

Assignment 4

keerthi Tiyyagura

2023-11-12

```
#Installing the required packages
```

```
library(flexclust)#Calling installed libraries
```

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

```
## Loading required package: grid
```

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

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```
library(cluster)
```

```
library(tidyverse)
```

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

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.3      v readr      2.1.4
```

```
## v forcats    1.0.0      v stringr    1.5.0
```

```
## v ggplot2    3.4.3      v tibble     3.2.1
```

```
## v lubridate  1.9.2      v tidyr      1.3.0
```

```
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()    masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
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(FactoMineR)
```

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

```
library(ggcorrplot)
```

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

```
#Import the "Pharmaceuticals" dataset  
Pharmaceu <- read.csv("C:/Users/keert/Downloads/Pharmaceuticals.csv")  
summary(Pharmaceu)
```

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

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.

```
Pharmac <- na.exclude(Pharmaceu) # It provides the data after removal of missing values.  
Pharmac
```

```
##      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	BMJ	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
```

#To cluster the 21 firms, the quantitative variables(1-9) are required.

```
row.names(Pharmac) <- Pharmac[,1]
Pharma1 <- Pharmac[,3:11]
head(Pharma1)
```

```
##      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
```

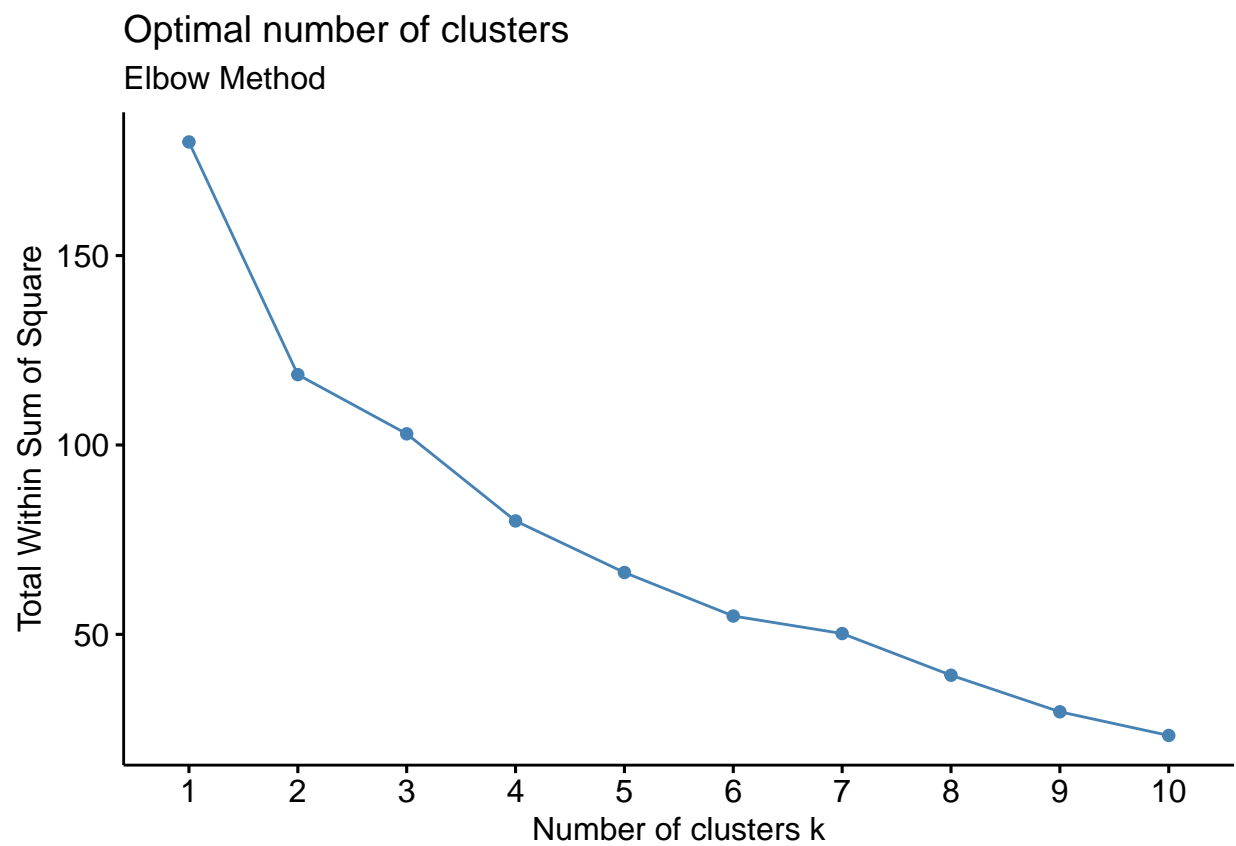
#Scale all dataframe's quantitative variables

```
Pharma2 <- scale(Pharma1)
head(Pharma2)
```

