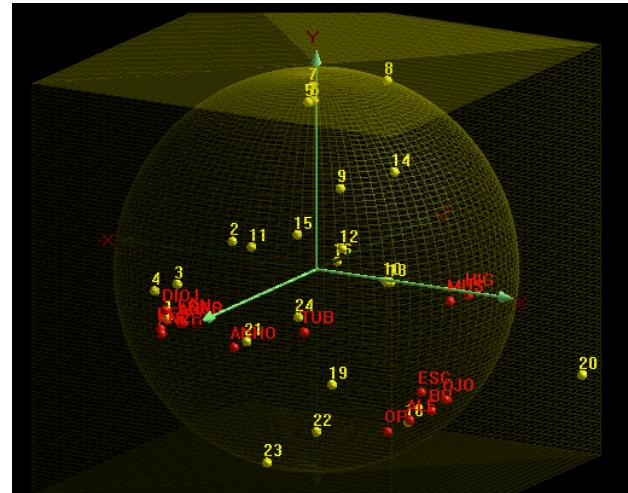




Métodos de Reducción de la Dimensión: *Análisis en Componentes Principales*

[Una introducción]

Análisis en Componentes Principales - ACP



- El ACP es el método más utilizado en Minería de Datos (en Francia).
 - Fue propuesto en 1933 por Hotelling.
 - La primera implementación computacional se dío en los años 60.
 - Fue aplicado para analizar encuestas de opinión pública por Jean Pages.
-
- **Objetivo:** Construir un pequeño número de nuevas variables (componentes) en las cuales se concentre la mayor cantidad posible de información, como se ilustra en la Figura 1

Los Datos

Se parte de una tabla de datos:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1j} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{i1} & \cdots & x_{ij} & \cdots & x_{im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nj} & \cdots & x_{nm} \end{pmatrix} \longleftrightarrow \text{individuo } i$$

↑
Variable j

Ejemplo:

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

Nubes de Puntos

INDIVIDUOS - FILAS

Luis	5.0	6.5	6.5	7.0	9.0
------	-----	-----	-----	-----	-----

$$\in \mathbb{R}^5$$

VARIABLES - COLUMNAS

Español
9.2
7.3
8.0
6.5
7.8
7.7
8.2
7.5
6.5
8.7

$$\in \mathbb{R}^{10}$$

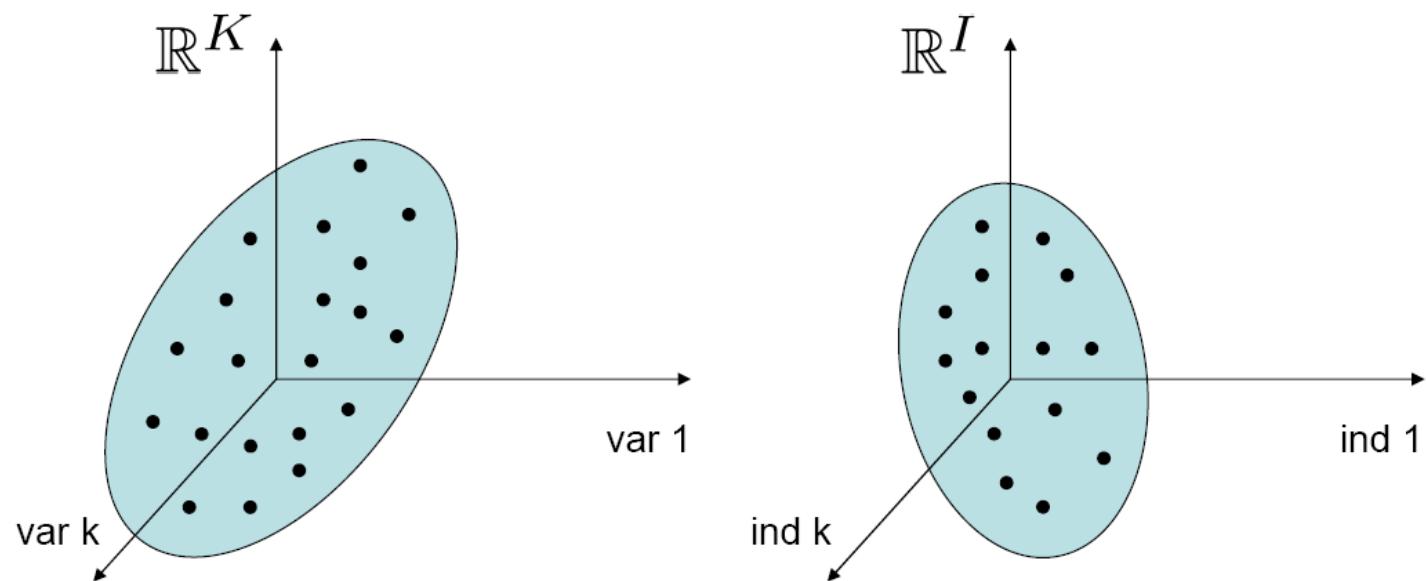
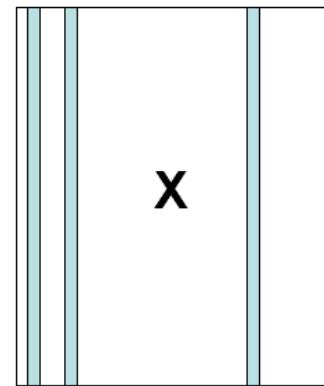
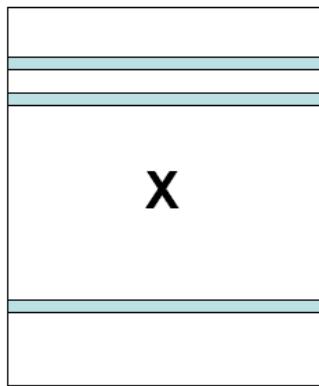


Tabla de Datos

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

$$\xrightarrow{\hspace{1cm}} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{np} \end{bmatrix}$$

100% de la información

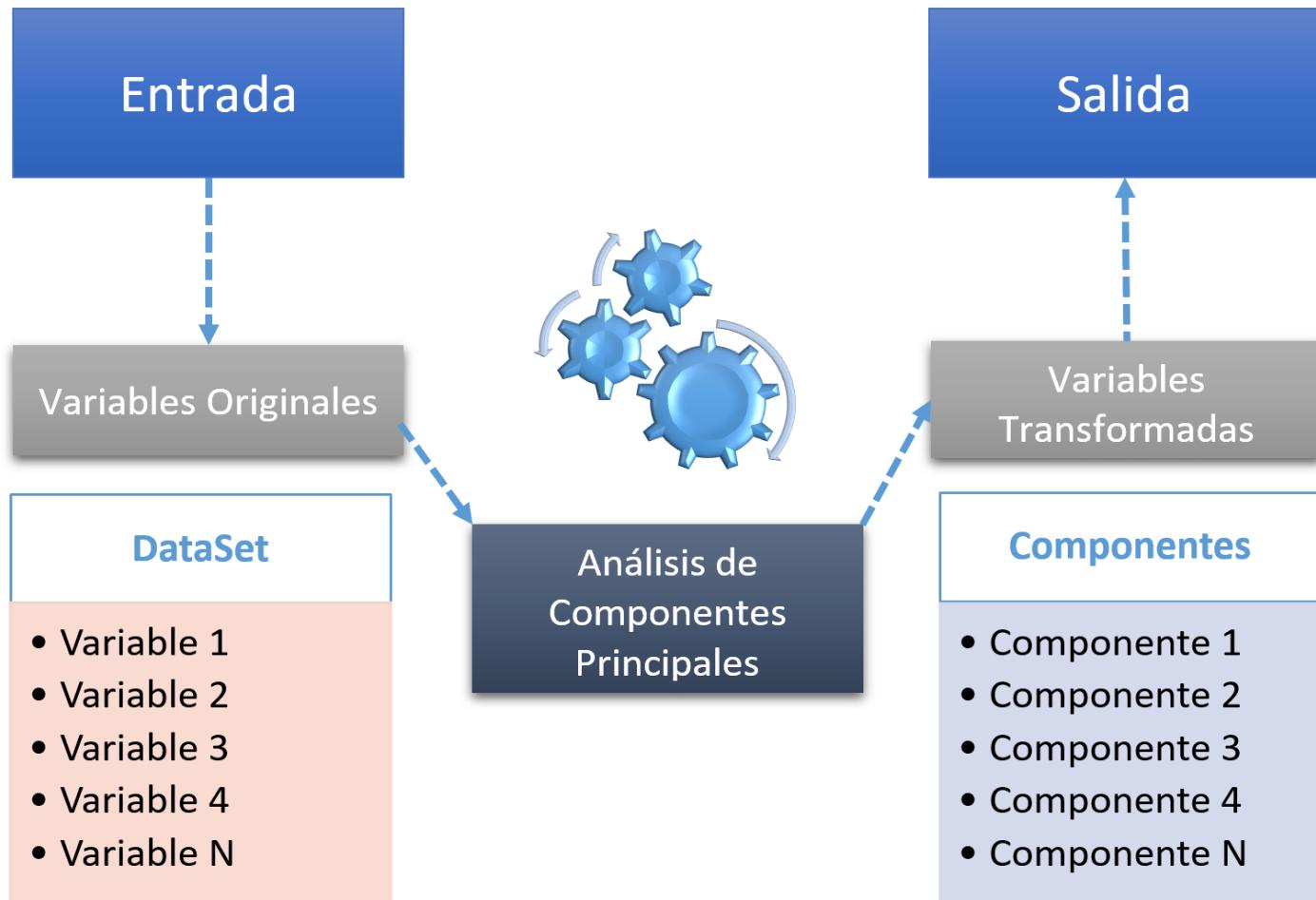
80%

16%

.....

0.02%

Transformación de las variables originales en componentes



DATOS

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

COMPONENTES

	Comp1	Comp2	Comp3	Comp4	Comp5
Lucia	0,3231	1,7725	1,1988	-0,055	0,0036
Pedro	0,6654	-1,6387	0,1455	-0,0231	-0,1234
Ines	1,0025	-0,5157	0,6289	0,5164	0,1429
Luis	-3,1721	-0,2628	-0,382	0,6778	-0,0625
Andres	-0,4889	1,3654	-0,8352	-0,1558	0,1234
Ana	1,7086	-1,0217	-0,1271	0,0668	0,0253
Carlos	0,0676	1,4623	-0,5062	-0,1179	0,0131
José	2,0119	-1,2759	-0,5422	-0,1978	0,0174
Sonia	-3,042	-1,2549	0,4488	-0,64	0,0379
Maria	0,9239	1,3694	-0,0293	-0,0715	-0,1777

DATOS

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

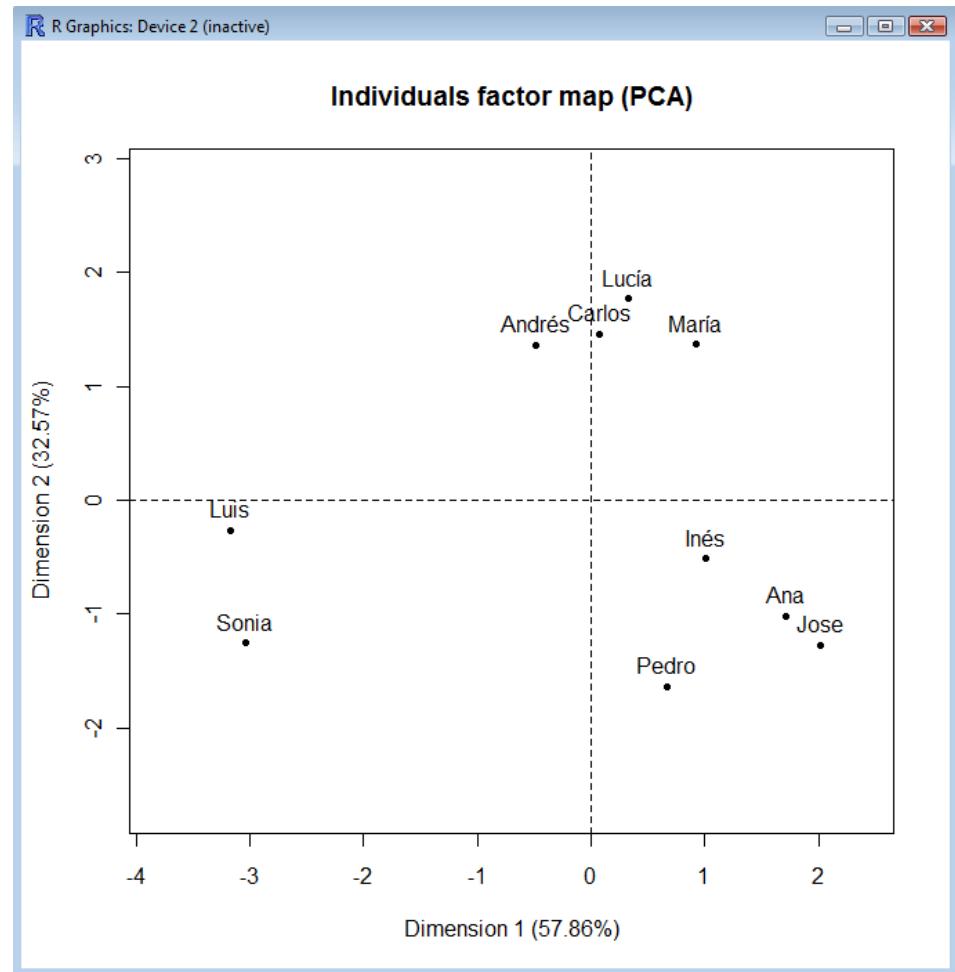
COMPONENTES

	Comp1	Comp2			
Lucia	0,3231	1,7725			
Pedro	0,6654	-1,6387			
Ines	1,0025	-0,5157			
Luis	-3,1721	-0,2628			
Andres	-0,4889	1,3654			
Ana	1,7086	-1,0217			
Carlos	0,0676	1,4623			
Jose	2,0119	-1,2759			
Sonia	-3,042	-1,2549			
Maria	0,9239	1,3694			

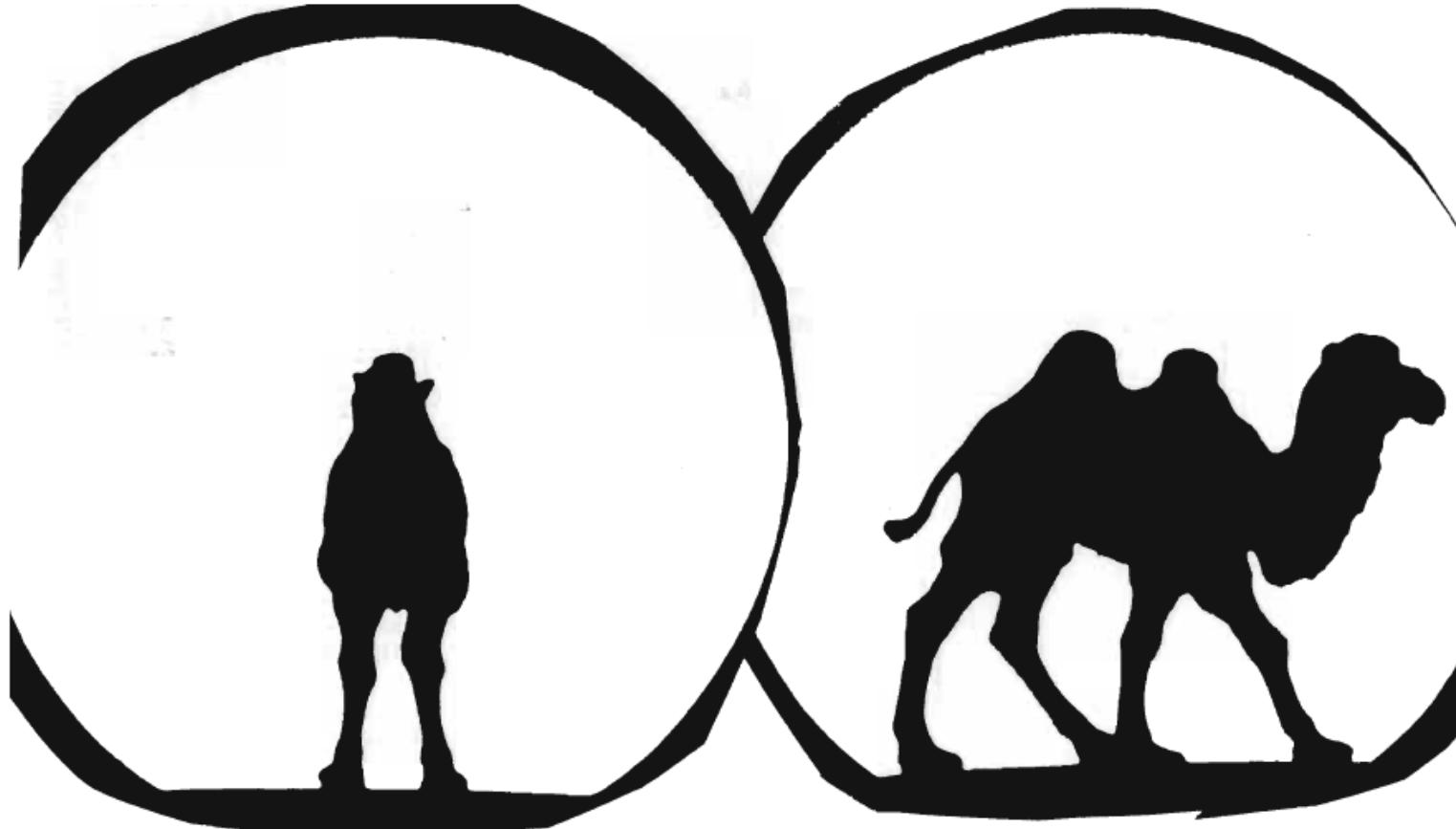
Plano Principal

COMPONENTES

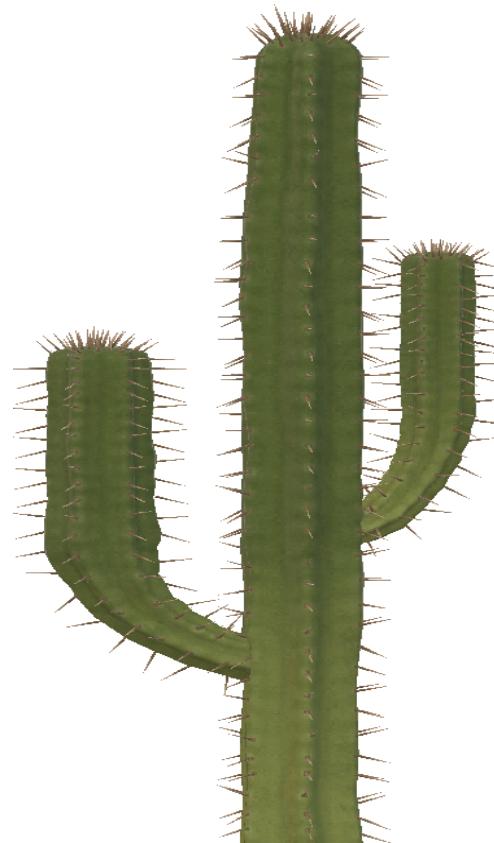
	Comp1	Comp2			
Lucia	0,3231	1,7725			
Pedro	0,6654	-1,6387			
Ines	1,0025	-0,5157			
Luis	-3,1721	-0,2628			
Andres	-0,4889	1,3654			
Ana	1,7086	-1,0217			
Carlos	0,0676	1,4623			
Jose	2,0119	-1,2759			
Sonia	-3,042	-1,2549			
Maria	0,9239	1,3694			



Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



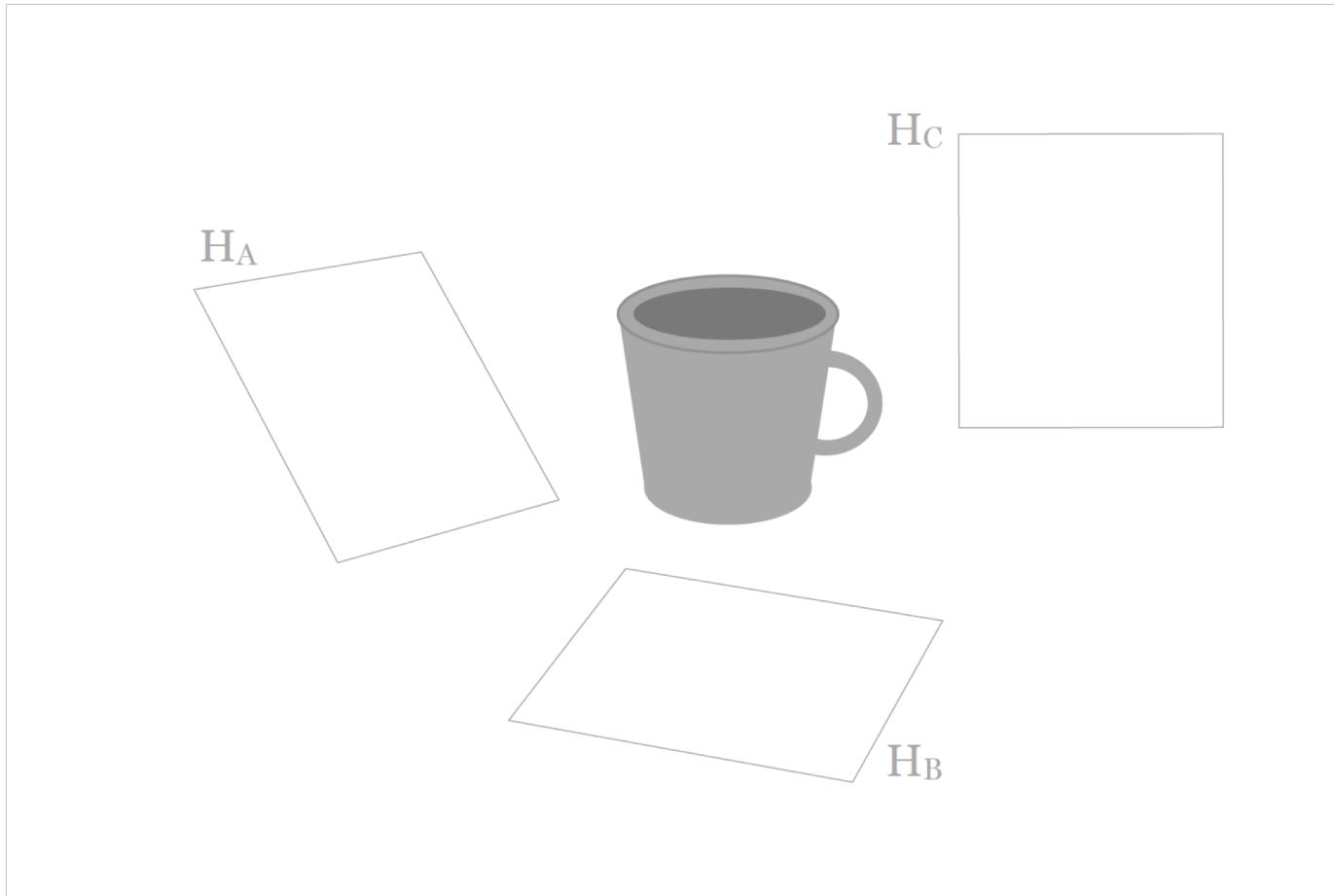
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



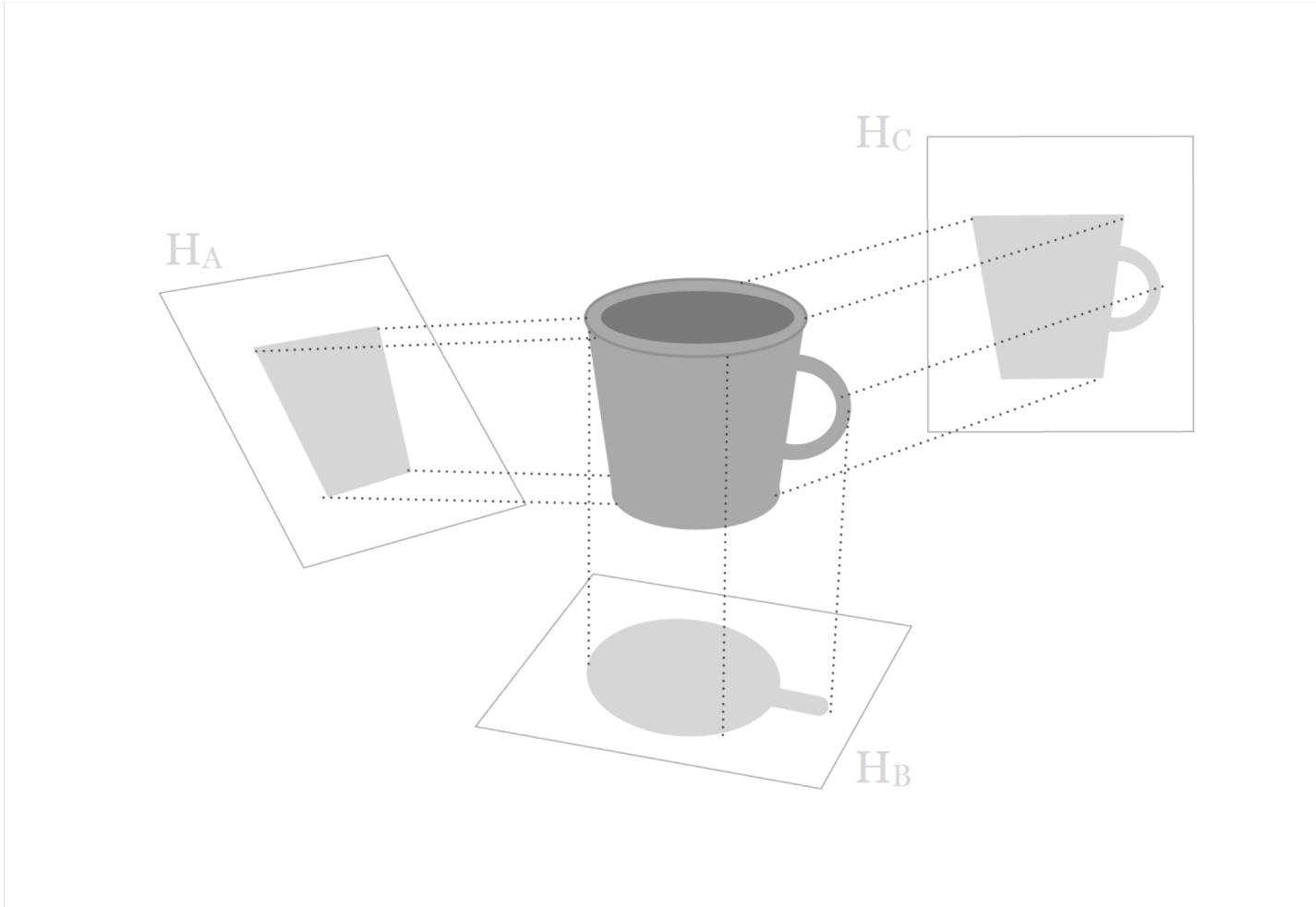
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



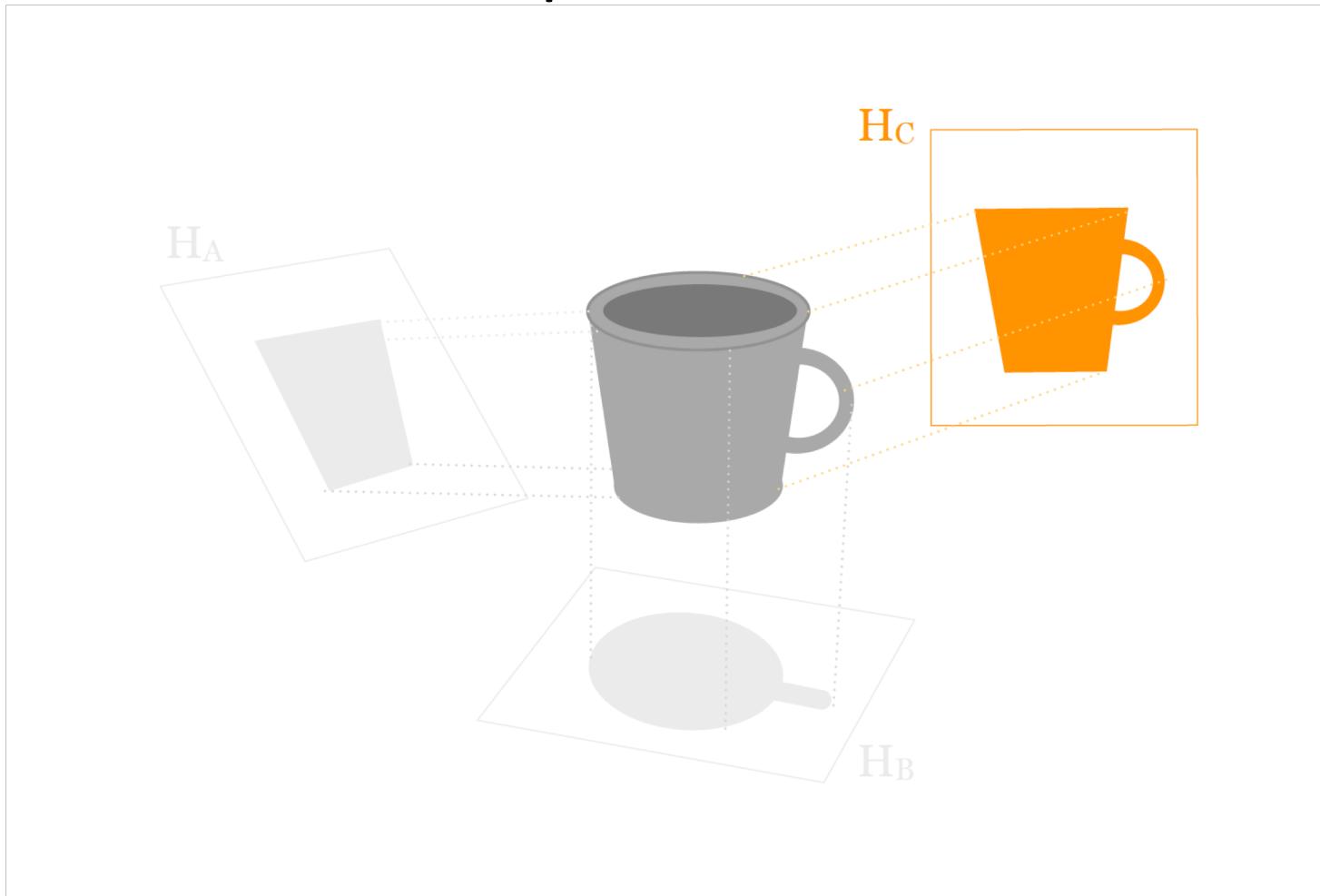
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



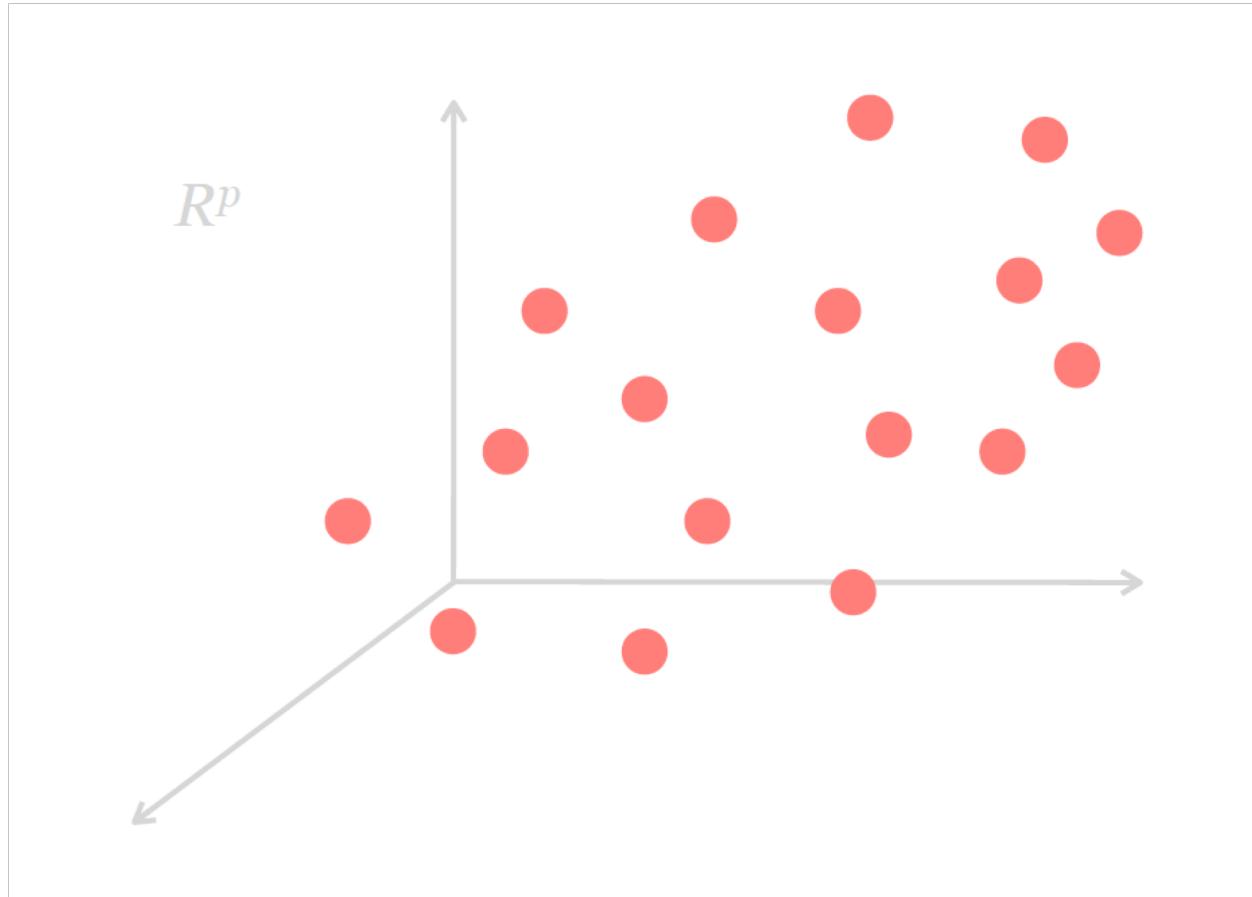
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



