Homework Assignment 1

Your Names Here

Assigned: Jan 18, 2018, Due Jan 26, 2018 11:59PM

This homework is due by 11:59PM on Thu Jan 26. To complete this assignment, follow these steps:

1. Download the `HW1.Rmd` file from Canvas.
2. Open `HW1.Rmd` in RStudio.
3. Replace the "Your Name Here" text in the `author:` field with names of the students in HW team.
4. Supply your solutions to the homework by editing `HW1.Rmd`.
5. When you have completed the homework and have **checked** that your code both runs in the Console and knits correctly when you click `Knit HTML`, rename the R Markdown file to `HW1_YourNamesHere.Rmd`, and submit on Canvas. (YourNameHere should be changed to your own names.) You should submit both the RMD file and its html output.
6. Run your code in the Console and Knit HTML frequently to check for errors.
7. You may find it easier to solve a problem by interacting only with the Console at first.
8. Note - you can insert an R block for writing your R code by either clicking the **insert** button or pressing Control-Alt-I.

Problem 1: Vectors

Let's first start working with vectors. I am providing you with the following vector:

```
somedata <- c(10.4, 9.8, 5.6, 4,9, 2,4, 3.1, 7.2, 6.4, 8.8, 12.5, 21.7)
```

I am asking you to do the tasks below. For each of these, you should insert a new R code block and type your code there.

What is the 5th value in vector `somedata`

Show all values except the 5th value in vector `somedata`

Calculate how many values there are in the vector `somedata`

Calculate the difference between the maximum value and the minimum value in the vector `somedata`

Show all values in `somedata` more than 10

Check whether the value 7.2 exists in the vector `somedata` or not. Code should output `TRUE` or `FALSE` depending upon whether the value exists or not in the vector.

Problem 2: Factors

Assume that we have collected address data from several students. We are storing the state of residence information in a vector named `student.state`. I have created the vector for you below.

```
student.state <- c("MI", "IL", "NY", "MI", "NY", "HI", "IL", "MI", "MI", "NY")
```

I am asking you to do the tasks below. Again do them in separate R code blocks.

Convert the contents of student.state into factors (look up command `as.factor`)

Count the number of levels in the factor created above (Useful command: `levels`)

Figure out which state(s) appears in the list only once (this is a hard one!)

Problem 3: Data frame basics

We will continue working with the nycflights13 dataset we looked at last class. Please be sure to keep the data file in the same directory as the RMD file.

First - read the data into an object named `nyc`.

```
# Write code below to import nycflightsjan13 dataset into an object named nyc
```

I now would like you to answer the following - all in their own separate R code blocks.

1. Let's first filter this data to small portions. For example: How many flights where there by United Airlines (code: UA) on Jan 12th 2013?
2. Lets focus on Arrival Delay. First thing we want to figure is: a) What was the average arrival delay in Jan 2013? and b) What was the maximum arrival delay? c) What was the median arrival delay. Note: You will need to make sure that you take care of NA values using the na.rm = TRUE option as we did in class.
3. Lets see if all airlines are equally terrible as far as flight arrival delays are concerned. For this question you will have to make sure that airline column is coded as a factor.

a) Calculate average arrival delays by airline (Hint: look up the command `tapply`)

b) Draw a Bar Plot of Average Arrival Delays for all the Airlines (Hint: command for making a Bar Plot is simply `barplot`)

c) Which airline has the highest average delay? Which airline has the smallest average delay? Are there airlines that actually have negative average delay?

That's it. Once you are done, make sure everything works and knits well and then you can uplaod the RMD flile and the html output to Canvas.

> I am **always** available to help you with this and other assignments. Feel free to send emails with your questions, post on Piazza or drop by my office. When asking questions it works best to just send me R commands as plain text in email rather than attaching the whole RMD file. -Sanjeev.