

# FML Assignment 4

2023-11-11

```
library(ggplot2)
library(factoextra)

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

library(flexclust)

## Loading required package: grid

## Loading required package: lattice

## Loading required package: modeltools

## Loading required package: stats4

library(cluster)
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
## ✓lubridate 1.9.3    ✓ tibble    3.2.1
## ✓ purrr    1.0.2    ✓ tidyverse 1.3.0

## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()   masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors

library(dplyr)
```

**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.**

```
ps <- read.csv("C:/Users/sriha/Desktop/Assignments/Fundamentals of Machine
Learning/Datasets/Pharmaceuticals.csv")
```

```
pharmaceuticals <- na.omit(ps)
```

pharmaceuticals

|       | Symbol | Name                                   | Market_Cap     | Beta       | PE_Ratio          | ROE  |
|-------|--------|--|----------------|------------|-------------------|------|
| ## 1  | ABT    | Abbott Laboratories                    | 68.44          | 0.32       | 24.7              | 26.4 |
| 11.8  |        | Allergan, Inc.                         | 7.58           | 0.41       | 82.5              | 12.9 |
| ## 2  | AGN    | Amersham plc                           | 6.30           | 0.46       | 20.7              | 14.9 |
| 5.5   |        | AstraZeneca PLC                        | 67.63          | 0.52       | 21.5              | 27.4 |
| ## 3  | AHM    | Aventis                                | 47.16          | 0.32       | 20.1              | 21.8 |
| 7.8   |        | Bayer AG                               | 16.90          | 1.11       | 27.9              | 3.9  |
| ## 4  | AZN    | Bristol-Myers Squibb Company           | 51.33          | 0.50       | 13.9              | 34.8 |
| 15.4  |        | Chattem, Inc                           | 0.41           | 0.85       | 26.0              | 24.1 |
| ## 5  | AVE    | Elan Corporation, plc                  | 0.78           | 1.08       | 3.6               | 15.1 |
| 7.5   |        | Eli Lilly and Company                  | 73.84          | 0.18       | 27.9              | 31.0 |
| ## 6  | BAY    | GlaxoSmithKline plc                    | 122.11         | 0.35       | 18.0              | 62.9 |
| 1.4   |        | IVAX Corporation                       | 2.60           | 0.65       | 19.9              | 21.4 |
| ## 7  | BMY    | Johnson & Johnson                      | 173.93         | 0.46       | 28.4              | 28.6 |
| 15.1  |        | MRX Medicis Pharmaceutical Corporation | 1.20           | 0.75       | 28.6              | 11.2 |
| ## 8  | CHTT   | Merck & Co., Inc.                      | 132.56         | 0.46       | 18.9              | 40.6 |
| 4.3   |        | Novartis AG                            | 96.65          | 0.19       | 21.6              | 17.9 |
| ## 9  | ELN    | Pfizer Inc                             | 199.47         | 0.65       | 23.6              | 45.6 |
| 5.1   |        | Pharmacia Corporation                  | 56.24          | 0.40       | 56.5              | 13.5 |
| ## 10 | LLY    | Schering-Plough Corporation            | 34.10          | 0.51       | 18.9              | 22.6 |
| 13.5  |        | Watson Pharmaceuticals, Inc.           | 3.26           | 0.24       | 18.4              | 10.2 |
| ## 11 | GSK    | Wyeth                                  | 48.19          | 0.63       | 13.1              | 54.9 |
| 20.3  |        | Asset_Turnover                         | Leverage       | Rev_Growth | Net_Profit_Margin |      |
| ## 12 | IVX    | Median                                 | Recommendation |            |                   |      |
| 6.8   |        |  |                |            |                   |      |
| ## 13 | JNJ    |  |                |            |                   |      |
| 16.3  |        |  |                |            |                   |      |
| ## 14 | MRX    |  |                |            |                   |      |
| 5.4   |        |  |                |            |                   |      |
| ## 15 | MRK    |  |                |            |                   |      |
| 15.0  |        |  |                |            |                   |      |
| ## 16 | NVS    |  |                |            |                   |      |
| 11.2  |        |  |                |            |                   |      |
| ## 17 | PFE    |  |                |            |                   |      |
| 19.2  |        |  |                |            |                   |      |
| ## 18 | PHA    |  |                |            |                   |      |
| 5.7   |        |  |                |            |                   |      |
| ## 19 | SGP    |  |                |            |                   |      |
| 13.3  |        |  |                |            |                   |      |
| ## 20 | WPI    |  |                |            |                   |      |
| 6.8   |        |  |                |            |                   |      |
| ## 21 | WYE    |  |                |            |                   |      |
| 13.4  |        |  |                |            |                   |      |

|       |          |          |       |      |          |
|-------|----------|----------|-------|------|----------|
| ## 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

```

## Considering Quantitative data

