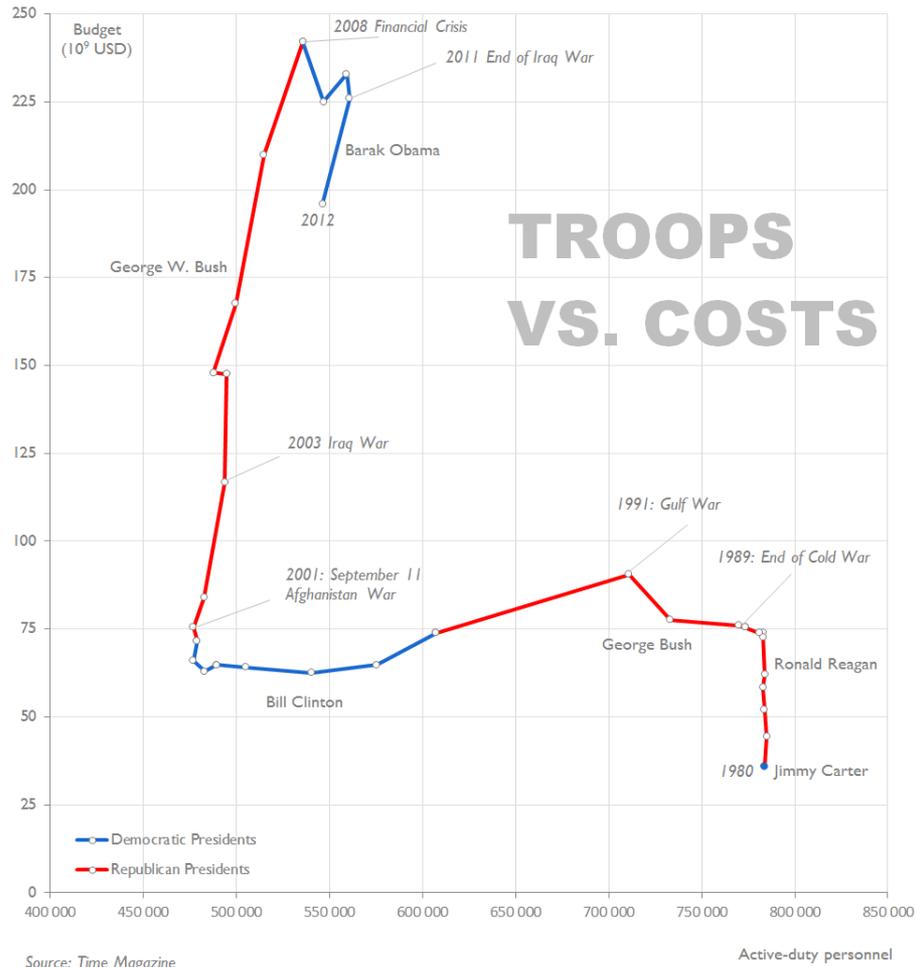


Time series of 2D data set



http://seawifs.gsfc.nasa.gov/SEAWIFS/BACKGROUND/Gallery/time_series.jpg

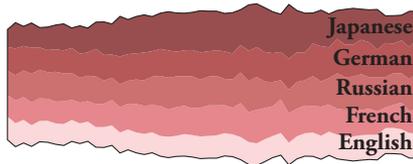
Connected Scatterplot



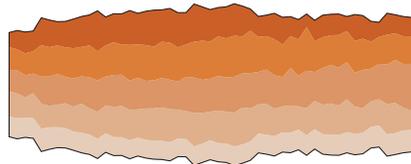
<http://bit.ly/1AADeyq>

Stream graphs

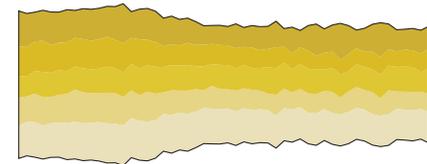
diseases of the circulatory system



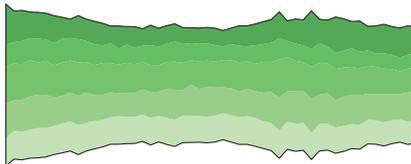
diseases of the digestive system



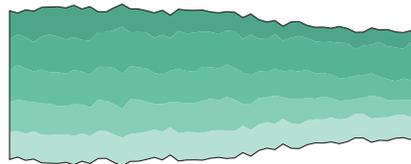
endocrine, nutritional and metabolic diseases



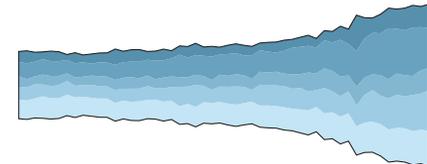
infectious and parasitic diseases



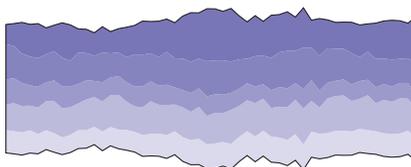
injury, poisoning and other external causes



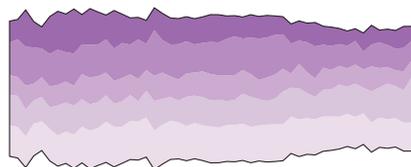
mental and behavioral disorders



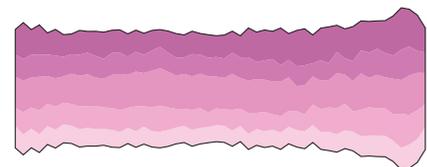
cancer (neoplasms)



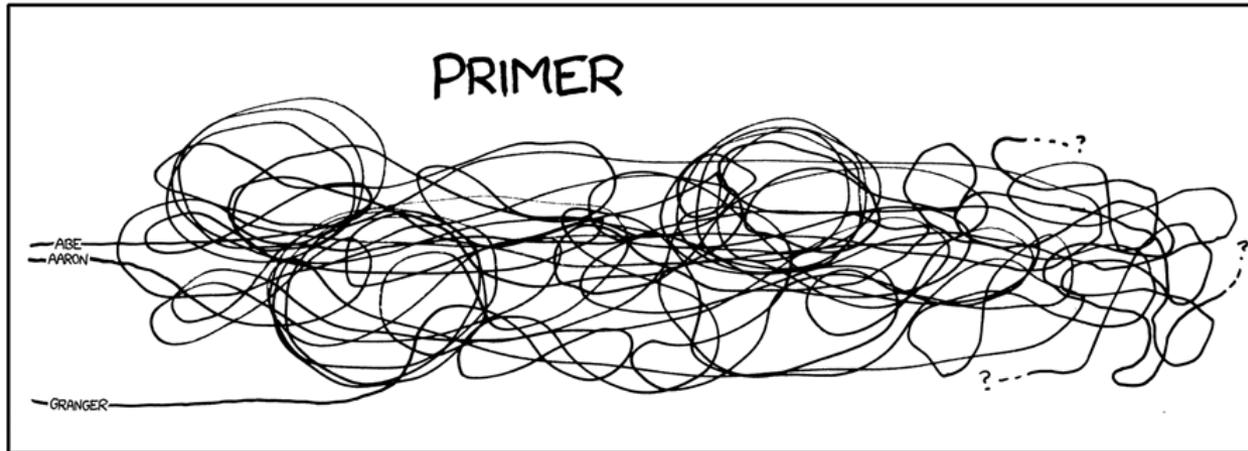
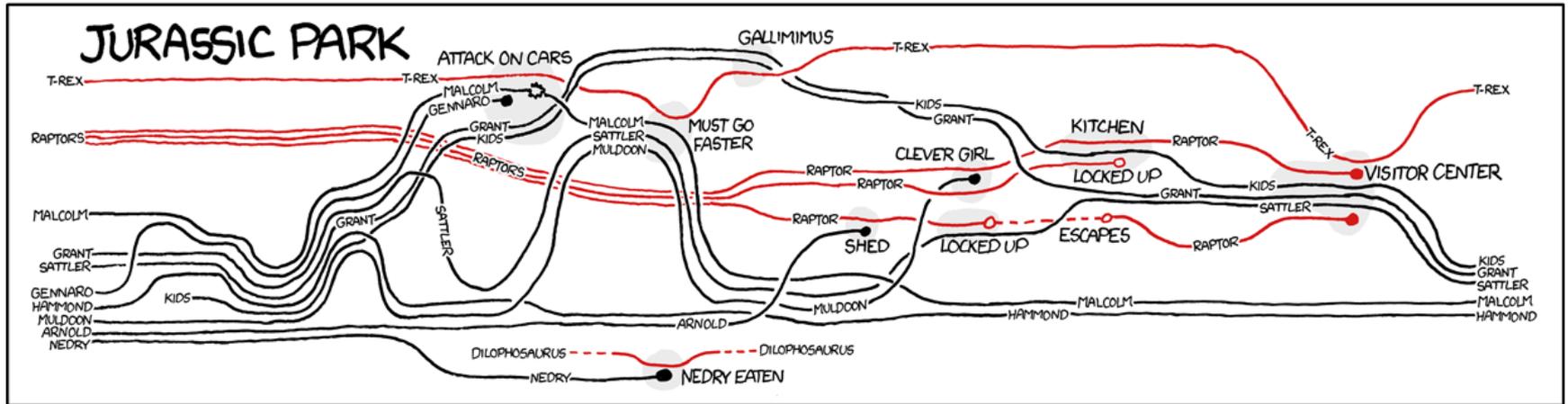
pregnancy and childbirth



diseases of the respiratory system



Storylines



Shape of Song

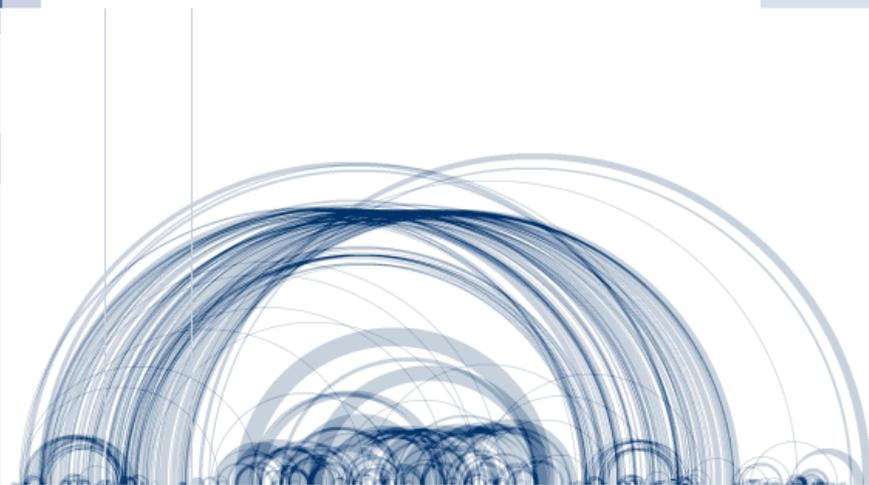
THE SHAPE OF SONG

[What the diagrams mean](#) | [Image gallery](#) | [Home](#) | [Contact](#)

Repertoire (648 pieces)

Composer	Title
Chopin	√ Fantasie Impr...
	Mazurka in C#m
	Mazurka in F#m
	Nocturne No....
	Prelude 1
	Revolutionary...
Choral arr. Ha...	My Funny Val...
Chris DeBurgh	Lady In Red
	Lady In Red (2)
chris isaak	somebody's c...
Christmas Ca...	Frosty the Sn...
	Greensleeves
	Joy To The W...
	Silent Night
	Silver Bells
	The 12 Days ...
	Winter Wonde...
Classic	You Are My S...
Clay Dale	Clay35b
Clinger	German Bovinia
	Nostalgia
	Twinkle Variat...
	Violin Plano ...
Clutch	Juggernaut
	Juggernaut
Copland	Appalachian ...
Couperin/Mo...	Frankfurter
Cradle of Filth	The Forest W...
cranberries	zombie
Cream	White room

Fantasie Impromptu. Chopin. Viewing track 1 of 1. Play (all tracks)



You can add any MIDI file on the web to the repertoire.

URL

Title Composer Add to repertoire

Enter the URL of a MIDI file and the title of the piece. Composer is optional but nice.

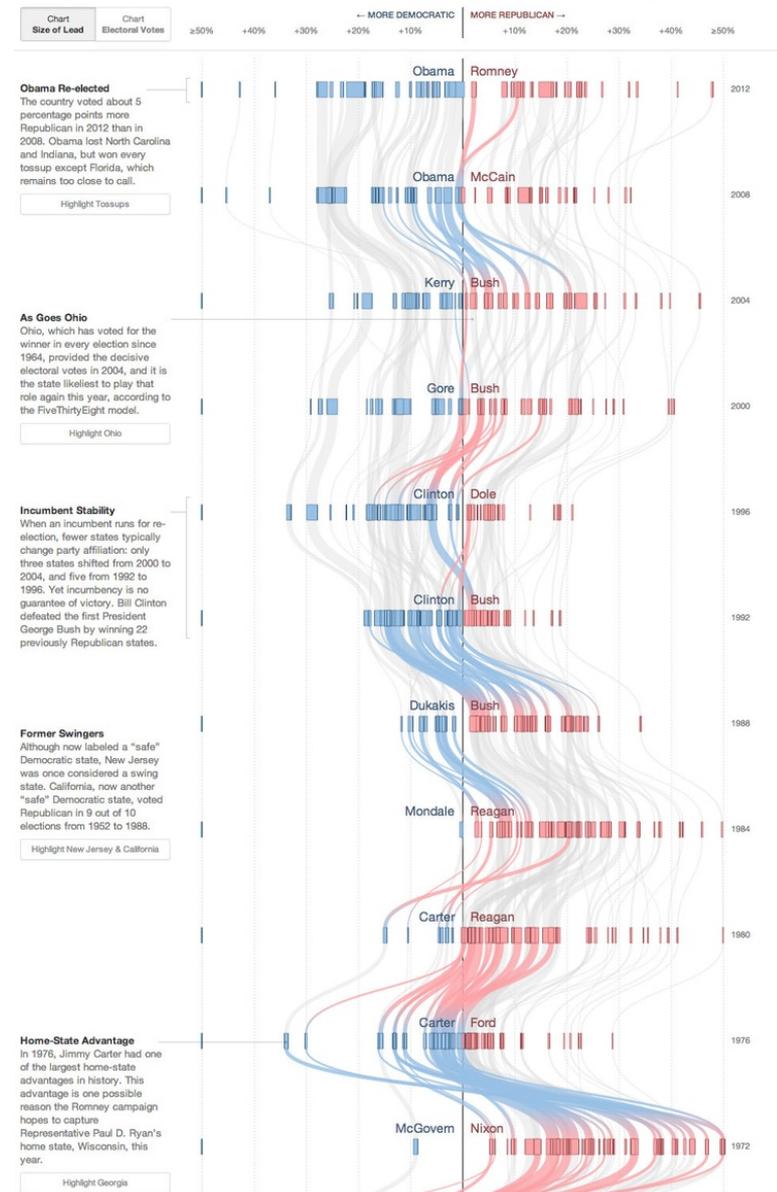
<http://www.turbulence.org/Works/song/mono.html>

Over the Decades, How States Have Shifted

Over the Decades, How States Have Shifted

Recent elections have placed a heavy emphasis on "swing states" – Ohio, Florida and the other competitive states. Yet in the past, many more states shifted between the Democratic and Republican parties. A look at how the states stacked up in the 2012 election and how they have shifted over past elections.

Each box represents a state sized by number of electoral votes.
Each curve shows how much it shifted left or right between elections.



<http://nyti.ms/Wr1dhZ>

Possible tools for temporal vis.

