MVA Final Project

Javier Ferrando Monsonis
Marcel Porta Valles
Mehmet Fatih ??agil

February 20, 2018

Libraries

```r
library(chemometrics)
library(DMwR)
library(mice)
library(missForest)
library(ggplot2)
library(graphics)
library(gridExtra)
library(Hmisc)
library(knitr)
library(FactoMineR)
library(DataExplorer)
```

```
## Warning: package 'DataExplorer' was built under R version 3.5.2
```

```r
library(factoextra)
library(expm)
library(fpc)
library(cluster)

theme_set(theme_bw())
setwd("/Users/JaviFerrando/Desktop/MVA-Project")
```

```r
heart_disease = read.csv("data/heart.csv")
columns <- colnames(heart_disease)
columns[1] <- "age"
colnames(heart_disease) <- columns

# Find missing variables
which(is.na(heart_disease))
```

```
## integer(0)
```

```r
kable(head(heart_disease))
```

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
57	1	0	140	192	0	1	148	0	0.4	1	0	1	1

```
describe(heart_disease)
```

```
## heart_disease
##
##  14  Variables      303  Observations
## --------------------------------------------------------------------------------
## age
##          n   missing  distinct      Info      Mean       Gmd       .05       .10
##        303         0        41     0.999     54.37     10.36      39.1      42.0
##        .25       .50       .75       .90       .95
##       47.5      55.0      61.0      66.0      68.0
##
## lowest : 29 34 35 37 38, highest: 70 71 74 76 77
## --------------------------------------------------------------------------------
## sex
##          n   missing  distinct      Info       Sum      Mean       Gmd
##        303         0         2     0.649       207    0.6832    0.4343
##
## --------------------------------------------------------------------------------
## cp
##          n   missing  distinct      Info      Mean       Gmd
##        303         0         4     0.866     0.967     1.105
##
## Value          0     1     2     3
## Frequency    143    50    87    23
## Proportion 0.472 0.165 0.287 0.076
## --------------------------------------------------------------------------------
## trestbps
##          n   missing  distinct      Info      Mean       Gmd       .05       .10
##        303         0        49     0.995     131.6     19.32       108       110
##        .25       .50       .75       .90       .95
##        120       130       140       152       160
##
## lowest :  94 100 101 102 104, highest: 174 178 180 192 200
## --------------------------------------------------------------------------------
## chol
##          n   missing  distinct      Info      Mean       Gmd       .05       .10
##        303         0       152         1     246.3     55.95     175.0     188.0
##        .25       .50       .75       .90       .95
##      211.0     240.0     274.5     308.8     326.9
##
## lowest : 126 131 141 149 157, highest: 394 407 409 417 564
## --------------------------------------------------------------------------------
## fbs
##          n   missing  distinct      Info       Sum      Mean       Gmd
##        303         0         2     0.379        45    0.1485    0.2538
##
## --------------------------------------------------------------------------------
## restecg
##          n   missing  distinct      Info      Mean       Gmd
##        303         0         3      0.76    0.5281    0.5274
##
## Value          0     1     2
## Frequency    147   152     4
```

```
## Proportion 0.485 0.502 0.013
## -------------------------------------------------------------------------------
## thalach
##        n    missing  distinct       Info       Mean        Gmd        .05        .10
##      303          0        91          1      149.6      25.77      108.1      116.0
##      .25        .50       .75        .90        .95
##    133.5      153.0     166.0      176.6      181.9
##
## lowest :  71  88  90  95  96, highest: 190 192 194 195 202
## -------------------------------------------------------------------------------
## exang
##        n    missing  distinct       Info        Sum       Mean        Gmd
##      303          0         2       0.66         99     0.3267     0.4414
##
## -------------------------------------------------------------------------------
## oldpeak
##        n    missing  distinct       Info       Mean        Gmd        .05        .10
##      303          0        40      0.964       1.04      1.225        0.0        0.0
##      .25        .50       .75        .90        .95
##      0.0        0.8       1.6        2.8        3.4
##
## lowest : 0.0 0.1 0.2 0.3 0.4, highest: 4.0 4.2 4.4 5.6 6.2
## -------------------------------------------------------------------------------
## slope
##        n    missing  distinct       Info       Mean        Gmd
##      303          0         3      0.798      1.399     0.6291
##
## Value           0     1     2
## Frequency      21   140   142
## Proportion  0.069 0.462 0.469
## -------------------------------------------------------------------------------
## ca
##        n    missing  distinct       Info       Mean        Gmd
##      303          0         5      0.795     0.7294      1.005
##
## Value           0     1     2     3     4
## Frequency     175    65    38    20     5
## Proportion  0.578 0.215 0.125 0.066 0.017
## -------------------------------------------------------------------------------
## thal
##        n    missing  distinct       Info       Mean        Gmd
##      303          0         4      0.778      2.314     0.6125
##
## Value           0     1     2     3
## Frequency       2    18   166   117
## Proportion  0.007 0.059 0.548 0.386
## -------------------------------------------------------------------------------
## target
##        n    missing  distinct       Info        Sum       Mean        Gmd
##      303          0         2      0.744        165     0.5446     0.4977
##
## -------------------------------------------------------------------------------
```

```
classVar <- lapply(heart_disease,class)    # class of each variable
factor_heart <- heart_disease
factor_heart$target <- as.factor(heart_disease$target)
factor_heart$sex <- as.factor(heart_disease$sex)
factor_heart$fbs <- as.factor(heart_disease$fbs)
factor_heart$exang <- as.factor(heart_disease$exang)
factor_heart$restecg <- as.factor(heart_disease$restecg)
factor_heart$thal <- as.factor(heart_disease$thal)
factor_heart$slope <- as.factor(heart_disease$slope)
factor_heart$cp <- as.factor(heart_disease$cp)
factor_heart$ca <- as.factor(heart_disease$ca)

#Outlier detection
############################
#Moutlier(heart_disease[,-14], quantile = 0.975, plot = TRUE, tol=1e-36) #Doesn't work


#Local Outlier Factor
outlier.scores <- lofactor(heart_disease[,-14], k=5)
plot(density(outlier.scores),main='Distribution of individuals local outlier factor scores')
```

