

ASSIGNMENT-4

vivek rao kathheragandla

2024-03-12

```
library("tidyverse")
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## Warning: package 'forcats' was built under R version 4.3.2
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.3      ✓ tibble     3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr      1.0.2
## — Conflicts ————— tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```
library("ISLR")
library("factoextra")
```

```
## Warning: package 'factoextra' was built under R version 4.3.3
```

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

```
library("flexclust")
```

```
## Warning: package 'flexclust' was built under R version 4.3.3
```

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

```
library("ggplot2")
library("dplyr")
library("caret")
```

```
## Warning: package 'caret' was built under R version 4.3.3
```

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

#applying the knit functions

```
data_pharma <- read.csv("C:\\Users\\KATHTHERAGANDLA VIVEK\\OneDrive\\Documents\\FML\\Assignment_4\\Pharmaceuticals (1).csv")
head(data_pharma)
```

```
## 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
```

```
summary(data_pharma)
```

```
##      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
##
##
##
```

##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 the pahrmaceuticals dataframe
Scaling_pharma_data <- scale(data_pharma[,3:11])
summary(Scaling_pharma_data)
```

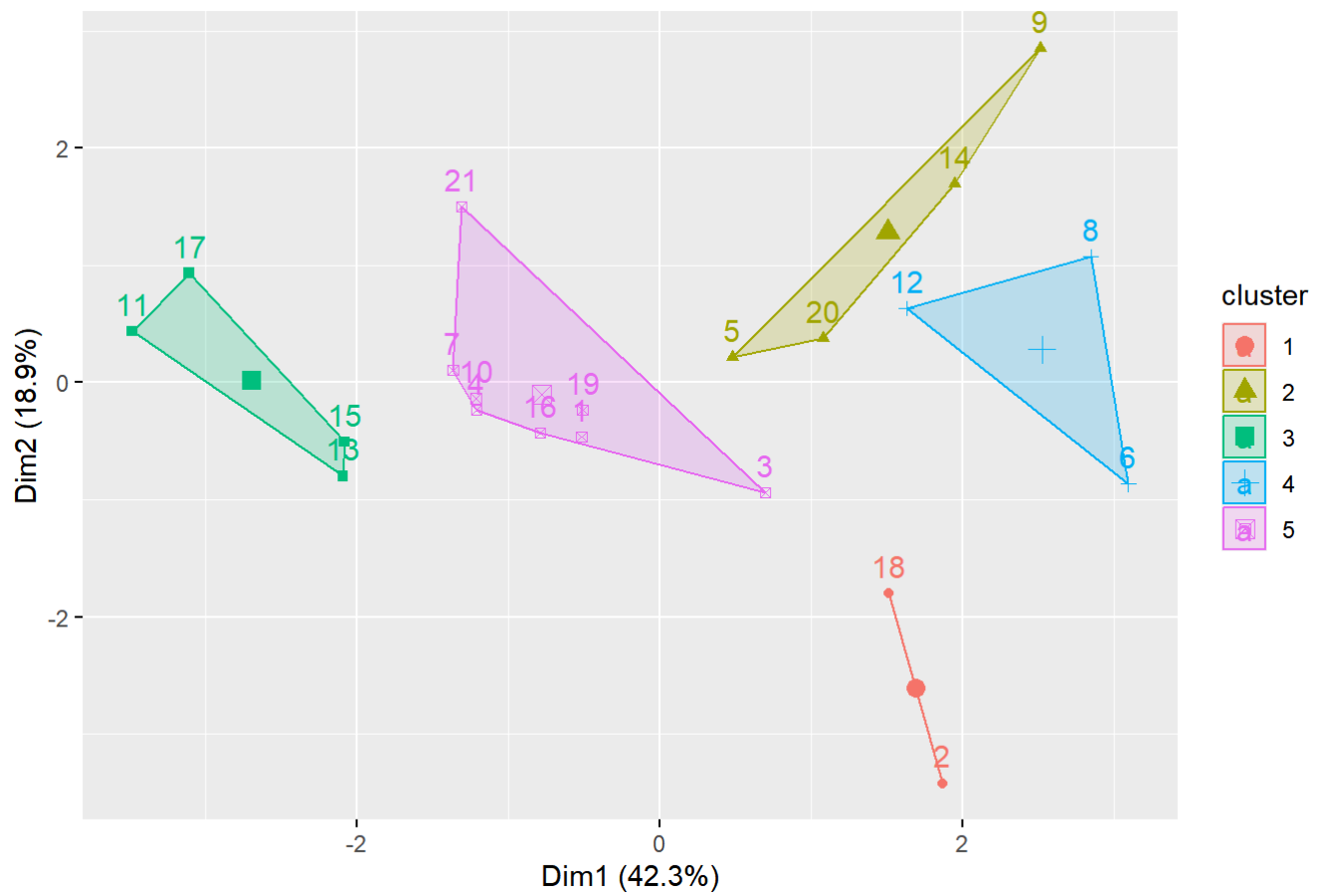
##	Market_Cap	Beta	PE_Ratio	ROE
##	Min. : -0.9768	Min. : -1.3466	Min. : -1.3404	Min. : -1.4515
##	1st Qu.: -0.8763	1st Qu.: -0.6844	1st Qu.: -0.4023	1st Qu.: -0.7223
##	Median : -0.1614	Median : -0.2560	Median : -0.2429	Median : -0.2118
##	Mean : 0.0000	Mean : 0.0000	Mean : 0.0000	Mean : 0.0000
##	3rd 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 range scaling for the the dataframe
range_pharma <- scale(data_pharma[,3:11])
#summarizing the scaled data frame
summary(range_pharma)
```

