Stock Data Visualization of Pharmaceutical Companies

MTH208A ,DSLab -1

Instructor :- Dr. Dootika Vats

Suchandra Sasmal , Debarghya Jana , Mrinmoy Saha , Mohammad Saqib Ansari

Abstract:

We all need medicine in some or more part of our lives, as humans we are always interested in gains of either material or spiritual. There are several pharmaceutical companies .We here tried to analyse the several companies performance with respect to the financial measure to make the task of the investor easier to make the better decision. So here we tried to simplify the task with our shiny app to see their performance of the **last ten years since 2013 to 2022.**

To better appreciate the data we need to understand the some terminologies related to stock market which are as follows.

Market Cap: The total worth of all a company's stock shares is referred to as market cap, also known as market capitalization.

1. **Large -Cap** : Those companies whose market capitalization is more than twenty thousand crore.
2. **Mid -Cap** : Those companies whose market capitalization is more than Five thousand crore and less than 20 thousand crore.
3. **Small-Cap** : Those companies whose market capitalization is less than five thousand crore.

CMP: CMP is essentially the Current Market Price.It is also known as Market Value.It is the most recent selling price of a stock, currency, commodity etc that is traded on an exchange and is the most reliable indicator of that security's present value.

PE: PE is Price-earnings ratio.It is also known as P/E ratio, P/E, or PER, is the ratio of a company's share (stock) price to the company's earnings per share. It is used for valuing companies and to find out whether they are overvalued or undervalued.

Industry PE: It is the average price-to-earning ratio of a particular sector or industry. It's used as a benchmark to compare the PE of a stock to the PE of an entire industry.

Equities: Equity refers to the shares in a company's ownership in the context of Stock Market Investments.It is the total amount of money that a shareholder is eligible to receive if all of a company's debts are paid off and its assets liquidated.When a company issues shares to the investors in return of money,these shares are called equities.The equity meaning in share market is nothing but these shares which investors can buy or sell.The equity market is also called a stock market where traders buy or sell shares.

PVB: It is the price to book value ratio, which compares the market and book value of the company.The higher the PBV, the more expensive the stock.

TTM: TTM(Trailing 12 months) is the term for the data from the past 12 consecutive months used for reporting financial figures. It is a measurement of a company's financial performance.

Operating Profit Margin: The operational margin calculates the profit an organization makes on each rupee of sales , after variable production costs, such as labor and raw materials, but before interest or taxes.It is computed by dividing operating income by net sales for a business.

Net Profit Margin: The net profit margin calculates the profit an organization makes on each rupee of sales,after variable production cost,such as labor and raw materials,but before interest or taxes.It is computed by dividing operating income by net sales for a business.

Return on Equity(ROE): A metric of financial performance known as return on equity(ROE) is obtained by dividing net income by shareholder's equity.ROE is referred to as the return on net assets since shareholder's equity is determined by company's debt from its assets.

Annual Revenue Sales: The total amount of money a business brings in each year from the sale of goods, services, assets, or capital is known as annual revenue. It does not take any of costs into consideration. Because of this, income statements frequently refer to revenue as "sales."

Introduction:

When covid 19 hit the world , the life came to the halt, many business, educational institute and industries were shut but the pharma companies were open so we may think it as the most profitable business but that is not the fact as per the Forbes 2022 list of top 50 companies and fortune list of 2021 , there are no mention of any pharmaceutical companies.

Here we get the information that top companies that are making better profit are mostly banking about 20 % , oil-petroleum and Info tech 10% as per Forbes 2022 list.

We extracted the data from the first four pages of **"MoneyWorks4Me"** of the **"top pharma companies in India"** .using the package rvest .

Data:

```
k <- load("/Users/debarghyajana/Downloads/Pharma_Companies (2).rdata")
```

The data contain 95 rows and 31 column , each row contains the information regarding the company and it financial measure on that basis the companies performance is analyse.

```
## [1] "The following variables are"
```

```
##  [1] "Name"                    "Headquarters"
##  [3] "Year"                    "MarketCap_Rs.Cr."
##  [5] "Type_Of_MarketCap"       "CMP"
##  [7] "CMP_Change_percent"      "Latest_PE"
##  [9] "Latest_PBV"              "FiveYearProfitGr_percent"
## [11] "ROCE_percent"            "Sales_Mar.13"
## [13] "Sales_Mar.14"            "Sales_Mar.15"
## [15] "Sales_Mar.16"            "Sales_Mar.17"
## [17] "Sales_Mar.18"            "Sales_Mar.19"
## [19] "Sales_Mar.20"            "Sales_Mar.21"
## [21] "Sales_Mar.22"            "TTM_Sales"
## [23] "BVPS"                    "TTM_BVPS"
## [25] "Adjusted_Net_Profit"     "TTM_ANP"
```

```
## [27] "Return_On_Equity"          "Operating_Profit_Margin"
## [29] "Net_Profit_Margin"         "FiftyTwo_Week_High"
## [31] "FiftyTwo_Week_Low"         "AllTime_High"
## [33] "AllTime_Low"               "Equity_Rs.Cr."
## [35] "Industry_PE"
```

Essentially , In the website , there are total 4 pages of data.So, we just separately scraped the website one by one page and then collect data page wise.There are **'NA'** written in some column of some variable to emphasize that the value is **0** . We replace the 'NA' with value **0** to make that column of that variable totally numerical. After this type of data cleaning procedure , total data of all pages had been gathered and make the final table with **95** observations and there are total **26** variables under which our total data was obtained. Then the total data is saved in **".Rdata"** object.Then we worked on the finally sorted data set. There were two columns year of establishment and headquarter was not present so we tried to extract it manually.

Obtaining of the data and Data Cleaning: Essentially , In the website , there are total 4 pages of data.So, we just separately scraped the website one by one page and then collect data page wise.There are **'NA'** written in some coloumn of some variable to emphasize that the value is **0** . We replace the 'NA' with value **0** to make that coloumn of that variable totally numerical.After this type of data cleaning procedure , total data of all pages had beed gathered and make the final table with **95** observations and there are total **26** variables under which our total data was obtained.Then the totaal data is saved in **".Rdata"** object.Then we worked on the finally sorted dataset.

Bias in the data :-

The term "bias" may be define as the deviation from the true value. Here as we have taken the data of the listed companies only which showed on the first four pages of the website "moneywork4me.com" the company might be showing the companies which might be bias on showing some companies as there are more than 3000 pharma companies in the India. So we may say there might be the selection bias. Here we tried to infer on the basis of the data only the actual values might be differ from our observation.

