K Means

Nisha Chaurasia

2023-03-13

##Loading Required Packages

```r
rm(list = ls()) #cleaning the environment
library(readr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v dplyr   1.0.10
## v tibble  3.1.8      v stringr 1.5.0
## v tidyr   1.3.0      v forcats 0.5.2
## v purrr   1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(caret)
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```r
library(knitr)
library(class)
library(ggplot2)
library(ggcorrplot)
library(dplyr)
library(e1071)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
library(caret)
library(factoextra)
```

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

```
library(cluster)
library(cowplot)
library(pander)
library(kernlab)
```

```
##
## Attaching package: 'kernlab'
##
## The following object is masked from 'package:purrr':
##
##     cross
##
## The following object is masked from 'package:ggplot2':
##
##     alpha
```

```
library(tidyr)
```

##Import Data "Pharmaceuticals.csv"

```
pharma <- read.csv("C:/Users/Chaur/OneDrive/Desktop/FML/Assignment_4_Kmeans/Pharmaceuticals.csv")
head(pharma)
```

```
##   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
```

##Understand the bank data structure

```
str(pharma) #21 obs. of  14 variables:
```

```
## 'data.frame':    21 obs. of  14 variables:
##  $ Symbol               : chr  "ABT" "AGN" "AHM" "AZN" ...
##  $ Name                 : chr  "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZeneca PL
```

```
##  $ Market_Cap           : num  68.44 7.58 6.3 67.63 47.16 ...
##  $ Beta                 : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...
##  $ PE_Ratio             : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...
##  $ ROE                  : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...
##  $ ROA                  : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...
##  $ Asset_Turnover       : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
##  $ Leverage             : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...
##  $ Rev_Growth           : num  7.54 9.16 7.05 15 26.81 ...
##  $ Net_Profit_Margin    : num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...
##  $ Median_Recommendation: chr  "Moderate Buy" "Moderate Buy" "Strong Buy" "Moderate Sell" ...
##  $ Location             : chr  "US" "CANADA" "UK" "UK" ...
##  $ Exchange             : chr  "NYSE" "NYSE" "NYSE" "NYSE" ...
```

```
summary(pharma)
```

```
##     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  
##                    
##                    
##                    
```

```
colMeans(is.na(pharma)) #No Missing data
```

```
##                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
```

#1.Use only the numerical variables (1 to 9) to cluster the 21 firms.

```
pharma2 <- pharma[,c(1,3:11)]
row.names(pharma2) <- pharma2[,1]
pharma2 <- pharma2[,-1]
head(pharma2)
```

```
##     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
```

```
str(pharma2) #Dropped "Name", "Median_Recommendation", "Location", "Exchange"
```

```
## 'data.frame':    21 obs. of  9 variables:
##  $ Market_Cap       : num  68.44 7.58 6.3 67.63 47.16 ...
##  $ Beta             : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...
##  $ PE_Ratio         : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...
##  $ ROE              : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...
##  $ ROA              : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...
##  $ Asset_Turnover   : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
##  $ Leverage         : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...
##  $ Rev_Growth       : num  7.54 9.16 7.05 15 26.81 ...
##  $ Net_Profit_Margin: num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...
```

##Normalizing the data by using Scale function.

```
set.seed(72)
pharma_Norm <- scale(pharma2)#normalizing the data by subtracting the mean of the data and dividing by
pandoc.table(head(pharma_Norm),style="grid", split.tables = Inf)# top 6 Observation from pharma_Norm
```

