```
stat.desc(Heart_disease_statlog)
                      age          sex           cp     trestbps         chol
nbr.val      2.700000e+02 270.00000000 270.00000000 2.700000e+02 2.700000e+02
nbr.null     0.000000e+00  87.00000000  20.00000000 0.000000e+00 0.000000e+00
nbr.na       0.000000e+00   0.00000000   0.00000000 0.000000e+00 0.000000e+00
min          2.900000e+01   0.00000000   0.00000000 9.400000e+01 1.260000e+02
max          7.700000e+01   1.00000000   3.00000000 2.000000e+02 5.640000e+02
range        4.800000e+01   1.00000000   3.00000000 1.060000e+02 4.380000e+02
sum          1.469700e+04 183.00000000 587.00000000 3.546300e+04 6.740800e+04
median       5.500000e+01   1.00000000   2.00000000 1.300000e+02 2.450000e+02
mean         5.443333e+01   0.67777778   2.17407407 1.313444e+02 2.496593e+02
SE.mean      5.543601e-01   0.02849347   0.05782064 1.087023e+00 3.145524e+00
CI.mean.0.95 1.091436e+00   0.05609856   0.11383854 2.140155e+00 6.192977e+00
var          8.297509e+01   0.21920694   0.90267107 3.190371e+02 2.671467e+03
std.dev      9.109067e+00   0.46819541   0.95009003 1.786161e+01 5.168624e+01
coef.var     1.673435e-01   0.69078011   0.43700904 1.359906e-01 2.070271e-01
                      fbs      restecg      thalach       exang      oldpeak
nbr.val      270.00000000 270.00000000 2.700000e+02 270.0000000 270.00000000
nbr.null     230.00000000 131.00000000 0.000000e+00 181.0000000  85.00000000
nbr.na         0.00000000   0.00000000 0.000000e+00   0.0000000   0.00000000
min            0.00000000   0.00000000 7.100000e+01   0.0000000   0.00000000
max            1.00000000   2.00000000 2.020000e+02   1.0000000   6.20000000
range          1.00000000   2.00000000 1.310000e+02   1.0000000   6.20000000
sum           40.00000000 276.00000000 4.041300e+04  89.0000000 283.50000000
median         0.00000000   2.00000000 1.535000e+02   0.0000000   0.80000000
mean           0.14814815   1.02222222 1.496778e+02   0.3296296   1.05000000
SE.mean        0.02165978   0.06072973 1.409821e+00   0.0286612   0.06969525
CI.mean.0.95   0.04264425   0.11956602 2.775686e+00   0.0564288   0.13721754
var            0.12666942   0.99578686 5.366504e+02   0.2217954   1.31150558
std.dev        0.35590648   0.99789121 2.316572e+01   0.4709516   1.14520984
coef.var       2.40236872   0.97619792 1.547706e-01   1.4287295   1.09067604
                    slope          ca         thal       target
nbr.val      270.00000000 270.0000000 270.00000000 270.00000000
nbr.null     130.00000000 160.0000000   0.00000000 150.00000000
nbr.na         0.00000000   0.0000000   0.00000000   0.00000000
min            0.00000000   0.0000000   1.00000000   0.00000000
max            2.00000000   3.0000000   3.00000000   1.00000000
range          2.00000000   3.0000000   2.00000000   1.00000000
sum          158.00000000 181.0000000 492.00000000 120.00000000
median         1.00000000   0.0000000   1.00000000   0.00000000
mean           0.58518519   0.6703704   1.82222222   0.44444444
SE.mean        0.03739057   0.0574437   0.05837143   0.03029677
CI.mean.0.95   0.07361539   0.1130964   0.11492295   0.05964895
var            0.37747487   0.8909404   0.91995043   0.24783147
std.dev        0.61438984   0.9438964   0.95914047   0.49782675
coef.var       1.04990668   1.4080222   0.52635757   1.12011019
dim(Heart_disease_statlog)
[1] 270  14
```

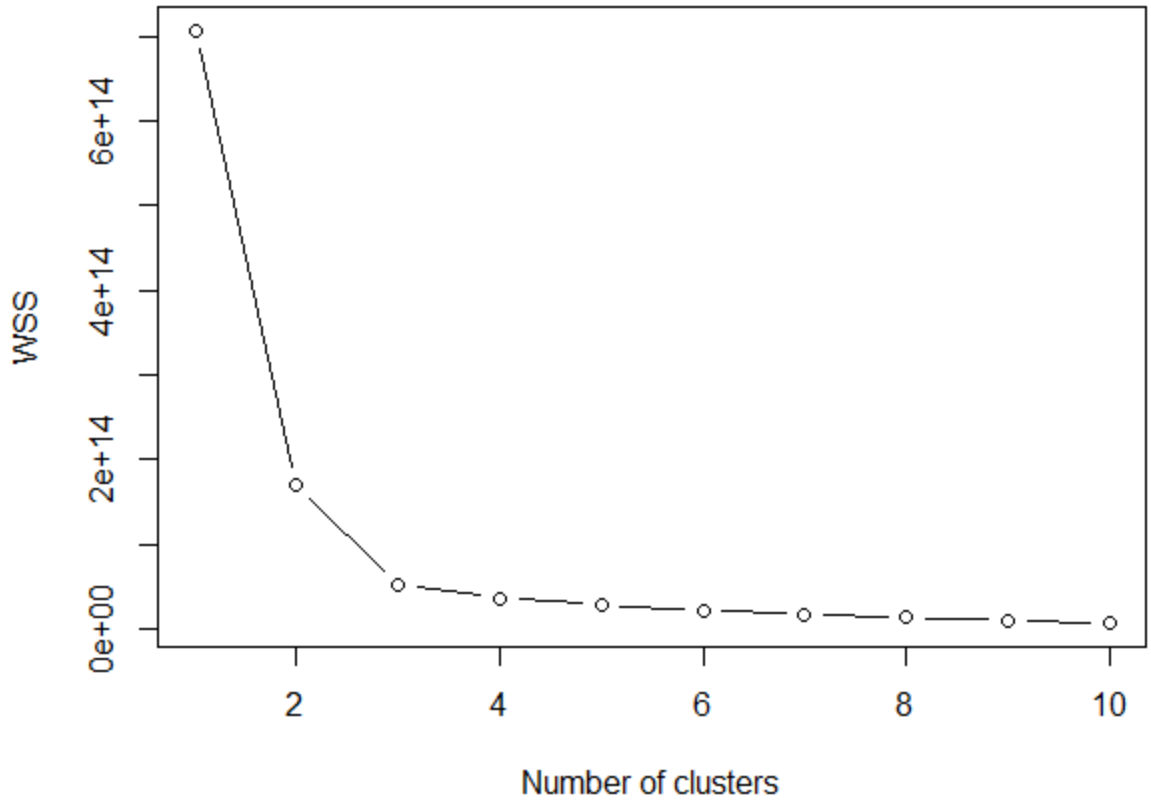
```
glimpse(Heart_disease_statlog)
Rows: 270
Columns: 14
$ age      <int> 70, 67, 57, 64, 74, 65, 56, 59, 60, 63, 59, 53, 44, 61, 57,
71, 4…
$ sex      <int> 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1,
1, 1,…
$ cp       <int> 3, 2, 1, 3, 1, 3, 2, 3, 3, 3, 3, 3, 2, 0, 3, 3, 3, 3, 0, 0,
3, 1,…
$ trestbps <int> 130, 115, 124, 128, 120, 120, 130, 110, 140, 150, 135, 142,
140, …
$ chol     <int> 322, 564, 261, 263, 269, 177, 256, 239, 293, 407, 234, 226,
235, …
$ fbs      <int> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,
0, 0,…
$ restecg  <int> 2, 2, 0, 0, 2, 0, 2, 2, 2, 2, 0, 2, 2, 0, 2, 0, 0, 2, 2, 0,
2, 2,…
$ thalach  <int> 109, 160, 141, 105, 121, 140, 142, 142, 170, 154, 161, 111,
180, …
$ exang    <int> 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 1,
1, 0,…
$ oldpeak  <dbl> 2.4, 1.6, 0.3, 0.2, 0.2, 0.4, 0.6, 1.2, 1.2, 4.0, 0.5, 0.0,
0.0, …
$ slope    <int> 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 2, 1, 0,
1, 1,…
$ ca       <int> 3, 0, 0, 1, 1, 0, 1, 1, 2, 3, 0, 0, 0, 2, 1, 0, 2, 0, 0, 0,
2, 0,…
$ thal     <int> 1, 3, 3, 3, 1, 3, 2, 3, 3, 3, 3, 3, 1, 1, 1, 1, 3, 3, 1, 3,
3, 1,…
$ target   <int> 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0,
1, 0,…
```

```
 >
```

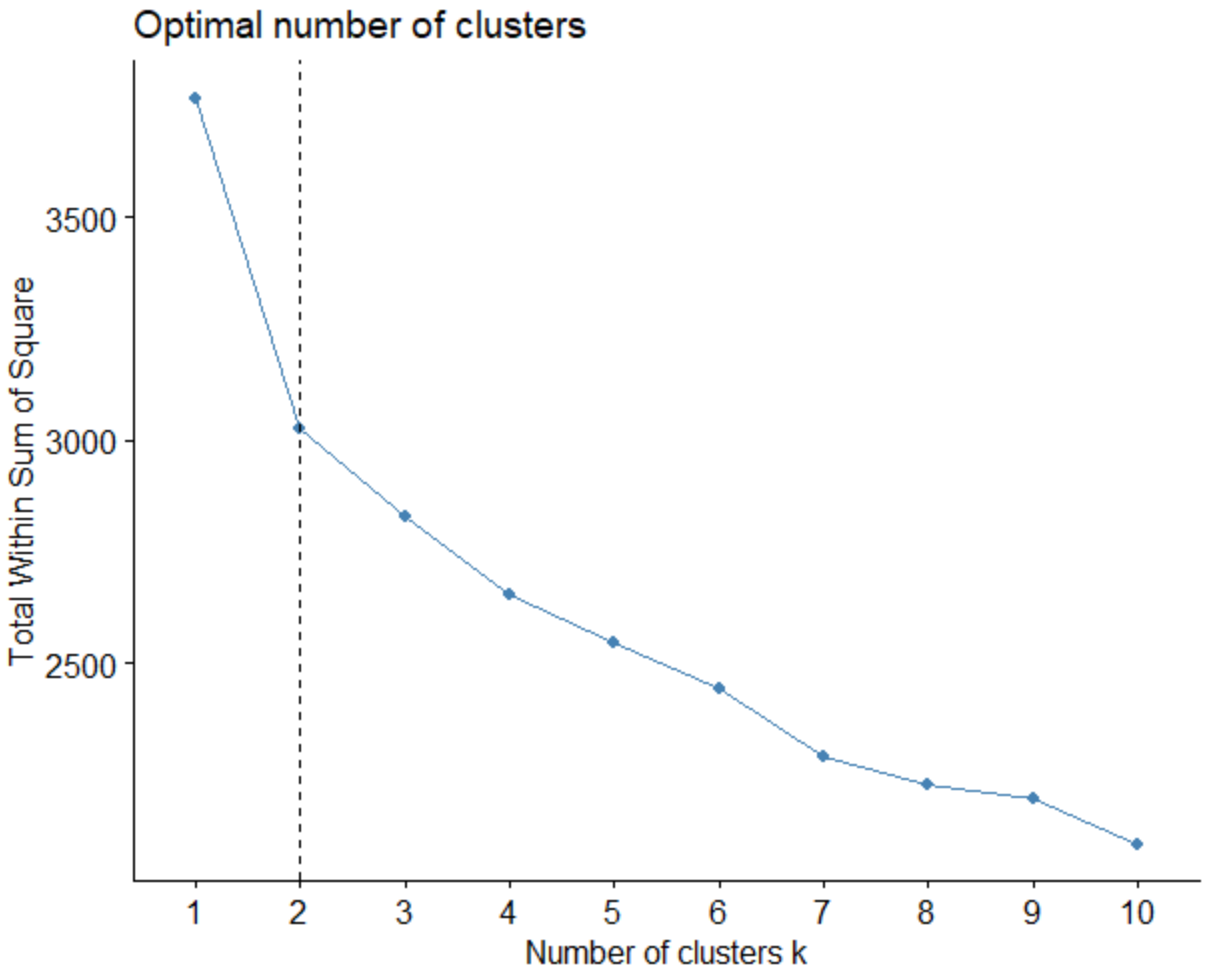
```
str(Heart_disease_statlog)
'data.frame':     270 obs. of  14 variables:
 $ age     : int  70 67 57 64 74 65 56 59 60 63 ...
 $ sex     : int  1 0 1 1 0 1 1 1 1 0 ...
 $ cp      : int  3 2 1 3 1 3 2 3 3 3 ...
 $ trestbps: int  130 115 124 128 120 120 130 110 140 150 ...
 $ chol    : int  322 564 261 263 269 177 256 239 293 407 ...
 $ fbs     : int  0 0 0 0 0 0 1 0 0 0 ...
 $ restecg : int  2 2 0 0 2 0 2 2 2 2 ...
 $ thalach : int  109 160 141 105 121 140 142 142 170 154 ...
```

```
 $ exang   : int  0 0 0 1 1 0 1 1 0 0 ...
 $ oldpeak : num  2.4 1.6 0.3 0.2 0.2 0.4 0.6 1.2 1.2 4 ...
 $ slope   : int  1 1 0 1 0 0 1 1 1 1 ...
 $ ca      : int  3 0 0 1 1 0 1 1 2 3 ...
 $ thal    : int  1 3 3 3 1 3 2 3 3 3 ...
 $ target  : int  1 0 1 0 0 0 1 1 1 1 ...
#column names
> colnames(Heart_disease_statlog)
 [1] "age"      "sex"      "cp"       "trestbps" "chol"     "fbs"
"restecg"
 [8] "thalach"  "exang"    "oldpeak"  "slope"    "ca"       "thal"     "target"
> #change column names
> colnames(Heart_disease_statlog) <- c("age", "sex", "chest_pain_type",
"resting_blood_pressure", "cholesterol", "fasting_blood_sugar", "rest_ecg",
"max_heart_rate_achieved",
+                                    "exercise_induced_angina",
"st_depression", "st_slope","major_vessels","thalassemia","target")
> colnames(Heart_disease_statlog)
 [1] "age"                     "sex"                     "chest_pain_type"
 [4] "resting_blood_pressure"  "cholesterol"             "fasting_blood_sugar"
 [7] "rest_ecg"                "max_heart_rate_achieved"
"exercise_induced_angina"
[10] "st_depression"           "st_slope"                "major_vessels"
[13] "thalassemia"             "target"
```

