# FML ASSIGNMENT 4

Sri Chandana

2023-11-19

```
# The necessary packages are loaded
library(caret)
```

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

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

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

```
library(dplyr)
```

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

```
library(tidyverse)
```

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

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
## v readr     2.1.4
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
#install.packages("cowplot")
library(cowplot)
```

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

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
#install.packages("flexclust")
library(flexclust)
```

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

```
## Loading required package: grid
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

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

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

```
#install.packages("NbClust")
library(NbClust)
```

```
# It imports the "Pharmaceuticals" dataset from the specified file path
Pharmacy <- read.csv("C://Users//srich//OneDrive//Desktop//FML//DATASETS//Pharmaceuticals.csv")
```

```r
# The "Pharmacy" dataset will be viewed
view(Pharmacy)
```

```r
#  It displays the first few rows of the "Pharmacy" dataset
head(Pharmacy)
```

```
##   Symbol                Name Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover
## 1    ABT Abbott Laboratories      68.44 0.32     24.7 26.4 11.8            0.7
## 2    AGN      Allergan, Inc.       7.58 0.41     82.5 12.9  5.5            0.9
## 3    AHM        Amersham plc       6.30 0.46     20.7 14.9  7.8            0.9
## 4    AZN     AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4            0.9
## 5    AVE             Aventis      47.16 0.32     20.1 21.8  7.5            0.6
## 6    BAY            Bayer AG      16.90 1.11     27.9  3.9  1.4            0.6
##   Leverage Rev_Growth Net_Profit_Margin Median_Recommendation Location Exchange
## 1     0.42       7.54              16.1          Moderate Buy       US     NYSE
## 2     0.60       9.16               5.5          Moderate Buy   CANADA     NYSE
## 3     0.27       7.05              11.2            Strong Buy       UK     NYSE
## 4     0.00      15.00              18.0         Moderate Sell       UK     NYSE
## 5     0.34      26.81              12.9          Moderate Buy   FRANCE     NYSE
## 6     0.00      -3.17               2.6                  Hold  GERMANY     NYSE
```

```r
# It displays the summary statistics for the "Pharmacy" dataset
summary(Pharmacy)
```

```
##     Symbol              Name             Market_Cap          Beta
##  Length:21          Length:21          Min.   :  0.41   Min.   :0.1800
##  Class :character   Class :character   1st Qu.:  6.30   1st Qu.:0.3500
##  Mode  :character   Mode  :character   Median : 48.19   Median :0.4600
##                                        Mean   : 57.65   Mean   :0.5257
##                                        3rd Qu.: 73.84   3rd Qu.:0.6500
##                                        Max.   :199.47   Max.   :1.1100
##     PE_Ratio          ROE             ROA        Asset_Turnover    Leverage
##  Min.   : 3.60   Min.   : 3.9   Min.   : 1.40   Min.   :0.3    Min.   :0.0000
##  1st Qu.:18.90   1st Qu.:14.9   1st Qu.: 5.70   1st Qu.:0.6    1st Qu.:0.1600
##  Median :21.50   Median :22.6   Median :11.20   Median :0.6    Median :0.3400
##  Mean   :25.46   Mean   :25.8   Mean   :10.51   Mean   :0.7    Mean   :0.5857
##  3rd Qu.:27.90   3rd Qu.:31.0   3rd Qu.:15.00   3rd Qu.:0.9    3rd Qu.:0.6000
##  Max.   :82.50   Max.   :62.9   Max.   :20.30   Max.   :1.1    Max.   :3.5100
##    Rev_Growth    Net_Profit_Margin Median_Recommendation   Location
##  Min.   :-3.17   Min.   : 2.6      Length:21              Length:21
##  1st Qu.: 6.38   1st Qu.:11.2      Class :character       Class :character
##  Median : 9.37   Median :16.1      Mode  :character       Mode  :character
##  Mean   :13.37   Mean   :15.7
##  3rd Qu.:21.87   3rd Qu.:21.1
##  Max.   :34.21   Max.   :25.5
##    Exchange
##  Length:21
##  Class :character
##  Mode  :character
##
##
##
```

