FML Assignment 4

Pardhu suryasriraj sunkara

2023-11-08

An equities analyst is studying the pharmaceutical industry and would like your help in exploring and understanding the financial data collected by her firm. Her main objective is to understand the structure of the pharmaceutical industry using some basic financial measures. Financial data gathered on 21 firms in the pharmaceutical industry are available in the file Pharmaceuticals.csv

For each firm, the following variables are recorded:

1. Market capitalization (in billions of dollars)
2. Beta
3. Price/earnings ratio
4. Return on equity
5. Return on assets
6. Asset turnover
7. Leverage
8. Estimated revenue growth
9. Net profit margin
10. Median recommendation (across major brokerages)
11. Location of firm's headquarters
12. Stock exchange on which the firm is listed

Use cluster analysis to explore and analyze the given dataset as follows:

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
library(factoextra)
```

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

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

```
library(ISLR)
library(cluster)
library(dbscan)
```

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

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

```
library(ggplot2)
library(gridExtra)
```

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

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

#Import and clean the data.

```
pharm <- read.csv("C:\\Users\\pardh\\OneDrive\\Desktop\\FML\\Assighnment-4\\Pharmaceuticals.cs
v")

dim(pharm) #Dimensions
```

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

1. 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.

The numeric variables in the data are from 3 to 11.

```
row.names(pharm) <- pharm[, 1]
numeric_pharma <- pharm[ , c(3:11)]

dim(numeric_pharma) #Dimensions
```

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

```
numeric_pharma <- na.omit(numeric_pharma) #Omitting missing values

summary(numeric_pharma)
```

```
##    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
```

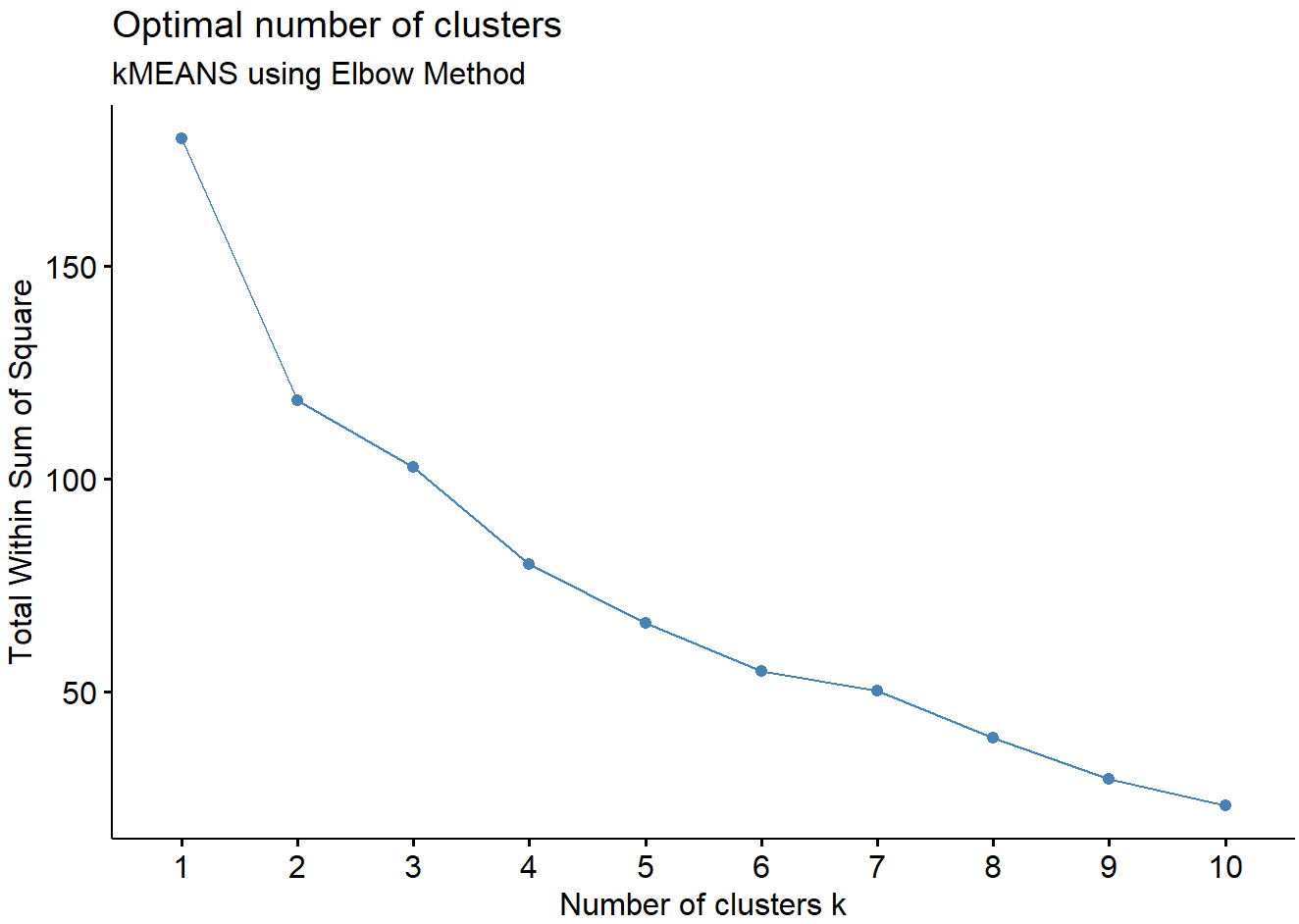
```
#Normalizing the given data
norm_pharmacy <- scale(numeric_pharma)
summary(norm_pharmacy)
```

```
##    Market_Cap           Beta            PE_Ratio            ROE
##  Min.   :-0.9768   Min.   :-1.3466   Min.   :-1.3404   Min.   :-1.4515
##  1st Qu.:-0.8763   1st Qu.:-0.6844   1st Qu.:-0.4023   1st Qu.:-0.7223
##  Median :-0.1614   Median :-0.2560   Median :-0.2429   Median :-0.2118
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.2762   3rd Qu.: 0.4841   3rd Qu.: 0.1495   3rd Qu.: 0.3450
##  Max.   : 2.4200   Max.   : 2.2758   Max.   : 3.4971   Max.   : 2.4597
##       ROA          Asset_Turnover       Leverage          Rev_Growth
##  Min.   :-1.7128   Min.   :-1.8451   Min.   :-0.74966   Min.   :-1.4971
##  1st Qu.:-0.9047   1st Qu.:-0.4613   1st Qu.:-0.54487   1st Qu.:-0.6328
##  Median : 0.1289   Median :-0.4613   Median :-0.31449   Median :-0.3621
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.00000   Mean   : 0.0000
##  3rd Qu.: 0.8430   3rd Qu.: 0.9225   3rd Qu.: 0.01828   3rd Qu.: 0.7693
##  Max.   : 1.8389   Max.   : 1.8451   Max.   : 3.74280   Max.   : 1.8862
##  Net_Profit_Margin
##  Min.   :-1.99560
##  1st Qu.:-0.68504
##  Median : 0.06168
##  Mean   : 0.00000
##  3rd Qu.: 0.82364
##  Max.   : 1.49416
```

Analysing the data with different clustering Methods

1. K-means clustering

