FML_Assignment _ 4

Jhansi Bussa

```r
knitr::opts_chunk$set(echo = TRUE)
#applying the knit functions
data_pharma <- read.csv("C:/Users/jbussa/Downloads/Pharmaceuticals.csv")
#reading pharmacueticals data
#loading library tidyverse for transforming data
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr      1.1.0     ✓ readr      2.1.4
## ✓ forcats    1.0.0     ✓ stringr    1.5.0
## ✓ ggplot2    3.4.1     ✓ tibble     3.1.8
## ✓ lubridate  1.9.2     ✓ tidyr      1.3.0
## ✓ purrr      1.0.1
## ── Conflicts ─────────────────────────────────────────
tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## i Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force
all conflicts to become errors
```

```r
# Using the factoextra library to begin data extraction and visualization

library(factoextra)
```

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

```r
#Loading ISLR libaray for statistical learning
library(ISLR)
#Loading the flexclust to implement cluster techniques
library(flexclust)
```

```
## Loading required package: grid
## Loading required package: lattice
## Loading required package: modeltools
## Loading required package: stats4
```

```r
#loading the library caret for training model
library(caret)
```

```
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
```

```
##
##      lift
```

##Question 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.

```
#Performing z-score scaling for pharmaceuticals dataframe
Scaled_pharma_data <- scale(data_pharma[,3:11])
summary(Scaled_pharma_data)

##    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
```

```
#Performing the range scaling for the the
dataframe

range_pharma_data <- scale(data_pharma[,3:11])
#summarizing the scaled data frame
summary(range_pharma_data)
##    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
```
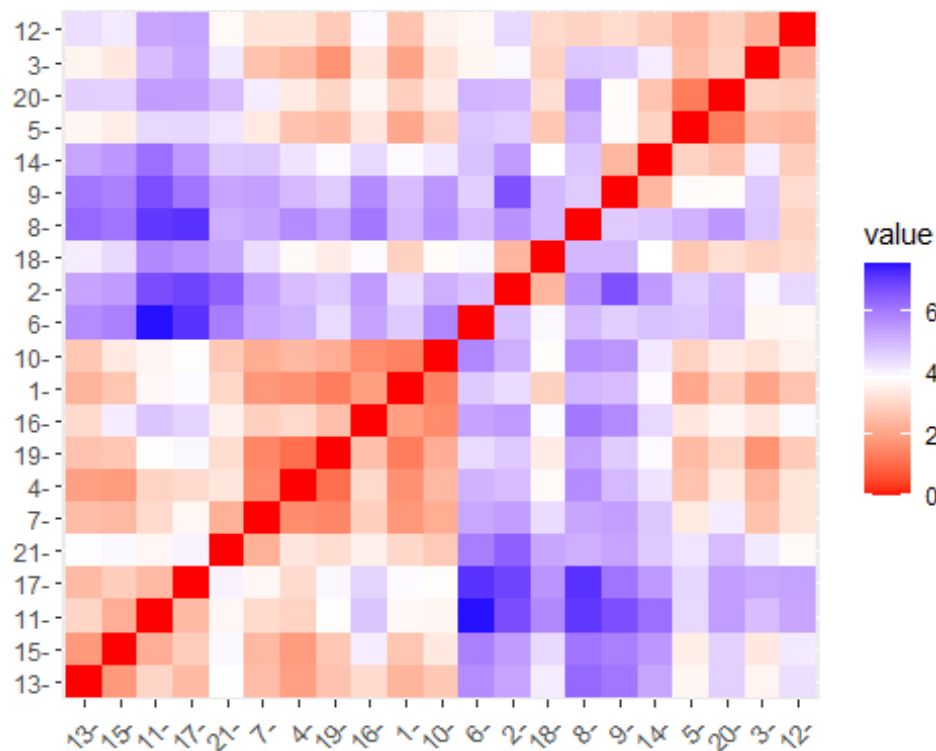
```
##  Mean    : 0.0000     Mean    : 0.0000     Mean    : 0.00000     Mean    : 0.0000
##  3rd Qu.: 0.8430     3rd Qu.: 0.9225     3rd Qu.: 0.01828     3rd Qu.: 0.7693
##  Max.    : 1.8389     Max.    : 1.8451     Max.    : 3.74280     Max.    : 1.8862
##  Net_Profit_Margin
##  Min.    :-1.99560
##  1st Qu.:-0.68504
##  Median : 0.06168
##  Mean    : 0.00000
##  3rd Qu.: 0.82364
##  Max.    : 1.49416
```

```r
#Finding the scaled pharmaceutical data's distance
distance <- get_dist(Scaled_pharma_data)
fviz_dist(distance) #visualizing the distance between rows of the distance
matrix
```
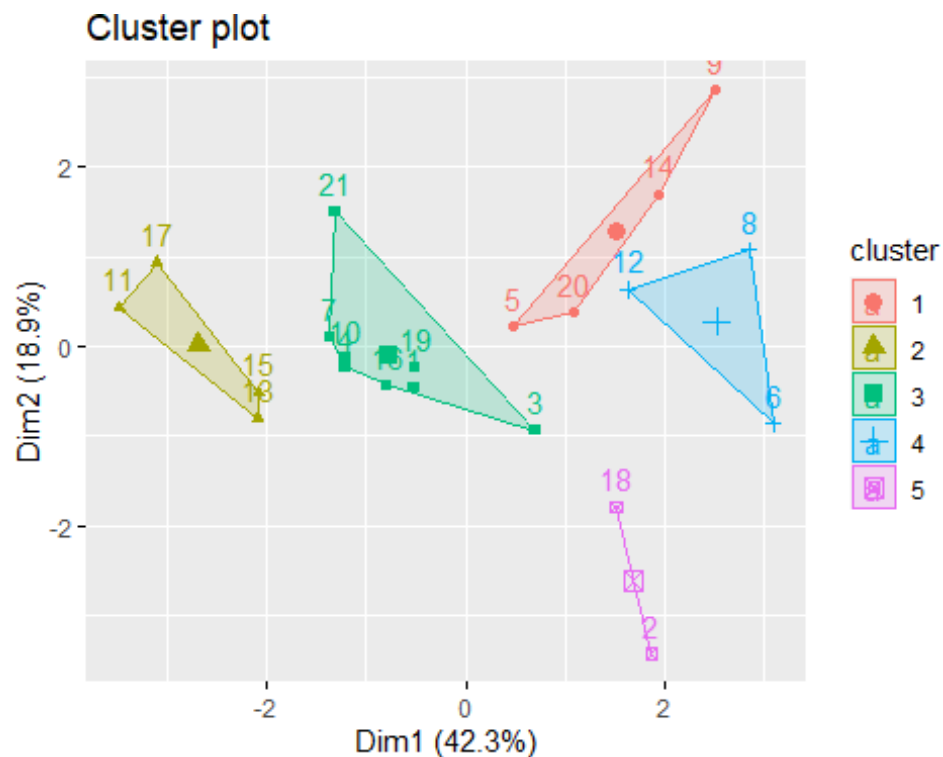


```r
#Applying K-means clustering for the scaled data
kmeans_1 <- kmeans(Scaled_pharma_data, centers = 5, nstart = 25)
#Visualizing the clusters on a graph
fviz_cluster(kmeans_1, data = Scaled_pharma_data)
```

