

Unit1_Assignment_StockDynamics

Anuranjan

23 September 2016

In this problem, we'll take a look at how the stock dynamics of companies: IBM, General Electric (GE), Procter and Gamble, Coca Cola, and Boeing have changed over time. Each data frame has two variables, described as follows:

Date: the date of the stock price, always given as the first of the 'month'.

StockPrice: the average stock price of the company in the given month.

Reading CSV files.

```
IBM = read.csv('IBMStock.csv')
GE = read.csv("GESTock.csv")
ProcterGamble = read.csv("ProcterGambleStock.csv")
CocaCola = read.csv("CocaColaStock.csv")
Boeing = read.csv("BoeingStock.csv")
summary(IBM)
```

```
##           Date           StockPrice
## 1/1/00 : 1   Min.      : 43.40
## 1/1/01 : 1   1st Qu.: 88.34
## 1/1/02 : 1   Median :112.11
## 1/1/03 : 1   Mean    :144.38
## 1/1/04 : 1   3rd Qu.:165.41
## 1/1/05 : 1   Max.     :438.90
## (Other):474
```

```
str(IBM)
```

```
## 'data.frame':    480 obs. of  2 variables:
## $ Date          : Factor w/ 480 levels "1/1/00","1/1/01",...: 11 171 211 251 29
1 331 371 411 451 51 ...
## $ StockPrice: num  360 347 327 320 270 ...
```

```
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
```

Since Date variable is stored as factor,
converting it to Date object.

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

```
## 'data.frame':   480 obs. of  2 variables:
##  $ Date      : Date, format: "1970-01-01" "1970-02-01" ...
##  $ StockPrice: num  360 347 327 320 270 ...
```

```
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
```

STR() function shows us that there are
480 observations.

Earliest year in our dataset.

```
c(min(IBM$Date), min(GE$Date), min(ProcterGamble$Date), min(CocaCola$Date), min(Boeing$Date))
```

```
## [1] "1970-01-01" "1970-01-01" "1970-01-01" "1970-01-01" "1970-01-01"
```

Latest year in our dataset.

```
c(max(IBM$Date), max(GE$Date), max(ProcterGamble$Date), max(CocaCola$Date) , ma
x(Boeing$Date))
```

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

Mean Stock Price

```
c(mean(IBM$StockPrice), mean(GE$StockPrice), mean(ProcterGamble$StockPrice), me
an(CocaCola$StockPrice) , mean(Boeing$StockPrice))
```

```
## [1] 144.37503 59.30350 77.70452 60.02973 46.59293
```

Max Stock Price

```
c(max(IBM$StockPrice), max(GE$StockPrice), max(ProcterGamble$StockPrice), max(C
ocaCola$StockPrice) , max(Boeing$StockPrice))
```

```
## [1] 438.9016 156.8437 149.6200 146.5843 107.2800
```

Min Stock Price

```
c(min(IBM$StockPrice), min(GE$StockPrice), min(ProcterGamble$StockPrice), min(C
ocaCola$StockPrice) , min(Boeing$StockPrice))
```

```
## [1] 43.395000 9.293636 46.884545 30.057143 12.736364
```

Median Stock Price

```
c(median(IBM$StockPrice), median(GE$StockPrice), median(ProcterGamble$StockPric
e), median(CocaCola$StockPrice) , median(Boeing$StockPrice))
```

```
## [1] 112.11460 55.81205 78.33608 51.43699 44.88340
```

