Prof. Rachel Skillman
POLI 170A: Applied Data Analysis for Political Science
Due: July 7, 2022 @ 11:59pm PST

Problem Set 1

Please write down all the names of students that you collaborated with on this problem set. Remember that all submitted work must be your own. Submit a single document (either a pdf or word doc) with your answers to the Gradescope upload link. The assignment is out of 100 points total and worth 10% of your final grade.

CONTENT REVIEW

Question 1. What are the 5 components that your research project will be asked to address? These are the 5 components we have talked about in each class that allow us to break research (both our own and the research of others) into manageable and accessible parts. Give a brief, 1-2 sentence description of each component. (15 points, 1 point for identifying each component and 2 points for each component description)

Question 2. Identify the 5 components in the following abstract. Remember that some components have multiple pieces (like identifying the broader importance of a topic or briefly reviewing relevant literature). (10 points, 2 points per component)

From Steven Liao's *Chinese Foreign Real Estate Investment and Local Voting in U.S. Presidential Elections* (2021)[1]

Abstract. Despite the increasing globalization of housing markets, little is known about its political implications. This study investigates whether rising Chinese investments in U.S. homes influenced local voting in recent U.S. presidential elections. Building on pocketbook/sociotropic voting and nativism theories, I develop hypotheses on the electoral consequences of foreign real estate investment through greater home demand and equity, improved local economies, and changing neighborhoods. Using difference-in-differences designs that combine a unique shock to Chinese capital outflows in 2013 with county-level measures of local attractiveness to Chinese investments, I find that greater exposure reduced Democratic vote shares in both the 2016 and 2020 presidential elections. Furthermore, an initially larger white population strengthened this effect while a larger college-educated population weakened it. In contrast, local equity gains, housing competition, or economic strength did not systematically influence the effect. Together, the results appear more consistent with the pro-conservative effects of nativism.

Question 3. How do we get from important but vague topics to specific, testable research questions? Let's work through the steps. (15 points total)
1) Give an example of a broad concept that is of interest to social scientists. (2 points)
2) Give an example of a property/variable that you could measure about that concept. (2 points)
3) Give an example of a way you could operationalize that property/variable so that other reasearchers could measure it the same way. (5 points)
4) What is the unit of analysis in the operationalization of the variable you specified? (2 points)
5) Give an example of values that the operationalized variable could take on. (2 points)
6) What type of measurement is your hypothesized variable? (Nominal, Ordinal, Continuous) (2 points)

---

[1]https://www.stevenliao.org/uploads/2/5/6/9/25699716/frei-elections.pdf

Question 4. Loading data into R. Using the dataset "RuralPopData_POLI170A.xlsx" (from the World Bank's World Development Indicators[2]) on Canvas, please complete the following steps. As your answer to this question, please copy and paste the code you used to complete each step. Remember to annotate/comment each line of code so that you and other researchers can understand your process and easily replicate your results. You can use the GUI functionality to accomplish some of the steps but you should "hard code" the code output from the GUI command into your R file. (20 points total)

1) Open a new R Script and set your working directory to be the folder where you have saved the data. (2 points)
2) Install the `readxl` package if you have not already. Load the `readxl` package. (2 points)
3) Load the dataset into R. When doing so, change the variable type of "Rural population (% of total population)" from character to numeric. This will automatically convert missing data (which the World Bank format records as ..) into `NA` (3 points)
4) Make a new version of the dataset called "rural_pop" (2 points)
5) Remove the old version of the dataset (2 points)
6) Rename the variables to be: (a) country; (b) iso; (c) rural_pop_percent (3 points)
7) Report the mean of rural_pop_percent for all observations (3 points)
8) What type of variable is rural_pop_percent? (3 points)

<div align="center">

FINAL PROJECT PROPOSAL

</div>

Each answer about your proposed final project only needs to be a few sentences.

If you are having trouble finding data for your final project, or coming up with a research question, check out these sources. Make sure to consider whether data is available to answer your research question before deciding on a question.

- https://databank.worldbank.org/source/world-development-indicators
- https://www.bea.gov/
- https://data.census.gov/cedsci/
- https://data.gov/
- https://electionstudies.org/data-center/
- https://cces.gov.harvard.edu/
- https://www.afrobarometer.org/data/data-sets/
- http://www.asianbarometer.org/data/data-release
- https://www.latinobarometro.org/latContents.jsp
- https://europa.eu/eurobarometer/surveys/browse/all
- https://www.kaggle.com/datasets

Question 5. Are you working with a partner on the final project, or by yourself? If you are working with a partner, please write their name here. (0 points)

Question 6. What is the research question you will answer for your final project? Make sure you indicate the significance of your proposed research topic. Why should we care about your phenomenon or relationship of interest? (5 points)

Question 7. What is the theory that you will argue answers your research question? Cite at least one academic article where another scholar has discussed your topic. If you have trouble finding academic literature, check out https://scholar.google.com/. (10 points)

---

[2]https://databank.worldbank.org/source/world-development-indicators

Question 8. What is the testable hypothesis your final project will assess to provide evidence that supports or does not support your theory? Make sure your hypothesis meets the criteria for a good hypothesis. (10 points)

   Note: Remember, unless you want to do a more advanced project, you should plan to do a cross-sectional analysis for your final project. That means you should plan to select 1 time period for both your independent and dependent variable (preferably the same time period unless you have a theoretical reason to choose different ones, e.g. if you expect there to be a lag in the effect of a policy). Alternatively, you can average data from multiple time periods into one aggregate measure of your variables of interest. In order to figure out the appropriate time period(s) for testing your hypothesis, you should carefully consider the data generating process and data availability and make an informed decision about the time span that makes sense for your proposed question and answer.

Question 9. What is the independent variable in your hypothesis? The dependent variable? What types of measurements are these variables? (5 points)

Question 10. What data will you use to test your hypothesis? You must include URLs of websites to indicate that you have looked for your data and verified that it exists and you may be able to get access to it. For 1 point of extra credit, include a screenshot of your ***downloaded*** data (either independent or dependent variable) opened in Excel or in R. You cannot get extra credit for simply screenshotting data on a website. (10 points with 1 possible extra point)