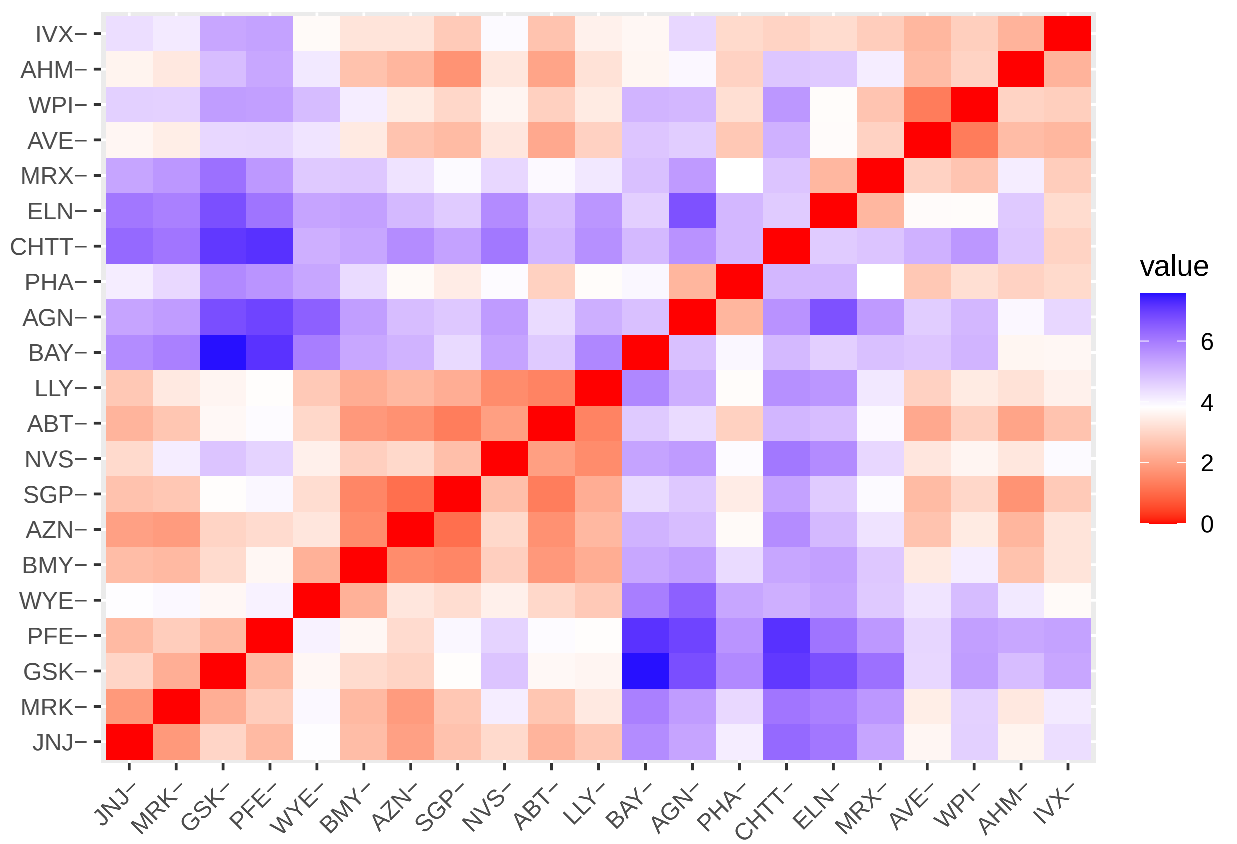
```
##
##
## +---------+------------+----------+----------+---------+---------+----------------+----------+-------
## |    | Market_Cap |   Beta   | PE_Ratio |   ROE   |   ROA   | Asset_Turnover | Leverage | Rev_G
## +=========+============+==========+==========+=========+=========+================+==========+=======
## | **ABT** |   0.1841   | -0.8013  | -0.04671 | 0.04009 | 0.2416  |       0        | -0.2121  |  -0.5
```

```
## +---------+-----------+----------+---------+---------+---------------+----------+------
## | **AGN** |  -0.8544   | -0.4507  |  3.497   | -0.8548 | -0.9423 |     0.9225     | 0.01828  |  -0.38
## +---------+-----------+----------+---------+---------+---------------+----------+------
## | **AHM** |  -0.8763   |  -0.256  |  -0.292  | -0.7223 | -0.5101 |     0.9225     | -0.4041  |  -0.57
## +---------+-----------+----------+---------+---------+---------------+----------+------
## | **AZN** |   0.1703   | -0.02226 | -0.2429  | 0.1064  | 0.9181  |     0.9225     | -0.7497  |   0.14
## +---------+-----------+----------+---------+---------+---------------+----------+------
## | **AVE** |   -0.179   | -0.8013  | -0.3287  | -0.2648 | -0.5664 |    -0.4613     | -0.3145  |   1.2
## +---------+-----------+----------+---------+---------+---------------+----------+------
## | **BAY** |  -0.6954   |  2.276   |  0.1495  | -1.451  | -1.713  |    -0.4613     | -0.7497  |   -1.4
## +---------+-----------+----------+---------+---------+---------------+----------+------
```

##Clustering the data by using euclidean distnace and plotting the graph
##Using Euclidean distance formula

$$distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2)}$$

```
pharma_distance <- get_dist(pharma_Norm) #By default uses Euclidean distance to compute the distances b
fviz_dist(pharma_distance, order = TRUE, show_labels = TRUE) #heatmap to visualize the distance
```




```
countries <- pharma[,c(1,2)]
unique(countries)
```

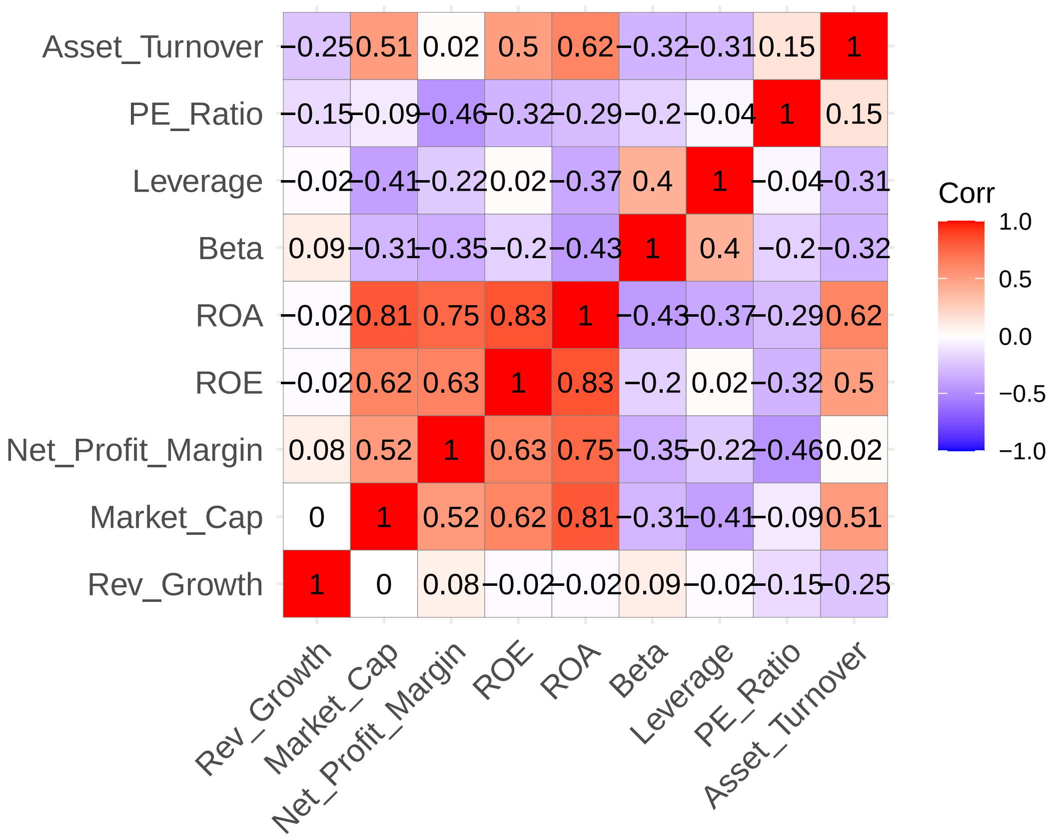
```
##    Symbol                              Name
## 1     ABT              Abbott Laboratories
```

```
## 2     AGN                      Allergan, Inc.
## 3     AHM                        Amersham plc
## 4     AZN                     AstraZeneca PLC
## 5     AVE                             Aventis
## 6     BAY                            Bayer AG
## 7     BMY         Bristol-Myers Squibb Company
## 8    CHTT                        Chattem, Inc
## 9     ELN              Elan Corporation, plc
## 10    LLY              Eli Lilly and Company
## 11    GSK                GlaxoSmithKline plc
## 12    IVX                   IVAX Corporation
## 13    JNJ                  Johnson & Johnson
## 14    MRX Medicis Pharmaceutical Corporation
## 15    MRK                  Merck & Co., Inc.
## 16    NVS                        Novartis AG
## 17    PFE                         Pfizer Inc
## 18    PHA               Pharmacia Corporation
## 19    SGP           Schering-Plough Corporation
## 20    WPI         Watson Pharmaceuticals, Inc.
## 21    WYE                              Wyeth
```

```
#The intensity of color changes as distances increases or decreases. Below heatmap represents the dista
```

