

# Proyecto AED: Análisis de la Calidad de Vida en España

Daniela Meriño, Víctor Mateu, Antonio Nova

2024-12-31

## Introducción

La calidad de vida es un concepto multidimensional que abarca aspectos económicos, sociales, ambientales y subjetivos, reflejando el bienestar general de las personas en una sociedad. En este trabajo, analizaremos la calidad de vida en España, desglosada en diversas dimensiones, a lo largo de varias comunidades autónomas y un periodo de años.

El análisis se centra en múltiples dimensiones de la calidad de vida, como las condiciones materiales, la salud, la educación, el trabajo, las relaciones sociales, la seguridad, el entorno ambiental y la gobernanza, entre otras. Además, el enfoque temporal nos permitirá evaluar tendencias, identificar patrones y observar disparidades entre las diferentes comunidades autónomas.

Los datos utilizados para este estudio provienen de un conjunto estructurado que organiza información clave de calidad de vida por año, comunidad autónoma y dimensión. Este formato permitirá realizar comparaciones entre las regiones y temporales, utilizando herramientas estadísticas y visualizaciones para descubrir las relaciones y diferencias significativas entre las regiones de España.

## Importación y Ordenación de los datos

En este apartado, se procede a la importación y preparación de los datos proporcionados por el Instituto Nacional de Estadística (INE) de España. Los datos corresponden al **Indicador Multidimensional de Calidad de Vida (IMCV)**. Este indicador tiene como objetivo ofrecer una visión integral del bienestar en las distintas comunidades autónomas mediante la evaluación de múltiples dimensiones, como salud, educación, trabajo, entre otras.

Nuestro conjunto de datos se organiza en nueve dimensiones. Estas abarcan aspectos clave como condiciones materiales de vida (renta, pobreza y seguridad económica), trabajo (empleo y calidad laboral), salud (esperanza de vida y determinantes como ejercicio y tabaquismo), educación (formación y abandono escolar), ocio y relaciones sociales (satisfacción y apoyo social), seguridad física (criminalidad y percepción de seguridad), gobernanza (confianza en instituciones y participación política), entorno (contaminación y acceso a zonas verdes) y experiencia de vida (satisfacción y propósito vital). Juntas, estas dimensiones ofrecen una visión del bienestar en las diferentes comunidades autónomas de España.

La naturaleza experimental de estos datos implica que están en proceso de consolidación metodológica y pueden incluir innovaciones en las fuentes de información. En este sentido, el primer paso es garantizar que los datos se encuentren correctamente estructurados y listos para su análisis.

En primer lugar, destacar que de este conjunto de datos utilizada tiene un número muy elevado de variables por lo que nosotros nos centraremos en los resultados globales de cada dimensión por comunidad autónoma por año.

Aunque, para entender un poco mejor de cómo se han obtenido los datos del conjunto que vamos a analizar, destacar que el cálculo del **Indicador Multidimensional de Calidad de Vida (IMCV)** se basa en la selección de indicadores representativos de nueve dimensiones clave, como condiciones materiales de vida,

salud, educación y trabajo, obtenidos de fuentes estadísticas como la Encuesta de Condiciones de Vida. Los valores de cada indicador se normalizan para permitir comparaciones homogéneas mediante el método Mazziotta-Pareto Ajustado (AMPI), que reescala los datos en un rango común y penaliza la variabilidad excesiva. Posteriormente, se agregan los valores normalizados de cada dimensión en un índice representativo, y finalmente, los índices de todas las dimensiones se combinan en un único valor global mediante una media ponderada. El resultado permite analizar la evolución temporal y las diferencias territoriales en la calidad de vida en España.

Una vez tenemos una idea básica sobre el dataset que vamos a analizar, nos centraremos en la importación y ordenación de este.

El fichero descargado a través del INE se obtuvo en formato .xlsx, por lo que la importación se tuvo que realizar a través de la función `read_excel` de la librería `readxl`. Se creó un código en R de manera que cumpla las condiciones de un conjunto de datos considerado *tidy*. Esto requiere que cada variable esté representada en una columna, cada observación de una variable está en una fila distinta, la primera fila incluye los nombres de las variables y los nombres deben ser representativos de las variables que almacenan.

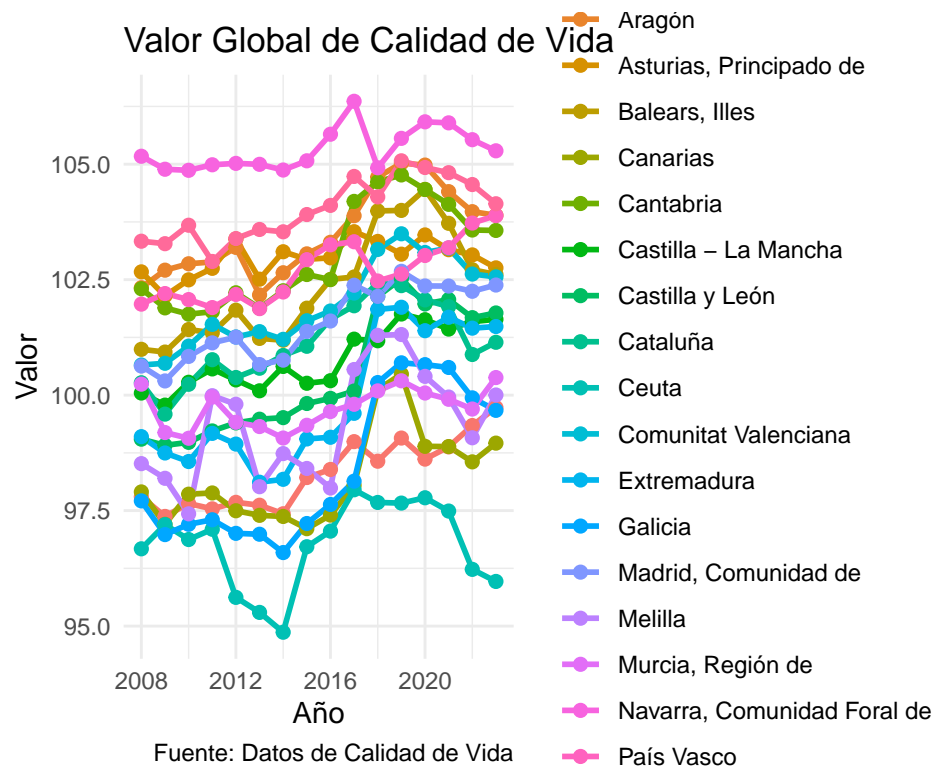
```
##           CCAA Anio  Dim      Valor
## 1 Andalucía 2008 dim1  98.19206
## 2 Andalucía 2008 dim2  95.79060
## 3 Andalucía 2008 dim3  97.64004
## 4 Andalucía 2008 dim4  96.31249
## 5 Andalucía 2008 dim5  98.45372
## 6 Andalucía 2008 dim6 100.29773
```

## Análisis de los datos

El análisis de datos es un proceso fundamental para extraer información relevante, identificar patrones y apoyar la toma de decisiones basada en evidencia. En este caso, los datos proporcionan información multidimensional que permite explorar la calidad de vida en diferentes comunidades autónomas a lo largo del tiempo. Este análisis combina técnicas descriptivas, como visualizaciones y resúmenes estadísticos, con enfoques más avanzados, como modelos estadísticos y exploraciones multivariantes.

En primer lugar, notamos que no hay valores faltantes en el conjunto de datos, sin embargo sí se observó que ciertos valores para diferentes dimensiones se repetían durante varios años. Investigando esto se observó que no todos los indicadores están disponibles todos los años. Por ejemplo, el caso más extremo es el de la dimensión Gobernanza y Derechos Básicos para el que solo hay datos en el año 2013 y, parcialmente para 2022. Los valores para los otros años, que ya venían en el conjunto de datos descargados del INE, simplemente fueron repetidos el resto de años posteriores a la nueva adquisición de datos en 2023. También se puede observar la imputación de estos valores para las dimensiones Ocio y relaciones sociales y Experiencia general de la vida.

