# Cluster Analysis: Pharmaceutical Industry & Cereals

MIS 749 – Spring 2015

Samuel Beckom and Roger Chen
3-24-2015

**14.2 Pharmaceutical Industry**

**Use only the quantitative variables (1-9) to cluster the 21 firms. 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, and so on.**

I used hierarchical clustering with Euclidean distance being used and with the variable values being normalized to their respective z-scores. I used Ward's method over complete linkage, single linkage, and centroid distance since it made the most distinct groups. Using the dendrogram in Figure 1 and setting 10 as the cutoff value for distance between clusters, I discovered the following four clusters:

- Cluster 1
  - AstraZeneca PLC
  - Schering-Plough Corporation
  - Bristol-Myers Squibb Company
  - Abbott Laboratories
  - Eli Lilly and Company
  - Novartis AG
  - Wyeth
- Cluster 2
  - Johnson & Johnson
  - Merck & Co., Inc.
  - GlaxoSmithKline plc
  - Pfizer Inc.
- Cluster 3
  - Elan Corporation, plc
  - Medicis Pharmaceutical Corporation
  - Aventis
  - Watson Pharmaceuticals, Inc.
  - Amersham plc
  - IVAX Corporation
- Cluster 4
  - Allergan, Inc.
  - Pharmacia Corporation
  - Bayer AG
  - Chattem, Inc.

