**Healthcare Cost Analysis**

**Business Scenario:**

A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyze the data to research on healthcare costs and their utilization.

**Expectation /Goals**

1. To record the patient statistics, the agency wants to find the age category  
   of people who frequent the hospital and has the maximum expenditure.

**Code:**

hospital\_cost <- read.csv("HospitalCosts.csv") #importing the dataset

View(hospital\_cost)

table(hospital\_cost$AGE)

max(summary(as.factor(hospital\_cost$AGE)))

#for visual plotting of the above result

hist(hospital\_cost$AGE,xlab = "Age", ylab = 'No. of Visits',col = "green",main = "Age wise Frequency of Paitents")

#for expendiature analysis

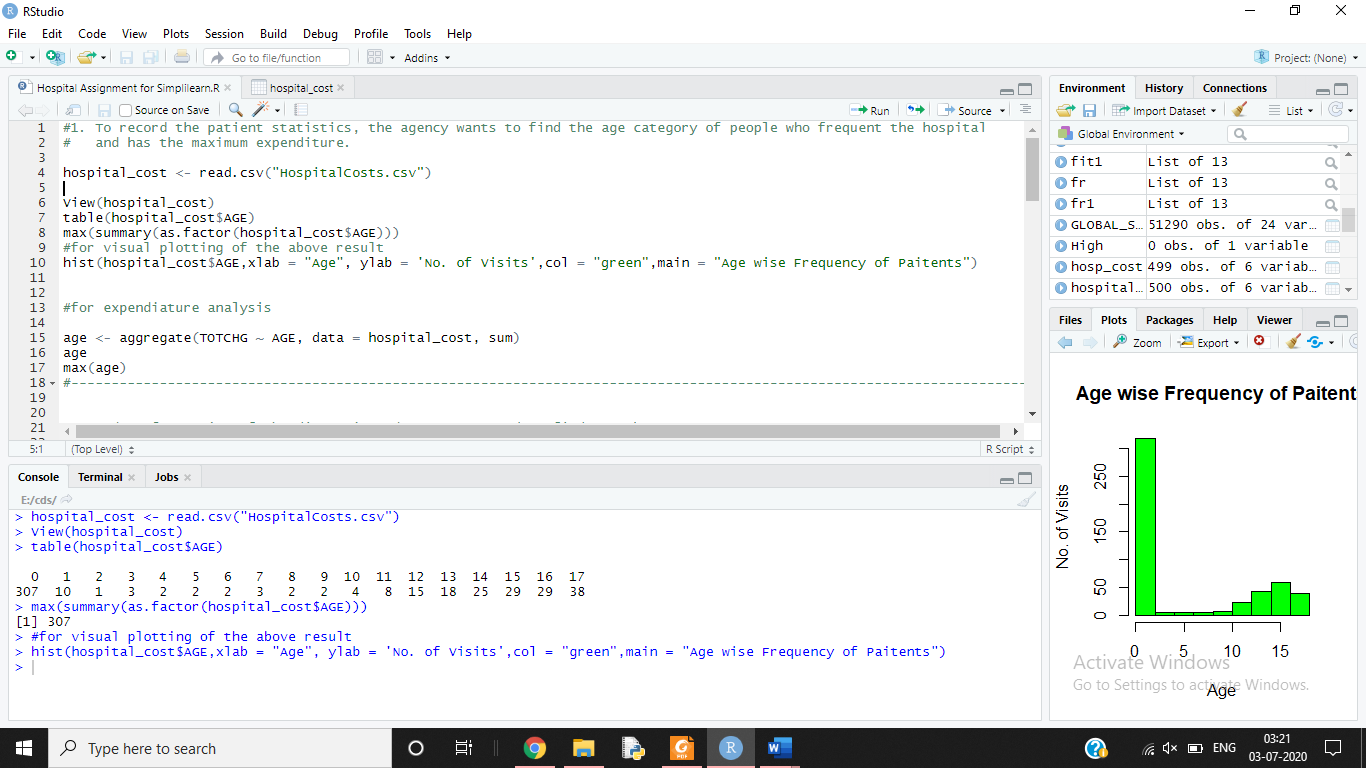
age <- aggregate(TOTCHG ~ AGE, data = hospital\_cost, sum)