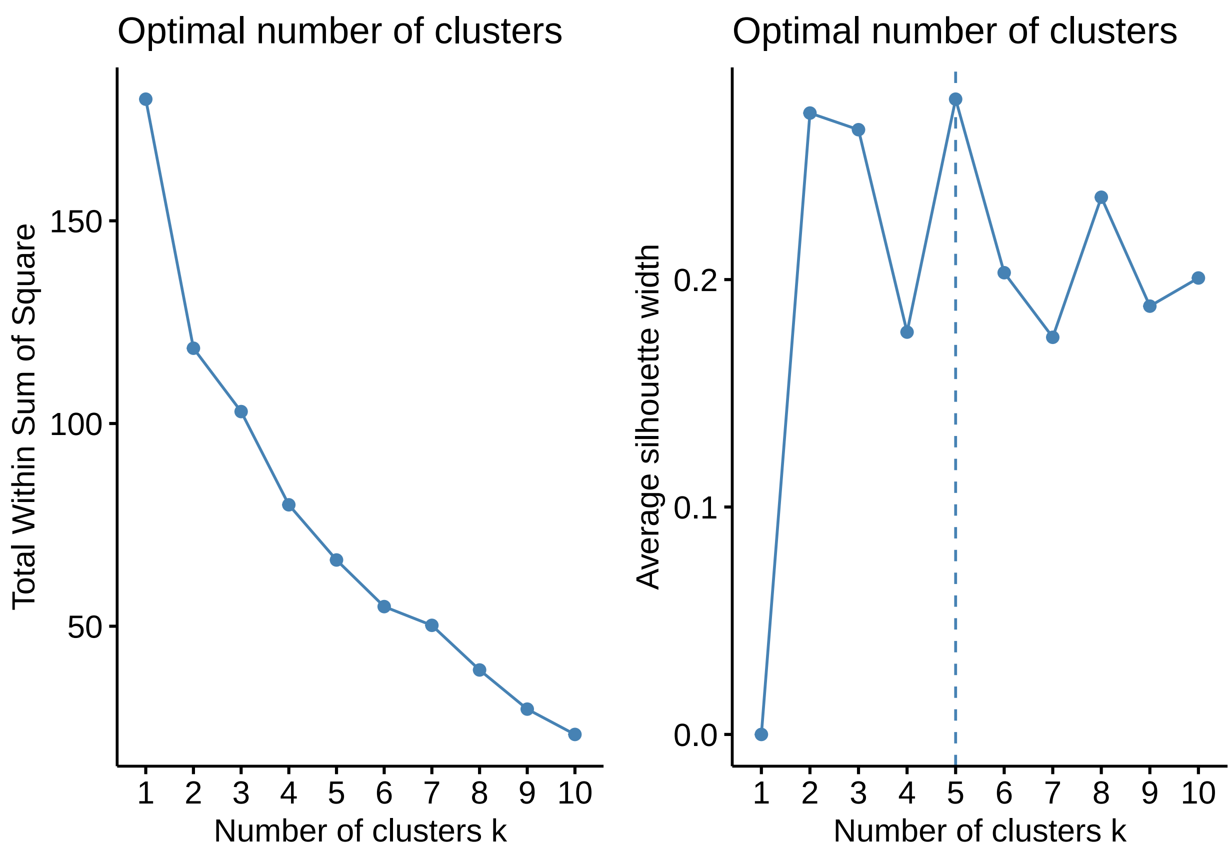
#To see if there is any correlation amoung the variables choosen for clustering

```
corr<-cor(pharma_Norm)
ggcorrplot(corr,outline.color = "grey50",lab = TRUE,hc.order = TRUE,type = "full") ##Return on Assets (
```



##Finding the number of cluster for grouping similar countries together. ##There are two main methods to find the value of K or number of cluster: Elbow chart and the Silhouette Method

```
Elbow_method <- fviz_nbclust(pharma_Norm, kmeans, method = "wss")
Silhouette <- fviz_nbclust(pharma_Norm, kmeans, method = "silhouette")
plot_grid(Elbow_method, Silhouette, nrow = 1) #The elbow method is giving value k = 6 however the silho
```

