In assignment, you will conduct regression analysis on your research topic. Before you can run regressions, you need to make sure your data is ready.

Note that we’ll start with preliminary analysis of your data and regression outcome. At this point, you do not need to write a full-blown section of the regression results.

Questions:

Answer Questions 1-6 in a word processor. Print out your answer and submit it in your binder.

1. Tabulate your dependent and independent variables. You should present the summary statistics in clear and readable table formats. If you have a large number of categorical variables, make sure you start a separate table and report the frequencies.

2. Is there anything in your summary statistics that requires special attention or explanation? If so, please make sure you explain them.

3. Present a scatter plot on your dependent variable and main explanatory variable.

4. Report your regression output in the table format presented in class. Make sure you present at least 3 different specifications of your choice and explain their differences in a short paragraph. You may need to try numerous permutations before you settle on these regressions. For example, non-linear (age squared), adding year and state dummies, different measures of income/wage, etc.

5. Interpret your main results with a paragraph or two. The “main results” refer to the variable you care about the most (spending on SAT scores, single mom status on education attainment, unemployment on crime/drug use, etc). You may want to talk about the sign as well as the magnitude.

6. Use a few sentences to talk about the coefficients of other variables.

7. Email me your .do file and your data file. If you are using Excel, email me your spreadsheet.
The first part of the assignment (therefore this section in your paper) would be a descriptive summary of all (dependent and independent) variables. You should present the summary statistics in a clear and readable table formats. If you have a large number of categorical variables, make sure you start a separate table and report the frequencies. In your table, you may want to highlight some stylized facts. Do not simply “report” everything that the audience can see. (for example, the mean of GDP per capita is $46,025 and the median is $50,000) Instead, focus on the general trend (if any) and the variables that deserves explanation (e.g. over 50% of the single mothers in the sample are African American).

The second part of the assignment is the regression part. Before you write your first draft, you will need to try a few different specifications. For example, there may be a few variables that you’re not sure of, and you’d like to run various regressions by including different variables. It is perfectly normal to run many regressions before you settle on one (or a few).

In addition, you may also run into numerous technical problems. The most common one would be multi-collinearity. If Stata or Excel automatically drops some of your variables, this may be the problem. When this occurs, you will need to make sure that you understand the reason behind this and modify your regression equation.

When you write your regression part, you should always start by saying the expected o