R Notebook

```r
options(repos = c(CRAN = "https://cran.r-project.org"))
library(tidyverse)  # data manipulation
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr    1.5.0
## v ggplot2    3.4.2      v tibble     3.2.1
## v lubridate  1.9.2      v tidyr      1.3.0
## v purrr      1.0.1
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
install.packages("factoextra")
```

```
## Installing package into 'C:/Users/Chris/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'factoextra' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##   C:\Users\Chris\AppData\Local\Temp\RtmpsdMxwR\downloaded_packages
```

```r
library(ISLR)
library (factoextra)
```

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

```r
pharma <- read.csv("C:\\Users\\Chris\\Desktop\\Jocelyn\\Machine Learning\\Assignment IV\\pharmaceuticals
```

```r
summary(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
##
##
##
```

```r
head(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
```

```r
## select columns with numerical variables
pharma_kmeans1 <- pharma[, c(2:11)]
pharma_kmeans <- pharma_kmeans1[, -1]
print(pharma_kmeans)
```

```
##    Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## 1       68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## 2        7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## 3        6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
## 4       67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## 5       47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 6       16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
## 7       51.33 0.50     13.9 34.8 15.1            0.9     0.57       2.70
## 8        0.41 0.85     26.0 24.1  4.3            0.6     3.51       6.38
## 9        0.78 1.08      3.6 15.1  5.1            0.3     1.07      34.21
## 10      73.84 0.18     27.9 31.0 13.5            0.6     0.53       6.21
## 11     122.11 0.35     18.0 62.9 20.3            1.0     0.34      21.87
```

```
## 12        2.60 0.65    19.9 21.4  6.8       0.6     1.45    13.99
## 13      173.93 0.46    28.4 28.6 16.3       0.9     0.10     9.37
## 14        1.20 0.75    28.6 11.2  5.4       0.3     0.93    30.37
## 15      132.56 0.46    18.9 40.6 15.0       1.1     0.28    17.35
## 16       96.65 0.19    21.6 17.9 11.2       0.5     0.06    -2.69
## 17      199.47 0.65    23.6 45.6 19.2       0.8     0.16    25.54
## 18       56.24 0.40    56.5 13.5  5.7       0.6     0.35    15.00
## 19       34.10 0.51    18.9 22.6 13.3       0.8     0.00     8.56
## 20        3.26 0.24    18.4 10.2  6.8       0.5     0.20    29.18
## 21       48.19 0.63    13.1 54.9 13.4       0.6     1.12     0.36
##    Net_Profit_Margin
## 1              16.1
## 2               5.5
## 3              11.2
## 4              18.0
## 5              12.9
## 6               2.6
## 7              20.6
## 8               7.5
## 9              13.3
## 10             23.4
## 11             21.1
## 12             11.0
## 13             17.9
## 14             21.3
## 15             14.1
## 16             22.4
## 17             25.2
## 18              7.3
## 19             17.6
## 20             15.1
## 21             25.5
```

