**BANA7038 Homework 1**

**Write a report to answer the following questions. Please organize the report to avoid printing large tables or plotting large figures. Limit the length of your report. Try to be as concise as possible. Clearly mark the question numbers in your report (e.g., in large font).**

**When answering these questions, lay out the following:**

**1. What you are trying to do? (Your goal)**

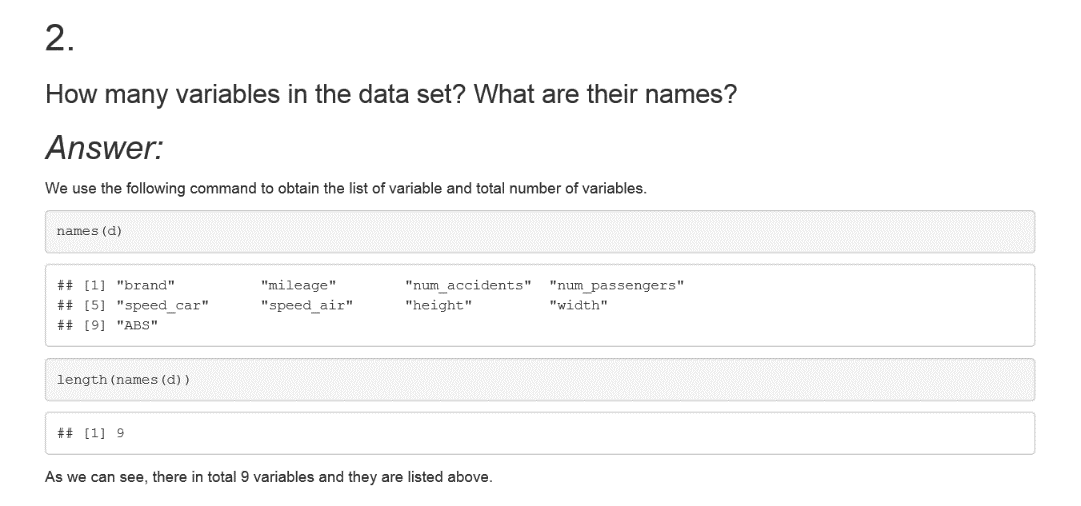
**2. The R code (how do you realize it?)**

**3. The R output**

**4. Your observations (What do you observe from the output?)**

**If necessary, you can repeat 1 through 4 for many times to answer one question fully.**

**Here is a sample answer that you can mimic.**



1. Import the CSV file ‘car\_r.csv’ using the function “read.table()” or “read.csv()”. Where to find the instruction on how to use the functions?

2. How many variables in the data set? What are their names?

3. How many observations in total? How many observations for Ford?

4. Calculate the mean for each of the car parameters (measures). Please also report the corresponding standard deviation.

5. Obtain the histogram for each of the car parameters.

6. Is there any missing value in the data set? If yes, which variable? What is the proportion of missing values?

7. Calculate the relative speed of the car (defined as = speed\_car + speed\_air, where speed\_car is always positive and speed\_air can be positive or negative). What is the average relative speed of the car?

8. How many cars have mileage less than 40000? How many cars have height less than 5? Please delete those observations (i.e., cars whose mileages are less than 40000 and cars whose heights are less than 5) and delete the observations that contain NAs from the original data set to form a new data set.

9. Divide the new data set (as obtained in Step 8) into three subsets: Ford, GM and Toyota. That is, create three new data sets containing only Ford, GM, and Toyota cars, respectively.