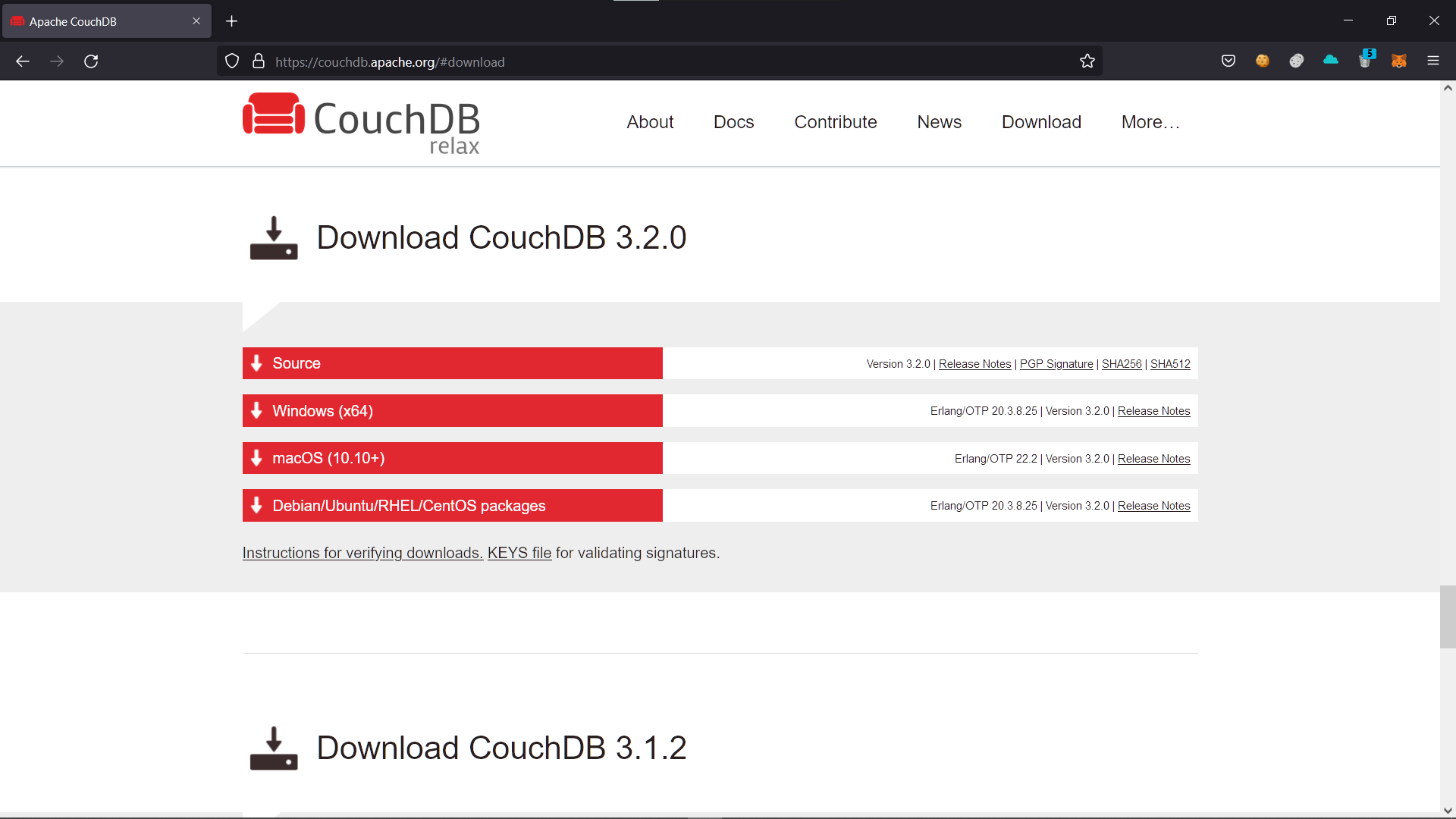
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Practical No | Topic | Date | Page No | Remark |
| 1 | Exp1: Data collection, Data curation and management for Unstructured data(NoSQL) with Couch DB. | 05/02/22 | 2-5 |  |
| 2 | Exp 2: Practical of Data collection, Data curation and management for Large-  scale Data system (such as MongoDB) | 29/01/22 | 6-10 |  |
| 3 | Exp:3 Dimension Reduction using Principal Component Analysis (PCA). | 23-11-2021 | 11-12 |  |
| 4 | Exp: 4. Practical of Clustering | 04-12-2021 | 13-15 |  |
| 5 | Exp: 5. Practical of Time-series forecasting | 11-12-2021 | 16-20 |  |
| 6 | Exp 6. Practical of Simple/Multiple Linear Regression | 15/01/22 | 21-24 |  |
| 7 | Exp 7. Practical of Logistics Regression | 18/12/21 | 25-27 |  |
| 8 | Exp: 8. Practical of Hypothesis testing | 30/11/2021 | 28 |  |
| 9 | Exp 9. Practical of Analysis of Variance | 22/02/22 | 29-30 |  |
| 10 | Exp 10. Practical of Decision Tree | 12/02/22 | 31-32 |  |

Data Science

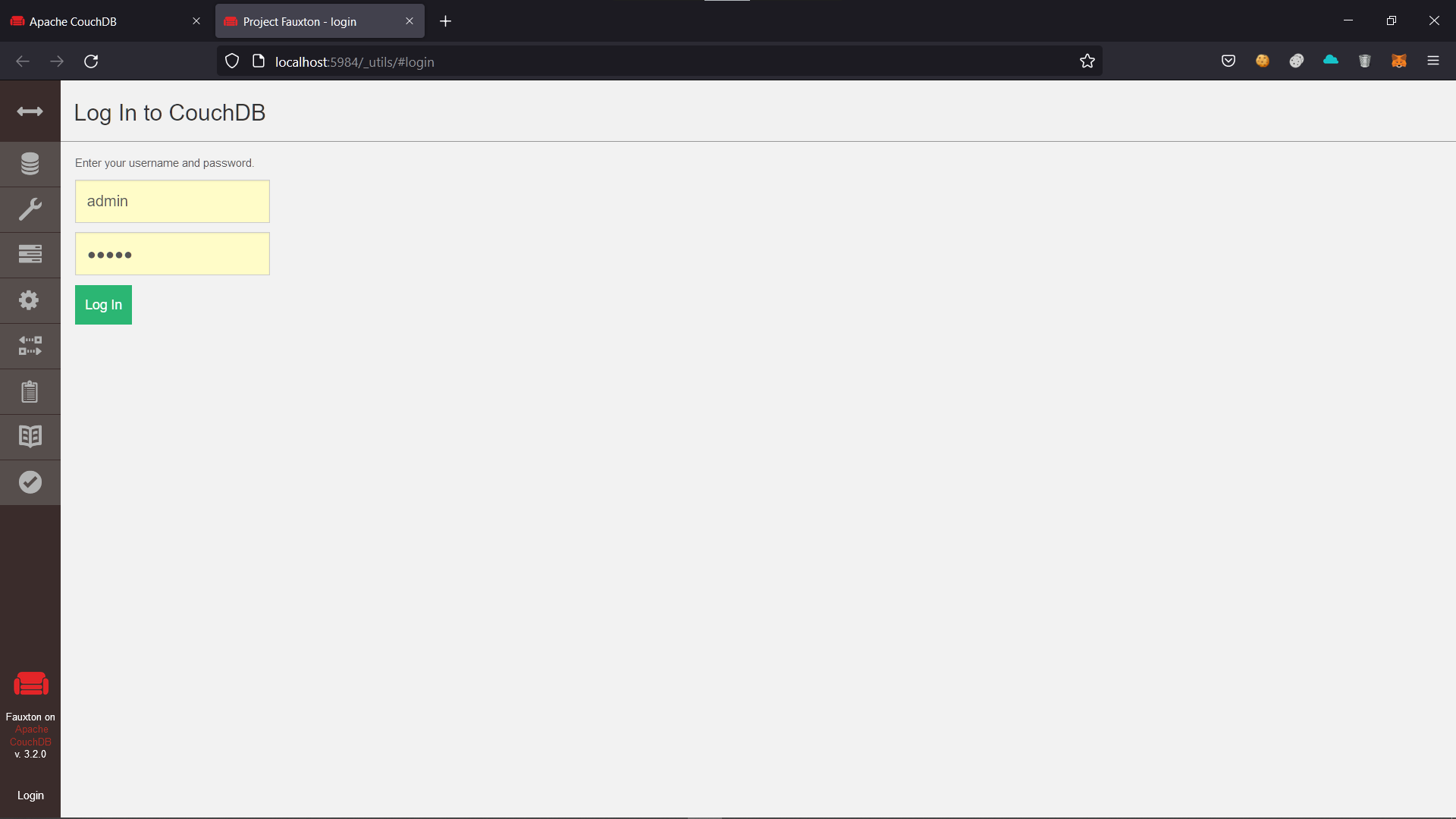
**PRACTICAL NO. 1**

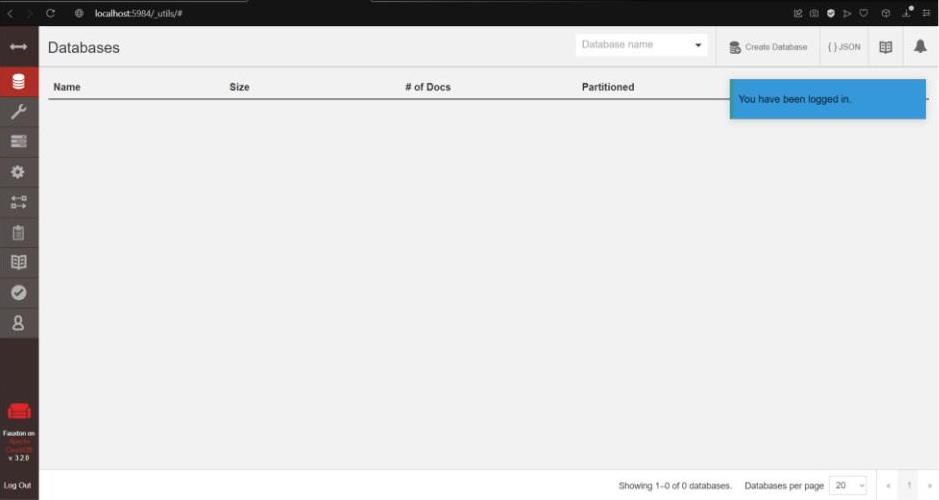
**Aim: Practical of Data collection, Data curation and management for Unstructured data (NoSQL) Download and install CouchDB**

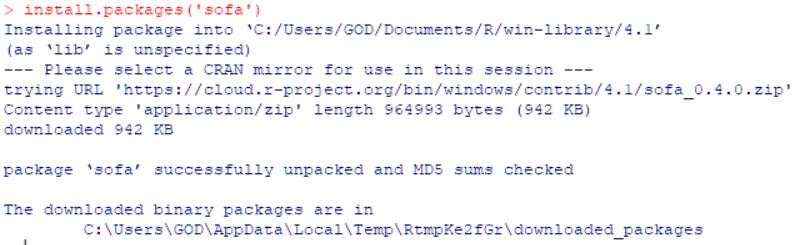
Set up and validate new login credientials



Now visit http://localhost:5984/\_utils and enter the credentials



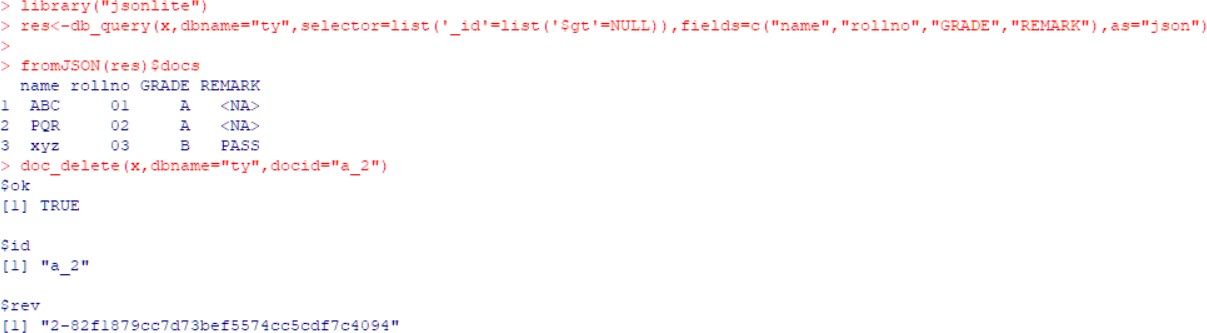


Now open r console Install Package

Connect to the database: Insert rows and commit changes:

Query Database



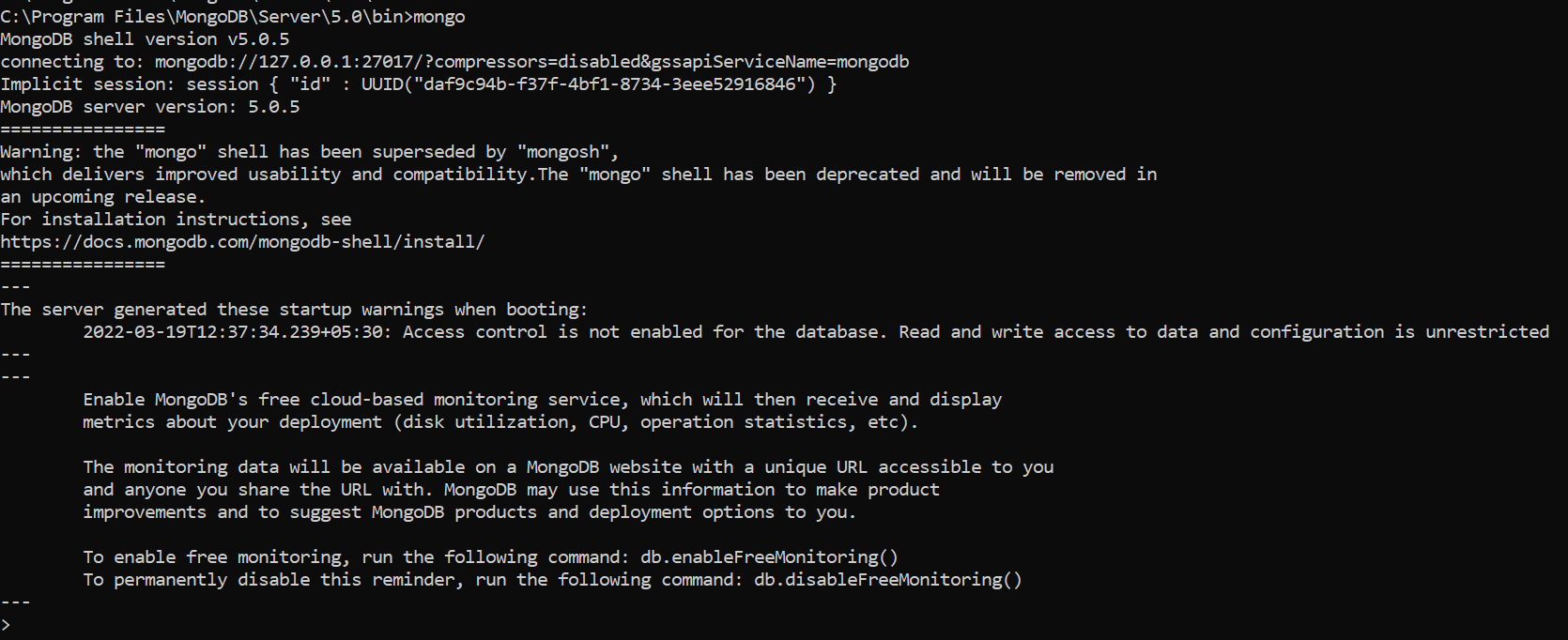
Delete Record:

Update a record:

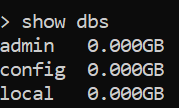


**PRACTICAL NO. 2**

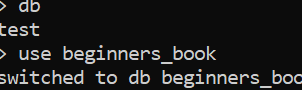
**Aim: Practical of Data collection, Data curation and management for Large-scale Data system (such as MongoDB)**



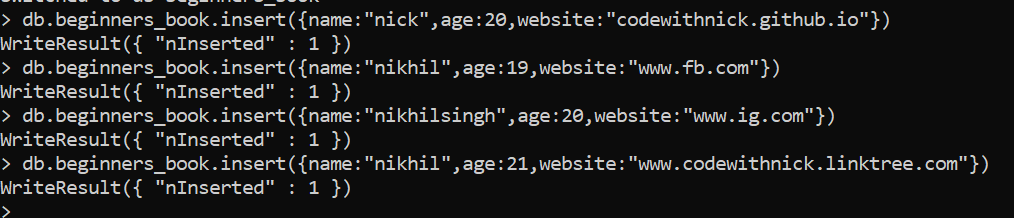
Show databases



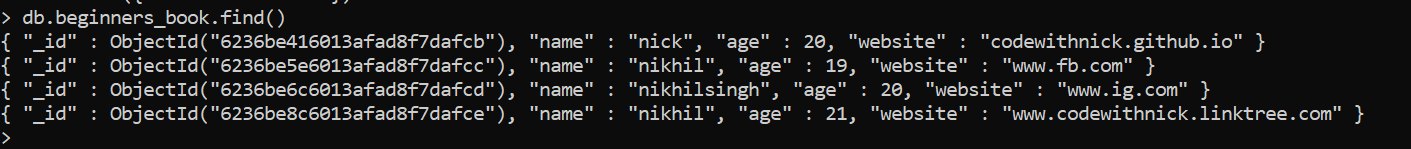
Switch to beginners\_book (collections)