Distribution of individuals local outlier factor scores
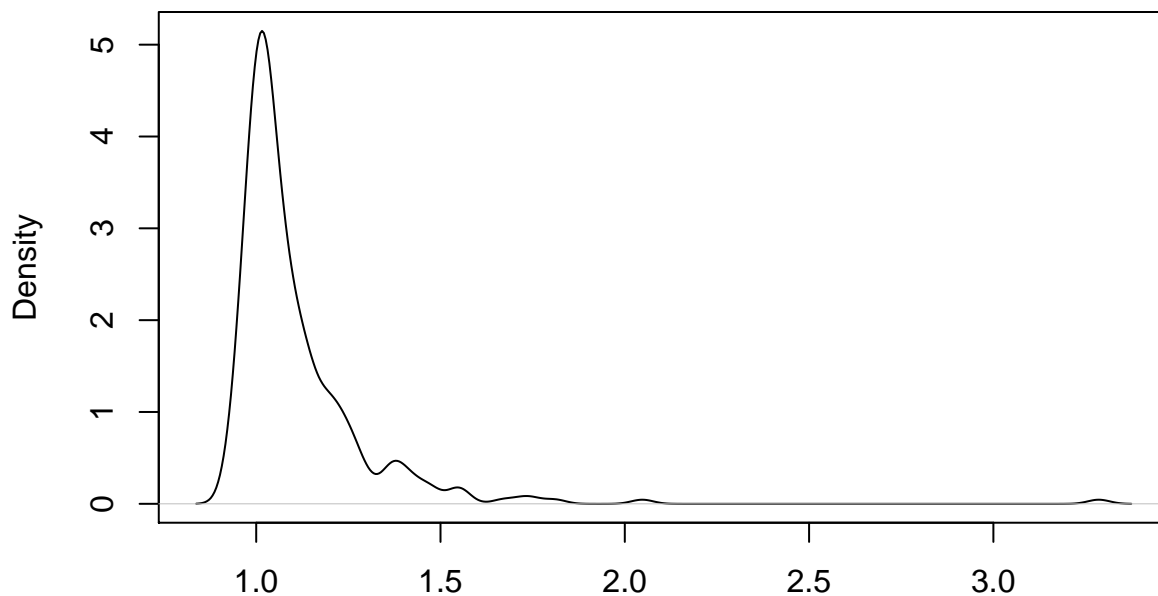
N = 303   Bandwidth = 0.02936

```
#Exploratory Data Analysis
#Density of heart presence/absence disease by age
g1 <- ggplot(data=heart_disease, aes(x=age, fill=as.factor(target)))+
  geom_density(alpha=.5)+
  ggtitle("Age") +
  scale_fill_manual(values = c('skyblue4', 'skyblue2'),name = "Disease", labels = c("Yes", "No"))

#Density of heart presence/absence disease by Max heart rate
g2 <- ggplot(data=heart_disease, aes(x=thalach, fill=as.factor(target)))+
```
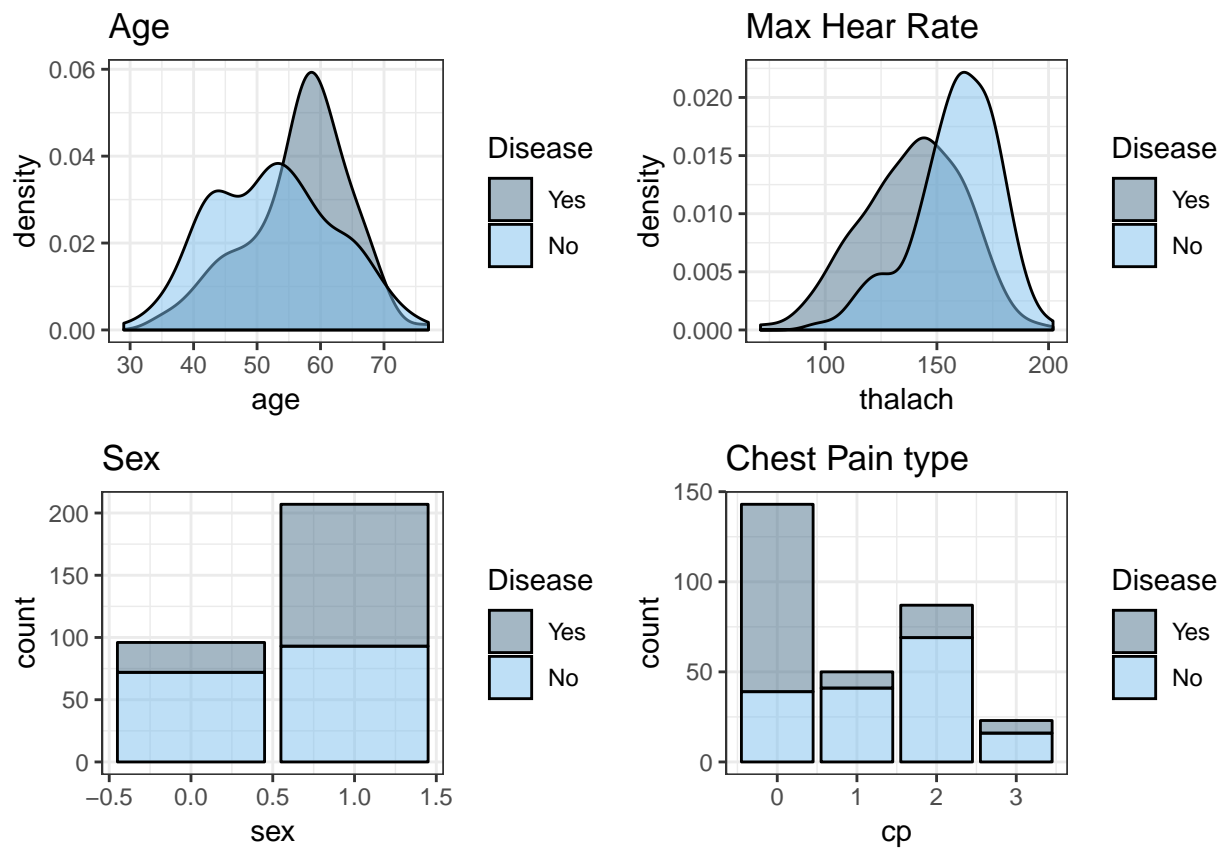
```
  geom_density(alpha=.5)+
  ggtitle("Max Hear Rate") +
  scale_fill_manual(values = c('skyblue4', 'skyblue2'),name = "Disease", labels = c("Yes", "No"))

#Density of heart presence/absence disease by sex
g3 <- ggplot(data=heart_disease, aes(x=sex, fill=as.factor(target)))+
    geom_bar(alpha=.5, color="black")+
    ggtitle("Sex") +
    scale_fill_manual(values = c('skyblue4', 'skyblue2'),name = "Disease", labels = c("Yes", "No"))

#Density of heart presence/absence disease by chest type
g4 <- ggplot(data=heart_disease, aes(x=cp, fill=as.factor(target)))+
  geom_bar(alpha=.5, color="black")+
  ggtitle("Chest Pain type") +
  scale_fill_manual(values = c('skyblue4', 'skyblue2'),name = "Disease", labels = c("Yes", "No"))

grid.arrange(g1, g2, g3, g4, ncol = 2)
```



```
plot_correlation(heart_disease)
```

Features	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
target	−0.23	−0.28	0.43	−0.14	−0.09	−0.03	0.14	0.42	−0.44	−0.43	0.35	−0.39	−0.34	1
thal	0.07	0.21	−0.16	0.06	0.1	−0.03	−0.01	−0.1	0.21	0.21	−0.1	0.15	1	−0.34
ca	0.28	0.12	−0.18	0.1	0.07	0.14	−0.07	−0.21	0.12	0.22	−0.08	1	0.15	−0.39
slope	−0.17	−0.03	0.12	−0.12	0	−0.06	0.09	0.39	−0.26	−0.58	1	−0.08	−0.1	0.35
oldpeak	0.21	0.1	−0.15	0.19	0.05	0.01	−0.06	−0.34	0.29	1	−0.58	0.22	0.21	−0.43
exang	0.1	0.14	−0.39	0.07	0.07	0.03	−0.07	−0.38	1	0.29	−0.26	0.12	0.21	−0.44
thalach	−0.4	−0.04	0.3	−0.05	−0.01	−0.01	0.04	1	−0.38	−0.34	0.39	−0.21	−0.1	0.42
restecg	−0.12	−0.06	0.04	−0.11	−0.15	−0.08	1	0.04	−0.07	−0.06	0.09	−0.07	−0.01	0.14
fbs	0.12	0.05	0.09	0.18	0.01	1	−0.08	−0.01	0.03	0.01	−0.06	0.14	−0.03	−0.03
chol	0.21	−0.2	−0.08	0.12	1	0.01	−0.15	−0.01	0.07	0.05	0	0.07	0.1	−0.09
trestbps	0.28	−0.06	0.05	1	0.12	0.18	−0.11	−0.05	0.07	0.19	−0.12	0.1	0.06	−0.14
cp	−0.07	−0.05	1	0.05	−0.08	0.09	0.04	0.3	−0.39	−0.15	0.12	−0.18	−0.16	0.43
sex	−0.1	1	−0.05	−0.06	−0.2	0.05	−0.06	−0.04	0.14	0.1	−0.03	0.12	0.21	−0.28
age	1	−0.1	−0.07	0.28	0.21	0.12	−0.12	−0.4	0.1	0.21	−0.17	0.28	0.07	−0.23

