FML_ASSIGNMENT_4

Niharika Matsa

2023-11-12

Loading the required libraries

```
library(flexclust)

## Loading required package: grid

## Loading required package: lattice

## Loading required package: modeltools

## Loading required package: stats4

library(cluster)
library(tidyverse)

## ── 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 1.9.3     ✔ 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 become errors

library(factoextra)

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

library(FactoMineR)
library(ggcorrplot)
```

1. 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.

```r
data = read.csv("Pharmaceuticals.csv")
data
```

```
##    Symbol                                Name Market_Cap Beta PE_Ratio  ROE
ROA
## 1     ABT                 Abbott Laboratories      68.44 0.32     24.7 26.4
11.8
## 2     AGN                      Allergan, Inc.       7.58 0.41     82.5 12.9
5.5
## 3     AHM                        Amersham plc       6.30 0.46     20.7 14.9
7.8
## 4     AZN                    AstraZeneca PLC      67.63 0.52     21.5 27.4
15.4
## 5     AVE                             Aventis      47.16 0.32     20.1 21.8
7.5
## 6     BAY                            Bayer AG      16.90 1.11     27.9  3.9
1.4
## 7     BMY        Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8
15.1
## 8    CHTT                        Chattem, Inc       0.41 0.85     26.0 24.1
4.3
## 9     ELN              Elan Corporation, plc       0.78 1.08      3.6 15.1
5.1
## 10    LLY              Eli Lilly and Company      73.84 0.18     27.9 31.0
13.5
## 11    GSK                GlaxoSmithKline plc     122.11 0.35     18.0 62.9
20.3
## 12    IVX                     IVAX Corporation       2.60 0.65     19.9 21.4
6.8
## 13    JNJ                   Johnson & Johnson     173.93 0.46     28.4 28.6
16.3
## 14    MRX Medicis Pharmaceutical Corporation       1.20 0.75     28.6 11.2
5.4
## 15    MRK                   Merck & Co., Inc.     132.56 0.46     18.9 40.6
15.0
## 16    NVS                         Novartis AG      96.65 0.19     21.6 17.9
11.2
## 17    PFE                          Pfizer Inc     199.47 0.65     23.6 45.6
19.2
## 18    PHA               Pharmacia Corporation      56.24 0.40     56.5 13.5
5.7
## 19    SGP         Schering-Plough Corporation      34.10 0.51     18.9 22.6
13.3
## 20    WPI       Watson Pharmaceuticals, Inc.       3.26 0.24     18.4 10.2
6.8
## 21    WYE                              Wyeth      48.19 0.63     13.1 54.9
13.4
##    Asset_Turnover Leverage Rev_Growth Net_Profit_Margin
Median_Recommendation
## 1             0.7     0.42       7.54              16.1          Moderate
```

```
Buy
## 2                 0.9     0.60       9.16                 5.5          Moderate
Buy
## 3                 0.9     0.27       7.05                11.2            Strong
Buy
## 4                 0.9     0.00      15.00                18.0         Moderate
Sell
## 5                 0.6     0.34      26.81                12.9          Moderate
Buy
## 6                 0.6     0.00      -3.17                 2.6
Hold
## 7                 0.9     0.57       2.70                20.6         Moderate
Sell
## 8                 0.6     3.51       6.38                 7.5          Moderate
Buy
## 9                 0.3     1.07      34.21                13.3         Moderate
Sell
## 10                0.6     0.53       6.21                23.4
Hold
## 11                1.0     0.34      21.87                21.1
Hold
## 12                0.6     1.45      13.99                11.0
Hold
## 13                0.9     0.10       9.37                17.9          Moderate
Buy
## 14                0.3     0.93      30.37                21.3          Moderate
Buy
## 15                1.1     0.28      17.35                14.1
Hold
## 16                0.5     0.06      -2.69                22.4
Hold
## 17                0.8     0.16      25.54                25.2          Moderate
Buy
## 18                0.6     0.35      15.00                 7.3
Hold
## 19                0.8     0.00       8.56                17.6
Hold
## 20                0.5     0.20      29.18                15.1         Moderate
Sell
## 21                0.6     1.12       0.36                25.5
Hold
##        Location Exchange
## 1            US     NYSE
## 2        CANADA     NYSE
## 3            UK     NYSE
## 4            UK     NYSE
## 5        FRANCE     NYSE
## 6       GERMANY     NYSE
## 7            US     NYSE
## 8            US   NASDAQ
```

```
## 9      IRELAND     NYSE
## 10          US     NYSE
## 11          UK     NYSE
## 12          US     AMEX
## 13          US     NYSE
## 14          US     NYSE
## 15          US     NYSE
## 16 SWITZERLAND     NYSE
## 17          US     NYSE
## 18          US     NYSE
## 19          US     NYSE
## 20          US     NYSE
## 21          US     NYSE
```

```
Pharmaceuticals = data[3:11]
head(Pharmaceuticals)
```

```
##   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
##   Net_Profit_Margin
## 1              16.1
## 2               5.5
## 3              11.2
## 4              18.0
## 5              12.9
## 6               2.6
```

