JUSTICE K S HEGDE
INSTITUTE OF MANAGEMENT
NITTE - 574110

Learning Financial Analysis in MBA with R

Laxmi Nayak*
Amrutia Meet**
Mohsin Ahmed***

Abstract

MBA students frequently learn financial analysis without being able to visualize the intricacies of financial algorithms; this changed with the availability of R, the free statistical data analysis language. In this paper we show how R can be used in selecting a portfolio based on the dynamic market. The selection of portfolio is crucial for an investor to get the desired return. The use of statistical tools for the selection of stocks in the portfolio is helpful in hedging the risk or to decide the level of risk the investor wants to be exposed to. Using the R language, an investor can visualize the performance of his/her portfolio by graphing the projected returns. It is a decision making tool in the selection and evaluation of the portfolio.

Keywords: Financial management, statistics, R, risk, portfolio evaluation, pedagogy

INTRODUCTION

MBA students had to rely on expensive and non-portable software like SPSS and Excel to do financial data analysis, until the R language came along. The R Language is a free open source programming language for Statistical computing and Graphics. It is a powerful language for data analysis and visualization. The basis for R language is the S language from Bell Labs. R language is named after the first letter of names of its two authors: Ross Ihaka and Robert Gentleman from New Zealand. Usually the C programming language is used to code computers; but C lacks high level statistical functions and graphics, whereas R has many built-in statistical functions. R is highly extensible wherein the user can alter the program characteristics according to his/her requirement. R is also user friendly and portable, running on a wide variety of machines.

This article discusses the application of R in the field of Statistics and Finance as taught in MBA. R is used in analyzing data and to make statistical inference. The common statistical functions in R are *Correlation, T-test,* and *Regression*. The analysis and calculation of the *return* on a Portfolio is important for every investor. The analysis of a *share* is vital before investing in it. The analysis helps the investor to pick the right *Share* at the right time. The analysis of the *Share's* past predicts it's future performance.

Details

We assume the user has installed R and R studio on his/her laptop running Windows, Mac or Linux. We are using R version 3.1.1 on Windows 7 in the screenshots below.

*Student, Justice K.S.Hegde Institute of Management, NMAMIT, Nitte 574110, India. Nyk.laxmi.06@gmail.com
**Student, Justice K.S.Hegde Institute of Management, NMAMIT, Nitte 574110, India. rg.amrutia@gmail.com
***Professor, Justice K.S.Hegde Institute of Management, NMAMIT, Nitte 574110, India. moshahmed@gmail.com

ISBN 978-93-5196-889

Share price analysis with R Studio

To analyze the *Share* price in R, we use the 'Quantmod' package. To get the package follow the steps as shown in the following diagrams:

