In this assignment, you will write the next section of your paper, data and methodology.

Before you begin, please make sure you revise your literature review and put it before your data and methodology section. When you revise your lit review, please think about making a transition towards data in the end. Some of you already did a good job on that. You will only be evaluated based on your data and methodology.

After you complete your assignment, please send me to file electronically and bring your binder to class. You do NOT need to print this out. I will grade it and insert the graded printout with my comments in your binder and return it to you.

The section is consisted of two parts: theory and data/methodology. The requirement for each part is described as the following.

**Economic Theory**

If you have **not** included theory in your literature review, you need to spend roughly a page or so explaining your theory. There are several options for this.

1. Some of you may choose to update your literature review by incorporate some theory into it. If this is what you do, mark these sections **RED** in the lit review.

2. You can also start a new section on theory.

In either case, you can simply expand the theory part on your proposal.

**Data and Methodology**

1. A brief description of the data sources. You should mention the source (name of survey/dataset, agency, etc), the time frame, the geographic region, and the sample size. Also briefly describe the dataset. If you have multiple sources, you need to specify every data source. Also, you will have
to put the correct citation in the reference section. Also explain whether this is cross-sectional, time series, or panel data.

2. Justification. Explain why you use this data source. For example, why this year, why this region? Why you only look at states, metropolitan areas, counties, G7 countries, or swing states? Why is this dataset particularly suitable to answer your research questions?

3. If you combine the sources, you should also include how you actually construct the dataset. In this case, you may experience some attrition (losing some observations). If this is the case, explain how this may affect your research outcome if at all. Also state the final number of observations (sample size).

4. The variables. You need to talk about what variables you are going to use.
   a. Define each variable if it is not obvious like unemployment rate or GDP growth rate. For example: property crime rate, time spent playing outside, energy consumption, etc.
   b. Specify the unit of each variable. (percentage, dollar, years, KW/Hour, etc) If it is unclear, you will be asked during your final presentation.
   c. For dummy variables, describe how they are set up. (ex. Female=1 and 0 otherwise, ethnicity variables, etc)

5. The type of regression. Example: OLS or probit or other.

6. The actual regression equation, dependent variable, independent variable(s) and other controls such as regional dummies.

7. The expected (theoretical prediction) of coefficients to independent variables. You should be very careful because in some cases, the sign may be ambiguous due to contradicting theories. If this is the case, make sure you explain both sides of the theory.

Looking forward:

The section following this one would be the summary statistics of your variables and your regression outcome. So you will need to get the data ready.

Your regression specification may also change, so you may need to come back and revise this section in the future.