Prof. Rai

**Template for Project-2: Cluster Analysis**

Names: Rongjia Jing
Yun Lin
Tianyi Zhang

1. (30) Use only the quantitative variables (a)-(i) to cluster the 21 firms. Provide results and justify the various choices made in conducting the cluster analysis such as weights accorded different variables, the specific clustering algorithm/s used, the number of clusters formed, etc.

As is shown in Figure 1-1, a hierarchical agglomerative cluster plot is drawn to show that 21 firms in the industry is grouped by four clusters, each cluster with company number of 11, 3, 1 and 6 respectively.

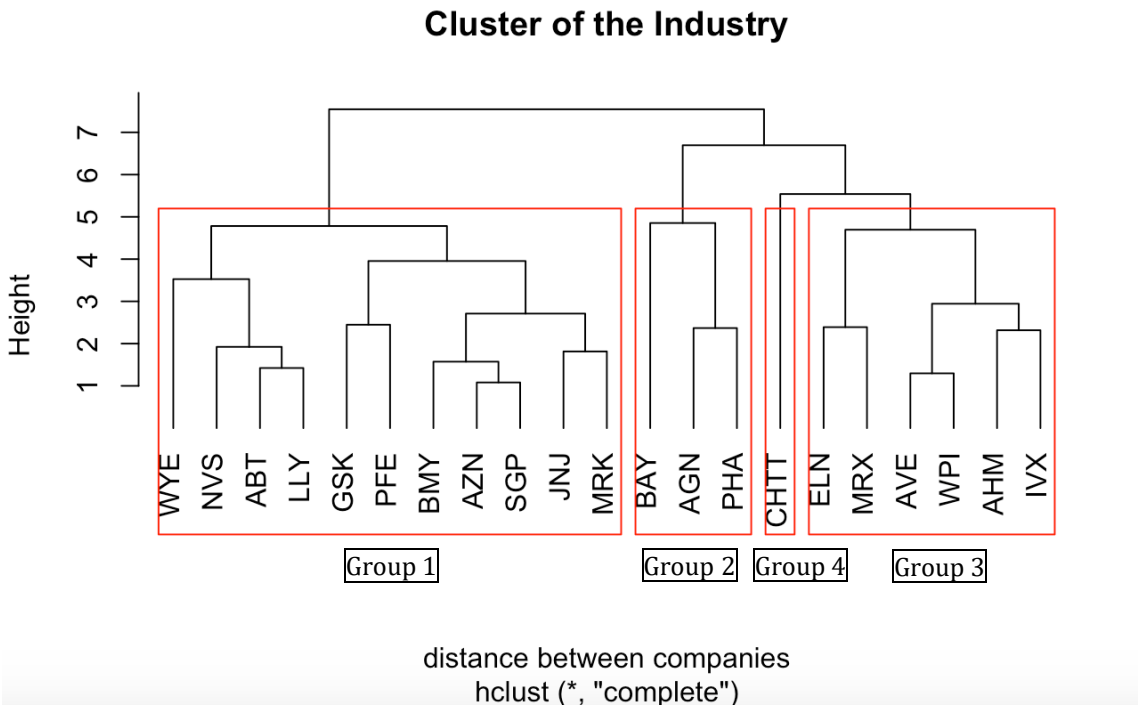


Figure 1-1. Cluster of the Industry

(1) specific clustering algorithm used

In the hierarchical agglomerative clustering, we use “complete” linkage and "average" linkage to compare the difference in the number of companies included in each

cluster of the two methods. As is seen in Figure 1-2, it is more balanced distributed when using “complete” linkage. Therefore, we choose to use “complete” linkage in clustering.

| | member.a | | | |
|---|---|---|---|---|
| member.c | 1 | 2 | 3 | 4 |
| 1 | 11 | 0 | 0 | 0 |
| 2 | 0 | 3 | 0 | 0 |
| 3 | 4 | 0 | 0 | 2 |
| 4 | 0 | 0 | 1 | 0 |

Figure 1-2. Distribution of groups with “complete” linkage and “average” linkage

(2) Choice of cluster number:

A scree plot is plotted to calculate the within square sum of different number of clusters. As is seen in Figure 1-3, there is a steep significant drop from point one to point 5. The graph normalizes at point 6 and 7 and then starts to fall gently.
Therefore, cluster number 4 and 5 are taken into consideration.

