MSMS 206: Practical 04

Ananda Biswas

April 10, 2025

- Question: Consider the "Swiss" dataset in MASS package of R. Perform the following clustering algorithms to divide the data-set into clusters.
- (a) k-means clustering algorithm to divide the data-set into 3 clusters;
- (b) Agglomerative Hierarchical Clustering.
 - $oldsymbol{\Theta}$ k—means clustering algorithm to divide the data-set into 3 clusters

```
library (MASS)
dim(swiss)
## [1] 47 6
head(swiss)
##
                Fertility Agriculture Examination Education Catholic
## Courtelary
                     80.2
                                  17.0
                                                15
                                                           12
                                                                  9.96
                     83.1
                                  45.1
                                                 6
                                                            9
                                                                 84.84
## Delemont
                                  39.7
                                                 5
                                                            5
## Franches-Mnt
                     92.5
                                                                 93.40
                                                            7
## Moutier
                     85.8
                                  36.5
                                                12
                                                                 33.77
                                  43.5
                                                17
## Neuveville
                     76.9
                                                           15
                                                                  5.16
## Porrentruy
                     76.1
                                  35.3
                                                 9
                                                            7
                                                                 90.57
##
                Infant.Mortality
                             22.2
## Courtelary
## Delemont
                             22.2
## Franches-Mnt
                             20.2
## Moutier
                             20.3
## Neuveville
                             20.6
## Porrentruy
                             26.6
```

```
kmeans(swiss, 3)
## K-means clustering with 3 clusters of sizes 11, 20, 16
##
## Cluster means:
    Fertility Agriculture Examination Education Catholic Infant. Mortality
## 1 58.30909
                 19.50909
                             25.72727
                                         23.000 22.21455
                                                                19.22727
## 2 68.32500
                 55.90500
                             17.05000
                                         7.850 7.55000
                                                                19.67000
## 3 80.55000 65.51875 9.43750 6.625 96.15000
                                                                20.77500
```

```
##
## Clustering vector:
   Courtelary Delemont Franches-Mnt Moutier Neuveville Porrentruy
                              2 2 3
    1
Broye
            3 3
##
             Glane Gruyere Sarine 3 3 3
                                               Aigle
                              Sarine Veveyse
            3
##
     3
                                       3
     Aubonne Avenches Cossonay Echallens Grandson Lausanne
##
                     2
                             2
                                       2
     2.
            2
##
   La Vallee
##
             Lavaux
                      Morges
                              Moudon
                                       Nyone
                                                 Orbe
                                        2
             2
                     2
                               2
      1
       Oron Payerne Paysd'enhaut Rolle Vevey Yverdon 2 2 2 2 1 2
##
##
       2
    Conthey Entremont
                     Herens Martigwy Monthey St Maurice
##
##
     3
            3
                      3
                             3
                                       3
                                              3
            Sion Boudry La Chauxdfnd Le Locle Neuchatel
3 2 1 1 1
##
     Sierre
##
   Val de Ruz ValdeTravers V. De Geneve Rive Droite Rive Gauche
                                  1
    2
            1 1
##
##
## Within cluster sum of squares by cluster:
## [1] 9116.894 5966.297 6532.906
## (between SS / total SS = 81.8 %)
## Available components:
##
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"
## [6] "betweenss" "size" "iter" "ifault"
```

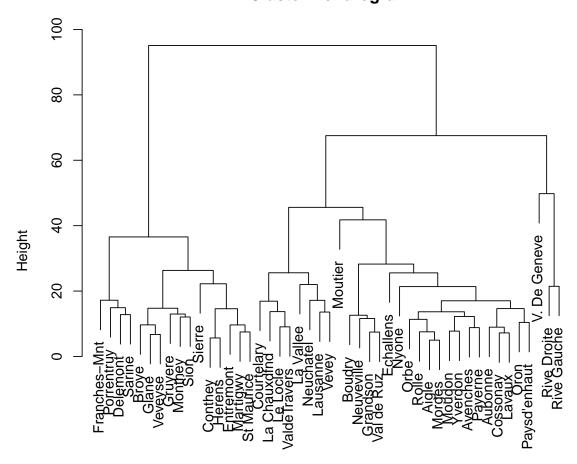
• Agglomerative Hierarchical Clustering

```
d <- dist(swiss)

x <- hclust(d, method = "average")

plot(x)</pre>
```

Cluster Dendrogram



d hclust (*, "average")