**Practical No:02**

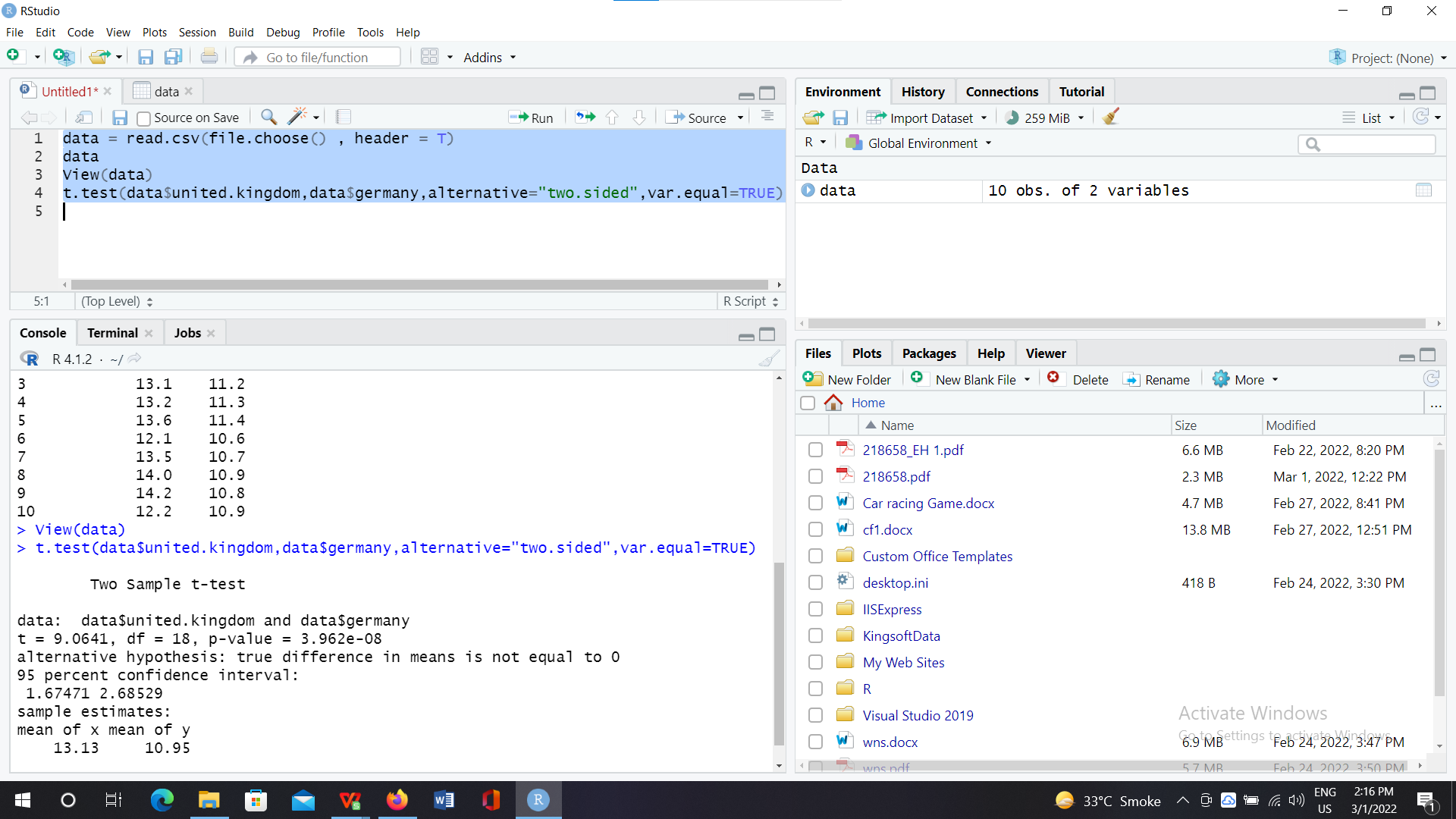
**Aim: Practical Of Hypothesis Testing**

**1) Independent ‘t’ Test**

A researcher studied education in the united kingdom and germany wanted to compare how many years of on average women in each country spent in school .The Researched obtained the random sample from both the countries. Test whether the average number of years spent in the school by women in 2 countries are equal or not.

|  |  |
| --- | --- |
| united kingdom | germany |
| 12.8 | 10.8 |
| 12.6 | 10.9 |
| 13.1 | 11.2 |
| 13.2 | 11.3 |
| 13.6 | 11.4 |
| 12.1 | 10.6 |
| 13.5 | 10.7 |
| 14 | 10.9 |
| 14.2 | 10.8 |
| 12.2 | 10.9 |

**Output:**



**Conclusion:**

When alternative = ‘two.sided’,

The p-value = 3.962e-08 which is less than 0.05 therefore **we reject the hypothesis H0**.

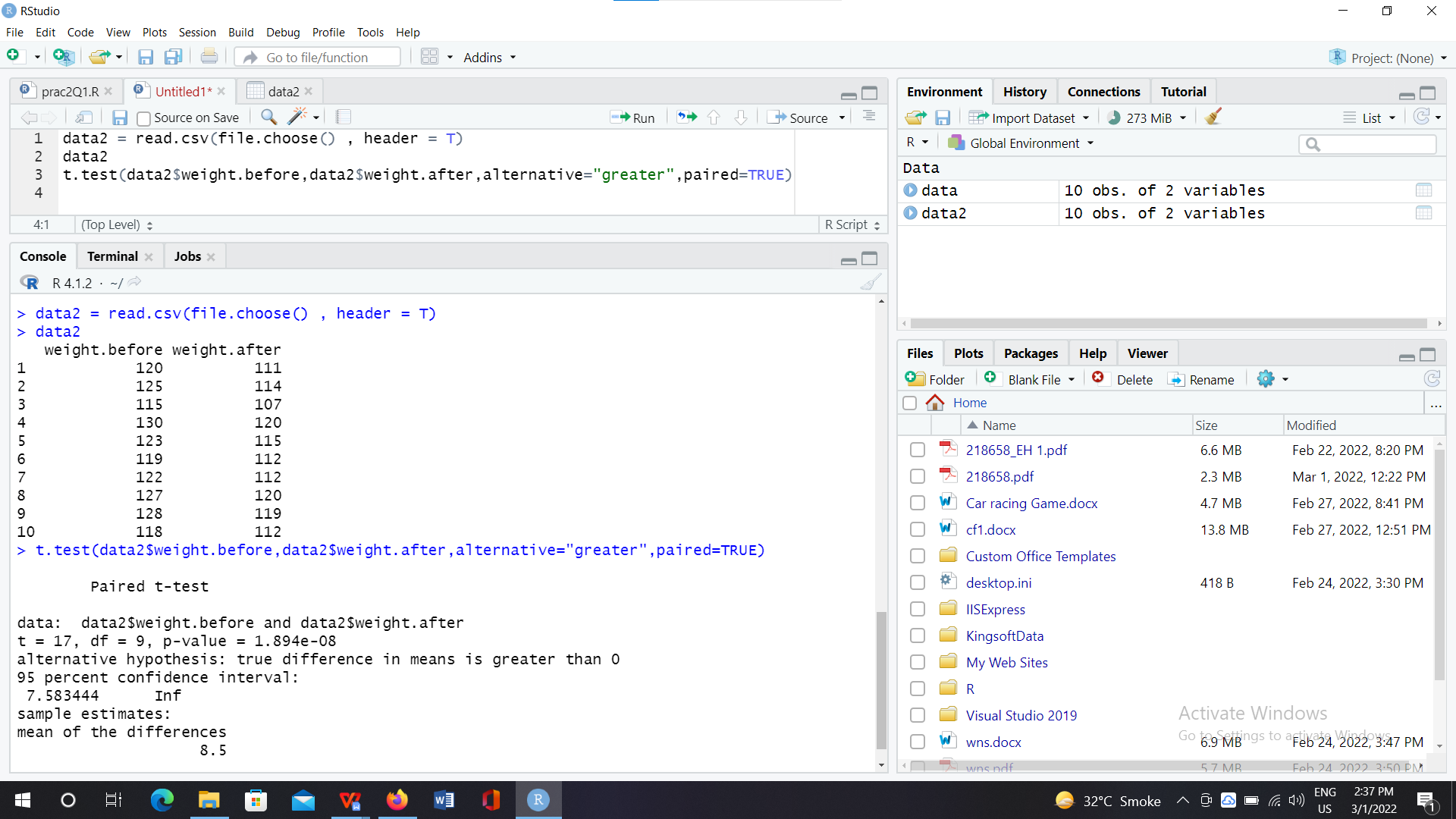
Therefore Average number of years spend in school by women in two contries are not equal 0.