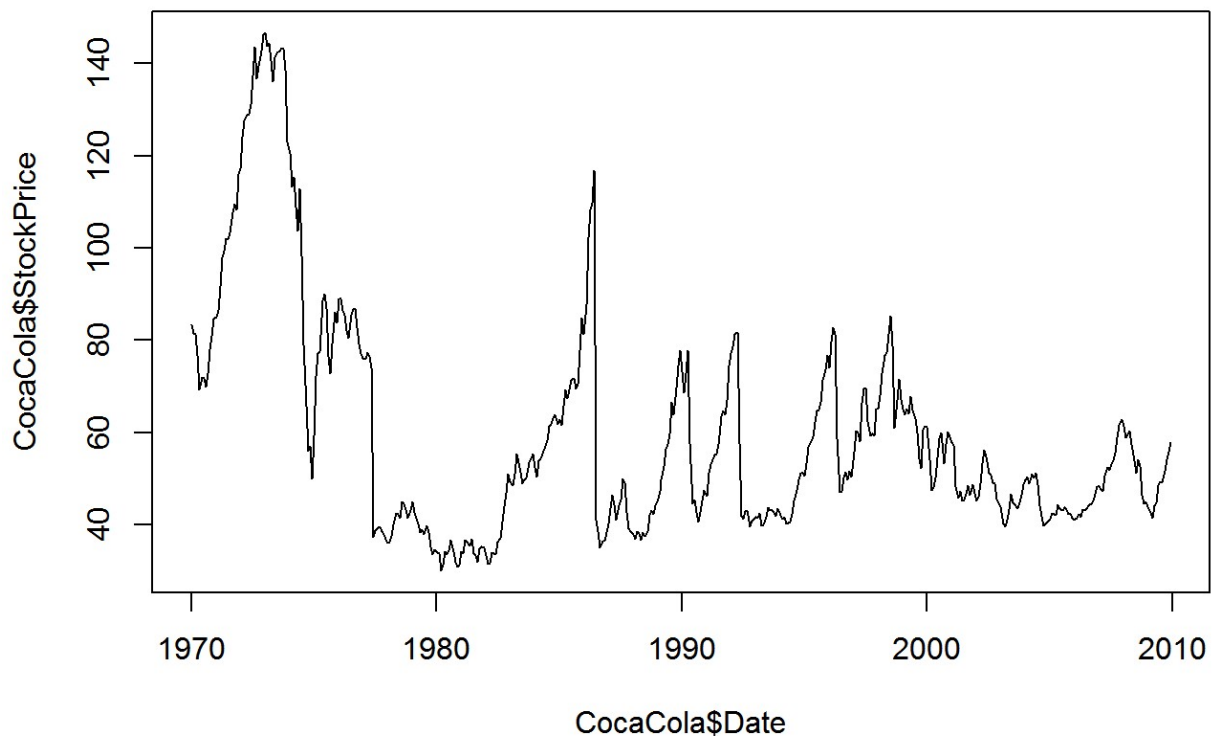
Std Dev of Stock Price

```
c(sd(IBM$StockPrice), sd(GE$StockPrice), sd(ProcterGamble$StockPrice), sd(CocaCola$StockPrice) , sd(Boeing$StockPrice))
```

```
## [1] 87.82208 23.99255 18.19414 25.16629 19.89184
```

Generating plot for CocaCola. Adding argument `type="l"` generates line instead of points.

