PUBPOL750 Data Analysis for Public Policy 1: Final Project

Due date: August 12 2022

Throughout the semester, you learned the basics of coding in R and exploratory data analysis in one and two dimensions. In the final project, you will practice applying these methods to a question of your interest and present the results as a research paper or a policy note.

Propose one interesting relationship between two variables, i.e., how does variable Y changes when X changes. You can work with two continuous variables, two categorical variables or one of each. You can choose more than one pair of variables if you want. For example, you could want to study how age relates to liking different political parties. In this case, you would select one independent variable (age) and several dependent variables (party ratings for each party). You could also have multiple independent variables (e.g., education and age) and one dependent variable. Feel free to experiment.

The paper's length can range from 5 to 20 pages. The paper can be written using RMarkdown or Word. You should have an introduction and a short theory section. This can either be combined, or two distinct sections. Discuss the substantive importance of your research question and your expectations of the relation(s). You should then present and discuss the dataset and the variables used. Then, you should present a univariate analysis of each variable. Then, you should present a bivariate analysis of the relation(s) of interest. Plots should figure pre-eminently, and you should comment on them. In the final section, present a discussion of your results. The paper, including the plots, should be formatted so that it does not look like a rough draft.

There are two main objectives. First you should apply the methods learned in the course. To this end, you need to load in data, clean it, analyze it, plot it and comment on the analyses. Second, you should embed this data analysis in a full research paper or a policy note. You can be more creative here. For example, once the quantitative analysis is completed feel free to add a longer discussion section where you use the quantitative analysis as the basis for something else.

Typically, social scientific research aiming at peer review publication takes the following form: introduction, literature review, data and methods, results, discussion. It's fine if you chose this format, but, here, it's also fine if you chose to do differently. For example, you could talk about a policy issue for which there is no data available (or at least not that you are aware of). In this case, it would be ok to analyze a related question (e.g. from a public opinion poll or from the CES) and then extrapolate in a longer discussion, where you formulate tentative positions, or policy recommendations based on some other source.

You can pick any dataset of interest to you. It does not have to be public policy related. The only criterion is that it allows you to answer a question of interest to you; the dataset should be interesting enough.