Lab1: Basic Descriptive Analytics with R

Rabea Radman

December 27, 2020

Importing Data into R

```
Boeing = read.csv("BoeingStock.csv")
head(Boeing)
```

```
##     Date StockPrice
## 1 1/1/70   27.85381
## 2 2/1/70   22.38105
## 3 3/1/70   23.10524
## 4 4/1/70   21.57136
## 5 5/1/70   18.93286
## 6 6/1/70   15.44318
```

```
CocaCola = read.csv("CocaColaStock.csv")
head(CocaCola)
```

```
##     Date StockPrice
## 1 1/1/70   83.36810
## 2 2/1/70   81.59105
## 3 3/1/70   81.33810
## 4 4/1/70   76.80591
## 5 5/1/70   69.27857
## 6 6/1/70   72.01545
```

```
GE = read.csv("GEStock.csv")
head(GE)
```

```
##     Date StockPrice
## 1 1/1/70   74.25333
## 2 2/1/70   69.97684
## 3 3/1/70   72.15857
## 4 4/1/70   74.25273
## 5 5/1/70   66.66524
## 6 6/1/70   67.59318
```

```
IBM = read.csv("IBMStock.csv")
head(IBM)
```

```
##     Date StockPrice
## 1 1/1/70   360.3190
## 2 2/1/70   346.7237
## 3 3/1/70   327.3457
## 4 4/1/70   319.8527
## 5 5/1/70   270.3752
## 6 6/1/70   267.2050
```

```
ProcterGamble = read.csv("ProcterGambleStock.csv")
head(ProcterGamble)
```

```
##     Date StockPrice
## 1 1/1/70  111.87429
## 2 2/1/70  111.45368
## 3 3/1/70  108.45143
## 4 4/1/70  106.28864
## 5 5/1/70   73.33286
## 6 6/1/70   48.31864
```

Clean Date Field

Viewing type of data

```
str(Boeing)
```

```
## 'data.frame':    480 obs. of  2 variables:
##  $ Date      : chr  "1/1/70" "2/1/70" "3/1/70" "4/1/70" ...
##  $ StockPrice: num  27.9 22.4 23.1 21.6 18.9 ...
```

```
str(CocaCola)
```

```
## 'data.frame':    480 obs. of  2 variables:
##  $ Date      : chr  "1/1/70" "2/1/70" "3/1/70" "4/1/70" ...
##  $ StockPrice: num  83.4 81.6 81.3 76.8 69.3 ...
```

```
str(ProcterGamble)
```

```
## 'data.frame':    480 obs. of  2 variables:
##  $ Date      : chr  "1/1/70" "2/1/70" "3/1/70" "4/1/70" ...
##  $ StockPrice: num  111.9 111.5 108.5 106.3 73.3 ...
```

```
str(GE)
```

```
## 'data.frame':    480 obs. of  2 variables:
##  $ Date      : chr  "1/1/70" "2/1/70" "3/1/70" "4/1/70" ...
##  $ StockPrice: num  74.3 70 72.2 74.3 66.7 ...
```

```
str(IBM)
```

```
## 'data.frame':    480 obs. of  2 variables:
##  $ Date      : chr  "1/1/70" "2/1/70" "3/1/70" "4/1/70" ...
##  $ StockPrice: num  360 347 327 320 270 ...
```

What is the data type of the Data variable?

Answer: Data Frame

Using as.Date() function to convert the dates

```
Boeing$Date = as.Date(Boeing$Date, "%m/%d/%y")
CocaCola$Date = as.Date(CocaCola$Date, "%m/%d/%y")
ProcterGamble$Date = as.Date(ProcterGamble$Date, "%m/%d/%y")
GE$Date = as.Date(GE$Date, "%m/%d/%y")
IBM$Date = as.Date(IBM$Date, "%m/%d/%y")
```

View data after converting dates

```
head(Boeing)
```

```
##         Date StockPrice
## 1 1970-01-01   27.85381
## 2 1970-02-01   22.38105
## 3 1970-03-01   23.10524
## 4 1970-04-01   21.57136
## 5 1970-05-01   18.93286
## 6 1970-06-01   15.44318
```

```
head(CocaCola)
```

```
##         Date StockPrice
## 1 1970-01-01   83.36810
## 2 1970-02-01   81.59105
## 3 1970-03-01   81.33810
## 4 1970-04-01   76.80591
## 5 1970-05-01   69.27857
## 6 1970-06-01   72.01545
```

```
head(ProcterGamble)
```

```
##         Date StockPrice
## 1 1970-01-01  111.87429
## 2 1970-02-01  111.45368
## 3 1970-03-01  108.45143
## 4 1970-04-01  106.28864
## 5 1970-05-01   73.33286
## 6 1970-06-01   48.31864
```

```
head(GE)
```

```
##         Date StockPrice
## 1 1970-01-01   74.25333
## 2 1970-02-01   69.97684
## 3 1970-03-01   72.15857
## 4 1970-04-01   74.25273
## 5 1970-05-01   66.66524
## 6 1970-06-01   67.59318
```

```
head(IBM)
```

```
##         Date StockPrice
## 1 1970-01-01   360.3190
## 2 1970-02-01   346.7237
## 3 1970-03-01   327.3457
## 4 1970-04-01   319.8527
## 5 1970-05-01   270.3752
## 6 1970-06-01   267.2050
```

Warm-up/Basic statistics Questions:

1. How many rows of data are in each dataset?

```
nrow(Boeing)
```

```
## [1] 480
```

```
nrow(CocaCola)
```

```
## [1] 480
```

```
nrow(ProcterGamble)
```

```
## [1] 480
```

```
nrow(GE)
```

```
## [1] 480
```

```
nrow(IBM)
```

```
## [1] 480
```

Answer: 480 rows for all datasets.

2. What is the earliest/latest year in our datasets?

```
summary(Boeing)
```

```
##       Date              StockPrice
##  Min.   :1970-01-01   Min.   : 12.74
##  1st Qu.:1979-12-24   1st Qu.: 34.64
##  Median :1989-12-16   Median : 44.88
##  Mean   :1989-12-15   Mean   : 46.59
##  3rd Qu.:1999-12-08   3rd Qu.: 57.21
##  Max.   :2009-12-01   Max.   :107.28
```

```
summary(CocaCola)
```

```
##       Date              StockPrice
##  Min.   :1970-01-01   Min.   : 30.06
##  1st Qu.:1979-12-24   1st Qu.: 42.76
##  Median :1989-12-16   Median : 51.44
##  Mean   :1989-12-15   Mean   : 60.03
##  3rd Qu.:1999-12-08   3rd Qu.: 69.62
##  Max.   :2009-12-01   Max.   :146.58
```

```
summary(ProcterGamble)
```

```
##       Date              StockPrice
##  Min.   :1970-01-01   Min.   : 46.88
##  1st Qu.:1979-12-24   1st Qu.: 62.48
##  Median :1989-12-16   Median : 78.34
##  Mean   :1989-12-15   Mean   : 77.70
##  3rd Qu.:1999-12-08   3rd Qu.: 89.47
##  Max.   :2009-12-01   Max.   :149.62
```

```
summary(GE)
```

```
##       Date              StockPrice
##  Min.   :1970-01-01   Min.   :  9.294
##  1st Qu.:1979-12-24   1st Qu.: 44.214
##  Median :1989-12-16   Median : 55.812
##  Mean   :1989-12-15   Mean   : 59.303
##  3rd Qu.:1999-12-08   3rd Qu.: 72.226
##  Max.   :2009-12-01   Max.   :156.844
```

```
summary(IBM)
```

```
##       Date              StockPrice
##  Min.   :1970-01-01   Min.   : 43.40
##  1st Qu.:1979-12-24   1st Qu.: 88.34
##  Median :1989-12-16   Median :112.11
##  Mean   :1989-12-15   Mean   :144.38
##  3rd Qu.:1999-12-08   3rd Qu.:165.41
##  Max.   :2009-12-01   Max.   :438.90
```