- Basic charting tools

- Raw

<http://raw.densitydesign.org/>

- TimelineJS

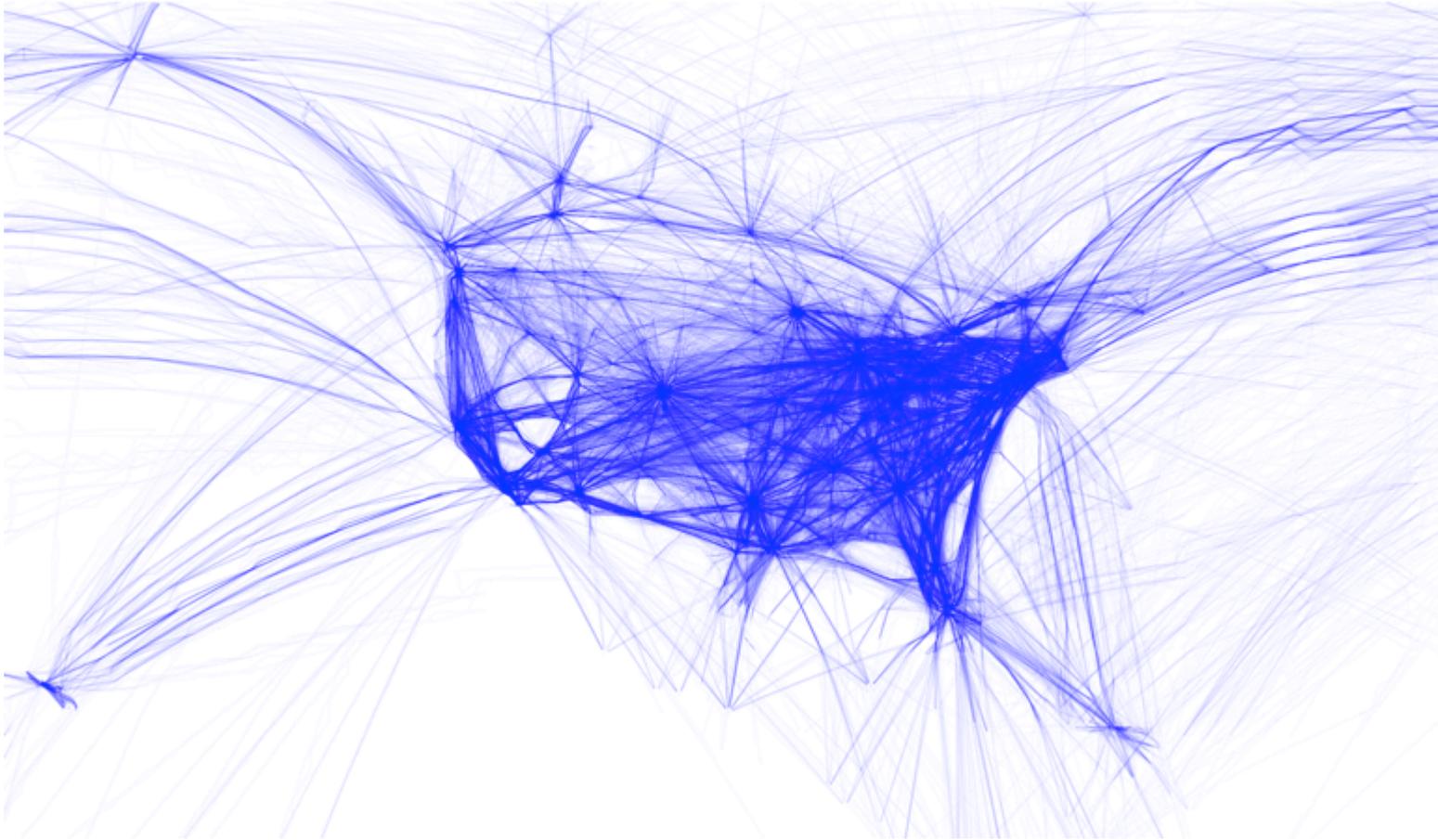
<http://timeline.knightlab.com/>

- Simile Timeline

<http://simile.mit.edu/>

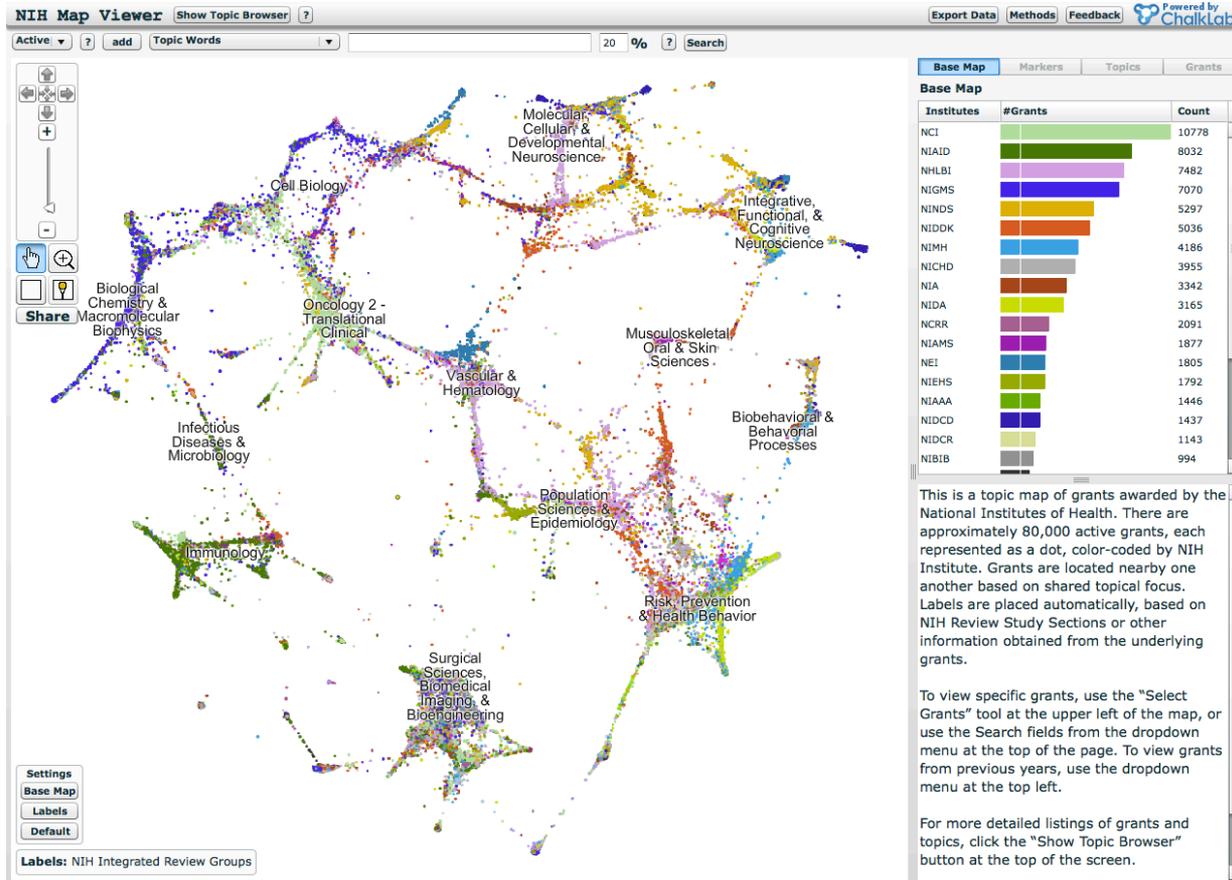
- D3

Edges



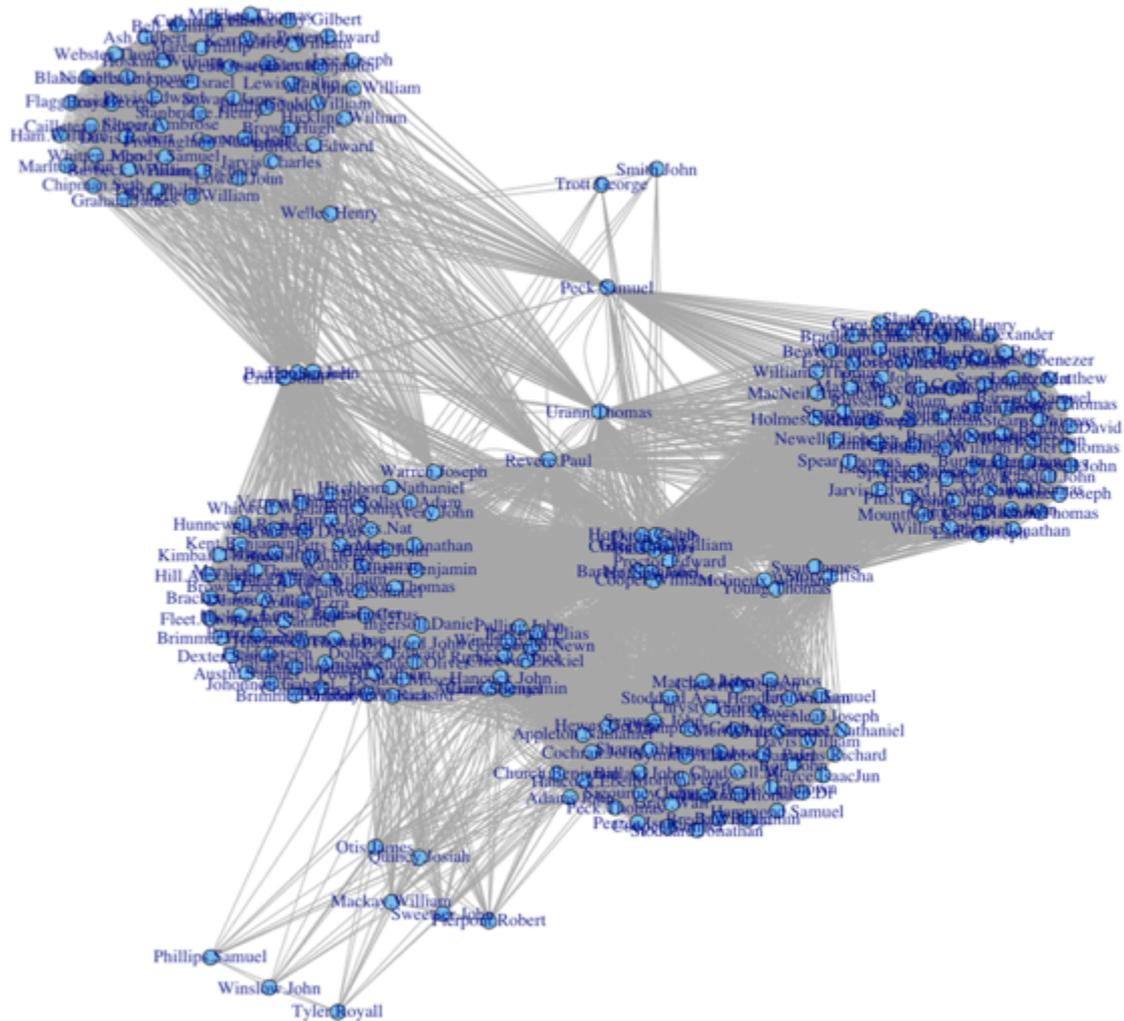
<http://www.aaronkoblin.com/work/flightpatterns/>

Nodes



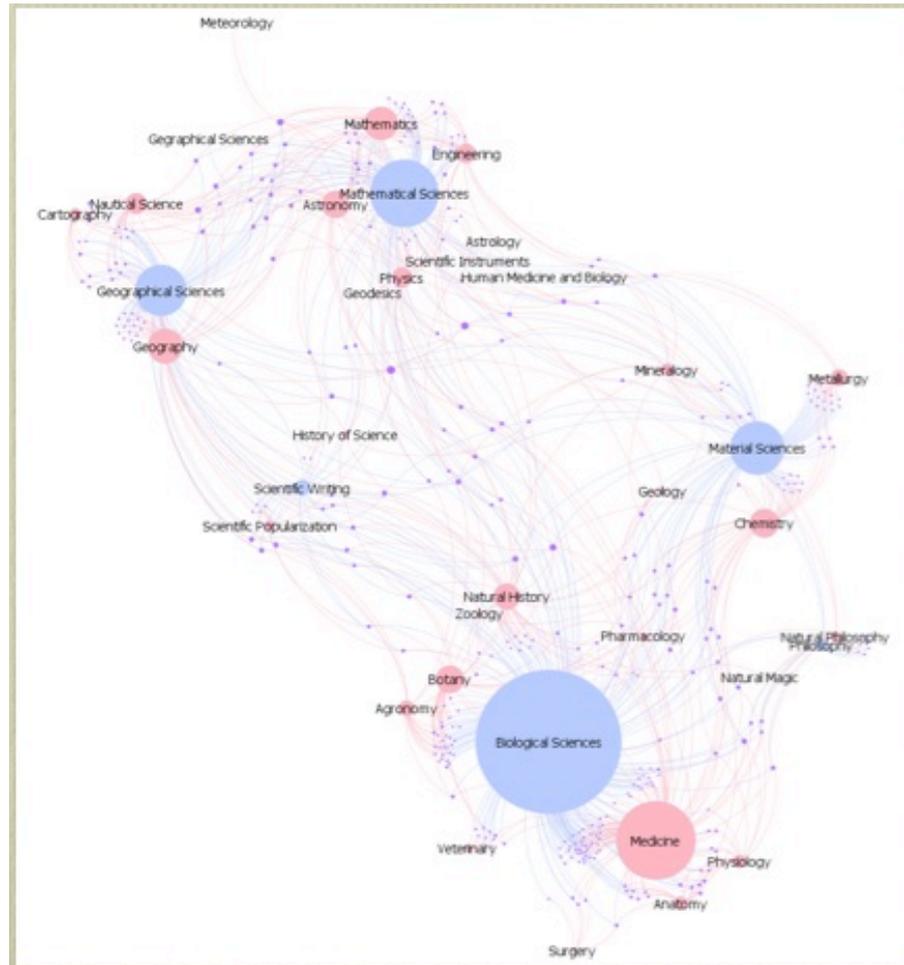
<http://nihmaps.org/vids.php>

Both



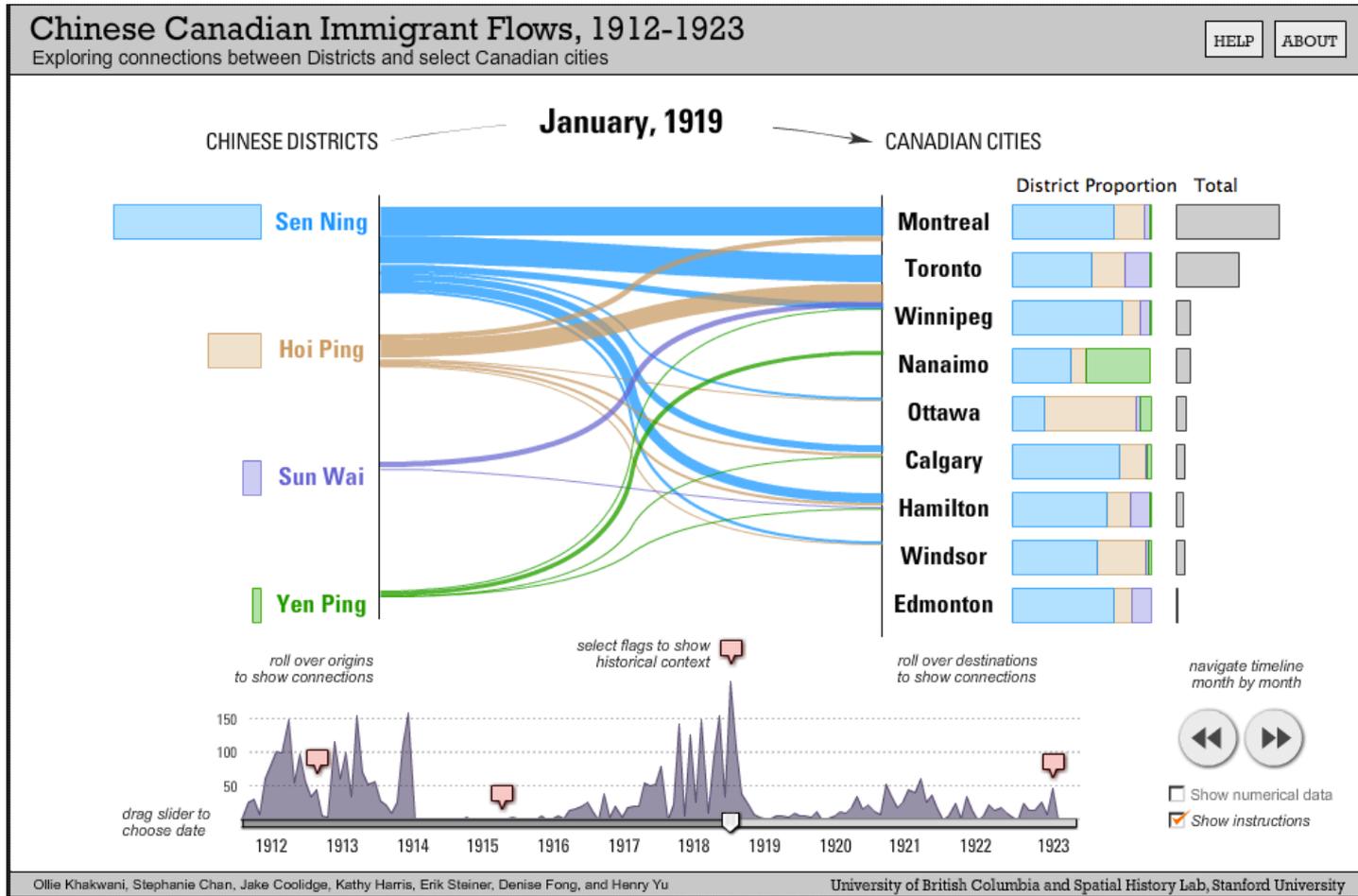
<http://kieranhealy.org/blog/archives/2013/06/09/using-metadata-to-find-paul-revere/>

With color and size coding



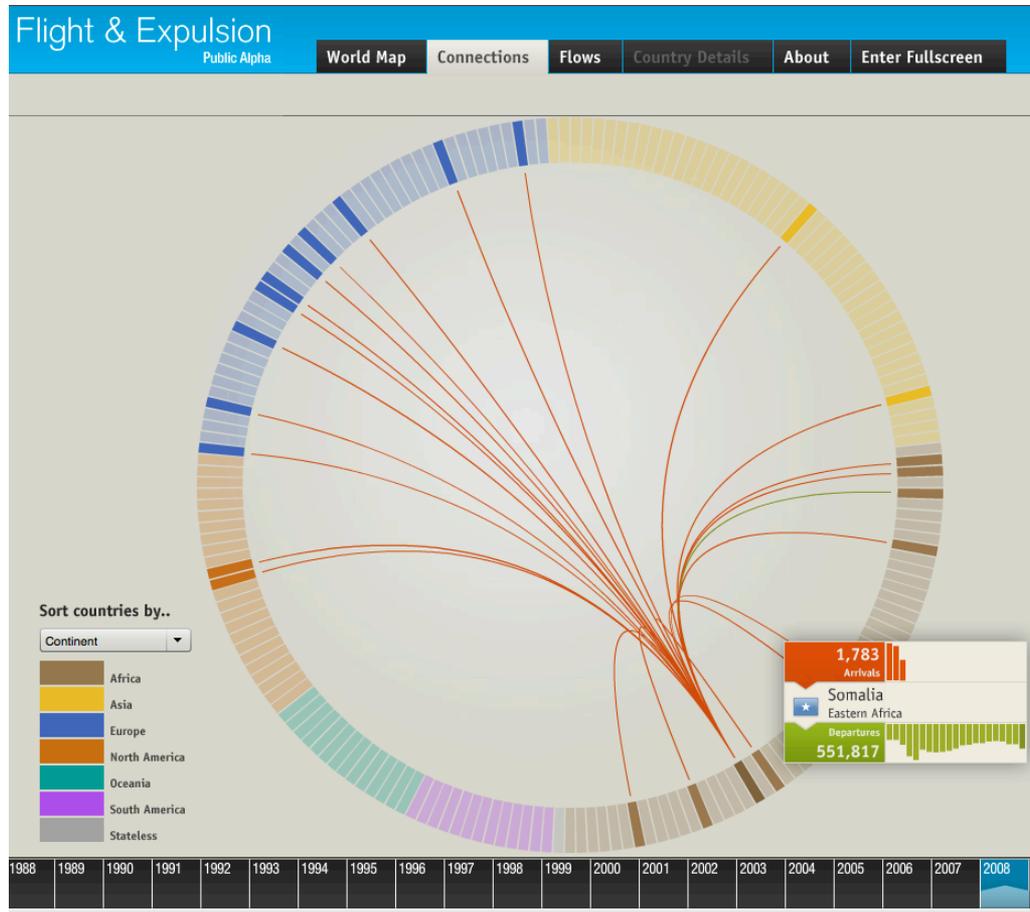
<http://republicofletters.stanford.edu/casestudies/spanishempire.html>