Insert records



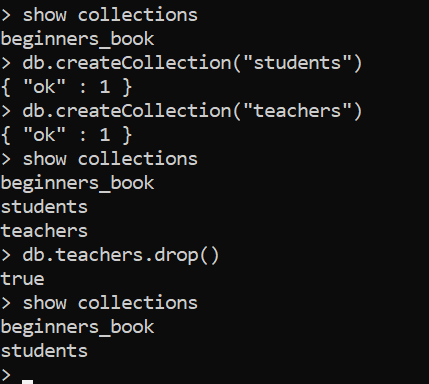
Look at records



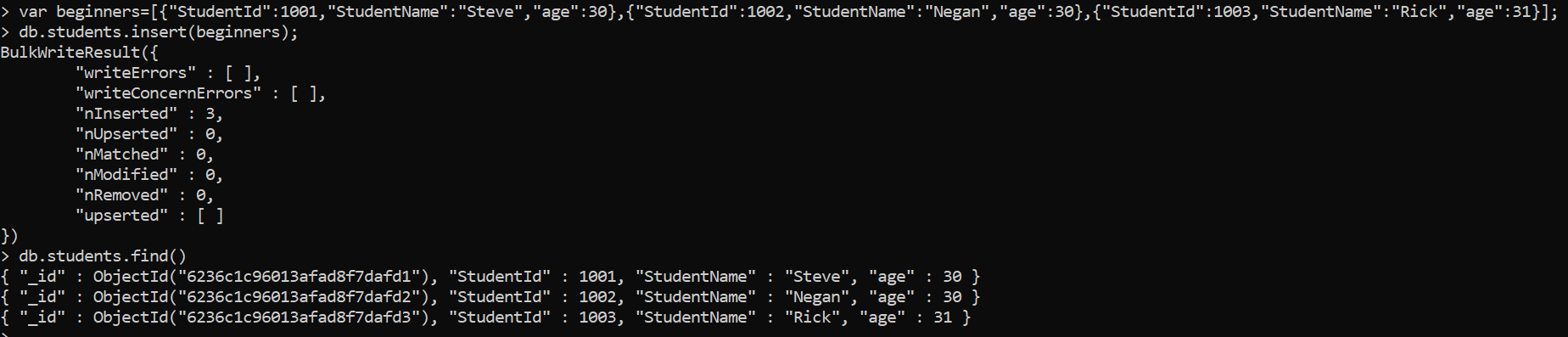
Insert dynamic records



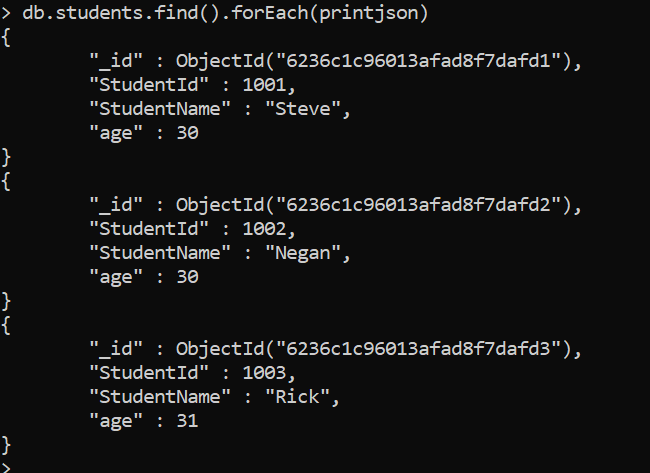
Create, view and drop collections



Bulk insert



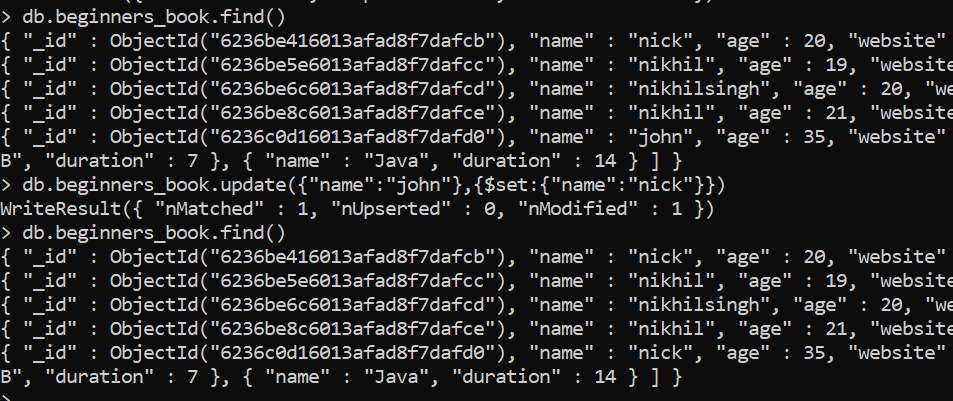
View each record

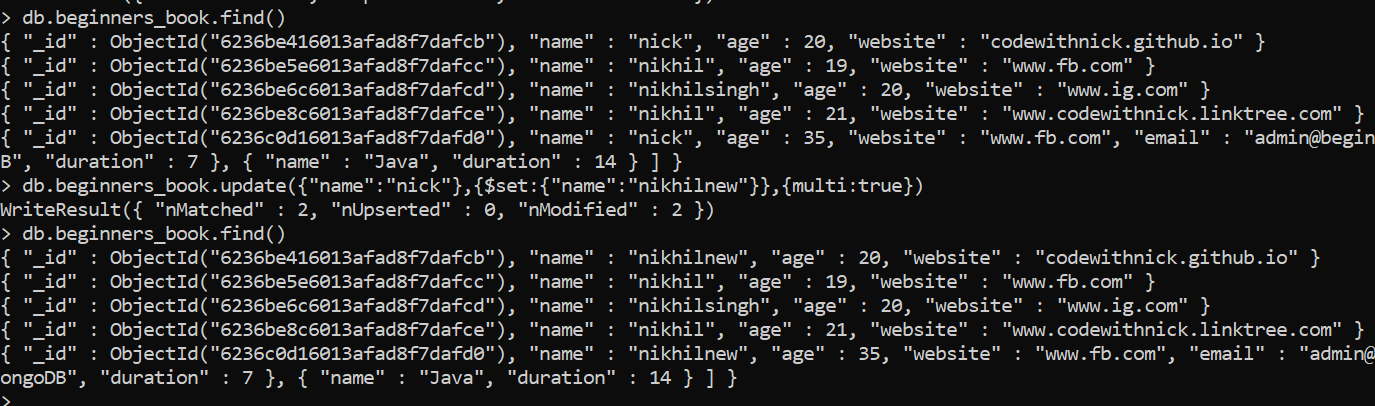


Search records according to attributes

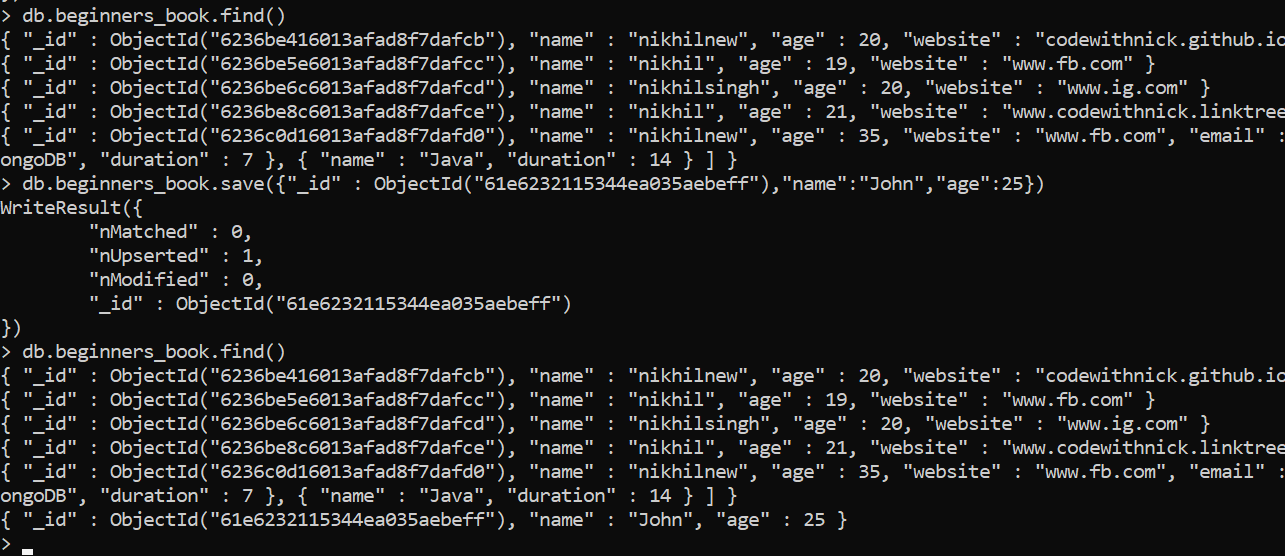


Update query

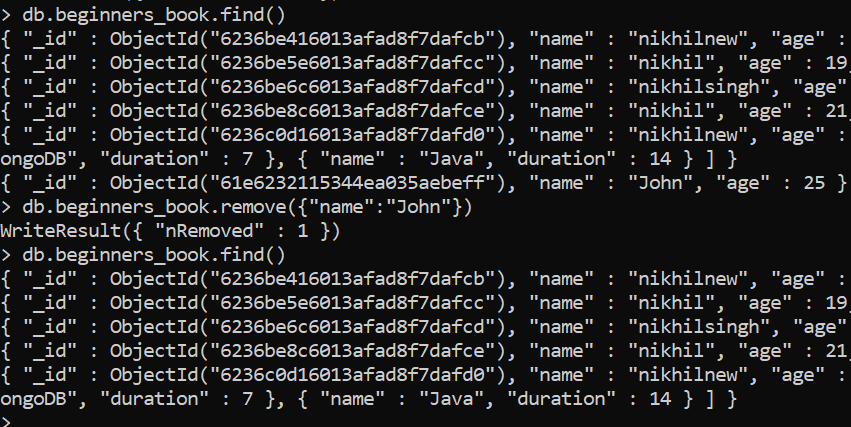




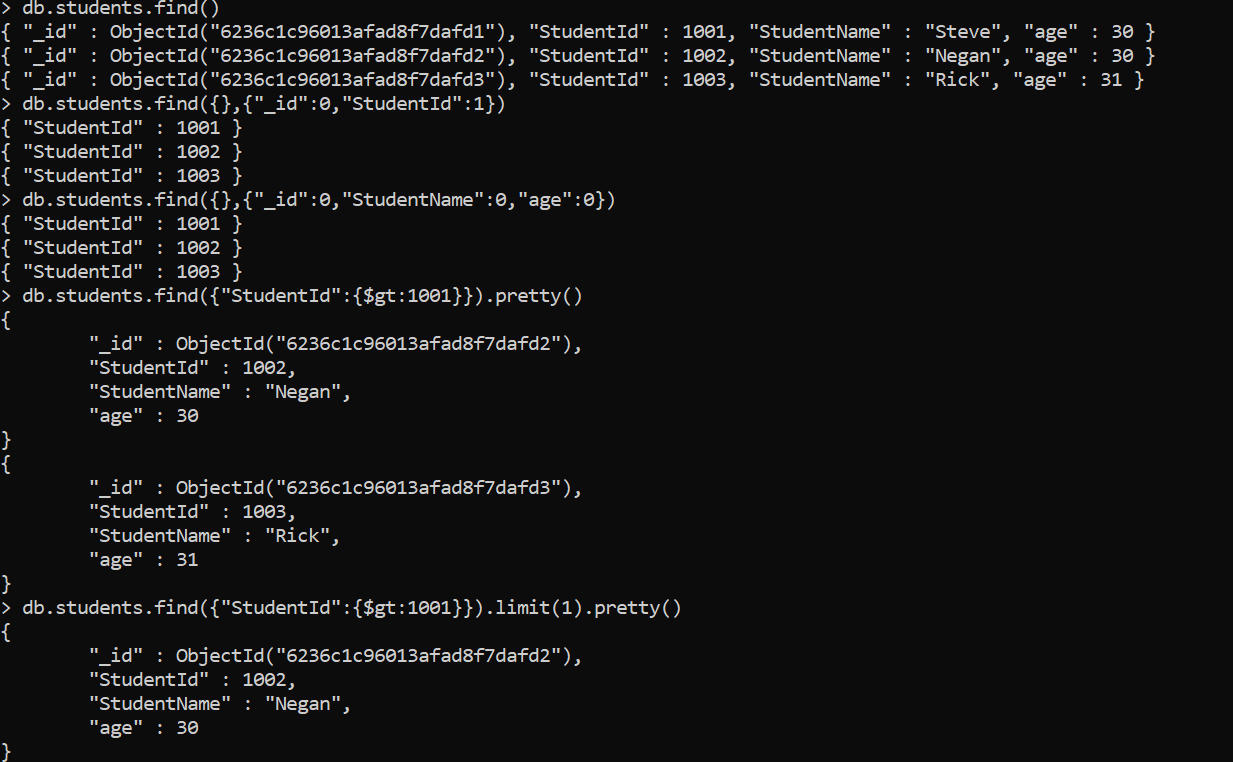
Save if not exists or update if exists

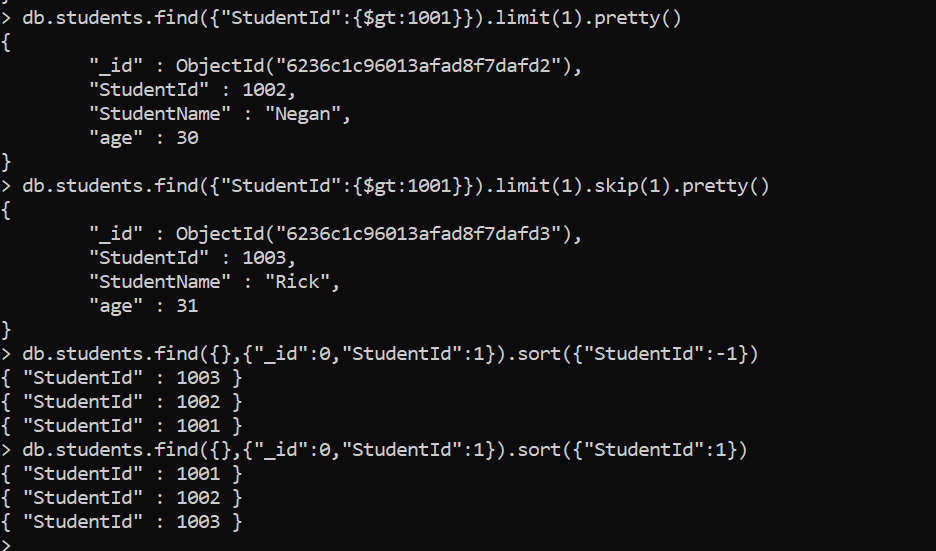