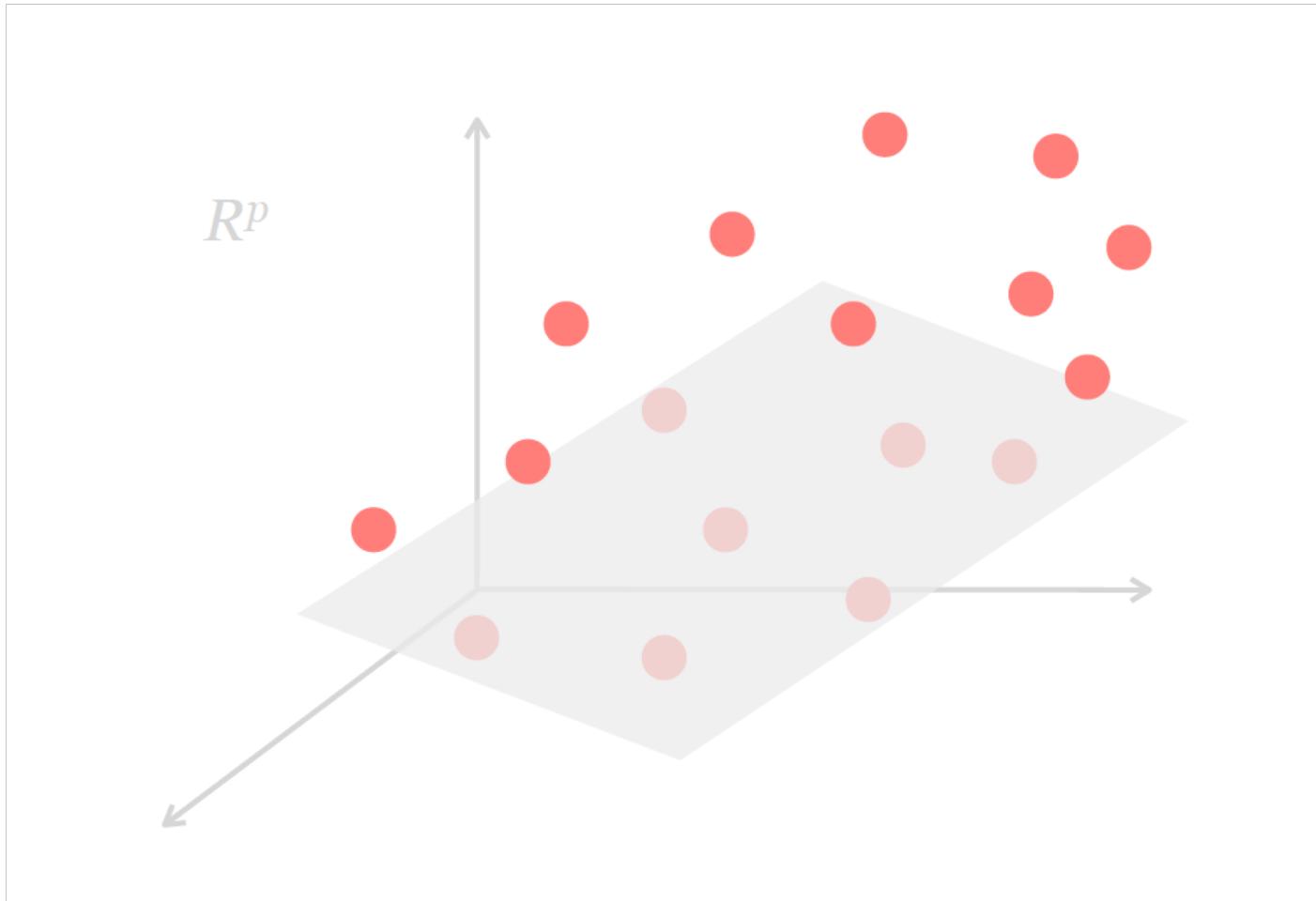
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



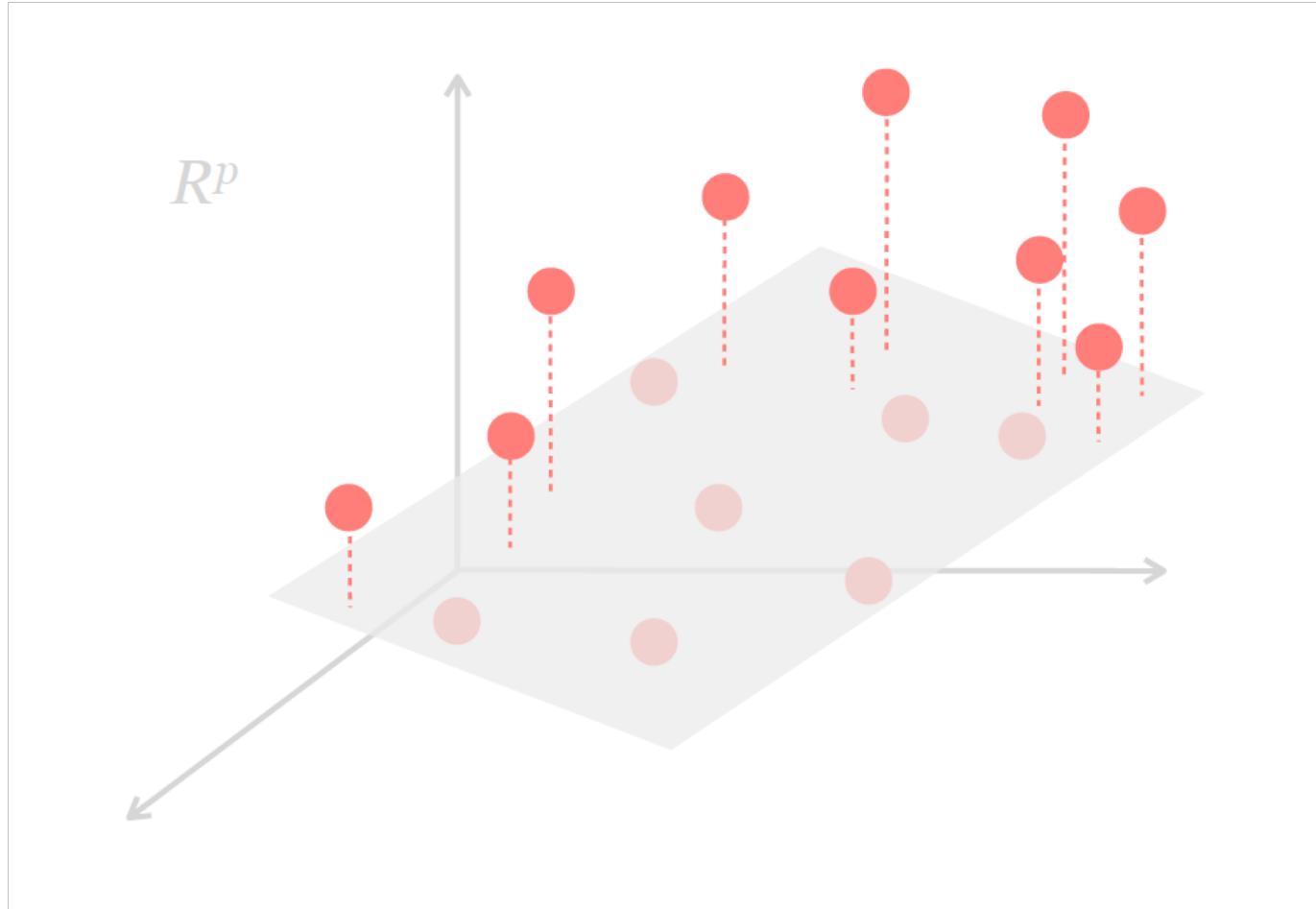
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



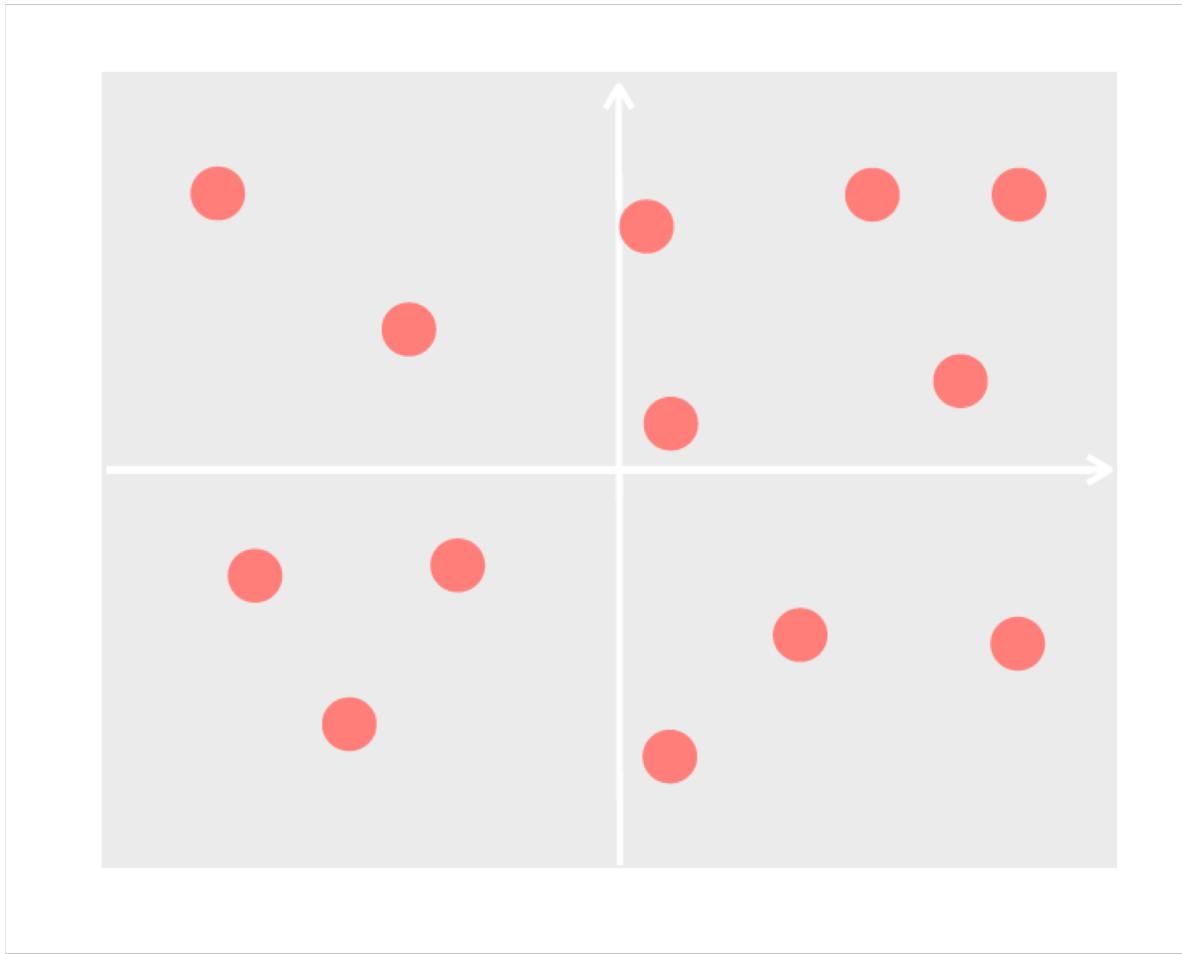
Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.

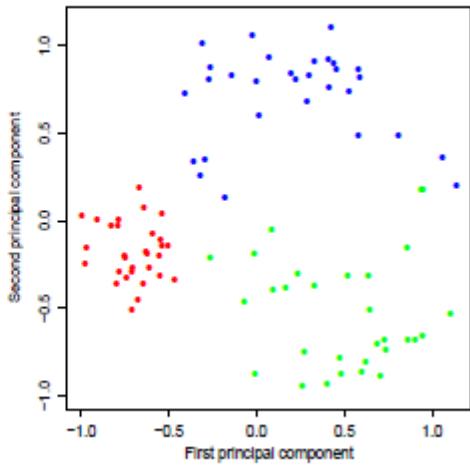
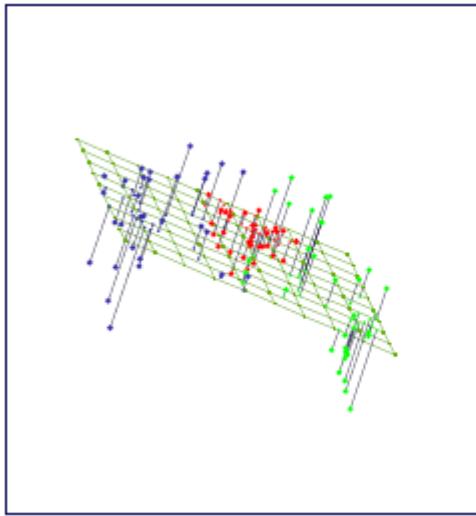
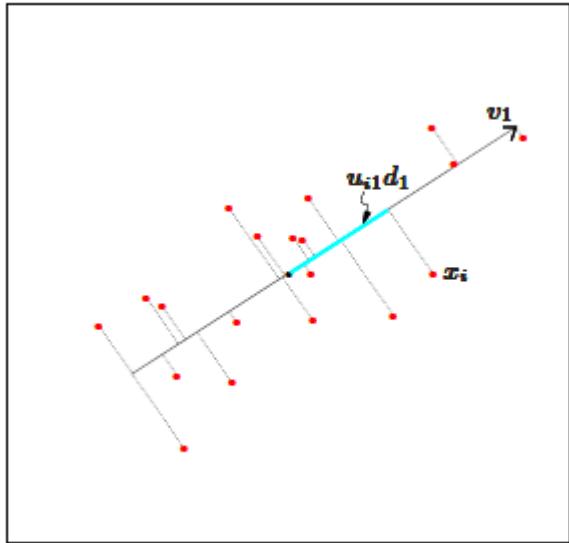


Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.



Objetivo: Encontrar el mejor plano (subespacio) para ver la nube de puntos.





DATOS

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

COMPONENTES

	Comp1	Comp2	Comp3	Comp4	Comp5
Lucia	0,3231	1,7725	1,1988	-0,055	0,0036
Pedro	0,6654	-1,6387	0,1455	-0,0231	-0,1234
Ines	1,0025	-0,5157	0,6289	0,5164	0,1429
Luis	-3,1721	-0,2628	-0,382	0,6778	-0,0625
Andres	-0,4889	1,3654	-0,8352	-0,1558	0,1234
Ana	1,7086	-1,0217	-0,1271	0,0668	0,0253
Carlos	0,0676	1,4623	-0,5062	-0,1179	0,0131
José	2,0119	-1,2759	-0,5422	-0,1978	0,0174
Sonia	-3,042	-1,2549	0,4488	-0,64	0,0379
Maria	0,9239	1,3694	-0,0293	-0,0715	-0,1777

DATOS

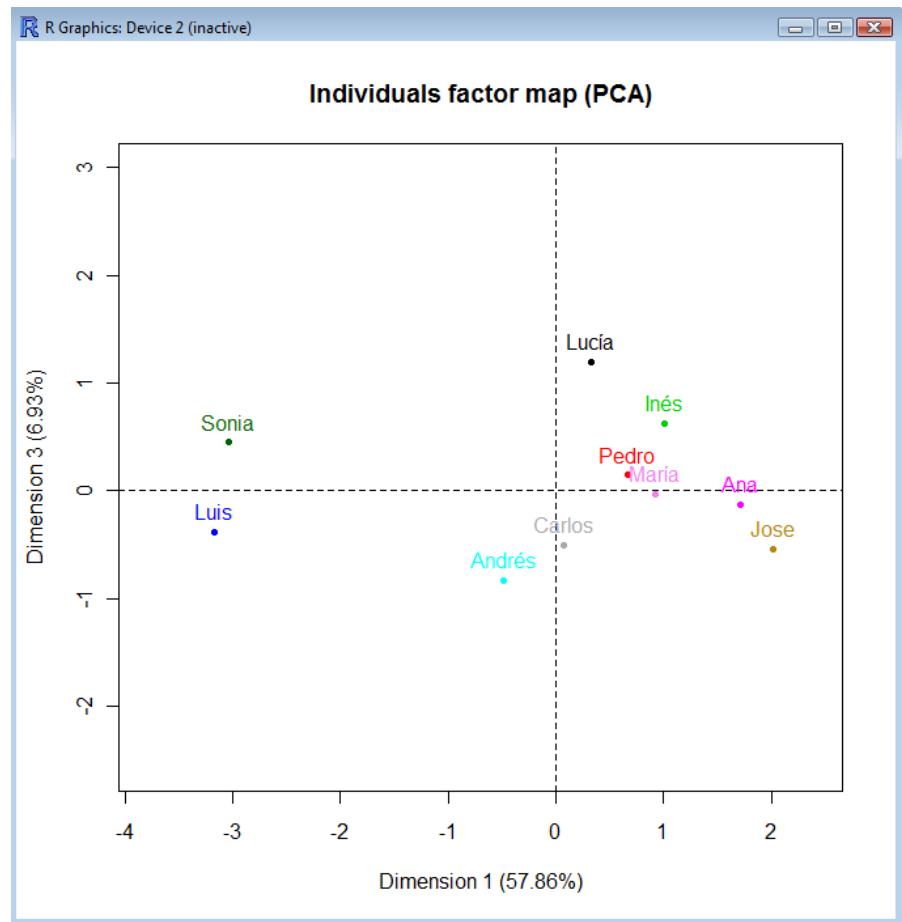
	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

