FML Assign4

Sushma

2023-10-27

```r
##Load the librabries

library(factoextra)
```

```
## Loading required package: ggplot2

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

```r
library(ggplot2)
library(tidyverse)
```

```
## -- 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 lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0

## -- 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(ISLR)
library(NbClust)
library(cluster)
```

```r
## Import the data from csv file.

Pharmaceuticals <- read.csv("C:/Users/niyas/Downloads/Pharmaceuticals.csv")
view(Pharmaceuticals)
```

a. Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made in conducting the cluster analysis, such as weights for different variables, the specific clustering algorithm(s) used, the number of clusters formed, and so on.

```r
##Create a new data frame 'A' by removing rows with missing values from 'Pharmaceuticals'
A <- na.omit(Pharmaceuticals)
##Generate a summary of the data in the 'A' data frame
summary(A)
```

```
##     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
#Set row names of the data frame 'A' to the values in its first column
row.names(A) <- A[,1]
#Create a new data frame 'Pharma' containing columns 3 to 11 from 'A'
Pharma <- A[,3:11]
#Display the rows of the 'Pharma' data frame
head(Pharma)
```

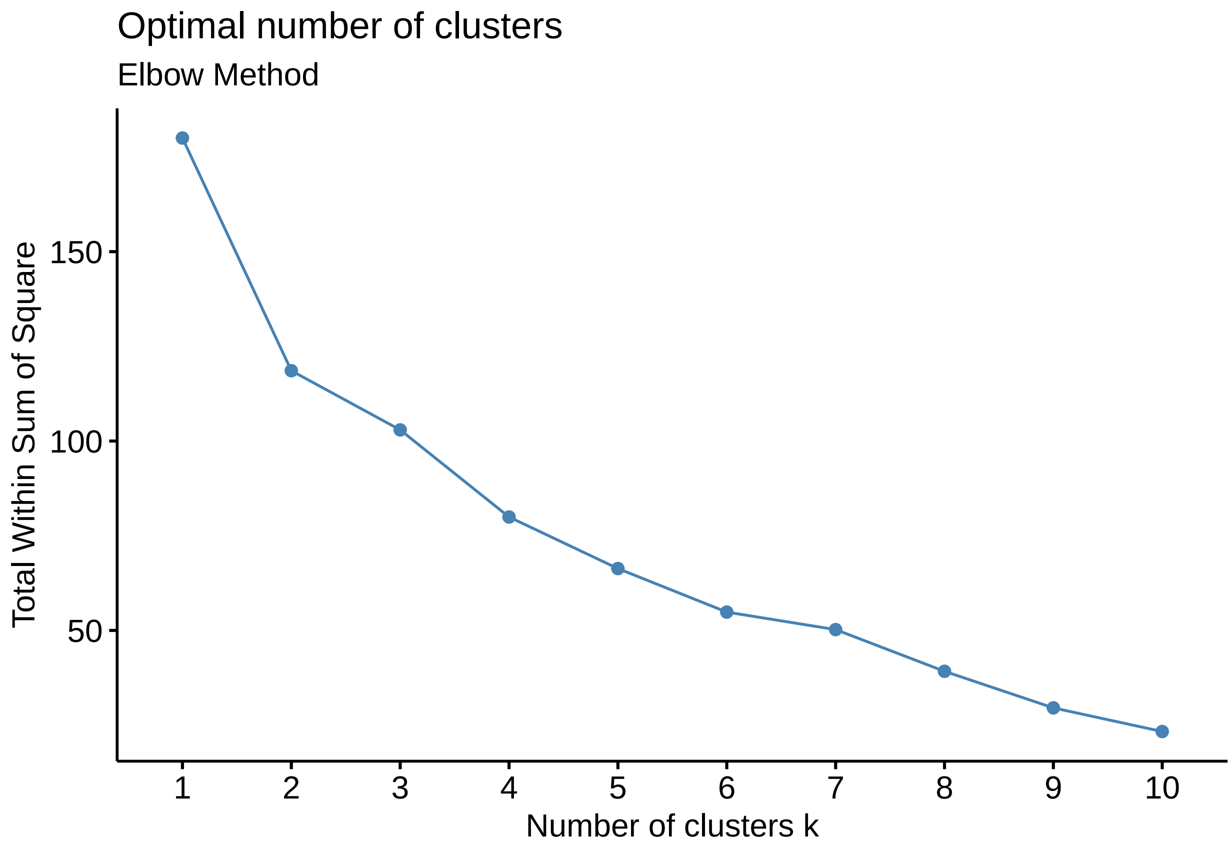
```
##     Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## ABT      68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## AGN       7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## AHM       6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
## AZN      67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## AVE      47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## BAY      16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
##     Net_Profit_Margin
## ABT              16.1
## AGN               5.5
## AHM              11.2
```