Remove



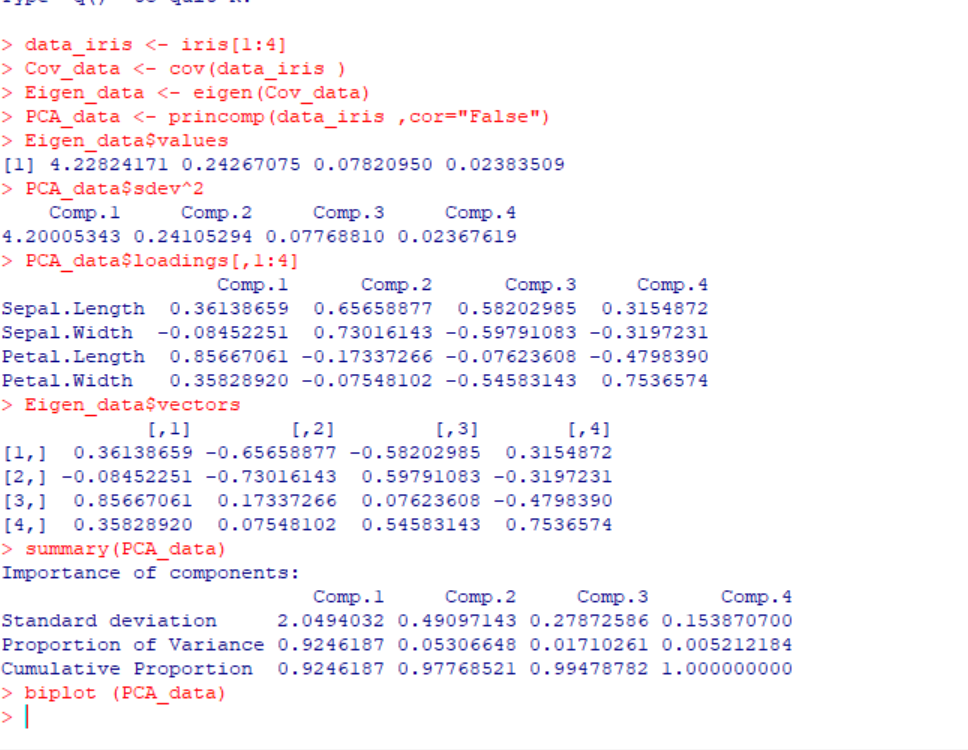
Finding using more attrs

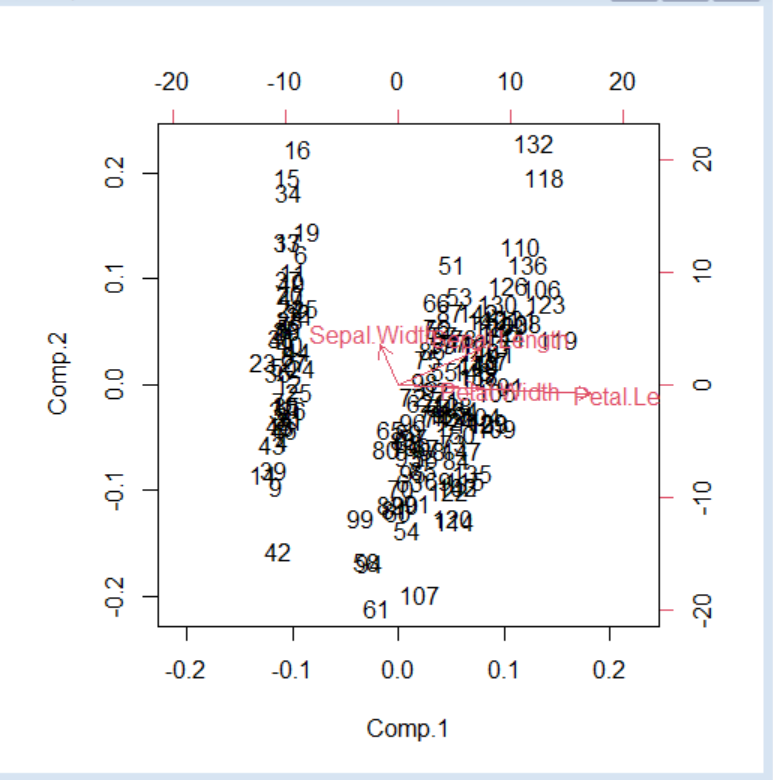


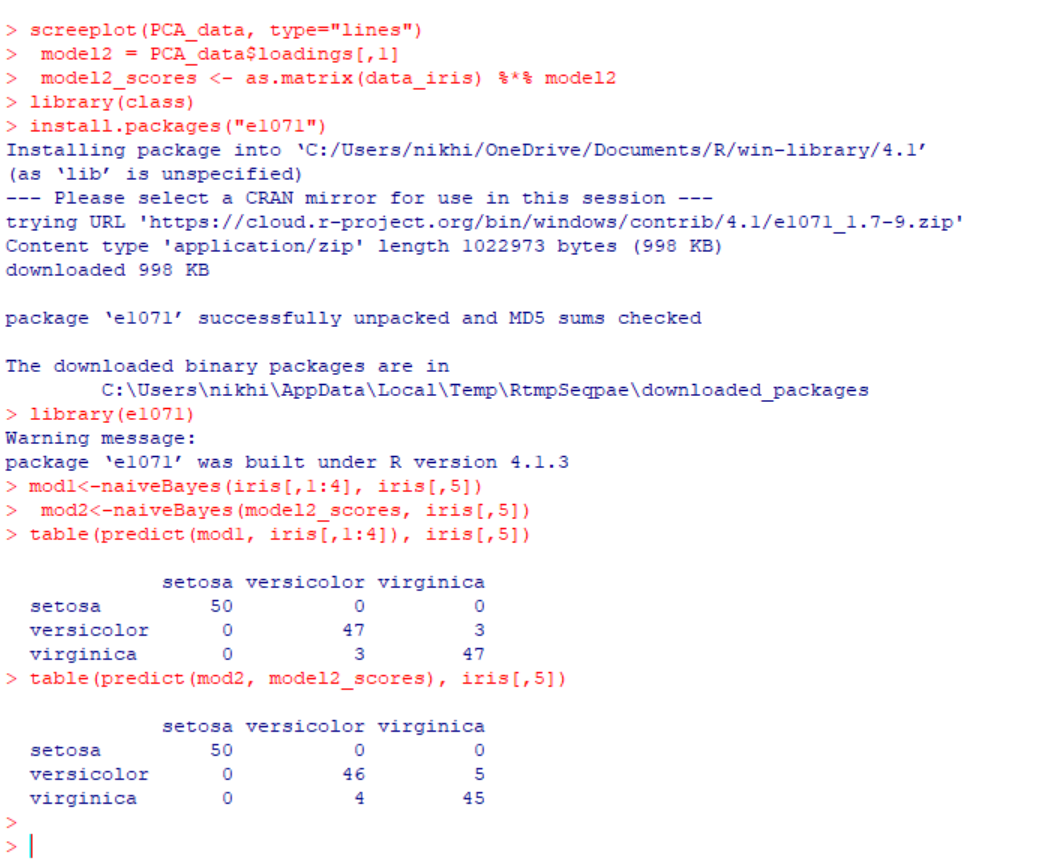


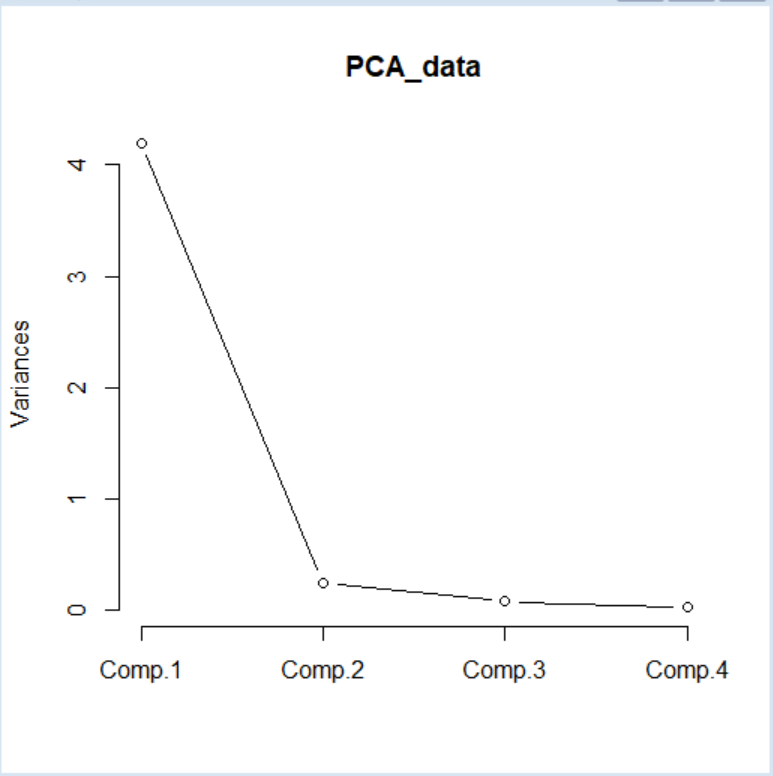
**PRACTICAL NO. 3**

**Aim: Practical of Principal Component Analysis**



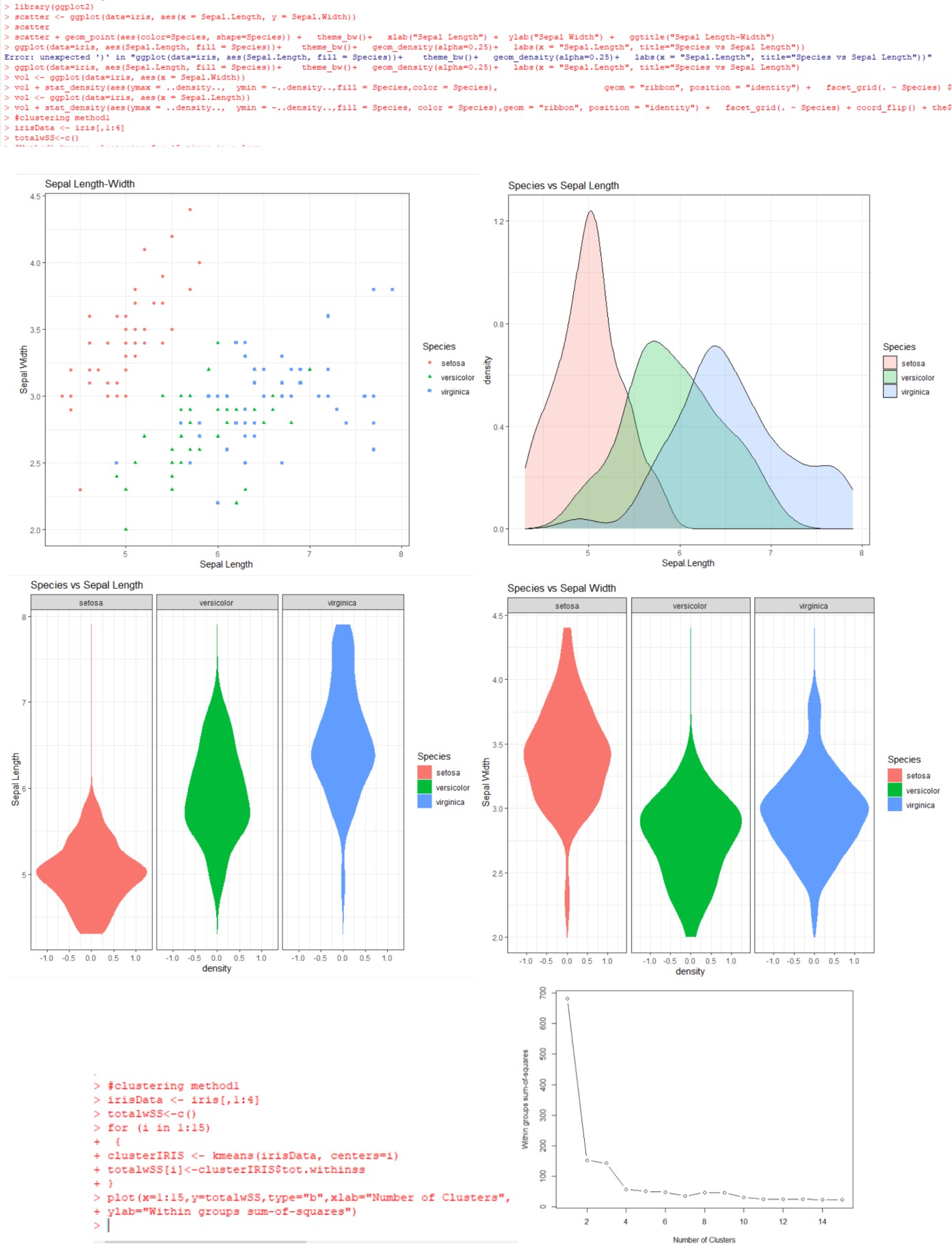


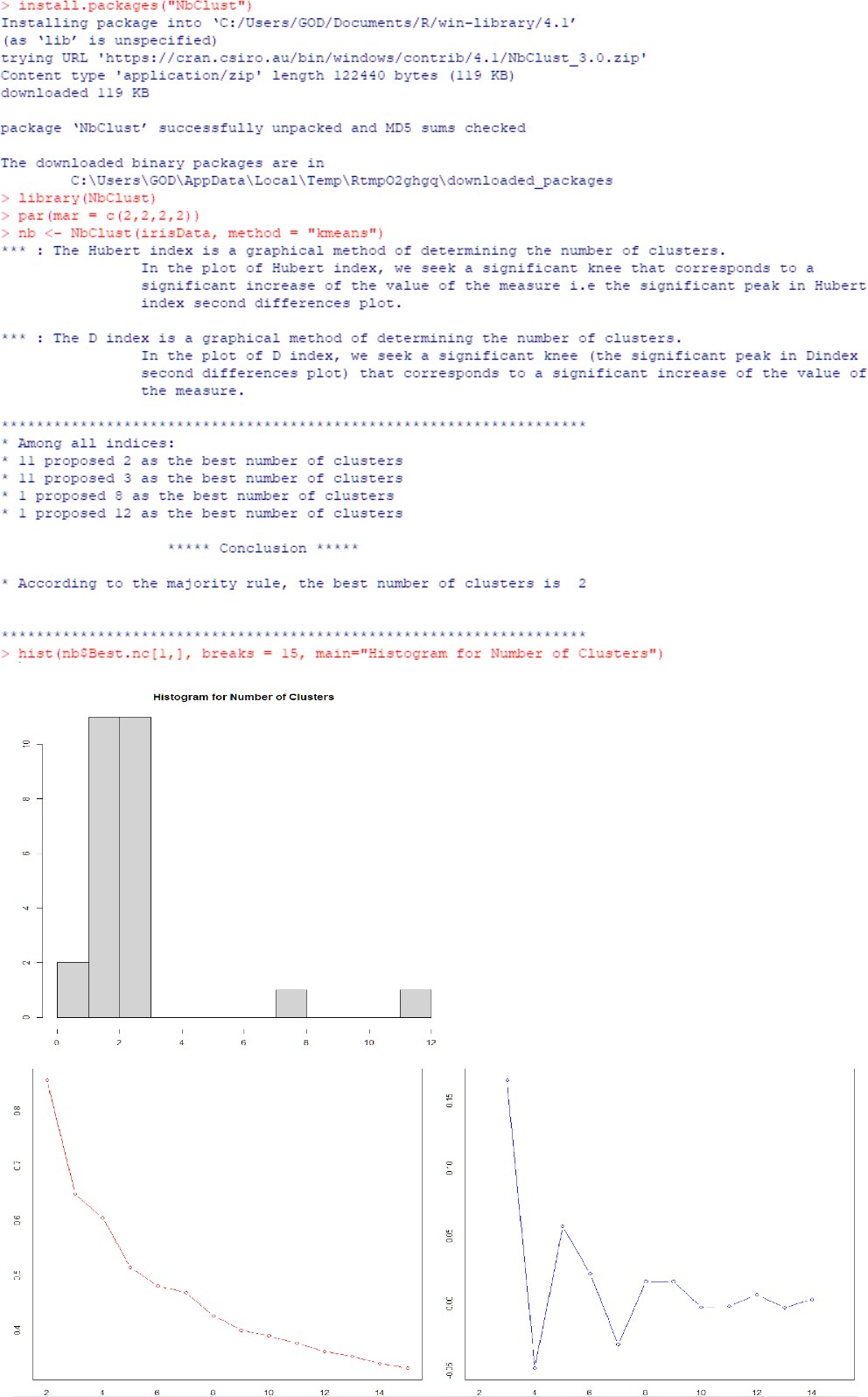


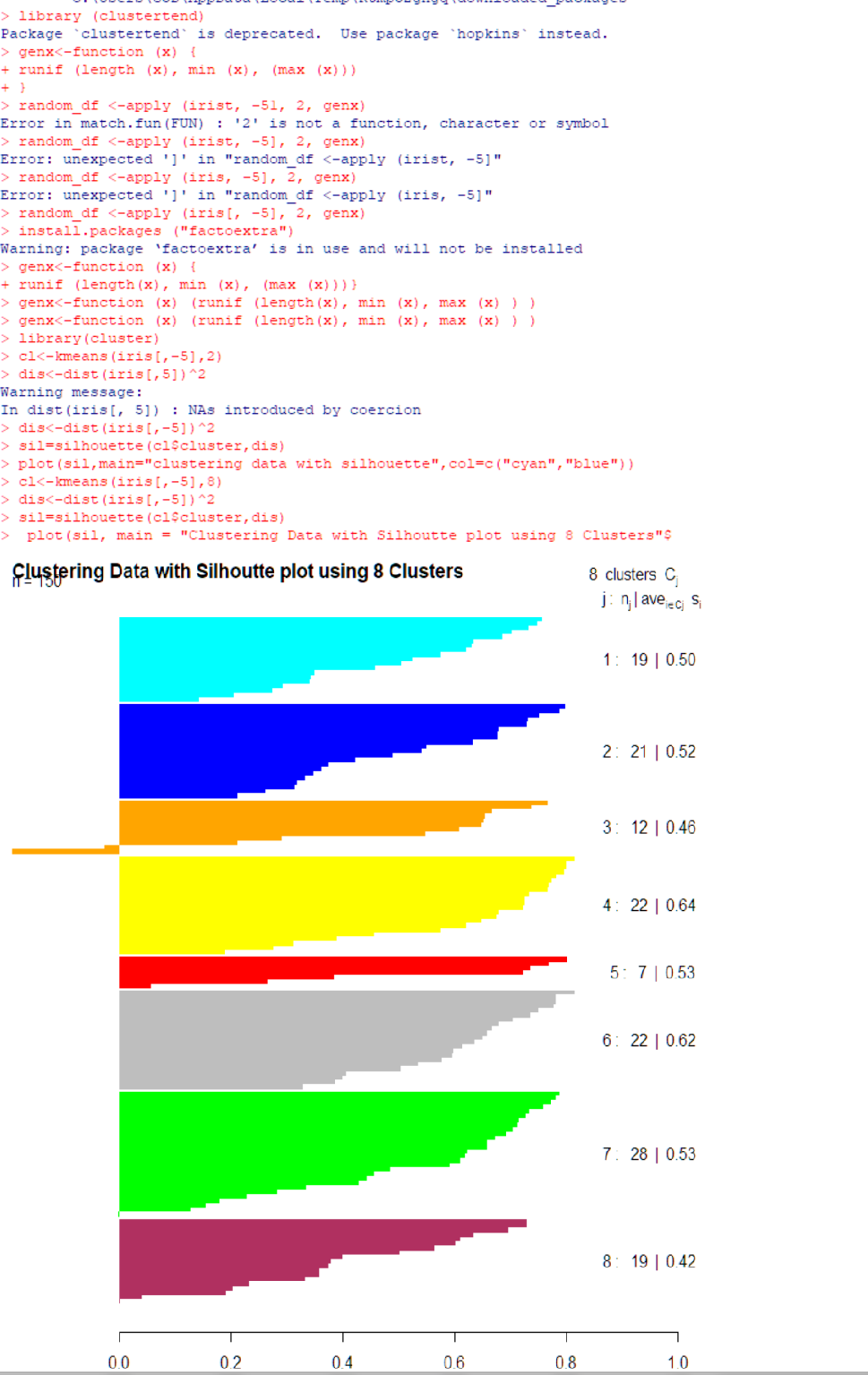


**PRACTICAL NO. 4**

**Aim: Practical of Clustering**







**PRACTICAL NO. 5**

**Aim: Practical of Time-series forecasting**

#load Data Airpassenger

data("AirPassengers")

