**Insurance Factors Identification**

**Question 1.** The committee is interested to know each field of the data collected through descriptive analysis to gain basic insights into the data set and to prepare for further analysis.

**Solution 1.**

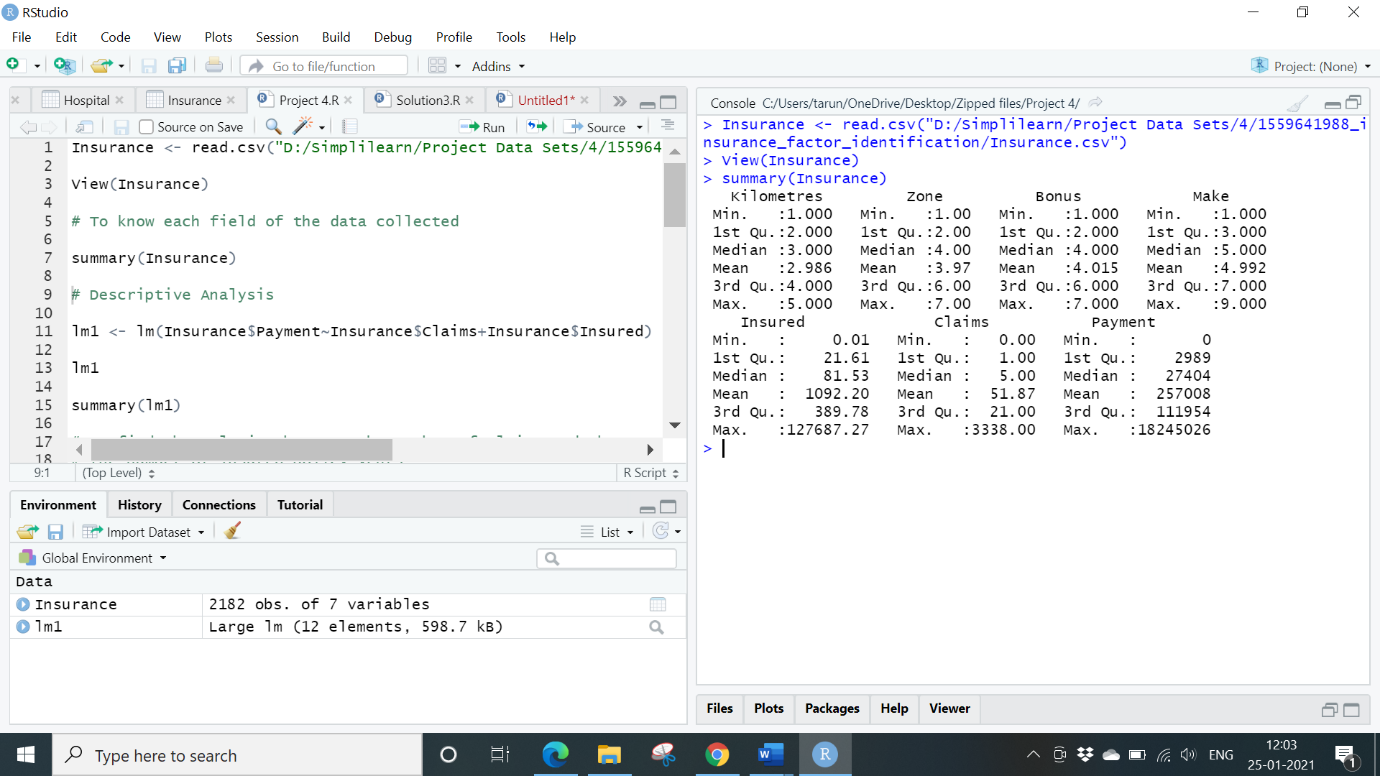
**Code:**

Insurance<read.csv("D:/Simplilearn/ProjectDataSets/4/1559641988\_insurance\_factor\_identification/Insurance.csv")

View (Insurance)

# To know each field of the data collected

summary (Insurance)



**Question 2.** The total value of payment by an insurance company is an important factor to be monitored. So, the committee has decided to find whether this payment is related to the number of claims and the number of insured policy years. They also want to visualize the results for better understanding.