# Plot WSS against the number of clusters
plot(1:10, wss, type = "b", xlab = "Number of clusters", ylab = "WSS")




```
#No. Clusters
> fviz_nbclust(df, kmeans, method = "wss") +
+  geom_vline(xintercept =2, linetype = 2)
```

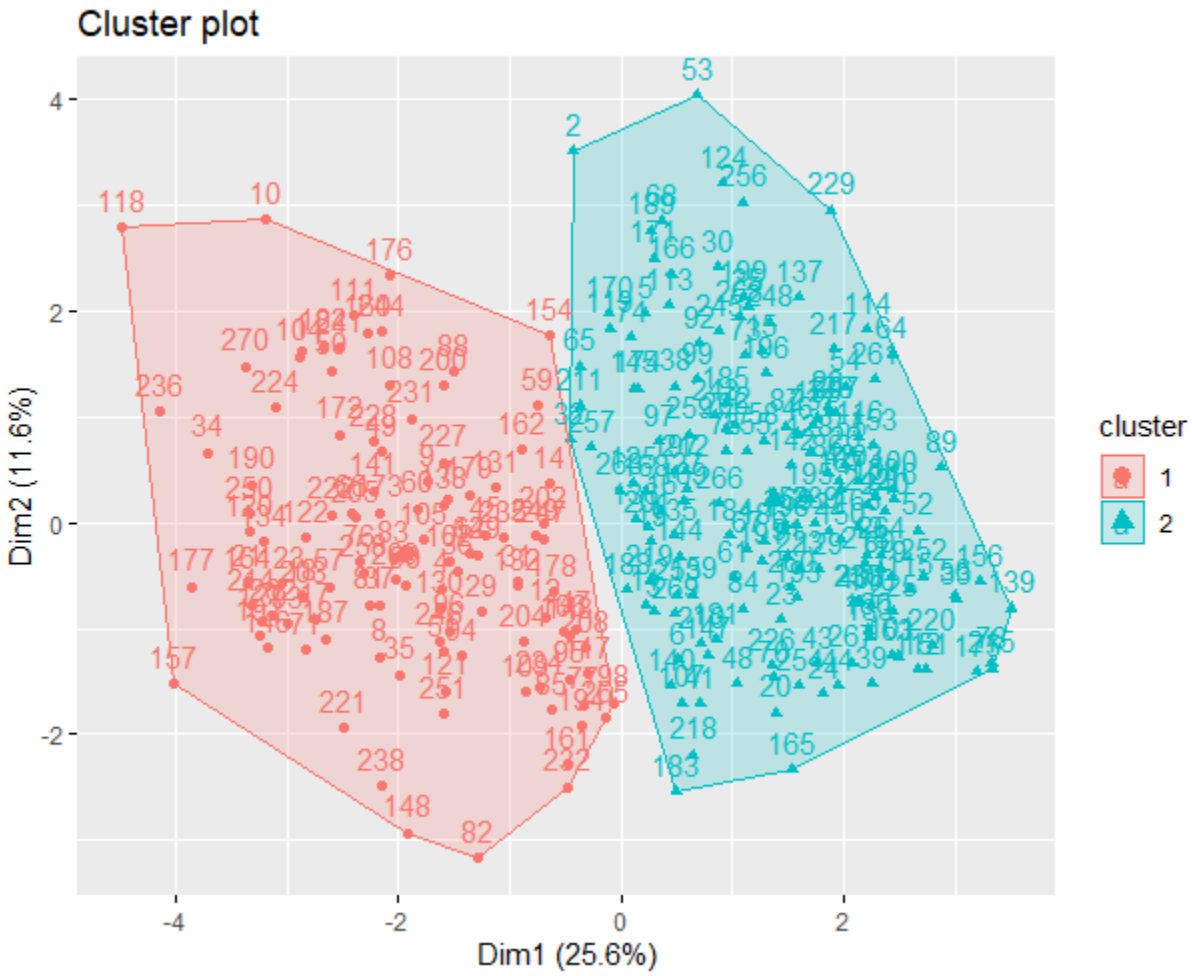



```
#Clustering
> #Compute k-means with k = 2
> set.seed(123)
> km.res <- kmeans(df, 2, nstart = 25)
> print(km.res)
K-means clustering with 2 clusters of sizes 112, 158

Cluster means:
         age        sex chest_pain_type resting_blood_pressure cholesterol
1  0.2876517  0.3449584       0.5685899             0.13867643  0.07620897
2 -0.2039050 -0.2445275      -0.4030511            -0.09830228 -0.05402155
  fasting_blood_sugar    rest_ecg max_heart_rate_achieved
exercise_induced_angina
1         -0.03995313  0.1656271              -0.6239636
0.6461363
2          0.02832120 -0.1174065               0.4423033
-0.4580207
  st_depression   st_slope major_vessels thalassemia     target
1     0.6400886  0.5443737     0.5857040   0.6973423  0.9724809
2    -0.4537337 -0.3858852    -0.4151826  -0.4943186 -0.6893535
```

```
Clustering vector:
  [1] 1 2 2 1 2 2 1 1 1 1 2 1 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 1 1 1 1
2 2
 [40] 2 2 2 2 2 1 2 1 2 1 1 1 2 2 2 2 2 1 2 1 1 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1
1 2
 [79] 2 2 1 1 1 2 1 2 2 1 2 1 2 2 1 1 1 1 2 1 2 2 2 1 2 1 1 1 2 1 1 2 1 2 2 2 2
2 1
[118] 1 2 1 1 1 1 2 2 2 1 2 2 1 1 1 2 1 2 2 2 1 2 2 1 2 1 2 2 1 2 1 1 2 2 2 2 1
2 2
[157] 1 2 2 1 1 1 2 1 2 2 2 2 2 2 2 1 1 2 2 1 1 1 1 2 2 1 2 2 2 2 1 2 2 1 2 1 2
1 2
[196] 2 2 2 2 1 2 1 1 1 1 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 1 2 2 1 1 2 2 1 1
2 1
[235] 1 1 2 1 2 2 1 2 2 1 2 1 1 2 1 1 1 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 1

Within cluster sum of squares by cluster:
[1] 1306.682 1719.196
 (between_SS / total_SS =  19.7 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
aggregate(df, by=list(cluster=km.res$cluster), mean)
  cluster        age        sex chest_pain_type resting_blood_pressure
cholesterol
1       1  0.2876517  0.3449584       0.5685899             0.13867643
0.07620897
2       2 -0.2039050 -0.2445275      -0.4030511            -0.09830228
-0.05402155
  fasting_blood_sugar   rest_ecg max_heart_rate_achieved
exercise_induced_angina
1         -0.03995313  0.1656271              -0.6239636
0.6461363
2          0.02832120 -0.1174065               0.4423033
-0.4580207
  st_depression   st_slope major_vessels thalassemia     target
1     0.6400886  0.5443737     0.5857040   0.6973423  0.9724809
2    -0.4537337 -0.3858852    -0.4151826  -0.4943186 -0.6893535
head(df)
           age        sex chest_pain_type resting_blood_pressure cholesterol
[1,] 1.7089201  0.6882217       0.8693133            -0.07527007   1.3996132
[2,] 1.3795779 -1.4476387      -0.1832185            -0.91506006   6.0817107
[3,] 0.2817705  0.6882217      -1.2357503            -0.41118607   0.2194151
[4,] 1.0502357  0.6882217       0.8693133            -0.18724207   0.2581101
[5,] 2.1480430 -1.4476387      -1.2357503            -0.63513007   0.3741952
[6,] 1.1600164  0.6882217       0.8693133            -0.63513007  -1.4057758
```

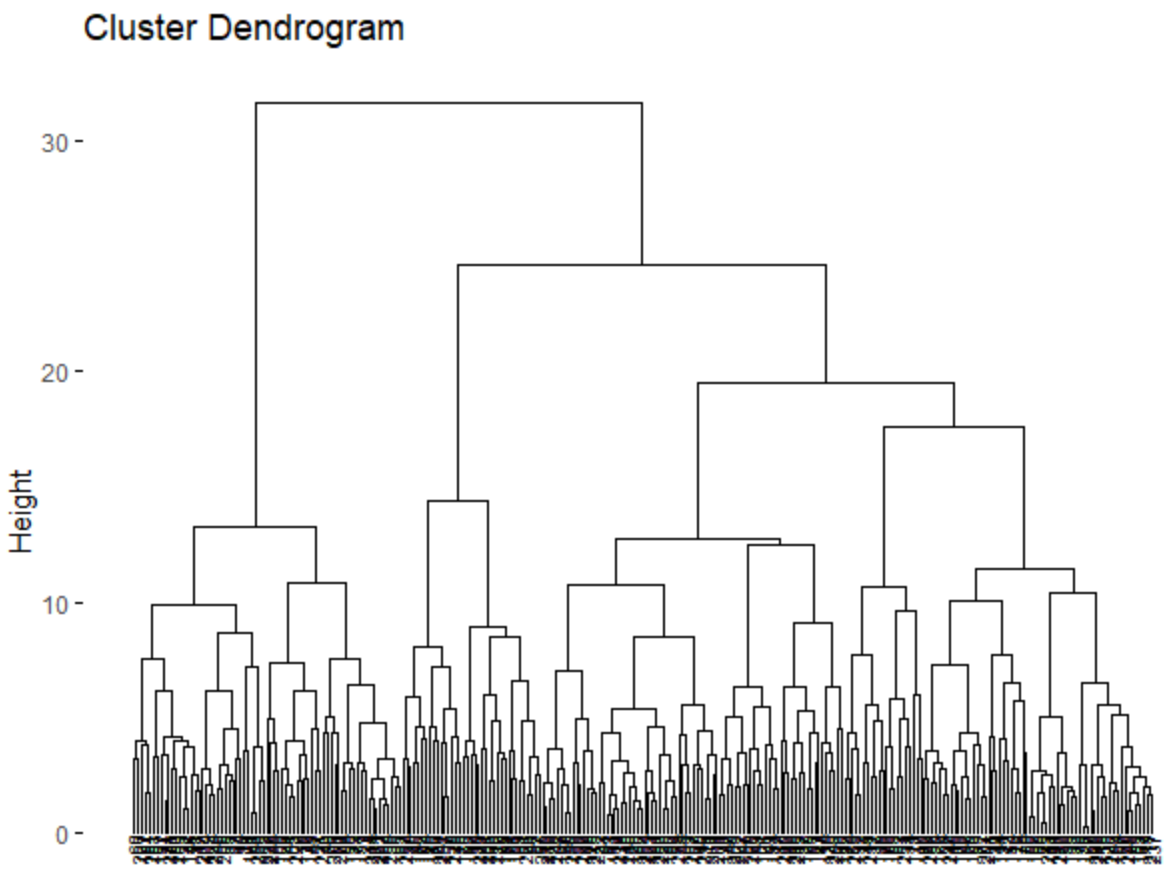
```
     fasting_blood_sugar   rest_ecg max_heart_rate_achieved
exercise_induced_angina
[1,]          -0.4162558  0.9798441              -1.7559473
-0.6999225
[2,]          -0.4162558  0.9798441               0.4455818
-0.6999225
[3,]          -0.4162558 -1.0243824              -0.3745957
-0.6999225
[4,]          -0.4162558 -1.0243824              -1.9286162
1.4234380
[5,]          -0.4162558  0.9798441              -1.2379404
1.4234380
[6,]          -0.4162558 -1.0243824              -0.4177629
-0.6999225
     st_depression   st_slope major_vessels thalassemia     target
[1,]     1.1788233  0.6751655     2.4680989   -0.857249  1.1159616
[2,]     0.4802613  0.6751655    -0.7102161    1.227951 -0.8927693
[3,]    -0.6549018 -0.9524656    -0.7102161    1.227951  1.1159616
[4,]    -0.7422221  0.6751655     0.3492223    1.227951 -0.8927693
[5,]    -0.7422221 -0.9524656     0.3492223   -0.857249 -0.8927693
[6,]    -0.5675816 -0.9524656    -0.7102161    1.227951 -0.8927693
km.res$cluster
  [1] 1 2 2 1 2 2 1 1 1 1 2 1 2 1 2 2 1 1 2 2 1 2 2 2 2 2 2 2 1 2 1 2 2 1 1 1 1
2 2
 [40] 2 2 2 2 2 1 2 1 2 1 1 1 2 2 2 2 2 1 2 1 1 2 1 2 2 2 1 2 2 2 2 1 2 2 2 2 1
1 2
 [79] 2 2 1 1 1 2 1 2 2 1 2 1 2 2 1 1 1 1 2 1 2 2 2 1 2 1 1 1 2 1 1 2 1 2 2 2 2
2 1
[118] 1 2 1 1 1 1 2 2 2 1 2 2 1 1 1 2 1 2 2 2 1 2 2 1 2 1 2 2 1 2 1 1 2 2 2 2 1
2 2
[157] 1 2 2 1 1 1 2 1 2 2 2 2 2 2 2 1 1 2 2 1 1 1 1 2 2 1 2 2 2 2 1 2 2 1 2 1 2
1 2
[196] 2 2 2 2 1 2 1 1 1 1 2 2 1 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 1 2 2 1 1 2 2 1 1
2 1
[235] 1 1 2 1 2 2 1 2 2 1 2 1 1 2 1 1 1 2 2 2 2 2 2 1 2 2 2 1 2 2 1 2 2 2 2 2 1
> km.res$size
[1] 112 158
> km.res$centers
         age        sex chest_pain_type resting_blood_pressure cholesterol
1  0.2876517  0.3449584       0.5685899             0.13867643  0.07620897
2 -0.2039050 -0.2445275      -0.4030511            -0.09830228 -0.05402155
  fasting_blood_sugar   rest_ecg max_heart_rate_achieved
exercise_induced_angina
1         -0.03995313  0.1656271              -0.6239636
0.6461363
2          0.02832120 -0.1174065               0.4423033
-0.4580207
  st_depression   st_slope major_vessels thalassemia     target
1     0.6400886  0.5443737     0.5857040   0.6973423  0.9724809
```

```
2    -0.4537337 -0.3858852    -0.4151826  -0.4943186 -0.6893535
fviz_cluster(km.res , data = df)
```



