

Proyecto

OT

2023-01-13

Aplicación de conglomerados al proyecto, para este método se necesita una matriz cuadrada por lo cual hemos tomado un extracto de los datos.

```
#install.packages("philentropy")  
#install.packages("readxl")  
library(philentropy)
```

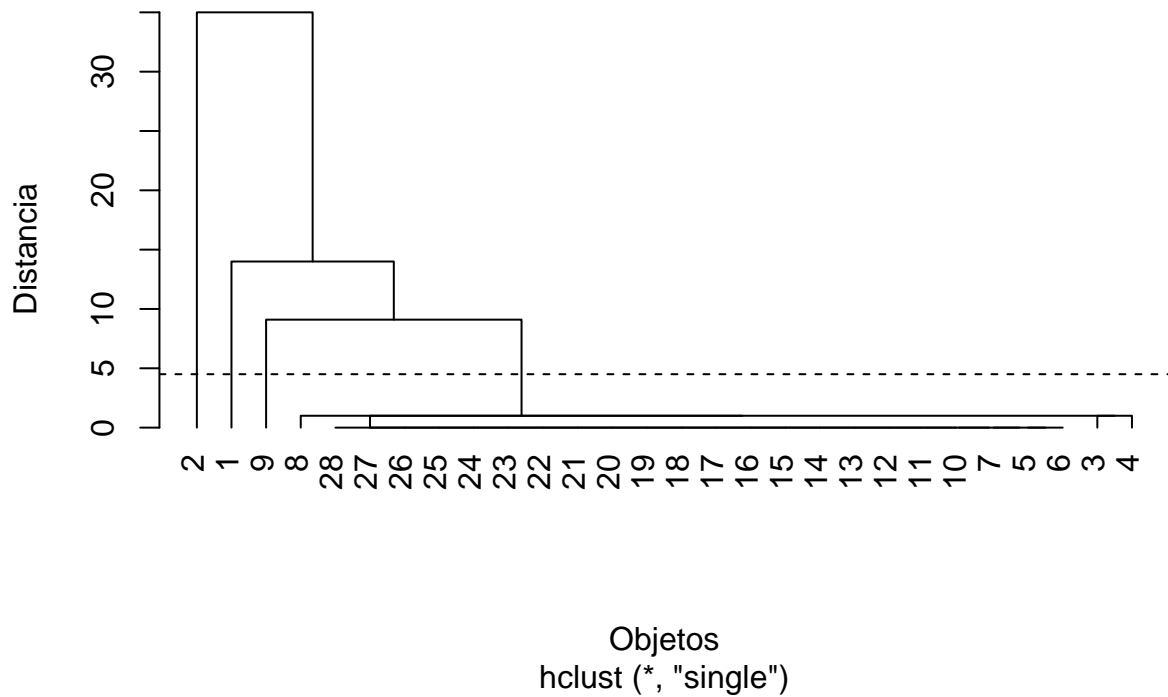
```
## Warning: package 'philentropy' was built under R version 4.2.2
```

```
datos <- readxl::read_excel("matrix.xlsx")  
df <- data.frame(datos)
```

Utilizamos la distancia default que es la distancia euclidiana y el método “*single*”

```
dimnames(df)=list(seq(1:28),seq(1:28))  
  
library(philentropy)  
y<- as.dist(df)  
  
cl <- hclust(y,method="single")  
plot(cl,hang = -1,main="Dendograma: método del vecino más próximo",xlab="Objetos",ylab="Distancia")  
abline(h=4.5,lty=2)
```

