Assignment_V

Meghana Udiga

11/30/2022

```
library(readr)
cereals <- read_csv("D:/Cereals.csv")
```

```
## Rows: 77 Columns: 16
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): name, mfr, type
## dbl (13): calories, protein, fat, sodium, fiber, carbo, sugars, potass, vita...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
Data_numerical <- data.frame(cereals[,4:16])
```

Libraries

```
library(cluster)
library(caret)
```

```
## Loading required package: ggplot2

## Loading required package: lattice
```

```
library(dendextend)
```

```
## Warning: package 'dendextend' was built under R version 4.2.2
```

```
##
## ---------------------
## Welcome to dendextend version 1.16.0
## Type citation('dendextend') for how to cite the package.
##
## Type browseVignettes(package = 'dendextend') for the package vignette.
## The github page is: https://github.com/talgalili/dendextend/
##
## Suggestions and bug-reports can be submitted at: https://github.com/talgalili/dendextend/issues
## You may ask questions at stackoverflow, use the r and dendextend tags:
##   https://stackoverflow.com/questions/tagged/dendextend
##
##  To suppress this message use:  suppressPackageStartupMessages(library(dendextend))
## ---------------------
```

```
##
## Attaching package: 'dendextend'

## The following object is masked from 'package:stats':
##
##     cutree
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.2.2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

Remove all cereals with missing values.

Removing missing values in present data.

```
missingvalues_removed <- na.omit(Data_numerical)
```

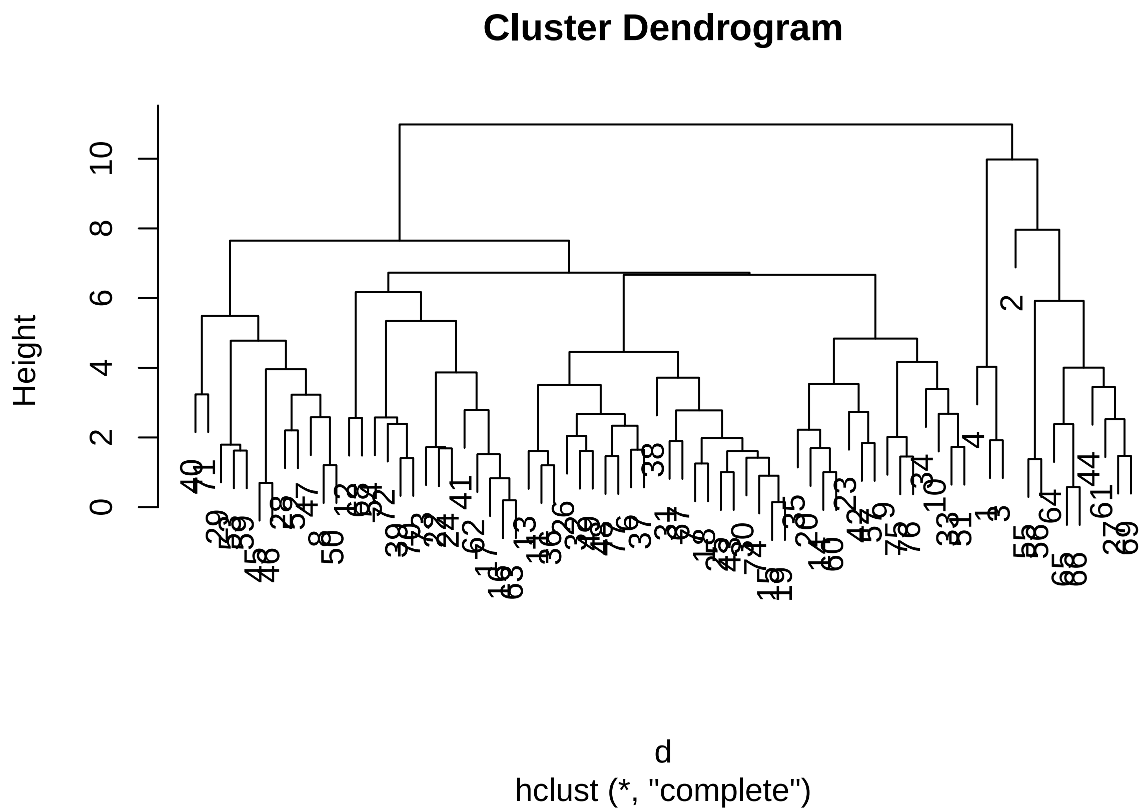
Data Normalization & Data Scaling:

```
Normalise <- scale(missingvalues_removed)
```

euclidean distance to measure the distance:

```
d <- dist(Normalise, method = "euclidean")
```

```
##Perform Hierarchical Clustering using complete linkage.
HC <- hclust(d, method = "complete")
plot(HC)
```




```
round(HC$height, 3)
```

```
##  [1]  0.143  0.196  0.575  0.698  0.828  0.904  1.003  1.004  1.201  1.203
## [11]  1.254  1.378  1.408  1.421  1.454  1.463  1.474  1.517  1.608  1.611
## [21]  1.616  1.625  1.650  1.687  1.692  1.720  1.730  1.795  1.839  1.897
## [31]  1.919  1.982  2.015  2.046  2.203  2.224  2.339  2.381  2.394  2.522
## [41]  2.563  2.574  2.579  2.668  2.682  2.734  2.776  2.787  3.229  3.236
## [51]  3.385  3.451  3.510  3.535  3.717  3.866  3.957  4.005  4.031  4.168
## [61]  4.456  4.779  4.839  5.342  5.488  5.920  6.169  6.669  6.731  7.650
## [71]  7.964  9.979 10.984
```

Determining Optimal Clusters:

We can also use agnes() function to perform clustering. Performing clustering using agnes() with single, complete, average and ward.

```
H_C_S <- agnes(Normalise, method = "single")
H_C_C <- agnes(Normalise, method = "complete")
H_C <- agnes(Normalise, method = "average")
H_C_W <- agnes(Normalise, method = "ward")
#Comparing the agglomerative coefficients for Single,complete, average and ward.
print(H_C_S$ac)
```

```
## [1] 0.6067859
```

```r
print(H_C_C$ac)
```

```
## [1] 0.8353712
```

```r
print(H_C$ac)
```

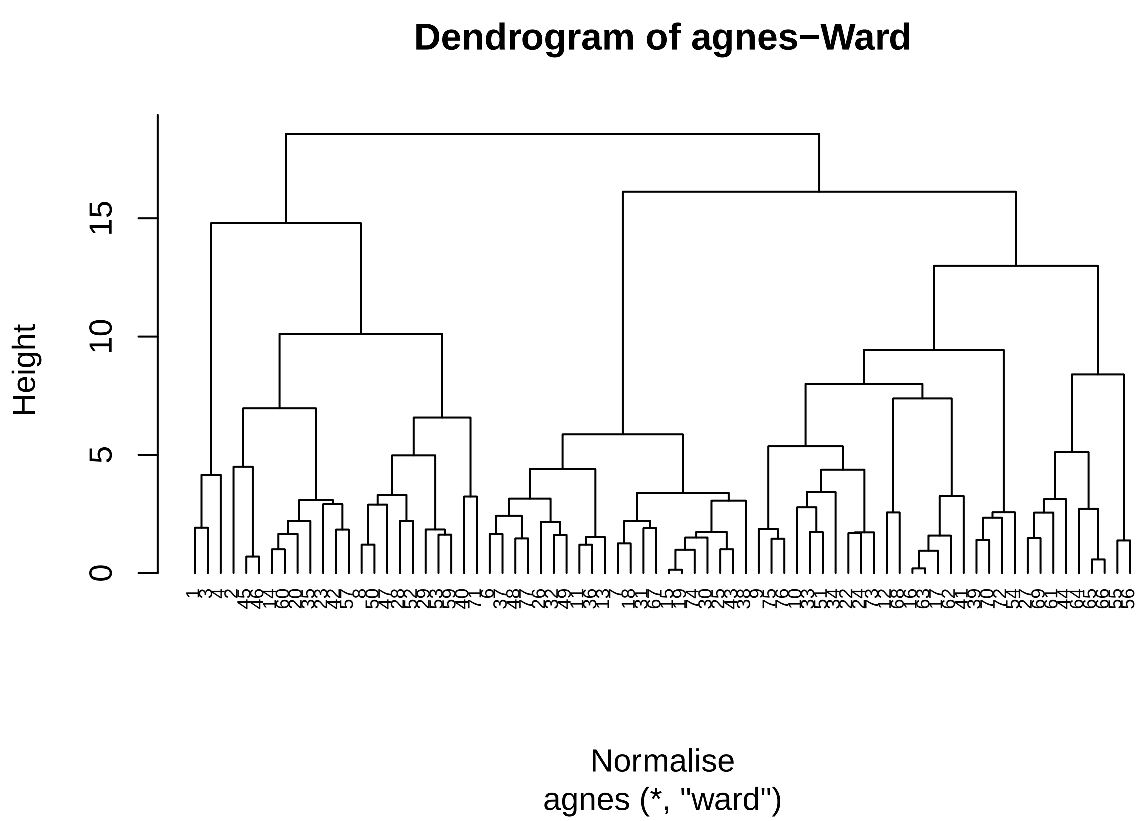
```
## [1] 0.7766075
```