summary(pharma_kmeans)

```
##    Market_Cap          Beta           PE_Ratio          ROE
##  Min.   :  0.41   Min.   :0.1800   Min.   : 3.60   Min.   : 3.9
##  1st Qu.:  6.30   1st Qu.:0.3500   1st Qu.:18.90   1st Qu.:14.9
##  Median : 48.19   Median :0.4600   Median :21.50   Median :22.6
##  Mean   : 57.65   Mean   :0.5257   Mean   :25.46   Mean   :25.8
##  3rd Qu.: 73.84   3rd Qu.:0.6500   3rd Qu.:27.90   3rd Qu.:31.0
##  Max.   :199.47   Max.   :1.1100   Max.   :82.50   Max.   :62.9
##       ROA         Asset_Turnover     Leverage        Rev_Growth
##  Min.   : 1.40   Min.   :0.3     Min.   :0.0000   Min.   :-3.17
##  1st Qu.: 5.70   1st Qu.:0.6     1st Qu.:0.1600   1st Qu.: 6.38
##  Median :11.20   Median :0.6     Median :0.3400   Median : 9.37
##  Mean   :10.51   Mean   :0.7     Mean   :0.5857   Mean   :13.37
##  3rd Qu.:15.00   3rd Qu.:0.9     3rd Qu.:0.6000   3rd Qu.:21.87
##  Max.   :20.30   Max.   :1.1     Max.   :3.5100   Max.   :34.21
##  Net_Profit_Margin
##  Min.   : 2.6
##  1st Qu.:11.2
##  Median :16.1
##  Mean   :15.7
```
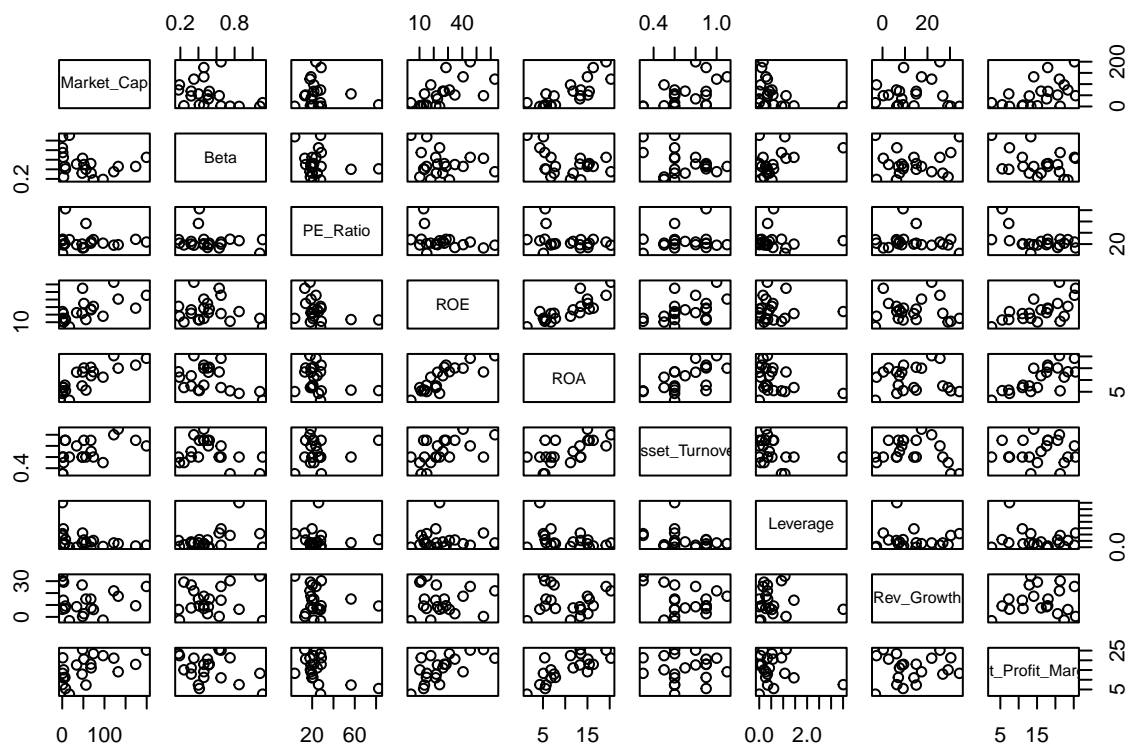
```
##  3rd Qu.:21.1
##  Max.   :25.5
```

```
##Data exploration - look for outliers, correlations between variables
boxplot(pharma_kmeans)
```
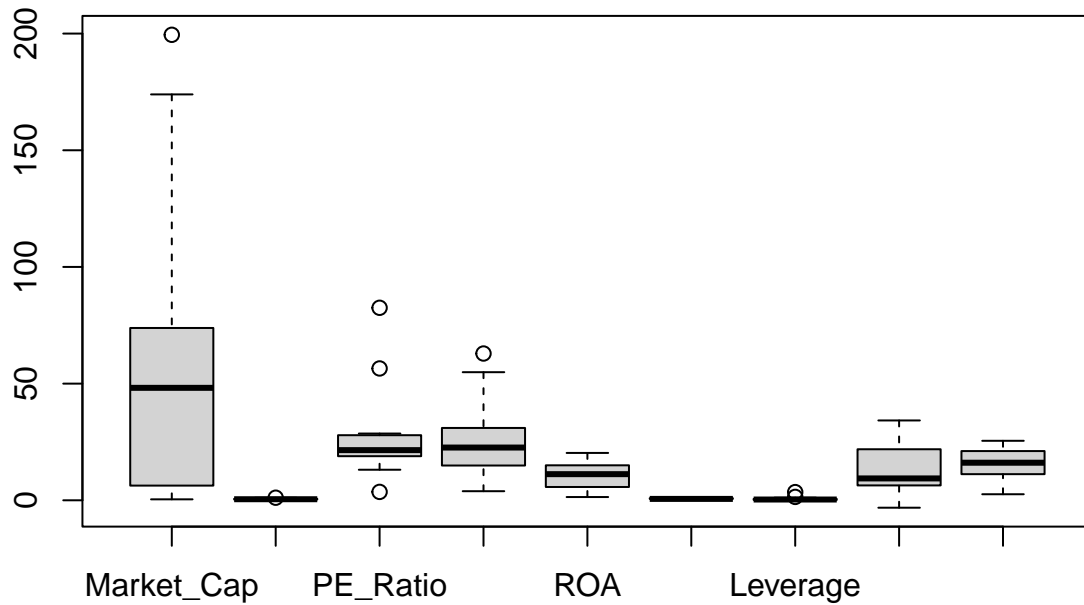


```
## Market_Cap, PE Ratio, and ROE variables have outliers
library(ggplot2)
pairs(pharma_kmeans)
```