Answer: Min Year = 1970, Max Year = 2009

3. For the period above what is the average stock price of Coca Cola?

```
cocacola_mean = mean(CocaCola$StockPrice)
cocacola_mean
```

```
## [1] 60.02973
```

Answer: 60.02973

4. What is the maximum price of IBM during this period?

```
max_price_IBM = max(IBM$StockPrice)
max_price_IBM
```

```
## [1] 438.9016
```

Answer: 438.9016

5. What is the standard deviation of P&G stock price over this period?

```
sd_PG = sd(ProcterGamble$StockPrice)
sd_PG
```

```
## [1] 18.19414
```

Answer: 18.19414

6. What is the median price of Boeing in the last 5 years for which we have data?

Get Max Date in UTC to be able to subtract dates (year)

```
Bng_maxDate = as.POSIXlt(max(Boeing$Date))
Bng_maxDate
```

```
## [1] "2009-12-01 UTC"
```

Subtract 5 years from the max date (to get the last 5 years)

```
Bng_maxDate$year = Bng_maxDate$year - 5
Bng_maxDate
```

```
## [1] "2004-12-01 UTC"
```

Slice to get data for last 5 years

```
Bng_Last5Yrs = Boeing[as.POSIXlt(Boeing$Date) > Bng_maxDate, ]
print(head(Bng_Last5Yrs))
```

```
##           Date StockPrice
## 421 2005-01-01   50.67450
## 422 2005-02-01   53.22474
## 423 2005-03-01   57.20773
## 424 2005-04-01   58.56143
## 425 2005-05-01   61.01381
## 426 2005-06-01   63.78818
```

```
print(tail(Bng_Last5Yrs))
```

```
##           Date StockPrice
## 475 2009-07-01   41.48273
## 476 2009-08-01   45.99429
## 477 2009-09-01   51.36286
## 478 2009-10-01   51.15909
## 479 2009-11-01   50.69650
## 480 2009-12-01   55.02864
```

Calculate median for the last 5 years

```
Bng_Last5Yrs_median = median((Bng_Last5Yrs$StockPrice))
print(Bng_Last5Yrs_median)
```

