

BIPLOT: Matriz Flores

Oscar Elí Bonilla Morales

2022-05-16

Selección de la base de datos

Para este ejemplo se utilizó la base de datos llamada “flores” de la paquetería “datos”, esta base contiene los datos sobre 3 tipos de plantas, datos como ancho y largo.

```
install.packages("datos")
library(datos)
str(flores)
```

```
## 'data.frame': 150 obs. of 5 variables:
## $ Largo.Sepalo: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Ancho.Sepalo: num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Largo.Petalo: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Ancho.Petalo: num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Especie : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

Para comenzar, se realizará un gráfico exploratorio de las variables numéricas. por lo que será necesario realizar la instalación del paquete “MultiBiplotR”. Cabe recalcar que este paquete puede usarse como una alternativa a la paquetería original de R para la creación de Biplots.

```
install.packages("MultiBiplotR")
library(MultiBiplotR)
```

Una vez la paquetería fue instalada proseguimos a realizar una copia de nuestros datos, esto para evitar realizar cambios dentro de la matriz original.

```
BD<-flores
```

Después de crear la copia podemos continuar con la exploración de la matriz

```
dim(BD)
```

```
## [1] 150 5
```

```
str(BD)
```

```
## 'data.frame': 150 obs. of 5 variables:
## $ Largo.Sepalo: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Ancho.Sepalo: num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Largo.Petalo: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Ancho.Petalo: num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Especie : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
colnames(BD)
```

```
## [1] "Largo.Sepalo" "Ancho.Sepalo" "Largo.Petalo" "Ancho.Petalo" "Especie"
```

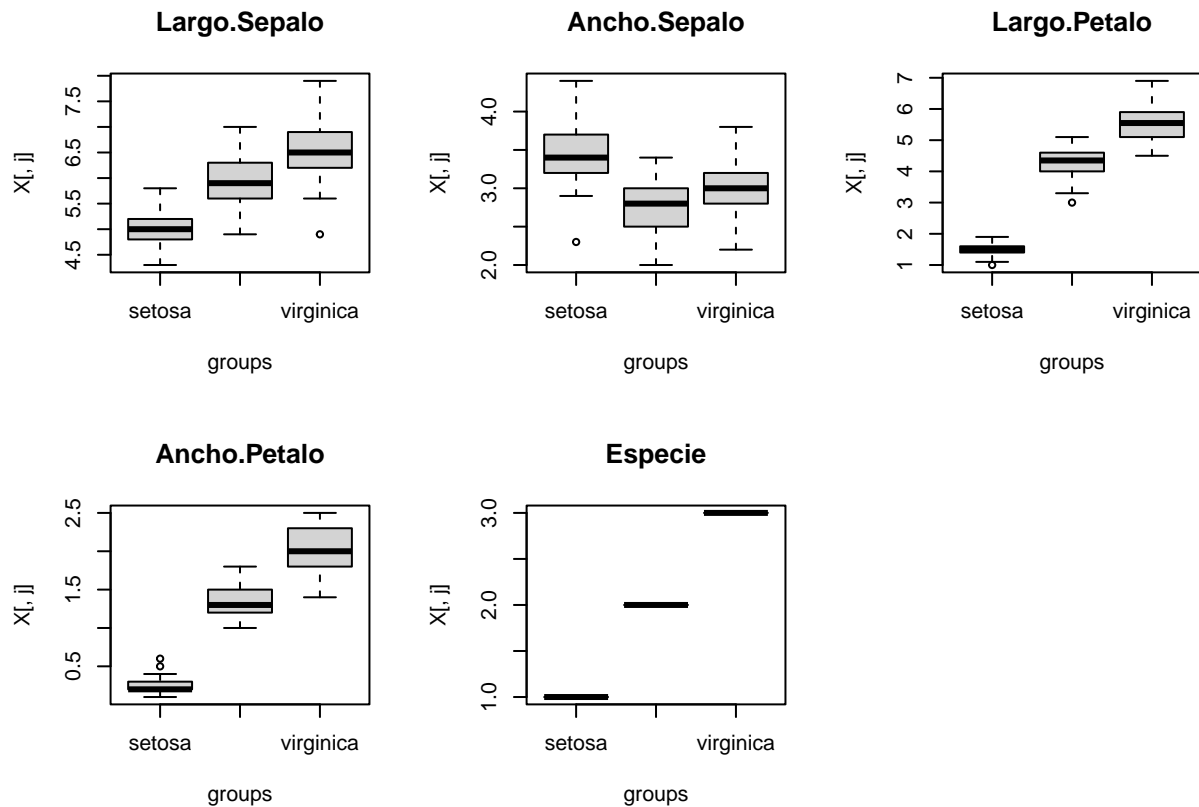
```
attach(BD)
```

La matriz seleccionada cuenta con 5 variables, 4 de ellas de tipo numerico y solo una de tipo factor

En este caso para la realización de gráficos de exploración optamos por un boxplot donde se vean reflejadas las variables.

