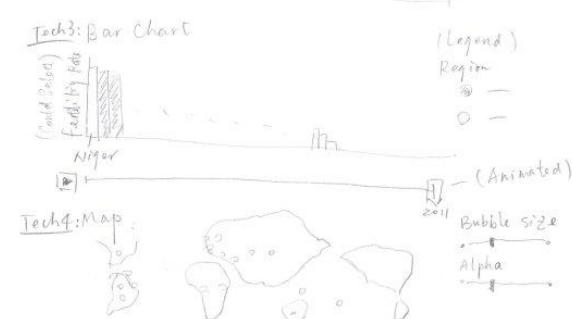
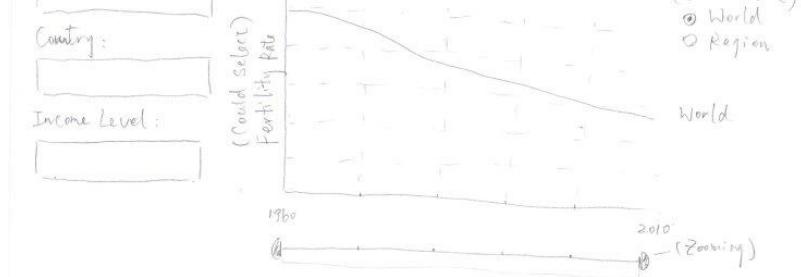
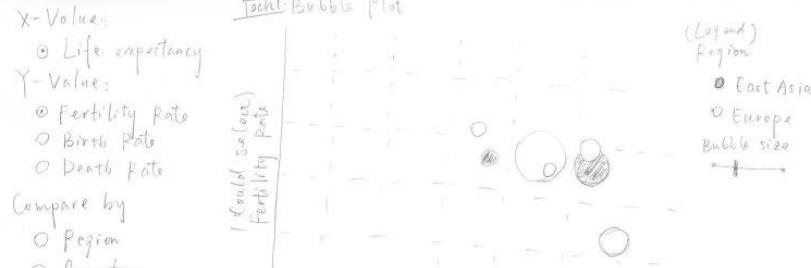


Multivariate and Time Series Data Visualization

Yannet Interian -- USF

Project Sketch Due Thursday 4/20



Agenda

- Multivariate
 - Bubble Chart
 - Heatmap
 - Scatterplot Matrix
 - Small Multiples
 - Parallel Coordinates
 - Spider Plots
- Time Series
 -

Bubble Charts

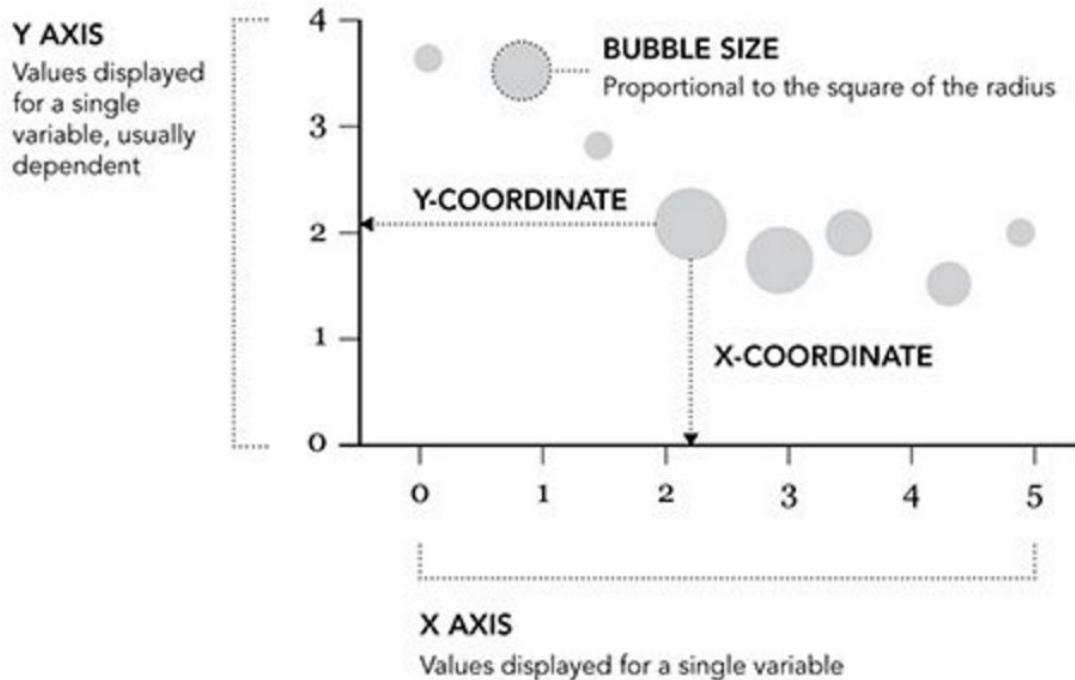
Bubble Charts: encoding data

X axis: continuous data

Y axis: continuous data

Bubble size: numerical data

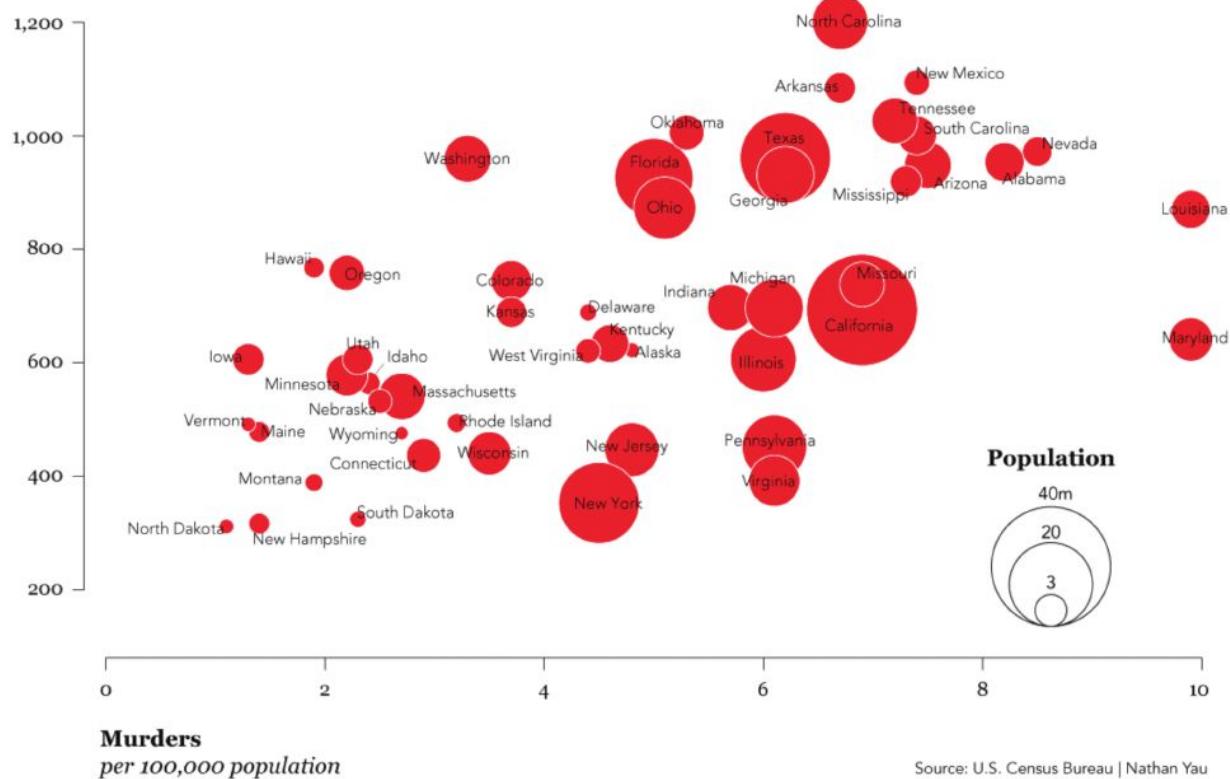
Color: numerical or categorical



MURDERS VERSUS BURGLARIES IN THE UNITED STATES

Burglaries

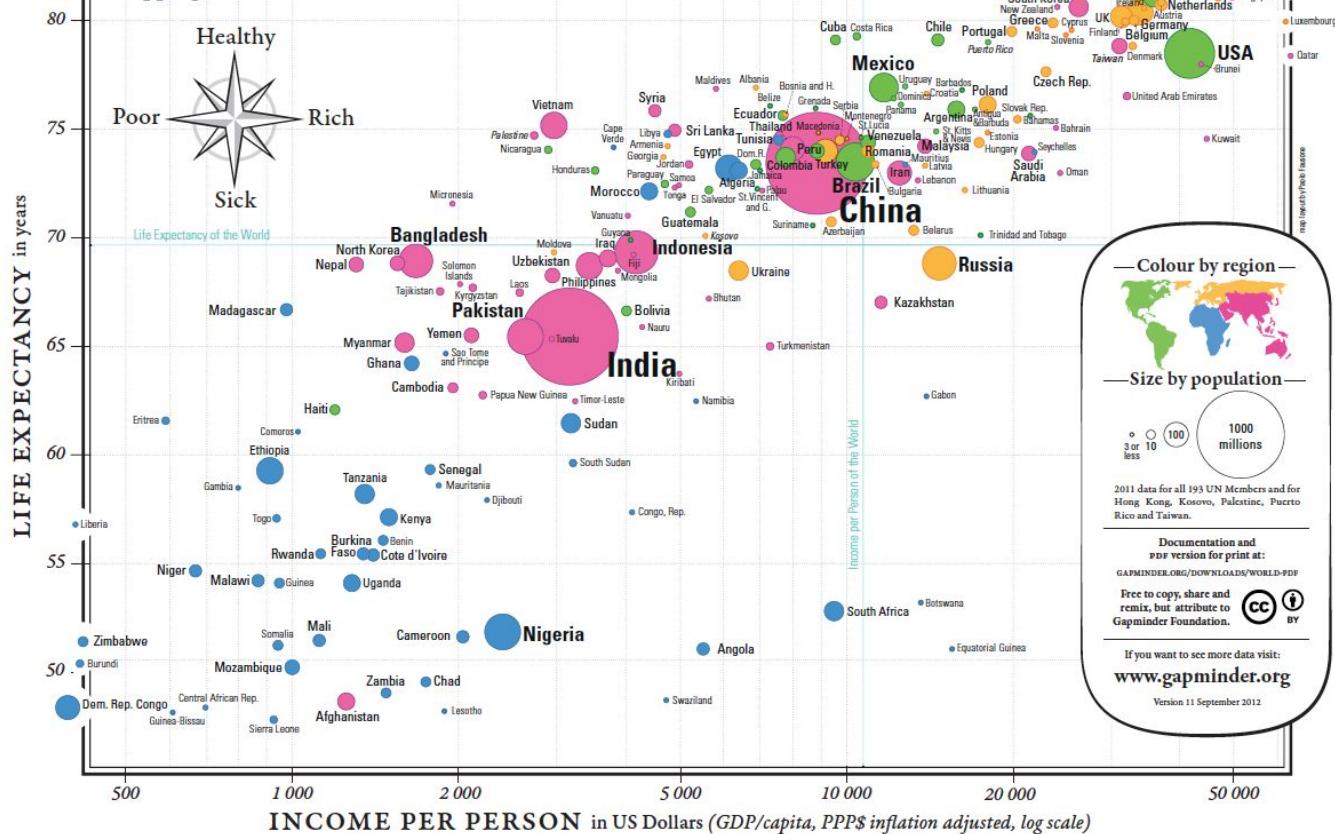
per 100,000 population



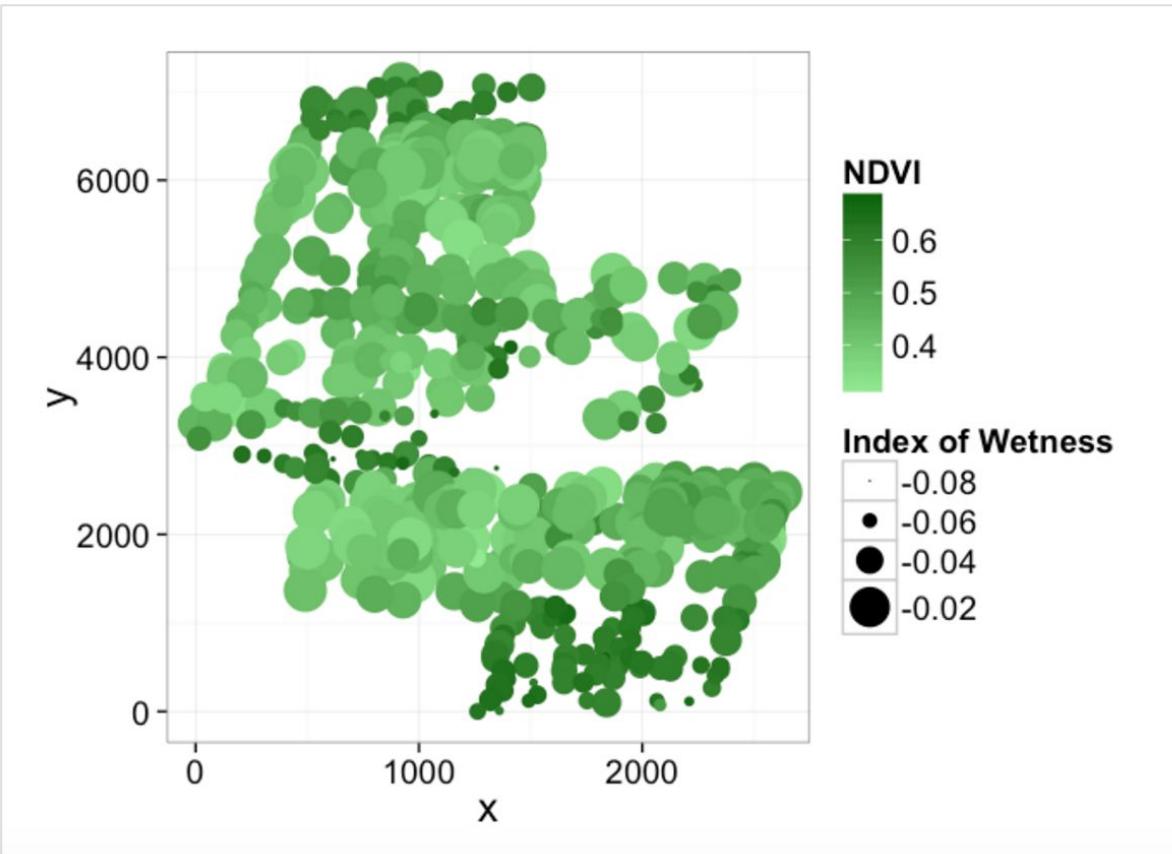
Source: U.S. Census Bureau | Nathan Yau

GAPMINDER WORLD 2012

Mapping the Wealth and Health of Nations



```
qplot(x, y, data=boreal, size=Wet, color=NDVI) +  
  theme_bw(base_size=18) +  
  scale_size_continuous("Index of Wetness", range=c(0,10)) +  
  scale_color_gradient("NDVI", low="lightgreen", high="darkgreen")
```



Discussion

- Able to encode four dimensions of data
 - Ideal if one dimension is categorical (color)
- Rough comparison possible
 - Beware comparing circle areas (bubbles should be sized by area)
- Obscuring data may be an issue
 - Large circles should be behind smaller ones
 - Issues increases with density

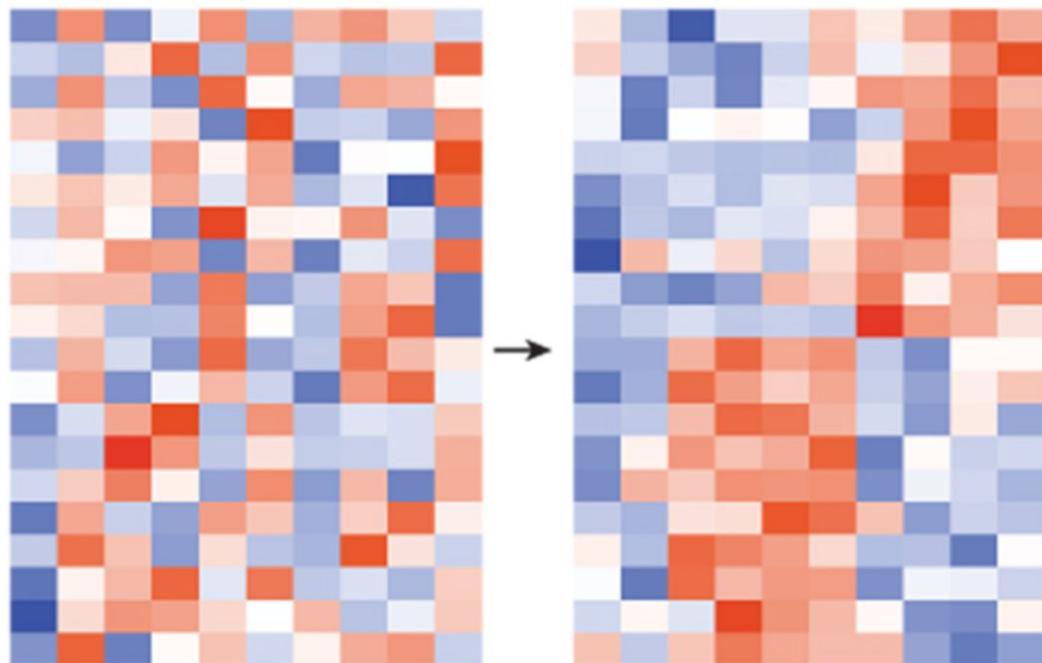
Heatmap

HeatMap: encoding data

- Encode Data
 - Horizontal position
 - Column from dataset
- Vertical position
 - Row from dataset
- Box color
 - Value from row, column in dataset
 - Numerical or categorical

HeatMap: good practice

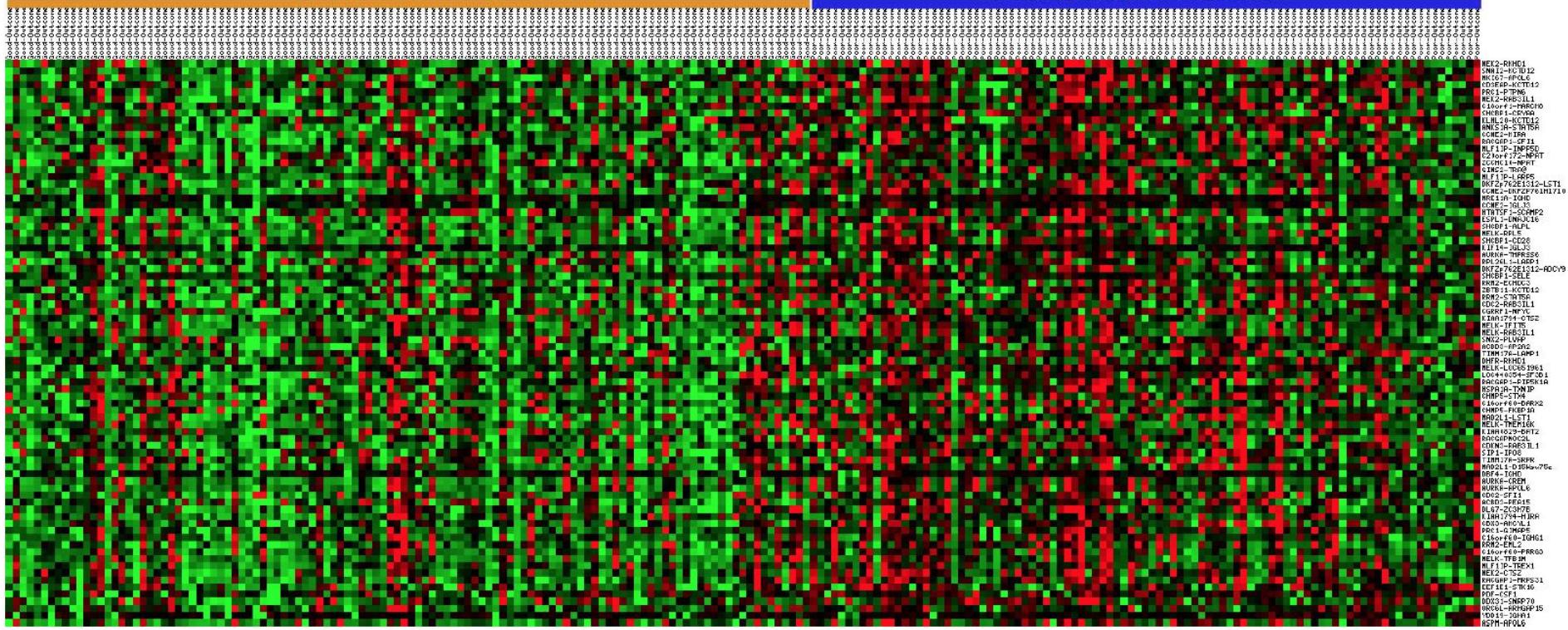
Clustering can dramatically affect our ability to see structure in heat maps