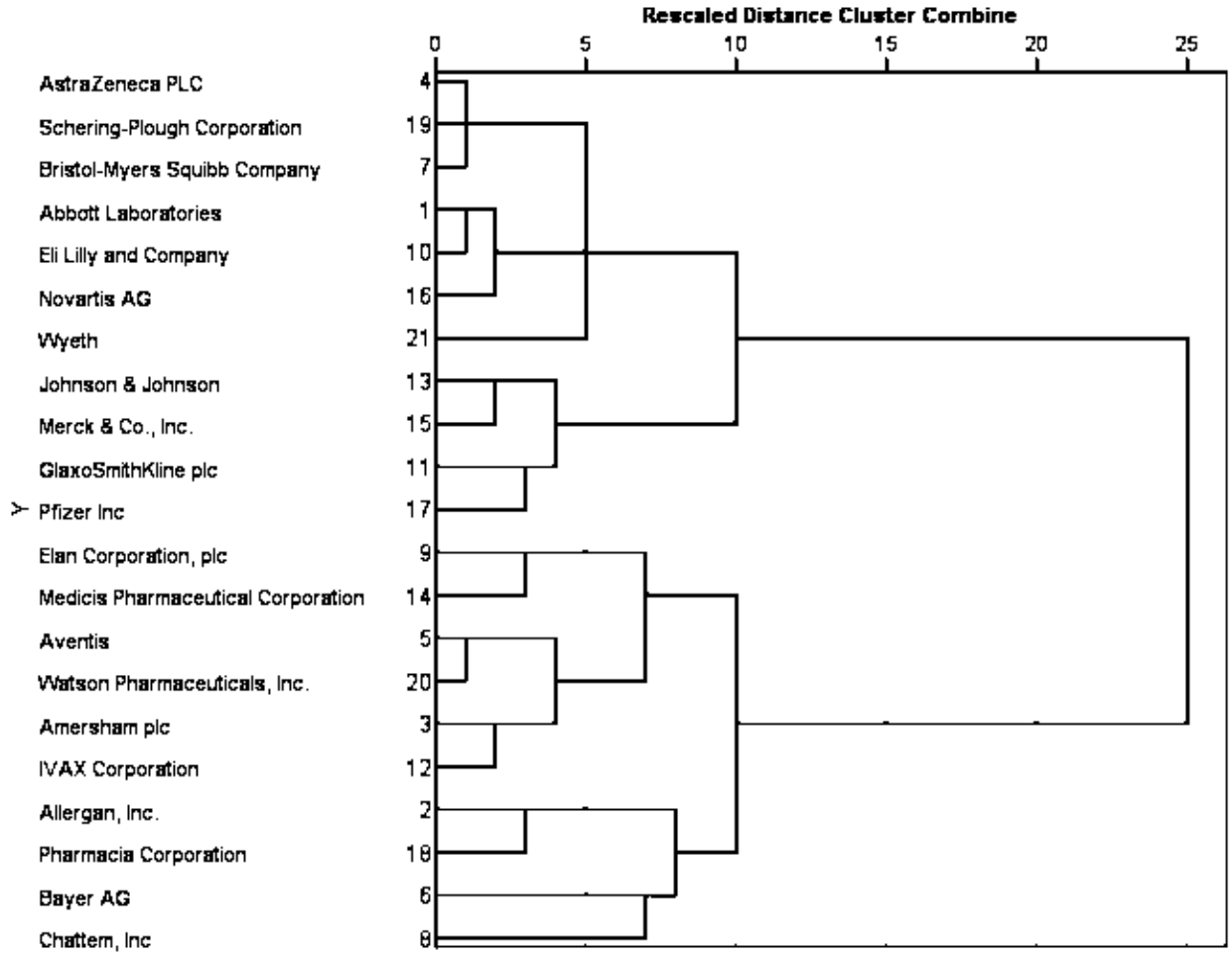

Figure 1. Dendrogram using Ward Linkage for Pharmaceutical Industry

**Summary Statistics**

The next step is to look at the summary statistics for each cluster to determine how these companies relate to each other so that a label can be applied to each cluster.

Table 1. Cluster 1 Summary Statistics

| | Market Cap. | Beta | PE Ratio | ROE | ROA | Asset Turnover | Leverage | Rev. Growth | Net Profit |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 62.88 | 0.41 | 20.23 | 30.71 | 13.39 | 0.71 | 0.39 | 5.38 | 20.51 |
| St. Dev. | 20.38 | 0.18 | 5.40 | 11.98 | 1.54 | 0.16 | 0.41 | 5.85 | 3.44 |
| Count | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| Min. | 34.10 | 0.18 | 13.10 | 17.90 | 11.20 | 0.50 | 0.00 | -2.69 | 16.10 |
| Max. | 96.65 | 0.63 | 27.90 | 54.90 | 15.40 | 0.90 | 1.12 | 15.00 | 25.50 |

Table 2. Cluster 2 Summary Statistics

| | Market Cap. | Beta | PE Ratio | ROE | ROA | Asset Turnover | Leverage | Rev. Growth | Net Profit |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 157.02 | 0.48 | 22.23 | 44.43 | 17.70 | 0.95 | 0.22 | 18.53 | 19.58 |
| St. Dev. | 36.08 | 0.12 | 4.79 | 14.23 | 2.47 | 0.13 | 0.11 | 6.97 | 4.72 |
| Count | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Min. | 122.11 | 0.35 | 18.00 | 28.60 | 15.00 | 0.80 | 0.10 | 9.37 | 14.10 |
| Max. | 199.47 | 0.65 | 28.40 | 62.90 | 20.30 | 1.10 | 0.34 | 25.54 | 25.20 |

Table 3. Cluster 3 Summary Statistics

| | Market Cap. | Beta | PE Ratio | ROE | ROA | Asset Turnover | Leverage | Rev. Growth | Net Profit |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 10.22 | 0.58 | 18.55 | 15.77 | 6.57 | 0.53 | 0.71 | 23.60 | 14.13 |
| St. Dev. | 18.20 | 0.31 | 8.17 | 4.92 | 1.10 | 0.23 | 0.51 | 10.64 | 3.82 |
| Count | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| Min. | 0.78 | 0.24 | 3.60 | 10.20 | 5.10 | 0.30 | 0.20 | 7.05 | 11.00 |
| Max. | 47.16 | 1.08 | 28.60 | 21.80 | 7.80 | 0.90 | 1.45 | 34.21 | 21.30 |