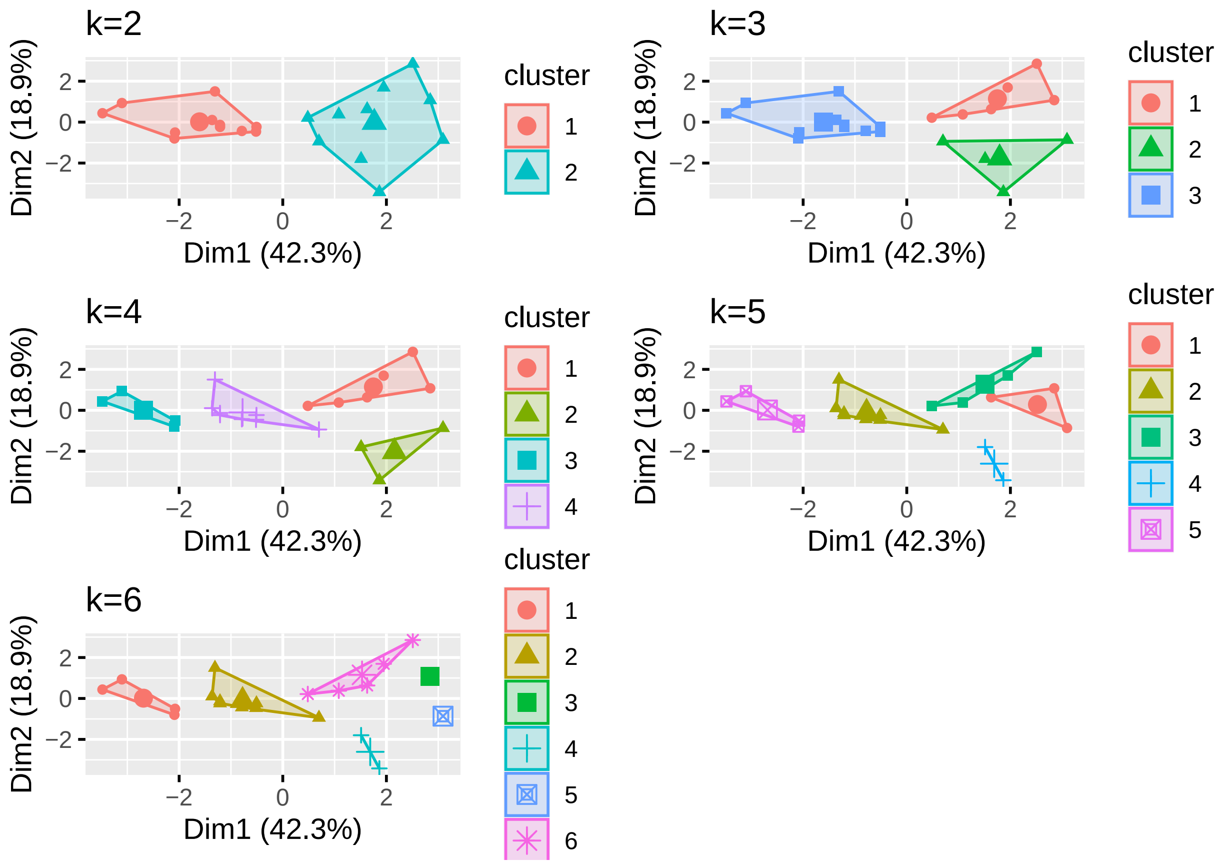


#Trying to find out the optimal value of k since elbow method is showing k = 2 or 6 and silhouette method is showing k= 5. will exploare all values from 2 to 6

```
k2<-kmeans(pharma_Norm,centers =2,nstart=25)
k3<-kmeans(pharma_Norm,centers =3,nstart=25)
k4<-kmeans(pharma_Norm,centers =4,nstart=25)
k5<-kmeans(pharma_Norm,centers =5,nstart=25)
k6<-kmeans(pharma_Norm,centers =6,nstart=25)
p1<-fviz_cluster(k2,geom = "point", data=pharma_Norm)+ggtitle("k=2")
p2<-fviz_cluster(k3,geom = "point", data=pharma_Norm)+ggtitle("k=3")
p3<-fviz_cluster(k4,geom = "point", data=pharma_Norm)+ggtitle("k=4")
p4<-fviz_cluster(k5,geom = "point", data=pharma_Norm)+ggtitle("k=5")
p5<-fviz_cluster(k6,geom = "point", data=pharma_Norm)+ggtitle("k=6")
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
grid.arrange(p1,p2,p3,p4,p5)#The value 5 has no overlap and also creating 5 different clusters
```



#Since value of K = 5 is making more sense will create 5 clusters for our analysis

```
pharma_Kmeans <- kmeans(pharma_Norm, centers = 5, nstart = 25)
pandoc.table(pharma_Kmeans$centers,style="grid", split.tables = Inf)
```

```
##
##
## +------------+---------+----------+---------+---------+----------------+----------+-------------+-----
## | Market_Cap |  Beta   | PE_Ratio |   ROE   |   ROA   | Asset_Turnover | Leverage | Rev_Growth  | Net_
## +============+=========+==========+=========+=========+================+==========+=============+=====
## |  -0.03142  | -0.4361 | -0.3172  |  0.195  | 0.4084  |     0.173      | -0.2745  |   -0.7042   |
## +------------+---------+----------+---------+---------+----------------+----------+-------------+-----
## |  -0.4393   | -0.4702 |   2.7    | -0.835  | -0.9235 |     0.2306     | -0.1417  |   -0.1168   |
## +------------+---------+----------+---------+---------+----------------+----------+-------------+-----
## |  -0.7602   | 0.2796  | -0.4774  | -0.7438 | -0.8107 |     -1.268     | 0.06308  |    1.518    |
## +------------+---------+----------+---------+---------+----------------+----------+-------------+-----
## |  -0.8705   |  1.341  | -0.05284 | -0.6184 | -1.193  |    -0.4613     |  1.366   |   -0.6913   |
## +------------+---------+----------+---------+---------+----------------+----------+-------------+-----
## |   1.696    | -0.1781 | -0.1985  |  1.235  |  1.35   |     1.153      | -0.4681  |   0.4672    |
## +------------+---------+----------+---------+---------+----------------+----------+-------------+-----
```

```
pharma_Kmeans$size #Size of the cluster
```