```r
print(H_C_W$ac)
```

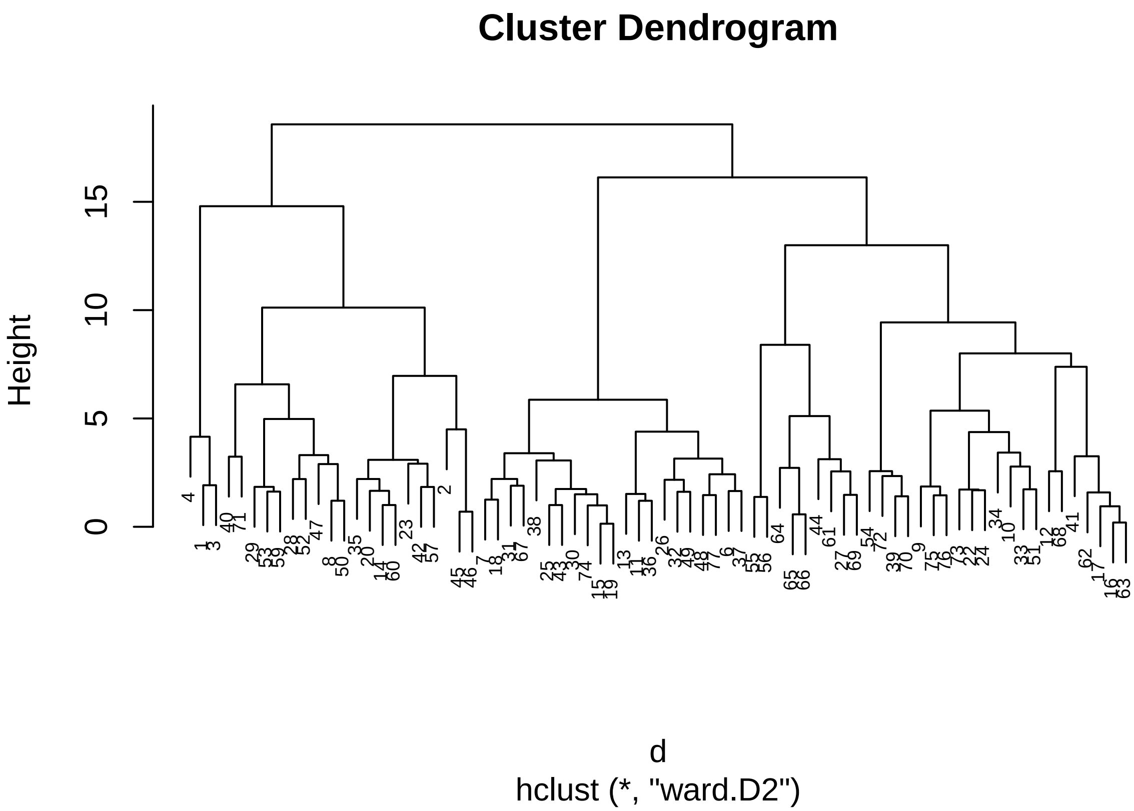
```
## [1] 0.9046042
```

According to the findings, the wards approach is the most effective with a value of 0.904. Using the Ward technique and cutting the dendrogram to plot the agnes. By measuring the distance, we may determine that k = 5.

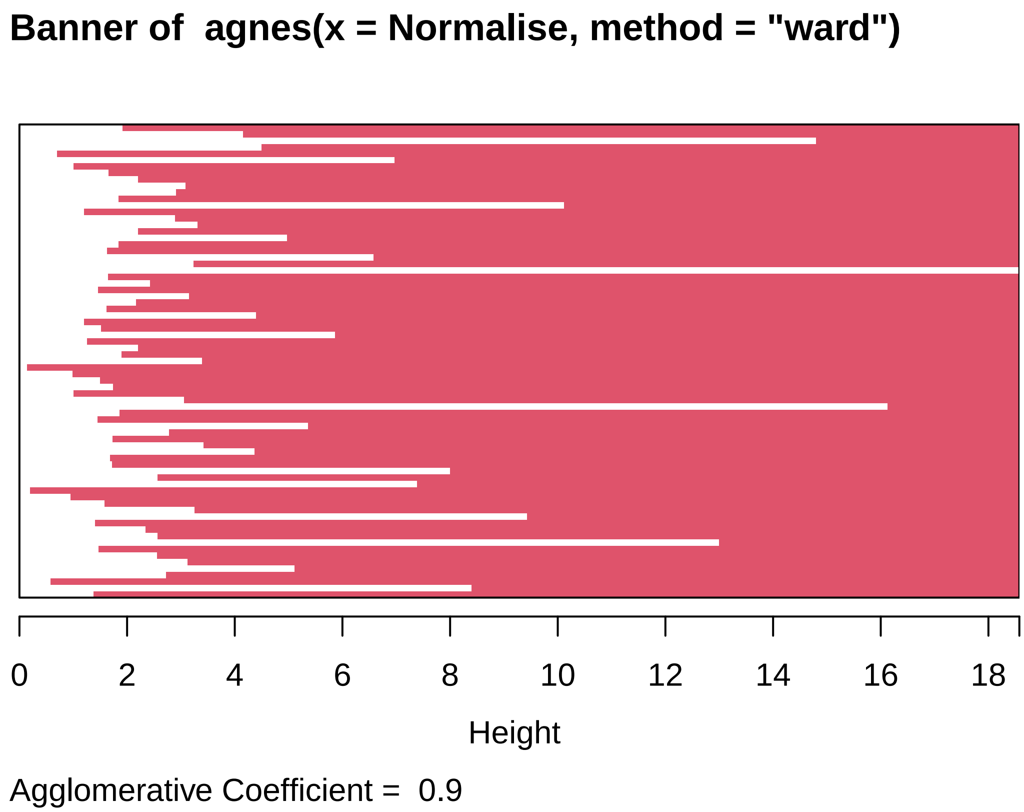
```r
pltree(H_C_W, cex = 0.6, hang = -1, main = "Dendrogram of agnes-Ward")
```




