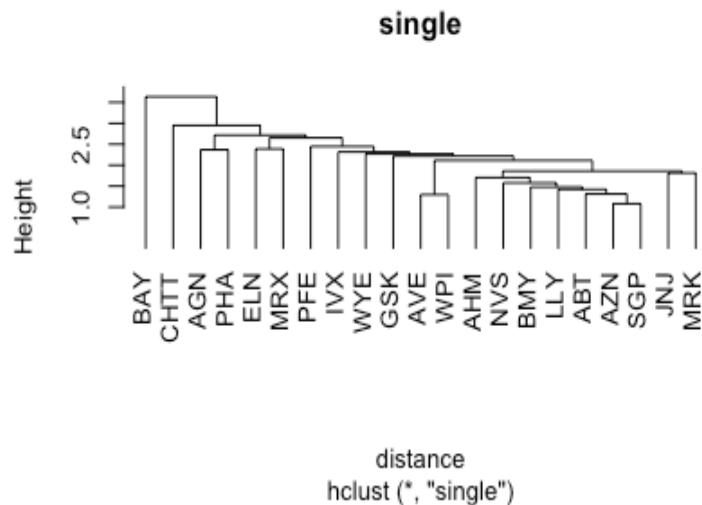
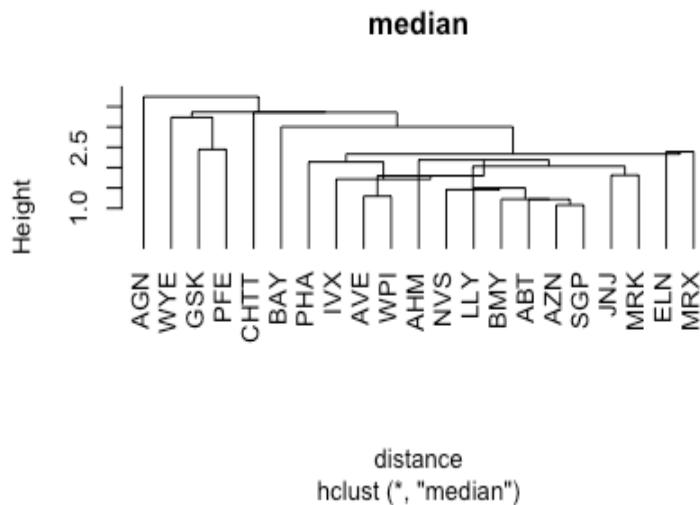
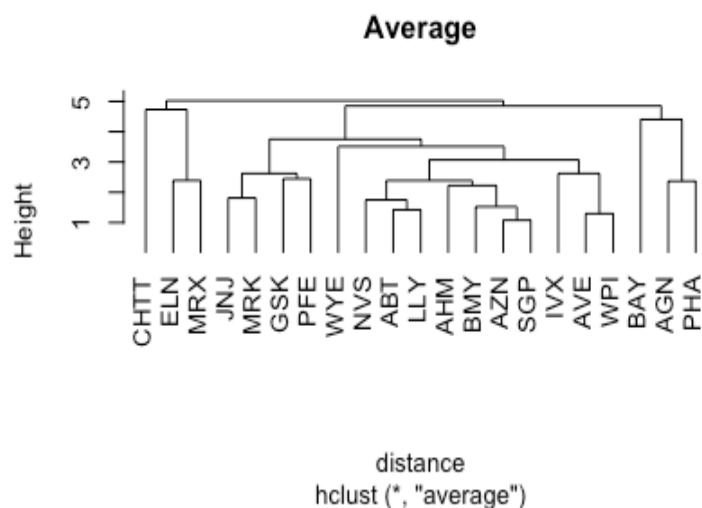
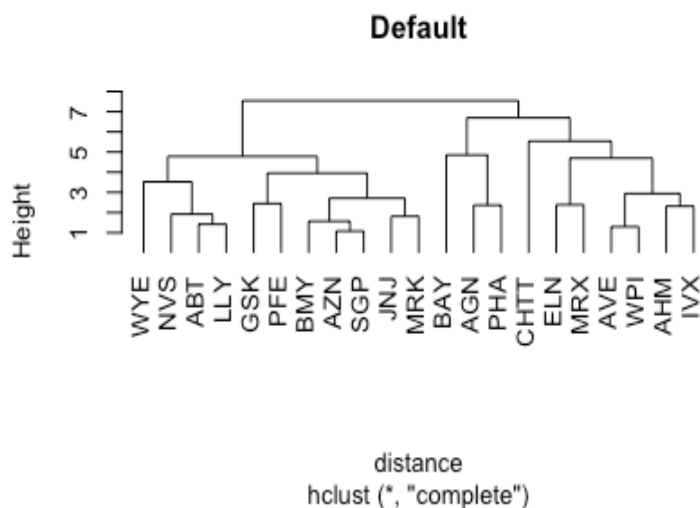