```
BX1<-BoxPlotPanel(BD, nrow=2, groups=BD$Especie)
```

```
## [1] 2
```



En posible apreciar en los gráficos la existencia de datos atipicos. De igual manera, de forma visual es posible observar cierta similitud entre la variable “virginica” y “versicolor” en el gráfico “Ancho.Sepalo”.

Filtrado de variables

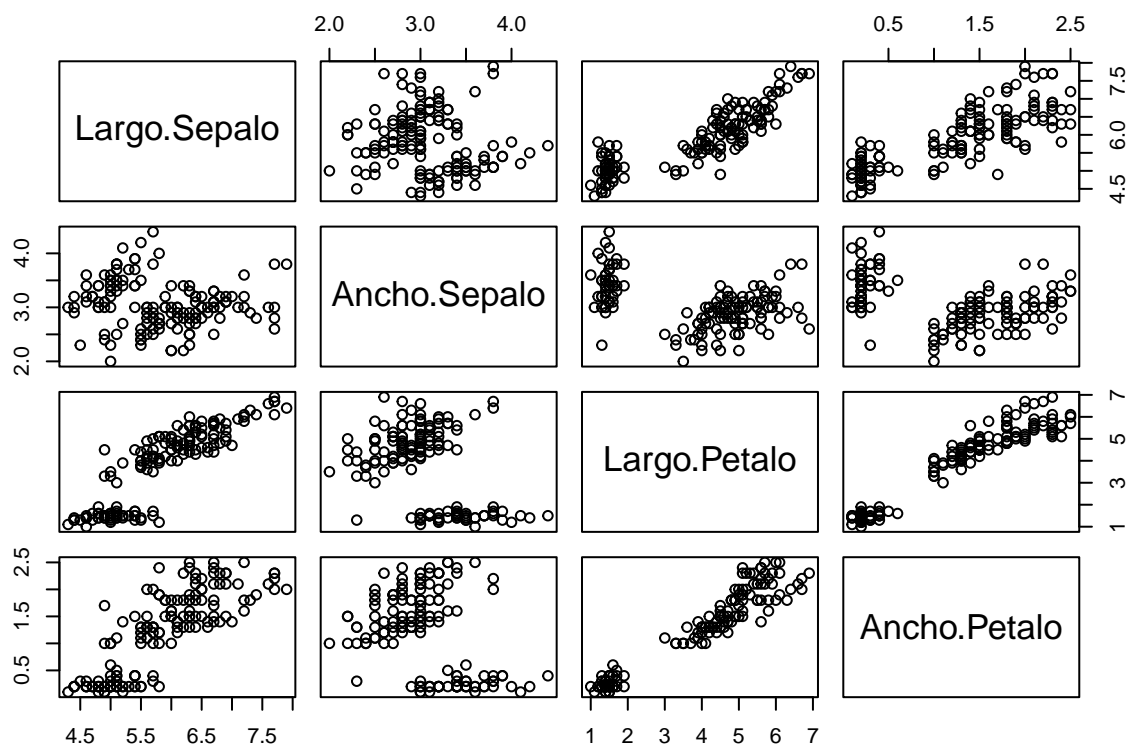
Para poder realizar un BIPLLOT es necesario utilizar solo las variables numéricas.

```
X<-BD[,1:4]
head(X)
```

```
##   Largo.Sepalo Ancho.Sepalo Largo.Petalo Ancho.Petalo
## 1          5.1          3.5          1.4          0.2
## 2          4.9          3.0          1.4          0.2
## 3          4.7          3.2          1.3          0.2
## 4          4.6          3.1          1.5          0.2
## 5          5.0          3.6          1.4          0.2
## 6          5.4          3.9          1.7          0.4
```

Gráfico de dispersión de las variables numéricas

```
# 2.- Generacion Plot  
PL1<-plot(X[,1:4])
```



En este gráfico es posible observar una dispersion considerablemente buena, en algunas de nuestras variables

Reduccion de la dimensionalidad

Procedemos a realizar un análisis de componentes principales para determinar cuales seran aquellos factores con mayor importancia

```
acpflores<-PCA.Analysis(X,Scaling = 5)  
summary(acpflores)
```

```
## ##### Principal Components Analysis #####  
##  
## Transformation of the raw data:  
## [1] "Standardize columns"  
##  
## Eigenvalues & Explained Variance (Inertia)  
##      Eigenvalue Exp. Var Cumulative  
## [1,]  434.85617   72.962    72.962  
## [2,]  136.19054   22.851    95.813  
## [3,]   21.86677    3.669    99.482  
##  
##  
## STRUCTURE OF THE PRINCIPAL COMPONENTS  
##      Dim 1 Dim 2 Dim 3  
## Largo.Sepalo  0.890 -0.361  0.276  
## Ancho.Sepalo -0.460 -0.883 -0.094
```

```
## Largo.Petaló  0.992 -0.023 -0.054
## Ancho.Petaló  0.965 -0.064 -0.243
```

Una vez realizado este análisis obtenemos una varianza explicada de: -0.361 para el primer factor y un -0.883 para el segundo, recordemos que en un análisis de componentes principales solo tomamos en consideración los 2 primeros componentes.

Presentación de tablas

