

# Laboratorio 4 Metodos de Agrupamiento

Grupo 6

## Integrantes

- Edwin Sanchez
- Stephanie Tamayo
- Andres Felipe Torres
- Fredy Urrea
- Sergio Velasquez
- Manuel Espitia

## Introduccion

Carfa de librerias y datos

```
library("FactoMineR")
library("dplyr")
library("kableExtra")
library(readxl)
library(FactoMineR)
library(factoextra)
library(kableExtra)
library(readr)
library(knitr)
library(tidyr)
library(dplyr)
library(tibble)

datos <- read_csv2("r14_Sci_Qs_Webometrics.csv")
#head(datos)
#str(datos)
#ecc<- read_csv2("ECC_completa_19426.csv")
```

Utilizar el archivo r\_14\_Sci\_Qs\_Webometrics.csv que se encuentrea en la carpeta conjunto de datos utilizados en los ejemplos y ejercicios para contruir una tabla con los promedios por país y llvaevar a cabo lo siguiente ejercci

## Punto 1

Construir una agrupación de los países con los indicadores de Scimago por el método de K- medias seleccionando el número de grupos K con el indicador de Carlinski-Hrabasz y tipificar los grupos obtenidos por los promedios y las varianzas como en el ejenmplo 7.4.1

## Punto 2

Construir una agrupación de los países con los indicadores del Scimago por el método de K - medias seleccionando el número de grupos K con el indicador de Hartigan y tipificar los grupos obtenidos por los promedios y las varianzas con en el ejemplo 7.5.1

## Punto 3

Comparar las agrupaciones obtenidas en los ejercicios 1 y 2 y explicar las diferencias, si las hay, tanto en número de grupos como en países que quedaron en cada ejercicio

## Punto 4

Repetir los ejercicios 1,2 y 3 Pero utilizando los indicadores de QS.

## Punto 5

Construir una agrupación jerárquica por países utilizando la salida del PCA con los indicadores del Scimago, tipificar los grupos con en el ejemplo 7.10.1

## Punto 6

Construir una agrupación jerárquica por países utilizando la salida del PCA con los indicadores de QS, tipificar los grupos como en el ejemplo 7.10.1

### Selección de los datos

```
datos_qs <- datos %>%
  select(Pais,
         QS.Ranking,
         QS.Puntuacion,
         QS.Reputacion.academica,
         QS.Reputacion.entre.empleadores,
         QS.Estudiantes.por.profesor,
         QS.PromArticulos.por.docente,
         QS.Citas.por.articulo,
         QS.Docentes.con.doctorado,
         QS.Impacto.Web)

promedios_pais <- datos_qs %>%
  group_by(Pais) %>%
  summarise(across(everything(), mean, na.rm = TRUE))

# Preparar datos numéricos
datos_numericos <- promedios_pais %>%
  select(-Pais)

rownames(datos_numericos) <- promedios_pais$Pais

datos_numericos_df <- as.data.frame(datos_numericos)
rownames(datos_numericos_df) <- promedios_pais$Pais
# Hacer ACP
```

```
res.PCA <- PCA(datos_numericos_df, ncp = 9, graph = FALSE, scale.unit = TRUE)

# Extraer y mostrar valores propios
eig <- round(res.PCA$eig, 2)

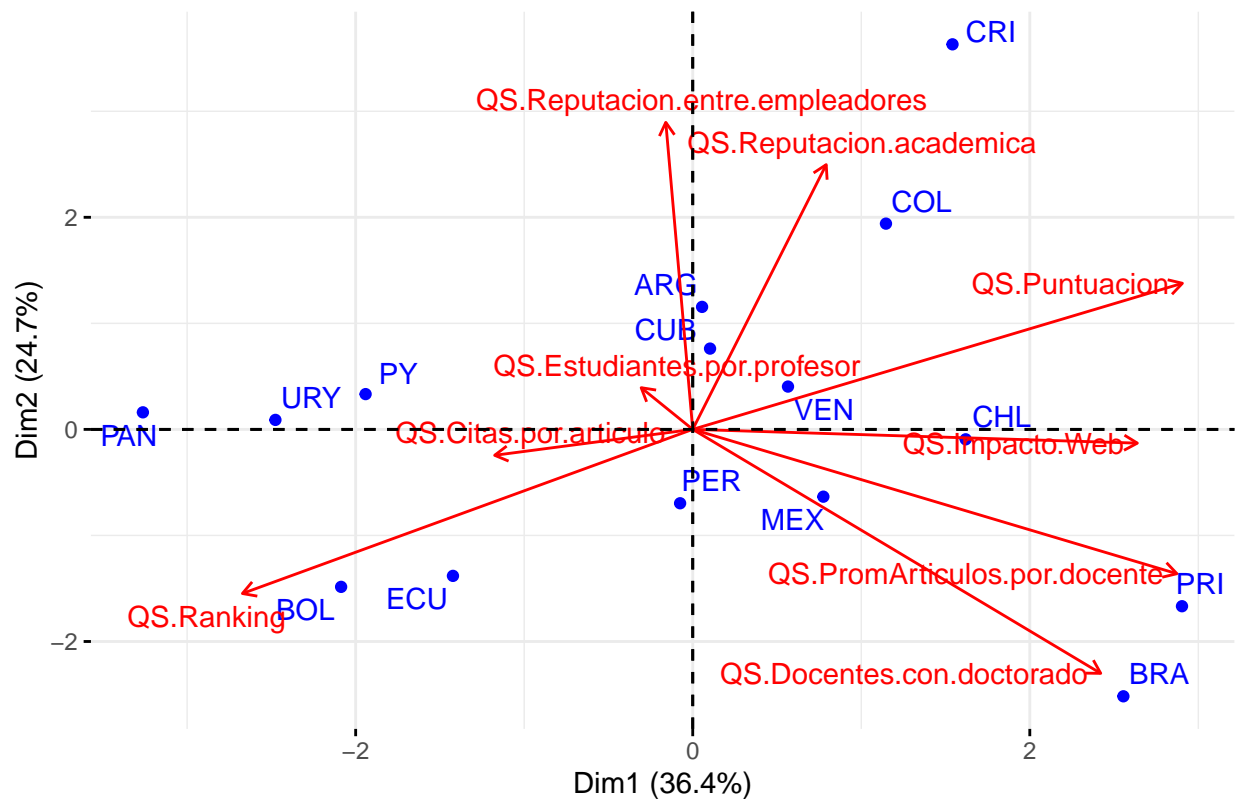
kable(eig, format = "latex", booktabs = TRUE,
      caption = "Valores propios y varianzas del ACP de indicadores del QS") %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

Table 1: Valores propios y varianzas del ACP de indicadores del QS

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.28	36.39	36.39
comp 2	2.22	24.67	61.06
comp 3	1.57	17.44	78.50
comp 4	1.01	11.18	89.68
comp 5	0.59	6.53	96.21
comp 6	0.18	1.99	98.20
comp 7	0.11	1.21	99.41
comp 8	0.05	0.54	99.96
comp 9	0.00	0.04	100.00