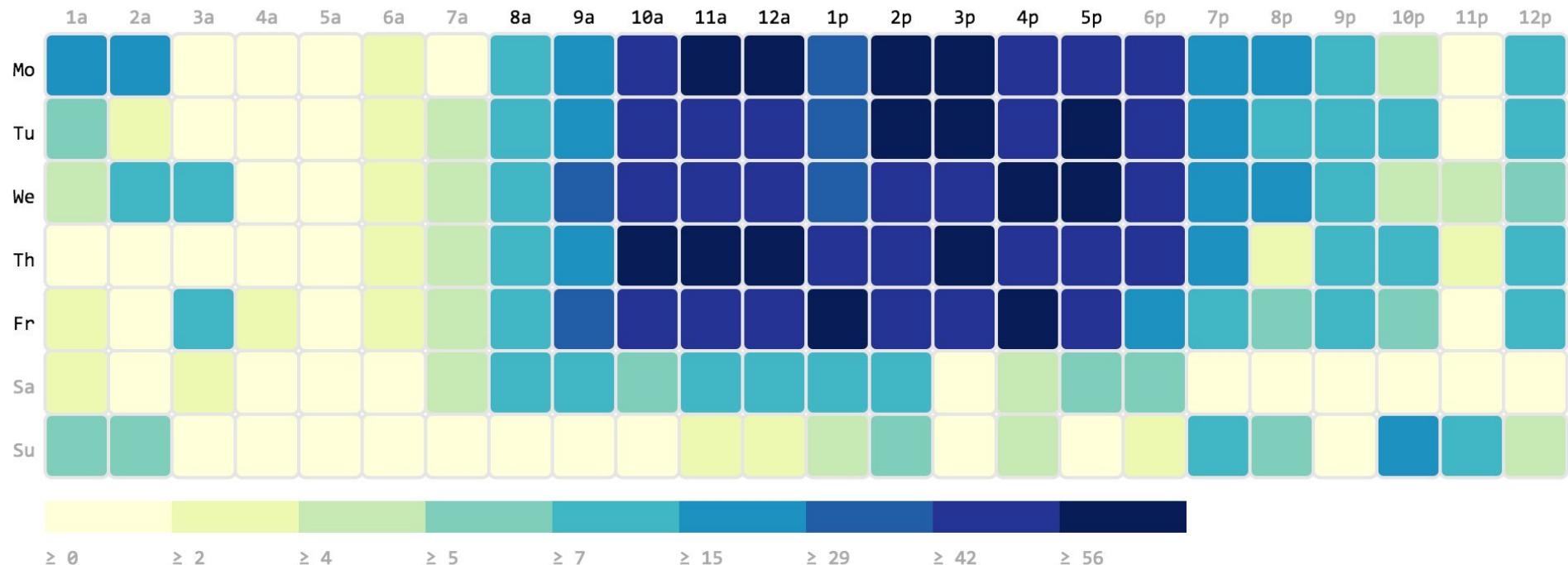
-1.3 1.7

Good-outcome patients

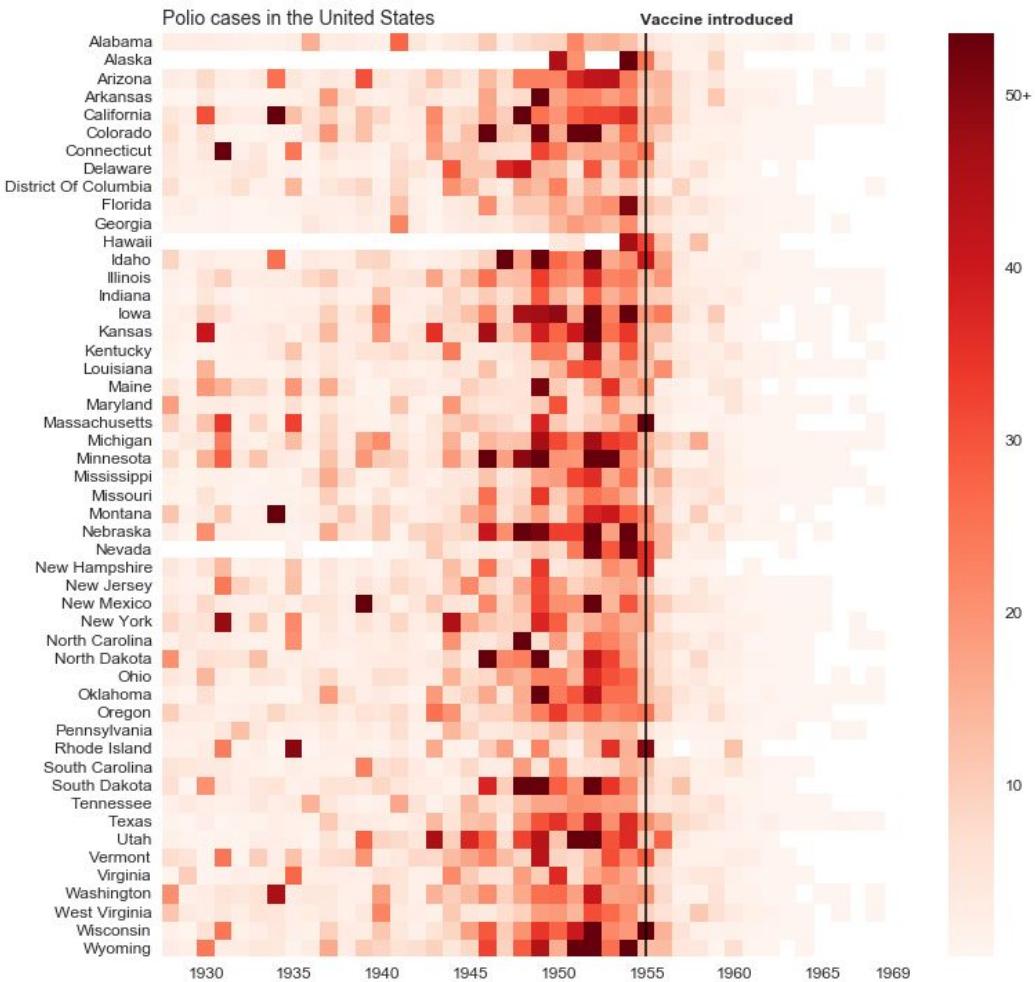
Poor-outcome patients



User sessions by day / hour

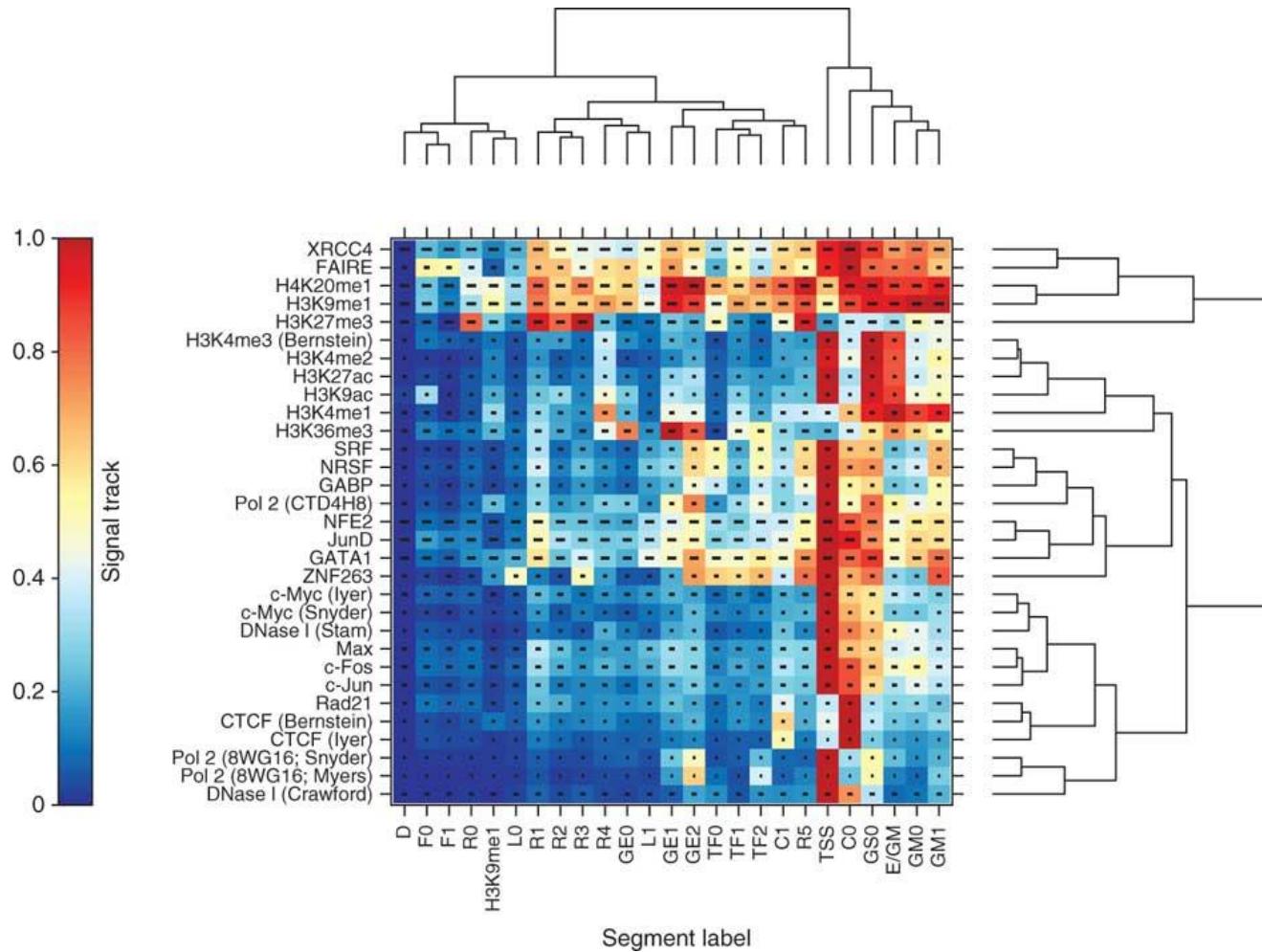


Polio cases in the US



Data source: Project TYCHO (tycho.pitt.edu) | Author: Randy Olson (randalolson.com / @randal_olson)

<http://www.randalolson.com/2016/03/04/revisiting-the-vaccine-visualizations/>



HeatMap: Discussion

- Essentially showing the data set, but replaces numbers by color values
- Good for certain types of data
 - Continuous data well-suited
 - Unordered categorical limited to 7 categories
- Encourages comparison and pattern finding
 - Sorting changes patterns!
 - Hierarchical clustering

Scatterplot Matrix

Scatterplot Matrix: Encode Data

- Scatterplot
 - Horizontal position maps to one variable
 - Vertical position maps to another variable
- Matrix of scatterplots
 - Each scatterplot focuses on one pair
 - Which pair is determined by row and column

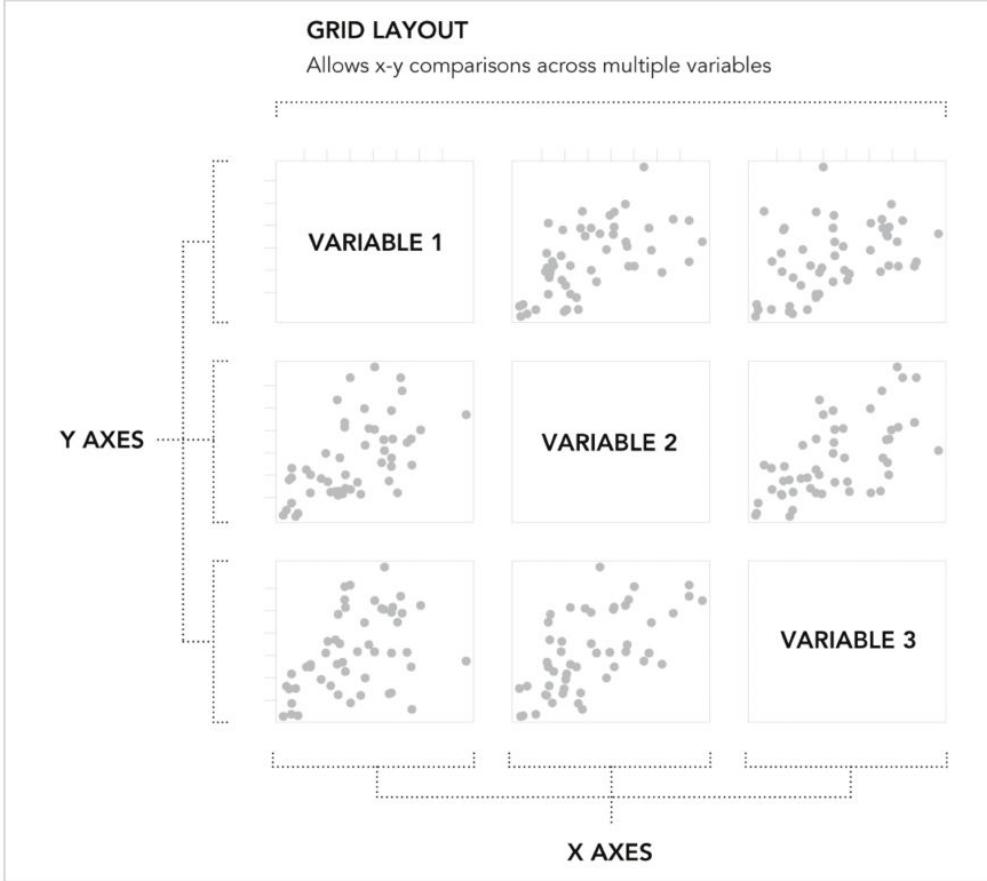


FIGURE 6-8 Scatterplot matrix framework

MURDERS VERSUS BURGLARIES IN THE UNITED STATES

States with higher murder rates tend to have higher burglary rates.

Burglaries

per 100,000 population

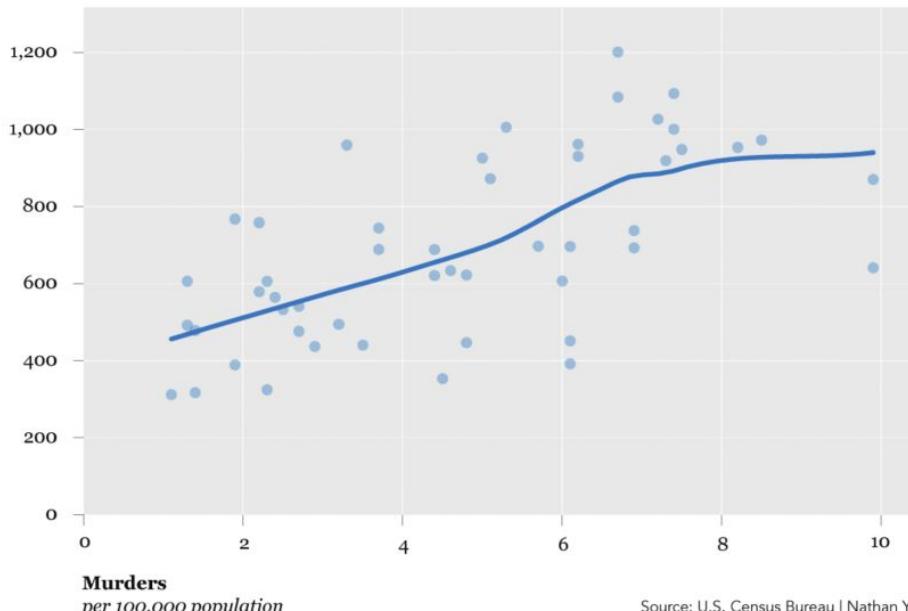


FIGURE 6-7 Revised scatterplot on murder versus burglary

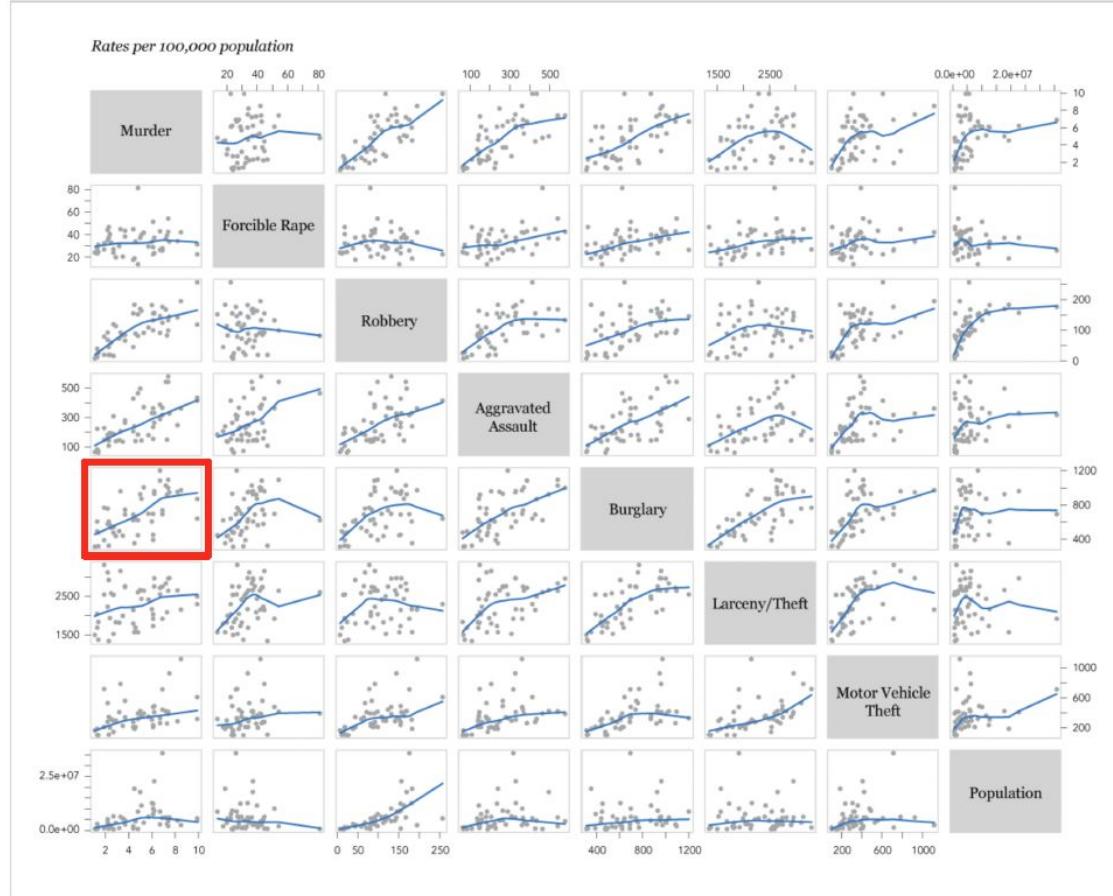
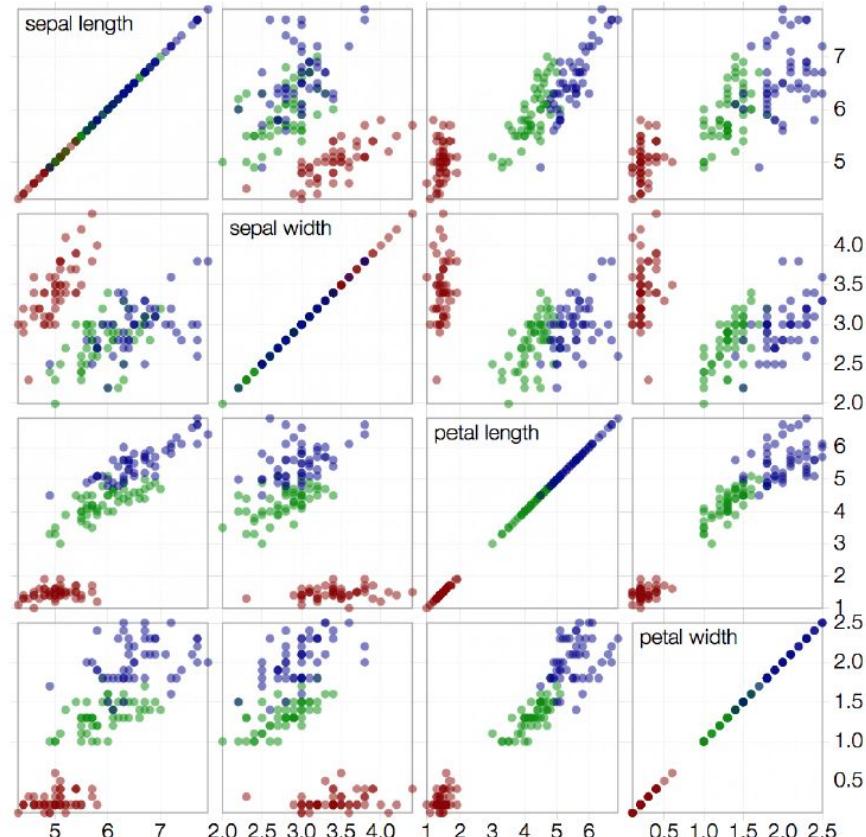


