ML bpalazzo\_5

Brandon Palazzo

11/30/2020

Data Cleaning and Manipulation

library(tidyverse)

## -- Attaching packages --------------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2 v purrr 0.3.4  
## v tibble 3.0.3 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(cluster)   
library(factoextra)

## Warning: package 'factoextra' was built under R version 4.0.3

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

library(dendextend)

## Warning: package 'dendextend' was built under R version 4.0.3

##   
## ---------------------  
## Welcome to dendextend version 1.14.0  
## 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  
## Or contact: <tal.galili@gmail.com>  
##   
## To suppress this message use: suppressPackageStartupMessages(library(dendextend))  
## ---------------------

##   
## Attaching package: 'dendextend'

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

library(caTools)

## Warning: package 'caTools' was built under R version 4.0.3

library(clValid)

## Warning: package 'clValid' was built under R version 4.0.3

cereals <- read.csv("Cereals.csv")  
  
row.names(cereals) <- cereals[,1]  
  
cereals <- cereals[,-1]  
  
str(cereals)

## 'data.frame': 77 obs. of 15 variables:  
## $ mfr : chr "N" "Q" "K" "K" ...  
## $ type : chr "C" "C" "C" "C" ...  
## $ calories: int 70 120 70 50 110 110 110 130 90 90 ...  
## $ protein : int 4 3 4 4 2 2 2 3 2 3 ...  
## $ fat : int 1 5 1 0 2 2 0 2 1 0 ...  
## $ sodium : int 130 15 260 140 200 180 125 210 200 210 ...  
## $ fiber : num 10 2 9 14 1 1.5 1 2 4 5 ...  
## $ carbo : num 5 8 7 8 14 10.5 11 18 15 13 ...  
## $ sugars : int 6 8 5 0 8 10 14 8 6 5 ...  
## $ potass : int 280 135 320 330 NA 70 30 100 125 190 ...  
## $ vitamins: int 25 0 25 25 25 25 25 25 25 25 ...  
## $ shelf : int 3 3 3 3 3 1 2 3 1 3 ...  
## $ weight : num 1 1 1 1 1 1 1 1.33 1 1 ...  
## $ cups : num 0.33 1 0.33 0.5 0.75 0.75 1 0.75 0.67 0.67 ...  
## $ rating : num 68.4 34 59.4 93.7 34.4 ...

cereals <- na.omit(cereals)  
  
cereals\_num <- cereals[, c(-1, -2, -12)]  
  
cereals\_num\_scale <- scale(cereals\_num)

Compute Euclidean distance

d <- dist(cereals\_num\_scale, method = "euclidean")  
  
  
hc\_com <- agnes(d, method = "complete")  
hc\_avg <- agnes(d, method = "average")  
hc\_sin <- agnes(d, method = "single")  
hc\_ward <- agnes(d, method = "ward")  
  
  
hc\_com$ac # 0.8469328

## [1] 0.8469328

hc\_avg$ac # 0.7881955

## [1] 0.7881955

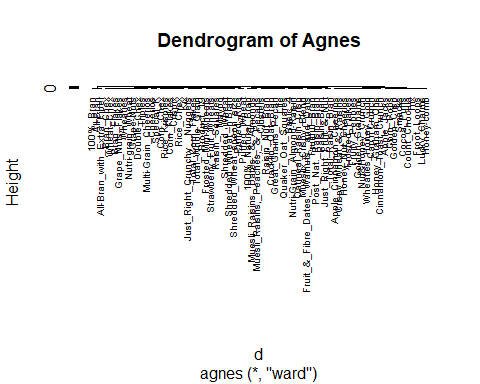
hc\_sin$ac # 0.6072384

## [1] 0.6072384

hc\_ward$ac # 0.9087265

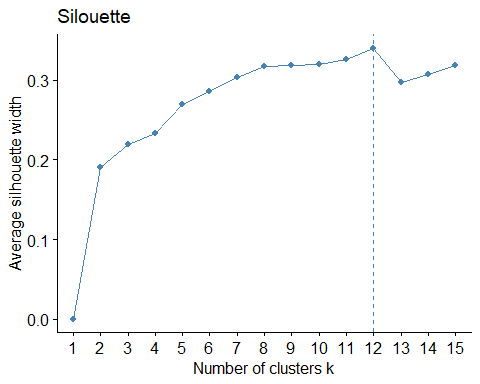
## [1] 0.9087265

#Based on the highest agglomerative coefficient, the best method to use for this problem is Ward's Method.  
  
