**#find the age category of the people who frequent the hospital and has the max expenditure.**

hosp\_cost<-read.table(file.choose(),sep=",",header=TRUE)

head(hosp\_cost)

summary(hosp\_cost)

head(hosp\_cost$AGE)

summary(hosp\_cost$AGE)

table(hosp\_cost$AGE)

hist(hosp\_cost$AGE)

summary(as.factor(hosp\_cost$AGE))

max(table(hosp\_cost$AGE))

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

which.max(table(hosp\_cost$AGE))

tapply(hosp\_cost$TOTCHG,hosp\_cost$AGE,sum)

max(tapply(hosp\_cost$TOTCHG,hosp\_cost$AGE,sum))

**# to find the diagnosis related group which has maximum hospitalization and expenditure**

diag<-as.factor(hosp\_cost$APRDRG)

diag

str(hosp\_cost$APRDRG)

summary(diag)

which.max(summary(diag))

diag1<-tapply(hosp\_cost$TOTCHG,diag,sum)

which.max(diag1)

max(diag1)

#**analyze if the race of the patient is related to the hospital cost**

hosp\_cost$RACE1<-as.factor(hosp\_cost$RACE)

summary(rc)

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

#annova model

result<-aov(hosp\_cost$TOTCHG~hosp\_cost$RACE)

summary(result)

**#analyze the severity of the hospital cost by age and gender for proper allocation of resources**

fit<-lm(hosp\_cost$TOTCHG~hosp\_cost$AGE+hosp\_cost$FEMALE)

summary(fit)

**#find if the length of the stay can be predicted from age, gender and race**

los<-lm(hosp\_cost$LOS~hosp\_cost$AGE+hosp\_cost$FEMALE+hosp\_cost$RACE)

summary(los)

#**find the variable that affects the hospital cost**

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

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

summary(var)