Table 4. Cluster 4 Summary Statistics

| | Market Cap. | Beta | PE Ratio | ROE | ROA | Asset Turnover | Leverage | Rev. Growth | Net Profit |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 20.28 | 0.69 | 48.23 | 13.60 | 4.23 | 0.68 | 1.12 | 6.84 | 5.73 |
| St. Dev. | 24.90 | 0.35 | 26.77 | 8.26 | 1.98 | 0.15 | 1.62 | 7.58 | 2.27 |
| Count | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| Min. | 0.41 | 0.40 | 26.00 | 3.90 | 1.40 | 0.60 | 0.00 | -3.17 | 2.60 |
| Max. | 56.24 | 1.11 | 82.50 | 24.10 | 5.70 | 0.90 | 3.51 | 15.00 | 7.50 |

**Labeling the Clusters from the Quantitative and Qualitative Data**

*Cluster 1*:
Table 1 shows that this cluster has a low mean beta (meaning that these companies' stock are less volatile than the market as a whole), has a relatively high mean return on equity (ROE) and mean return on assets (ROA), low mean leverage, and high mean asset turnover. Additionally, this cluster has the highest mean net profit value relative to the other three clusters. With regards to the qualitative data, most of the median recommendations from major brokerages is to either moderately buy or hold these companies' stock. This cluster is labeled Investor Friendly Cluster.

*Cluster 2*:
Table 2 shows cluster 2's summary statistics. This cluster has the highest mean market capitalization, ROE, ROA, and asset turnover. This cluster also has a relatively low mean leverage ratio, and has a relatively high mean net profit. Also the median recommendation from stock brokers is to either hold or moderately buy these companies' stock. Thus, this cluster is labeled the Pharmaceutical Industry Giants.

*Cluster 3*:
As can be seen from Table 3, this cluster has the second highest mean beta, has a high mean leverage ratio, and the highest mean revenue growth relative to the other clusters. This suggest that these companies are conducting relatively risky activities. When companies make risky investments (meaning high risk, high reward investments) it is likely due to the fact that they are focusing on growing the company. With regards to the qualitative data, there does not appear to be any meaningful pattern for this cluster. This company, then, is labeled Growth Based Pharmaceutical Companies.

*Cluster 4*:
Table 4 shows the summary statistics of cluster 4. This cluster has the highest mean beta relative to the other clusters. Additionally, this cluster has a low mean revenue growth and net profit. This cluster also has a mean leverage ratio of 1.12, meaning that this cluster has 12% more liabilities than assets. However, this cluster has the highest mean price per earnings (P/E) ratio which means that investors believe that the earnings per share for this will grow faster than the other clusters. Indeed, the median recommendation across the major brokerages is to either moderately buy or hold these companies' stock. This cluster is labeled the Undervalued Pharmaceutical Companies.

**14.3 Customer Rating of Breakfast Cereals**

The following are the dendrograms associated with the hierarchical clustering of the breakfast cereal dataset using single linkage, complete linkage, and cluster centroids.

