Assignment 4

MUNERAH

11/1/2021

```
## Load the file

library(readxl)
df <- read.csv("C:/Users/mnooo/Desktop/Datasets/Pharmaceuticals.csv")
```

```
#install.packages

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(ISLR)
library(utils)
library(caret)
```

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

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

```
library(factoextra)
```

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

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.4     v purrr   0.3.4
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::lift()   masks caret::lift()
```

```
library(cowplot)
library (psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
##Explore the DataSet

str(df)
```

```
## 'data.frame':    21 obs. of  14 variables:
##  $ Symbol               : chr  "ABT" "AGN" "AHM" "AZN" ...
##  $ Name                 : chr  "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZe
neca PLC" ...
##  $ 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 ...
##  $ Median_Recommendation: chr  "Moderate Buy" "Moderate Buy" "Strong Buy" "Moderate Sell" ...
##  $ Location             : chr  "US" "CANADA" "UK" "UK" ...
##  $ Exchange             : chr  "NYSE" "NYSE" "NYSE" "NYSE" ...
```

```
##Show first 6 rows
head(df)
```

```
##   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
```

```
##Descriptive Statistics
summary(df)
```

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

```
## number of rows and colmns
dim(df)
```

```
## [1] 21 14
```

```
###21 row and 14 colmns
```

```
##Running K-means Model
set.seed(123)
df1<-df[,3:11] ### Select only numerical variabls
dfnorm<-scale(df1)#Normalizing
k4<-kmeans(dfnorm, centers = 4, nstart = 25)# Running with clusters K=4
df1$Cluster <-k4$cluster  ### adding cluster to the data set
df1$MedianRec <-df[,c(12)] ### adding varials back to the dataset
df1$Location <-df[, c(13)]
df1$Exchange <-df[, c(14)]
Clus1<-df1[order(df1$Cluster),]
Clus1
```

```
##    Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## 11     122.11 0.35     18.0 62.9 20.3            1.0     0.34      21.87
## 13     173.93 0.46     28.4 28.6 16.3            0.9     0.10       9.37
## 15     132.56 0.46     18.9 40.6 15.0            1.1     0.28      17.35
## 17     199.47 0.65     23.6 45.6 19.2            0.8     0.16      25.54
## 1       68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## 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
## 7       51.33 0.50     13.9 34.8 15.1            0.9     0.57       2.70
## 10      73.84 0.18     27.9 31.0 13.5            0.6     0.53       6.21
## 16      96.65 0.19     21.6 17.9 11.2            0.5     0.06      -2.69
## 19      34.10 0.51     18.9 22.6 13.3            0.8     0.00       8.56
## 21      48.19 0.63     13.1 54.9 13.4            0.6     1.12       0.36
## 5       47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 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
## 12       2.60 0.65     19.9 21.4  6.8            0.6     1.45      13.99
## 14       1.20 0.75     28.6 11.2  5.4            0.3     0.93      30.37
## 20       3.26 0.24     18.4 10.2  6.8            0.5     0.20      29.18
## 2        7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## 6       16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
## 18      56.24 0.40     56.5 13.5  5.7            0.6     0.35      15.00
##    Net_Profit_Margin Cluster     MedianRec    Location Exchange
## 11              21.1       1          Hold          UK     NYSE
## 13              17.9       1  Moderate Buy          US     NYSE
## 15              14.1       1          Hold          US     NYSE
## 17              25.2       1  Moderate Buy          US     NYSE
## 1               16.1       2  Moderate Buy          US     NYSE
## 3               11.2       2    Strong Buy          UK     NYSE
## 4               18.0       2 Moderate Sell          UK     NYSE
## 7               20.6       2 Moderate Sell          US     NYSE
## 10              23.4       2          Hold          US     NYSE
## 16              22.4       2          Hold SWITZERLAND     NYSE
## 19              17.6       2          Hold          US     NYSE
## 21              25.5       2          Hold          US     NYSE
## 5               12.9       3  Moderate Buy      FRANCE     NYSE
## 8                7.5       3  Moderate Buy          US   NASDAQ
## 9               13.3       3 Moderate Sell     IRELAND     NYSE
## 12              11.0       3          Hold          US     AMEX
## 14              21.3       3  Moderate Buy          US     NYSE
## 20              15.1       3 Moderate Sell          US     NYSE
## 2                5.5       4  Moderate Buy      CANADA     NYSE
## 6                2.6       4          Hold     GERMANY     NYSE
## 18               7.3       4          Hold          US     NYSE
```