Dado el volumen de datos disponibles, procedemos a simplificar el análisis calculando un promedio de todas las dimensiones para obtener un valor global de calidad de vida por comunidad autónoma y por año. Este enfoque nos permite mantener la representatividad de todas las dimensiones sin perder información relevante. Una vez calculados los valores globales, clasificaremos las comunidades autónomas en clústeres que representen tres niveles: clase baja, media y alta, según su calidad de vida. Esto nos permitirá trabajar con menos datos de forma más manejable, conservando la esencia de la información para identificar patrones y realizar comparaciones significativas.



```
dim_all <- data %>%
  filter(Dim == "dim4")

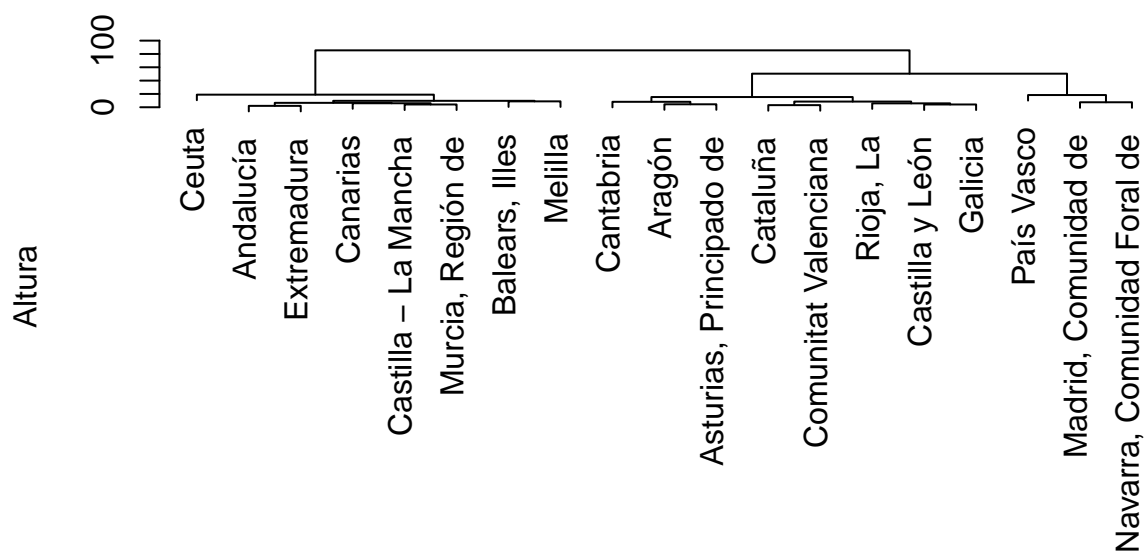
mat_dim <- dim_all %>%
  select(CCAA, Anio, Valor) %>%
  pivot_wider(names_from = Anio, values_from = Valor) %>%
  column_to_rownames("CCAA")

distancia <- dist(mat_dim, method = "euclidean")

cluster_jerarquico <- hclust(distancia, method = "ward.D2")

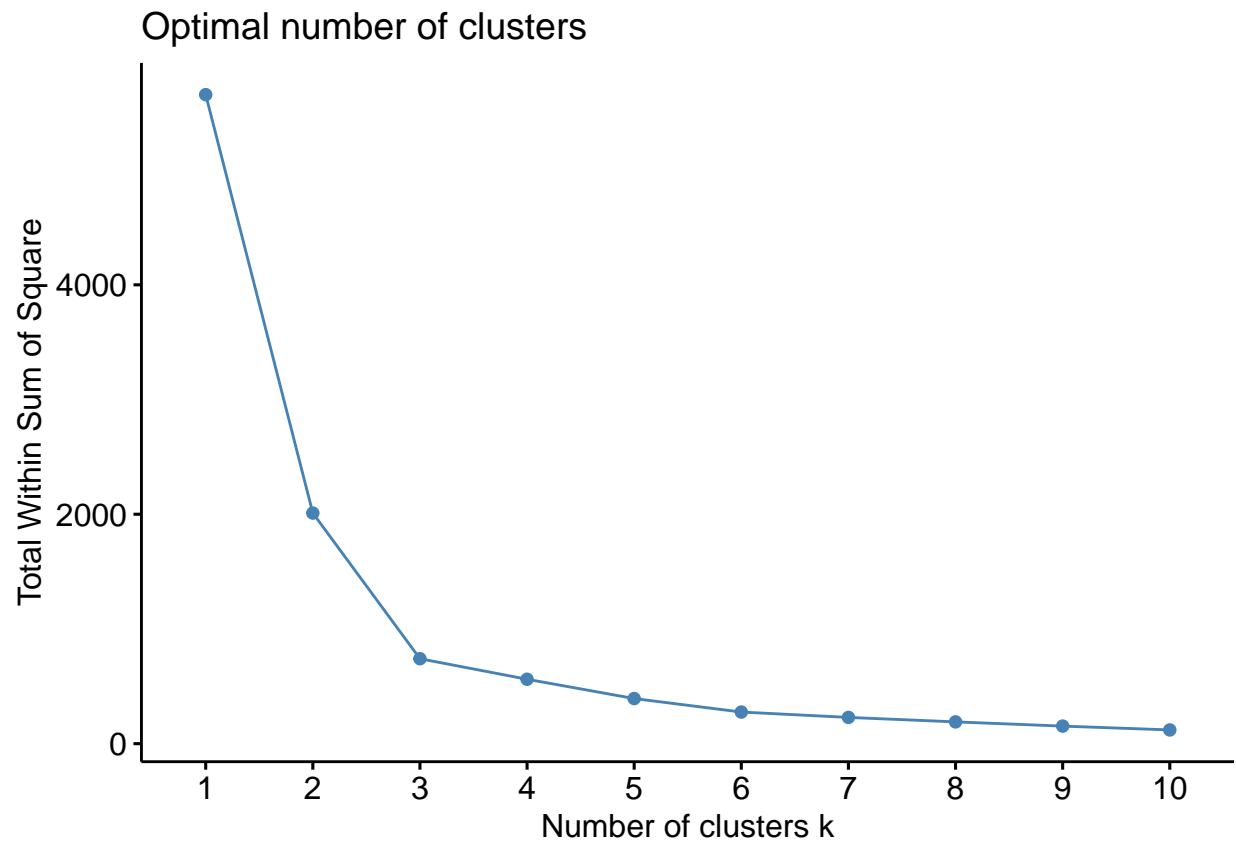
plot(cluster_jerarquico, main = "Clustering Jerárquico de las CCAA", xlab = "", ylab = "Altura")
```

## Clustering Jerárquico de las CCAA

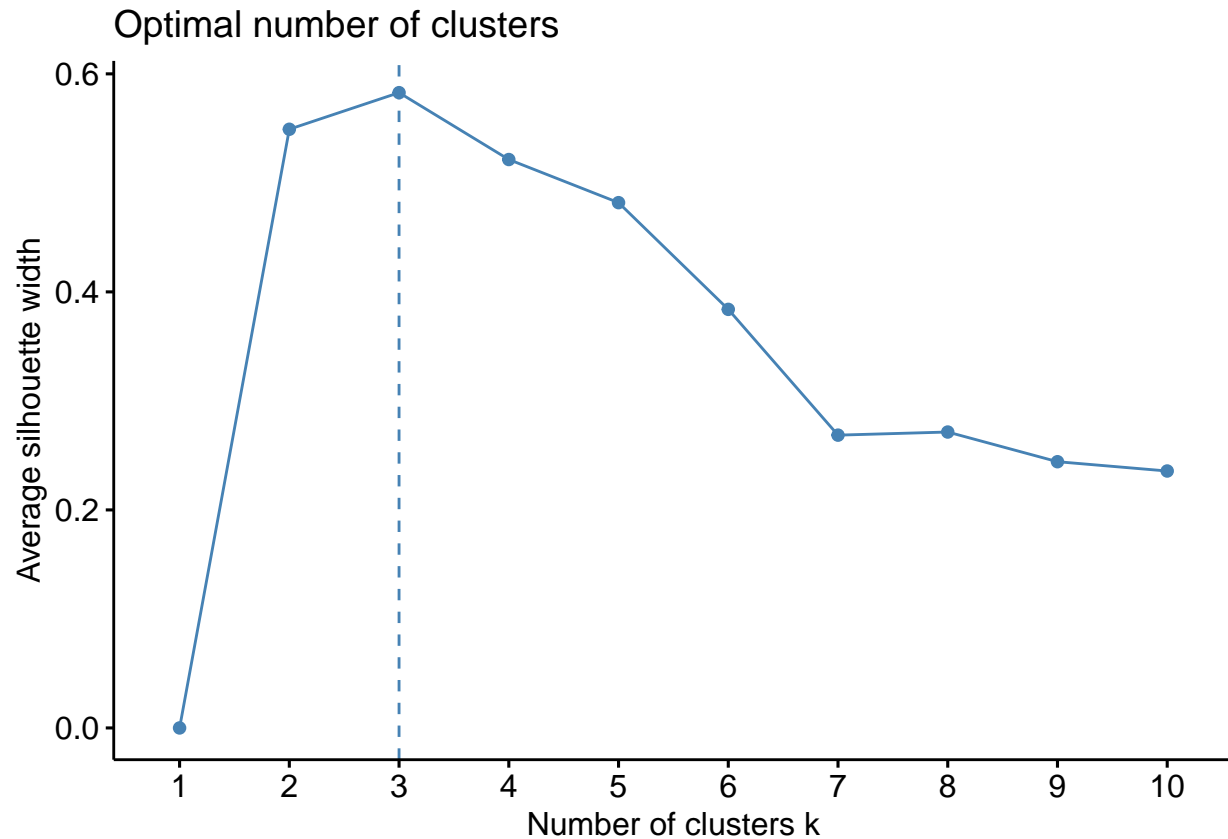


`hclust (*, "ward.D2")`

```
fviz_nbclust(mat_dim, hcut, method = "wss")
```



```
fviz_nbclust(mat_dim, hcut, method = "silhouette")
```



```

grupos <- cutree(cluster_jerarquico, k = 3)

dim_all <- dim_all %>%
  left_join(data.frame(CCAA = rownames(mat_dim), Cluster = grupos), by = "CCAA")

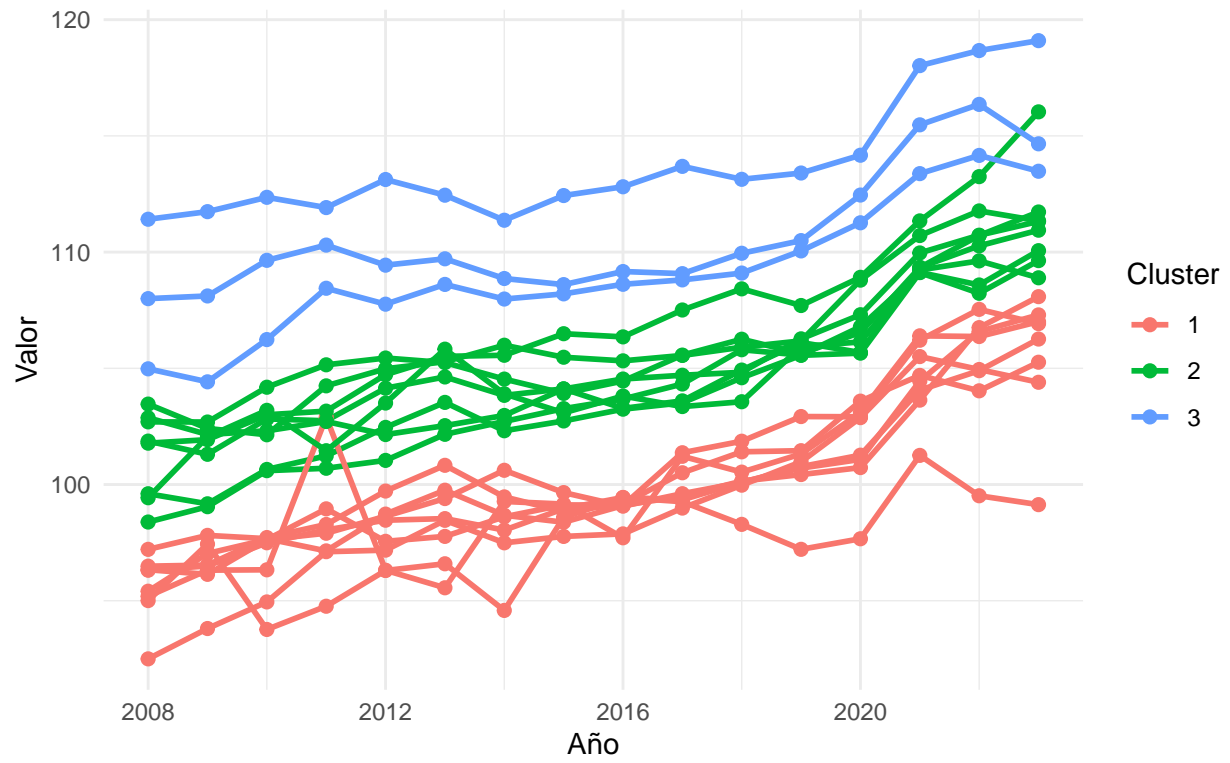
dim_all <- dim_all %>%
  arrange(Cluster, CCAA) # Ordenar primero por cluster y luego por CCAA

plot_cluster <- ggplot(dim_all, aes(x = Anio, y = Valor, color = factor(Cluster), group = CCAA)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Clustering de Entorno y Medioambiente",
    x = "Año",
    y = "Valor",
    color = "Cluster",
    caption = "Fuente: Datos de Calidad de Vida"
  ) +
  theme_minimal()

print(plot_cluster)

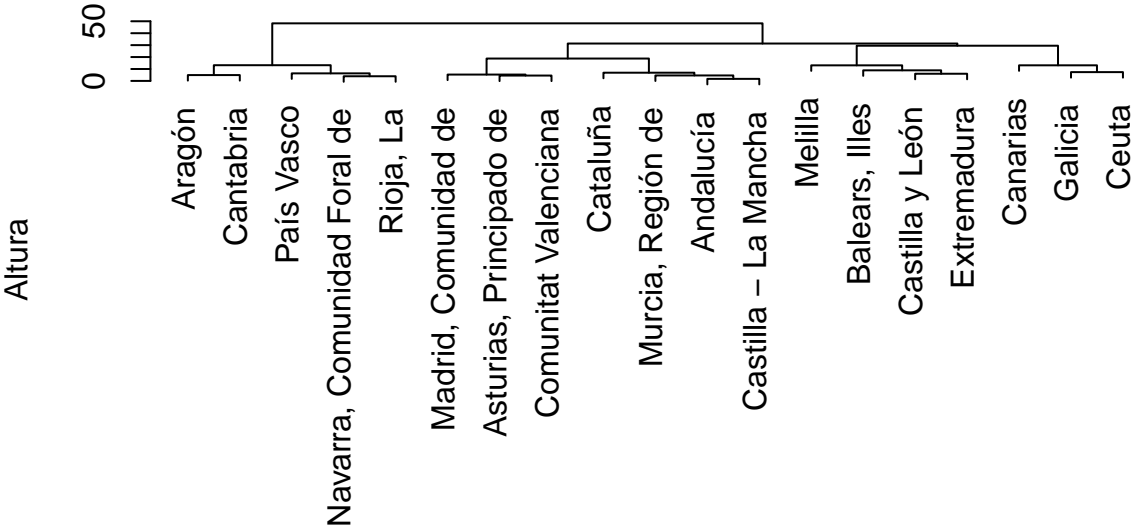
```