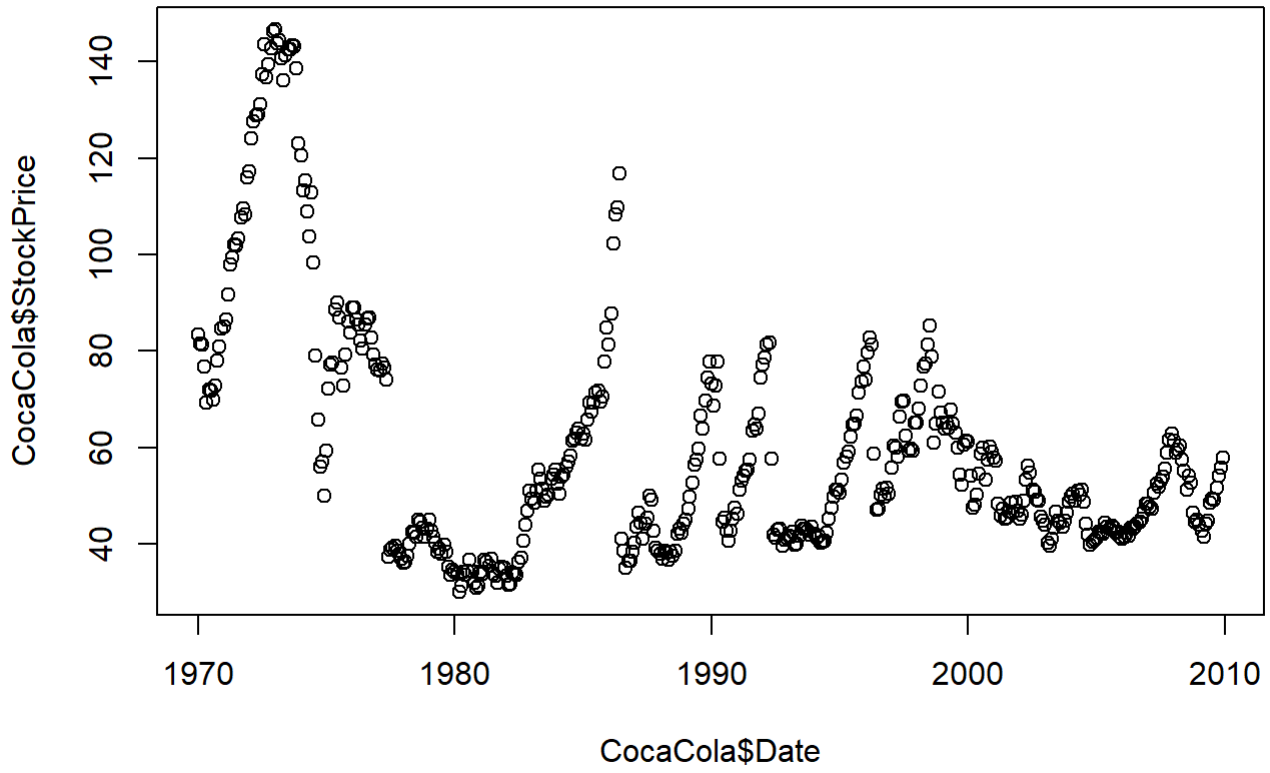
```
## [1] 69.67567
```

Answer: 69.67567

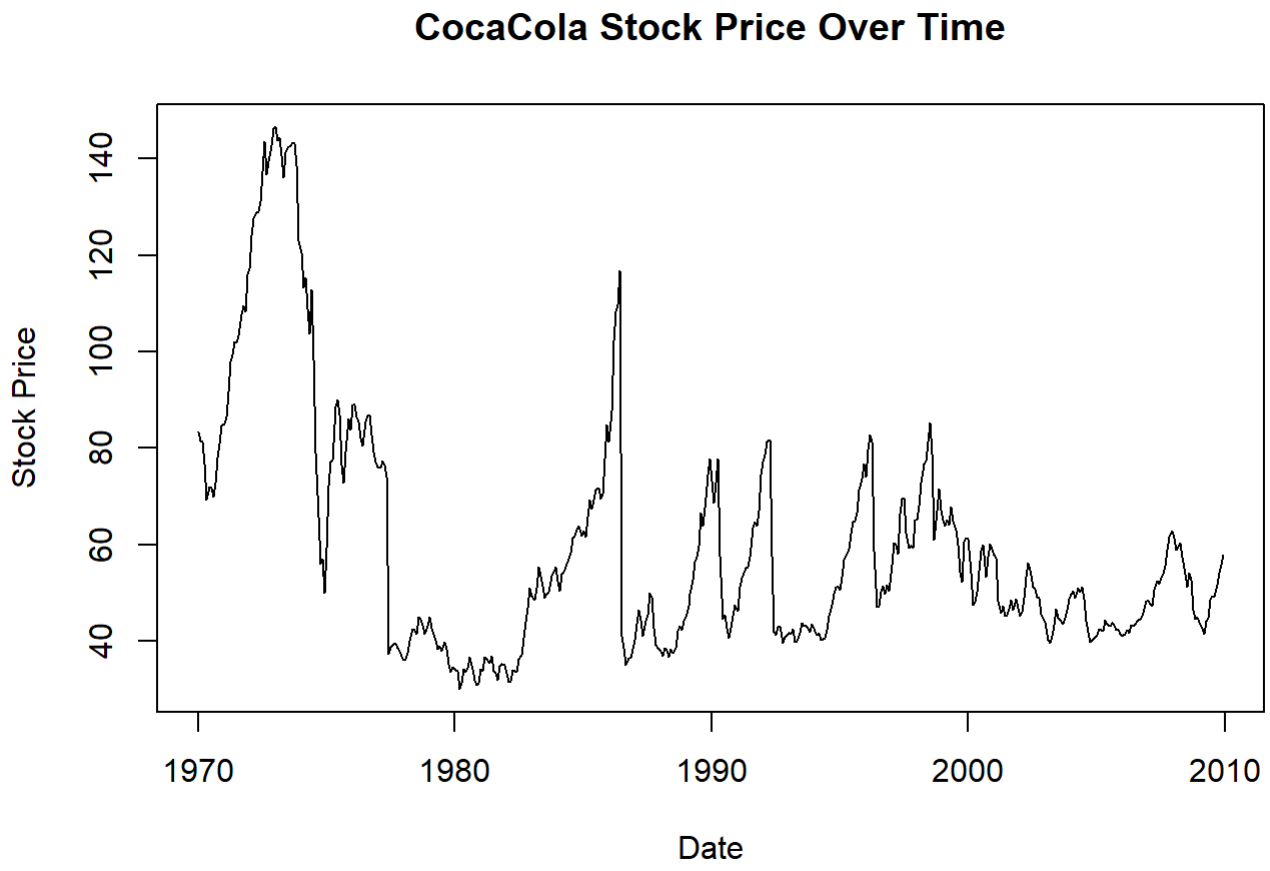
Basic Plotting Questions

Part 1

```
plot(CocaCola$Date, CocaCola$StockPrice)
```




```
plot(CocaCola$Date, CocaCola$StockPrice, main = "CocaCola Stock Price Over Time"
     , xlab = "Date", ylab = "Stock Price", type = "l")
```



1. Identify the year during which Coca-Cola had the highest/lowest stock price?

```
CocaCola[which.max(CocaCola$StockPrice),]
```

```
##          Date StockPrice
## 37 1973-01-01   146.5843
```

```
CocaCola[which.min(CocaCola$StockPrice),]
```

```
##           Date StockPrice
## 123 1980-03-01   30.05714
```

Answer: CocaCola Highest Price was 146.5843 in year = 1973

& lowest Price was = 30.05714 in year = 1980

2. What calendar year did it look to have the biggest (Year-over-Year) percentage increase?

```
CocaCola$Date = as.Date(CocaCola$Date, "%m/%d/%y")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

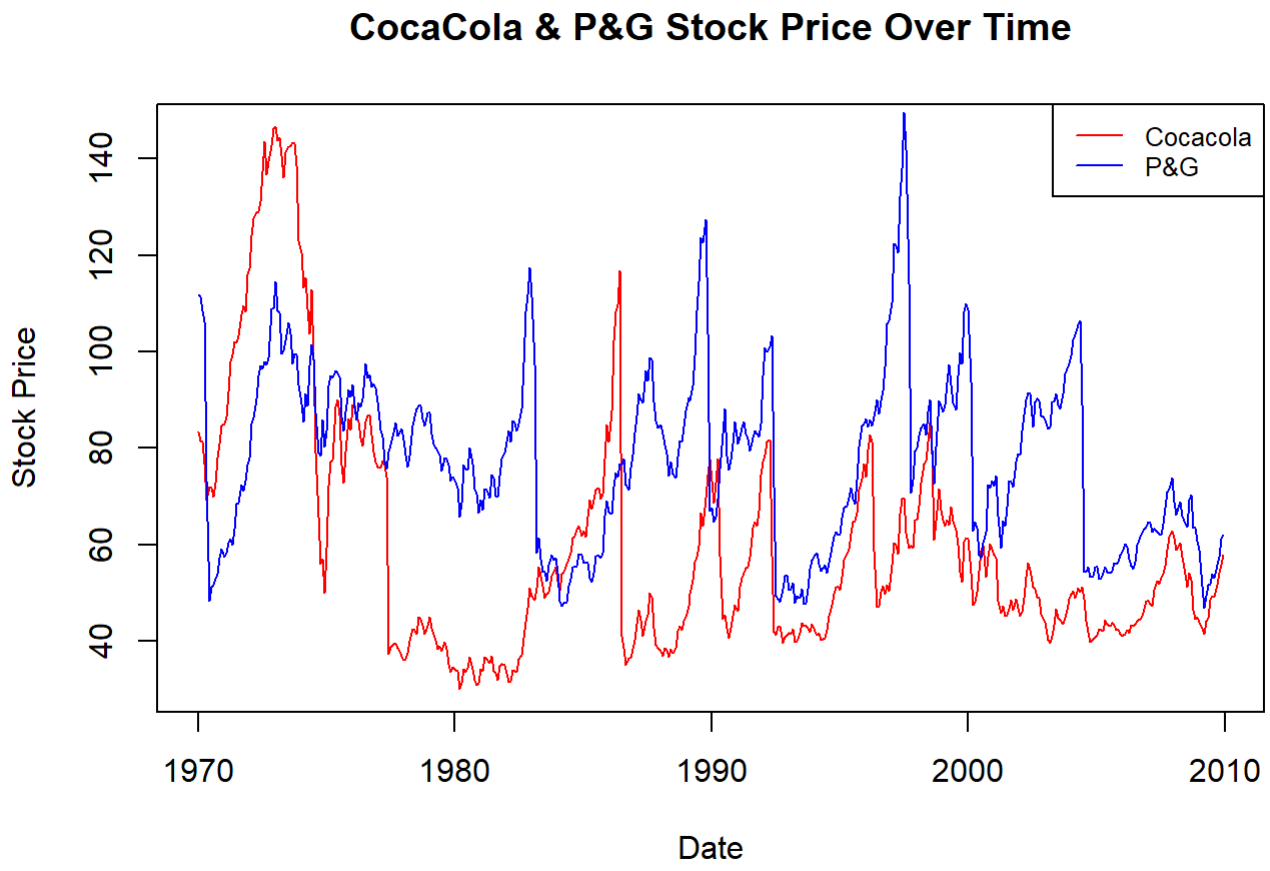
```
df2 = CocaCola%>%
  arrange(Date) %>%
  mutate(pct_chg = 100*((StockPrice - lag(StockPrice))/lag(StockPrice)))
df2[which.max(df2$pct_chg),]
```

```
##          Date StockPrice  pct_chg
## 62 1975-02-01   72.22158 21.90418
```

Answer: Biggest YOY percentage increase is year 1975

Part 2

```
plot(CocaCola$Date, CocaCola$StockPrice, type = "l",
     main = "CocaCola & P&G Stock Price Over Time", xlab = "Date",
     ylab = "Stock Price", col = "red")
lines(ProcterGamble$Date, ProcterGamble$StockPrice, col = "blue")
legend("topright",legend = c("Cocacola","P&G"), col = c("red","blue"),
       lty = 1, cex = 0.8)
```



1. In March of 2000 the stock market plummeted as the tech bubble burst. Using the plot above, which company's stock dropped more (relatively – i.e. percentage-wise)?

Answer: Cocacola

2. In the year 1983 which company stock was going up? Which was going down?