Cluster plot


```
print(kmeans_1)

## K-means clustering with 5 clusters of sizes 4, 4, 8, 3, 2
##
## Cluster means:
##      Market_Cap        Beta    PE_Ratio         ROE         ROA Asset_Turnover
## 1 -0.76022489   0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 2  1.69558112  -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 3 -0.03142211  -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
## 4 -0.87051511   1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 5 -0.43925134  -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
##       Leverage Rev_Growth Net_Profit_Margin
## 1  0.06308085  1.5180158      -0.006893899
## 2 -0.46807818  0.4671788       0.591242521
## 3 -0.27449312 -0.7041516       0.556954446
## 4  1.36644699 -0.6912914      -1.320000179
## 5 -0.14170336 -0.1168459      -1.416514761
##
## Clustering vector:
##  [1] 3 5 3 3 1 4 3 4 1 3 2 4 2 1 2 3 2 5 3 1 3
##
## Within cluster sum of squares by cluster:
## [1] 12.791257  9.284424 21.879320 15.595925  2.803505
##  (between_SS / total_SS =  65.4 %)
##
## Available components:
##
```
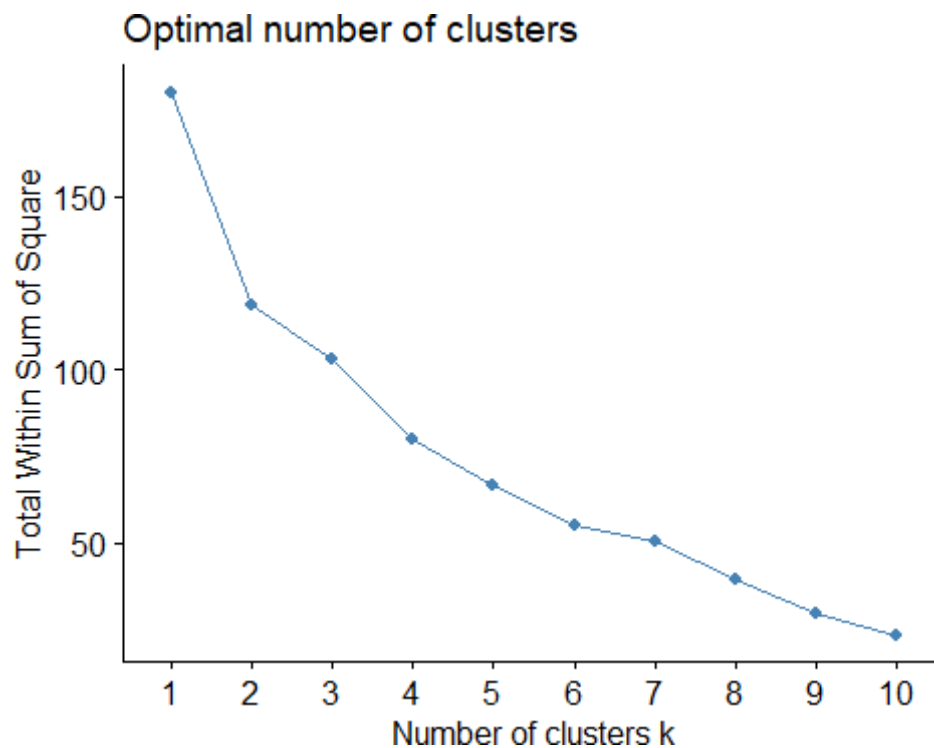
```
## [1] "cluster"       "centers"       "totss"          "withinss"
"tot.withinss"
## [6] "betweenss"     "size"          "iter"           "ifault"
```

