R PROJECT REPORT

#(1-5)

Load the data in RStudio using appropriate functions and tools.

install.packages("dplyr")

install.packages("ggplot2")

install.packages("lubridate")

library(dplyr)

library(ggplot2)

library(lubridate)

Britannia<-read.csv("C:\\Users\\khush\\Downloads\\britannia.data.csv")

str(Britannia)

colnames(Britannia)<-c("Date","Open.price","High.price","Low.Price","Close.Price","Volume.Traded")

colnames(Britannia)

View(Britannia)

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

#6)

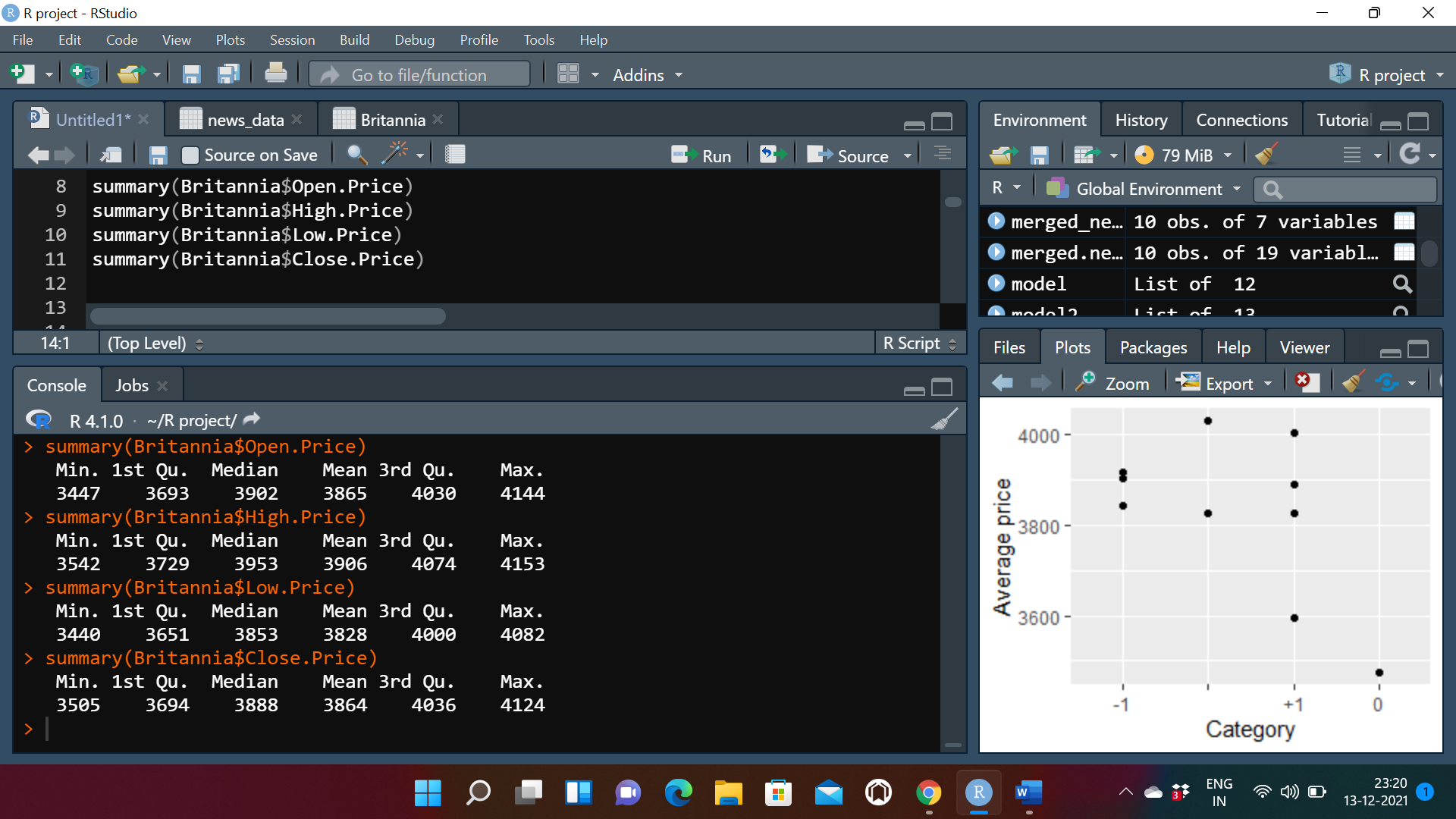
summary(Britannia$Open.Price)

summary(Britannia$High.Price)

summary(Britannia$Low.Price)

summary(Britannia$Close.Price)

summary(Britannia$Volume.Traded)



#(7-8)

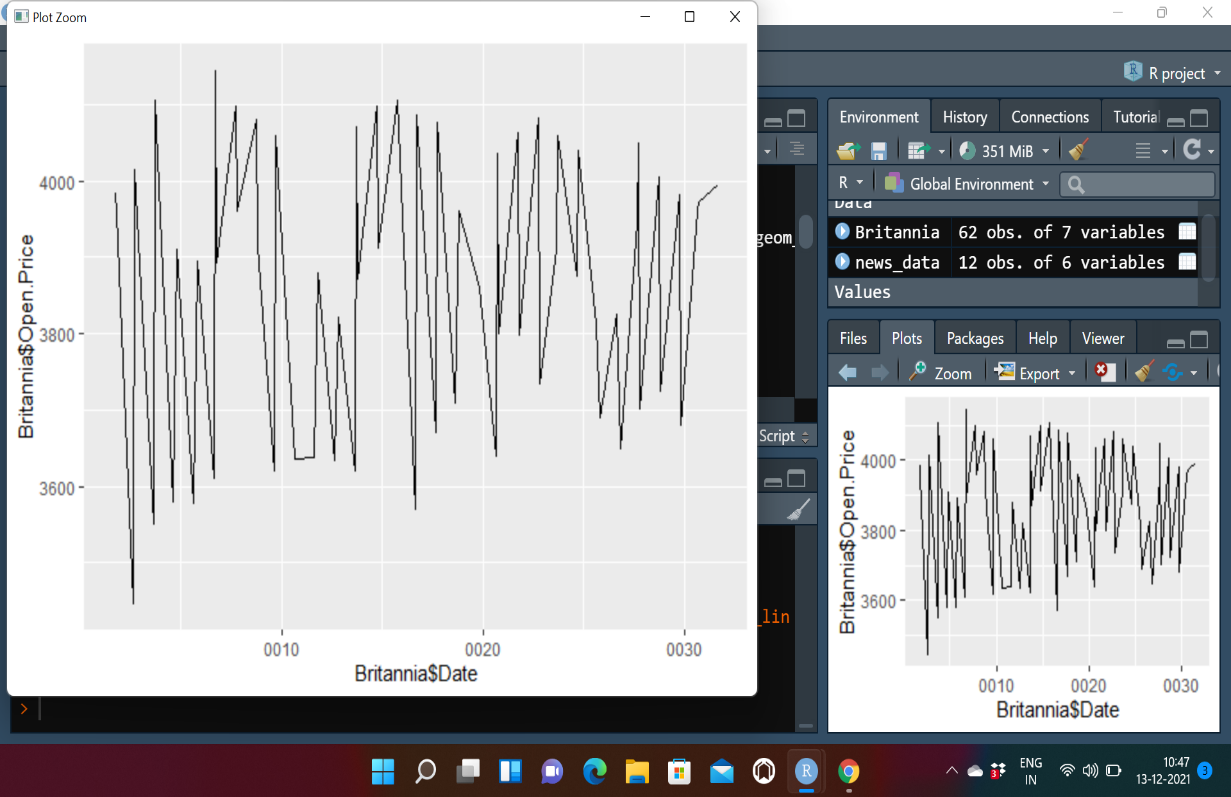
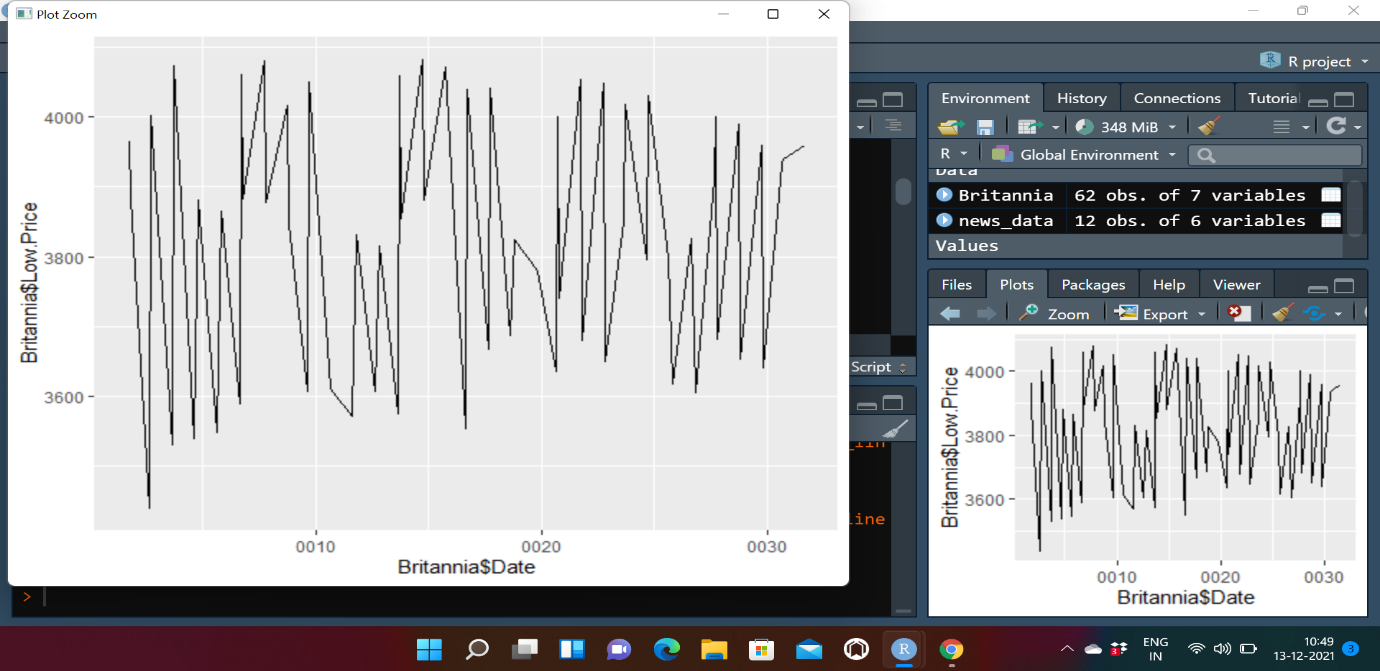
ggplot(data = Britannia,aes(y=Britannia$Open.Price,x=Britannia$Date)) + geom\_line()