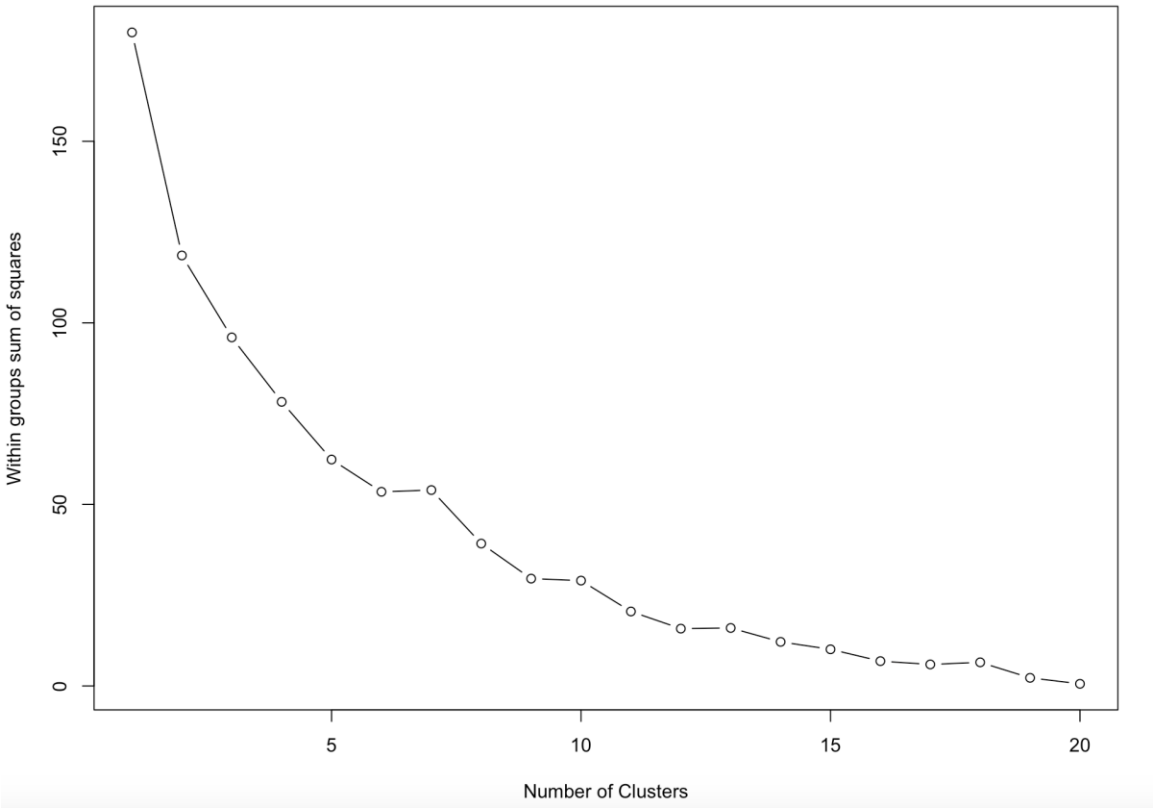


Figure 1-3. Scree Plot of clusters

If the industry is grouped by 5 clusters, two clusters have a size of 1, which is not as balanced as the distribution of group sizes in 4 clusters. Therefore, we choose to divide the 21 films into 4 clusters.

Membership of 5 clusters

| Group | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| number | 11 | 2 | 6 | 1 | 1 |

Membership of 4 clusters

| Group | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| number | 11 | 3 | 6 | 1 |

As is seen in the Silhouette plot (Figure 1-4), all the widths are positive number and the average width is 0.25.

