Course Description

Data visualization is at once both science and art. In this course we learn how to harness information to tell stories about the world and to synthesize, summarize, and communicate our findings to others. Using the statistical software program R (and RStudio), we will learn to read, tame, and tidy data so that we can analyze and visualize our data artfully and effectively. We will learn principles of good visualization and basic tenets of color theory, applying these principles to our mapping of data to plots. We do this to share and frame our data analyses persuasively and without bias. Class live lectures will also be recorded and shared to Canvas.

Course Objectives and Outcomes

1. Students will learn to gather, organize, transform, and “wrangle” data in R. These skills represent the fundamental first steps of programming and data analysis.

2. Students will be introduced to the basics of functional and iterative programming in order to facilitate better wrangling and visualization.

3. Students will learn to analyze data in ways fitting for good visualizations. Largely these skills will focus on descriptive statistics, measures of central tendency and dispersion, and regression coefficients.

4. Students will learn to visualize data in R through the language of the “tidyverse” and ggplot2, from basic two-variable plots to more complicated plots, animations and maps.

5. Students will learn how R works with other tools, applications and packages (Plotly and Shiny) to present highly professional visualizations.

6. Students will learn to write professional projects, papers, and other with embedded R code and graphics in Markdown.

7. Students will learn to gather data via web scraping and a basic introduction to interacting with APIs.
Highly Recommended (Not Required) Readings

  - This book can be found free online at: R for Data Science, by Hadley Wickham and Garrett Grolemund

  - This book can be found free online at: ggplot2, by Hadley Wickham

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 the videos. These videos will be made available through Canvas links.

Course Policy

We are live and in-person once again. While I intend to execute class as a more traditional, live course, we need to maintain the ability to flexible and adapt to potentially rapidly changing circumstances. The course adopts a “resilient learning” approach, mixing synchronous and asynchronous elements as needs require. Live teaching will be accompanied by videos created by me. Documents and videos related to live teaching sessions will be available in Canvas files for those who could not attend. Zoom virtual meetings will supplement classroom and office hours time. If any of you have difficulties with your computer, a virtual machine has been set up for QSS students to run RStudio. In addition, RStudio has an RStudio Cloud environment available for your use. Weekly assignments will allow you to practice the topics covered by my live lectures and both sets of videos. Three projects in the second half of the course will allow you to exercise your skills and creativity, producing your own visualizations without any explicit instructions. The final will be “Mission Impossible” style; you will have limited time to recreate a number of plots, many of which be based on limited or incomplete data.

Attendance

You are expected to attend class in person unless you have made alternative arrangements due to illness, medical reasons, or the need to isolate due to COVID-19. For the health and safety of our class community, please: do not attend class when you are sick, nor when you have been instructed by Student Health Services to stay home. You will be able to view recordings of class in Canvas if you are unable to attend.
Student Accessibility and Accommodations

Students requesting disability-related accommodations and services for this course are required to register with Student Accessibility Services (SAS; Getting Started with SAS webpage; student.accessibility.services@dartmouth.edu; 1-603-646-9900) and to request that an accommodation email be sent to me in advance of the need for an accommodation. Then, students should schedule a follow-up meeting with me to determine relevant details such as what role SAS or its Testing Center may play in accommodation implementation. This process works best for everyone when completed as early in the quarter as possible. If students have questions about whether they are eligible for accommodations or have concerns about the implementation of their accommodations, they should contact the SAS office. All inquiries and discussions will remain confidential.

Grading Policy

Final grades will be administered on a normal, non-pandemic Dartmouth scale. Nonetheless, I will be mindful of the special circumstances of the pandemic. Weekly assignments represent the largest chunk of your grade. Three weekly projects in the second half of the course allow you to Students experiencing difficulties finishing assignments are encouraged to reach out to me as soon as possible.

- **40%** of your underlying grade will be determined by nine exercises, one of which will be dropped.
- **25%** of your underlying grade will be determined by three data visualization projects.
- **25%** of your underlying grade will be determined by a coding final in which you reproduce a series of plots with limited instructions.
- **10%** of your underlying grade will be determined by participation in class and on class Slack channels.

Mental health

The academic environment at Dartmouth is challenging, terms are intensive, and the COVID-19 pandemic, persistent though less intense, has not made things easier. There are a number of resources available to you, on campus and remote, to support student wellness. Please follow the links below to the resources available.