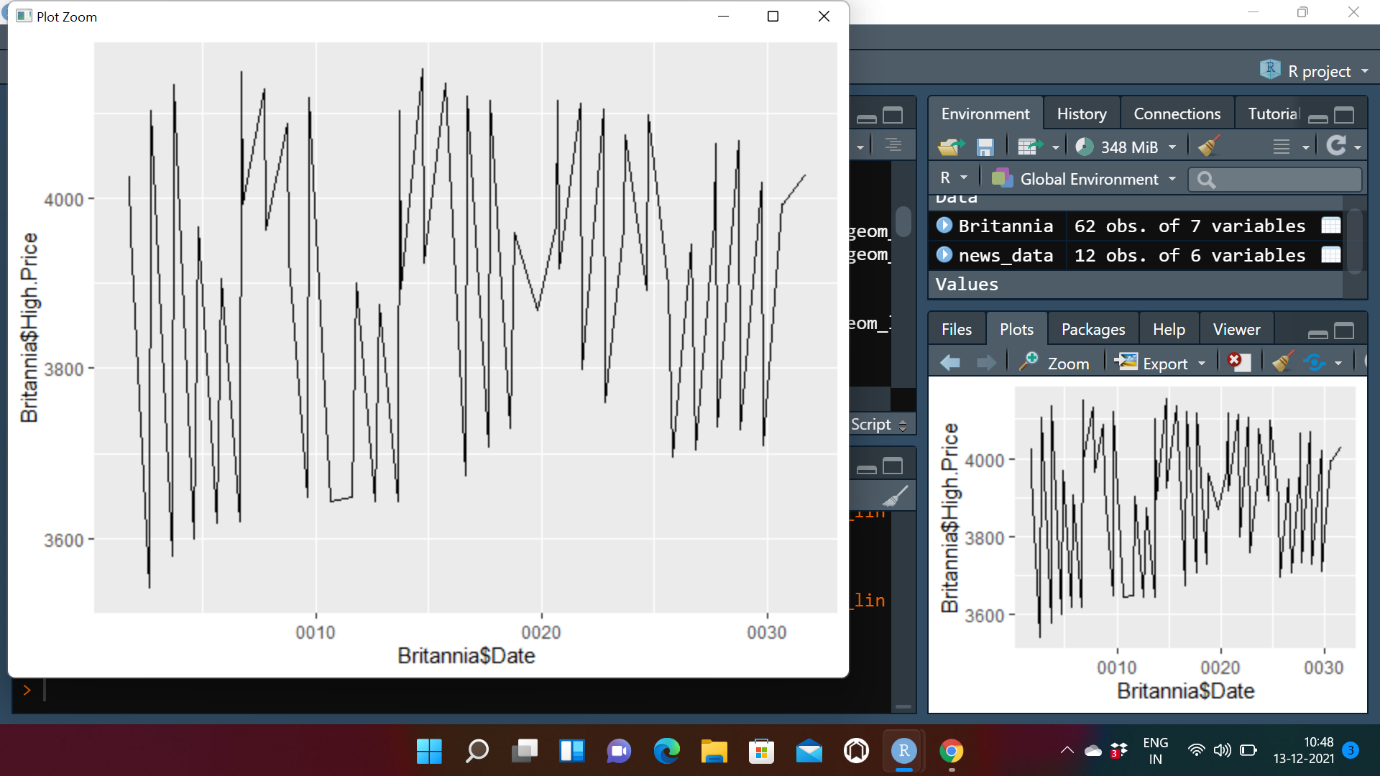
ggplot(data = Britannia,aes(y=Britannia$High.Price,x=Britannia$Date)) + geom\_line()

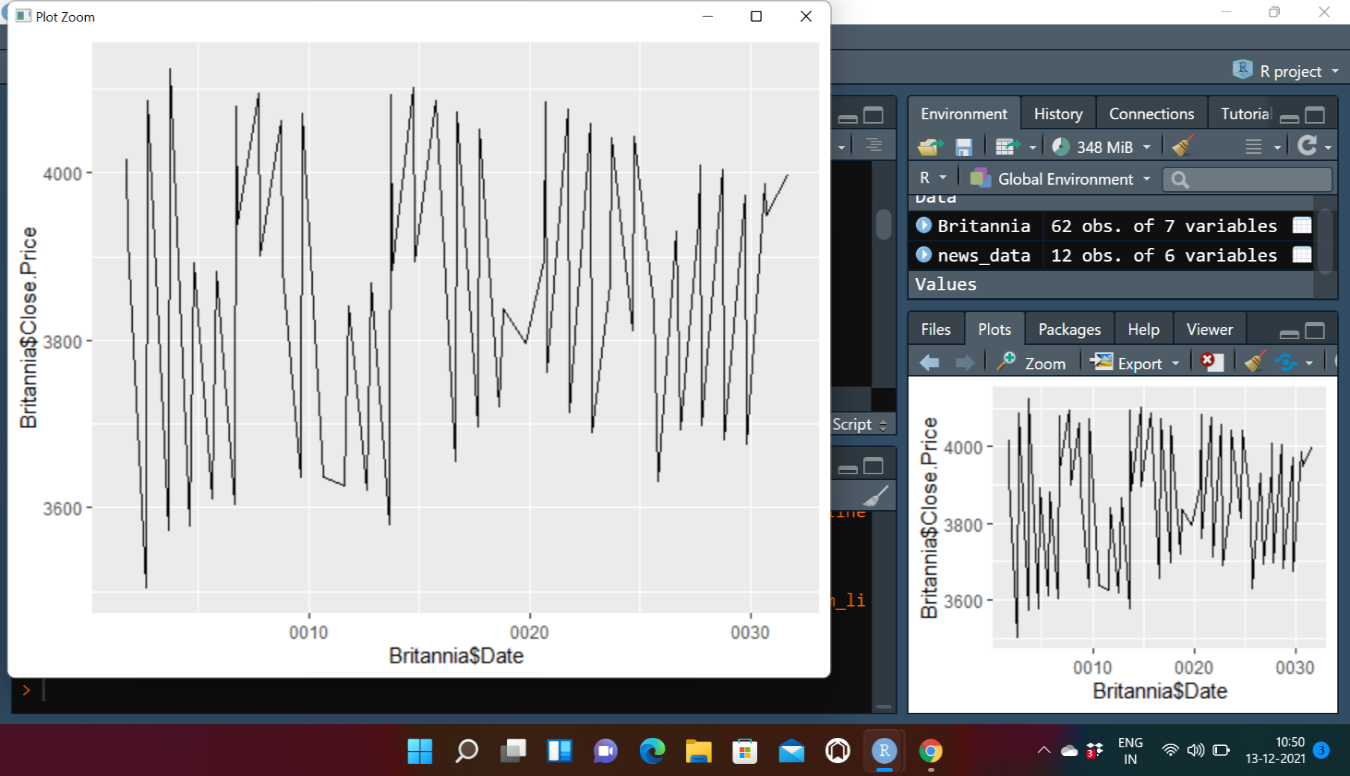
ggplot(data = Britannia,aes(y=Britannia$Low.Price,x=Britannia$Date)) + geom\_line()

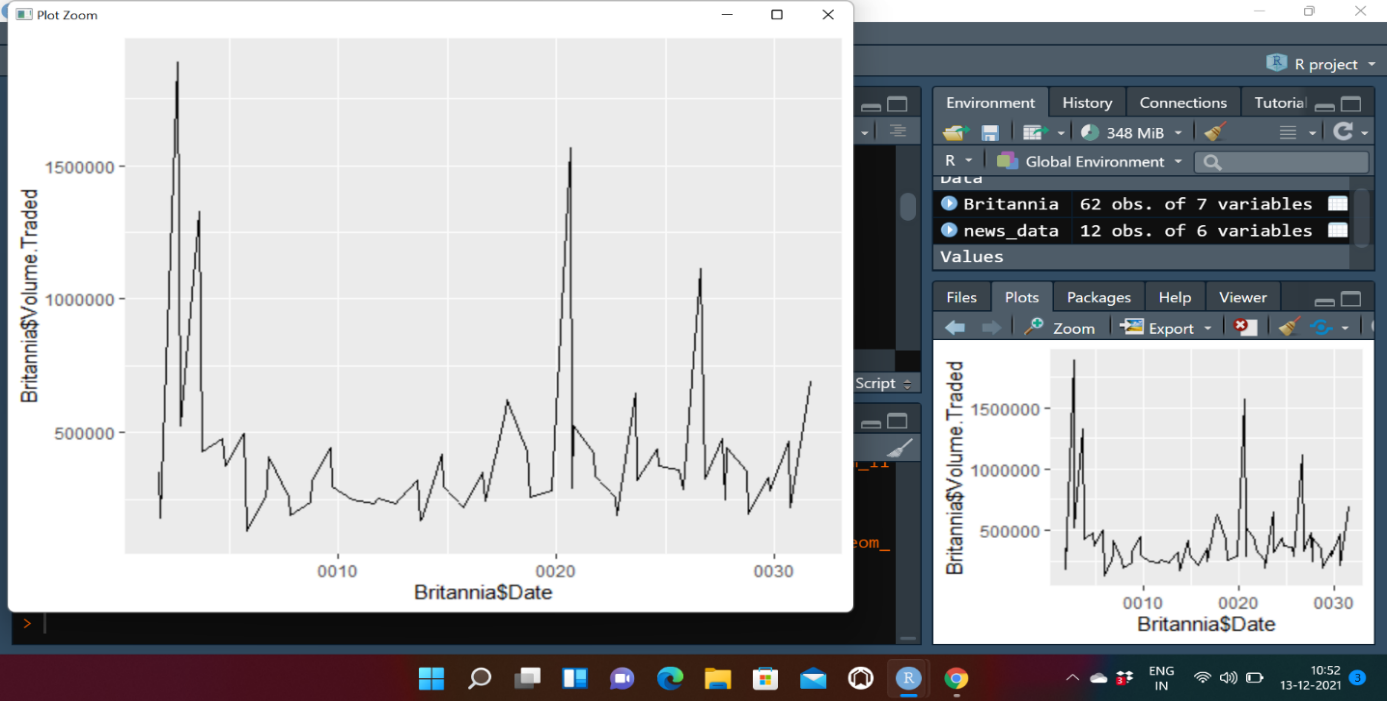
ggplot(data = Britannia,aes(y=Britannia$Close.Price,x=Britannia$Date)) + geom\_line()

