

Visualizing data with R and RStudio

ME 447/547

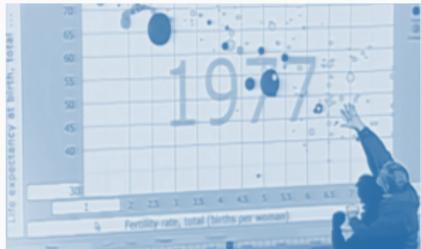
Richard Layton

March 2020

Rose-Hulman Institute of Technology

This week we encounter the three basic themes of the course

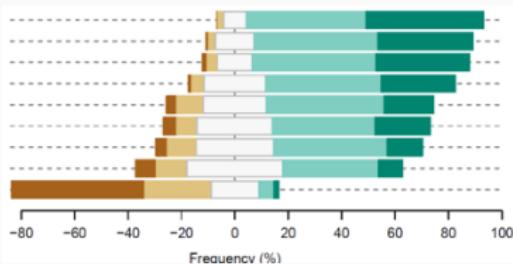
Rhetoric
(Mon)



Means
(T/W)

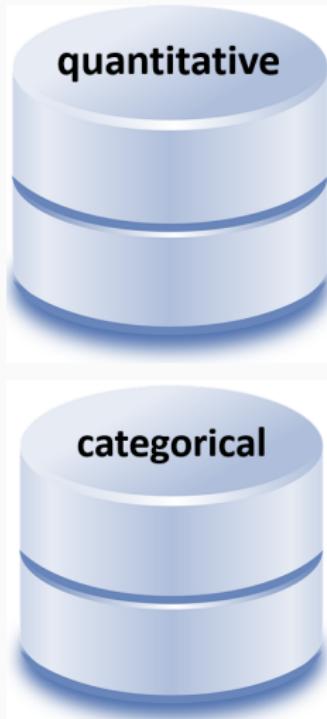


Repertoire
(Thu)



Repertoire

Graph design begins by understanding the data structure ...



Number of variables?

Continuous or discrete?

Number of variables?

Nominal or ordinal?

Number of levels each?

... and by knowing the prior art suited to that structure

62

strip plot

box and whisker plot

multiway

scatterplot

dot plot

line graph

conditioning plot

scatterplot matrix

63

parallel coordinate plot

cycle plot

mosaic plot

financial (OHLC) plot

diverging stacked bar

linked micromaps

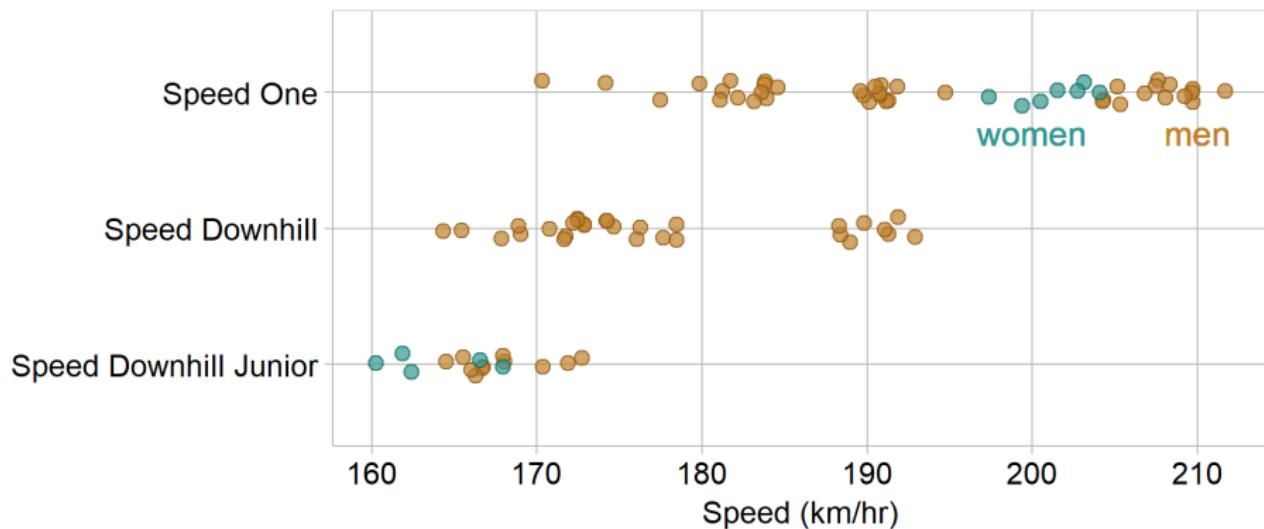
proportional symbol map

dot density map

Gallery – strip plot, jitter plot, or 1D scatterplot

Quantitative: speed (continuous), $N_{\text{obs}} = 91$

Categorical: event (nominal, 3 levels), sex (nominal, 2 levels)