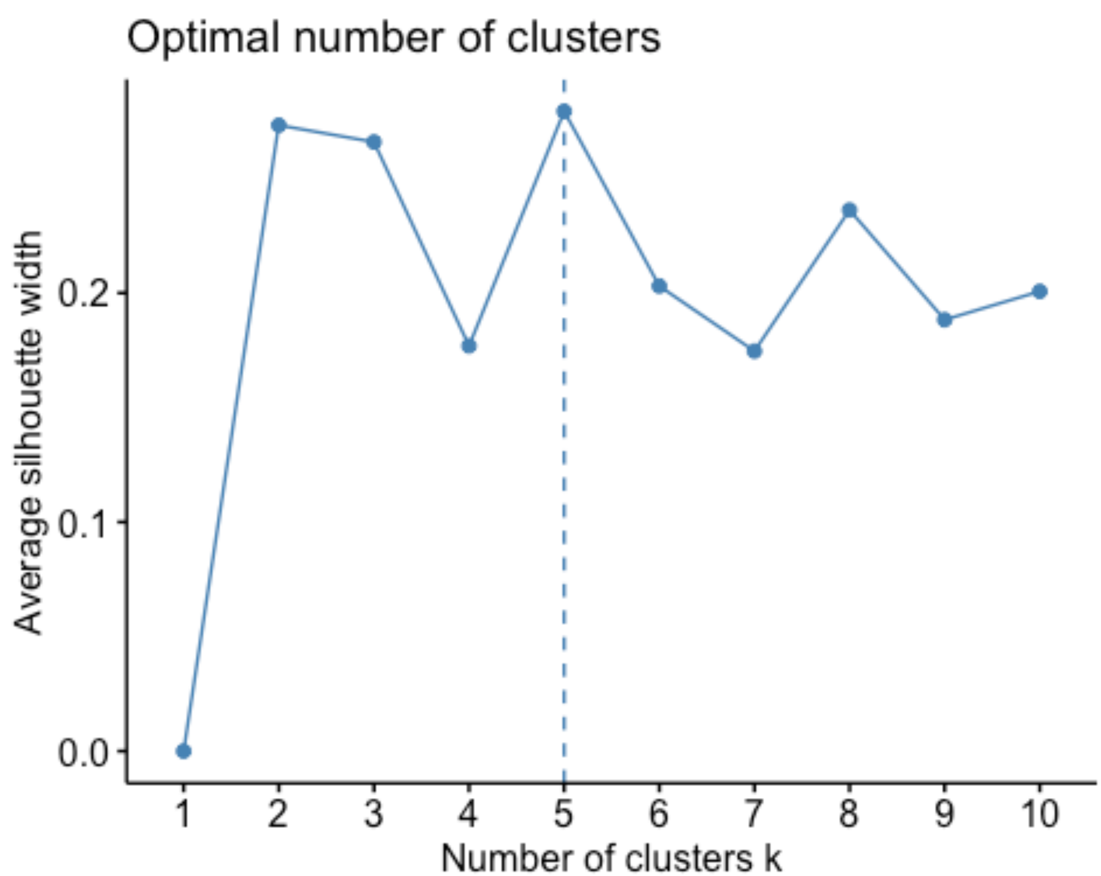
```
summary(Pharmaceuticals)
```

```
##    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
```

