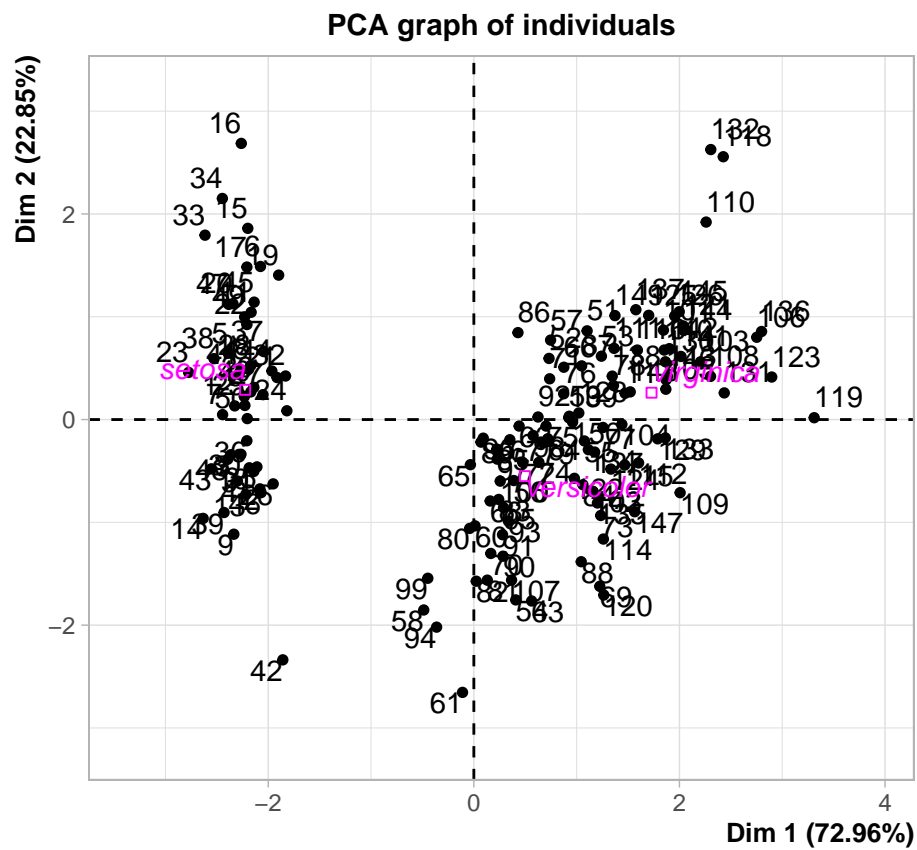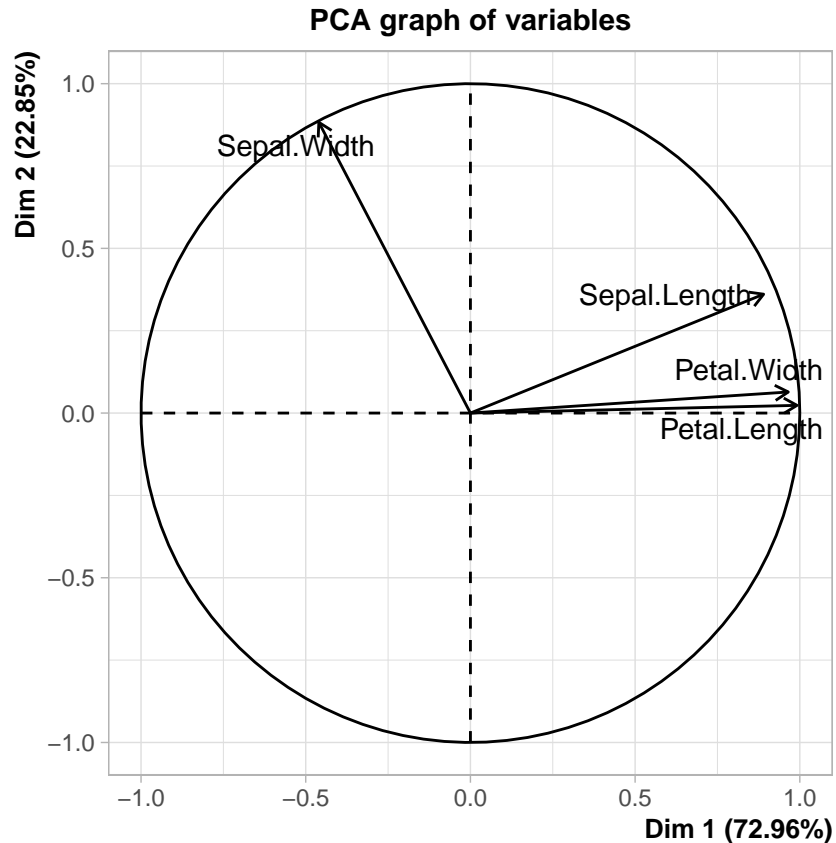# ciencia_dades_iris.R

jcortes

2023-11-30

```r
rm(list=ls())
```

```r
library(FactoMineR)
```

```r
d <- read.csv('https://raw.githubusercontent.com/jordicortes40/PE_Bloc_D/main/Dades/iris.csv')
# iris <- read.csv('../Dades/iris.csv')
```

```r
(pca.iris <- PCA(iris,quali.sup=c(5: 5)))
```