#The biggest difference between hierarchical and k-means clustering is that k-means requires a pre-specified amount of clusters and hierarchical does not. Hierarchical clustering puts the clusters in a hierarchical form, which makes it much easier to interpret. Hierarchical clustering is preferred when you don't have a specific set of clusters needed to be defined.   
  
pltree(hc\_ward, cex = 0.6, hang = -1, main = "Dendrogram of Agnes")



Determining how many clusters to use

fviz\_nbclust(cereals\_num\_scale, FUN = hcut, method = "silhouette", k.max = 15) + ggtitle("Silouette")



#Based on the silhouette method and studying the dendrogram,the ideal amount of clusters is 12.

cluster\_groups <- cutree(hc\_ward, k = 12)  
cluster\_groups

## 100%\_Bran 100%\_Natural\_Bran   
## 1 2   
## All-Bran All-Bran\_with\_Extra\_Fiber   
## 1 1   
## Apple\_Cinnamon\_Cheerios Apple\_Jacks   
## 3 3   
## Basic\_4 Bran\_Chex   
## 4 5   
## Bran\_Flakes Cap'n'Crunch   
## 5 3   
## Cheerios Cinnamon\_Toast\_Crunch   
## 6 3   
## Clusters Cocoa\_Puffs   
## 7 3   
## Corn\_Chex Corn\_Flakes   
## 8 8   
## Corn\_Pops Count\_Chocula   
## 3 3   
## Cracklin'\_Oat\_Bran Crispix   
## 7 8   
## Crispy\_Wheat\_&\_Raisins Double\_Chex   
## 3 5   
## Froot\_Loops Frosted\_Flakes   
## 3 3   
## Frosted\_Mini-Wheats Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats   
## 9 4   
## Fruitful\_Bran Fruity\_Pebbles   
## 4 3   
## Golden\_Crisp Golden\_Grahams   
## 3 3   
## Grape\_Nuts\_Flakes Grape-Nuts   
## 5 5   
## Great\_Grains\_Pecan Honey\_Graham\_Ohs   
## 7 3   
## Honey\_Nut\_Cheerios Honey-comb   
## 3 3   
## Just\_Right\_Crunchy\_\_Nuggets Just\_Right\_Fruit\_&\_Nut   
## 10 11   
## Kix Life   
## 8 7   
## Lucky\_Charms Maypo   
## 3 9   
## Muesli\_Raisins,\_Dates,\_&\_Almonds Muesli\_Raisins,\_Peaches,\_&\_Pecans   
## 2 2   
## Mueslix\_Crispy\_Blend Multi-Grain\_Cheerios   
## 4 5   
## Nut&Honey\_Crunch Nutri-Grain\_Almond-Raisin   
## 3 4   
## Nutri-grain\_Wheat Oatmeal\_Raisin\_Crisp   
## 5 4   
## Post\_Nat.\_Raisin\_Bran Product\_19   
## 4 10   
## Puffed\_Rice Puffed\_Wheat   
## 12 12   
## Quaker\_Oat\_Squares Raisin\_Bran   
## 7 4   
## Raisin\_Nut\_Bran Raisin\_Squares   
## 7 9   
## Rice\_Chex Rice\_Krispies   
## 8 8   
## Shredded\_Wheat Shredded\_Wheat\_'n'Bran   
## 9 9   
## Shredded\_Wheat\_spoon\_size Smacks   
## 9 3   
## Special\_K Strawberry\_Fruit\_Wheats   
## 6 9   
## Total\_Corn\_Flakes Total\_Raisin\_Bran   
## 10 11   
## Total\_Whole\_Grain Triples   
## 10 5   
## Trix Wheat\_Chex   
## 3 5   
## Wheaties Wheaties\_Honey\_Gold   
## 5 3

tapply(cereals$calories, cluster\_groups, mean)