Features

Correlation Meter −1.0 −0.5 0.0 0.5 1.0

```r
#PCA with categoriacal values
pca_facto <- factor_heart[, sapply(factor_heart, class) != "factor"]
#Some categorical values can be added as supplementary
#pca_facto$sex <- factor_heart$sex
#pca_facto$ca <- factor_heart$ca
pca_facto$disease <- heart_disease$target
pca_facto$disease[pca_facto$disease==0] <- "Yes"
pca_facto$disease[pca_facto$disease==1] <- "No"

pca_facto_heart <- PCA(pca_facto, quali.sup = 6, scale.unit = TRUE,  graph = TRUE)
```
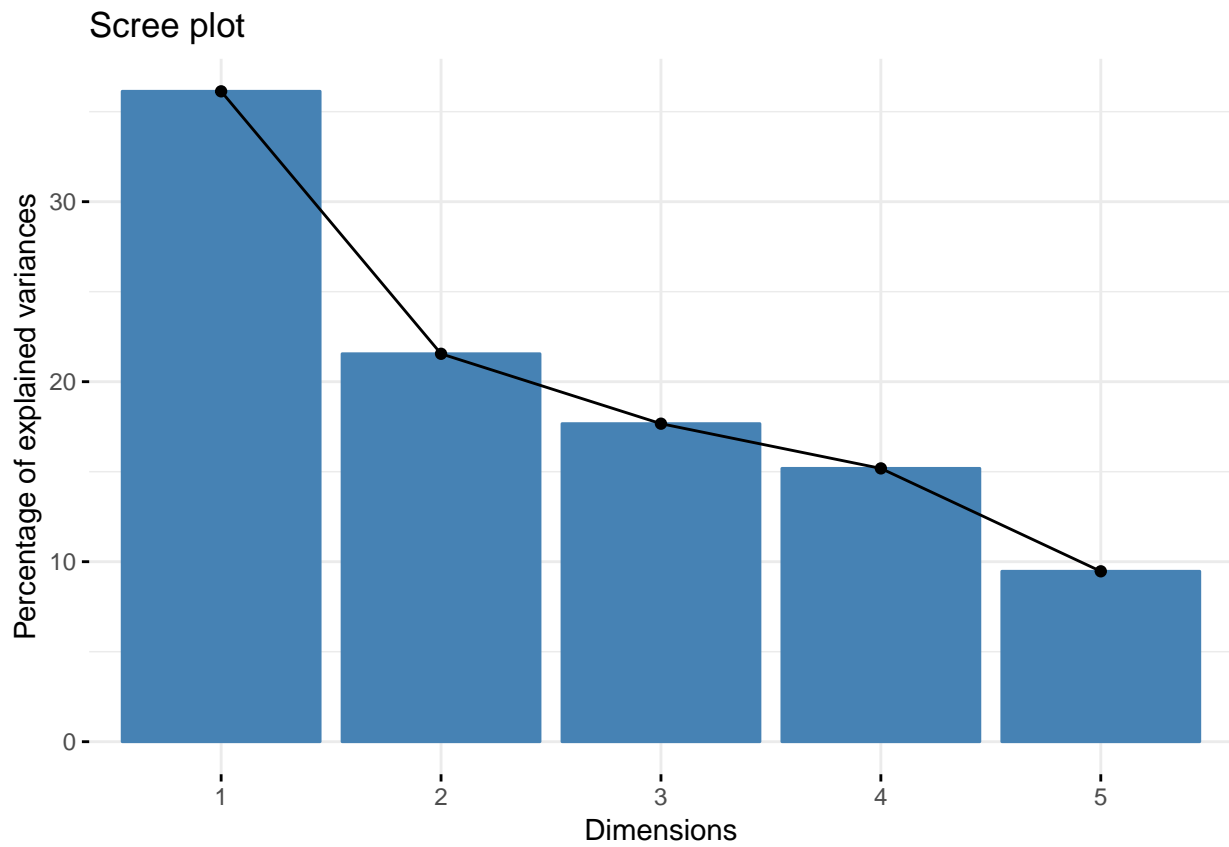
Individuals factor map (PCA)
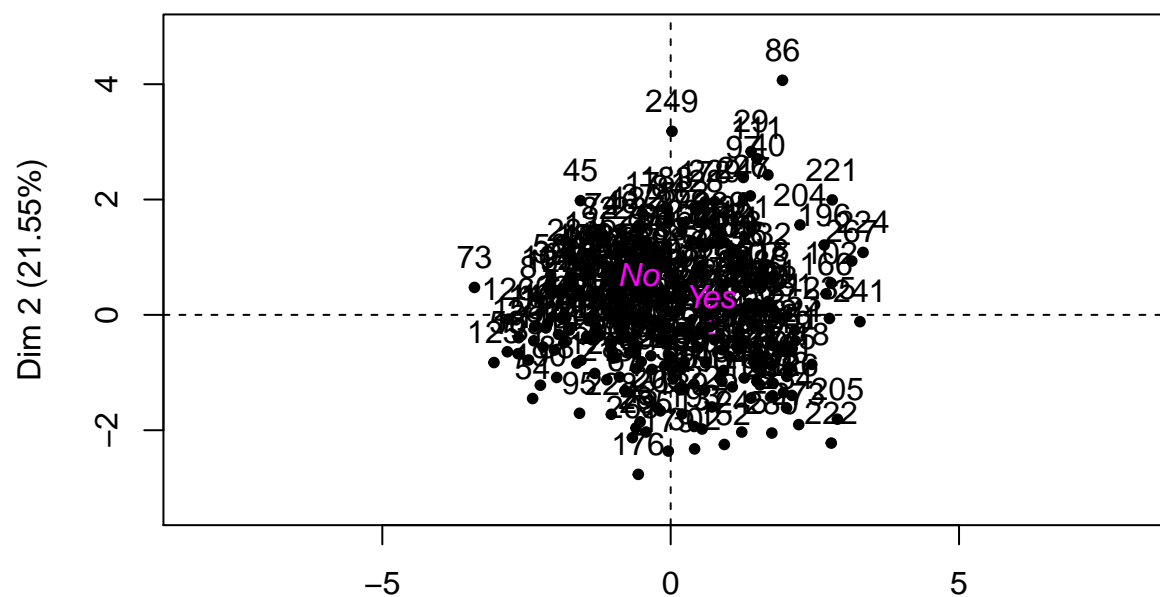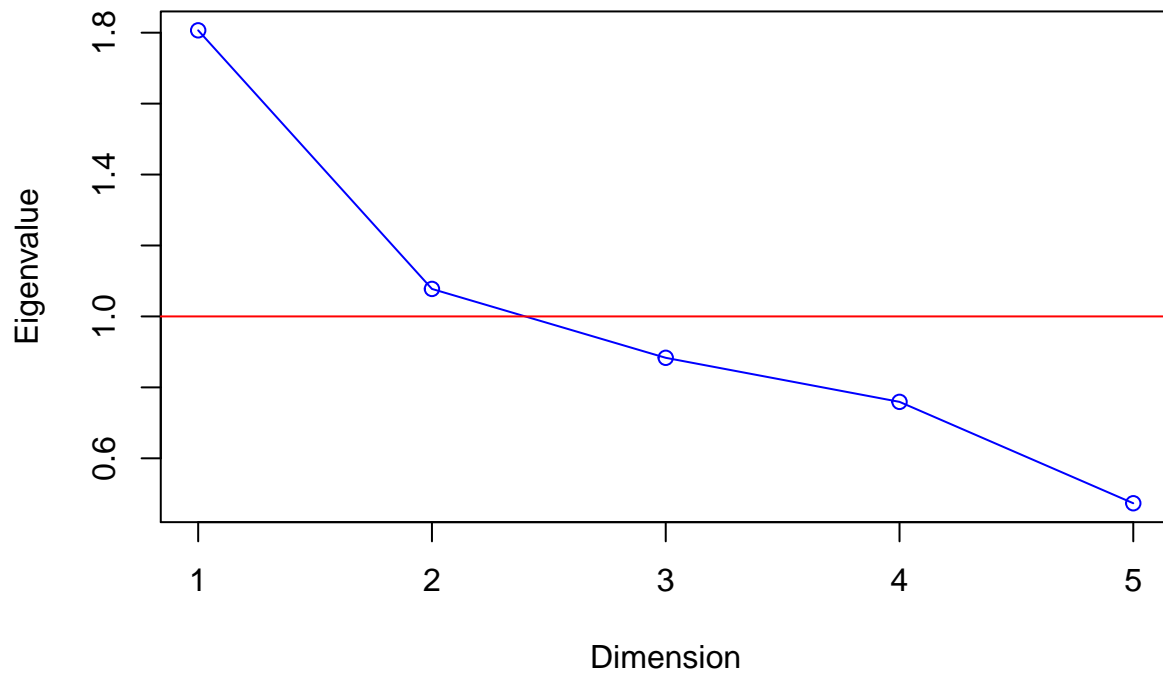