Bipartite graph, alluvial diagram



<http://stanford.io/1hCYwkd>

Circular layout/chord diagram



<http://www.niceone.org/lab/refugees/>

Tube Map

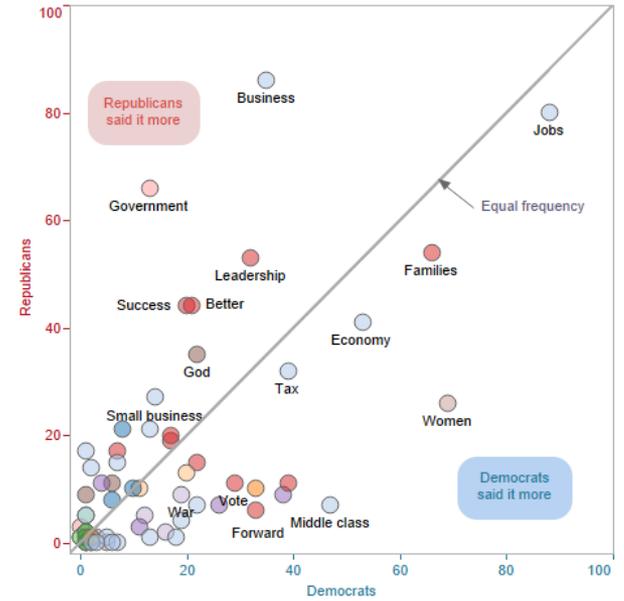
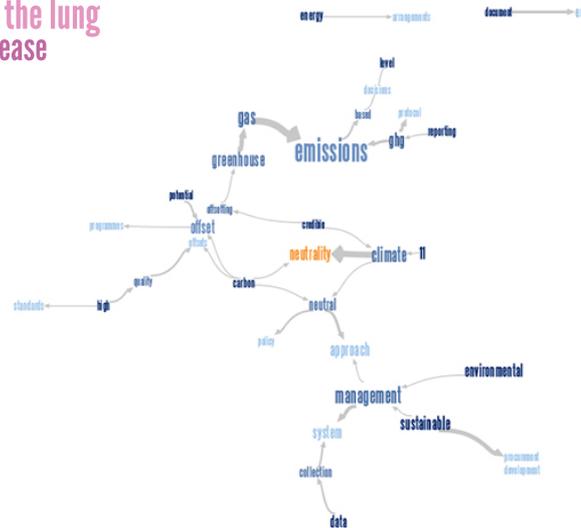
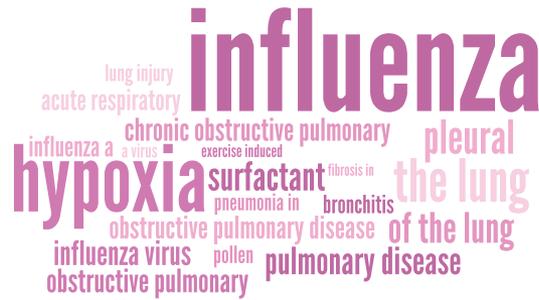


<http://diagrams.org/images/png/large/f00022.html>

Possible tools for network vis.

- D3
- Gephi
<http://gephi.org/>
- NodeXL
<http://nodexl.codeplex.com/>
- Pajek
<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
- Cytoscape
- Network Workbench/Sci²
<http://nwb.cns.iu.edu/>,
<https://sci2.cns.iu.edu/>
- VOSviewer
<http://www.vosviewer.com/>
- UCINET
<https://sites.google.com/site/ucinetsoftware/home>
- GUESS
<http://graphexploration.cond.org/>
- R
- SigmaJS
<http://sigmajs.org/>
- Circos
<http://circos.ca/>

Showing Text



http://guides.library.duke.edu/text_vis

Bubble Plot

The Words Speakers Use

Looking at the number of times words have been used by speakers at each party's convention suggests the different themes the parties have highlighted.

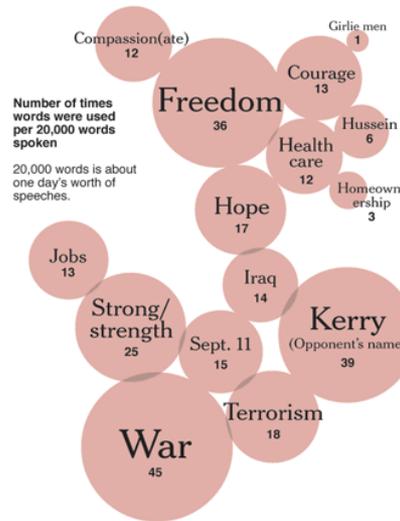
The Republican speakers have used more terms related to terrorism and the war in Iraq, while the Democratic speakers were more likely to mention health care or jobs.

In addition, Republicans were more likely to mention the opposing candidate by name, something the Democrats rarely did.

MATTHEW ERICSON

Republican Convention

Speakers on Monday, Tuesday and Wednesday



SOME OF THE REPUBLICAN SPEAKERS



Democratic Convention

Speakers on all days



SOME OF THE DEMOCRATIC SPEAKERS



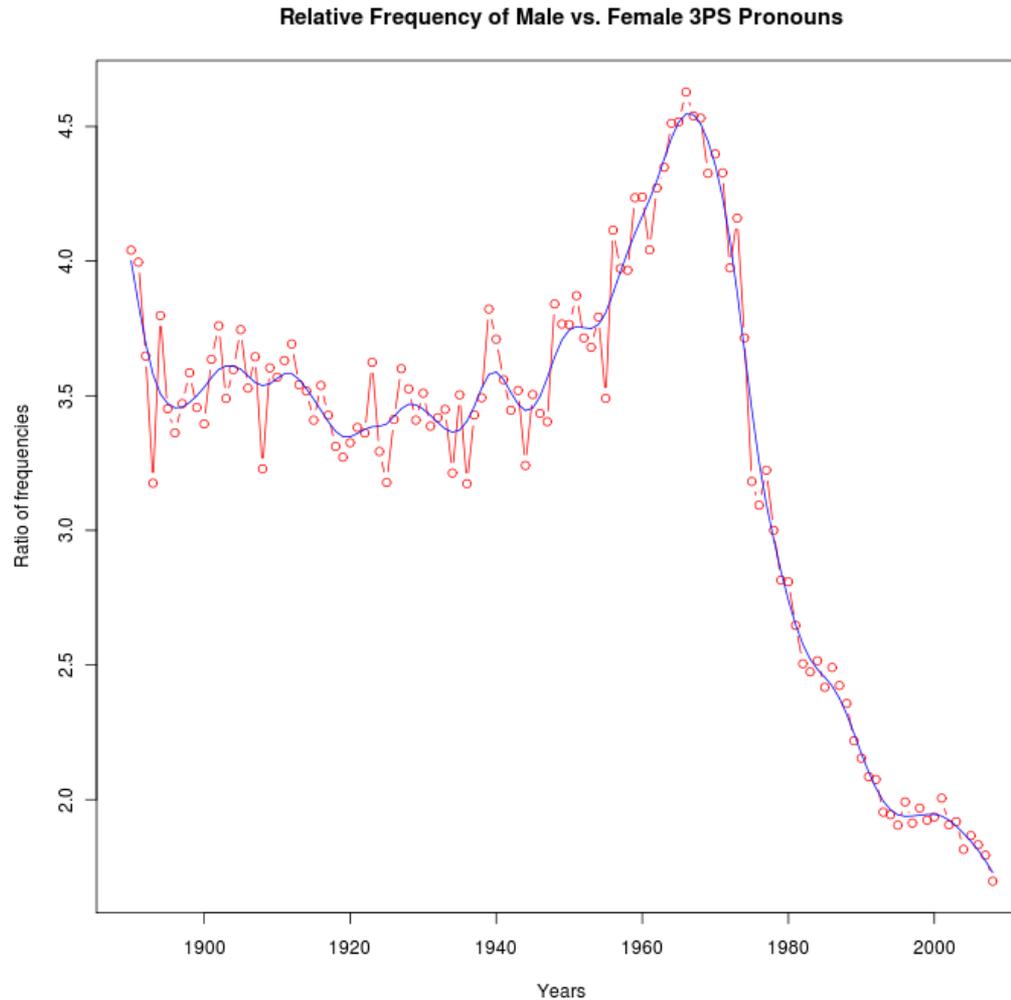
NUMBER OF TIMES SPEAKERS USED EACH WORD OR PHRASE

	John McCain	Rudolph W. Giuliani	Arnold Schwarzenegger	Laura Bush	Zell Miller	Dick Cheney	Bill Clinton	Barack Obama	John Edwards	John Kerry
Opponent's name		12			15	13				2
Hussein	1	6		2	2	3				
Terrorism	2	21	2						2	
Religion	1	1	1							1
Afghanistan	1	2		3	2	2	1		1	
Courage	5	1			1	1	1		1	
Homeownership				3		1	1			
Freedom	7	9	4	6	12	7	2	2	2	5
Compassion(ate)			2			1	1			
Sept. 11	4	11		2	1	3	2		3	4
War	18	13		7	8	10	6	6	10	14
Iraq	2	4	3	4	2	3	2	3	9	4
Education		1		4		1			2	3
Strong/strength	2	3	1	4		2	5		5	8
Troops		1	2		2	2		1	1	4
Economy			4			2	3	1	1	3
Hope	1	4	1	4		1		13	8	4
Health care		2		1		1	5	1	9	12
Jobs				1		1	6	5	6	12

Sources: Federal News Service transcripts of speeches; Republican National Convention

The New York Times

Frequencies over time



<http://languagelog.ldc.upenn.edu/nll/?p=4126>

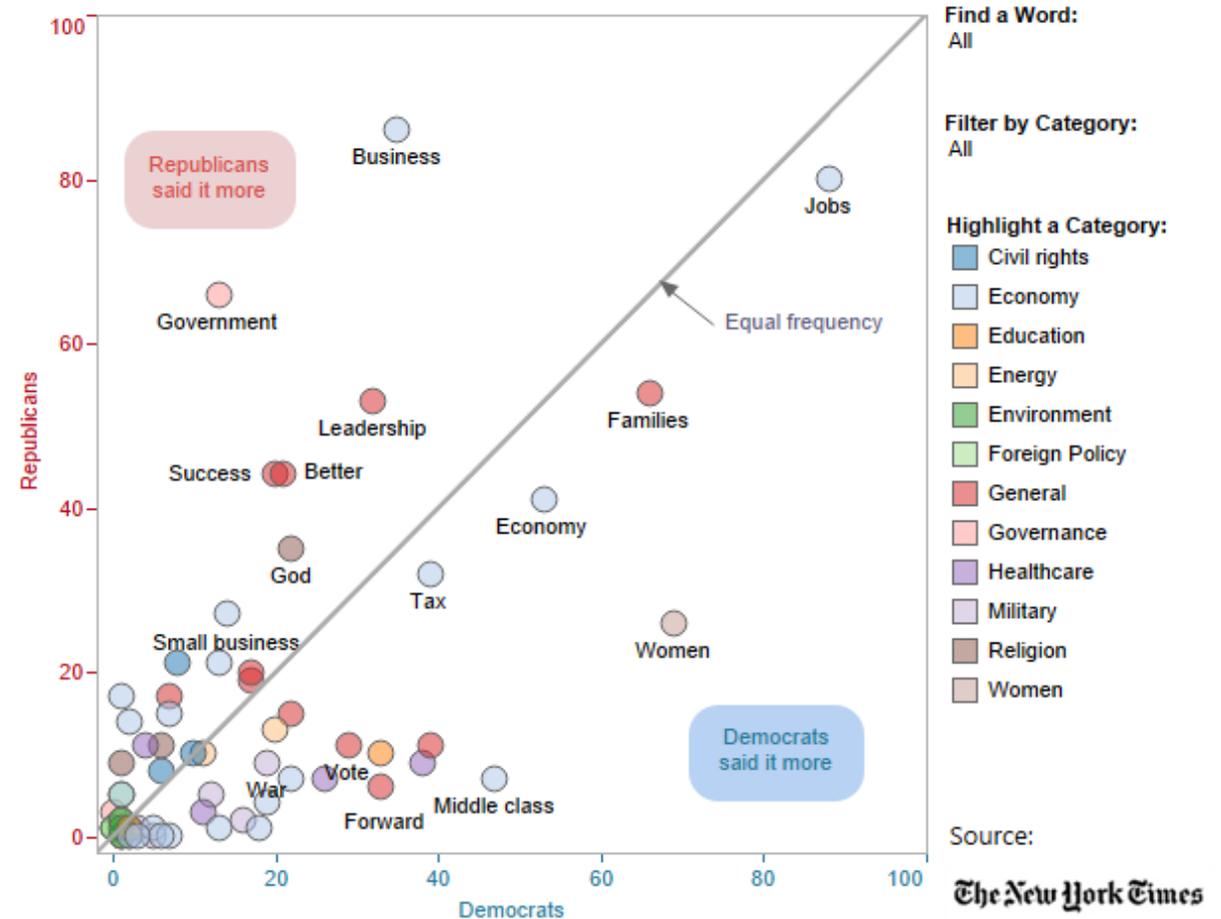
Scatter Plot

Ye Shall Know Them By Their Words...



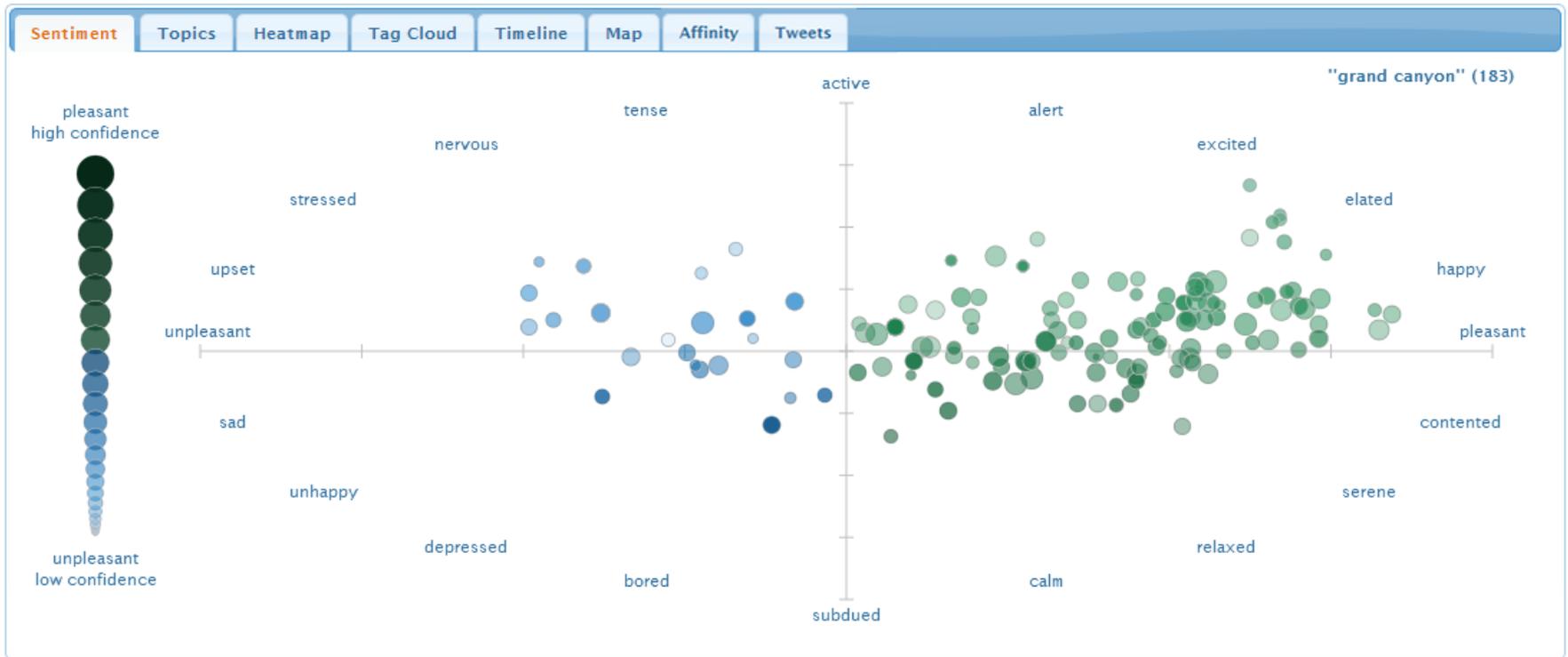
Compare how often speakers from the different parties used select words and phrases at their respective 2012 presidential nominating conventions. On each axis is plotted the count per 25,000 words.

