Assignment\_4

#load packages

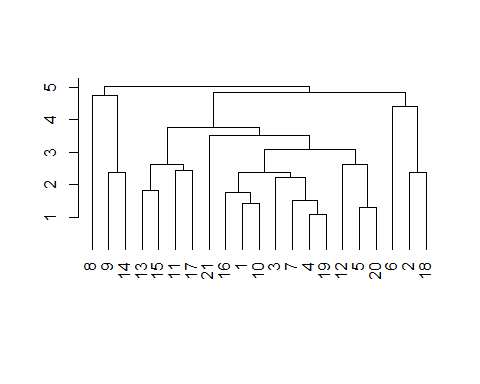
library(stats)

#load dataset

pharm.data<-read.csv("Pharmaceuticals.csv")

#Part a.

pharm.df<-pharm.data[,c(3:11)] #use only numerical variables to cluster the 21 firms.  
  
pharm.df.norm <- scale(pharm.df) #normalize the data so that the data is on a common scale.  
  
distance <- dist(pharm.df.norm, method="euclidean")  
hc<-hclust(distance, method="average", members=NULL)  
#measure distance using euclidean distance metric. I decided to use hierarchical clustering because the dataset is small. Also, it is easy to understand.  
  
plot(hc, hang=-1, ann=FALSE) # plot a dendrogram. The number of clusters formed appears to be 3. I determined this because it appears that there are 3 separations in the dendrogram.



#Part b.

hc3 <- cutree(hc, k=3)  
hcagg3 = aggregate(hc3, pharm.df, FUN=mean)  
hcagg3

## Market\_Cap Beta PE\_Ratio ROE ROA Asset\_Turnover Leverage Rev\_Growth  
## 1 16.90 1.11 27.9 3.9 1.4 0.6 0.00 -3.17  
## 2 7.58 0.41 82.5 12.9 5.5 0.9 0.60 9.16  
## 3 56.24 0.40 56.5 13.5 5.7 0.6 0.35 15.00  
## 4 0.41 0.85 26.0 24.1 4.3 0.6 3.51 6.38  
## 5 2.60 0.65 19.9 21.4 6.8 0.6 1.45 13.99  
## 6 6.30 0.46 20.7 14.9 7.8 0.9 0.27 7.05  
## 7 47.16 0.32 20.1 21.8 7.5 0.6 0.34 26.81  
## 8 0.78 1.08 3.6 15.1 5.1 0.3 1.07 34.21  
## 9 132.56 0.46 18.9 40.6 15.0 1.1 0.28 17.35  
## 10 3.26 0.24 18.4 10.2 6.8 0.5 0.20 29.18  
## 11 68.44 0.32 24.7 26.4 11.8 0.7 0.42 7.54  
## 12 34.10 0.51 18.9 22.6 13.3 0.8 0.00 8.56  
## 13 173.93 0.46 28.4 28.6 16.3 0.9 0.10 9.37  
## 14 67.63 0.52 21.5 27.4 15.4 0.9 0.00 15.00  
## 15 51.33 0.50 13.9 34.8 15.1 0.9 0.57 2.70  
## 16 122.11 0.35 18.0 62.9 20.3 1.0 0.34 21.87  
## 17 1.20 0.75 28.6 11.2 5.4 0.3 0.93 30.37  
## 18 96.65 0.19 21.6 17.9 11.2 0.5 0.06 -2.69  
## 19 73.84 0.18 27.9 31.0 13.5 0.6 0.53 6.21  
## 20 199.47 0.65 23.6 45.6 19.2 0.8 0.16 25.54  
## 21 48.19 0.63 13.1 54.9 13.4 0.6 1.12 0.36  
## Net\_Profit\_Margin x  
## 1 2.6 2  
## 2 5.5 2  
## 3 7.3 2  
## 4 7.5 3  
## 5 11.0 1  
## 6 11.2 1  
## 7 12.9 1  
## 8 13.3 3  
## 9 14.1 1  
## 10 15.1 1  
## 11 16.1 1  
## 12 17.6 1  
## 13 17.9 1  
## 14 18.0 1  
## 15 20.6 1  
## 16 21.1 1  
## 17 21.3 3  
## 18 22.4 1  
## 19 23.4 1  
## 20 25.2 1  
## 21 25.5 1

#Cluster 1 appears to consist of many firms with large market capitalization and high net profit margins. Also, all the firms are less risky than the market, with Betas of less than 1.  
  
#Cluster 2 firms have lower net profit margins, but relatively high P/E ratios.   
  
#Cluster 3 firms have small market capitalization, but mostly large revenue growth. They also have firms with higher market risk and leverage.

#Part c.

Med.Rec.Table = table(pharm.data$Median\_Recommendation, hc3)  
colnames(Med.Rec.Table)=c("cl1","cl2","cl3")  
Med.Rec.Table

## hc3  
## cl1 cl2 cl3  
## Hold 7 2 0  
## Moderate Buy 4 1 2  
## Moderate Sell 3 0 1  
## Strong Buy 1 0 0

Location.Table = table(pharm.data$Location, hc3)  
colnames(Location.Table)=c("cl1","cl2","cl3")  
Location.Table

## hc3  
## cl1 cl2 cl3  
## CANADA 0 1 0  
## FRANCE 1 0 0  
## GERMANY 0 1 0  
## IRELAND 0 0 1  
## SWITZERLAND 1 0 0  
## UK 3 0 0  
## US 10 1 2

Exchange.Table = table(pharm.data$Exchange, hc3)  
colnames(Exchange.Table)=c("cl1","cl2","cl3")  
Exchange.Table

## hc3  
## cl1 cl2 cl3  
## AMEX 1 0 0  
## NASDAQ 0 0 1  
## NYSE 14 3 2

#Cluster 1 appears to be the largest cluster, which is made up of 15 firms. The majority of firms in Cluster 1 have a hold recommendation, are located in the US, and traded on NYSE.  
  
#Cluster 2 consists of 3 firms. The majority have a hold recommendation. All are located in different countries and traded on NYSE.  
  
#Cluster 3 consists of 3 firms. The majority have a moderate buy recommendation, location in the US, and traded on NYSE.

#Part d

#Cluster 1 Name: Large Market Cap Cluster

#Cluster 2 Name: High P/E Ratio Cluster

#Cluster 3 Name: High Revenue Growth Cluster