Analytics Programming IN R

Presentation

STOCK PRICE ANALYSIS OF FMCG INDUSTRIES

Selected companies for the analysis

- DABUR
- HUL
- ITC
- BRITANNIA

Data collection

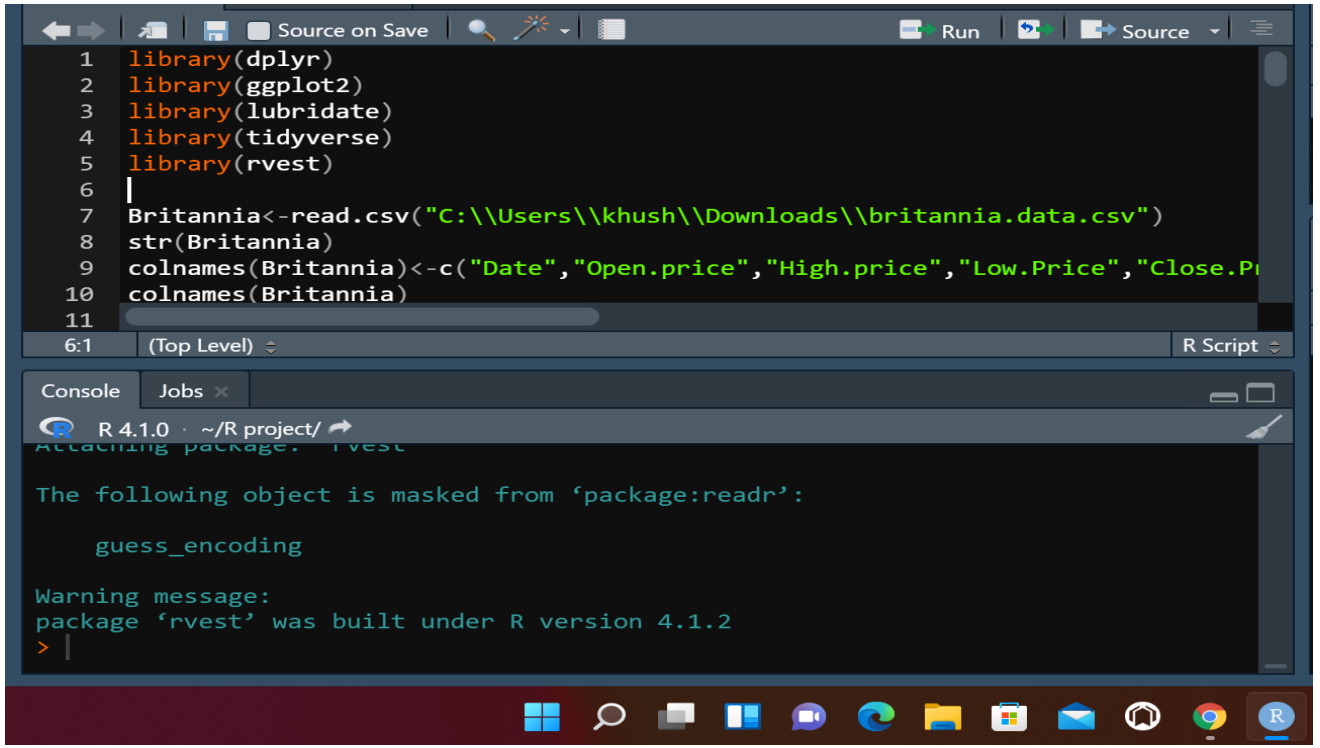
- Data sources : Data for stock price of companies is collected from NSE.
- Data Overview : The Stock prices data collected datewise for the period August 2021 to October 2021. The data collected include the date, Stock open price , Stock close price, day high , day low and Volume Traded for the Period mentioned above.
- Variables : Independent Variable is Date and Independent is Stock prices.

Data Pre-Processing

- Installing libraries and importing Dataset
- Firstly we have installed the packages with install.packages() and after that we use libraries() function to load it for use.
- After installing the libraries we load the company data.

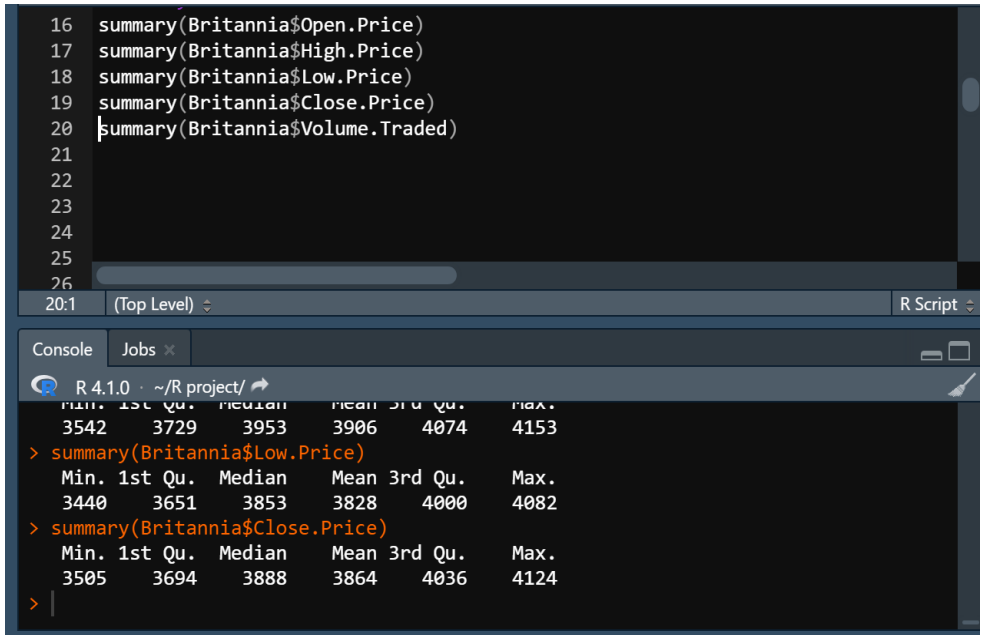
Required libraries

- Library(dplyr)
- Library(ggplot2)
- Library(lubridate)
- Library(tidyverse)
- Library(rvest)