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

```
pharma_Kmeans$withinss
```

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

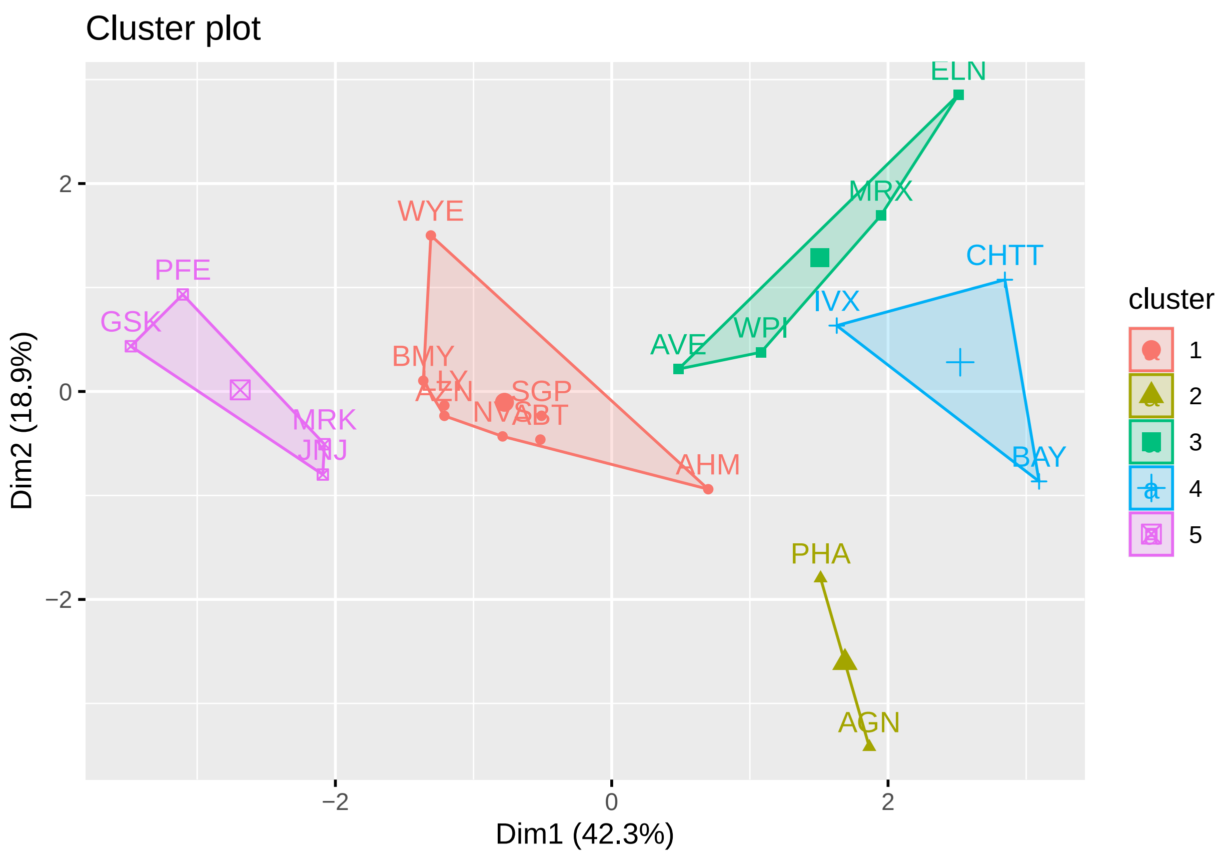
```
pharma_Kmeans$cluster[16]
```

```
## NVS
##   1
```

```
paste("Observation 16th is country NVS and belongs to cluster", pharma_Kmeans$cluster[16])
```

```
## [1] "Observation 16th is country NVS and belongs to cluster 1"
```

```
fviz_cluster(pharma_Kmeans, data = pharma_Norm)
```




```
#Understanding the results : the entire data is divided into 5 clusters. The "cluster 3" has most numbe
```

#Also using Kcca to get the clusters instead of Kmeans because K means uses the mean where as KCCA uses the KMedian

```
#using k-means with k=3 for making clusters
set.seed(180)
library(cluster)
library(flexclust)
```

```
## Loading required package: grid

## Loading required package: modeltools

## Loading required package: stats4

##
## Attaching package: 'modeltools'

## The following object is masked from 'package:kernlab':
##
##     prior

##
## Attaching package: 'flexclust'

## The following object is masked from 'package:kernlab':
##
##     kcca

## The following object is masked from 'package:e1071':
##
##     bclust
```

