R Notebook

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.2
```

```
## -- Attaching packages ---------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.0     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
## Warning: package 'tibble' was built under R version 3.4.4
```

```
## Warning: package 'tidyr' was built under R version 3.4.3
```

```
## Warning: package 'purrr' was built under R version 3.4.4
```

```
## Warning: package 'dplyr' was built under R version 3.4.4
```

```
## Warning: package 'forcats' was built under R version 3.4.3
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(skimr)
```

```
## Warning: package 'skimr' was built under R version 3.4.4
```

```
library(readr)
library(dplyr)
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

Read in all the required files

```
IBM = read_csv("IBMStock.csv")
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   StockPrice = col_double()
## )
```

```
GE = read_csv("GEStock.csv")
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   StockPrice = col_double()
```

```
## )
```

```
ProcterGamble = read_csv("ProcterGambleStock.csv")
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   StockPrice = col_double()
## )
```

```
CocaCola = read_csv("CocaColaStock.csv")
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   StockPrice = col_double()
## )
```

```
Boeing = read_csv("BoeingStock.csv")
```

```
## Parsed with column specification:
## cols(
##   Date = col_character(),
##   StockPrice = col_double()
## )
```

1.1. Summary Statistics

Convert the dates into a format that R can understand

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

```
## Warning in strptime(x, format, tz = "GMT"): unknown timezone 'zone/tz/
## 2018i.1.0/zoneinfo/America/Chicago'
```

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

The earliest year and latest year

```
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
```

```
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(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(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(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
```

Mean stock price of IBM over this time period

```
IBM %>%
  summarize(mean_stock_price = mean(StockPrice))
```

```
## # A tibble: 1 x 1
##   mean_stock_price
##              <dbl>
## 1             144.
```

Minimum stock price of GE over this time period

```
GE %>%
  summarize(min_stock_price = min(StockPrice))
```

```
## # A tibble: 1 x 1
##   min_stock_price
##             <dbl>
## 1            9.29
```

Maximum stock price of Coca-Cola over this time period

```
CocaCola %>%
  summarize(max_stock_price = max(StockPrice))
```

```
## # A tibble: 1 x 1
##   max_stock_price
##             <dbl>
## 1            147.
```

Median stock price of Boeing over this time period

```
Boeing %>%
  summarize(median_stock_price = median(StockPrice))
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
## # A tibble: 1 x 1
##   median_stock_price
##                <dbl>
## 1               44.9
```

Standard deviation of the stock price of PG over this time period

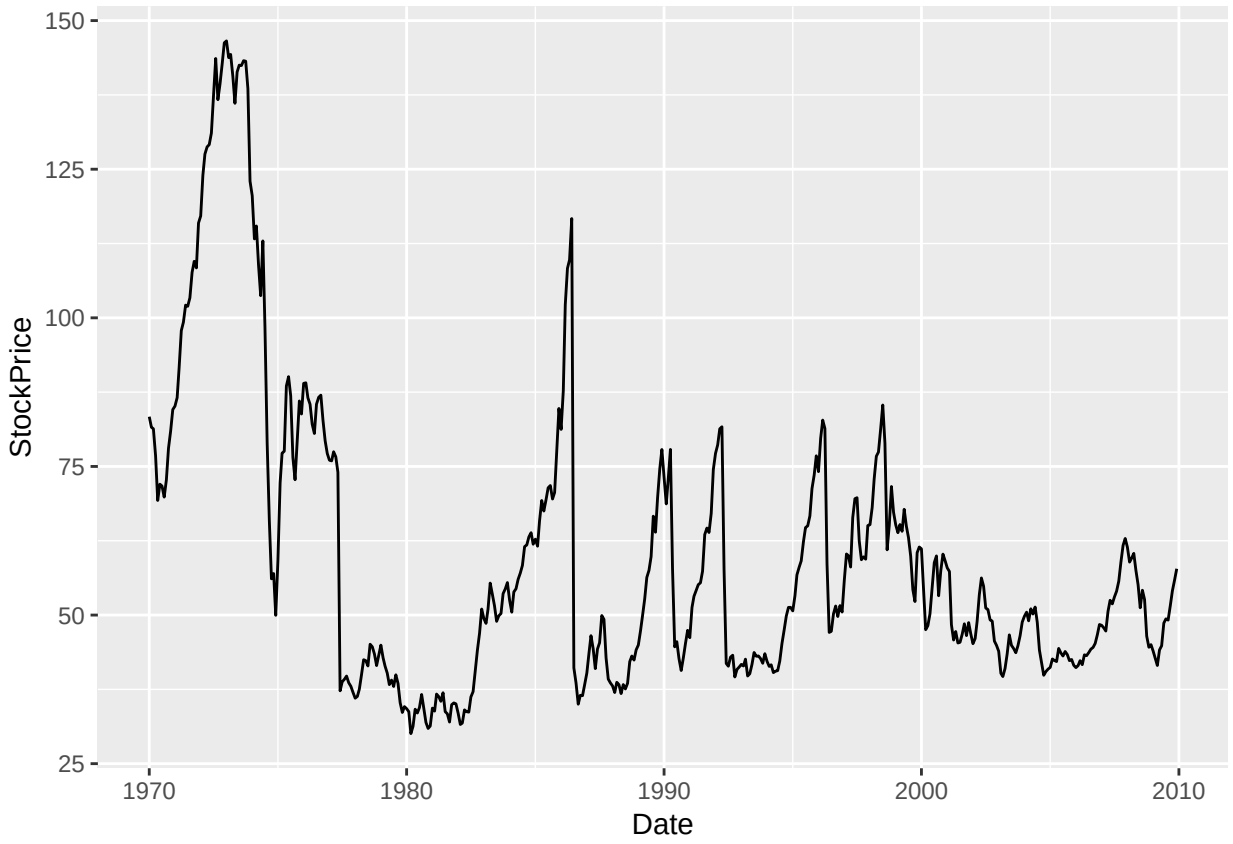
```
ProcterGamble %>%
  summarize(sd_stock_price = sd(StockPrice))
```

```
## # A tibble: 1 x 1
##   sd_stock_price
##            <dbl>
## 1           18.2
```

