Assignment4

Anujeeth Veerla

04/11/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
library(factoextra)
```

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

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

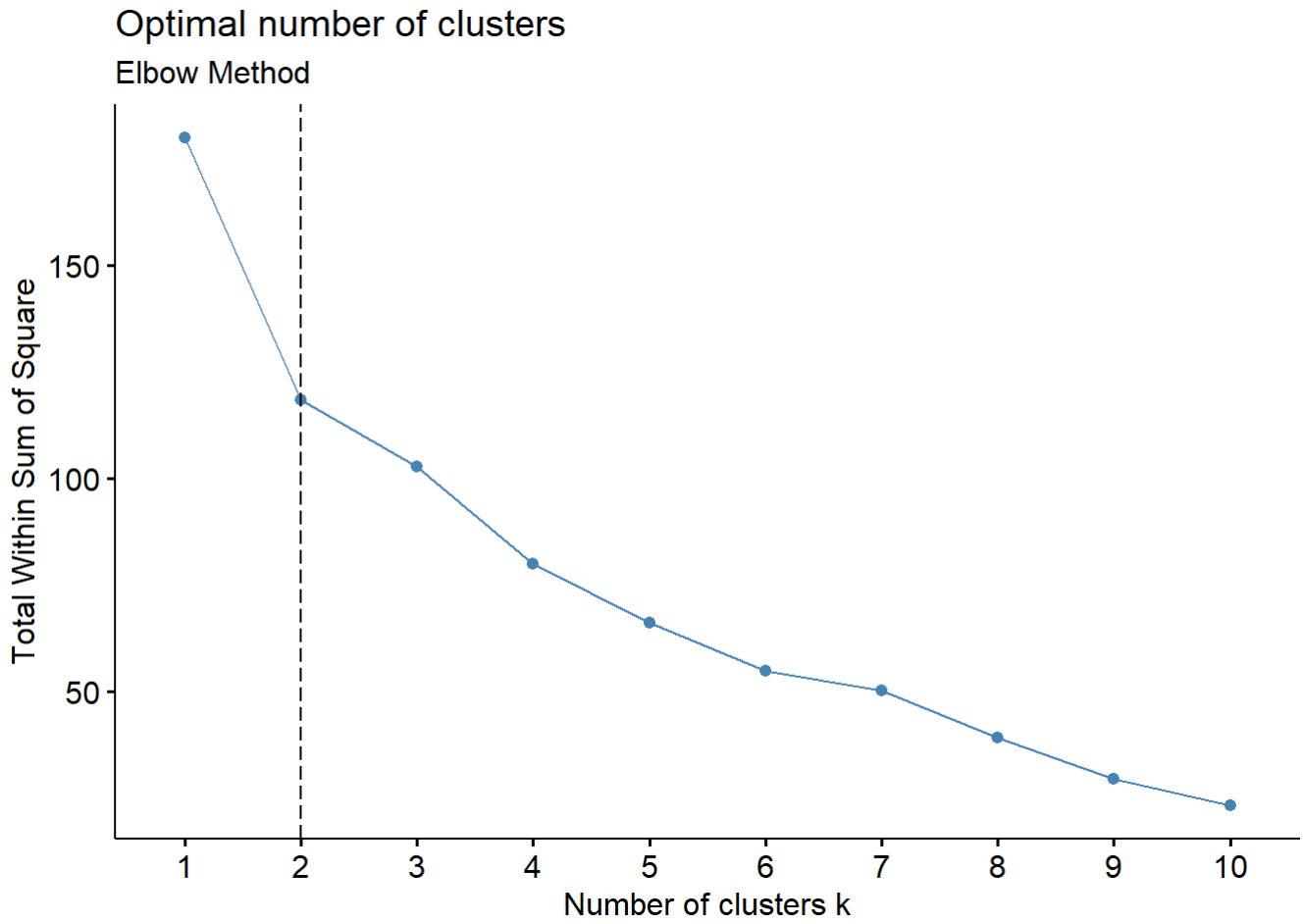
```
#Reading CSV file into Pharma variable
Pharma <- read.csv("Pharmaceuticals.csv")
head(Pharma)
```

