ROI on Hand Picked Stocks 2007-2020

Janis Corona

2/17/2020

```
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$prevDay <- prevDay[1:3303,1]
dailyChange <- as.data.frame(Close_noNAs$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

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$DailyVolume <- rowSums(Volume1, na.rm=TRUE)

dayVol <- as.data.frame(Volume1$DailyVolume)
colnames(dayVol) <- 'previousDayVolume'
zero <- as.data.frame(as.numeric(dayVol$previousDayVolume[1]))
colnames(zero) <- 'previousDayVolume'
prevDay1 <- rbind(zero,dayVol)
Volume1$prevDayVolume <- prevDay1[1:3303,1]

dailyVolumeChange <- as.data.frame(Volume1$DailyVolume-Volume1$prevDayVolume)
colnames(dailyVolumeChange) <- 'dailyVolumeChange'

Volume2 <- cbind(Volume1,dailyVolumeChange)
Volume2$VolumeRatioDaily2Initial <-
Volume2$DailyVolume/Volume2$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"              "DailyValue"
## [109] "prevDay"                "dailyValueChange"
## [111] "ROI_dollars"            "Date"
```

```
## [113] "DayOfWeek"                "Month"
## [115] "Year"                     "UE_monthlyRate"
## [117] "DailyVolume"              "prevDayVolume"
## [119] "dailyVolumeChange"        "VolumeRatioDaily2Initial"
```

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

```
Stocks$ValueRatioDaily2Initial <- Stocks$DailyValue/Stocks$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$DailyRatios_X_UE <-
Stocks$ValueRatioDaily2Initial*Stocks$VolumeRatioDaily2Initial*Stocks$UE_mont
hlyRate
```

Add an exponential calculation field based on the unemployment rate for rate, and using numeric day of the month for t, and k as the month.

```
Stocks$dayOfMonth <- day(Stocks$Date)
dayOfMonth <- day(Stocks$Date)
ue1 <- Stocks$UE_monthlyRate

Stocks$poisson <- (exp(-(ue1))*(ue1)^dayOfMonth)/(factorial(dayOfMonth))

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$TGT.Close[1:3302])
FTRa <- c(0, stocks1$FTR.Close[1:3302])
UBSIa <- c(0,stocks1$UBSI.Close[1:3302])
HDa <- c(0,stocks1$HD.Close[1:3302])
JPMa <- c(0,stocks1$JPM.Close[1:3302])
XOMa <- c(0,stocks1$XOM.Close[1:3302])
CVXa <- c(0,stocks1$CVX.Close[1:3302])
NSANYa <- c(0,stocks1$NSANY.Close[1:3302])
MGMa <- c(0,stocks1$MGM.Close[1:3302])
TEVAa <- c(0, stocks1$TEVA.Close[1:3302])
HSTa <- c(0, stocks1$HST.Close[1:3302])
WFCa <- c(0, stocks1$WFC.Close[1:3302])
WWEa <- c(0, stocks1$WWE.Close[1:3302])
INOa <- c(0,stocks1$INO.Close[1:3302])
SCEa <- c(0,stocks1$SCE.PB.Close[1:3302])
FFINa <- c(0,stocks1$FFIN.Close[1:3302])
GOOGa <- c(0,stocks1$GOOG.Close[1:3302])
WMa <- c(0,stocks1$WM.Close[1:3302])
ONCYa <- c(0,stocks1$ONCY.Close[1:3302])
Sa <- c(0,stocks1$S.Close[1:3302])
Fa <- c(0,stocks1$F.Close[1:3302])
ARWRa <- c(0,stocks1$ARWR.Close[1:3302])
COSTa <- c(0,stocks1$COST.Close[1:3302])
AALa <- c(0,stocks1$AAL.Close[1:3302])
JWNa <- c(0,stocks1$JWN.Close[1:3302])
NUSa <- c(0,stocks1$NUS.Close[1:3302])
ADDYYa <- c(0,stocks1$ADDYY.Close[1:3302])
KSSa <- c(0,stocks1$KSS.Close[1:3302])
MSFTa <- c(0,stocks1$MSFT.Close[1:3302])
LUVa <- c(0,stocks1$LUV.Close[1:3302])
HMCa <- c(0,stocks1$HMC.Close[1:3302])
PCGa <- c(0,stocks1$PCG.Close[1:3302])
DLTRa <- c(0,stocks1$DLTR.Close[1:3302])
```

```
KGJIa <- c(0,stocks1$KGJI.Close[1:3302])
NKEa <- c(0,stocks1$NKE.Close[1:3302])
AMZNa <- c(0,stocks1$AMZN.Close[1:3302])
ROSTa <- c(0,stocks1$ROST.Close[1:3302])
WMTa <- c(0,stocks1$WMT.Close[1:3302])
TJXa <- c(0,stocks1$TJX.Close[1:3302])
TMa <- c(0,stocks1$TM.Close[1:3302])
Ta <- c(0,stocks1$T.Close[1:3302])
JNJa <- c(0,stocks1$JNJ.Close[1:3302])
Ca <- c(0,stocks1$C.Close[1:3302])
EPDa <- c(0,stocks1$EPD.Close[1:3302])
VZa <- c(0,stocks1$VZ.Close[1:3302])
HRBa <- c(0,stocks1$HRB.Close[1:3302])
NFLXa <- c(0,stocks1$NFLX.Close[1:3302])
AAPa <- c(0,stocks1$AAP.Close[1:3302])
HOFTa <- c(0,stocks1$HOFT.Close[1:3302])
SIGa <- c(0,stocks1$SIG.Close[1:3302])
RRGBa <- c(0,stocks1$RRGB.Close[1:3302])
Ma <- c(0,stocks1$M.Close[1:3302])
JBLUa <- c(0,stocks1$JBLU.Close[1:3302])
```