**Solution 2.**

**Code.**

# Descriptive Analysis

lm1 <- lm (Insurance$Payment~Insurance$Claims+Insurance$Insured)

lm1

summary(lm1)

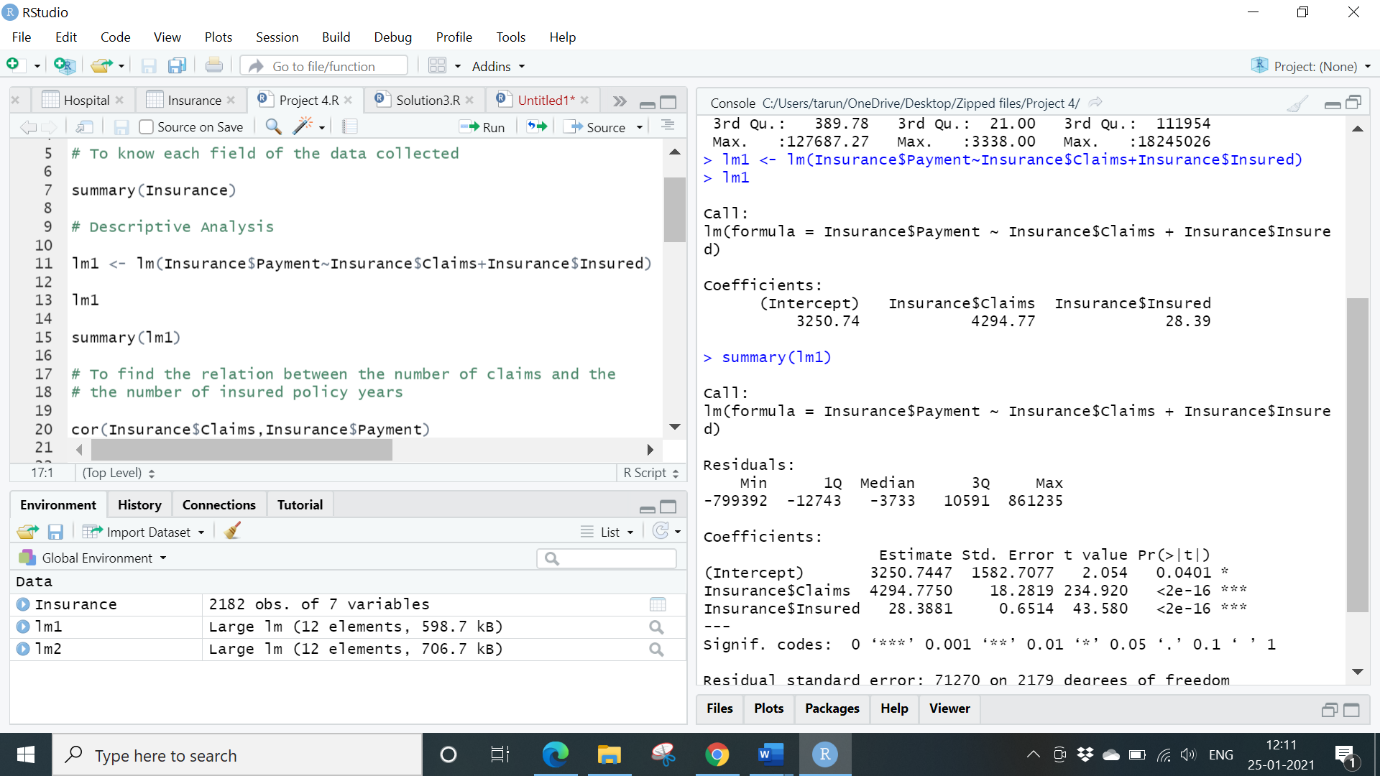
# To find the relation between the number of claims and the

