Dataset Analysis I

(Assignment 1)

Ankita Dasgupta

J014 Semester III

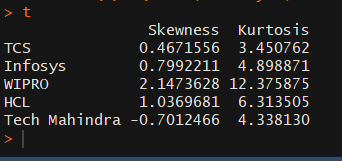
B.Tech Data Science

1st August, 2020

**Q. For the last 4 months stock market is not performing well because of COVID19. As analyst you must find how the IT sector (majorly in these five stocks-TCS, Infosys, Wipro, HCL Tech, Tech Mahindra) is performing in this period.**

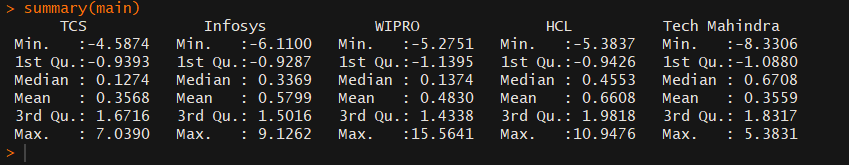
**A. Do you think it is worst affected by COVID19?**

**Ans.** We can comment on the effect of COVID 19 to the stocks of all five companies with the help of the values of skewness and kurtosis.



The negative value of skewness tells us that we have negative returns for the stocks belonging to Tech Mahindra.

Also, as the values for Kurtosis for all five companies is greater than three, we can say that the curve for their returns would be leptokurtic.

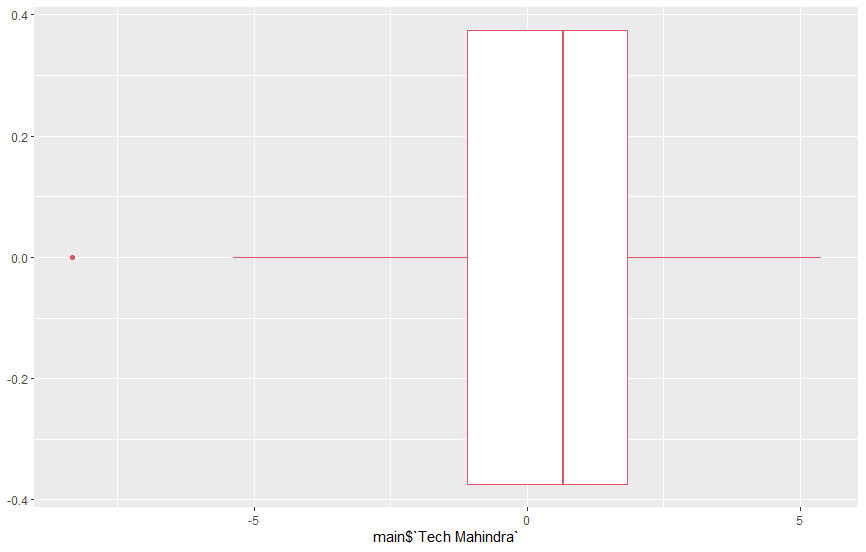


Overall, most of the IT companies would show a bell curve which is skewed towards the right meaning that they would be suffering losses due to the pandemic.

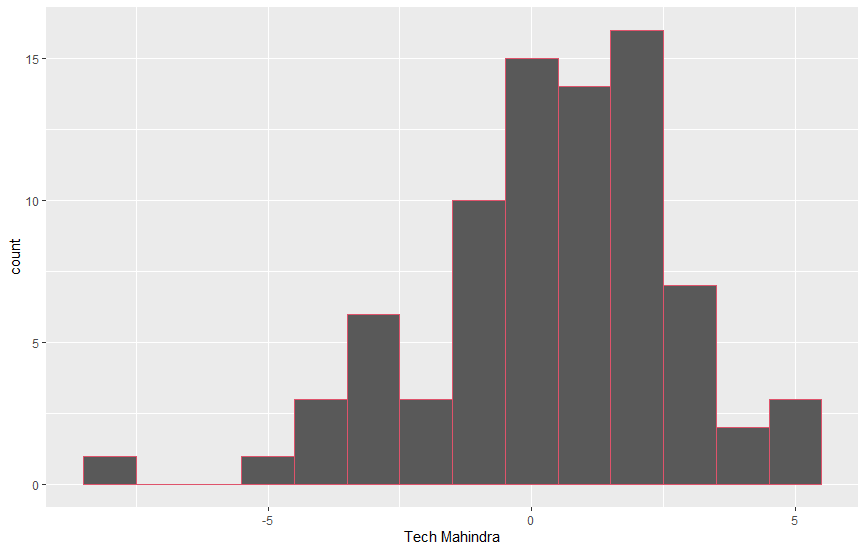
**B. If yes then among the stocks which stock in this sector is the worst performer and why you think so.**

**Ans.** To identify the worst performer we can compare the values of skewness. As the skewness of Tech Mahindra is negative, we can say that it is the worst performer.