```
##   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
```

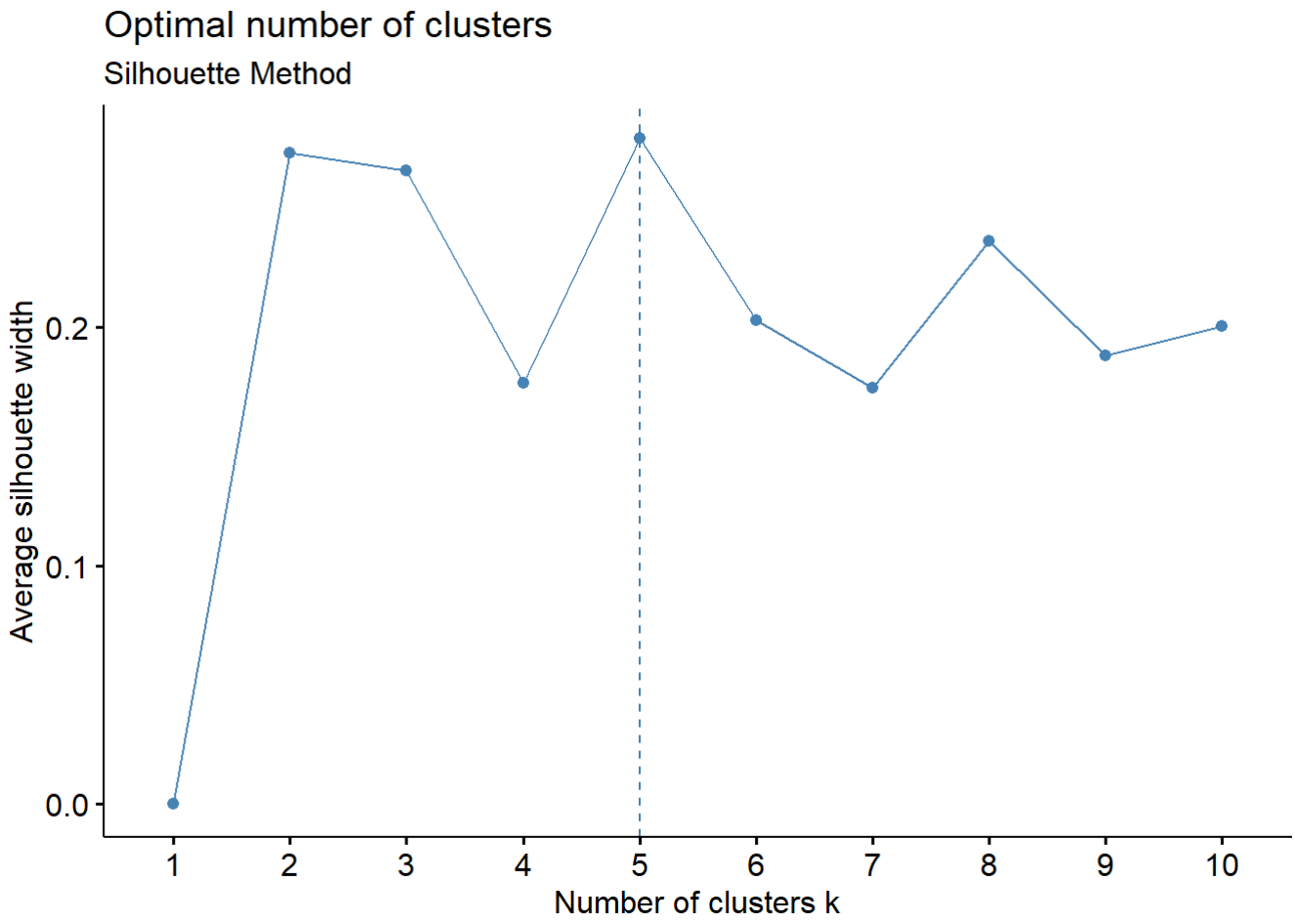
```
#Taking the quantitative variables (1-9) to cluster 21 firms
Pharma1<-Pharma[,3:11]
head(Pharma1)
```