```
k4$centers ### centroid values
```

```
##    Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1  1.69558112 -0.1780563 -0.1984582  1.2349879  1.3503431   1.153164e+00
## 2 -0.03142211 -0.4360989 -0.3172485  0.1950459  0.4083915   1.729746e-01
## 3 -0.82617719  0.4775991 -0.3696184 -0.5631589 -0.8514589  -9.994088e-01
## 4 -0.52462814  0.4451409  1.8498439 -1.0404550 -1.1865838   1.480297e-16
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.4680782  0.4671788         0.5912425
## 2 -0.2744931 -0.7041516         0.5569544
## 3  0.8502201  0.9158889        -0.3319956
## 4 -0.3443544 -0.5769454        -1.6095439
```

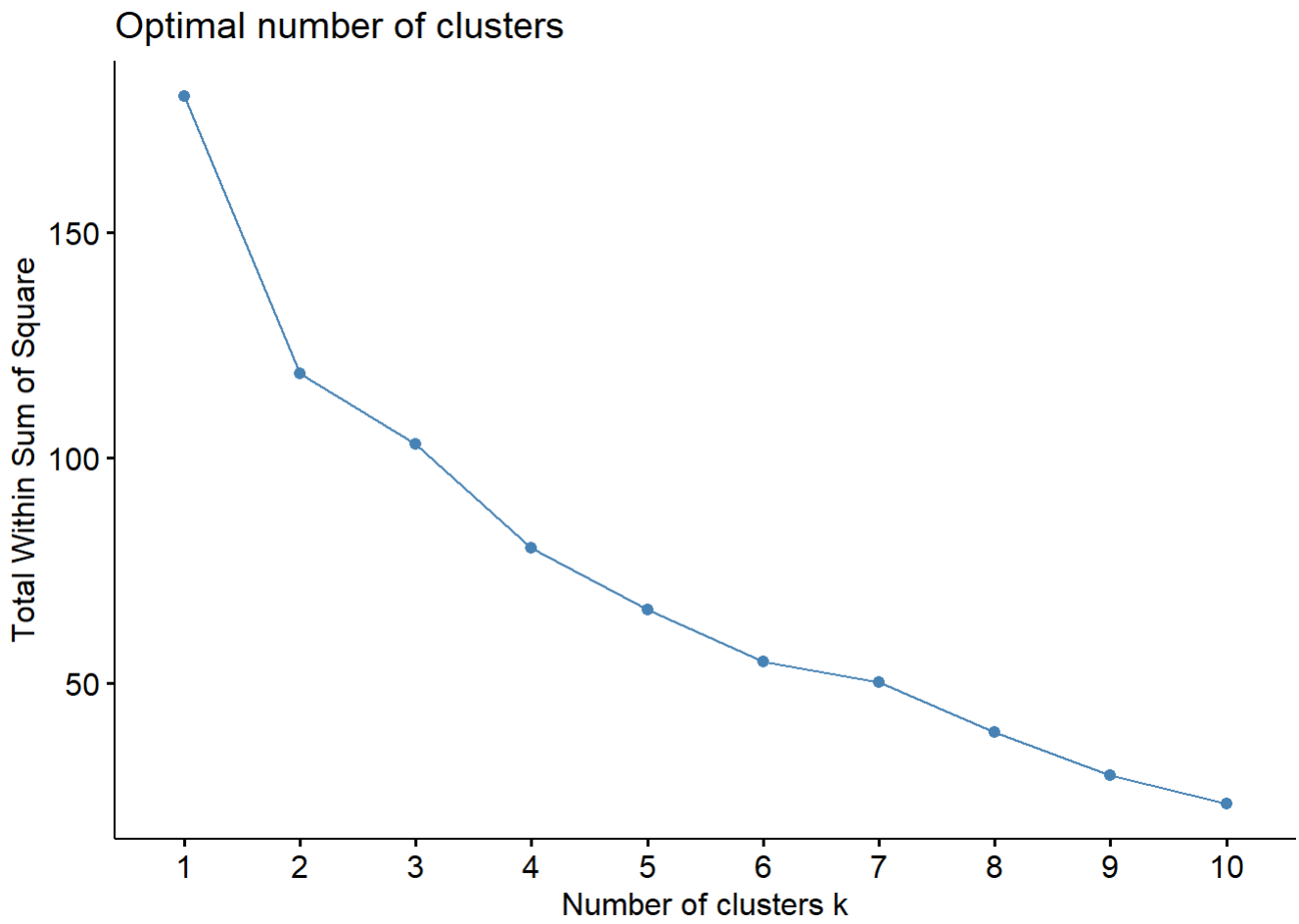
```
k4$size
```