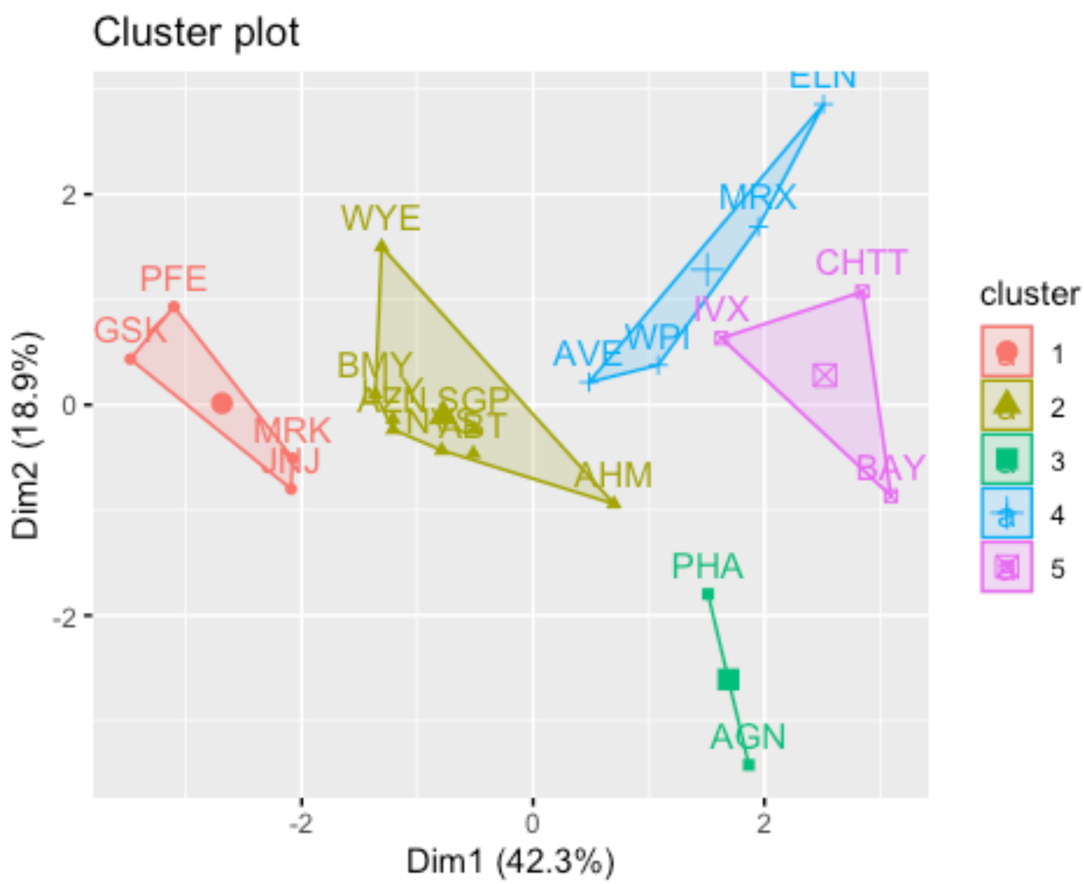
Normalizing the data

```
Pharma = scale(Pharmaceuticals)
row.names(Pharma) = data[,1]
distance = get_dist(Pharma)
correlation = cor(Pharma)

fviz_nbclust(Pharma, kmeans, method = "silhouette")
```