age

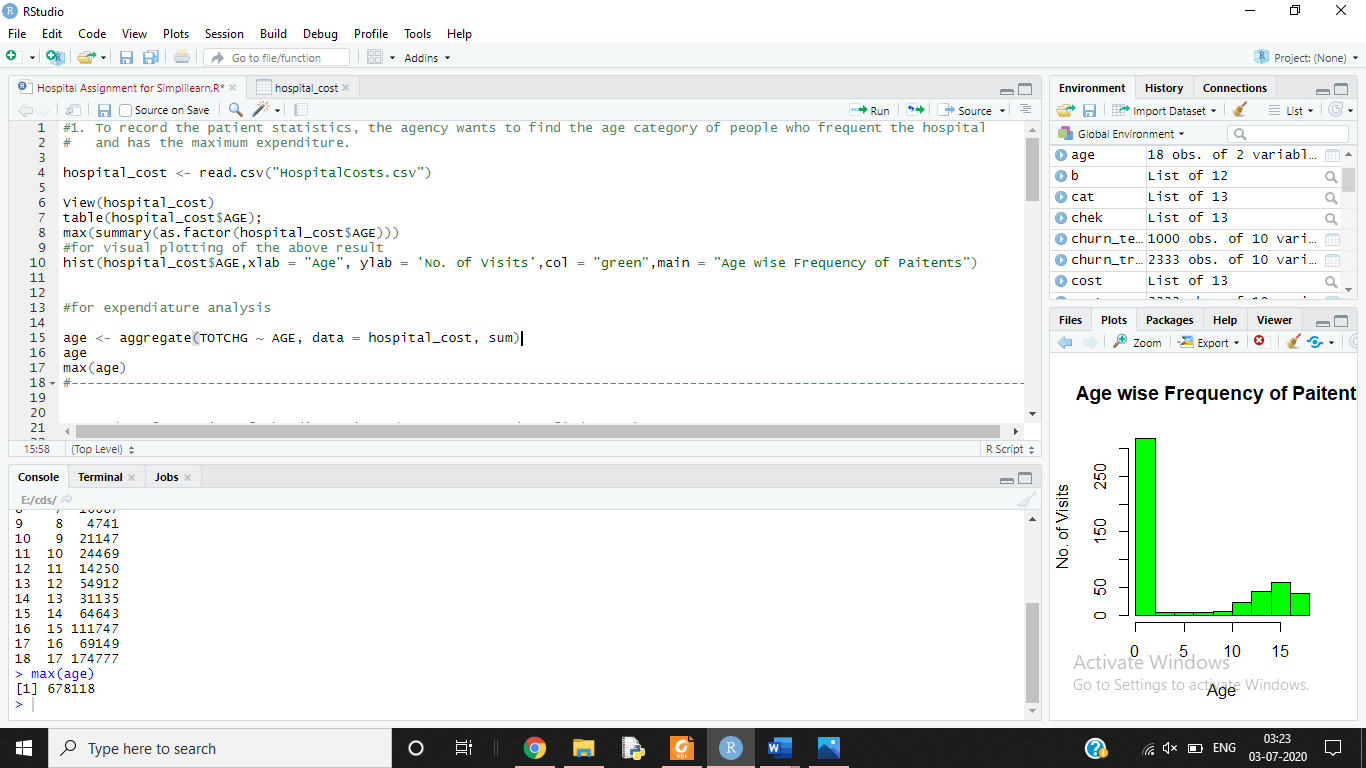
max(age)

**Output Screenshot:**

1. For finding the age category with max frequency of visit



1. Finding and Calculating the Expenditure



**Analysis:**

From the graph plotted it can be clearly seen that the INFANTS are having the maximum frequency of visits with the number of visits being = 307.

This was again re-verified using maximum of summary function which resulted in 307 and was later plotted using graph.

For **CALCULATING THE MAXIMUM EXPENDITURE** :

Aggregate function was used to sum up the expenditure of each age category and then maximum function is used to find out the maximum out of it, which also comes out to be the age group 0 [INFANTS].

