Assignment-4

Saurabh Patloori

2023-10-29

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

```r
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 1.9.2     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(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
```

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

```r
Pharma_data <- na.omit(Pharmaceuticals)
#Provides the data after removing the incomplete cases.
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
```

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 numercial values i.e., 3-11 columns from csv file
head(Pharma_data1)
```

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

```
##      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
#To determine the number of clusters to do the cluster analysis using Elbow Method
fviz_nbclust(Pharma_data2, kmeans, method = "wss")
```

Optimal number of clusters
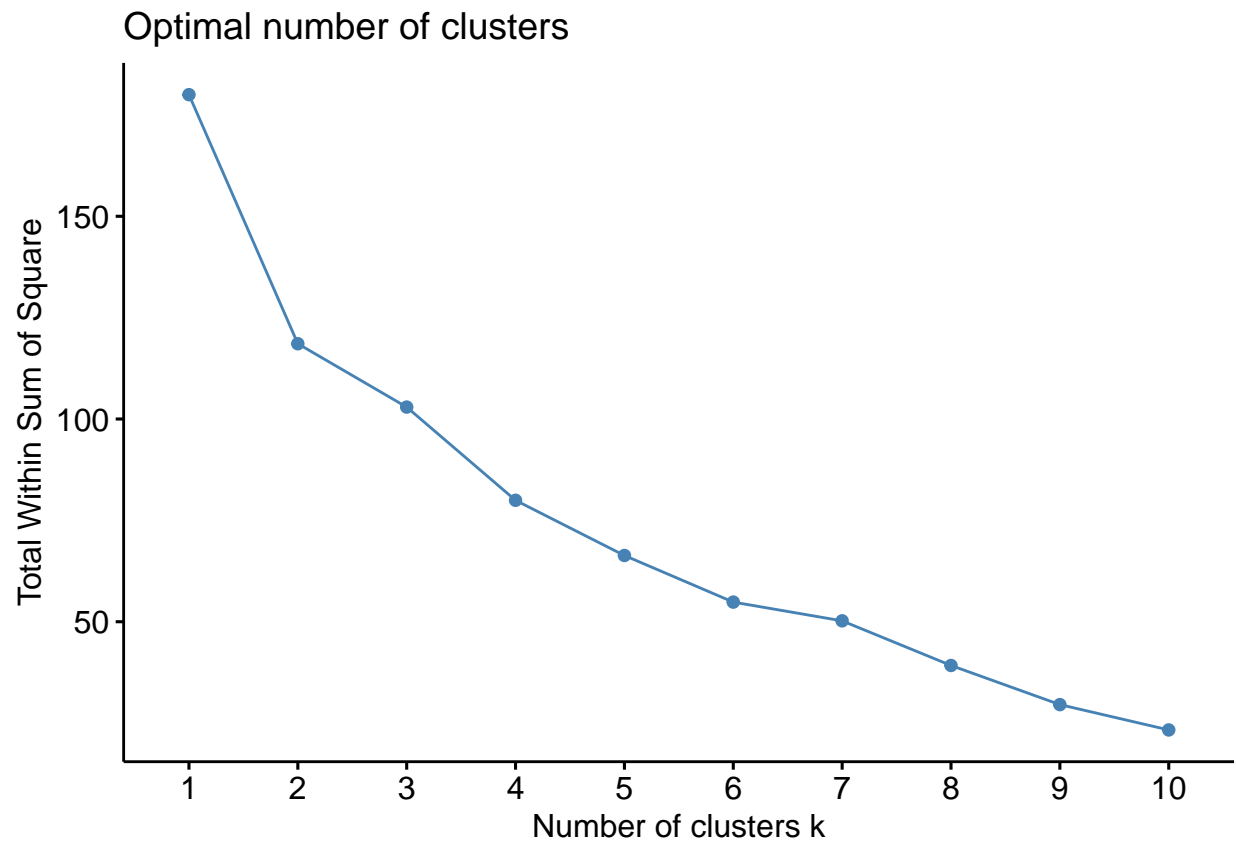