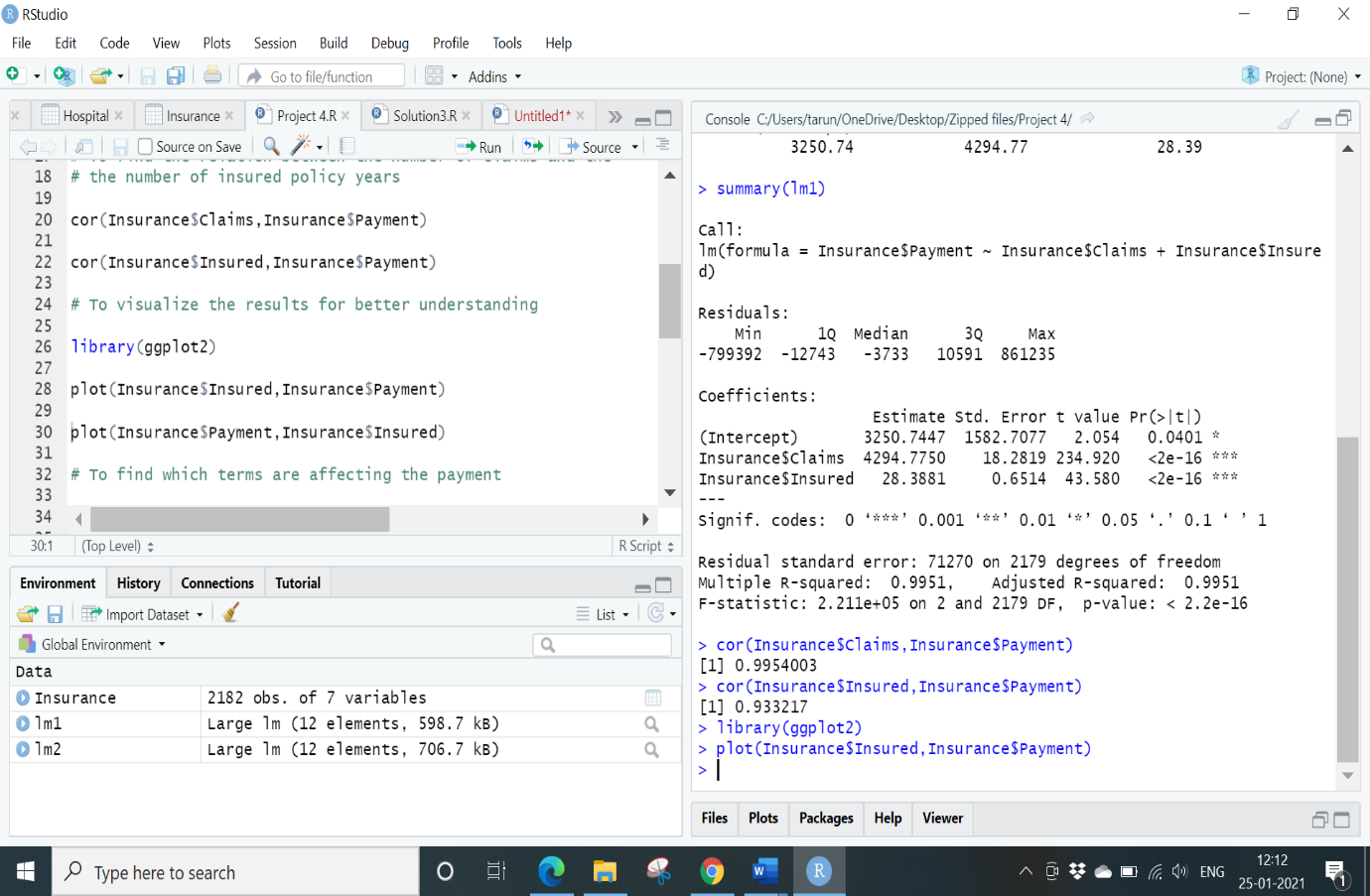
# the number of insured policy years

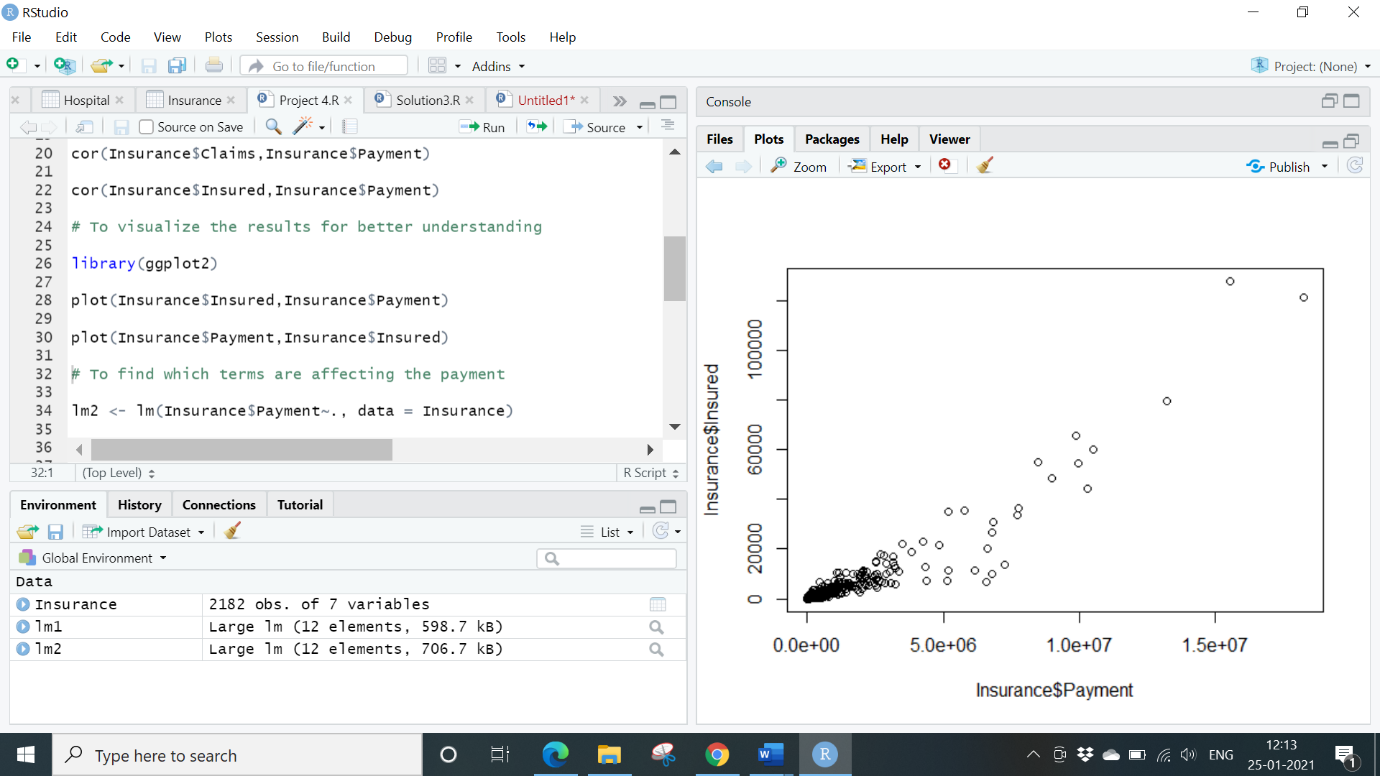