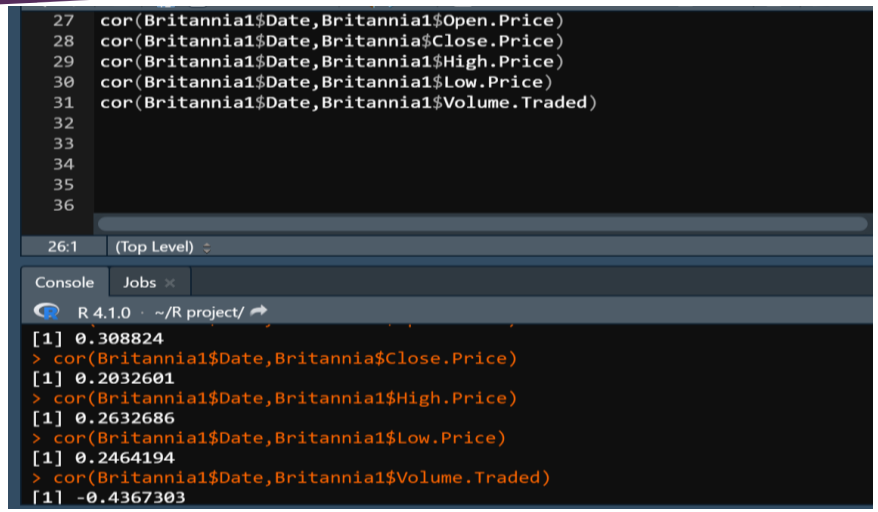
Summary of each Variables

- summary(Britannia$Open.Price)
- summary(Britannia$High.Price)
- summary(Britannia$Low.Price)
- summary(Britannia$Close.Price)
- summary(Britannia$Volume.Traded)



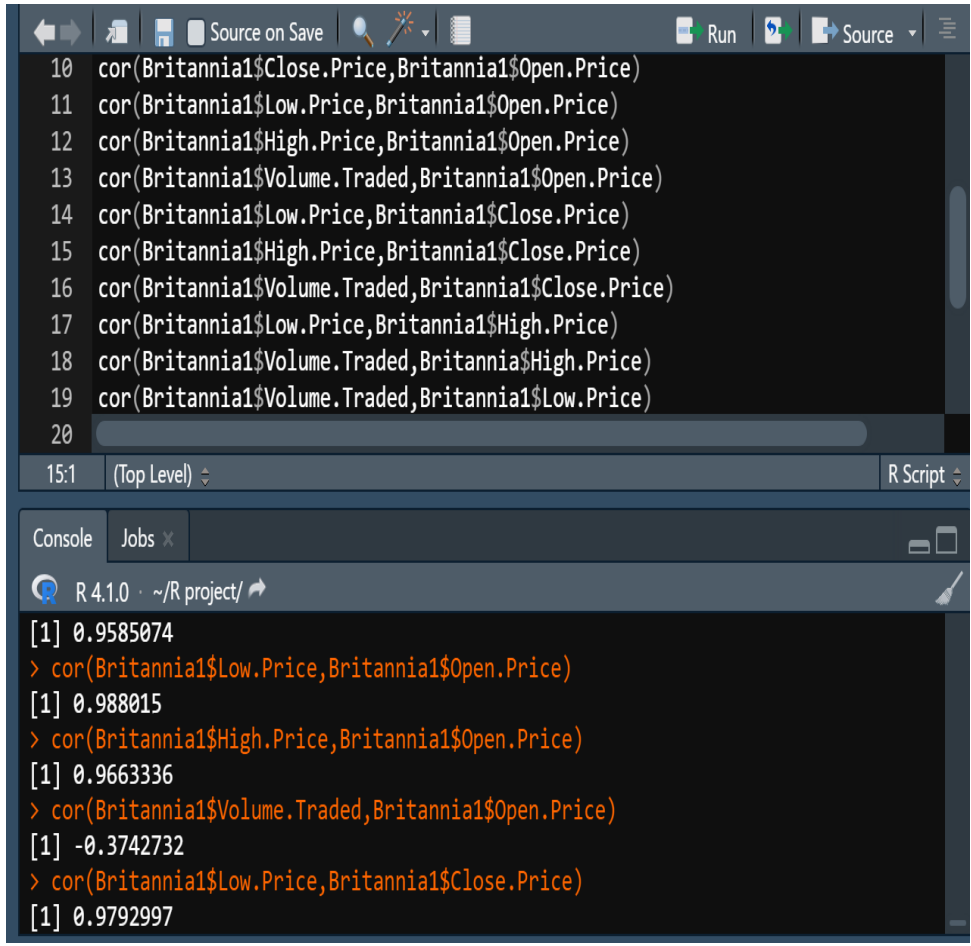
Corelation between date and other variables

- cor(Britannia1$Date,Britannia1$Open.Price)
- cor(Britannia1$Date,Britannia$Close.Price)
- cor(Britannia1$Date,Britannia1$High.Price)
- cor(Britannia1$Date,Britannia1$Low.Price)
- cor(Britannia1$Date,Britannia1$Volume.Traded)



