

Problem Statement

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 the healthcare costs and their utilization.

IMPORTING DATASETS

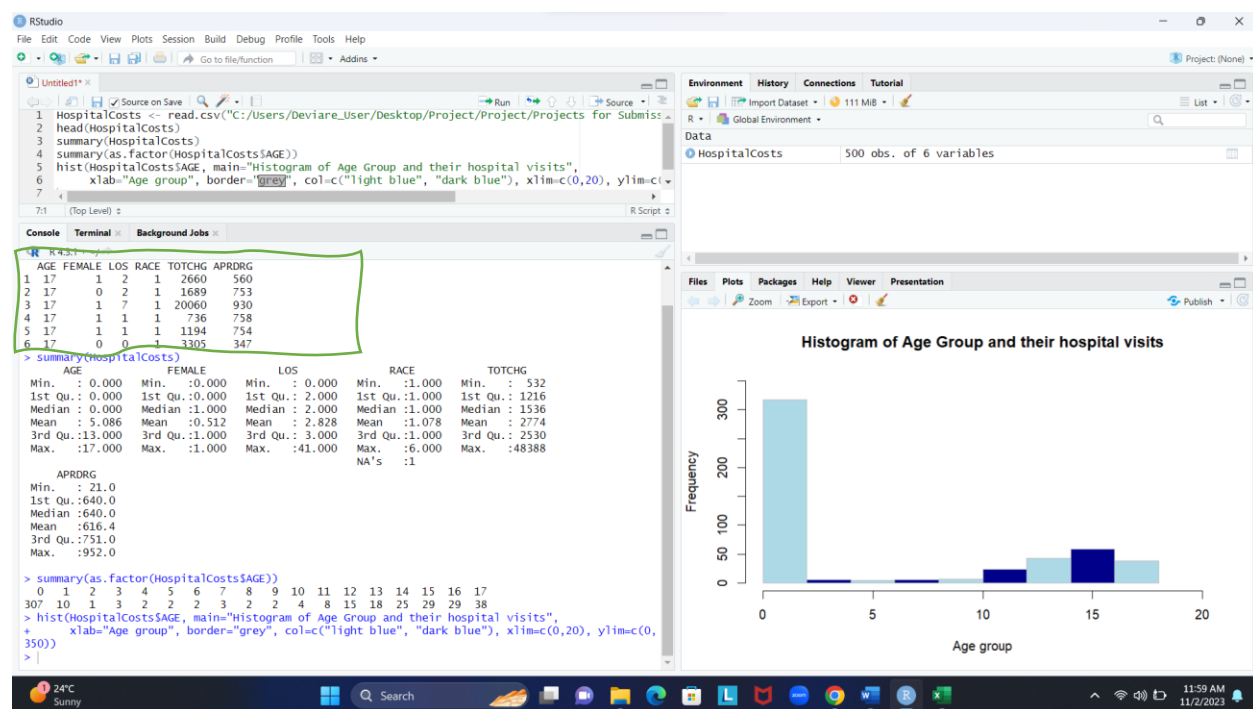
CODE

HospitalCosts <-

```
read.csv("C:/Users/Deviare_User/Desktop/Project/Project/Projects for Submission/Healthcare/Healthcare/HospitalCosts.csv")
```

```
head(HospitalCosts)
```

OUTPUT



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

```
summary(HospitalCosts)
```

```
summary(as.factor(HospitalCosts$AGE))
```

