KV\_Assignment4\_fml

Keerthana Vonteddu

2024-03-17

#Summary 1. The data has been standardized using range and Z-score scaling. 2. Five clusters with five centroids are the optimal outcome when k-means clustering is carried out using the silhouette technique. 3. The best k = 5 number we came up with in this instance is roughly 0.3, which is good. Higher values indicate higher quality. The silhouette ranges from -1 to +1. 4. The cluster names are categorized using the properties of the data set and the cluster summary: NetProfitMargin was higher in Cluster 1 and “High Market Cap & High ROE” Hold Cluster refers to Cluster 2. The third cluster is “Low Market Cap & Less ROA”. The fourth cluster is “Small Net Profit Margin & High PE Ratio.” The fifth cluster is “Moderate Buy & Moderate Sell”. #Problem Statement: An equities analyst is researching the pharmaceutical sector and needs your assistance in examining and deciphering the financial information that her company has gathered. Her primary goal is to use certain fundamental financial metrics to comprehend the pharmaceutical industry’s structure. The Pharmaceuticals.csv file contains financial information collected on 21 pharmaceutical companies. Get the Pharmaceuticals.csv file. The following variables are noted for every firm.

1. Market capitalization expressed in USD per billion. 2. Beta. The ratio of price to earnings. 4. Equity return. 5. Asset return. 6.Asset turnover. 7. Make use of leverage. 8. Projected growth in sales. 9. The net profit margin. 10. Median advice (for all major brokerages). 11. The location of the corporate office. 12. The stock exchange where the company is listed.

#Packages Required

#install.packages("flexclust")  
library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

library(ISLR)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

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: modeltools

## Loading required package: stats4

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ✔ purrr 1.0.2 ✔ tidyr 1.3.0  
## ✔ readr 2.1.4

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ purrr::lift() masks caret::lift()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(readr)

#keeping in mind that KV\_A4\_fml is our dataframe and loading information about medications

KV\_A4\_fml <- read.csv("/Users/keerthanavonteddu/Desktop/fml/Pharmaceuticals.csv")  
summary(KV\_A4\_fml)

## 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   
##   
##   
##

head(KV\_A4\_fml)

## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## 1 ABT Abbott Laboratories 68.44 0.32 24.7 26.4 11.8 0.7  
## 2 AGN Allergan, Inc. 7.58 0.41 82.5 12.9 5.5 0.9  
## 3 AHM Amersham plc 6.30 0.46 20.7 14.9 7.8 0.9  
## 4 AZN AstraZeneca PLC 67.63 0.52 21.5 27.4 15.4 0.9  
## 5 AVE Aventis 47.16 0.32 20.1 21.8 7.5 0.6  
## 6 BAY Bayer AG 16.90 1.11 27.9 3.9 1.4 0.6  
## Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation Location Exchange  
## 1 0.42 7.54 16.1 Moderate Buy US NYSE  
## 2 0.60 9.16 5.5 Moderate Buy CANADA NYSE  
## 3 0.27 7.05 11.2 Strong Buy UK NYSE  
## 4 0.00 15.00 18.0 Moderate Sell UK NYSE  
## 5 0.34 26.81 12.9 Moderate Buy FRANCE NYSE  
## 6 0.00 -3.17 2.6 Hold GERMANY NYSE

##Q1.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\_scaled.df <- scale(KV\_A4\_fml[,3:11])  
head(data\_scaled.df)

