# Assignment-5

Anjali Nangunuri

2023-11-27

```
# Load the data
cereals <- read.csv("C:/Users/Anjali/Desktop/Anjali_FML_Assignment 5/Cereals.csv")

# Remove rows with missing values
cereals <- na.omit(cereals)
```

```
install.packages("cluster")
```

```
## Installing package into 'C:/Users/Anjali/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##         binary source needs_compilation
## cluster  2.1.4  2.1.5              TRUE

## installing the source package 'cluster'

## Warning in install.packages("cluster"): installation of package 'cluster' had
## non-zero exit status
```

```
install.packages("factoextra")
```

```
## Installing package into 'C:/Users/Anjali/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'factoextra' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Anjali\AppData\Local\Temp\RtmpU33xMs\downloaded_packages
```

```
install.packages("tidyverse")
```

```
## Installing package into 'C:/Users/Anjali/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\Anjali\AppData\Local\Temp\RtmpU33xMs\downloaded_packages
```

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(cluster)
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.3.2
```

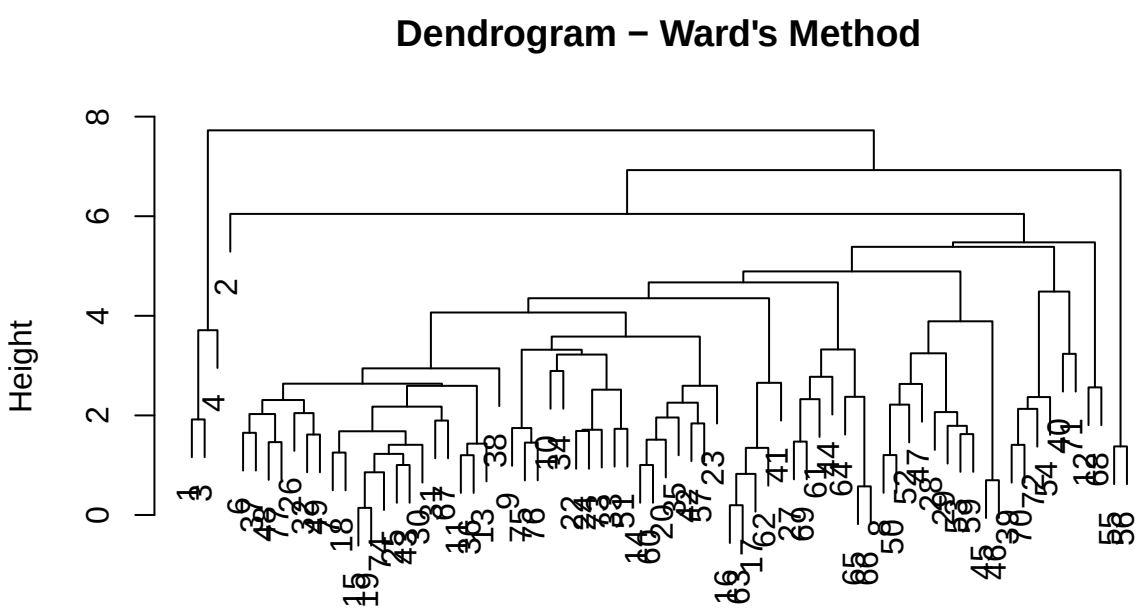
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```r
# Step 2: Hierarchical Clustering
# Identify numeric columns
numeric_columns <- sapply(cereals, is.numeric)

# Normalize only the numeric columns
normalized_cereals <- scale(cereals[, numeric_columns])

# Apply hierarchical clustering with Ward's method
ward_cluster <- agnes(normalized_cereals)

# Visualize dendrogram for Ward's method
plot(ward_cluster, which.plots = 2, main = "Dendrogram - Ward's Method")
```




```
# Step 3: Cluster Stability and Healthy Cereals
# Create cluster partitions A and B
set.seed(123)
partition_A <- sample(1:2, nrow(normalized_cereals), replace = TRUE)
partition_B <- 3 - partition_A

# Fit cluster on partition A
cluster_A <- cutree(ward_cluster, k = 3)

# Use cluster centroids from A to assign records in partition B
cluster_B <- cluster_A[partition_B]

# Assess cluster consistency
consistency <- sum(cluster_A == cluster_B) / length(cluster_B)

# Identify healthy cereals cluster
# (Based on your analysis and specific criteria for 'healthy')
# Example: Let's assume 'healthy' cereals have low sugar and high fiber
healthy_cereals_cluster <- cluster_A[which(cereals$sugars < 5 & cereals$fiber > 5)]

# Print or visualize the results
cat("Cluster Consistency:", consistency, "\n")
```

```
## Cluster Consistency: 0.5810811
```

```r
cat("Healthy Cereals Cluster:", healthy_cereals_cluster, "\n")
```

```
## Healthy Cereals Cluster: 1
```