Εργασία 9

2024-05-20

Διερεύνηση συνόλου δεδομένων

```
library(readr)
airtraffic <- read_csv("https://query.data.world/s/vn6i4q4nfumfsdvrthabatxg25otdd?dws=00000", show_col_
airtraffic$`Operating Airline IATA Code` = NULL
airtraffic$`Published Airline IATA Code` = NULL
airtraffic$Month = NULL
airtraffic$Year = NULL
airtraffic = unique(airtraffic)
summary(airtraffic)
```

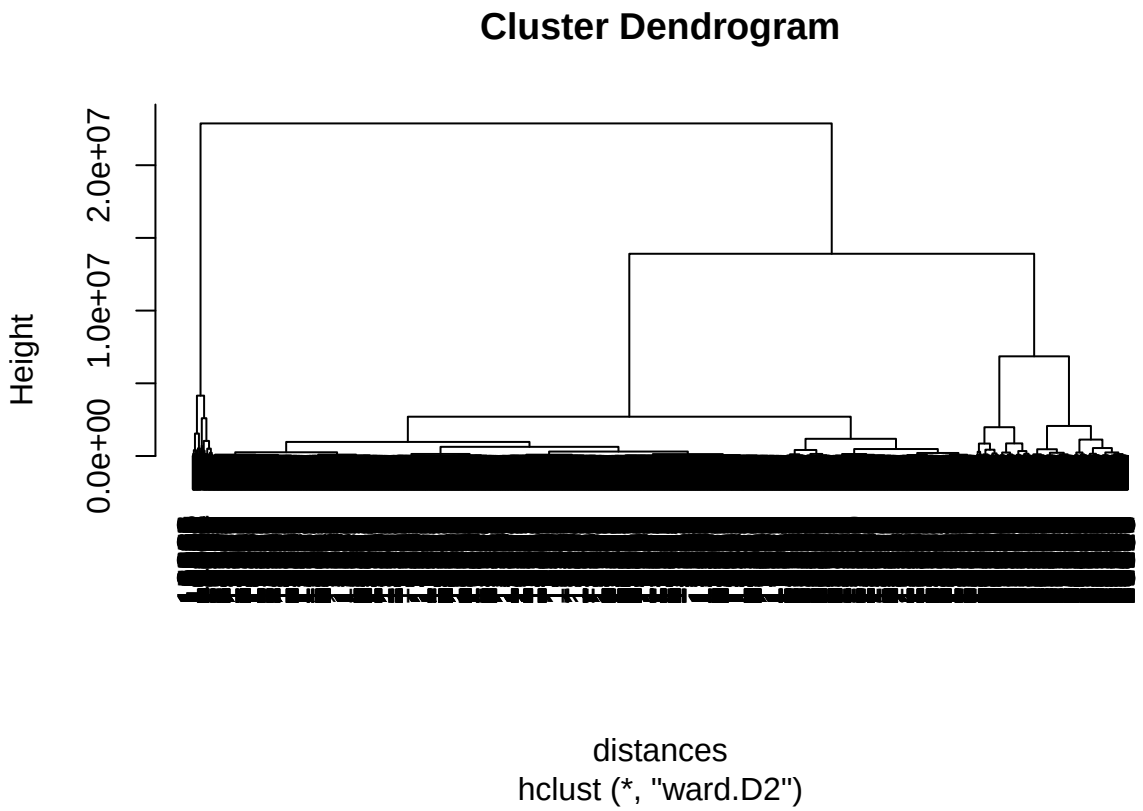
```
##  Activity Period  Operating Airline  Published Airline  GEO Summary
##  Min.   :200507   Length:15007       Length:15007       Length:15007
##  1st Qu.:200803   Class :character   Class :character   Class :character
##  Median :201011   Mode  :character   Mode  :character   Mode  :character
##  Mean   :201045
##  3rd Qu.:201308
##  Max.   :201603
##   GEO Region        Activity Type Code Price Category Code   Terminal
##  Length:15007       Length:15007       Length:15007        Length:15007
##  Class :character   Class :character   Class :character    Class :character
##  Mode  :character   Mode  :character   Mode  :character    Mode  :character
##
##
##
##  Boarding Area      Passenger Count  Adjusted Activity Type Code
##  Length:15007       Min.   :     1   Length:15007
##  Class :character   1st Qu.:  5374   Class :character
##  Mode  :character   Median :  9210   Mode  :character
##                     Mean   : 29241
##                     3rd Qu.: 21159
##                     Max.   :659837
##  Adjusted Passenger Count
##  Min.   :     1
##  1st Qu.:  5496
##  Median :  9354
##  Mean   : 29332
##  3rd Qu.: 21182
##  Max.   :659837
```

Ιεραρχική συσταδοποίηση

```
distances = dist(airtraffic[1:12], method = "euclidean")
```

```
## Warning in dist(airtraffic[1:12], method = "euclidean"): NAs introduced by
## coercion
```

```
clusterAirtraffic = hclust(distances, method = "ward.D2")
plot(clusterAirtraffic)
```




```
clusterGroups = cutree(clusterAirtraffic, k = 10)
cluster1 = subset(airtraffic , clusterGroups == 3)
View(cluster1)
```