```r
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 4.3.3
```

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

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

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

```r
library(tidyverse)
```

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

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

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
```

```
## -- 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
library(ggplot2)
library(readr)

# Load the dataset
pharmaceuticals_data <- read.csv("C:\\Users\\pakan\\Desktop\\FML\\Pharmaceuticals.csv")
p_data<-na.omit(pharmaceuticals_data)
summary(p_data)
```

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

After removing the missing values from the data, we need to focus on the quantitative variables (1-9) to group the 21 companies.

```
row.names(p_data) <- p_data[,1]
pharmaceuticals_data<- p_data[,3:11]
head(pharmaceuticals_data)
```

```
##     Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## ABT      68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## AGN       7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## AHM       6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
## AZN      67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## AVE      47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## BAY      16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
##     Net_Profit_Margin
## ABT              16.1
## AGN               5.5
## AHM              11.2
## AZN              18.0
## AVE              12.9
## BAY               2.6
```

All the numerical variables in the dataframe have been successfully scaled.

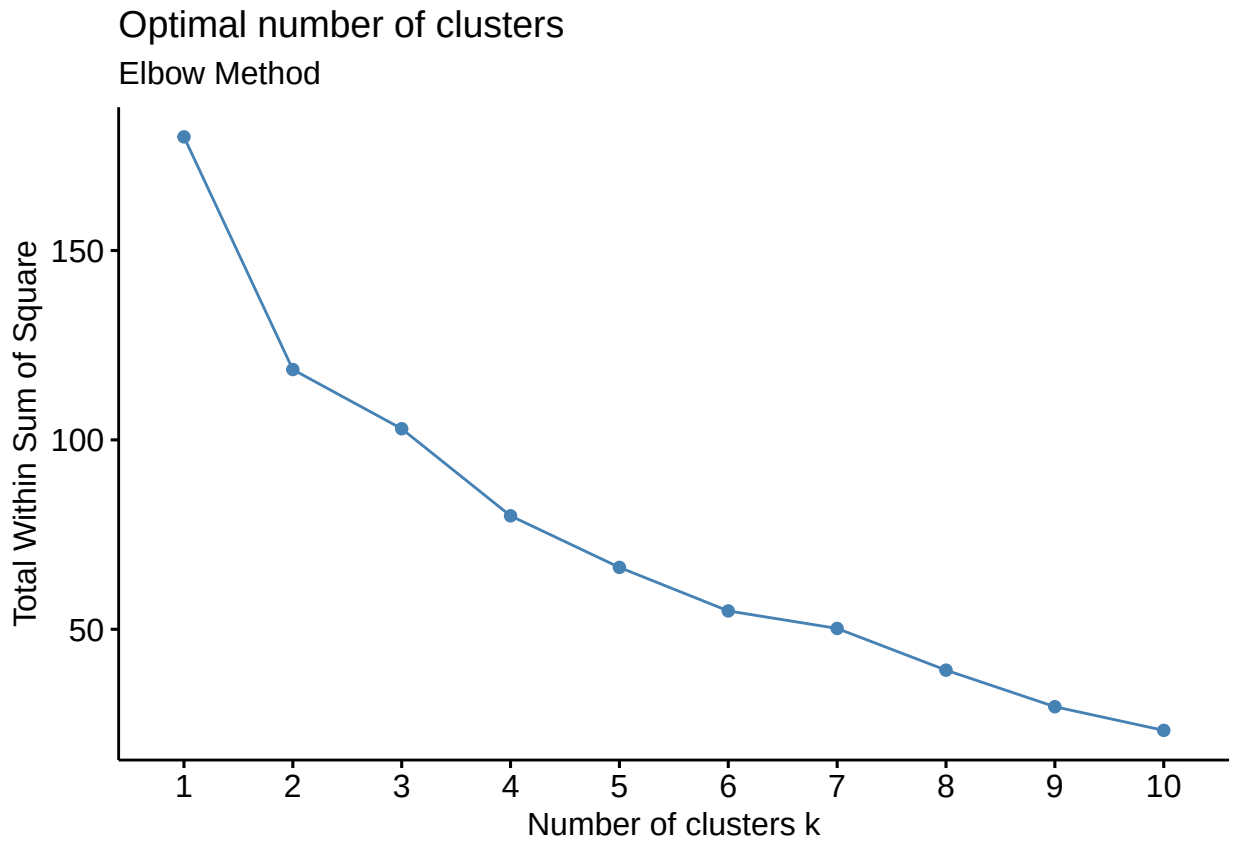
```
Pharma_data<-scale(pharmaceuticals_data)
head(Pharma_data)
```

