ROI on Hand Picked Stocks 2007-2020

Janis Corona

February 2020

This is a project that is for now analyzing some hand picked stock to see if a program can be written based on the analysis of how certain stocks perform from 2007-2020. It looks at cyclical patterns of highs and lows, adds in the DOW highs and lows, the unemployment highs and lows, then mean and median values of daily changes various date fields for day of the week and month. The idea is to get the best performing stocks, analyze them with subsets of the worst performing stocks, get the specific features of each stock to describe it as a profit or loss forecasted stock to invest in based on its current stats, and more.

It will then add in the public sentiments for the lows and highs or local minima and maxima of the stock in the best performing set to predict the best time to buy and best to sell respectively, so that you could buy at a low cost and sell at a high cost and keep trading to increase profits of the portfolio.

```
portfolio <- read.csv('all_portfolio_prices.csv', header=TRUE,
na.strings=c('',' '),
                      row.names=1)

portfolio$Date <- row.names(portfolio)

Vol <- grep('Volume', colnames(portfolio))
close <- grep('Close',colnames(portfolio))
Close <- portfolio[,close]
Volume <- portfolio[,Vol]
colnames(Close)

##  [1] "TGT.Close"    "FTR.Close"    "UBSI.Close"   "HD.Close"
"JPM.Close"
##  [6] "XOM.Close"    "CVX.Close"    "NSANY.Close"  "GNBT.Close"
"MGM.Close"
## [11] "TEVA.Close"   "HST.Close"    "FCAU.Close"   "WFC.Close"
"WWE.Close"
## [16] "INO.Close"    "QSR.Close"    "GRPN.Close"   "SCE.PB.Close"
"FFIN.Close"
## [21] "GOOG.Close"   "WM.Close"     "ONCY.Close"   "S.Close"
"GM.Close"
## [26] "F.Close"      "ASCCY.Close"  "ARWR.Close"   "COST.Close"
"AAL.Close"
## [31] "JWN.Close"    "CSSEP.Close"  "NUS.Close"    "AMC.Close"
"ADDYY.Close"
## [36] "KSS.Close"    "MSFT.Close"   "LUV.Close"    "HMC.Close"
"PCG.Close"
## [41] "DLTR.Close"   "KGJI.Close"   "NKE.Close"    "AMZN.Close"
```

```
"ROST.Close"
## [46] "TMUS.Close"   "WMT.Close"    "TJX.Close"    "TM.Close"
"PBYI.Close"
## [51] "T.Close"      "JNJ.Close"    "C.Close"      "EPD.Close"
"VZ.Close"
## [56] "HRB.Close"    "NFLX.Close"   "AAP.Close"    "HOFT.Close"
"SIG.Close"
## [61] "SDC.Close"    "RRGB.Close"   "M.Close"      "JBLU.Close"
"YELP.Close"
```

Remove NAs from the data. The colSums(is.na(Close)) isn't returning the columns with NAs, so this must be done manually.

```
Close_noNAs <- Close[,-c(9,13,17,18,25,27,32,34,46,50,61,65)]
Volume_noNAs <- Volume[,-c(9,13,17,18,25,27,32,34,46,50,61,65)]

Close_noNAs$SCE.PB.Close <- as.numeric(Close_noNAs$SCE.PB.Close)
Volume_noNAs$SCE.PB.Volume <- as.numeric(Volume_noNAs$SCE.PB.Volume)
```

Add in a value of the portfolio column for each day's closing price of all stock that don't have NAs.

```
Close_noNAs$DailyValue <- rowSums(Close_noNAs,na.rm=TRUE)
```

Add in a daily change column of the portfolio closing prices.

```
dayVal <- as.data.frame(Close_noNAs$DailyValue)
colnames(dayVal) <- 'previousDayValue'
zero <- as.data.frame(as.numeric(dayVal$previousDayValue[1]))
colnames(zero) <- 'previousDayValue'
prevDay <- rbind(zero,dayVal)
Close_noNAs$prevDay <- prevDay[1:length(prevDay$previousDayValue)-1,1]
dailyChange <- as.data.frame(Close_noNAs$DailyValue-Close_noNAs$prevDay)
colnames(dailyChange) <- 'dailyValueChange'

Close1 <- cbind(Close_noNAs,dailyChange)
```

Add a column that gives the return in dollars on initial dollars invested.

```
Close1$ROI_dollars <- Close1$DailyValue-Close1$DailyValue[1]
```

Add some date fields to look at the values by date, day of the week, month, and year in analyzing this data.

```
Close1$Date <- as.Date.character(row.names(Close1))

Close1$DayOfWeek <- weekdays(as.Date(Close1$Date))

month <- month(as.Date(Close1$Date))
Month <- month.abb[month]
Close1$Month <- Month
```

Add in the year of the Date column.

```
Year <- year(as.Date(Close1$Date))

Close1$Year <- Year

Close1$MonthYear <- paste(Close1$Month, Close1$Year, sep='-')
Close1$MonthYear <- as.factor(Close1$MonthYear)
```

Add in some unemployment information as a column to see how the portfolio is doing by date.

```
ue <- read.delim('BLS_unemploymentRates2007-2020.txt', sep=',',header=TRUE,
                 na.strings=c('',' '))
UE <- ue[,-14]#remove the empty 'Annual' column
```

Use tidyr to gather the month fields with their respective unemployment rates per month.

```
gatherMonths <- gather(UE, 'UE_Month', 'UE_monthlyRate',2:13)

gatherMonths$MonthYear <- paste(gatherMonths$UE_Month, gatherMonths$Year,
sep='-')
gatherMonths$MonthYear <- as.factor(gatherMonths$MonthYear)

UE2 <- gatherMonths[,3:4]
Close2 <- merge(Close1, UE2, by.x='MonthYear', by.y='MonthYear')
row.names(Close2) <- Close2$Date
colnames(Close2)[55:58] <- paste('portfolio',colnames(Close2)[55:58],
sep='_')

write.csv(Close2, 'ROI_UE_2007_2020.csv', row.names=FALSE)
```

Lets add in the volume of trades per day from the Volume_noNAs data set. But lets add in some fields for total portfolio trades per day,

```
Volume1 <- Volume_noNAs
Volume1$portfolio_DailyVolume <- rowSums(Volume1, na.rm=TRUE)

dayVol <- as.data.frame(Volume1$portfolio_DailyVolume)
colnames(dayVol) <- 'portfolio_previousDayVolume'
zero <- as.data.frame(as.numeric(dayVol$portfolio_previousDayVolume[1]))
colnames(zero) <- 'portfolio_previousDayVolume'
prevDay1 <- rbind(zero,dayVol)
Volume1$portfolio_prevDayVolume <-
  prevDay1[1:(length(prevDay1$portfolio_previousDayVolume)-1),1]

dailyVolumeChange <- as.data.frame(Volume1$portfolio_DailyVolume-
Volume1$portfolio_prevDayVolume)
colnames(dailyVolumeChange) <- 'portfolio_dailyVolumeChange'

Volume2 <- cbind(Volume1,dailyVolumeChange)
```

```r
Volume2$portfolio_VolumeRatioDaily2Initial <-
Volume2$portfolio_DailyVolume/Volume2$portfolio_prevDayVolume[1]

Volume2$Date <- as.Date(row.names(Volume2))

stocks <- cbind(Close2, Volume2)

Stocks <- stocks[,c(2:54,64:116,1,55:63,117:120)]
colnames(Stocks)
```

```
##   [1] "TGT.Close"                          "FTR.Close"
##   [3] "UBSI.Close"                         "HD.Close"
##   [5] "JPM.Close"                          "XOM.Close"
##   [7] "CVX.Close"                          "NSANY.Close"
##   [9] "MGM.Close"                          "TEVA.Close"
##  [11] "HST.Close"                          "WFC.Close"
##  [13] "WWE.Close"                          "INO.Close"
##  [15] "SCE.PB.Close"                       "FFIN.Close"
##  [17] "GOOG.Close"                         "WM.Close"
##  [19] "ONCY.Close"                         "S.Close"
##  [21] "F.Close"                            "ARWR.Close"
##  [23] "COST.Close"                         "AAL.Close"
##  [25] "JWN.Close"                          "NUS.Close"
##  [27] "ADDYY.Close"                        "KSS.Close"
##  [29] "MSFT.Close"                         "LUV.Close"
##  [31] "HMC.Close"                          "PCG.Close"
##  [33] "DLTR.Close"                         "KGJI.Close"
##  [35] "NKE.Close"                          "AMZN.Close"
##  [37] "ROST.Close"                         "WMT.Close"
##  [39] "TJX.Close"                          "TM.Close"
##  [41] "T.Close"                            "JNJ.Close"
##  [43] "C.Close"                            "EPD.Close"
##  [45] "VZ.Close"                           "HRB.Close"
##  [47] "NFLX.Close"                         "AAP.Close"
##  [49] "HOFT.Close"                         "SIG.Close"
##  [51] "RRGB.Close"                         "M.Close"
##  [53] "JBLU.Close"                         "TGT.Volume"
##  [55] "FTR.Volume"                         "UBSI.Volume"
##  [57] "HD.Volume"                          "JPM.Volume"
##  [59] "XOM.Volume"                         "CVX.Volume"
##  [61] "NSANY.Volume"                       "MGM.Volume"
##  [63] "TEVA.Volume"                        "HST.Volume"
##  [65] "WFC.Volume"                         "WWE.Volume"
##  [67] "INO.Volume"                         "SCE.PB.Volume"
##  [69] "FFIN.Volume"                        "GOOG.Volume"
##  [71] "WM.Volume"                          "ONCY.Volume"
##  [73] "S.Volume"                           "F.Volume"
##  [75] "ARWR.Volume"                        "COST.Volume"
##  [77] "AAL.Volume"                         "JWN.Volume"
##  [79] "NUS.Volume"                         "ADDYY.Volume"
```

```
##  [81] "KSS.Volume"                        "MSFT.Volume"
##  [83] "LUV.Volume"                        "HMC.Volume"
##  [85] "PCG.Volume"                        "DLTR.Volume"
##  [87] "KGJI.Volume"                       "NKE.Volume"
##  [89] "AMZN.Volume"                       "ROST.Volume"
##  [91] "WMT.Volume"                        "TJX.Volume"
##  [93] "TM.Volume"                         "T.Volume"
##  [95] "JNJ.Volume"                        "C.Volume"
##  [97] "EPD.Volume"                        "VZ.Volume"
##  [99] "HRB.Volume"                        "NFLX.Volume"
## [101] "AAP.Volume"                        "HOFT.Volume"
## [103] "SIG.Volume"                        "RRGB.Volume"
## [105] "M.Volume"                          "JBLU.Volume"
## [107] "MonthYear"                         "portfolio_DailyValue"
## [109] "portfolio_prevDay"                 "portfolio_dailyValueChange"
## [111] "portfolio_ROI_dollars"             "Date"
## [113] "DayOfWeek"                         "Month"
## [115] "Year"                              "UE_monthlyRate"
## [117] "portfolio_DailyVolume"             "portfolio_prevDayVolume"
## [119] "portfolio_dailyVolumeChange"
"portfolio_VolumeRatioDaily2Initial"
```

Add a value of stock daily to the initial value as a ratio.

```
Stocks$portfolio_ValueRatioDaily2Initial <-
  Stocks$portfolio_DailyValue/Stocks$portfolio_DailyValue[1]
```

Add a field that multiplies the daily value and daily volume ratios compared to the initial value and volume by the unemployment rate.

```
Stocks$portfolio_DailyRatios_X_UE <-

Stocks$portfolio_ValueRatioDaily2Initial*Stocks$portfolio_VolumeRatioDaily2In
itial*Stocks$UE_monthlyRate
```

Add an exponential calculation field based on the unemployment rate for rate, and using t=1/12 for 12 months, and a binary value of 1 or 2 where the daily change is positive is assigned a 1 and a negative is a 2. This will make those values decreasing daily have a lower poisson and those values increasing a higher poisson value. This is a modified poisson used for probability of an outcome occuring with a constant rate. Added to rank daily changes based on unemployment rate of each month.

```
Stocks <- Stocks[complete.cases(Stocks$UE_monthlyRate),]
Stocks$dayOfMonth <- day(Stocks$Date)
dayOfMonth <- day(Stocks$Date)
ue1 <- Stocks$UE_monthlyRate

incrDecr <- ifelse(Stocks$portfolio_dailyValueChange>0,1,2)

Stocks$portfolio_poisson <- round((exp(-
```

```
(ue1*1/12))*(ue1*1/12)^incrDecr)/(factorial(incrDecr)),5)

summary(Stocks$portfolio_poisson)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
## 0.03177 0.07392 0.22652 0.19506 0.29808 0.36217

write.csv(Stocks, 'StocksStats.csv', row.names=TRUE)
```

Make a daily ROI dollars column for each of the stocks in this set.

```
stocks1 <- Stocks[,1:53]
colnames(stocks1)

##  [1] "TGT.Close"    "FTR.Close"    "UBSI.Close"   "HD.Close"    
"JPM.Close"   
##  [6] "XOM.Close"    "CVX.Close"    "NSANY.Close"  "MGM.Close"   
"TEVA.Close"  
## [11] "HST.Close"    "WFC.Close"    "WWE.Close"    "INO.Close"   
"SCE.PB.Close"
## [16] "FFIN.Close"   "GOOG.Close"   "WM.Close"     "ONCY.Close"   "S.Close"
## [21] "F.Close"      "ARWR.Close"   "COST.Close"   "AAL.Close"   
"JWN.Close"   
## [26] "NUS.Close"    "ADDYY.Close"  "KSS.Close"    "MSFT.Close"  
"LUV.Close"   
## [31] "HMC.Close"    "PCG.Close"    "DLTR.Close"   "KGJI.Close"  
"NKE.Close"   
## [36] "AMZN.Close"   "ROST.Close"   "WMT.Close"    "TJX.Close"   
"TM.Close"    
## [41] "T.Close"      "JNJ.Close"    "C.Close"      "EPD.Close"   
"VZ.Close"    
## [46] "HRB.Close"    "NFLX.Close"   "AAP.Close"    "HOFT.Close"  
"SIG.Close"   
## [51] "RRGB.Close"   "M.Close"      "JBLU.Close"

stocks1$TGT_ROI_dollars <- stocks1$TGT.Close-stocks1$TGT.Close[1]
stocks1$FTR_ROI_dollars <- stocks1$FTR.Close-stocks1$FTR.Close[1]
stocks1$UBSI_ROI_dollars <- stocks1$UBSI.Close-stocks1$UBSI.Close[1]
stocks1$HD_ROI_dollars <- stocks1$HD.Close-stocks1$HD.Close[1]
stocks1$JPM_ROI_dollars <- stocks1$JPM.Close-stocks1$JPM.Close[1]

stocks1$XOM_ROI_dollars <- stocks1$XOM.Close-stocks1$XOM.Close[1]
stocks1$CVX_ROI_dollars <- stocks1$CVX.Close-stocks1$CVX.Close[1]
stocks1$NSANY_ROI_dollars <- stocks1$NSANY.Close-stocks1$NSANY.Close[1]
stocks1$MGM_ROI_dollars <- stocks1$MGM.Close-stocks1$MGM.Close[1]
stocks1$TEVA_ROI_dollars <- stocks1$TEVA.Close-stocks1$TEVA.Close[1]

stocks1$HST_ROI_dollars <- stocks1$HST.Close-stocks1$HST.Close[1]
stocks1$WFC_ROI_dollars <- stocks1$WFC.Close-stocks1$WFC.Close[1]
stocks1$WWE_ROI_dollars <- stocks1$WWE.Close-stocks1$WWE.Close[1]
stocks1$INO_ROI_dollars <- stocks1$INO.Close-stocks1$INO.Close[1]
```

```
stocks1$SCE.PB_ROI_dollars <- stocks1$SCE.PB.Close-stocks1$SCE.PB.Close[1]

stocks1$FFIN_ROI_dollars <- stocks1$FFIN.Close-stocks1$FFIN.Close[1]
stocks1$GOOG_ROI_dollars <- stocks1$GOOG.Close-stocks1$GOOG.Close[1]
stocks1$WM_ROI_dollars <- stocks1$WM.Close-stocks1$WM.Close[1]
stocks1$ONCY_ROI_dollars <- stocks1$ONCY.Close-stocks1$ONCY.Close[1]
stocks1$S_ROI_dollars <- stocks1$S.Close-stocks1$S.Close[1]

stocks1$F_ROI_dollars <- stocks1$F.Close-stocks1$F.Close[1]
stocks1$ARWR_ROI_dollars <- stocks1$ARWR.Close-stocks1$ARWR.Close[1]
stocks1$COST_ROI_dollars <- stocks1$COST.Close-stocks1$COST.Close[1]
stocks1$AAL_ROI_dollars <- stocks1$AAL.Close-stocks1$AAL.Close[1]
stocks1$JWN_ROI_dollars <- stocks1$JWN.Close-stocks1$JWN.Close[1]

stocks1$NUS_ROI_dollars <- stocks1$NUS.Close-stocks1$NUS.Close[1]
stocks1$HMC_ROI_dollars <- stocks1$HMC.Close-stocks1$HMC.Close[1]
stocks1$AMZN_ROI_dollars <- stocks1$AMZN.Close-stocks1$AMZN.Close[1]
stocks1$T_ROI_dollars <- stocks1$T.Close-stocks1$T.Close[1]
stocks1$HRB_ROI_dollars <- stocks1$HRB.Close-stocks1$HRB.Close[1]
stocks1$RRGB_ROI_dollars <- stocks1$RRGB.Close-stocks1$RRGB.Close[1]

stocks1$ADDYY_ROI_dollars <- stocks1$ADDYY.Close-stocks1$ADDYY.Close[1]
stocks1$PCG_ROI_dollars <- stocks1$PCG.Close-stocks1$PCG.Close[1]
stocks1$ROST_ROI_dollars <- stocks1$ROST.Close-stocks1$ROST.Close[1]
stocks1$JNJ_ROI_dollars <- stocks1$JNJ.Close-stocks1$JNJ.Close[1]
stocks1$NFLX_ROI_dollars <- stocks1$NFLX.Close-stocks1$NFLX.Close[1]
stocks1$M_ROI_dollars <- stocks1$M.Close-stocks1$M.Close[1]

stocks1$KSS_ROI_dollars <- stocks1$KSS.Close-stocks1$KSS.Close[1]
stocks1$DLTR_ROI_dollars <- stocks1$DLTR.Close-stocks1$DLTR.Close[1]
stocks1$WMT_ROI_dollars <- stocks1$WMT.Close-stocks1$WMT.Close[1]
stocks1$C_ROI_dollars <- stocks1$C.Close-stocks1$C.Close[1]
stocks1$AAP_ROI_dollars <- stocks1$AAP.Close-stocks1$AAP.Close[1]
stocks1$JBLU_ROI_dollars <- stocks1$JBLU.Close-stocks1$JBLU.Close[1]

stocks1$MSFT_ROI_dollars <- stocks1$MSFT.Close-stocks1$MSFT.Close[1]
stocks1$KGJI_ROI_dollars <- stocks1$KGJI.Close-stocks1$KGJI.Close[1]
stocks1$EPD_ROI_dollars <- stocks1$EPD.Close-stocks1$EPD.Close[1]
stocks1$TJX_ROI_dollars <- stocks1$TJX.Close-stocks1$TJX.Close[1]
stocks1$HOFT_ROI_dollars <- stocks1$HOFT.Close-stocks1$HOFT.Close[1]

stocks1$LUV_ROI_dollars <- stocks1$LUV.Close-stocks1$LUV.Close[1]
stocks1$NKE_ROI_dollars <- stocks1$NKE.Close-stocks1$NKE.Close[1]
stocks1$TM_ROI_dollars <- stocks1$TM.Close-stocks1$TM.Close[1]
stocks1$VZ_ROI_dollars <- stocks1$VZ.Close-stocks1$VZ.Close[1]
stocks1$SIG_ROI_dollars <- stocks1$SIG.Close-stocks1$SIG.Close[1]
```

These are the values of the stock the previous day that will be subtracted from each day to get the daily change from the day before in dollars.

```
TGTa <- c(0,stocks1$TGT.Close[1:(length(stocks1$TGT.Close)-1)])
FTRa <- c(0, stocks1$FTR.Close[1:(length(stocks1$TGT.Close)-1)])
UBSIa <- c(0,stocks1$UBSI.Close[1:(length(stocks1$TGT.Close)-1)])
HDa <- c(0,stocks1$HD.Close[1:(length(stocks1$TGT.Close)-1)])
JPMa <- c(0,stocks1$JPM.Close[1:(length(stocks1$TGT.Close)-1)])
XOMa <- c(0,stocks1$XOM.Close[1:(length(stocks1$TGT.Close)-1)])
CVXa <- c(0,stocks1$CVX.Close[1:(length(stocks1$TGT.Close)-1)])
NSANYa <- c(0,stocks1$NSANY.Close[1:(length(stocks1$TGT.Close)-1)])
MGMa <- c(0,stocks1$MGM.Close[1:(length(stocks1$TGT.Close)-1)])
TEVAa <- c(0, stocks1$TEVA.Close[1:(length(stocks1$TGT.Close)-1)])
HSTa <- c(0, stocks1$HST.Close[1:(length(stocks1$TGT.Close)-1)])
WFCa <- c(0, stocks1$WFC.Close[1:(length(stocks1$TGT.Close)-1)])
WWEa <- c(0, stocks1$WWE.Close[1:(length(stocks1$TGT.Close)-1)])
INOa <- c(0,stocks1$INO.Close[1:(length(stocks1$TGT.Close)-1)])
SCEa <- c(0,stocks1$SCE.PB.Close[1:(length(stocks1$TGT.Close)-1)])
FFINa <- c(0,stocks1$FFIN.Close[1:(length(stocks1$TGT.Close)-1)])
GOOGa <- c(0,stocks1$GOOG.Close[1:(length(stocks1$TGT.Close)-1)])
WMa <- c(0,stocks1$WM.Close[1:(length(stocks1$TGT.Close)-1)])
ONCYa <- c(0,stocks1$ONCY.Close[1:(length(stocks1$TGT.Close)-1)])
Sa <- c(0,stocks1$S.Close[1:(length(stocks1$TGT.Close)-1)])
Fa <- c(0,stocks1$F.Close[1:(length(stocks1$TGT.Close)-1)])
ARWRa <- c(0,stocks1$ARWR.Close[1:(length(stocks1$TGT.Close)-1)])
COSTa <- c(0,stocks1$COST.Close[1:(length(stocks1$TGT.Close)-1)])
AALa <- c(0,stocks1$AAL.Close[1:(length(stocks1$TGT.Close)-1)])
JWNa <- c(0,stocks1$JWN.Close[1:(length(stocks1$TGT.Close)-1)])
NUSa <- c(0,stocks1$NUS.Close[1:(length(stocks1$TGT.Close)-1)])
ADDYYa <- c(0,stocks1$ADDYY.Close[1:(length(stocks1$TGT.Close)-1)])
KSSa <- c(0,stocks1$KSS.Close[1:(length(stocks1$TGT.Close)-1)])
MSFTa <- c(0,stocks1$MSFT.Close[1:(length(stocks1$TGT.Close)-1)])
LUVa <- c(0,stocks1$LUV.Close[1:(length(stocks1$TGT.Close)-1)])
HMCa <- c(0,stocks1$HMC.Close[1:(length(stocks1$TGT.Close)-1)])
PCGa <- c(0,stocks1$PCG.Close[1:(length(stocks1$TGT.Close)-1)])
DLTRa <- c(0,stocks1$DLTR.Close[1:(length(stocks1$TGT.Close)-1)])
KGJIa <- c(0,stocks1$KGJI.Close[1:(length(stocks1$TGT.Close)-1)])
NKEa <- c(0,stocks1$NKE.Close[1:(length(stocks1$TGT.Close)-1)])
AMZNa <- c(0,stocks1$AMZN.Close[1:(length(stocks1$TGT.Close)-1)])
ROSTa <- c(0,stocks1$ROST.Close[1:(length(stocks1$TGT.Close)-1)])
WMTa <- c(0,stocks1$WMT.Close[1:(length(stocks1$TGT.Close)-1)])
TJXa <- c(0,stocks1$TJX.Close[1:(length(stocks1$TGT.Close)-1)])
TMa <- c(0,stocks1$TM.Close[1:(length(stocks1$TGT.Close)-1)])
Ta <- c(0,stocks1$T.Close[1:(length(stocks1$TGT.Close)-1)])
JNJa <- c(0,stocks1$JNJ.Close[1:(length(stocks1$TGT.Close)-1)])
Ca <- c(0,stocks1$C.Close[1:(length(stocks1$TGT.Close)-1)])
EPDa <- c(0,stocks1$EPD.Close[1:(length(stocks1$TGT.Close)-1)])
VZa <- c(0,stocks1$VZ.Close[1:(length(stocks1$TGT.Close)-1)])
HRBa <- c(0,stocks1$HRB.Close[1:(length(stocks1$TGT.Close)-1)])
NFLXa <- c(0,stocks1$NFLX.Close[1:(length(stocks1$TGT.Close)-1)])
AAPa <- c(0,stocks1$AAP.Close[1:(length(stocks1$TGT.Close)-1)])
HOFTa <- c(0,stocks1$HOFT.Close[1:(length(stocks1$TGT.Close)-1)])
SIGa <- c(0,stocks1$SIG.Close[1:(length(stocks1$TGT.Close)-1)])
```

```
RRGBa <- c(0,stocks1$RRGB.Close[1:(length(stocks1$TGT.Close)-1)])
Ma <- c(0,stocks1$M.Close[1:(length(stocks1$TGT.Close)-1)])
JBLUa <- c(0,stocks1$JBLU.Close[1:(length(stocks1$TGT.Close)-1)])
```

This creates the DailyChange per stock columns.

```
stocks1$TGT_dailyChange <- stocks1$TGT.Close-TGTa
stocks1$FTR_dailyChange <- stocks1$FTR.Close-FTRa
stocks1$UBSI_dailyChange <- stocks1$UBSI.Close-UBSIa
stocks1$HD_dailyChange <- stocks1$HD.Close-HDa
stocks1$JPM_dailyChange <- stocks1$JPM.Close-JPMa

stocks1$XOM_dailyChange <- stocks1$XOM.Close-XOMa
stocks1$CVX_dailyChange <- stocks1$CVX.Close-CVXa
stocks1$NSANY_dailyChange <- stocks1$NSANY.Close-NSANYa
stocks1$MGM_dailyChange <- stocks1$MGM.Close-MGMa
stocks1$TEVA_dailyChange <- stocks1$TEVA.Close-TEVAa

stocks1$HST_dailyChange <- stocks1$HST.Close-HSTa
stocks1$WFC_dailyChange <- stocks1$WFC.Close-WFCa
stocks1$WWE_dailyChange <- stocks1$WWE.Close-WWEa
stocks1$INO_dailyChange <- stocks1$INO.Close-INOa
stocks1$SCE.PB_dailyChange <- stocks1$SCE.PB.Close-SCEa

stocks1$FFIN_dailyChange <- stocks1$FFIN.Close-FFINa
stocks1$GOOG_dailyChange <- stocks1$GOOG.Close-GOOGa
stocks1$WM_dailyChange <- stocks1$WM.Close-WMa
stocks1$ONCY_dailyChange <- stocks1$ONCY.Close-ONCYa
stocks1$S_dailyChange <- stocks1$S.Close-Sa

stocks1$F_dailyChange <- stocks1$F.Close-Fa
stocks1$ARWR_dailyChange <- stocks1$ARWR.Close-ARWRa
stocks1$COST_dailyChange <- stocks1$COST.Close-COSTa
stocks1$AAL_dailyChange <- stocks1$AAL.Close-AALa
stocks1$JWN_dailyChange <- stocks1$JWN.Close-JWNa

stocks1$NUS_dailyChange <- stocks1$NUS.Close-NUSa
stocks1$HMC_dailyChange <- stocks1$HMC.Close-HMCa
stocks1$AMZN_dailyChange <- stocks1$AMZN.Close-AMZNa
stocks1$T_dailyChange <- stocks1$T.Close-Ta
stocks1$HRB_dailyChange <- stocks1$HRB.Close-HRBa
stocks1$RRGB_dailyChange <- stocks1$RRGB.Close-RRGBa

stocks1$ADDYY_dailyChange <- stocks1$ADDYY.Close-ADDYYa
stocks1$PCG_dailyChange <- stocks1$PCG.Close-PCGa
stocks1$ROST_dailyChange <- stocks1$ROST.Close-ROSTa
stocks1$JNJ_dailyChange <- stocks1$JNJ.Close-JNJa
stocks1$NFLX_dailyChange <- stocks1$NFLX.Close-NFLXa
stocks1$M_dailyChange <- stocks1$M.Close-Ma
```

```r
stocks1$KSS_dailyChange <- stocks1$KSS.Close-KSSa
stocks1$DLTR_dailyChange <- stocks1$DLTR.Close-DLTRa
stocks1$WMT_dailyChange <- stocks1$WMT.Close-WMTa
stocks1$C_dailyChange <- stocks1$C.Close-Ca
stocks1$AAP_dailyChange <- stocks1$AAP.Close-AAPa
stocks1$JBLU_dailyChange <- stocks1$JBLU.Close-JBLUa

stocks1$MSFT_dailyChange <- stocks1$MSFT.Close-MSFTa
stocks1$KGJI_dailyChange <- stocks1$KGJI.Close-KGJIa
stocks1$EPD_dailyChange <- stocks1$EPD.Close-EPDa
stocks1$TJX_dailyChange <- stocks1$TJX.Close-TJXa
stocks1$HOFT_dailyChange <- stocks1$HOFT.Close-HOFTa

stocks1$LUV_dailyChange <- stocks1$LUV.Close-LUVa
stocks1$NKE_dailyChange <- stocks1$NKE.Close-NKEa
stocks1$TM_dailyChange <- stocks1$TM.Close-TMa
stocks1$VZ_dailyChange <- stocks1$VZ.Close-VZa
stocks1$SIG_dailyChange <- stocks1$SIG.Close-SIGa
```

Combine the stocks1 stats of ROI and daily change in dollars per stock to the stocks stats data table.

```r
stocks2 <- stocks1[,-c(1:53)]
StocksSTATS <- cbind(Stocks,stocks2)
```

All the columns we now have are:

```r
StocksSTATS <- StocksSTATS[,c(1:106,125:230,107:124)]
colnames(StocksSTATS)
```

```
##   [1] "TGT.Close"                                "FTR.Close"
##   [3] "UBSI.Close"                               "HD.Close"
##   [5] "JPM.Close"                                "XOM.Close"
##   [7] "CVX.Close"                                "NSANY.Close"
##   [9] "MGM.Close"                                "TEVA.Close"
##  [11] "HST.Close"                                "WFC.Close"
##  [13] "WWE.Close"                                "INO.Close"
##  [15] "SCE.PB.Close"                             "FFIN.Close"
##  [17] "GOOG.Close"                               "WM.Close"
##  [19] "ONCY.Close"                               "S.Close"
##  [21] "F.Close"                                  "ARWR.Close"
##  [23] "COST.Close"                               "AAL.Close"
##  [25] "JWN.Close"                                "NUS.Close"
##  [27] "ADDYY.Close"                              "KSS.Close"
##  [29] "MSFT.Close"                               "LUV.Close"
##  [31] "HMC.Close"                                "PCG.Close"
##  [33] "DLTR.Close"                               "KGJI.Close"
##  [35] "NKE.Close"                                "AMZN.Close"
##  [37] "ROST.Close"                               "WMT.Close"
##  [39] "TJX.Close"                                "TM.Close"
##  [41] "T.Close"                                  "JNJ.Close"
```

```
##  [43] "C.Close"                          "EPD.Close"
##  [45] "VZ.Close"                         "HRB.Close"
##  [47] "NFLX.Close"                       "AAP.Close"
##  [49] "HOFT.Close"                       "SIG.Close"
##  [51] "RRGB.Close"                       "M.Close"
##  [53] "JBLU.Close"                       "TGT.Volume"
##  [55] "FTR.Volume"                       "UBSI.Volume"
##  [57] "HD.Volume"                        "JPM.Volume"
##  [59] "XOM.Volume"                       "CVX.Volume"
##  [61] "NSANY.Volume"                     "MGM.Volume"
##  [63] "TEVA.Volume"                      "HST.Volume"
##  [65] "WFC.Volume"                       "WWE.Volume"
##  [67] "INO.Volume"                       "SCE.PB.Volume"
##  [69] "FFIN.Volume"                      "GOOG.Volume"
##  [71] "WM.Volume"                        "ONCY.Volume"
##  [73] "S.Volume"                         "F.Volume"
##  [75] "ARWR.Volume"                      "COST.Volume"
##  [77] "AAL.Volume"                       "JWN.Volume"
##  [79] "NUS.Volume"                       "ADDYY.Volume"
##  [81] "KSS.Volume"                       "MSFT.Volume"
##  [83] "LUV.Volume"                       "HMC.Volume"
##  [85] "PCG.Volume"                       "DLTR.Volume"
##  [87] "KGJI.Volume"                      "NKE.Volume"
##  [89] "AMZN.Volume"                      "ROST.Volume"
##  [91] "WMT.Volume"                       "TJX.Volume"
##  [93] "TM.Volume"                        "T.Volume"
##  [95] "JNJ.Volume"                       "C.Volume"
##  [97] "EPD.Volume"                       "VZ.Volume"
##  [99] "HRB.Volume"                       "NFLX.Volume"
## [101] "AAP.Volume"                       "HOFT.Volume"
## [103] "SIG.Volume"                       "RRGB.Volume"
## [105] "M.Volume"                         "JBLU.Volume"
## [107] "TGT_ROI_dollars"                  "FTR_ROI_dollars"
## [109] "UBSI_ROI_dollars"                 "HD_ROI_dollars"
## [111] "JPM_ROI_dollars"                  "XOM_ROI_dollars"
## [113] "CVX_ROI_dollars"                  "NSANY_ROI_dollars"
## [115] "MGM_ROI_dollars"                  "TEVA_ROI_dollars"
## [117] "HST_ROI_dollars"                  "WFC_ROI_dollars"
## [119] "WWE_ROI_dollars"                  "INO_ROI_dollars"
## [121] "SCE.PB_ROI_dollars"               "FFIN_ROI_dollars"
## [123] "GOOG_ROI_dollars"                 "WM_ROI_dollars"
## [125] "ONCY_ROI_dollars"                 "S_ROI_dollars"
## [127] "F_ROI_dollars"                    "ARWR_ROI_dollars"
## [129] "COST_ROI_dollars"                 "AAL_ROI_dollars"
## [131] "JWN_ROI_dollars"                  "NUS_ROI_dollars"
## [133] "HMC_ROI_dollars"                  "AMZN_ROI_dollars"
## [135] "T_ROI_dollars"                    "HRB_ROI_dollars"
## [137] "RRGB_ROI_dollars"                 "ADDYY_ROI_dollars"
## [139] "PCG_ROI_dollars"                  "ROST_ROI_dollars"
## [141] "JNJ_ROI_dollars"                  "NFLX_ROI_dollars"
```

```
## [143] "M_ROI_dollars"                      "KSS_ROI_dollars"
## [145] "DLTR_ROI_dollars"                   "WMT_ROI_dollars"
## [147] "C_ROI_dollars"                      "AAP_ROI_dollars"
## [149] "JBLU_ROI_dollars"                   "MSFT_ROI_dollars"
## [151] "KGJI_ROI_dollars"                   "EPD_ROI_dollars"
## [153] "TJX_ROI_dollars"                    "HOFT_ROI_dollars"
## [155] "LUV_ROI_dollars"                    "NKE_ROI_dollars"
## [157] "TM_ROI_dollars"                     "VZ_ROI_dollars"
## [159] "SIG_ROI_dollars"                    "TGT_dailyChange"
## [161] "FTR_dailyChange"                    "UBSI_dailyChange"
## [163] "HD_dailyChange"                     "JPM_dailyChange"
## [165] "XOM_dailyChange"                    "CVX_dailyChange"
## [167] "NSANY_dailyChange"                  "MGM_dailyChange"
## [169] "TEVA_dailyChange"                   "HST_dailyChange"
## [171] "WFC_dailyChange"                    "WWE_dailyChange"
## [173] "INO_dailyChange"                    "SCE.PB_dailyChange"
## [175] "FFIN_dailyChange"                   "GOOG_dailyChange"
## [177] "WM_dailyChange"                     "ONCY_dailyChange"
## [179] "S_dailyChange"                      "F_dailyChange"
## [181] "ARWR_dailyChange"                   "COST_dailyChange"
## [183] "AAL_dailyChange"                    "JWN_dailyChange"
## [185] "NUS_dailyChange"                    "HMC_dailyChange"
## [187] "AMZN_dailyChange"                   "T_dailyChange"
## [189] "HRB_dailyChange"                    "RRGB_dailyChange"
## [191] "ADDYY_dailyChange"                  "PCG_dailyChange"
## [193] "ROST_dailyChange"                   "JNJ_dailyChange"
## [195] "NFLX_dailyChange"                   "M_dailyChange"
## [197] "KSS_dailyChange"                    "DLTR_dailyChange"
## [199] "WMT_dailyChange"                    "C_dailyChange"
## [201] "AAP_dailyChange"                    "JBLU_dailyChange"
## [203] "MSFT_dailyChange"                   "KGJI_dailyChange"
## [205] "EPD_dailyChange"                    "TJX_dailyChange"
## [207] "HOFT_dailyChange"                   "LUV_dailyChange"
## [209] "NKE_dailyChange"                    "TM_dailyChange"
## [211] "VZ_dailyChange"                     "SIG_dailyChange"
## [213] "MonthYear"                          "portfolio_DailyValue"
## [215] "portfolio_prevDay"                  "portfolio_dailyValueChange"
## [217] "portfolio_ROI_dollars"              "Date"
## [219] "DayOfWeek"                          "Month"
## [221] "Year"                               "UE_monthlyRate"
## [223] "portfolio_DailyVolume"              "portfolio_prevDayVolume"
## [225] "portfolio_dailyVolumeChange"
"portfolio_VolumeRatioDaily2Initial"
## [227] "portfolio_ValueRatioDaily2Initial"  "portfolio_DailyRatios_X_UE"
## [229] "dayOfMonth"                         "portfolio_poisson"
```

```r
write.csv(StocksSTATS, 'STOCKS_STATS.csv', row.names=TRUE)
```

Lets us pick one stock, look at the stats we added for that stock and then pull out some googled articles of that stock as a company in the news since 2007 till today's date of

Feb. 18, 2020 to compare the sentiments on the company with words that we will count the number of times the company is in the news, the comments by readers, zoom in on the dates of those articles, and see how the company behaved. Lets choose the highest ROI in dollars out of our stocks and compare it to the lowest ROI in dollars.

