FML Assignment 4

Pavan Kumar Mekala

2024-03-13

Summary

Questions - Answers

a. Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made in conducting the clust

In the process of clustering 21 corporations, it is imperative to incorporate all numerical variables, numbered from 1 to 9. These variables are integral as they reflect various financial dimensions such as profitability, market valuation, price-to-earnings ratio, return on equity, return on assets, and leverage, which collectively influence a firm's equity. Each variable was assigned an equal weight, signifying their uniform impact on the firm's financial standing. Market Capitalization: Reflects the overall size and market valuation of a company. Beta: Measures the volatility of a company's returns relative to market fluctuations. PE Ratio: Represents the relationship between a company's stock value and its earnings. ROE: Demonstrates a company's proficiency in generating profits from shareholder equity. ROA: Assesses an organization's ability to generate profits from its assets. Asset Turnover: Evaluates the effectiveness of a company in utilizing its assets to generate revenue. Leverage: Indicates the degree to which a company is financed through debt. Rev_Growth: Displays the rate of revenue growth over a specific period. Net Profit Margin: Reveals the proportion of revenue that translates into net income. I have given the Kmeans Algorithm some thought in order to cluster the dataset. And I used the optimal value of 5 from the silhouette technique to determine the number of clusters for the Kmeans clustering, and I clustered using the number of clusters of 2, as indicated by the Elbow method. However, since the points are closer to the centroids, the clusters created when the number of points is five are superior. The clusters identified through K-means are as follows: The first cluster, comprising four firms: AVE, WPI, MRX, ELN. The second cluster, with three firms: IVX, CHTT, BAY. The third cluster, including two firms: PHA, AGN. The fourth cluster, containing four firms: GSK, PFE, MRK, JNJ. The fifth cluster, the largest, encompassing eight firms: WYE, BMY, LLY, AZN, NVS, ABT, SGP, AHM. This clustering was based on the silhouette method's suggestion of five clusters, which proved more cohesive than the two clusters indicated by the elbow method.

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

The clusters, when examined in light of the numerical variables utilized for their formation, reveal distinct financial profiles: Cluster 1: Comprising AVE, WPI, MRX, ELN, this cluster is characterized by robust revenue growth and a high beta coefficient, suggesting a strong growth trajectory but lower asset efficiency and profitability. These firms may be in their nascent stages, likely investing significantly in expansion efforts. Their high beta and revenue growth imply potential for rapid earnings improvement. Cluster 2: Encompassing IVX, CHTT, BAY, these companies boast substantial market capitalization and solid returns on equity and assets, coupled with high asset turnover. Their low beta and profit-to-return ratios suggest established, stable operations with less efficient profit generation, highlighting their maturity and stability. Cluster 3: Featuring PHA, AGN, this cluster is marked by elevated price-to-earnings ratios and asset turnover, indicative of expected earnings acceleration despite historically low profitability. The combination of high valuation and low net profit margins points to a higher risk profile for investors. Cluster 4: Consisting of GSK, PFE, MRK, JNJ, this cluster stands out with the highest net profit margins and asset efficiency, demonstrating strong financial performance and low risk. The low beta and revenue growth indicate stable stock prices and modest revenue expansion, typical of mature, well-established entities. Cluster 5: Including WYE, BMY, LLY, AZN, NVS, ABT, SGP, AHM, this cluster is distinguished by a high beta and leverage, signaling higher investment risk due to stock price volatility and significant debt levels. However, these firms may offer higher returns in favorable market conditions.

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

Regarding the patterns with numerical variables 10 to 12, which were not used informing the clusters, the following observations can be made: Cluster 1: Generally recommended as a moderate buy or sell, these companies are based in France, Ireland, and the US, and are listed on the NYSE. Cluster 2: With a hold or moderate buy recommendation, these companies span Germany and the US and are listed across AMEX, NASDAQ, and NYSE. Cluster 3: Recommended as hold and moderate buy, these firms are located in the US and Canada, also listed on the NYSE. Cluster 4: Recommended as a hold and moderate buy, these UK and US-based companies are traded on the NYSE. Cluster 5: Carrying diverse recommendations from hold to strong buy, these firms are from Switzerland, the UK, and the US, with listings on the NYSE.

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

Appropriate name for each clusters are: Cluster1:High Growth Potential, Cluster2:High Risk High Beta, Cluster3: High Risk High Reward, Cluster4: Stability and Profitability, Cluster5: Low Risk High Profitability

```r
## Loading all the necessary Packages
library(ISLR)
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.3.3

## Loading required package: ggplot2

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

```r
library(caret)
```

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

```r
library(cluster)
library(class)
```

```
## Warning: package 'class' was built under R version 4.3.3
```

```r
library(e1071)
library(dplyr)
```

```
##
## 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
```

```r
library(klustR)
```

```
## Warning: package 'klustR' was built under R version 4.3.3
```

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

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'forcats' was built under R version 4.3.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v forcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.1
## v readr     2.1.5
```

```
## -- 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
```

```r
library(dbscan)
```

```
## Warning: package 'dbscan' was built under R version 4.3.3
```

```
##
## Attaching package: 'dbscan'
##
## The following object is masked from 'package:stats':
##
##     as.dendrogram
```

```r
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.3.3
```

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

Importing the data to R and checing by using dim() and print() functions

```r
Pharmaceuticals <- read.csv("Pharmaceuticals.csv")
dim(Pharmaceuticals)
```

```
## [1] 21 14
```

```r
print(Pharmaceuticals)
```

```
##    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
```

