BSDS 100: Intro to Data Science with R
Assignment 6

by Abbie M. Popa (University of San Francisco)

Directions: For all questions in this assignment, write complete sentences and fully answer any question that is asked, and use `R` to answer each question. Provide all `R` code and solutions by *knitting* your final RStudio file into a single file named `your_name_CA6.pdf`. Late assignments will automatically have 10 points deducted, if submitted within a week of the due date. Assignments submitted after the answer key is posted will not be accepted and will receive zero points.

1. (2 pts) What is the advantage of storing (and loading) data as a `.csv` file rather than a `.xlsx` file?

2. (6 pts) Name three different types of data sets that you may be interested in loading into `R` and describe at least one function that can be used to input each of the types of data set. (2 pts per example)

3. Answer the following questions.

   (a) (2 pts) Create a data frame that has the following four columns:
   - `Numbers`: the numbers 1 through 50, where each number is repeated twice in a row. (e.g. 1 1 2 2 3 3 ...)
   - `Logicals`: a vector of length 100 whose jth entry is TRUE if the jth entry of `Numbers` is even and FALSE if the jth entry of `Numbers` is odd.
   - `Rev.Numbers`: the vector `Numbers` but in reverse order.
   - `Weirdness`: the sum of `Logicals` and `Rev.Numbers`.

   (b) (2 pts) What are the data types for each of these columns?

   (c) (2 pts) Describe why the variable `Weirdness` is an `Integer` variable.

   (d) (2 pts) Save this data frame to any chosen directory as a `.RData` object named `MyDataFrame`.

   (e) (2 pts) Remove the data from your workspace, then reload `MyDataFrame` and print out the first 6 entries in each column of the data frame.

4. Load the `Airport` data that we investigated in the Input Output Lecture. Then write code to answer each of the following:

   (a) (2 pts) What are the names of the variables in this data set and what are their data types?

   (b) (2 pts) What is the mean and standard deviation of the longitude of these airports?

   (c) (2 pts) What is the minimum and maximum latitude of these airports?

   (d) (2 pts) Which airport has the minimum latitude? The maximum latitude?

   (e) (2 pts) Add a new observation (row) to this data frame. Add whatever you would like as the new input, but make sure that each variable maintains its original data type. (i.e. if the longitude variable is numeric, make sure that it remains numeric after the new observation is added).

   (f) (2 pts) Save your new data frame as a `.csv`, a `.txt`, and a `.RData` file.