

# Basic Analytics 1

*author: Dr. Marko Mitic*

## Problem Description: **STOCK DYNAMICS**

A stock market is where buyers and sellers trade shares of a company, and is one of the most popular ways for individuals and companies to invest money. The size of the world stock market is now estimated to be in the trillions. The largest stock market in the world is the New York Stock Exchange (NYSE), located in New York City. About 2,800 companies are listed on the NYSE. In this problem, we'll look at the monthly stock prices of five of these companies: IBM, General Electric (GE), Procter and Gamble, Coca Cola, and Boeing. The data used in this problem comes from Infochimps.

Firstly, download and read the following files into R, using the `read.csv` function: `IBMStock.csv`, `GEStock.csv`, `ProcterGambleStock.csv`, `CocaColaStock.csv`, and `BoeingStock.csv`. (Do not open these files in any spreadsheet software before completing this problem because it might change the format of the Date field.)

```
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")
```

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.

For example, the structure of the IBM dataset this can be verified as follows:

```
str(IBM)
```

```
## 'data.frame':   480 obs. of  2 variables:
##  $ Date       : Factor w/ 480 levels "1/1/00","1/1/01",...: 11 171 211 251 291 331 371 411 451 51 ...
##  $ StockPrice: num  360 347 327 320 270 ...
```

Before working with these data sets, the dates in the dataframes need to be converted into formats that R can understand. We can convert this to a “Date” object in R by using the following five commands (one for each data set):

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

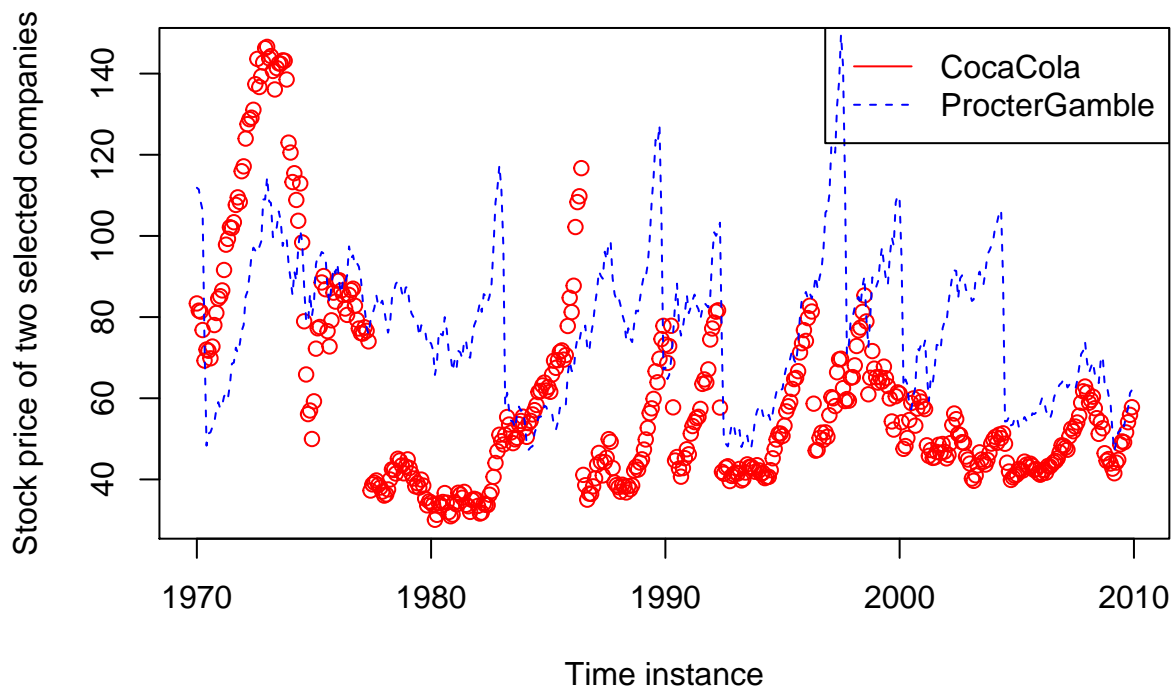
Statistical summary give us insights on the first and last year in each data frame:

```
summary(IBM)
```