```
fviz_pca_biplot(res.PCA,
  label = c("var", "ind"),
  repel = TRUE,
  col.var = "red",
  col.ind = "blue",
  title = "Biplot del PCA de indicadores QS (promedios por país)")
```

Biplot del PCA de indicadores QS (promedios por país)



```
# 1. Extraer las coordenadas de los países
coord_raw <- as.data.frame(res.PCA$ind$coord)
```

```
# 5. Generar la tabla
```

```
kable(coord_raw,
      format = "latex",
      booktabs = TRUE,
      caption = "Coordenadas de los países sobre las primeras 5 componentes del ACP de indicadores QS"
      label = "tab:coord_paises_qs") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

```
# 1. Extraer la matriz de contribuciones de variables
contrib <- as.data.frame(res.PCA$var$contrib)
```

```
kable(contrib,
      format = "latex",
      booktabs = TRUE,
      caption = "Contribuciones a la varianza de los primeros factores del ACP de los indicadores de QS"
      label = "tab:contrib_varianza_qs") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

```
# 1. Extraer la matriz de cosenos cuadrados de variables
cos2 <- as.data.frame(res.PCA$var$cos2)
```

Table 2: Coordenadas de los países sobre las primeras 5 componentes del ACP de indicadores QS

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9
ARG	0.0553692	1.1551279	0.5376874	-0.2423438	-0.4057592	-0.2460565	-0.0523415	0.0691996	-0.0016669
BOL	-2.0863655	-1.4850598	-1.7338709	2.0374641	-0.9915697	0.0435581	-0.1398526	0.1759736	0.0867538
BRA	2.5552991	-2.5170114	0.4809138	-1.1595736	-0.3775776	0.7791567	0.1710793	0.2440908	0.0663737
CHL	1.6195912	-0.0935815	-0.4362781	-0.2885591	-0.8209235	-0.5206462	0.7711150	-0.1149535	-0.0225526
COL	1.1465121	1.9404104	-0.9452651	-0.4361957	-1.1799837	-0.0853826	0.0650674	0.0095933	-0.0510178
CRI	1.5412683	3.6311439	0.8587740	1.3648181	0.5754167	0.3217094	0.0675189	-0.0373707	0.1067383
CUB	0.1026698	0.7618092	-2.2962401	-0.1709268	1.4625128	0.3365789	0.0125006	0.3850595	-0.0935406
ECU	-1.4227769	-1.3814022	-0.5028941	1.1548081	0.8543210	0.1260512	0.3864631	-0.4281607	-0.0395708
MEX	0.7749895	-0.6343816	0.0645598	-0.5463992	0.5598738	0.4251191	0.0195478	-0.3525444	0.0114495
PAN	-3.2624187	0.1608980	-0.2362558	-0.8618182	-0.0636272	-0.2764801	0.1394966	0.1017456	0.0233837
PER	-0.0750074	-0.6965576	0.2002001	-0.8246675	0.3751968	-0.2578245	-0.4997392	-0.2010767	0.0463708
PRI	2.9030301	-1.6674545	1.4046391	1.4829303	0.4680613	-0.7221585	-0.2781988	0.1734875	-0.0567592
PY	-1.9411105	0.3327833	2.2107585	0.4506918	-0.9577416	0.6763758	-0.2329368	-0.0204197	-0.1192181
URY	-2.4762331	0.0893699	1.8753030	-0.9796684	0.7678657	-0.3949733	0.1986696	0.2061328	0.0410664
VEN	0.5651827	0.4039060	-1.4820317	-0.9805602	-0.2660656	-0.2050275	-0.6283894	-0.2107567	0.0021898

Table 3: Contribuciones a la varianza de los primeros factores del ACP de los indicadores de QS

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9
QS.Ranking	18.4096694	9.1120547	10.2269074	2.1690103	0.1135665	3.1824880	3.0471884	1.283954	52.4551610
QS.Puntuacion	21.7669648	7.2234908	6.7546109	0.2442678	0.4891739	1.6243742	0.2182294	21.729552	39.9493360
QS.Reputacion.academica	1.6177817	23.7147299	5.3931871	5.3359980	45.0087617	6.4285710	0.8555042	10.904453	0.7410135
QS.Reputacion.entre.empleadores	0.0658472	31.8891733	1.8843646	1.5071666	39.6289742	0.6208831	0.0242128	22.857979	1.5213995
QS.Estudiantes.por.profesor	0.2412685	0.5879019	45.5594039	20.8737177	0.9454985	25.1172840	0.2394831	6.154541	0.2809010
QS.PromArticulos.por.docente	21.2819903	7.0444779	0.0006352	5.5698711	4.1722351	1.1283821	56.9065027	3.538527	0.3573788
QS.Citas.por.articulo	3.5623796	0.2270817	21.6801521	48.9902316	4.5311180	2.8023904	8.4179639	8.010008	1.7786746
QS.Docentes.con.doctorado	15.1100185	20.1371445	0.0716663	0.0005615	0.0703941	6.1063558	30.2865346	25.305839	2.9114863
QS.Impacto.Web	17.9440801	0.0639453	8.4290725	15.3091754	5.0402781	52.9892714	0.0043810	0.215147	0.0046492

## # 5. Generar la tabla LaTeX