```
df_merge = merge(CocaCola,ProcterGamble,all.x = TRUE, all.y = TRUE, by = "Date")
library(plotly)
```

```
## Loading required package: ggplot2
```

```
##
## Attaching package: 'plotly'
```

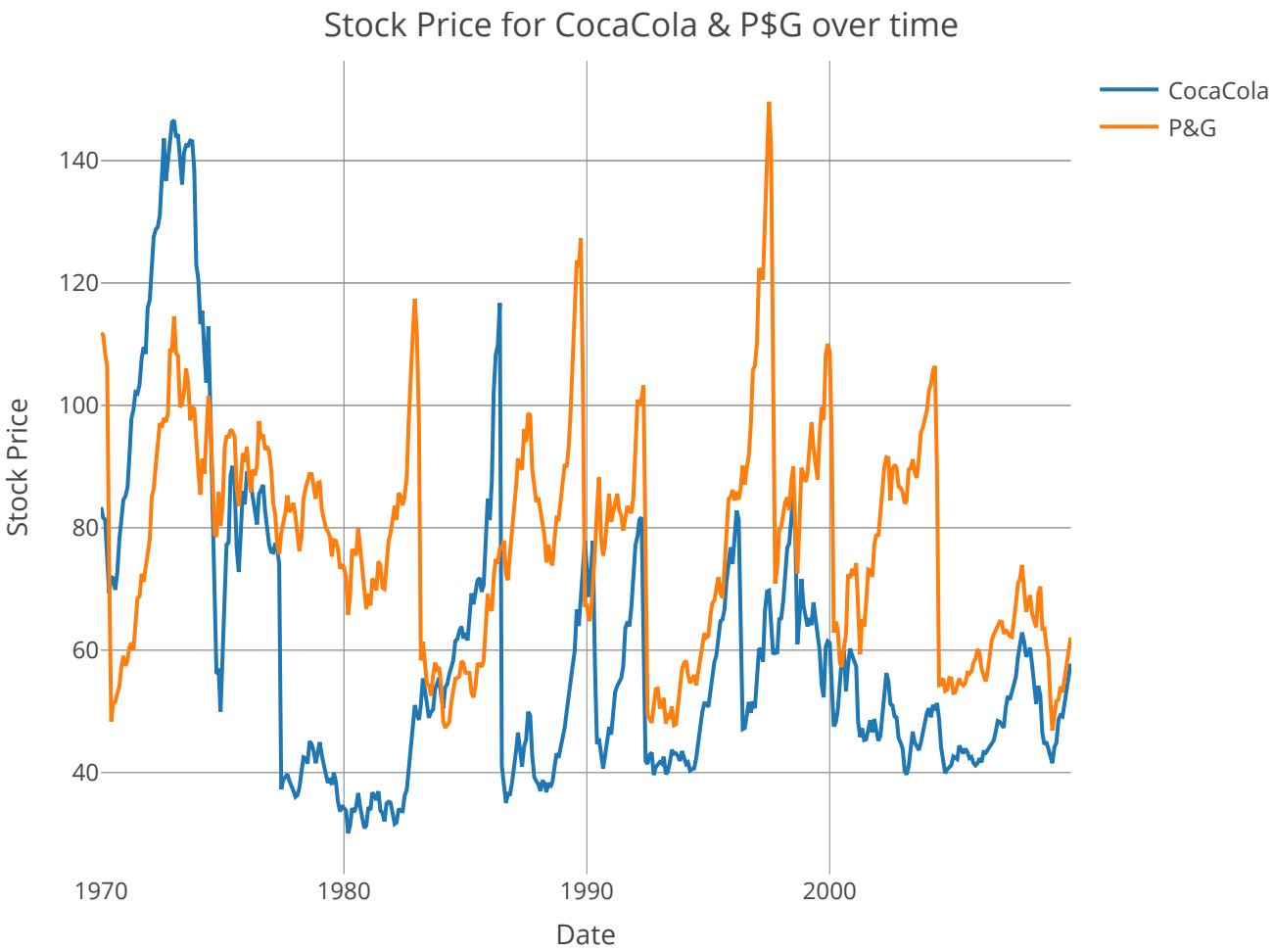
```
## The following object is masked from 'package:ggplot2':
##
##     last_plot
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
## The following object is masked from 'package:graphics':
##
##     layout
```

```
fig = plot_ly(df_merge)%>%
  add_trace(x = ~Date, y = ~StockPrice.x, name = "CocaCola", type = "scatter", mode = "lines")%
>%
  add_trace(x = ~Date, y =~StockPrice.y, name = "P&G", type = "scatter",
            mode = "lines")%>%
layout(title = 'Stock Price for CocaCola & P$G over time',
       xaxis = list(title = "Date", zeroline = FALSE),
       yaxis = list(side = 'left', title = 'Stock Price'))
fig
```

```
## Warning: `arrange_()` is deprecated as of dplyr 0.7.0.
## Please use `arrange()` instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```



Answer: From the interactive graph, in year 1983 CocaCola is going up and P&G is going down

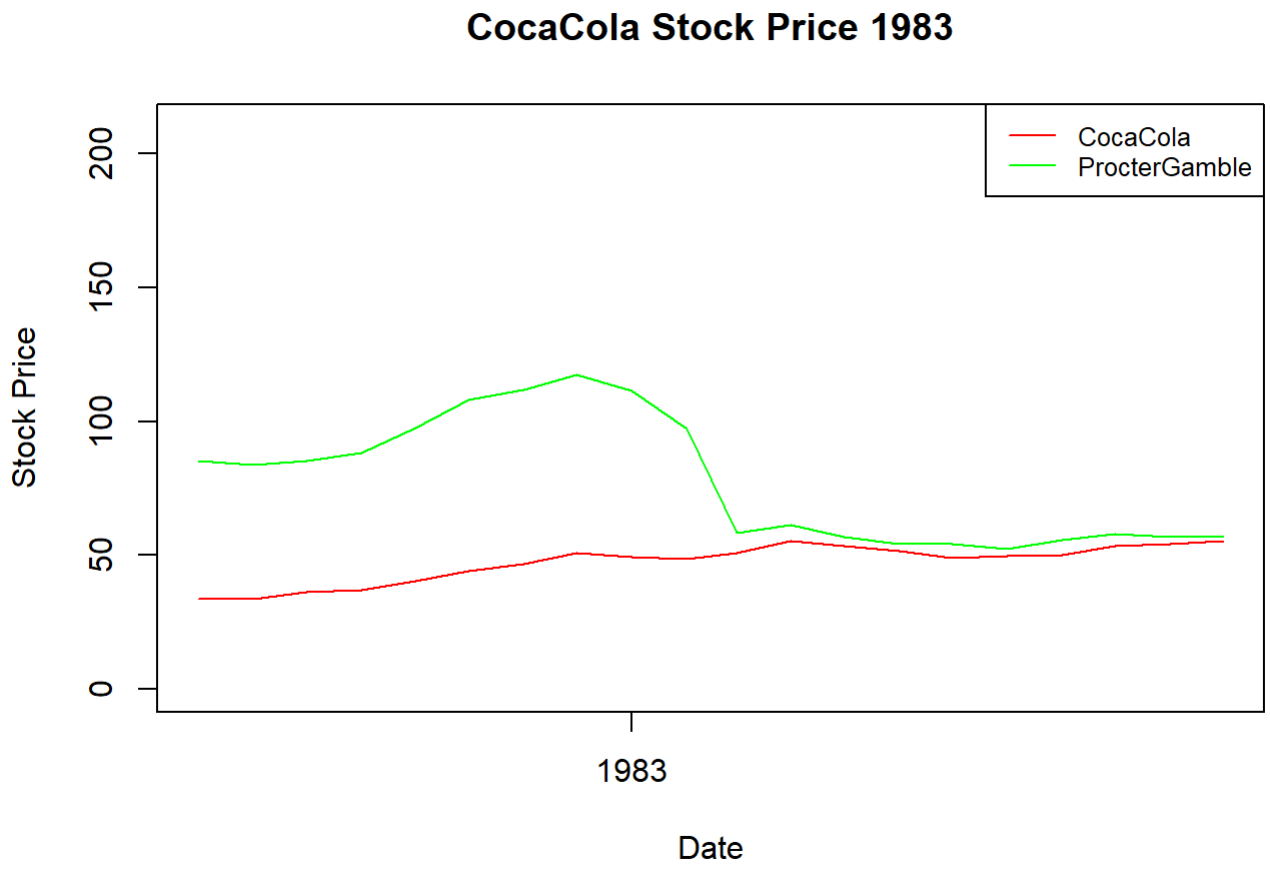
```
which(CocaCola$Date == "1982-01-01")
```

```
## [1] 145
```

```
which(CocaCola$Date == "1983-12-01")
```

```
## [1] 168
```

```
plot(CocaCola$Date[149:168], CocaCola$StockPrice[149:168], type="l", xlab = "Date",
     ylab="Stock Price", main="CocaCola Stock Price 1983", col="red", ylim=c(0,210))
lines(ProcterGamble$Date[149:168],ProcterGamble$StockPrice[149:168],col="green")
legend("topright",legend=c("CocaCola","ProcterGamble"),col=c("red","green"),
       lty=1, cex=0.8)
```



3. Across the entire time period shown in your plot which stock had a generally lower price?

Answer: From the interactive graph, CocaCola had a generally lower price.