## Clustering de Entorno y Medioambiente



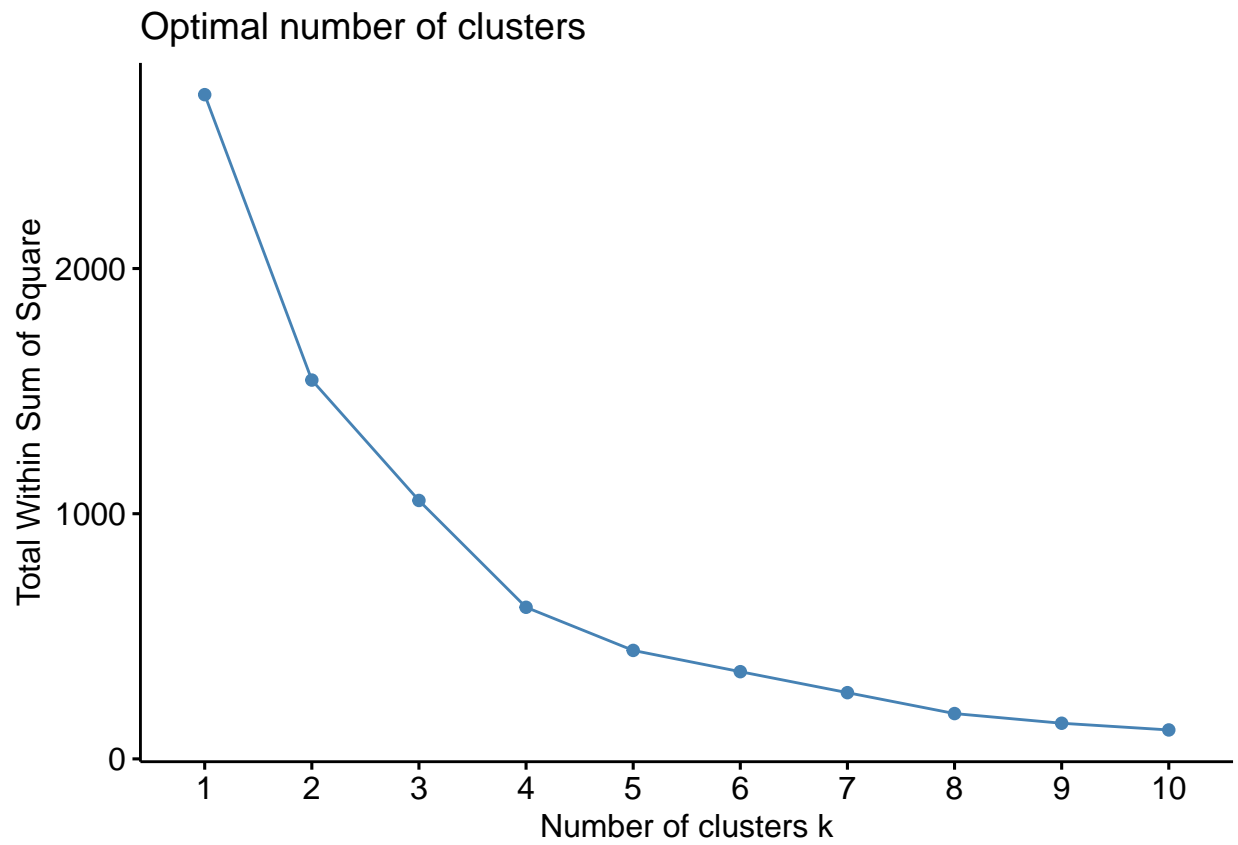
Fuente: Datos de Calidad de Vida

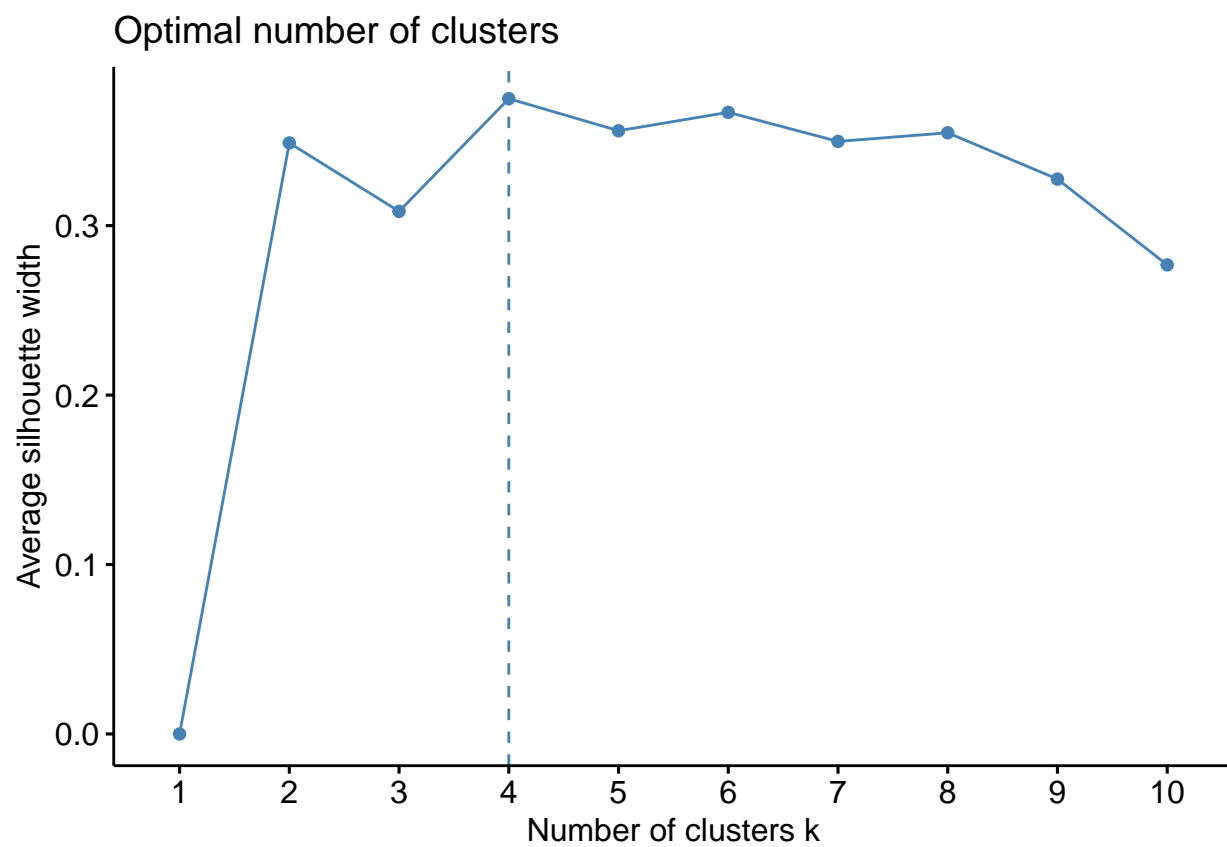
Clustering Jerárquico de las CCAA

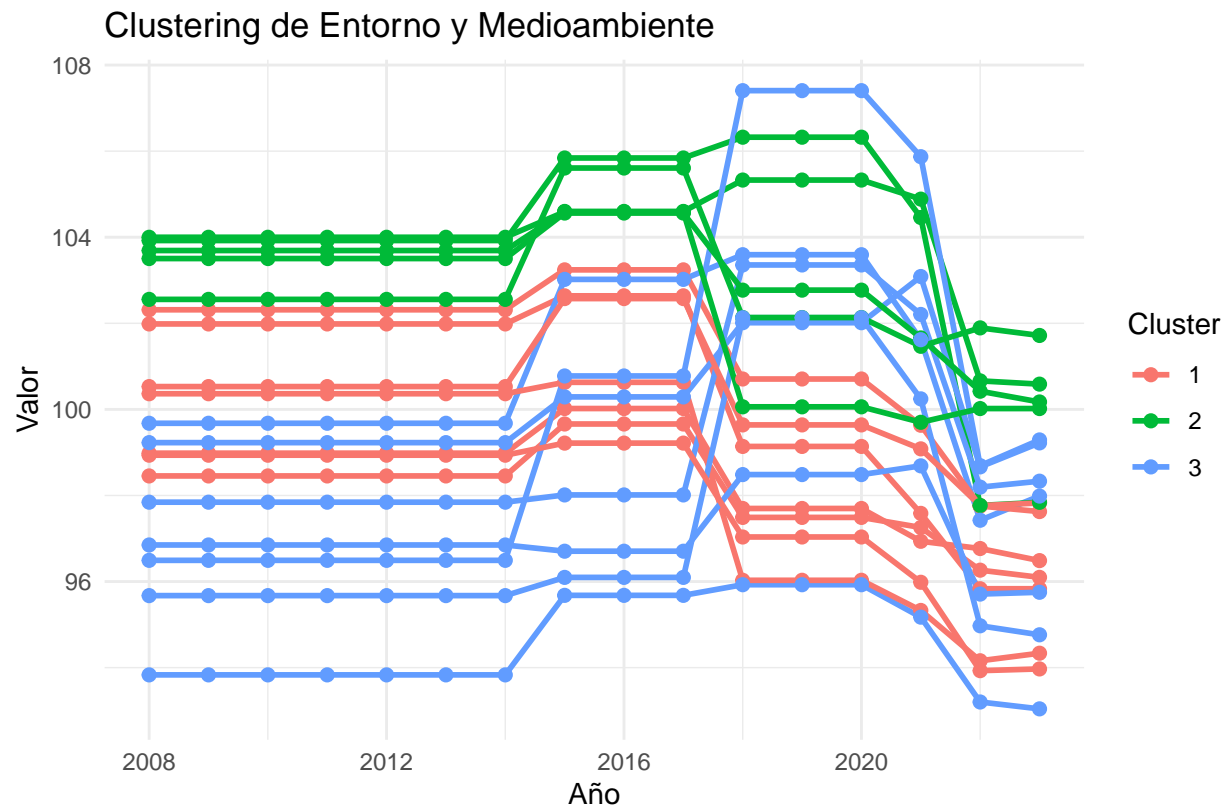


hclust (\*, "ward.D2")