FIGURE 6-9 Scatterplot matrix of crime rates

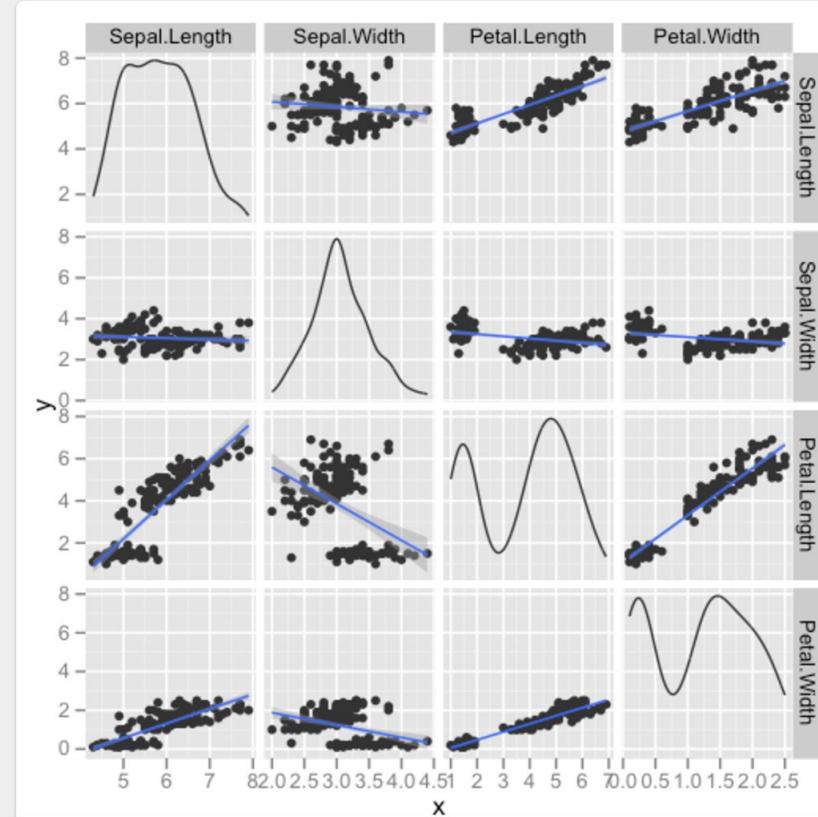
"Visualize This: The FlowingData Guide to Design, Visualization, and Statistics" by Nathan Yau, 2011.



- *Iris setosa*
- *Iris versicolor*
- *Iris virginica*

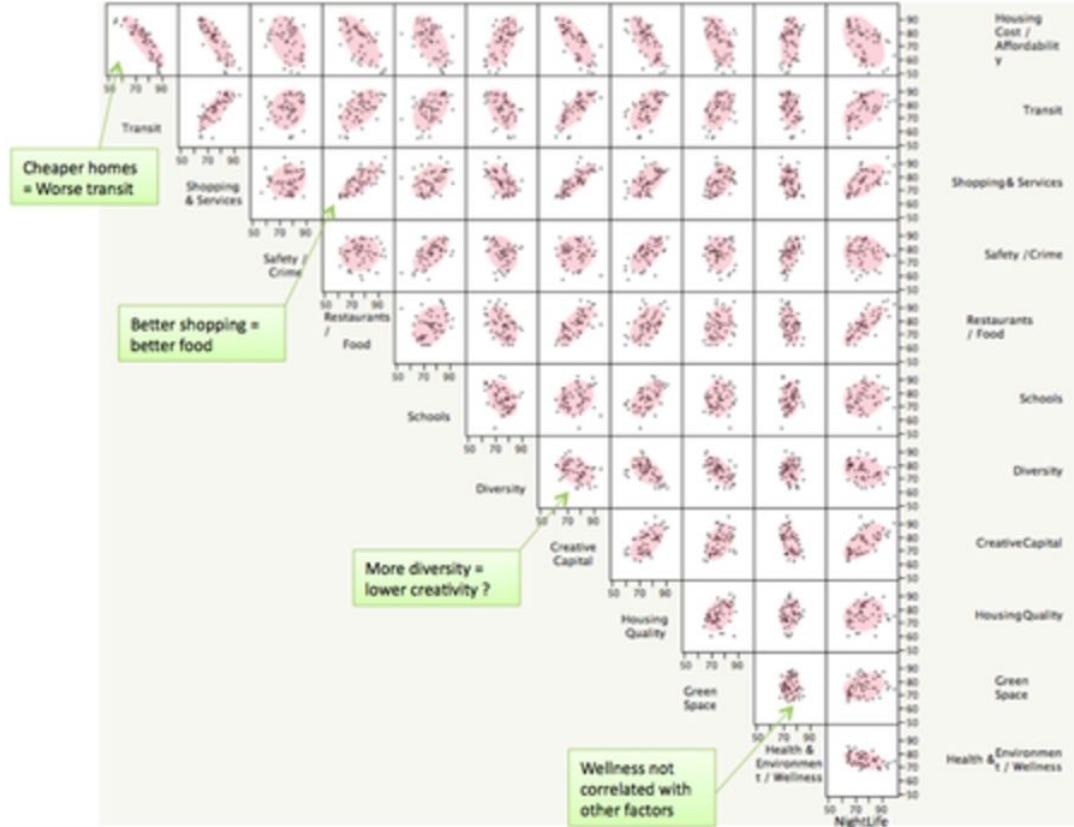
Edgar Anderson's *Iris* data set
scatterplot matrix

```
plotmatrix(iris[,1:4], colour="gray20") +  
geom_smooth(method="lm")
```



Correlations between “housing variables”

The Most Livable Neighborhoods in New York



<http://nymag.com/realestate/neighborhoods/2010/65374/>,

http://junkcharts.typepad.com/junk_charts/2010/06/the-scatterplot-matrix-a-great-tool.html

Scatterplot Matrix: Discussion

- Good for exploration and comparison
 - Can be a little overwhelming at first
- Works with numerical or ordered data
- Form of small-multiples plot
 - Multiple scatterplots

Small Multiples

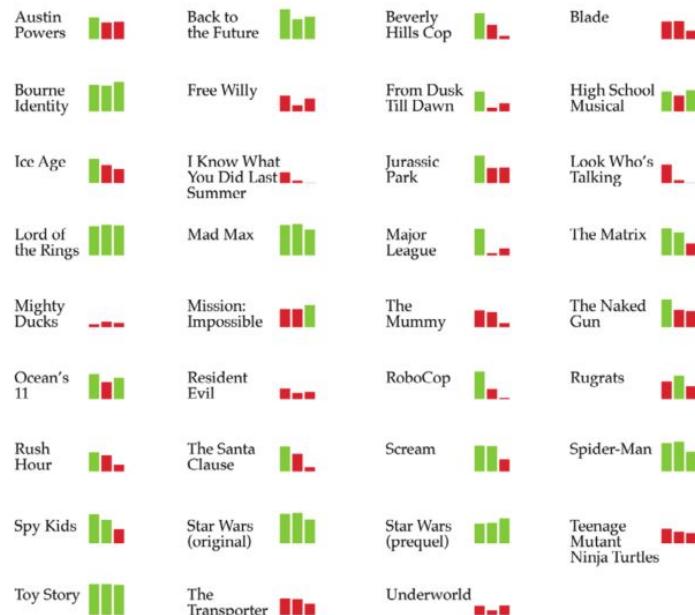
Small Multiples: Encode Data

- Group data by a variable to divide it into subsets
 - Usually a categorical variable
- Create a small plot for each subset
- Show all subset plots on same page
 - Encourages comparison

Fresh Originals / Rotten Finales

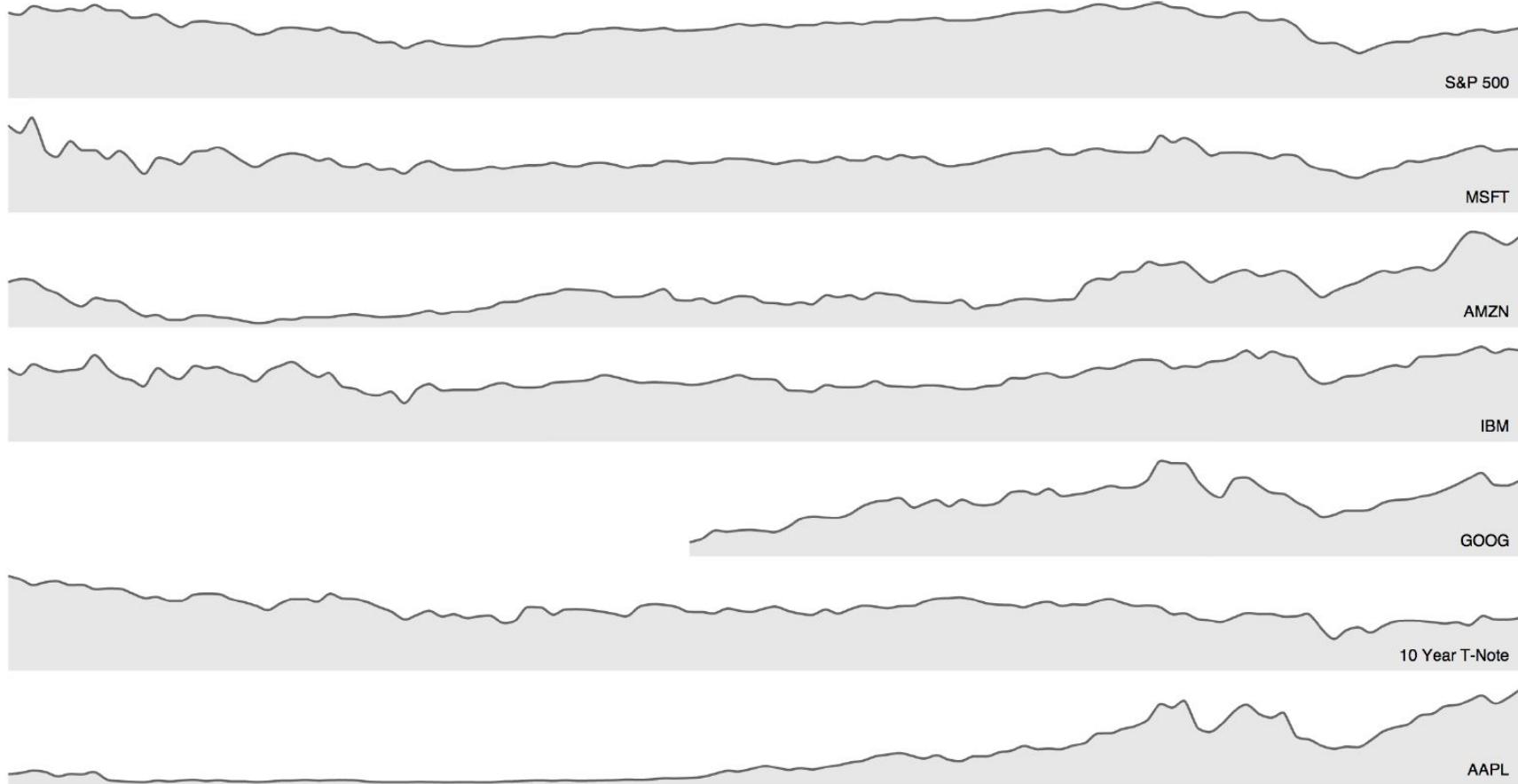
Out of the 35 selected trilogies, 23 of them had *fresh* originals; however, only 11 trilogies had a *fresh* finale.

■ Fresh (at least 60%)
■ Rotten



Source: Rotten Tomatoes, Wikipedia | By: FlowingData, <http://flowingdata.com>

FIGURE 6-40 Small multiples for ratings of trilogies



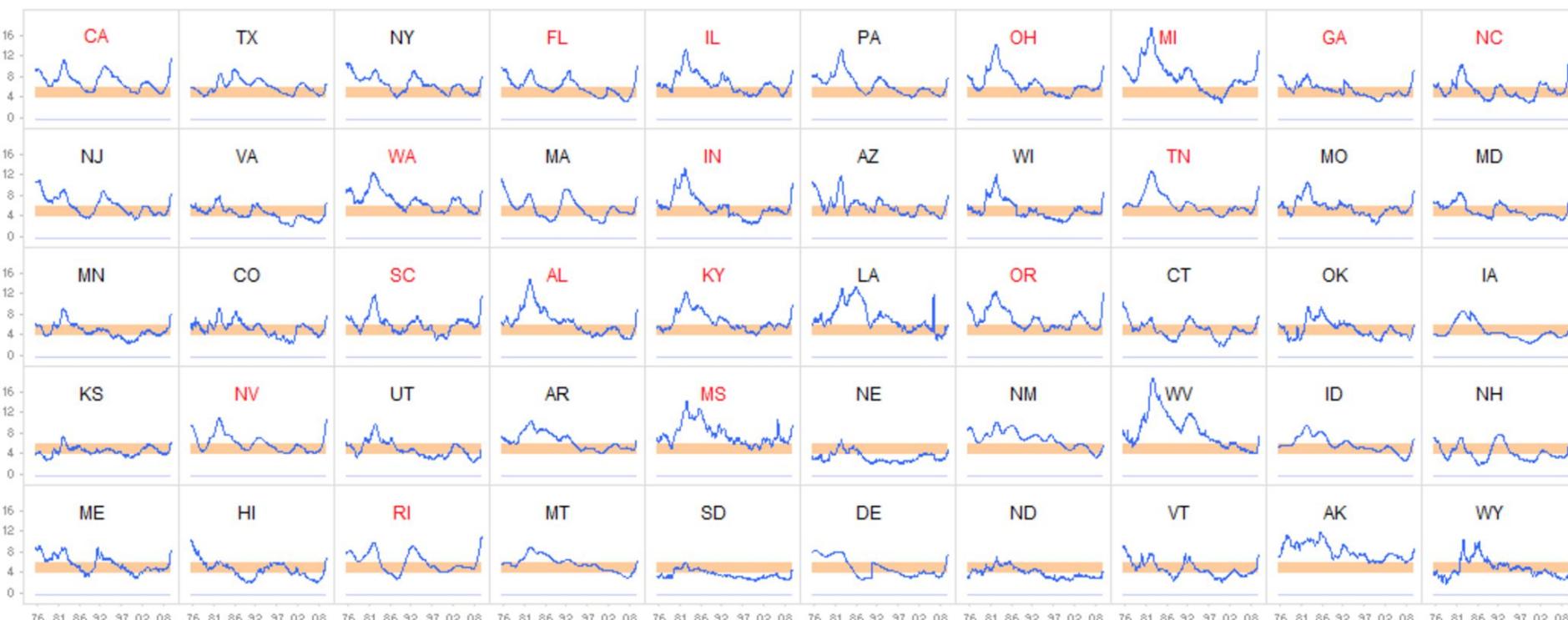
2000: State-level support (orange) or opposition (green) on school vouchers, relative to the national average of 45% support



State level support
(orange), opposition
(green) for school
vouchers relative to the
national average of 45%.

Orange and green colors correspond to states where support for vouchers was greater or less than the national average.
The seven ethnicreligious categories are mutually exclusive. "Evangelicals" includes Mormons as well as born-again Protestants.
Where a category represents less than 1% of the voters of a state, the state is left blank.

