Research Milestone 1: Research Proposal

Your submission should have two files:

1. **PDF file**. A 500-word narrative describing your proposed research project. Guidelines are provided below.
1. **R Script.** Upload the R script that loads your data files into R and generates the output included in your writeup.

Proposal Guidelines

1) Introduction and background.

- Discuss your motivation for picking your topic.
- Include at least at least 1 figure as motivation from a news article or report. Explain how it is related to your topic.

2) Data Overview.

- Provide a URL to where the data can be downloaded (csv or excel file is preferred)
- Load the data into Rstudio and include a screenshot of the data frame in your environment
- Provide a brief description of the data set:
    - unit of observation
    - time frame
    - number of observations

4) Variables
- Select at least four categorical variables and four numeric variables in your dataset that you plan on using for analysis
- Use the slice_head() function to show the first ten values of your selected variables. Copy the output to your writeup.
- For each categorical variable:
    a) Provide a description of the variable.
    b) The selected categorical variable should have no more than 4-5 levels (categories). If the selected variable has more than this amount, explain how you will combine/eliminate levels.
- For each quantitative variable:
    c) Provide a description of the variable

d) If the variable selected currently isn't quantitative, describe how you will convert it to a quantitative variable

- Bonus Variables:
    - Describe any geographic variables in your dataset (longitude/latitude, county or state names) that you can use for creating maps.
    - Describe any free response variables that you could use to conduct a text analysis (descriptions, comments, reviews)

3) Objectives and goals.
- State an objective/hypothesis that you wish to explore with your data.
- Describe at least four relationships you would like to explore in order to meet your objective.
    - The relationships described must include at least five of your selected eight variables

Data Requirements

- At least 1000 observations
- At least 4 categorical variables
- At least 4 quantitative variables

Not enough categorical/quantitative variables?

If you don't have enough categorical variables in your dataset, you can propose creating categorical variables from a quantitative variable.

Some examples:
- Create a categorical variable for decade (1970s, 1980s, 1990s) if you have a year numeric variable
- Create a categorical variable for income ranges if you have an income numeric variable (0-20k, 21k-40k, 40k and above)
- Create a region variable if you have state data (North, South, East, West)

If you don't have enough numeric variables in your dataset, you can propose creating numeric variables from a variable that contains free response text.

Some examples:
- Song lyrics: counting the number of times a song uses the word "love"
- Yelp review comments: count the number of times a review mentions the word "expensive"