ggplot(data = Britannia,aes(y=Britannia$Volume.Traded,x=Britannia$Date)) + geom\_line()









#9) Correlation Between Date and other Variables

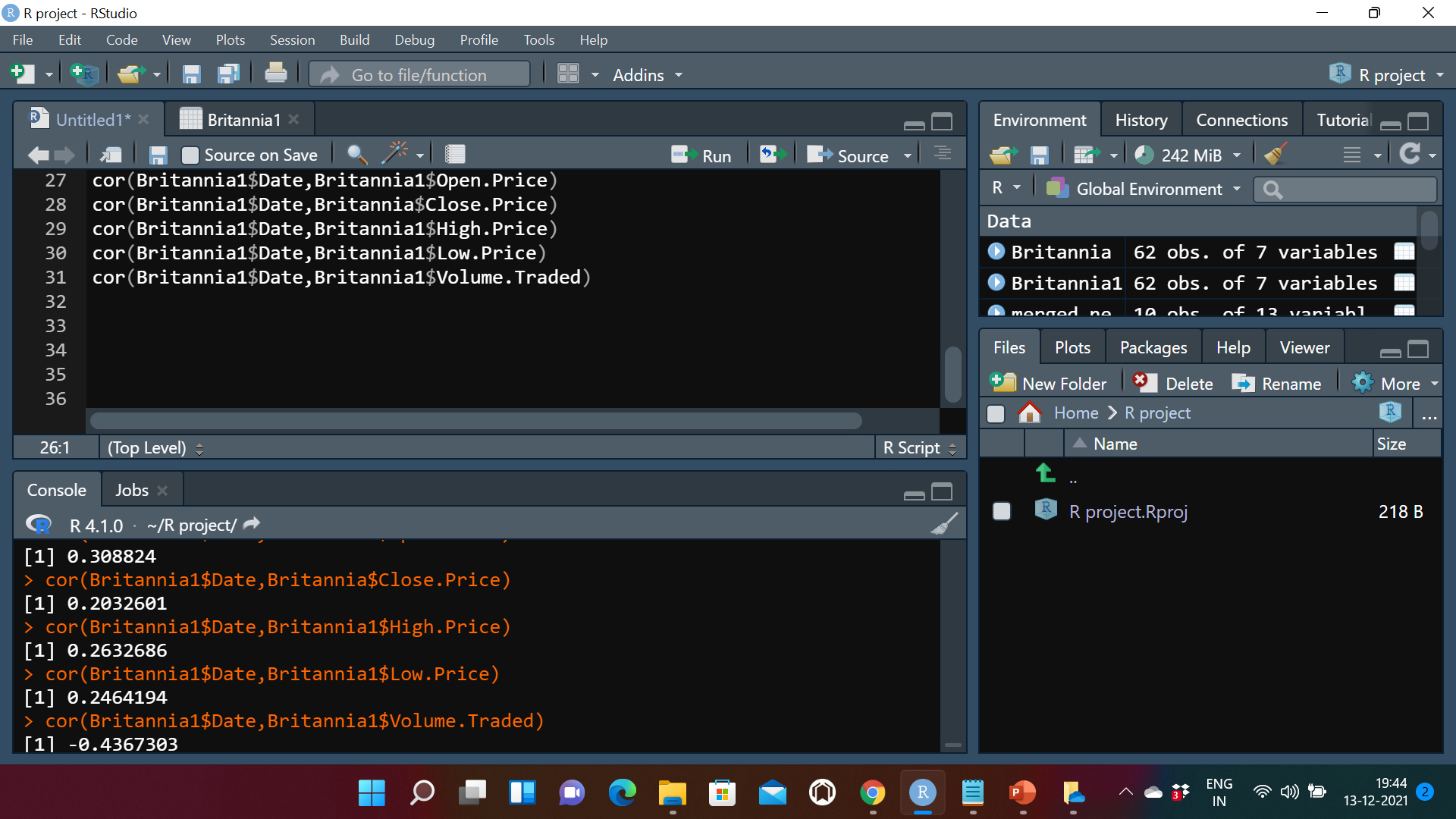
cor(Britannia1$Date,Britannia1$Open.Price)

cor(Britannia1$Date,Britannia$Close.Price)

cor(Britannia1$Date,Britannia1$High.Price)

cor(Britannia1$Date,Britannia1$Low.Price)

cor(Britannia1$Date,Britannia1$Volume.Traded)



Interpretation :

Correlation value is maximum with Open Price , hence it is evident that Open prices will vary maximum with the date, and Date and Volume traded have the weakest linear relationship.

#(9-10) Correlation 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)

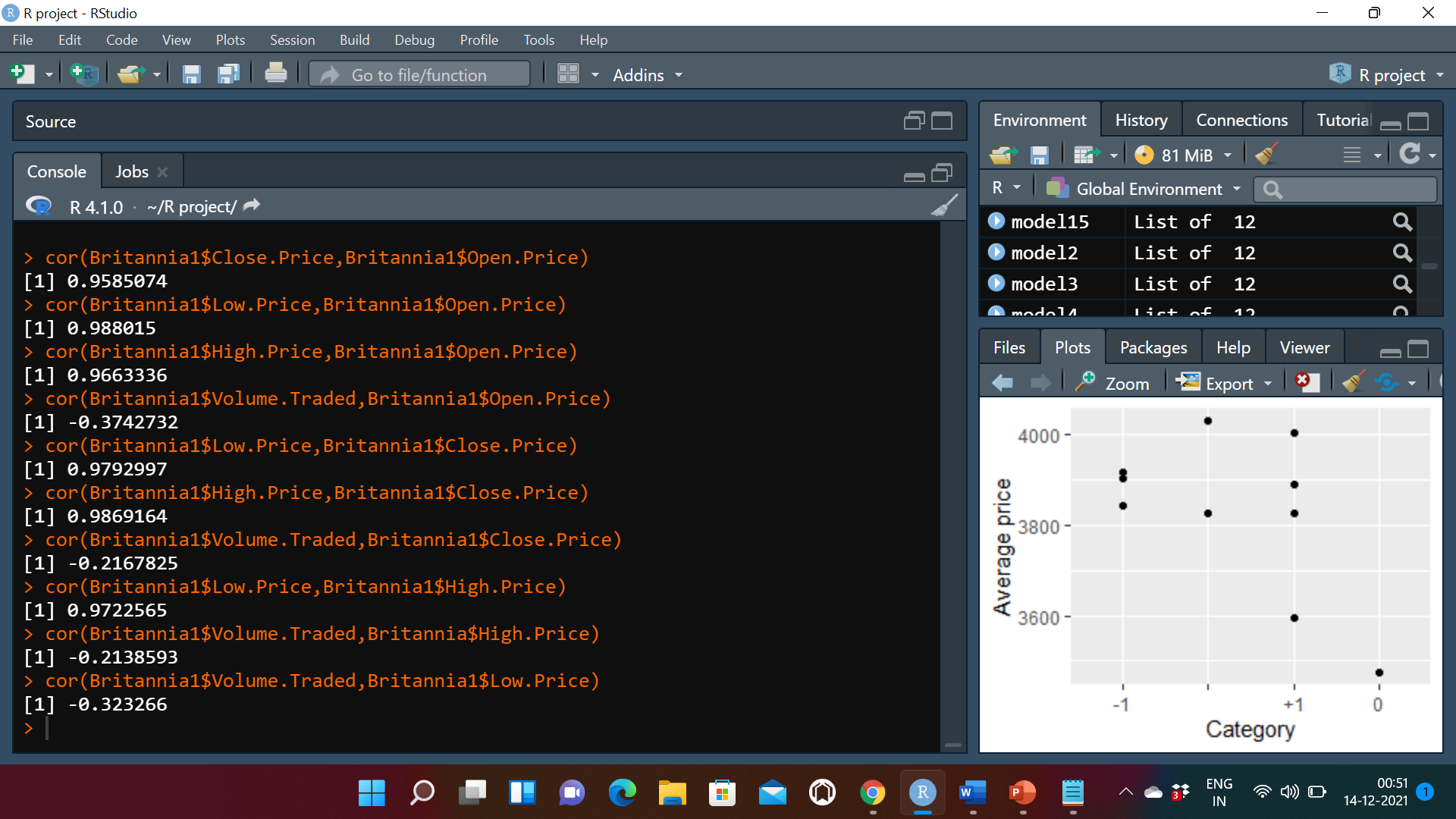
cor(Britannia1$High.Price,Britannia1$Close.Price)

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)



Interpretation:

From the above Values we can interpret that highest price and the close price have the strongest linear relationship, it means that when the share price was highest it was complemented by a higher highest price for that particular date

Whereas

Volume traded and low price have a weak linear relationship, according to this the volume traded was not much influenced by the low price of the share.