Democrats had more to say about *healthcare, military and women*, while Republicans mentioned *religion and governance* more frequently:



<http://tabsoft.co/1ByN2br>

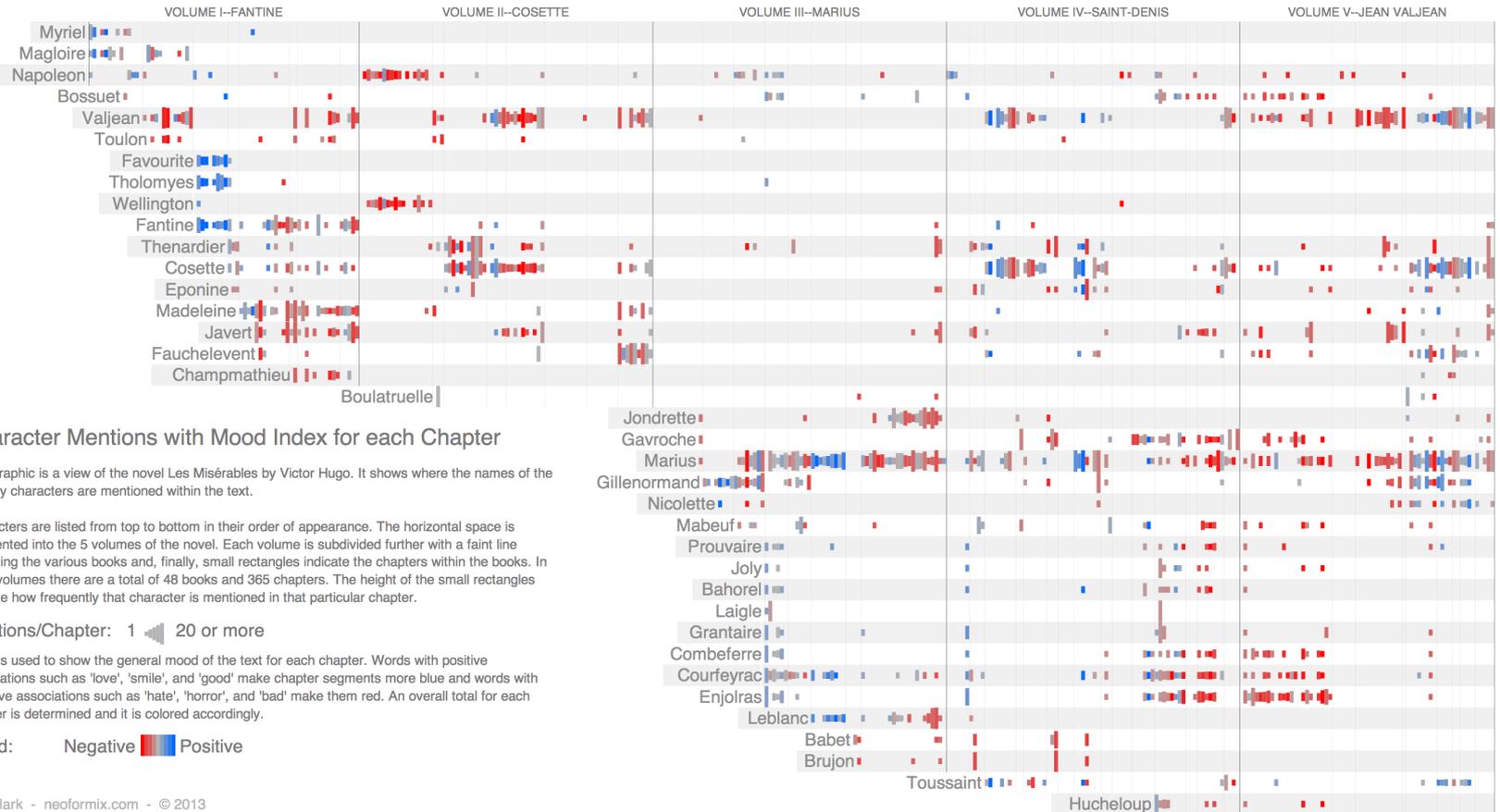
Sentiment analysis



http://www.csc.ncsu.edu/faculty/healey/tweet_viz/

Sentiment analysis

NOVEL VIEWS - Les Misérables - Character Mentions



Character Mentions with Mood Index for each Chapter

This graphic is a view of the novel *Les Misérables* by Victor Hugo. It shows where the names of the primary characters are mentioned within the text.

Characters are listed from top to bottom in their order of appearance. The horizontal space is segmented into the 5 volumes of the novel. Each volume is subdivided further with a faint line indicating the various books and, finally, small rectangles indicate the chapters within the books. In the 5 volumes there are a total of 48 books and 365 chapters. The height of the small rectangles indicate how frequently that character is mentioned in that particular chapter.

Mentions/Chapter: 1 ◀ 20 or more

Color is used to show the general mood of the text for each chapter. Words with positive associations such as 'love', 'smile', and 'good' make chapter segments more blue and words with negative associations such as 'hate', 'horror', and 'bad' make them red. An overall total for each chapter is determined and it is colored accordingly.

Mood: Negative ■ ■ Positive

Word Tree

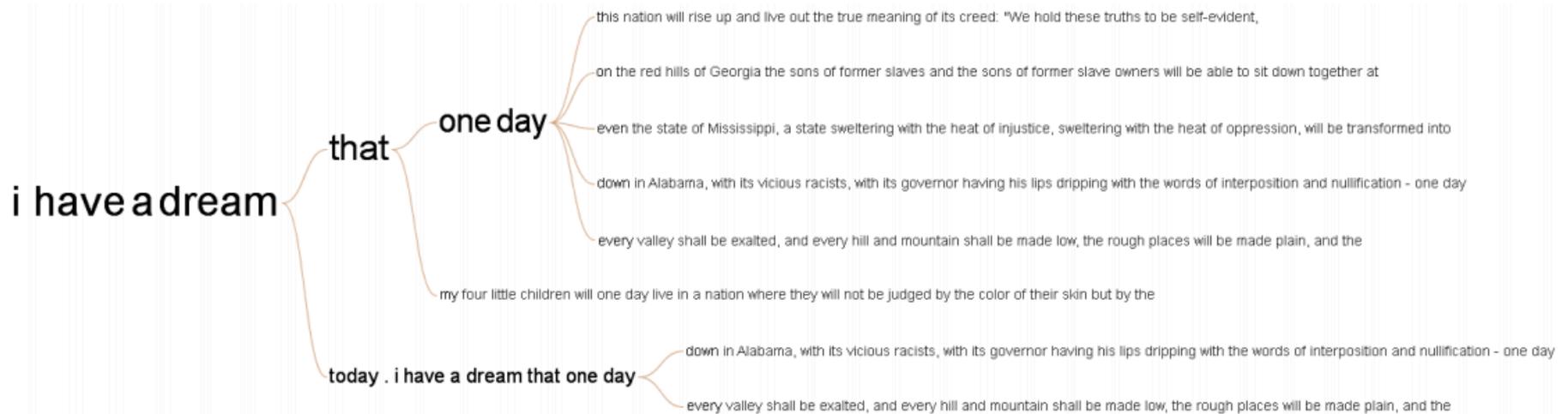
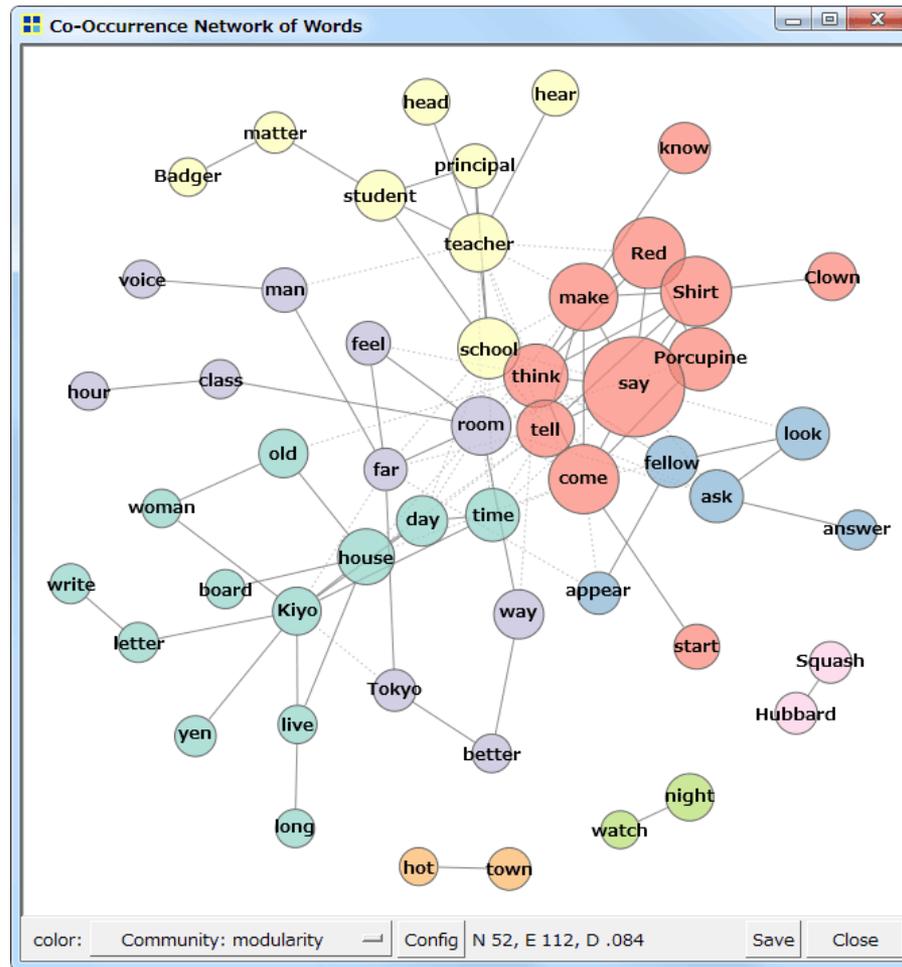


Fig 10: Word Tree showing all occurrences of "I have a dream" in Martin Luther King's historical speech.

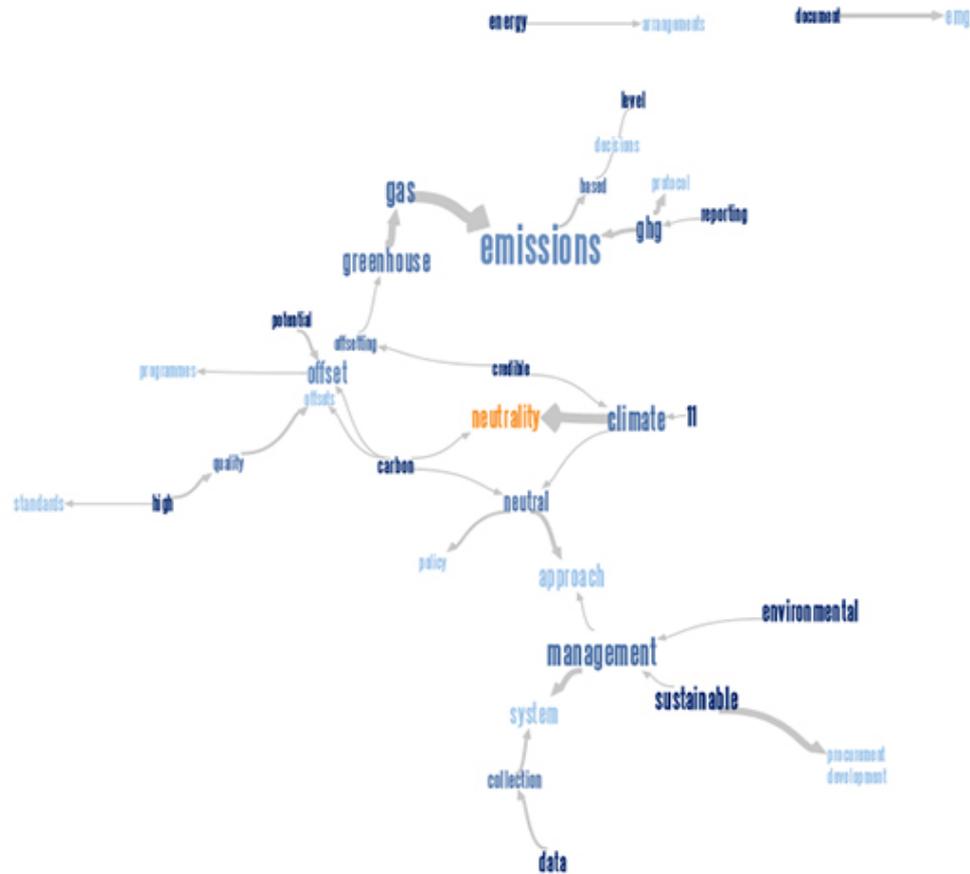
<http://hint.fm/projects/wordtree/>

Word co-occurrence network



http://en.wikipedia.org/wiki/Co-occurrence_networks

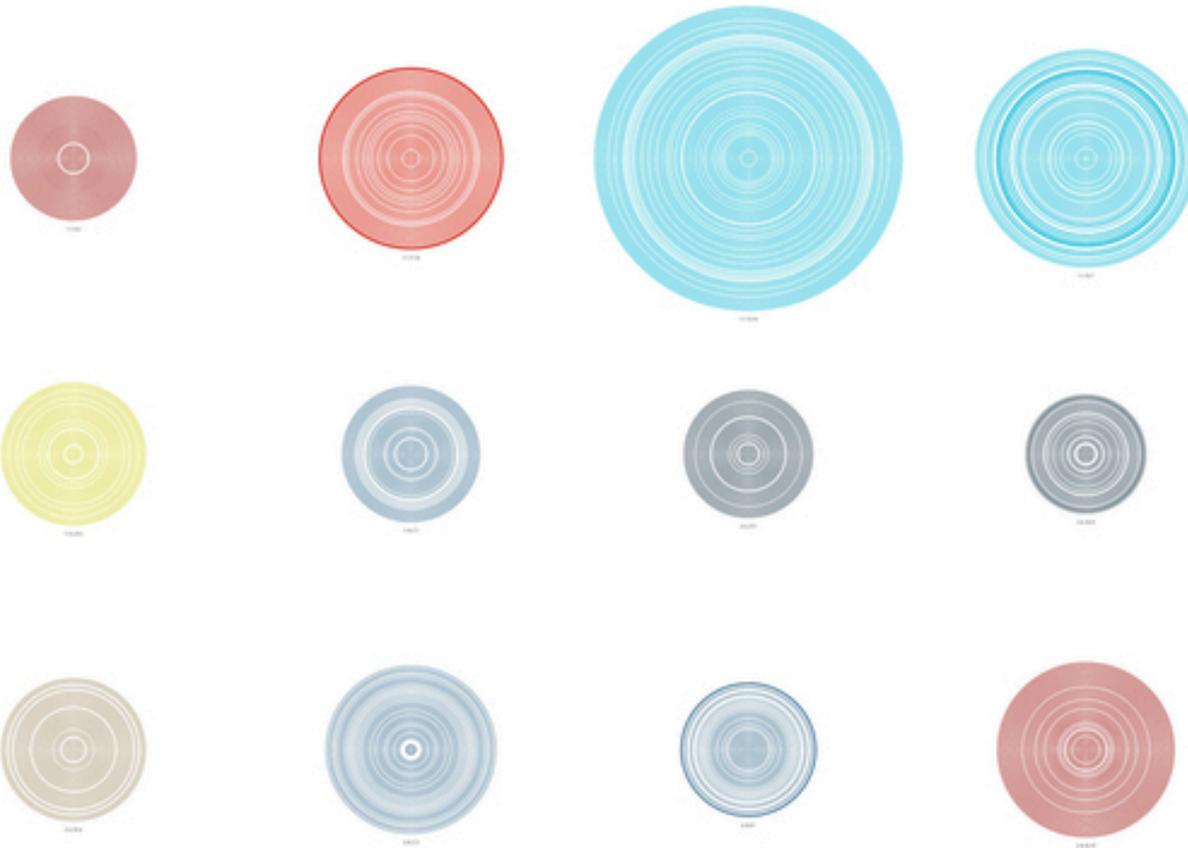
Phrasenet



<http://www-01.ibm.com/software/analytics/many-eyes/>

Rhythm Textures

Exploring ways of visually representing sentences by using their punctuation to create circular diagrams. Each word is represented by a line, and the thickness of the lines (and the space between the lines) radiating outwards from the center point provides a record of the pauses and emphasis created by the punctuation.