**2) Paired ‘t’ Test**

A health club advertised a weight reducing program and claimed that on an average a participant in an program loses weight in 6 months.A person wanted to, verify the claim ,the club allowed him to select randomly the records of 10 participants about their weights before and after the program.

|  |  |
| --- | --- |
| weight before | weight after |
| 120 | 111 |
| 125 | 114 |
| 115 | 107 |
| 130 | 120 |
| 123 | 115 |
| 119 | 112 |
| 122 | 112 |
| 127 | 120 |
| 128 | 119 |
| 118 | 112 |

**Output:**



**Conclusion:**

When alternative = ‘greater’,

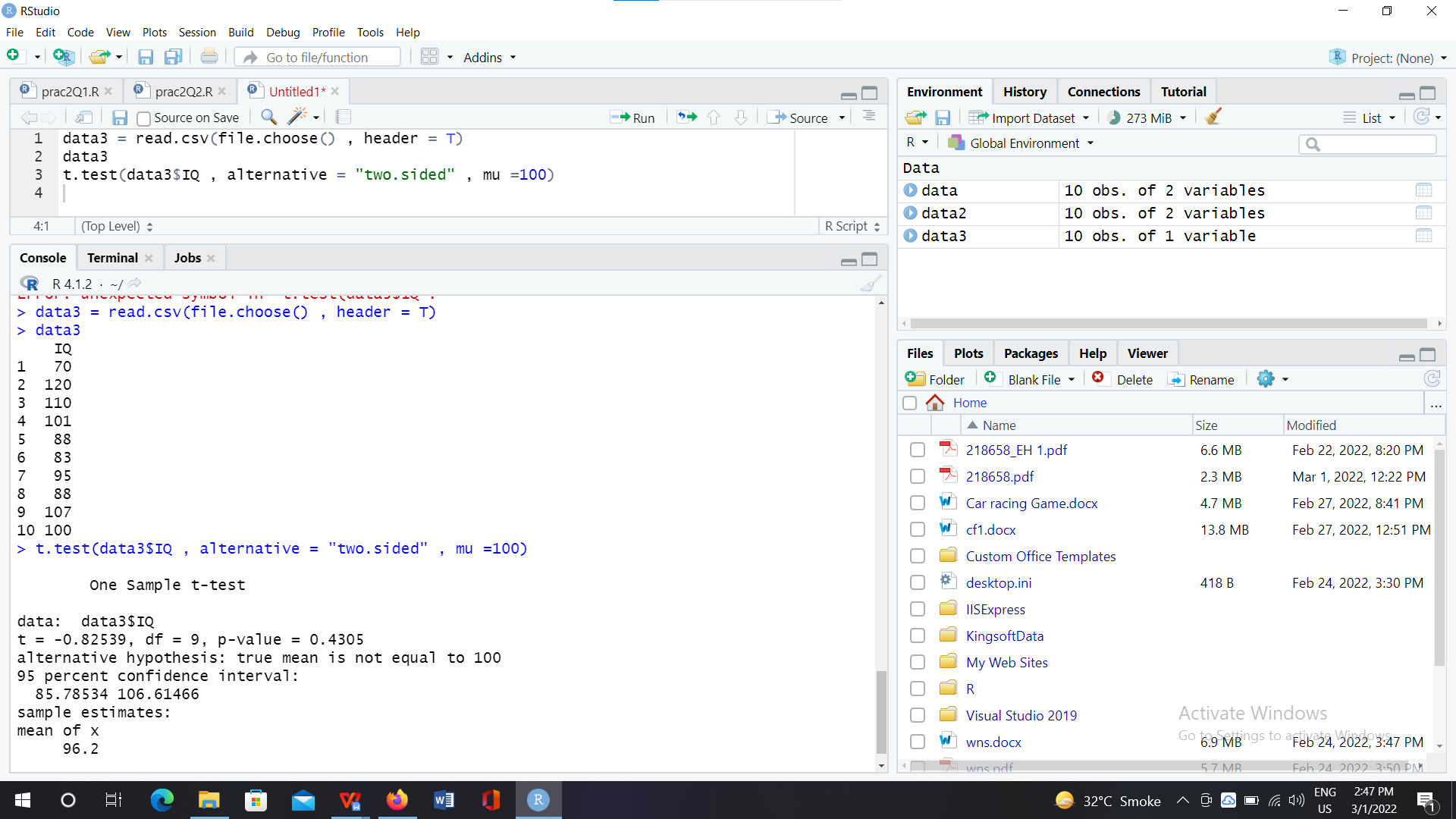
The p-value=1.894e-08 which is less than 0.05 therefore **we reject the hypothesis H0**.

1. **One Sample ‘t’ Test**

A random sample of 10 Students has the following IQ:

|  |
| --- |
| IQ |
| 70 |
| 120 |
| 110 |
| 101 |
| 88 |
| 83 |
| 95 |
| 88 |
| 107 |
| 100 |

**Output:**



**Conclusion:**

When alternative = two.sided,

The p-value=0.4305 which is less than 0.05 therefore **we reject the hypothesis H0**.Therefore population IQ

is not equal to 100.

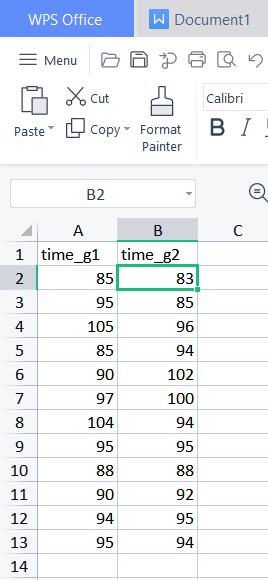
**Practical No:03**

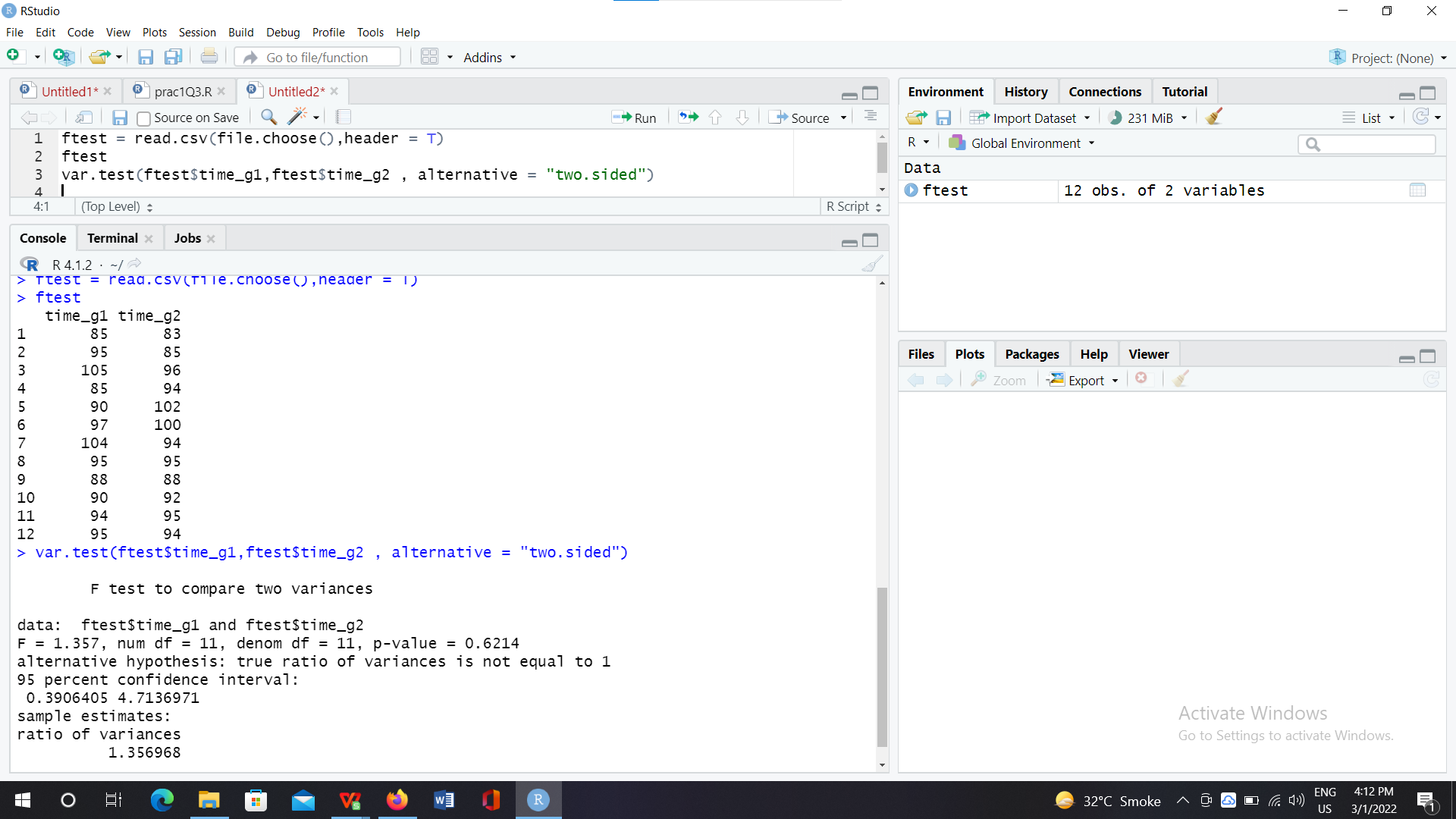
**Aim:Practical of Analysis of Variance**

**1. F-test(practical of analysis of data):**

Two samples are drawn from two normal populations. From the following data test whether the two

samples have the same variance.





**Conclusion:**

When alternative = two.sided ,

Here p = 0.4524, is greater than 0.05 so, **accept the hypothesis H0**.Therefore no significant difference in

variance of two group.

**2. One-way Anova:**

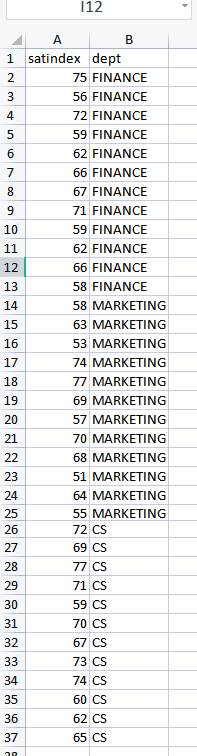
A company is accessing the difference in time to complete the task between three groups of employees.