#Correlation Test Between Date and other 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)

#Correlation Test With different Variables

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)

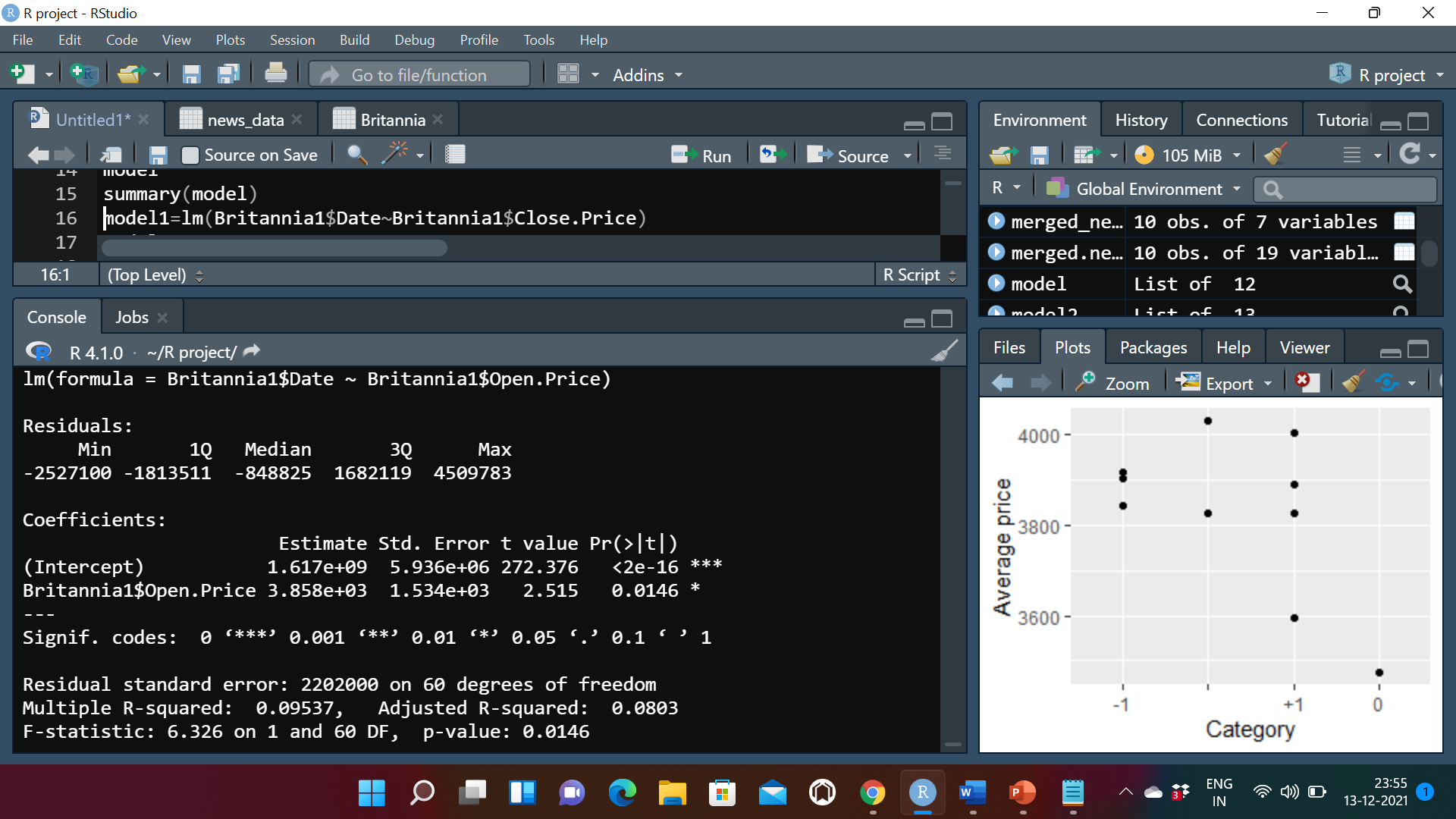
cor.test(Britannia1$Volume.Traded,Britannia1$Low.Price)

Regression for different variables in dataframe:

model=lm(Britannia1$Date~Britannia1$Open.Price)

model

summary(model)



model1=lm(Britannia1$Date~Britannia1$Close.Price)

model1

summary(model1)

R-squared: 0.04131

model2=lm(Britannia1$Date~Britannia1$Low.Price)

model2

summary(model2)

R-squared: 0.06072

model3=lm(Britannia1$Date~Britannia1$High.Price)

model3

summary(model3)

R-squared: 0.06931

model5=lm(Britannia1$Date~Britannia1$Volume.Traded)

model5

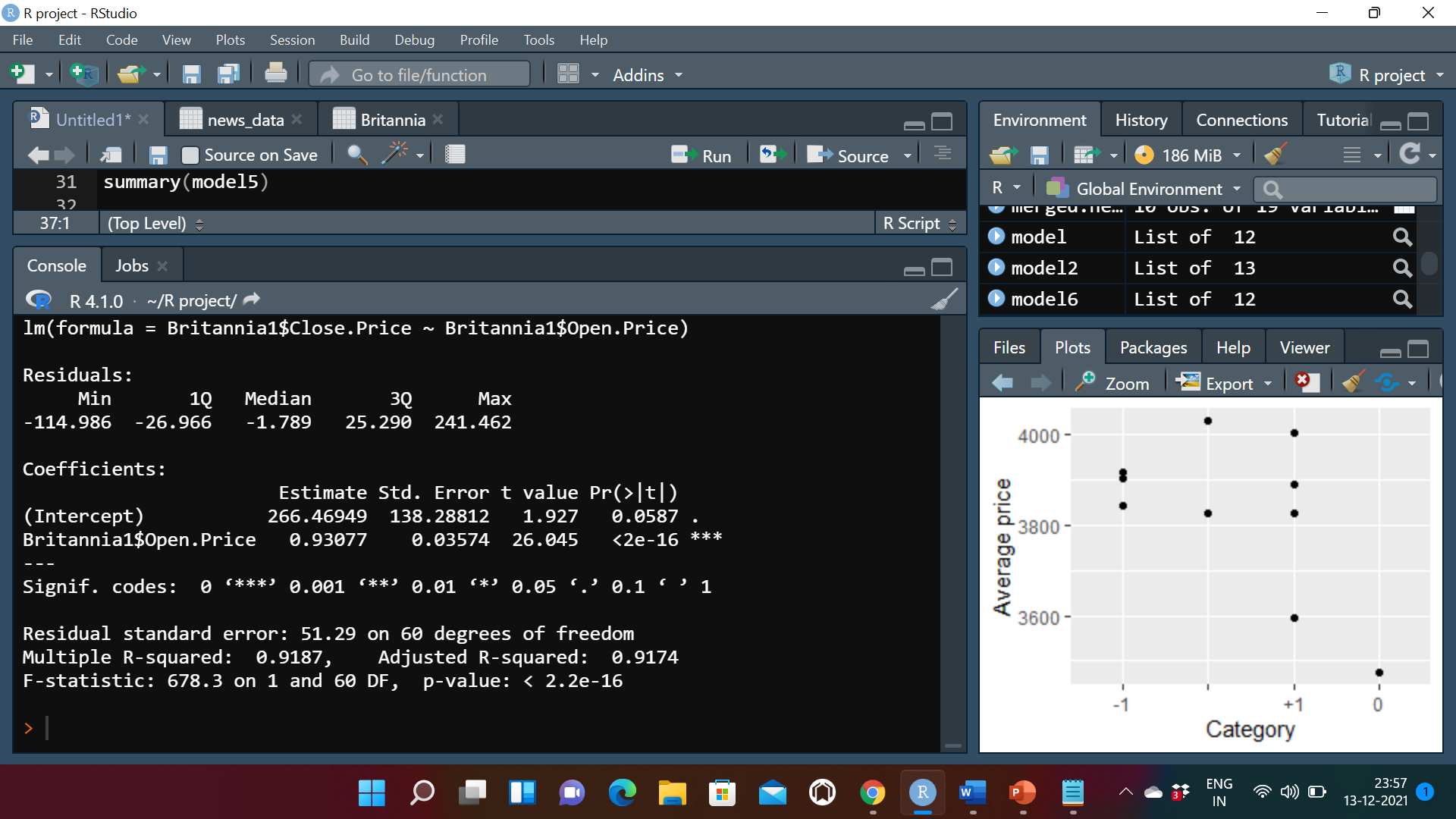
summary(model5)

R-squared: 0.1907

model6=lm(Britannia1$Close.Price~Britannia1$Open.Price)

model6

summary(model6)



model7=lm(Britannia1$Low.Price~Britannia1$Open.Price)

model7

summary(model7)

R-squared: 0.9762

model8=lm(Britannia1$High.Price~Britannia1$Open.Price)

model8

summary(model8)

R-squared: 0.9338

model9=lm(Britannia1$Volume.Traded~Britannia1$Open.Price)

model9

summary(model9)

R-squared: 0.1401

model10=lm(Britannia1$Low.Price~Britannia1$Close.Price)

model10

summary(model10)

R-squared: 0.959

model11=lm(Britannia1$High.Price~Britannia1$Close.Price)

model11

summary(model11)

R-squared: 0.974

model12=lm(Britannia1$Volume.Traded~Britannia1$Close.Price)

model12

summary(model12)

R-squared: 0.04699

model13=lm(Britannia1$Low.Price~Britannia1$High.Price)

model13

summary(model13)

R-squared: 0.9453

model14=lm(Britannia1$Volume.Traded~Britannia1$High.Price)

model14

summary(model14)

R-squared: 0.04574

model15=lm(Britannia1$Volume.Traded~Britannia1$Low.Price)

model15

summary(model15)

R-squared: 0.1045

Interpretation:

Highest Impacting Variable

Since R squared value of low price and open Price is the highest i.e, 0.9762 so low Price impacts open Price the most.