Dendograma: método del vecino más próximo

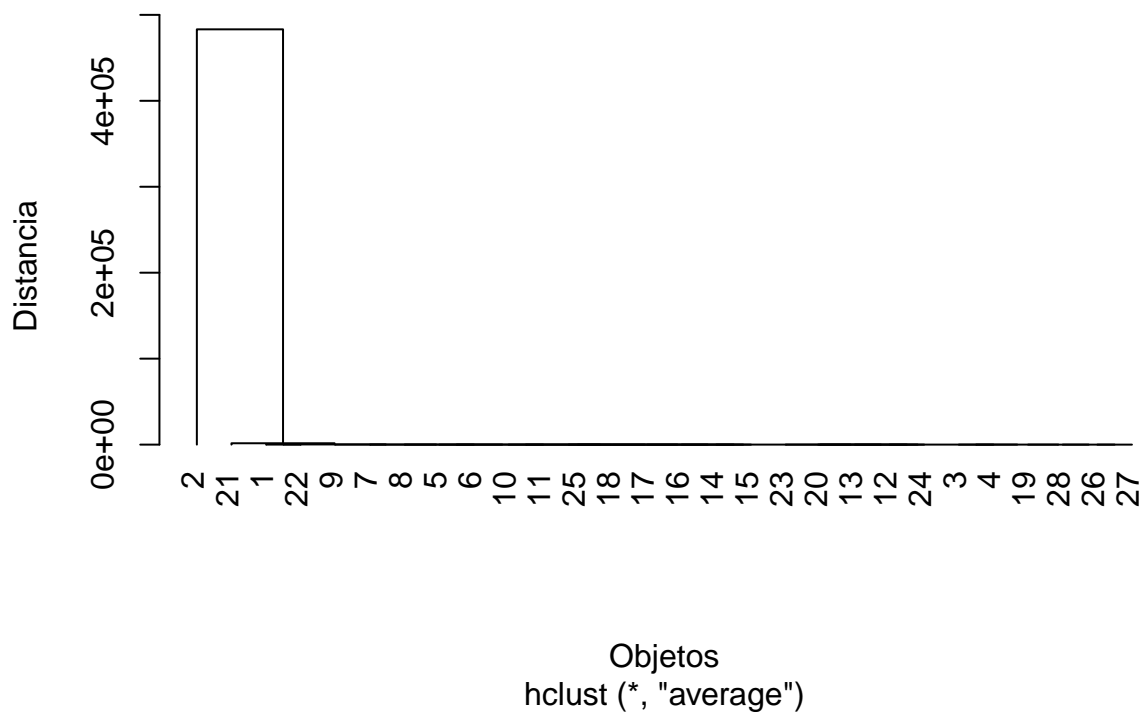


Podemos apreciar que se han generado 4 clusters a una distancia de 4,5.

También podemos realizarlo por el método “*ave*” o “*average*”.

```
c1 <- hclust(y,method="average")
plot(c1,hang = -1,main="Dendograma: método del promedio",xlab="Objetos",ylab="Distancia")
```

Dendograma: método del promedio



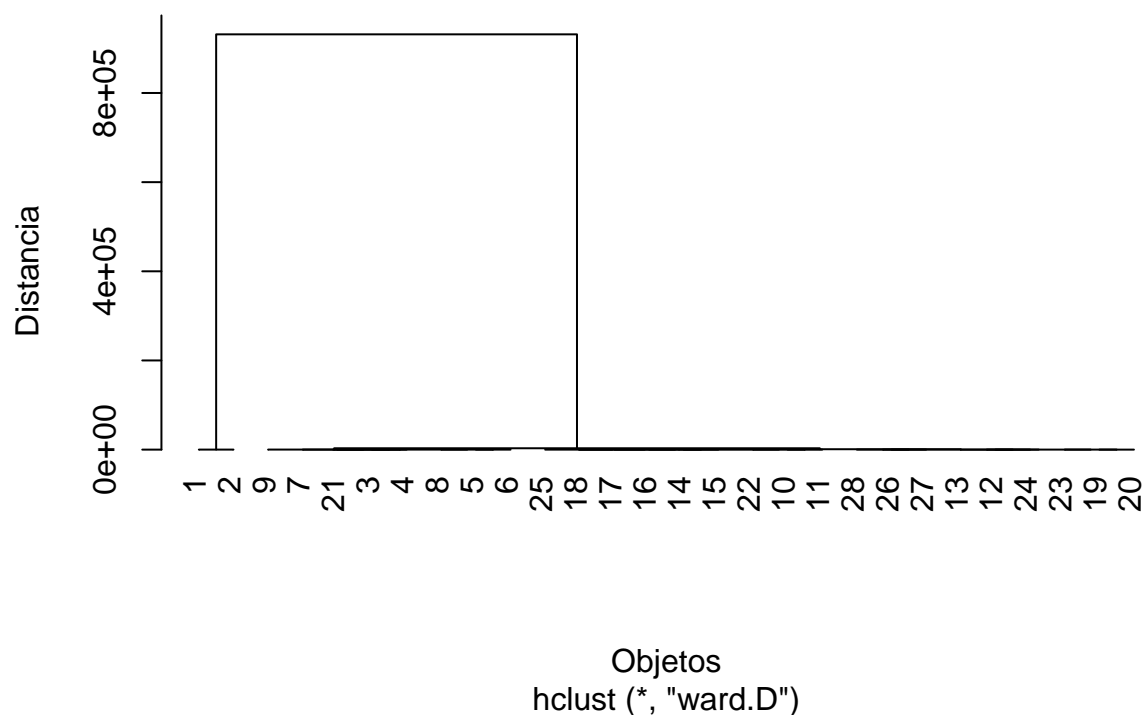
Por el método *ward*.

```
cl <- hclust(y,method="ward")
```

```
## The "ward" method has been renamed to "ward.D"; note new "ward.D2"
```

```
plot(cl, hang = -1, main="Dendrograma: método de Ward", xlab="Objetos", ylab="Distancia")
```

