Project Milestone 1 Requirements and Rubric

These are the requirements for the first project milestone. Use the template Rmd file and submit a knitted version to the assignment. Use this Rmd file as a beginning and build upon the work you do in this file for the next milestone and the final project. You can change aspects of your project (points 2-5 below) as you work on developing your project. Be aware that major changes can result in lost time.

1. **Project Metadata:** Change the file name to represent your project name. Fill out the top section of the Rmd file:
   —----------
   title: the working title of the project- must be descriptive, not generic.
   author: full first and last names of both partners as they appear in spire.
   date: yyyy-mm-dd
   output: pdf_document
   —----------
   +1 Title descriptive, not generic.
   +1 Names correctly entered.
   +1 A date was entered.
   +1 Project file was renamed.

2. **Description:** A paragraph that describes the project that must include: 1- a description of the domain, e.g. finance, health, astronomy, climate, etc., you are considering, and 2- one or more guiding question(s) you will investigate in the project. Use specific language.What trends will you explore/visualize, what outcomes will you predict/associate with input variables? Mention if your data involves a period of time/when it was recorded.
   +2 Area of interest clearly and adequately described.
   +1 Area of interest not clearly or incompletely described.
   +2 Guiding question(s) clearly and adequately described.
   +1 Guiding question(s) not clearly or incompletely described.

3. **Data Description and Links:** Identify at least one data set you have discovered that you could use for the project. Each data set should be clearly described: what are the variables, their data types, and what do they represent? For example, if a variable is a “metric” of some sort, explain what the metric is measuring. If there are variables you are not using in your analysis, you do not need to describe them, but mention the variables you are not using. Provide a link to each data source you plan to use. Describe how you will get the data and in what format it is received. Example: downloaded as csv file. If you need to perform any additional steps to acquire the data, such as if you are executing a query of some sort to get the data, include a description of that process.
   +1 All variables used in the analysis named and data type mentioned.
   +1 What each variable measures is provided.
   +1 Steps to acquire the data described.

+1 Link to each data set provided.

4. **Data Exploration:** Provide evidence you have downloaded and explored the data. Download the data and display a summary of a sample (not the whole thing) of the data or the part of the data you are planning to work with. A plot is also acceptable. The summary/plot should be produced in R. You might be working with more than one dataset, so show a summary of (some of) both data sets. ***Always state the size of any data you present including the number of observations and variables of interest.***
   +1 Evidence data set downloaded and loaded in R (such as a read.csv statement).
   +1 Summary or plot of variables of interest displayed in R.
   +1 Size of the data: number of observations and number of variables mentioned.

5. **Analysis Plan:** Mention at least ***three statistical techniques*** from this course that you plan to use in your analysis ***and state how they will be used*** in your analysis. This can be hypothesis testing, ANOVA, regression, classification modeling, clustering, time series or any other technique we covered or will cover in the class.
   +1 Three techniques mentioned.
   +1 Description of how each technique will be used to explore guiding questions.