##	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			

```
#applying K-means clustering for the scaled data
kmeans_1 <- kmeans(Scaling_pharma_data, centers = 5, nstart = 25)
#visualizing the clusters on agraph
fviz_cluster(kmeans_1, data = Scaling_pharma_data)
```

Cluster plot



```
#Assuming kmeans_1 contains the results of your k-means clustering
print(kmeans_1)
```

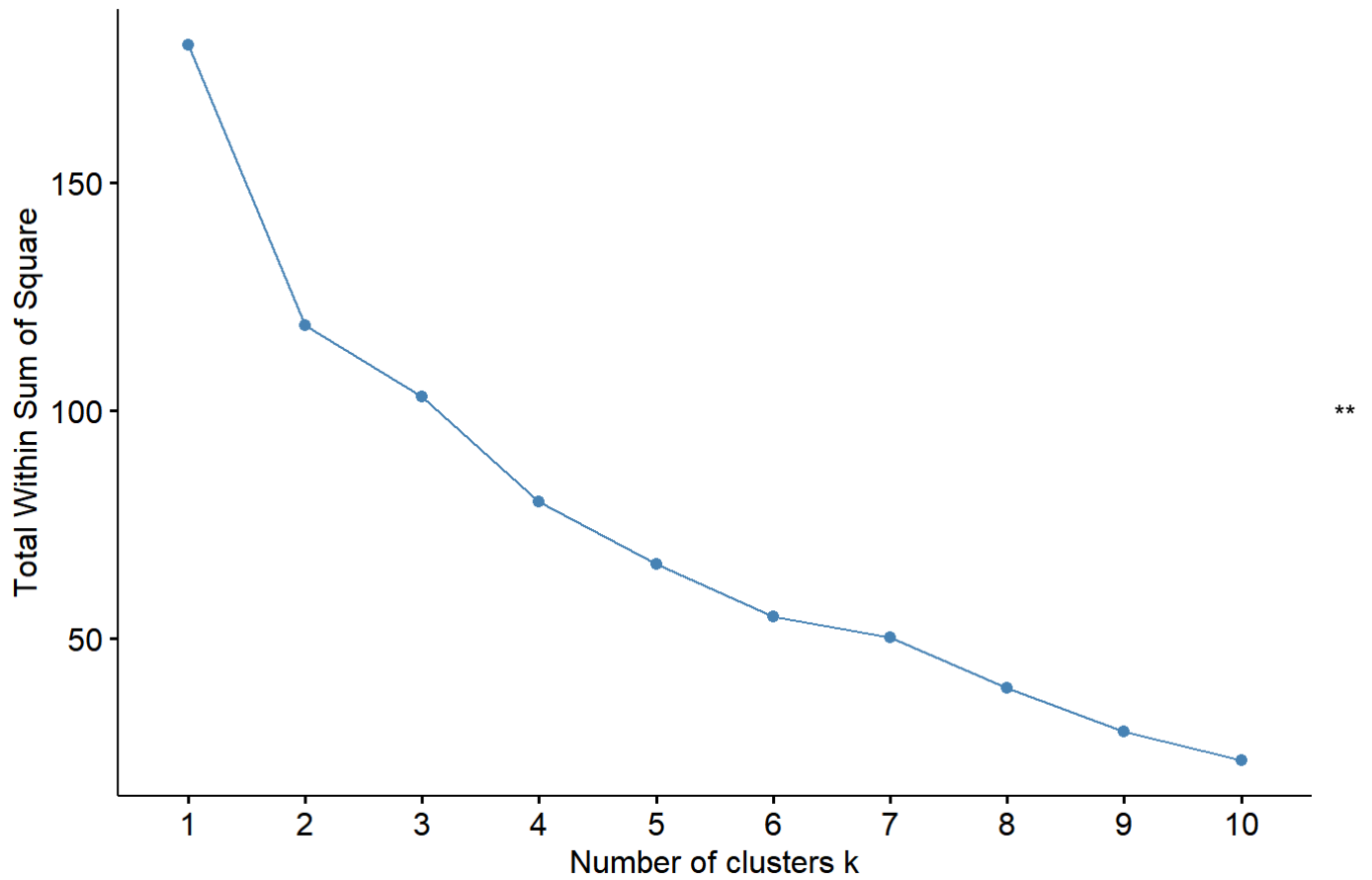
```
## K-means clustering with 5 clusters of sizes 2, 4, 4, 3, 8
##
## Cluster means:
##      Market_Cap      Beta      PE_Ratio      ROE      ROA Asset_Turnover
## 1 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951  0.2306328
## 2 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428 -1.2684804
## 3  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431  1.1531640
## 4 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478 -0.4612656
## 5 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915  0.1729746
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.14170336 -0.1168459      -1.416514761