```r
# Hierarchical clustering using ward method.
HC1 <- hclust(d, method = "ward.D2" )
plot(HC1,cex=0.6)
```

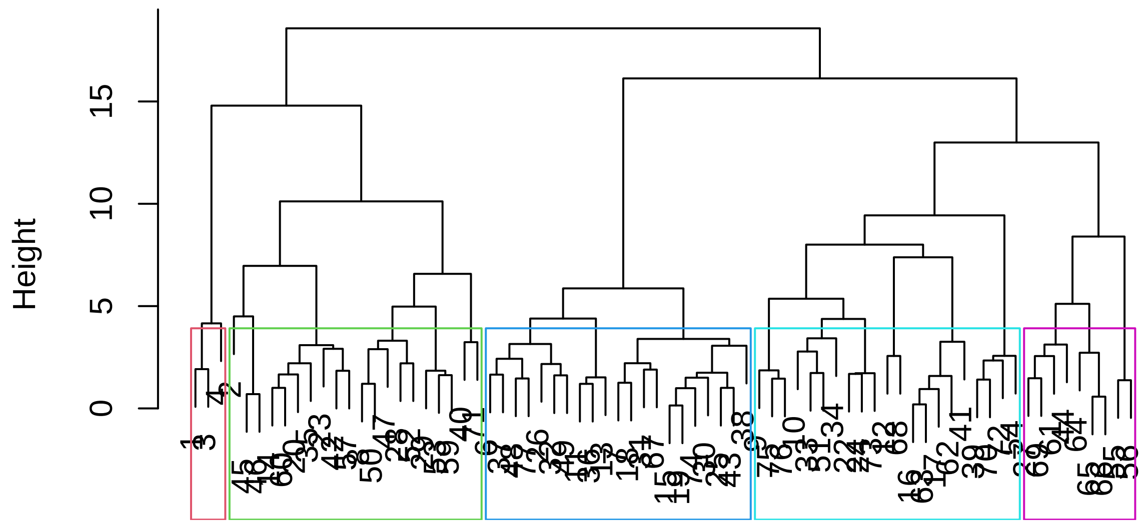