Monthly Unemployment Rates by State, Jan 1976 - Apr 2009

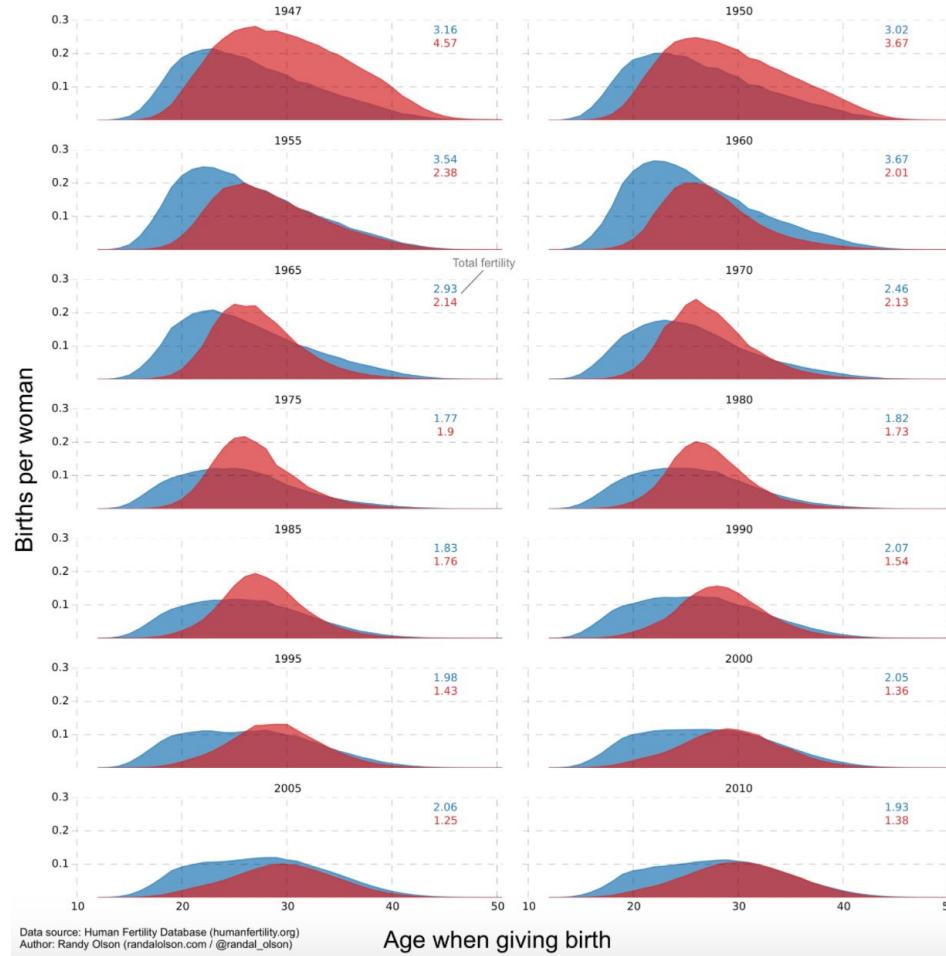


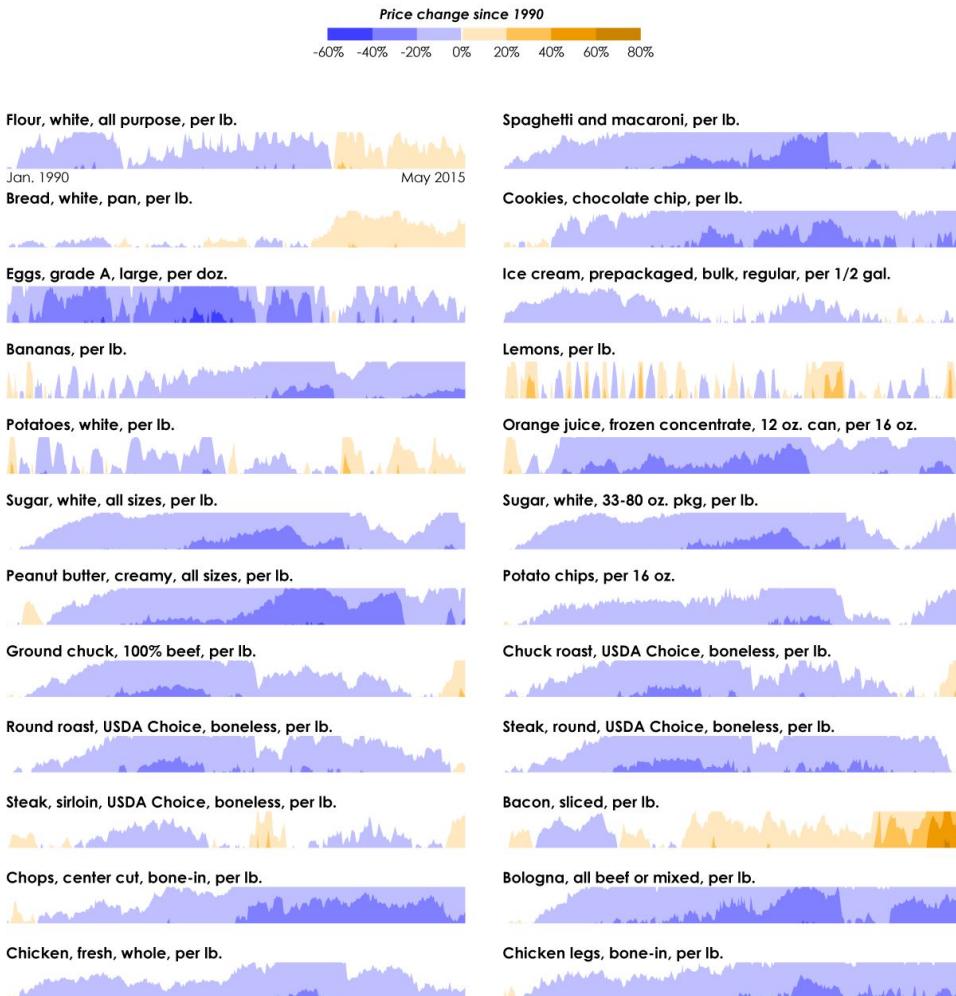
Source: Bureau of Labor Statistics

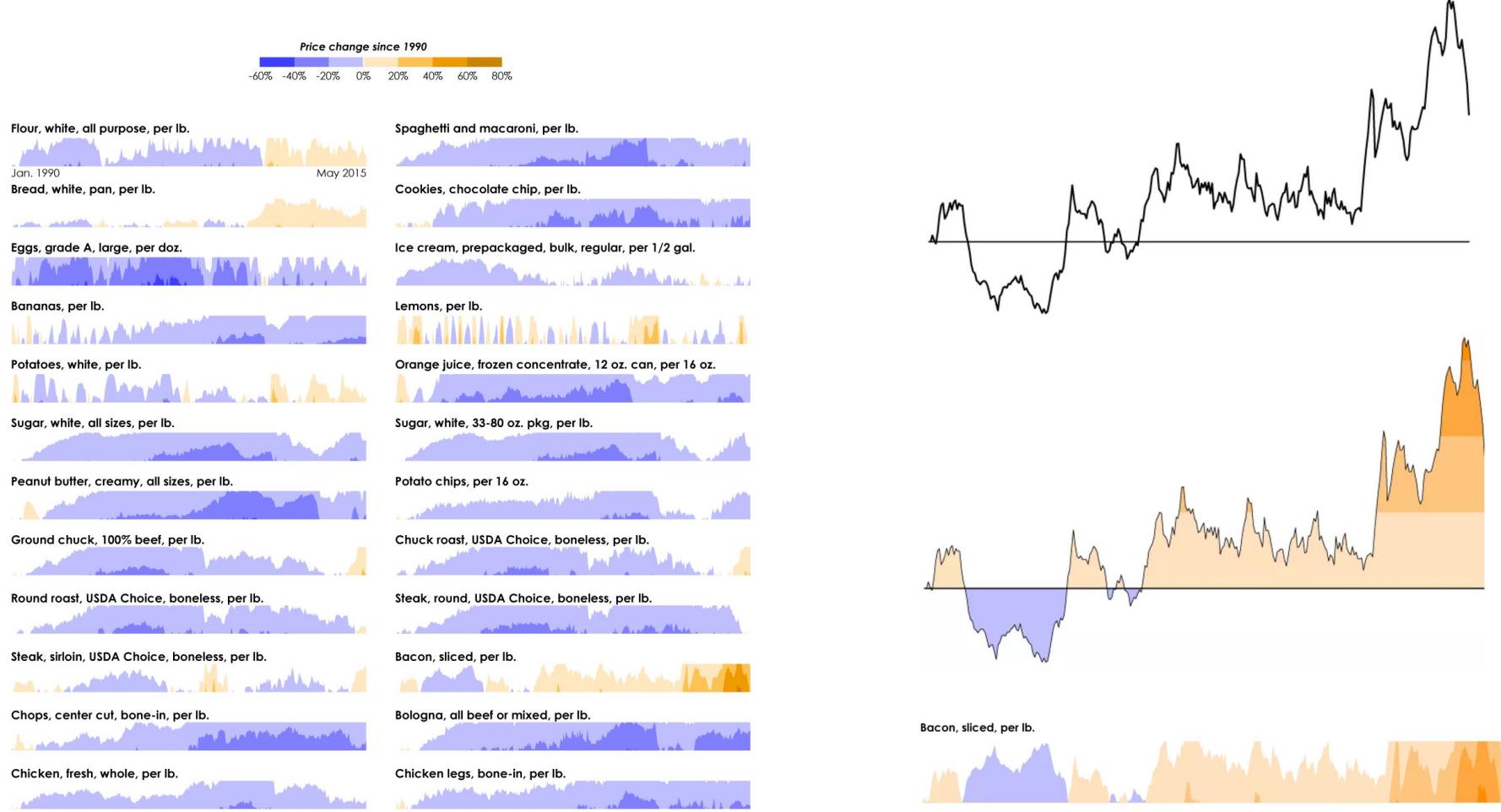
Notes: The orange band denotes a "normal" unemployment rate (4%-6%);

State code in red: unemployment rate in April 2009 is higher than the US average

Fertility in USA and Japan, 1947 - 2010







Small Multiples: Discussion

- Excellent for comparison
 - Depends on how to place the multiples
- Requires a variable to use for grouping
 - Discrete or categorical data
- Can use with any type of plot
- Harder to tell exact values

Parallel Coordinates

Parallel Coordinates: Encode Data

- Create one vertical line for every column
 - Numerical or ordered data
- Plot every row
 - x position is column
 - y position is value for that column
 - line connects values for a single row
- Picture an xy scatterplot, but putting both axis lines vertically

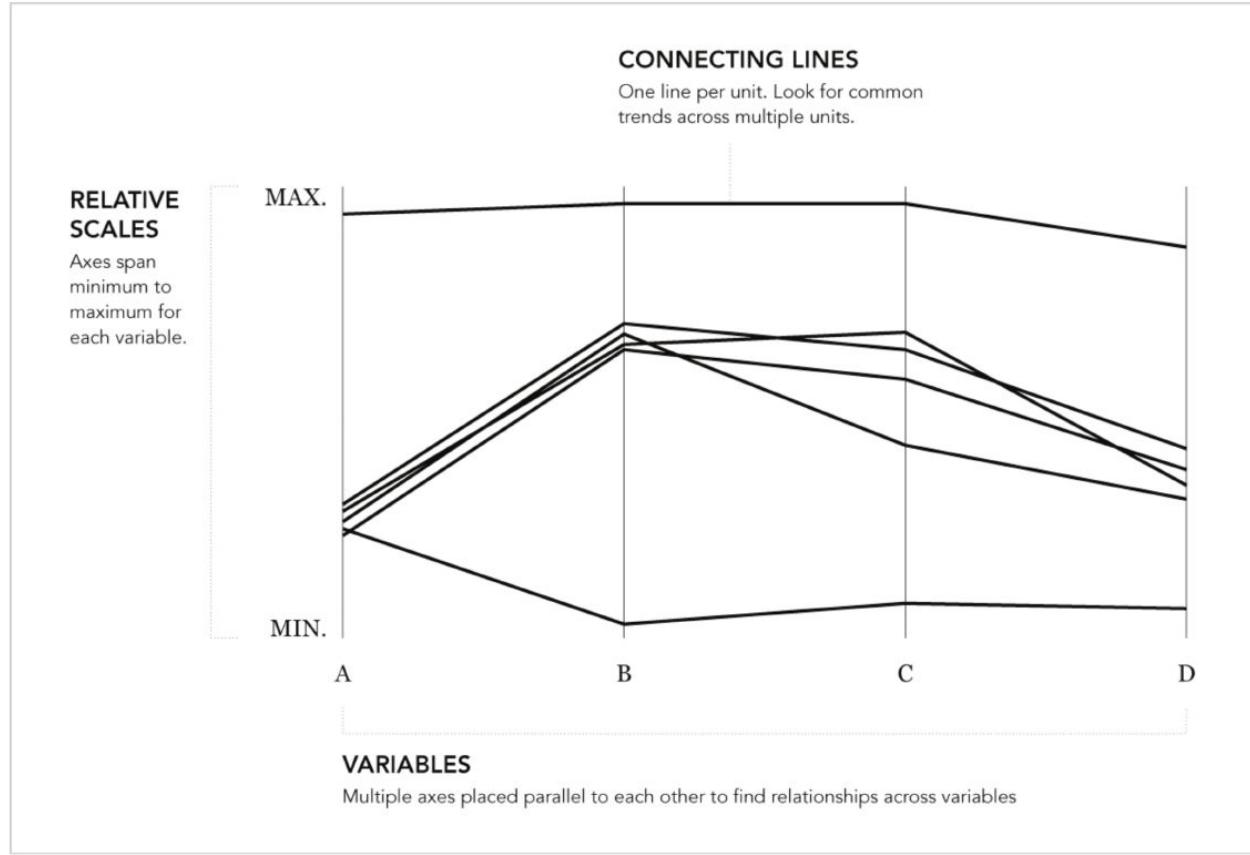


FIGURE 7-20 Parallel coordinates framework

EDUCATION IN THE UNITED STATES

States with higher SAT reading scores predictably also have relatively higher math and writing scores. However, this does not necessarily mean that students are better educated in these states. It's more likely an indicator for what percentage of graduates actually took the test.

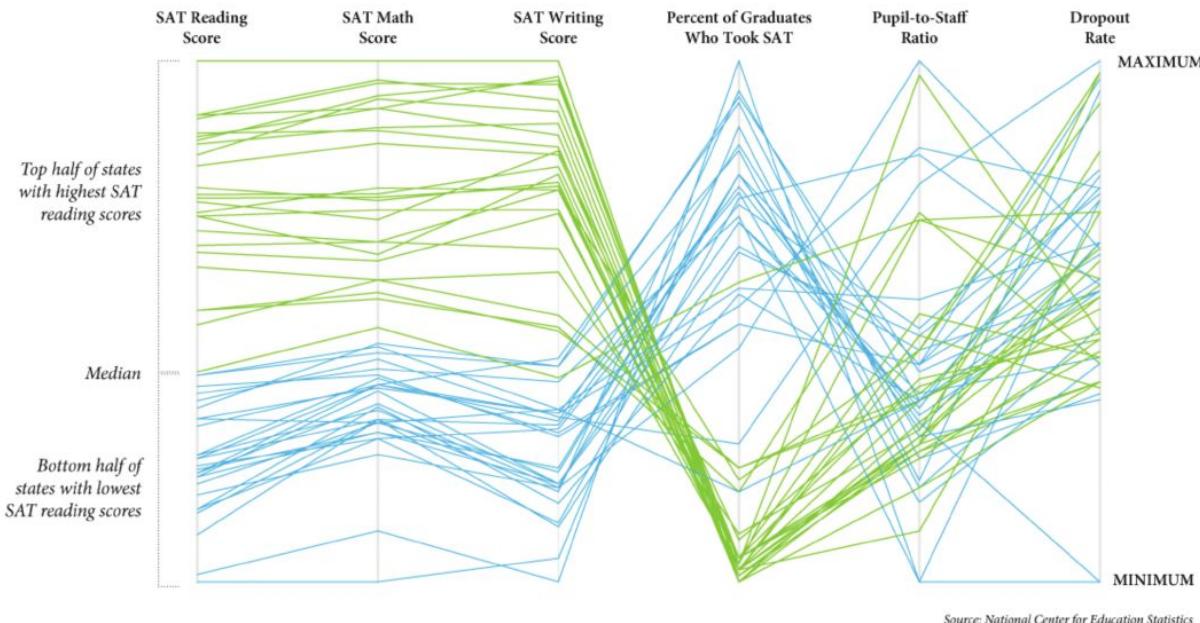
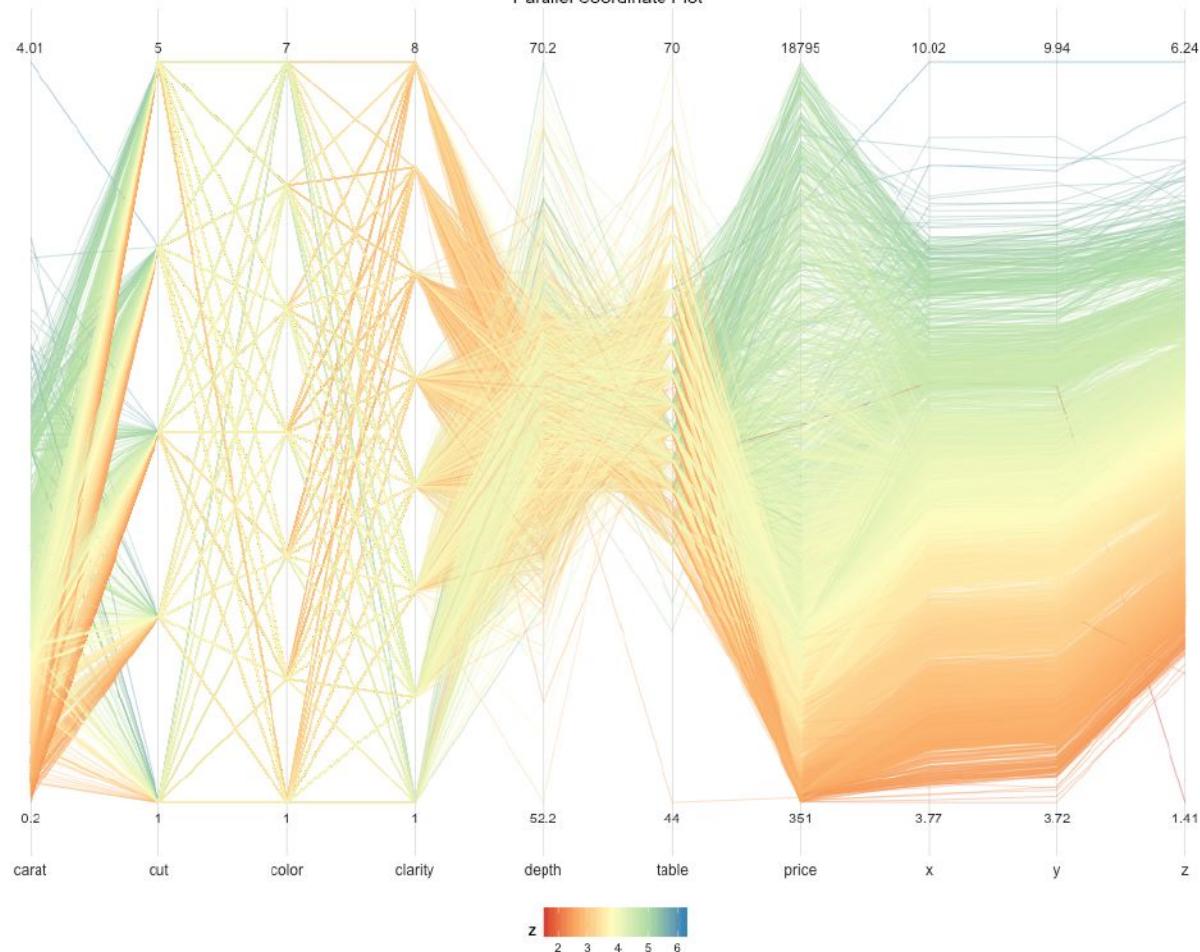


FIGURE 7-26 Standalone parallel coordinates plot on SAT scores

Parallel Coordinate Plot



Custom R and ggplot2 Implementation

Discussion

- Almost always requires interactivity
 - Choose column to color by
 - Choose how to sort
 - Highlighting (brushing)
- The order of the axes impacts how the reader understands the data.
 - a. Relationships between adjacent dimensions are easier to perceive than between non-adjacent dimensions.

Radar / Start Plot

Encode

Same as parallel coordinates but drawn radially

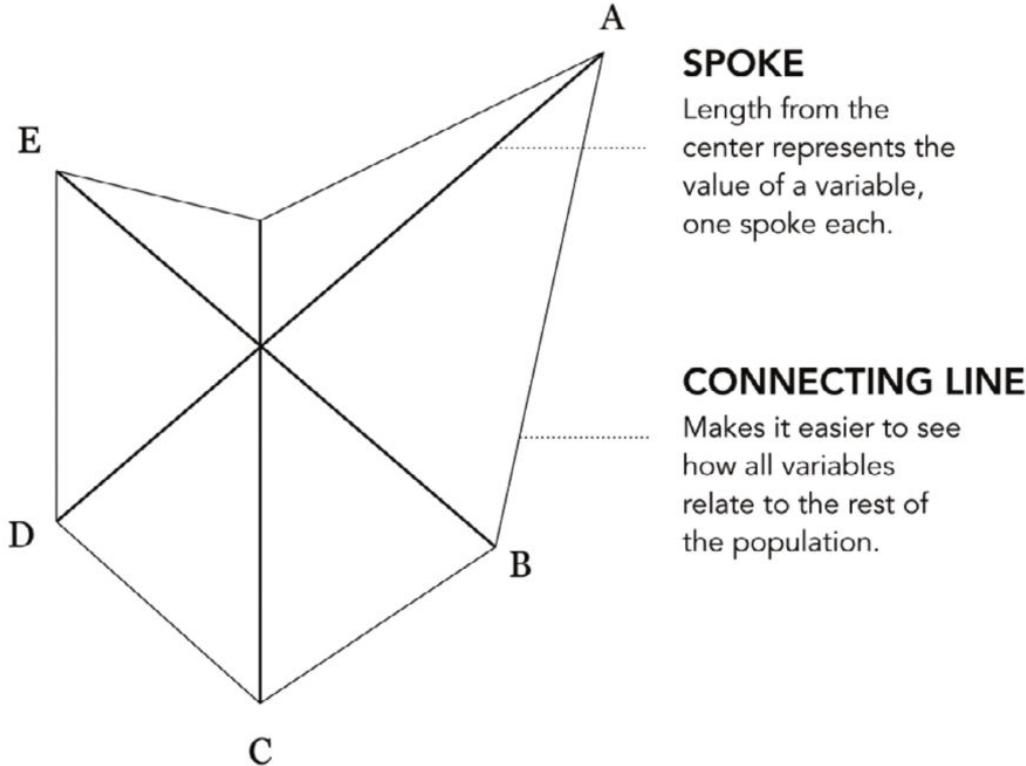


FIGURE 7-14 Star chart framework

"Visualize This: The FlowingData Guide to Design, Visualization, and Statistics" by Nathan Yau, 2011.