```
##By seeing the above graph from Elbow method, Graph is not
#clear to choose k=2 or 3 or 4 or 5.
#Silhouette method for determining no of clusters


fviz_nbclust(Pharma_data2, kmeans, method = "silhouette")
```

Optimal number of clusters
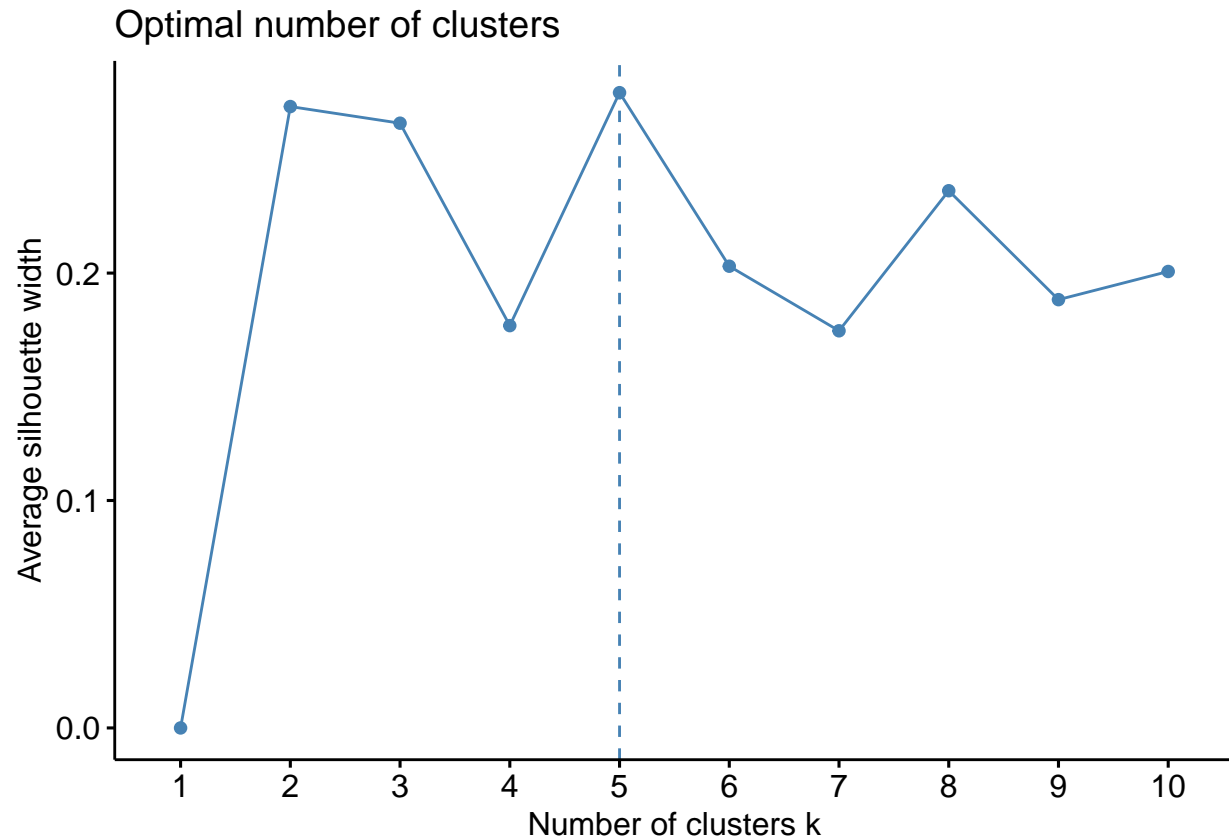

```r
#Applying K-means
set.seed(64060)
k_5<- kmeans(Pharma_data2,centers=5,nstart = 25)
```

