CSC 466 Lab 4 Report: Clustering

Otakar Andrysek | oandryse@calpoly.edu Nathan Johnson | njohns60@calpoly.edu

Introduction

In Lab 4 we implemented three different algorithms (K-means, Hierarchical Clustering, and DBSCAN) to perform unsupervised learning (clustering). These algorithms were then run against six datasets to calibrate ideal hyperparameters for each dataset and to discover strengths and weaknesses of each of the algorithms.

Study Design

For the **K-means** clustering algorithm we decided to mainly stick to the simpler implementation. We did implement the initial selection of centroids that was discussed in class, where each next centroid is the furthest point from all other centroids. We have two stopping conditions: when the new centroid is equal to the previous centroid or after 500 iterations. We are using euclidean distance as the distance measure, the centroids are computed as the means of their respective cluster, we didn't apply any transformations to the data, and did not address outliers.

As with K-means there were also some choices we had when implementing our **Agglomerative Hierarchical** Clustering Algorithm. Firstly, we decided to keep using euclidean distance as our distance measure. To calculate the distance between clusters we decided to use complete linkage. Complete linkage is very interesting because it computes the distance between clusters as the furthest distance between all points of the two clusters. This causes the clusters that are created to be tighter or closer together than a single or average link. We also decided to have our dendrogram in JSON format rather than XML format.

Our **DBSCAN** implementation was straightforward, mostly relying on a custom data structure, DBPoint. Upon reading the dataset all points are pushed into a list of DBPoints. This structure among other attributes holds the cluster information, point type, and neighbor DBpoint references. All distance values are Euclidean distances and the list of points is updated, once for neighborhood discovery, again for cluster formation, and finally for point type selection. Cluster and outlier information is outputted to the console.

Results

4clusters.csv

Best Algorithm: DBSCAN
Number of Clusters: 4
Average SSE of clusters: 124.2
Hyperparameters of best Algorithm: Epsilon: 5, MinPoints: 4
Observations:

- DBSCAN: handles this dataset very well. Four distinct clusters were found with some of the noise left as outliers (15% of the data). If epsilon was increased to reduce outlier the clusters would converge, failing the given expectation of creating four unique clusters. On visual observation, however both the four cluster and three clutter results appear reasonable.
- K-Means: To determine the best number of clusters we looked at the total SSE for each value of k, and after k reaches an optimal value the total SSE will show less improvement with an addition of another cluster. After testing k's we noticed that the 5th cluster is the last cluster that adds significant improvement to the total SSE. Even though the fifth cluster breaks up a visually obvious cluster, we felt it would be better to use the k that the algorithm thinks is better (k = 5), compared to the k we think is better. This will allow us to better compare the models because it gets rid of some subjectiveness.
- Hierarchical Clustering: To test hierarchical clustering we again looked at the total SSE for as clusters change. We evaluated thresholds until the number of clusters changed, and then recorded the total SSE for the new cluster. After looking at cluster sizes 1 - 6 we felt that using a threshold of 20 which gave 4 clusters was best. Having 5 clusters didn't lead to a significant enough decrease in total SSE.
- DBSCAN and K-Means both yielded almost the exact same clusters, except DBSCAN is able to detect and remove outliers while our K-Means did not. So that means that our K-Means had included the outliers in a cluster which would increase the cluster's SSE causing it to be worse.

AccidentsSet01.csv

Best Algorithm: DBSCAN / K-Mean
Number of Clusters: 3
Average SSE of clusters: 29.7
Hyperparameters of best Algorithm: Epsilon: 3, MinPoints: 3 // K: 3
Observations:

- DBSCAN: The algorithm performed very well. Based on visual observation all of the axis when compared appeared to be clustered properly, there was one obvious outlier which was properly unclustered.

- K-Means: We noticed that the fourth cluster had little improvement on the total SSE, which is why we decided on a k of 3. In actuality using a k of 3 allows for the data to be clustered into two clusters and gives the outlier its own cluster.
- Hierarchical Clustering: At a threshold of 5.5 we have 4 clusters and a considerable decrease in total SSE. Interestingly we had two different nodes with the exact same height of 5.0. And cutting the dendrogram at 5.0 created 6 clusters, which we found was too many. Since we couldn't cut the tree to create 5 clusters and look at the difference in total SSE, we have to use 5.5 as our threshold.