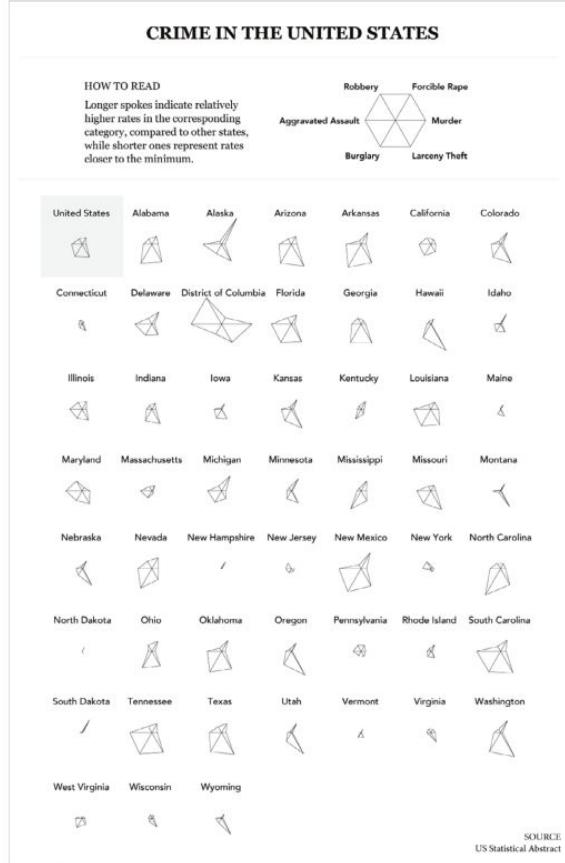
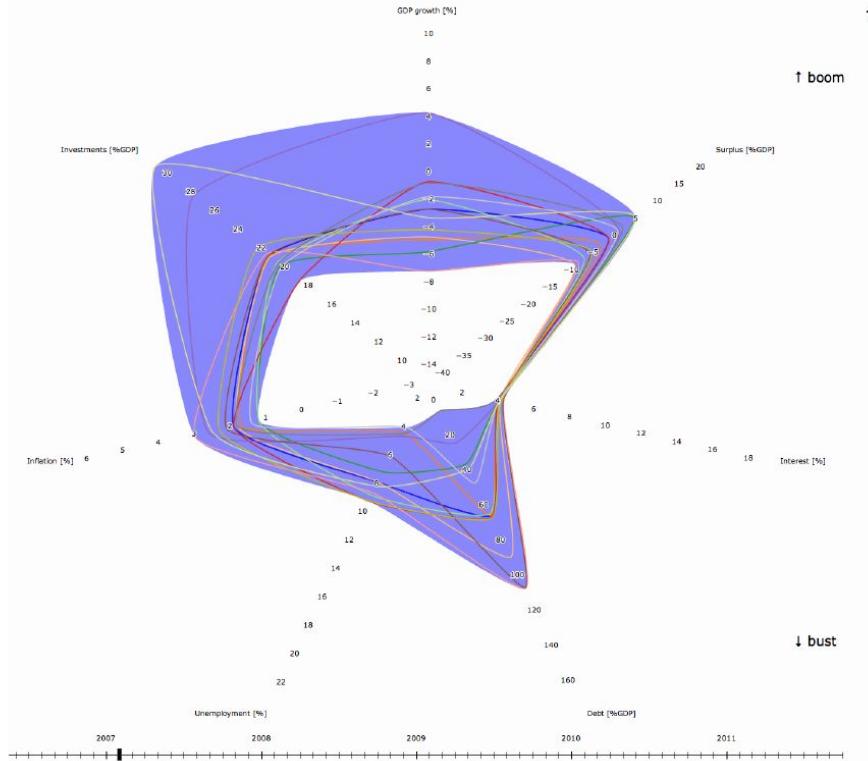


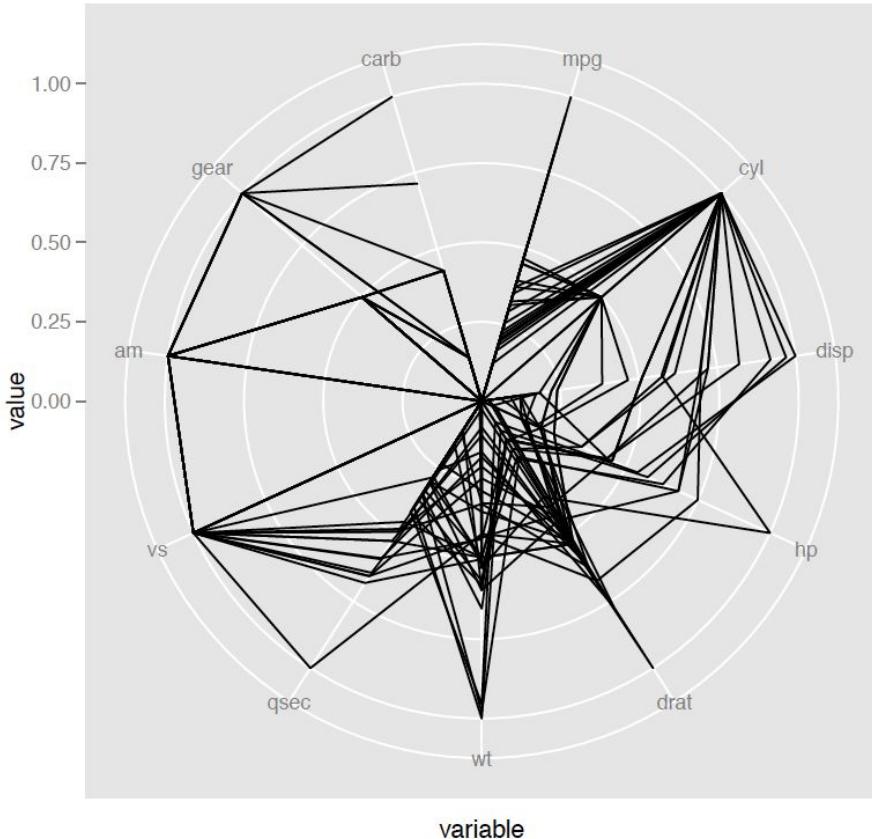
FIGURE 7-19 Series of star charts showing crime by state

"Visualize This: The FlowingData Guide to Design, Visualization, and Statistics" by Nathan Yau, 2011.

Demo:

<http://www.larsko.org/v/euc/>





<https://github.com/hadley/ggplot2/issues/516>

Discussion

- Gives a shape to the data
 - When plotted on top of each other, must have some ability to highlight and filter
- Works decently for small multiples?
 - Requires small number of rows
- Better for data with a circular aspect?
 - Monthly temperature time series

Time Series Visualization

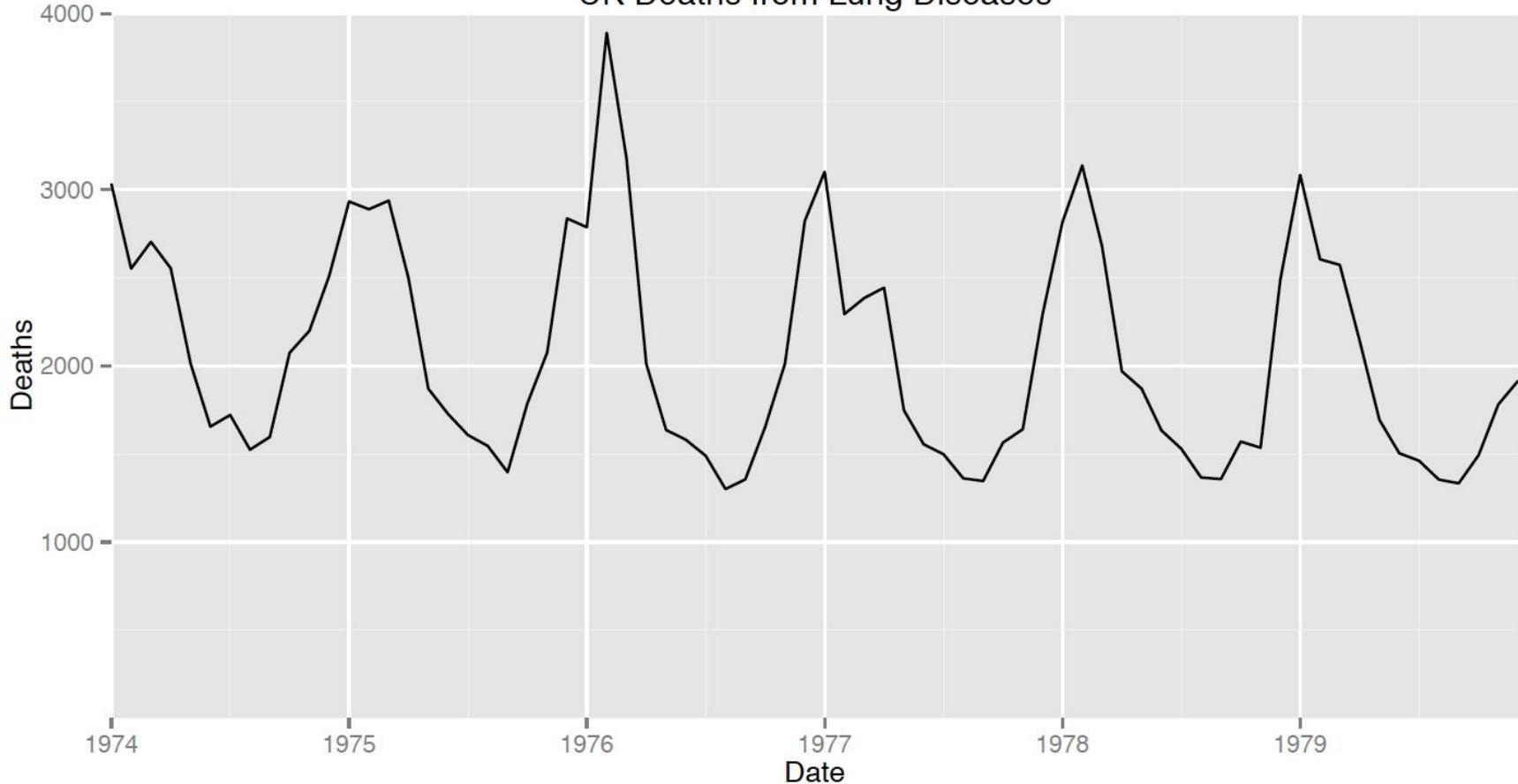
Time Series

- Series of values that change over time
- Values usually sampled at regular time intervals
- Examples include temperature, stock market, heart beat monitor, system logs, version history

Basic Techniques

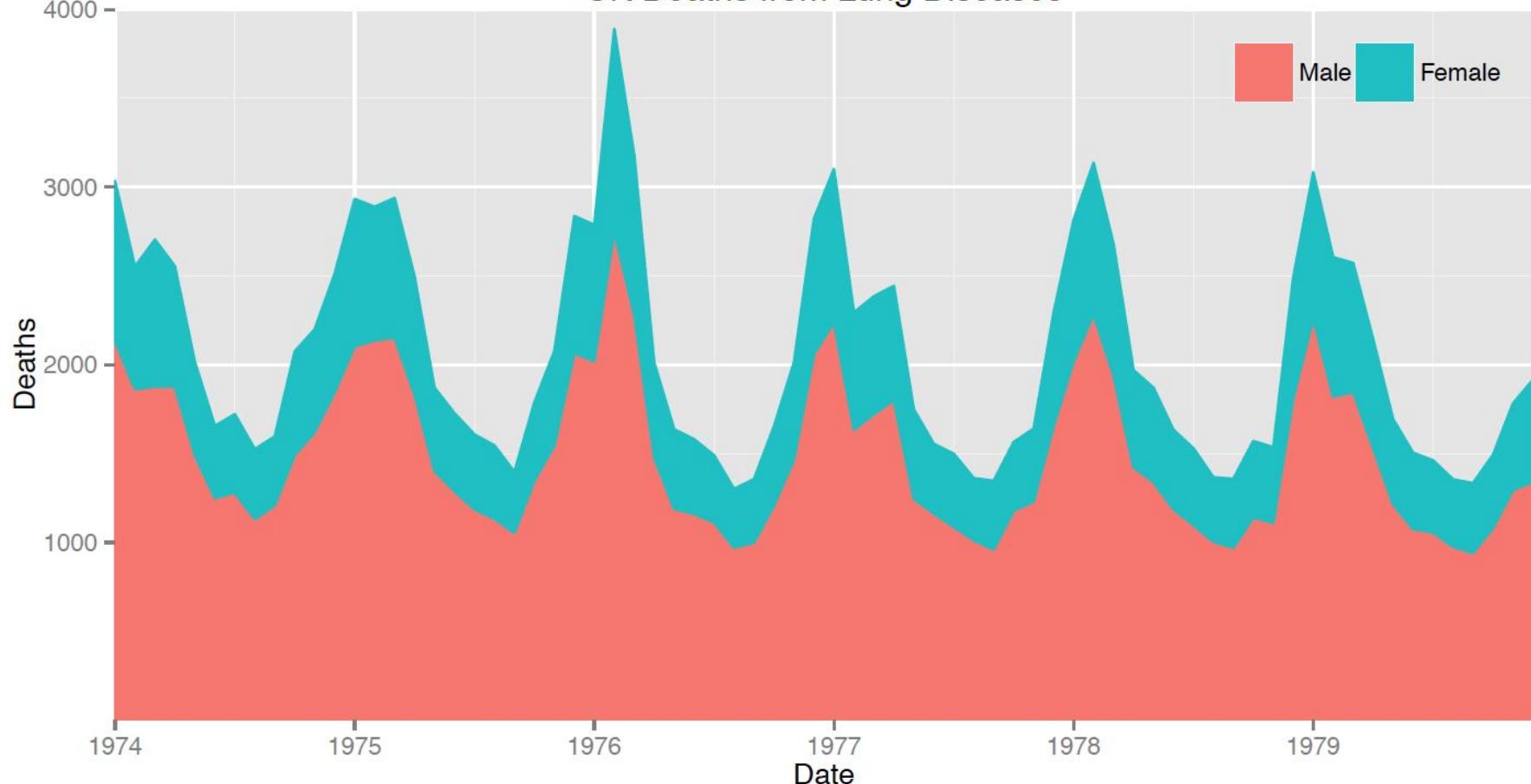
- Plot time on x-axis, value on y-axis
 - e.g. line, bar, area plots
- Plot two variables on x-axis and y-axis, show time via motion by animating changes over time
 - e.g. animated bubble plot, geospatial visualizations, linked views

