

# 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)
```

### Promedio de datos

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

### Ejecución de ACP e interpretación de valores propios y varianza explicada

```
# 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

La Tabla 1 muestra los valores propios (eigenvalues) y las porcentajes de varianza explicada, tanto individuales como acumuladas, para cada uno de los nueve componentes principales obtenidos a partir del Análisis de Componentes Principales (ACP) aplicado a los indicadores del ranking QS.

El primer componente explica el 36.39% de la varianza total de los datos, lo que indica que este eje concentra una gran parte de la información contenida en las variables originales. El segundo componente agrega un 24.67% adicional, por lo que juntos, los dos primeros componentes explican aproximadamente el 61.06% de la varianza total, lo cual ya representa una reducción significativa de la dimensión del problema manteniendo una buena cantidad de información.

Los primeros cuatro componentes alcanzan una varianza acumulada de 89.68%, lo cual es considerado muy satisfactorio en análisis exploratorios, ya que permite visualizar los datos en un espacio de menor dimensión (2D o 3D) sin una pérdida considerable de información.

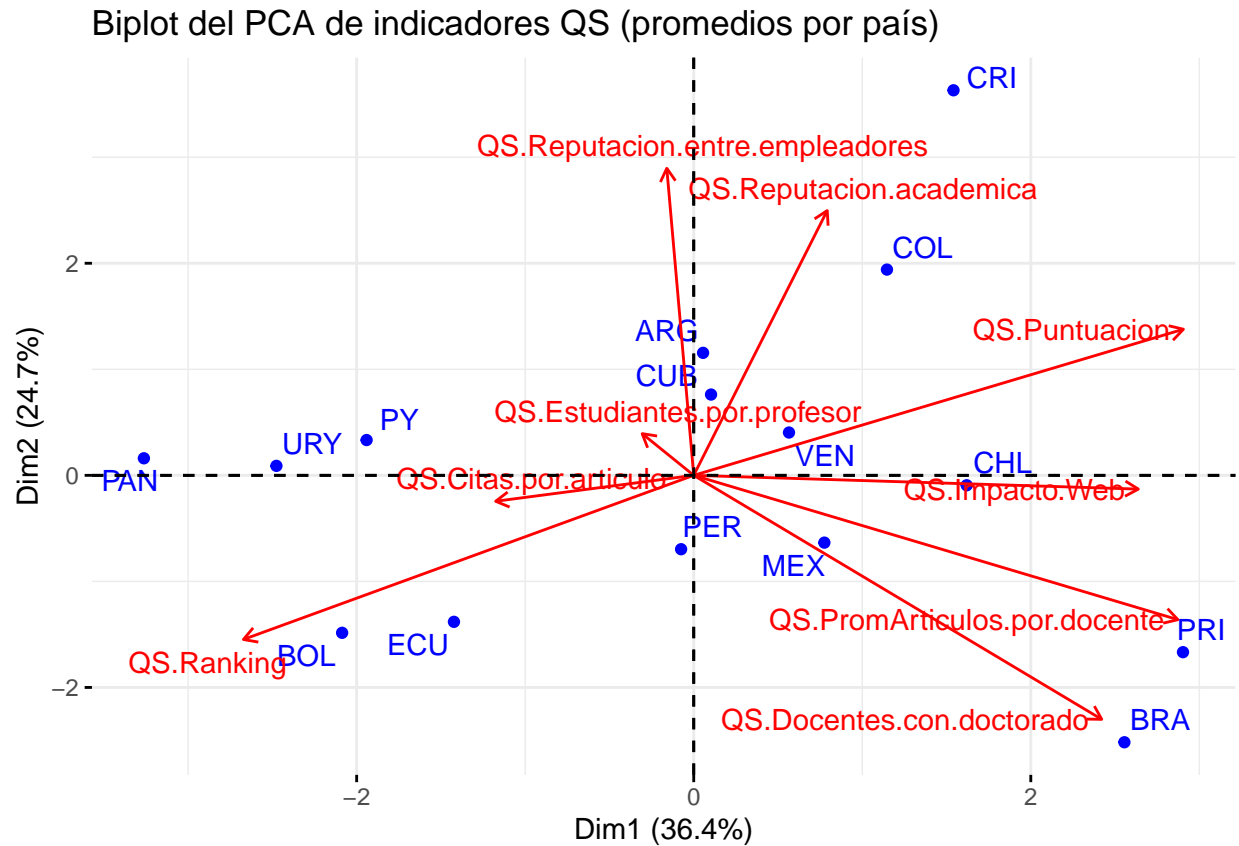
A pesar de que se calcularon los nueve componentes posibles (tantos como variables originales), a partir del componente 5 en adelante, la varianza explicada por cada componente es bastante baja (menor al 7%), indicando que su aporte a la explicación de la estructura de los datos es limitado.

Con base en esta información, decidimos trabajar inicialmente con los primeros seis componentes, que explican en conjunto el 98.20% de la varianza total. Sin embargo, dado que los componentes 5 y 6 apenas agregan un 6.53% y 1.99% respectivamente, una reducción a cuatro componentes principales podría ser más eficiente, conservando casi el 90% de la varianza total y simplificando aún más el análisis.

Esta selección permite un equilibrio adecuado entre simplificación del modelo y preservación de la información relevante, lo que resulta fundamental para la interpretación de los grupos (clusters) posteriores y para la visualización en planos factoriales.

## Biplot

```
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)")
```



El gráfico muestra un biplot correspondiente a los dos primeros componentes principales obtenidos del Análisis de Componentes Principales (ACP) aplicado a los promedios nacionales de los indicadores del ranking QS. El eje horizontal (Dim1) explica el 36.4% de la varianza total, mientras que el eje vertical (Dim2) representa un 24.7% adicional, sumando entre ambos un 61.1% de la variabilidad de los datos originales.

En este biplot:

Los puntos azules representan países, proyectados en el nuevo espacio definido por los dos primeros componentes.

Las flechas rojas representan los vectores de los indicadores QS, que indican la dirección y fuerza de su contribución a la varianza explicada.

Interpretación de los ejes Componente 1 (Dim1) parece estar fuertemente influido por indicadores como:

- QS.Puntuación
- QS.Docentes con doctorado
- QS.Promedio de artículos por docente

- QS.Impacto web

en contraste con QS.Ranking (que apunta en la dirección opuesta, indicando que un mejor ranking numérico -menor valor- se asocia negativamente con este componente).

Esto sugiere que Dim1 refleja una dimensión de capacidad investigativa, calidad docente y visibilidad, donde los países más hacia la derecha (por ejemplo, Chile, Puerto Rico, Brasil) presentan mejores resultados en estos aspectos.

Componente 2 (Dim2) está más relacionado con:

- QS.Reputación académica
- QS.Reputación entre empleadores
- QS.Estudiante por profesor

Estos factores definen una dimensión de prestigio percibido y atención individual al estudiante, lo que explica la posición destacada de países como Costa Rica y Colombia, más hacia la parte superior del gráfico.

Agrupación de países Países como Costa Rica (CRI) y Colombia (COL) se ubican en la región superior derecha del gráfico, lo que indica buenos desempeños tanto en indicadores de prestigio como en métricas duras de desempeño.

Chile (CHL) y Puerto Rico (PRI) se asocian principalmente a altos niveles de producción académica e indicadores institucionales, más alineados con el componente 1.