Variables factor map (PCA)
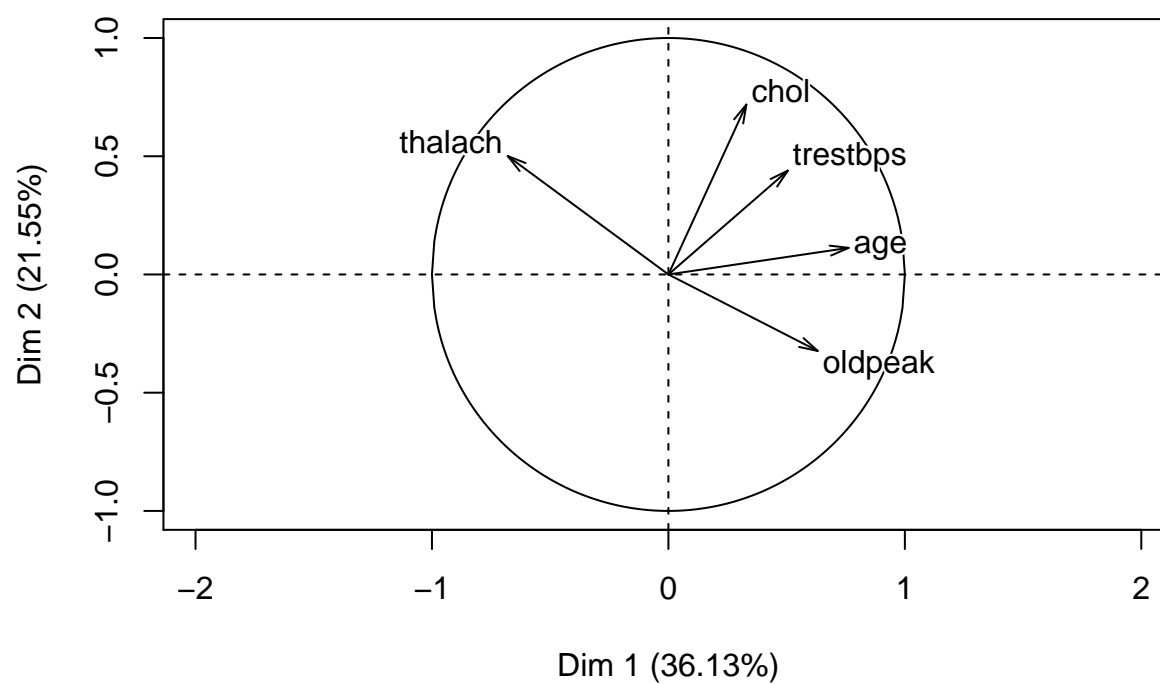

```
#Screeplots
fviz_screeplot(pca_facto_heart, addlabels = FALSE)
```

Scree plot


```
eigen_values <- pca_facto_heart$eig[,1]
plot(eigen_values, type="o", main="Screeplot",
     xlab='Dimension', ylab='Eigenvalue', col='blue')
abline(h=1,col="red")
```
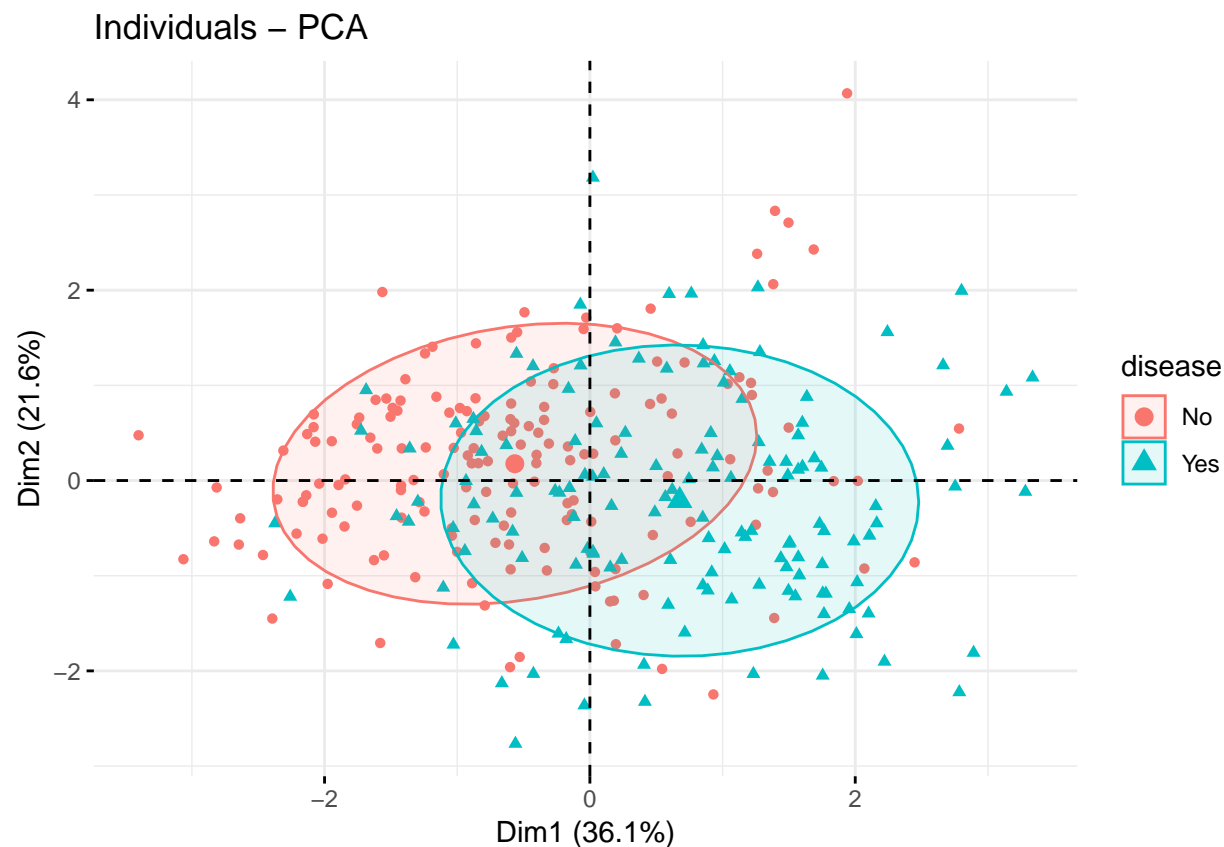
Screeplot


```
#Represented in Rp
#quali.sup -> Every modality is the centroide of the respective individuals having chosen that modality
fviz_pca_ind(pca_facto_heart, habillage = 6, geom = "point", label="quali",addEllipses =TRUE, ellipse.le
```