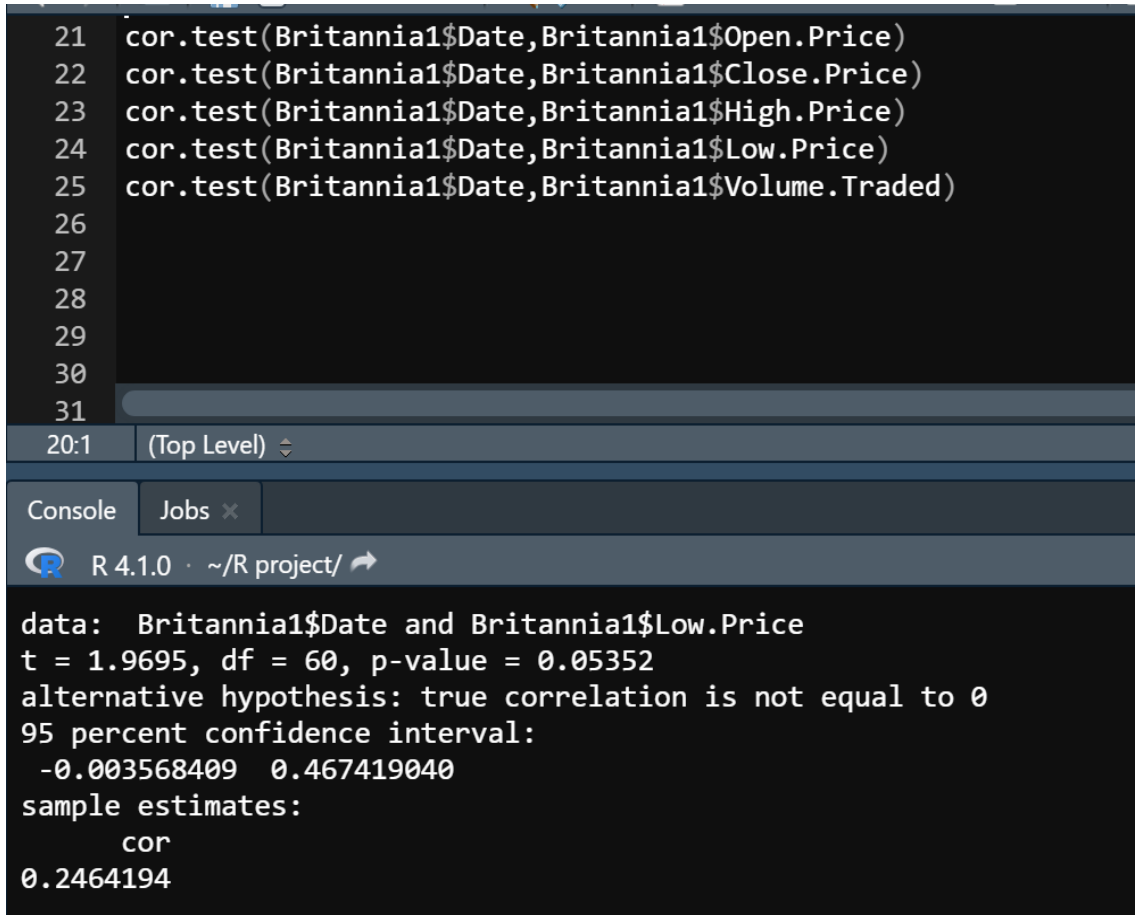
Corelation with different Variables

- cor(Britannia1$Close.Price,Britannia1$Open.Price)
- cor(Britannia1$Low.Price,Britannia1$Open.Price)
- cor(Britannia1$High.Price,Britannia1$Open.Price)
- cor(Britannia1$Volume.Traded,Britannia1$Open.Price)
- cor(Britannia1$Low.Price,Britannia1$Close.Price)
- cor(Britannia1$High.Price,Britannia1$Close.Price)
- cor(Britannia1$Volume.Traded,Britannia1$Close.Price)
- cor(Britannia1$Low.Price,Britannia1$High.Price)
- cor(Britannia1$Volume.Traded,Britannia$High.Price)
- cor(Britannia1$Volume.Traded,Britannia1$Low.Price)



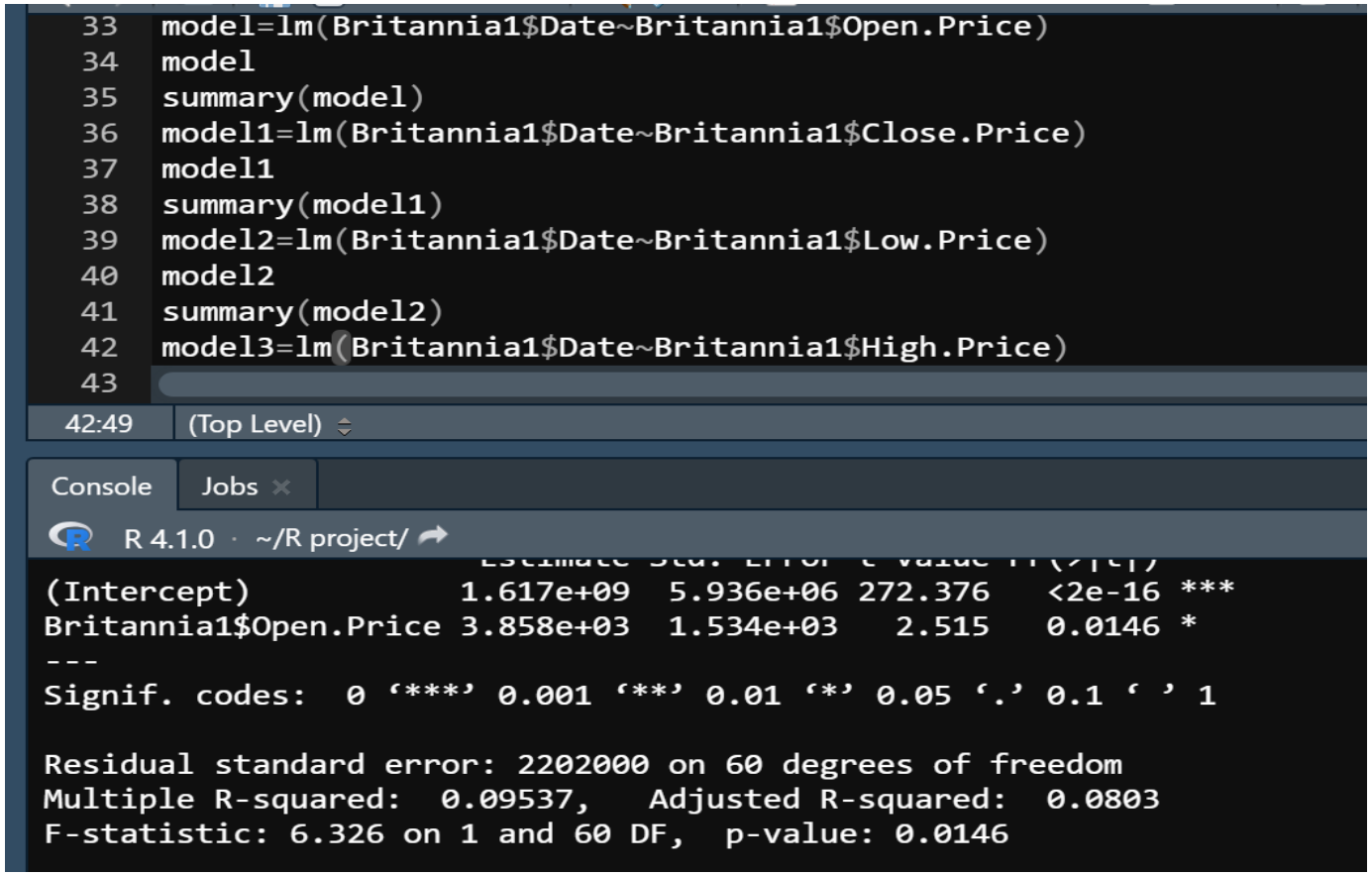
Corelation test with different Variables