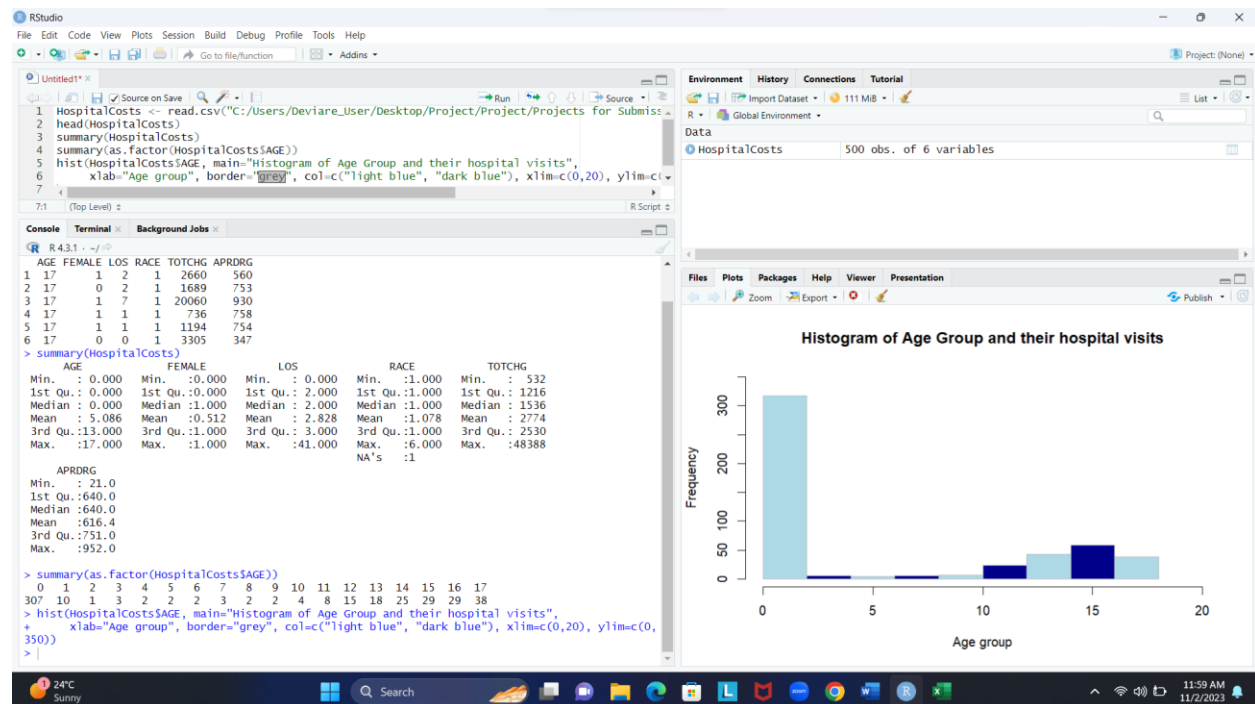
```
hist(HospitalCosts$AGE, main="Histogram of Age Group and their hospital visits",  
xlab="Age group", border="grey", col=c("light blue", "dark blue"), xlim=c(0,20),  
ylim=c(0,350))
```