```

row.names(pharmaceuticals) <- pharmaceuticals[,1]

pharma <- pharmaceuticals[,3:11]

t(t(names(pharma)))

##      [,1]
## [1,] "Market_Cap"
## [2,] "Beta"
## [3,] "PE_Ratio"
## [4,] "ROE"
## [5,] "ROA"
## [6,] "Asset_Turnover"
## [7,] "Leverage"
## [8,] "Rev_Growth"
## [9,] "Net_Profit_Margin"

head(pharma)

##   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

```

## Normalising the Data

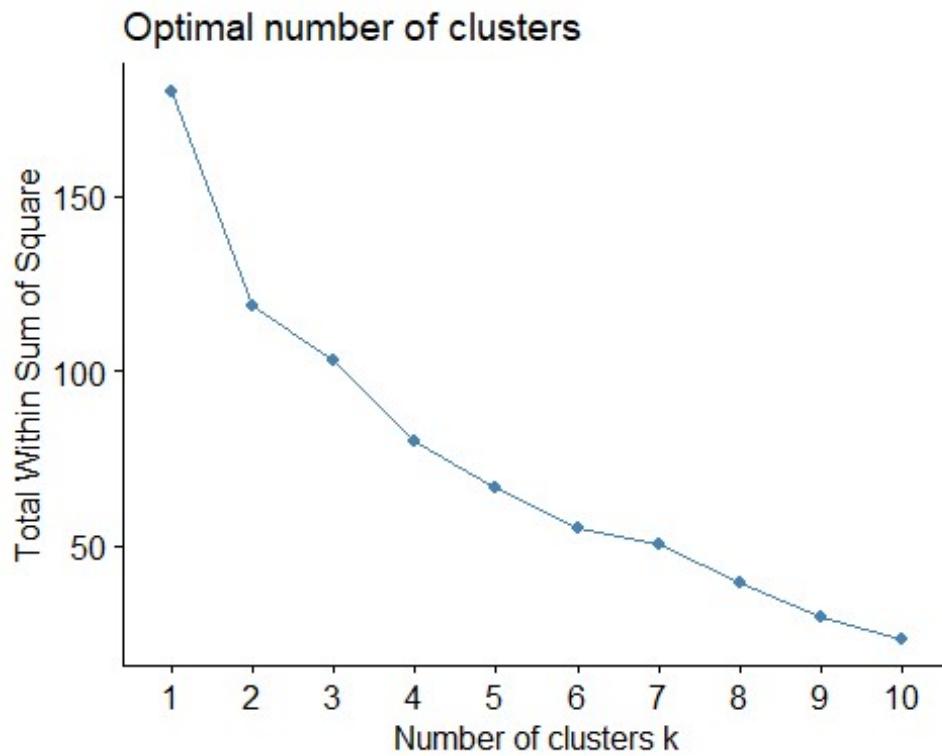
```
pharma1<-scale(pharma)

head(pharma1)

