**CS544 Module 3 Assignment**

Abraham Cepeda Oseguera

U75425818

**Part 1**

1. Barplot of the frequencies for the last digit.

Chart, bar chart

Description automatically generated

1. Barplot of the frequencies for the first digit

Chart, bar chart, histogram

Description automatically generated

1. What inferences do you draw from these two plots?

In the last digit bar plot there is an almost equal frequency for the number 1, 3, 7, and 9. While there is only one case for both 2 and 5. On the other hand, the first digit bar plot show a similar frequency for numbers from 1 to 9. However, there is a slight decline in frequency as the first digit increases, which reflects that as the number increases there is less chance of being a prime number.

**Part 2**

1. For which state were the highest number of quarters produced by each mint? For which state were the lowest number of quarters produced by each mint?

|  |  |  |
| --- | --- | --- |
| Mint | Max | Min |
| Denver Mint | Connecticut | Virginia |
| Philly Mint | Oklahoma | Iowa |

1. What is the value of the total coins in dollars?

$ 8,699,400,000

1. Produce the following barplot from the data using the R barplot function with the data for the two mints as a matrix. Write any two striking inferences you can observe by looking at the plot.

Chart, histogram

Description automatically generated

1. Show the scatter plot of the number of coins between the two mints. Write any two inferences you can observe looking at the plot.

Chart, scatter chart

Description automatically generated

1. Show the side-by-side box plots for the two mints. Write any two inferences for each of the box plots.

Chart, box and whisker chart

Description automatically generated

1. Using R code, what states would be considered as outliers for each of the two mints. Use the five number summary function to derive the outlier bounds

**Denver Mint**: Connecticut, Virginia

**Philly Mint**: Connecticut, Massachusetts , Maryland, South Carolina, New Hampshire, Virginia, New York, North Carolina