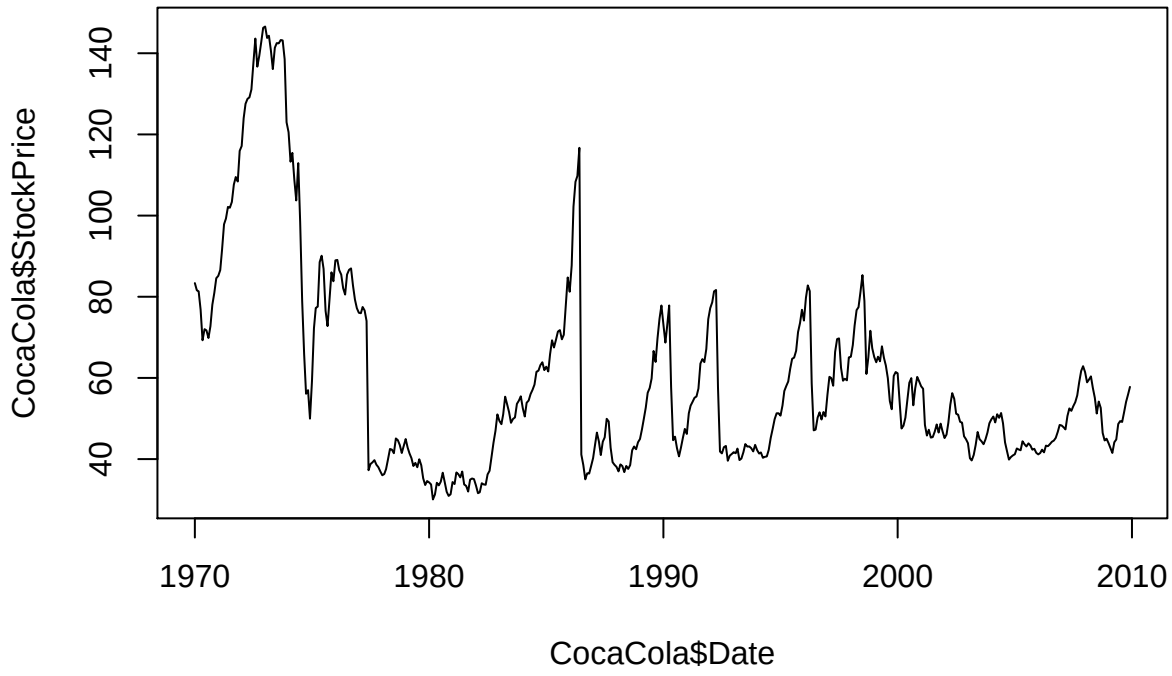
2.1) Visualizing Stock Dynamics

Plot the stock price over time for CocaCola

```
ggplot(CocaCola) +
  geom_line(aes(Date, StockPrice))
```