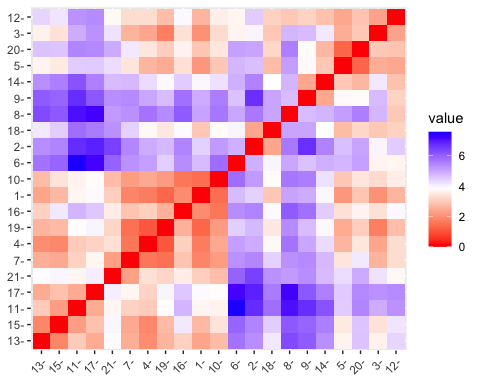
## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## [1,] 0.1840960 -0.80125356 -0.04671323 0.04009035 0.2416121 -5.121077e-16  
## [2,] -0.8544181 -0.45070513 3.49706911 -0.85483986 -0.9422871 9.225312e-01  
## [3,] -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700 9.225312e-01  
## [4,] 0.1702742 -0.02225704 -0.24290879 0.10638147 0.9181259 9.225312e-01  
## [5,] -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461 -4.612656e-01  
## [6,] -0.6953818 2.27578267 0.14948233 -1.45146000 -1.7127612 -4.612656e-01  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## [1,] -0.2120979 -0.5277675 0.06168225  
## [2,] 0.0182843 -0.3811391 -1.55366706  
## [3,] -0.4040831 -0.5721181 -0.68503583  
## [4,] -0.7496565 0.1474473 0.35122600  
## [5,] -0.3144900 1.2163867 -0.42597037  
## [6,] -0.7496565 -1.4971443 -1.99560225

summary(data\_scaled.df)

## Market\_Cap Beta PE\_Ratio ROE   
## Min. :-0.9768 Min. :-1.3466 Min. :-1.3404 Min. :-1.4515   
## 1st Qu.:-0.8763 1st Qu.:-0.6844 1st Qu.:-0.4023 1st Qu.:-0.7223   
## Median :-0.1614 Median :-0.2560 Median :-0.2429 Median :-0.2118   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.2762 3rd Qu.: 0.4841 3rd Qu.: 0.1495 3rd Qu.: 0.3450   
## Max. : 2.4200 Max. : 2.2758 Max. : 3.4971 Max. : 2.4597   
## ROA Asset\_Turnover Leverage Rev\_Growth   
## Min. :-1.7128 Min. :-1.8451 Min. :-0.74966 Min. :-1.4971   
## 1st Qu.:-0.9047 1st Qu.:-0.4613 1st Qu.:-0.54487 1st Qu.:-0.6328   
## Median : 0.1289 Median :-0.4613 Median :-0.31449 Median :-0.3621   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.00000 Mean : 0.0000   
## 3rd Qu.: 0.8430 3rd Qu.: 0.9225 3rd Qu.: 0.01828 3rd Qu.: 0.7693   
## Max. : 1.8389 Max. : 1.8451 Max. : 3.74280 Max. : 1.8862   
## Net\_Profit\_Margin   
## Min. :-1.99560   
## 1st Qu.:-0.68504   
## Median : 0.06168   
## Mean : 0.00000   
## 3rd Qu.: 0.82364   
## Max. : 1.49416

#Distance Measurement in Data Analysis for Scaled Medicine

distance.df <- get\_dist(data\_scaled.df)  
fviz\_dist(distance.df)



#Scale the data and arrange it using K-means clustering.

kmeans.\_1.df <- kmeans(data\_scaled.df,centers = 5,nstart = 25)  
kmeans.\_1.df

## K-means clustering with 5 clusters of sizes 8, 3, 4, 4, 2  
##   
## 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 1.69558112 -0.1780563 -0.19845823 1.2349879 1.3503431 1.1531640  
## 5 -0.43925134 -0.4701800 2.70002464 -0.8349525 -0.9234951 0.2306328  
## 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 -0.46807818 0.4671788 0.591242521  
## 5 -0.14170336 -0.1168459 -1.416514761  
##   
## Clustering vector:  
## [1] 1 5 1 1 3 2 1 2 3 1 4 2 4 3 4 1 4 5 1 3 1  
##   
## Within cluster sum of squares by cluster:  
## [1] 21.879320 15.595925 12.791257 9.284424 2.803505  
## (between\_SS / total\_SS = 65.4 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
## [6] "betweenss" "size" "iter" "ifault"

