USF UNIVERSITY OF SAN FRANCISCO

CHANGE THE WORLD FROM HERE

Multivariate Data Visualization

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science

Multivariate Data

- Multiple variables (columns) of data
  - 2 variables (x, y) is bivariate
  - 3 variables (x, y, z) is trivariate…
- Mostly concerned with 4 or more variables
- Variables can be of any type (or combination of types)

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

Data Types

- **Numerical**
  - Continuous (-11.2, -1.3, 0.4, 4.8, 14.9, …)
  - Discrete (-2, -1, 0, 1, 2, …)
- **Categorical**
  - Ordered (January, February, March, April, …)
  - Unordered (Red, Blue, Green, Purple, …)

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

Data Types

- **Special/Structured**
  - Time (1998-01-27 12:00, 1999-04-13 22:10, …)
  - GPS (37.77679 latitude, -122.45117 longitude, …)
  - Social Security Numbers (555-55-5555, …)
- **Unstructured/Semi-Structured**
  - Free-Form Text (Twitter Feed, Screenplay, …)

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

Outline

- Bubble Chart
- Heatmap
- Scatterplot Matrix
- Small Multiples
- Parallel Coordinates
- Other Techniques

- *Encoding Data*
- *Examples*
- *Discussion*

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

BUBBLE CHART

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

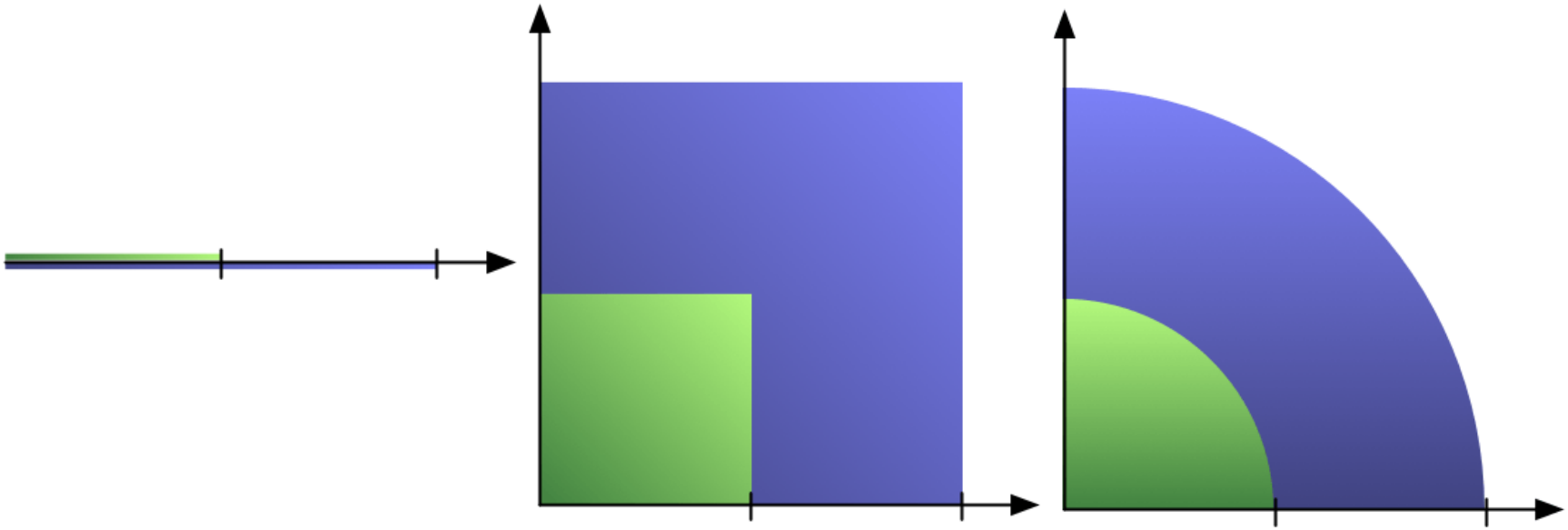
Encoding Data

- **Horizontal position**
  – Continuous data
- **Vertical position**
  – Continuous data
- **Circle area** *(not radius)*
  – Numerical data
- **Circle color**
  – Numerical or categorical

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

Linear vs Quadratic Change



https://eagereyes.org/blog/2008/linear-vs-quadratic-change

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

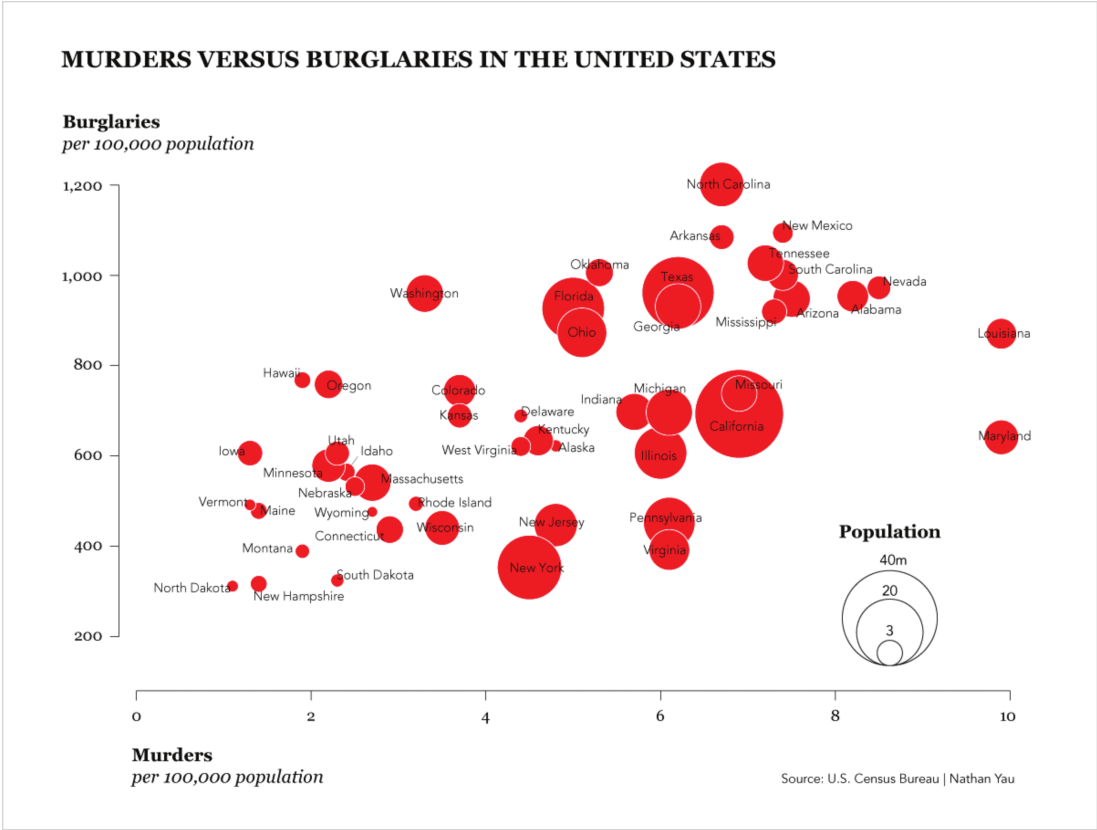
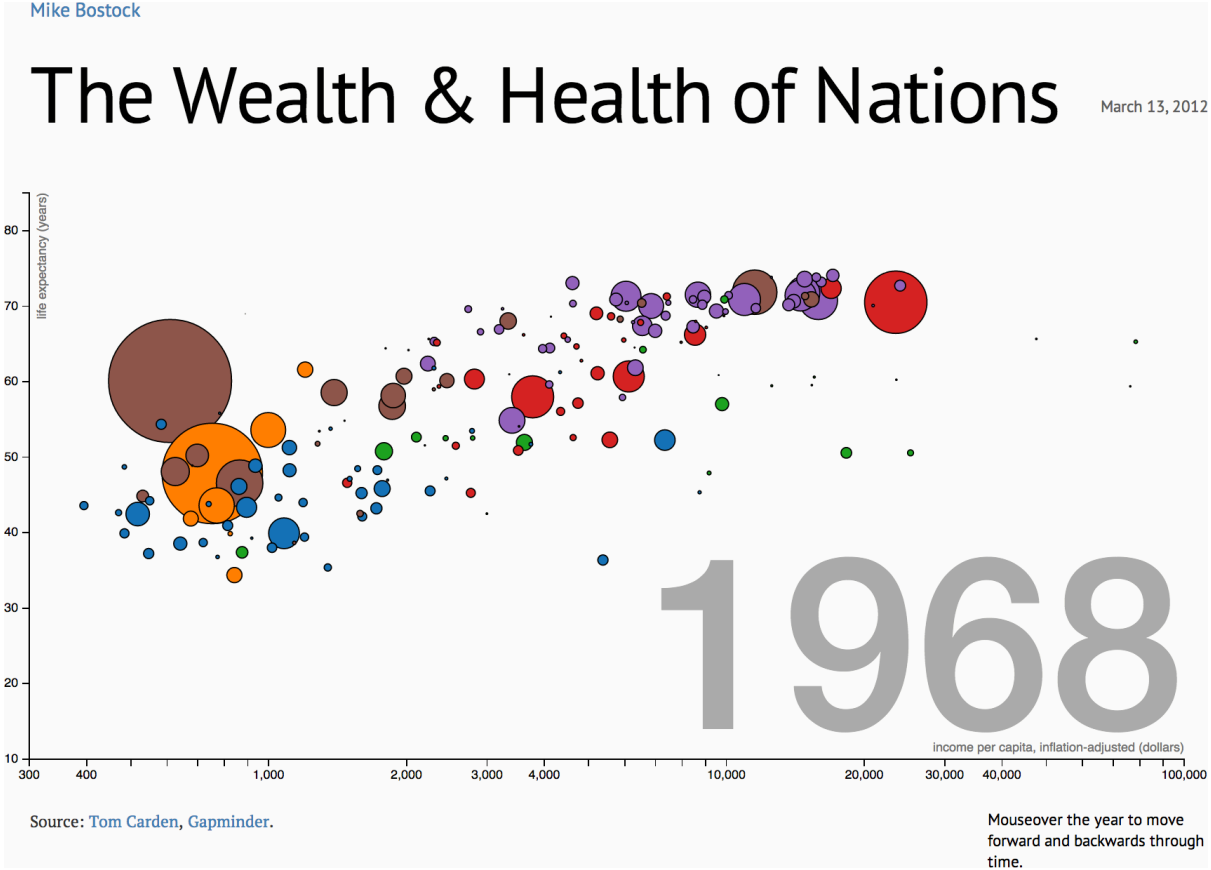