```
summary(acpflores)
```

```
## ##### Principal Components Analysis #####
##
## Transformation of the raw data:
## [1] "Standardize columns"
##
## Eigenvalues & Explained Variance (Inertia)
##      Eigenvalue Exp. Var Cumulative
## [1,]  434.85617   72.962    72.962
## [2,]  136.19054   22.851    95.813
## [3,]   21.86677    3.669   99.482
##
##
## STRUCTURE OF THE PRINCIPAL COMPONENTS
##           Dim 1  Dim 2  Dim 3
## Largo.Sepalo  0.890 -0.361  0.276
## Ancho.Sepalo -0.460 -0.883 -0.094
## Largo.Petaló  0.992 -0.023 -0.054
## Ancho.Petaló  0.965 -0.064 -0.243
```

Contenido del objeto acpflores

```
names(acpflores)
```

```
## [1] "Title"           "Type"            "call"
## [4] "Non_Scaled_Data" "alpha"           "Dimension"
## [7] "Means"           "Medians"         "Deviations"
## [10] "Minima"          "Maxima"          "P25"
## [13] "P75"             "GMean"           "Initial_Transformation"
## [16] "Scaled_Data"     "nrows"           "ncols"
## [19] "nrowsSup"        "ncolsSup"        "dim"
## [22] "EigenValues"     "Inertia"         "CumInertia"
## [25] "EV"              "Structure"       "RowCoordinates"
## [28] "ColCoordinates"  "RowContributions" "ColContributions"
## [31] "Scale_Factor"   "ClusterType"     "Clusters"
## [34] "ClusterColors"   "ClusterNames"
```

Gráfico de ACP

```
acp1<-plot(acpflores, ShowBox=FALSE)
```

Principal Components Analysis (Dim 1 (73 %)- 2 (22.9 %))

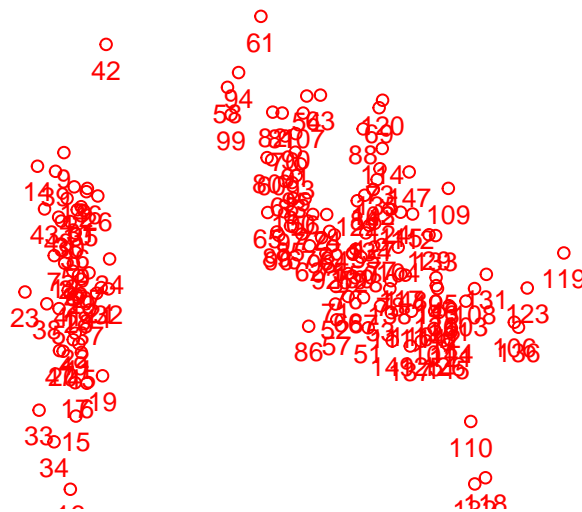


Gráfico de barras ACP