```

kable(cos2,
      format = "latex",
      booktabs = TRUE,
      caption = "Cosenos cuadrados de las variables sobre los primeros 5 factores del ACP de indicadores",
      label = "tab:cos2_qs") %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))

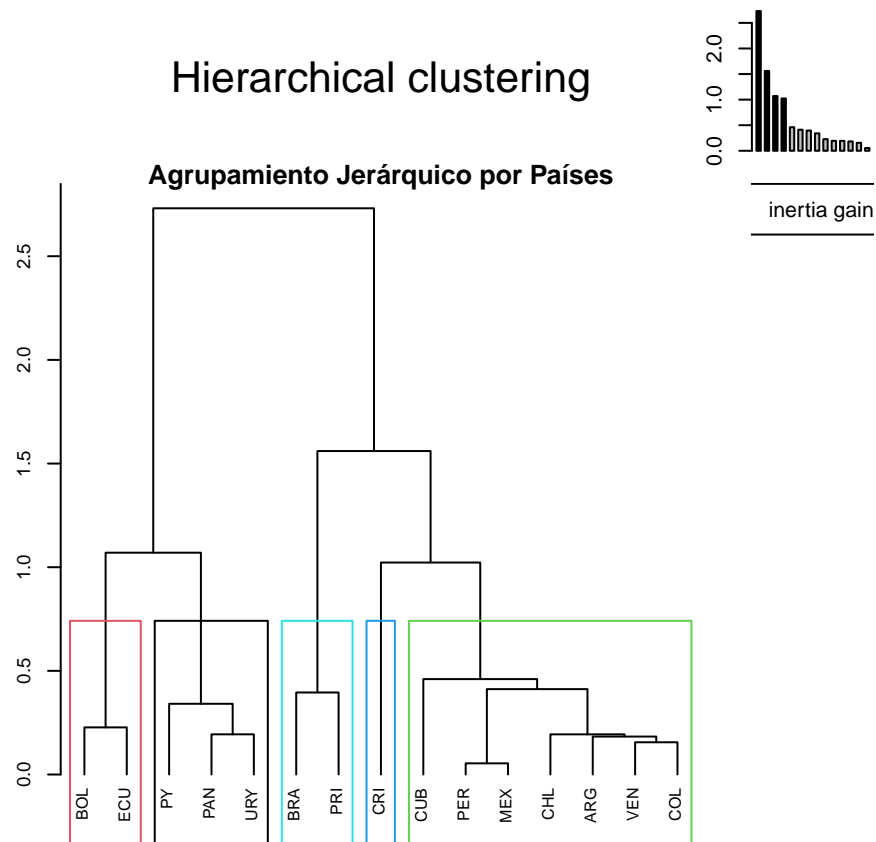
```

Table 4: Cosenos cuadrados de las variables sobre los primeros 5 factores del ACP de indicadores de QS

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9
QS.Ranking	0.6029745	0.2022794	0.1605446	0.0218325	0.0006671	0.0056887	0.0033317	0.0006284	0.0020530
QS.Puntuacion	0.7129365	0.1603550	0.1060356	0.0024587	0.0028736	0.0029036	0.0002386	0.0106348	0.0015636
QS.Reputacion.academica	0.0529874	0.5264456	0.0846636	0.0537103	0.2644008	0.0114910	0.0009354	0.0053368	0.0000290
QS.Reputacion.entre.empleadores	0.0021567	0.7079109	0.0295812	0.0151706	0.2327976	0.0011098	0.0000265	0.0111871	0.0000595
QS.Estudiantes.por.profesor	0.0079023	0.0130509	0.7152031	0.2101075	0.0055543	0.0448970	0.0002618	0.0030121	0.0000110
QS.PromArticulos.por.docente	0.6970521	0.1563811	0.0000100	0.0560644	0.0245095	0.0020170	0.0622202	0.0017318	0.0000140
QS.Citas.por.articulo	0.1166791	0.0050410	0.3403405	0.4931185	0.0266177	0.0050093	0.0092040	0.0039202	0.0000696
QS.Docentes.con.doctorado	0.4949006	0.4470265	0.0011250	0.0000057	0.0004135	0.0109151	0.0331146	0.0123851	0.0001140
QS.Impacto.Web	0.5877250	0.0014195	0.1323217	0.1540968	0.0296088	0.0947179	0.0000048	0.0001053	0.0000002

```
res.hcpc <- HCPC(res.PCA, nb.clust = -1, min = 3, max = 6, proba = 0.1, graph = FALSE)
```

```
# Mejorar con ajustes de etiquetas y estilo base R
plot(res.hcpc,
     choice = "tree",
     axes   = c(1, 2),
     main   = "Agrupamiento Jerárquico por Países",
     cex    = 0.8,      # tamaño de texto de etiquetas
     cex.main = 1.2,    # tamaño del título
     cex.axis = 0.9,    # tamaño ejes
     cex.lab = 1        # tamaño etiquetas de ejes
)
```



```
# 1. Extraer la asignación de cada país desde HCPC
```

```
df_clust <- res.hcpc$data.clust %>%
  rownames_to_column("Pais") %>% # pasa los rownames (códigos) a la columna Pais
  select(Pais, clust)
```

```
# 2. Agrupar y resumir: contar y pegar la lista de países
```

```
tabla_composicion <- df_clust %>%
  group_by(clust) %>%
  summarise(
    N_paises = n(),
    Paises   = paste(Pais, collapse = ", ")
  ) %>%
  arrange(clust) %>%
```

```

rename(
  Cluster = clust,
  `# Países` = N_paises
)

