**PROJECT**

**ON**

**HEALTHCARE COST**

**Background and Objective:**

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.

Here is a detailed description of the given dataset:

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Age | Age of the patient discharged |
| Female | A binary variable that indicates if the patient is female |
| Los | Length of stay in days |
| Race | Race of the patient (specified numerically) |
| Totchg | Hospital discharge costs |
| Aprdrg | All Patient Refined Diagnosis Related Groups |

SOLUTION :

CODE:

Q1

**HC<-read.csv(file.choose(),header=T)**

**head(HC)**

**#Graphical Representation(no. of occurance of age)**

**hist(HC$AGE)**

**summary(as.factor(HC$AGE))**

**#RESULT**

**#infants(0-1) have the max. frequency of Hospital visit ie.307**

**#Age catagory that has max. hospitalisation**

**tapply(HC$TOTCHG,HC$AGE,sum)**

**which.max(tapply(HC$TOTCHG,HC$AGE,sum))**

**#RESULT**

**#max expenditure also by infant,678118**

**Q2**

**diags<-as.factor(HC$APRDRG)**

**summary(diags)**

**which.max(summary(diags))**

**#Catagory 640 has the max. entries of hospitalisation**

**tapply(HC$TOTCHG,diags,sum)**

**which.max(tapply(HC$TOTCHG,diags,sum))**

**#also the hospital cost is max for 640**

**max(tapply(HC$TOTCHG,diags,sum))**

**#max cost is 437978**

**Q3**

**#Ho:the race of patient is related to hospitalisation cost**

**#Ha:not related**

**rc<-as.factor(HC$RACE)**

**summary(rc)**

**#to omit NA value**

**HCna<-na.omit(HC)**

**MODElanv<-aov(HC$TOTCHG~HC$RACE)**

**summary(MODElanv)**

**#Ho rejected as per pvalue 68%**

**#so,there is no relation between RACE and hospitalisation cost**

**Q4**

**MDLlm1<-lm(HC$TOTCHG~HC$AGE+HC$FEMALE)**

**summary(MDLlm1)**

**#pvalue is very less,so age is important factor for hospital cost**

**#gender also important factor**

**Q5**

**MDLlm2<-lm(HC$LOS~HC$AGE+HC$FEMALE+HC$RACE)**

**summary(MDLlm2)**

**#loc has no linear relationship with age,gender&race**

**Q6**

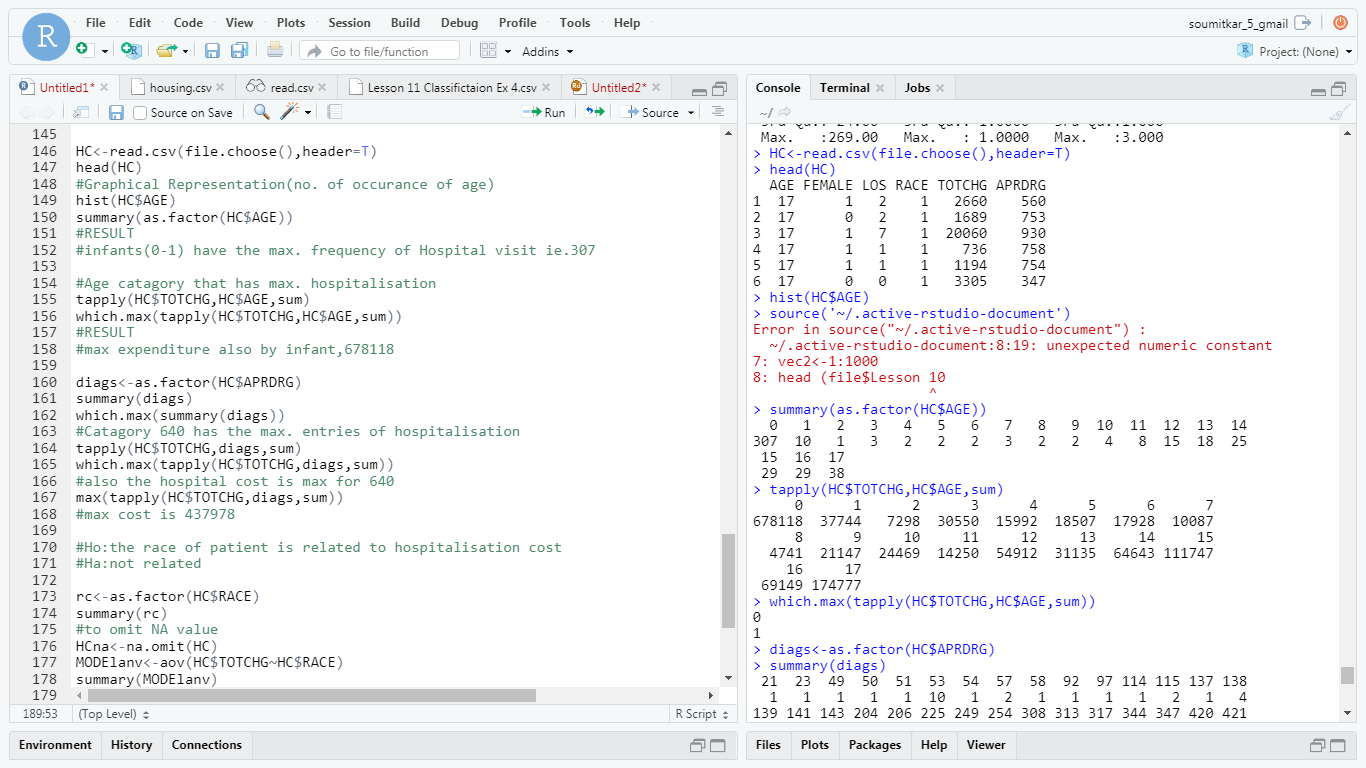
**MDLlm3<-lm(TOTCHG~.,data=HCna)**

**summary(MDLlm3)**

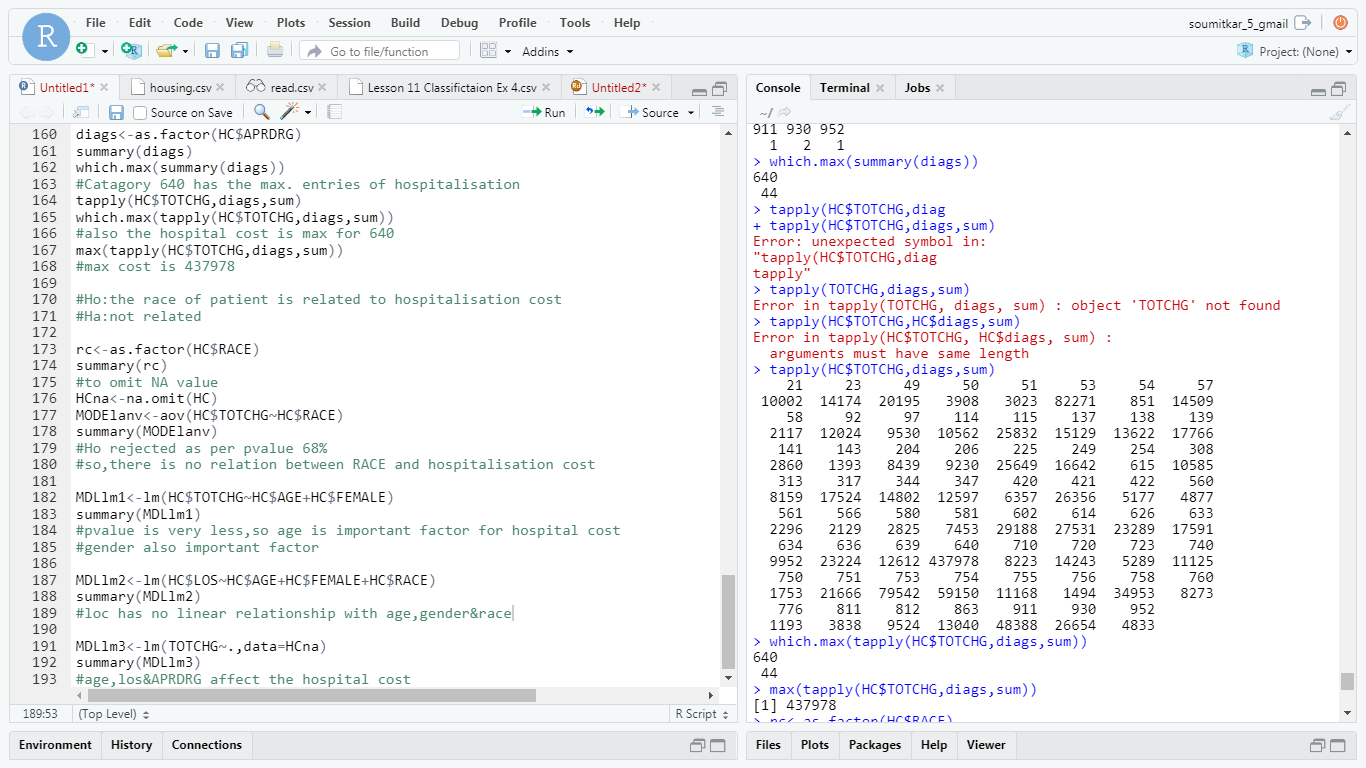
**#age,los&APRDRG affect the hospital cost**