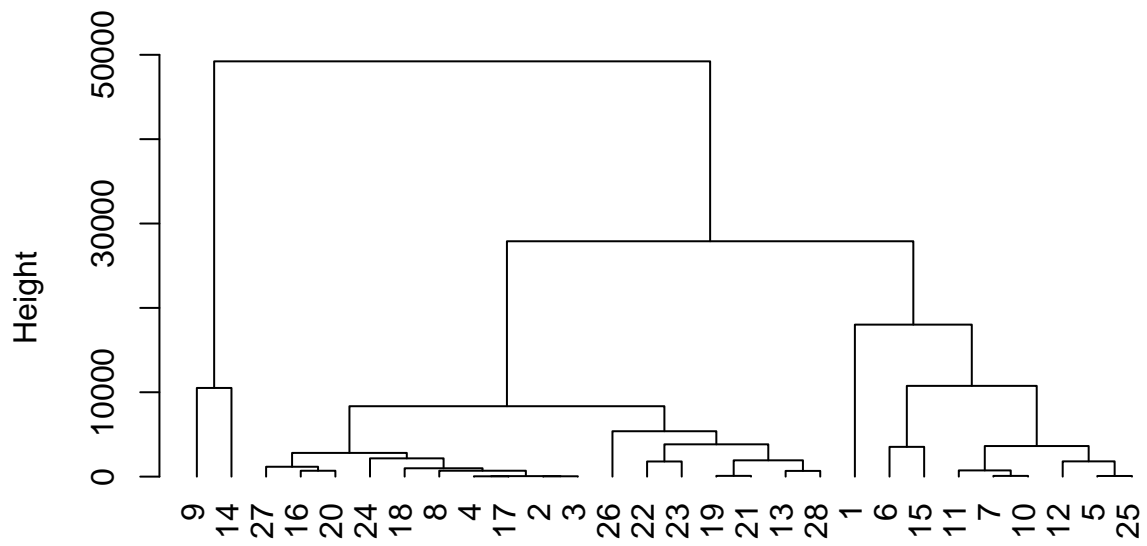
Dendrograma: método de Ward



Esta es una manera de ver como podemos calcular las diferentes distancias en nuestra matriz de dstos.

```
#distancia maxima entre dos componentes de X eY
dd <-dist(df,method = "maximum")
#distacia de manhattan
dd <-dist(df,method = "manhattan")
#distancia de canberra
dd <-dist(df,method = "canberra")
#distancia de minkowski
dd <-dist(df,method = "minkowski")
#distancia euclidiana
dd <-dist(df)
#conformacion de los conglomerados
cl <- hclust(dd)
plot(cl,hang=-1)
```

Cluster Dendrogram



dd
hclust (*, "complete")