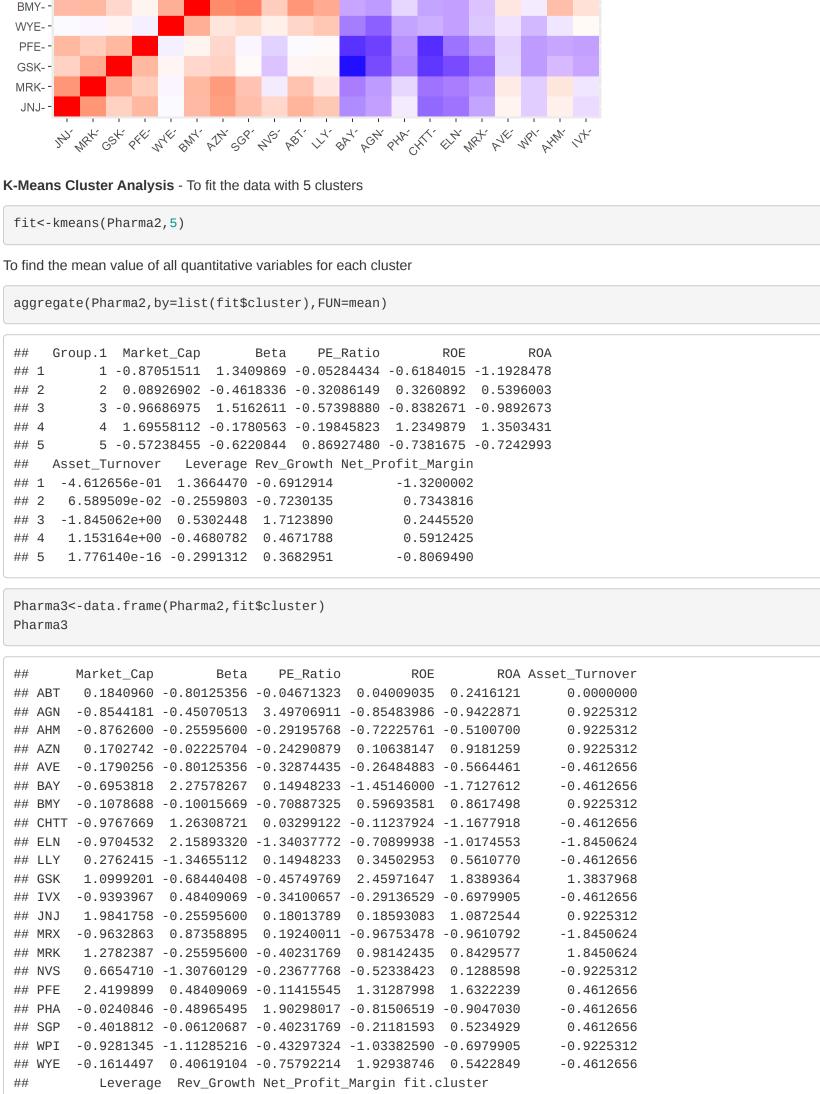
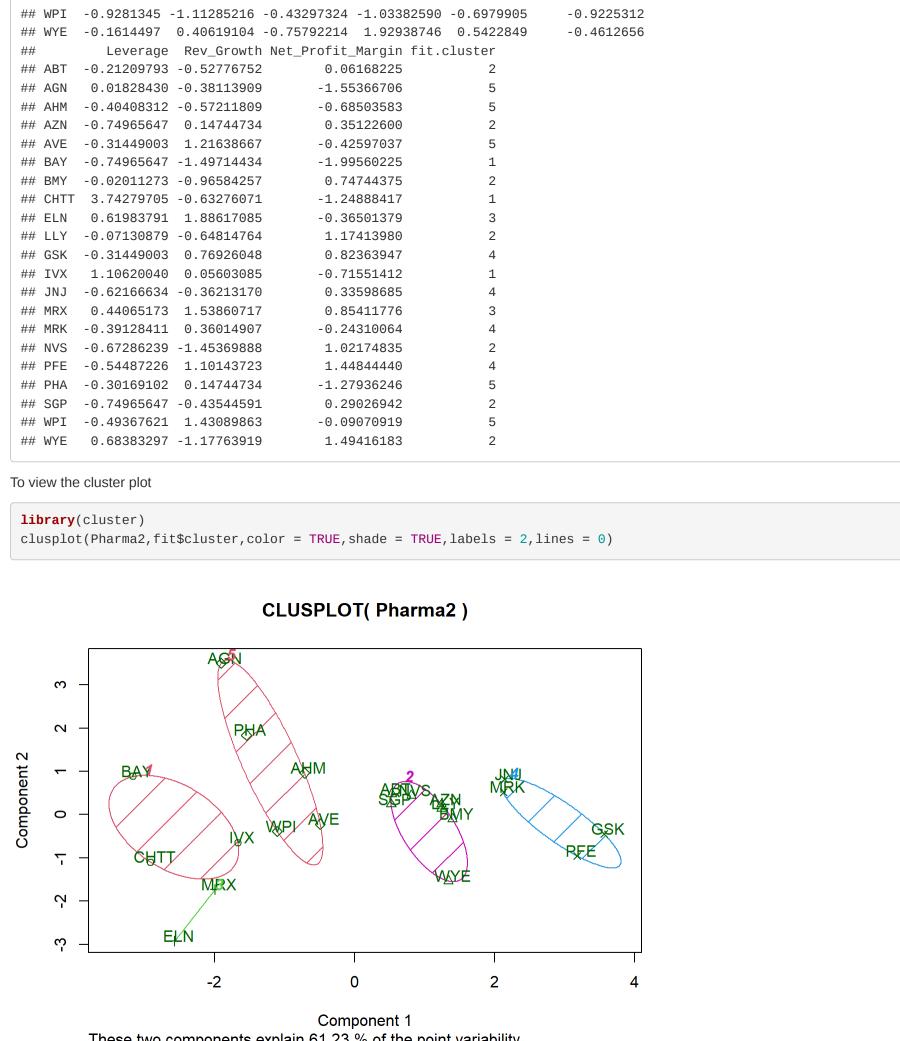
```
Assignment 4_FML_VD
VaishnaviD
2024-03-30
Problem Statement -
An equities analyst is studying the pharmaceutical industry and would like help in exploring and understanding the financial data collected by her
firm. Her main objective is to understand the structure of the pharmaceutical industry using some basic financial measures. Financial data gathered
on 21 firms in the pharmaceutical industry are available in the file Pharmaceuticals.csv. This can be done using k-Means for clustering.
To use cluster analysis to explore and analyze the given dataset
To load required library, specify the file location and showing
dimensions
 library(factoextra)
 ## Warning: package 'factoextra' was built under R version 4.3.3
 ## Loading required package: ggplot2
 ## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
 library(ggplot2)
 library(tidyverse)
 ## Warning: package 'tidyverse' was built under R version 4.3.3
                                                                 — tidyverse 2.0.0 —
 ## — Attaching core tidyverse packages —
 ## \checkmark dplyr 1.1.4 \checkmark readr 2.1.5
 ## \checkmark forcats 1.0.0 \checkmark stringr 1.5.1
 ## ✓ lubridate 1.9.3 ✓ tibble 3.2.1
 ## ✓ purrr 1.0.2 ✓ tidyr
 ## — Conflicts —
                                                           – tidyverse_conflicts() —
 ## * dplyr::filter() masks stats::filter()
 ## * dplyr::lag() masks stats::lag()
 ## i Use the conflicted package (<a href="http://conflicted.r-lib.org/">http://conflicted.r-lib.org/</a>) to force all conflicts to become errors
 library(readr)
 Pharmaceuticals <- read.csv("C:/Users/Vaishnavi/OneDrive - Kent State University/FML/Assignment 4/Pharmaceutical
 s.csv")
 summary(Pharmaceuticals)
                                           Market_Cap
        Symbol
                           Name
                                                                  Beta
                                           Min. : 0.41 Min. :0.1800
