Assignment4

krupa

2023-11-12

Installing Required Libraries

```
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(flexclust)
```

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

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

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

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

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

```
library(FactoMineR)
```

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

```
library(ggcorrplot)
```

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

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

```
library(factoextra)
```

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

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

```
library(cluster)
```

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

Printing pharma

```
pharma= read.csv("Pharmaceuticals.csv")
pharma_filter = pharma[3:11]
head(pharma_filter)

##   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
```

Reason for Choosing Variables

Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth, and Net_Profit_Margin are common financial metrics used to evaluate and compare company performance. These variables provide a comprehensive picture of a company's financial health, profitability, and efficiency.

1. Market_cap:

The value ranges from 0.41 to 199.47. The overall size and valuation of pharmaceutical firms are indicated.

2. Beta:

varies between 0.18 and 1.11. The sensitivity of a firm's returns to market fluctuations is measured.

3. PE_Ratio:

Ranges from 3.6 to 82.5 and represents the value of a company's stock in relation to its earnings.

4. ROE:

The scale runs from 3.9 to 62.9. Indicates how well a company uses shareholder equity to generate profit.

5. ROA:

The value ranges from 0.3 to 1.1. This metric assesses a company's ability to generate profit from its assets.

6. Asset_Turnover:

The value ranges from 0.5 to 1.1. Represents how effectively a company uses its assets to generate revenue.

7. Leverage:

The value ranges from 0 to 3.51. The extent to which a company uses debt to finance its operations.

8. Rev_Growth:

The value ranges from -3.17 to 34.21. The percentage change in revenue over a specific time period.

9.Net_Profit_Margin:

The value ranges from 2.6 to 25.54. The percentage of revenue that is converted into profit.

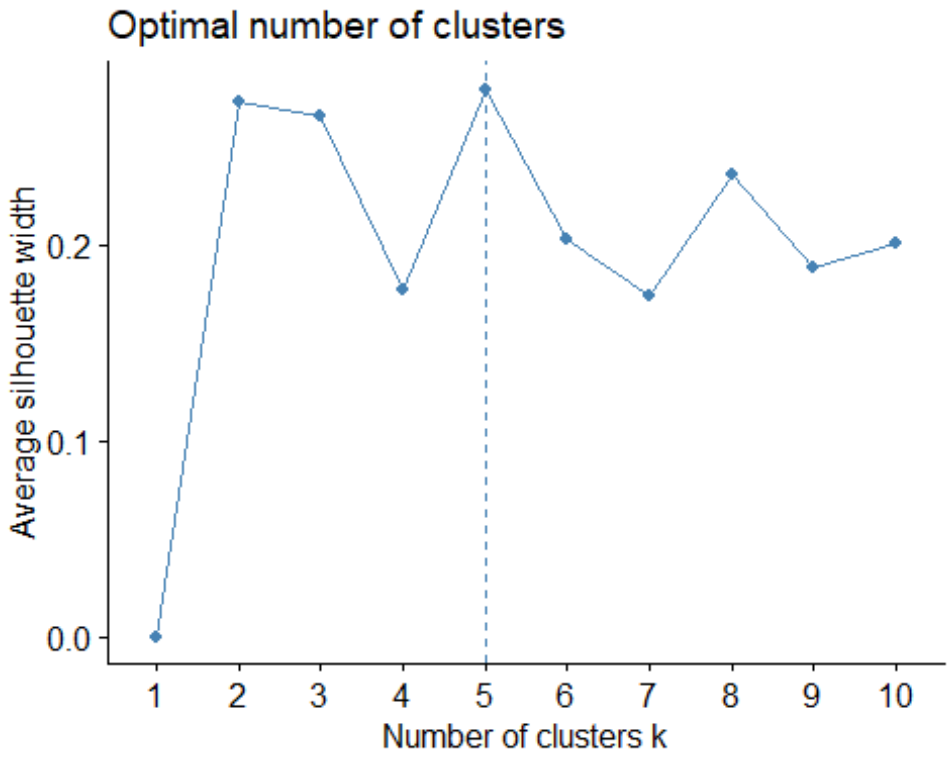
Normalising pharma

```
pharma_norm = scale(pharma_filter)
row.names(pharma_norm) = pharma[,1]
distance = get_dist(pharma_norm)
corr = cor(pharma_norm)
```

Normalization

The numerical variables must be normalized to ensure that each variable contributes proportionally to the clustering process. Because these variables may have varying units or scales, normalizing them prevents one variable from dominating the clustering based on its magnitude. Market_Cap, for example, is in the hundreds, whereas Beta is a fraction between 0 and 1.