Questions arrising from dataset:-

Our main goal is to make some data analysis by data visualization method. So, questions which I want to give answer are shorted below.

1)Which pharmacy company has significant amount of growth ? So, we make some questions to get an idea about the best growing pharmacy company in India on a particular moment that is on particular timepoint on a day (Note: Date: 2nd October , 2.06 pm).Here We just want to observe that which company has highest and lowest value with respet to all variables. By getting answers of this question we can get idea of such a company which have significant amount of growth on that day and then make conclusions about them.

We have done this analysis by using the help of **Barplot**.

2)Which Market capatalization types of business are frequent?

3)To what extent pharma companies came into existence after independence ?

4)Which State has the highest concentration of businesses?

5)Which campanies giving better returns to their investors?

6)Which companies showed the best growth in last five years?

7)To compare the companies sales performance during 2013 and 2022.

8)To compare financial measures to analyze the company performance.

Visualizations:

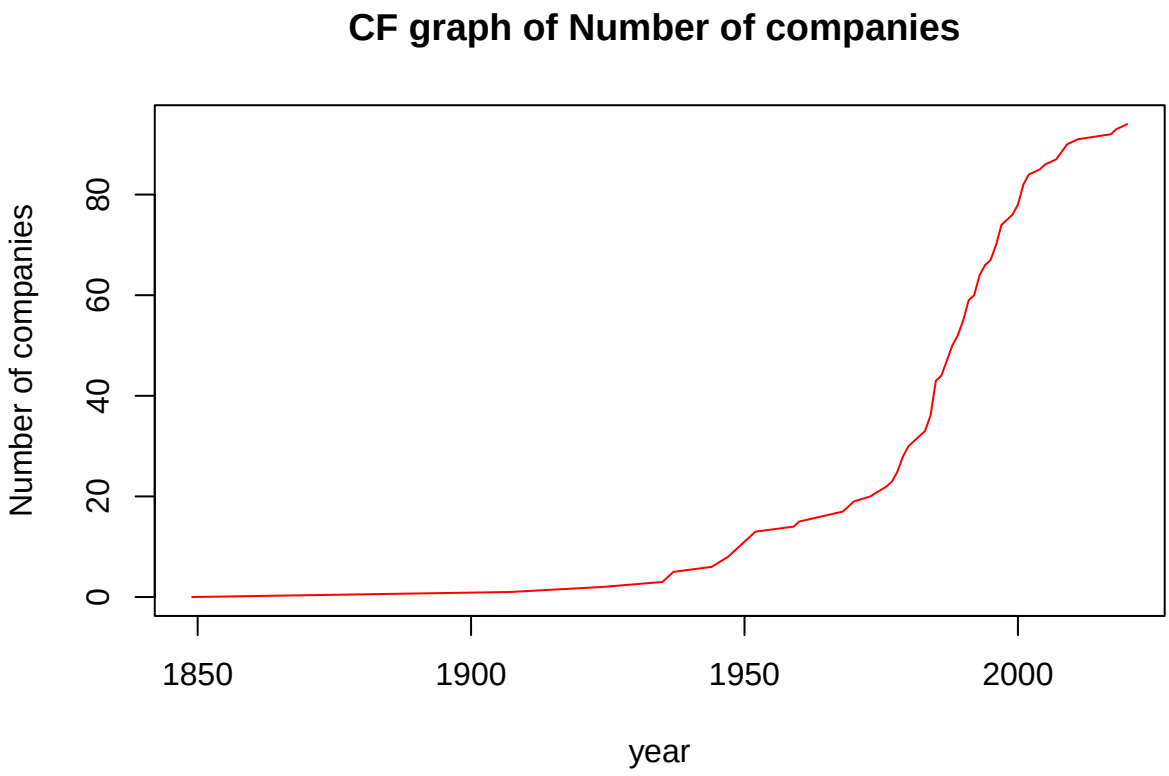
Here we will make some plots that will essentially gives answers of our intended queries.

Now we are going to plot a **cummulative frequency curve** of Number of companies vs. Year.

```
k <- load("/Users/debarghyajana/Downloads/Pharma_Companies (2).rdata")

year <- Stock_Data$Year
year<- sort(year)
x<-c()

for(i in year){
 add<- length(year[year<i])
 x<-c(x,add)
}
x
```

```
##  [1]  0  1  2  3  4  5  6  6  8  8 10 11 12 13 14 15 15 17 18 19 20 20 22 23 23
## [26] 25 25 25 28 28 30 31 31 33 33 33 36 36 36 36 36 36 36 43 44 44 44 47 47 47
## [51] 50 50 52 52 52 55 55 55 55 59 60 60 60 60 64 64 66 67 67 67 70 70 70 70 74
## [76] 75 76 76 78 78 78 78 82 82 84 85 86 87 87 87 90 91 92 93 94
```

```
plot(x=year,y=x,col="red",ylab="Number of companies", xlab="year",
     type="l",main="CF graph of Number of companies ")
```



By observing the cummulative frequency curve we can conclude that It is obviously a monotonically increasing