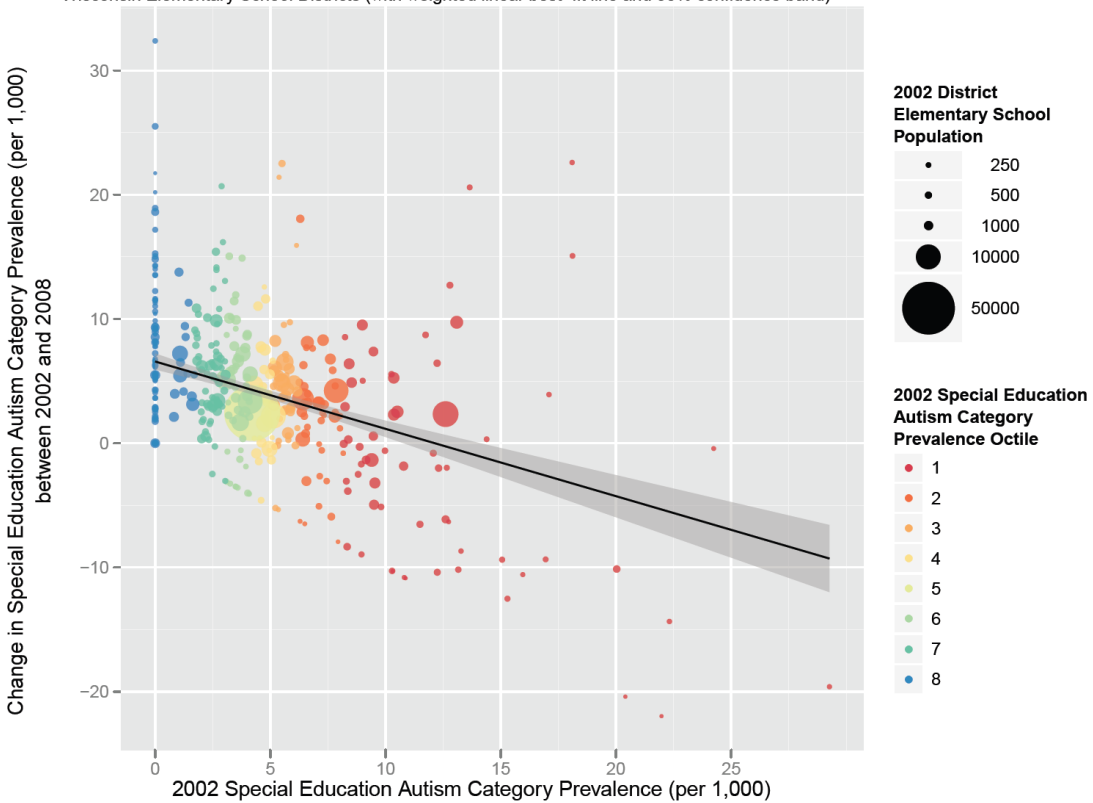
FIGURE 6-15 Bubble plot showing crime in the United States

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



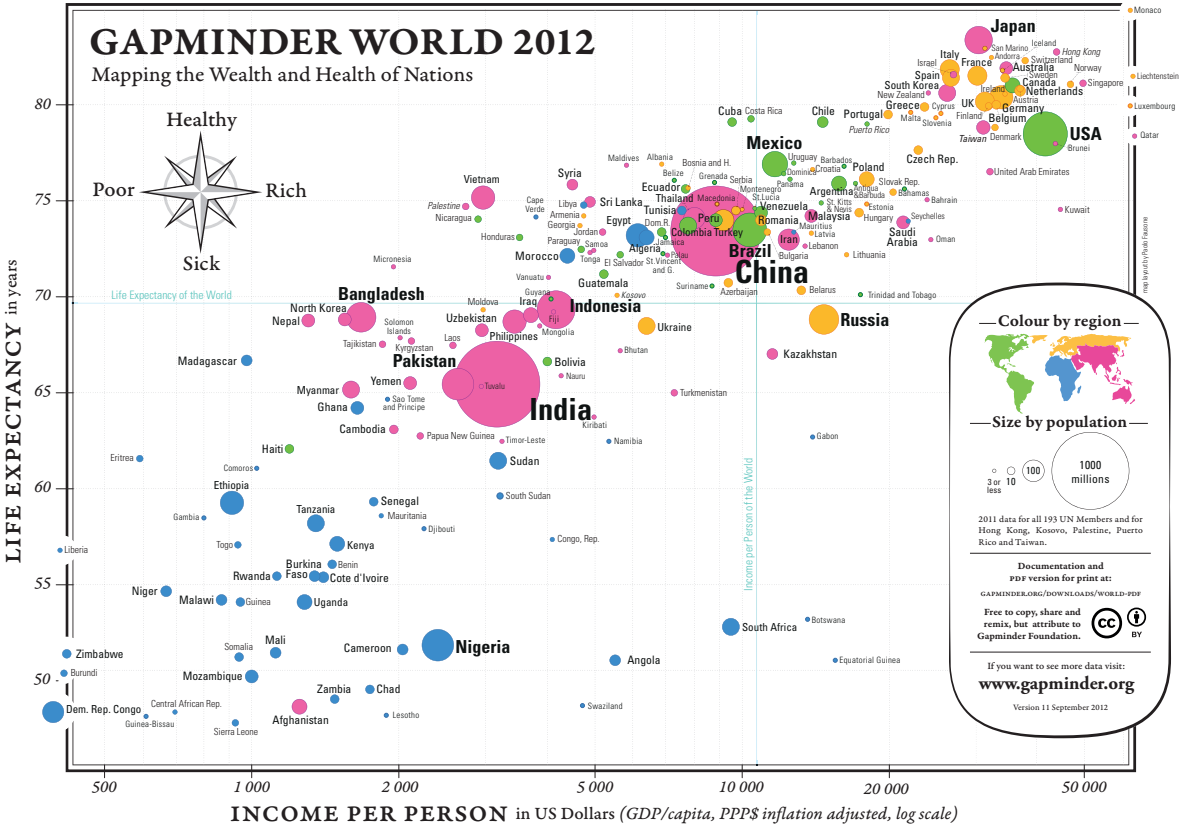
http://bost.ocks.org/mike/nations/

Figure 3. Change in Special Education Autism Category Prevalence between 2002 and 2008 vs Baseline (2002) Prevalence, Wisconsin Elementary School Districts (with weighted linear best−fit line and 95% confidence band)



http://www.matthewmaenner.com/blog/?p=150

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://www.gapminder.org/downloads/world-pdf/

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

Discussion

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

HEATMAP

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

Encode Data

- **Horizontal and vertical position**
  - Row/column position in data table
  - Discrete or categorical variable
- **Box color**
  - Value from row, column in dataset
  - May be numerical or categorical

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

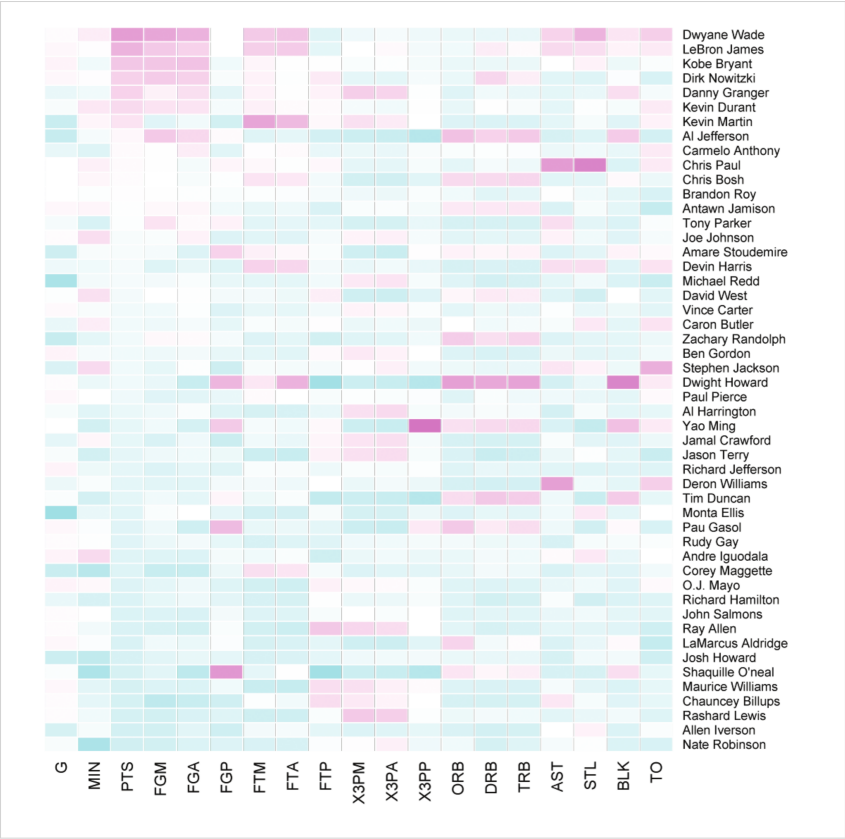
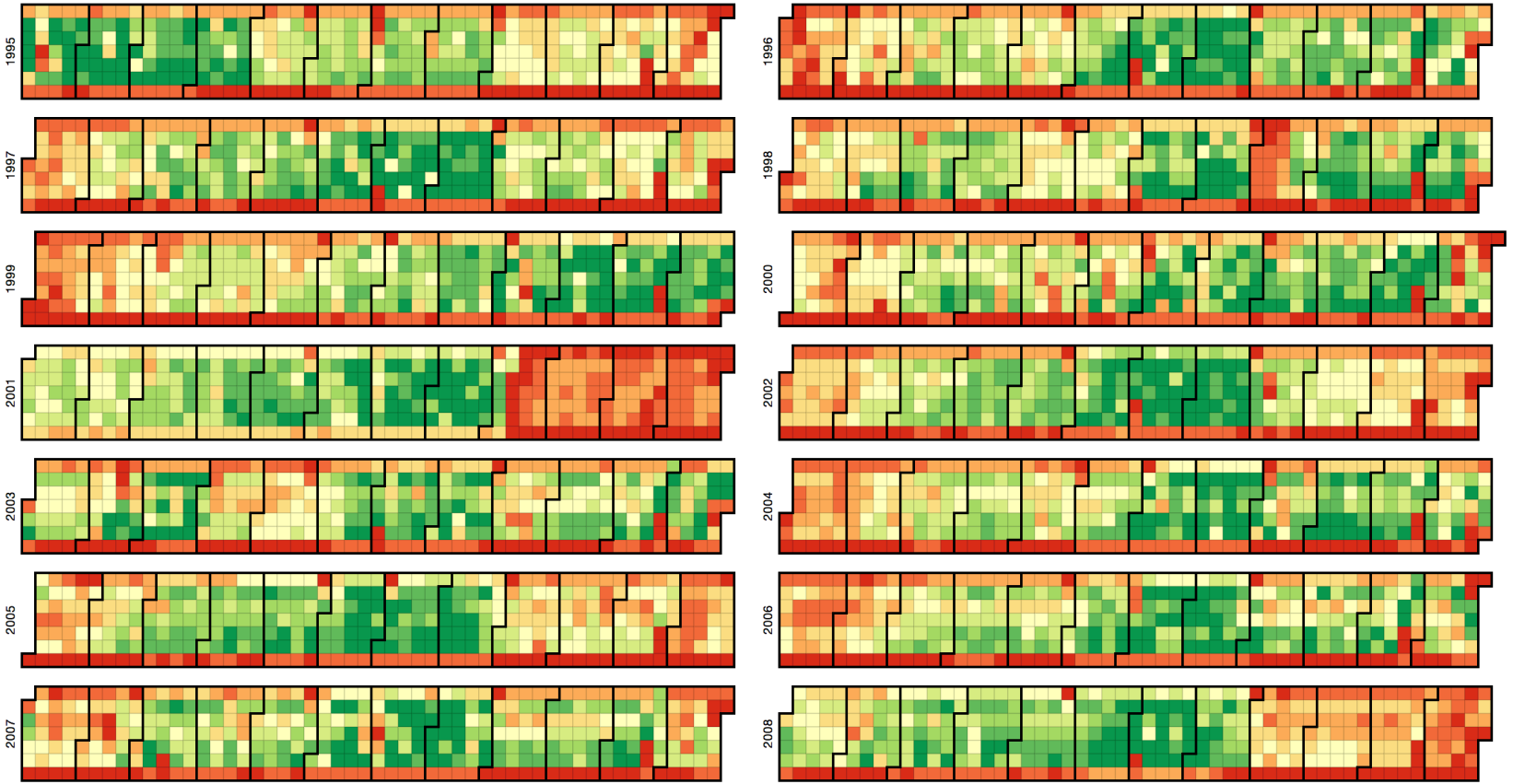