# 3. Verificar en R
#print(tabla_composicion)

# 4. Generar tabla LaTeX
kable(tabla_composicion,
  format = "latex",
  booktabs = TRUE,
  caption = "Composición de los grupos resultantes del HCPC",
  label = "tab:composicion_grupos") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))

```

Table 5: Composición de los grupos resultantes del HCPC

Cluster	# Países	Países
1	3	PAN, PY, URY
2	2	BOL, ECU
3	7	ARG, CHL, COL, CUB, MEX, PER, VEN
4	1	CRI
5	2	BRA, PRI

```

desc1 <- as.data.frame(res.hcpc$desc.var$quanti[[1]])

# Renombrar columnas con nombres más legibles
colnames(desc1) <- c(
  # Nombre de la variable
  "V-test", # Estadístico v-test
  "Media en Grupo 1", # Media dentro del grupo
  "Media Global", # Media general
  "SD en Grupo 1", # Desviación estándar en el grupo
  "SD Global", # Desviación estándar global
  "P-valor" # Valor p de la prueba
)

# Generar la tabla en LaTeX
kable(desc1,
  format = "latex",
  booktabs = TRUE,
  caption = "Características principales del Grupo 1 (ACP indicadores QS)",
  label = "tab:desc_grupo1") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))

desc1 <- as.data.frame(res.hcpc$desc.var$quanti[[2]])

# Renombrar columnas con nombres más legibles
colnames(desc1) <- c(
  # Nombre de la variable
  "V-test", # Estadístico v-test

```

Table 6: Características principales del Grupo 1 (ACP indicadores QS)

	V-test	Media en Grupo 1	Media Global	SD en Grupo 1	SD Global	P-valor
QS.Estudiantes.por.profesor	2.305460	71.508333	47.51372	13.116090	19.471087	0.0211408
QS.Puntuacion	-1.653133	50.541667	56.91801	3.836901	7.216019	0.0983037
QS.Docentes.con.doctorado	-1.906793	18.533333	43.49244	4.800058	24.488349	0.0565473
QS.PromArticulos.por.docente	-1.977893	4.258333	25.92898	2.901604	20.497607	0.0479408
QS.Impacto.Web	-3.098552	59.208333	77.94462	5.184686	11.312504	0.0019447

```

"Media en Grupo 2", # Media dentro del grupo
"Media Global",    # Media general
"SD en Grupo 2",   # Desviación estándar en el grupo
"SD Global",       # Desviación estándar global
"P-valor"          # Valor p de la prueba
)

# Generar la tabla en LaTeX
kable(desc1,
      format = "latex",
      booktabs = TRUE,
      caption = "Características principales del Grupo 2 (ACP indicadores QS)",
      label = "tab:desc_grupo2") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))

```

Table 7: Características principales del Grupo 2 (ACP indicadores QS)

	V-test	Media en Grupo 2	Media Global	SD en Grupo 2	SD Global	P-valor
QS.Ranking	1.974189	106.33333	79.62590	5.666667	19.854123	0.0483602
QS.Reputacion.academica	-1.652284	45.86667	62.04669	7.966667	14.371486	0.0984766
QS.Puntuacion	-1.986624	47.15000	56.91801	2.350000	7.216019	0.0469641
QS.Estudiantes.por.profesor	-2.213239	18.15000	47.51372	12.150000	19.471087	0.0268812

```

desc1 <- as.data.frame(res.hcpc$desc.var$quanti[[3]])

# Renombrar columnas con nombres más legibles
colnames(desc1) <- c(
  # Nombre de la variable
  "V-test",          # Estadístico v-test
  "Media en Grupo 3", # Media dentro del grupo
  "Media Global",     # Media general
  "SD en Grupo 3",    # Desviación estándar en el grupo
  "SD Global",        # Desviación estándar global
  "P-valor"          # Valor p de la prueba
)

# Generar la tabla en LaTeX
kable(desc1,
      format = "latex",
      booktabs = TRUE,
      caption = "Características principales del Grupo 3 (ACP indicadores QS)",
      label = "tab:desc_grupo2") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))

```



Table 8: Características principales del Grupo 3 (ACP indicadores QS)

	V-test	Media en Grupo 3	Media Global	SD en Grupo 3	SD Global	P-valor
QS.Reputacion.academica	1.756266	69.25816	62.04669	9.394464	14.37149	0.0790430
QS.Citas.por.articulo	-2.792955	52.47206	69.68927	11.453714	21.57580	0.0052229

```
desc1 <- as.data.frame(res.hcpc$desc.var$quanti[[4]])

# Renombrar columnas con nombres más legibles
colnames(desc1) <- c(
  # Nombre de la variable
  "V-test",          # Estadístico v-test
  "Media en Grupo 4", # Media dentro del grupo
  "Media Global",     # Media general
  "SD en Grupo 4",    # Desviación estándar en el grupo
  "SD Global",        # Desviación estándar global
  "P-valor"           # Valor p de la prueba
)

# Generar la tabla en LaTeX
kable(desc1,
  format = "latex",
  booktabs = TRUE,
  caption = "Características principales del Grupo 4 (ACP indicadores QS)",
  label = "tab:desc_grupo5") %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

Table 9: Características principales del Grupo 4 (ACP indicadores QS)

