PLSC 349: Visualization of Political and Social Data

Fall 2021
Preliminary syllabus subject to change

Course Personnel and Office Hours

- Instructor: Alex Coppock
  - Email: alex.coppock@yale.edu
  - Office location: Rosenkranz Hall Room 135
  - Office hours: Tuesdays, 1pm - 4pm. Sign up here: https://calendly.com/acoppock/office-hours
- Teaching assistant: Deepika Padmanabhan
  - Email: deepika.padmanabhan@yale.edu
  - Office location: To be announced
  - Office hours: To be announced
- Teaching assistant: Shikhar Singh
  - shikhar.singh@yale.edu
  - Office location: To be announced
  - Office hours: To be announced

Course Meetings

- Lectures: Mondays and Wednesdays, 3:30 - 4:20 Eastern. Location to be announced
- Sections: Locations and times to be announced
- In person attendance at all lectures and one section a week is mandatory. The sessions will not be broadcast or recorded.

Description

This course is a one-semester introduction to data visualization with a focus on political and social data. Our main textbook will be *The Visual Display of Quantitative Information* by Edward Tufte, a foundational book that explores the history of data visualization and offers a perspective on how graphs should be constructed. We will also learn from other visualization pioneers not included in Tufte's review such as W.E.B. Du Bois and Florence Nightingale. Our secondary textbook will be the ggplot2 book by Hadley Wickham, an indispensable

resource for constructing statistical graphs in the programming language R. The main goal of the course will be to learn to communicate both *what we know* and *why we think we know it* through excellent data visualization.

The course is organized around five **threads**:

- Theory of data visualization
- History of data visualization
- Wrangling and visualizing data using code
- Artistic abstraction
- Data visualizations of social and political phenomena

Books

- Tufte, Edward. 1983. The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press. This is the only book you have to buy and it is available for purchase at the bookstore. We will read almost every chapter, but not in chapter order. My suggestion is to read it through cover to cover some nice afternoon, then when we zero on particular chapters throughout the semester, you'll have the full context. I've read it dozens of times and still find things I hadn't noticed (or just forgot) every time.
- Wickham, Hadley, Danielle Navarro and Thomas Lin Pedersen. 2021. ggplot2: Elegant Graphics for Data Analysis. 3rd (in progress) ed. Springer. Free online at: `https://ggplot2-book.org/`. This book is somewhere between a software manual and friendly tutorial. It 100% works best if you "follow along" in the sense of copying and pasting the code from and running it as you read.

Prerequisites

An introductory course in statistics and probability of any kind would be sufficient. Throughout the course, we will be visualizing data summaries and measures of uncertainty about those summaries, so background familiarity with data summaries like the mean, variance, standard deviation, or the regression line and uncertainty measures like a standard error or confidence interval will be helpful. If you do not have this background but are willing to put in some extra effort to learn about them (not so tough, I promise!), you are welcome in the course. Finally, you do *not* need to know how to program in R before taking the course (my preferred coding style will be covered extensively in section), but you do need to be up for learning how.

Software

We will be using the open-source statistical software `R`. In addition to `R`, please also download and install RStudio. Pro-tip: change your computer's defaults so that .r scripts open in RStudio, not R.

- Download `R` here: `www.r-project.org`
- Download RStudio here: `www.rstudio.org`

Weekly problem sets

The weekly problem sets will reinforce visualization concepts and help you get familiar with R and ggplot. You may work with others, but everyone should write their own code and answers on their own.

Final Project

Since this class is interdisciplinary, you have wide latitude for the final project. You can write in any of the five tracks. You should pick something you care about, so keep thinking throughout the first few weeks of the course about what you're inspired to write about. The final project should be maybe 3,000 to 4,000 words, plus many visualizations.

Grading Policy:

Weekly problem sets (75%), Final project (25%).

Academic Honesty:

To ensure that you do not accidentally violate Yale's academic honesty policies, please review these sites:

- Academic Honesty: http://bit.ly/2a6uTC5
- Understanding and Avoiding Plagiarism: http://bit.ly/29VnoN1

I would like to emphasize that it is a violation of the honesty policy to:

- Copy another student's problem set, just changing a few words here and there. Collaboration is encouraged, but at some point relying too much on your partner becomes a violation of academic integrity. Most cases are clear-cut; for cases that are ambiguous, ask.
- Copy and paste whole blocks of code from your partner that you didn't have a hand in writing.
- Copy whole sentences from the internet.

It is *not* a violation of the honesty policy to:

- Copy code from websites like stackoverflow or other online forums. This is not cheating, it's learning. Part of what makes it learning is that understanding code off the internet well enough to use it usually means that you at least sort of understand it. If you do copy such code, please include a link to the forum or site where you obtained the code in the comments. This is good practice anyway, as you will often forget where code came from.
- Discuss the problem sets with your partners and compare answers.