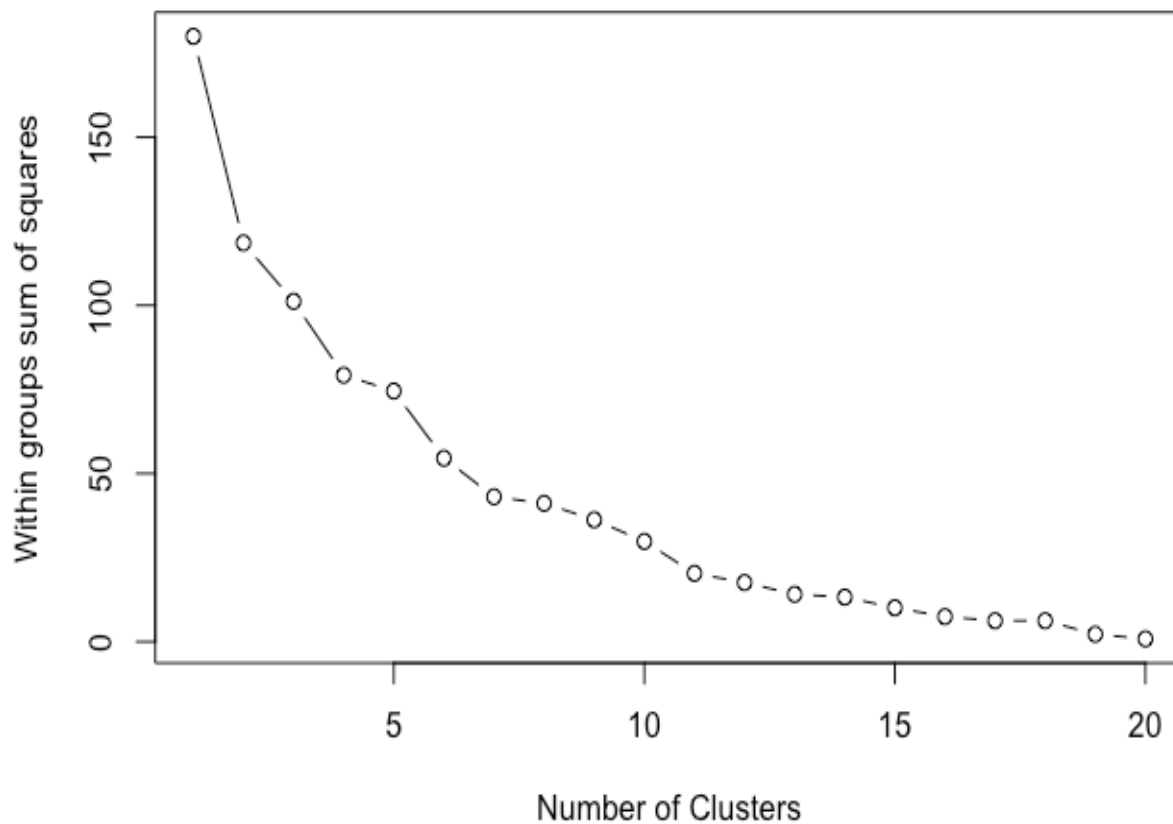
Cluster Analysis by Amber.C

Use only the quantitative variables (a)-(i) to cluster the 21 firms. rms. Decide about appropriate number of clusters.