	V-test	Media en Grupo 4	Media Global	SD en Grupo 4	SD Global	P-valor
QS.Puntuacion	1.986136	71.25	56.91801	0	7.216019	0.0470182
QS.Reputacion.academica	1.757182	87.30	62.04669	0	14.371486	0.0788869
QS.Reputacion.entre.empleadores	1.718007	82.70	61.79570	0	12.167760	0.0857953
QS.Ranking	-2.071404	38.50	79.62590	0	19.854123	0.0383211

```
desc1 <- as.data.frame(res.hcpc$desc.var$quanti[[5]])

# Renombrar columnas con nombres más legibles
colnames(desc1) <- c(
  # Nombre de la variable
  "V-test",          # Estadístico v-test
  "Media en Grupo 5", # Media dentro del grupo
  "Media Global",     # Media general
  "SD en Grupo 5",    # Desviación estándar en el grupo
  "SD Global",        # Desviación estándar global
  "P-valor"           # Valor p de la prueba
)

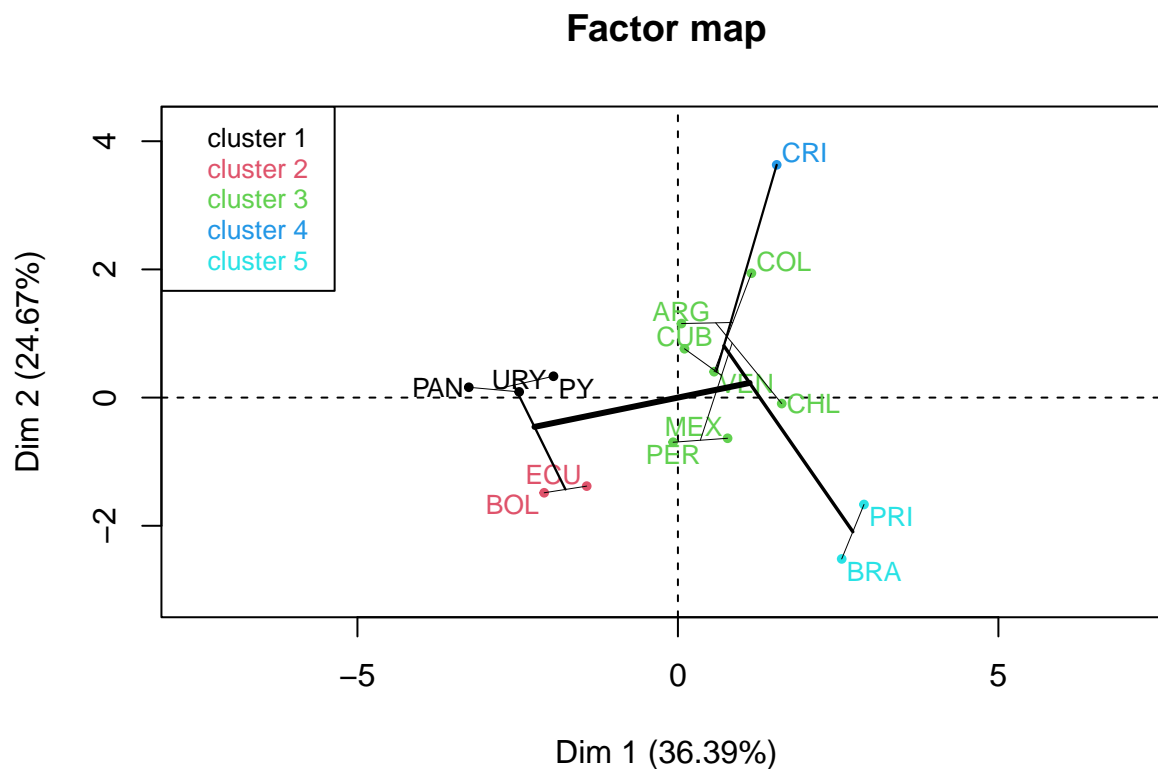
# Generar la tabla en LaTeX
kable(desc1,
  format = "latex",
```

```
booktabs = TRUE,
caption = "Características principales del Grupo 5 (ACP indicadores QS)",
label = "tab:desc_grupo5") %>%
kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

Table 10: Características principales del Grupo 5 (ACP indicadores QS)

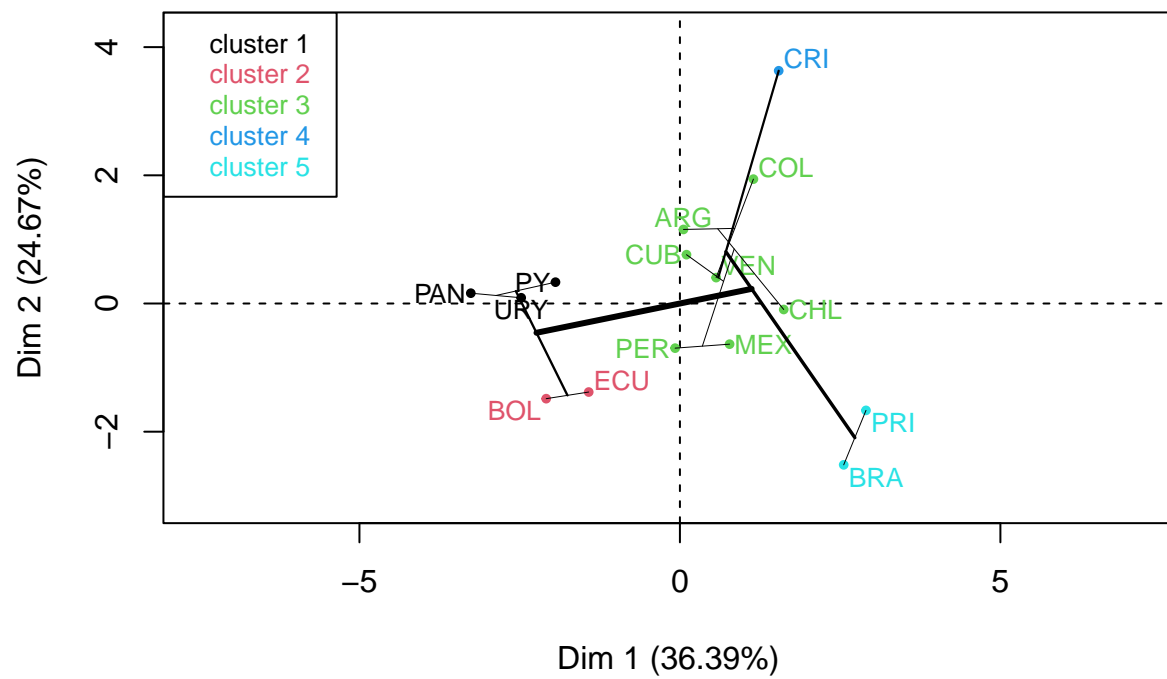
	V-test	Media en Grupo 5	Media Global	SD en Grupo 5	SD Global	P-valor
QS.Docentes.con.doctorado	3.157814	96.18372	43.49244	3.8162791	24.48835	0.0015896
QS.PromArticulos.por.docente	2.590126	62.10465	25.92898	10.7046512	20.49761	0.0095941
QS.Reputacion.entre.empleadores	-1.887451	46.14697	61.79570	0.1469697	12.16776	0.0590997

