Main Assignment 2: Choose your own adventure with causal inference

For this final main assignment you'll work in teams to apply a dataset to a **causal** research question. The choice of dataset with accompanying research question will be up to you. Your final submission will be a report of no more than 1,000 words that introduces your research question, the data you use to answer it, your analysis where you talk about what you find, and a conclusion where you summarize some high-level implications. This will be due on Canvas May 1 by 4 pm.

Accessing and Picking a Dataset

Before you can get to the analysis, step one will be to pick a dataset. For this assignment you'll have your pick of numerous datasets that are available in the `{qss}` R package (link here). This R package is a companion to the textbook *Quantitative Social Science*. To install the package, you can run the following code:

```
devtools::install_github("kosukeimai/qss-package", build_vignettes = TRUE)
```

You can then browse the datasets available in the package by writing:

```
library(qss)
data(package = "qss")
```

To download and look at the help file for at a specific dataset from `{qss}` you can write something like the following which downloads and opens the help file for the `resume` dataset:

```
data("resume")
?resume
```

Unfortunately, longer descriptions of the data are not available in the help files. You would need a copy of the *QSS* textbook in order to see the full summary of each dataset. However,

since I have a copy, once you identify a dataset (or a few) that you want to consider using, I can help you get more details.

Instructions for Your Write-up

Once you've finalized your choice of dataset, you can begin the rest of the work of completing the main assignment. Your final paper should adhere to the following outline:

1. **Introduction (research question & why care):** Start by stating your research question as a question (e.g., What is the effect of social pressure on turning out to vote?). Then explain why this research question is important.
2. **Data and Design (what you did)**: Explain how you will answer your research question. Summarize the data you'll use, why it helps you answer your research question, and what variables in it you will use. (Don't just list these as bullet points with the raw variable names from data. Talk about them in a way that would be easily digestible for a non-technical audience.) Also, describe what kind of research design you will use to analyze the data to answer your research question. Does the data come from an experiment? Is it observational? Some datasets might be consistent with multiple analysis choices. Pick one and justify your decision. You must choose one of the four research designs we discussed for causal inference: a randomized controlled trial, selection on observables, regression discontinuity, or difference-in-differences.
3. **Analysis (what you found)**: Summarize the results of your analysis and talk about the answer that you've come up with for your research question. This section should contain exactly *three data visualizations* (no more and no less). The first should show the distribution of the outcome variable in your data. Depending on the type of variable it is, this may be a histogram or a density plot, or perhaps a bar chart. If the variable is distributed geographically, you might even consider showing its distribution with a map (here are some lecture notes I created for DPR 101 students on drawing maps). The second figure should show the distribution of the causal variable of interest (all the same details with the first data viz apply to this one as well). The third figure should show the results from your analysis. This can be in the form of a coefficient plot or a plot summarizing predictions. Whatever it is, it should clearly show the estimate of the causal effect of interest. A clear motivation and description of what's going on in each data visualization should appear in a paragraph before the data visualization. Each figure should be numbered 1, 2, and 3, respectively and have a clear description in the title. There should be no sub-figures (e.g., don't have figures labeled number 0, or 1a and 1b, etc.).
4. **Conclusion (implications)**: Talk about the implications of your findings. Don't just reiterate. Think big.

Some additional rules:

- Your submission should be in the form of word document or pdf.

- No code blocks, messages, or warnings should appear in your text.
- All sources you use should be referenced using in-text hyperlinks.
- Make sure you include all group member names in your report.