```
plot(H_C_W)
```




```
rect.hclust(H_C_W, k=5, border = 2:10)
```

Dendrogram of agnes(x = Normalise, method = "ward")



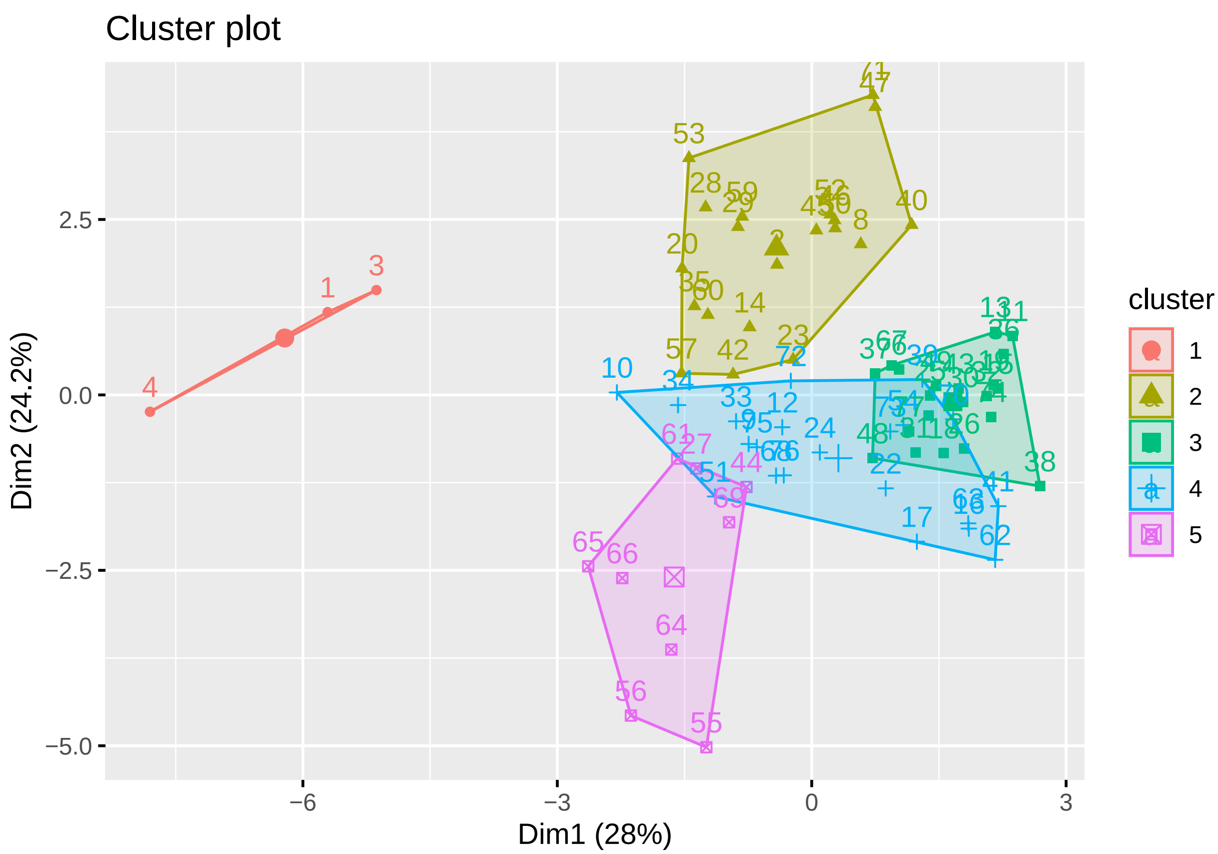
Normalise
Agglomerative Coefficient = 0.9

```
subgrp <- cutree(HC1, k = 5)
table(subgrp)
```

```
## subgrp
##  1  2  3  4  5
##  3 20 21 21  9
```

```
dataframe <- as.data.frame(cbind(Normalise,subgrp))
```

```
##To visualiZe the results in scatter plot.
fviz_cluster(list(data = Normalise, cluster = subgrp))
```



Selecting the cluster that is best cereal for breakfast, which will have high protein, fiber and low in sugar, sodium. Choosing the Cluster of Healthy Cereals.

