ASSIGNMENT 4 FML

Shivani Haridas Pitla

2022-11-06

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.2

## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.2.2

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

```
library(ISLR)
library(gridExtra)
```

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

```
library(cluster)
library(dplyr)
```

```
PHARMACEUTICALS=read.csv("C:/Users/shiva/Downloads/Pharmaceuticals.csv")
```

```
#a. Use only the numerical variables (1 to 9) to cluster the 21 firms. Justify the various choices made

#choosing the numerical variables and removing the Null Values from the dataset.
colSums(is.na(PHARMACEUTICALS))
```

```
##                  Symbol                   Name            Market_Cap
##                       0                      0                     0
##                    Beta               PE_Ratio                   ROE
##                       0                      0                     0
##                     ROA         Asset_Turnover              Leverage
##                       0                      0                     0
##              Rev_Growth      Net_Profit_Margin Median_Recommendation
##                       0                      0                     0
##                Location               Exchange
##                       0                      0
```

```
row.names(PHARMACEUTICALS)<- PHARMACEUTICALS[,1]
PHARMACEUTICALS1<- PHARMACEUTICALS[, 3:11]
head(PHARMACEUTICALS1)
```

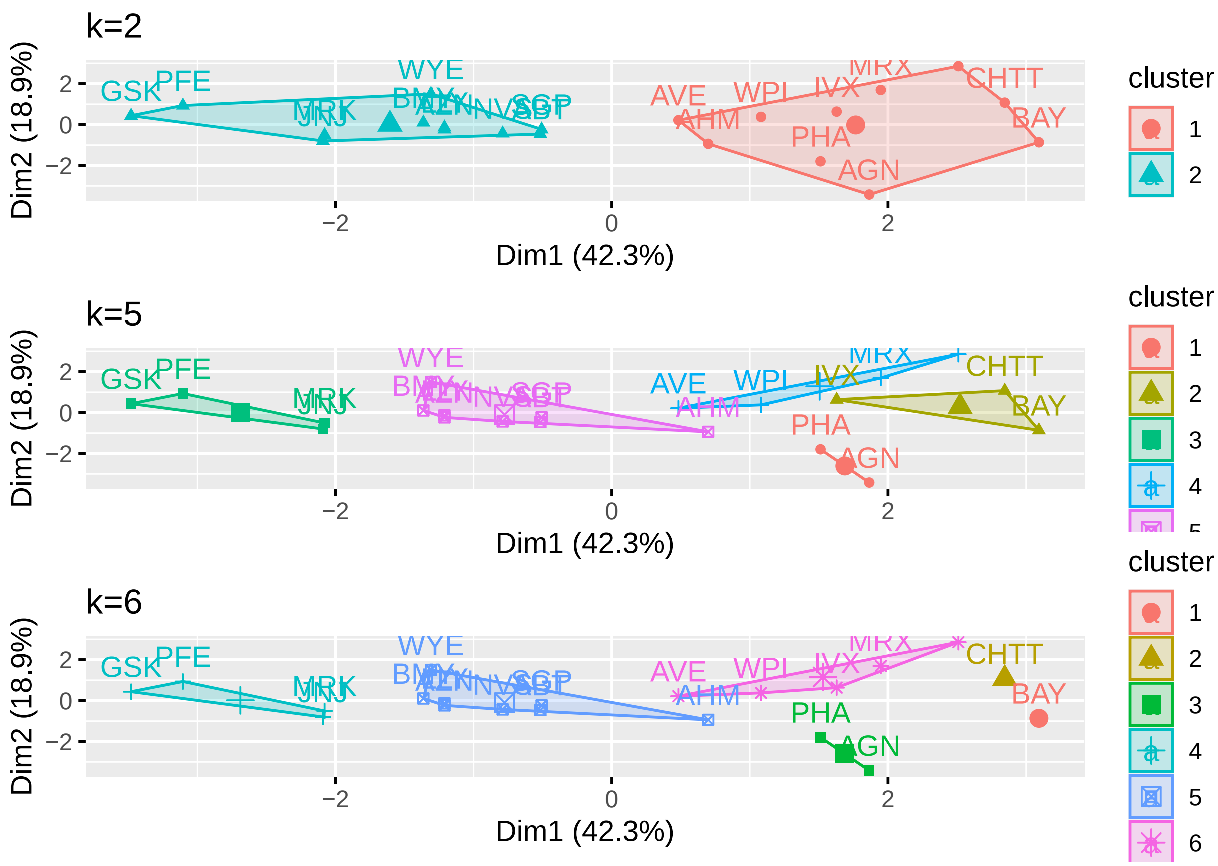
```
##     Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## ABT      68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## AGN       7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## AHM       6.30 0.46     20.7 14.9  7.8            0.9     0.27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