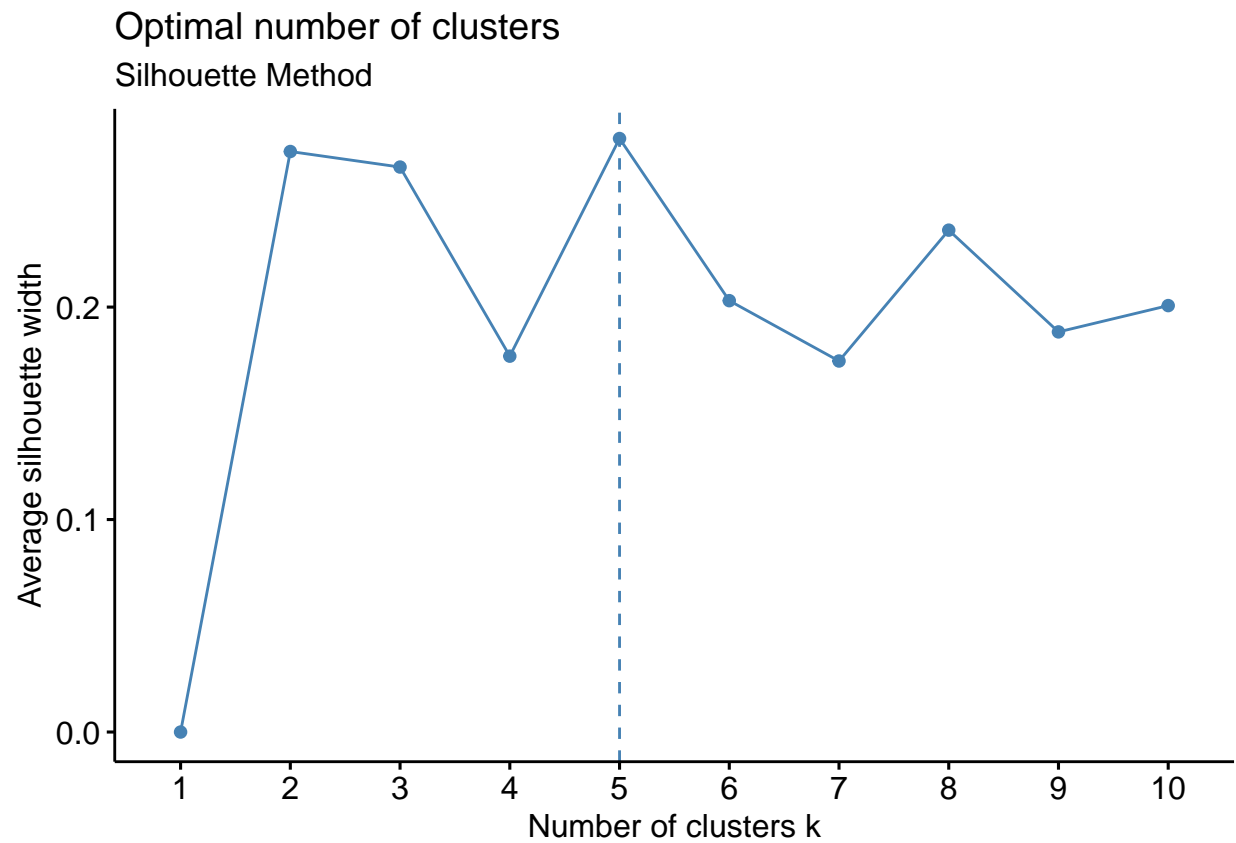
```
##      Market_Cap      Beta      PE_Ratio      ROE      ROA Asset_Turnover
## ABT  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121  0.0000000
## AGN -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871  0.9225312
## AHM -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700  0.9225312
## AZN  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259  0.9225312
## AVE -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461 -0.4612656
## BAY -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612 -0.4612656
##      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
```