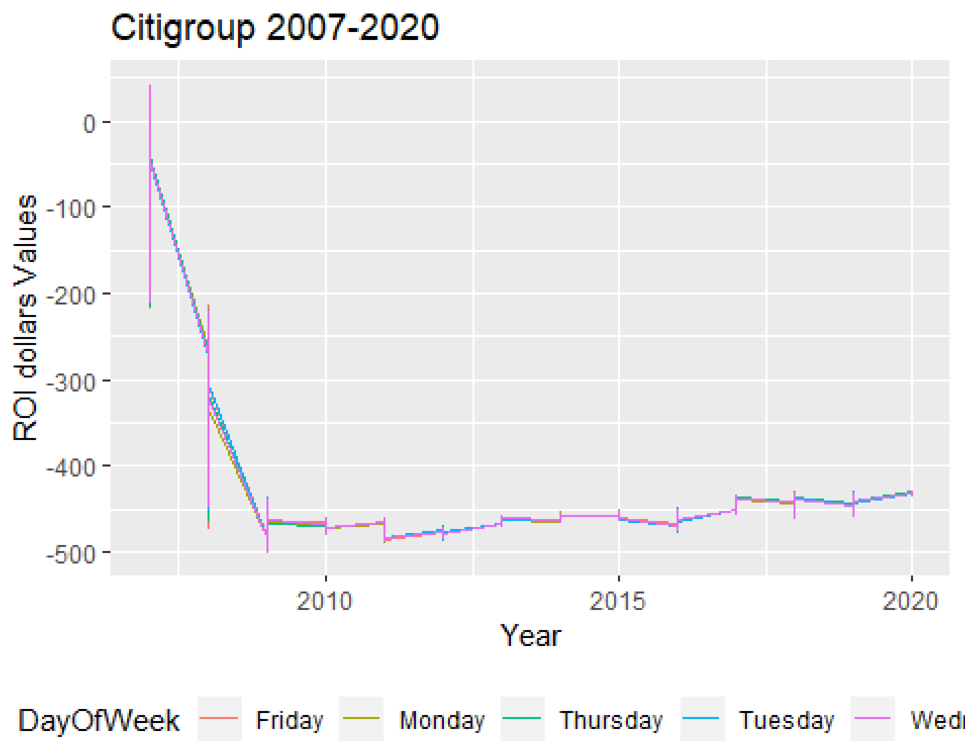
```
m <- StocksSTATS[order(StocksSTATS$Date,
decreasing=FALSE)[length(StocksSTATS$Date)], 107:159]
t <- as.data.frame(t(m))
colnames(t) <- row.names(m)
t$StockROI <- row.names(t)

Troi <- t[order(t$'2020-01-31', decreasing=TRUE),]

mostLeast <- rbind(head(Troi,3),tail(Troi,3))
mostLeast <- na.omit(mostLeast)
mostLeast

##                      2020-01-31           StockROI
## AMZN_ROI_dollars       1968.300   AMZN_ROI_dollars
## GOOG_ROI_dollars       1205.821   GOOG_ROI_dollars
## SCE.PB_ROI_dollars      679.000 SCE.PB_ROI_dollars
## MGM_ROI_dollars         -40.520    MGM_ROI_dollars
## FTR_ROI_dollars        -225.200    FTR_ROI_dollars
## C_ROI_dollars          -436.090      C_ROI_dollars
```

The above table shows the three highest returns on investment and the three lowest since Jan 3, 2007 to Jan 31, 2020. Lets use the lowest stock for now (C is Citigroup bank), because AMZN (Amazon) is always in the news and it would fluctuate a lot I would think, but we could look at the quartiles for each and get the news releases of each date where the stock was in that quartile range, look at the median ROI, the min and max too, and cross referencing with the other stat fields.

```
amzn <- grep('AMZN', colnames(StocksSTATS))
c <- grep('^C[.|_]', colnames(StocksSTATS))
C_stock <- StocksSTATS[,c(c,213:230)]
amzn_stock <- StocksSTATS[,c(amzn,213:230)]
```

Citigroup is our C_stock table and Amazon is our amzn_stock table. Lets look at the daily ratios of volume and ROI in dollars times the unemployment rate column and the day of the week and day of the year and poisson columns.

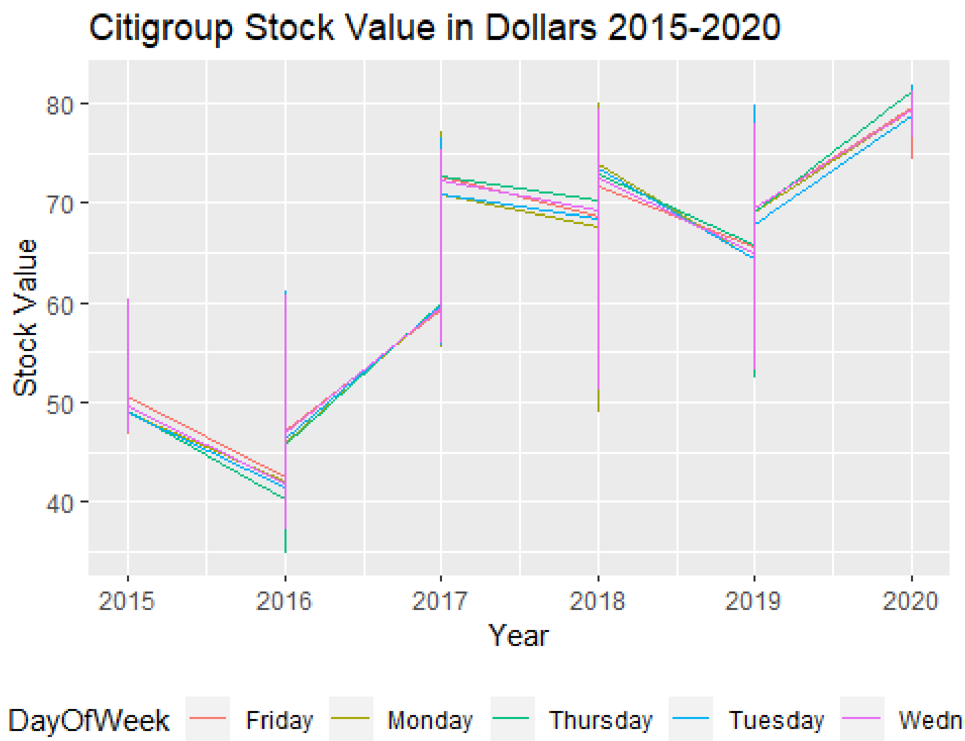
```
ggplot(data = C_stock, aes(x=Year, y=C_ROI_dollars,group=DayOfWeek)) +
  geom_line(aes(color=DayOfWeek))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Citigroup 2007-2020')+
  ylab('ROI dollars Values')

## Warning in pal_name(palette, type): Unknown palette paired
```



We can see from the plot above that buying Citigroup stock anywhere before 2010, was a bad idea. But we also see that the stock would have been good to buy around 2010-2016, as it overall increased its return on investment in dollars initially invested.

Lets look at the years from 2016-2020 to see this plotted Citigroup stock.

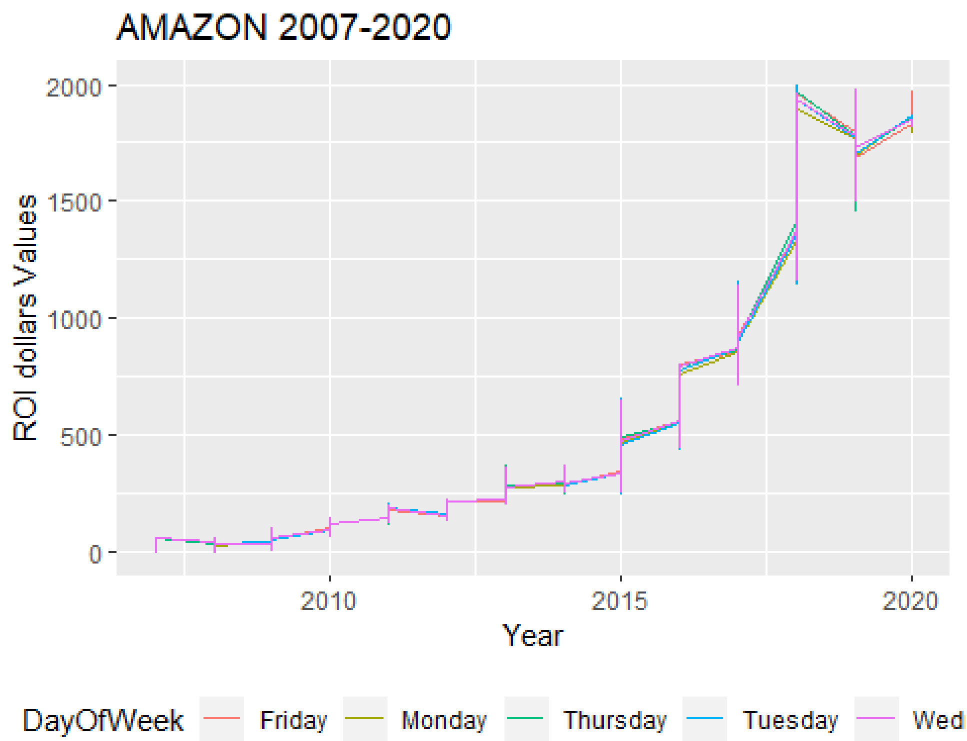
```
y2015plus <- subset(C_stock, C_stock$Year>2014)

ggplot(data = y2015plus, aes(x=Year, y=C.Close,group=DayOfWeek)) +
  geom_line(aes(color=DayOfWeek))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Citigroup Stock Value in Dollars 2015-2020')+
  ylab('Stock Value')

## Warning in pal_name(palette, type): Unknown palette paired
```



We see from the above plot that Citigroup was good to buy at the start of 2016 or 2019 if you want to see an increase all year long, but in 2017-2018 it decreased.Overall, if investing since 2016, the stock increased from the high $40 to the mid-high $70 range. This would be good to cross reference with unemployment rates and the news articles online text mined for public sentiment on Citigroup.

Lets look at amazon for the same quick plotted analysis as done with Citigroup.

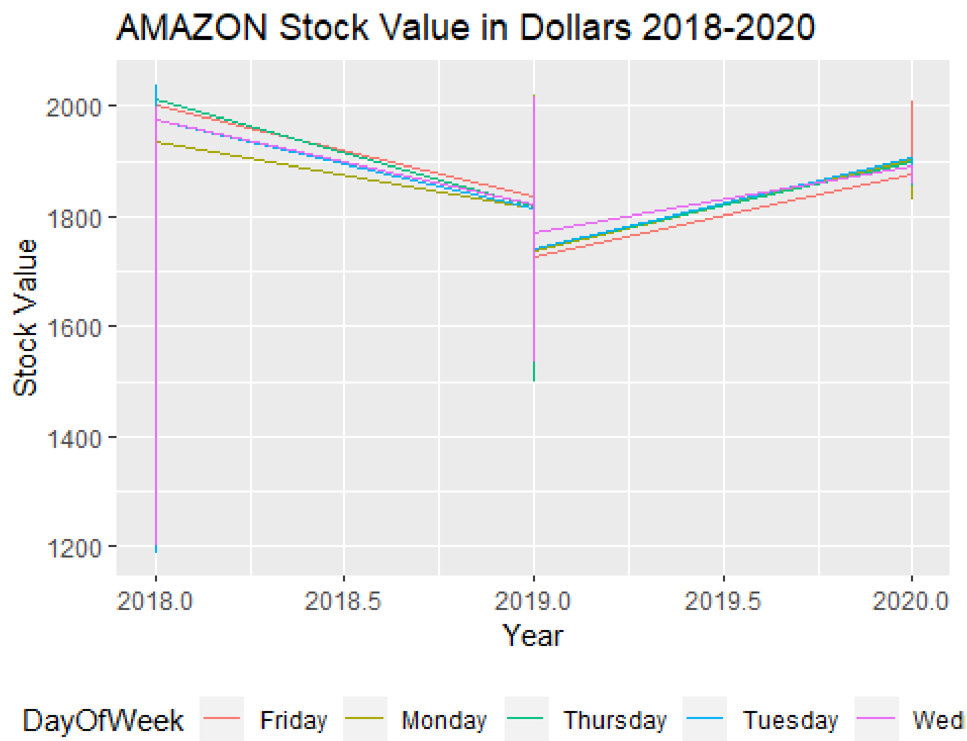
```
ggplot(data = amzn_stock, aes(x=Year, y=AMZN_ROI_dollars,group=DayOfWeek)) +
  geom_line(aes(color=DayOfWeek))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('AMAZON 2007-2020')+
  ylab('ROI dollars Values')

## Warning in pal_name(palette, type): Unknown palette paired
```



We can see from the plot above that buying AMAZON stock anywhere before 2010, was a great idea. But we also see that the stock would have been good to buy around 2010-2018 or 2019 but not in 2018, as it overall increased its return on investment in dollars initially invested.In 2018, you bought high and it decreased the entire year. This would be great to see what happened in 2018 with the value. So we will.

Lets look at the years from 2018-2020 to see this plotted Citigroup stock.

```
y2015plus <- subset(amzn_stock, amzn_stock$Year>2017)

ggplot(data = y2015plus, aes(x=Year, y=AMZN.Close,group=DayOfWeek)) +
  geom_line(aes(color=DayOfWeek))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('AMAZON Stock Value in Dollars 2018-2020')+
  ylab('Stock Value')

## Warning in pal_name(palette, type): Unknown palette paired
```



The chart above shows how the value in dollars and day of the week from 2018-2020 decreases in 2018 and increases in 2019. If you bought in 2018, you lost money the entire year, but you gained it back in 2019 plus some additional earnings.

Lets group by the day of the month in this time series of the Citigroup stock and get the median value for the volumne of stocks traded for Citigroup by days 1-31 of the month.

```
v1 <- as.vector(colnames(C_stock)[2])
Citi <- C_stock %>% group_by(dayOfMonth) %>% summarise_at(vars(v1), median,
                                                                   na.rm=T)
Citi <- as.data.frame(Citi)
colnames(Citi)[2] <- 'Citi_Median_Volume'
Citi <- Citi[order(Citi$Citi_Median_Volume, decreasing=T),]
headTail_Citi_volume <- rbind(head(Citi,3), tail(Citi,3))
headTail_Citi_volume

##    dayOfMonth Citi_Median_Volume
## 16         16           22388100
## 31         31           22302200
## 3           3           21221500
## 25         25           17960700
## 20         20           17548500
## 2           2           17134600
```

From the above table we see that the most volume of trades for Citigroup is at the middle and end of the month, and the lowest volume of trades are at the beginning of the new month and the third week of the month.

Lets look at the statistics of citigroup.

```
summary(C_stock)

##     C.Close          C.Volume          C_ROI_dollars    C_dailyChange
##  Min.   : 10.20   Min.   :  1005100   Min.   :-500.3   Min.   :-298.300
##  1st Qu.: 41.80   1st Qu.: 13019600   1st Qu.:-468.7   1st Qu.:  -0.680
##  Median : 51.49   Median : 19493900   Median :-459.0   Median :  -0.010
##  Mean   : 93.38   Mean   : 26987469   Mean   :-417.1   Mean   :   0.021
##  3rd Qu.: 69.46   3rd Qu.: 33280800   3rd Qu.:-441.0   3rd Qu.:   0.650
##  Max.   :552.50   Max.   :377263800   Max.   :  42.0   Max.   : 510.500
##
##     MonthYear    portfolio_DailyValue portfolio_prevDay
##  Aug-2007:  23   Min.   :1229         Min.   :1229
##  Aug-2011:  23   1st Qu.:2821         1st Qu.:2821
##  Aug-2012:  23   Median :3542         Median :3541
##  Aug-2016:  23   Mean   :3988         Mean   :3986
##  Aug-2017:  23   3rd Qu.:5104         3rd Qu.:5104
##  Aug-2018:  23   Max.   :7910         Max.   :7910
##  (Other) :3155
##  portfolio_dailyValueChange portfolio_ROI_dollars      Date
##  Min.   :-1014.322          Min.   :-1748.9       Min.   :2007-01-03
##  1st Qu.:  -39.065          1st Qu.: -157.4       1st Qu.:2010-04-12
##  Median :    2.276          Median :  563.9       Median :2013-07-18
##  Mean   :    1.475          Mean   : 1009.6       Mean   :2013-07-16
##  3rd Qu.:   43.517          3rd Qu.: 2126.4       3rd Qu.:2016-10-21
##  Max.   : 1025.453          Max.   : 4931.7       Max.   :2020-01-31
##
##   DayOfWeek             Month                Year      UE_monthlyRate
##  Length:3293        Length:3293        Min.   :2007   Min.   : 3.500
##  Class :character   Class :character   1st Qu.:2010   1st Qu.: 4.600
##  Mode  :character   Mode  :character   Median :2013   Median : 5.600
##                                        Mean   :2013   Mean   : 6.282
##                                        3rd Qu.:2016   3rd Qu.: 8.200
##                                        Max.   :2020   Max.   :10.000
##
##  portfolio_DailyVolume portfolio_prevDayVolume portfolio_dailyVolumeChange
##  Min.   :1.133e+08     Min.   :1.133e+08       Min.   :-714176400
##  1st Qu.:3.370e+08     1st Qu.:3.370e+08       1st Qu.: -50722061
##  Median :4.194e+08     Median :4.196e+08       Median :    250560
##  Mean   :4.752e+08     Mean   :4.753e+08       Mean   :    -55791
##  3rd Qu.:5.716e+08     3rd Qu.:5.716e+08       3rd Qu.:  50561500
##  Max.   :1.611e+09     Max.   :1.611e+09       Max.   : 620907605
##
##  portfolio_VolumeRatioDaily2Initial portfolio_ValueRatioDaily2Initial
##  Min.   :0.1981                     Min.   :0.4236
##  1st Qu.:0.5891                     1st Qu.:0.9720
##  Median :0.7333                     Median :1.2206
##  Mean   :0.8307                     Mean   :1.3742
##  3rd Qu.:0.9992                     3rd Qu.:1.7591
```

```
##  Max.   :2.8163                     Max.   :2.7259
##
##  portfolio_DailyRatios_X_UE   dayOfMonth    portfolio_poisson
##  Min.   : 0.9658           Min.   : 1.00   Min.   :0.03177
##  1st Qu.: 4.4923           1st Qu.: 8.00   1st Qu.:0.07392
##  Median : 5.6528           Median :16.00   Median :0.22652
##  Mean   : 6.4285           Mean   :15.74   Mean   :0.19506
##  3rd Qu.: 7.8497           3rd Qu.:23.00   3rd Qu.:0.29808
##  Max.   :24.2627           Max.   :31.00   Max.   :0.36217
##
```

From the above summary statistics of Citigroup, we see the min, quantiles, median, mean, and max numeric values as well as length and class type for the non-numeric features of this data set.

Some interesting insights into the above table are that considering an initial investment of 510 USD, the return on the initial investment in dollars is almost the entire amount invested but not quite. Definitely about 80% from the quantile and statistics on the ROI column.

The daily changes fluctuated from a loss of 298 USD in one day to a profit of 510 USD on another day. These are good indicators of where to look on these days, to see if the public sentiment on these dates for Citigroup would indicate more people getting rid of their Citi stock or buying up more of it.

Also, the max and min volume of stock is much more and less respectively than the median volume of trades for this Citigroup stock. These dates for information would also be an interesting place to start to find a pattern with buying/selling stock and combining web scraped text from news articles and comments about Citigroup on those dates.

First, we should grab those points of interest in the data and create a table to compare these values.

```
C_stock_minmaxValueChanges <- subset(C_stock,

C_stock$C_dailyChange==min(C_stock$C_dailyChange) |

C_stock$C_dailyChange==max(C_stock$C_dailyChange) |

C_stock$C.Volume==min(C_stock$C.Volume) |

C_stock$C.Volume==max(C_stock$C.Volume))
C_stock_minmaxValueChanges
```

```
##            C.Close  C.Volume C_ROI_dollars C_dailyChange MonthYear
## 2007-04-02  510.50   2282100          0.00    510.500000  Apr-2007
## 2013-04-02   44.11   1005100       -466.39      0.320000  Apr-2013
## 2015-12-28   52.38 377263800       -458.12     -0.329998  Dec-2015
## 2008-06-02  214.60  15302800       -295.90   -298.300018  Jun-2008
##            portfolio_DailyValue portfolio_prevDay
```

```
portfolio_dailyValueChange
## 2007-04-02             2901.650          2891.963
9.686608
## 2013-04-02             3433.938          3354.901
79.037872
## 2015-12-28             5005.455          4984.970
20.485009
## 2008-06-02             3120.541          3144.698                 -
24.157199
##            portfolio_ROI_dollars       Date DayOfWeek Month Year
UE_monthlyRate
## 2007-04-02             -76.28907 2007-04-02    Monday   Apr 2007
4.5
## 2013-04-02             455.99978 2013-04-02   Tuesday   Apr 2013
7.6
## 2015-12-28            2027.51641 2015-12-28    Monday   Dec 2015
5.0
## 2008-06-02             142.60220 2008-06-02    Monday   Jun 2008
5.6
##            portfolio_DailyVolume portfolio_prevDayVolume
## 2007-04-02             572035712               572035712
## 2013-04-02             258084601               330998801
## 2015-12-28             975152259               752607802
## 2008-06-02             464823559               265152951
##            portfolio_dailyVolumeChange portfolio_VolumeRatioDaily2Initial
## 2007-04-02                           0                          1.0000000
## 2013-04-02                   -72914200                          0.4511687
## 2015-12-28                   222544457                          1.7047052
## 2008-06-02                   199670608                          0.8125779
##            portfolio_ValueRatioDaily2Initial portfolio_DailyRatios_X_UE
## 2007-04-02                          1.000000                   4.500000
## 2013-04-02                          1.183444                   4.057888
## 2015-12-28                          1.725038                  14.703404
## 2008-06-02                          1.075437                   4.893707
##            dayOfMonth portfolio_poisson
## 2007-04-02          2           0.25773
## 2013-04-02          2           0.33619
## 2015-12-28         28           0.27468
## 2008-06-02          2           0.06828
```

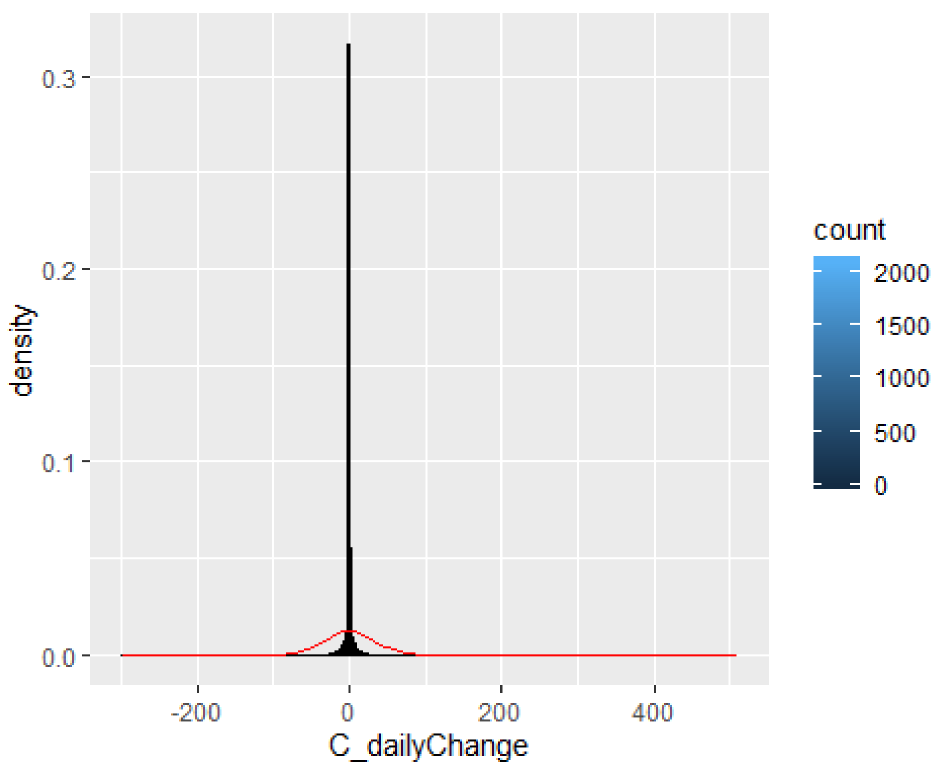
From the above information, Monday is the day of the week with the highest and lowest daily change, as well as the highest volume of trade. Tuesday is the day with the lowest volume of trade. The dates to pull an internet search of news articles about Citigroup to analyze public sentiment on Citi stock are:

- April 2, 2007
- April 2, 2013
- December 28, 2015
- June 2, 2008

This should be interesting to see what type of articles are available on line with a google search of those dates and citigroup.

Lets see if there are any other outlier dates to examine by looking at the standard deviation of the daily change on Citigroup stock. We want to see if there are any days where the stock has a daily change more than or less than this amount times three then times two. Because most values will be within the standard deviation for the Gaussian curve.

```
gg <- ggplot(C_stock, aes(x=C_dailyChange))
gg <- gg + geom_histogram(binwidth=2, colour="black",
                          aes(y=..density.., fill=..count..))
gg <- gg + stat_function(fun=dnorm,
                         color="red",
                         args=list(mean=mean(C_stock$C_dailyChange),
                                   sd=sd(C_stock$C_dailyChange)))

gg
```




```
sdC <- sd(C_stock$C_dailyChange)
out <- sdC*3
sdC;out

## [1] 32.16953

## [1] 96.50858
```

The standard error for the daily change in dollars is 32.17 USD and our threshold to find dates outside this normal range of daily change dollar values is 96.51 USD.

Lets add another column to this data set called threshold3 for those daily change values inside the threshold and those outside the threshold.

```
C_stock$Threshold3 <- ifelse(C_stock$C_dailyChange < out, 'inside','outside')

C_outer_SD <- subset(C_stock, C_stock$Threshold3=='outside')
summary(C_outer_SD)
```

```
##     C.Close          C.Volume          C_ROI_dollars      C_dailyChange
##  Min.   :330.6   Min.   : 2282100   Min.   :-179.90   Min.   :266.2
##  1st Qu.:471.2   1st Qu.:13456250   1st Qu.: -39.30   1st Qu.:399.6
##  Median :510.6   Median :19551450   Median :   0.15   Median :441.4
##  Mean   :488.2   Mean   :30425167   Mean   : -22.32   Mean   :424.4
##  3rd Qu.:542.8   3rd Qu.:35952375   3rd Qu.:  32.27   3rd Qu.:475.4
##  Max.   :552.5   Max.   :81343800   Max.   :  42.00   Max.   :510.5
##
##     MonthYear portfolio_DailyValue portfolio_prevDay
portfolio_dailyValueChange
##  Apr-2007:1   Min.   :2724         Min.   :2744      Min.   :-85.034
##  Aug-2007:1   1st Qu.:2899         1st Qu.:2878      1st Qu.: -4.048
##  Dec-2007:1   Median :2974         Median :2942      Median : -1.393
##  Feb-2007:1   Mean   :3104         Mean   :3044      Mean   : 59.150
##  Jan-2007:1   3rd Qu.:3343         3rd Qu.:3076      3rd Qu.: 20.755
##  Jul-2007:1   Max.   :3656         Max.   :3619      Max.   :734.207
##  (Other) :6
##  portfolio_ROI_dollars      Date              DayOfWeek
##  Min.   :-253.961      Min.   :2007-01-03   Length:12
##  1st Qu.: -79.356      1st Qu.:2007-03-25   Class :character
##  Median :  -4.371      Median :2007-06-16   Mode  :character
##  Mean   : 125.597      Mean   :2007-06-17
##  3rd Qu.: 364.923      3rd Qu.:2007-09-10
##  Max.   : 677.926      Max.   :2007-12-03
##
##     Month                 Year       UE_monthlyRate  portfolio_DailyVolume
##  Length:12          Min.   :2007   Min.   :4.400   Min.   :2.160e+08
##  Class :character   1st Qu.:2007   1st Qu.:4.500   1st Qu.:3.962e+08
##  Mode  :character   Median :2007   Median :4.600   Median :4.644e+08
##                     Mean   :2007   Mean   :4.617   Mean   :5.398e+08
##                     3rd Qu.:2007   3rd Qu.:4.700   3rd Qu.:6.314e+08
##                     Max.   :2007   Max.   :5.000   Max.   :1.005e+09
##
##  portfolio_prevDayVolume portfolio_dailyVolumeChange
##  Min.   :198190500       Min.   :-197842207
##  1st Qu.:387785669       1st Qu.: -23781530
##  Median :564614969       Median :  26069930
##  Mean   :528884214       Mean   :  10878309
##  3rd Qu.:594041737       3rd Qu.:  70618878
```

```
##  Max.   :971072459       Max.   : 124348468
##
##  portfolio_VolumeRatioDaily2Initial portfolio_ValueRatioDaily2Initial
##  Min.   :0.3776                     Min.   :0.9388
##  1st Qu.:0.6926                     1st Qu.:0.9989
##  Median :0.8118                     Median :1.0248
##  Mean   :0.9436                     Mean   :1.0696
##  3rd Qu.:1.1038                     3rd Qu.:1.1521
##  Max.   :1.7576                     Max.   :1.2599
##
##  portfolio_DailyRatios_X_UE   dayOfMonth   portfolio_poisson  Threshold3
##  Min.   :1.654              Min.   :1.00   Min.   :0.04659   Length:12
##  1st Qu.:3.696              1st Qu.:1.00   1st Qu.:0.05008   Class
:character
##  Median :4.400              Median :1.00   Median :0.05454   Mode
:character
##  Mean   :4.641              Mean   :1.75   Mean   :0.13836
##  3rd Qu.:5.116              3rd Qu.:2.25   3rd Qu.:0.25948
##  Max.   :8.297              Max.   :4.00   Max.   :0.26474
##
```

We can see from the above statistics on the subset of Citigroup stock that are outside this threshold that there are 12 dates to select in the range of Jan 2007 through Sep 2008. So we will add those dates to our data set of text scraped news articles on Citigroup.

```
NLP_dates_Citi <- rbind(C_stock_minmaxValueChanges, C_outer_SD[,-23])
NLP_dates_Citi

##             C.Close  C.Volume C_ROI_dollars C_dailyChange MonthYear
## 2007-04-02   510.50   2282100      0.000000    510.500000  Apr-2007
## 2013-04-02    44.11   1005100   -466.389999      0.320000  Apr-2013
## 2015-12-28    52.38 377263800   -458.119999     -0.329998  Dec-2015
## 2008-06-02   214.60  15302800   -295.899994   -298.300018  Jun-2008
## 2007-04-021  510.50   2282100      0.000000    510.500000  Apr-2007
## 2007-08-01   468.50  13495700    -42.000000    397.800003  Aug-2007
## 2007-12-03   330.60  81343800   -179.899994    266.250008  Dec-2007
## 2007-02-01   547.30  80864600     36.799988    467.409989  Feb-2007
## 2007-01-03   552.50  43508100     42.000000    488.520000  Jan-2007
## 2007-07-02   516.40  32822200      5.900024    441.990020  Jul-2007
## 2007-06-01   545.10  23057000     34.599976    473.939972  Jun-2007
## 2007-03-01   510.80   8981300      0.299988    440.769989  Mar-2007
## 2007-05-01   542.00  13337900     31.500000    479.779999  May-2007
## 2007-11-01   385.10  33433800   -125.399994    322.950004  Nov-2007
## 2007-10-01   477.20  16045900    -33.299988    402.080009  Oct-2007
## 2007-09-04   472.10  15929600    -38.399994    400.240005  Sep-2007
##             portfolio_DailyValue portfolio_prevDay
portfolio_dailyValueChange
## 2007-04-02             2901.650          2891.963
9.686608
## 2013-04-02             3433.938          3354.901
```

```
79.037872
## 2015-12-28                5005.455          4984.970                   
20.485009
## 2008-06-02                3120.541          3144.698                 -
24.157199
## 2007-04-021               2901.650          2891.963                   
9.686608
## 2007-08-01                2778.299          2781.133                  -
2.834138
## 2007-12-03                2723.978          2743.972                 -
19.993872
## 2007-02-01                3279.015          3281.965                  -
2.949476
## 2007-01-03                2977.939          2977.939                   
0.000000
## 2007-07-02                2969.196          2946.619                   
22.576765
## 2007-06-01                3003.989          3006.774                  -
2.785581
## 2007-03-01                2889.381          2896.725                  -
7.344424
## 2007-05-01                2957.539          2937.392                   
20.147648
## 2007-11-01                3534.398          3619.433                 -
85.034241
## 2007-10-01                3655.864          3611.738                   
44.126353
## 2007-09-04                3571.178          2836.972                   
734.206543
##              portfolio_ROI_dollars       Date DayOfWeek Month Year
## 2007-04-02              -76.289072 2007-04-02    Monday   Apr 2007
## 2013-04-02              455.999776 2013-04-02   Tuesday   Apr 2013
## 2015-12-28             2027.516411 2015-12-28    Monday   Dec 2015
## 2008-06-02              142.602196 2008-06-02    Monday   Jun 2008
## 2007-04-021             -76.289072 2007-04-02    Monday   Apr 2007
## 2007-08-01             -199.639490 2007-08-01 Wednesday   Aug 2007
## 2007-12-03             -253.960930 2007-12-03    Monday   Dec 2007
## 2007-02-01              301.076786 2007-02-01  Thursday   Feb 2007
## 2007-01-03                0.000000 2007-01-03 Wednesday   Jan 2007
## 2007-07-02               -8.742779 2007-07-02    Monday   Jul 2007
## 2007-06-01               26.049900 2007-06-01    Friday   Jun 2007
## 2007-03-01              -88.557542 2007-03-01  Thursday   Mar 2007
## 2007-05-01              -20.399119 2007-05-01   Tuesday   May 2007
## 2007-11-01              556.459753 2007-11-01  Thursday   Nov 2007
## 2007-10-01              677.925528 2007-10-01    Monday   Oct 2007
## 2007-09-04              593.239860 2007-09-04   Tuesday   Sep 2007
##              UE_monthlyRate portfolio_DailyVolume portfolio_prevDayVolume
## 2007-04-02              4.5             572035712               572035712
## 2013-04-02              7.6             258084601               330998801
## 2015-12-28              5.0             975152259               752607802
```

```
## 2008-06-02            5.6            464823559                 265152951
## 2007-04-021           4.5            572035712                 572035712
## 2007-08-01            4.6            686001371                 572681959
## 2007-12-03            5.0           1005429691                 971072459
## 2007-02-01            4.5            933350159                 809001691
## 2007-01-03            4.6            613250413                 565411759
## 2007-07-02            4.7            460278863                 658121070
## 2007-06-01            4.6            381151267                 397701502
## 2007-03-01            4.4            215973129                 198190500
## 2007-05-01            4.4            314742689                 233827359
## 2007-11-01            4.7            468477291                 563818179
## 2007-10-01            4.7            401234791                 446710205
## 2007-09-04            4.7            425224899                 358038171
##             portfolio_dailyVolumeChange portfolio_VolumeRatioDaily2Initial
## 2007-04-02                            0                          1.0000000
## 2013-04-02                    -72914200                          0.4511687
## 2015-12-28                    222544457                          1.7047052
## 2008-06-02                    199670608                          0.8125779
## 2007-04-021                           0                          1.0000000
## 2007-08-01                    113319412                          1.1992282
## 2007-12-03                     34357232                          1.7576345
## 2007-02-01                    124348468                          1.6316292
## 2007-01-03                     47838654                          1.0720492
## 2007-07-02                   -197842207                          0.8046331
## 2007-06-01                    -16550235                          0.6663068
## 2007-03-01                     17782629                          0.3775518
## 2007-05-01                     80915330                          0.5502151
## 2007-11-01                    -95340888                          0.8189651
## 2007-10-01                    -45475414                          0.7014156
## 2007-09-04                     67186728                          0.7433538
##             portfolio_ValueRatioDaily2Initial portfolio_DailyRatios_X_UE
## 2007-04-02                          1.0000000                   4.500000
## 2013-04-02                          1.1834435                   4.057888
## 2015-12-28                          1.7250378                  14.703404
## 2008-06-02                          1.0754368                   4.893707
## 2007-04-021                         1.0000000                   4.500000
## 2007-08-01                          0.9574896                   5.281943
## 2007-12-03                          0.9387687                   8.250061
## 2007-02-01                          1.1300522                   8.297218
## 2007-01-03                          1.0262916                   5.061081
## 2007-07-02                          1.0232786                   3.869810
## 2007-06-01                          1.0352692                   3.173112
## 2007-03-01                          0.9957719                   1.654204
## 2007-05-01                          1.0192614                   2.467577
## 2007-11-01                          1.2180652                   4.688499
## 2007-10-01                          1.2599262                   4.153540
## 2007-09-04                          1.2307408                   4.299916
##             dayOfMonth portfolio_poisson
## 2007-04-02           2           0.25773
## 2013-04-02           2           0.33619
```

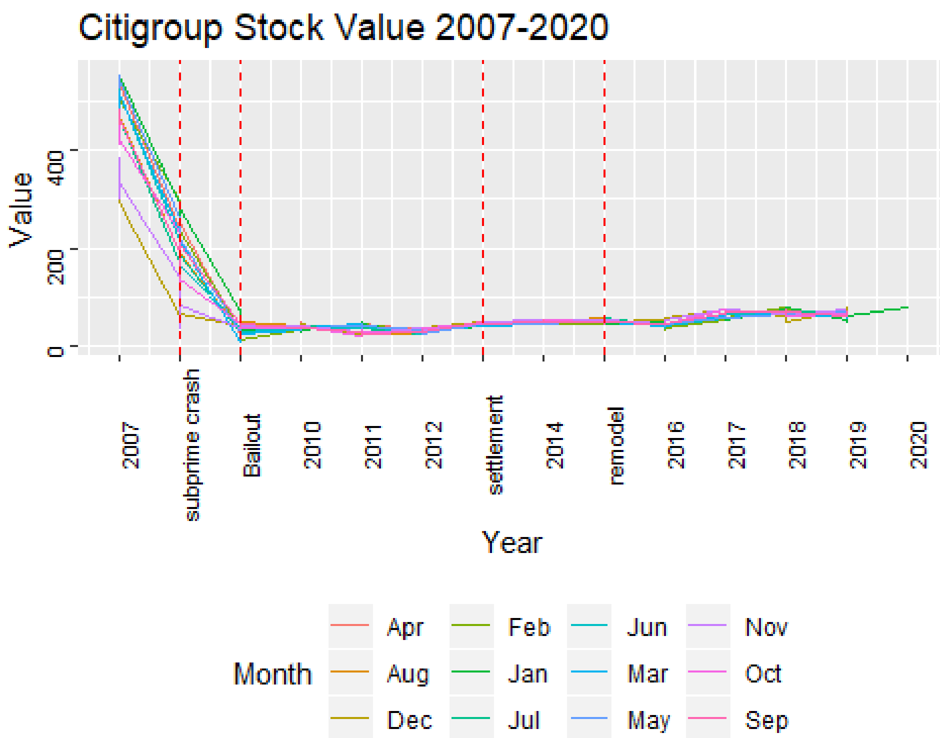
```
## 2015-12-28          28          0.27468
## 2008-06-02           2          0.06828
## 2007-04-021          2          0.25773
## 2007-08-01           1          0.05008
## 2007-12-03           3          0.05723
## 2007-02-01           1          0.04833
## 2007-01-03           3          0.05008
## 2007-07-02           2          0.26474
## 2007-06-01           1          0.05008
## 2007-03-01           1          0.04659
## 2007-05-01           1          0.25411
## 2007-11-01           1          0.05184
## 2007-10-01           1          0.26474
## 2007-09-04           4          0.26474
```

I am going to pull the data from these dates with the Google Search for the specific date on Citigroup stock, put it in a table with the date, the article title, reference, article content, and the comments if available.

Note: when searching the internet, there were limited articles and most were about Citi's involvement in the sub-prime mortgage crisis of 2007-2008, and a bailout of Citigroup by the US. For the month and years of the two dates not in or around 2007-2008, there are only two for April 2013 and December 2015. Where Citi settled a lawsuit for covering up bad mortgage loans in August 2012 and a person reported on a forum about FICO scores how he was approved for a 4600 USD credit card with Citi. There isn't enough data to rely on the web for NLP on Citigroup for these time frames.

Lets plot this as a simple line chart of the value of the stock over the years.

```
ggplot(data = C_stock, aes(x=Year, y=C.Close, group=Month)) +
  geom_line(aes(color=Month))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+

scale_x_continuous(breaks=c(2007,2008,2009,2010,2011,2012,2013,2014,2015,2016
,2017,2018,2019,2020),
                     labels=c(2007,'subprime
crash','Bailout',2010,2011,2012,'settlement',2014,'remodel',2016,2017,2018,20
19,2020))+
  theme(axis.text = element_text(colour = "black", angle=90, size =
rel(.75)))+
  geom_vline(xintercept=c(2008,2009,2013,2015), linetype='dashed',
color='red')+
  ggtitle('Citigroup Stock Value 2007-2020')+
  ylab('Value')

## Warning in pal_name(palette, type): Unknown palette paired
```



We could pull based on the keywords: 'settlement', 'bail-out', 'sub-prime loans', but we would only get the obvious negative sentiment for these keywords. A New York Times article posted an article in Dec 2015 about the remodeling that Citigroup was doing to their offices, but the full article would have to be purchased. The fact that they spent money on remodeling could have some public sentiment of either they aren't distributing their profits to shareholders or they are making enough profits to spend money on remodeling, which is also reported at the end of the year in 2015 to write off for that tax year. Although, I was told by an accountant that some corporations and small businesses have a different tax year and a quick search on Google returned the fiscal year is any consecutive 12-month business cycle that usually ends at the end of each quarter.