Now we can surely say that the, **The Hospital Expenditure is Directly proportional to the No. of Hospital visits , i.e. more the number of visits more will be the expenditure. (**As proven above in the analysis**).**

1. In order of severity of the diagnosis and treatments and to find out the  
   expensive treatments, the agency wants to find the diagnosis related group  
   that has maximum hospitalization and expenditure

**Code:**

t <- table(hospital\_cost$APRDRG)

d <- as.data.frame(t)

names(d)[1] = 'Diagnosis Group'

d

which.max(table(hospital\_cost$APRDRG))

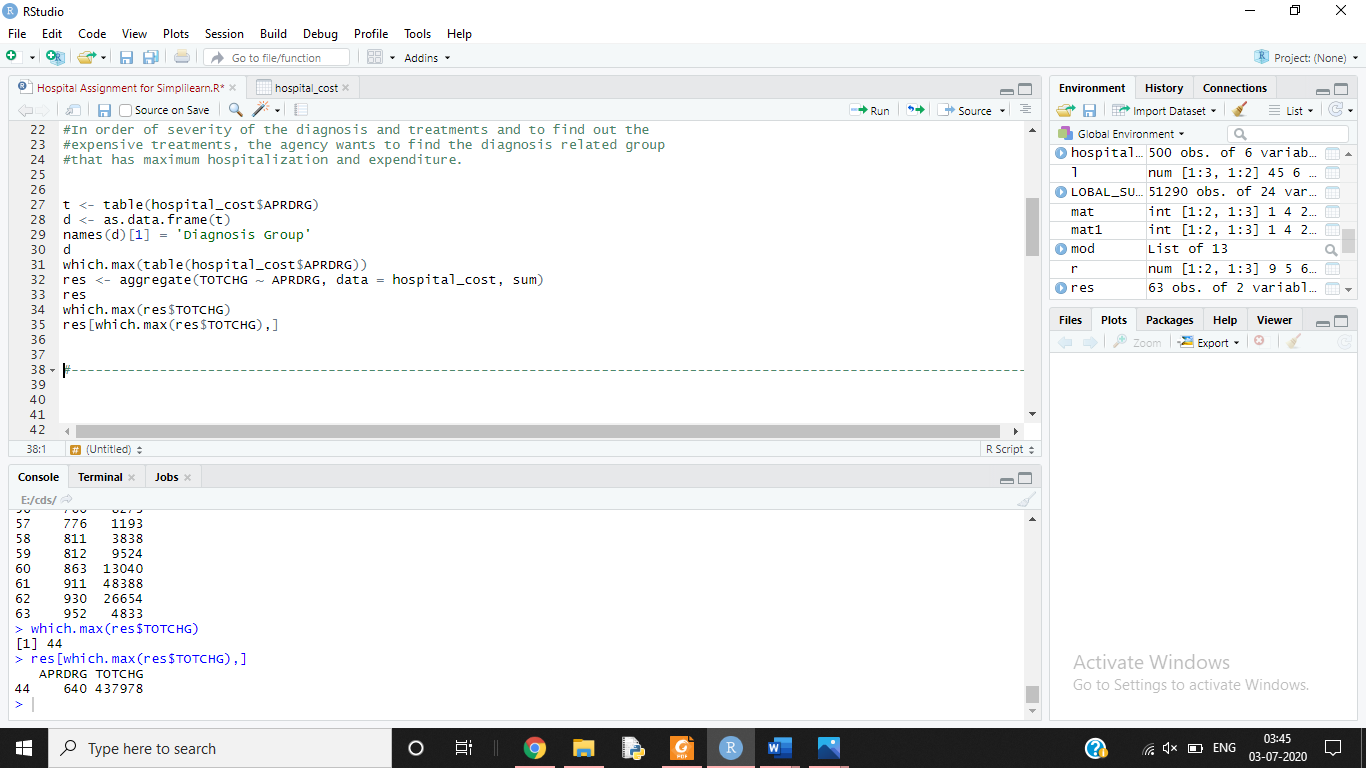
res <- aggregate(TOTCHG ~ APRDRG, data = hospital\_cost, sum)

res

which.max(res$TOTCHG)

res[which.max(res$TOTCHG),]

**Output Screenshot:**



**Analysis:**

The column APRDRG from the data set was used along with functions like which.max and aggregate and was also converted into a DF(Dataframe). It was seen that the category “640” was having the maximum Hospitalization around 267 out of the total 500 along with the highest hospitalization cost amounting: **437978**

1. To make sure that there is no malpractice, the agency needs to analyze if  
   the race of the patient is related to the hospitalization costs.