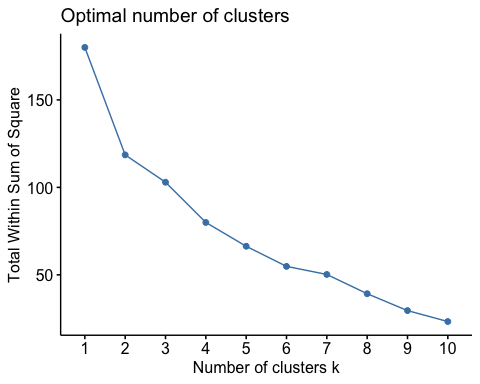
fviz\_cluster(kmeans.\_1.df,data = data\_scaled.df) #graphical Representation of the Cluster



Overview: Information on nine attributes pertaining to each of the twenty-one pharmaceutical businesses can be acquired by applying the summary algorithm on the k-means data. There are five different clusters in total, each having five centroid points: 2, 3, 4, 8, and 9.

##Q2.Interpret the clusters with respect to the numerical variables used in forming the clusters

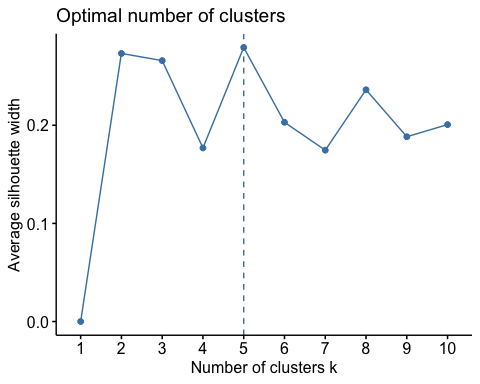
#graphing the total number of clusters versus the sum of the values  
fviz\_nbclust(data\_scaled.df,kmeans,method = "wss")



This figure makes it evident that, rather than taking the form of an elbow, the graph is significantly flattening at k=4 and k=6.

#By graphing the number of clusters against the typical silence width, you can determine the ideal number of clusters.

fviz\_nbclust(data\_scaled.df,kmeans,method = "silhouette")



We can quickly determine that five clusters is the best and most effective option for this k-means clustering by using the Silhoutte graph.

##Q3.Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters).

#utilizing the data shown in the clusters to get the mean value.

#executing the aggregate functionality of the pharmaceutical data.  
aggregate(KV\_A4\_fml[3:11],by=list(cluster = kmeans.\_1.df$cluster),mean)

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

#Use Cbind to merge the data frames.  
shot.data <- cbind(KV\_A4\_fml,cluster = kmeans.\_1.df$cluster)  
tibble(shot.data)

## # A tibble: 21 × 15  
## Symbol Name Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 ABT Abbott … 68.4 0.32 24.7 26.4 11.8 0.7 0.42  
## 2 AGN Allerga… 7.58 0.41 82.5 12.9 5.5 0.9 0.6   
## 3 AHM Amersha… 6.3 0.46 20.7 14.9 7.8 0.9 0.27  
## 4 AZN AstraZe… 67.6 0.52 21.5 27.4 15.4 0.9 0   
## 5 AVE Aventis 47.2 0.32 20.1 21.8 7.5 0.6 0.34  
## 6 BAY Bayer AG 16.9 1.11 27.9 3.9 1.4 0.6 0   
## 7 BMY Bristol… 51.3 0.5 13.9 34.8 15.1 0.9 0.57  
## 8 CHTT Chattem… 0.41 0.85 26 24.1 4.3 0.6 3.51  
## 9 ELN Elan Co… 0.78 1.08 3.6 15.1 5.1 0.3 1.07  
## 10 LLY Eli Lil… 73.8 0.18 27.9 31 13.5 0.6 0.53  
## # ℹ 11 more rows  
## # ℹ 6 more variables: Rev\_Growth <dbl>, Net\_Profit\_Margin <dbl>,  
## # Median\_Recommendation <chr>, Location <chr>, Exchange <chr>, cluster <int>

#recapitulating the detailed division by cluster

by(shot.data,factor(shot.data$cluster),summary)

