Assignment\_5

2022-11-20

##Inputting and Setting up Data/Packages

mydata <- read.csv("C:\\Users\\bucol\\Downloads\\Cereals.csv")  
library(tidyverse) # data manipulation

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(factoextra) # clustering algorithms & visualization

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

library(cluster)

## Warning: package 'cluster' was built under R version 4.2.2

library(caret)

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

set.seed(49)  
  
  
summary(mydata)

## 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   
##

#We see there is some missing data that needs to be removed

cerealscomplete <- na.omit(mydata)  
cerealscomplete #this has removed the cereals with incomplete data

## name mfr type calories protein fat sodium  
## 1 100%\_Bran N C 70 4 1 130  
## 2 100%\_Natural\_Bran Q C 120 3 5 15  
## 3 All-Bran K C 70 4 1 260  
## 4 All-Bran\_with\_Extra\_Fiber K C 50 4 0 140  
## 6 Apple\_Cinnamon\_Cheerios G C 110 2 2 180  
## 7 Apple\_Jacks K C 110 2 0 125  
## 8 Basic\_4 G C 130 3 2 210  
## 9 Bran\_Chex R C 90 2 1 200  
## 10 Bran\_Flakes P C 90 3 0 210  
## 11 Cap'n'Crunch Q C 120 1 2 220  
## 12 Cheerios G C 110 6 2 290  
## 13 Cinnamon\_Toast\_Crunch G C 120 1 3 210  
## 14 Clusters G C 110 3 2 140  
## 15 Cocoa\_Puffs G C 110 1 1 180  
## 16 Corn\_Chex R C 110 2 0 280  
## 17 Corn\_Flakes K C 100 2 0 290  
## 18 Corn\_Pops K C 110 1 0 90  
## 19 Count\_Chocula G C 110 1 1 180  
## 20 Cracklin'\_Oat\_Bran K C 110 3 3 140  
## 22 Crispix K C 110 2 0 220  
## 23 Crispy\_Wheat\_&\_Raisins G C 100 2 1 140  
## 24 Double\_Chex R C 100 2 0 190  
## 25 Froot\_Loops K C 110 2 1 125  
## 26 Frosted\_Flakes K C 110 1 0 200  
## 27 Frosted\_Mini-Wheats K C 100 3 0 0  
## 28 Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats P C 120 3 2 160  
## 29 Fruitful\_Bran K C 120 3 0 240  
## 30 Fruity\_Pebbles P C 110 1 1 135  
## 31 Golden\_Crisp P C 100 2 0 45  
## 32 Golden\_Grahams G C 110 1 1 280  
## 33 Grape\_Nuts\_Flakes P C 100 3 1 140  
## 34 Grape-Nuts P C 110 3 0 170  
## 35 Great\_Grains\_Pecan P C 120 3 3 75  
## 36 Honey\_Graham\_Ohs Q C 120 1 2 220  
## 37 Honey\_Nut\_Cheerios G C 110 3 1 250  
## 38 Honey-comb P C 110 1 0 180  
## 39 Just\_Right\_Crunchy\_\_Nuggets K C 110 2 1 170  
## 40 Just\_Right\_Fruit\_&\_Nut K C 140 3 1 170  
## 41 Kix G C 110 2 1 260  
## 