```
## AZN              18.0
## AVE              12.9
## BAY               2.6
```

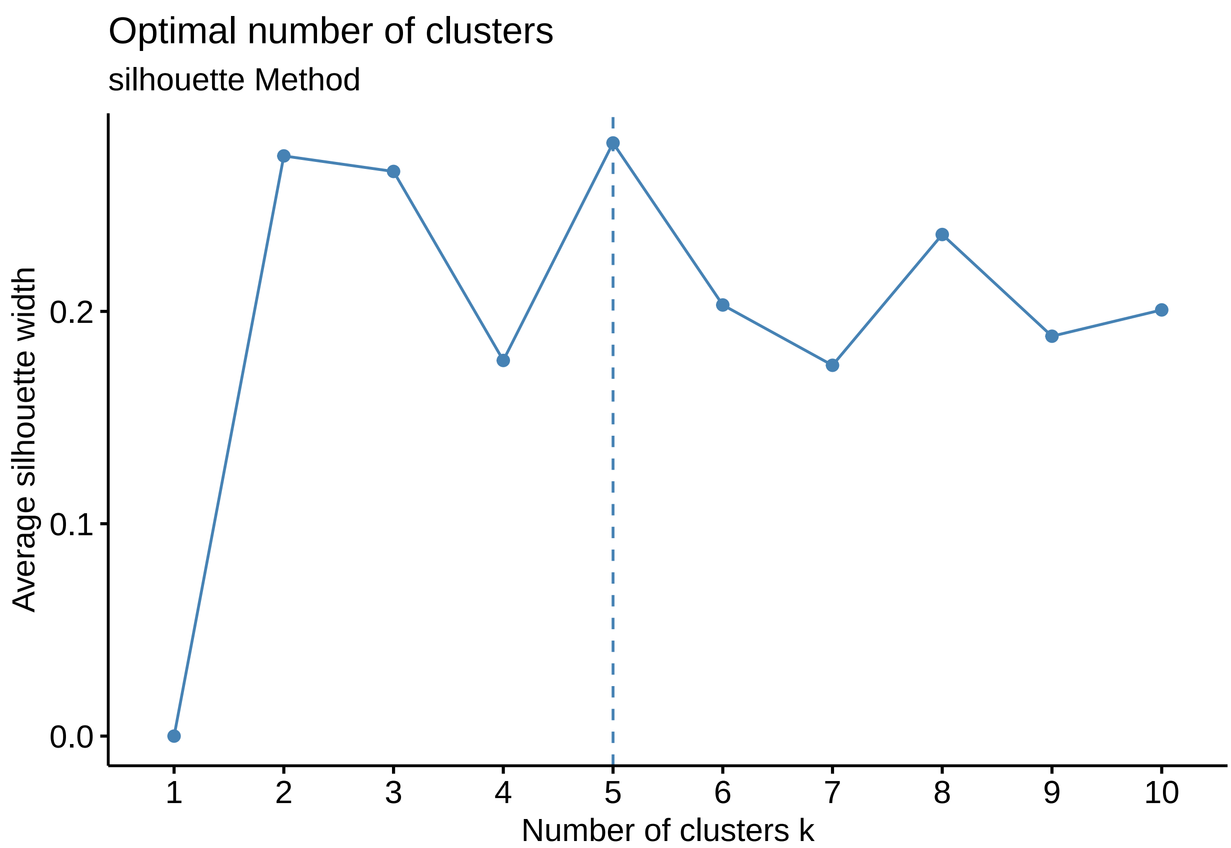
```r
# Scale the data in the 'Pharma' data frame to standardize variables
Pharma1 <- scale (Pharma)
# Display the rows of the scaled 'Pharma1' data frame
head(Pharma1)
```

```
##     Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## ABT  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## AGN -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## AHM -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## AZN  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## AVE -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## BAY -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
##       Leverage Rev_Growth Net_Profit_Margin
## ABT -0.2120979 -0.5277675        0.06168225
## AGN  0.0182843 -0.3811391       -1.55366706
## AHM -0.4040831 -0.5721181       -0.68503583
## AZN -0.7496565  0.1474473        0.35122600
## AVE -0.3144900  1.2163867       -0.42597037
## BAY -0.7496565 -1.4971443       -1.99560225
```