By developing and observing hierarchical agglomerative clusters using four different methods including default, average median and single linkage, I think default method is the most appropriate linkage method.

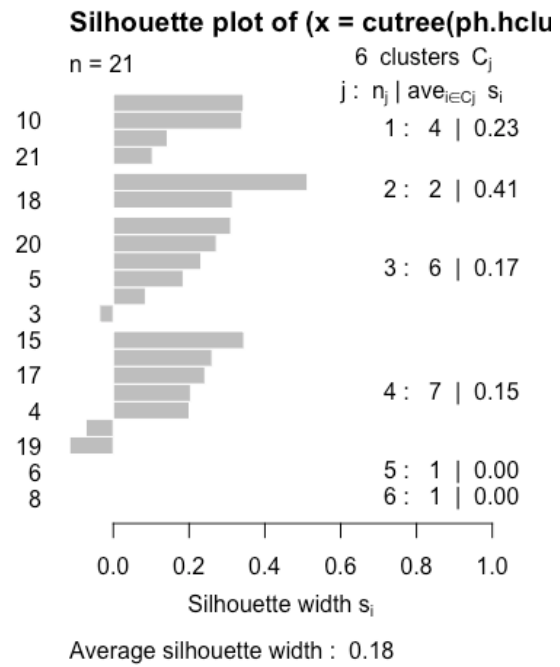
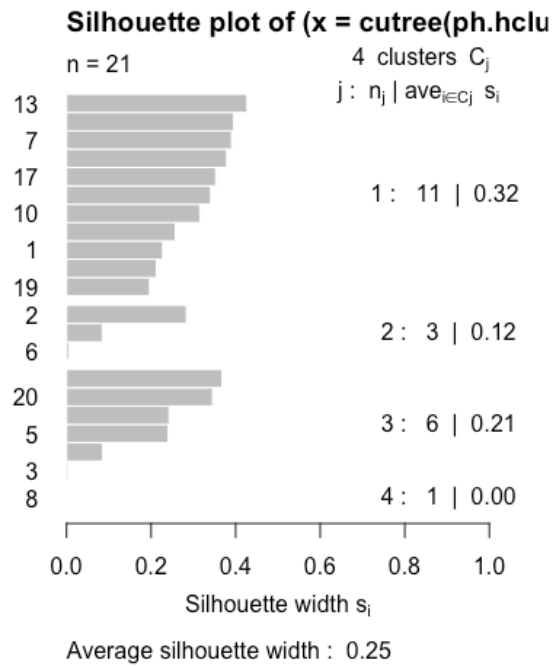
Based on default linkage (maximum distance) method, we developed scree plot for checking within cluster variability. As is shown below, within groups the sum of squares experienced continuous sharp drop from forming 1 to 4 clusters, an insignificant drop from 4 to 5 and a relatively large drop from 5 to 6.