## factor(shot.data$cluster): 1  
## Symbol Name Market\_Cap Beta   
## Length:8 Length:8 Min. : 6.30 Min. :0.1800   
## Class :character Class :character 1st Qu.:44.67 1st Qu.:0.2875   
## Mode :character Mode :character Median :59.48 Median :0.4800   
## Mean :55.81 Mean :0.4138   
## 3rd Qu.:69.79 3rd Qu.:0.5125   
## Max. :96.65 Max. :0.6300   
## PE\_Ratio ROE ROA Asset\_Turnover   
## Min. :13.10 Min. :14.90 Min. : 7.80 Min. :0.5000   
## 1st Qu.:17.65 1st Qu.:21.43 1st Qu.:11.65 1st Qu.:0.6000   
## Median :21.10 Median :26.90 Median :13.35 Median :0.7500   
## Mean :20.29 Mean :28.74 Mean :12.69 Mean :0.7375   
## 3rd Qu.:22.38 3rd Qu.:31.95 3rd Qu.:13.90 3rd Qu.:0.9000   
## Max. :27.90 Max. :54.90 Max. :15.40 Max. :0.9000   
## Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation  
## Min. :0.0000 Min. :-2.690 Min. :11.20 Length:8   
## 1st Qu.:0.0450 1st Qu.: 2.115 1st Qu.:17.23 Class :character   
## Median :0.3450 Median : 6.630 Median :19.30 Mode :character   
## Mean :0.3713 Mean : 5.591 Mean :19.35   
## 3rd Qu.:0.5400 3rd Qu.: 7.795 3rd Qu.:22.65   
## Max. :1.1200 Max. :15.000 Max. :25.50   
## Location Exchange cluster   
## Length:8 Length:8 Min. :1   
## Class :character Class :character 1st Qu.:1   
## Mode :character Mode :character Median :1   
## Mean :1   
## 3rd Qu.:1   
## Max. :1   
## ------------------------------------------------------------   
## factor(shot.data$cluster): 2  
## Symbol Name Market\_Cap Beta   
## Length:3 Length:3 Min. : 0.410 Min. :0.65   
## Class :character Class :character 1st Qu.: 1.505 1st Qu.:0.75   
## Mode :character Mode :character Median : 2.600 Median :0.85   
## Mean : 6.637 Mean :0.87   
## 3rd Qu.: 9.750 3rd Qu.:0.98   
## Max. :16.900 Max. :1.11   
## PE\_Ratio ROE ROA Asset\_Turnover Leverage   
## Min. :19.90 Min. : 3.90 Min. :1.400 Min. :0.6 Min. :0.000   
## 1st Qu.:22.95 1st Qu.:12.65 1st Qu.:2.850 1st Qu.:0.6 1st Qu.:0.725   
## Median :26.00 Median :21.40 Median :4.300 Median :0.6 Median :1.450   
## Mean :24.60 Mean :16.47 Mean :4.167 Mean :0.6 Mean :1.653   
## 3rd Qu.:26.95 3rd Qu.:22.75 3rd Qu.:5.550 3rd Qu.:0.6 3rd Qu.:2.480   
## Max. :27.90 Max. :24.10 Max. :6.800 Max. :0.6 Max. :3.510   
## Rev\_Growth Net\_Profit\_Margin Median\_Recommendation Location   
## Min. :-3.170 Min. : 2.600 Length:3 Length:3   
## 1st Qu.: 1.605 1st Qu.: 5.050 Class :character Class :character   
## Median : 6.380 Median : 7.500 Mode :character Mode :character   
## Mean : 5.733 Mean : 7.033   
## 3rd Qu.:10.185 3rd Qu.: 9.250   
## Max. :13.990 Max. :11.000   
## Exchange cluster   
## Length:3 Min. :2   
## Class :character 1st Qu.:2   
## Mode :character Median :2   
## Mean :2   
## 3rd Qu.:2   
## Max. :2   
## ------------------------------------------------------------   
## factor(shot.data$cluster): 3  
## Symbol Name Market\_Cap Beta   
## Length:4 Length:4 Min. : 0.780 Min. :0.2400   
## Class :character Class :character 1st Qu.: 1.095 1st Qu.:0.3000   
## Mode :character Mode :character Median : 2.230 Median :0.5350   
## Mean :13.100 Mean :0.5975   
## 3rd Qu.:14.235 3rd Qu.:0.8325   
## Max. :47.160 Max. :1.0800   
## PE\_Ratio ROE ROA Asset\_Turnover   
## Min. : 3.60 Min. :10.20 Min. :5.100 Min. :0.300   
## 1st Qu.:14.70 1st Qu.:10.95 1st Qu.:5.325 1st Qu.:0.300   
## Median :19.25 Median :13.15 Median :6.100 Median :0.400   
## Mean :17.68 Mean :14.57 Mean :6.200 Mean :0.425   
## 3rd Qu.:22.23 3rd Qu.:16.77 3rd Qu.:6.975 3rd Qu.:0.525   
## Max. :28.60 Max. :21.80 Max. :7.500 Max. :0.600   
## Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation  