Fuente: Datos de Calidad de Vida

```
dim_all <- data %>%
  filter(Dim == "dim6")

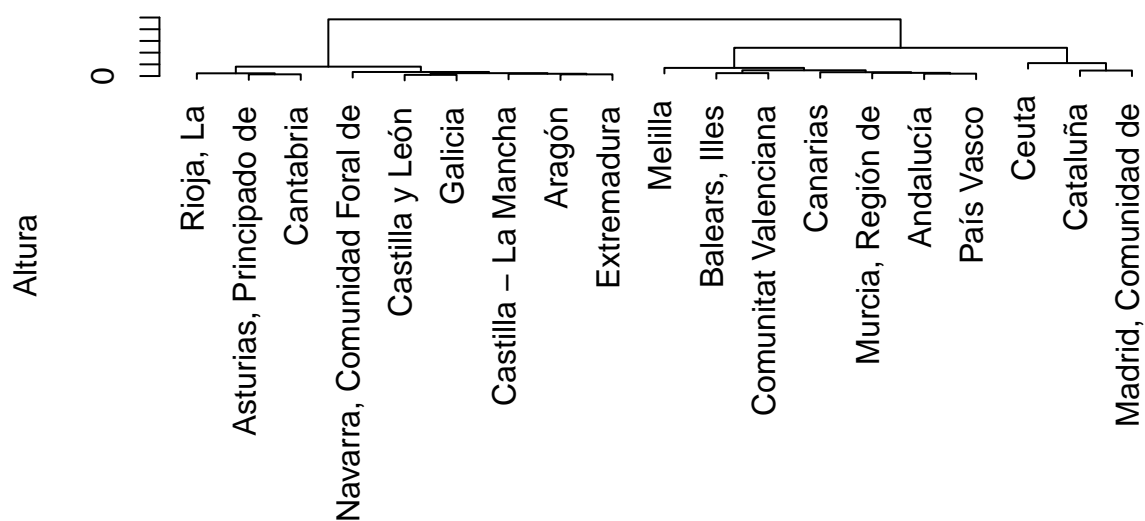
mat_dim <- dim_all %>%
  select(CCAA, Anio, Valor) %>%
  pivot_wider(names_from = Anio, values_from = Valor) %>%
  column_to_rownames("CCAA")

distancia <- dist(mat_dim, method = "euclidean")

cluster_jerarquico <- hclust(distancia, method = "ward.D2")

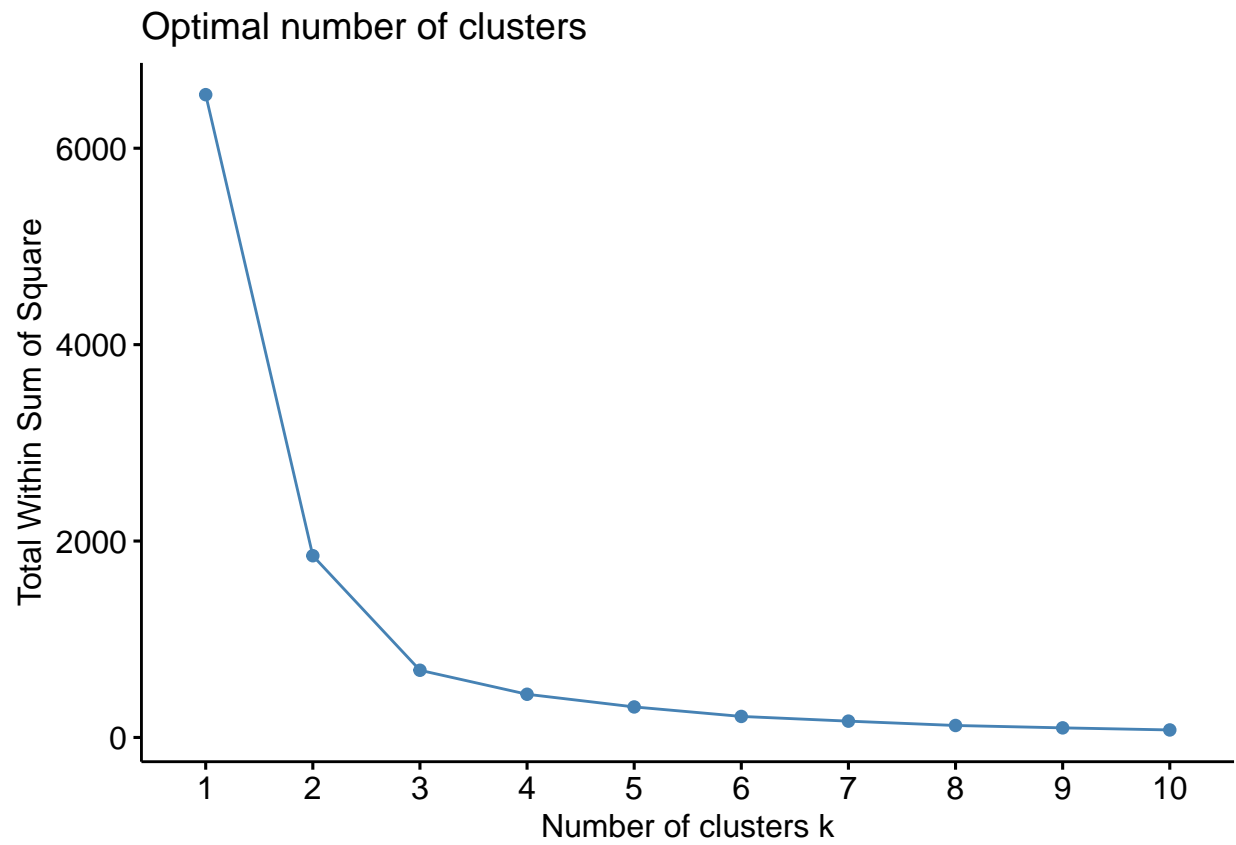
plot(cluster_jerarquico, main = "Clustering Jerárquico de las CCAA", xlab = "", ylab = "Altura")
```