Scatter plots show correlations between variables

```
correlation_matrix<-cor(pharma_kmeans)
print(correlation_matrix)
```

```
##                    Market_Cap         Beta    PE_Ratio          ROE         ROA
## Market_Cap        1.000000000 -0.31250762 -0.08798317  0.61952576  0.80908852
## Beta             -0.312507620  1.00000000 -0.19716312 -0.20273345 -0.42583638
## PE_Ratio         -0.087983169 -0.19716312  1.00000000 -0.32205434 -0.29207790
## ROE               0.619525759 -0.20273345 -0.32205434  1.00000000  0.83168600
## ROA               0.809088517 -0.42583638 -0.29207790  0.83168600  1.00000000
## Asset_Turnover    0.507917513 -0.32069694  0.14974635  0.49612507  0.61977107
## Leverage         -0.408937481  0.40116206 -0.03985770  0.01560562 -0.36535802
## Rev_Growth        0.003788982  0.08807135 -0.15499183 -0.01905389 -0.02118403
## Net_Profit_Margin 0.516711077 -0.34546582 -0.46240116  0.63395830  0.74875756
##                   Asset_Turnover    Leverage   Rev_Growth Net_Profit_Margin
## Market_Cap            0.50791751 -0.40893748  0.003788982        0.51671108
## Beta                 -0.32069694  0.40116206  0.088071348       -0.34546582
## PE_Ratio              0.14974635 -0.03985770 -0.154991834       -0.46240116
## ROE                   0.49612507  0.01560562 -0.019053892        0.63395830
## ROA                   0.61977107 -0.36535802 -0.021184032        0.74875756
## Asset_Turnover        1.00000000 -0.30817546 -0.253024565        0.01862763
## Leverage             -0.30817546  1.00000000 -0.021881004       -0.22135214
## Rev_Growth           -0.25302457 -0.02188100  1.000000000        0.08478937
## Net_Profit_Margin     0.01862763 -0.22135214  0.084789374        1.00000000
```
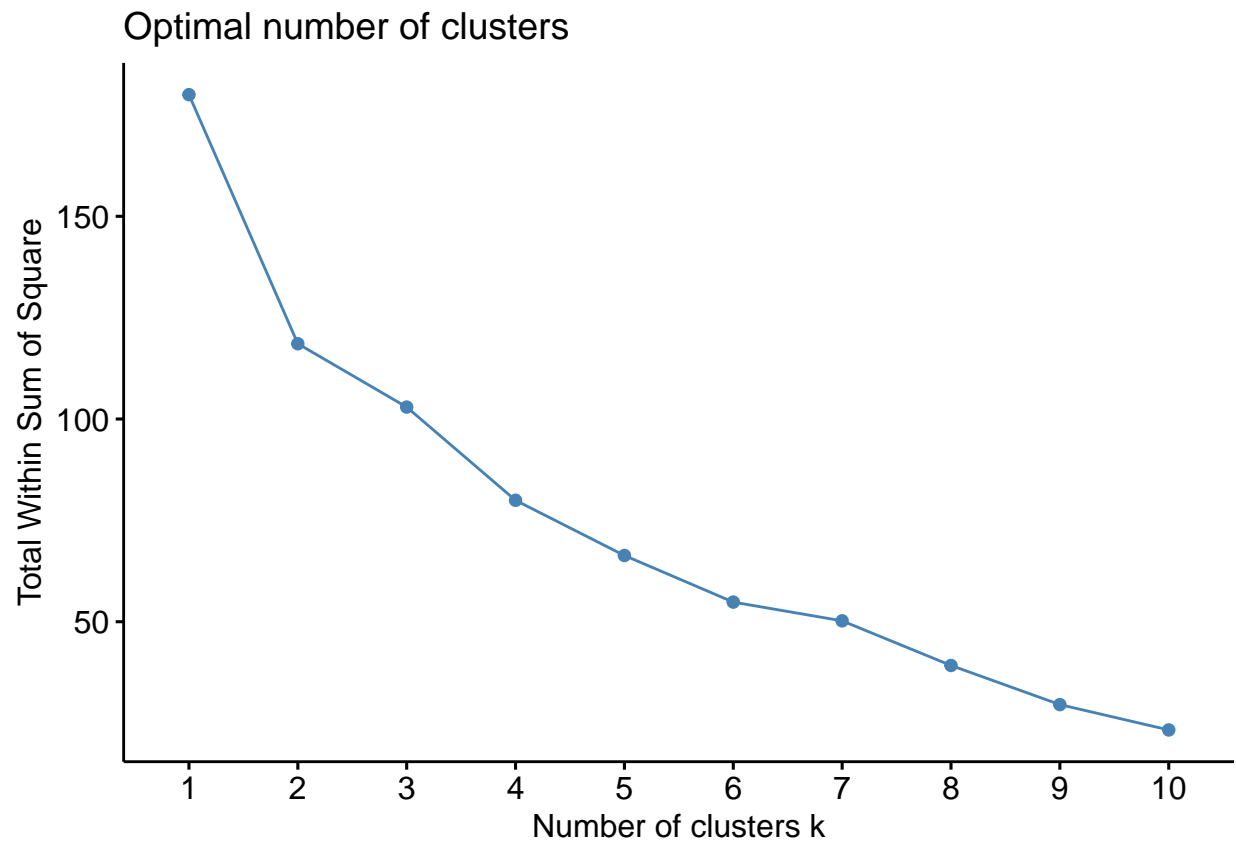
5

```
## Market cap, ROA and ROE, ROA have correlation co-efficient >.8, suggesting these variables are highly
```
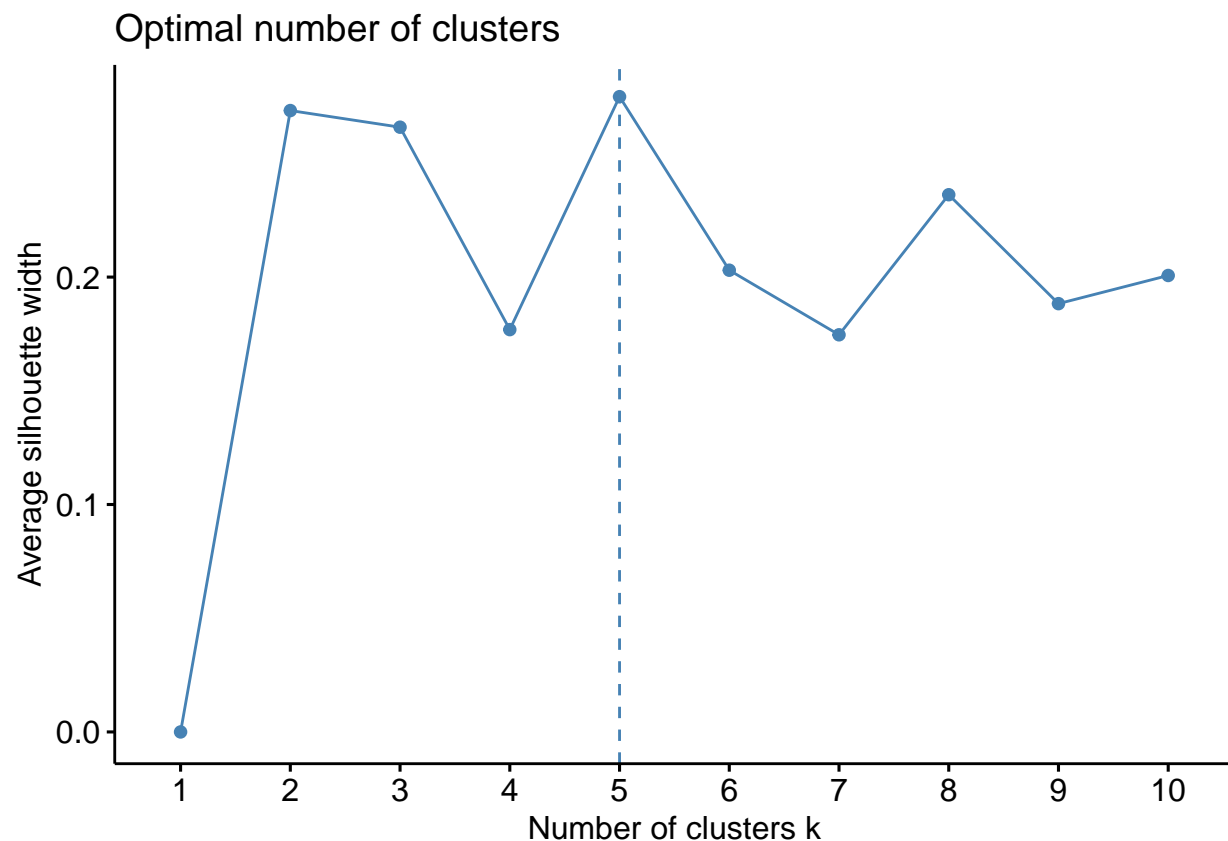
```
## scale data
# Scaling the data frame (z-score)
set.seed(54)
pharma_scaled <- scale(pharma_kmeans)
distance <- get_dist(pharma_scaled)
fviz_dist(distance)
```