## Min. :0.200 Min. :26.81 Min. :12.90 Length:4   
## 1st Qu.:0.305 1st Qu.:28.59 1st Qu.:13.20 Class :character   
## Median :0.635 Median :29.77 Median :14.20 Mode :character   
## Mean :0.635 Mean :30.14 Mean :15.65   
## 3rd Qu.:0.965 3rd Qu.:31.33 3rd Qu.:16.65   
## Max. :1.070 Max. :34.21 Max. :21.30   
## Location Exchange cluster   
## Length:4 Length:4 Min. :3   
## Class :character Class :character 1st Qu.:3   
## Mode :character Mode :character Median :3   
## Mean :3   
## 3rd Qu.:3   
## Max. :3   
## ------------------------------------------------------------   
## factor(shot.data$cluster): 4  
## Symbol Name Market\_Cap Beta   
## Length:4 Length:4 Min. :122.1 Min. :0.3500   
## Class :character Class :character 1st Qu.:129.9 1st Qu.:0.4325   
## Mode :character Mode :character Median :153.2 Median :0.4600   
## Mean :157.0 Mean :0.4800   
## 3rd Qu.:180.3 3rd Qu.:0.5075   
## Max. :199.5 Max. :0.6500   
## PE\_Ratio ROE ROA Asset\_Turnover   
## Min. :18.00 Min. :28.60 Min. :15.00 Min. :0.800   
## 1st Qu.:18.68 1st Qu.:37.60 1st Qu.:15.97 1st Qu.:0.875   
## Median :21.25 Median :43.10 Median :17.75 Median :0.950   
## Mean :22.23 Mean :44.42 Mean :17.70 Mean :0.950   
## 3rd Qu.:24.80 3rd Qu.:49.92 3rd Qu.:19.48 3rd Qu.:1.025   
## Max. :28.40 Max. :62.90 Max. :20.30 Max. :1.100   
## Leverage Rev\_Growth Net\_Profit\_Margin Median\_Recommendation  
## Min. :0.100 Min. : 9.37 Min. :14.10 Length:4   
## 1st Qu.:0.145 1st Qu.:15.36 1st Qu.:16.95 Class :character   
## Median :0.220 Median :19.61 Median :19.50 Mode :character   
## Mean :0.220 Mean :18.53 Mean :19.57   
## 3rd Qu.:0.295 3rd Qu.:22.79 3rd Qu.:22.12   
## Max. :0.340 Max. :25.54 Max. :25.20   
## Location Exchange cluster   
## Length:4 Length:4 Min. :4   
## Class :character Class :character 1st Qu.:4   
## Mode :character Mode :character Median :4   
## Mean :4   
## 3rd Qu.:4   
## Max. :4   
## ------------------------------------------------------------   
## factor(shot.data$cluster): 5  
## Symbol Name Market\_Cap Beta   
## Length:2 Length:2 Min. : 7.58 Min. :0.4000   
## Class :character Class :character 1st Qu.:19.75 1st Qu.:0.4025   
## Mode :character Mode :character Median :31.91 Median :0.4050   
## Mean :31.91 Mean :0.4050   
## 3rd Qu.:44.08 3rd Qu.:0.4075   
## Max. :56.24 Max. :0.4100   
## PE\_Ratio ROE ROA Asset\_Turnover Leverage   
## Min. :56.5 Min. :12.90 Min. :5.50 Min. :0.600 Min. :0.3500   
## 1st Qu.:63.0 1st Qu.:13.05 1st Qu.:5.55 1st Qu.:0.675 1st Qu.:0.4125   
## Median :69.5 Median :13.20 Median :5.60 Median :0.750 Median :0.4750   
## Mean :69.5 Mean :13.20 Mean :5.60 Mean :0.750 Mean :0.4750   
## 3rd Qu.:76.0 3rd Qu.:13.35 3rd Qu.:5.65 3rd Qu.:0.825 3rd Qu.:0.5375   
## Max. :82.5 Max. :13.50 Max. :5.70 Max. :0.900 Max. :0.6000   
## Rev\_Growth Net\_Profit\_Margin Median\_Recommendation Location   
## Min. : 9.16 Min. :5.50 Length:2 Length:2   
## 1st Qu.:10.62 1st Qu.:5.95 Class :character Class :character   
## Median :12.08 Median :6.40 Mode :character Mode :character   
## Mean :12.08 Mean :6.40   
## 3rd Qu.:13.54 3rd Qu.:6.85   
## Max. :15.00 Max. :7.30   
## Exchange cluster   
## Length:2 Min. :5   
## Class :character 1st Qu.:5   
## Mode :character Median :5   
## Mean :5   
## 3rd Qu.:5   
## Max. :5