FIGURE 7-3 Default heatmap ordered by points per game

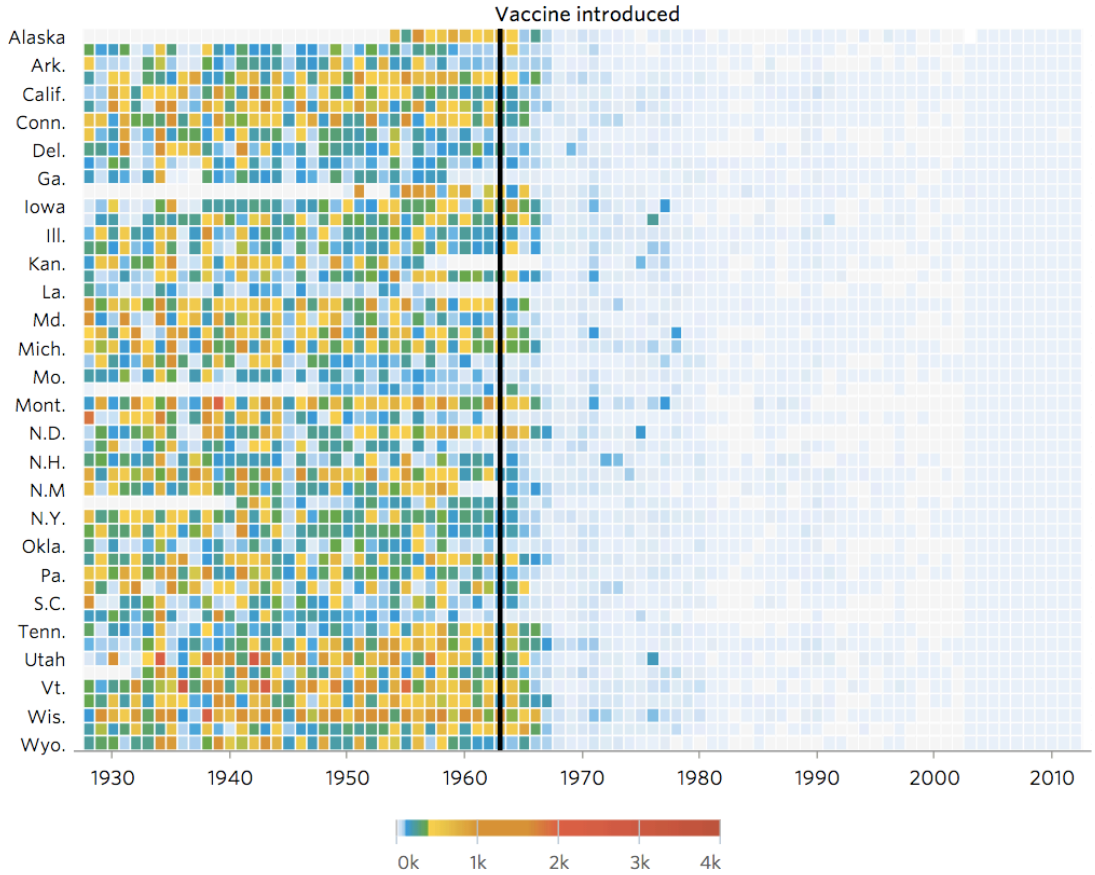
"Visualize This: The FlowingData Guide to Design, Visualization, and Statistics" by Nathan Yau, 2011.

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



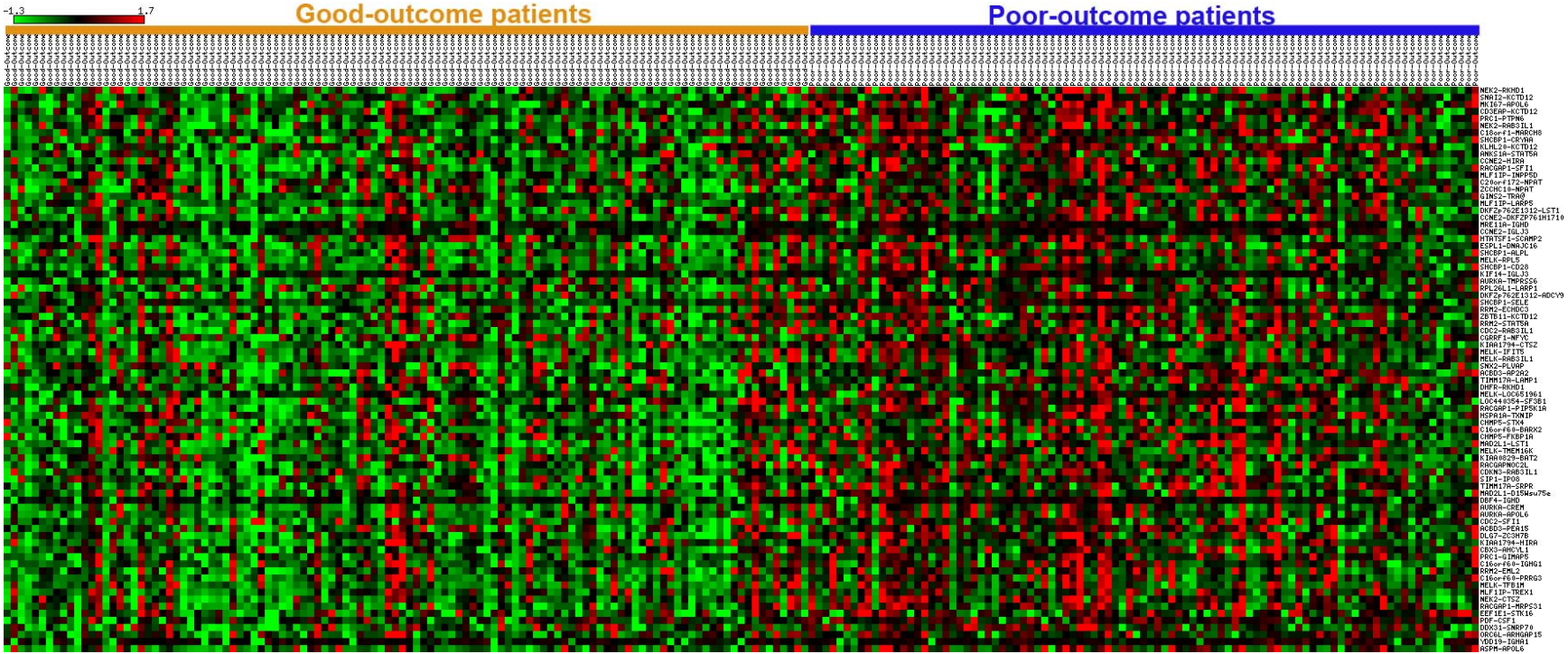
http://mbostock.github.io/d3/talk/20111018/calendar.html

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://graphics.wsj.com/infectious-diseases-and-vaccines/

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://www.biomedcentral.com/1471-2105/9/125