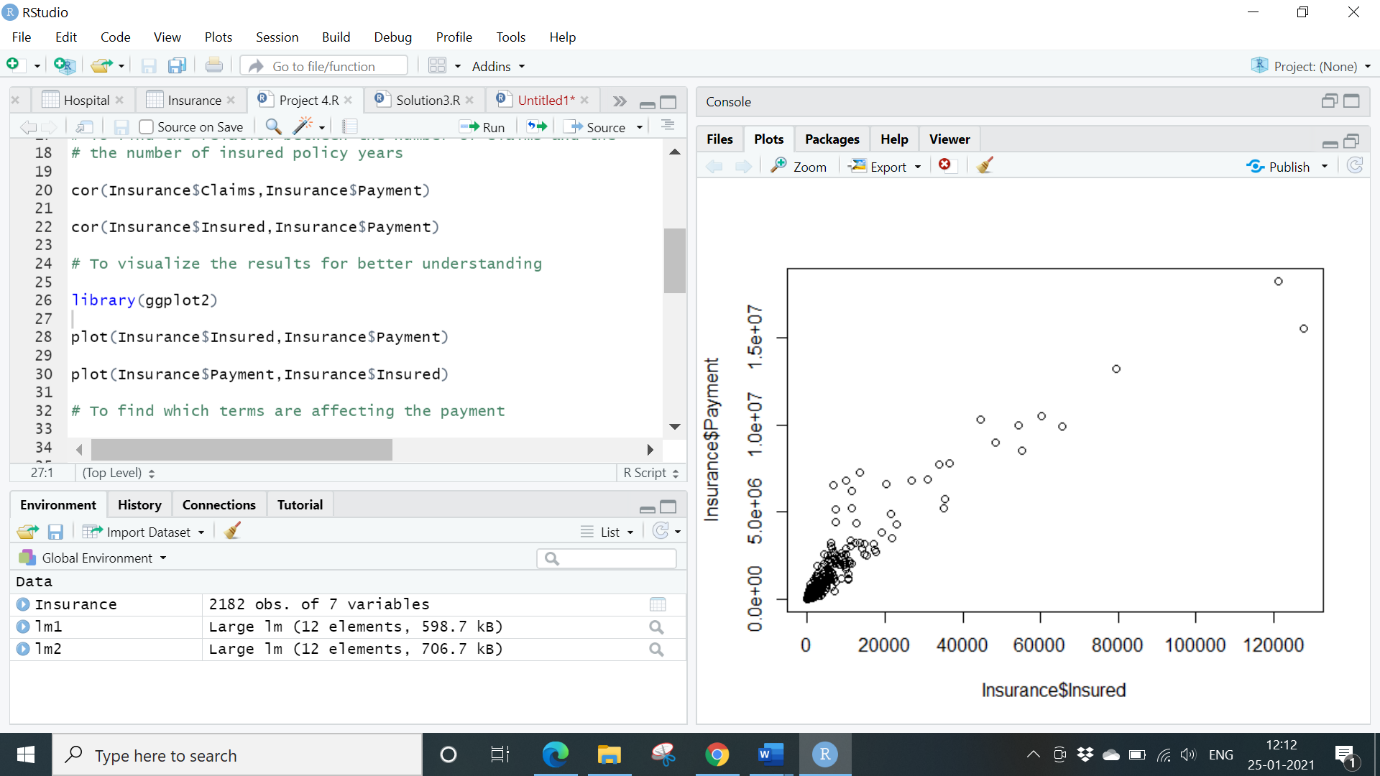
cor () – To find the co-relation between the terms