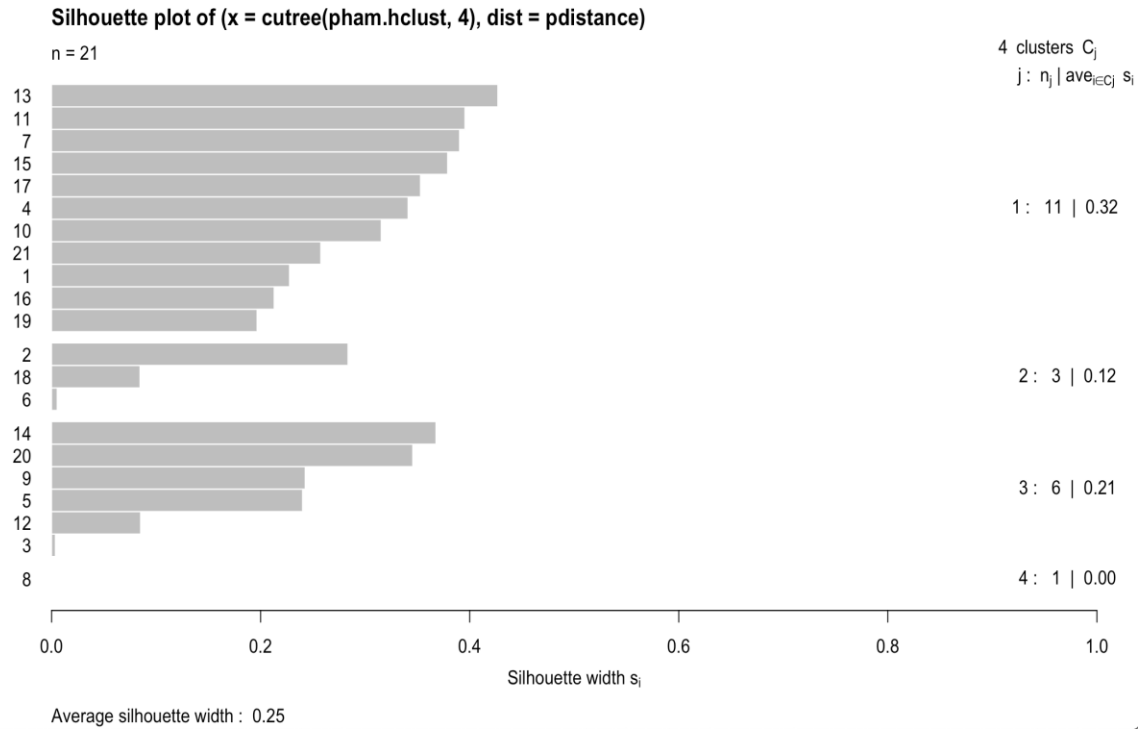


Figure 1-4. Silhouette plot of cluster

(3) weights accorded different variables

The cluster plot shows the impact and significance of the variables play in 4 clusters. In Figure 1-5, the two components consisting of all the quantitative variables explain 61.23% of the point variability.