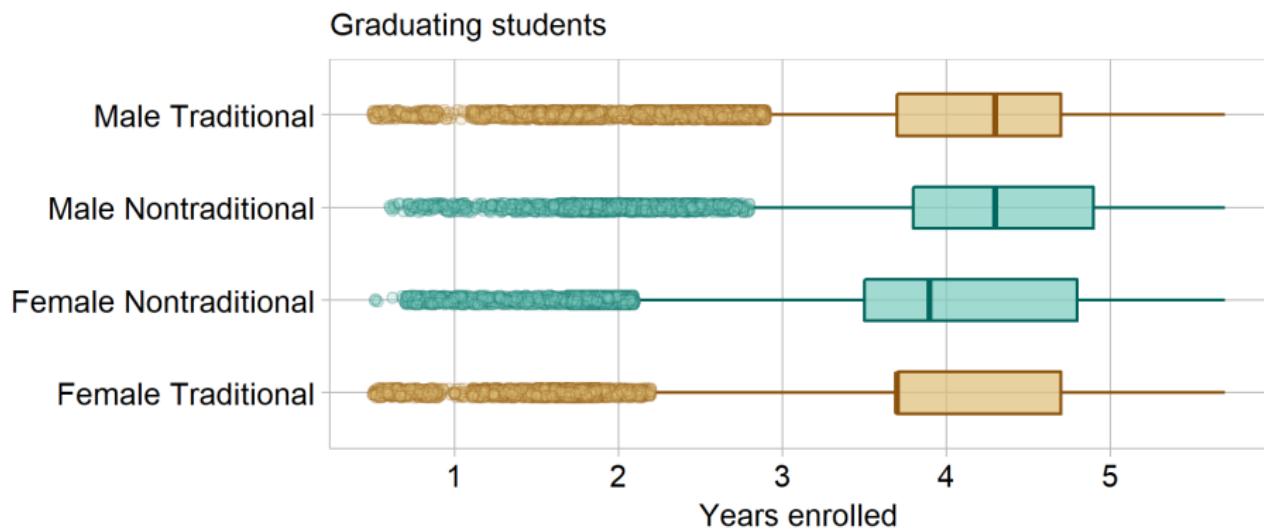


Data source (Unwin, 2015)

Gallery – box and whisker or box plot

Quantitative: Years enrolled (continuous), $N_{\text{obs}} = 269057$

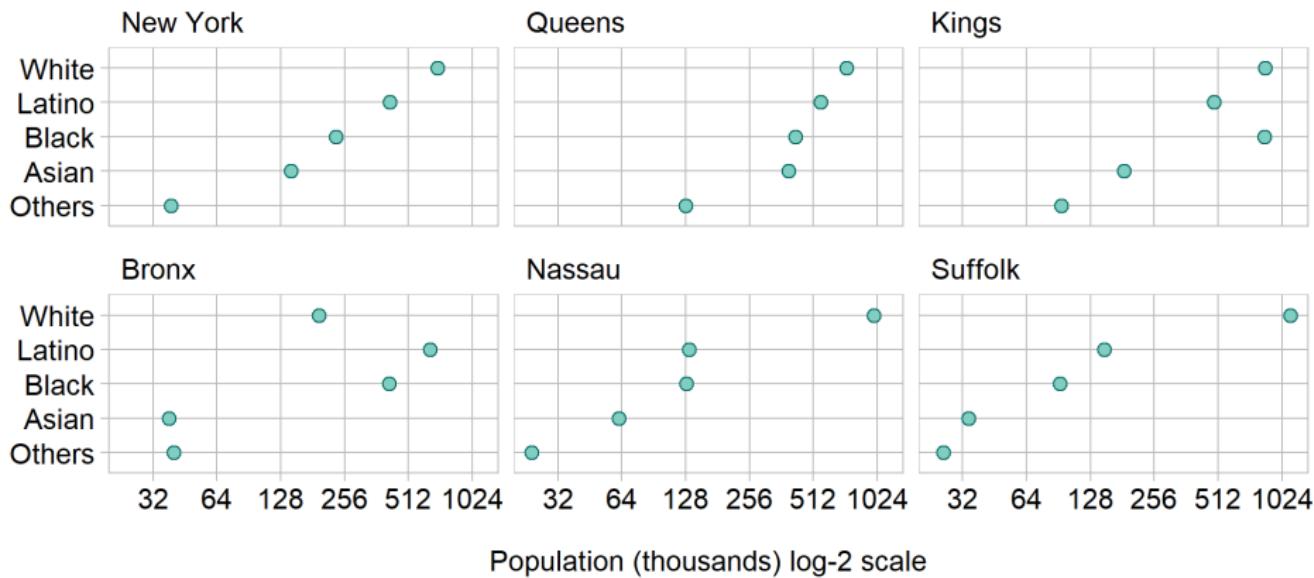
Categorical: Path + sex (nominal, 4 levels)



Gallery – multiway

Quantitative: Population (continuous), $N_{\text{obs}} = 30$

Categorical: Race/ethnicity (nominal, 5L) & county (nominal, 6L)