#Median the calculation

Recm.table <- table(shot.data$cluster,shot.data$Median\_Recommendation)  
names(dimnames(Recm.table)) <- c("Cluster","Recommendation")  
Recm.table <- addmargins(Recm.table)  
Recm.table

## Recommendation  
## Cluster Hold Moderate Buy Moderate Sell Strong Buy Sum  
## 1 4 1 2 1 8  
## 2 2 1 0 0 3  
## 3 0 2 2 0 4  
## 4 2 2 0 0 4  
## 5 1 1 0 0 2  
## Sum 9 7 4 1 21

There are 21 suggestions total: 4 holds, 7 moderate buys, 4 moderate sells, and 1 strong buy. Cluster 5 combines the purchase and sell recommendations that are in opposition to the other three. Clusters 1, 2, and 3 contain all of the information about mod purchases and holds.For Cluster 4, there is a moderate buy and a moderate sell recommendation.

#A cluster breakdown of the organization’s headquarters based on the aggregated data.

location\_table <- table(shot.data$cluster,shot.data$Location)  
names(dimnames(location\_table)) <- c("Cluster","Location")  
location\_table <- addmargins(location\_table)  
location\_table

## Location  
## Cluster CANADA FRANCE GERMANY IRELAND SWITZERLAND UK US Sum  
## 1 0 0 0 0 1 2 5 8  
## 2 0 0 1 0 0 0 2 3  
## 3 0 1 0 1 0 0 2 4  
## 4 0 0 0 0 0 1 3 4  
## 5 1 0 0 0 0 0 1 2  
## Sum 1 1 1 1 1 3 13 21

In total, there are 21 companies: 13 are located in the United States, 3 in the United Kingdom, and 1 in each of Canada, France, Germany, Ireland, and Switzerland. Group 5 includes the United States, the United Kingdom, and Switzerland. Cluster 2 includes the US and Canada. Cluster 3 includes the US and Britain, whereas Cluster 4 includes France, Ireland, and the US.