```
acp2<-princomp(X, cor=TRUE, score=TRUE)
plot(acp2)
```

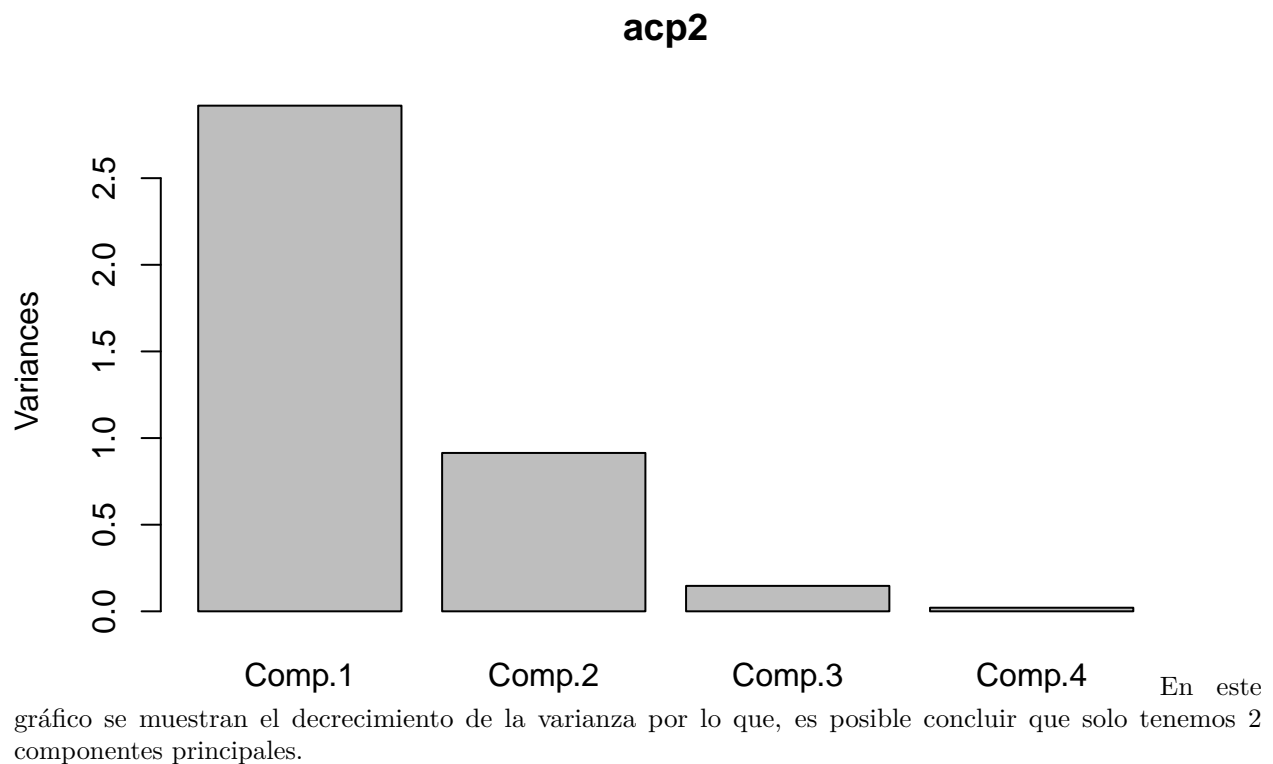
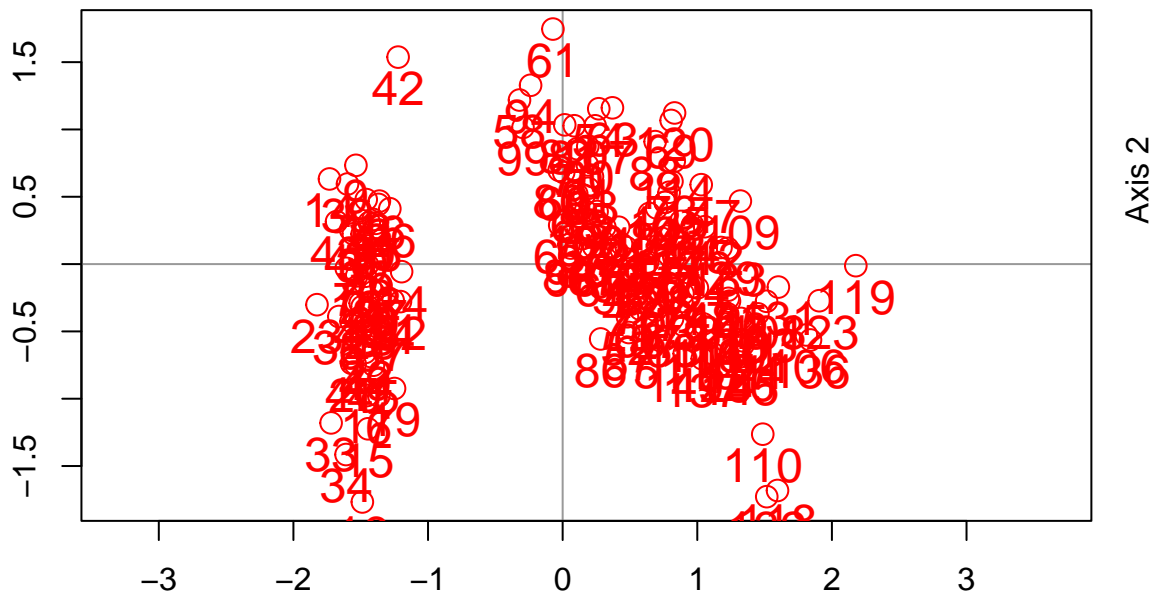


Gráfico circular de correlacion

```
acp3<-plot(acpflores, CorrelationCircle=TRUE,  
           ShowAxis=TRUE, CexInd=1.5)
```

Principa

Principal Components Analysis (Dim 1 (73 %)– 2 (22.9 %))



Adición de grupos al biplot definido por especie

```
acpflores1<-AddCluster2Biplot(acpflores, ClusterType="us",  
                               Groups = BD$Especie)
```

Gráfico con poligonos

```
acp4<-plot(acpflores1, PlotClus=TRUE,  
           ClustCenters=TRUE, margin=0.05,  
           CexInd=0.7, ShowBox=TRUE)
```

Principal Components Analysis (Dim 1 (73 %)- 2 (22.9 %))

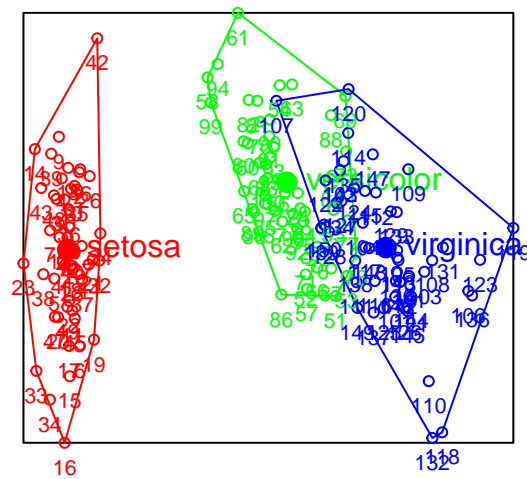


Gráfico con elipses