1. Go to Tools

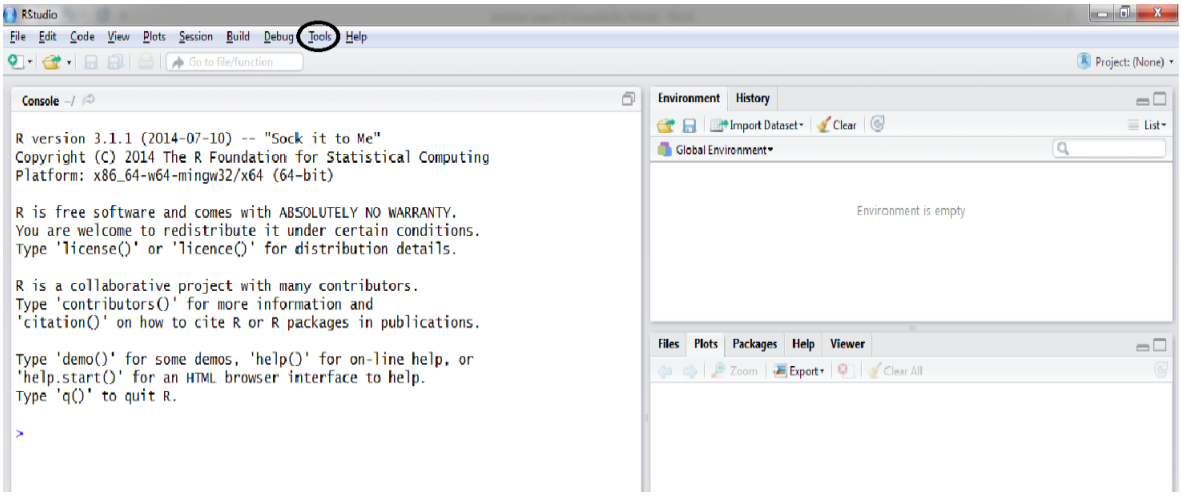

Figure 1:

2. Select install packages from the drop down list.

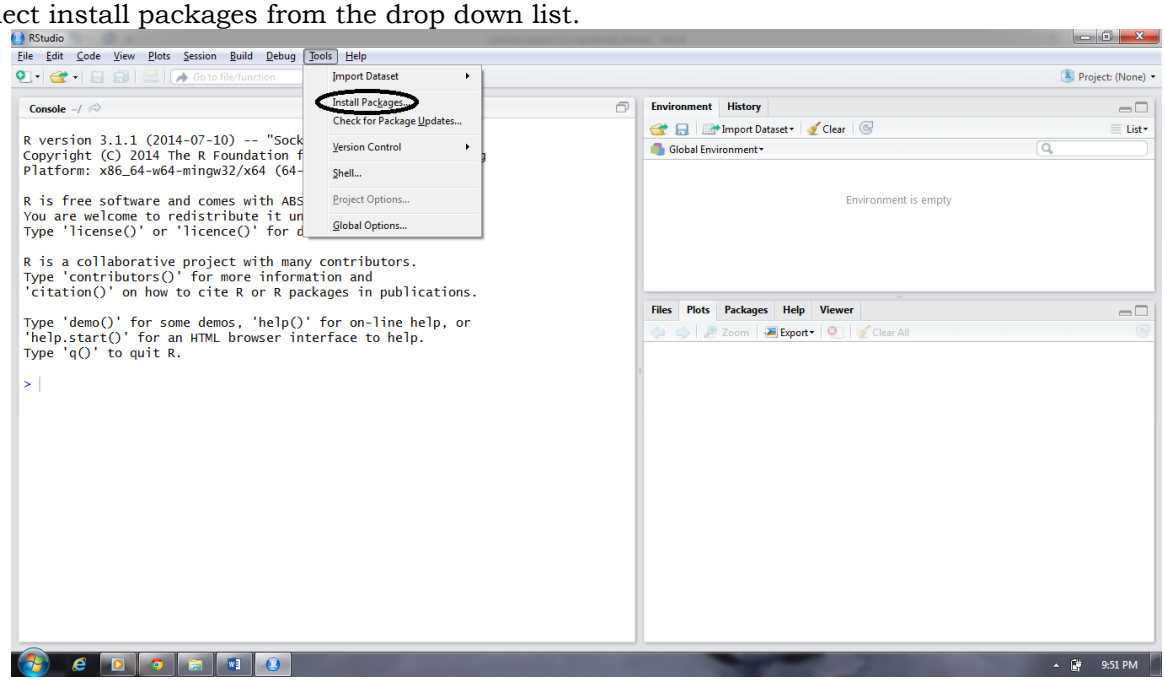

Figure 2

ISBN 978-93-5196-889

3. In the install package window type 'Quantmod'

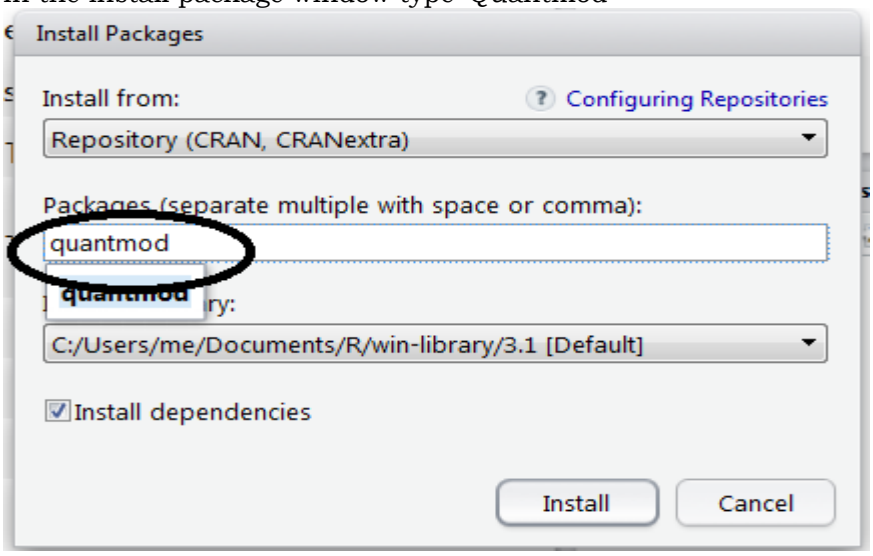

Figure 3

4. This installs the package and it is ready to use.

The next step is to get the Share prices. To get the Share price type the following code in the console window.

```
library(quantmod)
getSymbols("TCS.BO")
```

TCS.BO ×

1954 observations of 6 variables

	row.names	TCS.BO.Open	TCS.BO.High	TCS.BO.Low	TCS.BO.Close	TCS.BO.Volume	TCS.BO.Adjusted
1	2007-01-02	1250.00	1256.00	1222.00	1248.55	202800	543.10
2	2007-01-03	1200.00	1287.50	1200.00	1281.30	359500	557.34
3	2007-01-04	1300.00	1300.00	1252.10	1263.50	400800	549.60
4	2007-01-05	1210.00	1271.00	1210.00	1262.55	384100	549.19
5	2007-01-08	1265.00	1270.00	1217.05	1245.60	586700	541.81
6	2007-01-09	1250.00	1263.00	1227.00	1255.45	331800	546.10
7	2007-01-10	1255.00	1275.00	1230.00	1234.10	303700	536.81
8	2007-01-11	1245.00	1297.90	1205.00	1277.20	611800	555.56
9	2007-01-12	1300.00	1330.00	1282.00	1325.30	940400	576.48
10	2007-01-15	1339.00	1367.00	1305.10	1327.90	2133600	577.61
11	2007-01-16	1399.00	1399.00	1320.15	1327.85	993600	577.59
12	2007-01-17	1332.90	1334.70	1280.00	1305.90	410100	568.04

Displayed 1000 rows of 1954 (954 omitted)

Figure 4

Here we consider the Share of TCS. The extension .BO indicates the prices are from the BSE. Quantmod takes the prices from *Yahoo Finance* by default. It can be taken from any site of the choice of the user. The following code will give the share price from the date since the site is recording the share prices. In this case it gives prices from 1st January, 2007. If you want any specific period then type the following code:

ISBN 978-93-5196-889

```
library(quantmod)
getSymbols("TCS.BO")
startT  <- "2010-01-01"
endT    <- "2012-01-01"
rangeT  <- paste(startT,"::",endT,sep ="")
tTCS.BO   <- TCS.BO[,6][rangeT]
```

tTCS.BO

	row.names	TCS.BO.Adjusted
1	2010-01-01	694.68
2	2010-01-04	696.49
3	2010-01-05	695.74
4	2010-01-06	679.90
5	2010-01-07	662.16
6	2010-01-08	649.05
7	2010-01-11	661.79
8	2010-01-12	694.08
9	2010-01-13	730.35
10	2010-01-14	724.75
11	2010-01-15	733.64
12	2010-01-18	740.87
13	2010-01-19	723.26