```
##   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
```

```
#here, scaling quantitative variables
Pharma2<-scale(Pharma1)
#we use two methods to calculate K value
#Calculating K value using Elbow method
fviz_nbclust(Pharma2, kmeans, method = "wss") + geom_vline(xintercept = 2, linetype = 5) + la
bs(subtitle = "Elbow Method")
```



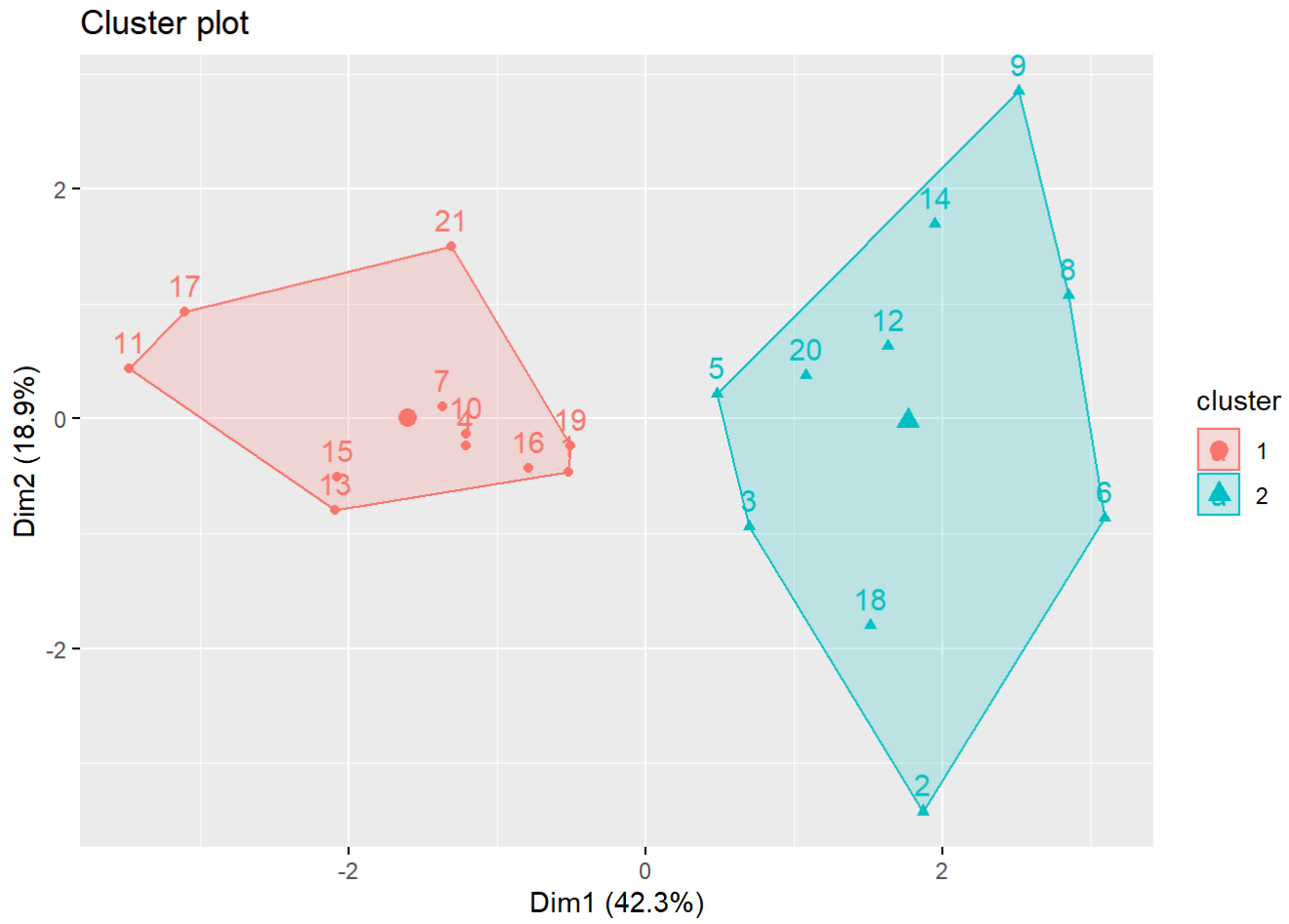

```
#Calculating K value using Silhoutte method
fviz_nbclust(Pharma2, kmeans, method = "silhouette") + labs(subtitle = "Silhouette Method")
```




```
set.seed(1)
k5<-kmeans(Pharma2, centers = 2, nstart = 25)
k5$centers
```

```
##   Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
## 2 -0.7407208  0.3945061  0.3039863 -0.7222576 -0.9178575     -0.5073922
##     Leverage Rev_Growth Net_Profit_Margin
## 1 -0.3331068 -0.2902163         0.6823310
## 2  0.3664175  0.3192379        -0.7505641
```

```
#Visualizing clusters
fviz_cluster(k5, data=Pharma2)
```




```
fit<-kmeans(Pharma2,2)
#Getting mean values of quantitative variables
aggregate(Pharma2, by=list(fit$cluster), FUN=mean)
```

