

In-class Coding Exercises

Session 2

Go to the Final Projects section of the Data Wrangling With R course website. Peruse the different data sources for your upcoming midterm/final project and perform some initial exploration. After today, you should decide which dataset you'd like to use.

Getting Started

Once you've selected your preferred dataset (or maybe you can use this exploration to decide), practice what you've learned by:

1. Importing the data
2. Identifying and reviewing the codebook (if available) or website of origin
3. Learn about the data by:
 - Assessing dimensions
 - Viewing the head and tail of the data
 - Identifying the data types of each variable
 - Identifying missing data
 - Computing summary statistics for the variables
 - Check for duplicate rows or columns (You may need to google this! We haven't discussed duplicate rows/columns yet.)
4. Learn about the data visually by plotting:
 - Histograms
 - Bar charts
 - Box plots
 - Scatter plots

Where To Go From Here

By the end of class, identify which dataset you will use for your midterm/final project. Start thinking about what type of questions you would want to ask and answer of these datasets with your final project. Come to class next week prepared with the data you plan to use and the research questions you want to use to guide your data analysis. As you develop questions think about storyboarding in order to build a cohesive story that will use data to create insights which will lead to action.