

Dendrograma

Karina Itzel Rodríguez Conde

2022-05-10

Introducción

El Dendrograma también es conocido como *árbol ultramétrico* y es un diagrama de árbol que contiene un punto llamado **raíz** y n puntos extremos. En él, se muestran los grupos que se crearon al formar conglomerados mediante ciertas similitudes. Las categorías o grupos en los que se dividen, forman otros subgrupos.

Paquetería y librería a utilizar

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

```
library(cluster.datasets)
```

Matriz de datos

La base de datos a utilizar es *all.us.city.crime.1970* y está precargada en R en la paquetería **cluster.datasets**, contiene los registros del crimen de la ciudad junto con las estadísticas de población, con 24 observaciones y 10 variables.

```
data("all.us.city.crime.1970")
```

Tratamiento y exploración de la matriz

```
AMM=all.us.city.crime.1970  
head(AMM)
```

##	city	population	white.change	black.population	murder	rape	robbery
## 1	Anaheim	1420	50.8	39	2.7	21.9	94
## 2	Baltimore	2071	-21.4	501	13.2	34.9	564
## 3	Boston	2754	-16.5	151	4.4	14.8	136
## 4	Buffalo	1349	-20.7	118	5.7	13.7	145
## 5	Chicago	6979	-18.6	1306	12.9	25.4	363
## 6	Cincinnati	1385	-17.2	156	6.4	16.8	120
##	assault	burglary	car.theft				
## 1	103	1607	377				
## 2	396	1351	701				
## 3	95	1054	984				
## 4	111	862	448				
## 5	233	830	708				
## 6	107	912	348				

1.- Dimensión de la base

```
dim(AMM)
```

```
## [1] 24 10
```

2.- Tipo de variables

```
str(AMM)
```

```
## 'data.frame': 24 obs. of 10 variables:
## $ city : chr "Anaheim" "Baltimore" "Boston" "Buffalo" ...
## $ population : num 1420 2071 2754 1349 6979 ...
## $ white.change : num 50.8 -21.4 -16.5 -20.7 -18.6 -17.2 -26.5 14.2 -29.1 25.5 ...
## $ black.population: num 39 501 151 118 1306 ...
## $ murder : num 2.7 13.2 4.4 5.7 12.9 6.4 14.5 18.4 14.7 16.9 ...
## $ rape : num 21.9 34.9 14.8 13.7 25.4 16.8 18.7 41 31.1 27.1 ...
## $ robbery : num 94 564 136 145 363 120 288 206 649 335 ...
## $ assault : num 103 396 95 111 233 107 132 338 223 183 ...
## $ burglary : num 1607 1351 1054 862 830 ...
## $ car.theft : num 377 701 984 448 708 ...
```

3.- Saber si existen datos nulos

```
anyNA(AMM)
```

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

Esta matriz no contiene datos nulos.

4.- Resumen de los datos

```
summary(AMM)
```

```
##      city      population      white.change      black.population
## Length:24      Min.   : 1268      Min.   : -39.400      Min.   : 39.0
## Class :character 1st Qu.: 1416      1st Qu.: -20.875      1st Qu.: 117.5
## Mode  :character Median : 2024      Median : -13.450      Median : 302.0
##              Mean   : 2932      Mean   : -8.304      Mean   : 452.8
##              3rd Qu.: 2923      3rd Qu.: 6.750      3rd Qu.: 585.5
##              Max.   :11529      Max.   : 50.800      Max.   :2080.0
##      murder      rape      robbery      assault
## Min.   : 2.600      Min.   : 5.70      Min.   : 53.0      Min.   : 63.0
## 1st Qu.: 4.400      1st Qu.:16.40      1st Qu.:142.8      1st Qu.:106.8
## Median : 9.350      Median :20.20      Median :243.0      Median :157.0
## Mean   : 9.188      Mean   :23.18      Mean   :277.9      Mean   :187.8
## 3rd Qu.:13.525      3rd Qu.:28.10      3rd Qu.:351.8      3rd Qu.:232.2
## Max.   :18.400      Max.   :50.00      Max.   :665.0      Max.   :421.0
##      burglary      car.theft
## Min.   : 499      Min.   : 348.0
## 1st Qu.: 854      1st Qu.: 523.2
## Median :1333      Median : 684.0
## Mean   :1313      Mean   : 679.6
## 3rd Qu.:1660      3rd Qu.: 795.8
## Max.   :2164      Max.   :1208.0
```

5.- Cálculo de la matriz de distancia de Mahalonobis

```
dist.AMM<-dist(AMM[,2:10])
```

6.- Convertir los resultados del cálculo de la distancia a una matriz de datos y me indique 3 dígitos.

```
round(as.matrix(dist.AMM)[1:10, 1:10],3)
```