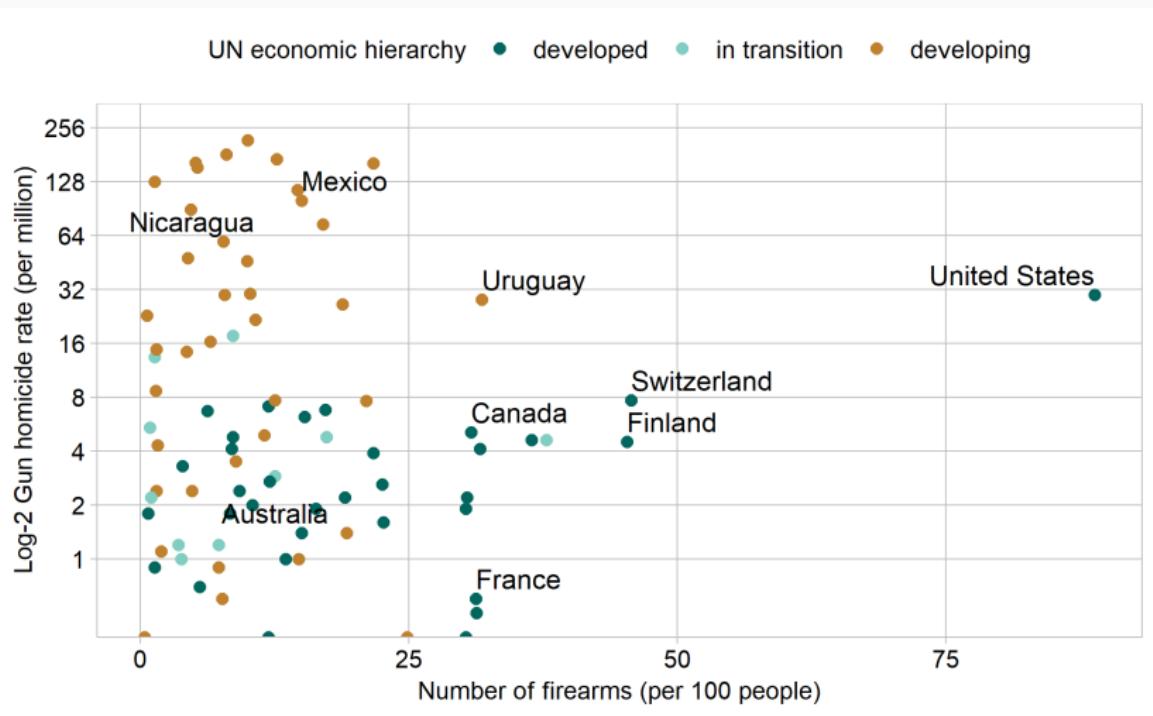


Adapted from (Robbins, 2013)

Gallery – scatterplot

Quantitative: Gun homicides & gun ownership (continuous), $N_{\text{obs}} = 90$

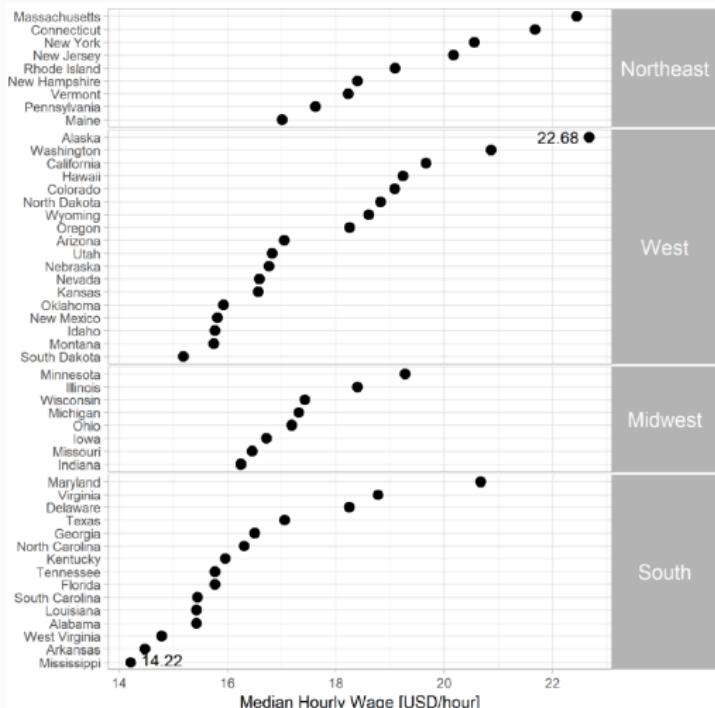
Categorical: Country (nominal, 90L) & economic hierarchy (nominal, 3L)



Gallery – Cleveland dot plot

Quantitative: 2016 median hourly wage (continuous), $N_{\text{obs}} = 50$

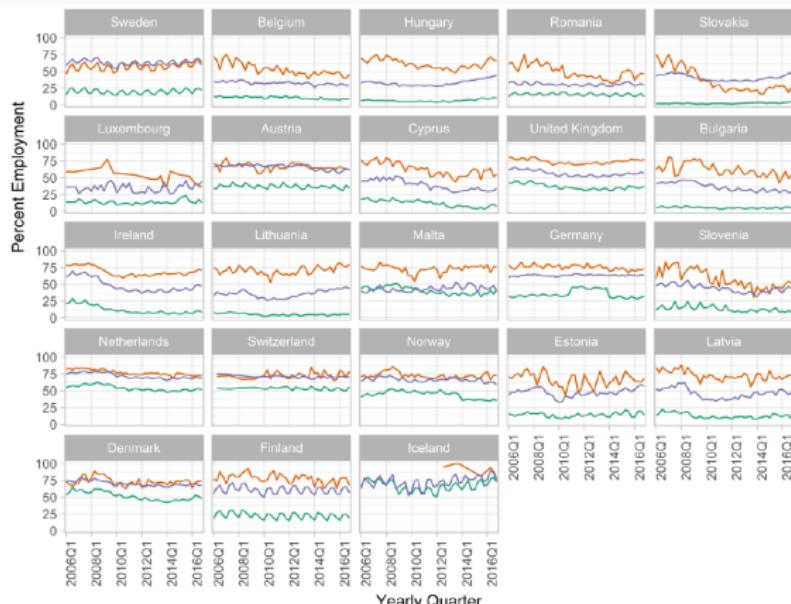
Categorical: State (nominal, 50 levels) & region (nominal, 4 levels)



