# Final Project

**Due at 11:59pm on December 11.**

## Project Description

For your final project, you will work individually or in groups of two to put together a report based on web data. You will come up with an exploratory research question, obtain the data using the tools we have discussed in this class, create effective graphics using that data, perform exploratory analyses, and report your conclusions. This report should be done in an R Markdown or Quarto document, and version controlled and tracked using Git.

Your project must include the following:

- **Data from the web obtained using APIs or web scraping**. You may also use a publicly available **SQL database**. Note: This does not mean you cannot use also data from other sources! Think about how you might combine multiple data sources.
- Appropriate **data management**, such as using **dplyr** or **SQL**.
- A **visualization** that helps answer the question you are trying to answer.
- Some form of **data exploration** or **interactive graphics,** such as shiny graphics, text analysis, or clustering.

The order in which you present these might vary. One possible structure for your report might be:

- **Introduction**: What is the question you are trying to answer?
- **Data**: What is the dataset you are obtaining? How did you obtain it? What are the characteristics of the dataset?
- **Analysis**: What is a visualization that helps answer the question? What other exploration did you do to answer the questions?
- **Conclusion**: What is your conclusion based on your analysis? What are possible limitations of your analysis?

### Projects from Previous Classes

Some examples of projects from previous classes:

- Labor market-related Google search queries and unemployment figures in South Africa 2015-2019
- Topic Modeling of New York Times Article Headlines During the Spanish Flu and COVID-19 Pandemic
- Sentiment Analysis of Tweets on Georgia Abortion Law
- Pre- and Post-Debate Democratic Primary Data: Twitter, Google Trends, and Polls

## Submissions and Deadlines

### Final Project Proposal (5% of final grade)

**Due: November 12 at 11:59pm.**

Please provide a one paragraph summary of what you plan on doing for your final project. You should describe the question that you are interested in investigating, the data sources for how you will explore this, and any analysis techniques that you will use, such as descriptions of types of visualizations. Your proposal must have the following components:

- A clear **question of interest**. This must be something you can answer with the data source you propose.
- The **methodology** for how you will obtain the data to answer the question. For example, this might be an API or web scraping.
- **A clear description of the data you will use**. For example, include the unit of analysis, number of rows, number of variables, types of variables, and so on. This will vary depending on the type of data you have (e.g., web scraped text data vs. data from APIs).

### Final Presentation (10% of final grade)

**Due: December 3 at 11:59pm. Presented on December 4 in class.**

Approximately 3-5 slides total:

- A **cover slide** with your title and name(s)
- A 1-2 **description slide(s)** with text and statistics (maybe a small visualization)

- A 1-2 **visualization slide(s)**

The three slides must be in PDF format. Your presentation must be 3-5 minutes long. There are many presentations to go through. Please do not go over.

## Final Project (25% of final grade)

**Due: December 11 at 11:59pm.**

You must submit:

- The clean **PDF file of your report**. This should not show any code unless it is necessary to demonstrate something. Include the link to your GitHub repository containing the files below. You do not need to submit anything else here.

In your GitHub repository, make sure the following are included.

- The **Rmd or qmd file** used to generate the report.
- Any additional files needed to replicate your code. **It should be possible to compile your Rmd or qmd file.**