42 Life Q C 100 4 2 150  
## 43 Lucky\_Charms G C 110 2 1 180  
## 44 Maypo A H 100 4 1 0  
## 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  
## 48 Multi-Grain\_Cheerios G C 100 2 1 220  
## 49 Nut&Honey\_Crunch K C 120 2 1 190  
## 50 Nutri-Grain\_Almond-Raisin K C 140 3 2 220  
## 51 Nutri-grain\_Wheat K C 90 3 0 170  
## 52 Oatmeal\_Raisin\_Crisp G C 130 3 2 170  
## 53 Post\_Nat.\_Raisin\_Bran P C 120 3 1 200  
## 54 Product\_19 K C 100 3 0 320  
## 55 Puffed\_Rice Q C 50 1 0 0  
## 56 Puffed\_Wheat Q C 50 2 0 0  
## 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  
## 61 Raisin\_Squares K C 90 2 0 0  
## 62 Rice\_Chex R C 110 1 0 240  
## 63 Rice\_Krispies K C 110 2 0 290  
## 64 Shredded\_Wheat N C 80 2 0 0  
## 65 Shredded\_Wheat\_'n'Bran N C 90 3 0 0  
## 66 Shredded\_Wheat\_spoon\_size N C 90 3 0 0  
## 67 Smacks K C 110 2 1 70  
## 68 Special\_K K C 110 6 0 230  
## 69 Strawberry\_Fruit\_Wheats N C 90 2 0 15  
## 70 Total\_Corn\_Flakes G C 110 2 1 200  
## 71 Total\_Raisin\_Bran G C 140 3 1 190  
## 72 Total\_Whole\_Grain G C 100 3 1 200  
## 73 Triples G C 110 2 1 250  
## 74 Trix G C 110 1 1 140  
## 75 Wheat\_Chex R C 100 3 1 230  
## 76 Wheaties G C 100 3 1 200  
## 77 Wheaties\_Honey\_Gold G C 110 2 1 200  
## fiber carbo sugars potass vitamins shelf weight cups rating  
## 1 10.0 5.0 6 280 25 3 1.00 0.33 68.40297  
## 2 2.0 8.0 8 135 0 3 1.00 1.00 33.98368  
## 3 9.0 7.0 5 320 25 3 1.00 0.33 59.42551  
## 4 14.0 8.0 0 330 25 3 1.00 0.50 93.70491  
## 6 1.5 10.5 10 70 25 1 1.00 0.75 29.50954  
## 7 1.0 11.0 14 30 25 2 1.00 1.00 33.17409  
## 8 2.0 18.0 8 100 25 3 1.33 0.75 37.03856  
## 9 4.0 15.0 6 125 25 1 1.00 0.67 49.12025  
## 10 5.0 13.0 5 190 25 3 1.00 0.67 53.31381  
## 11 0.0 12.0 12 35 25 2 1.00 0.75 18.04285  
## 12 2.0 17.0 1 105 25 1 1.00 1.25 50.76500  
## 13 0.0 13.0 9 45 25 2 1.00 0.75 19.82357  
## 14 2.0 13.0 7 105 25 3 1.00 0.50 40.40021  
## 15 0.0 12.0 13 55 25 2 1.00 1.00 22.73645  
## 16 0.0 22.0 3 25 25 1 1.00 1.00 41.44502  
## 17 1.0 21.0 2 35 25 1 1.00 1.00 45.86332  
## 18 1.0 13.0 12 20 25 2 1.00 1.00 35.78279  
## 19 0.0 12.0 13 65 25 2 1.00 1.00 22.39651  
## 20 4.0 10.0 7 160 25 3 1.00 0.50 40.44877  
## 22 1.0 21.0 3 30 25 3 1.00 1.00 46.89564  
## 23 2.0 11.0 10 120 25 3 1.00 0.75 36.17620  
## 24 1.0 18.0 5 80 25 3 1.00 0.75 44.33086  
## 25 1.0 11.0 13 30 25 2 1.00 1.00 32.20758  
