# Untitled

2023-10-30

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
Pharmaceuticals <- read.csv("C:\\Users\\13308\\OneDrive\\Documents\\64060_-spavliga\\Pharmaceuticals.csv

#A

data.df <- Pharmaceuticals


data.df <- data.df[,-14]
data.df <- data.df[,-13]
data.df <- data.df[,-12]
data.df <- data.df[,-1]
data.df <- data.df[,-1]
str(data.df)
```

```
## 'data.frame':    21 obs. of  9 variables:
##  $ Market_Cap       : num  68.44 7.58 6.3 67.63 47.16 ...
##  $ Beta             : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...
##  $ PE_Ratio         : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...
##  $ ROE              : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...
##  $ ROA              : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...
##  $ Asset_Turnover   : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
##  $ Leverage         : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...
##  $ Rev_Growth       : num  7.54 9.16 7.05 15 26.81 ...
##  $ Net_Profit_Margin: num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...
```

```
datanorm <- sapply(data.df, scale)

str(datanorm)
```

```
##  num [1:21, 1:9] 0.184 -0.854 -0.876 0.17 -0.179 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr [1:9] "Market_Cap" "Beta" "PE_Ratio" "ROE" ...
```

```r
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.2.3
```

```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

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

```r
library(flexclust)
```

```
## Warning: package 'flexclust' was built under R version 4.2.3
```

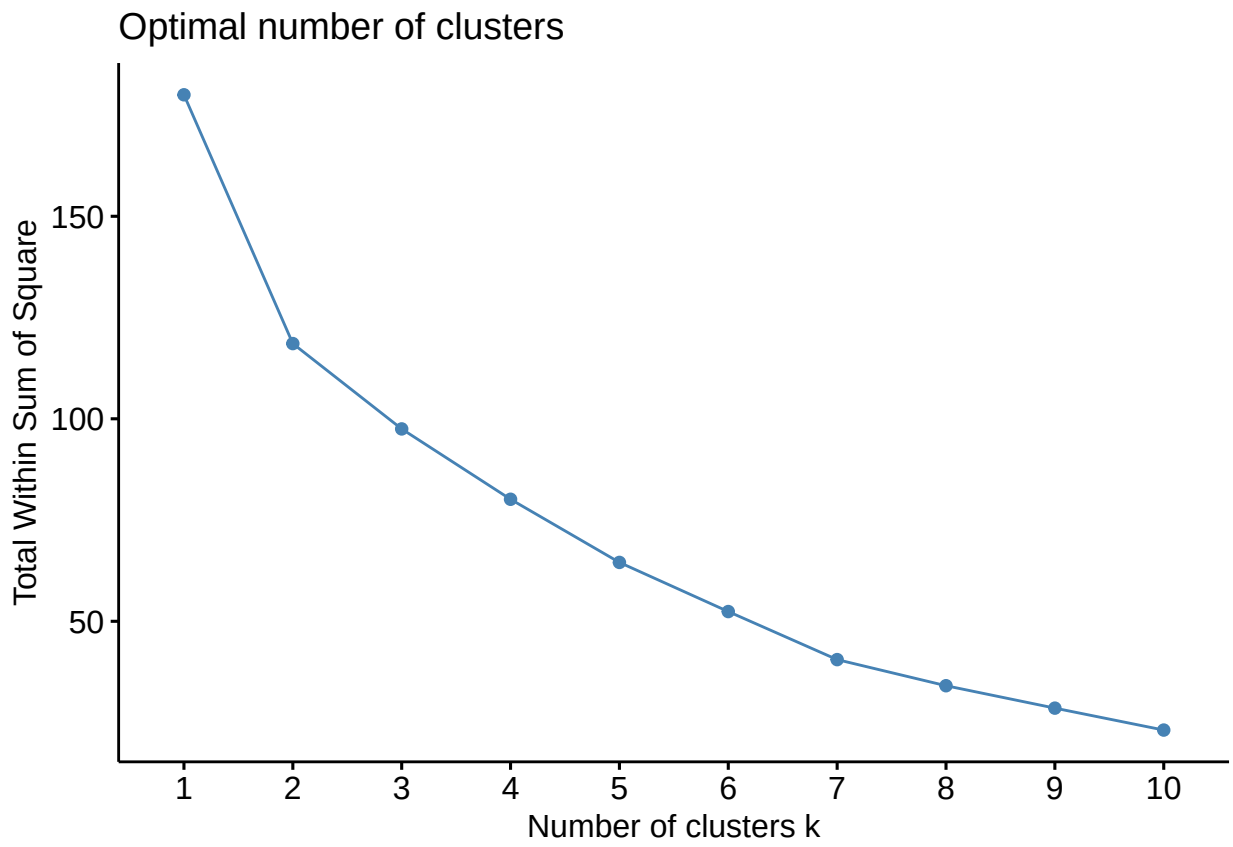
```
## Loading required package: grid
```

```
## Loading required package: lattice
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```r
fviz_nbclust(datanorm, FUNcluster = hcut, method = "wss")
```