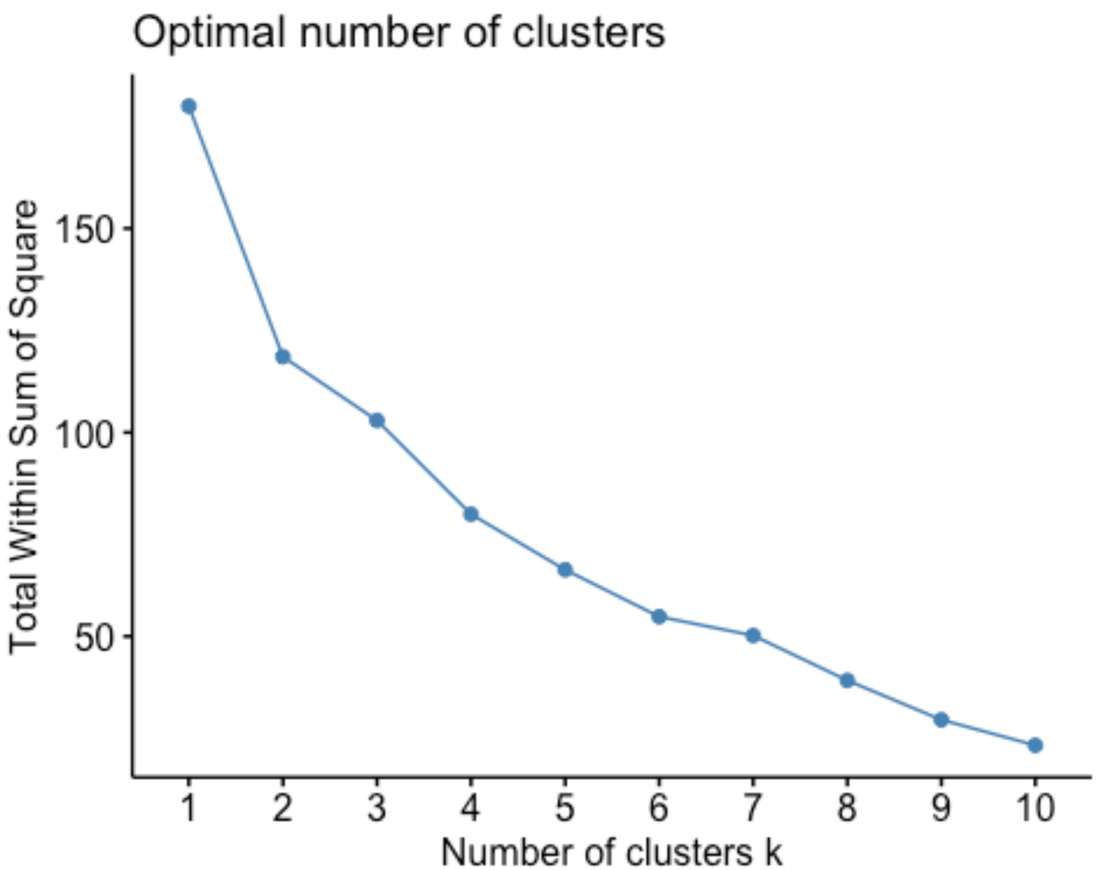

```
set.seed(69)
k5 = kmeans(Pharma, centers = 5, nstart = 30)
k5$size

## [1] 4 8 2 4 3

k5$centers

##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 2 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
```

```
## 3 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
## 4 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 5 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.46807818  0.4671788       0.591242521
## 2 -0.27449312 -0.7041516       0.556954446
## 3 -0.14170336 -0.1168459      -1.416514761
## 4  0.06308085  1.5180158      -0.006893899
## 5  1.36644699 -0.6912914      -1.320000179

fviz_cluster(k5, data = Pharma)
```



elbow

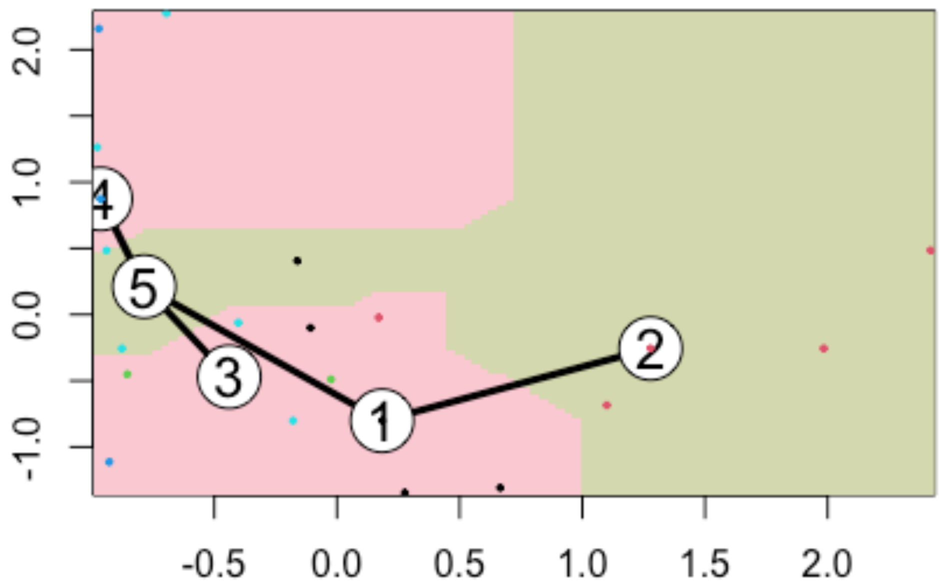
```
fviz_nbclust(Pharma, kmeans, method = "wss")
```



Manhattan

```
set.seed(50)
k51 = kcca(Pharma, k = 5, kccaFamily("kmedians"))
k51

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

clusters_index = predict(k51)
dist(k51@centers)

##          1        2        3        4
## 2 2.558034
## 3 4.451230 4.795056
## 4 4.222539 4.954336 4.589219
## 5 2.645989 3.581581 3.351236 2.857647
```

```r
image(k51)
points(Pharma, col = clusters_index, pch = 20, cex = 0.5)
```