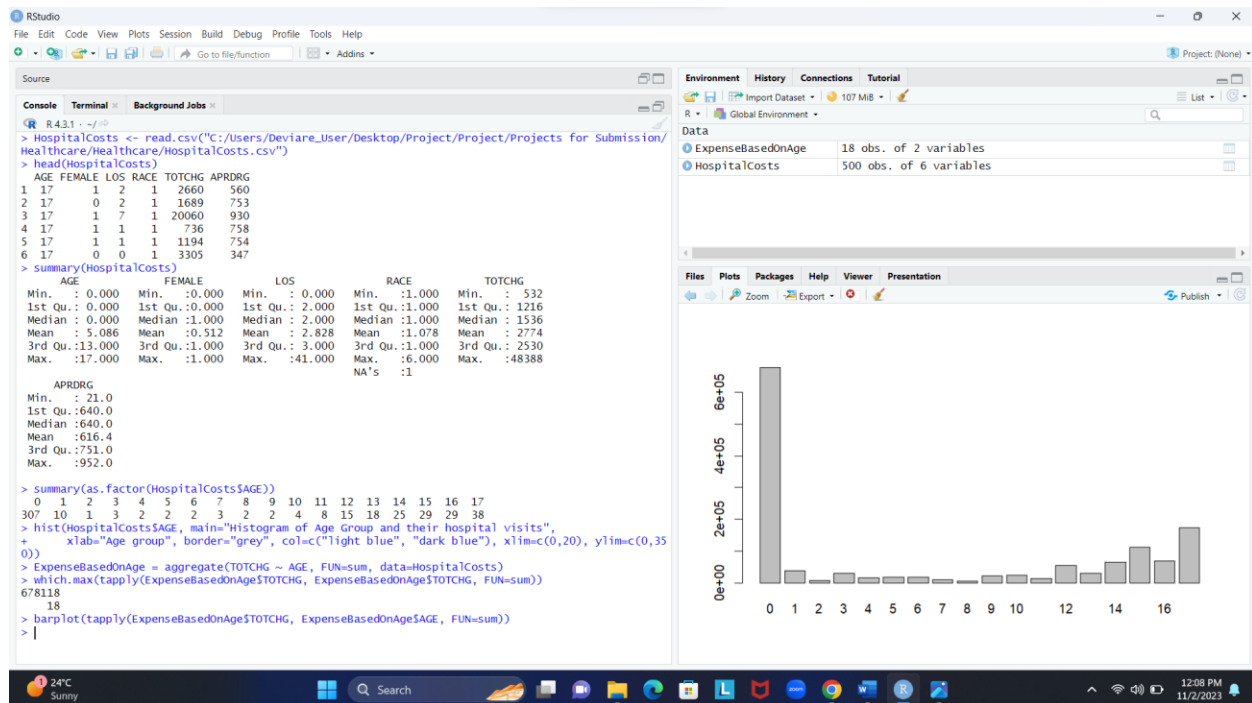
```
ExpenseBasedOnAge = aggregate(TOTCHG ~ AGE, FUN=sum, data=HospitalCosts)
```

```
which.max(tapply(ExpenseBasedOnAge$TOTCHG, ExpenseBasedOnAge$AGE, FUN=sum))
```

```
barplot(tapply(ExpenseBasedOnAge$TOTCHG, ExpenseBasedOnAge$AGE, FUN=sum))
```

OUTPUT





2. 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.

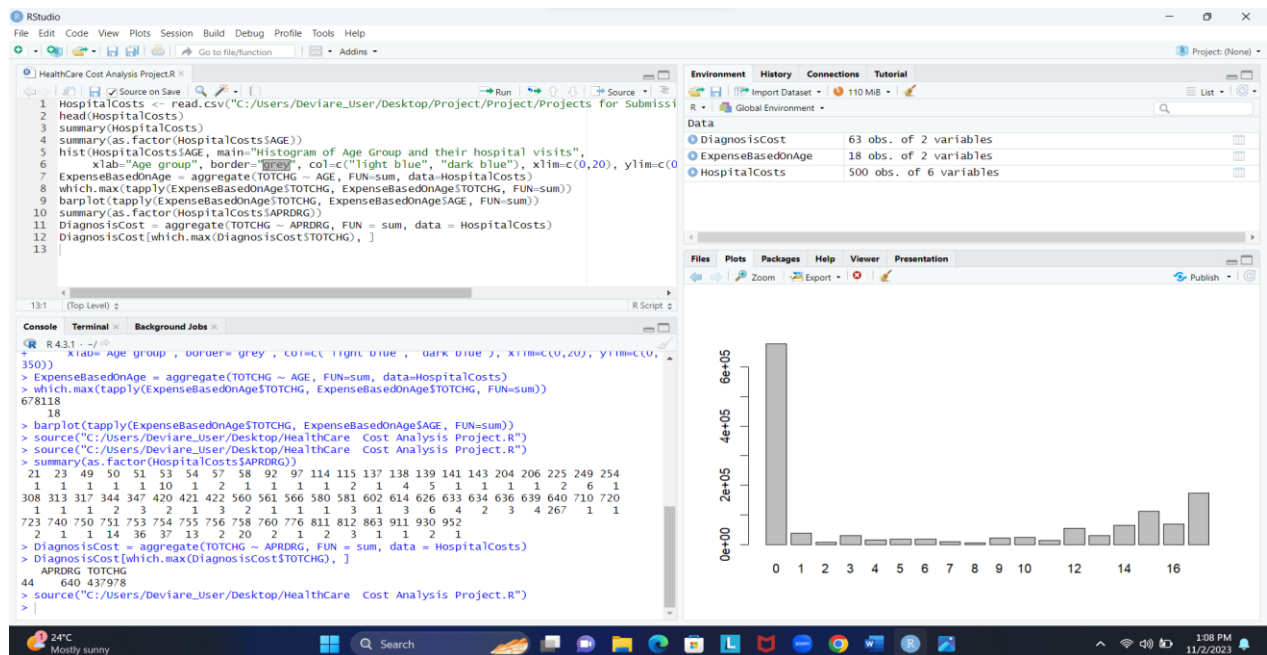
CODE

```
summary(as.factor(HospitalCosts$APRDRG))
```

```
DiagnosisCost = aggregate(TOTCHG ~ APRDRG, FUN = sum, data = HospitalCosts)
```

```
DiagnosisCost[which.max(DiagnosisCost$TOTCHG), ]
```

