cluster jerarquico y kmedias 2

Jaime Isaac Peña Mejía

23/4/2021

Cluster jerárquico.

```
clust <- read.table("task9data2.txt", quote="\"", comment.char="")
clust
```

```
##               V1  V2  V3
## 1      Australia 2.5 211
## 2        Austria 3.9 167
## 3    Belgium/Lux 2.9 131
## 4         Canada 2.4 191
## 5        Denmark 2.9 220
## 6        Finland 0.8 297
## 7         France 9.1  71
## 8        Iceland 0.8 211
## 9        Ireland 0.7 300
## 10         Italy 7.9 107
## 11   Netherlands 1.8 167
## 12   New_Zealand 1.9 266
## 13        Norway 0.8 227
## 14         Spain 6.5  86
## 15        Sweden 1.6 207
## 16   Switzerland 5.8 115
## 17       England 1.3 285
## 18 United_States 1.2 199
## 19       Germany 2.7 172
```

```
# Give row names and column names
cnames <- c("Pais","wine","heart")
colnames(clust)<-cnames
rownames(clust)<-clust$Pais
clust
```

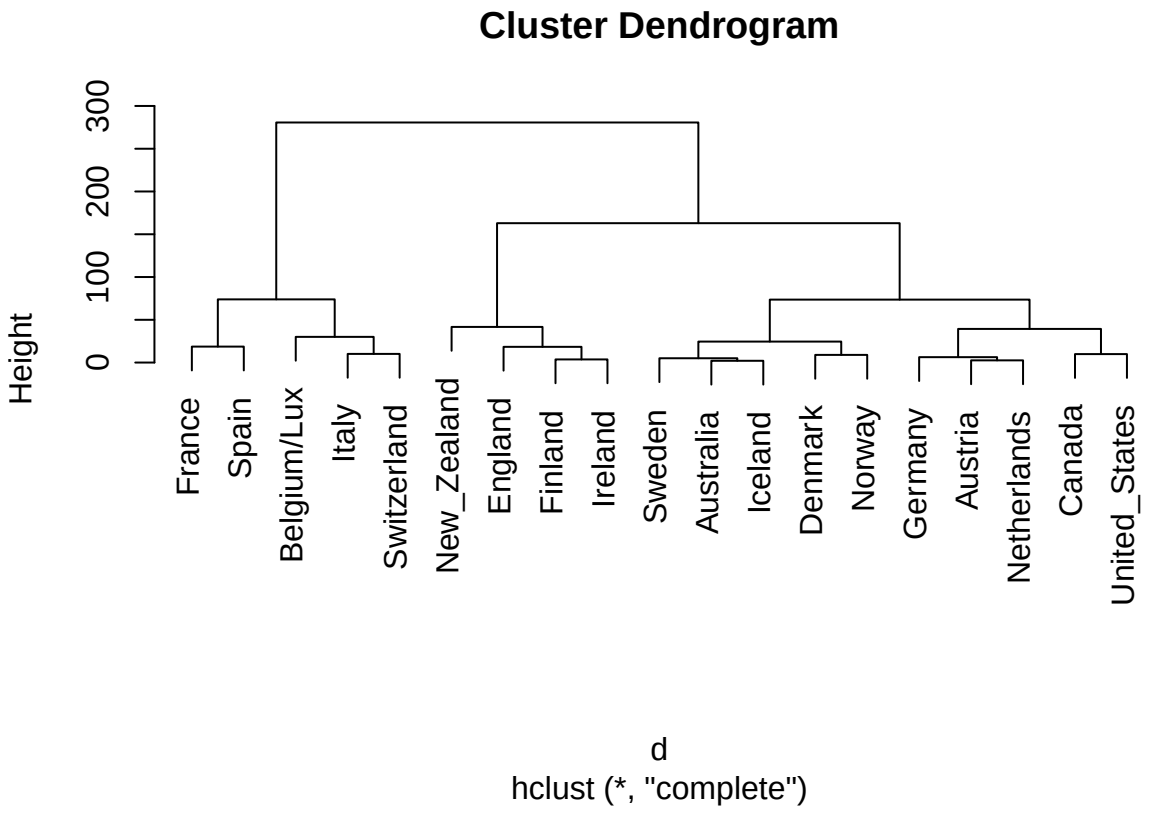
```
##                       Pais wine heart
## Australia         Australia  2.5   211
## Austria             Austria  3.9   167
## Belgium/Lux     Belgium/Lux  2.9   131
## Canada               Canada  2.4   191
## Denmark             Denmark  2.9   220
## Finland             Finland  0.8   297
## France               France  9.1    71
## Iceland             Iceland  0.8   211
## Ireland             Ireland  0.7   300
```

```
## Italy                Italy  7.9   107
## Netherlands     Netherlands  1.8   167
## New_Zealand     New_Zealand  1.9   266
## Norway               Norway  0.8   227
## Spain                 Spain  6.5    86
## Sweden               Sweden  1.6   207
## Switzerland     Switzerland  5.8   115
## England             England  1.3   285
## United_States United_States  1.2   199
## Germany             Germany  2.7   172
```