Brasil (BRA) destaca especialmente por sus docentes con doctorado, aunque su reputación relativa parece menor (proyectado hacia abajo).

En cambio, países como Ecuador (ECU), Bolivia (BOL) y Panamá (PAN) aparecen en la región izquierda del gráfico, lo que sugiere menores puntuaciones en la mayoría de los indicadores, incluyendo reputación y producción académica.

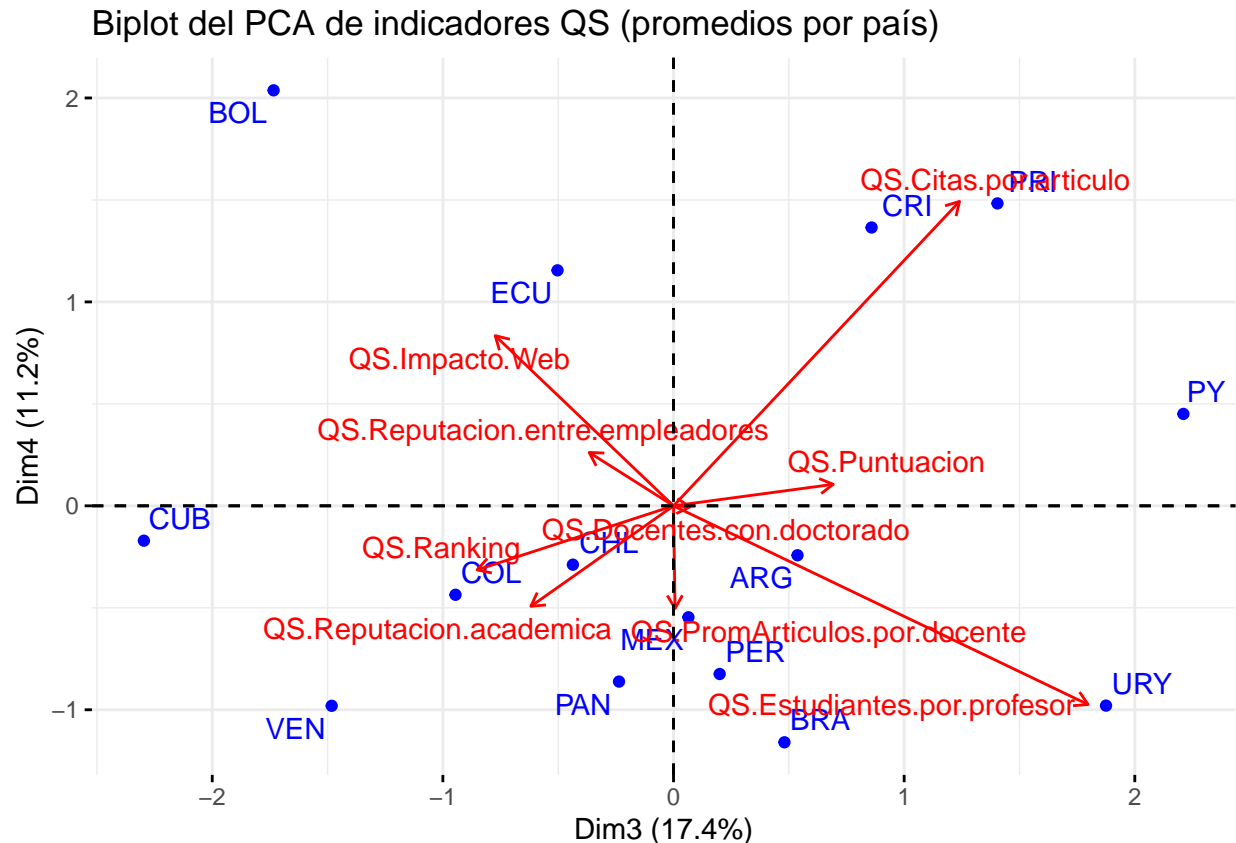
Algunos países, como Argentina (ARG) y Cuba (CUB), se ubican cerca del centro, indicando un desempeño intermedio o balanceado, sin destacar fuertemente en una dirección específica.

Relación entre variables Las flechas que apuntan en direcciones similares están positivamente correlacionadas (por ejemplo, QS.Puntuación y QS.Impacto Web).

Flechas en direcciones opuestas indican correlaciones negativas, como ocurre entre QS.Ranking y la mayoría de los demás indicadores.

La longitud de las flechas indica la contribución del indicador al plano mostrado: flechas más largas como las de QS.Puntuación o QS.Docentes con doctorado tienen mayor peso en la construcción de los componentes visualizados.

```
fviz_pca_biplot(res.PCA,
  label = c("var", "ind"),
  axes  = c(3, 4),
  repel = TRUE,
  col.var = "red",
  col.ind = "blue",
  title = "Biplot del PCA de indicadores QS (promedios por país)")
```



El gráfico presenta un biplot que muestra la proyección de los promedios por país en el espacio definido por el tercer (Dim3) y cuarto (Dim4) componente principal del Análisis de Componentes Principales (ACP) sobre los indicadores del ranking QS. En conjunto, estos dos componentes explican aproximadamente un 28.6% de la variabilidad total: 17.4% por parte de Dim3 y 11.2% por parte de Dim4.

En este biplot:

Los puntos azules representan los países, posicionados de acuerdo con su proyección en los componentes 3 y 4.

Las flechas rojas representan las variables originales (indicadores QS), indicando tanto su dirección como su contribución relativa a estos componentes.

**Dimensión 3 (Dim3)** El eje horizontal (Dim3) está influenciado por varios indicadores, pero destacan:

- QS.Estudiante.por.profesor
- QS.Ranking (en dirección opuesta) en menor medida

Esto sugiere que Dim3 refleja un eje de productividad y atención al estudiante, donde valores altos están asociados con mayor producción académica por docente y menor número de estudiantes por profesor (mejor atención), mientras que valores negativos tienden a indicar peores posiciones en el ranking (números mayores).

**Dimensión 4 (Dim4)** El eje vertical (Dim4) está principalmente asociado con:

- QS.Citas.por.Artículo
- En menor medida, con QS.Impacto.Web

Esto puede interpretarse como una dimensión de visibilidad e impacto académico, donde valores positivos representan países cuyas instituciones tienen un mayor número de citas por artículo, reflejando reconocimiento y alcance de la producción científica.

Interpretación por países Costa Rica (CRI) y Paraguay (PY) se destacan en la parte superior derecha del gráfico, lo que indica un buen desempeño en indicadores como citas por artículo, y puntuación QS.

Uruguay (URY) también aparece con una fuerte proyección positiva en Dim3, sugiriendo buen desempeño en promedio de artículos por docente y estudiantes por profesor.

Bolivia (BOL) y Ecuador (ECU) presentan valores positivos sobre Dim4, indicando un buen posicionamiento en cuanto a citas por artículo o visibilidad, a pesar de que no destaquen en otros indicadores.

Por otro lado, Venezuela (VEN), Cuba (CUB) y Colombia (COL) tienen coordenadas negativas en ambos componentes, lo que indica menores valores relativos en estos indicadores.

Argentina (ARG), México (MEX) y Panamá (PAN) se encuentran cerca del origen, lo que indica un perfil más neutral o balanceado en relación con estas dimensiones específicas del PCA.