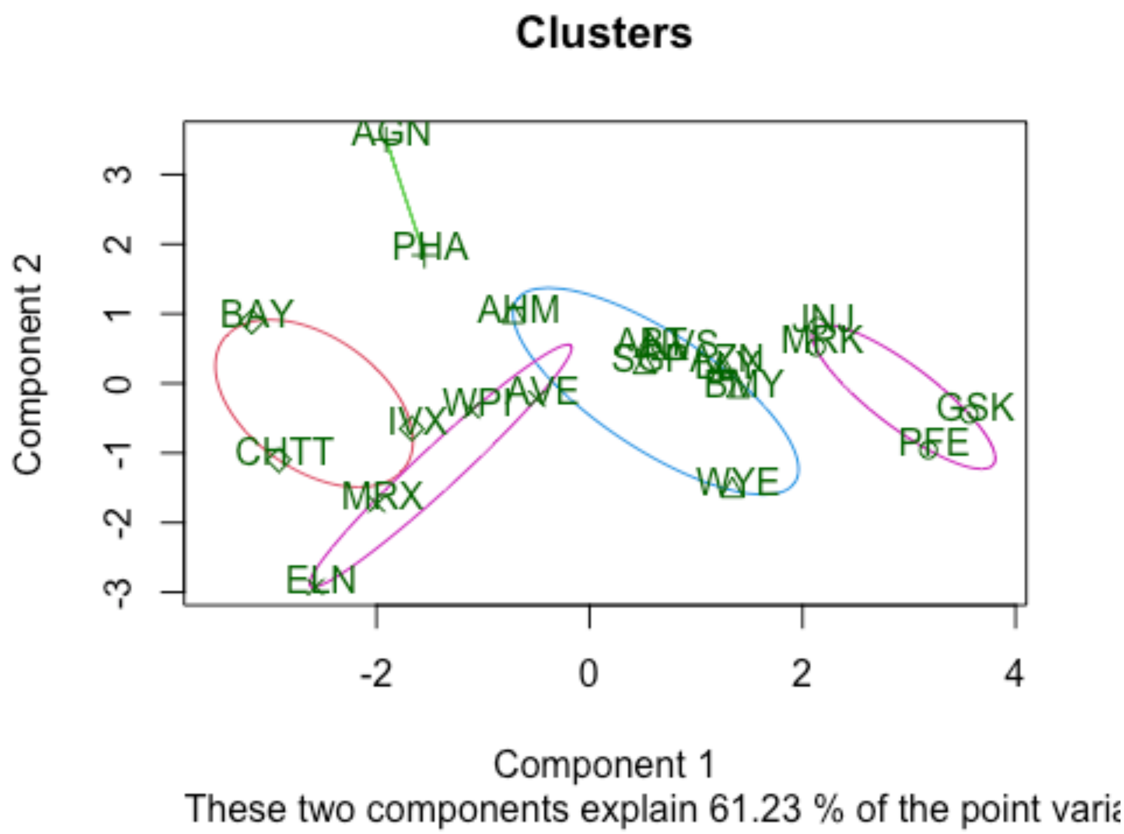
2.

Interpret the clusters with respect to the numerical variables used in forming the clusters. Is there a pattern in the clusters with respect to the numerical variables (10 to 12)? (those not used in forming the clusters)

```r
Pharmaceuticals %>% mutate(Cluster = k5$cluster) %>% group_by(Cluster) %>%
summarise_all("mean")

## # A tibble: 5 × 10
##   Cluster Market_Cap  Beta PE_Ratio   ROE   ROA Asset_Turnover Leverage
##     <int>      <dbl> <dbl>    <dbl> <dbl> <dbl>          <dbl>    <dbl>
## 1       1     157.   0.48      22.2  44.4 17.7           0.95     0.22
## 2       2      55.8  0.414     20.3  28.7 12.7           0.738    0.371
## 3       3      31.9  0.405     69.5  13.2  5.6           0.75     0.475
## 4       4      13.1  0.598     17.7  14.6  6.2           0.425    0.635
## 5       5       6.64 0.87      24.6  16.5  4.17          0.6      1.65
## # ℹ 2 more variables: Rev_Growth <dbl>, Net_Profit_Margin <dbl>
```

Interpretation:

```r
clusplot(Pharma,k5$cluster, main = "Clusters", color = TRUE, labels = 3,
lines = 0)
```



Below is the Cluster naming based on the companies:

Cluster 1: ELN, MRX, WPI and AVE

Cluster 2: AGN and PHA

Cluster 3: AHM,WYE,BMY,AZN, LLY, ABT, NVS and SGP

Cluster 4: BAY, CHTT and IVX

Cluster 5: JNJ, MRK, PFE and GSK

Interpretation

Cluster 1 - Best: Cluster stands out with the best Net Profit Margin, the lowest PE ratio, and rapid sales growth. This cluster is considered a strong candidate for purchase or holding as a reserve.

