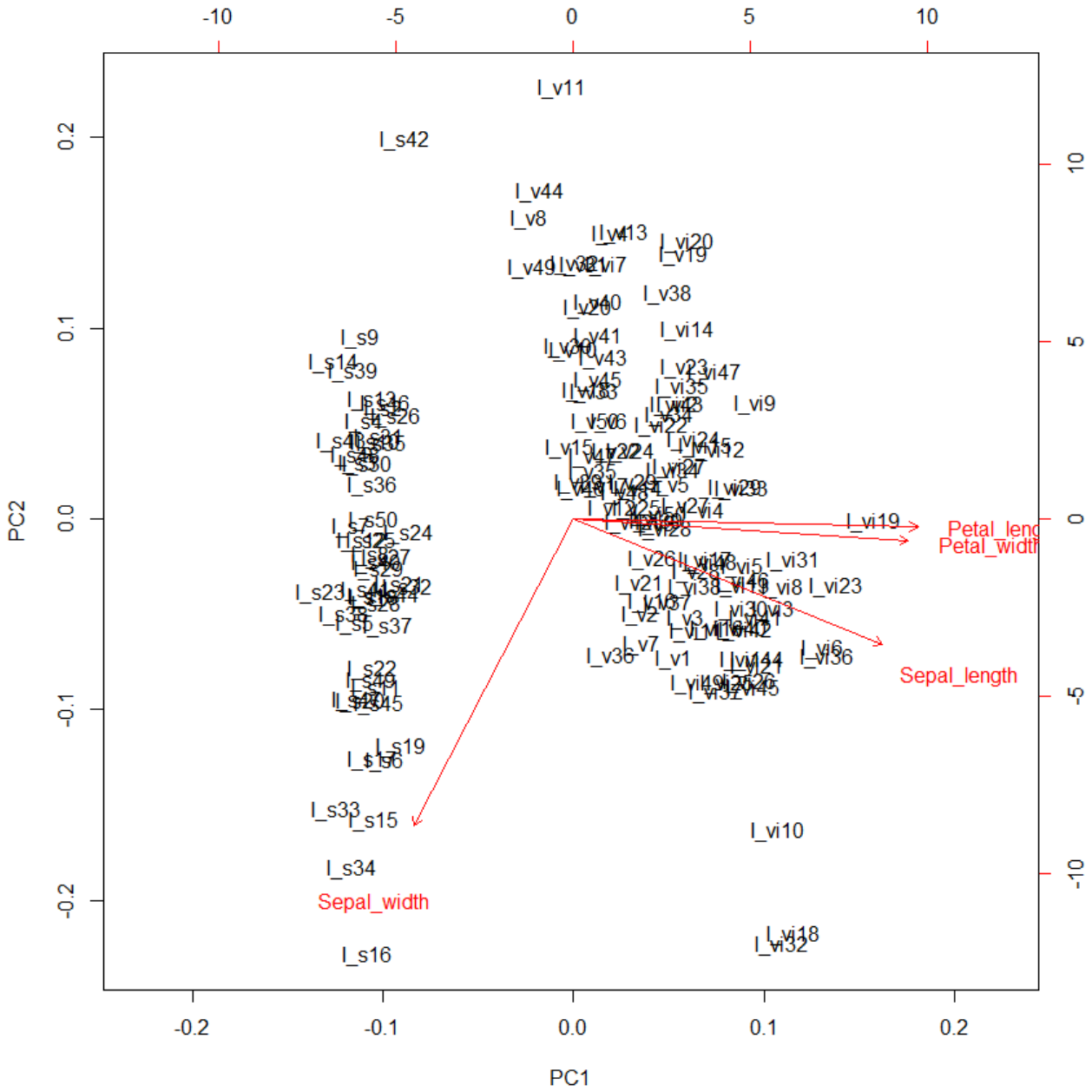
3-Práctica de Análisis BIPLOT en R

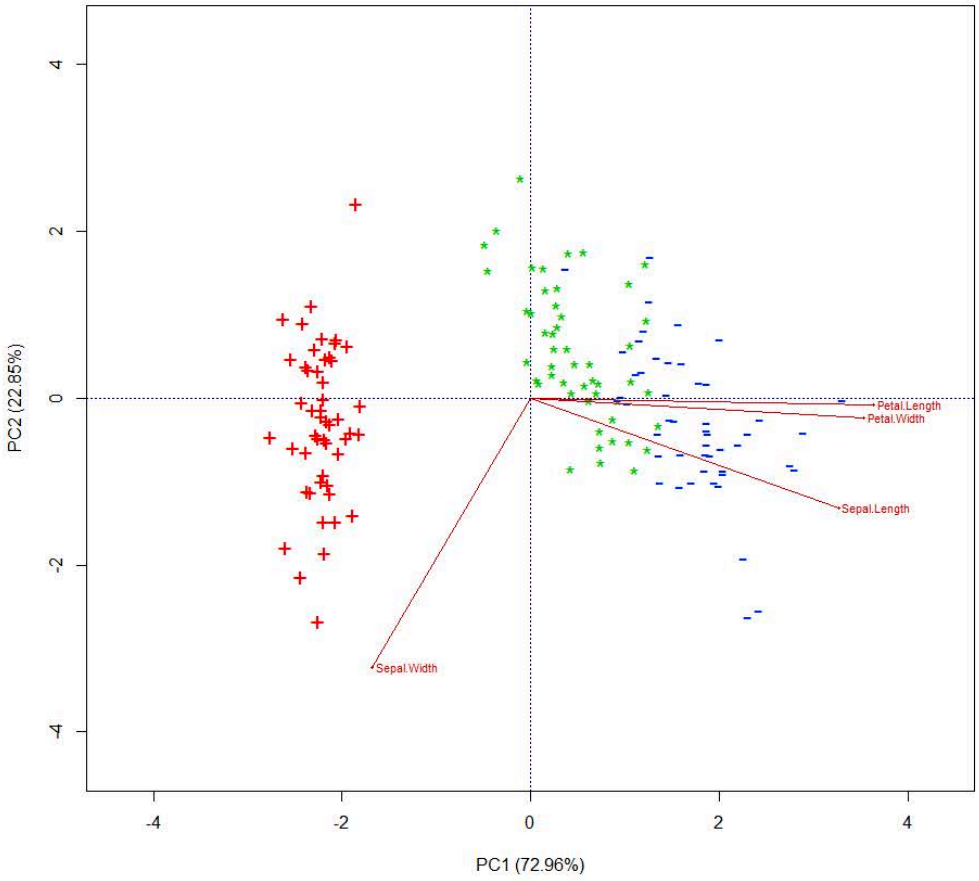
Vamos a trabajar en esta ocasión con los datos sobre lirios de Fisher (Fisher's Iris data). The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimetres.

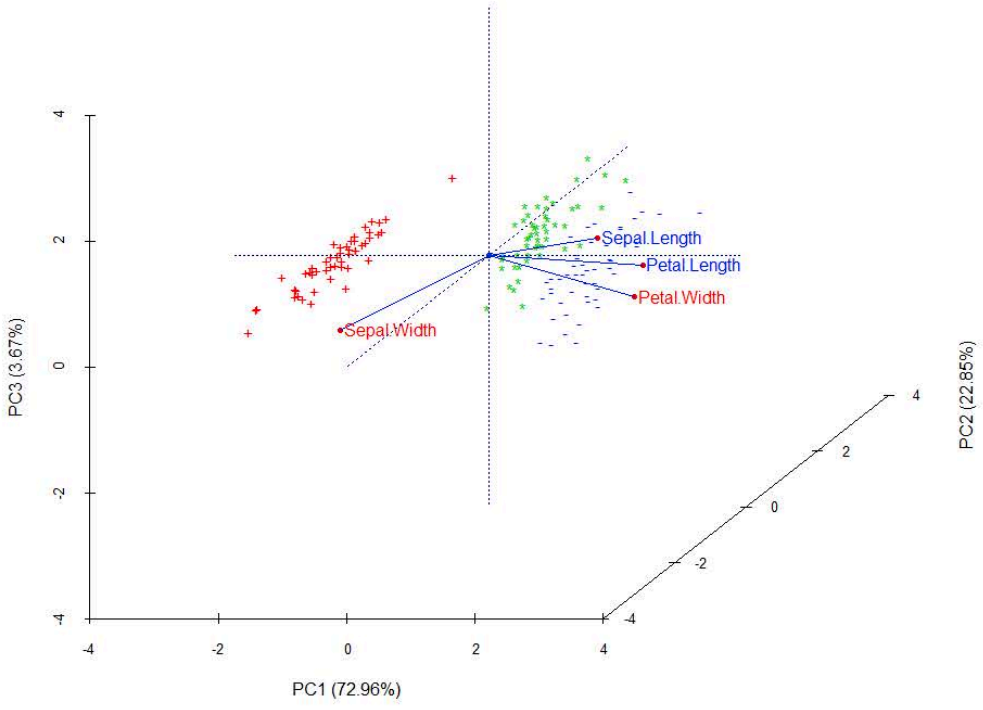
Utilizaréis el paquete bpca, (Biplot de datos multivariantes basado en Análisis de Componentes Principales), que representa gráficamente biplot en 2 y 3 dimensiones, y tendréis que llegar a los siguientes gráficos, por supuesto previamente debéis leer y entender la documentación de dicho paquete.