Relación entre variables QS.Citas.por.Artículo está casi exclusivamente alineada con Dim4, lo que refuerza su papel como indicador de impacto académico.

Algunas variables como QS.Reputación.académica o QS.Impacto.Web tienen una menor contribución a estos componentes, como se aprecia en la corta longitud de sus vectores.

## Coordenadas

```
# 1. Extraer las coordenadas de los países (primeras 4 dimensiones)
coord_raw <- as.data.frame(res.PCA$ind$coord)[, 1:4]

# 2. Redondear a 2 decimales
coord_raw <- round(coord_raw, 2)

# 3. Generar la tabla
kable(coord_raw,
      format = "latex",
      booktabs = TRUE,
      caption = "Coordenadas de los países sobre las primeras 4 componentes del ACP de indicadores QS"
      label = "tab:coord_paises_qs") %>%
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

La Tabla ?? presenta las coordenadas de los países latinoamericanos sobre las primeras cuatro componentes principales obtenidas del Análisis de Componentes Principales (ACP) aplicado a los indicadores QS. Estas coordenadas reflejan la posición relativa de cada país en el nuevo espacio reducido, sintetizando la información multivariada de los indicadores originales.

La Dimensión 1 (Dim.1) concentra la mayor parte de la varianza explicada por estas cuatro dimensiones, y permite observar una clara separación entre países como Brasil (2.56), Puerto Rico (2.90) y Chile (1.62), que tienen valores positivos elevados, indicando un mejor desempeño relativo en los indicadores que más contribuyen a esta dimensión. Por el contrario, países como Panamá (-3.26), Uruguay (-2.48) y Bolivia (-2.09) se ubican en el extremo opuesto, lo que sugiere un perfil marcadamente diferente, posiblemente asociado a puntuaciones más bajas en esos indicadores.

En la Dimensión 2 (Dim.2), se destacan Costa Rica (3.63) y Venezuela (2.21) con valores positivos altos, lo que indica que capturan una variabilidad distinta a la de Dim.1, probablemente relacionada con indicadores específicos como citas por artículo o impacto web. Por su parte, México (-2.30) y Puerto Rico (-1.67) tienen valores negativos significativos en esta dimensión.

Las Dimensiones 3 y 4 (Dim.3 y Dim.4) explican menor proporción de la variabilidad total, pero aún aportan información relevante sobre las diferencias estructurales entre países. Por ejemplo, Uruguay (1.88) y Puerto Rico (1.40) presentan coordenadas positivas altas en Dim.3, mientras que países como Brasil (-2.52) y Bolivia (-1.73) se ubican en valores negativos, reflejando distintos patrones en indicadores como promedio

Table 2: Coordenadas de los países sobre las primeras 4 componentes del ACP de indicadores QS

	Dim.1	Dim.2	Dim.3	Dim.4
ARG	0.06	1.16	0.54	-0.24
BOL	-2.09	-1.49	-1.73	2.04
BRA	2.56	-2.52	0.48	-1.16
CHL	1.62	-0.09	-0.44	-0.29
COL	1.15	1.94	-0.95	-0.44
CRI	1.54	3.63	0.86	1.36
CUB	0.10	0.76	-2.30	-0.17
ECU	-1.42	-1.38	-0.50	1.15
MEX	0.77	-0.63	0.06	-0.55
PAN	-3.26	0.16	-0.24	-0.86
PER	-0.08	-0.70	0.20	-0.82
PRI	2.90	-1.67	1.40	1.48
PY	-1.94	0.33	2.21	0.45
URY	-2.48	0.09	1.88	-0.98
VEN	0.57	0.40	-1.48	-0.98

de artículos por docente o porcentaje de docentes con doctorado. En Dim.4, Bolivia (2.04), Puerto Rico (1.48) y Ecuador (1.15) destacan por sus valores positivos, lo que sugiere que esta dimensión podría estar capturando variabilidad en indicadores secundarios pero aún discriminantes.

*# 1. Extraer la matriz de contribuciones de variables*

```
contrib <- as.data.frame(res.PCA$var$contrib)[, 1:4]
```

```
contrib <- round(contrib, 2)
```

```
kable(contrib,
```

```
  format = "latex",
```

```
  booktabs = TRUE,
```

```
  caption = "Contribuciones a la varianza de los primeros 4 factores del ACP de los indicadores de
```

```
  label = "tab:contrib_varianza_qs") %>%
```

```
  kable_styling(latex_options = c("striped", "scale_down", "hold_position"))
```

Table 3: Contribuciones a la varianza de los primeros 4 factores del ACP de los indicadores de QS

	Dim.1	Dim.2	Dim.3	Dim.4
QS.Ranking	18.41	9.11	10.23	2.17
QS.Puntuacion	21.77	7.22	6.75	0.24
QS.Reputacion.academica	1.62	23.71	5.39	5.34
QS.Reputacion.entre.empleadores	0.07	31.89	1.88	1.51
QS.Estudiantes.por.profesor	0.24	0.59	45.56	20.87
QS.PromArticulos.por.docente	21.28	7.04	0.00	5.57
QS.Citas.por.articulo	3.56	0.23	21.68	48.99
QS.Docentes.con.doctorado	15.11	20.14	0.07	0.00
QS.Impacto.Web	17.94	0.06	8.43	15.31

La Tabla ?? presenta las contribuciones porcentuales de cada indicador QS a la varianza explicada por las primeras cuatro dimensiones del Análisis de Componentes Principales (ACP). Estas contribuciones permiten



interpretar el significado de cada dimensión en función de los indicadores originales, proporcionando así una guía para entender los factores latentes que estructuran los datos.

**Dimensión 1 (Dim.1): Calidad académica y producción científica** La primera dimensión concentra las mayores contribuciones de QS.Puntuación (21.77%), QS.PromArtículos.por.docente (21.28%), QS.Impacto.Web (17.94%), y QS.Docentes.con.doctorado (15.11%). Esta dimensión parece capturar un componente general de rendimiento académico y productividad investigativa, asociado a la puntuación global del ranking QS, el impacto digital y la formación del cuerpo docente.

**Dimensión 2 (Dim.2): Reputación institucional** En la segunda dimensión destacan QS.Reputación.entre.empleadores (31.89%), QS.Reputación.académica (23.71%) y QS.Docentes.con.doctorado (20.14%), sugiriendo que esta dimensión agrupa aspectos vinculados a la imagen externa de la institución, tanto desde la percepción académica como desde el sector laboral. Esta dimensión también representa el prestigio institucional más allá de métricas cuantitativas.

**Dimensión 3 (Dim.3): Relación docente-estudiante y visibilidad académica** La tercera dimensión está dominada por QS.Estudiantes.por.profesor (45.56%), junto con una contribución notable de QS.Impacto.Web (8.43%). Esto sugiere que esta dimensión representa factores estructurales del entorno de enseñanza, posiblemente asociados al tamaño relativo del cuerpo docente y la eficiencia educativa.

**Dimensión 4 (Dim.4): Citación y desempeño académico relativo** La cuarta dimensión presenta una contribución sobresaliente de QS.Citas.por.artículo (48.99%), lo cual indica que esta dimensión se relaciona principalmente con la visibilidad e impacto de la producción científica. También contribuyen en menor medida QS.Estudiantes.por.profesor (20.87%) y QS.Impacto.Web (15.31%), lo que sugiere una dimensión complementaria centrada en la influencia académica a nivel internacional.

