Análisis Biplot

Parte II

BIPLOT DE COMPONENTES PRINCIPALES

Datos originales $\mathbf{X}_{(nxp)}$ *(Matriz como desviaciones respecto a la media de columnas)*

Preserva columnas $\mathbf{X^TX}_{(pxp)}$ $\mathbf{XX^T}_{(nxn)}$ Preserva filas

$$\mathbf{r(X^TX)}=p \qquad \mathbf{r(XX^T)}=p$$

$$\mathbf{X'X=ULU'} \qquad \mathbf{XX'=U^*LU^{*\prime}}$$

U(pxp) vector característico de X'X

U*(nxp) vector característico de XX'

Gabriel K.R. & Odoroff, C.L. (1986) The biplot for Exploration and Diagnosis: examples and software, Dept. of Stat. Rep. #86/03, University of Rochester, Rochester, New York.

Continuación…

Vectores característicos escalados por $L^{1/2}$=diag(λ)

$$V=UL^{1/2} \qquad V^*=U^*L^{1/2}$$

$(VV'=X'X)$ $\qquad$ $(V^*V^{*\prime}= XX')$

Matriz de correlación

Vectores característicos escalados por $L^{-1/2}$=diag(λ)

$$W=UL^{-1/2} \qquad W^*=U^*L^{-1/2}$$

$$\boxed{X=U^*L^{1/2}U'}$$

ver (ii)

Construcción de X(nxp)

Continuación…

$$\mathbf{X=AB'}$$

Representación gráfica de X en dos dimensiones

$$\mathbf{X=U*L^{c/2}L^{(1-c)/2}U'}$$

$$0 \leq c \leq 1$$

$$\boxed{\mathbf{A=U*L^{c/2}} \text{ y } \mathbf{B'=L^{(1-c)/2}U'}}$$

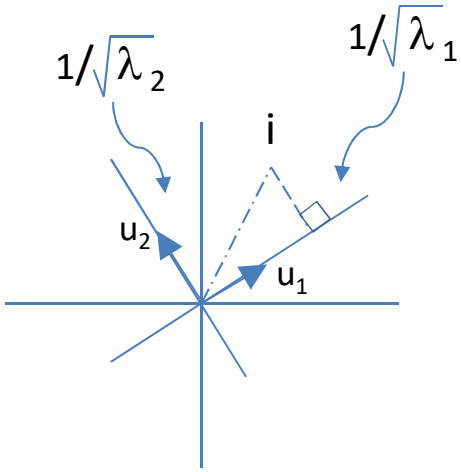
C=0.5	SQ biplot	$\mathbf{A=U*L^{1/4}}$ y $\mathbf{B'=L^{1/4}U'}$	Para diagnóstico
C=0	GH biplot	$\mathbf{A=U*}$ y $\mathbf{B'=L^{1/2}U'}$	Preserva la métrica de columna
C=1	JK biplot	$\mathbf{A=V^*=Z}$ y $\mathbf{B'=U'}=Y^*$	Preserva la métrica de filas

ENFOQUE GEOMÉTRICO CON ESCALAMIENTO

Análisis en R^p:

$\underset{px1}{X_i}$ (observación i)

$$\text{Comp}_{\overline{u}_1} x_i = F_1(i) = x'_i\left(\frac{1}{\sqrt{\lambda_1}}\overline{u}_1\right)$$



$$F_\alpha() = \underset{(nxp)}{X}\ \frac{1}{\sqrt{\lambda_\alpha}}\ \underset{(px1)}{\overline{u}_\alpha}$$

$$F() = \left(X\ \frac{1}{\sqrt{\lambda_1}}\overline{u}_1 \quad \dots \quad X\ \frac{1}{\sqrt{\lambda_p}}\overline{u}_p \right)$$

$$F() = X\left(\frac{1}{\sqrt{\lambda_1}}\overline{u}_1 \quad \dots \quad \frac{1}{\sqrt{\lambda_p}}\overline{u}_p \right)$$

ENFOQUE GEOMÉTRICO CON ESCALAMIENTO
Con enfoque matricial

Análisis en R^p:

$$F() = XUL^{-1/2}$$

$$L^{-1/2} = \begin{pmatrix} 1/\sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & 1/\sqrt{\lambda_\alpha} & \\ & & & \ddots \\ & & & & 1/\sqrt{\lambda_p} \end{pmatrix}$$

Scores componentes principales para las observaciones en R_p (X'X)

$$\underset{(nxp)}{Y} = XUL^{-1/2} = XW$$

SQ-Biplot (c=0.5)