Creating a transpose of the dataframe

```
t(t(names(Pharmaceuticals)))
```

```
##       [,1]
##  [1,] "Symbol"
##  [2,] "Name"
##  [3,] "Market_Cap"
##  [4,] "Beta"
##  [5,] "PE_Ratio"
##  [6,] "ROE"
##  [7,] "ROA"
##  [8,] "Asset_Turnover"
##  [9,] "Leverage"
## [10,] "Rev_Growth"
## [11,] "Net_Profit_Margin"
## [12,] "Median_Recommendation"
## [13,] "Location"
## [14,] "Exchange"
```

Removing unwanted columns from the list and checked it by using dim() and summary() function

```
row.names(Pharmaceuticals) <- Pharmaceuticals[,1]
C_Data <- Pharmaceuticals[,3:11]
dim(C_Data)
```

```
## [1] 21  9
```

```r
summary(C_Data)
```

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

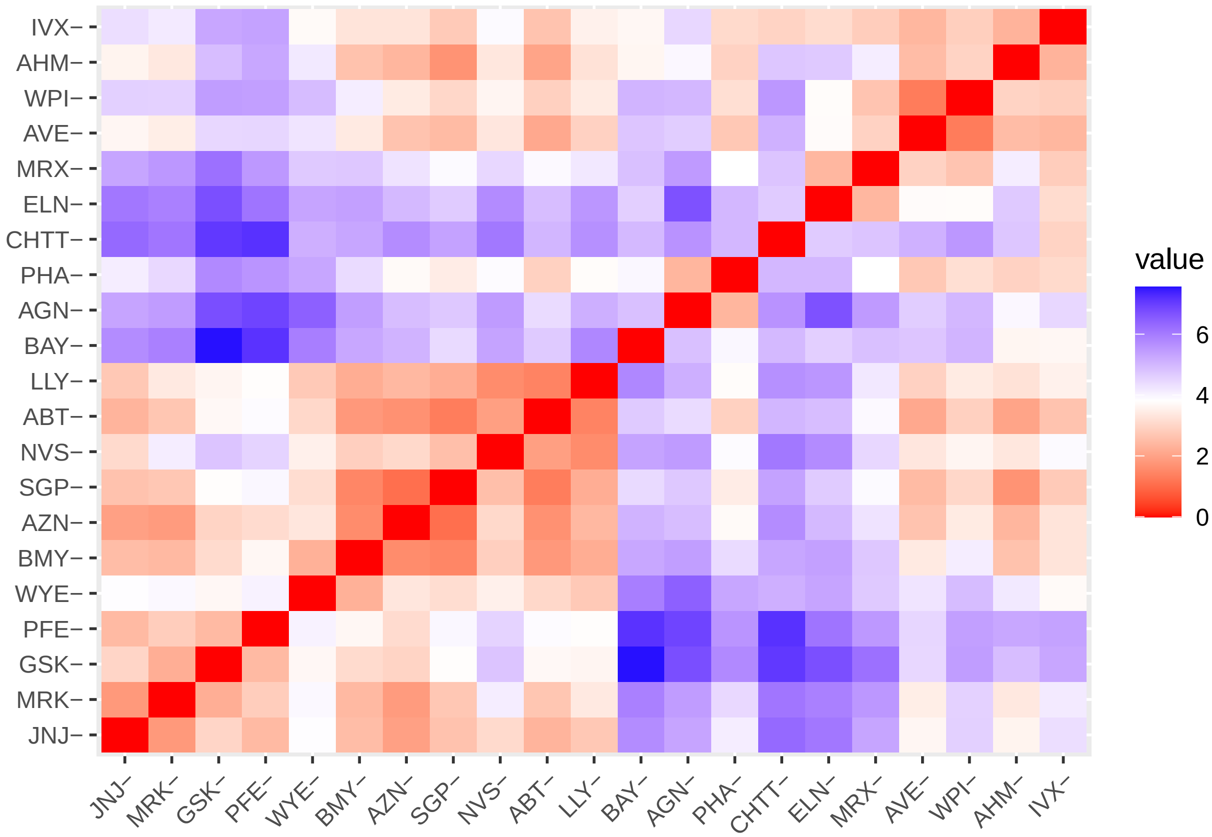
Scaling the data by using scale() function and checked it by using head() function

```r
Scaling_data <- scale(C_Data)
head(Scaling_data)
```

```
##     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
```