## 26 1.0 14.0 11 25 25 1 1.00 0.75 31.43597  
## 27 3.0 14.0 7 100 25 2 1.00 0.80 58.34514  
## 28 5.0 12.0 10 200 25 3 1.25 0.67 40.91705  
## 29 5.0 14.0 12 190 25 3 1.33 0.67 41.01549  
## 30 0.0 13.0 12 25 25 2 1.00 0.75 28.02576  
## 31 0.0 11.0 15 40 25 1 1.00 0.88 35.25244  
## 32 0.0 15.0 9 45 25 2 1.00 0.75 23.80404  
## 33 3.0 15.0 5 85 25 3 1.00 0.88 52.07690  
## 34 3.0 17.0 3 90 25 3 1.00 0.25 53.37101  
## 35 3.0 13.0 4 100 25 3 1.00 0.33 45.81172  
## 36 1.0 12.0 11 45 25 2 1.00 1.00 21.87129  
## 37 1.5 11.5 10 90 25 1 1.00 0.75 31.07222  
## 38 0.0 14.0 11 35 25 1 1.00 1.33 28.74241  
## 39 1.0 17.0 6 60 100 3 1.00 1.00 36.52368  
## 40 2.0 20.0 9 95 100 3 1.30 0.75 36.47151  
## 41 0.0 21.0 3 40 25 2 1.00 1.50 39.24111  
## 42 2.0 12.0 6 95 25 2 1.00 0.67 45.32807  
## 43 0.0 12.0 12 55 25 2 1.00 1.00 26.73451  
## 44 0.0 16.0 3 95 25 2 1.00 1.00 54.85092  
## 45 3.0 16.0 11 170 25 3 1.00 1.00 37.13686  
## 46 3.0 16.0 11 170 25 3 1.00 1.00 34.13976  
## 47 3.0 17.0 13 160 25 3 1.50 0.67 30.31335  
## 48 2.0 15.0 6 90 25 1 1.00 1.00 40.10596  
## 49 0.0 15.0 9 40 25 2 1.00 0.67 29.92429  
## 50 3.0 21.0 7 130 25 3 1.33 0.67 40.69232  
## 51 3.0 18.0 2 90 25 3 1.00 1.00 59.64284  
## 52 1.5 13.5 10 120 25 3 1.25 0.50 30.45084  
## 53 6.0 11.0 14 260 25 3 1.33 0.67 37.84059  
## 54 1.0 20.0 3 45 100 3 1.00 1.00 41.50354  
## 55 0.0 13.0 0 15 0 3 0.50 1.00 60.75611  
## 56 1.0 10.0 0 50 0 3 0.50 1.00 63.00565  
## 57 2.0 14.0 6 110 25 3 1.00 0.50 49.51187  
## 59 5.0 14.0 12 240 25 2 1.33 0.75 39.25920  
## 60 2.5 10.5 8 140 25 3 1.00 0.50 39.70340  
## 61 2.0 15.0 6 110 25 3 1.00 0.50 55.33314  
## 62 0.0 23.0 2 30 25 1 1.00 1.13 41.99893  
## 63 0.0 22.0 3 35 25 1 1.00 1.00 40.56016  
## 64 3.0 16.0 0 95 0 1 0.83 1.00 68.23588  
## 65 4.0 19.0 0 140 0 1 1.00 0.67 74.47295  
## 66 3.0 20.0 0 120 0 1 1.00 0.67 72.80179  
## 67 1.0 9.0 15 40 25 2 1.00 0.75 31.23005  
## 68 1.0 16.0 3 55 25 1 1.00 1.00 53.13132  
## 69 3.0 15.0 5 90 25 2 1.00 1.00 59.36399  
## 70 0.0 21.0 3 35 100 3 1.00 1.00 38.83975  
## 71 4.0 15.0 14 230 100 3 1.50 1.00 28.59278  
## 72 3.0 16.0 3 110 100 3 1.00 1.00 46.65884  
## 73 0.0 21.0 3 60 25 3 1.00 0.75 39.10617  
## 74 0.0 13.0 12 25 25 2 1.00 1.00 27.75330  
## 75 3.0 17.0 3 115 25 1 1.00 0.67 49.78744  
## 76 3.0 17.0 3 110 25 1 1.00 1.00 51.59219  
## 77 1.0 16.0 8 60 25 1 1.00 0.75 36.18756