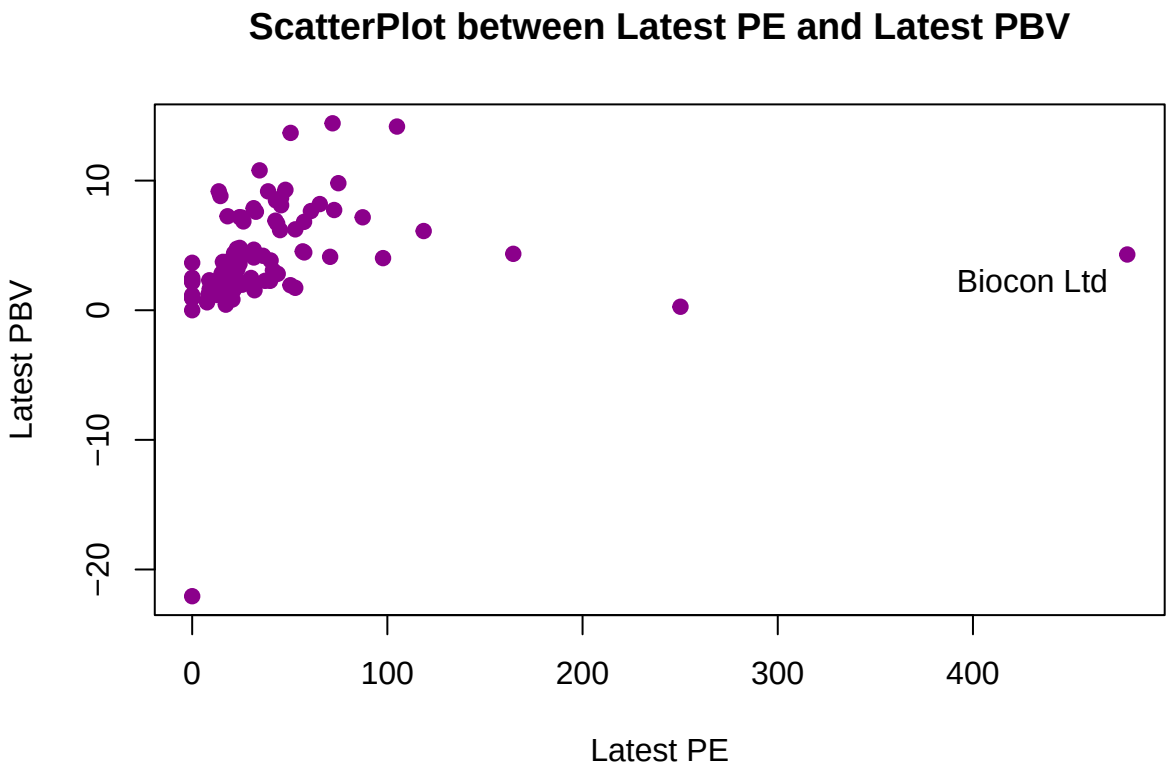
curve and Number of companies established is increasing year by yea.But If we see carefully the plot we can conclude that **Number of companies established are significantly increases year by year after Independence.**

Now we are going to analyze the performances of the company with respect to the different financial measures. So, first we are making the scatterplot between PE and PBV to see whether there is any linear association between PE and PBV.Also,We are interested in the nature of each company in terms of the pair (PE,PBV) to observe that whether the Indian Pharmacy companies follow the similar nature or not.Also We are interested in noticing that if there is any outlier in this plot.

```
library(ggplot2)
install.packages("corrplot", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/gh/x_3x_yf907z7c2w3yx7k55r00000gn/T//RtmpnfmLQ1/downloaded_packages
```

```
plot(Stock_Data$Latest_PE,Stock_Data$Latest_PBV,pch=19,col="magenta4",main="ScatterPlot between Latest
x=Stock_Data$Latest_PE[Stock_Data$Latest_PE>400]
text(x,y=2,pos=2,expression("Biocon Ltd"))
```



From the Scatterplot we observe that there is a clusture in the top left corner . So, the nature of the most of the companies are same in relation between PE and PBV.Also, we can see that "Biocon Ltd" is an exception in this particular category which gives outlier value in relation of PE and PBV in this scatter plot.

we are making the scatterplot between PE and ROCE Percent to see whether there is any linear association between PE and ROCE Percent.Also,We are interested in the nature of each company in terms of the pair (PE,ROCE Percent) to observe that whether the Indian Pharmacy companies follow the similar nature or not.Also We are interested in noticing that if there is any outlier in this plot.

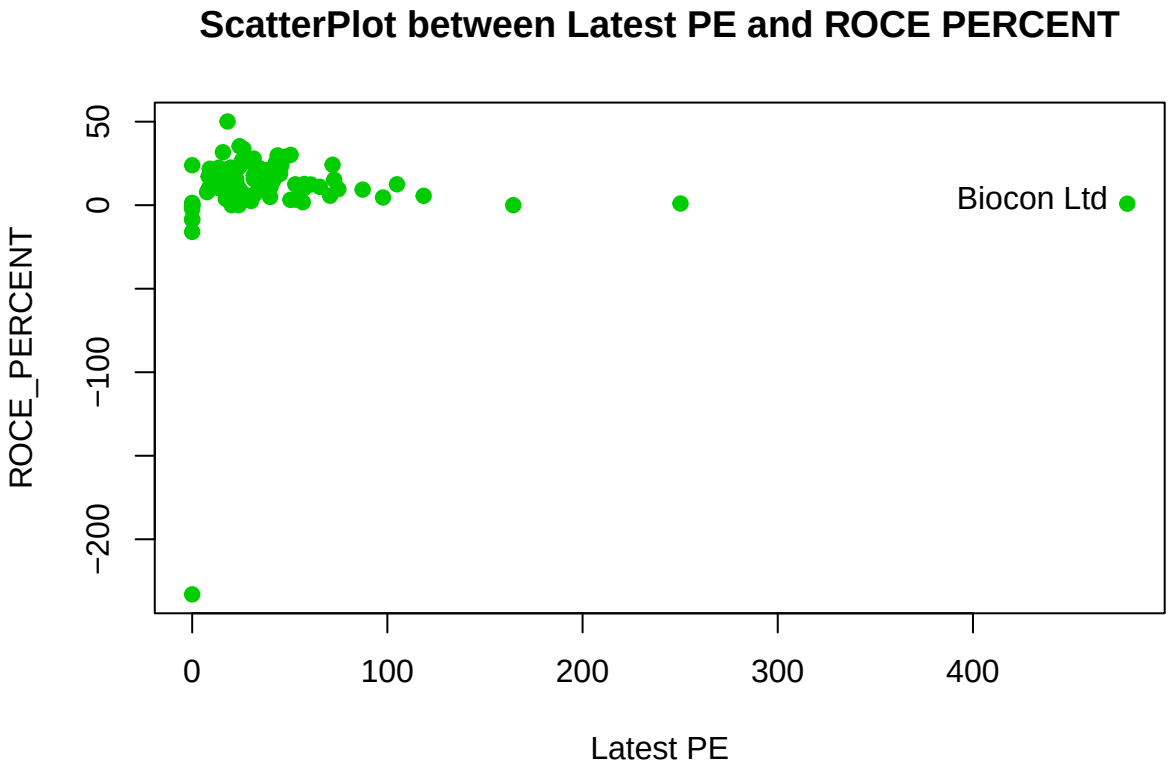
```
library(ggplot2)
install.packages("corrplot", repos = "http://cran.us.r-project.org")
```

```
##
## The downloaded binary packages are in
##  /var/folders/gh/x_3x_yf907z7c2w3yx7k55r00000gn/T//RtmpnfmLQ1/downloaded_packages
```

```
plot(Stock_Data$Latest_PE,Stock_Data$ROCE_percent,pch=19,col="green3",main="ScatterPlot between Latest P
x=Stock_Data$Latest_PE[Stock_Data$Latest_PE>400]
Stock_Data$Latest_PE > 400 & Stock_Data$ROCE_percent > 0
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
text(x,y=2,pos=2,expression("Biocon Ltd"))
```