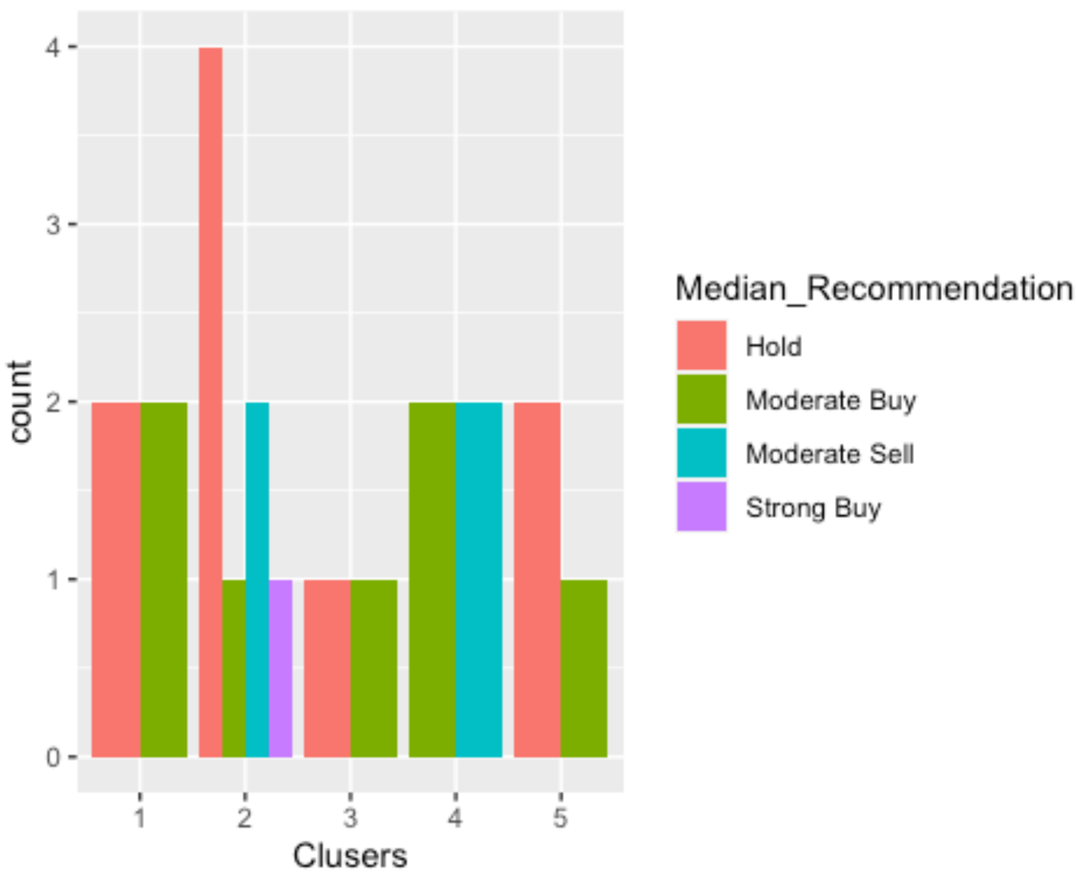
Cluster 2 - Substantial Risk: Cluster 2 is characterized by a notably high PE ratio, signaling potential overvaluation. Investors should approach this cluster with caution due to the elevated valuation.

Cluster 3 - Pursue: This Cluster represents a moderate-risk category. While not as extreme as some other clusters, careful consideration is still advised for entities in this group.

