Assignment 4

Abi

2022-11-06

Importing Dataset

```
Pharma <- read.csv("C:/Users/abinaya/Downloads/Pharmaceuticals.csv")
View(Pharma)
```

Loading Packages

```
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
```

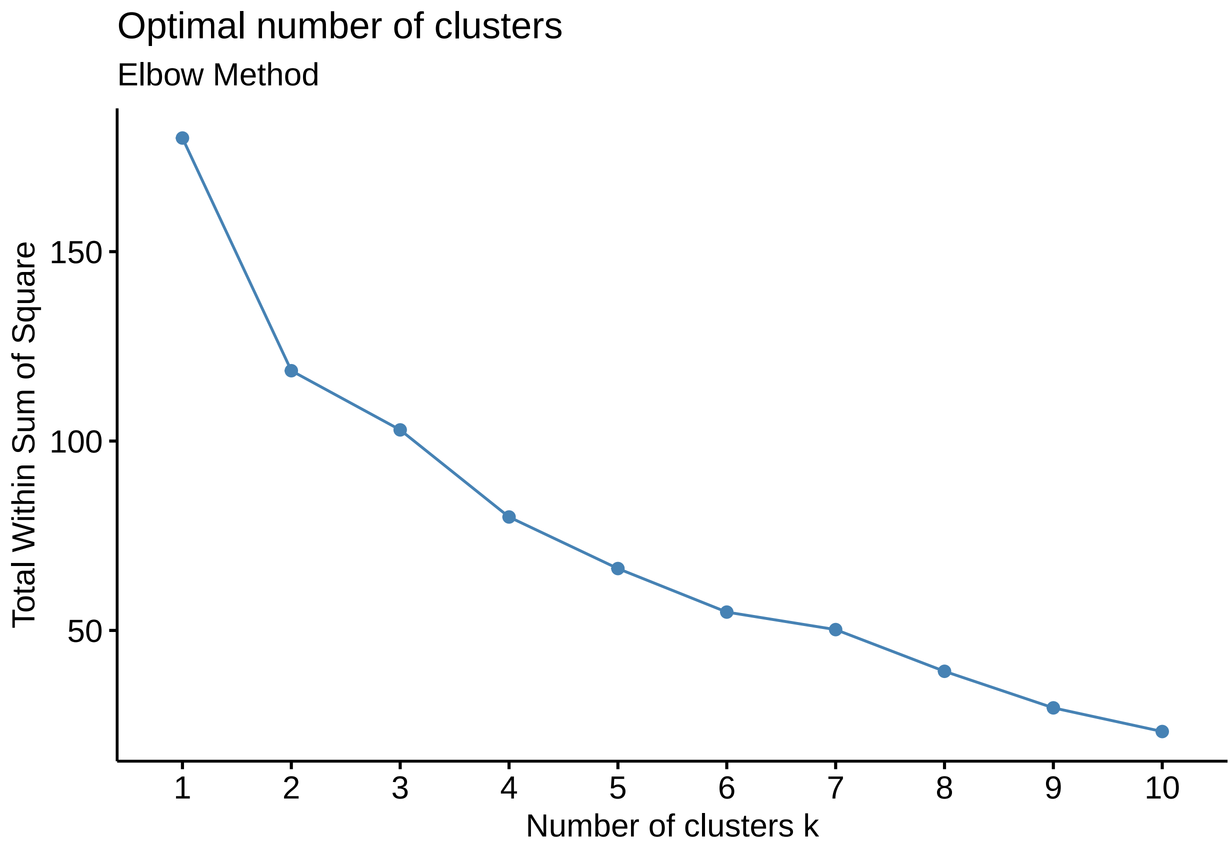
```r
library(cluster)
library(dplyr)
```