```
## [1] 4 8 6 3
```

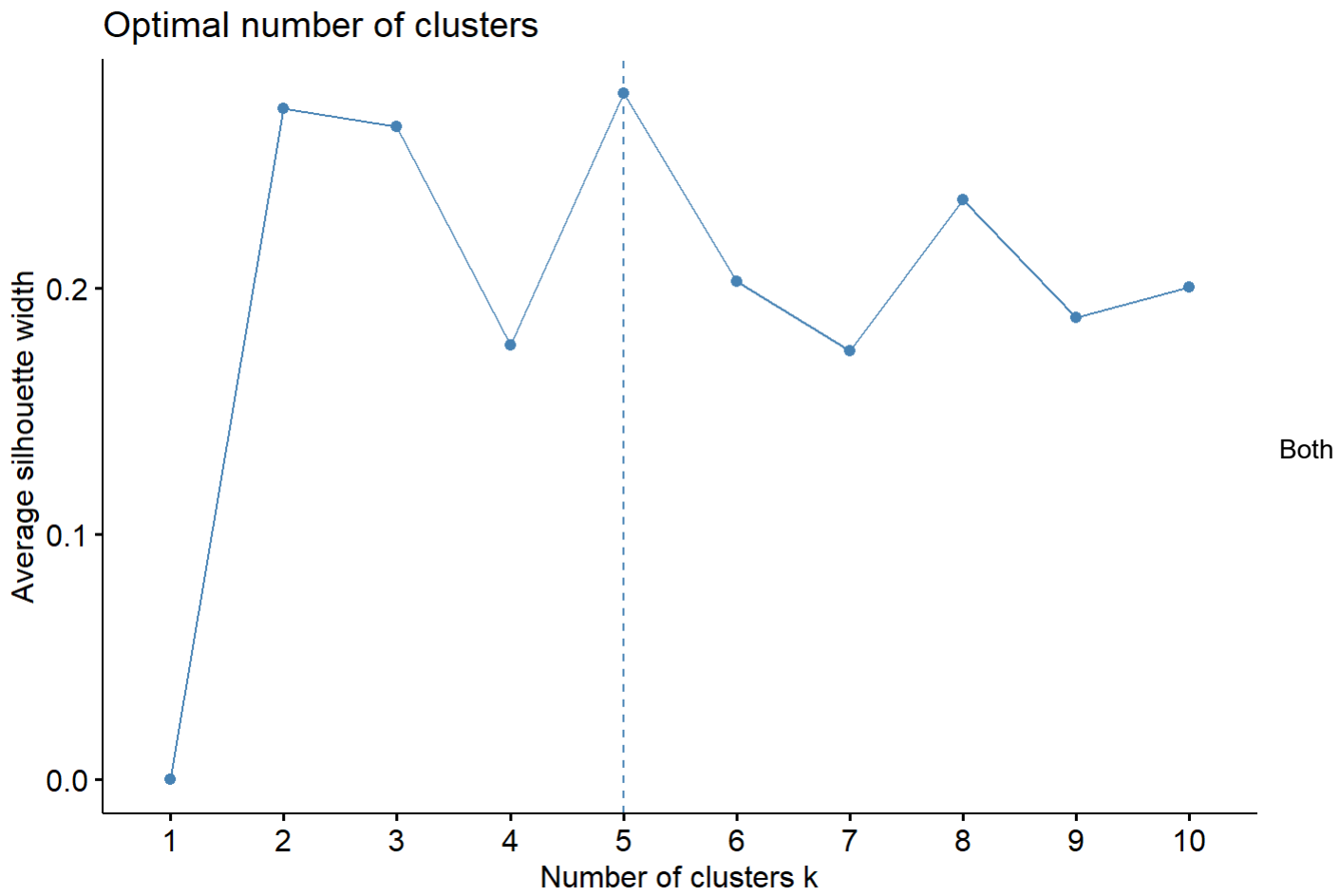
```
table(k4$cluster) ## show the size for each cluster
```

```
##
## 1 2 3 4
## 4 8 6 3
```

```
### Choosing the best K using elbow chart

set.seed(123)
df1<-df[,3:11]
dfnorm<-scale(df1)
fviz_nbclust(dfnorm, kmeans, method = "wss")
```




```
### Choosing the best K using Silhouette Method
set.seed(123)
df1<-df[,3:11]
dfnorm<-scale(df1)
fviz_nbclust(dfnorm, kmeans, method = "silhouette")
```



Both

Graphs shows the optimal K value is equal 5

```
##Running K-means Model using the optimal k value k=5
set.seed(123)
df1<-df[,3:11] ### Select only numerical variabls
dfnorm<-scale(df1)#Normalizing
k5<-kmeans(dfnorm, centers = 5, nstart = 25)# Running with clusters K=5
df1$Cluster <-k5$cluster
df1$MedianRec <-df[,c(12)]
df1$Location <-df[, c(13)]
df1$Exchange <-df[, c(14)]
Clus2<-df1[order(df1$Cluster),]
Clus2
```