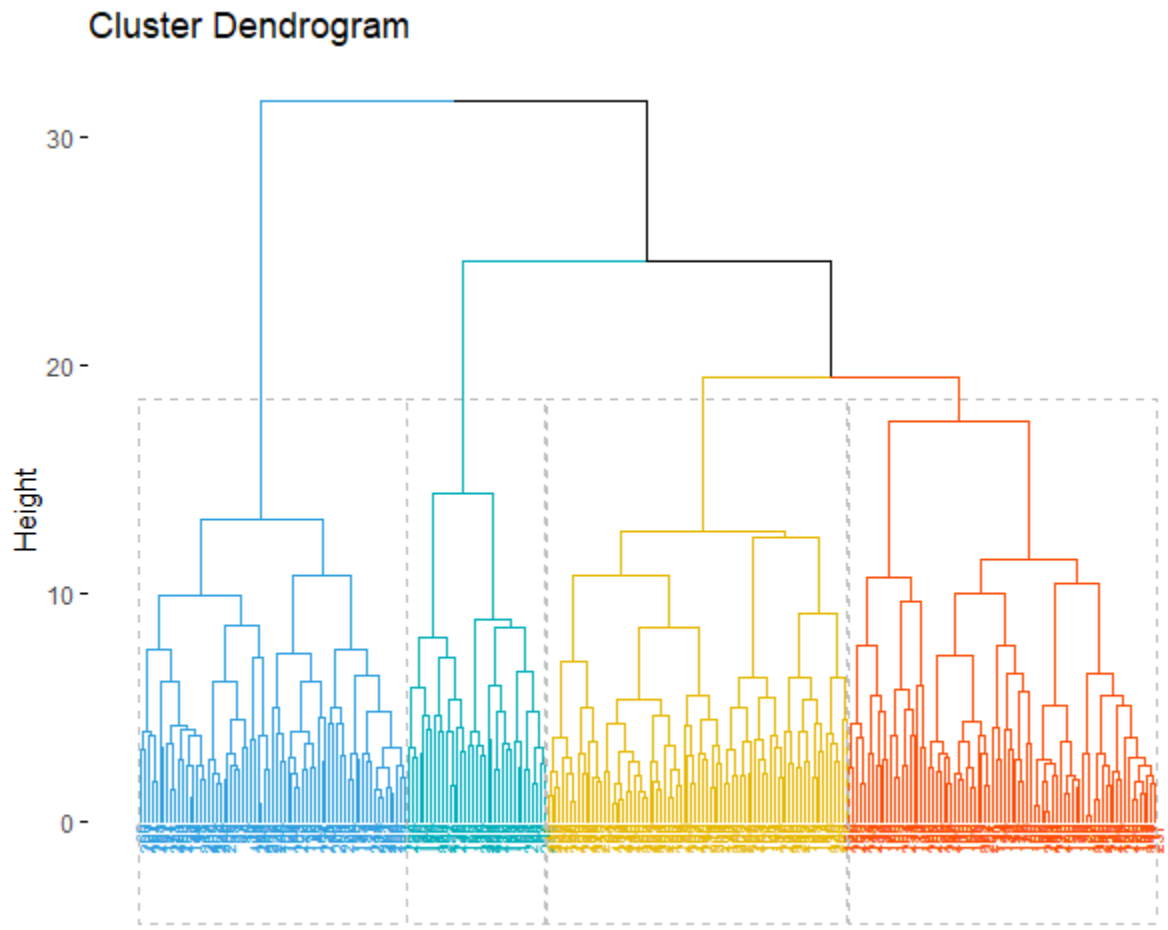

```
fviz_cluster(km.res , data = df)
> fviz_cluster(km.res, data = df,
+              palette = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07"),
+              ellipse.type = "euclid", # Concentration ellipse
+              star.plot = TRUE, # Add segments from centroids to items
+              repel = TRUE, # Avoid label overplotting (slow)
+              ggtheme = theme_minimal()
```


```
df <- scale(hd)
> res.dist <- dist(df, method = "euclidean")
> res.hc <- hclust(d = res.dist, method = "ward.D2")
> fviz_dend(res.hc, cex = 0.5)
```

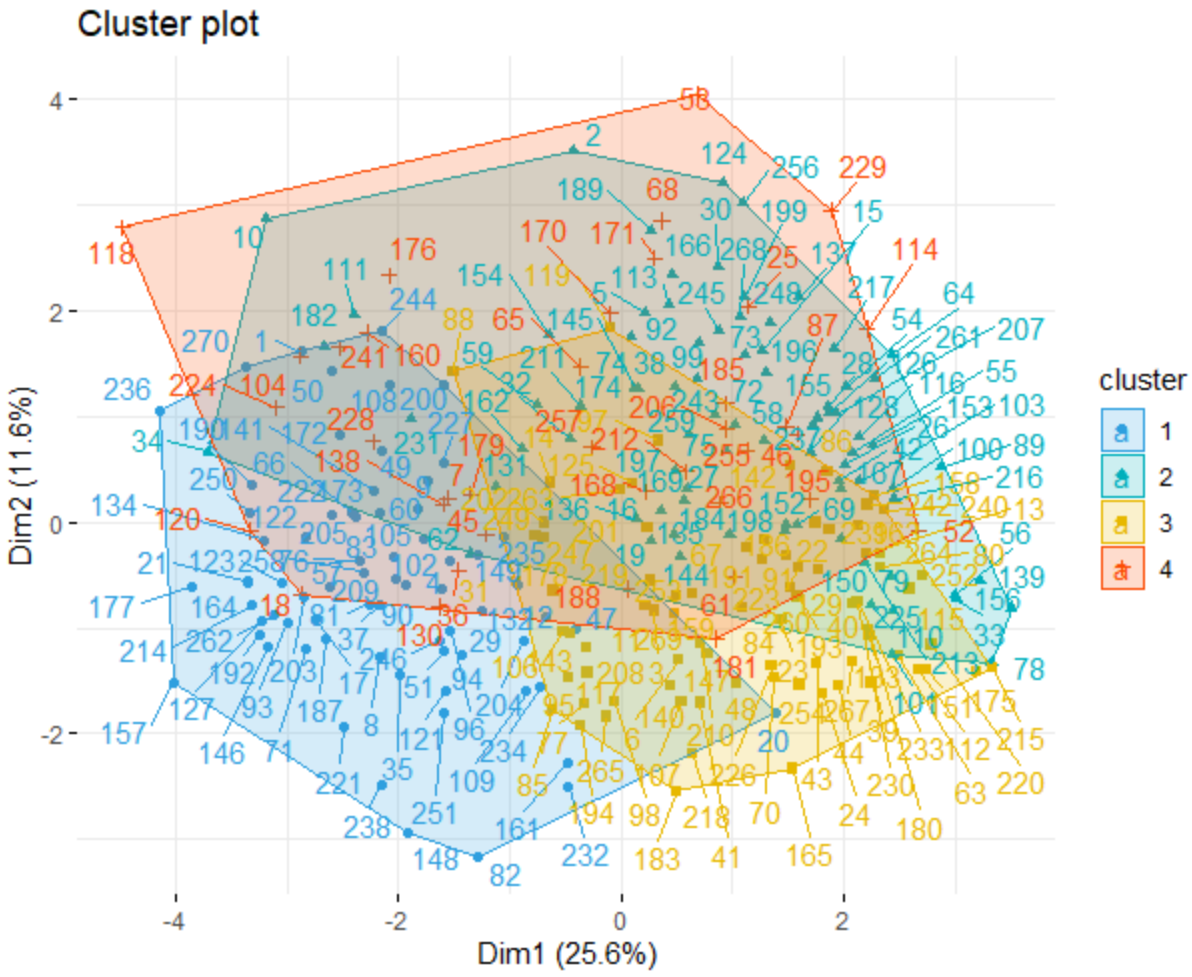



```
# Validation
> res.coph <- cophenetic(res.hc)
> cor(res.dist, res.coph)
[1] 0.4872433
res.coph <- cophenetic(res.hc)
> cor(res.dist, res.coph)
[1] 0.4872433
> res.hc2 <- hclust(res.dist, method = "average")
> cor(res.dist, cophenetic(res.hc2))
[1] 0.6869141
> grp <- cutree(res.hc, k = 4)
> table(grp)
grp
 1  2  3  4
71 82 80 37
> rownames(df)[grp == 1]
NULL
# Customize dendrogram
> fviz_dend(res.hc, k = 4, cex = 0.5, k_colors = c("#2E9FDF", "#00AFBB",
"#E7B800", "#FC4E07"),
```

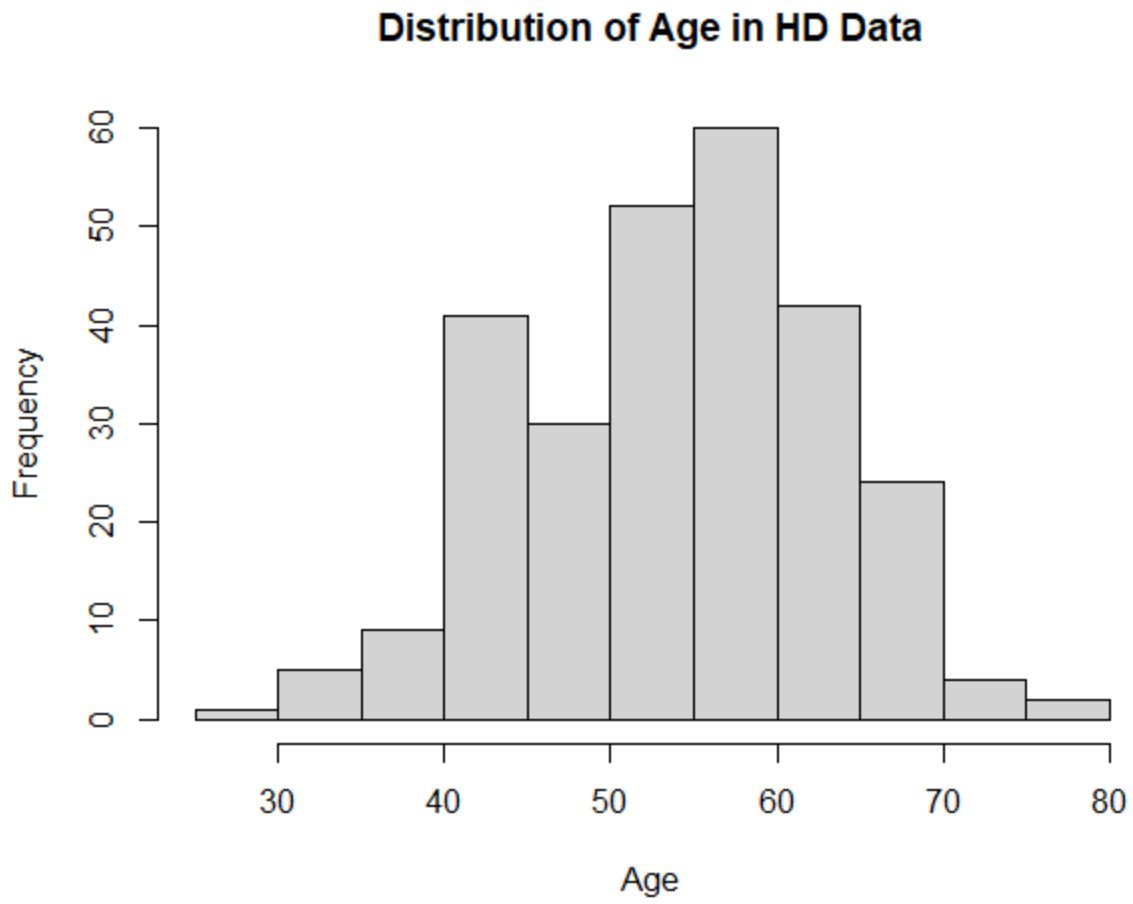
```
+           color_labels_by_k = TRUE, rect = TR
```



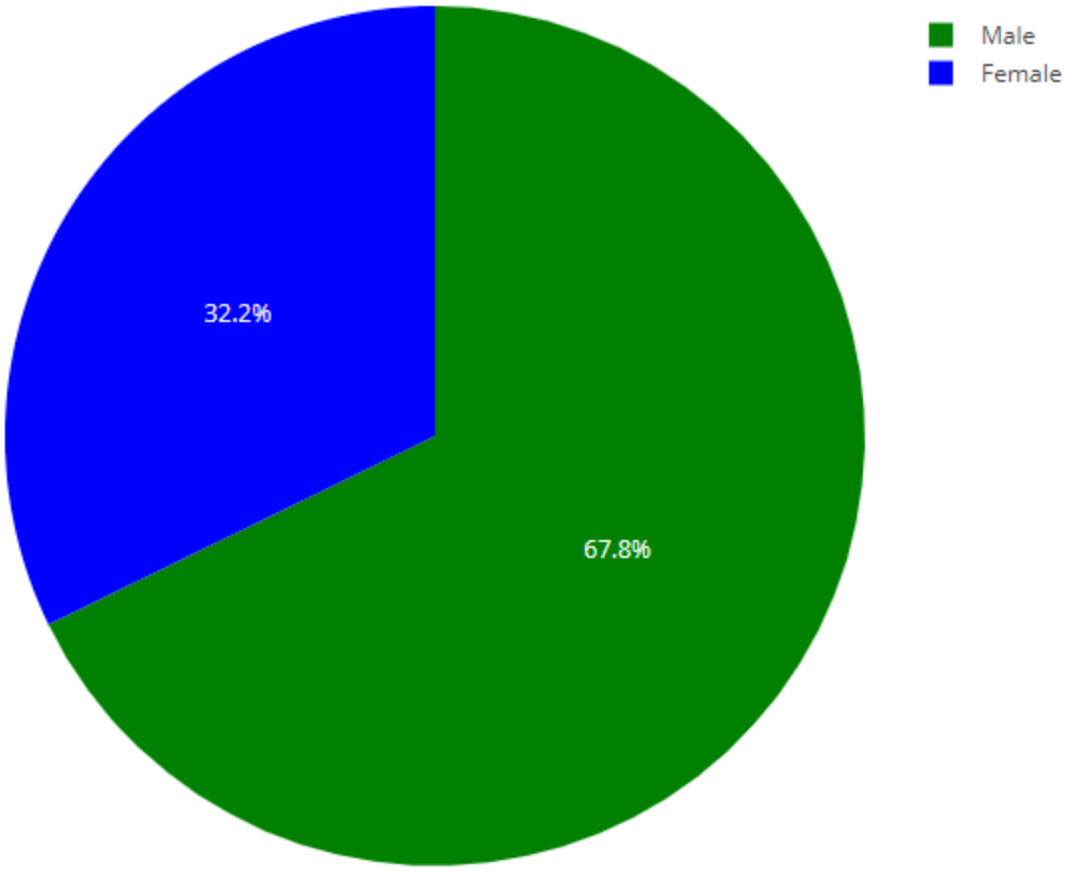

```
# PCA #################
fviz_cluster(list(data = df, cluster = grp),
             palette = c("#2E9FDF", "#00AFBB", "#E7B800", "#FC4E07"),
             ellipse.type = "convex",
             repel = TRUE,
             show.clust.cent = FALSE, ggtheme = theme_minimal())
```