We can see that the volume of trades is highest in December 2015 from our dates, but we should compare this to which quantile this number is within for the volume of trades of Citi stock.

```
summary(C_stock$C.Volume)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##   1005100  13019600  19493900  26987469  33280800 377263800
```

We already know that this is the date that the most trades in stock of Citi occured as it is the reason we added this date to our NLP data set of dates to pull information from the web for. The above will refresh the comparisons of the trade volume to this date.

It looks like public sentiment thinks Citi is going back to its old bail-out days of 2007-2008 and not a trust-worthy stock for their personal portfolios. But they are still around, and the fact that people that have a less than trust-worthy credit profile were given a credit card with a high value could indicate some people also consider that they are building a new demographic of people to invest in by earning the trust of those who have sub-par trust worthiness with credit. And, yet some other investors could also think this is a bad move to make as it depends on those same people realizing their mistakes and not making them again. Which really turns into the reason some stocks are volatile to begin with and possibly a reason to understand Game Theory, a class I dropped in my undergrad college. But nonetheless I am a data scientist with other coventional and non-conventional ways of extracting useful information, and this approach uses my math and analytic skills to fully understand the stock market and certain stocks and trends with public sentiment.

On this highest trade day, the daily change in dollars was still within the standard error by only dropping 0.33 USD. Where the standard error is 32.00 USD.

Of note is whether or not those making these trades are doing so to lower their Capital Gains at the end of the year, because there is a slight loss on it to balance out the portfolio. Also, this is the end of the year, possibly the last trading day of the year as it is. Lets look at all monthYear dates equal to Dec-2015 to see if there are any other dates past Dec 28, 2015.

```
dec2015 <- subset(C_stock, C_stock$MonthYear=='Dec-2015')
tail(dec2015)
```

```
##            C.Close  C.Volume C_ROI_dollars C_dailyChange MonthYear
## 2015-12-23   52.63  93423000       -457.87      0.620003  Dec-2015
## 2015-12-24   52.71 119108100       -457.79      0.079998  Dec-2015
## 2015-12-28   52.38 377263800       -458.12     -0.329998  Dec-2015
## 2015-12-29   52.98 281369700       -457.52      0.599999  Dec-2015
## 2015-12-30   52.30  62625000       -458.20     -0.680001  Dec-2015
## 2015-12-31   51.75  49092600       -458.75     -0.549999  Dec-2015
##            portfolio_DailyValue portfolio_prevDay
portfolio_dailyValueChange
## 2015-12-23             4998.690          4968.045
30.64500
## 2015-12-24             4984.970          4998.690                  -
13.72002
## 2015-12-28             5005.455          4984.970
20.48501
## 2015-12-29             4738.190          5005.455                 -
267.26507
## 2015-12-30             4800.285          4738.190
62.09506
## 2015-12-31             4707.685          4800.285                  -
92.59999
##            portfolio_ROI_dollars       Date DayOfWeek Month Year
UE_monthlyRate
## 2015-12-23              2020.751 2015-12-23 Wednesday   Dec 2015
5
## 2015-12-24              2007.031 2015-12-24  Thursday   Dec 2015
```

```
5
## 2015-12-28              2027.516 2015-12-28    Monday   Dec 2015
5
## 2015-12-29              1760.251 2015-12-29   Tuesday   Dec 2015
5
## 2015-12-30              1822.346 2015-12-30 Wednesday   Dec 2015
5
## 2015-12-31              1729.746 2015-12-31  Thursday   Dec 2015
5
##            portfolio_DailyVolume portfolio_prevDayVolume
## 2015-12-23             903674159               619024059
## 2015-12-24             752607802               903674159
## 2015-12-28             975152259               752607802
## 2015-12-29            1248436459               975152259
## 2015-12-30             534260059              1248436459
## 2015-12-31             504630159               534260059
##            portfolio_dailyVolumeChange portfolio_VolumeRatioDaily2Initial
## 2015-12-23                   284650100                          1.5797513
## 2015-12-24                  -151066357                          1.3156658
## 2015-12-28                   222544457                          1.7047052
## 2015-12-29                   273284200                          2.1824450
## 2015-12-30                  -714176400                          0.9339628
## 2015-12-31                   -29629900                          0.8821655
##            portfolio_ValueRatioDaily2Initial portfolio_DailyRatios_X_UE
## 2015-12-23                          1.722706                  13.607238
## 2015-12-24                          1.717978                  11.301424
## 2015-12-28                          1.725038                  14.703404
## 2015-12-29                          1.632930                  17.818897
## 2015-12-30                          1.654330                   7.725412
## 2015-12-31                          1.622417                   7.156201
##            dayOfMonth portfolio_poisson Threshold3
## 2015-12-23         23           0.27468     inside
## 2015-12-24         24           0.05723     inside
## 2015-12-28         28           0.27468     inside
## 2015-12-29         29           0.05723     inside
## 2015-12-30         30           0.27468     inside
## 2015-12-31         31           0.05723     inside
```

We now know that Dec-28-2015 is not the last trading day of the year, because the 29th through 31st for Tuesday through Thursday are also trading days. There was a fluctuation in dollars earned and lost all under a dollar. Some useful information to add in would be who or where are these trades derived. Are they financial advisors, trust fund managers, independent investors, foreign or national investors, are they hobbyists just playing the stock market on an e-trade, are they educated, experienced, and so on?

To get this information we could first find out how much it costs for a hobbyist to make a trade online from e-trade or similar and whether or not this information is shared on demographics of the stock ownership. We could also look at the American Survey on Census data from the census bureau for numer of financial workers there are and how many people graduated with a BS, MS, or Phd in Finance or Economics. If there is location data on where these stock owners live attach this information gathered to it to make a better inference on this stock and what motivates the trades. Any volunteers?

For now, we will just continue with what we have on hand for Citi. We can answer the question of whether or not, historically there are more trades in December than any other month in our data by grouping by month year and getting the median trades per month and year.

```
Citi_trades_monthYear <- C_stock %>% group_by(MonthYear) %>%
  summarise_at(vars(colnames(C_stock[2])), mean)
Citi_trades_monthYear <-
Citi_trades_monthYear[order(Citi_trades_monthYear$C.Volume,decreasing=TRUE),]
Citi_trades_monthYear
```

```
## # A tibble: 157 x 2
##    MonthYear   C.Volume
##    <fct>          <dbl>
##  1 Dec-2011  102284343.
##  2 Dec-2012   97253820
##  3 Feb-2007   94010711.
##  4 Feb-2008   80151765
##  5 Dec-2019   79458262.
##  6 Aug-2019   72849682.
##  7 Feb-2015   70393405.
##  8 Dec-2015   67380332.
##  9 Jan-2010   64943774.
## 10 Jan-2012   63211745
## # … with 147 more rows
```

From the above table ordered from most trades to least trades per month and year by mean number of trades per month, we see that December is in the top 10 month years of high trades in 2011,2012, 2015, and 2019. February has the next highest trades but the years are the same years of the sub-prime mortgage crisis that Citigroup was involved in, but also in 2015. looking at the next top ten months we see that Dec, Jan, and Feb are in the highest mean of the trades per day grouped by month and year. What do we know about Jan and Feb outside of the assumption about December being the last day of the tax year to offset capital gains with capital losses?

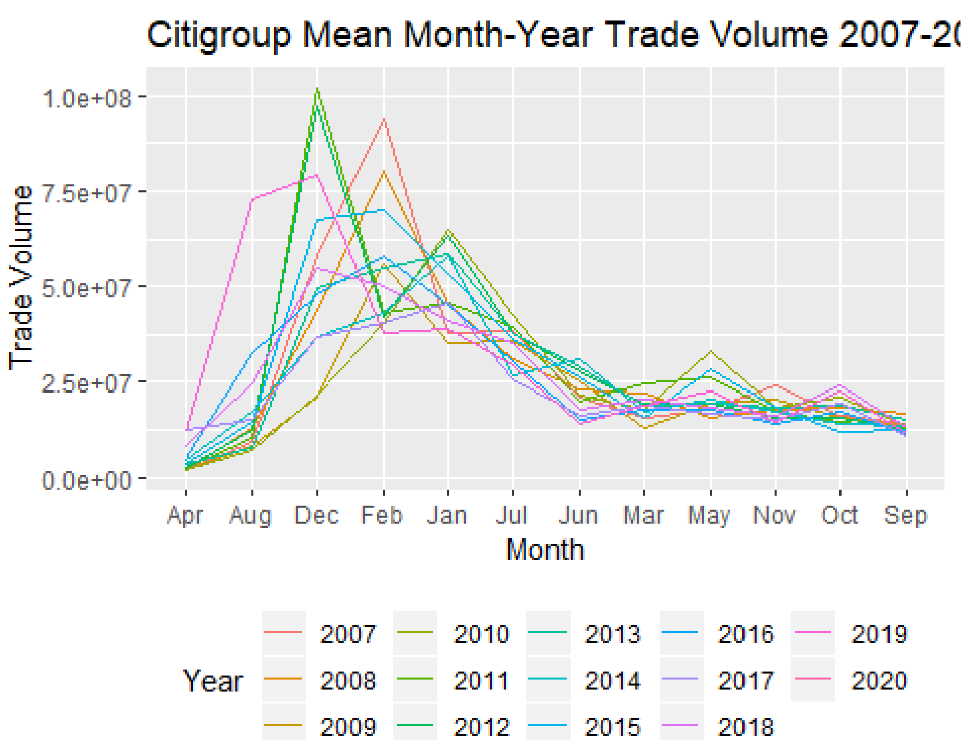
Well, I know that being a student, some people get their student loans around winter quarter in January and that many people expecting tax refunds get their refunds in February. We would have to see if there are any other assumptions about these months. But we would be able to ascertain if students receiving an education are investing, and if consumers with tax refunds are using some of that money to invest.There are certainly other assumptions that could be made for why the last month of the year and the first two months of the first quarter are high trade volume days. But for now lets stick with these assumptions.

July starts to show up in the following set of ten top month years from 21-30, as the 30th highest trade month year. Jan and Feb are still in the top 40 high volume trade month years, while June shows up three times in the 30-40 top high volume trade month and years. July could also be the start of the third quarter and the remaining balance on student loans made. Lets see where September/October show up in these top ordered volumes. They are near the end of the top trade months.

So, possibly this indicates no ties to student loan payments, but tax refunds could be likely for February being a high trade month. We definitely know December is a top trade day.

Lets plot this data.

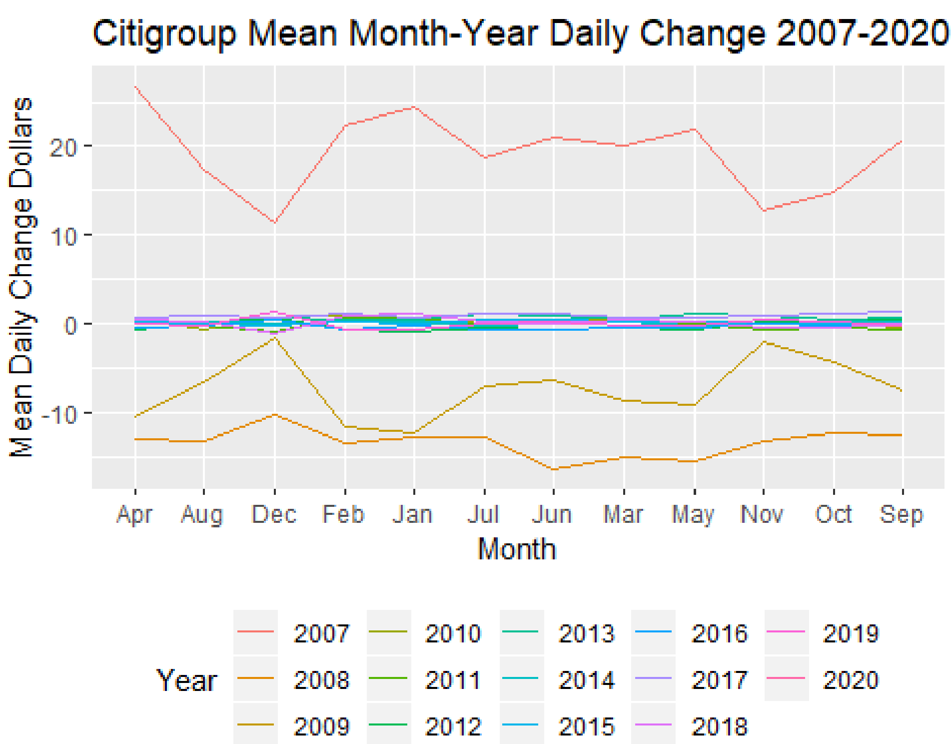
```
Citi_trades_monthYear$Month <- gsub('-[0-
9]{4}','',Citi_trades_monthYear$MonthYear)
Citi_trades_monthYear$Year <- gsub('[a-zA-z]{3}-
','',Citi_trades_monthYear$MonthYear)

ggplot(data = Citi_trades_monthYear, aes(x=Month, y=C.Volume,group=Year)) +
  geom_line(aes(color=Year))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Citigroup Mean Month-Year Trade Volume 2007-2020')+
  ylab('Trade Volume')

## Warning in pal_name(palette, type): Unknown palette paired
```



We can see that December is definitely the highest trading month, then February as the next highest, and January as the third highest trading month.

Lets look at the daily change mean values per month, by grouping by MonthYear and taking the mean value of the daily change, order by highest to smallest, and plot.

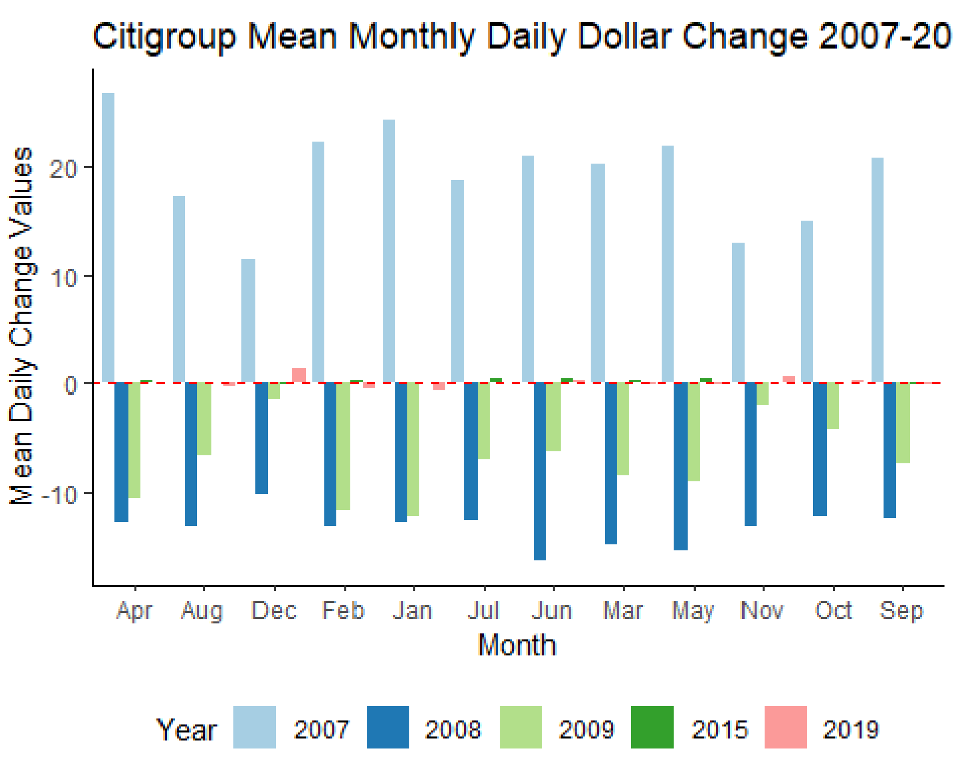
```
Citi_meanMonthly_dailyChange <- C_stock %>% group_by(MonthYear) %>%
  summarise_at(vars(as.vector(colnames(C_stock))[4]), mean)

Citi_meanMonthly_dailyChange$Year <-
  gsub('[a-zA-Z]{3}-','',Citi_meanMonthly_dailyChange$MonthYear)
Citi_meanMonthly_dailyChange$Month <-
  gsub('-[0-9]{4}','',Citi_meanMonthly_dailyChange$MonthYear)


ggplot(data = Citi_meanMonthly_dailyChange, aes(x=Month,
y=C_dailyChange,group=Year)) +
  geom_line(aes(color=Year))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Citigroup Mean Month-Year Daily Change 2007-2020')+
  ylab('Mean Daily Change Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



From the above line chart, it is not obvious what years those years having almost no change are.The year 2007 is at the top with the highest positive mean daily change values fluctuating to around 20 USD per day. While the years 2008 and 2009 have the highest negative mean of daily change values per month with average daily decreases around a daily loss of 5-15 USD.

Lets make a bar chart of 2007, 2008, 2009, 2015, and 2019 of this data on mean daily value changes per month.

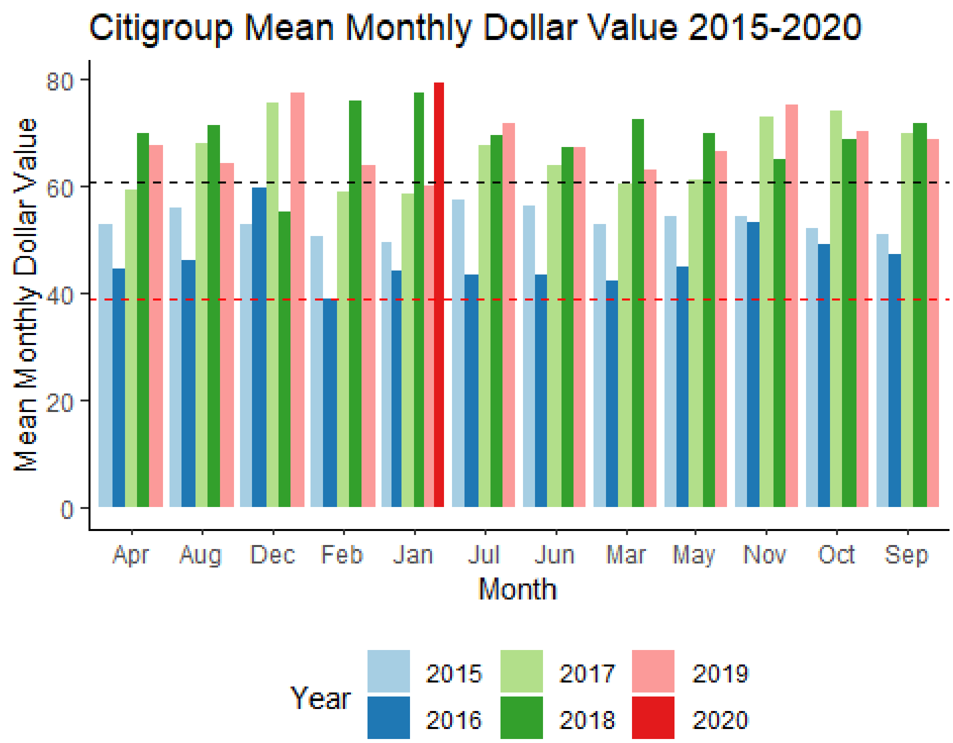
```
y4 <- subset(Citi_meanMonthly_dailyChange,
                              Citi_meanMonthly_dailyChange$Year==2008 |
                              Citi_meanMonthly_dailyChange$Year==2009 |
                              Citi_meanMonthly_dailyChange$Year==2007 |
                              Citi_meanMonthly_dailyChange$Year==2015 |
                              Citi_meanMonthly_dailyChange$Year==2019)
ggplot(data = y4, aes(x=Month, y=C_dailyChange,fill=Year)) +
  geom_bar(stat='identity', position=position_dodge())+
  scale_y_continuous()+
  scale_fill_brewer(palette='Paired') +
  geom_hline(yintercept=0, linetype="dashed", color = "red")+
  theme_classic()+
  theme(legend.position="bottom")+
  ggtitle('Citigroup Mean Monthly Daily Dollar Change 2007-2019')+
  ylab('Mean Daily Change Values')
```



From the above, we can see the Citigroup stock had increases per day in value from the previous day in 2007, but that in 2008 and 2009 those daily increases turned to daily decreases from day to day as the sub-prime loans collapsed that Citigroup held. And in 2015 and 2019 years after Citigroup's bailout there was a mean monthly daily change value next to nothing as the daily change from day to day fluctuated around zero dollars for the month.

This could mean it is gaining strength and remains as is safe to buy as it increases. But lets look at the years 2015-2019 to see how the value of the Citigroup stock has faired by month year to confirm this assertion just made.

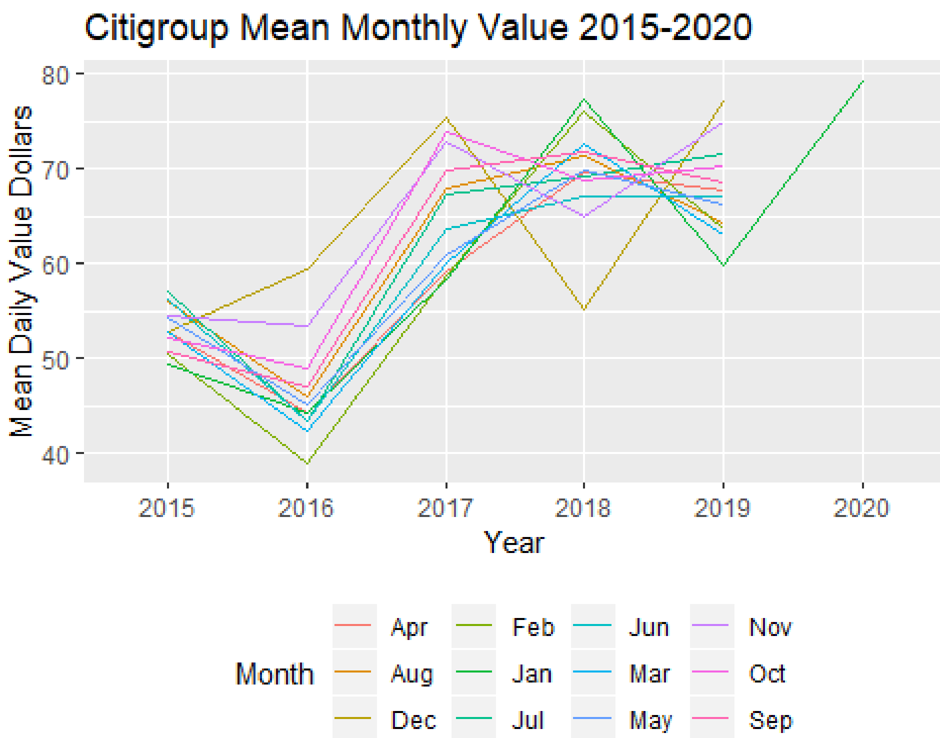
```
y4value <- subset(C_stock, C_stock$Year>2014)
y4valMY <- y4value %>% group_by(MonthYear) %>%
  summarise_at(vars(as.vector(colnames(y4value)[1])), mean)

y4valMY$Year <- gsub('[a-zA-Z]{3}-','', y4valMY$MonthYear)
y4valMY$Month <- gsub('-[0-9]{4}','', y4valMY$MonthYear)


ggplot(data = y4valMY, aes(x=Month, y=C.Close,fill=Year)) +
  geom_bar(stat='identity', position=position_dodge())+
  scale_y_continuous()+
  scale_fill_brewer(palette='Paired') +
  geom_hline(yintercept=min(y4valMY$C.Close), linetype="dashed", color =
"red")+
  geom_hline(yintercept=mean(y4valMY$C.Close), linetype="dashed", color =
"black")+
  theme_classic()+
  theme(legend.position="bottom")+
  ggtitle('Citigroup Mean Monthly Dollar Value 2015-2020')+
  ylab('Mean Monthly Dollar Value')
```



From the above bar chart, we can see that the minimum value is the dashed red line which occured in February 2016. And that every month since 2016 has been above this minimum value. It has almost double from it's minimum value in January and February 2020.The mean value from 2015-2020 (Jan-Feb) is just above 60 USD which is 1 1/2 times its minimum value.

Lets look at the line chart of this by years 2015-2020.

```
ggplot(data = y4valMY, aes(x=Year, y=C.Close,group=Month)) +
  geom_line(aes(color=Month))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Citigroup Mean Monthly Value 2015-2020')+
  ylab('Mean Daily Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



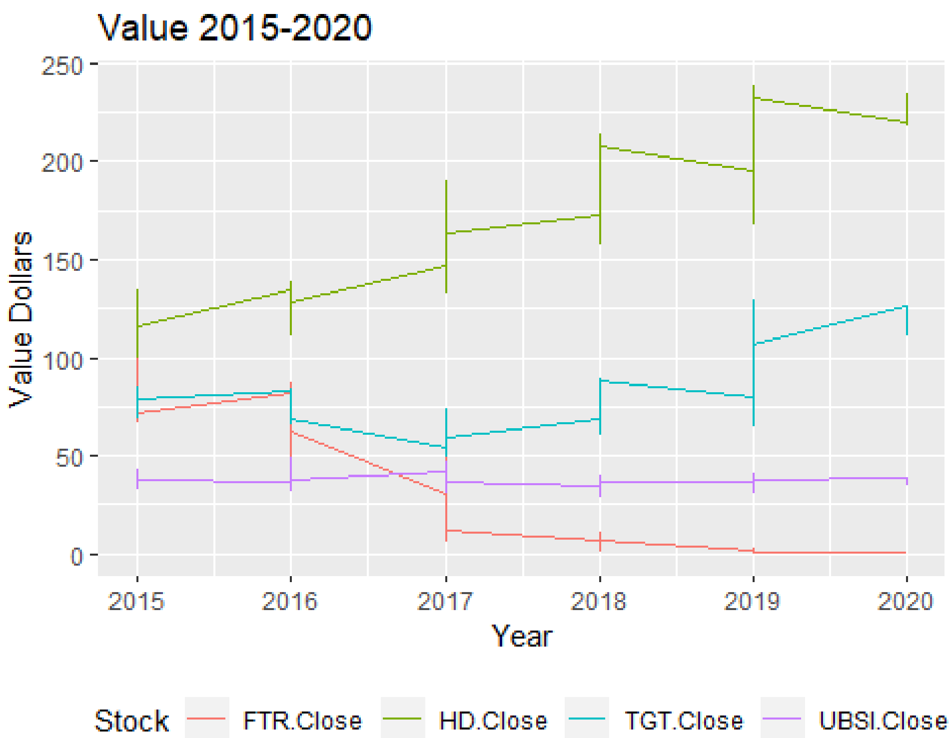
The above line chart of the mean monthly dollar value of the Citigroup stock show that all months move the same direction of decreasing in 2015, increasing in 2016, except for in 2017 and 2018 where 3-6 months decreased and 6-9 months increased monthly mean values. The span of 2019 through 2020 can't be analyzed yet, but January increased since the year prior. Overall, since 2015 the value has increased from 50-60 USD to between 75-80 USD. This could make it a good stock to have in your portfolio as it has steadily been increasing since it's historical rough patches of the sub-prime mortgage loan accounts, the public bailout, and the lawsuit settlement payout. But nothing has been in the news about them to discourage investors from dropping this stock from their stock folder.

We saw that Citigroup is maintaining its current value and slightly increasing over the last four years. Lets start subset sampling stocks and look at the changes they have made in value over the last four years. And see if we notice anything we want to further exploit.

```
Value1 <- StocksSTATS[,c(1:53,160:230)]
Value2 <- subset(Value1, Year>2014)

sub1 <- Value2[,c(1:4,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
```

```
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')
```

```
## Warning in pal_name(palette, type): Unknown palette paired
```



The first four stocks in our set of 53 is shown in the line chart above from 2015-2020.

From the above line chart, it is obvious that over the last five years, the pink line for FTR is a terrible stock as it has been on the decline, but we would have to look at it further to see why it has been decreasing in value since 2015.

The olive color line for HD indicates it has been on a steady increase from the 120-125 USD range in 2015 to the 220-225 USD range in 2020.

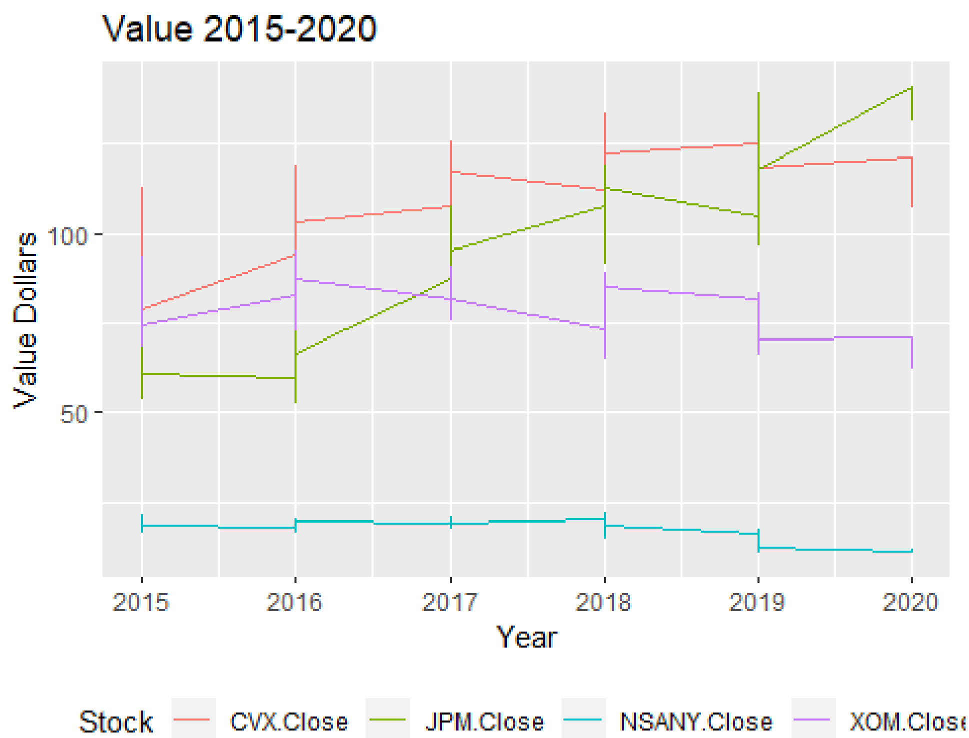
Also, increasing steadily is the blue line for TGT, which started at 75-80 in 2015 and is at 125 in 2020 in value.

The purple line for UBSI has been maintaining steadily from 45 range to 45 range over five years. ***

Lets look at the next four stocks.

```
sub1 <- Value2[,c(5:8,115)]

sub1tidy <- gather(sub1, 'Stock','Value',1:4)
```

```
ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```

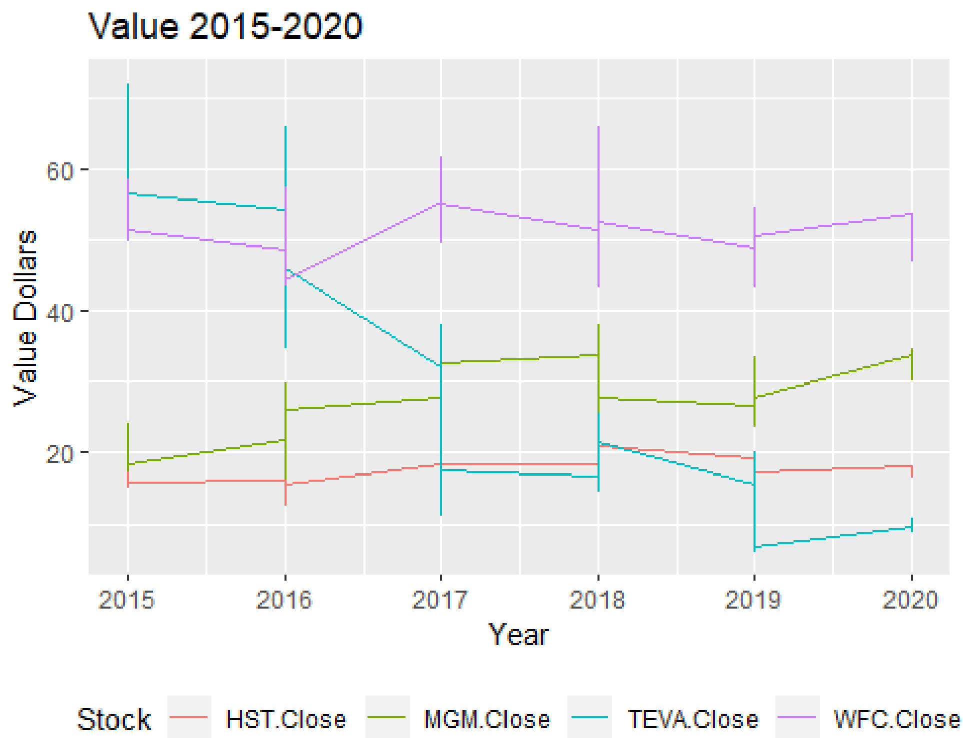


From the above subset of the next four stock in our 53 stocks, we can see that there are two stocks increasing significantly for JPM and CVX. We also note that the XOM and NSANY stocks have decreased over the last five years. ***

Now for the next four stocks.

```
sub1 <- Value2[,c(9:12,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')
```

```
## Warning in pal_name(palette, type): Unknown palette paired
```



The above line chart shows the third subset of four stocks of our 53 stocks.

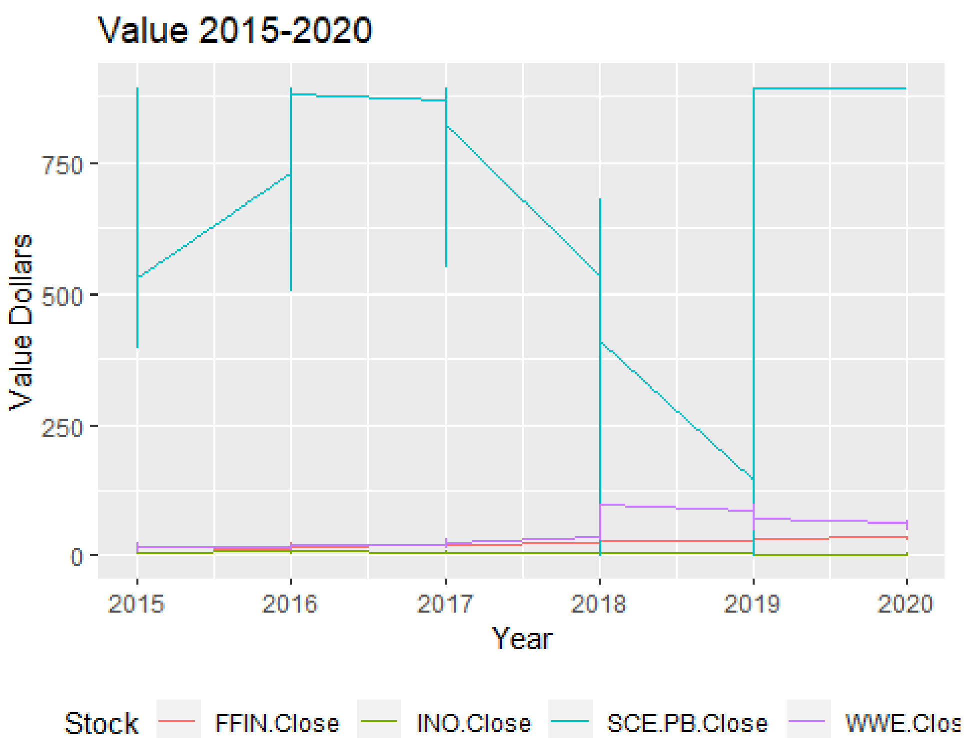
The MGM stock has increased significantly since 2005, and slight increases are shown for WFC and HST though not significantly. There is some cyclical movements in the WFC with 2016 giving a steady increase all year, then declining 2017-2019, and ending with a steady increase in 2019.

The TEVA stock has had a huge loss over the last five years, with the last year showing an an increase slightly. It started at the 55 range in 2015 and is at the 10 range in 2020. This could indicate that it is a good time to buy TEVA, since it is priced low and shows an increase in the last year, where the last four years it has been decreasing annually for each year. This would require further analysis for why it has been decreasing over the last five years. ***

Now for the next four stocks in our subset four.

```
sub1 <- Value2[,c(13:16,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
```

```
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



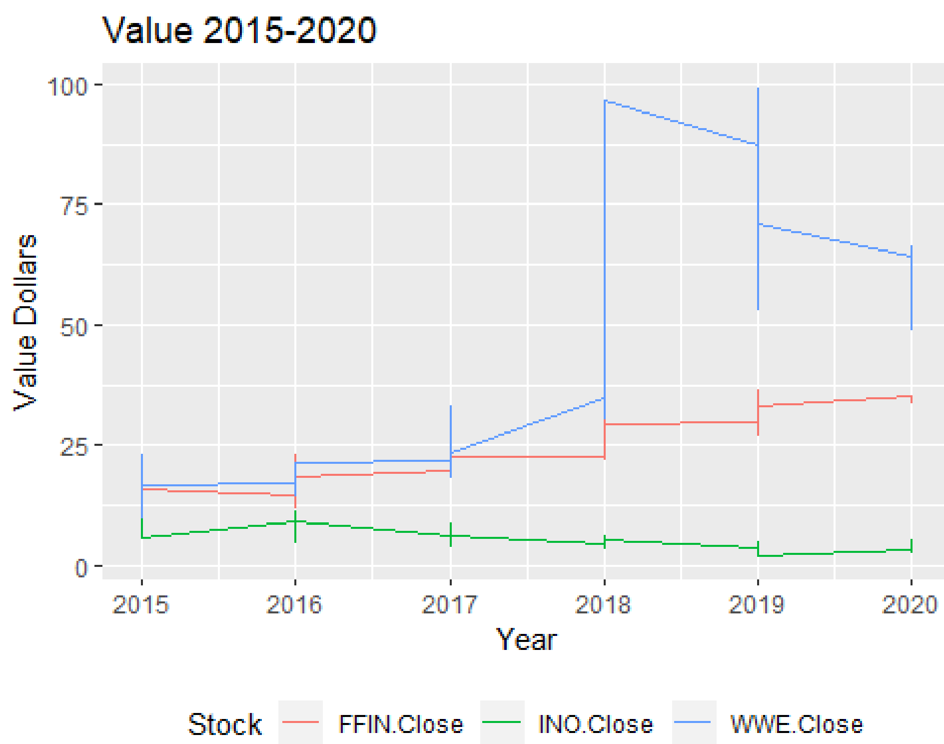
The above line chart shows that SCE.PB is on its own scale that outweighs the scale of the other smaller valued stocks, there is also volatility and cyclical movements in SCE.PB which makes it a good choice to further analyze with timelines of web article events that could have triggered these changes in value of a steady increase in 2015, a high jump increase in 2016, then a steep decline throughout 2017 and 2018, then a huge jump of an increase to the same level at 2016. This is a utility company so government contracts could be involved with all that entails, and possible fires causing damage and settlements in the declining years. But for now it is just speculation and assumptions.

The other stocks are getting limited spotlight above, and they need their own scale as SCE.PB pushed down their scaled visual line charts.

