Deliverable 2: Proposals with summary statistics

Due by Friday, Nov. 4th at 11:59pm

Submit a knitted R Markdown file with 6 sections (including output, and where applicable):

1. **State research question(s) and motivation**
   - What do you hope to learn from the proposed analysis? What are the policy implications of your potential results?
2. **Policy background**
   - What policy context should we know? Make sure we understand what your "treatment" entails, and what factors may have contributed to the treatment (in the absence of random assignment)? Describe the nature of your policy/treatment variation.
     - Some examples: you might be looking at cross-sectional variation between individuals or entities in exposure to a time-invariant policy, time variation shared by all individuals or entities, or differential exposure within entities or subgroups over time.
   - What are the (potential) mechanisms linking your treatment or policy-relevant variable(s) of interest to outcomes? It may seem obvious to you, but make sure we know why might you expect to see any relationship between your policy/treatment variables and outcomes.
3. **Data description**
   - Describe your data sources, how was your input data generated, the unit of observation and population represented by your sample(s).
   - Describe any key data restructuring or subsamples of the input data that you plan to focus on.
     - For example, do you need to consider restricting your input data to a subset of the sample? Do you plan to restructure/aggregate your input data to explore relationships between variables with a different unit of analysis than in your input data?
   - Describe how you will measure key variables (outcomes, policy/treatment variables) and any control variables that may be particularly important to account for (explain why).
4. **Preliminary exploratory analysis**
   - Show descriptive stats to summarize the distribution of your key X and Y variables, using tables and/or charts. One of the most important goals of this deliverable is to make sure you have enough sample variation for you to work with, so you should prioritize this preliminary EDA and report results, however preliminary.
     - Depending on your research design, this could mean describing describe the *sample variation* in your candidate X and Y variables over time and/or between relevant groups.
     - Keep in mind whether you will be analyzing panel or cross-sectional data and the nature of the variation in your policy/treatment variable. This will inform the EDA tools that make the most sense for you.
5. **Empirical strategy**

- Carefully describe the explanatory analysis you are planning on working towards. What is the policy variation you'll be exploiting in any regression analysis to come and potential threats to internal validity? Another way to think about this: what sort of comparisons will you be making with your planned regression analysis?
    - An example: you might be constructing an entity-year panel and exploiting policy variation over time within entities by using entity fixed effects. What threats to internal validity, such as omitted variable bias, remain to be addressed?
- Outline key steps to prepare the data for analysis (data cleaning, recoding, merging, appending, aggregation, etc.).
- Highlight key issues or limitations you need to address – be specific about how you plan to solve programming obstacles or fill critical data gaps!

6. **Appendix**

- Include your coding work to-date for importing, cleaning, recoding, restructuring and joining input data sources.

Tips:

- use code chunks to generate and present summary statistics
- don't clutter your write-up w/code (you can include more in an Appendix)