Least Impacting variable

Since R squared value of Date and close price is the lowest i.e, 0.04131 so date impact close price the least.

#11) Web Scraping

install.packages("rvest")

install.packages("dplyr")

library(rvest)

library(dplyr)

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")

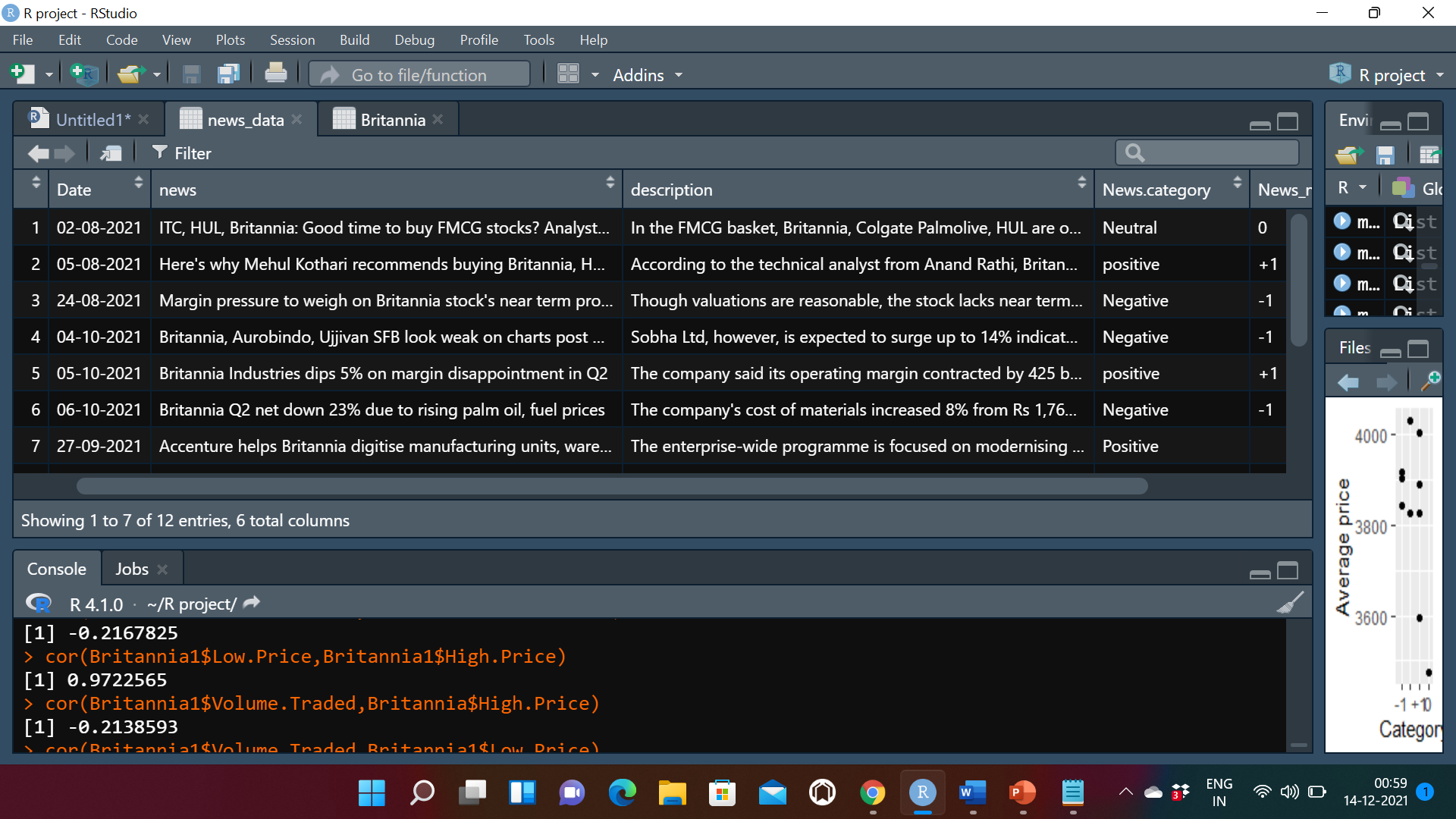
news\_data<-read.csv(file.choose())

View(news\_data)

names(news\_data)

#12) 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"," ")))



#13)

library(dplyr)

news\_data$Date<-as\_date(news\_data$Date)

merged\_news <- inner\_join(Britannia,news\_data,by.x=Date,by.y=Date)

merged\_news<-subset(merged\_news,select=- c(High.Price,Low.Price,Volume.Traded,news,X,News.category))

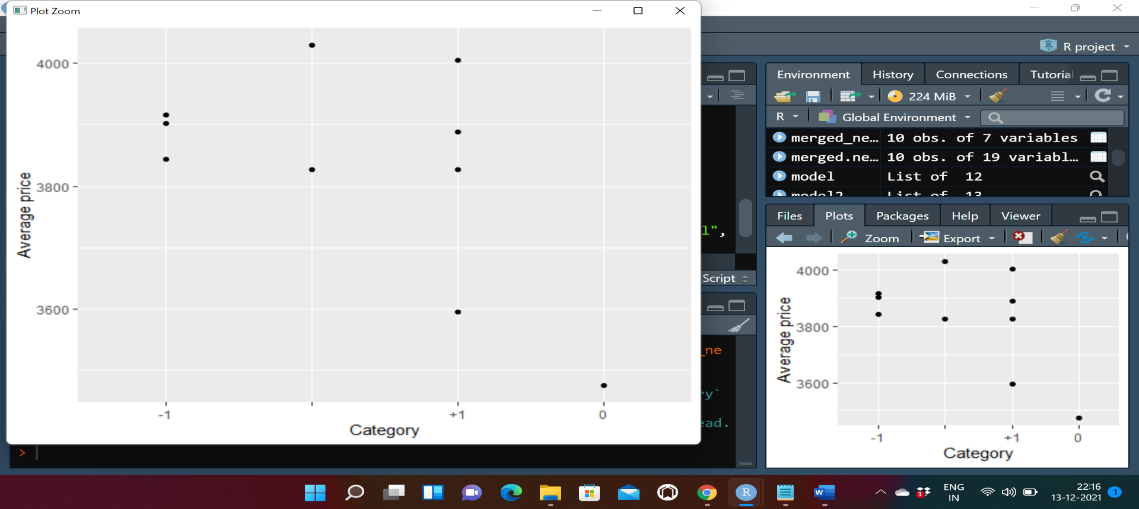
merged\_news$Average.Price<- ((merged\_news$Open.Price+merged\_news$Close.Price)/2)

model2=lm(merged\_news$News\_new.category~merged\_news$Average.Price)

summary(model2)

#14) Scatter Plot

ggplot(data = merged\_news,aes(y=merged\_news$Average.Price,x=merged\_news$News\_new.category)) + geom\_point() + labs(y="Average price",x="Category")



Scatter Plot created for the Average price of share and News category on particular date.

Conclusion from Graph: when the news related to company was positive,The average share price on that particular date was on the higher side whereas the average share price drop significantly when the news related to company was negative or neutral.

#(1-5)

Load the data in RStudio using appropriate functions and tools.

library(dplyr)

library(ggplot2)

library(tidyverse)

library(lubridate)

install.packages("readxl")

library("readxl")

ITC2 <- readxl::read\_xlsx(file.choose())

View(ITC2)

colnames(ITC2)<-c("Date","Open Price","High Price","Low Price","Close Price","No. of Trades")

ITC2$Date<-as.Date(ITC2$Date, format = "%d%m%y")

str(ITC2)

View(ITC2)

ITC2$Date<- as.numeric(ITC2$Date)