The Boxplot for the returns of Tech Mahindra also shows the same.



(Q3-Q2) < (Q2-Q1)

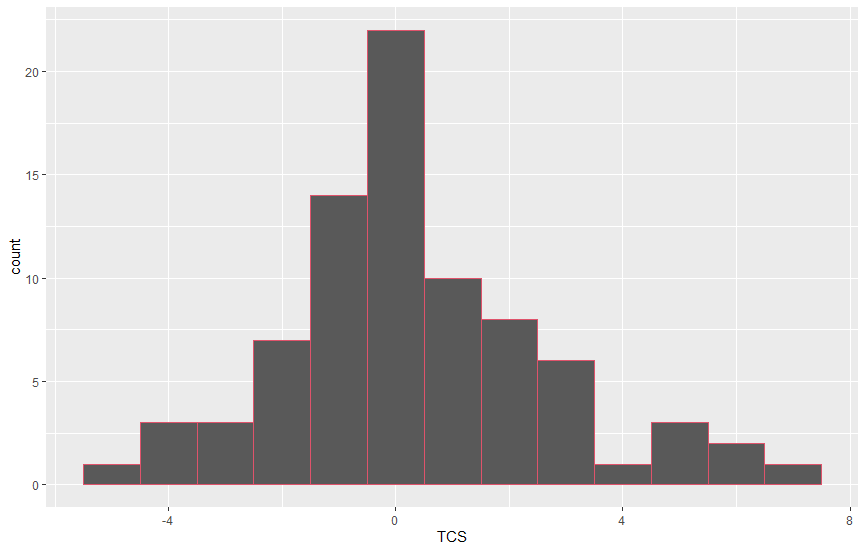


The histogram for the returns of Tech Mahindra also shows the same inference.

**C. If No then among the stocks which stocks in this sector is the best performer and why you think so.**

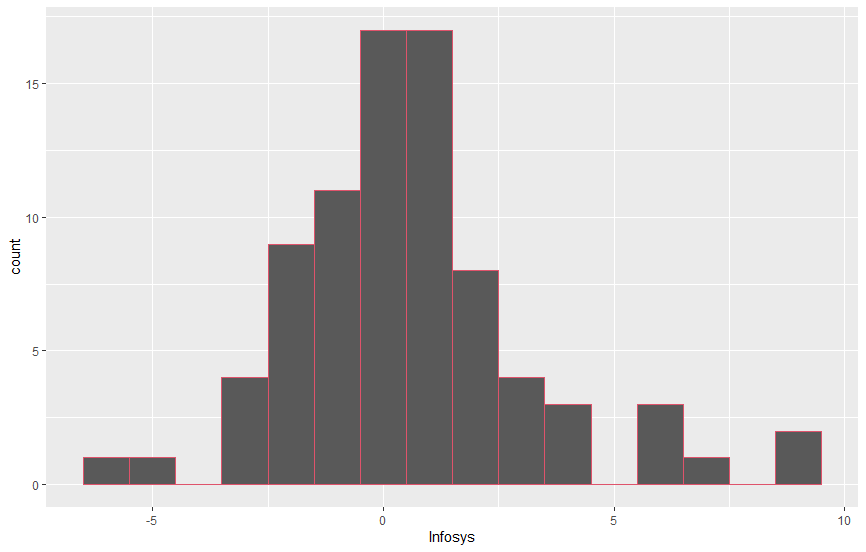
**Ans.** With the help of the skewness as well as the histograms we can find the best performer.

* TCS:



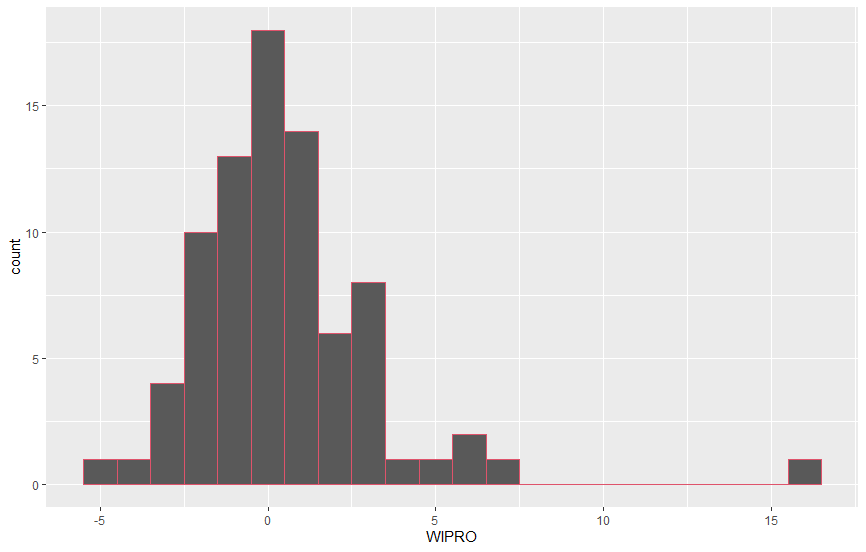
Skewness: 0.4671556

* Infosys:



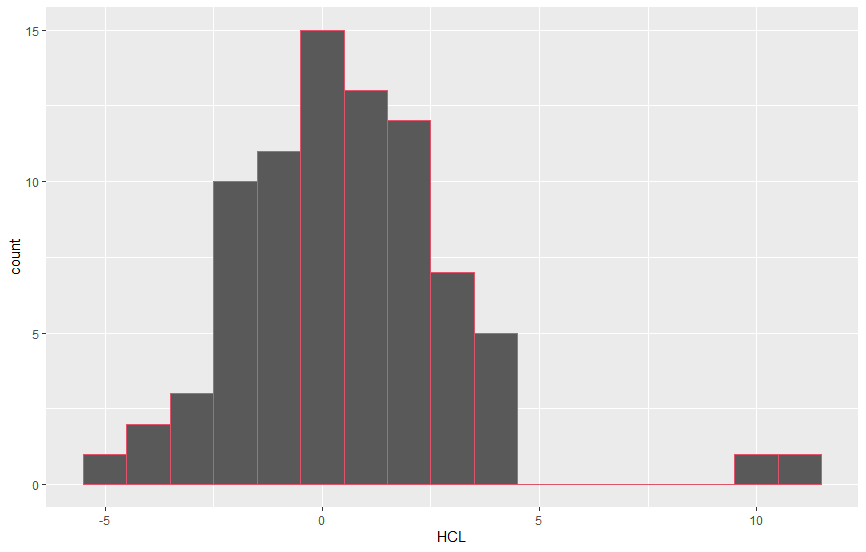
Skewness: 0.7992211

* WIPRO:



Skewness: 2.1473628

* HCL:



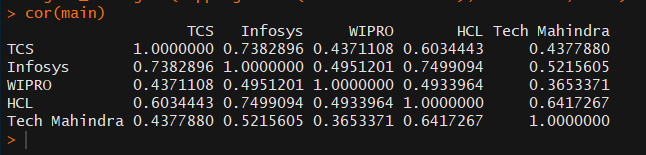
Skewness: 1.0369681

As the magnitude of skewness for the returns of WIPRO are the highest, we can say that will have the most positive returns i.e., WIPRO is the best performer.

**D. Do think the stocks are correlated with each other?**

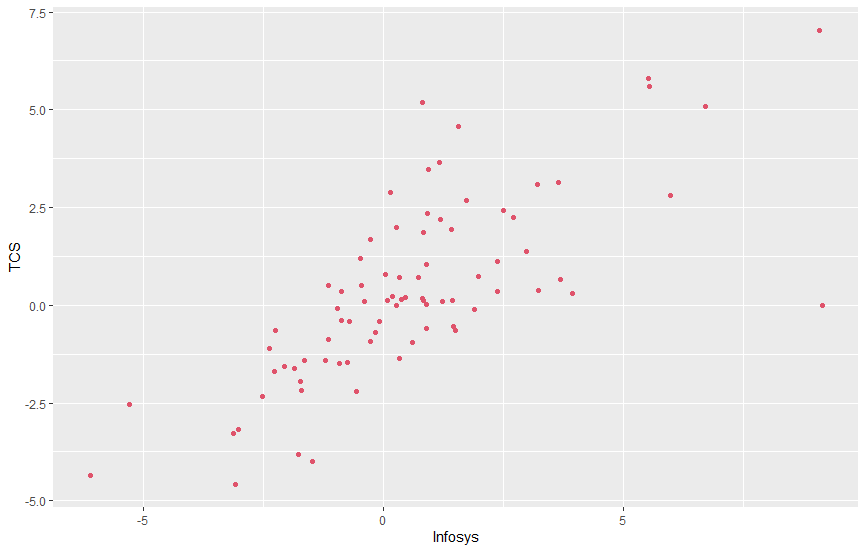
**Ans.** Correlation is defined as interdependence between variable quantities.

We can find the correlation between all each of the five companies with the help of the correlation matrix.