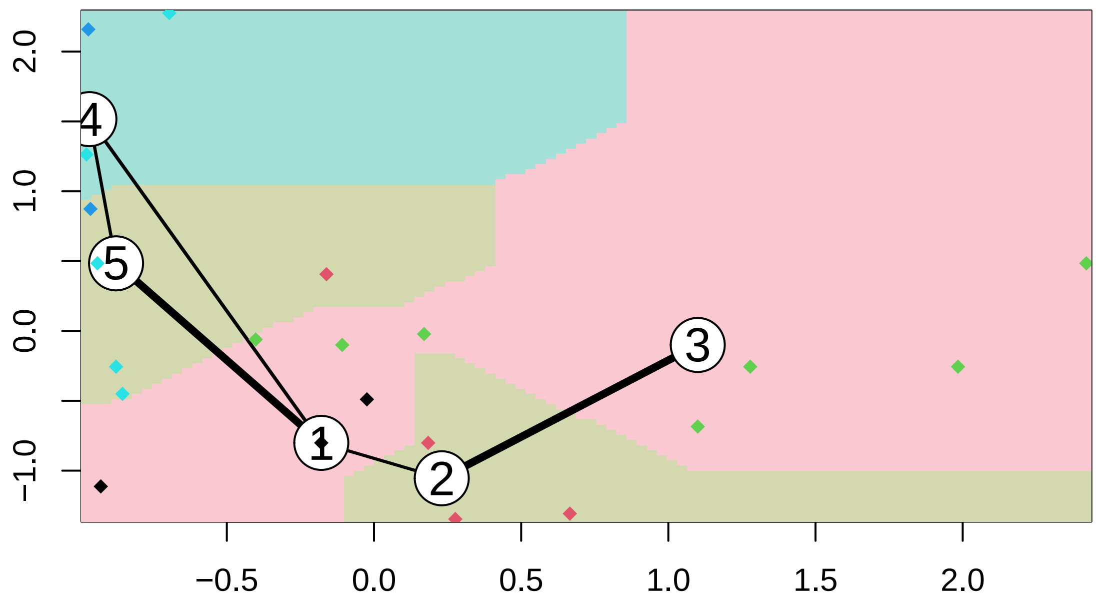
```
pharma_KCCA_3 <- kcca(pharma_Norm, k = 5, kccaFamily("kmedians"))
```

```
## Found more than one class "kcca" in cache; using the first, from namespace 'kernlab'

## Also defined by 'flexclust'

## Found more than one class "kcca" in cache; using the first, from namespace 'kernlab'

## Also defined by 'flexclust'
```

```
pharma_KCCA_3
```

```
## kcca object of family 'kmedians'
##
## call:
## kcca(x = pharma_Norm, k = 5, family = kccaFamily("kmedians"))
##
## cluster sizes:
##
## 1 2 3 4 5
## 3 4 7 2 5
```

```
clusters_index <- predict(pharma_KCCA_3)
image(pharma_KCCA_3)
points(pharma_Norm, col = clusters_index, pch = 18, cex = 1)
```




```
#KCCA and K-means clustering is the type of problem they are used to solve. KCCA is used for finding th
```

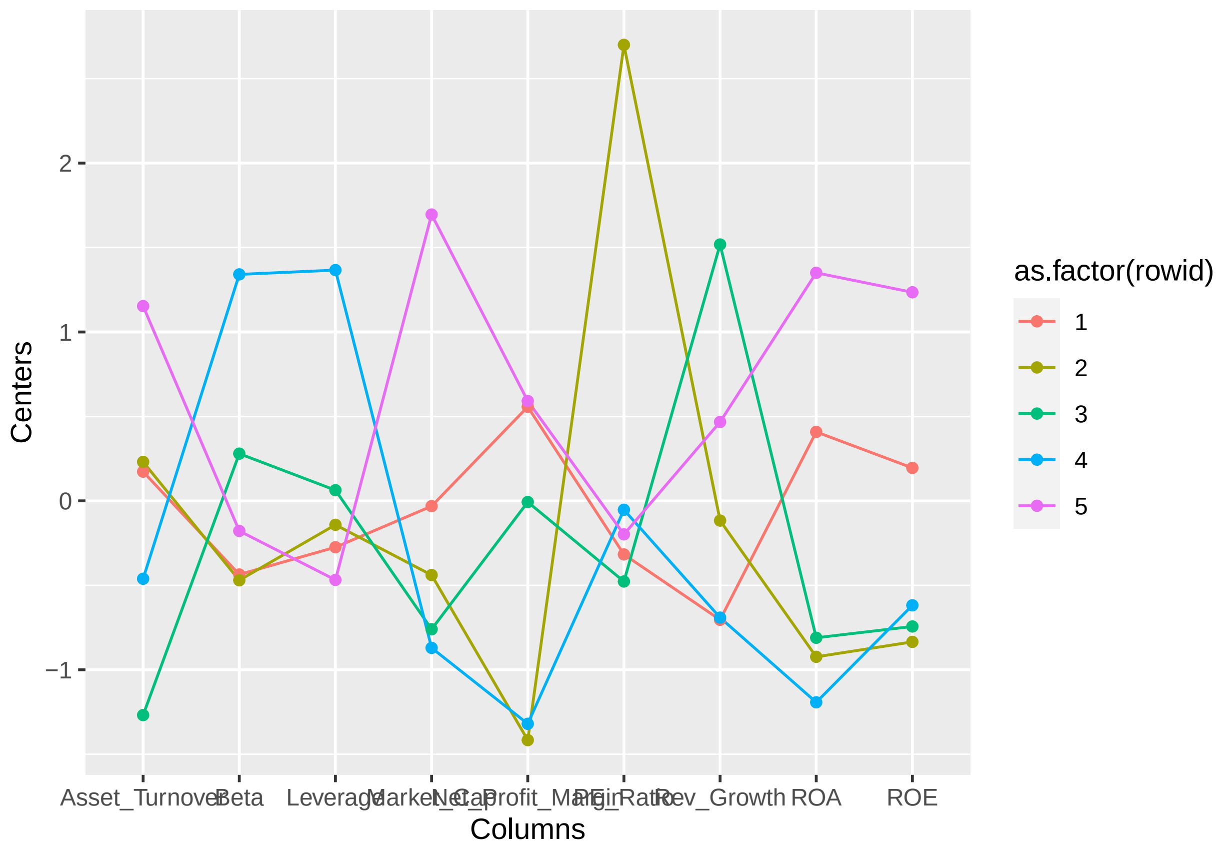
#Will Continue with cluster created by Kmeans since its more accurate for unsupervised learning method

```
#graphical plotting of data grouped in clusters
Centroid_1 <- data.frame(pharma_Kmeans$centers) %>% rowid_to_column() %>% gather('Columns', 'Centers', -
print(Centroid_1)
```

```
##    rowid         Columns      Centers
## 1      1      Market_Cap -0.031422109
## 2      2      Market_Cap -0.439251341
## 3      3      Market_Cap -0.760224892
## 4      4      Market_Cap -0.870515113
## 5      5      Market_Cap  1.695581115
## 6      1            Beta -0.436098941
## 7      2            Beta -0.470180039
## 8      3            Beta  0.279604106
## 9      4            Beta  1.340986857
## 10     5            Beta -0.178056346
## 11     1        PE_Ratio -0.317248516
```

```
## 12     2          PE_Ratio  2.700024643
## 13     3          PE_Ratio -0.477423799
## 14     4          PE_Ratio -0.052844340
## 15     5          PE_Ratio -0.198458234
## 16     1               ROE  0.195045857
## 17     2               ROE -0.834952524
## 18     3               ROE -0.743802224
## 19     4               ROE -0.618401510
## 20     5               ROE  1.234987906
## 21     1               ROA  0.408391543
## 22     2               ROA -0.923495091
## 23     3               ROA -0.810742783
## 24     4               ROA -1.192847826
## 25     5               ROA  1.350343113
## 26     1    Asset_Turnover  0.172974602
## 27     2    Asset_Turnover  0.230632802
## 28     3    Asset_Turnover -1.268480411
## 29     4    Asset_Turnover -0.461265604
## 30     5    Asset_Turnover  1.153164010
## 31     1          Leverage -0.274493115
## 32     2          Leverage -0.141703357
## 33     3          Leverage  0.063080849
## 34     4          Leverage  1.366446992
## 35     5          Leverage -0.468078185
## 36     1        Rev_Growth -0.704151557
## 37     2        Rev_Growth -0.116845875
## 38     3        Rev_Growth  1.518015830
## 39     4        Rev_Growth -0.691291399
## 40     5        Rev_Growth  0.467178770
## 41     1 Net_Profit_Margin  0.556954446
## 42     2 Net_Profit_Margin -1.416514761
## 43     3 Net_Profit_Margin -0.006893899
## 44     4 Net_Profit_Margin -1.320000179
## 45     5 Net_Profit_Margin  0.591242521
```

```
ggplot(Centroid_1, aes(x = Columns, y = Centers, color = as.factor(rowid))) + geom_line(aes(group = as.
```