```
fviz_nbclust(norm_pharmacy, kmeans, method = "wss") + labs(subtitle = "kMEANS using Elbow Metho
d")
```



From the graph, the optimal number of cluster k can be considered as 2 because the bent on the line is clearly seen at point 2.

therefore, K = 2

```
set.seed(159)
k = 2
k2 <- kmeans(norm_pharmacy, centers = k, nstart = 21)
```

```
#Finding the centroids
k2$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
```

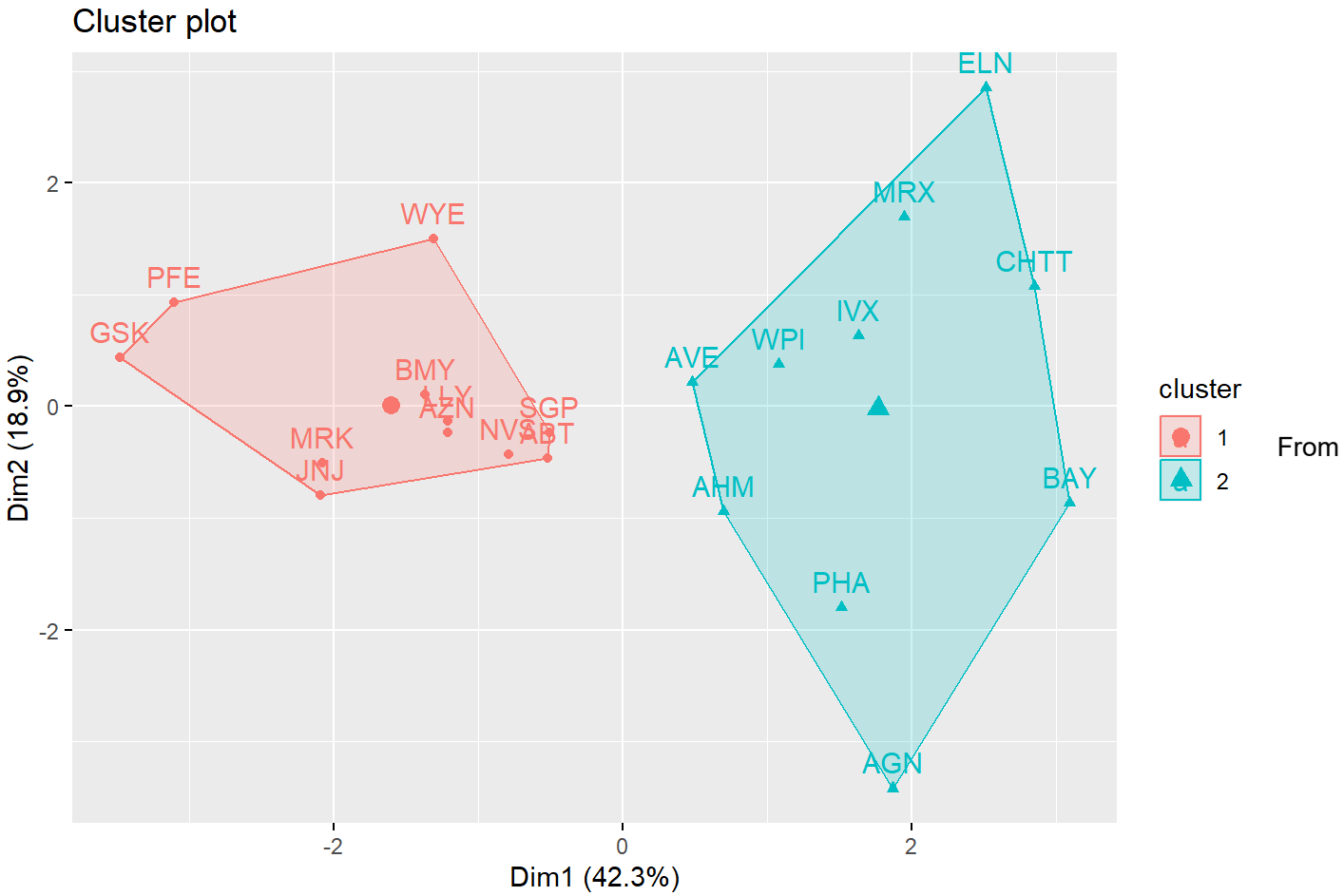
```
#Finding the size of each cluster
k2$size
```

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