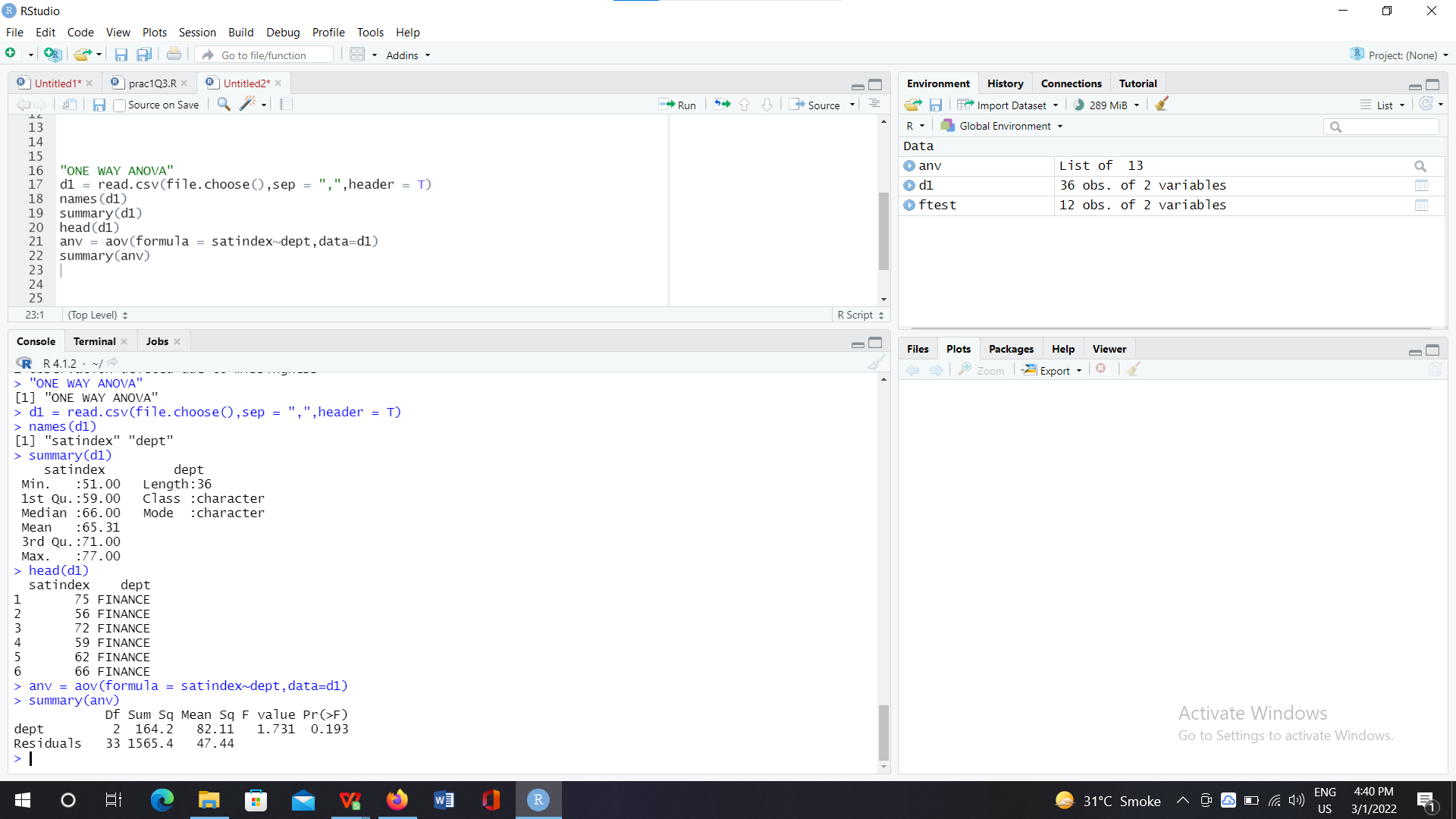
State the hypothesis and do the variance test for given dataset: (One way ANOVA)

Group-1: Fianance

Group-2: Marketing

Group-3: CS





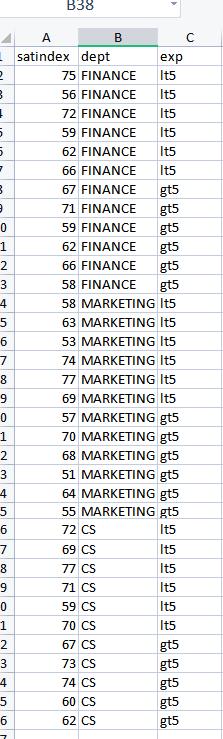
**Conclusion:**

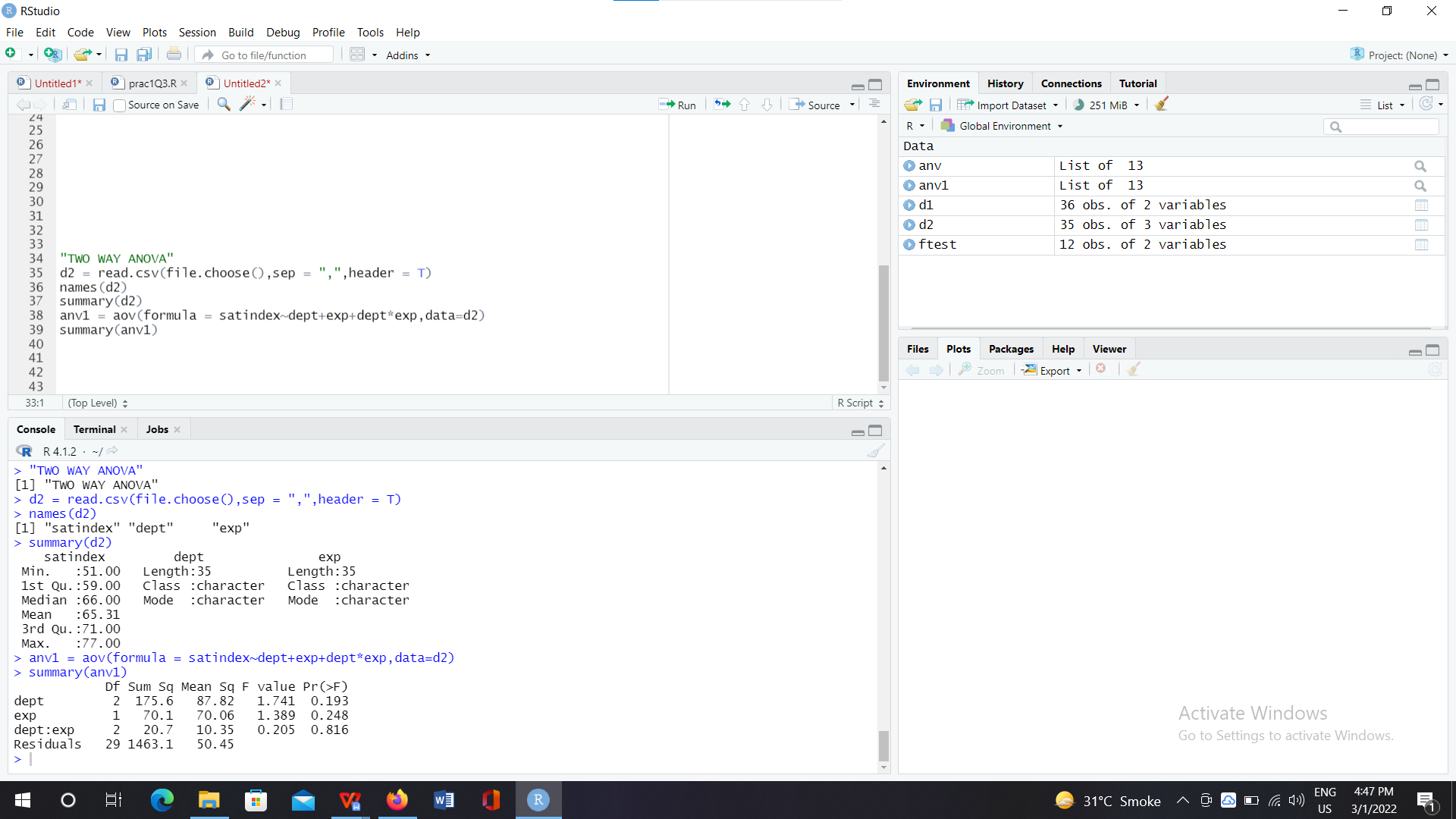
Here p = 0.193 which is greater than 0.05, **therefore accept the hypothesis H0**.

Therefore, the three group of employees require same amount of time to perform the task.

**3. Two-way Anova:**

Test whether the given factors have a significant impact on the dataset.





**Conclusion:**

The p-value for both the factors is greater than 0.05 **we accept the hypothesis H0**.

Therefore, both the factors viz. dept and exp have a statistically significant impact on the dataset.