```r
#a. Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made

# Calculates the column wise mean of missing values in the "Pharmacy" dataset
colMeans(is.na(Pharmacy))
```

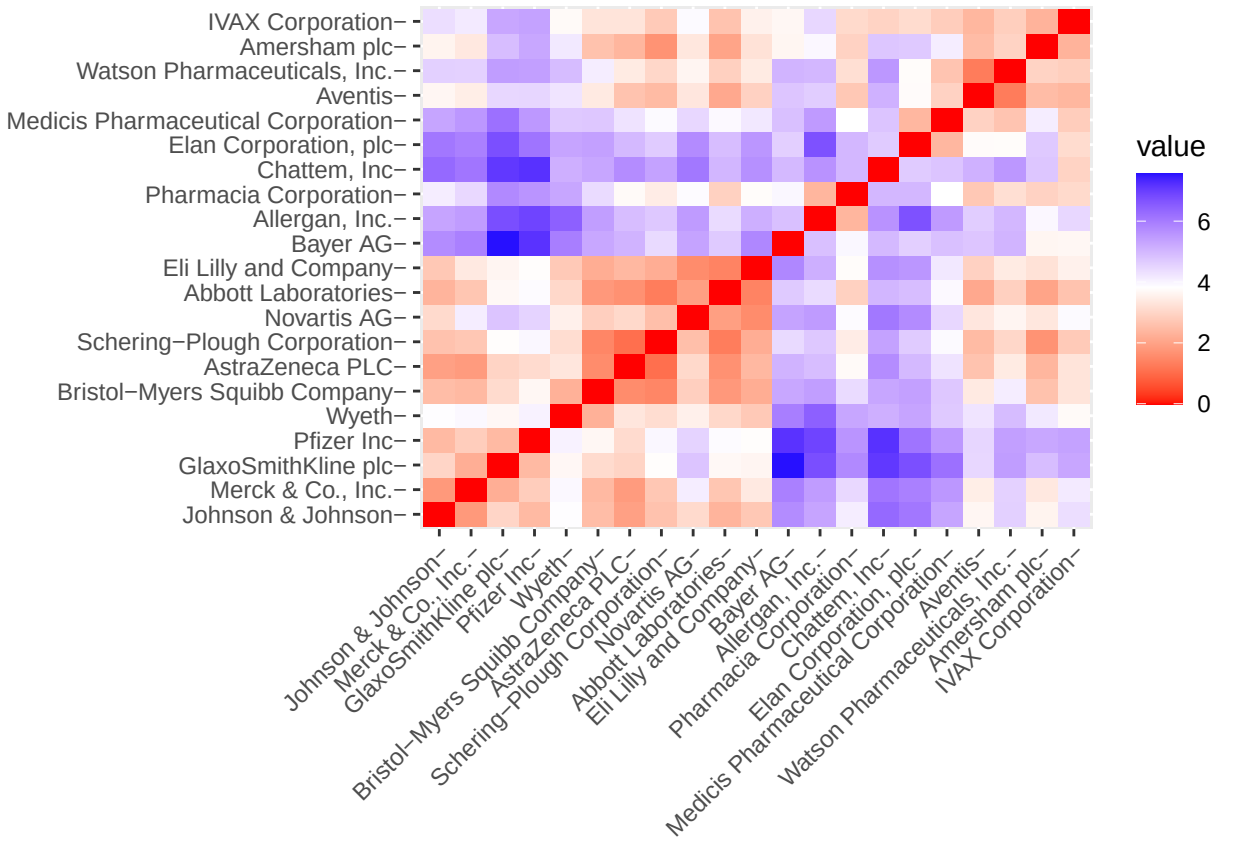
```
##                Symbol                  Name            Market_Cap
##                     0                     0                     0
##                  Beta              PE_Ratio                   ROE
##                     0                     0                     0
##                   ROA        Asset_Turnover              Leverage
##                     0                     0                     0
##            Rev_Growth     Net_Profit_Margin Median_Recommendation
##                     0                     0                     0
##              Location              Exchange
##                     0                     0
```

```r
# Sets row names of "Pharmacy" to the values in the second column.
row.names(Pharmacy) <- Pharmacy[,2]
# Removes the second column from the "Pharmacy" dataset
Pharmacy <- Pharmacy[,-2]
#  Removes the first column and columns 11 to 13 from the updated "Pharmacy" dataset
Pharmacy.1 <- Pharmacy[,-c(1,11:13)]
```