```
Newdatacereals <- cereals
Newdatacereals_omission <- na.omit(Newdatacereals)
Clust <- cbind(Newdatacereals_omission, subgrp)
Clust[Clust$subgrp==1,]
```

```
##                        name mfr type calories protein fat sodium fiber carbo
## 1                 100%_Bran   N    C       70       4   1    130    10     5
## 3                  All-Bran   K    C       70       4   1    260     9     7
## 4 All-Bran_with_Extra_Fiber   K    C       50       4   0    140    14     8
##   sugars potass vitamins shelf weight cups   rating subgrp
## 1      6    280       25     3      1 0.33 68.40297      1
## 3      5    320       25     3      1 0.33 59.42551      1
## 4      0    330       25     3      1 0.50 93.70491      1
```

```
Clust[Clust$subgrp==2,]
```

```
##                                        name mfr type calories protein fat sodium
## 2                         100%_Natural_Bran   Q    C      120       3   5     15
## 8                                   Basic_4   G    C      130       3   2    210
## 14                                 Clusters   G    C      110       3   2    140
## 20                       Cracklin'_Oat_Bran   K    C      110       3   3    140
## 23                    Crispy_Wheat_&_Raisins   G    C      100       2   1    140
```

```
## 28 Fruit_&_Fibre_Dates,_Walnuts,_and_Oats   P    C      120       3   2    160
## 29                          Fruitful_Bran   K    C      120       3   0    240
## 35                      Great_Grains_Pecan   P    C      120       3   3     75
## 40                  Just_Right_Fruit_&_Nut   K    C      140       3   1    170
## 42                                    Life   Q    C      100       4   2    150
## 45       Muesli_Raisins,_Dates,_&_Almonds   R    C      150       4   3     95
## 46      Muesli_Raisins,_Peaches,_&_Pecans   R    C      150       4   3    150
## 47                    Mueslix_Crispy_Blend   K    C      160       3   2    150
## 50               Nutri-Grain_Almond-Raisin   K    C      140       3   2    220
## 52                    Oatmeal_Raisin_Crisp   G    C      130       3   2    170
## 53                   Post_Nat._Raisin_Bran   P    C      120       3   1    200
## 57                      Quaker_Oat_Squares   Q    C      100       4   1    135
## 59                             Raisin_Bran   K    C      120       3   1    210
## 60                         Raisin_Nut_Bran   G    C      100       3   2    140
## 71                       Total_Raisin_Bran   G    C      140       3   1    190
##    fiber carbo sugars potass vitamins shelf weight cups   rating subgrp
## 2    2.0   8.0      8    135        0     3   1.00 1.00 33.98368      2
## 8    2.0  18.0      8    100       25     3   1.33 0.75 37.03856      2
## 14   2.0  13.0      7    105       25     3   1.00 0.50 40.40021      2
## 20   4.0  10.0      7    160       25     3   1.00 0.50 40.44877      2
## 23   2.0  11.0     10    120       25     3   1.00 0.75 36.17620      2
## 28   5.0  12.0     10    200       25     3   1.25 0.67 40.91705      2
## 29   5.0  14.0     12    190       25     3   1.33 0.67 41.01549      2
## 35   3.0  13.0      4    100       25     3   1.00 0.33 45.81172      2
## 40   2.0  20.0      9     95      100     3   1.30 0.75 36.47151      2
## 42   2.0  12.0      6     95       25     2   1.00 0.67 45.32807      2
## 45   3.0  16.0     11    170       25     3   1.00 1.00 37.13686      2
## 46   3.0  16.0     11    170       25     3   1.00 1.00 34.13976      2
## 47   3.0  17.0     13    160       25     3   1.50 0.67 30.31335      2
## 50   3.0  21.0      7    130       25     3   1.33 0.67 40.69232      2
## 52   1.5  13.5     10    120       25     3   1.25 0.50 30.45084      2
## 53   6.0  11.0     14    260       25     3   1.33 0.67 37.84059      2
## 57   2.0  14.0      6    110       25     3   1.00 0.50 49.51187      2
## 59   5.0  14.0     12    240       25     2   1.33 0.75 39.25920      2
## 60   2.5  10.5      8    140       25     3   1.00 0.50 39.70340      2
## 71   4.0  15.0     14    230      100     3   1.50 1.00 28.59278      2
```