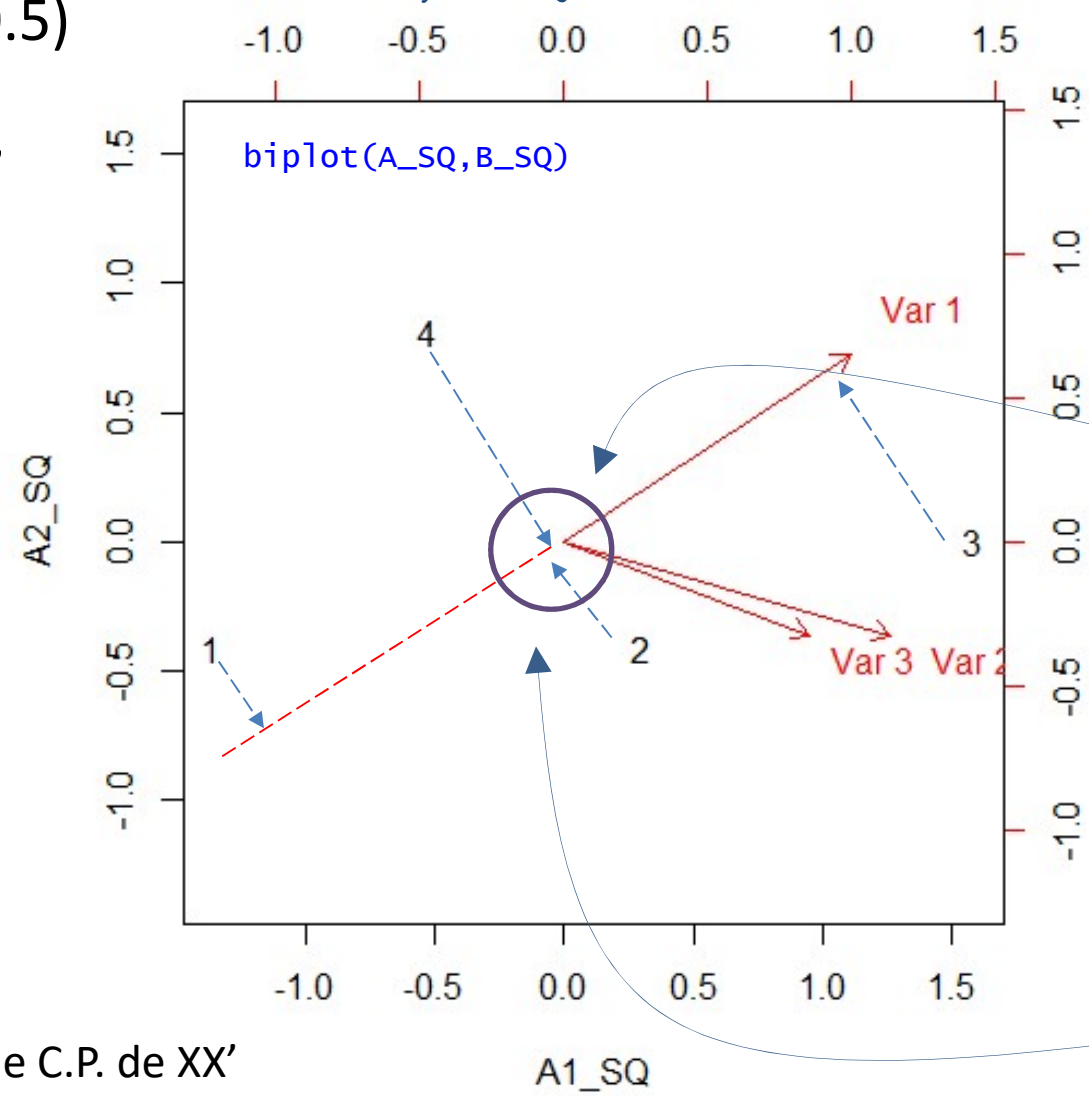
$F(nxp)=U*L^{1/4}$

$G(pxp)= L^{1/4}\,U'$

$$F=\begin{pmatrix} 1.36 & -0.41 \\ -0.3 & -0.41 \\ -1.58 & 0 \\ 0.53 & 0.82 \end{pmatrix}$$

$$G=\begin{pmatrix} -1.25 & 0.82 \\ -1.42 & -0.41 \\ -1.07 & -0.41 \end{pmatrix}$$

Gráfica objeto-variables



Se proyectan filas sobre columnas

Son aprox. Iguales a cero en la primera columna de **X**

Var 1

$$X=\begin{pmatrix} -2 & -2 & -1 \\ 0 & 1 & 0 \\ 2 & 2 & 2 \\ 0 & -1 & -1 \end{pmatrix}$$

F: scores objetos de C.P. de XX'

G: scores vector de C.P. de X'X

SQ-Biplot (c=0.5)

$F(nxp)=U*L^{1/4}$

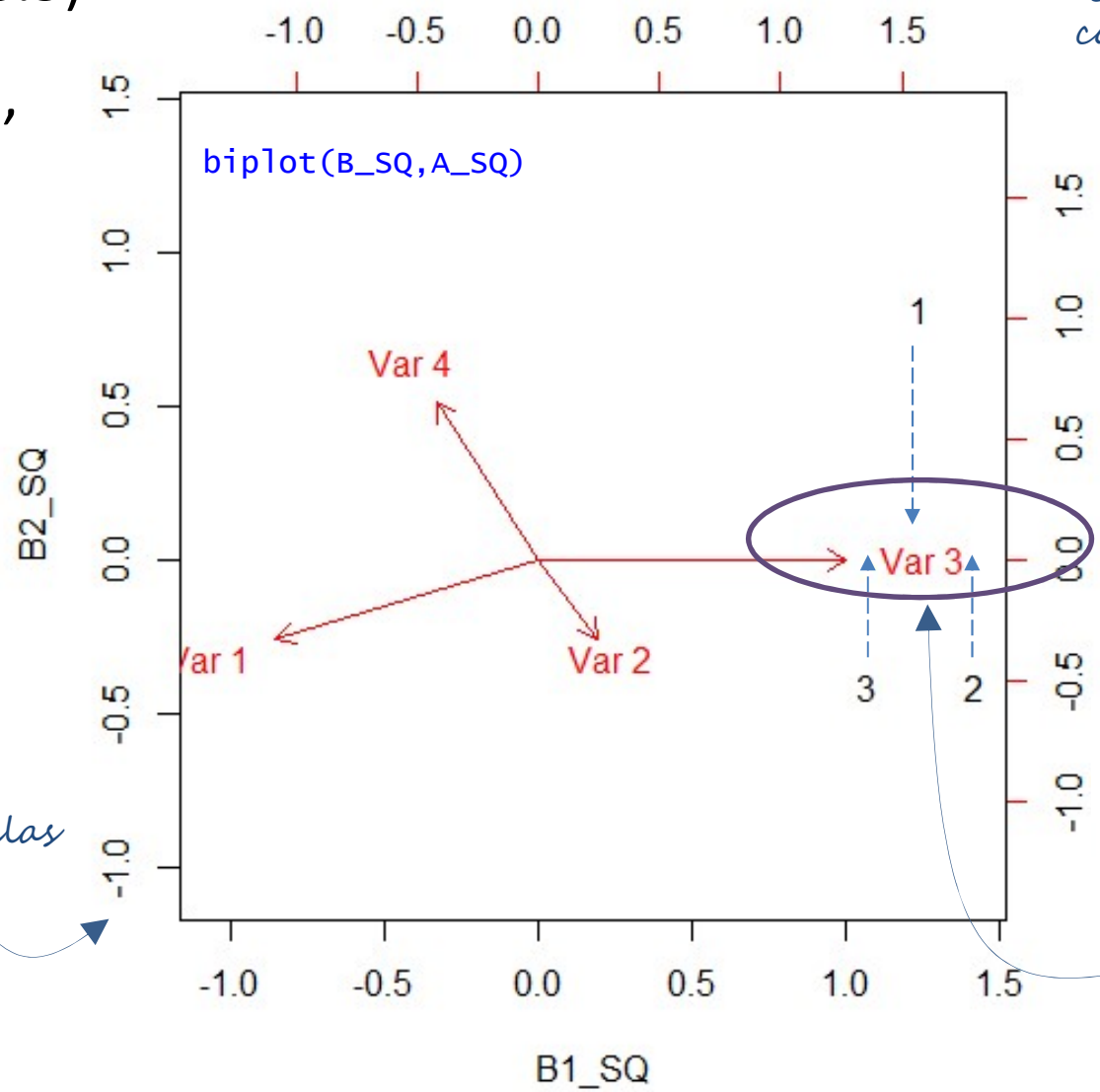
$G(pxp)= L^{1/4}\,U'$

$$F=\begin{pmatrix} 1.36 & -0.41 \\ -0.3 & -0.41 \\ -1.58 & 0 \\ 0.53 & 0.82 \end{pmatrix}$$

$$G=\begin{pmatrix} -1.25 & 0.82 \\ -1.42 & -0.41 \\ -1.07 & -0.41 \end{pmatrix}$$