UK Deaths from Lung Diseases



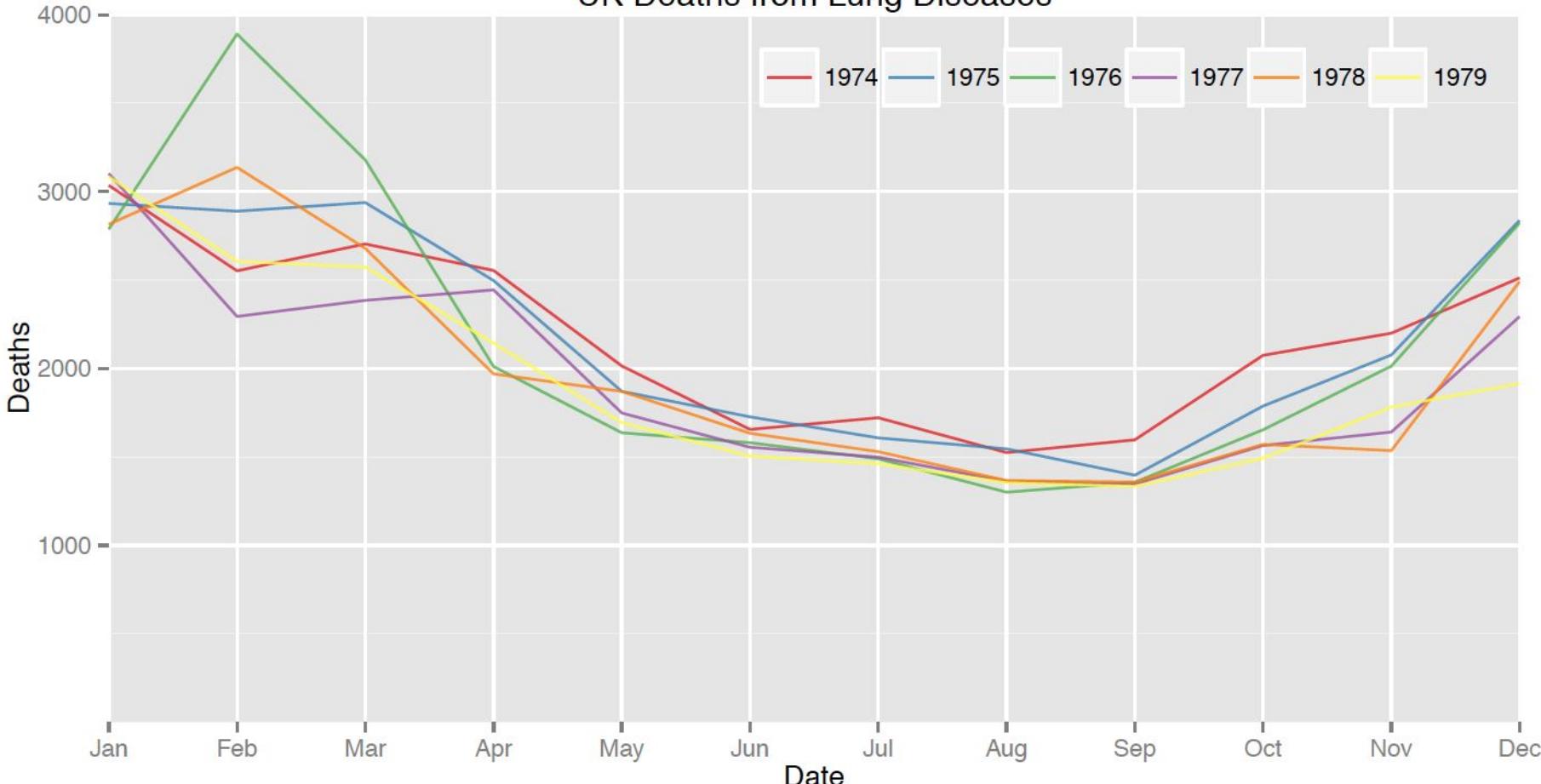
Line Plot

UK Deaths from Lung Diseases



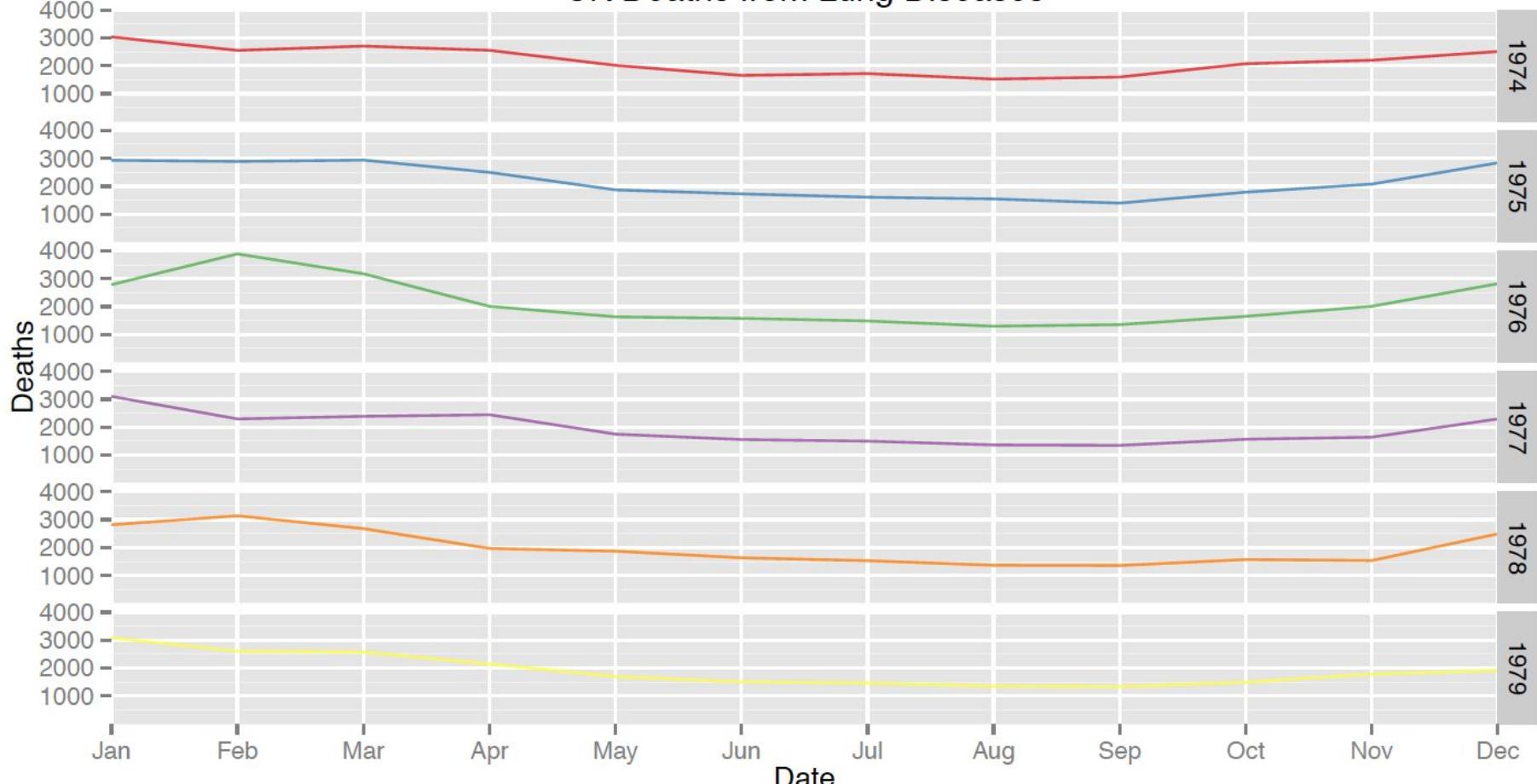
Stacked Area Plot

UK Deaths from Lung Diseases



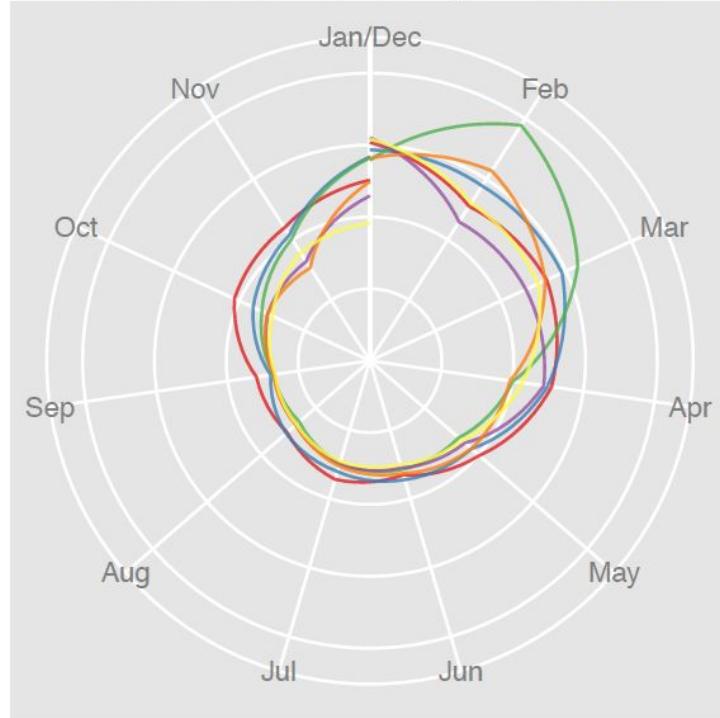
Multi-Line Plot

UK Deaths from Lung Diseases



Small Multiples Plot

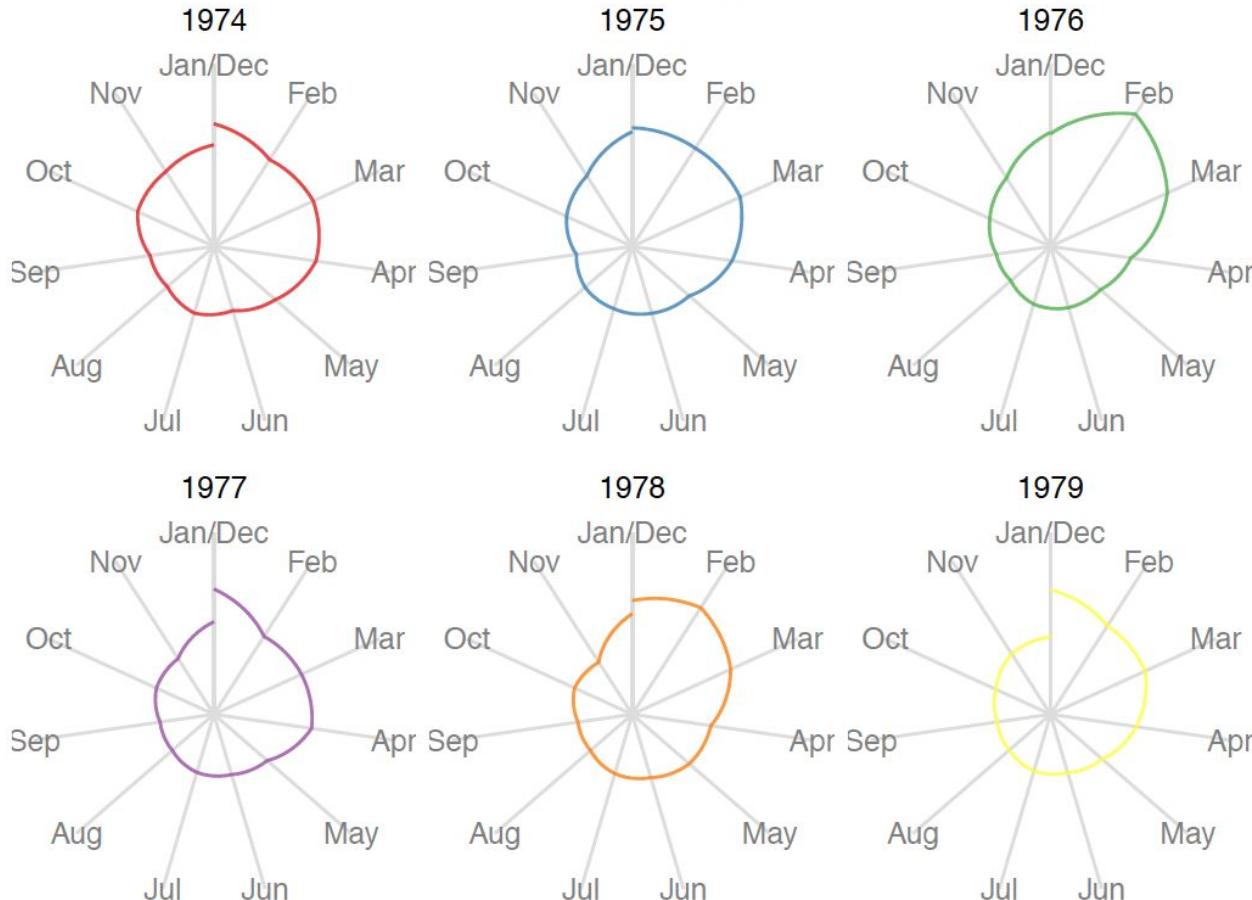
UK Deaths from Lung Diseases



— 1974 — 1975 — 1976 — 1977 — 1978 — 1979

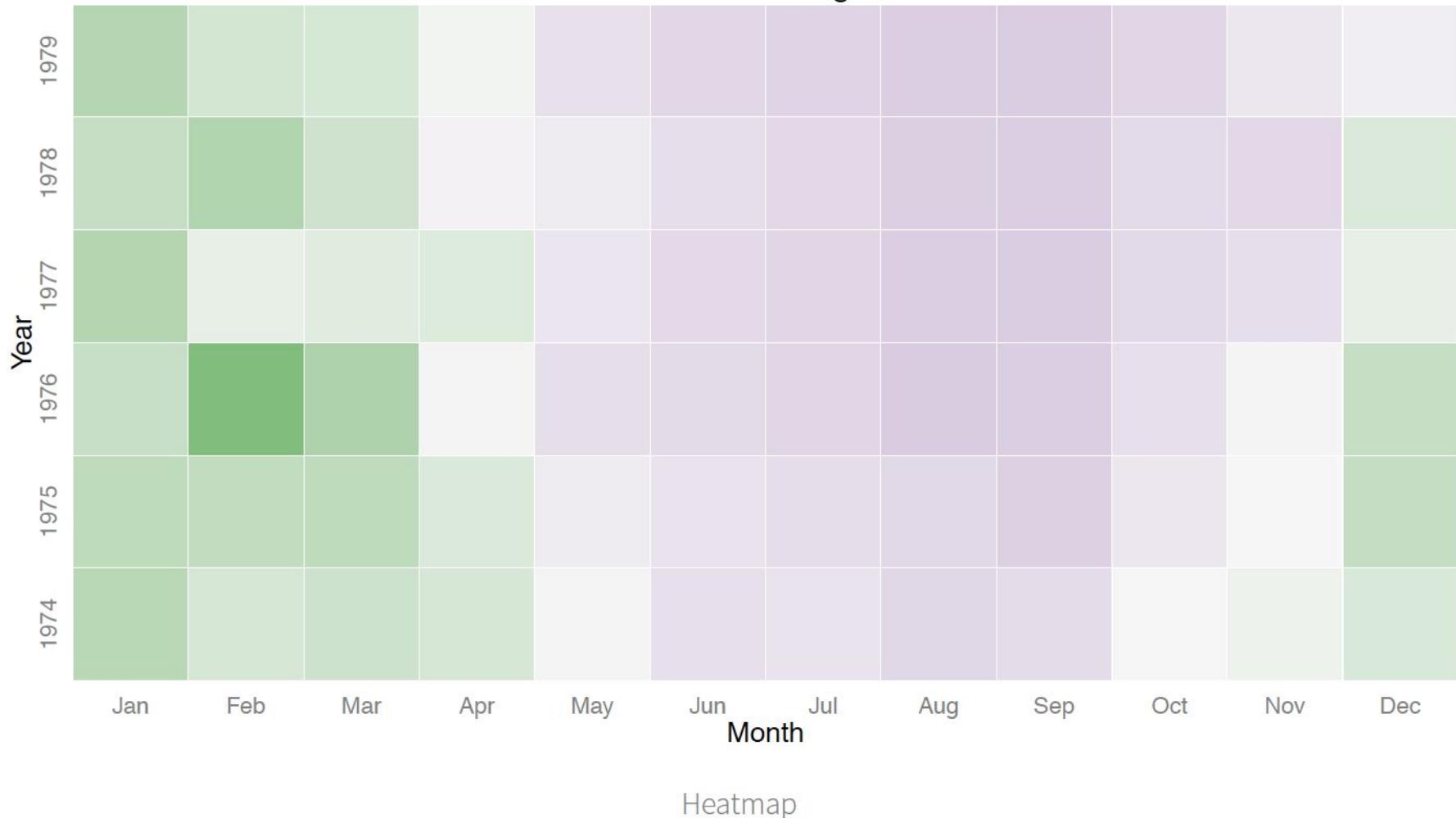
Star Plot*

UK Deaths from Lung Diseases

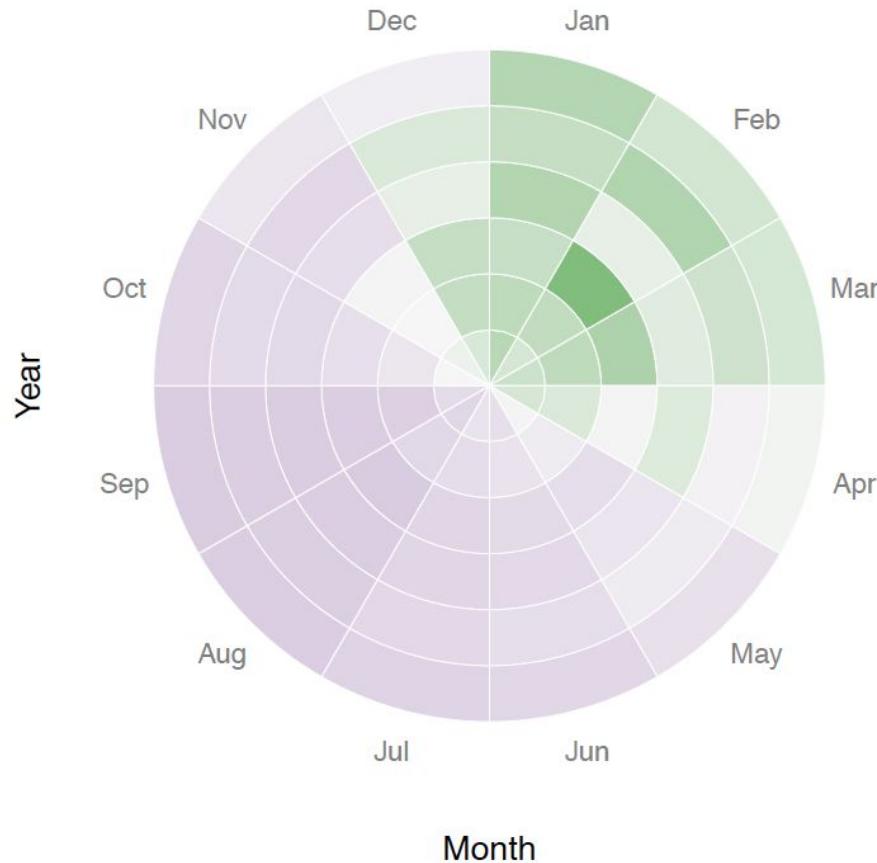


Small Multiple Star Plots*

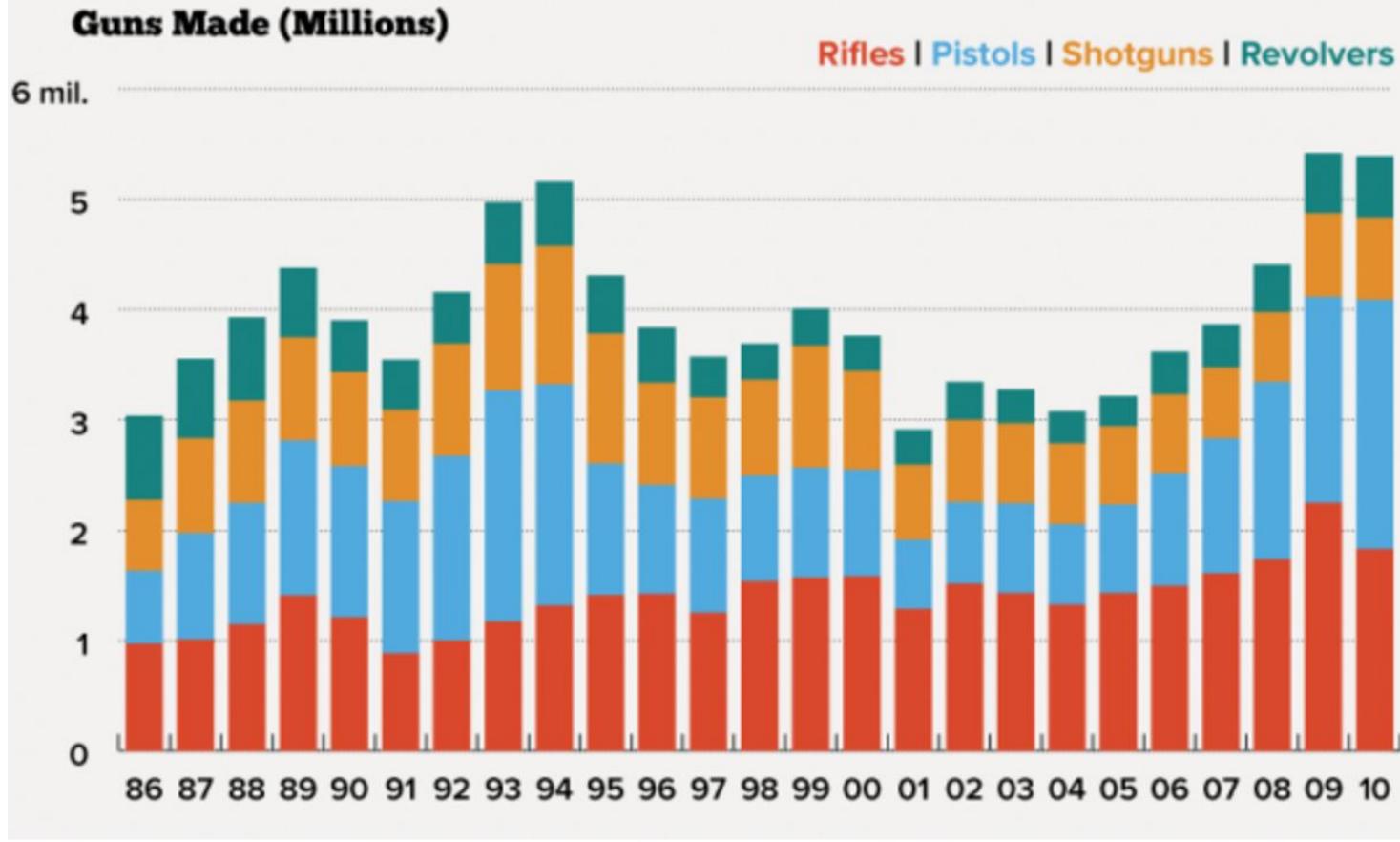
UK Deaths from Lung Diseases



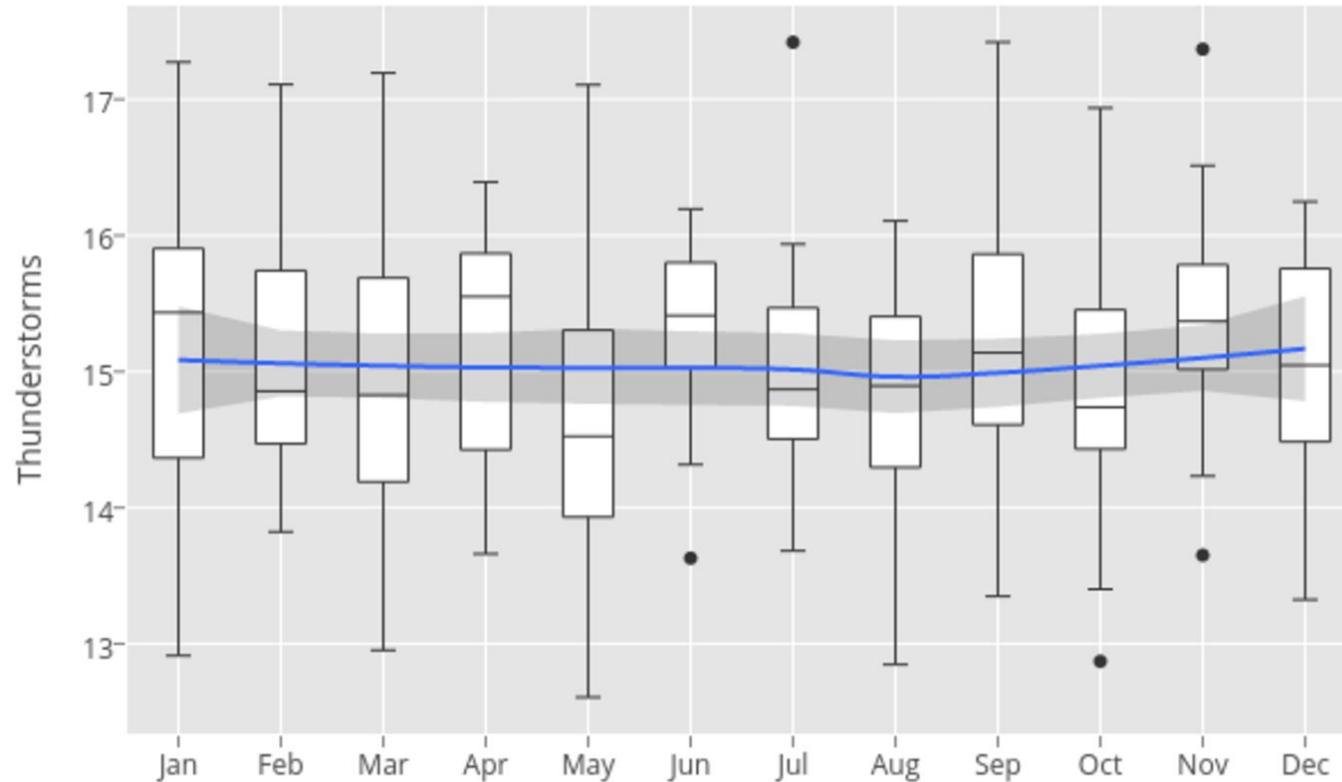
UK Deaths from Lung Diseases



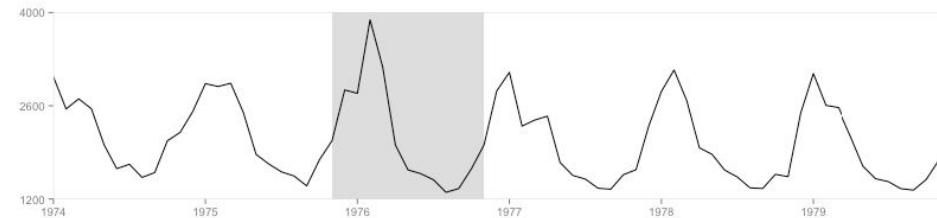
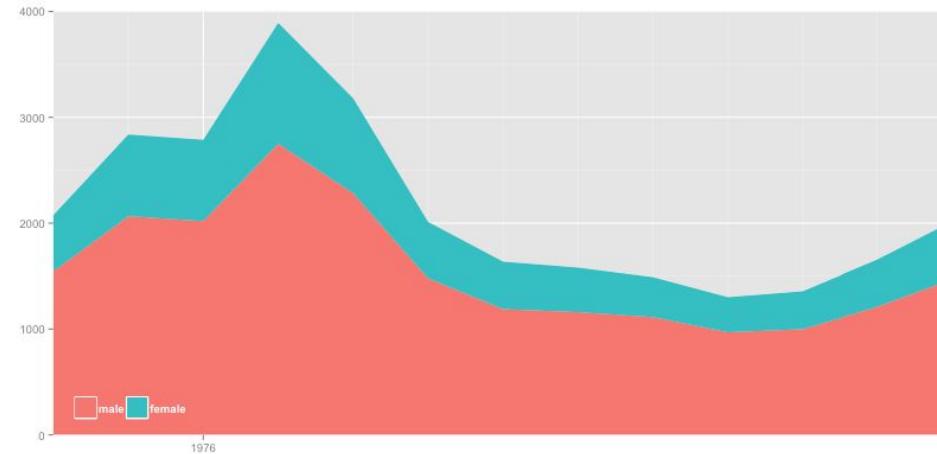
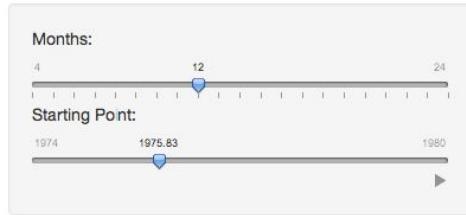
