ejercicio 7

Javier

30/11/2020

```r
library(cluster)
library(stats)
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.3

csv_data <- read.csv2("divorce_data.csv")
summary(csv_data)

##     DIVYEAR          momint          dadint         momclose
## Min.   :89.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:89.00   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000
## Median :90.00   Median :1.000   Median :2.000   Median :2.000
## Mean   :89.68   Mean   :1.809   Mean   :2.522   Mean   :1.809
## 3rd Qu.:90.00   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.000
## Max.   :90.00   Max.   :4.000   Max.   :4.000   Max.   :4.000
##                                 NA's   :1
##    depression        livewith       gethitched
## Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:2.000   1st Qu.:1.000   1st Qu.:2.000
## Median :3.000   Median :1.000   Median :2.000
## Mean   :2.851   Mean   :1.326   Mean   :2.213
## 3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:2.000
## Max.   :4.000   Max.   :2.000   Max.   :8.000
##                 NA's   :1

csv_data$DIVYEAR<-ifelse(csv_data$DIVYEAR==89, 0, 1)
table(csv_data$livewithmom)

## < table of extent 0 >

csv_data[csv_data$livewithmom==9, ]

## [1] DIVYEAR    momint    dadint    momclose   depression livewith
## gethitched
## <0 rows> (or 0-length row.names)

# Cluster con distribución normal
normal_dist <- scale(csv_data)
cluster <- hclust(dist(normal_dist))
cluster

##
## Call:
```
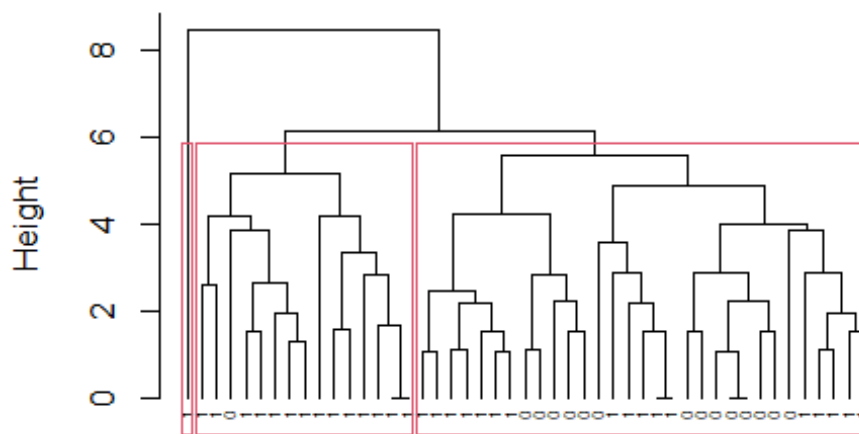
```
## hclust(d = dist(normal_dist))
##
## Cluster method   : complete
## Distance         : euclidean
## Number of objects: 47

# Año en el que se divorciaron (columna como label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$DIVYEAR)
rect.hclust(cluster,k=3)
```
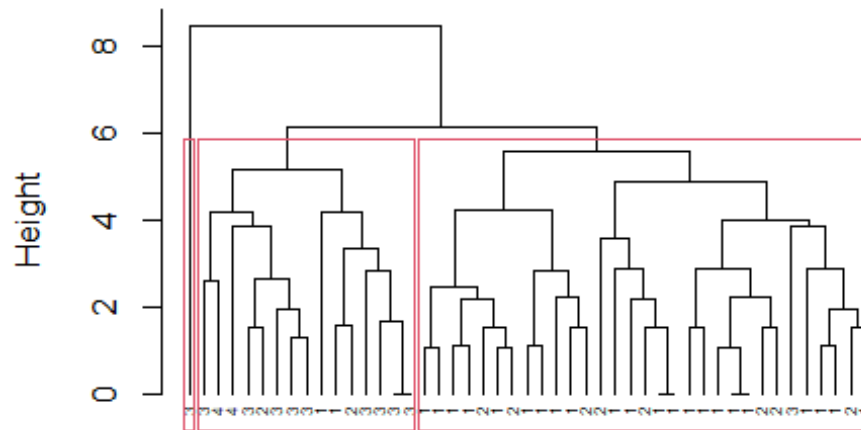