## 2  0.06308085  1.5180158      -0.006893899
## 3 -0.46807818  0.4671788       0.591242521
## 4  1.36644699 -0.6912914      -1.320000179
## 5 -0.27449312 -0.7041516       0.556954446
##
## Clustering vector:
## [1] 5 1 5 5 2 4 5 4 2 5 3 4 3 2 3 5 3 1 5 2 5
##
## Within cluster sum of squares by cluster:
## [1] 2.803505 12.791257 9.284424 15.595925 21.879320
## (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 summary function performed on the k-means data gives the information on 9 variables for the 21 pharmaceutical companies. There are a total of 5 clusters with 5 centeroid points and the sizes of the clusters are (2, 3, 4, 4, 8).

##QUESTION-B - Interpret the clusters with respect to the numerical variables used in forming the clusters.

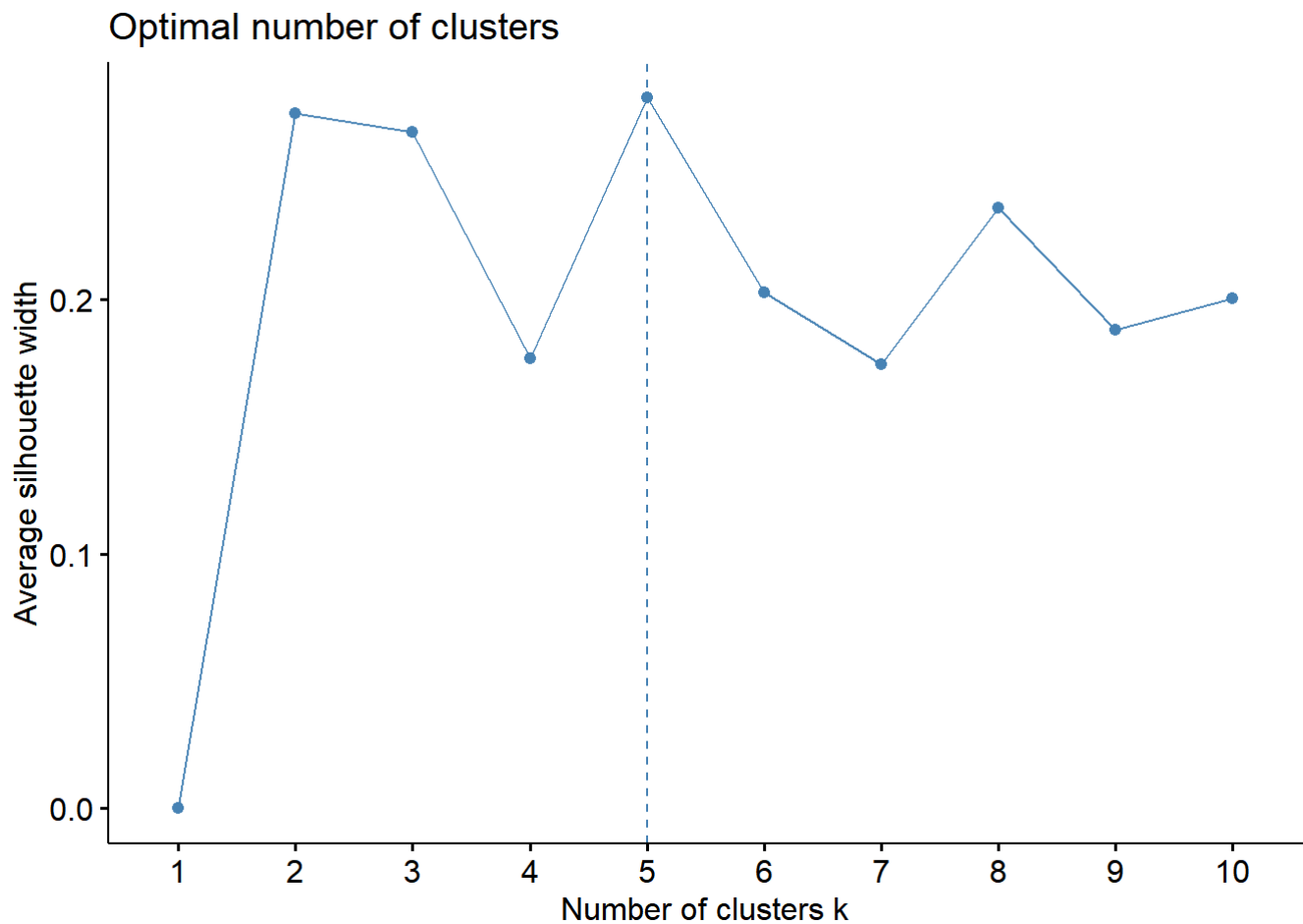
```
#plotting the number of clusters vs the total value
fviz_nbclust(Scaling_pharma_data, kmeans, method = "wss")
```

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



##By judging from the silhouette graph we easily conclude that 5 clusters is the best and optimal solution for this k-means clustering.

##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)