COMPONENTES

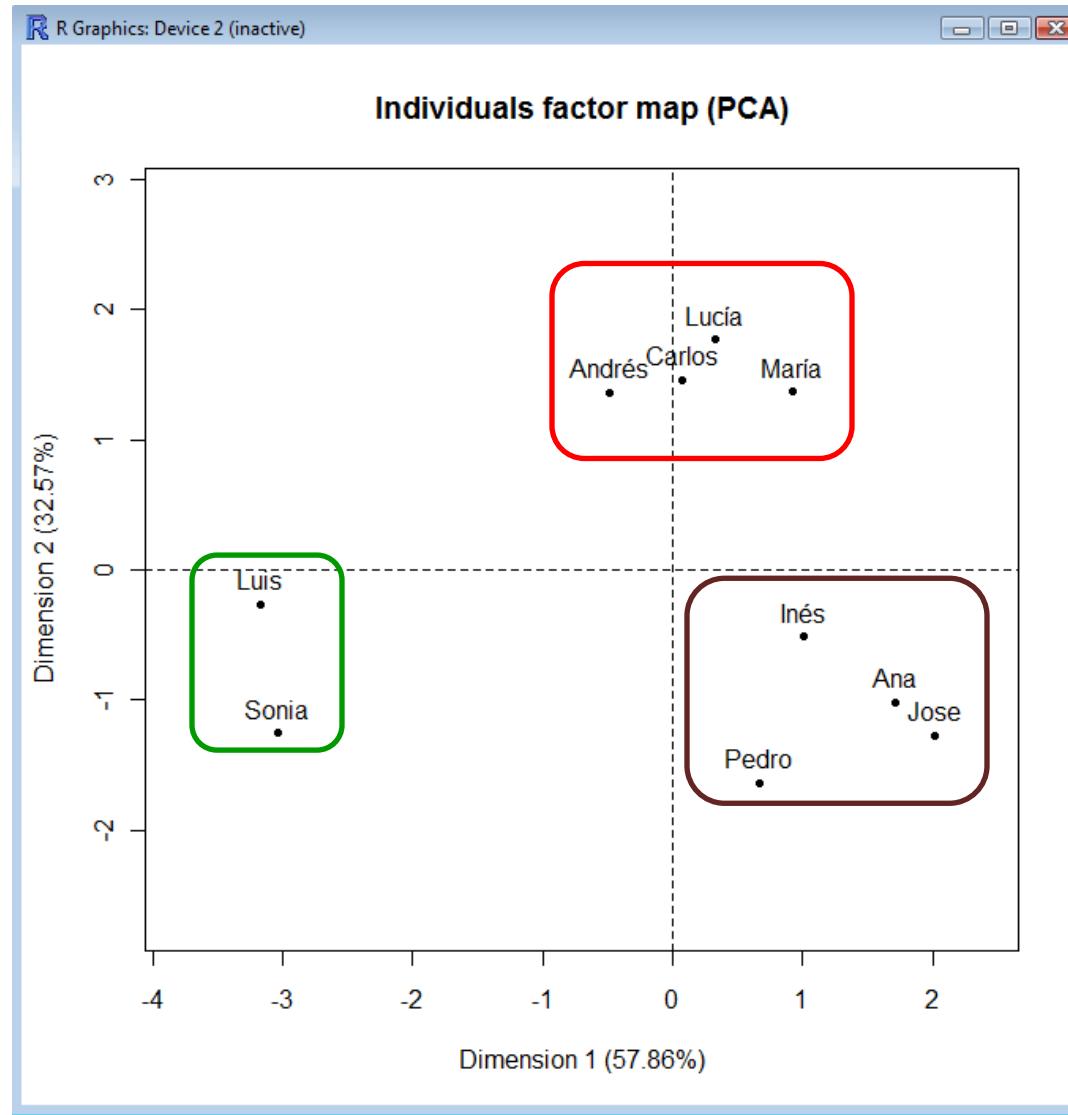
	Comp1	Comp2	Comp3	Comp4	Comp5
Lucia	0,3231		1,1988		
Pedro	0,6654		0,1455		
Ines	1,0025		0,6289		
Luis	-3,1721		-0,382		
Andres	-0,4889		-0,8352		
Ana	1,7086		-0,1271		
Carlos	0,0676		-0,5062		
Jose	2,0119		-0,5422		
Sonia	-3,042		0,4488		
Maria	0,9239		-0,0293		

COMPONENTES

	Comp1		Comp3	
Lucia	0,3231		1,1988	
Pedro	0,6654		0,1455	
Ines	1,0025		0,6289	
Luis	-3,1721		-0,382	
Andres	-0,4889		-0,8352	
Ana	1,7086		-0,1271	
Carlos	0,0676		-0,5062	
Jose	2,0119		-0,5422	
Sonia	-3,042		0,4488	
Maria	0,9239		-0,0293	



Análisis de clústeres o conglomerados



1. Se identifican los clústeres de individuos. Como se puede observar, se marcaron los 3 clústeres o conglomerados.
2. La siguiente tabla presenta la composición de los clústeres mencionados:

Cluster	Individuos
1	{Luis, Sonia}
2	{Ana, Inés, José, Pedro}
3	{Andrés, Carlos, Lucía, María}

Variables en las columnas

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

Nubes de Puntos

INDIVIDUOS - FILAS

Luis	5.0	6.5	6.5	7.0	9.0
------	-----	-----	-----	-----	-----

$$\in \mathbb{R}^5$$

VARIABLES - COLUMNAS

Español
9.2
7.3
8.0
6.5
7.8
7.7
8.2
7.5
6.5
8.7

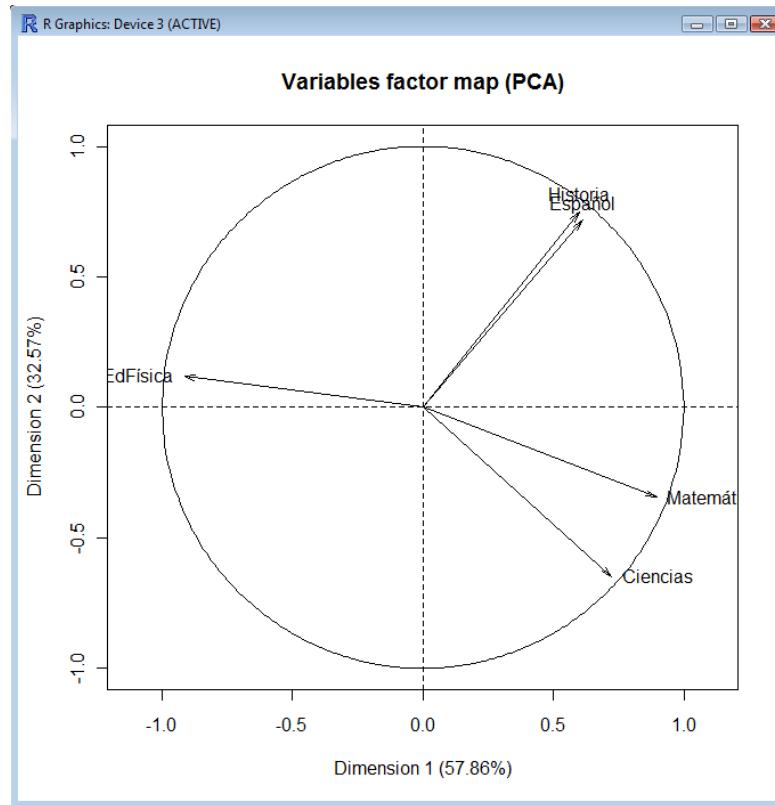
$$\in \mathbb{R}^{10}$$

	Matemáticas	Ciencias	Español	Historia	EdFísica
Lucía	7.0	6.5	9.2	8.6	8.0
Pedro	7.5	9.4	7.3	7.0	7.0
Inés	7.6	9.2	8.0	8.0	7.5
Luis	5.0	6.5	6.5	7.0	9.0
Andrés	6.0	6.0	7.8	8.9	7.3
Ana	7.8	9.6	7.7	8.0	6.5
Carlos	6.3	6.4	8.2	9.0	7.2
José	7.9	9.7	7.5	8.0	6.0
Sonia	6.0	6.0	6.5	5.5	8.7
María	6.8	7.2	8.7	9.0	7.0

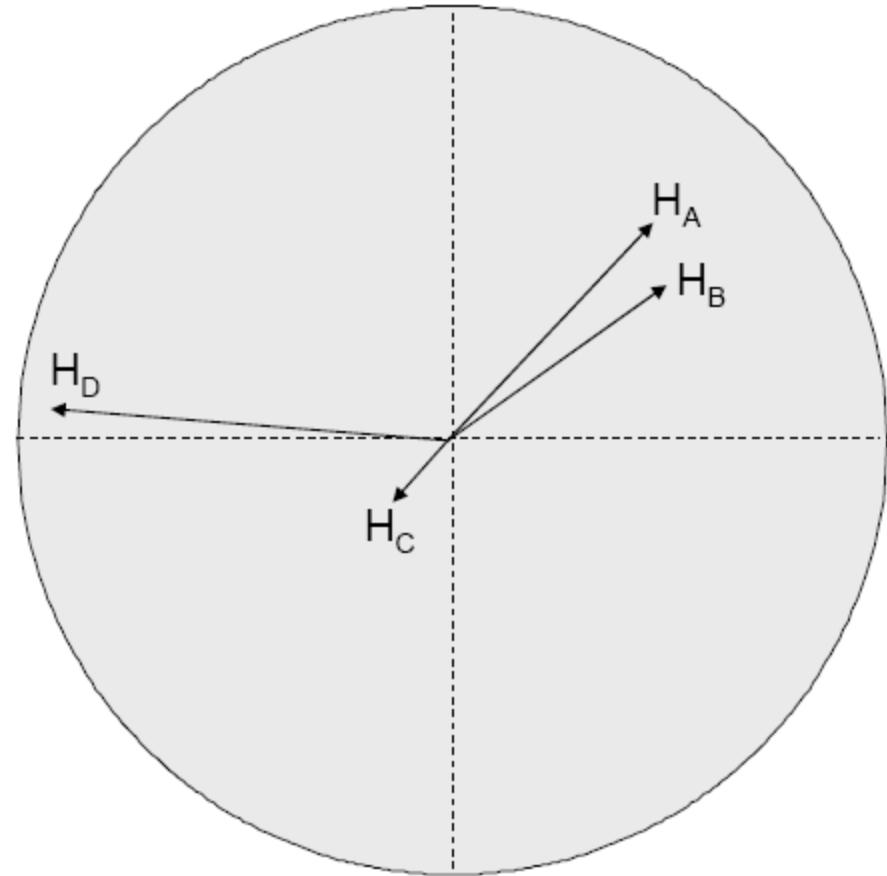
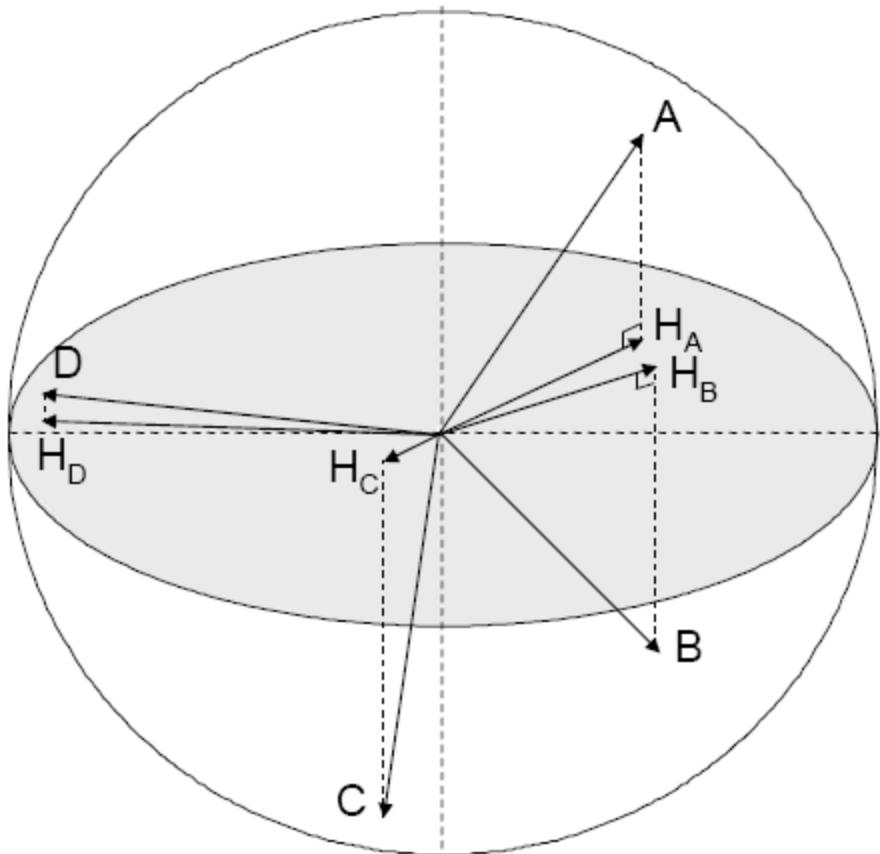
	Comp1	Comp2	Comp3	Comp4	Comp5
Matemáticas	0.8957980	-0.3452036	0.25797931	-0.09146818	0.05882803
Ciencias	0.7227976	-0.6483946	0.02384033	0.23587773	-0.03068234
Español	0.6108931	0.7173206	0.33102532	-0.02454152	-0.04561456
Historia	0.5999227	0.7484701	-0.23206345	0.15639747	0.03964443
EdFísica	-0.9139265	0.1196373	0.34065108	0.18315368	0.02892890

	Matemáticas	Ciencias	Español	Historia	EdFísica	
	Lucía	7.0	6.5	9.2	8.6	8.0
	Pedro	7.5	9.4	7.3	7.0	7.0
	Inés	7.6	9.2	8.0	8.0	7.5
	Luis	5.0	6.5	6.5	7.0	9.0
	Andrés	6.0	6.0	7.8	8.9	7.3
	Ana	7.8	9.6	7.7	8.0	6.5
	Carlos	6.3	6.4	8.2	9.0	7.2
	José	7.9	9.7	7.5	8.0	6.0
	Sonia	6.0	6.0	6.5	5.5	8.7
	Maria	6.8	7.2	8.7	9.0	7.0

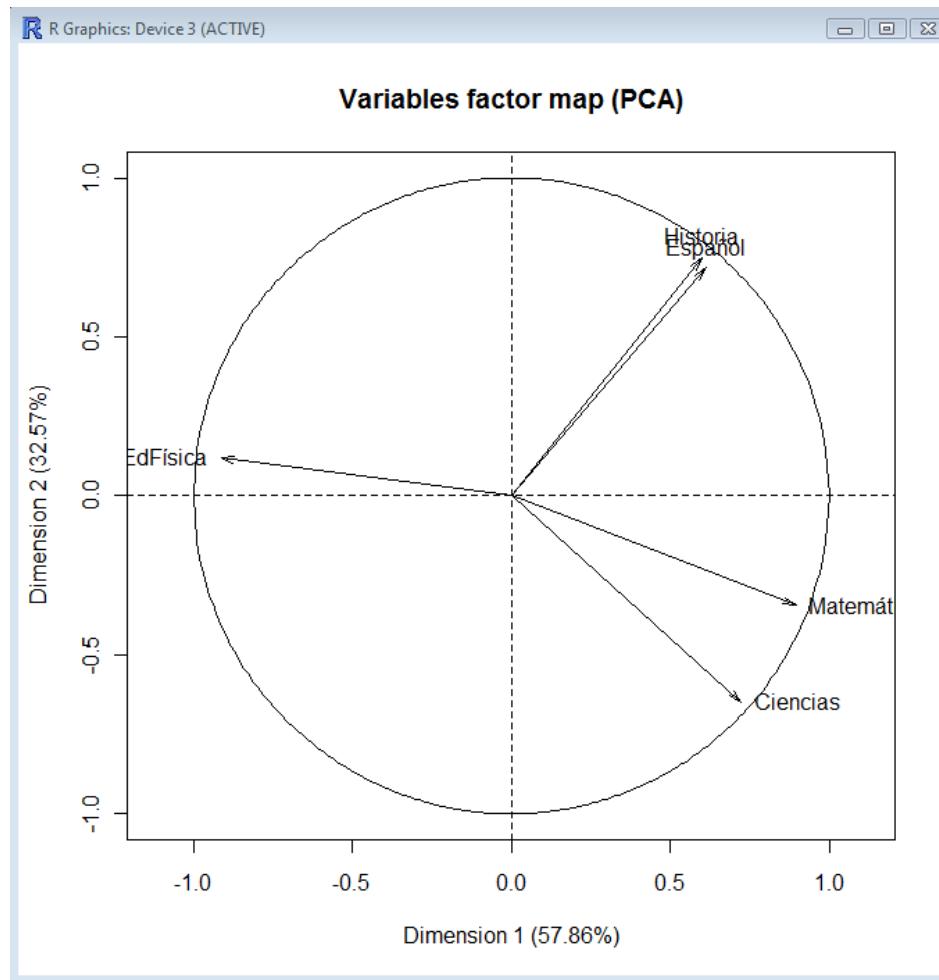
	Comp1	Comp2			
Matemáticas	0.8957980	-0.3452036			
Ciencias	0.7227976	-0.6483946			
Español	0.6108931	0.7173206			
Historia	0.5999227	0.7484701			
EdFísica	-0.9139265	0.1196373			



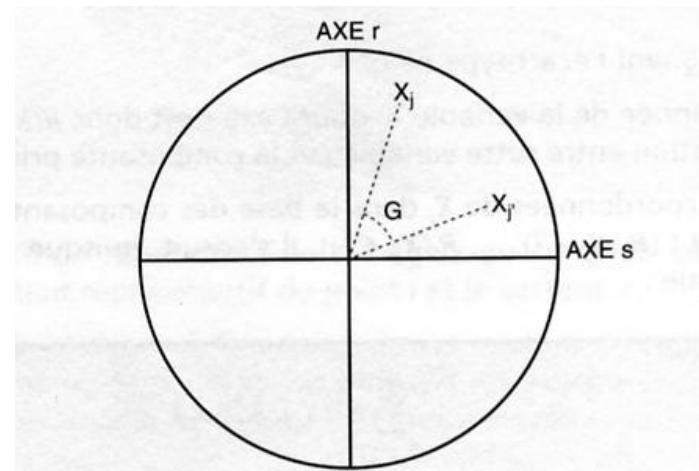
En el espacio de las variables



Círculo de Correlaciones



INTERPRETACIÓN



- Si X^j y $X^{j'}$ están cercanas entre si, entonces X^j y $X^{j'}$ son fuerte y positivamente correlacionadas.
- Si el ángulo entre X^j y $X^{j'}$ es cercano a los 90° entonces NO existe ninguna correlación entre ambas variables.
- Si X^j y $X^{j'}$ están opuestas al vértice (origen) entonces existe una correlación fuerte y negativa entre X^j y $X^{j'}$.