```r
#Visualizing the output
#centroids
k_5$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
fviz_cluster(k_5,data = Pharma_data2) # to Visualize the clusters
```

Cluster plot
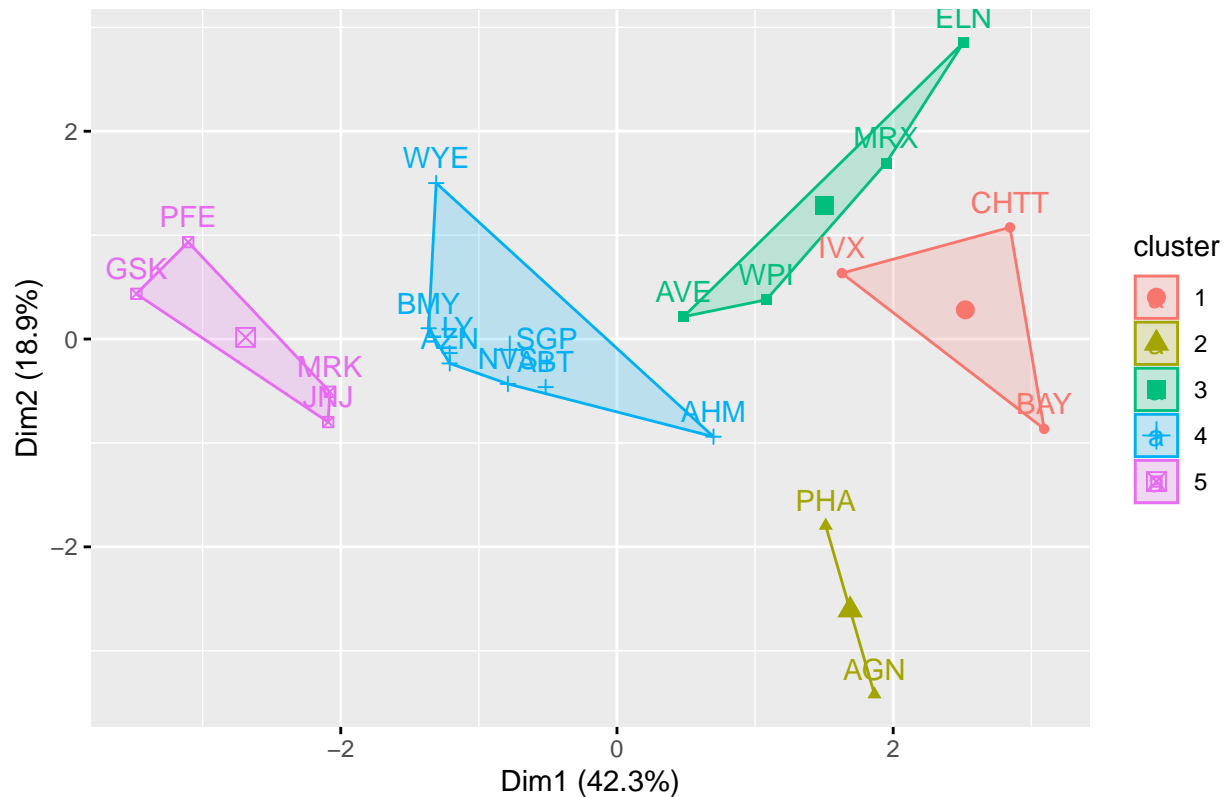

```
k_5
```

```
## K-means clustering with 5 clusters of sizes 3, 2, 4, 8, 4
##
## Cluster means:
##     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
##
## Clustering vector:
## ABT AGN AHM AZN AVE BAY BMY CHTT ELN LLY GSK IVX JNJ MRX MRK NVS
##   4   2   4   4   3   1   4    1   3   4   5   1   5   3   5   4
## PFE PHA SGP WPI WYE
##   5   2   4   3   4
##
## Within cluster sum of squares by cluster:
## [1] 15.595925  2.803505 12.791257 21.879320  9.284424
```

6

