

Assignment_5

Dengcheng Chen

04/17/2022

Data Preprocessing. Remove all cereals with missing values

```
library(readr)
C.df <- read.csv('Cereals.csv', header=TRUE)
C.df <- na.omit(C.df)
rownames(C.df) <- C.df$name
C.df$name = NULL
C.df <- C.df[,3:15]
C.df <- scale(C.df)
summary(C.df)
```

```
##      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
```

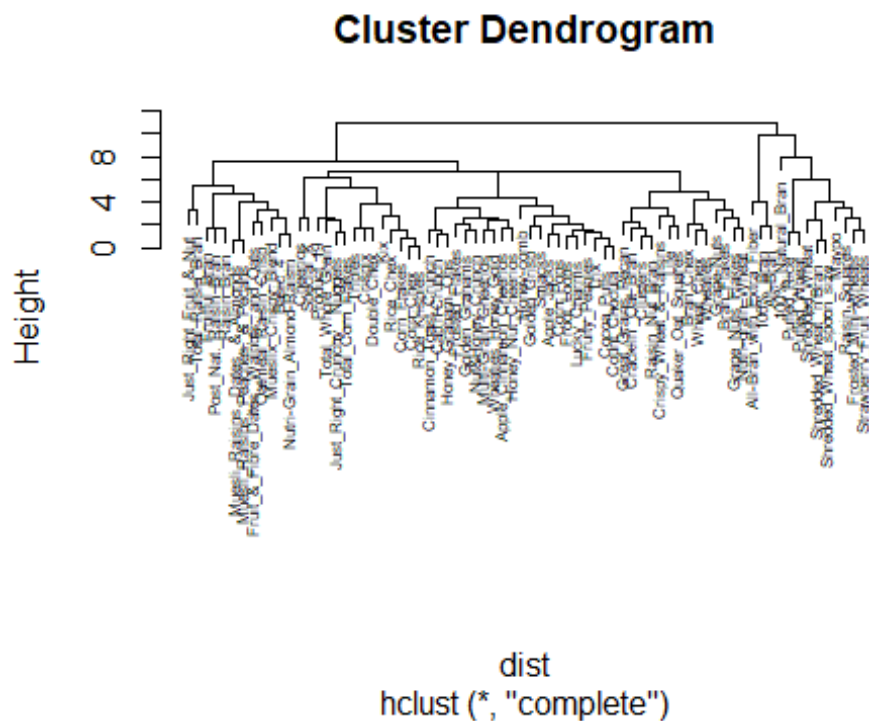
```
library(factoextra)

## Warning: ³Ï°ü'factoextra'ÊÇÕÃR°æ±¼4.1.3 À¼°ÔìµÄ

## 载入需要的程辑包: ggplot2

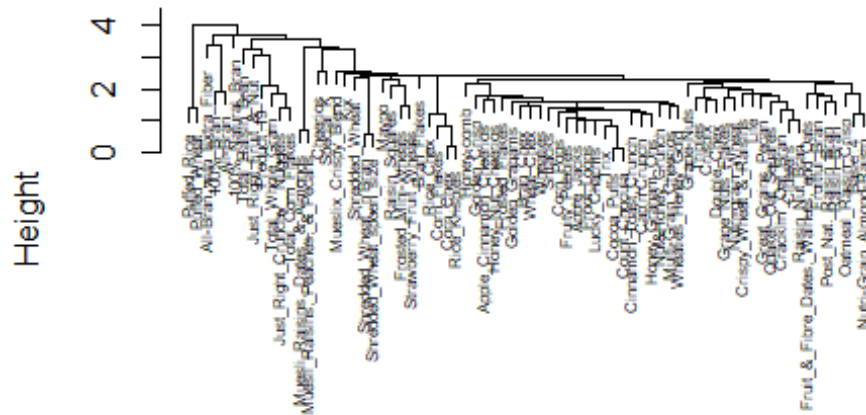
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

dist <- dist(C.df, method = "euclidean")
HC_C <- hclust(dist, method = "complete")
HC_S <- hclust(dist, method = "single")
HC_A <- hclust(dist, method = "average")
HC_W <- hclust(dist, method = "ward.D2")
plot(HC_C, cex=0.5, hang=0.1)
```