Ejemplo: Estudiantes

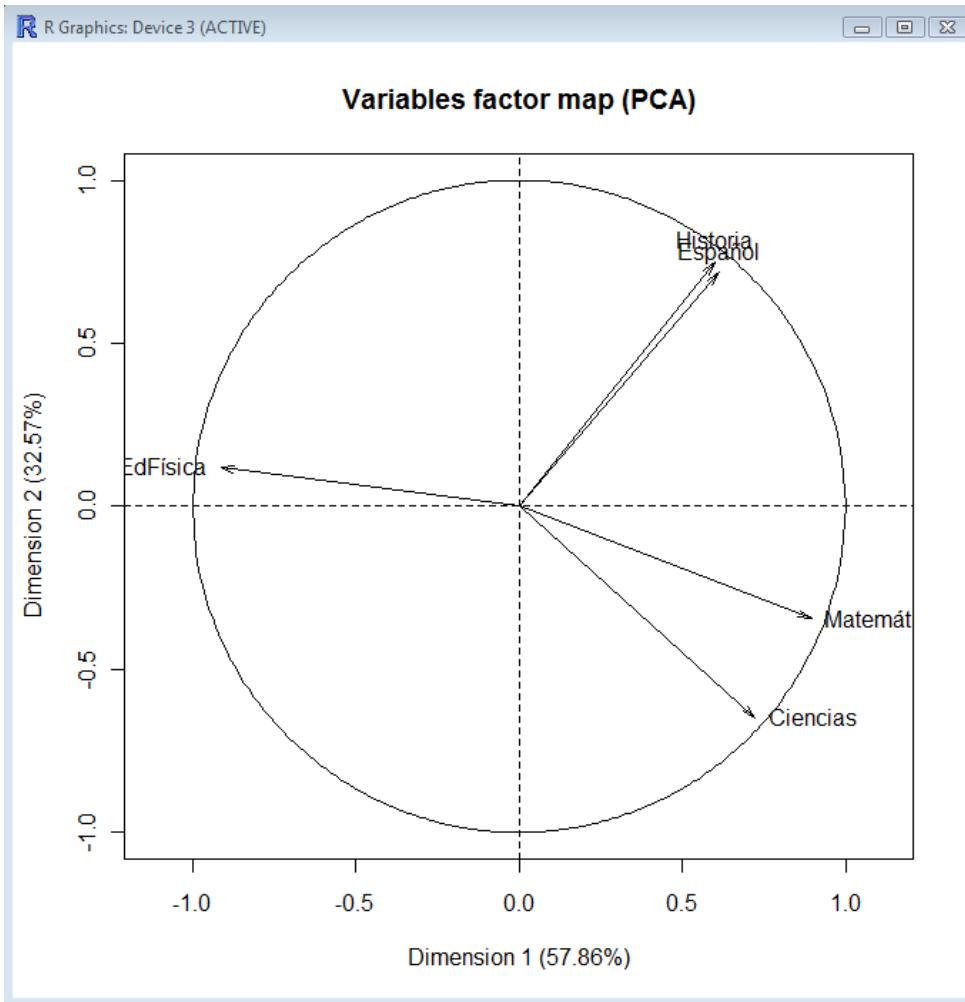


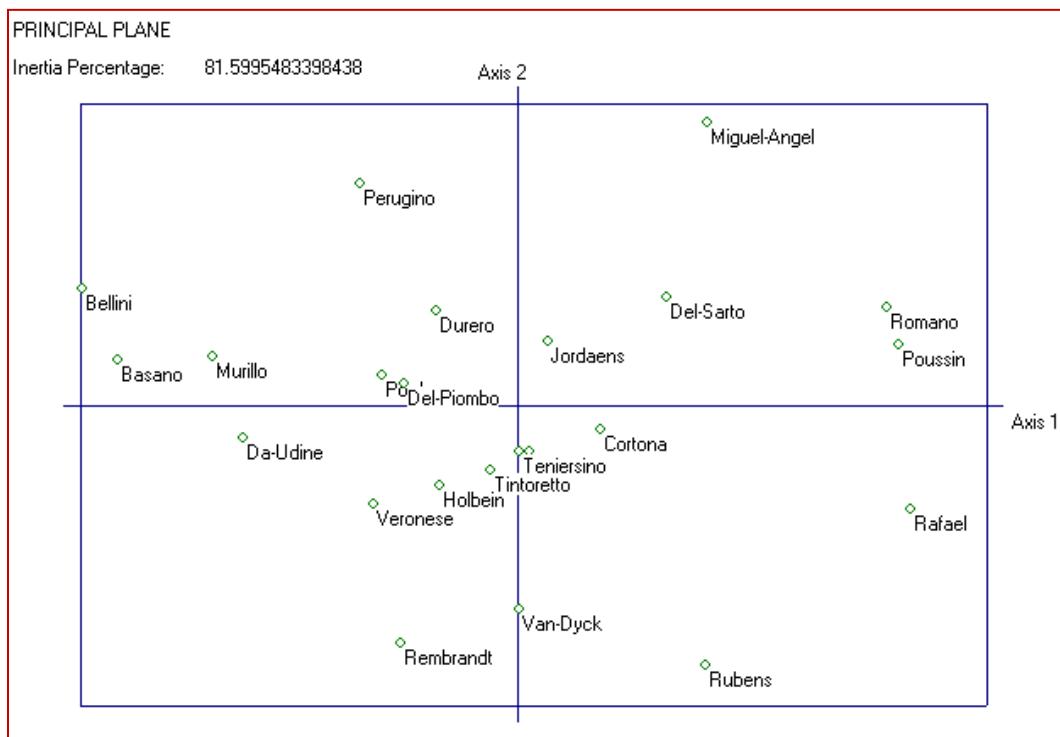
Tabla con las correlaciones presentes en el círculo

Variable	Matemática	Ciencias	Español	Historia	Ed. Física
Matemática	-	Positive	None	None	Negative
Ciencias		-	None	None	Negative
Español			-	Positive	None
Historia				-	None
Ed. Física					-

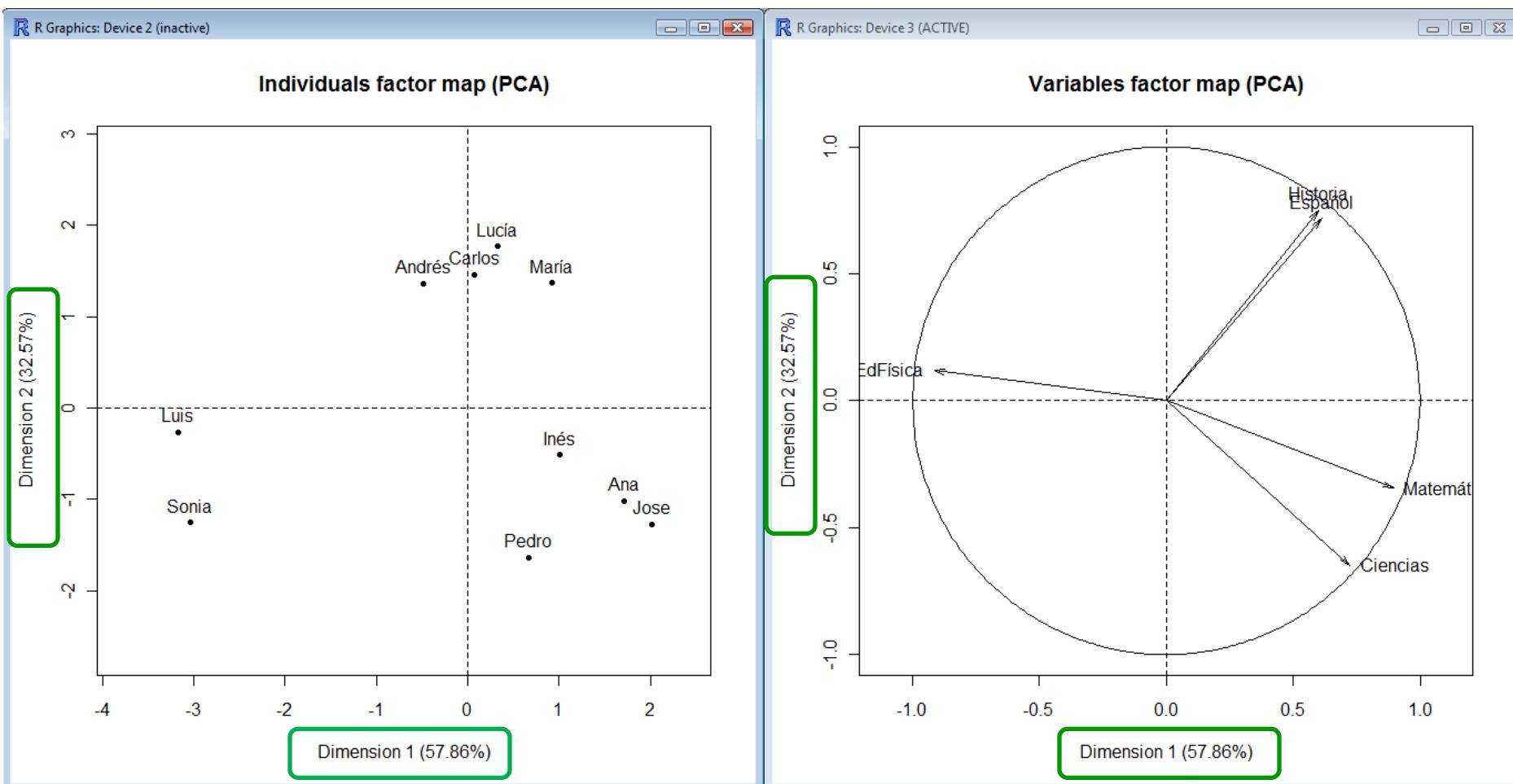
Calidad de los Gráficos

- Por ejemplo, la **inercia** explicada por el plano principal, ejes 1 y 2 es:

$$\frac{\beta_1 + \beta_2}{m}.$$

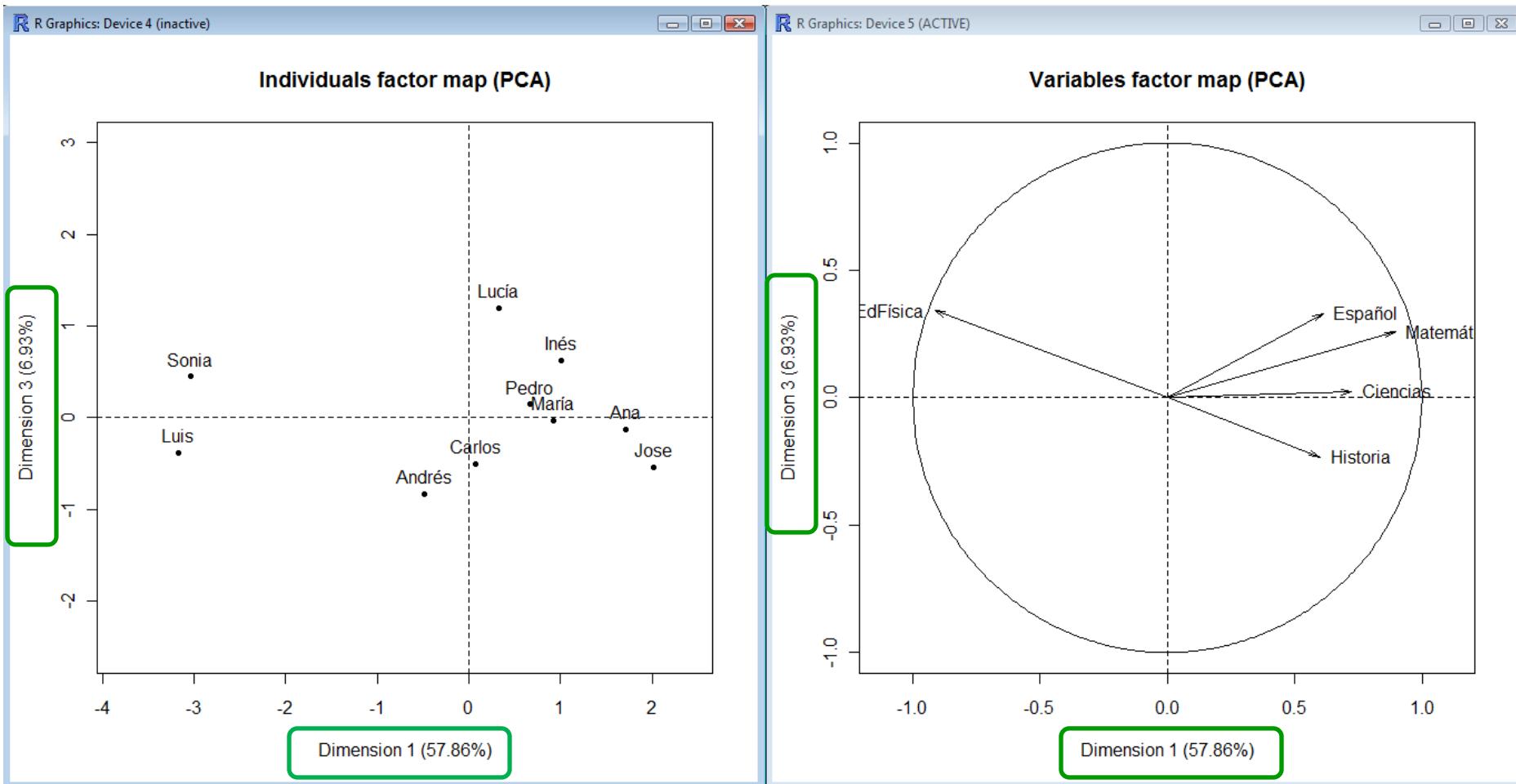


Inercia Explicada = 90.43%



$$\text{Inercia} = 57.86 + 32.57 = 90.43$$

Inercia Explicada = 64.79%



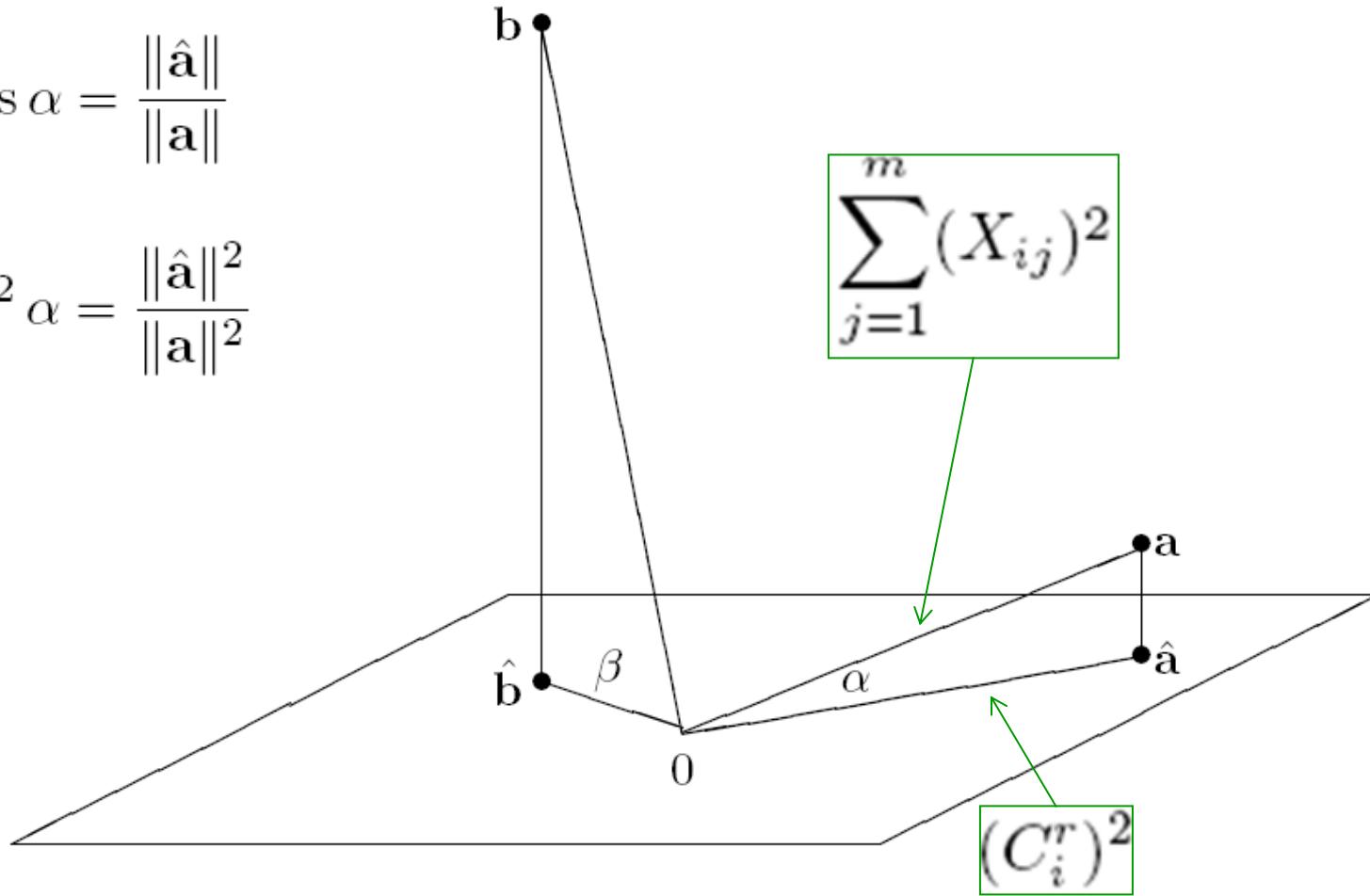
$$\text{Inercia} = 57.86 + 6.93 = 64.79$$

Calidad de la representación de individuos y variables

1. Solamente los INDIVIDUOS y VARIABLES **BIEN** representados en el plano principal y en círculo de correlaciones, respectivamente, pueden ser interpretados.
2. ¿Cómo saber que un INDIVIDUO o VARIABLE está **BIEN** representado?