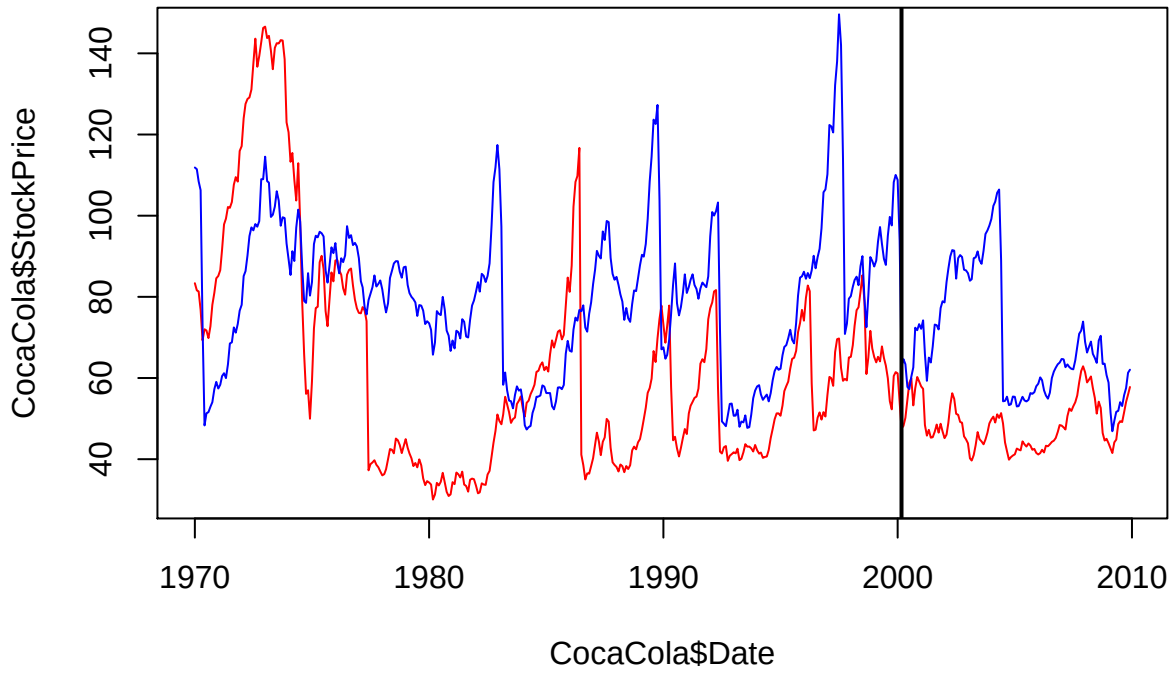

```
plot(CocaCola$Date, CocaCola$StockPrice, type = "l")
```