Figure 5

This code gives the Share price from 1st January 2010 to 1st January 2012.
To plot the graph of the *Share,* add this line of code:

```
plot(tTCS.BO, col="red")
```

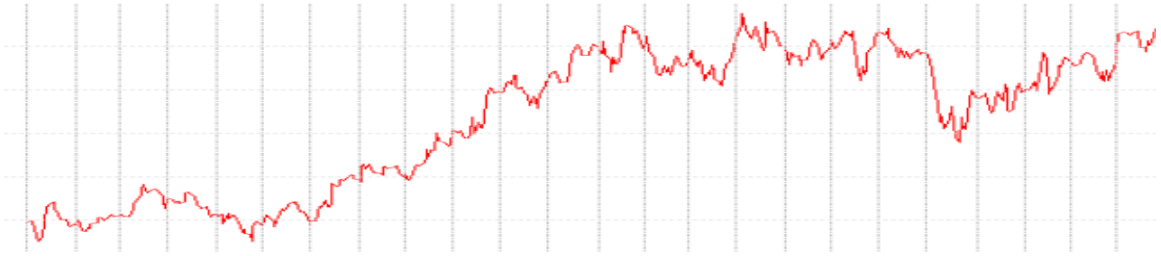

Figure 6

To compare and analyze the growth of one Share with another Share, we can add the other Share in the same graph, with this code:

```
library(quantmod)
getSymbols("TCS.BO")
getSymbols("INFY.BO")
startT  <- "2010-01-01"
endT    <- "2012-01-01"
rangeT  <- paste(startT,"::",endT,sep ="")
tTCS.BO   <- TCS.BO[,6][rangeT]
tINFY.BO   <- INFY.BO[,6][rangeT]
```


ISBN 978-93-5196-889

```
plot(tTCS.BO, col="red")
par(new=TRUE)
plot(tINFY.BO, col="blue")
```

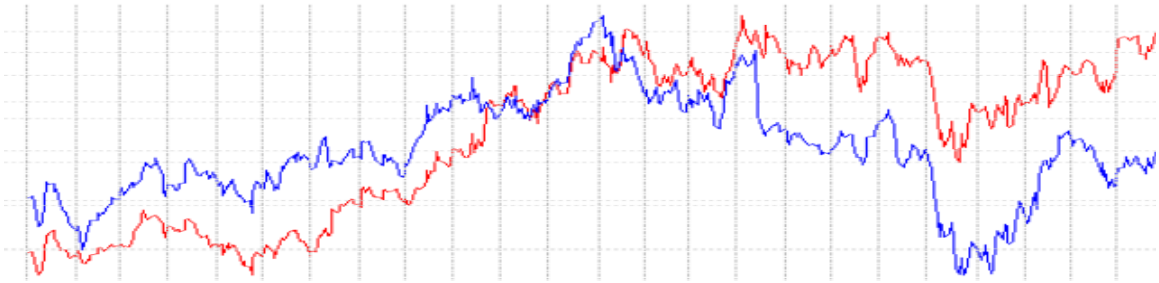

Figure 7

This code will plot the graph of the share price of TCS and INFOSYS in the same graph for the time period specified. This way we can view and analyze multiple; and compare the returns generated by the Shares for the period specified.

To obtain the CAGR (*annual return)* of your portfolio type the following code

```
cagr <- (last_price/base_price)^(1/n)-1
```

Here the *last_price* can be the price at which you will sell the Share or even the latest price of Share. The *base_price* can be the *buy price* or the price which you select as base; 'n' is the number of years you held the Share. This code will give the return generated by the Share annually.