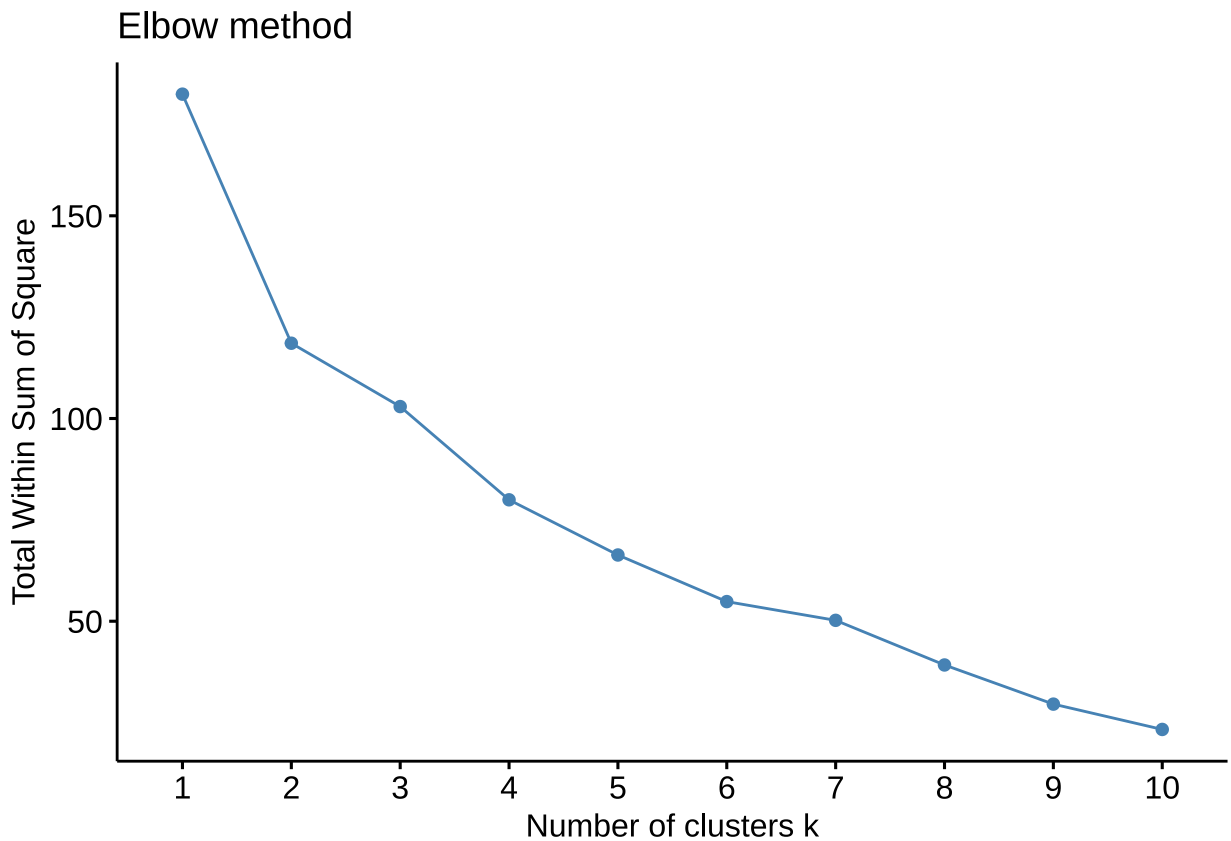
Measuring and visualizing the distance between each variable

```r
Dist_data <- get_dist(Scaling_data)
Visualize_data <- fviz_dist(Dist_data)
Visualize_data
```