```
##   Group.1 Market_Cap       Beta   PE_Ratio        ROE        ROA Asset_Turnover
## 1       1 -0.7407208  0.3945061  0.3039863 -0.7222576 -0.9178575     -0.5073922
## 2       2  0.6733825 -0.3586419 -0.2763512  0.6565978  0.8344159      0.4612656
##     Leverage Rev_Growth Net_Profit_Margin
## 1  0.3664175  0.3192379        -0.7505641
## 2 -0.3331068 -0.2902163         0.6823310
```

```
Pharma3<-data.frame(Pharma2, fit$cluster)
Pharma3
```

```
##    Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## 1   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## 2  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## 3  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## 4   0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## 5  -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## 6  -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
## 7  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## 8  -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## 9  -0.9704532  2.15893320 -1.34037772 -0.70899938 -1.0174553     -1.8450624
## 10  0.2762415 -1.34655112  0.14948233  0.34502953  0.5610770     -0.4612656
## 11  1.0999201 -0.68440408 -0.45749769  2.45971647  1.8389364      1.3837968
## 12 -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.6979905     -0.4612656
## 13  1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544      0.9225312
## 14 -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -1.8450624
## 15  1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577      1.8450624
## 16  0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -0.9225312
## 17  2.4199899  0.48409069 -0.11415545  1.31287998  1.6322239      0.4612656
## 18 -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.9047030     -0.4612656
## 19 -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929      0.4612656
## 20 -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905     -0.9225312
## 21 -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849     -0.4612656
##       Leverage  Rev_Growth Net_Profit_Margin fit.cluster
## 1  -0.21209793 -0.52776752        0.06168225           2
## 2   0.01828430 -0.38113909       -1.55366706           1
## 3  -0.40408312 -0.57211809       -0.68503583           1
## 4  -0.74965647  0.14744734        0.35122600           2
## 5  -0.31449003  1.21638667       -0.42597037           1
## 6  -0.74965647 -1.49714434       -1.99560225           1
## 7  -0.02011273 -0.96584257        0.74744375           2
## 8   3.74279705 -0.63276071       -1.24888417           1
## 9   0.61983791  1.88617085       -0.36501379           1
## 10 -0.07130879 -0.64814764        1.17413980           2
## 11 -0.31449003  0.76926048        0.82363947           2
## 12  1.10620040  0.05603085       -0.71551412           1
## 13 -0.62166634 -0.36213170        0.33598685           2
## 14  0.44065173  1.53860717        0.85411776           1
## 15 -0.39128411  0.36014907       -0.24310064           2
## 16 -0.67286239 -1.45369888        1.02174835           2
## 17 -0.54487226  1.10143723        1.44844440           2
## 18 -0.30169102  0.14744734       -1.27936246           1
## 19 -0.74965647 -0.43544591        0.29026942           2
## 20 -0.49367621  1.43089863       -0.09070919           1
## 21  0.68383297 -1.17763919        1.49416183           2
```

```
library(cluster)
#Visualizing data
clusplot(Pharma2, fit$cluster, color= TRUE, shades=TRUE, labels=2, lines=0)
```