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

# Using numerical variables and removing the Null Value
colSums(is.na(Pharma))
```

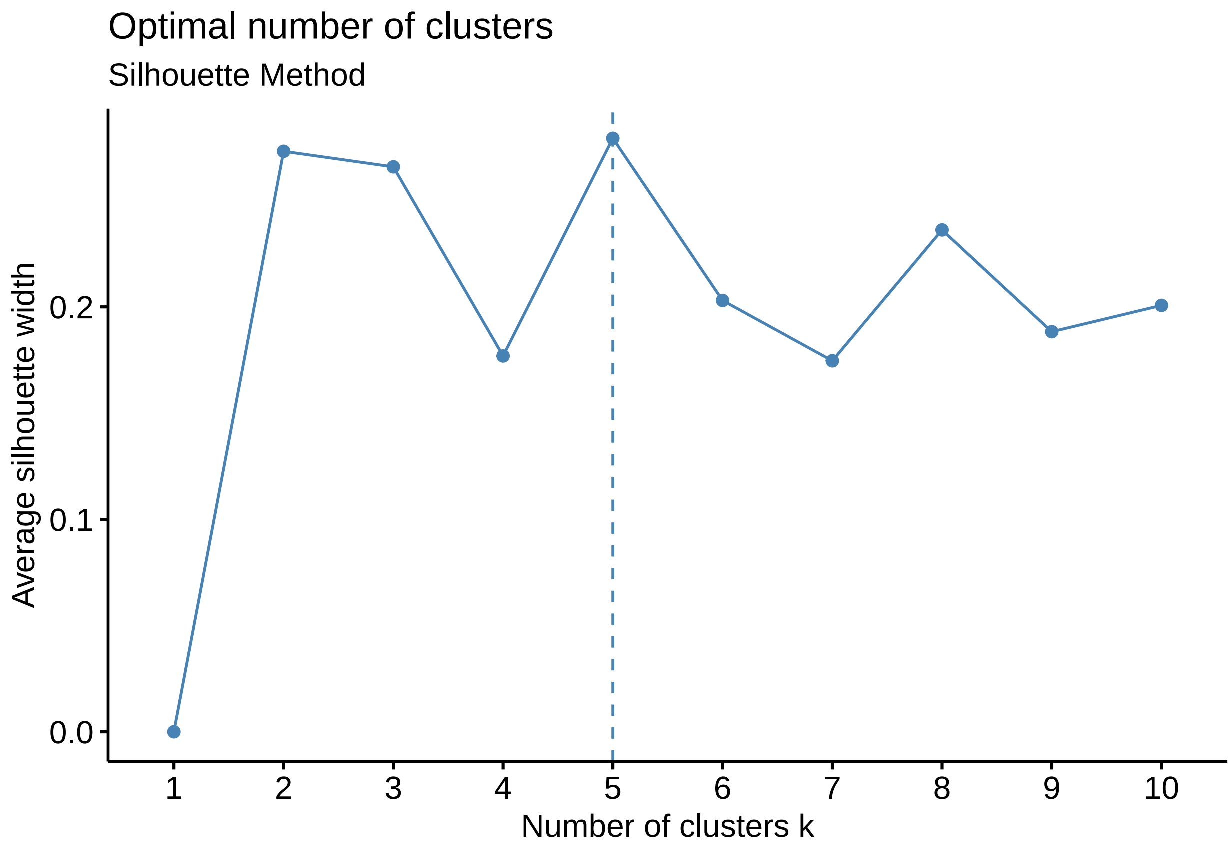
```
##                Symbol                  Name            Market_Cap
##                     0                     0                     0
##                  Beta              PE_Ratio                   ROE
##                     0                     0                     0
##                  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

```r
row.names(Pharma)<- Pharma[,1]
Pharma1<- Pharma[, 3:11]
view(Pharma1)
# Scaling dataset
Pharma_Data <- scale(Pharma1)
view(Pharma_Data)
```