cor (Insurance$Claims, Insurance$Payment)

cor (Insurance$Insured, Insurance$Payment)

****

****

****

Both the terms Insured and claims were positively co-related with payment.

Scatter plot even denotes the same.

**Question 3.** The committee wants to figure out the reasons for insurance payment increase and decrease. So, they have decided to find whether distance, location, bonus, make, and insured amount or claims are affecting the payment or all or some of these are affecting it.

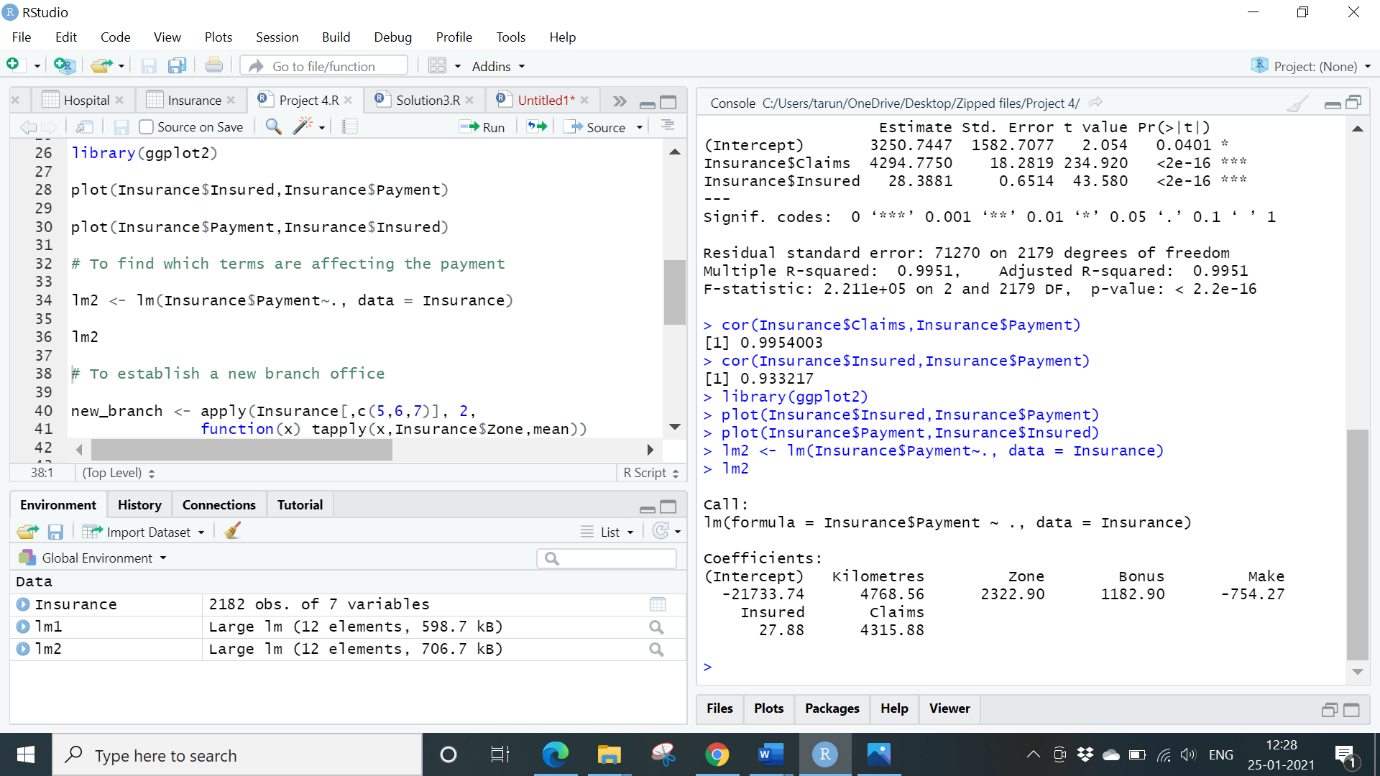
**Solution 3.**

**Code:**

# To find which terms are affecting the payment

lm2 <- lm (Insurance$Payment~., data = Insurance)

