FORBES 2000 GLOBAL LIST-Version 2017  
Statistical Review   
BY “R”

Spring 2018

**introduction**

The data handling and manipulation techniques explanation is the main goal of our final R project and because the students in our class are specialized in different area I have decided to choose a data set that most of them can have sense of those data. The Forbes 2000 list for year 2016 collected by “Forbes Magazine” have been used for this purpose.

**Methodology and resulst**

Total review of Data analysis has been started by using str(), dim(),nrow(),ncol(),…and then continued by more complex analysis. Like most of real data sets we have missing data that should be deal with and I have done this after the primary review. You can see that there are blank spaces in Sector column that I changed it to a new level as “unknown” .first I changes Sector class From Factor to character and after replacing unknown instead of blank space I changed it back to its original Factor type.

Data manipulation is the next step that has been done. R software is one of the strongest one for this phase .by using sort(),subset(),attach(),…I give a better view from the Forbes 2000 Global data set.by sub setting you can get a better understanding about these big companies, for example by *>UKcomp <- subset(Forbes2000, Country == "United Kingdom")* and then *>dim(Ukcomp)* you can see 91 or %4.6=(91/2000) of these companies are UK based company. May be someone is interested in companies with highest sales :

*>order\_sales<-order(Forbes2000$Sales,decreasing = TRUE)*