```
#Calculating the mean value from the actual data that is plotted in the clusters
#performing 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	31.910000	0.40500	69.5000	13.20000	5.600000	0.7500
## 2	2	13.100000	0.59750	17.6750	14.57500	6.200000	0.4250
## 3	3	157.017500	0.48000	22.2250	44.42500	17.700000	0.9500
## 4	4	6.636667	0.87000	24.6000	16.46667	4.166667	0.6000
## 5	5	55.810000	0.41375	20.2875	28.73750	12.687500	0.7375

##	Leverage	Rev_Growth	Net_Profit_Margin
## 1	0.475000	12.080000	6.400000
## 2	0.635000	30.142500	15.650000
## 3	0.220000	18.532500	19.575000
## 4	1.653333	5.733333	7.033333
## 5	0.371250	5.591250	19.350000

```
#performing 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      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>
```

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

```

## factor(temp_data$cluster): 1
##      Symbol      Name      Market_Cap      Beta
## Length:2      Length:2      Min.   : 7.58      Min.   :0.4000
## Class :character Class :character 1st Qu.:19.75      1st Qu.:0.4025
## Mode  :character Mode  :character Median :31.91      Median :0.4050
##                                     Mean  :31.91      Mean  :0.4050
##                                     3rd Qu.:44.08      3rd Qu.:0.4075
##                                     Max.   :56.24      Max.   :0.4100
##      PE_Ratio      ROE      ROA      Asset_Turnover      Leverage
## Min.   :56.5      Min.   :12.90      Min.   :5.50      Min.   :0.600      Min.   :0.3500
## 1st Qu.:63.0      1st Qu.:13.05      1st Qu.:5.55      1st Qu.:0.675      1st Qu.:0.4125
## Median :69.5      Median :13.20      Median :5.60      Median :0.750      Median :0.4750
## Mean   :69.5      Mean   :13.20      Mean   :5.60      Mean   :0.750      Mean   :0.4750
## 3rd Qu.:76.0      3rd Qu.:13.35      3rd Qu.:5.65      3rd Qu.:0.825      3rd Qu.:0.5375
## Max.   :82.5      Max.   :13.50      Max.   :5.70      Max.   :0.900      Max.   :0.6000
##      Rev_Growth      Net_Profit_Margin Median_Recommendation      Location
## Min.   : 9.16      Min.   :5.50      Length:2      Length:2
## 1st Qu.:10.62      1st Qu.:5.95      Class :character Class :character
## Median :12.08      Median :6.40      Mode  :character Mode  :character
## Mean   :12.08      Mean   :6.40
## 3rd Qu.:13.54      3rd Qu.:6.85
## Max.   :15.00      Max.   :7.30
##      Exchange      cluster
## Length:2      Min.   :1
## Class :character 1st Qu.:1
## Mode  :character Median :1
##                                     Mean  :1
##                                     3rd Qu.:1
##                                     Max.   :1
## -----
## factor(temp_data$cluster): 2
##      Symbol      Name      Market_Cap      Beta
## Length:4      Length:4      Min.   : 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.   :2
## Class :character Class :character 1st Qu.:2