- cor.test(Britannia1$Date,Britannia1$Open.Price)
- cor.test(Britannia1$Date,Britannia1$Close.Price)
- cor.test(Britannia1$Date,Britannia1$High.Price)
- cor.test(Britannia1$Date,Britannia1$Low.Price)
- cor.test(Britannia1$Date,Britannia1$Volume.Traded)
- cor.test(Britannia1$Close.Price,Britannia1$Open.Price)
- cor.test(Britannia1$Low.Price,Britannia1$Open.Price)
- cor.test(Britannia1$High.Price,Britannia1$Open.Price)
- cor.test(Britannia1$Volume.Traded,Britannia1$Open.Price)
- cor.test(Britannia1$Low.Price,Britannia1$Close.Price)
- cor.test(Britannia1$High.Price,Britannia1$Close.Price)
- cor.test(Britannia1$Volume.Traded,Britannia1$Close.Price)
- cor.test(Britannia1$Low.Price,Britannia1$High.Price)
- cor.test(Britannia1$Volume.Traded,Britannia1$High.Price)
- cor.test(Britannia1$Volume.Traded,Britannia1$Low.Price)



Linear Regression

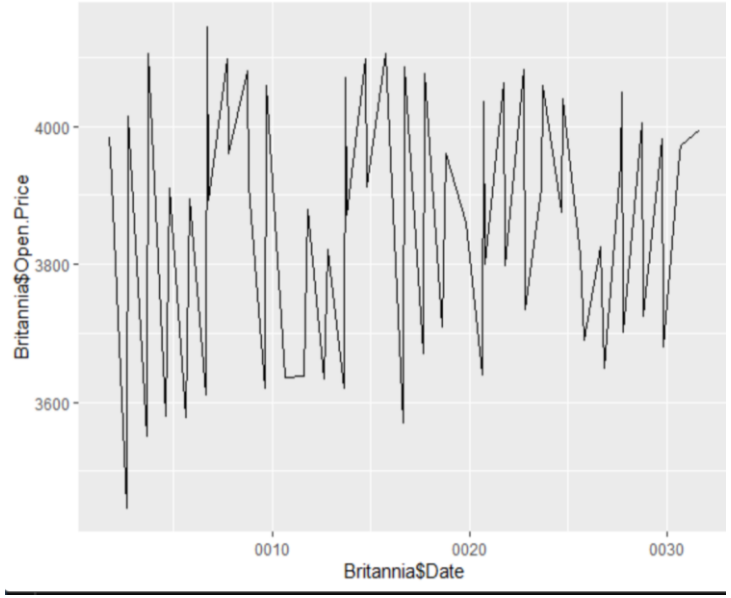
```
model=lm(Britannia1$Date~Britannia1$Open.Price)
model
summary(model)
model1=lm(Britannia1$Date~Britannia1$Close.Price)
model1
summary(model1)
model2=lm(Britannia1$Date~Britannia1$Low.Price)
model2
summary(model2)
model3=lm(Britannia1$Date~Britannia1$High.Price)
model3
summary(model3)
model4=lm(Britannia1$Date~Brritannia1$Close.Price)
model4
summary(model4)
```



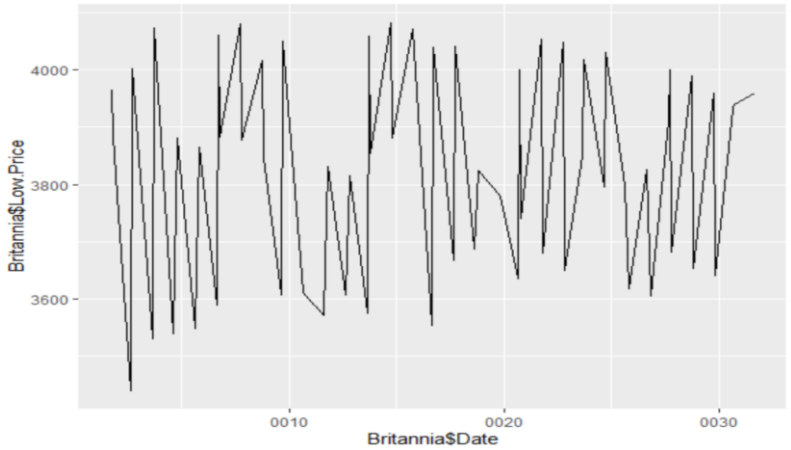
Data Visualization(using ggplot2)

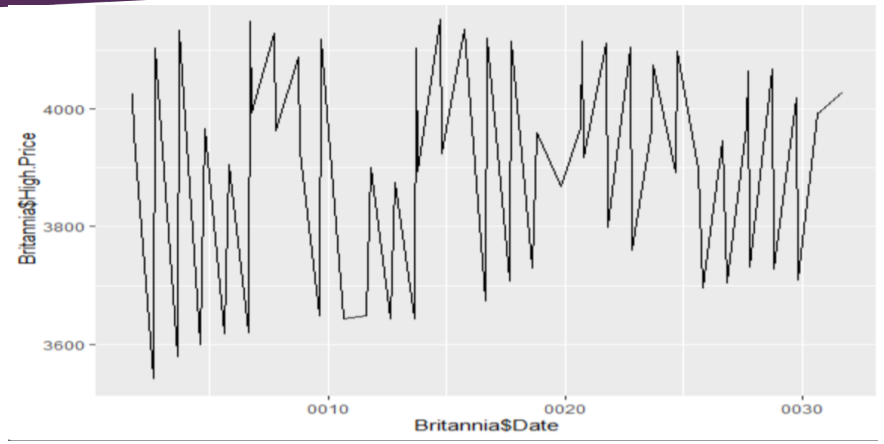
```
Britannia$Date<- as.Date(Britannia$Date,format = "%m/%d/%y")
```

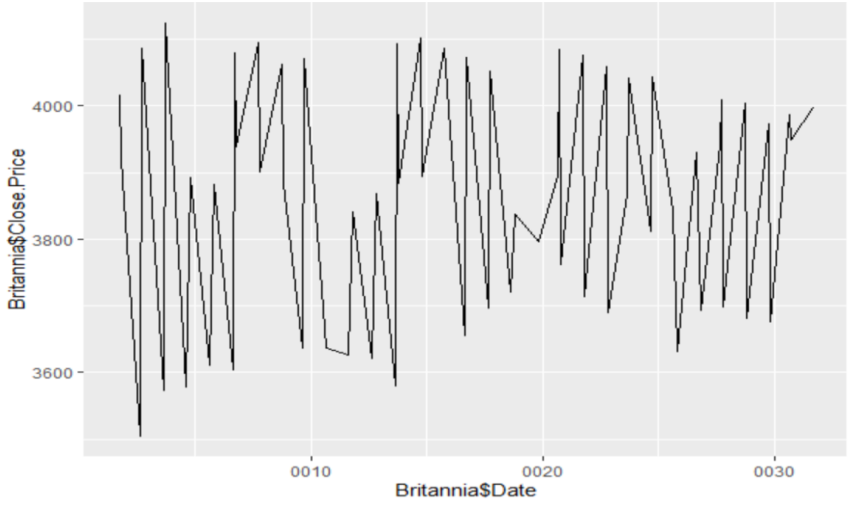
- ggplot(data = Britannia,aes(y=Britannia$Open.Price,x=Britannia$Date)) + geom_line()
- ggplot(data = Britannia,aes(y=Britannia$High.Price,x=Britannia$Date)) + geom_line()
- ggplot(data = Britannia,aes(y=Britannia$Low.Price,x=Britannia$Date)) + geom_line()
- ggplot(data = Britannia,aes(y=Britannia$Close.Price,x=Britannia$Date)) + geom_line()
- ggplot(data = Britannia,aes(y=Britannia$Volume.Traded,x=Britannia$Date)) + geom_line()