```
##    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
## 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
## 7       51.33 0.50     13.9 34.8 15.1            0.9     0.57       2.70
## 10      73.84 0.18     27.9 31.0 13.5            0.6     0.53       6.21
## 16      96.65 0.19     21.6 17.9 11.2            0.5     0.06      -2.69
## 19      34.10 0.51     18.9 22.6 13.3            0.8     0.00       8.56
## 21      48.19 0.63     13.1 54.9 13.4            0.6     1.12       0.36
## 6       16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
## 8        0.41 0.85     26.0 24.1  4.3            0.6     3.51       6.38
## 12       2.60 0.65     19.9 21.4  6.8            0.6     1.45      13.99
## 2        7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## 18      56.24 0.40     56.5 13.5  5.7            0.6     0.35      15.00
## 11     122.11 0.35     18.0 62.9 20.3            1.0     0.34      21.87
## 13     173.93 0.46     28.4 28.6 16.3            0.9     0.10       9.37
## 15     132.56 0.46     18.9 40.6 15.0            1.1     0.28      17.35
## 17     199.47 0.65     23.6 45.6 19.2            0.8     0.16      25.54
## 5       47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 9        0.78 1.08      3.6 15.1  5.1            0.3     1.07      34.21
## 14       1.20 0.75     28.6 11.2  5.4            0.3     0.93      30.37
## 20       3.26 0.24     18.4 10.2  6.8            0.5     0.20      29.18
##    Net_Profit_Margin Cluster     MedianRec    Location Exchange
## 1               16.1       1  Moderate Buy          US     NYSE
## 3               11.2       1    Strong Buy          UK     NYSE
## 4               18.0       1 Moderate Sell          UK     NYSE
## 7               20.6       1 Moderate Sell          US     NYSE
## 10              23.4       1          Hold          US     NYSE
## 16              22.4       1          Hold SWITZERLAND     NYSE
## 19              17.6       1          Hold          US     NYSE
## 21              25.5       1          Hold          US     NYSE
## 6                2.6       2          Hold     GERMANY     NYSE
## 8                7.5       2  Moderate Buy          US   NASDAQ
## 12              11.0       2          Hold          US     AMEX
## 2                5.5       3  Moderate Buy      CANADA     NYSE
## 18               7.3       3          Hold          US     NYSE
## 11              21.1       4          Hold          UK     NYSE
## 13              17.9       4  Moderate Buy          US     NYSE
## 15              14.1       4          Hold          US     NYSE
## 17              25.2       4  Moderate Buy          US     NYSE
## 5               12.9       5  Moderate Buy      FRANCE     NYSE
## 9               13.3       5 Moderate Sell     IRELAND     NYSE
## 14              21.3       5  Moderate Buy          US     NYSE
## 20              15.1       5 Moderate Sell          US     NYSE
```

```
k5$centers ## show cenroid valus
```

```
##    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 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
## 4  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 5 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.27449312 -0.7041516       0.556954446
## 2  1.36644699 -0.6912914      -1.320000179
## 3 -0.14170336 -0.1168459      -1.416514761
## 4 -0.46807818  0.4671788       0.591242521
## 5  0.06308085  1.5180158      -0.006893899
```

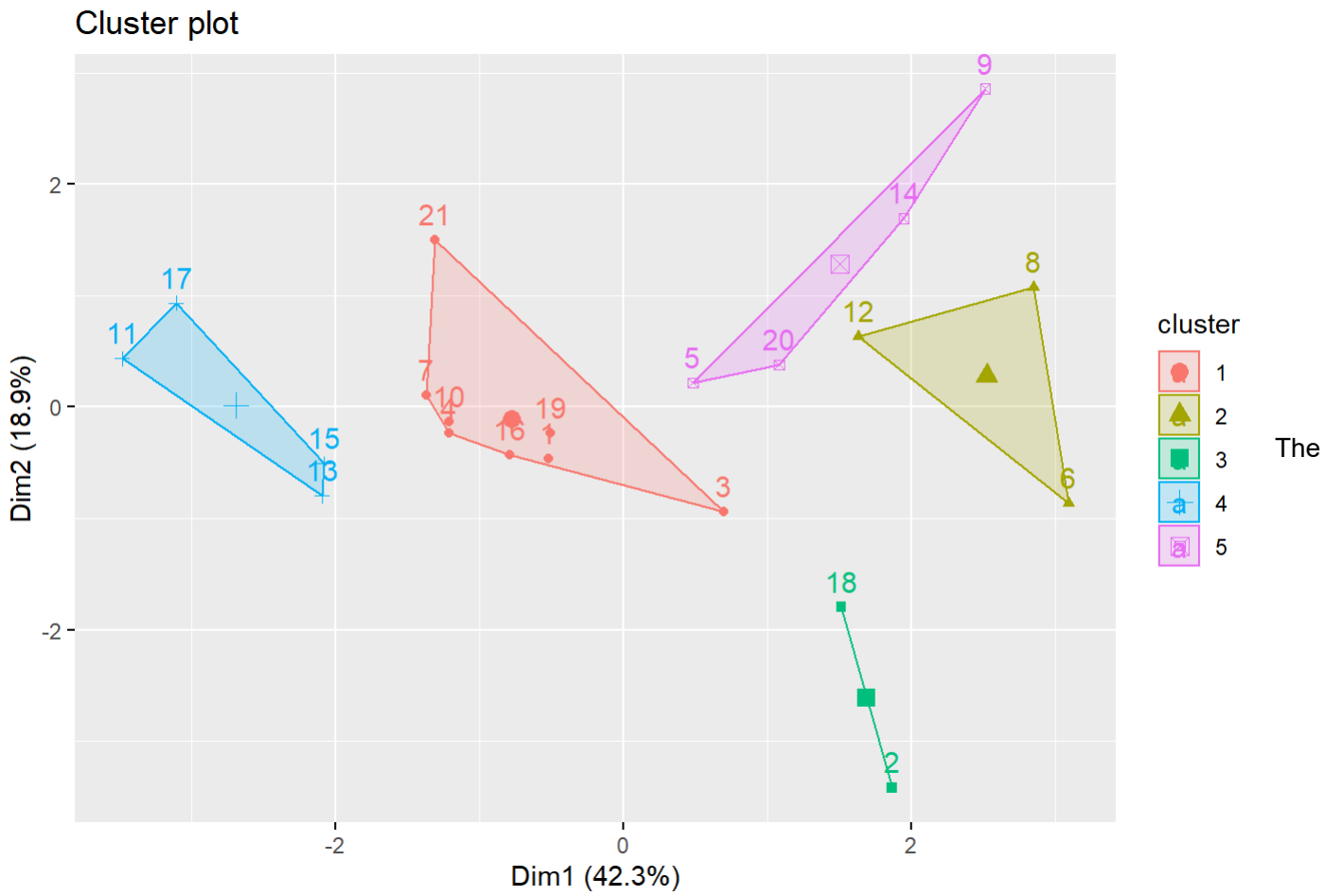
```
k5$size
```

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