```
acp5<-plot(acpflores1, PlotClus=TRUE, ClustCenters=TRUE,
margin=0.05, CexInd=0.7, TypeClus="el",
ShowBox=F)
```

Principal Components Analysis (Dim 1 (73 %)- 2 (22.9 %))

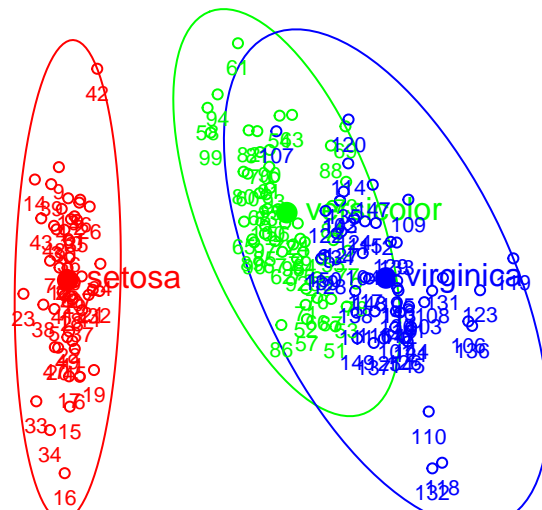
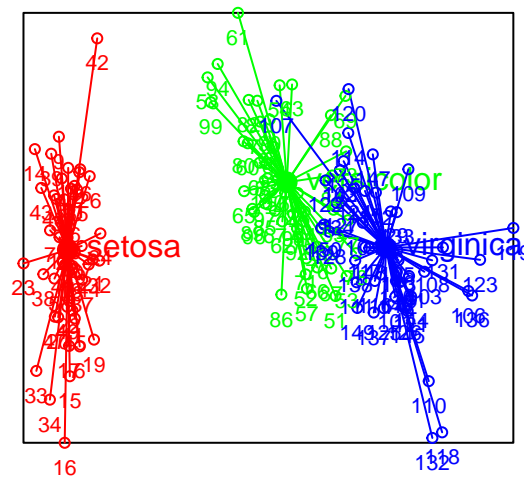


Gráfico con estrellas

```
acp6<-plot(acpflores1, PlotClus=TRUE, ClustCenters=TRUE,
margin=0.05, CexInd=0.7, TypeClus="st",
ShowBox=TRUE)
```

Principal Components Analysis (Dim 1 (73 %)- 2 (22.9 %))



CREACIÓN

DEL BIPLLOT

alpha=

0:GH

1:JK

2:HJ

Predeterminado JK

Los distintos tipos de graficos se encuentran ya intengrados en la paqueteria “MultBiplotR”, por lo que dependiendo del tipo de biplot que se dea obtener, ya ses un jk, hj o gj, será necesario especificarlo dentro del código.

```
bipflores<-PCA.Biplot(X, alpha = 1, Scaling = 5)
```

Al realizar un summary de la funcion “bipflores” nos brinda una gran cantidad de datos descriptivos de nuestros datos y del análisis de componentes.

Valores propios

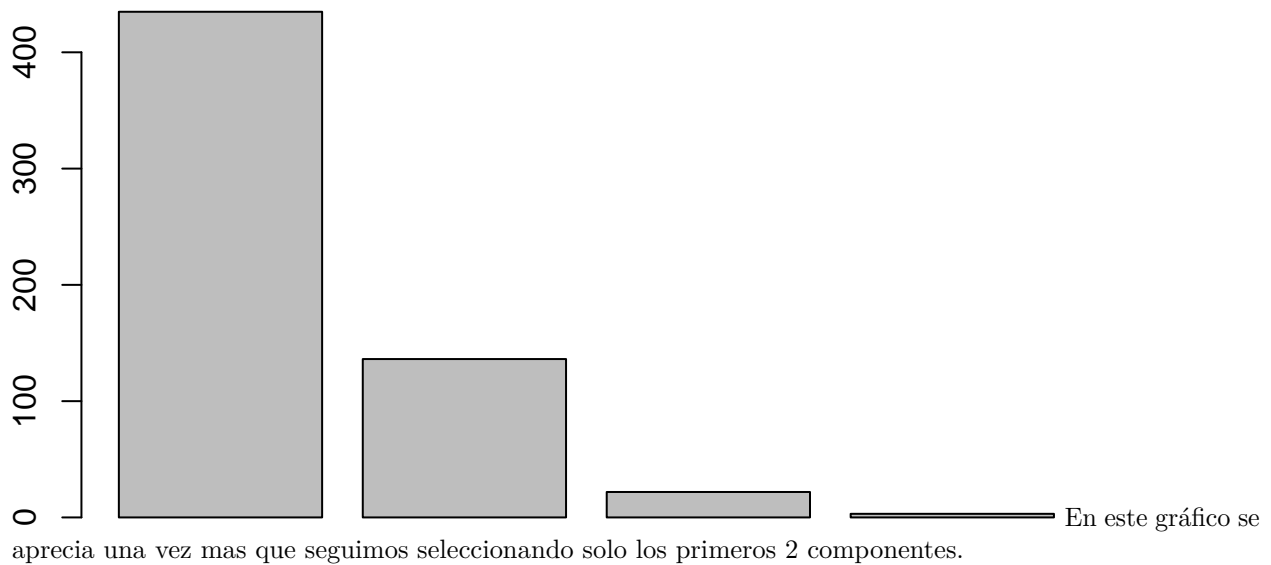
Obtenemos los valores propios

```
bipflores$EigenValues
```

```
## [1] 434.856175 136.190540 21.866774 3.086511
```

screepplot

```
SC<-barplot(bipflores$EigenValues)
```

Vectores propios

Obtenemos los vectores propios

```
bipflores$EV
```

```
##           [,1]      [,2]      [,3]
## [1,]  0.5210659 -0.37741762  0.7195664
## [2,] -0.2693474 -0.92329566 -0.2443818
## [3,]  0.5804131 -0.02449161 -0.1421264
## [4,]  0.5648565 -0.06694199 -0.6342727
```

Tabla de inercias