```r
# Visualize the Elbow Method for determining the optimal number of clusters (k) in k-means clustering a

fviz_nbclust(Pharma1, kmeans, method = "wss") + labs(subtitle = "Elbow Method")
```



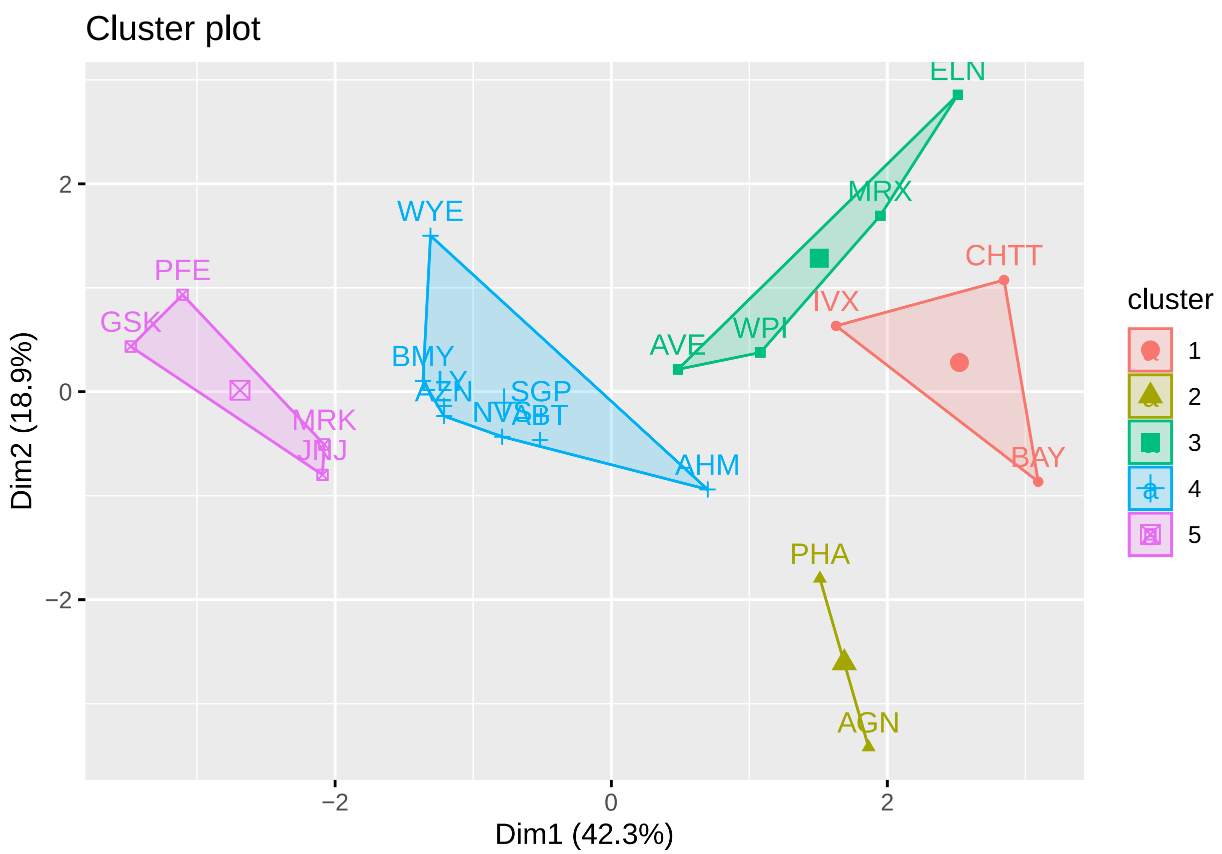