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



Choosing 5 clusters

Silhouette analysis compares the similarity of an object to its own cluster to other clusters. It displays a graphical representation of cluster quality for various k values.

```
set.seed(1)
k5 = kmeans(pharma_norm, centers = 5, nstart = 25)
k5$centers
```

```
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
## 2 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951      0.2306328
## 3 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915      0.1729746
## 4 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478     -0.4612656
## 5  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
##      Leverage Rev_Growth Net_Profit_Margin
## 1  0.06308085  1.5180158      -0.006893899
## 2 -0.14170336 -0.1168459      -1.416514761
## 3 -0.27449312 -0.7041516       0.556954446
## 4  1.36644699 -0.6912914      -1.320000179
## 5 -0.46807818  0.4671788       0.591242521
```

Selecting K-means

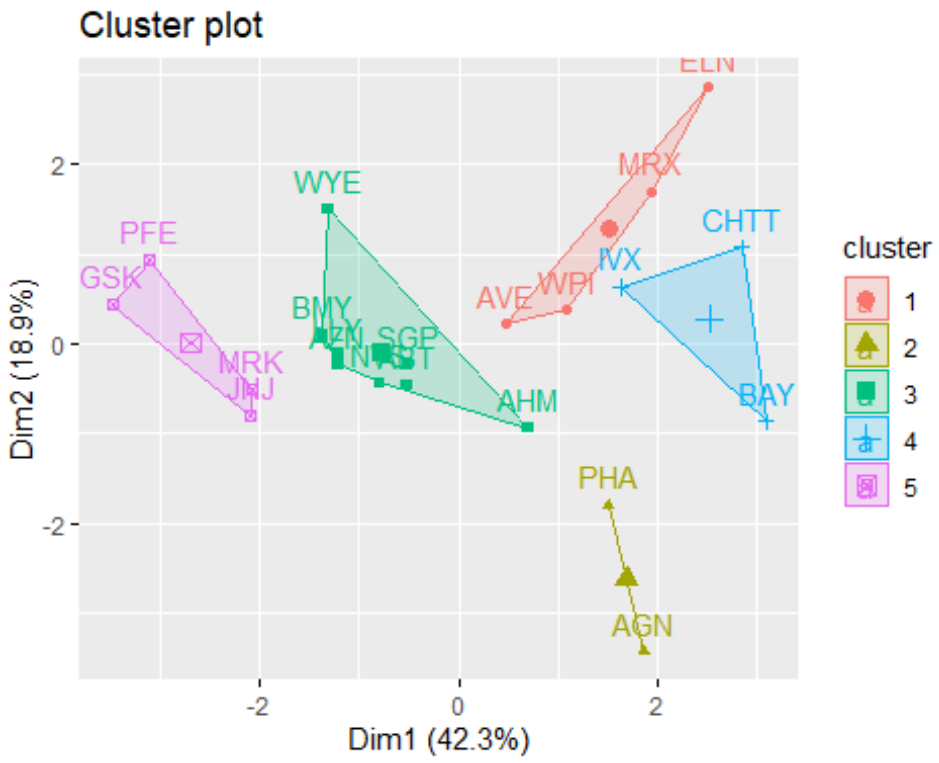
K-means is often used in exploratory pharma analysis to identify patterns and groupings within the pharma, and K-means clustering can provide insights into the financial profiles

of pharmaceutical firms. It can identify groups of firms with similar financial characteristics, which can help with strategic decision-making or investment analysis.

```
k5$size
```

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

```
fviz_cluster(k5, data = pharma_norm)
```



Appropriate Name:

Cluster 1 - Growth Pioneers

Cluster 2 - Speculative Stars

Cluster 3 - Market Titans

Cluster 4 - Beta Bulls

Cluster 5 - Blue Chip Triumph

Interpretation of Clusters based on Cluster Variables:

Cluster 1: AVE, WPI, MRX, and ELN have moderate Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth, and Net_Profit_Margin values. Recognized for its exceptional Net Profit Margin, the lowest PE ratio, and rapid sales growth, Cluster 1 emerges as an optimal choice for strategic investment or holding as a reserve.