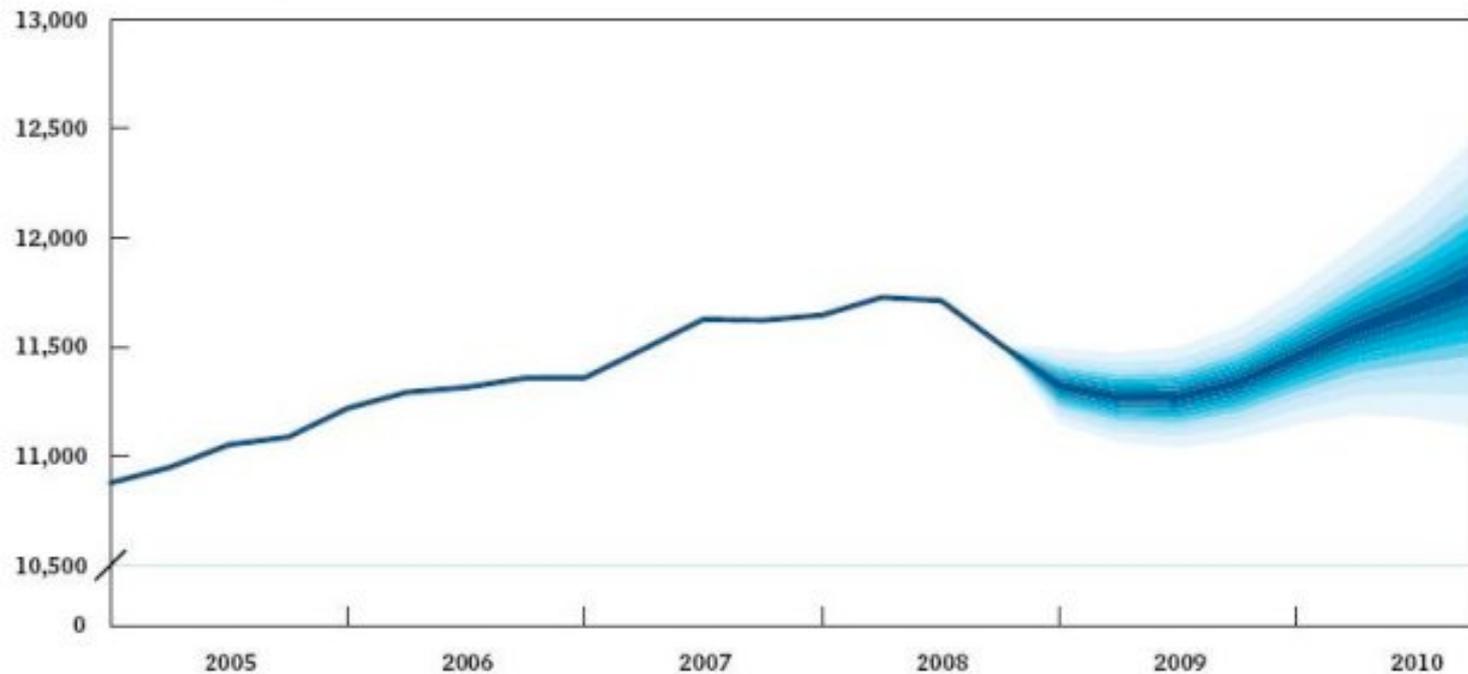
<http://itsbeenreal.co.uk/index.php?/wwwwords/rhythm-textures/>

VISUALIZING UNCERTAINTY

Projections

Uncertainty in Projections of Real GDP

(Billions of 2000 dollars)



<http://peltiertech.com/WordPress/excel-fan-chart-showing-uncertainty-in-projections/>

Missing data

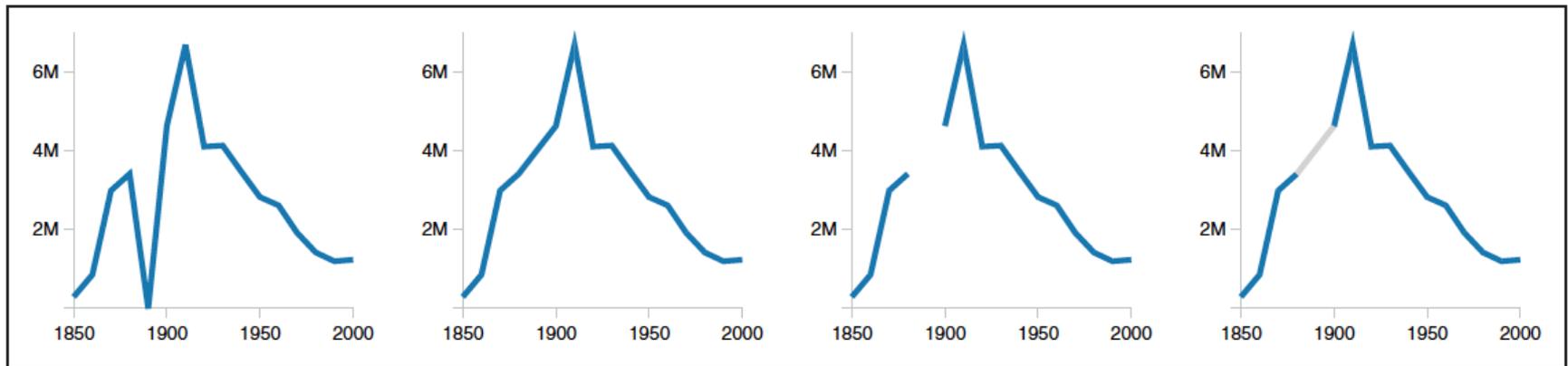
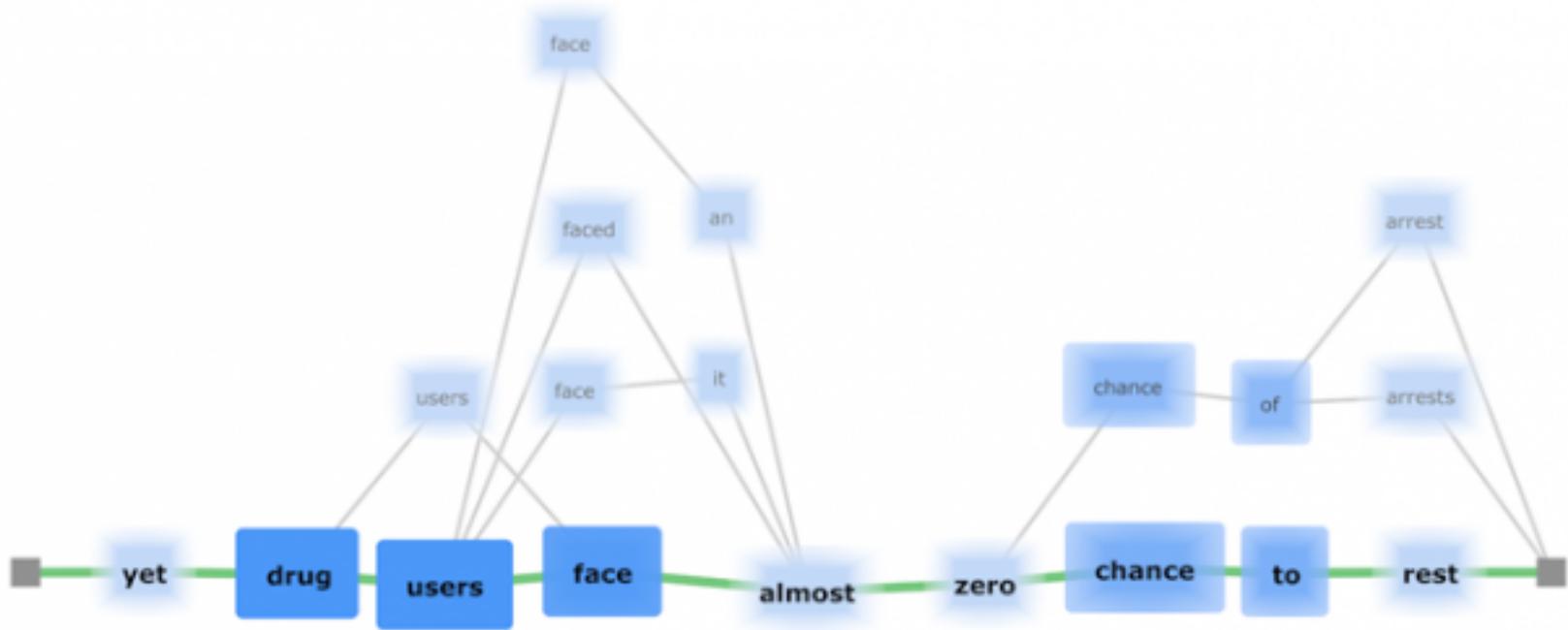


Figure 4. Alternative representations of missing data in a line chart. The data are U.S. census counts of people working as 'Farm Laborers'; values from 1890 are missing due to records being burned in a fire. (a) Missing data is treated as a zero value. (b) Missing data is ignored, resulting in a line segment that interpolates the missing value. (c) Missing data is omitted from the chart. (d) Missing data is explicitly interpolated and rendered in gray.

Alternative solutions



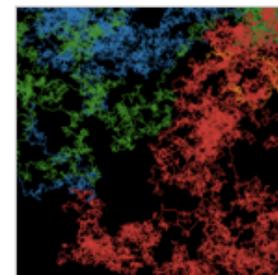
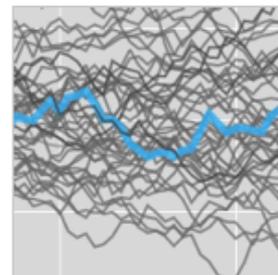
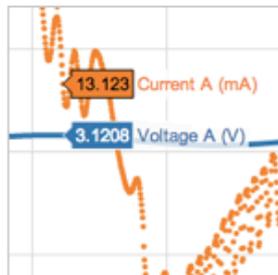
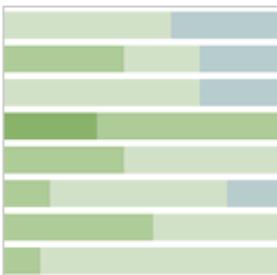
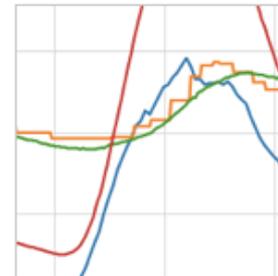
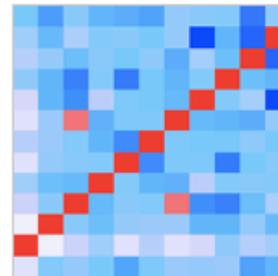
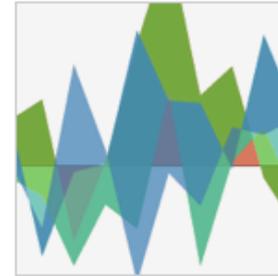
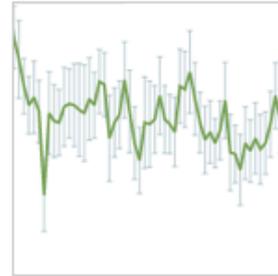
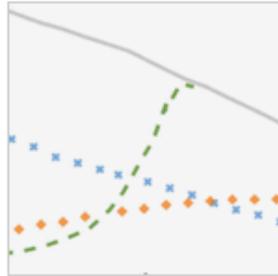
<http://vialab.science.uoit.ca/portfolio/lattice-uncertainty-visualization-understanding-machine-translation-and-speech-recognition>

Take-away

Uncertainty is blue.

TOOLS THAT DON'T NEED INSTALLATION

Plot.ly



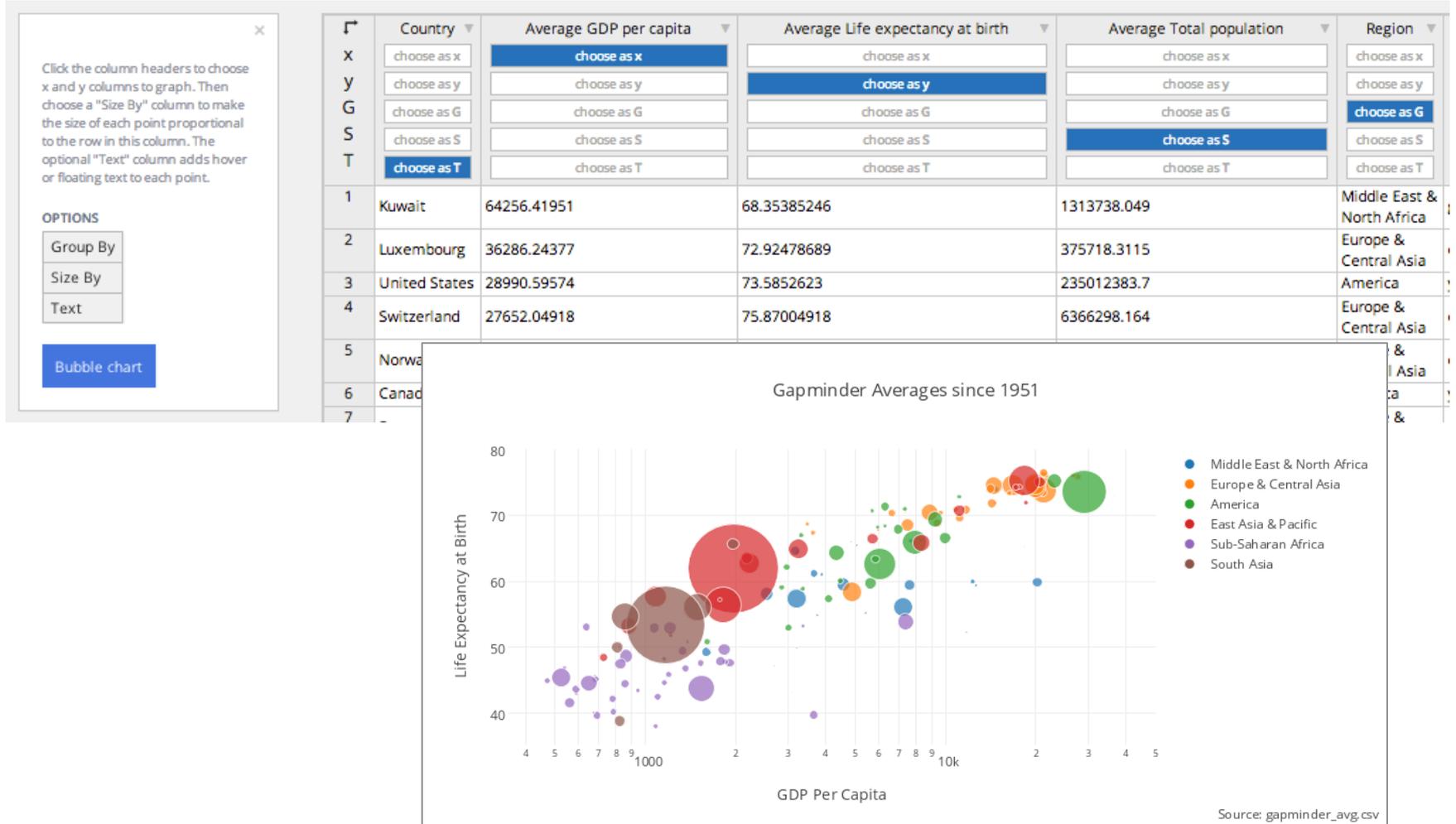
<https://plot.ly/>

Plot.ly

- Browser based (or Excel add-in)
- Makes wide variety of chart types
- Allows for python, MATLAB, R, etc. syntax
- Makes charts that are hosted/shareable

<https://plot.ly/>

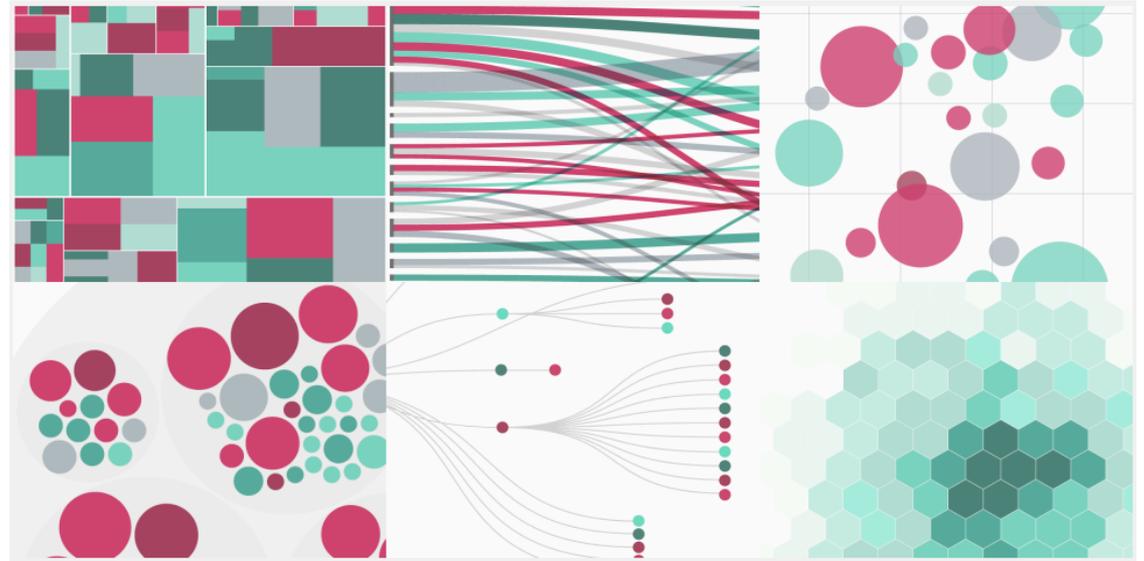
Example: Bubble chart



Raw

Has visualizations to show:

- Numbers
- Relationships
- Hierarchies



<http://raw.densitydesign.org/>

Raw

- Paste in a data table (.csv, .tsv, copied from Excel)
- Select chart type
- Drag column headers to different chart attributes
- Save out image or SVG code

<http://raw.densitydesign.org/>

Example: Alluvial Diagram

STEPS

Drag numbers, strings, dates here

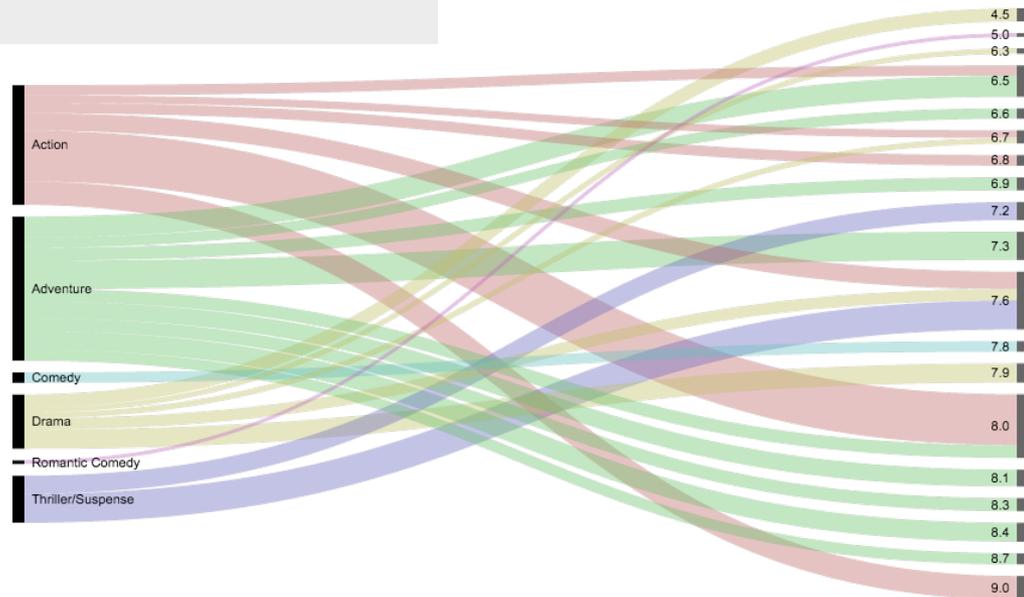
Genre string dimension.title()