Data Visualization from 1995-2005:

First stock price of the year 1995 sits in row position: 301

```
which(CocaCola$Date == "1995-01-01")
```

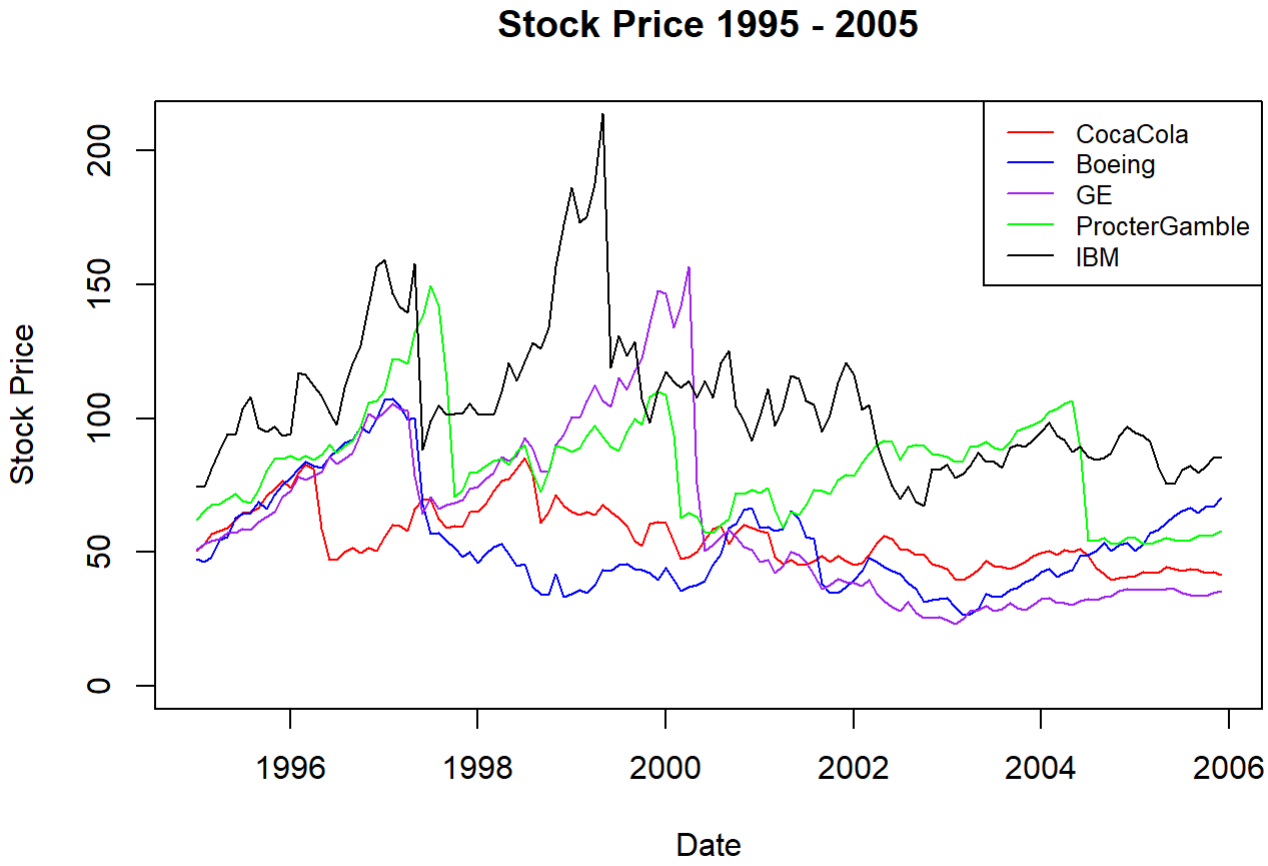
```
## [1] 301
```

Last stock price of the year 2005 sits in row position: 432

```
which(CocaCola$Date == "2005-12-01")
```

```
## [1] 432
```

```
plot(CocaCola$Date[301:432], CocaCola$StockPrice[301:432], type="l",
     xlab = "Date",ylab="Stock Price", main="Stock Price 1995 - 2005", col="red",
     ylim=c(0,210))
lines(Boeing$Date[301:432],Boeing$StockPrice[301:432],col="blue")
lines(GE$Date[301:432],GE$StockPrice[301:432],col="purple")
lines(ProcterGamble$Date[301:432],ProcterGamble$StockPrice[301:432],col="green")
lines(IBM$Date[301:432],IBM$StockPrice[301:432],col="black")
legend("topright",legend=c("CocaCola","Boeing","GE","ProcterGamble","IBM"),
       col=c("red","blue","purple","green","black"), lty=1, cex=0.8)
```



####

1. Which stock price fell the most right after the tech bubble of March 2000?

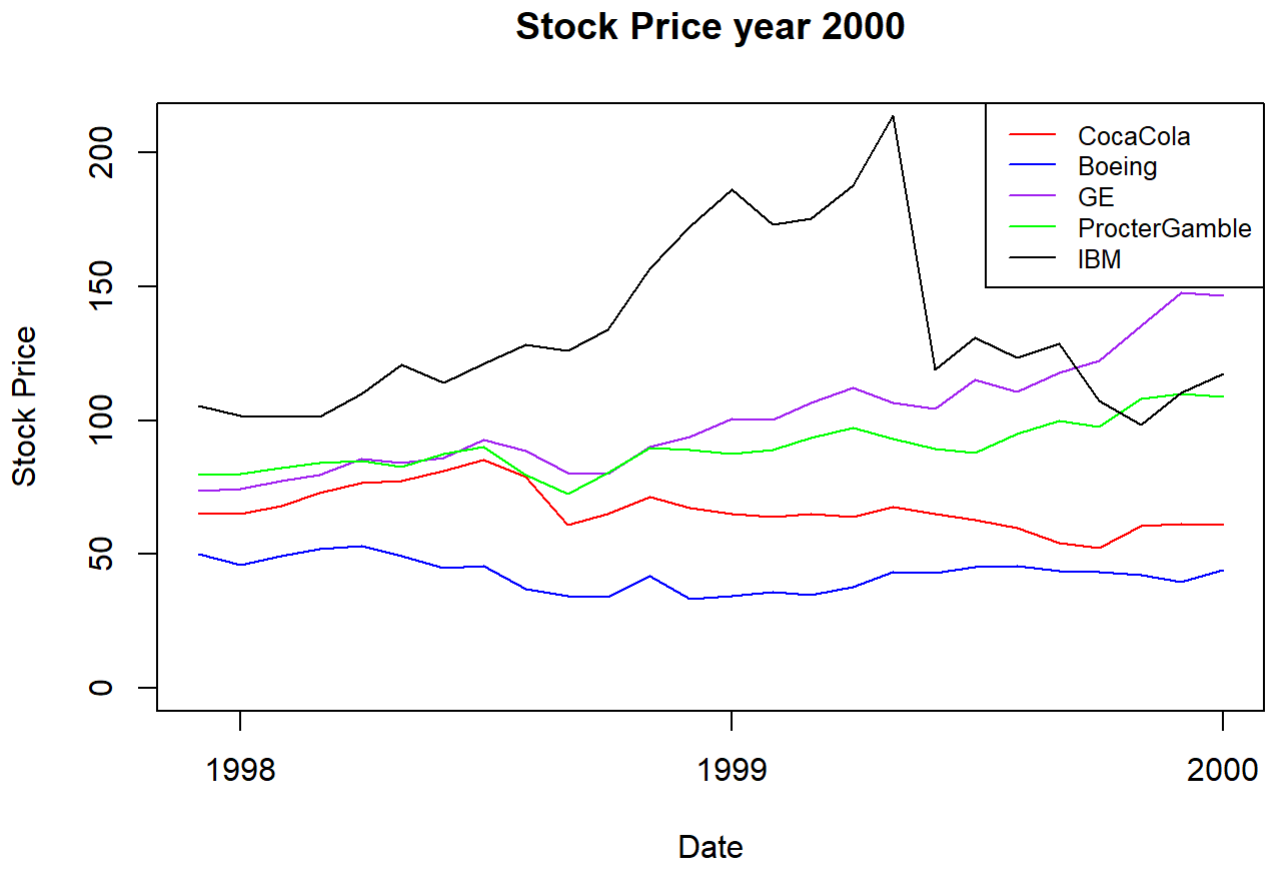
```
which(CocaCola$Date == "2000-01-01")
```

```
## [1] 361
```

```
which(CocaCola$Date == "2000-12-01")
```

```
## [1] 372
```

```
plot(CocaCola$Date[361:336], CocaCola$StockPrice[361:336], type="l",
     xlab = "Date",ylab="Stock Price", main="Stock Price year 2000",
     col="red", ylim=c(0,210))
lines(Boeing$Date[361:336],Boeing$StockPrice[361:336],col="blue")
lines(GE$Date[361:336],GE$StockPrice[361:336],col="purple")
lines(ProcterGamble$Date[361:336],ProcterGamble$StockPrice[361:336],col="green")
lines(IBM$Date[361:336],IBM$StockPrice[361:336],col="black")
legend("topright",legend=c("CocaCola","Boeing","GE","ProcterGamble","IBM"),
       col=c("red","blue","purple","green","black"), lty=1, cex=0.8)
```



