FML Assignment-4

RAJEEV VARMA

2022-11-03

My\_Pharmaceuticals <- read.csv("C:/Users/RAJEEV VARMA/Downloads/Pharmaceuticals.csv")  
summary(My\_Pharmaceuticals)

## 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   
##   
##   
##

str(My\_Pharmaceuticals)

## 'data.frame': 21 obs. of 14 variables:  
## $ Symbol : chr "ABT" "AGN" "AHM" "AZN" ...  
## $ Name : chr "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZeneca PLC" ...  
## $ Market\_Cap : num 68.44 7.58 6.3 67.63 47.16 ...  
## $ Beta : num 0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...  
## $ PE\_Ratio : num 24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...  
## $ ROE : num 26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...  
## $ ROA : num 11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...  
## $ Asset\_Turnover : num 0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...  
## $ Leverage : num 0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...  
## $ Rev\_Growth : num 7.54 9.16 7.05 15 26.81 ...  
## $ Net\_Profit\_Margin : num 16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...  
## $ Median\_Recommendation: chr "Moderate Buy" "Moderate Buy" "Strong Buy" "Moderate Sell" ...  
## $ Location : chr "US" "CANADA" "UK" "UK" ...  
## $ Exchange : chr "NYSE" "NYSE" "NYSE" "NYSE" ...

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.4.1   
## ✔ readr 2.1.2 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(cluster)  
library(gridExtra)

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

library(readr)  
library(dplyr)  
library(caret)

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(factoextra)

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

#a)  
#Removing Null Values and selecting the Numercial variables.  
colSums(is.na(My\_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

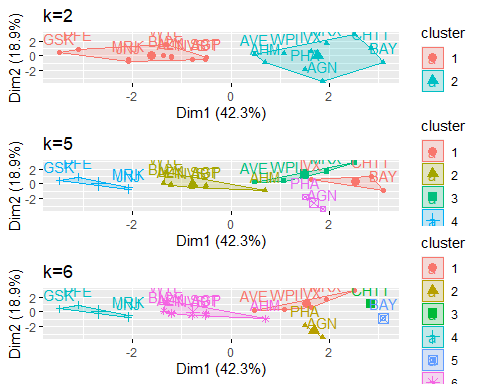
row.names(My\_Pharmaceuticals)<- My\_Pharmaceuticals[,1]  
My\_Pharmaceuticals.data.num<- My\_Pharmaceuticals[, 3:11]  
head(My\_Pharmaceuticals.data.num)

## 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 Normalizing the dataset  
My\_Pharmaceuticals.scale <- scale(My\_Pharmaceuticals.data.num)  
head(My\_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

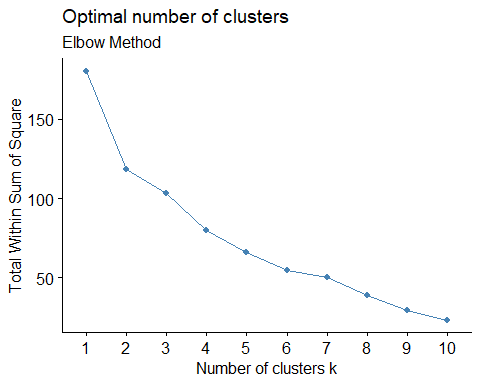
normal.data <- as.data.frame(scale(My\_Pharmaceuticals.data.num))  
  
#Computing K-means clustering for different centers   
kmeans.1 <- kmeans(My\_Pharmaceuticals.scale, centers = 2, nstart = 30)  
kmeans.2<- kmeans(My\_Pharmaceuticals.scale, centers = 5, nstart = 30)  
kmeans.3<- kmeans(My\_Pharmaceuticals.scale, centers = 6, nstart = 30)  
Plot.1<-fviz\_cluster(kmeans.1, data = My\_Pharmaceuticals.scale)+ggtitle("k=2")  
plot.2<-fviz\_cluster(kmeans.2, data = My\_Pharmaceuticals.scale)+ggtitle("k=5")  
plot.3<-fviz\_cluster(kmeans.3, data = My\_Pharmaceuticals.scale)+ggtitle("k=6")  
grid.arrange(Plot.1,plot.2,plot.3, nrow = 3)



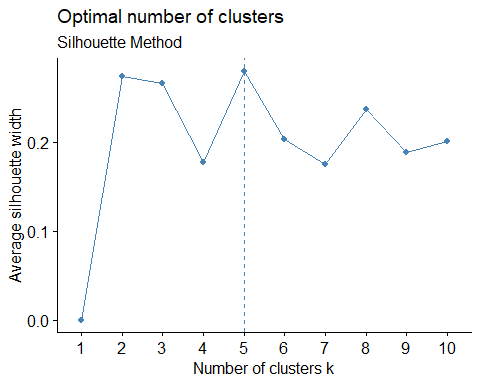
distance<- dist(My\_Pharmaceuticals.scale, method = "euclidean")  
fviz\_dist(distance)