Now for the next four stocks in our subset four.

```
sub1 <- Value2[,c(13,14,16,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:3)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
```

```
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



From the above line chart, we see that WWE had a huge jump in 2018 of an increase from the 40 range to the 90 range but then decreased during 2018 and 2019 to a price still much higher at the 60 range than its starting value in 2015 of the 20 range.

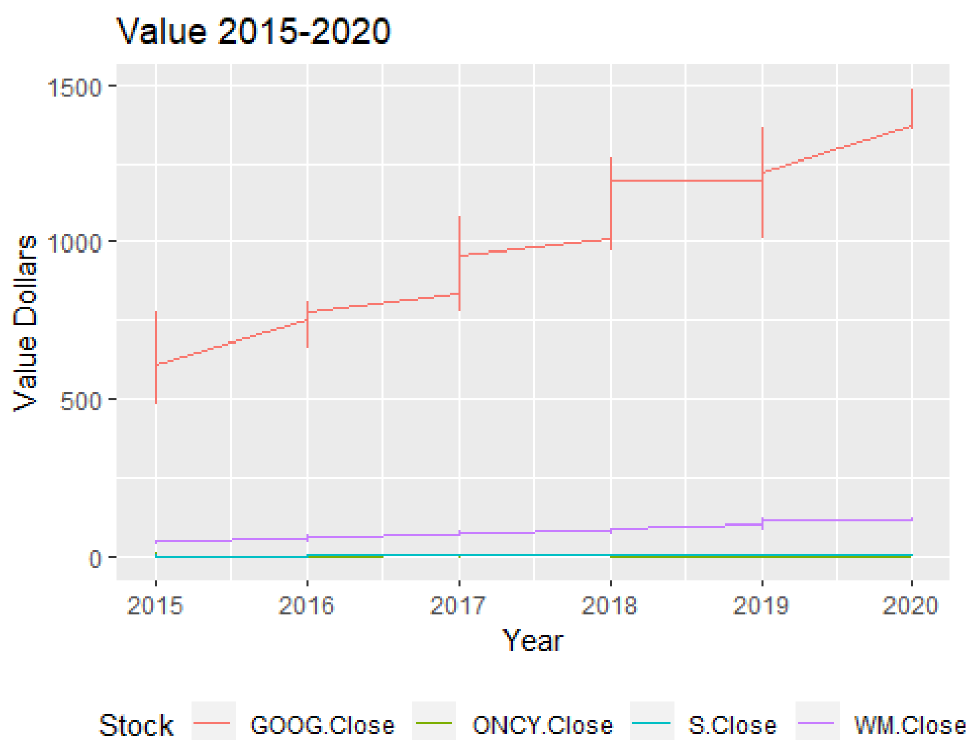
The FFIN stock has been steadily increasing over the last five years with a flat line on the value in 2017 and 2018.

The INO stock has declined since 2016 after an increasing year in 2015, but lost only slightly in value over a five year span returning no profits over that time span.

Now for the next stocks in our subset.

```
sub1 <- Value2[,c(17:20,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
```

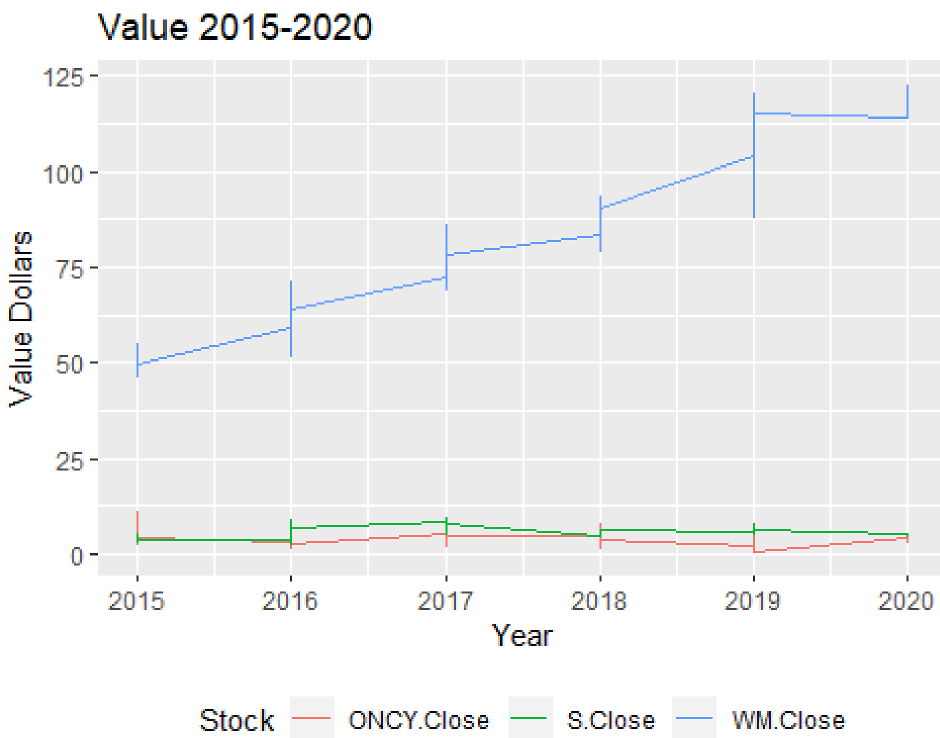
```
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



In the above subset of stocks, Google out scales the other three stocks and shows that it has been increasing steadily every year, except 2018 where it is almost the same price all year.

Lets look at the other three stocks that our on a lower scaled value to analyze them.

```
sub1 <- Value2[,c(18:20,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:3)


ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  #geom_hline(yintercept=m15, color='red')
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```

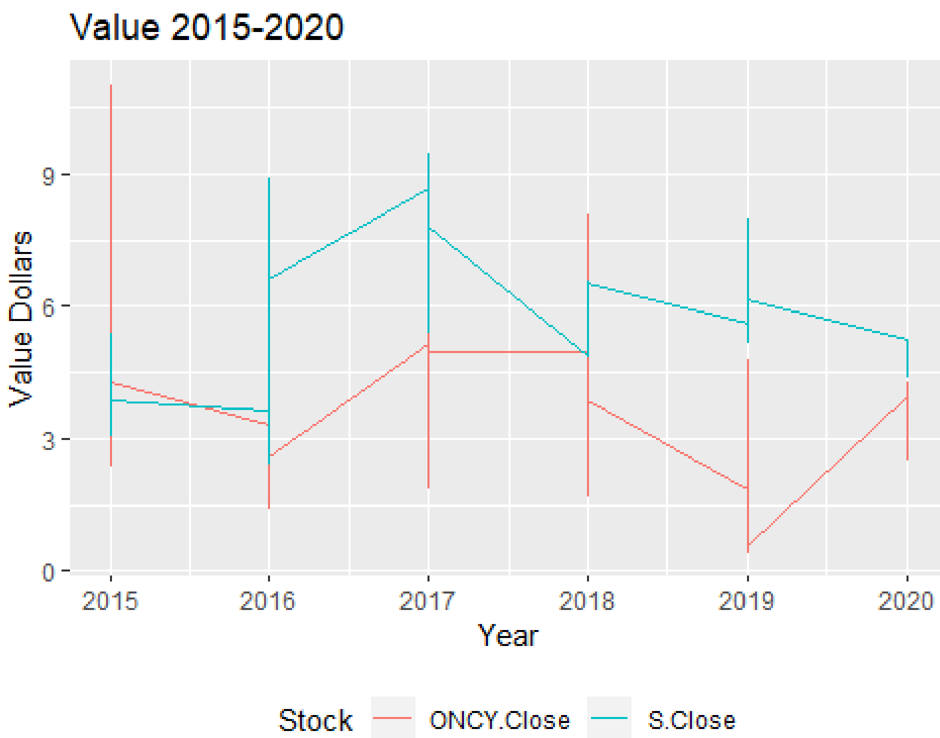


The line chart above shows that WM has increased significantly every year since 2015, with a slight decrease in 2019, but overall has increased from the 50 range in 2015 to the 113 range in 2020.

The ONCY and S stocks have had slight increases and decreases in the last five years but look like they have increased slightly overall from 2015-2020.

Lets look at S and ONCY stocks more closely.

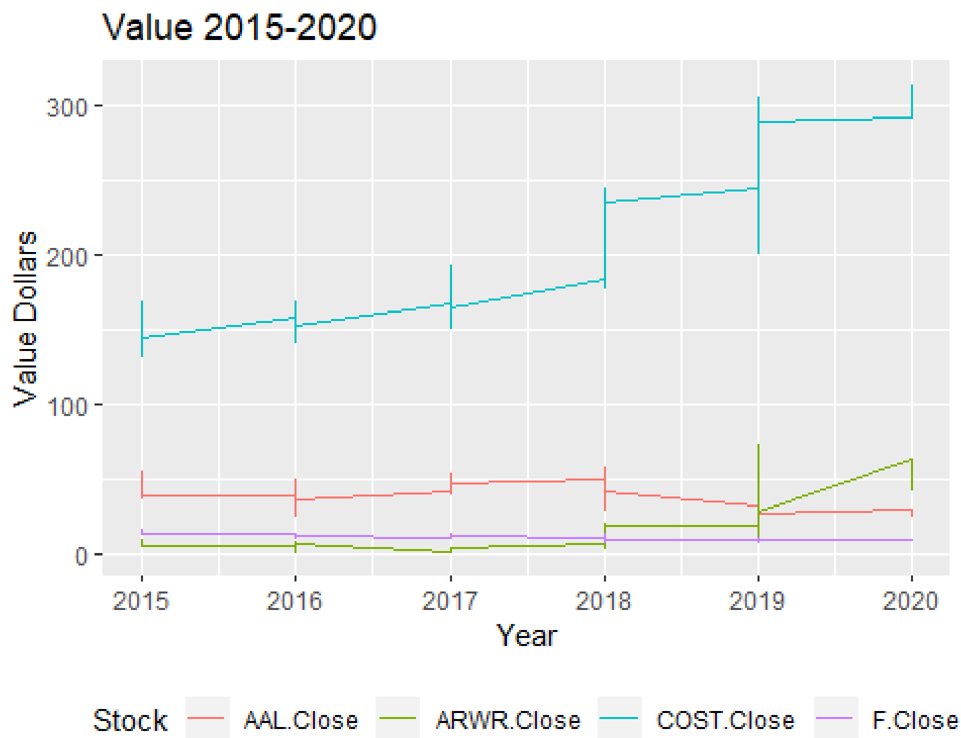
```
sub1 <- Value2[,c(19:20,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:2)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



It looks like these two stocks, ONCY and S, have had cyclical patterns in the last five years, and if that is true, then S stock hasn't reached its cyclical minimum and ONCY stock hasn't reached it cyclical maximum. And if this is not the case then there are some triggers in the value of this stock in 2016, where they both increased, then steadily decreased in 2017. A global minimum in the last five years is seen in 2019 for ONCY stock, while the global maximums for both stock is in 2017. The start of 2016 showed both stocks had a local minima while S stock had its global minima this year, but only for this last five year period.

Now for the next stocks in our subset.

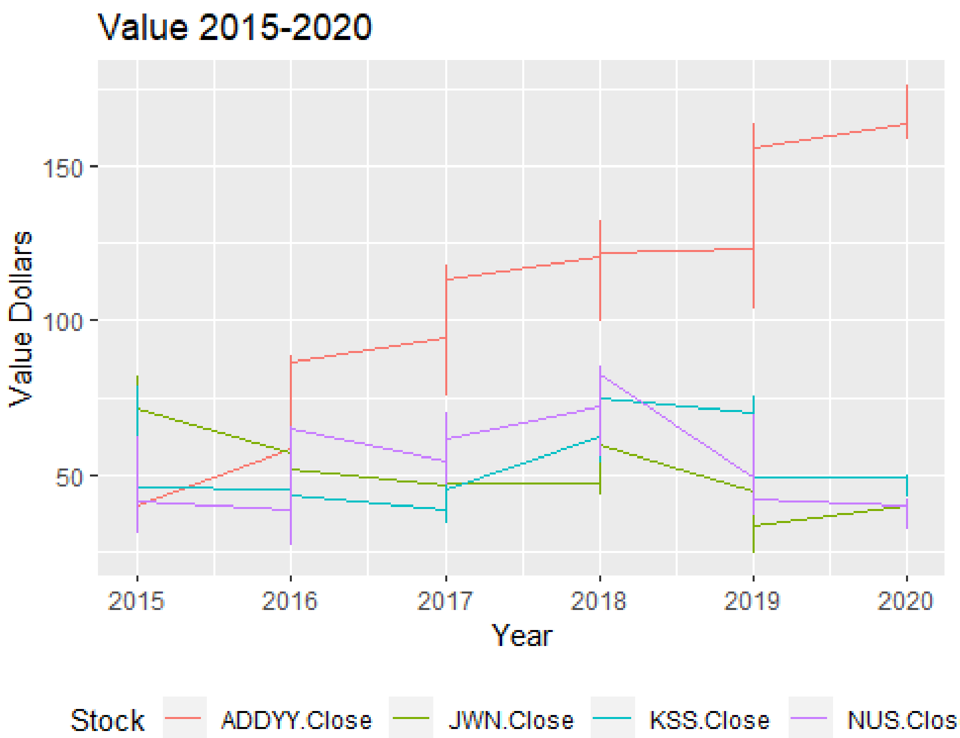
```
sub1 <- Value2[,c(21:24,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above subset shows that ARWR and COST stock have been increasing the last two years, but ARWR stock had some near flat changes in value for years 2015, 2016, and 2017. The purple line for Ford is relatively maintaining value, but no increases or decreases of note for Ford in the last five years. The AAL stock had a global maxima in 2018 but overall decreased in value slightly in the last five years. ***

Now for the next stocks in our subset.

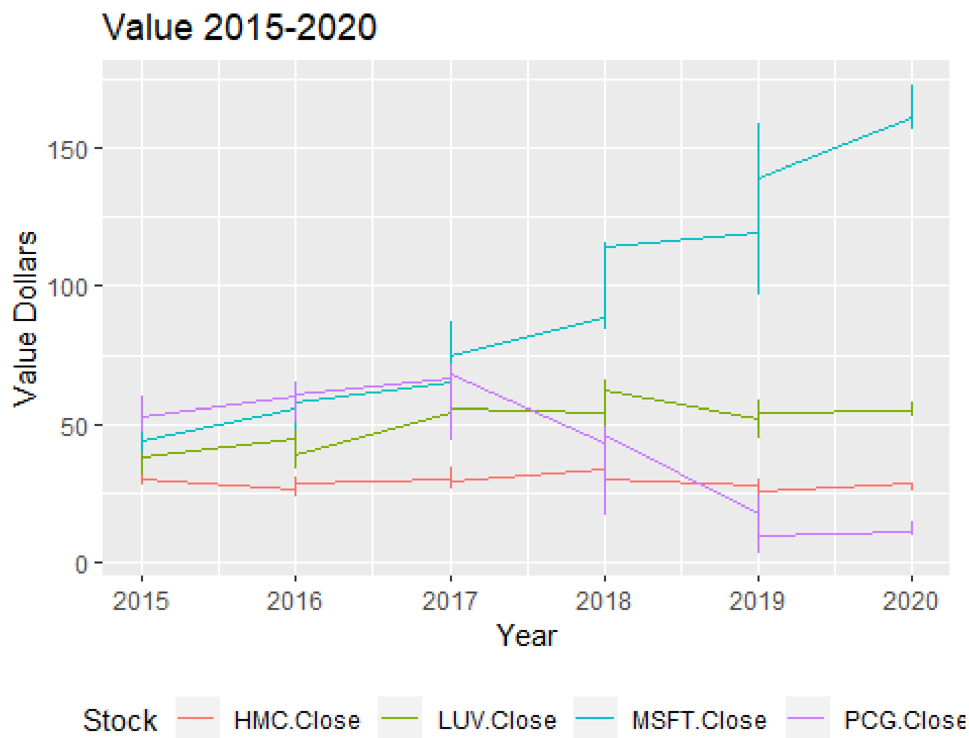
```
sub1 <- Value2[,c(25:28,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above line chart shows that ADDYY has been significantly increasing over the last five years it jumped from the 40 USD range to the 165 USD range in 2020. The other three stocks all moved together with slightly different rates of increase and decrease. But the JWN stock lost value over the last five years, while KSS and NUS stocks both increased only marginally after some cyclical rise and falls in value. ***

Now for the next stocks in our subset.

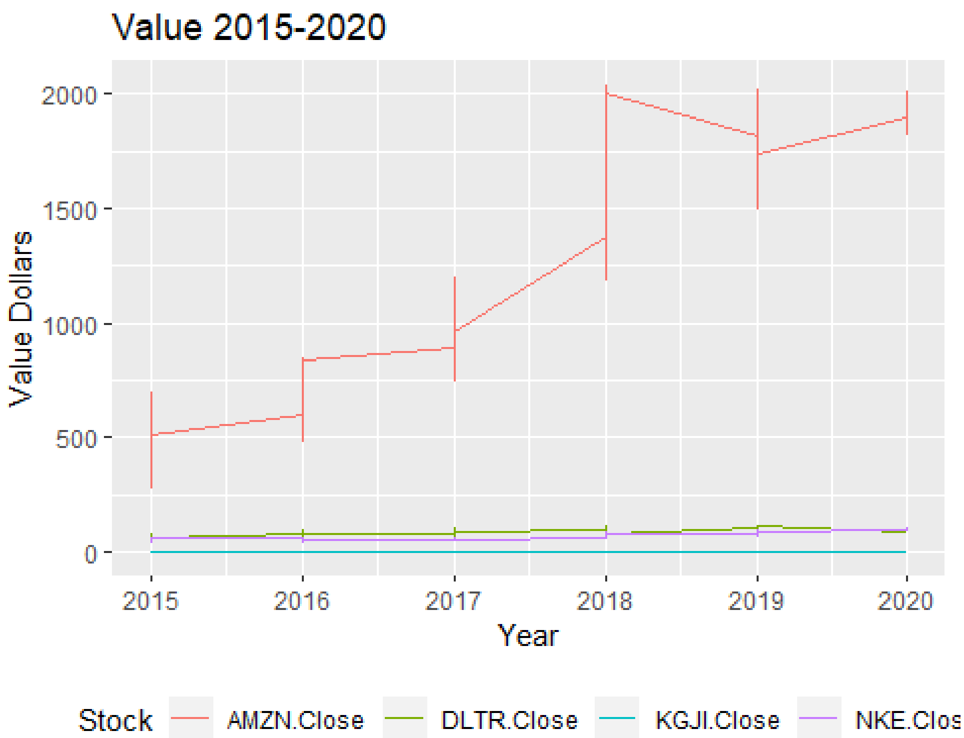
```
sub1 <- Value2[,c(29:32,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above line chart shows that MSFT increased steadily the last five years with none of the years having declining values in stock. PCG stock had a local maxima in 2017 but a local minima in 2019 which led to an overall loss in value from 2015-2020. The LUV stock is the olive colored stock that had an increase overall in value by about 10 USD. And the HMC stock slightly stayed the same and may have decreased marginally in the last five years. ***

Now for the next stocks in our subset.

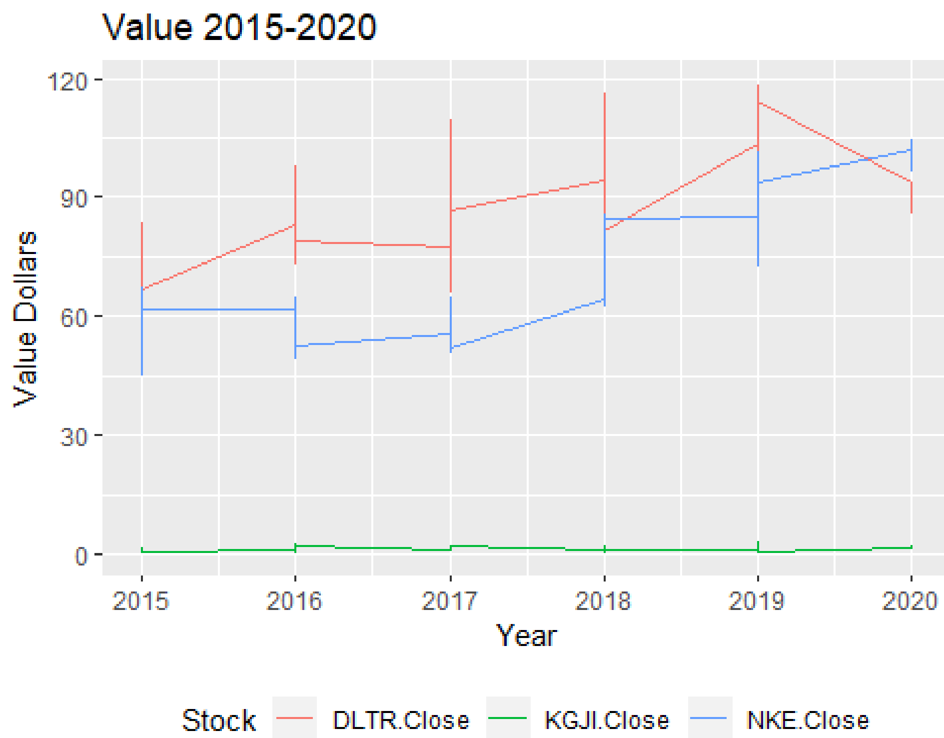
```
sub1 <- Value2[,c(33:36,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above line chart shows that AMZN stock is on its own scale and has saw an overall huge jump in value in the last five years, with every year increasing, except in 2018 where it decreased from its local maxima at the start of 2018. Its value in 2015 was in the 500 USD range and at the start of 2020 was in the 1700-1800 USD range.

Lets look at the scale more appropriate for the other three stocks of DLTR, KGJI, and NKE.

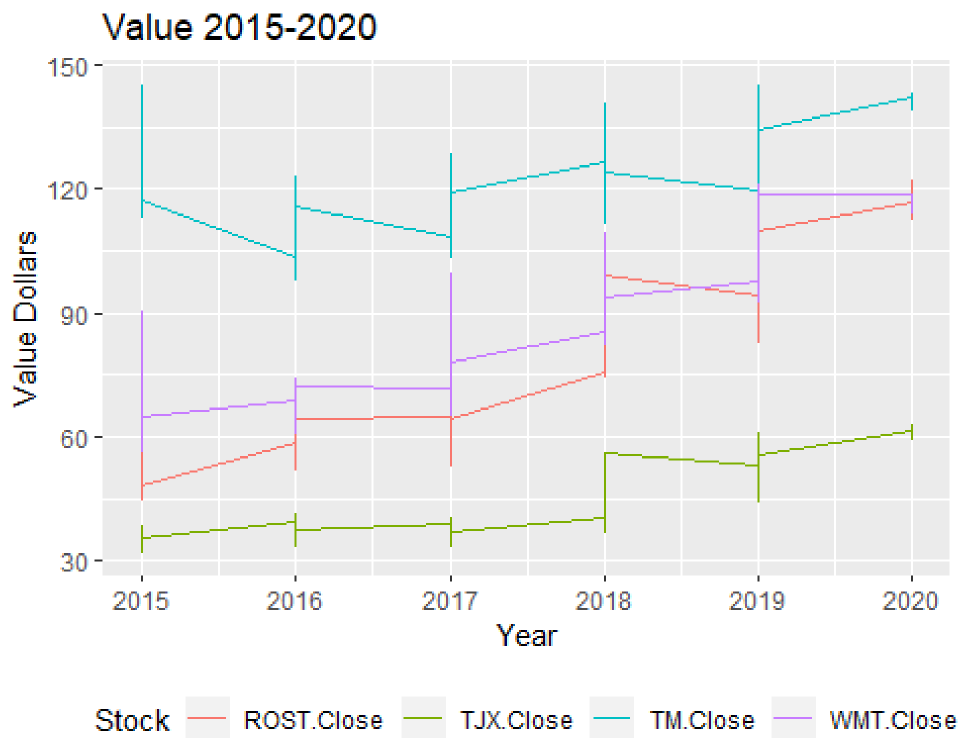
```
sub1 <- Value2[,c(33:35,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:3)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above line chart shows the smaller scale value changes by year for DLTR, KGJI, and NKE. Both NKE and DLTR stocks have increased in value over the last five years, while DLTR did see a decreasing value throughout the last year of 2019. The KGJI stock showed marginal changes in value over the last five years, with no significant local minimas or local maximas.It does look like a slight increase overall from 2015-2020 for KGJI stock. ***

Now for the next stocks in our subset.

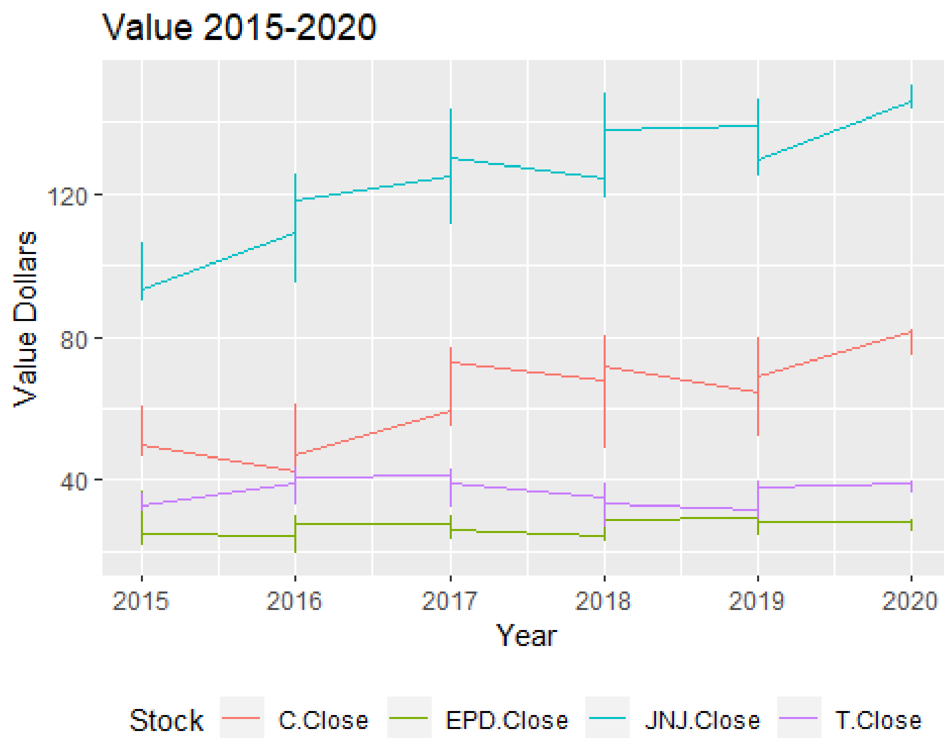
```
sub1 <- Value2[,c(37:40,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



In the above line chart we see that all of the stocks increased noticeably in the last five years. The TM stock had some years that decreased in 2015, 2016, and 2018, but always starts the new year at a higher value than the year before. In 2018 WMT increased, while the other three stocks of TJX, TM, and ROST saw slight decreases. ***

Now for the next stocks in our subset.

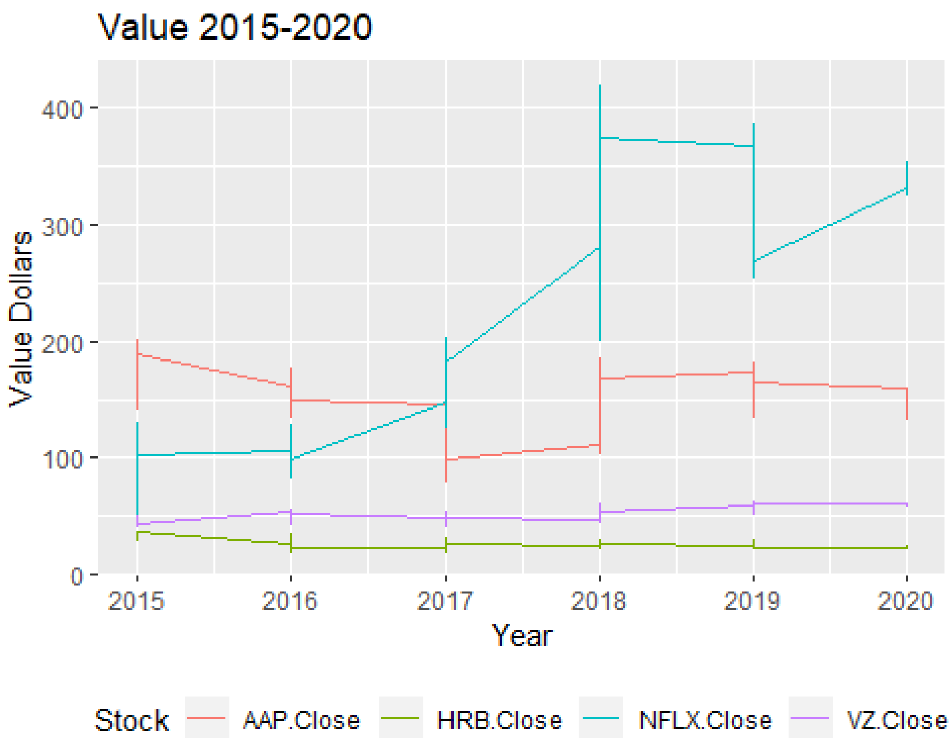
```
sub1 <- Value2[,c(41:44,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above line chart also shows an overall increase in value over the last five years with significant jumps in value for C and JNJ stocks. In 2017, there were some decreases in value throughout the year for all these stocks of C, EPD, JNJ, and T stocks, but in two years they all started 2019 at the same values of 2017 and saw increasing values throughout 2019. ***

Now for the next stocks in our subset.

```
sub1 <- Value2[,c(45:48,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```

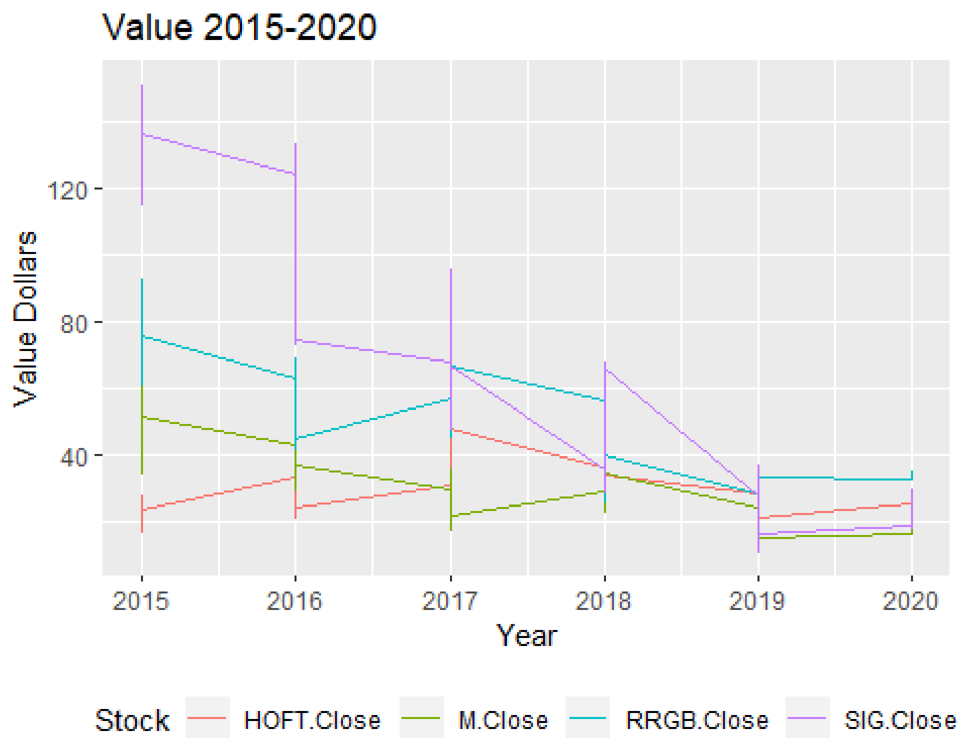


The above line chart shows that NFLX increased significantly while HRB and AAP saw losses over the last five years. VZ stock saw a slight increase in value over the last five years. In 2017 Netflix saw a huge inrease, while in 2018 it stayed somewhat stagnant with a sharp drop in value at the start of 2019 that saw an increasing year throughout 2019.

In 2017, there was a sharp drop in value for AAP, but by the start of 2018 the value increased to a value above the start of 2017.

Now for the last five stocks in our subset.

```
sub1 <- Value2[,c(49:53,115)]
sub1tidy <- gather(sub1, 'Stock','Value',1:4)

ggplot(data = sub1tidy, aes(x=Year, y=Value,group=Stock)) +
  geom_line(aes(color=Stock))+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('Value 2015-2020')+
  ylab('Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



Our last set of stock show that RRGB and SIG saw significant losses over the last five years, while M stock showed a smaller loss. HOFT stock saw an increase over the last five years, but only marginally or slightly. In 2017 M stock saw an increasing year for its value after having two years from 2015-2016 see decreasing values throughout those years. M stock and HOFT stock seemed to be negatively correlated for years 2015-2018, with both stocks having different rates of decrease in 2018 and an increase in value of similar rates of increase in 2019. All of these stocks decreased at different rates in 2018, and increased at different rates in 2019. ***

Lets group by the year and get the mean values over the last five years for each stock value.

```
Value3 <- Value2[,c(1:53,112,115)]

yearMeans <- Value3 %>% group_by(Year) %>%
  summarise_at(vars(as.vector(colnames(Value3)[1:53])), mean)

yearMeansTidy <- gather(yearMeans,'Stock','YearMeanValue',2:54)

stock5yrMeans <- yearMeansTidy %>% group_by(Stock) %>%
  summarise_at(vars(as.vector(colnames(yearMeansTidy)[3])), mean)
colnames(stock5yrMeans)[2] <- 'stock5yrMeans'

Stock5year <- merge(stock5yrMeans,yearMeansTidy, by.x='Stock', by.y='Stock')

stock5yrOrdered <- Stock5year[with(Stock5year, order(Stock, Year)),]
```

Lets add a field that shows if the stock had an increase of 10% during the year and a field that shows if it decreased

```
ymn <- stock5yrOrdered$YearMeanValue
YMN <- c(ymn[1],ymn[1:length(ymn)-1])

stc2 <- stock5yrOrdered$Stock
STC2 <- c('xyz',stc2[1:length(stc2)-1])

STC3 <- ifelse(stc2==STC2, 1,0)

stock5yrOrdered$Direction5yr10PercentChange <- ifelse(STC3==1 &
stock5yrOrdered$YearMeanValue-YMN >  .10*YMN,'up10',
                                              ifelse(STC3==1 &
stock5yrOrdered$YearMeanValue-YMN <= -0.10*YMN, 'down10',
                                              ifelse(STC3==1 &
stock5yrOrdered$YearMeanValue-YMN <= 0, 'down', ifelse(STC3==1 &
stock5yrOrdered$YearMeanValue-YMN > 0, 'up', ''))))

show1 <- cbind(head(stock5yrOrdered), tail(stock5yrOrdered))
show1

##       Stock stock5yrMeans Year YearMeanValue Direction5yr10PercentChange
## 2 AAL.Close      38.67371 2015      45.12210
## 3 AAL.Close      38.67371 2016      38.18385                      down10
## 4 AAL.Close      38.67371 2017      47.49072                        up10
## 1 AAL.Close      38.67371 2018      42.80195                        down
## 5 AAL.Close      38.67371 2019      30.87933                      down10
## 6 AAL.Close      38.67371 2020      27.56429                      down10
##       Stock stock5yrMeans Year YearMeanValue Direction5yr10PercentChange
## 2 XOM.Close        78.737 2015      82.82845
## 3 XOM.Close        78.737 2016      86.21968                          up
## 4 XOM.Close        78.737 2017      81.86159                        down
## 1 XOM.Close        78.737 2018      79.95570                        down
## 5 XOM.Close        78.737 2019      73.73464                        down
## 6 XOM.Close        78.737 2020      67.82191                        down

length(unique(stock5yrOrdered$Stock))

## [1] 53
```

Lets get these subsets of stocks that within the time span of 2015-2020 increased by more than 10% annually, decreased by 10% or more annually, decreased, or increased.

```
Stocks10PercentAnnualDecrease2015_2020 <- subset(stock5yrOrdered,
stock5yrOrdered$Direction5yr10PercentChange=='down10')

stocks10Decr <- Stocks10PercentAnnualDecrease2015_2020 %>% group_by(Stock)
%>% count(n=n())
colnames(stocks10Decr)[2] <- 'nTimesDecr10_5yr'
```

```
stocks10Decr <- stocks10Decr[,-3]

Stocks10PercentAnnualIncrease2015_2020 <- subset(stock5yrOrdered,
stock5yrOrdered$Direction5yr10PercentChange=='up10')

stocks10Incr <- Stocks10PercentAnnualIncrease2015_2020 %>% group_by(Stock)
%>% count(n=n())
colnames(stocks10Incr)[2] <- 'nTimesIncr10_5yr'
stocks10Incr <- stocks10Incr[,-3]

StocksAnnualIncrease2015_2020 <- subset(stock5yrOrdered,
stock5yrOrdered$Direction5yr10PercentChange=='up')

StocksIncrZerobase <- StocksAnnualIncrease2015_2020 %>% group_by(Stock) %>%
count(n=n())
colnames(StocksIncrZerobase)[2] <- 'nTimesIncrFromZero_5yrs'
StocksIncrZerobase <- StocksIncrZerobase[,-3]

StocksAnnualDecrease2015_2020 <- subset(stock5yrOrdered,
stock5yrOrdered$Direction5yr10PercentChange=='down')

StocksDecrZerobase <- StocksAnnualDecrease2015_2020 %>% group_by(Stock) %>%
count(n=n())
colnames(StocksDecrZerobase)[2] <- 'nTimesDecrFromZero_5yrs'
StocksDecrZerobase <- StocksDecrZerobase[,-3]
```

Lets merge these sets together with outer joins.

```
Stocks5yrChanges_outerJoin <- merge(stocks10Decr,stocks10Incr, by.x='Stock',
by.y='Stock', all=TRUE)

Stocks5yrChanges_outerJoin1 <-
merge(Stocks5yrChanges_outerJoin,StocksDecrZerobase, by.x='Stock',
by.y='Stock', all=TRUE)

Stocks5yrChanges_outerJoin2 <-
merge(Stocks5yrChanges_outerJoin1,StocksIncrZerobase, by.x='Stock',
by.y='Stock', all=TRUE)

stock_5yr_stats_2015_2020 <-
merge(stock5yrOrdered,Stocks5yrChanges_outerJoin2, by.x='Stock',
by.y='Stock', all=TRUE)

length(unique(stock_5yr_stats_2015_2020$Stock))

## [1] 53
```

Write this file out to analyze those stocks having decreased and increased the most in the last 5 years.

```
write.csv(stock_5yr_stats_2015_2020,'stocks_STATS_N_Changes.csv',
row.names=FALSE)
```

Lets attach the stock name to this data set above by reading in the file with the names on it when hand picking these stocks by searching manually in finance.yahoo.com.

```
stockNames <- read.csv('yahooStockBasket.csv', header=T, sep=',',
na.strings=c('',' '))
stock_5yr_stats_2015_2020$Stock <- gsub('[.]Close','',
stock_5yr_stats_2015_2020$Stock)
stockNames$stock <- gsub('-','.', stockNames$stock)

stock_5yr_stats_2015_2020$Stock <- as.factor(stock_5yr_stats_2015_2020$Stock)
StockNames_STATS_2015_2020 <- merge(stockNames,stock_5yr_stats_2015_2020,
                                    by.x='stock', by.y='Stock')

StockNames_STATS_2015_2020$nTimesDecr10_5yr <-
  ifelse(is.na(StockNames_STATS_2015_2020$nTimesDecr10_5yr==TRUE),
                          0,StockNames_STATS_2015_2020$nTimesDecr10_5yr)

StockNames_STATS_2015_2020$nTimesIncr10_5yr <-
  ifelse(is.na(StockNames_STATS_2015_2020$nTimesIncr10_5yr==TRUE),
                          0,StockNames_STATS_2015_2020$nTimesIncr10_5yr)

StockNames_STATS_2015_2020$nTimesDecrFromZero_5yrs <-
  ifelse(is.na(StockNames_STATS_2015_2020$nTimesDecrFromZero_5yrs==TRUE),

0,StockNames_STATS_2015_2020$nTimesDecrFromZero_5yrs)

StockNames_STATS_2015_2020$nTimesIncrFromZero_5yrs <-
  ifelse(is.na(StockNames_STATS_2015_2020$nTimesIncrFromZero_5yrs==TRUE),

0,StockNames_STATS_2015_2020$nTimesIncrFromZero_5yrs)

StockNames_STATS_2015_2020$Direction5yr10PercentChange <-

ifelse(StockNames_STATS_2015_2020$Direction5yr10PercentChange=='',0,StockName
s_STATS_2015_2020$Direction5yr10PercentChange)

write.csv(StockNames_STATS_2015_2020, 'StockNames_STATS_2015_2020.csv',
row.names=FALSE)

show2 <-
rbind(head(StockNames_STATS_2015_2020,3),tail(StockNames_STATS_2015_2020,3))
show2

##     stock
## 1     AAL
## 2     AAL
## 3     AAL
```

```
## 316   XOM
## 317   XOM
## 318   XOM
##
stockInfo
## 1   American Airlines Group Inc. (AAL)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 2   American Airlines Group Inc. (AAL)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 3   American Airlines Group Inc. (AAL)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 316                Exxon Mobil Corporation (XOM)\nNYSE - NYSE Delayed
Price. Currency in USD
## 317                Exxon Mobil Corporation (XOM)\nNYSE - NYSE Delayed
Price. Currency in USD
## 318                Exxon Mobil Corporation (XOM)\nNYSE - NYSE Delayed
Price. Currency in USD
##     stockExchange stock5yrMeans Year YearMeanValue
Direction5yr10PercentChange
## 1          Nasdaq      38.67371 2018      42.80195
down
## 2          Nasdaq      38.67371 2017      47.49072
up10
## 3          Nasdaq      38.67371 2020      27.56429
down10
## 316          NYSE      78.73700 2017      81.86159
down
## 317          NYSE      78.73700 2019      73.73464
down
## 318          NYSE      78.73700 2015      82.82845
0
##     nTimesDecr10_5yr nTimesIncr10_5yr nTimesDecrFromZero_5yrs
## 1                  3                1                       1
## 2                  3                1                       1
## 3                  3                1                       1
## 316                0                0                       4
## 317                0                0                       4
## 318                0                0                       4
##     nTimesIncrFromZero_5yrs
## 1                         0
## 2                         0
## 3                         0
## 316                       1
## 317                       1
## 318                       1
```

```
length(unique(StockNames_STATS_2015_2020$stock))
```

```
## [1] 53
```

Lets the mean annual unemployment rates using the original table to combine with this table of the n times a stock increases/decreases per year in the last five years.