lm2

****

**Question 4.** The insurance company is planning to establish a new branch office, so they are interested to find at what location, kilometre, and bonus level their insured amount, claims, and payment gets increased.

**Solution 4.**

**Code:**

# To establish a new branch office

new\_branch <- apply (Insurance [, c (5,6,7)], 2,

function(x) tapply (x, Insurance$Zone, mean))

new\_branch

# to find at what location, kilometre, and bonus level their

# insured amount, claims, and payment gets increased.

high\_claims <- apply (Insurance [, c (5,6,7)],2,

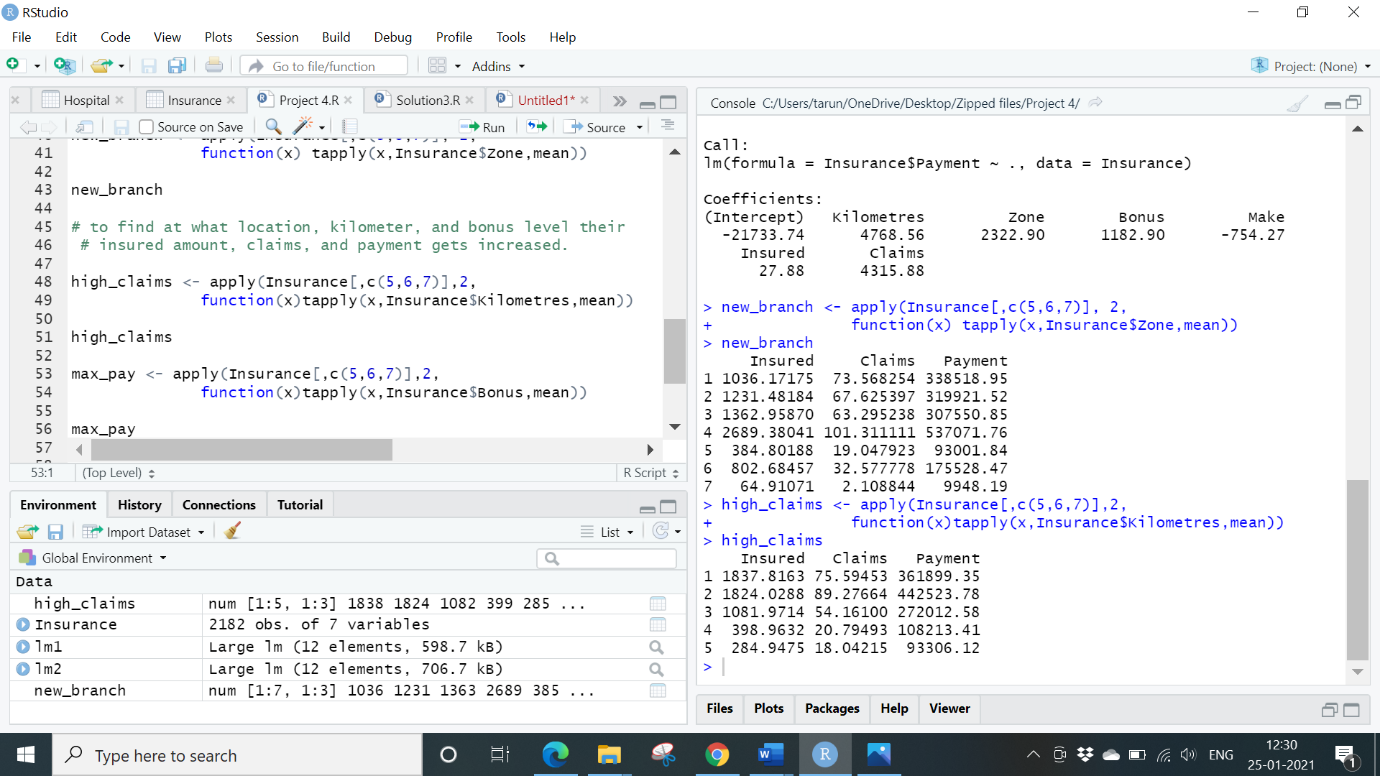
function(x)tapply (x, Insurance$Kilometres, mean))