```
Inercias<-data.frame(paste("Eje",1:length(bipflores$EigenValues)),
                     bipflores$EigenValues, bipflores$Inertia,
                     bipflores$CumInertia)

colnames(Inercias)<-c("Eje", "Valor Propio",
                     "Inercia", "Inercia acumulada")
```

Markdown

Se realizan los comandos para una mejor presentación de las tablas

```
library(knitr)
kable(Inercias)
```

Eje	Valor Propio	Inercia	Inercia acumulada
Eje 1	434.856175	72.962	72.962
Eje 2	136.190540	22.851	95.813
Eje 3	21.866774	3.669	99.482
Eje 4	3.086511	0.518	100.000

tabla contribucion de columnas

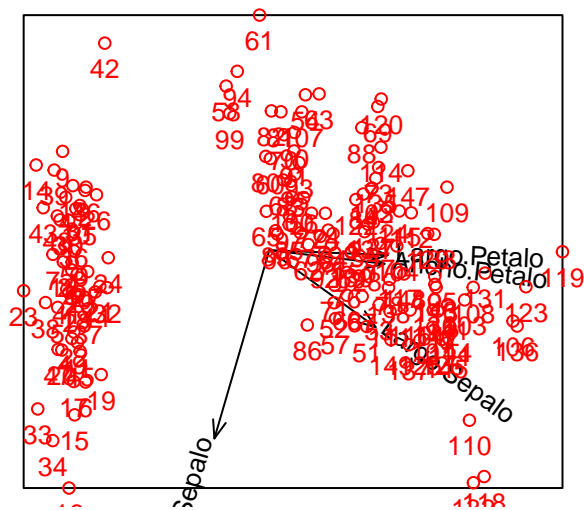
```
kable(bipflores$ColContributions)
```

	Dim 1	Dim 2	Dim 3
Largo.Sepalo	79.24	13.02	7.60
Ancho.Sepalo	21.17	77.92	0.88
Largo.Petalo	98.32	0.05	0.30
Ancho.Petalo	93.12	0.41	5.90

Grafico

```
plot(bipflores, ShowBox=TRUE)
```

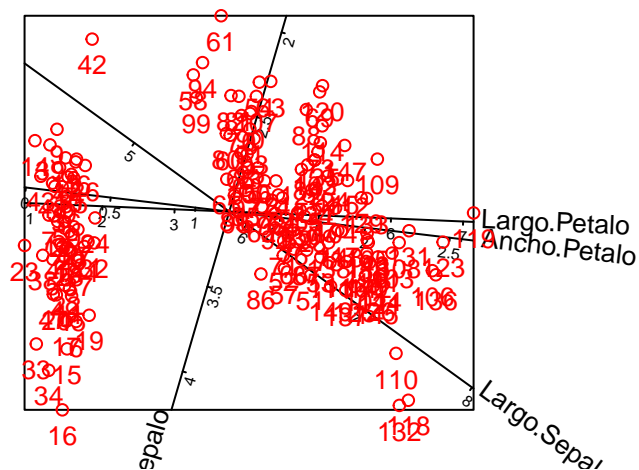
PCA Biplot (Dim 1 (73 %)- 2 (22.9 %))



Prolongacion de vectores linea recta

```
BP1<-plot(bipflores, mode="s",  
margin=0.1, ShowBox=TRUE)
```

PCA Biplot (Dim 1 (73 %)- 2 (22.9 %))



Prolongacion de vectores con flechas y linea punteada

```
BP2<-plot(bipflores, mode="ah", margin=0.05,
          ShowBox=TRUE)
```

PCA Biplot (Dim 1 (73 %)- 2 (22.9 %))

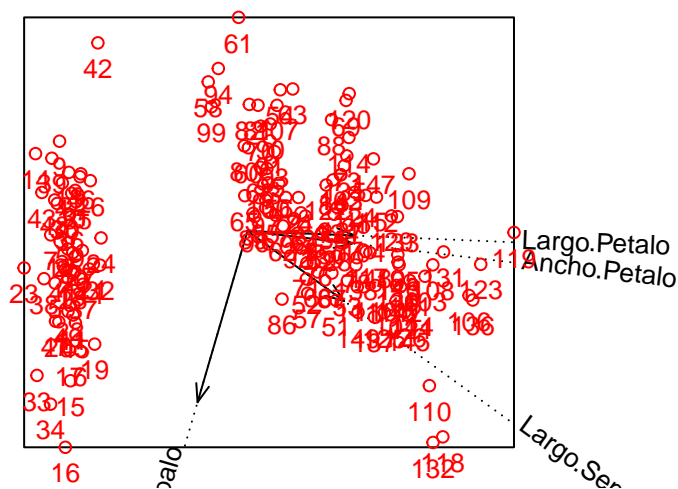


Grafico circular correlaciones

