> ## CS 544 Module 3 Assignment

> ## Hao Wu

> ## Part 1)

> df <- read.csv("http://people.bu.edu/kalathur/datasets/myPrimes.csv")

> ##Question a

> table(df$LastDigit)

1 2 3 5 7 9

306 1 310 1 308 303

> barplot(table(df$LastDigit),

+ col = "black",

+ xlab = "Type", ylab = "Frequency")

> ##Question b

> table(df$FirstDigit)

1 2 3 4 5 6 7 8 9

160 146 139 139 131 135 125 127 127

> barplot(table(df$FirstDigit),

+ col = "black",

+ xlab = "Type", ylab = "Frequency")

> ## Part 2)

> us\_quarters <- read.csv("http://people.bu.edu/kalathur/datasets/us\_quarters.csv")

> ##Question a

> Denver\_max\_State <- us\_quarters[which.max(us\_quarters$DenverMint),1]

> Denver\_max\_State

[1] "Connecticut"

> Philly\_max\_State <- us\_quarters[which.max(us\_quarters$PhillyMint),1]

> Philly\_max\_State

[1] "Virginia"

> Denver\_min\_State <- us\_quarters[which.min(us\_quarters$DenverMint),1]

> Denver\_min\_State

[1] "Oklahoma"

> Philly\_min\_State <- us\_quarters[which.min(us\_quarters$PhillyMint),1]

> Philly\_min\_State

[1] "Iowa"

> ##Question b

> TotalCoin <- (sum(us\_quarters$DenverMint) + sum(us\_quarters$PhillyMint))/4

> TotalCoin

[1] 8699400

> ##Question c

> a <- cbind(us\_quarters$DenverMint,us\_quarters$PhillyMint)

> barplot(a ~ us\_quarters$State,

+ col = c("blue","grey"),

+ legend = c("DenverMint","PhillyMint"),beside = TRUE)

> ##Question d

> colnames(a) <- c("DenverMint","PhillyMint")

> plot(a)

> ##Question c

> boxplot(a)

> ##Question d

> Denver <- fivenum(us\_quarters$DenverMint)

> us\_quarters$State[which(us\_quarters$DenverMint < (Denver[2] - 1.5\*(Denver[4] - Denver[2])) |

+ us\_quarters$DenverMint > (Denver[4] + 1.5\*(Denver[4] - Denver[2])))]

[1] "Connecticut" "Virginia"

> Philly <- fivenum(us\_quarters$PhillyMint)

> us\_quarters$State[which(us\_quarters$PhillyMint < (Philly[2] - 1.5\*(Philly[4] - Philly[2])) |

+ us\_quarters$PhillyMint > (Philly[4] + 1.5\*(Philly[4] - Philly[2])))]

[1] "Connecticut" "Massachusetts" "Maryland" "South Carolina" "New Hampshire" "Virginia"

[7] "New York" "North Carolina"

> ##Part 3)

> stocks <- read.csv("http://people.bu.edu/kalathur/datasets/stocks.csv")

> head(stocks)

Date MSFT AAPL GOOG FB AMZN TSLA

1 2021-01-01 216 130 1787 269 3162 816

2 2021-01-08 211 128 1740 246 3127 845

3 2021-01-15 223 136 1891 273 3307 845

4 2021-01-22 237 136 1863 265 3238 835

5 2021-01-29 240 137 2062 266 3331 850

6 2021-02-05 242 134 2096 270 3262 812

> ##Question a

> pairs(stocks[,2:7])

> ##Question b

> round(cor(stocks[,2:7]),2)

MSFT AAPL GOOG FB AMZN TSLA

MSFT 1.00 0.90 0.95 0.68 0.64 0.71

AAPL 0.90 1.00 0.79 0.54 0.59 0.73

GOOG 0.95 0.79 1.00 0.85 0.67 0.47

FB 0.68 0.54 0.85 1.00 0.66 0.05

AMZN 0.64 0.59 0.67 0.66 1.00 0.34

TSLA 0.71 0.73 0.47 0.05 0.34 1.00

> ##Question c

> summary(stocks)

Date MSFT AAPL GOOG FB AMZN TSLA

Length:53 Min. :211.0 Min. :120.0 Min. :1740 Min. :246.0 Min. :2978 Min. : 572.0

Class :character 1st Qu.:242.0 1st Qu.:129.0 1st Qu.:2262 1st Qu.:299.0 1st Qu.:3248 1st Qu.: 664.0