$$\cos \alpha = \frac{\|\hat{\mathbf{a}}\|}{\|\mathbf{a}\|}$$

$$\cos^2 \alpha = \frac{\|\hat{\mathbf{a}}\|^2}{\|\mathbf{a}\|^2}$$



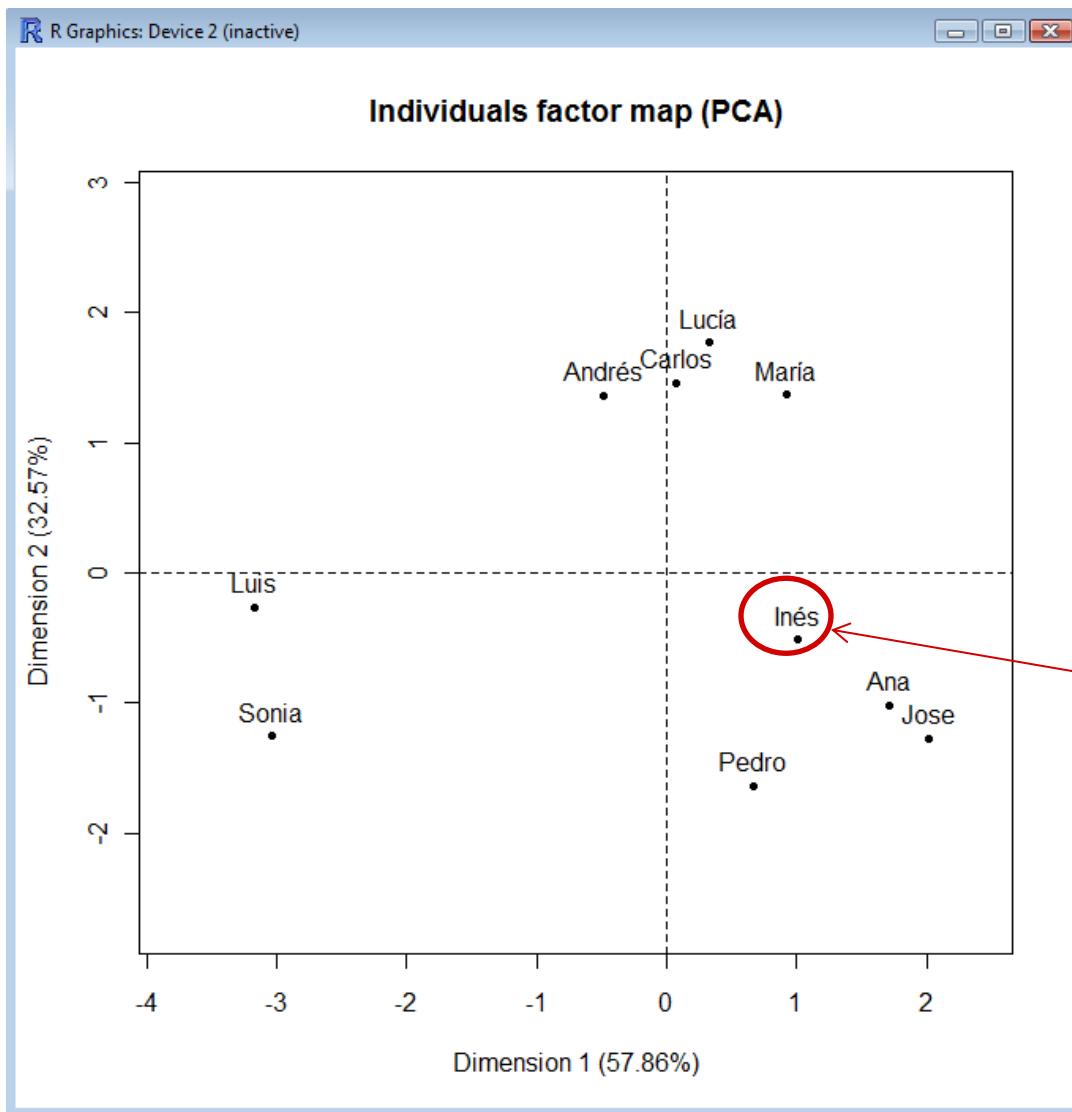
NOTA: Si el $\cos^2 \alpha$ es “cercano” a 1 la representación del individuo será muy buena o si multiplicado por 100 es mayor a 60 la representación es buena.

Ejemplo $\text{Cos}^2(\alpha)$: Estudiantes

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Luía	0.022270827	0.670420670	0.3066598387	0.0006458478	2,82E+00
Pedro	0.139905502	0.848430539	0.0066865270	0.0001680781	4,81E+03
Inés	0.514468899	0.136122895	0.2024397137	0.1365196756	1,04E+04
Luis	0.936851990	0.006429392	0.0135836048	0.0427712757	3,64E+02
Andés	0.084139511	0.656353715	0.2456037030	0.0085448999	5,36E+03
Ana	0.732686110	0.261979570	0.0040527949	0.0011209894	1,61E+02
Carlos	0.001892733	0.886081139	0.1061921889	0.0057625700	7,14E+01
Jose	0.673612108	0.270910359	0.0489165039	0.0065104446	5,06E+01
Sonia	0.808829929	0.137636943	0.0176072367	0.0358004434	1,25E+02
María	0.308554271	0.677869212	0.0003109770	0.0018464085	1,14E+04

Ejemplo $\text{Cos}^2(\alpha)$: Estudiantes

	Dim.1	Dim.2	SUMA*100
Luía	0,022270827	0,67042067	69,27
Pedro	0,139905502	0,848430539	98,83
Inés	0,514468899	0,136122895	65,06
Luis	0,93685199	0,006429392	94,33
Andrés	0,084139511	0,656353715	74,05
Ana	0,73268611	0,26197957	99,47
Carlos	0,001892733	0,886081139	88,80
Jose	0,673612108	0,270910359	94,45
Sonia	0,808829929	0,137636943	94,65
María	0,308554271	0,677869212	98,64



Todos los individuos están bien representados en este plano.

Inés es la persona que tiene la menos buena representación en este plano.

Calidad de representación de cada variable

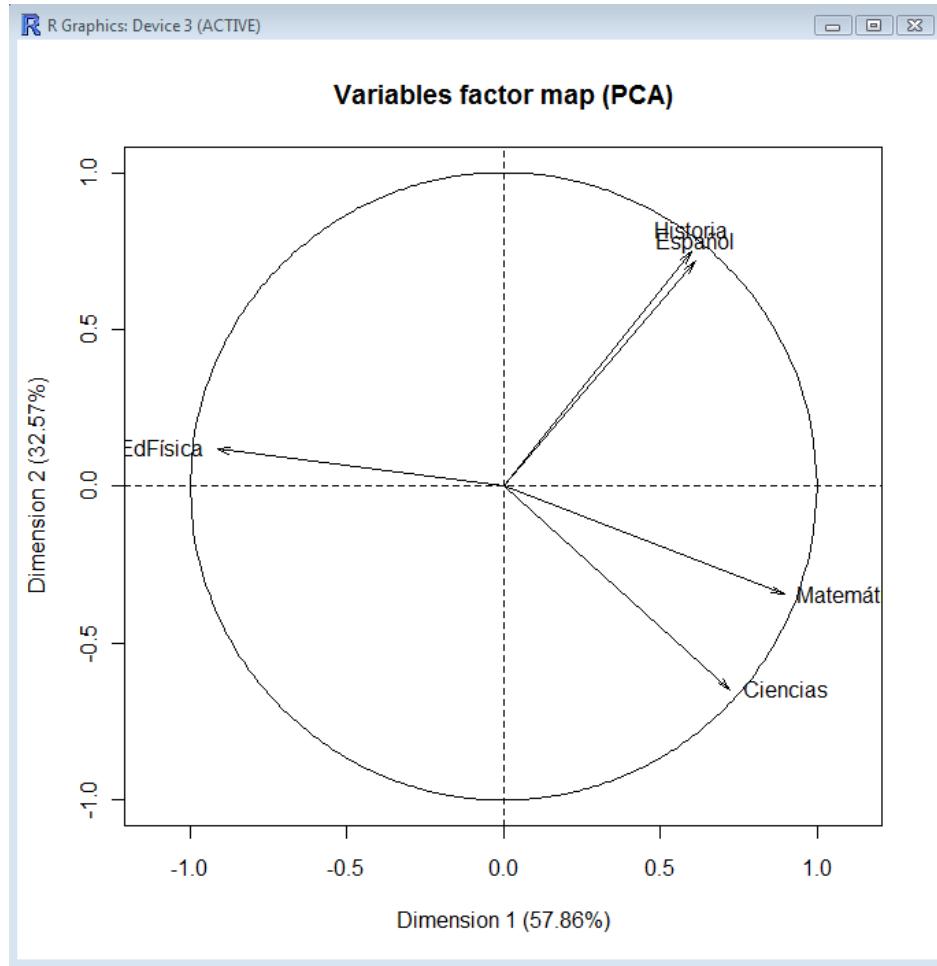
- ✓ La calidad de la representación de una variable sobre el círculo de correlaciones, será también medida con el **coseno al cuadrado** del ángulo la variable y su proyección. Ahora bien, recuérdese que entre variables, el coseno es igual a una correlación, por lo que serán las **correlaciones cercanas a 1** las que impliquen la calidad de la representación de las variables.
- ✓ Es decir estarán bien representadas aquellas variables que queden ubicadas cerca de la frontera o borde del círculo de correlaciones

Ejemplo $\text{Cos}^2(\alpha)$: Estudiantes

	Dim1	Dim2	Dim3	Dim4	Dim5
Matemáticas	0,802454	0,1191655	0,066553324	0,008366429	0,003460737
Ciencias	0,5224364	0,42041555	0,000568361	0,055638305	0,000941406
Español	0,3731904	0,51454884	0,109577763	0,000602286	0,002080688
Historia	0,3599073	0,56020745	0,053853443	0,02446017	0,001571681
EdFísica	0,8352616	0,01431309	0,116043157	0,03354527	0,000836881

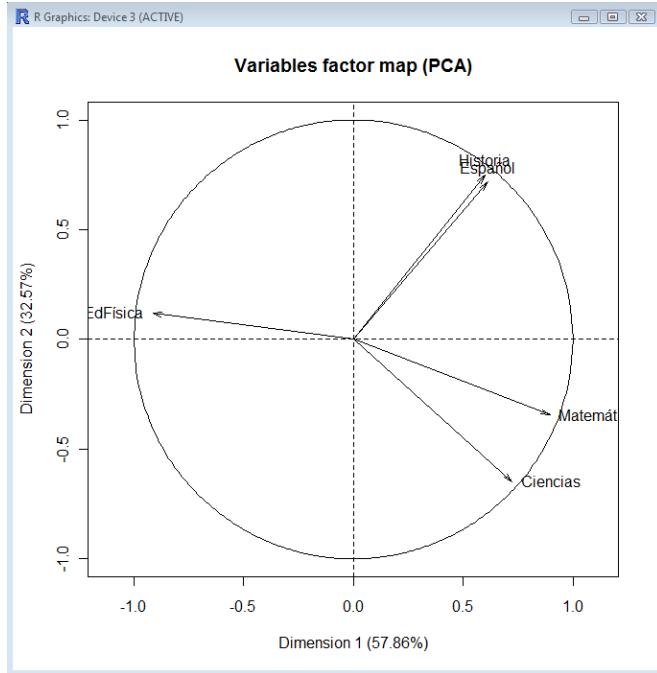
	Dim1	Dim2	Suma
Matemáticas	0,802454	0,1191655	0,9216195
Ciencias	0,5224364	0,42041555	0,94285195
Español	0,3731904	0,51454884	0,88773924
Historia	0,3599073	0,56020745	0,92011475
EdFísica	0,8352616	0,01431309	0,84957469

Ejemplo: Estudiantes



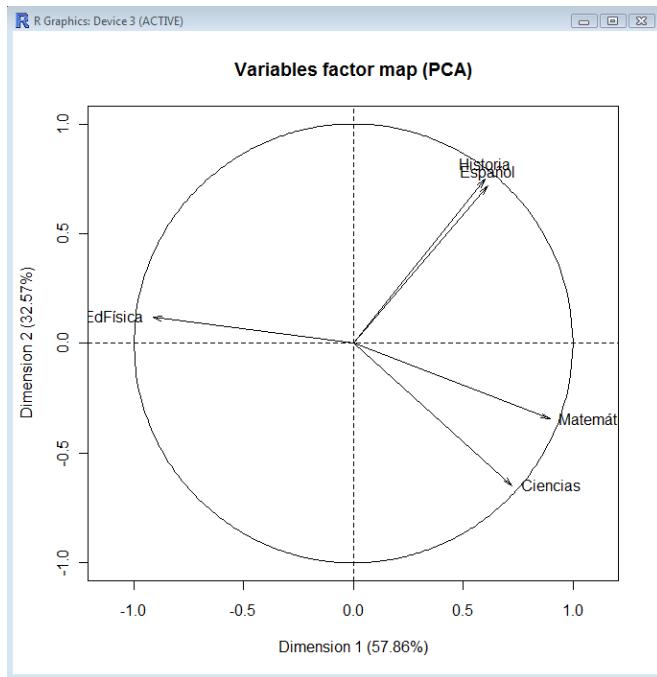
Todas las variables están bien representadas en este círculo de correlaciones

Ejemplo: Estudiantes



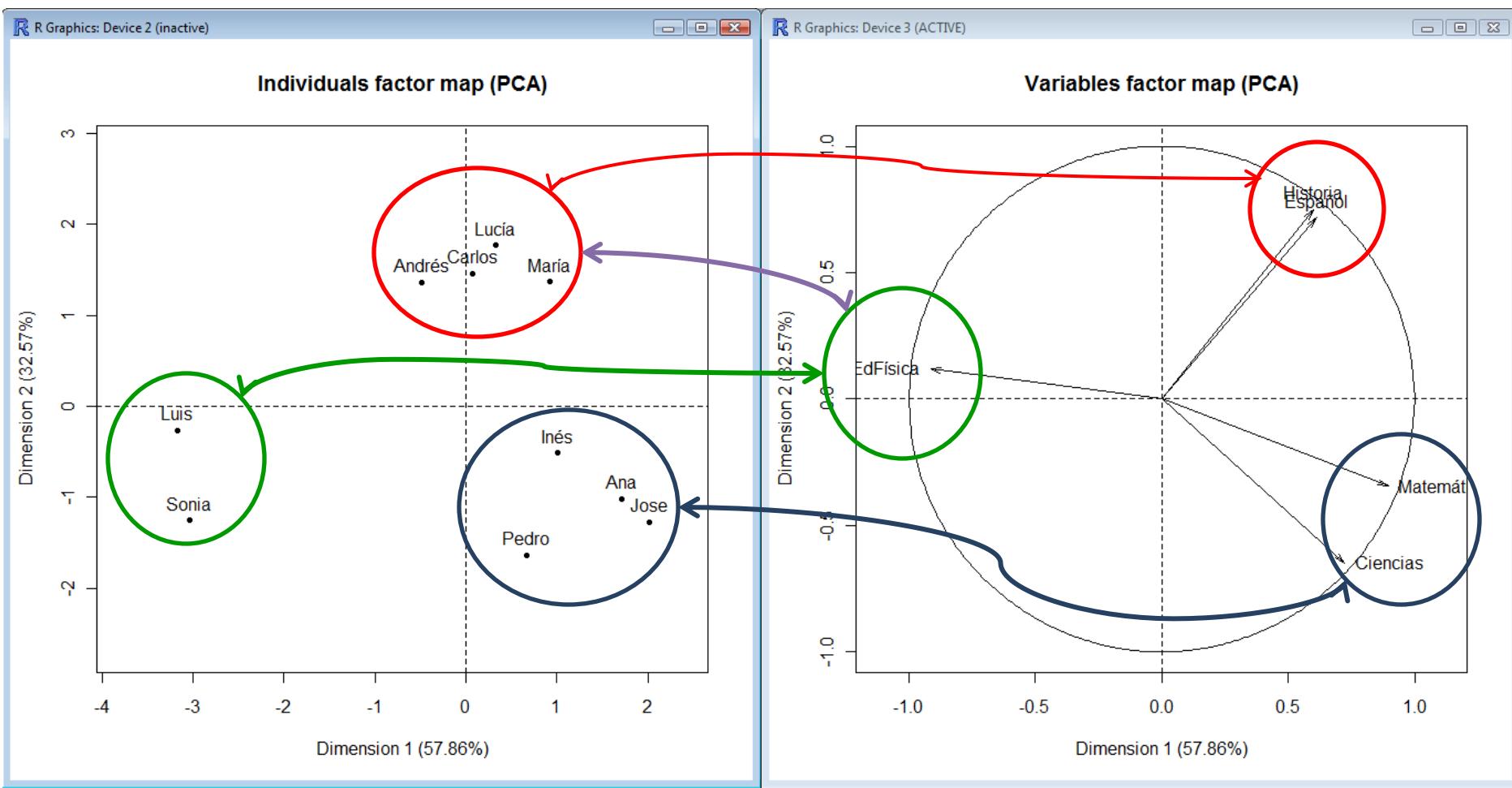
Un ángulo de apenas $1,71^\circ$ representa una clara correlación **positiva** entre las variables Español e Historia, es decir, que las notas en estas materias se comportan de manera similar en los estudiantes.

Ejemplo: Estudiantes



Ángulos cercanos a 180° , como este caso de $166,38^\circ$ entre Matemática y Ed. Física denotan correlaciones **negativas**. Entonces al crecer variable Matemáticas, la variable Educación Física va a disminuir y viceversa.