'
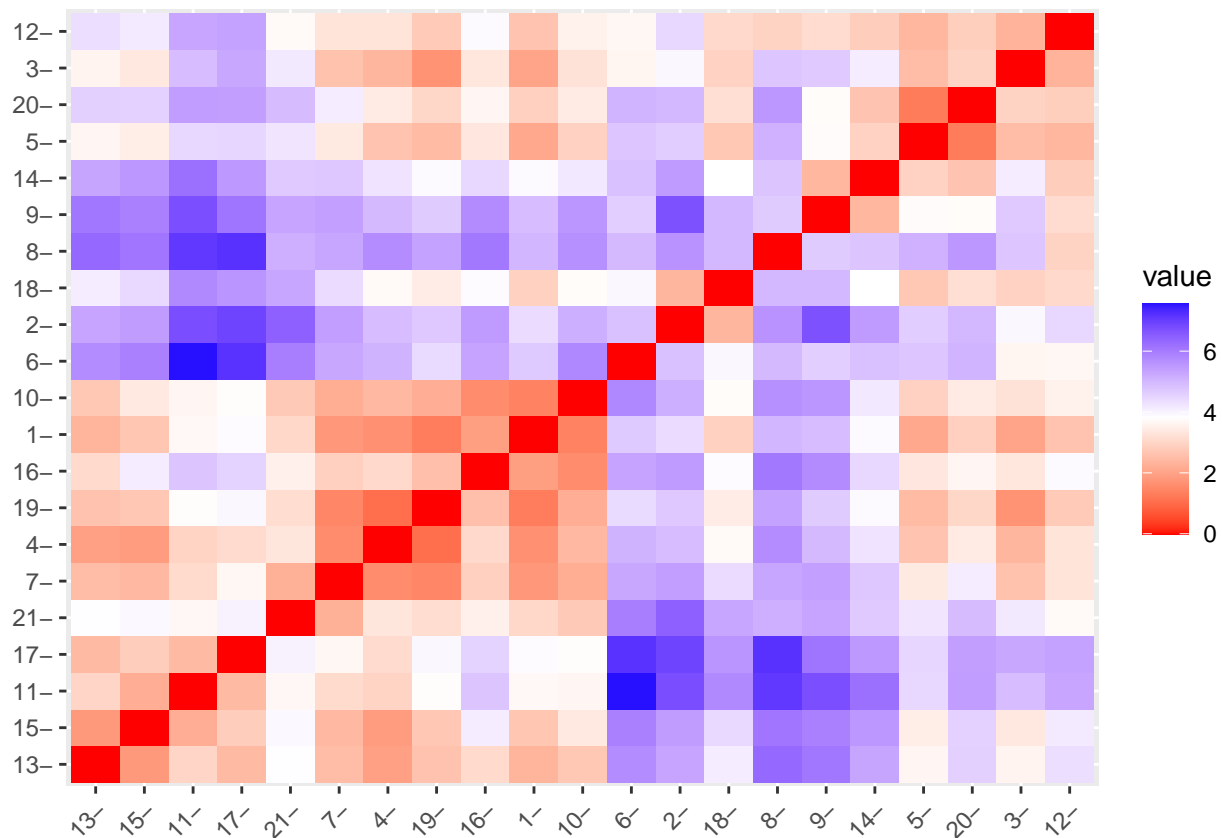
```
## Determine best K
fviz_nbclust(pharma_scaled, kmeans, method = "wss")
```