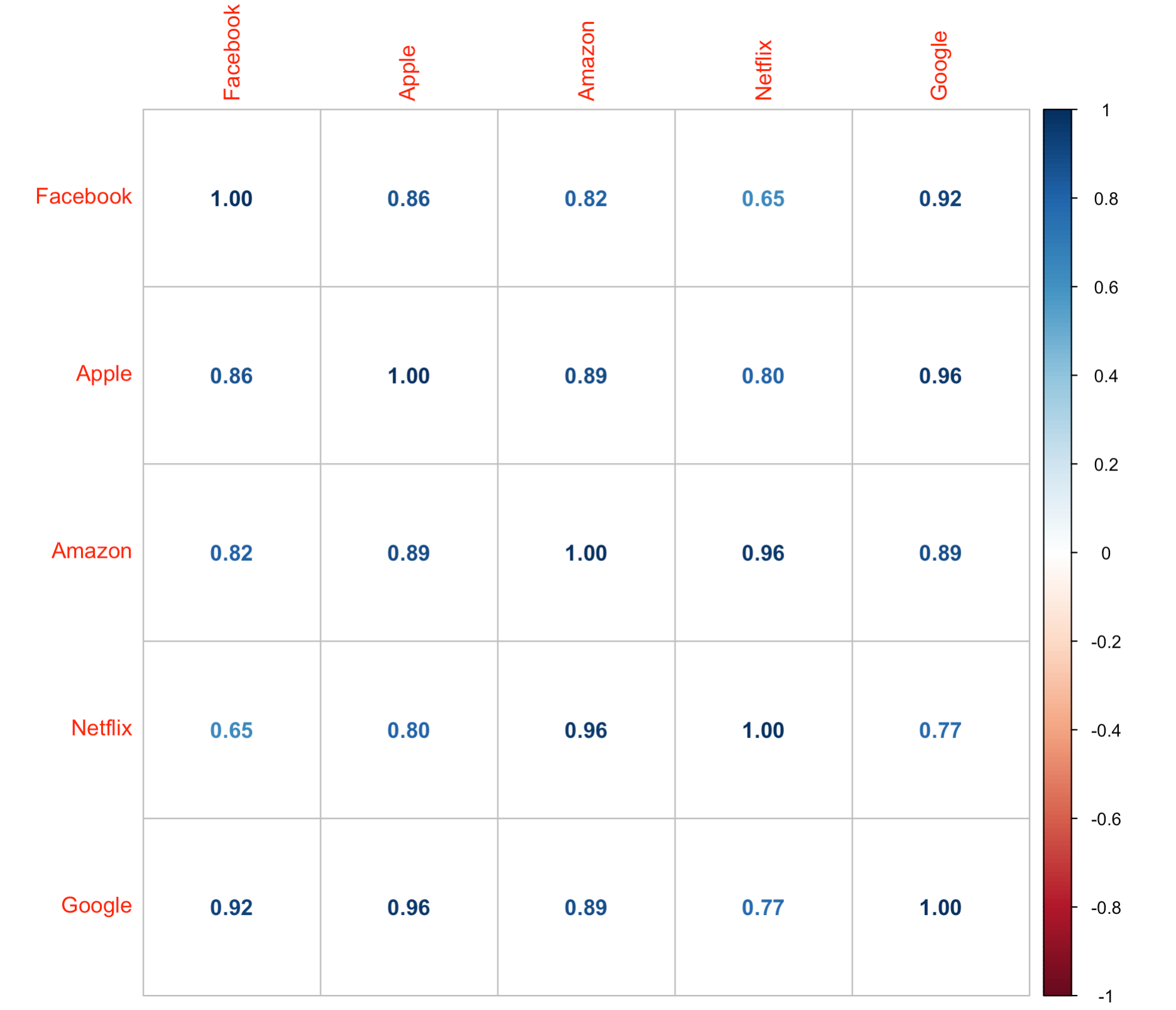
**Part 3**

1. Show the pair wise plots for all the 5 stocks in the dataset in a single plot.

A picture containing text, keyboard

Description automatically generated

1. Show the correlation matrix for the 5 stocks in the dataset.



1. Provide at least 4 interpretations of the results.

By seeing the pairwise plots of the stocks, there appears to be a general linear relationship among the stocks. Some relationships appear to be stronger than others, for example, the relationship of Netflix and Google is stronger than the relationship between Netflix and Facebook. After analyzing the correlation matrix, it is clear that there is a general strong relationship between the stock’s prices of all companies. The strongest relationship is between the tech giants Google and Apple with a correlation of 0.96.

**Part 4**

1. Show the default histogram of the student scores. Save the result of the histogram into a variable. Using only the counts and breaks property of this variable, write the R code to produce the following output. The code for the following output should not refer to the individual scores.

Text

Description automatically generated

1. Using the breaks option of the histogram, show the histogram and the custom output as shown below so that students in the range (70,90] get an A grade, (50,70] get a B grade, and (30-50] get a C grade. The code for the following output should not refer to the individual scores.

Text

Description automatically generated

**Extra credit 1**

1. Construct parallel boxplots of this set of data.

Chart, box and whisker chart

Description automatically generated

1. Do the data indicate that females or males had the greater mean number of texts? Explain in detail (Shape, Outliers, Center, Spread; Conclusion).

**Shape:** The shape of the female boxplot appears to be right skewed since the distance from Q1 to the median is much smaller than the distance between the median and Q3. The shape of the male boxplot appears to be normally distributed.

**Outliers:** In the female graph, there is a clear outlier a 1098, while the male does not have any outlier.

**Center:** The center of the female data appears to be between 170 and 200 with a median of 175 but a mean of 326. The center of the male data is located between 170 and 190 with a median of 183 and a mean of 174

**Spread:** The range of the female data is from 15 to 1098 with and interquartile range of 228. While the range of the male data goes from 3 to 337 with and interquartile range of 201.

**Conclusion:** After analyzing both boxplots, it is clear that on average females send more texts than males. This occurs even if the extreme outlier in the female data is removed.

**Extra credit 2**

**Chart, histogram

Description automatically generated**

1. The approximate median age of the Liberian population falls in which of these intervals: 0–4, 15–19, 30–34, 40–44? Explain

The median age of the Liberian population would likely fall in the range of 15-19 given that an important portion of the population lies between the ages of 0 and 14.

1. Explain why it is impossible to calculate the mean age of either population.

Given this data, it would be impossible to calculate the mean because the frequencies are given in ranges of age. Therefore, we cannot know the frequency of each age within each range.

1. Which country has more children younger than 10 years of age? Explain

Canada = 0.9+0.9+1+0.9 = 3.7 millions

Liberia = 310+270+315+270 = 1,165 thousands = 1.165 million

After doing the sum of the ranges 0-4 and 5-9 of both male and female of each country, we can conclude that Canada has significantly more children under the age of 10.

1. Does the population pyramid indicate that Canadian men or Canadian women live longer? Explain.

The population pyramid of Canada indicates that women live longer than men because from the age of 60 onwards the population of each range of each is larger in women than in men.

1. In 2010, Liberia had recently come out of a civil war with the extensive use of child soldiers. How is this visible in the population pyramid?

There is a significant decrease in the age range of 15-19 compared to age range of 10-14 and 20-24, which suggest that sadly thousands of children in the age range of 15-19 died during the civil war.