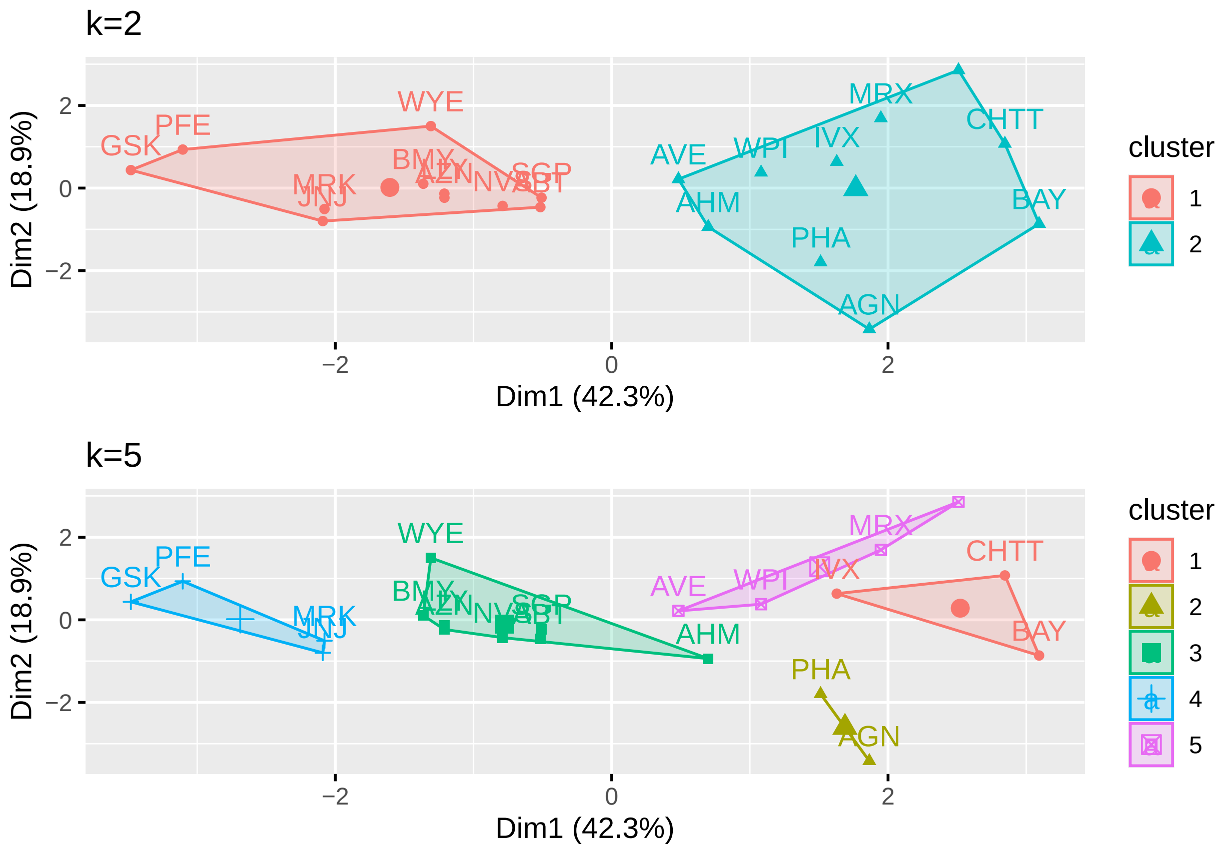
```r
# Estimating the number of clusters
# Here we are using WSS method to scale data, to calculate the value of k

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



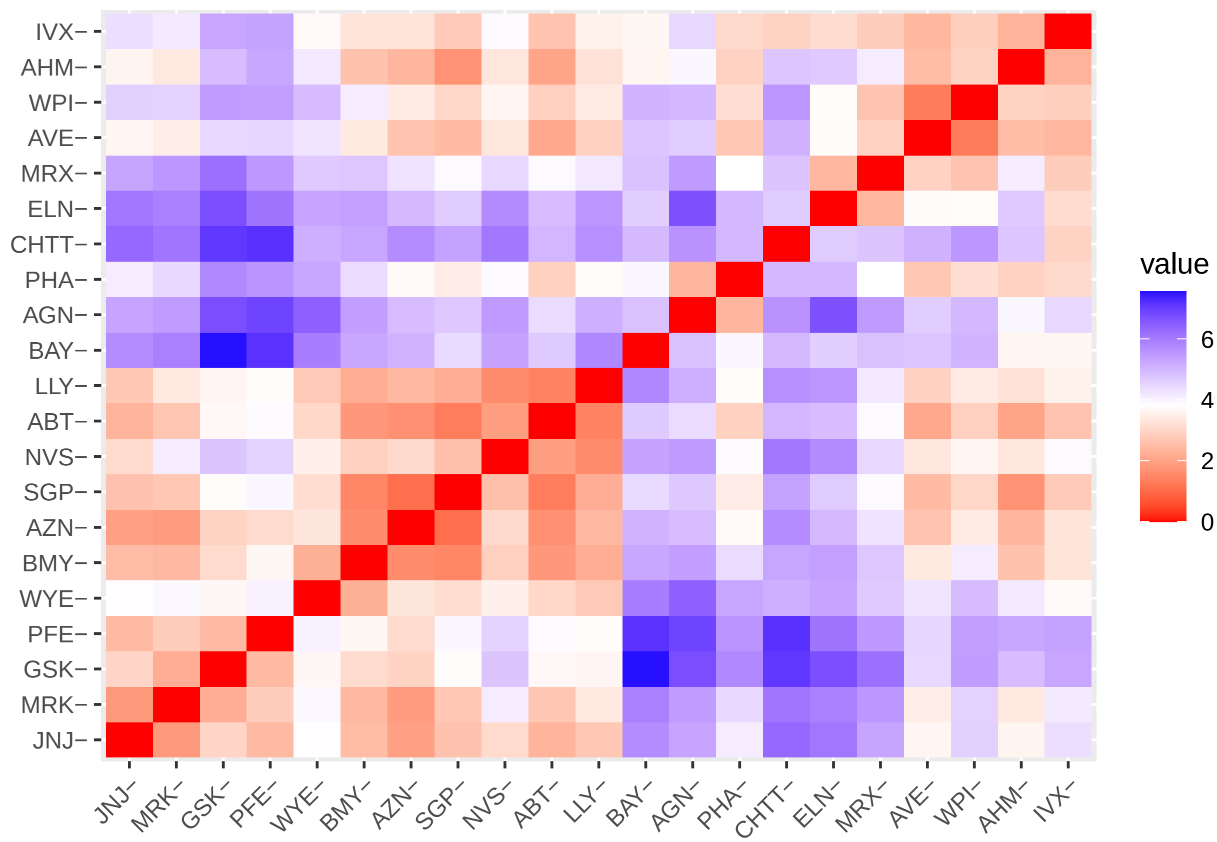