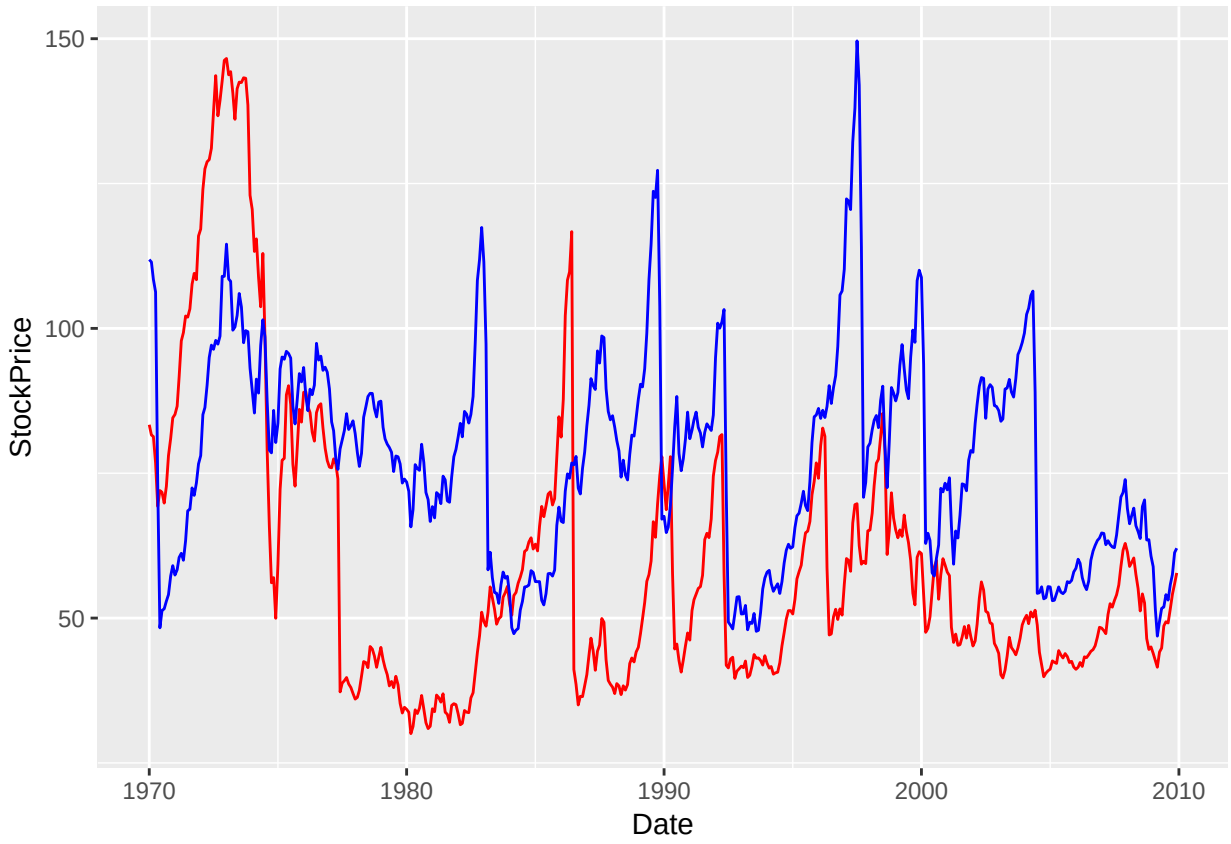
2.1) Visualizing Stock Dynamics

Add a line for P&G

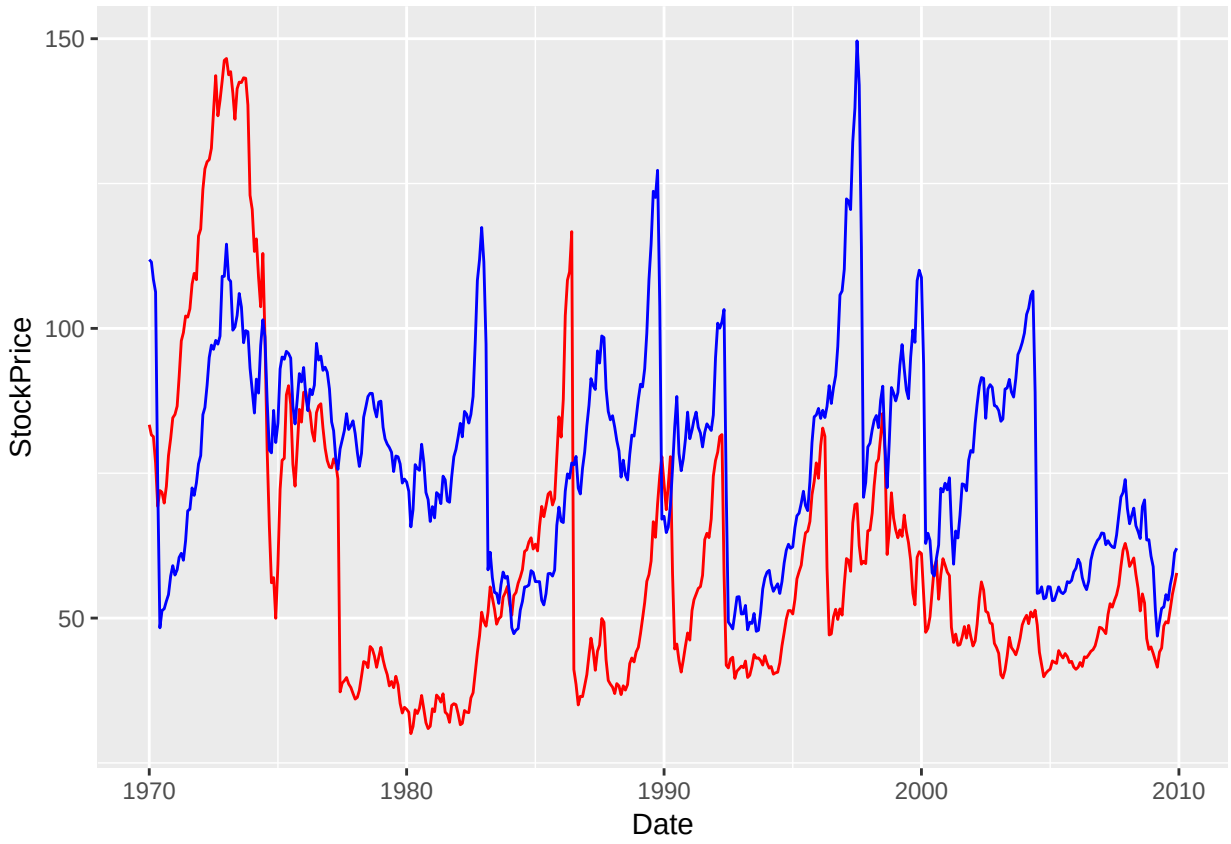
```
plot(CocaCola$Date, CocaCola$StockPrice, type = "l", col = "red")
lines(ProcterGamble$Date, ProcterGamble$StockPrice, col = "blue")
# to see which company's stock dropped more in March of 2000
abline(v = as.Date(c("2000-03-01")), lwd = 2)
```




```
ggplot() +
  geom_line(data = CocaCola, col = "red", aes(Date, StockPrice)) +
  geom_line(data = ProcterGamble, col = "blue", aes(Date, StockPrice))
```