Gráfica punto-vector

Son aprox. Iguales como la tercera fila de **X**

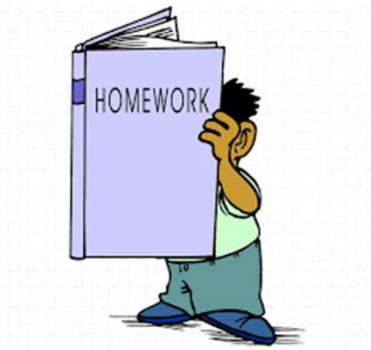


$$X=\begin{pmatrix} -2 & -2 & -1 \\ 0 & 1 & 0 \\ 2 & 2 & 2 \\ 0 & -1 & -1 \end{pmatrix}$$

Se proyectan columnas sobre filas

VISUALIZACIÓN E INTERPRETACIÓN

```
library("FactoMineR")
res.pca <- PCA(decathlon2.active, graph = FALSE)
print(res.pca)
```




```
   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$contrib"    "contributions of the individuals"
11 "$call"           "summary statistics"
12 "$call$centre"    "mean of the variables"
13 "$call$ecart.type" "standard error of the variables"
14 "$call$row.w"     "weights for the individuals"
15 "$call$col.w"     "weights for the variables"
```

Gráfico de variables

```
fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 50))
```

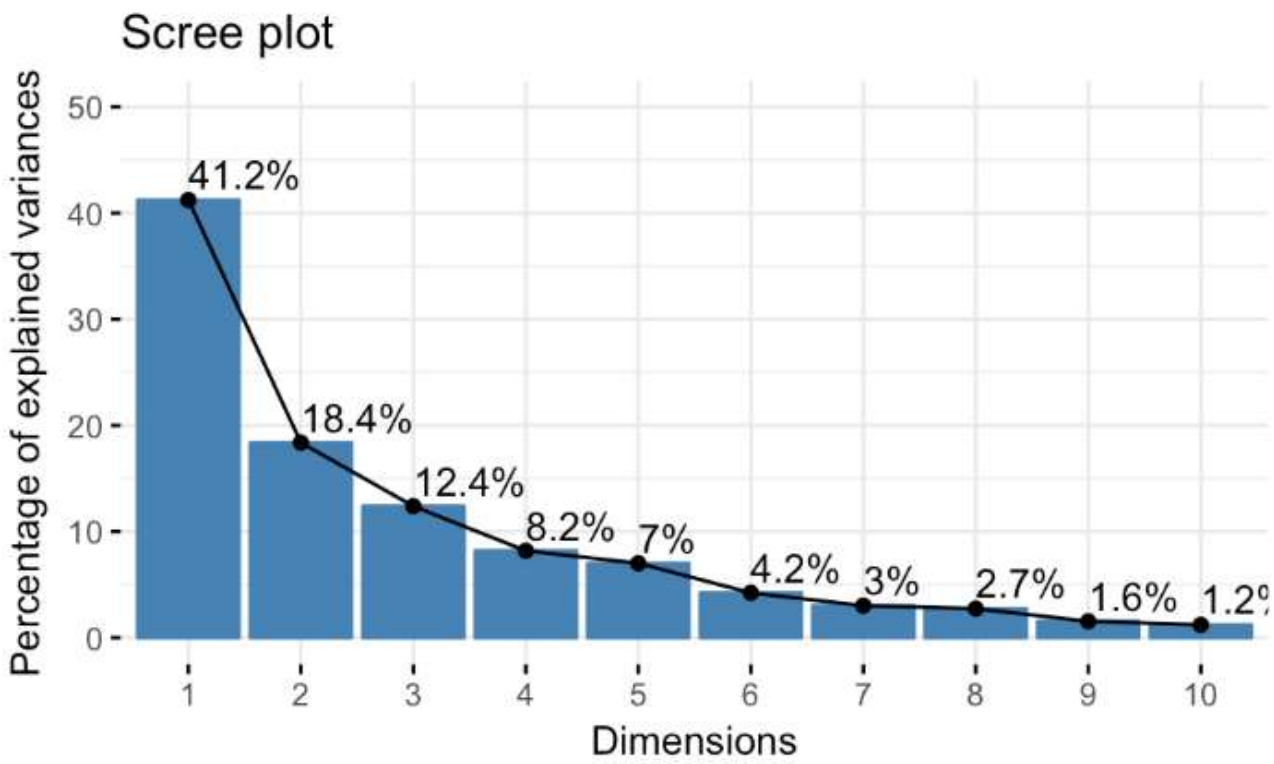


Gráfico de variables (continuación…)

```
get_eigenvalue(res.pca)
```

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.1242133	41.242133	41.24213
Dim.2	1.8385309	18.385309	59.62744
Dim.3	1.2391403	12.391403	72.01885
Dim.4	0.8194402	8.194402	80.21325
Dim.5	0.7015528	7.015528	87.22878
Dim.6	0.4228828	4.228828	91.45760
Dim.7	0.3025817	3.025817	94.48342
Dim.8	0.2744700	2.744700	97.22812
Dim.9	0.1552169	1.552169	98.78029
Dim.10	0.1219710	1.219710	100.00000