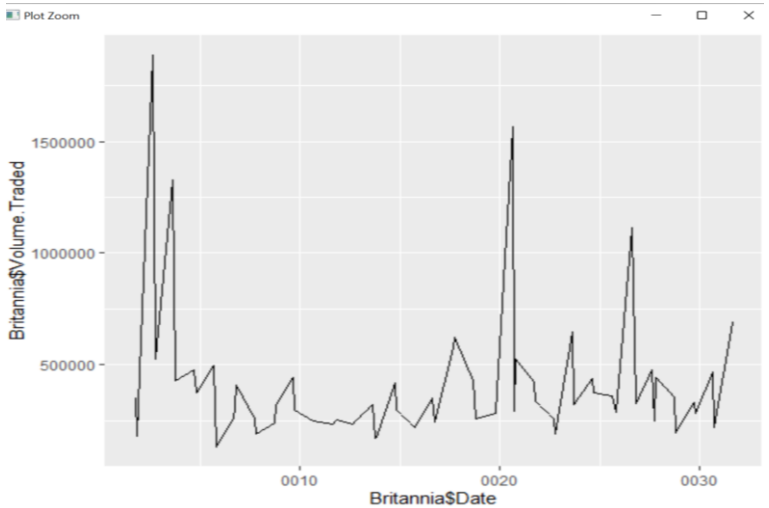


Data Visualization









Web Scraping

```
url="https://www.business-standard.com/company/britannia-inds-93/news"
page1=read_html(url)
page1
news1=page1%>%html_nodes("h2 a")%>%html_text()
news1
date1=page1%>%html_nodes("h2+ p")%>%html_text()
date1
desc1=page1%>%html_nodes(".company-news-listing-txt p+p")%>%html_text()
desc1
df=data.frame(date1,news1,desc1)
View(df)
colnames(df)<-c("Date","news","description")
write.csv(df,"newdf.csv")
```

Add a new variable Category which have numeric value

```
news_data<-read.csv(file.choose())

View(news_data)

names(news_data)

news_data$News.category

news_data$News_new.category=ifelse(news_data$News.category=="positive",
"+1", ifelse(news_data$News.Category=="Negative","-
1",ifelse(news_data$News.Category=="Neutral", "0"," ")))
```

Summary of each Variables

- summary(ITC2$`Open Price`).
- summary(ITC2$`High Price`).
- summary(ITC2$`Low Price`)
- summary(ITC2$`Close Price`)
- summary(ITC2$`Total Traded Quantity`)

By using this function we were able to find the average mean price of ITC share, the inter-quartile range, and the maximum and minimum price the share has reached for different variables

```
> View(ITC2)
> summary(ITC2$`Open Price`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  205.0   210.1   217.1   224.8   236.2   261.4
> summary(ITC2$`High Price`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  206.8   211.6   228.2   227.4   240.8   265.3
> summary(ITC2$`Low Price`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  204.3   208.7   216.3   222.2   235.3   257.4
> summary(ITC2$`Close Price`)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  204.6   209.8   223.2   224.5   237.2   262.6
> summary(ITC2$`Total Traded Quantity`)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  2737937  12287469  18934170  26823750  27167158 149238901
>
```

Corelation between date and other variables

- cor(ITC2$Date,ITC2$`Open Price`)
- cor(ITC2$Date, ITC2$`High Price`)
- cor(ITC2$Date,ITC2$`Low Price`)
- cor(ITC2$Date,ITC2$`Close Price`)
- cor(ITC2$Date, ITC2$`Total Traded Quantity`)

Correlation value is maximum with Open price, hence it is evident that Open price will vary maximum with the date, followed by Low price.

```
> ITC2$Date<- as.numeric(ITC2$Date)
> cor(ITC2$Date,ITC2$`Open Price`)
[1] 0.8336983
> cor(ITC2$Date, ITC2$`High Price`)
[1] 0.8275886
> cor(ITC2$Date,ITC2$`Low Price`)
[1] 0.8332108
> cor(ITC2$Date,ITC2$`Close Price`)
[1] 0.8183432
> cor(ITC2$Date, ITC2$`Total Traded Quantity`)
[1] 0.2576858
```

Corelation with different Variables

- cor(ITC2$`Open Price`, ITC2$`Close Price`)
- cor(ITC2$`Open Price` ,ITC2$`High Price`)
- cor(ITC2$`Open Price`,ITC2$`Low Price`)
- cor(ITC2$`Close Price`, ITC2$`Low Price`)
- cor(ITC2$`Close Price`, ITC2$`High Price`)
- cor(ITC2$`High Price`, ITC2$`Low Price`)
- cor(ITC2$`Open Price`, ITC2$`Total Traded Quantity`)
- cor(ITC2$`Close Price`,ITC2$`Total Traded Quantity`)
- cor(ITC2$`High Price`, ITC2$`Total Traded Quantity`)
- cor(ITC2$`Low Price`, ITC2$`Total Traded Quantity`)

```
> cor(ITC2$`Open Price`, ITC2$`Close Price`)
[1] 0.9681214
> cor(ITC2$`Open Price` ,ITC2$`High Price`)
[1] 0.984748
> cor(ITC2$`Open Price`,ITC2$`Low Price`)
[1] 0.9869802
> cor(ITC2$`Close Price`, ITC2$`Low Price`)
[1] 0.9862493
> cor(ITC2$`Close Price`, ITC2$`High Price`)
[1] 0.9885074
> cor(ITC2$`High Price`, ITC2$`Low Price`)
[1] 0.9817918
> cor(ITC2$`Open Price`, ITC2$`Total Traded Quantity`)
[1] 0.3472648
> cor(ITC2$`Close Price`,ITC2$`Total Traded Quantity`)
[1] 0.4446186
> cor(ITC2$`High Price`, ITC2$`Total Traded Quantity`)
[1] 0.4712786
> cor(ITC2$`Low Price`, ITC2$`Total Traded Quantity`)
[1] 0.3355868
```