#making a data frame with the combined data and building the exchange table to show a summary of the stock exchange values for each cluster.

exc\_table <- table(shot.data$cluster,shot.data$Exchange)  
names(dimnames(exc\_table)) <- c("Cluster","Exchange")  
exc\_table <- addmargins(exc\_table)  
exc\_table

## Exchange  
## Cluster AMEX NASDAQ NYSE Sum  
## 1 0 0 8 8  
## 2 1 1 1 3  
## 3 0 0 4 4  
## 4 0 0 4 4  
## 5 0 0 2 2  
## Sum 1 1 19 21

There are a total of 21 companies: 1 Amex, 1 Nasdaq, and 19 NYSE. The NYSE is the lone entity in Cluster 5. Cluster 2 comprises all three. The sole organization in clusters 1, 3, and 4 is NYSE. ##Q3. Provide an appropriate name for each cluster using any or all of the variables in the dataset.

Cluster 1: “High Market Cap - High ROE” greater Net-profit Margin is made up of all US companies listed on the NYSE that are either held or acquired.

Cluster 2 is known as the “Hold Cluster” since it only consists of companies that are listed on the New York Stock Exchange.

Cluster 3: On stock exchanges, this is referred to as “Low Market Cap & Less ROA” and is either buy or hold.

Cluster 4: “High PE ratio / Small Net Profit Margin” is the category this falls under.

Cluster 5: This cluster may be referred to as “Moderate Buy & Moderate Sell.”

#Range the scaling

data\_range <- scale(KV\_A4\_fml[,3:11])  
summary(data\_range)

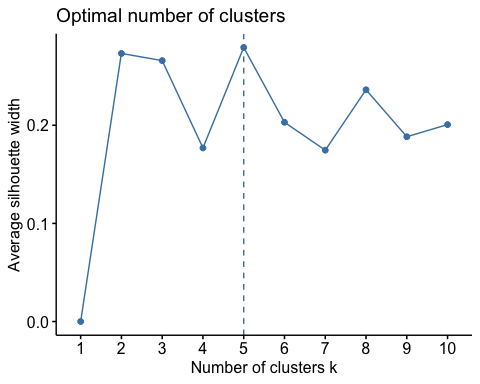
## Market\_Cap Beta PE\_Ratio ROE   
## Min. :-0.9768 Min. :-1.3466 Min. :-1.3404 Min. :-1.4515   
## 1st Qu.:-0.8763 1st Qu.:-0.6844 1st Qu.:-0.4023 1st Qu.:-0.7223   
## Median :-0.1614 Median :-0.2560 Median :-0.2429 Median :-0.2118   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.2762 3rd Qu.: 0.4841 3rd Qu.: 0.1495 3rd Qu.: 0.3450   
## Max. : 2.4200 Max. : 2.2758 Max. : 3.4971 Max. : 2.4597   
## ROA Asset\_Turnover Leverage Rev\_Growth   
## Min. :-1.7128 Min. :-1.8451 Min. :-0.74966 Min. :-1.4971   
## 1st Qu.:-0.9047 1st Qu.:-0.4613 1st Qu.:-0.54487 1st Qu.:-0.6328   
## Median : 0.1289 Median :-0.4613 Median :-0.31449 Median :-0.3621   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.00000 Mean : 0.0000   
## 3rd Qu.: 0.8430 3rd Qu.: 0.9225 3rd Qu.: 0.01828 3rd Qu.: 0.7693   
## Max. : 1.8389 Max. : 1.8451 Max. : 3.74280 Max. : 1.8862   
## Net\_Profit\_Margin   
## Min. :-1.99560   
## 1st Qu.:-0.68504   
## Median : 0.06168   
## Mean : 0.00000   
## 3rd Qu.: 0.82364   
## Max. : 1.49416

head(data\_range)

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover  
## [1,] 0.1840960 -0.80125356 -0.04671323 0.04009035 0.2416121 -5.121077e-16  
## [2,] -0.8544181 -0.45070513 3.49706911 -0.85483986 -0.9422871 9.225312e-01  
## [3,] -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700 9.225312e-01  
## [4,] 0.1702742 -0.02225704 -0.24290879 0.10638147 0.9181259 9.225312e-01  
## [5,] -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461 -4.612656e-01  
## [6,] -0.6953818 2.27578267 0.14948233 -1.45146000 -1.7127612 -4.612656e-01  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## [1,] -0.2120979 -0.5277675 0.06168225  
## [2,] 0.0182843 -0.3811391 -1.55366706  
## [3,] -0.4040831 -0.5721181 -0.68503583  
## [4,] -0.7496565 0.1474473 0.35122600  
## [5,] -0.3144900 1.2163867 -0.42597037  
## [6,] -0.7496565 -1.4971443 -1.99560225