 ## Length:21
                 Length:21
    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
                                         R0A
 ##
                                                  Asset_Turnover Leverage
            : 3.60 Min.
                           : 3.9 Min.
                                                           :0.3
                                         : 1.40
     1st Qu.:18.90
                     1st Qu.:14.9
                                    1st Qu.: 5.70
                                                                    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. :62.9 Max. :20.30
                                                   Max. :1.1
                                                                    Max. :3.5100
 ##
            :82.50
       Rev_Growth
                     Net_Profit_Margin Median_Recommendation Location
          :-3.17
                     Min. : 2.6
                                       Length:21
                                                              Length:21
     1st Qu.: 6.38
                                                             Class :character
                     1st Qu.:11.2
                                       Class :character
                                    Mode :character
     Median : 9.37
                     Median :16.1
                                                             Mode :character
                     Mean :15.7
     Mean :13.37
     3rd Qu.:21.87
                     3rd Qu.:21.1
     Max. :34.21
       Exchange
     Length:21
     Class :character
     Mode :character
 ##
 ##
To confirm that all the data has been properly imported
 head(Pharmaceuticals)
                            Name Market_Cap Beta PE_Ratio ROE ROA Asset_Turnover
                                      68.44 0.32
 ## 1
         ABT Abbott Laboratories
                                                     24.7 26.4 11.8
               Allergan, Inc.