Gallery – line graph

Quantitative: Percent employment (continuous), $N_{\text{obs}} = 1656$

Categorical: Country (nominal, 23L), education (ord, 3L), quarter (ord, 24L)

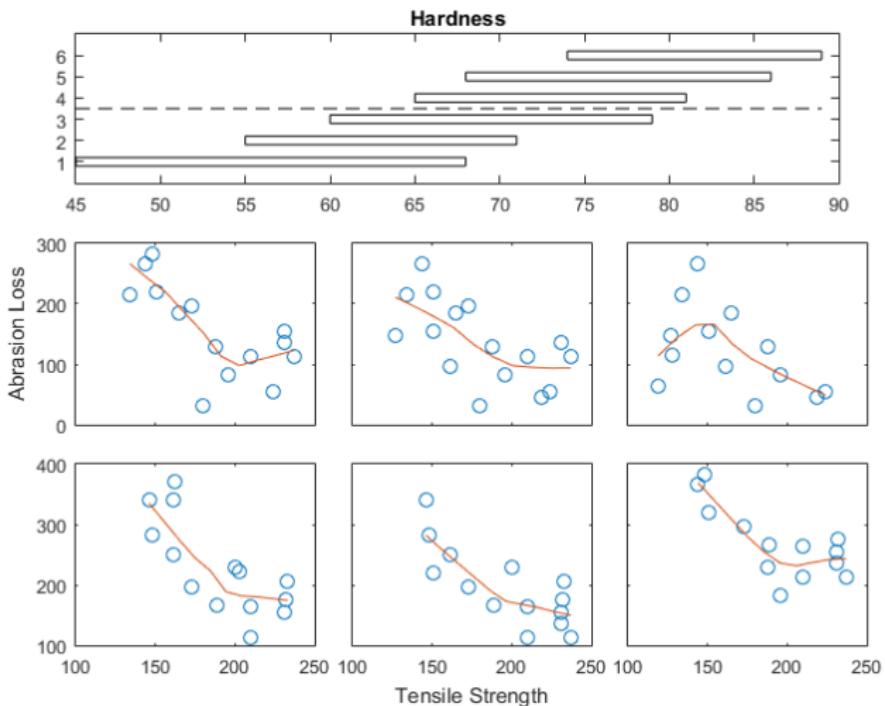


Education Level Completed

- Less than primary, primary and lower secondary education (levels 0-2)
- Upper secondary and post-secondary non-tertiary education (levels 3 and 4)
- Tertiary education (levels 5-8)