```
k2$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
```

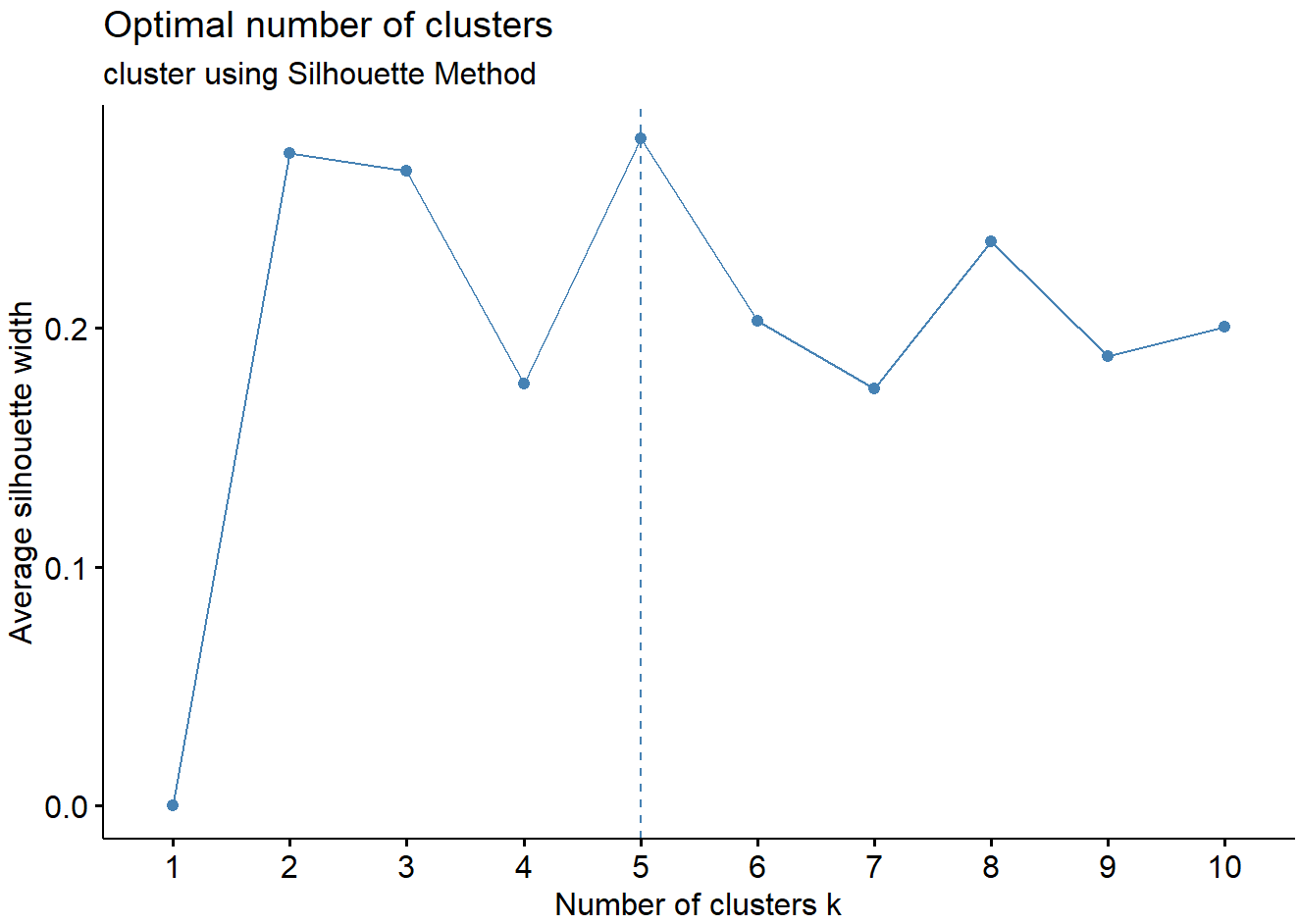
```
fviz_cluster(k2, norm_pharmacy)
```



the above graph, cluster-1 includes 11 firms and cluster-2 includes 10 firms. There are no outliers. This suggests that all the firms were succesfully assigned to the cluster.

#Similarly the silhouttee method can be used for taking out the optimal number of clusters

```
fviz_nbclust(norm_pharmacy, kmeans, method = "silhouette") + labs(subtitle = "cluster using Silh
ouette Method ")
```



From the above graph, the line is cutting the graph at point 5. This tells that the optimal number under this method can be taken as 5.(k = 5)

Running K-means clustering, taking k = 5

```
set.seed(159)
#k = 5
k5 <- kmeans(norm_pharmacy, centers = 5, nstart = 21)
```

```
#Finding the centroids
k5$centers
```

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

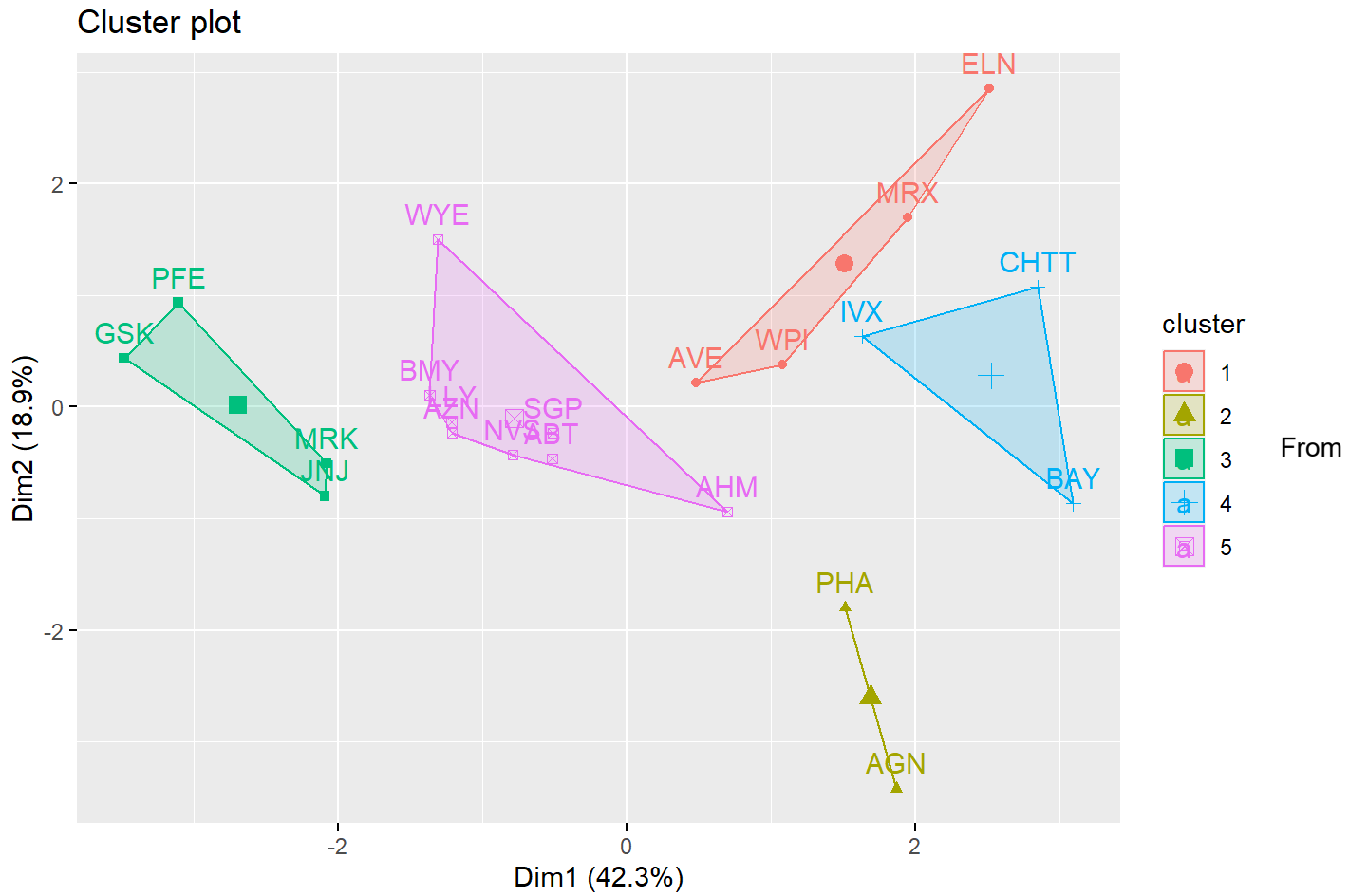
```
#Finding the size of each cluster
k5$size
```

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