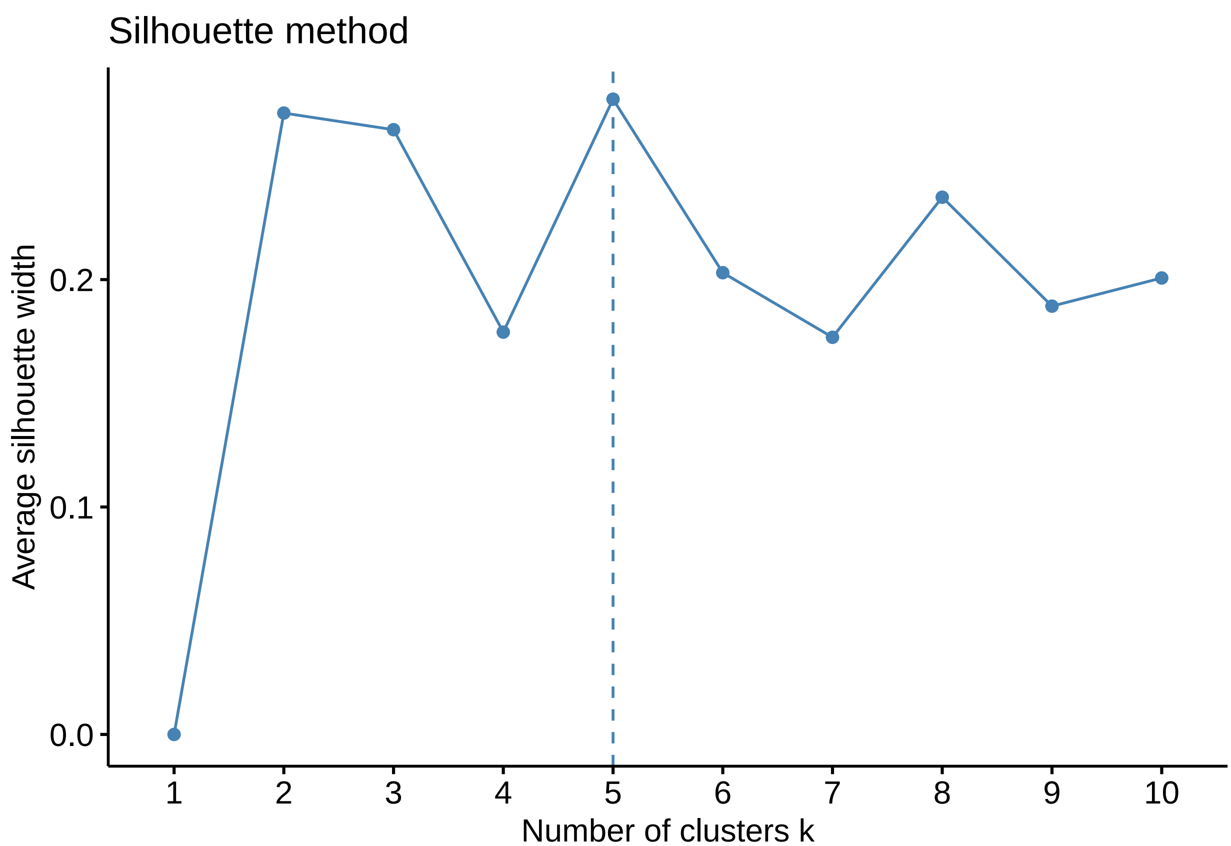


a.Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made in conducting the cluste

```
# sum of squares method
fviz_nbclust(Scaling_data, kmeans, method = "wss") + ggtitle("Elbow method")
```




```
# silhouette method
fviz_nbclust(Scaling_data, kmeans, method = "silhouette") + ggtitle("Silhouette method")
```



We must take into account the value of 2 since the curve was bent (an elbow) at point 2 according to the plot of the WSS (sum of squares) or Elbow technique. However, because of the graphical representation's lack of crispness, it is still ambiguous.

```r
# Considering the value of k=2
k <- 2
set.seed(333)
# kmeans algorithm
k_wss <- kmeans(Scaling_data, centers = k, nstart = 21)
k_wss
```

```
## K-means clustering with 2 clusters of sizes 11, 10
##
## Cluster means:
##    Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
## 2 -0.7407208  0.3945061  0.3039863 -0.7222576 -0.9178575     -0.5073922
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3331068 -0.2902163         0.6823310
## 2  0.3664175  0.3192379        -0.7505641
##
## Clustering vector:
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK  NVS
##    1    2    2    1    2    2    1    2    2    1    1    2    1    2    1    1
##  PFE  PHA  SGP  WPI  WYE
##    1    2    1    2    1
```

```
##
## Within cluster sum of squares by cluster:
## [1] 43.30886 75.26049
##  (between_SS / total_SS =  34.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
# For getting the centroids of the clusters
cat("The Centers of the clustes are", "\n")
```

```
## The Centers of the clustes are
```

```
k_wss$centers
```

```
##   Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
## 2 -0.7407208  0.3945061  0.3039863 -0.7222576 -0.9178575     -0.5073922
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3331068 -0.2902163         0.6823310
## 2  0.3664175  0.3192379        -0.7505641
```

```
# Getting the size of each cluster
cat("The size of each cluster is", "\n")
```

```
## The size of each cluster is
```

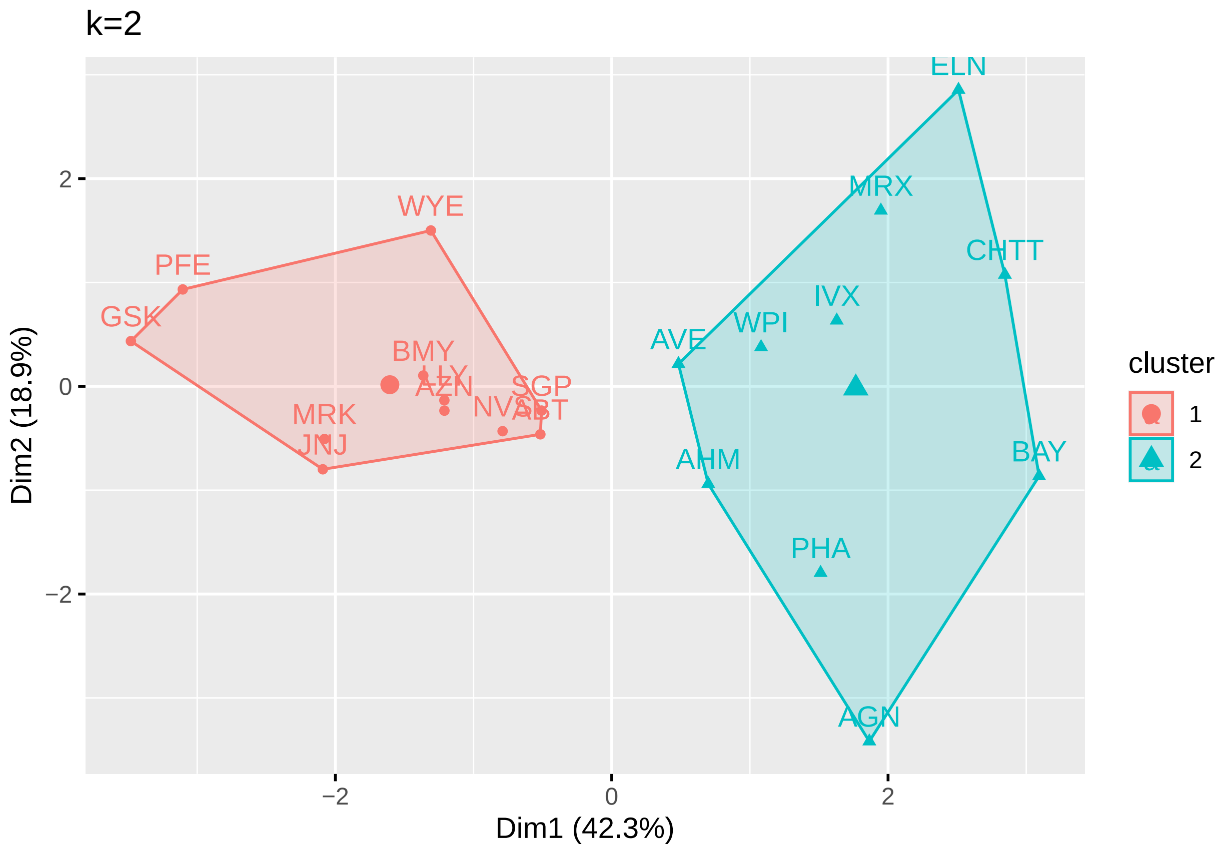
```
k_wss$size
```

```
## [1] 11 10
```

```
# Getting the points that are related to their corresponded cluster
k_wss$cluster
```

```
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK  NVS
##    1    2    2    1    2    2    1    2    2    1    1    2    1    2    1    1
##  PFE  PHA  SGP  WPI  WYE
##    1    2    1    2    1
```

```
# Visualization of clusters
fviz_cluster(k_wss, data = Scaling_data) + ggtitle("k=2")
```



