FML - Assignment-4

Rashed Syed

2023-11-20

```
library(tidyverse)  # For Data manipulation
```

```
## -- 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 1.9.3     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
```

```
library(factoextra)  # Used for clustering algorithms and visualization
```

```
## 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
```

```
library(dplyr)
library(ggplot2)
library(cluster)
```

```
Pharmaceuticals <- read.csv("C:/Users/kunal/OneDrive/Desktop/Rashed/Pharmaceuticals.csv")
```

```
# Task 1: Use only the numerical variables (1 to 9) to cluster the 21 firms.
# Prior to clustering data, remove the missing data and rescale variables for comparability.

Pharma_data <- na.omit(Pharmaceuticals)
Pharma_data
```

```
##    Symbol                              Name Market_Cap Beta PE_Ratio  ROE  ROA
## 1     ABT             Abbott Laboratories      68.44 0.32     24.7 26.4 11.8
## 2     AGN                  Allergan, Inc.       7.58 0.41     82.5 12.9  5.5
## 3     AHM                    Amersham plc       6.30 0.46     20.7 14.9  7.8
## 4     AZN                AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4
## 5     AVE                        Aventis      47.16 0.32     20.1 21.8  7.5
## 6     BAY                      Bayer AG      16.90 1.11     27.9  3.9  1.4
## 7     BMY  Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8 15.1
```