```
k5$cluster
```

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

```
fviz_cluster(k5, norm_pharmacy)
```



the above cluster graph, Cluster 1 includes 5 firms, cluster 2 includes 2 firms, cluster 3 includes 4 firms, cluster 4 includes 3 firms and cluster 5 includes 8 firms in Total. This clustering will be the best one as it is clustered considering all the numerical data to know the equity such as Market capital, net profit, return on assets, asset turnover, etc.

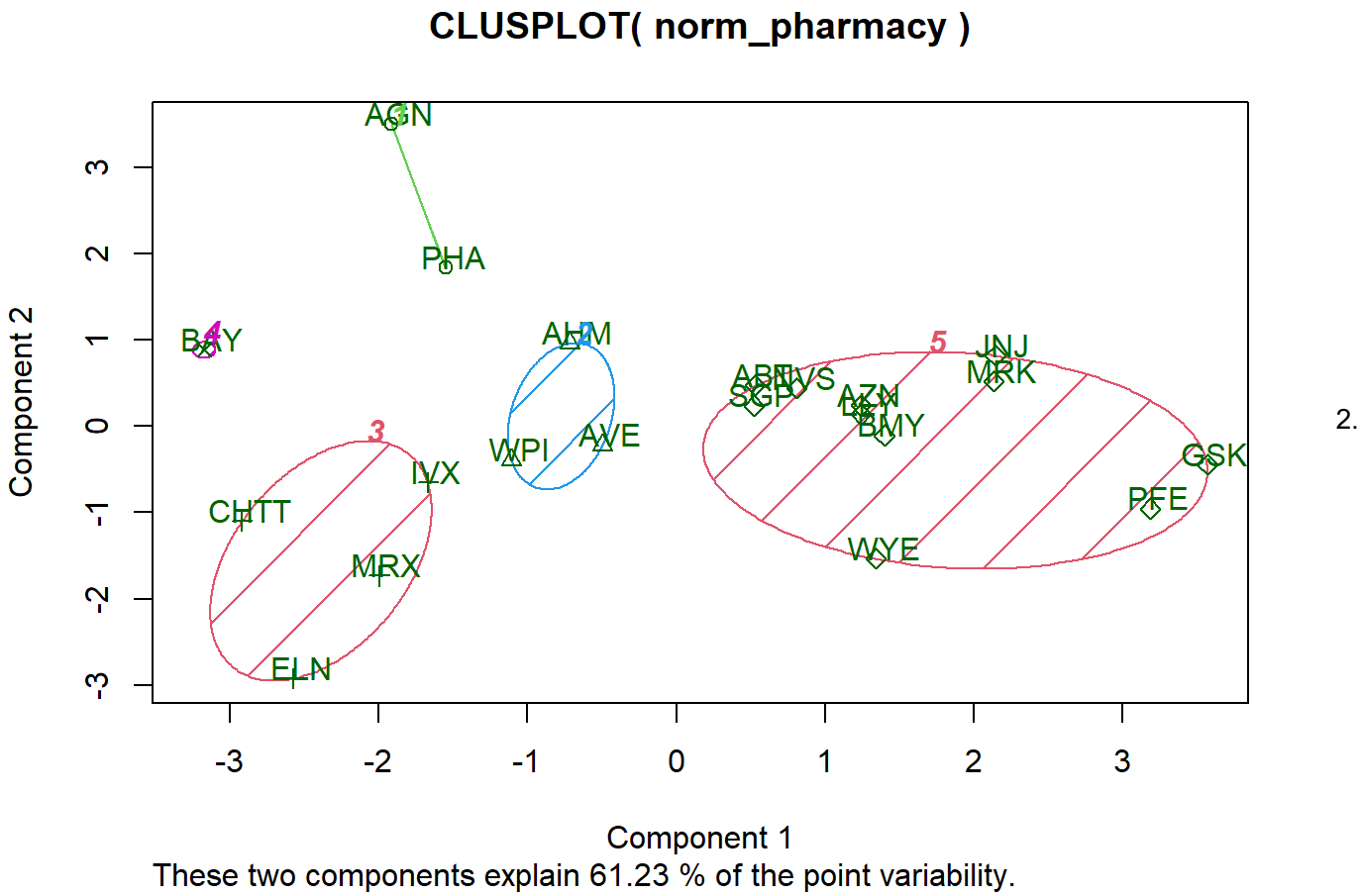
#Fitting the data with 5 clusters.

```
fit_data <- kmeans(norm_pharmacy, 5)
```

```
#calculating the mean of all the variables in each cluster
aggregate(norm_pharmacy, by = list(fit_data$cluster), FUN = mean)
```

```
##   Group.1 Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1       1 -0.4392513 -0.4701800  2.7000246 -0.8349525 -0.9234951      0.2306328
## 2       2 -0.6611400 -0.7233539 -0.3512251 -0.6736441 -0.5915022     -0.1537552
## 3       3 -0.9624758  1.1949250 -0.3639982 -0.5200697 -0.9610792     -1.1531640
## 4       4 -0.6953818  2.2757827  0.1494823 -1.4514600 -1.7127612     -0.4612656
## 5       5  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.1417034 -0.1168459        -1.4165148
## 2 -0.4040831  0.6917224        -0.4005718
## 3  1.4773718  0.7120120        -0.3688236
## 4 -0.7496565 -1.4971443        -1.9956023
## 5 -0.3331068 -0.2902163         0.6823310
```

