```r
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))
```

```r
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
```

```r
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")
```