#Now we need to make the cereal names the row names  
row.names(cerealscomplete) <- cerealscomplete[,1]  
#now we must remove the names from the dataset  
cerealscomplete <- cerealscomplete[,-1]  
  
#I'm going to remove the first two columns because they're categorical.   
cerealscomplete <- cerealscomplete[,-1]  
cerealscomplete <- cerealscomplete[,-1]  
cerealscomplete

## calories protein fat sodium fiber carbo  
## 100%\_Bran 70 4 1 130 10.0 5.0  
## 100%\_Natural\_Bran 120 3 5 15 2.0 8.0  
## All-Bran 70 4 1 260 9.0 7.0  
## All-Bran\_with\_Extra\_Fiber 50 4 0 140 14.0 8.0  
## Apple\_Cinnamon\_Cheerios 110 2 2 180 1.5 10.5  
## Apple\_Jacks 110 2 0 125 1.0 11.0  
## Basic\_4 130 3 2 210 2.0 18.0  
## Bran\_Chex 90 2 1 200 4.0 15.0  
## Bran\_Flakes 90 3 0 210 5.0 13.0  
## Cap'n'Crunch 120 1 2 220 0.0 12.0  
## Cheerios 110 6 2 290 2.0 17.0  
## Cinnamon\_Toast\_Crunch 120 1 3 210 0.0 13.0  
## Clusters 110 3 2 140 2.0 13.0  
## Cocoa\_Puffs 110 1 1 180 0.0 12.0  
## Corn\_Chex 110 2 0 280 0.0 22.0  
## Corn\_Flakes 100 2 0 290 1.0 21.0  
## Corn\_Pops 110 1 0 90 1.0 13.0  
## Count\_Chocula 110 1 1 180 0.0 12.0  
## Cracklin'\_Oat\_Bran 110 3 3 140 4.0 10.0  
## Crispix 110 2 0 220 1.0 21.0  
## Crispy\_Wheat\_&\_Raisins 100 2 1 140 2.0 11.0  
## Double\_Chex 100 2 0 190 1.0 18.0  
## Froot\_Loops 110 2 1 125 1.0 11.0  
## Frosted\_Flakes 110 1 0 200 1.0 14.0  
## Frosted\_Mini-Wheats 100 3 0 0 3.0 14.0  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats 120 3 2 160 5.0 12.0  
## Fruitful\_Bran 120 3 0 240 5.0 14.0  
## Fruity\_Pebbles 110 1 1 135 0.0 13.0  
## Golden\_Crisp 100 2 0 45 0.0 11.0  
## Golden\_Grahams 110 1 1 280 0.0 15.0  
## Grape\_Nuts\_Flakes 100 3 1 140 3.0 15.0  
## Grape-Nuts 110 3 0 170 3.0 17.0  
## Great\_Grains\_Pecan 120 3 3 75 3.0 13.0  
## Honey\_Graham\_Ohs 120 1 2 220 1.0 12.0  
## Honey\_Nut\_Cheerios 110 3 1 250 1.5 11.5  
## Honey-comb 110 1 0 180 0.0 14.0  
## Just\_Right\_Crunchy\_\_Nuggets 110 2 1 170 1.0 17.0  
## Just\_Right\_Fruit\_&\_Nut 140 3 1 170 2.0 20.0  
## Kix 110 2 1 260 0.0 21.0  
## Life 100 4 2 150 2.0 12.0  
## Lucky\_Charms 110 2 1 180 0.0 12.0  
## Maypo 100 4 1 0 0.0 16.0  
## Muesli\_Raisins,\_Dates,\_&\_Almonds 150 4 3 95 3.0 16.0  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans 150 4 3 150 3.0 16.0  
## Mueslix\_Crispy\_Blend 160 3 2 150 3.0 17.0  
## Multi-Grain\_Cheerios 100 2 1 220 2.0 15.0  
## Nut&Honey\_Crunch 120 2 1 190 0.0 15.0  
## Nutri-Grain\_Almond-Raisin 140 3 2 220 3.0 21.0  
## Nutri-grain\_Wheat 90 3 0 170 3.0 18.0  
## Oatmeal\_Raisin\_Crisp 130 3 2 170 1.5 13.5  
## Post\_Nat.