#finding the class name

class(AirPassengers)

#Data In time series format

#start of time series

start(AirPassengers)

#Exp:05

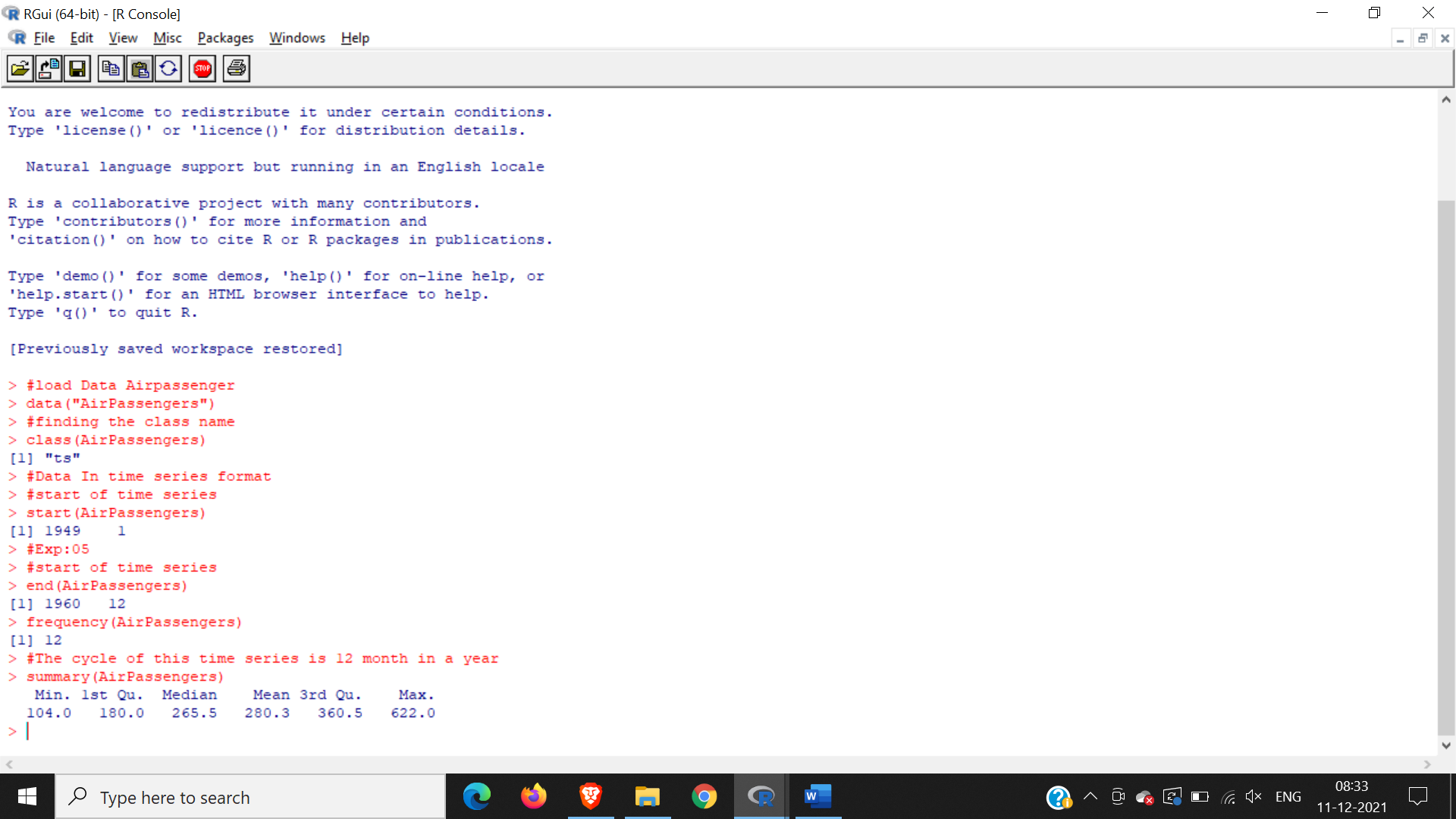
#start of time series

end(AirPassengers)

frequency(AirPassengers)

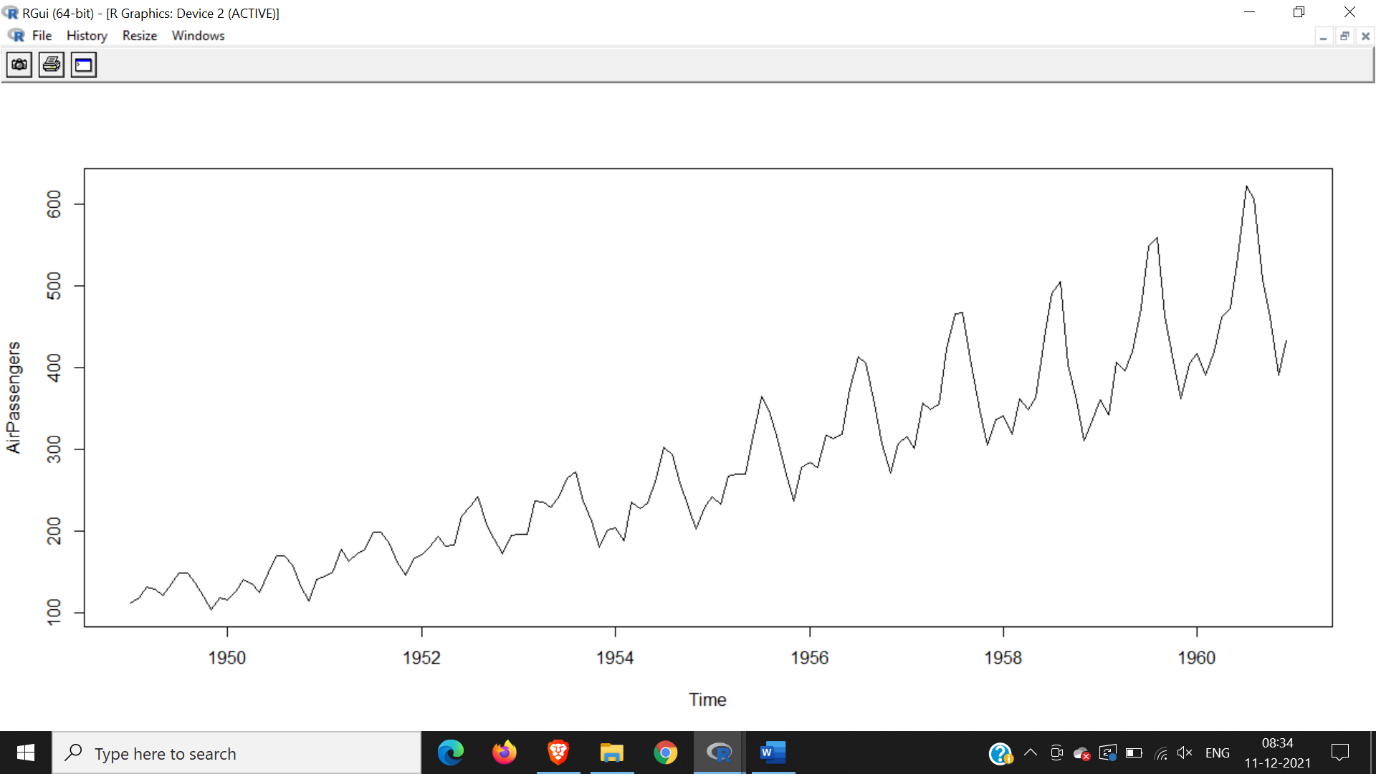
#The cycle of this time series is 12 month in a year

summary(AirPassengers)



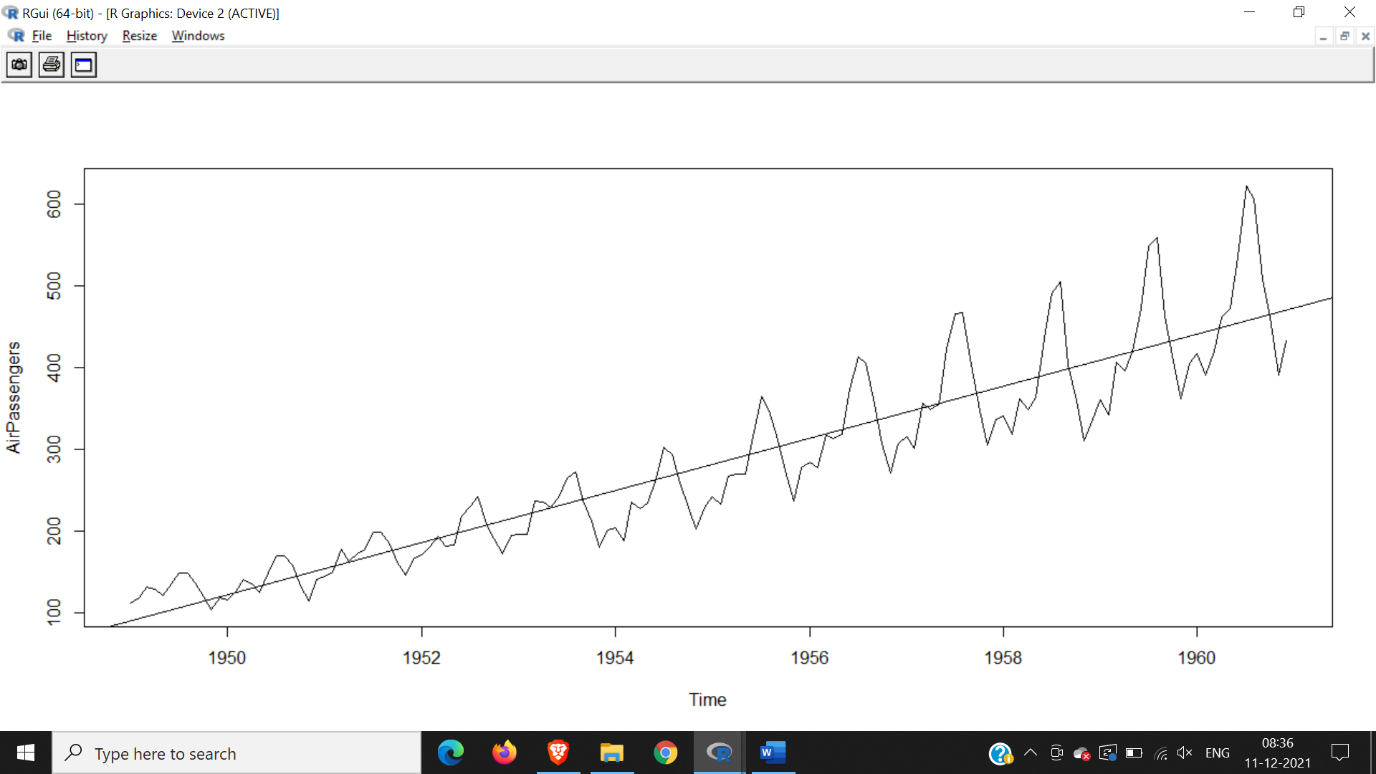
#the number of passenger are distributed across the spectrum

plot(AirPassengers)



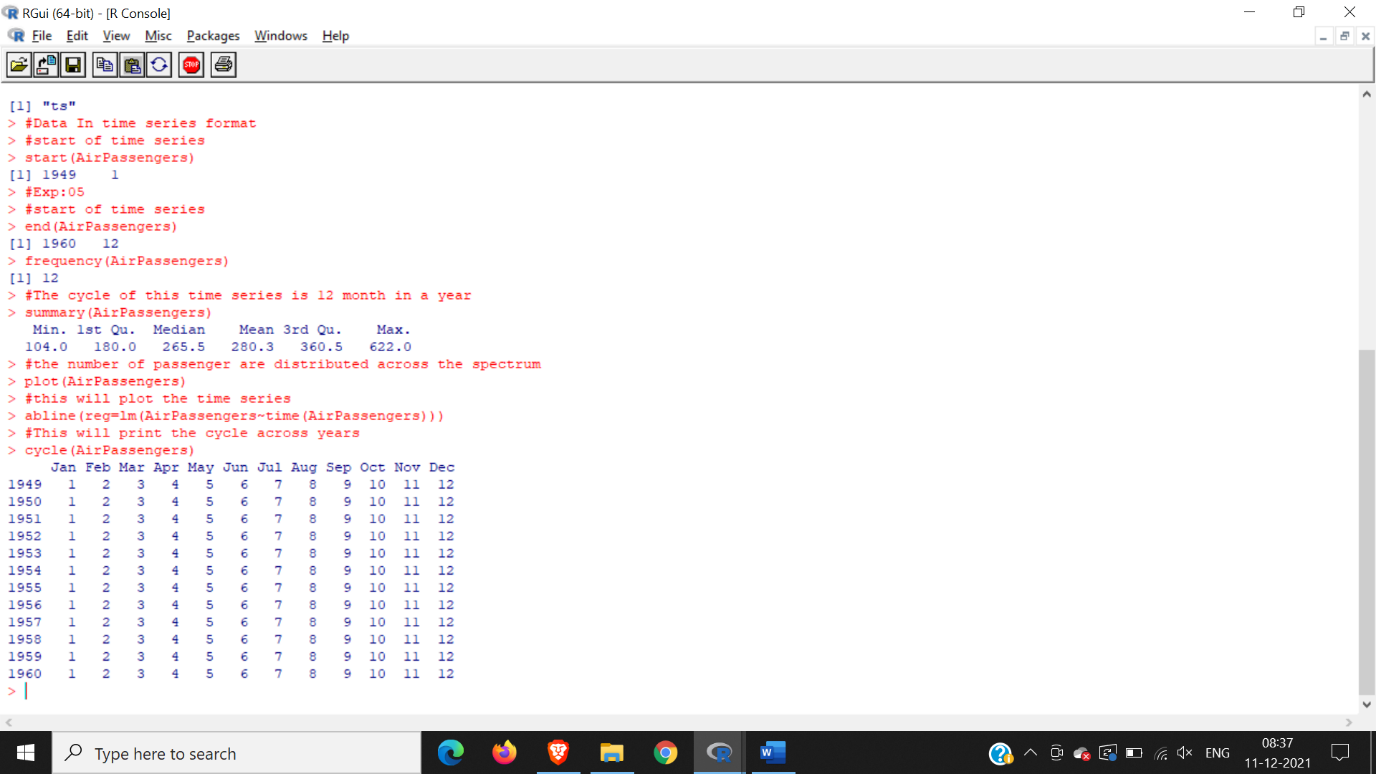
#this will plot the time series

abline(reg=lm(AirPassengers~time(AirPassengers)))



#This will print the cycle across years

cycle(AirPassengers)



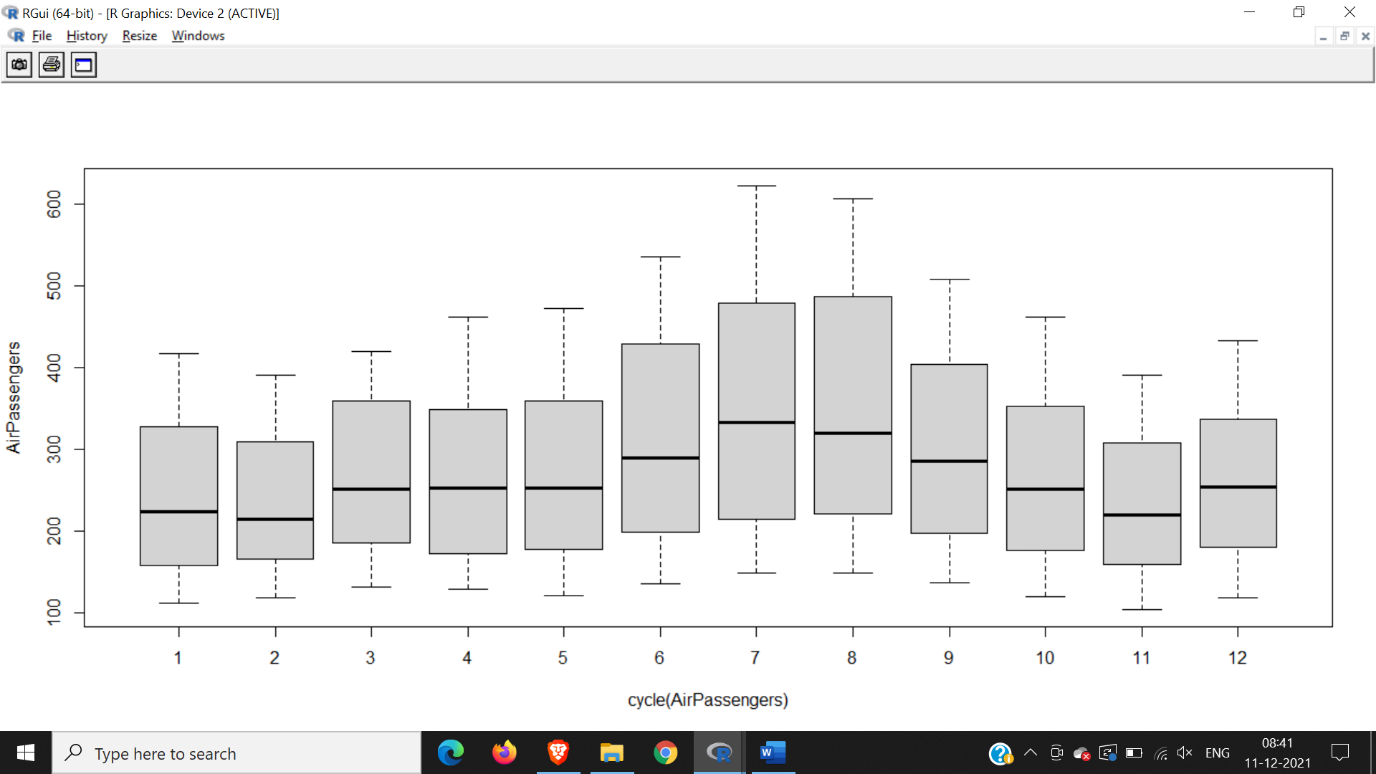
#this wil aggreate the cycles and display a year on year trend

plot(aggregate(AirPassengers,FUN=mean))