From this kmeans clustering's result, which has a k value of 2. By considering all of the numerical variables—which are financial indicators that must be taken into account in order to determine equity because equity is dependent upon market capital—we can see that 11 companies fall into one cluster and the remaining 10 companies fall into another cluster. asset turnover, return on assets, net profit, etc. Furthermore, we can observe from the clusters that a few of the locations, such as AGN, ELN, GSK, etc., are distant from the centroids, indicating that not enough clusters were gathered. The maximum average silhouette width is at point 5, as shown by the silhouette method plot, thus we must take the valus of K as 5 into consideration.

```
k <- 5
set.seed(333)
# kmeans alogrithm
k_pav <- kmeans(Scaling_data, centers = k, nstart = 20)
k_pav
```

```
## K-means clustering with 5 clusters of sizes 4, 3, 2, 4, 8
##
## Cluster means:
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 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 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
##      Leverage Rev_Growth Net_Profit_Margin
## 1  0.06308085  1.5180158      -0.006893899
## 2  1.36644699 -0.6912914      -1.320000179
```

```
## 3 -0.14170336 -0.1168459      -1.416514761
## 4 -0.46807818  0.4671788       0.591242521
## 5 -0.27449312 -0.7041516       0.556954446
##
## Clustering vector:
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK  NVS
##    5    3    5    5    1    2    5    2    1    5    4    2    4    1    4    5
##  PFE  PHA  SGP  WPI  WYE
##    4    3    5    1    5
##
## Within cluster sum of squares by cluster:
## [1] 12.791257 15.595925  2.803505  9.284424 21.879320
##  (between_SS / total_SS =  65.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
## For getting the centroids of the clusters
cat("The Centers of the clustes are", "\n")
```

```
## The Centers of the clustes are
```

```
k_pav$centers
```

```
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 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 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
##      Leverage Rev_Growth Net_Profit_Margin
## 1  0.06308085  1.5180158      -0.006893899
## 2  1.36644699 -0.6912914      -1.320000179
## 3 -0.14170336 -0.1168459      -1.416514761
## 4 -0.46807818  0.4671788       0.591242521
## 5 -0.27449312 -0.7041516       0.556954446
```

```
# Getting the size of each cluster
cat("The size of each cluster is", "\n")
```

```
## The size of each cluster is
```