**SCREENSHOTS**

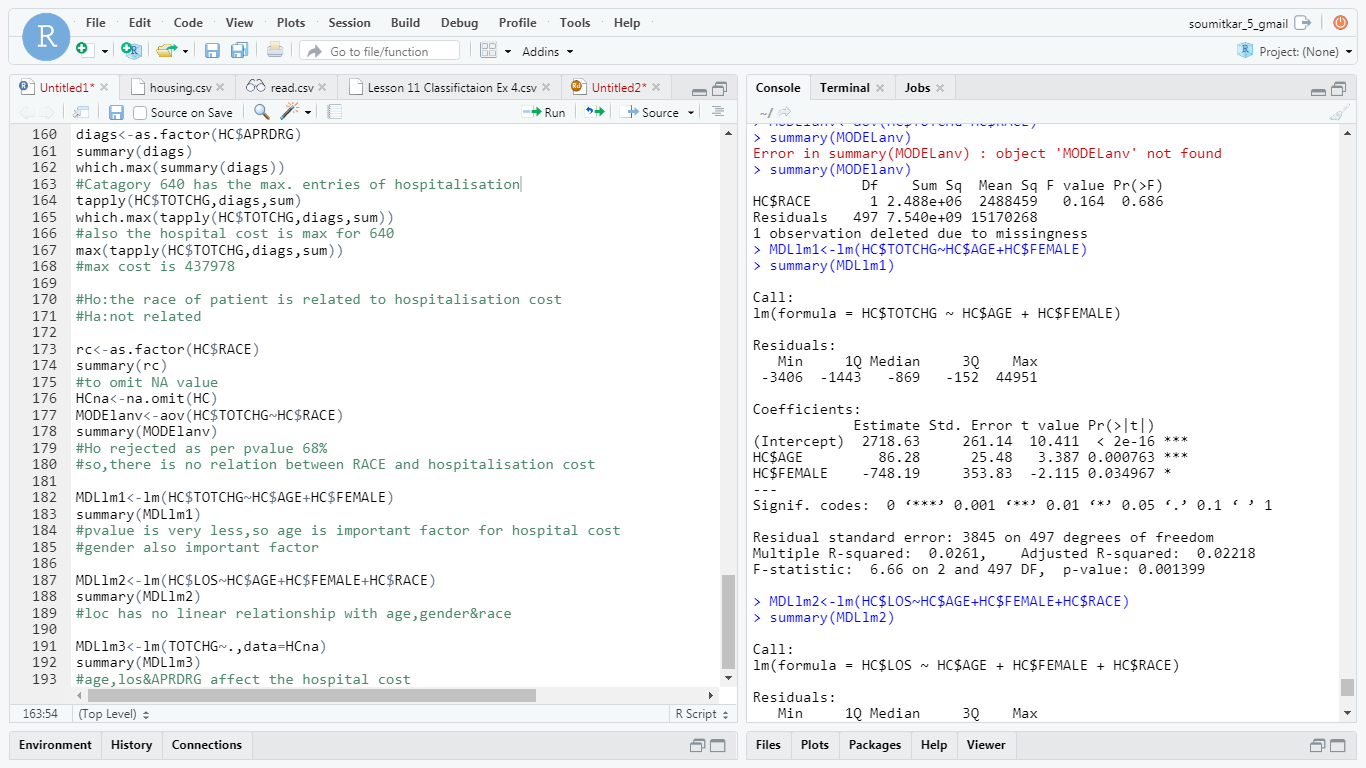
**SCREENSHOT 1**

****

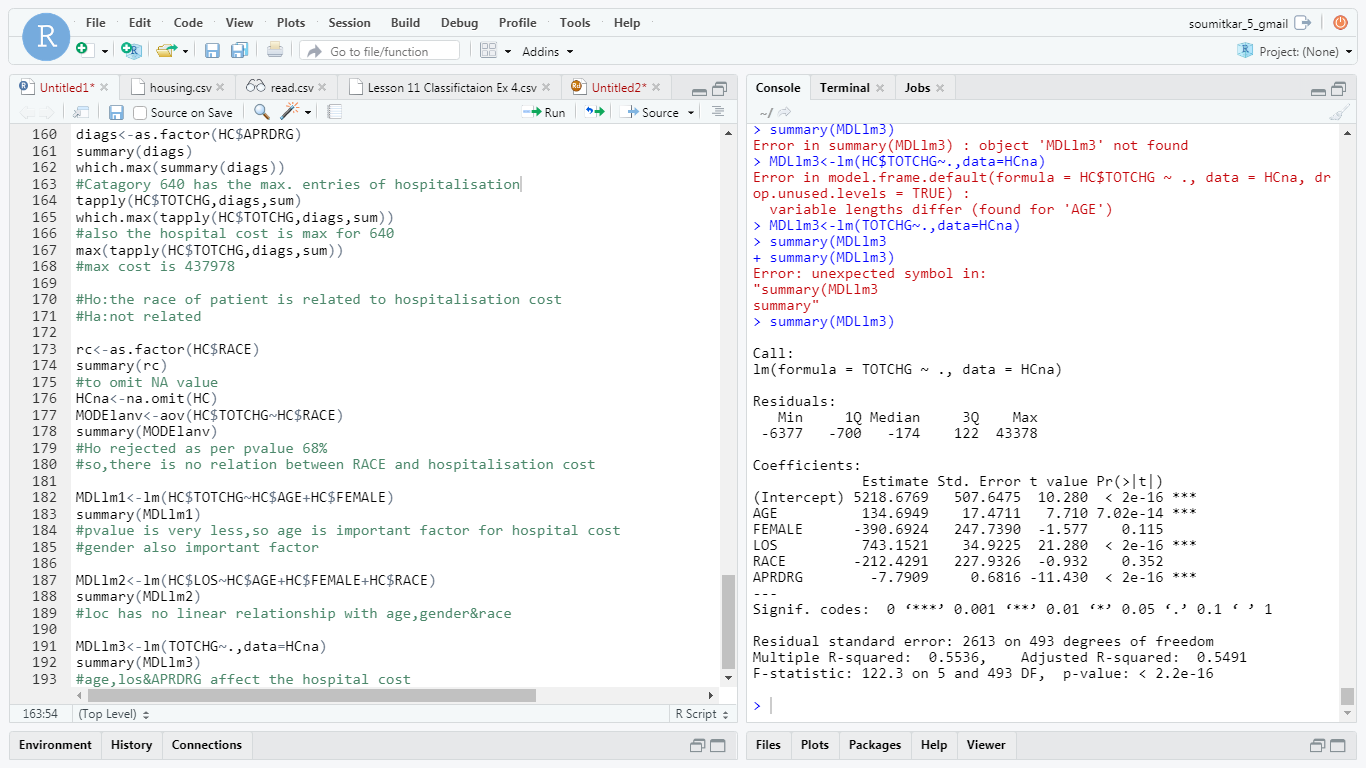
**SCREENSHOT 2**

****

**SCREENSHOT3**

****

**SCREENSHOT4**

****

**Analyze the Healthcare cost and Utilization in Wisconsin hospitals.**

**Submitted By :**

**NAME : Soumit Kar**

**Mail : soumitkar.5@gmail.com**