##OVERVIEW of the K-means Clustering:-

The 21 pharmaceutical businesses' nine variables are provided with information by the summary function applied to the k-means data. Five clusters in total with five centroid points have the following sizes: (2, 3, 4, 4, 8, 10).
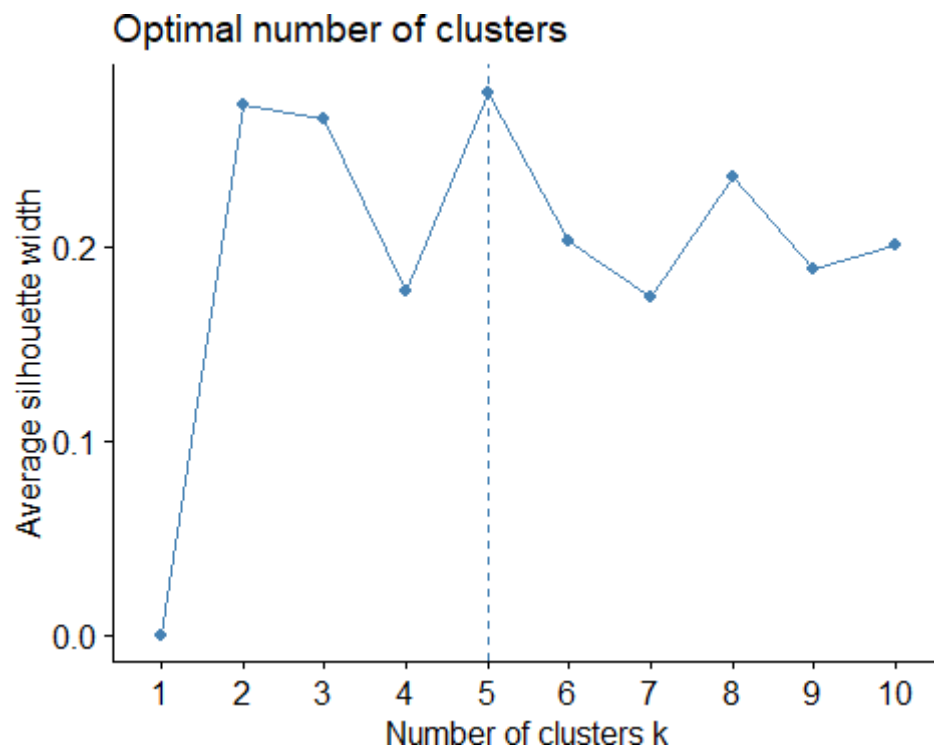
##Question b - Interpret the clusters with respect to the numerical variables used in forming the clusters

```
#plotting number of clusters vs the total value
fviz_nbclust(Scaled_pharma_data, kmeans, method = "wss")
```


Optimal number of clusters

** Here in this plot we can clearly see that the graph is not forming an elbow shape and it is getting flattened slightly at k=4 and k=6

```
#figuring out the number of optimal clusters by plotting the number
of clusters
#against average silhoutte width
fviz_nbclust(Scaled_pharma_data, kmeans, method = "silhouette")
```

Optimal number of clusters

The silhoutte graph easily leads us to the conclusion that the best and optimal solution for this k-means clustering is five clusters..