```
GC<-CorrelationCircle(bipflores)
```

PCA Biplot – Correlation Circle

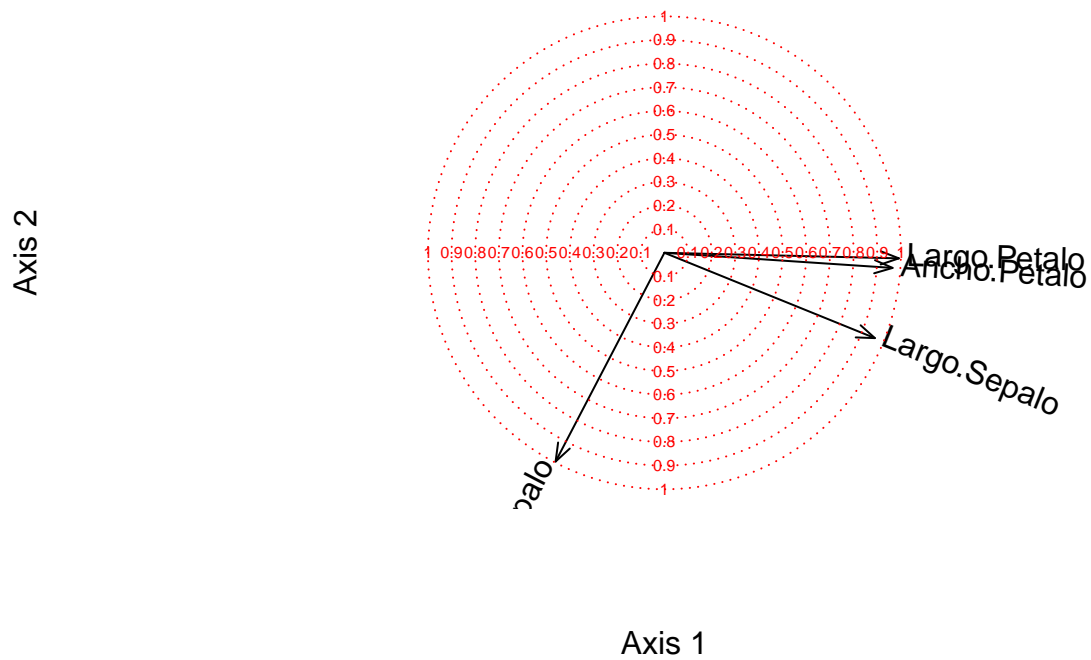
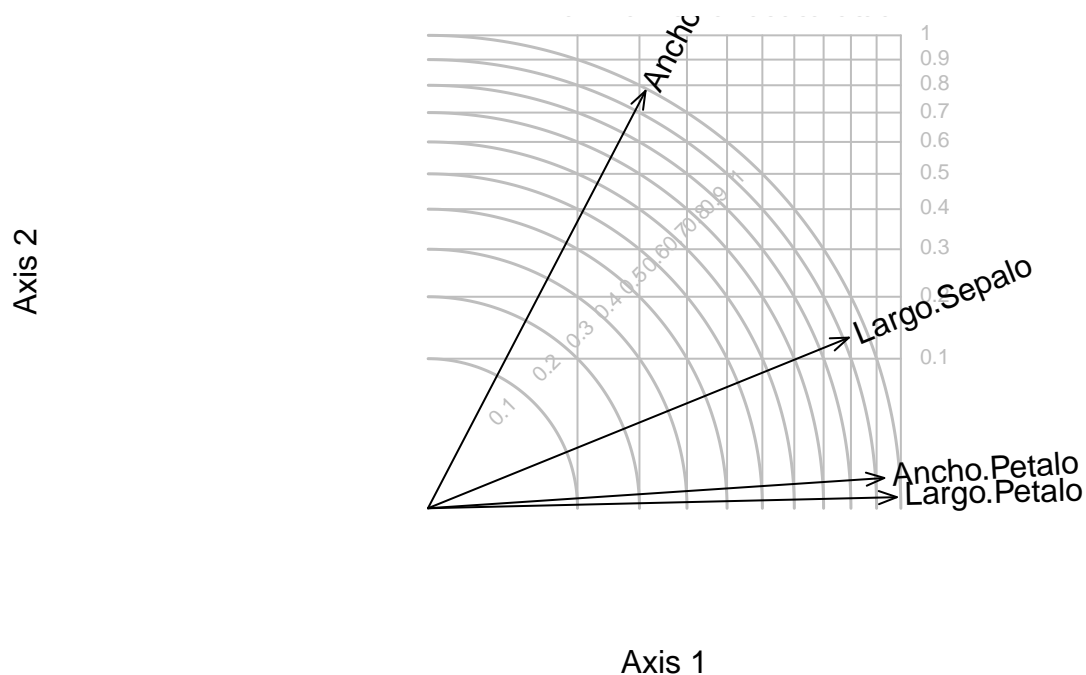


Grafico contribuciones de los vectores Calidad de representacion eje 1, 2 y 1+2