```
Clust[Clust$subgrp==3,]
```

```
##                        name mfr type calories protein fat sodium fiber carbo
## 6  Apple_Cinnamon_Cheerios   G    C      110       2   2    180   1.5  10.5
## 7              Apple_Jacks   K    C      110       2   0    125   1.0  11.0
## 11            Cap'n'Crunch   Q    C      120       1   2    220   0.0  12.0
## 13    Cinnamon_Toast_Crunch   G    C      120       1   3    210   0.0  13.0
## 15              Cocoa_Puffs   G    C      110       1   1    180   0.0  12.0
## 18                Corn_Pops   K    C      110       1   0     90   1.0  13.0
## 19            Count_Chocula   G    C      110       1   1    180   0.0  12.0
## 25              Froot_Loops   K    C      110       2   1    125   1.0  11.0
## 26           Frosted_Flakes   K    C      110       1   0    200   1.0  14.0
## 30           Fruity_Pebbles   P    C      110       1   1    135   0.0  13.0
## 31             Golden_Crisp   P    C      100       2   0     45   0.0  11.0
## 32           Golden_Grahams   G    C      110       1   1    280   0.0  15.0
## 36          Honey_Graham_Ohs   Q    C      120       1   2    220   1.0  12.0
## 37        Honey_Nut_Cheerios   G    C      110       3   1    250   1.5  11.5
```

```
## 38              Honey-comb   P    C      110       1   0    180   0.0  14.0
## 43            Lucky_Charms   G    C      110       2   1    180   0.0  12.0
## 48    Multi-Grain_Cheerios   G    C      100       2   1    220   2.0  15.0
## 49        Nut&Honey_Crunch   K    C      120       2   1    190   0.0  15.0
## 67                  Smacks   K    C      110       2   1     70   1.0   9.0
## 74                    Trix   G    C      110       1   1    140   0.0  13.0
## 77     Wheaties_Honey_Gold   G    C      110       2   1    200   1.0  16.0
##    sugars potass vitamins shelf weight cups   rating subgrp
## 6      10     70       25     1      1 0.75 29.50954      3
## 7      14     30       25     2      1 1.00 33.17409      3
## 11     12     35       25     2      1 0.75 18.04285      3
## 13      9     45       25     2      1 0.75 19.82357      3
## 15     13     55       25     2      1 1.00 22.73645      3
## 18     12     20       25     2      1 1.00 35.78279      3
## 19     13     65       25     2      1 1.00 22.39651      3
## 25     13     30       25     2      1 1.00 32.20758      3
## 26     11     25       25     1      1 0.75 31.43597      3
## 30     12     25       25     2      1 0.75 28.02576      3
## 31     15     40       25     1      1 0.88 35.25244      3
## 32      9     45       25     2      1 0.75 23.80404      3
## 36     11     45       25     2      1 1.00 21.87129      3
## 37     10     90       25     1      1 0.75 31.07222      3
## 38     11     35       25     1      1 1.33 28.74241      3
## 43     12     55       25     2      1 1.00 26.73451      3
## 48      6     90       25     1      1 1.00 40.10596      3
## 49      9     40       25     2      1 0.67 29.92429      3
## 67     15     40       25     2      1 0.75 31.23005      3
## 74     12     25       25     2      1 1.00 27.75330      3
## 77      8     60       25     1      1 0.75 36.18756      3
```

```
Clust[Clust$subgrp==4,]
```

```
##                           name mfr type calories protein fat sodium fiber carbo
## 9                    Bran_Chex   R    C       90       2   1    200     4    15
## 10                 Bran_Flakes   P    C       90       3   0    210     5    13
## 12                    Cheerios   G    C      110       6   2    290     2    17
## 16                   Corn_Chex   R    C      110       2   0    280     0    22
## 17                 Corn_Flakes   K    C      100       2   0    290     1    21
## 22                     Crispix   K    C      110       2   0    220     1    21
## 24                 Double_Chex   R    C      100       2   0    190     1    18
## 33           Grape_Nuts_Flakes   P    C      100       3   1    140     3    15
## 34                  Grape-Nuts   P    C      110       3   0    170     3    17
## 39 Just_Right_Crunchy__Nuggets   K    C      110       2   1    170     1    17
## 41                         Kix   G    C      110       2   1    260     0    21
## 51           Nutri-grain_Wheat   K    C       90       3   0    170     3    18
## 54                  Product_19   K    C      100       3   0    320     1    20
## 62                   Rice_Chex   R    C      110       1   0    240     0    23
## 63               Rice_Krispies   K    C      110       2   0    290     0    22
## 68                   Special_K   K    C      110       6   0    230     1    16
## 70           Total_Corn_Flakes   G    C      110       2   1    200     0    21
## 72           Total_Whole_Grain   G    C      100       3   1    200     3    16
## 73                     Triples   G    C      110       2   1    250     0    21
## 75                  Wheat_Chex   R    C      100       3   1    230     3    17
## 76                    Wheaties   G    C      100       3   1    200     3    17
```

