Assignment 4

2022-11-06

library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

library(tidyverse)

## ── Attaching packages  
## ───────────────────────────────────────  
## tidyverse 1.3.2 ──

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

library(factoextra)

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

set.seed(123)

getwd()

## [1] "/Users/bharathreddy/Downloads"

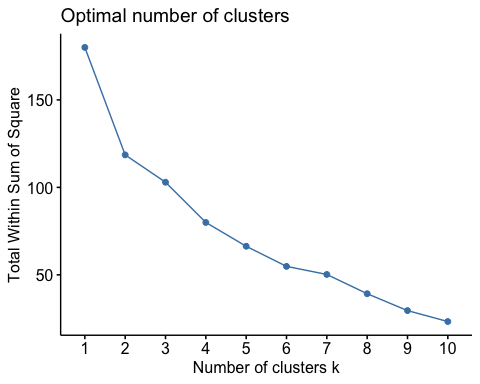
setwd("/Users/bharathreddy/Downloads")  
k\_mean<-read.csv("Pharmaceuticals.csv")

#1.a. 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.

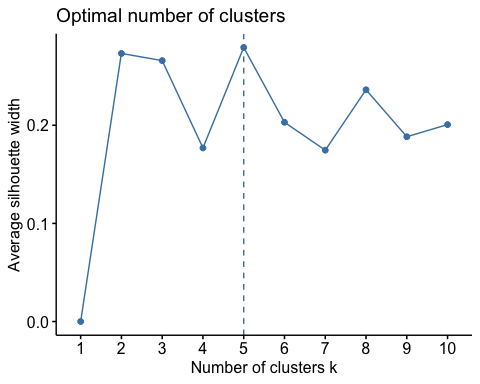
data\_frame <- k\_mean[,c(3:11)]  
norm\_scale <- scale(data\_frame)

#finding optiomal k value

fviz\_nbclust(norm\_scale, kmeans, method = "wss")



fviz\_nbclust(norm\_scale, kmeans, method = "silhouette")



# found kmeans using above methods

df1 <- kmeans(norm\_scale, centers = 2, nstart = 25)  
df1$centers

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 0.6733825 -0.3586419 -0.2763512 0.6565978 0.8344159 0.4612656  
## 2 -0.7407208 0.3945061 0.3039863 -0.7222576 -0.9178575 -0.5073922  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.3331068 -0.2902163 0.6823310  
## 2 0.3664175 0.3192379 -0.7505641

df2 <- kmeans(norm\_scale, centers = 5, nstart = 25)  
df1$centers

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 0.6733825 -0.3586419 -0.2763512 0.6565978 0.8344159 0.4612656  
## 2 -0.7407208 0.3945061 0.3039863 -0.7222576 -0.9178575 -0.5073922  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 -0.3331068 -0.2902163 0.6823310  
## 2 0.3664175 0.3192379 -0.7505641

# plotting cluster

fviz\_cluster(df1, data= norm\_scale)



fviz\_cluster(df2, data= norm\_scale)



#using silhouette method where k=5 to interpret clusters.

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

cluster<- df2$cluster  
cluster\_group<-cbind(k\_mean, cluster)  
cluster\_group

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

# finding mean value using aggregate function.

aggregate(cluster\_group[,-c(1,2,12:14)],by=list(cluster\_group$cluster),FUN="mean")

## Group.1 Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 1 6.636667 0.87000 24.6000 16.46667 4.166667 0.6000  
## 2 2 31.910000 0.40500 69.5000 13.20000 5.600000 0.7500  
## 3 3 55.810000 0.41375 20.2875 28.73750 12.687500 0.7375  
## 4 4 157.017500 0.48000 22.2250 44.42500 17.700000 0.9500  
## 5 5 13.100000 0.59750 17.6750 14.57500 6.200000 0.4250  
## Leverage Rev\_Growth Net\_Profit\_Margin cluster  
## 1 1.653333 5.733333 7.033333 1  
## 2 0.475000 12.080000 6.400000 2  
## 3 0.371250 5.591250 19.350000 3  
## 4 0.220000 18.532500 19.575000 4  
## 5 0.635000 30.142500 15.650000 5

#2.Interpret the clusters with respect to the numerical variables used in forming the clusters. # cluster 1 has highest market\_cap,asset\_turnover,ROE,ROA where Beta,PE\_RATIO has lower values. # cluster 2 has highest PE\_RATIO and lowest in ASSET\_TURNOVER and Beta. # cluster 3 has highest PE\_RATIO,ROE and Market\_cap where as Asset\_turnover is lowest of all. # cluster 4 has the lowest Leverage,Beta and highest Market\_cap,ROE. # cluster 5 has lowest Beta ,leverage and highets PE\_ratio,Market\_cap.

#3.Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in # cluster 1 with lowest in beta and PE\_RATIO has majority of HOLD and moderate sell. # cluster 2 with highest PE\_RATIO has both HOLD and moderate buy. # cluster 3 with lower Asset\_turnover has equal moderate buy and moderate sell pattern. # cluster 4 with highest market\_cap has same praportion of moderate buy and HOLD. # cluster 5 with lowest Beta has HOLD and moderate buy.

#4.Provide an appropriate name for each cluster using any or all of the variables in the dataset. # cluster 1 has most of holds hence Hold cluster. # cluster 2 with lowest Asset\_turnover is Hold or moderate buy cluster. # cluster 3 with equal pattern is either moderate buy or moderate sell cluster. # cluster 4 will be moderate buy/Hold cluster. # cluster 4 with lowest Beta highest Market\_cap cluster or Hold/moderate buy cluster.