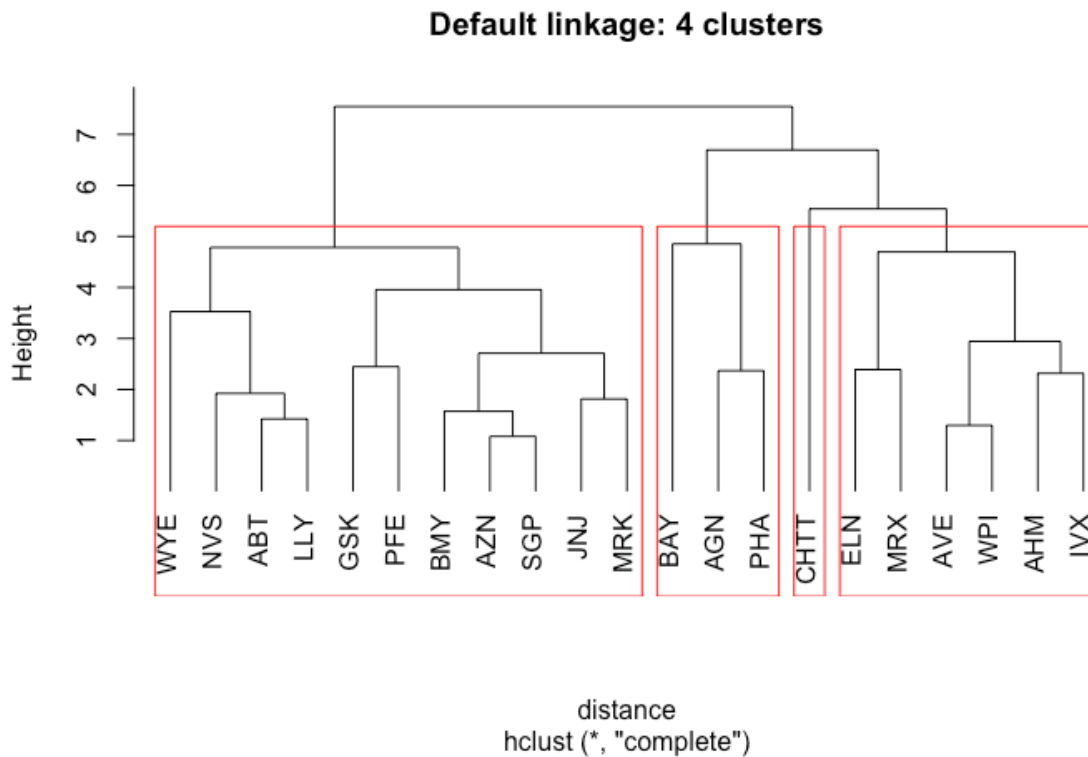


In choosing between 4 and 6 clusters, I first observed the cluster membership. For 4 clusters, there are 11,3,6,1 companies in each cluster respectively. For 6 clusters, it generated more outliers with 4,2,6,7,1,1 company in each cluster. In this sense, dividing into 4 clusters is a better choice.

Furthermore, we looked into the Silhouette Plots for both methods. Noticed that there are negative values for three firms when dividing into 6 clusters, while 4 clusters plot provided all positive values and a higher average silhouette width. Therefore, companies fit the characteristics of each cluster better with 4 clusters.



To conclude, I divide the 21 companies into 4 clusters, each with 11,3,1,6 members, as is shown below.



Interpret the clusters with respect to the quantitative variables that were used in forming the clusters.

```
member = cutree(ph.hclust,4)
table(member)
aggregate(nor,list(member),mean)
```

Hierarchical agglomerative clustering

	Group.1	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover
1	1	0.6733825	-0.3586419	-0.27635122	0.6565978	0.8344159	4.612656e-01
2	2	-0.5246281	0.4451409	1.84984387	-1.0404550	-1.1865838	1.480297e-16
3	3	-0.8094260	0.2244252	-0.42377658	-0.6648053	-0.7418386	-7.687760e-01
4	4	-0.9767669	1.2630872	0.03299122	-0.1123792	-1.1677918	-4.612656e-01
		Leverage	Rev_Growth	Net_Profit_Margin			
1		-0.3331068	-0.2902163	0.6823310			
2		-0.3443544	-0.5769454	-1.6095439			
3		0.1590734	0.9259960	-0.2380209			
4		3.7427970	-0.6327607	-1.2488842			

1. high profit margin, low leverage
2. low profit margin, low leverage
3. high profit margin, high leverage
4. low profit margin, high leverage

To define the cluster. The highest value and the second-highest value defined as high.

According to the hierarchical agglomerative clustering method, the leverage and net profit margin have the largest range which contributed to the variance between clusters. Cluster 1 is grouped by the highest profit margin which contributes to the high return on asset and low leverage. The low leverage means that the companies in cluster 1 are mostly financed by equity. Most of the profit will flow into the equity which makes the market capitalization the largest.

The cluster 2 is characterized by the lowest profit margin and low leverage which means that the return on equity is also low. The companies in cluster 2 are not worth investing in because they cannot create value for stockholders due to the low ROE.

Because the companies in the cluster 3 have a relatively high-profit margin and relatively high leverage. Most firms may have a larger interest burden which decreases the net profit a little bit. So the return on asset is the second largest.