```
# Scaling and Normalisation the dataset(PARMACEUTICALS).
PHARMACEUTICALS_SCALE <- scale(PHARMACEUTICALS1)
head(PHARMACEUTICALS_SCALE)
```

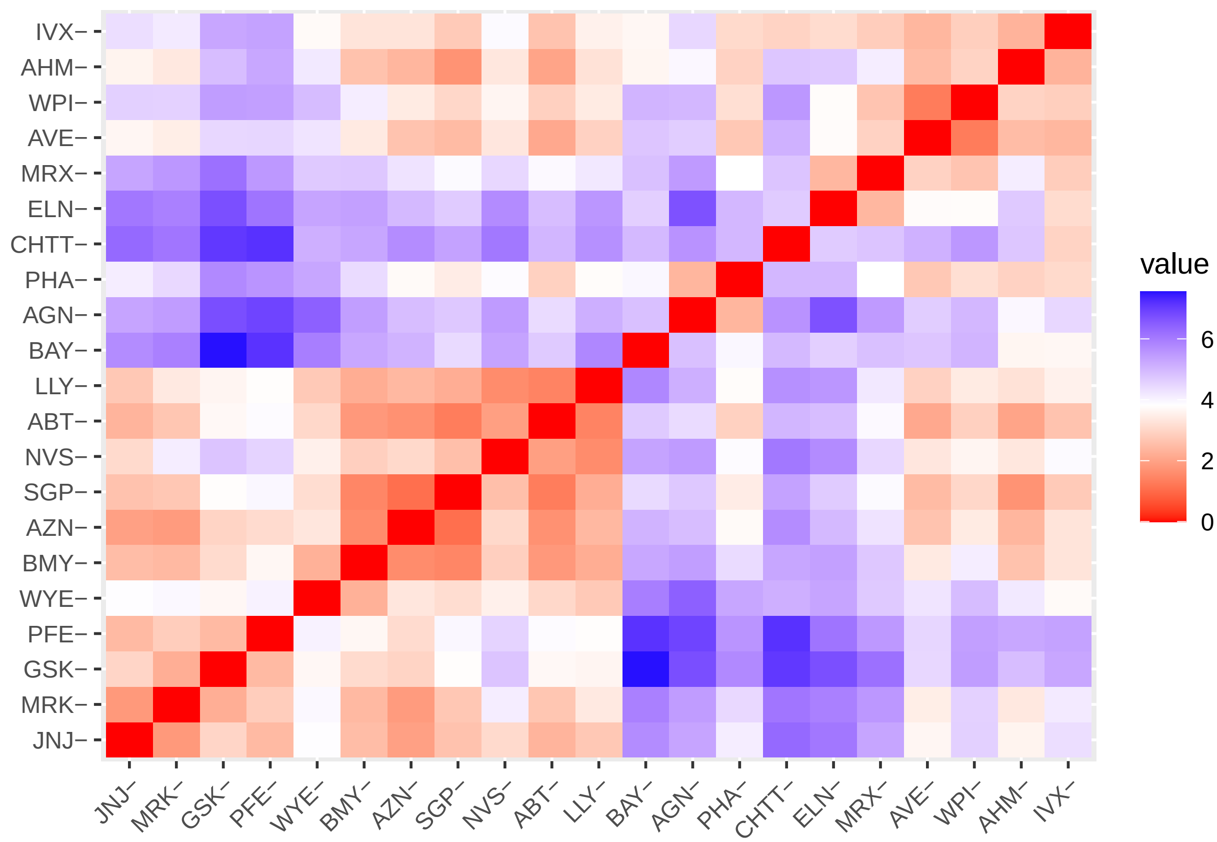
```
##     Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## ABT  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## AGN -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## AHM -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      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
```