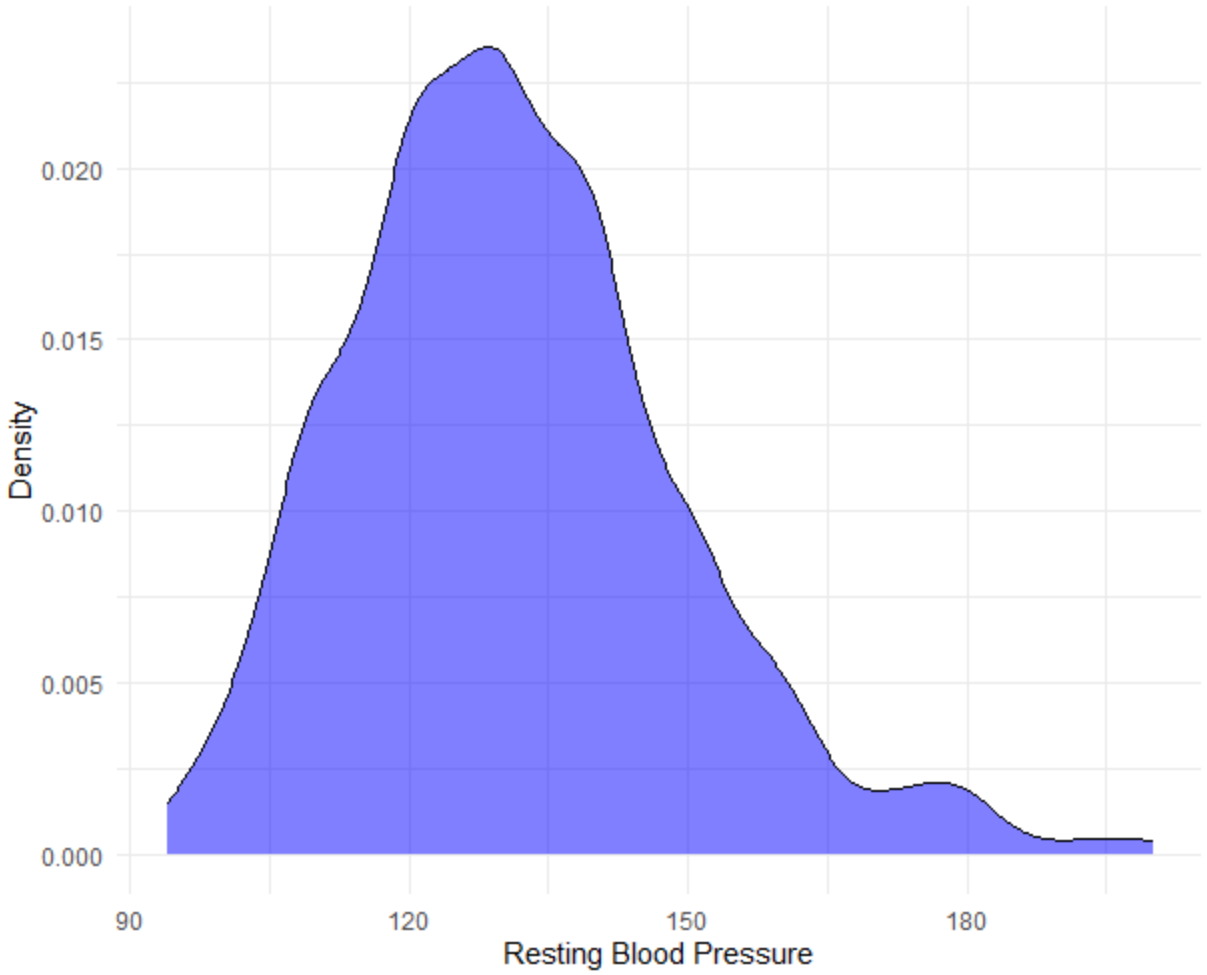
#age
hist(hd$age, xlab = "Age", main = "Distribution of Age in HD Data")



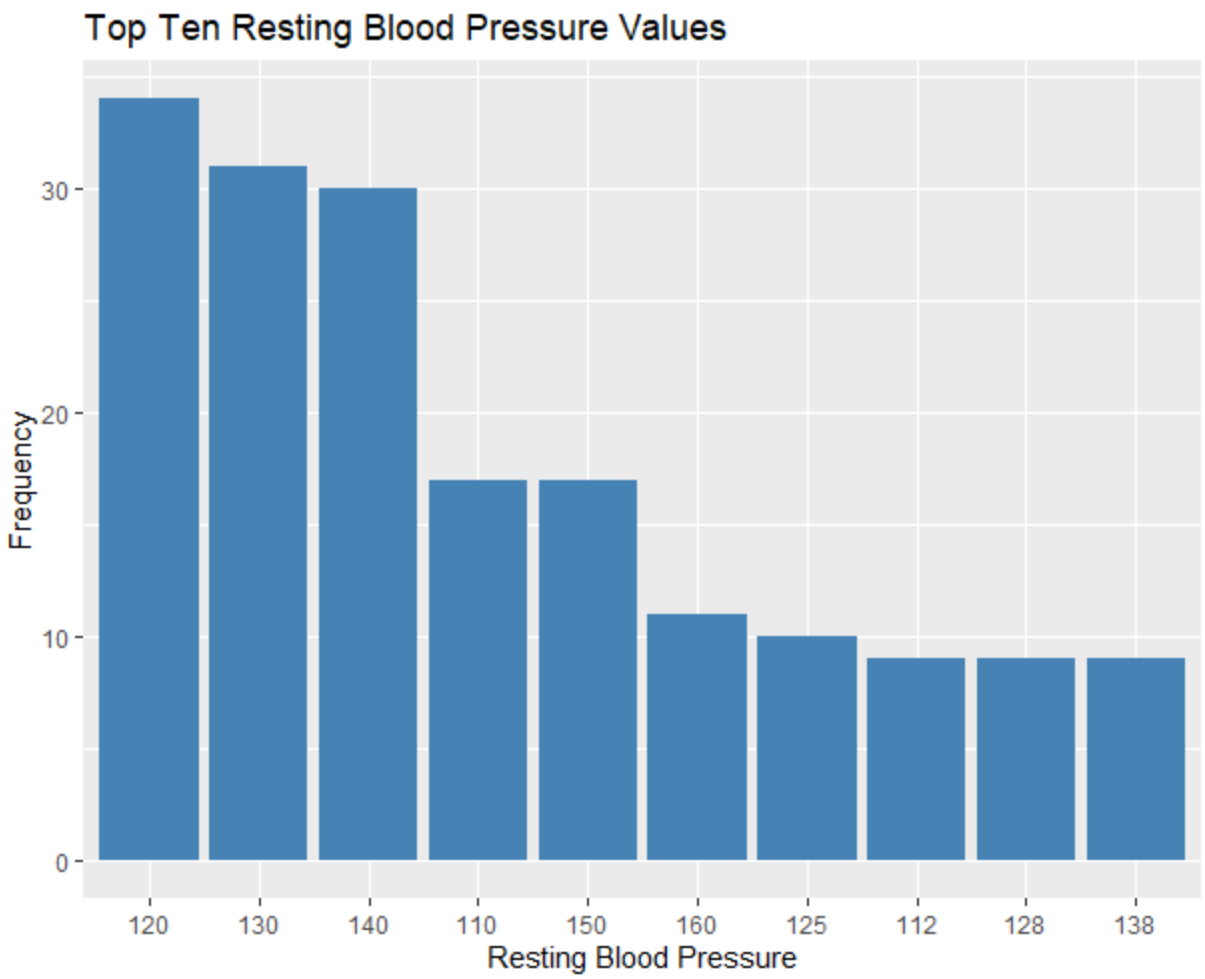

```
male_count <- nrow(hd[hd$sex == 1,])
female_count <- nrow(hd[hd$sex == 0,])
labels <- c('Male', 'Female')
data <- c(male_count, female_count)
colors <- c('green', 'blue')

fig <- plot_ly(labels = labels, values = data, type = 'pie', marker = list(colors = colors))
fig
```



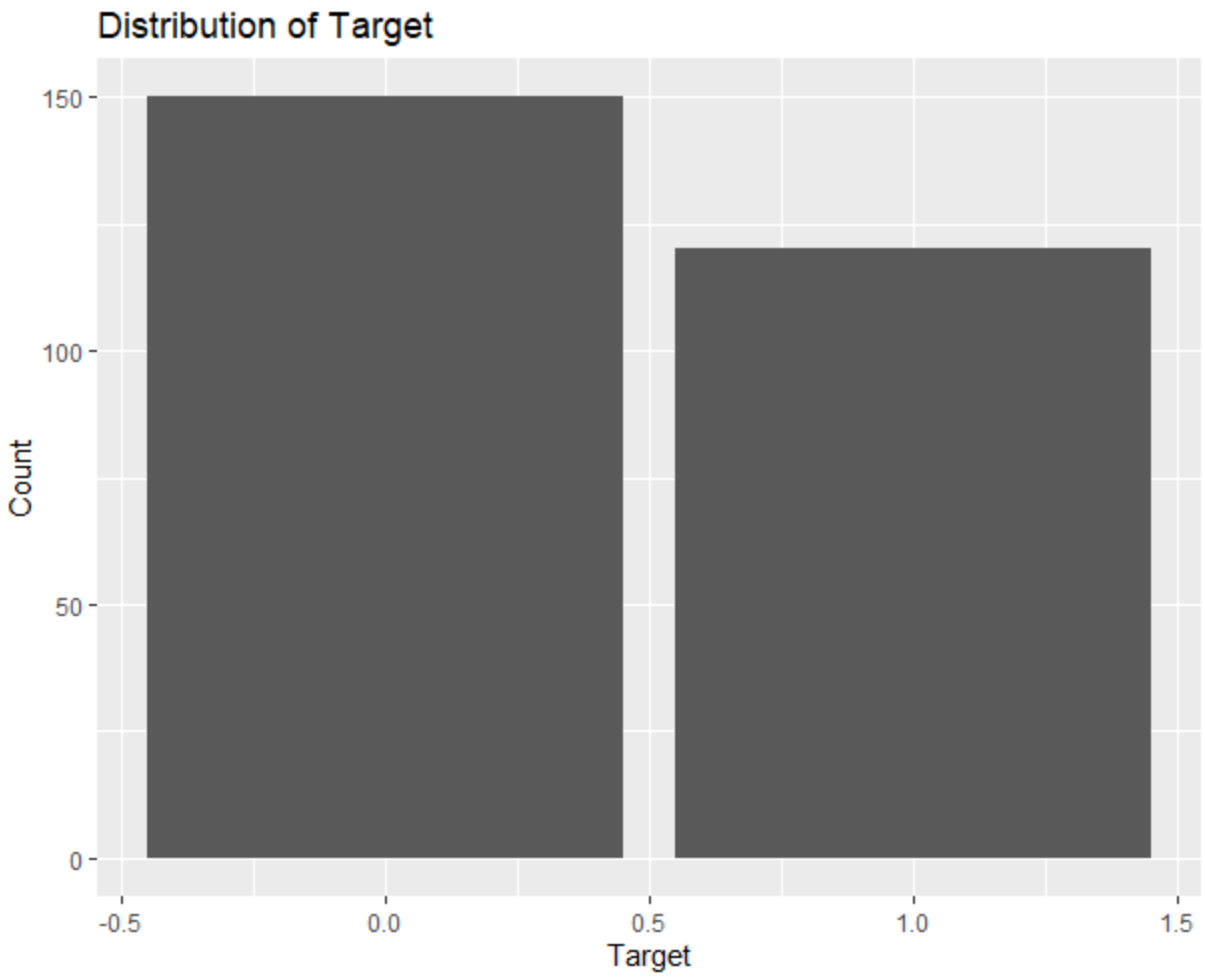

```
ggplot(hd, aes(resting_blood_pressure)) +
  geom_density(fill = "blue", alpha = 0.5) +
  labs(x = "Resting Blood Pressure", y = "Density") +
  theme_minimal()
```