```r
# Scaling the data using the silhouette method, to yields the cluster count.
fviz_nbclust(Pharma_Data, kmeans, method = "silhouette") + labs(subtitle = "Silhouette Method")
```




```
# Computing K-means clustering for multiple centers using a range of K values, then comparing the outco
k2 <- kmeans(Pharma_Data, centers = 2, nstart = 25)
k5 <- kmeans(Pharma_Data, centers = 5, nstart = 25)

Plot1 <- fviz_cluster(k2, data = Pharma_Data)+ggtitle("k=2")
Plot2 <- fviz_cluster(k5, data = Pharma_Data)+ggtitle("k=5")

grid.arrange(Plot1,Plot2, nrow = 2)
```




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




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

```
##   Group.1 Market_Cap        Beta    PE_Ratio        ROE         ROA
## 1       1 -0.4392513 -0.47018004  2.70002464 -0.8349525 -0.92349509
## 2       2  0.9547543 -0.06120687 -0.35764816  1.0818081  1.10336187
## 3       3 -0.1799275 -0.81238208 -0.22714308 -0.3387161 -0.04563784
## 4       4 -0.8705151  1.34098686 -0.05284434 -0.6184015 -1.19284783
## 5       5 -0.9668697  1.51626107 -0.57398880 -0.8382671 -0.98926727
##   Asset_Turnover   Leverage  Rev_Growth Net_Profit_Margin
## 1      0.2306328 -0.1417034 -0.11684587        -1.4165148
## 2      0.8566361 -0.2797499 -0.01818848         0.7082574
## 3     -0.1976853 -0.4168821 -0.14141325         0.1923035
## 4     -0.4612656  1.3664470 -0.69129140        -1.3200002
## 5     -1.8450624  0.5302448  1.71238901         0.2445520
```