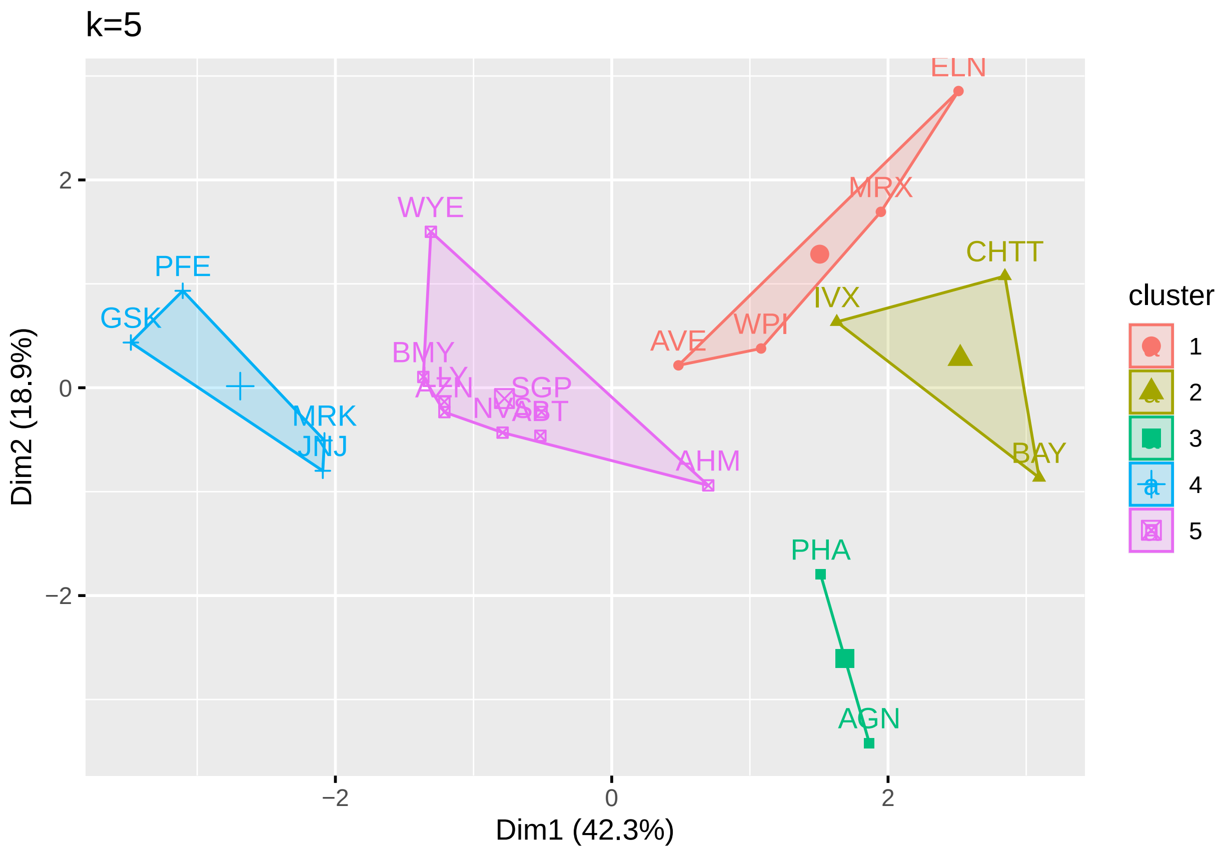
```
k_pav$size
```

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

```
# Getting the points that are related to their corresponded cluster
k_pav$cluster
```

```
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK  NVS
##    5    3    5    5    1    2    5    2    1    5    4    2    4    1    4    5
##  PFE  PHA  SGP  WPI  WYE
##    4    3    5    1    5
```

```r
# Visualization of clusters
fviz_cluster(k_pav, data = Scaling_data) + ggtitle("k=5")
```



From this kmeans clustering's result, which has a k value of 5. As we can see, there are 4 companies in the first cluster, 3 in the second, 2 in the third, and 4 in the fourth cluster, with the remainder falling under the fifth cluster. All numerical variables are taken into account because they are the financial measures that must be taken into account in order to determine equity, since equity is dependent on factors such as market capital, net profit, return on assets, asset turnover, etc. And this shows that the spots are significantly closer to the centroids.Perhaps this cluster is the greatest as well.

B.Interpret the clusters with respect to the numerical variables used informing the clusters.

```r
# Creating table by using clusters
Select_Columns <- Pharmaceuticals[,c(2:11)]
clustering_data <- Select_Columns %>%
  mutate(clustering=k_pav$cluster) %>% arrange(clustering, ascending = TRUE)
clustering_data
```

```
##                                     Name Market_Cap Beta PE_Ratio  ROE  ROA
## AVE                               Aventis      47.16 0.32     20.1 21.8  7.5
## ELN               Elan Corporation, plc       0.78 1.08      3.6 15.1  5.1
## MRX  Medicis Pharmaceutical Corporation       1.20 0.75     28.6 11.2  5.4
## WPI        Watson Pharmaceuticals, Inc.       3.26 0.24     18.4 10.2  6.8
## BAY                              Bayer AG      16.90 1.11     27.9  3.9  1.4
## CHTT                         Chattem, Inc       0.41 0.85     26.0 24.1  4.3
## IVX                       IVAX Corporation       2.60 0.65     19.9 21.4  6.8
## AGN                        Allergan, Inc.       7.58 0.41     82.5 12.9  5.5
## PHA               Pharmacia Corporation      56.24 0.40     56.5 13.5  5.7
## GSK                 GlaxoSmithKline plc     122.11 0.35     18.0 62.9 20.3
## JNJ                   Johnson & Johnson     173.93 0.46     28.4 28.6 16.3
## MRK                   Merck & Co., Inc.     132.56 0.46     18.9 40.6 15.0
## PFE                          Pfizer Inc     199.47 0.65     23.6 45.6 19.2
## ABT                 Abbott Laboratories      68.44 0.32     24.7 26.4 11.8
## AHM                        Amersham plc       6.30 0.46     20.7 14.9  7.8
## AZN                     AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4
## BMY        Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8 15.1
## LLY                 Eli Lilly and Company      73.84 0.18     27.9 31.0 13.5
## NVS                         Novartis AG      96.65 0.19     21.6 17.9 11.2
## SGP         Schering-Plough Corporation      34.10 0.51     18.9 22.6 13.3
## WYE                               Wyeth      48.19 0.63     13.1 54.9 13.4
##      Asset_Turnover Leverage Rev_Growth Net_Profit_Margin clustering
## AVE             0.6     0.34      26.81              12.9          1
## ELN             0.3     1.07      34.21              13.3          1
## MRX             0.3     0.93      30.37              21.3          1
## WPI             0.5     0.20      29.18              15.1          1
## BAY             0.6     0.00      -3.17               2.6          2
## CHTT            0.6     3.51       6.38               7.5          2
## IVX             0.6     1.45      13.99              11.0          2
## AGN             0.9     0.60       9.16               5.5          3
## PHA             0.6     0.35      15.00               7.3          3
## GSK             1.0     0.34      21.87              21.1          4
## JNJ             0.9     0.10       9.37              17.9          4
## MRK             1.1     0.28      17.35              14.1          4
## PFE             0.8     0.16      25.54              25.2          4
## ABT             0.7     0.42       7.54              16.1          5
## AHM             0.9     0.27       7.05              11.2          5
## AZN             0.9     0.00      15.00              18.0          5
## BMY             0.9     0.57       2.70              20.6          5
## LLY             0.6     0.53       6.21              23.4          5
## NVS             0.5     0.06      -2.69              22.4          5
## SGP             0.8     0.00       8.56              17.6          5
## WYE             0.6     1.12       0.36              25.5          5
```

```r
cat("The list of firms with correspoding to clusters are")
```

```
## The list of firms with correspoding to clusters are
```

```r
clustering_data[,c(1,11)]
```

```
##                                     Name clustering
## AVE                               Aventis          1
```

```
## ELN                Elan Corporation, plc         1
## MRX  Medicis Pharmaceutical Corporation         1
## WPI        Watson Pharmaceuticals, Inc.         1
## BAY                            Bayer AG         2
## CHTT                       Chattem, Inc         2
## IVX                    IVAX Corporation         2
## AGN                      Allergan, Inc.         3
## PHA               Pharmacia Corporation         3
## GSK                 GlaxoSmithKline plc         4
## JNJ                   Johnson & Johnson         4
## MRK                   Merck & Co., Inc.         4
## PFE                          Pfizer Inc         4
## ABT                 Abbott Laboratories         5
## AHM                        Amersham plc         5
## AZN                     AstraZeneca PLC         5
## BMY        Bristol-Myers Squibb Company         5
## LLY                Eli Lilly and Company         5
## NVS                         Novartis AG         5
## SGP         Schering-Plough Corporation         5
## WYE                               Wyeth         5
```

```
# Mean of all numeric variables
aggregate(Scaling_data, by=list(k_pav$cluster), FUN=mean)
```