```
table(k5$cluster)
```

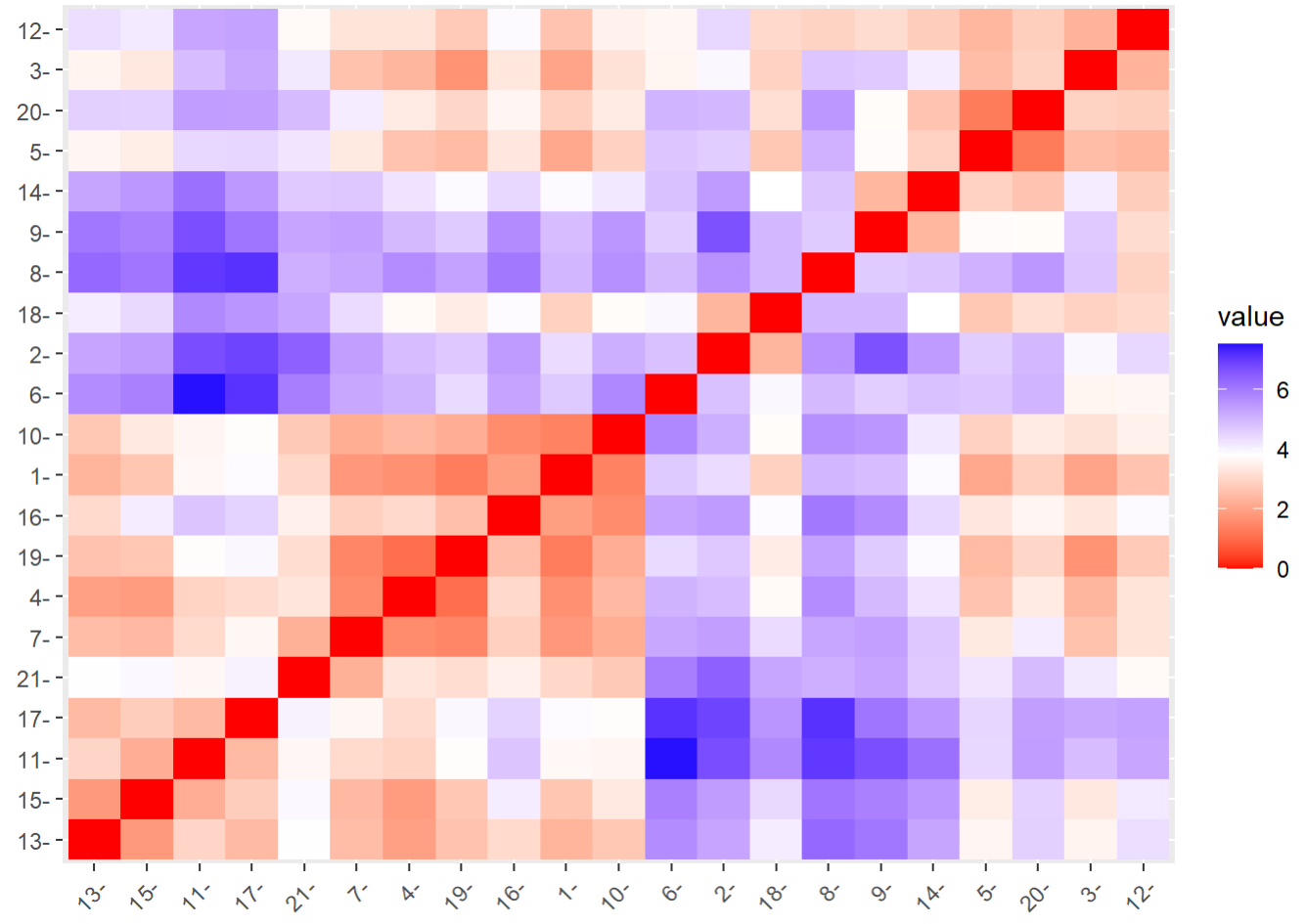
```
##
## 1 2 3 4 5
## 8 3 2 4 4
```

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



The points in the center of the shapes shows centroid points . I have tried Restart points grater than 25 , and the centroid values remains the same which means 25 restart points is goo enough .