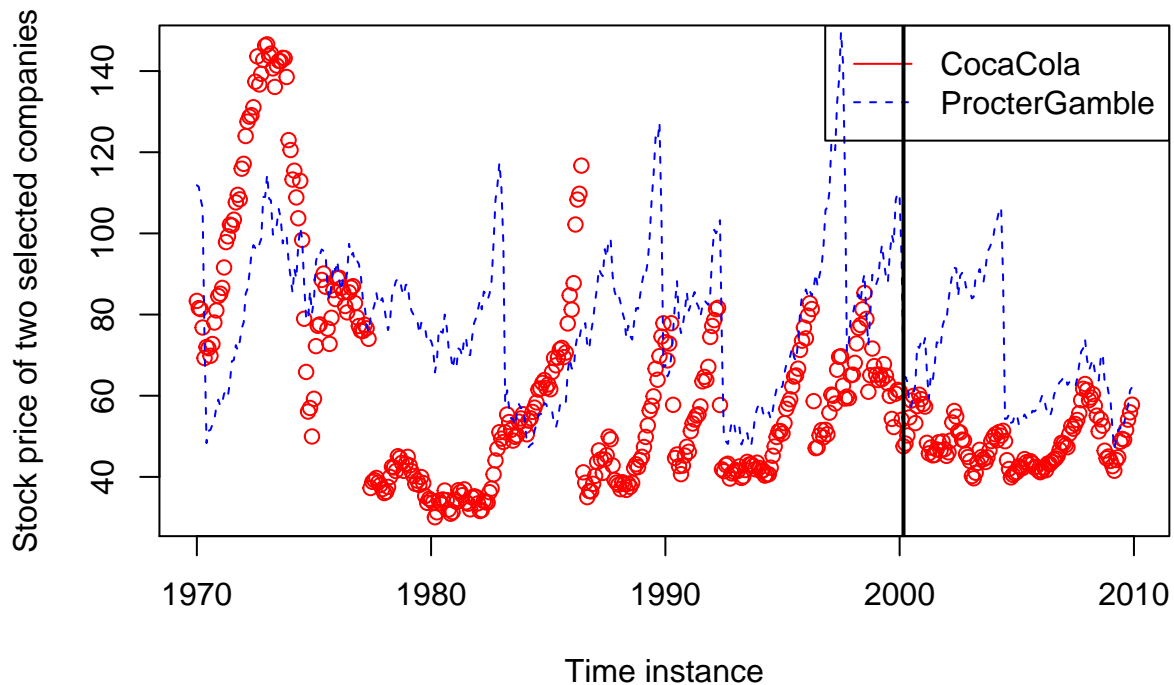
```
##      Date      StockPrice
## Min.   :1970-01-01  Min.   : 43.40
## 1st Qu.:1979-12-24  1st Qu.: 88.34
## Median :1989-12-16  Median :112.11
## Mean   :1989-12-15  Mean    :144.38
## 3rd Qu.:1999-12-08  3rd Qu.:165.41
## Max.   :2009-12-01  Max.    :438.90
```

Mean, median and quantile stock values can be observed using this command. Next, let's plot a graph to visualize stock price over the years (for CocaCola and Procter&Gamble companies).

```
plot(CocaCola$Date,CocaCola$StockPrice, col="red",
      xlab="Time instance", ylab="Stock price of two selected companies")
lines(ProcterGamble$Date,ProcterGamble$StockPrice, col="Blue", lty=2)
legend("topright", # places a legend at the appropriate place
      c("CocaCola", "ProcterGamble"), # puts text in the legend
      lty=c(1,2), # gives the legend appropriate symbols (lines)
      lwd=c(1,1),col=c("red","blue"))
```



From this plot, one can easily answer the questions about the highest and lowest stock price of a particular company. Using command `abline` as `abline(v=as.Date(c("2000-03-01")), lwd=2)`, the price around specific dates be further analyzed.