Individuals – PCA


```
plot.PCA(pca_facto_heart, quali.sup = 6, scale.unit = TRUE,choix = 'ind',label="quali")
```

```
## Warning in plot.window(...): "quali.sup" is not a graphical parameter

## Warning in plot.window(...): "scale.unit" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "quali.sup" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "scale.unit" is not a graphical
## parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "quali.sup" is
## not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "scale.unit"
## is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "quali.sup" is
## not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "scale.unit"
## is not a graphical parameter

## Warning in box(...): "quali.sup" is not a graphical parameter

## Warning in box(...): "scale.unit" is not a graphical parameter

## Warning in title(...): "quali.sup" is not a graphical parameter

## Warning in title(...): "scale.unit" is not a graphical parameter

## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "quali.sup" is not a graphical parameter
```

```
## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "scale.unit" is not a graphical parameter

## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "quali.sup" is not a graphical parameter

## Warning in int_abline(a = a, b = b, h = h, v = v, untf = untf, ...):
## "scale.unit" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "quali.sup" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "scale.unit" is not
## a graphical parameter

## Warning in text.default(xy, labels, cex = cex, ...): "quali.sup" is not a
## graphical parameter

## Warning in text.default(xy, labels, cex = cex, ...): "scale.unit" is not a
## graphical parameter
```
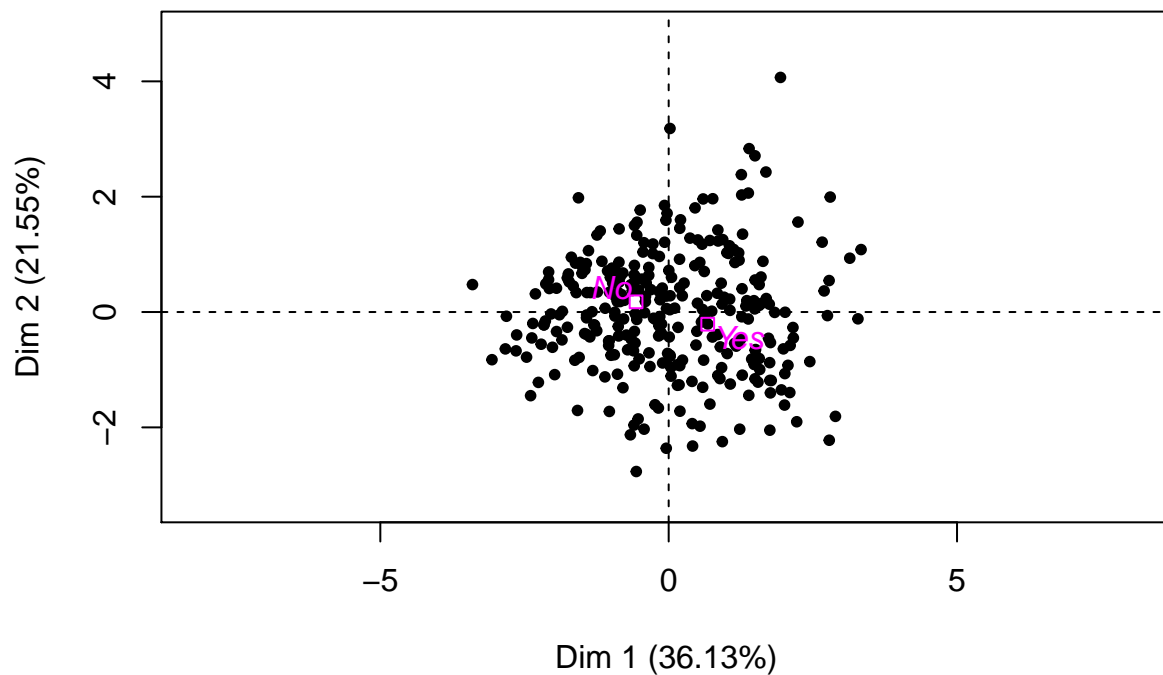
## Individuals factor map (PCA)


```r
#Represented in Rn
#Projection of variables, show correlation between principal components
fviz_pca_var(pca_facto_heart, geom = c("arrow", "text"), col.var = "cos2")#By quality of representation
```

11

Variables – PCA
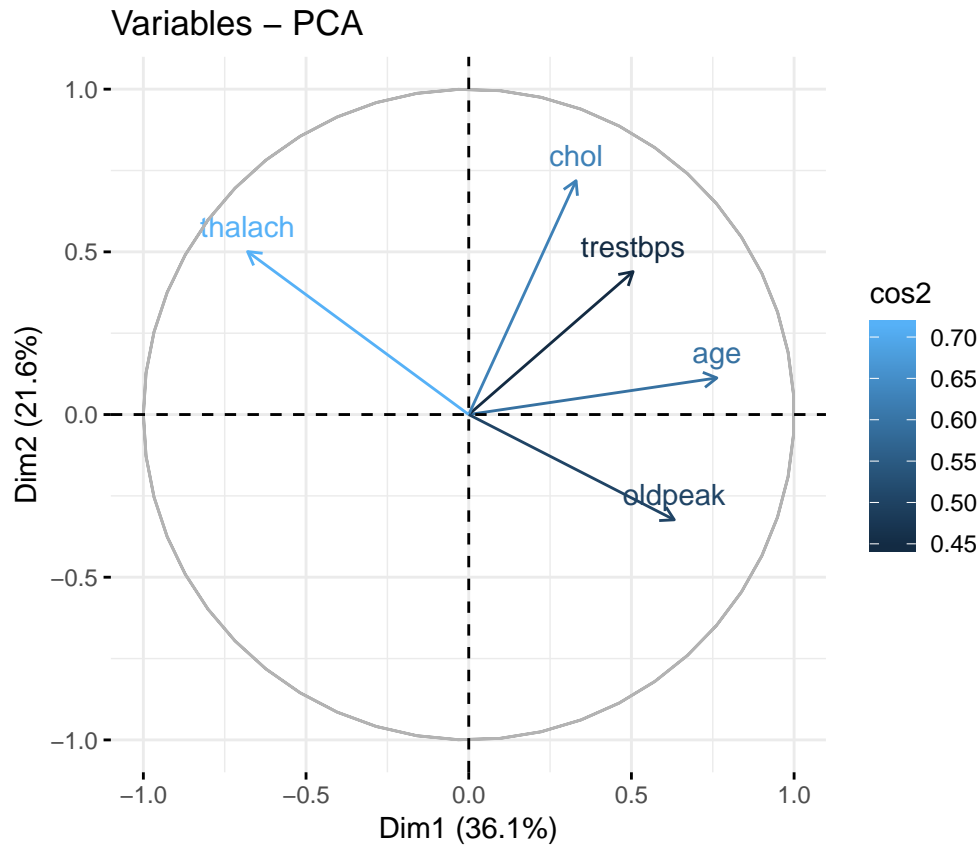
```
proj_indiv <- pca_facto_heart$ind$coord[,1:2] #individual projections on 1st factorial plane
#Clustering
hc_ward = hclust(dist(proj_indiv),method = "ward.D")
plot(hc_ward, main= "HC using Ward Agglomeration method", xlab="",sub="",cex=.9,  labels=FALSE)
abline(h=60)
rect.hclust(hc_ward, k = 3, border = 2:6)
```
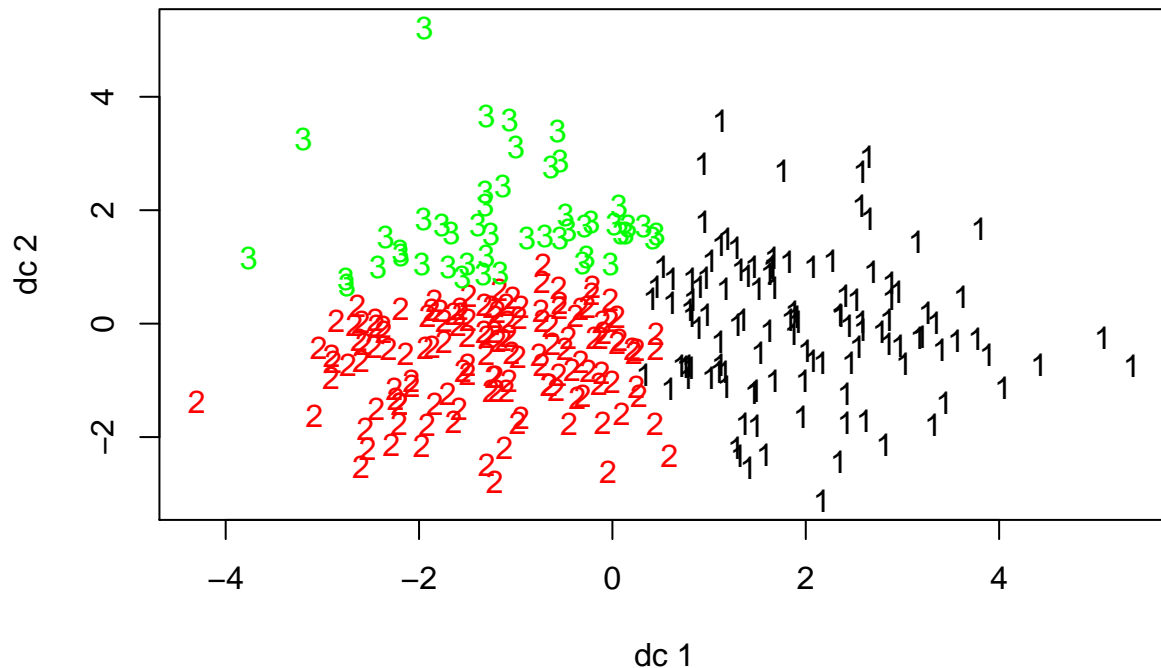
HC using Ward Agglomeration method
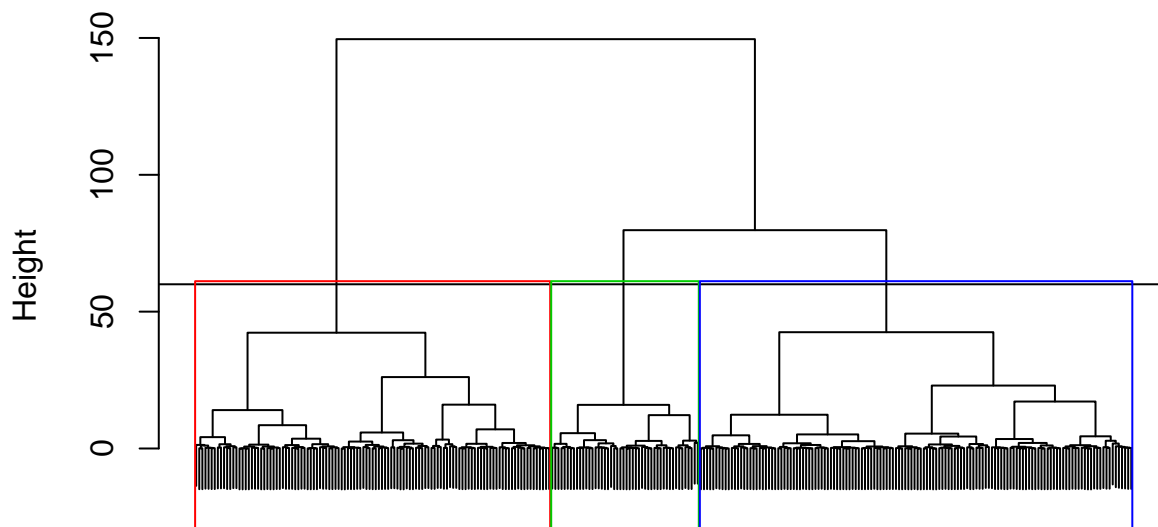
```r
#Association of individuals to clusters
classes <- cutree(hc_ward, h=50) #Depending on the height, number of clusters is chosen
plotcluster(proj_indiv, classes,main="Projections of individuals in Hierarchical Clustering of 3 classes
```