**Practical No:04**

**Aim:Practical of Decision Tree**

"PassengerId:type should be integers

Survived:Survived or Not

Pclass:Class of Travel

Name:Name of Passenger

Sex:Gender

Age:Age of Passengers

SibSp:Number of Sibling/Spouse aboard

Parch:Number of Parent/Child aboard

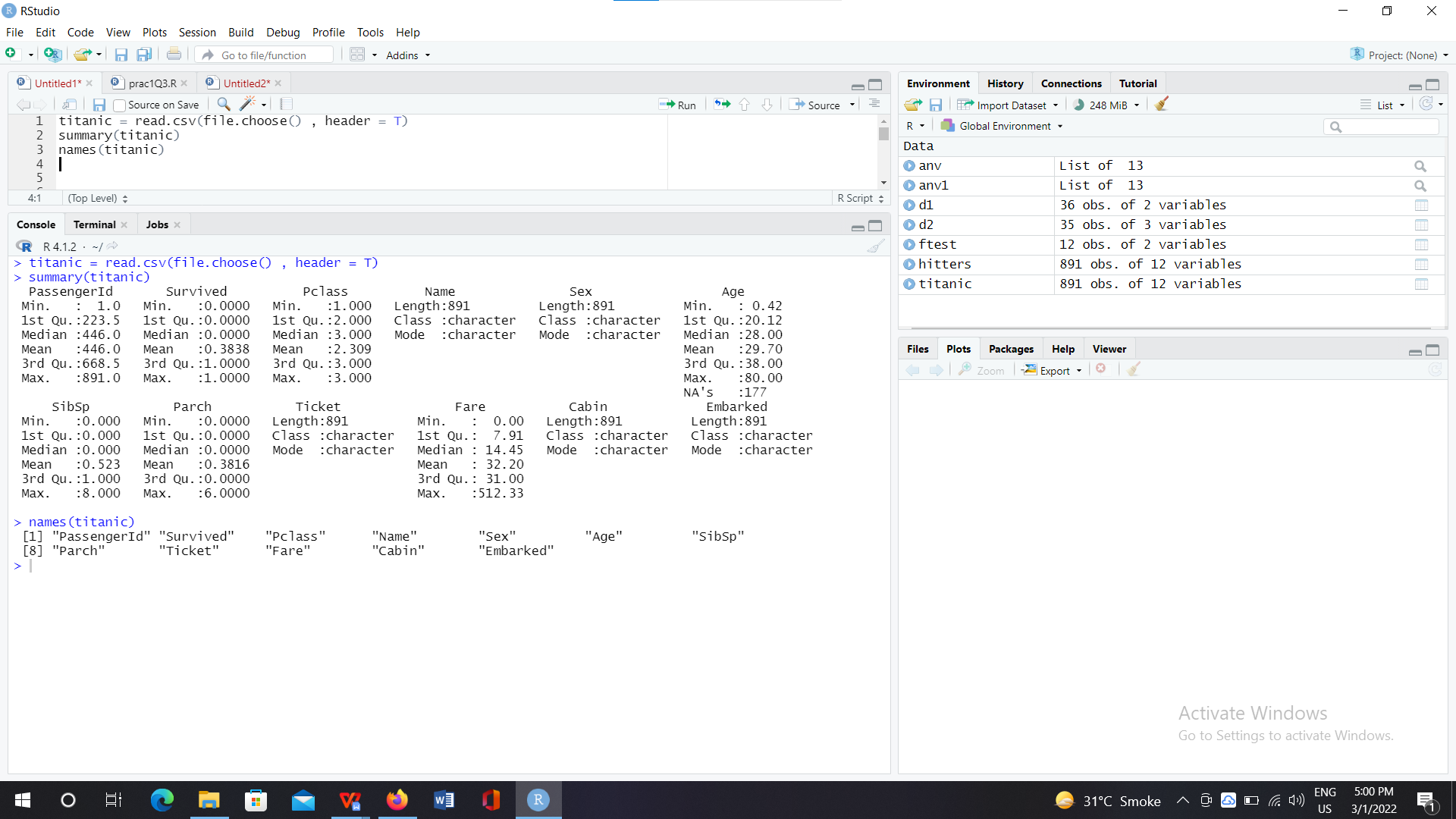
Ticket , Fare , Cabin

Embarked:The port in which a passenger has embarked. C - Cherbourg, S - Southampton, Q= Queenstown"

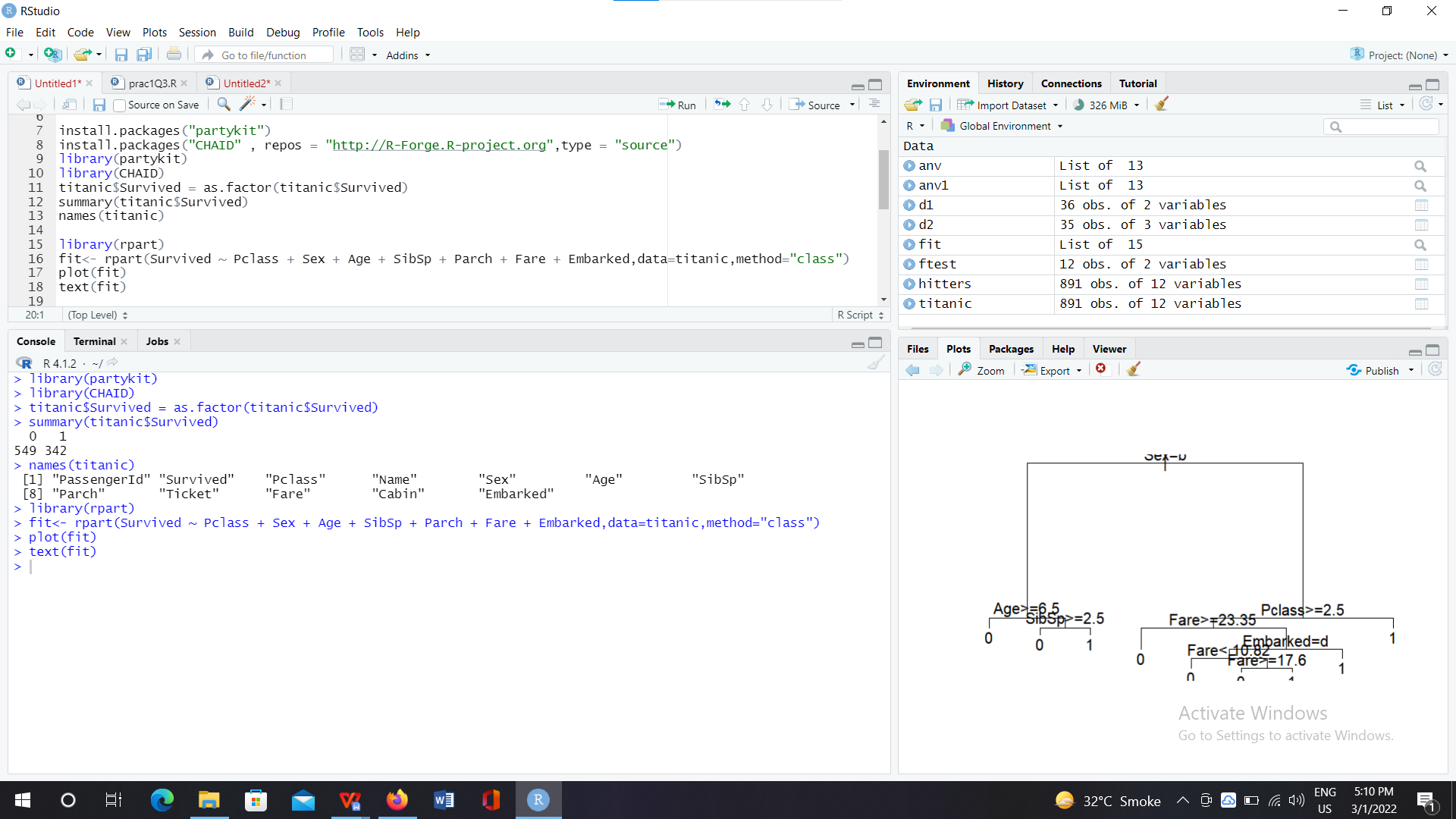
**1) Exploring the dataset. The purpose of this dataset is to predict which people are**

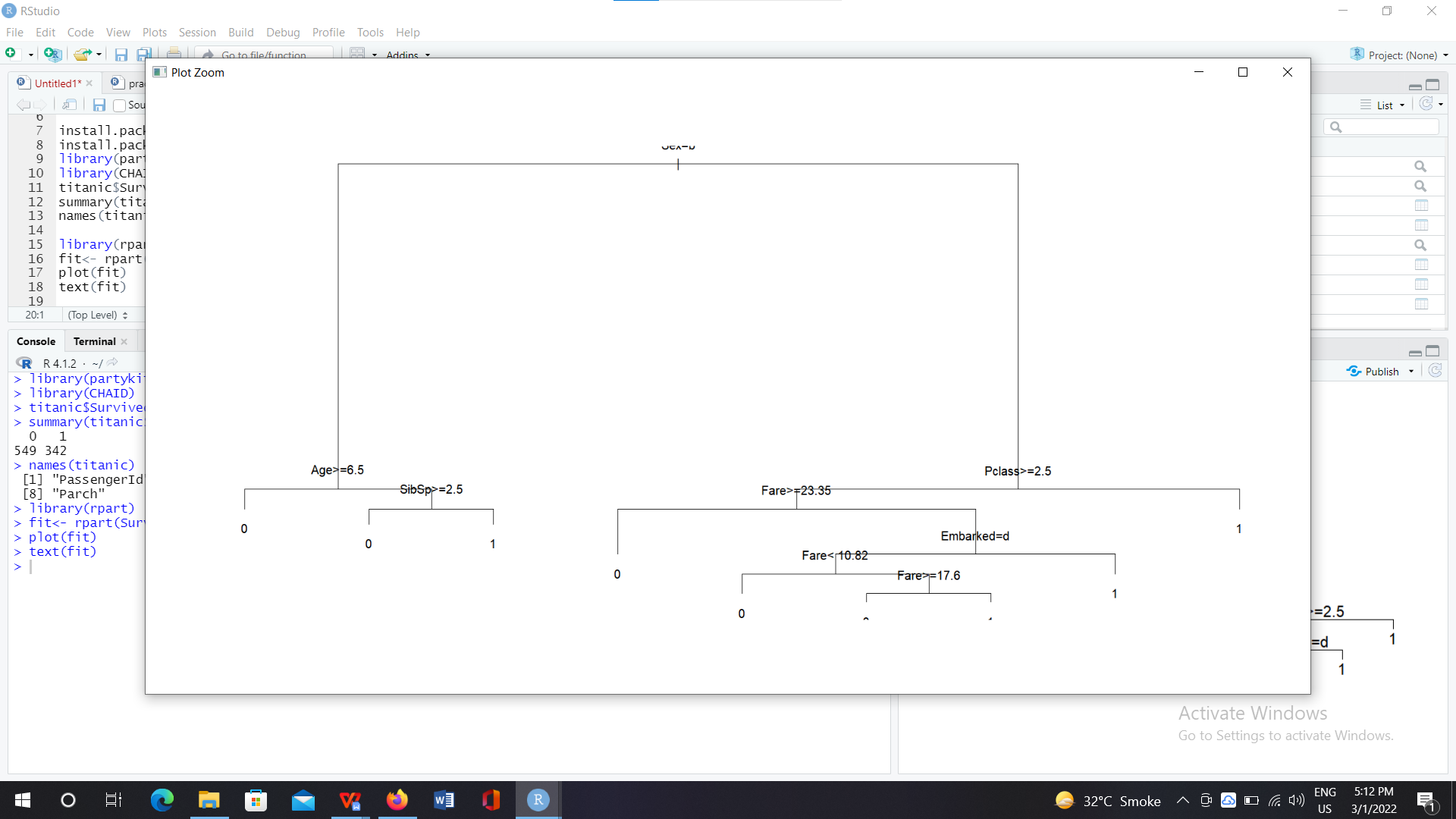
**more likely to survive after the collision with the iceberg. The dataset contains 12**