#Elbow method

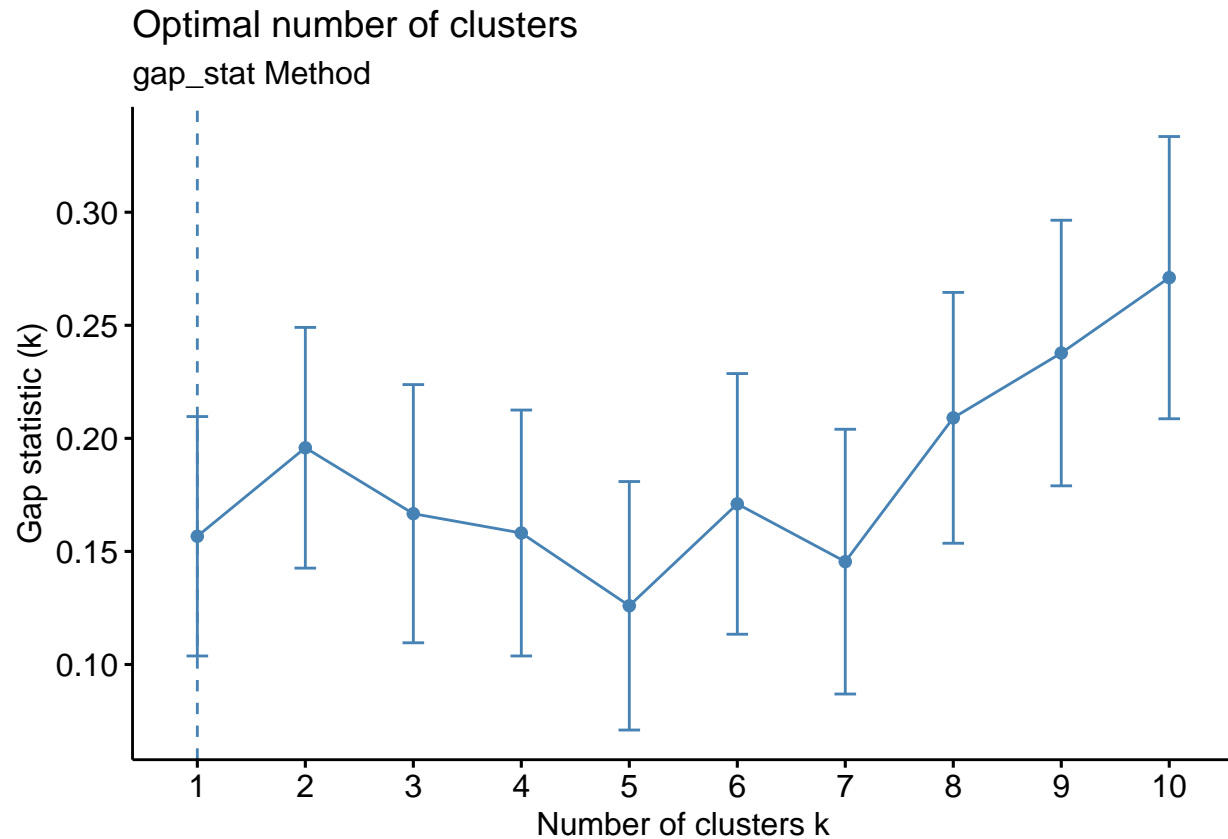
```
# Using Elbow method, calculate the no.of clusters to do the cluster analysis.  
fviz_nbclust(Pharma2, kmeans, method = "wss") + labs(subtitle = "Elbow Method")
```



```
# Using silhouette method to calculate no.of clusters  
fviz_nbclust(Pharma2, kmeans, method = "silhouette") + labs(subtitle = "Silhouette Method")
```



```
# Also using the gap_stat method for calculating the no. of clusters  
fviz_nbclust(Pharma2, kmeans, method = "gap_stat") + labs(subtitle = "gap_stat Method")
```