The cluster 4 is grouped by the low-profit margin and high leverage which means most of the little profit needs to pay the interest. The stocks of the firms in cluster 4 are not good investments. The highest beta indicated that the companies' value is very fluctuated by the market.

K means:

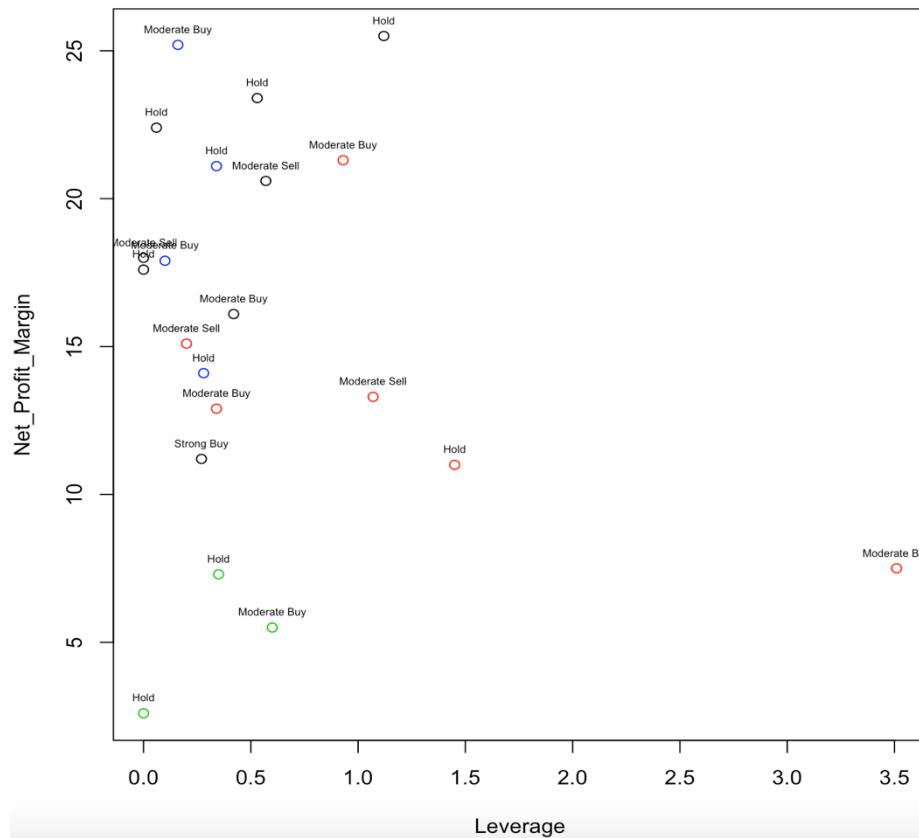
	Market_Cap	Beta	PE_Ratio	ROE	ROA	Asset_Turnover	Leverage
1	-0.03142211	-0.4360989	-0.3172485	0.1950459	0.4083915	1.729746e-01	-0.2744931
2	-0.82617719	0.4775991	-0.3696184	-0.5631589	-0.8514589	-9.994088e-01	0.8502201
3	-0.52462814	0.4451409	1.8498439	-1.0404550	-1.1865838	1.480297e-16	-0.3443544
4	1.69558112	-0.1780563	-0.1984582	1.2349879	1.3503431	1.153164e+00	-0.4680782
	Rev_Growth	Net_Profit_Margin					
1	-0.7041516	0.5569544					
2	0.9158889	-0.3319956					
3	-0.5769454	-1.6095439					
4	0.4671788	0.5912425					

In the k-means method, the clusters mostly are defined by market capitalization and ROA which have the largest range. The higher the ROA, the higher the ROE.

