

Assignment 4

Madhukar Heerekar

2022-03-21

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.1.3
```

```
## Loading required package: ggplot2
```

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

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

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6      v dplyr   1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## v purrr   0.3.4
```

```
## Warning: package 'readr' was built under R version 4.1.3
```

```
## Warning: package 'forcats' was built under R version 4.1.3
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(ISLR)
pharmacy <- read.csv("C://Users//heere//OneDrive//Desktop/FML ASSIGNMENT 4//pharmacy.csv")
View(pharmacy)
#TASK1
#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.
x <- na.omit(pharmacy)
summary(x)
```

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

```
row.names(x) <- x[,1]
Pharmacy1 <- x[,3:11]
head(Pharmacy1)
```

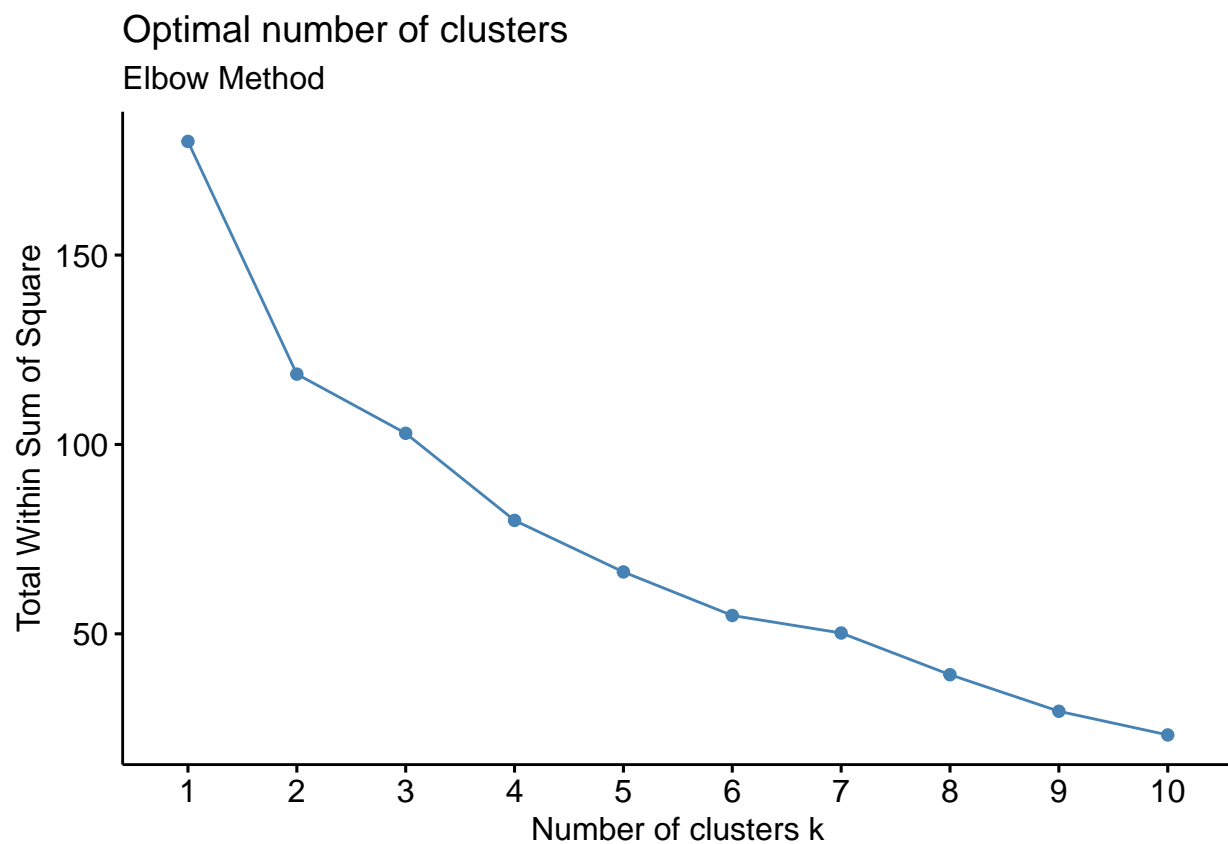
```
##      Market_Cap Beta PE_Ratio ROE ROA Asset_Turnover Leverage Rev_Growth
## ABT      68.44 0.32      24.7 26.4 11.8      0.7      0.42      7.54
## AGN      7.58 0.41      82.5 12.9 5.5      0.9      0.60      9.16
## AHM      6.30 0.46      20.7 14.9 7.8      0.9      0.27      7.05
## AZN      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
```

```
Pharmacy2 <- scale(Pharmacy1)
head(Pharmacy2)
```