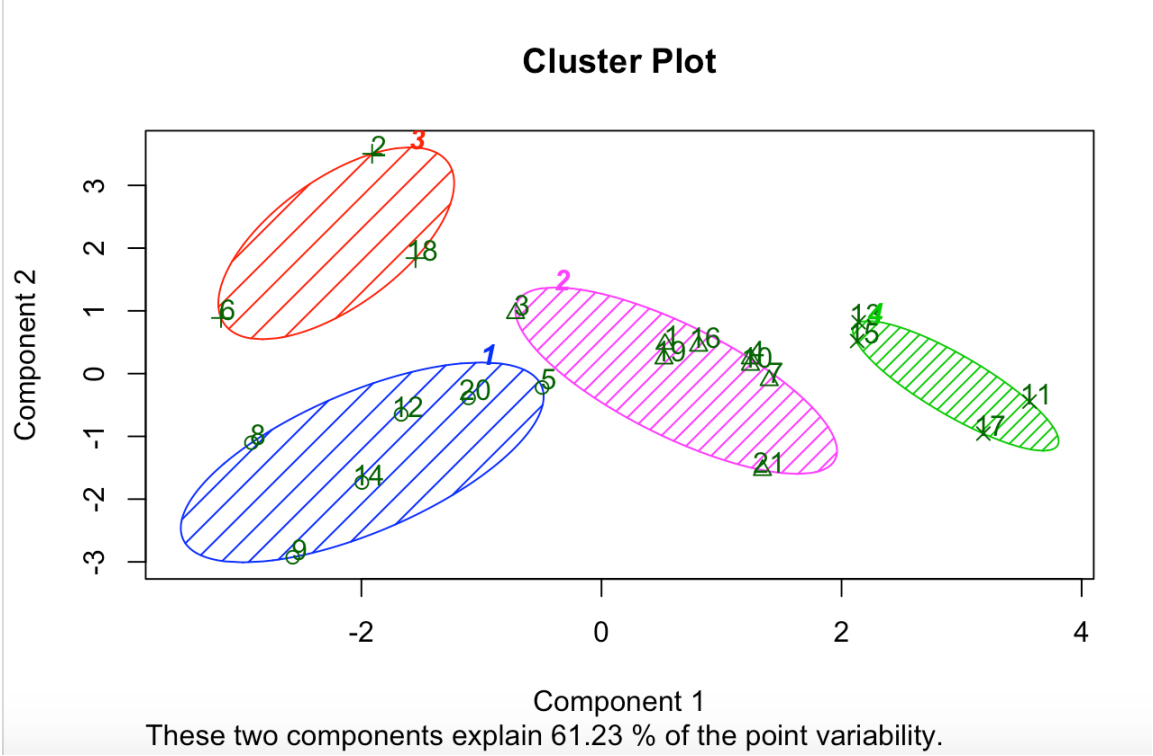


Figure 1-5. Cluster plot

The table shows the variations of 9 variables in four clusters with the normalized data. And we can see the variable “leverage” has the biggest variation between 4 clusters while “PE ratio” and “Net profit margin” has large variations as well. This means that these four clusters were mainly divided according to each company’s leverage, PE ratio, and Net profit margin.

| Group | Market_ Cap | Beta | PE_Ratio | ROE | ROA | Asset_ Turnover | Leverage | Rev_ Growth | Net_Profit_ Margin |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 0.7 | -0.4 | -0.3 | 0.7 | 0.8 | 0.5 | -0.3 | -0.3 | 0.7 |
| **2** | -0.5 | 0.4 | 1.8 | -1.0 | -1.2 | 0.0 | -0.3 | -0.6 | -1.6 |
| **3** | -0.8 | 0.2 | -0.4 | -0.7 | -0.7 | -0.8 | 0.2 | 0.9 | -0.2 |
| **4** | -1.0 | 1.3 | 0.0 | -0.1 | -1.2 | -0.5 | 3.7 | -0.6 | -1.2 |

Table 1-1. Variations with normalized data

2. (25) Interpret the clusters with respect to the quantitative variables that were used in forming the clusters.

| Group | Market_ Cap | Beta | PE_Ratio | ROE | ROA | Asset_ Turnover | Leve rage | Rev_ Growth | Net_Profit_ Margin |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 97.1 | 0.4 | 21.0 | 35.7 | 15.0 | 0.8 | 0.3 | 10.2 | 20.2 |
| **2** | 26.9 | 0.6 | 55.6 | 10.1 | 4.2 | 0.7 | 0.3 | 7.0 | 5.1 |
| **3** | 10.2 | 0.6 | 18.6 | 15.8 | 6.6 | 0.5 | 0.7 | 23.6 | 14.1 |
| **4** | 0.4 | 0.9 | 26.0 | 24.1 | 4.3 | 0.6 | 3.5 | 6.4 | 7.5 |

Table 2-1. Mean aggregation (using raw data)

From the table above we can see that some variables such as “Market Capitalization” and “PE ratio” have large differences between each group while some other variables such as “Asset turnover” and “Beta” don’t. Group 1 got the highest mean of Market capitalization, ROE, ROA, and Net profit margin but its PE ratio was not the highest. And at the same time, Group 4 got the lowest mean of Market Capitalization but the second highest mean of PE ratio and ROE.