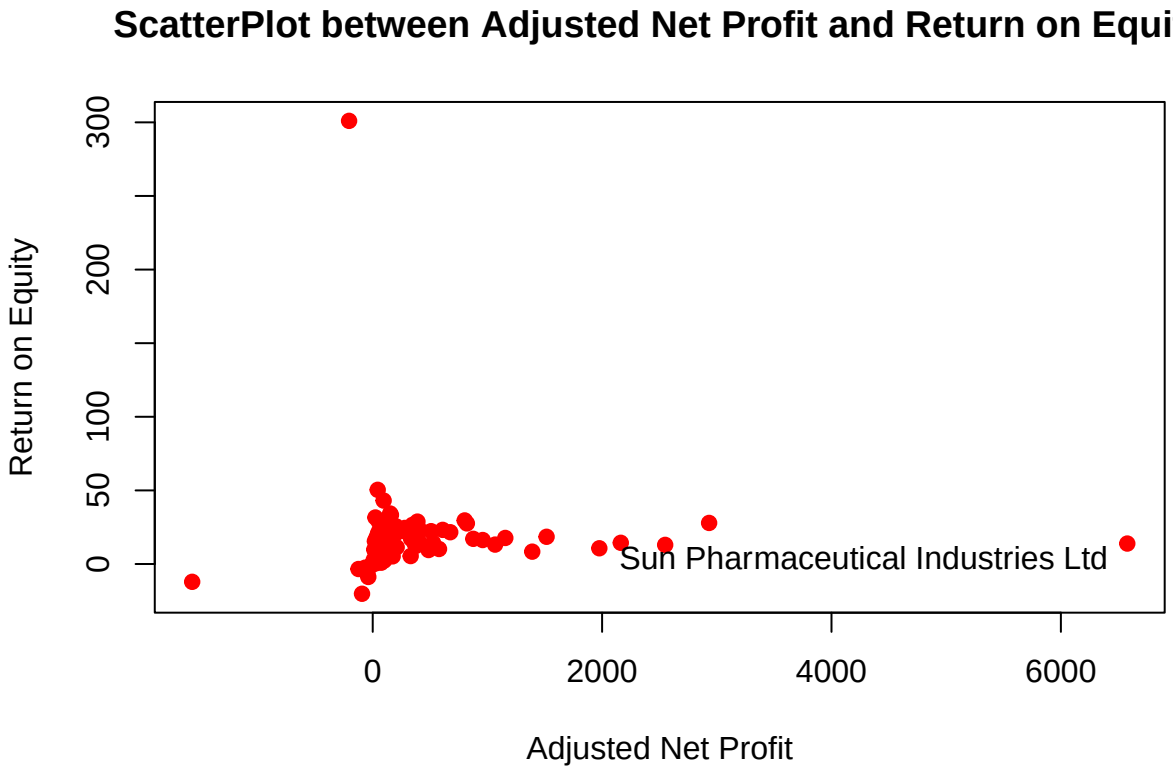


From the Scatterplot we observe that there is a clusture in the top left corner . So, the nature of the most of the companies are same in relation between PE and ROCE Percent.Also, we can see that "Biocon Ltd" is an exception in this particular category which gives outlier value in relation of PE and ROCE Percent in this scatter plot.

we are making the scatterplot between PE and ROCE Percent to see whether there is any linear association between PE and ROCE Percent.Also,We are interested in the nature of each company in terms of the pair (Adjusted Net Profit , Return on Equity) to observe that whether the Indian Pharmacy companies follow the similar nature or not.Also We are interested in noticing that if there is any outlier in this plot.

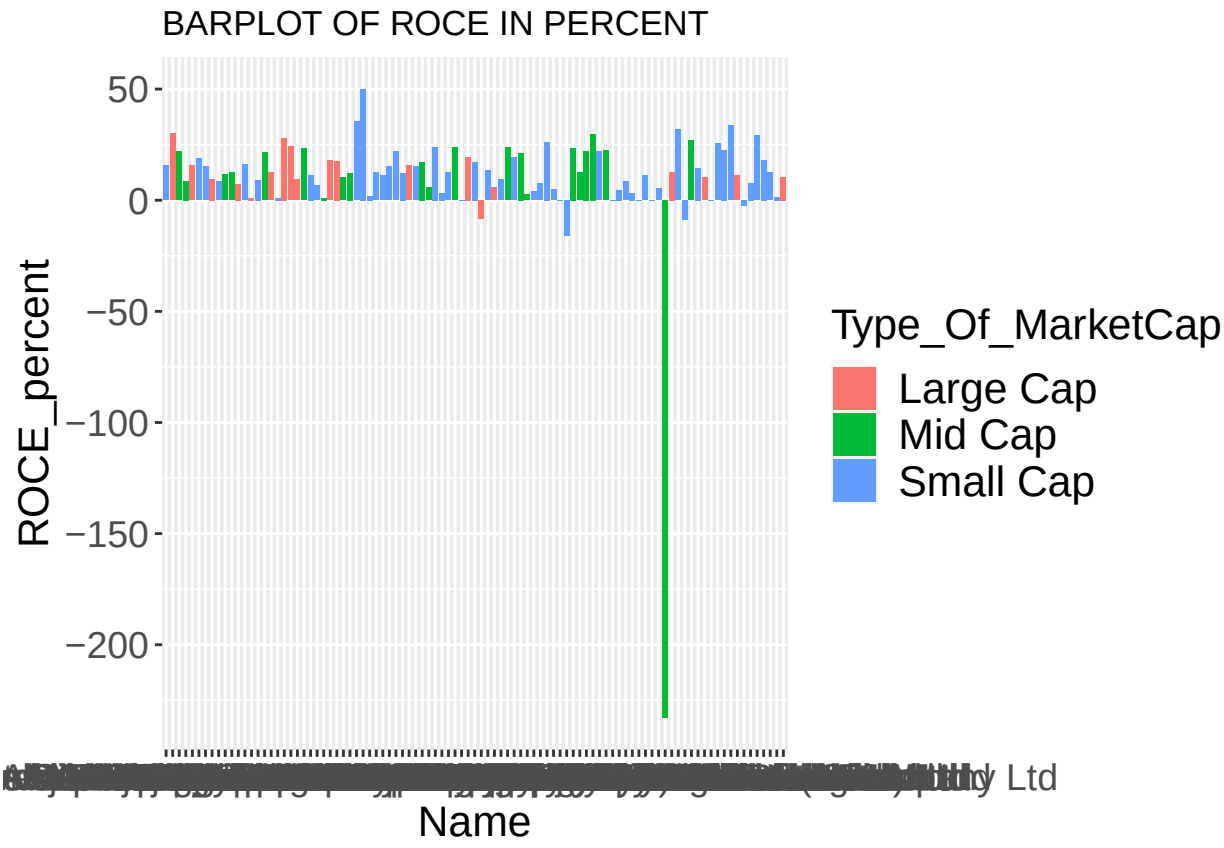
```
plot(Stock_Data$Adjusted_Net_Profit,Stock_Data$Return_On_Equity,pch=19,col="red",main="ScatterPlot betwe
x=Stock_Data$Adjusted_Net_Profit[Stock_Data$Adjusted_Net_Profit>6000]
```

```
text(x,y=2,pos=2,expression("Sun Pharmaceutical Industries Ltd"))
```