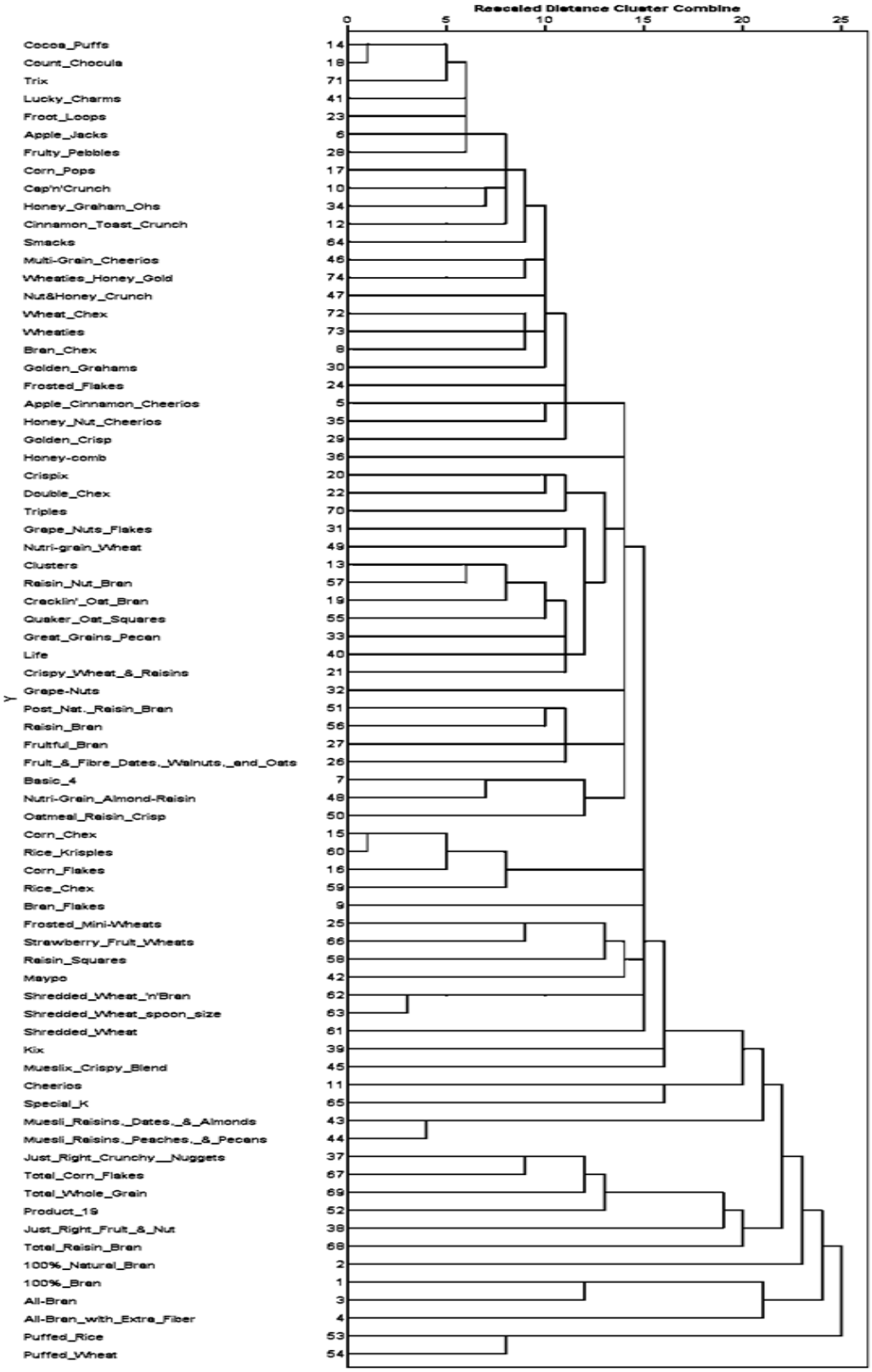


Figure 2. Dendrogram for Breakfast Cereals Using Single Linkage.