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

```
# Determining the mean value based on the data that is plotted in the
clusters itself
#perfroming the aggregate function for the pharma data
aggregate(data_pharma[3:11], by=list(cluster=kmeans_1$cluster), mean)

##   cluster Market_Cap    Beta PE_Ratio     ROE      ROA Asset_Turnover
## 1       1  13.100000 0.59750  17.6750 14.57500  6.200000         0.4250
## 2       2 157.017500 0.48000  22.2250 44.42500 17.700000         0.9500
## 3       3  55.810000 0.41375  20.2875 28.73750 12.687500         0.7375
## 4       4   6.636667 0.87000  24.6000 16.46667  4.166667         0.6000
## 5       5  31.910000 0.40500  69.5000 13.20000  5.600000         0.7500
##   Leverage Rev_Growth Net_Profit_Margin
## 1 0.635000   30.142500         15.650000
## 2 0.220000   18.532500         19.575000
## 3 0.371250    5.591250         19.350000
## 4 1.653333    5.733333          7.033333
## 5 0.475000   12.080000          6.400000

#perfroming the merging of the data frames using cbind
temp_data <- cbind(data_pharma, cluster = kmeans_1$cluster)
tibble(temp_data)
```

```
## # A tibble: 21 × 15
##    Symbol Name          Marke…¹  Beta PE_Ra…²   ROE   ROA Asset…³ Lever…⁴
Rev_G…⁵
##    <chr>  <chr>          <dbl> <dbl>   <dbl> <dbl> <dbl>   <dbl>   <dbl>
<dbl>
##  1 ABT    Abbott Labo…   68.4   0.32    24.7  26.4  11.8     0.7    0.42
7.54
##  2 AGN    Allergan, I…    7.58  0.41    82.5  12.9   5.5     0.9    0.6
9.16
##  3 AHM    Amersham plc    6.3   0.46    20.7  14.9   7.8     0.9    0.27
7.05
##  4 AZN    AstraZeneca…   67.6   0.52    21.5  27.4  15.4     0.9    0
15
##  5 AVE    Aventis        47.2   0.32    20.1  21.8   7.5     0.6    0.34
26.8
##  6 BAY    Bayer AG       16.9   1.11    27.9   3.9   1.4     0.6    0
-3.17
##  7 BMY    Bristol-Mye…   51.3   0.5     13.9  34.8  15.1     0.9    0.57
2.7
##  8 CHTT   Chattem, Inc    0.41  0.85    26    24.1   4.3     0.6    3.51
6.38
##  9 ELN    Elan Corpor…    0.78  1.08     3.6  15.1   5.1     0.3    1.07
34.2
## 10 LLY    Eli Lilly a…   73.8   0.18    27.9  31    13.5     0.6    0.53
6.21
## # … with 11 more rows, 5 more variables: Net_Profit_Margin <dbl>,
## #   Median_Recommendation <chr>, Location <chr>, Exchange <chr>, cluster
<int>,
## #   and abbreviated variable names ¹Market_Cap, ²PE_Ratio, ³
Asset_Turnover,
## #   ⁴Leverage, ⁵Rev_Growth
```

```
#Summarizing the detailed beakdown by cluster
by(temp_data, factor(temp_data$cluster), summary)
```

```
## factor(temp_data$cluster): 1
##     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.    1
## Class :character   Class :character   1st Qu.:1
## Mode  :character   Mode  :character   Median  1
##                                       Mean    1
##                                       3rd Qu.:1
##                                       Max.    1
## ----------------------------------------------------------------------------
## factor(temp_data$cluster): 2
##     Symbol            Name             Market_Cap         Beta
## Length:4         Length:4         Min.    :122.1   Min.    :0.3500
## Class :character   Class :character   1st Qu.:129.9   1st Qu.:0.4325
## Mode  :character   Mode  :character   Median :153.2   Median :0.4600
##                                       Mean    :157.0   Mean    :0.4800
##                                       3rd Qu.:180.3   3rd Qu.:0.5075
##                                       Max.    :199.5   Max.    :0.6500
##     PE_Ratio          ROE              ROA          Asset_Turnover
## Min.    :18.00   Min.    :28.60   Min.    :15.00   Min.    :0.800
## 1st Qu.:18.68   1st Qu.:37.60   1st Qu.:15.97   1st Qu.:0.875
## Median :21.25   Median :43.10   Median :17.75   Median :0.950
## Mean    :22.23   Mean    :44.42   Mean    :17.70   Mean    :0.950
## 3rd Qu.:24.80   3rd Qu.:49.92   3rd Qu.:19.48   3rd Qu.:1.025
## Max.    :28.40   Max.    :62.90   Max.    :20.30   Max.    :1.100
##     Leverage        Rev_Growth      Net_Profit_Margin Median_Recommendation
## Min.    :0.100   Min.    : 9.37   Min.    :14.10    Length:4
## 1st Qu.:0.145   1st Qu.:15.36   1st Qu.:16.95    Class :character
## Median :0.220   Median :19.61   Median :19.50    Mode  :character
## Mean    :0.220   Mean    :18.53   Mean    :19.57
## 3rd Qu.:0.295   3rd Qu.:22.79   3rd Qu.:22.12
## Max.    :0.340   Max.    :25.54   Max.    :25.20
##     Location         Exchange           cluster
## Length:4         Length:4         Min.    2
## Class :character   Class :character   1st Qu.:2
## Mode  :character   Mode  :character   Median  2
##                                       Mean    2
##                                       3rd Qu.:2
##                                       Max.    2
## ----------------------------------------------------------------------------
## factor(temp_data$cluster): 3
##     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.   3
##  Class :character  Class :character  1st Qu.:3
##  Mode  :character  Mode  :character  Median  3
##                                      Mean    3
##                                      3rd Qu.:3
##                                      Max.    3
## -----------------------------------------------------------------------------
## factor(temp_data$cluster): 4
##     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.   4
##  Class :character    1st Qu.:4
##  Mode  :character    Median :4
##                      Mean   4
##                      3rd Qu.:4
##                      Max.   4
##
## factor(temp_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 calculation
recommend_table <- table(temp_data$cluster, temp_data$Median_Recommendation)
```

```r
names(dimnames(recommend_table)) <- c("Cluster", "Recommendation")
recommend_table <- addmargins(recommend_table)
recommend_table
```

