Group Project Instructions
Data and Society Fall 2021

All students in the class will complete a group project to produce an Rmarkdown report using data to analyze violence against women (VAW) indices and data sources. A large proportion of each student's grade will come from this paper and from assignments related to it, such as the in-class presentation. As such, it is important to start early on this assignment and make steady progress on it throughout the semester.

The aim of this project is to understand how social/political/economic factors are associated with the OECD's Violence Against Women data. Our primary data source comes from a 2019 OECD study of violence against women using three indicators (attitudes, laws and experience of violence). You can access these three primary outcomes here for the year 2019 and approximately 2019:

https://data.oecd.org/inequality/violence-against-women.htm#indicator-chart

These 3 variables can be downloaded as CSVs and consequently imported into a project in Rstudio. That project should be created from a group repo I will create for this assignment.

You will submit this paper as a group. There are two main components to the project:

1) Exploratory analysis of these 3 indicators and the patterns they reveal about violence against women across countries. It's also important to note limitations in the data, such as issues with what the data does or does not capture that is important to understanding violence against women. **You should produce at least 4 plots that help capture what we can learn (or not) from these data, as well as emphasize any issues of data quality**.
2) Answering an inferential question concerning what factors we think might **cause** increases or decreases in violence against women using one of the indicators mentioned above (it's fine if you use more than 1 if you want to). I expect you to have another **4 plots** that compare relationships between these indicators and other variables. You also need to perform **at least 2** regression analyses that try to model these indicators and provide a helpful measure of **uncertainty** in expressing the relationship between variables of interest.

For #1, you can use only the three measures, or if you want you can compare them to other measures of violence against women. There are other data sets available you can use as reference points if you wish, such as via the World Bank (https://data.worldbank.org/indicator/SG.VAW.1549.ZS), the United Nations (https://unstats.un.org/unsd/gender/vaw/), as well as efforts by independent researchers (https://journals.sagepub.com/doi/full/10.1177/08862605211037424). Note that you will need to merge these datasets in R if you want to use them.

For #2, you need to select between 1 and 3 variables that you think might matter for explaining the variation in violence against women across countries. These can be economic, political, social or even other types of factors if you prefer. You can draw these factors from what is available in the OECD data (i.e., select other data types from https://data.oecd.org/). You can also use other sources, such as World Bank, IMF, UN and WHO data. I also recommend searching for data on Github and also in Harvard's Dataverse (https://dataverse.harvard.edu/), a repository of data produced by scholars.

Your research question should take the following format:

How do [variables] appear to reduce or increase violence against women?

The specific analyses you will need to do involve using the provided data to examine associations via plotting methods and also basic statistics (i.e., producing tables of averages, checking correlation coefficients, etc). The aim here is to understand what associations exist and what might explain them. Your regression models should predict one or more of the violence against women indicators with these variables and any other "control" variables you think might be potential confounding variables.

Even though it is unlikely you will be able to rule out all threats to inference, I do want you to be open and clear about what limitations there are to these analyses and do what you can can to account for them. It is more important to be transparent about issues you cannot resolve than it is to present findings that seem more rigorous.

Format

Your Rmarkdown file, which is included in the repo as template.Rmd, needs to follow a research paper format. This means you need to have 1) a clear thesis/hypothesis (which answers the question given to you above), 2) a set of clear points that support your thesis/interpretation, and 3) an essay with clear structure and supporting paragraphs. Please read the following guide for writing political science research papers so you can learn the correct structure of the paper:

https://sitesmedia.s3.amazonaws.com/politicalsci/files/2012/08/Guidelines-on-Writing-Research-Papers.pdf

Note that you will need to start with doing some literature review on your topic, then develop your own theory that best explains the outcome/answers the research question. However, I do not expect you to do a thorough or exhaustive investigation of what other scholars have done--this is still an exploratory research paper, and it is fine to follow intuitions.

There is no minimum page length for this assignment, but I would ask that the paper remain under 7,000 words. The word count should be included right by the title of the paper. **You may**

use any citation format, but citations should be included in-text (by footnotes or parenthetical citations) for any sources you use. Please do not use end notes. To save time, you can consider using Rstudio's Visual Editor mode which can import citations from Zotero (see this page for more info: https://rstudio.github.io/visual-markdown-editing/).

The Rmarkdown file uses the distill article format (you may need to install the R package distill to knit the file). For more information about this format and what styles/features you can change, see this site: https://rstudio.github.io/distill/basics.html.

Your final submission will be an Rmarkdown file committed and pushed to your Github repo, which I will then download and knit. The code in the file must be properly documented and clear so I can follow what is going on.

Paper Steps:

My recommendation is to first write up a 2-3 page outline for your project/paper outlining your argument and what analyses you want to conduct. This outline should identify 1) the research question you have chosen, 2) a list of resources culled from scholarly and other works (may also be datasets), and 3) a brief overview of how you want to do the analysis / any initial findings you have about the topic. This outline should be as specific as possible about the possible analyses you want to conduct.

Consider the following resource for a helpful way to build your outline:

https://writingcenter.fas.harvard.edu/pages/outlining

You will also need to do an in-class presentation during the last two weeks of class (Dec. 9th - 14th). I will provide more guidance closer to the due date, but these will be 10-minute summaries (i.e., fairly short) of your research paper for the class. These slides should have appropriate visualizations and other results to share with the class.