Assignment 1

Due 02/01/2019 by beginning of class

Instructions:
You may work with others on these problems but you must turn in your own work. You may type your solutions in any way you like (LaTeX, Markdown, Office, etc...) as long as you present your work clearly and in an organized way.

Unless otherwise specified, you must hand in a printed copy of your work at the beginning of class.

Problems:

1 Complete the course "Introduction to R" from Datacamp (`https://www.datacamp.com/courses/free-introduction-to-r`). You are only expected to complete the free part of their course, and is designed to give you practice with writing basic R commands and familiarity with R types especially dataframes. Download and print the statement of accomplishment and hand this in for credit.

2 Download the toy data from the website: `https://cgrudz.github.io/teaching/stat_757_2019/assignments/19_02_01/data_19_02_01.zip`

The comma delimited file data.csv contains information from the event data recorder in a car driven by two drivers, Alan and Barbara. Records are sampled every 30 seconds during a trip, and include information on the location, speed, and heading of the vehicle. Each record also includes a field called "event", which denotes whether the driver exceeded a pre-defined acceleration threshold at any point during that 30-second period. We assume that these events are correlated with risky driving behavior.

The file data-dictionary.txt contains a brief description of each variable included in the data set.

The data is too long to analyze by hand, but luckily we can use R to compute the following:

2.1 Identify how many total "events" Allen had. Explain how you extracted and calculated the total number of Alan's events data from the dataframe.

2.2 Identify how many events Allan had relative to the *total number of measurements* of their trips. Explain how you extracted the total number of measurements of Allan's trips from the dataframe.

2.3 Identify how many events Allan had relative to the *total number of trips* they took. Explain how you extracted the total number of trips Allan took from the dataframe.

2.4 Find the mean and median speed of all of Allan's trips. Explain how you computed these summary statistics.

2.5 Repeat 2.1 but for Barbara.

2.6 Repeat 2.2 but for Barbara.

2.7 Repeat 2.3 but for Barbara.

2.8 Does Allan or Barbara seem to be a more risky driver? Why? Extra credit if you provide additional quantitative analysis of the data.

Hint 1: There are *many ways* to solve problem 2 above, but I haven't shown you exactly how to do this yet. It is your job to do some research on how to access different sub-vectors of dataframes. R has abundant documentation available — try some resources on my website or search for your own.

Hint 2: Did you know that the type "LOGICAL" can be used to index/ extract slices of vectors? Try setting up a statement based on what you want to be "TRUE".