```
##     Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## ABT  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## AGN -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## AHM -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## AZN  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## AVE -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## BAY -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
##       Leverage Rev_Growth Net_Profit_Margin
## ABT -0.2120979 -0.5277675        0.06168225
```

```
## AGN  0.0182843 -0.3811391       -1.55366706
## AHM -0.4040831 -0.5721181       -0.68503583
## AZN -0.7496565  0.1474473        0.35122600
## AVE -0.3144900  1.2163867       -0.42597037
## BAY -0.7496565 -1.4971443       -1.99560225
```

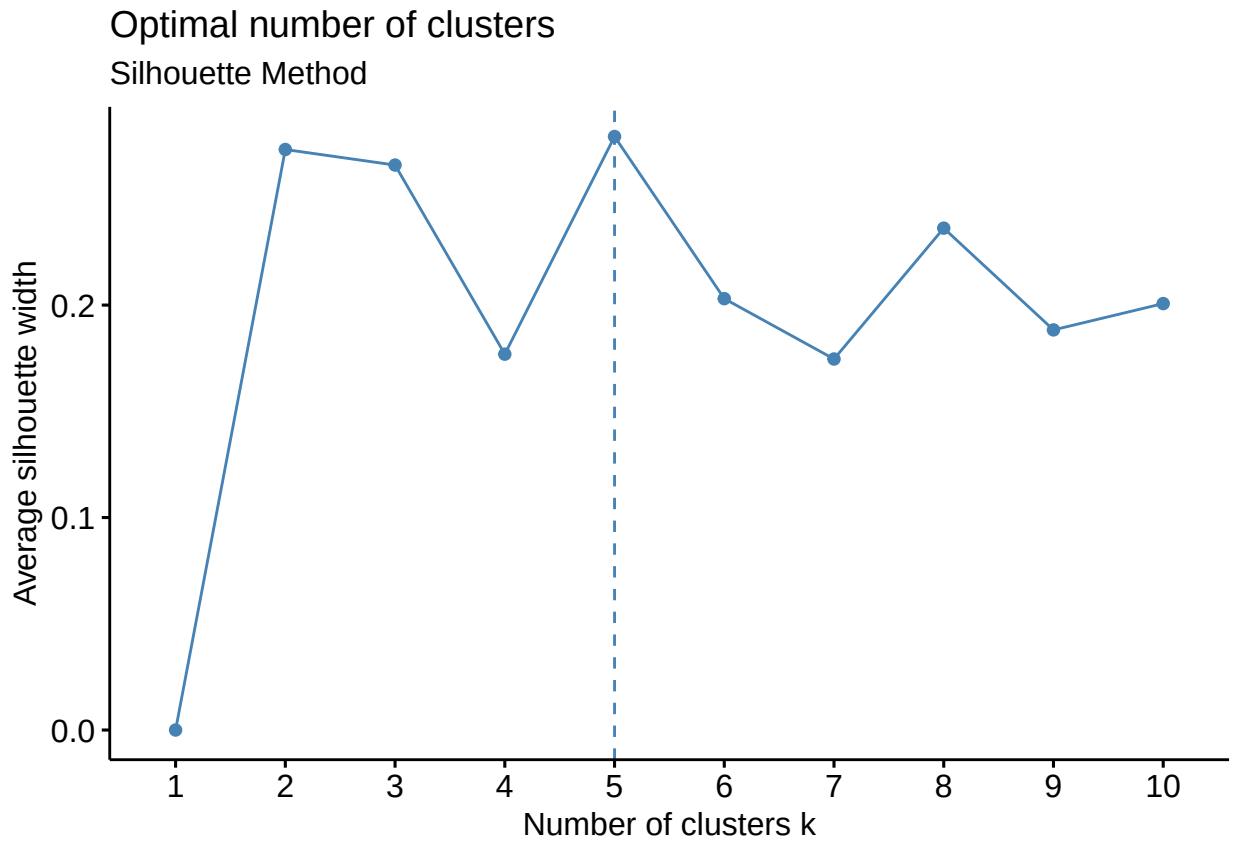
Now, let's move on to determining the number of clusters for the Elbow Method cluster analysis.