```
# 1. Extraer la matriz de cosenos cuadrados de variables
cos2 <- as.data.frame(res.PCA$var$cos2)[, 1:4]

# 5. Generar la tabla LaTeX
cos2 <- round(cos2, 2)

kable(cos2,
      format = "latex",
      booktabs = TRUE,
      caption = "Cosenos cuadrados de las variables sobre los primeros 4 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 4 factores del ACP de indicadores de QS

	Dim.1	Dim.2	Dim.3	Dim.4
QS.Ranking	0.60	0.20	0.16	0.02
QS.Puntuacion	0.71	0.16	0.11	0.00
QS.Reputacion.academica	0.05	0.53	0.08	0.05
QS.Reputacion.entre.empleadores	0.00	0.71	0.03	0.02
QS.Estudiantes.por.profesor	0.01	0.01	0.72	0.21
QS.PromArticulos.por.docente	0.70	0.16	0.00	0.06
QS.Citas.por.articulo	0.12	0.01	0.34	0.49
QS.Docentes.con.doctorado	0.49	0.45	0.00	0.00
QS.Impacto.Web	0.59	0.00	0.13	0.15

Calidad de representación de los indicadores QS en los primeros factores La Tabla ?? presenta los cosenos cuadrados ( $\cos^2$ ) de cada variable respecto a las primeras cuatro componentes principales del Análisis de Componentes Principales (ACP). El coseno cuadrado representa la proporción de la varianza de cada variable

que es explicada por cada componente, y se interpreta como una medida de calidad de representación de la variable en ese plano factorial.

Variables mejor representadas en la primera dimensión Las variables QS.Puntuación (0.71), QS.PromArtículos.por.docente (0.70), QS.Ranking (0.60) y QS.Impacto.Web (0.59) muestran altos  $\cos^2$  sobre la primera dimensión, lo que confirma que están fuertemente asociadas al componente de desempeño académico general y producción científica ya identificado previamente.

Variables dominantes en la segunda dimensión La Reputación entre empleadores (0.71) y la Reputación académica (0.53) presentan sus mayores  $\cos^2$  en la segunda dimensión, lo que indica que esta componente captura especialmente aspectos de percepción externa e institucional. La variable Docentes con doctorado también tiene una carga importante compartida entre la primera y segunda dimensión (0.49 y 0.45, respectivamente), reforzando su doble vínculo con desempeño y prestigio.

Variables representadas por la tercera dimensión La variable Estudiantes por profesor presenta un  $\cos^2$  muy elevado en la tercera dimensión (0.72), lo que indica que este componente está fuertemente determinado por factores estructurales del entorno educativo. De forma secundaria, Citas por artículo (0.34) también se relaciona con esta dimensión.

Variables asociadas a la cuarta dimensión La variable más representada en la cuarta dimensión es Citas por artículo (0.49), lo que indica que esta dimensión se refiere principalmente al impacto de la producción académica. También tiene cierta influencia la variable Estudiantes por profesor (0.21).

## Agrupación de países con HCPC

Se aplicó el método HCPC (Clustering Jerárquico sobre Componentes Principales) para agrupar los países según sus características en los indicadores QS. El análisis se realizó sobre los resultados del ACP, lo que permite trabajar con una versión reducida y más clara de la información.

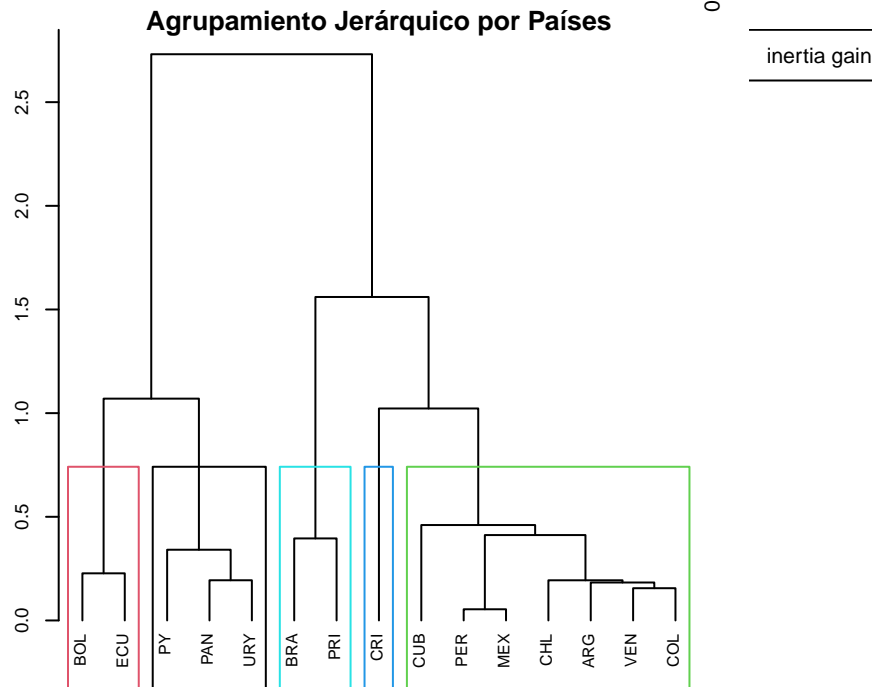
El número de grupos se determinó automáticamente por el método, considerando entre 3 y 6 posibles clústeres, y utilizando un nivel de significancia del 10% para evaluar la estabilidad de las agrupaciones. Esto permitió identificar países con perfiles similares, facilitando su comparación.

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

## Dendograma

```
# 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
)
```

# Hierarchical clustering



Dendrograma del Agrupamiento Jerárquico La gráfica muestra el dendrograma resultante del análisis de agrupamiento jerárquico (HCPC) aplicado a los países según los indicadores QS.

En ella se identifican cinco grupos de países, representados por colores distintos. Estos grupos reúnen países con perfiles similares en términos de calidad educativa y desempeño en los indicadores analizados. Por ejemplo:

- Grupo rojo: incluye a Bolivia y Ecuador, países con características similares que podrían reflejar puntajes más bajos en varios indicadores.
- Grupo verde: agrupa a Chile, Argentina, Venezuela y Colombia, que podrían compartir un perfil intermedio o equilibrado en los distintos factores.

La distancia vertical entre ramas indica el nivel de similitud entre los países: mientras más baja es la unión, mayor es la semejanza. Esto permite visualizar qué países comparten características comunes y cuáles se diferencian claramente del resto.