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

There are 21 suggestions in total: one strong buy, seven moderate buys, nine holds, and four moderate sells. Cluster 5 has all four suggestions, including the opposing advice on buys and sells. Clusters 1, 2, and 3 only contain information about mod purchases and holds. For Cluster 4, recommendations for a moderate buy and a moderate sell are provided.

```r
#Location of firm headquarter's breakdown of clusters based on the merged
data
location_table <- table(temp_data$cluster, temp_data$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      1       0       1           0  0  2   4
##     2        0      0       0       0           0  1  3   4
##     3        0      0       0       0           1  2  5   8
##     4        0      0       1       0           0  0  2   3
##     5        1      0       0       0           0  0  1   2
##     Sum      1      1       1       1           1  3 13  21
```

There are a total of 21 firms: 13 are located in the United States, 3 in the United Kingdom, and 1 in each of Canada, France, Germany, Ireland, and Switzerland. In Cluster 5, the US, UK, and Switzerland are all mentioned. US and Germany are in Cluster 2. USA and Canada are located in Group 1. Cluster 3 includes the US and the UK. Ireland, France, and the US comprise Cluster 4.

```r
#summarizing the stock exchange values for each cluster
#creating a data frame for the merged data and initializing the exchange
table
exchange_table <- table(temp_data$cluster, temp_data$Exchange)
names(dimnames(exchange_table)) <- c("Cluster", "Exchange")
exchange_table <- addmargins(exchange_table)
exchange_table
```

```
##         Exchange
## Cluster AMEX NASDAQ NYSE Sum
##     1      0      0    4   4
```

```
##     2          0          0     4     4
##     3          0          0     8     8
##     4          1          1     1     3
##     5          0          0     2     2
##     Sum        1          1    19    21
```

##In total, there are 21 companies: 1 Amex, 1 Nasdaq, and 19 NYSE. Just the NYSE is in Cluster 5. Cluster 2 includes all three. The only thing in clusters 1, 3, and 4 is NYSE.

####In total, there are 21 companies: 1 Amex, 1 Nasdaq, and 19 NYSE. Just the NYSE is in Cluster 5. Cluster 2 includes all three. The only thing in clusters 1, 3, and 4 is NYSE.

##Question d: Using any or all of the variables in the dataset, give each cluster a suitable name.

Answer: Cluster 1: - The term "Small_Net_Profit_Margin-High PE ratio" can be used to refer to cluster 3. They're all NAM businesses.

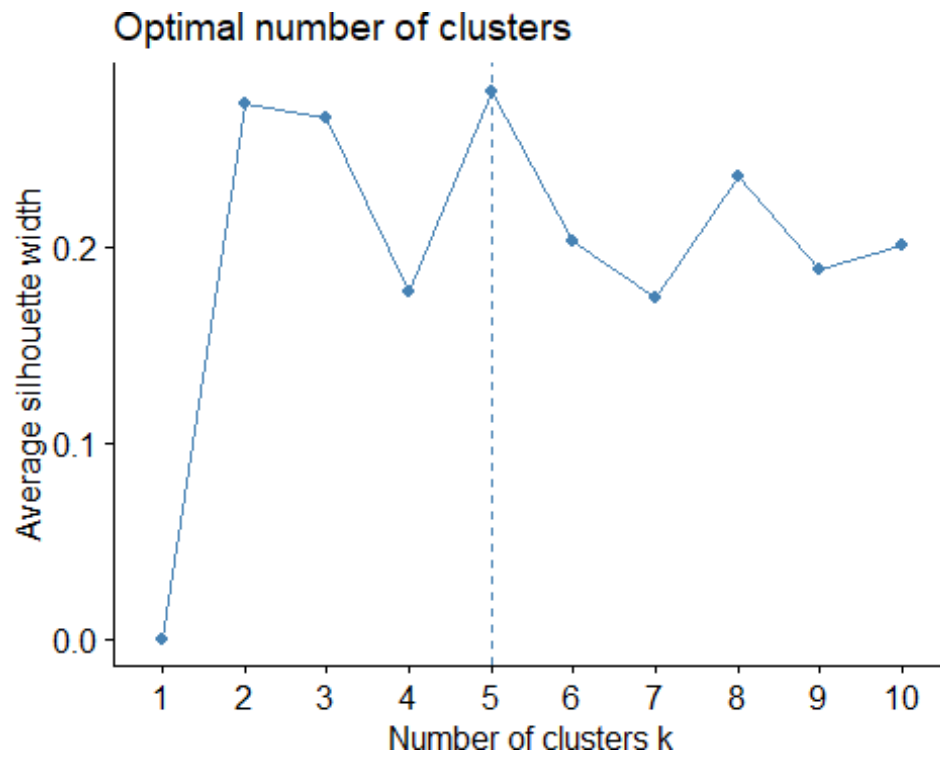Cluster 2: "Low_Market_Cap & Less_ROA" - Hold or Buy exchanges is the name of cluster number two.