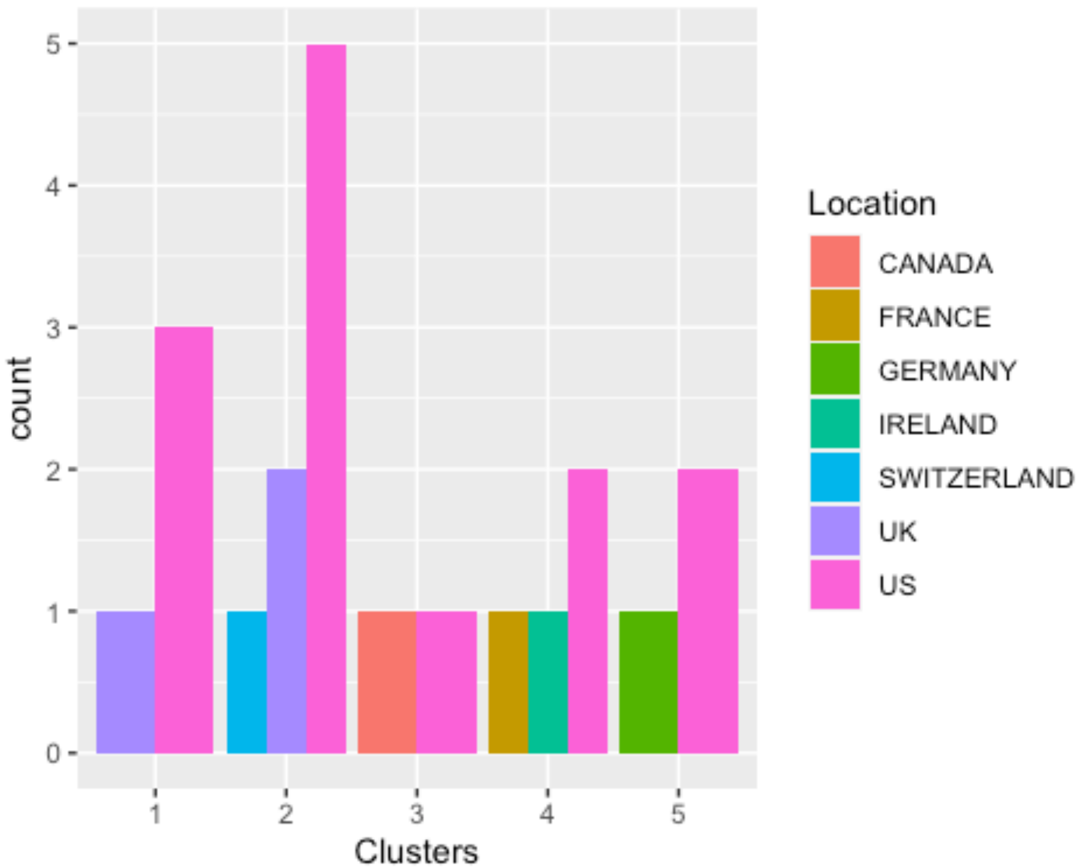
Cluster 4 - Deadly, Despite Excellent PE Ratio
Despite having an excellent PE ratio, Cluster 4 is marked by exceptionally high risk, driven by elevated leverage, poor Net Profit Margin, and very low revenue growth. Ownership of entities in this cluster is considered highly risky.

Cluster 5 - Fortune Overall Metrics This Cluster showcases robust market capitalization, ROI, ROA, asset turnover, and Net Profit Margin. With a moderately valued PE ratio, entities in this cluster are deemed favorable for purchase and retention. The substantial revenue growth of 18.5% adds to the attractiveness of this cluster.

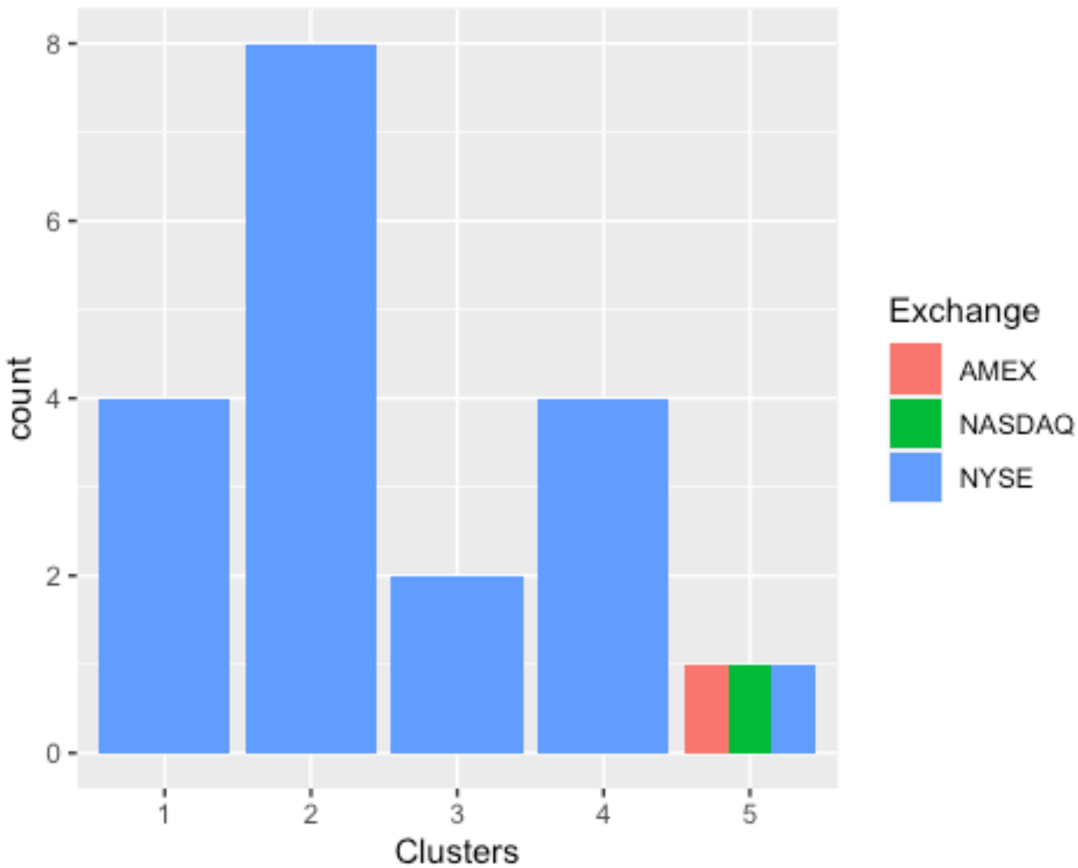
```
Pharmaceuticals1 = data[12:14] %>% mutate(Clusters = k5$cluster)
ggplot(Pharmaceuticals1,mapping=aes(factor(Clusters),fill=Median_Recommendati
on))+geom_bar(position = 'dodge')+labs(x='Clusers')
```




```
ggplot(Pharmaceuticals1, mapping = aes(factor(Clusters),fill =
Location))+geom_bar(position = 'dodge')+labs(x = 'Clusters')
```




```r
ggplot(Pharmaceuticals1, mapping = aes(factor(Clusters), fill =
Exchange))+geom_bar(position = 'dodge')+labs(x = 'Clusters')
```