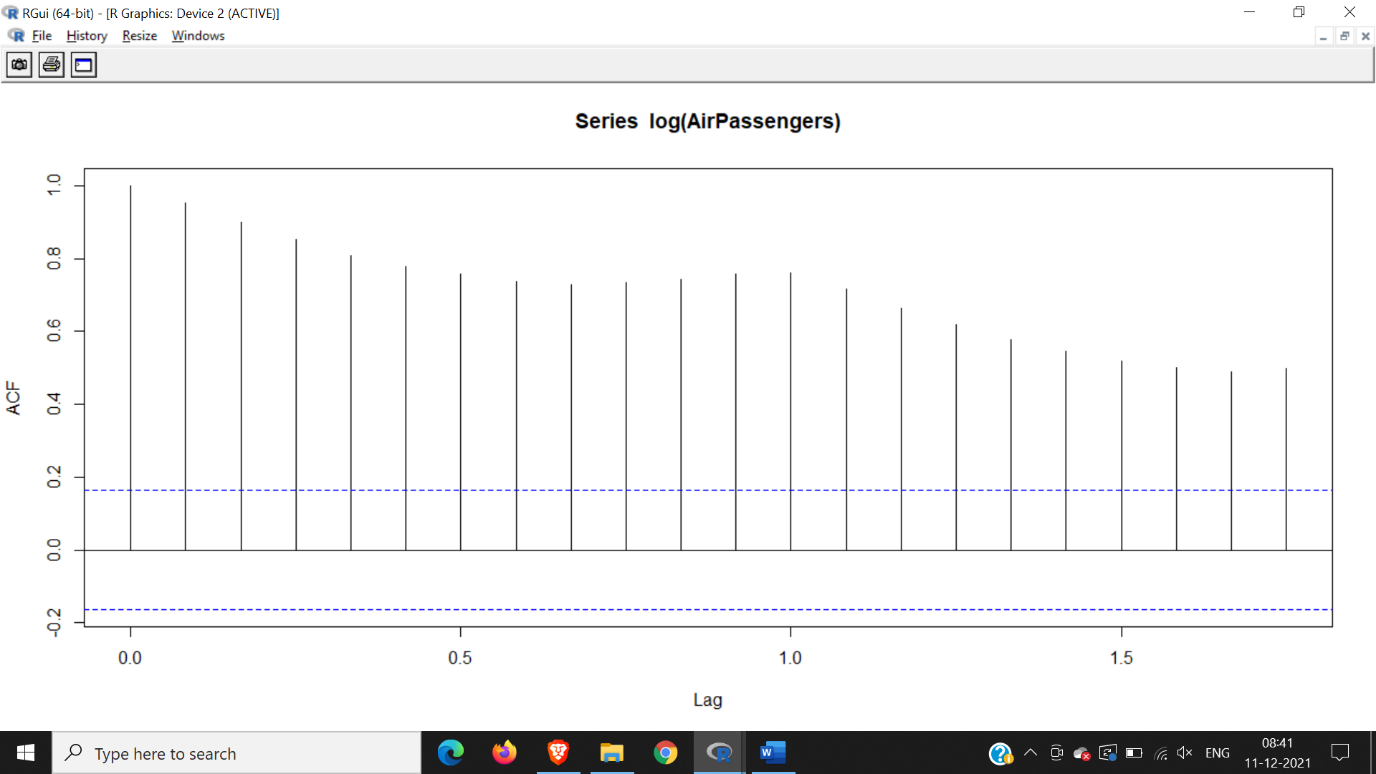


#box plot across month will give us a sense on seasonal effect

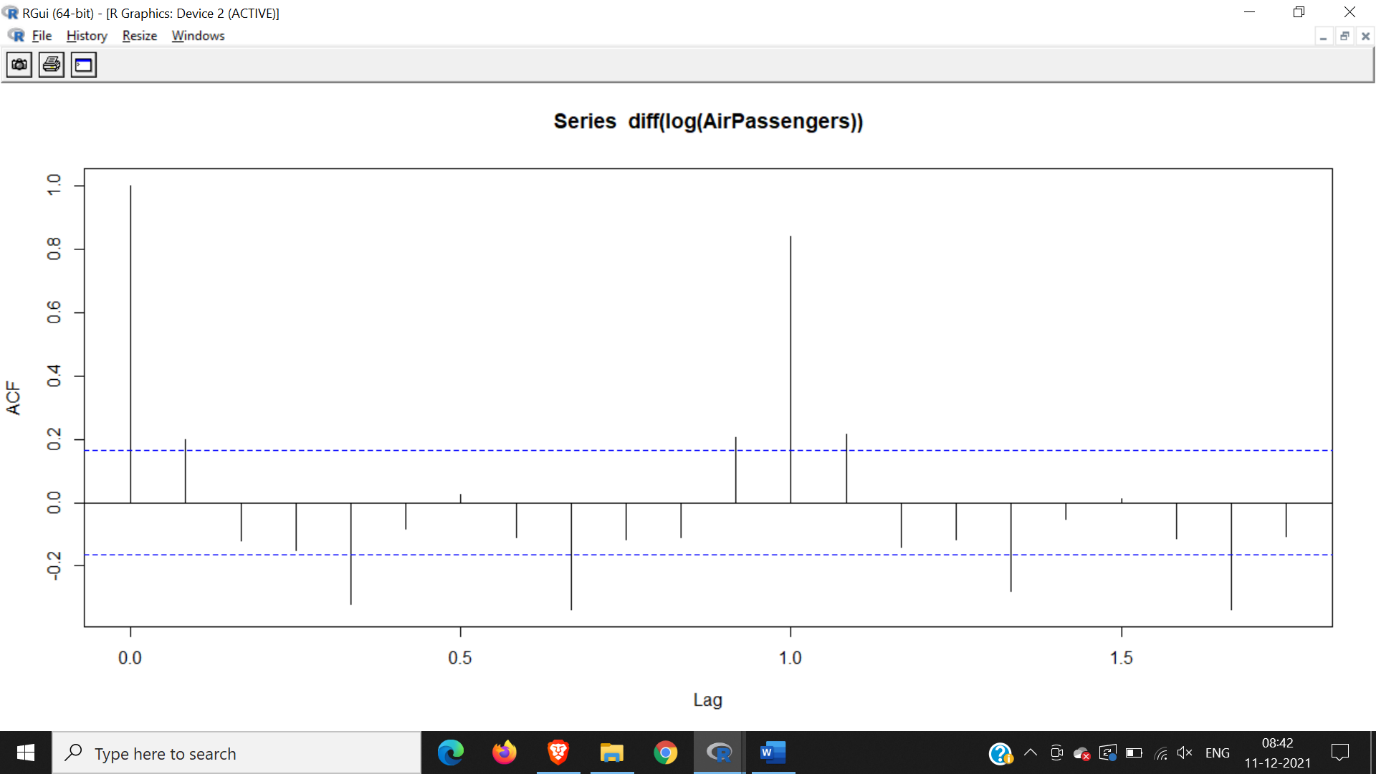
boxplot(AirPassengers~cycle(AirPassengers))



acf(log(AirPassengers))



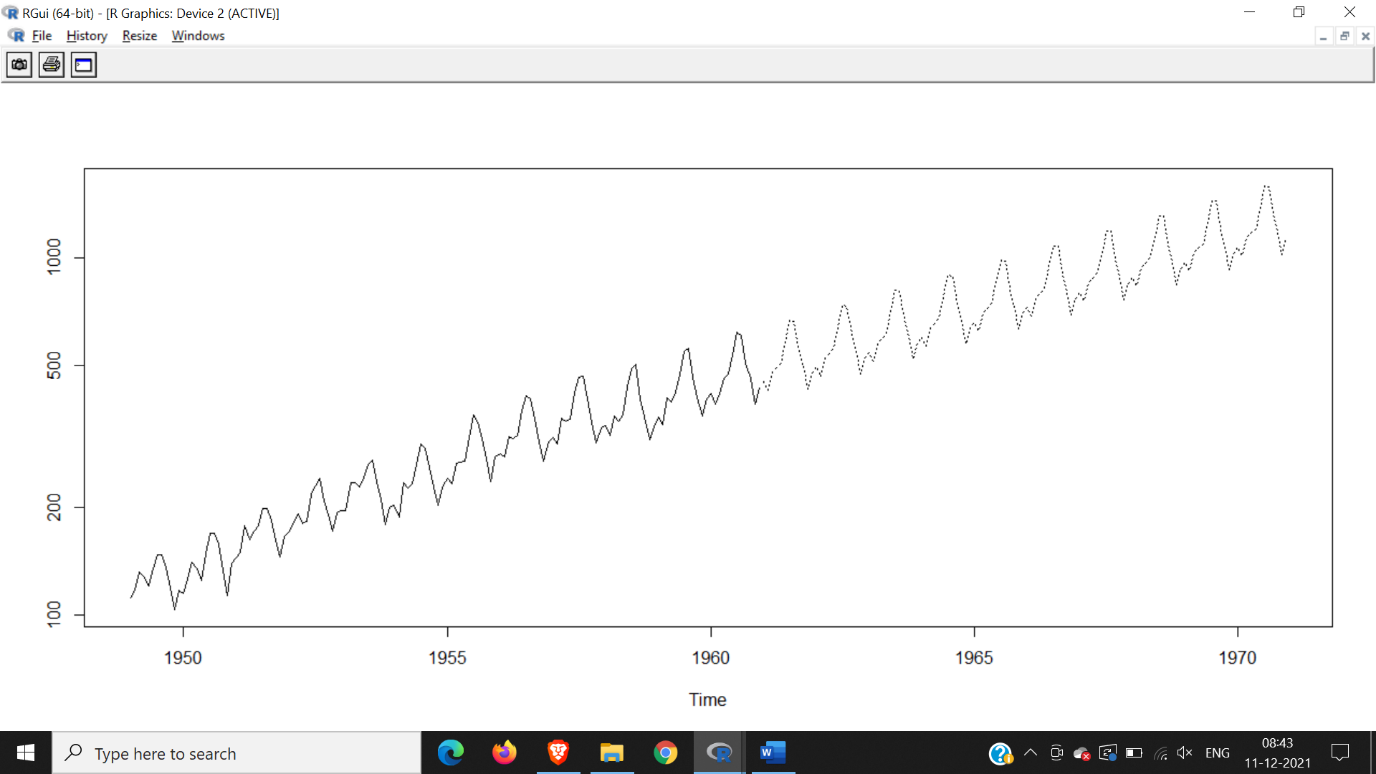
acf(diff(log(AirPassengers)))



(fit <- arima (log(AirPassengers), c(0, 1, 1), seasonal = list (order= c(0, 1,1),period=12)))

pred <- predict(fit, n.ahead = 10\*12)

ts.plot(AirPassengera,2.718^pred$pred,log="y",lty=c(1,3))

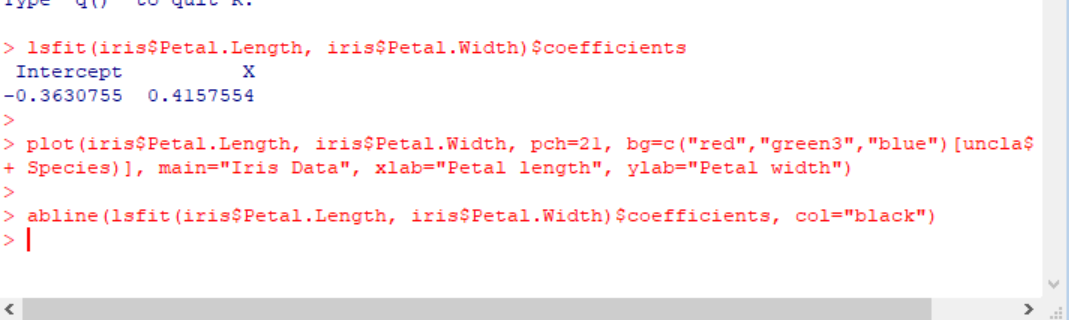


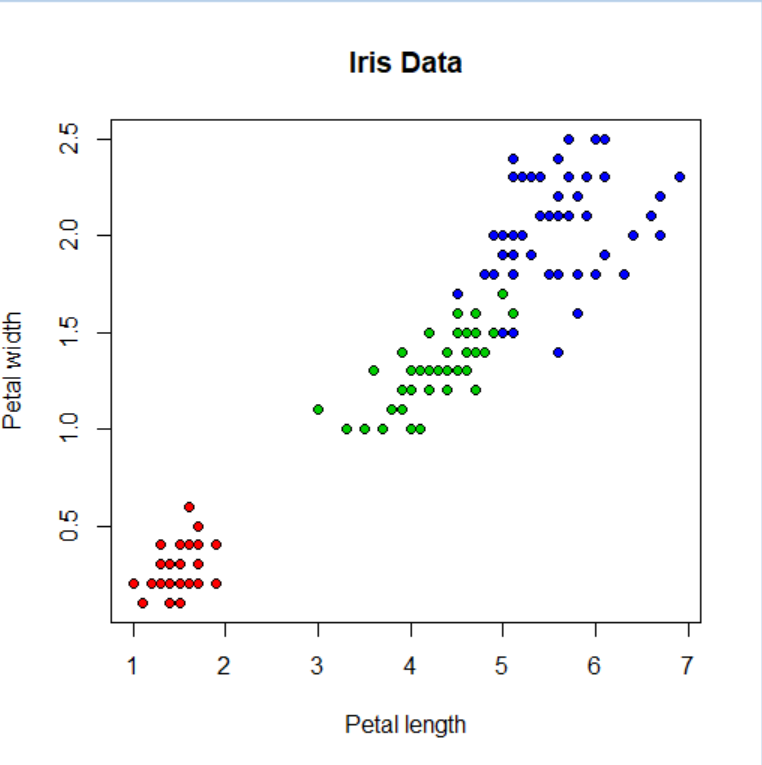
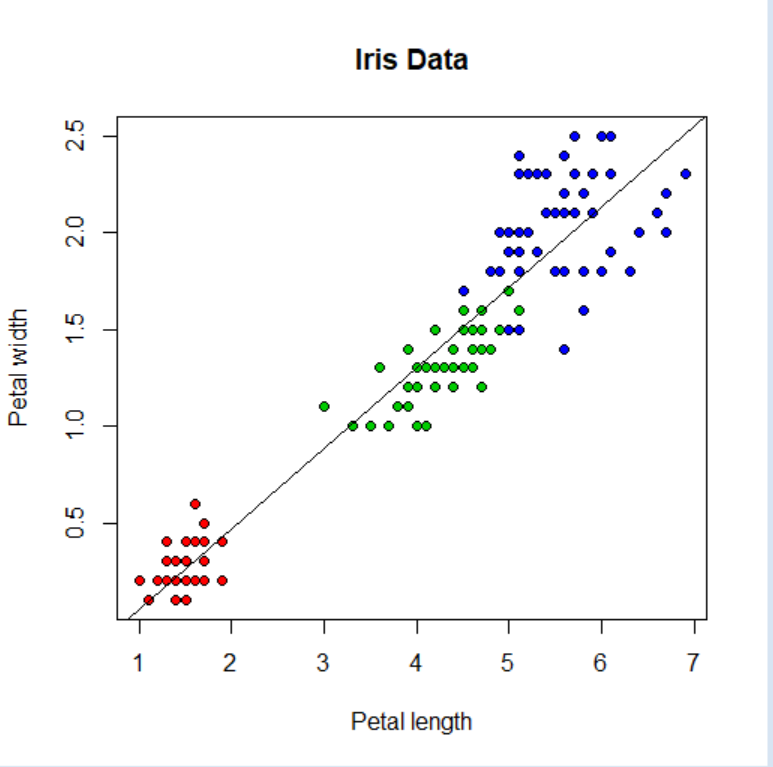
Hence we conclude there will be a rise in passengers