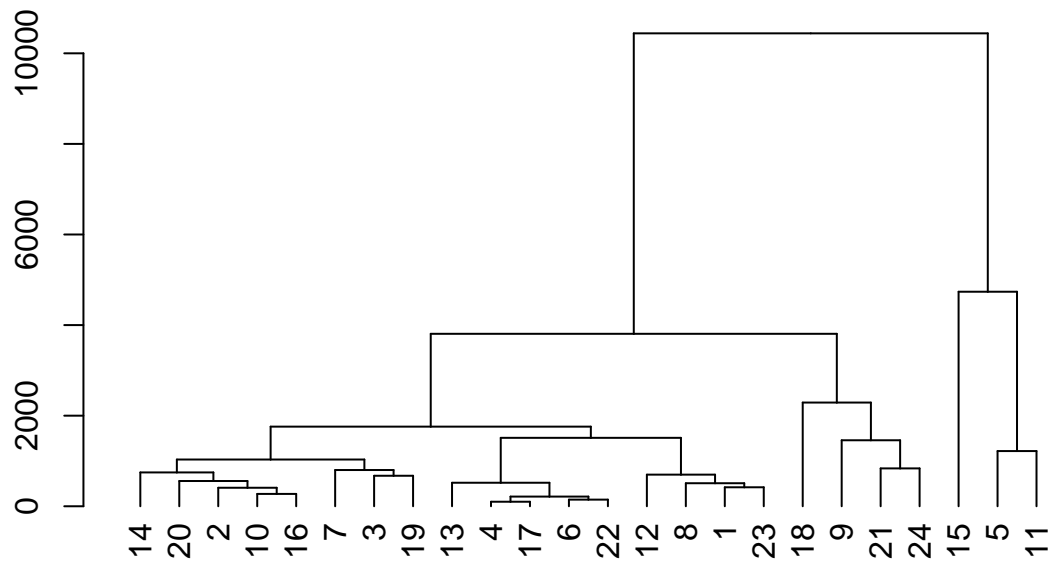
```
##           1           2           3           4           5           6           7           8
## 1      0.000 1058.300 1572.504  761.053 5771.937  710.025 1360.983  422.109
## 2  1058.300    0.000 1015.666 1108.330 5007.499 1089.954  838.957  723.851
## 3  1572.504 1015.666    0.000 1516.454 4402.472 1516.329  799.884 1398.734
## 4   761.053 1108.330 1516.454    0.000 5765.368  126.108 1077.855  808.872
## 5  5771.937 5007.499 4402.472 5765.368    0.000 5729.459 5034.933 5578.453
## 6   710.025 1089.954 1516.329  126.108 5729.459    0.000 1127.835  776.364
## 7  1360.983  838.957  799.884 1077.855 5034.933 1127.835    0.000 1133.271
## 8   422.109  723.851 1398.734  808.872 5578.453  776.364 1133.271    0.000
## 9  2982.502 2248.152 1919.985 3192.638 3069.266 3151.014 2537.780 2769.177
## 10  807.648  390.186  994.894 1030.500 5124.629 1005.952  856.442  522.849
##           9           10
## 1  2982.502  807.648
## 2  2248.152  390.186
## 3  1919.985  994.894
## 4  3192.638 1030.500
## 5  3069.266 5124.629
## 6  3151.014 1005.952
## 7  2537.780  856.442
## 8  2769.177  522.849
## 9    0.000 2315.381
## 10 2315.381    0.000
```