##      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 for determining number of clusters

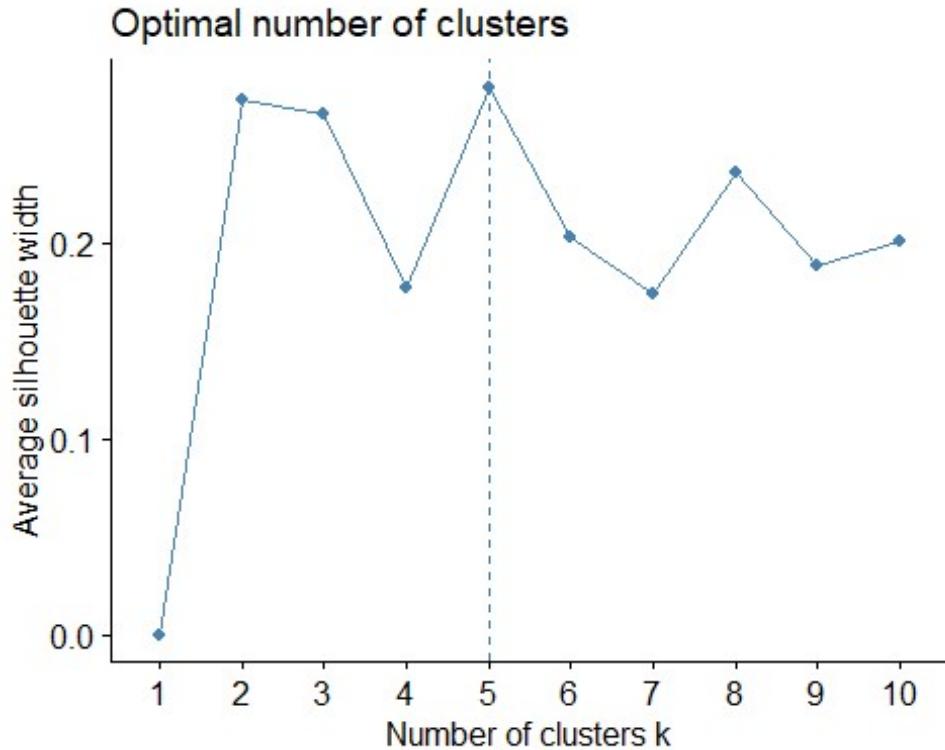
```
fviz_nbclust(pharma1, kmeans, method = "wss")
```



```
# From the graph we cannot determine K value, Hence using silhouette method
```

```
#Silhouette method for determining number of clusters
```

```
fviz_nbclust(pharma1, kmeans, method = "silhouette")
```



## K means clustering

```
set.seed(150)
```

```
k5 <- kmeans(pharma1, centers = 5, nstart = 25)
```

#Visualize the output

```
k5$centers #centroids
```

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

# Visualizing the clusters

```
fviz_cluster(k5,data = pharma1)
```

Cluster plot



#Cluster 2 - WPI, MRX, ELN, AVE (low PE\_Ratio, high ROE, low ROA, low Net\_Profit\_Margin, high Rev\_Growth)

#Cluster 3 - AHM, SGP, WYE, BMY, AZN, ABT, NVS, LLY ( low Market\_Cap, low Beta, low PE\_Ratio, high Leverage, high Rev\_Growth.)

#Cluster 4 - JNJ, MRK, PFE, GSK (High Market\_Cap, ROE, ROA, Asset\_Turnover Ratio and low Beta/PE Ratio)