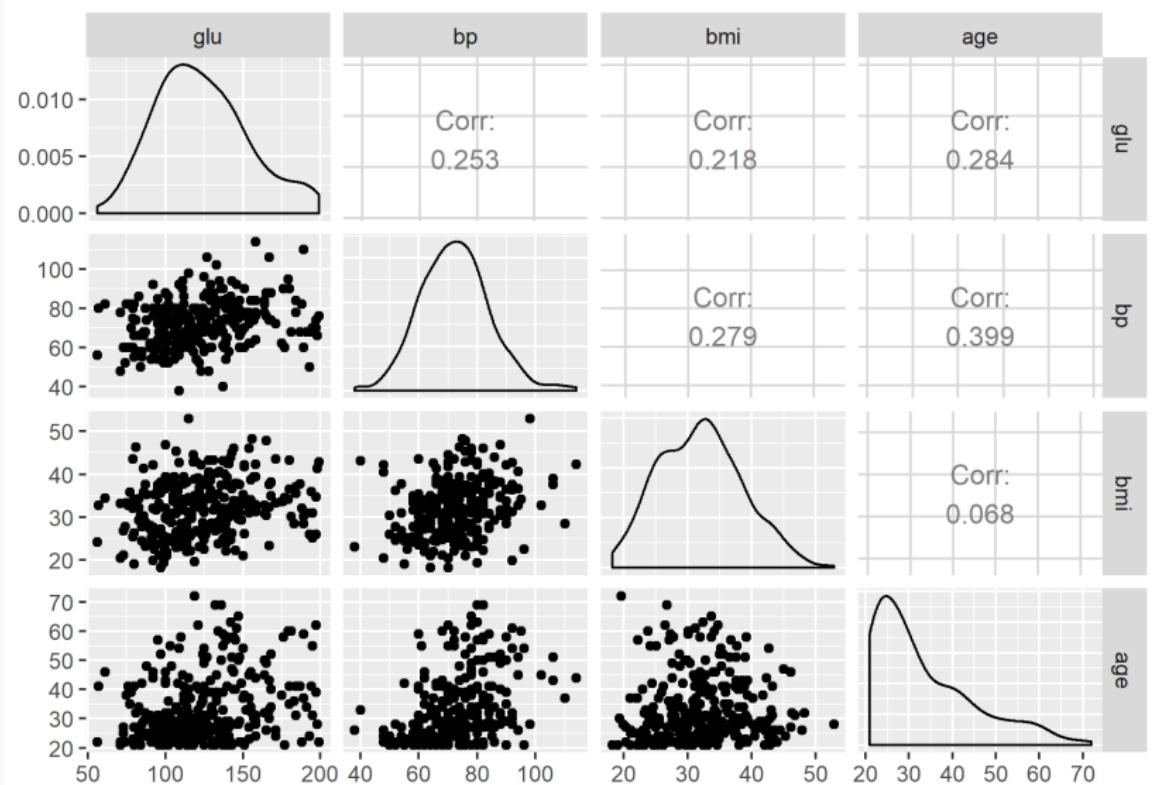
Gallery – conditioning plot

Quantitative: Rubber abrasion loss, tensile strength, & hardness
(all continuous), $N_{\text{obs}} = 30$



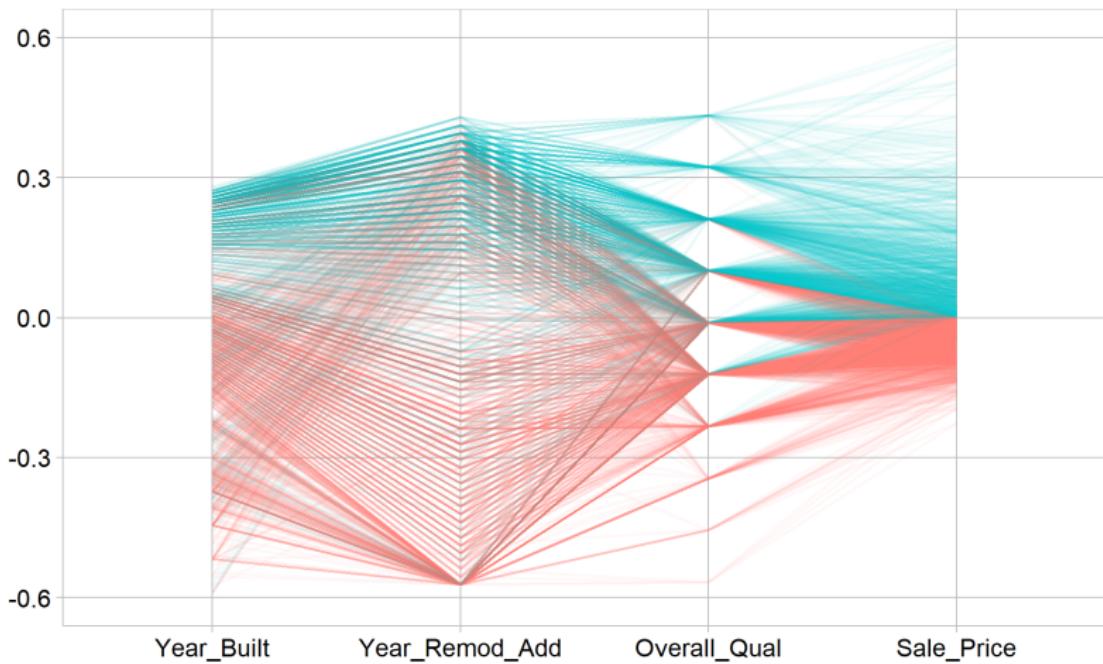
Gallery – scatterplot matrix

Quantitative: glucose, blood pressure, BMI, age (continuous), $N_{\text{obs}} = 300$



Gallery – parallel coordinate

Quantitative: Year built, remodeled, & sale price (continuous),
quality (discrete) $N_{\text{obs}} = 2930$