```
fviz_nbclust(Pharma_data, kmeans, method = "wss") + labs(subtitle = "Elbow Method")
```



Determining the number of clusters using the silhouette method

```
fviz_nbclust(Pharma_data, kmeans, method = "silhouette")+ labs(subtitle = "Silhouette Method")
```

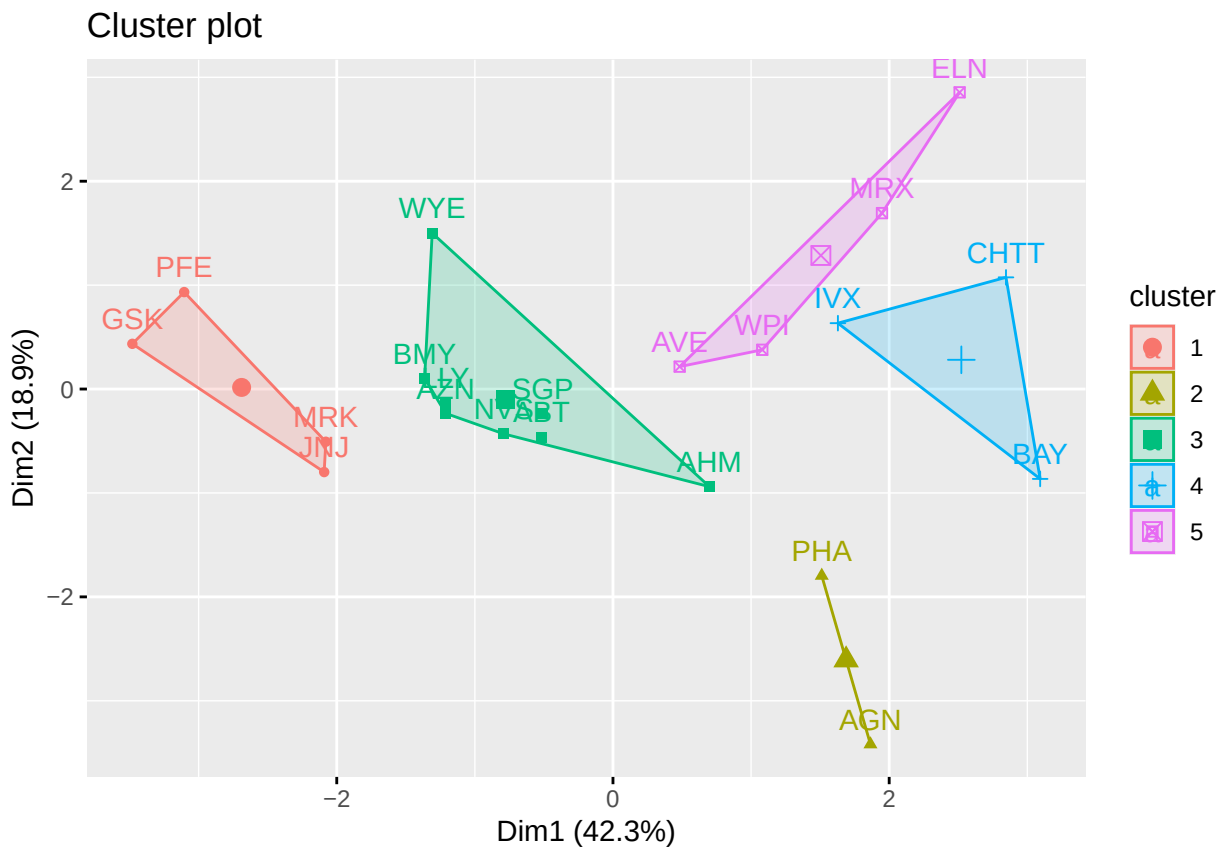


Based on the plots provided, it is clear that there are 5 clusters, which effectively showcase the variations in the data.