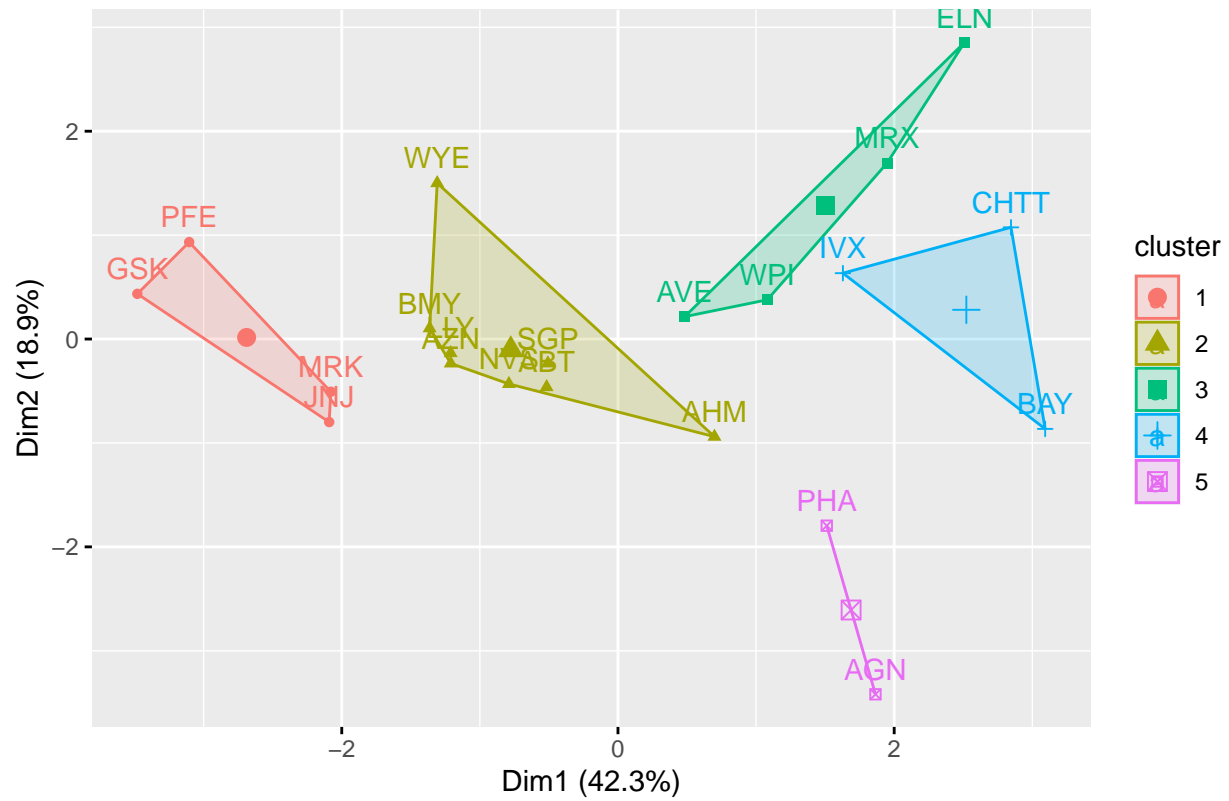
#In all the above plots,no.of clusters=5,which is sufficient to display the data variations.

```
set.seed(10)
k5<- kmeans(Pharma2,centers=5,nstart = 25)
#Visualizing the output
k5$centers# for centroids
```

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

```
fviz_cluster(k5,data = Pharma2) # to Visualize the clusters
```