## Clustering Jerárquico de las CCAA

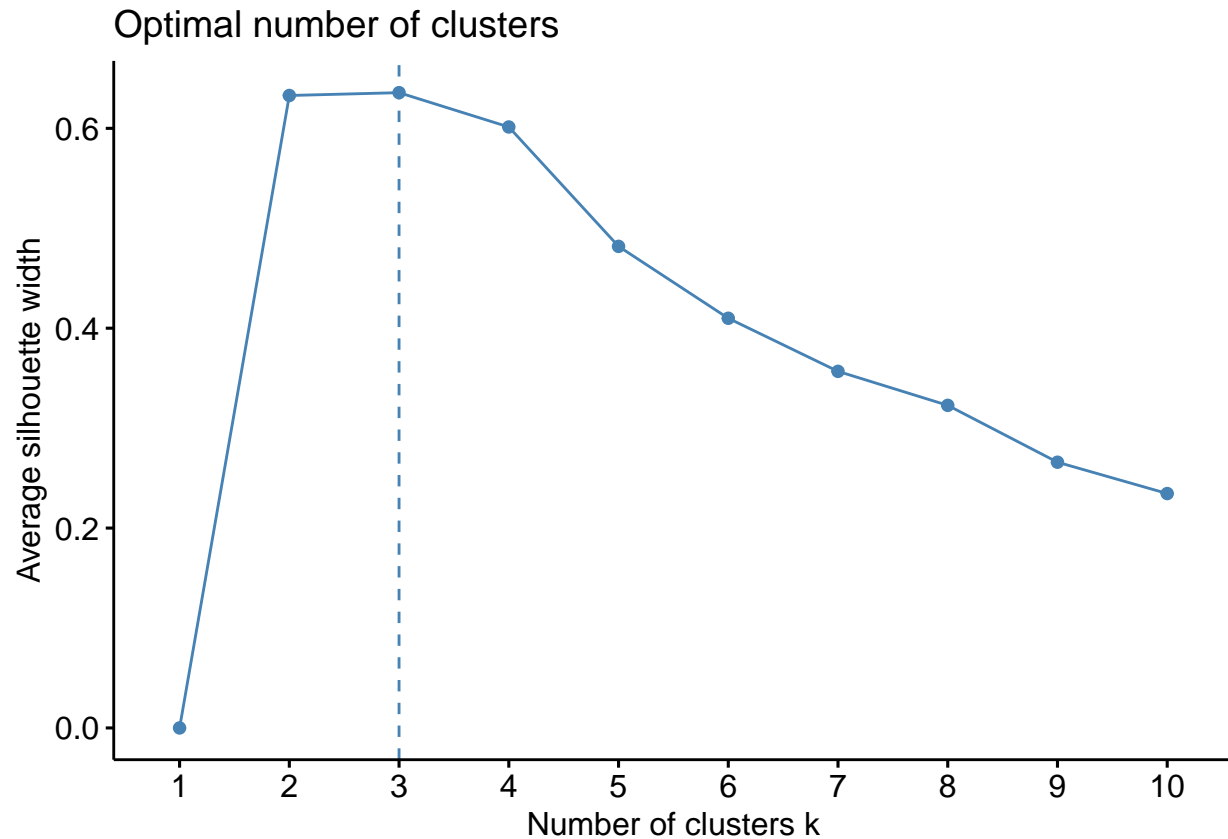


`hclust (*, "ward.D2")`

```
fviz_nbclust(mat_dim, hcut, method = "wss")
```



```
fviz_nbclust(mat_dim, hcut, method = "silhouette")
```



```

grupos <- cutree(cluster_jerarquico, k = 3)

dim_all <- dim_all %>%
  left_join(data.frame(CCAA = rownames(mat_dim), Cluster = grupos), by = "CCAA")

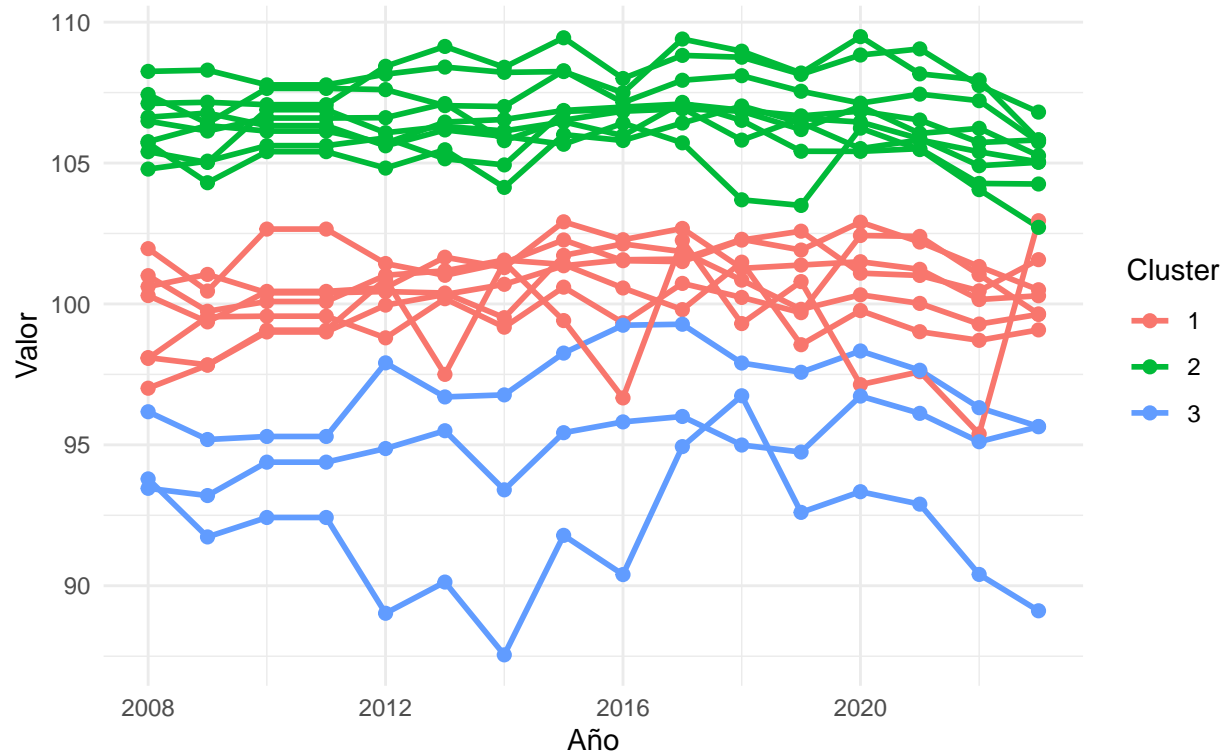
dim_all <- dim_all %>%
  arrange(Cluster, CCAA) # Ordenar primero por cluster y luego por CCAA

plot_cluster <- ggplot(dim_all, aes(x = Anio, y = Valor, color = factor(Cluster), group = CCAA)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Clustering de Entorno y Medioambiente",
    x = "Año",
    y = "Valor",
    color = "Cluster",
    caption = "Fuente: Datos de Calidad de Vida"
  ) +
  theme_minimal()

print(plot_cluster)

```

## Clustering de Entorno y Medioambiente



Fuente: Datos de Calidad de Vida

```
dim_all <- data %>%
  filter(Dim == "dim7")

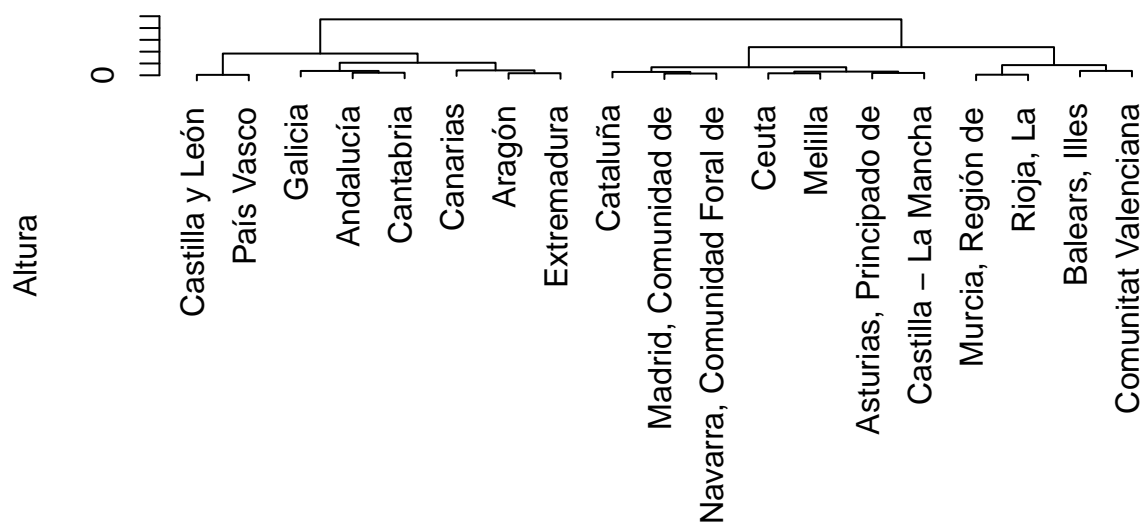
mat_dim <- dim_all %>%
  select(CCAA, Anio, Valor) %>%
  pivot_wider(names_from = Anio, values_from = Valor) %>%
  column_to_rownames("CCAA")

distancia <- dist(mat_dim, method = "euclidean")

cluster_jerarquico <- hclust(distancia, method = "ward.D2")

plot(cluster_jerarquico, main = "Clustering Jerárquico de las CCAA", xlab = "", ylab = "Altura")
```

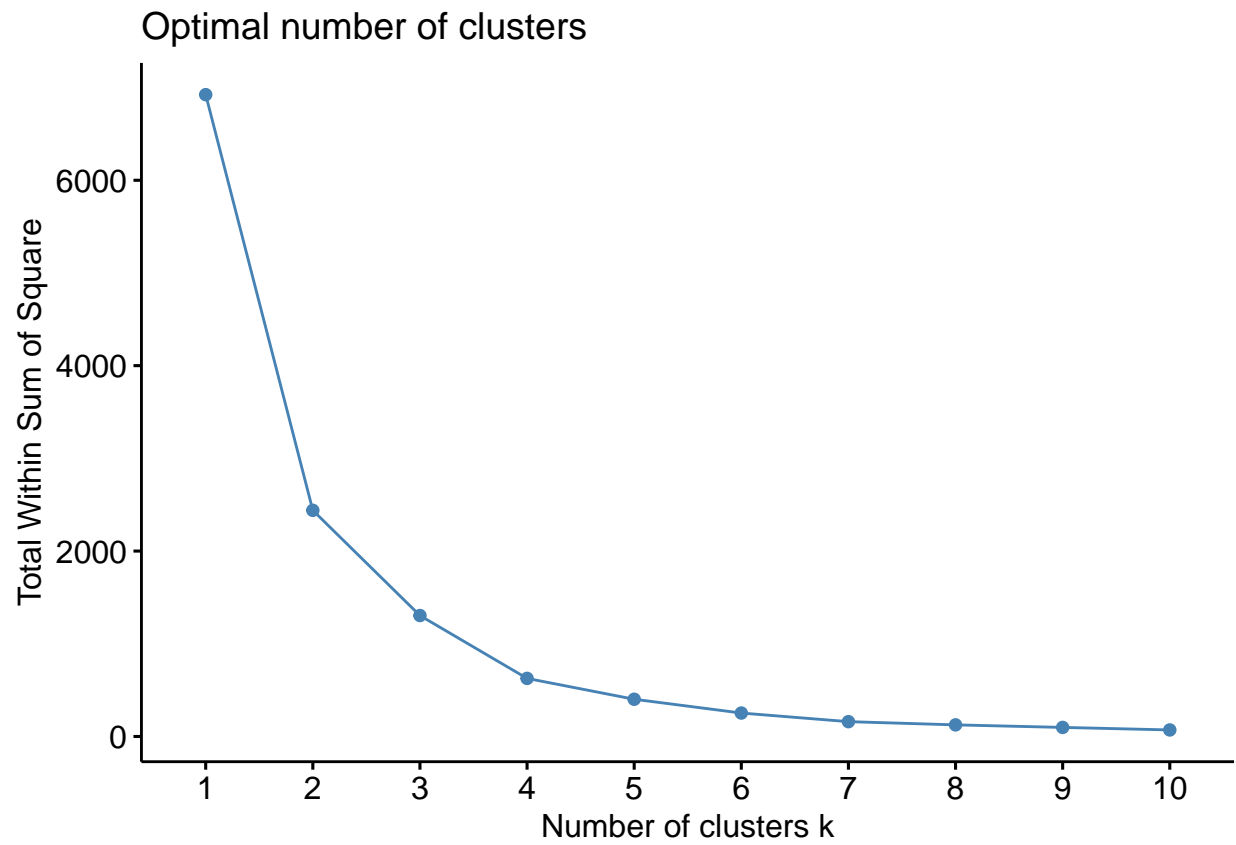
## Clustering Jerárquico de las CCAA



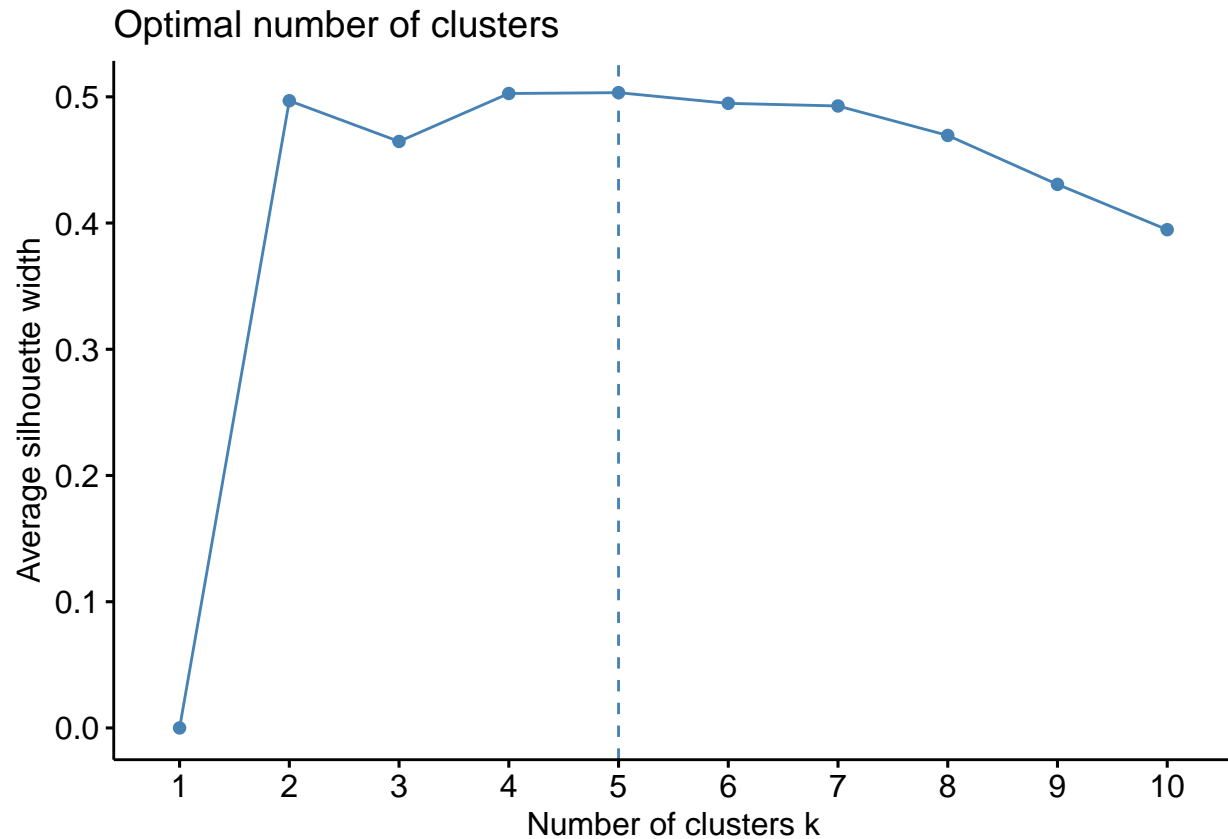
`hclust (*, "ward.D2")`

```
fviz_nbclust(mat_dim, hcut, method = "wss")
```





```
fviz_nbclust(mat_dim, hcut, method = "silhouette")
```