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

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.

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

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

All the columns we now have are:

```
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] "MonthYear"                "DailyValue"
## [109] "prevDay"                  "dailyValueChange"
## [111] "ROI_dollars"              "Date"
## [113] "DayOfWeek"                "Month"
## [115] "Year"                     "UE_monthlyRate"
## [117] "DailyVolume"              "prevDayVolume"
## [119] "dailyVolumeChange"        "VolumeRatioDaily2Initial"
## [121] "ValueRatioDaily2Initial"  "DailyRatios_X_UE"
## [123] "dayOfMonth"               "poisson"
## [125] "TGT_ROI_dollars"          "FTR_ROI_dollars"
## [127] "UBSI_ROI_dollars"         "HD_ROI_dollars"
## [129] "JPM_ROI_dollars"          "XOM_ROI_dollars"
## [131] "CVX_ROI_dollars"          "NSANY_ROI_dollars"
## [133] "MGM_ROI_dollars"          "TEVA_ROI_dollars"
## [135] "HST_ROI_dollars"          "WFC_ROI_dollars"
## [137] "WWE_ROI_dollars"          "INO_ROI_dollars"
## [139] "SCE.PB_ROI_dollars"       "FFIN_ROI_dollars"
## [141] "GOOG_ROI_dollars"         "WM_ROI_dollars"
## [143] "ONCY_ROI_dollars"         "S_ROI_dollars"
## [145] "F_ROI_dollars"            "ARWR_ROI_dollars"
## [147] "COST_ROI_dollars"         "AAL_ROI_dollars"
## [149] "JWN_ROI_dollars"          "NUS_ROI_dollars"
## [151] "HMC_ROI_dollars"          "AMZN_ROI_dollars"
## [153] "T_ROI_dollars"            "HRB_ROI_dollars"
## [155] "RRGB_ROI_dollars"         "ADDYY_ROI_dollars"
## [157] "PCG_ROI_dollars"          "ROST_ROI_dollars"
## [159] "JNJ_ROI_dollars"          "NFLX_ROI_dollars"
## [161] "M_ROI_dollars"            "KSS_ROI_dollars"
## [163] "DLTR_ROI_dollars"         "WMT_ROI_dollars"
## [165] "C_ROI_dollars"            "AAP_ROI_dollars"
## [167] "JBLU_ROI_dollars"         "MSFT_ROI_dollars"
## [169] "KGJI_ROI_dollars"         "EPD_ROI_dollars"
## [171] "TJX_ROI_dollars"          "HOFT_ROI_dollars"
## [173] "LUV_ROI_dollars"          "NKE_ROI_dollars"
## [175] "TM_ROI_dollars"           "VZ_ROI_dollars"
## [177] "SIG_ROI_dollars"          "TGT_dailyChange"
## [179] "FTR_dailyChange"          "UBSI_dailyChange"
## [181] "HD_dailyChange"           "JPM_dailyChange"
## [183] "XOM_dailyChange"          "CVX_dailyChange"
## [185] "NSANY_dailyChange"        "MGM_dailyChange"
## [187] "TEVA_dailyChange"         "HST_dailyChange"
## [189] "WFC_dailyChange"          "WWE_dailyChange"
## [191] "INO_dailyChange"          "SCE.PB_dailyChange"
## [193] "FFIN_dailyChange"         "GOOG_dailyChange"
## [195] "WM_dailyChange"           "ONCY_dailyChange"
## [197] "S_dailyChange"            "F_dailyChange"
## [199] "ARWR_dailyChange"         "COST_dailyChange"
## [201] "AAL_dailyChange"          "JWN_dailyChange"
## [203] "NUS_dailyChange"          "HMC_dailyChange"
## [205] "AMZN_dailyChange"         "T_dailyChange"
```

```
## [207] "HRB_dailyChange"          "RRGB_dailyChange"
## [209] "ADDYY_dailyChange"        "PCG_dailyChange"
## [211] "ROST_dailyChange"         "JNJ_dailyChange"
## [213] "NFLX_dailyChange"         "M_dailyChange"
## [215] "KSS_dailyChange"          "DLTR_dailyChange"
## [217] "WMT_dailyChange"          "C_dailyChange"
## [219] "AAP_dailyChange"          "JBLU_dailyChange"
## [221] "MSFT_dailyChange"         "KGJI_dailyChange"
## [223] "EPD_dailyChange"          "TJX_dailyChange"
## [225] "HOFT_dailyChange"         "LUV_dailyChange"
## [227] "NKE_dailyChange"          "TM_dailyChange"
## [229] "VZ_dailyChange"           "SIG_dailyChange"
```