## 1 2 3 4 5 6 7 8   
## 63.33333 140.00000 110.95238 130.00000 99.00000 110.00000 106.66667 108.33333   
## 9 10 11 12   
## 91.42857 105.00000 140.00000 50.00000

Stability Testing

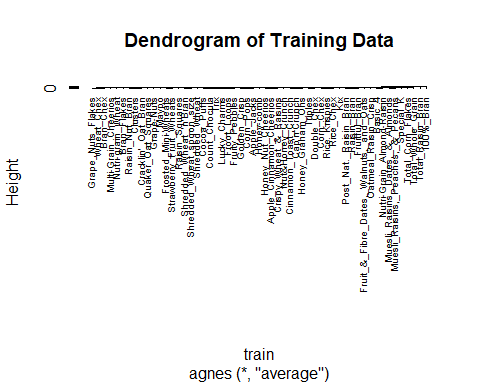
smp\_size <- floor(0.75 \* nrow(cereals\_num\_scale))  
  
  
set.seed(123)  
train\_ind <- sample(seq\_len(nrow(cereals\_num\_scale)), size = smp\_size)  
  
train <- cereals\_num\_scale[train\_ind, ]  
test <- cereals\_num\_scale[-train\_ind, ]  
  
train

## calories protein fat  
## Grape\_Nuts\_Flakes -0.3541153 0.4522084 0.0000000  
## Post\_Nat.\_Raisin\_Bran 0.6537514 0.4522084 0.0000000  
## Cocoa\_Puffs 0.1498180 -1.4068705 0.0000000  
## Total\_Corn\_Flakes 0.1498180 -0.4773310 0.0000000  
## Maypo -0.3541153 1.3817478 0.0000000  
## Oatmeal\_Raisin\_Crisp 1.1576848 0.4522084 0.9932203  
## Muesli\_Raisins,\_Dates,\_&\_Almonds 2.1655516 1.3817478 1.9864405  
## Wheat\_Chex -0.3541153 0.4522084 0.0000000  
## Frosted\_Mini-Wheats -0.3541153 0.4522084 -0.9932203  
## Raisin\_Nut\_Bran -0.3541153 0.4522084 0.9932203  
## Fruity\_Pebbles 0.1498180 -1.4068705 0.0000000  
## Special\_K 0.1498180 3.2408266 -0.9932203  
## Bran\_Flakes -0.8580487 0.4522084 -0.9932203  
## Golden\_Crisp -0.3541153 -0.4773310 -0.9932203  
## Honey\_Nut\_Cheerios 0.1498180 0.4522084 0.0000000  
## Bran\_Chex -0.8580487 -0.4773310 0.0000000  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats 0.6537514 0.4522084 0.9932203  
## Basic\_4 1.1576848 0.4522084 0.9932203  
## Triples 0.1498180 -0.4773310 0.0000000  
## Shredded\_Wheat\_'n'Bran -0.8580487 0.4522084 -0.9932203  
## Cracklin'\_Oat\_Bran 0.1498180 0.4522084 1.9864405  
## Honey-comb 0.1498180 -1.4068705 -0.9932203  
## Trix 0.1498180 -1.4068705 0.0000000  
## Corn\_Pops 0.1498180 -1.4068705 -0.9932203  
## Total\_Raisin\_Bran 1.6616182 0.4522084 0.0000000  
## Kix 0.1498180 -0.4773310 0.0000000  
## Cinnamon\_Toast\_Crunch 0.6537514 -1.4068705 1.9864405  
## Corn\_Chex 0.1498180 -0.4773310 -0.9932203  
## Grape-Nuts 0.1498180 0.4522084 -0.9932203  
## Raisin\_Bran 0.6537514 0.4522084 0.0000000  
## Shredded\_Wheat\_spoon\_size -0.8580487 0.4522084 -0.9932203  
## Quaker\_Oat\_Squares -0.3541153 1.3817478 0.0000000  
## Lucky\_Charms 0.1498180 -0.4773310 0.0000000  
## Cap'n'Crunch 0.6537514 -1.4068705 0.9932203  
## Froot\_Loops 0.1498180 -0.4773310 0.0000000  
## Fruitful\_Bran 0.6537514 0.4522084 -0.9932203  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans 2.1655516 1.3817478 1.9864405  
## Nutri-grain\_Wheat -0.8580487 0.4522084 -0.9932203  
## Multi-Grain\_Cheerios -0.3541153 -0.4773310 0.0000000  
## Strawberry\_Fruit\_Wheats -0.8580487 -0.4773310 -0.9932203  
## Honey\_Graham\_Ohs 0.6537514 -1.4068705 0.9932203  
## Shredded\_Wheat -1.3619821 -0.4773310 -0.9932203  
## Apple\_Cinnamon\_Cheerios 0.1498180 -0.4773310 0.9932203  
## Rice\_Chex 0.1498180 -1.4068705 -0.9932203  
## Nutri-Grain\_Almond-Raisin 1.6616182 0.4522084 0.9932203  
## Clusters 0.1498180 0.4522084 0.9932203  
## Count\_Chocula 0.1498180 -1.4068705 0.0000000  
## 100%\_Bran -1.8659155 1.3817478 0.0000000  
## Rice\_Krispies 0.1498180 -0.4773310 -0.9932203  
## Raisin\_Squares -0.8580487 -0.4773310 -0.9932203  
## Apple\_Jacks 0.1498180 -0.4773310 -0.9932203  
## Crispy\_Wheat\_&\_Raisins -0.3541153 -0.4773310 0.0000000  
## Nut&Honey\_Crunch 0.6537514 -0.4773310 0.0000000  
## Total\_Whole\_Grain -0.3541153 0.4522084 0.0000000  
## Double\_Chex -0.3541153 -0.4773310 -0.9932203  
## sodium fiber carbo  
## Grape\_Nuts\_Flakes -0.27020566 0.34015322 0.06944832  
## Post\_Nat.\_Raisin\_Bran 0.45469653 1.57808790 -0.95838683  
## Cocoa\_Puffs 0.21306247 -0.89778146 -0.70142805  
## Total\_Corn\_Flakes 0.45469653 -0.89778146 1.61120105  
## Maypo -1.96164410 -0.89778146 0.32640711  
## Oatmeal\_Raisin\_Crisp 0.09224544 -0.27881412 -0.31598986  
## Muesli\_Raisins,\_Dates,\_&\_Almonds -0.81388230 0.34015322 0.32640711  
## Wheat\_Chex 0.81714763 0.34015322 0.58336590  
## Frosted\_Mini-Wheats -1.96164410 0.34015322 -0.18751047  
## Raisin\_Nut\_Bran -0.27020566 0.13383078 -1.08686623  
## Fruity\_Pebbles -0.33061417 -0.89778146 -0.44446926  
## Special\_K 0.81714763 -0.48513656 0.32640711  
## Bran\_Flakes 0.57551356 1.16544301 -0.44446926  
## Golden\_Crisp -1.41796746 -0.89778146 -0.95838683  
## Honey\_Nut\_Cheerios 1.05878169 -0.27881412 -0.82990744  
## Bran\_Chex 0.45469653 0.75279812 0.06944832  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats -0.02857160 1.16544301 -0.70142805  
## Basic\_4 0.57551356 -0.07249167 0.84032469  
## Triples 1.05878169 -0.89778146 1.61120105  
## Shredded\_Wheat\_'n'Bran -1.96164410 0.75279812 1.09728348  
## Cracklin'\_Oat\_Bran -0.27020566 0.75279812 -1.21534562  
## Honey-comb 0.21306247 -0.89778146 -0.18751047  
## Trix -0.27020566 -0.89778146 -0.44446926  
## Corn\_Pops -0.87429082 -0.48513656 -0.44446926  
## Total\_Raisin\_Bran 0.33387950 0.75279812 0.06944832  
## Kix 1.17959872 -0.89778146 1.61120105  
## Cinnamon\_Toast\_Crunch 0.57551356 -0.89778146 -0.44446926  
## Corn\_Chex 1.42123279 -0.89778146 1.86815984  
## Grape-Nuts 0.09224544 0.34015322 0.58336590  
## Raisin\_Bran 0.57551356 1.16544301 -0.18751047  
## Shredded\_Wheat\_spoon\_size -1.96164410 0.34015322 1.35424227  
## Quaker\_Oat\_Squares -0.33061417 -0.07249167 -0.18751047  
## Lucky\_Charms 0.21306247 -0.89778146 -0.70142805  
## Cap'n'Crunch 0.69633060 -0.89778146 -0.70142805  
## Froot\_Loops -0.45143121 -0.48513656 -0.95838683  
## Fruitful\_Bran 0.93796466 1.16544301 -0.18751047  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans -0.14938863 0.34015322 0.32640711  
## Nutri-grain\_Wheat 0.09224544 0.34015322 0.84032469  
## Multi-Grain\_Cheerios 0.69633060 -0.07249167 0.06944832  
## Strawberry\_Fruit\_Wheats -1.78041856 0.34015322 0.06944832  
## Honey\_Graham\_Ohs 0.69633060 -0.48513656 -0.70142805  
## Shredded\_Wheat -1.96164410 0.34015322 0.32640711  
## Apple\_Cinnamon\_Cheerios 0.21306247 -0.27881412 -1.08686623  
## Rice\_Chex 0.93796466 -0.89778146 2.12511863  
## Nutri-Grain\_Almond-Raisin 0.69633060 0.34015322 1.61120105  
## Clusters -0.27020566 -0.07249167 -0.44446926  
## Count\_Chocula 0.21306247 -0.89778146 -0.70142805  
## 100%\_Bran -0.39102269 3.22866747 -2.50013957  
## Rice\_Krispies 1.54204982 -0.89778146 1.86815984  
## Raisin\_Squares -1.96164410 -0.07249167 0.06944832  
## Apple\_Jacks -0.45143121 -0.48513656 -0.95838683  
## Crispy\_Wheat\_&\_Raisins -0.27020566 -0.07249167 -0.95838683  
## Nut&Honey\_Crunch 0.33387950 -0.89778146 0.06944832  
## Total\_Whole\_Grain 0.45469653 0.34015322 0.32640711  
## Double\_Chex 0.33387950 -0.48513656 0.84032469  
## sugars potass vitamins  
## Grape\_Nuts\_Flakes -0.48360961 -0.19065695 -0.1818422  
## Post\_Nat.\_Raisin\_Bran 1.58103142 2.27835060 -0.1818422  
## Cocoa\_Puffs 1.35162686 -0.61391539 -0.1818422  
## Total\_Corn\_Flakes -0.94241873 -0.89608768 3.1822385  
## Maypo -0.94241873 -0.04957081 -0.1818422  
## Oatmeal\_Raisin\_Crisp 0.66341318 0.30314456 -0.1818422  
## Muesli\_Raisins,\_Dates,\_&\_Almonds 0.89281774 1.00857529 -0.1818422  
## Wheat\_Chex -0.94241873 0.23260148 -0.1818422  
## Frosted\_Mini-Wheats -0.02480049 0.02097226 -0.1818422  
## Raisin\_Nut\_Bran 0.20460407 0.58531685 -0.1818422  
## Fruity\_Pebbles 1.12222230 -1.03717383 -0.1818422  
## Special\_K -0.94241873 -0.61391539 -0.1818422  
## Bran\_Flakes -0.48360961 1.29074758 -0.1818422  
## Golden\_Crisp 1.81043598 -0.82554461 -0.1818422  
## Honey\_Nut\_Cheerios 0.66341318 -0.12011388 -0.1818422  
## Bran\_Chex -0.25420505 0.37368763 -0.1818422  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats 0.66341318 1.43183372 -0.1818422  
## Basic\_4 0.20460407 0.02097226 -0.1818422  
## Triples -0.94241873 -0.54337232 -0.1818422  