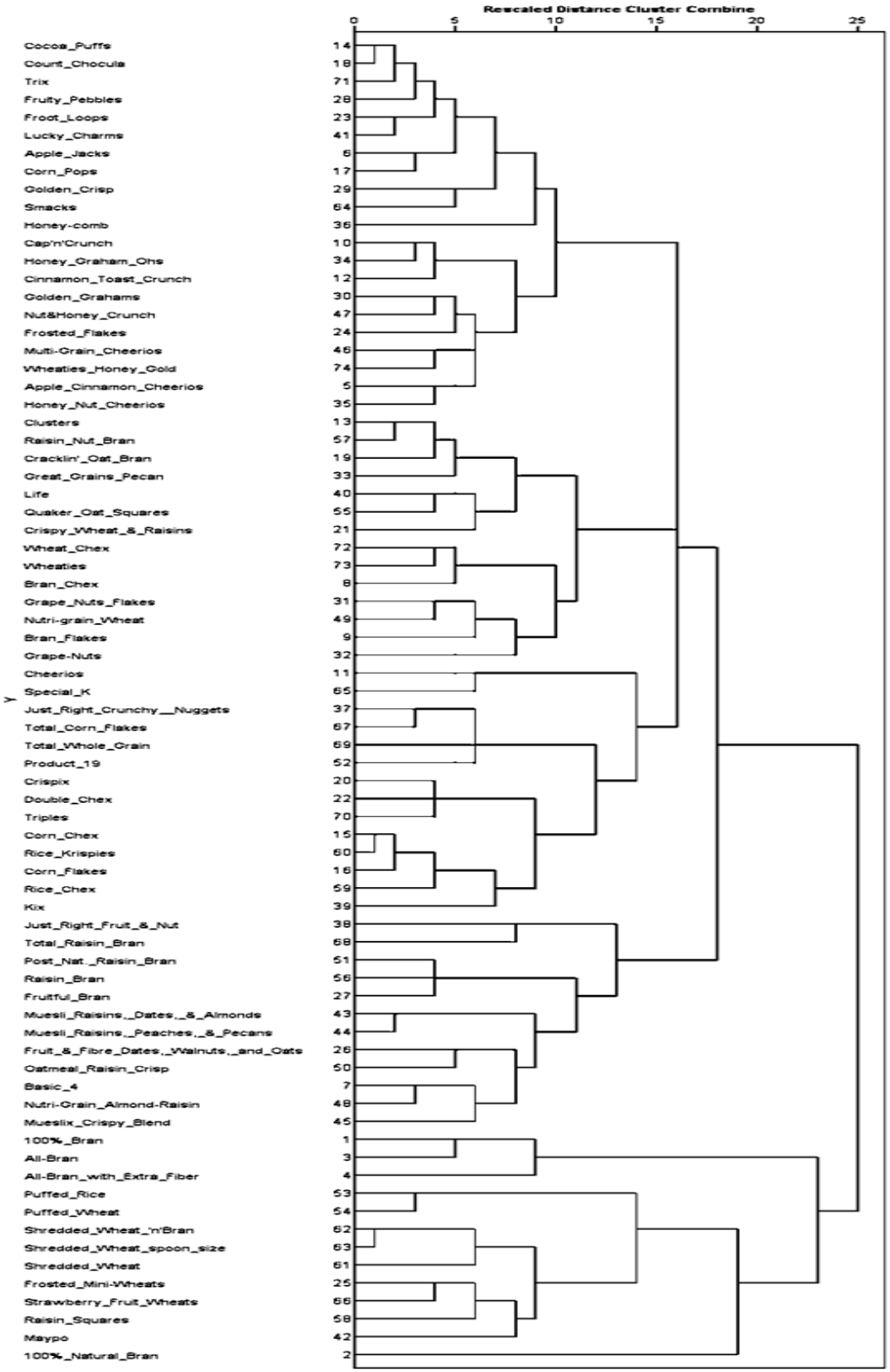


Figure 3. Dendrogram for Breakfast Cereals Using Complete Linkage.