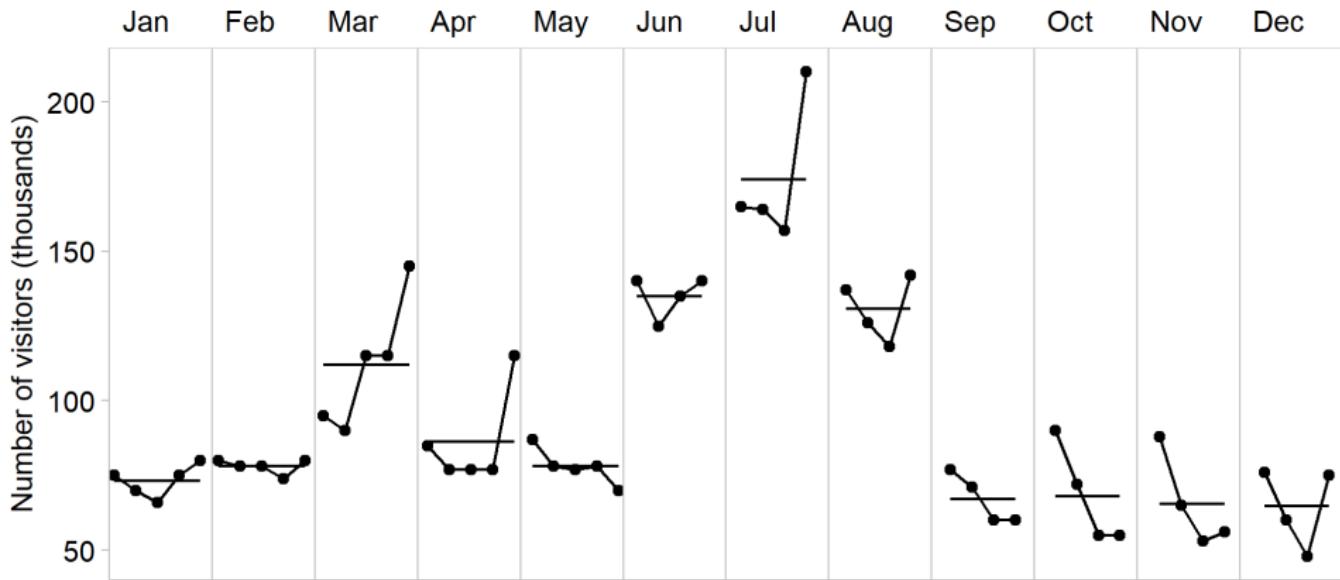


Gallery – cycle plot

Quantitative: Number of visitors (continuous), $N_{\text{obs}} = 53$

Categorical: Month (ordinal, 12 levels), year (ordinal, 5 levels)

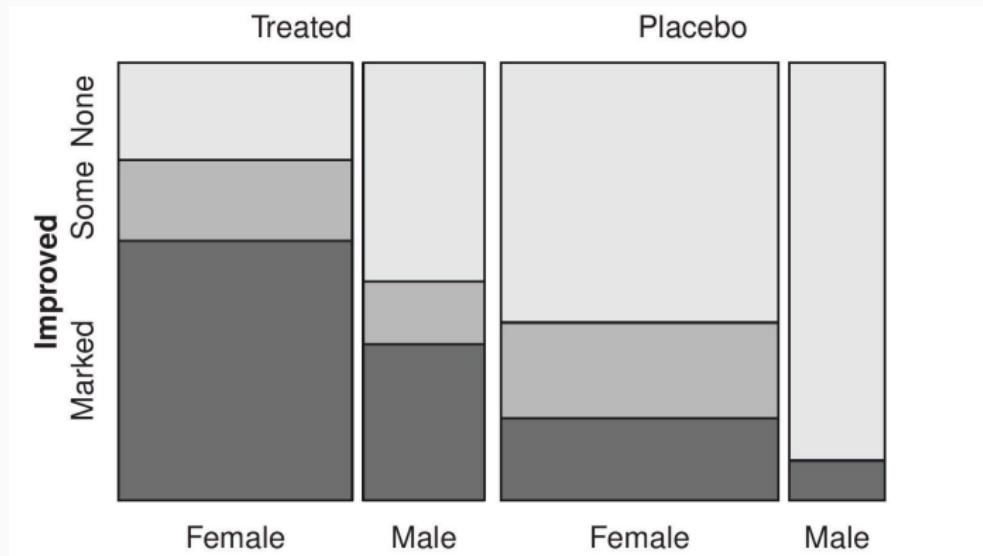
St. Louis Science Center attendance, 1998 to 2002



Gallery – mosaic plot

Quantitative: Frequency (continuous), $N_{\text{obs}} = 84$

Categorical: Sex (nomi, 2L), treatment (nomi, 2L), outcome (ordi, 3L)

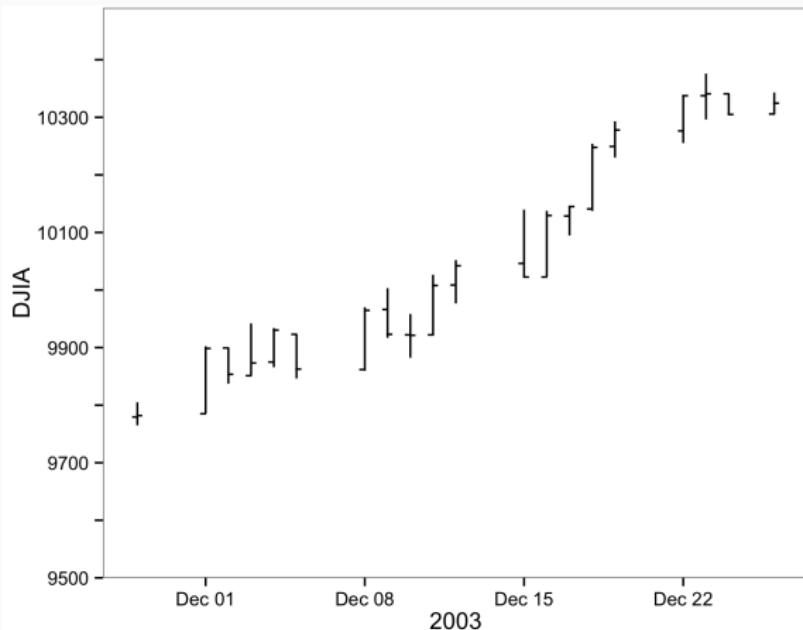


David Meyer, Achim Zeileis, and Kurt Hornik (2017) *vcd: Visualizing Categorical Data*, R package version 1.4-4,
arthritis treatment data.