Mode :character Median :276.0 Median :141.0 Median :2584 Median :330.0 Median :3334 Median : 732.0

Mean :276.5 Mean :141.5 Mean :2524 Mean :321.2 Mean :3349 Mean : 786.7

3rd Qu.:302.0 3rd Qu.:149.0 3rd Qu.:2876 3rd Qu.:344.0 3rd Qu.:3463 3rd Qu.: 845.0

Max. :341.0 Max. :178.0 Max. :3014 Max. :378.0 Max. :3731 Max. :1230.0

> apply(round(cor(stocks[,2:7]),2),2,mean)

MSFT AAPL GOOG FB AMZN TSLA

0.8133333 0.7583333 0.7883333 0.6300000 0.6500000 0.5500000

> ##Question d

> cm <- round(cor(stocks[,2:7]),2)

> cm

MSFT AAPL GOOG FB AMZN TSLA

MSFT 1.00 0.90 0.95 0.68 0.64 0.71

AAPL 0.90 1.00 0.79 0.54 0.59 0.73

GOOG 0.95 0.79 1.00 0.85 0.67 0.47

FB 0.68 0.54 0.85 1.00 0.66 0.05

AMZN 0.64 0.59 0.67 0.66 1.00 0.34

TSLA 0.71 0.73 0.47 0.05 0.34 1.00

> for(i in 1:ncol(cm)){

+ Top3 <- as.matrix(head(sort(cm[i,-i],decreasing = TRUE),n=3))

+ ##Use sort method get only one row for each stock, but without the target stock

+ ##Use head method to get Largest 3 correlation and store it in Top variable

+ cat("Top 3 for Stock",colnames(cm)[i], "\n",rownames(Top3),"\n",Top3[,1],"\n")

+ }

Top 3 for Stock MSFT

GOOG AAPL TSLA

0.95 0.9 0.71

Top 3 for Stock AAPL

MSFT GOOG TSLA

0.9 0.79 0.73

Top 3 for Stock GOOG

MSFT FB AAPL

0.95 0.85 0.79

Top 3 for Stock FB

GOOG MSFT AMZN

0.85 0.68 0.66

Top 3 for Stock AMZN

GOOG FB MSFT

0.67 0.66 0.64

Top 3 for Stock TSLA

AAPL MSFT GOOG

0.73 0.71 0.47

> ##Part 4)

> scores <- read.csv("http://people.bu.edu/kalathur/datasets/scores.csv")

> head(scores)

Student Score

1 S1 77

2 S2 53

3 S3 57

4 S4 59

5 S5 59

6 S6 63

> ##Question 1

> hist(scores$Score,breaks = 8)

> hist\_score <- hist(scores$Score,breaks = 8)

> hist\_score

$breaks

[1] 35 40 45 50 55 60 65 70 75 80 85

$counts

[1] 3 4 10 13 17 27 13 8 3 2

$density

[1] 0.006 0.008 0.020 0.026 0.034 0.054 0.026 0.016 0.006 0.004

$mids

[1] 37.5 42.5 47.5 52.5 57.5 62.5 67.5 72.5 77.5 82.5

$xname

[1] "scores$Score"

$equidist

[1] TRUE

attr(,"class")

[1] "histogram"

> for(i in 1:length(hist\_score$counts)){

+ cat(hist\_score$counts[i],"students in range","(",hist\_score$breaks[i],",",hist\_score$breaks[i+1],"]","\n")

+ }

3 students in range ( 35 , 40 ]

4 students in range ( 40 , 45 ]

10 students in range ( 45 , 50 ]

13 students in range ( 50 , 55 ]

17 students in range ( 55 , 60 ]

27 students in range ( 60 , 65 ]

13 students in range ( 65 , 70 ]

8 students in range ( 70 , 75 ]

3 students in range ( 75 , 80 ]

2 students in range ( 80 , 85 ]

> ##Question 2

> hist(scores$Score,breaks =c(30,50,70,90))

> hist2 <- hist(scores$Score,breaks =c(30,50,70,90))

> for(i in 1:length(hist2$counts)) {

+ Grade <- c("C","B","A")

+ cat(hist2$counts[i],"students in",Grade[i]," range","(",hist2$breaks[i],",",hist2$breaks[i+1],"]","\n")

+ }

17 students in C range ( 30 , 50 ]

70 students in B range ( 50 , 70 ]

13 students in A range ( 70 , 90 ]