Discussion

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

SCATTERPLOT MATRIX

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

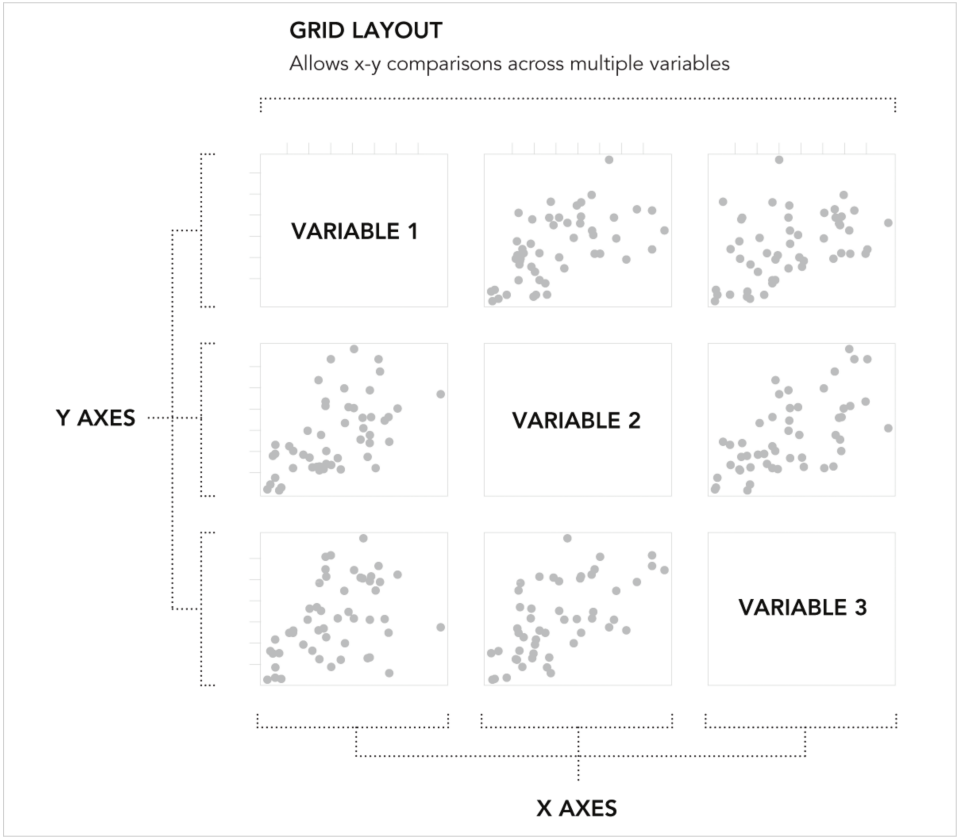


FIGURE 6-8 Scatterplot matrix framework

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

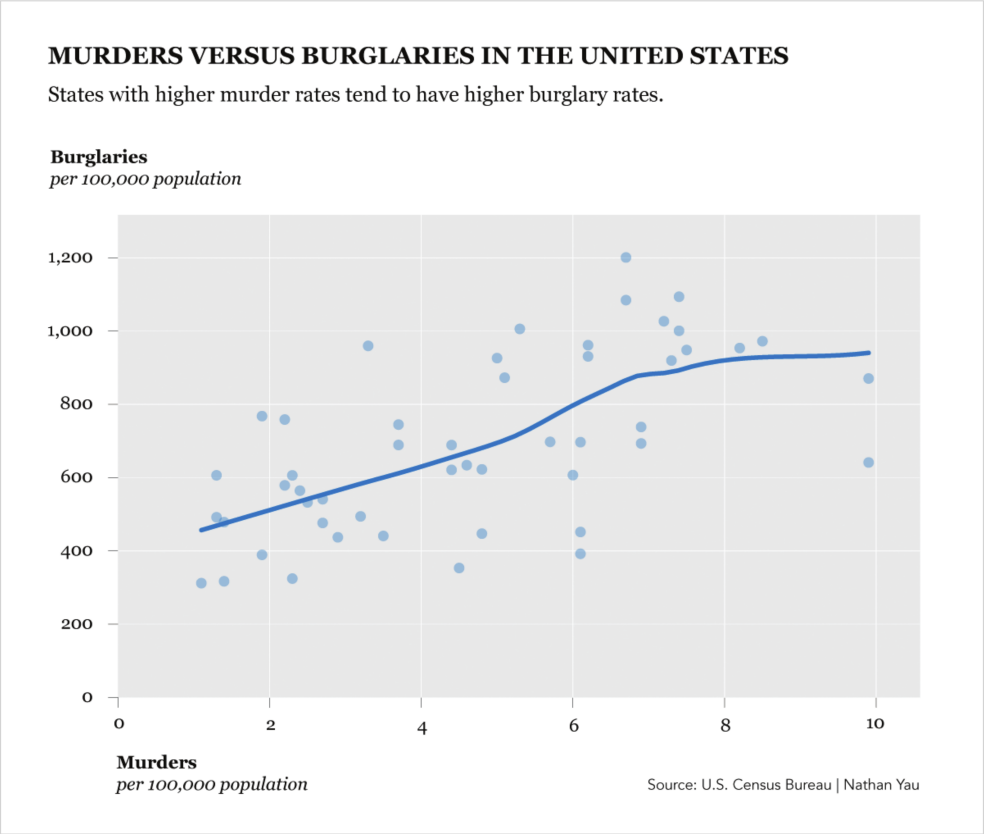


FIGURE 6-7 Revised scatterplot on murder versus burglary

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

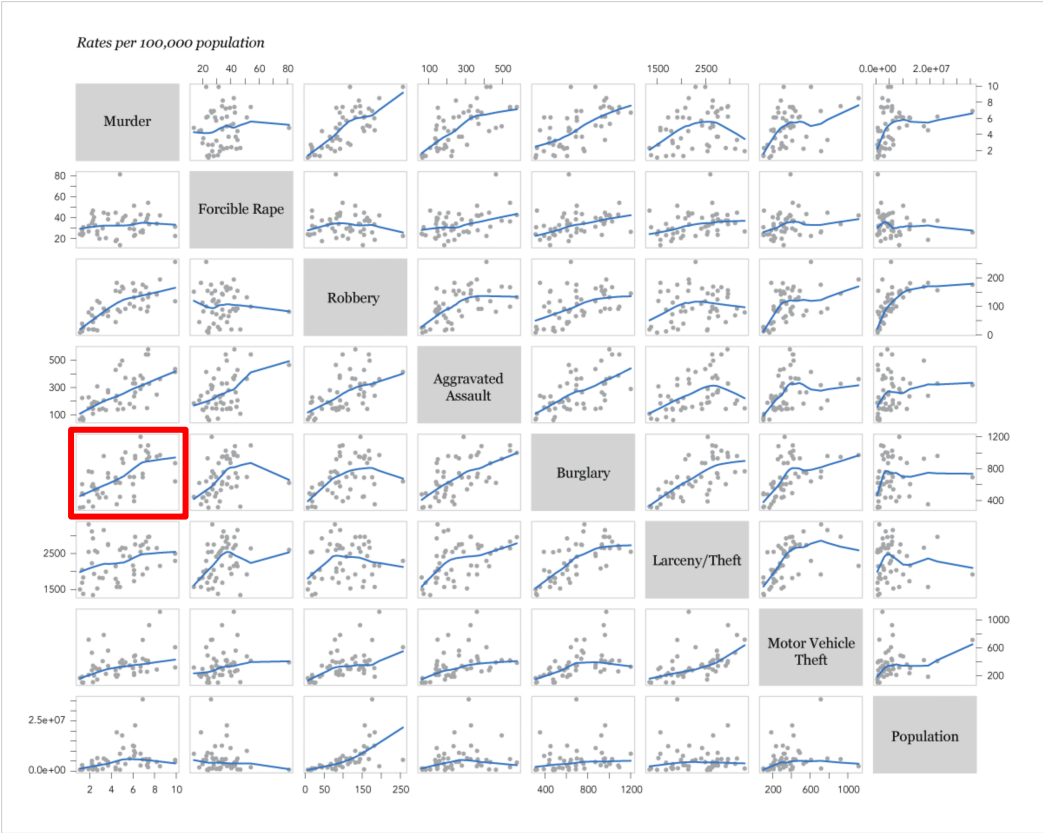
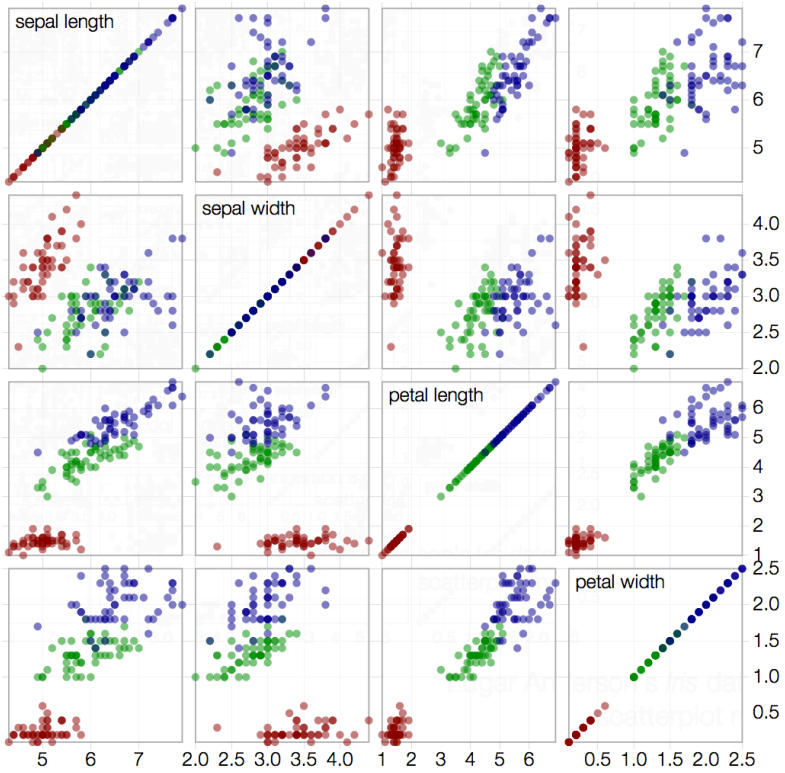


FIGURE 6-9 Scatterplot matrix of crime rates

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

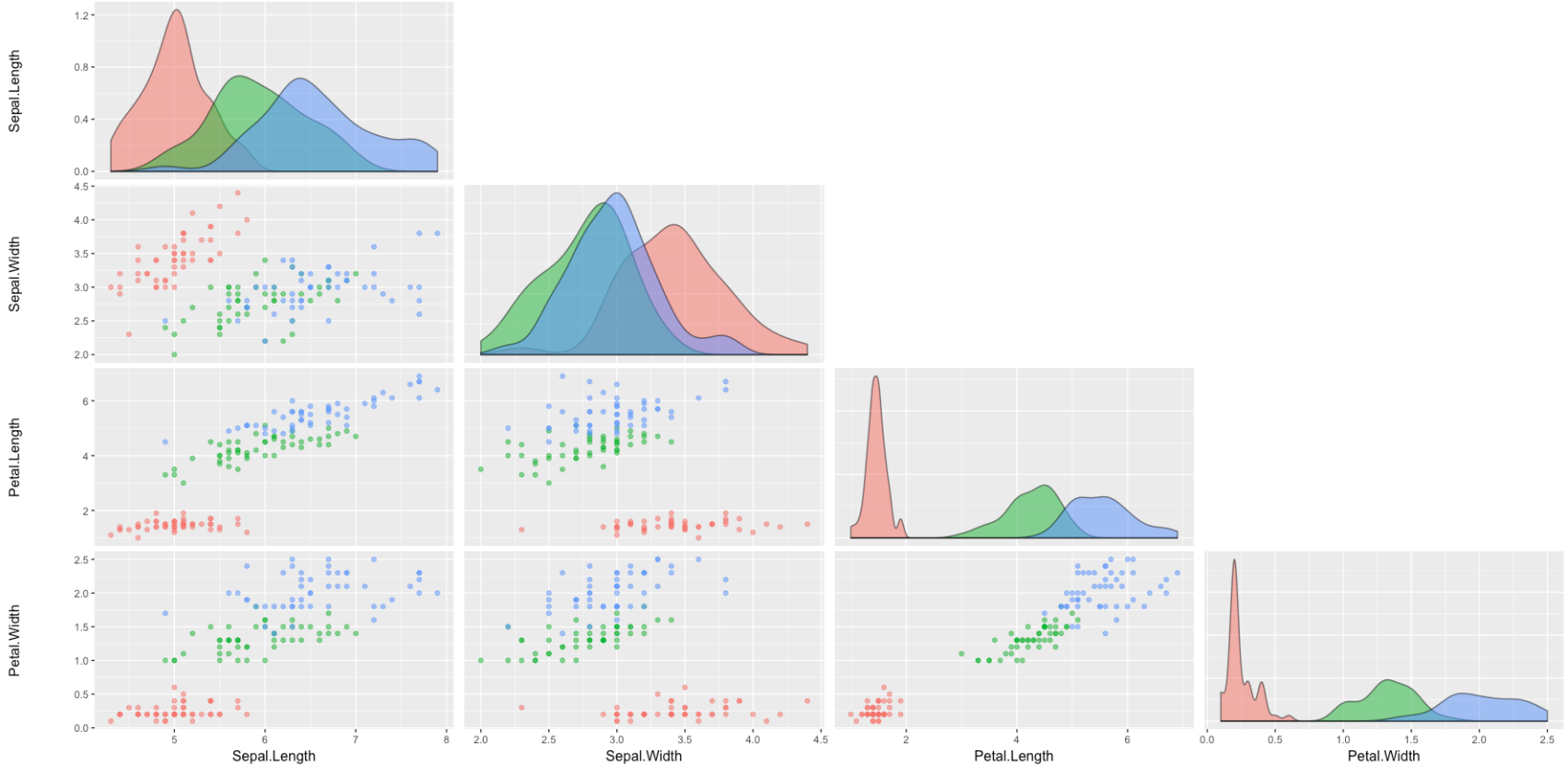


- *Iris setosa*
- *Iris versicolor*
- *Iris virginica*

Edgar Anderson's *Iris* data set
scatterplot matrix

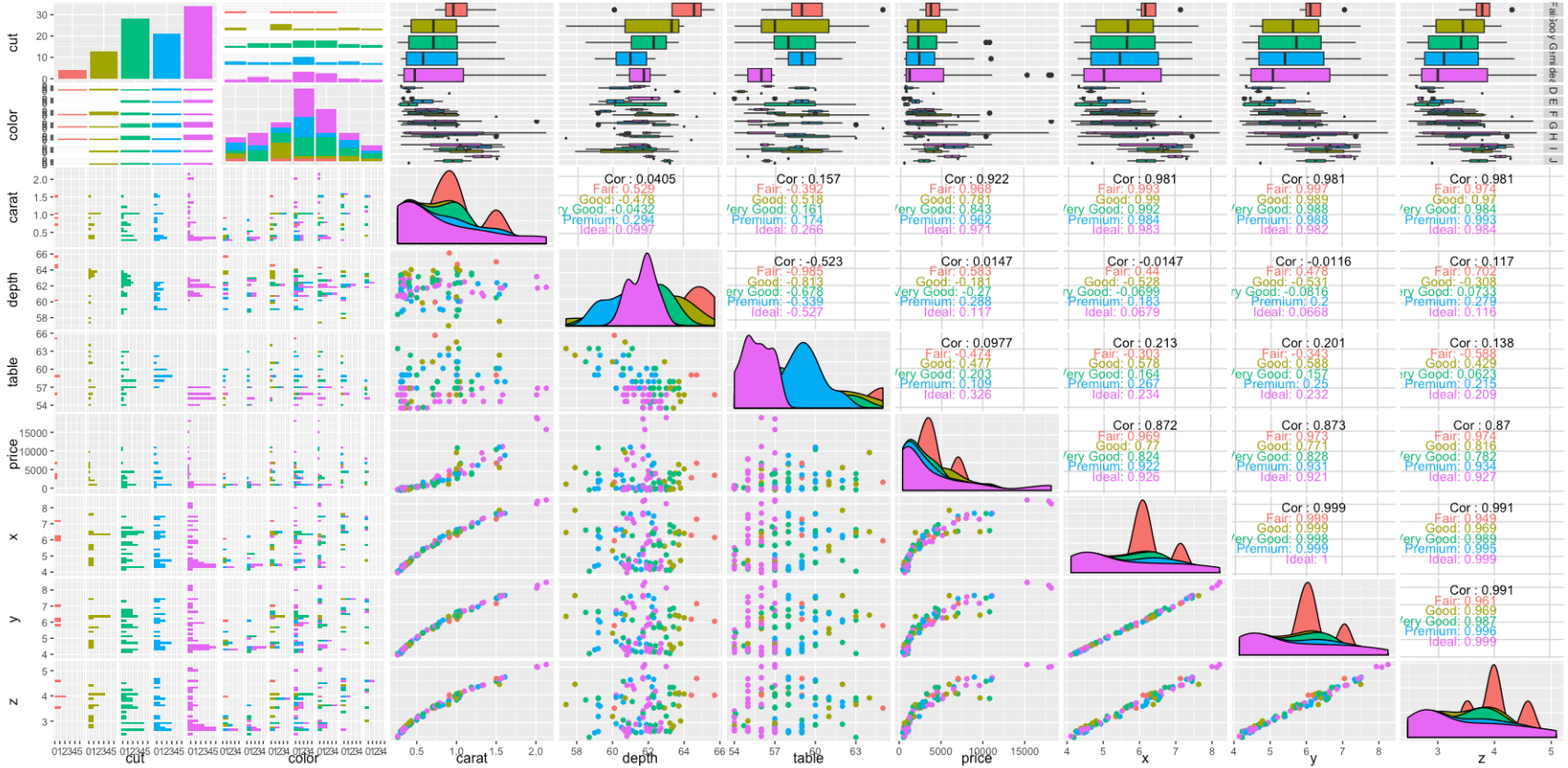
http://mbostock.github.com/d3/talk/20111116/iris-splom.html