**PCA graph of individuals**

PCA graph of variables


```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 150 individuals, described by 5 variables
## *The results are available in the following objects:
##
##    name                 description
## 1  "$eig"               "eigenvalues"
## 2  "$var"               "results for the variables"
## 3  "$var$coord"         "coord. for the variables"
## 4  "$var$cor"           "correlations variables - dimensions"
## 5  "$var$cos2"          "cos2 for the variables"
## 6  "$var$contrib"       "contributions of the variables"
## 7  "$ind"               "results for the individuals"
## 8  "$ind$coord"         "coord. for the individuals"
## 9  "$ind$cos2"          "cos2 for the individuals"
## 10 "$ind$contrib"       "contributions of the individuals"
## 11 "$quali.sup"         "results for the supplementary categorical variables"
## 12 "$quali.sup$coord"   "coord. for the supplementary categories"
## 13 "$quali.sup$v.test"  "v-test of the supplementary categories"
## 14 "$call"              "summary statistics"
## 15 "$call$centre"       "mean of the variables"
## 16 "$call$ecart.type"   "standard error of the variables"
## 17 "$call$row.w"        "weights for the individuals"
## 18 "$call$col.w"        "weights for the variables"
```
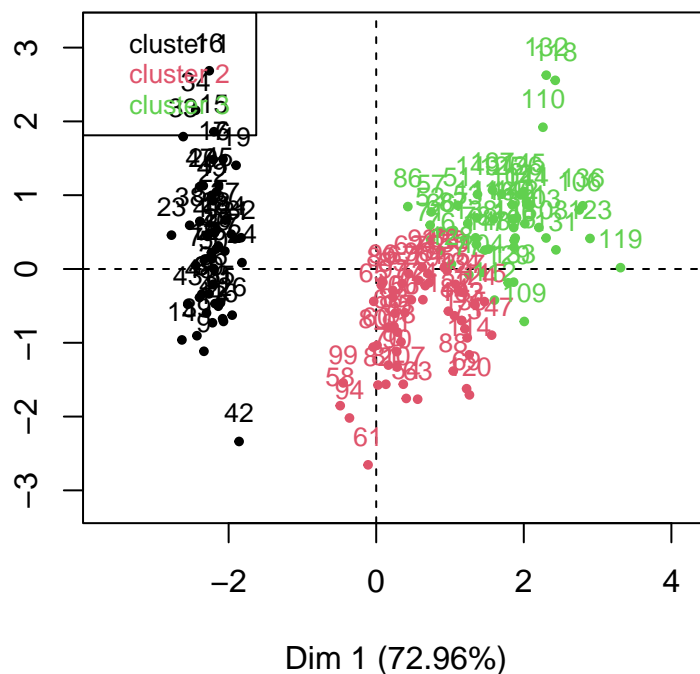
```
(hcpc.iris <- HCPC(pca.iris,nb.clust=-1))
```

Hierarchical Clustering

Hierarchical Classification

inertia gain
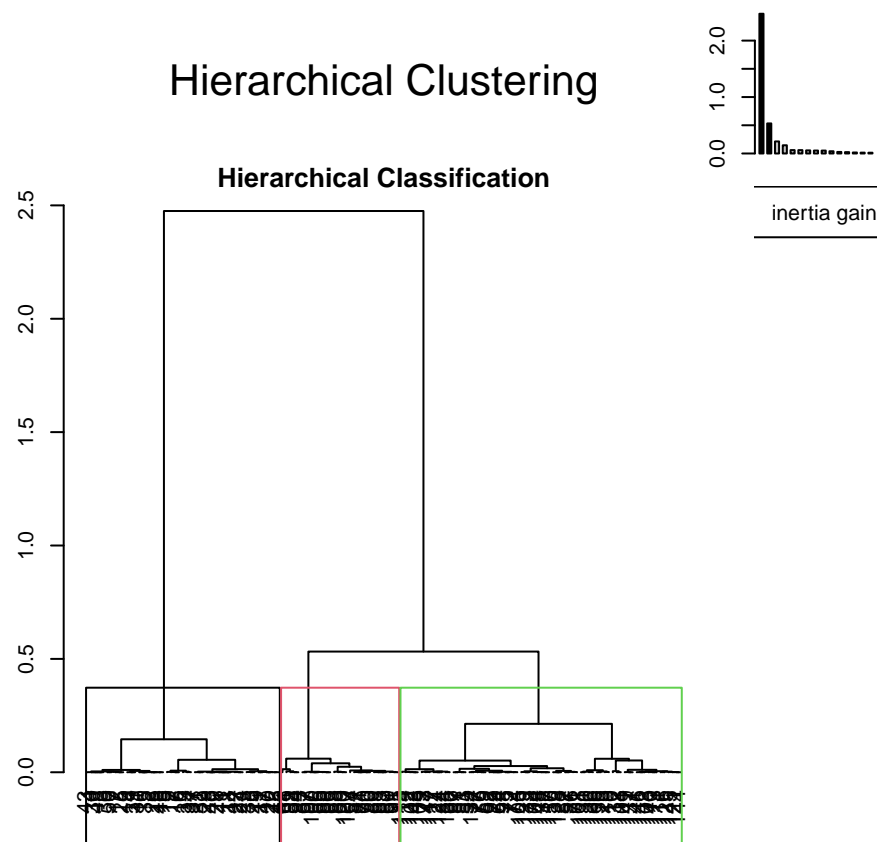
Hierarchical clustering on the factor map