OUTPUT



3. 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

```
summary(as.factor(HospitalCosts$RACE))
```

```
raceInfluence=lm(TOTCHG~ RACE, data=HospitalCosts)
```

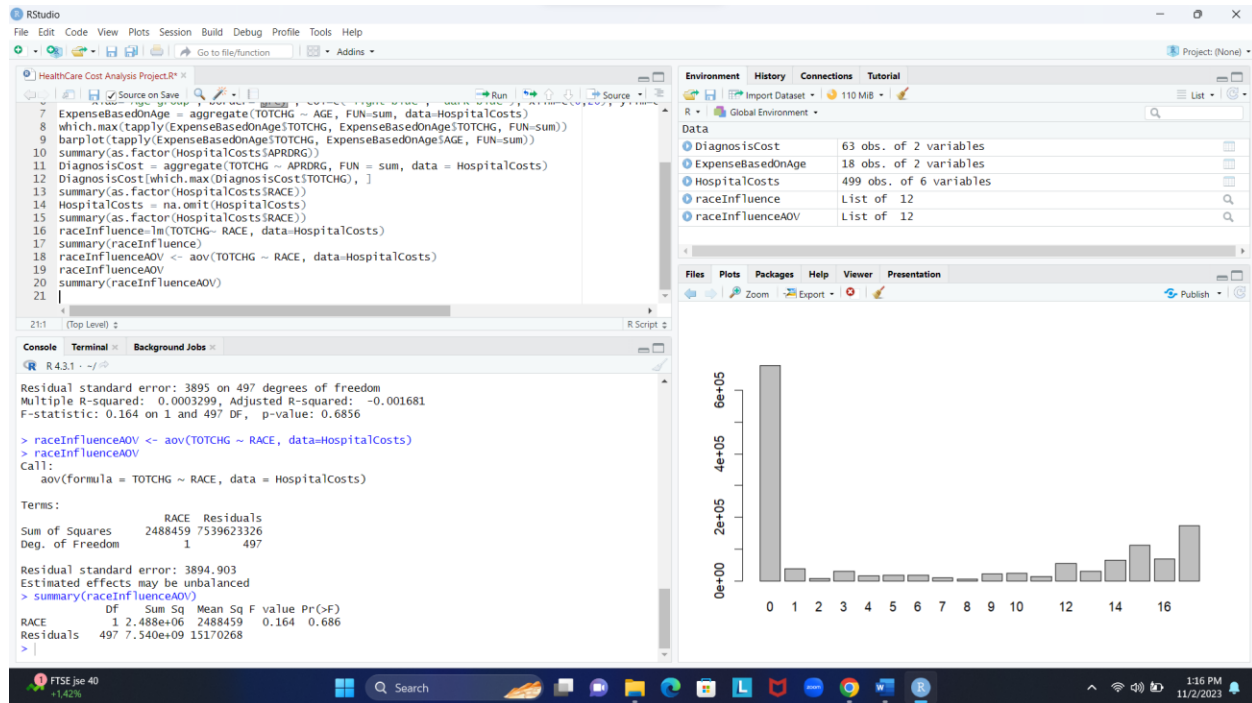
```
summary(raceInfluence)
```

```
raceInfluenceAOV <- aov(TOTCHG ~ RACE, data=HospitalCosts)
```

```
raceInfluenceAOV
```

```
summary(raceInfluenceAOV)
```

