

Dendograma: EJemplo Idiomas

Oscar Elí Bonilla Morales

2022-05-24

Cargamos librerías

```
install.packages("cluster.datasets")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)

library("cluster.datasets")
```

Matriz de datos

```
data("languages.spoken.europe")
```

Cambiamos el nombre de la matriz

```
LSE=languages.spoken.europe
head(LSE)
```

```
##      country finnish swedish danish norwegian english german dutch flemish
## 1 West Germany      0      0      0          0      21    100     2      1
## 2      Italy        0      0      0          0       5      3      0      0
## 3      France        0      2      3          0      10      7      1      1
## 4 Netherlands      0      0      0          0      41     47    100    100
## 5      Belgium      0      0      0          0      14     15      0     59
## 6    Luxemburg      0      0      0          0      31    100      4      1
##  french italian spanish portuguese
## 1     10       2       1          0
## 2     11     100       1          0
## 3    100     12       7          1
## 4     16       2       2          0
## 5     44       2       1          0
## 6     92     10       0          0
```

```
#Exploración de la matriz
```

```
dim(LSE)
```

```
## [1] 16 13
```

```
str(LSE)
```

```
## 'data.frame':   16 obs. of  13 variables:
```

```
## $ country : chr "West Germany" "Italy" "France" "Netherlands" ...
## $ finnish : num 0 0 0 0 0 0 0 0 0 0 ...
## $ swedish : num 0 0 2 0 0 0 0 0 0 0 ...
## $ danish : num 0 0 3 0 0 0 0 0 0 0 ...
## $ norwegian : num 0 0 0 0 0 0 0 0 0 0 ...
## $ english : num 21 5 10 41 14 31 100 9 18 21 ...
## $ german : num 100 3 7 47 15 100 7 0 100 83 ...
## $ dutch : num 2 0 1 100 0 4 0 0 1 1 ...
## $ flemish : num 1 0 1 100 59 1 0 0 1 2 ...
## $ french : num 10 11 100 16 44 92 15 10 4 64 ...
## $ italian : num 2 100 12 2 2 10 3 1 2 23 ...
## $ spanish : num 1 1 7 2 1 0 2 2 1 3 ...
## $ portuguese: num 0 0 1 0 0 0 0 100 0 1 ...
```

```
anyNA(LSE)
```

```
## [1] FALSE
```

Calculo de la matriz de distancia de Mahalonobis

Para poder realizar esta función es necesario solo utilizar variables numéricas, por lo que realizaremos una modificación en en nuestro data set

```
dist.LSE<-dist(LSE[,2:12])
```

Convertir los resultados del Calculo de la distancia a una matriz de datos y me indique 3 digitos.

```
round(as.matrix(dist.LSE)[1:12, 1:12],3)
```

```
##      1      2      3      4      5      6      7      8      9
## 1  0.000 138.834 130.461 150.502 108.619 83.024 122.156 100.752 6.782
## 2 138.834 0.000 125.527 181.279 119.996 157.366 135.893 99.136 138.683
## 3 130.461 125.527 0.000 171.339 82.037 96.068 124.282 91.165 134.454
## 4 150.502 181.279 171.339 0.000 119.243 166.643 158.376 152.545 151.937
## 5 108.619 119.996 82.037 119.243 0.000 115.165 108.554 69.921 110.481
## 6 83.024 157.366 96.068 166.643 115.165 0.000 139.316 131.567 89.370
## 7 122.156 135.893 124.282 158.376 108.554 139.316 0.000 91.428 124.491
## 8 100.752 99.136 91.165 152.545 69.921 131.567 91.428 0.000 100.603
## 9 6.782 138.683 134.454 151.937 110.481 89.370 124.491 100.603 0.000
## 10 60.432 124.109 85.697 154.425 93.638 36.892 121.754 102.171 65.909
## 11 127.973 148.324 142.151 175.602 126.937 153.330 117.919 109.745 128.464
## 12 123.576 149.913 141.602 175.083 128.152 147.689 124.888 114.140 124.165
##      10      11      12
## 1  60.432 127.973 123.576
## 2 124.109 148.324 149.913
## 3 85.697 142.151 141.602
## 4 154.425 175.602 175.083
## 5 93.638 126.937 128.152
## 6 36.892 153.330 147.689
## 7 121.754 117.919 124.888
## 8 102.171 109.745 114.140
## 9 65.909 128.464 124.165
```