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://gettinggeneticsdone.blogspot.com/2011/07/scatterplot-matrices-in-r.html

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



https://cran.r-project.org/web/packages/GGally/vignettes/ggpairs.html

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

SMALL MULTIPLES

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

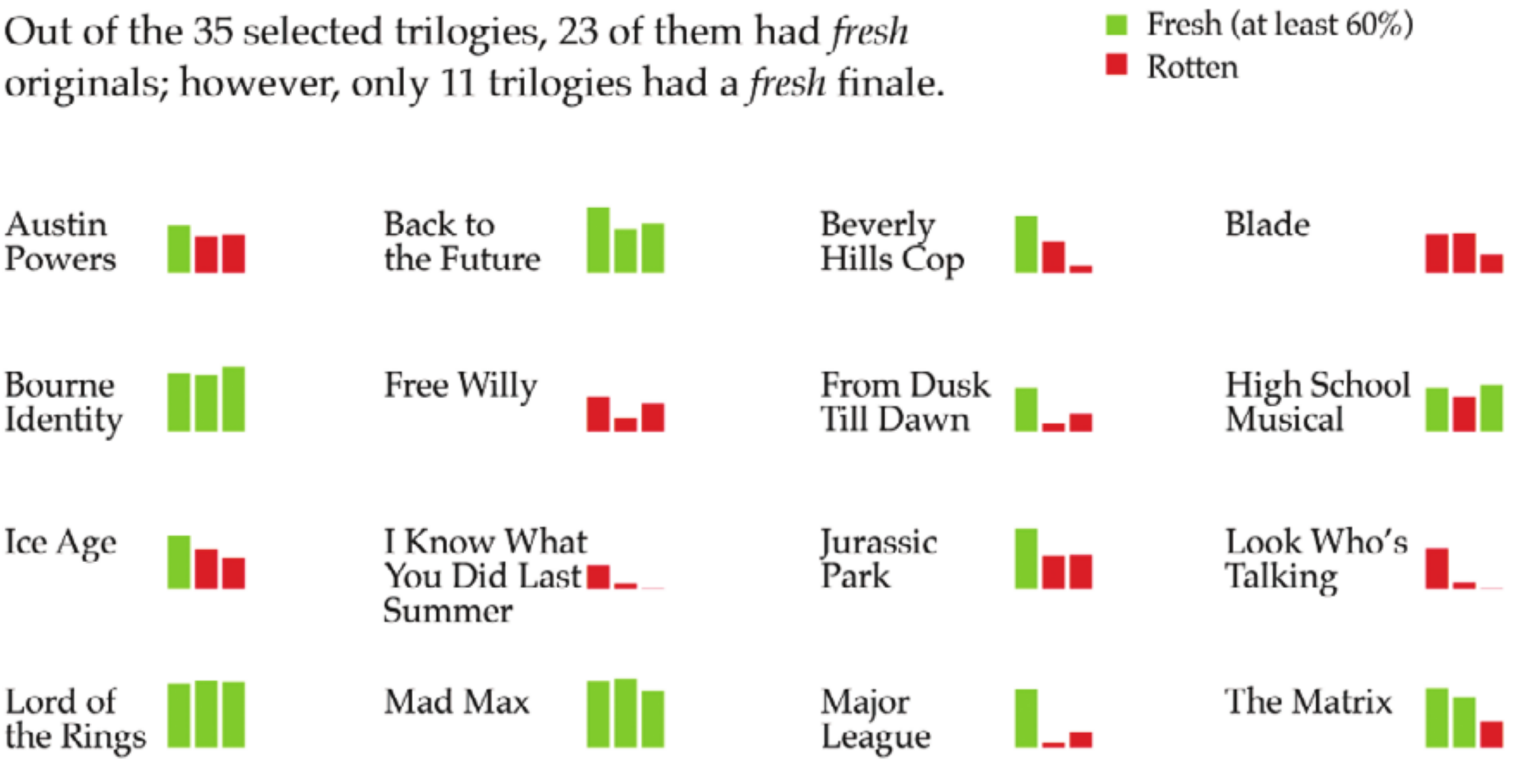
Encode Data

- 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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

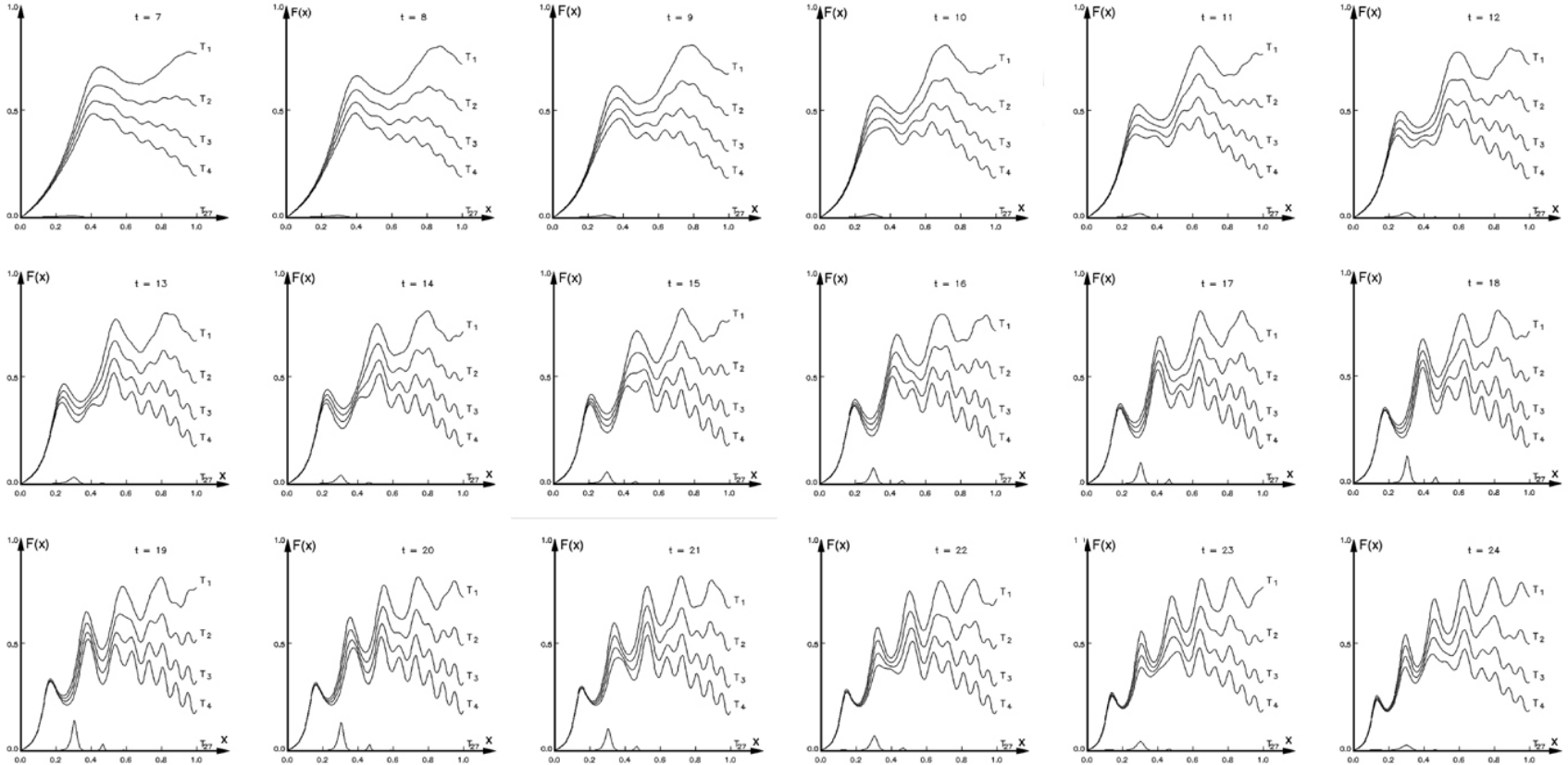
Fresh Originals / Rotten Finales

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



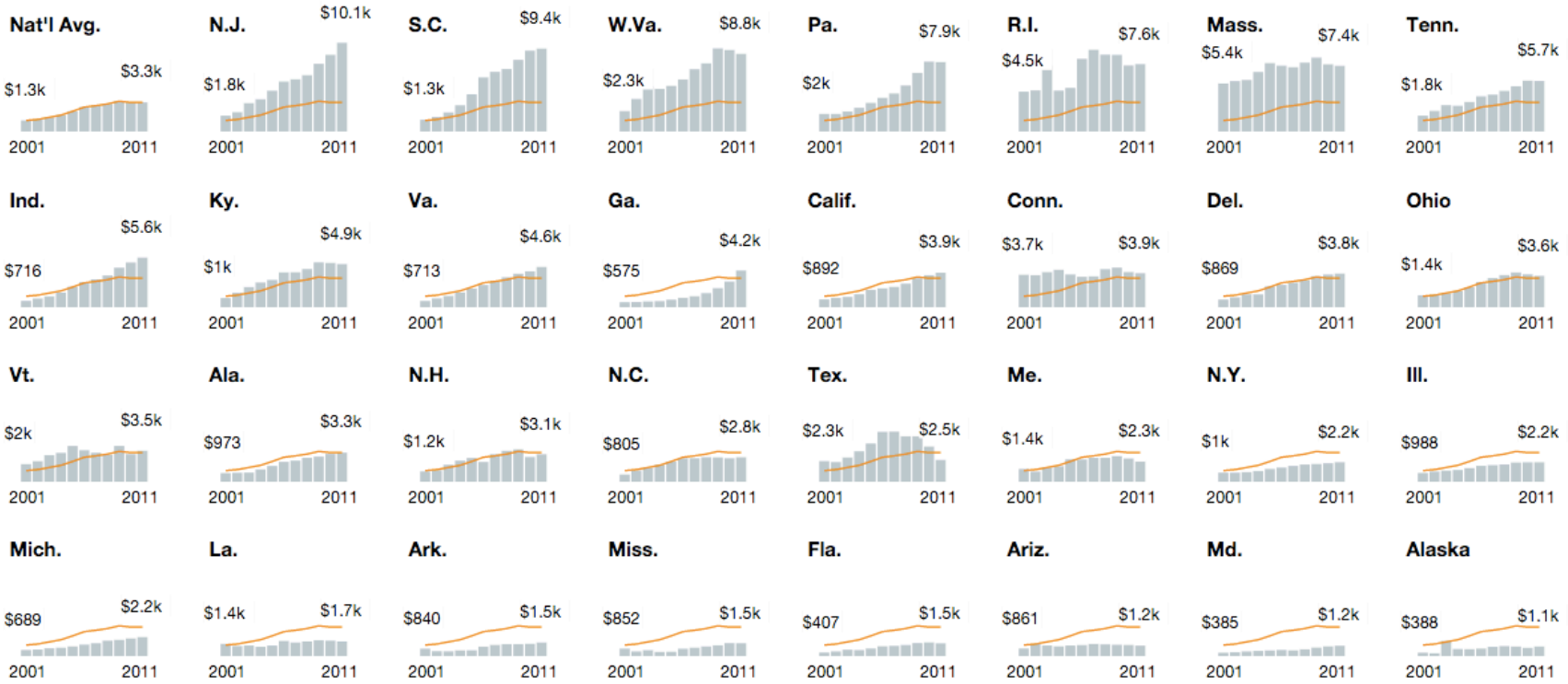
"Visualize This: The FlowingData Guide to Design, Visualization, and Statistics" by Nathan Yau, 2011.