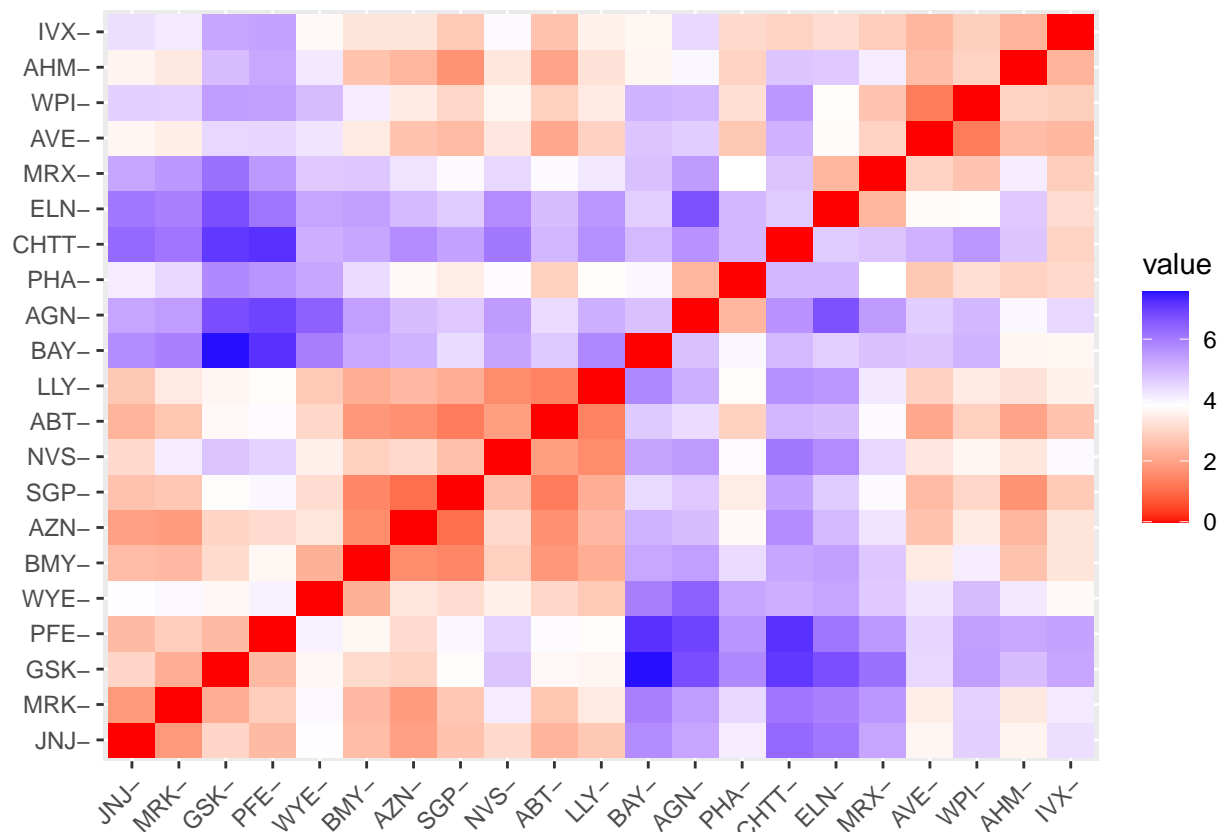
Cluster plot



```
k5$size
```

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

```
Distance<- dist(Pharma2, method = "euclidean")
fviz_dist(Distance)
```

```
#To Fit the data with 5 clusters Using K-Means Cluster Analysis
```

```
Fit<-kmeans(Pharma2,5)
```

```
#calculate the mean of all quantitative variables in each cluster
```

```
aggregate(Pharma2,by=list(Fit$cluster),FUN=mean)
```

```
##      Group.1 Market_Cap      Beta  PE_Ratio      ROE      ROA
## 1         1  0.6733825 -0.3586419 -0.27635122  0.6565978  0.8344159
## 2         2 -0.9767669  1.2630872  0.03299122 -0.1123792 -1.1677918
## 3         3 -0.5246281  0.4451409  1.84984387 -1.0404550 -1.1865838
## 4         4 -0.7307042 -0.4214928 -0.34867046 -0.5780744 -0.6181243
## 5         5 -0.9668697  1.5162611 -0.57398880 -0.8382671 -0.9892673
##      Asset_Turnover      Leverage Rev_Growth Net_Profit_Margin
## 1  4.612656e-01 -0.33310678 -0.2902163      0.6823310
## 2 -4.612656e-01  3.74279705 -0.6327607     -1.2488842
## 3  1.480297e-16 -0.34435439 -0.5769454     -1.6095439
## 4 -2.306328e-01 -0.02651224  0.5327995     -0.4793074
## 5 -1.845062e+00  0.53024482  1.7123890      0.2445520
```

```
Pharma3<-data.frame(Pharma2,Fit$cluster)
Pharma3
```

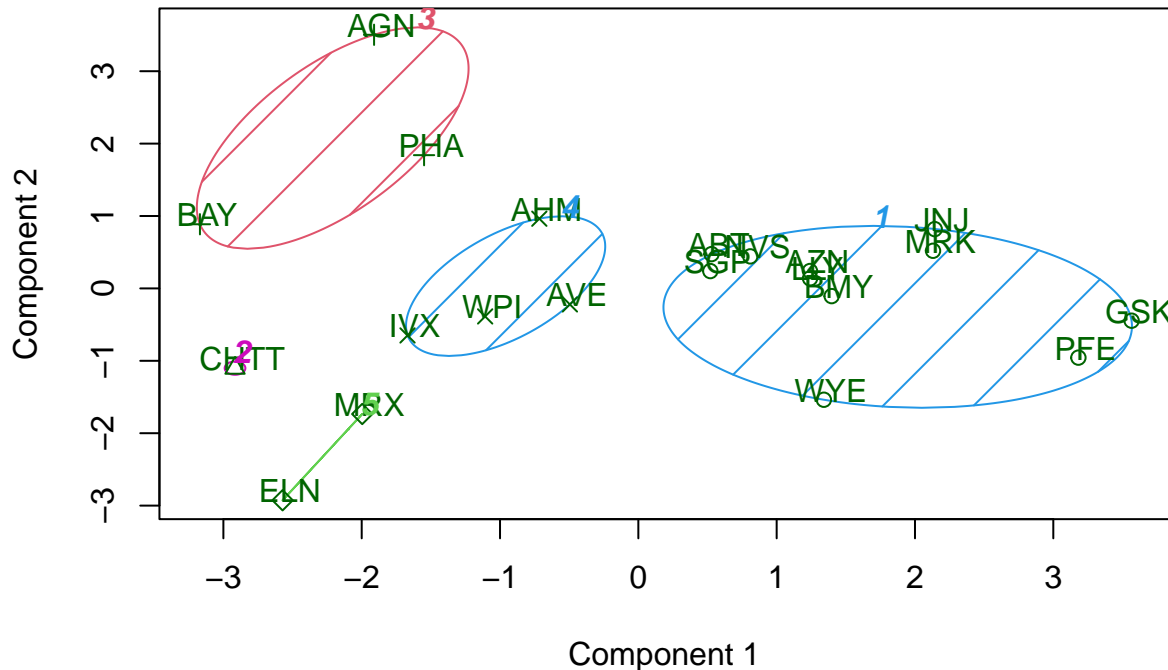
```
##      Market_Cap      Beta  PE_Ratio      ROE      ROA Asset_Turnover
```