```

```

## Mode :character Mode :character Median :2
##                                     Mean  :2
##                                     3rd Qu.:2
##                                     Max.   :2
## -----
## factor(temp_data$cluster): 3
##      Symbol          Name          Market_Cap          Beta
## Length:4          Length:4          Min.    :122.1    Min.    :0.3500
## Class :character   Class :character   1st Qu.:129.9    1st Qu.:0.4325
## Mode  :character   Mode  :character   Median :153.2    Median :0.4600
##                                     Mean  :157.0    Mean  :0.4800
##                                     3rd Qu.:180.3    3rd Qu.:0.5075
##                                     Max.   :199.5    Max.   :0.6500
##      PE_Ratio      ROE          ROA          Asset_Turnover
## Min.    :18.00    Min.    :28.60    Min.    :15.00    Min.    :0.800
## 1st Qu.:18.68    1st Qu.:37.60    1st Qu.:15.97    1st Qu.:0.875
## Median :21.25    Median :43.10    Median :17.75    Median :0.950
## Mean    :22.23    Mean    :44.42    Mean    :17.70    Mean    :0.950
## 3rd Qu.:24.80    3rd Qu.:49.92    3rd Qu.:19.48    3rd Qu.:1.025
## Max.    :28.40    Max.    :62.90    Max.    :20.30    Max.    :1.100
##      Leverage      Rev_Growth      Net_Profit_Margin Median_Recommendation
## Min.    :0.100    Min.    : 9.37    Min.    :14.10    Length:4
## 1st Qu.:0.145    1st Qu.:15.36    1st Qu.:16.95    Class:character
## Median :0.220    Median :19.61    Median :19.50    Mode  :character
## Mean    :0.220    Mean    :18.53    Mean    :19.57
## 3rd Qu.:0.295    3rd Qu.:22.79    3rd Qu.:22.12
## Max.    :0.340    Max.    :25.54    Max.    :25.20
##      Location      Exchange      cluster
## Length:4          Length:4          Min.    :3
## Class :character   Class :character   1st Qu.:3
## Mode  :character   Mode  :character   Median :3
##                                     Mean    :3
##                                     3rd Qu.:3
##                                     Max.    :3
## -----
## factor(temp_data$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: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. :5
## Class :character Class :character 1st Qu.:5
## Mode :character Mode :character Median :5
## Mean :5
## 3rd Qu.:5
## Max. :5
```

#median calculation

```
recommend_table <- table(temp_data$cluster, temp_data$Median_Recommendation)
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 1 1 0 0 2
## 2 0 2 2 0 4
## 3 2 2 0 0 4
## 4 2 1 0 0 3
## 5 4 1 2 1 8
## Sum 9 7 4 1 21
```

##One strong buy, seven moderate buys, nine holds, and four moderate sells make the total number of 21 recommendations. All four recommendations, including the opposite advice on buys and sells, are mixed together in Cluster 5. Only mod purchase and hold information can be found in Clusters 1, 2, and 3.Both a moderate buy and moderate sell recommendation are present for Cluster 4.

```
#Location of firm headquarter's breakdown of clusters based on the merged data
location_data <- table(temp_data$cluster, temp_data$Location)
names(dimnames(location_data)) <- c("Cluster", "Location")
location_data <- addmargins(location_data)
location_data
```

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

##There are 21 firms in all, with 13 in the US, 3 in the UK, and 1 each in Canada, France,Germany,Ireland, and Switzerland. US, UK, and Switzerland are all featured in Cluster 5. Germany and the US are in Cluster 2. US and Canada are in Cluster 1. US and Britain are in Cluster 3.The US, France, and Ireland make up Cluster 4.

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

##There are 21 companies overall, divided into 1 Amex, 1 Nasdaq, and 19 NYSE. Cluster 5 just has the NYSE. All three are in Cluster 2. clusters 1,3,4 all contains only NYSE.

##QUESTION-D - Provide an appropriate name for each cluster using any or all of the variables in the dataset.

ANSWER :- Cluster 1: - The cluster 3 can be named as “Small_Net_Profit_Margin-High PE ratio”. All are NAM companies.

Cluster 2: The cluster number 2 can be named “Low_Market_Cap & Less_ROA” - Hold or Buy exchanges.