Projections of individuals in Hierarchical Clustering of 3 classes


```r
get_centroids <- function(classes, n_classes){
  centroids <- NULL
  for(k in 1:n_classes){
    centroids <- rbind(centroids, colMeans(proj_indiv[classes == k, , drop = FALSE]))
  }
  return(centroids)
}
centroids <- get_centroids(classes, 3)
```
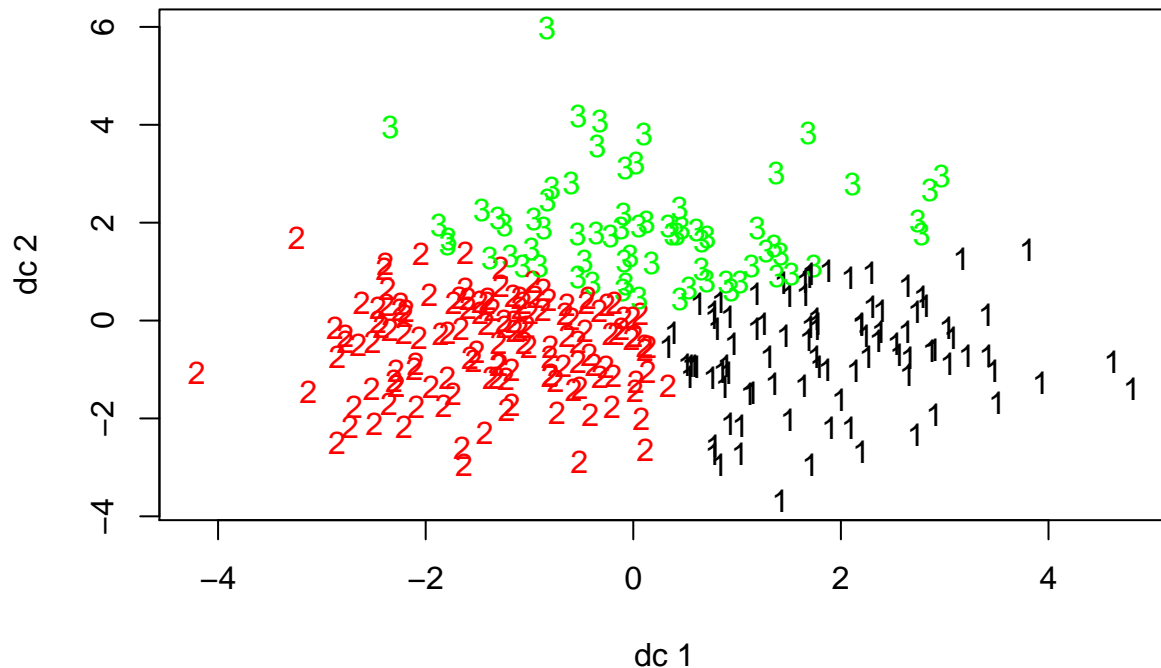
```r
#k_mean needs centroid of clusters
k_mean <- kmeans(proj_indiv, centroids)
plotcluster(proj_indiv, k_mean$cluster,main="Projections of individuals in K-means Clustering of 3 class
```

Projections of individuals in K–means Clustering of 3 classes


```
cal_idx_before <- calinhara(proj_indiv,classes,cn=max(classes))
cal_idx_after <- calinhara(proj_indiv,k_mean$cluster,cn=max(k_mean$cluster))

print(cal_idx_before)
```

```
## [1] 198.1154
```

```
print(cal_idx_after)
```

```
## [1] 226.1952
```

```
#Improvement
```

```
Calinski_Harabassza <- function (projections, hc, kind, n_classes){
  classes <- cutree(hc, k=n_classes)
  centroids <- get_centroids(classes, n_classes)
  if(kind=='hc'){
    index <- calinhara(proj_indiv,classes,cn=max(classes))
  }
  if(kind=='kmeans'){
    kmeans_classes <- kmeans(proj_indiv, centers = centroids)$cluster
    index <-calinhara(proj_indiv,kmeans_classes,cn=max(kmeans_classes))
  }
  return(index)
}
get_indexes <- function(until, kind){
  indexes <- c()
  for (n_classes in 2:until){
    indexes <- c(indexes, Calinski_Harabassza(proj_indiv, hc_ward, kind, n_classes))
  }
  return(indexes)
```
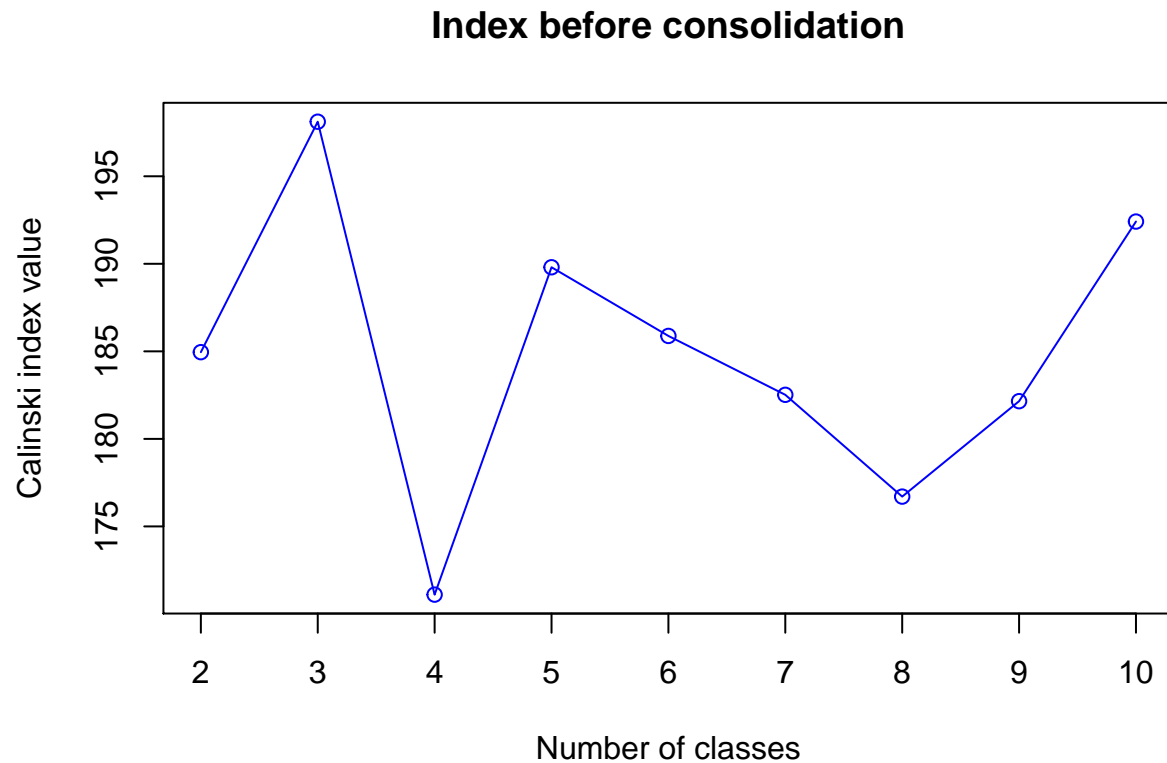
```
}
```

```
indexes_before <- get_indexes(10, 'hc')
plot(indexes_before, type = "o", xlab = 'Number of classes', ylab = 'Calinski index value'
, main = 'Index before consolidation', col = 'blue', xaxt
= "n")
axis(1, at=1:9, labels = c(2, 3, 4, 5, 6, 7,8,9,10))
```