Rating IMDB number x

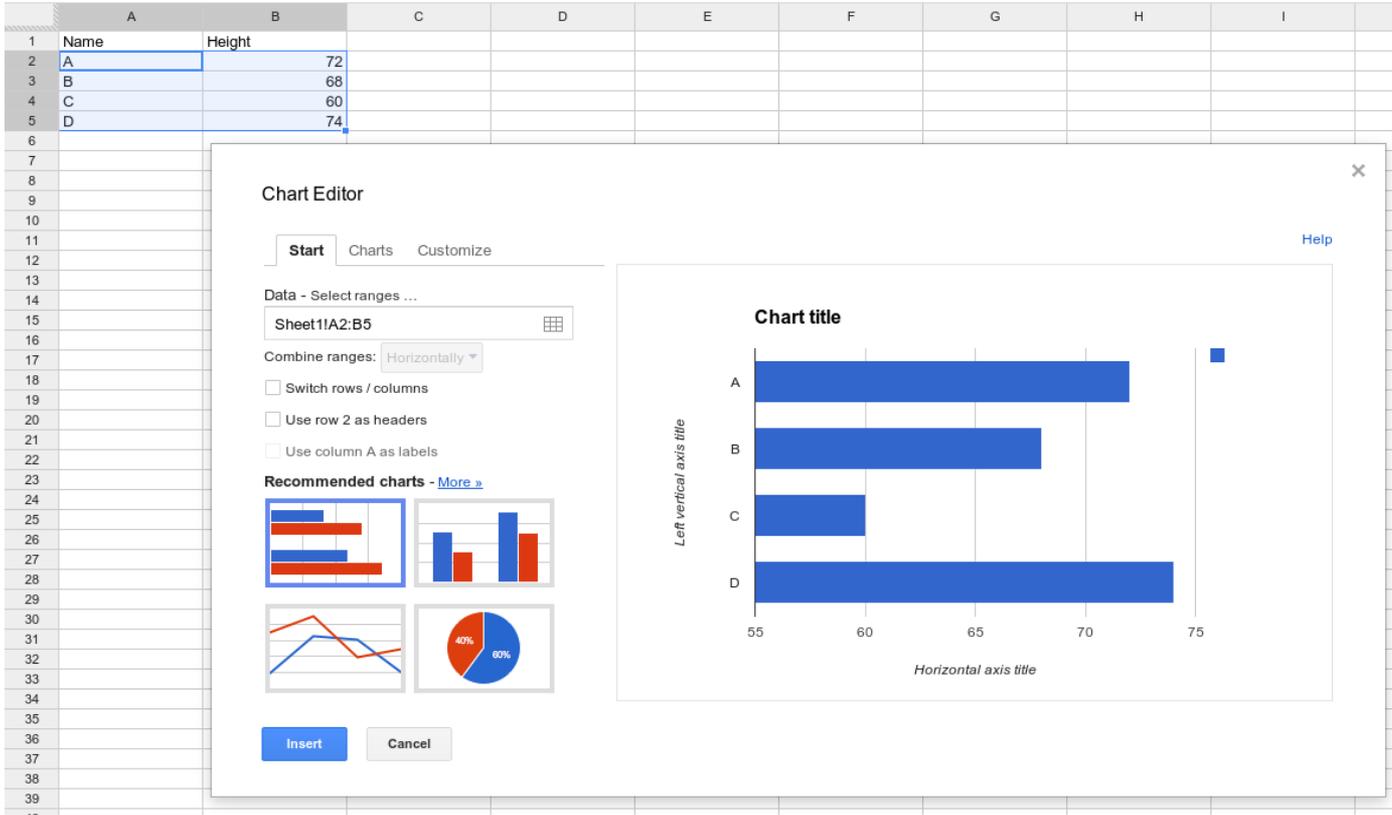
SIZE

Drag numbers here

Total Domestic Box Office number x



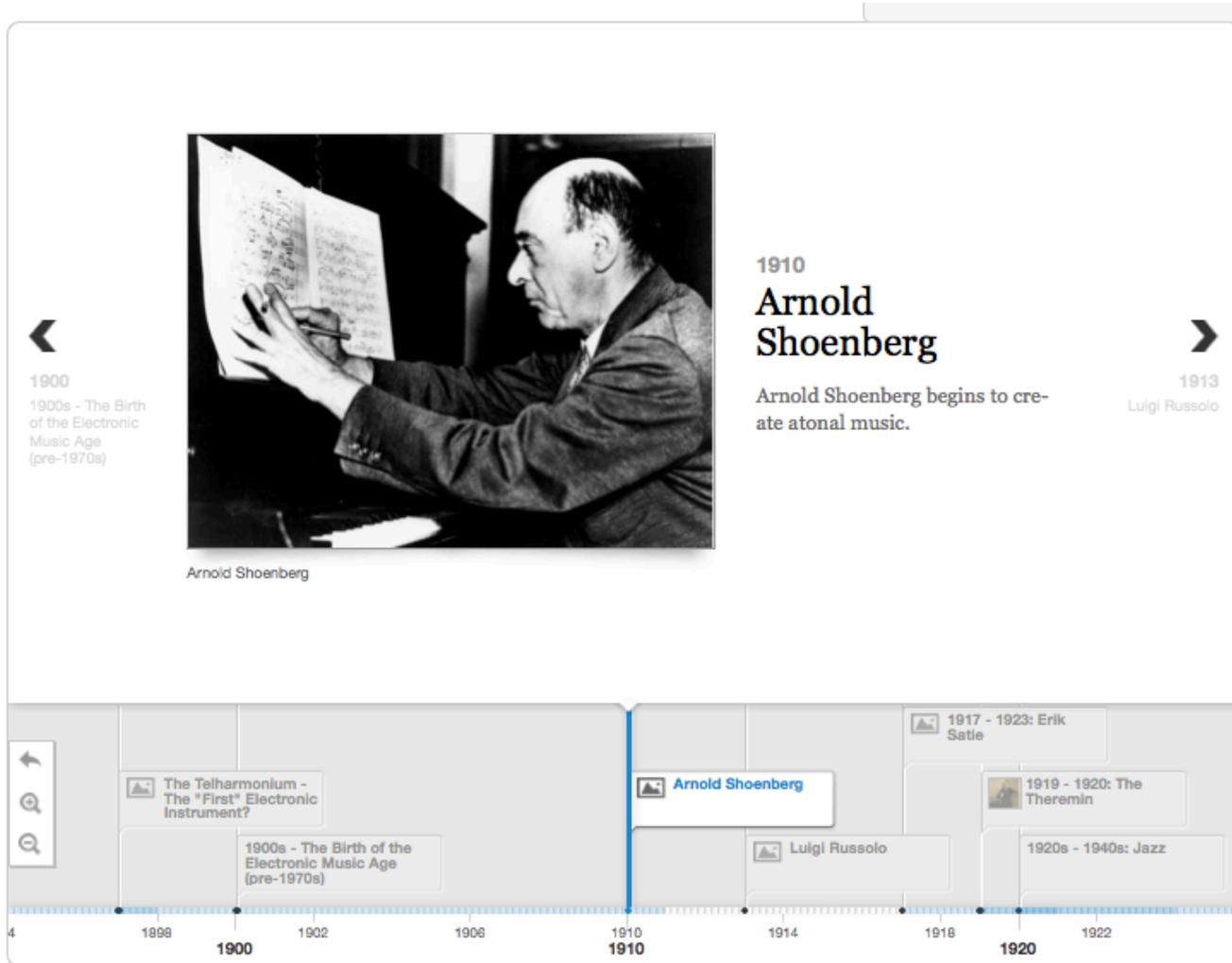
Google Spreadsheets



-  Line
-  Area
-  Column
-  Bar
-  Scatter
-  Pie
-  Map
-  Trend
-  More

<https://drive.google.com/>

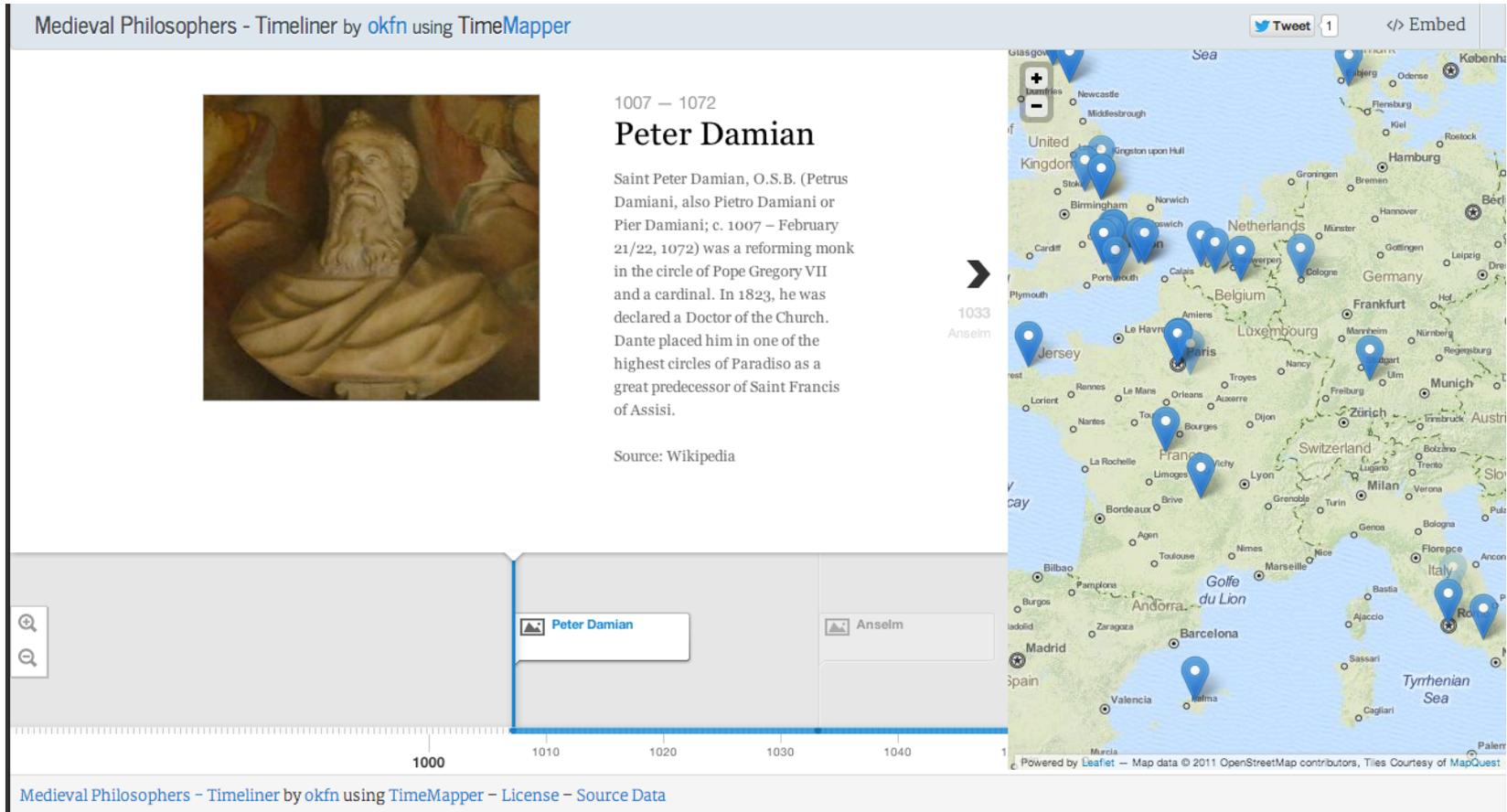
TimelineJS



<http://timeline.knightlab.com/>

Timeliner

Medieval Philosophers - Timeliner by okfn using TimeMapper



1007 — 1072

Peter Damian

Saint Peter Damian, O.S.B. (Petrus Damiani, also Pietro Damiani or Pier Damiani; c. 1007 – February 21/22, 1072) was a reforming monk in the circle of Pope Gregory VII and a cardinal. In 1823, he was declared a Doctor of the Church. Dante placed him in one of the highest circles of Paradiso as a great predecessor of Saint Francis of Assisi.

Source: Wikipedia

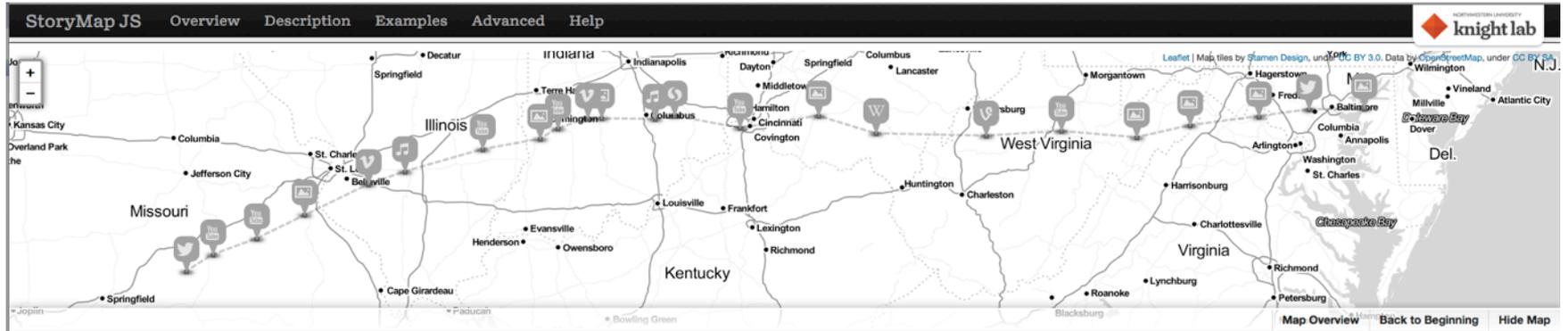
1033 Anselm

1000 1010 1020 1030 1040

Medieval Philosophers - Timeliner by okfn using TimeMapper - License - Source Data

<http://timemapper.okfnlabs.org/>

StoryMapJS



Mean US_Mean_Center_of_Population_1790-2010

US Manifest Destiny - Mean center of United States population

The mean center of U.S. population is determined by the United States Census Bureau from the results of each census. Defined as the point at which an imaginary, flat, weightless, and rigid map of the United States would balance perfectly if weights of identical value were placed on it so that each weight represented the location of one person on the date of the census.

This is an overview or title slide to show all the points in your story routed on your map.



Also, GitHub auto-rendering

- 3D Files

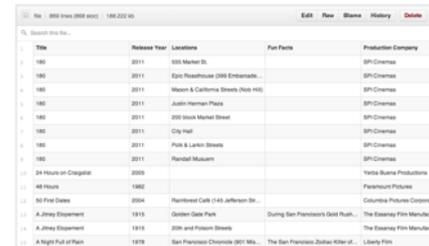
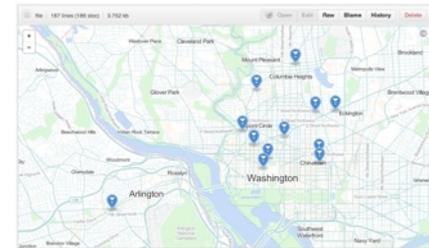
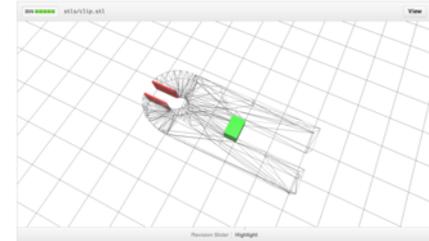
<https://help.github.com/articles/3d-file-viewer>

- GeoJSON/TopoJSON

<https://help.github.com/articles/mapping-geojson-files-on-github>

- CSV/TSV

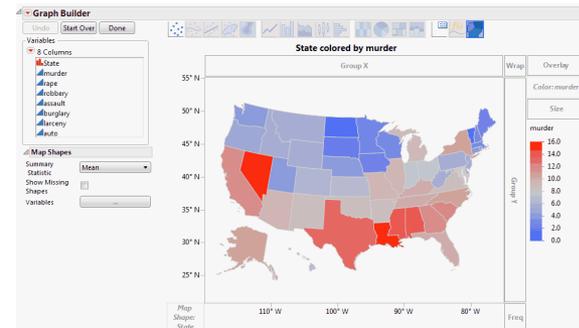
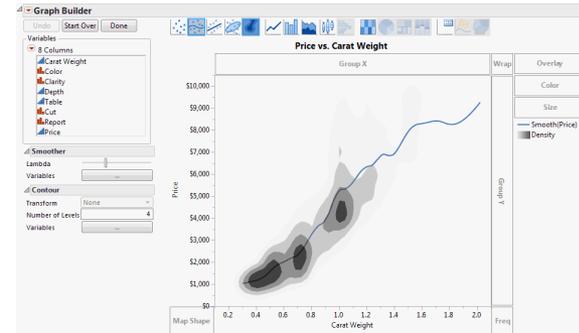
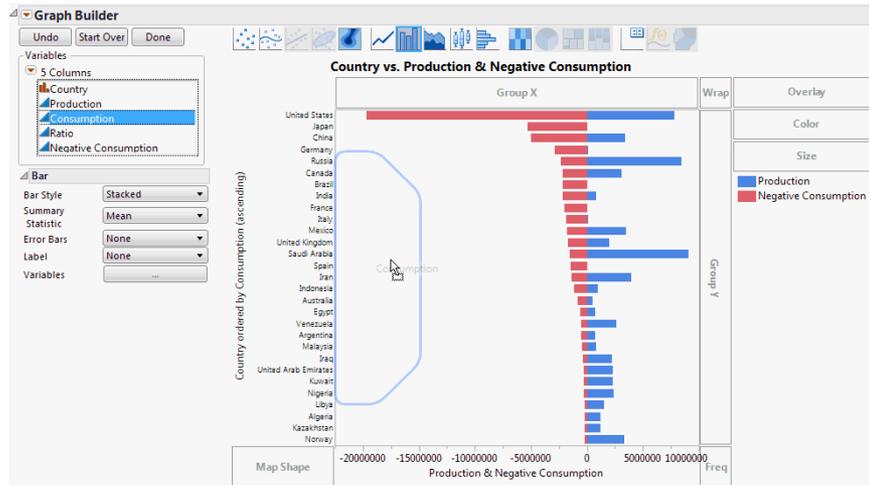
<https://help.github.com/articles/rendering-csv-and-tsv-data>