ISBN 978-93-5196-889

Correlation

R is also used in computing correlation. Correlation method is used in finance for the purpose of *hedging* the risk involved in investing in the stocks. The risk involved in investing in one stock can be mitigated by investing in another stock which has negative correlation with it. Similarly when the correlation is positive, risk increases as two stocks move up and down together. In the example below, the closing price of different stocks is taken into consideration to determine the correlation between them. The code to find the correlation between different stocks is as shown below. The closing share price of companies in different sectors is considered, viz: Infosys, ITC, ONGC, Sun pharma, Ashok Leyland, IRB, Jet Airways.

```
library(quantmod)
getSymbols(c("INFY.BO","ITC.BO","ONGC.BO","SUNPHARMA.BO","ASHOKLEY.BO","IRB
.BO","JETAIRWAYS.BO"))
Data<cbind(diff(log(Cl(INFY.BO))),diff(log(Cl(ITC.BO))),diff(log(Cl(ONGC.BO))),diff(log(Cl(S
UNPHARMA.BO))),diff(log(Cl(ASHOKLEY.BO))),diff(log(Cl(IRB.BO))),diff(log(Cl(JETAIRWA
YS.BO))))
cor(Data,use="pairwise.complete.obs")
corrplot(cor(Data,use="pairwise.complete.obs"))
```

The correlation output is as shown below.

	INFY.BO.Close	ITC.BO.Close	ONGC.BO.Close	SUNPHARMA.BO.Close
INFY.BO.Close	1.00000000	0.16817190	0.1437487	0.04613895
ITC.BO.Close	0.16817190	1.00000000	0.1824896	0.09797613
ONGC.BO.Close	0.14374873	0.18248963	1.0000000	0.09154710
SUNPHARMA.BO.Close	0.04613895	0.09797613	0.0915471	1.00000000
ASHOKLEY.BO.Close	0.16770256	0.16077122	0.2648574	0.06595595
IRB.BO.Close	0.10615793	0.13379150	0.2803564	0.12810188
JETAIRWAYS.BO.Close	0.12355553	0.15401877	0.2522722	0.06426276

	ASHOKLEY.BO.Close	IRB.BO.Close	JETAIRWAYS.BO.Close
INFY.BO.Close	0.16770256	0.1061579	0.12355553
ITC.BO.Close	0.16077122	0.1337915	0.15401877
ONGC.BO.Close	0.26485742	0.2803564	0.25227217
SUNPHARMA.BO.Close	0.06595595	0.1281019	0.06426276
ASHOKLEY.BO.Close	1.00000000	0.3345593	0.34666089
IRB.BO.Close	0.33455930	1.0000000	0.29930768
JETAIRWAYS.BO.Close	0.34666089	0.2993077	1.00000000