#trying out or researching various techniques to see if they could also provide more effective clustering?

fviz\_nbclust(data\_range,FUN = kmeans,method = "silhouette")



kmeans\_2.df <- kcca(data\_scaled.df,k=5,kccaFamily("kmeans"))  
kmeans\_2.df

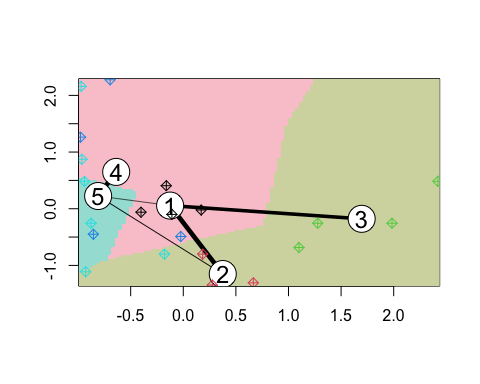
## kcca object of family 'kmeans'   
##   
## call:  
## kcca(x = data\_scaled.df, k = 5, family = kccaFamily("kmeans"))  
##   
## cluster sizes:  
##   
## 1 2 3 4 5   
## 4 3 4 4 6

clusters(kmeans\_2.df)

## [1] 2 4 5 1 5 4 1 4 5 2 3 5 3 5 3 2 3 4 1 5 1

#Using the predict() function.

index\_cluster.df <- predict(kmeans\_2.df)  
image(kmeans\_2.df)  
points(data\_scaled.df,col = index\_cluster.df,pch = 9,cex = 1.0)

 Here, the kcca method is utilized to perform a kmeans cluster on k = 5, as opposed to the kmeans function from basic R. While the clustering has the same size as the underlying R approach, the assignment between the points is different. The clustering graph shows that there is more ambiguity in the categorization than we would like, especially between groups 1, 2, and 3.

#PLOTTING K-MEDIANS AND CLUSTERS.

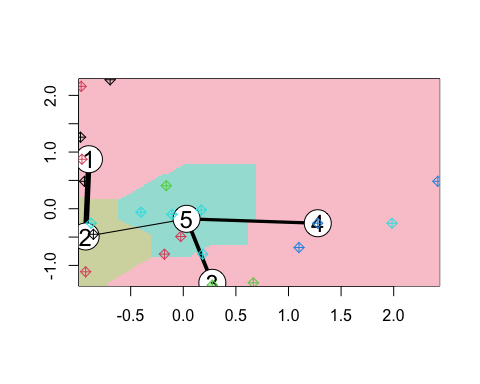
kmeans\_2.df <- kcca(data\_scaled.df,k = 5,kccaFamily("kmedians"))  
kmeans\_2.df

## kcca object of family 'kmedians'   
##   
## call:  
## kcca(x = data\_scaled.df, k = 5, family = kccaFamily("kmedians"))  
##   
## cluster sizes:  
##   
## 1 2 3 4 5   
## 4 5 3 3 6

clusters(kmeans\_2.df)

## [1] 5 1 5 5 2 1 5 1 2 3 4 1 5 2 4 3 4 2 5 2 3

index\_cluster.df <- predict(kmeans\_2.df)  
image(kmeans\_2.df)  
points(data\_scaled.df,col = index\_cluster.df,pch = 9,cex = 1.0)

 The five clusters have the following sizes when the KCC is converted from Kmeans to Kmedian: 4, 5, 3, 3, and 6.However, the clustering is less clear. Even though it’s unclear if a better cluster exists, we want to look over the new data to see if there are any other methods or tools we might utilize to improve the visual cluster.