Index before consolidation

Number of classes

```
indexes_after <- get_indexes(10, 'kmeans')
plot(indexes_after, type = "o", xlab = 'Number of classes', ylab = 'Calinski index value'
, main = 'Index after consolidation', col = 'blue', xaxt
= "n")
axis(1, at=1:9, labels = c(2, 3, 4, 5, 6, 7,8,9,10))
```

Index after consolidation
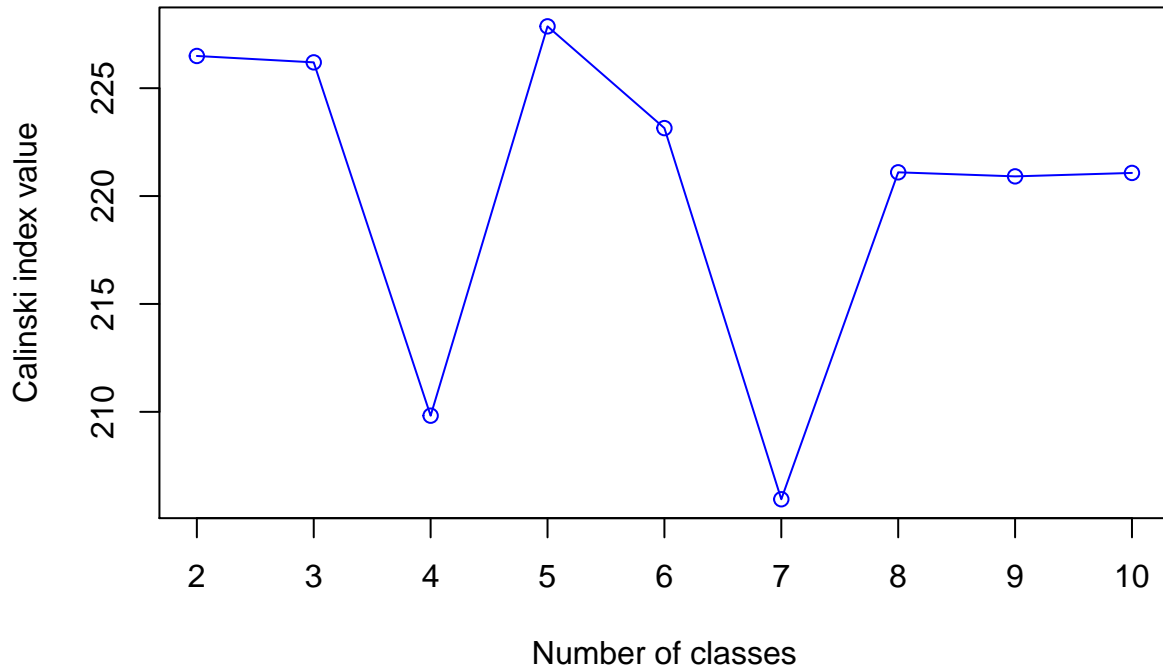

```
first_factorial <- proj_indiv
df <- data.frame(first_factorial, Class = as.factor(k_mean$cluster))
catdes(df, num.var = length(df), proba = 0.05, row.w = NULL)
```

```
##
## Link between the cluster variable and the quantitative variables
## ================================================================
##             Eta2      P-value
## Dim.1 0.6316289 8.766734e-66
## Dim.2 0.5503737 8.467640e-53
##
## Description of each cluster by quantitative variables
## =====================================================
## $`1`
##          v.test Mean in category  Overall mean sd in category Overall sd
## Dim.1   9.014585        1.0477037  2.852366e-15       0.8712730   1.344089
## Dim.2 -10.516103       -0.9439256 -3.090499e-16       0.6923393   1.038050
##            p.value
## Dim.1 1.976176e-19
## Dim.2 7.282354e-26
##
## $`2`
##         v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 -13.79634        -1.151379 2.852366e-15       0.7373691   1.344089
##            p.value
## Dim.1 2.681253e-43
##
## $`3`
##         v.test Mean in category  Overall mean sd in category Overall sd
## Dim.2 10.780147        1.1748120 -3.090499e-16       0.7675912   1.038050
```

```
## Dim.1  6.454662          0.9108084  2.852366e-15      0.8858076   1.344089
##              p.value
## Dim.2 4.272002e-27
## Dim.1 1.084604e-10
```

```r
factor_heart$disease[factor_heart$target==0] <- "Yes"
factor_heart$disease[factor_heart$target==1] <- "No"
factor_heart$target <- NULL

factor_heart2 <- factor_heart
factor_heart2$age<-cut(factor_heart2$age, seq(0,80,10), right=FALSE)
factor_heart2$age <- paste("Age", factor_heart2$age, sep="_")
min(factor_heart2$oldpeak)
```

```
## [1] 0
```

```r
factor_heart2$oldpeak<-cut(factor_heart2$oldpeak, seq(0,7,1), right=FALSE)
factor_heart2$oldpeak <- paste("Oldp", factor_heart2$oldpeak, sep="_")
factor_heart2$thalach<-cut(factor_heart2$thalach, seq(70,220,20), right=FALSE)
factor_heart2$thalach <- paste("thalach", factor_heart2$thalach, sep="_")
factor_heart2$trestbps<-cut(factor_heart2$trestbps, seq(80,220,20), right=FALSE)
factor_heart2$trestbps <- paste("thres", factor_heart2$trestbps, sep="_")
factor_heart2$chol<-cut(factor_heart2$chol, seq(100,600,100), right=FALSE)
factor_heart2$chol <- paste("Col", factor_heart2$chol, sep="_")
#factor_heart2$age <- NULL
kable(head(factor_heart2))
```