```
set.seed(120)
k5_clusters<- kmeans(Pharma_data,centers=5,nstart = 25)
#Visualize the output
k5_clusters$centers  #centroids
```

```
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 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 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.46807818  0.4671788       0.591242521
## 2 -0.14170336 -0.1168459      -1.416514761
## 3 -0.27449312 -0.7041516       0.556954446
## 4  1.36644699 -0.6912914      -1.320000179
## 5  0.06308085  1.5180158      -0.006893899
```

```
fviz_cluster(k5_clusters,data = Pharma_data) # to Visualize the clusters
```

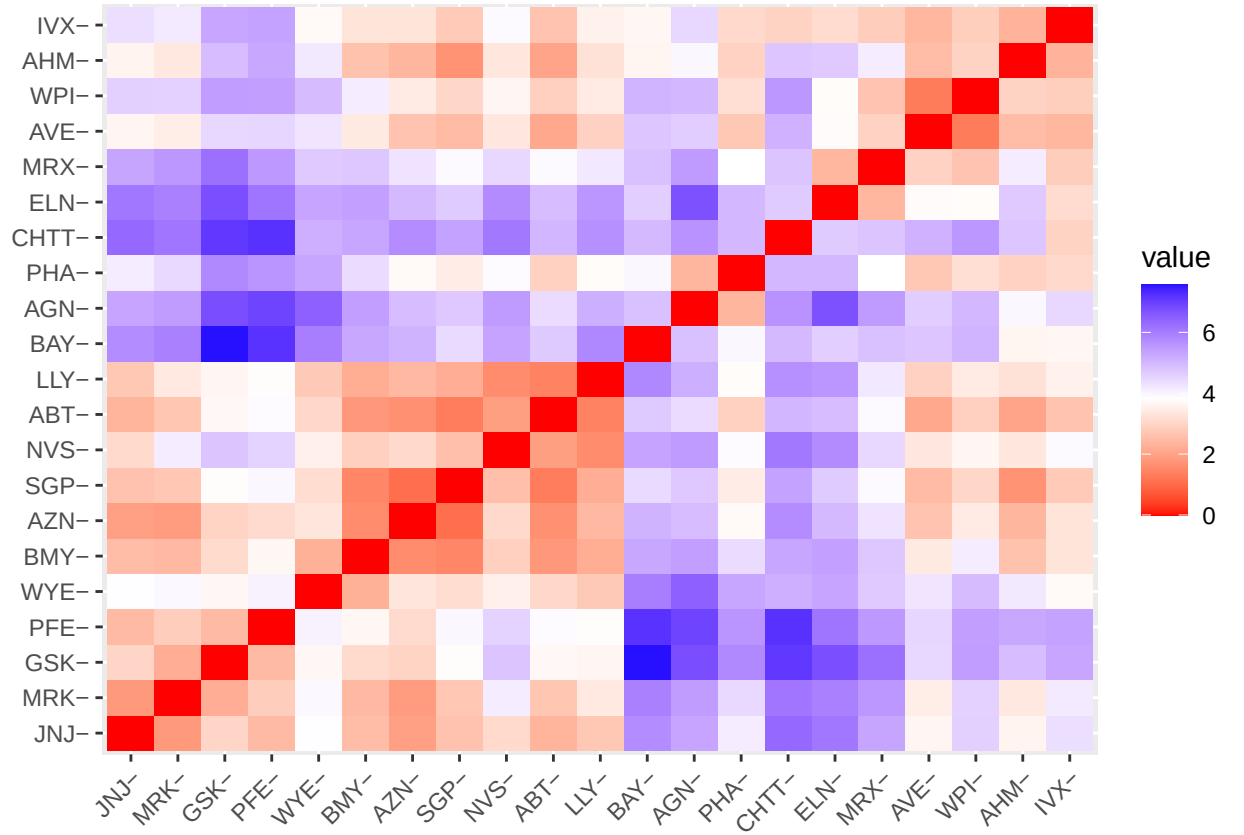