```
# Using several values of K, computing K-means clustering for various centers, and comparing the result
kmeans.1 <- kmeans(PHARMACEUTICALS_SCALE, centers = 2, nstart = 25)
kmeans.2<- kmeans(PHARMACEUTICALS_SCALE, centers = 5, nstart = 25)
```

```
kmeans.3<- kmeans(PHARMACEUTICALS_SCALE, centers = 6, nstart = 25)
Plot.1<-fviz_cluster(kmeans.1, data = PHARMACEUTICALS_SCALE)+ggtitle("k=2")
plot.2<-fviz_cluster(kmeans.2, data = PHARMACEUTICALS_SCALE)+ggtitle("k=5")
plot.3<-fviz_cluster(kmeans.3, data = PHARMACEUTICALS_SCALE)+ggtitle("k=6")
grid.arrange(Plot.1,plot.2,plot.3, nrow = 3)
```




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




```
Aggregate.data<-kmeans(PHARMACEUTICALS_SCALE,5)
aggregate(PHARMACEUTICALS_SCALE, by=list(Aggregate.data$cluster), FUN=mean)
```

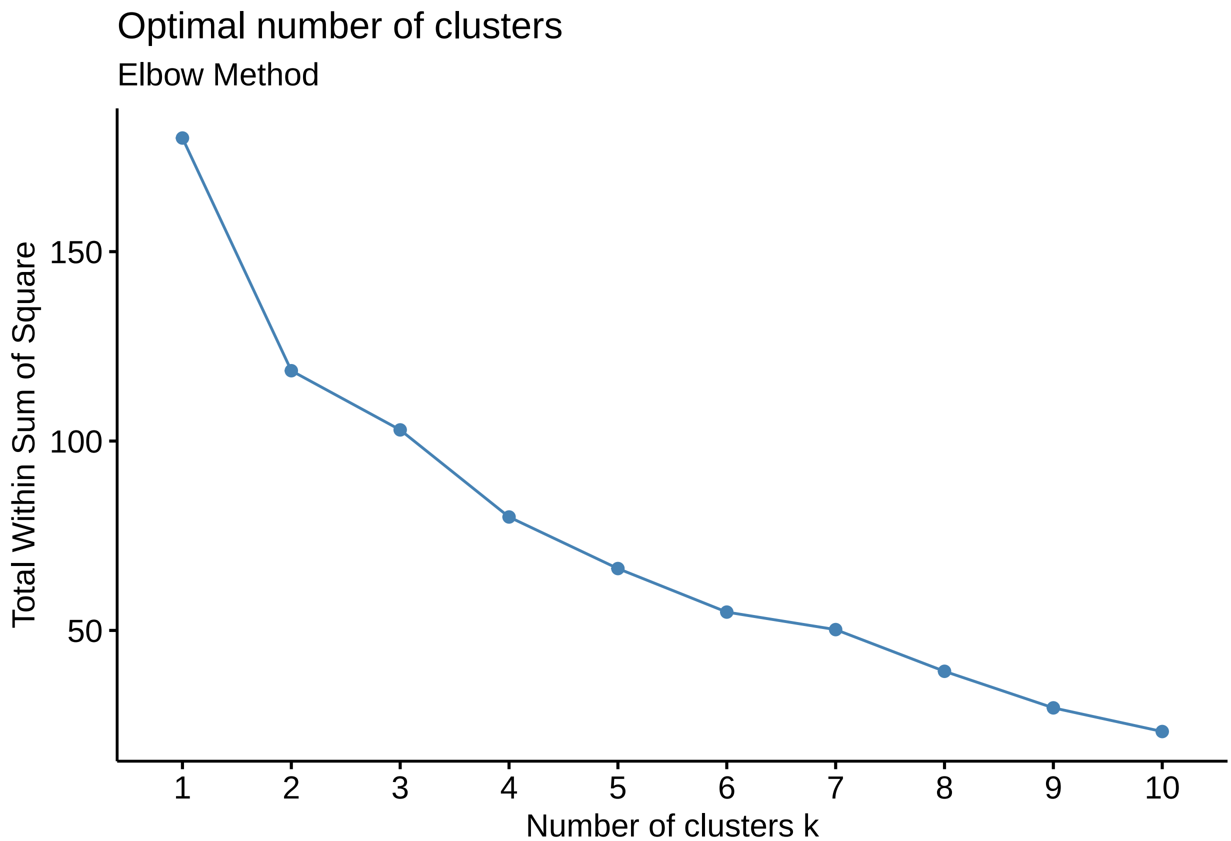
```
##   Group.1  Market_Cap       Beta    PE_Ratio        ROE        ROA
## 1       1 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428
## 2       2 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915
## 3       3 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
## 4       4  1.69558112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

```
aggregate_Data1 <- data.frame(PHARMACEUTICALS_SCALE, Aggregate.data$cluster)
aggregate_Data1
```

