Tutorial 2 STQD6214

1. It is believed that the pressure of the tires affects the gas mileage of a car. 14 samples were collected consisting of two variables, the pressure (in pounds per sq. inch) and mileage (in thousands).
   a) Using the `data.frame()` function, create a new data frame named `TIRES` using the following data:

```
x <- c(30,30,31,31,32,32,33,33,34,34,35,35,36,36)
y <- c(29.5,30.2,32.1,34.5,36.3,35,38.2,37.6,37.7,36.1,
       33.6,34.2,26.8,27.4)
```

   b) Using `names()`, change the variable names in the data frame such that the variable `x` is changed to `pressure`, and the variable `y` is changed to `mileage`.
   c) Print out the first 4 rows.
   d) Print out the rows in the dataset with mileage is greater than 35.
   e) Print out your current working directory.
   f) Using the `write.csv()` function, save the data frame into a CSV file. Verify the file.

2. A teacher collected a sample of 10 students to study the relationships between students' test scores, their IQ, how many hours they study each week, their GPA, and their class. The file "TestScore.csv" contains the dataset.
   a) Using `read.table()` or `read.csv()` functions, import the dataset into R, and assign it to a data frame named `TestScore`.
   b) What are the available variables in the data frame?
   c) Print out the 8^{th} row of the data frame.
   d) Using the `describe()` function in `psych` package, print out the summary statistics for the data frame.
   e) Calculate the mean, median, and variance for the GPAs of the students.
   f) Calculate the first and third quartiles for the GPAs of students, and its interquartile range.
   g) Calculate skewness and kurtosis for the GPAs of the students. Is the data skewed to the left or to the right? Is the data platykurtic or leptokurtic?
   h) Print out all rows containing students from Class B. Assign these values to another data frame.
   i) Calculate the mean, median, and variance for the IQ of students in Class B.
   j) Calculate the range for the IQ of students in Class B.
   k) Using `quantile()`, what is minimum value for the top 15% of students' study hours?
   l) Find the mode for the study hours of the students.

3. To study the relationships between sales price, appraised value, and improvement value of residential properties in four neighbourhoods (Hyde Park, Cheval, Hunter's Green, and Davis Isles) in the city of Tampa, Florida, a sample of 176 observations were collected. The file "property_sales.csv" contains the observations in the study.
   a) Import the dataset into R.
   b) Print out the name of all the available variables in the dataset.
   c) Print out the first 6 rows of the dataset.
   d) Using the `table()` function on the neighbourhood variable, print out how many observations there are for each neighbourhood. Which neighourhood has the highest frequency?
   e) What is the highest and lowest sales price in the dataset? Calculate the value for its range.
   f) Calculate the mean, median and variance for the sales price of all the properties.
   g) Calculate the mean, median and variance for the sales price of the properties in Davis Isles.
   h) Calculate skewness and kurtosis for the sales price in Davis Isles. Is the data skewed to the left or to the right? Is the data platykurtic or leptokurtic?
   i) Calculate the first, second, and third quartiles for the appraised land value for the properties in Cheval. Calculate its interquartile range.