**Code:**

hosp\_cost <- na.omit(hospital\_cost)

hosp\_cost$RACE <- as.factor(hosp\_cost$RACE)

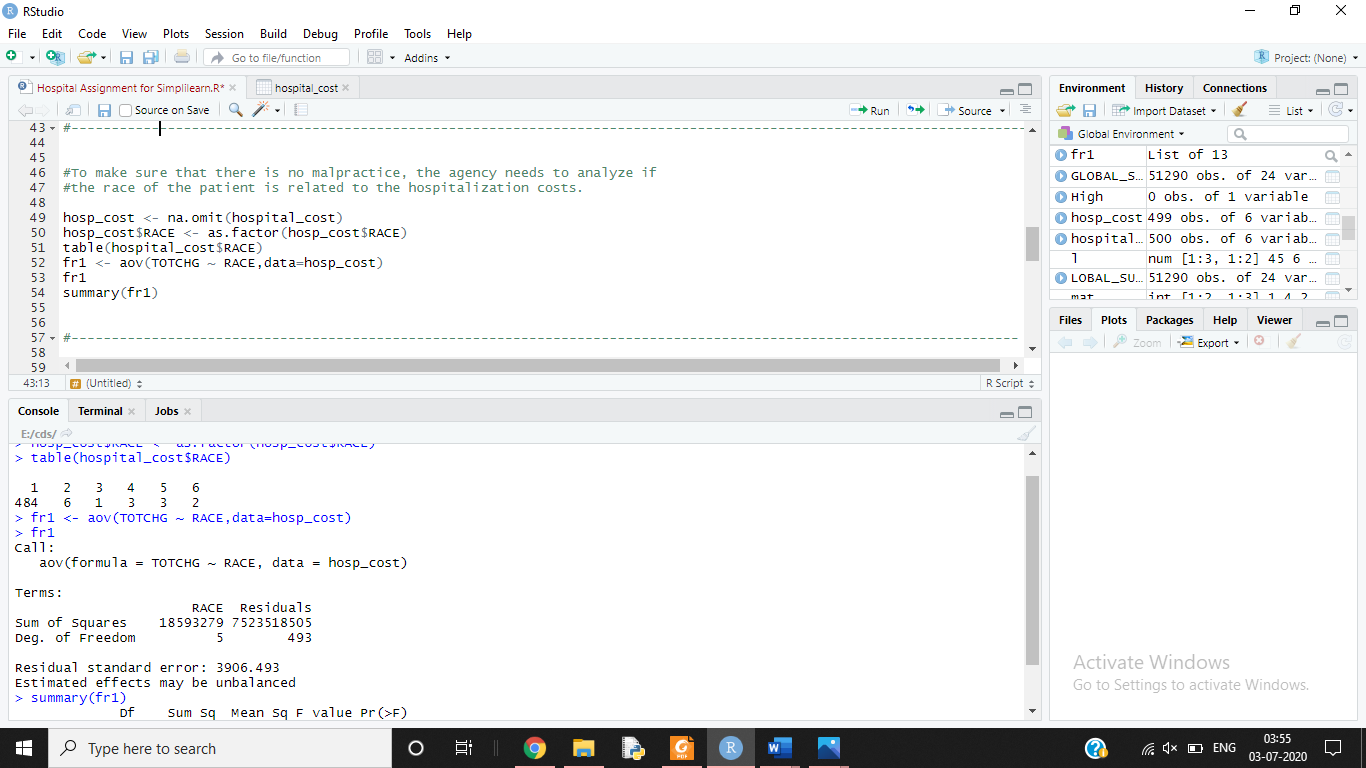
table(hospital\_cost$RACE)

fr1 <- aov(TOTCHG ~ RACE,data=hosp\_cost)

fr1

summary(fr1)

**Output Screenshot:**



**Analysis:**

Firstly, we have to omit the NA values in order to maintain high accuracy in the results and then we’ll do factorization of the Race variable. Then we’ll used aov function in order to check if Race made any impact on the Hospitalization cost.

The lower **F value** indicated that variation between hospital cost of different Race is quite smaller as compared to variation among the same Race.

The higher P value showed that there is no relationship between the hospital cost and the Race.