```
clusplot(norm_pharmacy, fit_data$cluster, color = TRUE, shade = TRUE, labels = 2, lines = 0)
```

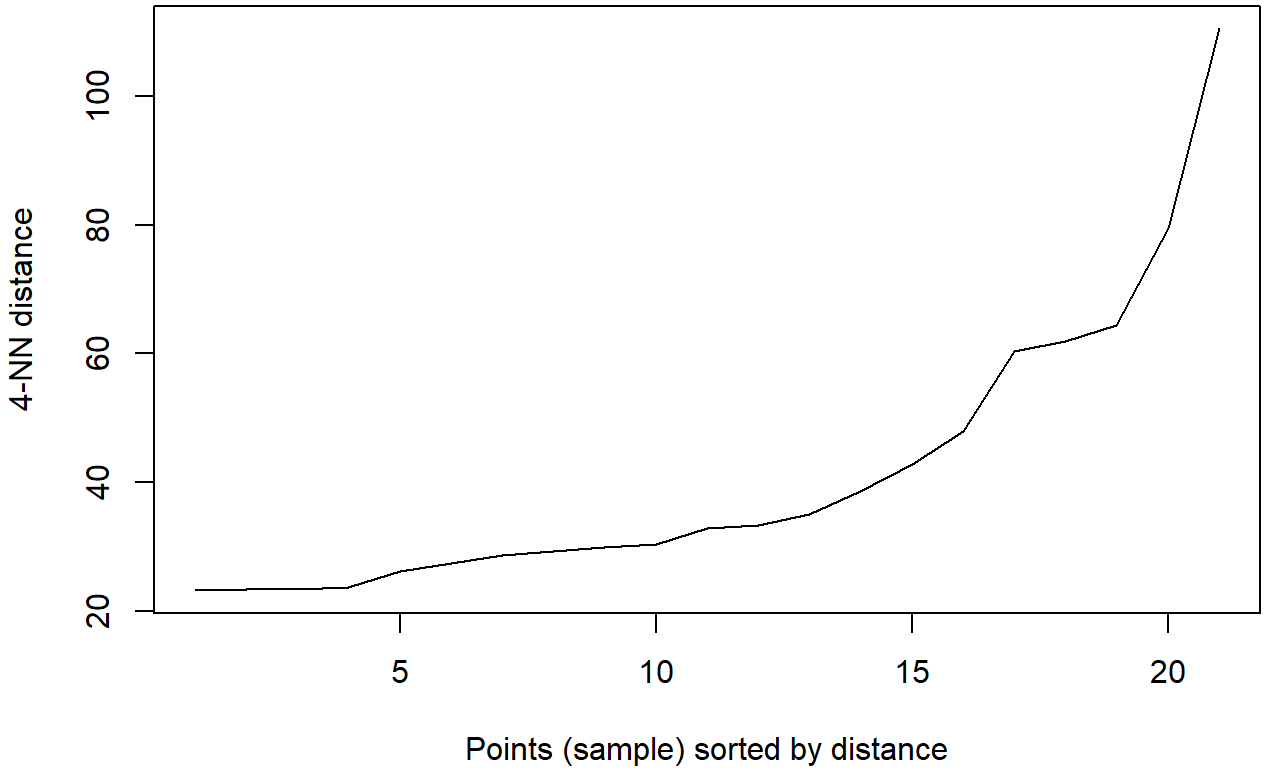


2.

DBSCAN clustering

Determining the optimal ‘eps’ value

```
dbscan::kNNdistplot(numeric_pharma, k = 4)
```



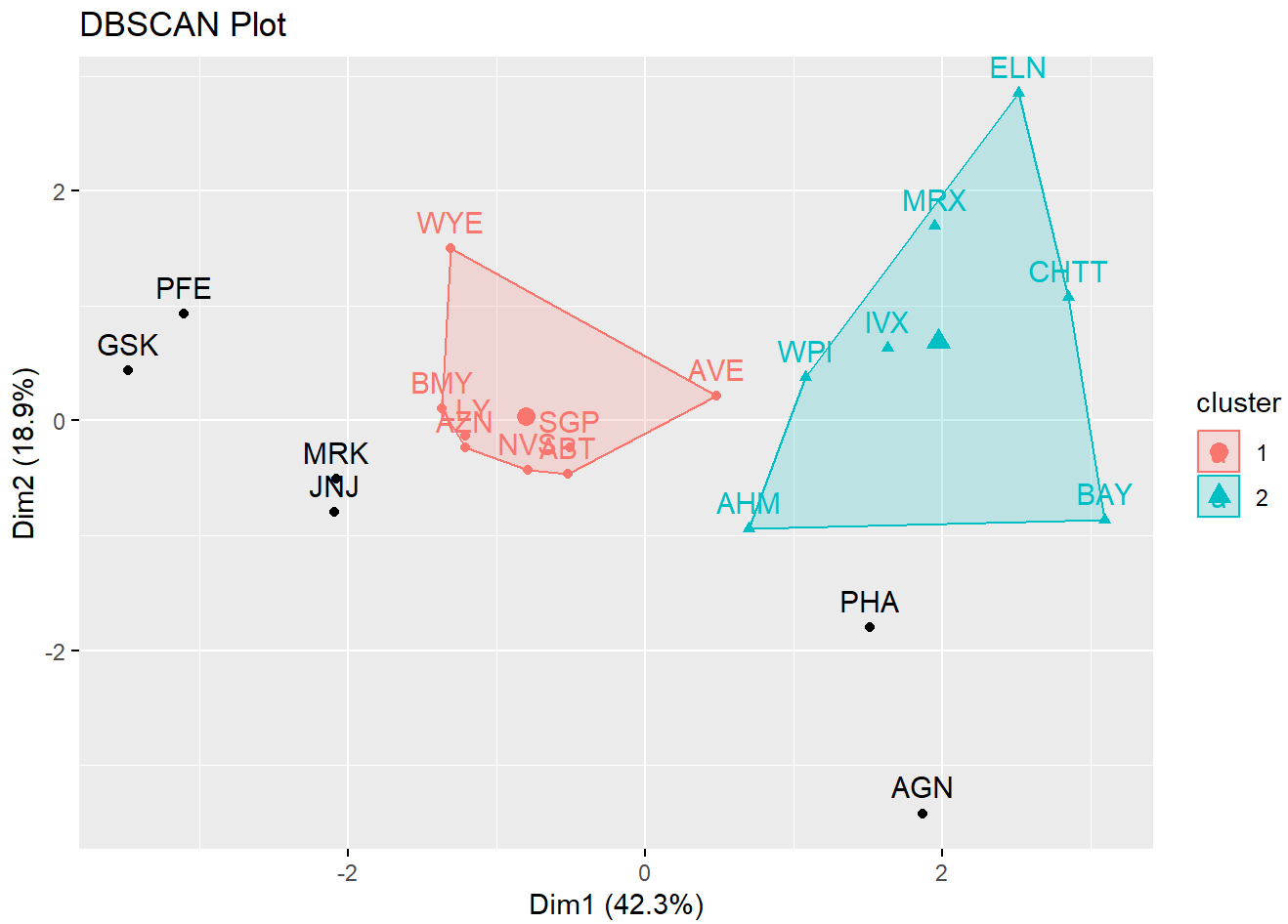
The knn-dist plot above is used to find the optimal value of radius for DBSCAN clustering. In the graph the curve was bent at a distance between 20 to 40. SO, the EPS value can be at 30 at 4 minimum points.