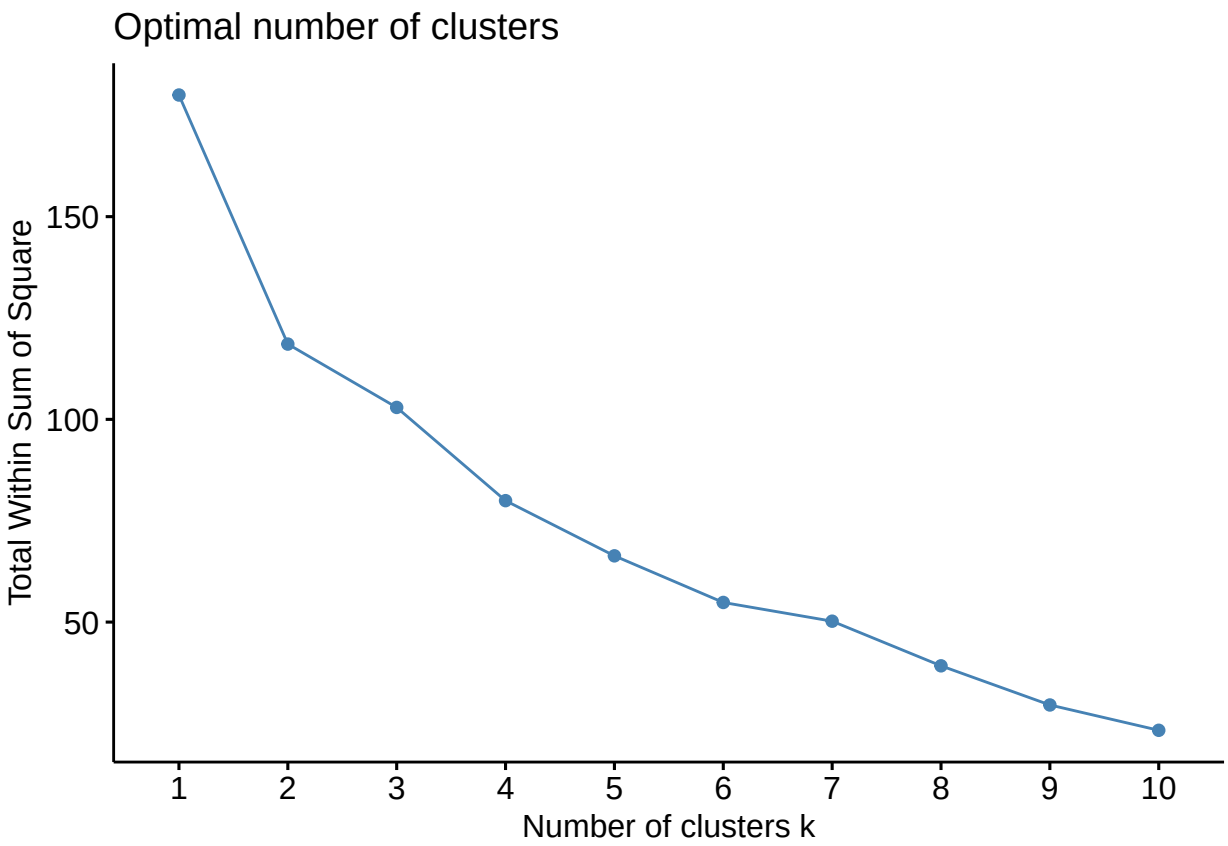
```
## 8    CHTT                        Chattem, Inc       0.41 0.85     26.0 24.1  4.3
## 9     ELN              Elan Corporation, plc       0.78 1.08      3.6 15.1  5.1
## 10    LLY              Eli Lilly and Company      73.84 0.18     27.9 31.0 13.5
## 11    GSK                GlaxoSmithKline plc     122.11 0.35     18.0 62.9 20.3
## 12    IVX                   IVAX Corporation       2.60 0.65     19.9 21.4  6.8
## 13    JNJ                  Johnson & Johnson     173.93 0.46     28.4 28.6 16.3
## 14    MRX Medicis Pharmaceutical Corporation       1.20 0.75     28.6 11.2  5.4
## 15    MRK                  Merck & Co., Inc.     132.56 0.46     18.9 40.6 15.0
## 16    NVS                        Novartis AG      96.65 0.19     21.6 17.9 11.2
## 17    PFE                         Pfizer Inc     199.47 0.65     23.6 45.6 19.2
## 18    PHA              Pharmacia Corporation      56.24 0.40     56.5 13.5  5.7
## 19    SGP        Schering-Plough Corporation      34.10 0.51     18.9 22.6 13.3
## 20    WPI       Watson Pharmaceuticals, Inc.       3.26 0.24     18.4 10.2  6.8
## 21    WYE                             Wyeth      48.19 0.63     13.1 54.9 13.4
##    Asset_Turnover Leverage