Answer: From the graph it is apparent that GE fell the most right after the tech bubble of march 2000

2. What stock had the highest maximum price between 1995-2005?

Answer: IBM

3. A few years before the tech bubble of 1997, there was another stock market crash trigged by economic crisis in Asia in October of 1997. If you compare stock prices from

September 1997 to November 1997, which companies saw a decrease in price? Which company experienced the biggest decrease?

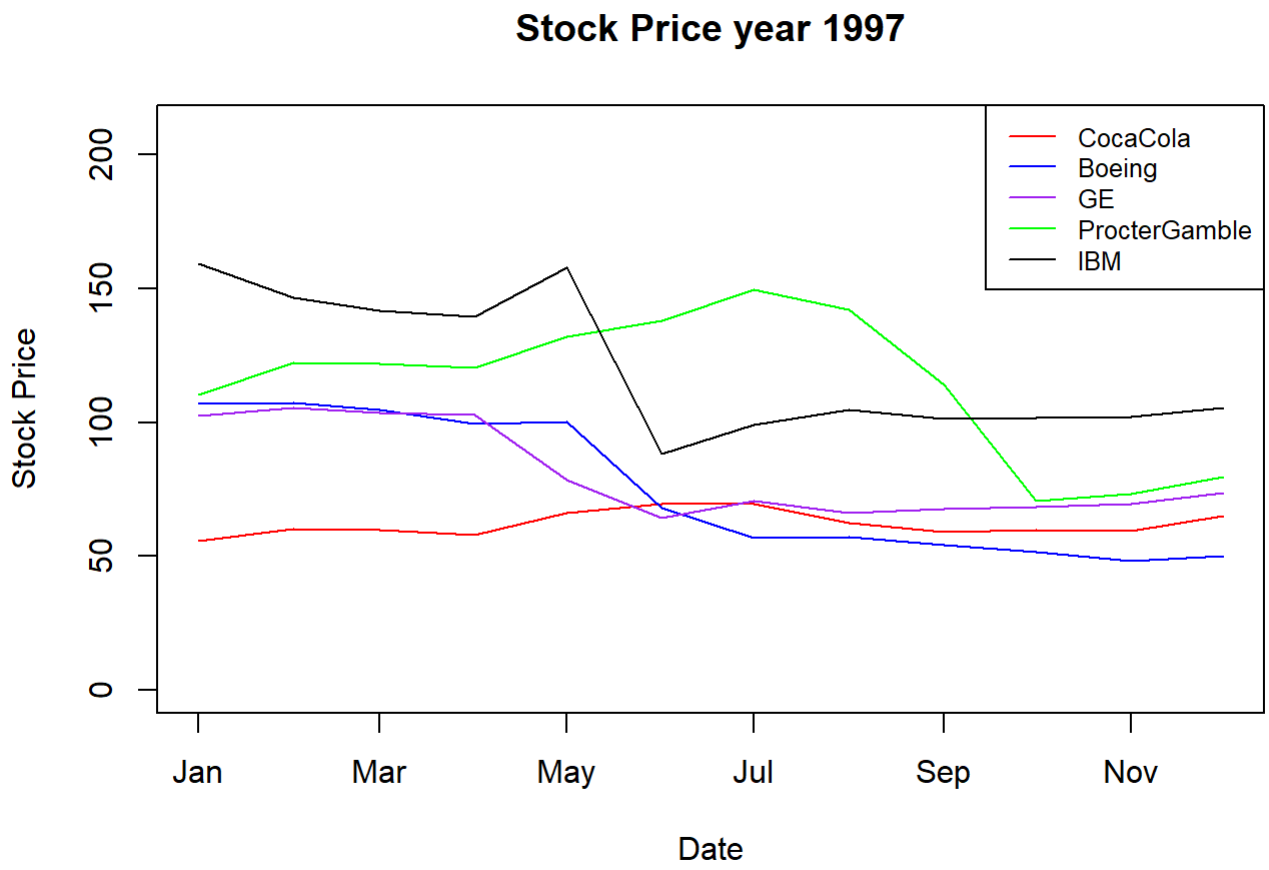
```
which(CocaCola$Date == "1997-01-01")
```

```
## [1] 325
```

```
which(CocaCola$Date == "1997-12-01")
```

```
## [1] 336
```

```
plot(CocaCola$Date[325:336], CocaCola$StockPrice[325:336], type="l",
     xlab = "Date",ylab="Stock Price", main="Stock Price year 1997",
     col="red", ylim=c(0,210))
lines(Boeing$Date[325:336],Boeing$StockPrice[325:336],col="blue")
lines(GE$Date[325:336],GE$StockPrice[325:336],col="purple")
lines(ProcterGamble$Date[325:336],ProcterGamble$StockPrice[325:336],
      col="green")
lines(IBM$Date[325:336],IBM$StockPrice[325:336],col="black")
legend("topright",legend=c("CocaCola","Boeing","GE","ProcterGamble","IBM"),
       col=c("red","blue","purple","green","black"), lty=1, cex=0.8)
```



```
CocaCola_pct_chng = CocaCola[335,]-CocaCola[333,]
IBM_pct_chng = IBM[335,]-IBM[333,]
GE_pct_chng = GE[335,]-GE[333,]
PG_pct_chng = ProcterGamble[335,]-ProcterGamble[333,]
Boeing_pct_chng = Boeing[335,]-Boeing[333,]

CocaCola_pct_chng
```

```
##        Date StockPrice
## 335 61 days 0.09275689
```

```
IBM_pct_chng
```

```
##        Date StockPrice
## 335 61 days  0.7346116
```

```
GE_pct_chng
```

```
##        Date StockPrice
## 335 61 days   1.935213
```

```
PG_pct_chng
```

```
##        Date StockPrice
## 335 61 days  -40.65787
```

```
Boeing_pct_chng
```

```
##        Date StockPrice
## 335 61 days  -5.759524
```

which companies saw a decrease in price?