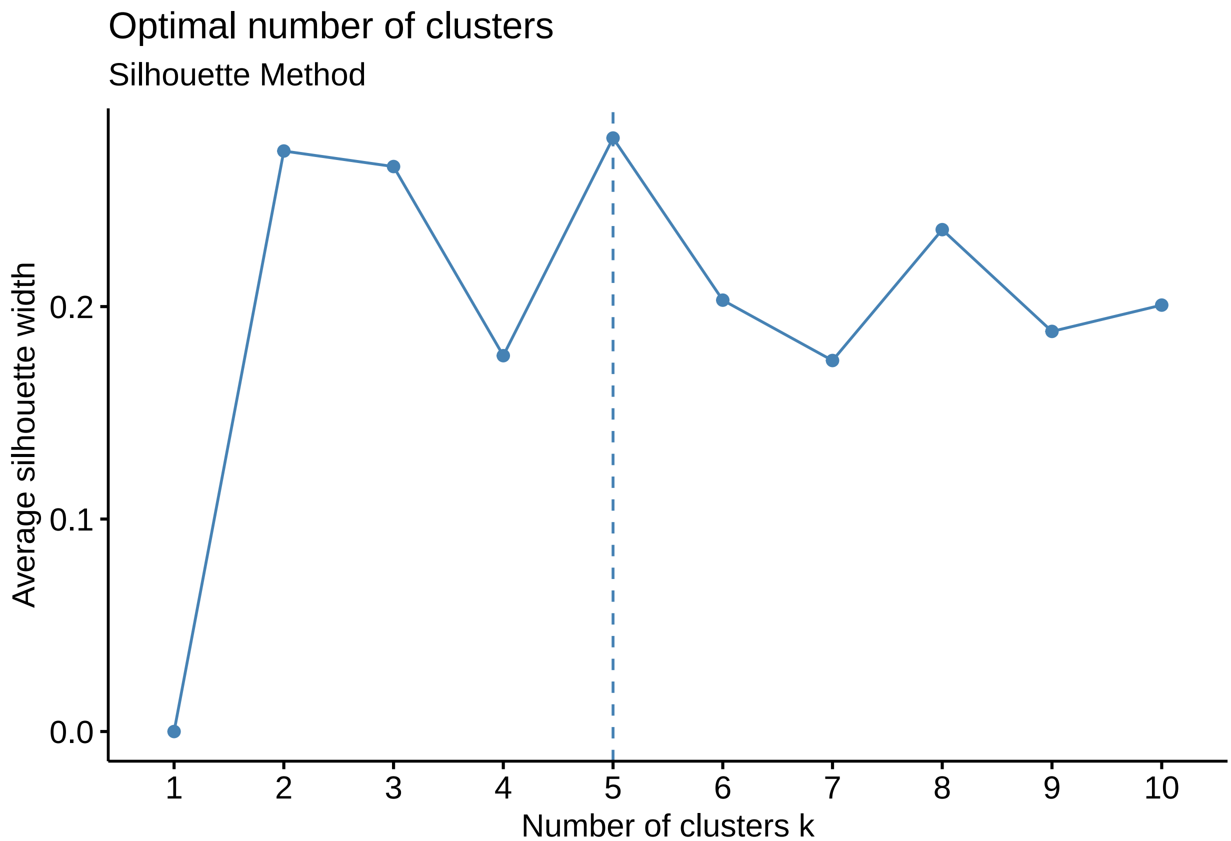
```
##      Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## ABT   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## AGN  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## AHM  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -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 Aggregate.data.cluster
## ABT  -0.21209793 -0.52776752        0.06168225                      2
## AGN   0.01828430 -0.38113909       -1.55366706                      5
## AHM  -0.40408312 -0.57211809       -0.68503583                      2
## AZN  -0.74965647  0.14744734        0.35122600                      2
## AVE  -0.31449003  1.21638667       -0.42597037                      1
## BAY  -0.74965647 -1.49714434       -1.99560225                      3
## BMY  -0.02011273 -0.96584257        0.74744375                      2
## CHTT  3.74279705 -0.63276071       -1.24888417                      3
## ELN   0.61983791  1.88617085       -0.36501379                      1
## LLY  -0.07130879 -0.64814764        1.17413980                      2
## GSK  -0.31449003  0.76926048        0.82363947                      4
## IVX   1.10620040  0.05603085       -0.71551412                      3
## 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                      2
## PFE  -0.54487226  1.10143723        1.44844440                      4
## PHA  -0.30169102  0.14744734       -1.27936246                      5
## SGP  -0.74965647 -0.43544591        0.29026942                      2
## WPI  -0.49367621  1.43089863       -0.09070919                      1
## WYE   0.68383297 -1.17763919        1.49416183                      2
```

```r
# estimating how many clusters there are
# To calculate the value of k, the data are scaled using the elbow method.
fviz_nbclust(PHARMACEUTICALS_SCALE, FUNcluster = kmeans, method = "wss") + labs(subtitle = "Elbow Metho
```




```
# The number of clusters is calculated by scaling the data using the silhouette method.
fviz_nbclust(PHARMACEUTICALS_SCALE,FUNcluster = kmeans,method = "silhouette")+labs(subtitle="Silhouette
```