```
ue$Annual <- round(rowMeans(ue[,2:13], na.rm=T),2)
ue_15_20 <- ue[9:14,c(1,14)]
colnames(ue_15_20)[2] <- 'Annual_UE'
```

Now, combine the unemployment and the newest stats with counts table.

```
stock_5yrs_ue <- merge(ue_15_20,StockNames_STATS_2015_2020, by.x='Year',
by.y='Year')
```

Add in a boolean field to show if the YearMeanValue is greater than the Stock5yrMeans column as a 1 if true and a 0 if not.

```
stock_5yrs_ue$YearMeanGreaterThan5yrMean <-
ifelse(stock_5yrs_ue$YearMeanValue >

stock_5yrs_ue$stock5yrMeans,1,0)

write.csv(stock_5yrs_ue,'stock_2015-2020_ue.csv',row.names=FALSE)
```

Make separate portfolios for each of the stocks that increased by more than 10% annually more than at least 1 time, decreased more than 10% annually more than at least 1 time, then get the mean value of the YearMeanValue column. Compare this to the portfolio of the stocks that never decreased more than 10% annually.

```
sub_D10 <- subset(StockNames_STATS_2015_2020,
StockNames_STATS_2015_2020$nTimesDecr10_5yr > 0)


D10_2015 <- subset(sub_D10, sub_D10$Year==2015)
D10_2020 <- subset(sub_D10, sub_D10$Year==2020)

md10_2015 <- mean(D10_2015$YearMeanValue)
md10_2020 <- mean(D10_2020$YearMeanValue)
md10_2015

## [1] 67.18499

md10_2020

## [1] 65.57645

ROI_D10 <- md10_2020/md10_2015
ROI_D10

## [1] 0.9760581
```

```
d10_startValue <- md10_2015*length(unique(D10_2015$stock))
d10_endValue <- md10_2020*length(unique(D10_2020$stock))
d10_startValue

## [1] 2149.92

d10_endValue

## [1] 2098.446
```

The above values show the 2015 average stock value of those stocks that decreased more than 10 percent in the last five years more than once was 67 USD. And in 2020 those stocks decreased in value to 66 USD giving it a five year ROI in the last five years of 0.976, or a decline of 2.4 percent in value. The 2015 value of this portfolio of stocks was 2150 USD, and in 2020 the portfolio value of the stocks was 2098 USD showing the dollar decrease over five years.

```
sub_nvr_D10 <- subset(StockNames_STATS_2015_2020,
StockNames_STATS_2015_2020$nTimesDecr10_5yr == 0)

nD10_2015 <- subset(sub_nvr_D10, sub_nvr_D10$Year==2015)
nD10_2020 <- subset(sub_nvr_D10, sub_nvr_D10$Year==2020)

mnD10_2015 <- mean(nD10_2015$YearMeanValue)
mnD10_2020 <- mean(nD10_2020$YearMeanValue)
mnD10_2015

## [1] 109.2018

mnD10_2020

## [1] 272.5558

ROI_nD10 <- mnD10_2020/mnD10_2015
ROI_nD10

## [1] 2.495891

nD10_startValue <- mnD10_2015*length(unique(nD10_2015$stock))
nD10_endValue <- mnD10_2020*length(unique(nD10_2020$stock))
nD10_startValue

## [1] 2293.238

nD10_endValue

## [1] 5723.671
```

The above numbers show the mean stock value of those stocks that never decreased by more than 10 percent in 2015-2020. The 2015 average stock price of these stocks was 109 USD, and in 2020 the average price was 273 USD. This was a ROI of 2.49 or 249 percent, which means it more than doubled in value over the last five years. The 2015 portfolio price of these specific stock were 2293 USD and in 2020 the portfolio price was 5724 USD. This shows that having a stock that never decreases by more than 10 percent in five years could be a good stock to buy.

Lets now do the reverse and look at those stocks that increased more than 10% at least three times in the last five years of 2015-2020 and compare the means.

```
sub_I10 <- subset(StockNames_STATS_2015_2020,
                  StockNames_STATS_2015_2020$nTimesIncr10_5yr > 3)

sub_nvr_I10 <- subset(StockNames_STATS_2015_2020,
                      StockNames_STATS_2015_2020$nTimesIncr10_5yr == 0)

m2015 <- subset(sub_I10, sub_I10$Year==2015)
m2020 <- subset(sub_I10, sub_I10$Year==2020)

pm_2015 <- mean(m2015$YearMeanValue)
pm_2020 <- mean(m2020$YearMeanValue)

ROI_incr10_3x <- pm_2020/pm_2015
ROI_incr10_3x

## [1] 2.553548

I10_3_startValue <- pm_2015*length(unique(m2015$stock))
I10_3_endValue <- pm_2020*length(unique(m2020$stock))
I10_3_startValue

## [1] 823.0797

I10_3_endValue

## [1] 2101.773

mn_2015 <- subset(sub_nvr_I10, sub_nvr_I10$Year==2015)
mn_2020 <- subset(sub_nvr_I10, sub_nvr_I10$Year==2020)

pmn_2015 <- mean(mn_2015$YearMeanValue)
pmn_2020 <- mean(mn_2020$YearMeanValue)

ROI_nvr10 <- pmn_2020/pmn_2015
ROI_nvr10

## [1] 0.5342204

nI10_startValue <- pmn_2015*length(unique(mn_2015$stock))
nI10_endValue <- pmn_2020*length(unique(mn_2020$stock))
nI10_startValue

## [1] 502.048
```

```
nI10_endValue

## [1] 268.2043
```

From the above, we can see that those stocks that never increased by more than 10 percent during the last five years lost almost half their 2015 start value of 502 USD in 2015 and 268 USD in 2020 and having a ROI ratio of 0.53. On the other hand, the stocks that increased by more than 10 percent at least three times during the last five years had a ROI ratio of 2.55, a 2015 portfolio value of 823 USD and a 2020 portfolio value of 2102 USD.

Now lets look at those stocks that increased at least one time in the last five years but never by more than 10 percent.

```
sub_Iz <- subset(StockNames_STATS_2015_2020,
StockNames_STATS_2015_2020$nTimesIncrFromZero_5yr > 0)

Iz_2015 <- subset(sub_Iz, sub_Iz$Year==2015)
Iz_2020 <- subset(sub_Iz, sub_Iz$Year==2020)

Iz_2015 <- subset(sub_Iz, sub_Iz$Year==2015)
Iz_2020 <- subset(sub_Iz, sub_Iz$Year==2020)

m_Iz_2015 <- mean(Iz_2015$YearMeanValue)
m_Iz_2020 <- mean(Iz_2020$YearMeanValue)
m_Iz_2015

## [1] 100.1005

m_Iz_2020

## [1] 194.7926

ROI_Iz <- m_Iz_2020/m_Iz_2015
ROI_Iz

## [1] 1.945971

p_Iz_2015 <- m_Iz_2015*length(unique(sub_Iz$stock))
p_Iz_2020 <- m_Iz_2020*length(unique(sub_Iz$stock))
p_Iz_2015

## [1] 3503.518

p_Iz_2020

## [1] 6817.742
```

From the data above, the 2015 average stock price of 100 USD for the stock that had an increasing year at least one time in the last five years but not by more than 10 percent of the last year value. The 2020 average stock price increased to 195 USD, with an ROI of 1.95

or 195 percent. The 2015 portfolio value was 3504 USD and in 2020 the portfolio value increased to 6818 USD. This makes sense that those stocks that increase are good to have as they are making you money, but even if they don't increase by more than 10 percent in any year, when combined with other increasing stock they can nearly double your investment over five years.

Here are the stocks that never increased in the last five years.

```
sub_nvr_Iz <- subset(StockNames_STATS_2015_2020,
StockNames_STATS_2015_2020$nTimesIncrFromZero_5yr == 0)

nIz_2015 <- subset(sub_nvr_Iz, sub_nvr_Iz$Year==2015)
nIz_2020 <- subset(sub_nvr_Iz, sub_nvr_Iz$Year==2020)

m_nIz_2015 <- mean(nIz_2015$YearMeanValue)
m_nIz_2020 <- mean(nIz_2020$YearMeanValue)
m_nIz_2015

## [1] 52.2022

m_nIz_2020

## [1] 55.79861

ROI_nIz <- m_nIz_2020/m_nIz_2015
ROI_nIz

## [1] 1.068894

nIz_startValue <- m_nIz_2015*length(unique(nIz_2015$stock))
nIz_endValue <- m_nIz_2020*length(unique(nIz_2020$stock))
nIz_startValue

## [1] 939.6396

nIz_endValue

## [1] 1004.375
```

From the above we have a portfolio of stock that never increased from zero, but might have increased by more than 10 percent.This variable was designed to capture exactly those stocks that did not increase by more than 10 percent but did increase some. This portfolio has a 2015 average stock value of 52 USD and this value increases to 56 USD in 2020 with an ROI of 1.07 or a five year interest of 7 percent. The 2015 portfolio value was 940 USD and in 2020 the portfolio value was 1004 USD. ***

Lets get the entire 53 stock portfolio mean value in 2015 and compare to the same 53 stock portfolio mean value in 2020.

```
p2015 <- subset(StockNames_STATS_2015_2020,
StockNames_STATS_2015_2020$Year==2015)
p2020 <- subset(StockNames_STATS_2015_2020,
StockNames_STATS_2015_2020$Year==2020)

pm2015 <- mean(p2015$YearMeanValue)
pm2020 <- mean(p2020$YearMeanValue)

pm2015

## [1] 83.83316

pm2020

## [1] 147.5871

ROI_all <- pm2020/pm2015
ROI_all

## [1] 1.760486

pm2015*length(unique(StockNames_STATS_2015_2020$stock))

## [1] 4443.157

pm2020*length(unique(StockNames_STATS_2015_2020$stock))

## [1] 7822.117
```

The portfolio mean was 84 USD in 2015 and 147 USD in 2020. The 2015 portfolio was valued at 4443 USD and in 2020 at 7822 USD for all stocks. The ROI is 1.76, which is good because you almost doubled the loan with all 53 of these stocks in five years spanning 2015-2020.

```
76/5

## [1] 15.2
```

So, with a return of 76% on top of the value invested in 2015 figuratively for this example, that is 15.2% annual interest earned each of five years. This is called pooling, that the wins over compensate for the losses and it is used in health insurance companies as well as financial portfolios like 401k investment tools.

What would the ROI be for all stocks that increased during the last five years by more than 10 per cent?

```
incr_10_2015 <- subset(sub_I10, sub_I10$Year==2015)
mean_incr_10_2015 <- mean(incr_10_2015$YearMeanValue)

incr_10_2020 <- subset(sub_I10, sub_I10$Year==2020)
```

```
mean_incr_10_2020 <- mean(incr_10_2020$YearMeanValue)

mean_incr_10_2015

## [1] 117.5828

mean_incr_10_2020

## [1] 300.2533

ROI_Incr_10 <- mean_incr_10_2020/mean_incr_10_2015
ROI_Incr_10

## [1] 2.553548

value2015 <- mean_incr_10_2015*length(unique(sub_I10$stock))
value2020 <- mean_incr_10_2020*length(unique(sub_I10$stock))

value2015

## [1] 823.0797

value2020

## [1] 2101.773
```

The **return on investment** is more than doubled to 2102 USD over five years from a value of 823 USD in 2015 by selecting only the stocks in this portfolio of stocks that increased by more than 10% at least once in the last 5 years. The return of the ratio of the mean value in 2020 to 2015 is 2.55, which means the portfolio more than doubled.

But how do we or how can we know what stocks to select now that will increase many times as long as we have the investment? Can machine learning be built from this data set to find the stocks in this sample that produce good indicating features of other stocks that could be profitable to buy? We will develop this as we progress through this portfolio. We would also want indicators that would tell if certain stocks look like a good prospect but are actually going to be on a steady decline that translates to financial loss as long as you own them.

There are four sets of counts for those that increased more than 10%, decreased more than 10%, increased more than zero but less than 10%, and decreased more than zero but less than 10% within the five year span from 2015-2020. Lets see if there is a better subset of choices for a better market portfolio.

Lets add a five year poisson column using lambda=(unemployment rate), time=(nTimesIncr10_5yr), and k=(YearMeanGreaterThan5yrMean).We will use the best subset so far of the stocks that increased by more than 10% annually in at least 3 out of the last five years.

```r
ue2 <- stock_5yrs_ue$Annual_UE
t <- stock_5yrs_ue$nTimesIncr10_5yr
k <- stock_5yrs_ue$YearMeanGreaterThan5yrMean
stock_5yrs_ue$poisson5yrUE <- round((exp(-ue2*t)*(ue2*t)^k)/(factorial(k)),5)
```

Lets get a subset of those stocks that have cyclical patterns within five years, so that we have three years the stock increases more than 10% exactly 3 times, and two years where the stock decreases less than 10% exactly 2 times. Separately, get the stocks it increases greater than 10% exactly 3 times, and decreases more than 10% exactly 2 times. Also get the reverse of these values

```r
cyclical <- subset(stock_5yrs_ue, stock_5yrs_ue$nTimesIncr10_5yr==3 &
(stock_5yrs_ue$nTimesDecr10_5yr==2 |
stock_5yrs_ue$nTimesDecrFromZero_5yrs==2))

cyclical2 <- subset(stock_5yrs_ue, stock_5yrs_ue$nTimesIncrFromZero_5yrs >=2
& (stock_5yrs_ue$nTimesDecr10_5yr >= 2 |
stock_5yrs_ue$nTimesDecrFromZero_5yrs >= 2))

c1 <- as.character(unique(cyclical$stock))
c2 <- as.character(unique(cyclical2$stock))
cycle <- c(c1,c2)
cycle1 <- as.data.frame(cycle)
colnames(cycle1) <- 'Stock'

portCycle <- merge(cycle1,stock_5yrs_ue, by.x='Stock', by.y='stock')
portCycle_2015 <- subset(portCycle, Year==2015)
portCycle_2020 <- subset(portCycle, Year==2020)

pc_mean2015 <- mean(portCycle_2015$YearMeanValue)
pc_mean2020 <- mean(portCycle_2020$YearMeanValue)

pc_mean2015

## [1] 36.228

pc_mean2020

## [1] 37.91673

ROI_pc <- pc_mean2020/pc_mean2015
ROI_pc

## [1] 1.046614

startValue <- pc_mean2015*length(unique(portCycle_2015$Stock))
endValue <- pc_mean2020*length(unique(portCycle_2020$Stock))

startValue

## [1] 434.736
```

```
endValue

## [1] 455.0007
```

The above shows that the **cyclical stocks that have highs and lows the time span of the loan aren't great investments**, as these stocks started at 435 USD in 2015 but ended with a portfolio value of 455 USD over a five year time span from 2015-2020. The ratio of average stock in 2020 to average stock in 2015 is 1.04, which means it earned 4 perent interest over 5 years or less than 1 percent interest a year. This is equivalent to most bank savings accounts. It is good they at least stayed the same and didn't cause the portfolio to lose money, and we can assume those stocks that do decrease continuously will be the stocks that lose money. We should look at the columns we added earlier that calculated the ROI dollars for each stock and see the average number of times the stock closes at a decreasing value over the span of the original data. Then use those outcomes to rank the stock a poor, average, good, or great stock to buy.

Lets use the StocksSTATS table with the 230 columns of ROI for each stock from the start in 2007 throughout 2020 and the daily changes for each stock for the same time span. We could add cumulative sum columns to each stock or just plot the daily changes for each of the 53 stocks and see if we notice any patterns and compare the the final recording ROI from the initial investment. Maybe see if some of these stocks are good to jump on, like a wave to increase value of the portfolio, or drop the stock at some point to keep the portfolio from dropping in value.

```
dailyChange <- grep('dailyChange',colnames(StocksSTATS))
DailyChanges <- StocksSTATS[,c(dailyChange,218:222,229)]
summary(DailyChanges)

##  TGT_dailyChange     FTR_dailyChange     UBSI_dailyChange
##  Min.   :-66.41001   Min.   :-81.00000   Min.   :-16.2500
##  1st Qu.: -0.51000   1st Qu.: -0.90000   1st Qu.: -0.3500
##  Median :  0.03000   Median :  0.00000   Median :  0.0000
##  Mean   :  0.03247   Mean   :  0.00026   Mean   :  0.0115
##  3rd Qu.:  0.57000   3rd Qu.:  0.75000   3rd Qu.:  0.3300
##  Max.   : 59.93000   Max.   :235.68000   Max.   : 34.8000
##  HD_dailyChange       JPM_dailyChange     XOM_dailyChange
##  Min.   :-199.42000   Min.   :-88.18999   Min.   :-23.84000
##  1st Qu.:   -0.43999   1st Qu.: -0.51000   1st Qu.: -0.58000
##  Median :    0.06000   Median :  0.02000   Median :  0.00000
##  Mean   :    0.07046   Mean   :  0.03574   Mean   :  0.02144
##  3rd Qu.:    0.67000   3rd Qu.:  0.58000   3rd Qu.:  0.61000
##  Max.   :   57.01999   Max.   : 48.24000   Max.   : 76.16000
##  CVX_dailyChange     NSANY_dailyChange    MGM_dailyChange
##  Min.   :-56.84000   Min.   :-15.639999   Min.   :-73.43000
##  1st Qu.: -0.78000   1st Qu.: -0.170000   1st Qu.: -0.30000
##  Median :  0.08000   Median :  0.000000   Median :  0.01000
##  Mean   :  0.03602   Mean   :  0.003793   Mean   :  0.00842
##  3rd Qu.:  0.88000   3rd Qu.:  0.170000   3rd Qu.:  0.32000
```

```
##  Max.   : 74.83000   Max.   : 21.540001   Max.   : 71.58000
##  TEVA_dailyChange    HST_dailyChange      WFC_dailyChange
##  Min.   :-30.11000   Min.   :-13.010280   Min.   :-18.84000
##  1st Qu.: -0.38000   1st Qu.: -0.190000   1st Qu.: -0.37000
##  Median : -0.02000   Median :  0.010000   Median :  0.00000
##  Mean   :  0.00209   Mean   :  0.005251   Mean   :  0.01532
##  3rd Qu.:  0.34000   3rd Qu.:  0.200000   3rd Qu.:  0.38000
##  Max.   : 38.18000   Max.   : 26.206558   Max.   : 34.01000
##  WWE_dailyChange     INO_dailyChange      SCE.PB_dailyChange
##  Min.   :-69.75000   Min.   :-11.520000   Min.   :-890.0000
##  1st Qu.: -0.18000   1st Qu.: -0.120000   1st Qu.: -19.0000
##  Median :  0.01000   Median :  0.000000   Median :   0.0000
##  Mean   :  0.02161   Mean   :  0.000623   Mean   :   0.2712
##  3rd Qu.:  0.21000   3rd Qu.:  0.080000   3rd Qu.:  22.0000
##  Max.   : 67.69000   Max.   : 13.320000   Max.   : 886.0000
##  FFIN_dailyChange     GOOG_dailyChange     WM_dailyChange
##  Min.   :-28.191665   Min.   :-1170.0303   Min.   :-84.98000
##  1st Qu.: -0.120000   1st Qu.:   -3.1314   1st Qu.: -0.23000
##  Median :  0.006667   Median :    0.2989   Median :  0.05000
##  Mean   :  0.010121   Mean   :    0.3702   Mean   :  0.03492
##  3rd Qu.:  0.155000   3rd Qu.:    4.4200   3rd Qu.:  0.34000
##  Max.   :  9.840002   Max.   :  344.4899   Max.   : 34.65000
##  ONCY_dailyChange    S_dailyChange        F_dailyChange
##  Min.   :-25.08000   Min.   :-14.830000   Min.   :-6.200001
##  1st Qu.: -0.28500   1st Qu.: -0.100000   1st Qu.:-0.120000
##  Median :  0.00000   Median :  0.000000   Median : 0.000000
##  Mean   :  0.00017   Mean   :  0.001874   Mean   : 0.002782
##  3rd Qu.:  0.19000   3rd Qu.:  0.090000   3rd Qu.: 0.120000
##  Max.   : 41.23000   Max.   : 19.299999   Max.   : 9.780000
##  ARWR_dailyChange    COST_dailyChange     AAL_dailyChange
##  Min.   :-49.50000   Min.   :-245.26999   Min.   :-42.66000
##  1st Qu.: -0.20000   1st Qu.:  -0.60000   1st Qu.: -0.36000
##  Median :  0.00000   Median :   0.10000   Median :  0.00000
##  Mean   :  0.00856   Mean   :   0.08749   Mean   :  0.00819
##  3rd Qu.:  0.20000   3rd Qu.:   0.90000   3rd Qu.:  0.37000
##  Max.   : 50.91000   Max.   :  93.68001   Max.   : 45.05000
##  JWN_dailyChange     NUS_dailyChange     HMC_dailyChange
##  Min.   :-38.08000   Min.   :-96.59000   Min.   :-13.20000
##  1st Qu.: -0.54000   1st Qu.: -0.41000   1st Qu.: -0.31000
##  Median :  0.03000   Median :  0.02000   Median :  0.00000
##  Mean   :  0.01022   Mean   :  0.01292   Mean   :  0.00792
##  3rd Qu.:  0.58000   3rd Qu.:  0.52000   3rd Qu.:  0.30000
##  Max.   : 53.74000   Max.   : 95.71001   Max.   : 34.69000
##  AMZN_dailyChange     T_dailyChange        HRB_dailyChange
##  Min.   :-1939.1100   Min.   :-15.26000   Min.   :-14.630001
##  1st Qu.:   -2.4300   1st Qu.: -0.22000   1st Qu.: -0.220001
##  Median :    0.3200   Median :  0.02000   Median :  0.010000
##  Mean   :    0.5272   Mean   :  0.01149   Mean   :  0.007173
##  3rd Qu.:    4.1200   3rd Qu.:  0.24000   3rd Qu.:  0.230001
##  Max.   : 1078.1600   Max.   : 39.46000   Max.   : 21.260000
```

```
##  RRGB_dailyChange   ADDYY_dailyChange    PCG_dailyChange
##  Min.   :-43.2600   Min.   :-138.60000   Min.   :-36.91000
##  1st Qu.: -0.5200   1st Qu.:  -0.45000   1st Qu.: -0.34000
##  Median :  0.0000   Median :   0.04000   Median :  0.03000
##  Mean   :  0.0101   Mean   :   0.04728   Mean   :  0.00304
##  3rd Qu.:  0.5600   3rd Qu.:   0.57001   3rd Qu.:  0.35000
##  Max.   : 42.7500   Max.   :  48.06000   Max.   : 49.44000
##  ROST_dailyChange     JNJ_dailyChange     NFLX_dailyChange
##  Min.   :-109.62500   Min.   :-87.02000   Min.   :-368.0886
##  1st Qu.:  -0.19000   1st Qu.: -0.38000   1st Qu.:  -0.4286
##  Median :   0.02750   Median :  0.03000   Median :   0.0143
##  Mean   :   0.03336   Mean   :  0.03929   Mean   :   0.0813
##  3rd Qu.:   0.32500   3rd Qu.:  0.50000   3rd Qu.:   0.6243
##  Max.   :  35.09000   Max.   : 60.10000   Max.   : 216.5200
##  M_dailyChange       KSS_dailyChange     DLTR_dailyChange
##  Min.   :-35.46000   Min.   :-31.74000   Min.   :-98.37333
##  1st Qu.: -0.40000   1st Qu.: -0.59000   1st Qu.: -0.32667
##  Median :  0.02000   Median :  0.01000   Median :  0.02333
##  Mean   :  0.00472   Mean   :  0.01508   Mean   :  0.03467
##  3rd Qu.:  0.44000   3rd Qu.:  0.61000   3rd Qu.:  0.46000
##  Max.   : 45.06000   Max.   : 77.49000   Max.   : 37.19000
##  WMT_dailyChange     C_dailyChange      AAP_dailyChange
##  Min.   :-74.62000   Min.   :-298.300   Min.   :-131.92001
##  1st Qu.: -0.39000   1st Qu.:  -0.680   1st Qu.:  -0.74000
##  Median :  0.04000   Median :  -0.010   Median :   0.04000
##  Mean   :  0.03604   Mean   :   0.021   Mean   :   0.05023
##  3rd Qu.:  0.46000   3rd Qu.:   0.650   3rd Qu.:   0.85001
##  Max.   : 47.40000   Max.   : 510.500   Max.   :  87.86001
##  JBLU_dailyChange     MSFT_dailyChange     KGJI_dailyChange
##  Min.   :-10.390000   Min.   :-140.49000   Min.   :-7.820000
##  1st Qu.: -0.139999   1st Qu.:  -0.29000   1st Qu.:-0.020000
##  Median :  0.000000   Median :   0.03000   Median : 0.000000
##  Mean   :  0.005087   Mean   :   0.04222   Mean   : 0.000182
##  3rd Qu.:  0.140000   3rd Qu.:   0.39000   3rd Qu.: 0.020000
##  Max.   : 14.690000   Max.   :  56.19000   Max.   : 9.090000
##  EPD_dailyChange      TJX_dailyChange     HOFT_dailyChange
##  Min.   :-14.510001   Min.   :-53.93750   Min.   :-23.229999
##  1st Qu.: -0.170000   1st Qu.: -0.14250   1st Qu.: -0.240000
##  Median :  0.015000   Median :  0.02000   Median :  0.000000
##  Mean   :  0.008679   Mean   :  0.01693   Mean   :  0.006511
##  3rd Qu.:  0.195000   3rd Qu.:  0.21250   3rd Qu.:  0.250000
##  Max.   : 15.875000   Max.   : 21.21500   Max.   : 22.400002
##  LUV_dailyChange    NKE_dailyChange     TM_dailyChange      VZ_dailyChange
##  Min.   :-42.8500   Min.   :-88.72500   Min.   :-47.61000   Min.   :-
25.92523
##  1st Qu.: -0.2100   1st Qu.: -0.23250   1st Qu.: -0.84000   1st Qu.: -
0.29000
##  Median :  0.0100   Median :  0.02750   Median : -0.02000   Median :
0.02814
##  Mean   :  0.0164   Mean   :  0.02852   Mean   :  0.04084   Mean   :
```

```
0.01833
##  3rd Qu.:  0.2500   3rd Qu.:  0.31750   3rd Qu.:  0.85999   3rd Qu.:
0.31888
##  Max.   : 25.0600   Max.   : 29.47000   Max.   :126.92000   Max.   :
35.31607
##  SIG_dailyChange          Date              DayOfWeek            Month
##  Min.   :-75.22000   Min.   :2007-01-03   Length:3293        Length:3293
##  1st Qu.: -0.61000   1st Qu.:2010-04-12   Class :character   Class
:character
##  Median :  0.03000   Median :2013-07-18   Mode  :character   Mode
:character
##  Mean   :  0.00509   Mean   :2013-07-16
##  3rd Qu.:  0.61000   3rd Qu.:2016-10-21
##  Max.   : 50.72000   Max.   :2020-01-31
##       Year      UE_monthlyRate     dayOfMonth
##  Min.   :2007   Min.   : 3.500   Min.   : 1.00
##  1st Qu.:2010   1st Qu.: 4.600   1st Qu.: 8.00
##  Median :2013   Median : 5.600   Median :16.00
##  Mean   :2013   Mean   : 6.282   Mean   :15.74
##  3rd Qu.:2016   3rd Qu.: 8.200   3rd Qu.:23.00
##  Max.   :2020   Max.   :10.000   Max.   :31.00
```

```
dailyChangesColSums <- as.data.frame(colSums(DailyChanges[1:53]))
colnames(dailyChangesColSums) <- 'avgDailyChange_2007_2020'
row.names(dailyChangesColSums) <-
gsub('_dailyChange','',row.names(dailyChangesColSums))
head(dailyChangesColSums,5)
```

```
##      avgDailyChange_2007_2020
## TGT                    106.91
## FTR                      0.87
## UBSI                    37.87
## HD                     232.02
## JPM                    117.69
```

The DOW Industrial Jones average was also downloaded from Yahoo Finance to see a bigger picture of these daily changes by adding in the change in the DOW. We will upload it to our data and put the daily change values into a new column with the Close of the DOW daily.

```
dow <- read.csv('DOW.csv', sep=',', header=T, na.strings=c('',' '))
head(dow)
```

```
##         Date     Open     High      Low    Close Adj.Close    Volume
## 1 2007-01-03 12459.54 12580.35 12404.82 12474.52  12474.52 327200000
## 2 2007-01-04 12473.16 12510.41 12403.86 12480.69  12480.69 259060000
## 3 2007-01-05 12480.05 12480.13 12365.41 12398.01  12398.01 235220000
## 4 2007-01-08 12392.01 12445.92 12337.37 12423.49  12423.49 223500000
## 5 2007-01-09 12424.77 12466.43 12369.17 12416.60  12416.60 225190000
## 6 2007-01-10 12417.00 12451.61 12355.63 12442.16  12442.16 226570000
```

Lets keep the date, close, and volume columns.

```
dow1 <- dow[,c(1,5,7)]
colnames(dow1) <- c('Date','DOW_Daily_Close','DOW_Daily_Volume')
head(dow1)
```

```
##         Date DOW_Daily_Close DOW_Daily_Volume
## 1 2007-01-03        12474.52        327200000
## 2 2007-01-04        12480.69        259060000
## 3 2007-01-05        12398.01        235220000
## 4 2007-01-08        12423.49        223500000
## 5 2007-01-09        12416.60        225190000
## 6 2007-01-10        12442.16        226570000
```

Now add in a daily change column to the dow1 table.

```
dow_a <- dow1$DOW_Daily_Close
dow_b <- c(0,dow_a)
dow_c <- dow_b[1:(length(dow_b)-1)]
dow1$DOW_Daily_Change <- dow1$DOW_Daily_Close-dow_c
head(dow1)
```

```
##         Date DOW_Daily_Close DOW_Daily_Volume DOW_Daily_Change
## 1 2007-01-03        12474.52        327200000     12474.519531
## 2 2007-01-04        12480.69        259060000         6.170899
## 3 2007-01-05        12398.01        235220000       -82.680664
## 4 2007-01-08        12423.49        223500000        25.480468
## 5 2007-01-09        12416.60        225190000        -6.890625
## 6 2007-01-10        12442.16        226570000        25.560547
```

Lets attach the daily change of the DOW to the table of daily changes per stock we made earlier and compare.

```
dow1$Date <- as.Date(dow1$Date)
DailyChanges2 <- merge(DailyChanges, dow1, by.x='Date', by.y='Date')
colnames(DailyChanges2)
```

```
##  [1] "Date"                "TGT_dailyChange"     "FTR_dailyChange"
##  [4] "UBSI_dailyChange"    "HD_dailyChange"      "JPM_dailyChange"
##  [7] "XOM_dailyChange"     "CVX_dailyChange"     "NSANY_dailyChange"
## [10] "MGM_dailyChange"     "TEVA_dailyChange"    "HST_dailyChange"
## [13] "WFC_dailyChange"     "WWE_dailyChange"     "INO_dailyChange"
## [16] "SCE.PB_dailyChange" "FFIN_dailyChange"    "GOOG_dailyChange"
## [19] "WM_dailyChange"      "ONCY_dailyChange"    "S_dailyChange"
## [22] "F_dailyChange"       "ARWR_dailyChange"    "COST_dailyChange"
## [25] "AAL_dailyChange"     "JWN_dailyChange"     "NUS_dailyChange"
## [28] "HMC_dailyChange"     "AMZN_dailyChange"    "T_dailyChange"
## [31] "HRB_dailyChange"     "RRGB_dailyChange"    "ADDYY_dailyChange"
## [34] "PCG_dailyChange"     "ROST_dailyChange"    "JNJ_dailyChange"
## [37] "NFLX_dailyChange"    "M_dailyChange"       "KSS_dailyChange"
## [40] "DLTR_dailyChange"    "WMT_dailyChange"     "C_dailyChange"
```

```
## [43] "AAP_dailyChange"    "JBLU_dailyChange"    "MSFT_dailyChange"
## [46] "KGJI_dailyChange"   "EPD_dailyChange"     "TJX_dailyChange"
## [49] "HOFT_dailyChange"   "LUV_dailyChange"     "NKE_dailyChange"
## [52] "TM_dailyChange"     "VZ_dailyChange"      "SIG_dailyChange"
## [55] "DayOfWeek"          "Month"               "Year"
## [58] "UE_monthlyRate"     "dayOfMonth"          "DOW_Daily_Close"
## [61] "DOW_Daily_Volume"   "DOW_Daily_Change"
```

Lets add an indicator for increasing or decreasing unemployment rate per month.

```
DailyChanges2$lastMonth_UE_rate <-
  c(DailyChanges2$UE_monthlyRate[1],
    DailyChanges2$UE_monthlyRate[1:length(DailyChanges2$UE_monthlyRate)-1])

DailyChanges2$increasingMonthly_UE_rate <-
ifelse((DailyChanges2$UE_monthlyRate-DailyChanges2$lastMonth_UE_rate) > 0, 1,
0)
```

Save this file to csv.

```
write.csv(DailyChanges2, 'DailyChanges_UE_DOW_07_20.csv', row.names=FALSE)
```

Lets see a summary of our date with summaries when the unemployment rate increased the next month and the DOW daily changes increased the next day and separately a subset of the DOW decreasing daily.This will see if the DOW is affected by the increasing unemployment rate or not. And also show which stocks are increasing when the DOW is decreasing and unemployment rate increasing to indicate great public sentiment for those stocks during poor public sentiment about the state of the economy.

