Data Visualization

Robert A. Cooper, Instructor
ICPSR Summer Program 2020

E-mail: robert.a.cooper@dartmouth.edu
Office Hours: 5:00 - 7:00 PM daily via Zoom

Class Hours: Asynchronous

TA: Shana Scogin
Email: sscogin@nd.edu
Office Hours: 3:00 - 5:00 PM via Zoom

Course Description

Data visualization is at once both art and science. The computer revolution has generated a virtually infinite universe of data points on any topic one would care to study. As budding professionals and researchers we all need to learn how to harness information to synthesize, summarize, and communicate what we care about and know about to others. Using the statistical software program \texttt{R} (and \texttt{RStudio}), we will learn to read, tame, and tidy data so that we can analyze and visualize our data artfully and effectively.

Course Objectives

1. Students will learn to gather, organize, transform, and “wrangle” data in \texttt{R}.
2. Students will learn to analyze data in ways both useful and fit for good visualizations.
3. Students will learn the nuts and bolts of visualizing data in \texttt{R}, especially through the language of the “tidyverse” and \texttt{ggplot2}.
4. Students will learn how \texttt{R} works with other tools, applications and packages to present highly professional visualizations.
5. Students will learn to write professional projects, papers, and other with embedded \texttt{R} code and graphics in Markdown.

Highly Recommended (Not Required) Readings


**Required Video Viewings**

I will be sharing my own videos, tailored to the set of skills most useful to completing assignments as well as to broader research endeavors. Matching or approximately matching slides will attend some if not most of the videos. Videos will be made available through links and a YouTube channel.

Pending approval, we will be also be using courses from “tracks” of Datacamp lectures as supplementary lectures for the course. Largely we will draw lectures from the tracks on Data Visualization and on Tidyverse, though other videos will be used piecemeal. Assuming we are approved, you will need to sign up for a Datacamp account, which will be free through my instructor account and an invite link on Canvas. Details to follow.

**Course Policy**

The virtual classroom has the capacity to present a fun and challenging learning environment. Often you will be working alone, but there will certainly be occasion for you to work in teams, whether through Zoom meetings or Canvas discussion boards. The course adopts an asynchronous, “flipped classroom” approach. Assignments will allow you to practice the topics covered by both sets of videos. The assignments are designed to cover the material and to challenged with puzzles. Some of the puzzles are easy, others less so. Zoom virtual meetings will replace classroom and office hours time. Office hours will be held during the regularly scheduled class time, 3:00 - 7:00 P.M., split between the instructor and teaching assistant.

**Grading Policy**

There are four assignments for the two-week workshop, with each assignment accounting for 25% of the total grade in the workshop.

**Disabilities Policy**

Students requesting disability-related accommodations and services for this course are encouraged to schedule a phone/video meeting with me. This conversation will help to establish what supports are built into my online course. In order for accommodations to be authorized, students are required to consult with Services for Student Disabilities ([https://ssd.umich.edu/](https://ssd.umich.edu/), email: ssd-office@umich.edu; ) and to email me their SSD accommodation form. We will then work together with SSD if accommodations need to be modified based on the online learning environment. If students have questions about whether they are eligible for accommodations, they should contact the SSD office. All inquiries and discussions will remain confidential.

**Religious Observances**

Some students may wish to take part in religious observances that occur during this academic term. If you have a religious observance that conflicts with your participation in the course, please
virtually meet with me before the end of the second week of the term to discuss appropriate accommodations.

**Schedule**

**Week 01, 07/06 - 07/10:**

- **Day 1:** Introduction & Basic Data Work
  - Install R and RStudio
  - Basics: Objects, Vectors, & Matrices
- **Day 2:** Data transformation continued
  - Factors, Data Frames, & Lists
  - Intro to the tidyverse: filter, select, arrange, mutate
- **Day 3:** Plots in ggplot2
  - tidyverse continued: group_by, summarize, stringr functions
  - Plotting with ggplot2, Part 1: geoms and aesthetics
- **Day 4:** Exploring different geoms
  - Plotting with ggplot2: choosing the “right” geoms
  - Principles of good visualization
- **Day 5:** Fine-tuning visualizations
  - More data transformation in dplyr: pivot functions, join functions
  - Plotting with ggplot2: tricks and color theory

**Week 02, 07/13 - 07/17:**

- **Day 6:** Geospatial Data
  - Spatial data objects: sp, sf, and spatial data frame classes
  - tmap as a ggplot2 alternative
- **Day 7:** Geospatial Data
  - Mapping continued: layers, class conversions, and joins
- **Day 8:** Animated plots with ggnanimate or magick
  - Plotting with ganimate
  - Plotting with magick as an alternative
- **Day 9:** Pulling Data from the Web
  - Data Scraping with rvest
  - Introduction to Plotly and Shiny (time permitting)
- **Day 10:** Finale
  - Wrap-up: great visualizations v. the “Slides of Shame”
  - Supplementary Materials