

Assignment_5

Kevin Gardner

Due 4/17/2022

Following is the link to my GitHub account:

https://github.com/Kgardner22/64060_-kgardner

IMPORT THE DATA:

```
cereals.df <- read.csv('C:/R/MyData/Cereals.csv', header = T, sep = ',')

summary(cereals.df)
```

##	name	mfr	type	calories
##	Length:77	Length:77	Length:77	Min. : 50.0
##	Class :character	Class :character	Class :character	1st Qu.:100.0
##	Mode :character	Mode :character	Mode :character	Median :110.0
##				Mean :106.9
##				3rd Qu.:110.0
##				Max. :160.0
##				
##	protein	fat	sodium	fiber
##	Min. :1.000	Min. :0.000	Min. : 0.0	Min. : 0.000
##	1st Qu.:2.000	1st Qu.:0.000	1st Qu.:130.0	1st Qu.: 1.000
##	Median :3.000	Median :1.000	Median :180.0	Median : 2.000
##	Mean :2.545	Mean :1.013	Mean :159.7	Mean : 2.152
##	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:210.0	3rd Qu.: 3.000
##	Max. :6.000	Max. :5.000	Max. :320.0	Max. :14.000
##				
##	carbo	sugars	potass	vitamins
##	Min. : 5.0	Min. : 0.000	Min. : 15.00	Min. : 0.00
##	1st Qu.:12.0	1st Qu.: 3.000	1st Qu.: 42.50	1st Qu.: 25.00
##	Median :14.5	Median : 7.000	Median : 90.00	Median : 25.00
##	Mean :14.8	Mean : 7.026	Mean : 98.67	Mean : 28.25
##	3rd Qu.:17.0	3rd Qu.:11.000	3rd Qu.:120.00	3rd Qu.: 25.00

```
## Max. :23.0 Max. :15.000 Max. :330.00 Max. :100.00
## NA's :1 NA's :1 NA's :2
## shelf weight cups rating
## Min. :1.000 Min. :0.50 Min. :0.250 Min. :18.04
## 1st Qu.:1.000 1st Qu.:1.00 1st Qu.:0.670 1st Qu.:33.17
## Median :2.000 Median :1.00 Median :0.750 Median :40.40
## Mean :2.208 Mean :1.03 Mean :0.821 Mean :42.67
## 3rd Qu.:3.000 3rd Qu.:1.00 3rd Qu.:1.000 3rd Qu.:50.83
## Max. :3.000 Max. :1.50 Max. :1.500 Max. :93.70
##
```

REMOVE ALL CEREALS WITH MISSING VALUES:

```
cereals.df <- na.omit(cereals.df) #Remove NA (missing) values
```

SET ROW NAMES IN THE DATAFRAME:

```
# set row names to the name column
row.names(cereals.df) <- cereals.df[,1]

# remove the name column as a variable
cereals.df <- cereals.df[,-1]
```

NORMALIZE THE DATA:

```
# normalize all numeric variables (Columns 3 - 15)
cereals.df.norm <- apply(cereals.df[,c(3:15)], scale)

# add row names: cereals
row.names(cereals.df.norm) <- row.names(cereals.df)
```

APPLY HIERARCHICAL CLUSTERING USING AGNES AND FOUR LINKAGE MEASURES:

Apply hierarchical clustering to the data using Euclidean distance to the normalized measurements.

```
# Load the cluster package so we can use agnes
library(cluster)

# compute normalized distance based on all numeric variables
d.norm <- dist(cereals.df.norm, method = "euclidean")

hc_single <- agnes(d.norm, method = "single") # uses single linkage
hc_complete <- agnes(d.norm, method = "complete") #uses complete linkage
hc_average <- agnes(d.norm, method = "average") #uses average linkage
hc_ward <- agnes(d.norm, method = "ward") #uses Ward's method
```