```
dow_up_ue_up <- subset(DailyChanges2,
DailyChanges2$increasingMonthly_UE_rate==1 &
                         DailyChanges2$DOW_Daily_Change >= 0)

dow_down_ue_up <- subset(DailyChanges2,
DailyChanges2$increasingMonthly_UE_rate==1 &
                         DailyChanges2$DOW_Daily_Change < 0)
```

Summary of the DOW up and unemployment up:

```
summary(dow_up_ue_up)

##       Date            TGT_dailyChange    FTR_dailyChange  UBSI_dailyChange
##  Min.   :2007-04-02   Min.   :-46.9000   Min.   :-81.00   Min.   :-10.700
##  1st Qu.:2009-04-16   1st Qu.:-13.2650   1st Qu.:-40.95   1st Qu.: -4.000
##  Median :2013-01-02   Median : -6.5100   Median : -3.15   Median : -1.120
##  Mean   :2013-03-11   Mean   : -0.7819   Mean   : 13.59   Mean   :  0.489
##  3rd Qu.:2017-02-15   3rd Qu.:  7.5500   3rd Qu.: 10.20   3rd Qu.:  3.860
##  Max.   :2020-01-02   Max.   : 59.9300   Max.   :235.68   Max.   : 34.800
##  HD_dailyChange      JPM_dailyChange   XOM_dailyChange   CVX_dailyChange
##  Min.   :-198.220   Min.   :-83.210   Min.   :-23.840   Min.   :-40.880
##  1st Qu.:  -6.020   1st Qu.: -7.530   1st Qu.: -7.665   1st Qu.:-15.290
```

```
##  Median :   2.790   Median :  1.840   Median : -2.420   Median :  3.410  
##  Mean   :  -8.685   Mean   : -2.735   Mean   :  1.085   Mean   : -1.455  
##  3rd Qu.:  22.855   3rd Qu.:  7.700   3rd Qu.:  6.285   3rd Qu.:  8.180  
##  Max.   :  41.820   Max.   : 48.240   Max.   : 76.160   Max.   : 74.830  
##  NSANY_dailyChange  MGM_dailyChange   TEVA_dailyChange  HST_dailyChange   
##  Min.   :-11.7900   Min.   :-57.280   Min.   :-23.680   Min.   :-13.0103  
##  1st Qu.: -4.2400   1st Qu.: -8.315   1st Qu.: -6.195   1st Qu.: -4.3800  
##  Median :  0.2600   Median : -0.780   Median : -1.280   Median :  0.1700  
##  Mean   :  0.1368   Mean   : -1.060   Mean   :  3.095   Mean   :  0.1865  
##  3rd Qu.:  2.1000   3rd Qu.:  5.685   3rd Qu.:  9.995   3rd Qu.:  4.6449  
##  Max.   : 21.5400   Max.   : 71.580   Max.   : 37.410   Max.   : 26.2066  
##  WFC_dailyChange   WWE_dailyChange   INO_dailyChange    SCE.PB_dailyChange
##  Min.   :-18.840   Min.   :-55.090   Min.   :-10.8800   Min.   :-878.000  
##  1st Qu.: -7.195   1st Qu.: -4.690   1st Qu.: -1.8000   1st Qu.:-414.000  
##  Median : -2.110   Median :  0.290   Median : -1.0000   Median :  16.000  
##  Mean   : -1.618   Mean   :  1.587   Mean   :  0.3177   Mean   :   6.774  
##  3rd Qu.:  3.285   3rd Qu.:  4.820   3rd Qu.:  3.2150   3rd Qu.: 362.500  
##  Max.   : 34.010   Max.   : 67.690   Max.   : 13.3200   Max.   : 857.000  
##  FFIN_dailyChange   GOOG_dailyChange   WM_dailyChange    ONCY_dailyChange 
##  Min.   :-27.0050   Min.   :-1170.03   Min.   :-81.960   Min.   :-15.105  
##  1st Qu.: -0.1725   1st Qu.:  -72.16   1st Qu.: -1.530   1st Qu.: -4.037  
##  Median :  0.8500   Median :   32.93   Median :  2.970   Median : -0.760  
##  Mean   : -0.6908   Mean   :  -41.75   Mean   : -2.227   Mean   :  2.146  
##  3rd Qu.:  3.0767   3rd Qu.:  168.05   3rd Qu.: 10.125   3rd Qu.:  5.035  
##  Max.   :  9.8400   Max.   :  267.46   Max.   : 34.650   Max.   : 28.690  
##  S_dailyChange      F_dailyChange     ARWR_dailyChange  COST_dailyChange 
##  Min.   :-14.8300   Min.   :-5.520   Min.   :-49.500   Min.   :-245.270  
##  1st Qu.: -4.0250   1st Qu.:-1.775   1st Qu.: -8.215   1st Qu.:  -8.480  
##  Median : -0.1800   Median :-0.600   Median : -0.470   Median :  12.580  
##  Mean   :  0.3581   Mean   : 0.129   Mean   :  1.236   Mean   :  -8.776  
##  3rd Qu.:  1.8500   3rd Qu.: 0.800   3rd Qu.:  9.550   3rd Qu.:  24.215  
##  Max.   : 19.3000   Max.   : 9.780   Max.   : 49.000   Max.   :  76.860  
##  AAL_dailyChange    JWN_dailyChange   NUS_dailyChange   HMC_dailyChange  
##  Min.   :-28.6000   Min.   :-36.350   Min.   :-42.910   Min.   :-9.7300  
##  1st Qu.: -6.9450   1st Qu.:-16.185   1st Qu.:-13.845   1st Qu.:-4.3500  
##  Median :  0.0500   Median : -1.520   Median : -2.010   Median :-0.9600  
##  Mean   : -0.2045   Mean   : -1.158   Mean   : -4.774   Mean   : 0.7997  
##  3rd Qu.:  5.3500   3rd Qu.:  7.755   3rd Qu.:  4.405   3rd Qu.: 3.1850  
##  Max.   : 45.0500   Max.   : 53.740   Max.   : 20.330   Max.   :34.6900  
##  AMZN_dailyChange   T_dailyChange       HRB_dailyChange  
RRGB_dailyChange
##  Min.   :-1939.11   Min.   :-13.89000   Min.   :-14.6300   Min.   :-33.04  
##  1st Qu.:   -3.79   1st Qu.: -5.51000   1st Qu.: -3.9850   1st Qu.:-19.39  
##  Median :   40.42   Median : -0.20000   Median : -0.1100   Median :  1.02  
##  Mean   :  -33.22   Mean   :  0.02839   Mean   : -0.5958   Mean   : -4.03  
##  3rd Qu.:  216.99   3rd Qu.:  3.72000   3rd Qu.:  2.0500   3rd Qu.:  7.43  
##  Max.   :  899.08   Max.   : 39.46000   Max.   : 21.2600   Max.   : 39.10  
##  ADDYY_dailyChange  PCG_dailyChange   ROST_dailyChange    JNJ_dailyChange  
##  Min.   :-127.610   Min.   :-24.670   Min.   :-104.5000   Min.   :-87.020  
##  1st Qu.: -12.300   1st Qu.: -2.170   1st Qu.:   0.4012   1st Qu.: -8.945  
```

```
##  Median :   3.770   Median :  0.700   Median :   1.9225   Median :  0.850
##  Mean   :  -5.184   Mean   :  3.108   Mean   :  -4.3563   Mean   : -2.412
##  3rd Qu.:  16.255   3rd Qu.:  6.950   3rd Qu.:   6.1050   3rd Qu.: 11.580
##  Max.   :  44.310   Max.   : 49.440   Max.   :  35.0900   Max.   : 60.100
##  NFLX_dailyChange   M_dailyChange      KSS_dailyChange   DLTR_dailyChange
##  Min.   :-342.307   Min.   :-34.4400   Min.   :-31.740   Min.   :-95.987
##  1st Qu.:  -3.334   1st Qu.:-14.3050   1st Qu.: -9.950   1st Qu.: -3.310
##  Median :   2.124   Median :  1.1500   Median : -1.500   Median :  2.647
##  Mean   :  -4.954   Mean   : -0.9623   Mean   :  1.922   Mean   : -5.528
##  3rd Qu.:  26.558   3rd Qu.: 10.6000   3rd Qu.: 10.290   3rd Qu.:  8.771
##  Max.   : 210.520   Max.   : 45.0600   Max.   : 77.490   Max.   : 24.700
##  WMT_dailyChange   C_dailyChange      AAP_dailyChange     JBLU_dailyChange
##  Min.   :-73.960   Min.   :-294.400   Min.   :-127.1000   Min.   :-9.9600
##  1st Qu.: -6.650   1st Qu.: -17.600   1st Qu.:  -4.7650   1st Qu.:-3.5800
##  Median :  0.540   Median :  -1.570   Median :   4.3000   Median :-0.0200
##  Mean   : -3.403   Mean   :  11.267   Mean   :   0.7365   Mean   :-0.1635
##  3rd Qu.:  8.530   3rd Qu.:   9.565   3rd Qu.:  23.3150   3rd Qu.: 1.6400
##  Max.   : 47.400   Max.   : 510.500   Max.   :  87.8600   Max.   :11.9000
##  MSFT_dailyChange   KGJI_dailyChange  EPD_dailyChange    TJX_dailyChange
##  Min.   :-140.490   Min.   :-1.9700   Min.   :-14.5100   Min.   :-52.1600
##  1st Qu.:  -4.260   1st Qu.:-0.6050   1st Qu.: -3.1300   1st Qu.: -0.6175
##  Median :   6.110   Median :-0.3200   Median : -0.1400   Median :  1.8675
##  Mean   :  -3.038   Mean   : 0.2361   Mean   : -0.5506   Mean   : -1.7903
##  3rd Qu.:  14.225   3rd Qu.: 0.5200   3rd Qu.:  2.9700   3rd Qu.:  4.8825
##  Max.   :  56.190   Max.   : 7.5600   Max.   : 15.8750   Max.   : 21.2150
##  HOFT_dailyChange   LUV_dailyChange   NKE_dailyChange   TM_dailyChange
##  Min.   :-19.6400   Min.   :-40.830   Min.   :-81.700   Min.   :-41.950
##  1st Qu.: -4.4550   1st Qu.: -6.485   1st Qu.: -1.359   1st Qu.:-19.645
##  Median : -2.1300   Median : -1.450   Median :  2.765   Median : -6.110
##  Mean   : -0.5397   Mean   : -2.059   Mean   : -2.018   Mean   : -2.171
##  3rd Qu.:  3.1850   3rd Qu.:  6.880   3rd Qu.: 11.512   3rd Qu.:  4.820
##  Max.   : 20.9100   Max.   : 21.740   Max.   : 29.470   Max.   :126.920
##  VZ_dailyChange     SIG_dailyChange     DayOfWeek             Month
##  Min.   :-20.6356   Min.   :-55.5000   Length:31          Length:31
##  1st Qu.: -3.7856   1st Qu.:-19.9650   Class :character   Class :character
##  Median : -0.1100   Median : -1.1300   Mode  :character   Mode  :character
##  Mean   : -0.1209   Mean   : -0.4039
##  3rd Qu.:  4.5250   3rd Qu.: 17.5700
##  Max.   : 35.3161   Max.   : 50.7200
##       Year      UE_monthlyRate    dayOfMonth    DOW_Daily_Close
##  Min.   :2007   Min.   :3.600   Min.   :1.000   Min.   : 7762
##  1st Qu.:2009   1st Qu.:4.450   1st Qu.:1.000   1st Qu.:11253
##  Median :2013   Median :5.100   Median :1.000   Median :13668
##  Mean   :2013   Mean   :6.142   Mean   :1.581   Mean   :16184
##  3rd Qu.:2016   3rd Qu.:8.500   3rd Qu.:2.000   3rd Qu.:20192
##  Max.   :2020   Max.   :9.900   Max.   :4.000   Max.   :28869
##  DOW_Daily_Volume    DOW_Daily_Change  lastMonth_UE_rate
##  Min.   : 74050000   Min.   :  2.471   Min.   :3.500
##  1st Qu.:148865000   1st Qu.: 36.360   1st Qu.:4.350
##  Median :213700000   Median :114.949   Median :5.000
```

```
##  Mean   :217756129   Mean   :129.052   Mean   :5.977
##  3rd Qu.:307650000   3rd Qu.:200.670   3rd Qu.:8.300
##  Max.   :388480000   Max.   :348.580   Max.   :9.800
##  increasingMonthly_UE_rate
##  Min.   :1
##  1st Qu.:1
##  Median :1
##  Mean   :1
##  3rd Qu.:1
##  Max.   :1
```

Summary of the DOW down and unemployment down:

```
summary(dow_down_ue_up)
```

```
##       Date             TGT_dailyChange   FTR_dailyChange   UBSI_dailyChange
##  Min.   :2007-12-03   Min.   :-47.520   Min.   :-72.150   Min.   :-16.220
##  1st Qu.:2008-10-17   1st Qu.:-20.555   1st Qu.:-53.700   1st Qu.: -5.790
##  Median :2009-09-01   Median : -7.550   Median : -5.250   Median : -0.850
##  Mean   :2011-04-15   Mean   : -8.380   Mean   :-11.453   Mean   : -3.033
##  3rd Qu.:2013-09-02   3rd Qu.: -1.405   3rd Qu.:  2.175   3rd Qu.:  1.535
##  Max.   :2019-10-01   Max.   : 22.310   Max.   :192.700   Max.   :  4.450
##  HD_dailyChange     JPM_dailyChange   XOM_dailyChange    CVX_dailyChange
##  Min.   :-199.420   Min.   :-64.650   Min.   :-19.6700   Min.   :-29.880
##  1st Qu.:  -7.195   1st Qu.: -7.430   1st Qu.:-10.5250   1st Qu.:-14.040
##  Median :   0.670   Median : -0.300   Median :  0.5000   Median : -5.660
##  Mean   :  -6.023   Mean   : -5.855   Mean   : -0.9374   Mean   : -2.875
##  3rd Qu.:  11.375   3rd Qu.:  4.255   3rd Qu.:  7.5250   3rd Qu.:  8.705
##  Max.   :  55.200   Max.   :  7.560   Max.   : 24.9600   Max.   : 34.630
##  NSANY_dailyChange MGM_dailyChange  TEVA_dailyChange  HST_dailyChange
##  Min.   :-15.640   Min.   :-73.43   Min.   :-13.220   Min.   :-12.247
##  1st Qu.: -7.670   1st Qu.:-53.83   1st Qu.: -3.925   1st Qu.: -9.090
##  Median : -3.450   Median : -5.75   Median :  2.760   Median : -2.130
##  Mean   : -3.421   Mean   :-18.56   Mean   :  3.083   Mean   : -3.406
##  3rd Qu.:  1.000   3rd Qu.:  3.68   3rd Qu.:  6.490   3rd Qu.:  1.055
##  Max.   : 10.350   Max.   : 56.48   Max.   : 38.180   Max.   :  4.810
##  WFC_dailyChange   WWE_dailyChange   INO_dailyChange   SCE.PB_dailyChange
##  Min.   :-18.260   Min.   :-55.790   Min.   :-8.8000   Min.   :-879.0
##  1st Qu.: -7.755   1st Qu.: -3.205   1st Qu.:-2.6800   1st Qu.:-608.5
##  Median : -5.520   Median : -1.320   Median :-1.0000   Median : -58.0
##  Mean   : -2.947   Mean   : -3.239   Mean   :-0.1611   Mean   :-103.1
##  3rd Qu.:  3.100   3rd Qu.:  0.610   3rd Qu.: 3.0900   3rd Qu.: 285.0
##  Max.   :  9.350   Max.   : 15.740   Max.   : 7.2000   Max.   : 729.0
##  FFIN_dailyChange   GOOG_dailyChange  WM_dailyChange     ONCY_dailyChange
##  Min.   :-24.2067   Min.   :-848.61   Min.   :-84.980   Min.   :-19.950
##  1st Qu.: -0.1592   1st Qu.: -60.75   1st Qu.: -3.010   1st Qu.: -8.645
##  Median :  1.0133   Median :  25.88   Median : -1.300   Median : -5.700
##  Mean   : -0.1642   Mean   : -35.57   Mean   : -2.921   Mean   : -1.166
##  3rd Qu.:  1.8629   3rd Qu.:  79.80   3rd Qu.:  3.305   3rd Qu.:  2.708
##  Max.   :  3.6650   Max.   : 165.49   Max.   : 25.400   Max.   : 34.010
```

```
##  S_dailyChange     F_dailyChange    ARWR_dailyChange  COST_dailyChange
##  Min.   :-11.980   Min.   :-5.380   Min.   :-31.200   Min.   :-227.820
##  1st Qu.:-10.695   1st Qu.:-3.500   1st Qu.:-25.150   1st Qu.: -14.105
##  Median : -2.190   Median :-1.710   Median : -5.500   Median :   5.930
##  Mean   : -3.012   Mean   :-1.140   Mean   : -9.518   Mean   :  -6.886
##  3rd Qu.:  1.510   3rd Qu.: 1.135   3rd Qu.: -0.890   3rd Qu.:  14.720
##  Max.   :  9.050   Max.   : 4.780   Max.   : 15.420   Max.   :  55.300
##  AAL_dailyChange   JWN_dailyChange   NUS_dailyChange   HMC_dailyChange
##  Min.   :-33.490   Min.   :-32.290   Min.   :-75.570   Min.   :-13.200
##  1st Qu.: -9.750   1st Qu.:-19.460   1st Qu.: -8.070   1st Qu.: -6.905
##  Median : -6.380   Median :-12.070   Median : -1.180   Median : -2.430
##  Mean   : -4.933   Mean   : -8.785   Mean   : -3.667   Mean   : -2.399
##  3rd Qu.:  3.420   3rd Qu.:  3.345   3rd Qu.:  3.035   3rd Qu.:  1.020
##  Max.   : 20.530   Max.   : 12.880   Max.   : 41.580   Max.   :  9.830
##  AMZN_dailyChange   T_dailyChange     HRB_dailyChange    RRGB_dailyChange
##  Min.   :-1685.38   Min.   :-15.260   Min.   :-10.1000   Min.   :-29.840
##  1st Qu.:  -14.44   1st Qu.: -9.550   1st Qu.: -2.5000   1st Qu.:-21.235
##  Median :   11.82   Median : -2.150   Median : -0.0300   Median : -7.640
##  Mean   :  -58.88   Mean   : -2.689   Mean   :  0.1426   Mean   : -6.002
##  3rd Qu.:   57.96   3rd Qu.:  3.460   3rd Qu.:  2.2100   3rd Qu.:  4.840
##  Max.   :  254.85   Max.   :  9.980   Max.   : 14.5700   Max.   : 20.720
##  ADDYY_dailyChange  PCG_dailyChange    ROST_dailyChange    JNJ_dailyChange
##  Min.   :-114.700   Min.   :-36.9100   Min.   :-99.39750   Min.   :-
60.6500
##  1st Qu.: -14.550   1st Qu.: -6.8350   1st Qu.:  0.06875   1st Qu.: -
7.9600
##  Median :   0.950   Median :  0.5500   Median :  2.10500   Median :
2.6200
##  Mean   :  -4.267   Mean   : -0.7774   Mean   : -2.43053   Mean   : -
0.2811
##  3rd Qu.:   9.335   3rd Qu.:  2.4800   3rd Qu.:  3.72000   3rd Qu.:
9.0450
##  Max.   :  45.600   Max.   : 36.1400   Max.   : 12.96000   Max.   :
31.4200
##  NFLX_dailyChange     M_dailyChange     KSS_dailyChange   DLTR_dailyChange
##  Min.   :-290.35286   Min.   :-23.290   Min.   :-31.300   Min.   :-91.7233
##  1st Qu.:  -0.00286   1st Qu.:-17.480   1st Qu.:-16.285   1st Qu.: -1.3917
##  Median :   1.10714   Median :-13.550   Median : -7.260   Median :  3.6000
##  Mean   : -12.92707   Mean   : -6.327   Mean   : -7.433   Mean   : -0.2938
##  3rd Qu.:   3.03357   3rd Qu.:  5.530   3rd Qu.:  2.655   3rd Qu.:  4.8650
##  Max.   :  30.90000   Max.   : 14.980   Max.   : 20.100   Max.   : 28.9500
##  WMT_dailyChange   C_dailyChange     AAP_dailyChange    JBLU_dailyChange
##  Min.   :-66.390   Min.   :-298.30   Min.   :-102.160   Min.   :-10.390
##  1st Qu.: -3.830   1st Qu.:-201.40   1st Qu.:  -5.170   1st Qu.: -4.100
##  Median :  5.480   Median : -91.20   Median :   2.260   Median : -0.200
##  Mean   :  1.438   Mean   : -97.17   Mean   :   1.378   Mean   : -1.680
##  3rd Qu.: 10.335   3rd Qu.:   2.37   3rd Qu.:   7.135   3rd Qu.:  0.445
##  Max.   : 17.570   Max.   : 266.25   Max.   :  46.750   Max.   :  2.850
##  MSFT_dailyChange   KGJI_dailyChange  EPD_dailyChange    TJX_dailyChange
##  Min.   :-104.940   Min.   :-3.7200   Min.   :-12.7250   Min.   :-47.8125
```

```
##  1st Qu.:  -9.850   1st Qu.:-0.6300   1st Qu.: -3.2000   1st Qu.: -0.2062
##  Median :  -1.670   Median :-0.3400   Median : -0.0800   Median :  1.0200
##  Mean   :  -5.594   Mean   :-0.4168   Mean   : -0.3142   Mean   : -1.1578
##  3rd Qu.:   3.825   3rd Qu.:-0.1200   3rd Qu.:  2.5650   3rd Qu.:  2.5537
##  Max.   :  30.260   Max.   : 1.4100   Max.   :  6.8200   Max.   :  5.8875
##  HOFT_dailyChange  LUV_dailyChange   NKE_dailyChange   TM_dailyChange
##  Min.   :-15.410   Min.   :-38.580   Min.   :-67.750   Min.   :-47.61
##  1st Qu.: -5.810   1st Qu.: -3.530   1st Qu.: -2.794   1st Qu.:-30.57
##  Median : -1.770   Median : -0.010   Median :  0.765   Median : -7.82
##  Mean   : -3.428   Mean   : -1.265   Mean   : -2.084   Mean   :-12.20
##  3rd Qu.:  0.075   3rd Qu.:  2.575   3rd Qu.:  2.931   3rd Qu.:  5.45
##  Max.   :  4.060   Max.   : 16.720   Max.   : 17.240   Max.   : 19.53
##  VZ_dailyChange    SIG_dailyChange    DayOfWeek            Month
##  Min.   :-17.933   Min.   :-51.310   Length:19          Length:19
##  1st Qu.: -7.713   1st Qu.:-18.580   Class :character   Class :character
##  Median : -1.522   Median :-15.490   Mode  :character   Mode  :character
##  Mean   : -2.523   Mean   : -3.959
##  3rd Qu.:  1.765   3rd Qu.: 15.620
##  Max.   :  9.990   Max.   : 42.810
##       Year      UE_monthlyRate   dayOfMonth    DOW_Daily_Close
##  Min.   :2007   Min.   : 3.6   Min.   :1.000   Min.   : 6763
##  1st Qu.:2008   1st Qu.: 5.7   1st Qu.:1.000   1st Qu.: 9415
##  Median :2009   Median : 6.7   Median :1.000   Median :12290
##  Mean   :2011   Mean   : 6.9   Mean   :1.789   Mean   :12912
##  3rd Qu.:2014   3rd Qu.: 8.1   3rd Qu.:3.000   3rd Qu.:14973
##  Max.   :2019   Max.   :10.0   Max.   :3.000   Max.   :26573
##  DOW_Daily_Volume    DOW_Daily_Change  lastMonth_UE_rate
##  Min.   : 75630000   Min.   :-679.95   Min.   :3.500
##  1st Qu.:131975000   1st Qu.:-241.33   1st Qu.:5.550
##  Median :199090000   Median : -59.98   Median :6.500
##  Mean   :213377895   Mean   :-147.22   Mean   :6.668
##  3rd Qu.:263550000   3rd Qu.: -23.12   3rd Qu.:7.750
##  Max.   :568670000   Max.   :  -5.18   Max.   :9.800
##  increasingMonthly_UE_rate
##  Min.   :1
##  1st Qu.:1
##  Median :1
##  Mean   :1
##  3rd Qu.:1
##  Max.   :1
```

From the above subset of stock daily changes during a time of increasing monthly unemployment rate and decreasing DOW daily value, there are only three stocks that all had increasing daily median and mean values for those time periods: TEVA, WMT, and AAP. There are some stocks that only had median increasing values: HD, XOM, FFIN, GOOG, COST, AMZN, ADDY, PCG, ROST, JNJ, NFLX, DLTR, TJX, and NKE. One stock only had an increasing daily change mean value but not median value: HRB.

Lets look at these stocks that increased during decreasing public outlook on economy assumed from decreasing DOW value (losses in investments/future/retirement) and increasing unemployment (more people not working) from month before.

```
stocksGood <- subset(stockNames, stockNames$stock == 'TEVA' |
                       stockNames$stock == 'WMT'|
                       stockNames$stock == 'AAP'|
                       stockNames$stock == 'HD'|
                       stockNames$stock == 'XOM'|
                       stockNames$stock == 'FFIN'|
                       stockNames$stock == 'GOOG'|
                       stockNames$stock == 'COST'|
                       stockNames$stock == 'AMZN'|
                       stockNames$stock == 'ADDY'|
                       stockNames$stock == 'PCG'|
                       stockNames$stock == 'ROST'|
                       stockNames$stock == 'JNJ'|
                       stockNames$stock == 'NFLX'|
                       stockNames$stock == 'DLTR'|
                       stockNames$stock == 'TJX'|
                       stockNames$stock == 'NKE'|
                       stockNames$stock == 'HRB')
stocksGood$stockInfo
```

```
##  [1] The Home Depot, Inc. (HD)\nNYSE - NYSE Delayed Price. Currency in USD
##  [2] Exxon Mobil Corporation (XOM)\nNYSE - NYSE Delayed Price. Currency in
USD
##  [3] Teva Pharmaceutical Industries Limited (TEVA)\nNYSE - NYSE Delayed
Price. Currency in USD
##  [4] First Financial Bankshares, Inc. (FFIN)\nNasdaqGS - NasdaqGS Real
Time Price. Currency in USD
##  [5] Alphabet Inc. (GOOG)\nNasdaqGS - NasdaqGS Real Time Price. Currency
in USD
##  [6] Costco Wholesale Corporation (COST)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
##  [7] PG&E Corporation (PCG)\nNYSE - NYSE Delayed Price. Currency in USD
##  [8] Dollar Tree, Inc. (DLTR)\nNasdaqGS - NasdaqGS Real Time Price.
Currency in USD
##  [9] NIKE, Inc. (NKE)\nNYSE - NYSE Delayed Price. Currency in USD
## [10] Amazon.com, Inc. (AMZN)\nNasdaqGS - NasdaqGS Real Time Price.
Currency in USD
## [11] Ross Stores, Inc. (ROST)\nNasdaqGS - NasdaqGS Real Time Price.
Currency in USD
## [12] Walmart Inc. (WMT)\nNYSE - NYSE Delayed Price. Currency in USD
## [13] The TJX Companies, Inc. (TJX)\nNYSE - NYSE Delayed Price. Currency in
USD
## [14] Johnson & Johnson (JNJ)\nNYSE - NYSE Delayed Price. Currency in USD
## [15] H&R Block, Inc. (HRB)\nNYSE - NYSE Delayed Price. Currency in USD
## [16] Netflix, Inc. (NFLX)\nNasdaqGS - NasdaqGS Real Time Price. Currency
in USD
```

```
## [17] Advance Auto Parts, Inc. (AAP)\nNYSE - NYSE Delayed Price. Currency
in USD
## 65 Levels: adidas AG (ADDYY)\nOther OTC - Other OTC Delayed Price.
Currency in USD ...
```

From the above, the stocks of auto parts, cheap department and goods, health and beauty products, Nike sports shoes for people wanting to workout and not spend money to occupy time or to predict an increase in low crime robberies (assumptions made by real person not AI), Google for job searches, Amazon because ever expanding and employing many workers, costco for middle class workers and families, Ross and TJ Maxx for low cost business/dress attire and goods, electric company, fuel, home improvement/repair stores, and low cost movie entertainment at home.

Split the summaries of each table to show those that have mean positive values.

```
S <- as.data.frame(summary(dow_up_ue_up))
S1 <- as.data.frame(summary(dow_down_ue_up))
S <- S[-(1:6),-1]
S1 <- S1[-(1:6),-1]

S$Freq <- as.character(S$Freq)
S1$Freq <- as.character(S1$Freq)

s_a <- strsplit(S$Freq, ':')
s_b <- strsplit(S1$Freq, ':')

S$Stat <- lapply(s_a, '[',1)
S1$Stat <- lapply(s_b, '[',1)

S$Stat <- as.vector(S$Stat)

S$StatValue <- as.numeric(lapply(s_a, '[',2))
S1$StatValue <- as.numeric(lapply(s_b, '[',2))

S_mean <- S[grep('Mean', S$Stat),]
S1_mean <- S1[grep('Mean', S1$Stat),]

Dow_up_ue_up_meanPos <- subset(S_mean, S_mean$StatValue >= 0)
Dow_down_ue_up_meanPos <- subset(S1_mean, S1_mean$StatValue >= 0)

Dow_up_ue_up_meanPos <- Dow_up_ue_up_meanPos[grep('dailyChange',
Dow_up_ue_up_meanPos$Var2),]
Dow_down_ue_up_meanPos <- Dow_down_ue_up_meanPos[grep('dailyChange',

Dow_down_ue_up_meanPos$Var2),]
colnames(Dow_up_ue_up_meanPos)[1] <- 'DOW_up_UE_up'
colnames(Dow_down_ue_up_meanPos)[1] <- 'DOW_down_UE_down'

S_Median <- S[grep('Median', S$Stat),]
```

```
S1_Median <- S1[grep('Median', S1$Stat),]

Dow_up_ue_up_MedianPos <- subset(S_Median, S_Median$StatValue >= 0)
Dow_down_ue_up_MedianPos <- subset(S1_Median, S1_Median$StatValue >= 0)

Dow_up_ue_up_MedianPos <- Dow_up_ue_up_MedianPos[grep('dailyChange',
Dow_up_ue_up_MedianPos$Var2),]
Dow_down_ue_up_MedianPos <- Dow_down_ue_up_MedianPos[grep('dailyChange',

Dow_down_ue_up_MedianPos$Var2),]
colnames(Dow_up_ue_up_MedianPos)[1] <- 'DOW_up_UE_up'
colnames(Dow_down_ue_up_MedianPos)[1] <- 'DOW_down_UE_down'
```

Write those tables to csv to use as needed. We should test how well the same amount invested in the original 53 stocks over the span from 2007-2020 did to the same amount of money using different weights on those stocks whose value of daily changes was positive for the median and mean values separately when the DOW was up and unemployment was up and also when the DOW was down and unemployment was up. The Stat column is a list and won't print to csv without removing it, and the Freq column also has the statistic being evaluated.

```
write.csv(Dow_up_ue_up_meanPos[,-3], 'Dow_up_ue_up_meanPos.csv',
row.names=FALSE)
write.csv(Dow_down_ue_up_meanPos[,-3],'Dow_down_ue_up_meanPos.csv',
row.names=FALSE)
write.csv(Dow_up_ue_up_MedianPos[,-3],'Dow_up_ue_up_MedianPos.csv',
row.names=FALSE)
write.csv(Dow_down_ue_up_MedianPos[,-3],'Dow_down_ue_up_MedianPos.csv',
row.names=FALSE)
```

Now, will make a vector of those stocks that have positive medians and means when the DOW is up or down when unemployment is up.

```
Dow_up_med <- as.data.frame(Dow_up_ue_up_MedianPos$DOW_up_UE_up)
DOW_up_mean <- as.data.frame(Dow_up_ue_up_meanPos$DOW_up_UE_up)
DOW_down_med <- as.data.frame(Dow_down_ue_up_MedianPos$DOW_down_UE_down)
DOW_down_mean <- as.data.frame(Dow_down_ue_up_meanPos$DOW_down_UE_down)

colnames(Dow_up_med) <- 'Dow_up_median'
colnames(DOW_up_mean) <- 'DOW_up_mean'
colnames(DOW_down_med) <- 'DOW_down_median'
colnames(DOW_down_mean) <- 'DOW_down_mean'

DOW_up_mean$DOW_up_mean <- gsub('_dailyChange','', DOW_up_mean$DOW_up_mean)
DOW_down_mean$DOW_down_mean <- gsub('_dailyChange', '',
DOW_down_mean$DOW_down_mean)
Dow_up_med$Dow_up_median <- gsub('_dailyChange', '',
Dow_up_med$Dow_up_median)
DOW_down_med$DOW_down_median <- gsub('_dailyChange', '',
DOW_down_med$DOW_down_median)
```

Lets add the values to these subsets of all 53 original stocks.

```
StockValues <- Close2
StockValues <- StockValues[,-c(1,55:58,60:63)]

colnames(StockValues) <- gsub('.Close','', colnames(StockValues))
StockValues$total <- rowSums(StockValues[1:53])

portfolio53 <- StockValues[order(StockValues$Date,decreasing=FALSE),]

portfolio53 <- portfolio53[c(1,3303),]
portfolio53

##                TGT    FTR  UBSI     HD    JPM   XOM    CVX NSANY   MGM
TEVA
## 2007-01-03  57.18 215.40 39.05  41.07  48.07 74.11  70.97 24.16 57.57
31.26
## 2020-02-14 116.63   0.57 34.36 245.03 137.46 60.65 110.08  9.46 31.52
12.22
##                  HST   WFC   WWE   INO SCE.PB      FFIN     GOOG     WM
ONCY
## 2007-01-03 24.04307 35.74 16.18 13.16    280  6.996666  232.922  37.03
21.945
## 2020-02-14 16.91000 48.22 44.93  4.15    893 34.580002 1520.740 125.75
2.590
##                S    F  ARWR   COST  AAL   JWN   NUS  ADDYY   KSS   MSFT
LUV
## 2007-01-03 19.04 7.51 44.00  52.84 56.3 51.39 18.54  25.00 67.08  29.86
15.52
## 2020-02-14  8.69 8.10 41.27 318.31 29.2 40.28 30.45 156.11 44.47 185.35
57.97
##              HMC   PCG  DLTR KGJI       NKE    AMZN     ROST    WMT
TJX
## 2007-01-03 39.71 47.27 10.23 1.10  12.20875   38.70   7.6225  47.55
7.1675
## 2020-02-14 26.78 16.20 88.68 1.19 103.54000 2134.87 121.7800 117.89
63.3800
##                TM     T    JNJ      C    EPD       VZ   HRB       NFLX
AAP
## 2007-01-03 135.30 34.95  66.40 552.50 14.485 35.30673 23.20   3.801429
35.58
## 2020-02-14 140.15 38.25 150.13  78.79 26.270 58.51000 22.38 380.399994
133.59
##             HOFT   SIG  RRGB     M  JBLU       Date    total
## 2007-01-03 15.47 47.74 35.00 37.51 15.20 2007-01-03 2977.939
## 2020-02-14 22.25 26.07 35.21 16.67 21.27 2020-02-14 8193.300

profit_all <- 8193-2978
profit_all
```

```
## [1] 5215
```

With all 53 stocks from January 1, 2007 throughout February 14, 2020, the portfoliio initially cost 2978 USD and was valued at 8193 USD at the end of that time span. Lets see how much the portfolio is worth when using only the stocks in our subsets of stocks that had positive values when the DOW was up or down but unemployment was increasing. The profit earned was 5215 USD with this portfolio.

```
p53 <- gather(StockValues, 'stock','stockValue', 1:53)
```

The positive median value stock when the DOW was up and unemployment was up.

```
p1 <- merge(Dow_up_med,p53, by.x='Dow_up_median','stock')
P1 <- p1 %>% group_by(Date) %>% summarise_at(vars(stockValue), sum)
P1[c(1,3303),]

## # A tibble: 2 x 2
##   Date       stockValue
##   <date>          <dbl>
## 1 2007-01-03      1294.
## 2 2020-02-14      7189.
```

The initial stock value for the portfolio of stock that had a positive median value when the DOW was up and unemployment was up started at 1294 USD and ended with a value of 7189 USD. Lets weight this portfolio so that we can see the profits in dollars if the initial investment was the same amount as the investment of all 53 stocks.

```
P1_i <- P1$stockValue[1]
P1_l <- P1$stockValue[3303]
P1_i

## [1] 1294.482

P1_l

## [1] 7189.25

profit1 <- P1_l-P1_i
profit1

## [1] 5894.768

r1 <- P1_l/P1_i
r1

## [1] 5.553767

p53i <- portfolio53$total[1]
p53i

## [1] 2977.939
```

```
finalValue_P1 <- p53i*r1
finalValue_P1

## [1] 16538.78

total_P1_profit <- finalValue_P1 - p53i
total_P1_profit

## [1] 13560.84

unique(p1$Dow_up_median)

##  [1] "AAL"     "AAP"     "ADDYY"   "AMZN"    "COST"    "CVX"     "DLTR"    "FFIN"
##  [9] "GOOG"    "HD"      "HST"     "JNJ"     "JPM"     "M"       "MSFT"    "NFLX"
## [17] "NKE"     "NSANY"   "PCG"     "ROST"    "RRGB"    "SCE.PB"  "TJX"     "WM"
## [25] "WMT"     "WWE"
```

From the above values, the initial investment was 1294 USD, and the final value of these stock were 7189 USD in our time series. The profit in dollars earned was 5895 USD. The ratio of final value to initial value was 5.55. The total profit if the same investment amount made as with all 53 stocks was made on these stocks (2978 USD) that had a median positive value when the DOW was up and unemployment was increasing took the initial amount times the ratio of final/initial value added to the difference in the initial price invested in all stocks times the ratio of final/initial stock in this portfolio of 26 stocks. The final value of this portfolio is 16539 USD with profits earned of 13561 USD. ***

The positive median value stock when the DOW was down and unemployment was up.

```
p2 <- merge(DOW_down_med,p53, by.x='DOW_down_median','stock')
P2 <- p2 %>% group_by(Date) %>% summarise_at(vars(stockValue), sum)
P2[c(1,3303),]

## # A tibble: 2 x 2
##   Date       stockValue
##   <date>          <dbl>
## 1 2007-01-03       741.
## 2 2020-02-14      5658.

P2_i <- P2$stockValue[1]
P2_l <- P2$stockValue[3303]
P2_i

## [1] 740.7288

P2_l

## [1] 5658.1

profit2 <- P2_l-P2_i
profit2

## [1] 4917.371
```

```
r2 <- P2_l/P2_i
r2

## [1] 7.638558

p53i <- portfolio53$total[1]
p53i

## [1] 2977.939

total_P2_Value <- p53i*r2
total_P2_Value

## [1] 22747.16

total_P2_profit <- total_P2_Value - p53i
total_P2_profit

## [1] 19769.22

unique(p2$DOW_down_median)

##  [1] "AAP"   "ADDYY" "AMZN"  "COST"  "DLTR"  "FFIN"  "GOOG"  "HD"    "JNJ"
## [10] "NFLX"  "NKE"   "PCG"   "ROST"  "TEVA"  "TJX"   "WMT"   "XOM"
```

The portfolio of stock that had positive median values of daily change when the DOW was down and unemployment was higher than the month before is shown above. There are 17 stocks in this portfolio. The initial value was 741 USD with a final value of 5658 USD and a profit of 4917 USD earned as is. If the same amount was invested in just these stocks as was invested in the entire portfolio of 53 stocks of 2978 USD, then the ratio of final/initial value would be used on adding additional stocks in this portfolio at a ratio of 7.64. The total end value would be 22747 USD with profits earned of 19769 USD.

The positive mean value stock when the DOW was up and unemployment was up.

```
p3 <- merge(DOW_down_mean,p53, by.x='DOW_down_mean','stock')
P3 <- p3 %>% group_by(Date) %>% summarise_at(vars(stockValue), sum)
P3[c(1,3303),]

## # A tibble: 2 x 2
##   Date       stockValue
##   <date>          <dbl>
## 1 2007-01-03       138.
## 2 2020-02-14       286.

P3_i <- P3$stockValue[1]
P3_l <- P3$stockValue[3303]
P3_i

## [1] 137.59
```

```
P3_l

## [1] 286.08

profit3 <- P3_l-P3_i
profit3

## [1] 148.49

r3 <- P3_l/P3_i
r3

## [1] 2.079221

p53i <- portfolio53$total[1]
p53i

## [1] 2977.939

total_P3_Value <- (p53i)*r3
total_P3_Value

## [1] 6191.792

total_P3_profit <- total_P3_Value - p53i
total_P3_profit

## [1] 3213.853

unique(p3$DOW_down_mean)

## [1] "AAP"  "HRB"  "TEVA" "WMT"
```

The above portfolio shows those stocks who had a positive mean value of daily changes when the DOW was down and unemployment was increased more than the month before. There are only four stock in this portfolio. The initial value was 138 USD and the final value was 286 USD. The profits in dollars was 149 USD as is. The ratio of final/initial value is 2.08. When investing the same amount of 2978 USD as was used in the portfolio of 53 stock, the dollars earned were 6192 USD, with profit earned in dollars of 3213 USD.

The positive mean value stock when the DOW was down and unemployment was up.

```
p4 <- merge(DOW_up_mean,p53, by.x='DOW_up_mean','stock')
P4 <- p4 %>% group_by(Date) %>% summarise_at(vars(stockValue), sum)
P4[c(1,3303),]

## # A tibble: 2 x 2
##   Date       stockValue
##   <date>          <dbl>
## 1 2007-01-03      1588.
## 2 2020-02-14      1476.
```

```
P4_i <- P4$stockValue[1]
P4_l <- P4$stockValue[3303]
P4_i

## [1] 1588.048

P4_l

## [1] 1476.17

profit4 <- P4_l-P4_i
profit4

## [1] -111.8781

r4 <- P4_l/P4_i
r4

## [1] 0.9295499

p53i <- portfolio53$total[1]
p53i

## [1] 2977.939

total_P4_Value <- p53i*r4
total_P4_Value

## [1] 2768.143

total_P4_profit <- total_P4_Value - p53i
total_P4_profit

## [1] -209.7959

unique(p4$DOW_up_mean)

##  [1] "AAP"    "ARWR"   "C"      "F"      "FTR"    "HMC"    "HST"    "INO"
##  [9] "KGJI"   "KSS"    "NSANY"  "ONCY"   "PCG"    "S"      "SCE.PB" "T"
## [17] "TEVA"   "UBSI"   "WWE"    "XOM"
```

The above shows those stock in the portfolio that had a positive mean daily change when the DOW was up and unemployment was up. There are 20 stocks in this portfolio. The initial value of this portfolio was 1588 USD and the final value was less at 1476 USD. The loss was 112 USD with a final/initial ratio of 0.93. When investing the same amount as the initial portfolio of 2978 USD, the final portfolio value is a loss of 210 USD.

Lets make a data table of this information.

```
du1 <- as.data.frame( c(P1_i, P2_i,P3_i,P4_i))
du2 <- as.data.frame( c(P1_l,P2_l,P3_l,P4_l))
du3 <-
as.data.frame(c(length(unique(p1$Dow_up_median)),length(unique(p2$DOW_down_me
```

```
dian)),
         length(unique(p3$DOW_down_mean)), length(unique(p4$DOW_up_mean))))
du4 <- as.data.frame( c(profit1, profit2, profit3, profit4))

colnames(du1) <- 'initialValue'
colnames(du2) <- 'finalValue'
colnames(du3) <- 'numberStocksInPortfolio'
colnames(du4) <- 'profitInitialValue'

du5 <- as.data.frame(c(p53i,p53i,p53i,p53i))
colnames(du5) <- 'ifInitialInvestmentAsAll53Made'

du6 <- as.data.frame(c(finalValue_P1,
total_P2_Value,total_P3_Value,total_P4_Value))
colnames(du6) <- 'finalValueIfSame53StockInvestment'

du7 <- as.data.frame(c(total_P1_profit, total_P2_profit, total_P3_profit,
total_P4_profit))
colnames(du7) <- 'totalProfitSame53StockInvestment'

du8 <- as.data.frame(c(r1,r2,r3,r4))
colnames(du8) <- 'ratioFinal_2_Initial'

DOW_UE <- cbind(du1,du2,du8,du4,du3,du5,du6,du7)
row.names(DOW_UE) <-
c('Dow_up_median','DOW_down_median','DOW_down_mean','DOW_up_mean')

write.csv(DOW_UE, 'DOW_UE.csv', row.names=TRUE)

DOW_UE

##                 initialValue finalValue ratioFinal_2_Initial
profitInitialValue
## Dow_up_median      1294.4819    7189.25            5.5537665
5894.7682
## DOW_down_median     740.7288    5658.10            7.6385583
4917.3713
## DOW_down_mean       137.5900     286.08            2.0792208
148.4900
## DOW_up_mean        1588.0481    1476.17            0.9295499          -
111.8781
##                 numberStocksInPortfolio ifInitialInvestmentAsAll53Made
## Dow_up_median                        26                       2977.939
## DOW_down_median                      17                       2977.939
## DOW_down_mean                         4                       2977.939
## DOW_up_mean                          20                       2977.939
##                 finalValueIfSame53StockInvestment
## Dow_up_median                           16538.776
## DOW_down_median                         22747.158
```

```
## DOW_down_mean                                  6191.792
## DOW_up_mean                                    2768.143
##                   totalProfitSame53StockInvestment
## Dow_up_median                               13560.8372
## DOW_down_median                             19769.2190
## DOW_down_mean                                3213.8533
## DOW_up_mean                                  -209.7959
```

In summary of evaluating the stocks that had positive mean and median values when the unemployment rate was more than the previous month, but the DOW was either increasing or decreasing on that day from the previous day, the best portfolio of stocks was the one with the highest profit. The return on investment ratio was 7.63 for this portfolio, with an initial investment of 741 USD it returned 5658 USD, but when the same investment amount was distributed to this portfolio as the entire portfolio of 53 stocks of 2978 USD, the profits made were 19769 USD from 2007 through 2020.

The portfolio of stock that performed the worst with a loss of 111 USD on an initial investment of 1588 USD and a final to initial value ratio of 0.93 was the portfolio of stock that had a positive daily change mean value when the DOW was up and unemployment rate up. When this portfolio had the 2978 USD invested in it as the original portfolio of 53 stocks it saw a loss of 210 USD from 2007-2020.

The portfolio of stocks all having a positive median value of daily changes did much better than the positive mean value stock portfolios when the DOW was up or down and unemployment rate increased from the previous month from 2007-2020.

The question arises when asked on how to distribute the remaining dollars of the initial investment of the original portfolio of all 53 stocks, when that value you want to invest is 2978 USD but the portfolio of single stocks have a set value of 138-1588 USD for the four portfolios.

We would use an even weighted distribution if we could buy partial stocks, but it is likely we will not be able to. If it is the case that we could distribute the weights of the remaining balance to buy partial stocks then you would take that remaining balance and divide by the number of stock in the portfolio. We could do that now to see how much each of the weights are in investment dollars of each stock.

