CS544 Module 3 Assignment (Self Graded)

Part 1) 10 points

Use the *primes* (UsingR) dataset. Use the diff function to compute the differences between successive primes. Show the frequencies of these differences. Show the barplot of these differences.

Part 2) 15 points

Use the *coins* (UsingR) dataset. Do not use explicit loops for any calculations. Do not hard code the denominations in the solution. The solution should work for any denominations.

a) How many coins are there of each denomination?
b) What is the total value of the coins for each denomination?
c) What is the total value of all the coins?
d) Show the barplot for the number of coins by year.

Part 3) 15 points

Use the *south* (UsingR) dataset.

a) Show the stem plot of the data. What do you interpret from this plot?
b) Show the five number summary of the data. Calculate the lower and upper ends of the outlier ranges. What are the outliers in the data?
c) Show the horizontal boxplot of the data along with the appropriate labels on the plot.

Part 4) 10 points

Use the *pi2000* (UsingR) dataset.

a) How many times each of the digits 0 to 9 occur in this dataset?
b) Show the percentages of their frequencies.
c) Show the histogram of the data.

Part 5) 20 points

Suppose that a football (NFL), basketball (NBA), and hockey (NHL) games are being shown at the same time. Consider the two-way summarized data shown below showing the preferences of men and women what sport they wish to watch.

Gender	Sport NFL	NBA	NHL
Men	25	10	15
Women	20	40	30

a) Using cbind, create the matrix for the above data.
b) Set the row names for the data.
c) Set the column names for the data.
d) Now, add the dimension variables Gender and Sport to the data.
e) Show the marginal distributions for the Gender and the Sport.
f) Show the result of adding margins to the data.
g) Show the proportional data separately for Gender and Sport. Interpret the results.
h) Using appropriate colors, show the mosaic plot for the data. Also show the barplot for Gender and Sport separately with the bars side by side. Add legend to the plots.

Part 6) 10 points
Use the midsize (UsingR) dataset.

a) Show the pair wise plots for all the variables.
b) Provide at least 4 interpretations of the results.

Part 7) 20 points
Use the MLBattend (UsingR) dataset.

a) Extract the wins for the teams BAL, BOS, DET, LA, PHI into the respective vectors.
b) Create a data frame of five columns using these vectors. Use the team names for the columns
c) Show the boxplot of the data frame.
d) Provide at least 5 interpretations of the results.