```r
# Checks the dimensions of the "Pharmacy" dataset
dim(Pharmacy)
```

```
## [1] 21 13
```

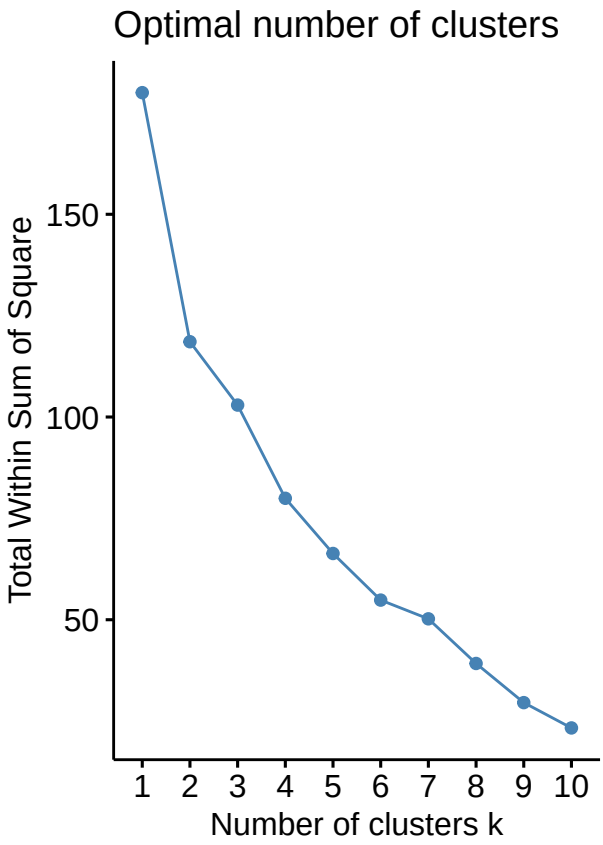
```r
# Standardizes the columns of "Pharmacy.1" using the scale function
norm.Pharmacy.1 <- scale(Pharmacy.1)
# Calculates the distance matrix based on the standardized data
dist <- get_dist(norm.Pharmacy.1)
# Visualizes the distance matrix using function
fviz_dist(dist)
```

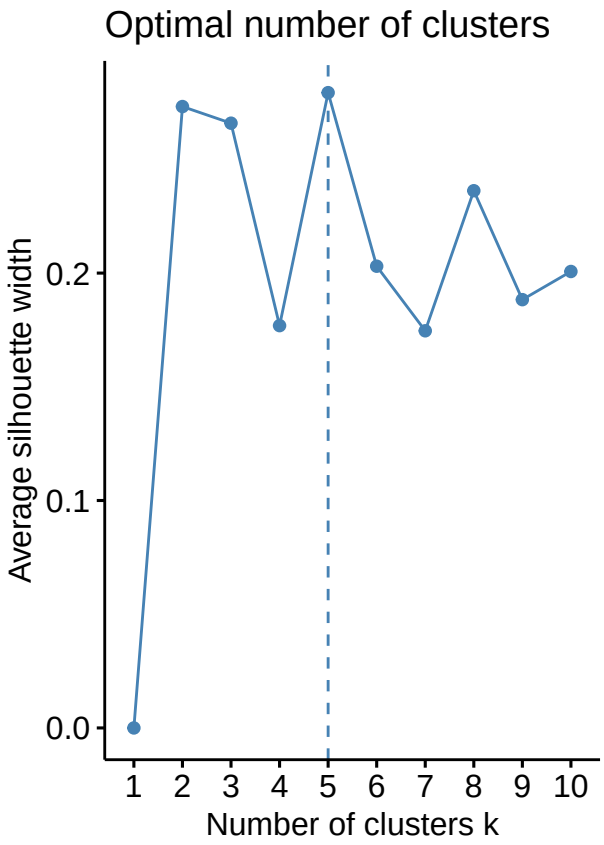



```
# The chart shows how the color intensity changes as we move across distances. As expected,the diagonal

# For finding the best K Value: The Elbow chart and the Silhouette Method are effective ways to decide
```

```
# Calculates Within Cluster Sum of Squares (WSS) for different numbers of clusters using the k-means al
WSS <- fviz_nbclust(norm.Pharmacy.1, kmeans, method = "wss")
# Calculates Silhouette scores for different numbers of clusters using the k-means algorithm
Sil <- fviz_nbclust(norm.Pharmacy.1, kmeans, method = "silhouette")
# Displays the plots of WSS and Silhouette scores
plot_grid(WSS, Sil)
```