Also, we are having around 484 out of 500 values of data for the Race 1 which makes it less reliable and it can be concluded that enough data was not available for a fair analysis to find the relationship between the hospital cost and patient’s Race.

1. To properly utilize the costs, the agency has to analyze the severity of the  
   hospital costs by age and gender for proper allocation of resources.

**Code:**

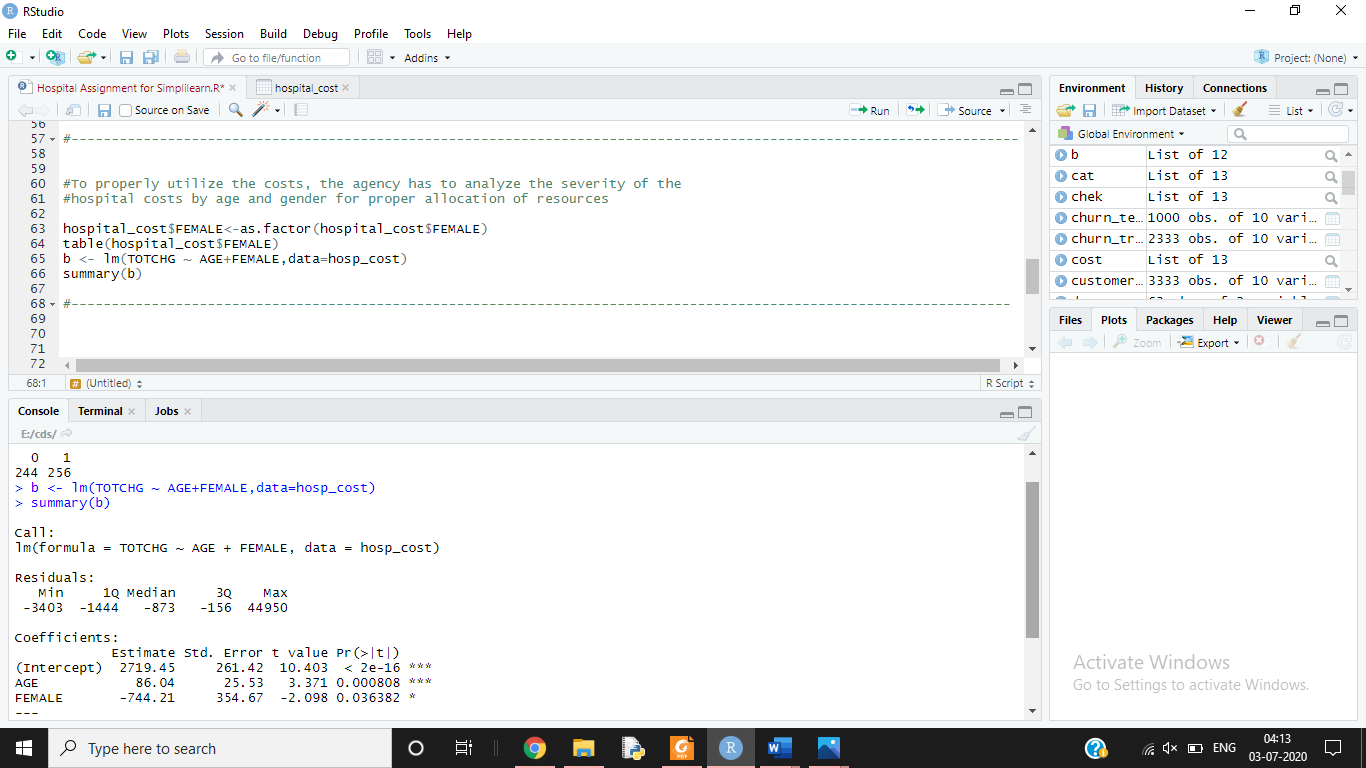
hospital\_cost$FEMALE<-as.factor(hospital\_cost$FEMALE)

table(hospital\_cost$FEMALE)

b <- lm(TOTCHG ~ AGE+FEMALE,data=hosp\_cost)

summary(b)

**Output Screenshot:**



**Analysis:**

For the analysis of severity of hospital cost, we’ll use Linear Regression with hospital discharge cost (TOTCHG) and will find the discharge cost in accordance to AGE and Female (they’ll act as the dependent variable).