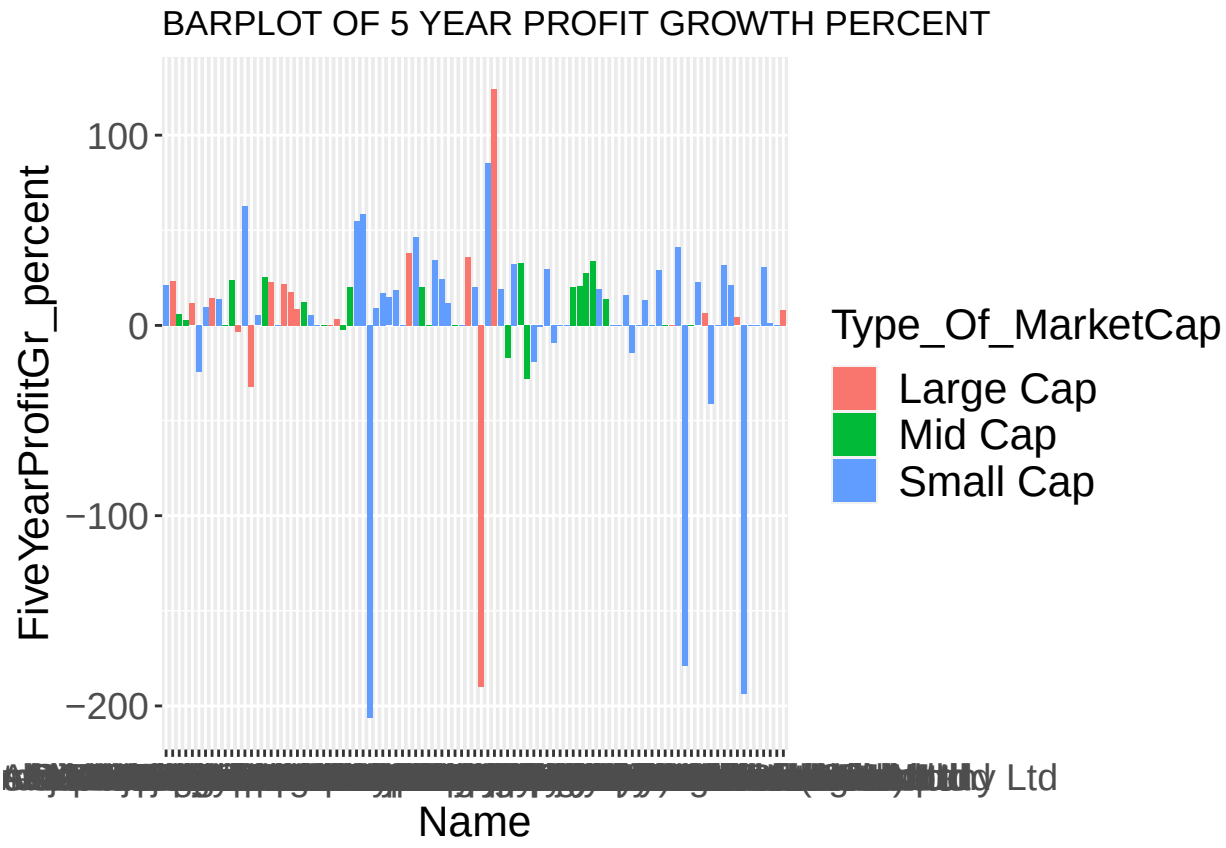
From the Scatterplot we observe that there is a clusture in the bottom. So, the nature of the most of the companies are same in relation between Adjusted Net Profit and Return on Equity.Also, we can see that "Sun Pharmaceutical Ltd" is an exception in this particular category which gives outlier value in relation of Adjusted Net Profit and Return on Equity in this scatter plot.

```
library(ggplot2)
ggplot(Stock_Data,aes(x=Name,y=ROCE_percent))+geom_bar(stat="identity",aes(fill=Type_Of_MarketCap))+labs
```



Here, We can see that ROCE IN PERCENT is moderately high in small cap companies on average rather than Large cap and Mid cap companies. Also, ROCE IN PERCENT is consistently around 25 in the Small Cap Companies.So, the nature of Small Cap Companies are similar in terms of ROCE in Percent. Moreover ROCE in Percent for maximum midcap companies are below 25. There is one extremely low ROCE in Percent value , which belong to a Midcap Company , detected as an outlier value and that breaks the nature of ROCE in percent values of Midcap Companies.

```
library(ggplot2)
 ggplot(Stock_Data,aes(x=Name,y=FiveYearProfitGr_percent))+geom_bar(stat="identity",aes(fill=Type_Of_Ma
```