Cluster 2: PHA and AGN have lower Market_Cap, Beta, and PE_Ratio values.Highlighted by a notably high PE ratio, Cluster 2 issues a cautionary signal regarding potential overvaluation. Investors are advised to approach this cluster with careful consideration, acknowledging the elevated valuation.

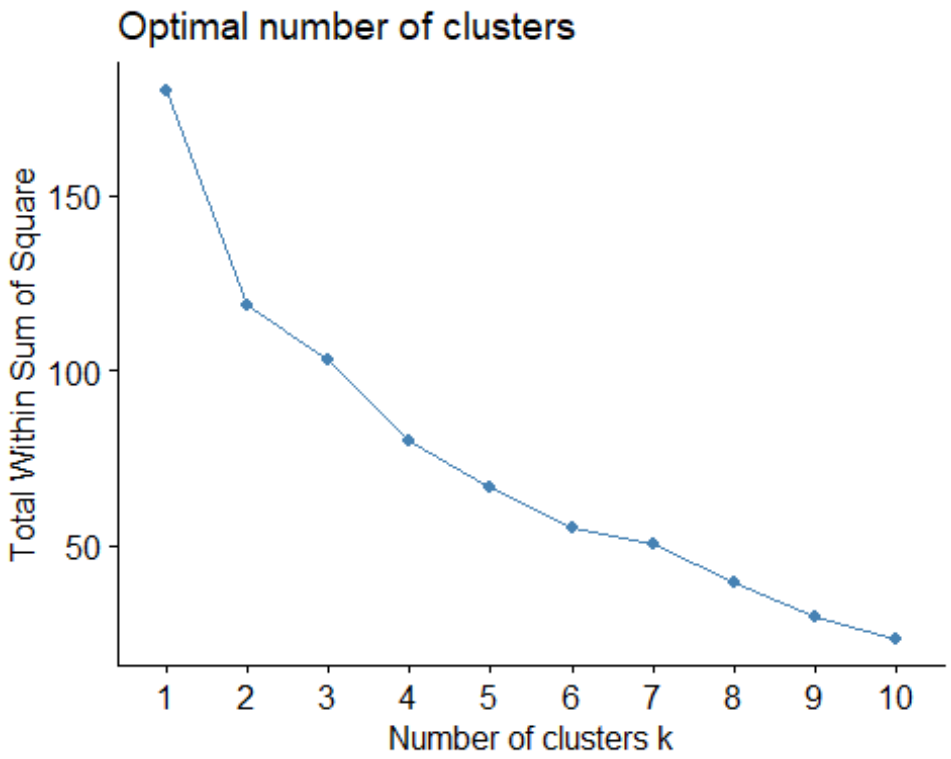
Cluster 3: WYE, BMY, AZN, SGP, AHM, LLY,NVS, and ABT have higher Market_Cap, Beta, PE_Ratio, Rev_Growth, and Net_Profit_Margin values than the other clusters. Cluster 3 epitomizes a moderate-risk category. While not as extreme as other clusters, entities in this group demand thoughtful consideration, striking a balance between risk and potential return.

Cluster 4: IVX, CHTT, and BAY have lower Market_Cap and PE_Ratio values.Despite showcasing an excellent PE ratio, Cluster 4 carries significant risk due to elevated leverage, poor Net Profit Margin, and very low revenue growth. Ownership of entities in this cluster is regarded as highly precarious.