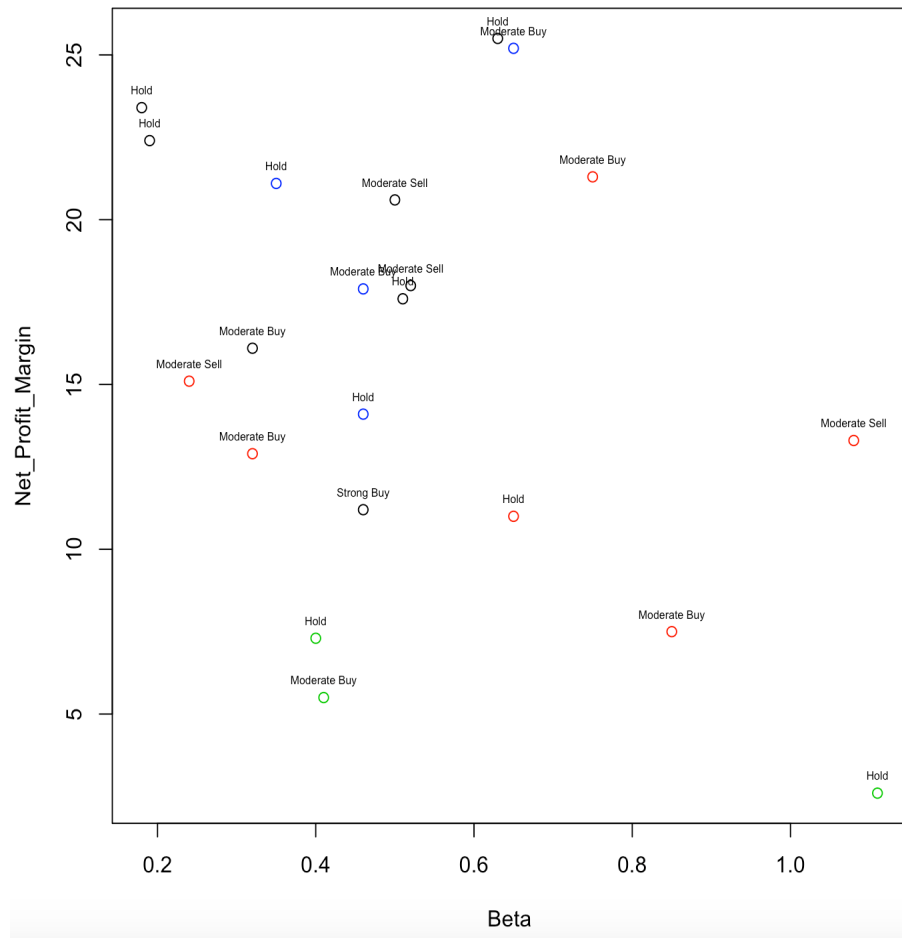
The companies in cluster 1 have relatively high market capitalization and relatively high returns on the asset which is indicated by the high net profit margin. The companies in cluster 1 have the lowest beta which means the companies are quite independent of the market. Cluster 2 and 3 are characterized by low market capitalization and low return on assets. The firms in cluster 3 have the highest PE ratio mainly because of the lowest net profit margin. The stock price of these firms will always be overvalued. The market capitalization can reflect the expectation of the value of the firms. Because the beta can measure the risk of a firm. The firms in cluster 4 have the highest net profit margin and relatively low beta, investors should invest the firms in cluster 4 compared to other clusters.

In conclusion, investors should not invest in companies with low-profit margins and low leverage. They should invest in firms that have a high net profit margin and relatively low beta.

Analyze the pattern in the clusters with respect to the qualitative variables (j)-(l) (that were not used in forming the clusters).



When considering firms' net profit margin and leverage, the recommendations are not that obvious in k-mean clusters. For the firms who have low leverage and low net profit margin in green, the recommendation is "hold". The investors can buy the stocks of the firms who have low leverage and moderate net profit margin because most of the net profit will flow into equity. For the firms who have low leverage and high net profit margin, the investors should hold the stocks.



For the companies that have high beta and low net profit margins, the recommendations are “sell”. The investors should buy firms who have low beta and moderate-high net profit margins. Firms with a high net profit margin and moderate low beta should be held.

Carrying out cluster analysis using self-organizing maps. Provided with approach, visualizations and interpretation.