OUTPUT



4. 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.

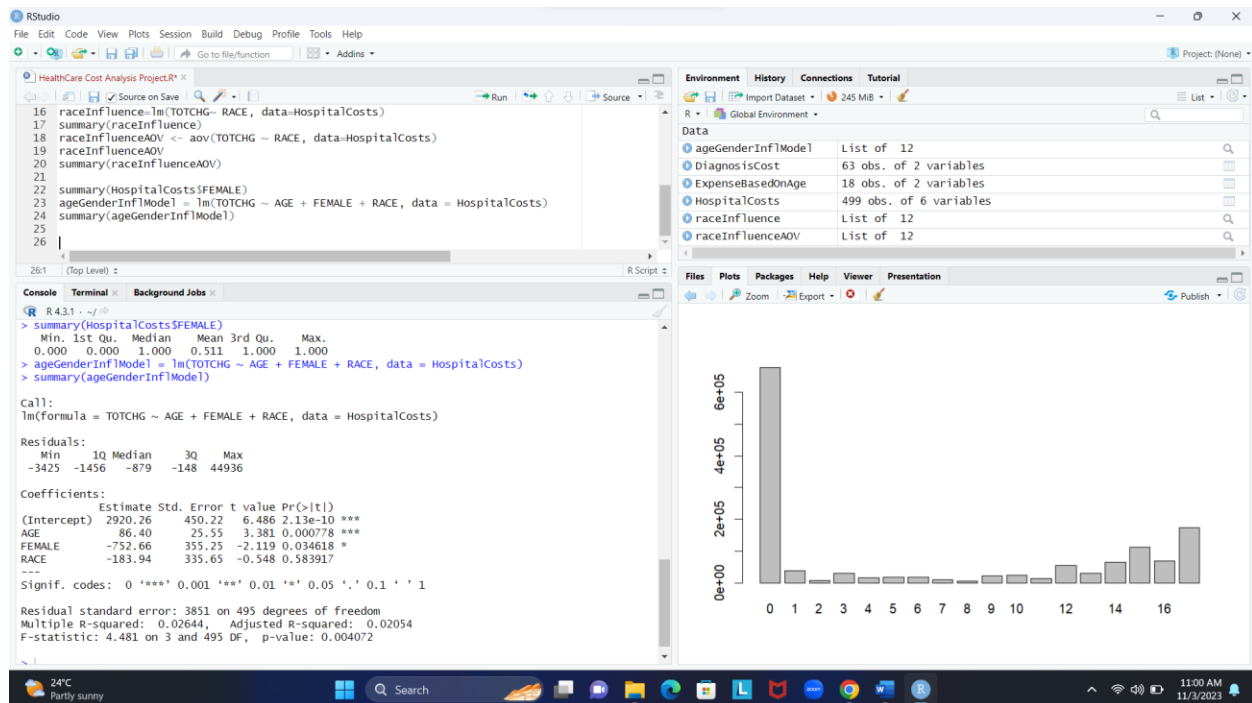
CODE

```
summary(HospitalCosts$FEMALE)
```

```
ageGenderInflModel = lm(TOTCHG ~ AGE + FEMALE + RACE, data = HospitalCosts)
```

```
summary(ageGenderInflModel)
```