U.S. Gun Manufacturing By Type: 1986-2010



Box plot with Smoothed Fit



UK Deaths from Lung Diseases



Animated Linked Views

SHOWCASE

Non-Animated Visualizations



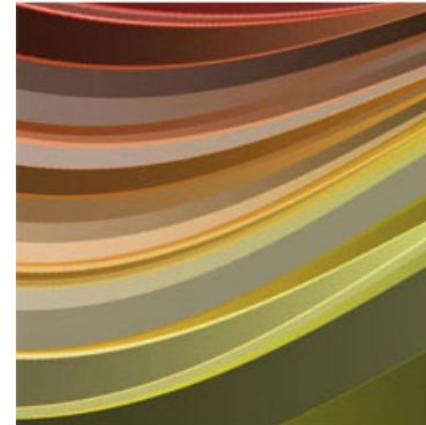
WINTER



SPRING

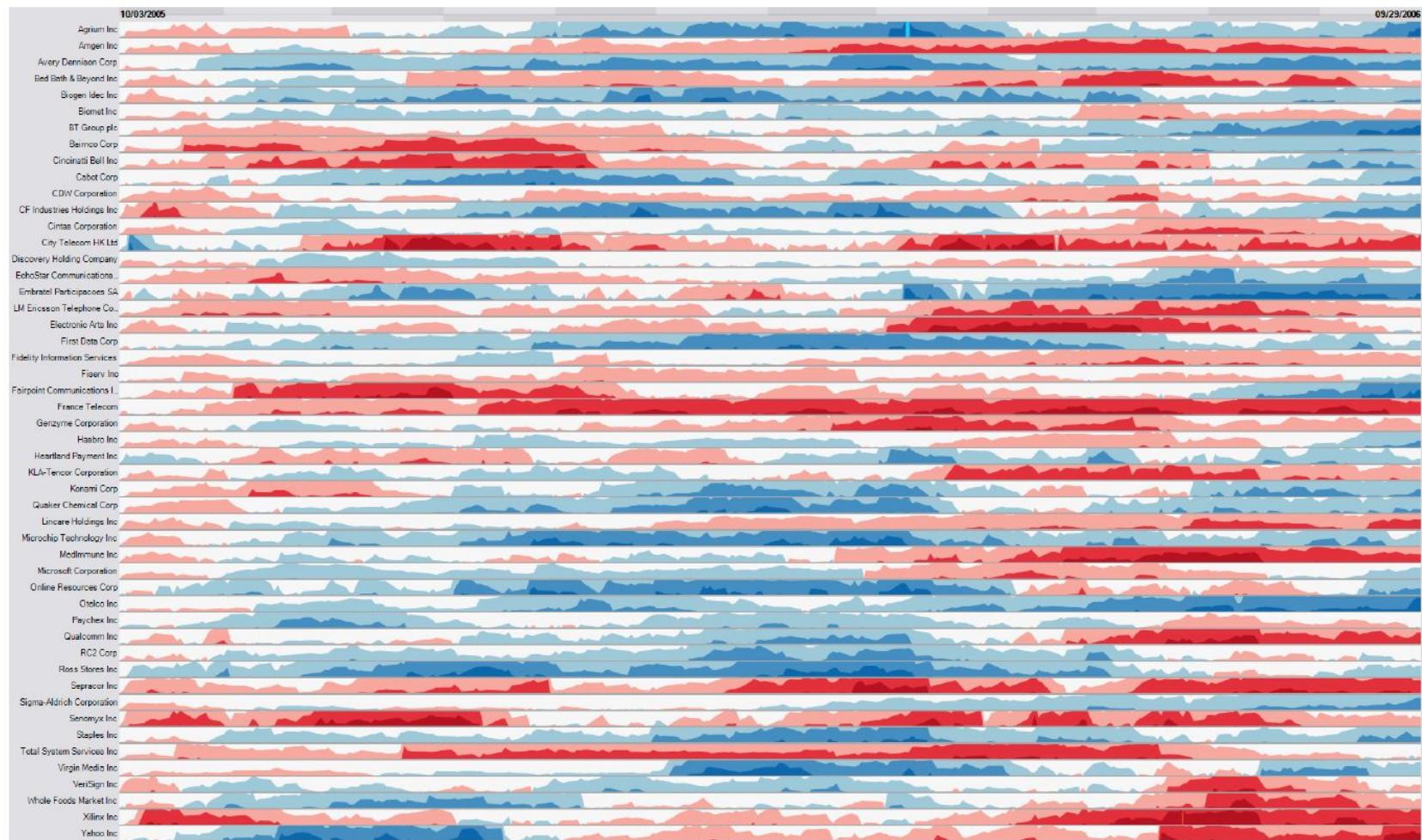


FALL



SUMMER





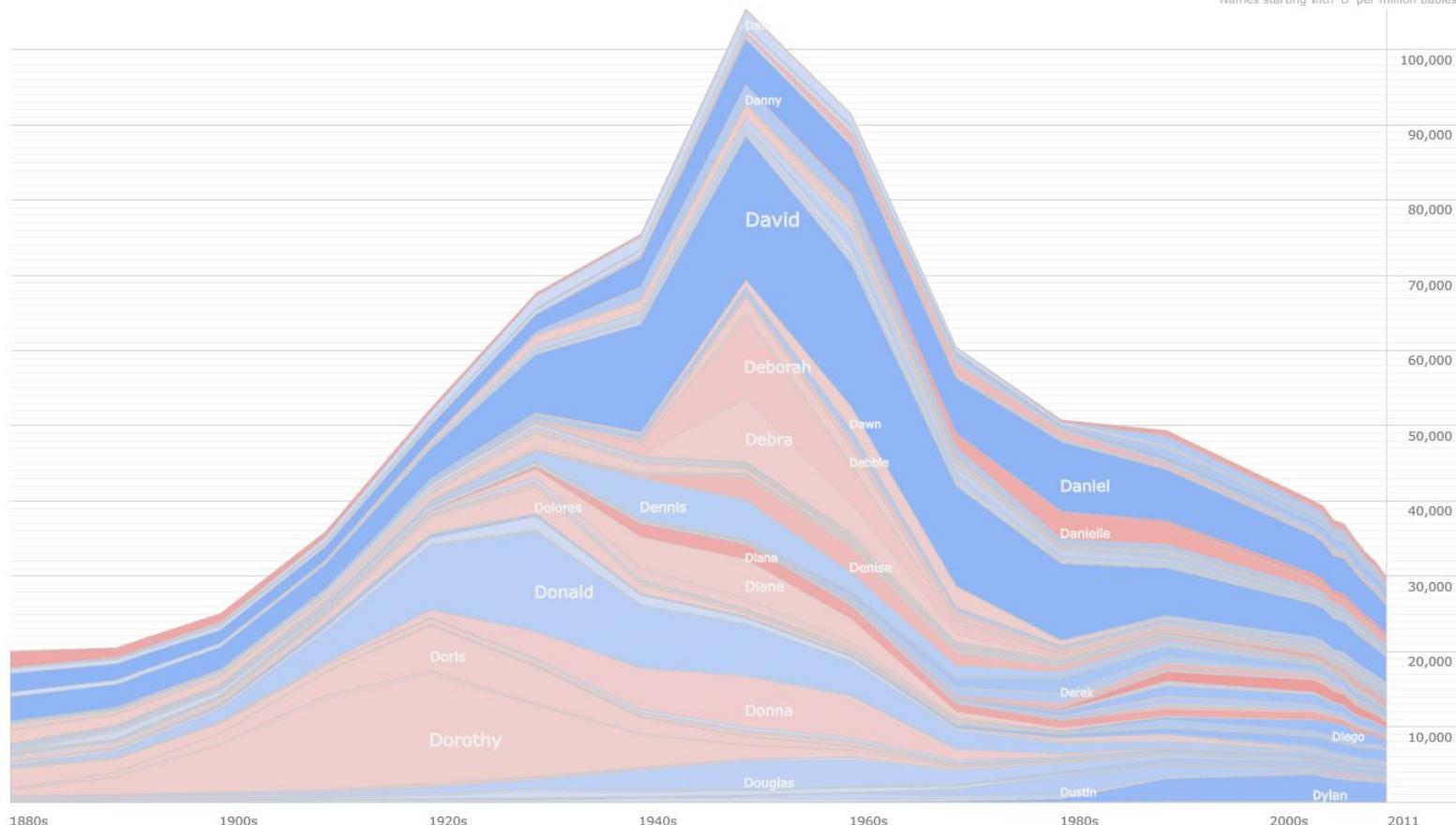
Baby Name > D X

Both Boys Girls

2010 rank: boys 1000 500 100 25 1

girls 1000 500 100 25 1

Names starting with 'D' per million babies



1880s

1900s

1920s

1940s

1960s

1980s

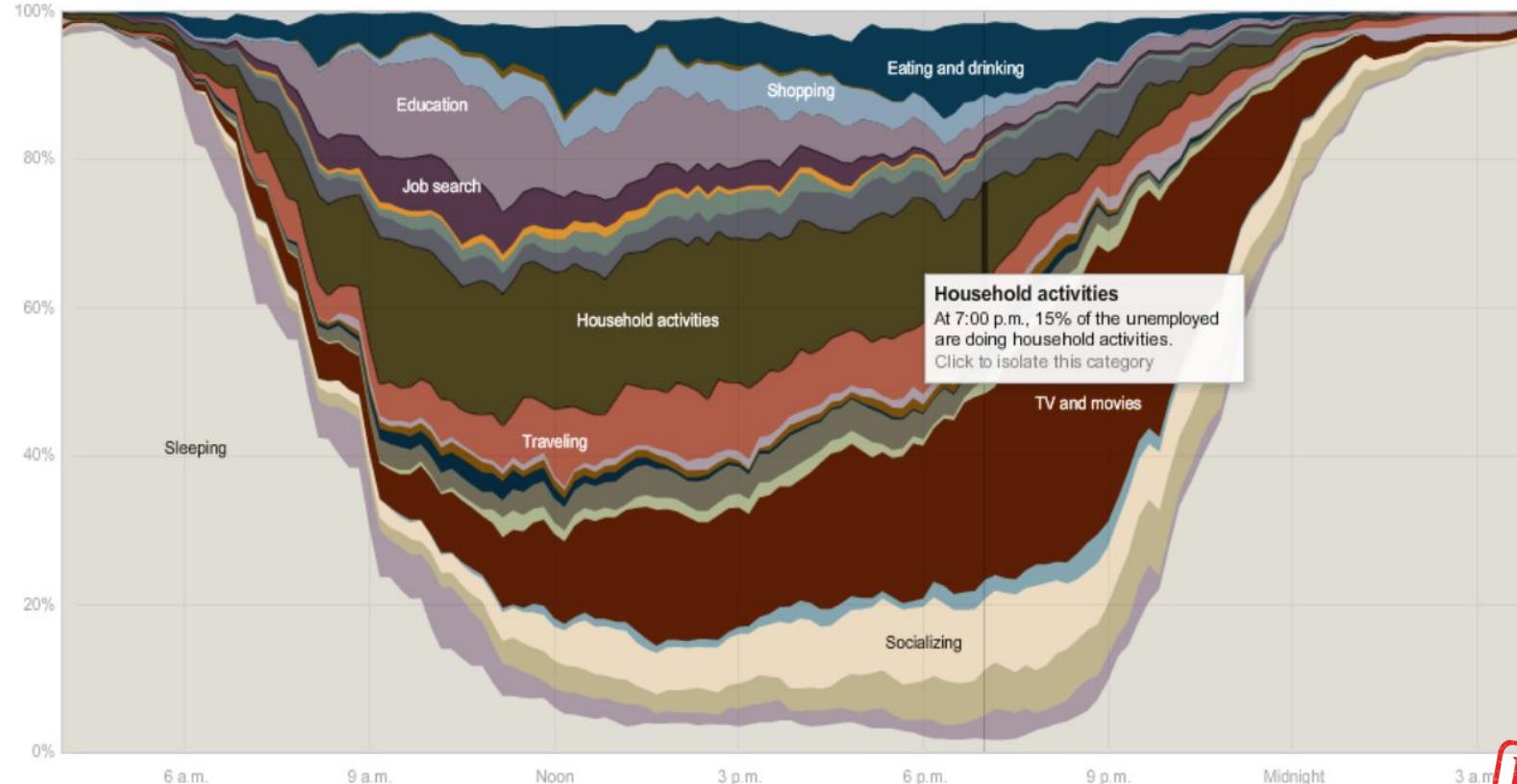
2000s

2011

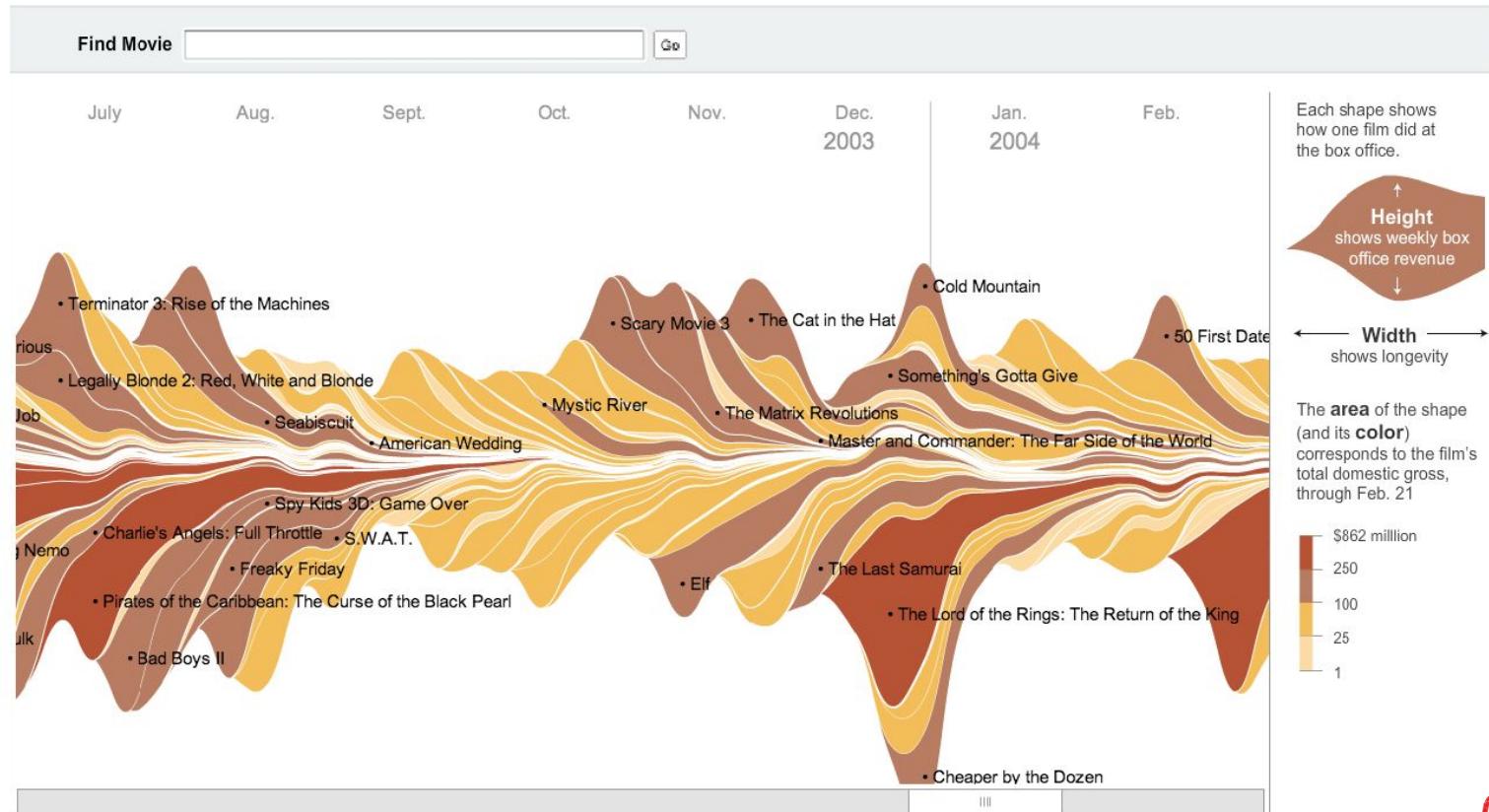
The unemployed

On average, the unemployed spend about a half-hour looking for work. They tidy the house, do laundry and yard work for more than two hours, about an hour more than the employed.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children