```
DOW_UE$EvenRemainingWeightsUSD <- (DOW_UE$ifInitialInvestmentAsAll53Made-
DOW_UE$initialValue)/
  (DOW_UE$numberStocksInPortfolio)
DOW_UE$EvenWeightsUSD <-
(DOW_UE$ifInitialInvestmentAsAll53Made)/(DOW_UE$numberStocksInPortfolio)
DOW_UE[,c(1,5,9,10)]
```

```
##                   initialValue numberStocksInPortfolio
EvenRemainingWeightsUSD
## Dow_up_median      1294.4819                      26
64.74834
```

```
## DOW_down_median     740.7288                      17
131.60058
## DOW_down_mean       137.5900                       4
710.08715
## DOW_up_mean        1588.0481                      20
69.49453
##                  EvenWeightsUSD
## Dow_up_median         114.5361
## DOW_down_median       175.1729
## DOW_down_mean         744.4847
## DOW_up_mean           148.8969
```

The above table shows even weights after one stock of each is bought and the remaining money from 2978 USD is dispursed equally to each stock in the portfolio in the 'EvenRemainingWeightsUSD' column. The value of the even weights on the amount of dollars to invest in each stock from the total 2978 USD is in the 'EvenWeightUSD' column.

If you are not allowed to buy partial stocks, then you would have to rank the stocks in each portfolio so that more money is spent on the forecasted higher yielding stock.

So, we found a subset of stock in the portfolio that did outstanding, and we want to buy those stocks to make a profit, but we also want to look at the characteristics of those stocks and see what features they have or properties in the data that could make any other stock fit a description of a 'good stock to buy' category. Some features that come to mind are, are they all increasing, are they cyclical, how many local maxima and local minima each of these stocks have, what the sentiment in the internet search engines provide for these stocks, do they market, are they politically motivated such as Nike with the footbal player protesting police abuse of black males, are they part of larger business mergers such as talks of Tmobile getting bought out by Sprint, or how Frontier bought a portion of Verizon, and so on.

Also, we want to look at this as a careless surfer looking for intervals of small waves to buy close to the local minima on these stocks, ride it out and sell it close to its local maxima to simulate how exploiting the stocks in the short run can lead to more profits. We could all do this and just like this line of code is in 1920s, a depression could follow if we all did this. Like a huge crash. But we're all blind, arrogant data scientists in charge of our own way of thinking and we want to see what happens. So lets do it. Did I lose you on the analogy? Which one? Its ok, you'll find yourself for the next part of this data exploration.

Lets look at our 17 stocks belonging to the best subset and see what qualities each stock has by first adding a feature column on the stock brand with an internet search and looking at the local minima and maxima of each stock.

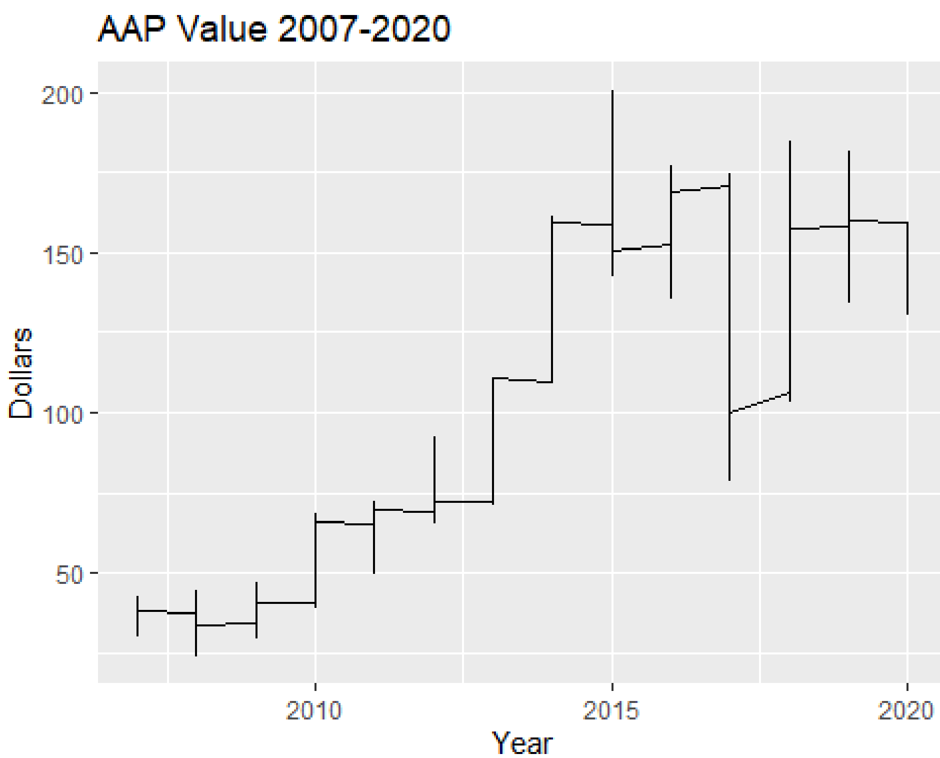
```
set17 <- merge(DOW_down_med, stockNames, by.x='DOW_down_median',
by.y='stock')
set17

##    DOW_down_median
## 1              AAP
## 2            ADDYY
## 3             AMZN
## 4             COST
## 5             DLTR
## 6             FFIN
## 7             GOOG
## 8               HD
## 9              JNJ
## 10            NFLX
## 11             NKE
## 12             PCG
## 13            ROST
## 14            TEVA
## 15             TJX
## 16             WMT
## 17             XOM
##
stockInfo
## 1                         Advance Auto Parts, Inc. (AAP)\nNYSE - NYSE Delayed
Price. Currency in USD
## 2                            adidas AG (ADDYY)\nOther OTC - Other OTC Delayed
Price. Currency in USD
## 3                      Amazon.com, Inc. (AMZN)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 4          Costco Wholesale Corporation (COST)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 5                     Dollar Tree, Inc. (DLTR)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 6     First Financial Bankshares, Inc. (FFIN)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 7                         Alphabet Inc. (GOOG)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 8                              The Home Depot, Inc. (HD)\nNYSE - NYSE Delayed
Price. Currency in USD
## 9                                Johnson & Johnson (JNJ)\nNYSE - NYSE Delayed
Price. Currency in USD
## 10                        Netflix, Inc. (NFLX)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
## 11                                    NIKE, Inc. (NKE)\nNYSE - NYSE Delayed
Price. Currency in USD
## 12                              PG&E Corporation (PCG)\nNYSE - NYSE Delayed
Price. Currency in USD
## 13                   Ross Stores, Inc. (ROST)\nNasdaqGS - NasdaqGS Real Time
Price. Currency in USD
```

```
## 14     Teva Pharmaceutical Industries Limited (TEVA)\nNYSE - NYSE Delayed
Price. Currency in USD
## 15                     The TJX Companies, Inc. (TJX)\nNYSE - NYSE Delayed
Price. Currency in USD
## 16                                Walmart Inc. (WMT)\nNYSE - NYSE Delayed
Price. Currency in USD
## 17                     Exxon Mobil Corporation (XOM)\nNYSE - NYSE Delayed
Price. Currency in USD
##    stockExchange
## 1           NYSE
## 2      Other OTC
## 3         Nasdaq
## 4         Nasdaq
## 5         Nasdaq
## 6         Nasdaq
## 7         Nasdaq
## 8           NYSE
## 9           NYSE
## 10        Nasdaq
## 11          NYSE
## 12          NYSE
## 13        Nasdaq
## 14          NYSE
## 15          NYSE
## 16          NYSE
## 17          NYSE
```

Lets search these companies and add in a feature that gives the number of results for each company.

```
set17$numberSearchReturnMillions <- round(c(0.446, 1.99,541, 0.780, 0.379,
0.0465, 3.51, 0.410, 45.1,
                                      4.46, 4.28, 0.00417, 4.11, 0.00392,
2.59, 1.7, 1.14),4)

closing17 <- Close2[,-c(1,55:58,60)]
colnames(closing17) <- gsub('.Close','',colnames(closing17))
colnames(closing17) <- gsub('.PB','', colnames(closing17))
close17 <- gather(closing17, 'stock','stockValue',1:53)
Close17 <- merge(DOW_down_med, close17, by.x='DOW_down_median', by.y='stock')
Close17 <- Close17[order(Close17$Date),]


dow17 <- dow[,-c(2:4,6,7)]
dow17$Date <- as.Date(dow17$Date)
colnames(dow17)[2] <- 'DOW_Close'
Close17_dow <- merge(Close17, dow17, by.x='Date', by.y='Date')

aap <- subset(Close17_dow, Close17_dow$DOW_down_median=='AAP' )
aap1 <- subset(aap, aap$Year==2017|
                aap$Year==2018|
                aap$Year==2019)
```

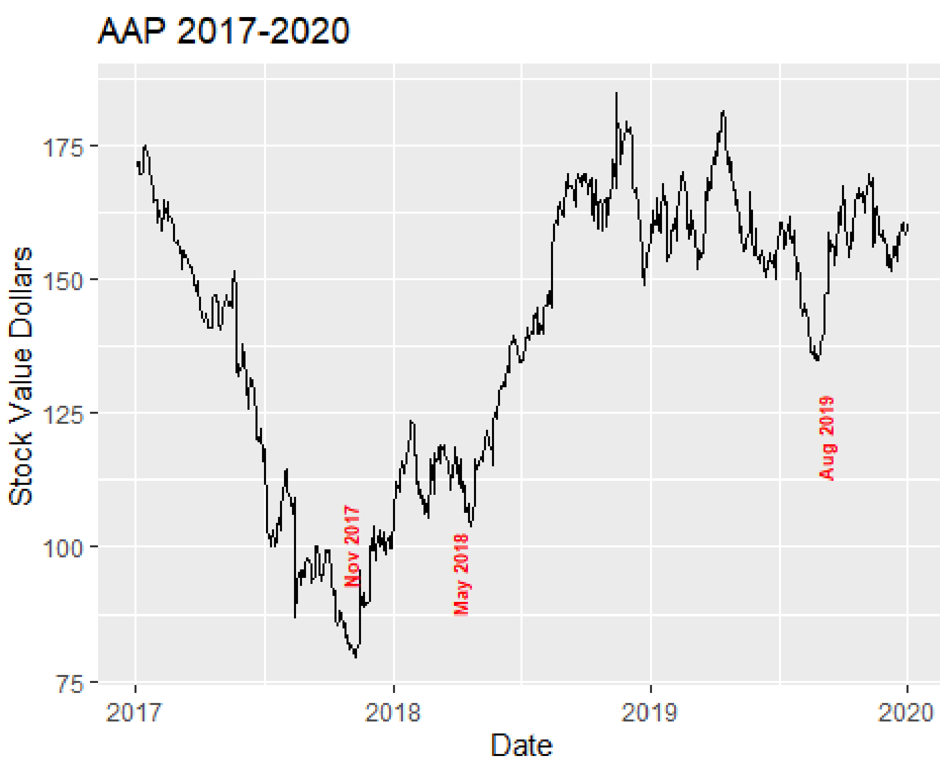
```r
ggplot(data = aap, aes(x=Year, y=stockValue)) +
  geom_line()+
  scale_y_continuous()+
  scale_fill_brewer(palette="paired") +
  theme(legend.position="bottom")+
  ggtitle('AAP Value 2007-2020')+
  ylab('Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```



The above shows that AAP had a huge drop in 2017 at a local minimum, the other minimum is in 2008 and is the global minimum for this stock. The next chart shows the years 2017-2020 to zoom in on this loss.

```r
annotation1 <- data.frame(
   x = c(as.Date('2017-11-01'),as.Date('2018-04-04'),as.Date('2019-09-04')),
   y = c(100,95,120),
   label = c("Nov 2017", "May 2018","Aug 2019"))

gg1 <- ggplot(data = aap1, aes(x=Date, y=stockValue)) +
       geom_line()+
       scale_y_continuous()+
       scale_fill_brewer(palette="paired") +
       theme(legend.position="bottom")+
       ggtitle('AAP 2017-2020')+
         geom_text(data=annotation1, aes( x=x, y=y, label=label),
```

```
,
            color="red",
            size=2.5 , angle=90 , fontface="bold")+
        ylab('Stock Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired

gg1
```



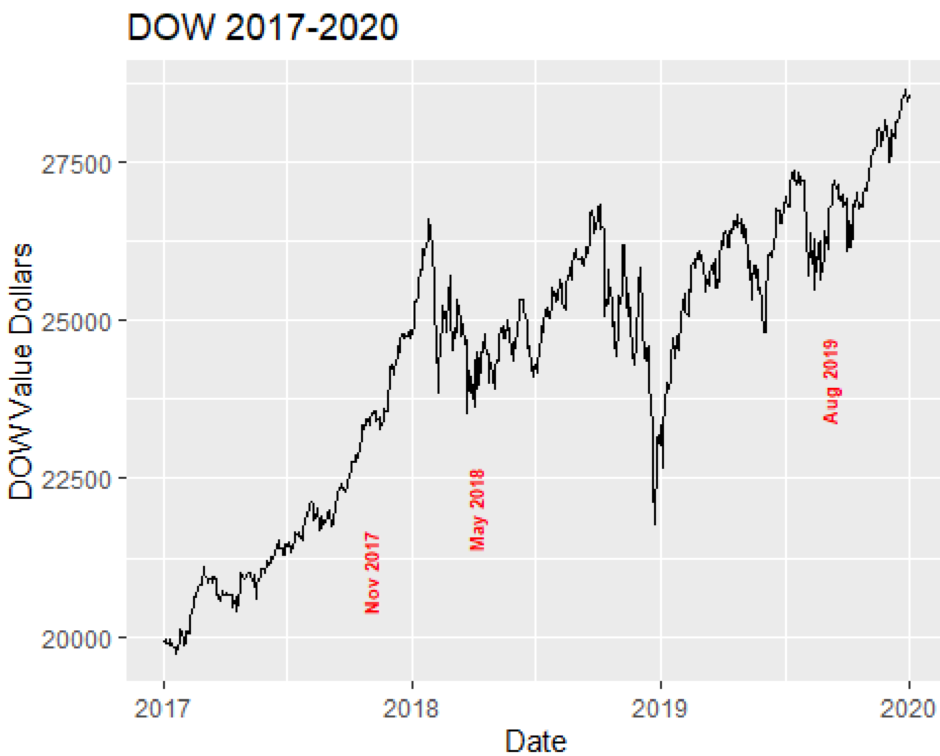
The chart above shows that AAP had a decreasing year in 2017 down to its minimum value in the fourth quarter of the year, then increased throughout 2018 until 2019. There was also another local minima in the third quarter of 2019 for AAP before it began increasing. Something could have happened in the first quarter of 2017 to cause it to decrease and another thing in 2018. We could check the internet for articles around that time. Lets look at the summary stats for this table.

```
summary(aap1)

##       Date                DOW_down_median        Month                  Year
##  Min.   :2017-01-03   Length:754         Length:754         Min.   :2017
##  1st Qu.:2017-10-02   Class :character   Class :character   1st Qu.:2017
##  Median :2018-07-02   Mode  :character   Mode  :character   Median :2018
##  Mean   :2018-07-02                                         Mean   :2018
##  3rd Qu.:2019-04-02                                         3rd Qu.:2019
##  Max.   :2019-12-31                                         Max.   :2019
##  UE_monthlyRate    stockValue       DOW_Close
##  Min.   :3.500   Min.   : 79.38   Min.   :19732
```

```
##   1st Qu.:3.700   1st Qu.:116.19   1st Qu.:22421
##   Median :3.900   Median :149.94   Median :24879
##   Mean   :3.965   Mean   :140.15   Mean   :24397
##   3rd Qu.:4.200   3rd Qu.:161.47   3rd Qu.:26061
##   Max.   :4.700   Max.   :184.72   Max.   :28645
```

The above summary statistics show a low unemployment rate for this period of time spanning three years from 2017 to 2020 with unemployment ranging from 3.5 to 4.7. The DOW had closing values ranging from 19732 USD to 28645 USD. Lets plot this date range for the DOW and see if they move together.

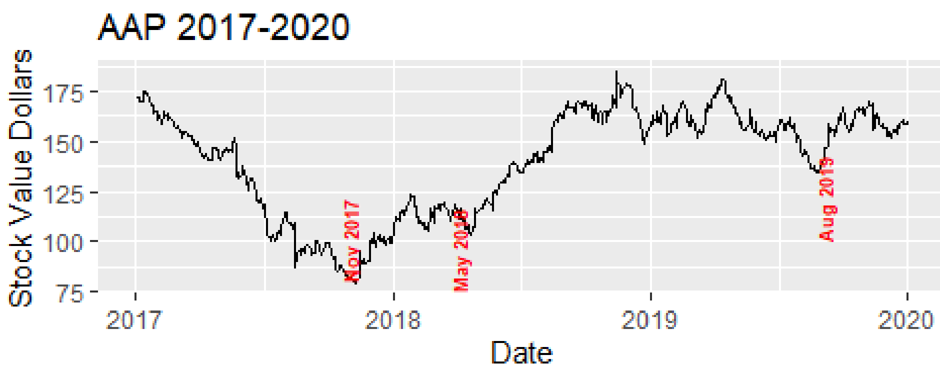
```
annotation2 <- data.frame(
   x = c(as.Date('2017-11-01'),as.Date('2018-04-04'),as.Date('2019-09-04')),
   y = c(21000,22000,24000),
   label = c("Nov 2017", "May 2018","Aug 2019"))

gg2 <- ggplot(data = aap1, aes(x=Date, y=DOW_Close)) +
       geom_line()+
       scale_y_continuous()+
       scale_fill_brewer(palette="paired") +
       theme(legend.position="bottom")+
       ggtitle('DOW 2017-2020')+
       ylab('DOW Value Dollars')+
       geom_text(data=annotation2, aes( x=x, y=y, label=label),
,
           color="red",
           size=2.5 , angle=90 , fontface="bold")

## Warning in pal_name(palette, type): Unknown palette paired

gg2
```

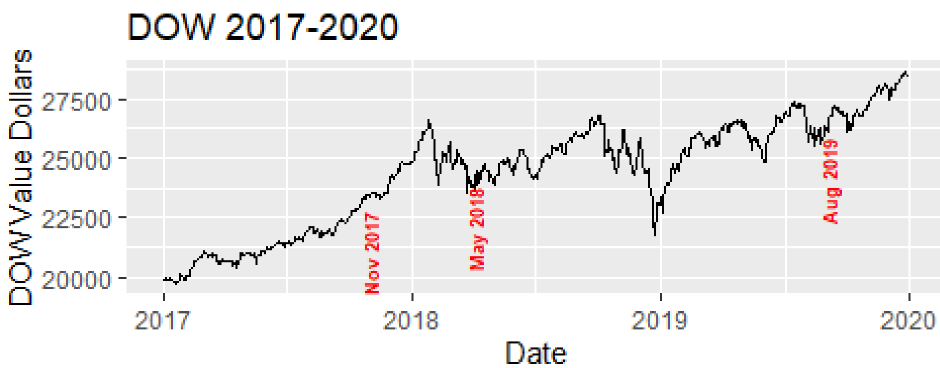


The above chart shows the increasing values of the DOW the same years as the AAP saw a decreasing year in 2017 to its local minimum in the fourth quarter of 2017, and then an increase until 2019 when it stabilized around the same value till 2020. But the DOW increased the time that AAP was decreasing and saw a local minima in the end of 2018 right when AAP reached a smaller local minima and also in the third quarter when AAP also had a local minima.

Lets look at these charts on top of each other.

```
grid.arrange(gg1, gg2, nrow = 2)
```

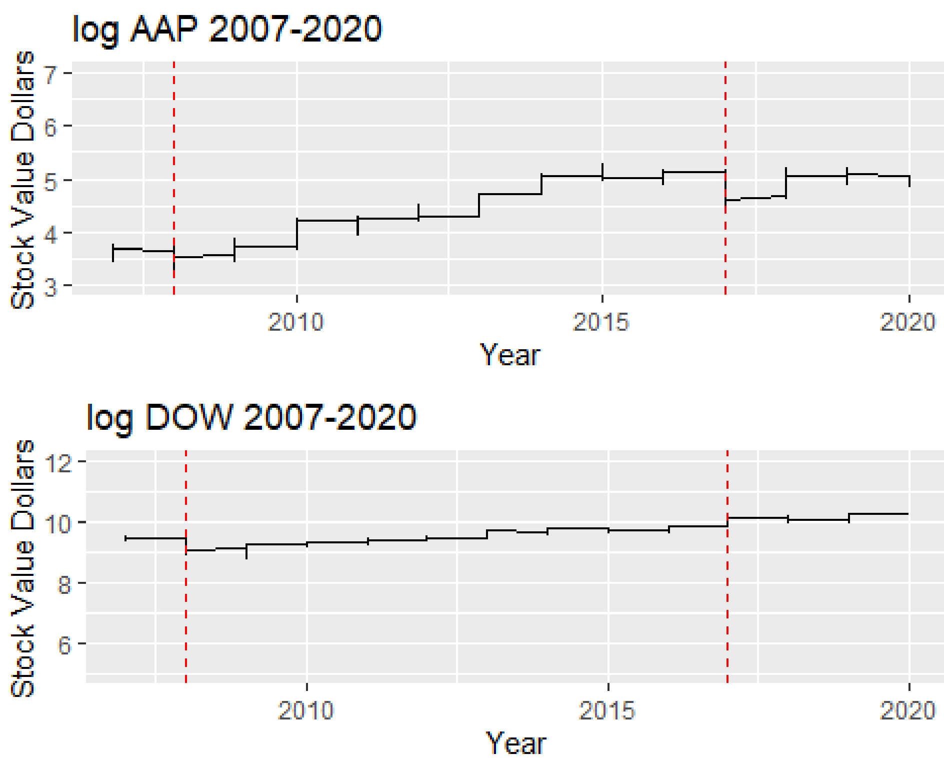




Lets see what happened in 2017 for AAP and in 2018 to cause it to decrease then increase respectively… just declining sales, plans to expand and open more stores, and public outcry on the sales declines of Advance Auto Parts. So, this could mean that because the DOW was doing great and increasing, investors thought to take their money out of the after market car parts stores, or maybe they thought they were hurting because of Amazon Prime taking their business. Those are some possibilities. When looking at stocks in the DOW Jones industrial average, AAP isn't listed as one of these stocks, so it could be that people took their money out of the after markets car parts stock as it was declining and put it into any of the stocks that belong to the DOW, because it increased in value in 2017 while AAP decreased, then they both moved together around and after 2018.

```
aaplog <- aap
aaplog$logAAP <- log1p(aaplog$stockValue)
aaplog$logDOW <- log1p(aaplog$DOW_Close)

g1 <- ggplot(data = aaplog, aes(x=Year, y=logAAP)) +
      geom_line()+
      scale_y_continuous(limits=c(3,7))+
      scale_fill_brewer(palette="paired") +
      theme(legend.position="bottom")+
      geom_vline(xintercept=c(2008,2017), linetype='dashed', color='red')+
      ggtitle('log AAP 2007-2020')+
      ylab('Stock Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired
```

```
g2 <- ggplot(data = aaplog, aes(x=Year, y=logDOW)) +
      geom_line()+
      scale_y_continuous(limits=c(5,12))+
      scale_fill_brewer(palette="paired") +
      theme(legend.position="bottom")+
      geom_vline(xintercept=c(2008,2017), linetype='dashed', color='red')+
      ggtitle('log DOW 2007-2020')+
      ylab('Stock Value Dollars')

## Warning in pal_name(palette, type): Unknown palette paired

grid.arrange(g1, g2, nrow = 2)
```



The above chart shows the log scale of the value + 1 so that there aren't any natural log errors in scaling. There is also an added few lines to show the years of 2008 and 2017 when AAP had decreasing years in stock value.

Lets look at the DOW over the years with the other 16 stocks in this portfolio of stocks that proved most profitable when the DOW was down and unemployment was up using the set of stocks with positive median values under those constraints.

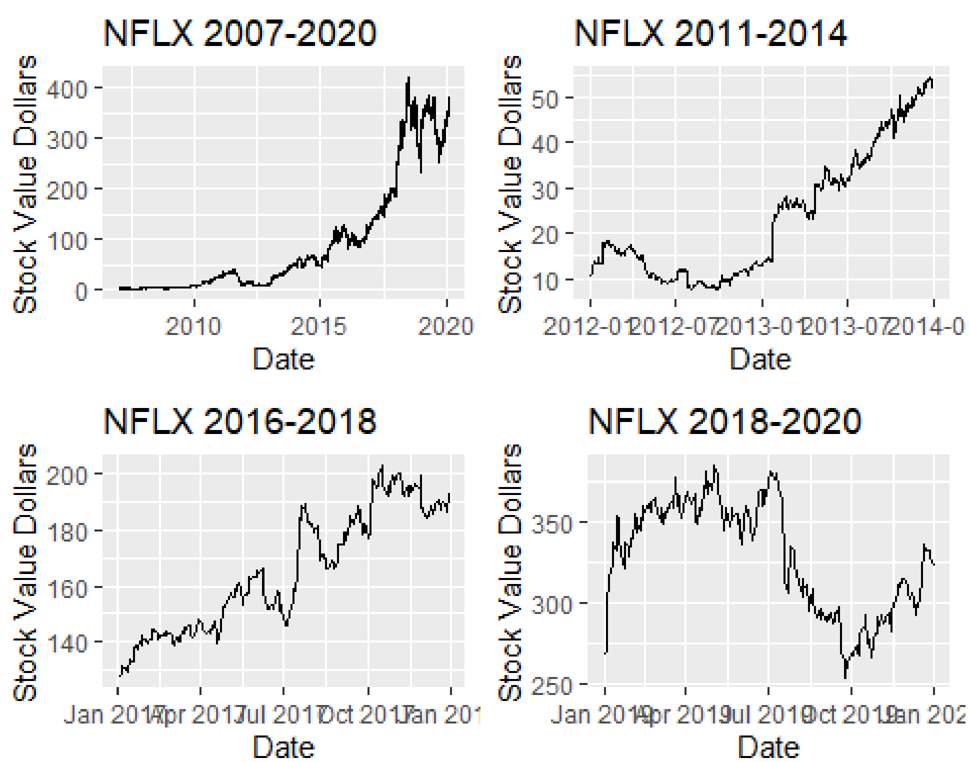
```
dROI17 <- Close17_dow %>% group_by(DOW_down_median) %>%
  summarise_at(vars(stockValue), mean)
colnames(dROI17)[2] <- 'avgStockValue'
```

```r
start17 <- subset(Close17_dow, Close17_dow$Date=='2007-01-03')
final17 <- subset(Close17_dow, Close17_dow$Date=='2020-02-14')

start17 <- start17[order(start17$DOW_down_median),]
final17 <- final17[order(final17$DOW_down_median),]
dROI17 <- dROI17[order(dROI17$DOW_down_median),]

DOW_ROI <- as.data.frame(final17$DOW_Close/start17$DOW_Close)
colnames(DOW_ROI) <- 'DOW_ROI'

colnames(start17)[6] <- 'startValue'
colnames(final17)[6] <- 'finalValue'

dROI17$startValue <- start17$startValue
dROI17$finalValue <- final17$finalValue
dROI17$DOW_ROI <- DOW_ROI$DOW_ROI
dROI17$stock_ROI <- dROI17$finalValue/dROI17$startValue
dROI17

## # A tibble: 17 x 6
##    DOW_down_median avgStockValue startValue finalValue DOW_ROI stock_ROI
##    <chr>                   <dbl>      <dbl>      <dbl>   <dbl>     <dbl>
##  1 AAP                      97.3      35.6       134.     2.36     3.75
##  2 ADDYY                    58.4      25         156.     2.36     6.24
##  3 AMZN                    551.       38.7      2135.     2.36    55.2
##  4 COST                    123.       52.8       318.     2.36     6.02
##  5 DLTR                     52.8      10.2        88.7    2.36     8.67
##  6 FFIN                     14.7       7.00       34.6    2.36     4.94
##  7 GOOG                    559.      233.       1521.     2.36     6.53
##  8 HD                       90.9      41.1       245.     2.36     5.97
##  9 JNJ                      91.1      66.4       150.     2.36     2.26
## 10 NFLX                     92.4       3.80      380.     2.36   100.
## 11 NKE                      39.7      12.2       104.     2.36     8.48
## 12 PCG                      44.1      47.3        16.2    2.36     0.343
## 13 ROST                     40.8       7.62      122.     2.36    16.0
## 14 TEVA                     41.1      31.3        12.2    2.36     0.391
## 15 TJX                      26.0       7.17       63.4    2.36     8.84
## 16 WMT                      69.8      47.5       118.     2.36     2.48
## 17 XOM                      81.4      74.1        60.7    2.36     0.818
```

Netflix killed the return on investment with more than 100 fold profits. Lets look at Netflix to see the highs and lows of this stock since 2007 and through till 2020.

```r
nflx <- subset(Close17_dow, Close17_dow$DOW_down_median=='NFLX')
nflx1 <- subset(nflx, nflx$Year > 2011 & nflx$Year < 2014)
nflx2 <- subset(nflx, nflx$Year > 2016 & nflx$Year < 2018)
nflx3 <- subset(nflx, nflx$Year > 2018 & nflx$Year < 2020)

gg3 <- ggplot(data = nflx, aes(x=Date, y=stockValue)) +
       geom_line()+
```

```
        scale_y_continuous()+
        theme(legend.position="bottom")+
        ggtitle('NFLX 2007-2020')+
        ylab('Stock Value Dollars')
gg4 <- ggplot(data = nflx1, aes(x=Date, y=stockValue)) +
        geom_line()+
        scale_y_continuous()+
        theme(legend.position="bottom")+
        ggtitle('NFLX 2011-2014')+
        ylab('Stock Value Dollars')
gg5 <- ggplot(data = nflx2, aes(x=Date, y=stockValue)) +
        geom_line()+
        scale_y_continuous()+
        theme(legend.position="bottom")+
        ggtitle('NFLX 2016-2018')+
        ylab('Stock Value Dollars')
gg6 <- ggplot(data = nflx3, aes(x=Date, y=stockValue)) +
        geom_line()+
        scale_y_continuous()+
        theme(legend.position="bottom")+
        ggtitle('NFLX 2018-2020')+
        ylab('Stock Value Dollars')

grid.arrange(gg3, gg4, gg5,gg6, nrow = 2)
```



Netflix was certainly a great stock to invest in at around 3 USD in 2007 and at around 325 USD in 2020. We see that Netflix was on the up and up almost its entire course growing to more than 100 times its initial starting value in 2007. There were some lows such as in 2012 there was small dip in the curve, then in July and August-September 2017, and also in the third quarter of 2019. But it still performed amazingly. The stock dreams of riches are made of and conartists use to get more money from people on risky start up penny stocks. But lets put out all we know about Netflix.

- Netflix was first heard from the author of this tutorial in 2003 when some roommate of a guy the author dated bragged about how awesome Netflix is to cost $7/month and you can rent new movies mailed to your home for no additional charge. This roommate also bought a flat screen tv for 7000 USD before they became ubiquitously priced from 200-500 USD five years later.
- The minimum wage for workers around CA in this time period was also about the cost of the Netflix monthly membership. Many tv shows started being options to rent from sources such as premium cable tv shows like Dexter around 2007 or so.
- I pulled the cord on cable due to high costs and got a Netflix membership for around 8-10 USD in about 2014. Which was also around the price of minimum wage at that time.
- cell phones became very great and needed personal items with fast wifi and internet streamings still at a cost that beat cable tv and home phone lines at this same time period.
- Netflix got more innovative, they started adding more Netflix produced shows in 2016 that made memes on instagram and facebook, the top social media platforms of the time in the 2010s like with 'Orange is the New Black' (I couldn't watch, but saw many memes on).

So, given what you know and what you scanned above, isn't it no surprise that a stock that out competes alternative forms of entertainment, is low cost in price and able to be taken mobile or use anywhere and at any time for next to nothing in cost as an hour of a consumer's 160-200 hour work month if working minimum wage and full time. Maybe its time to add another feature, like federal minimum wage rates to this data. We will in fact do this later, but for now we will add another feature that compares the ROI ration to that of the DOW Jones industrial average, and try to pull the most striking features out of that group.

```
dROI17$stockBeatsDOW <- ifelse(dROI17$stock_ROI > dROI17$DOW_ROI, 'Yes',
'No')
dROI17[,c(1,7)]
```

```
## # A tibble: 17 x 2
##    DOW_down_median stockBeatsDOW
##    <chr>           <chr>
##  1 AAP             Yes
##  2 ADDYY           Yes
##  3 AMZN            Yes
```

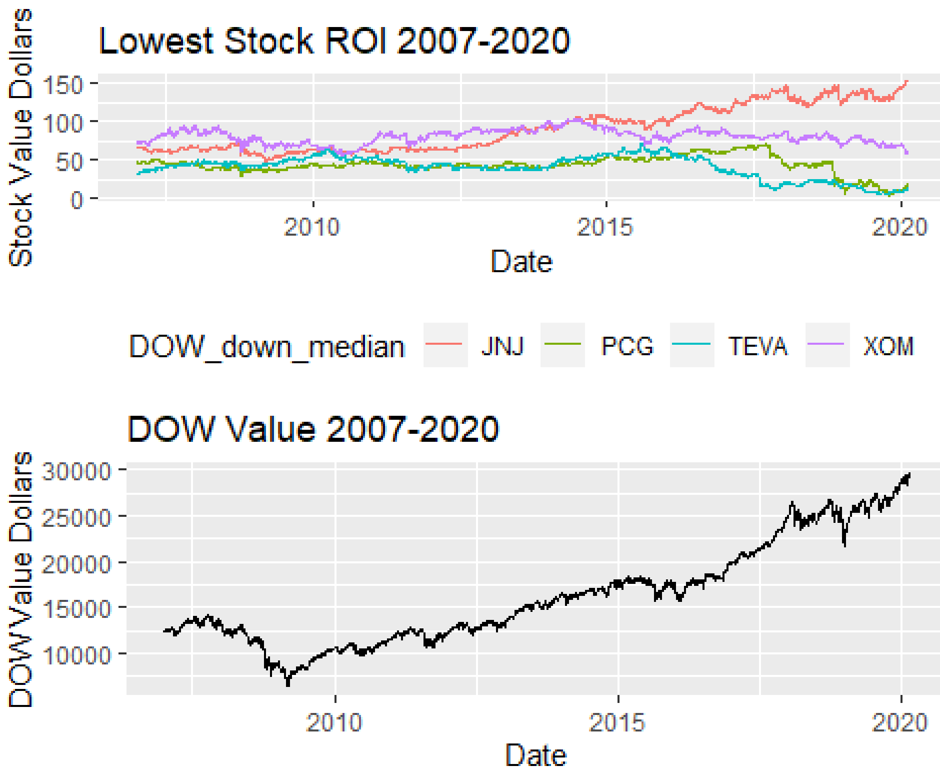
```
##   4 COST             Yes
##   5 DLTR             Yes
##   6 FFIN             Yes
##   7 GOOG             Yes
##   8 HD               Yes
##   9 JNJ              No
## 10 NFLX             Yes
## 11 NKE              Yes
## 12 PCG              No
## 13 ROST             Yes
## 14 TEVA             No
## 15 TJX              Yes
## 16 WMT              Yes
## 17 XOM              No
```

Looking at the above chart of the stocks that beat the DOW in return on investment ratios of final stock value to initial stock value (2007-2020), Exxon Mobil, Johnson & Johnson, Teva Pharmaceuticals Industries, and Pacific Gas and Electric lossed money or had lower returns than that of the DOW. Of note, the pharmaceuticals might make you go on a wild tangent to know which company is supplying us with flu vaccines annually. If so, Sanofi (SNY) is the US's largest supplier and they aren't in this analysis. But it would be interesting to see when they make the most money, considering we have a CoV-19 flu contagion globally as of Feb. 2020. While, it is safe to say the companies that low income consumers love or live by did well, such as: Walmart, Adidas, Nike, Home Depot, Netflix, Amazon, Advance Auto Parts, TJ Maxx, Ross, Dollar Tree, and Costco. Costco is more of a middle class or small business store because you have to have cash, or used to have cash or money in your checking account to buy their goods and services with your atm card. They may have changed this. First Financial bank also did well, and it was selected because it was one of the banks available when hand picking these stocks. I don't use it and don't know anyone who does, and that is because I am on the West Coast, and this bank originates out of the East Coast. This could be an indicator that the East Coast is picking up in business and getting more home loans, business loans, etc. Than the more West Coast known banks like Citi, Chase, Bank of America, JP Morgan Chase. As these other banks did not perform well for median positive daily changes in stock prices during increasing unemployment and decreasing DOW values.