Answer: Boeing , and P&G

Which company experienced the biggest decrease?

Answer: P&G

4. Which stock seemed to provide the best return (i.e. increase in price) between 2004-2005?

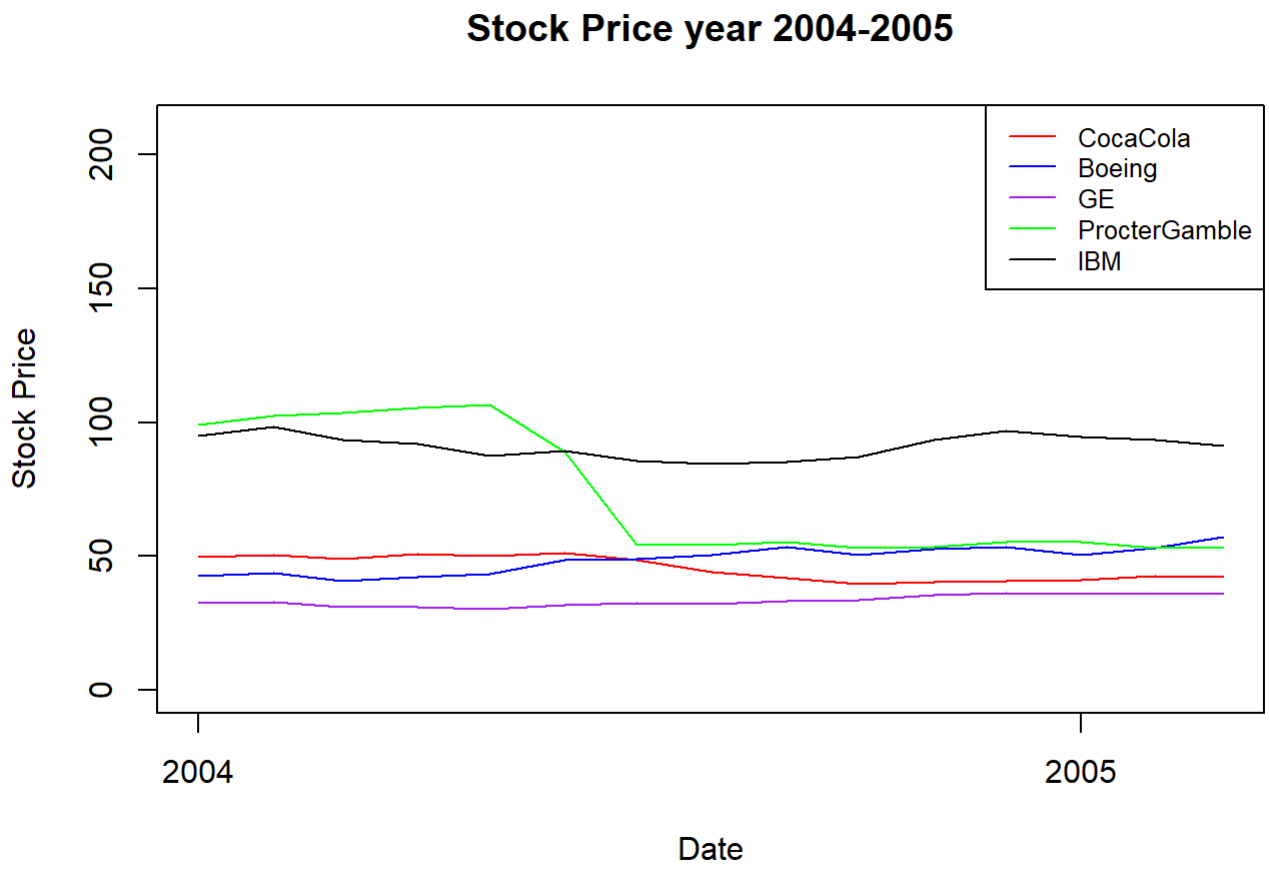
```
which(CocaCola$Date == "2004-01-01")
```

```
## [1] 409
```

```
which(CocaCola$Date == "2005-01-01")
```

```
## [1] 421
```

```
plot(CocaCola$Date[409:423], CocaCola$StockPrice[409:423], type="l",
     xlab = "Date",ylab="Stock Price", main="Stock Price year 2004-2005", col="red",
     ylim=c(0,210))
lines(Boeing$Date[409:423],Boeing$StockPrice[409:423],col="blue")
lines(GE$Date[409:423],GE$StockPrice[409:423],col="purple")
lines(ProcterGamble$Date[409:423],ProcterGamble$StockPrice[409:423],col="green")
lines(IBM$Date[409:423],IBM$StockPrice[409:423],col="black")
legend("topright",legend=c("CocaCola","Boeing","GE","ProcterGamble","IBM"),
       col=c("red","blue","purple","green","black"), lty=1, cex=0.8)
```



Answer: Boeing

5. Between 1995-2005, which company had the biggest delta between the maximum and minimum stock price?

```
which(CocaCola$Date == "1995-01-01")
```

```
## [1] 301
```

```
which(CocaCola$Date == "2005-01-01")
```

```
## [1] 421
```

```
Boeing_delta = max(Boeing[301:432,]$StockPrice)-min(Boeing[301:432,]$StockPrice)
GE_delta = max(GE[301:432,]$StockPrice)-min(GE[301:432,]$StockPrice)
IBM_delta = max(IBM[301:432,]$StockPrice)-min(IBM[301:432,]$StockPrice)
PG_delta = max(ProcterGamble[301:432,]$StockPrice)-min(ProcterGamble[301:432,]$StockPrice)
CocaCola_delta = max(CocaCola[301:432,]$StockPrice)-min(CocaCola[301:432,]$StockPrice)

cat("Boeing Delta = ", Boeing_delta,"\n")
```

```
## Boeing Delta =  80.66905
```

```
cat("GE Delta =",GE_delta,"\n")
```

```
## GE Delta = 133.7779
```

```
cat("IBM Delta =",IBM_delta,"\n")
```

```
## IBM Delta = 146.7631
```

```
cat("PG Delta =",PG_delta,"\n")
```

```
## PG Delta = 96.62526
```

```
cat("CocaCola Delta =",CocaCola_delta,"\n")
```

```
## CocaCola Delta = 45.67552
```

Answer: IBM = 146.7631 had the highest delta between the maximum and minimum stock price

6. Which two companies' stock price seem to be the most correlated (i.e. move up/down together)?

```
cat("CocaCola & IBM Correlation = ", cor(CocaCola$StockPrice, IBM$StockPrice),"\n")
```

```
## CocaCola & IBM Correlation =  0.7227479
```

```
cat("CocaCola & GE Correlation = ", cor(CocaCola$StockPrice, GE$StockPrice),"\n")
```

```
## CocaCola & GE Correlation =  0.1775435
```

```
cat("CocaCola & P&G Correlation = ", cor(CocaCola$StockPrice, ProcterGamble$StockPrice),"\n")
```

```
## CocaCola & P&G Correlation =  0.3320132
```

```
cat("CocaCola & Boeing Correlation = ", cor(CocaCola$StockPrice, Boeing$StockPrice),"\n")
```

```
## CocaCola & Boeing Correlation =  -0.3305256
```

```
cat("IBM & GE Correlation = ", cor(IBM$StockPrice, GE$StockPrice),"\n")
```

```
## IBM & GE Correlation =  0.1098373
```

```
cat("IBM & P&G Correlation = ", cor(IBM$StockPrice, ProcterGamble$StockPrice),"\n")
```

```
## IBM & P&G Correlation =  0.3163998
```

```
cat("IBM & Boeing Correlation = ", cor(IBM$StockPrice, Boeing$StockPrice),"\n")
```

```
## IBM & Boeing Correlation =  -0.3675284
```

```
cat("GE & P&G Correlation = ", cor(GE$StockPrice, ProcterGamble$StockPrice),"\n")
```

```
## GE & P&G Correlation =  0.1865324
```

```
cat("GE & Boeing Correlation = ", cor(GE$StockPrice, Boeing$StockPrice),"\n")
```

```
## GE & Boeing Correlation =  -0.09634328
```

```
cat("P&G & Boeing Correlation = ", cor(ProcterGamble$StockPrice,Boeing$StockPrice),"\n")
```

```
## P&G & Boeing Correlation =  -0.0862052
```