Cluster 3: The fourth cluster is titled "High Market Cap - more RoE - more RoA - High Asset Turnover - more NetProfitMargin"; all of the US companies in this cluster are NYSE-listed and can be bought or held.

Cluster 4: This cluster is defined by "high revenue growth - mixed recommendation & low RoE & minimum asset turnover & least PE ratio." All are NYSE-listed American or European corporations.
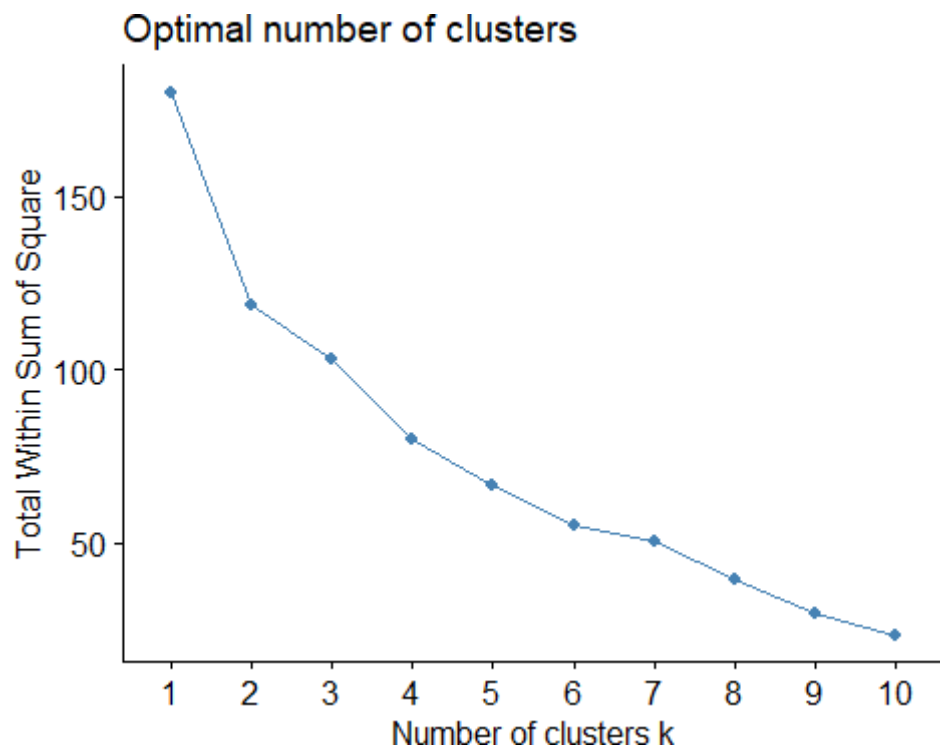
Cluster 5: "Least_Revenue_growth" is the name of the fifth cluster. It was largely made up of NYSE-listed US corporations.