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

#view of the cluster plot

```
clusplot(Pharma2,Fit$cluster,color = TRUE,shade = TRUE,labels = 2,lines = 0)
```

CLUSPLOT(Pharma2)



These two components explain 61.23 % of the point variability.

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)?

Based on mean values:

JNJ, MRK, PFE, and GSK comprise Cluster 1. The highest metrics in this cluster are Market_cap, ROA, ROE, and Asset_Turnover; the lowest are Beta and PE_Ratio.

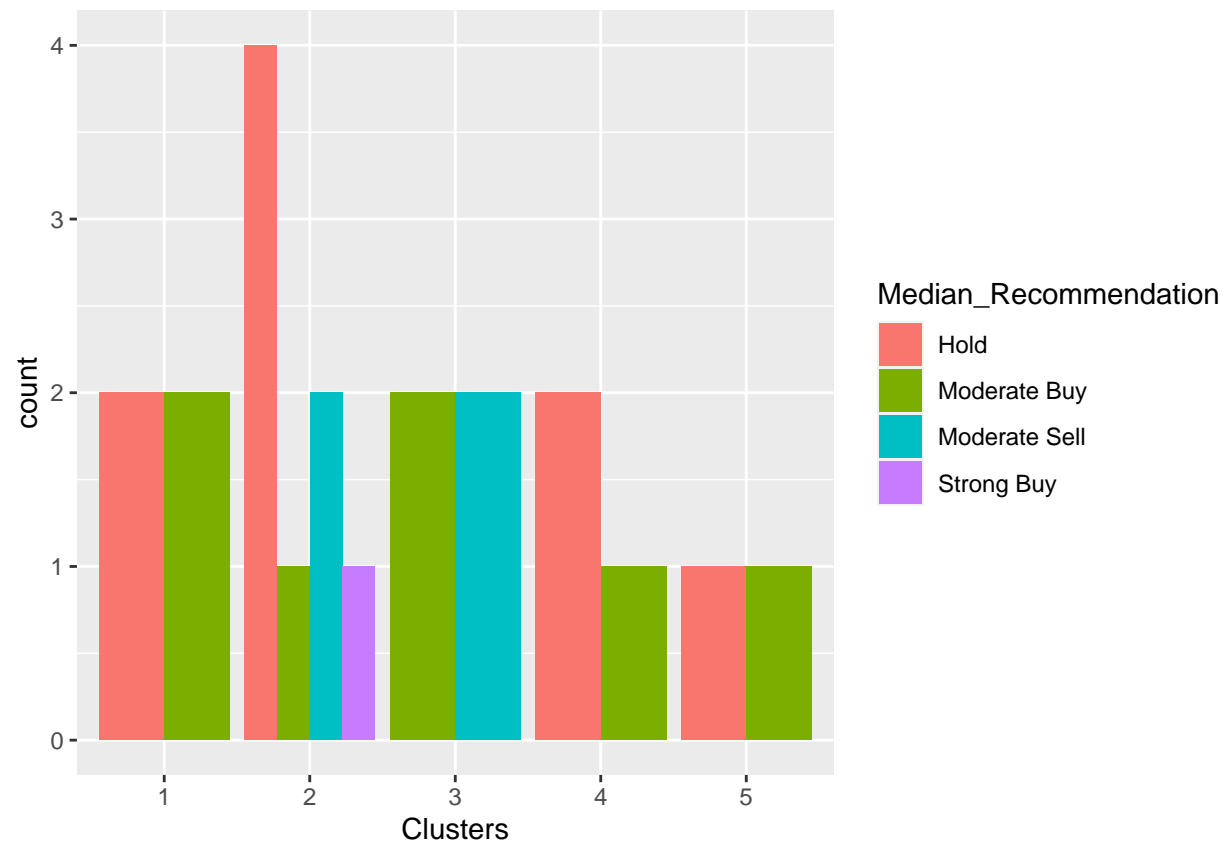
Cluster 2 - AHM, WPI, AVE ~Cluster 2 has the lowest PE Ratio, Asset Turnover, and the highest Rev_Growth.

