**ASSIGNMENT 2**

Q1. Explain Type-1 error and Type-2 error.

**Ans: Type 1 error is rejecting the null hypothesis when it is true. It is also called as level of significance. It is referred as α. Type 2 error is accepting the null when it is false.**

Q2. Import the data named “cars” already present in R-datasets. And perform the following:

1. Basic EDA

**library(readr)**

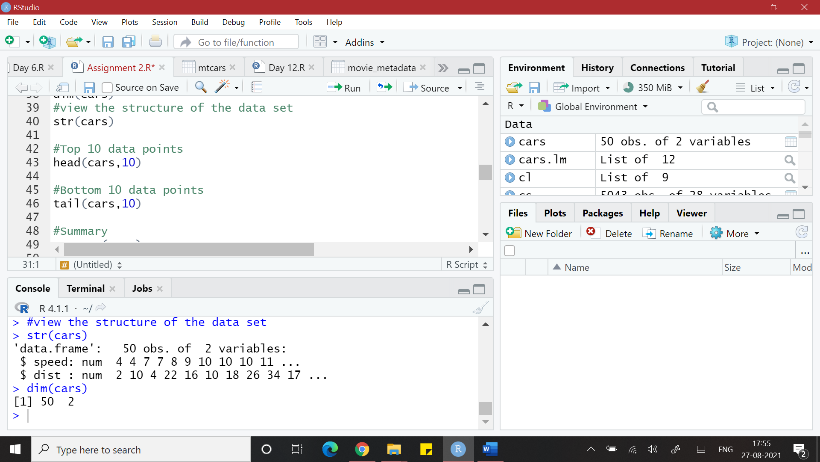
**data(cars)**

**Dim(cars)**

**#** **50 2**

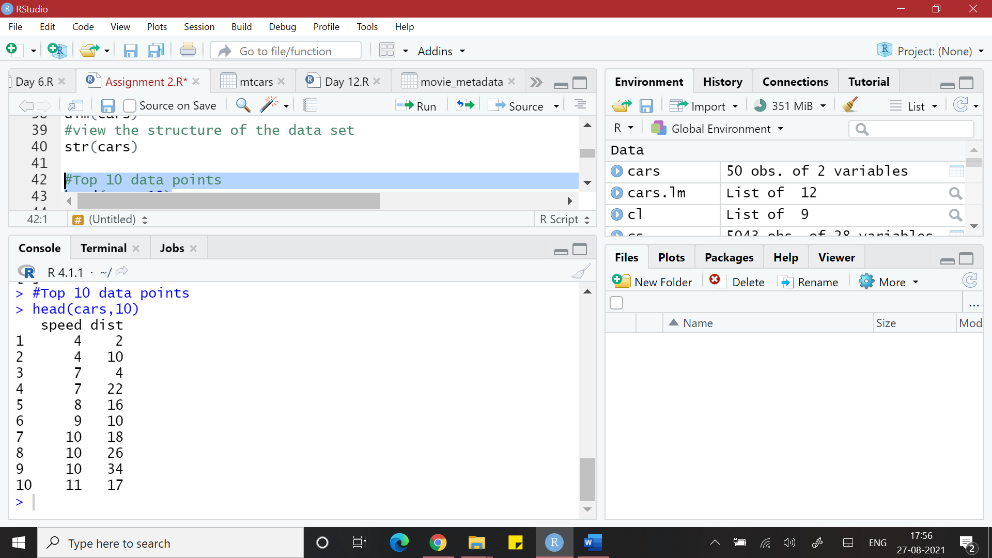
**#view the structure of the data set**

**str(cars)**



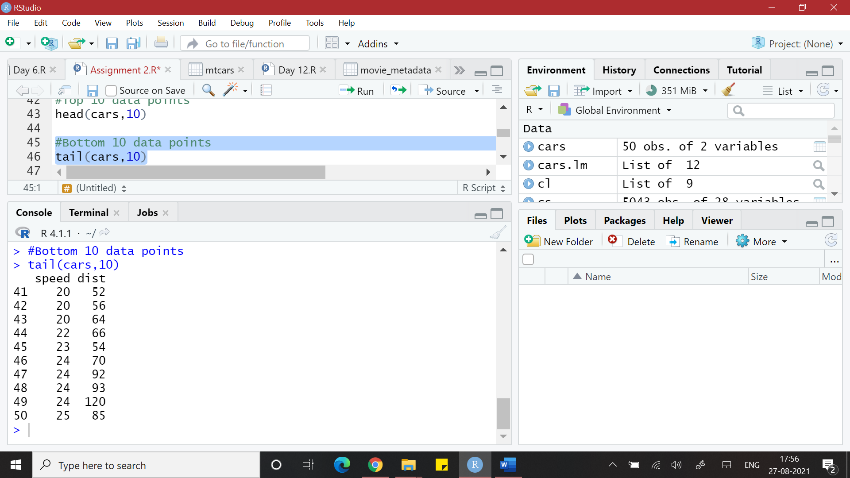
**#Top 10 data points**

**head(cars,10)**



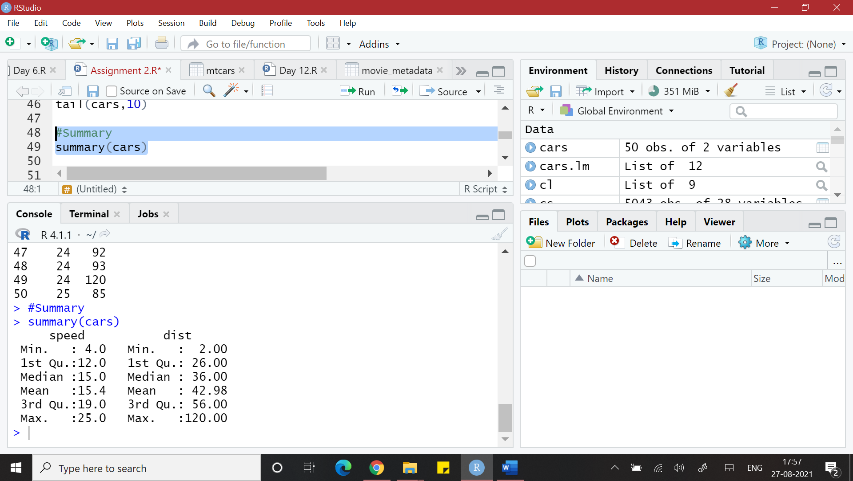
**#Bottom 10 data points**

**tail(mtcars,10)**

****

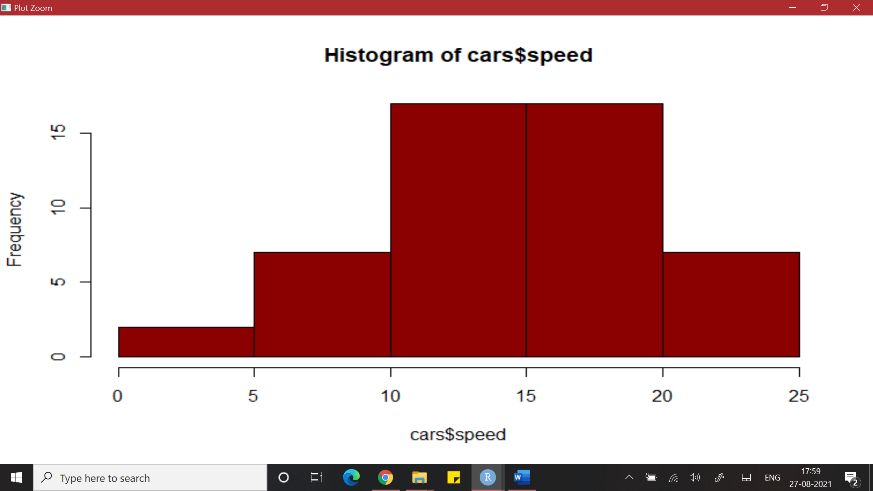