*>Forbes2000[order\_sales[1:3] ,c("Company","Sales", "Profits", "Assets")]*

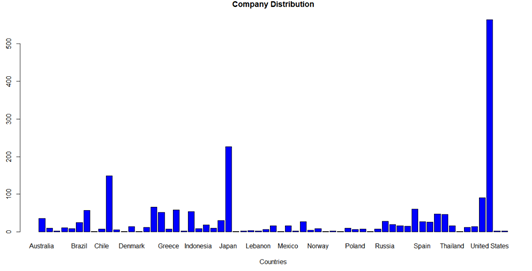
*Company Sales Profits Assets*

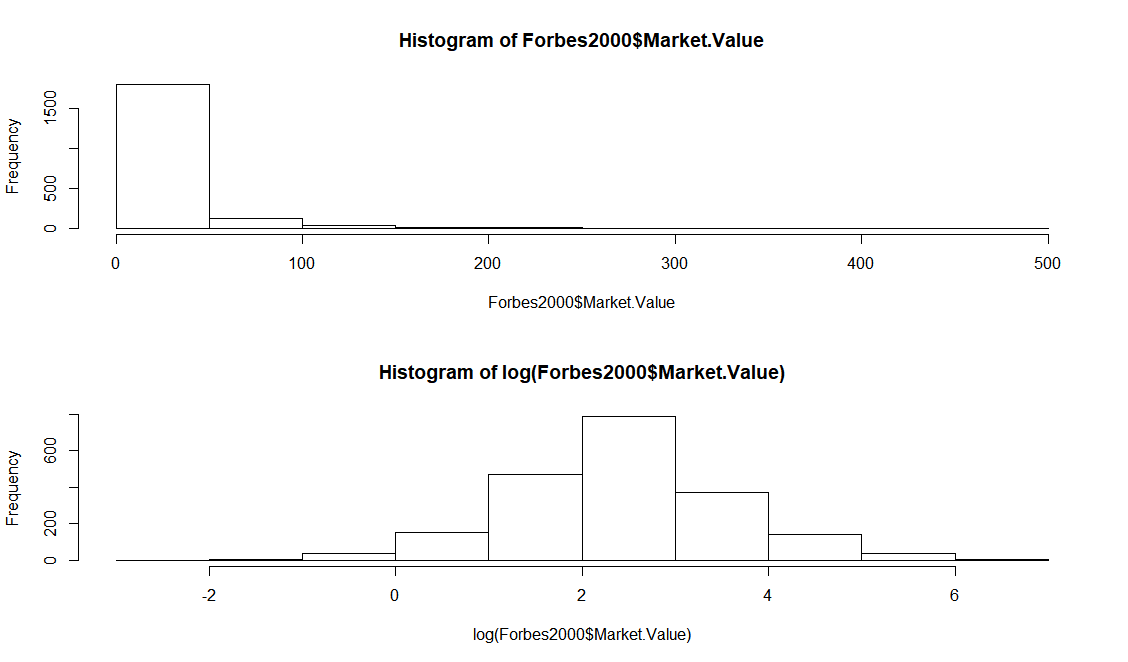
*20 Wal-Mart Stores 476.5 16.0 204.8*

*11 Royal Dutch Shell 451.4 16.4 357.5*

*29 Sinopec-China Petroleum 445.3 10.9 228.4*

From the *>summary(Forbess2000)*  output we can immediately see that most of the companies are situated in the US and that most of the companies are working in the banking

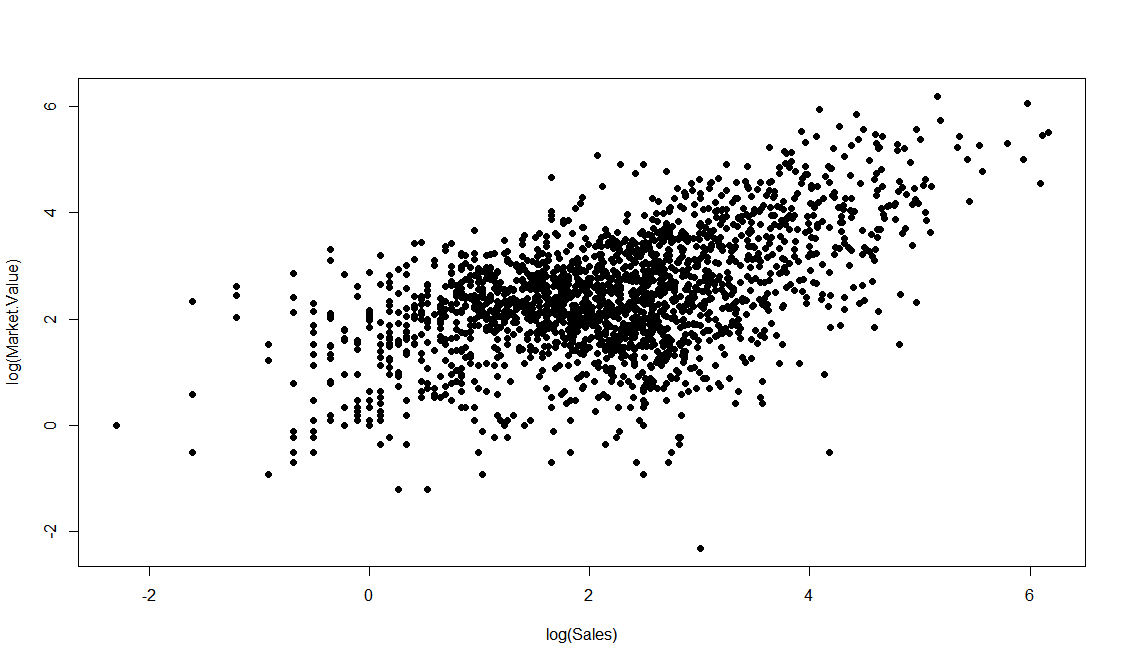
Sector. Also by using a bar plot you can see the most of these companies are in US ,Japan and China.

The degree of skewness of a distribution can be investigated by constructing

histograms using the hist function.

*>hist(log(Forbes2000$Market.Value))*

we use logarithmic version of our data for more symmetric view.

****

Scatterplot of the logarithms of market value and sales.

*>plot(log(Market.Value) ~ log(Sales), data = Forbes2000,pch = 20)*

**Conclusion**

Reading data into R is possible in many different ways, including direct connections

to data base engines. Tabular data are handled by data.frames in R,and the usual data

manipulation techniques such as sorting, ordering or sub-setting can be performed by simple R statements. An overview on data stored in a data.frame can be given mainly by two functions (summary and str). Simple graphics such as histograms and scatterplots can be constructed by applying the appropriate R functions (hist and plot).

**Refrences**

https://www.r-project.org/other-docs.html, http:// www.forbes.com

“R for Data Analytics”, Metro college hand out