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. Lets also use the poisson formula that chose the day of the month, because some people might want to buy stocks on pay day around the 1st or 15th for most, or also some on every Friday or every other Friday which would also use the day of the week.

```
m <- StocksSTATS[order(StocksSTATS$Date, decreasing=FALSE)[3303], 124:176]
t <- as.data.frame(t(m))
colnames(t) <- row.names(m)
t$StockROI <- row.names(t)

Troi <- t[order(t$'2020-02-14', decreasing=TRUE),]

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

##                    2020-02-14           StockROI
## AMZN_ROI_dollars     2094.450   AMZN_ROI_dollars
## GOOG_ROI_dollars     1292.331   GOOG_ROI_dollars
## SCE.PB_ROI_dollars    679.000 SCE.PB_ROI_dollars
## FTR_ROI_dollars      -225.180    FTR_ROI_dollars
## C_ROI_dollars        -431.710      C_ROI_dollars
```

The above table shows the three highest returns on investment and the three lowest since Jan 3, 2007 to Feb 14, 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.

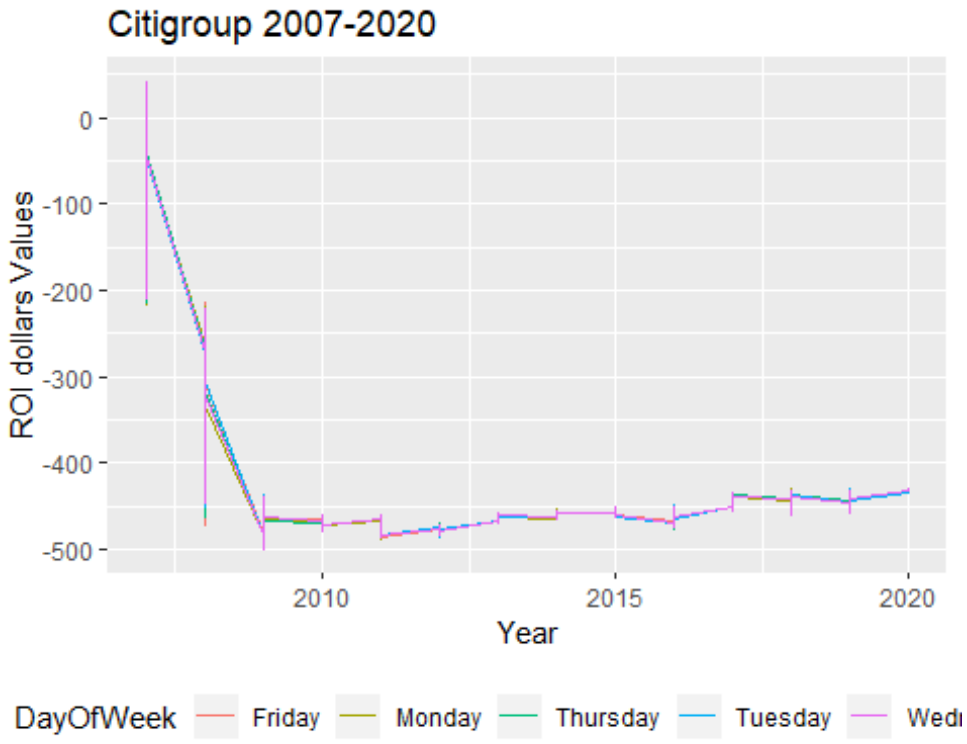
```
amzn <- grep('AMZN', colnames(StocksSTATS))
c <- grep('^C[.|_]', colnames(StocksSTATS))
```

```
C_stock <- StocksSTATS[,c(c,107,112:116,123:124)]
amzn_stock <- StocksSTATS[,c(amzn,107,112:116,123:124)]
```