```
plot.HCPC(res.hcpc, axes = c(1,2), choice = "map")
```



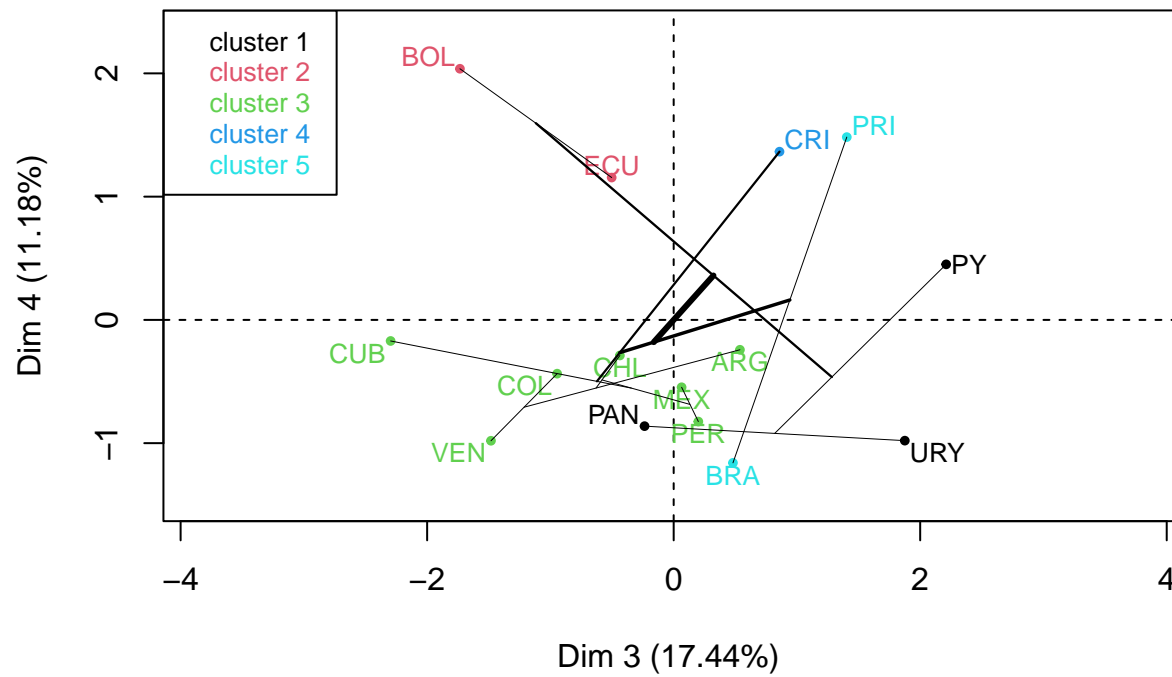
```
plot.HCPC(res.hcpc, choice = "map", axes = c(1, 2), title = "Distribución de los grupos sobre el plano")
```

## Distribución de los grupos sobre el plano 1-2 (ACP indicadores QS)



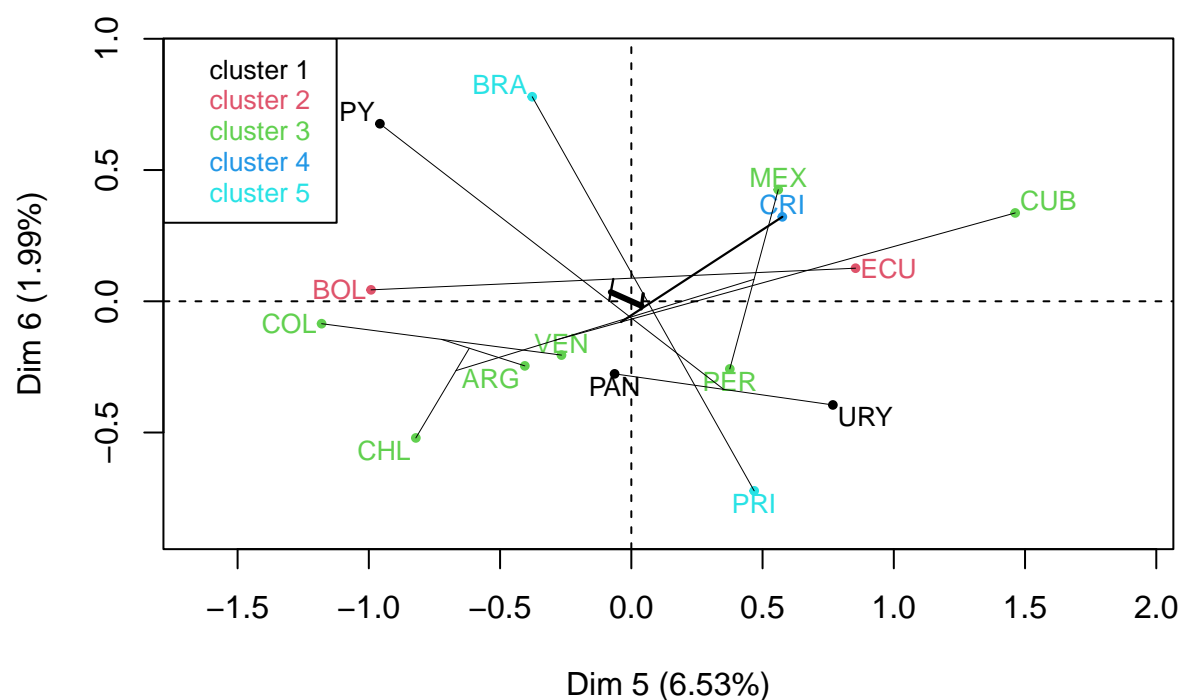
```
plot.HCPC(res.hcpc, choice = "map", axes = c(3, 4), title = "Distribución de los grupos sobre el plano 3-4")
```

## Distribución de los grupos sobre el plano 3-4 (ACP indicadores QS)



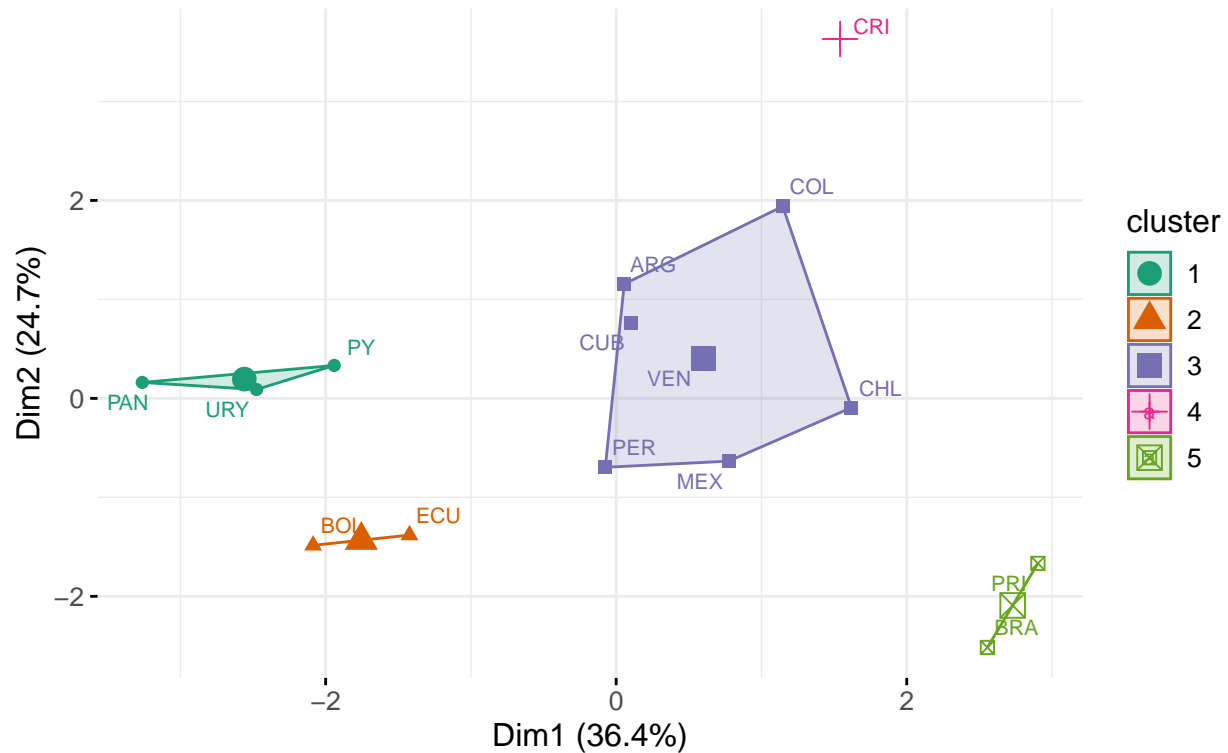
```
plot.HCPC(res.hcpc, choice = "map", axes = c(5, 6), title = "Distribución de los grupos sobre el plano 3-4")
```

## Distribución de los grupos sobre el plano 5-6 (ACP indicadores QS)



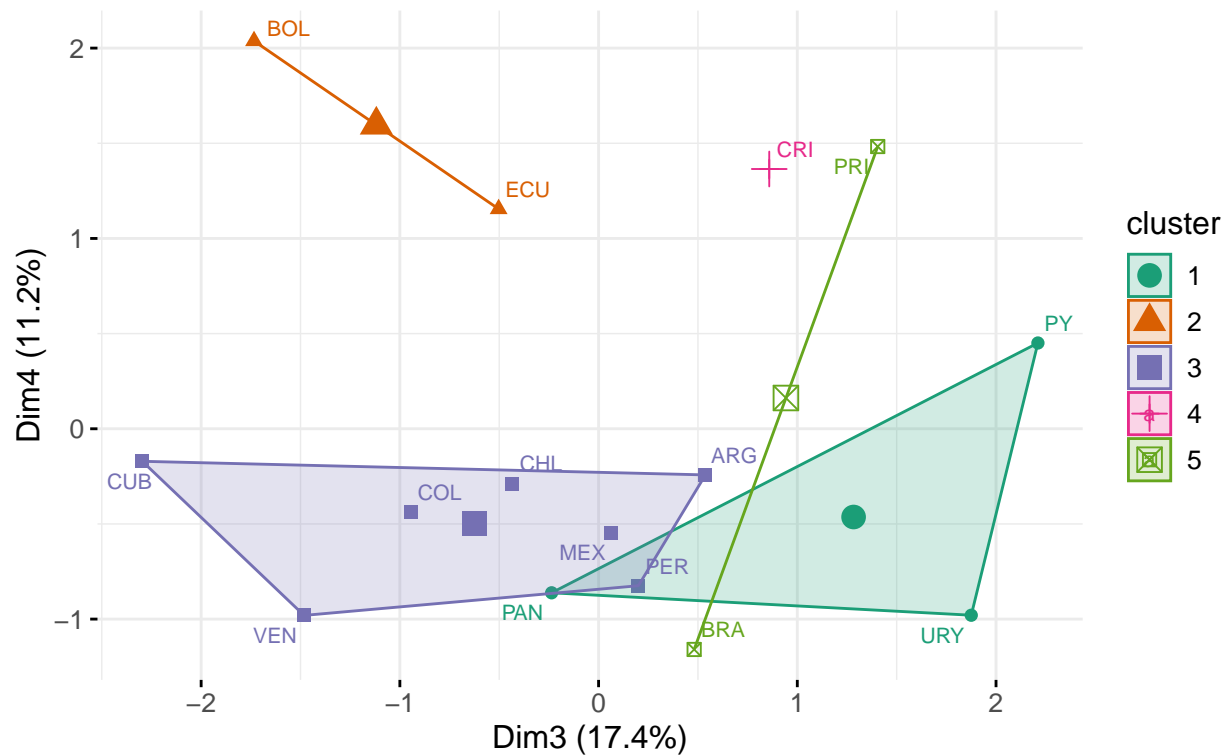
```
fviz_cluster(res.hcpc,