**variables and 891 observations.**

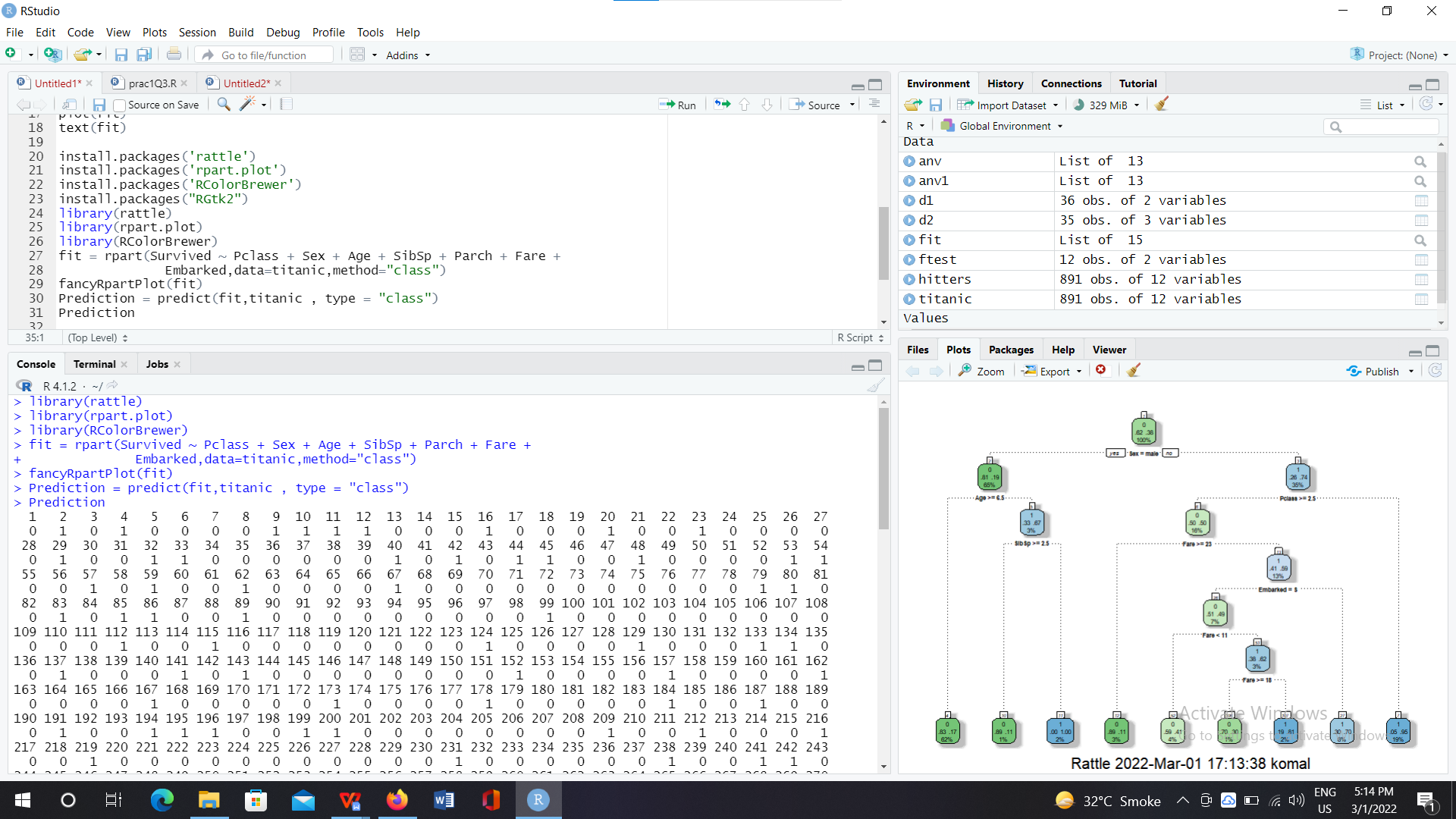


1. **From the following summary we can say that almost 549 passengers are dead and 342 have survived.**





1. **In the following prediction we can see the passenger Id with the value of 0 and 1. Where 0 means passenger with that passenger Id is dead and 1 means passenger with that passenger Id survived.**



**4) At the top, it is the overall probability of survival. It shows the proportion of passengers that survived the crash.**

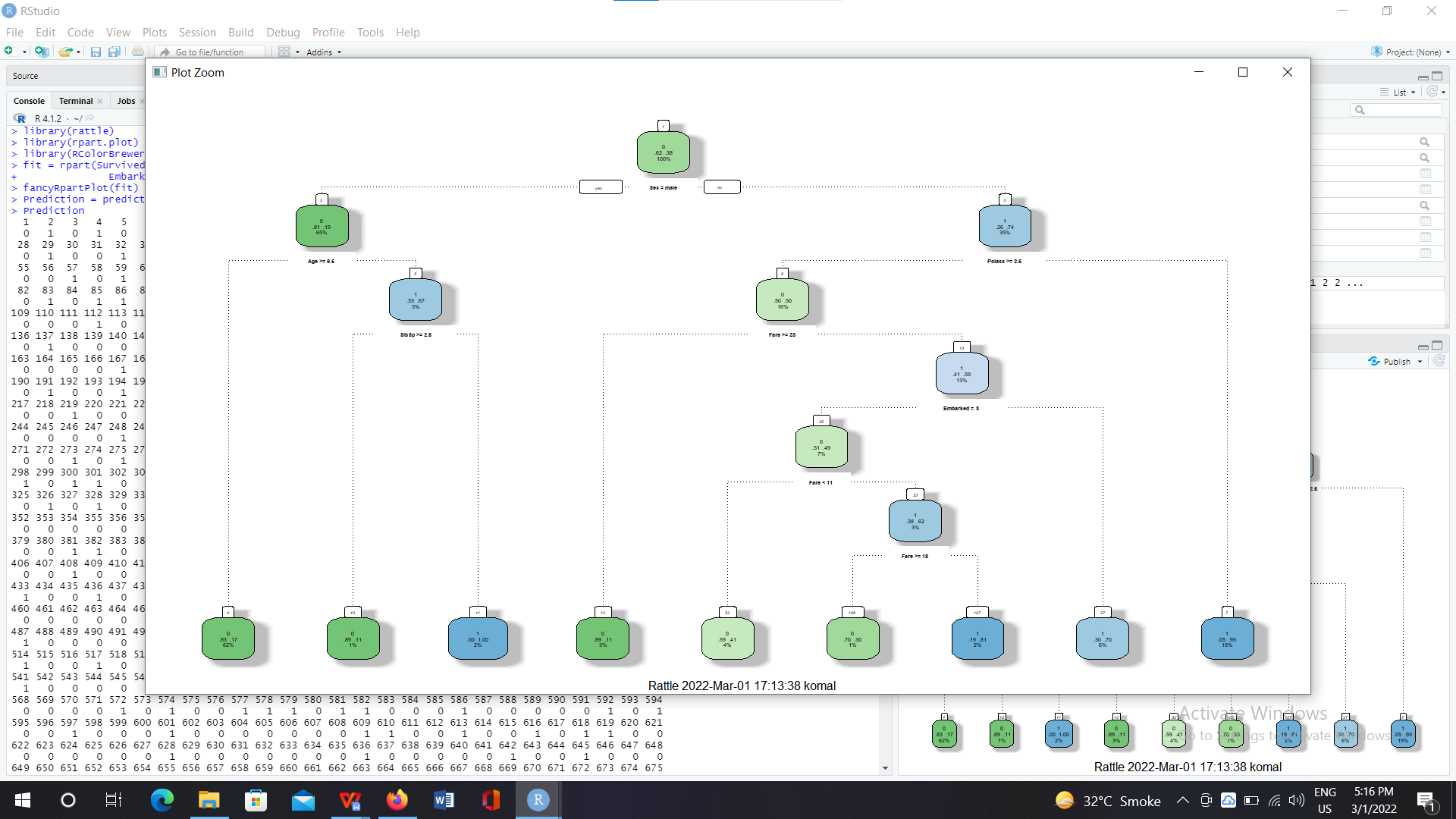
**Node-0: 38% of passengers survived and 62% of passengers are not.**

**Node-1: This node asks whether the gender of the passenger is male.**

**If yes, then you go down to the root's left child node. 65% are males with a survival probability of**

**19%. Node-2: In the second node, you ask if the male passenger is above 6.5 years old. If yes, then the chance of survival is 17 percent.**

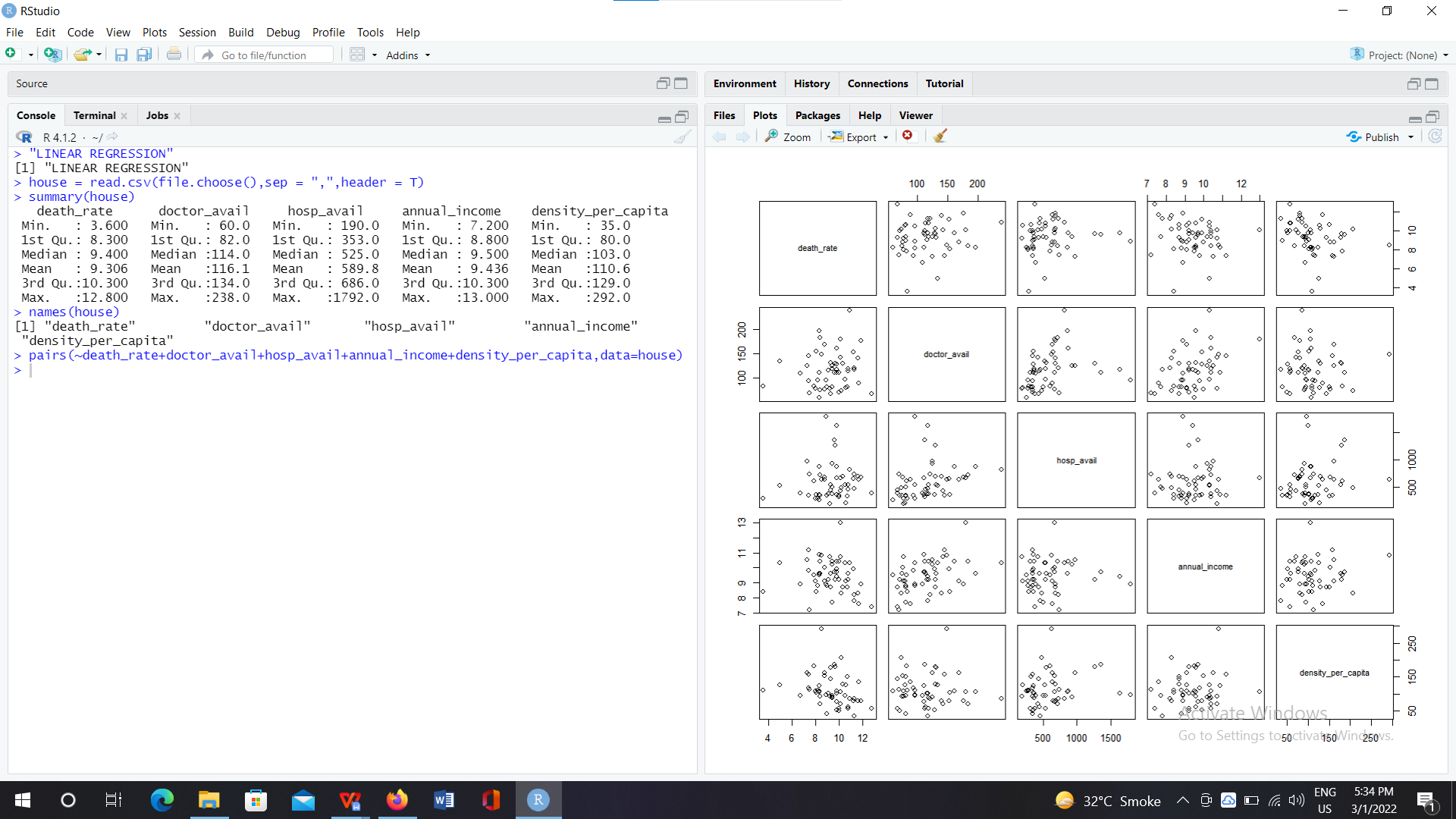
**This way we can calculate the likelihood of survival.**