AccidentsSet03.csv

Best Algorithm: DBSCAN
Number of Clusters: 4
Average SSE of clusters: 22.2
Hyperparameters of best Algorithm: Epsilon: 1, MinPoints: 5
Observations:

- DBSCAN: On this dataset the algorithm did a fair job grouping data into four clusters with only 8% outliers. Given the 5 dimensional data it was difficult to verify proper clustering, mean SSE for all clusters was used as the main measure for accuracy. (22.1)
- K-Means: We found that the 6th cluster is the first cluster that doesn't significantly improve the total SSE. However after doing some investigating of some planes in the data it seems that 5 clusters was overclustering so we decided a k of 4 was best. (122)
- Hierarchical Clustering: At a threshold of 3.5 we split the dendrogram into 4 clusters which didn't relate to a significant decrease in total SSE, therefore we decided to use a threshold of 5 which split the data into 3 clusters.

Iris.csv

Best Algorithm: K-Means
Number of Clusters: 3
Average SSE of clusters: 26.3
Hyperparameters of best Algorithm: K: 3
Observations:

- DBSCAN: On this dataset the algorithm struggles. Due to the relative density of the data the algorithm would constantly bounce between excessively merging clusters or leaving too many outliers. Lots of time was spent in attempts to tune the hyperparameters, eventually the given fact that there should be three clusters was honored above the 'performance' of the algorithm. As a result 50% of the data was found to be an outlier. Mean cluster SSE was 4.83. Additional tuning might have lended better results, but at this point it was obvious this is not the ideal algorithm for this dataset.
- K-Means: In the iris dataset there is one distinct cluster, and then two clusters that have overlap. This is really well shown with the by k-means, because the most significant decrease in total SSE is from 1 to 2 clusters. While adding a 3rd cluster shows significant improvement in SSE, it isn't as significant as it should be knowing that the

dataset is built of three clusters. But since we know the iris dataset has three clusters of breeds we would specify a k of 3.

- Hierarchical Clustering: We observed that using a threshold of 4 creates 3 clusters which gives us the last additional cluster with a significant decrease in total SSE. Interestingly this was our only model that showed that the iris dataset should have 3 clusters.

Mammal_milk.csv

Best Algorithm: DBSCAN
Number of Clusters: 4
Average SSE of clusters: 28.5
Hyperparameters of best Algorithm: Epsilon: 5, MinPoints: 3
Observations:

- DBSCAN: It was difficult to tune ideal hyperparameters for this dataset as there were no obvious dense clusters. Due to the fairly homogeneous distribution of data the algorithm had some large SSE values on clusters, cluster sizes were very inconsistent, and there was 12% of the remaining data as outliers. There were four clusters generated. SSE Mean: 28.5
- K-Means: Adding the 5th cluster was the first cluster that didn't show significant improvement in the total SSE. Since it was hard to visualize the data we decided to use a k of 4 over 3.
- Hierarchical Clustering: Looking at the total SSE dropping the threshold to 20 creates three clusters which is a significant reduction of total SSE compared to two clusters. And dropping the threshold to 16 which produces 4 clusters doesn't significantly drop the total SSE so we decided on using 20 as the threshold.

Many_clusters.csv

Best Algorithm: DBSCAN
Number of Clusters: 6
Average SSE of clusters: 199.1
Hyperparameters of best Algorithm: Epsilon: 6, MinPoints: 6
Observations:

- DBSCAN: This is the ideal dataset for this algorithm, with the important note that not every datapoint is classified. DBSCAN did an excellent job at capturing all the dense point groups. Even the less dense groups were well identified with some tuning of the epsilon parameter. Overall six groups were identified with 15% of the data remaining as outliers (these were mostly points scattered sparsely between the other clusters).
- K-Means: Because this data doesn't have clear clusters it is harder to determine k from looking at the total SSE as the improvements aren't as proportionally large compared to other datasets. That being said, we noticed that the 7th cluster is the first cluster that doesn't significantly improve the total SSE, so we decided on using k = 6.
- Hierarchical Clustering: Using a threshold of 28 split the dendrogram into 5 clusters, which was the last significant decrease in total SSE. Interestingly this is different from

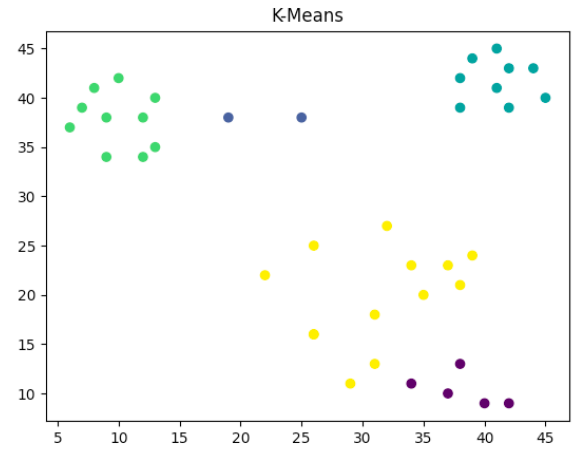
the models and as can be seen during our visual section this wasn't the best method of clustering compared to the other observed outputs.