**#Summary**

**summary(mtcars)**

****

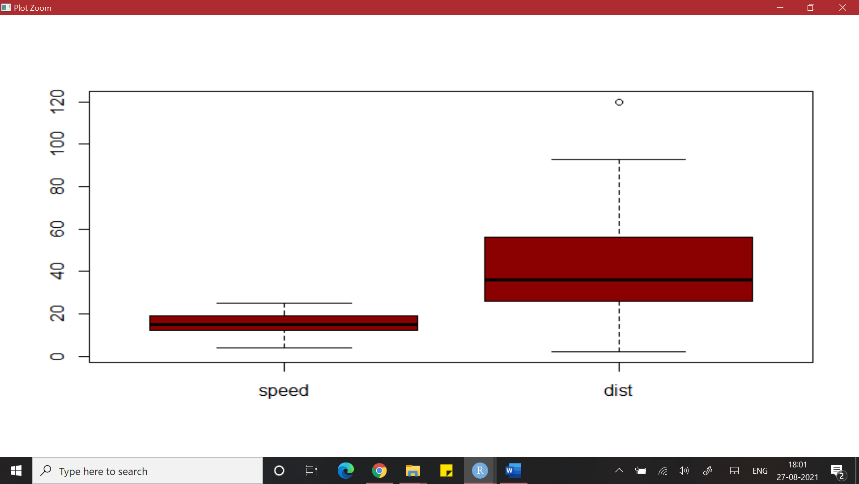
**#Histogram**

**hist(cars$speed ,col = "dark red")**

****

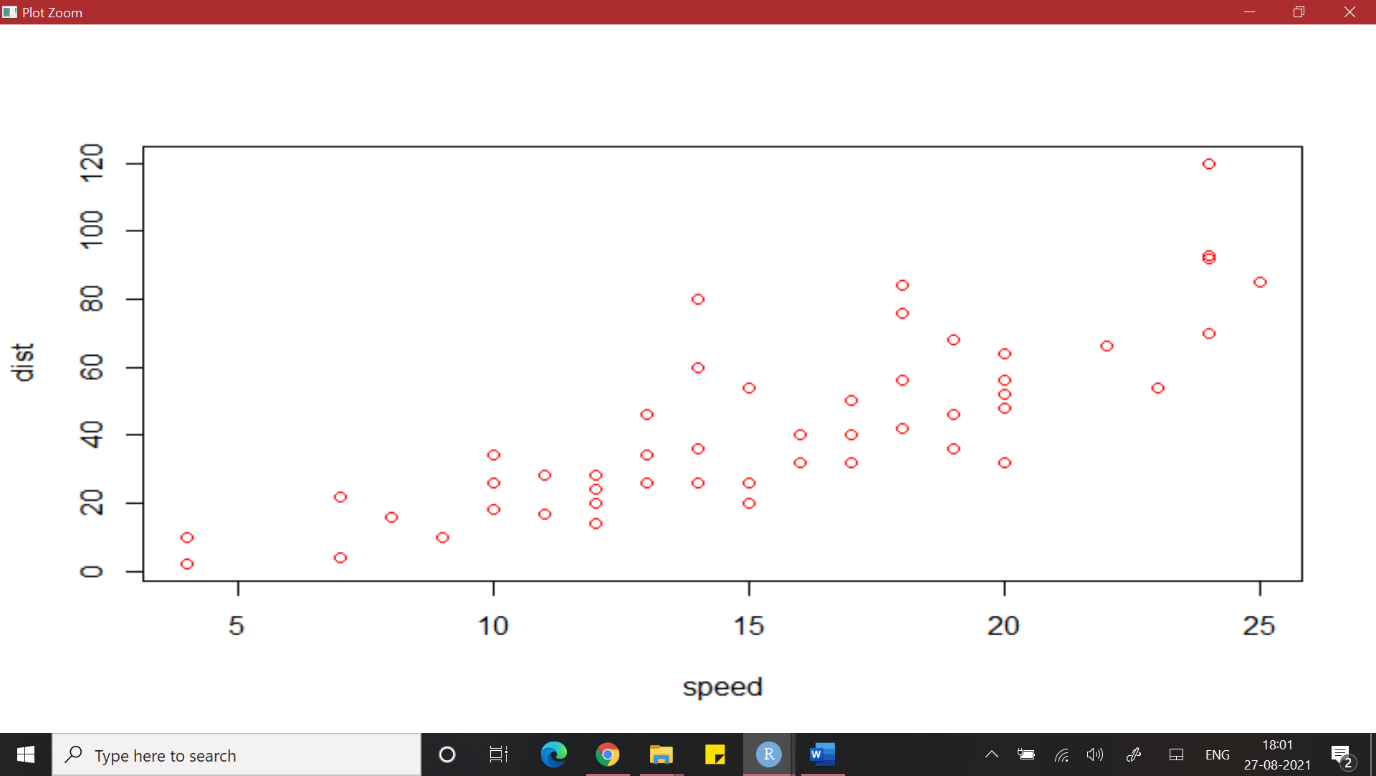
**#Box plot**

**boxplot(cars, col = "dark red")**

****

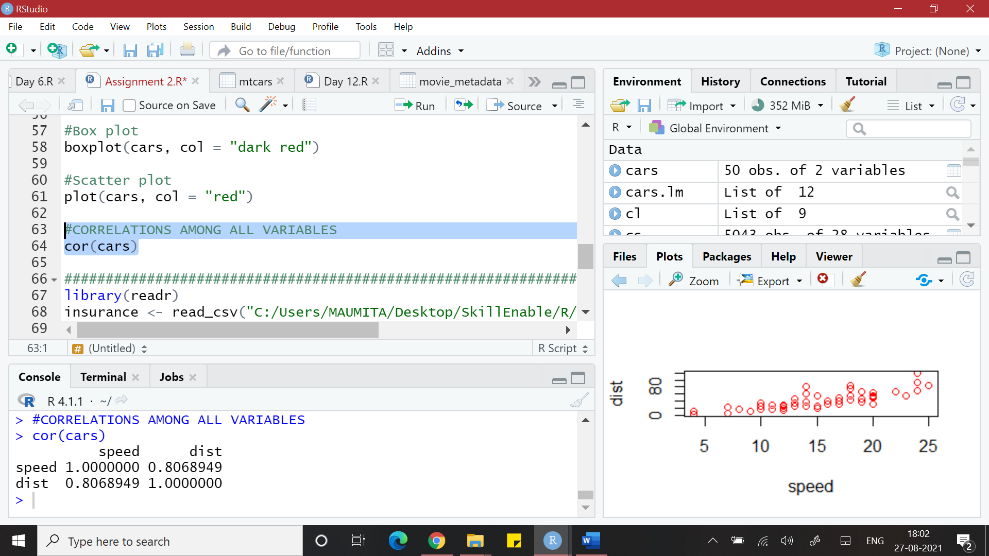
**#Scatter plot**

**plot(cars, col = "red")**

****

**#CORRELATIONS AMONG