7.- Cálculo del dendrograma

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

8.- Generación del dendrograma

```
plot(dend.AMM)
```

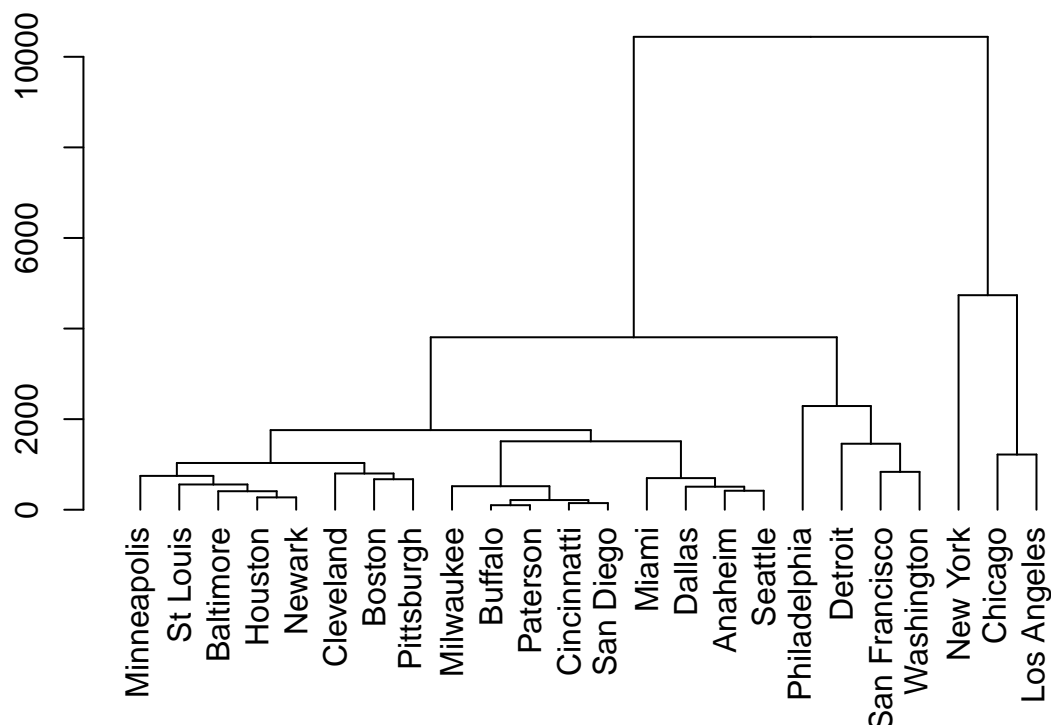


9.- Agregar etiquetas al gráfico

```
AMM.nombres=AMM
rownames(AMM.nombres)= AMM.nombres$city
AMM.nombres=AMM.nombres[, -1]
```

10.- Construcción del nuevo gráfico

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



Modificando el dendrograma

Paquetería y librería a utilizar

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

```
library(dendextend)
```

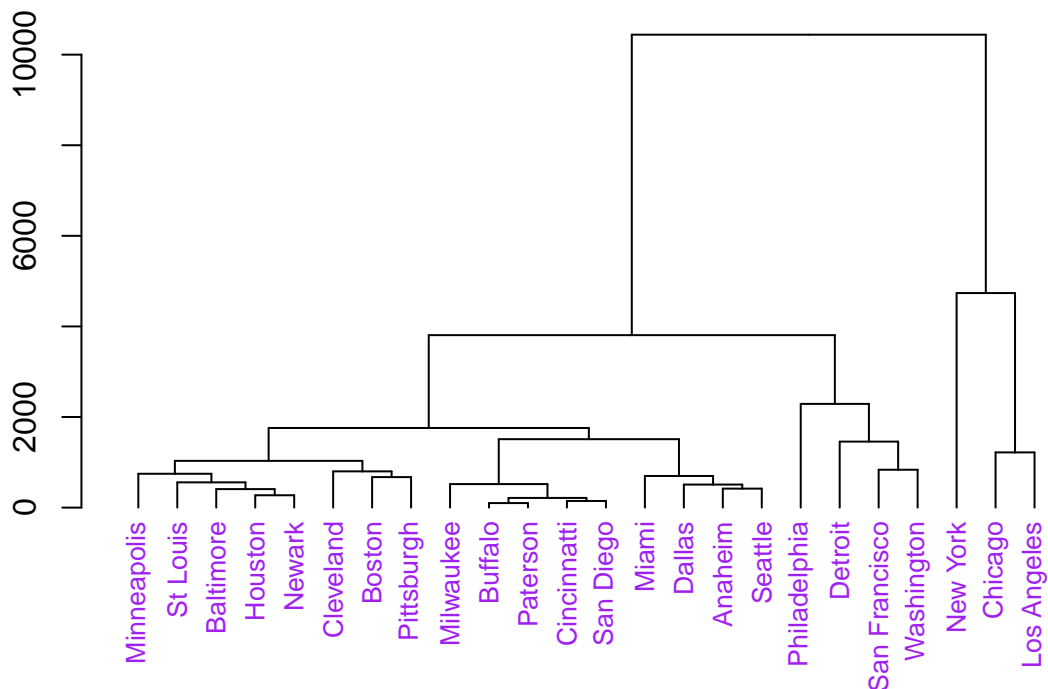
1.- Guardar las etiquetas en un objeto “L”

```
L=labels(dend.AMM)  
labels(dend.AMM)=AMM$city[L]
```

2.- Cambiar el tamaño de las etiquetas

```
dend.AMM %>%  
  set(what="labels_col", "purple") %>%  
  set(what="labels_cex", 0.8) %>%  
  plot(main="Dendrograma de ciudades")
```

Dendrograma de ciudades



El dendrograma anterior muestra aquellas ciudades que pertenecen o no a un mismo grupo de clasificación. Las ciudades de Minneapolis, St. Louis, Baltimore, Houston y Newark conforman un sólo grupo, mientras que, las ciudades de Cleveland, Boston y Pittsburgh conforman otro grupo. Estas ciudades y todas las demás que están clasificadas por grupos, tienen una o más características similares que hacen que sean de la misma clasificación.