Por el procedimiento basado en la función prcomp, obtendríamos el siguiente gráfico:

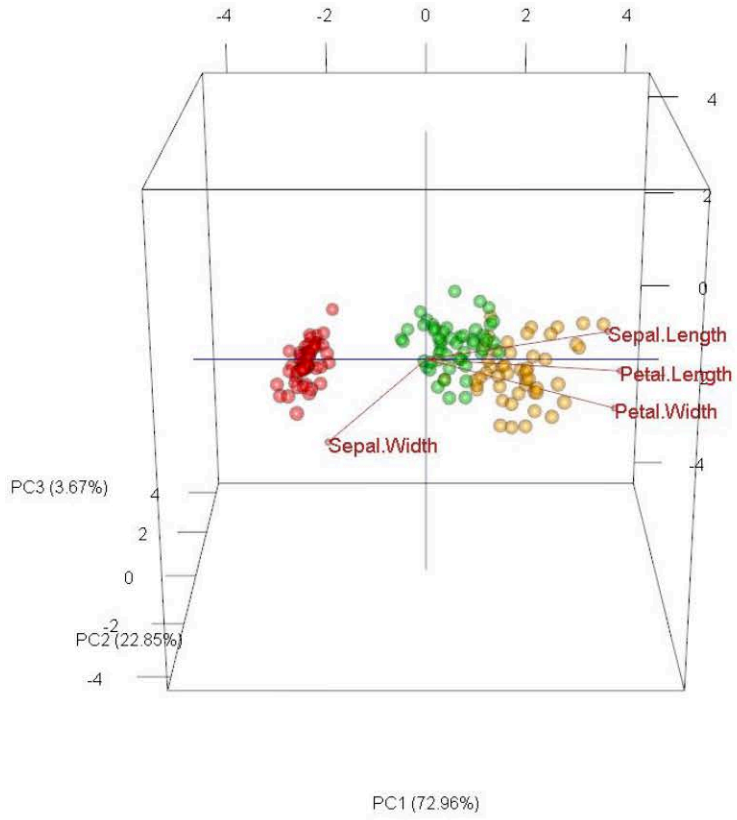


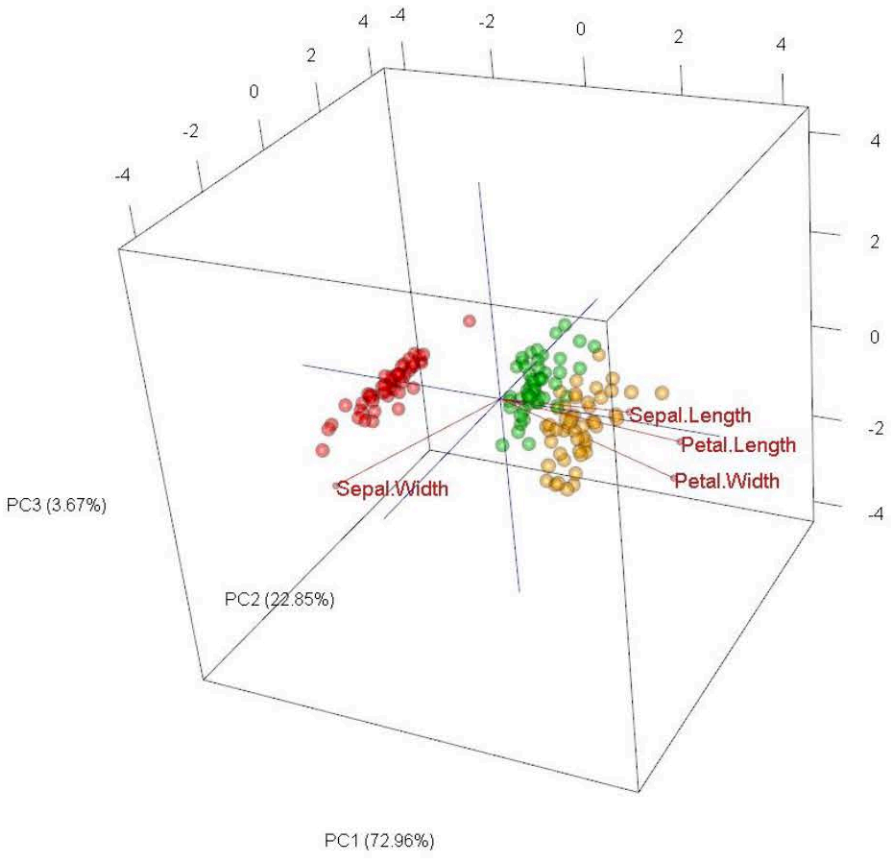
Y con el paquete bpca, obtenemos los gráficos de abajo en 2 y 3 dimensiones, donde las 3 especies de lirios, setosa, versicolor y virginica, aparecen con colores diferentes, los porcentajes de los ejes factoriales corresponden a los obtenidos por la función prcomp con datos centrados y normalizados.


PC2 (22.85%)
PC1 (72.96%)
Petal.Length
Petal.Width
Sepal.Length
Sepal.Width


PC3 (3.67%)
PC1 (72.96%)
PC2 (22.85%)
Sepal.Length
Petal.Length
Petal.Width
Sepal.Width

Permite también obtener gráficos en 3d dinámicos, es decir que puedes interactuar con el gráfico, desplazando los ejes en el espacio





```
> X <- read.table("iris.txt", dec=".")

> X <- as.matrix(X)

> X
```

	Sepal_length	Sepal_width	Petal_length	Petal_width
I_s1	5.1	3.5	1.4	0.2
I_s2	4.9	3.0	1.4	0.2
I_s3	4.7	3.2	1.3	0.2
I_s4	4.6	3.1	1.5	0.2
I_s5	5.0	3.6	1.4	0.2
I_s6	5.4	3.9	1.7	0.4
I_s7	4.6	3.4	1.4	0.3
I_s8	5.0	3.4	1.5	0.2
I_s9	4.4	2.9	1.4	0.2
I_s10	4.9	3.1	1.5	0.1
I_s11	5.4	3.7	1.5	0.2
I_s12	4.8	3.4	1.6	0.2
I_s13	4.8	3.0	1.4	0.1
I_s14	4.3	3.0	1.1	0.1
I_s15	5.8	4.0	1.2	0.2
I_s16	5.7	4.4	1.5	0.4
I_s17	5.4	3.9	1.3	0.4
I_s18	5.1	3.5	1.4	0.3
I_s19	5.7	3.8	1.7	0.3
I_s20	5.1	3.8	1.5	0.3
I_s21	5.4	3.4	1.7	0.2
I_s22	5.1	3.7	1.5	0.4
I_s23	4.6	3.6	1.0	0.2
I_s24	5.1	3.3	1.7	0.5
I_s25	4.8	3.4	1.9	0.2
I_s26	5.0	3.0	1.6	0.2
I_s27	5.0	3.4	1.6	0.4
I_s28	5.2	3.5	1.5	0.2
I_s29	5.2	3.4	1.4	0.2