Gráfico de variables

```
fviz_pca_var(res.pca)
```

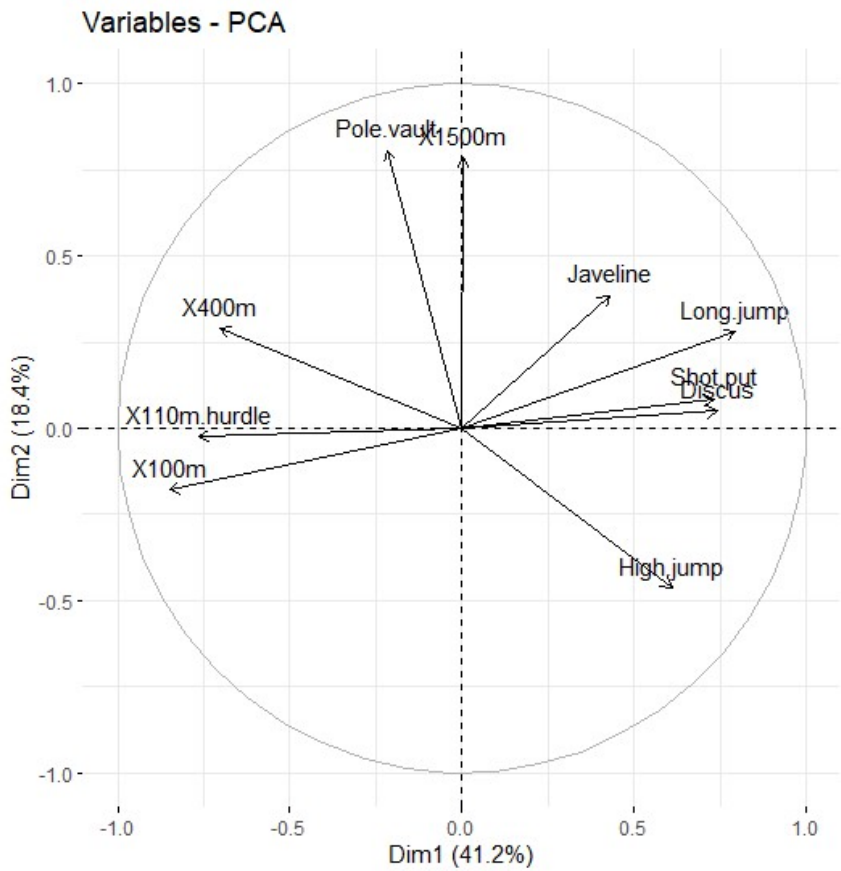


Gráfico de variables (continuación…)

```
fviz_pca_var(res.pca,
 col.var = "cos2",
# Color según calidad de representación
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
 repel = TRUE # Evita superposition de texto
)
```

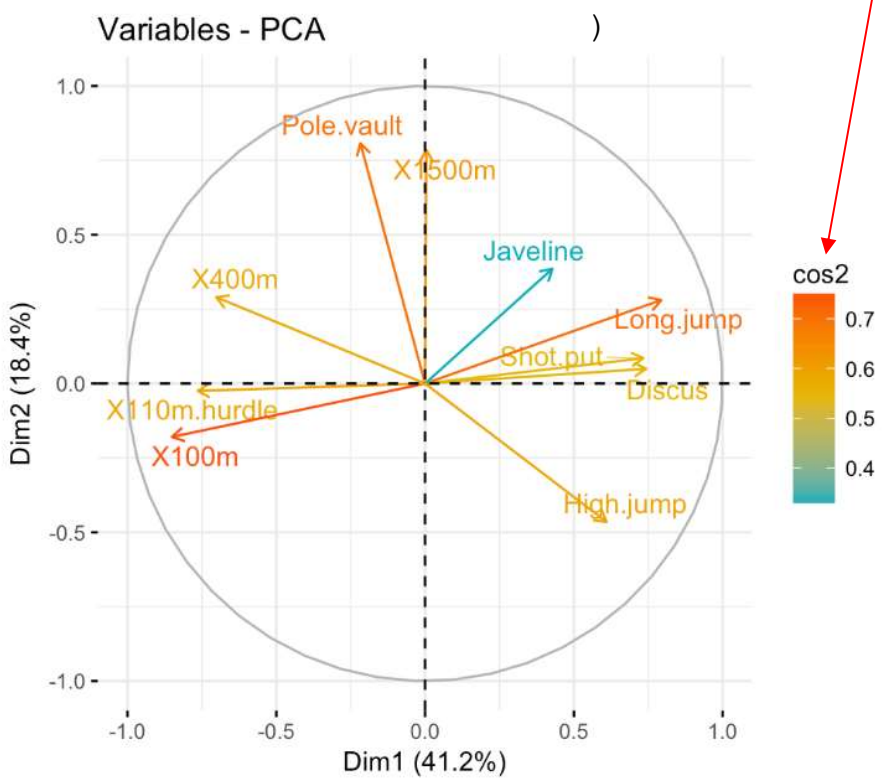


Gráfico de variables (continuación…)

```
fviz_pca_var(res.pca,
 col.var = "contrib",
# Color según calidad de representación
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
 repel = TRUE # Evita superposition de texto
)
```

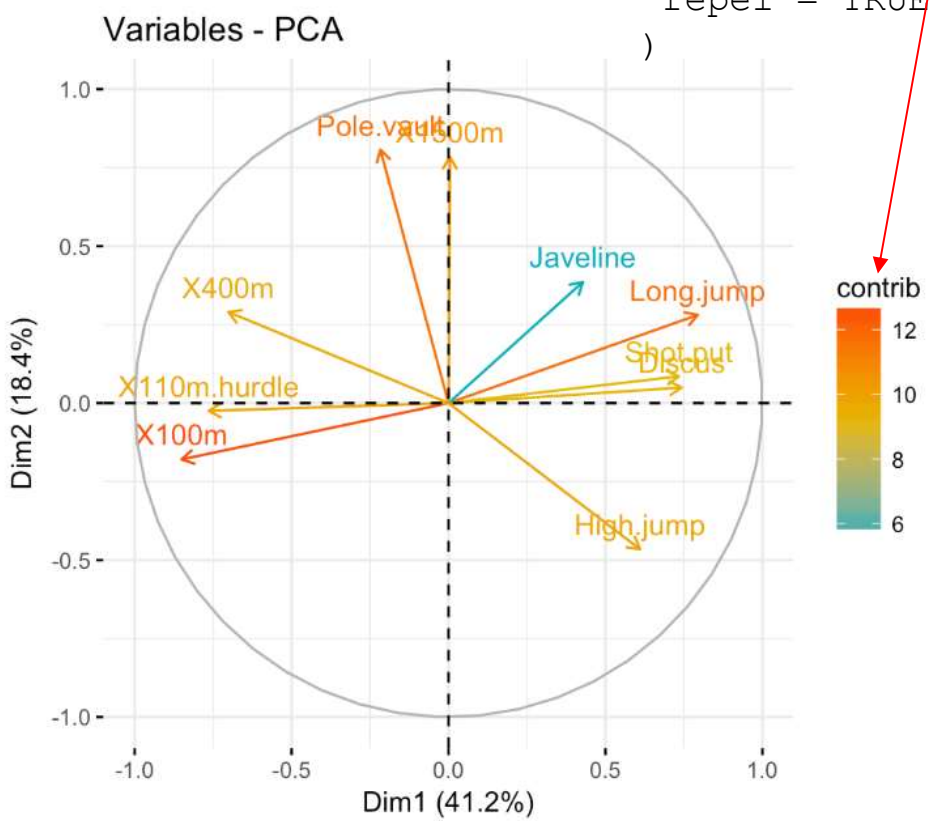
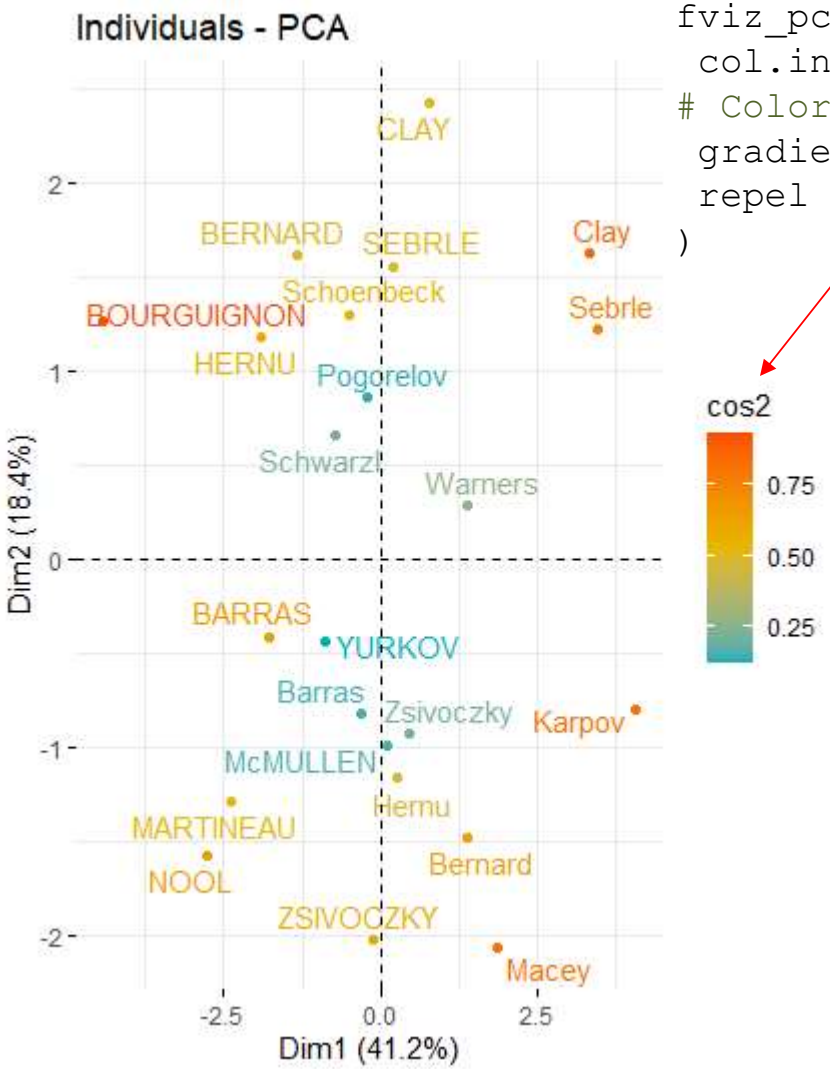