Answer: CocaCola & IBM (correlation = 0.7227479) seem to be the most correlated

Monthly Trend Analysis

1. For IBM, compare the average stock price for each month to the its overall average stock price and identify all the months for which IBM historically had a higher stock price (we call this over- indexing)? Which month over-indexed the most?

```
IBM_monthly_mean = tapply(IBM$StockPrice, months(IBM$Date), mean)
IBM_monthly_mean
```

```
##     April    August  December  February   January      July      June     March
##  152.1168  140.1455  140.7593  152.6940  150.2384  139.0670  139.0907  152.4327
##       May  November   October September
##  151.5022  138.0187  137.3466  139.0885
```

identify all the months for which IBM historically had a higher stock price?

```
sort(IBM_monthly_mean)
```

```
##   October  November      July September      June    August  December   January
##  137.3466  138.0187  139.0670  139.0885  139.0907  140.1455  140.7593  150.2384
##       May     April     March  February
##  151.5022  152.1168  152.4327  152.6940
```

Answer: January, February, March, April, May

Which month over-indexed the most?

```
max(IBM_monthly_mean)
```

```
## [1] 152.694
```

Answer: February = 152.6940

2. Repeat the tapply() function you used to solve the last question for each of the 4 remaining companies. Do any of two or more companies have their highest stock price in the same months as each other? Which companies and months does this happen for?

```
Boeing_monthly_mean = tapply(Boeing$StockPrice, months(Boeing$Date), mean)
sort(IBM_monthly_mean)
```

```
##   October  November      July September      June    August  December   January
##  137.3466  138.0187  139.0670  139.0885  139.0907  140.1455  140.7593  150.2384
##       May     April     March  February
##  151.5022  152.1168  152.4327  152.6940
```

```
GE_monthly_mean = tapply(GE$StockPrice, months(GE$Date), mean)
sort(GE_monthly_mean)
```

```
##   October September      June    August      July  November  December       May
##  56.23897  56.23913  56.46844  56.50315  56.73349  57.28879  59.10217  60.87135
##   January  February     March     April
##  62.04511  62.52080  63.15055  64.48009
```

```
CocaCola_monthly_mean = tapply(CocaCola$StockPrice, months(CocaCola$Date), mean)
sort(CocaCola_monthly_mean)
```

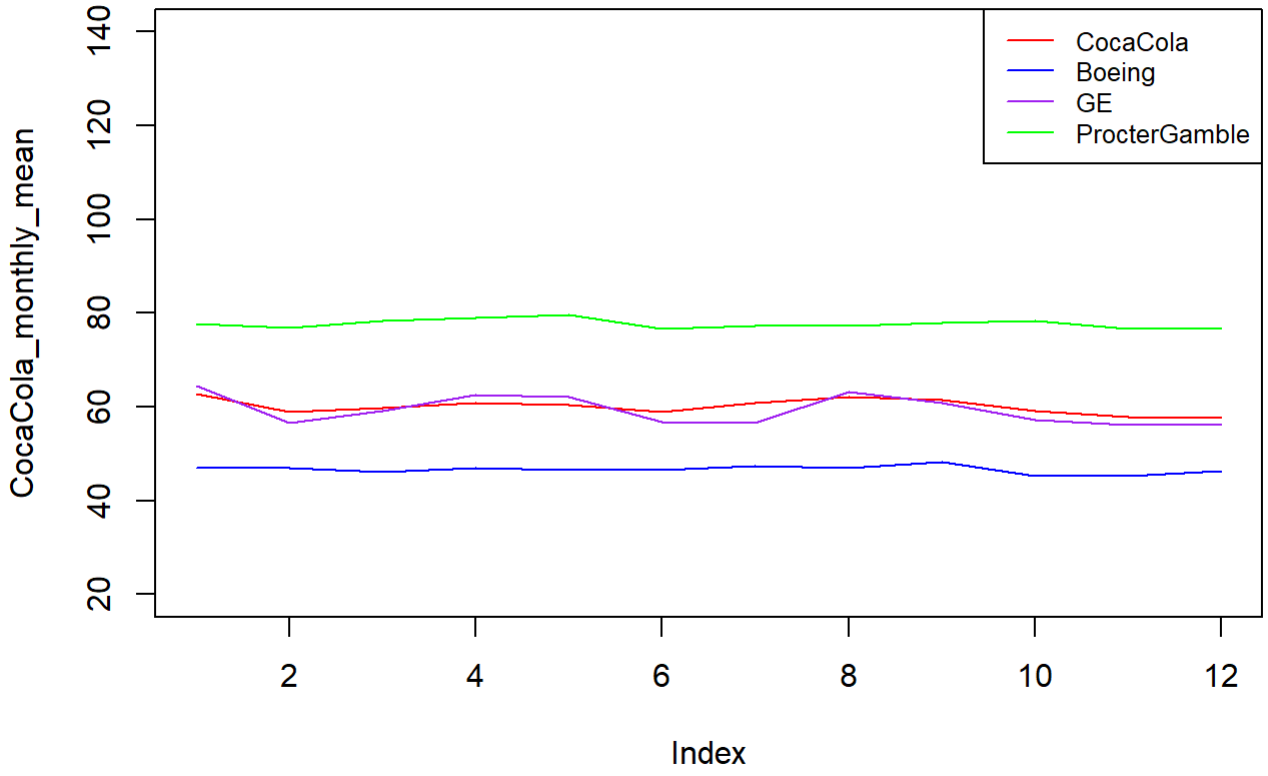
```
## September   October    August      July  November  December   January  February
##  57.60024  57.93887  58.88014  58.98346  59.10268  59.73223  60.36849  60.73475
##      June       May     March     April
##  60.81208  61.44358  62.07135  62.68888
```

```
ProcterGamble_monthly_mean = tapply(ProcterGamble$StockPrice, months(ProcterGamble$Date), mean)
sort(ProcterGamble_monthly_mean)
```

```
## September      July   October    August     March      June     April       May
##  76.62385  76.64556  76.67903  76.82266  77.34761  77.39275  77.68671  77.85958
##  December  November  February   January
##  78.29661  78.45610  79.02575  79.61798
```

2. Do any of two or more companies have their highest stock price in the same months as each other?

```
plot(CocaCola_monthly_mean, type = "l", col = "red", ylim = c(20,140))
lines(GE_monthly_mean, col ="purple")
lines(Boeing_monthly_mean, col ="blue")
lines(ProcterGamble_monthly_mean, col ="green")
legend("topright",legend=c("CocaCola","Boeing","GE","ProcterGamble"),
       col=c("red","blue","purple","green"), lty=1, cex=0.8)
```



2. Do any of two or more companies have their highest stock price in the same months as each other?Which companies and months does this happen for?

Answer: GE & CocaCola in the month of April

3. What trend do you see for the months of December vs January for each company? Is there an over-arching trend that applies to all companies when comparing all historical December vs January stock prices?

Answer: Yes, all stock prices have an over-arching trend which applies to all companies

by looking at the mean values from December & January