I_s30	4.7	3.2	1.6	0.2
I_s31	4.8	3.1	1.6	0.2
I_s32	5.4	3.4	1.5	0.4
I_s33	5.2	4.1	1.5	0.1
I_s34	5.5	4.2	1.4	0.2
I_s35	4.9	3.1	1.5	0.2
I_s36	5.0	3.2	1.2	0.2
I_s37	5.5	3.5	1.3	0.2
I_s38	4.9	3.6	1.4	0.1
I_s39	4.4	3.0	1.3	0.2
I_s40	5.1	3.4	1.5	0.2
I_s41	5.0	3.5	1.3	0.3
I_s42	4.5	2.3	1.3	0.3
I_s43	4.4	3.2	1.3	0.2
I_s44	5.0	3.5	1.6	0.6
I_s45	5.1	3.8	1.9	0.4
I_s46	4.8	3.0	1.4	0.3
I_s47	5.1	3.8	1.6	0.2
I_s48	4.6	3.2	1.4	0.2
I_s49	5.3	3.7	1.5	0.2
I_s50	5.0	3.3	1.4	0.2
I_v1	7.0	3.2	4.7	1.4
I_v2	6.4	3.2	4.5	1.5
I_v3	6.9	3.1	4.9	1.5
I_v4	5.5	2.3	4.0	1.3
I_v5	6.5	2.8	4.6	1.5
I_v6	5.7	2.8	4.5	1.3
I_v7	6.3	3.3	4.7	1.6
I_v8	4.9	2.4	3.3	1.0
I_v9	6.6	2.9	4.6	1.3
I_v10	5.2	2.7	3.9	1.4
I_v11	5.0	2.0	3.5	1.0
I_v12	5.9	3.0	4.2	1.5
I_v13	6.0	2.2	4.0	1.0
I_v14	6.1	2.9	4.7	1.4
I_v15	5.6	2.9	3.6	1.3
I_v16	6.7	3.1	4.4	1.4
I_v17	5.6	3.0	4.5	1.5
I_v18	5.8	2.7	4.1	1.0
I_v19	6.2	2.2	4.5	1.5

I_v20	5.6	2.5	3.9	1.1
I_v21	5.9	3.2	4.8	1.8
I_v22	6.1	2.8	4.0	1.3
I_v23	6.3	2.5	4.9	1.5
I_v24	6.1	2.8	4.7	1.2
I_v25	6.4	2.9	4.3	1.3
I_v26	6.6	3.0	4.4	1.4
I_v27	6.8	2.8	4.8	1.4
I_v28	6.7	3.0	5.0	1.7
I_v29	6.0	2.9	4.5	1.5
I_v30	5.7	2.6	3.5	1.0
I_v31	5.5	2.4	3.8	1.1
I_v32	5.5	2.4	3.7	1.0
I_v33	5.8	2.7	3.9	1.2
I_v34	6.0	2.7	5.1	1.6
I_v35	5.4	3.0	4.5	1.5
I_v36	6.0	3.4	4.5	1.6
I_v37	6.7	3.1	4.7	1.5
I_v38	6.3	2.3	4.4	1.3
I_v39	5.6	3.0	4.1	1.3
I_v40	5.5	2.5	4.0	1.3
I_v41	5.5	2.6	4.4	1.2
I_v42	6.1	3.0	4.6	1.4
I_v43	5.8	2.6	4.0	1.2
I_v44	5.0	2.3	3.3	1.0
I_v45	5.6	2.7	4.2	1.3
I_v46	5.7	3.0	4.2	1.2
I_v47	5.7	2.9	4.2	1.3
I_v48	6.2	2.9	4.3	1.3
I_v49	5.1	2.5	3.0	1.1
I_v50	5.7	2.8	4.1	1.3