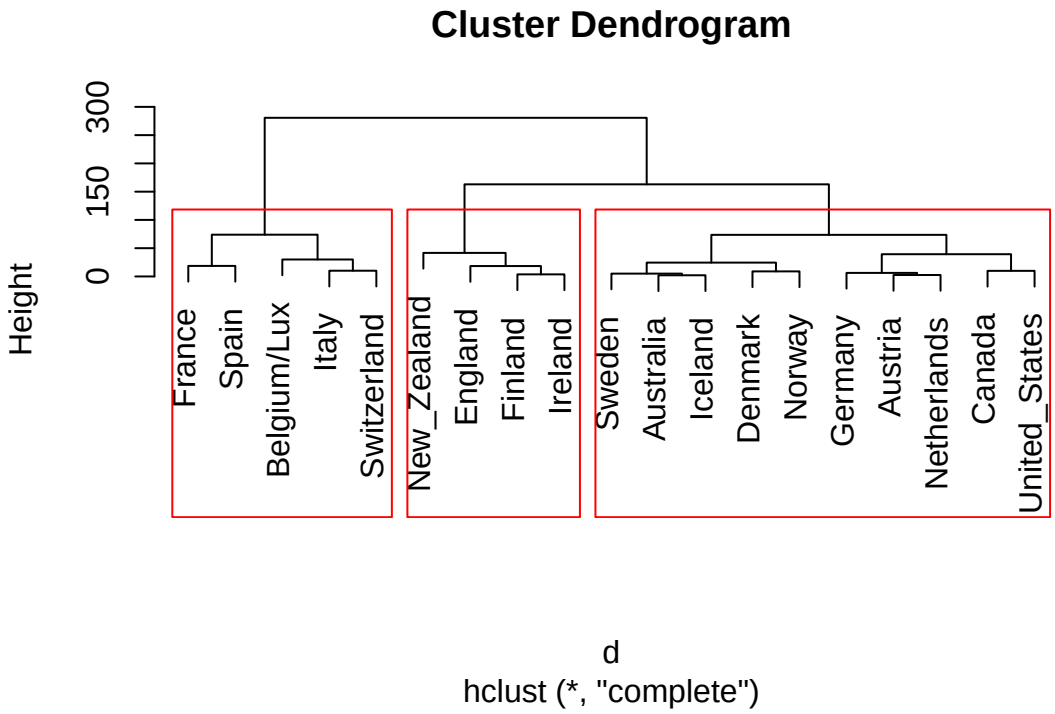
```
# Hierarchical Clustering
# Create distance matrix
d <- dist(clust, method = "euclidean")
```

```
## Warning in dist(clust, method = "euclidean"): NAs introducidos por coerción
```

