

Final assignment

Introduction to Quantitative Social Science (PLSC 30500)

Fall quarter 2021

For your final assignment, you will manipulate and analyze data, produce a written report, and give a short presentation at the end of the quarter. Your submission should produce some social scientific insight about the topic you chose. It should also display your abilities in the main skills we aimed to teach in the class: data wrangling, data visualization, regression, and inference.

Written report

Your written report should be in the form of a research report written in RMarkdown. Please submit both the Rmd file and output knitted to html or pdf. Your report should include all of the code you wrote to analyze the data and the figures your analysis produces.

Your report should also include “narrative” sections explaining what you did and interpreting the output of your analysis. In these narrative sections:

- Include a summary of your dataset, describing how many observations the dataset has, what the observations represent, and what some key variables are. Make sure to tell us what the source for your data is.
- If you are manipulating/tidying your data, you should describe what changes you are making to the data and why.
- If you are running analyses, clearly describe what analyses you are running and on what variables.

Figures should be clearly labeled and formatted.

You should aim for your report to be 10-15 pages long when knitted to pdf. Try to minimize extraneous output (e.g. unnecessary warning messages), but include content that helps a reader understand your data and what you did to it.

Try to demonstrate most of the following skills in your report:

- Creating/recoding variables
- Merging datasets (not possible for all datasets)
- Reshaping a dataset (pivot_longer, pivot_wider)
- Visualizing the data using multiple aesthetics in a single plot
- Reporting a difference in means and interpreting results, including the estimate and standard errors
- Running a regression and interpreting results, including relevant coefficients, standard errors, and p-values
- Running a few alternative specifications for the same regression, presenting the results in a figure or table, and discussing how the results change
- Including an interaction in a regression and interpreting it
- Including a non-linearity in a regression and interpreting it
- Producing more than one estimate of a standard error and interpreting it

Some suggestions:

- Choose a topic that you are interested in.
- Choose a dataset that includes variables whose relationship might be interesting, e.g. GDP per capita and human rights scores and level of industrialization. This ensures there is a regression to run.
- Consider using a replication dataset from a published article. If you do, make sure to give credit to the original paper in your report. Learn from the original replication code, but make sure your analysis is original in some way. (Please provide the replication code from the original article along with your submission.)
- Alternatively, use another publicly available dataset, like CCES, CSES, or VDem.
- When making plots in ggplot, use the labs() command to provide informative labels for your aesthetics. (By default ggplot just displays the variable name.) But otherwise don't worry about changing the defaults to make your figures beautiful (unless you enjoy doing that).
- Use your data skills to check for any errors in the data you are using. Are there any values that are impossible? If using a replication dataset, did the original analysis make any coding mistakes?

Presentation

In the last two class sessions, you will make a five minute presentation to the class. In your presentation, you should describe your data, discuss any relevant challenges in data wrangling, and explain one to two figures from your written report. We will have you send us your figures before the class in which you are presenting, and will not make use of any additional slides.