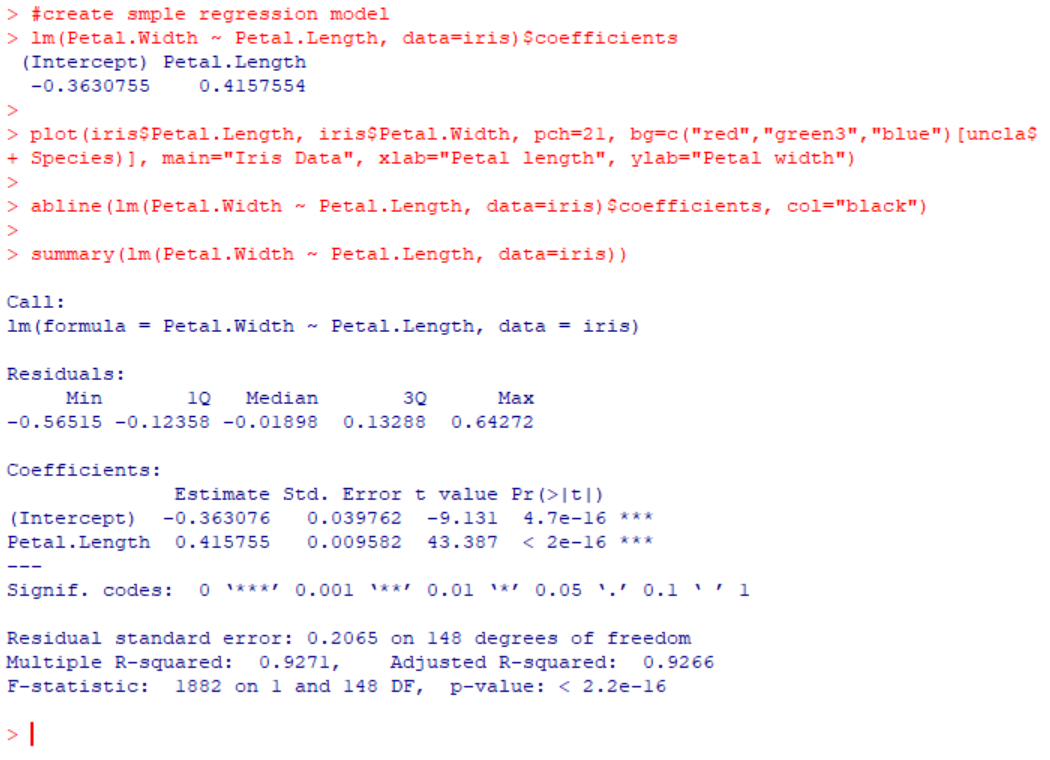
**PRACTICAL NO. 6**

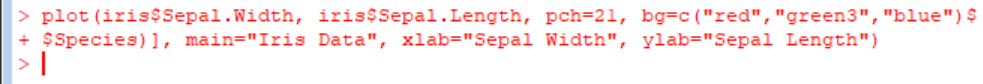
**Aim: Practical of Simple/Multiple Linear Regression**

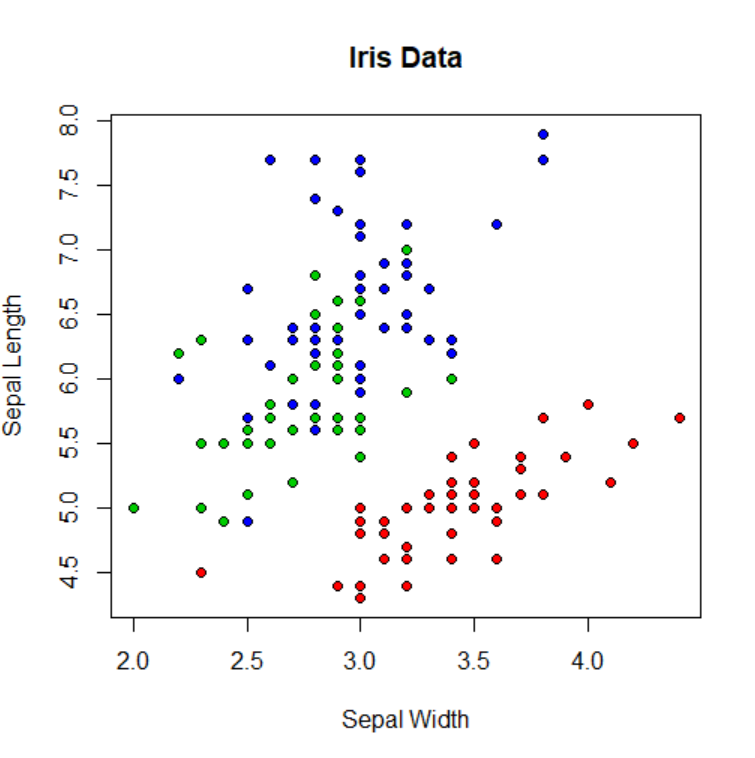
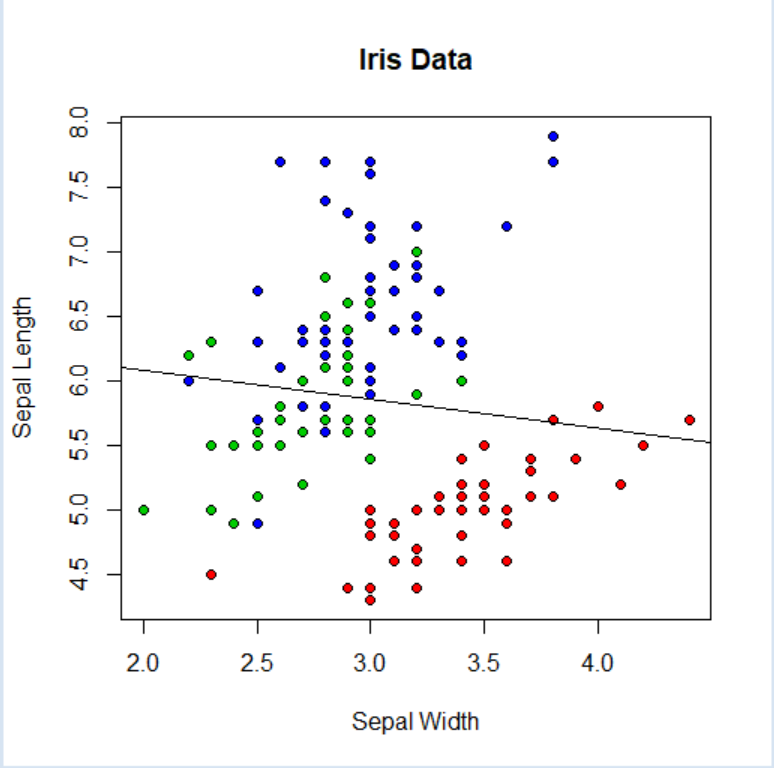
**SIMPLE :**

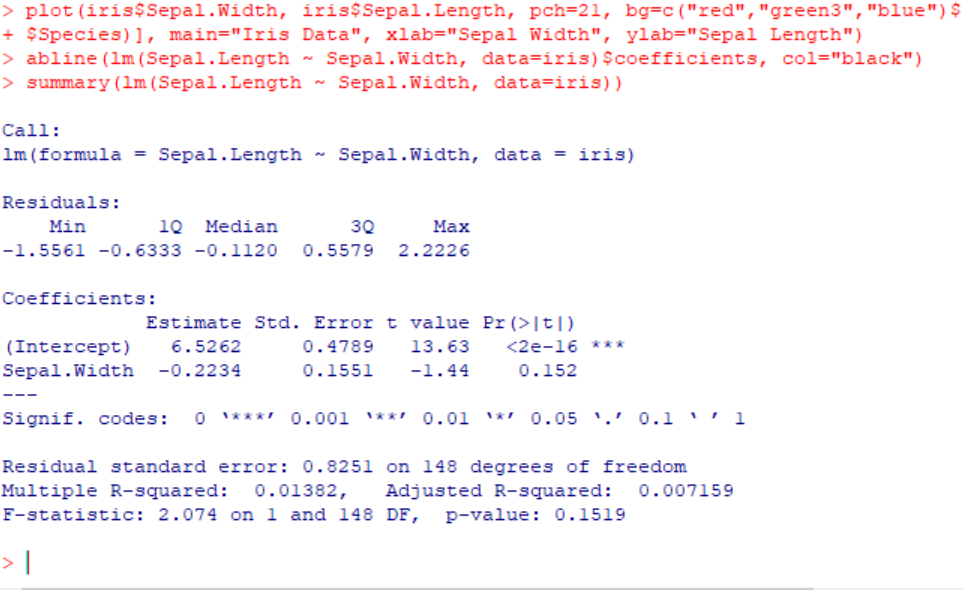






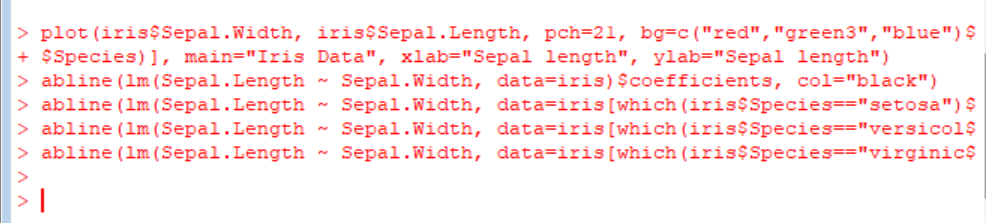
 

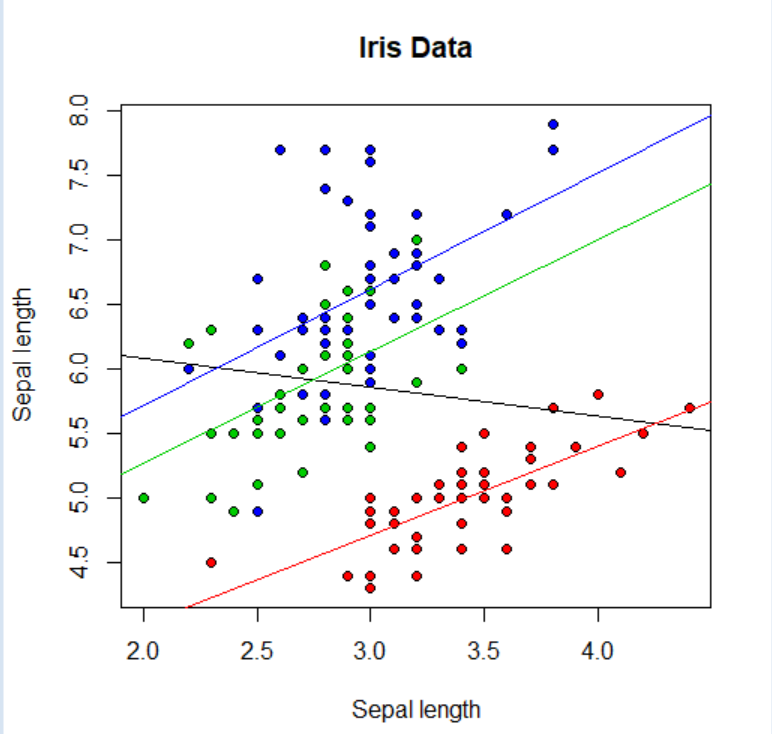


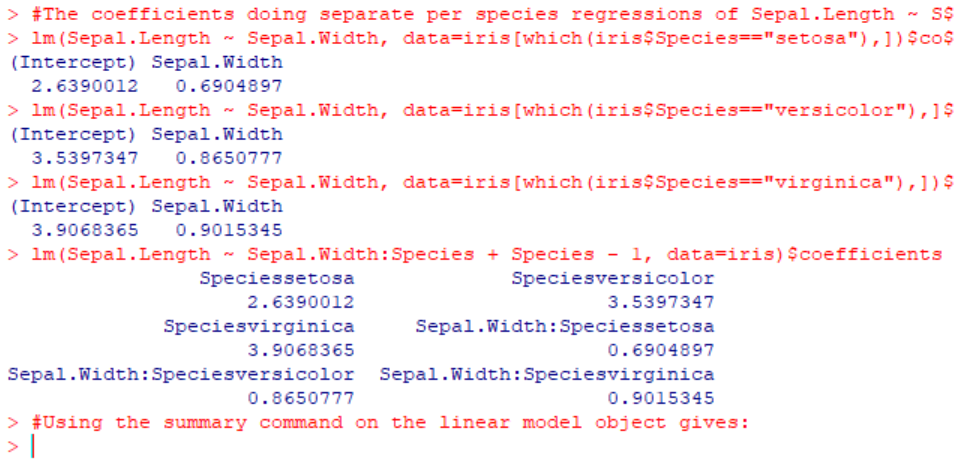
**Multiple:**

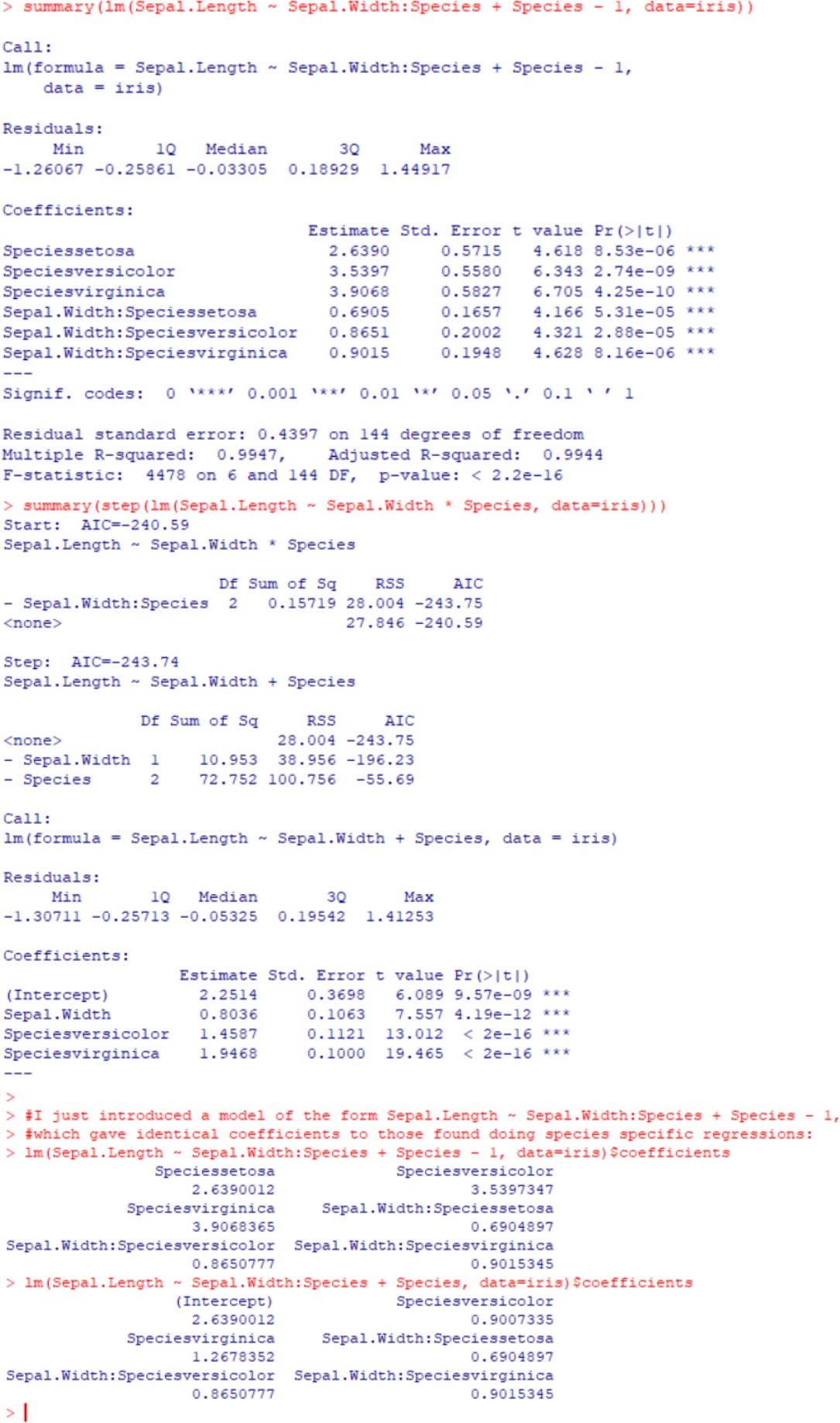
#What happens if we divide the data up by species, and run three separate linear regressions?