Optimal number of clusters


```r
fviz_nbclust(pharma_scaled, kmeans, method = "silhouette")
```

Optimal number of clusters

```
## Market cap, ROA and ROE, ROA have correlation co-efficient >.8, meaning their R2 values are approxima
pharma_scaled_df<-data.frame(pharma_scaled)
print(pharma_scaled_df)
```

```
##      Market_Cap        Beta    PE_Ratio         ROE         ROA Asset_Turnover
## 1     0.1840960 -0.80125356 -0.04671323  0.04009035   0.2416121      0.0000000
## 2    -0.8544181 -0.45070513  3.49706911 -0.85483986  -0.9422871      0.9225312
## 3    -0.8762600 -0.25595600 -0.29195768 -0.72225761  -0.5100700      0.9225312
## 4     0.1702742 -0.02225704 -0.24290879  0.10638147   0.9181259      0.9225312
## 5    -0.1790256 -0.80125356 -0.32874435 -0.26484883  -0.5664461     -0.4612656
## 6    -0.6953818  2.27578267  0.14948233 -1.45146000  -1.7127612     -0.4612656
## 7    -0.1078688 -0.10015669 -0.70887325  0.59693581   0.8617498      0.9225312
## 8    -0.9767669  1.26308721  0.03299122 -0.11237924  -1.1677918     -0.4612656
## 9    -0.9704532  2.15893320 -1.34037772 -0.70899938  -1.0174553     -1.8450624
## 10    0.2762415 -1.34655112  0.14948233  0.34502953   0.5610770     -0.4612656
## 11    1.0999201 -0.68440408 -0.45749769  2.45971647   1.8389364      1.3837968
## 12   -0.9393967  0.48409069 -0.34100657 -0.29136529  -0.6979905     -0.4612656
## 13    1.9841758 -0.25595600  0.18013789  0.18593083   1.0872544      0.9225312
## 14   -0.9632863  0.87358895  0.19240011 -0.96753478  -0.9610792     -1.8450624
## 15    1.2782387 -0.25595600 -0.40231769  0.98142435   0.8429577      1.8450624
## 16    0.6654710 -1.30760129 -0.23677768 -0.52338423   0.1288598     -0.9225312
## 17    2.4199899  0.48409069 -0.11415545  1.31287998   1.6322239      0.4612656
## 18   -0.0240846 -0.48965495  1.90298017 -0.81506519  -0.9047030     -0.4612656
```

```
## 19 -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929       0.4612656
## 20 -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905      -0.9225312
## 21 -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849      -0.4612656
##       Leverage   Rev_Growth Net_Profit_Margin
## 1  -0.21209793 -0.52776752        0.06168225
## 2   0.01828430 -0.38113909       -1.55366706
## 3  -0.40408312 -0.57211809       -0.68503583
## 4  -0.74965647  0.14744734        0.35122600
## 5  -0.31449003  1.21638667       -0.42597037
## 6  -0.74965647 -1.49714434       -1.99560225
## 7  -0.02011273 -0.96584257        0.74744375
## 8   3.74279705 -0.63276071       -1.24888417
## 9   0.61983791  1.88617085       -0.36501379
## 10 -0.07130879 -0.64814764        1.17413980
## 11 -0.31449003  0.76926048        0.82363947
## 12  1.10620040  0.05603085       -0.71551412
## 13 -0.62166634 -0.36213170        0.33598685
## 14  0.44065173  1.53860717        0.85411776
## 15 -0.39128411  0.36014907       -0.24310064
## 16 -0.67286239 -1.45369888        1.02174835
## 17 -0.54487226  1.10143723        1.44844440
## 18 -0.30169102  0.14744734       -1.27936246
## 19 -0.74965647 -0.43544591        0.29026942
## 20 -0.49367621  1.43089863       -0.09070919
## 21  0.68383297 -1.17763919        1.49416183
```

```r
weighted_ROA<-pharma_scaled_df$ROA*.33
pharma_scaled_df$ROA<-weighted_ROA
##check ROA is replaced by weighted ROA
print(pharma_scaled_df)
```

```
##       Market_Cap        Beta     PE_Ratio         ROE        ROA Asset_Turnover
## 1    0.1840960 -0.80125356 -0.04671323  0.04009035  0.07973199      0.0000000
## 2   -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.31095476      0.9225312
## 3   -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.16832309      0.9225312
## 4    0.1702742 -0.02225704 -0.24290879  0.10638147  0.30298156      0.9225312
## 5   -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.18692722     -0.4612656
## 6   -0.6953818  2.27578267  0.14948233 -1.45146000 -0.56521121     -0.4612656
## 7   -0.1078688 -0.10015669 -0.70887325  0.59693581  0.28437743      0.9225312
## 8   -0.9767669  1.26308721  0.03299122 -0.11237924 -0.38537128     -0.4612656
## 9   -0.9704532  2.15893320 -1.34037772 -0.70899938 -0.33576026     -1.8450624
## 10   0.2762415 -1.34655112  0.14948233  0.34502953  0.18515540     -0.4612656
## 11   1.0999201 -0.68440408 -0.45749769  2.45971647  0.60684903      1.3837968
## 12  -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.23033686     -0.4612656
## 13   1.9841758 -0.25595600  0.18013789  0.18593083  0.35879395      0.9225312
## 14  -0.9632863  0.87358895  0.19240011 -0.96753478 -0.31715613     -1.8450624
## 15   1.2782387 -0.25595600 -0.40231769  0.98142435  0.27817605      1.8450624
## 16   0.6654710 -1.30760129 -0.23677768 -0.52338423  0.04252373     -0.9225312
## 17   2.4199899  0.48409069 -0.11415545  1.31287998  0.53863388      0.4612656
## 18  -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.29855200     -0.4612656
## 19  -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.17275264      0.4612656
## 20  -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.23033686     -0.9225312
## 21  -0.1614497  0.40619104 -0.75792214  1.92938746  0.17895402     -0.4612656
##       Leverage   Rev_Growth Net_Profit_Margin
```

```
## 1   -0.21209793 -0.52776752         0.06168225
## 2    0.01828430 -0.38113909        -1.55366706
## 3   -0.40408312 -0.57211809        -0.68503583
## 4   -0.74965647  0.14744734         0.35122600
## 5   -0.31449003  1.21638667        -0.42597037
## 6   -0.74965647 -1.49714434        -1.99560225
## 7   -0.02011273 -0.96584257         0.74744375
## 8    3.74279705 -0.63276071        -1.24888417
## 9    0.61983791  1.88617085        -0.36501379
## 10  -0.07130879 -0.64814764         1.17413980
## 11  -0.31449003  0.76926048         0.82363947
## 12   1.10620040  0.05603085        -0.71551412
## 13  -0.62166634 -0.36213170         0.33598685
## 14   0.44065173  1.53860717         0.85411776
## 15  -0.39128411  0.36014907        -0.24310064
## 16  -0.67286239 -1.45369888         1.02174835
## 17  -0.54487226  1.10143723         1.44844440
## 18  -0.30169102  0.14744734        -1.27936246
## 19  -0.74965647 -0.43544591         0.29026942
## 20  -0.49367621  1.43089863        -0.09070919
## 21   0.68383297 -1.17763919         1.49416183
```

##kmeans using euclidian distance

```
k5 <- kmeans(pharma_scaled_df, centers = 5, nstart = ) # k = 5, number of restarts = 25

# Visualize the output

k5$centers # output the centers
```

```
##    Market_Cap        Beta   PE_Ratio         ROE         ROA Asset_Turnover
## 1 -0.1346592   0.1530172 -0.7333977   1.2631616  0.23166572    2.306328e-01
## 2 -0.1361523  -0.7136164 -0.2291138  -0.2830789  0.02469477   -5.765820e-02
## 3  1.6955811  -0.1780563 -0.1984582   1.2349879  0.44561323    1.153164e+00
## 4 -0.5246281   0.4451409  1.8498439  -1.0404550 -0.39157266    1.480297e-16
## 5 -0.9624758   1.1949250 -0.3639982  -0.5200697 -0.31715613   -1.153164e+00
##      Leverage Rev_Growth Net_Profit_Margin
## 1   0.3318601 -1.0717409         1.1208028
## 2  -0.4584789 -0.1053057         0.2121688
## 3  -0.4680782  0.4671788         0.5912425
## 4  -0.3443544 -0.5769454        -1.6095439
## 5   1.4773718  0.7120120        -0.3688236
```

```
k5$size # Number of companies in each cluster
```

```
## [1] 2 8 4 3 4
```

```
fviz_cluster(k5, data = pharma_scaled_df) # Visualize the output
```