Title	Release Year	Location	Fun Facts	Production Company
1	190	2011	555 Market St	SPI-Cinema
2	190	2011	Epic Resonance (300 Embarcadero)	SPI-Cinema
3	190	2011	Maen & California Streets (Jazz Hill)	SPI-Cinema
4	190	2011	Golden Gate Park	SPI-Cinema
5	190	2011	200 Stock Market Street	SPI-Cinema
6	190	2011	City Hall	SPI-Cinema
7	190	2011	Pink & Larkin Streets	SPI-Cinema
8	190	2011	Randall Museum	SPI-Cinema
9	24 Hours on Craigslist	2008		Verde Burns Productions
10	48 Hours	1982		Paramount Pictures
11	50 First Dates	2004	Marinwood Oaks (140 Jefferson St.)	Columbia TriStar Home Entertainment
12	3,000 Hours	1916	Golden Gate Park	Stuyvesant-Hudson Film Studios
13	A Time to Die	1915	200 and Mission Streets	The Essanay Film Manufacturing Company
14	A Night Full of Rain	1919	San Francisco Chronicle (301 W. ...)	The Essanay Film Manufacturing Company

SOFTWARE APPLICATIONS

JMP Pro



<https://oit.duke.edu/comp-print/software/license/detail.php?id=4>

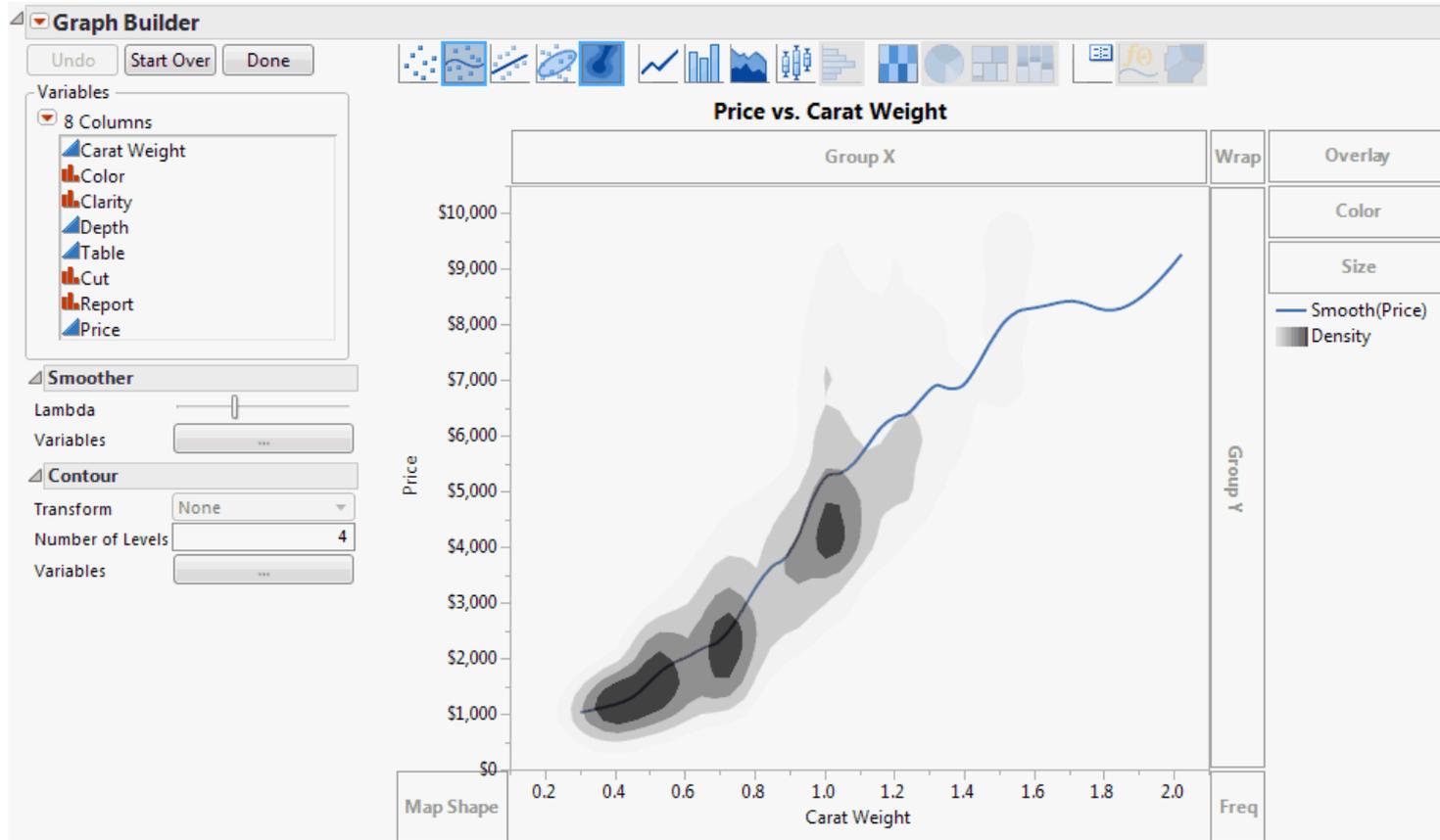
http://www.jmp.com/support/help/Essential_Graphing.shtml

JMP: Essential Graphing

- Overlay Plots
- Scatterplot 3D
- Contour Plots
- Bubble Plots
- Parallel Plots
- Cell Plots
- Treemaps
- Scatterplot Matrix
- Ternary Plots
- Summary Charts
- Create Maps

http://www.jmp.com/support/help/Essential_Graphing.shtml

Example: Contour Plot

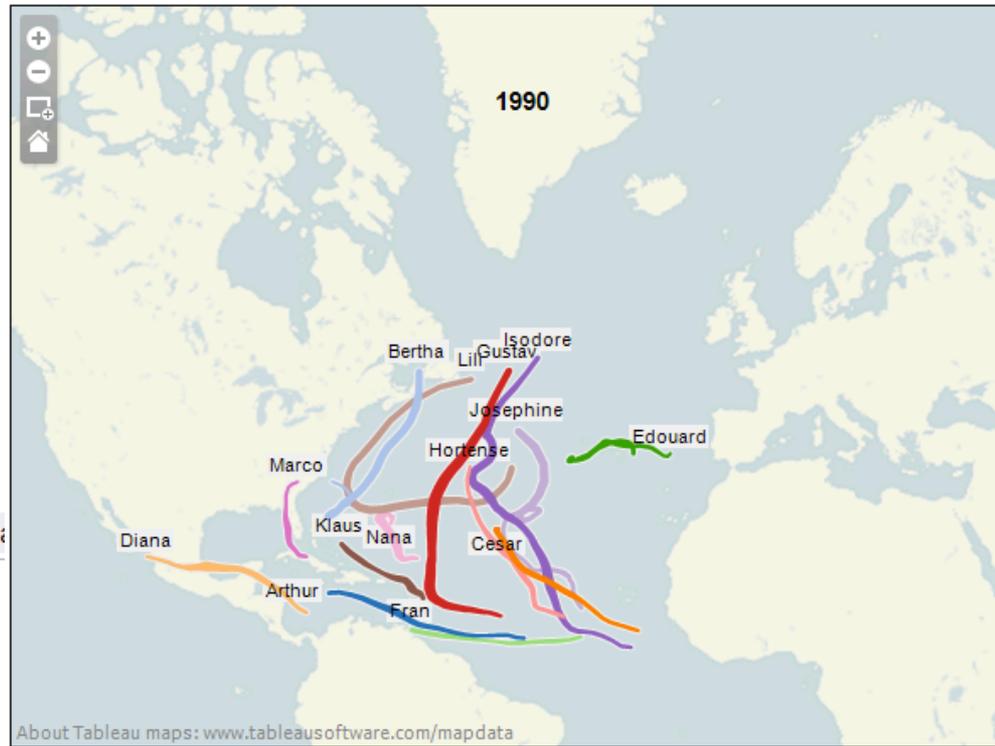
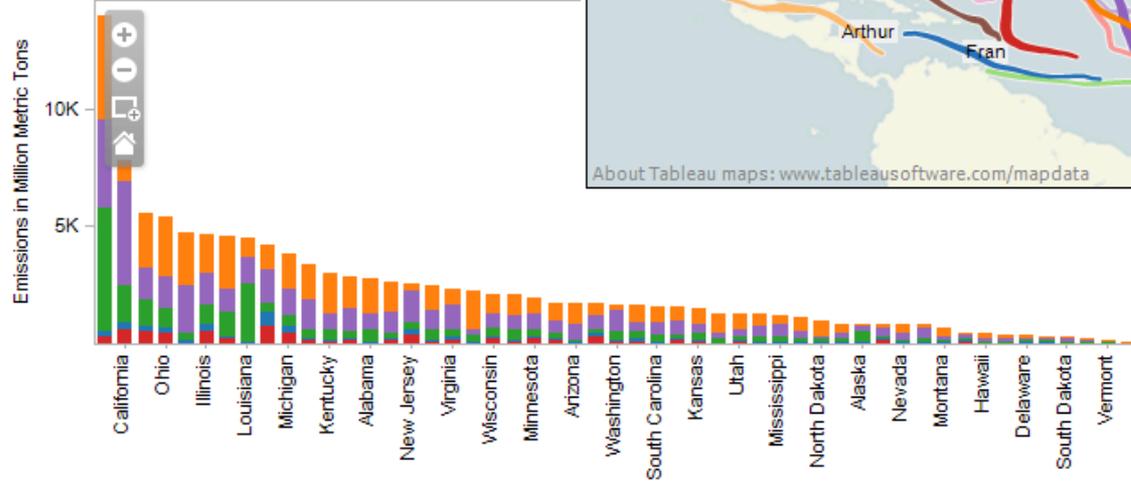


JMP Pro

- Statistical software
- Drag-and-drop chart builder
- Good charting options, including a basic map
- Can save code for all charts
(good for reproducibility)
- Can save vector graphics from charts
(good for print publications and graphic design work)

Tableau

Emissions by State



What can Tableau make?

- Text tables
a grid representing variables by size and color
- Heat maps
a grid representing variables by text and color
- **Maps (symbol, filled)**
- Pie charts
- Horizontal bars
- Stacked bars
- Side-by-side bars
- **Treemap**
a grid representing variables by size

- Circle views
- Side-by-side circles
- Lines/Area charts
- Lines/Area charts (discrete)
- Dual lines
- Dual combination
- Scatter plots
- **Histogram**
- **Box-and-whisker**
- **Gantt**
- **Bullet graphs**
- **Packed bubbles/Word cloud**



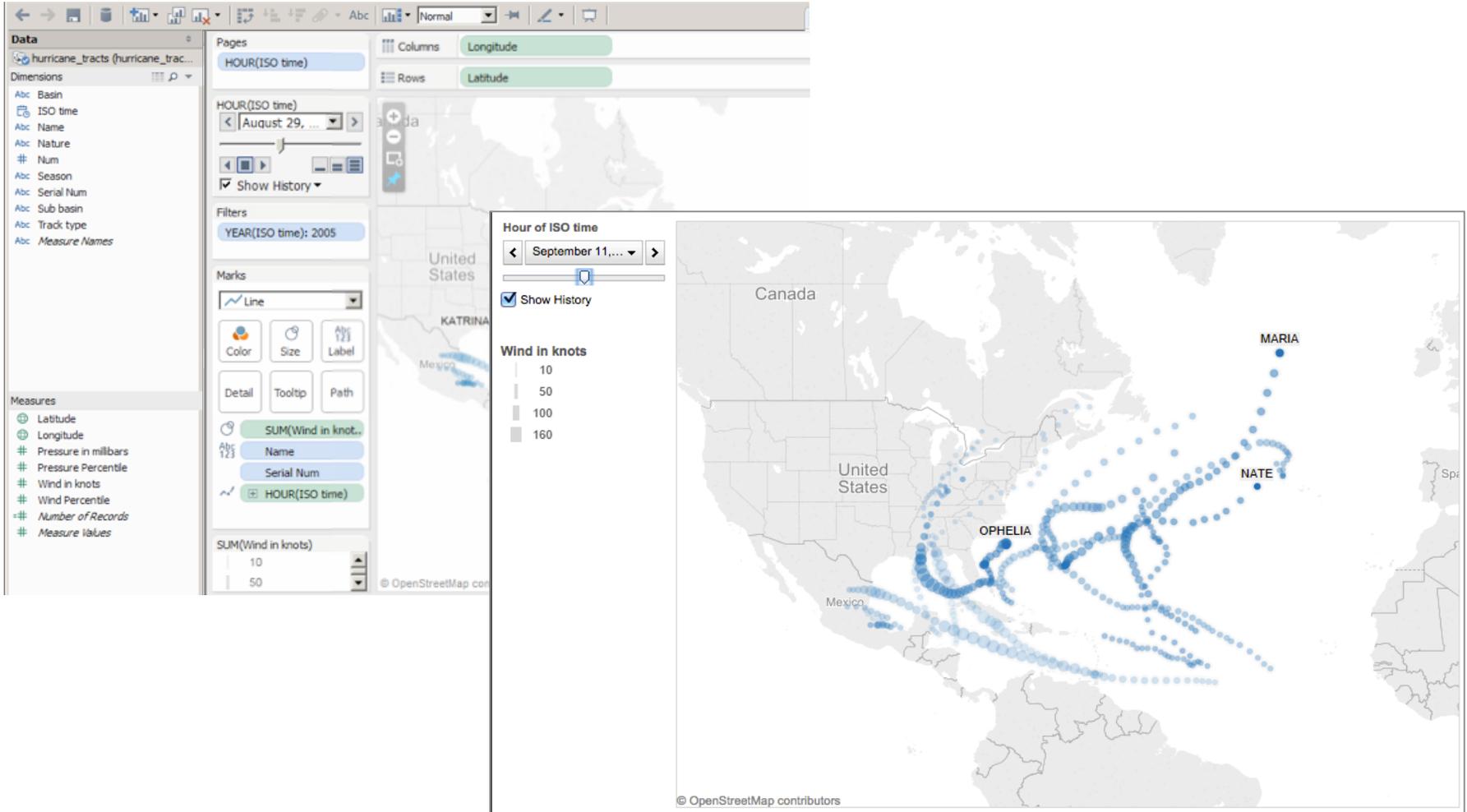
Tableau Desktop

- Built specifically for visualization
- Can create interactive charts and dashboards
- Can post to the web
(but make sure data are safe to share)
- Not great for print charts
(basically have to take screenshots)

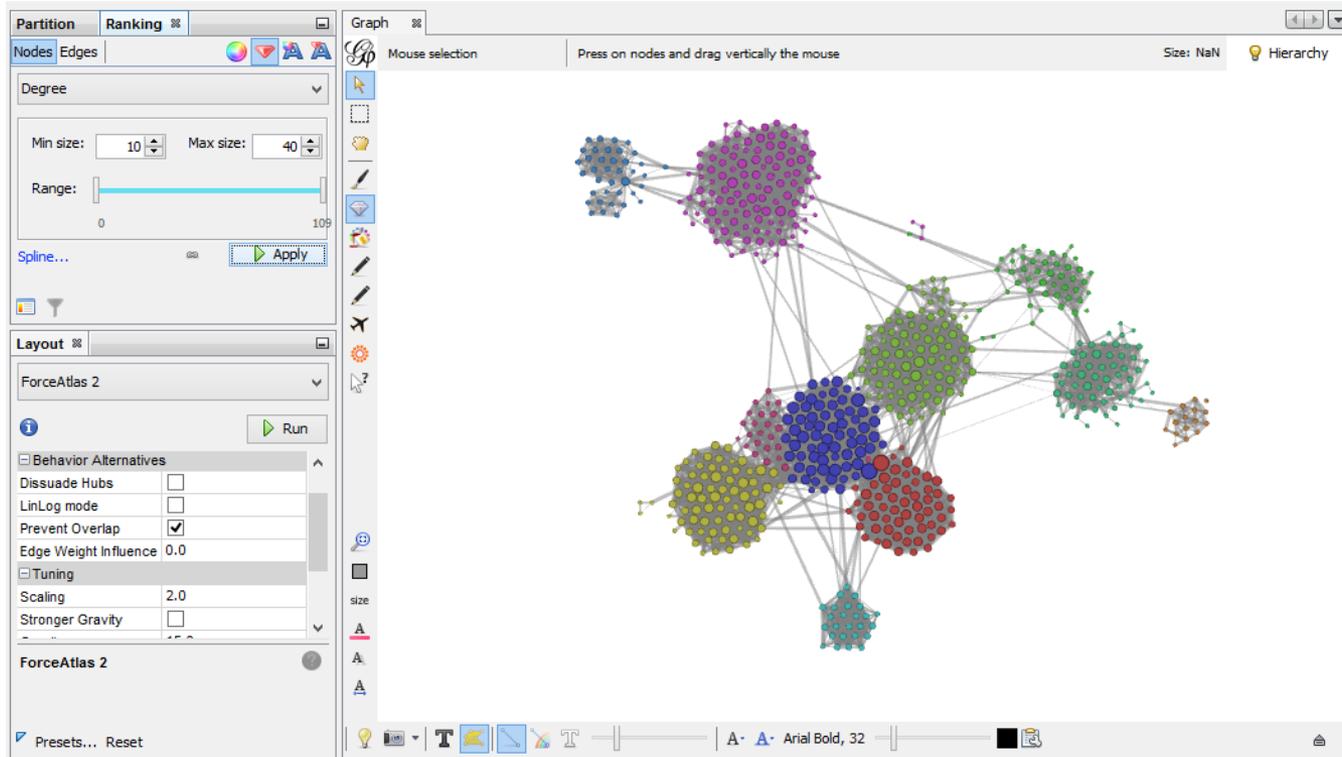
Free for students:

<http://www.tableau.com/academic/students>

Example: Animated Map



Gephi



http://bit.ly/gephi_workshop

Data formats

- Confusing number of choices
- GEXF supports many program features, but a pain to write by hand
- Spreadsheet is convenient and supports important features

	Edge List/Matrix Structure	XML Structure	Edge Weight	Attributes	Visualization Attributes	Attribute Default Value	Hierarchical Graphs	Dynamics
CSV	Yes	Yes						
DL Ucinet	Yes		Yes					
DOT Graphviz		Yes		Yes				
GDF		Yes	Yes	Yes	Yes			
GEXF		Yes	Yes	Yes	Yes	Yes	Yes	Yes
GML		Yes	Yes	Yes	Yes			
GraphML		Yes	Yes	Yes	Yes	Yes		
NET Pajek	Yes		Yes	Yes				
TLP Tulip								
VNA Netdraw		Yes	Yes					
Spreadsheet*		Yes	Yes					Yes

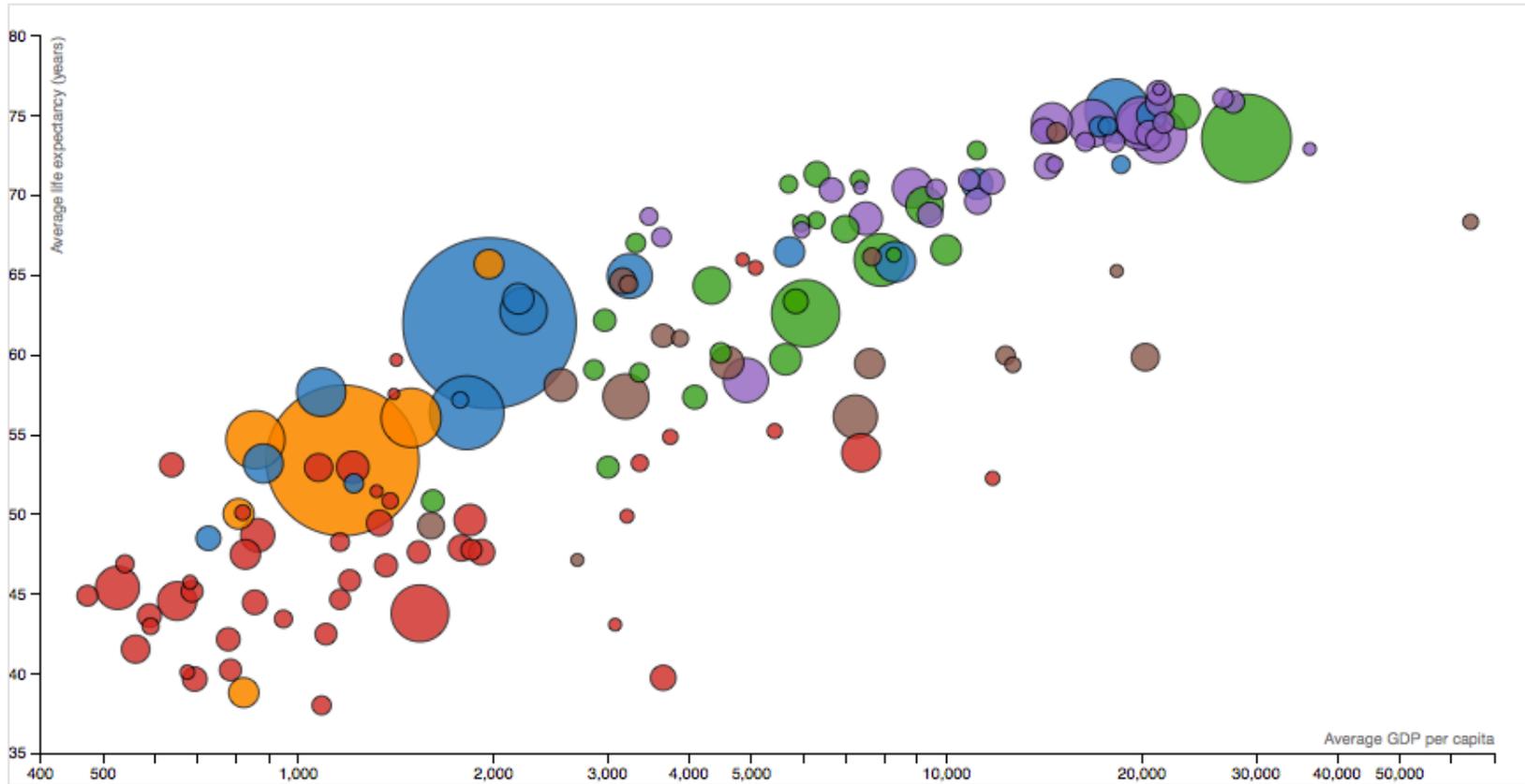
<https://gephi.org/users/supported-graph-formats/>

In addition to network visualization, Gephi can calculate:

- Degree (when directed, in-degree and out-degree)
- Diameter
 - Betweenness Centrality
 - Closeness Centrality
 - Eccentricity
- Density
- Clustering/Modularity

ADVANCED TOOLS

D3.js



<http://d3js.org>

<http://bl.ocks.org/dukevis/8782982>

About D3

- JavaScript library
- Fairly low level; building with rectangles and circles and lines, instead of pre-made chart structures*
- Basic functioning makes it easy to join HTML elements with data points

*D3 Middleware

Basic line/area chart:

- xCharts ~ 10 lines?
<http://tenxer.github.io/xcharts/>
- Rickshaw (specifically for time series) ~ 16 lines
<http://code.shutterstock.com/rickshaw/>
- NVD3 ~ 31 lines
<http://nvd3.org/>
- Vega ~ 57 lines
<http://trifacta.github.io/vega/>

http://chimera.labs.oreilly.com/books/1230000000345/ch02.html#_tools_built_with_d3



*D3 Middleware, cont'd.

- DC (<http://dc-js.github.io/dc.js/>)
good for dashboards (includes Crossfilter)
- D3plus (<http://d3plus.org/>)
good for tool tips and info panels
- Dimplejs (<http://dimplejs.org/>)
good for annotations, very pretty

D3 Resources

- Interactive Data Visualization for the Web
<http://chimera.labs.oreilly.com/books/1230000000345>
- Tutorial and Cheat Sheet, c. 2012
www.jeromecukier.net/blog/2012/10/15/d3-tutorial-at-visweek-2012/
- D3 Tips and Tricks
<https://leanpub.com/D3-Tips-and-Tricks/read>

When to use D3

- Need for customized chart types
(<http://bl.ocks.org/mbostock>)
- Want to use JavaScript
- Have only a low number of data points or elements
(SVG vs. HTML5 Canvas)
- Want to have it on your résumé

D3 workshop tomorrow!

Visualization in d3

Thursday, March 19, 7-9pm

Edge Workshop Room (Bostock 1st Floor)

<http://library.duke.edu/events/data/event.do?id=6817&occur=15231>

(Workshop is full, but if there are no-shows you could try to sneak in.)

Python

- Bokeh

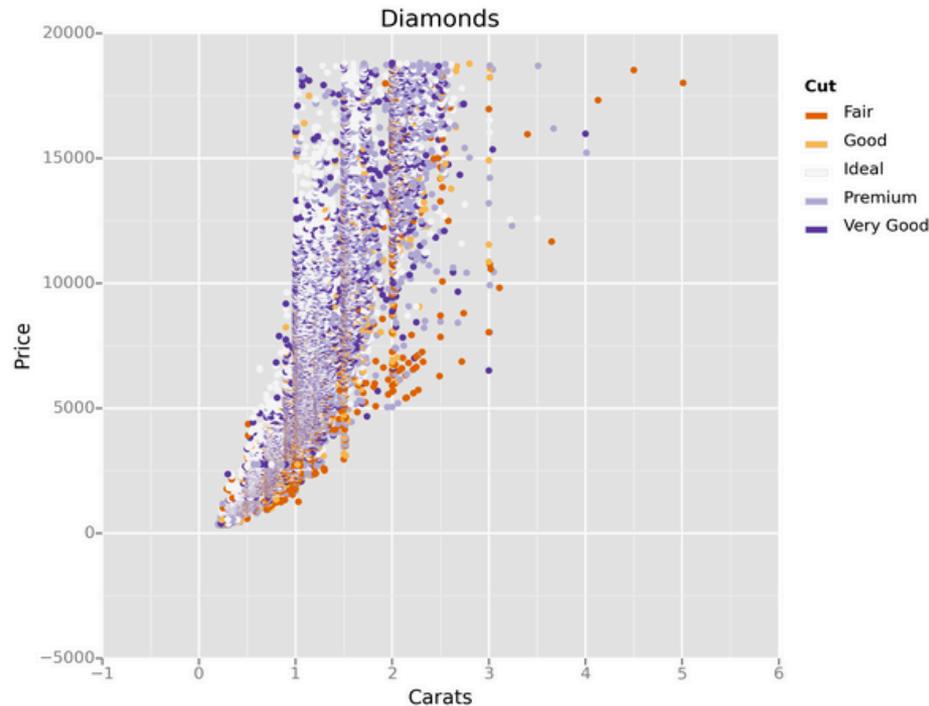
<http://bokeh.pydata.org>

web visualizations with big datasets



Python

- ggplot2 for python
<http://ggplot.yhathq.com/>
includes good graphical principles

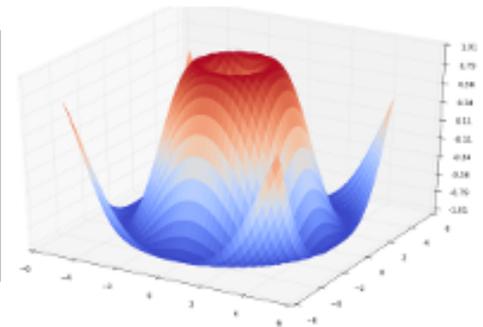
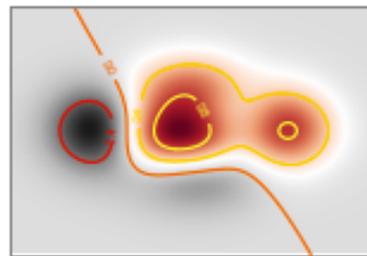
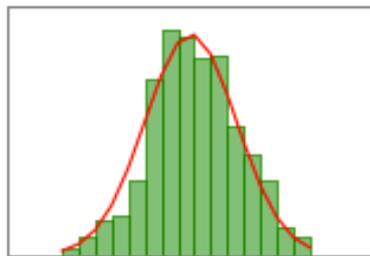
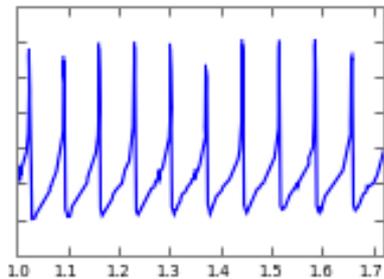


Python

- Anaconda

<https://store.continuum.io/cshop/anaconda/>

good for installing many data analysis packages, including matplotlib



<http://matplotlib.org/>

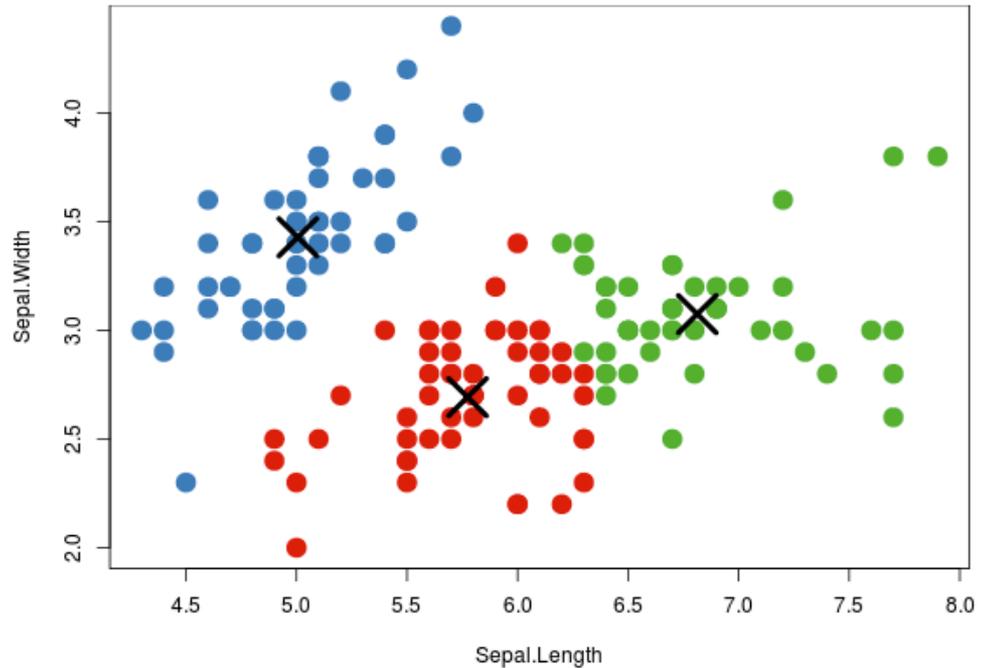
R Shiny

Iris k-means clustering

X Variable
Sepal.Length ▼

Y Variable
Sepal.Width ▼

Cluster count
3



<http://shiny.rstudio.com/>

<https://vm-manage.oit.duke.edu/>

<http://www.shinyapps.io/>

ELK stack



<http://www.elasticsearch.org/>

ELK stack

- Elasticsearch
 - “flexible and powerful open source, distributed, real-time search and analytics engine” · full-text search (lucene) plus fast queries and many built-in aggregations for large data (time-based and stats w/facets)
- Logstash
 - “helps you take logs and other time based event data from any system and store it in a single place” · parse
- Kibana
 - “Elasticsearch’s data visualization engine” · sharable dashboards for real-time, interactive visual exploration

ELK stack

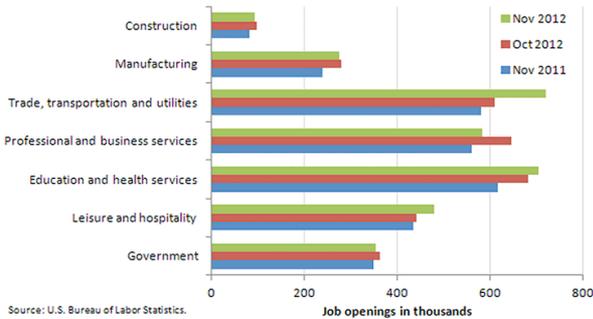
- Open source, but company builds APIs for all major languages
- Potential end-to-end solution for storage, plus monitoring by both developers and customers
- Geared towards large time-based, geo-spatial, and textual data
- Free for academic use
- Security product is pay only

MORE TIPS

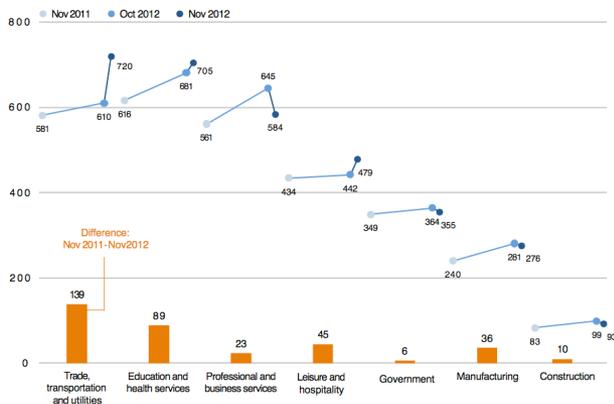
Good Chart Makeover Examples

The Why Axis chart remakes
<http://thewhyaxis.info/remakes/>

Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted



Job openings by industry, November 2011, October 2012 and November 2012, seasonally adjusted (Thousands of jobs)

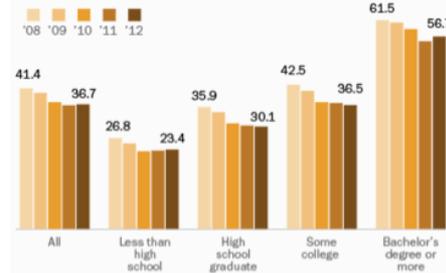


Storytelling With Data visual makeovers:
<http://www.storytellingwithdata.com/search/label/Visual%20Makeover>

ORIGINAL

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



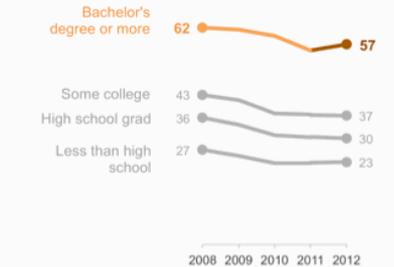
Note: Marriage eligible includes the newly married plus those widowed, divorced or never married at interview.
 Source: US Census

PEW RESEARCH CENTER

COLE's MAKEOVER

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced or never married at interview.
 Source: US Census

PEW RESEARCH CENTER

On the web

- Bad examples:

WTF Viz, <http://wtfviz.net/>

- Good examples:

Thumbs Up Viz, <http://thumbsupviz.com/>

- Ask for help:

Help Me Viz, <http://helpmeviz.com/>

More on Data Visualization

Visual communication:

<http://guides.library.duke.edu/visualcomm>

Data visualization:

<http://guides.library.duke.edu/datavis/>

Top 10 dos and don'ts for charts and graphs:

<http://guides.library.duke.edu/topten>

GETTING HELP

Data and Visualization Services

- Data collections, LibGuides, etc.
<http://library.duke.edu/data/>
- Blog (tutorials, announcements, etc.)
<http://blogs.library.duke.edu/data/>
- Walk-in consultations
<http://library.duke.edu/data/about/schedule>
(or by appointment – askdata@duke.edu)
- Data and Visualization Lab in the Edge
<http://library.duke.edu/data/about/lab>
(fast hardware, diverse software)
- Additional workshops
<http://library.duke.edu/data/news/>
(listserv – dvs-announce@duke.edu)

QUESTIONS? SUGGESTIONS?

angela.zoss@duke.edu

http://twitter.com/duke_vis