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

```
# Adding clusters to the scaled data
Scaling_data2 <- data.frame(Scaling_data, k_pav$cluster)
Scaling_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
## 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 k_pav.cluster
## ABT  -0.21209793 -0.52776752        0.06168225             5
## AGN   0.01828430 -0.38113909       -1.55366706             3
## AHM  -0.40408312 -0.57211809       -0.68503583             5
## AZN  -0.74965647  0.14744734        0.35122600             5
## AVE  -0.31449003  1.21638667       -0.42597037             1
## BAY  -0.74965647 -1.49714434       -1.99560225             2
## BMY  -0.02011273 -0.96584257        0.74744375             5
## CHTT  3.74279705 -0.63276071       -1.24888417             2
## ELN   0.61983791  1.88617085       -0.36501379             1
## LLY  -0.07130879 -0.64814764        1.17413980             5
## GSK  -0.31449003  0.76926048        0.82363947             4
## IVX   1.10620040  0.05603085       -0.71551412             2
## JNJ  -0.62166634 -0.36213170        0.33598685             4
## MRX   0.44065173  1.53860717        0.85411776             1
## MRK  -0.39128411  0.36014907       -0.24310064             4
## NVS  -0.67286239 -1.45369888        1.02174835             5
## PFE  -0.54487226  1.10143723        1.44844440             4
## PHA  -0.30169102  0.14744734       -1.27936246             3
## SGP  -0.74965647 -0.43544591        0.29026942             5
## WPI  -0.49367621  1.43089863       -0.09070919             1
## WYE   0.68383297 -1.17763919        1.49416183             5
```

After comparing all the mean values of the numeric variables from the cluster, I Conclude that Cluster 1: Comprising AVE, WPI, MRX, ELN, this cluster is characterized by robust revenue growth and a high beta coefficient, suggesting a strong growth trajectory but lower asset efficiency and profitability. These firms may be in their nascent stages, likely investing significantly in expansion efforts. Their high beta and revenue growth imply potential for rapid earnings improvement. Cluster 2: Encompassing IVX, CHTT, BAY, these companies boast substantial market capitalization and solid returns on equity and assets, coupled with high asset turnover. Their low beta and profit-to-return ratios suggest established, stable operations with less efficient profit generation, highlighting their maturity and stability. Cluster 3: Featuring PHA, AGN, this cluster is marked by elevated price-to-earnings ratios and asset turnover, indicative of expected earnings acceleration despite historically low profitability. The combination of high valuation and low net profit margins points to a higher risk profile for investors. Cluster 4: Consisting of GSK, PFE, MRK, JNJ, this cluster stands out with the highest net profit margins and asset efficiency, demonstrating strong financial performance and low risk. The low beta and revenue growth indicate stable stock prices and modest revenue expansion, typical of mature, well-established entities. Cluster 5: Including WYE, BMY, LLY, AZN, NVS, ABT, SGP, AHM, this cluster is distinguished by a high beta and leverage, signaling higher investment risk due to stock price volatility and significant debt levels. However, these firms may offer higher returns in favorable market conditions. ## Is there a pattern in the clusters with respect to the numerical variables (10 to 12) (those not used in forming the clusters)

```r
# Adding clusters to the data
D_Pattern <- Pharmaceuticals[12:14] %>% mutate(Clustering = k_pav$cluster)
D_Pattern
```