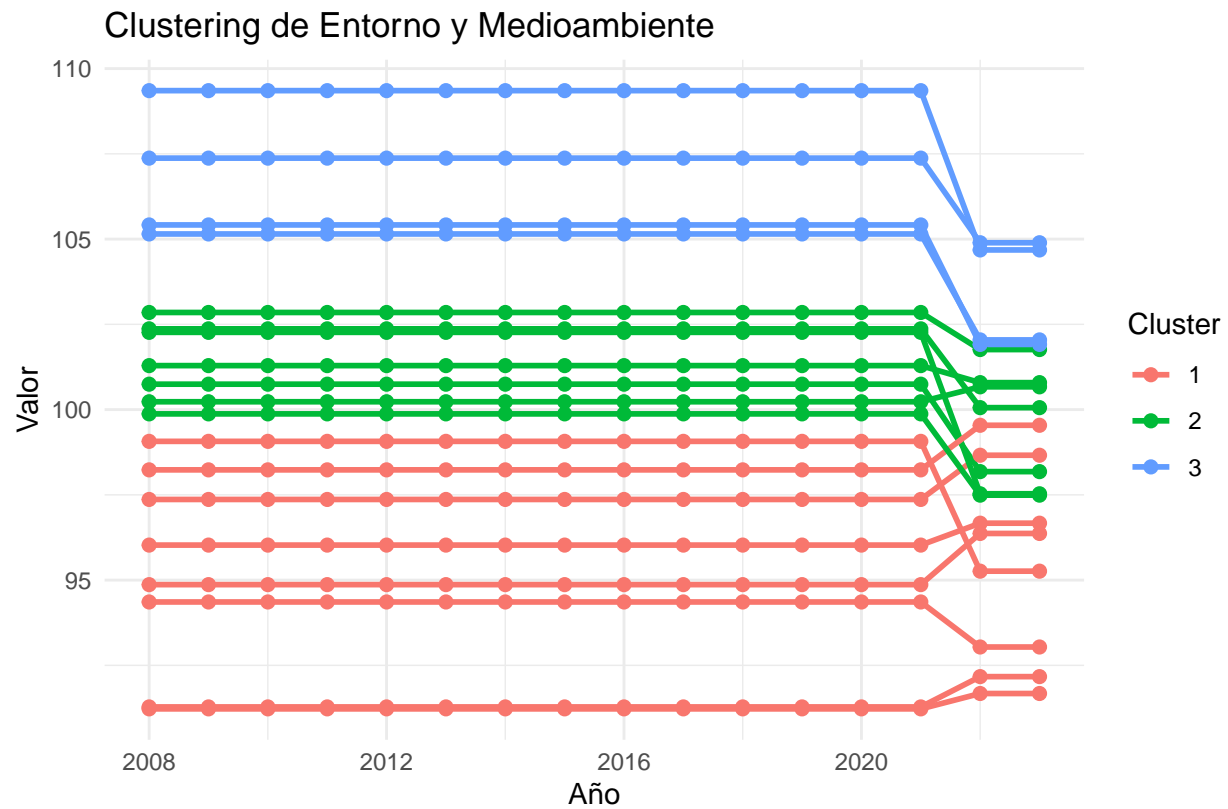
```
grupos <- cutree(cluster_jerarquico, k = 3)

dim_all <- dim_all %>%
  left_join(data.frame(CCAA = rownames(mat_dim), Cluster = grupos), by = "CCAA")

dim_all <- dim_all %>%
  arrange(Cluster, CCAA) # Ordenar primero por cluster y luego por CCAA

plot_cluster <- ggplot(dim_all, aes(x = Anio, y = Valor, color = factor(Cluster), group = CCAA)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Clustering de Entorno y Medioambiente",
    x = "Año",
    y = "Valor",
    color = "Cluster",
    caption = "Fuente: Datos de Calidad de Vida"
  ) +
  theme_minimal()

print(plot_cluster)
```



Fuente: Datos de Calidad de Vida

```
dim_all <- data %>%
  filter(Dim == "dim8")

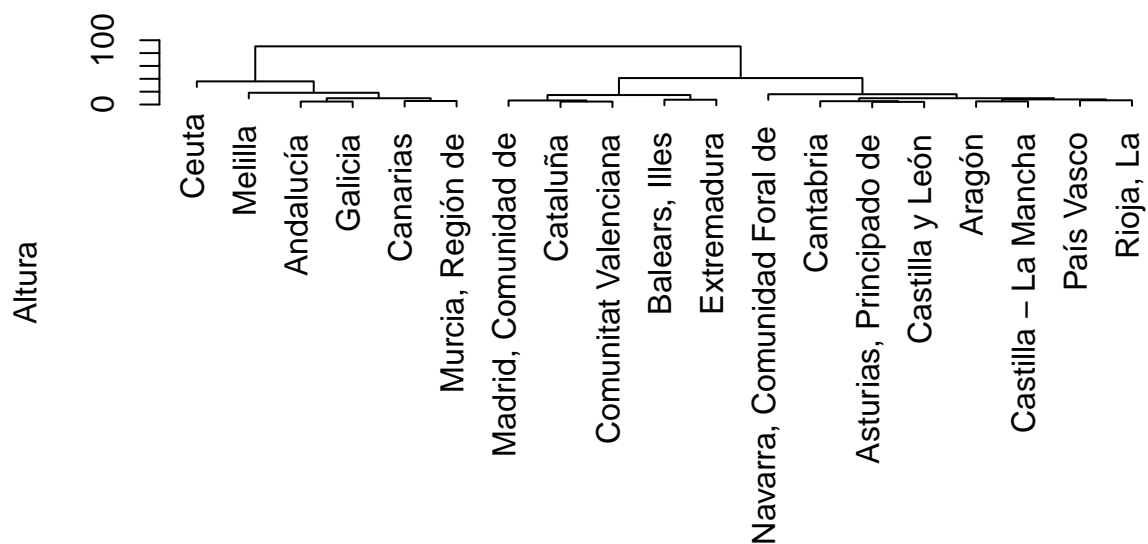
mat_dim <- dim_all %>%
  select(CCAA, Anio, Valor) %>%
  pivot_wider(names_from = Anio, values_from = Valor) %>%
  column_to_rownames("CCAA")

distancia <- dist(mat_dim, method = "euclidean")

cluster_jerarquico <- hclust(distancia, method = "ward.D2")

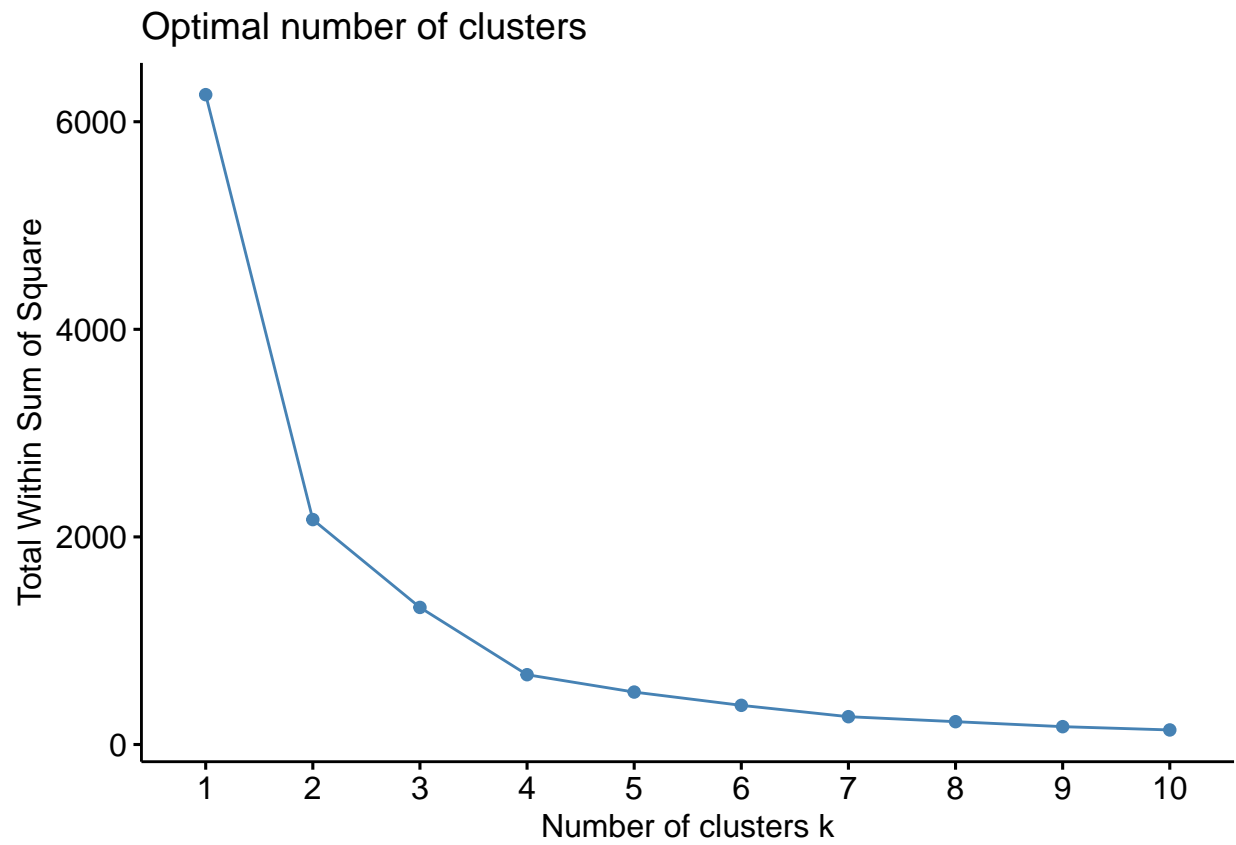
plot(cluster_jerarquico, main = "Clustering Jerárquico de las CCAA", xlab = "", ylab = "Altura")
```