```
### #kmeans clustering, using euclidean distance
distance<- dist(dfnorm, method = "euclidean")
fviz_dist(distance)
```




```
## For easier handling , will apply k=3
set.seed(123)
df1<-df[,3:11] ### Select only numerical variabls
dfnorm<-scale(df1)#Normalizing
k3<-kmeans(dfnorm, centers =3, nstart = 25)# Running with clusters K=3
df1$Cluster <-k3$cluster
df1$MedianRec <-df[,c(12)]
df1$Location <-df[, c(13)]
df1$Exchange <-df[, c(14)]
Clus3<-df1[order(df1$Cluster),]
Clus3
```

```
##    Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## 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
## 6       16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
## 18      56.24 0.40     56.5 13.5  5.7            0.6     0.35      15.00
## 1       68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## 4       67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## 7       51.33 0.50     13.9 34.8 15.1            0.9     0.57       2.70
## 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
## 13     173.93 0.46     28.4 28.6 16.3            0.9     0.10       9.37
## 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
## 19      34.10 0.51     18.9 22.6 13.3            0.8     0.00       8.56
## 21      48.19 0.63     13.1 54.9 13.4            0.6     1.12       0.36
## 5       47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 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
## 12       2.60 0.65     19.9 21.4  6.8            0.6     1.45      13.99
## 14       1.20 0.75     28.6 11.2  5.4            0.3     0.93      30.37
## 20       3.26 0.24     18.4 10.2  6.8            0.5     0.20      29.18
##    Net_Profit_Margin Cluster     MedianRec    Location Exchange
## 2                5.5       1  Moderate Buy      CANADA     NYSE
## 3               11.2       1    Strong Buy          UK     NYSE
## 6                2.6       1          Hold     GERMANY     NYSE
## 18               7.3       1          Hold          US     NYSE
## 1               16.1       2  Moderate Buy          US     NYSE
## 4               18.0       2 Moderate Sell          UK     NYSE
## 7               20.6       2 Moderate Sell          US     NYSE
## 10              23.4       2          Hold          US     NYSE
## 11              21.1       2          Hold          UK     NYSE
## 13              17.9       2  Moderate Buy          US     NYSE
## 15              14.1       2          Hold          US     NYSE
## 16              22.4       2          Hold SWITZERLAND     NYSE
## 17              25.2       2  Moderate Buy          US     NYSE
## 19              17.6       2          Hold          US     NYSE
## 21              25.5       2          Hold          US     NYSE
## 5               12.9       3  Moderate Buy      FRANCE     NYSE
## 8                7.5       3  Moderate Buy          US   NASDAQ
## 9               13.3       3 Moderate Sell     IRELAND     NYSE
## 12              11.0       3          Hold          US     AMEX
## 14              21.3       3  Moderate Buy          US     NYSE
## 20              15.1       3 Moderate Sell          US     NYSE
```

```
k3$centers ## show cenroid values
```

```
##   Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.6125361  0.2698666  1.3143935 -0.9609057 -1.0174553      0.2306328
## 2  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
## 3 -0.8261772  0.4775991 -0.3696184 -0.5631589 -0.8514589     -0.9994088
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3592866 -0.5757385        -1.3784169
## 2 -0.3331068 -0.2902163         0.6823310
## 3  0.8502201  0.9158889        -0.3319956
```

```
k3$size
```

```
## [1]  4 11  6
```

```
table(k3$cluster)
```

```
##
##  1  2  3
##  4 11  6
```

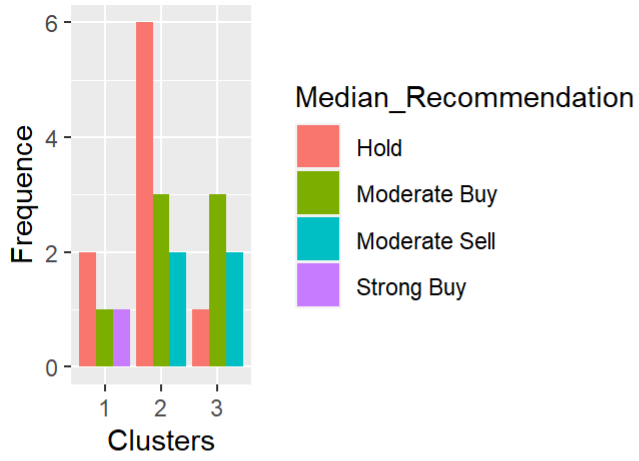
```
##Statistics Summary By each Cluster