Preliminary Schedule

Week 1

Wednesday, September 1

- Introduction: what do we see when we visualize data?
- Reading: Tufte (1983): Chapter 1: Graphical Excellence
- Reading: The syllabus, top to bottom!
- Supplemental reading: Anscombe (1973) for the origin of Anscombe's famous quartet

Friday, September 3 (Labor day make-up class

- Theory: Tufte on multidimensionality
- History: William Playfair, inventor of the line, bar, area, and pie charts
- Code: Grammar of graphics
- Reading: Wickham et al. (2021): Chapters 1 - 2 (https://ggplot2-book.org)
- Supplemental reading: Friendly et al. (2010) for history of the first known statistical graph
- Supplemental reading: Friendly and Denis (2005) for history of the scatterplot
- Assigned: Problem set 1: Introduction to ggplot2

Week 2

Monday, September 6

- NO CLASS, LABOR DAY

Wednesday, September 8

- Theory: Data, Code, Plot.
- Reading: Tufte (1983): Chapter 1: Graphical Excellence for a second time.
- Reading: Wickham et al. (2021): Chapter 3 (https://ggplot2-book.org)
- Supplemental reading: Blair et al. (2022), Chapter 2 (available here: https://book.declaredesign.org/what-is-a-research-design.html), which describes how recorded data are the result of data strategies, not "objective truth."
- Due: Problem Set 1: Introduction to ggplot2
- Assigned: Problem Set 2: The *Women's Table*

Week 2 Section

Tufte (1983): Part II: Chapter 4 - 9

Week 3

Monday, September 13

- Theory: The Lie Factor
- Reading: Tufte (1983): Chapter 2: Graphical Integrity
- Supplemental reading: Witt (2019) on how to set the range of the Y axis
- Supplemental reading: Correll et al. (2020) for some interesting variations on "broken axes"
- Supplemental reading: Huff (1954), Chapter 5 on the "Gee-Whiz Graph", though the whole book is a blast

Wednesday, September 15

- Visualizing the Pandemic
- Reading: skim these experiments on log versus linear scales: Ryan and Evers (2020); Sevi et al. (2020); Romano et al. (2020)
- Watching: Watch this video on the financial times: https://www.ft.com/video/9a72a9d4-8db1-4615-8333-4b73ae3ddff8
- Due: Problem Set 2: The *Women's Table*
- Assigned: Problem Set 3: The Lie Factor

Week 3 Section

Week 4

Monday, September 20

- : History: Florence Nightengale
- : Artistic Abstraction: Agnes Martin
- : Looking: Martin (1973): On a Clear Day (https://www.moma.org/collection/works/63682)
- Reading: Wickham et al. (2021): Chapter 13 through Chapter 16 (https://ggplot2-book.org)

Wednesday, September 22

- Dear Data and grouping
- Reading: Lupi and Posavec (2016): Introduction, plus skim through projects (on canvas)
- Reading: Wickham et al. (2021): Chapter 4 (https://ggplot2-book.org)
- Due: Problem Set 3: The Lie Factor
- Assigned: Problem Set 4: Dear Data

Week 4 Section

Week 5

Monday, September 27

- History: W.E.B. DuBois
- Reading: W.E.B. Du Bois Center (2018), pages 7-43

Wednesday, September 29

- Theory: scales
- Reading: Wickham et al. (2021): Chapter 10 (https://ggplot2-book.org)
- Due: Problem Set 4: Dear Data
- Assigned: Problem Set 5: Du Bois Redesign and Homage

Week 5 Section

Week 6

Monday, October 4

- Theory: Data-ink ratio
- Art: Mona Chalabi
- Reading: Tufte (1983), Chapter 4: Data-ink and Graphical Redesign

Wednesday, October 6

- Tufte's redesigns
- Reading: Tufte (1983), Chapter 6: Data-Ink Maximization and Graphical Design
- Due: Problem Set 5: Du Bois Redesign and Homage
- Assigned: Problem Set 6: Data-ink ratio

Week 6 Section

Week 7

Monday, October 11

- Theory: Uncertainty

Wednesday, October 13

- Topic: Visualizing national elections
- Due: Problem Set 6: Data-ink ratio

Week 7 Section

Week 8

Monday, October 18

- Art: Sol Lewitt
- Assigned: Problem Set 7: Sol LeWitt

Wednesday, October 20

- NO CLASS OCTOBER RECESS

Week 8 Section

- NO SECTION OCTOBER RECESS

Week 9

Monday, October 25

- Theory: Small multiples
- Reading: Tufte (1983), Chapter 8: Data Density and Small Multiples.
- Reading: Wickham et al. (2021): Chapter 17 (https://ggplot2-book.org)

Wednesday, October 27

- Theory: Perception experiments
- Reading: Cleveland (1985)
- Reading: Coppock (2020)
- Due: Problem Set 7: Sol LeWitt
- Assigned: Problem Set 8: Design a perception experiment