```
k5_clusters
```

```
## K-means clustering with 5 clusters of sizes 4, 2, 8, 3, 4
##
## Cluster means:
##    Market_Cap       Beta    PE_Ratio        ROE        ROA Asset_Turnover
## 1  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431      1.1531640
## 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 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428     -1.2684804
##      Leverage Rev_Growth Net_Profit_Margin
## 1 -0.46807818  0.4671788       0.591242521
## 2 -0.14170336 -0.1168459      -1.416514761
## 3 -0.27449312 -0.7041516       0.556954446
## 4  1.36644699 -0.6912914      -1.320000179
## 5  0.06308085  1.5180158      -0.006893899
##
## Clustering vector:
##  ABT  AGN  AHM  AZN  AVE  BAY  BMY CHTT  ELN  LLY  GSK  IVX  JNJ  MRX  MRK  NVS
##    3    2    3    3    5    4    3    4    5    3    1    4    1    5    1    3
##  PFE  PHA  SGP  WPI  WYE
##    1    2    3    5    3
##
## Within cluster sum of squares by cluster:
## [1]  9.284424  2.803505 21.879320 15.595925 12.791257
```

```
##  (between_SS / total_SS =  65.4 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

```
distance<- dist(Pharma_data, method = "euclidean")
fviz_dist(distance)
```



K-Means Cluster Analysis - here i am fitting data with 5 clusters

```
fit_5c<-kmeans(Pharma_data,5)
```

Now, we can determine the average value of all numerical variables for each cluster.

```
aggregate(Pharma_data,by=list(fit_5c$cluster),FUN=mean)
```

```
##   Group.1  Market_Cap       Beta    PE_Ratio        ROE        ROA
## 1       1 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
## 2       2  0.08926902 -0.4618336 -0.32086149  0.3260892  0.5396003
## 3       3 -0.96686975  1.5162611 -0.57398880 -0.8382671 -0.9892673
## 4       4  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431
## 5       5 -0.57238455 -0.6220844  0.86927480 -0.7381675 -0.7242993
##   Asset_Turnover   Leverage Rev_Growth Net_Profit_Margin
```