DEMO



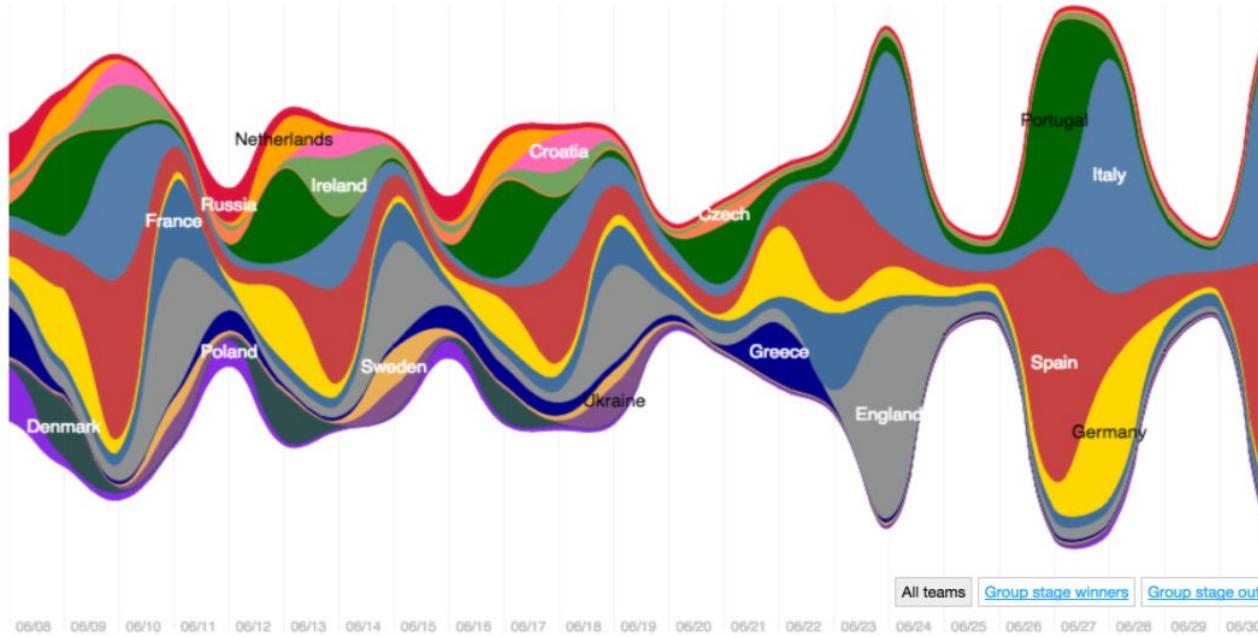
DEMO



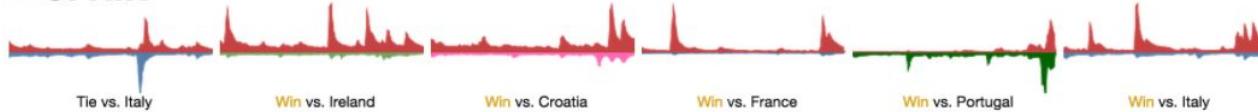
#Euro2012

A summary for the action on Twitter during the European football tournament. [Tweet](#) 1,810

The Streamgraph below shows volume of Tweets during the #Euro2012 period. Click on a team's name to see details.

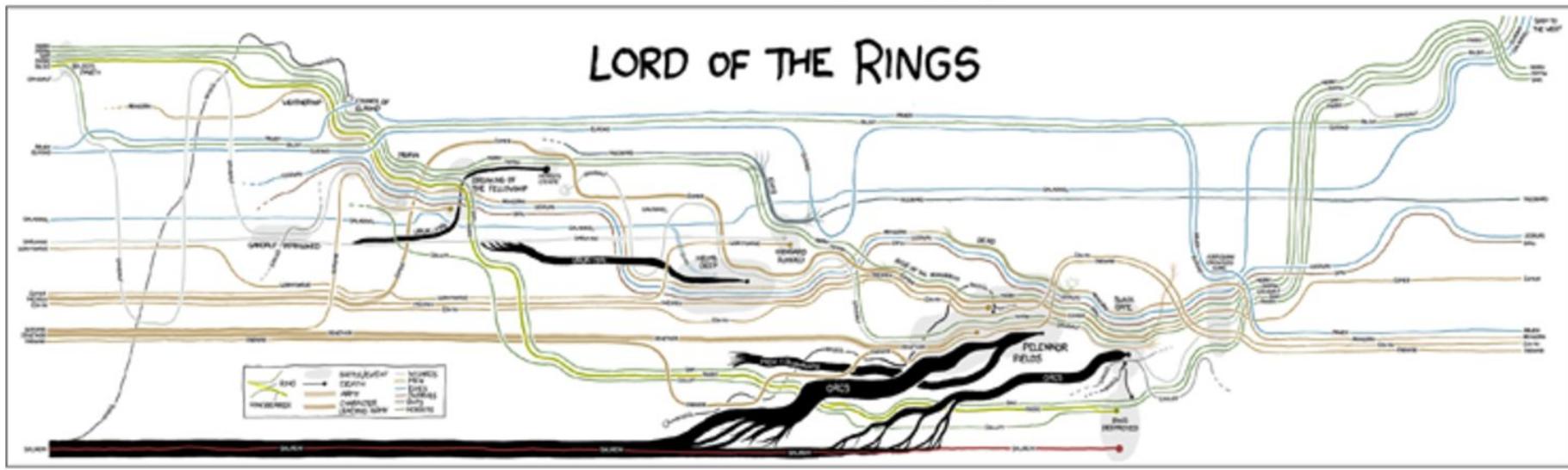


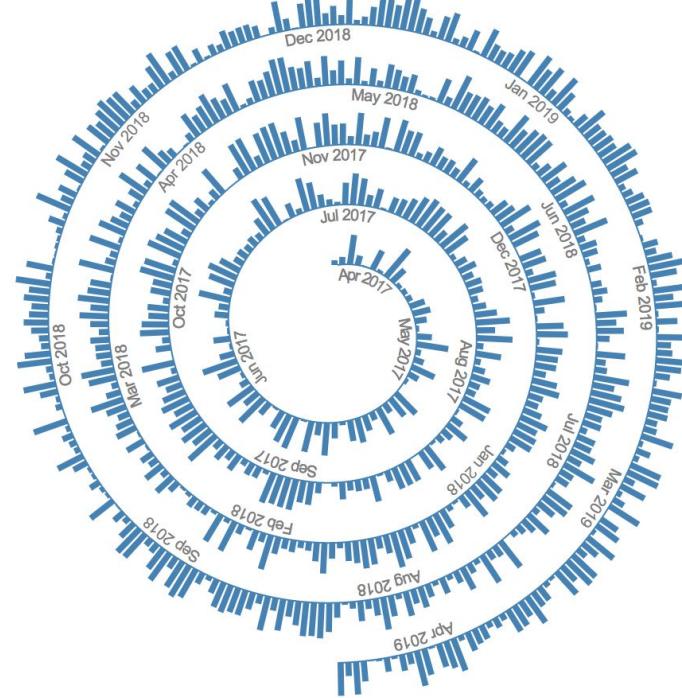
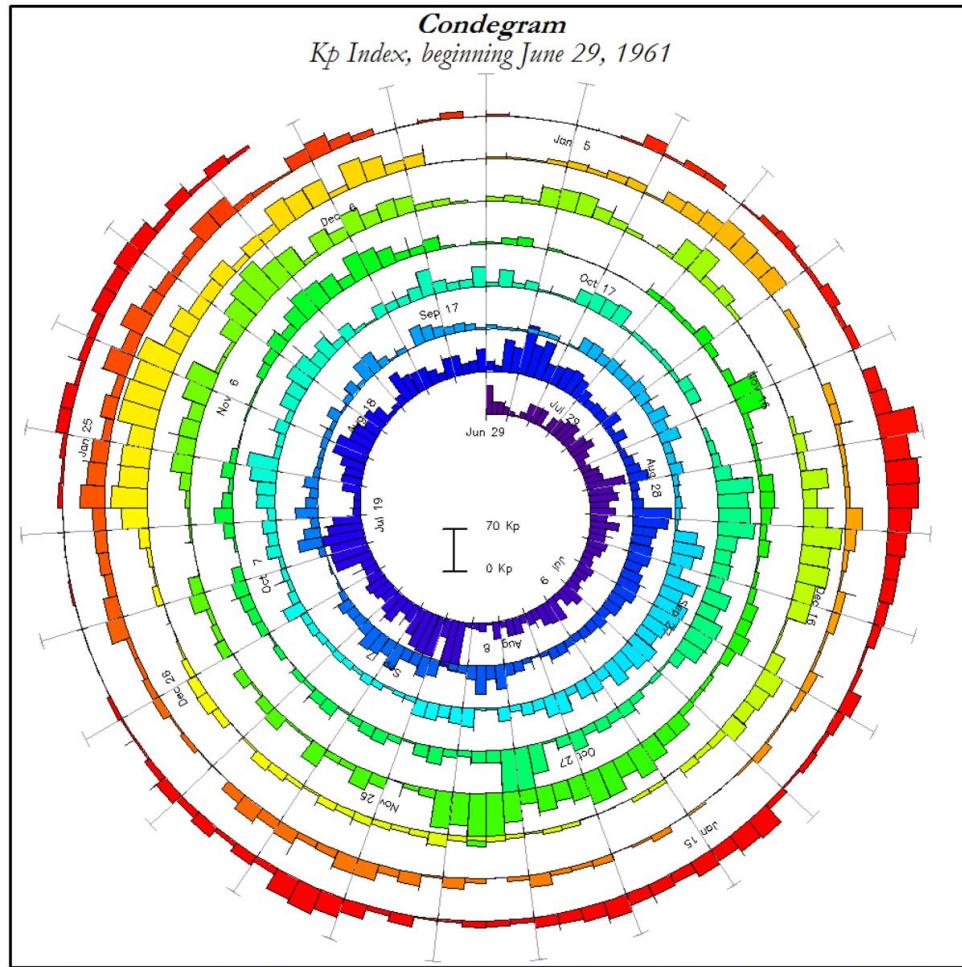
■ SPAIN



THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.

LORD OF THE RINGS





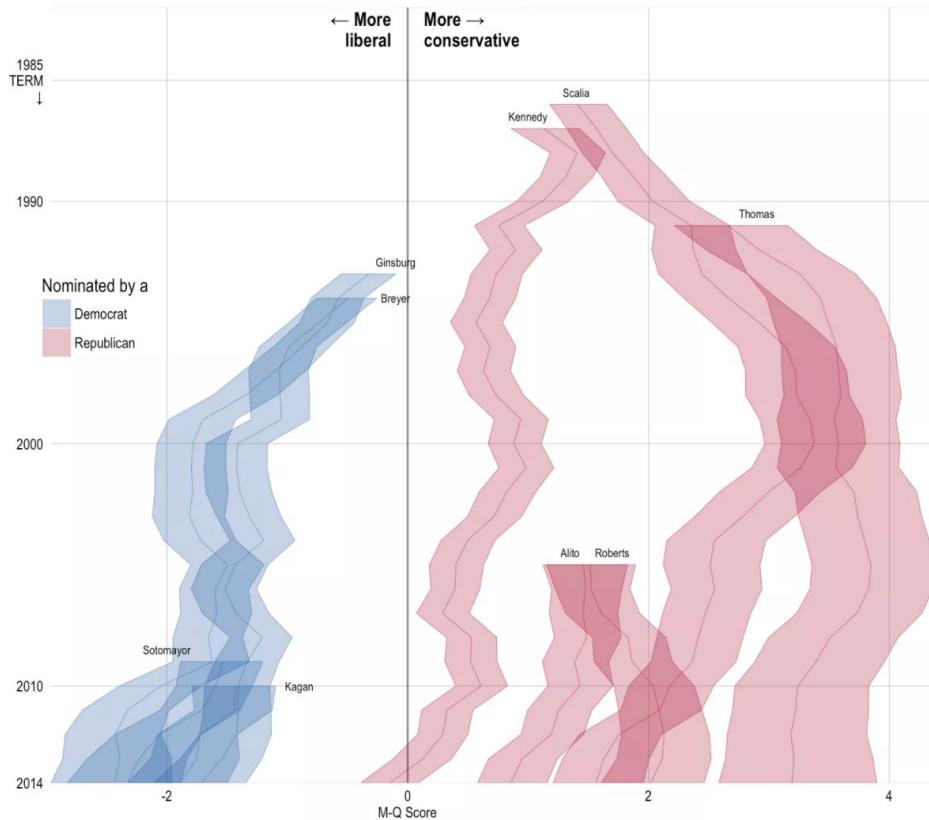
https://en.wikipedia.org/wiki/File:Condegram_Spiral_Plot.png,
<http://blocks.org/larsenmtl/222043d93a41d48b58d2bfa1e3d4f708>

Making time series rivers

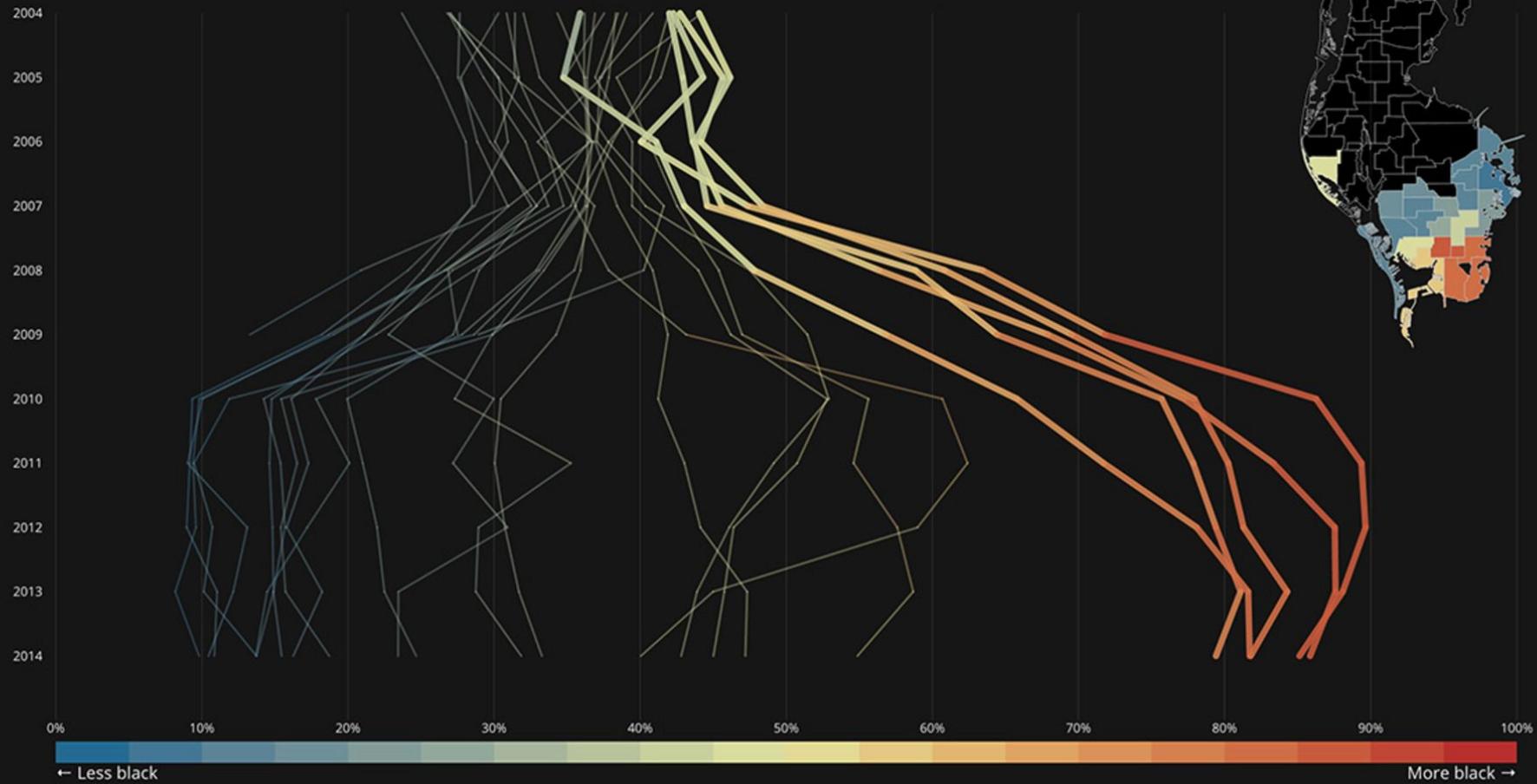


Martin-Quinn scores for selected justices, 1985-2014

Ribbon band derived from mean plus one standard deviation. Inner line is the M-Q median.



Today, **Campbell Park, Fairmount Park, Lakewood, Maximo and Melrose** are the **most segregated** schools in Pinellas.



04 April 2015

01 October 2015

May 31 Jun 07 Jun 14 Jun 21 Jun 28 Jul 05 Jul 12 Jul 19 Jul 26 Aug 02 Aug 09 Aug 16 Aug 23 Aug 30 Sep 06

Lorem (14)



Ipsum (64)



Dolor (82)



Sit (82)



Amet (46)



Consectetur (91)



Adipisicing (6)



elit (14)



Eiusmod tempor (102)



Incididunt (31)



May 31 Jun 07 Jun 14 Jun 21 Jun 28 Jul 05 Jul 12 Jul 19 Jul 26 Aug 02 Aug 09 Aug 16 Aug 23 Aug 30 Sep 06

SHOWCASE

Animated Visualizations

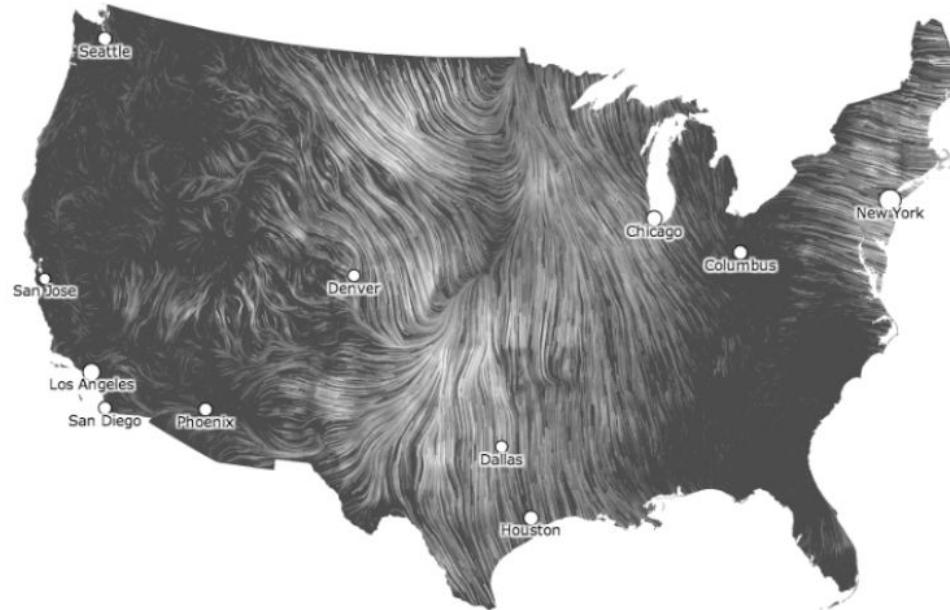
wind map

February 17, 2013

11:59 pm EST

(time of forecast download)

top speed: 39.4 mph
average: 8.8 mph



Gallery



An invisible, ancient source of energy surrounds us—energy that powered the first explorations of the world, and that may be a key to the future. This map shows you the delicate tracery of wind flowing over the US.

[Donate to Hurricane Sandy relief efforts here.](#)

NEW: wind map prints are available from [Point.B Studio](#).

<http://windmap.info>

DEMO