\_Raisin\_Bran 120 3 1 200 6.0 11.0  
## Product\_19 100 3 0 320 1.0 20.0  
## Puffed\_Rice 50 1 0 0 0.0 13.0  
## Puffed\_Wheat 50 2 0 0 1.0 10.0  
## Quaker\_Oat\_Squares 100 4 1 135 2.0 14.0  
## Raisin\_Bran 120 3 1 210 5.0 14.0  
## Raisin\_Nut\_Bran 100 3 2 140 2.5 10.5  
## Raisin\_Squares 90 2 0 0 2.0 15.0  
## Rice\_Chex 110 1 0 240 0.0 23.0  
## Rice\_Krispies 110 2 0 290 0.0 22.0  
## Shredded\_Wheat 80 2 0 0 3.0 16.0  
## Shredded\_Wheat\_'n'Bran 90 3 0 0 4.0 19.0  
## Shredded\_Wheat\_spoon\_size 90 3 0 0 3.0 20.0  
## Smacks 110 2 1 70 1.0 9.0  
## Special\_K 110 6 0 230 1.0 16.0  
## Strawberry\_Fruit\_Wheats 90 2 0 15 3.0 15.0  
## Total\_Corn\_Flakes 110 2 1 200 0.0 21.0  
## Total\_Raisin\_Bran 140 3 1 190 4.0 15.0  
## Total\_Whole\_Grain 100 3 1 200 3.0 16.0  
## Triples 110 2 1 250 0.0 21.0  
## Trix 110 1 1 140 0.0 13.0  
## Wheat\_Chex 100 3 1 230 3.0 17.0  
## Wheaties 100 3 1 200 3.0 17.0  
## Wheaties\_Honey\_Gold 110 2 1 200 1.0 16.0  
## sugars potass vitamins shelf weight cups  
## 100%\_Bran 6 280 25 3 1.00 0.33  
## 100%\_Natural\_Bran 8 135 0 3 1.00 1.00  
## All-Bran 5 320 25 3 1.00 0.33  
## All-Bran\_with\_Extra\_Fiber 0 330 25 3 1.00 0.50  
## Apple\_Cinnamon\_Cheerios 10 70 25 1 1.00 0.75  
## Apple\_Jacks 14 30 25 2 1.00 1.00  
## Basic\_4 8 100 25 3 1.33 0.75  
## Bran\_Chex 6 125 25 1 1.00 0.67  
## Bran\_Flakes 5 190 25 3 1.00 0.67  
## Cap'n'Crunch 12 35 25 2 1.00 0.75  
## Cheerios 1 105 25 1 1.00 1.25  
## Cinnamon\_Toast\_Crunch 9 45 25 2 1.00 0.75  
## Clusters 7 105 25 3 1.00 0.50  
## Cocoa\_Puffs 13 55 25 2 1.00 1.00  
## Corn\_Chex 3 25 25 1 1.00 1.00  
## Corn\_Flakes 2 35 25 1 1.00 1.00  
## Corn\_Pops 12 20 25 2 1.00 1.00  
## Count\_Chocula 13 65 25 2 1.00 1.00  
## Cracklin'\_Oat\_Bran 7 160 25 3 1.00 0.50  
## Crispix 3 30 25 3 1.00 1.00  
## Crispy\_Wheat\_&\_Raisins 10 120 25 3 1.00 0.75  
## Double\_Chex 5 80 25 3 1.00 0.75  
## Froot\_Loops 13 30 25 2 1.00 1.00  
## Frosted\_Flakes 11 25 25 1 1.00 0.75  
## Frosted\_Mini-Wheats 7 100 25 2 1.00 0.80  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats 10 200 25 3 1.25 0.67  
## Fruitful\_Bran 12 190 25 3 1.33 0.67  
## Fruity\_Pebbles 12 25 25 2 1.00 0.75  
## Golden\_Crisp 15 40 25 1 1.00 0.88  
## Golden\_Grahams 9 45 25 2 1.00 0.75  
## Grape\_Nuts\_Flakes 5 85 25 3 1.00 0.88  
## Grape-Nuts 3 90 25 3 1.00 0.25  
## Great\_Grains\_Pecan 4 100 25 3 1.00 0.33  
## Honey\_Graham\_Ohs 11 45 25 2 1.00 1.00  
## Honey\_Nut\_Cheerios 10 90 25 1 1.00 0.75  
## Honey-comb 11 35 25 1 1.00 1.33  
## Just\_Right\_Crunchy\_\_Nuggets 6 60 100 3 1.00 1.00  
## Just\_Right\_Fruit\_&\_Nut 9 95 100 3 1.30 0.75  
## Kix 3 40 25 2 1.00 1.50  
## Life 6 95 25 2 1.00 0.67  
## Lucky\_Charms 12 55 25 2 1.00 1.00  
## Maypo 3 95 25 2 1.00 1.00  
## Muesli\_Raisins,\_Dates,\_&\_Almonds 11 170 25 3 1.00 1.00  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans 11 170 25 3 1.00 1.00  
## Mueslix\_Crispy\_Blend 13 160 25 3 1.50 0.67  
## Multi-Grain\_Cheerios 6 90 25 1 1.00 1.00  
## Nut&Honey\_Crunch 9 40 25 2 1.00 0.67  
## Nutri-Grain\_Almond-Raisin 7 130 25 3 1.33 0.67  
## Nutri-grain\_Wheat 2 90 25 3 1.00 1.00  
## Oatmeal\_Raisin\_Crisp 10 120 25 3 1.25 0.50  
## Post\_Nat.