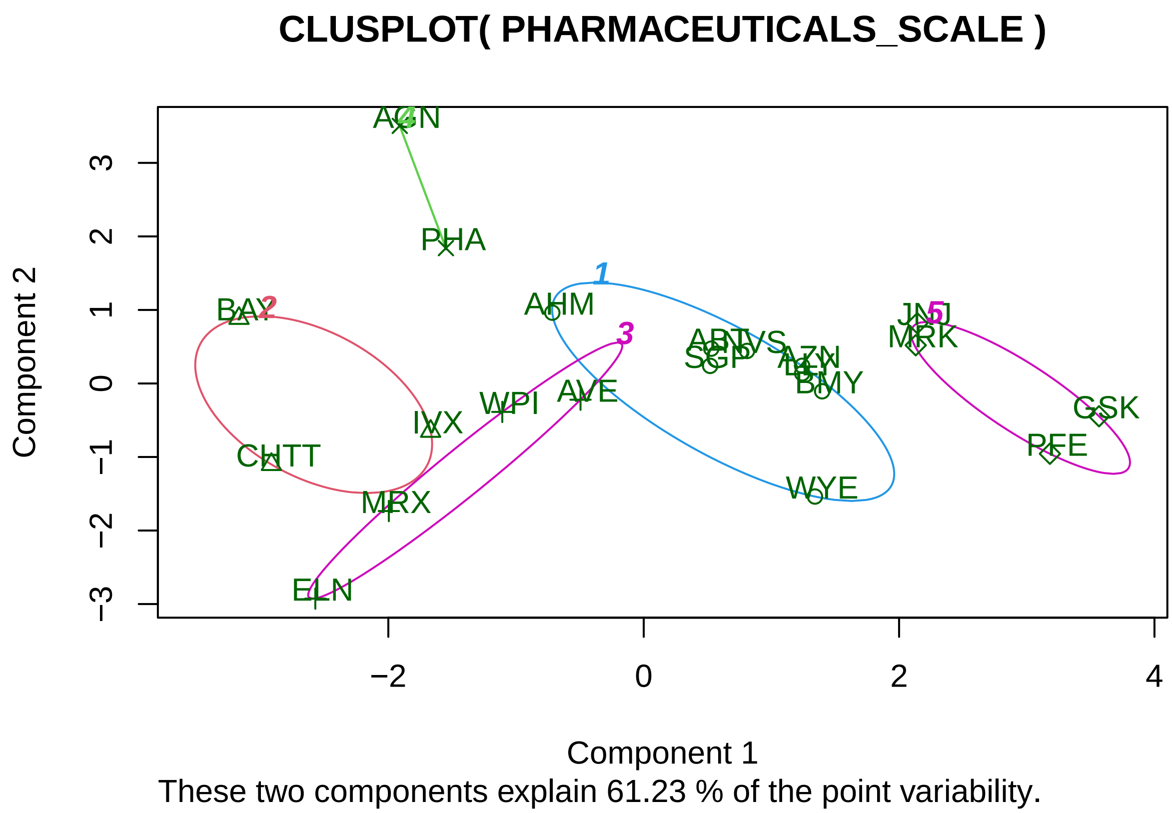



```
# Final analysis and Extracting results using 5 clusters and Visualize the results
set.seed(300)
FINALCLUSTER<- kmeans(PHARMACEUTICALS_SCALE, 5, nstart = 25)
print(FINALCLUSTER)
```

```
## K-means clustering with 5 clusters of sizes 8, 3, 4, 2, 4
##
## Cluster means:
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
## 2 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 3 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 4 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
## 5  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.27449312 -0.7041516       0.556954446
## 2  1.36644699 -0.6912914      -1.320000179
## 3  0.06308085  1.5180158      -0.006893899
## 4 -0.14170336 -0.1168459      -1.416514761
## 5 -0.46807818  0.4671788       0.591242521
##
## Clustering vector:
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK  NVS
##    1    4    1    1    3    2    1    2    3    1    5    2    5    3    5    1
##  PFE  PHA  SGP  WPI  WYE
##    5    4    1    3    1
```

```
##
## Within cluster sum of squares by cluster:
## [1] 21.879320 15.595925 12.791257  2.803505  9.284424
##  (between_SS / total_SS =  65.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
clusplot(PHARMACEUTICALS_SCALE,FINALCLUSTER$cluster, color = TRUE, labels = 2,lines = 0)
```