```
Pharma_Data1 <- data.frame(Pharma_Data, Aggre$cluster)
Pharma_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
```

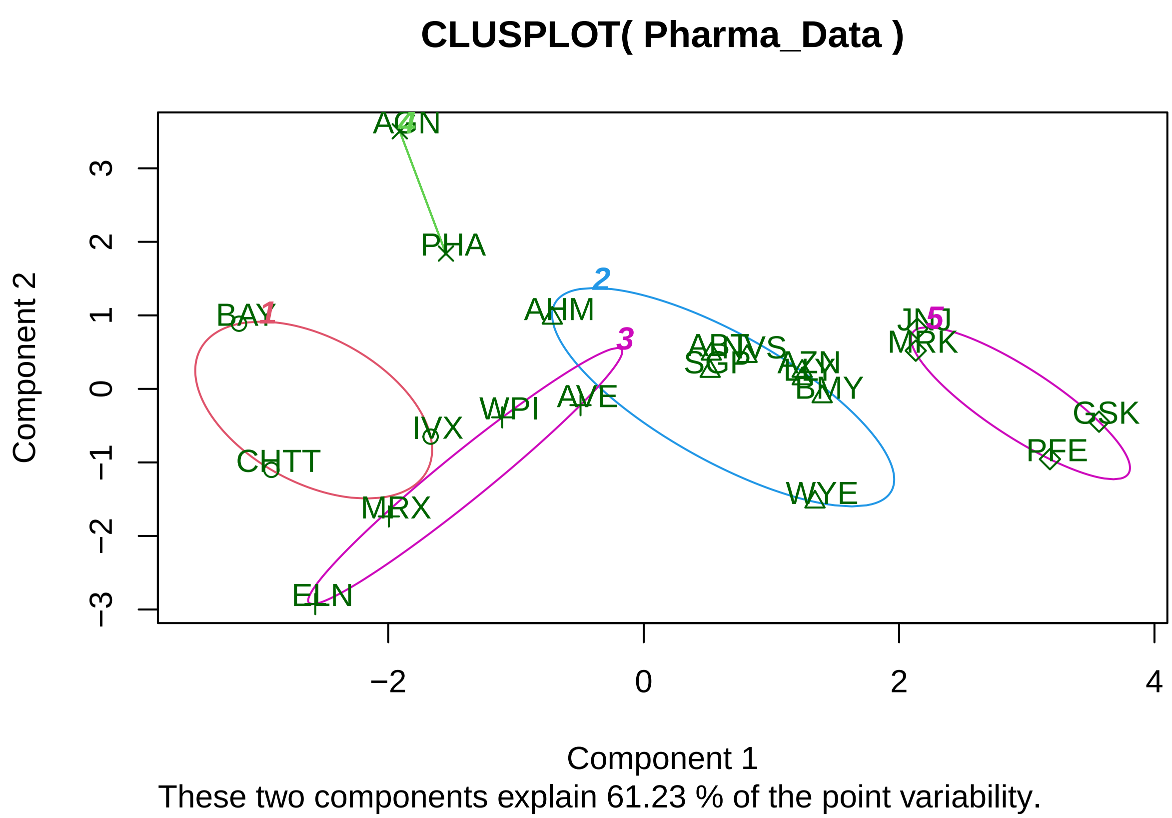
```
## BAY  -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
## BMY  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## CHTT -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## ELN  -0.9704532  2.15893320 -1.34037772 -0.70899938 -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 Aggre.cluster
## ABT  -0.21209793 -0.52776752        0.06168225             3
## AGN   0.01828430 -0.38113909       -1.55366706             1
## AHM  -0.40408312 -0.57211809       -0.68503583             3
## AZN  -0.74965647  0.14744734        0.35122600             2
## AVE  -0.31449003  1.21638667       -0.42597037             3
## BAY  -0.74965647 -1.49714434       -1.99560225             4
## BMY  -0.02011273 -0.96584257        0.74744375             2
## CHTT  3.74279705 -0.63276071       -1.24888417             4
## ELN   0.61983791  1.88617085       -0.36501379             5
## LLY  -0.07130879 -0.64814764        1.17413980             3
## GSK  -0.31449003  0.76926048        0.82363947             2
## IVX   1.10620040  0.05603085       -0.71551412             4
## JNJ  -0.62166634 -0.36213170        0.33598685             2
## MRX   0.44065173  1.53860717        0.85411776             5
## MRK  -0.39128411  0.36014907       -0.24310064             2
## NVS  -0.67286239 -1.45369888        1.02174835             3
## PFE  -0.54487226  1.10143723        1.44844440             2
## PHA  -0.30169102  0.14744734       -1.27936246             1
## SGP  -0.74965647 -0.43544591        0.29026942             3
## WPI  -0.49367621  1.43089863       -0.09070919             3
## WYE   0.68383297 -1.17763919        1.49416183             2
```