##Also Trying or Investigating Whether or Not Other Algorithms Can Perform Better Clustering -?

```
fviz_nbclust(range_pharma_data, FUN = kmeans, method = "silhouette")
```

Optimal number of clusters


```
fviz_nbclust(range_pharma_data, kmeans, method = "wss")
```

Optimal number of clusters

PLOTTING THE KMEANS AND THE CLUSTERS

```
kmeans_2 = kcca(Scaled_pharma_data, k=5, kccaFamily("kmeans"))
kmeans_2

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

clusters(kmeans_2)

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

#Applying the predict() function
clusters_index <- predict(kmeans_2)
image(kmeans_2)
points(Scaled_pharma_data, col=clusters_index, pch=19, cex=1.0)
```
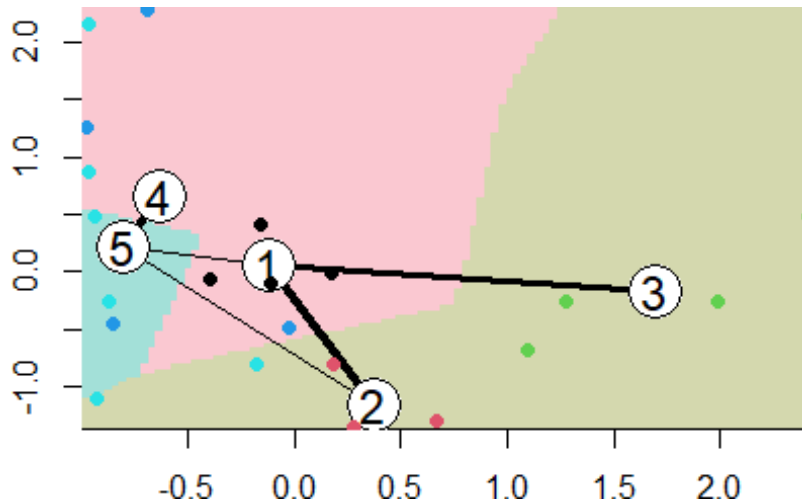

Here, we use the kcca algorithm to perform a kmeans cluster on k = 5, instead of the kmeans function from base R. Compared to the base R method, the clustering has the same size but a different point assignment. In particular, the grouping between clusters 1, 2, and 3 is not as distinct as we would like, as the clustering graph shows.
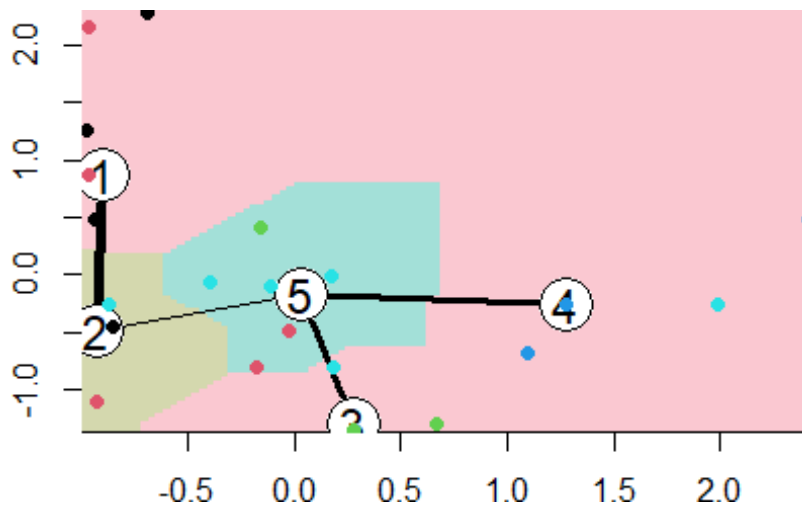
PLOTTING THE KMEDIANS AND THE CLUSTERS.

```
kmeans_2 = kcca(Scaled_pharma_data, k=5, kccaFamily("kmedians"))
kmeans_2

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

clusters(kmeans_2) #clusteing

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

clusters_index <- predict(kmeans_2)
image(kmeans_2)
points(Scaled_pharma_data, col=clusters_index, pch=19, cex=1.0)
```


##The sizes of the five clusters are 4, 5, 3, 3, and 6 if we convert from kmeans to kmedian in KCC. However, the clustering is less pronounced. Though it's unclear if a better cluster exists, we are looking into the additional data to see if there are any other methods or resources we can use to improve the visual cluster.