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://www.sv.vt.edu/classes/ESM4714/methods/CogVizCmp.html

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://projects.propublica.org/graphics/ambulances

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

PARALLEL COORDINATES

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

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

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

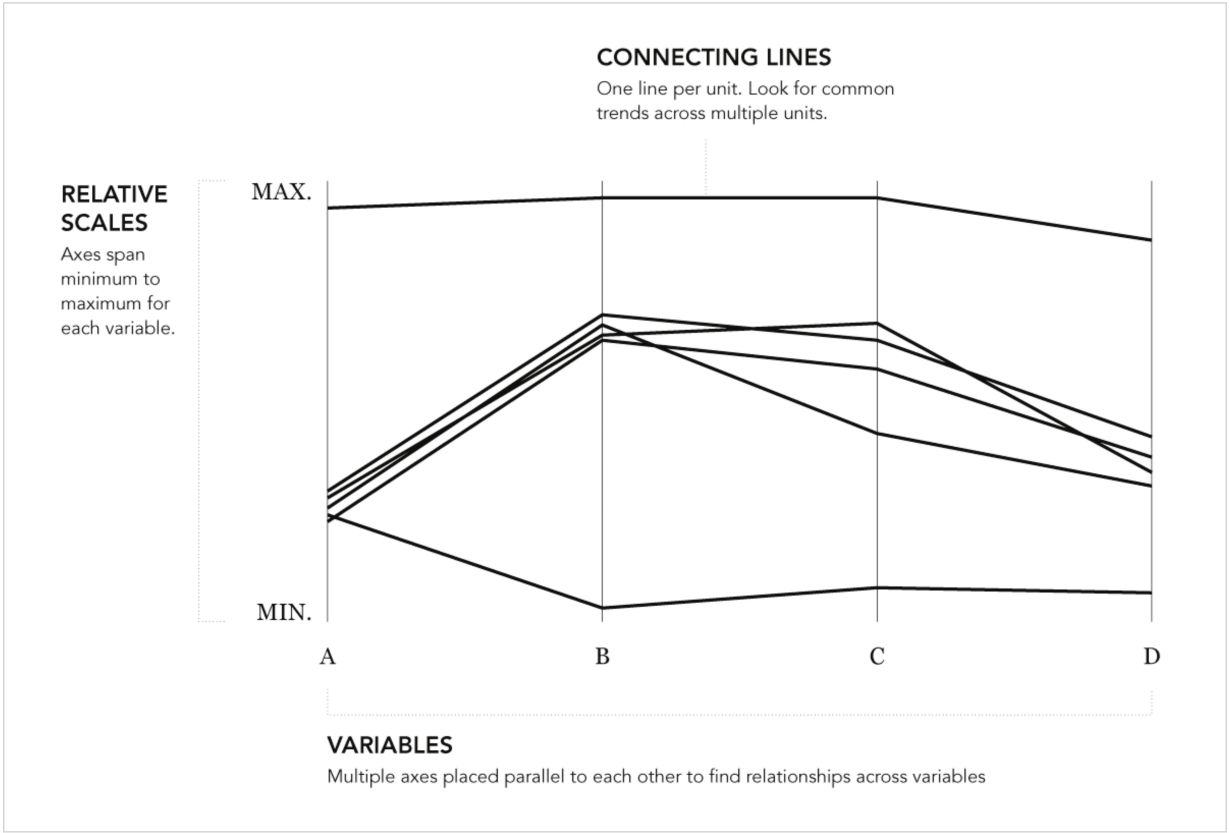


FIGURE 7-20 Parallel coordinates framework

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

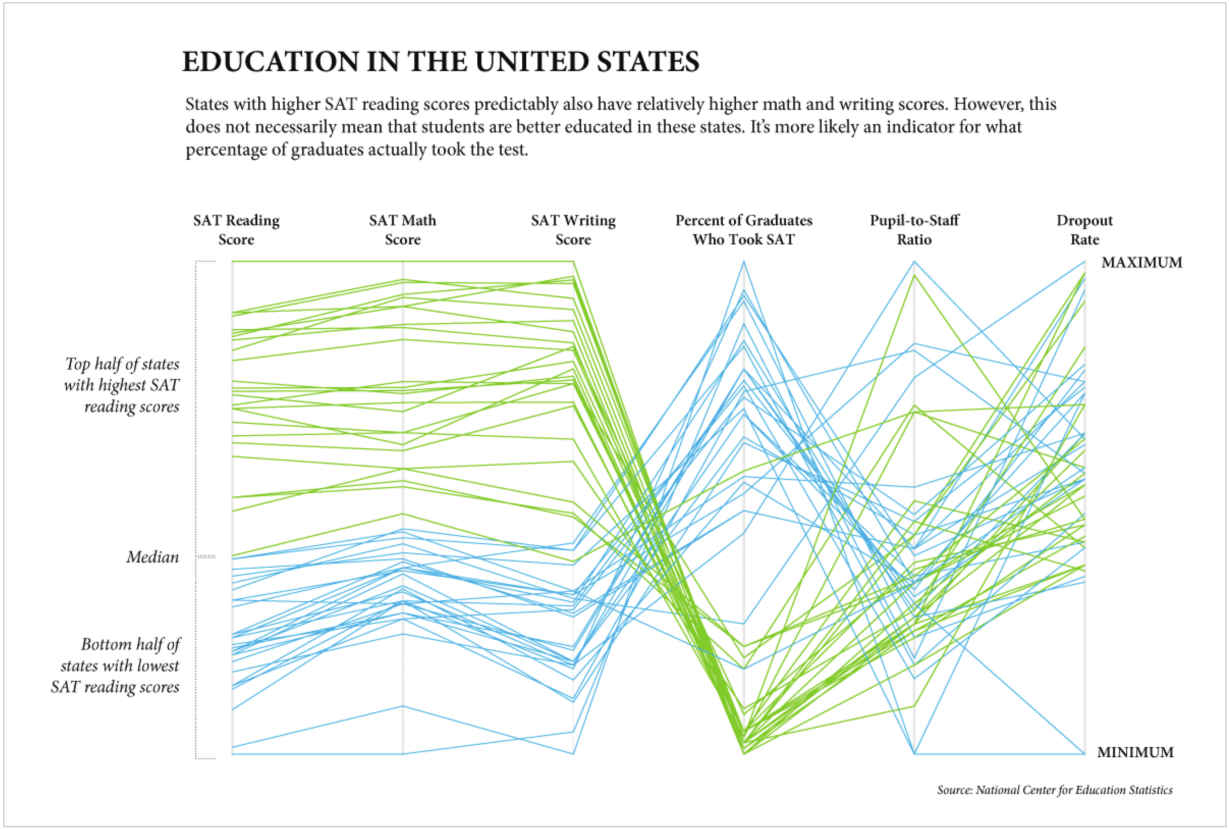
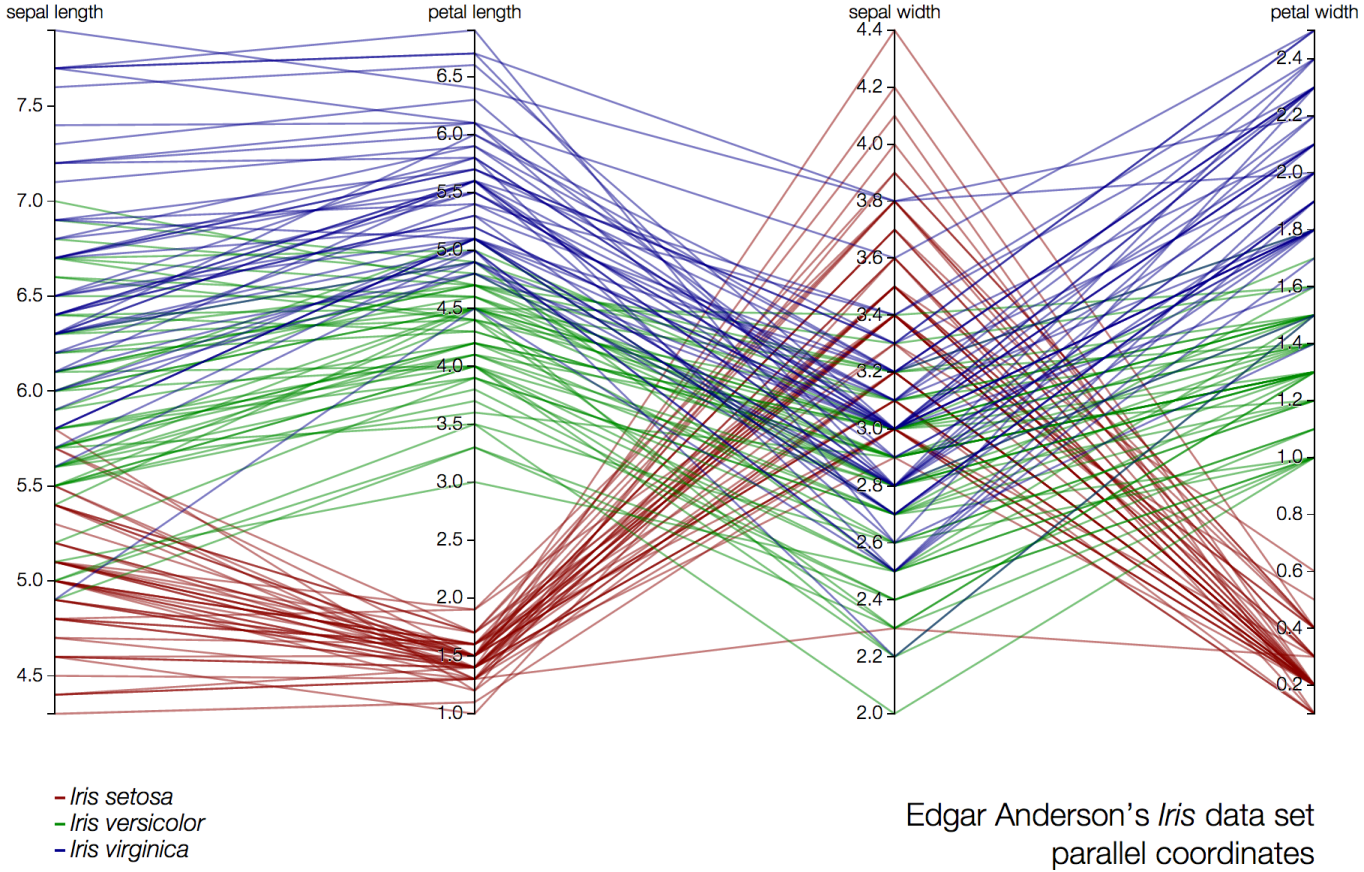


FIGURE 7-26 Standalone parallel coordinates plot on SAT scores

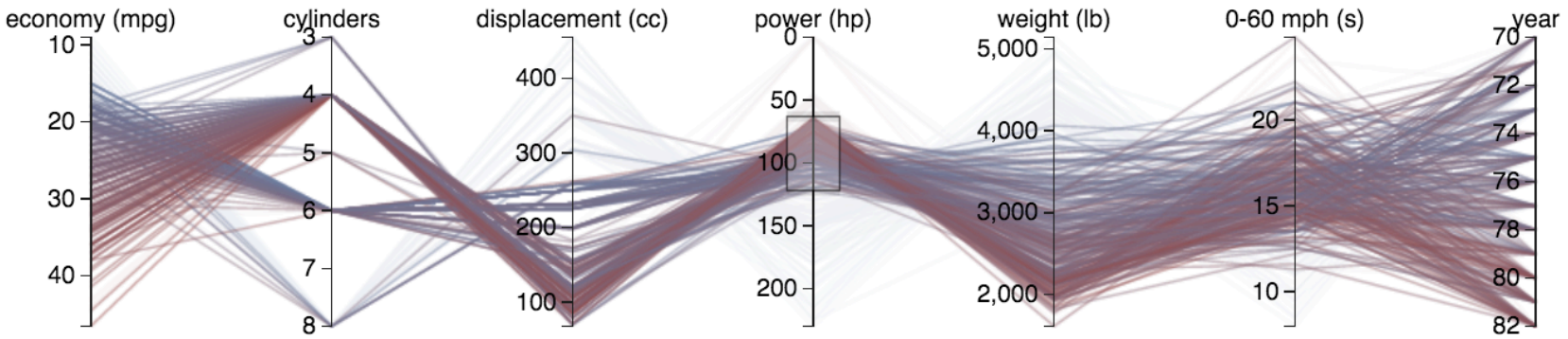
"Visualize This: The FlowingData Guide to Design, Visualization, and Statistics" by Nathan Yau, 2011.