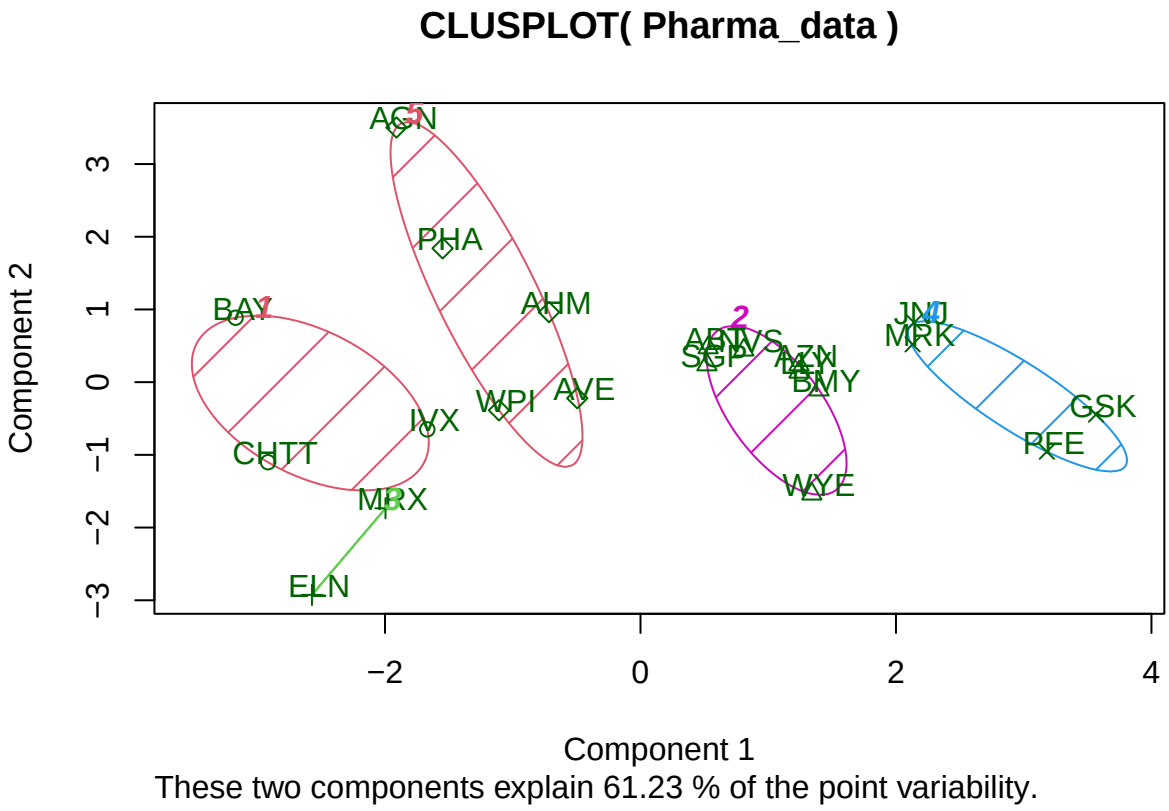
```
## 1  -4.612656e-01  1.3664470 -0.6912914        -1.3200002
## 2   6.589509e-02 -0.2559803 -0.7230135         0.7343816
## 3  -1.845062e+00  0.5302448  1.7123890         0.2445520
## 4   1.153164e+00 -0.4680782  0.4671788         0.5912425
## 5   1.776140e-16 -0.2991312  0.3682951        -0.8069490
```

```
Pharma3<-data.frame(Pharma_data,fit_5c$cluster)
Pharma3
```

```
##      Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## ABT   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## AGN  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## AHM  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## AZN   0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## AVE  -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## BAY  -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
## BMY  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## CHTT -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## ELN  -0.9704532  2.15893320 -1.34037772 -0.70899938 -1.0174553     -1.8450624
## LLY   0.2762415 -1.34655112  0.14948233  0.34502953  0.5610770     -0.4612656
## GSK   1.0999201 -0.68440408 -0.45749769  2.45971647  1.8389364      1.3837968
## IVX  -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.6979905     -0.4612656
## JNJ   1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544      0.9225312
## MRX  -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -1.8450624
## MRK   1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577      1.8450624
## NVS   0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -0.9225312
## PFE   2.4199899  0.48409069 -0.11415545  1.31287998  1.6322239      0.4612656
## PHA  -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.9047030     -0.4612656
## SGP  -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929      0.4612656
## WPI  -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905     -0.9225312
## WYE  -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849     -0.4612656
##        Leverage  Rev_Growth Net_Profit_Margin fit_5c.cluster
## ABT  -0.21209793 -0.52776752       0.06168225              2
## AGN   0.01828430 -0.38113909      -1.55366706              5
## AHM  -0.40408312 -0.57211809      -0.68503583              5
## AZN  -0.74965647  0.14744734       0.35122600              2
## AVE  -0.31449003  1.21638667      -0.42597037              5
## BAY  -0.74965647 -1.49714434      -1.99560225              1
## BMY  -0.02011273 -0.96584257       0.74744375              2
## CHTT  3.74279705 -0.63276071      -1.24888417              1
## ELN   0.61983791  1.88617085      -0.36501379              3
## LLY  -0.07130879 -0.64814764       1.17413980              2
## GSK  -0.31449003  0.76926048       0.82363947              4
## IVX   1.10620040  0.05603085      -0.71551412              1
## JNJ  -0.62166634 -0.36213170       0.33598685              4
## MRX   0.44065173  1.53860717       0.85411776              3
## MRK  -0.39128411  0.36014907      -0.24310064              4
## NVS  -0.67286239 -1.45369888       1.02174835              2
## PFE  -0.54487226  1.10143723       1.44844440              4
## PHA  -0.30169102  0.14744734      -1.27936246              5
## SGP  -0.74965647 -0.43544591       0.29026942              2
## WPI  -0.49367621  1.43089863      -0.09070919              5
## WYE   0.68383297 -1.17763919       1.49416183              2
```