# Estimate the number of clusters(Elbow Method is used for scaling the data to determine the value of k)  
fviz\_nbclust(normal.data, FUNcluster = kmeans, method = "wss") + labs(subtitle = "Elbow Method")



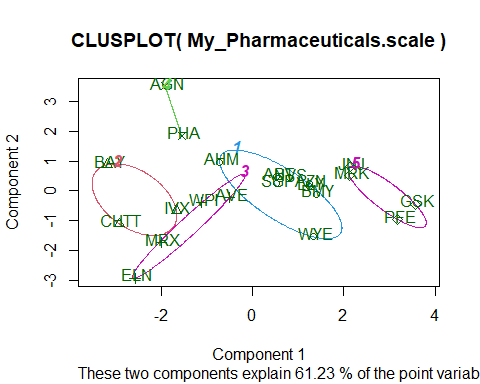
# Silhouette Method is used to determine total number of clusters  
fviz\_nbclust(normal.data,FUNcluster = kmeans,method = "silhouette")+labs(subtitle="Silhouette Method")



# Final analysis and Extracting the results using 5 clusters and Visualizing the results  
set.seed(300)  
final.Cluster<- kmeans(My\_Pharmaceuticals.scale, 5, nstart = 25)  
print(final.Cluster)

## 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"

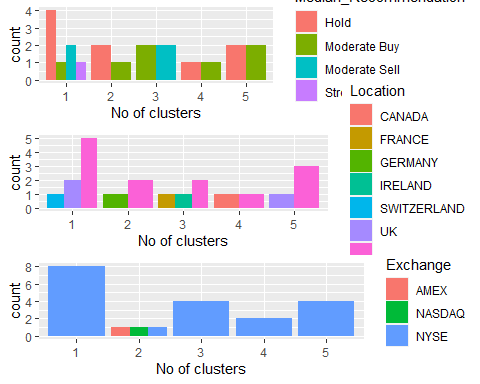
clusplot(My\_Pharmaceuticals.scale,final.Cluster$cluster, color = TRUE, labels = 2,lines = 0)



#b)  
#Cluster 1 - AHM,SGP,WYE,BMY,AZN, ABT, NVS, LLY   
#Lowest Market.Cap,highest Beta,lowest PE.Ratio,highest Leverage,highest Rev.Growth.  
  
#Cluster 2 - BAY, CHTT, IVX   
#Lowest Rev.Growth,highest Net.Profit.Margin  
  
#Cluster 3 - WPI, MRX,ELN,AVE   
#Highest PE.Ratio,lowest.ROE,lowest ROA,lowest Net.Profit.Margin  
  
#Cluster 4 - AGN, PHA   
#Lowest Beta,lowest Asset.Turnover  
  
#Cluster 5 - JNJ, MRK, PFE,GSK   
#Highest Market.Cap and lowest Beta/PE Ratio  
  
My\_Pharmaceuticals.Cluster <- My\_Pharmaceuticals[,c(12,13,14)]%>% mutate(clusters = final.Cluster$cluster)%>% arrange(clusters, ascending = TRUE)  
My\_Pharmaceuticals.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

#c)  
plot1<-ggplot(My\_Pharmaceuticals.Cluster, mapping = aes(factor(clusters), fill=Median\_Recommendation))+geom\_bar(position = 'dodge')+labs(x ='No of clusters')  
plot2<- ggplot(My\_Pharmaceuticals.Cluster, mapping = aes(factor(clusters),fill = Location))+geom\_bar(position = 'dodge')+labs(x ='No of clusters')  
plot3<- ggplot(My\_Pharmaceuticals.Cluster, mapping = aes(factor(clusters),fill = Exchange))+geom\_bar(position = 'dodge')+labs(x ='No of clusters')  
grid.arrange(plot1, plot2, plot3)

 #As per <graph:->

#Cluster 1 :The Hold median is highest in this cluster,It also contains separate Hold, Moderate Buy, Moderate Sell, and Strong Buy medians. They are listed on the NYSE and they come from the US, UK, and Switzerland.

#Cluster 2: Even though the firms are evenly divided throughout AMEX, NASDAQ, and NYSE, has a distinct Hold and Moderate Buy median, as well as a different count between the US and Germany.

#Cluster 3: It is listed on the NYSE, has separate counts for France, Ireland, and the US, and has equal moderate buy and sell medians.

#Cluster 4: It dispersed throughout the US and UK, as well as being listed in, has the identical hold and moderate buy medians

#Cluster 5: solely listed on the NYSE, equally dispersed in the US and Canada, with Hold and Moderate Buy medians.

#d)  
#Cluster1-Hold Cluster  
#Cluster2-Hold Cluster  
#Cluster3-Buy-Sell Cluster  
#Cluster4-Hold-Buy Cluster  
#Cluster5-Hold-Buy Cluster