#6). #summary of Columns of ITC

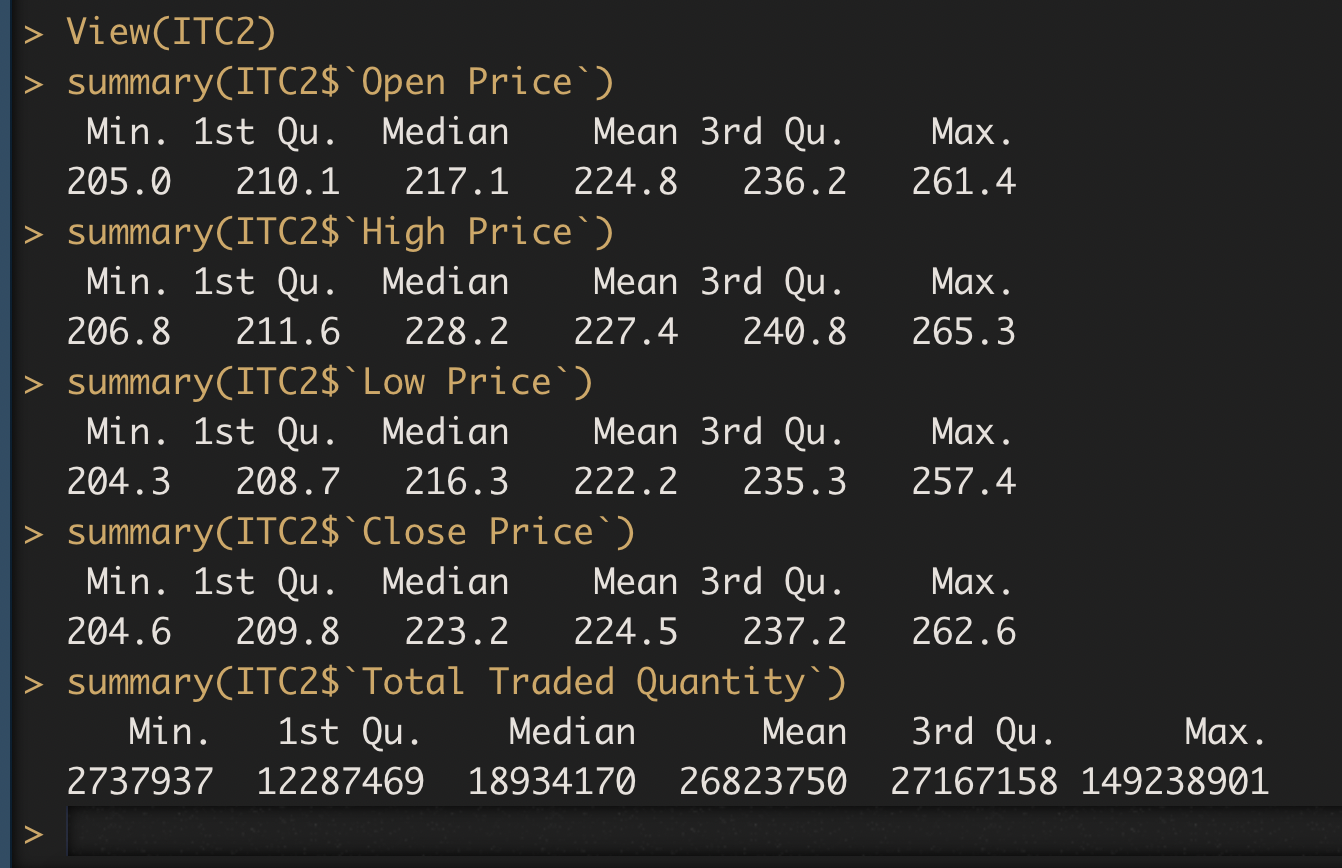
summary(ITC2$`Open Price`)

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**

#(7-8)

#linechart Between date and open price

library(ggplot2)

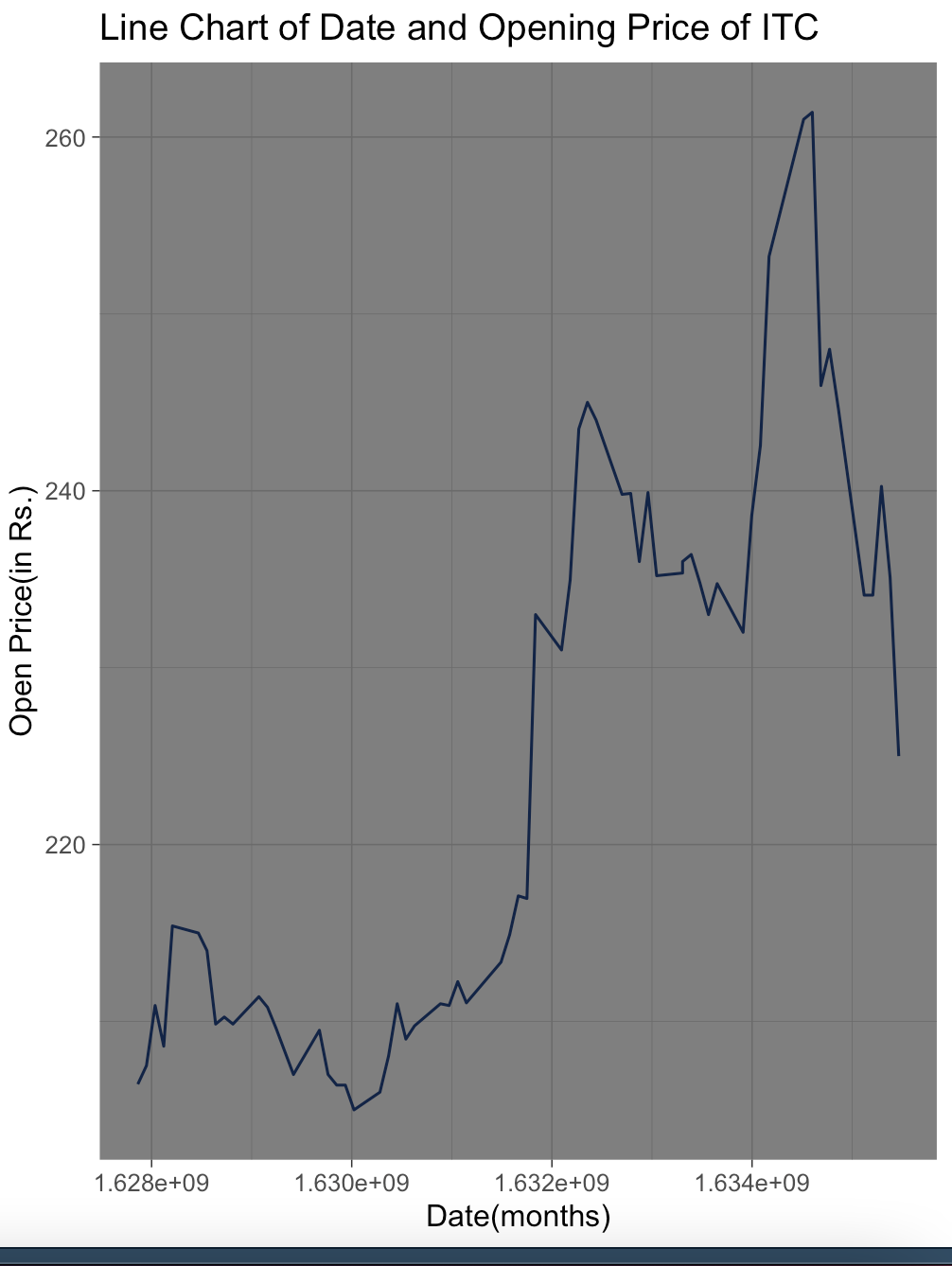
ggplot(ITC2) +

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()



#linechart Between date and closing price

library(ggplot2)

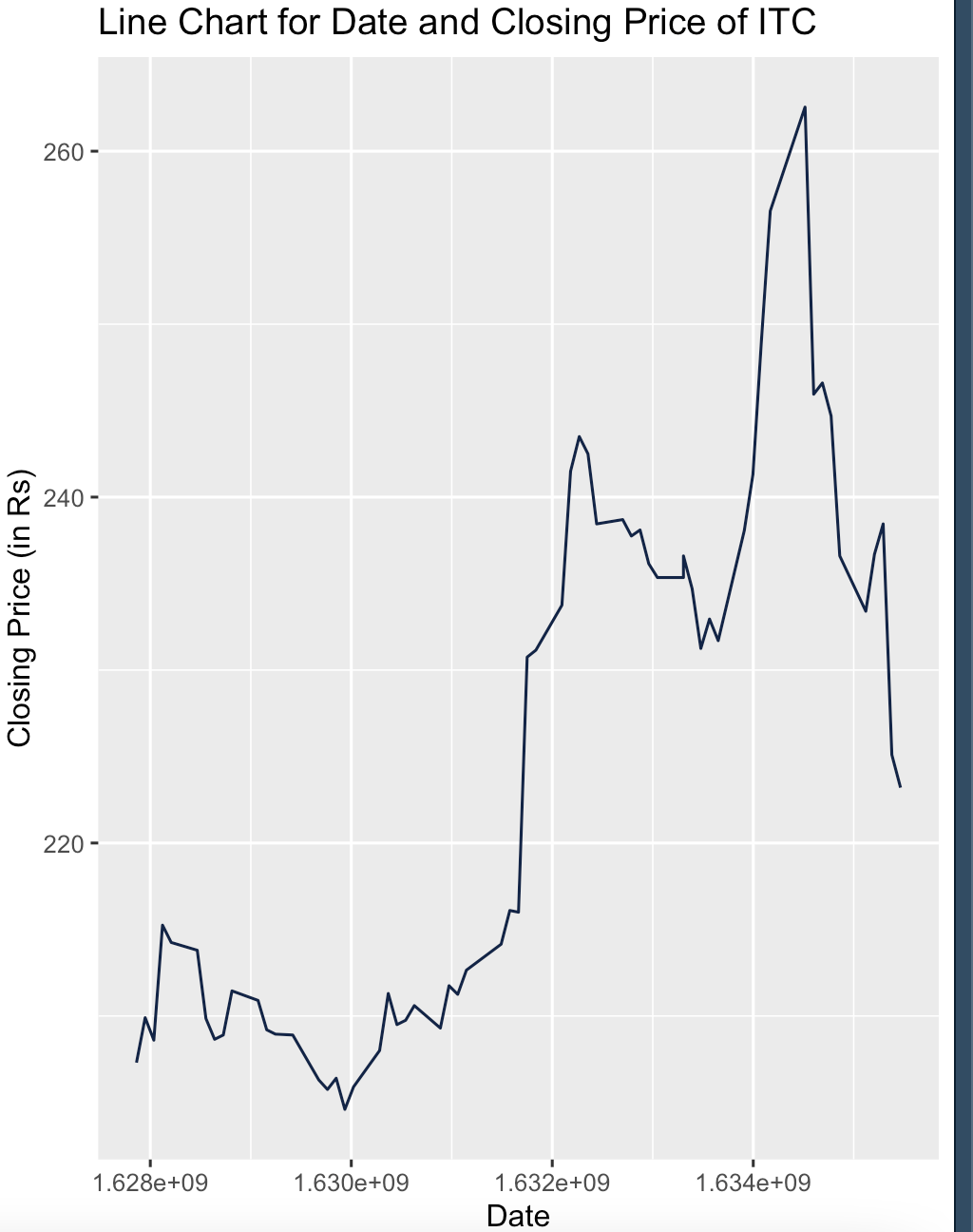
ggplot(ITC2) +

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)