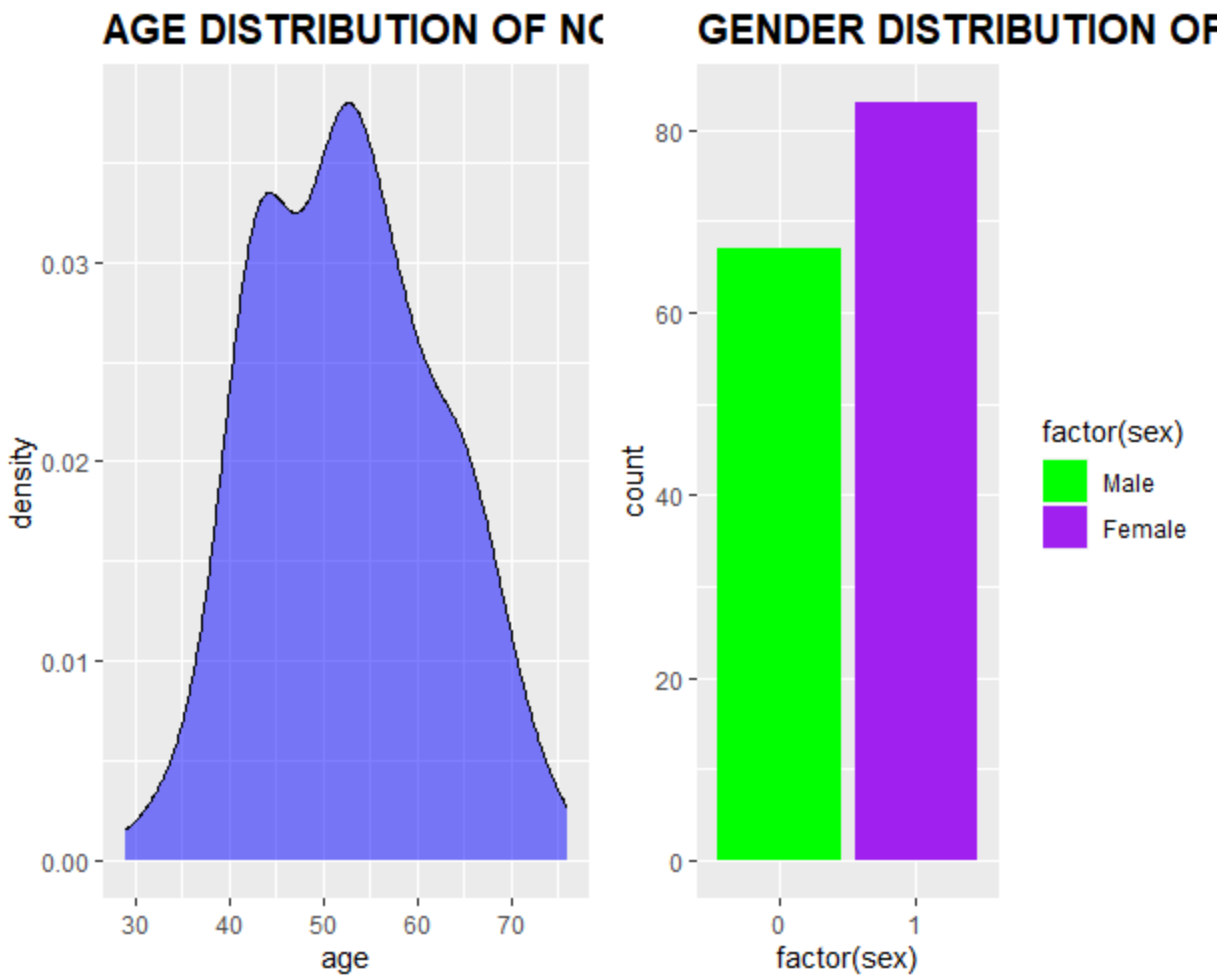

```
rest_blood_press_top10 <- head(sort(table(hd$resting_blood_pressure),
decreasing = TRUE), 10)
> ggplot(data.frame(rest_blood_press_top10), aes(x = Var1, y = Freq)) +
+  geom_bar(stat = "identity", fill = "steelblue") +
+  ggtitle("Top Ten Resting Blood Pressure Values") +
+  xlab("Resting Blood Pressure") +
+  ylab("Frequency")
```



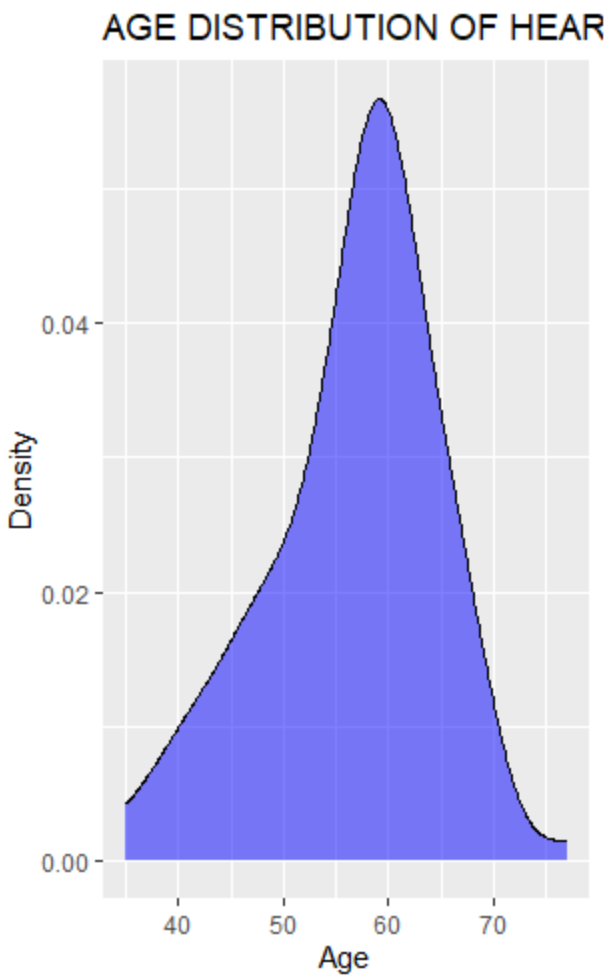

```
#dist of target
ggplot(hd, aes(x = target)) +
  geom_bar() +
  xlab("Target") +
  ylab("Count") +
  ggtitle("Distribution of Target")
```

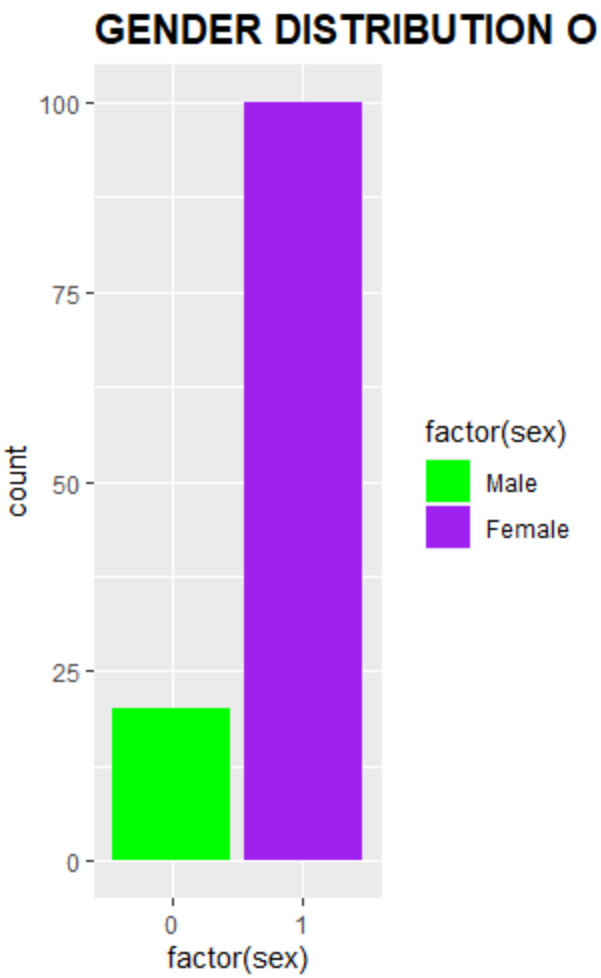



```
# plotting normal patients
fig <- ggplot(df_0, aes(x = age)) +
  geom_density(fill = "blue", alpha = 0.5) +
  ggtitle("AGE DISTRIBUTION OF NORMAL PATIENTS") +
  theme(plot.title = element_text(face = "bold", size = 15))

fig2 <- ggplot(df_0, aes(x = factor(sex), fill = factor(sex))) +
  geom_bar() +
  ggtitle("GENDER DISTRIBUTION OF NORMAL PATIENTS") +
  theme(plot.title = element_text(face = "bold", size = 15)) +
  scale_fill_manual(values = c("green", "purple"), labels = c("Male", "Female"))

gridExtra::grid.arrange(fig, fig2, ncol = 2)
```




```
# Plot age distribution of heart disease patients
fig3 <-ggplot(hd_1, aes(x = age)) +
  geom_density(fill = "blue", alpha = 0.5) +
  ggtitle("AGE DISTRIBUTION OF HEART DISEASE PATIENTS") +
  xlab("Age") +
  ylab("Density")

# Plot gender distribution of heart disease patients
fig4<-ggplot(hd_1, aes(x = factor(sex), fill = factor(sex))) +
  geom_bar() +
  ggtitle("GENDER DISTRIBUTION OF HEART DISEASE PATIENTS") +
  theme(plot.title = element_text(face = "bold", size = 15)) +
  scale_fill_manual(values = c("green", "purple"), labels = c("Male", "Female"))
gridExtra::grid.arrange(fig3, fig4, ncol = 2)
```





#dummy variables creation

```
dummy_variables <- model.matrix(~ chest_pain_type + rest_ecg + st_slope +
thalassemia - 1, data = hd)
> print(dummy_variables)
    chest_pain_typetypical angina chest_pain_typeatypical angina
1                               0                               0
2                               0                               0
3                               0                               1
4                               0                               0
5                               0                               1
6                               0                               0
7                               0                               0
8                               0                               0
9                               0                               0
10                              0                               0
11                              0                               0
12                              0                               0
13                              0                               0
14                              1                               0
```

15	0	0
16	0	0
17	0	0
18	0	0
19	1	0
20	1	0
21	0	0
22	0	1
23	0	0
24	0	0
25	0	1
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	1	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	1
45	0	0
46	0	0
47	0	0
48	0	0
49	0	1
50	0	0
51	0	0
52	0	1
53	0	0
54	0	1
55	0	1
56	0	1
57	0	0
58	0	0
59	0	0
60	0	1
61	0	0
62	0	0
63	0	0

64	1	0
65	1	0
66	0	0
67	0	0
68	0	1
69	0	0
70	0	0
71	0	0
72	0	0
73	0	1
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	1
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	1	0
87	0	1
88	1	0
89	0	1
90	0	0

	chest_pain_typenon-angina pain	chest_pain_typeasymptomatic
1	0	1
2	1	0
3	0	0
4	0	1
5	0	0
6	0	1
7	1	0
8	0	1
9	0	1
10	0	1
11	0	1
12	0	1
13	1	0
14	0	0
15	0	1
16	0	1
17	0	1
18	0	1
19	0	0
20	0	0
21	0	1

22	0	0
23	0	1
24	0	1
25	0	0
26	1	0
27	0	1
28	1	0
29	1	0
30	1	0
31	1	0
32	0	1
33	1	0
34	0	1
35	0	1
36	0	1
37	0	1
38	0	0
39	1	0
40	0	1
41	0	1
42	0	1
43	1	0
44	0	0
45	1	0
46	1	0
47	1	0
48	0	1
49	0	0
50	0	1
51	0	1
52	0	0
53	1	0
54	0	0
55	0	0
56	0	0
57	0	1
58	1	0
59	0	1
60	0	0
61	1	0
62	0	1
63	1	0
64	0	0
65	0	0
66	0	1
67	0	1
68	0	0
69	1	0
70	1	0

71	0	1
72	0	1
73	0	0
74	1	0
75	0	1
76	0	1
77	0	1
78	1	0
79	1	0
80	0	0
81	0	1
82	0	1
83	0	1
84	1	0
85	0	1
86	0	0
87	0	0
88	0	0
89	0	0
90	0	1

	rest_ecgAbnormality in ST-T wave	rest_ecgleft ventricular hypertrophy
1	0	1
2	0	1
3	0	0
4	0	0
5	0	1
6	0	0
7	0	1
8	0	1
9	0	1
10	0	1
11	0	0
12	0	1
13	0	1
14	0	0
15	0	1
16	0	0
17	0	0
18	0	1
19	0	1
20	0	0
21	0	1
22	0	1
23	0	0
24	0	0
25	0	1
26	0	0
27	0	1
28	0	1

29	0	1
30	0	1
31	0	1
32	0	1
33	0	0
34	0	1
35	0	1
36	0	1
37	0	1
38	0	1
39	0	0
40	0	1
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	1
47	0	1
48	0	1
49	0	0
50	0	1
51	0	0
52	0	0
53	0	1
54	0	0
55	0	1
56	0	0
57	0	1
58	0	0
59	0	0
60	0	1
61	0	0
62	0	0
63	0	0
64	0	0
65	0	1
66	0	1
67	0	1
68	0	1
69	0	0
70	0	0
71	0	0
72	0	0
73	0	1
74	1	0
75	0	0
76	0	1
77	0	1

78			0	0
79			0	0
80			0	0
81			0	0
82			0	0
83			0	1
84			0	1
85			0	0
86			0	1
87			0	1
88			0	1
89			0	0
90			0	1

	st_slopeflat	st_slopedownsloping	thalassemiafixed	defect
1	1	0	1	
2	1	0	0	
3	0	0	0	
4	1	0	0	
5	0	0	1	
6	0	0	0	
7	1	0	0	
8	1	0	0	
9	1	0	0	
10	1	0	0	
11	1	0	0	
12	0	0	0	
13	0	0	1	
14	1	0	1	
15	0	0	1	
16	1	0	1	
17	1	0	0	
18	0	1	0	
19	1	0	1	
20	0	0	0	
21	1	0	0	
22	1	0	1	
23	1	0	1	
24	0	0	1	
25	0	0	1	
26	0	0	1	
27	1	0	1	
28	0	0	1	
29	1	0	0	
30	0	0	1	
31	1	0	0	
32	0	0	0	
33	0	0	1	
34	0	1	0	
35	1	0	0	