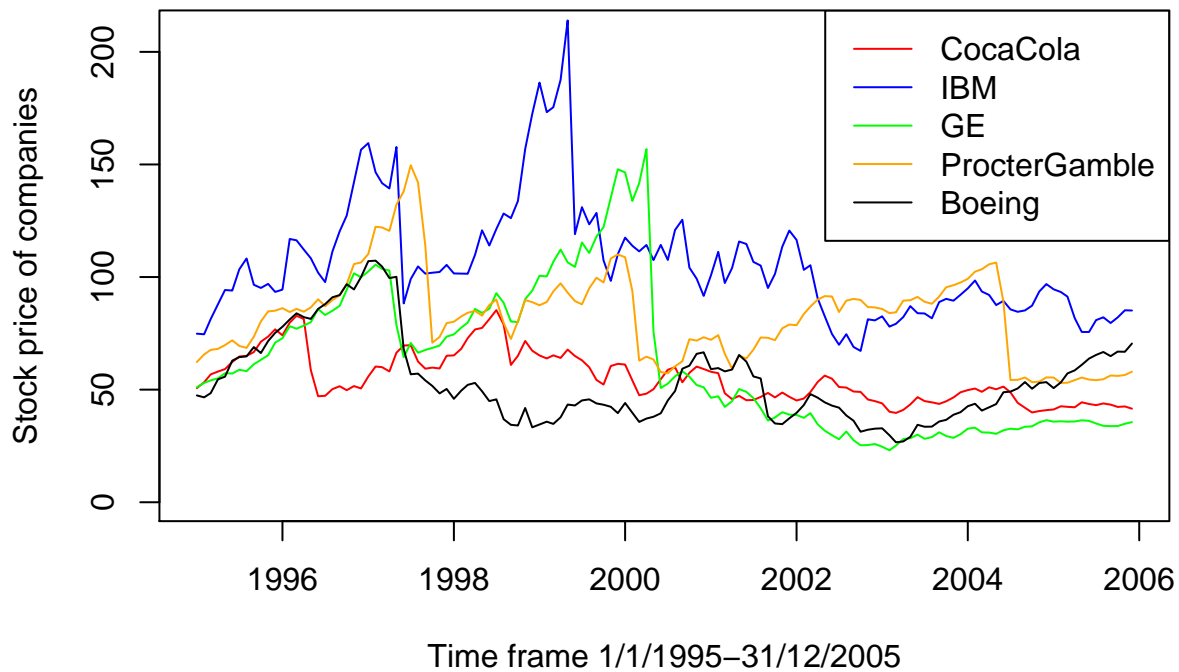


From new plot we can observe the stock price dropout of these selected companies during the technology bubble burst in March 2000. It is obvious that the Procter&Gamble company was affected more by this event.

The stock price change during time period 1995-2005 can be obtained with:

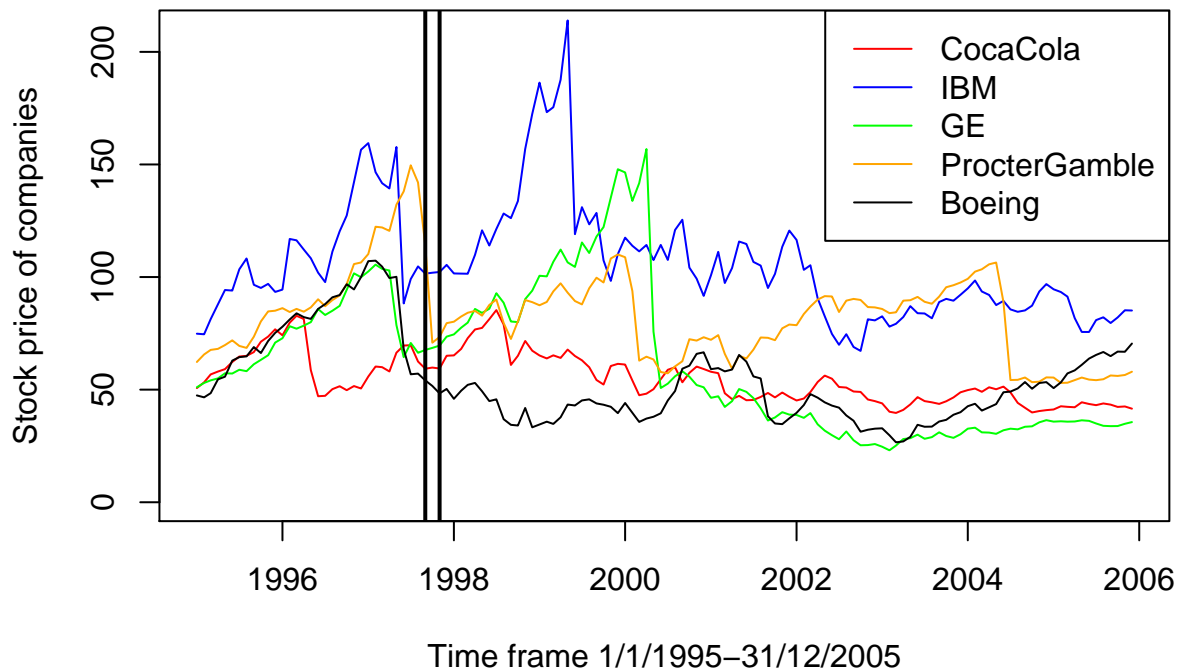
```
plot(CocaCola$Date[301:432], CocaCola$StockPrice[301:432], type="l", col="red", ylim=c(0,210),
     xlab="Time frame 1/1/1995-31/12/2005", ylab="Stock price of companies")
lines(IBM$Date[301:432], IBM$StockPrice[301:432], col="Blue", lty=1)
lines(GE$Date[301:432], GE$StockPrice[301:432], col="Green", lty=1)
lines(ProcterGamble$Date[301:432], ProcterGamble$StockPrice[301:432], col="Orange", lty=1)
lines(Boeing$Date[301:432], Boeing$StockPrice[301:432], col="Black", lty=1)

legend("topright", # places a legend at the appropriate place
      c("CocaCola", "IBM", "GE", "ProcterGamble", "Boeing"), # puts text in the legend
      lty=c(1,1,1,1,1), # gives the legend appropriate symbols (lines)
      lwd=c(1,1,1,1,1), col=c("red", "blue", "green", "orange", "black"))
```



One can notice (with `abline` command) that stock price of GE company has fell the most right after March 2000. Also, it can be observed that stock of IBM reached the highest price in period 1995-2005.

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, it can be concluded that Boeing and Procter&Gamble saw a decreasing trend in their stock price:



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

Finally, let us analyze stock price behaviour by month. We can calculate mean prices (for example for IBM) using `tapply` function:

```
tapply(IBM$StockPrice, months(IBM$Date), mean, na.rm=TRUE)
```

```
##      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
```

We can conclude that in 1995-2005 time period, stock price in first 5 Months (January-May) are above the overall average. Therefore, these are the months in which we want to sell our stocks. Also using `tapply` we conclude that every company's average stock price is higher at the beginning of the next year (by comparing December-January prices).

```
tapply(CocaCola$StockPrice, months(CocaCola$Date), mean, na.rm=TRUE)
```

```
##      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
```

```
tapply(GE$StockPrice, months(GE$Date), mean, na.rm=TRUE)
```

```
##      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, na.rm=TRUE)
```

```
##      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, na.rm=TRUE)
```

```
##      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
```