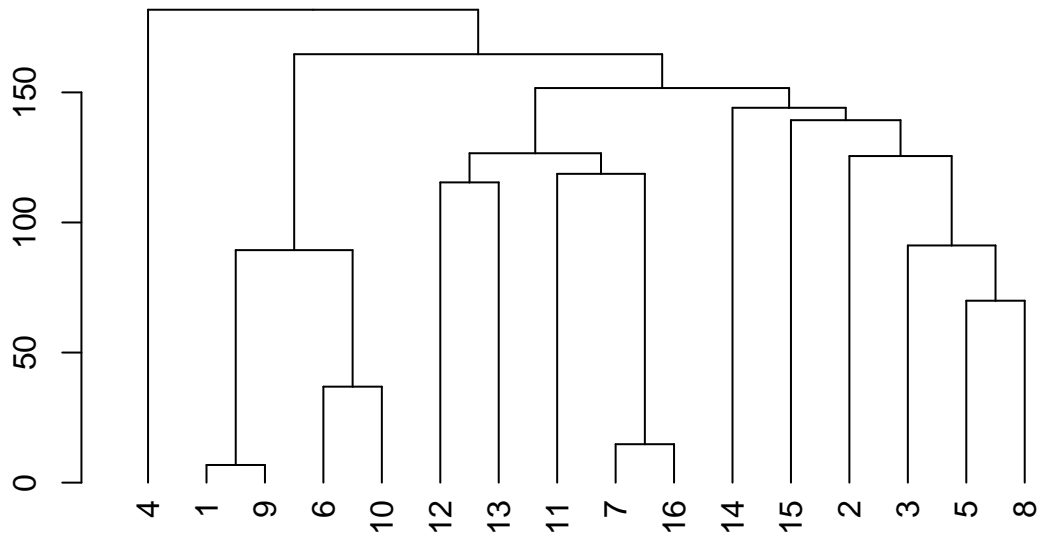
```
## 10  0.000 133.981 129.240
## 11 133.981  0.000 120.241
## 12 129.240 120.241  0.000
```

Calculo del dendrograma

```
dend.LSE<-as.dendrogram(hclust(dist.LSE))
```

Generacion del dendrograma

```
plot(dend.LSE)
```

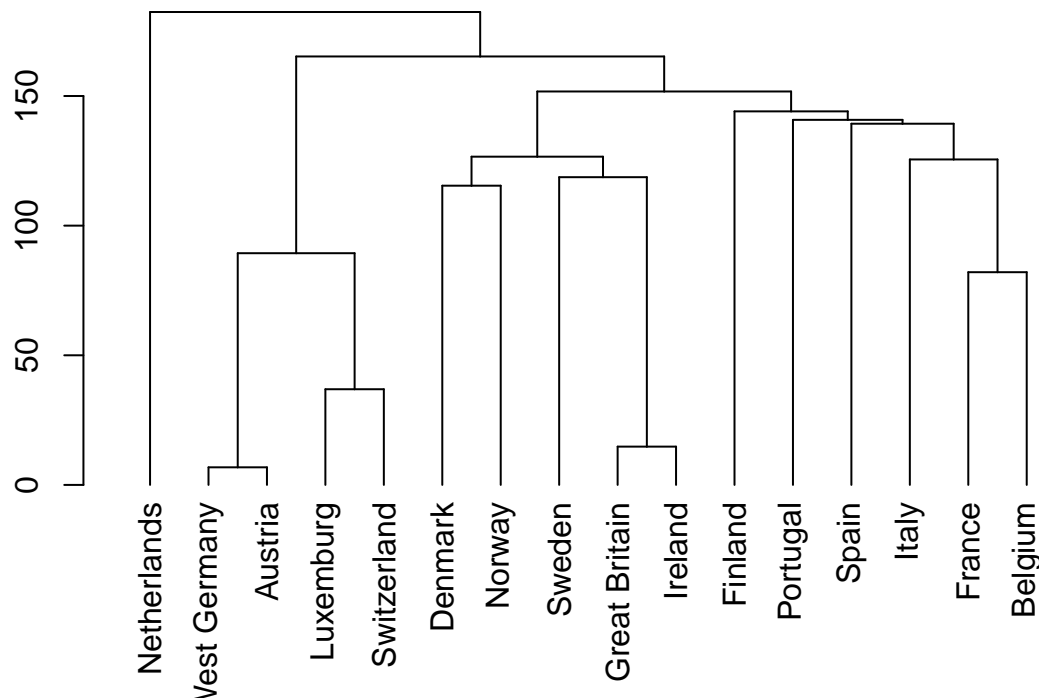


Agregar etiquetas al Grafico

```
LSE.nombres=LSE
rownames(LSE.nombres)= LSE.nombres$country
LSE.nombres=LSE.nombres[, -1]
```