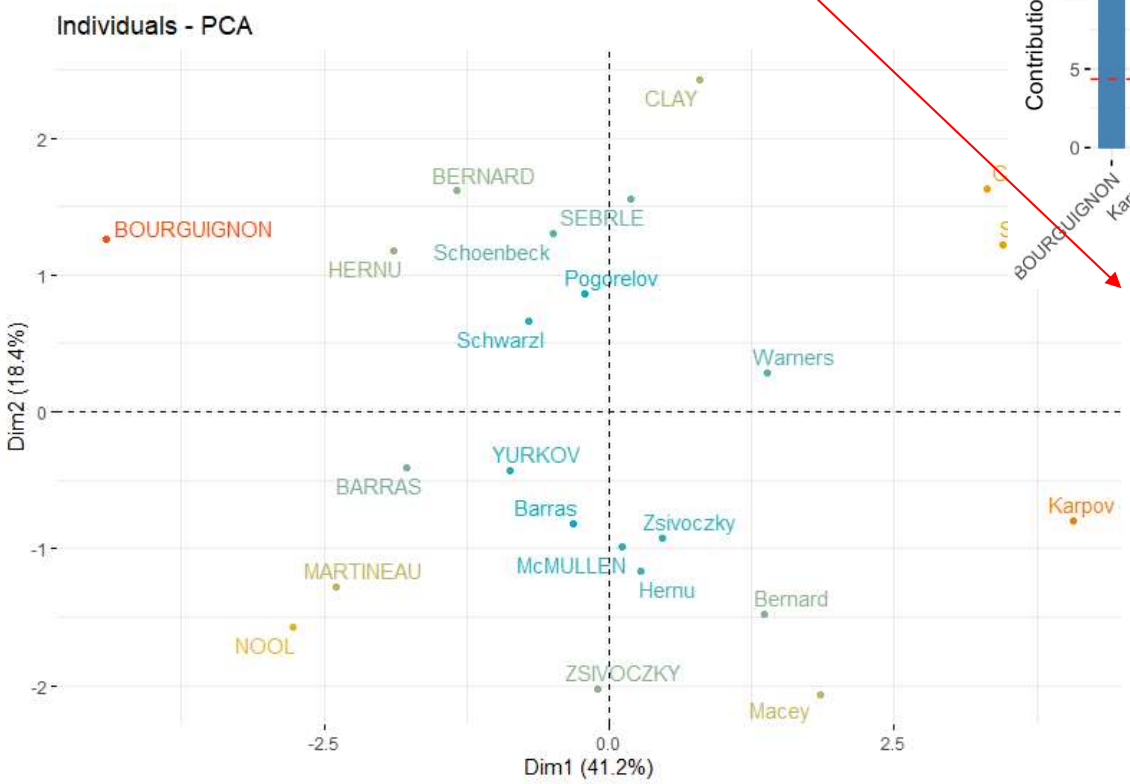
Gráfico de individuos (continuación…)

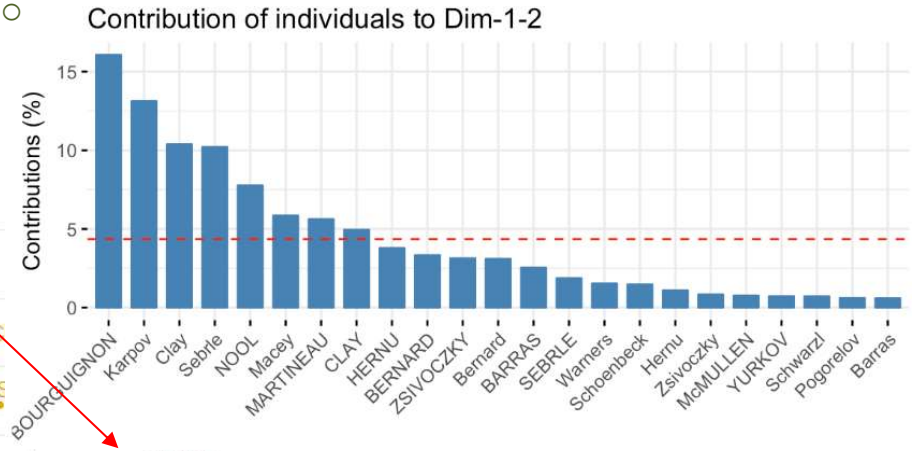



```
fviz_pca_ind(res.pca,
 col.ind = "cos2",
# Color según calidad de representación
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
 repel = TRUE # Evita superposition de texto
)
```