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))+
  #geom_point()+
  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.

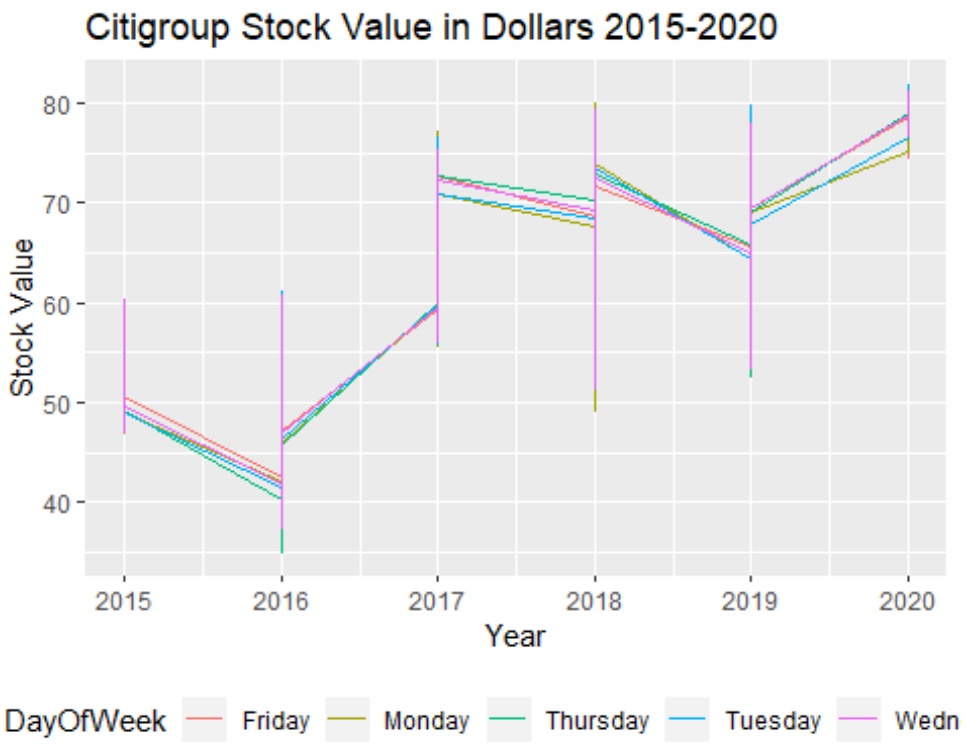
Lets look at the years from 2016-2020 to see this plotted Citigroup stock.