```
# The charts indicate different optimal values for k, the Elbow Method suggests k=2, while the Silhouet
```

```
# Set the seed for reproducibility
# Performs k-means clustering on the normalized "Pharmacy.1" data with 5 centers
# Displays the cluster centers obtained from the k-means clustering
set.seed(123)
KMeans.Pharmacy.Opt <- kmeans(norm.Pharmacy.1, centers = 5, nstart = 50)
KMeans.Pharmacy.Opt$centers
```

```
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
## 2 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 3 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
## 4  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 5 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.27449312 -0.7041516       0.556954446
## 2  1.36644699 -0.6912914      -1.320000179
## 3 -0.14170336 -0.1168459      -1.416514761
## 4 -0.46807818  0.4671788       0.591242521
## 5  0.06308085  1.5180158      -0.006893899
```

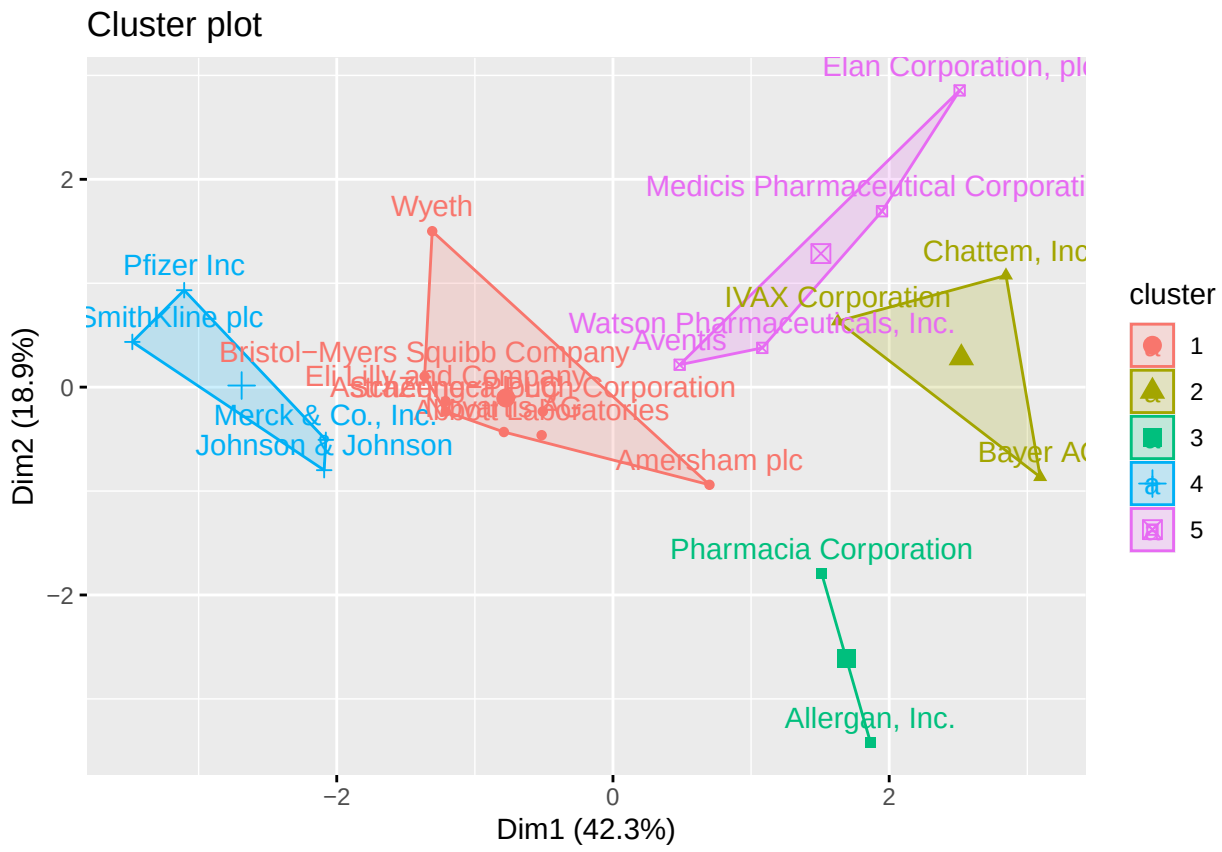
```
# Display the size of each cluster
KMeans.Pharmacy.Opt$size
```