ISBN 978-93-5196-889

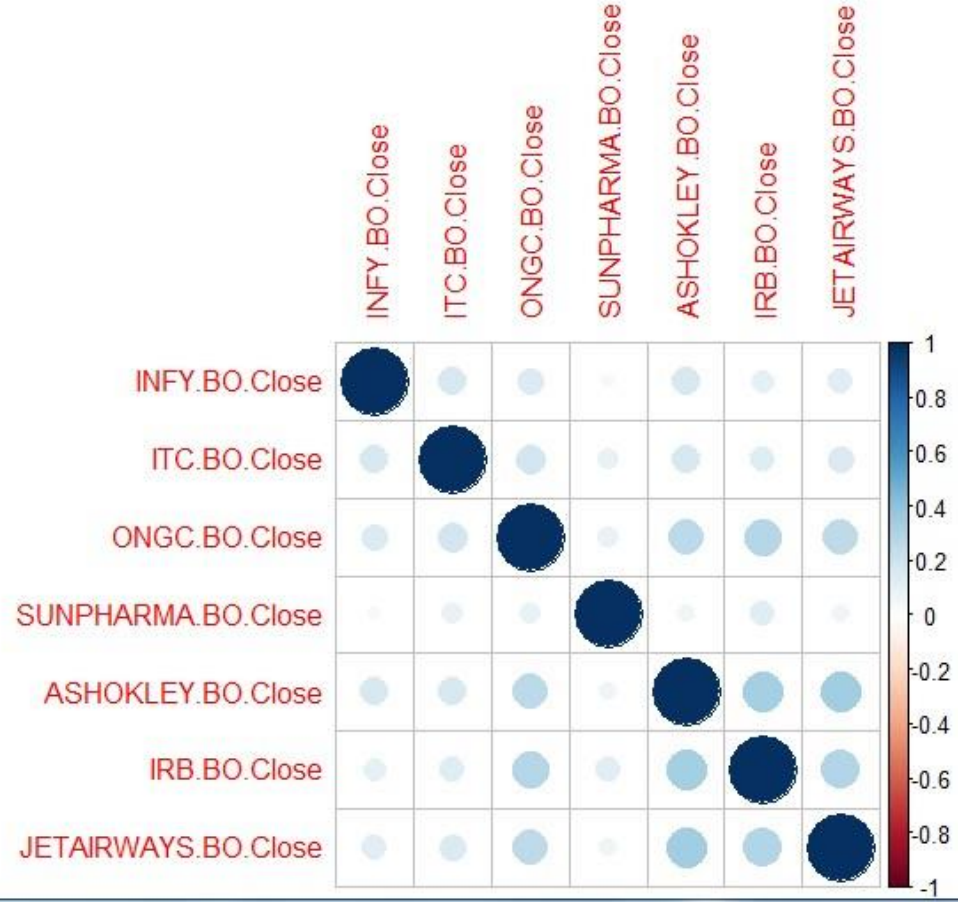

Figure 8: The correlation that exists between different stocks.

Figure 8 shows the correlation between the different stocks. This method of finding the correlation between different stocks using R will help the investor in deciding which shares to buy in his/her portfolio to mitigate the risks – in hedging. They can change the portfolio based on the changes in the market.

From the above correlation calculation, we notice the correlation coefficient between Infosys and Sun pharma is close to zero since the stocks move in unrelated directions. During the recession time, the stock of Infosys fell whereas the stock of Sun pharma remained unaffected since the company is in the pharma sector which doesn't get affected by the IT recession. Similarly, the stocks of Jet Airways and Ashok Leyland move in the same direction, being in the airline and automobile sector respectively.

Regression

R language can also be used for obtaining the regression line of the given data. Here we show a simple example to obtain the regression line of the two variables whose values are normally distributed random number, generated with the R function 'rnorm'. Among

ISBN 978-93-5196-889

the two variables, one is dependent and another is independent variable. The R code to obtain the value dependent variable y from the independent variable x is given below:

```
x <- rnorm(100)
y <- x + rnorm(100)
> plot( x, y)
> abline( lm(y ~ x), col = "red" )
```