```
# Visualize the Silhouette Method for determining the optimal number of clusters (k) in k-means cluster

fviz_nbclust(Pharma1, kmeans, method = "silhouette") + labs(subtitle = "silhouette Method")
```



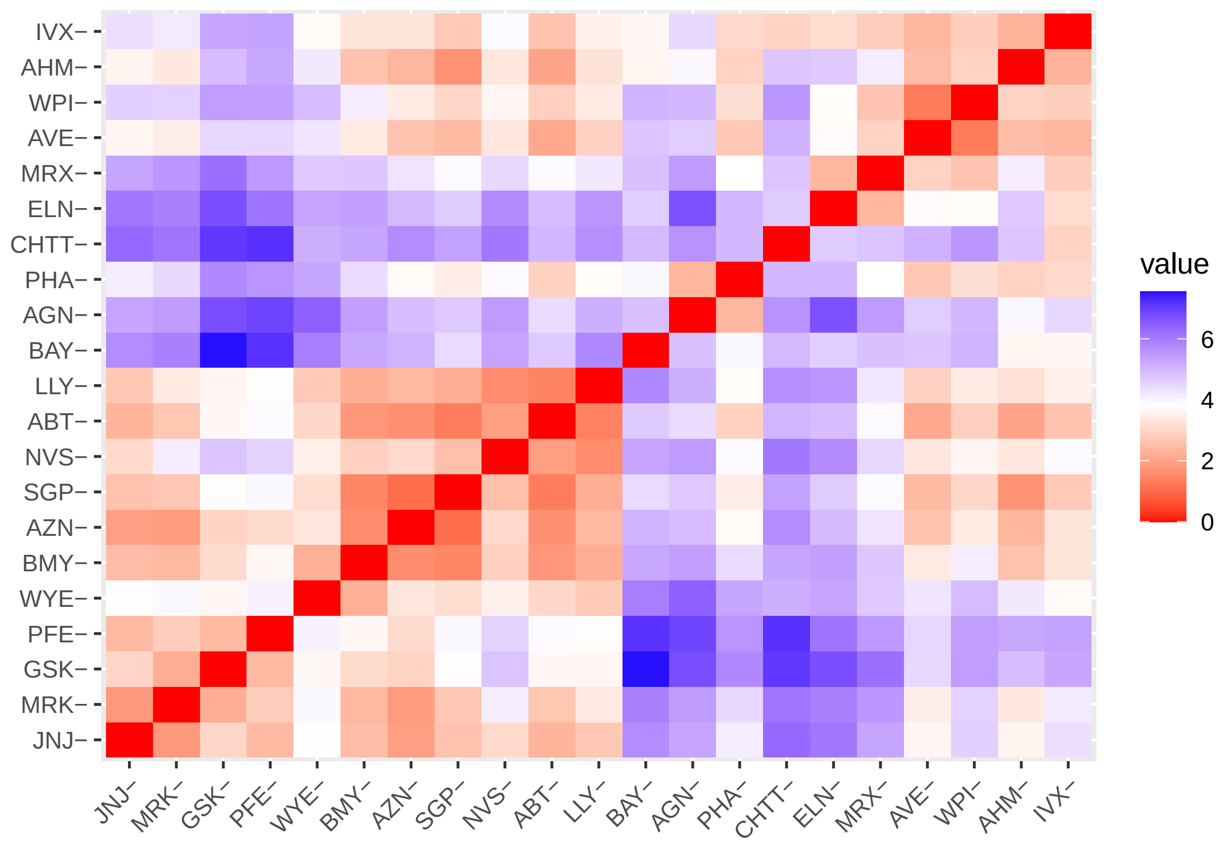

```r
# Set the seed for reproducibility at 64060
set.seed(64060)

# Perform k-means clustering with 5 clusters and 25 different starting points
k5 <- kmeans(Pharma1, centers = 5, nstart = 25)

# Display the cluster centers for the 5 clusters
k5$centers
```

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

```r
# Visualize the results of k-means clustering using the 'fviz_cluster' function.

fviz_cluster(k5, data = Pharma1)
```




```r
# Calculate the Euclidean distance matrix between observations in the 'Pharma1' dataset
distance <- dist(Pharma1, method = "euclidean")