OUTPUT



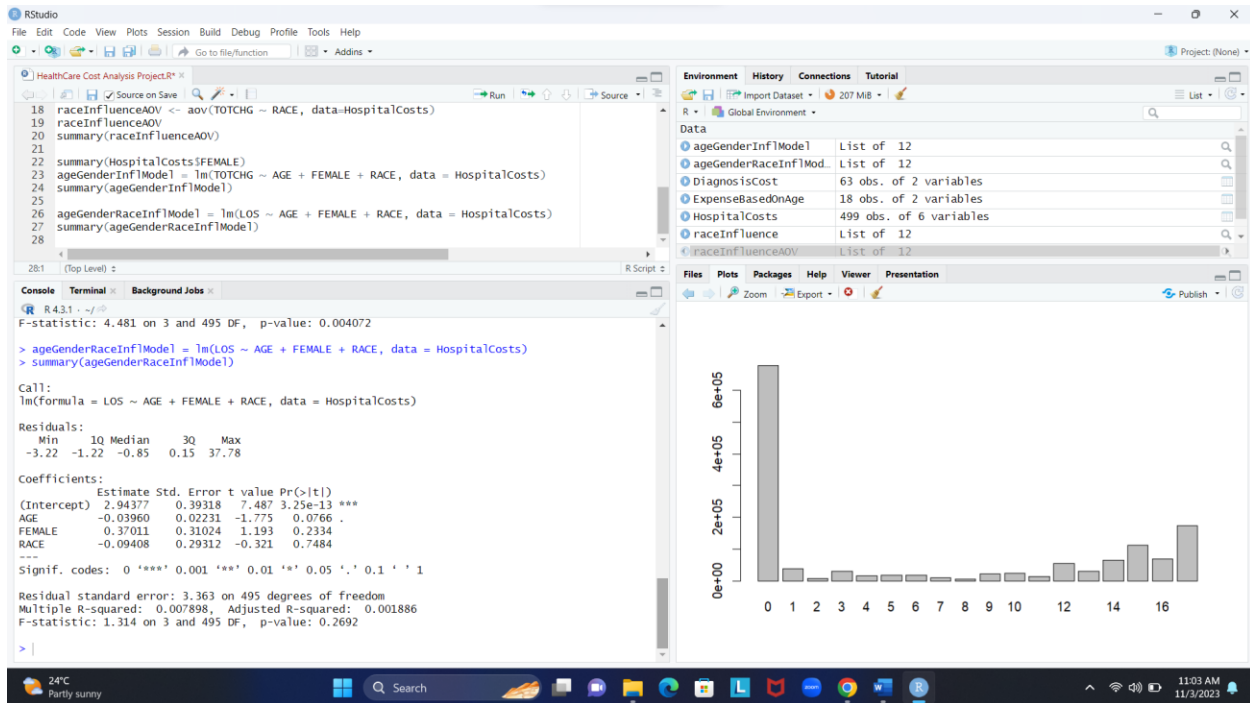
5. Since the length of stay is the crucial factor for inpatients, the agency wants to find if the length of stay can be predicted from age, gender, and race.

CODE

```
ageGenderRaceInflModel = lm(LOS ~ AGE + FEMALE + RACE, data = HospitalCosts)
```

```
summary(ageGenderRaceInflModel)
```

OUTPUT



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

CODE

```
HospitalCostModel = lm(TOTCHG ~ AGE + FEMALE + LOS + APRDRG, data = HospitalCosts)
```

```
summary(HospitalCostModel)
```

```
hcm1 = lm(TOTCHG ~ AGE + FEMALE + LOS + APRDRG, data = HospitalCosts)
```

```
summary(hcm1)
```

```
hcm2 = lm(TOTCHG ~ AGE + LOS + APRDRG, data = HospitalCosts)
```

```
summary(hcm2)
```

```
hcm3 = lm(TOTCHG ~ AGE + LOS, data = HospitalCosts)
```

```
summary(hcm3)
```


OUTPUT