```
## [1] 8 3 2 4 4
```

```
# Display the within-cluster sum of squares
KMeans.Pharmacy.Opt$withinss
```

```
## [1] 21.879320 15.595925  2.803505  9.284424 12.791257
```

```
# Visualize the k-means clusters using a scatter plot
fviz_cluster(KMeans.Pharmacy.Opt, data = norm.Pharmacy.1)
```




```
# Using the dataset,we identified five clusters based on their proximity to core points. Cluster 4 stan
# On the other hand, Cluster 5 is characterized by a low Asset Turnover.Examining the size of each clus
# The within-cluster sum of squared distances provides insights into data dispersion: Cluster 1 (21.9)
```

```
#b. Interpret the clusters with respect to the numerical variables used in forming the clusters.

# Set the seed for reproducibility
# Performs k-means clustering on the normalized "Pharmacy.1" data with 3 clusters
# Displays the cluster centers

set.seed(123)
KMeans.Pharmacy <- kmeans(norm.Pharmacy.1, centers = 3, nstart = 50)
KMeans.Pharmacy$centers
```

```
##   Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.6125361  0.2698666  1.3143935 -0.9609057 -1.0174553      0.2306328
## 2  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
## 3 -0.8261772  0.4775991 -0.3696184 -0.5631589 -0.8514589     -0.9994088
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3592866 -0.5757385        -1.3784169
## 2 -0.3331068 -0.2902163         0.6823310
## 3  0.8502201  0.9158889        -0.3319956
```

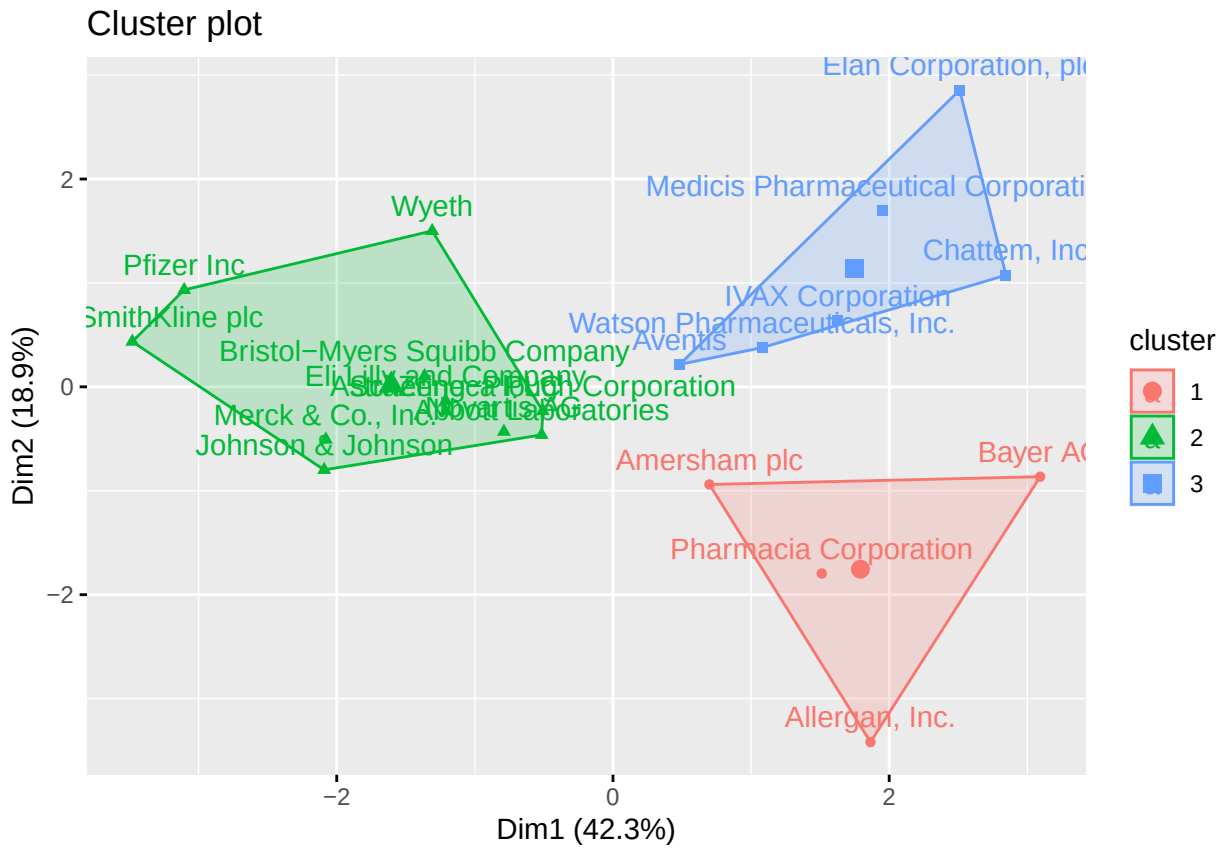
```
# Displays the sizes of each cluster obtained from the k-means clustering.
KMeans.Pharmacy$size
```