Gallery – financial (OHLC) plot

Quantitative: Opening, high, low, closing price (continuous), $N_{\text{obs}} = 20$

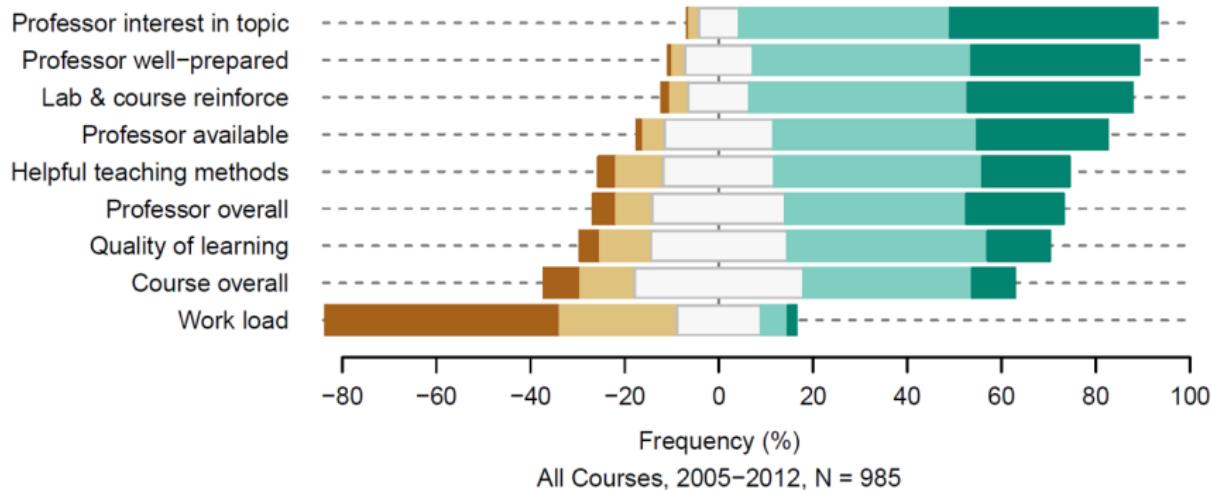
Categorical: Date (ordinal, 20 levels)



Gallery – diverging stacked bar

Quantitative: Frequency (continuous), $N_{\text{obs}} = 985$

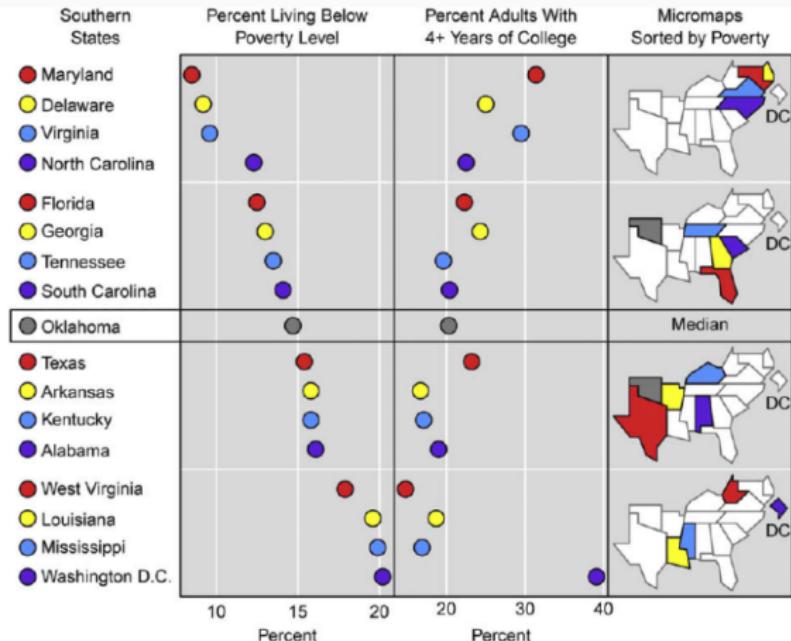
Categorical: Survey questions (nominal, 7L), responses (ordinal, 5L)



Gallery – linked micromaps

Quantitative: Percent poverty, percent college (continuous), $N_{\text{obs}} = 17$

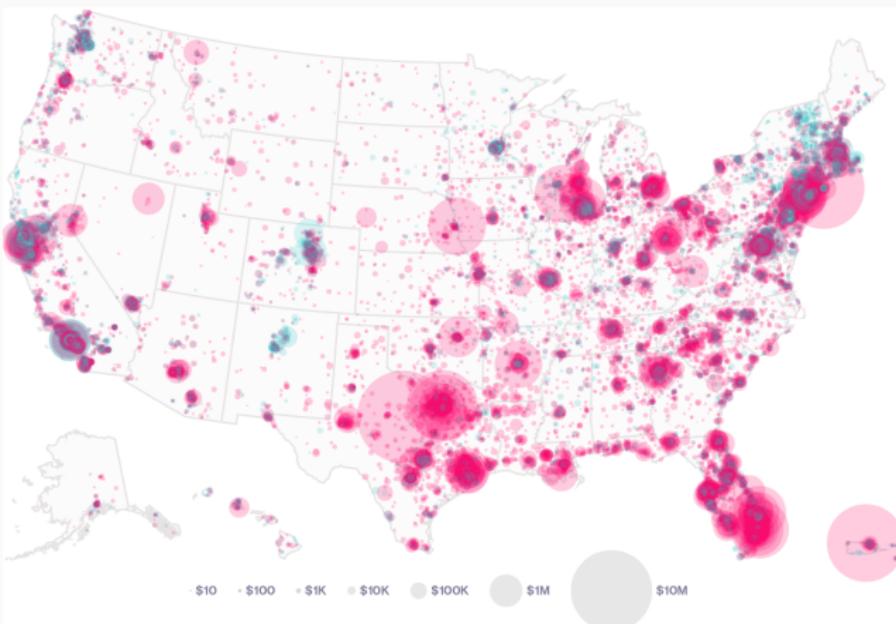
Categorical: State and geographic location (nominal, 17L)



