

36-721
Statistical Graphics and Visualization

Fall 2015
6.0 units

Course Syllabus (DRAFT)

Instructor: Jerzy Wiecezorek **TA:** ???
jerzy@cmu.edu ...
<http://www.stat.cmu.edu/~jwieczor/> ...
Office hours: ??? Office hours: ???

Class meetings: ???

Exam dates: ???

Website: <http://www.cmu.edu/blackboard/>

Prerequisites: None

Supplemental Textbooks: Not required, but suggested:

- *The Visual Display of Quantitative Information* by Edward Tufte
- *The Functional Art* by Alberto Cairo
- *Fundamental Statistical Concepts in Presenting Data* by Rafe Donahue, free at <http://biostat.mc.vanderbilt.edu/wiki/Main/RafeDonahue>
- *The Grammar of Graphics*, 2nd edition, by Leland Wilkinson, free (via CMU campus library resources) at <http://link.springer.com/>
- *ggplot2* by Hadley Wickham, free (via CMU campus library resources) at <http://link.springer.com/>
- *Interactive Data Visualization for the Web* by Scott Murray
- ...

Software:

- R, available on campus lab computers or at <http://cran.r-project.org/>
- RStudio, available on campus lab computers or at <http://www.rstudio.com/>
- D3.js, available at <http://d3js.org/>
- Inkscape, available at <https://inkscape.org/>
- Tableau, free trial available at <http://www.tableau.com/>

Course objectives

An effective graphic is a powerful tool for analyzing data and communicating insights. By tapping into the human brain's efficient visual processing centers, a good statistical graphic can quickly provide a rich understanding of the data. Upon completing this course, you should be able to:

- Produce clean, effective, self-contained statistical graphics with R
- Apply principles of human visual perception to design and critique statistical graphics
- Use the Grammar of Graphics to develop visualizations flexibly instead of relying on pre-existing templates
- Create interactive data visualizations with `shiny` or `D3.js`
- Follow the growing research literature on visualization and perception

Official course listing

“Graphical displays of quantitative information take on many forms to help us understand both data and models. This course will serve to introduce the student to the most common forms of graphical displays and their uses and misuses. Students will learn both how to create these displays and how to understand them. The class will also cover some principles of visual perception and estimation. We will start with univariate and bivariate data, looking at some commonly used graphs and, after discussing their advantages/disadvantages, then turning to more sophisticated tools. We will then explore some three-dimensional tools, group structure/clustering, and projections of higher dimensional data. As time permits, the course will consider some more advanced graphical models such as statistical maps, networks, and the usage of icons.”

Assignments, grading, etc.

TBD

Tentative schedule

Date	Topic	Readings
Week 1	Dataviz Basics for Everyone	
Mon	Introduction, history, classic examples; installing R and <code>shiny</code>	Wainer
Wed	Best practices for core 1D & 2D charts (and tables); base R plots (bar, box, histogram, scatter, line, KDE); <code>xtable</code> , <code>knitr</code>	Tufte, Few
Week 2	Dataviz Highlights for Statisticians	
Mon	Grammar of Graphics (foundation of <code>ggplot2</code> and D3.js); <code>ggplot2</code> , Tableau	Wilkinson, Bertin Wickham
Wed	Graphical principles for statisticians; binwidths/bandwidths for histograms and KDE; visual diagnostics (for regression, MCMC, ...); plotting math functions (contours, 3D densities, ...)	Donahue, Cleveland
Week 3	Harnessing the Brain	
Mon	Dataviz as ‘external cognition’; cognitive principles; preattentive processing and perceptual tasks; weaknesses of pies, 3D bars, glyphs, etc.	Ware, Cleveland
Wed	Color theory; graphic design; layout; visual style; Inkscape (or Illustrator)	Cairo
Week 4	Communicating with Your Audience	
Mon	Communicating the story; sketching; titles & annotations	Cairo
Wed	Conveying statistical ideas visually to a lay audience; demos (sampling variation; how OLS works; ...); <code>nullabor</code> for visual hypothesis tests	
Week 5	Special Topics 1	
Mon	Animation & interactive graphics; interaction design; affordances; Shneiderman’s mantra; brushing and selection; <code>shiny</code> , <code>animation</code> , <code>rgl</code> ; D3.js	Shneiderman
Wed	Maps; principles of cartography; map projections	
Week 6	Special Topics 2	
Mon	High-dim. data: multidim. scaling, proj. pursuit, ...	Cook
Wed	Trees and networks; hivemaps, ...	
Week 7	Research in Dataviz	
Mon	Doing dataviz research; usability, psych., anthro. studies; other fields (‘infoviz,’ data art, scientific illustration, ...)	Cleveland, Heer
Wed	Open Qs: displaying uncertainty in maps & rankings; mult. comparisons, survey weighting, missing data, ...	
Week 8	Finals Week	
Mon	Presentations	
Wed[?]	Presentations[?]	