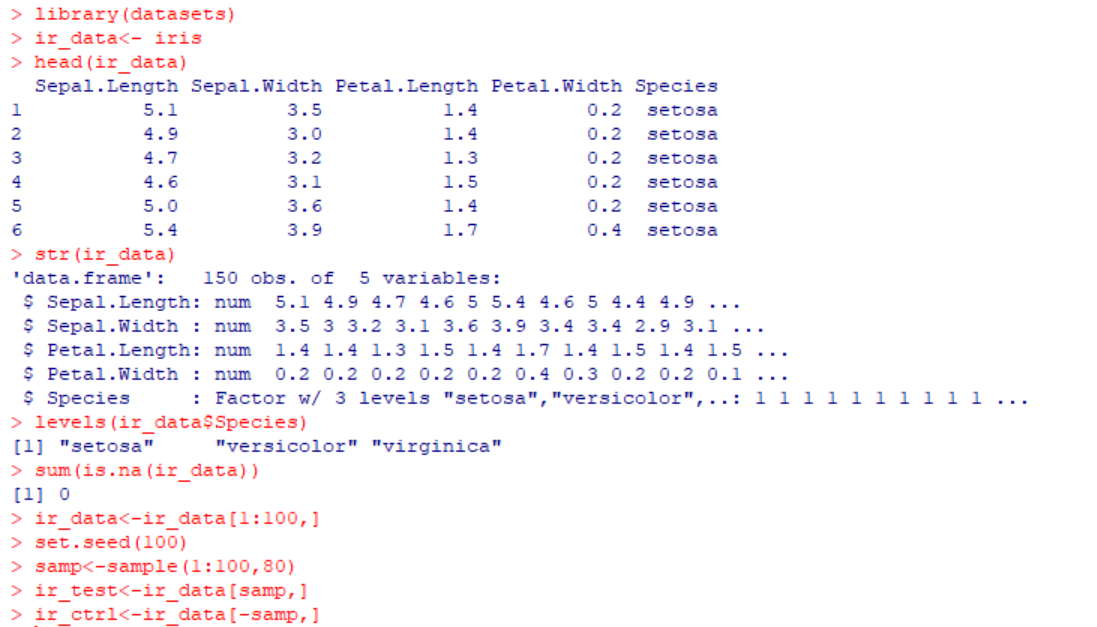


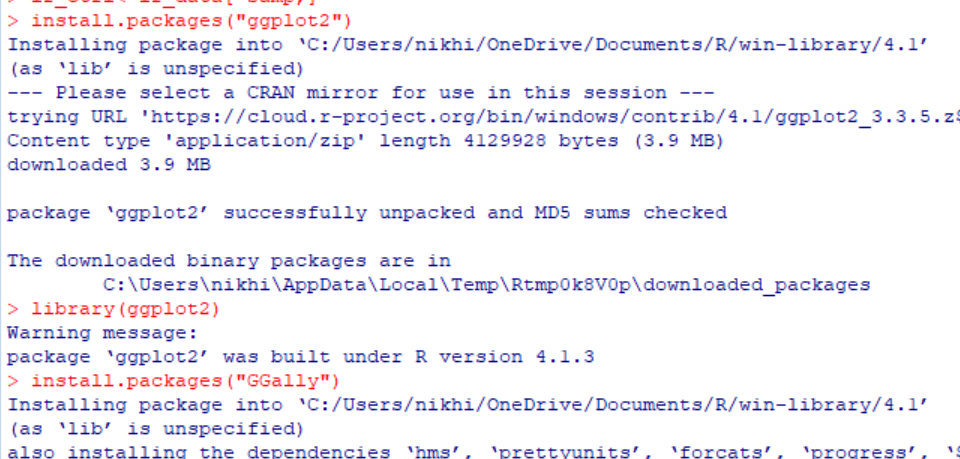
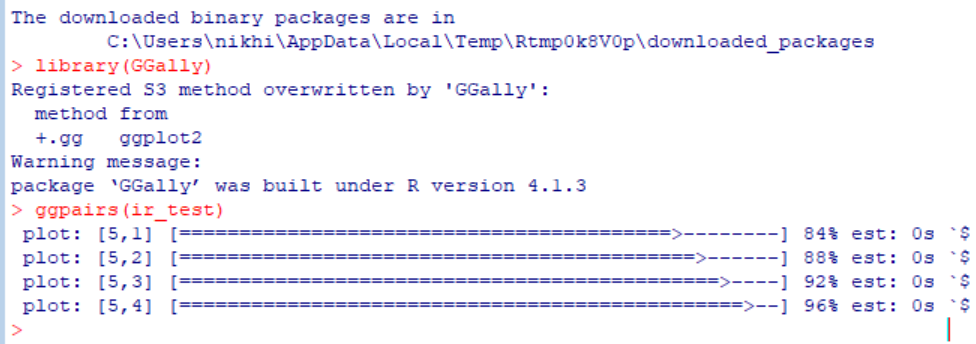


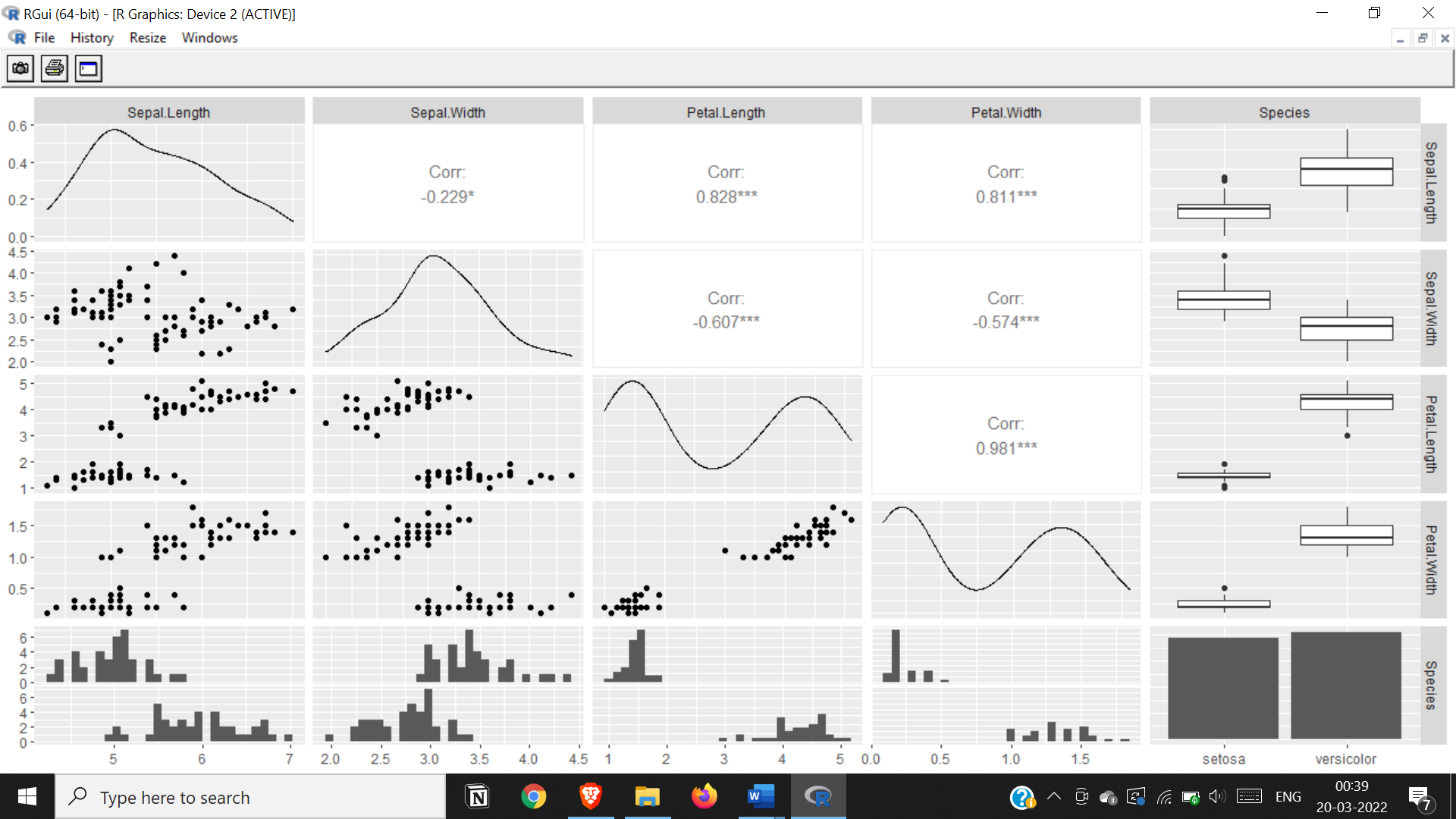


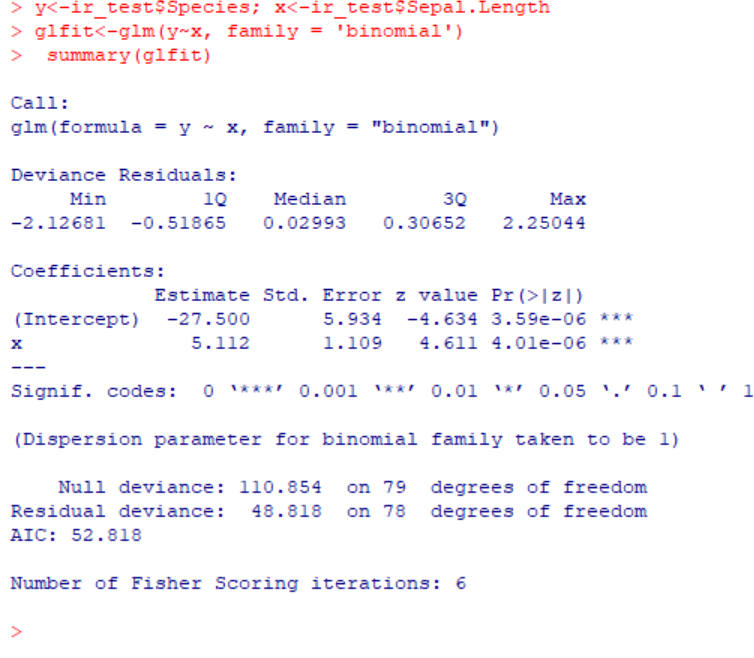
**PRACTICAL NO. 7**

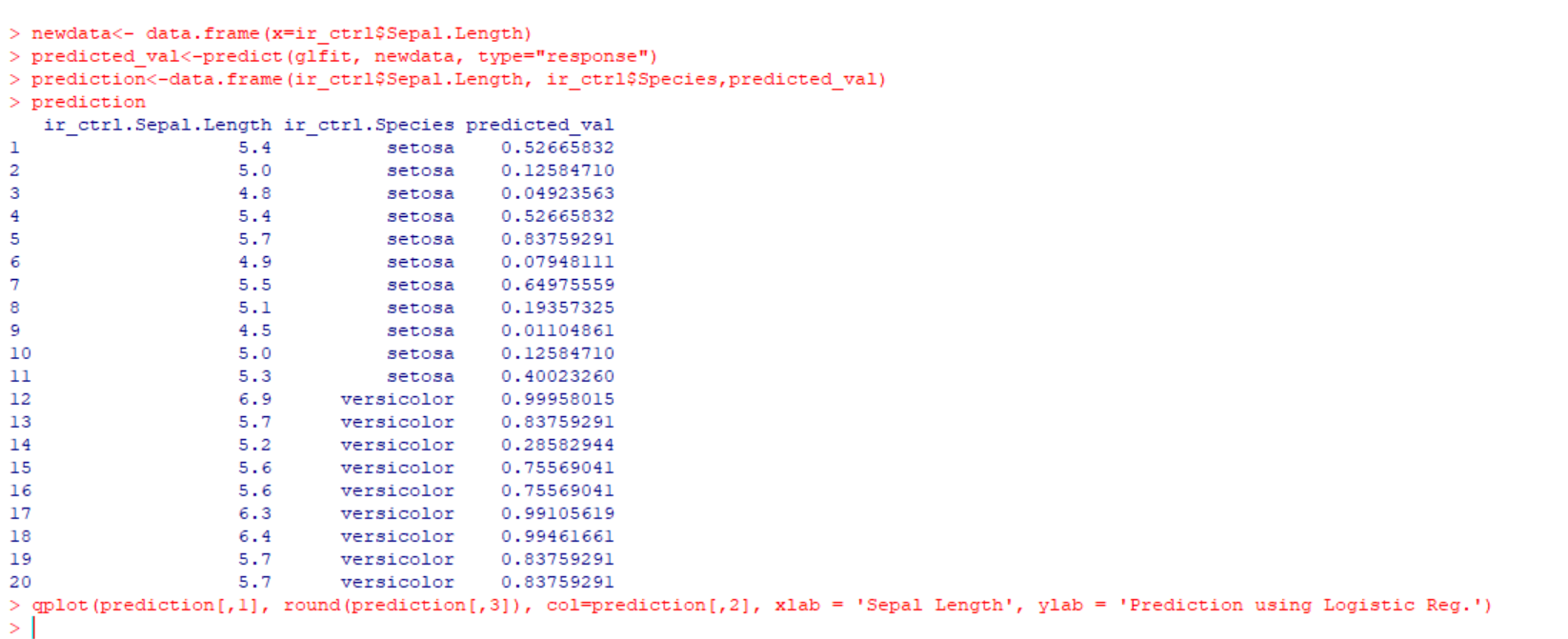
**Aim: Practical of Logistics Regression**