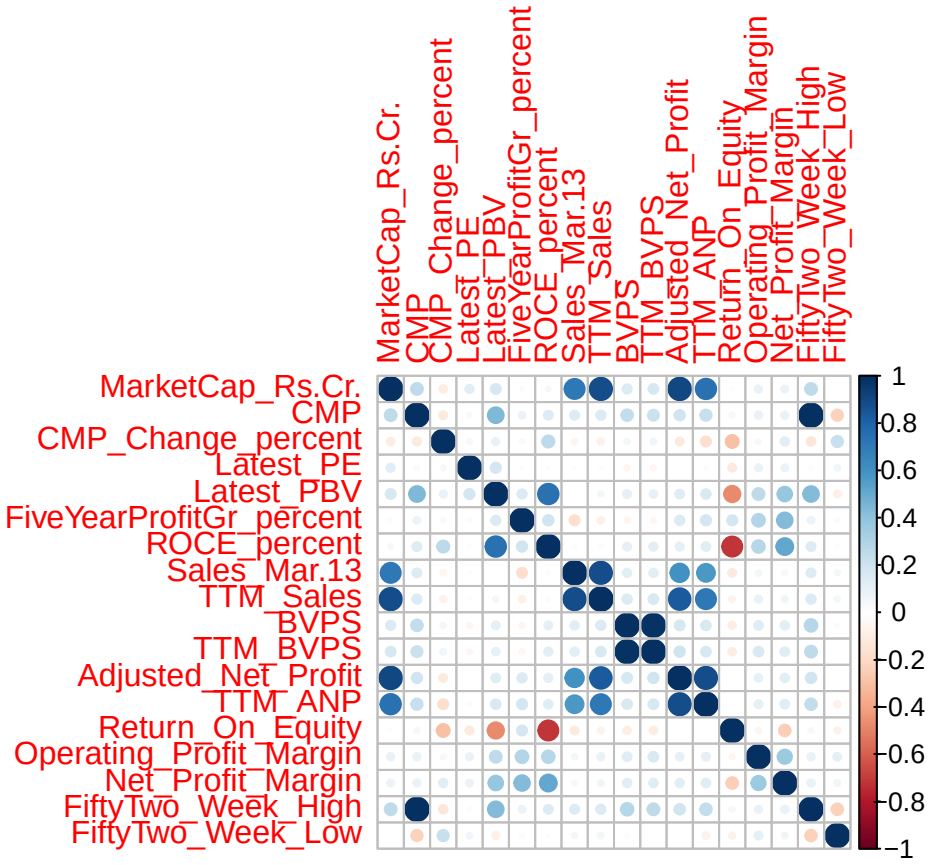


By observing this barplot we can conclude that majority of Small Cap companies in 5 year profit growth percent having very low percent of 5 year profit and also they don't have any similarities in terms of profit growth. But in Mid Cap companies we can see their profit growth percent is more or less similar to each other.So We can say Mid cap companies show similar behaviour in respect of Profit Growth percent.

```
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```
dat<- Stock_Data[,c(4,6:12,22:31)]
corrplot(cor(as.matrix(dat)))
```

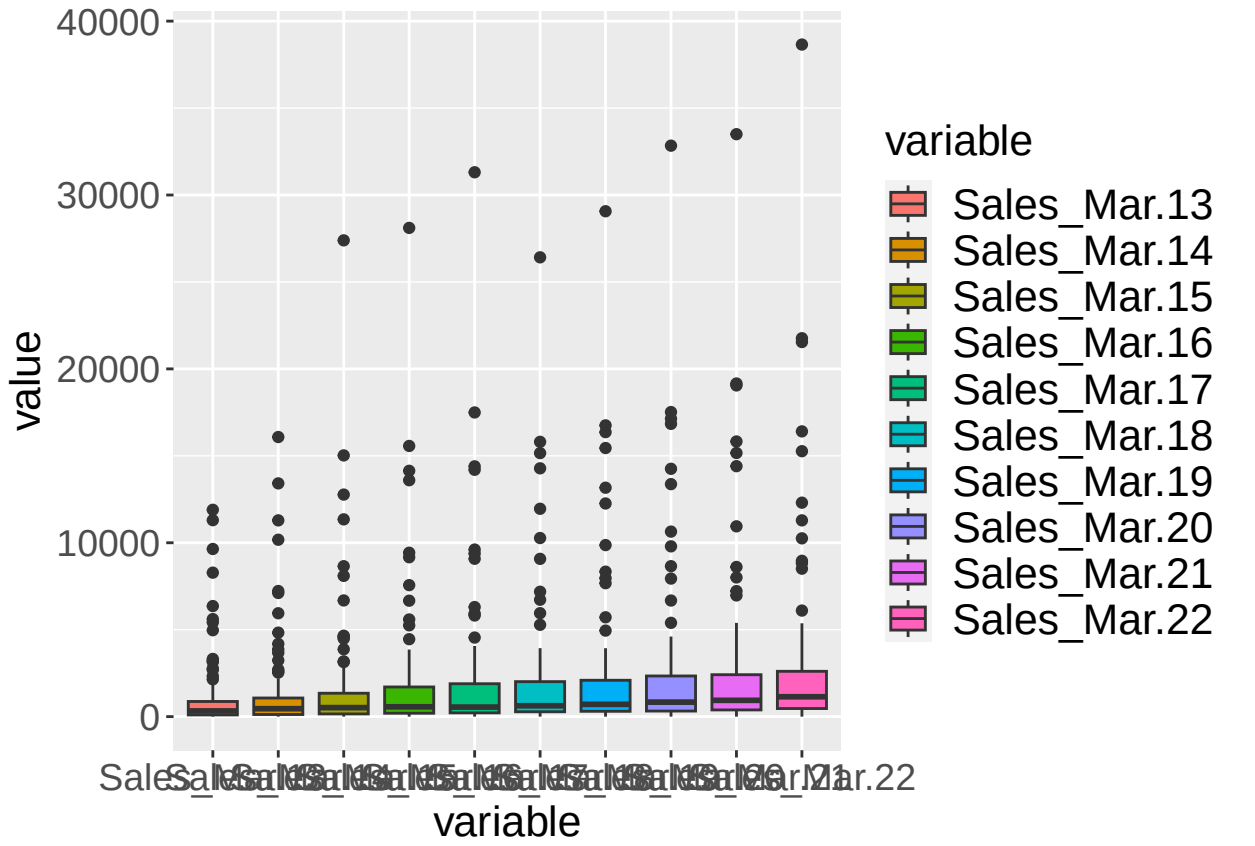


By a careful observation we can see there is a **very high positive correlation** in between CMP and Fifty two week high (correlation coeficient 0.8).That means there is a very strong linear associationship between CMP and Fift Two week high.So, we can say their rate of increaising and decreasing in values are almost same and proportional to each other. TTM Sales and Market Cap are nearly uncorrelated. They have very low positive correlation.So, we can conclude that they merely depend on each other.

```
library(reshape)

s1<-Stock_Data[,c(12:21)]
s<-melt(s1)
```

```
## Using  as id variables
```

```
ggplot(s,aes(x=variable,y=value))+geom_boxplot(aes(fill=variable))+theme(axis.title = element_text(size
```



From year **2013 to 2022** if we observe these **boxplots** for various years ,especially the **outlier** value of sales , we see that all outlier values in various years are **greater than 25000**. Also , there is a nearly linear relationship between these values and which is **monotonically increasing except year 2018**.