COMPARE AGGLOMERATIVE COEFFICIENTS

```
print(hc_single$ac)

## [1] 0.6067859
```

```
print(hc_complete)

## Call:      agnes(x = d.norm, method = "complete")
## Agglomerative coefficient: 0.8353712
## Order of objects:
## [1] 100%_Bran
## [2] All-Bran
## [3] All-Bran_with_Extra_Fiber
## [4] 100%_Natural_Bran
## [5] Frosted_Mini-Wheats
## [6] Strawberry_Fruit_Wheats
## [7] Raisin_Squares
## [8] Maypo
## [9] Shredded_Wheat
## [10] Shredded_Wheat_'n'Bran
## [11] Shredded_Wheat_spoon_size
## [12] Puffed_Rice
## [13] Puffed_Wheat
## [14] Apple_Cinnamon_Cheerios
## [15] Honey_Nut_Cheerios
## [16] Multi-Grain_Cheerios
## [17] Wheaties_Honey_Gold
## [18] Frosted_Flakes
## [19] Golden_Grahams
## [20] Nut&Honey_Crunch
## [21] Cap'n'Crunch
## [22] Honey_Graham_Ohs
## [23] Cinnamon_Toast_Crunch
## [24] Apple_Jacks
## [25] Corn_Pops
## [26] Cocoa_Puffs
## [27] Count_Chocula
## [28] Trix
## [29] Fruity_Pebbles
## [30] Froot_Loops
## [31] Lucky_Charms
## [32] Golden_Crisp
## [33] Smacks
## [34] Honey-comb
## [35] Bran_Chex
## [36] Wheat_Chex
## [37] Wheaties
## [38] Bran_Flakes
## [39] Grape_Nuts_Flakes
## [40] Nutri-grain_Wheat
## [41] Grape-Nuts
## [42] Clusters
## [43] Raisin_Nut_Bran
## [44] Cracklin'_Oat_Bran
## [45] Great_Grains_Pecan
```

```

## [46] Crispy_Wheat_&_Raisins
## [47] Life
## [48] Quaker_Oat_Squares
## [49] Cheerios
## [50] Special_K
## [51] Corn_Chex
## [52] Rice_Krispies
## [53] Corn_Flakes
## [54] Rice_Chex
## [55] Kix
## [56] Crispix
## [57] Double_Chex
## [58] Triples
## [59] Just_Right_Crunchy__Nuggets
## [60] Total_Corn_Flakes
## [61] Total_Whole_Grain
## [62] Product_19
## [63] Basic_4
## [64] Nutri-Grain_Almond-Raisin
## [65] Mueslix_Crispy_Blend
## [66] Fruit_&_Fibre_Dates,_Walnuts,_and_Oats
## [67] Oatmeal_Raisin_Crisp
## [68] Muesli_Raisins,_Dates,_&_Almonds
## [69] Muesli_Raisins,_Peaches,_&_Pecans
## [70] Fruitful_Bran
## [71] Post_Nat._Raisin_Bran
## [72] Raisin_Bran
## [73] Just_Right_Fruit_&_Nut
## [74] Total_Raisin_Bran
## Height (summary):
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1431  1.6076  2.3389  2.9321  3.7169 10.9839
##
## Available components:
## [1] "order"      "height"     "ac"         "merge"      "diss"       "call"
## [7] "method"     "order.lab"

print(hc_average)

## Call:      agnes(x = d.norm, method = "average")
## Agglomerative coefficient: 0.7766075
## Order of objects:
##  [1] 100%_Bran
##  [2] All-Bran
##  [3] All-Bran_with_Extra_Fiber
##  [4] 100%_Natural_Bran
##  [5] Apple_Cinnamon_Cheerios
##  [6] Honey_Nut_Cheerios
##  [7] Multi-Grain_Cheerios
##  [8] Wheaties_Honey_Gold