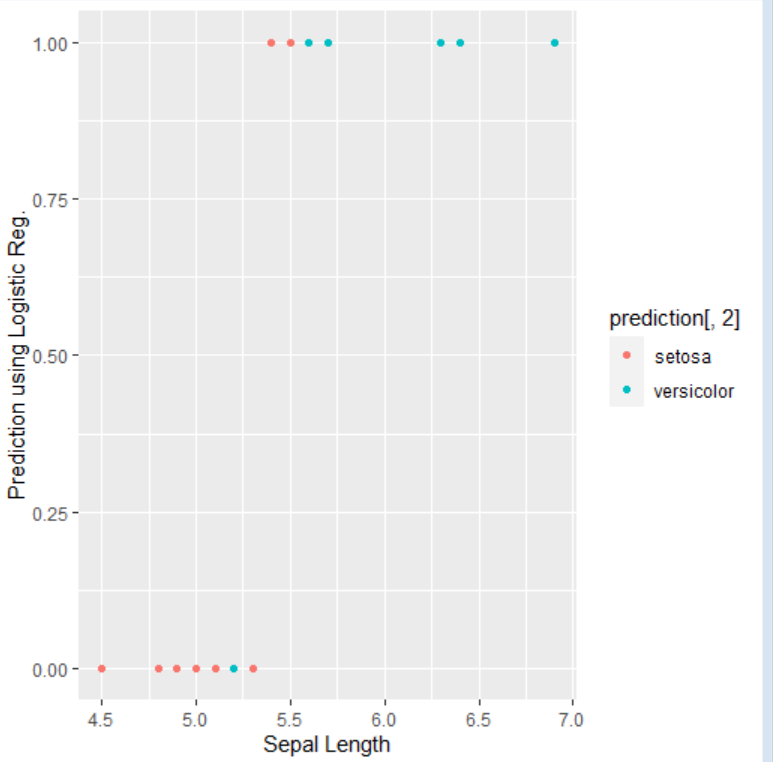






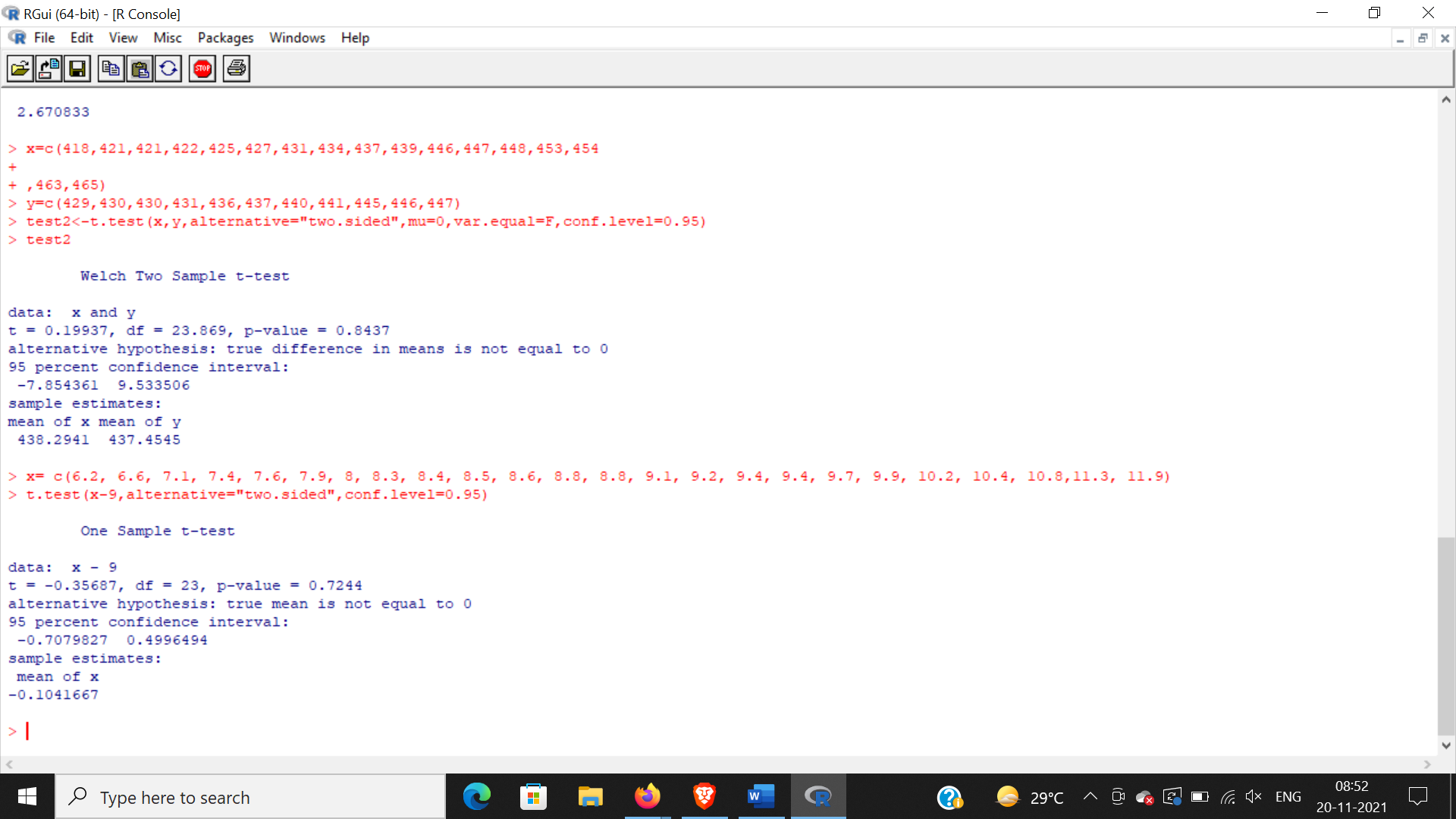


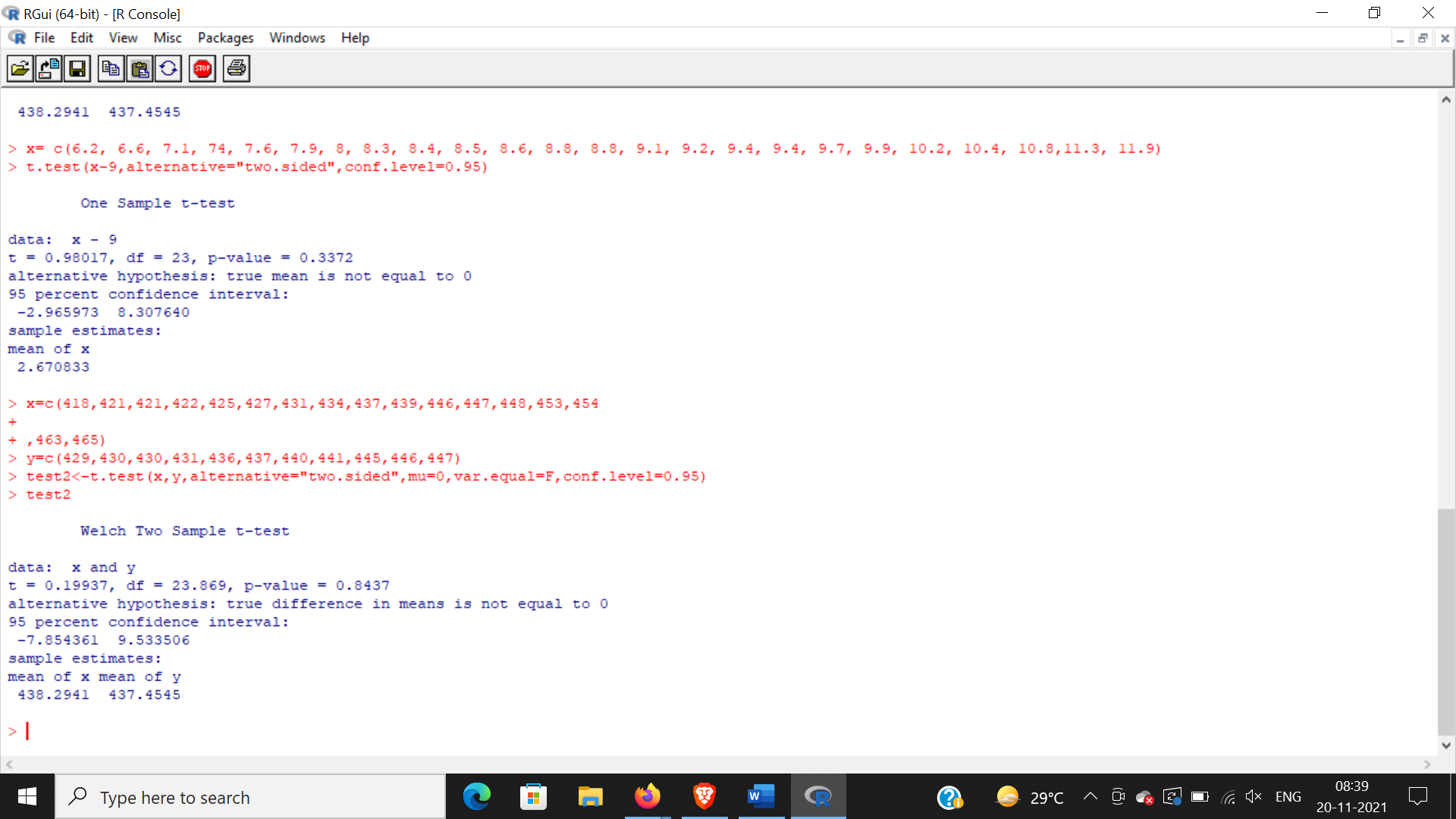


**Practical 8**

**Aim: practical of Hypothesis**

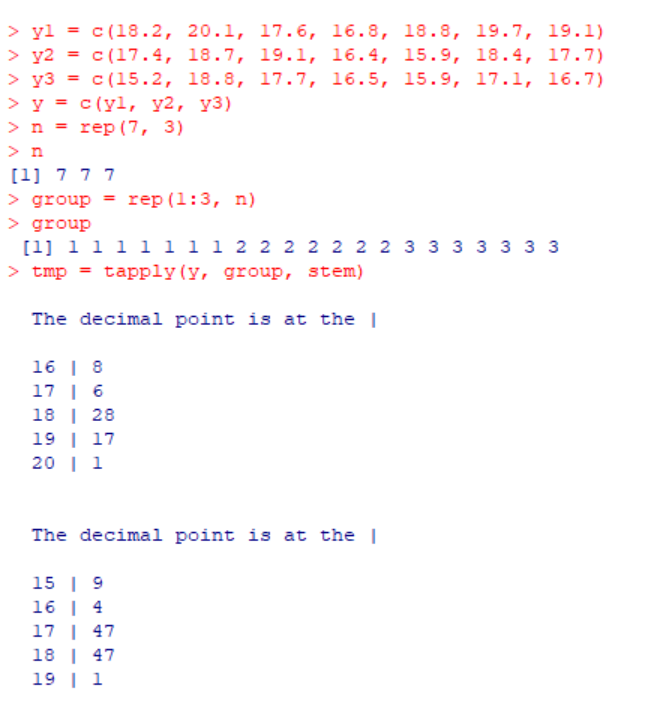
#t.test(dataset,dataset,alternative method, mu value, var.equal=F, conf.level=0.95)

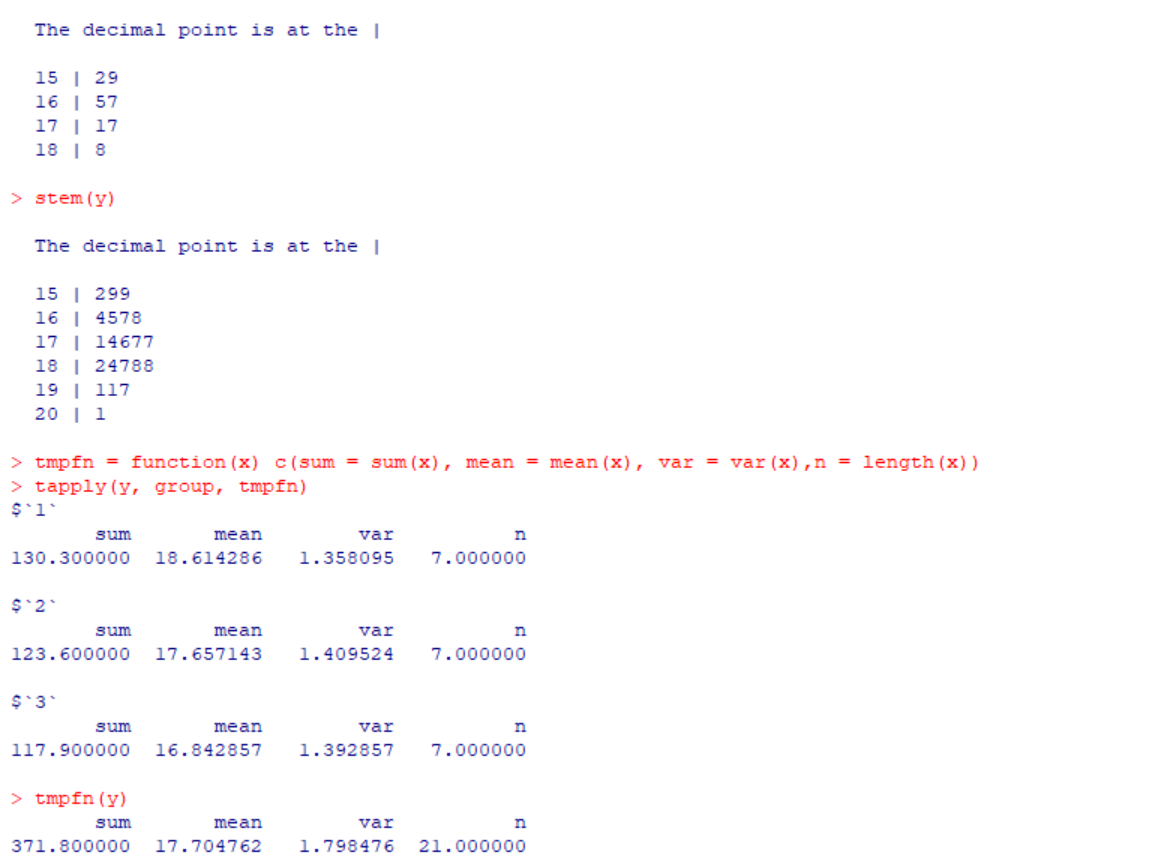


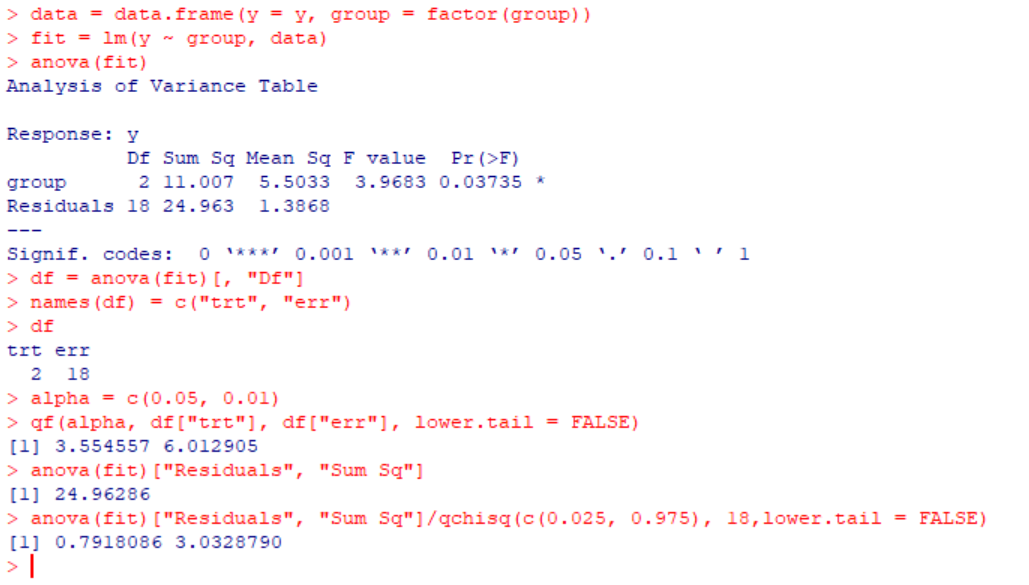


**PRACTICAL NO. 9**

**Aim: Practical of Analysis of Variance**

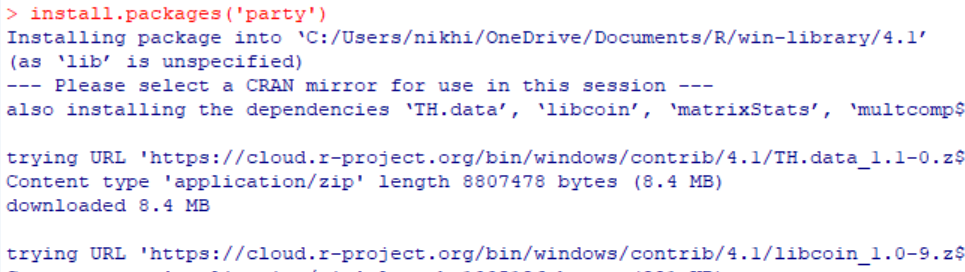


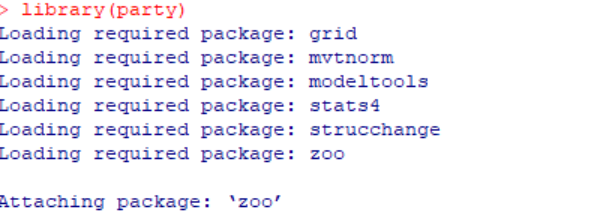


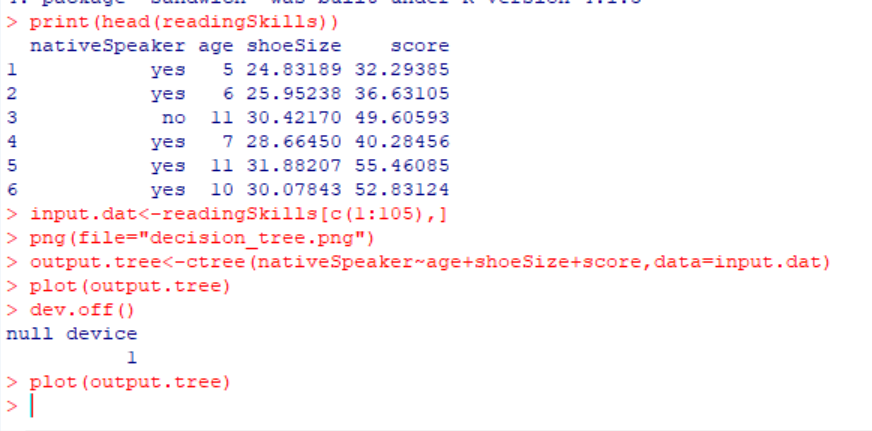


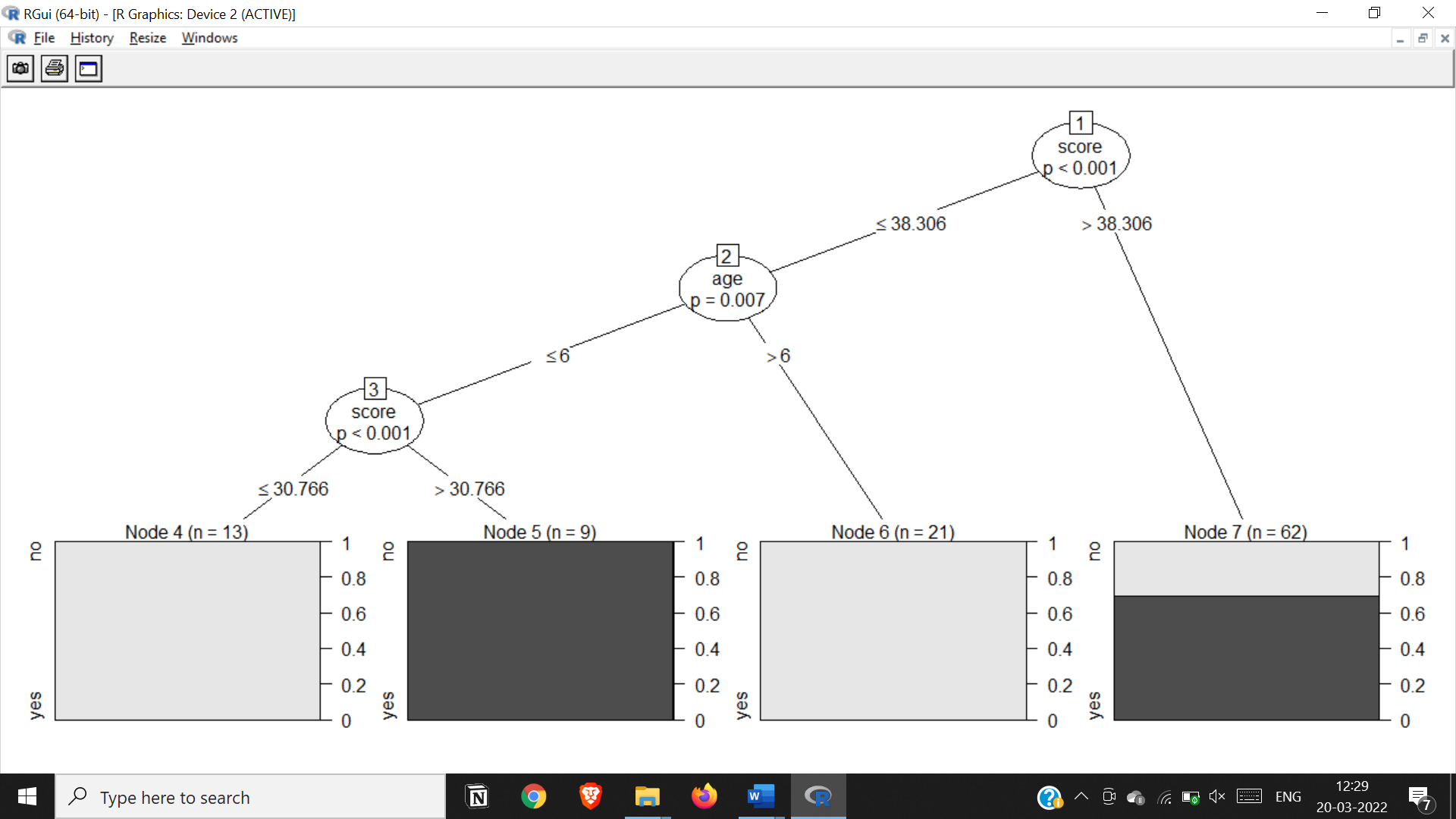
**PRACTICAL NO. 10**

**Aim: Practical of Decision Tree**









Conclusion: We conclude that anyone whose reading skill score is less than 38.3 and age is more than 6 is a native speaker