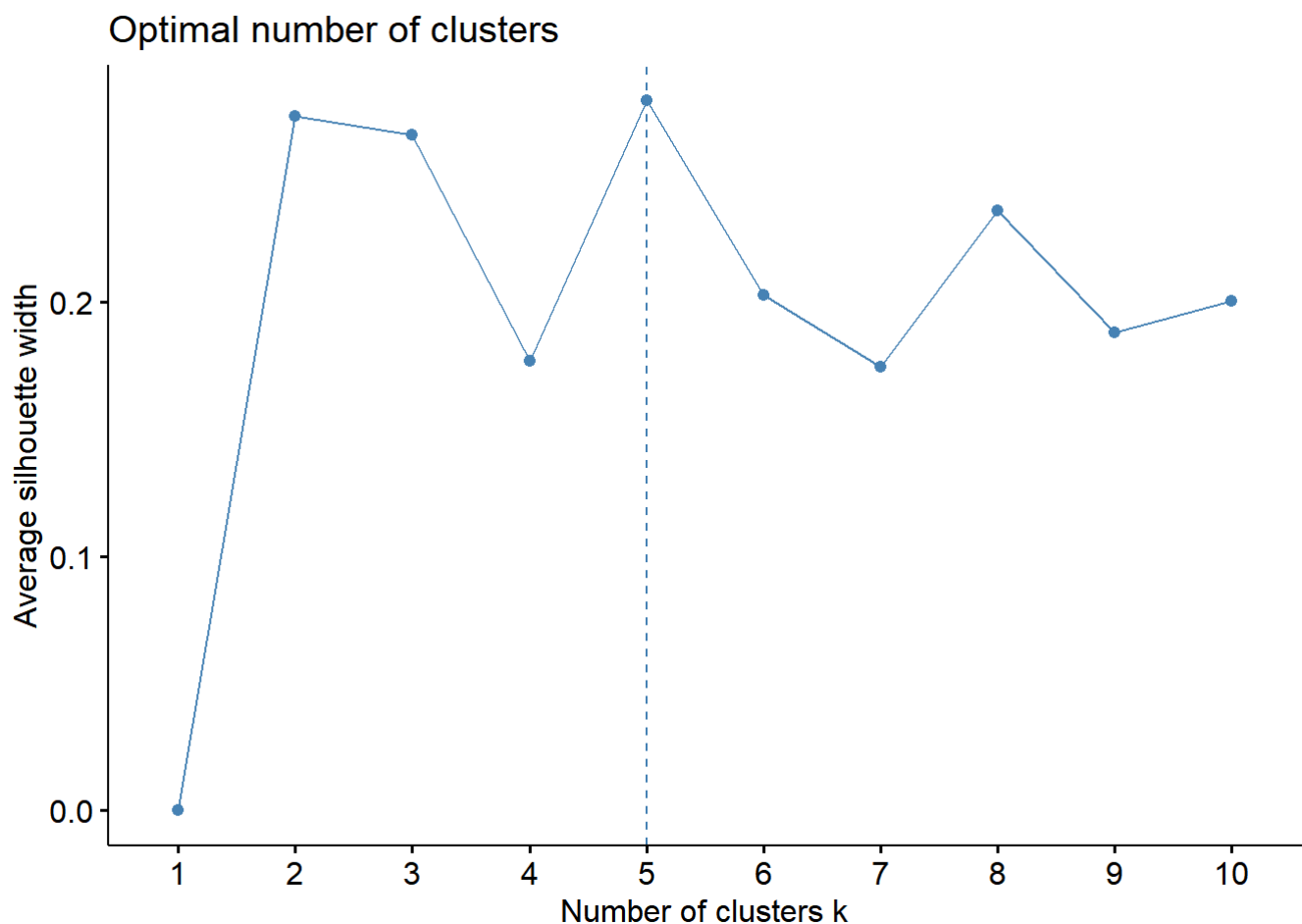
Cluster 3: The cluster 4 can be named “High Market Cap - more RoE - more RoA- High Asset Turnover- more NetProfitMargin” - All are the Hold or Buy US companies that are part of NYSE.

Cluster 4: The cluster 4 can be named as “least PE ratio & low RoE & Minimum Asset Turnover- High revenue growth - mixed recommendation. All are US or European companies that belongs to NYSE.

Cluster 5: The cluster number 5 can be named as “Least_Revenue_growth”. It mostly comprised of US companies and all are NYSE.