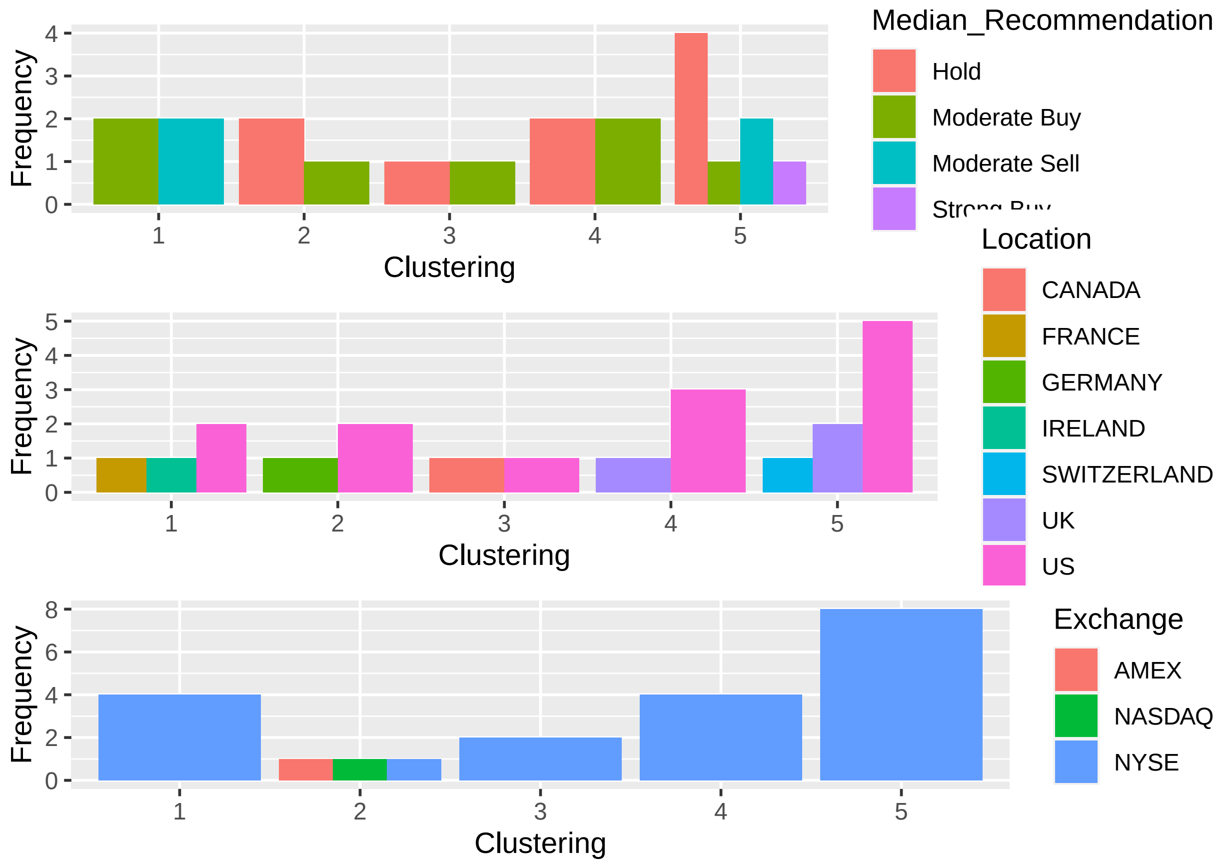
```
##      Median_Recommendation    Location Exchange Clustering
## ABT           Moderate Buy          US     NYSE          5
## AGN           Moderate Buy      CANADA     NYSE          3
## AHM             Strong Buy          UK     NYSE          5
## AZN          Moderate Sell          UK     NYSE          5
## AVE           Moderate Buy      FRANCE     NYSE          1
## BAY                   Hold     GERMANY     NYSE          2
## BMY          Moderate Sell          US     NYSE          5
## CHTT          Moderate Buy          US   NASDAQ          2
## ELN          Moderate Sell     IRELAND     NYSE          1
## LLY                   Hold          US     NYSE          5
## GSK                   Hold          UK     NYSE          4
## IVX                   Hold          US     AMEX          2
## JNJ           Moderate Buy          US     NYSE          4
## MRX           Moderate Buy          US     NYSE          1
## MRK                   Hold          US     NYSE          4
## NVS                   Hold SWITZERLAND     NYSE          5
## PFE           Moderate Buy          US     NYSE          4
## PHA                   Hold          US     NYSE          3
## SGP                   Hold          US     NYSE          5
## WPI          Moderate Sell          US     NYSE          1
## WYE                   Hold          US     NYSE          5
```

```r
# Plotting data with median recommendation
Median_Recommendation <- ggplot(D_Pattern, mapping = aes(factor(Clustering), fill = Median_Recommendati

# Plotting data with location
Location <- ggplot(D_Pattern, mapping = aes(factor(Clustering), fill = Location)) + geom_bar(position =

# Plotting data with exchange
Exchange <- ggplot(D_Pattern, mapping = aes(factor(Clustering), fill = Exchange)) + geom_bar(position =

grid.arrange(Median_Recommendation, Location, Exchange)
```



Cluster 1: Generally recommended as a moderate buy and sell, these companies are based in France, Ireland, and the US, and are listed on the NYSE. Cluster 2: With a hold or moderate buy recommendation, these companies span Germany and the US and are listed across AMEX, NASDAQ, and NYSE. Cluster 3: Recommended as hold and moderate buy, these firms are located in the US and Canada, also listed on the NYSE. Cluster 4: Recommended as a hold and moderate buy, these UK and US-based companies are traded on the NYSE. Cluster 5: Carrying diverse recommendations from hold to strong buy, these firms are from Switzerland, the UK, and the US, with listings on the NYSE.

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

Appropriate name for each clusters are: Cluster1:High Growth Potential, Cluster2:High Risk High Beta, Cluster3: High Risk High Reward, Cluster4: Stability and Profitability, Cluster5: Low Risk High Profitability.