## Composicion de lo grupos

```
# 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

Composición de los grupos del análisis HCPC La Tabla 5 presenta la composición de los cinco grupos resultantes del análisis de clustering jerárquico con análisis de componentes principales (HCPC), aplicado a los indicadores QS. Cada clúster agrupa países con perfiles similares según sus resultados en los factores principales del ACP.

Grupo 1: Incluye a Panamá, Paraguay y Uruguay. Estos países comparten características intermedias o balanceadas en los indicadores QS.

Grupo 2: Formado por Bolivia y Ecuador, posiblemente los países con los niveles más bajos en varios indicadores, lo que los diferencia claramente del resto.

Grupo 3: Es el grupo más numeroso, con siete países (Argentina, Chile, Colombia, Cuba, México, Perú y Venezuela), lo que sugiere una región con perfiles educativos relativamente parecidos, probablemente con desempeño intermedio-alto en algunos factores.

Grupo 4: Compuesto únicamente por Costa Rica, lo que indica que este país presenta un perfil educativo particular y distintivo frente al resto.

Grupo 5: Incluye a Brasil y Puerto Rico, países que podrían destacar por su puntuación en ciertos indicadores específicos, como reputación académica o producción científica.

## Carateristicas pricipales de los grupos

### Grupo 1

```
desc1 <- as.data.frame(res.hcpc$desc.var$quanti[[1]])
desc1 <- round(desc1, 2)
# 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"))
```

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	2.31	71.51	47.51	13.12	19.47	0.02
QS.Puntuacion	-1.65	50.54	56.92	3.84	7.22	0.10
QS.Docentes.con.doctorado	-1.91	18.53	43.49	4.80	24.49	0.06
QS.PromArticulos.por.docente	-1.98	4.26	25.93	2.90	20.50	0.05
QS.Impacto.Web	-3.10	59.21	77.94	5.18	11.31	0.00

### Análisis de las características principales del Grupo 1

La Tabla 6 presenta las variables cuantitativas más relevantes que caracterizan al Grupo 1, según el análisis de clasificación jerárquica sobre los factores del análisis de componentes principales (ACP) de los indicadores QS. Se incluyen medidas de tendencia central y dispersión, así como el estadístico V-test y el valor p asociado, que indican la significancia de cada variable en la definición del grupo.

Se observa que la variable QS.Estudiantes.por.profesor presenta un V-test positivo (2.31) y un valor p de 0.02, lo que indica que este grupo se distingue por tener una relación estudiantes/profesor significativamente mayor que la media general (71.51 frente a 47.51). Esta característica puede estar asociada a instituciones con menor capacidad docente relativa.

Por el contrario, variables como QS.Puntuación, QS.Docentes.con.doctorado, QS.PromArtículos.por.docente y QS.Impacto.Web muestran V-tests negativos, lo que implica que el Grupo 1 presenta valores inferiores a la media general en estos aspectos. Especialmente relevante es el caso de QS.Impacto.Web, con un V-test de -3.10 y un valor p < 0.01, lo cual sugiere un menor posicionamiento o visibilidad en línea de las instituciones pertenecientes a este grupo.

En resumen, el Grupo 1 se caracteriza principalmente por una mayor cantidad de estudiantes por docente y por desempeños inferiores al promedio en indicadores relacionados con calidad académica, investigación y visibilidad digital, siendo estas diferencias estadísticamente significativas en varios casos.

## Grupo 2

```
desc2 <- as.data.frame(res.hcpc$desc.var$quanti[[2]])
desc2 <- round(desc2, 2)
# Renombrar columnas con nombres más legibles
colnames(desc2) <- c(
  # Nombre de la variable
  "V-test",          # Estadístico v-test
  "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(desc2,
  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	1.97	106.33	79.63	5.67	19.85	0.05
QS.Reputacion.academica	-1.65	45.87	62.05	7.97	14.37	0.10
QS.Puntuacion	-1.99	47.15	56.92	2.35	7.22	0.05
QS.Estudiantes.por.profesor	-2.21	18.15	47.51	12.15	19.47	0.03

### Análisis de las características principales del Grupo 2

La Tabla 7 muestra las variables cuantitativas más destacadas que definen al Grupo 2 en el análisis de clasificación jerárquica sobre componentes principales. Este grupo se compone de países cuyas universidades presentan un perfil diferenciado respecto al promedio general.

En primer lugar, se observa que el indicador QS.Ranking tiene un V-test positivo de 1.97 y un valor p de 0.05, lo que indica que las instituciones de este grupo ocupan posiciones más bajas en el ranking QS (valores numéricos más altos corresponden a peor posición), en comparación con el promedio (106.33 frente a 79.63). Esto sugiere un menor reconocimiento internacional.

Además, los indicadores QS.Reputación académica, QS.Puntuación y QS.Estudiantes por profesor presentan V-tests negativos, lo cual indica que las instituciones de este grupo tienen valores por debajo de la media global en estos aspectos. En particular, QS.Estudiantes por profesor muestra una diferencia significativa (V-test = -2.21,  $p = 0.03$ ), con una media de 18.15 frente a 47.51, lo que refleja una mayor concentración de docentes por estudiante, asociada a mejor atención académica.

En resumen, el Grupo 2 está compuesto por universidades que, a pesar de tener un menor posicionamiento general en el ranking QS, destacan por una proporción docente/estudiante favorable, aunque presentan debilidades en reputación académica y puntuación global.

### Grupo 3

```
desc3 <- as.data.frame(res.hcpc$desc.var$quanti[[3]])
desc3 <- round(desc3, 2)
# Renombrar columnas con nombres más legibles
colnames(desc3) <- 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(desc3,
  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	1.76	69.26	62.05	9.39	14.37	0.08
QS.Citas.por.articulo	-2.79	52.47	69.69	11.45	21.58	0.01

### Análisis de las características principales del Grupo 3

La Tabla 8 presenta los indicadores más representativos del Grupo 3, conformado por países cuyas universidades muestran un perfil mixto en cuanto a reputación e impacto académico.

El indicador QS.Reputación académica presenta un V-test positivo de 1.76, con un valor p de 0.08, lo que sugiere que, en promedio, las universidades de este grupo son mejor valoradas académicamente (69.26) que el promedio general (62.05), aunque esta diferencia no es estadísticamente significativa al nivel del 5%.

Por otro lado, QS.Citas por artículo muestra un V-test negativo y significativo de -2.79 ( $p = 0.01$ ), lo cual indica que estas universidades tienen un menor impacto de sus publicaciones científicas en comparación con el promedio global (52.47 frente a 69.69).

En conjunto, el Grupo 3 se caracteriza por tener universidades con buena reputación académica pero menor visibilidad científica en términos de citas por publicación. Esto podría reflejar instituciones reconocidas por su tradición o calidad docente, pero con menor producción o proyección en investigación.

### Grupo 4

```
desc4 <- as.data.frame(res.hcpc$desc.var$quanti[[4]])
desc4 <- round(desc4, 2)
# Renombrar columnas con nombres más legibles
colnames(desc4) <- 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(desc4,
      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	1.99	71.25	56.92	0	7.22	0.05
QS.Reputacion.academica	1.76	87.30	62.05	0	14.37	0.08
QS.Reputacion.entre.empleadores	1.72	82.70	61.80	0	12.17	0.09
QS.Ranking	-2.07	38.50	79.63	0	19.85	0.04

#### Análisis de las características principales del Grupo 4

La Tabla 9 muestra los indicadores que caracterizan al Grupo 4, conformado por un único país (Costa Rica), cuyas universidades presentan un perfil destacable en la mayoría de los aspectos evaluados.

En primer lugar, se observa que este grupo tiene una puntuación promedio QS significativamente mayor que el promedio global (71.25 vs. 56.92), con un V-test de 1.99 y un valor p de 0.05, indicando una diferencia significativa al 10%.