                                     7.58 0.41
                                                     82.5 12.9 5.5
 ## 2
                 Amersham plc
 ## 3
         AHM
                                     6.30 0.46
                                                     20.7 14.9 7.8
                                                                                0.9
         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
                        Bayer AG
                                      16.90 1.11
                                                     27.9 3.9 1.4
      Leverage Rev_Growth Net_Profit_Margin Median_Recommendation Location Exchange
 ## 1
          0.42
                   7.54
                                       16.1
                                                     Moderate Buy
 ## 2
          0.60
                     9.16
                                        5.5
                                                     Moderate Buy
                                                                     CANADA
                                                    Strong Buy
 ## 3
          0.27
                     7.05
                                       11.2
                                                    Moderate Sell
 ## 4
          0.00
                    15.00
                                       18.0
                                                                         UK
                                                                                NYSE
                    26.81
                                       12.9
 ## 5
          0.34
                                                     Moderate Buy
                                                                     FRANCE
                                                                                NYSE
 ## 6
          0.00
                    -3.17
                                        2.6
                                                              Hold GERMANY
                                                                                NYSE
 tail(Pharmaceuticals)
       Symbol
                                      Name Market_Cap Beta PE_Ratio ROE ROA
 ## 16
                               Novartis AG
                                                96.65 0.19
                                                                21.6 17.9 11.2
                                Pfizer Inc
                                                                23.6 45.6 19.2
 ## 17
          PFE
                                               199.47 0.65
 ## 18
                                                                56.5 13.5 5.7
          PHA
                     Pharmacia Corporation 56.24 0.40
          SGP Schering-Plough Corporation 34.10 0.51
 ## 19
                                                               18.9 22.6 13.3
          WPI Watson Pharmaceuticals, Inc.
 ## 20
                                                3.26 0.24
                                                                18.4 10.2 6.8
                                     Wyeth
                                                                13.1 54.9 13.4
 ## 21
                                                48.19 0.63
       Asset_Turnover Leverage Rev_Growth Net_Profit_Margin Median_Recommendation
 ## 16
                  0.5
                          0.06
                                  -2.69
                                                       22.4
                  0.8
                                                                     Moderate Buy
 ## 17
                          0.16
                                    25.54
                                                       25.2
                          0.35
                                    15.00
                                                       7.3
 ## 19
                  0.8
                          0.00
                                     8.56
                                                       17.6
                                                                              Hold
 ## 20
                  0.5
                          0.20
                                    29.18
                                                       15.1
                                                                     Moderate Sell
                                     0.36
                                                        25.5
 ##
          Location Exchange
 ## 16 SWITZERLAND
                       NYSE
 ## 17
                       NYSE
 ## 18
                US
                       NYSE
 ## 19
                US
                       NYSE
 ## 21
                US
                       NYSE
 dim(Pharmaceuticals)
 ## [1] 21 14
Qustion (a): 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.
Answer (a): We have to remove missing data and rescale variables for comparability before grouping the data.
 pharma <- na.omit(Pharmaceuticals) #Doing this will remove all the missing values in the data
 pharma
 ##
       Symbol
                                             Name Market_Cap Beta PE_Ratio ROE
          ABT
                             Abbott Laboratories
 ## 1
                                                       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
          \mathsf{AHM}
                                     Amersham plc
                                                        6.30 0.46
                                                                      20.7 14.9 7.8
 ## 4
                                                       67.63 0.52
                                                                      21.5 27.4 15.4
          AZN
                                 AstraZeneca PLC
 ## 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
                                                                      13.9 34.8 15.1
 ##
          BMY
                    Bristol-Myers Squibb Company
                                                       51.33 0.50
 ## 8
                                                                      26.0 24.1 4.3
         CHTT
                                     Chattem, Inc
                                                       0.41 0.85
 ## 9
                           Elan Corporation, plc
                                                        0.78 1.08
          ELN
                                                                       3.6 15.1 5.1
                                                       73.84 0.18
 ## 10
                           Eli Lilly and Company
                                                                      27.9 31.0 13.5
 ## 11
                             GlaxoSmithKline plc
                                                                      18.0 62.9 20.3
          GSK
                                                      122.11 0.35
 ## 12
                                                       2.60 0.65
          IVX
                                IVAX Corporation
                                                                      19.9 21.4 6.8
 ##
    13
          JNJ
                                Johnson & Johnson
                                                      173.93 0.46
                                                                      28.4 28.6 16.3
 ## 14
                                                                      28.6 11.2 5.4
          MRX Medicis Pharmaceutical Corporation
                                                       1.20 0.75
 ## 15
                               Merck & Co., Inc.