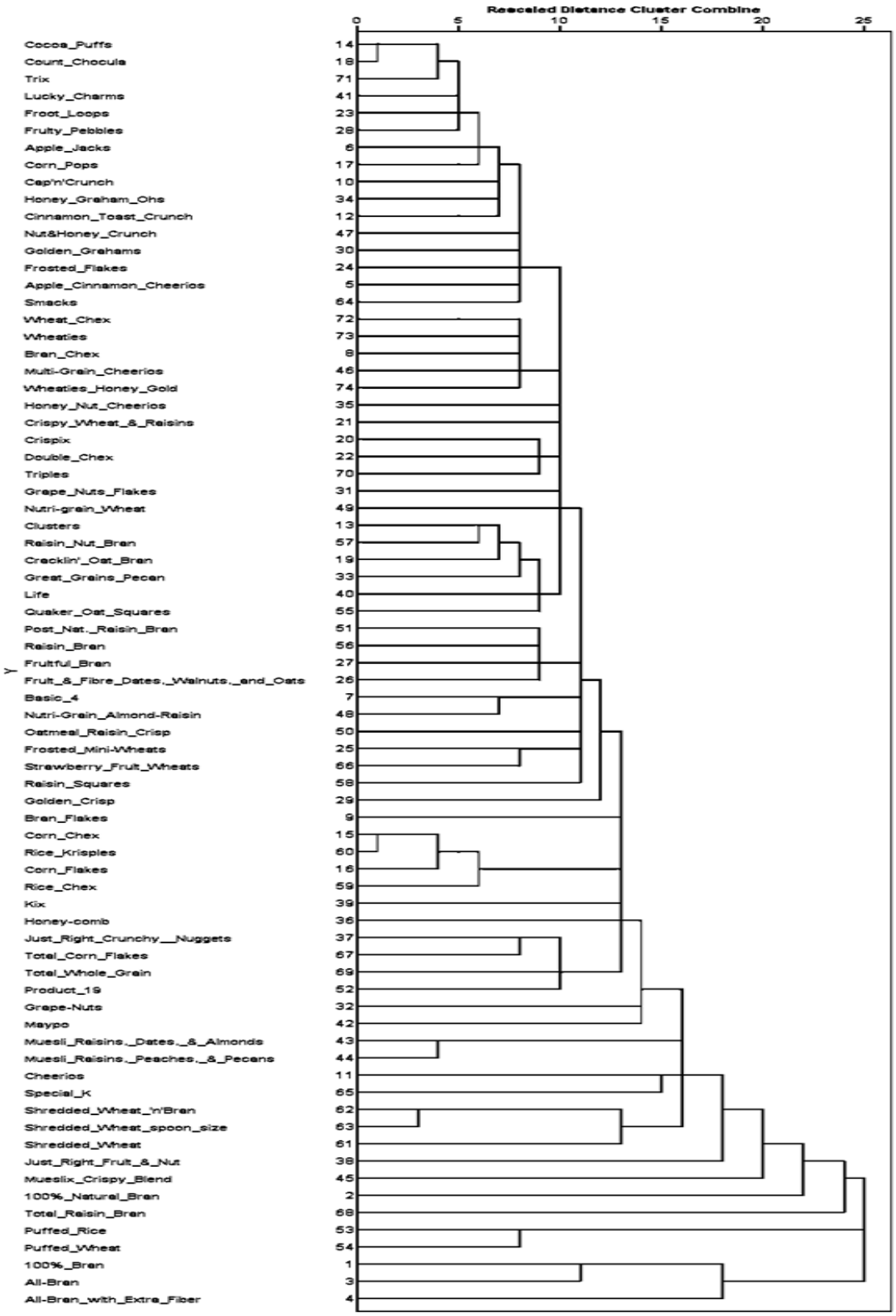


Figure 5. Dendrogram for Breakfast Cereals Using Centroid Linkage

**Which method leads to the most insightful or meaningful clusters?**

All of the methods appear to be clustering the same cereals, for the most part. The most meaningful method appears to be the method using complete linkage.

**Choose one of the methods. How many cluster would you use? What distance is used for this cutoff?**

I would use the complete linkage method, since as stated above, this methods appears to be the most meaningful. I would use six clusters which is associated with a cutoff value of 10. Each cereal associated with each cluster is as follows:

- Cluster 1
  - Apple Cinnamon Cheerios
  - Apple Jacks
  - Cap'n'Crunch
  - Cinnamon Toast Crunch
  - Cocoa Puffs
  - Corn Pops
  - Count Chocula
  - Froot Loops
  - Frosted Flakes
  - Fruity Pebbles
  - Golden Crisp
  - Golden Grahams
  - Honey Graham Ohs
  - Honey Nut Cheerios
  - Honey-Comb
  - Lucky Charms
  - Multi-Grain Cheerios
  - Nut & Honey Crunch
  - Smacks
  - Trix
  - Wheaties Honey Gold