ALL VARIABLES**

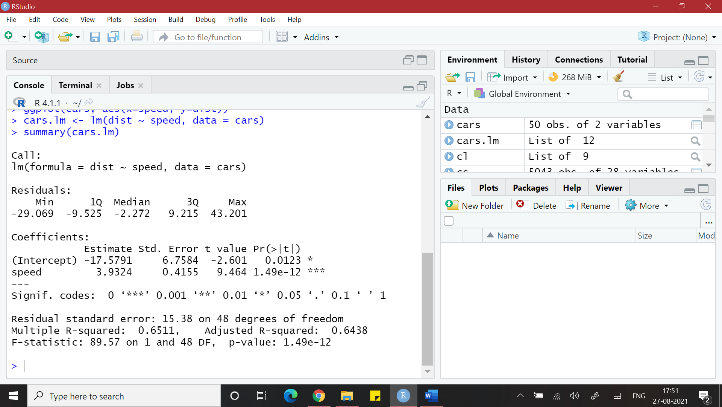
**cor(cars)**



1. Predict the value of distance covered by the car on the basis of speed.

**cars.lm <- lm(dist ~ speed, data = cars)**

**summary(cars.lm)**

****

1. Name the ML algorithm used.

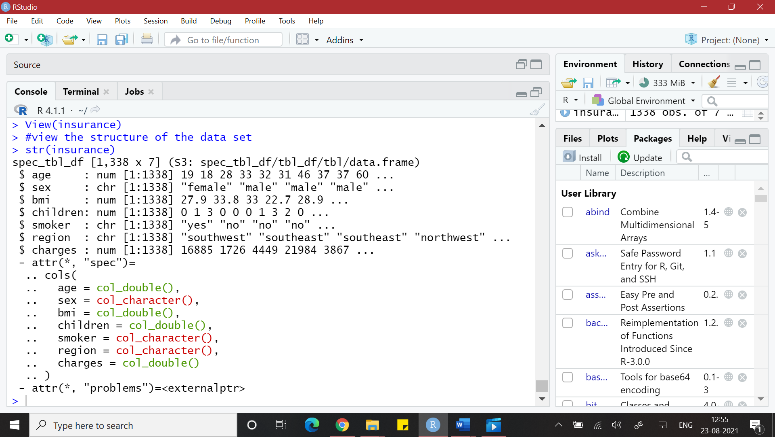
**LINEAR REGRESSION**

Q3. Import the file insurance.csv and perform the following:

1. Explore the data with your statistics concepts and visualization

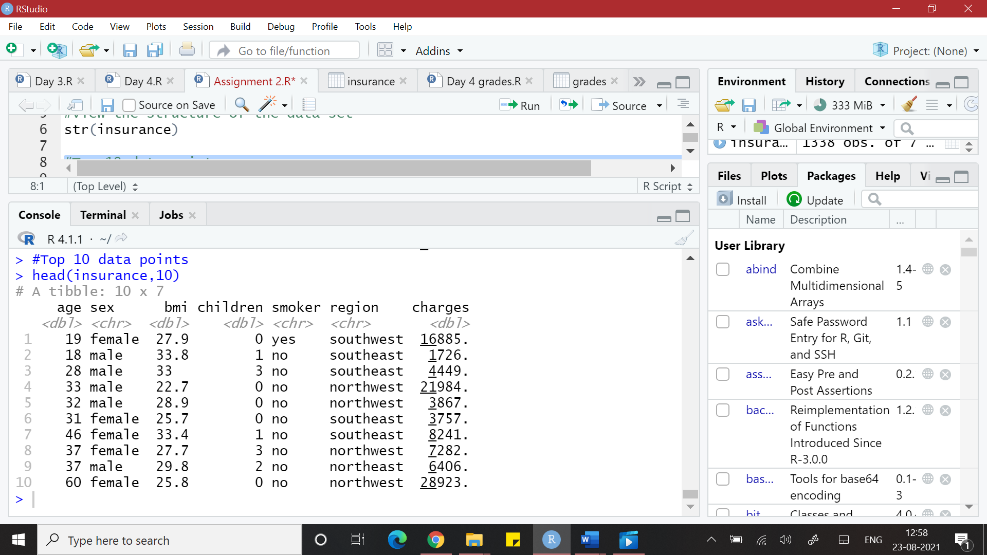
**#view the structure of the data set**

**str(insurance)**



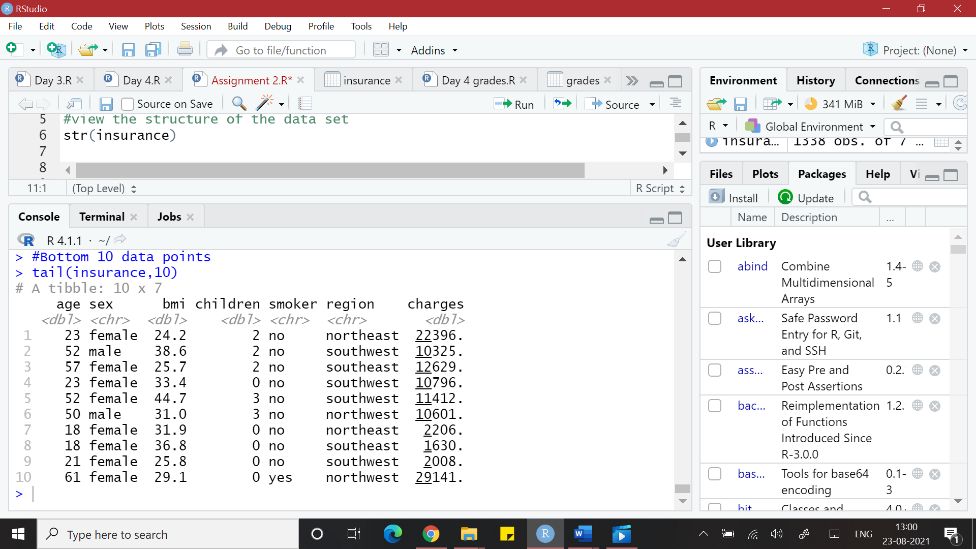
**#Top 10 data points**

**head(insurance,10)**



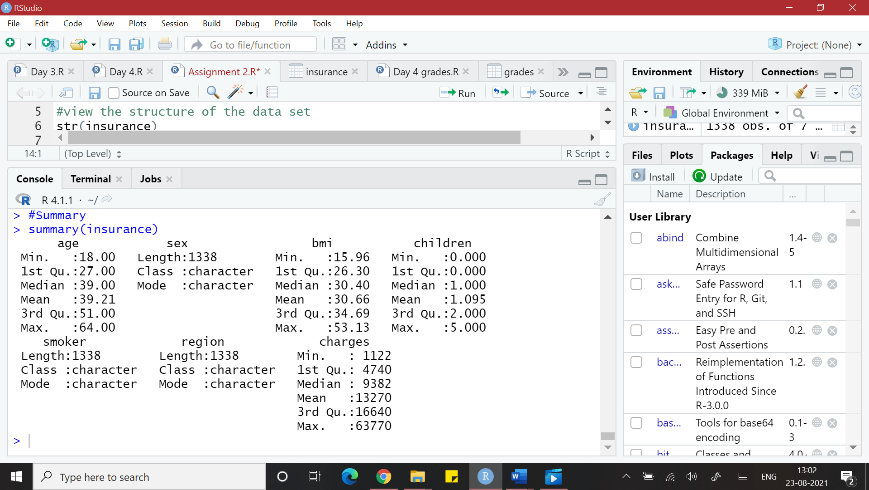
