library(xlsx)  
library(cluster)  
library(ggplot2)  
  
  
abc<-read.xlsx("C:/Users/Priya/Desktop/5th sem/DataSet/heart\_disease\_male.xls",1,header = TRUE)  
  
head(abc) #gives the first 6 values of a dataset  
tail(abc)  
str(abc)  
  
summary(abc)  
  
#Remove the outlier "?" in rest\_electro  
abc<-abc[!(abc$rest\_electro=='?'),]  
  
summary(abc)  
str(abc)  
  
#dotplot  
ggplot(abc,aes(age,fill=disease))+geom\_dotplot(binwidth = 0.80)  
ggplot(abc,aes(rest\_bpress,fill=disease))+geom\_dotplot(binwidth = 2.00)  
ggplot(abc,aes(max\_heart\_rate,fill=disease))+geom\_dotplot(binwidth=1.50)  
ggplot(abc,aes(blood\_sugar,fill=disease))+geom\_dotplot(binwidth=0.01)  
ggplot(abc,aes(age,fill=chest\_pain))+geom\_dotplot(binwidth=0.90)  
ggplot(abc,aes(chest\_pain,fill=disease))+geom\_dotplot(binwidth=0.050)  
ggplot(abc,aes(rest\_electro,fill=disease))+geom\_dotplot(binwidth=0.0250)  
  
  
#plot using points  
ggplot(abc,aes(age,max\_heart\_rate,color=disease))+geom\_point()  
ggplot(abc,aes(age,exercice\_angina,color=disease))+geom\_point()  
ggplot(abc,aes(age,max\_heart\_rate,color=chest\_pain))+geom\_point()  
ggplot(abc,aes(max\_heart\_rate,chest\_pain,color=disease))+geom\_point()  
  
  
#change categorical values into numerical  
abc$chest\_pain<-sapply(abc$chest\_pain,function(x) ifelse(x=="asympt",1,x))  
abc$chest\_pain<-sapply(abc$chest\_pain,function(x) ifelse(x=="atyp\_angina",2,x))  
abc$chest\_pain<-sapply(abc$chest\_pain,function(x) ifelse(x=="non\_anginal",3,x))  
abc$chest\_pain<-sapply(abc$chest\_pain,function(x) ifelse(x=="typ\_angina",4,x))  
  
  
abc$disease<-sapply(abc$disease,function(x) ifelse(x=="positive",2,1))  
  
  
abc$blood\_sugar<-sapply(abc$blood\_sugar,function(x) ifelse(x=="t",2,1))  
  
abc$rest\_electro<-sapply(abc$rest\_electro,function(x) ifelse(x=="left\_vent\_hyper",1,x))  
abc$rest\_electro<-sapply(abc$rest\_electro,function(x) ifelse(x=="st\_t\_wave\_abnormality",2,x))  
abc$rest\_electro<-sapply(abc$rest\_electro,function(x) ifelse(x=="normal",3,x))  
  
  
abc$exercice\_angina<-sapply(abc$exercice\_angina,function(x) ifelse(x=="yes",2,x))  
abc$exercice\_angina<-sapply(abc$exercice\_angina,function(x) ifelse(x=="no",1,x))  
  
  
cor(abc$age,abc$max\_heart\_rate)  
#cor(abc[c("rest\_bpress","disease")])  
cor(abc[c("age","disease")])  
#cor(abc[c("rest\_bpress","blood\_sugar")])  
cor(abc[c("max\_heart\_rate","disease")])  
cor(abc[c("max\_heart\_rate","blood\_sugar")])  
cor(abc[c("max\_heart\_rate","rest\_electro")])  
  
#to scale values   
#abc<-scale(abc)  
dataFrame<-as.data.frame(abc)  
dataFrame  
#for age and chestpain  
pamx<-pam(dataFrame[c(2,8)],2)  
pamx  
pamx$medoids  
inf1\_a\_c<-pamx$clustering  
  
  
abc<-cbind(abc,inf1\_a\_c)  
View(abc)  
  
clusplot(pamx)  
   
#age and heart disease   
pamx1<-pam(dataFrame[c(1,8)],2)  
pamx1  
pamx1$medoids  
inf2\_a\_c<-pamx1$clustering  
  
  
abc<-cbind(abc,inf2\_a\_dis)  
clusplot(pamx1)  
  
#rest\_b press and disease  
rest\_dis<-pam(dataFrame[c(3,8)],2)  
rest\_dis  
rest\_dis$medoids  
inf3\_r\_dise<-rest\_dis$clustering  
  
  
abc<-cbind(abc,inf3\_r\_dise)  
clusplot(rest\_dis)  
  
#rest\_b press ,blood sugar and disease  
r\_sug\_dis<-pam(dataFrame[c(3,8)],2)  
r\_sug\_dis  
r\_sug\_dis$medoids  
inf4\_r\_sug\_dis<-r\_sug\_dis$clustering  
  
  
abc<-cbind(abc,inf4\_r\_sug\_dis)  
clusplot(r\_sug\_dis)  
  
#getting heart attack and max\_heart rate are related   
heart<-pam(dataFrame[c(6,8)],2)  
heart  
heart$medoids  
inf<-heart$clustering  
  
  
abc<-cbind(abc,inf)  
clusplot(heart)  
  
#chest pain in dif age groups   
chest<-pam(dataFrame[c(1,2)],4)  
chest  
chest$medoids  
inf1<-chest$clustering  
  
  
abc<-cbind(abc,inf1)  
clusplot(chest)  
#plot(abc,pamx$clustering)  
#ggplot(pamx,aes(age,chest\_pain,color=disease))+geom\_point()  
#clusplot(pam(abc[c("age","disease")],2),xlab="age",ylab="disease",main="Graph")