```

```
## [9] Frosted_Flakes
## [10] Golden_Grahams
## [11] Nut&Honey_Crunch
## [12] Apple_Jacks
## [13] Corn_Pops
## [14] Cocoa_Puffs
## [15] Count_Chocula
## [16] Trix
## [17] Froot_Loops
## [18] Lucky_Charms
## [19] Fruity_Pebbles
## [20] Golden_Crisp
## [21] Smacks
## [22] Cap'n'Crunch
## [23] Honey_Graham_Ohs
## [24] Cinnamon_Toast_Crunch
## [25] Honey-comb
## [26] Bran_Chex
## [27] Wheat_Chex
## [28] Wheaties
## [29] Bran_Flakes
## [30] Grape-Nuts
## [31] Crispix
## [32] Double_Chex
## [33] Triples
## [34] Grape_Nuts_Flakes
## [35] Nutri-grain_Wheat
## [36] Clusters
## [37] Raisin_Nut_Bran
## [38] Cracklin'_Oat_Bran
## [39] Great_Grains_Pecan
## [40] Life
## [41] Quaker_Oat_Squares
## [42] Crispy_Wheat_&_Raisins
## [43] Corn_Chex
## [44] Rice_Krispies
## [45] Corn_Flakes
## [46] Rice_Chex
## [47] Kix
## [48] Frosted_Mini-Wheats
## [49] Strawberry_Fruit_Wheats
## [50] Raisin_Squares
## [51] Maypo
## [52] Shredded_Wheat
## [53] Shredded_Wheat_'n'Bran
## [54] Shredded_Wheat_spoon_size
## [55] Basic_4
## [56] Nutri-Grain_Almond-Raisin
## [57] Oatmeal_Raisin_Crisp
## [58] Mueslix_Crispy_Blend
```

```

## [59] Fruit_&_Fibre_Dates,_Walnuts,_and_Oats
## [60] Fruitful_Bran
## [61] Post_Nat._Raisin_Bran
## [62] Raisin_Bran
## [63] Muesli_Raisins,_Dates,_&_Almonds
## [64] Muesli_Raisins,_Peaches,_&_Pecans
## [65] Just_Right_Crunchy__Nuggets
## [66] Total_Corn_Flakes
## [67] Total_Whole_Grain
## [68] Product_19
## [69] Just_Right_Fruit_&_Nut
## [70] Total_Raisin_Bran
## [71] Cheerios
## [72] Special_K
## [73] Puffed_Rice
## [74] Puffed_Wheat
## Height (summary):
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1431  1.4633  2.0666  2.4461  2.9445  7.7243
##
## Available components:
## [1] "order"      "height"     "ac"         "merge"      "diss"       "call"
## [7] "method"     "order.lab"

print(hc_ward)

## Call:      agnes(x = d.norm, method = "ward")
## Agglomerative coefficient: 0.9046042
## Order of objects:
## [1] 100%_Bran
## [2] All-Bran
## [3] All-Bran_with_Extra_Fiber
## [4] 100%_Natural_Bran
## [5] Muesli_Raisins,_Dates,_&_Almonds
## [6] Muesli_Raisins,_Peaches,_&_Pecans
## [7] Clusters
## [8] Raisin_Nut_Bran
## [9] Cracklin'_Oat_Bran
## [10] Great_Grains_Pecan
## [11] Crispy_Wheat_&_Raisins
## [12] Life
## [13] Quaker_Oat_Squares
## [14] Basic_4
## [15] Nutri-Grain_Almond-Raisin
## [16] Mueslix_Crispy_Blend
## [17] Fruit_&_Fibre_Dates,_Walnuts,_and_Oats
## [18] Oatmeal_Raisin_Crisp
## [19] Fruitful_Bran
## [20] Post_Nat._Raisin_Bran
## [21] Raisin_Bran