Cluster 5: GSK, PFE, MRK, and JNJ have higher Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Rev_Growth, and Net_Profit_Margin values.cluster 5 unfolds with formidable market capitalization, ROI, ROA, asset turnover, and Net Profit Margin. Entities in this cluster, featuring a moderately valued PE ratio, are highly favorable for both purchase and retention. The substantial revenue growth of 18.5% further amplifies the appeal of this cluster.

Elbow

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



Manhattan

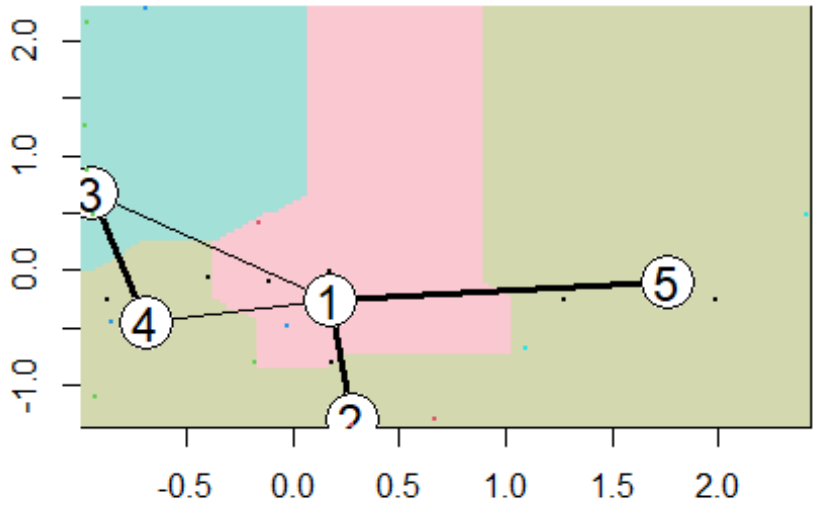
```
set.seed(1)
k51 = kcca(pharma_norm, k=5, kccaFamily("kmedians"))
k51

## kcca object of family 'kmedians'
##
## call:
## kcca(x = pharma_norm, k = 5, family = kccaFamily("kmedians"))
##
## cluster sizes:
##
## 1 2 3 4 5
## 7 3 6 3 2

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

##          1        2        3        4
## 2 2.150651
## 3 3.513242 4.146567
## 4 3.878726 4.246051 3.388339
## 5 3.018500 3.737739 5.124420 6.043691

image(k51)
points(pharma_norm, col=clusters_index, pch=19, cex=0.3)
```