I_vi1	6.3	3.3	6.0	2.5
I_vi2	5.8	2.7	5.1	1.9
I_vi3	7.1	3.0	5.9	2.1
I_vi4	6.3	2.9	5.6	1.8
I_vi5	6.5	3.0	5.8	2.2
I_vi6	7.6	3.0	6.6	2.1
I_vi7	4.9	2.5	4.5	1.7
I_vi8	7.3	2.9	6.3	1.8
I_vi9	6.7	2.5	5.8	1.8
I_vi10	7.2	3.6	6.1	2.5
I_vi11	6.5	3.2	5.1	2.0
I_vi12	6.4	2.7	5.3	1.9
I_vi13	6.8	3.0	5.5	2.1
I_vi14	5.7	2.5	5.0	2.0
I_vi15	5.8	2.8	5.1	2.4
I_vi16	6.4	3.2	5.3	2.3
I_vi17	6.5	3.0	5.5	1.8
I_vi18	7.7	3.8	6.7	2.2
I_vi19	7.7	2.6	6.9	2.3
I_vi20	6.0	2.2	5.0	1.5
I_vi21	6.9	3.2	5.7	2.3
I_vi22	5.6	2.8	4.9	2.0
I_vi23	7.7	2.8	6.7	2.0
I_vi24	6.3	2.7	4.9	1.8
I_vi25	6.7	3.3	5.7	2.1
I_vi26	7.2	3.2	6.0	1.8
I_vi27	6.2	2.8	4.8	1.8
I_vi28	6.1	3.0	4.9	1.8
I_vi29	6.4	2.8	5.6	2.1
I_vi30	7.2	3.0	5.8	1.6
I_vi31	7.4	2.8	6.1	1.9
I_vi32	7.9	3.8	6.4	2.0
I_vi33	6.4	2.8	5.6	2.2
I_vi34	6.3	2.8	5.1	1.5
I_vi35	6.1	2.6	5.6	1.4
I_vi36	7.7	3.0	6.1	2.3
I_vi37	6.3	3.4	5.6	2.4
I_vi38	6.4	3.1	5.5	1.8
I_vi39	6.0	3.0	4.8	1.8
I_vi40	6.9	3.1	5.4	2.1
I_vi41	6.7	3.1	5.6	2.4
I_vi42	6.9	3.1	5.1	2.3
I_vi43	5.8	2.7	5.1	1.9
I_vi44	6.8	3.2	5.9	2.3
I_vi45	6.7	3.3	5.7	2.5

```
I_vi46          6.7         3.0          5.2         2.3
I_vi47          6.3         2.5          5.0         1.9
I_vi48          6.5         3.0          5.2         2.0
I_vi49          6.2         3.4          5.4         2.3
I_vi50          5.9         3.0          5.1         1.8

> X_pca <- prcomp(X, center=T, scale=T)

> X_pca
Standard deviations:
[1] 1.7083611 0.9560494 0.3830886 0.1439265

Rotation:
                    PC1          PC2         PC3         PC4
Sepal_length  0.5210659 -0.37741762  0.7195664  0.2612863