```

```
## [22] Just_Right_Fruit_&_Nut
## [23] Total_Raisin_Bran
## [24] Apple_Cinnamon_Cheerios
## [25] Honey_Nut_Cheerios
## [26] Multi-Grain_Cheerios
## [27] Wheaties_Honey_Gold
## [28] Frosted_Flakes
## [29] Golden_Grahams
## [30] Nut&Honey_Crunch
## [31] Cap'n'Crunch
## [32] Honey_Graham_Ohs
## [33] Cinnamon_Toast_Crunch
## [34] Apple_Jacks
## [35] Corn_Pops
## [36] Golden_Crisp
## [37] Smacks
## [38] Cocoa_Puffs
## [39] Count_Chocula
## [40] Trix
## [41] Fruity_Pebbles
## [42] Froot_Loops
## [43] Lucky_Charms
## [44] Honey-comb
## [45] Bran_Chex
## [46] Wheat_Chex
## [47] Wheaties
## [48] Bran_Flakes
## [49] Grape_Nuts_Flakes
## [50] Nutri-grain_Wheat
## [51] Grape-Nuts
## [52] Crispix
## [53] Double_Chex
## [54] Triples
## [55] Cheerios
## [56] Special_K
## [57] Corn_Chex
## [58] Rice_Krispies
## [59] Corn_Flakes
## [60] Rice_Chex
## [61] Kix
## [62] Just_Right_Crunchy__Nuggets
## [63] Total_Corn_Flakes
## [64] Total_Whole_Grain
## [65] Product_19
## [66] Frosted_Mini-Wheats
## [67] Strawberry_Fruit_Wheats
## [68] Raisin_Squares
## [69] Maypo
## [70] Shredded_Wheat
## [71] Shredded_Wheat_'n'Bran
```

```
## [72] Shredded_Wheat_spoon_size
## [73] Puffed_Rice
## [74] Puffed_Wheat
## Height (summary):
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  0.1431  1.5858  2.3422  3.6092  4.1559 18.5749
##
## Available components:
## [1] "order"      "height"     "ac"         "merge"      "diss"       "call"
## [7] "method"     "order.lab"
```

COMPARE THE RESULTS:

Compare the results of the Agglomerative coefficients (AC) of the four methods.

Single Linkage (hc_single): AC = 0.6067859

Complete Linkage (hc_complete): AC = 0.8353712

Average Linkage (hc_average): AC = 0.7766075

Ward's Method (hc_ward): AC = 0.9046042

In comparing the Agglomerative coefficients (AC), we see Ward's Method is the best linkage method as it results in the highest Agglomerative coefficient value.

PLOT THE DENDROGRAM:

To help answer the question of “how many clusters would you choose”, let's first review the dendrogram using Ward's Method for hierarchical clustering since this proved to be the best linkage method.

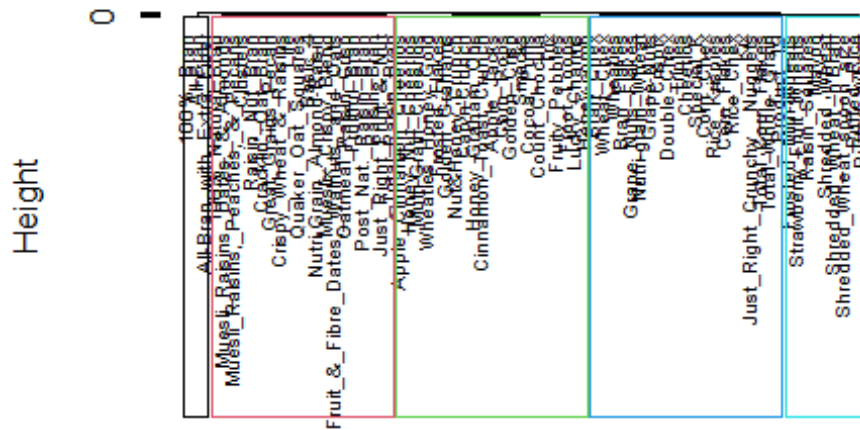
```
pltree(hc_ward, cex = 0.6, hang = -1, main = "Dendrogram of AGNES")
```


[illegible]

In reviewing the setting a cutoff of

```
pltree(hc_ward, cex = 0.6, hang = -1, main = "Dendrogram of AGNES")
rect.hclust(hc_ward, k=5, border = 1:5)
```

Dendrogram of AGNES



d.norm
agnes (*, "ward")

COMPUTE CLUSTER MEMBERSHIP BY "CUTTING" THE DENDROGRAM

```
memb <- cutree(hc_ward, k = 5)
memb

## [1] 1 2 1 1 3 3 2 4 4 3 4 3 2 3 4 4 3 3 2 4 2 4 3 3 5 2 2 3 3 3 4 4 2 3 3
## [39] 4 2 3 5 2 2 2 3 3 2 4 2 2 4 5 5 2 2 2 5 4 4 5 5 5 3 4 5 4 2 4 4 3 4 4
3
```