                                                      132.56 0.46
                                                                      18.9 40.6 15.0
                                      Novartis AG
 ## 16
          NVS
                                                       96.65 0.19
                                                                      21.6 17.9 11.2
 ## 17
          PFE
                                      Pfizer Inc
                                                      199.47 0.65
                                                                      23.6 45.6 19.2
                                                                      56.5 13.5 5.7
 ## 18
                           Pharmacia Corporation
          PHA
                                                      56.24 0.40
 ## 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
                                           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
                                     9.16
                                                         5.5
                                                                      Moderate Buy
 ## 3
                  0.9
                          0.27
                                     7.05
                                                        11.2
                                                                        Strong Buy
                  0.9
 ## 4
                          0.00
                                    15.00
                                                        18.0
                                                                     Moderate Sell
 ## 5
                  0.6
                          0.34
                                     26.81
                                                        12.9
                                                                      Moderate Buy
 ## 6
                                                                              Hold
                  0.6
                          0.00
                                     -3.17
                                                         2.6
                  0.9
                                     2.70
                                                        20.6
 ## 7
                          0.57
                                                                     Moderate Sell
 ## 8
                  0.6
                                     6.38
                                                        7.5
                                                                      Moderate Buy
                          3.51
                  0.3
                                     34.21
                                                        13.3
                                                                     Moderate Sell
 ## 9
                          1.07
                  0.6
                                                                              Hold
 ## 10
                          0.53
                                     6.21
                                                        23.4
 ## 11
                  1.0
                          0.34
                                     21.87
                                                        21.1
                                                                              Hold
 ## 12
                  0.6
                          1.45
                                     13.99
                                                        11.0
                                                                              Hold
 ## 13
                  0.9
                                     9.37
                                                        17.9
                                                                      Moderate Buy
                  0.3
                                                                      Moderate Buy
 ## 14
                          0.93
                                     30.37
                                                        21.3
 ## 15
                  1.1
                          0.28
                                    17.35
                                                        14.1
                                                                              Hold
 ## 16
                  0.5
                          0.06
                                     -2.69
                                                        22.4
                                                                              Hold
                  0.8
 ## 17
                          0.16
                                    25.54
                                                        25.2
                                                                      Moderate Buy
                  0.6
                                                                              Hold
 ## 18
                          0.35
                                    15.00
                                                        7.3
 ## 19
                  0.8
                          0.00
                                     8.56
                                                        17.6
                                                                              Hold
 ## 20
                  0.5
                                                        15.1
                                                                     Moderate Sell
                          0.20
                                     29.18
                  0.6
                                     0.36
                                                        25.5
                                                                              Hold
 ##
    21
                          1.12
 ##
          Location Exchange
                       NYSE
 ## 1
                US
 ## 2
            CANADA
                       NYSE
 ## 3
                UK
                       NYSE
                       NYSE
 ## 4
                UK
 ## 5
            FRANCE
                       NYSE
                       NYSE
           GERMANY
 ## 6
 ## 7
                US
                       NYSE
                     NASDAQ
 ## 9
           IRELAND
                       NYSE
                       NYSE
 ## 10
                US
 ## 11
                UK
                       NYSE
                US
 ## 12
                       AMEX
                US
                       NYSE
 ## 13
                US
                       NYSE
 ## 14
 ## 15
                US
                       NYSE
 ## 16 SWITZERLAND
                       NYSE
                       NYSE
 ## 17
                US
 ## 18
                US
                       NYSE
 ## 19
                US
                       NYSE
                       NYSE
 ## 20
                US
 ## 21
                US
                       NYSE
Having now eliminated the missing values from the data, we will gather only the quantitative variables (i.e., 1-9) in order to group the 21
companies.
 row.names(pharma) <- pharma[,1]</pre>
 Pharma1<- pharma[,3:11]</pre>
 head(Pharma1)
        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
 ## AGN
              7.58 0.41
                            82.5 12.9 5.5
                                                      0.9
                                                               0.60
                                                                          9.16
              6.30 0.46
                            20.7 14.9 7.8
                                                      0.9
                                                               0.27
                                                                          7.05
 ## AHM
             67.63 0.52
                            21.5 27.4 15.4
                                                      0.9
                                                              0.00
                                                                         15.00
 ## AZN
 ## 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
 ## AGN
                     5.5
 ## AHM
                     11.2
 ## AZN
                     18.0
                     12.9
 ## AVE
 ## BAY
                      2.6
All of the quantitative variables in the dataframe are now scaled.