# Visualizing the distance matrix using the 'fviz_dist' function
fviz_dist(distance)
```




```r
# Set the CRAN mirror to a specific location
options(repos = c(CRAN = "https://cran.rstudio.com/"))
```

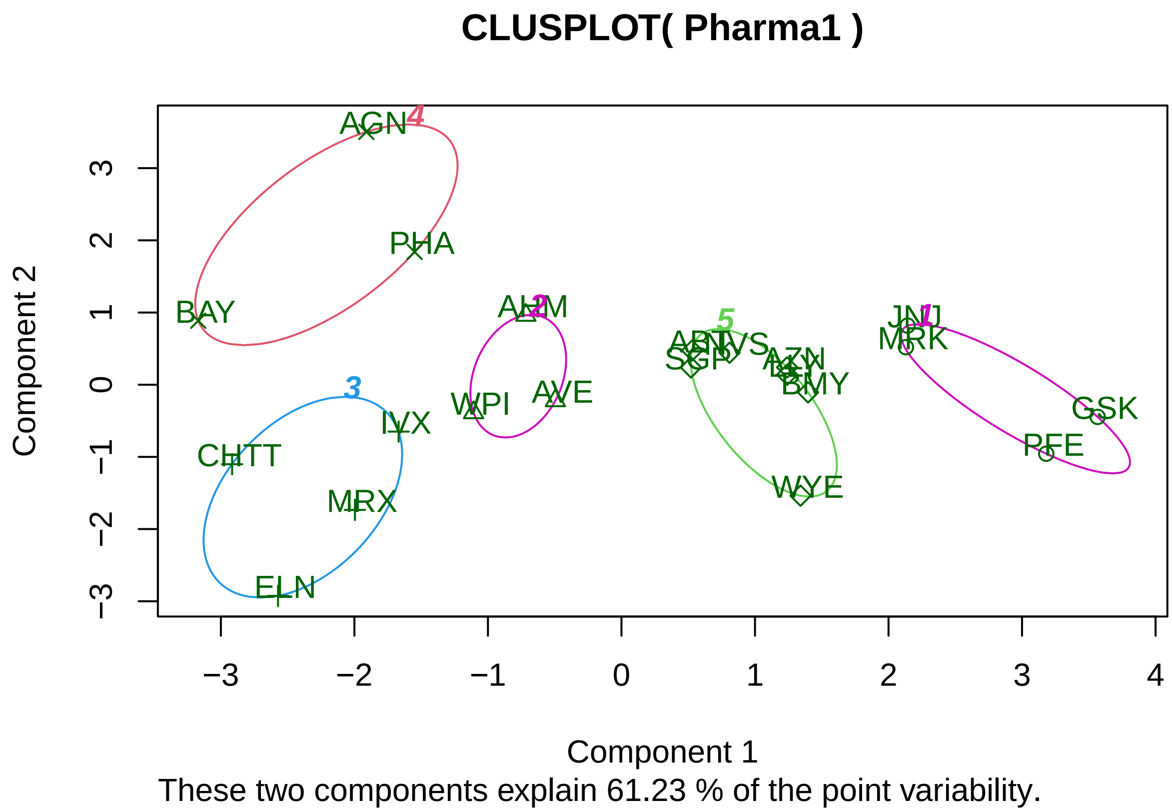
```r
# Performing k-means clustering on the 'Pharma1' dataset to create 5 clusters
fit <-kmeans(Pharma1, 5)

# Calculate and aggregate the mean values of variables within each cluster
aggregate(Pharma1, by = list(fit$cluster), FUN=mean)
```

```
##   Group.1  Market_Cap       Beta   PE_Ratio        ROE        ROA
## 1       1  1.69558112 -0.1780563 -0.1984582  1.2349879  1.3503431
## 2       2 -0.66114002 -0.7233539 -0.3512251 -0.6736441 -0.5915022
## 3       3 -0.96247577  1.1949250 -0.3639982 -0.5200697 -0.9610792
## 4       4 -0.52462814  0.4451409  1.8498439 -1.0404550 -1.1865838
## 5       5  0.08926902 -0.4618336 -0.3208615  0.3260892  0.5396003
##   Asset_Turnover   Leverage Rev_Growth Net_Profit_Margin
## 1   1.153164e+00 -0.4680782  0.4671788         0.5912425
## 2  -1.537552e-01 -0.4040831  0.6917224        -0.4005718
## 3  -1.153164e+00  1.4773718  0.7120120        -0.3688236
## 4   1.480297e-16 -0.3443544 -0.5769454        -1.6095439
## 5   6.589509e-02 -0.2559803 -0.7230135         0.7343816
```

```r
# Create a new data frame 'Pharma2' by adding cluster assignments to 'Pharma1'
Pharma2 <- data.frame(Pharma1, fit$cluster)
Pharma2
```