Linda Pickle & Daniel Carr (2010) Visualizing health data with micromaps, *Spatial and spatio-temporal epidemiology*, Vol. 1, pp. 143–50. <https://bit.ly/2H967PH>

Gallery – proportional symbol map

Categorical: Contribution (ordinal, 7 levels), party (nominal, 2 levels),
ZIP code location (nominal, 42k levels), $N_{\text{obs}} = 42k$



Zach Mider, Christopher Cannon, and Adam Pearce (Sep 15, 2015) Here's exactly where the candidates' cash came from, <https://www.bloomberg.com/politics/graphics/2015-presidential-money-map/>

Gallery – dot density map

Quantitative: One dot per person, $N_{\text{obs}} = 308M$

Categorical: Race/ethnicity (nominal, 5L), geospatial location (nominal)

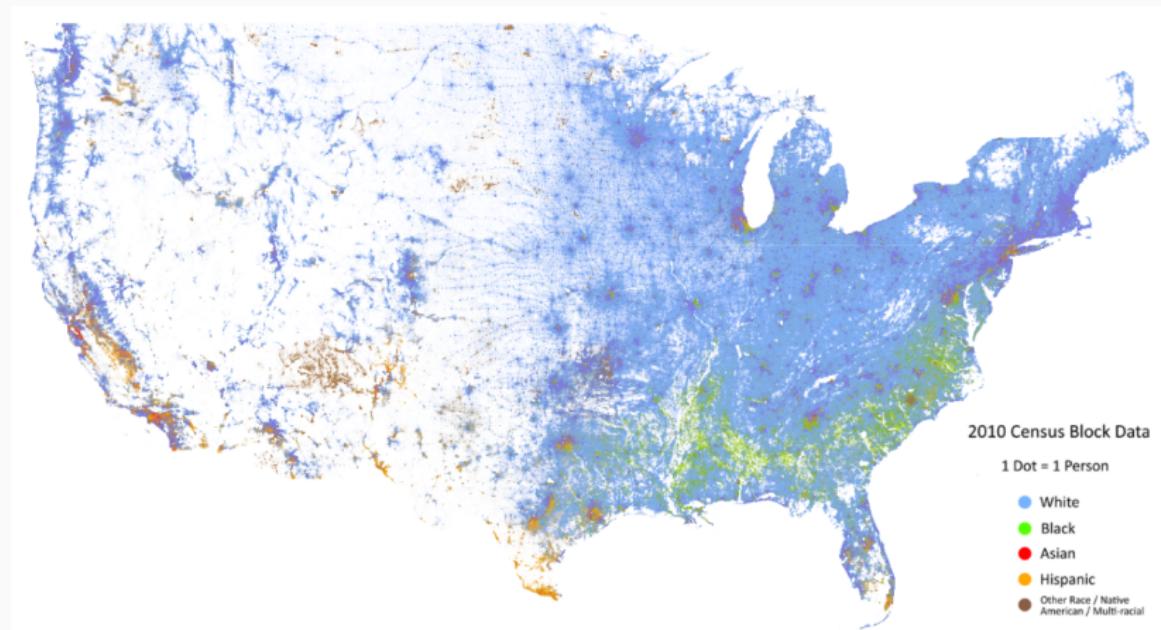
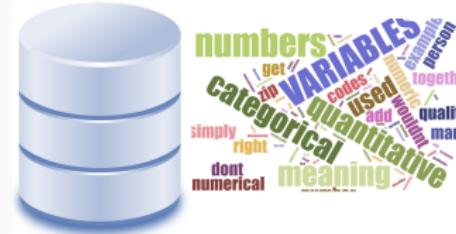
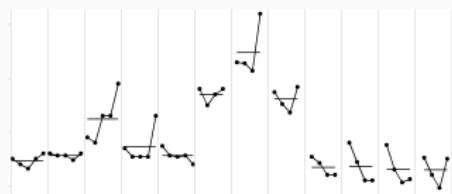


Image Copyright, 2013, Weldon Cooper Center for Public Service, Rector and Visitors of the University of Virginia
(Dustin A. Cable, creator) <http://demographics.virginia.edu/DotMap/>

Implications for the designer



Grasp the **data structure** first



Explore **data-suitable** designs



Align the **visual and verbal** arguments

References

Robbins N (2013) *Creating More Effective Graphs*. Chart House, Wayne, NJ

Unwin A (2015) *GDAdata: Datasets for the book Graphical Data Analysis with R*. R package version 0.93

<https://CRAN.R-project.org/package=GDAdata>