Edgar Anderson's *Iris* data set
parallel coordinates

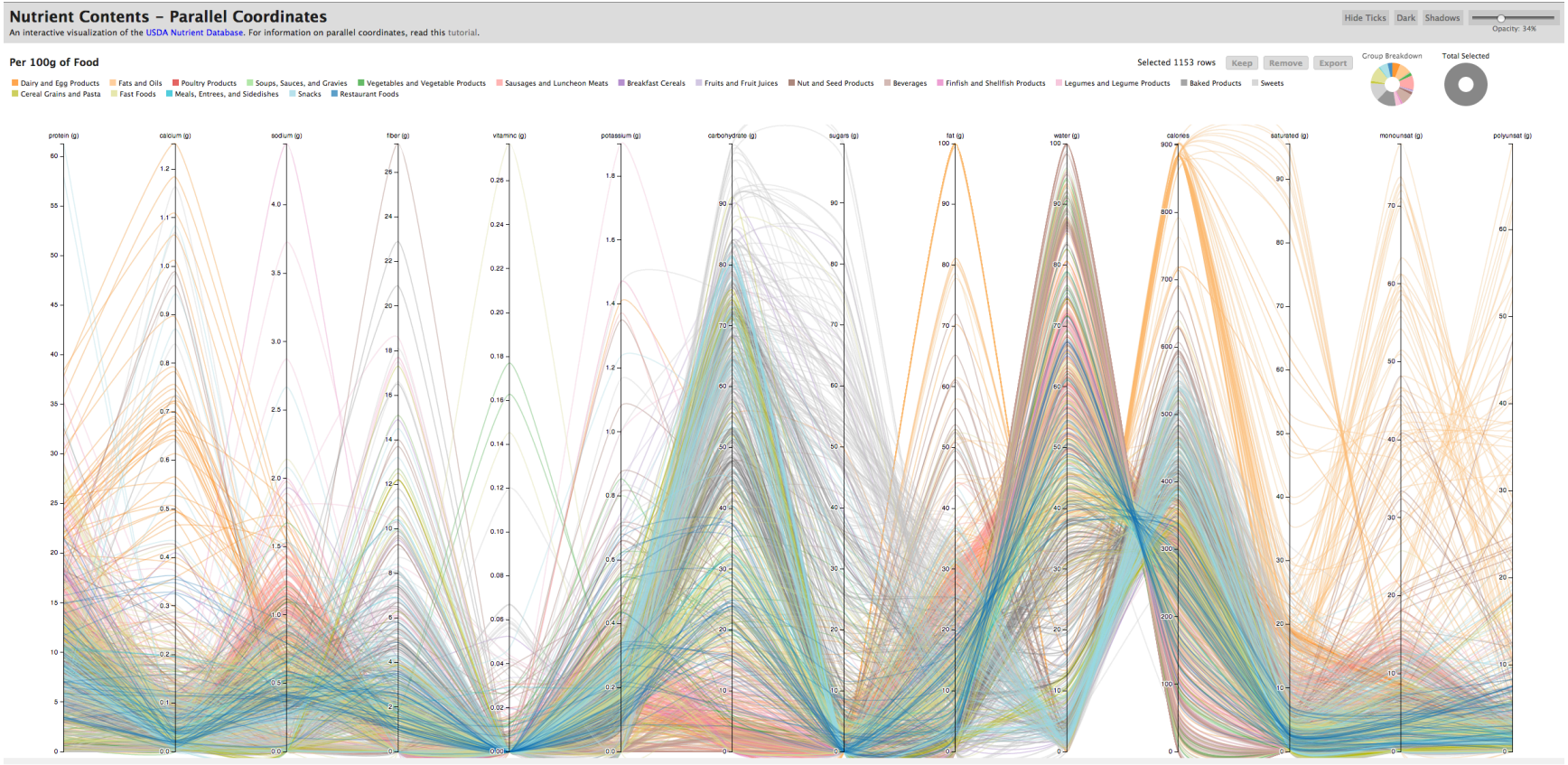
http://mbostock.github.com/d3/talk/20111116/iris-parallel.html

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



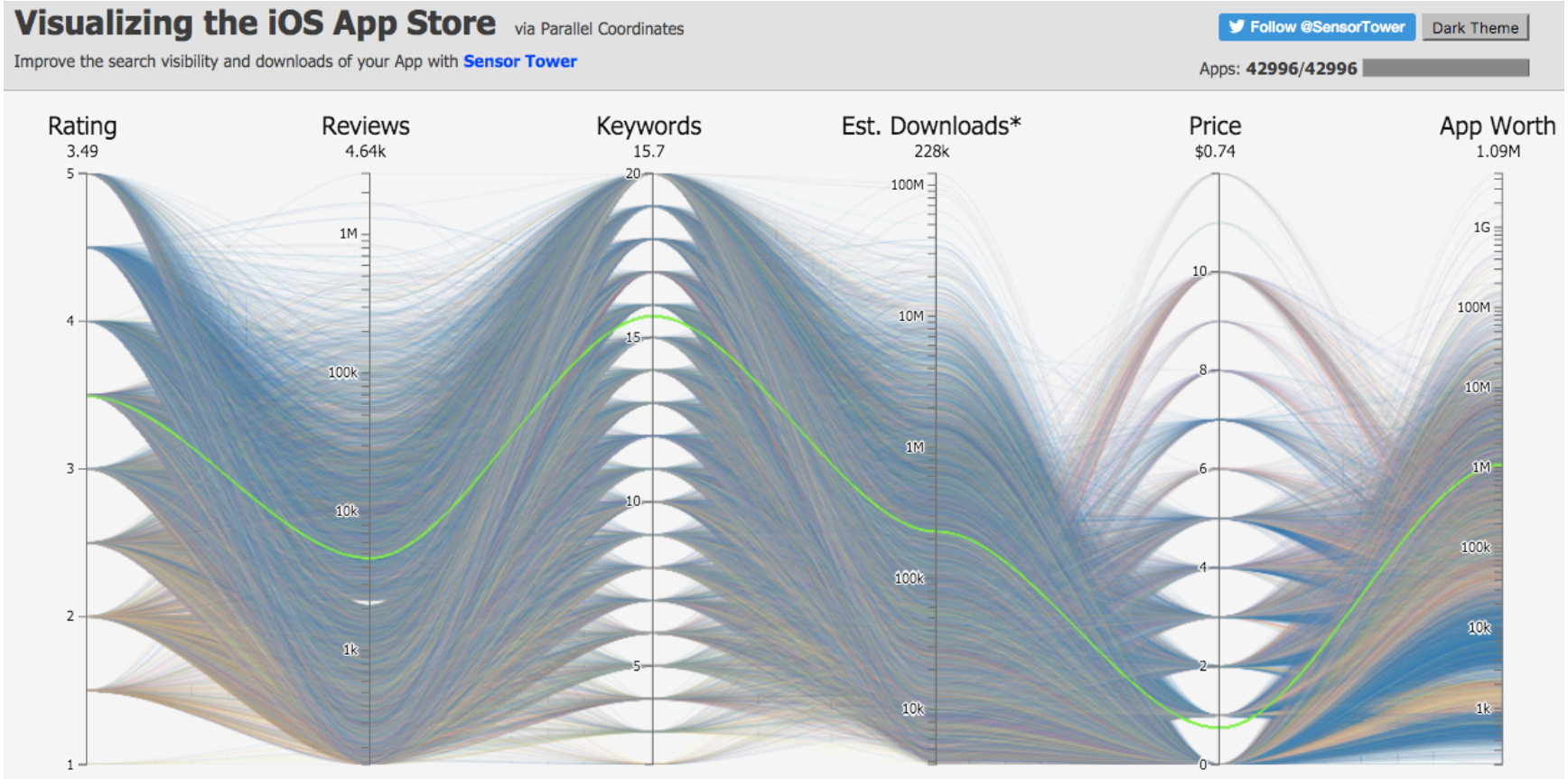
https://syntagmatic.github.io/parallel-coordinates/examples/brushing.html

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



http://exposedata.com/parallel/

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO



https://sensortower.com/visualizing-the-ios-app-store

Discussion

- Long startup time
- Almost always requires interactivity
  - Choose column to color by
  - Choose how to sort
  - Highlighting (brushing)
  - Clustering
- Very high density and data ink ratio, low lie factor