And in real life, people should better invest in companies involved in Group 1 if they are risk aversion while others who want more gain by taking higher risk should invest in Group 3. Since Group 1 companies have a relatively low beta with a relatively high Net profit margin and ROE which means that these companies might develop more steadily, while Group 3 have a relatively high beta but relatively high Revenue growth which means that there will be better development in the future.

3. (20) Is there a pattern in the clusters with respect to the qualitative variables (j)-(l) (that were not used in forming the clusters)?

| Count (Recommendation) | Hold | % of Group | Moderate Buy | % of Group | Moderat e Sell | % of Group | Strong Buy | % of Group | Grand Total |
|---|---|---|---|---|---|---|---|---|---|
| Group1 | 6 | 54.55% | 3 | 27.27% | 2 | 18.18% | | | 11 |
| Group2 | 2 | 66.67% | 1 | 33.33% | | | | | 3 |
| Group3 | | | 1 | 100.00% | | | | | 1 |
| Group4 | 1 | 16.67% | 2 | 33.33% | 2 | 33.33% | 1 | 16.67% | 6 |

Table 3-1. Count (Recommendation) in each group

The Recommendation column can be divided into 4 parts:

For Group 1, most companies are recommended to be hold or sold because their revenue growth rate are already high, and thus are likely to decrease in growth rate. This is because companies in group 1 are mature companies, which have less potential compared to growing companies.

For Group 3, AVE is recommended as ‘moderate-buy’ because they are to be acquired by a big company in 2005, the stock price will boom after M&A. AHM is recommended as ‘strong buy’ probably because their revenue growth will increase dramatically from 7%. The recommendations are divided because those companies are at growing stage, so their market opportunities and risk factors are diverse and hard to determine.

For Group 4, CHTT is rated as ‘moderate-buy’ since it has a high beta meaning its return is determined by the market return.

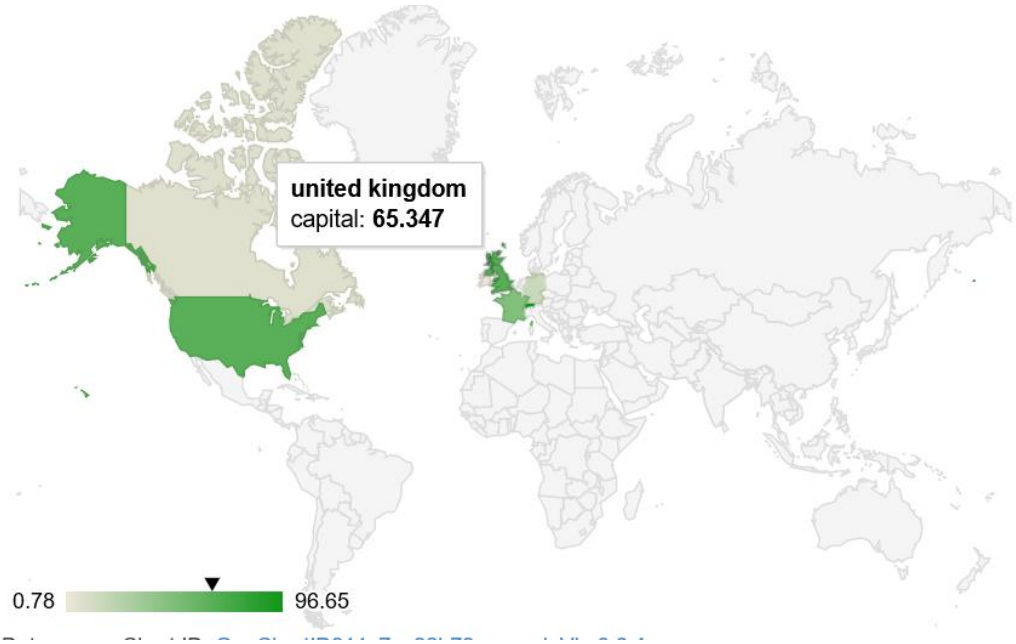


Figure 3-1. World Map of Capitalization

From the world map, the US and UK have the most market capitalization.