## Shredded\_Wheat\_'n'Bran -1.63063240 0.58531685 -1.3032024  
## Cracklin'\_Oat\_Bran -0.02480049 0.86748914 -0.1818422  
## Honey-comb 0.89281774 -0.89608768 -0.1818422  
## Trix 1.12222230 -1.03717383 -0.1818422  
## Corn\_Pops 1.12222230 -1.10771690 -0.1818422  
## Total\_Raisin\_Bran 1.58103142 1.85509216 3.1822385  
## Kix -0.94241873 -0.82554461 -0.1818422  
## Cinnamon\_Toast\_Crunch 0.43400862 -0.75500154 -0.1818422  
## Corn\_Chex -0.94241873 -1.03717383 -0.1818422  
## Grape-Nuts -0.94241873 -0.12011388 -0.1818422  
## Raisin\_Bran 1.12222230 1.99617831 -0.1818422  
## Shredded\_Wheat\_spoon\_size -1.63063240 0.30314456 -1.3032024  
## Quaker\_Oat\_Squares -0.25420505 0.16205841 -0.1818422  
## Lucky\_Charms 1.12222230 -0.61391539 -0.1818422  
## Cap'n'Crunch 1.12222230 -0.89608768 -0.1818422  
## Froot\_Loops 1.35162686 -0.96663076 -0.1818422  
## Fruitful\_Bran 1.12222230 1.29074758 -0.1818422  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans 0.89281774 1.00857529 -0.1818422  
## Nutri-grain\_Wheat -1.17182329 -0.12011388 -0.1818422  
## Multi-Grain\_Cheerios -0.25420505 -0.12011388 -0.1818422  
## Strawberry\_Fruit\_Wheats -0.48360961 -0.12011388 -0.1818422  
## Honey\_Graham\_Ohs 0.89281774 -0.75500154 -0.1818422  
## Shredded\_Wheat -1.63063240 -0.04957081 -1.3032024  
## Apple\_Cinnamon\_Cheerios 0.66341318 -0.40228617 -0.1818422  
## Rice\_Chex -1.17182329 -0.96663076 -0.1818422  
## Nutri-Grain\_Almond-Raisin -0.02480049 0.44423070 -0.1818422  
## Clusters -0.02480049 0.09151534 -0.1818422  
## Count\_Chocula 1.35162686 -0.47282925 -0.1818422  
## 100%\_Bran -0.25420505 2.56052289 -0.1818422  
## Rice\_Krispies -0.94241873 -0.89608768 -0.1818422  
## Raisin\_Squares -0.25420505 0.16205841 -0.1818422  
## Apple\_Jacks 1.58103142 -0.96663076 -0.1818422  
## Crispy\_Wheat\_&\_Raisins 0.66341318 0.30314456 -0.1818422  
## Nut&Honey\_Crunch 0.43400862 -0.82554461 -0.1818422  
## Total\_Whole\_Grain -0.94241873 0.16205841 3.1822385  
## Double\_Chex -0.48360961 -0.26120003 -0.1818422  
## weight cups rating  
## Grape\_Nuts\_Flakes -0.2008324 0.24766475 0.69155685  
## Post\_Nat.\_Raisin\_Bran 1.9501886 -0.64324039 -0.32287913  
## Cocoa\_Puffs -0.2008324 0.75675340 -1.39915514  
## Total\_Corn\_Flakes -0.2008324 0.75675340 -0.25168258  
## Maypo -0.2008324 0.75675340 0.88922515  
## Oatmeal\_Raisin\_Crisp 1.4287290 -1.36444931 -0.84945049  
## Muesli\_Raisins,\_Dates,\_&\_Almonds -0.2008324 0.75675340 -0.37302488  
## Wheat\_Chex -0.2008324 -0.64324039 0.52841741  
## Frosted\_Mini-Wheats -0.2008324 -0.09172768 1.13821301  