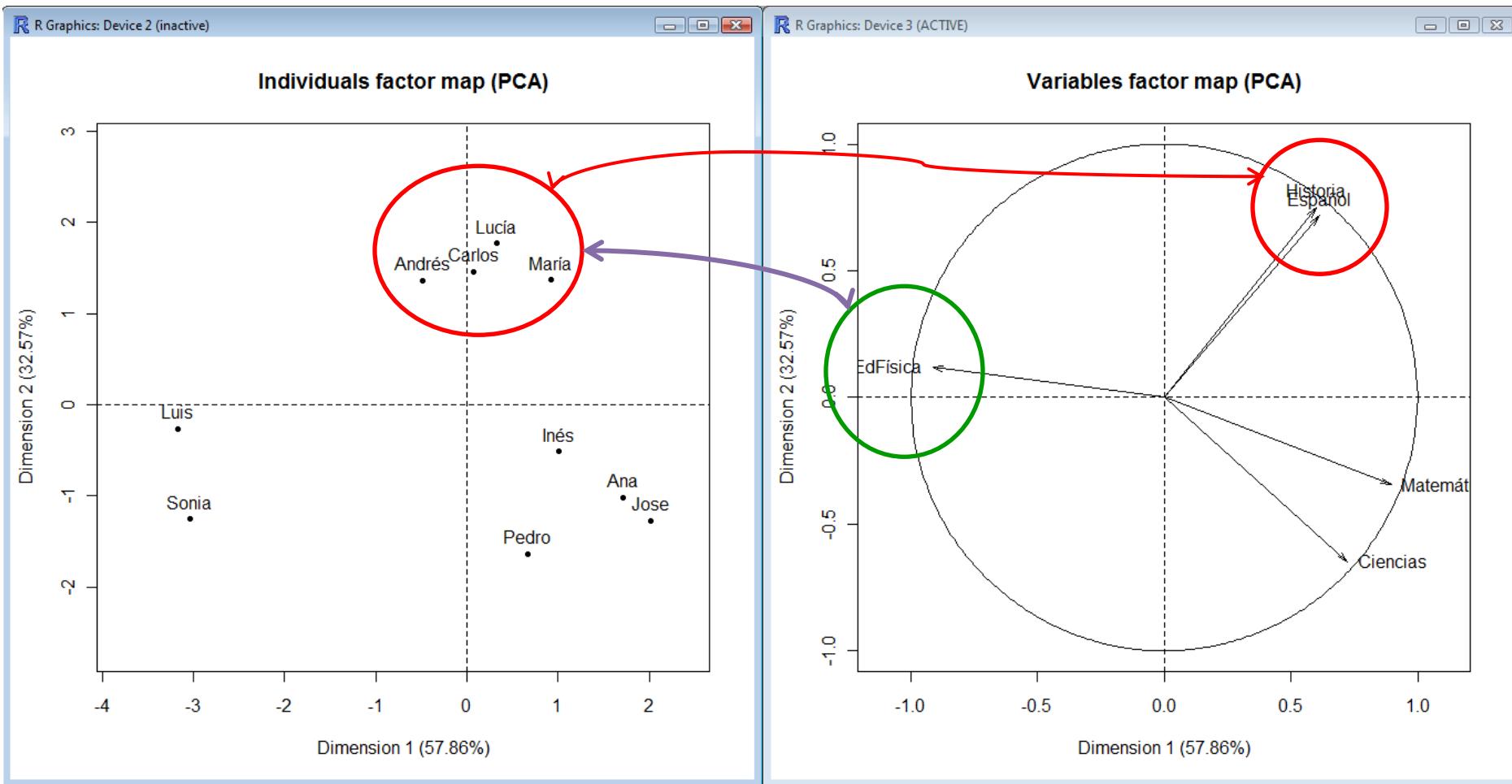
Sobreposición de Gráficos



Sobreposición de Gráficos

- El clúster 1 (Luis y Sonia), se ve fuertemente impactado –de manera positiva– por las notas de educación física, es decir, son buenos deportistas. Recordando las correlaciones entre variables, podemos afirmar que son malos en ciencias y matemáticas, pues son variables inversamente correlacionadas con respecto a educación física. También se puede apreciar que son malos en español e historia, pues se oponen negativamente a estas variables en el plano.
- El clúster 2 (Ana, Inés, José y Pedro) parece ser el opuesto del clúster 1. Sus integrantes se destacan en las ciencias y las matemáticas, sin embargo son malos en deportes. Con respecto a las materias de español e historia, por estar de forma perpendicular o correlación nula, no es correcto asumir nada sobre su comportamiento.
- El clúster 3 (Andrés, Carlos, Lucía y María) se caracteriza por agrupar a los estudiantes destacados en el área de español e historia. Se puede inferir también que si bien no son excelentes en deportes, al menos no son los peores, pues se encuentran en la mitad del eje X.

Influencia de dos o más variables

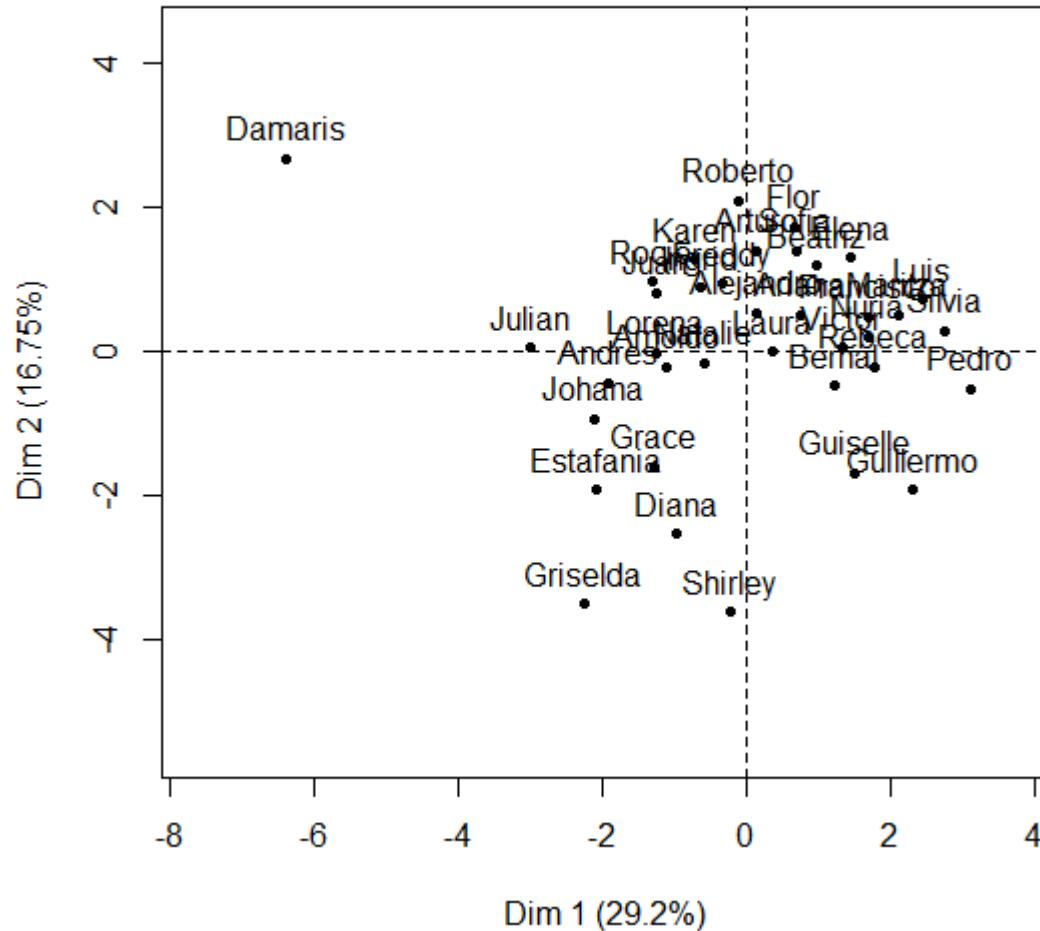


Ejemplo: Servicio al Cliente

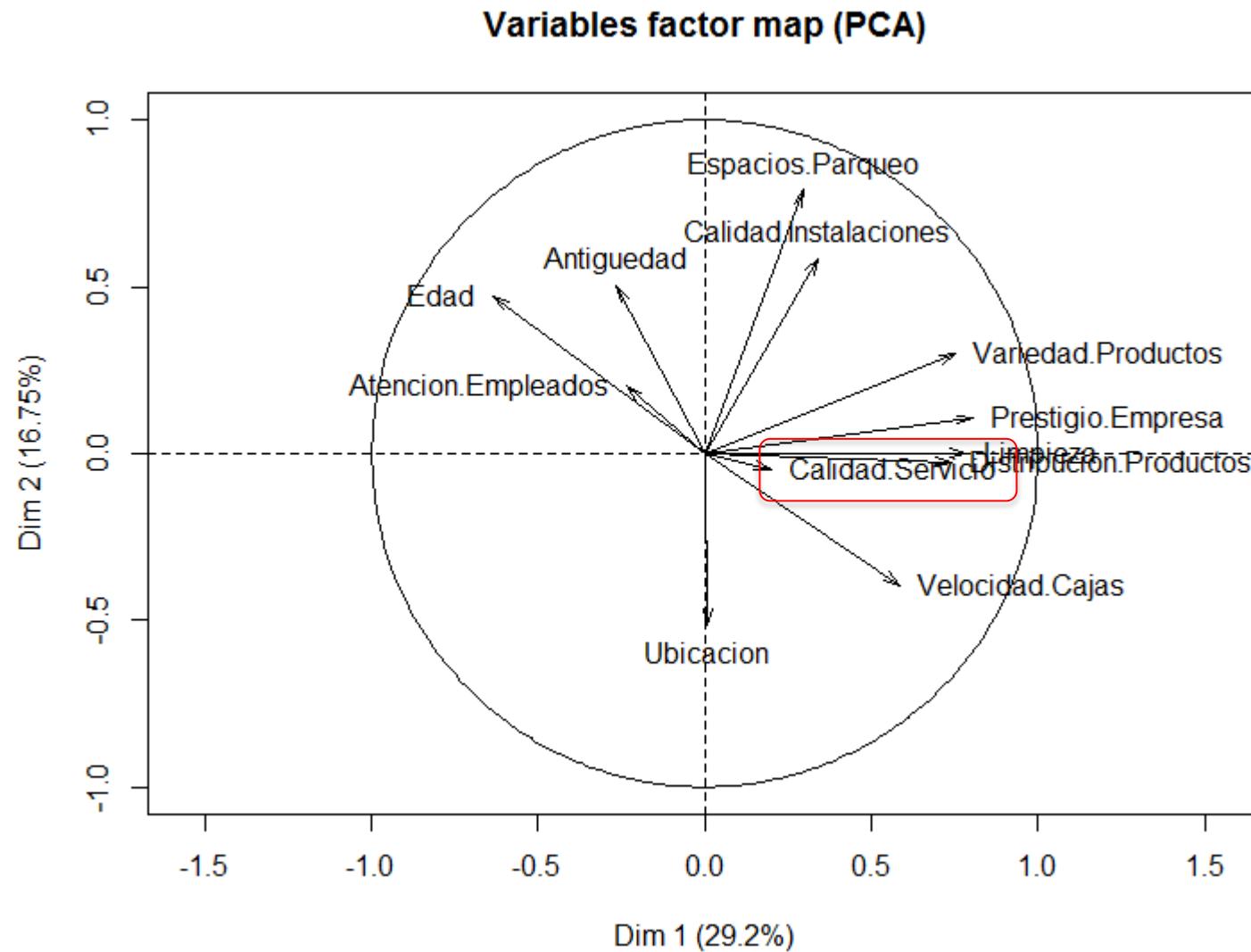
Nombre Cliente	Edad	Antiguedad	Espacios Parqueo	Velocidad Cajas	Distribucion Productos	Atencion Empleados	Calidad Instalaciones	Ubicacion	Limpieza	Variedad Productos	Prestigio Empresa	Calidad Servicio
Ariana	25	01	7,6	7,6	7,8	9,7	5,0	9,0	7,6	5,6	7,0	6,6
Guiselle	24	00	4,8	9,0	7,2	10,0	2,0	9,6	6,8	8,4	9,8	5,4
Francisco	28	07	6,8	8,4	7,6	8,7	2,7	9,2	6,2	9,0	9,6	8,5
Griselda	23	00	3,4	7,8	9,0	10,0	1,0	10,0	4,4	4,0	2,8	5,4
Damaris	49	06	7,0	3,2	1,2	10,0	4,0	9,0	1,4	4,8	2,6	3,3
Johana	32	04	5,6	7,8	6,8	10,0	3,0	10,0	5,0	4,2	4,2	7,2
Bernal	26	00	6,2	8,0	6,6	9,3	3,3	8,6	7,8	6,4	9,6	6,5
Freddy	23	04	5,6	6,8	6,2	9,7	4,0	6,8	6,8	7,4	5,6	4,5
Estafania	25	04	4,6	8,0	3,8	10,0	1,7	9,8	5,0	4,4	6,0	7,6
Laura	29	00	5,4	6,4	8,8	9,7	6,7	10,0	5,6	6,2	8,4	6,5
Arnoldo	32	13	4,2	7,8	7,6	10,0	1,3	9,6	7,4	6,4	7,0	3,9
Beatriz	27	05	8,0	9,4	5,6	8,7	4,0	7,4	7,2	6,6	7,6	5,4
Rebeca	22	00	6,8	9,0	9,0	10,0	5,0	9,4	6,0	7,0	9,4	4,6
Sofia	31	05	7,2	7,6	7,6	10,0	5,3	9,6	7,6	7,0	8,8	2,4
Ingrid	27	05	6,2	7,8	5,0	10,0	5,0	9,0	4,2	7,6	7,4	4,0
Rocio	42	04	6,2	9,0	6,2	10,0	3,0	7,4	6,2	6,2	5,2	5,7
Karen	43	05	6,6	7,8	5,6	9,7	3,7	8,8	7,2	6,2	8,0	5,4
Luis	24	04	7,8	7,2	9,6	9,0	3,0	9,4	8,0	9,4	8,0	8,8
Pedro	21	01	6,2	9,6	8,4	10,0	5,3	10,0	9,0	8,4	9,6	6,0
Lorena	30	04	6,8	7,8	4,0	9,3	2,7	9,2	6,8	6,2	4,6	4,5
Elena	31	05	7,0	7,4	8,0	9,3	3,3	8,6	7,8	9,2	8,4	5,6
Julian	40	05	5,0	7,8	3,4	9,7	3,3	9,2	4,2	6,0	5,0	1,5
Natalie	34	01	6,4	5,6	3,4	10,0	0,7	10,0	8,4	7,2	9,6	4,6
Shirley	24	00	3,8	9,6	6,0	7,7	1,0	9,4	6,4	6,0	5,2	3,0
Andres	38	03	6,0	7,8	3,6	9,3	2,3	9,6	4,8	5,8	6,4	6,9
Alejandro	26	06	6,2	8,6	8,0	9,3	5,0	8,6	5,6	6,6	6,6	3,2
Grace	28	00	6,0	7,8	3,0	10,0	0,7	10,0	7,0	6,2	5,8	5,5
Nuria	21	00	6,8	8,0	8,6	9,3	3,7	7,8	7,8	8,0	7,4	2,3
Flor	24	01	7,6	7,2	7,4	10,0	6,0	8,0	6,2	8,0	6,6	2,5
Roberto	41	08	6,8	8,8	6,4	10,0	5,0	8,6	8,2	7,4	6,6	3,2
Victor	27	03	7,6	8,2	9,4	9,7	2,0	9,4	7,8	7,0	8,2	5,2
Arturo	34	05	7,2	7,8	8,6	9,7	3,0	7,4	6,8	5,8	8,2	6,5
Maritza	24	04	6,8	8,2	7,2	10,0	3,7	9,4	9,6	9,0	8,2	5,8
Diana	24	01	3,4	8,2	3,8	10,0	0,7	9,2	7,6	6,6	6,4	2,6
Juan	23	04	6,0	6,6	4,0	10,0	3,0	6,8	6,2	6,8	4,6	8,1
Guillermo	19	01	4,8	9,6	9,2	9,0	4,0	9,6	7,6	7,2	8,8	4,8
Silvia	25	02	6,0	9,0	8,2	10,0	6,0	9,6	9,2	9,2	8,6	5,5

Plano principal

Individuals factor map (PCA)