describeBy(x=Clus3,group="Cluster",skew = FALSE)
```

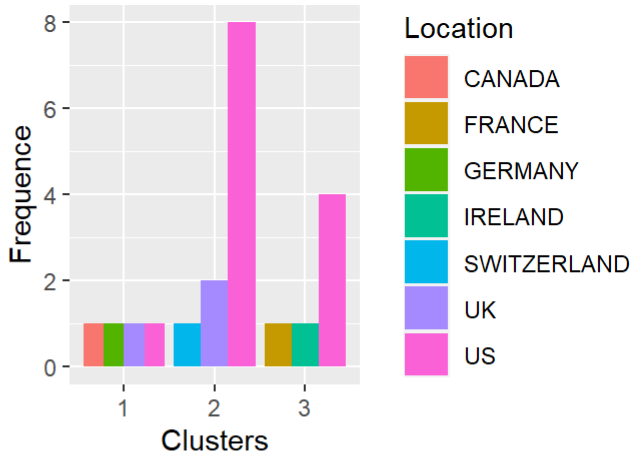
```
##
##  Descriptive statistics by group
## Cluster: 1
##                   vars n  mean    sd   min   max range    se
## Market_Cap           1 4 21.75 23.47  6.30 56.24 49.94 11.74
## Beta                 2 4  0.60  0.34  0.40  1.11  0.71  0.17
## PE_Ratio             3 4 46.90 28.33 20.70 82.50 61.80 14.16
## ROE                  4 4 11.30  5.00  3.90 14.90 11.00  2.50
## ROA                  5 4  5.10  2.68  1.40  7.80  6.40  1.34
## Asset_Turnover       6 4  0.75  0.17  0.60  0.90  0.30  0.09
## Leverage             7 4  0.30  0.25  0.00  0.60  0.60  0.12
## Rev_Growth           8 4  7.01  7.57 -3.17 15.00 18.17  3.79
## Net_Profit_Margin    9 4  6.65  3.60  2.60 11.20  8.60  1.80
## Cluster             10 4  1.00  0.00  1.00  1.00  0.00  0.00
## MedianRec*          11 4  1.75  0.96  1.00  3.00  2.00  0.48
## Location*           12 4  2.50  1.29  1.00  4.00  3.00  0.65
## Exchange*           13 4  1.00  0.00  1.00  1.00  0.00  0.00
## ------------------------------------------------------------
## Cluster: 2
##                   vars  n  mean    sd   min    max  range    se
## Market_Cap           1 11 97.11 53.81 34.10 199.47 165.37 16.22
## Beta                 2 11  0.43  0.16  0.18   0.65   0.47  0.05
## PE_Ratio             3 11 20.95  5.04 13.10  28.40  15.30  1.52
## ROE                  4 11 35.70 13.96 17.90  62.90  45.00  4.21
## ROA                  5 11 14.95  2.83 11.20  20.30   9.10  0.85
## Asset_Turnover       6 11  0.80  0.18  0.50   1.10   0.60  0.06
## Leverage             7 11  0.33  0.33  0.00   1.12   1.12  0.10
## Rev_Growth           8 11 10.16  8.90 -2.69  25.54  28.23  2.68
## Net_Profit_Margin    9 11 20.17  3.74 14.10  25.50  11.40  1.13
## Cluster             10 11  2.00  0.00  2.00   2.00   0.00  0.00
## MedianRec*          11 11  1.64  0.81  1.00   3.00   2.00  0.24
## Location*           12 11  2.64  0.67  1.00   3.00   2.00  0.20
## Exchange*           13 11  1.00  0.00  1.00   1.00   0.00  0.00
## ------------------------------------------------------------
## Cluster: 3
##                   vars n  mean    sd   min   max range   se
## Market_Cap           1 6  9.23 18.61  0.41 47.16 46.75 7.60
## Beta                 2 6  0.65  0.32  0.24  1.08  0.84 0.13
## PE_Ratio             3 6 19.43  8.71  3.60 28.60 25.00 3.56
## ROE                  4 6 17.30  5.93 10.20 24.10 13.90 2.42
## ROA                  5 6  5.98  1.23  4.30  7.50  3.20 0.50
## Asset_Turnover       6 6  0.48  0.15  0.30  0.60  0.30 0.06
## Leverage             7 6  1.25  1.20  0.20  3.51  3.31 0.49
## Rev_Growth           8 6 23.49 10.85  6.38 34.21 27.83 4.43
## Net_Profit_Margin    9 6 13.52  4.61  7.50 21.30 13.80 1.88
## Cluster             10 6  3.00  0.00  3.00  3.00  0.00 0.00
## MedianRec*          11 6  2.17  0.75  1.00  3.00  2.00 0.31
## Location*           12 6  2.50  0.84  1.00  3.00  2.00 0.34
## Exchange*           13 6  2.50  0.84  1.00  3.00  2.00 0.34
```

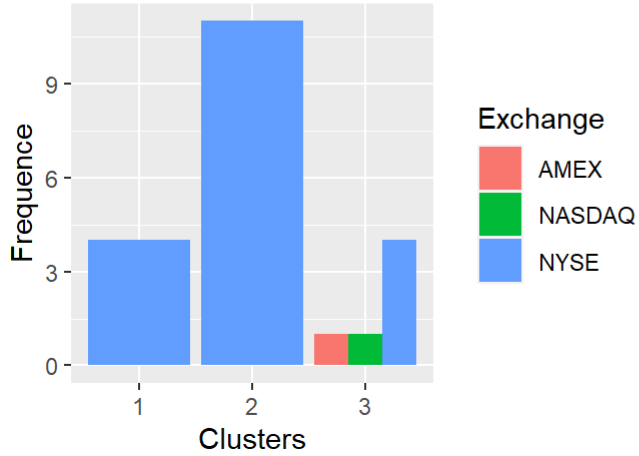
```
###link clusters with non-numerical variabls

