```
# Healthcare cost analysis


# To read an excel file
library(readxl)

# Location of excel file
Hospital <- read_excel("hospitalcosts.xlsx")

# data set
View(Hospital)

# To visualize the data
library(dplyr)
library(ggplot2)

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


maxTotalCharge = max(Hospital$TOTCHG)
dfAge = dplyr::filter(Hospital,TOTCHG == maxTotalCharge)
maximumHospitalChargeAge = dfAge["AGE"]
dfAge
maximumHospitalChargeAge


#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.


dfDiagnosisGroup = dplyr::filter(Hospital,TOTCHG ==
maxTotalCharge)
dfDiagnosisGroupMax = dfDiagnosisGroup["APRDRG"]

dfDiagnosisGroupMax

#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.
```

```
data <- table(Hospital$RACE,Hospital$TOTCHG)
plot(Hospital$TOTCHG,Hospital$RACE,type="l",col="red")
par(TRUE)
lines(Hospital$TOTCHG,Hospital$RACE,col="green")

#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.


predHospitalStayData <- lm(formula = LOS ~ AGE + FEMALE + RACE,
data = Hospital)
summary(predHospitalStayData)

#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.


predHospitalCharges <- lm(formula = TOTCHG ~ .,data = Hospital)
summary(predHospitalCharges)

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

# Hospital Charges not depend on Gender and Race(because
pr(>|t|)>0.05),
# highly dependable on Age, Length of hospital stay,
# All Patient Refined Diagnosis Related Groups (because
pr(>|t|)<0.05)
```