We used unsupervised maps to carry out cluster analysis. The 21 observations of 9 variables are divided into 4 nodes as shown in the codes plot. The observations that are most similar are classified into the same node. According to figure 4-2, the second node contains the most observations, and almost half of the observations fall in this node. As illustrated by figure 4-1, observations in the second node have high values in PE ratio and Beta with moderate leverage and revenue growth. However, these observations have much lower values in the rest of the variables, compared with those from other nodes. Such a result provides us with the following interpretation on the pharmaceutical industry. In general, the growth expectations for the industry are high as most of the observations have high PE ratios. In fact, the

industry meets investors' expectations, and develops in the right direction as most companies achieve high revenue growth. According to figure 4-1, we can tell that most companies in the pharmaceutical industry are highly levered, which means that they industry relies more on debt than equity. However, the industry with high Beta shows its potential volatility. Although the companies achieve high revenue growth, the profits made by them are low relative to the revenue on average. This observation can be proved by most observations' low ROE values. Despite the high revenue growth, the industry is not efficient in making profits.

Figure 4-1
2 by 2 of Pharmaceutical Industry Data

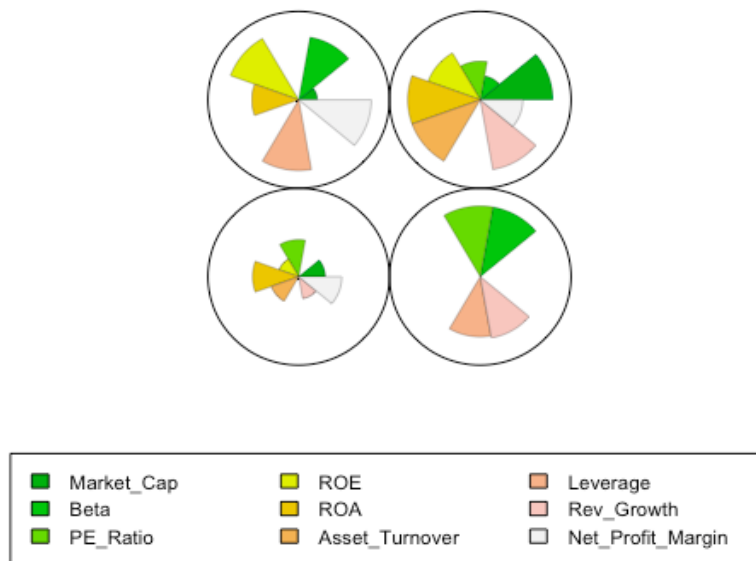
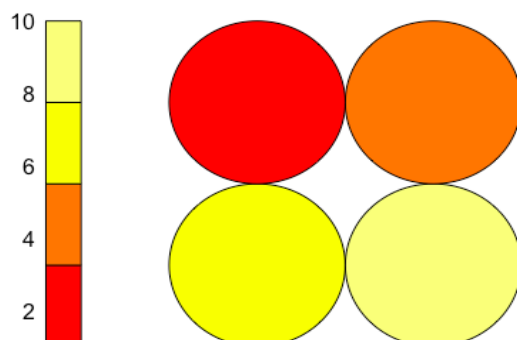


Figure 4-2

2 by 2 of Pharmaceutical Industry Data



Providing an appropriate name for each cluster using variables in the dataset after finalizing clusters.

Based on the hierarchical agglomerative clustering and self-organizing map, we divided the company into 4 groups.

The first cluster includes the companies that have low profit margin, high leverage and low PE ratio. The cluster is named as **startup cluster** with the stocks of startup companies because they use more debt when financing.

The second cluster includes the companies that have low profit margin, low leverage and high beta. The cluster is named as **growth cluster** with growth stock. The companies have mostly become public recently and cannot earn high profit, yet which is very similar to the pharmaceutical industry.

The third cluster is named as **value stocks** because the companies in the cluster mostly have high profit margin, low leverage and low PE ratio. The companies already went public for a long time and are more sensitive to the market.

The fourth cluster is named as **mature cluster** because the companies have high profit margin, high leverage and moderate PE ratio. The mature technologies help the companies make high profit margin and ROE.