high\_claims

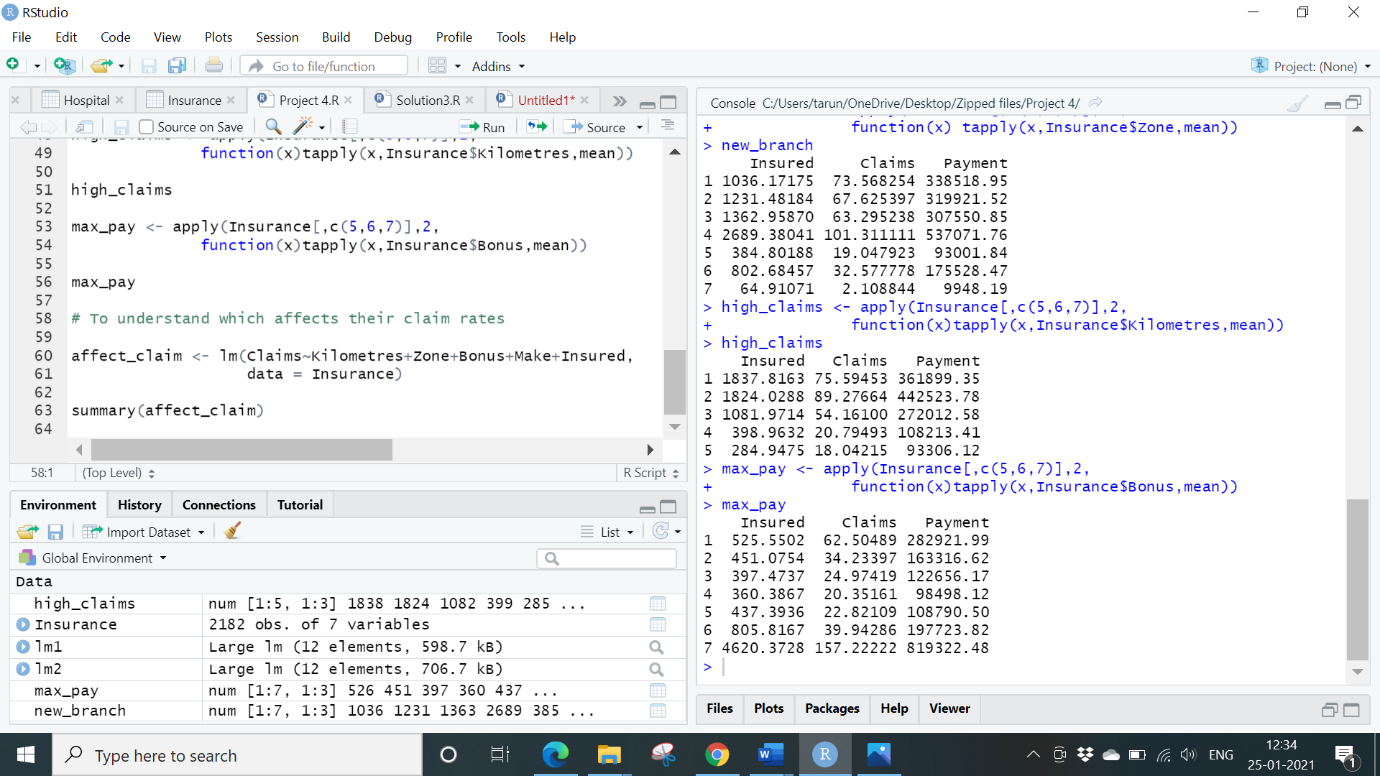
max\_pay <- apply (Insurance [, c (5,6,7)],2,

function(x)tapply (x, Insurance$Bonus, mean))

max\_pay



4 has highest number of claims and payment, 1-4 have more insured, claims and payments.