## Raisin\_Nut\_Bran -0.2008324 -1.36444931 -0.19014120  
## Fruity\_Pebbles -0.2008324 -0.30384795 -1.02225423  
## Special\_K -0.2008324 0.75675340 0.76669214  
## Bran\_Flakes -0.2008324 -0.64324039 0.77969576  
## Golden\_Crisp -0.2008324 0.24766475 -0.50730289  
## Honey\_Nut\_Cheerios -0.2008324 -0.30384795 -0.80517325  
## Bran\_Chex -0.2008324 -0.64324039 0.48087533  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats 1.4287290 -0.64324039 -0.10366038  
## Basic\_4 1.9501886 -0.30384795 -0.38002951  
## Triples -0.2008324 -0.30384795 -0.23269772  
## Shredded\_Wheat\_'n'Bran -0.2008324 -0.64324039 2.28743193  
## Cracklin'\_Oat\_Bran -0.2008324 -1.36444931 -0.13702824  
## Honey-comb -0.2008324 2.15674718 -0.97118798  
## Trix -0.2008324 0.75675340 -1.04166919  
## Corn\_Pops -0.2008324 0.75675340 -0.46951197  
## Total\_Raisin\_Bran 3.0582904 0.75675340 -0.98185009  
## Kix -0.2008324 2.87795610 -0.22308231  
## Cinnamon\_Toast\_Crunch -0.2008324 -0.30384795 -1.60671768  
## Corn\_Chex -0.2008324 0.75675340 -0.06603869  
## Grape-Nuts -0.2008324 -2.42505066 0.78377123  
## Raisin\_Bran 1.9501886 -0.30384795 -0.22179377  
## Shredded\_Wheat\_spoon\_size -0.2008324 -0.64324039 2.16834997  
## Quaker\_Oat\_Squares -0.2008324 -1.36444931 0.50878106  
## Lucky\_Charms -0.2008324 0.75675340 -1.11426481  
## Cap'n'Crunch -0.2008324 -0.30384795 -1.73360655  
## Froot\_Loops -0.2008324 0.75675340 -0.72427057  
## Fruitful\_Bran 1.9501886 -0.64324039 -0.09664548  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans -0.2008324 0.75675340 -0.58658904  
## Nutri-grain\_Wheat -0.2008324 0.75675340 1.23068291  
## Multi-Grain\_Cheerios -0.2008324 0.75675340 -0.16145563  
## Strawberry\_Fruit\_Wheats -0.2008324 0.75675340 1.21081332  
## Honey\_Graham\_Ohs -0.2008324 0.75675340 -1.46080340  
## Shredded\_Wheat -1.3089342 0.75675340 1.84299757  
## Apple\_Cinnamon\_Cheerios -0.2008324 -0.30384795 -0.91652483  
## Rice\_Chex -0.2008324 1.30826610 -0.02656845  
## Nutri-Grain\_Almond-Raisin 1.9501886 -0.64324039 -0.11967375  
## Clusters -0.2008324 -1.36444931 -0.14048876  
## Count\_Chocula -0.2008324 0.75675340 -1.42337774  
## 100%\_Bran -0.2008324 -2.08565823 1.85490376  
## Rice\_Krispies -0.2008324 0.75675340 -0.12909114  
## Raisin\_Squares -0.2008324 -1.36444931 0.92358705  
## Apple\_Jacks -0.2008324 0.75675340 -0.65539984  
## Crispy\_Wheat\_&\_Raisins -0.2008324 -0.30384795 -0.44147911  
## Nut&Honey\_Crunch -0.2008324 -0.64324039 -0.88697142  
## Total\_Whole\_Grain -0.2008324 0.75675340 0.30548275  
## Double\_Chex -0.2008324 -0.30384795 0.13959735