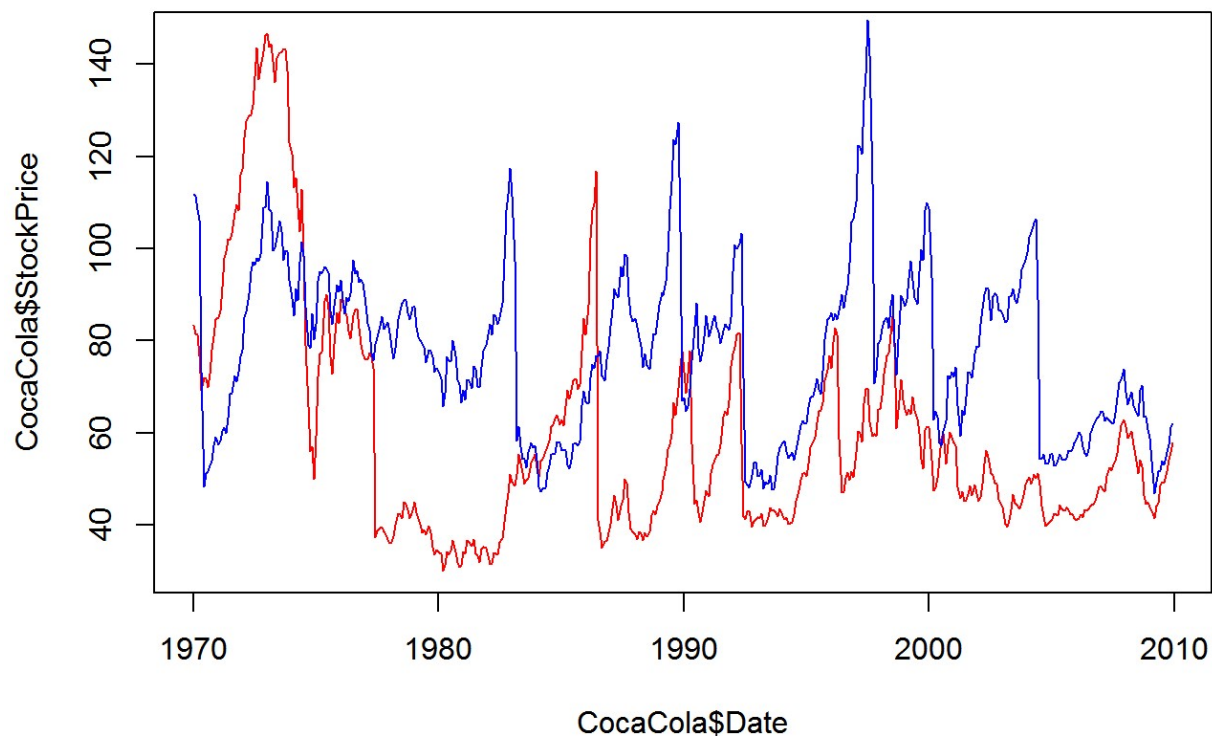
```
plot(CocaCola$Date,CocaCola$StockPrice, type = "l")
```



Adding a new line to the plot by using the lines function instead of the plot function. Adding colours to the lines as well to distinguish better.

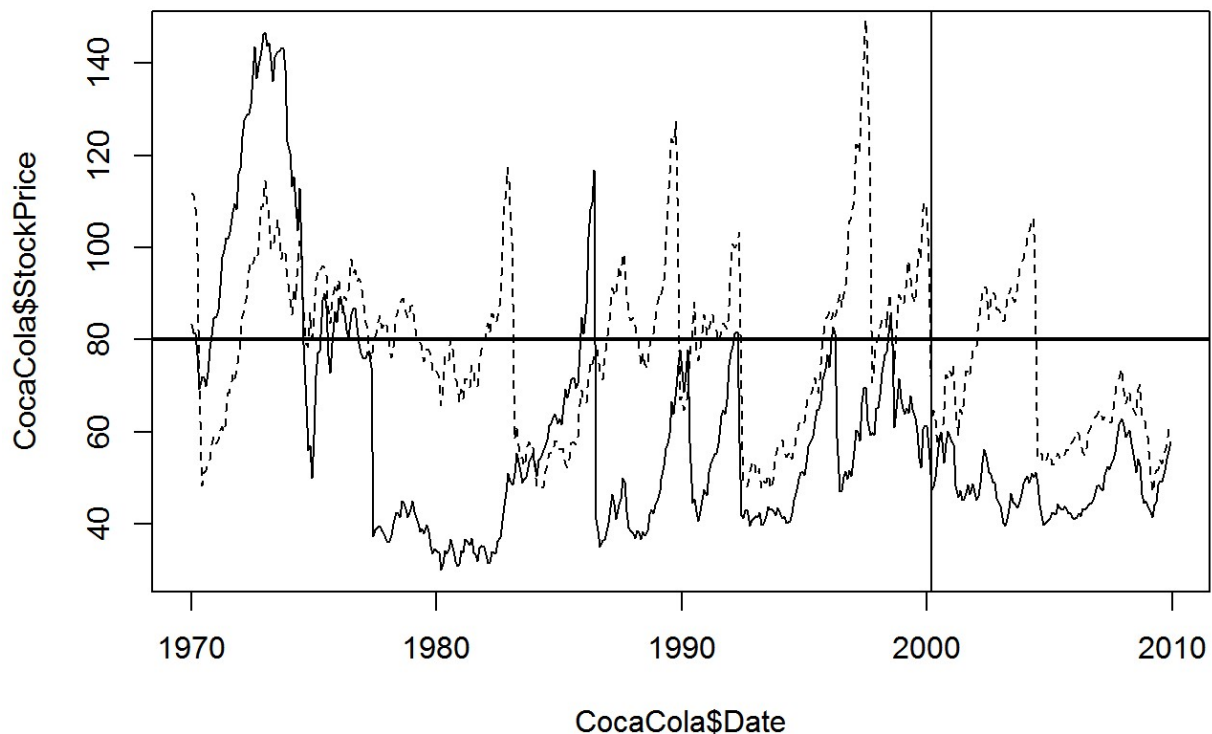
Note: In R markdown, we need to execute 'plot' and 'lines' command in same R-box, otherwise it would give error "plot.new has not been called yet."

```
plot(CocaCola$Date, CocaCola$StockPrice, type = "l", col="red")  
lines(ProcterGamble$Date, ProcterGamble$StockPrice, col="blue")
```