## Clustering Jerárquico de las CCAA

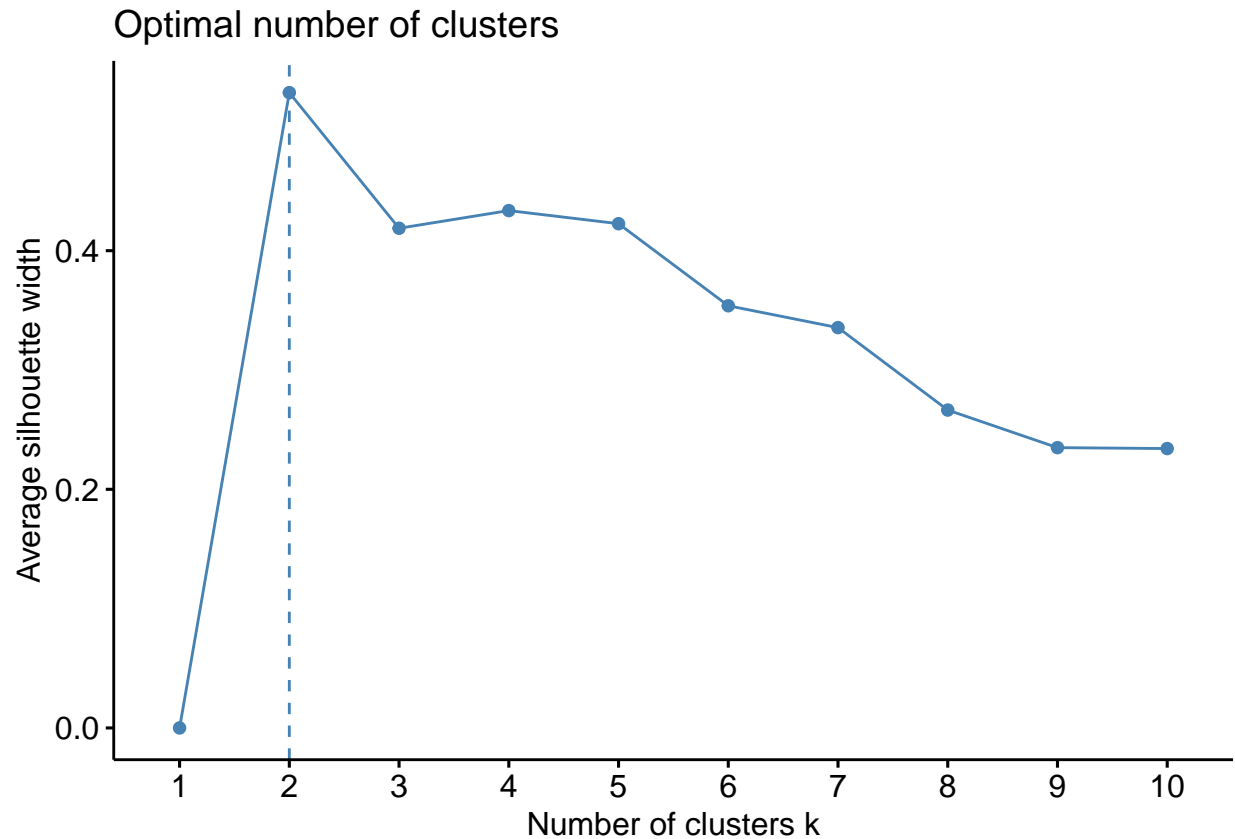


`hclust (*, "ward.D2")`

```
fviz_nbclust(mat_dim, hcut, method = "wss")
```



```
fviz_nbclust(mat_dim, hcut, method = "silhouette")
```



```
grupos <- cutree(cluster_jerarquico, k = 3)

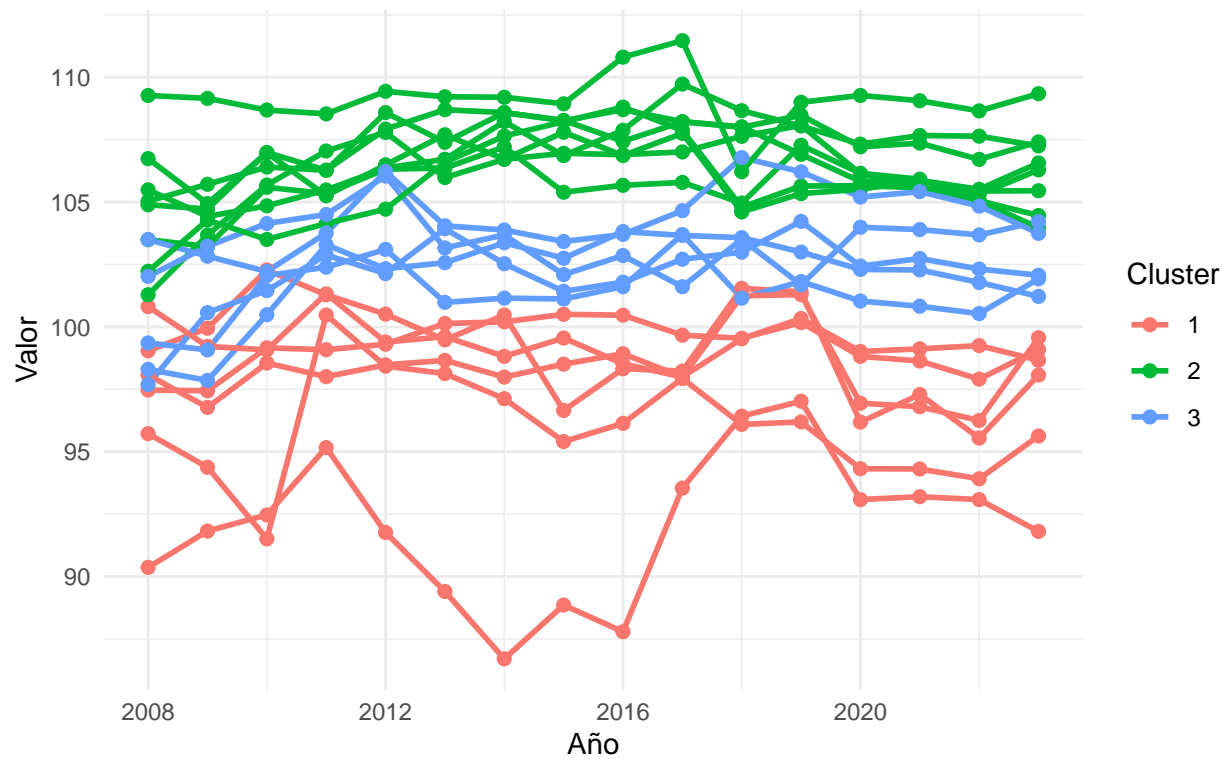
dim_all <- dim_all %>%
  left_join(data.frame(CCAA = rownames(mat_dim), Cluster = grupos), by = "CCAA")

dim_all <- dim_all %>%
  arrange(Cluster, CCAA) # Ordenar primero por cluster y luego por CCAA

plot_cluster <- ggplot(dim_all, aes(x = Anio, y = Valor, color = factor(Cluster), group = CCAA)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Clustering de Entorno y Medioambiente",
    x = "Año",
    y = "Valor",
    color = "Cluster",
    caption = "Fuente: Datos de Calidad de Vida"
  ) +
  theme_minimal()

print(plot_cluster)
```

## Clustering de Entorno y Medioambiente



Fuente: Datos de Calidad de Vida

```
dim_all <- data %>%
  filter(Dim == "dim9")

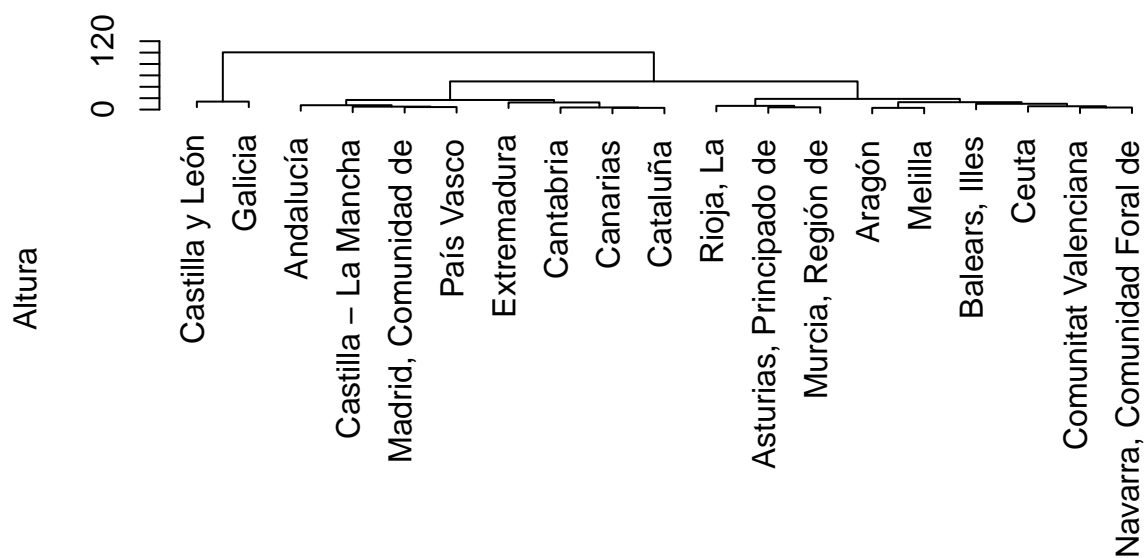
mat_dim <- dim_all %>%
  select(CCAA, Anio, Valor) %>%
  pivot_wider(names_from = Anio, values_from = Valor) %>%
  column_to_rownames("CCAA")

distancia <- dist(mat_dim, method = "euclidean")

cluster_jerarquico <- hclust(distancia, method = "ward.D2")

plot(cluster_jerarquico, main = "Clustering Jerárquico de las CCAA", xlab = "", ylab = "Altura")
```