Also, if we consider the **COVID times likely in between 2020 and 2021** , we can see that When the covid time was ongoing sales value inreases dramatically from 2019 to 2020 and remained nearly similar or specifically slightly increases in 2021.It's a strong sign that **sale of medicines increases** in covid time.

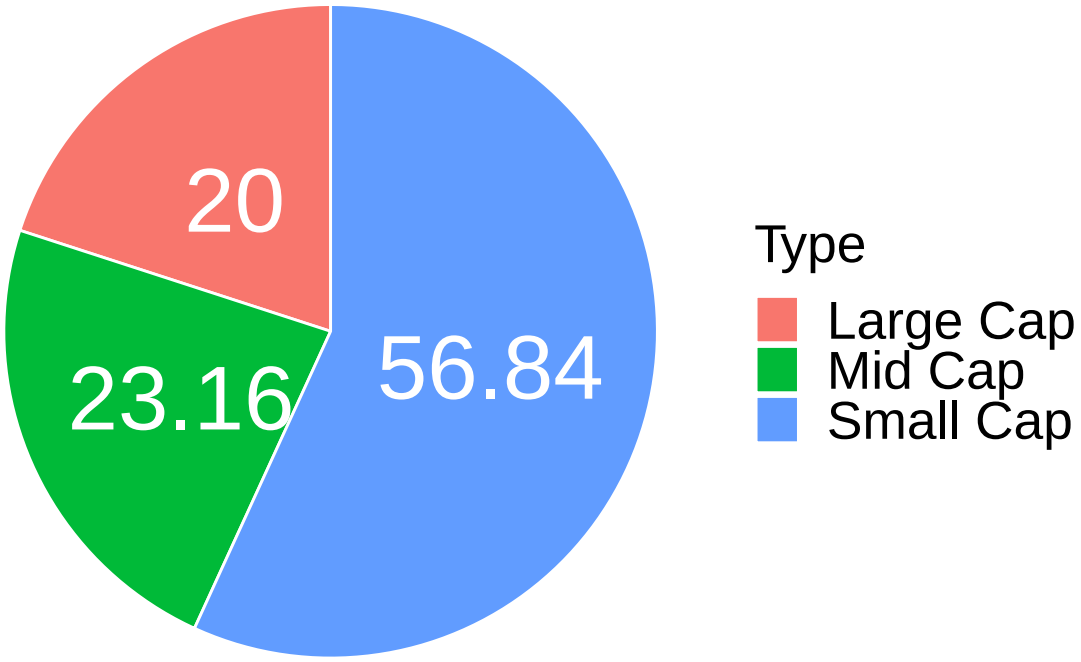
Moreover, from 2021 to 2022 we can observe that sales increased in a very large manner.So, It can also considered as the **effect of COVID 19** ,as our data covers upto the March 22.We all know that , **In the middle of year 2021 , 2nd wave of Covid happened and Third wave of covid happened in India in February to March time ,2022**.

So, This gives the extra effect in this higher values of sale on 2022 , as medicines needed more than normal time on that time by hypothesis. So, It is the **clear indication of having trend** in the dataset.

```
library(ggplot2)
a1 <- length(which(Stock_Data$Type_Of_MarketCap=="Large Cap"))

a2 <- length(which(Stock_Data$Type_Of_MarketCap=="Mid Cap"))

a3 <- length(which(Stock_Data$Type_Of_MarketCap=="Small Cap"))

Type  <- c("Large Cap","Mid Cap","Small Cap")

Percent <- 100*c(a1/nrow(Stock_Data),a2/nrow(Stock_Data),a3/nrow(Stock_Data))

d <- data.frame(Type,Percent)
```

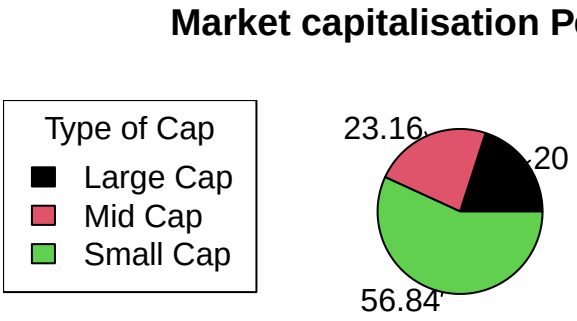
```
ggplot (d, aes(x="", y=Percent, fill=Type)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar("y") +
  theme_void() +
  geom_text(aes( label = round(Percent,2)),position=position_stack(vjust=0.5),size=12, color = "white")+
```

PIE CHART OF TYPE OF MARKET CAP



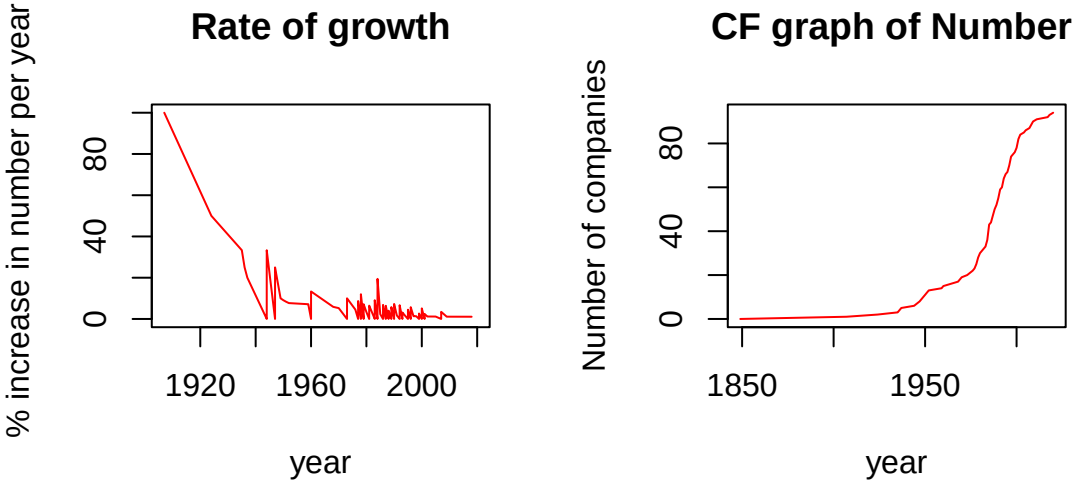
So, From the pie chart it is clear that Maximum number of companies in the Indian Pharmacy Market are Small Cap Companies(about 56.84 %) and Minimum number of companies in the Indian Pharmacy Market are Large Cap Companies(about 20 %).