CHTT, ELN, MRX, and IVX 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.

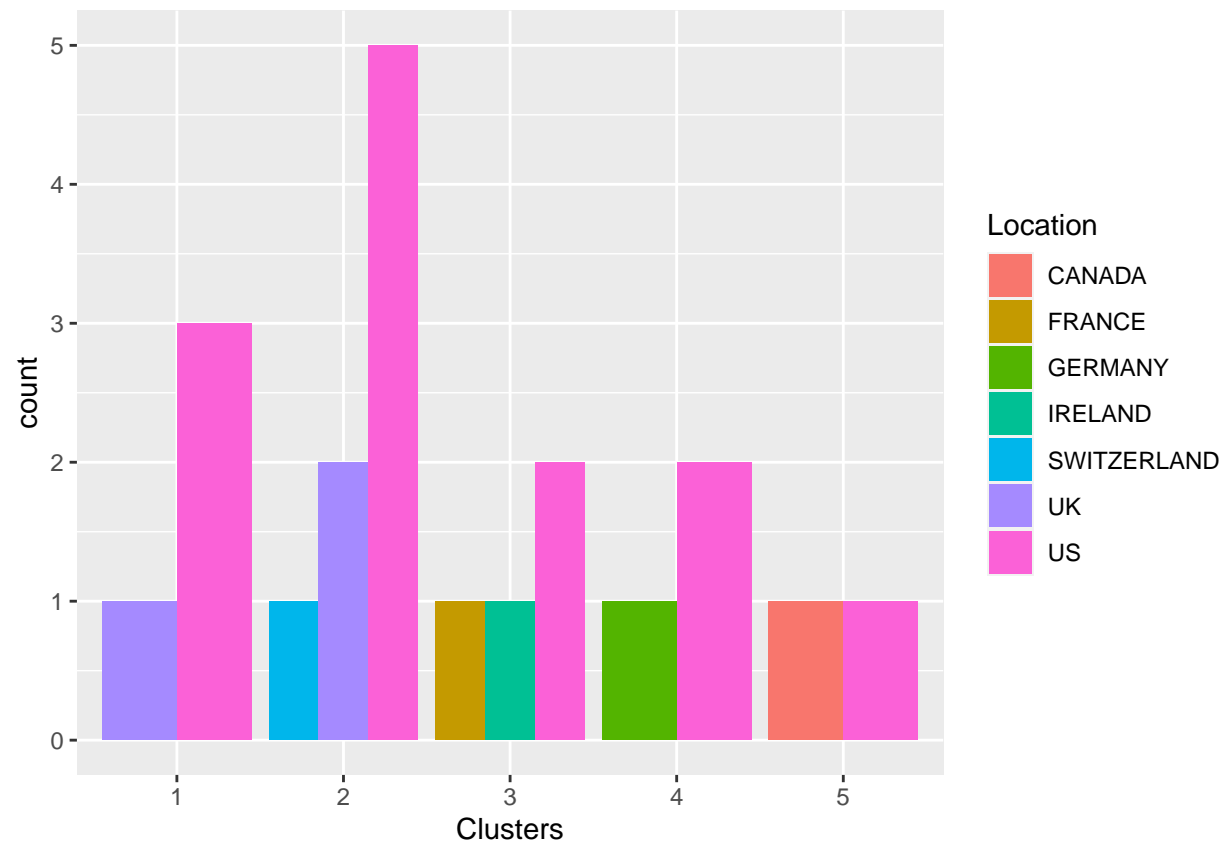
BAY, PHA, and AGN make up Cluster 4, which has the lowest leverage and asset turnover ratios and the highest PE ratio.