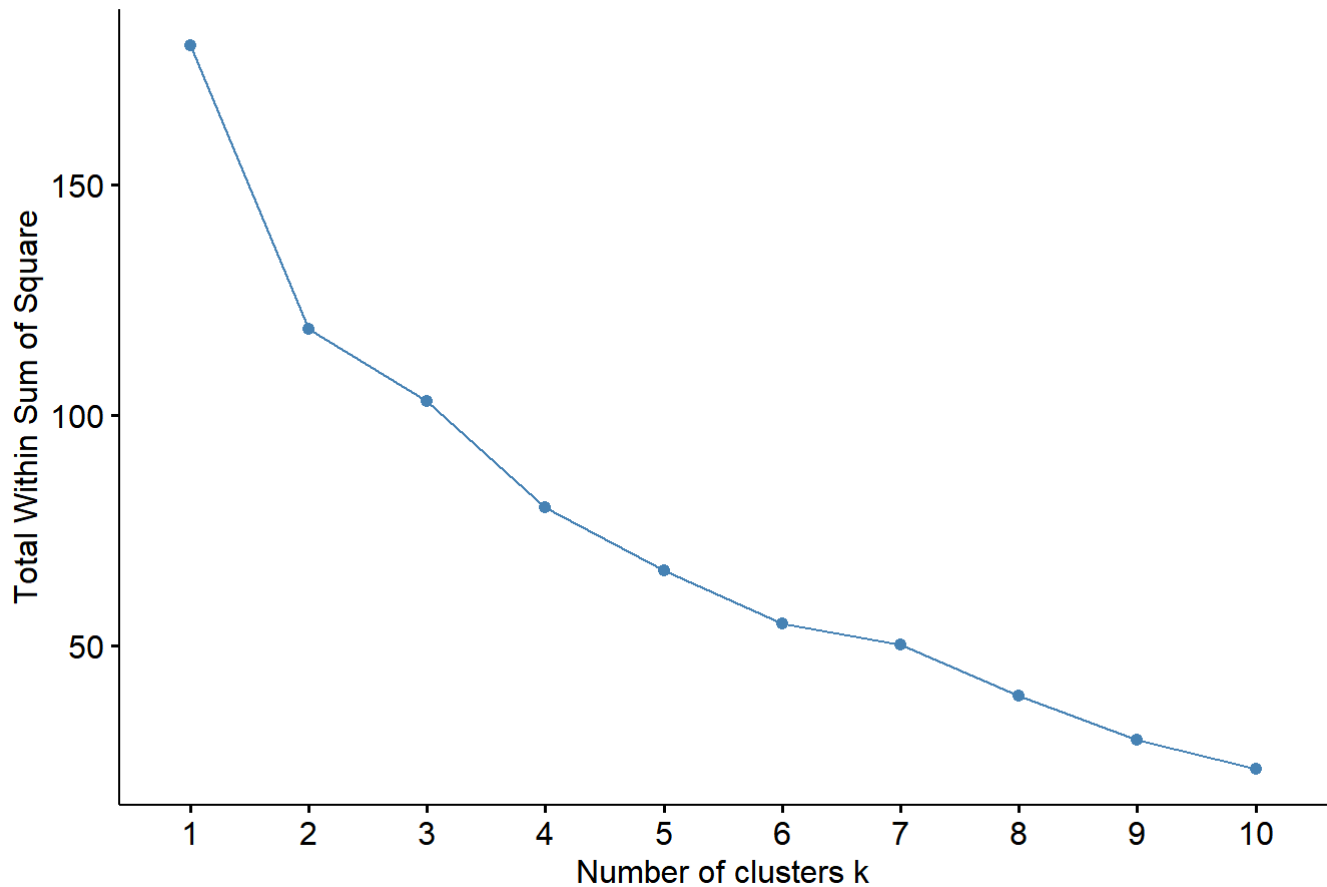
##Also Trying or Exploring the other algorithms whether they can perform better clustering or not?

```
#Assuming range_pharma is your pharmaceutical dataset
#Visualize the number of clusters using the silhouette method with k-means
fviz_nbclust(range_pharma, FUN = kmeans, method = "silhouette")
```



```
#Assuming range_pharma_data is your pharmaceutical dataset
#Visualize the number of clusters using the within-cluster sum of squares method with k-means
fviz_nbclust(range_pharma, kmeans, method = "wss")
```

Optimal number of clusters



PLOTTING THE KMEANS AND THE CLUSTERS

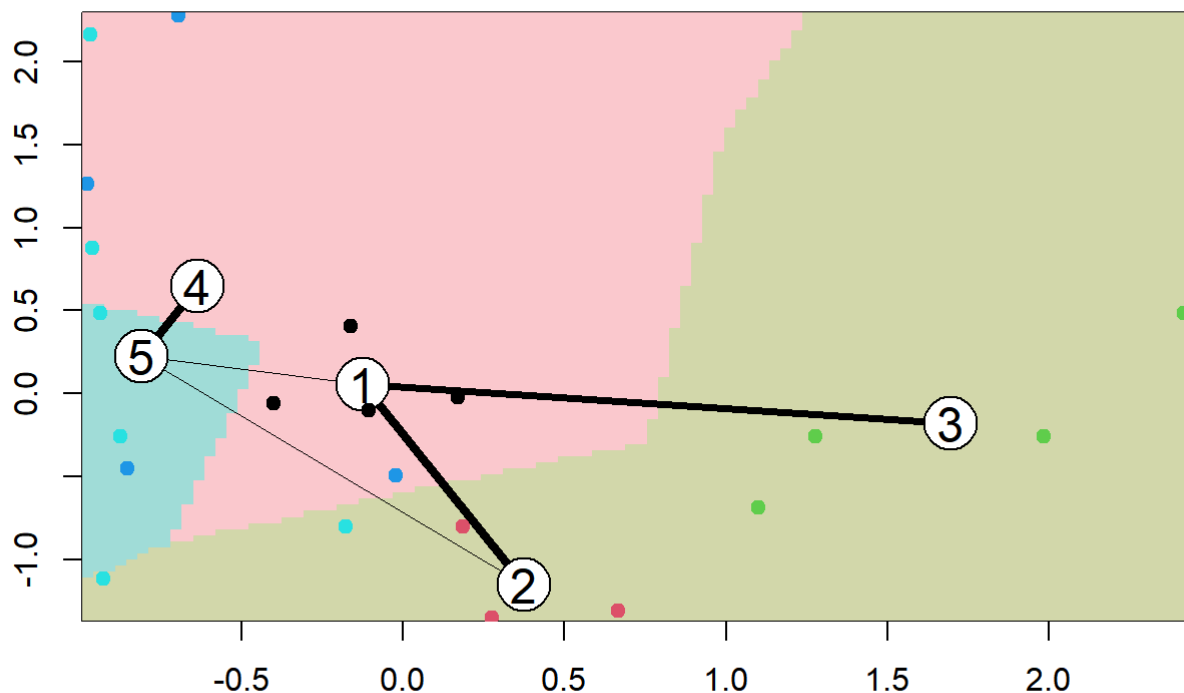
```
# Perform kernel k-means clustering
kmeans_2 = kcca(Scaling_pharma_data, k=5, kccaFamily("kmeans"))
kmeans_2
```