Además, la reputación académica (87.30) y la reputación entre empleadores (82.70) también superan considerablemente los promedios generales (62.05 y 61.80, respectivamente), con V-test de 1.76 y 1.72, y valores p de 0.08 y 0.09. Aunque no alcanzan significancia al 5%, sí son relevantes al nivel del 10%, evidenciando una percepción favorable tanto en el ámbito académico como laboral.

Finalmente, el indicador QS.Ranking muestra un V-test negativo de -2.07 ( $p = 0.04$ ), lo que indica que este país ocupa una mejor posición en el ranking (es decir, un valor numérico más bajo) que el promedio de los países del estudio (38.50 vs. 79.63), y esta diferencia es significativa incluso al nivel del 5%.

En conjunto, el Grupo 4 representa un perfil sobresaliente, con universidades reconocidas por su calidad, tanto en evaluación global como en reputación académica y profesional, y que además ocupan posiciones destacadas en los rankings internacionales.

#### Grupo 5

```

desc5 <- as.data.frame(res.hcpc$desc.var$quanti[[5]])
desc5 <- round(desc5, 2)
# Renombrar columnas con nombres más legibles
colnames(desc5) <- 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(desc5,
      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	3.16	96.18	43.49	3.82	24.49	0.00
QS.PromArticulos.por.docente	2.59	62.10	25.93	10.70	20.50	0.01
QS.Reputacion.entre.empleadores	-1.89	46.15	61.80	0.15	12.17	0.06

### Análisis de las características principales del Grupo 5

La Tabla 10 describe los indicadores más representativos del Grupo 5, conformado por Brasil y Puerto Rico. Este grupo se caracteriza por un perfil académico fuertemente vinculado a la formación investigativa y la productividad científica.

El indicador más destacable es la proporción de docentes con doctorado, cuya media en el grupo (96.18%) es más del doble que el promedio global (43.49%). Esta diferencia es altamente significativa (V-test = 3.16,  $p = 0.00$ ), lo que refleja un alto nivel de formación académica en las instituciones de estos países.

Además, el grupo también presenta una mayor producción científica por docente, con un promedio de 62.10 artículos, frente a 25.93 a nivel general. Esta diferencia es significativa al 5% (V-test = 2.59,  $p = 0.01$ ), lo que reafirma el perfil de investigación intensiva de este grupo.

En contraste, se observa que la reputación entre empleadores es menor en comparación con el promedio (46.15 vs. 61.80), con un V-test de -1.89 y un valor  $p$  de 0.06. Aunque no es una diferencia significativa al 5%, sí lo es al 10%, lo que sugiere que, pese a su fortaleza académica y científica, estas universidades podrían mejorar su posicionamiento en el mercado laboral.

En resumen, el Grupo 5 se distingue por contar con instituciones altamente calificadas en términos de formación doctoral y productividad académica, aunque con una reputación entre empleadores relativamente más baja que el promedio regional.

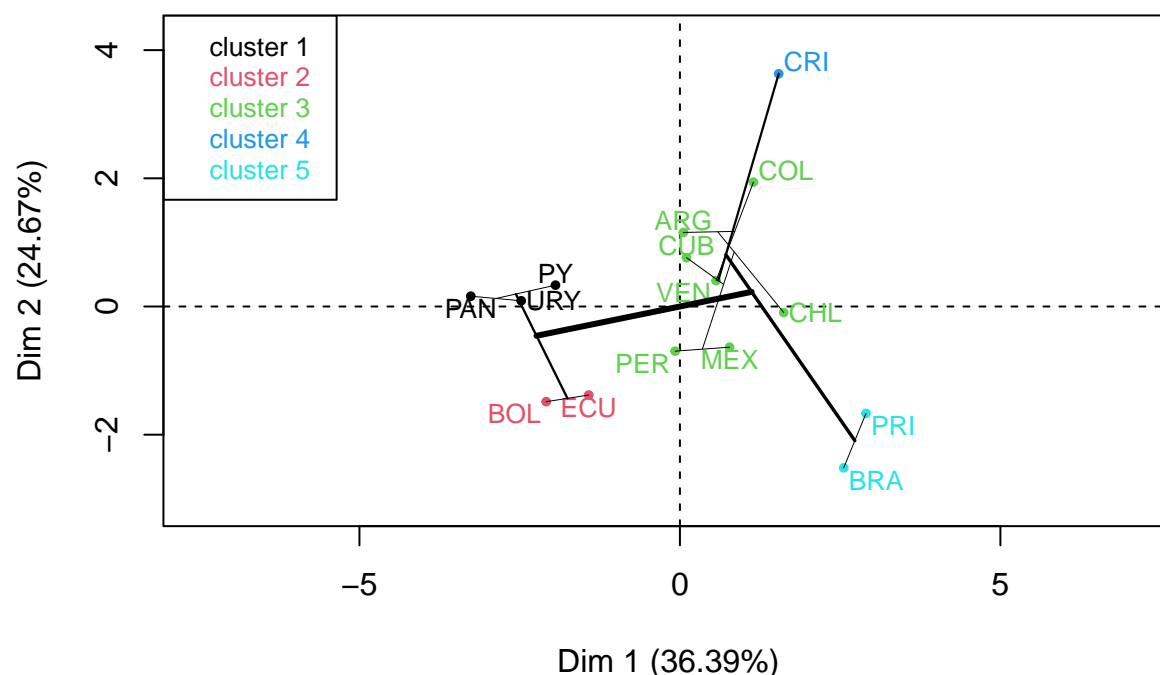
### Distribución de los grupos en el plano factorial (ACP sobre indicadores QS)

```

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)



Distribución de los grupos en el plano factorial (ACP sobre indicadores QS)

La Figura muestra la proyección de los países sobre los dos primeros componentes principales del Análisis de Componentes Principales (ACP), que explican conjuntamente un 61.06% de la variabilidad total de los datos (Dimensión 1: 36.39%, Dimensión 2: 24.67%). Esta visualización permite identificar claramente la separación de los grupos definidos mediante el análisis de conglomerados (clusters), lo que respalda la validez de la clasificación obtenida.

Los países agrupados en el Grupo 5 (Brasil y Puerto Rico, en color celeste) se ubican en el extremo derecho del plano, lo que indica altos valores en los indicadores relacionados con la producción académica y formación de alto nivel, como el porcentaje de docentes con doctorado y el número de artículos por docente.

En contraste, los países del Grupo 2 (Bolivia y Ecuador, en rosado) se sitúan en la región izquierda del gráfico, asociados a valores más bajos en los indicadores QS, particularmente en reputación y producción científica.

El Grupo 3 (verde), conformado por Argentina, Colombia, Cuba, Chile, México y Perú, se concentra cerca del centro del plano, lo cual sugiere un perfil intermedio o equilibrado en cuanto a desempeño QS. Esta posición cercana al origen implica que estos países no presentan valores extremos en las dimensiones analizadas.

Por su parte, el Grupo 4 (Costa Rica, en azul oscuro), aparece claramente separado hacia la parte superior del gráfico, reflejando puntuaciones elevadas en los indicadores de reputación académica y entre empleadores, así como en la puntuación general QS.

