Lesson 12 Data Visualization Basics

Goals for Today

- We're going to build simple graphical visualizations within R

Readings for Today

Required

- Grolemund and Wickham, R for Data Science - Chapter 3, Data Visualizations
    - I want you to work and follow along with the code *while reading this chapter*. It will make much more sense if you do. Pay attention to *how the functions* generate iterative plot elements (i.e., what does `ggplot()` do vs. `geom_point()`).
- *Towards Data Science*, A Comprehensive Guide to the Grammar of Graphics for Effective Visualization of Multi-dimensional Data
    - This is good supplemental reading to the above chapters.

Optional

- ggplot for plots and graphs. An introduction to data visualization using R programming (YouTube)
- How to use ggplot2 in R | A Beginner's RStudio Tutorial

Important Links and Files

- The file we worked on in class is available here

Information about the Final Project

We're just now beginning Week 6, so it's time to start focusing on your final project!

The Basics

- **Due Date for Written Component**: 2021-11-12
- **In-Class Presentations**: 2021-11-16
- **Course Grade Weight**: 40%

Guidelines and Expectations

- The written component is expected to contain the following:
  - A description of a dataset and a data-driven examination of its properties.
  - Some form of a hypothesis or question you'd like to investigate with the data.
  - Statistical tests - both assumptions, outputs, and interpretations - that either confirm or fail to confirm this hypothesis.
  - Appropriate visualizations to support your claims or investigations.
  - Context and interpretation.
- The in-class presentation is expected to be a distilled, simplified version of your written component that focuses much more on cleanly presenting your findings vs. exhaustively defending your arguments/hypotheses.
- There is no hard requirement on length. That said, I want the analysis to be sufficiently complex and thorough. You should expect to have the write-up be between **5-10 pages in length**. The presentation should be somewhere around **10 minutes in length** and contain visualizations you'll present/share with the class.

Interesting Data Sets

You can choose any data set and question that you like. If you need some ideas, check out these links below. Note: **remember the rules about plagiarism**. Do not pass off others' work as your own. It's bad.

- Kaggle Machine Learning Competition Datasets

- Piktochart
- Awesome Public Datasets
- data.world
- US Census Data
- Tableau Public Datasets