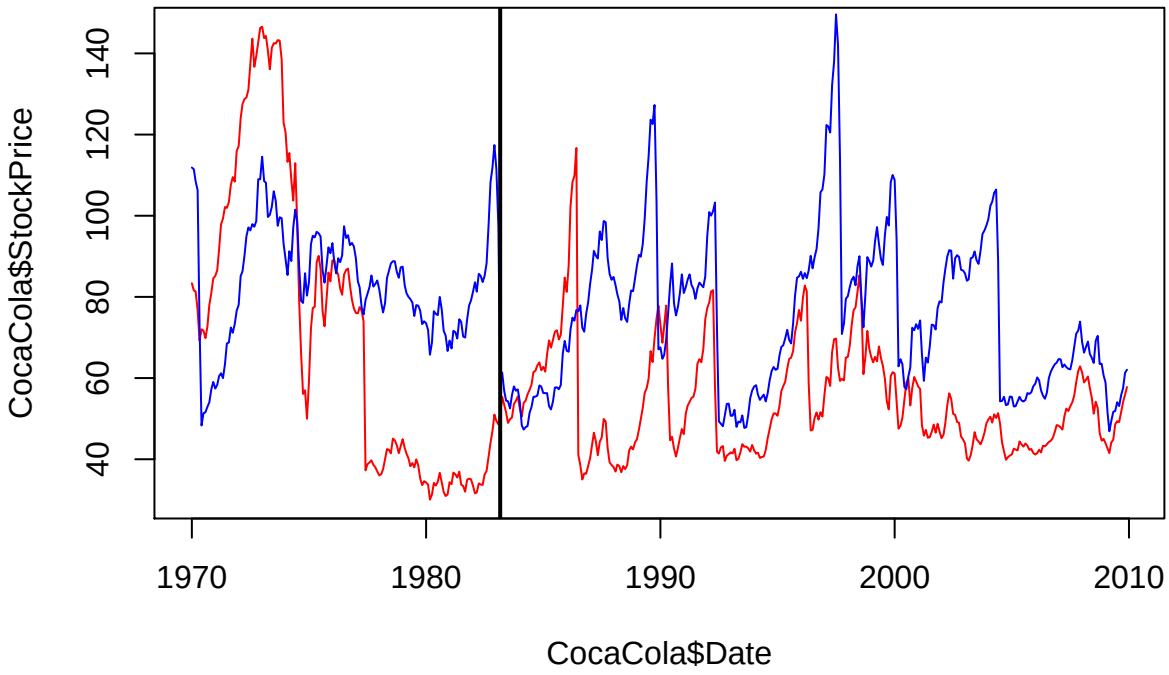



```
plot1 = ggplot() +
  geom_line(data = CocaCola, col = "red", aes(Date, StockPrice))

plot2 = plot1 +
  geom_line(data = ProcterGamble, col = "blue", aes(Date, StockPrice))

plot2
```



Around 1983, the stock for one of these companies (Coca-Cola or Procter and Gamble) was going up, while the other was going down. Which one was going up?

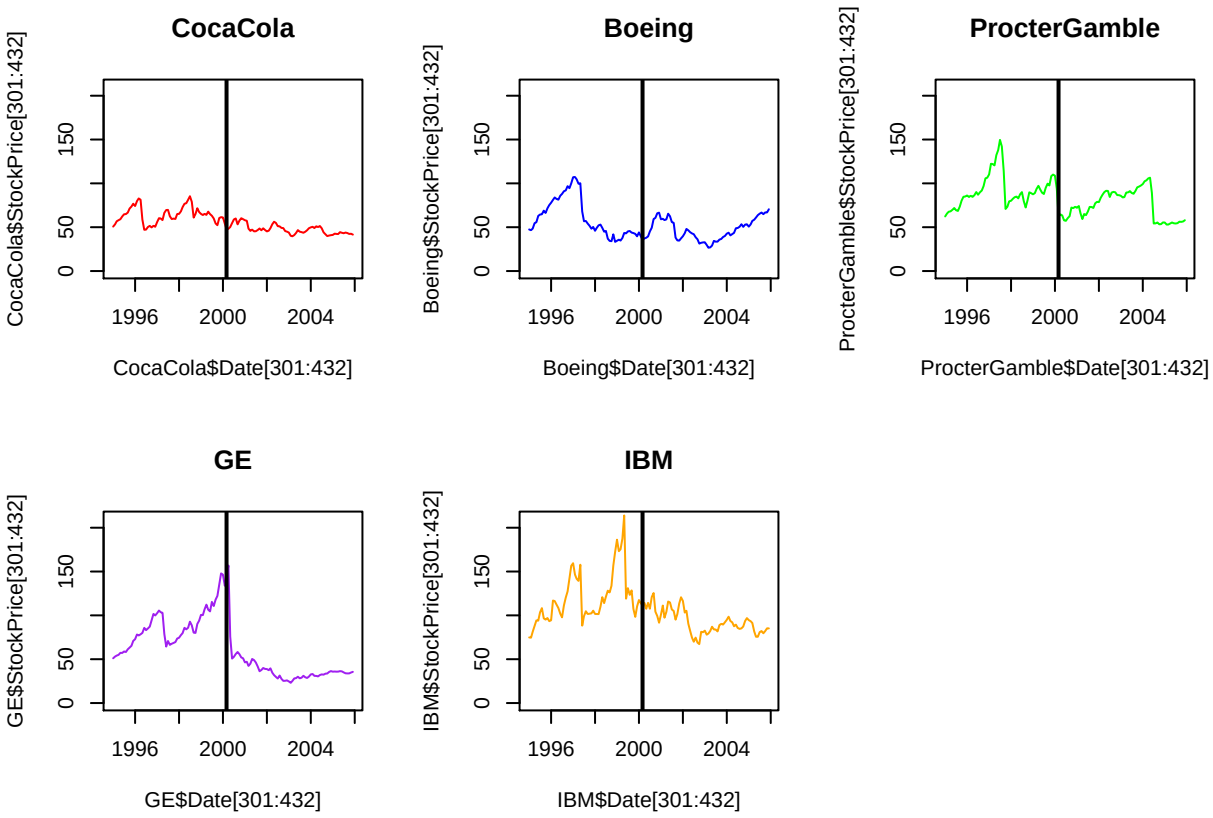
```
plot(CocaCola$Date, CocaCola$StockPrice, type = "l", col = "red")
lines(ProcterGamble$Date, ProcterGamble$StockPrice, col = "blue")
# to see which company's stock dropped more in March of 2000
abline(v = as.Date(c("1983-03-01")), lwd = 2)
```