```
##       Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## ABT   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## AGN  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## AHM  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## AZN   0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## AVE  -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## BAY  -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
## BMY  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## CHTT -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## ELN  -0.9704532  2.15893320 -1.34037772 -0.70899938 -1.0174553     -1.8450624
## LLY   0.2762415 -1.34655112  0.14948233  0.34502953  0.5610770     -0.4612656
## GSK   1.0999201 -0.68440408 -0.45749769  2.45971647  1.8389364      1.3837968
## IVX  -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.6979905     -0.4612656
## JNJ   1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544      0.9225312
## MRX  -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -1.8450624
## MRK   1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577      1.8450624
## NVS   0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -0.9225312
## PFE   2.4199899  0.48409069 -0.11415545  1.31287998  1.6322239      0.4612656
## PHA  -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.9047030     -0.4612656
## SGP  -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929      0.4612656
## WPI  -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905     -0.9225312
## WYE  -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849     -0.4612656
##          Leverage  Rev_Growth Net_Profit_Margin fit.cluster
## ABT  -0.21209793 -0.52776752        0.06168225           5
## AGN   0.01828430 -0.38113909       -1.55366706           4
## AHM  -0.40408312 -0.57211809       -0.68503583           2
## AZN  -0.74965647  0.14744734        0.35122600           5
## AVE  -0.31449003  1.21638667       -0.42597037           2
## BAY  -0.74965647 -1.49714434       -1.99560225           4
## BMY  -0.02011273 -0.96584257        0.74744375           5
## CHTT  3.74279705 -0.63276071       -1.24888417           3
## ELN   0.61983791  1.88617085       -0.36501379           3
## LLY  -0.07130879 -0.64814764        1.17413980           5
## GSK  -0.31449003  0.76926048        0.82363947           1
## IVX   1.10620040  0.05603085       -0.71551412           3
## JNJ  -0.62166634 -0.36213170        0.33598685           1
## MRX   0.44065173  1.53860717        0.85411776           3
## MRK  -0.39128411  0.36014907       -0.24310064           1
## NVS  -0.67286239 -1.45369888        1.02174835           5
## PFE  -0.54487226  1.10143723        1.44844440           1
## PHA  -0.30169102  0.14744734       -1.27936246           4
## SGP  -0.74965647 -0.43544591        0.29026942           5
## WPI  -0.49367621  1.43089863       -0.09070919           2
## WYE   0.68383297 -1.17763919        1.49416183           5
```

```
# Create a cluster plot to visualize the clusters formed by k-means clustering
clusplot(Pharma1, fit$cluster, color = TRUE, shade = FALSE, labels = 2, lines = 0)
```



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

#Among the companies that comprise Cluster 1 are #JNJ, MRK, PFE, and GSK; these companies have the largest market capitalizations and use financing to run their operations efficiently. (lower than 0.47 leverage).

#Due to their lowest asset turnover and beta values, the stocks of Cluster 2 companies, AHM, WPI, and AVE, have the potential to outperform the current market benchmark.

#They are the least capitalized company on the market, have the fastest revenue growth in Cluster 3, and are unable to even raise capital to support their operations. (MRX, CHTT, LVX, ELN). These business stocks' strong returns can be attributed in part to their high beta values.

#Cluster 4: AGN, BAY, RHA Because of their highest expense to earnings ratio, they are the lowest earning. Additionally, their Return on Equity is less than 1, which suggests that it is unlikely that investing in these companies will yield the highest returns.

#The group is composed of #Cluster-5 ABT, SGP, NVS, AZN, BMY, and WYE. They have the lowest rate of sales development, the highest asset turnover, and the highest net profit margin. These businesses are prospering as a result of their growth.

c. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters)

#The stocks in Cluster-1 have a mediocre personality; they are neither strong nor have they recently produced noteworthy returns.

#The businesses in Cluster-2 are evenly distributed over the world. Despite their sound technical foundation, the media has largely embraced their concepts.

#Cluster 3-Despite having a high leverage ratio, they are only moderately advised due to the security of their finances.

#Shares in Group-4 The media claims that should be preserved because they will eventually turn into priceless assets.

#Cluster No. 5: It is advised that companies having a high net profit margin stay in the cluster for a long time.

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

#Cluster 1: A workable strategy (since these are reputable stocks).

#Cluster-2 is a collection of gold miners, despite their low beta, the market is very bullish on them.

#The original configuration, or #Cluster-3 (stocks with solid financial and other fundamentals).

#Cluster-4: The original setup (stocks with solid fundamentals, including financials).

#cluster 5 is the recurring cluster. Adding the stocks to the portfolio is highly recommended because a significant net profit margin indicates that the business is performing well.