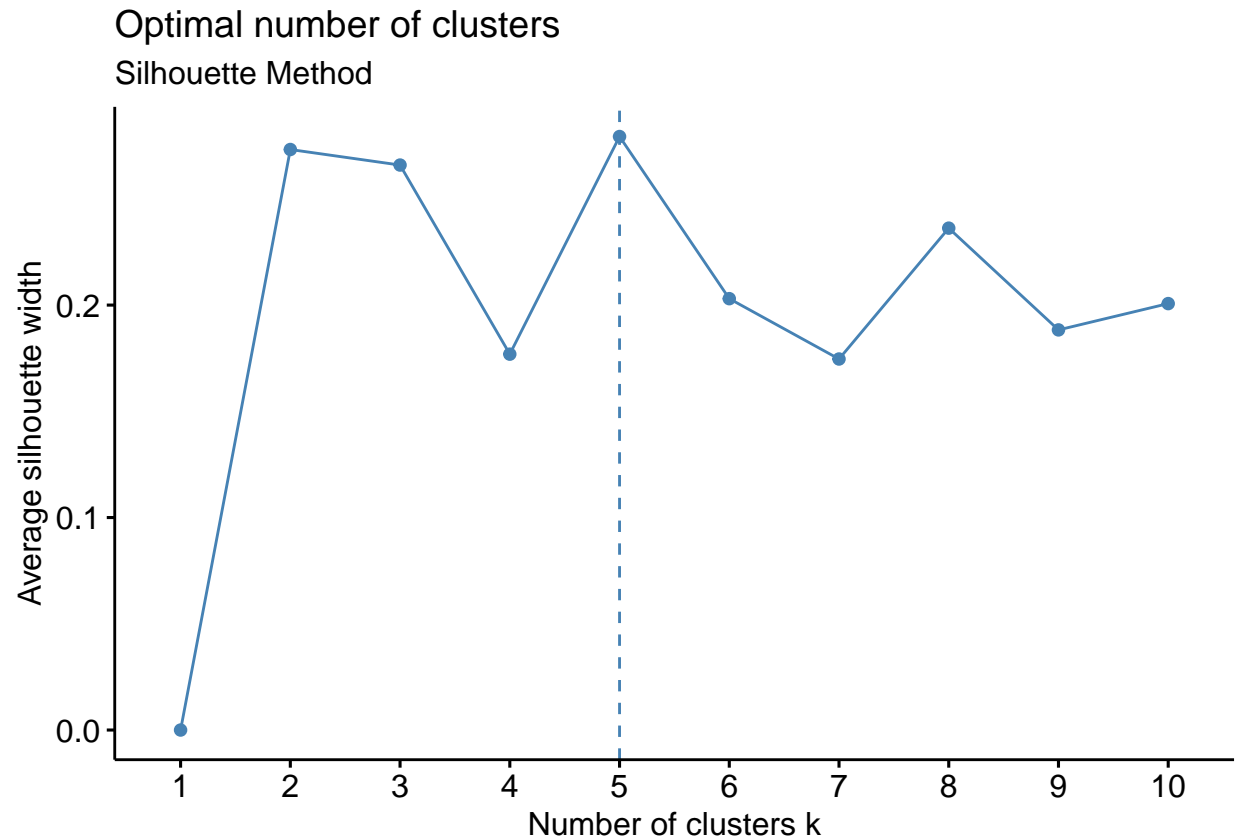
```
##      Market_Cap      Beta      PE_Ratio      ROE      ROA      Asset_Turnover
## ABT 0.1840960 -0.80125356 -0.04671323 0.04009035 0.2416121 0.0000000
```

```
## 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
```

```
fviz_nbclust(Pharmacy2, kmeans, method = "wss") + labs(subtitle = "Elbow Method")
```



```
fviz_nbclust(Pharmacy2, kmeans, method = "silhouette") + labs(subtitle = "Silhouette Method")
```