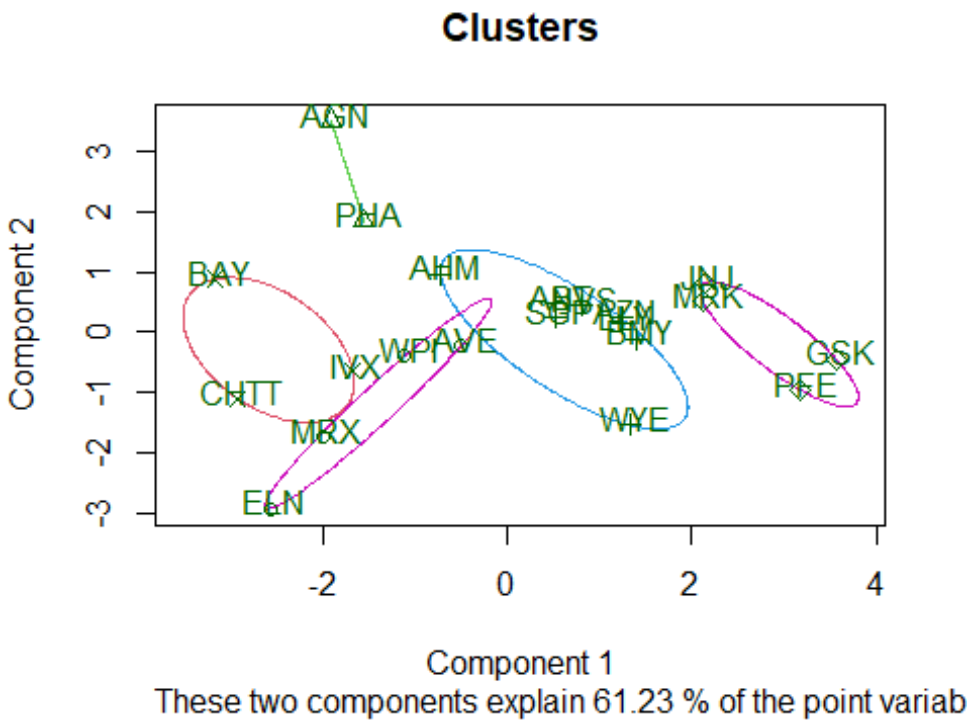


q2

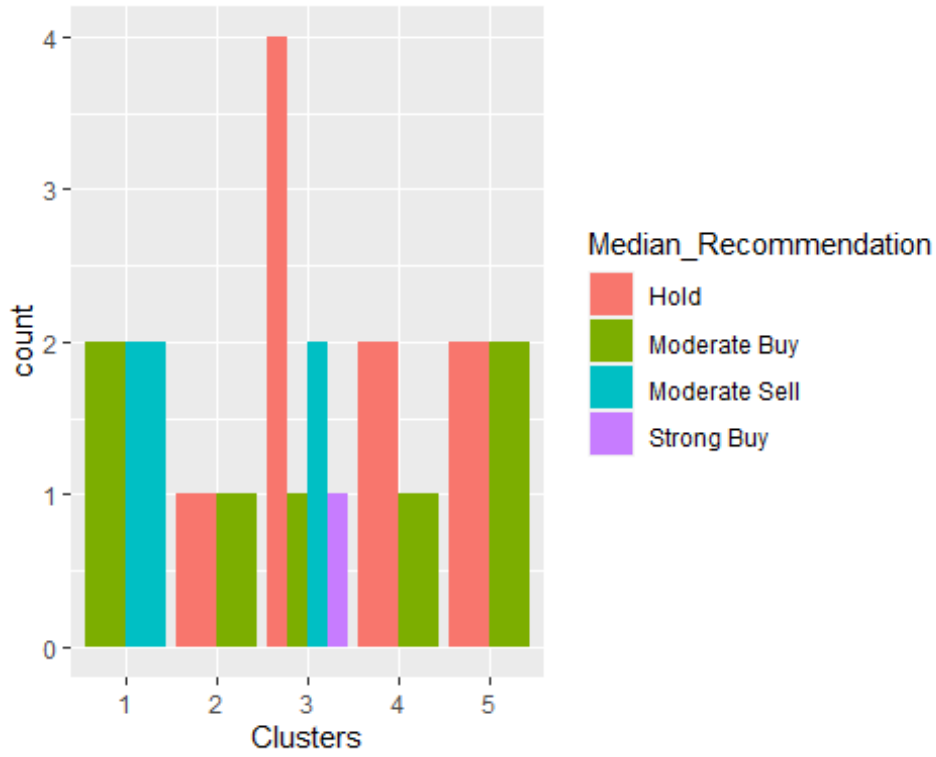
```
pharma_filter %>% mutate(Cluster = k5$cluster) %>% group_by(Cluster) %>%
summarise_all("mean")
```

```
## # 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      13.1  0.598     17.7  14.6  6.2           0.425    0.635
## 2       2      31.9  0.405     69.5  13.2  5.6           0.75     0.475
## 3       3      55.8  0.414     20.3  28.7 12.7           0.738    0.371
## 4       4       6.64 0.87      24.6  16.5  4.17          0.6      1.65
## 5       5     157.   0.48      22.2  44.4 17.7           0.95     0.22
## # ℹ 2 more variables: Rev_Growth <dbl>, Net_Profit_Margin <dbl>
```