**#Bottom 10 data points**

**tail(insurance,10)**



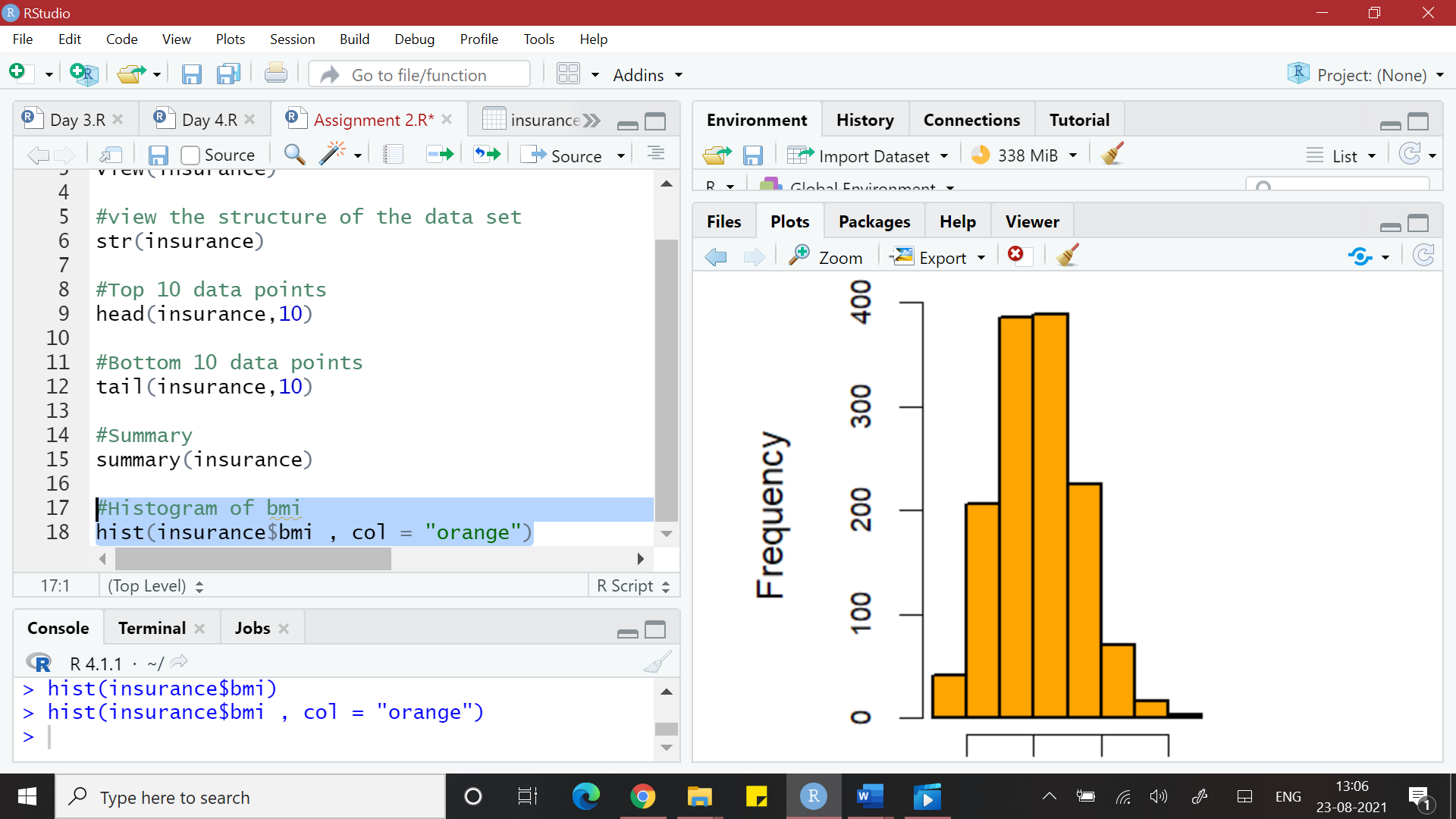
**#Summary**

**summary(insurance)**



**#Histogram of bmi**

**hist(insurance$bmi , col = "orange")**

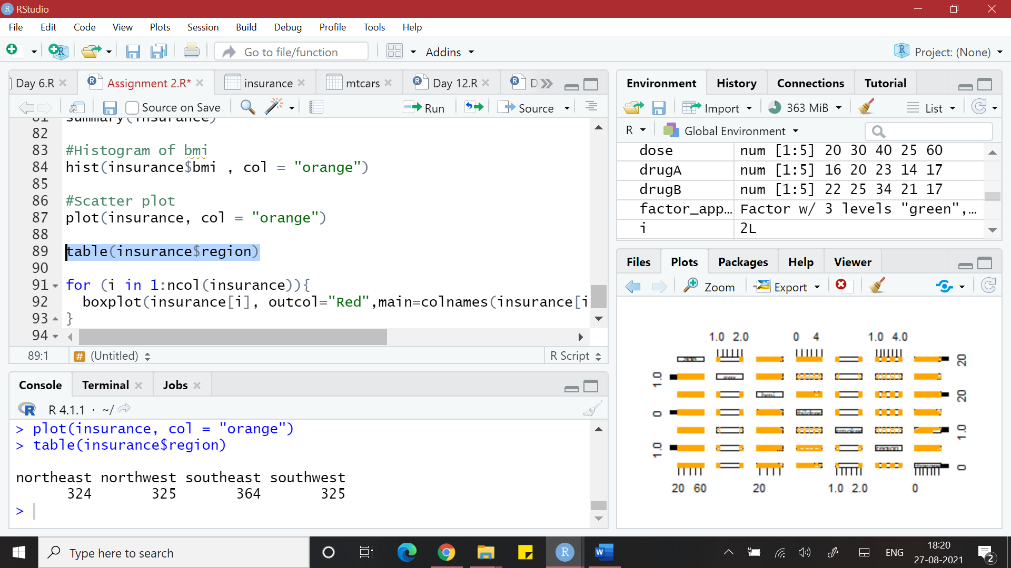


**#Scatter plot**

**plot(insurance, col = "orange")**



**table(insurance$region)**

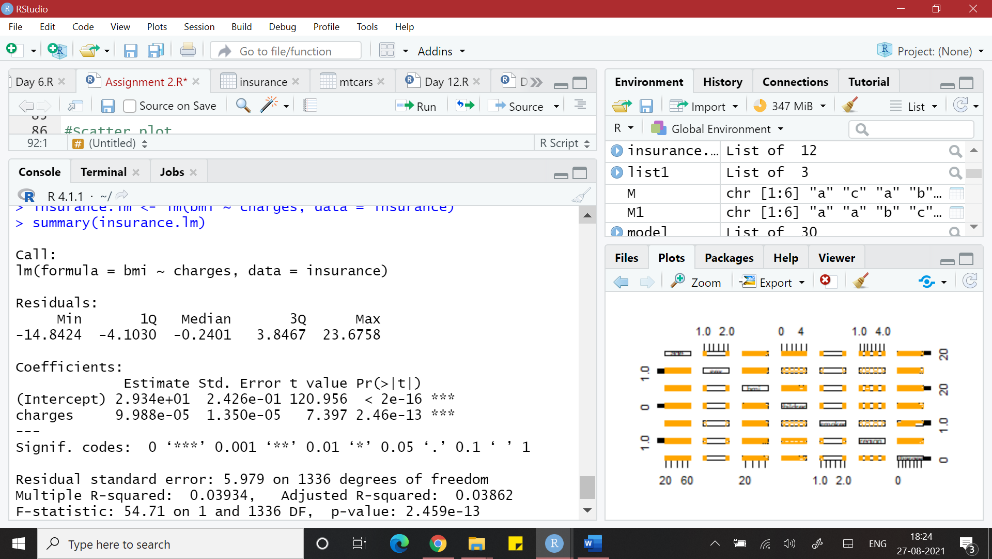


b. Predict the value of the cost incurred by the company given by the parameter “charges”

which depends on the other parameters.

**insurance.lm <- lm(bmi ~ charges, data = insurance)**

**summary(insurance.lm)**



c.Name the ML algorithm used.

**LINEAR REGRESSION**

1. Explain the results obtained

OR

Q4. Import the file Placement\_Data\_Full\_Class.csv and perform the following:

1. Explore the data with your statistics concepts and visualization
2. Predict “ salary” earned by students which depends on the other parameters.
3. Name the ML algorithm used.
4. Explain the results obtained