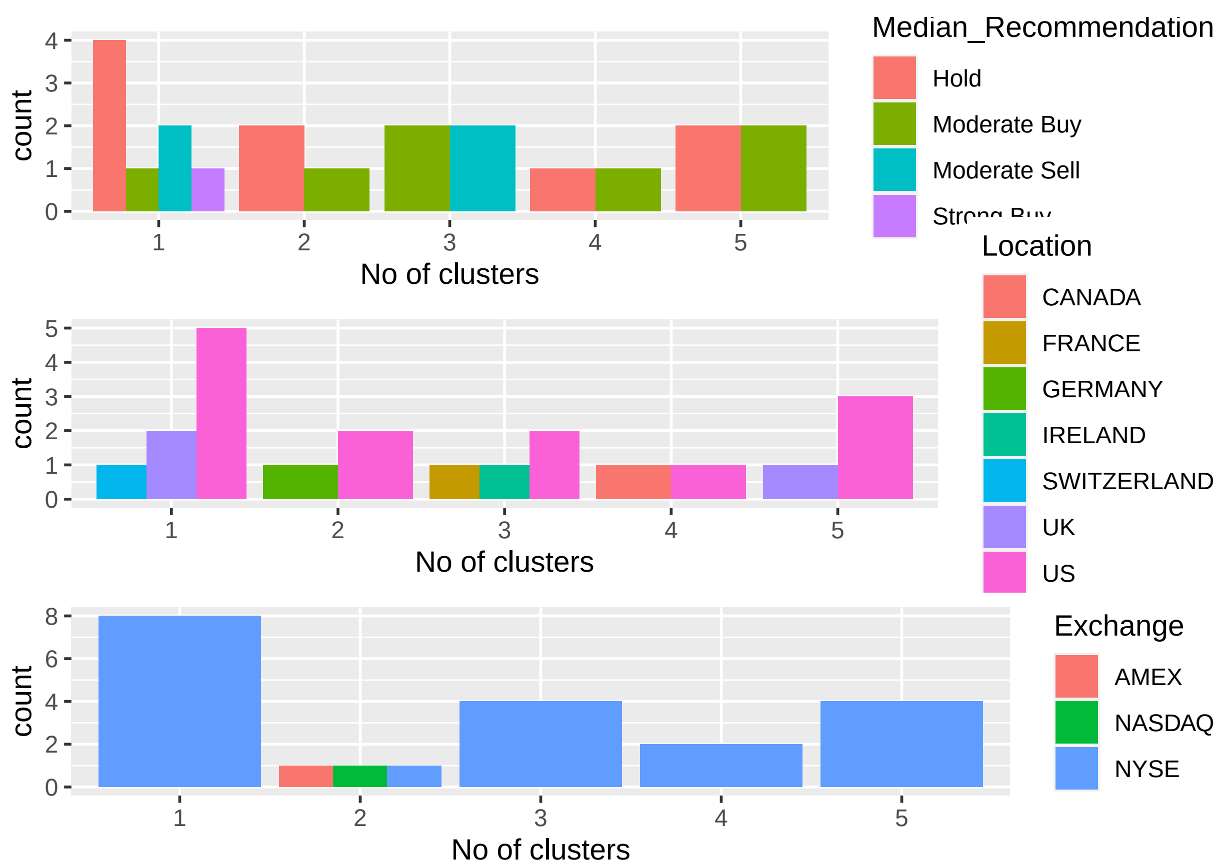

```
#b) Interpret the clusters with respect to the numerical variables used in forming the clusters.
#Cluster 1 consists of the stocks AHM, SGP, WYE, BMY, AZN, ABT, NVS, and LLY (lowest Market Cap, lowest
#Cluster 2 (lowest Rev Growth, highest Beta and levearge, lowest Net Profit Margin) is composed of the
#Cluster3 Lowest PE Ratio, Highest ROE, Lowest ROA, Lowest Net Profit Margin, Highest Rev Growth: WPI,
#cluster4 AGN, PHA (highest PE Ratio, lowest Asset Turnover, and lowest Beta)
#cluster5 JNJ, MRK, PFE, and GSK(Highest Market Cap, ROE, ROA, Asset Turnover Ratio, and Lowest Beta/PE

PHARMA_CLUSTER <- PHARMACEUTICALS[,c(12,13,14)]%>% mutate(clusters = FINALCLUSTER$cluster)%>% arrange(c
PHARMA_CLUSTER
```

```
##      Median_Recommendation    Location Exchange clusters
## ABT           Moderate Buy          US     NYSE        1
## AHM             Strong Buy          UK     NYSE        1
```