#Cluster 5 - BAY, CHTT, IVX (low Rev\_Growth, high Beta and leverage, low Net\_Profit\_Margin)

```
print(k5)

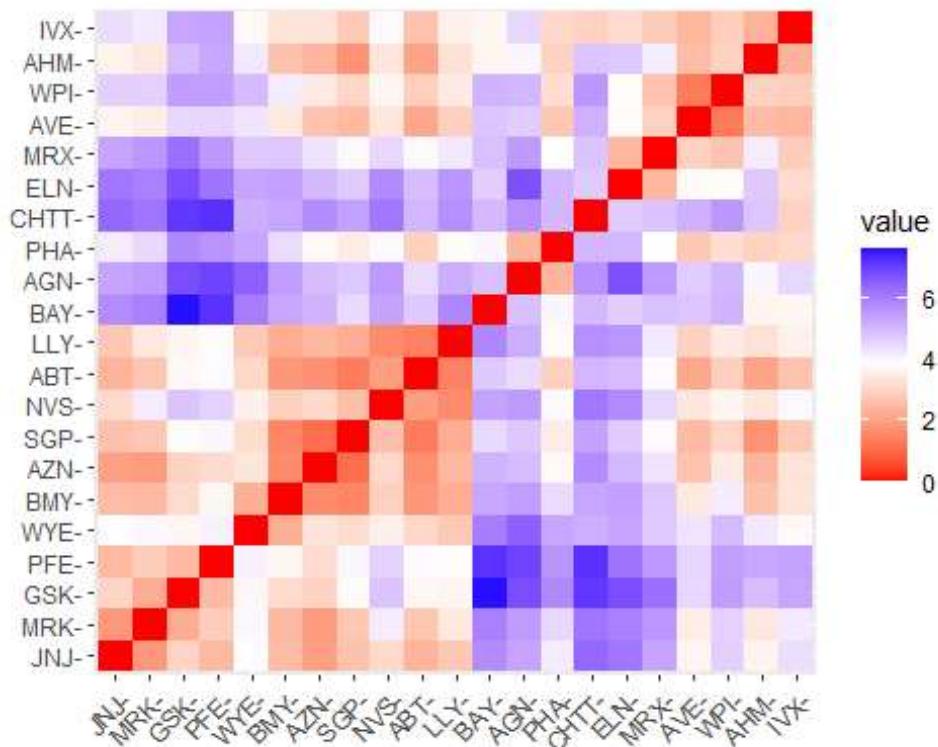
## K-means clustering with 5 clusters of sizes 2, 4, 8, 4, 3
##
## Cluster means:
##   Market_Cap      Beta    PE_Ratio       ROE       ROA Asset_Turnover
## 1 -0.43925134 -0.4701800  2.7002464 -0.8349525 -0.9234951     0.2306328
## 2 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428    -1.2684804
## 3 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915     0.1729746
## 4  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431     1.1531640
## 5 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478    -0.4612656
##   Leverage Rev_Growth Net_Profit_Margin
## 1 -0.14170336 -0.1168459      -1.416514761
## 2  0.06308085  1.5180158      -0.006893899
```

```

## 3 -0.27449312 -0.7041516      0.556954446
## 4 -0.46807818  0.4671788      0.591242521
## 5  1.36644699 -0.6912914     -1.320000179
##
## Clustering vector:
## ABT   AGN   AHM   AZN   AVE   BAY   BMY   CHTT   ELN   LLY   GSK   IVX   JNJ   MRX   MRK
NVS
##    3     1     3     3     2     5     3     5     2     3     4     5     4     2     4
3
## PFE   PHA   SGP   WPI   WYE
##    4     1     3     2     3
##
## Within cluster sum of squares by cluster:
## [1] 2.803505 12.791257 21.879320  9.284424 15.595925
## (between_SS / total_SS =  65.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"       "totss"        "withinss"
## "tot.withinss"
## [6] "betweenss"     "size"          "iter"         "ifault"
dis <- dist(pharma1, method = "euclidean")

fviz_dist(dis)