A

pattern can be observed in the median suggestions.

The most of the clusters/companies are listed on the NYSE and are based in the United States, but other than that, there doesn't appear to be any discernible pattern among the clusters, locations, or exchanges.

Cluster Interpretation according to variables:

Cluster 1

Median Suggestion An average buy and sell suggestion is given for Cluster 1. **Location** There are three places in Cluster 1, the most notable being the United States. **Exchange** NYSE is the only one cluster in exchange.

Cluster 2

Median Suggestion Cluster 2 has a low hold and a low purchase. **Location** The United States and Canada are the only two locations in Cluster 2, and they are dispersed equally. **Exchange** NYSE is the only one cluster in exchange.

Cluster 3

Median Suggestion Cluster 3 has an extremely strong hold. **Location** Cluster 3 has three locations, and is dominated by the United States, followed by the United Kingdom and

Switzerland. **Exchange** There is only one exchange in Cluster 3, the NYSE, and it has a big user base.

Cluster 4

Median Suggestion With a low buy rating, cluster 4 is rated as strongly held. **Location** The US is ranked higher than Germany in two locations in Cluster 4. **Exchange** Three equally distributed exchanges (AMEX, NASDAQ and NYSE) are located in Cluster 4.

Cluster 5

Median Suggestion A high buy and high hold rating are assigned to Cluster 5, based on the median recommendation. **Location** There are two locations for Cluster 5, with a significant majority of the United States and the United Kingdom. **Exchange** NYSE is the only one cluster in exchange.

3.Provide an appropriate name for each cluster using any or all of the variables in the dataset.

To name the clusters i have considered all the numerical variables below is the interpretations:

Cluster 1: High Profitability & Growth Leaders This cluster excels in Net Profit Margin, has the lowest PE ratio, and experiences rapid sales growth. It is named for its emphasis on profitability and growth potential.

Cluster 2: High Beta, Elevated PE Warning Characterized by a notably high Beta and a warning for an elevated PE ratio, Cluster 2 is named for its emphasis on market sensitivity and the cautionary signal regarding valuation.

Cluster 3: Moderate Risk, Balanced Metrics Representing a moderate-risk category, Cluster 3 is named for its balance across various metrics. It avoids extremes and may offer a balanced risk-return profile.

Cluster 4: High Risk, Low Profitability Despite a strong PE ratio, Cluster 4 carries high risk due to elevated leverage, poor Net Profit Margin, and low revenue growth. It is named for its high-risk nature and lower profitability.

Cluster 5: Robust Metrics & Growth Potential Cluster 5 is named for its robust market capitalization, strong Return on Equity (ROE), Return on Assets (ROA), and growth potential indicated by substantial revenue growth. It represents entities with solid fundamentals and growth prospects.