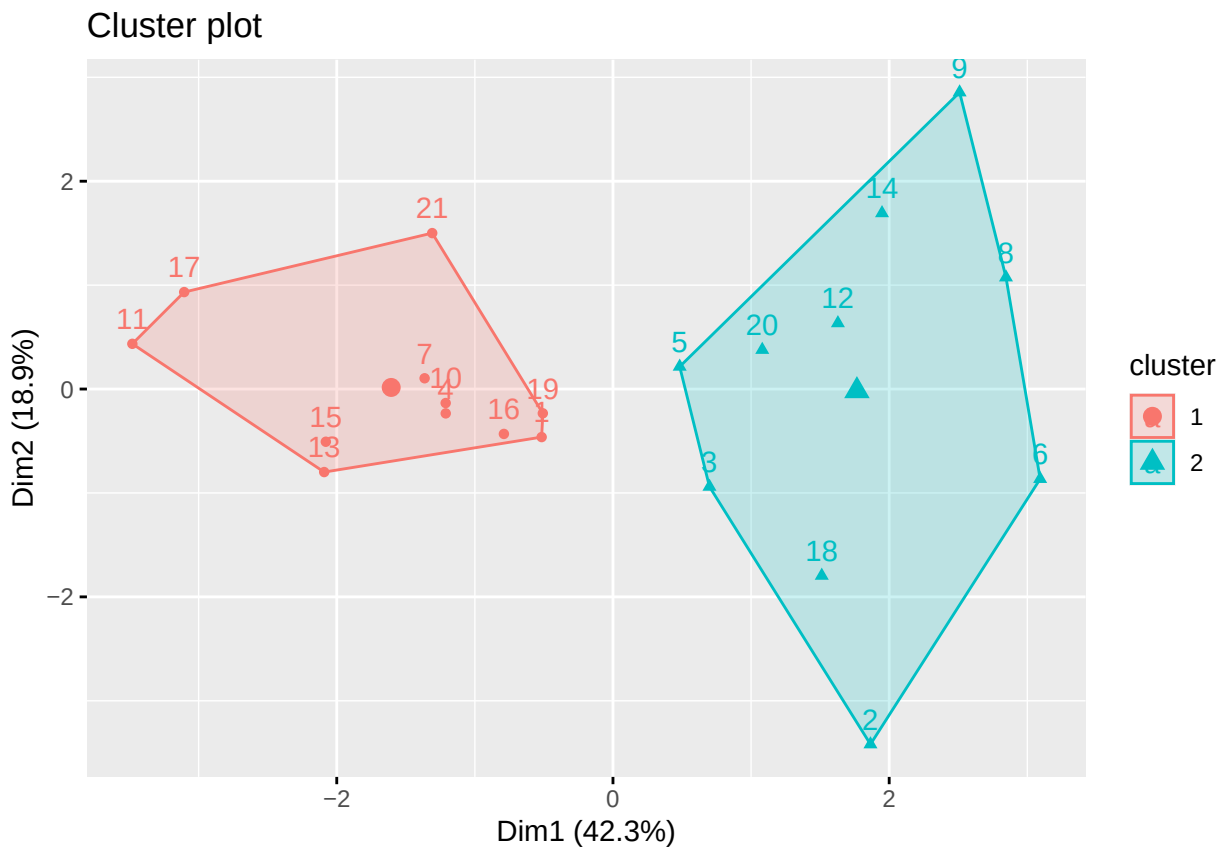


```r
k <- 2

datanormkmeans <- kmeans(datanorm, centers = k)
datanormkmeans
```

```
## K-means clustering with 2 clusters of sizes 11, 10
##
## Cluster means:
##   Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
## 2 -0.7407208  0.3945061  0.3039863 -0.7222576 -0.9178575     -0.5073922
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3331068 -0.2902163         0.6823310
## 2  0.3664175  0.3192379        -0.7505641
##
## Clustering vector:
##  [1] 1 2 2 1 2 2 1 2 2 1 1 2 1 2 1 1 1 2 1 2 1
##
## Within cluster sum of squares by cluster:
## [1] 43.30886 75.26049
##  (between_SS / total_SS =  34.1 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```r
fviz_cluster(datanormkmeans, data = datanorm)
```




```
# Based on the results of the fviz_nbclust function (elbow method), 2 clusters
# appeared to be the best option. Beyond
# preprocessing the data, I took a more straightfoward approach. I just used
# euclidean distance (default for kmeans()) as given
# the nature of our data, there isn't really a need for anything beyond it.

#B
# As a whole, cluster 1 appears to be characterized by being more profitable and
# generally healthier than cluster 2. Specifically noting the values of
# Market_Cap, ROE, ROA and Net_Profit_Margin, cluster 1 appears to be better
# performing than cluster 2. The only exception is Rev_Growth, but
# Net_Profit_Margin also favors cluster 1.

#C
# There does not appear to be any patterns, however it is worth noting that the
# 4 "Sells" in median recommendations are split equally between the two clusters.
# Based on the k-means results, I would have anticipated the "Sells" favoring
# cluster 2. There were no notable findings in location or exchange.

#D
# As a whole, the companies in cluster 1 are characterized by better
# performances in all areas aside from estimated revenue growth.
# Overall, I would name cluster 1 "High performing, high profit companies."
# Cluster 2 is characterized by lower profit margins,
# lower market cap, lower asset turnover, higher beta and so on. While
# the estimated revenue growth does favor cluster 2, only 1 out of 9 variables
```

```
# measured indicates cluster 2 performing better than cluster 1. In contrast to
# cluster 1's name, I would name cluster 2 "Lower performing, lower
# profit companies."
```