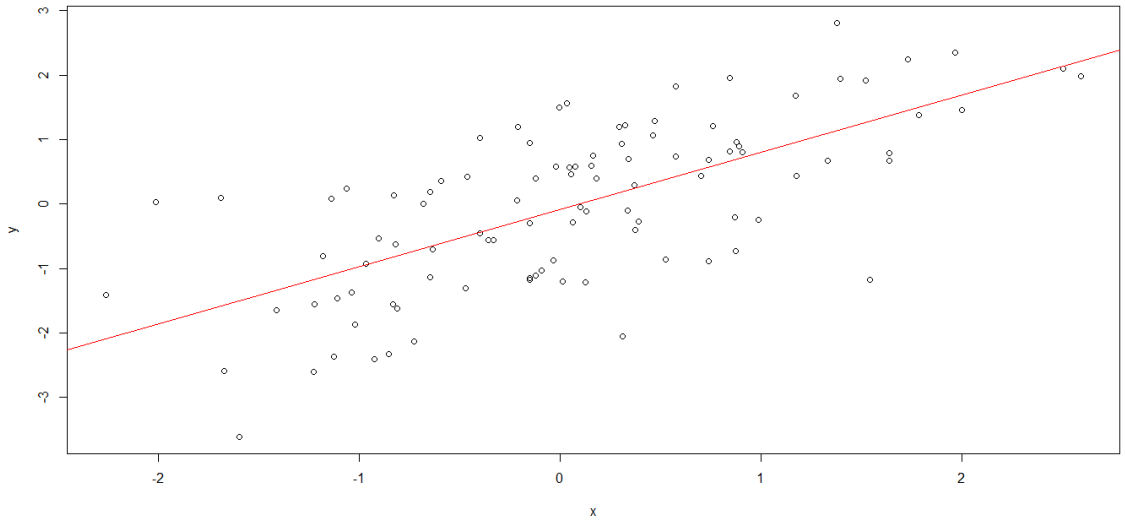

Figure 9: The regression line of dependent variable y.

Plotting the values of x and y where the numbers are generated randomly, we get a regression as the straight line in Figure 9.

In the next example, we do regression on real sales data. The sales data for the last 10 years is known and we want to estimate the sales for 11th year using R. First, we read the data, and assign it to a variable called 'data'. The values are plotted using the 'plot' function. Using '*abline'* function, the regression line which satisfies both sales and year is obtained. The output is shown below as a graph.

```
read.csv(file.choose())
data<-read.csv(file.choose())
data
   year sales
1     1  1000
2     2  1300
3     3  1800
4     4  2000
5     5  2000
6     6  2000
7     7  2200
8     8  2600
9     9  2900
10   10  3200
11   11    NA
View(data)
plot(data$sales ~ data$year)
```


ISBN 978-93-5196-889

```
z<-lm(data$sales ~ data$year)
abline(data$sales ~ data$year)
abline(z,col='red')
```

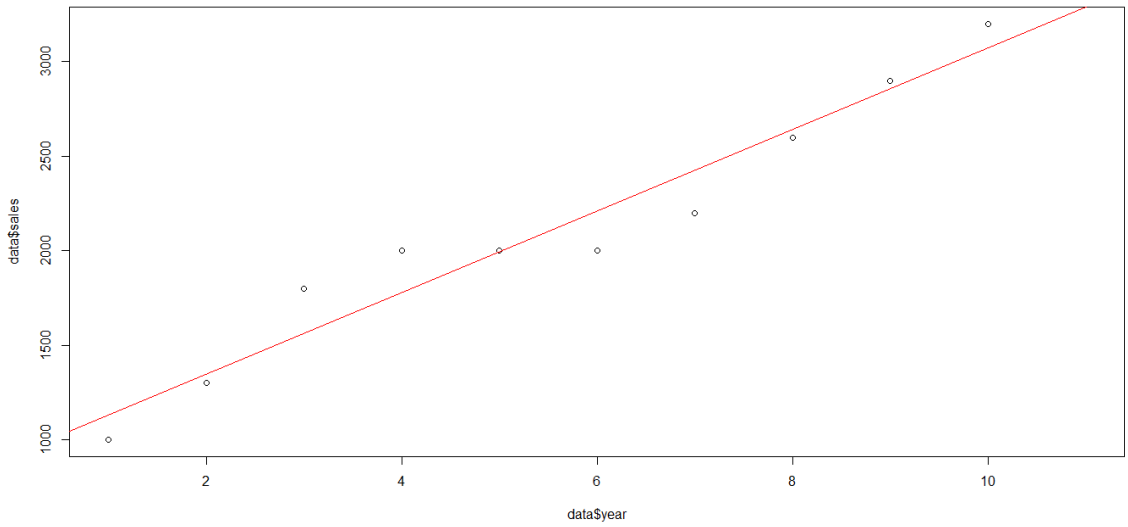

Figure 10: Showing the sales of 10 consecutive years.