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

ggplot(data = y2015plus, aes(x=Year, y=C.Close,group=DayOfWeek)) +
```

```
  geom_line(aes(color=DayOfWeek))+
  #geom_point()+
  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
```

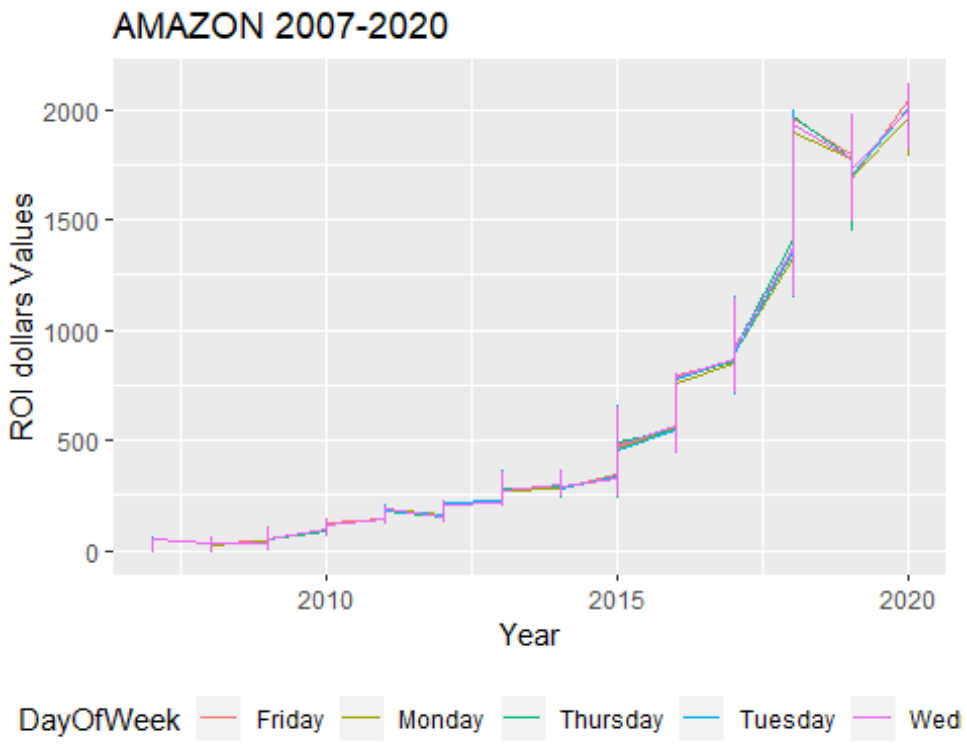


We see from the above plot that Citigroup was good to buy at the start of 2016 or 2019 if you want to see an increase all year long, but in 2017-2018 it decreased.Overall, if investing since 2016, the stock increased from the high $40 to the mid-high $70 range. This would be good to cross reference with 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))+
  #geom_point()+
  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
```



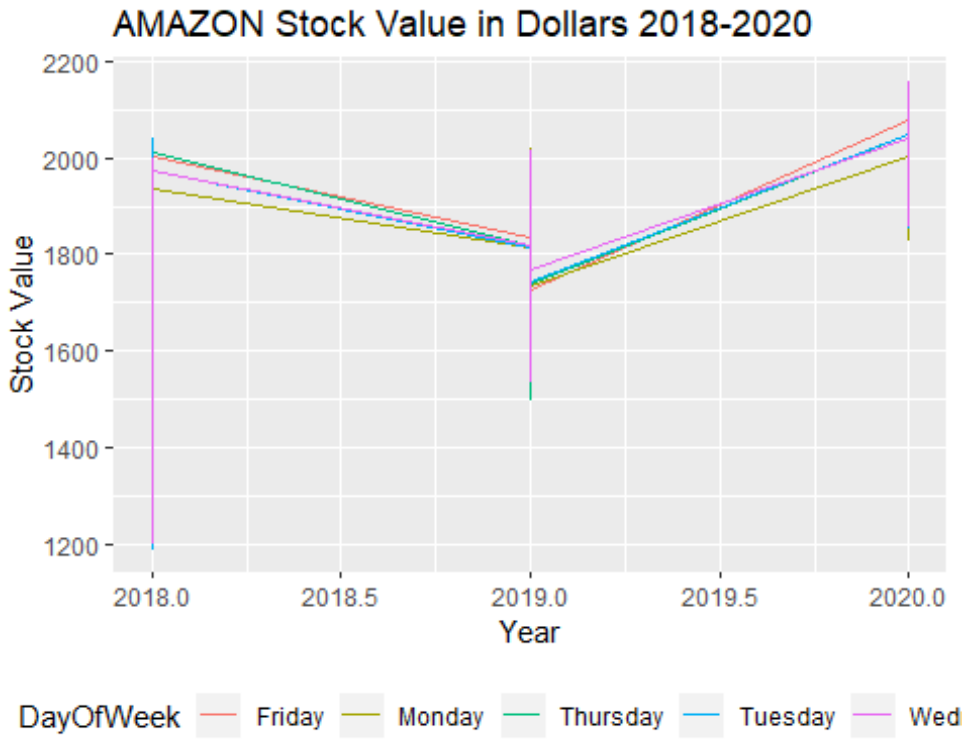
We can see from the plot above that buying AMAZON stock anywhere before 2010, was a great idea. But we also see that the stock would have been good to buy around 2010-2018 or 2019 but not in 2018, as it overall increased its return on investment in dollars initially invested.In 2018, you bought high 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))+
  #geom_point()+
  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])
poisson_Citi <- C_stock %>% group_by(dayOfMonth) %>% summarise_at(vars(v1),
median,
                                                                  na.rm=T)
poisson_Citi <- as.data.frame(poisson_Citi)
colnames(poisson_Citi)[2] <- 'Citi_Median_Volume'
poisson_Citi <- poisson_Citi[order(poisson_Citi$Citi_Median_Volume,
decreasing=T),]
headTail_Citi_volume <- rbind(head(poisson_Citi,3), tail(poisson_Citi,3))
headTail_Citi_volume