df3<-  df %>% select(c(12,13,14)) %>%
    mutate(Cluster =k3$cluster)
df3
```

```
##    Median_Recommendation    Location Exchange Cluster
## 1           Moderate Buy          US     NYSE       2
## 2           Moderate Buy      CANADA     NYSE       1
## 3             Strong Buy          UK     NYSE       1
## 4          Moderate Sell          UK     NYSE       2
## 5           Moderate Buy      FRANCE     NYSE       3
## 6                   Hold     GERMANY     NYSE       1
## 7          Moderate Sell          US     NYSE       2
## 8           Moderate Buy          US   NASDAQ       3
## 9          Moderate Sell     IRELAND     NYSE       3
## 10                  Hold          US     NYSE       2
## 11                  Hold          UK     NYSE       2
## 12                  Hold          US     AMEX       3
## 13          Moderate Buy          US     NYSE       2
## 14          Moderate Buy          US     NYSE       3
## 15                  Hold          US     NYSE       2
## 16                  Hold SWITZERLAND     NYSE       2
## 17          Moderate Buy          US     NYSE       2
## 18                  Hold          US     NYSE       1
## 19                  Hold          US     NYSE       2
## 20         Moderate Sell          US     NYSE       3
## 21                  Hold          US     NYSE       2
```

```
#plotting clusters with non-numerical variabls

med_rec <- ggplot(df3, mapping = aes(factor(Cluster), fill=Median_Recommendation)) +
  geom_bar(position = 'dodge') +
  labs(x='Clusters', y='Frequence')

location<-ggplot(df3, mapping = aes(factor(Cluster), fill=Location)) +
  geom_bar(position = 'dodge') +
  labs(x='Clusters', y='Frequence')

exchange<- ggplot(df3, mapping = aes(factor(Cluster), fill=Exchange)) +
 geom_bar(position = 'dodge') +
  labs(x='Clusters', y='Frequence')
plot_grid(med_rec, location,exchange)
```







There is no clear trend for those categorical variables

```
###  Observations from the statistical summary

####Cluster 1 has low average of revenue growth and profit margin, with relatively high Beta , s
o it is not recommended

####Cluster 2  has the highest mean for market capitalization = 97.11 , which looks interesting
 to invest at first look , but it has also high standard deviation which means some of the point
s has extreme values needed more investigation  .

####Cluster 3 has the highest average of revenue growth and profit margin with relatively few st
andard deviation, so it is recommended
```

```
### Rename cluster values

Clus3["Cluster"][Clus3["Cluster"] == "1"] <- "not recommended"
Clus3["Cluster"][Clus3["Cluster"] == "2"] <- "needed more investigation"
Clus3["Cluster"][Clus3["Cluster"] == "3"] <- "Recommended"
Clus3
```

```
##    Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## 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
## 6       16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
## 18      56.24 0.40     56.5 13.5  5.7            0.6     0.35      15.00
## 1       68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## 4       67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## 7       51.33 0.50     13.9 34.8 15.1            0.9     0.57       2.70
## 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
## 13     173.93 0.46     28.4 28.6 16.3            0.9     0.10       9.37
## 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
## 19      34.10 0.51     18.9 22.6 13.3            0.8     0.00       8.56
## 21      48.19 0.63     13.1 54.9 13.4            0.6     1.12       0.36
## 5       47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 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
## 12       2.60 0.65     19.9 21.4  6.8            0.6     1.45      13.99
## 14       1.20 0.75     28.6 11.2  5.4            0.3     0.93      30.37
## 20       3.26 0.24     18.4 10.2  6.8            0.5     0.20      29.18
##    Net_Profit_Margin                   Cluster     MedianRec    Location
## 2                5.5           not recommended  Moderate Buy      CANADA
## 3               11.2           not recommended    Strong Buy          UK
## 6                2.6           not recommended          Hold     GERMANY
## 18               7.3           not recommended          Hold          US
## 1               16.1 needed more investigation  Moderate Buy          US
## 4               18.0 needed more investigation Moderate Sell          UK
## 7               20.6 needed more investigation Moderate Sell          US
## 10              23.4 needed more investigation          Hold          US
## 11              21.1 needed more investigation          Hold          UK
## 13              17.9 needed more investigation  Moderate Buy          US
## 15              14.1 needed more investigation          Hold          US
## 16              22.4 needed more investigation          Hold SWITZERLAND
## 17              25.2 needed more investigation  Moderate Buy          US
## 19              17.6 needed more investigation          Hold          US
## 21              25.5 needed more investigation          Hold          US
## 5               12.9               Recommended  Moderate Buy      FRANCE
## 8                7.5               Recommended  Moderate Buy          US
## 9               13.3               Recommended Moderate Sell     IRELAND
## 12              11.0               Recommended          Hold          US
## 14              21.3               Recommended  Moderate Buy          US
## 20              15.1               Recommended Moderate Sell          US
##    Exchange
## 2      NYSE
## 3      NYSE
## 6      NYSE
## 18     NYSE
## 1      NYSE
## 4      NYSE
## 7      NYSE
## 10     NYSE
```

```
## 11     NYSE
## 13     NYSE
## 15     NYSE
## 16     NYSE
## 17     NYSE
## 19     NYSE
## 21     NYSE
## 5      NYSE
## 8    NASDAQ
## 9      NYSE
## 12     AMEX
## 14     NYSE
## 20     NYSE
```