In advance of the upcoming exam, I would like you to start thinking about your project. For this project, you will be analyzing a data set of your choice to look at the relationship between variables. I have separately attached a list of potential datasets and you can start looking for appropriate ones during lab on Thursday. You are allowed to talk to others in the class/your TA about this question, but should not be planning the same project as someone else you speak to and should follow the class rules on integrity as listed on the syllabus. You will not be allowed to work with others on the rest of the exam. I will provide more details about that at the end of the week or early next week.

The below question will appear on Exam 1 and be worth 30 points:

Please describe the data that you plan to use for your project in paragraph form. In describing this data, please answer the following questions (Please note that you will not be making any actual visualizations here or doing any actual analysis, just saying what you plan to do):

1. What is your proposed topic? Why is it important? (5 pts.)
2. Where did you find the data you plan to use? Why is it appropriate for this project? (5 pts.)
3. What are your main variables of interest? Why are they appropriate for this project? Please describe 3-4 variables here and include at least one response variable and several explanatory variables. (10 pts.)
   1. Please try to find a continuous numerical variable or an ordinal categorical variable with multiple categories that you can use as your response variable.
4. What would be some appropriate visualizations that you could make of this data using ggplot? Why would these be appropriate given the data you have? (10 pts.)