```
dbscan_cluster <- dbscan::dbscan(numeric_pharma, eps = 30, minPts = 4)
```

```
dbscan_cluster
```

```
## DBSCAN clustering for 21 objects.
## Parameters: eps = 30, minPts = 4
## Using euclidean distances and borderpoints = TRUE
## The clustering contains 2 cluster(s) and 6 noise points.
##
## 0 1 2
## 6 8 7
##
## Available fields: cluster, eps, minPts, dist, borderPoints
```

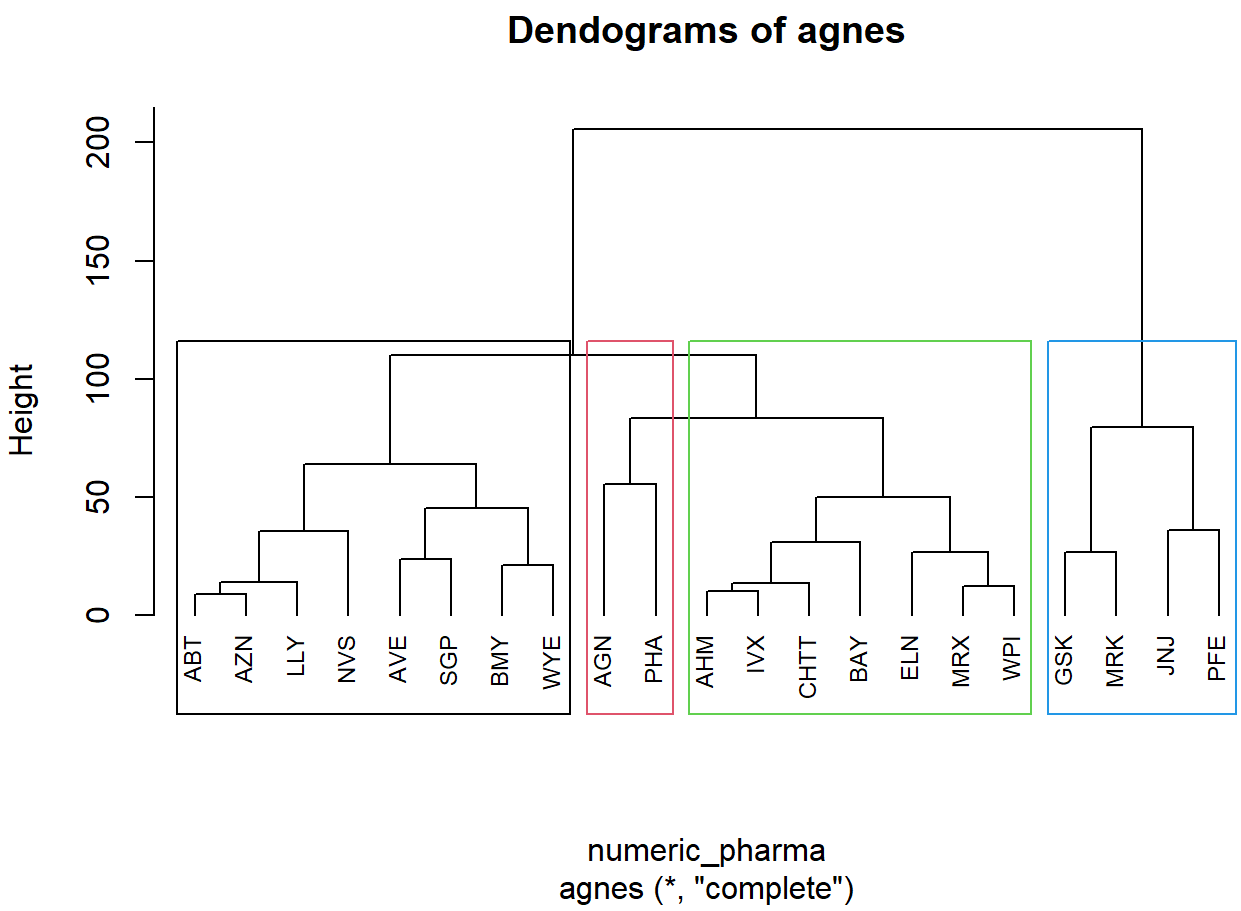
```
fviz_cluster(dbscan_cluster, numeric_pharma) + ggtitle("DBSCAN Plot")
```



From the above plot, we can see that only 2 clusters are formed one with 8 firms and other with 7 firms. The remaining other firms are plotted outside the clusters which can be considered as outliers. A good cluster method should adjust all the points to the cluster. Therefore, DBSCAN is not a good form of clustering analysis.

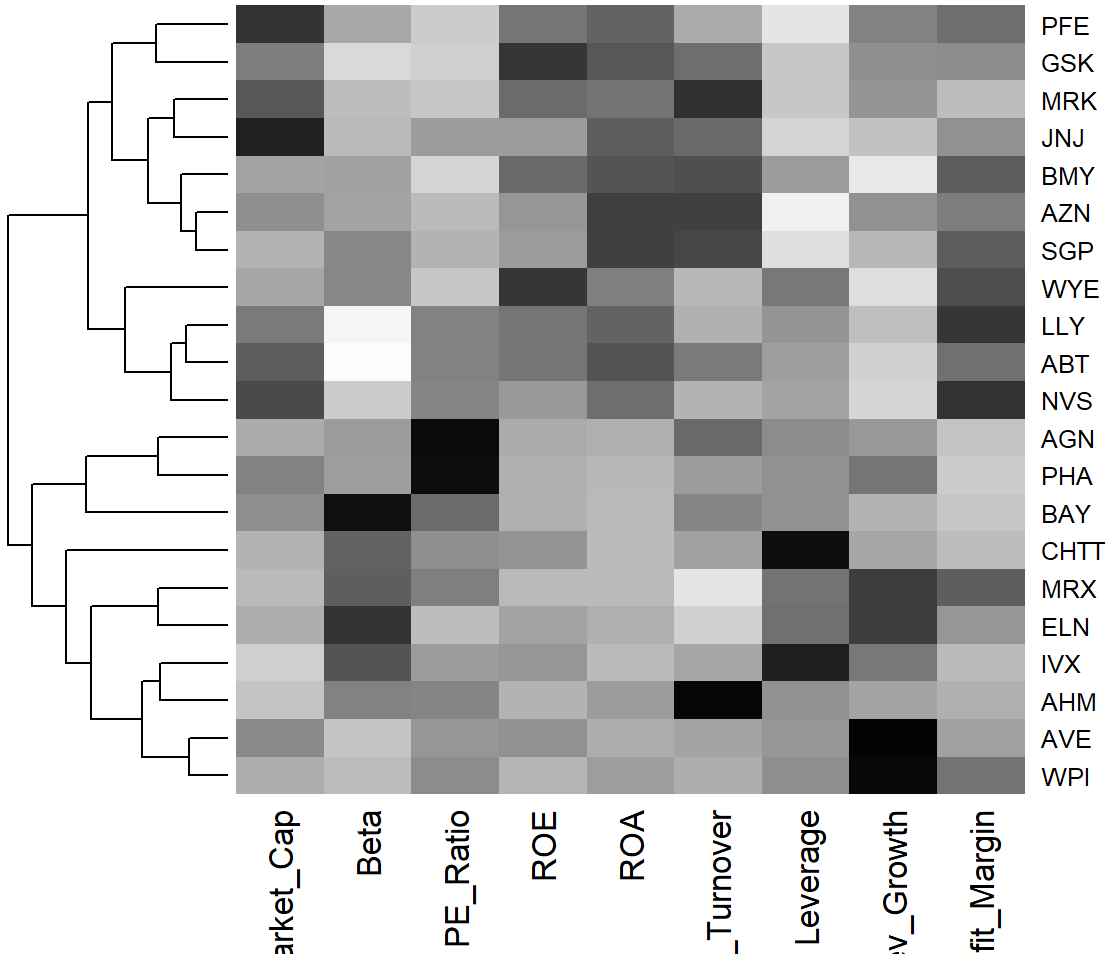
3. Hierarchical clustering