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

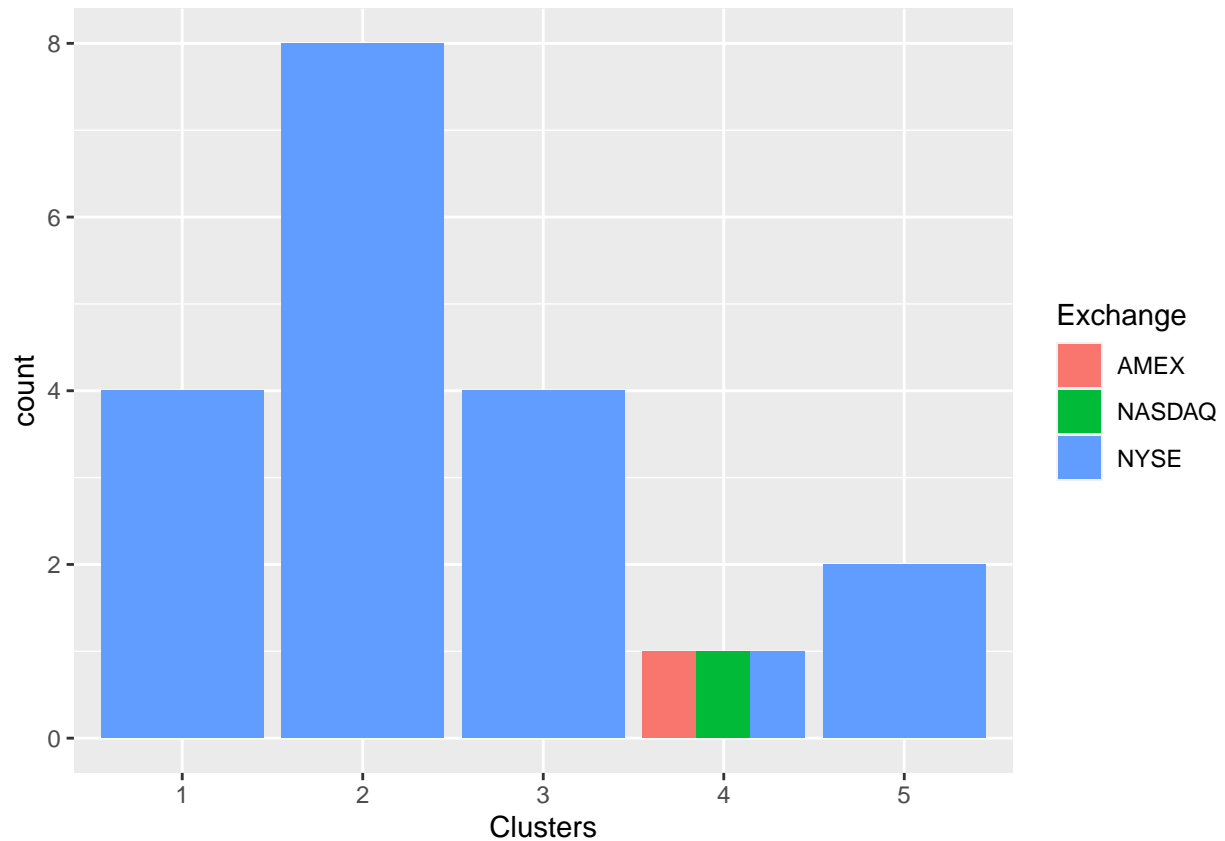
```
clustdata <- Pharmac[12:14] %>% mutate(Clusters=k5$cluster)
ggplot(clustdata, mapping = aes(factor(Clusters), fill = Median_Recommendation))+geom_bar(position='dodge')
```



```
ggplot(clustdata, mapping = aes(factor(Clusters), fill = Location)) + geom_bar(position = 'dodge') + labs(x = 'Clusters', y = 'count')
```



```
ggplot(clustdata, mapping = aes(factor(Clusters), fill = Exchange)) + geom_bar(position = 'dodge') + labs(x = 'Clusters', y = 'count')
```



The variable and clusters there is a trend in the median recommendations. There doesn't seem to be any discernable pattern among the clusters, locations, or exchanges other than the fact that the majority of the clusters/companies are listed on the NYSE and situated in the United States

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

Cluster 1: Top Buying

Cluster 2: Significant Risk

Cluster 3: Attempt it

Cluster 4: Very Dangerous or Runaway

Cluster 5: A Perfect Asset