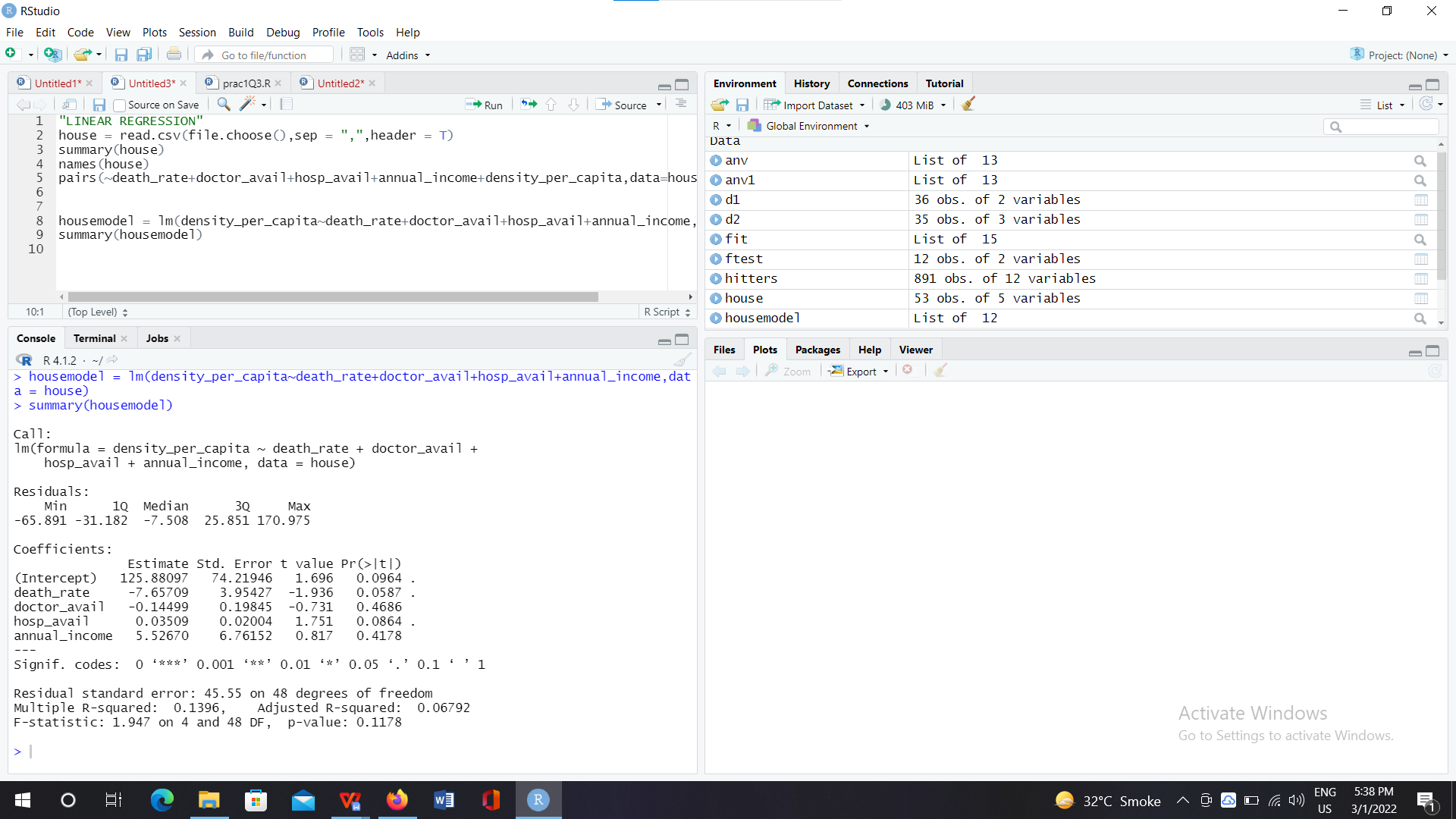
**Practical No:05**

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

**1.Scatter plot:**



1. **Fit a model:**

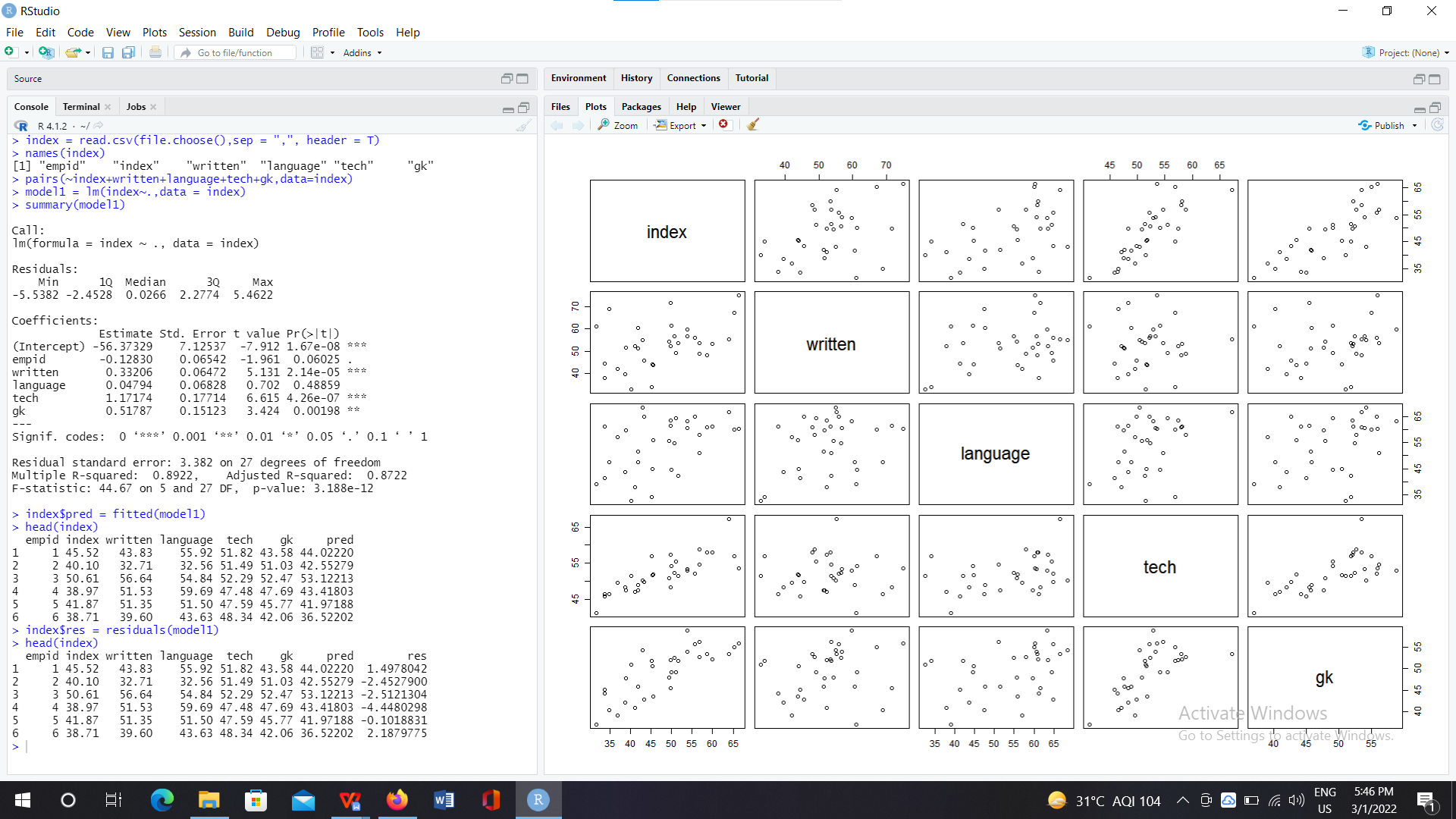


**Conclusion:**

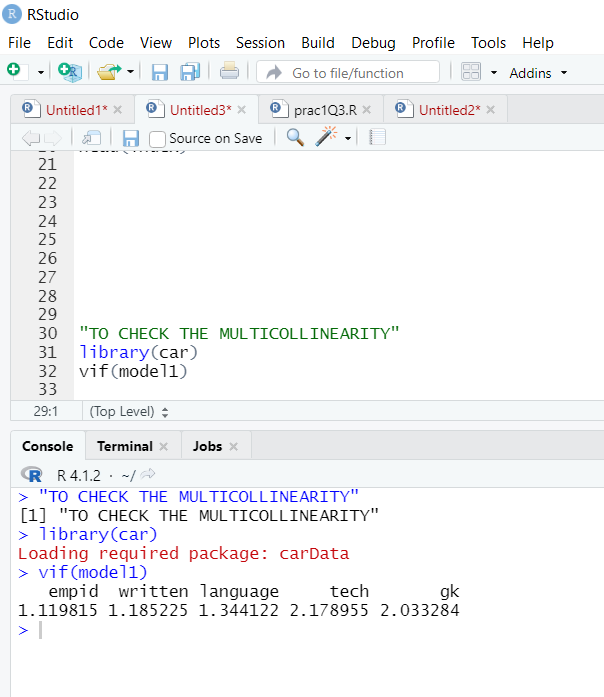
As value of multiple r(square) is 0.8922