```
#Dissimilarity matrix
d <- dist(numeric_pharma,method = "euclidean")

#Hierarchical clustering using complete Linkage(Agglomerative clustering)
hc1 <- agnes(numeric_pharma, method = "complete")

#Plot the obtained dendogram
pltree(hc1, cex = 0.75, hang = -1, main = "Dendograms of agnes")
rect.hclust(hc1, k = 4, border = 1:4)
```



In Hierarchical clustering, 4 clusters are formed. 1st cluster includes 8 firms, 2nd cluster includes 2 firms, 3rd cluster includes 7 firms and 4th cluster includes 4 firms. From the analysis of the dendogram, it is suggestible not to choose hierarchical clustering because the cluster is divided either with too many points or with too less points.

```
heatmap(as.matrix(norm_pharmacy), Colv = NA, hclustfun = hclust,
        col=rev(paste("gray",1:99,sep="")))
```



Out of all these Clustering methods, k-means clustering is the better option to consider because the points are distributed to all the clusters completely with out any outliers and also consider all the variables effecting the data.

---

2. Interpret the clusters with respect to the numerical variables used in forming the clusters. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)?

```
#creating the table with clusters
clust_data <- pharm[,c(2:11)]  %>%
                  mutate(cluster = k5$cluster) %>% arrange(cluster, ascending = T)

clust_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
## 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
## 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
## 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 cluster
## 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
## AGN             0.9     0.60       9.16               5.5       2
## PHA             0.6     0.35      15.00               7.3       2
## GSK             1.0     0.34      21.87              21.1       3
## JNJ             0.9     0.10       9.37              17.9       3
## MRK             1.1     0.28      17.35              14.1       3
## PFE             0.8     0.16      25.54              25.2       3
## BAY             0.6     0.00      -3.17               2.6       4
## CHTT            0.6     3.51       6.38               7.5       4
## IVX             0.6     1.45      13.99              11.0       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
```

```
# calculate the mean of all numerical variables
aggregate(norm_pharmacy, by=list(k5$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 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951
## 3       3  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431
## 4       4 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
## 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      0.2306328 -0.14170336 -0.1168459      -1.416514761
## 3      1.1531640 -0.46807818  0.4671788       0.591242521
## 4     -0.4612656  1.36644699 -0.6912914      -1.320000179
## 5      0.1729746 -0.27449312 -0.7041516       0.556954446
```

```
#Adding cluster to normalized data
norm_pharm1 <- data.frame(norm_pharmacy, k5$cluster)
norm_pharm1
```

```
##      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 k5.cluster
## ABT  -0.21209793 -0.52776752        0.06168225          5
## AGN   0.01828430 -0.38113909       -1.55366706          2
## 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          4
## BMY  -0.02011273 -0.96584257        0.74744375          5
## CHTT  3.74279705 -0.63276071       -1.24888417          4
## ELN   0.61983791  1.88617085       -0.36501379          1
## LLY  -0.07130879 -0.64814764        1.17413980          5
## GSK  -0.31449003  0.76926048        0.82363947          3
## IVX   1.10620040  0.05603085       -0.71551412          4
## JNJ  -0.62166634 -0.36213170        0.33598685          3
## MRX   0.44065173  1.53860717        0.85411776          1
## MRK  -0.39128411  0.36014907       -0.24310064          3
## NVS  -0.67286239 -1.45369888        1.02174835          5
## PFE  -0.54487226  1.10143723        1.44844440          3
## PHA  -0.30169102  0.14744734       -1.27936246          2
## SGP  -0.74965647 -0.43544591        0.29026942          5
## WPI  -0.49367621  1.43089863       -0.09070919          1
## WYE   0.68383297 -1.17763919        1.49416183          5
```

By comparing the mean values of all numeric variables from the clusters.

Cluster1 with the firms AVE, WPI, MRX, ELN has high revenue growth and beta value. but have low asset turnover, return on equity and return on asset.And the market capitalization is also relatively low.

Cluster2 with firms PHA, AGN has high Price or earnings ratio and asset turnover, but have low net profit margin, return on equity and return on asset. and the market capitalization is also relatively low.

Cluster3 with firms IVX, CHTT, BAY has high market capitalization, return on equity, Return on assets and Asset turnover. but they have lowest Beta and profit to return Ratio.

Cluster4 with firms WYE, BMY, LLY, AZN, NVS, ABT, SGP, AHM has high beta value and leverage. but have lowest net profit margin, market capitalization. And relatively low return on equity, return on asset, revenue growth.

Cluster5 with firms GSK, PFE, MRK, JNJ has highest net profit margin, asset turnover, return on equity, Return on assets. but have lowest Beta, profit to return Ratio, revenue growth. these features shows that these companies have high financial performance and low risk.

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

```
# Add the clusters to the data
clus_data1 <- pharm[12:14] %>% mutate(Clusters = k5$cluster)
clus_data1
```

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

Based on mean values:

```
filter(clus_data1, clus_data1$Clusters==1)
```

```
##     Median_Recommendation Location Exchange Clusters
## AVE          Moderate Buy   FRANCE     NYSE        1
## ELN         Moderate Sell  IRELAND     NYSE        1
## MRX          Moderate Buy       US     NYSE        1
## WPI         Moderate Sell       US     NYSE        1
```

Cluster 1 - AVE, ELN, MRX, and WPI comprise Cluster 1. The highest metrics in this cluster are Market_cap, ROA, ROE, and Asset_Turnover; the lowest are Beta and PE_Ratio.

```
filter(clus_data1, clus_data1$Clusters==2)
```

```
##     Median_Recommendation Location Exchange Clusters
## AGN          Moderate Buy   CANADA     NYSE        2
## PHA                  Hold       US     NYSE        2
```

Cluster 2 - AGN, PHA make up Cluster 2 has the lowest PE Ratio, Asset Turnover, and the highest Rev_Growth.