- **Which Market capitalization types of business are frequent?** So if we look upon the percentage distribution of the market value of the companies. Here we can see the most dominant companies in terms of capitalisation are small cap.



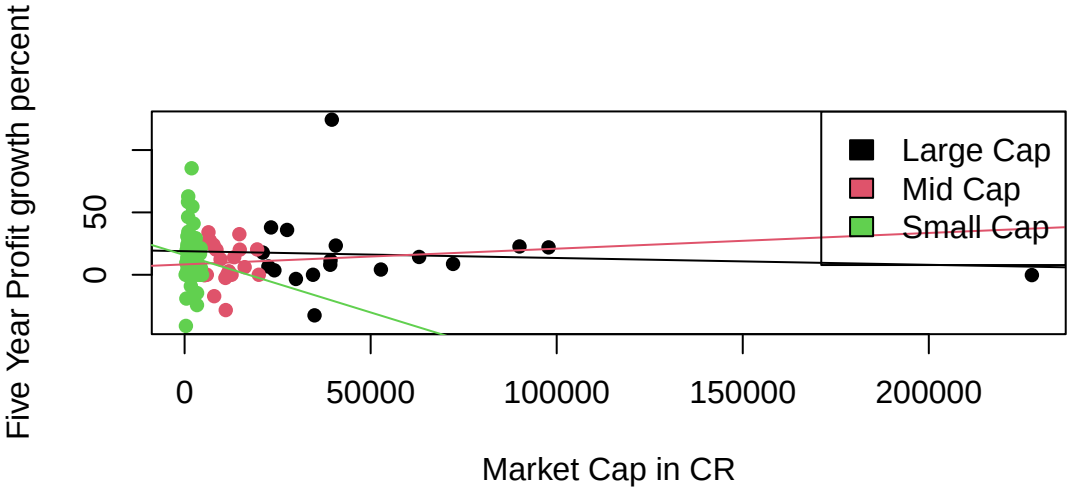
- **Which State has the highest concentration of businesses?** The answer of this question is that Mumbai has 27 company out of total 95 and Andhra Pradesh has the least 2 out of 95total companies as per the data. It also tells that Mumbai is the hub of Pharma industry as it is also the financial capital of India.
- **We tried to find at what rate the pharma industry is growing with years?** It is very important

for to know the scope of the industry growth so here we tried to visualise its growth with respect to time of the establishment from the available data.



We see here that number of companies have increase but still the rate of growth is decrease and the average growth rate of the industry is 5.6.

- **What is the profit growth rate of the companies during the last five years?**



Here the dataset is fitted by suitable straight line. So, we just fit the data set with respect to three types of companies by suitable linear approximations with the help of Linear Regression.

From the plot we can conclude that if we want to invest money for long term about five years in pharmaceutical industry so the best option if you want to stay in the safe side so invest in mid cap or large cap companies as their slope is far better than smallcap.

Shiny App :-

1.Pie Chart :

This shows the pie chart of the type of the market cap i.e., it shows the percentage of the companies having large ,mid and small market capabilities.

2.Analysis Of Stock Variables :

Here, all the stock variables are listed in the side bar panel, and upon clicking on any of them, we will get

two see Barplot, Boxplot and a Histogram along with the summary of the chosen variable. We also have two types of colour options in the barplot and histogram, according to the type of market cap and the year of eastablishment The names of companies having the highest and lowest values of the chosen variable is also written below the barplot.default(

3.Yearly Sales Analysis :

Here we can chose any number of years in the side bar panel to view boxplots, overlapped histogram and frequency polygon so that the sales of different years can be compared.

4.Comparison of Yearly Sales for Different Companies:

Here, the left side bar panel contains the names of companies and the right side bar panel contains years. Upon choosing any number of companies, we get a line diagram for sales for each of those selected companies for all the years mentioned,helping us to compare companies on the basis of sales in these 10 years. On clicking any of the years,in the right side bar panel, the company having the highest and lowest sales among the selected companies for that particulat year comes up as a text.

Final Conclusions :-

1) There is **tremandous increase** in the number of Pharma Companies since independence.
2) The majority of companies headquarter are in Mumbai.
3) The most of companies have shown **upward trend and few went down during the span of 2013 to 2022.**
4) The small cap companies are more in numbers.
5) Small cap companies have **potential to give better returns with a little greater risk** .
6) We can see a **linear trend** in the data of sales over years.
7) We can observe a **huge increment in the sales from 2019 to 2020.** So,We can hypothetically say that **Sales of medicines significantly increased due to COVID 19 1st Wave** in India.
8) Also there is a huge increament of sales from 2021 to 2022. We can also hypothecally say that It was due to Increase in sale of medicine due to **COVID 19 2nd wave and 3rd wave.**

So, here we analyse the growth rate of pharma company is 5.6 which is more than total industry growth rate 1.59% but average growth rate of pharma company is quite higher.The most of the companies are situated in Mumbai, Maharashtra.There is greater risk in small cap companies in investing as compare to mid-cap and large cap but ROCE of small cap is far better for the short duration.

References :-

1) Nifty Pharma Stocks - Nifty Pharma Index Companies | MoneyWorks4me.com(Data)
2) https://www.investopedia.com/(Terminologies)
3) https://en.wikipedia.org/wiki/List_of_largest_companies_in_India(list
4) https://shiny.rstudio.com/gallery/#demos.
5) https://shiny.rstudio.com/tutorial/written-tutorial/lesson1/