Lets plot those four stocks in this portfolio that did worse than the DOW.

```
low4 <- subset(Close17_dow, Close17_dow$DOW_down_median == 'JNJ'|
                  Close17_dow$DOW_down_median == 'PCG'|
                  Close17_dow$DOW_down_median == 'TEVA'|
                  Close17_dow$DOW_down_median == 'XOM')
gg7 <- ggplot(data = low4, aes(x=Date, y=stockValue, group=DOW_down_median))
+
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
       scale_fill_brewer(palette="Spectral") +

       theme(legend.position="bottom")+
```

```r
       ggtitle('Lowest Stock ROI 2007-2020')+
       ylab('Stock Value Dollars')
gg8 <- ggplot(data = low4, aes(x=Date, y=DOW_Close)) +
       geom_line()+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('DOW Value 2007-2020')+
       ylab('DOW Value Dollars')

grid.arrange(gg7,gg8, nrow = 2)
```

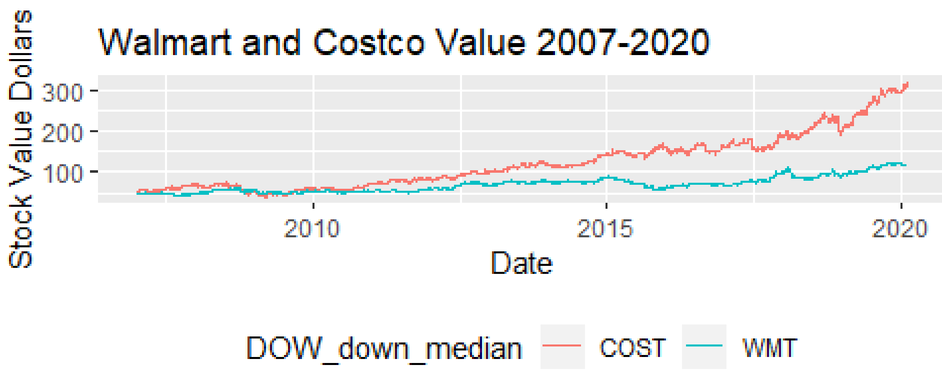


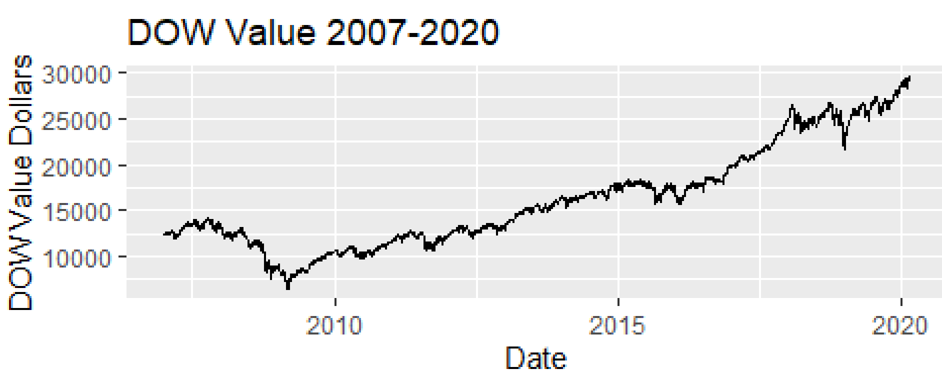
From the above charts, It looks like Johnson & Johnson has started to move upward with the DOW starting around 2013. The TEVA stock seems to be negatively correlated with the DOW and Exxon Mobil was positively correlated with the DOW from 2007 to about 2015, then started moving in the opposite direction after 2015. Exxon supplies fuels to automobiles, while PCG supplies electricity to hybrid and electric vehicles in certain US regions. Yet, both started moving opposite directions with the DOW after 2015.

Lets now compare Costco and Walmart to each other and the DOW.

```r
wal_Cost <- subset(Close17_dow, Close17_dow$DOW_down_median=='WMT' |
                        Close17_dow$DOW_down_median=='COST')

gg9 <- ggplot(data = wal_Cost, aes(x=Date, y=stockValue,
group=DOW_down_median)) +
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
```

```
       theme(legend.position="bottom")+
       ggtitle('Walmart and Costco Value 2007-2020')+
       ylab('Stock Value Dollars')

grid.arrange(gg9,gg8, nrow = 2)
```



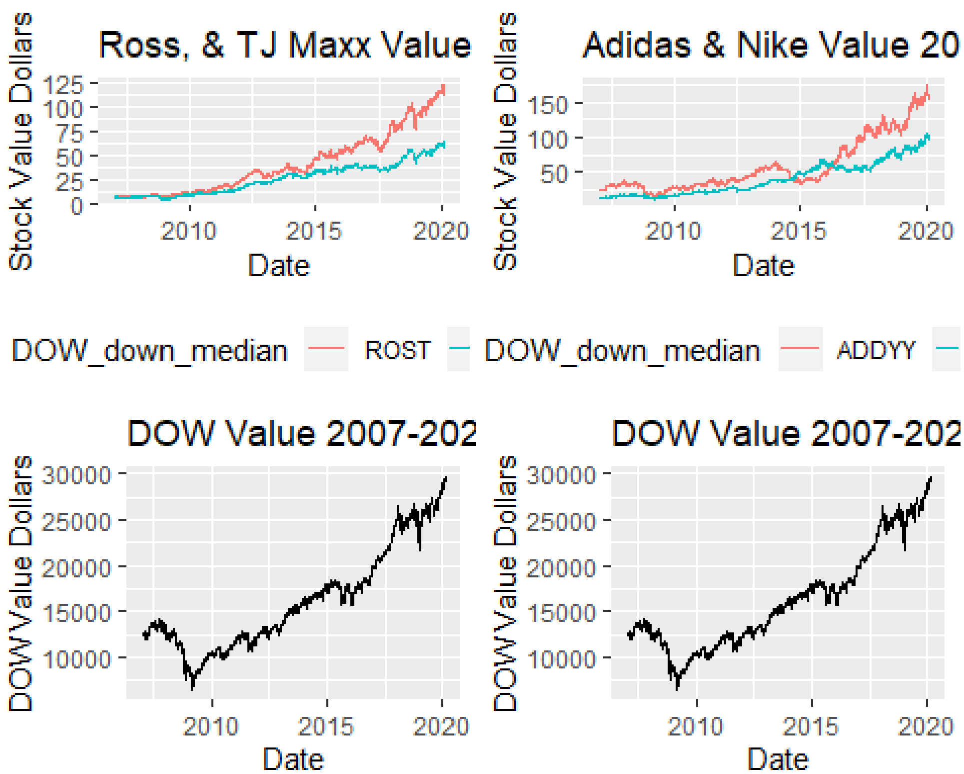


It is interesting to note that Costco and the DOW seem to be identical curves for the direction they move while Walmart seems also be increasing when the DOW does but at a much lower rate over time.

Lets now compare TJ Maxx, Ross, Nike, and Adidas to the DOW.

```
Retail <- subset(Close17_dow, Close17_dow$DOW_down_median == 'TJX' |
                   Close17_dow$DOW_down_median == 'ROST' )

Shoes <- subset(Close17_dow, Close17_dow$DOW_down_median == 'NKE' |
                   Close17_dow$DOW_down_median == 'ADDYY' )


gg10 <- ggplot(data = Retail, aes(x=Date, y=stockValue,
group=DOW_down_median)) +
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('Ross, & TJ Maxx Value 2007-2020')+
       ylab('Stock Value Dollars')
```

```
gg11 <- ggplot(data = Shoes, aes(x=Date, y=stockValue,
group=DOW_down_median)) +
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('Adidas & Nike Value 2007-2020')+
       ylab('Stock Value Dollars')

grid.arrange(gg10, gg11, gg8, gg8, nrow =2)
```

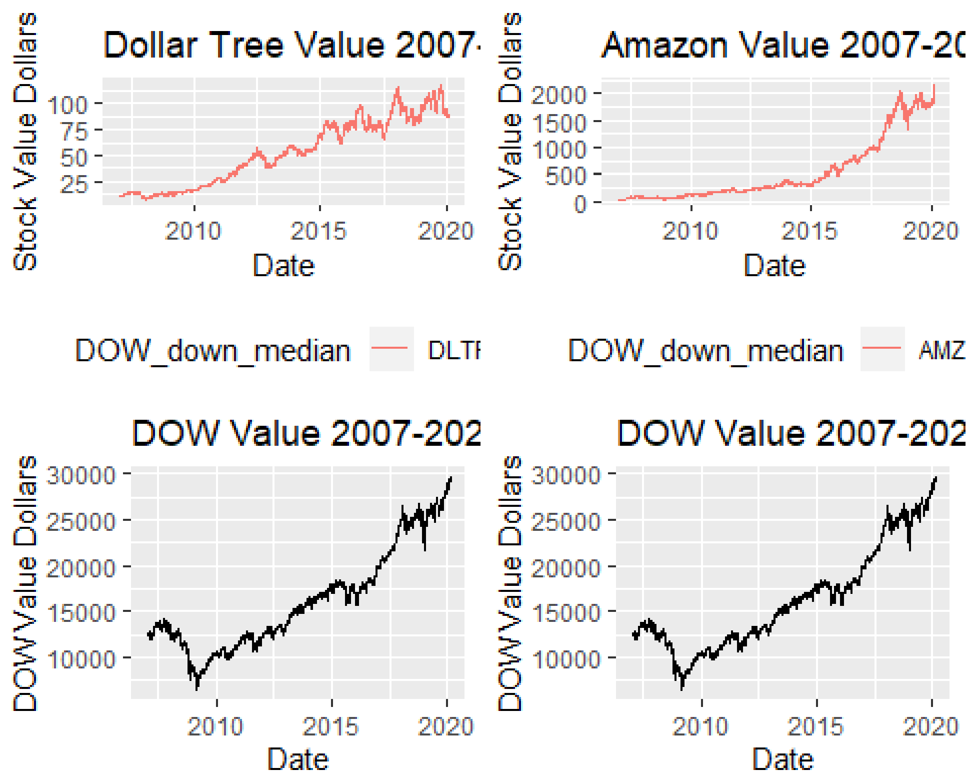


The DOW is plotted below each of the two plots of either Ross & TJ Maxx, or Adidas & Nike. Both Ross and Adidas did better than TJ Maxx and Nike. Except, that Nike did do better than Adidas between 2015 and 2016.

Now, lets look at Dollar Tree and Amazon compared to each other and the DOW in this time span.

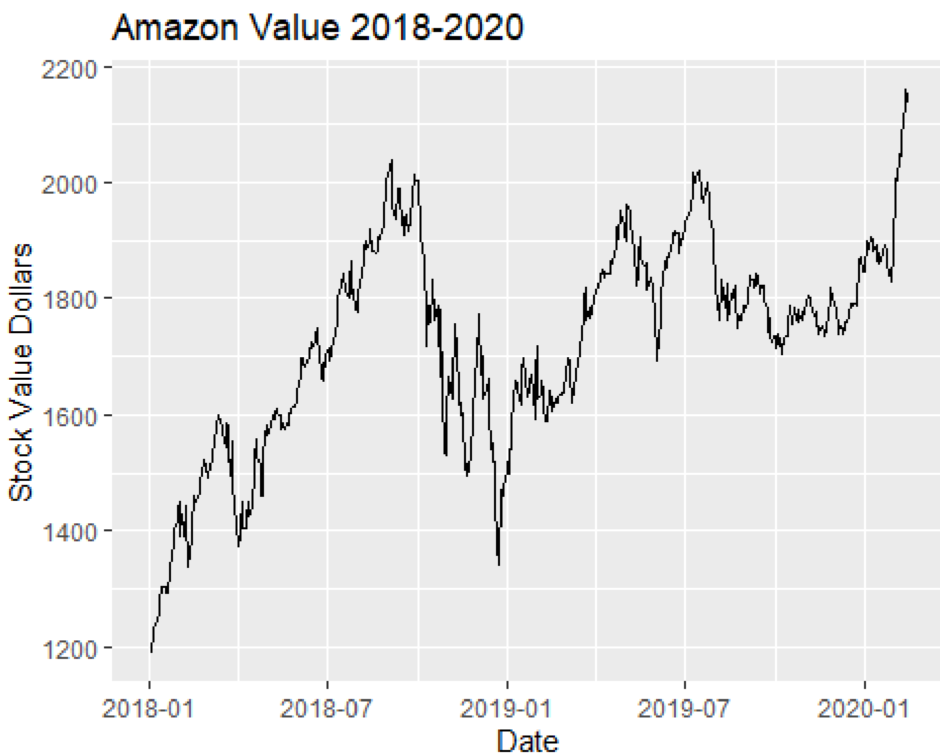
```
DLRv <- subset(Close17_dow, Close17_dow$DOW_down_median == 'DLTR' )
AMZNv <- subset(Close17_dow, Close17_dow$DOW_down_median == 'AMZN' )


gg12 <- ggplot(data = DLRv, aes(x=Date, y=stockValue, group=DOW_down_median))
+
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('Dollar Tree Value 2007-2020')+
```

```
       ylab('Stock Value Dollars')

gg13 <- ggplot(data = AMZNv, aes(x=Date, y=stockValue,
group=DOW_down_median)) +
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('Amazon Value 2007-2020')+
       ylab('Stock Value Dollars')

grid.arrange(gg12, gg13, gg8, gg8, nrow =2)
```



The above charts show Dollar Tree compared to Amazon and both compared to the DOW between 2007 and 2020. The Dollar Tree seems to be cyclical but overall increasing, while Amazon was a steady increase over the years except in 2018 where it had a decrease.

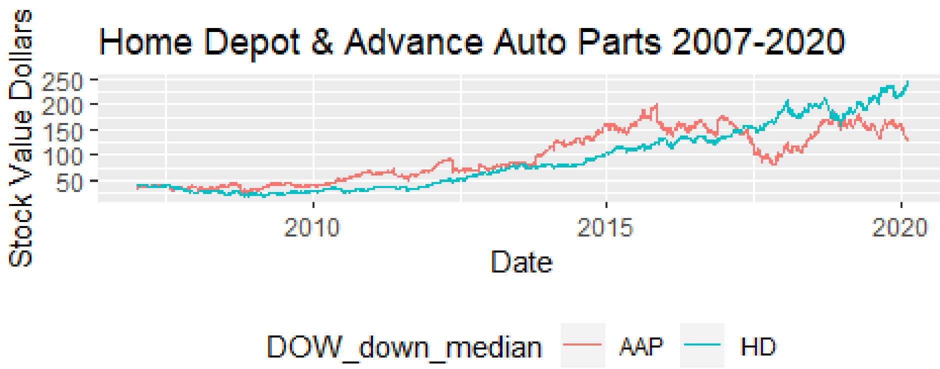
```
AMZNv2 <- subset(Close17_dow, Close17_dow$Year > 2017 &
                   Close17_dow$DOW_down_median == 'AMZN')

gg14 <- ggplot(data = AMZNv2, aes(x=Date, y=stockValue)) +
       geom_line()+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('Amazon Value 2018-2020')+
       ylab('Stock Value Dollars')
gg14
```

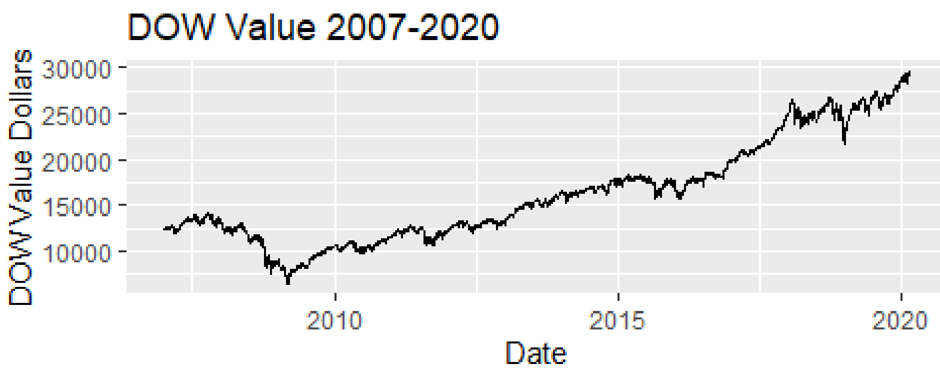


We see a drop in value of Amazon after September 2018 until about January 2019 when it increases until just before the end of the 2nd quarter in 2019 then it drops in the 1st quarter staying low before increasing to a global maximum in January 2020.

Lets look at Home Depot and Advance Auto parts now.

```
HDv <- subset(Close17_dow, Close17_dow$DOW_down_median=='HD' |
                Close17_dow$DOW_down_median=='AAP')

gg15 <- ggplot(data = HDv, aes(x=Date, y=stockValue, group=DOW_down_median))
+
       geom_line(aes(color=DOW_down_median))+
       scale_y_continuous()+
       theme(legend.position="bottom")+
       ggtitle('Home Depot & Advance Auto Parts 2007-2020')+
       ylab('Stock Value Dollars')

grid.arrange(gg15, gg8, nrow =2)
```



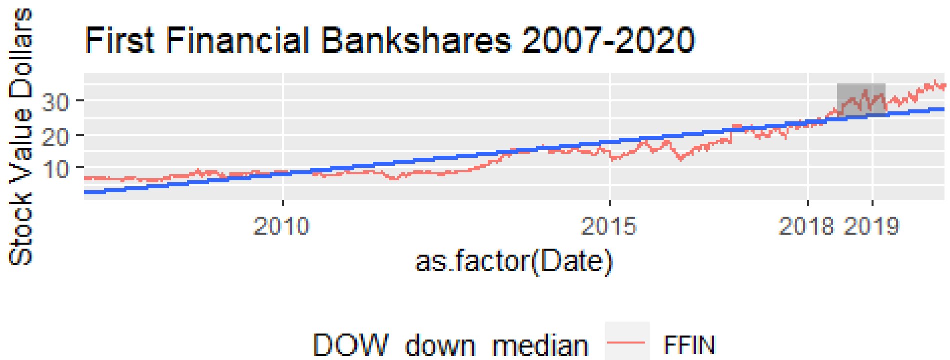


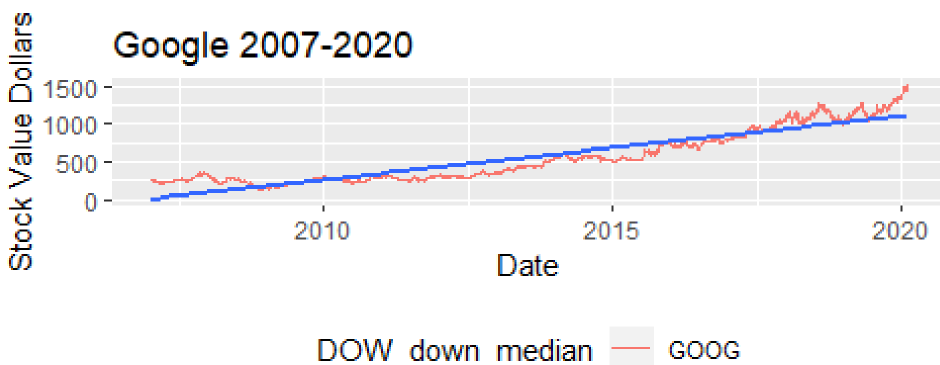
The charts above show that Home Depot and the DOW move together by increasing the time span from 2007-2020. Advance Auto Parts showed it had been increasing from 2007-2016, but then declined to a local minimum in the middle of 2017 where it then increased and remained a steady value up to 2020.

Lets also look at Google and First Financial Bankshares.

```
ffin <- subset(Close17_dow, Close17_dow$DOW_down_median=='FFIN')
goog <- subset(Close17_dow, Close17_dow$DOW_down_median=='GOOG')

gg16 <- ggplot(data=ffin, aes(x=as.factor(Date), y=stockValue,
group=DOW_down_median))+
        geom_line(aes(color=DOW_down_median))+
        geom_smooth(method = "lm")+
        annotate("rect", xmin = "2018-07-01", xmax = "2019-03-31", ymin = 25,
ymax = 35,
        alpha = .4)+
        scale_x_discrete(breaks=c("2010-01-04","2015-01-02","2018-01-05",
                                  "2019-01-03"),
                         labels=c("2010","2015", "2018","2019"))+
        scale_y_continuous()+
        theme(legend.position="bottom")+
        ggtitle('First Financial Bankshares 2007-2020')+
        ylab('Stock Value Dollars')

gg17 <- ggplot(data = goog, aes(x=Date, y=stockValue, group=DOW_down_median))
```

```
+
      geom_line(aes(color=DOW_down_median))+
      geom_smooth(method = "lm")+
      scale_y_continuous()+
      theme(legend.position="bottom")+
      ggtitle('Google 2007-2020')+
      ylab('Stock Value Dollars')

grid.arrange(gg16, gg17,nrow =2)
```



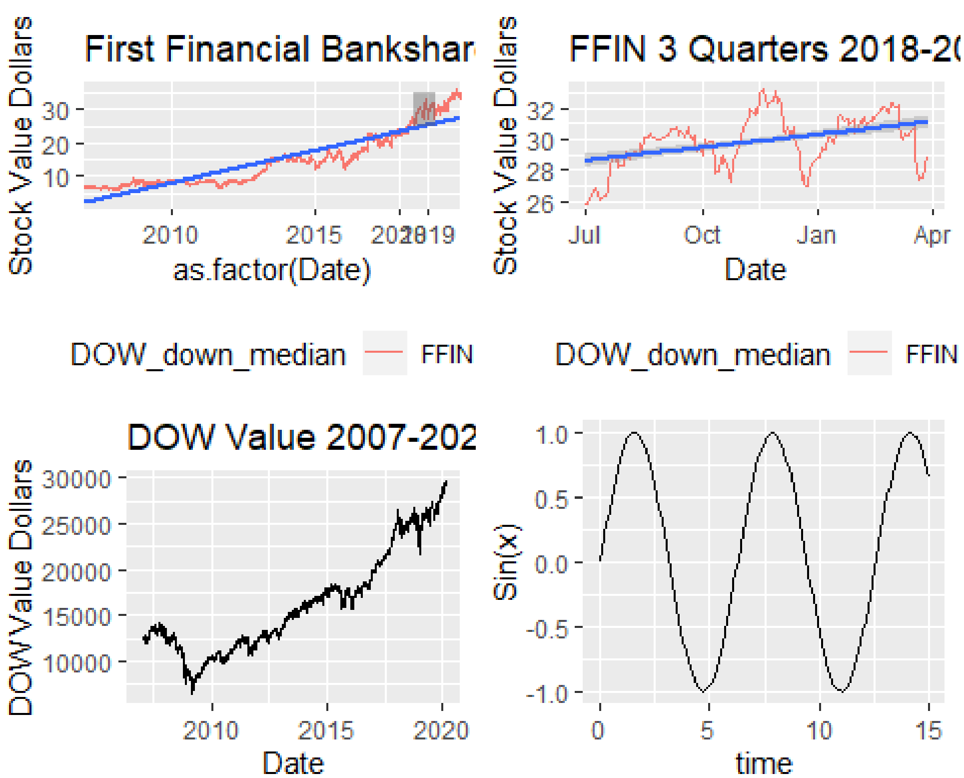


The above chart shows Google and First Financial Bankshares from 2007-2020. First Financial has many highs and lows between 2015-2020 but is overall increasing, while Google is steadily increasing up till 2017 when it has a few highs and lows until 2020. Both have increased overall as indicated by the linear trendlines added to the two linear plots above for Google and First Financial Bankshares.

We know that some of these plots have time periods that are cyclical, but so far we know that overall, if we want to earn the most on our portfolios we would like to buy low and sell high. These stocks are increasing to rates higher than when they start a cyclical pattern. Lets examine the First Financial Bankshares stock when it sees these cyclical patterns further for some strategy development. There are three peaks or highs for FFIN after about 2017, so lets plot this.

```
ffin2 <- subset(ffin, ffin$Year==2018 & ffin$Month=='Jul'|
                      ffin$Year==2018 & ffin$Month=='Aug'|
                      ffin$Year==2018 & ffin$Month=='Sep'|
```

```
                     ffin$Year==2018 & ffin$Month=='Oct'|
                     ffin$Year==2018 & ffin$Month=='Nov'|
                     ffin$Year==2018 & ffin$Month=='Dec'|
                     ffin$Year==2019 & ffin$Month=='Jan'|
                     ffin$Year==2019 & ffin$Month=='Feb'|
                     ffin$Year==2019 & ffin$Month=='Mar')



t <- seq(0,15,0.1)
y <- sin(t)
ty <- qplot(t,y, geom='path', xlab='time', ylab='Sin(x)')

gg18 <- ggplot(data=ffin2, aes(x=Date, y=stockValue, group=DOW_down_median))+
        geom_line(aes(color=DOW_down_median))+
        geom_smooth(method = "lm")+
        scale_y_continuous()+
        theme(legend.position="bottom")+
        ggtitle('FFIN 3 Quarters 2018-2019')+
        ylab('Stock Value Dollars')

grid.arrange(gg16,gg18,gg8, ty,nrow=2)
```



The above chart shows the sine curve to illustrate the highs and lows of the daily value changes for the First Financial Bankshares linear plot. If the curve was distorted at certain time intervals startind in the 3rd quarter of 2018 and ending before the 2nd quarter of

2019, this sine curve would show the three waves this stock saw. We are interested in buying at the beginning of the 3rd quarter when the price is low, then selling before the start of the 4th quarter when the stock is high, then buying again when it sees a dip in price after about the 1st month of the 4th quarter, then selling in the middle of the 4th quarter when high, buying at the end of the 4th quarter when low, and selling when high in the middle of the 1st quarter of 2019, and buying when low at the end of the 1st quarter, to maximize profits.

These quick analysis of the stock that did well when the DOW was decreasing and unemployment was increasing was interesting to look at. But now, lets look at these stocks and find out if we can indicate when a stock is good by setting a threshold for the number of minimums the stock has and if it decreases by more than its value in the last quarter and also the last two quarters then compare this to the local maxima where it increases to more than a set threshold than it was valued at in the last quarter and also the last two quarters. Count the number of times this happens and compare to the data on the stock ROI. This will require adding in more columns to calculate the daily change compared to a median value of the stock in each quarter. We have the years, the months, and the stock values in the table we are currently using.

We can do this by using time lags with the dplyr package. We should already have this package loaded. We have been using for median and mean calculations when grouping by stock. The data set we should use could be this one, Close17_dow, or we could spread out the stock names back to being columns and add the stock lags for each stock for 7 days, 30 days, 90 days, and 180 days. Don't get too excited, we have 53 stocks to do this for, or we could just do it with these 17 stocks that made the most profit from having positive median values when unemployment was increasing and the DOW was decreasing. I vote we do the 17. ***

We are going to create time lags of 7,30,60,90, 120, 150, and 180 days to see if there are any rolling stock values that could indicate when to buy or sell and pin point are threshold values and possibly work this into an automated program later with continuous uploads of these stocks in monitoring our portfolio of stock.

```
Close17_dow <- Close17_dow[with(Close17_dow, order(DOW_down_median, Date)),]

Laap <- subset(Close17_dow, Close17_dow$DOW_down_median=='AAP')
Laddyy <- subset(Close17_dow, Close17_dow$DOW_down_median=='ADDYY')
Lamzn <- subset(Close17_dow, Close17_dow$DOW_down_median=='AMZN')
Lcost <- subset(Close17_dow, Close17_dow$DOW_down_median=='COST')
Ldltr <- subset(Close17_dow, Close17_dow$DOW_down_median=='DLTR')
Lffin <- subset(Close17_dow, Close17_dow$DOW_down_median=='FFIN')
Lgoog <- subset(Close17_dow, Close17_dow$DOW_down_median=='GOOG')
Lhd <- subset(Close17_dow, Close17_dow$DOW_down_median=='HD')
Ljnj <- subset(Close17_dow, Close17_dow$DOW_down_median=='JNJ')
Lnflx <- subset(Close17_dow, Close17_dow$DOW_down_median=='NFLX')
Lnke <- subset(Close17_dow, Close17_dow$DOW_down_median=='NKE')
```

```
Lpcg <- subset(Close17_dow, Close17_dow$DOW_down_median=='PCG')
Lrost <- subset(Close17_dow, Close17_dow$DOW_down_median=='ROST')
Lteva <- subset(Close17_dow, Close17_dow$DOW_down_median=='TEVA')
Ltjx <- subset(Close17_dow, Close17_dow$DOW_down_median=='TJX')
Lwmt <- subset(Close17_dow, Close17_dow$DOW_down_median=='WMT')
Lxom <- subset(Close17_dow, Close17_dow$DOW_down_median=='XOM')

aapL7 <- lag(Laap$stockValue,7)
addyyL7 <- lag(Laddyy$stockValue,7)
amznL7 <- lag(Lamzn$stockValue,7)
costL7 <- lag(Lcost$stockValue,7)
dltrL7 <- lag(Ldltr$stockValue,7)
ffinL7 <- lag(Lffin$stockValue,7)
googL7 <- lag(Lgoog$stockValue,7)
hdL7 <- lag(Lhd$stockValue,7)
jnjL7 <- lag(Ljnj$stockValue,7)
nflxL7 <- lag(Lnflx$stockValue,7)
nkeL7 <- lag(Lnke$stockValue,7)
pcgL7 <- lag(Lpcg$stockValue,7)
rostL7 <- lag(Lrost$stockValue,7)
tevaL7 <- lag(Lteva$stockValue,7)
tjxL7 <- lag(Ltjx$stockValue,7)
wmtL7 <- lag(Lwmt$stockValue,7)
xomL7 <- lag(Lxom$stockValue,7)

Close17_dow$lag7 <- c(aapL7,addyyL7,amznL7,costL7,dltrL7,ffinL7,googL7,

hdL7,jnjL7,nflxL7,nkeL7,pcgL7,rostL7,tevaL7,
                                  tjxL7,wmtL7,xomL7)

aapL30 <- lag(Laap$stockValue,30)
addyyL30 <- lag(Laddyy$stockValue,30)
amznL30 <- lag(Lamzn$stockValue,30)
costL30 <- lag(Lcost$stockValue,30)
dltrL30 <- lag(Ldltr$stockValue,30)
ffinL30 <- lag(Lffin$stockValue,30)
googL30 <- lag(Lgoog$stockValue,30)
hdL30 <- lag(Lhd$stockValue,30)
jnjL30 <- lag(Ljnj$stockValue,30)
nflxL30 <- lag(Lnflx$stockValue,30)
nkeL30 <- lag(Lnke$stockValue,30)
pcgL30 <- lag(Lpcg$stockValue,30)
rostL30 <- lag(Lrost$stockValue,30)
tevaL30 <- lag(Lteva$stockValue,30)
tjxL30 <- lag(Ltjx$stockValue,30)
wmtL30 <- lag(Lwmt$stockValue,30)
xomL30 <- lag(Lxom$stockValue,30)

Close17_dow$lag30 <-
```

```
c(aapL30,addyyL30,amznL30,costL30,dltrL30,ffinL30,googL30,

hdL30,jnjL30,nflxL30,nkeL30,pcgL30,rostL30,tevaL30,
                                  tjxL30,wmtL30,xomL30)


aapL60 <- lag(Laap$stockValue,60)
addyyL60 <- lag(Laddyy$stockValue,60)
amznL60 <- lag(Lamzn$stockValue,60)
costL60 <- lag(Lcost$stockValue,60)
dltrL60 <- lag(Ldltr$stockValue,60)
ffinL60 <- lag(Lffin$stockValue,60)
googL60 <- lag(Lgoog$stockValue,60)
hdL60 <- lag(Lhd$stockValue,60)
jnjL60 <- lag(Ljnj$stockValue,60)
nflxL60 <- lag(Lnflx$stockValue,60)
nkeL60 <- lag(Lnke$stockValue,60)
pcgL60 <- lag(Lpcg$stockValue,60)
rostL60 <- lag(Lrost$stockValue,60)
tevaL60 <- lag(Lteva$stockValue,60)
tjxL60 <- lag(Ltjx$stockValue,60)
wmtL60 <- lag(Lwmt$stockValue,60)
xomL60 <- lag(Lxom$stockValue,60)

Close17_dow$lag60 <-
c(aapL60,addyyL60,amznL60,costL60,dltrL60,ffinL60,googL60,

hdL60,jnjL60,nflxL60,nkeL60,pcgL60,rostL60,tevaL60,
                                  tjxL60,wmtL60,xomL60)

aapL90 <- lag(Laap$stockValue,90)
addyyL90 <- lag(Laddyy$stockValue,90)
amznL90 <- lag(Lamzn$stockValue,90)
costL90 <- lag(Lcost$stockValue,90)
dltrL90 <- lag(Ldltr$stockValue,90)
ffinL90 <- lag(Lffin$stockValue,90)
googL90 <- lag(Lgoog$stockValue,90)
hdL90 <- lag(Lhd$stockValue,90)
jnjL90 <- lag(Ljnj$stockValue,90)
nflxL90 <- lag(Lnflx$stockValue,90)
nkeL90 <- lag(Lnke$stockValue,90)
pcgL90 <- lag(Lpcg$stockValue,90)
rostL90 <- lag(Lrost$stockValue,90)
tevaL90 <- lag(Lteva$stockValue,90)
tjxL90 <- lag(Ltjx$stockValue,90)
wmtL90 <- lag(Lwmt$stockValue,90)
xomL90 <- lag(Lxom$stockValue,90)

Close17_dow$lag90 <-
```

```
c(aapL90,addyyL90,amznL90,costL90,dltrL90,ffinL90,googL90,

hdL90,jnjL90,nflxL90,nkeL90,pcgL90,rostL90,tevaL90,
                                  tjxL90,wmtL90,xomL90)

aapL120 <- lag(Laap$stockValue,120)
addyyL120 <- lag(Laddyy$stockValue,120)
amznL120 <- lag(Lamzn$stockValue,120)
costL120 <- lag(Lcost$stockValue,120)
dltrL120 <- lag(Ldltr$stockValue,120)
ffinL120 <- lag(Lffin$stockValue,120)
googL120 <- lag(Lgoog$stockValue,120)
hdL120 <- lag(Lhd$stockValue,120)
jnjL120 <- lag(Ljnj$stockValue,120)
nflxL120 <- lag(Lnflx$stockValue,120)
nkeL120 <- lag(Lnke$stockValue,120)
pcgL120 <- lag(Lpcg$stockValue,120)
rostL120 <- lag(Lrost$stockValue,120)
tevaL120 <- lag(Lteva$stockValue,120)
tjxL120 <- lag(Ltjx$stockValue,120)
wmtL120 <- lag(Lwmt$stockValue,120)
xomL120 <- lag(Lxom$stockValue,120)

Close17_dow$lag120 <-
c(aapL120,addyyL120,amznL120,costL120,dltrL120,ffinL120,googL120,

hdL120,jnjL120,nflxL120,nkeL120,pcgL120,rostL120,tevaL120,
                                  tjxL120,wmtL120,xomL120)

aapL150 <- lag(Laap$stockValue,150)
addyyL150 <- lag(Laddyy$stockValue,150)
amznL150 <- lag(Lamzn$stockValue,150)
costL150 <- lag(Lcost$stockValue,150)
dltrL150 <- lag(Ldltr$stockValue,150)
ffinL150 <- lag(Lffin$stockValue,150)
googL150 <- lag(Lgoog$stockValue,150)
hdL150 <- lag(Lhd$stockValue,150)
jnjL150 <- lag(Ljnj$stockValue,150)
nflxL150 <- lag(Lnflx$stockValue,150)
nkeL150 <- lag(Lnke$stockValue,150)
pcgL150 <- lag(Lpcg$stockValue,150)
rostL150 <- lag(Lrost$stockValue,150)
tevaL150 <- lag(Lteva$stockValue,150)
tjxL150 <- lag(Ltjx$stockValue,150)
wmtL150 <- lag(Lwmt$stockValue,150)
xomL150 <- lag(Lxom$stockValue,150)

Close17_dow$lag150 <-
c(aapL150,addyyL150,amznL150,costL150,dltrL150,ffinL150,googL150,
```

```
hdL150,jnjL150,nflxL150,nkeL150,pcgL150,rostL150,tevaL150,
                                  tjxL150,wmtL150,xomL150)

aapL180 <- lag(Laap$stockValue,180)
addyyL180 <- lag(Laddyy$stockValue,180)
amznL180 <- lag(Lamzn$stockValue,180)
costL180 <- lag(Lcost$stockValue,180)
dltrL180 <- lag(Ldltr$stockValue,180)
ffinL180 <- lag(Lffin$stockValue,180)
googL180 <- lag(Lgoog$stockValue,180)
hdL180 <- lag(Lhd$stockValue,180)
jnjL180 <- lag(Ljnj$stockValue,180)
nflxL180 <- lag(Lnflx$stockValue,180)
nkeL180 <- lag(Lnke$stockValue,180)
pcgL180 <- lag(Lpcg$stockValue,180)
rostL180 <- lag(Lrost$stockValue,180)
tevaL180 <- lag(Lteva$stockValue,180)
tjxL180 <- lag(Ltjx$stockValue,180)
wmtL180 <- lag(Lwmt$stockValue,180)
xomL180 <- lag(Lxom$stockValue,180)

Close17_dow$lag180 <-
c(aapL180,addyyL180,amznL180,costL180,dltrL180,ffinL180,googL180,

hdL180,jnjL180,nflxL180,nkeL180,pcgL180,rostL180,tevaL180,
                                  tjxL180,wmtL180,xomL180)
```

Save this new table by writing it to csv file. Then we will see how many times each stock is lower than 7,30,60,90,120,150, and 180 days prior in stock value prices for each stock. See if we can use this to automate a data set that selects the stock as good or bad to buy, or good or bad to buy/sell. We could also use this information to create a machine learning data set that will use this information for those stocks that are good/bad at certain points in time to predict what its price will be or if it will return a profit. We already know four of these stocks didn't return a profit, but they are in this portfolio of 17 stocks whose median values were positive when the DOW was decreasing and unemployment was increasing. The other 13 stocks returned a profit, and some substantially such as Netflix with 100 fold increased value, and Amazon with 55 fold increased value.

```
write.csv(Close17_dow, 'Close17_dow_lags.csv', row.names=FALSE)

head(Close17_dow,10)

##            Date DOW_down_median Month Year UE_monthlyRate stockValue
DOW_Close
## 1   2007-01-03             AAP   Jan 2007            4.6      35.58
12474.52
## 18  2007-01-04             AAP   Jan 2007            4.6      35.81
12480.69
## 35  2007-01-05             AAP   Jan 2007            4.6      35.02
```

```
12398.01
## 52  2007-01-08             AAP   Jan 2007            4.6      35.14
12423.49
## 69  2007-01-09             AAP   Jan 2007            4.6      35.44
12416.60
## 86  2007-01-10             AAP   Jan 2007            4.6      35.49
12442.16
## 103 2007-01-11             AAP   Jan 2007            4.6      36.40
12514.98
## 120 2007-01-12             AAP   Jan 2007            4.6      36.20
12556.08
## 137 2007-01-16             AAP   Jan 2007            4.6      36.20
12582.59
## 154 2007-01-17             AAP   Jan 2007            4.6      36.35
12577.15
##      lag7 lag30 lag60 lag90 lag120 lag150 lag180
## 1      NA    NA    NA    NA     NA     NA     NA
## 18     NA    NA    NA    NA     NA     NA     NA
## 35     NA    NA    NA    NA     NA     NA     NA
## 52     NA    NA    NA    NA     NA     NA     NA
## 69     NA    NA    NA    NA     NA     NA     NA
## 86     NA    NA    NA    NA     NA     NA     NA
## 103    NA    NA    NA    NA     NA     NA     NA
## 120 35.58    NA    NA    NA     NA     NA     NA
## 137 35.81    NA    NA    NA     NA     NA     NA
## 154 35.02    NA    NA    NA     NA     NA     NA
```

```
tail(Close17_dow,10)
```

```
##             Date DOW_down_median Month Year UE_monthlyRate stockValue
DOW_Close
## 55998 2020-02-03             XOM   Feb 2020             NA      60.73
28399.81
## 56015 2020-02-04             XOM   Feb 2020             NA      59.97
28807.63
## 56032 2020-02-05             XOM   Feb 2020             NA      62.73
29290.85
## 56049 2020-02-06             XOM   Feb 2020             NA      61.88
29379.77
## 56066 2020-02-07             XOM   Feb 2020             NA      61.47
29102.51
## 56083 2020-02-10             XOM   Feb 2020             NA      59.96
29276.82
## 56100 2020-02-11             XOM   Feb 2020             NA      60.53
29276.34
## 56117 2020-02-12             XOM   Feb 2020             NA      61.27
29551.42
## 56134 2020-02-13             XOM   Feb 2020             NA      60.93
29423.31
## 56151 2020-02-14             XOM   Feb 2020             NA      60.65
```

```
29398.08
##        lag7 lag30 lag60 lag90 lag120 lag150 lag180
## 55998 66.77 69.87 73.09 71.14  69.63  76.63  76.36
## 56015 66.32 69.39 71.49 71.35  70.49  76.56  75.91
## 56032 64.74 69.94 73.01 70.97  67.65  75.72  75.90
## 56049 64.65 70.29 70.77 71.48  67.25  76.44  76.25
## 56066 64.11 70.02 70.34 70.61  68.30  76.13  75.56
## 56083 64.79 70.13 69.37 68.95  69.45  76.48  73.79
## 56100 62.12 69.89 68.80 67.15  69.03  76.43  74.10
## 56117 60.73 69.48 68.50 67.98  69.72  77.51  72.61
## 56134 59.97 69.78 69.19 68.97  69.57  77.57  72.16
## 56151 62.73 70.90 68.52 68.02  67.49  77.63  71.97
```