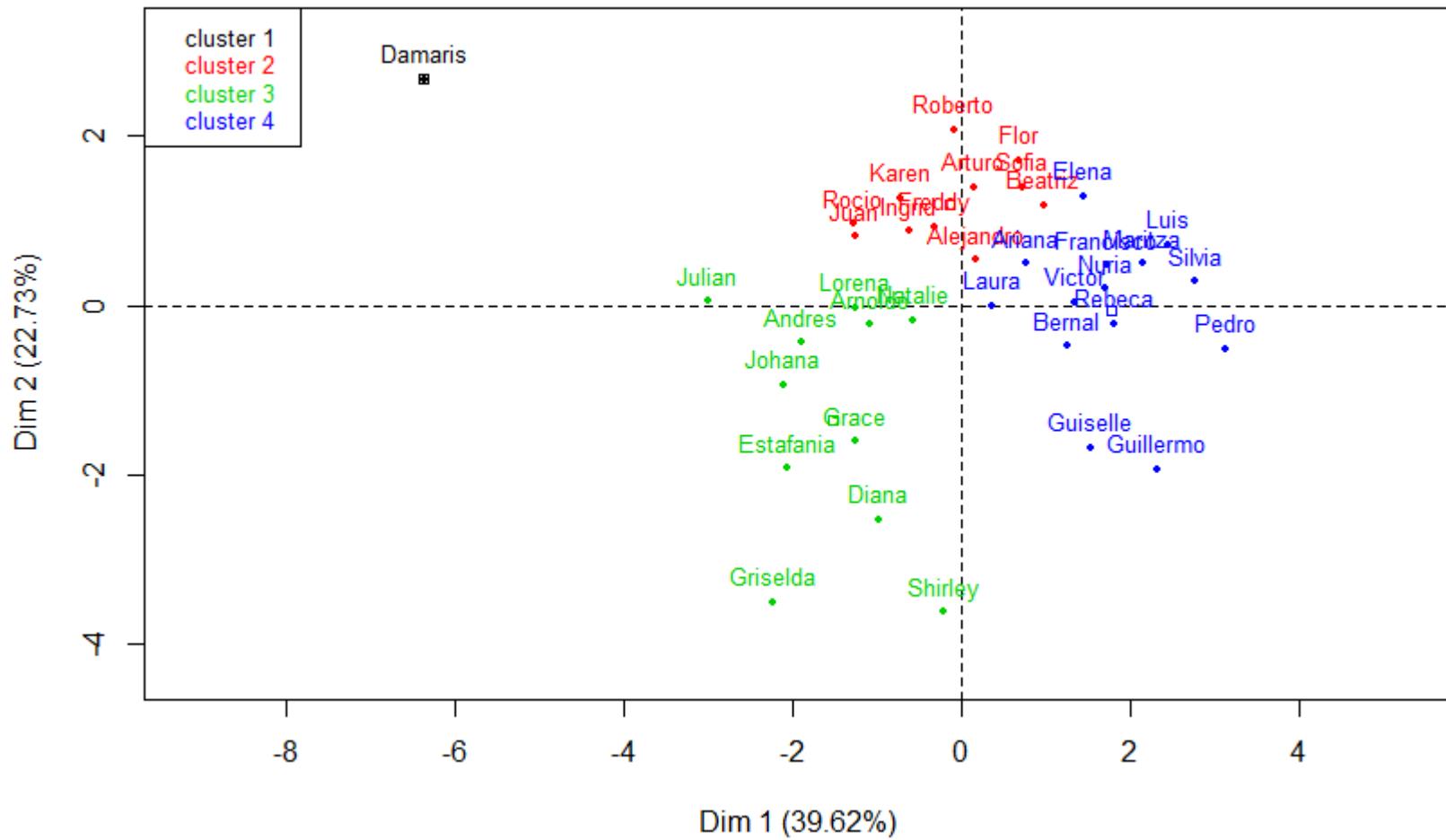


Círculo de Correlaciones



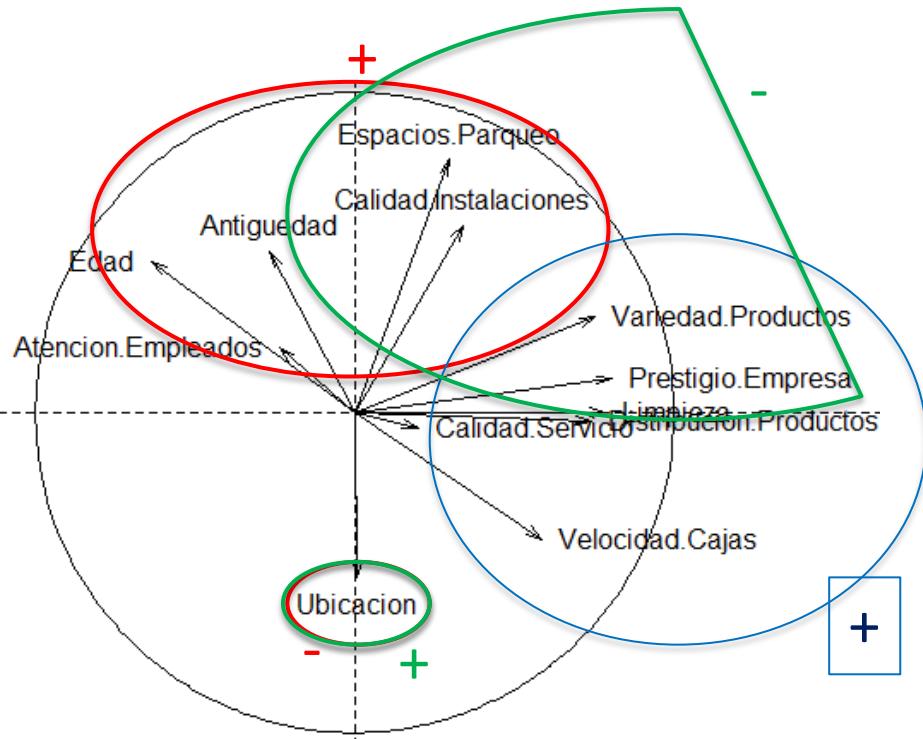
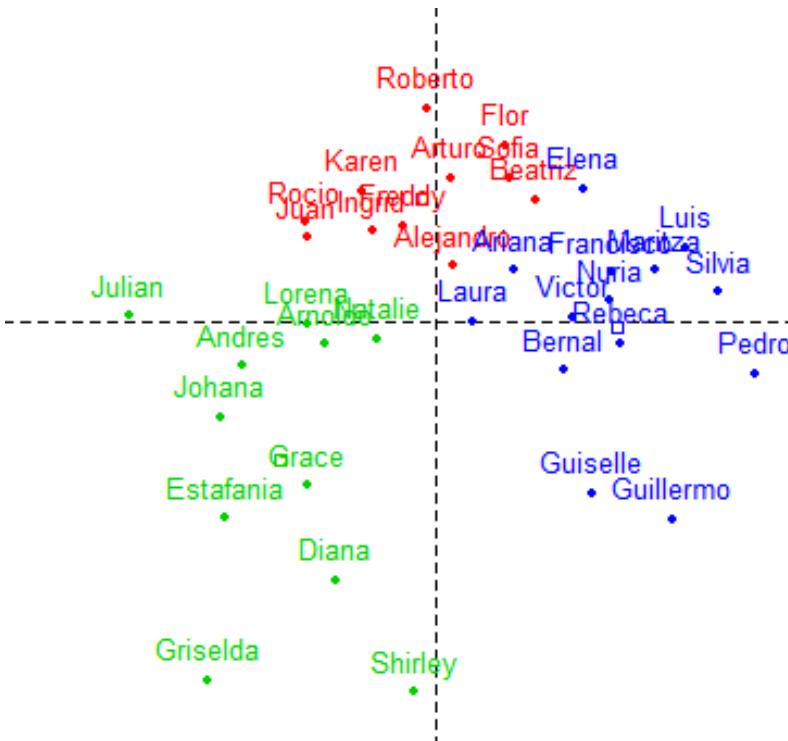
Análisis de los conglomerados

Factor map



Nombre Cliente	Edad	Antiguedad	Espacios Parqueo	Velocidad Cajas	Distribucion Productos	Atencion Empleados	Calidad Instalaciones	Ubicacion	Limpieza	Variedad Productos	Prestigio Empresa	Calidad Servicio
Ariana	25	01	7,6	7,6	7,8	9,7	5,0	9,0	7,6	5,6	7,0	6,6
Guiselle	24	00	4,8	9,0	7,2	10,0	2,0	9,6	6,8	8,4	9,8	5,4
Francisco	28	07	6,8	8,4	7,6	8,7	2,7	9,2	6,2	9,0	9,6	8,5
Griselda	23	00	3,4	7,8	9,0	10,0	1,0	10,0	4,4	4,0	2,8	5,4
Damaris	49	06	7,0	3,2	1,2	10,0	4,0	9,0	1,4	4,8	2,6	3,3
Johana	32	04	5,6	7,8	6,8	10,0	3,0	10,0	5,0	4,2	4,2	7,2
Bernal	26	00	6,2	8,0	6,6	9,3	3,3	8,6	7,8	6,4	9,6	6,5
Freddy	23	04	5,6	6,8	6,2	9,7	4,0	6,8	6,8	7,4	5,6	4,5
Estafania	25	04	4,6	8,0	3,8	10,0	1,7	9,8	5,0	4,4	6,0	7,6
Laura	29	00	5,4	6,4	8,8	9,7	6,7	10,0	5,6	6,2	8,4	6,5
Arnoldo	32	13	4,2	7,8	7,6	10,0	1,3	9,6	7,4	6,4	7,0	3,9
Beatriz	27	05	8,0	9,4	5,6	8,7	4,0	7,4	7,2	6,6	7,6	5,4
Rebeca	22	00	6,8	9,0	9,0	10,0	5,0	9,4	6,0	7,0	9,4	4,6
Sofia	31	05	7,2	7,6	7,6	10,0	5,3	9,6	7,6	7,0	8,8	2,4
Ingrid	27	05	6,2	7,8	5,0	10,0	5,0	9,0	4,2	7,6	7,4	4,0
Rocio	42	04	6,2	9,0	6,2	10,0	3,0	7,4	6,2	6,2	5,2	5,7
Karen	43	05	6,6	7,8	5,6	9,7	3,7	8,8	7,2	6,2	8,0	5,4
Luis	24	04	7,8	7,2	9,6	9,0	3,0	9,4	8,0	9,4	8,0	8,8
Pedro	21	01	6,2	9,6	8,4	10,0	5,3	10,0	9,0	8,4	9,6	6,0
Lorena	30	04	6,8	7,8	4,0	9,3	2,7	9,2	6,8	6,2	4,6	4,5
Elena	31	05	7,0	7,4	8,0	9,3	3,3	8,6	7,8	9,2	8,4	5,6
Julian	40	05	5,0	7,8	3,4	9,7	3,3	9,2	4,2	6,0	5,0	1,5
Natalie	34	01	6,4	5,6	3,4	10,0	0,7	10,0	8,4	7,2	9,6	4,6
Shirley	24	00	3,8	9,6	6,0	7,7	1,0	9,4	6,4	6,0	5,2	3,0
Andres	38	03	6,0	7,8	3,6	9,3	2,3	9,6	4,8	5,8	6,4	6,9
Alejandro	26	06	6,2	8,6	8,0	9,3	5,0	8,6	5,6	6,6	6,6	3,2
Grace	28	00	6,0	7,8	3,0	10,0	0,7	10,0	7,0	6,2	5,8	5,5
Nuria	21	00	6,8	8,0	8,6	9,3	3,7	7,8	7,8	8,0	7,4	2,3
Flor	24	01	7,6	7,2	7,4	10,0	6,0	8,0	6,2	8,0	6,6	2,5
Roberto	41	08	6,8	8,8	6,4	10,0	5,0	8,6	8,2	7,4	6,6	3,2
Victor	27	03	7,6	8,2	9,4	9,7	2,0	9,4	7,8	7,0	8,2	5,2
Arturo	34	05	7,2	7,8	8,6	9,7	3,0	7,4	6,8	5,8	8,2	6,5
Maritza	24	04	6,8	8,2	7,2	10,0	3,7	9,4	9,6	9,0	8,2	5,8
Diana	24	01	3,4	8,2	3,8	10,0	0,7	9,2	7,6	6,6	6,4	2,6
Juan	23	04	6,0	6,6	4,0	10,0	3,0	6,8	6,2	6,8	4,6	8,1
Guillermo	19	01	4,8	9,6	9,2	9,0	4,0	9,6	7,6	7,2	8,8	4,8
Silvia	25	02	6,0	9,0	8,2	10,0	6,0	9,6	9,2	9,2	8,6	5,5

Interpretación de los Clústeres



En Python

ACP con el paquete “prince”

En Mac (Terminal):

```
pip install git+https://github.com/MaxHalford/Prince
```

En Windows (Anaconda Prompt):

```
pip install git+https://github.com/MaxHalford/Prince
```

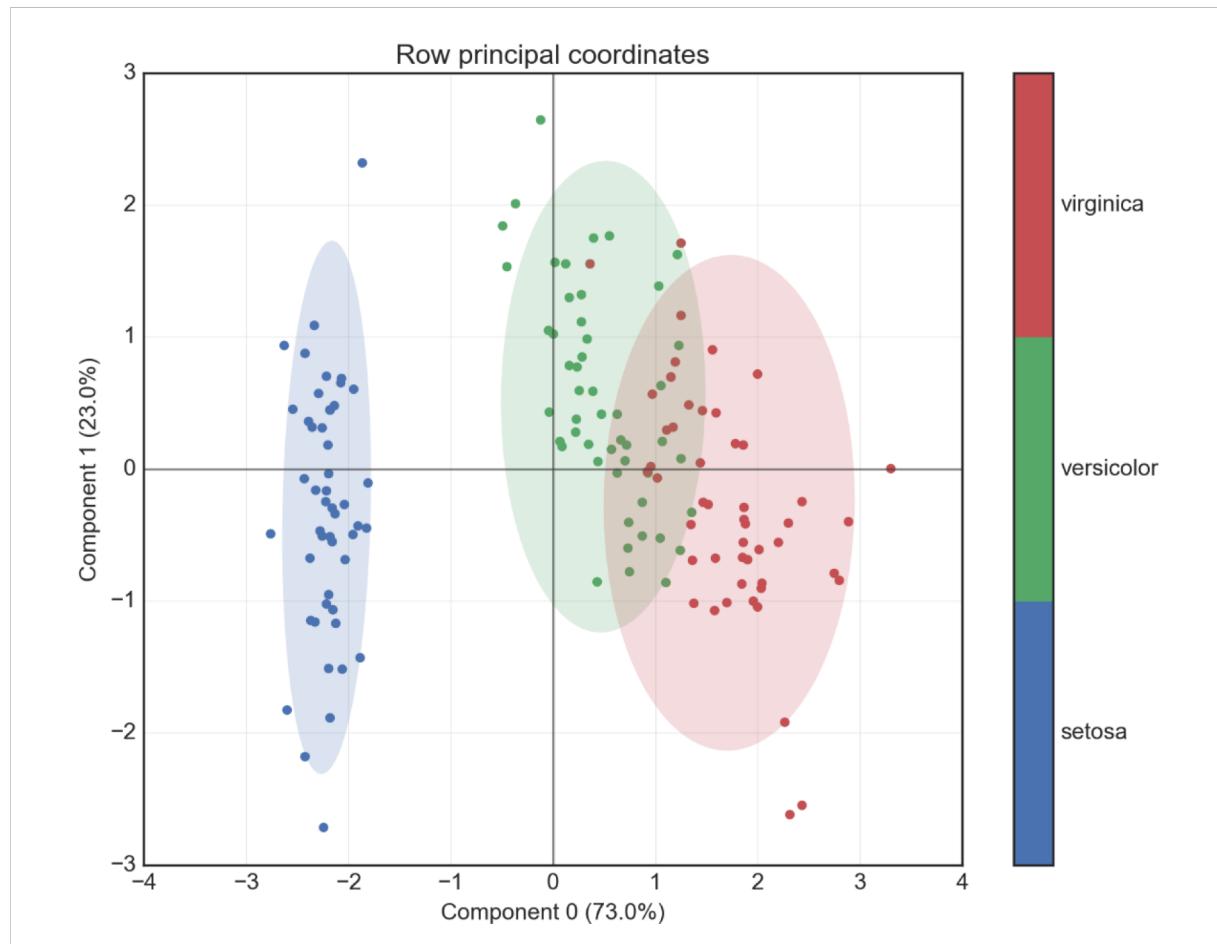
Prince en github: + <https://github.com/MaxHalford/prince>

Instalando paquetes

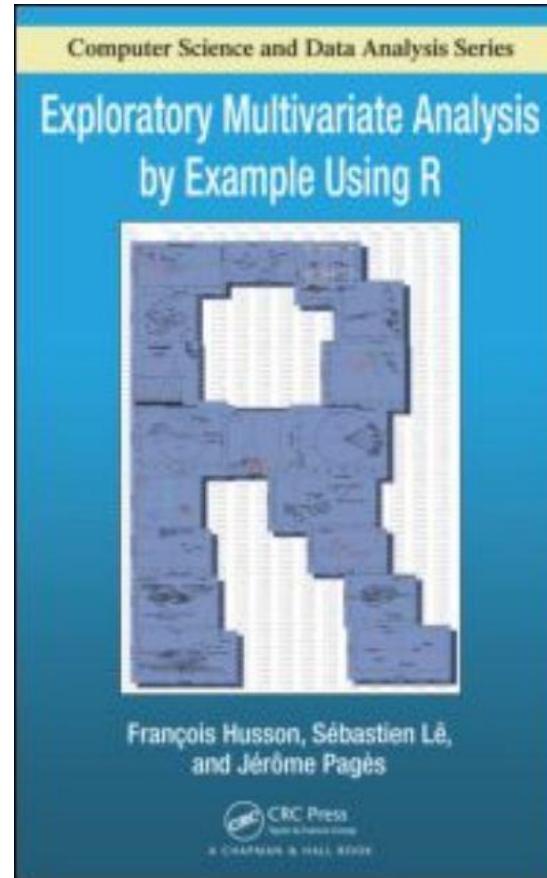
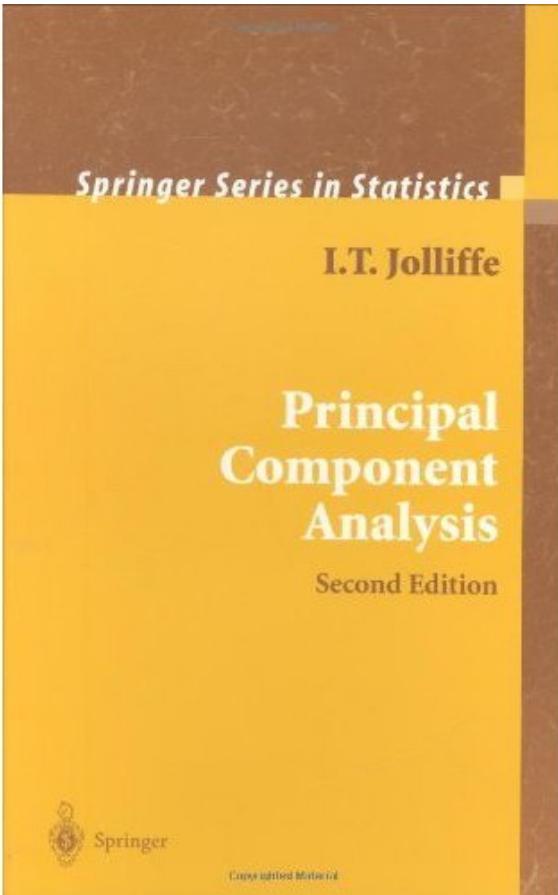
<https://docs.python.org/3/installing/>

```
import prince
```

En Python



¿Dónde obtener más información?





oldemar rodíguez

CONSULTOR en M1NER14 D4T0S

Gracias ...