Lab Report Assignment

1. Download movies dataset from the blackboard. Read the dataset in R. Assign it to a variable name ‘movie’ [6 marks]
2. Perform a query to retrieve the following. [6 marks]
   - First 10 lines of the data
   - Summary of data
3. Visualize each of the attribute using histogram and fill in missing value in the dataset. If the data is normally distributed, use mean to replace missing value. If the data is skewed, use median to replace missing value. [36 marks]
   a. Score
   b. Budget
   c. Votes
   d. Budget
   e. Gross
   f. Runtime
4. Compute the descriptive statistics for the score and budget [16 marks]
   a. Min
   b. Mean
   c. Median
   d. Standard Deviation
5. Print the name and genre of movies which score is higher than mean score [6 marks]
6. Visualize the genre and runtime using stripchart. Which genre has the highest runtime? Please include proper header and label [10 marks]
7. Convert the genre data to numerical. [5 marks]
8. Calculate the correlation between the numerical data in the dataset. Which pair of data has highest correlation? [15 marks]