 Pharma2<-scale(Pharma1)
 head(Pharma2)
        Market_Cap
                          Beta
                                  PE_Ratio
                                                               ROA Asset_Turnover
 ## ABT 0.1840960 -0.80125356 -0.04671323 0.04009035 0.2416121
                                                                        0.0000000
                                                                        0.9225312
 ## AGN -0.8544181 -0.45070513 3.49706911 -0.85483986 -0.9422871
 ## 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
                                                                       -0.4612656
 ## AVE -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461
 ## 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
To calculate the number of clusters for the Elbow Method cluster analysis.
 fviz_nbclust(Pharma2, kmeans, method = "wss") + labs(subtitle = "Elbow Method")
        Optimal number of clusters
        Elbow Method
 Total Within Sum of Square
    50
                   2
                                          5
            1
                           3
                                                  6
                                                                 8
                                                                         9
                                                                                10
                                     Number of clusters k
To calculate the number of clusters using the silhouette method
 fviz_nbclust(Pharma2, kmeans, method = "silhouette")+ labs(subtitle = "Silhouette Method")
       Optimal number of clusters
       Silhouette Method
Average silhouette width
   0.0
                   2
                           3
                                                  6
                                                         7
                                                                 8
                                                                         9
                                                                                10
                                  4
                                          5
                                     Number of clusters k
It can be seen from the plots above that there are 5 clusters which is sufficient to display the changes in the data.
 set.seed(120)
 k5<- kmeans(Pharma2, centers=5, nstart = 25)</pre>
 #Visualize the output
 k5$centers #centroids
                                PE_Ratio
                                                R0E
       Market_Cap
                        Beta
                                                            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,data = Pharma2) # to Visualize the clusters
     Cluster plot
   2 -
                           WYE
                                                                    CHTT
                                                                               cluster
 Dim2 (18.9%)
                                                        PHA
   -2 -
                                   Dim1 (42.3%)
 k5
 ## K-means clustering with 5 clusters of sizes 4, 2, 8, 3, 4
 ## Cluster means:
       Market_Cap
                        Beta
                                PE_Ratio
                                                            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
                               -1.416514761
 ## 2 -0.14170336 -0.1168459
 ## 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
         PHA
               SGP WPI WYE
 ##
```







These two components explain 61.23 % of the point variability. Question (b): Interpret the clusters with respect to the numerical variables used in forming the clusters. Answer(b): While analysis we can observe that the mean values of all quantitative variables for each cluster Cluster 1 - BAY, CHTT, IVX Cluster 2 - ABT, AZN, BMY, LLY, NVS, SGP, WYE Cluster 3 - ELN, MRX Cluster 4 - JNJ, MRK, PFE, GSK Cluster 5 - AGN, AHM, AVE, PHA, WPI Cluster 1 has highest Beta, Leverage and lowest Market\_Cap, ROE, ROA, Leverage, Rev\_Growth, Net\_Profit\_Margin Cluster 2 has highest Net\_Profit\_Margin and lowest Beta. Cluster 3 has highest Rev\_Growth and lowest PE\_Ratio, Asset\_Turnover. Cluster 4 has highest Market\_Cap, ROE, ROA, Asset\_Turnover Cluster 5 has highest PE\_Ratio. Question(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 is a pattern in the clusters with respect to Media recommendation variable. Cluster 1 - highest Beta, highest Leverage has mostly Moderate Buy Recommendation. Cluster 2 - highest Net\_Profit\_Margin has mostly Hold Recommendation Cluster 3 - lowest PE\_Ratio and lowest Asset\_Turnover has Hold Recommendation Cluster 4 - highest Market\_Cap, highest ROE, highest ROA, highest Asset\_Turnover has equal Hold and Moderate Buy Recommendation Cluster 5 - highest PE\_Ratio has the Strong Buy Recommendation, because high PE\_Ratio indicates the company is growing fast. We can observe a pattern among the clusters with respect to variables(10 to 12) Clusters 1,4 has mostly Moderate Buy Recommendation and Clusters 2,3,4 has Hold Recommendation Question(d): Provide an appropriate name for each cluster using any or all of the variables in the dataset. **Answer(d):** Cluster1 - high Beta, Leverage cluster or Buy Cluster. Cluster2 - high Net\_Profit\_Margin cluster or high hold cluster.

Cluster3 - Low PE\_Ratio, Asset\_Turnover cluster or hold cluster.

Cluster5 - high PE\_Ratio cluster or high Buy cluster.

Cluster4 - Moderate Buy cluster