So 89% of variation in index is explained by the model and 11% is not explained by the model

1. **check for global testing** :



1. **check the multicollinearity:**

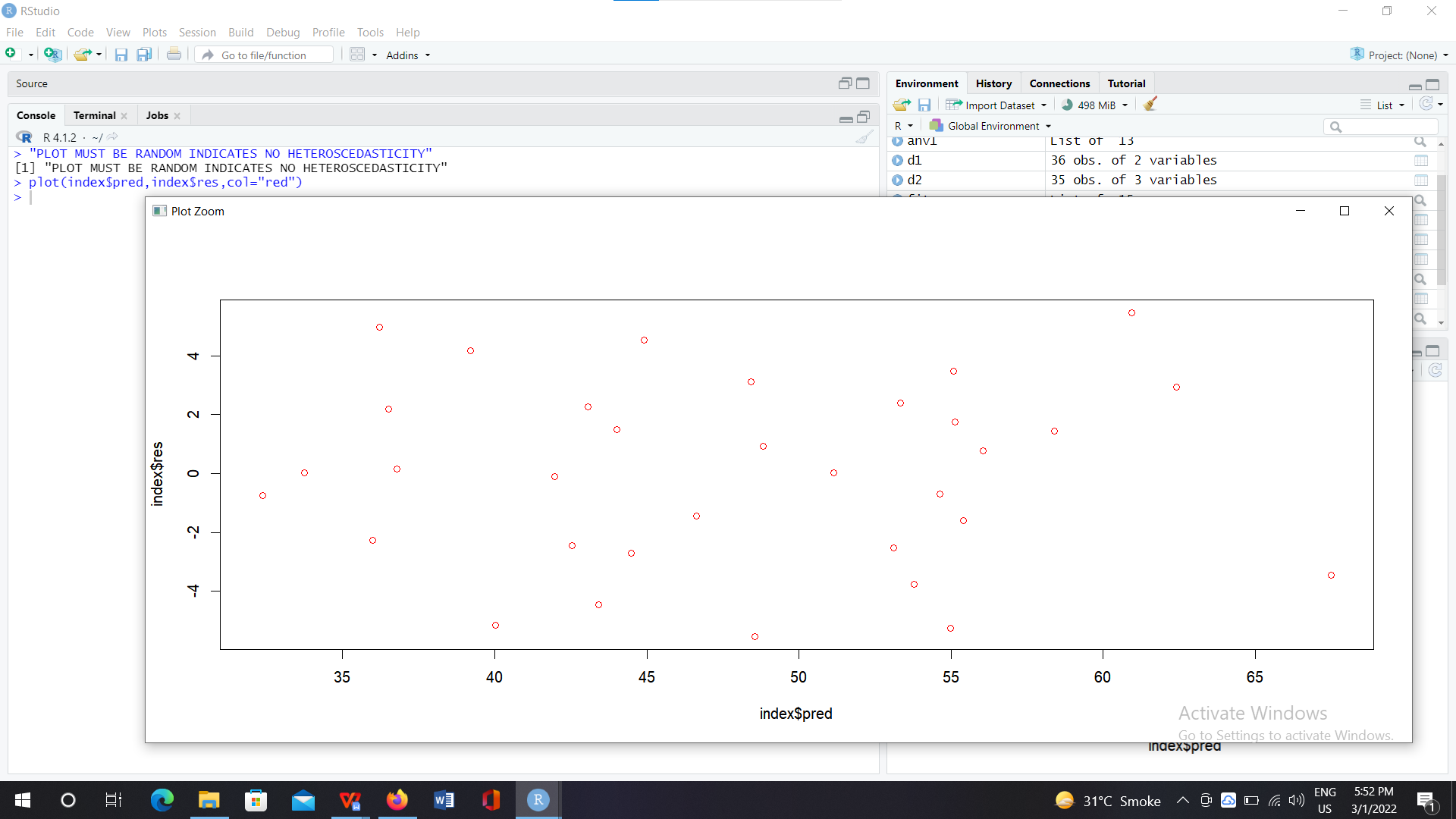


**Conclusion:**

As all VIF are less than 5

Multicollinearity is not present.

1. **check heteroscedasticity**



**Conclusion:**

since errors are generated randomly. There is no heteroscedasticity.

1. **Shapiro test to check normality of errors.**



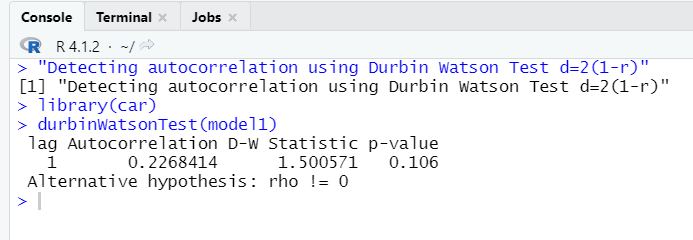
**Conclusion:**

Ho: There is homoscedasticity

H1: there is no constant variance

As P-value is greater than 0.05 we accept Ho.

