NYC DATA SCIENCE ACADEMY

NYC Data Science Bootcamp

Introduction to R Part I

***** Save all your code to *yourname*.R and push it to the homework Github repository.

Note: you may need to pull from origin before you push it to Github.

Question #1:

Use the TimesSquareSignage.csv in the homework folder and import it into R. Then check the following features of the dataset:

1. The number of observations and the number of variables
2. The type of each variable
3. How many missing values are there in the dataset?
4. Which rows (people) have missing value? Which columns (variables) include missing value?

Question #2:

From the Time Square dataset, we'd like to extract specific information about advertising in Midtown Manhattan. Obtain the following data frames and save them in CSV files:

1. Observations from Upper Broadway
2. Observations with greater-than-average square footage
3. The name, address and location of the top observations in terms of total square footage

Question #3:

1. From your RStudio, import the built-in data by running data(cars)
2. Print the first 5 lines from *cars*.

3. Randomly generate a vector as long as the the number of rows in *cars*, and have elements *NY, CA* or *CT*. Call the vector *state.*

4. Add *state* to the data frame *cars* as a new column. Again name the column *state.*

5. Create a new column *ratio* whose value is the ratio *dist/speed.* Then compute the average and standard deviation.