```
filter(clus_data1, clus_data1$Clusters==3)
```

```
##     Median_Recommendation Location Exchange Clusters
## GSK                  Hold       UK     NYSE        3
## JNJ          Moderate Buy       US     NYSE        3
## MRK                  Hold       US     NYSE        3
## PFE          Moderate Buy       US     NYSE        3
```

Cluster 3 - GSK, JNJ, MRK, and PFE make up Cluster 3; it has the lowest Market Cap, ROE, ROA, Leverage, Rev Growth, and Net Profit Margin, and the highest Beta and Leverage

```
filter(clus_data1, clus_data1$Clusters==4)
```

```
##      Median_Recommendation Location Exchange Clusters
## BAY                   Hold  GERMANY     NYSE        4
## CHTT          Moderate Buy       US   NASDAQ        4
## IVX                   Hold       US     AMEX        4
```

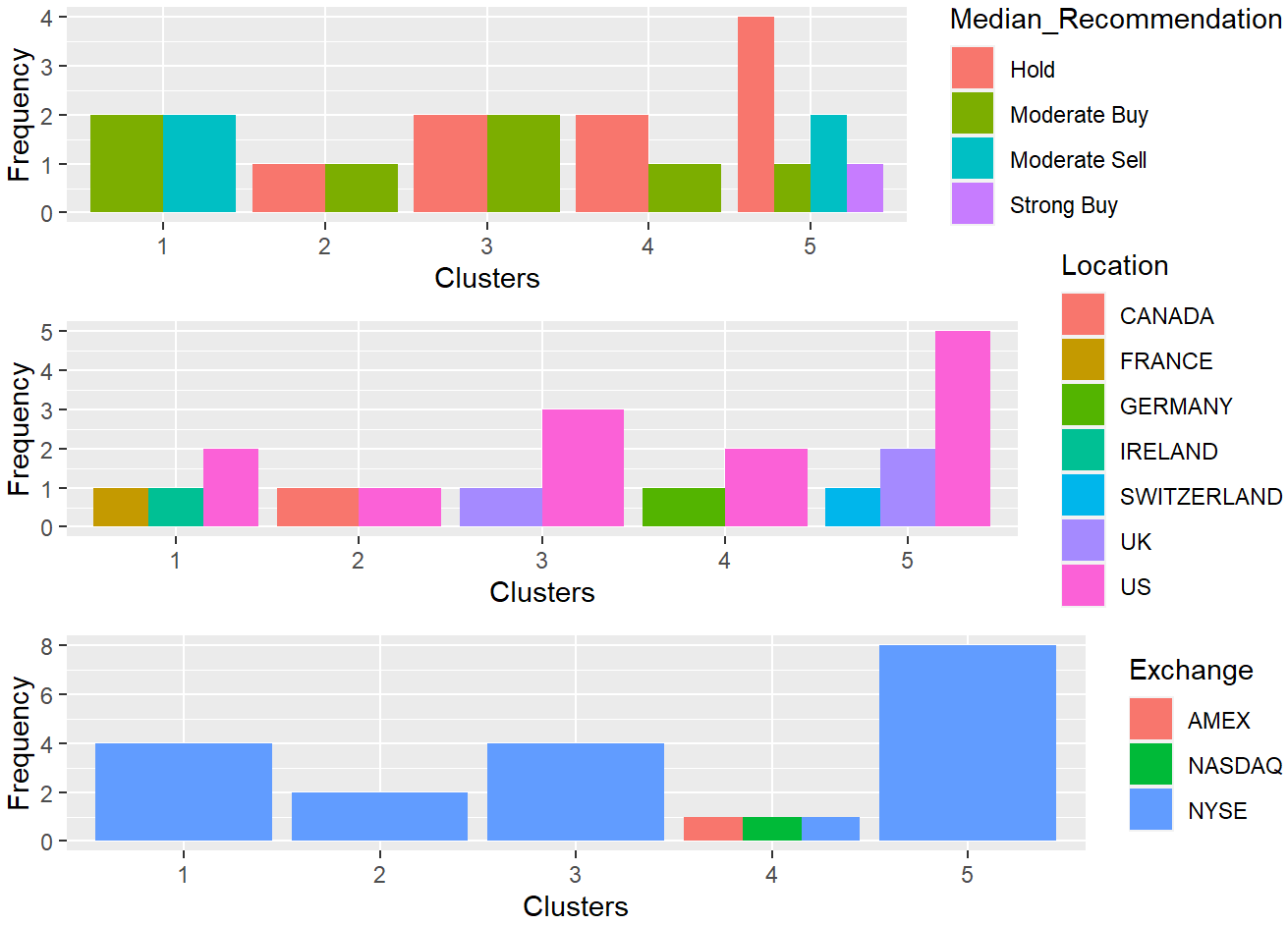
Cluster 4 - BAY, CHTT, and IVX make up Cluster 4, which has the lowest leverage and asset turnover ratios and the highest PE ratio.

```
filter(clus_data1, clus_data1$Clusters==5)
```

```
##     Median_Recommendation    Location Exchange Clusters
## ABT          Moderate Buy          US     NYSE        5
## AHM             Strong Buy         UK     NYSE        5
## AZN         Moderate Sell          UK     NYSE        5
## BMY         Moderate Sell          US     NYSE        5
## LLY                  Hold          US     NYSE        5
## NVS                  Hold SWITZERLAND     NYSE        5
## SGP                  Hold          US     NYSE        5
## WYE                  Hold          US     NYSE        5
```

Cluster 5: ABT, AHM, AZN, BMY, NVS, SGP, LLY, WYE ~ Cluster 5 has the lowest leverage, beta, and the highest Net Profit Margin.

```
# Plot the data with Median_Recommendation
recommendation <- ggplot(clus_data1, mapping = aes(factor(Clusters), fill =Median_Recommendatio
n)) + geom_bar(position='dodge') + labs(x ='Clusters',y = 'Frequency')

# Plot the data with Location
location <- ggplot(clus_data1, mapping = aes(factor(Clusters), fill = Location)) + geom_bar(posi
tion = 'dodge') + labs(x='Clusters',y = 'Frequency')

# Plot the data with Exchange
exchange <- ggplot(clus_data1, mapping = aes(factor(Clusters), fill = Exchange)) + geom_bar(posi
tion = 'dodge') + labs(x='Clusters',y = 'Frequency')

grid.arrange(recommendation, location, exchange)
```



Cluster1, Recommended as Moderate Buy and Moderate Sell from Locations France, Ireland and US and was listed under NYSE.

Cluster2, Recommended as Hold and Moderate Buy from Locations US and canada, and listed under NYSE.

Cluster3, Recommended as Hold and Moderate Buy from Locations UK and US, and listed under NYSE.

Cluster4, Recommended as Hold and Moderate Buy from Locations Germany and US and listed under AMEX, NASDAQ and NYSE.

Cluster5, Recommended Hold, Moderate Sell, Strong Buy & Moderate Buy from Locations Switzerland, UK and US and listed under NYSE

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

Appropriate names for each cluster

cluster-1 : High Buy (as it is the high growth potential cluster)

cluster-2 : Significant risk (as this cluster shows High risk and simultaneously high reward)

cluster-3 : Go for it (As this cluster shows the stability and profitability)

cluster-4 : Avoid or Runway (as this shows the high risk and high beta cluster)

cluster-5 : Perfect cluster (THis cluster shows low risk and also high profitability cluster)