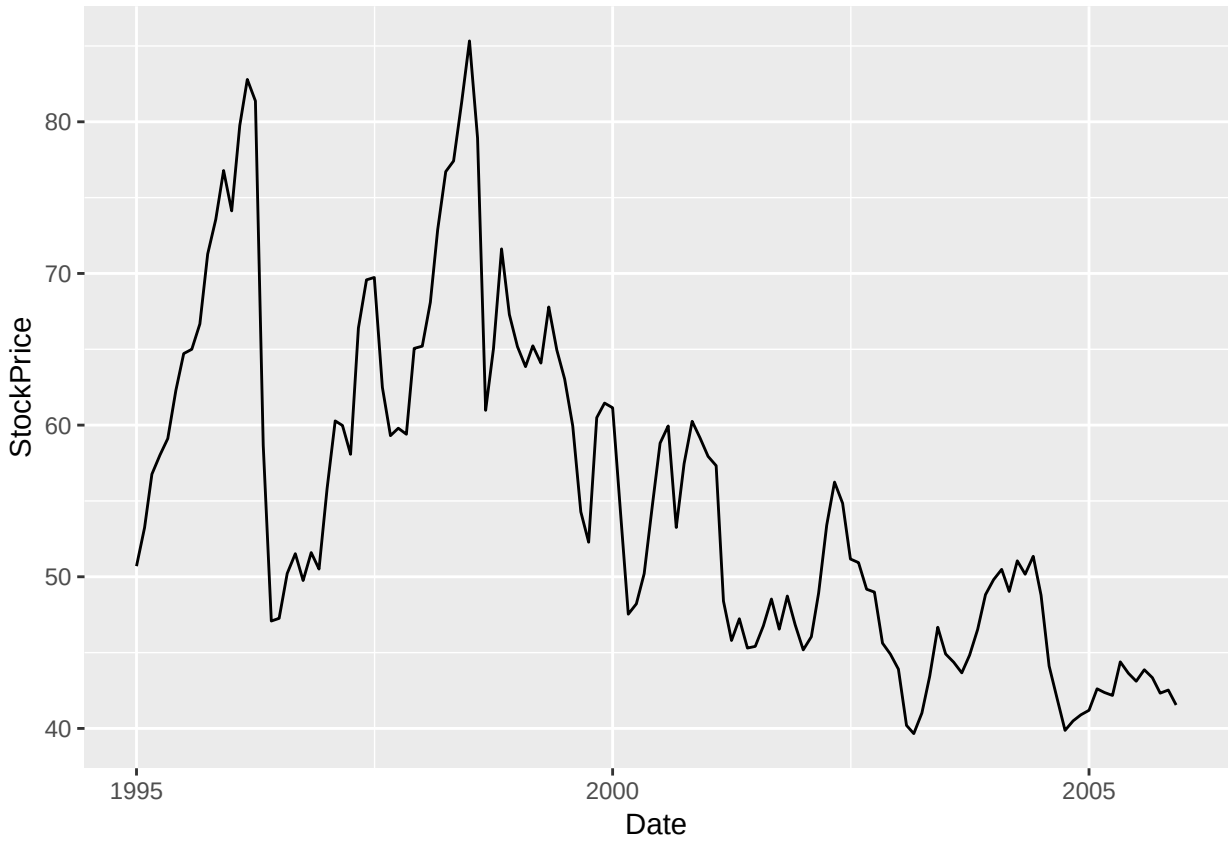
3.1) Viasualizing Stock Dynamics 1995-2005

Let's take a look at how stock prices changed from 1995-2005 for all companies

```
par(mfrow = c(2,3))

plot1 = plot(CocaCola$Date[301:432], CocaCola$StockPrice[301:432], type = "l", col = "red", ylim = c(0,
abline(v = as.Date(c("2000-03-01")), lwd = 2)

plot2 = plot(Boeing$Date[301:432], Boeing$StockPrice[301:432], type = "l", col = "blue", ylim = c(0, 21
abline(v = as.Date(c("2000-03-01")), lwd = 2)


plot3 = plot(ProcterGamble$Date[301:432], ProcterGamble$StockPrice[301:432], type = "l", col = "green",
abline(v = as.Date(c("2000-03-01")), lwd = 2)

plot4 = plot(GE$Date[301:432], GE$StockPrice[301:432], type = "l", col = "purple", ylim = c(0, 210), ma
abline(v = as.Date(c("2000-03-01")), lwd = 2)

plot5 = plot(IBM$Date[301:432], IBM$StockPrice[301:432], type = "l", col = "orange", ylim = c(0, 210), m
abline(v = as.Date(c("2000-03-01")), lwd = 2)

# GE stock fell the most right after the technology burble burst in March 2000
# IBM stock reaches the highest value in the time period 1995-2005
```