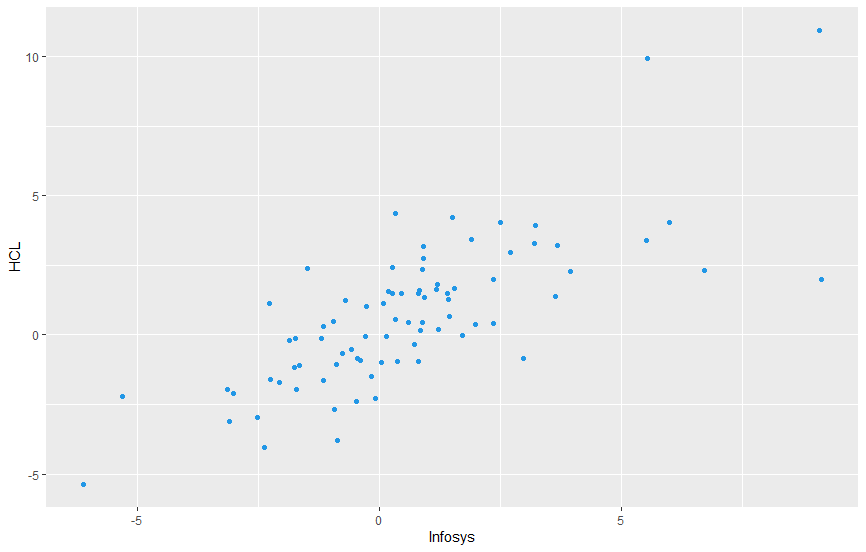


With the help of the correlation matrix and scatter plots we can say that,

* None of the stocks are negatively correlated to each other.
* Infosys and TCS are highly correlated.



* HCL and Infosys are highly correlated.



**Code:**

**install.packages("ISLR")**

**library(quantmod)**

**prices1<- getSymbols("TCS.BO", source="yahoo", auto.assign=FALSE, return.class="xts",from="2020-04-01")[,4]**

**return1<-na.omit(diff(log(prices1)))\*100**

**View(return1)**

**prices2<- getSymbols("INFY.BO", source="yahoo", auto.assign=FALSE, return.class="xts",from="2020-04-01")[,4]**

**return2<-na.omit(diff(log(prices2)))\*100**

**View(return2)**

**prices3<- getSymbols("WIPRO.BO", source="yahoo", auto.assign=FALSE, return.class="xts",from="2020-04-01")[,4]**

**return3<-na.omit(diff(log(prices3)))\*100**

**View(return3)**

**prices4<- getSymbols("HCLTECH.BO", source="yahoo", auto.assign=FALSE, return.class="xts",from="2020-04-01")[,4]**

**return4<-na.omit(diff(log(prices4)))\*100**

**View(return4)**

**prices5<- getSymbols("TECHM.BO", source="yahoo", auto.assign=FALSE, return.class="xts",from="2020-04-01")[,4]**

**return5<-na.omit(diff(log(prices5)))\*100**

**View(return5)**

**main<-data.frame(return1, return2, return3, return4, return5)**

**View(main)**

**colSums(is.na(main))**

**install.packages("moments")**

**library(moments)**

**v1=c(skewness(main$TCS.BO.Close), skewness(main$INFY.BO.Close), skewness(main$WIPRO.BO.Close), skewness(main$HCLTECH.BO.Close), skewness(main$TECHM.BO.Close))**

**v2=c(kurtosis(main$TCS.BO.Close), kurtosis(main$INFY.BO.Close), kurtosis(main$WIPRO.BO.Close), kurtosis(main$HCLTECH.BO.Close), kurtosis(main$TECHM.BO.Close))**

**t=data.frame(v1,v2)**

**colnames(t)=c('Skewness','Kurtosis')**

**rownames(t)=c('TCS','Infosys','WIPRO','HCL','Tech Mahindra')**

**t**

**colnames(main)=c('TCS','Infosys','WIPRO','HCL','Tech Mahindra')**

**install.packages("tidyverse")**

**library(tidyverse)**

**summary(main)**

**ggplot(data= main)+**

**geom\_boxplot(mapping = aes(x=`Tech Mahindra`),col=2)**

**ggplot(data= main)+**

**geom\_histogram(mapping = aes(x=`TCS`),binwidth=1,col=2)**

**ggplot(data= main)+**

**geom\_histogram(mapping = aes(x=`Infosys`),binwidth=1,col=2)**

**ggplot(data= main)+**

**geom\_histogram(mapping = aes(x=`WIPRO`),binwidth=1,col=2)**

**ggplot(data= main)+**

**geom\_histogram(mapping = aes(x=`HCL`),binwidth=1,col=2)**

**ggplot(data= main)+**

**geom\_histogram(mapping = aes(x=`Tech Mahindra`),binwidth=1,col=2)**

**cor(main)**

**ggplot(data= main)+**

**geom\_point(mapping = aes(x=`Infosys`,y=`TCS`),col=2)**

**ggplot(data= main)+**

**geom\_point(mapping = aes(x=`Infosys`,y=`HCL`),col=4)**