Week 9 Section

Week 10

Monday, November 1

- Model-in-data-space

Wednesday, November 3

- Basics of experimental analysis
- Due: Problem Set 8: Design a perception experiment
- Assigned: Problem Set 9: Administer a perception experiment

Week 10 Section

Week 11

Monday, November 8

- Theory: Animation

Wednesday, November 10

- : Theory: Maps
- Due: Problem Set 9: Administer a perception experiment
- Assigned: Problem Set 10: Analyze a perception experiment
- Assigned: Final project proposals

Week 11 Section

Week 12

Monday, November 15

- Theory: Networks
- Due: Problem Set 10: Analyze a perception experiment (NOTE THIS IS EARLIER THAN USUAL!)

Wednesday, November 17

- Discuss perception experiments
- Due: Final project proposals
- Assigned Problem Set 11: Bad Graph Homage and Redesign

Week 11 Section

- No section!

Thanksgiving

Monday, November 22: NO CLASS THANKSGIVING

Wednesday, November 24: NO CLASS THANKSGIVING

Week 13

Monday, November 29

- Graph criticism
- Reading: Tufte (1983): Chapter 5: Chartjunk: Vibrations, Grids, and Ducks
- Reading: Tufte (1983): Chapter 9: Aesthetics and Technique in Data Graphical Design

Wednesday, December 1

- Topic: To be announced
- Due Problem Set 11: Bad Graph Homage and Redesign

Week 13 Section

Week 14

Monday, December 6

- Discuss Bad Graph Homage and Redesigns

Wednesday, December 8

- Wrap up lecture
- Final project analysis plan due

Exam Week

Tuesday, December 21

- Final project due to canvas at 9am (our university-mandated exam time)

References

Anscombe, Francis J. 1973. "Graphs in statistical analysis." *The american statistician* 27(1):17–21. 4

Blair, Graeme, Alexander Coppock and Macartan Humphreys. 2022. *Research Design: Declare, Diagnose, Redesign*. Princeton, NJ: Princeton University Press. Under advance contract. 4

Cleveland, William S. 1985. *The Elements of Graphing Data*. Vol. 2 Wadsworth Advanced Books and Software Monterey, CA. 7

Coppock, Alexander. 2020. Visualize As You Randomize: Design-based Statistical Graphs for Randomized Experiments. In *Advances in Experimental Political Science*, ed. James N. Druckman and Donald P. Green. New York: Cambridge University Press. 7

Correll, Michael, Enrico Bertini and Steven Franconeri. 2020. Truncating the y-axis: Threat or menace? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. pp. 1–12. 5

Friendly, Michael and Daniel Denis. 2005. "The early origins and development of the scatterplot." *Journal of the History of the Behavioral Sciences* 41(2):103–130. 4

Friendly, Michael, Pedro Valero-Mora and Joaquín Ibáñez Ulargui. 2010. "The first (known) statistical graph: Michael Florent van Langren and the "Secret" of Longitude." *The American Statistician* 64(2):174–184. 4

Huff, Darrell. 1954. *How to Lie with Statistics*. WW Norton & Company. 5

Lupi, Giorgia and Stefanie Posavec. 2016. *Dear Data*. Chronicle Books. 5

Romano, Alessandro, Chiara Sotis, Goran Dominioni and Sebastián Guidi. 2020. "The scale of COVID-19 graphs affects understanding, attitudes, and policy preferences." *Health economics* 29(11):1482–1494. 5

Ryan, William H and Ellen RK Evers. 2020. "Graphs with logarithmic axes distort lay judgments." *Behavioral Science & Policy* 6(2):13–23. 5

Sevi, Semra, Marco Mendoza Aviña, Gabrielle Péloquin-Skulski, Emmanuel Heisbourg, Paola Vegas, Maxime Coulombe, Vincent Arel-Bundock, Peter John Loewen and André Blais. 2020. "Logarithmic versus linear visualizations of COVID-19 cases do not affect citizens' support for confinement." *Canadian Journal of Political Science/Revue canadienne de science politique* 53(2):385–390. 5

Tufte, Edward. 1983. *The Visual Display of Quantitative Information.* Cheshire, Connecticut: Graphics Press. 4, 5, 6, 7, 8

W.E.B. Du Bois Center. 2018. *WEB Du Bois's Data Portraits: Visualizing Black America.* Chronicle Books. 6

Wickham, Hadley, Danielle Navarro and Thomas Lin Pedersen. 2021. *ggplot2: Elegant Graphics for Data Analysis.* 3rd (in progress) ed. Springer.
URL: *https://ggplot2-book.org/* 4, 5, 6, 7

Witt, Jessica K. 2019. "Graph construction: an empirical investigation on setting the range of the y-axis." *Meta-psychology* . 5