Gráfico de individuos (continuación…)

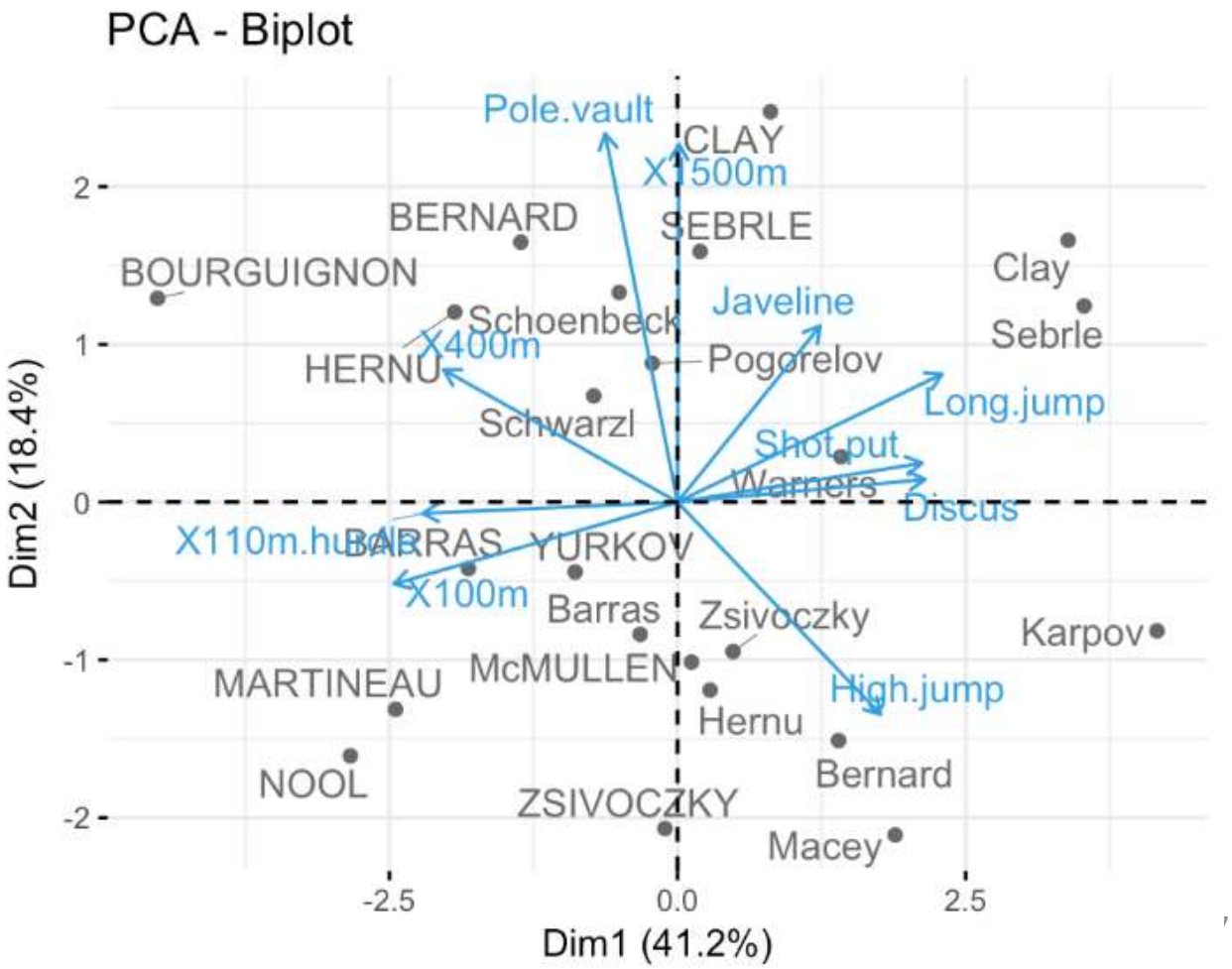
```
fviz_pca_ind(res.pca,
 col.ind = "contrib",
# Color según calidad de representación
 gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
 repel = TRUE # Evita superposition de texto
)
```





BIPLOT

```
fviz_pca_biplot(res.pca, repel = TRUE,
 col.var = "#2E9FDF", # Variables color
 col.ind = "#696969" # Individuals color )
```



RECURSOS ANALÍTICOS

- ✓ Objetos activos
- ✓ Variables activas
- ✓ Objetos suplementarios (pasivos)
- ✓ Variables cuantitativas suplementarias (pasivas)
- ✓ Variables cualitativas suplementarias (pasivas)
- ✓ Calidad de representación de los objetos a la CP
- ✓ Calidad de representación de las variables a la CP

Dataset: `decathlon2`
Paquete: `factoextra`

10 variables

23 objetos

4 objetos

name	100m	Long.jump	//	Javeline	1500m	Rank	Points	Competition
SEBRLE	11.04	7.58		63.19	291.7	1	8217	Decastar
CLAY	10.76	7.4		60.15	301.5	2	8122	Decastar
Macey	10.89	7.47		58.46	265.42	4	8414	OlympicG
Warners	10.62	7.74		55.39	278.05	5	8343	OlympicG
\\								
Zsivoczky	10.91	7.14		63.45	269.54	6	8287	OlympicG
Hernu	10.97	7.19		57.76	264.35	7	8237	OlympicG
Pogorelov	10.95	7.31		53.45	287.63	11	8084	OlympicG
Schoenbeck	10.9	7.3		60.89	278.82	12	8077	OlympicG
Barras	11.14	6.99		64.55	267.09	13	8067	OlympicG
KARPOV	11.02	7.3		50.31	300.2	3	8099	Decastar
WARNERS	11.11	7.6		51.77	278.1	6	8030	Decastar
Nool	10.8	7.53		61.33	276.33	8	8235	OlympicG
Drews	10.87	7.38		51.53	274.21	19	7926	OlympicG

Active individuals
Active variables
Supplementary **quantitative** variables
Supplementary **qualitative** variable
Supplementary individuals

PREDICCIÓN DE VARIABLES SUPLEMENTARIAS

$$\mathbf{X=AB'}$$

Representación gráfica de X en dos dimensiones

$$\mathbf{X=U*L^{c/2}L^{(1-c)/2}U'}$$

$$0 \leq c \leq 1$$

C=1 JK biplot $\mathbf{A=V^*}$ y $\mathbf{B'=U'}=Y^*$

Por lo que se obtiene una buena representación por filas

Preserva la métrica de filas

$$Z=AB'_p+e$$

$$\boxed{B'_p=(A'A)^{-1}A'Z}$$

Z: valores reales de variables suplementarias

A: valores marcadores de fila

$\mathbf{B'_p}$: valores marcadores de predicción de **B'**

$$B'_p=(A'A)^{-1}A'Z$$

"Las coordenadas de las variables serán predecidos usando la información de los parámetros obtenidos por componentes principals (variables activas) y las variables suplementarias"