##Question3 : Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters)

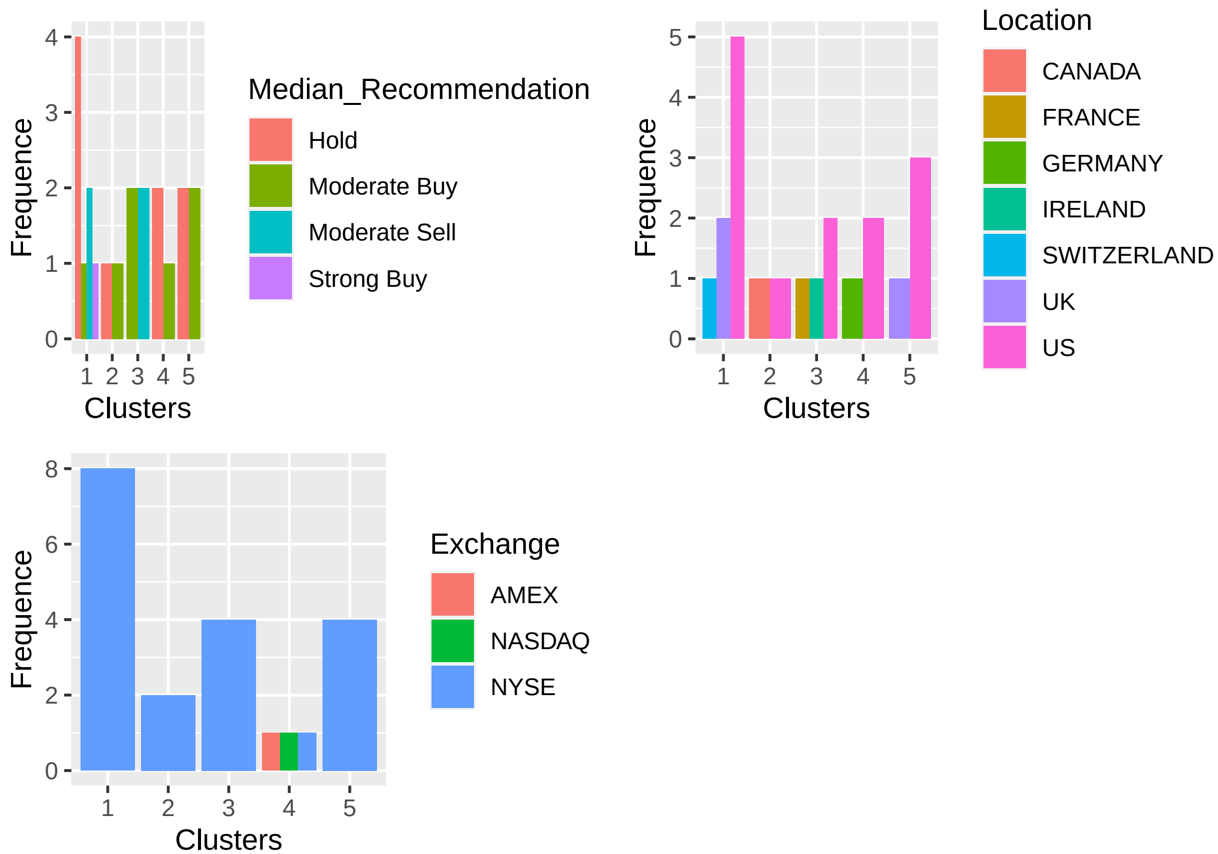
```
Pharma_Pattern <-  pharma %>% select(c(12,13,14)) %>% mutate(Cluster = pharma_Kmeans$cluster)
print(Pharma_Pattern) #The remaining three category to be considered are Stock Exchange, Location, and M
```

```
##    Median_Recommendation    Location Exchange Cluster
## 1           Moderate Buy          US     NYSE       1
## 2           Moderate Buy      CANADA     NYSE       2
## 3             Strong Buy          UK     NYSE       1
## 4          Moderate Sell          UK     NYSE       1
## 5           Moderate Buy      FRANCE     NYSE       3
## 6                   Hold     GERMANY     NYSE       4
## 7          Moderate Sell          US     NYSE       1
## 8           Moderate Buy          US   NASDAQ       4
## 9          Moderate Sell     IRELAND     NYSE       3
## 10                  Hold          US     NYSE       1
## 11                  Hold          UK     NYSE       5
## 12                  Hold          US     AMEX       4
## 13          Moderate Buy          US     NYSE       5
## 14          Moderate Buy          US     NYSE       3
## 15                  Hold          US     NYSE       5
## 16                  Hold SWITZERLAND     NYSE       1
## 17          Moderate Buy          US     NYSE       5
## 18                  Hold          US     NYSE       2
## 19                  Hold          US     NYSE       1
```

```
## 20          Moderate Sell           US      NYSE       3
## 21                   Hold           US      NYSE       1
```