```
## Warning in plot.window(...): "shades" is not a graphical parameter
```

```
## Warning in plot.xy(xy, type, ...): "shades" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "shades" is not a
## graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "shades" is not a
## graphical parameter
```

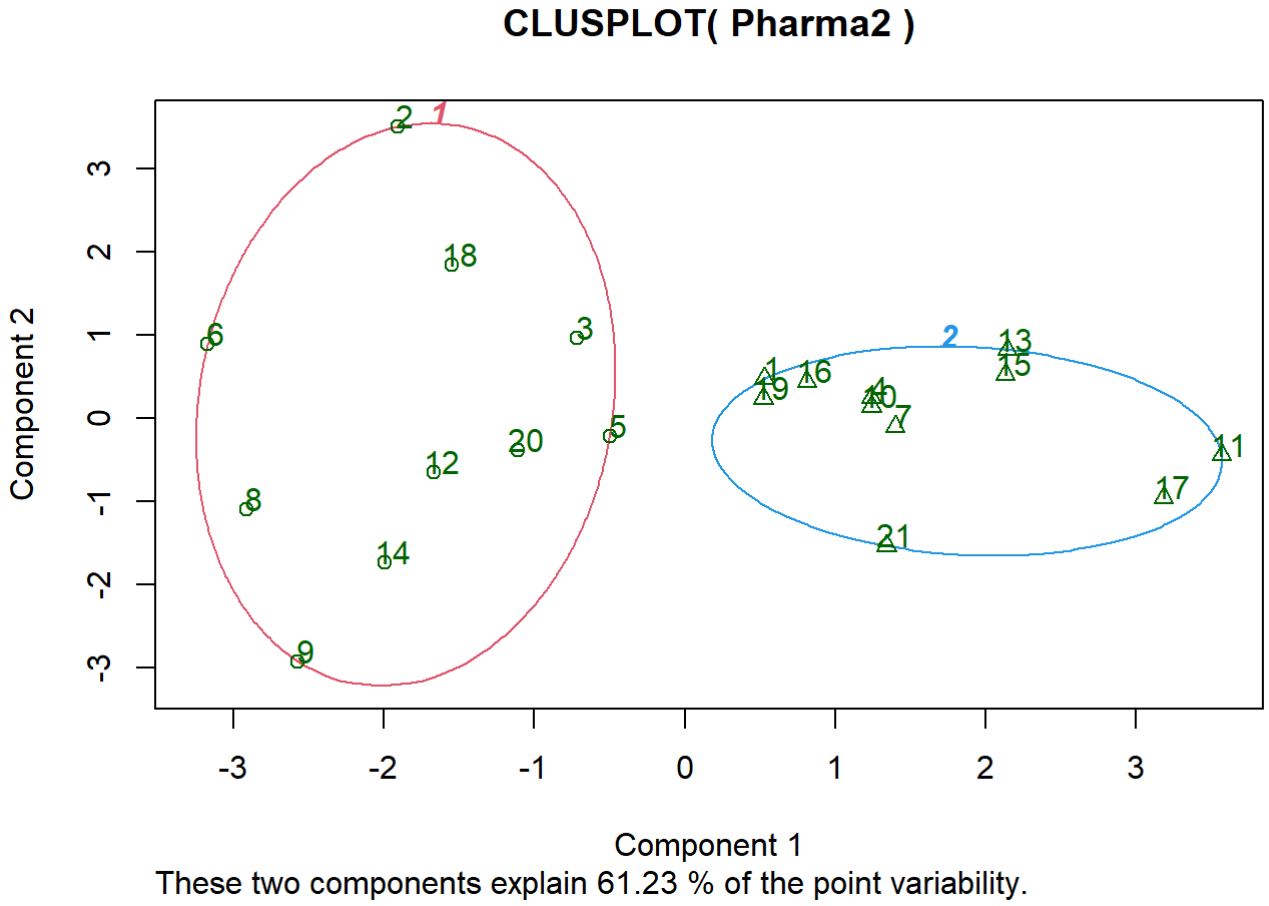
```
## Warning in box(...): "shades" is not a graphical parameter
```

```
## Warning in title(...): "shades" is not a graphical parameter
```

```
## Warning in polygon(z[[k]], density = if (shade) density[k] else 0, col =
## col.clus[jInd[i]], : "shades" is not a graphical parameter

## Warning in polygon(z[[k]], density = if (shade) density[k] else 0, col =
## col.clus[jInd[i]], : "shades" is not a graphical parameter
```

```
## Warning in plot.xy(xy.coords(x, y), type = type, ...): "shades" is not a
## graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "shades" is not a
## graphical parameter
```

```
## Warning in text.default(xy, labels = labs, ...): "shades" is not a graphical
## parameter

## Warning in text.default(xy, labels = labs, ...): "shades" is not a graphical
## parameter
```




```
#b) Interpret the clusters with respect to the numerical variables used in forming the cluste
rs.
#Cluster 1 - 2, 3, 5,6,8, 9,12,14,18,20
#Cluster 2 -  1, 4,7,10,11,13,15,16,17,19, 21
#Observing mean values of numerical variables
#Cluster 1 has low Market_Cap, high Beta, high PE_Ratio, low ROE, low ROA, low Asset_Turnove
r, high leverage, high Rev_Growth low Net_PRofit_Margin
#Cluster 2 has high Market_Cap, low Beta, low PE_Ratio, high ROE, high ROA,high Asset_Turnove
r, low leverage, low Rev_Growth, high Net_Profit_Margin



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

#Cluster 1 has mostly Moderate Buy recommendation
#Cluster 2 has HOld recommendation

# d) Provide an appropriate name for each cluster using any or all of the variables in the da
taset.

#Cluster 1 has low capped companies
#Cluster 2 has high market capped companies
```