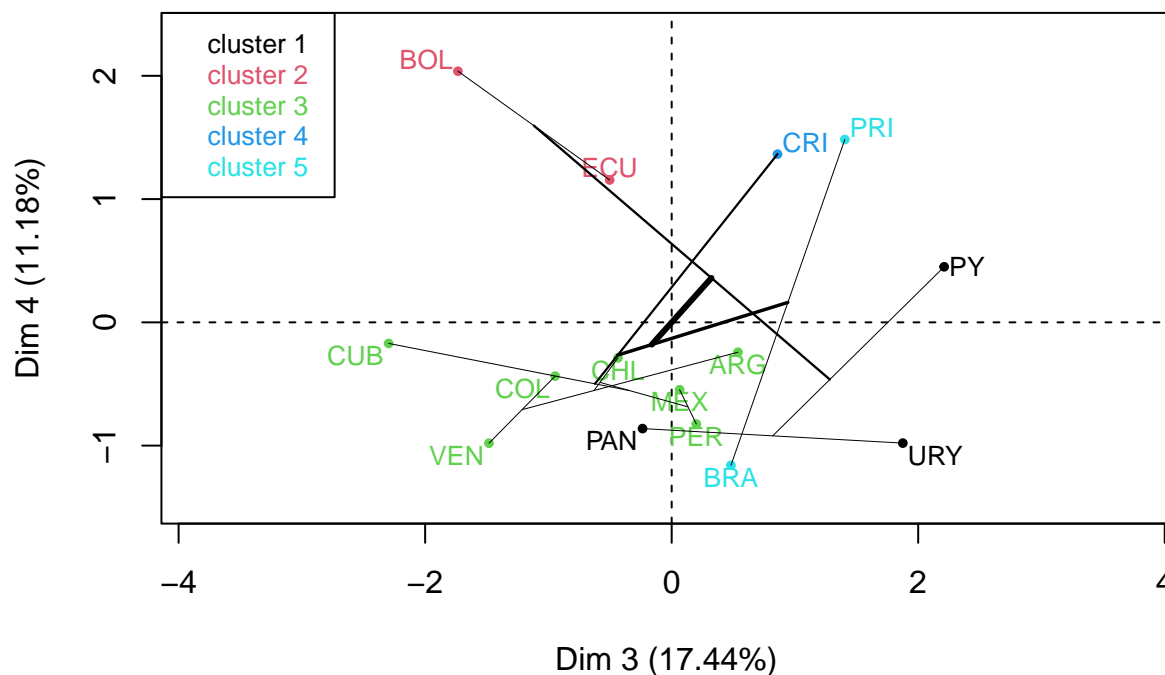
Finalmente, el Grupo 1 (Panamá, Paraguay y Uruguay, en negro) se ubica en el cuadrante inferior izquierdo del plano, indicando valores bajos o moderados en la mayoría de los indicadores QS, sin destacar significativamente en ningún eje principal.

En resumen, el gráfico evidencia una buena discriminación entre los grupos identificados, lo que refuerza la utilidad del ACP como herramienta exploratoria complementaria al análisis de conglomerados. Además,

ofrece una representación visual clara del posicionamiento relativo de los países latinoamericanos según su desempeño en los 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)



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

La Figura presenta la distribución de los países latinoamericanos sobre el plano definido por la tercera y cuarta dimensiones principales del Análisis de Componentes Principales (ACP) aplicado a los indicadores QS. Estas dimensiones explican conjuntamente un 28.62% de la variabilidad total de los datos (Dimensión 3: 17.44%, Dimensión 4: 11.18%).

Aunque estas dimensiones adicionales no capturan tanta varianza como las dos primeras, resultan útiles para observar patrones más sutiles en la diferenciación de los grupos. En este plano, también se visualiza una buena separación entre algunos clusters, lo cual refuerza la validez de la segmentación.

El Grupo 2 (Bolivia y Ecuador, en rosado) se mantiene bien diferenciado, posicionándose en el cuadrante superior izquierdo, lo cual sugiere que presentan valores particulares en las dimensiones 3 y 4, posiblemente relacionados con indicadores específicos como la percepción entre empleadores o baja visibilidad internacional.

El Grupo 5 (Brasil y Puerto Rico, en celeste) vuelve a destacarse, aunque ahora aparecen más dispersos a lo largo del eje horizontal. Puerto Rico se posiciona hacia la parte superior derecha, mientras Brasil se ubica hacia el extremo inferior derecho. Esto podría estar reflejando diferencias internas dentro del grupo en cuanto a aspectos secundarios de calidad académica o visibilidad.

El Grupo 3 (Argentina, Colombia, Chile, Cuba, México, Perú y Venezuela, en verde) se concentra mayoritariamente en la región central izquierda del gráfico, con una ligera dispersión en el eje vertical. Esta distribución confirma nuevamente que estos países mantienen un perfil intermedio y relativamente homogéneo en las dimensiones secundarias analizadas.

El Grupo 4 (Costa Rica, en azul oscuro) aparece más alejado del centro, ubicado en el cuadrante superior derecho. Este posicionamiento sugiere un comportamiento atípico o destacable del país en ciertos indicadores que no fueron tan visibles en las dos primeras dimensiones, reafirmando su singularidad en el contexto regional.

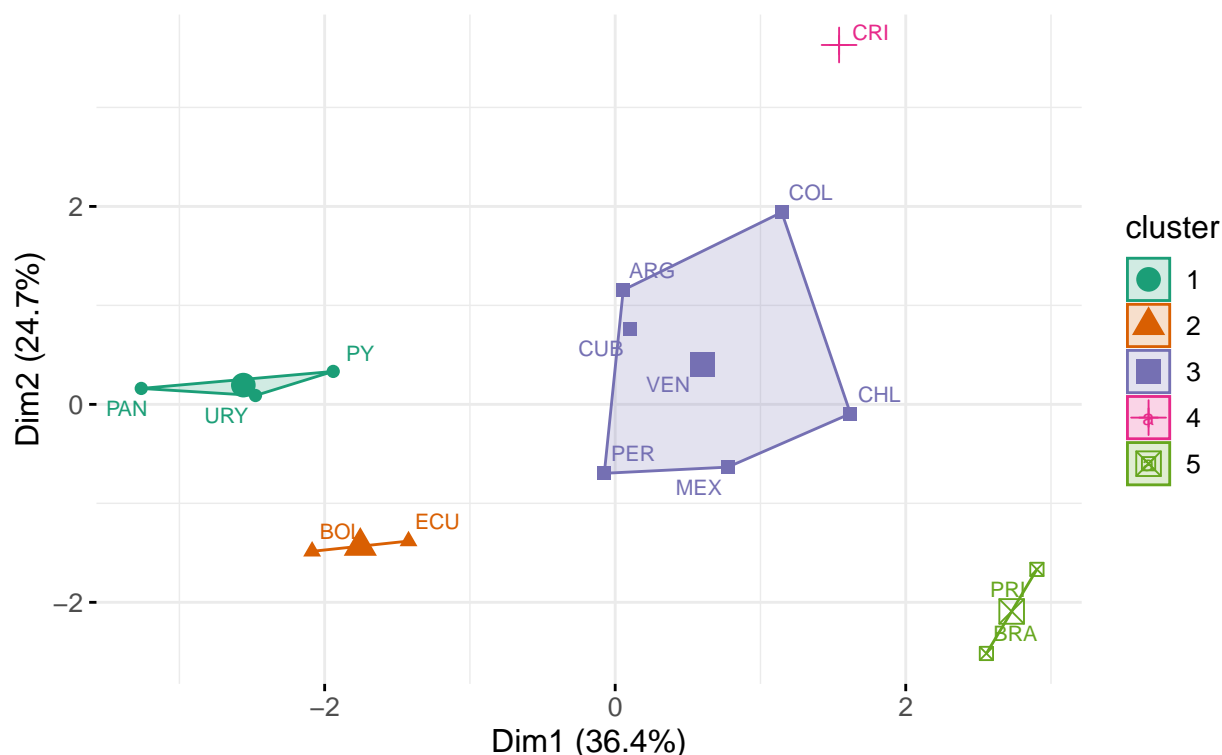
El Grupo 1 (Panamá, Paraguay y Uruguay, en negro) muestra una mayor dispersión en este plano. Paraguay y Uruguay se ubican hacia el extremo derecho del gráfico, mientras Panamá se encuentra más cerca del centro. Esto podría indicar diferencias más marcadas dentro del grupo en las dimensiones analizadas, especialmente en aquellos indicadores que explican menor varianza global pero que siguen siendo relevantes para la diferenciación.