To find the sales for 11th year, we call Plot along with the limit of x and y. In this case the limit for x=11 and y=4000. By substituting the value of x=11 in the equation, we can get the sales value. After we get the sales of 11th year, we can plot it, as shown below.

```
z
Call:
lm(formula = data$sales ~ data$year)
Coefficients:
(Intercept)    data$year
      913.3        215.8
plot(data$year , data$sales, xlim=c(0,11), ylim=c(0,4000))
data$sales[11] <- 913.3 + 215.8 * 11
abline(z)
913.3   +    215.8 * 11
[1] 3287.1
plot(data$year , data$sales, xlim=c(0,12), ylim=c(0,4000))
913.3   +    215.8 * 12
[1] 3502.9
abline(z,col='red')
```


ISBN 978-93-5196-889

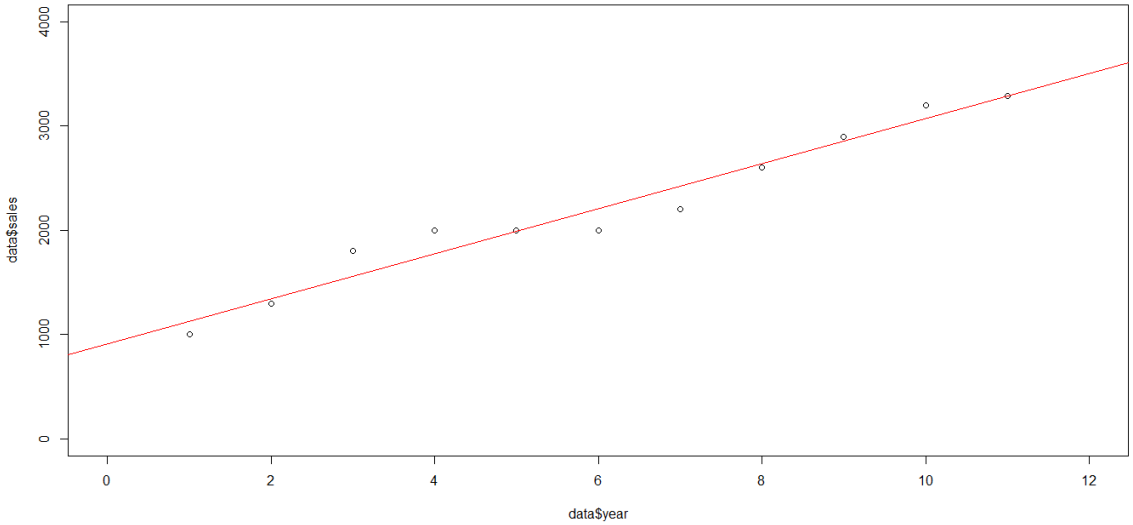

Figure 11: Showing the sales of 12th year.

Conclusion

R is a powerful langauge which can be used for data analysis. Its applications are vast, but here we focused on Statistics and Finance. R gives the output after every step than giving the final output, unlike Excel. R scripts can be automated to run without human intervention. This can help in analyzing large amounts of data quickly. Another important benefit of R is that it does not make any assumption about the data and options, unlike Excel. Therefore R provides better analysis and learning tool for MBA students than any other.

Acknowledgements

We thank our director Dr. K. Sankaran for constant encouragement on our work; Prof. Krishna Prasad for help with the APA referencing style; Prof. Krishna Kumar for encouraging us to use R.

References

1. *Statistic Methods*, retrieved 10 Nov 2014, from http://statmethods.net/STATS/
2. "*R for Traders*" retrieved 10 Nov 2014, from http://www.rfortraders.com/
3. *Statistical Analysis with R*, Wikibooks, retrieved 10 Nov 2014, from http://en.wikibooks.org/wiki/Statistical_Analysis:_an_Introduction_using_R/R_basics
4. *Correlation matrices through R QuantMod*, Retrieved 10 Nov 2014,from https://www.quora.com/How-do-I-generate-correlation-matrices-through-R-QuantMod-for-stocks

ISBN 978-93-5196-889

5. "*R for MBA*" by Kavya M N, Nishita Rai, Jovita M, Nikitha F, Laxmi N, Meet A, Techbugs workshop, JKSHIM, NITTE, India 8/2014.

ISBN 978-93-5196-889