1. **Detecting autocorrelation using Durbin Watson Test d=2(1-r):**



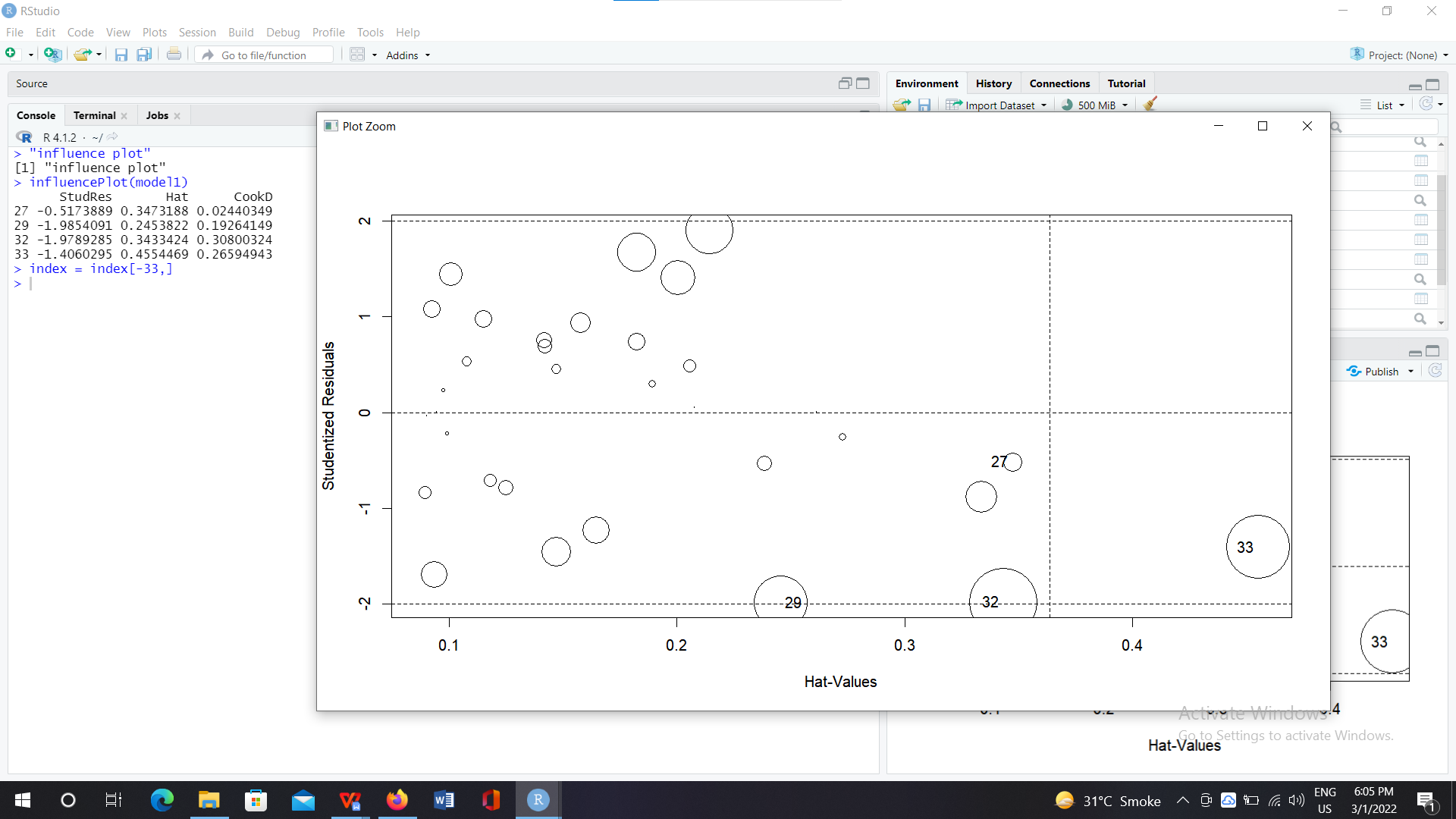
**Conclusion:**

Ho: Auto correlation is not present among errors

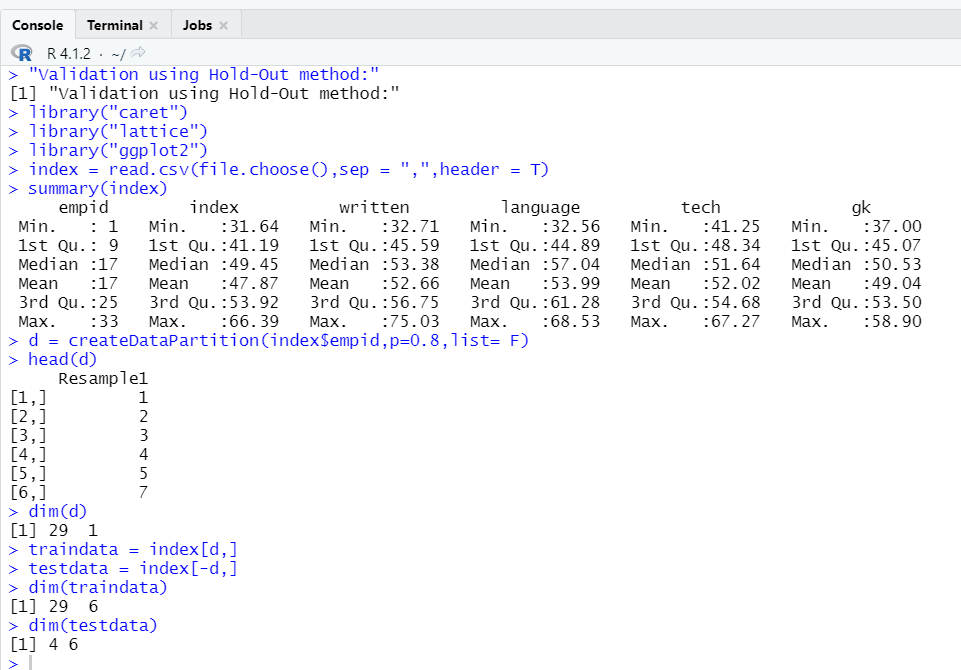
H1: not Ho

As P-value is greater than 0.05 we accept Ho.

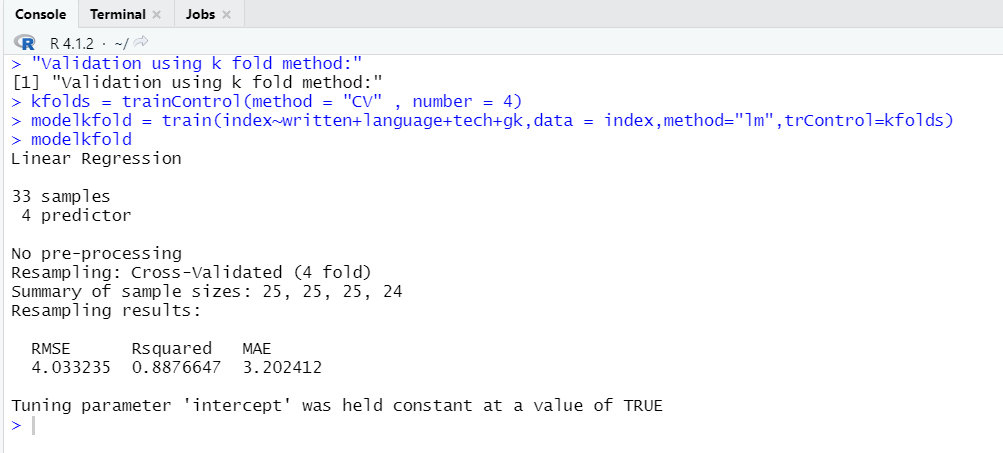
1. **influence plot**



1. **Validation using Hold-Out method:**

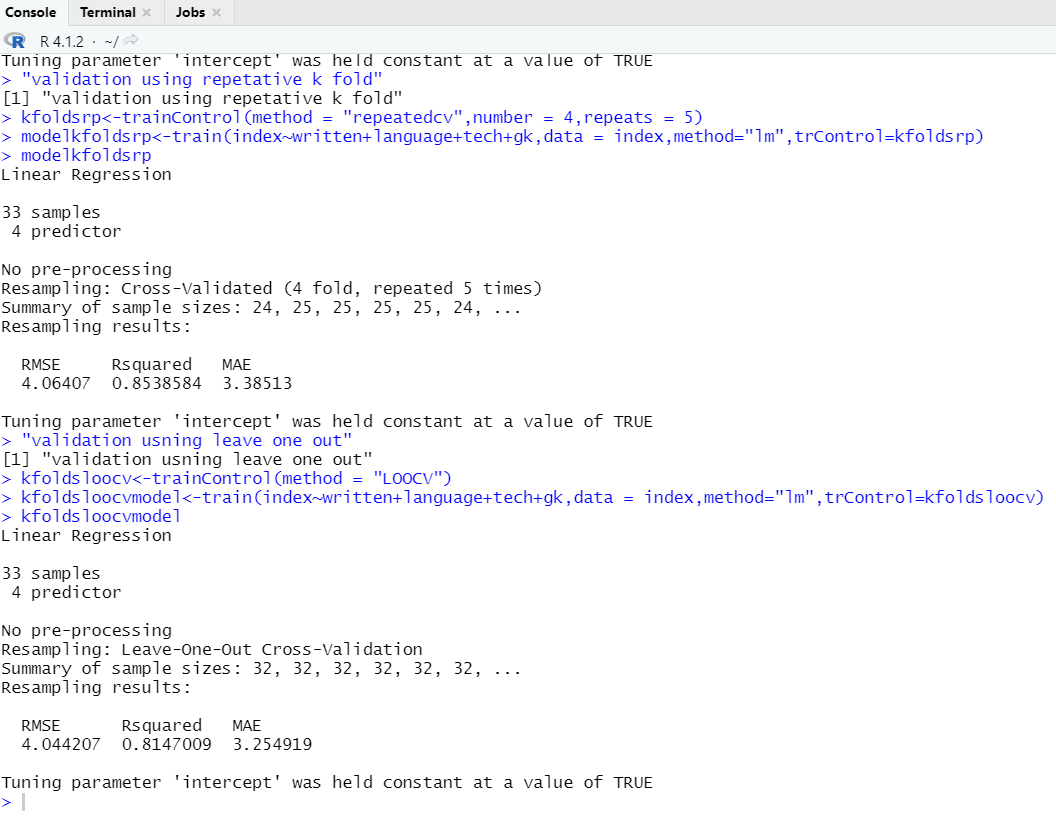


1. **Validation using k fold method:**



**Conclusion:**

As the value of RMSE is sufficiently large the model is stable.



1. **Model selection forward method.**

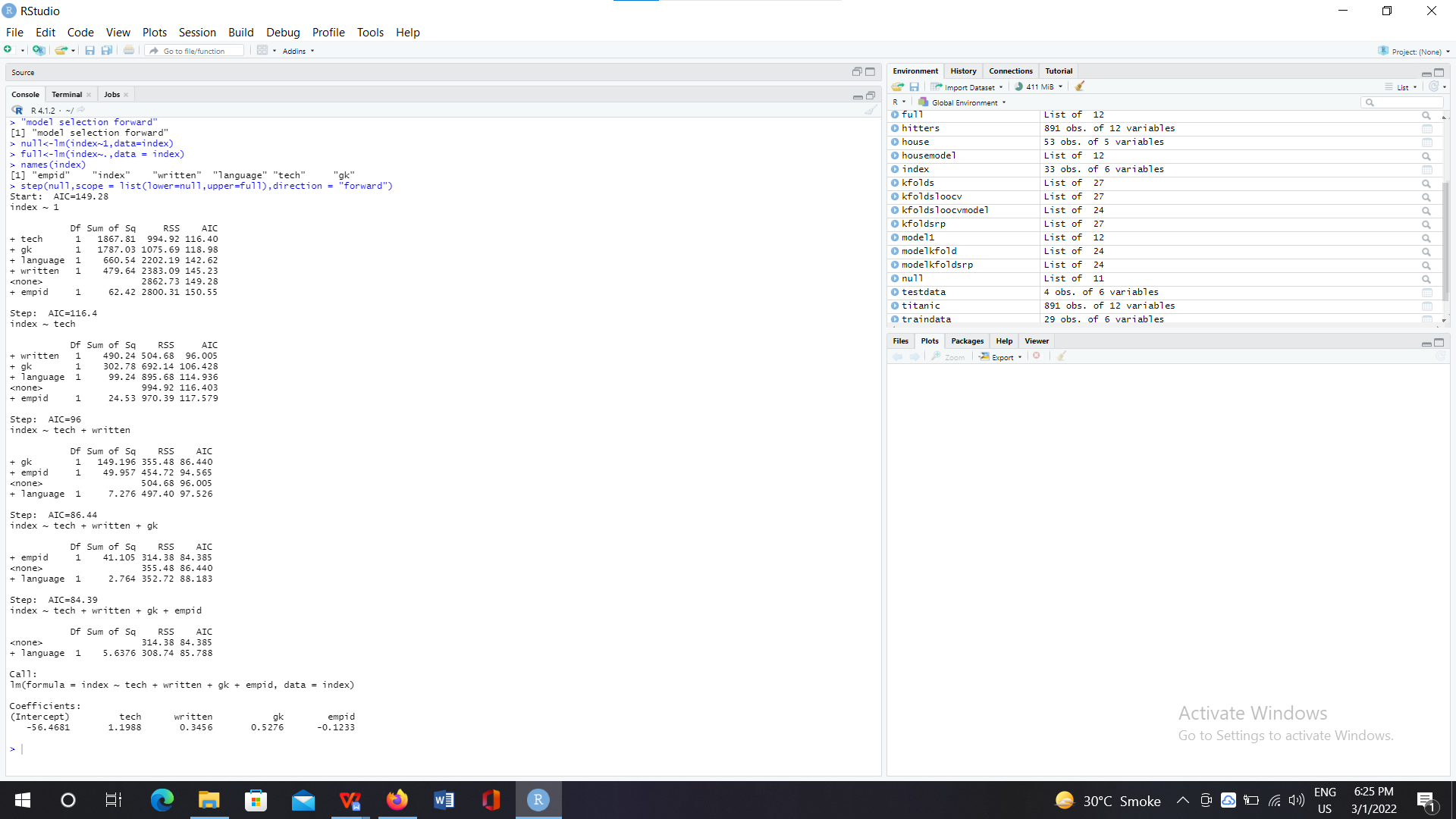
"model selection forward"

null<-lm(index~1,data=index)

full<-lm(index~.,data = index)

names(index)

step(null,scope = list(lower=null,upper=full),direction = "forward")



1. **Model selection backward method:**