En conjunto, esta proyección complementa la información observada en el plano 1-2, revelando matices adicionales en el perfil académico y de reputación QS de los países analizados. También permite observar qué países, dentro de un mismo grupo, presentan mayores similitudes o diferencias en los componentes menos dominantes del análisis.

## Distribución de los grupos en el plano factorial

```
fviz_cluster(res.hcpc,
             geom = c("point", "text"),      # Puntos y texto
             pointsize = 2,
             labelsiz = 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)



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

La figura presenta la proyección de los países latinoamericanos en el plano definido por las dos primeras componentes principales obtenidas del Análisis de Componentes Principales (ACP) sobre los indicadores QS. Este plano capta una porción significativa de la variabilidad total (61.1%), con la Dimensión 1 explicando el 36.4% y la Dimensión 2 el 24.7%. Esta representación permite observar de manera clara la agrupación y diferenciación de los países en función de los indicadores considerados.

El Grupo 5 (Brasil y Puerto Rico, en verde claro con bordes) se destaca por estar claramente separado del resto, ocupando el extremo derecho del gráfico en ambas dimensiones. Esta ubicación sugiere que estos países presentan los valores más altos en los indicadores más representativos del primer componente, posiblemente relacionados con reputación académica, empleabilidad o citación internacional.

El Grupo 4 (Costa Rica, en rosado) se encuentra aislado en el cuadrante superior derecho. Su posición tan alejada del resto indica un perfil único entre los países analizados. Costa Rica probablemente sobresale en algunos indicadores específicos que combinan alta percepción académica y empleabilidad, lo que justifica su agrupación independiente.

El Grupo 3 (en azul violeta), compuesto por países como Argentina, Colombia, Chile, México, Venezuela, Perú, Cuba y en parte Uruguay, se concentra en la zona central del gráfico, formando un polígono más compacto. Esta distribución sugiere que estos países tienen un desempeño intermedio y homogéneo en los principales indicadores QS, sin presentar extremos muy altos ni bajos.

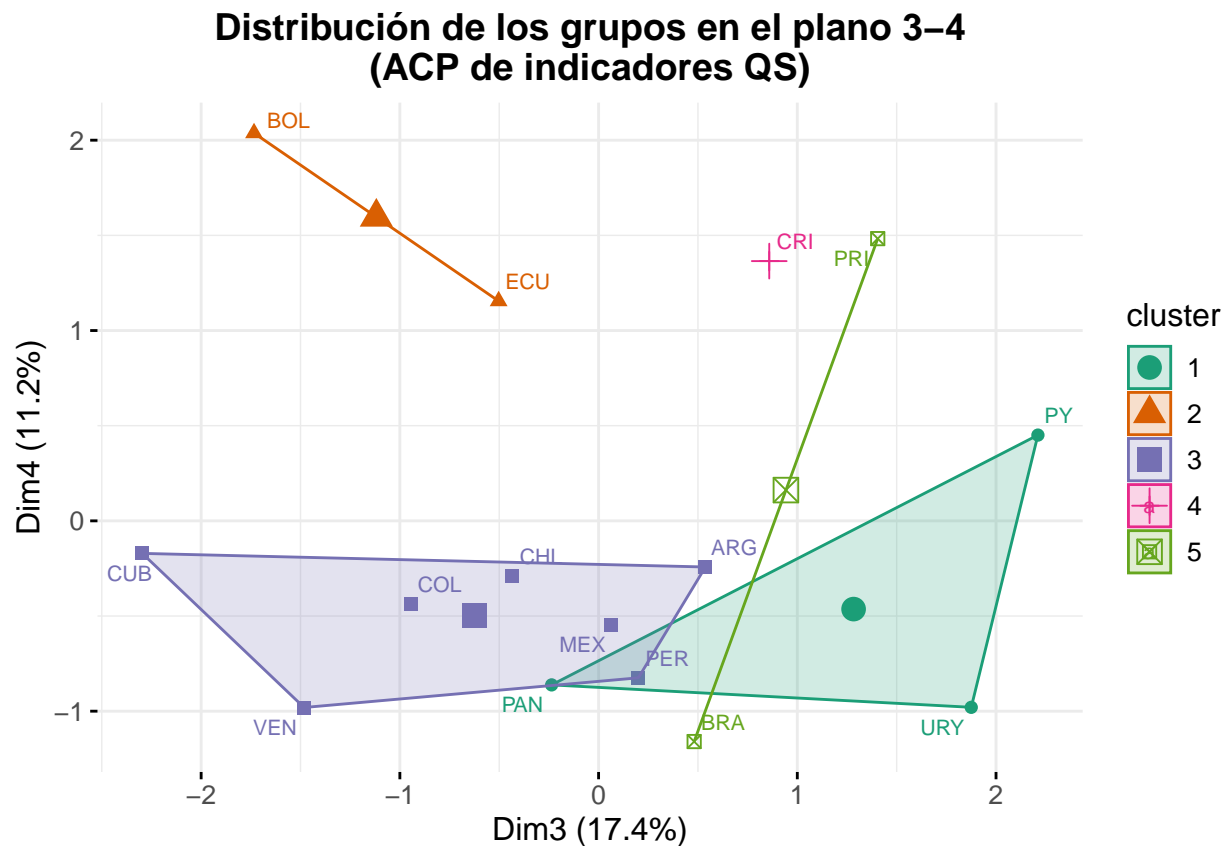
El Grupo 2 (Bolivia y Ecuador, en naranja) se posiciona en la parte inferior izquierda del gráfico, indicando bajos valores en los componentes principales. Esta ubicación podría reflejar menores niveles de reputación internacional, visibilidad académica o citación científica.

El Grupo 1 (Paraguay, Uruguay y Panamá, en verde oscuro) aparece en el cuadrante izquierdo, más cercano al origen y ligeramente disperso. Aunque relativamente cercanos entre sí, presentan un perfil algo más moderado

respecto a los indicadores QS, sin destacar especialmente en los factores principales pero manteniéndose diferenciados de los extremos.

En conjunto, esta representación permite observar con claridad la diferenciación estructural entre los grupos formados. Los clústeres reflejan tanto similitudes como contrastes importantes entre los países, revelando una segmentación coherente basada en características académicas y de percepción internacional.

```
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)
  )
)
```



## **Punto 7**

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