- Cluster 2
  - Bran Chex
  - Bran Flakes
  - Clusters
  - Cracklin' Oat Bran
  - Crispy Wheat & Raisins
  - Grape Nut Flakes
  - Grape-Nuts
  - Great Grains Pecan
  - Life

  - Nutri-Grain Wheat
  - Quaker Oat Squares
  - Raisin Nut Bran
  - Wheat Chex
  - Wheaties

- Cluster 3
  - Cheerios
  - Corn Chex
  - Corn Flakes
  - Crispix
  - Double Chex
  - Just Right Crunchy Nuggets
  - Kix
  - Product 19
  - Rice Chex
  - Rice Krispies
  - Special K
  - Total Corn Flakes
  - Total Whole Grain
  - Triples

- Cluster 4
  - Basic 4
  - Fruit & Fibre Dates, Walnuts, and Oats
  - Fruitful Bran
  - Just Right Fruit & Nut
  - Muesli Raisins, Dates, & Almonds
  - Muesli Raisins, Peaches, & Pecans
  - Mueslix Crispy Blend
  - Nutri-Grain Almond-Raisin
  - Oatmeal Raisin Crisp
  - Post Nat. Raisin Bran
  - Raisin Bran
  - Total Raisin Bran

- Cluster 5
  - 100% Bran
  - All-Bran
  - All-Bran with Extra Fiber
  - Frosted Mini-Wheats
  - Maypo
  - Puffed Rice
  - Puffed Wheat

  - Raisin Squares
  - Shredded Wheat
  - Shredded Wheat'n'Bran
  - Shredded Wheat Spoon Size
  - Strawberry Fruit Wheats

- Cluster 6
  - Natural Bran.

Tables 7-12 show the summary statistics for the variables associated with the six clusters used to determine the appropriate labels for each cluster. For the interest of space and since these variables were not considered when determining the cluster labels, variables shelf, weight, cups, and rating are not included in the following tables.

Table 7. Summary Statistics for Cluster 1

| | Calories | Protein | Fat | Sodium | Fiber | Carbo. | Sugar | Potass | Vitamins |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 110.95 | 1.52 | 1.00 | 172.38 | 0.57 | 12.62 | 11.29 | 45.95 | 25.00 |
| St. Dev. | 5.39 | 0.60 | 0.77 | 58.47 | 0.66 | 1.75 | 2.28 | 20.04 | 0.00 |
| Count | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 | 21 |
| Min. | 100 | 1 | 0 | 45 | 0 | 9 | 6 | 20 | 25 |
| Max. | 120 | 3 | 3 | 280 | 2 | 16 | 15 | 90 | 25 |

Table 8. Summary Statistics for Cluster 2

| | Calories | Protein | Fat | Sodium | Fiber | Carbo. | Sugar | Potass | Vitamins |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 101.43 | 3.00 | 1.29 | 160.00 | 2.96 | 13.96 | 5.36 | 116.79 | 25.00 |
| St. Dev. | 8.64 | 0.55 | 0.99 | 39.95 | 0.89 | 2.62 | 2.24 | 29.46 | 0.00 |
| Count | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| Min. | 90 | 2 | 0 | 75 | 2 | 10 | 2 | 85 | 25 |
| Max. | 120 | 4 | 3 | 230 | 5 | 18 | 10 | 190 | 25 |

Table 9. Summary Statistics for Cluster 3

| | Calories | Protein | Fat | Sodium | Fiber | Carbo. | Sugar | Potass | Vitamins |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 107.14 | 2.64 | 0.50 | 245.00 | 0.79 | 19.71 | 3.07 | 53.21 | 46.43 |
| St. Dev. | 4.69 | 1.50 | 0.65 | 45.36 | 0.89 | 2.40 | 1.21 | 27.50 | 35.16 |
| Count | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 | 14 |
| Min. | 100 | 1 | 0 | 170 | 0 | 16 | 1 | 25 | 25 |
| Max. | 110 | 6 | 2 | 320 | 3 | 23 | 6 | 110 | 100 |