36	0	0	0
37	0	0	0
38	0	0	1
39	0	0	1
40	0	0	1
41	0	0	0
42	0	0	1
43	0	0	1
44	0	0	0
45	1	0	0
46	0	0	1
47	0	0	1
48	0	0	1
49	1	0	0
50	1	0	0
51	1	0	0
52	0	0	1
53	0	0	1
54	0	0	1
55	1	0	1
56	0	0	1
57	1	0	1
58	0	0	1
59	1	0	1
60	1	0	0
61	0	0	0
62	1	0	0
63	0	0	1
64	0	0	1
65	0	1	0
66	1	0	0
67	0	0	1
68	0	0	1
69	1	0	1
70	0	0	1
71	1	0	0
72	0	0	1
73	0	0	1
74	1	0	1
75	0	0	1
76	1	0	0
77	1	0	1
78	0	0	1
79	1	0	1
80	0	0	1
81	1	0	0
82	1	0	0
83	0	0	0
84	1	0	1

85	1	0	0
86	0	0	1
87	0	0	1
88	0	1	0
89	0	0	1
90	1	0	0

	thalassemianormal	blood flow	thalassemiareversible	defect
1		0		0
2		0		1
3		0		1
4		0		1
5		0		0
6		0		1
7		1		0
8		0		1
9		0		1
10		0		1
11		0		1
12		0		1
13		0		0
14		0		0
15		0		0
16		0		0
17		0		1
18		0		1
19		0		0
20		0		1
21		0		1
22		0		0
23		0		0
24		0		0
25		0		0
26		0		0
27		0		0
28		0		0
29		0		1
30		0		0
31		0		1
32		1		0
33		0		0
34		0		1
35		0		1
36		0		1
37		0		1
38		0		0
39		0		0
40		0		0
41		0		1
42		0		0

43	0	0
44	0	1
45	1	0
46	0	0
47	0	0
48	0	0
49	1	0
50	0	1
51	1	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	1
61	0	1
62	0	1
63	0	0
64	0	0
65	1	0
66	1	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	1
72	0	0
73	0	0
74	0	0
75	0	0
76	0	1
77	0	0
78	0	0
79	0	0
80	0	0
81	0	1
82	0	1
83	0	1
84	0	0
85	1	0
86	0	0
87	0	0
88	0	1
89	0	0
90	0	1

[reached getOption("max.print") -- omitted 180 rows]

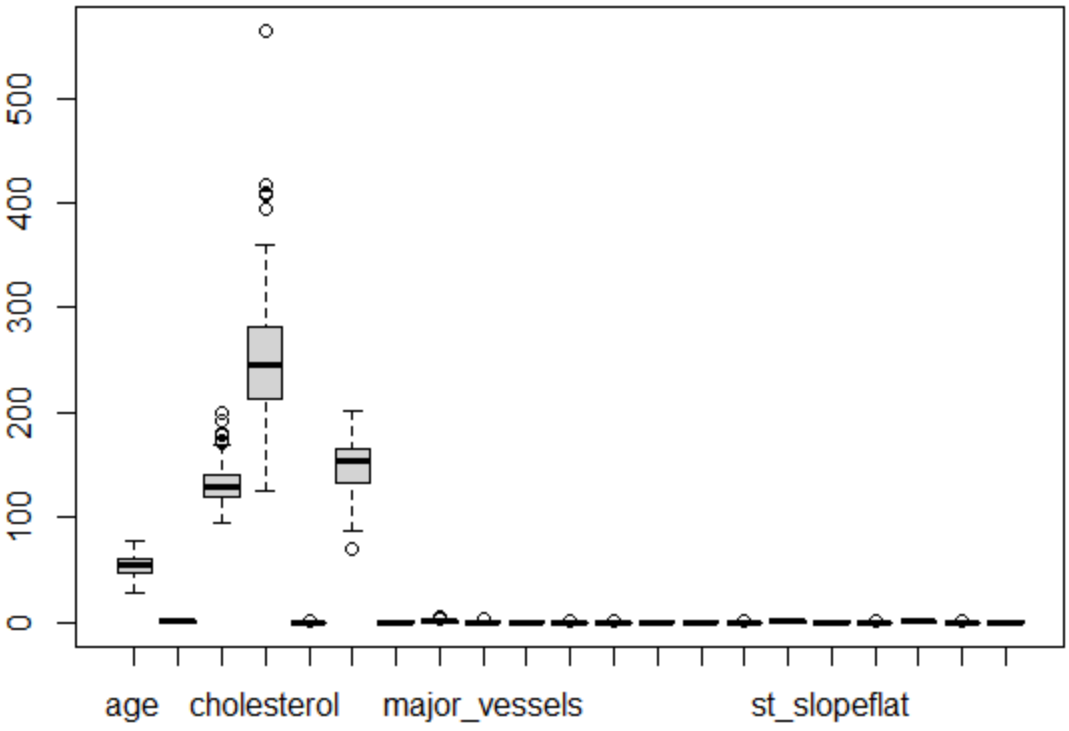
```
attr(,"assign")
 [1] 1 1 1 1 2 2 3 3 4 4 4
attr(,"contrasts")
attr(,"contrasts")$chest_pain_type
[1] "contr.treatment"

attr(,"contrasts")$rest_ecg
[1] "contr.treatment"

attr(,"contrasts")$st_slope
[1] "contr.treatment"

attr(,"contrasts")$thalassemia
[1] "contr.treatment"

> hd <- cbind(hd, dummy_variables)
> colnames(hd)
 [1] "age"                                  "sex"
 [3] "chest_pain_type"                      "resting_blood_pressure"
 [5] "cholesterol"                          "fasting_blood_sugar"
 [7] "rest_ecg"                             "max_heart_rate_achieved"
 [9] "exercise_induced_angina"              "st_depression"
[11] "st_slope"                             "major_vessels"
[13] "thalassemia"                          "target"
[15] "chest_pain_typetypical angina"        "chest_pain_typeatypical angina"
[17] "chest_pain_typenon-angina pain"       "chest_pain_typeasymptomatic"
[19] "rest_ecgAbnormality in ST-T wave"     "rest_ecgleft ventricular
hypertrophy"
[21] "st_slopeflat"                         "st_slopedownsloping"
[23] "thalassemiafixed defect"              "thalassemianormal blood flow"
[25] "thalassemiareversible defect"
```

Update the main dataset 'hd' by keeping only non-factor columns

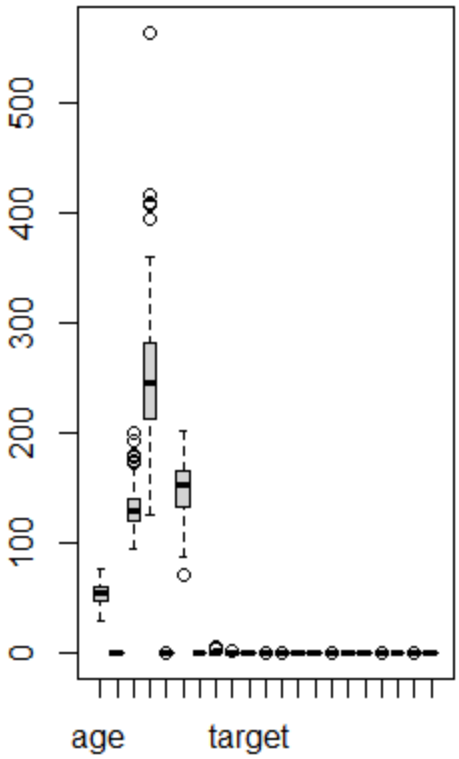
```
Identify the factor columns
> factor_columns <- sapply(hd, is.factor)
> hd <- hd[, !factor_columns]
> str(hd)
'data.frame':     270 obs. of  21 variables:
 $ age                                 : int  70 67 57 64 74 65 56 59 60 63
...
 $ sex                                 : int  1 0 1 1 0 1 1 1 1 0 ...
 $ resting_blood_pressure              : int  130 115 124 128 120 120 130 110
140 150 ...
 $ cholesterol                         : int  322 564 261 263 269 177 256 239
293 407 ...
 $ fasting_blood_sugar                 : int  0 0 0 0 0 0 1 0 0 0 ...
 $ max_heart_rate_achieved             : int  109 160 141 105 121 140 142 142
170 154 ...
 $ exercise_induced_angina             : int  0 0 0 1 1 0 1 1 0 0 ...
 $ st_depression                       : num  2.4 1.6 0.3 0.2 0.2 0.4 0.6 1.2
1.2 4 ...
 $ major_vessels                       : int  3 0 0 1 1 0 1 1 2 3 ...
 $ target                              : int  1 0 1 0 0 0 1 1 1 1 ...
 $ chest_pain_type_typical_angina      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ chest_pain_type_atypical_angina     : num  0 0 1 0 1 0 0 0 0 0 ...
 $ chest_pain_type_non_angina_pain     : num  0 1 0 0 0 0 1 0 0 0 ...
 $ chest_pain_type_asymptomatic        : num  1 0 0 1 0 1 0 1 1 1 ...
 $ rest_ecg_Abnormality_in_ST_T_wave   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ rest_ecgleft_ventricular_hypertrophy: num  1 1 0 0 1 0 1 1 1 1 ...
 $ st_slopeflat                        : num  1 1 0 1 0 0 1 1 1 1 ...
 $ st_slopedownsloping                 : num  0 0 0 0 0 0 0 0 0 0 ...
 $ thalassemia_fixed_defect            : num  1 0 0 0 1 0 0 0 0 0 ...
 $ thalassemia_normal_blood_flow       : num  0 0 0 0 0 0 1 0 0 0 ...
 $ thalassemia_reversible_defect       : num  0 1 1 1 0 1 0 1 1 1 ...
```

#data modelling/outliers

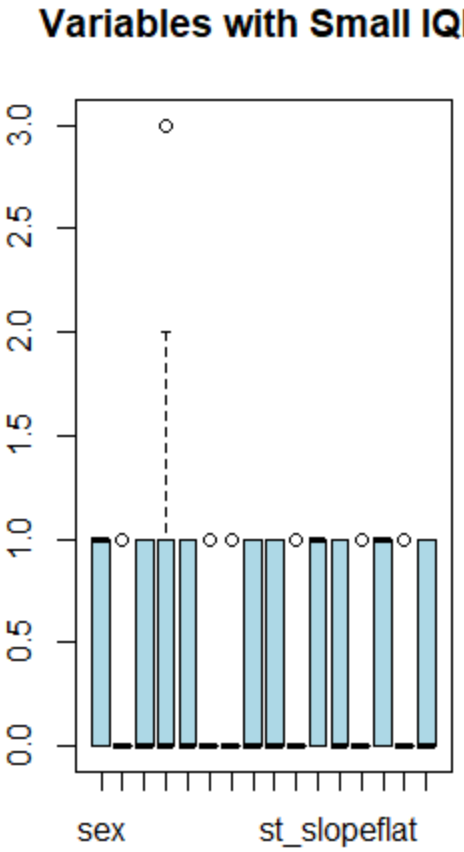
boxplot(hd)
# Plot the boxplot of the dataset without outliers
boxplot(hd)

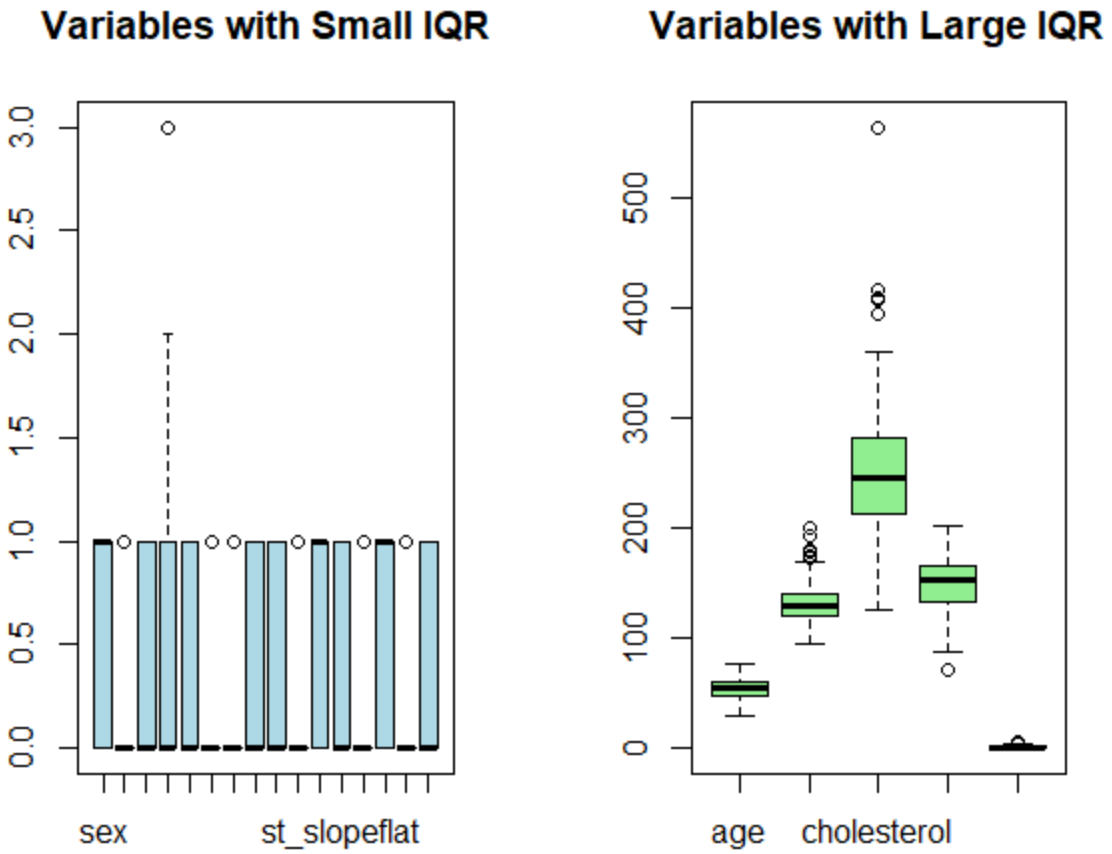



```
# Plot the boxplot of the dataset without outliers
boxplot(hd)
```




```
# Calculate the IQR for each variable
iqr_vals <- apply(hd, 2, IQR)

# Determine the threshold for splitting the variables
iqr_threshold <- quantile(iqr_vals, 0.5)

# Split the variables into two groups
small_iqr_vars <- names(iqr_vals[iqr_vals <= iqr_threshold])
large_iqr_vars <- names(iqr_vals[iqr_vals > iqr_threshold])

# Plot the boxplots for the variables with a small IQR
par(mfrow=c(1,2))
boxplot(hd[, small_iqr_vars], main="Variables with Small IQR", col="lightblue")

# Plot the boxplots for the variables with a large IQR
boxplot(hd[, large_iqr_vars], main="Variables with Large IQR", col="lightgreen")
```