```



```

# From the data the center is defined after 25 restarts and there are 5
clusters.

fit <- kmeans(pharma1,5)

#Finding the mean value of variables

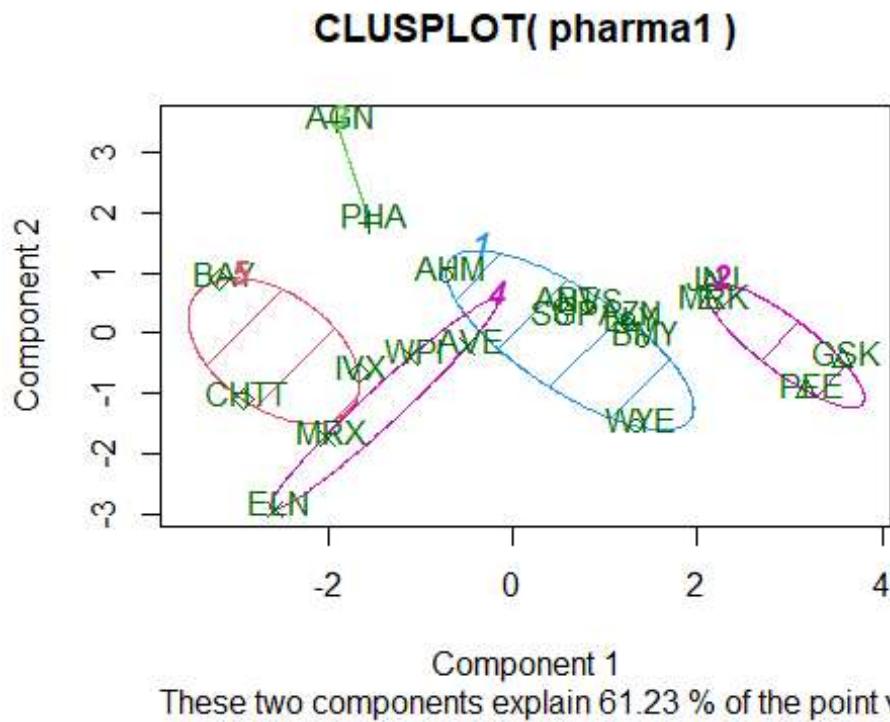
aggregate(pharma1,by=list(fit$cluster),FUN=mean)

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

# Cluster Plot

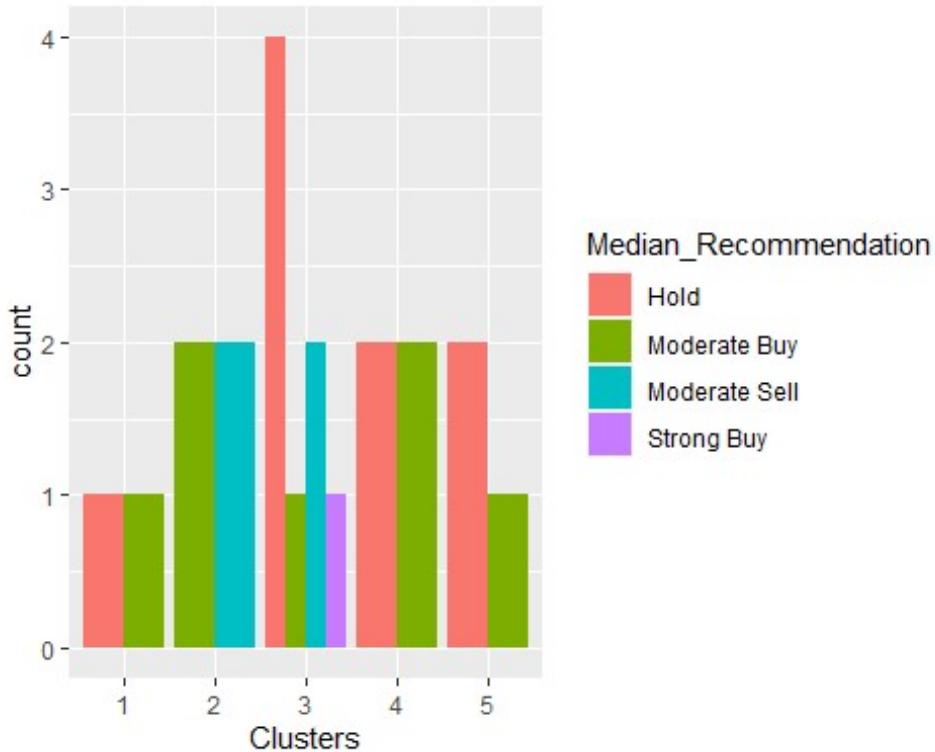
clusplot(pharma1,fit$cluster,color =
TRUE,shade = TRUE,labels = 2,lines = 0)

```



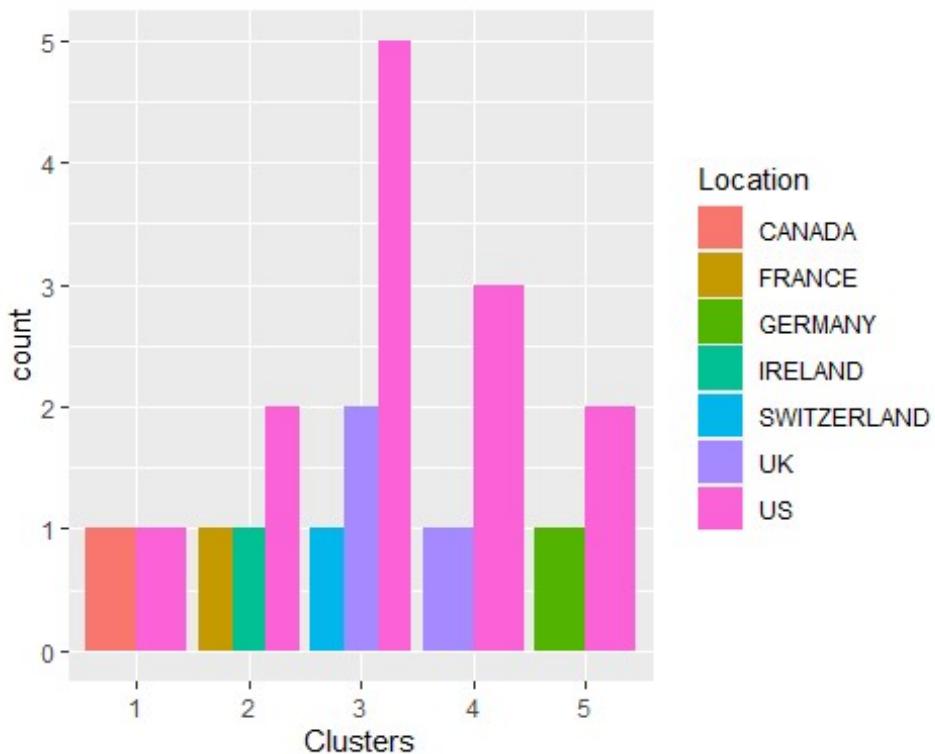
**2 Interpret the clusters with respect to the numerical variables used in forming the clusters. Is there a pattern in the clusters with respect to the numerical variables?**

```
clust_data <- ps[12:14]
clust_data$Clusters <- k5$cluster
ggplot(clust_data, aes(factor(Clusters), fill= Median_Recommendation))+
geom_bar(position='dodge')+labs(x='Clusters')
```