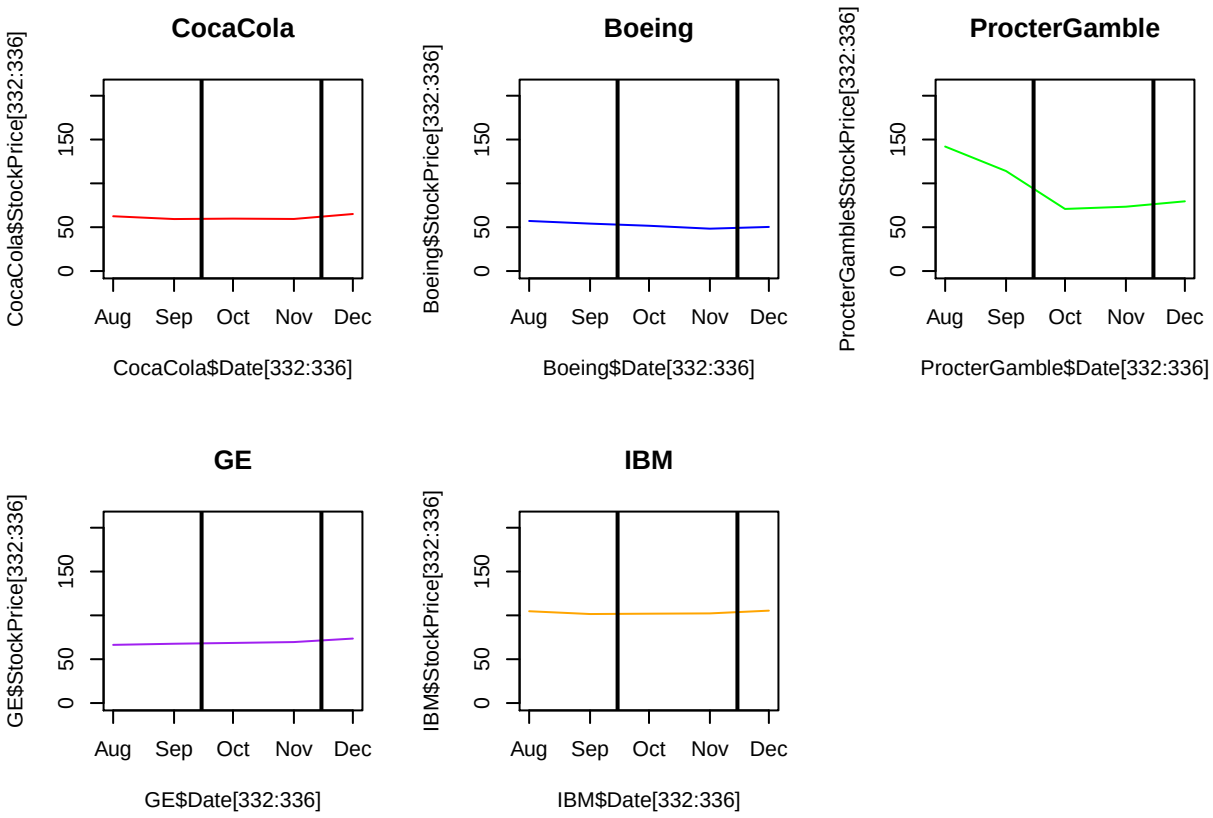



```
CocaCola_1995_2005 = CocaCola[301:432,]
ggplot(CocaCola_1995_2005) +
  geom_line(aes(Date, StockPrice))
```



3.3 In October of 1997, there was a global stock market crash that was caused by an economic crisis in Asia. Comparing September 1997 to November 1997, which companies saw a decreasing trend in their stock price? (Select all that apply.)