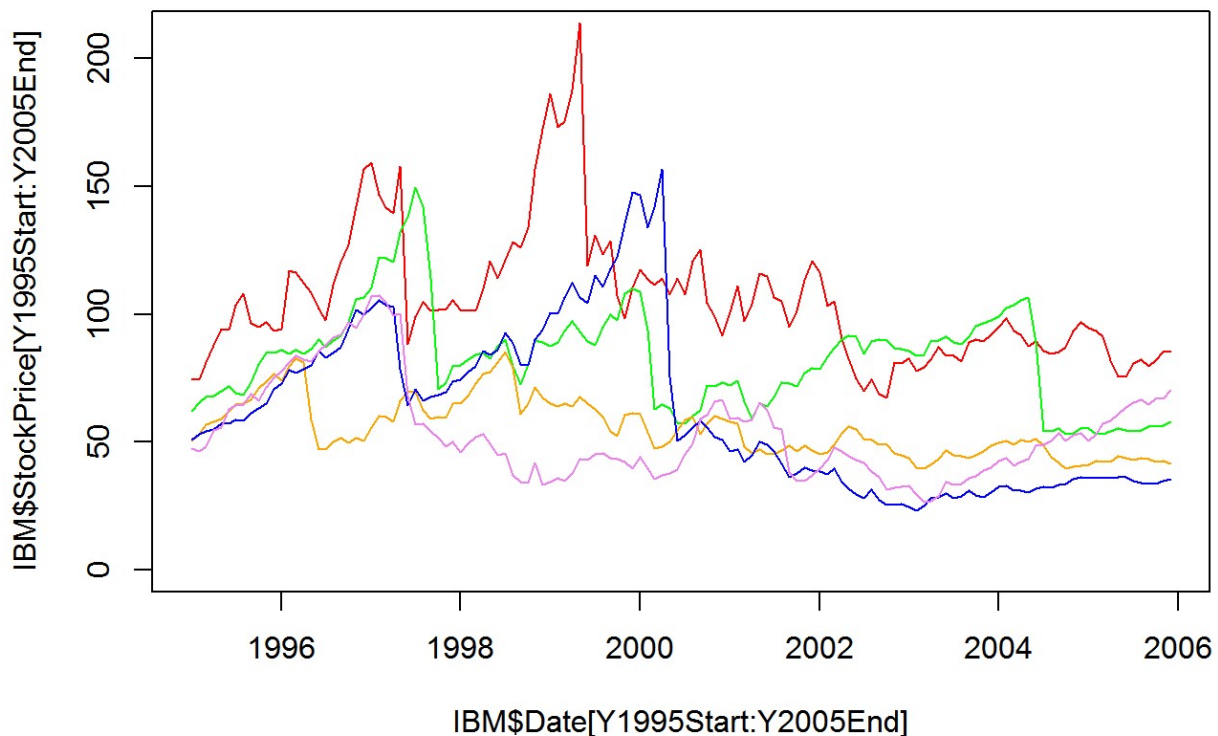
Instead we could also have changed the line type of one to distinguish better. We can draw vertical/horizontal line to make comparisons

```
plot(CocaCola$Date,CocaCola$StockPrice, type = "l")  
lines(ProcterGamble$Date, ProcterGamble$StockPrice,lty=2)  
abline(v=as.Date(c("2000-03-01")), lwd=1)  
abline(h=80, lwd=2)
```