```
# Final analysis, results extraction utilizing 5 groupings, and visualization of the outcomes
set.seed(125)
F_Cluster<- kmeans(Pharma_Data, 5, nstart = 25)
print(F_Cluster)
```

```
## K-means clustering with 5 clusters of sizes 3, 8, 4, 2, 4
##
## Cluster means:
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 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 -0.43925134 -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  1.36644699 -0.6912914      -1.320000179
## 2 -0.27449312 -0.7041516       0.556954446
## 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
##    2    4    2    2    3    1    2    1    3    2    5    1    5    3    5    2
##  PFE  PHA  SGP  WPI  WYE
##    5    4    2    3    2
##
## Within cluster sum of squares by cluster:
## [1] 15.595925 21.879320 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(Pharma_Data,F_Cluster$cluster, color = TRUE, labels = 2,lines = 0)
```

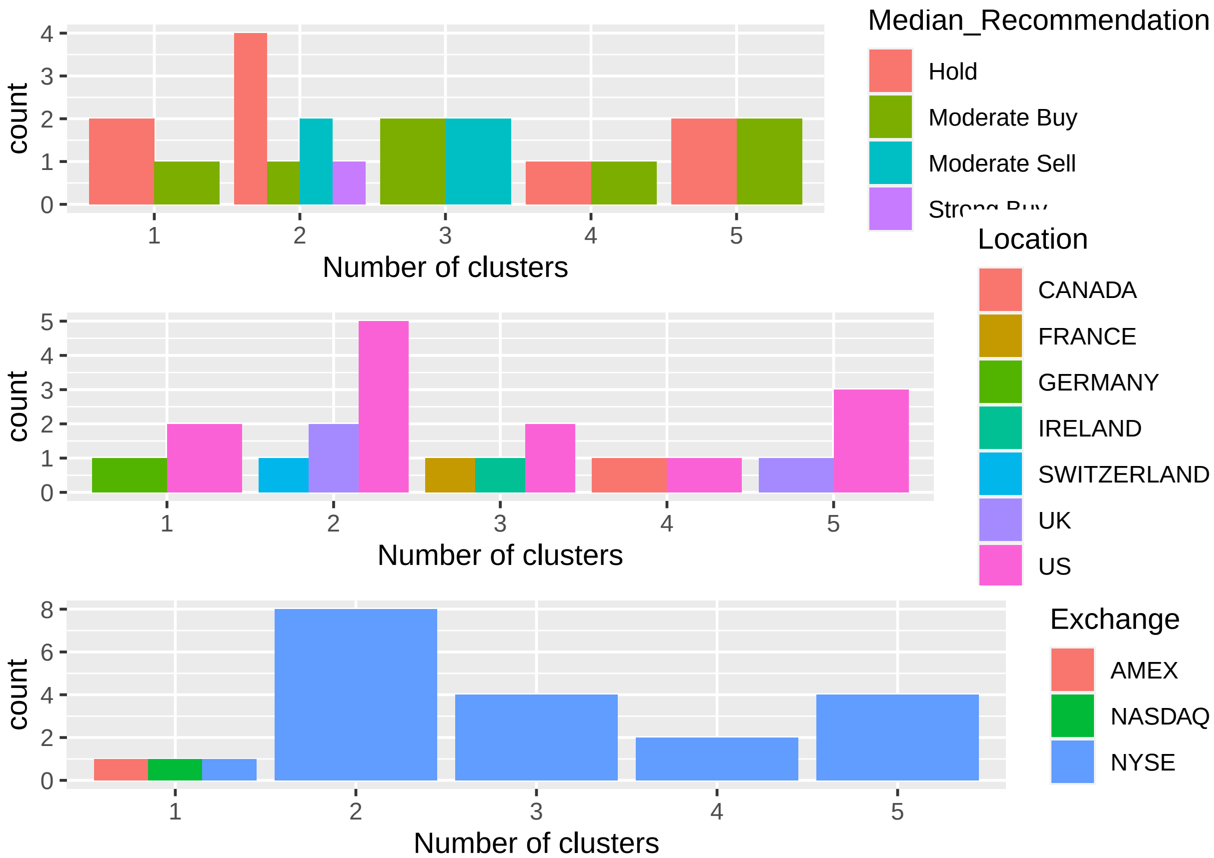