Cluster Dendrogram

dist(normal_dist)
hclust (*, "complete")

```
# Relaccion con la madre (columna como label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$momint, main = "Relaccion con
la Madre")
rect.hclust(cluster,k=3)
```
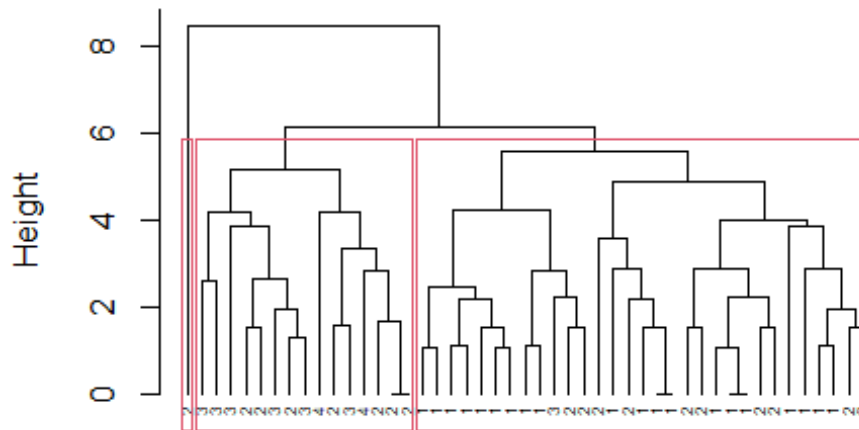
Relaccion con la Madre

dist(normal_dist)
hclust (*, "complete")

```r
# Cercania con la madre (columna como Label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$momclose,main = "Cercania con
la Madre")
rect.hclust(cluster,k=3)
```
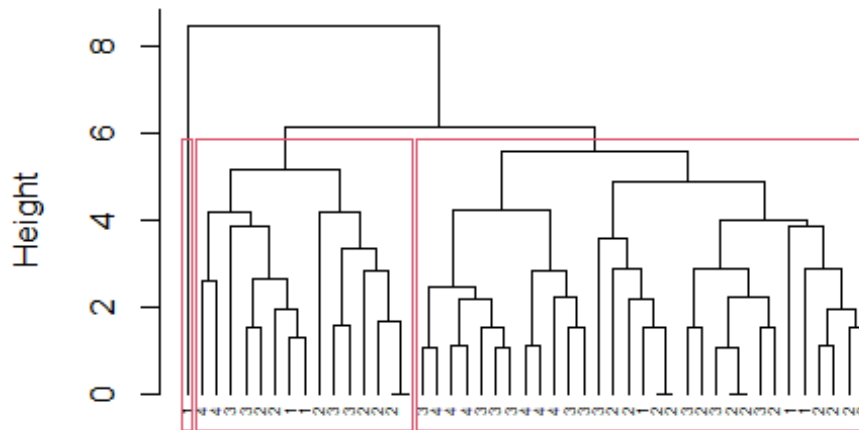
Cercania con la Madre

dist(normal_dist)
hclust (*, "complete")

```
# Relaccion con la padre (columna como label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$dadint,main = "Relaccion con
el Padre")
rect.hclust(cluster,k=3)
```
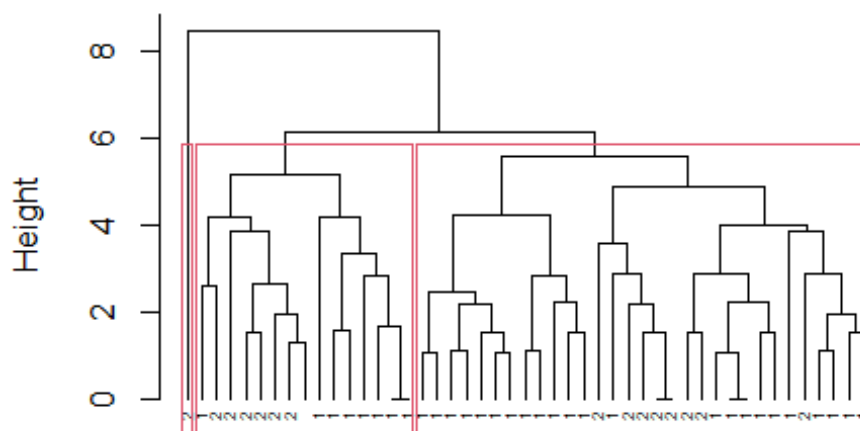
Relaccion con el Padre