Sepal_width  -0.2693474 -0.92329566 -0.2443818 -0.1235096
Petal_length  0.5804131 -0.02449161 -0.1421264 -0.8014492
Petal_width   0.5648565 -0.06694199 -0.6342727  0.5235971

> summary(X_pca)
Importance of components:
                          PC1    PC2     PC3     PC4
Standard deviation     1.7084 0.9560 0.38309 0.14393
Proportion of Variance 0.7296 0.2285 0.03669 0.00518
Cumulative Proportion  0.7296 0.9581 0.99482 1.00000

> biplot(X_pca) #1° gráfico

> summary(bpca(X))
 Eigenvalue(s):                          20.85321 11.67007 4.676192 1.756847
  - Considered on reduction:             20.85321 11.67007
  - Variance retained by each:           0.7296245 0.2285076
  - Cumulative variance retained:        0.7296245 0.9581321
  - Prop. of total variance retained:    0.958

> plot(bpca(iris[-5]), #2° gráfico

+ var.factor=.3, var.cex=.7,
+ obj.names=FALSE, obj.cex=1.5,
+ obj.col=c('red', 'green3', 'blue')[unclass(iris$Species)],
+ obj.pch=c('+', '*', '-')[unclass(iris$Species)])

> # 3d static #3° gráfico

> plot(bpca(iris[-5], d=1:3),
+ var.factor=.2, var.color=c('blue', 'red'), var.cex=1,
+ obj.names=FALSE, obj.cex=1,
+ obj.col=c('red', 'green3', 'blue')[unclass(iris$Species)],
 obj.pch=c('+', '*', '-')[unclass(iris$Species)])

> # 3d dynamic #4° y 5° gráfico

> plot(bpca(iris[-5], method='hj', d=1:3), rgl.use=TRUE,
+ var.col='brown', var.factor=.3, var.cex=1.2,
+ obj.names=FALSE, obj.cex=.8,
+ obj.col=c('red', 'green3', 'orange')[unclass(iris$Species)],
+ simple.axes=FALSE, box=TRUE)
```