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope
Age_[60,70)	1	3	thres_[140,160)	Col_[200,300)	1	0	thalach_[150,170)	0	Oldp_[2,3)	0
Age_[30,40)	1	2	thres_[120,140)	Col_[200,300)	0	1	thalach_[170,190)	0	Oldp_[3,4)	0
Age_[40,50)	0	1	thres_[120,140)	Col_[200,300)	0	0	thalach_[170,190)	0	Oldp_[1,2)	2
Age_[50,60)	1	1	thres_[120,140)	Col_[200,300)	0	1	thalach_[170,190)	0	Oldp_[0,1)	2
Age_[50,60)	0	0	thres_[120,140)	Col_[300,400)	0	1	thalach_[150,170)	1	Oldp_[0,1)	2
Age_[50,60)	1	0	thres_[140,160)	Col_[100,200)	0	1	thalach_[130,150)	0	Oldp_[0,1)	1

```r
mcaHeart <- MCA(factor_heart2,ncp=7,
              #quanti.sup=c(10),
              quali.sup=c(14),
              excl=NULL,
              graph = FALSE,
              level.ventil = 0.00,
              axes = c(1,2),
              row.w = NULL,
              method="Indicator",
              na.method="NA",
              tab.disj=NULL)

# mcaHeart <- MCA(factor_heart,ncp=7,
#              quanti.sup=c(1,4,5,8,10),
#              quali.sup=c(14),
#              excl=NULL,
#              graph = FALSE,
#              level.ventil = 0.00,
#              axes = c(1,2),
```

```
#              row.w = NULL,
#              method="Indicator",
#              na.method="NA",
#              tab.disj=NULL)
summary(mcaHeart)
```

```
##
## Call:
## MCA(X = factor_heart2, ncp = 7, quali.sup = c(14), excl = NULL,
##     graph = FALSE, level.ventil = 0, axes = c(1, 2), row.w = NULL,
##     method = "Indicator", na.method = "NA", tab.disj = NULL)
##
##
## Eigenvalues
##                         Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance                0.264   0.142   0.134   0.129   0.126   0.119
## % of var.               7.805   4.192   3.961   3.803   3.736   3.513
## Cumulative % of var.    7.805  11.997  15.958  19.761  23.496  27.009
##                         Dim.7   Dim.8   Dim.9  Dim.10  Dim.11  Dim.12
## Variance                0.110   0.108   0.106   0.101   0.097   0.096
## % of var.               3.263   3.188   3.131   2.981   2.869   2.844
## Cumulative % of var.   30.273  33.460  36.592  39.573  42.442  45.286
##                        Dim.13  Dim.14  Dim.15  Dim.16  Dim.17  Dim.18
## Variance                0.093   0.090   0.088   0.085   0.081   0.079
## % of var.               2.759   2.661   2.587   2.521   2.390   2.332
## Cumulative % of var.   48.045  50.706  53.293  55.814  58.204  60.536
##                        Dim.19  Dim.20  Dim.21  Dim.22  Dim.23  Dim.24
## Variance                0.079   0.077   0.075   0.072   0.070   0.068
## % of var.               2.322   2.283   2.213   2.140   2.079   1.997
## Cumulative % of var.   62.858  65.141  67.355  69.495  71.574  73.571
##                        Dim.25  Dim.26  Dim.27  Dim.28  Dim.29  Dim.30
## Variance                0.066   0.064   0.063   0.060   0.057   0.055
## % of var.               1.950   1.891   1.850   1.782   1.693   1.635
## Cumulative % of var.   75.521  77.412  79.262  81.043  82.736  84.371
##                        Dim.31  Dim.32  Dim.33  Dim.34  Dim.35  Dim.36
## Variance                0.054   0.052   0.050   0.046   0.044   0.040
## % of var.               1.596   1.543   1.491   1.364   1.291   1.192
## Cumulative % of var.   85.968  87.511  89.002  90.366  91.657  92.849
##                        Dim.37  Dim.38  Dim.39  Dim.40  Dim.41  Dim.42
## Variance                0.039   0.037   0.034   0.031   0.029   0.026
## % of var.               1.138   1.085   1.008   0.916   0.852   0.771
## Cumulative % of var.   93.987  95.072  96.080  96.996  97.848  98.619
##                        Dim.43  Dim.44
## Variance                0.024   0.023
## % of var.               0.712   0.669
## Cumulative % of var.   99.331 100.000
##
## Individuals (the 10 first)
##                   Dim.1    ctr    cos2    Dim.2    ctr    cos2    Dim.3    ctr
## 1             |   0.527  0.347  0.055 |  0.515  0.618  0.052 |  0.515  0.652
## 2             |  -0.428  0.229  0.039 |  0.110  0.028  0.003 |  0.790  1.537
## 3             |  -0.681  0.579  0.257 |  0.057  0.008  0.002 | -0.091  0.020
## 4             |  -0.718  0.645  0.376 | -0.104  0.025  0.008 |  0.229  0.129
## 5             |  -0.262  0.086  0.043 |  0.111  0.029  0.008 | -0.379  0.354
```

```
## 6              |   0.229   0.065   0.020 |  -0.107   0.027   0.004 |   0.451   0.500
## 7              |  -0.206   0.053   0.026 |   0.191   0.085   0.022 |  -0.453   0.505
## 8              |  -0.656   0.538   0.274 |  -0.209   0.101   0.028 |   0.359   0.318
## 9              |  -0.194   0.047   0.014 |   0.300   0.210   0.032 |   0.204   0.102
## 10             |  -0.484   0.293   0.126 |  -0.003   0.000   0.000 |   0.079   0.015
##                  cos2
## 1               0.052 |
## 2               0.132 |
## 3               0.005 |
## 4               0.038 |
## 5               0.090 |
## 6               0.076 |
## 7               0.124 |
## 8               0.082 |
## 9               0.015 |
## 10              0.003 |
##
## Categories (the 10 first)
##                 Dim.1    ctr   cos2 v.test   Dim.2    ctr   cos2 v.test
## Age_[20,30) | -1.800   0.311  0.011 -1.800 |   0.241  0.010  0.000   0.241 |
## Age_[30,40) | -1.140   1.873  0.068 -4.521 |  -0.337  0.305  0.006 -1.338 |
## Age_[40,50) | -0.641   2.839  0.128 -6.214 |  -0.230  0.679  0.016 -2.228 |
## Age_[50,60) |  0.158   0.299  0.017  2.298 |  -0.011  0.003  0.000 -0.165 |
## Age_[60,70) |  0.532   2.178  0.102  5.540 |   0.246  0.865  0.022  2.559 |
## Age_[70,80) |  0.271   0.071  0.003  0.870 |   0.310  0.172  0.003  0.996 |
## sex_0       | -0.294   0.798  0.040 -3.481 |   0.588  5.932  0.160  6.955 |
## sex_1       |  0.136   0.370  0.040  3.481 |  -0.273  2.751  0.160 -6.955 |
## cp_0        |  0.638   5.586  0.363 10.474 |  -0.164  0.687  0.024 -2.692 |
## cp_1        | -0.957   4.396  0.181 -7.390 |   0.035  0.011  0.000  0.271 |
##               Dim.3    ctr   cos2 v.test
## Age_[20,30)   4.476  3.794  0.066  4.476 |
## Age_[30,40)   1.347  5.156  0.095  5.344 |
## Age_[40,50)   0.305  1.271  0.029  2.962 |
## Age_[50,60)   0.059  0.081  0.002  0.854 |
## Age_[60,70) -0.618  5.780  0.137 -6.429 |
## Age_[70,80) -0.459  0.398  0.007 -1.472 |
## sex_0       -0.721  9.462  0.241 -8.538 |
## sex_1        0.335  4.388  0.241  8.538 |
## cp_0        -0.031  0.026  0.001 -0.505 |
## cp_1         0.247  0.579  0.012  1.911 |
##
## Categorical variables (eta2)
##              Dim.1 Dim.2 Dim.3
## age        | 0.260 0.038 0.287 |
## sex        | 0.040 0.160 0.241 |
## cp         | 0.421 0.030 0.047 |
## trestbps   | 0.119 0.313 0.068 |
## chol       | 0.014 0.140 0.228 |
## fbs        | 0.018 0.053 0.002 |
## restecg    | 0.096 0.068 0.002 |
## thalach    | 0.519 0.119 0.314 |
## exang      | 0.336 0.061 0.001 |
## oldpeak    | 0.507 0.476 0.192 |
##
```

```
## Supplementary categories
##                 Dim.1   cos2  v.test    Dim.2   cos2  v.test     Dim.3
## No       |     -0.619  0.458 -11.767 |   0.142  0.024   2.694 |  -0.040
## Yes      |      0.740  0.458  11.767 |  -0.170  0.024  -2.694 |   0.048
##                  cos2  v.test
## No              0.002  -0.766 |
## Yes             0.002   0.766 |
##
## Supplementary categorical variables (eta2)
##                 Dim.1 Dim.2 Dim.3
## disease   | 0.458 0.024 0.002 |
```

Individuals – MCA

Categories – MCA

Variables – MCA

MCA – Biplot

Scree plot