dist(normal_dist)
hclust (*, "complete")

```
# Con quien vive (columna como label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$livewith, main = "Tipos de
Cohabitacion")
rect.hclust(cluster,k=3)
```
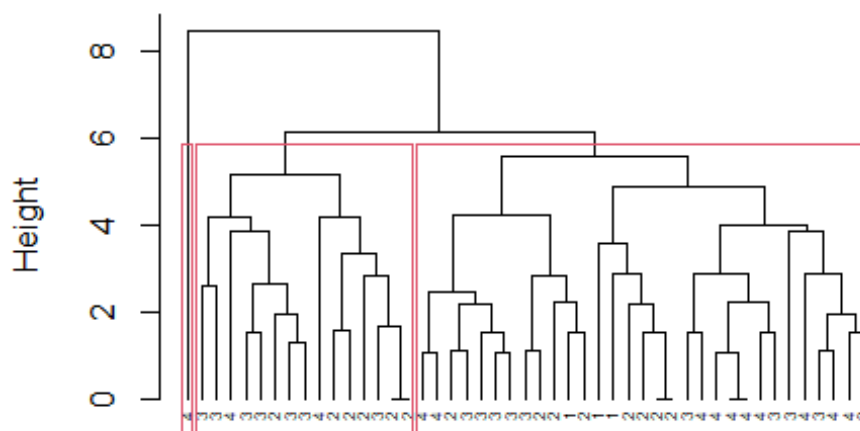
Tipos de Cohabitacion

dist(normal_dist)
hclust (*, "complete")

```r
# Sentimiento de depresion (columna como label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$depression, main =
"Sentimiento de Depresion")
rect.hclust(cluster,k=3)
```
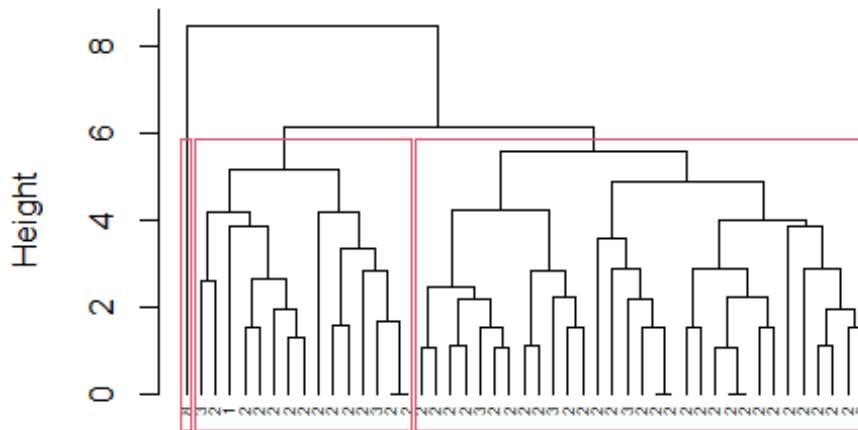
Sentimiento de Depresion

dist(normal_dist)
hclust (*, "complete")

```r
# Plan de matrimonio(columna como label)
plot(cluster,hang=-1,cex=.5,labels=csv_data$gethitched, main = "Plan de
Matrimonio")
rect.hclust(cluster,k=3)
```

Plan de Matrimonio

dist(normal_dist)
hclust (*, "complete")

```
distribucion <- as.data.frame(lapply(csv_data,scale))
str(distribucion)

## 'data.frame':    47 obs. of  7 variables:
##  $ DIVYEAR   : num  0.677 0.677 -1.445 0.677 -1.445 ...
##  $ momint    : num  1.258 1.258 -0.854 1.258 -0.854 ...
##  $ dadint    : num  -0.557 -0.557 0.511 1.578 0.511 ...
##  $ momclose  : num  0.225 1.401 -0.951 1.401 0.225 ...
##  $ depression: num  0.164 -0.937 1.265 0.164 -2.038 ...
##  $ livewith  : num  -0.688 1.422 -0.688 -0.688 -0.688 ...
##  $ gethitched: num  0.846 -0.229 -0.229 0.846 -0.229 ...

# Creacion de semilla
set.seed(321)

# Distribution Cluster
cluster_dist <- kmeans(distribucion[,colSums(is.na(distribucion))==0],3)
cluster_dist$size

## [1] 15 20 12

cluster_dist$centers

##       DIVYEAR      momint    momclose   depression   gethitched
## 1   0.5358437   1.1170605   1.0089641   0.01717647   0.27271371
## 2  -0.2777055  -0.4840346  -0.5978738   0.71438512  -0.17482624
## 3  -0.2069621  -0.5896013  -0.2647489  -1.21211246  -0.04951506
```

```r
# Creamos el grafico
par(mfrow=c(1, 1), mar=c(4, 4, 4, 2))
myColors <- c("blue", "red", "green", "orange", "gray", "violet",
"yellow")
barplot(t(cluster_dist$centers), beside = TRUE, xlab="cluster",
ylab="value", col = myColors)
legend("topright", ncol=2, legend = c("DIVYEAR", "momint", "dadint",
"momclose", "depression", "livewithmom", "gethitched"), fill = myColors)
```