Corelation test between date and other Variables

- cor.test(ITC2$Date, ITC2$`Open Price`)
- cor.test(ITC2$Date, ITC2$`Close Price`)
- cor.test(ITC2$Date, ITC2$`High Price`)
- cor.test(ITC2$Date, ITC2$`Low Price`)
- cor.test(ITC2$Date, ITC2$`Total Traded Quantity`)

```
> cor.test(ITC2$Date, ITC2$`Open Price`)

        Pearson's product-moment correlation

data:  ITC2$Date and ITC2$`Open Price`
t = 11.791, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7384732 0.8963186
sample estimates:
      cor
0.8336983

> cor.test(ITC2$Date, ITC2$`Close Price`)

        Pearson's product-moment correlation

data:  ITC2$Date and ITC2$`Close Price`
t = 11.121, df = 61, p-value = 2.621e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7156890 0.8863872
sample estimates:
```

```
> cor.test(ITC2$`Open Price`, ITC2$`Close Price`)

        Pearson's product-moment correlation

data:  ITC2$`Open Price` and ITC2$`Close Price`
t = 30.187, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.947671 0.980659
sample estimates:
      cor
0.9681214

> cor.test(ITC2$`Open Price` ,ITC2$`High Price`)

        Pearson's product-moment correlation

data:  ITC2$`Open Price` and ITC2$`High Price`
t = 44.205, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9748273 0.9907772
sample estimates:
      cor
0.984748
```

```
> cor.test(ITC2$Date, ITC2$`High Price`)

        Pearson's product-moment correlation

data:  ITC2$Date and ITC2$`High Price`
t = 11.515, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7293814 0.8923744
sample estimates:
      cor
0.8275886

> cor.test(ITC2$Date, ITC2$`Low Price`)

        Pearson's product-moment correlation

data:  ITC2$Date and ITC2$`Low Price`
t = 11.769, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.7377465 0.8960042
sample estimates:
```

Corelation test with different Variables

- cor.test(ITC2$`Open Price`, ITC2$`Close Price`)
- cor.test(ITC2$`Open Price` ,ITC2$`High Price`)
- cor.test(ITC2$`Open Price`, ITC2$`Low Price`)
- cor.test(ITC2$`Close Price`, ITC2$`Low Price`)
- cor.test(ITC2$`Close Price`, ITC2$`High Price`)
- cor.test(ITC2$`High Price`, ITC2$`Low Price`)
- cor.test(ITC2$`Open Price`, ITC2$`Total Traded Quantity`)
- cor.test(ITC2$`Close Price`, ITC2$`Total Traded Quantity`)
- cor.test(ITC2$`High Price`, ITC2$`Total Traded Quantity`)
- cor.test(ITC2$`Low Price`, ITC2$`Total Traded Quantity`)

```
> cor.test(ITC2$`Close Price`, ITC2$`High Price`)

        Pearson's product-moment correlation

data:  ITC2$`Close Price` and ITC2$`High Price`
t = 51.071, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9810086 0.9930556
sample estimates:
      cor
0.9885074

> cor.test(ITC2$`High Price`, ITC2$`Low Price`)

        Pearson's product-moment correlation

data:  ITC2$`High Price` and ITC2$`Low Price`
t = 40.367, df = 61, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9699774 0.9889831
sample estimates:
      cor
0.9817918
```

```
> cor.test(ITC2$`Open Price`, ITC2$`Total Traded Quantity`)

        Pearson's product-moment correlation

data:  ITC2$`Open Price` and ITC2$`Total Traded Quantity`
t = 2.8922, df = 61, p-value = 0.005296
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1088667 0.5478894
sample estimates:
      cor
0.3472648

> cor.test(ITC2$`Close Price`, ITC2$`Total Traded Quantity`)

        Pearson's product-moment correlation

data:  ITC2$`Close Price` and ITC2$`Total Traded Quantity`
t = 3.8769, df = 61, p-value = 0.0002615
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2212238 0.6236786
sample estimates:
      cor
0.4446186
```

Linear Regression

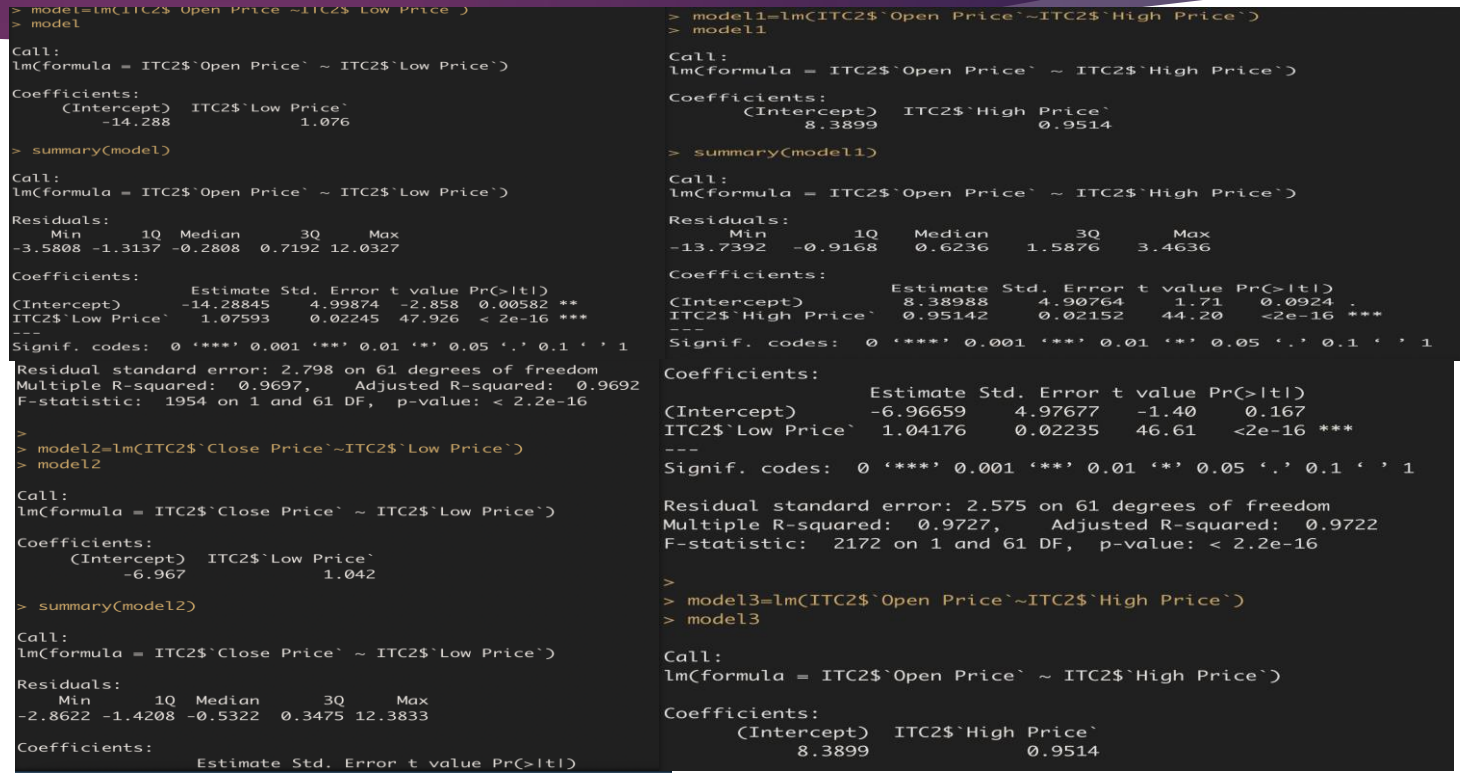
- model=lm(ITC2$`Open Price`~ITC2$`Low Price`)
- model
- summary(model)