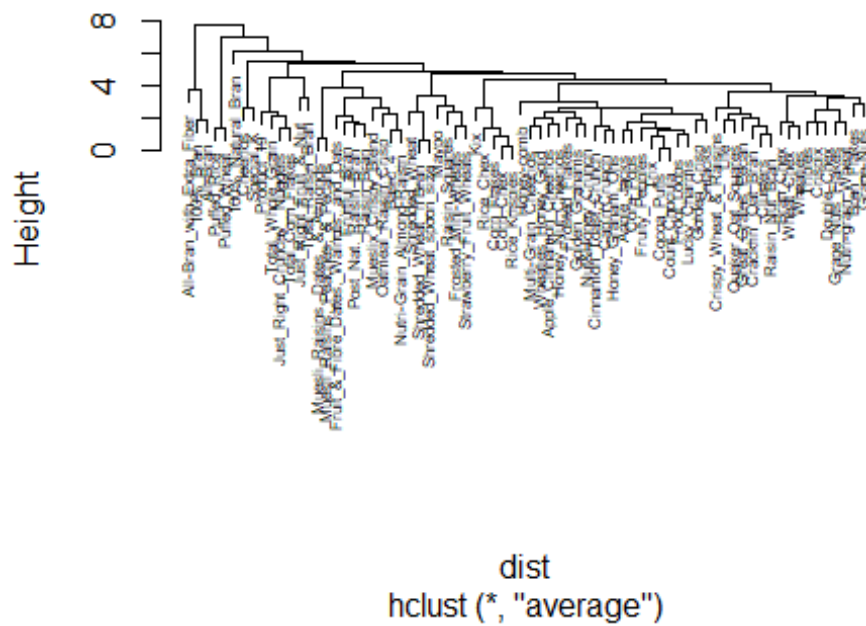
```
plot(HC_S, cex=0.5, hang=0.1)
```

Cluster Dendrogram



```
plot(HC_A, cex=0.5, hang=0.1)
```

Cluster Dendrogram



```
plot(HC_W, cex=0.5, hang=0.1)
```

[illegible]

Q2How many

```
plot(HC_W, cex=0.5, hang=0.1)
rect.hclust(HC_W, k = 5, border = 1:5)
```

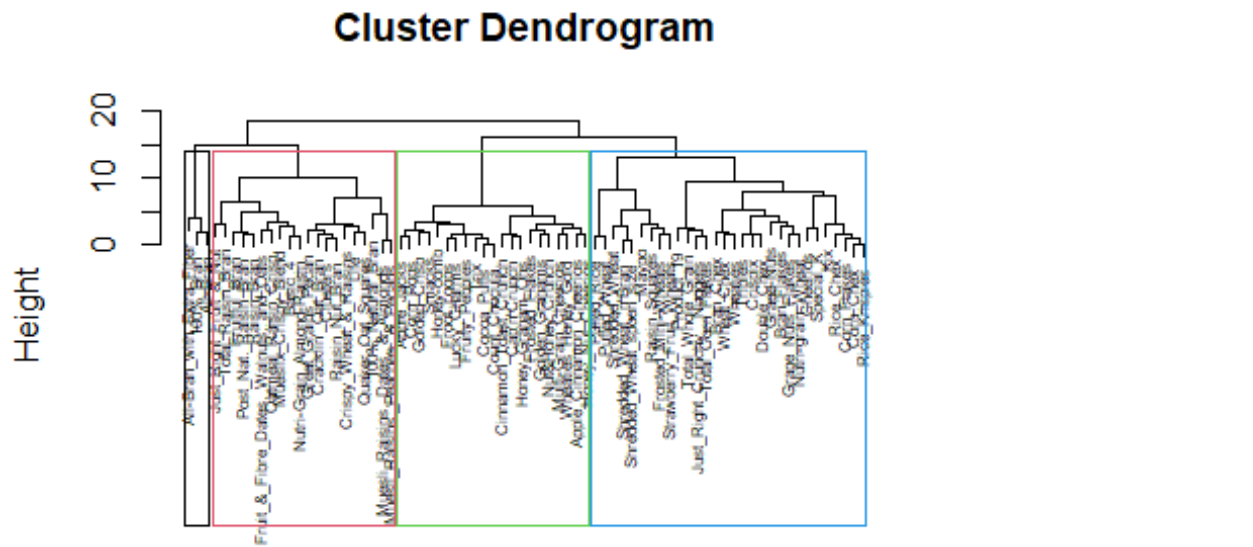
```
plot(HC_W, cex=0.5, hang=0.1)
rect.hclust(HC_W, k = 5, border = 1:5)
```

Height

dist
hclust (*, "ward.D2")

Wine & Fruit, Wine & Fruit
Just_Turn_01, Just_Turn_01
Post_Nut_01, Post_Nut_01
Fruit & Fibre, Dates, Walnuts, Pine Apples
Walnut, Orange, Apple
Nutri-Guard, Avocado, Apple
Cocoa, Cocoa
Crispy, Walnut & Raisins
Crispy, Orange, Strawberry
Muesli & Raisins, Dates, & Strawberries
Apple, Raisins
Garden, Raisins
Honey, Cocoa
Lucky, Raisins
Fruity, Raisins
Cocoa, Raisins
Cinnamon, Raisins, Cocoa
Honey, Raisins, Cocoa
Nutri-Guard, Raisins
Walnuts, Raisins, Cocoa
Apple, Raisins, Cocoa
Smoother, Raisins
Smoother, Raisins, Walnut
Smoother, Raisins, Walnut
Fruited Raisins, Raisins
Smoother, Raisins
Just_Right_Toddler, Raisins
Walnut, Raisins
Dough, Raisins
Ginger, Raisins
Nutri-Guard, Raisins
Raisins, Raisins
Raisins, Raisins
Cocoa, Raisins

```
plot(HC_W, cex=0.5, hang=0.1)
rect.hclust(HC_W, k = 4, border = 1:4)
```



```
dist
hclust (*, "ward.D2")
```

```
plot(HC_W, cex=0.5, hang=0.1)
rect.hclust(HC_W, k = 6, border = 1:6)
```


Q3The elementary public schools would like to choose a set of cereals to include in their daily cafeterias. Every day a different cereal is offered, but all cereals should support a healthy diet. For this goal, you are requested to find a cluster of “healthy cereals.”

For elementary public schools, the definition of healthy should be high protein, high vitamin, and reasonable calories. So, I will choose the 1st cluster counting from the right as the “healthy cereals”