Variables with Small IQR
3.0
2.5
2.0
1.5
1.0
0.5
0.0
sex
st_slopeflat



#logistic regression
model <- glm(target ~ age + sex + chest_pain_type_typical_angina + chest_pain_type_atypical_angina + chest_pain_type_non_angina_pain + chest_pain_type_asymptomatic + resting_blood_pressure + cholesterol + fasting_blood_sugar + rest_ecg_Abnormality_in_ST_T_wave + rest_ecgleft_ventricular_hypertrophy + max_heart_rate_achieved + exercise_induced_angina + st_depression + st_slopeflat + st_slopedownsloping + major_vessels + thalassemia_fixed_defect + thalassemia_normal_blood_flow + thalassemia_reversible_defect, data = hd)
summary(model)

```
Call:
glm(formula = target ~ age + sex + chest_pain_type_typical_angina +
    chest_pain_type_atypical_angina + chest_pain_type_non_angina_pain +
    chest_pain_type_asymptomatic + resting_blood_pressure + cholesterol +
    fasting_blood_sugar + rest_ecg_Abnormality_in_ST_T_wave +
    rest_ecgleft_ventricular_hypertrophy + max_heart_rate_achieved +
    exercise_induced_angina + st_depression + st_slopeflat +
    st_slopedownsloping + major_vessels + thalassemia_fixed_defect +
    thalassemia_normal_blood_flow + thalassemia_reversible_defect,
```

```
    data = hd)

Coefficients: (2 not defined because of singularities)
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                         0.3563048  0.3083393   1.156 0.248960
age                                -0.0013692  0.0028054  -0.488 0.625930
sex                                 0.1650129  0.0514051   3.210 0.001500 **
chest_pain_type_typical_angina     -0.2948531  0.0883215  -3.338 0.000971 ***
chest_pain_type_atypical_angina    -0.1746684  0.0686272  -2.545 0.011522 *
chest_pain_type_non_angina_pain    -0.2159668  0.0557626  -3.873 0.000137 ***
chest_pain_type_asymptomatic               NA         NA      NA       NA
resting_blood_pressure              0.0021268  0.0013058   1.629 0.104625
cholesterol                         0.0004800  0.0004381   1.096 0.274285
fasting_blood_sugar                -0.0403410  0.0621366  -0.649 0.516783
rest_ecg_Abnormality_in_ST_T_wave   0.1123395  0.2538099   0.443 0.658427
rest_ecgleft_ventricular_hypertrophy 0.0701175 0.0439581   1.595 0.111949
max_heart_rate_achieved            -0.0022791  0.0011785  -1.934 0.054239 .
exercise_induced_angina             0.0876399  0.0526335   1.665 0.097141 .
st_depression                       0.0478209  0.0251295   1.903 0.058187 .
st_slopeflat                        0.1046971  0.0542299   1.931 0.054657 .
st_slopedownsloping                 0.0160250  0.1057981   0.151 0.879728
major_vessels                       0.1251126  0.0256798   4.872 1.96e-06 ***
thalassemia_fixed_defect           -0.2276216  0.0533227  -4.269 2.79e-05 ***
thalassemia_normal_blood_flow      -0.1668977  0.0998752  -1.671 0.095956 .
thalassemia_reversible_defect              NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.1160807)

    Null deviance: 66.667  on 269  degrees of freedom
Residual deviance: 29.136  on 251  degrees of freedom
AIC: 205.09

Number of Fisher Scoring iterations: 2
```

#stepwise regression

```
Start:  AIC=205.09
target ~ age + sex + chest_pain_type_typical_angina +
chest_pain_type_atypical_angina +
    chest_pain_type_non_angina_pain + chest_pain_type_asymptomatic +
    resting_blood_pressure + cholesterol + fasting_blood_sugar +
    rest_ecg_Abnormality_in_ST_T_wave + rest_ecgleft_ventricular_hypertrophy +
    max_heart_rate_achieved + exercise_induced_angina + st_depression +
    st_slopeflat + st_slopedownsloping + major_vessels +
thalassemia_fixed_defect +
    thalassemia_normal_blood_flow + thalassemia_reversible_defect
```

```
Step:  AIC=205.09
target ~ age + sex + chest_pain_type_typical_angina +
chest_pain_type_atypical_angina +
    chest_pain_type_non_angina_pain + chest_pain_type_asymptomatic +
    resting_blood_pressure + cholesterol + fasting_blood_sugar +
    rest_ecg_Abnormality_in_ST_T_wave + rest_ecgleft_ventricular_hypertrophy +
    max_heart_rate_achieved + exercise_induced_angina + st_depression +
    st_slopeflat + st_slopedownsloping + major_vessels +
thalassemia_fixed_defect +
    thalassemia_normal_blood_flow


Step:  AIC=205.09
target ~ age + sex + chest_pain_type_typical_angina +
chest_pain_type_atypical_angina +
    chest_pain_type_non_angina_pain + resting_blood_pressure +
    cholesterol + fasting_blood_sugar + rest_ecg_Abnormality_in_ST_T_wave +
    rest_ecgleft_ventricular_hypertrophy + max_heart_rate_achieved +
    exercise_induced_angina + st_depression + st_slopeflat +
    st_slopedownsloping + major_vessels + thalassemia_fixed_defect +
    thalassemia_normal_blood_flow

                                       Df Deviance    AIC
- st_slopedownsloping                   1   29.139 203.11
- rest_ecg_Abnormality_in_ST_T_wave     1   29.159 203.30
- age                                   1   29.164 203.34
- fasting_blood_sugar                   1   29.185 203.54
- cholesterol                           1   29.276 204.38
<none>                                      29.136 205.09
- rest_ecgleft_ventricular_hypertrophy  1   29.432 205.81
- resting_blood_pressure                1   29.444 205.93
- exercise_induced_angina               1   29.458 206.05
- thalassemia_normal_blood_flow         1   29.460 206.08
- st_depression                         1   29.557 206.96
- st_slopeflat                          1   29.569 207.07
- max_heart_rate_achieved               1   29.570 207.08
- chest_pain_type_atypical_angina       1   29.888 209.97
- sex                                   1   30.332 213.95
- chest_pain_type_typical_angina        1   30.430 214.82
- chest_pain_type_non_angina_pain       1   30.877 218.76
- thalassemia_fixed_defect              1   31.252 222.01
- major_vessels                         1   31.892 227.49

Step:  AIC=203.11
target ~ age + sex + chest_pain_type_typical_angina +
chest_pain_type_atypical_angina +
    chest_pain_type_non_angina_pain + resting_blood_pressure +
    cholesterol + fasting_blood_sugar + rest_ecg_Abnormality_in_ST_T_wave +
    rest_ecgleft_ventricular_hypertrophy + max_heart_rate_achieved +
```

```
    exercise_induced_angina + st_depression + st_slopeflat +
    major_vessels + thalassemia_fixed_defect + thalassemia_normal_blood_flow

                                       Df Deviance    AIC
- rest_ecg_Abnormality_in_ST_T_wave     1   29.161 201.32
- age                                   1   29.167 201.37
- fasting_blood_sugar                   1   29.186 201.54
- cholesterol                           1   29.276 202.38
<none>                                      29.139 203.11
- rest_ecgleft_ventricular_hypertrophy  1   29.443 203.91
- resting_blood_pressure                1   29.450 203.98
- thalassemia_normal_blood_flow         1   29.461 204.08
- exercise_induced_angina               1   29.461 204.09
+ st_slopedownsloping                   1   29.136 205.09
- max_heart_rate_achieved               1   29.583 205.20
- st_slopeflat                          1   29.638 205.70
- st_depression                         1   29.745 206.67
- chest_pain_type_atypical_angina       1   29.890 207.98
- sex                                   1   30.333 211.96
- chest_pain_type_typical_angina        1   30.431 212.83
- chest_pain_type_non_angina_pain       1   30.879 216.77
- thalassemia_fixed_defect              1   31.269 220.16
- major_vessels                         1   31.948 225.96

Step:  AIC=201.32
target ~ age + sex + chest_pain_type_typical_angina +
chest_pain_type_atypical_angina +
    chest_pain_type_non_angina_pain + resting_blood_pressure +
    cholesterol + fasting_blood_sugar + rest_ecgleft_ventricular_hypertrophy +
    max_heart_rate_achieved + exercise_induced_angina + st_depression +
    st_slopeflat + major_vessels + thalassemia_fixed_defect +
    thalassemia_normal_blood_flow

                                       Df Deviance    AIC
- age                                   1   29.186 199.55
- fasting_blood_sugar                   1   29.209 199.76
- cholesterol                           1   29.297 200.57
<none>                                      29.161 201.32
- rest_ecgleft_ventricular_hypertrophy  1   29.451 201.99
- thalassemia_normal_blood_flow         1   29.485 202.30
- exercise_induced_angina               1   29.485 202.30
- resting_blood_pressure                1   29.502 202.46
+ rest_ecg_Abnormality_in_ST_T_wave     1   29.139 203.11
+ st_slopedownsloping                   1   29.159 203.30
- max_heart_rate_achieved               1   29.624 203.57
- st_slopeflat                          1   29.679 204.07
- st_depression                         1   29.781 205.00
- chest_pain_type_atypical_angina       1   29.915 206.21
- sex                                   1   30.340 210.02
```

```
- chest_pain_type_typical_angina        1   30.467 211.15
- chest_pain_type_non_angina_pain       1   30.894 214.91
- thalassemia_fixed_defect              1   31.269 218.17
- major_vessels                         1   31.948 223.96

Step:  AIC=199.55
target ~ sex + chest_pain_type_typical_angina + chest_pain_type_atypical_angina
+
    chest_pain_type_non_angina_pain + resting_blood_pressure +
    cholesterol + fasting_blood_sugar + rest_ecgleft_ventricular_hypertrophy +
    max_heart_rate_achieved + exercise_induced_angina + st_depression +
    st_slopeflat + major_vessels + thalassemia_fixed_defect +
    thalassemia_normal_blood_flow

                                       Df Deviance    AIC
- fasting_blood_sugar                   1   29.239 198.04
- cholesterol                           1   29.307 198.67
<none>                                      29.186 199.55
- rest_ecgleft_ventricular_hypertrophy  1   29.470 200.17
- resting_blood_pressure                1   29.503 200.47
- thalassemia_normal_blood_flow         1   29.519 200.61
- exercise_induced_angina               1   29.526 200.67
+ age                                   1   29.161 201.32
+ rest_ecg_Abnormality_in_ST_T_wave     1   29.167 201.37
+ st_slopedownsloping                   1   29.184 201.53
- max_heart_rate_achieved               1   29.633 201.66
- st_slopeflat                          1   29.711 202.37
- st_depression                         1   29.812 203.28
- chest_pain_type_atypical_angina       1   29.935 204.39
- sex                                   1   30.413 208.67
- chest_pain_type_typical_angina        1   30.530 209.71
- chest_pain_type_non_angina_pain       1   30.942 213.32
- thalassemia_fixed_defect              1   31.302 216.44
- major_vessels                         1   32.036 222.71

Step:  AIC=198.04
target ~ sex + chest_pain_type_typical_angina + chest_pain_type_atypical_angina
+
    chest_pain_type_non_angina_pain + resting_blood_pressure +
    cholesterol + rest_ecgleft_ventricular_hypertrophy +
max_heart_rate_achieved +
    exercise_induced_angina + st_depression + st_slopeflat +
    major_vessels + thalassemia_fixed_defect + thalassemia_normal_blood_flow

                                       Df Deviance    AIC
- cholesterol                           1   29.360 197.16
<none>                                      29.239 198.04
- rest_ecgleft_ventricular_hypertrophy  1   29.511 198.54
- resting_blood_pressure                1   29.527 198.69
```

```
- exercise_induced_angina               1   29.568 199.06
- thalassemia_normal_blood_flow         1   29.581 199.18
+ fasting_blood_sugar                   1   29.186 199.55
+ age                                   1   29.209 199.76
+ rest_ecg_Abnormality_in_ST_T_wave     1   29.218 199.85
+ st_slopedownsloping                   1   29.239 200.04
- max_heart_rate_achieved               1   29.683 200.11
- st_slopeflat                          1   29.800 201.17
- st_depression                         1   29.896 202.04
- chest_pain_type_atypical_angina       1   30.013 203.09
- sex                                   1   30.455 207.04
- chest_pain_type_typical_angina        1   30.665 208.90
- chest_pain_type_non_angina_pain       1   31.172 213.32
- thalassemia_fixed_defect              1   31.325 214.65
- major_vessels                         1   32.037 220.71

Step:  AIC=197.16
target ~ sex + chest_pain_type_typical_angina + chest_pain_type_atypical_angina
+
    chest_pain_type_non_angina_pain + resting_blood_pressure +
    rest_ecgleft_ventricular_hypertrophy + max_heart_rate_achieved +
    exercise_induced_angina + st_depression + st_slopeflat +
    major_vessels + thalassemia_fixed_defect + thalassemia_normal_blood_flow

                                       Df Deviance    AIC
<none>                                      29.360 197.16
+ cholesterol                           1   29.239 198.04
- rest_ecgleft_ventricular_hypertrophy  1   29.699 198.26
- resting_blood_pressure                1   29.714 198.39
- exercise_induced_angina               1   29.715 198.40
+ fasting_blood_sugar                   1   29.307 198.67
- thalassemia_normal_blood_flow         1   29.747 198.69
+ rest_ecg_Abnormality_in_ST_T_wave     1   29.339 198.96
+ age                                   1   29.347 199.03
- max_heart_rate_achieved               1   29.785 199.04
+ st_slopedownsloping                   1   29.360 199.15
- st_slopeflat                          1   29.946 200.49
- st_depression                         1   29.988 200.87
- chest_pain_type_atypical_angina       1   30.130 202.14
- sex                                   1   30.469 205.16
- chest_pain_type_typical_angina        1   30.846 208.48
- chest_pain_type_non_angina_pain       1   31.321 212.61
- thalassemia_fixed_defect              1   31.522 214.34
- major_vessels                         1   32.302 220.94
```