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

```r
#(c)Is there a pattern in the clusters with respect to the numerical variables (10 to 12)?
plot1<-ggplot(PHARMA_CLUSTER, mapping = aes(factor(clusters), fill=Median_Recommendation))+geom_bar(posi
plot2<- ggplot(PHARMA_CLUSTER, mapping = aes(factor(clusters),fill = Location))+geom_bar(position = 'doc
plot3<- ggplot(PHARMA_CLUSTER, mapping = aes(factor(clusters),fill = Exchange))+geom_bar(position = 'doc
grid.arrange(plot1, plot2, plot3)
```



```
#Given the graph:
#Cluster 1: The Hold median, which also includes distinct Hold, Moderate Buy, Moderate Sell, and Strong
#Cluster 2 features a distinct Hold and Moderate Buy median as well as a varied count between the US an
#Cluster 3 is traded on the NYSE, has distinct counts for France, Ireland, and the US, and has median b
#Cluster 4: has the same hold and moderate buy medians and is distributed throughout the US and UK in a
#Cluster 5: only listed on the NYSE, evenly distributed across the US and Canada, with medians of Hold
#Regarding the media recommendation variable, the clusters exhibit a certain pattern:
#Hold Recommendation is present in Clusters 1 and 2.
#All of Clusters 3, 4, and 5 have a moderate purchase recommendation.
```

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

#Cluster 1 :-  HIGH HOLD CLUSTER
#Cluster 2 :- HOLD CLUSTER
#Cluster 3 :- BUY-SELL CLUSTER
#Cluster 4 :- HOLD-BUY CLUSTER
#Cluster 5 :- HOLD-BUY CLUSTER
```