Cluster plot
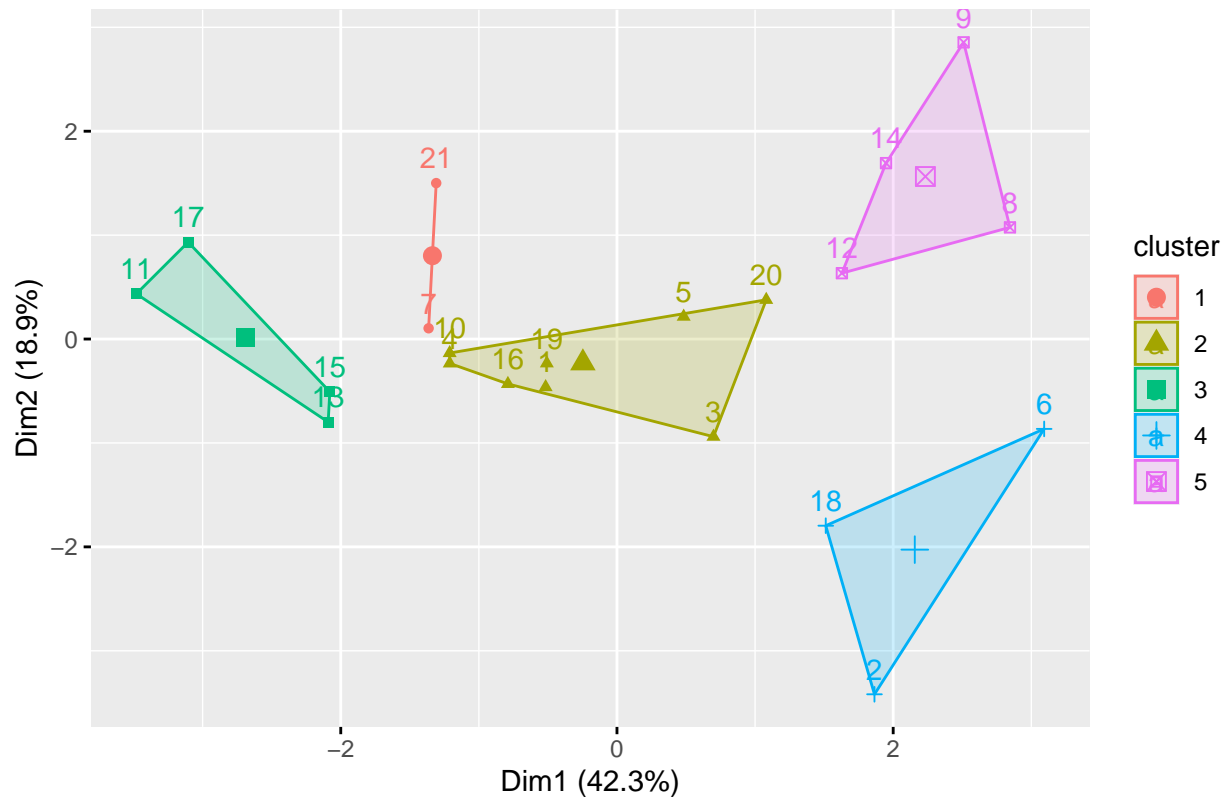

```r
k5$cluster[1:21] # Identify the cluster of each company
```

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

```r
install.packages("flexclust")
```

```
## Installing package into 'C:/Users/Chris/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)
```

```
## package 'flexclust' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'flexclust'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Chris\AppData\Local\R\win-library\4.3\00LOCK\flexclust\libs\x64\flexclust.dll
## to
## C:\Users\Chris\AppData\Local\R\win-library\4.3\flexclust\libs\x64\flexclust.dll:
## Permission denied
```

```
## Warning: restored 'flexclust'
```

```
##
## The downloaded binary packages are in
##  C:\Users\Chris\AppData\Local\Temp\RtmpsdMxwR\downloaded_packages
```

```r
library(flexclust)
```

```
## Loading required package: grid

## Loading required package: lattice

## Loading required package: modeltools

## Loading required package: stats4
```

```r
set.seed(54)
#kmeans clustering, using manhattan distance
k5_manhat = kcca(pharma_scaled_df, k=5, kccaFamily("kmedians"))
k5_manhat
```

```
## kcca object of family 'kmedians'
##
## call:
## kcca(x = pharma_scaled_df, k = 5, family = kccaFamily("kmedians"))
##
## cluster sizes:
##
## 1 2 3 4 5
## 3 2 8 5 3
```

```r
k5manhat_cluster_assignments <- k5_manhat@cluster
```

```r
k5manhat_cluster_assignments <- k5_manhat@cluster
print(k5manhat_cluster_assignments)
```