#logit----------

Logit model summary

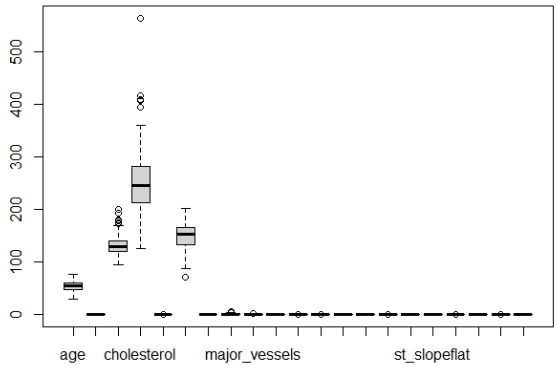
```
exp(coef(modelo))
```

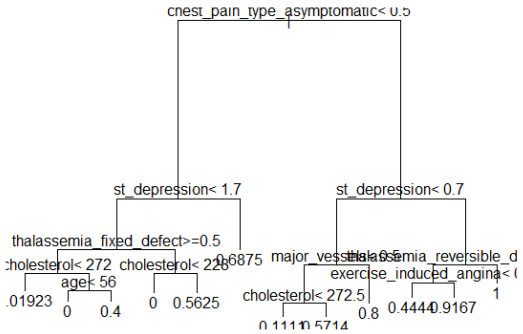
```
                          (Intercept) rest_ecgleft_ventricular_hypertrophy
                            0.3007351                             1.6194866
```

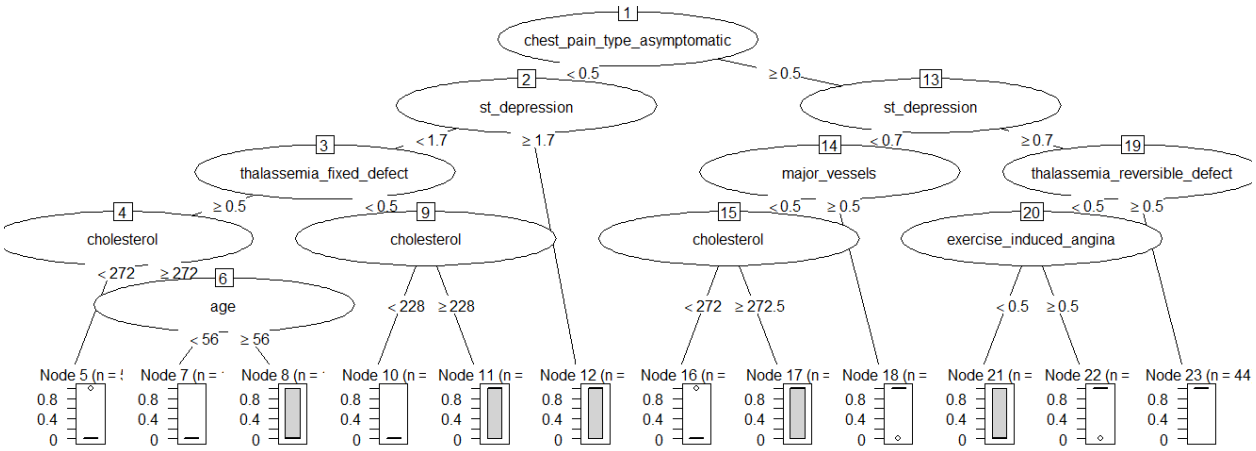
```
                        st_slopeflat               exercise_induced_angina
                           2.0457940                             2.5038577
                         cholesterol              resting_blood_pressure
                           1.0078829                             1.0162559
             max_heart_rate_achieved                                   age
                           0.9767305                             0.9734884
                                 sex                        st_depression
                           3.7316554                             1.4603150
     chest_pain_type_non_angina_pain                       major_vessels
                           0.3138164                             3.1648695
            thalassemia_fixed_defect
                           0.2519735
# Odds ratios
> exp(coef(modelo))
                         (Intercept) rest_ecgleft_ventricular_hypertrophy
                           0.3007351                             1.6194866
                        st_slopeflat               exercise_induced_angina
                           2.0457940                             2.5038577
                         cholesterol              resting_blood_pressure
                           1.0078829                             1.0162559
             max_heart_rate_achieved                                   age
                           0.9767305                             0.9734884
                                 sex                        st_depression
                           3.7316554                             1.4603150
     chest_pain_type_non_angina_pain                       major_vessels
                           0.3138164                             3.1648695
            thalassemia_fixed_defect
                           0.2519735
# Make predictions
# Prediction of membership probabilities
probabilities <-  modelo %>% predict(vali, type = "response")
> head(probabilities)
         1          2          3          7          9         16
0.98876376 0.53962563 0.36042895 0.84681050 0.95325935 0.04176075
# Assigning the observations to the class
#converting target into factor
vali$target <- as.factor(vali$target)
> contrasts(vali$target)
  1
0 0
1 1
> predicted.classes <- ifelse(probabilities > 0.5, 1, 0)
> table(predicted.classes, vali$target)

predicted.classes  0  1
                0 42  5
                1  6 28
# Model accuracy
```

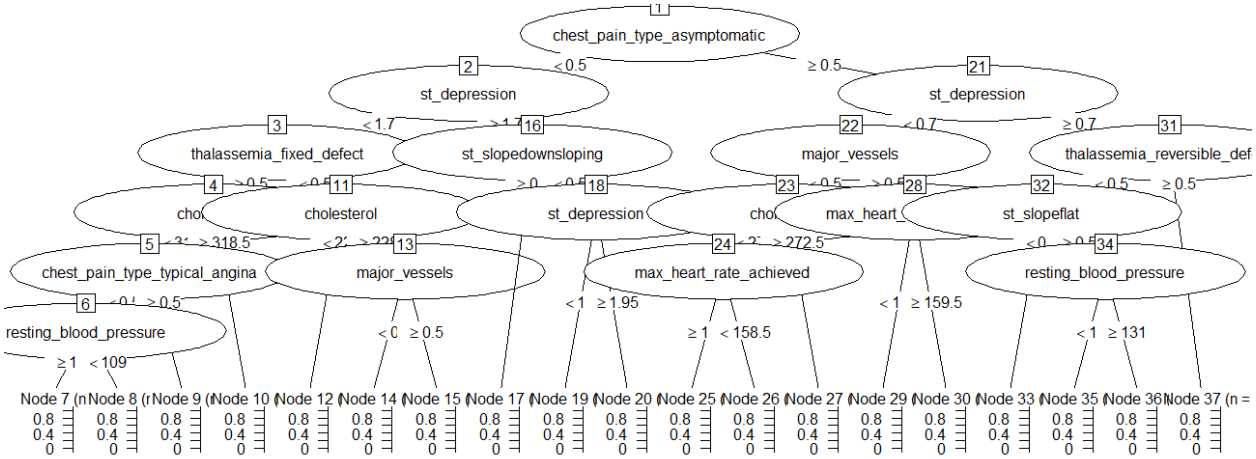
```
> # Classification prediction accuracy is about 78% and misclassification error
rate is 22%.
> head(predicted.classes)
 1  2  3  7  9 16
 1  1  0  1  1  0
> mean(predicted.classes == vali$target)
[1] 0.8641975
```

Plot the regression tree



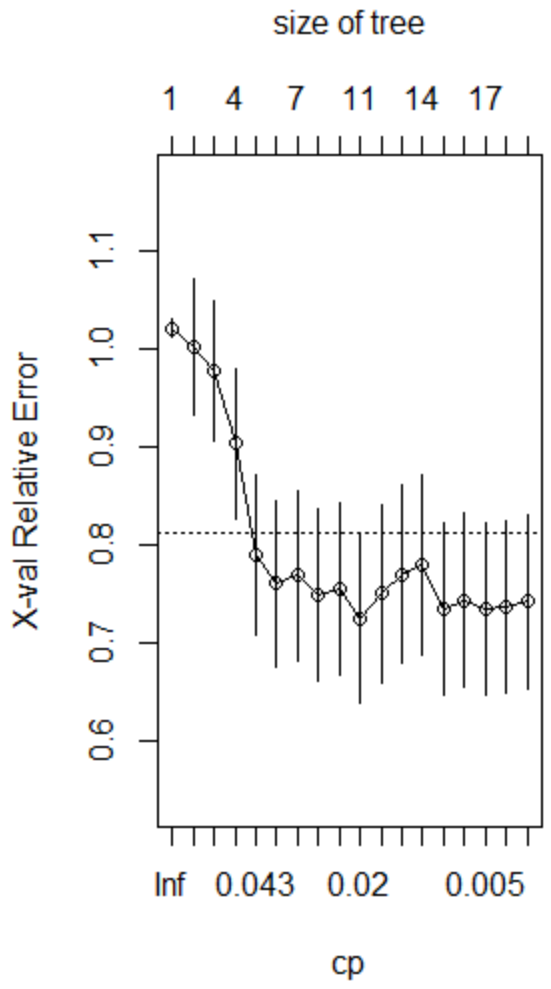






Display the results of cross-validation

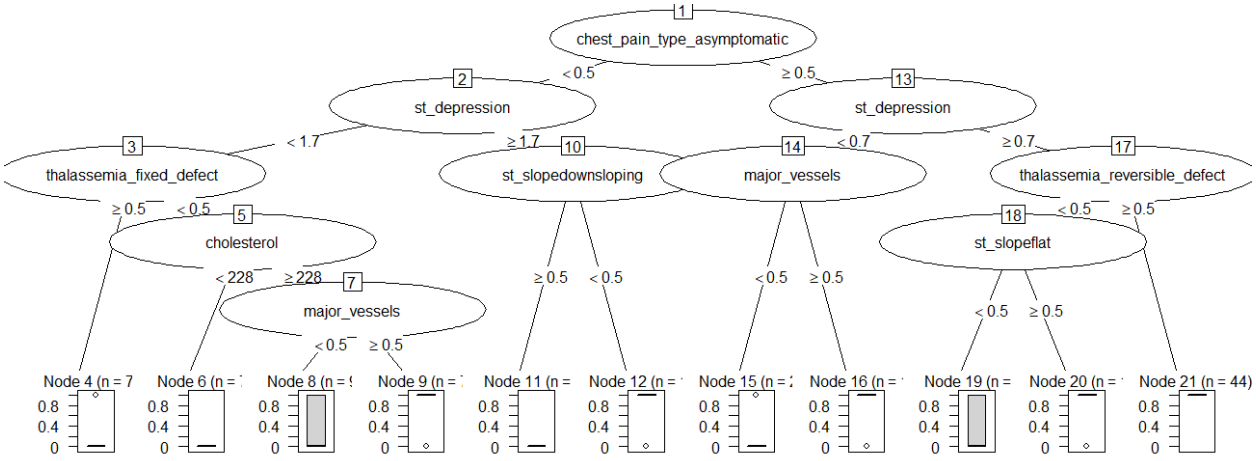
```
printcp(fit)

Regression tree:
rpart(formula = target ~ ., data = hd, subset = inTrain, method = "anova",
    minsplit = 10, cp = 0)

Variables actually used in tree construction:
 [1] chest_pain_type_asymptomatic   chest_pain_type_typical_angina
 [3] cholesterol                    major_vessels
 [5] max_heart_rate_achieved        resting_blood_pressure
 [7] st_depression                  st_slopedownsloping
 [9] st_slopeflat                   thalassemia_fixed_defect
[11] thalassemia_reversible_defect

Root node error: 53.833/216 = 0.24923

n= 216

          CP nsplit rel error  xerror      xstd
1  0.2587733      0   1.00000 1.02048 0.0088318
2  0.0963563      1   0.74123 1.00224 0.0696628
3  0.0742110      2   0.64487 0.97656 0.0714395
4  0.0546130      3   0.57066 0.90272 0.0760996
5  0.0335443      4   0.51605 0.78991 0.0820179
6  0.0324184      5   0.48250 0.75994 0.0846068
7  0.0286209      6   0.45008 0.76846 0.0867143
8  0.0221005      7   0.42146 0.74887 0.0874477
9  0.0200686      9   0.37726 0.75481 0.0872614
10 0.0198378     10   0.35719 0.72497 0.0870957
11 0.0128602     11   0.33736 0.75024 0.0906180
12 0.0109060     12   0.32450 0.76931 0.0905396
13 0.0097523     13   0.31359 0.77876 0.0921617
14 0.0071075     14   0.30384 0.73383 0.0881077
15 0.0056612     15   0.29673 0.74314 0.0892185
16 0.0043291     16   0.29107 0.73403 0.0880764
17 0.0041280     17   0.28674 0.73628 0.0883749
18 0.0000000     18   0.28261 0.74191 0.0889941


 >

# Visualize cross-validation results
> plotcp(fit)
```



Visualize cross-validation results
plotcp(fit)

Prune the tree
fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"]
pfit <- prune(fit, cp = fit$cptable[which.min(fit$cptable[,"xerror"]),"CP"]) # from cptable

# Plot the pruned regression tree
plot(as.party(pfit))




```
# Compare the performance of the pruned tree with the full tree on the
validation data
> pred_v_tree <- predict(fit, newdata = vali)
> pred_v_ptree <- predict(pfit, newdata = vali)
> accuracy(pred_v_tree, vali$target)
                ME      RMSE       MAE  MPE MAPE
Test set -0.120341 0.4342757 0.2872281 -Inf  Inf
> accuracy(pred_v_ptree, vali$target)
                 ME      RMSE       MAE  MPE MAPE
Test set -0.1663716 0.4453206 0.2967639 -Inf  Inf
```