Office of the Undergraduate Dean ([https://students.dartmouth.edu/undergraduate-deans/](https://students.dartmouth.edu/undergraduate-deans/))
Counseling and Human Development ([http://www.dartmouth.edu/~chd/](http://www.dartmouth.edu/~chd/))
Student Wellness Center ([https://students.dartmouth.edu/wellness-center/](https://students.dartmouth.edu/wellness-center/))

Academic Dishonesty Policy

Students are responsible for understanding the academic integrity rules at Dartmouth. Explanations of integrity rules and principles can be found at [http://www.dartmouth.edu/~uja/](http://www.dartmouth.edu/~uja/). Ignorance of the Academic Honor Principle will not be considered an excuse if a violation occurs. Beyond any penalties imposed as a consequence of an Academic Honor Principle investigation, any student who is found to have cheated or plagiarized on any assignment will receive a failing grade in the
class. Details on citing sources are available at http://www.dartmouth.edu/~writing/sources. Please see me immediately if you have any questions or concerns.

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, 09/11 - 09/15:** Class Intro, Intro to R, Objects, Data

- Introduction to R: objects, classes, coercion, etc.
- Atomic vectors: numerics, integers, factors, logicals, characters, and strings
- Reading in data with base R, haven, foreign, etc.

**Week 02, 09/18 - 09/22:** Tidyverse Introduction and Basic Plotting

- *tidyverse* "verb" functions
  - filter, mutate, arrange, select, etc.
  - Scales and facets: setting domains and ranges, titles and labels, and splitting plots by variable groups
- *tidyverse* data wrangling and descriptive statistics
  - group_by and summarize
  - Line plots, histograms, boxplots
  - The “grammar of graphics”, layers, and geometries
  - Aesthetics, attributes, and mapping

**Week 03, 09/25 - 09/29:** Continuation of *tidyverse*, color theory, and theory of visualization

- Principles of visualization and color theory
  - Legends, negative space/white space, and “best practices”
  - Themes and controlling plot details
  - Hue v. value v. intensity, color palettes, RColorBrewer, color blind palettes

**Week 04, 10/02 - 10/06:** More Plots, More Skills

- Plotting principles: different levels of data
  - Scatterplots vs. bar plots, histograms, and density plots
  - factor, recode, rename, reorder and the forcats package
  - pivot_wider and pivot_longer
  - case_when, factors, dates, strings and string patterns

**Week 05, 10/09 - 10/13:** More Geoms, More Data Skills

- Discriminating between plotting choices
  - Grouped mutates and “window” functions
  - Smoothing, grouping, sums, quantiles
  - Coordinates, zoom, aspect, scales, flipped coordinates, polar coordinates, pie charts, rose plots
Week 06, 10/16 - 10/20: Intermediate plots and plotting

- **PROJECT 1**
  - Continued plot variations: non-polar coordinate plots and plots of uncertainty
    - Proportions, pie charts, waffle charts, bar plots
    - Bars, dot plots, column plots, facets, flipped coordinates, point charts

Week 07, 10/23 - 10/27: Plots Upon Plots

- **PROJECT 2**
  - Continued expansion of our plotting choices
    - density plots, histograms, kernel density estimator, histograms-as-density plots, rugs
    - boxplots, jittering, violin plots, ridgeline plots (density plots by groups)
  - Animated Plots with gganimate and magick GIFs
    - transitions, easings, entries and exits, framing, saving as GIFs, mp4s

Week 08, 10/30 - 11/03: Mapping Data onto Space

- **FINAL PROJECT**
  - Geomapping
    - Background maps, ggmap, polygons, choropleths, raster and heatmaps
    - sp objects and tmap, an excellent alternative to the tidyverse
    - Raster data and raster maps: a comparison to other mapping data forms
    - Reading in shapefiles, coordinate reference systems (CRS), merging/joining data with shapefiles

Week 09, 11/06 - 11/10: Data Scraping: rvest, rSelenium, SelectorGadget, etc.

- Grabbing data from the web vs. interacting with an API
  - APIs, access tokens
  - GET and POST requests, httr, http, rate-limiting
  - JSON and XML, XPATH v. CSS (focusing mostly on CSS)

Week 10, 11/13 - 11/17: Flex Week & Beginning of Exam Week

- TBD (last day of class: November 14th)
- Final Exam period begins November 17th

Week 11, 11/20 - 11/24:

- End of Finals period (November 22nd)