SOC 456/503 & POL S 528
Data Brief #1

Data Overview and Background

Due Date:

Week 10, Thursday, November 4, 2021 by 11:59pm

Please upload your assignment (.Rmd and .html or .pdf files) to the specified eClass assignment page. You may knit your R Markdown file into either an HTML or PDF file format. Please include your R Markdown file code (.Rmd) and the output (.html or .pdf).

Directions:

Use your project dataset and its associated codebook to complete this assignment.

All assignments must be submitted as two files.

1. A .Rmd file of your R code for each question.
2. A compiled .html or .pdf file of your code and output.

Spelling, grammar, organization, and mechanics will be graded, so make sure to proofread.

Note:

Some questions will require only R Code/Output. Other answers will require some interpretation and writing. This should be obvious within the text of each question. Remember that you can also add text to R Markdown documents between the R code chunks. Check the R Markdown Cheat Sheet and R Markdown Book for more information.

Data Brief #1

Data Brief #1 will provide information about your data and measures, along with a discussion of descriptive statistics for your key variables. It will also present your revised research question and briefly discuss your topic.

Data Brief #1 should be divided into the three numbered parts discussed below. You are welcome to also include subsections with subheadings under each part.

1. Research Topic and Question

   In Part 1, you will review your research question and topic. This part of the data brief should present a clear and concise description of the topic and research question. I expect that many of you will be presenting revised questions based on my feedback on your Topic Proposals and Literature Summaries. This question will likely require 1-2 paragraphs to fully answer.

2. Data

   In Part 2, you will discuss your data. This part of the data brief should provide a good description of your dataset, addressing aspects like data collection procedures, sample sizes, response rates, and coverage of areas and groups. A lot of this information will likely come from your codebooks. Make sure to discuss missing data within your dataset and how you plan to address it.

   With your data discussion, you should also provide R code for reading-in your data and subsetting it (if necessary for your analyses). This code can be brief. The full output here is not necessary. It just helps in case I want to try and replicate your findings.

3. Key Measures

   In Part 3, you will describe your key measures or variables. This part of the data brief should include a discussion of each variable and how you will be incorporating it into your project. For instance, is the variable an outcome or predictor variable? Will you be using it to compare groups? How is the variable coded? Did you have to recode the variable? If so, what did you do?

   What do the individual variables look like? This section should also include descriptive statistics for each variable. These should be calculated in R and presented using words, tables, and at least one graph. Useful descriptive statistics include frequencies and percentages for nominal and ordinal variables; and means, medians, and standard deviations for continuous or interval-ratio variables. Useful graphs might include bar graphs and histograms. Although there is no maximum or minimum for the number of variables to include, I expect that most students will have 5-10 variables for their project.

You will be graded on both the content and organization of the Data Brief. Content will obviously vary across project topics, but all data briefs should address each of the three areas. All data briefs should also be well-organized and clearly written. The best data briefs will describe the research topic and question, discuss the dataset and how it will help to answer the question, and provide and interpret a set of descriptive statistics for key variables. Presentation matters for this assignment, so make sure to nicely format your text, tables, and graphs. Data Brief #1 is worth 15% of your grade.