##    dayOfMonth Citi_Median_Volume
## 16         16           22388100
## 31         31           22302200
## 3           3           21522100
## 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.   :-298.3000
##  1st Qu.: 41.80   1st Qu.: 13043950   1st Qu.:-468.7   1st Qu.:  -0.6750
##  Median : 51.59   Median : 19535800   Median :-458.9   Median :  -0.0100
##  Mean   : 93.33   Mean   : 27022488   Mean   :-417.2   Mean   :   0.0209
##  3rd Qu.: 69.58   3rd Qu.: 33314650   3rd Qu.:-440.9   3rd Qu.:   0.6500
##  Max.   :552.50   Max.   :377263800   Max.   :  42.0   Max.   : 510.5000
##
##     MonthYear         Date              DayOfWeek            Month
##  Aug-2007:  23   Min.   :2007-01-03   Length:3303        Length:3303
##  Aug-2011:  23   1st Qu.:2010-04-14   Class :character   Class :character
##  Aug-2012:  23   Median :2013-07-25   Mode  :character   Mode  :character
##  Aug-2016:  23   Mean   :2013-07-24
##  Aug-2017:  23   3rd Qu.:2016-11-01
##  Aug-2018:  23   Max.   :2020-02-14
##  (Other) :3165
##       Year      UE_monthlyRate     dayOfMonth       poisson
##  Min.   :2007   Min.   : 3.500   Min.   : 1.00   Min.   :0.000000
##  1st Qu.:2010   1st Qu.: 4.600   1st Qu.: 8.00   1st Qu.:0.000000
##  Median :2013   Median : 5.600   Median :16.00   Median :0.000563
##  Mean   :2013   Mean   : 6.282   Mean   :15.72   Mean   :0.032501
##  3rd Qu.:2016   3rd Qu.: 8.200   3rd Qu.:23.00   3rd Qu.:0.047937
##  Max.   :2020   Max.   :10.000   Max.   :31.00   Max.   :0.215785
##                 NA's   :10                       NA's   :10
```

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
Date
## 2007-04-02  510.50   2282100          0.00    510.500000  Apr-2007 2007-
04-02
## 2013-04-02   44.11   1005100       -466.39      0.320000  Apr-2013 2013-
04-02
## 2015-12-28   52.38 377263800       -458.12     -0.329998  Dec-2015 2015-
12-28
## 2008-06-02  214.60  15302800       -295.90   -298.300018  Jun-2008 2008-
06-02
##            DayOfWeek Month Year UE_monthlyRate dayOfMonth      poisson
## 2007-04-02    Monday   Apr 2007            4.5          2 1.124786e-01
## 2013-04-02   Tuesday   Apr 2013            7.6          2 1.445304e-02
## 2015-12-28    Monday   Dec 2015            5.0         28 8.232787e-13
## 2008-06-02    Monday   Jun 2008            5.6          2 5.798250e-02
```

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.