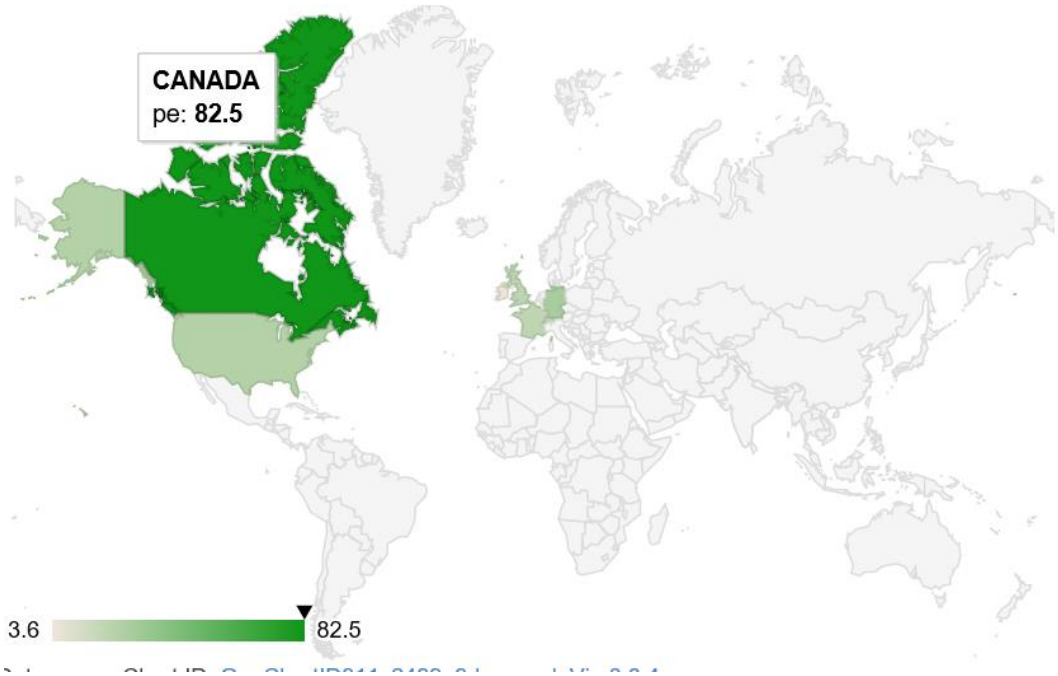


Figure 3-2. World Map of P/E

Canada has the most P/E ratio.

| | Row Labels | Average of Market_Cap | Average of Beta | Average of PE_Ratio | Average of ROE | Average of ROA | Average of Asset_Turnover | Average of Leverage | Average of Rev_Growth | Average of Net_Profit_Margin |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | AMEX_average | 2.60000 | 0.6500000 | 19.90000 | 21.40000 | 6.80000 | 0.6000000 | 1.4500000 | 13.99000 | 11.00000 |
| 2 | NASDAQ_average | 0.41000 | 0.8500000 | 26.00000 | 24.10000 | 4.30000 | 0.6000000 | 3.5100000 | 6.38000 | 7.50000 |
| 3 | NYSE_average | 63.56158 | 0.5021053 | 25.72632 | 26.11579 | 11.03684 | 0.7105263 | 0.3863158 | 13.70632 | 16.37368 |

Table 3-2. Average ratios Group by Stock Exchange

For Exchange column, there are only two companies in AMEX and NASDAQ (Group 3). They have lower market capitalization, return on equity and profit margin because they are at earlier stage than those companies in NYSE. But they have higher Beta and leverage, meaning they have higher return related to the market, and they rely on debt to acquire cash.