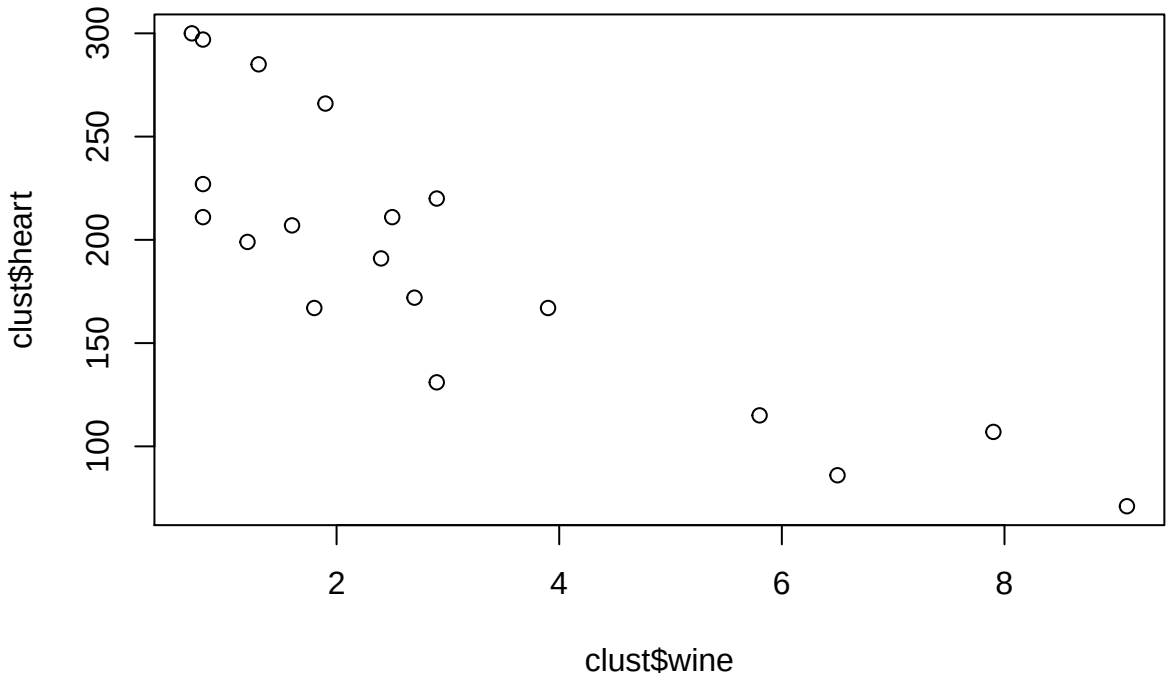
```
fit <- hclust(d)
# display dendogram
plot(fit)
```




```
plot(fit)
groups<-cutree(fit, k=5) # cut tree into 5 clusters
# draw dendogram with red borders around the 5 clusters
rect.hclust(fit,k=3,border="red")
```




```
plot(clust$wine,clust$heart)
text(x=clust$wine, y=clust$heart, labels=clust$country, col=groups)
```



k means en R de forma corta

```
# K means
clustk <- kmeans(clust[,c("wine","heart")], centers=3, nstart=10)
clustk
```

```
## K-means clustering with 3 clusters of sizes 4, 5, 10
##
## Cluster means:
##    wine heart
## 1 1.175 287.0
## 2 6.440 102.0
## 3 2.060 197.2
##
## Clustering vector:
##     Australia       Austria   Belgium/Lux        Canada       Denmark
##             3             3             2             3             3
##       Finland        France       Iceland       Ireland         Italy
##             1             2             3             1             2
##   Netherlands   New_Zealand        Norway         Spain        Sweden
##             3             1             3             2             3
##   Switzerland       England United_States       Germany
##             2             1             3             3
##
## Within cluster sum of squares by cluster:
## [1]  714.9075 2274.1520 4394.6040
##  (between_SS / total_SS =  91.2 %)
##
```

```
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
plot(clust$wine, clust$heart, xlab="wine", ylab="heart")
text(x=clust$wine, y=clust$heart, labels=clust$country,col=clustk$cluster+1)
```