```
##    sugars potass vitamins shelf weight cups   rating subgrp
## 9       6    125       25     1      1 0.67 49.12025      4
## 10      5    190       25     3      1 0.67 53.31381      4
## 12      1    105       25     1      1 1.25 50.76500      4
## 16      3     25       25     1      1 1.00 41.44502      4
## 17      2     35       25     1      1 1.00 45.86332      4
## 22      3     30       25     3      1 1.00 46.89564      4
## 24      5     80       25     3      1 0.75 44.33086      4
## 33      5     85       25     3      1 0.88 52.07690      4
## 34      3     90       25     3      1 0.25 53.37101      4
## 39      6     60      100     3      1 1.00 36.52368      4
## 41      3     40       25     2      1 1.50 39.24111      4
## 51      2     90       25     3      1 1.00 59.64284      4
## 54      3     45      100     3      1 1.00 41.50354      4
## 62      2     30       25     1      1 1.13 41.99893      4
## 63      3     35       25     1      1 1.00 40.56016      4
## 68      3     55       25     1      1 1.00 53.13132      4
## 70      3     35      100     3      1 1.00 38.83975      4
## 72      3    110      100     3      1 1.00 46.65884      4
## 73      3     60       25     3      1 0.75 39.10617      4
## 75      3    115       25     1      1 0.67 49.78744      4
## 76      3    110       25     1      1 1.00 51.59219      4
```

```
Clust[Clust$subgrp==5,]
```

```
##                         name mfr type calories protein fat sodium fiber carbo
## 27       Frosted_Mini-Wheats   K    C      100       3   0      0     3    14
## 44                     Maypo   A    H      100       4   1      0     0    16
## 55               Puffed_Rice   Q    C       50       1   0      0     0    13
## 56              Puffed_Wheat   Q    C       50       2   0      0     1    10
## 61            Raisin_Squares   K    C       90       2   0      0     2    15
## 64            Shredded_Wheat   N    C       80       2   0      0     3    16
## 65    Shredded_Wheat_'n'Bran   N    C       90       3   0      0     4    19
## 66 Shredded_Wheat_spoon_size   N    C       90       3   0      0     3    20
## 69   Strawberry_Fruit_Wheats   N    C       90       2   0     15     3    15
##    sugars potass vitamins shelf weight cups   rating subgrp
## 27      7    100       25     2   1.00 0.80 58.34514      5
## 44      3     95       25     2   1.00 1.00 54.85092      5
## 55      0     15        0     3   0.50 1.00 60.75611      5
## 56      0     50        0     3   0.50 1.00 63.00565      5
## 61      6    110       25     3   1.00 0.50 55.33314      5
## 64      0     95        0     1   0.83 1.00 68.23588      5
## 65      0    140        0     1   1.00 0.67 74.47295      5
## 66      0    120        0     1   1.00 0.67 72.80179      5
## 69      5     90       25     2   1.00 1.00 59.36399      5
```

Calculating mean ratings to determine the best cluster.

```
mean(Clust[Clust$subgrp==1,"rating"])
```

```
## [1] 73.84446
```

```r
mean(Clust[Clust$subgrp==2,"rating"])
```

```
## [1] 38.26161
```

```r
mean(Clust[Clust$subgrp==3,"rating"])
```

```
## [1] 28.84825
```

```r
mean(Clust[Clust$subgrp==4,"rating"])
```

```
## [1] 46.46513
```

```r
mean(Clust[Clust$subgrp==5,"rating"])
```

```
## [1] 63.0184
```

It is advisable to choose cluster 1 and the cereals in cluster 1 for a healthy diet, as we can see that the mean ratings for the subgrp==1 are the highest (73.84).