According to the P-value and Significant levels, Age has more impact than gender.

The number of Females and Males are almost equal and on an average, **Females incur lesser Hospital Costs than Males**(Based on the negative coefficient values).

1. Since the length of stay is the crucial factor for inpatients, the agency wants  
   to find if the length of stay can be predicted from age, gender, and race.

**Code:**

hospital\_cost$RACE<-as.factor(hospital\_cost$RACE)

table(hosp\_cost$LOS)

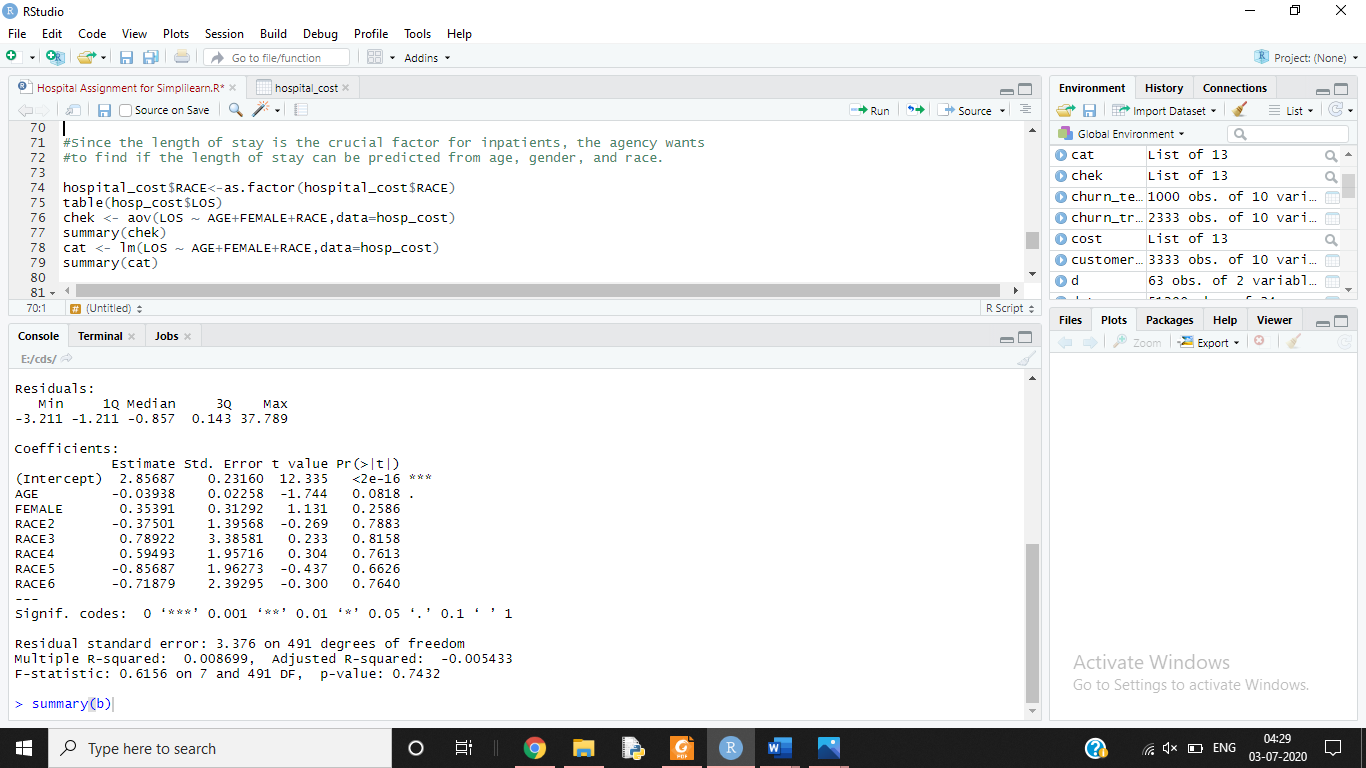
chek <- aov(LOS ~ AGE+FEMALE+RACE,data=hosp\_cost)

summary(chek)

cat <- lm(LOS ~ AGE+FEMALE+RACE,data=hosp\_cost)

summary(cat)

**Output Screenshot:**



**Analysis:**

Simply by using Linear Regression we can find if Length of Stay is dependent on Age, Gender or Race and they’ll act as the independent variable here.

