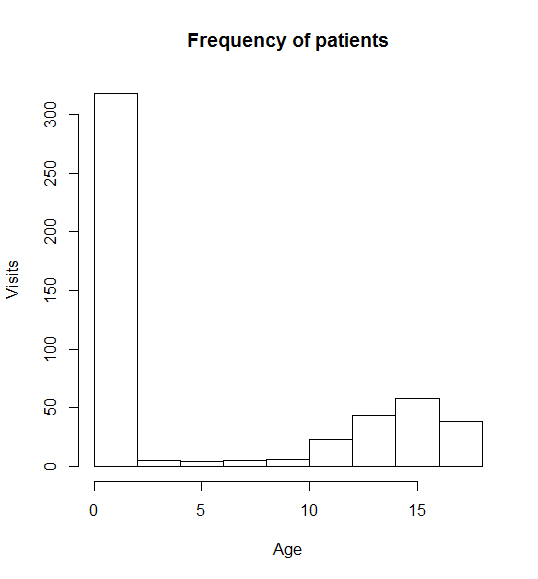
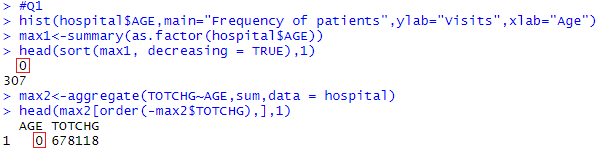
HEALTHCARE COST ANALYSYS

**Domain:** Healthcare

**Solution:**

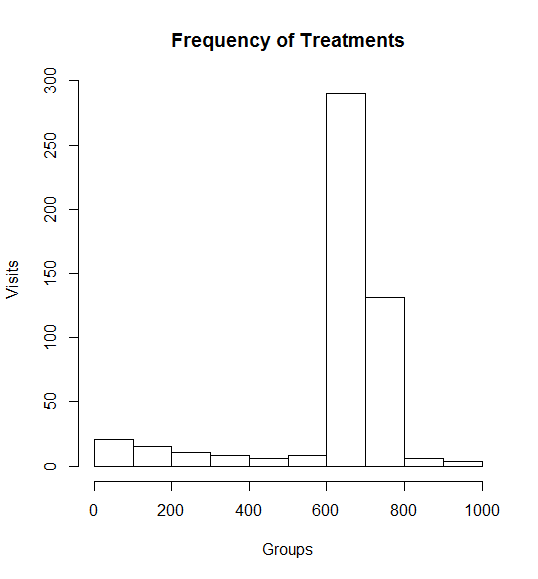
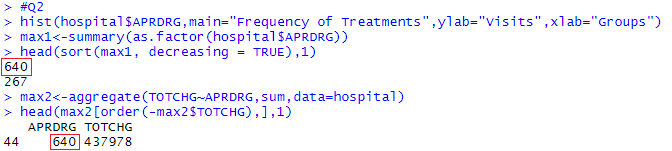
1. To record the patient statistics, the agency wants to find the age category of people who frequent the hospital and has the maximum expenditure.



From the histogram above we can see that children below age 5 have the highest frequency of visit.

The comparison of max of summary() and aggregate() function leads us to the conclusion that children of age **0-1** frequent the hospital and has the max expenditure. So it can be interpreted that visit and cost are directly proportional

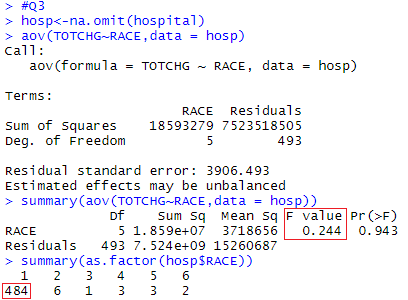
2. In order of severity of the diagnosis and treatments and to find out the expensive treatments, the agency wants to find the diagnosis-related group that has maximum hospitalization and expenditure.



From the histogram above we can see that the diagnostic-related group which falls between 600 - 800 has the highest frequency of visit.

The comparison of max of summary() and aggregate() function leads us to the conclusion that patients falling under diagnostic-related group **640** has maximum hospitalization and expenditure.

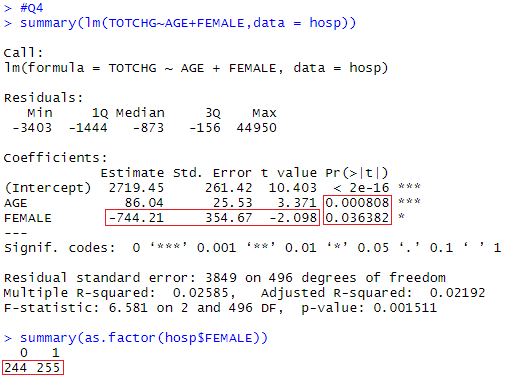
3. To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.



The output of the ANOVA test shows a very low F value implying that variation with respect to race is very less, and a very high P value implying that **cost and race are independent**.

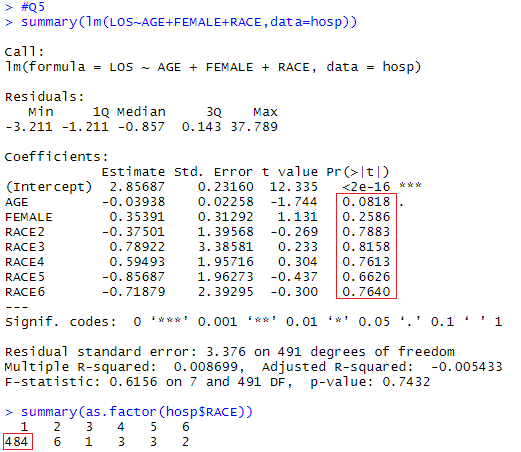
However as majority of the observations belongs to RACE – 1 this prediction may not be accurate.

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



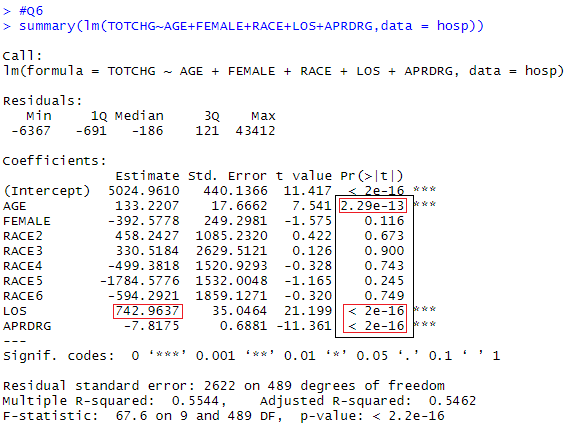
If we compare the P values obtained from the Linear Regression Model we can say that age has more weightage than gender. We can also see that there are almost equal males and females, and the coefficient of female being negative means that the **cost for hospitalization for females is less than males**.

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.



Since all the P values of the independent variables are high there exists no linear relationship among them, therefore we are **unable to predict length of stay from age, gender, and race**.

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



By looking at the P values we can see that **age, length of stay and diagnostic-related group affect the hospital costs**, further we can see the positive coefficient of length of stay implying that each increase in LOS increases the TOTCHG by 742 in value.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*