- model1=lm(ITC2$`Open Price`~ITC2$`High Price`)
- model1
- summary(model1)

- model2=lm(ITC2$`Close Price`~ITC2$`Low Price`)
- model2
- summary(model2)

- model3=lm(ITC2$`Open Price`~ITC2$`High Price`)
- model3
- summary(model3)

- model4=lm(ITC2$`High Price`~ITC2$`Total Traded Quantity`)
- model4
- summary(model4)



```
> model=lm(ITC2$`Open Price`~ITC2$`Low Price`)
> model

Call:
lm(formula = ITC2$`Open Price` ~ ITC2$`Low Price`)

Coefficients:
     (Intercept)  ITC2$`Low Price`
         -14.288            1.076

> summary(model)

Call:
lm(formula = ITC2$`Open Price` ~ ITC2$`Low Price`)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5808 -1.3137 -0.2808  0.7192 12.0327

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -14.28845    4.99874  -2.858  0.00582 **
ITC2$`Low Price`   1.07593    0.02245  47.926  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.798 on 61 degrees of freedom
Multiple R-squared:  0.9697,    Adjusted R-squared:  0.9692
F-statistic:  1954 on 1 and 61 DF,  p-value: < 2.2e-16

>
> model2=lm(ITC2$`Close Price`~ITC2$`Low Price`)
> model2

Call:
lm(formula = ITC2$`Close Price` ~ ITC2$`Low Price`)

Coefficients:
     (Intercept)  ITC2$`Low Price`
          -6.967            1.042

> summary(model2)

Call:
lm(formula = ITC2$`Close Price` ~ ITC2$`Low Price`)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8622 -1.4208 -0.5322  0.3475 12.3833

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
```

```
> model1=lm(ITC2$`Open Price`~ITC2$`High Price`)
> model1

Call:
lm(formula = ITC2$`Open Price` ~ ITC2$`High Price`)

Coefficients:
     (Intercept)  ITC2$`High Price`
          8.3899            0.9514

> summary(model1)

Call:
lm(formula = ITC2$`Open Price` ~ ITC2$`High Price`)

Residuals:
     Min       1Q   Median       3Q      Max
-13.7392  -0.9168   0.6236   1.5876   3.4636

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      8.38988    4.90764    1.71   0.0924 .
ITC2$`High Price`  0.95142    0.02152   44.20   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -6.96659    4.97677   -1.40    0.167
ITC2$`Low Price`   1.04176    0.02235   46.61   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.575 on 61 degrees of freedom
Multiple R-squared:  0.9727,    Adjusted R-squared:  0.9722
F-statistic:  2172 on 1 and 61 DF,  p-value: < 2.2e-16

>
> model3=lm(ITC2$`Open Price`~ITC2$`High Price`)
> model3

Call:
lm(formula = ITC2$`Open Price` ~ ITC2$`High Price`)

Coefficients:
     (Intercept)  ITC2$`High Price`
          8.3899            0.9514
```

Highest Impacting Variable

Since R squared value of Open price and Low Price is the highest i.e, **0.9741** so Opening Price impacts Lowest Price the most.

Least Impacting variable

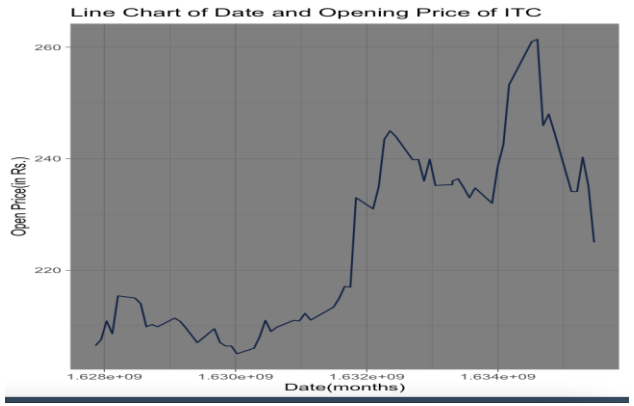
Since R squared value of High price and total traded quantity is the lowest i.e, **0.2221** , so High price impact total traded quantity the least.

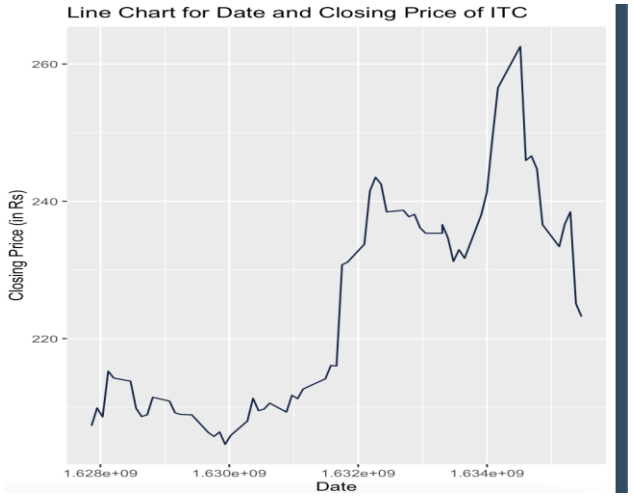
Data Visualization(using ggplot2)

```
ggplot(ITC2) +
  aes(x = Date, y = `Open Price`) +
  geom_line(size = 0.5, colour = "#112446") +
  labs(x = "Date(months)", y = "Open Price(in Rs.)", title = "Line Chart of Date and Opening Price of ITC") +
  theme_dark()
ggplot(ITC2) +
  aes(x = Date, y = `Close Price`) +
  geom_line(size = 0.5, colour = "#112446") +
  labs(x = "Date", y = "Closing Price (in Rs)", title = "Line Chart for Date and Closing Price of ITC") +
  theme_gray()

#linechart Between Date and Highest Price
library(ggplot2)

ggplot(ITC2) +
  aes(x = Date, y = `High Price`) +
  geom_line(size = 0.5, colour = "#112446") +
  labs(x = "Date", y = "Highest Price(in Rs)", title = "Date and Highest Price of ITC") +
  theme_gray()
```