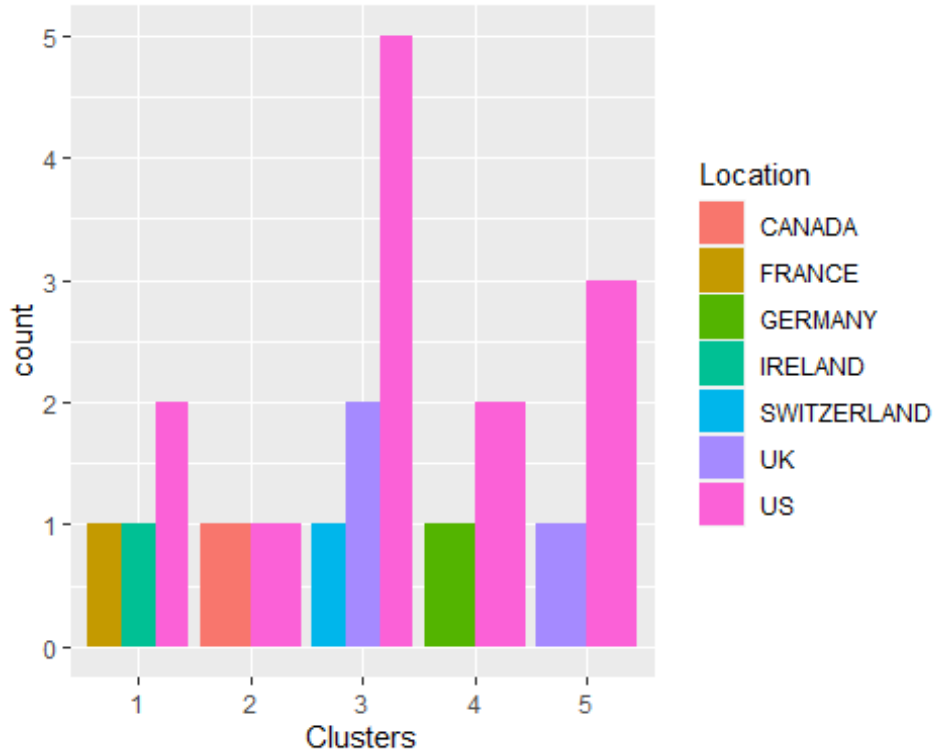
```
clusplot(pharma_norm,k5$cluster, main="Clusters",color = TRUE, labels =
3,lines = 0)
```



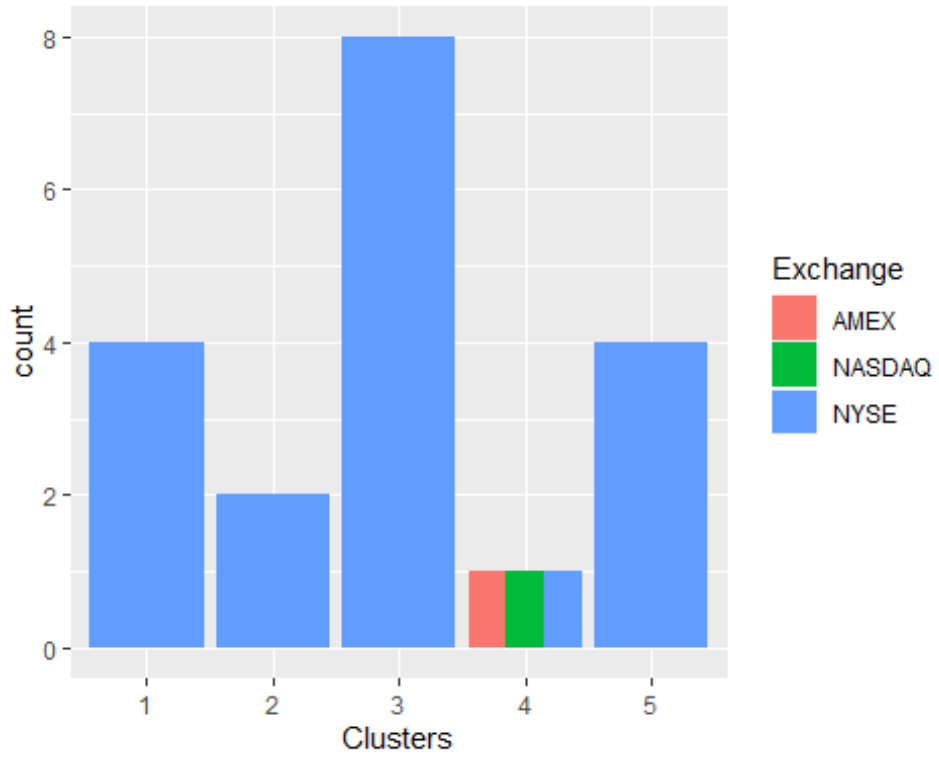

```
clust_pharma = pharma[12:14] %>% mutate(Clusters=k5$cluster)
ggplot(clust_pharma, mapping = aes(factor(Clusters),
fill =Median_Recommendation))+geom_bar(position='dodge')+labs(x ='Clusters')
```