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

OTHER TECHNIQUES

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

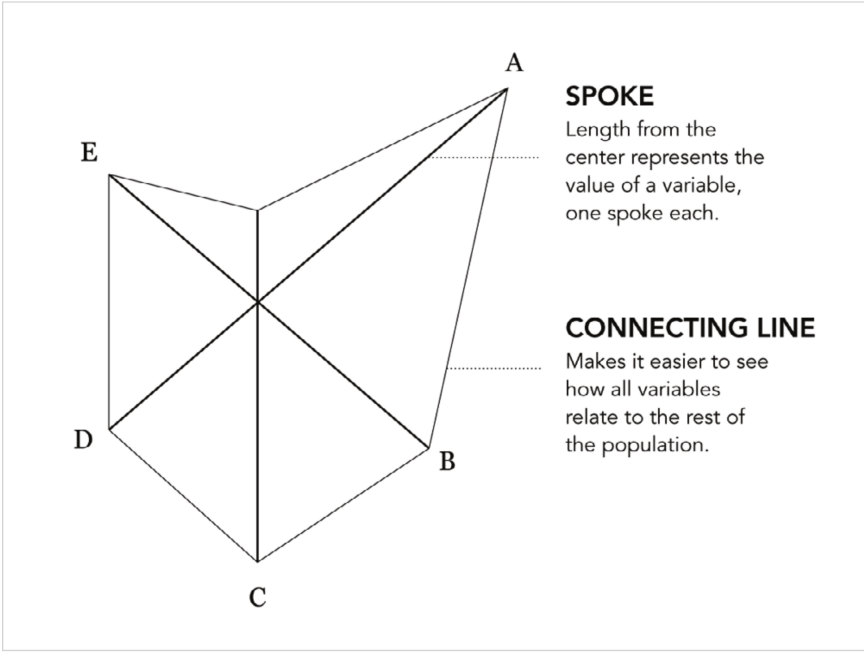


FIGURE 7-14 Star chart framework

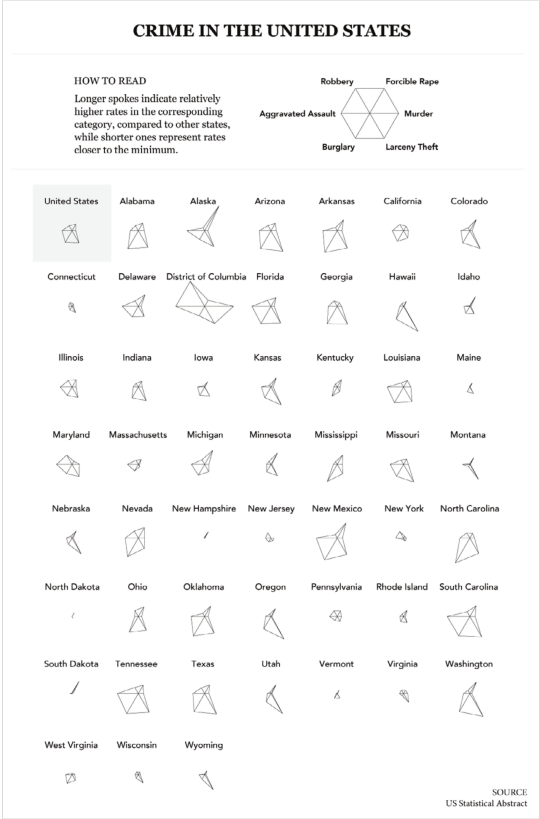
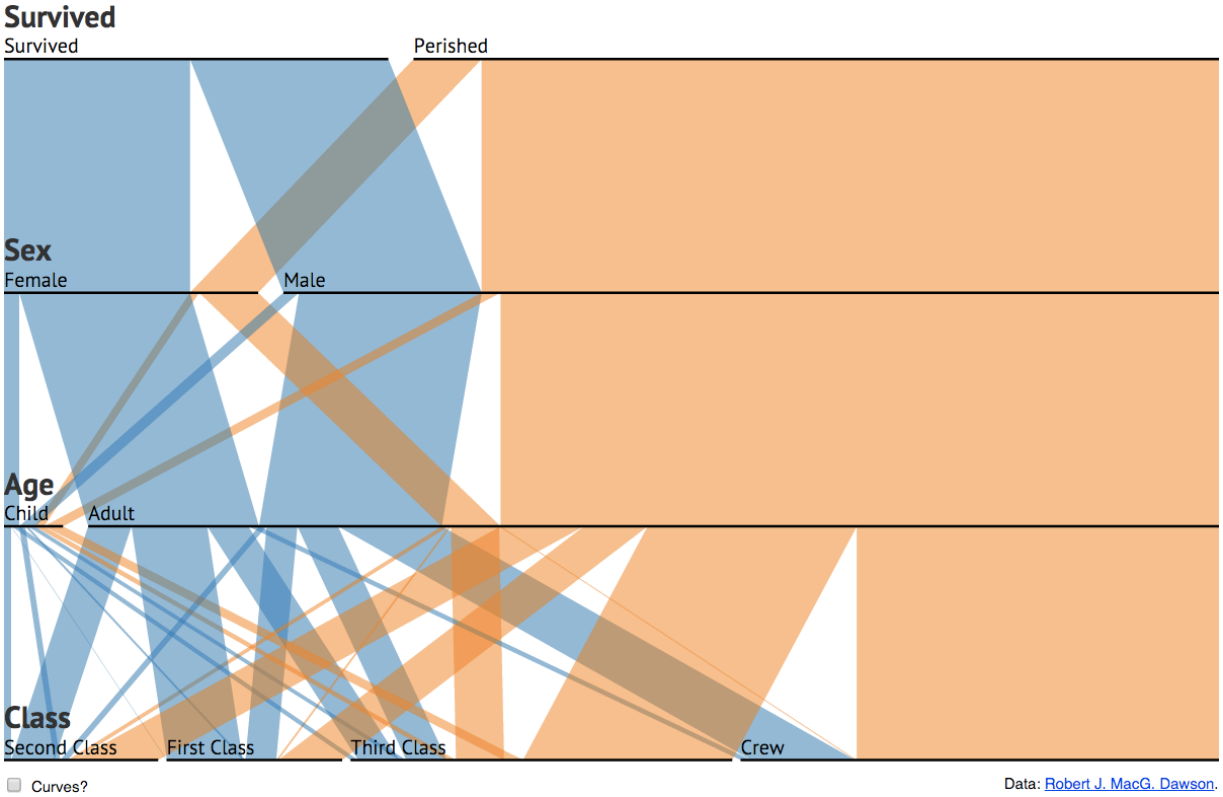


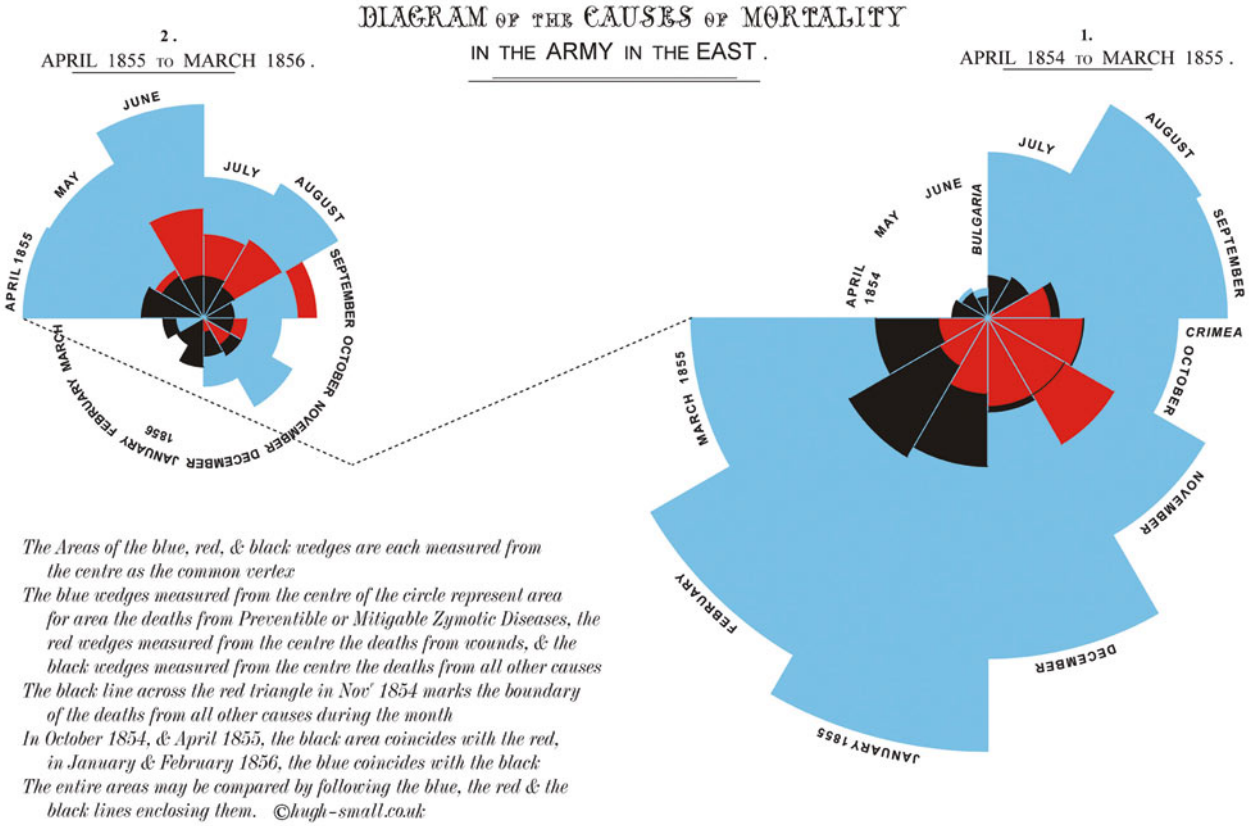
FIGURE 7-19 Series of star charts showing crime by state

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

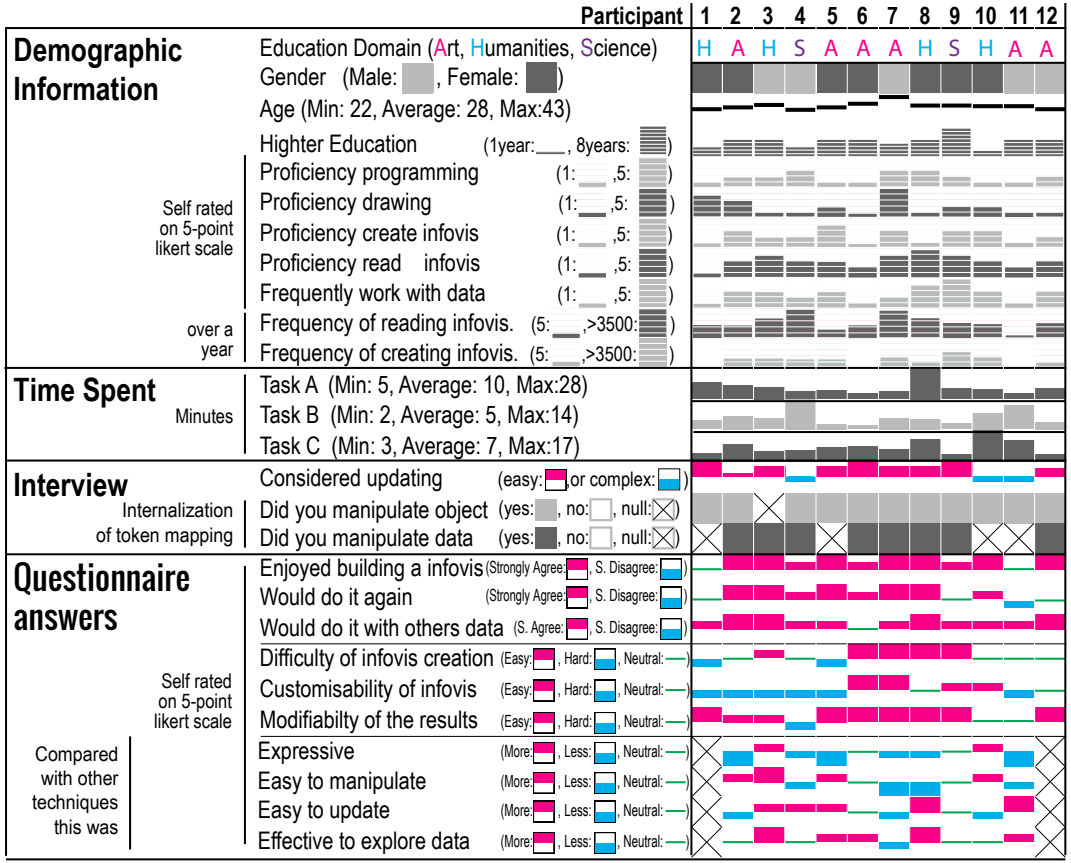
Titanic Survivors



http://eagereyes.org/parallel-sets and https://www.jasondavies.com/parallel-sets/



http://dd.dynamicdiagrams.com/2008/01/nightingales-rose/



http://bertifier.com/ and http://constructive.gforge.inria.fr/#!questionnaires.md

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

CONCLUSION

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

References

- Nathan Yau, **Visualize This: The FlowingData Guide to Design, Visualization, and Statistics**, Wiley Publishing, 2011.
- Mike Bostock, **Data Driven Documents** (D3.js)
  - https://github.com/mbostock/d3/wiki/Gallery
  - http://bl.ocks.org/mbostock
- Other references sourced on slides

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science
UNIVERSITY OF SAN FRANCISCO

CHANGE THE WORLD FROM HERE

Sophie Engle ◆ sjengle@cs.usfca.edu
Department of Computer Science