Proposal Guidelines (100 points):

(20 points) Summarize and motivate your proposed project

In Part A of this project, I wanted to see if there was a correlation or relationship between overall homelessness in the United States with energy consumption, production and expenditure. I was also curious to see if states that consumed, produced, and expended more on energy had higher GDP per capita than other states. My motivation behind this study was the rising prices for electricity and gas in Ann Arbor and the growing population of homeless people in Ann Arbor. I became curious to see if there is a relationship that could possibly explain this.

My datasets:
https://www.kaggle.com/adamschroeder/homelessness
This dataset shows every state (categorized by state abbreviations), the count of homeless people in that state, and the year of which this number has been recorded. This also includes a description of their status in homelessness. My second dataset is called United States energy, census, and gdp 2010-2014.
https://www.kaggle.com/lislejoem/us_energy_census_gdp_10-14
This dataset shows each state (categorized by state abbreviations), the total number of consumption, production and expenditure in the millions of BTU and dollars per year.
https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1
My third dataset is called US household income statistics. I will be grabbing this data from the US census bureau for the year of 2014.

(80 points) Proposed analysis

1. An exploratory analysis on which states consume, produce and expend the most energy. I want to see if the specific differences on these energy categories by creating bar graphs for each.

2. I want to see the overall summarization of data for income for each state. I will do this by finding the mean, median, IQR and show other data statistics to see key differences for each state.

3. I also want to graph a timeseries to see if homelessness in the United States has increased over the years my dataset provides. This is to see if there is a trend in the growing homeless population.

4. I am curious to see the number of homeless people recorded in each state against each other. I will create a bar graph of each state to show that number in order to see the difference in homeless people count for each state.

5. Similarly, I want to see a time series trend in my three energy variables. If there is a growing trend (which I assume there is) for energy, then this can help show a correlation of some sort.

6. I will then plot a scatter plot with homelessness and energy production to see if there is a relationship. I will also use a line of best fit in order to show if there is some sort of correlational data.

7. Similarly, I will do the same for energy production with a line of best fit and scatterplot.

8. I will do the same analysis for energy expenditure as well to show if there is or there is not a relationship between homelessness and expenditure.