```
ColContributionPlot(bipflores, AddSigns2Labs = FALSE)
```

PCA Biplot – Contribution Plot



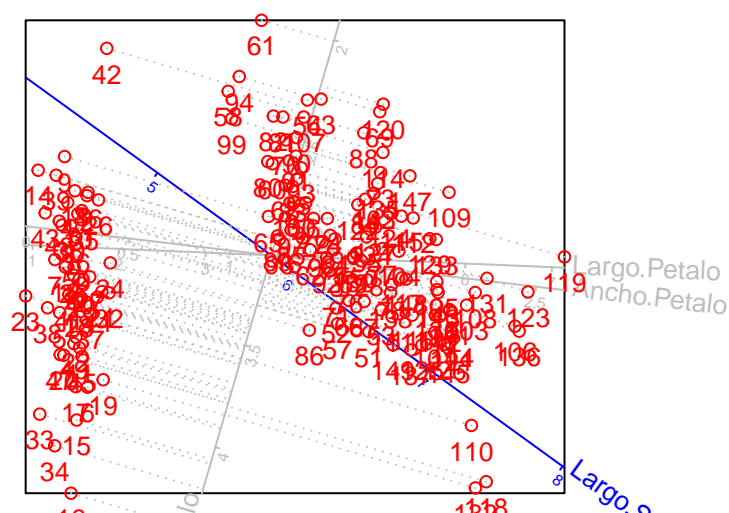
apreciar que la variable Largo.Sepalo contribuye mas al factor 1 que al 2.

Es posible

Proyeccion individuos sobre una variable dp= selecciona la variable

```
BP3<-plot(bipflores, dp=2, mode="s",
          ColorVar=c("blue", rep("grey",17)),
          ShowBox=TRUE)
```

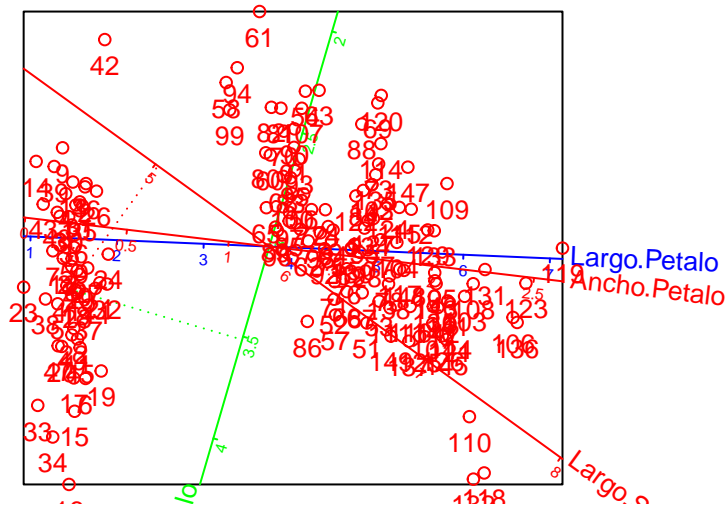
PCA Biplot (Dim 1 (73 %)– 2 (22.9 %))



#Proyeccion de ind sobre todas las variables PredPoints= individuo

```
BP4<-plot(bipflores, PredPoints=1, mode="s",
          ColorVar=1:18, ShowBox=TRUE)
```

PCA Biplot (Dim 1 (73 %)- 2 (22.9 %))



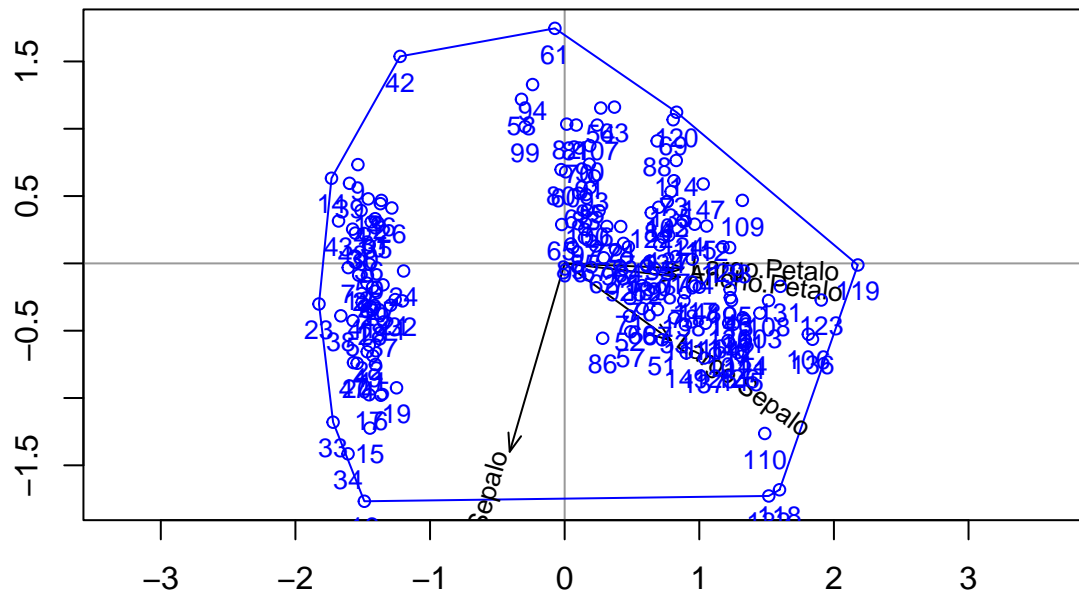
Agregar cluster Jerarquico con datos originales metodo ward.D

```
bipvino=AddCluster2Biplot(bipflores, NGroups=4,
                           ClusterType="hi",
                           method="ward.D",
                           Original=TRUE)
```

Cluster aplicado al biplot

```
clusBP<-plot(bipflores, PlotClus=TRUE, ShowAxis=TRUE)
```

PCA Biplot (Dim 1 (73 %)- 2 (22.9 %))



clusBP

NULL