\_Raisin\_Bran 14 260 25 3 1.33 0.67  
## Product\_19 3 45 100 3 1.00 1.00  
## Puffed\_Rice 0 15 0 3 0.50 1.00  
## Puffed\_Wheat 0 50 0 3 0.50 1.00  
## Quaker\_Oat\_Squares 6 110 25 3 1.00 0.50  
## Raisin\_Bran 12 240 25 2 1.33 0.75  
## Raisin\_Nut\_Bran 8 140 25 3 1.00 0.50  
## Raisin\_Squares 6 110 25 3 1.00 0.50  
## Rice\_Chex 2 30 25 1 1.00 1.13  
## Rice\_Krispies 3 35 25 1 1.00 1.00  
## Shredded\_Wheat 0 95 0 1 0.83 1.00  
## Shredded\_Wheat\_'n'Bran 0 140 0 1 1.00 0.67  
## Shredded\_Wheat\_spoon\_size 0 120 0 1 1.00 0.67  
## Smacks 15 40 25 2 1.00 0.75  
## Special\_K 3 55 25 1 1.00 1.00  
## Strawberry\_Fruit\_Wheats 5 90 25 2 1.00 1.00  
## Total\_Corn\_Flakes 3 35 100 3 1.00 1.00  
## Total\_Raisin\_Bran 14 230 100 3 1.50 1.00  
## Total\_Whole\_Grain 3 110 100 3 1.00 1.00  
## Triples 3 60 25 3 1.00 0.75  
## Trix 12 25 25 2 1.00 1.00  
## Wheat\_Chex 3 115 25 1 1.00 0.67  
## Wheaties 3 110 25 1 1.00 1.00  
## Wheaties\_Honey\_Gold 8 60 25 1 1.00 0.75  
## rating  
## 100%\_Bran 68.40297  
## 100%\_Natural\_Bran 33.98368  
## All-Bran 59.42551  
## All-Bran\_with\_Extra\_Fiber 93.70491  
## Apple\_Cinnamon\_Cheerios 29.50954  
## Apple\_Jacks 33.17409  
## Basic\_4 37.03856  
## Bran\_Chex 49.12025  
## Bran\_Flakes 53.31381  
## Cap'n'Crunch 18.04285  
## Cheerios 50.76500  
## Cinnamon\_Toast\_Crunch 19.82357  
## Clusters 40.40021  
## Cocoa\_Puffs 22.73645  
## Corn\_Chex 41.44502  
## Corn\_Flakes 45.86332  
## Corn\_Pops 35.78279  
## Count\_Chocula 22.39651  
## Cracklin'\_Oat\_Bran 40.44877  
## Crispix 46.89564  
## Crispy\_Wheat\_&\_Raisins 36.17620  
## Double\_Chex 44.33086  
## Froot\_Loops 32.20758  
## Frosted\_Flakes 31.43597  
## Frosted\_Mini-Wheats 58.34514  
## Fruit\_&\_Fibre\_Dates,\_Walnuts,\_and\_Oats 40.91705  
## Fruitful\_Bran 41.01549  
## Fruity\_Pebbles 28.02576  
## Golden\_Crisp 35.25244  
## Golden\_Grahams 23.80404  
## Grape\_Nuts\_Flakes 52.07690  
## Grape-Nuts 53.37101  
## Great\_Grains\_Pecan 45.81172  
## Honey\_Graham\_Ohs 21.87129  
## Honey\_Nut\_Cheerios 31.07222  
## Honey-comb 28.74241  
## Just\_Right\_Crunchy\_\_Nuggets 36.52368  
## Just\_Right\_Fruit\_&\_Nut 36.47151  
## Kix 39.24111  
## Life 45.32807  
## Lucky\_Charms 26.73451  
## Maypo 54.85092  
## Muesli\_Raisins,\_Dates,\_&\_Almonds 37.13686  
## Muesli\_Raisins,\_Peaches,\_&\_Pecans 34.13976  
## Mueslix\_Crispy\_Blend 30.31335  
## Multi-Grain\_Cheerios 40.10596  
## Nut&Honey\_Crunch 29.92429  
## Nutri-Grain\_Almond-Raisin 40.69232  
## Nutri-grain\_Wheat 59.64284  
## Oatmeal\_Raisin\_Crisp 30.45084  
## Post\_Nat.\_Raisin\_Bran 37.84059  
## Product\_19 41.50354  
## Puffed\_Rice 60.75611  
## Puffed\_Wheat 63.00565  
## Quaker\_Oat\_Squares 49.51187  
## Raisin\_Bran 39.25920  
## Raisin\_Nut\_Bran 39.70340  
## Raisin\_Squares 55.33314  
## Rice\_Chex 41.99893  
## Rice\_Krispies 40.56016  
## Shredded\_Wheat 68.23588  
## Shredded\_Wheat\_'n'Bran 74.47295  
## Shredded\_Wheat\_spoon\_size 72.80179  
## Smacks 31.23005  
## Special\_K 53.13132  
## Strawberry\_Fruit\_Wheats 59.36399  
## Total\_Corn\_Flakes 38.83975  
## Total\_Raisin\_Bran 28.59278  
## Total\_Whole\_Grain 46.65884  
## Triples 39.10617  
## Trix 27.75330  
## Wheat\_Chex 49.78744  
## Wheaties 51.59219  
## Wheaties\_Honey\_Gold 36.18756