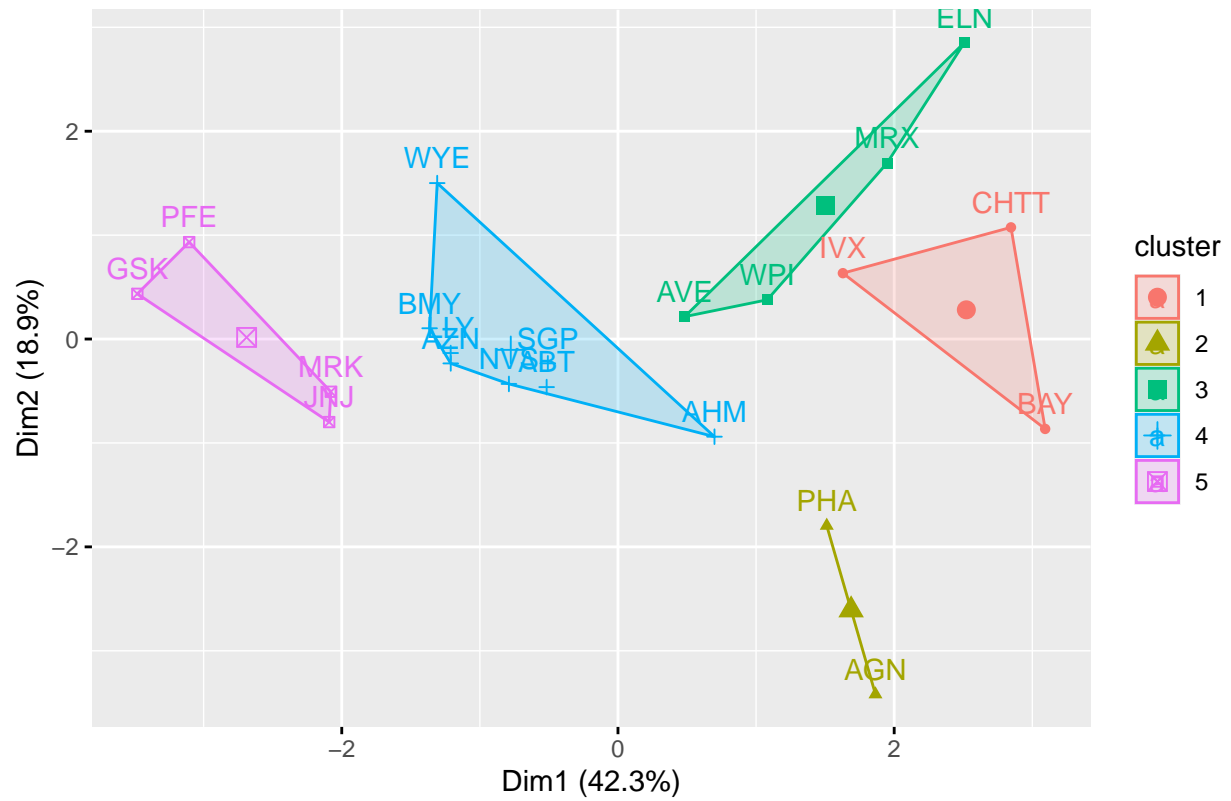
```
set.seed(64060)
```

```
k5 <- kmeans(Pharmacy2, centers = 5, nstart = 25)
k5$centers
```