```
##   (between_SS / total_SS =  65.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
distance<- dist(Pharma_data2, method = "euclidean")
fviz_dist(distance)
```
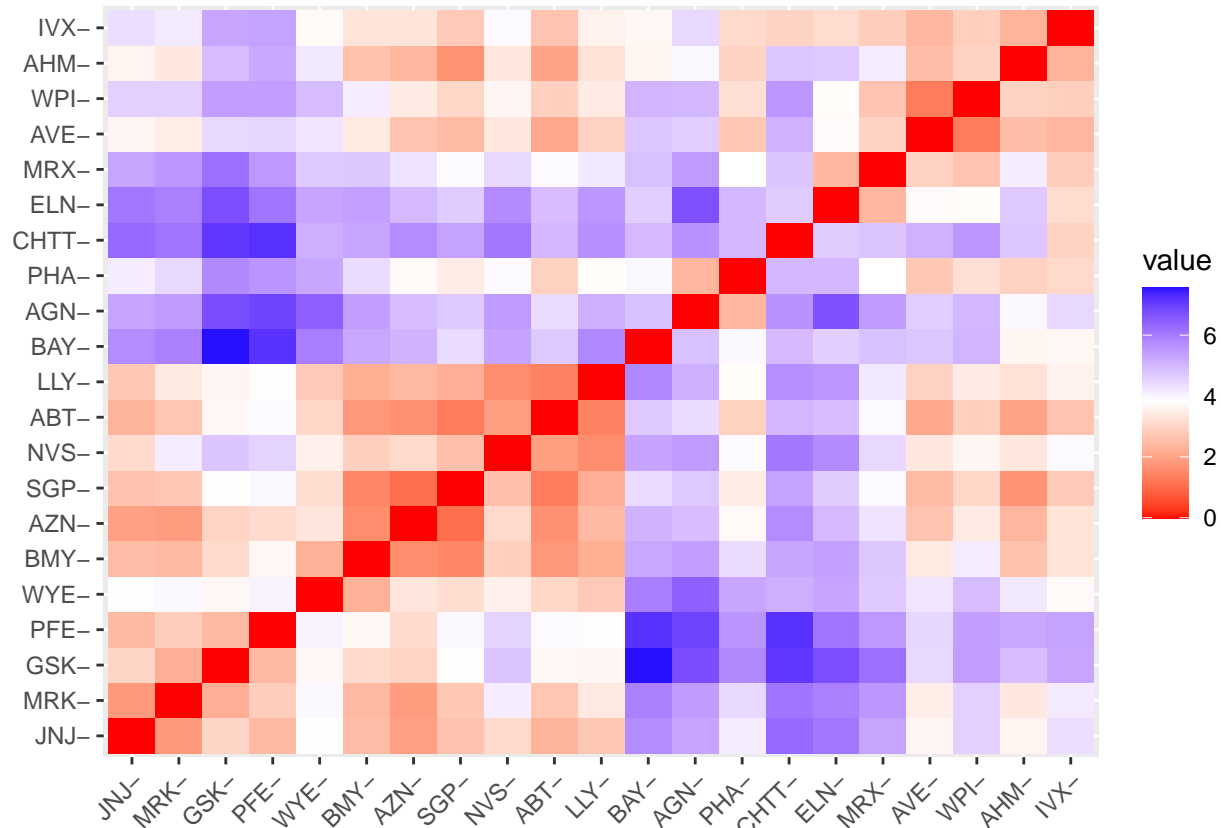


```
## I can see there are 5 clusters and the center is defined after 25 restarts
#which is determined in kmeans.
#K - Means Cluster Analysis- Fit the data with 5 clusters
fit<-kmeans(Pharma_data2,5)
```

```
#Finding the mean value of all quantitative variables for each cluster
aggregate(Pharma_data2,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
```