#Since we are dealing with very different variables, we must normalize the data

normcereal <- scale(cerealscomplete)  
summary(normcereal) #now our data is scaled.

## calories protein fat sodium   
## Min. :-2.8738 Min. :-1.40687 Min. :-0.9932 Min. :-1.9616   
## 1st Qu.:-0.3541 1st Qu.:-0.47733 1st Qu.:-0.9932 1st Qu.:-0.3306   
## Median : 0.1498 Median :-0.01256 Median : 0.0000 Median : 0.2131   
## Mean : 0.0000 Mean : 0.00000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.1498 3rd Qu.: 0.45221 3rd Qu.: 0.0000 3rd Qu.: 0.6661   
## Max. : 2.6695 Max. : 3.24083 Max. : 3.9729 Max. : 1.9045   
## fiber carbo sugars potass   
## Min. :-0.89778 Min. :-2.50014 Min. :-1.6306 Min. :-1.1783   
## 1st Qu.:-0.79462 1st Qu.:-0.70143 1st Qu.:-0.9424 1st Qu.:-0.8079   
## Median :-0.07249 Median :-0.05903 Median :-0.0248 Median :-0.1201   
## Mean : 0.00000 Mean : 0.00000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.34015 3rd Qu.: 0.58337 3rd Qu.: 0.8928 3rd Qu.: 0.3031   
## Max. : 4.87925 Max. : 2.12512 Max. : 1.8104 Max. : 3.2660   
## vitamins shelf weight cups   
## Min. :-1.3032 Min. :-1.4617 Min. :-3.4600 Min. :-2.4251   
## 1st Qu.:-0.1818 1st Qu.:-1.1612 1st Qu.:-0.2008 1st Qu.:-0.6432   
## Median :-0.1818 Median :-0.2599 Median :-0.2008 Median :-0.3038   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.:-0.1818 3rd Qu.: 0.9420 3rd Qu.:-0.2008 3rd Qu.: 0.7568   
## Max. : 3.1822 Max. : 0.9420 Max. : 3.0583 Max. : 2.8780   
## rating   
## Min. :-1.7336   
## 1st Qu.:-0.7071   
## Median :-0.1510   
## Mean : 0.0000   
## 3rd Qu.: 0.5807   
## Max. : 3.6578

#Now we will use Agnes to compare single, complete, and average linkage and Ward.

set.seed(17)  
hc\_single <- agnes(normcereal, method = "single")  
hc\_complete <- agnes(normcereal, method = "complete")  
hc\_average <- agnes(normcereal, method = "average")  
hc\_ward <- agnes(normcereal, method = "ward")  
  
#The we compare agglomerative coefficients  
print(hc\_single$ac) #0.6067859

## [1] 0.6067859

print(hc\_complete$ac) #0.8353712

## [1] 0.8353712

print(hc\_average$ac) #0.7766075

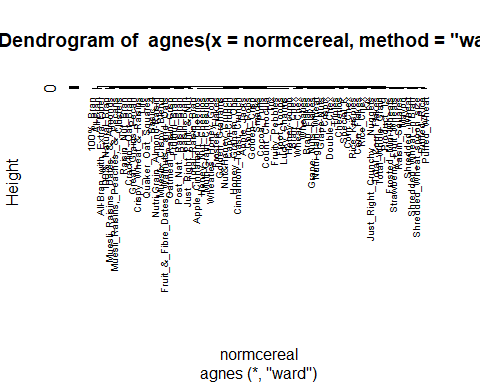
## [1] 0.7766075

print(hc\_ward$ac) #0.9046042 (this one is the highest)