Factor map

Dim 1 (72.96%)
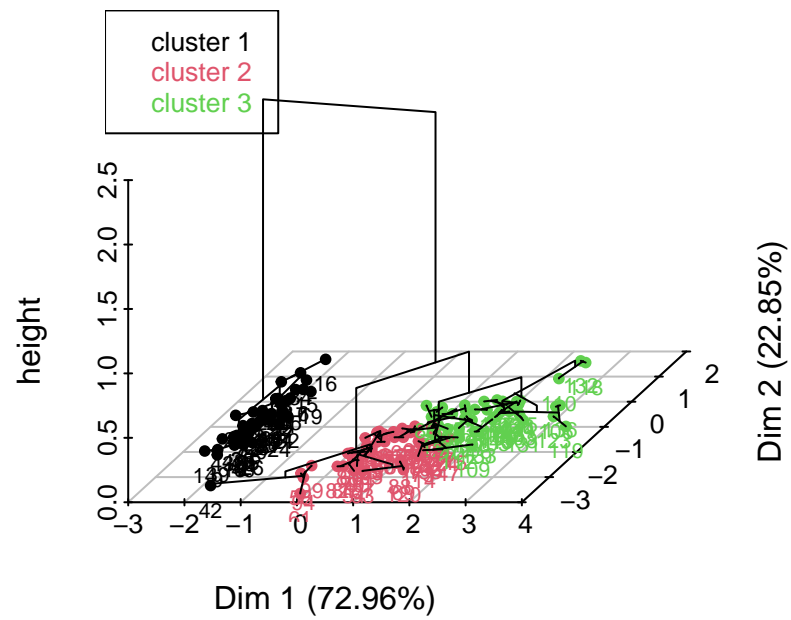
```
## **Results for the Hierarchical Clustering on Principal Components**
##     name
## 1  "$data.clust"
## 2  "$desc.var"
## 3  "$desc.var$quanti.var"
## 4  "$desc.var$quanti"
## 5  "$desc.var$test.chi2"
## 6  "$desc.axes$category"
## 7  "$desc.axes"
## 8  "$desc.axes$quanti.var"
## 9  "$desc.axes$quanti"
## 10 "$desc.ind"
## 11 "$desc.ind$para"
## 12 "$desc.ind$dist"
## 13 "$call"
## 14 "$call$t"
##     description
## 1  "dataset with the cluster of the individuals"
## 2  "description of the clusters by the variables"
## 3  "description of the cluster var. by the continuous var."
## 4  "description of the clusters by the continuous var."
## 5  "description of the cluster var. by the categorical var."
## 6  "description of the clusters by the categories."
## 7  "description of the clusters by the dimensions"
## 8  "description of the cluster var. by the axes"
## 9  "description of the clusters by the axes"
```

```
## 10 "description of the clusters by the individuals"
## 11 "parangons of each clusters"
## 12 "specific individuals"
## 13 "summary statistics"
## 14 "description of the tree"
```

```r
table(hcpc.iris$data.clust[,ncol(hcpc.iris$data.clust)],iris$Species)
```

```
## 
##     setosa versicolor virginica
## 1      50          0         0
## 2       0         39        14
## 3       0         11        36
```

```r
iris2 <- iris[,1:4]
km.iris <- kmeans(iris2,centers=3,nstart=10)
table(km.iris$cluster,iris$Species)
```

```
## 
##     setosa versicolor virginica
## 1       0         48        14
## 2       0          2        36
## 3      50          0         0
```

```r
# Representacio en components principals del K-means ---------------------------
pr.comp <- princomp(iris2)
x <- pr.comp$scores[,1]
y <- pr.comp$scores[,2]
plot(x,y,pch=19,col=km.iris$cluster)
```