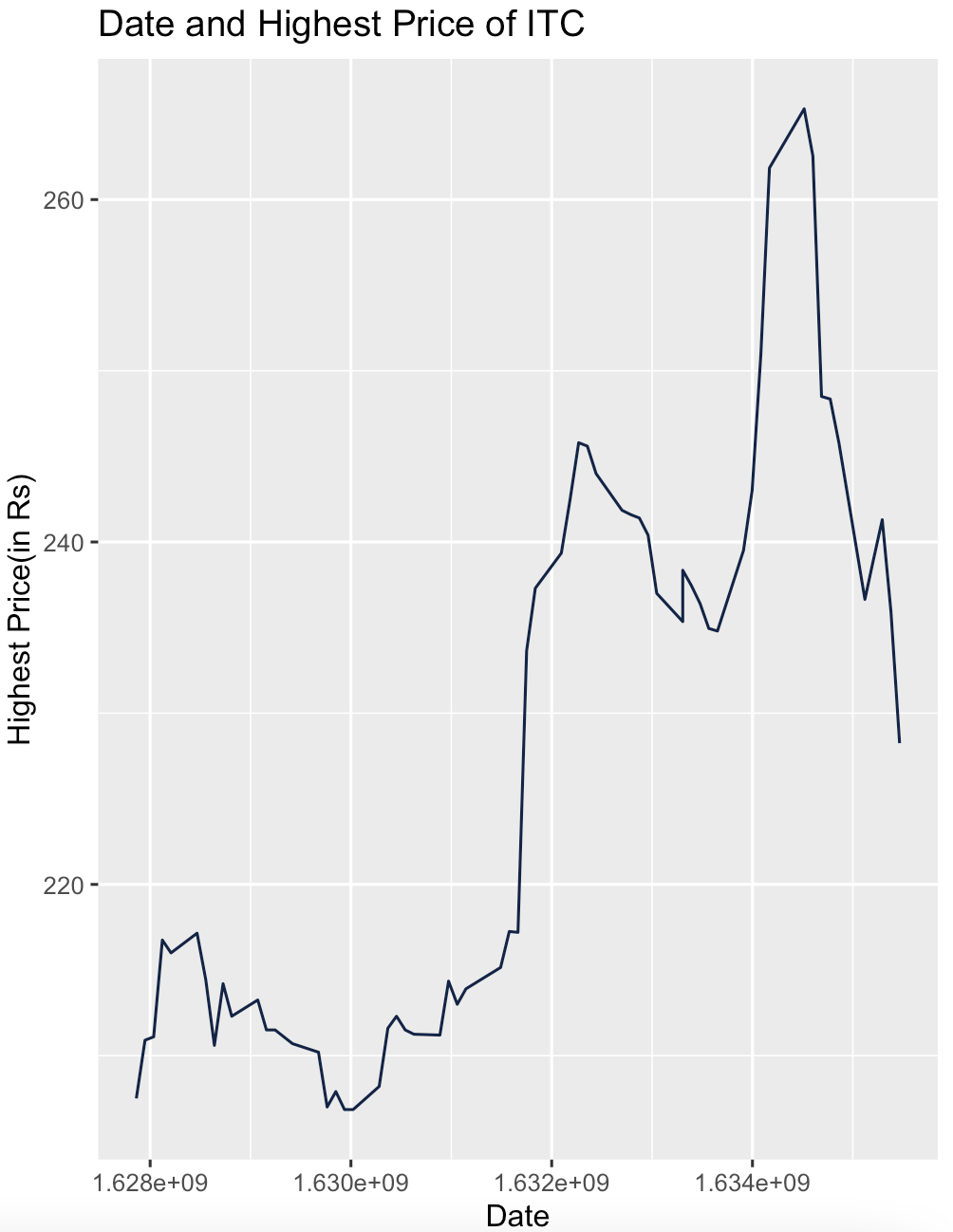
ggplot(ITC2) +

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()



#linechart Between Date and Lowest Price

library(ggplot2)

ggplot(ITC2) +

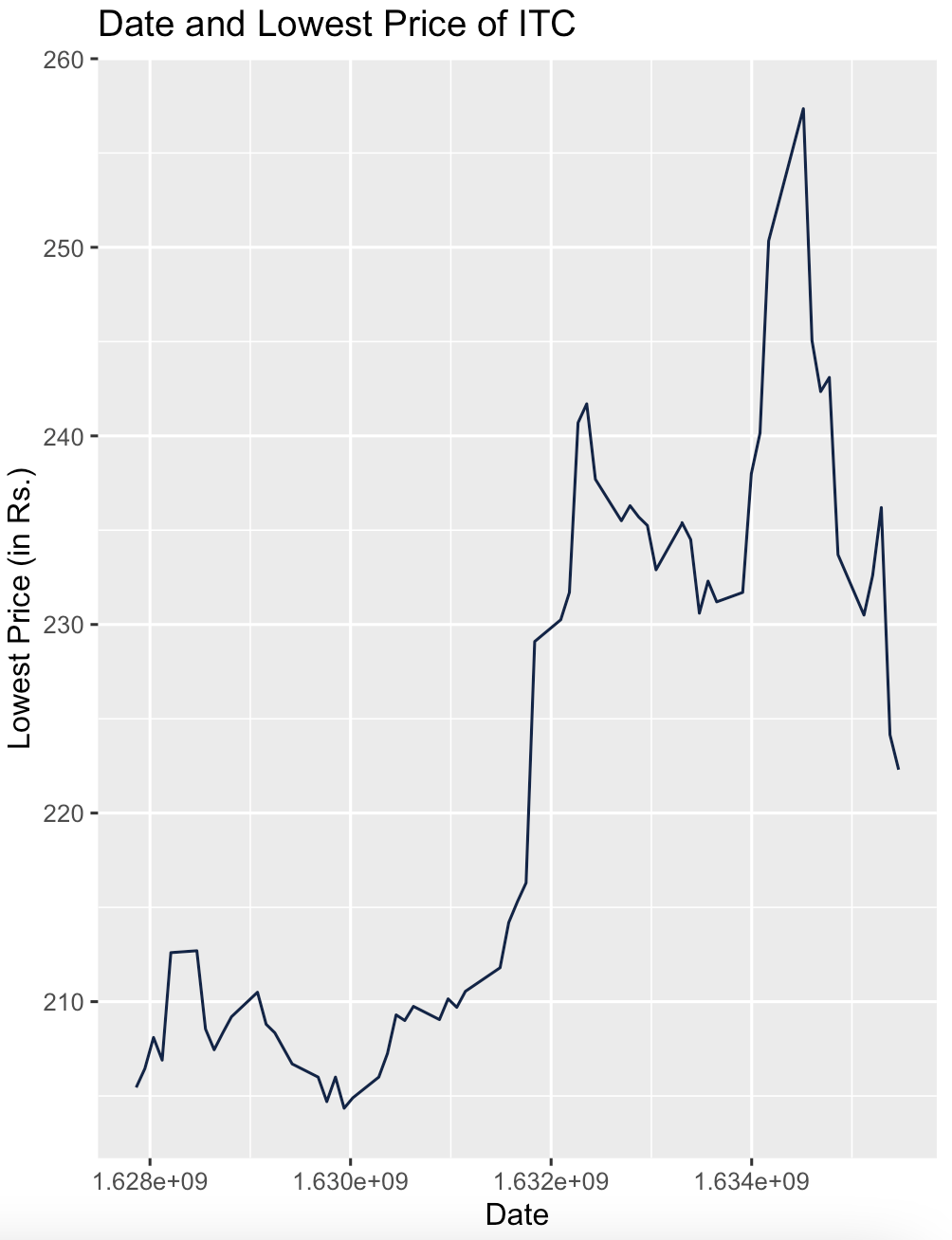
aes(x = Date, y = `Low Price`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(x = "Date",

y = "Lowest Price (in Rs.)", title = "Date and Lowest Price of ITC") +

theme\_gray()



#linechart Between Date and Volume Traded

ggplot(ITC2) +

aes(x = Date, y = `Total Traded Quantity`) +

geom\_line(size = 0.5, colour = "#112446") +

labs(

x = "Date",

y = "Total Traded Quantity",

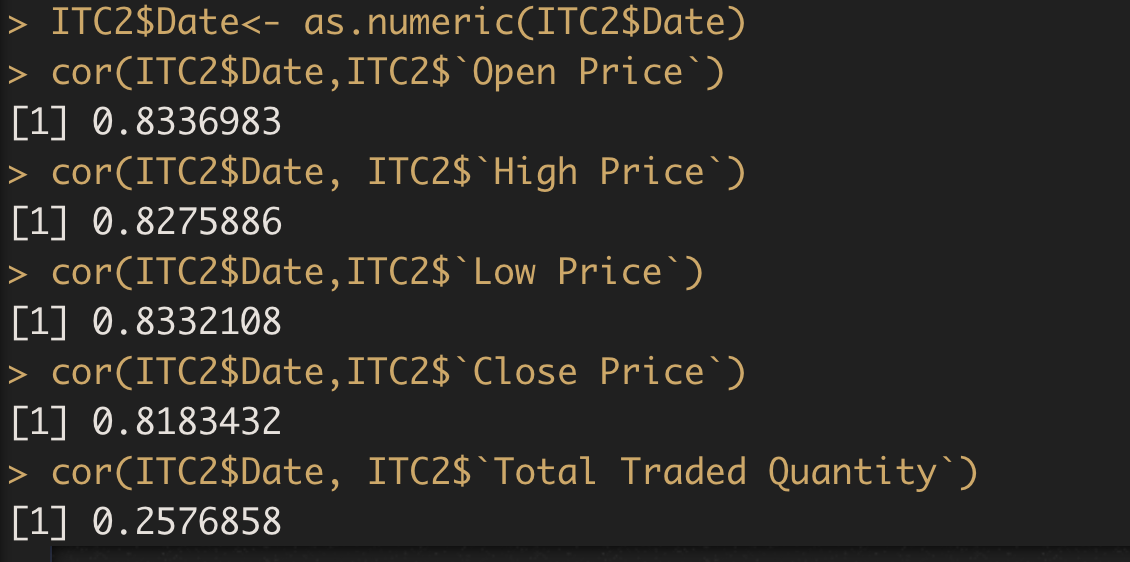
title = "Date and Volume Traded of ITC"