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

```
# Extract cluster assignments from the kernel k-means clustering object
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(Scaling_pharma_data, col=clusters_index, pch=19, cex=1.0)
```



Instead of utilising the kmeans function from base R, we perform a kmeans cluster on $k = 5$ here utilising the kcca method. In comparison to the base R method, the clustering has the same size but a distinct assignment between the points. The clustering graph shows that there is more ambiguity in the grouping than we would want, especially between groups 1, 2, and 3.

PLOTTING THE KMEDIANS AND THE CLUSTERS.

```
# Perform kernel k-means clustering with k-medians algorithm
kmeans_2 = kcca(Scaling_pharma_data, k=5, kccaFamily("kmedians"))
kmeans_2
```

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

```
# Extract cluster assignments from the kernel k-means clustering object
clusters(kmeans_2)
```

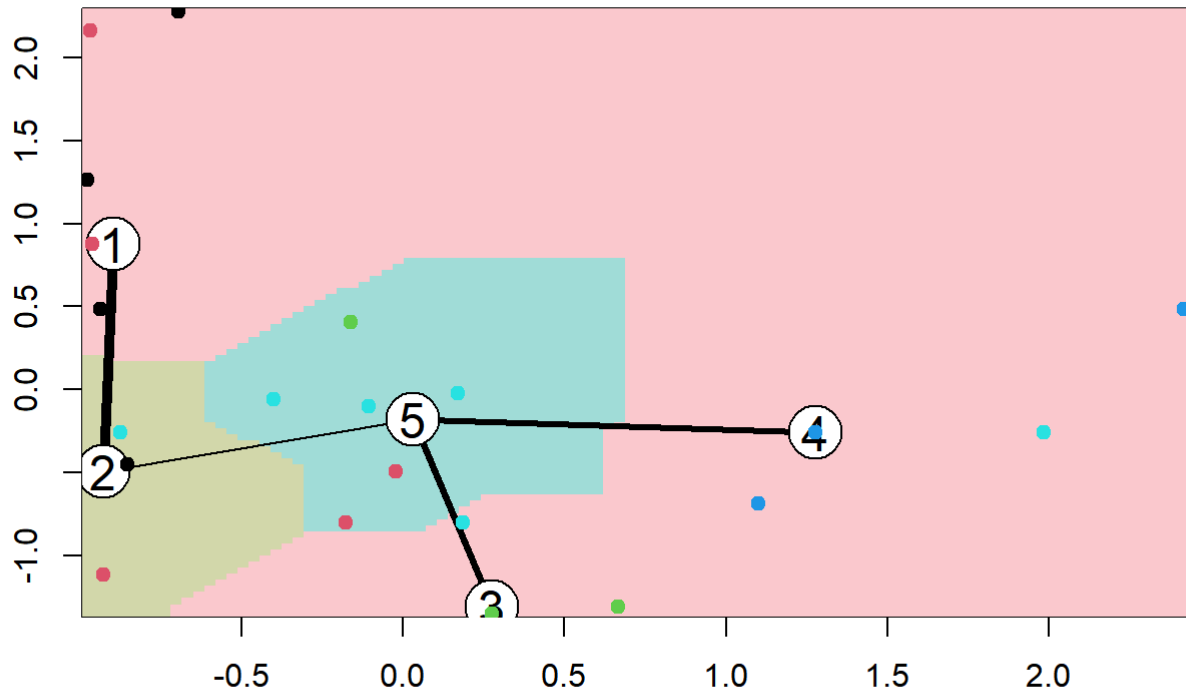
```
## [1] 5 1 5 5 2 1 5 1 2 3 4 1 5 2 4 3 4 2 5 2 3
```



```

# Predict cluster assignments for each data point
clusters_index <- predict(kmeans_2)
# Plot the image representation of the clustering
image(kmeans_2)
# Add points to the plot with cluster assignments
points(Scaling_pharma_data, col=clusters_index, pch=19, cex=1.0)

```



##The five clusters have sizes of 4, 5, 3, 3, and 6 if we convert the KCC to Kmeans. However, the clustering is less noticeable. Although we are looking into the additional data to see if there are any better methods or resources we can employ to improve the visual cluster, it's not obvious if a better cluster is truly possible.