## [1] 0.9046042

#Based on the agglomerative coefficients, the ward method is the best. #we will now plot the dendrogram

plot\_ward <- pltree(hc\_ward, cex = 0.6, hang = -1)

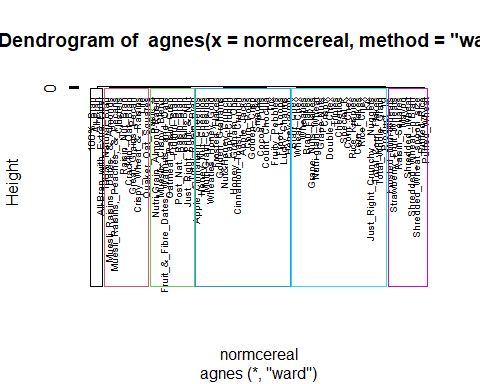


plot\_ward

## NULL

#Looking at the dendrogram, I think 6 clusters seems pretty good.

pltree(hc\_ward, cex = 0.6, hang = -1)  
rect.hclust(hc\_ward,k=6,border = 1:6)

 #Now we check their stability. #First, we partition.

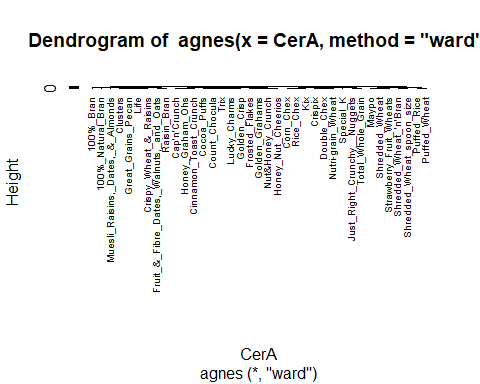
set.seed(17)  
  
#Creating partition set of 50%  
partset <- createDataPartition(normcereal[,"calories"], p=0.5, list = FALSE)  
CerA <- normcereal[partset,] #this is the first 50%  
CerB <- normcereal[-partset,]#second set  
  
#Now we'll run the AGNES to both sets.   
a\_ward <-agnes(CerA, method = "ward")  
print(a\_ward$ac) #0.8498682

## [1] 0.8498682

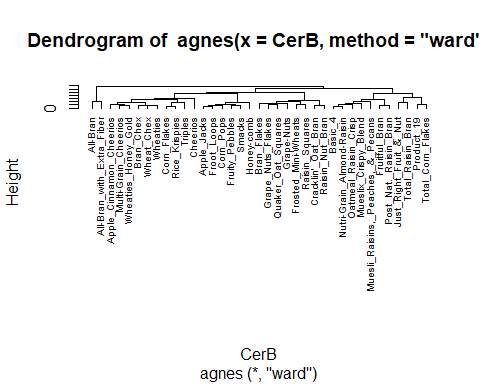
b\_ward <-agnes(CerB, method = "ward")  
print(b\_ward$ac) #0.8383771

## [1] 0.8383771

#we can note that the agglomerative coefficients are very similar.   
  
plot\_a <- pltree(a\_ward, cex=0.6, hang = -1)



plot\_b <- pltree(b\_ward, cex=0.6, hang = -1)

 If I compare these partitioned dendrograms to the original, we can see that most cereals remain in the same cluster. This tells us that it is pretty stable.

#Senario: Find cluster of healthy cereal. Okay, since we want healthy cereals we should be considering the nutrional items in this data set and their importance. We probably are looking at some that should be high and some that would be better low. high: protein, fiber, potassium, vitamins, and honestly rating since we want the kids to actually eat it.  
low: fat, sodium now calories and carbs are also important, but I dont think we specifically want them high or low since it needs to be balanced between healthy and filling. For the purpose of this analysis, I would to fet rid of shelf.  
To be honest, looking at this data, its a bit unbalanced because of the serving sizes which I assume is what cups is. If we really wanted to consider these on an equal playing field, we should probably make a ratio of the nutritional aspect per its cups so that all the nutritional data is scaled similarly. ie, a cereal doesn’t look healthier because its just less food. Also, due to the importance of scale here, you would want to manipulate your data depending on health benefit importance, so normalizing the data would not be the first option since we do care about the actual values of the nutritional facts.