Data Visualization(using ggplot2)

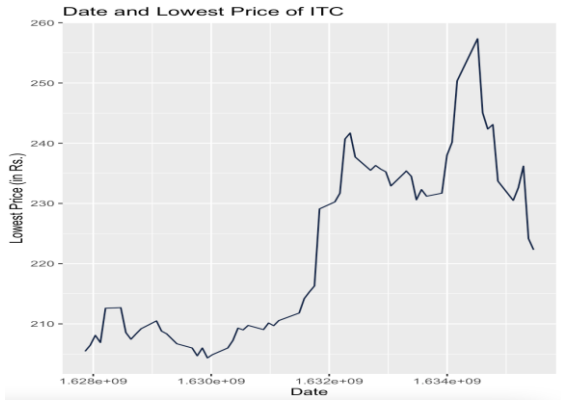
```
#linechart Between Date and Highest Price

library(ggplot2)

ggplot(ITC2) +
  aes(x = Date, y = `High Price`) +
  geom_line(size = 0.5, colour = "#112446") +
  labs(x = "Date", y = "Highest Price(in Rs)", title = "Date and Highest Price of ITC") +
  theme_gray()
```


```
#linechart Between Date and Lowest Price
library(ggplot2)

ggplot(ITC2) +
  aes(x = Date, y = `Low Price`) +
  geom_line(size = 0.5, colour = "#112446") +
  labs(x = "Date",
       y = "Lowest Price (in Rs.)", title = "Date and Lowest Price of ITC") +
  theme_gray()
```



Web Scraping

- url1="https://www.business-standard.com/advance-search?type=news&c-range=range&range=bwn_dates&from_date=01-08-2021&to_date=31-10-2021&c-headline=headline&c-cname=cname&cname=ITC&company=301"
- page=read_html(url1)
- page
- news1=page%>%html_nodes("h2 a")%>%html_text()
- news1
- date1=page%>%html_nodes(".fL+ p")%>%html_text()
- date1
- desc1=page%>%html_nodes("h2+ p")%>%html_text()
- desc1
- df=data.frame(date1,news1,desc1)
- View(df)
- colnames(df)<-c("date","news","description")

- write.csv(df,"ITCNewscsv1")

Add a new variable Category which have numeric value

- data=read.csv(file.choose(),header=TRUE,sep=",")
- df=as.data.frame(data)
- df$News_category<-ifelse(df$category=="positive", "+1", ifelse(df$category=="negative","-1",
- ifelse(df$category=="neutral", "0", "")))
- View(df)

	DATE	NEWS	DESCRIPTION	category	News_category
1	11-Aug-21	ITC to launch super app this year to tap small farmers...	The product will offer a wide range of agricultural sol...	positive	+1
2	12-Aug-21	ITC to invest $2 billion across segments in medium-t...	Super app 'ITC-MAARS' among new growth areas; inv...	neutral	0
3	03-Sep-21	ITC, Sun Pharma, Tata Steel: Strategies for FY22 Sens...	Here are trading strategies for some of the best and ...	negative	-1
4	07-Sep-21	Here's why CLSA is bullish on ITC; sees the stock Rs 2...	Given the road ahead for the next few years, CLSA fin...	positive	+1
5	15-Sep-21	ITC surges 8%, hits seven-month high; inches toward...	Despite today's sharpest rally, the stock has under-p...	negative	-1
6	16-Sep-21	ITC can rally 30% to Rs 300 levels in the days ahead, c...	The short-to-medium term outlook for the stock will ...	neutral	0
7	20-Sep-21	ITC hits 52-week high as Jefferies raises target; stock...	Analysts believe ITC's cigarettes business will fully rec...	positive	+1
8	22-Sep-21	ITC regains Rs 3-trillion market cap; stock hits 21-m...	In the past one week, ITC has outperformed the mark...	positive	+1
9	29-Sep-21	Earnings of cement, FMCG, aviation cos at risk as oil, ...	In a recent report, analysts at Goldman Sachs suggest...	negative	-1
10	14-Oct-21	ITC nears two-year high, up 11% in four days; can hit ...	Analysts believe ITC's cigarettes business will fully rec...	neutral	0
11	20-Oct-21	FMCG stks dip amid waning demand concern; ITC, HU...	The S&P BSE FMCG index, which fell 1.62 per cent in i...	negative	-1
12	26-Oct-21	ITC Q2 preview: Analysts see 13% YoY revenue jump; ...	On the bourses, the stock of the cigarette-to-hotels c...	positive	+1
13	27-Oct-21	ITC consolidated PAT up 10% to Rs 3,763 crore in Sep...	Revenues from operations were at Rs 14,844.38 crore...	positive	+1
14	28-Oct-21	ITC dips 5% post Q2 results; stock corrects 15% from ...	A stable taxation policy remains key to sustaining ste...	neutral	0
15	29-Oct-21	Weak cigarette volumes, margins in Q2 weigh on ITC,...	Decision by the panel on tobacco tax remains the key...	negative	-1

JOINING BOTH THE DATA SET

- library(dplyr)
- merged_news <- inner_join(ITC2,df,by.x=Date,by.y=DATE)
- merged_news<-subset(merged_news,select=-c(HighPrice,LowPrice,TotalTradedQuantity,Turnover,No.of Trades,Deliverable Qty,X..Dly.Qt.to.Traded.Qty,NEWS,DESCRIPTION,category))
- merged_news<-subset(merged_news,select=-c(PrevClose),LastPrice)

- merged_news<-subset(merged_news,select=-c(PrevClose,LastPrice))
- model=lm(merged_news$New_category~merged_news$AveragePrice)
- summary(model)