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



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



Interpretation of Clusters based on Variables 10 to 12:

Cluster 1:

Median Recommendation: Cluster 1 has a moderate buy and moderate sell recommendation.

Location: Cluster 1 has three locations, the most prominent of which is the United States.

Exchange: Cluster 1 only has one exchange, which is the NYSE.

Cluster 2:

Median Recommendation: Cluster 2 has a low hold and a low buy, according to the median recommendation.

Location: Cluster 2 has only two locations (the United States and Canada) that are evenly distributed.

Exchange: Cluster 2 only has one exchange, which is the NYSE.

Cluster 3:

Median Recommendation: Cluster 3 is a very strong hold.

Location: Cluster 3 has three locations, with the United States outnumbering the United Kingdom and Switzerland.

Exchange: Cluster 3 has only one exchange, the NYSE, which has a large number of participants.

Cluster 4:

Median Recommendation: Cluster 4 has a strong hold rating and a low buy rating.

Location: Cluster 4 has two locations where the US ranks higher than Germany.

Exchange: Cluster 4 is home to three exchanges (AMEX, NASDAQ, and NYSE), which are all evenly distributed.

Cluster 5:

Median Recommendation: Cluster 5 has a high hold and a high buy, according to the median recommendation.

Location: Cluster 5 has two locations, with the US outnumbering the UK by a large margin.

Exchange: Cluster 5 only has one exchange, which is NYSE.

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

Appropriate Name:

Cluster 1 - Growth Pioneers

Cluster 2 - Speculative Stars

Cluster 3 - Market Titans

Cluster 4 - Beta Bulls

Cluster 5 - Blue Chip Triumph

Interpretation of Clusters based on Cluster Variables:

Cluster 1: AVE, WPI, MRX, and ELN have moderate Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Leverage, Rev_Growth, and Net_Profit_Margin values. Recognized for its exceptional Net Profit Margin, the lowest PE ratio, and rapid sales growth, Cluster 1 emerges as an optimal choice for strategic investment or holding as a reserve.

Cluster 2: PHA and AGN have lower Market_Cap, Beta, and PE_Ratio values.Highlighted by a notably high PE ratio, Cluster 2 issues a cautionary signal regarding potential overvaluation. Investors are advised to approach this cluster with careful consideration, acknowledging the elevated valuation.

Cluster 3: WYE, BMY, AZN, SGP, AHM, LLY,NVS, and ABT have higher Market_Cap, Beta, PE_Ratio, Rev_Growth, and Net_Profit_Margin values than the other clusters. Cluster 3

epitomizes a moderate-risk category. While not as extreme as other clusters, entities in this group demand thoughtful consideration, striking a balance between risk and potential return.

Cluster 4: IVX, CHTT, and BAY have lower Market_Cap and PE_Ratio values.Despite showcasing an excellent PE ratio, Cluster 4 carries significant risk due to elevated leverage, poor Net Profit Margin, and very low revenue growth. Ownership of entities in this cluster is regarded as highly precarious.

Cluster 5: GSK, PFE, MRK, and JNJ have higher Market_Cap, Beta, PE_Ratio, ROE, ROA, Asset_Turnover, Rev_Growth, and Net_Profit_Margin values.cluster 5 unfolds with formidable market capitalization, ROI, ROA, asset turnover, and Net Profit Margin. Entities in this cluster, featuring a moderately valued PE ratio, are highly favorable for both purchase and retention. The substantial revenue growth of 18.5% further amplifies the appeal of this cluster.