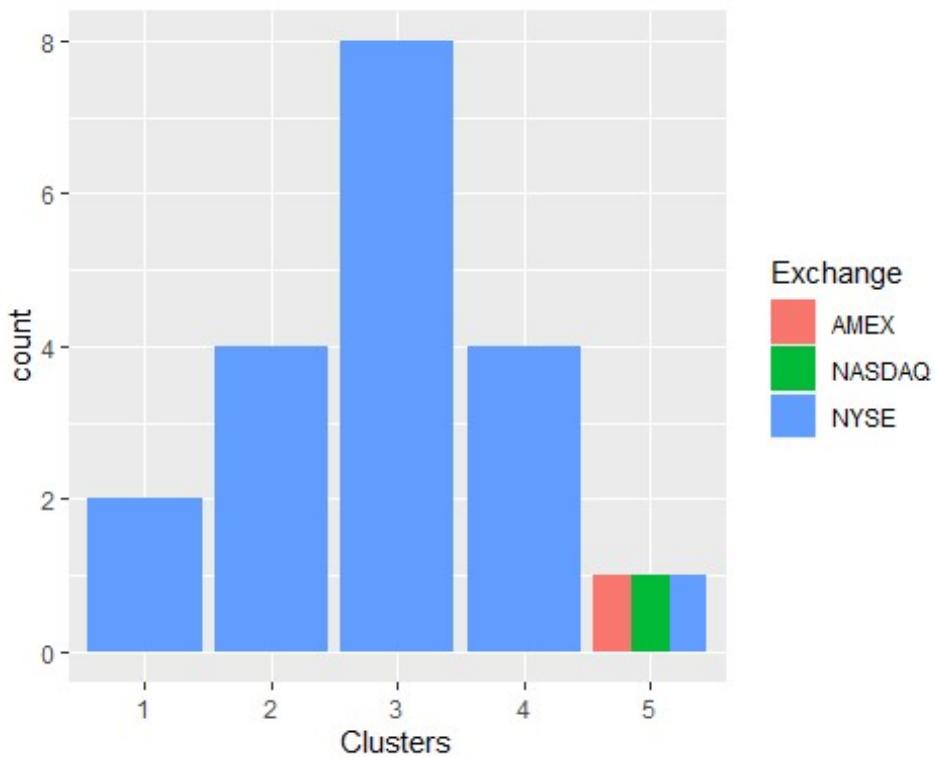


# From the above graph the cluster 1 is having low profit and it has moderate leverage. The graph determines same level of moderate buy and Hold. # In cluster 2, it is having high revenue growth there might be a potential growth in these companies because it has equal level of moderate buy and sell # In cluster 3, There is high level of Hold and also the strong large amount of profit which will attract the investors to buy more of this cluster. # In cluster 4, There is high amount of market capitalization and also high profit. It represents high potential of the company and has equal level of hold and buy. # In cluster 5, The graph represents high debt that leads to high leverage.

```
ggplot(clust_data, aes(factor(Clusters), fill= Location))+
geom_bar(position='dodge')+labs(x='Clusters')
```



```
ggplot(clust_data, aes(factor(Clusters), fill= Exchange))+
  geom_bar(position='dodge')+labs(x='Clusters')
```



```
# 3. Provide an appropriate name for each cluster using any or all of the  
variables in the dataset.  
  
# Cluster 1: Significant Risk  
  
# Cluster 2: Potential Growth  
  
# Cluster 3: Perfect Asset  
  
# Cluster 4: Long term investment  
  
# Cluster 5: Short term investment
```