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

```r
install.packages("cluster")
```

```
## Installing package into 'C:/Users/Chris/AppData/Local/R/win-library/4.3'
## (as 'lib' is unspecified)

## package 'cluster' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'cluster'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying
## C:\Users\Chris\AppData\Local\R\win-library\4.3\00LOCK\cluster\libs\x64\cluster.dll
## to C:\Users\Chris\AppData\Local\R\win-library\4.3\cluster\libs\x64\cluster.dll:
## Permission denied

## Warning: restored 'cluster'

##
## The downloaded binary packages are in
##   C:\Users\Chris\AppData\Local\Temp\RtmpsdMxwR\downloaded_packages
```
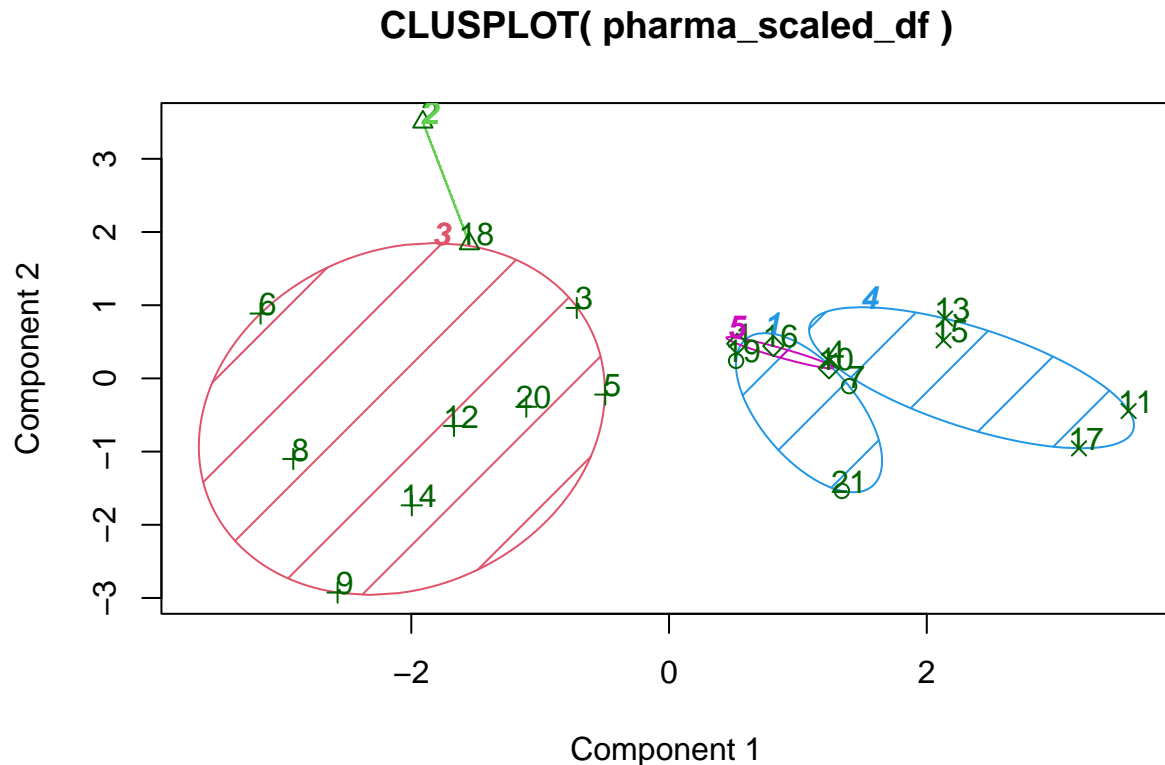
```
library(cluster)
cluster_plot <- clusplot(pharma_scaled_df, k5_manhat@cluster, color = TRUE, shade = TRUE, labels = 2, l
```

CLUSPLOT(pharma_scaled_df)

Component 1

These two components explain 61.23 % of the point variability.

```
centroids<-k5_manhat@centers
print(centroids)
```

```
##        Market_Cap         Beta    PE_Ratio         ROE         ROA Asset_Turnover
## [1,] -0.1614497 -0.06120687 -0.70887325  0.59693581  0.17895402      0.4612656
## [2,] -0.4392513 -0.47018004  2.70002464 -0.83495252 -0.30475338      0.2306328
## [3,] -0.9337656  0.67883982 -0.31035102 -0.71562850 -0.27374650     -0.4612656
## [4,]  1.2782387 -0.25595600 -0.24290879  0.98142435  0.35879395      0.9225312
## [5,]  0.2762415 -1.30760129 -0.04671323  0.04009035  0.07973199     -0.4612656
##        Leverage Rev_Growth Net_Profit_Margin
## [1,] -0.02011273 -0.9658426          0.7474438
## [2,] -0.14170336 -0.1168459         -1.4165148
## [3,]  0.06308085  0.6362088         -0.5555031
## [4,] -0.54487226  0.3601491          0.3512260
## [5,] -0.21209793 -0.6481476          1.0217484
```

##Opt for Manhattan distance measure as it is better suited for datasets with outliers

##b. - see excel sheet for detail

##Cluster 1. characterized by lowest P:E and low revenue growth, but relatively high net profit margin.

```
##Cluster 2. Characterized by extremely high P:E but low net profit margin and otherwise average statist
##Cluster 3. Characterized by the lowest market cap (lowest worth by stock market), but highest leverage
##Cluster 4. Characterized by being the largest companies by market cap, with the lowest leverage but h
## Cluster 5. Cluster with the second highest average market cap, but otherwise fairly average and stab

#C. No strong patterns in the clusters related to non-numerical variables. Cluster 3, the cluster with

#D.
#Cluster 1: Stable but low revenue growth
#Cluster 2: Overpriced or taking a gamble
#Cluster 3: Bright future
#Cluster 4: Big and getting bigger
#Cluster 5: Big and stable
```