```
##To view the cluster plot

clusplot(Pharma_data2,fit$cluster,color =
          TRUE,shade = TRUE,labels = 2,lines = 0)
```


CLUSPLOT(Pharma_data2)
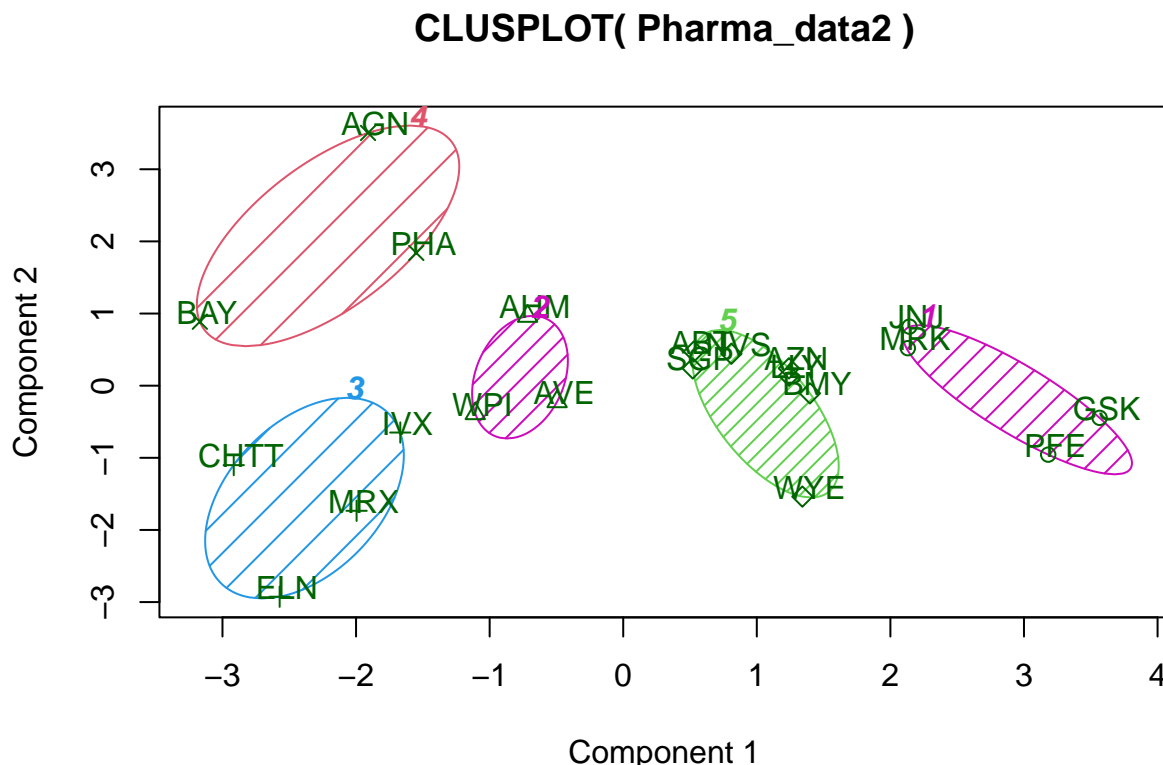
Component 1

These two components explain 61.23 % of the point variability.

#Task 2: Analyze the clusters in relation to the numerical variables used in their creation. AGN, PHA, and BAY are in #Cluster_1, and they have the highest PE_Ratio. #By looking at each cluster's mean values for every quantitative variable. The ROE percentage is low. The companies in #Cluster_2 with the largest market capitalization and good leverage values are #JNJ, MRK, GSK, and PFE. The three stocks in #Cluster_3 with the lowest beta and asset turnover are #AHM, AVE, and WPI. Task 3: Are there any patterns among the clusters concerning the numerical variables (10–12)? *(those not used in the formation of clusters) #Because cluster 1 has the highest PE Ratio, it should be held in compliance with media recommendations. #For cluster 2: It has the biggest market capitalization and a good leverage value. Besides, they can be pretty aggressive.#In cluster 3, the lowest asset turnover and beta are observed. Nonetheless, positive media endorsements predominate.#For cluster 4: Considering the high leverage ratio, they have a cautious recommendation. #Cluster 5: They have the lowest revenue growth, the highest asset turnover, and the biggest net profit margin.

#Task 4: Give each cluster a fitting name using some or all of the variables in the dataset. #Cluster 1: Hold cluster; they have respectable numbers. #4 Group 2: Light Buy or Hold a Cluster. #Cluster 3: Buying or Selling #Cluster 4: Buy Cluster—it is comparatively stable.