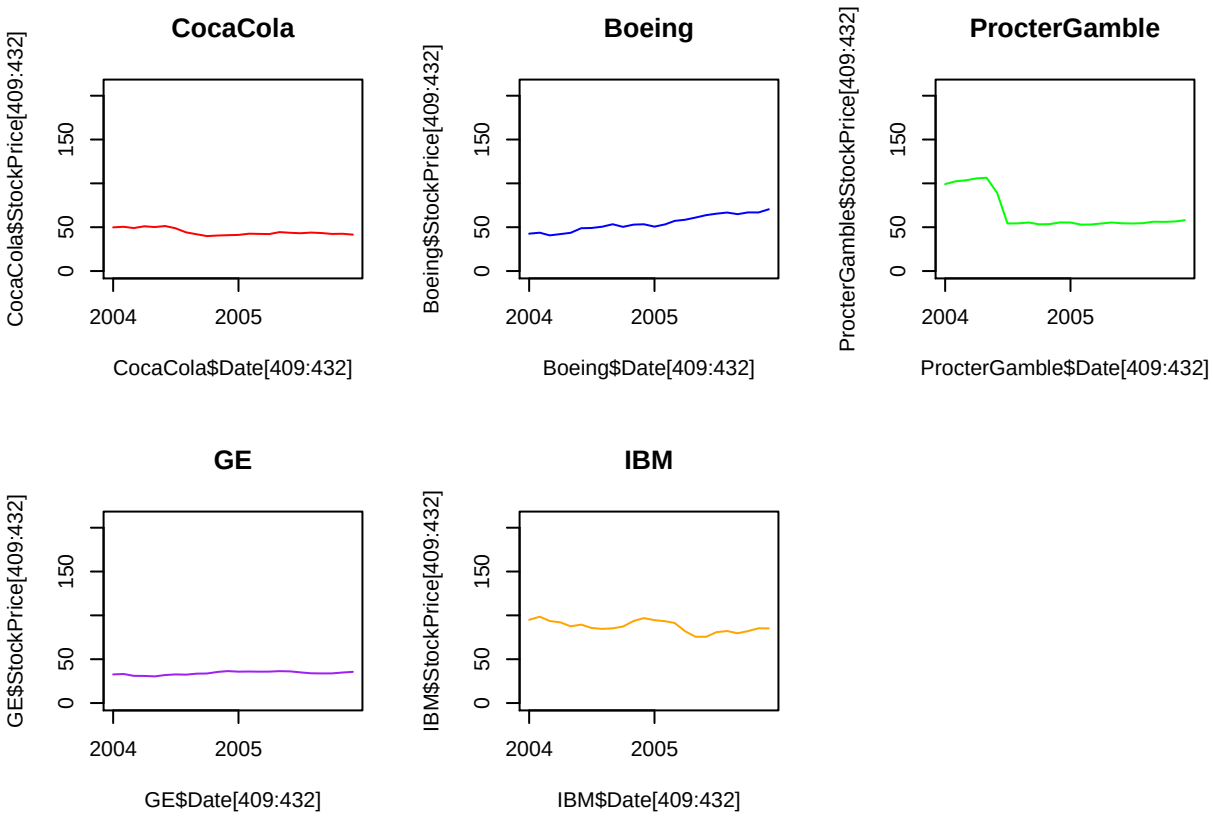
```
par(mfrow = c(2,3))

plot1 = plot(CocaCola$Date[332:336], CocaCola$StockPrice[332:336], type = "l", col = "red", ylim = c(0,
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot2 = plot(Boeing$Date[332:336], Boeing$StockPrice[332:336], type = "l", col = "blue", ylim = c(0, 21
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot3 = plot(ProcterGamble$Date[332:336], ProcterGamble$StockPrice[332:336], type = "l", col = "green",
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot4 = plot(GE$Date[332:336], GE$StockPrice[332:336], type = "l", col = "purple", ylim = c(0, 210), ma
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot5 = plot(IBM$Date[332:336], IBM$StockPrice[332:336], type = "l", col = "orange", ylim = c(0, 210), m
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)
```



3.4

In the last two years of this time period (2004 and 2005) which stock seems to be performing the best, in terms of increasing stock price?

```
par(mfrow = c(2,3))

plot1 = plot(CocaCola$Date[409:432], CocaCola$StockPrice[409:432], type = "l", col = "red", ylim = c(0,
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot2 = plot(Boeing$Date[409:432], Boeing$StockPrice[409:432], type = "l", col = "blue", ylim = c(0, 21
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot3 = plot(ProcterGamble$Date[409:432], ProcterGamble$StockPrice[409:432], type = "l", col = "green",
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot4 = plot(GE$Date[409:432], GE$StockPrice[409:432], type = "l", col = "purple", ylim = c(0, 210), ma
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)

plot5 = plot(IBM$Date[409:432], IBM$StockPrice[409:432], type = "l", col = "orange", ylim = c(0, 210), m
abline(v = as.Date(c("1997-09-15")), lwd = 2)
abline(v = as.Date(c("1997-11-15")), lwd = 2)
```



4.1 Monthly Trends

Lastly, let's see if stocks tend to be higher or lower during certain months. Use the tapply command to calculate the mean stock price of IBM, sorted by months. To sort by months, use

months(IBM$Date)

as the second argument of the tapply function.

For IBM, compare the monthly averages to the overall average stock price. In which months has IBM historically had a higher stock price (on average)?

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

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

```
mean(IBM$StockPrice)
```

```
## [1] 144.375
```

4.2 Repeat the tapply function from the previous problem for each of the other four companies, and use the output to answer the remaining questions.

General Electric and Coca-Cola both have their highest average stock price in the same month. Which month is this? (Answer: April)

```
tapply(Boeing$StockPrice, months(Boeing$Date), mean)
```

```
##     April    August  December  February   January      July      June
##  47.04686  46.86311  46.17315  46.89223  46.51097  46.55360  47.38525
##     March       May  November   October September
##  46.88208  48.13716  45.14990  45.21603  46.30485
```

```
mean(Boeing$StockPrice)
```

```
## [1] 46.59293
```

```
tapply(CocaCola$StockPrice, months(CocaCola$Date), mean)
```

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

```
mean(CocaCola$StockPrice)
```

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

```
tapply(GE$StockPrice, months(GE$Date), mean)
```

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

```
mean(GE$StockPrice)
```

```
## [1] 59.3035
```

```
tapply(ProcterGamble$StockPrice, months(ProcterGamble$Date), mean)
```

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

```
mean(ProcterGamble$StockPrice)
```

```
## [1] 77.70452
```

For the months of December and January, every company's average stock is higher in on emonth and lower in the other. In which month are the stock prices lower?