Visualizing stock prices changed from 1995-2005 for all five companies.

```
Y1995Start=which(IBM$Date=='1995-01-01')
Y2005End=which(IBM$Date=='2005-12-01')
plot(IBM$Date[Y1995Start:Y2005End], IBM$StockPrice[Y1995Start:Y2005End], type
="l", col="red", ylim=c(0,210))
lines(ProcterGamble$Date[Y1995Start:Y2005End], ProcterGamble$StockPrice[Y1995St
art:Y2005End], col="green" )
lines(CocaCola$Date[Y1995Start:Y2005End], CocaCola$StockPrice[Y1995Start:Y2005E
nd] , col="orange")
lines(GE$Date[Y1995Start:Y2005End], GE$StockPrice[Y1995Start:Y2005End], col="bl
ue")
lines(Boeing$Date[Y1995Start:Y2005End], Boeing$StockPrice[Y1995Start:Y2005En
d], col="violet")
```



```
#abline(v=as.Date(c("1997-09-01", "1997-11-01")), lwd=.1)
```

By looking at this plot, we can see that the stock for General Electric falls significantly more than the other stocks after the technology bubble burst and IBM has the highest value, around 1999.

Stocks trends during each month

```
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
```

```
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
```

```
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
```

```
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
```

```
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
```

- For IBM, we can see that the price has historically been higher than average during January - May, and lower than average during the remaining months.
- General Electric has an average stock price of 64.48 in April, which is higher than any other month.

- Coca-Cola has an average stock price of 62.69 in April, which is higher than any other month.
- Having lower stock prices in December is a trend that holds for all five companies.

After seeing these trends, we are ready to buy stock in certain months and sell it in others! But, we should be careful, because one really good or really bad year could skew the average to show a trend that is not really there in general.

Trying the above summary in the same graph. Not working, need to try more.

```
plotStockByMonth <- function(company, LineCol) {
  stockByMonth = tapply(company$StockPrice, format(company$Date, "%m"), mean)
  stockByMonthDF = as.data.frame(stockByMonth)
  names(stockByMonthDF) = c("AvgStockPrice")
  stockByMonthDF$month = names(stockByMonth)
  if(count==0) {
    plot(stockByMonthDF$month, stockByMonthDF$AvgStockPrice, col=LineCol, type =
"l")
    count <- count + 1
  } else {
    lines(stockByMonthDF$month, stockByMonthDF$AvgStockPrice, col=LineCol)
    count <- count + 1
  }
}

count=1
plot.new()
plotStockByMonth(IBM, "red")
plotStockByMonth(GE, "orange")
plotStockByMonth(ProcterGamble, "green")
plotStockByMonth(Boeing, "blue")
plotStockByMonth(CocaCola, "yellow")
```

count

[1] 6