```
## [1]  4 11  6
```

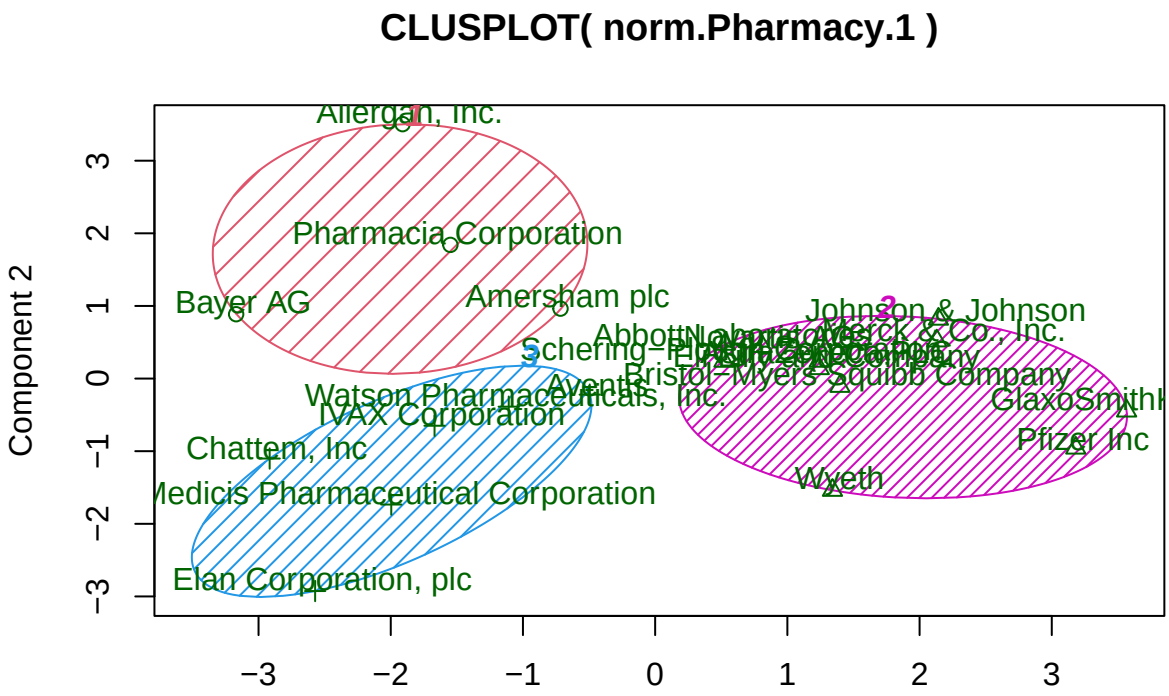
```
# Displays the within-cluster sum of squares for each cluster
KMeans.Pharmacy$withinss
```

```
## [1] 20.54199 43.30886 32.14336
```

```
# Visualize the k-means clusters using a scatter plot
fviz_cluster(KMeans.Pharmacy, data = norm.Pharmacy.1)
```