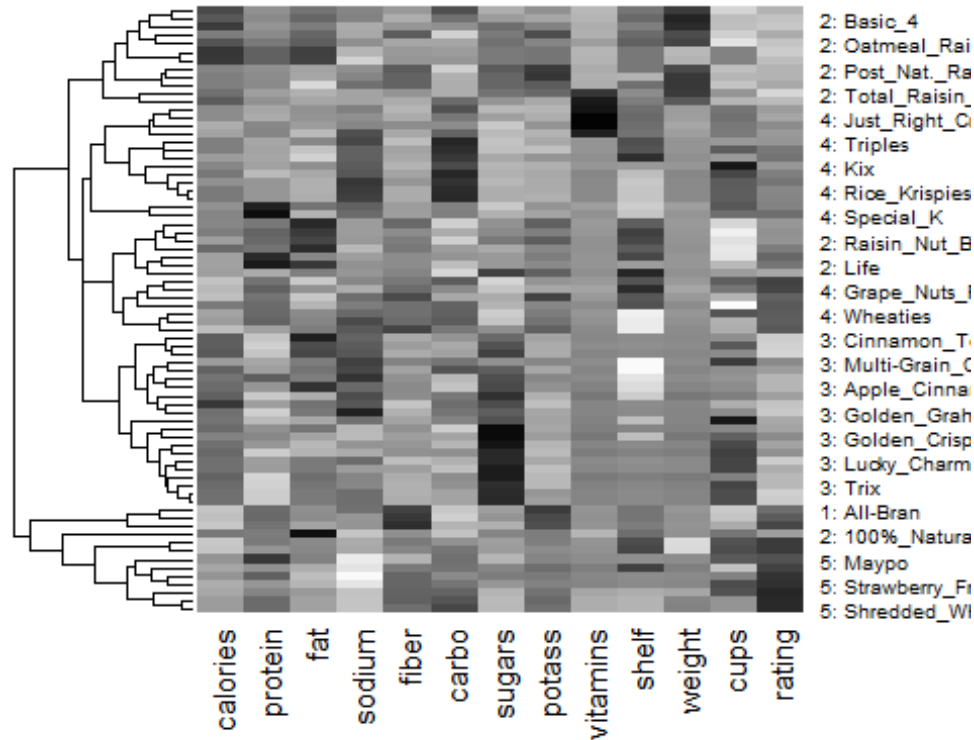
CREATING A HEATMAP

The following will create a heatmap for the 74 cereals (in rows). The rows are sorted by the five clusters from the Ward linkage clustering. Darker cells denote higher values within a column

```
# set labels as cluster membership and cereal name
row.names(cereals.df.norm) <- paste(memb, ":", row.names(cereals.df), sep =
"")

# plot heatmap
# rev() reverses the color mapping to large = dark
heatmap(as.matrix(cereals.df.norm), Colv = NA,
        col=rev(paste("gray",1:99,sep="")))

```



INTERPRETING THE DATA:

In reviewing the 13 variables for all 74 cereals by cluster, we can identify the following cluster commonalities:

Cluster 1: High in Fiber and Potassium with good customer satisfaction rating
 Cluster 2: High in Calories and Fat (and lowest in Vitamins)
 Cluster 3: High in Sugars and Calories
 Cluster 4: High in Carbohydrates ("Carbs") and slightly higher in Sodium
 Cluster 5: Higher in Protein, Fiber, Carbs, and Potassium and highest in Customer Satisfaction Rating

The elementary public schools would like to identify a group/cluster of "healthy cereals" to include in their daily cafeterias. To make this recommendation, we first need to understand which factors are desirable to consider a cereal "healthy".

A google search reveals "healthy cereals" should include: * Whole grains * High in Fiber, protein and nutrients/vitamins * Carbs are a main source of energy and help fuel our brains and vital organs

Based on our above interpretation of the five (5) clusters, the healthiest group of cereals is:

Cluster 5:

This cluster has higher levels of protein, fiber, carbs, and potassium while having the highest customer satisfaction ratings. There are nine cereals in this grouping providing a good selection of "healthy cereals" for elementary children.