#To visualize the distribution of businesses grouped by clusters and to identify any trends in the data, utilizing bar charts

```
Median_Recom <- ggplot(Pharma_Pattern, mapping = aes(factor(Cluster), fill=Median_Recommendation)) +
  geom_bar(position = 'dodge') + labs(x='Clusters', y='Frequence')

Location_0 <- ggplot(Pharma_Pattern, mapping = aes(factor(Cluster), fill=Location)) + geom_bar(position

Exchange_0 <- ggplot(Pharma_Pattern, mapping = aes(factor(Cluster), fill=Exchange)) +
geom_bar(position = 'dodge') + labs(x='Clusters', y='Frequence')
plot_grid(Median_Recom, Location_0, Exchange_0)
```




```
#The clustering analysis suggests that the companies in each cluster have similar characteristics in te

#Cluster -1 is dominated by American-based companies listed on the New York Stock Exchange, and they ha
#Cluster -2 has a mix of American and Canadian companies listed on the NYSE, and they have a moderate b
#Cluster -3 has companies from various locations listed on the NYSE, and they have a moderate buy or se
#Cluster -4 has companies from Germany and the USA listed on stock exchange markets other than NYSE (AM
#Cluster -5 has companies from the UK and USA, and they have a partially hold and buy recommendation fo
```

#4.Naming for each cluster using the variables in the dataset.

#Based on the entire analysis and looking at the characterstics of the clusters, 21 pharmaceutical industries can be categorized into 5 different groups:

#Cluster 1 - “Stable - efficient companies”: company with normal levels across financial metrics can be considered that the company is operating efficiently and effectively within its industry and competitive landscape. Also it is dominated by American-based companies listed on the New York Stock Exchange, and they have a spread advice to keep their stock, suggesting that they are stable and relatively low-risk investments

#Cluster 2 - “Overpriced - Risky companies”: since it has high price-to-earnings (PE) ratio and a low net profit margin means that the market is valuing the company’s stock at a premium compared to its current earnings, even though the company’s net profit margin is relatively low. which means investors are willing to pay a high price for each dollar of earnings the company generates, despite the fact that the company is not generating a high level of profit compared to its revenue.Such companies can be risky, as they may not be able to meet the market’s expectations and may experience a decline in stock price in the future.

#Cluster 3 - “Growth oriented - Low risky companies”: A company with low asset turnover and high revenue growth may indicate that the company has significant growth potential but is not yet operating at optimal efficiency. Investors should consider the company’s industry and competitive landscape, as well as its ability to sustain high revenue growth over the long term. It’s also important to evaluate the company’s profitability, as high revenue growth may not necessarily lead to higher profits if the company is not utilizing its assets efficiently.Also,these are the companies from various locations listed on the NYSE, and they have a moderate buy or sell recommendation, suggesting that they may have some growth potential

#Cluster 4 - “Debt-ridden - very risky companies”: Companies with high leverage and low net profit margin & ROA may indicate that the company is taking on a significant amount of debt to finance its operations, while not generating a sufficient level of profitability or returns on assets. This can be a concerning signal for investors, as the company may struggle to meet its debt obligations and may experience financial distress in the long term.Also,listed on stock exchange markets other than NYSE (AMEX and NASDAQ), and they have a hold or moderate buy recommendation.

#Cluster 5 - “Established - profitable companies”: Companies with high market capitalization are typically large and well-established companies that have a significant market presence and a strong financial position. High market capitalization means that the company has a large number of outstanding shares and a high stock price, resulting in a high valuation.Also,they have a partially hold and buy recommendation for their stocks listed on the NYSE.