Rev_Growth Net_Profit_Margin Median_Recommendation
## 1             0.7     0.42       7.54              16.1          Moderate Buy
## 2             0.9     0.60       9.16               5.5          Moderate Buy
## 3             0.9     0.27       7.05              11.2            Strong Buy
## 4             0.9     0.00      15.00              18.0         Moderate Sell
## 5             0.6     0.34      26.81              12.9          Moderate Buy
## 6             0.6     0.00      -3.17               2.6                 Hold
## 7             0.9     0.57       2.70              20.6         Moderate Sell
## 8             0.6     3.51       6.38               7.5          Moderate Buy
## 9             0.3     1.07      34.21              13.3         Moderate Sell
## 10            0.6     0.53       6.21              23.4                 Hold
## 11            1.0     0.34      21.87              21.1                 Hold
## 12            0.6     1.45      13.99              11.0                 Hold
## 13            0.9     0.10       9.37              17.9          Moderate Buy
## 14            0.3     0.93      30.37              21.3          Moderate Buy
## 15            1.1     0.28      17.35              14.1                 Hold
## 16            0.5     0.06      -2.69              22.4                 Hold
## 17            0.8     0.16      25.54              25.2          Moderate Buy
## 18            0.6     0.35      15.00               7.3                 Hold
## 19            0.8     0.00       8.56              17.6                 Hold
## 20            0.5     0.20      29.18              15.1         Moderate Sell
## 21            0.6     1.12       0.36              25.5                 Hold
##       Location Exchange
## 1           US     NYSE
## 2       CANADA     NYSE
## 3           UK     NYSE
## 4           UK     NYSE
## 5       FRANCE     NYSE
## 6      GERMANY     NYSE
## 7           US     NYSE
## 8           US   NASDAQ
## 9      IRELAND     NYSE
## 10          US     NYSE
## 11          UK     NYSE
## 12          US     AMEX
## 13          US     NYSE
## 14          US     NYSE
## 15          US     NYSE
## 16 SWITZERLAND     NYSE
## 17          US     NYSE
```

```
## 18          US     NYSE
## 19          US     NYSE
## 20          US     NYSE
## 21          US     NYSE
```

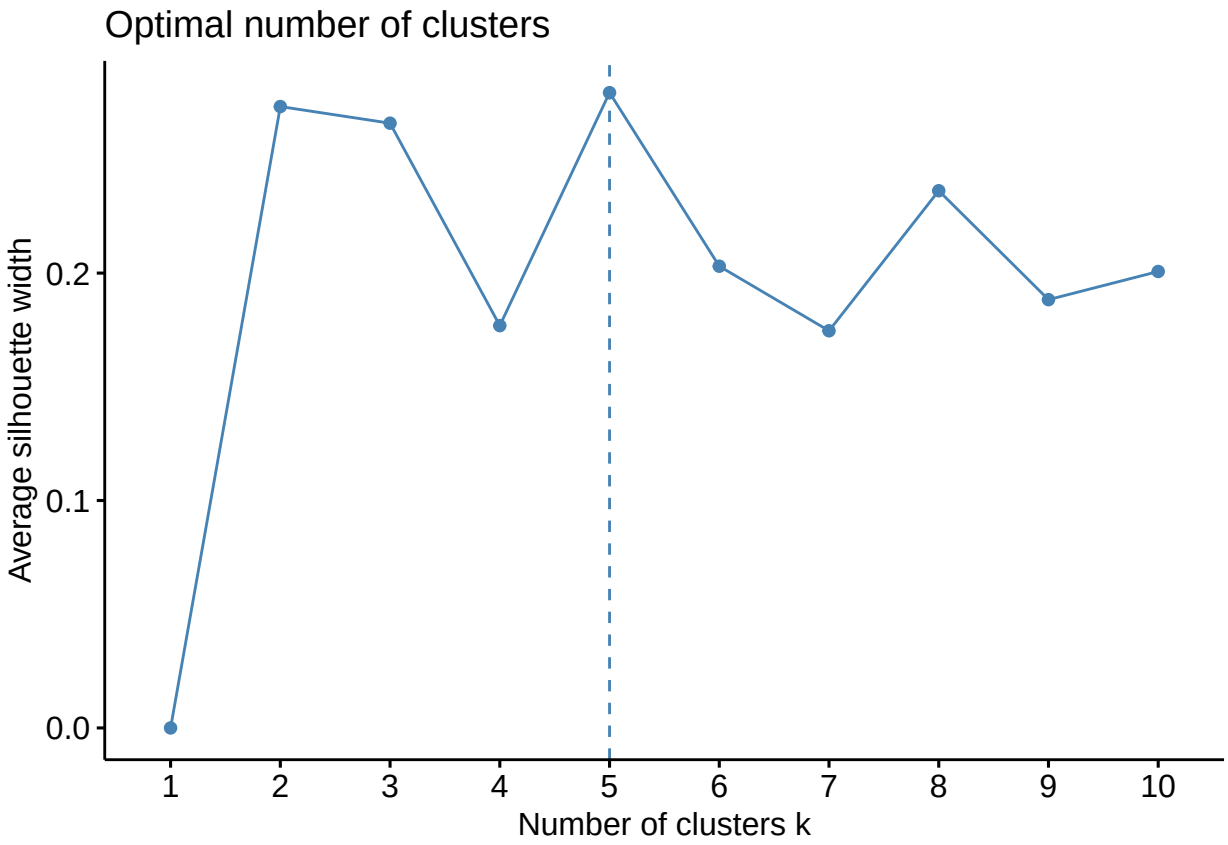
```r
# Taking the quantitative variables(1-9) to cluster the 21 firms
row.names(Pharma_data) <- Pharma_data[, 1]
Pharma_data1 <- Pharma_data[, 3:11]  # Considering only numerical values (columns 3-11)
```

```r
# Normalizing the data frame with scale method
Pharma_data2 <- scale(Pharma_data1)