Construimos de nuevo el Grafico

```
plot(as.dendrogram(hclust(dist(LSE.nombres))))
```



Interpretación idiomas

En nuestro dendrograma podemos decir que se divide en 3 clusters principales esto quiere decir que entre nuestros 3 clusters existirán países que comparten los mismos idiomas, por ejemplo, si creamos grupos podríamos decir que en el grupo 1 el cual incluye (alemania del este, Austria, Luxemburgo y Suiza) el idioma mas hablado es el alemán, mientras que en el grupo 2 (Dinamarca, Noruega, Suecia, Gran Bretaña e Irlanda) aunque también está presente el alemán predomina más el inglés.

Modificar el dendrograma

```
install.packages("dendextend")
library(dendextend)
```

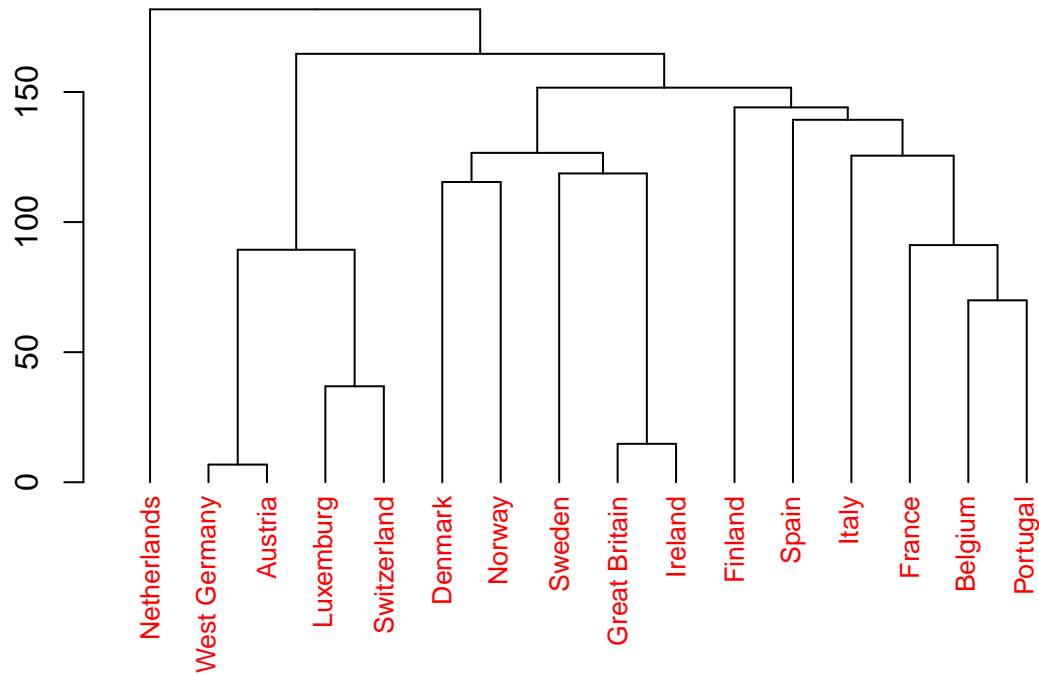
Guardar las etiquetas en un objeto “L”

```
L=labels(dend.LSE)
labels(dend.LSE)=LSE$country[L]
```

Cambiar el tamaño de las etiquetas

```
dend.LSE %>%
  set(what="labels_col", "red") %>% #Colores etiqueta
  set(what="labels_cex", 0.8) %>%
  plot(main="Dendrograma de Idiomas en Europa")
```

Dendrograma de Idiomas en Europa



Dendrograma de Circulo

```
install.packages("circlize")
library("circlize")

circlize_dendrogram(dend.LSE, labels_track_height=NA,
                    dend_track_height=0.1)
```