Table 10. Summary Statistics for Cluster 4

| | Calories | Protein | Fat | Sodium | Fiber | Carbo. | Sugar | Potass | Vitamins |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 135.00 | 3.17 | 1.67 | 180.42 | 3.54 | 15.63 | 10.92 | 172.08 | 37.50 |
| St. Dev. | 13.82 | 0.39 | 0.89 | 39.45 | 1.44 | 3.02 | 2.23 | 54.42 | 29.19 |
| Count | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Min. | 120 | 3 | 0 | 95 | 1.5 | 11 | 7 | 95 | 25 |
| Max. | 160 | 4 | 3 | 240 | 6 | 21 | 14 | 260 | 100 |

Table 11. Summary Statistics for Cluster 5

| | Calories | Protein | Fat | Sodium | Fiber | Carbo. | Sugar | Potass | Vitamins |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 77.50 | 2.83 | 0.25 | 45.42 | 4.33 | 13.17 | 2.67 | 145.42 | 14.58 |
| St. Dev. | 19.13 | 1.03 | 0.45 | 85.05 | 4.36 | 4.73 | 2.93 | 104.85 | 12.87 |
| Count | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| Min. | 50 | 1 | 0 | 0 | 0 | 5 | 0 | 15 | 0 |
| Max. | 100 | 4 | 1 | 260 | 14 | 20 | 7 | 330 | 25 |

**Labeling the Clusters**

*Cluster 1*:
From Table 7, we see this group is characterized by a high level of sugar. The label for the cluster is, Sugary Cereals.

*Cluster 2*:
Table 8 shows that this cluster is relatively low in calories, sodium, fat, and sugar. This cluster is labeled Healthy Cereals.

*Cluster 3*:
Looking at Table 9, we see that this cluster is highest in sodium relative to all other clusters. Thus, this cluster is labeled Sodium-Rich Cereals.

*Cluster 4*:
Table 10 shows that this cluster is the highest in calories. This cluster is labeled the Calorie Rich Cluster.

*Cluster 5*:
Cluster five, as can be seen from Table 11 is highest in fiber more than any other cluster, thus the label High Fiber Cluster.

*Cluster 6*:
The final cluster is a singleton, 100% Natural Bran.

**Recommendation for a set of five healthy cereals**

The data should be standardized because this converts all of the values to the same scale so that no one variable has more influence on the model when measuring the Euclidean distance between two observations. If we drop variables shelf, weight, cups, and rating, then cluster instead around calories, protein, fat, sodium, fiber, carbohydrates, sugar, potassium, and vitamins, using a cutoff value of five, then the Healthy Cereal Cluster is reduced to the following five cereals:

1. Cluster
2. Grape Nuts Flakes
3. Life
4. Quaker Oat Squares

5. Raisin Nut Bran

Table 12 shows that this reduced healthy cereal cluster's summary statistics still have similar properties to the larger healthy cereal cluster determined above. Additionally, as can be seen from above all five of these cereals are in the larger healthy cereals cluster.

Table 12. Summary Statistics for the Five Recommended Healthy Cereals

| | Calories | Protein | Fat | Sodium | Fiber | Carbo. | Sugar | Potass | Vitamins |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 102.00 | 3.40 | 1.60 | 141.00 | 2.30 | 12.90 | 6.40 | 107.00 | 25.00 |
| St. Dev. | 4.47 | 0.55 | 0.55 | 5.48 | 0.45 | 1.75 | 1.14 | 20.80 | 0.00 |
| Count | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| Min. | 100 | 3 | 1 | 135 | 2 | 10.5 | 5 | 85 | 25 |
| Max. | 110 | 4 | 2 | 150 | 3 | 15 | 8 | 140 | 25 |