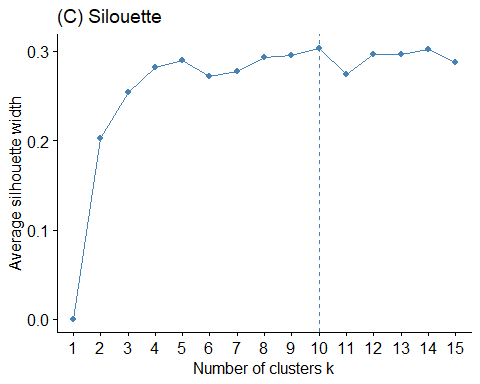
hc2\_avg <- agnes(train, method = "average")  
  
hc2\_avg$ac # 0.7492276

## [1] 0.7492276

pltree(hc2\_avg, cex = 0.6, hang = -1, main = "Dendrogram of Training Data")



fviz\_nbclust(train, FUN = hcut, method = "silhouette", k.max = 15) + ggtitle("(C) Silouette")



cluster\_groups2 <- cutree(hc2\_avg, k = 10)  
  
cluster\_groups2

## Grape\_Nuts\_Flakes Post\_Nat.\_Raisin\_Bran   
## 1 2   
## Cocoa\_Puffs Total\_Corn\_Flakes   
## 3 4   
## Maypo Oatmeal\_Raisin\_Crisp   
## 5 2   
## Muesli\_Raisins,\_Dates,\_&\_Almonds Wheat\_Chex   
## 6 1   
## Frosted\_Mini-Wheats Raisin\_Nut\_Bran   
## 5 1   
## Fruity\_Pebbles Special\_K   
## 3 7   
## Bran\_Flakes Golden\_Crisp   
## 1 3   
## Honey\_Nut\_Cheerios Bran\_Chex   
## 3 1   
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats Basic\_4   
## 2 2   
## Triples Shredded\_Wheat\_'n'Bran   
## 8 5   
## Cracklin'\_Oat\_Bran Honey-comb   
## 1 3   
## Trix Corn\_Pops   
## 3 3   
## Total\_Raisin\_Bran Kix   
## 9 8   
## Cinnamon\_Toast\_Crunch Corn\_Chex   
## 3 8   
## Grape-Nuts Raisin\_Bran   
## 1 2   
## Shredded\_Wheat\_spoon\_size Quaker\_Oat\_Squares   
## 5 1   
## Lucky\_Charms Cap'n'Crunch   
## 3 3   
## Froot\_Loops Fruitful\_Bran   
## 3 2   
## Muesli\_Raisins,\_Peaches,\_&\_Pecans Nutri-grain\_Wheat   
## 6 1   
## Multi-Grain\_Cheerios Strawberry\_Fruit\_Wheats   
## 1 5   
## Honey\_Graham\_Ohs Shredded\_Wheat   
## 3 5   
## Apple\_Cinnamon\_Cheerios Rice\_Chex   
## 3 8   
## Nutri-Grain\_Almond-Raisin Clusters   
## 2 1   
## Count\_Chocula 100%\_Bran   
## 3 10   
## Rice\_Krispies Raisin\_Squares   
## 8 5   
## Apple\_Jacks Crispy\_Wheat\_&\_Raisins   
## 3 3   
## Nut&Honey\_Crunch Total\_Whole\_Grain   
## 3 4   
## Double\_Chex   
## 8

#Comparing the assignment of clusters from the original set of data to the training set, which had 75% of the original cereals, I was able to find a lot of similarities between the data sets. For example, cereals such as Cap "n" Crunch and Apple Jacks, both were in the same cluster for both sets of data. The clusters between each set may be different numbers, but they still have the same cereals with common similarities. These clusters are very stable based on the similarities between the data sets.

Elementary Healthy Cereal Choices

# The data should not be normalized at first because the school would have constraints on what a cereal should have for each attribute, calories, fats, carbs, etc, to be considered healthy. If we normalize too early, we won't be able to tell if each cereal meets the minimum requirements to be considered "healthy" by the school's standards. Once all the cereals that don't meet the schools standards are filtered, then the data set can be normalized. Once the data is normalized, you can perform a typical cluster analysis.

Comparing Hierarchical and K-means Clustering

#Hierarchical and k-means clustering both group data points into clusters with similar attributes. Hierarchical clustering has the advantage of not needing to pre-specify the amount of clusters needed for the clustering algorithm to function, where as k-means needs a pre-determined amount of clusters. This gives the user the ability to interpret the dendrogram in many different ways. It is important to make sure that you have a great understanding of the data before hand to make better interpretations. The best way to compare the algorithms is to compare the clusters of both and see which data points are grouped with which.