4. (20) Provide an appropriate name for each cluster using any/all of the variables in the dataset.

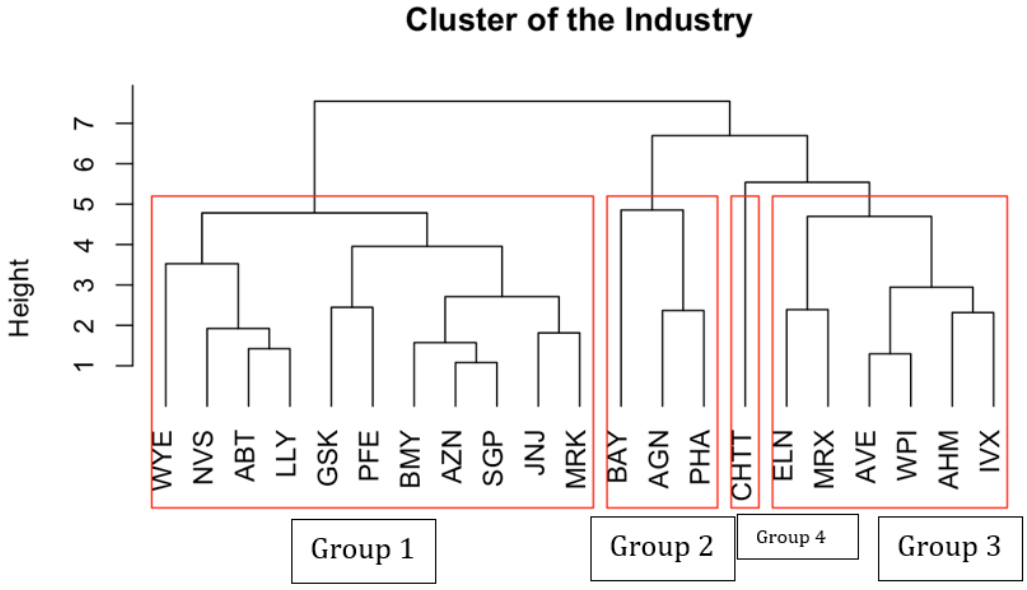


Figure 4-1. Cluster of the Industry

| Group | Market_ Cap | Beta | PE_Ratio | ROE | ROA | Asset_ Turnover | Leverage | Rev_ Growth | Net_Profit_ Margin |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 97.1 | 0.4 | 21.0 | 35.7 | 15.0 | 0.8 | 0.3 | 10.2 | 20.2 |
| **2** | 26.9 | 0.6 | 55.6 | 10.1 | 4.2 | 0.7 | 0.3 | 7.0 | 5.1 |
| **3** | 10.2 | 0.6 | 18.6 | 15.8 | 6.6 | 0.5 | 0.7 | 23.6 | 14.1 |
| **4** | 0.4 | 0.9 | 26.0 | 24.1 | 4.3 | 0.6 | 3.5 | 6.4 | 7.5 |

Table 4-1. Mean aggregation (using raw data)

Group 4: Start-up
Group 4 consists of Chattem Inc on NASDAQ, indicating it's at initial stage. A pharmaceutical start-up usually has its medicine in development, so it needs a lot of development fund and does not have enough cash flow from net income. So Chattem has the highest leverage and lowest revenue growth compared to other groups.

Group 3: Growing Companies
This group of companies have the highest growth rates of revenue. At growing stage, pharmaceutical companies have just promoted finished products to the market and received sales boom. But since there are a lot of risks such as regulation risk and imitation risk, their stock price/earnings ratio is still lower than developed companies.

Group 2: Over-priced Companies
The three companies have the lowest average return on equity, return on assets and margin but have the highest price/earnings ratio. So it is obvious that they are overvalued by investors.

Group 1: Mature Companies
Those companies represent top players in the pharmaceutical industry. They have the highest margins, indicating that their products are formalized and mass-produced. Highest average ROE and ROA means they efficiently utilize their assets to create profits. Lowest average Beta means their business models are well established and are unlikely to be affected by the market.