For all the Independent Variables, the P-value is quite hight which shows that there is no linear relationship between them and this can be concluded that we cannot predict the Length of Stay of a patient based on their Age, Gender or Race.

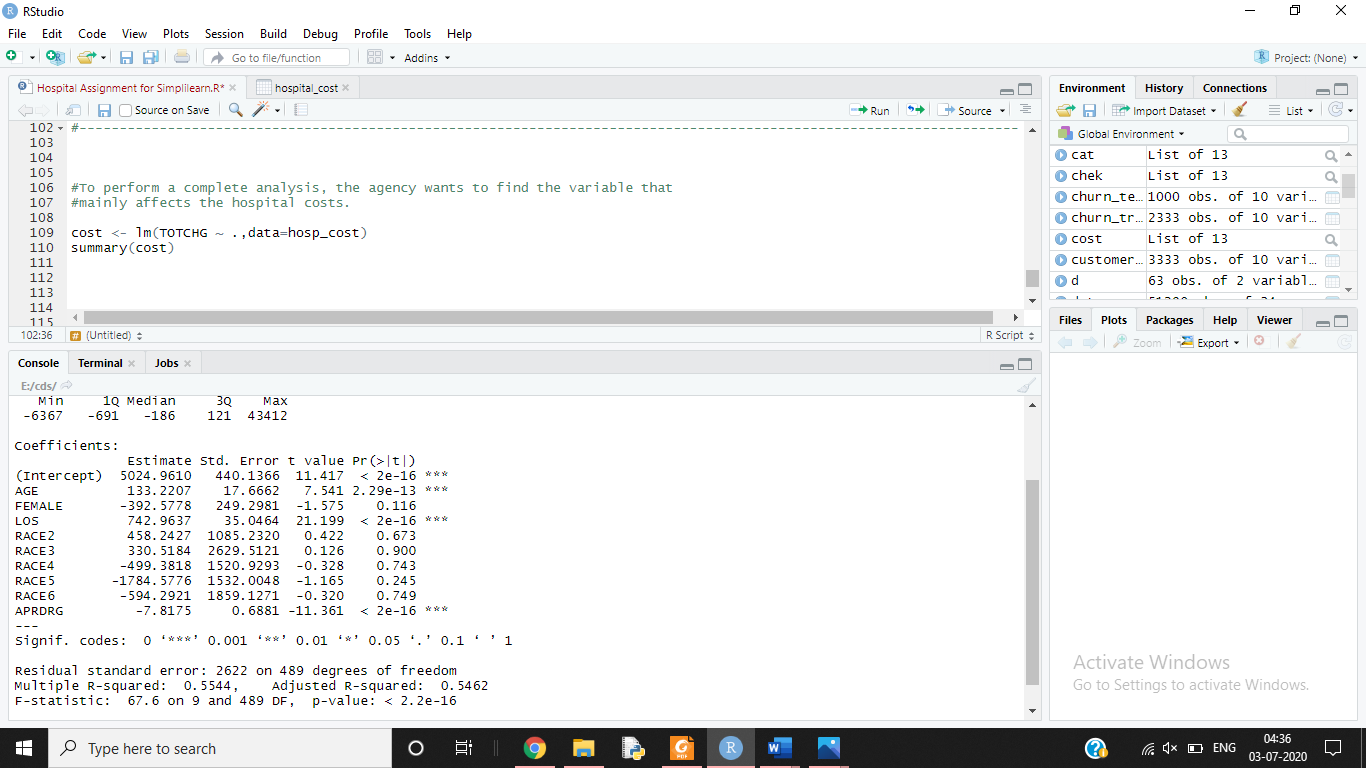
1. To perform a complete analysis, the agency wants to find the variable that  
   mainly affects the hospital costs.

**Code:**

cost <- lm(TOTCHG ~ .,data=hosp\_cost)

summary(cost)

**Output Screenshot:**



**Analysis:**

Again, by using linear regression on the TOTCHG(Hospital cost) we can check on which other variable it depends.

It was found that Age and Length of Stay affects the Hospital Cost.

Also, there is a positive relationship between Length of Stay and Hospital cost, so this can be concluded that with an increase in Length of stay by 1 , the Hospital cost increases by 742.9. (approximately 743).