2 has higher claims and payments than all.

**Question 5.** The committee wants to understand what affects their claim rates so as to decide the right premiums for a certain set of situations. Hence, they need to find whether the insured amount, zone, kilometre, bonus, or make affects the claim rates and to what extent.

**Solution 5.**

Dependent Variable as Claims and rest as Independent variables

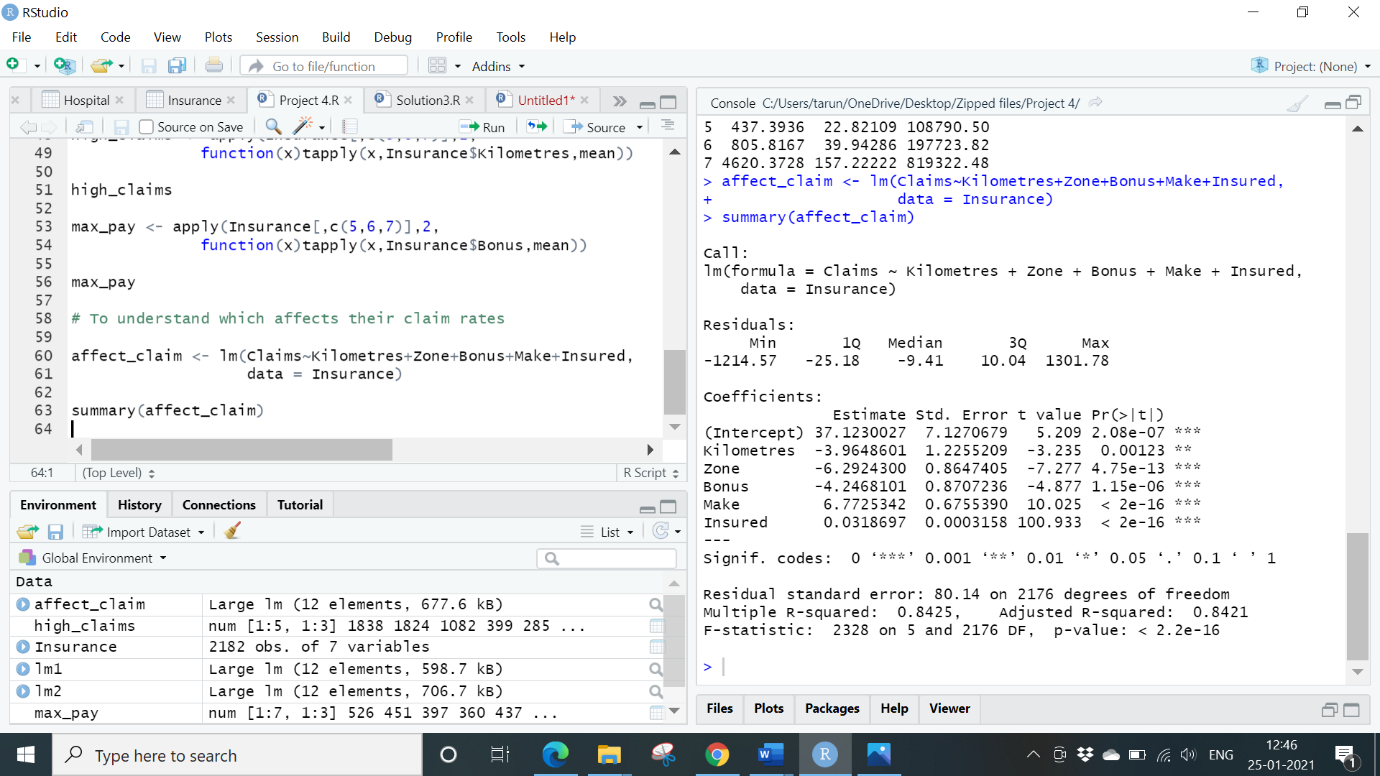
**Code:**

# To understand which affects their claim rates

affect\_claim <- lm (Claims~Kilometres+Zone+Bonus+Make+Insured,

data = Insurance)

summary(affect\_claim)



P values of independent variables are making an impact on the claims.