```
clusplot(norm.Pharmacy.1,KMeans.Pharmacy$cluster,color = TRUE,shade =TRUE, labels=2,lines=0)
```

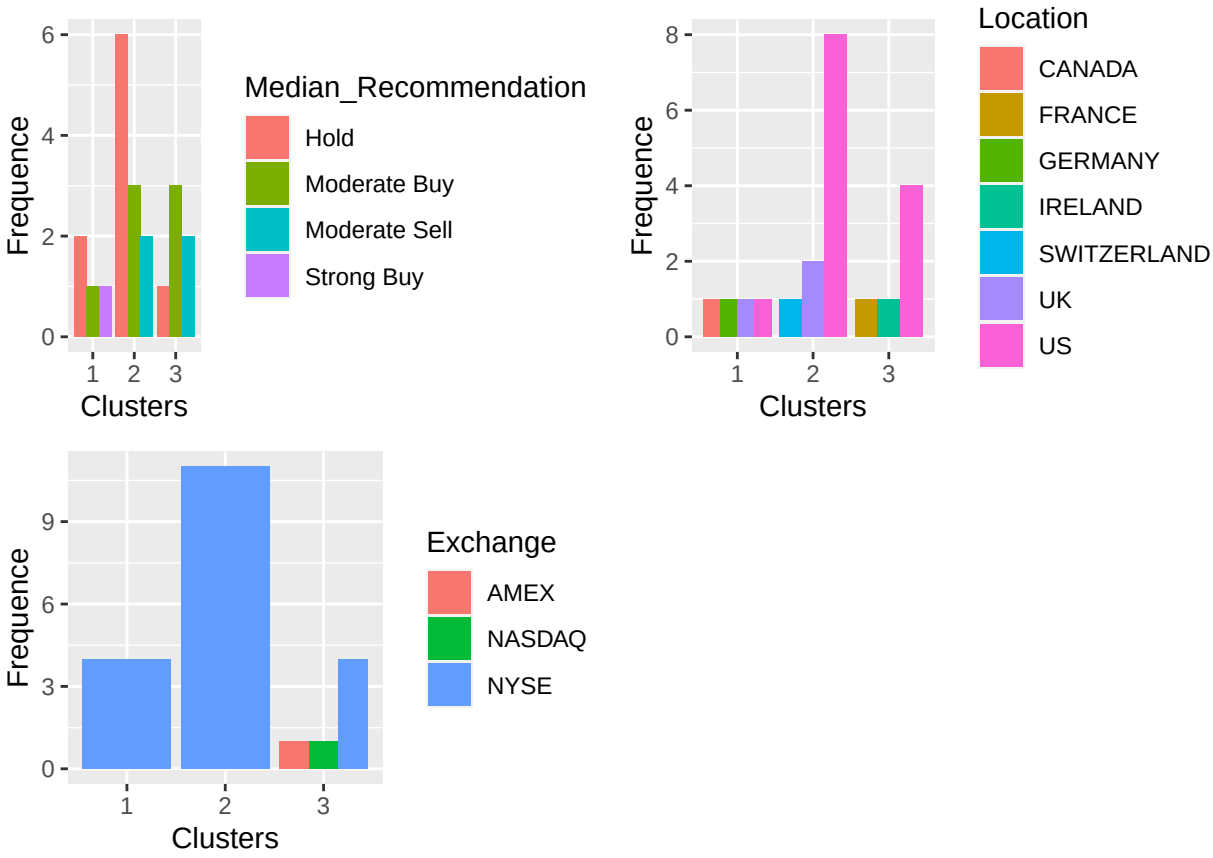



These two components explain 61.23 % of the point variability.

```
#c. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)?

# To explore patterns in the data for the last three categorical variables-Median Recommendation, Locat

Pharmacy.2 <-  Pharmacy%>% select(c(11,12,13)) %>%
    mutate(Cluster = KMeans.Pharmacy$cluster)
Med_Recom <- ggplot(Pharmacy.2, mapping = aes(factor(Cluster), fill=Median_Recommendation)) +
  geom_bar(position = 'dodge') +
  labs(x='Clusters', y='Frequence')
Loc <- ggplot(Pharmacy.2, mapping = aes(factor(Cluster), fill=Location)) +
  geom_bar(position = 'dodge') +
  labs(x='Clusters', y='Frequence')
Ex <- ggplot(Pharmacy.2, mapping = aes(factor(Cluster), fill=Exchange)) +
  geom_bar(position = 'dodge') +
  labs(x='Clusters', y='Frequence')
plot_grid(Med_Recom, Loc, Ex)
```




```
# The chart makes it clear that most companies in cluster 3 are from the United States, and all of them
```

```
#d. Provide an appropriate name for each cluster using any or all of the variables in the dataset.

#1) Cluster 1 - Global Giants: These companies are considered "overvalued international firms" because
#2) Cluster 2 - Growth Prospects: This group is labeled as "growing and leveraged firms" due to "Modera
#3) Cluster 3 - Stable US Companies: Companies in this cluster are characterized as "mature US firms" s
```