To view the cluster plot

```
library(cluster)
clusplot(Pharma_data,fit_5c$cluster,color = TRUE,shade = TRUE,labels = 2,lines = 0)
```



(b): Interpret the clusters with respect to the numerical variables used in forming the clusters.

solution(b):

By examining the average values of all numerical variables for each cluster

Cluster 1 - IVX, BAY, CHTT

Cluster 2 - LLY, NVS, SGP,WYE,ABT, AZN, BMY

Cluster 3 - MRX, ELN

Cluster 4 - PFE, GSK,JNJ, MRK

Cluster 5 - AVE, PHA, WPI,AGN, AHM

Cluster 1: Cluster 1 has the highest Beta value , Leverage value and lowest Market_Cap value , ROE, ROA, Leverage, Rev_Growth, Net_Profit_Margin Cluster 2: Cluster 2 has highest Net_Profit_Margin and lowest Beta. Cluster 3: Here cluster 3 has highest Rev_Growth and lowest PE_Ratio, Asset_Turnover. Cluster 4: Cluster 4 has highest Market_Cap, ROE, ROA,Asset_Turnover Cluster 5: Cluster 5 has higher PE_Ratio.

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

Answer(c):

There appears to be a noticeable trend in the clusters when it comes to the Media recommendation variable.

Cluster 1, which has the highest Beta and highest Leverage, is mostly recommended as a Moderate Buy.

Cluster 2, which has the highest net profit margin, predominantly has hold recommendations.

Cluster 3, which has the lowest PE_Ratio and lowest Asset_Turnover, is given a Hold Recommendation.

Cluster 4 has the highest market capitalization, return on equity, return on assets, and asset turnover. It is equally recommended to hold or moderately buy.

Cluster 5, which has the highest PE_Ratio, is strongly recommended as a great investment opportunity. This is due to the fact that a high PE_Ratio suggests that the company is experiencing rapid growth.

I noticed a pattern among the clusters when it comes to variables (10 to 12).

Clusters 1 and 4 are primarily associated with a Moderate Buy Recommendation.

Clusters 2, 3, and 4 are recommended to be held.

(d): Provide an appropriate name for each cluster using any or all of the variables in the dataset.

Answer(d): Cluster1 - cluster with high Beta and leverage, or cluster to consider buying.

Cluster2 - cluster with a high net profit margin or a high hold.

Cluster3 - Cluster with low PE ratio and high asset turnover, suitable for holding.

Cluster4 - Recommended Buy cluster

Cluster5 - cluster with a high PE_Ratio (or) high Buy cluster.