VARIABLES SUPLEMENTARIAS:

VARIABLES SUPLEMENTARIAS (Escala RANGO): Columnas 11 y 12

VARIABLES SUPLEMENTARIAS (Escala NOMINAL): Columna 13 (nominal)

OBJETOS O INDIVIDUOS ACTIVOS:

VARIABLES ACTIVAS:

se utilizan en el cálculo de las components principales.

```
library("factoextra")
data(decathlon2)
decathlon2.active <- decathlon2[1:23, 1:10]
```

OBJETOS Y VARIABLES ACTIVAS:

> decathlon2.active

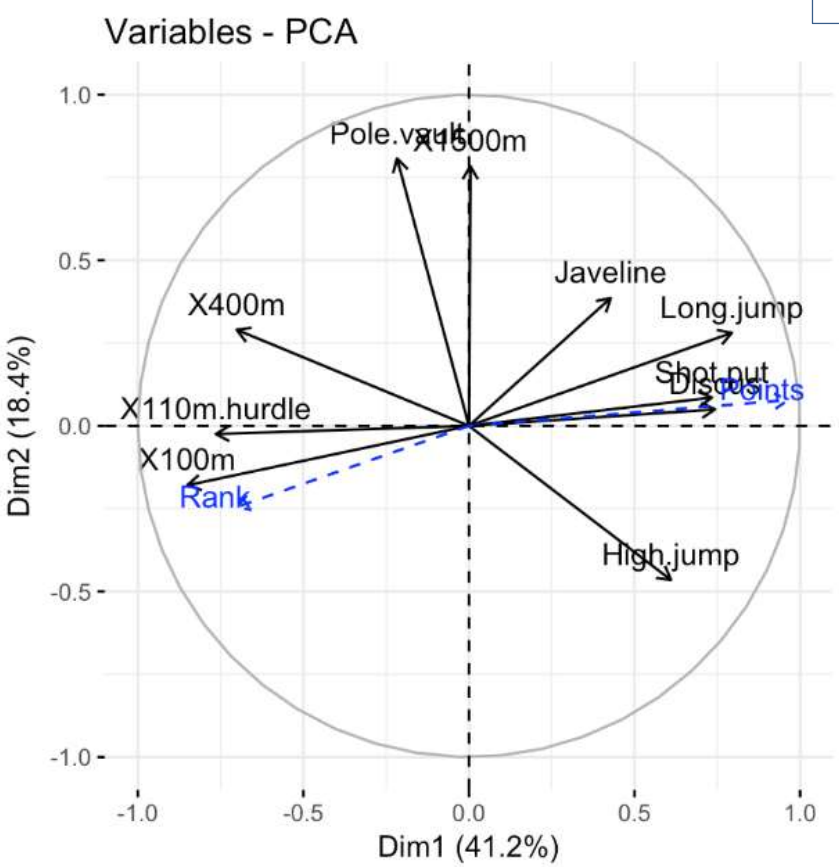
	X100m	Long.jump	Shot.put	High.jump	X400m	X110m.hurdle	Discus	Pole.vault	Javeline	X1500m
SEBRLE	11.04	7.58	14.83	2.07	49.81	14.69	43.75	5.02	63.19	291.70
CLAY	10.76	7.40	14.26	1.86	49.37	14.05	50.72	4.92	60.15	301.50
BERNARD	11.02	7.23	14.25	1.92	48.93	14.99	40.87	5.32	62.77	280.10
YURKOV	11.34	7.09	15.19	2.10	50.42	15.31	46.26	4.72	63.44	276.40
ZSIVOCZKY	11.13	7.30	13.48	2.01	48.62	14.17	45.67	4.42	55.37	268.00
McMULLEN	10.83	7.31	13.76	2.13	49.91	14.38	44.41	4.42	56.37	285.10
MARTINEAU	11.64	6.81	14.57	1.95	50.14	14.93	47.60	4.92	52.33	262.10
HERNU	11.37	7.56	14.41	1.86	51.10	15.06	44.99	4.82	57.19	285.10
BARRAS	11.33	6.97	14.09	1.95	49.48	14.48	42.10	4.72	55.40	282.00
NOOL	11.33	7.27	12.68	1.98	49.20	15.29	37.92	4.62	57.44	266.60
BOURGUIGNON	11.36	6.80	13.46	1.86	51.16	15.67	40.49	5.02	54.68	291.70
Sebrle	10.85	7.84	16.36	2.12	48.36	14.05	48.72	5.00	70.52	280.01
Clay	10.44	7.96	15.23	2.06	49.19	14.13	50.11	4.90	69.71	282.00
Karpov	10.50	7.81	15.93	2.09	46.81	13.97	51.65	4.60	55.54	278.11
Macey	10.89	7.47	15.73	2.15	48.97	14.56	48.34	4.40	58.46	265.42
Warners	10.62	7.74	14.48	1.97	47.97	14.01	43.73	4.90	55.39	278.05
Zsivoczky	10.91	7.14	15.31	2.12	49.40	14.95	45.62	4.70	63.45	269.54
Hernu	10.97	7.19	14.65	2.03	48.73	14.25	44.72	4.80	57.76	264.35
Bernard	10.69	7.48	14.80	2.12	49.13	14.17	44.75	4.40	55.27	276.31
Schwarzl	10.98	7.49	14.01	1.94	49.76	14.25	42.43	5.10	56.32	273.56
Pogorelov	10.95	7.31	15.10	2.06	50.79	14.21	44.60	5.00	53.45	287.63
Schoenbeck	10.90	7.30	14.77	1.88	50.30	14.34	44.41	5.00	60.89	278.82
Barras	11.14	6.99	14.91	1.94	49.41	14.37	44.83	4.60	64.55	267.09