```
# Question b
#Interpret the clusters in light of the numerical variables that were utilized to create them.
```

```r
#Cluster 1: BAY, CHTT, and IVX- lowest Rev Growth, highest Beta and levearge, lowest Net Profit Margin.
#Cluster 2: AHM, SGP, WYE, BMY, AZN, ABT, NVS, and LLY- lowest Market Cap, lowest Beta, lowest PE Ratio
#Cluster 3:WPI, MRX, ELN, AVE- Lowest PE Ratio, Highest ROE, Lowest ROA, Lowest Net Profit Margin, High
#Cluster 4:AGN, PHA-highest PE Ratio, lowest Asset Turnover, and lowest Beta.
#Cluster 5:JNJ, MRK, PFE, and GSK-Highest Market Cap, ROE, ROA, Asset Turnover Ratio, and Lowest Beta/P

P_Cluster <- Pharma[,c(12,13,14)]%>% mutate(clusters = F_Cluster$cluster)%>% arrange(clusters, ascendin
P_Cluster
```

```
##      Median_Recommendation    Location Exchange clusters
## BAY                   Hold     GERMANY     NYSE        1
## CHTT          Moderate Buy          US   NASDAQ        1
## IVX                   Hold          US     AMEX        1
## ABT           Moderate Buy          US     NYSE        2
## AHM             Strong Buy          UK     NYSE        2
## AZN          Moderate Sell          UK     NYSE        2
## BMY          Moderate Sell          US     NYSE        2
## LLY                   Hold          US     NYSE        2
## NVS                   Hold SWITZERLAND     NYSE        2
## SGP                   Hold          US     NYSE        2
## WYE                   Hold          US     NYSE        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(P_Cluster, mapping = aes(factor(clusters), fill=Median_Recommendation))+geom_bar(position
plot2<- ggplot(P_Cluster, mapping = aes(factor(clusters),fill = Location))+geom_bar(position = 'dodge')
plot3<- ggplot(P_Cluster, mapping = aes(factor(clusters),fill = Exchange))+geom_bar(position = 'dodge')
grid.arrange(plot1, plot2, plot3)
```




```
#Given the graph:
#Cluster 1: Despite the fact that the firms are evenly divided among AMEX, NASDAQ, and NYSE, it has a d
#Cluster 2: The highest median in this cluster is the Hold median, which also comprises unique Hold, Mo
#Cluster 3: It has equal moderate buy and moderate sell medians, and distinct counts for France, Irelan
#Cluster 4: It is equally scattred in the US and Canada, with Hold and Moderate Buy medians, which is e
#Cluster 5: It is uniformly distributed across the US and UK, with medians of Hold and Moderate Buy, wh
#Considering the media recommendation variable, the clusters demonstrate a specific pattern:
#Cluster 1 and Cluster 2 has Hold Recommendation.
#Cluster 3, Cluster 4 and Cluster 5 has moderate buy Recommendation.
```

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

#Cluster 1 : Strong Hold cluster
#Cluster 2 : Strong Buy cluster
#Cluster 3 : Tolerable cluster
#Cluster 4 : Fair Buy cluster
#Cluster 5 : Sustained cluster
```