  geom = c("point", "text"),      # Puntos y texto
  pointsize = 2,
  labelsize = 8,
  ellipse.type = "convex",
  palette = "Dark2",
  repel = TRUE,
  show.clust.cent = TRUE,
  axes = c(1, 2),
  ggtheme = theme_minimal(),
  main = "Distribución de los grupos en el plano 1-2\n(ACP de indicadores QS)" ) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10)
  )
)
```

## Distribución de los grupos en el plano 1-2 (ACP de indicadores QS)

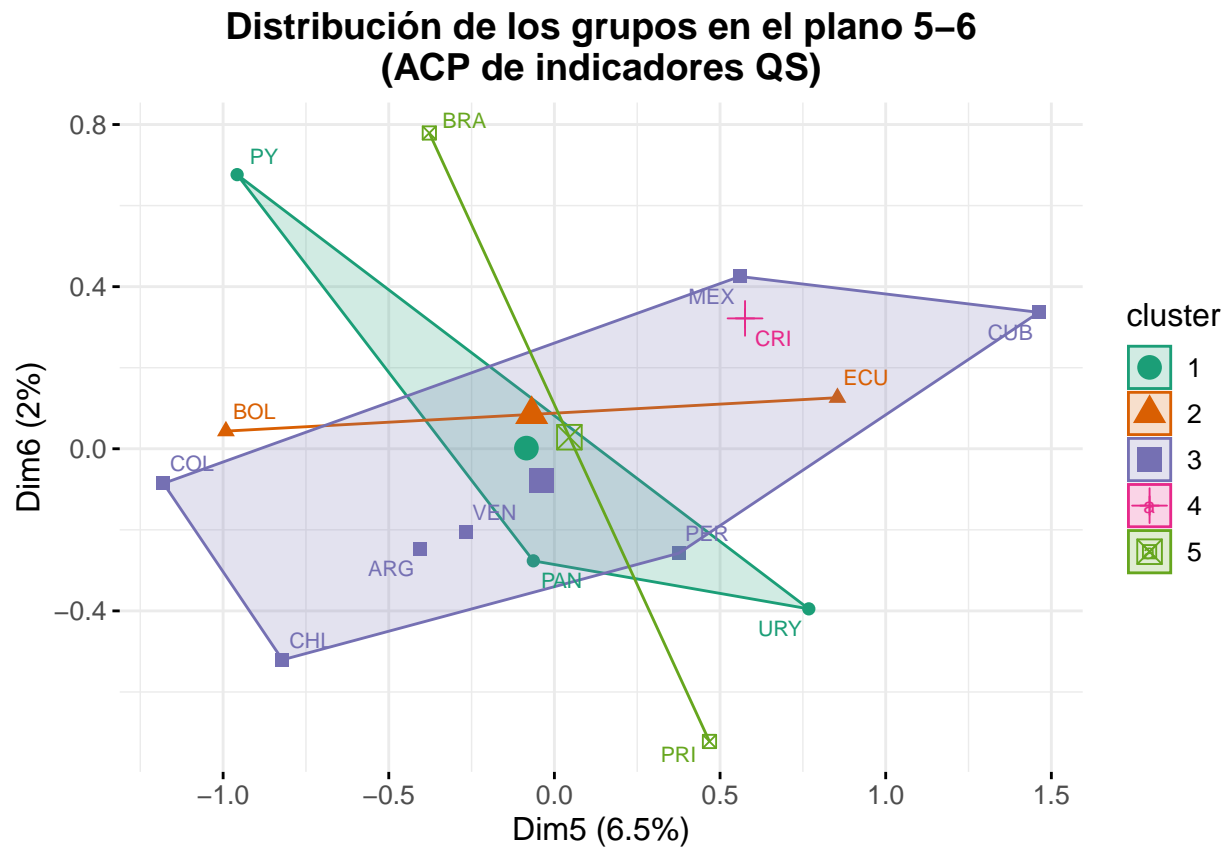


```
fviz_cluster(res.hcpc,
  geom = c("point", "text"),      # Puntos y texto
  pointsize = 2,
  labelsize = 8,
  ellipse.type = "convex",
  palette = "Dark2",
  repel = TRUE,
  show.clust.cent = TRUE,
  axes = c(3, 4),
  ggtheme = theme_minimal(),
  main = "Distribución de los grupos en el plano 3-4\n(ACP de indicadores QS)" ) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10)
  )
```

## Distribución de los grupos en el plano 3-4 (ACP de indicadores QS)



```
fviz_cluster(res.hcpc,
  geom = c("point", "text"),      # Puntos y texto
  pointsize = 2,
  labelsize = 8,
  ellipse.type = "convex",
  palette = "Dark2",
  repel = TRUE,
  show.clust.cent = TRUE,
  axes = c(5, 6),
  ggtheme = theme_minimal(),
  main = "Distribución de los grupos en el plano 5-6\n(ACP de indicadores QS)" ) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    legend.title = element_text(size = 12),
    legend.text = element_text(size = 10),
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10)
  )
```



## Punto 7

Comparar las agrupaciones de los ejercicios y explicar las diferencias, si las hay, tanto en número de grupos como en países que quedaron en cada ejercicio