ELEMENTOS SUPLEMENTARIOS

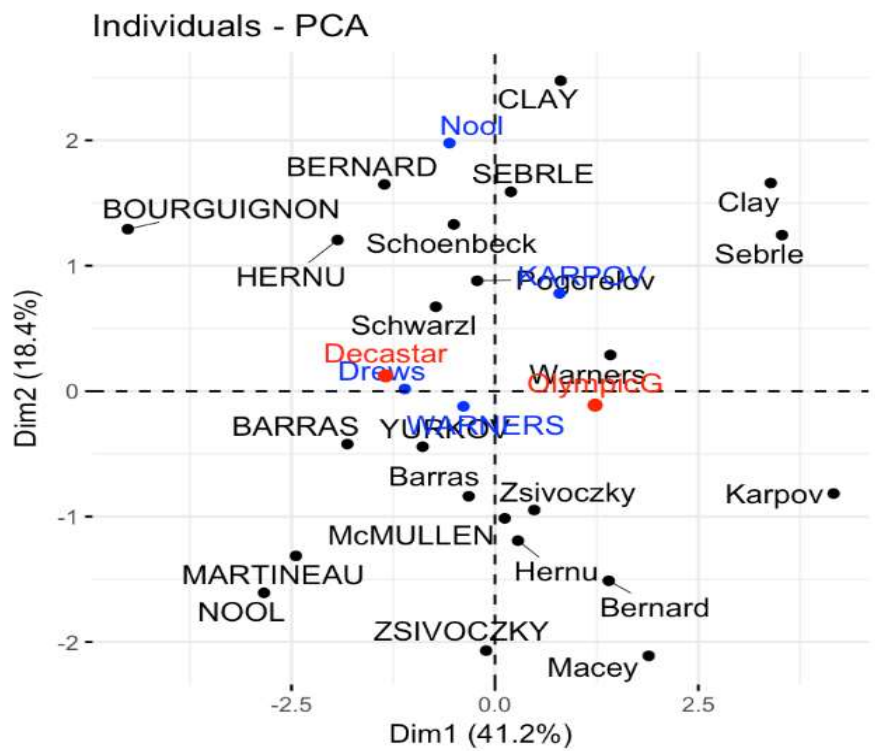
```
res.pca.sup <- PCA(decathlon2, ind.sup = 24:27,
quanti.sup = 11:12, quali.sup = 13, graph=FALSE)
```

X : a data frame. Las filas son individuos/objetos y columnas son variables numéricas.
ind.sup : vector numérico que especifica el índice de fila de los individuos suplementarios.
quanti.sup, quali.sup : vector numérico que especifica los índices de las variables cuantitativas y cualitativas respectivamente.
graph : valor lógico. Si es TRUE se muestra un gráfico.

```
fviz_pca_var(res.pca.sup)
```

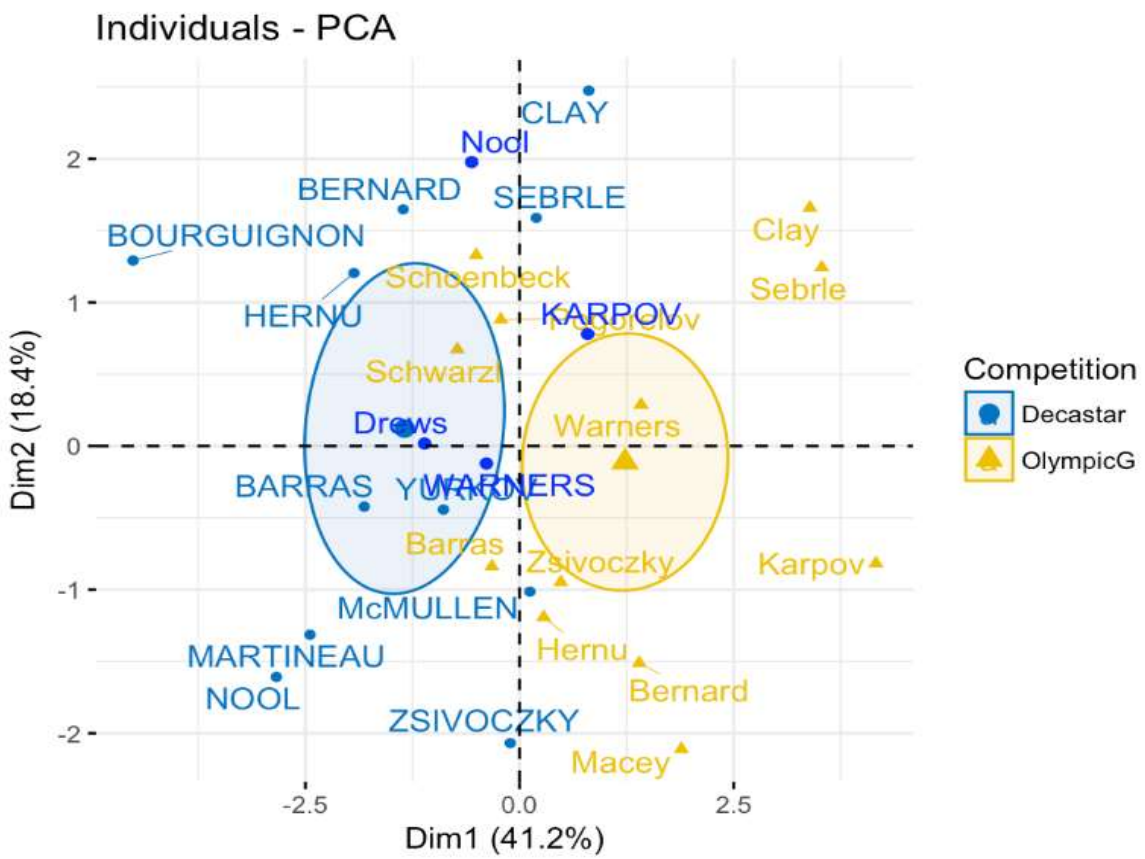



```
p <- fviz_pca_ind(res.pca.sup, col.ind.sup
= "blue", repel = TRUE)
p <- fviz_add(p,
res.pca.sup$quali.sup$coord, color = "red")
p
```



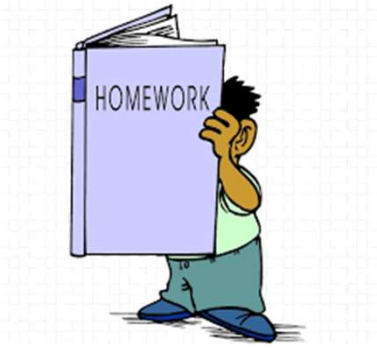
Individuos por variable cualitativa

```
fviz_pca_ind(res.pca, habillage = 13,
 addEllipses =TRUE, ellipse.type = "confidence",
 palette = "jco", repel = TRUE)
```



El argumento **habillage** se usa para especificar el índice de la variable cualitativa suplementaria.

ESTUDIAR LA SIGUIENTE INFORMACIÓN:



TESIS: APLICACIÓN DE METODOS MULTIVARIADOS PARA EL ESTUDIO DE LA ASOCIACION DE FACTORES DE RIESGO CARDIOVASCULAR Y DETERMINANTES SOCIALES: EVOLUCION EN EL TIEMPO EN CHILE

NOTAS: Applied Correspondence Analysis

Optimal representation of supplementary variables in biplots form PCA - GRAFFELMAN

"La evidencia grafica será más clara en la medida que las dos primeras componentes principales expliquen la mayor proporción de inercia."