) +

theme\_gray()

#9

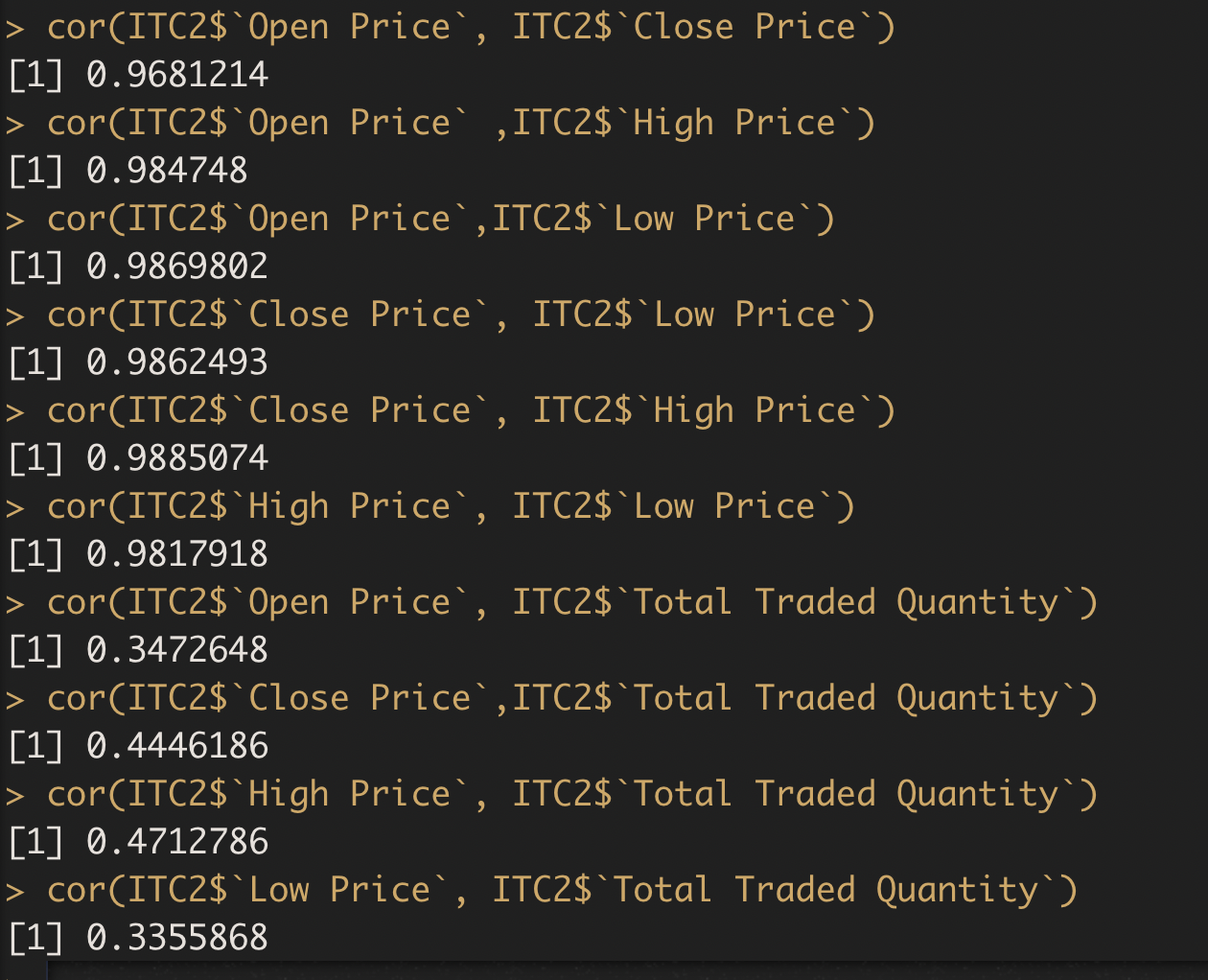
* cor(ITC2$Date,ITC2$`Open Price`)
* cor(ITC2$Date, ITC2$`High Price`)
* cor(ITC2$Date,ITC2$`Low Price`)
* cor(ITC2$Date,ITC2$`Close Price`)
* cor(ITC2$Date, ITC2$`Total Traded Quantity`)



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

#10

* 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`)



**From the above Values we can interpret that highest price and the close price have the strongest linear relationship, it means that when the share price was highest it was complemented by a higher highest price for that particular date**

**Whereas**

**Total Traded Quantity and low price have a weak linear relationship, according to this the volume traded was not much influenced by the low price of the share.**

#Correlation 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`)

#Correlation 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`)

Regression for different variables in dataframe:

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

model

summary(model)

R-squared: 0.9741

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

model1

summary(model1)

R-squared: 0.9697

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

model2

summary(model2)

R-squared: 0.9727

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

model3

summary(model3)

R-squared: 0.9697

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

model4

summary(model4)

R-squared: 0.2221

model5=lm(ITC2$`High Price`~ITC2$`Low Price`)

model5

summary(model5)

R-squared: 0.9639

model6=lm(ITC2$`Open Price`~ITC2$`Total Traded Quantity`)

model6

summary(model6)

R-squared: 0.1206

model7=lm(ITC2$`Close Price`~ITC2$`Total Traded Quantity`)

model7

summary(model7)

R-squared: 0.1977

model8=lm(ITC2$`Low Price`~ITC2$`Total Traded Quantity`)

model8

summary(model8)

R-squared: 0.1126

model9=lm(ITC2$Date~ITC2$`Open Price`)

model9

summary(model9)

R-squared: 0.6951

Interpretation:

Highest Impacting Variable

Since R squared value of low price and open Price is the highest i.e, 0.9741 so low Price impacts open Price the most.

Least Impacting variable

Since R squared value of Total traded quantity and low price is the lowest i.e, 0.1126 so Total traded quantity impact low price the least.

#11) Web Scraping

##scrap data of news

install.packages("rvest")

install.packages("dplyr")

library(rvest)

library(dplyr)

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)

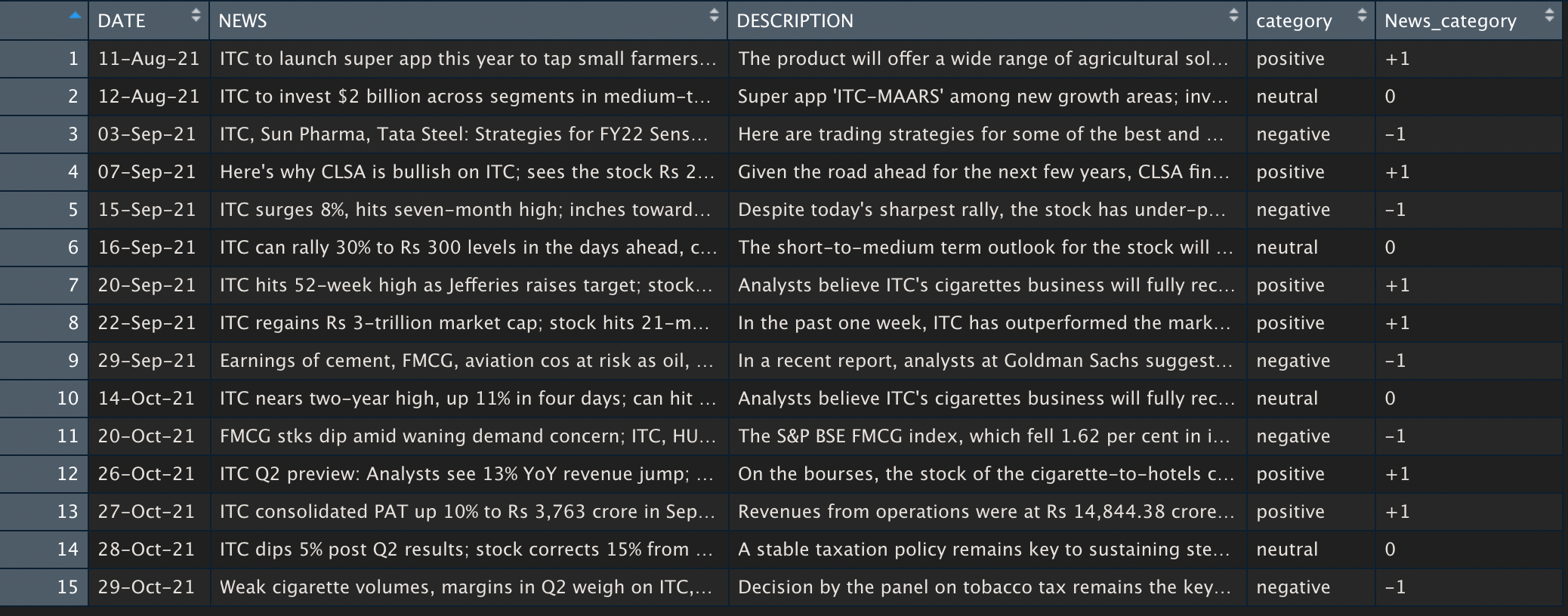
colnames(df)<-c("date","news","description")

write.csv(df,"ITCNewscsv1")

View(df)

#12)

* 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)



#13

* 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)

#14

ggplot(data = merged\_news,aes(y=merged\_news$AveragePrice,x=merged\_news$News\_category)) + geom\_point() + labs(y="Average price",x=" News\_category)) ")