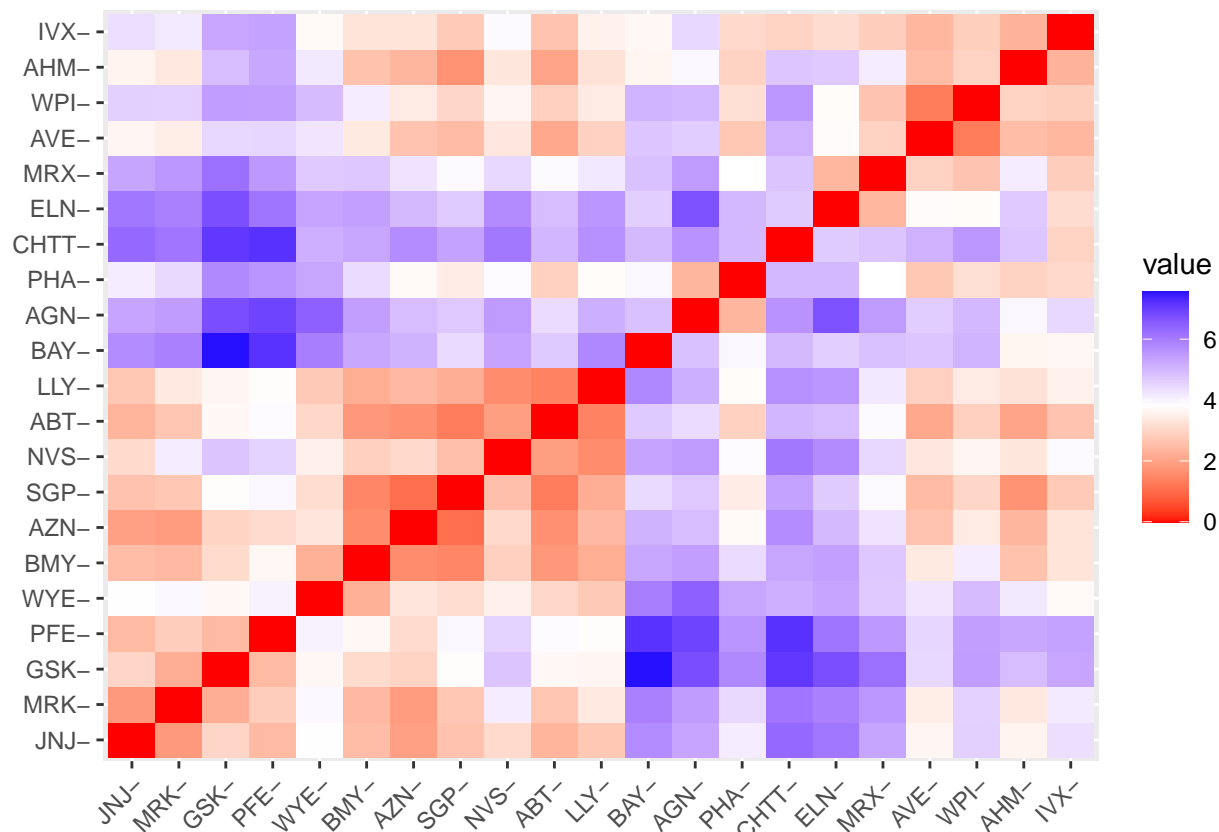
```
##      Market_Cap      Beta    PE_Ratio      ROE      ROA Asset_Turnover
## 1 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478   -0.4612656
## 2 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951    0.2306328
## 3 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428   -1.2684804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

```
fviz_cluster(k5, data = Pharmacy2)
```

Cluster plot



```
distance <- dist(Pharmacy2, method = "euclidean")
fviz_dist(distance)
```



```
fit <- kmeans(Pharmacy2, 5)
aggregate(Pharmacy2, by=list(fit$cluster), FUN=mean)
```