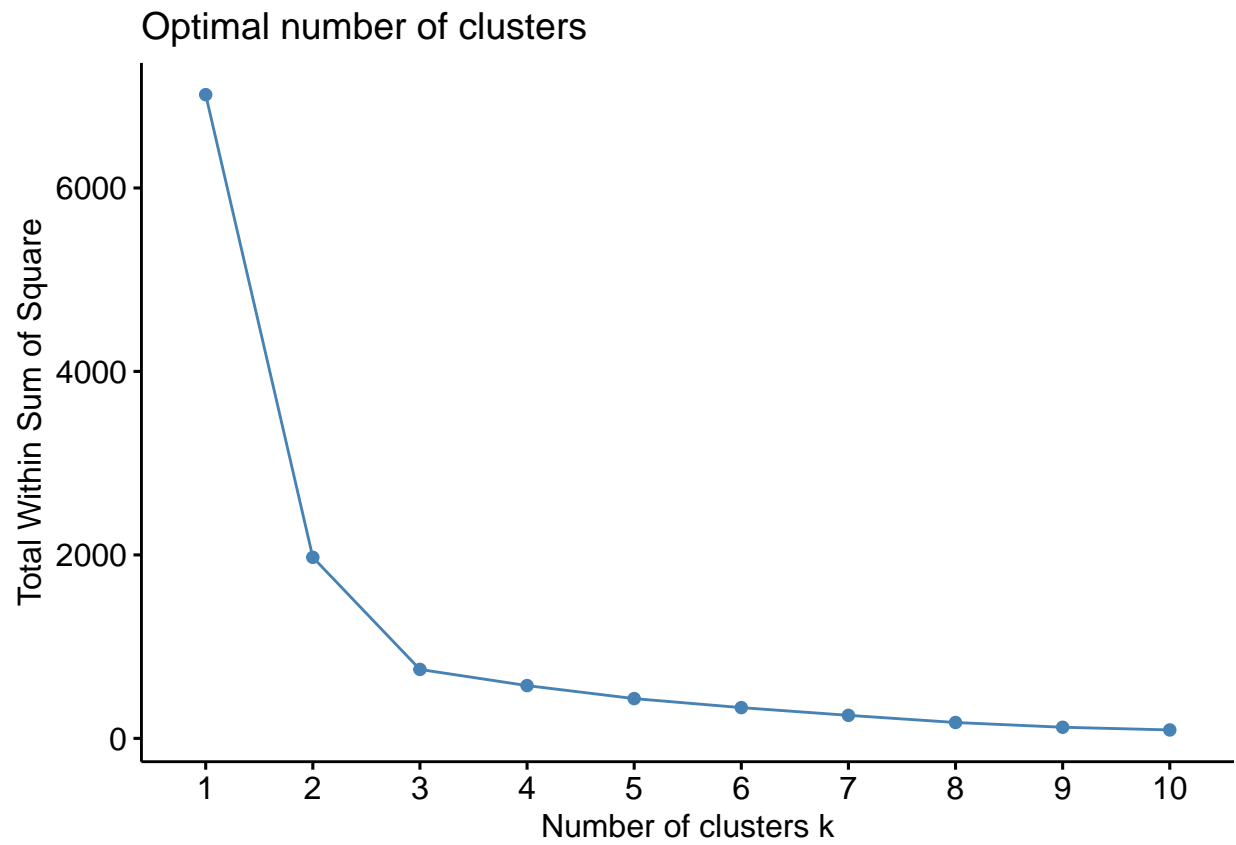
## Clustering Jerárquico de las CCAA



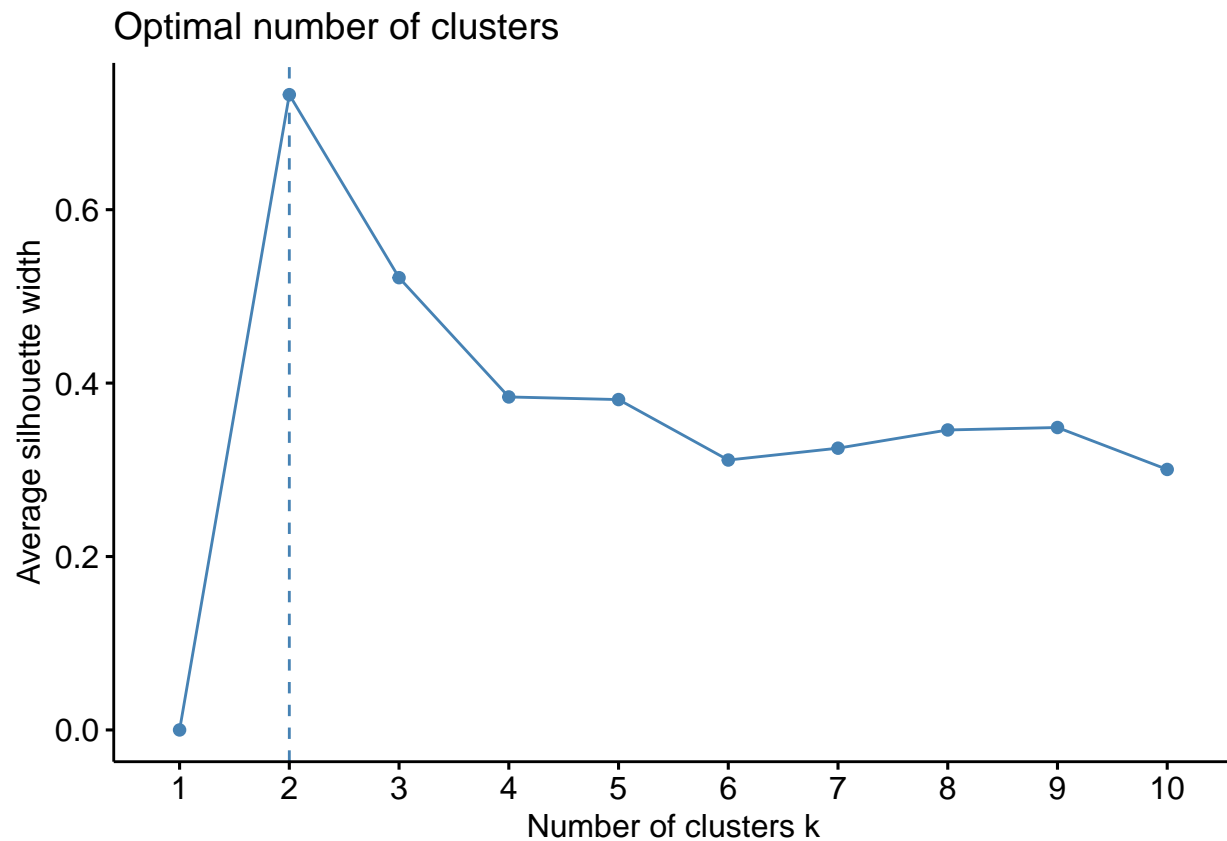
`hclust (*, "ward.D2")`

```
fviz_nbclust(mat_dim, hcut, method = "wss")
```





```
fviz_nbclust(mat_dim, hcut, method = "silhouette")
```



```

grupos <- cutree(cluster_jerarquico, k = 3)

dim_all <- dim_all %>%
  left_join(data.frame(CCAA = rownames(mat_dim), Cluster = grupos), by = "CCAA")

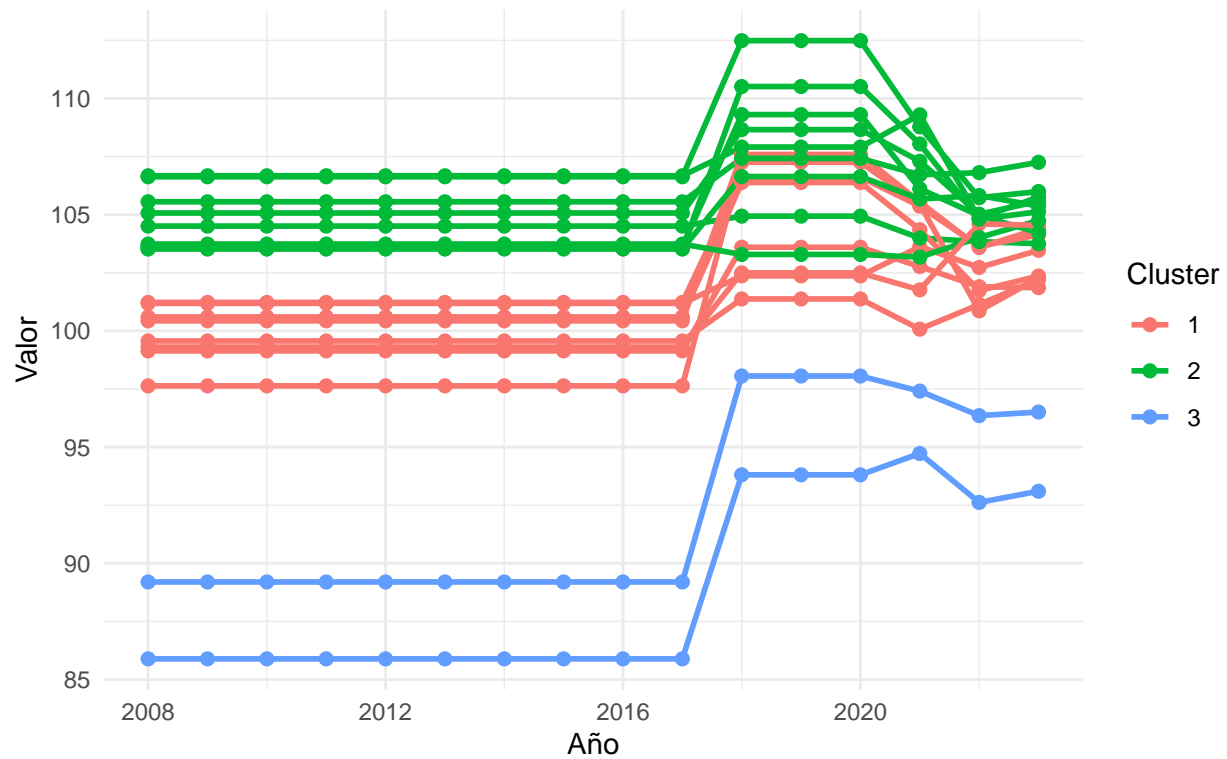
dim_all <- dim_all %>%
  arrange(Cluster, CCAA) # Ordenar primero por cluster y luego por CCAA

plot_cluster <- ggplot(dim_all, aes(x = Anio, y = Valor, color = factor(Cluster), group = CCAA)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Clustering de Entorno y Medioambiente",
    x = "Año",
    y = "Valor",
    color = "Cluster",
    caption = "Fuente: Datos de Calidad de Vida"
  ) +
  theme_minimal()

print(plot_cluster)

```

## Clustering de Entorno y Medioambiente



Fuente: Datos de Calidad de Vida

## matriz de correlacion

```
# Cargar librerías
library(dplyr)
library(reshape2)
library(ggcorrplot)

# Cargar los datos
data <- read.csv("Datos_Tidy.csv")

# Obtener lista de CCAA únicas
ccaa_list <- unique(data$CCAA)

# Crear un bucle para calcular y guardar las matrices de correlación
for (ccaa in ccaa_list) {
  # Filtrar los datos para la CCAA actual
  ccaa_data <- data %>%
    filter(CCAA == ccaa)

  # Reorganizar los datos en formato ancho (dimensiones como columnas)
  ccaa_wide <- ccaa_data %>%
    select(Anio, Dim, Valor) %>%
    spread(key = Dim, value = Valor)
```

```

# Calcular la matriz de correlación
cor_matrix <- cor(ccaa_wide %>% select(-Anio), use = "complete.obs")

# Crear el gráfico de la matriz de correlación
plot <- ggcorrplot(cor_matrix,
  method = "circle",
  type = "lower",
  lab = TRUE,
  lab_size = 3,
  colors = c("red", "white", "blue"),
  title = paste("Matriz de Correlación -", ccaa),
  tl.cex = 10,
  tl.srt = 45)

# Guardar el gráfico en un archivo
ggsave(filename = paste0("matriz_correlacion_", ccaa, ".png"),
  plot = plot,
  width = 8,
  height = 6)

# Mostrar un mensaje de progreso
print(paste("Matriz de correlación para", ccaa, "guardada."))
}

```

```

## [1] "Matriz de correlación para Andalucía guardada."

## [1] "Matriz de correlación para Aragón guardada."

## [1] "Matriz de correlación para Asturias, Principado de guardada."

## [1] "Matriz de correlación para Balears, Illes guardada."

## [1] "Matriz de correlación para Canarias guardada."

## [1] "Matriz de correlación para Cantabria guardada."

## [1] "Matriz de correlación para Castilla y León guardada."

## [1] "Matriz de correlación para Castilla - La Mancha guardada."

## [1] "Matriz de correlación para Cataluña guardada."

## [1] "Matriz de correlación para Comunitat Valenciana guardada."

## [1] "Matriz de correlación para Extremadura guardada."

## [1] "Matriz de correlación para Galicia guardada."

## [1] "Matriz de correlación para Madrid, Comunidad de guardada."

```

```
## [1] "Matriz de correlación para Murcia, Región de guardada."

## [1] "Matriz de correlación para Navarra, Comunidad Foral de guardada."

## [1] "Matriz de correlación para País Vasco guardada."

## [1] "Matriz de correlación para Rioja, La guardada."

## [1] "Matriz de correlación para Ceuta guardada."

## [1] "Matriz de correlación para Melilla guardada."
```