```

```r
# Determine the number of clusters using Elbow Method
fviz_nbclust(Pharma_data2, kmeans, method = "wss")
```

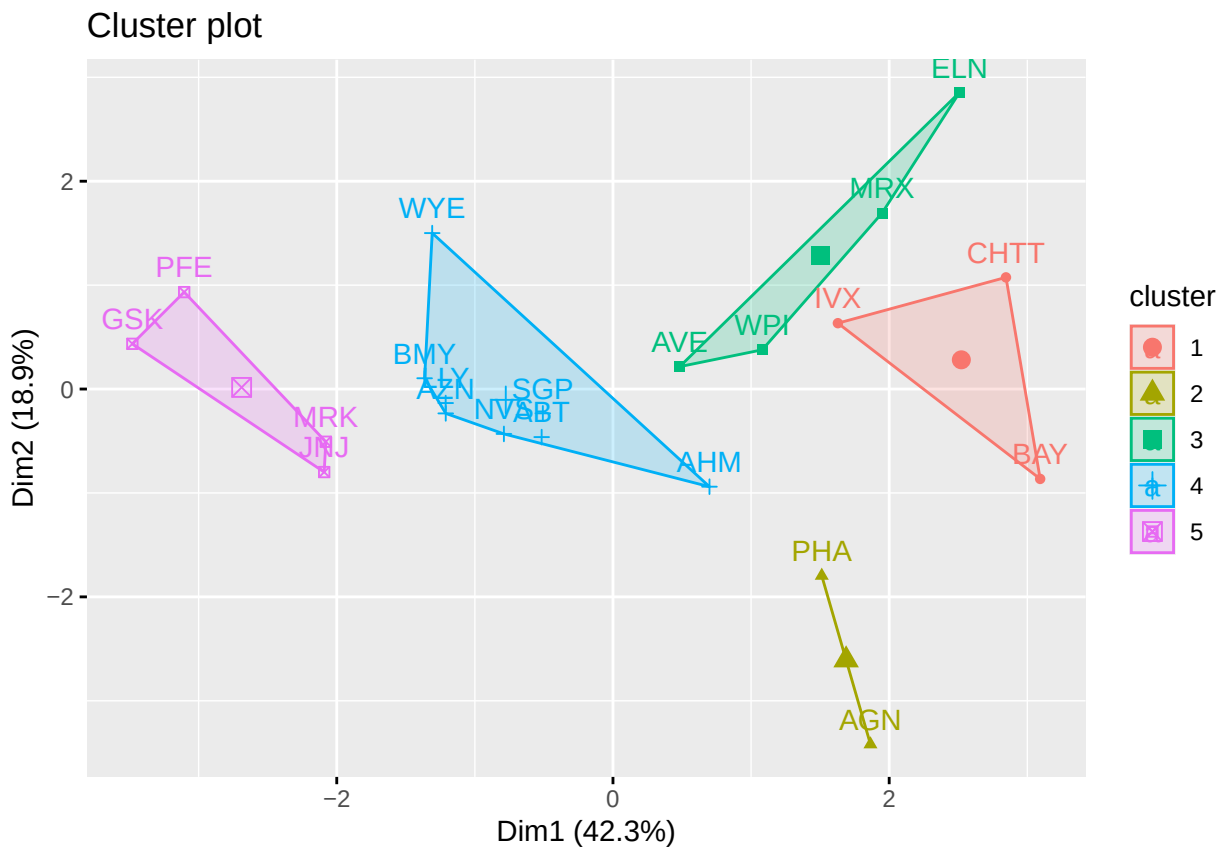



```r
# Determine the number of clusters using Silhouette method
fviz_nbclust(Pharma_data2, kmeans, method = "silhouette")
```

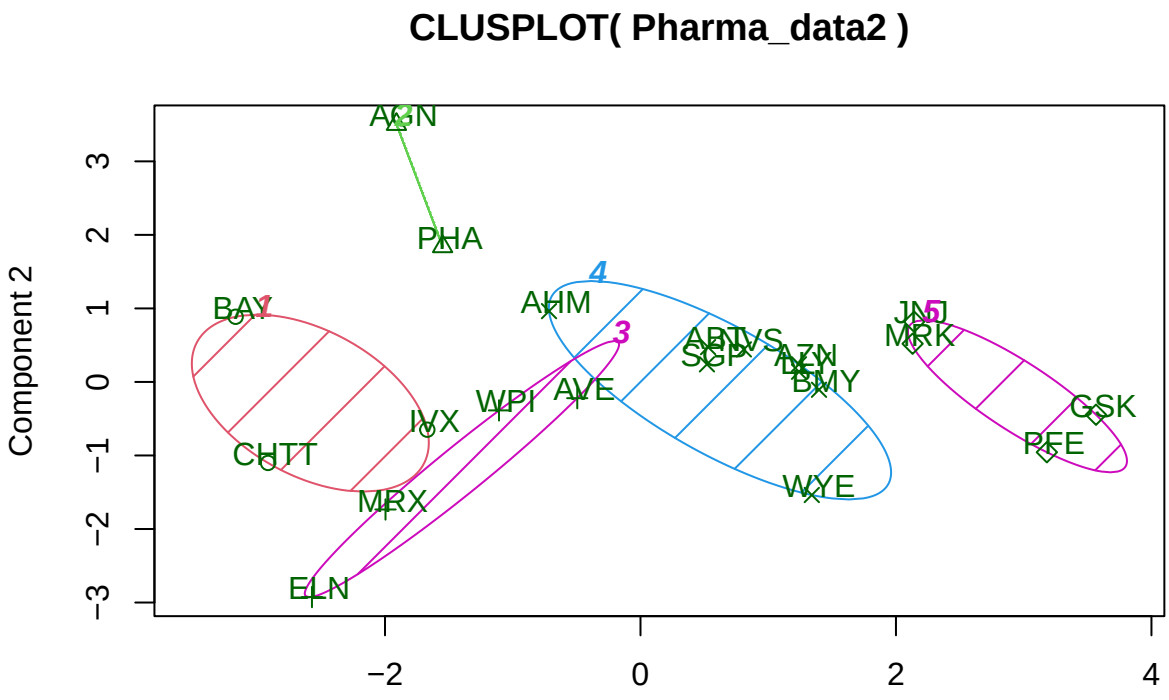



```
# Applying K-means with k=5
set.seed(64060)
k_5 <- kmeans(Pharma_data2, centers = 5, nstart = 25)
```

```
# Visualizing the clusters
fviz_cluster(k_5, data = Pharma_data2)
```




```r
# Task 2: Interpret the clusters with respect to the numerical variables used in forming the clusters.
# By noticing the mean values of all quantitative variables for each cluster

# ... (Cluster interpretation comments)

# Task 3: Is there a pattern in the clusters with respect to the numerical variables (10 to 12)?
# (those not used in forming the clusters)

# Note: It seems there is no analysis related to variables 10 to 12 in the provided code.

# Task 4: Provide an appropriate name for each cluster using any or all of the variables in the dataset
# ... (Cluster naming comments)

# Viewing the cluster plot
clusplot(Pharma_data2, k_5$cluster, color = TRUE, shade = TRUE, labels = 2, lines = 0)
```



These two components explain 61.23 % of the point variability.