Assignment\_4

Yannick Lokpe

11/9/2021

# Importing and viewing dataset

p <- read.csv("Pharmaceuticals.csv",header = TRUE)  
  
head(p)

## 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

# A

# Use only numerical variables (1 to 9 ) to cluster the 21 firms

p1<- p[,3:11]  
head(p1)

## 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

# Scaling all the quantitative variables in the dataframe

ps<- scale(p1)  
head(ps)

## 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  
## Leverage Rev\_Growth Net\_Profit\_Margin  
## [1,] -0.2120979 -0.5277675 0.06168225  
## [2,] 0.0182843 -0.3811391 -1.55366706  
## [3,] -0.4040831 -0.5721181 -0.68503583  
## [4,] -0.7496565 0.1474473 0.35122600  
## [5,] -0.3144900 1.2163867 -0.42597037  
## [6,] -0.7496565 -1.4971443 -1.99560225

# Determing number of clusters for the cluster analysis

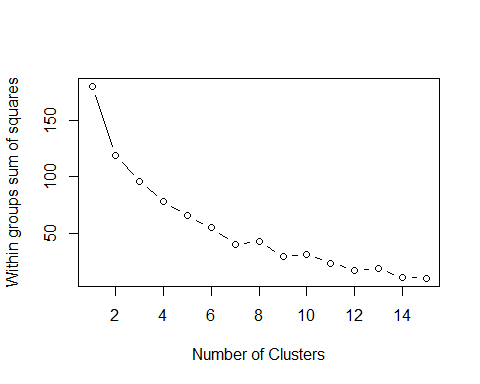
wss <-(nrow(ps)-1)\*sum(apply(ps,2,var))  
wss

## [1] 180

for (i in 2:15)wss[i]<-sum(kmeans(ps,centers=i)$withinss)  
wss

## [1] 180.00000 118.56934 95.99420 78.24600 65.61035 54.84345 40.53936  
## [8] 42.50679 29.55677 31.42706 23.60399 17.51296 18.85666 11.14352  
## [15] 10.25863

plot(1:15,wss,type="b",xlab="Number of Clusters",ylab="Within groups sum of squares")

 # In this plot we can notice, there are 5 clusters sufficient to capture most variations in the data

# K-Means Cluster Analysis

# Fiting the data with 5 clusters

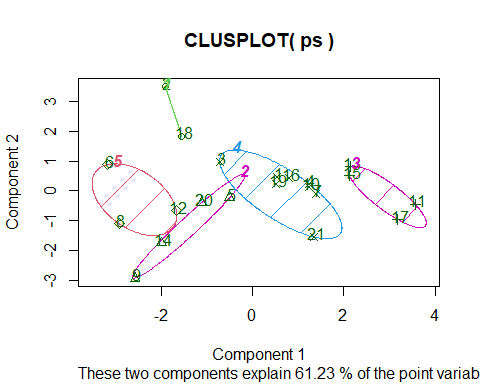
fit <- kmeans(ps,5)  
aggregate(ps,by=list(fit$cluster),FUN=mean)

## Group.1 Market\_Cap Beta PE\_Ratio ROE ROA  
## 1 1 -0.43925134 -0.4701800 2.70002464 -0.8349525 -0.9234951  
## 2 2 -0.76022489 0.2796041 -0.47742380 -0.7438022 -0.8107428  
## 3 3 1.69558112 -0.1780563 -0.19845823 1.2349879 1.3503431  
## 4 4 -0.03142211 -0.4360989 -0.31724852 0.1950459 0.4083915  
## 5 5 -0.87051511 1.3409869 -0.05284434 -0.6184015 -1.1928478  
## Asset\_Turnover Leverage Rev\_Growth Net\_Profit\_Margin  
## 1 0.2306328 -0.14170336 -0.1168459 -1.416514761  
## 2 -1.2684804 0.06308085 1.5180158 -0.006893899  
## 3 1.1531640 -0.46807818 0.4671788 0.591242521  
## 4 0.1729746 -0.27449312 -0.7041516 0.556954446  
## 5 -0.4612656 1.36644699 -0.6912914 -1.320000179

ps1<-data.frame(ps,fit$cluster)  
head(ps1)

## 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  
## Leverage Rev\_Growth Net\_Profit\_Margin fit.cluster  
## 1 -0.2120979 -0.5277675 0.06168225 4  
## 2 0.0182843 -0.3811391 -1.55366706 1  
## 3 -0.4040831 -0.5721181 -0.68503583 4  
## 4 -0.7496565 0.1474473 0.35122600 4  
## 5 -0.3144900 1.2163867 -0.42597037 2  
## 6 -0.7496565 -1.4971443 -1.99560225 5

library(cluster)  
clusplot(ps,fit$cluster,color=TRUE,shade=TRUE,labels=2,lines=0)



# B

# Interpreting the clusters with respect to the numerical variables used in forming the clusters

Cluster 1 Row 1,3,4,7,10,16,19,21 Cluster 2 Row 5,9,14,20 Cluster 3 Row 2,18 Cluster 4 Row 6,8,12 Cluster 5 Row 11,13,15,17

Using the output aggregate(ps,by=list(fit$cluster),FUN=mean), we can conclude that Cluster 1 has the lowest Beta, lowest leverage, and the highest Net\_Profit\_Margin Cluster 2 has the lowest PE\_Ratio, lowest Asset\_Turnover, and the highest Rev\_growth Cluster 3 has the highest PE\_Ratio, and a low Net\_Profit\_Margin Cluster 4 has the lowest Market\_Cap,highest Beta,lowest ROE,lowest ROA,highest leverage,lowest Rev\_growth, and the lowest Net\_Profit\_margin Cluster 5 has the highest Market\_cap,highest ROE,highest ROA,highest Asset\_Turnover,and a high Net\_Profit\_Margin

# C

Is there any pattern in the clusters with respect to the numerical variables ( 10 to 12)?

With the respect to the numerical variable (10 to 12), there is a pattern in the clusters as Median\_Recommendation. Cluster 1 with highest Net\_Profit\_Margin, lowest Beta, and lowest Leverage mostly have hold as a Median\_Recommendation Cluster 2 with the highest Rev\_growth and lowest Asset\_turnover and lowest PE\_Ratio has a moderate in sell and buy as a Median\_Recommendation Cluster 3 with the highest PE\_Ratio and low Net\_Profit\_Margin has a moderate buy and hold as Median\_Recommendation Cluster 4 has mostly hold as Median\_Recommendation Cluster 5 has mostly hold and moderate buy as Median\_Recommendation

# D

Naming each cluster using any or all of the variables in the data

Cluster 1 low Beta,Leverage, high Net\_Profit\_margin cluster Cluster 2 High Rev\_growth, Low PE\_Ratio, Asset\_Turnover Cluster 3 High PE\_Ratio, Low Net\_Profit\_Margin Cluster 4 high Beta,Leverage, Low Market\_cap,ROE,ROA,Rev\_growth,Net\_Profit\_Margin Cluster 5 high Market\_Cap,ROE,ROA,Asset\_Turnover,Net\_Profit\_Margin