```
##   Group.1 Market_Cap      Beta  PE_Ratio      ROE      ROA
## 1      1  1.69558112 -0.1780563 -0.1984582  1.2349879  1.3503431
## 2      2 -0.66114002 -0.7233539 -0.3512251 -0.6736441 -0.5915022
## 3      3 -0.96247577  1.1949250 -0.3639982 -0.5200697 -0.9610792
## 4      4 -0.52462814  0.4451409  1.8498439 -1.0404550 -1.1865838
## 5      5  0.08926902 -0.4618336 -0.3208615  0.3260892  0.5396003
##   Asset_Turnover  Leverage Rev_Growth Net_Profit_Margin
## 1  1.153164e+00 -0.4680782  0.4671788      0.5912425
## 2 -1.537552e-01 -0.4040831  0.6917224     -0.4005718
## 3 -1.153164e+00  1.4773718  0.7120120     -0.3688236
## 4  1.480297e-16 -0.3443544 -0.5769454     -1.6095439
## 5  6.589509e-02 -0.2559803 -0.7230135      0.7343816
```

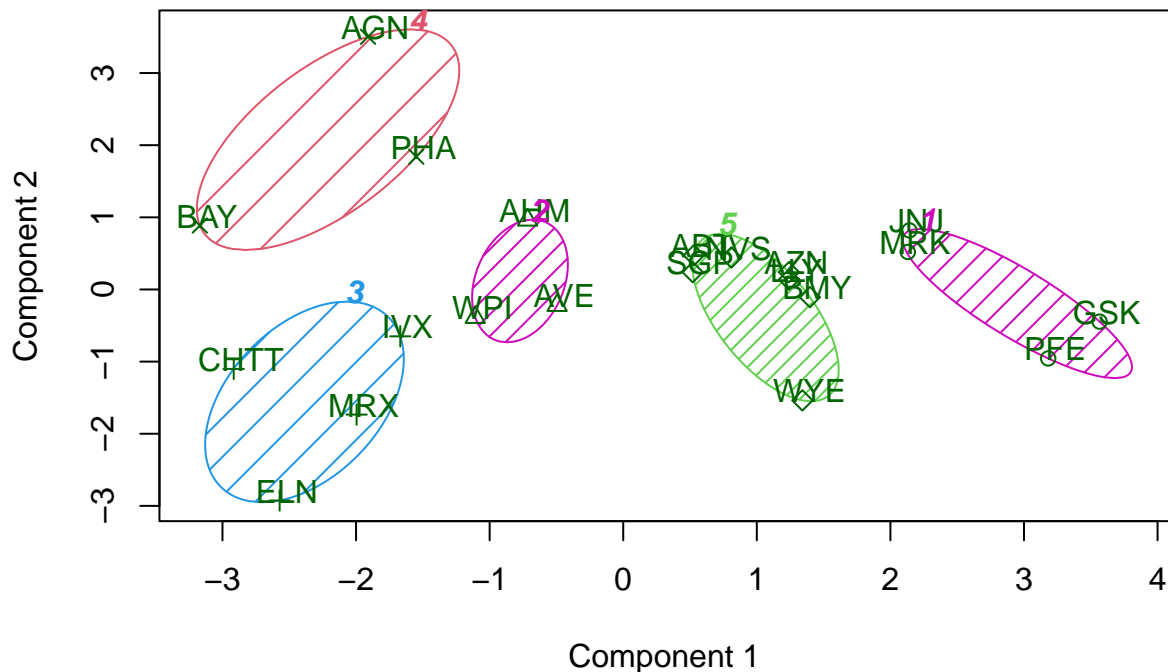
```
pharmacy3 <- data.frame(Pharmacy2, fit$cluster)
pharmacy3
```

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

```
library(cluster)
clusplot(Pharmacy2, fit$cluster, color = TRUE, shade = TRUE,
         labels = 2, lines = 0)
```

CLUSPLOT(Pharmacy2)



#Task2

#B) Interpret the clusters with respect to numerical variables used in formatting the clusters.

#cluster 1 - JNJ,MRK,PFE,GSK-- They have the largest market capitalization, and the corporations are ef

#cluster 2 - AHM,WPI,AVE--They have the lowest asset turnover and the lowest beta, which means that the

#cluster 3 - CHTT,MRX,LVX,ELN--They have the smallest market capitalization, do not rely on debt to sup

#cluster 4 - AGN,BAY,RHA--- These have the largest cost-to-income ratio, making them less profitable.

#Their Return on Equity is also less than one, indicating that investing in these stocks will not be as

#cluster 5 - ABT,SGP,NVS,AZN,BMY,WYE--Their asset turnover is the largest, their revenue growth is the

#Task3

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

#Each cluster was manually filtered to detect patterns in media suggestions, location, and exchange.

#cluster1--The stocks are moderate in character, which means they are neither weak nor have had strong

#cluster2--In terms of geographical distribution, the stocks are well-balanced. Their principles are so

#cluster3--Because of their financial soundness, they are moderately recommended despite their high lev

#cluster4--These are the stocks that, according to the media, should be held because they will eventual

#cluster5--The cluster contains companies with a high net profit margin that are advised to be kept for

#Task 4: Provide an appropriate name for each cluster using any or all of the variables in the dataset.
#Cluster1-- Profitable clusture
#cluster2-- Gold mine clusture
#cluster3-- primary clusture
#clusture4-- Bear clusture
#clusture5-- continuing clusture