- Since DBSCAN was able to look past all the outliers and find the dense groups it did the best at clustering. If we wanted to cluster all the points into groups K-Means did a better job.

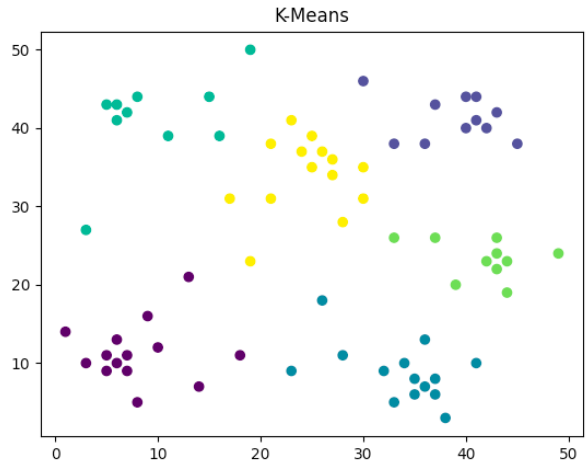
Visualization

K-MEANS

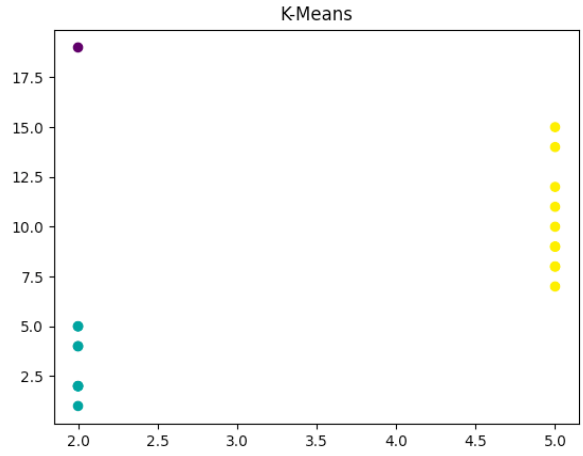
4clusters.csv

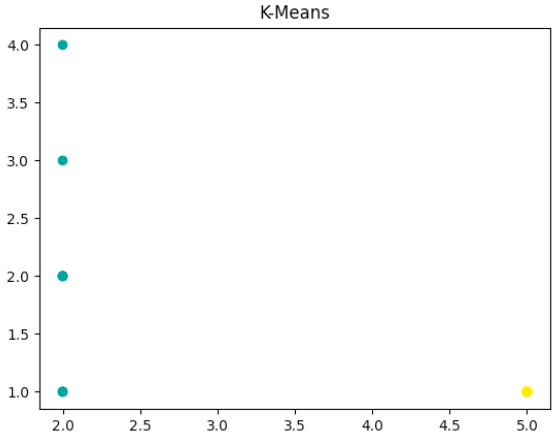


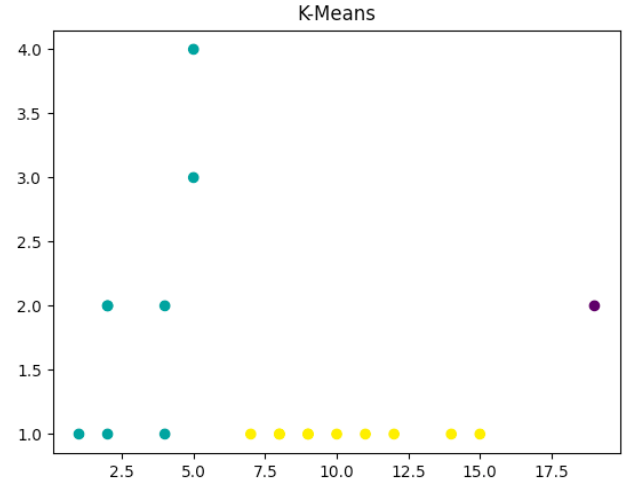
many_clusters.csv



AccidentsSet01.csv (images comparing in order: axis 1,2, axis 1,3, axis 2,3)

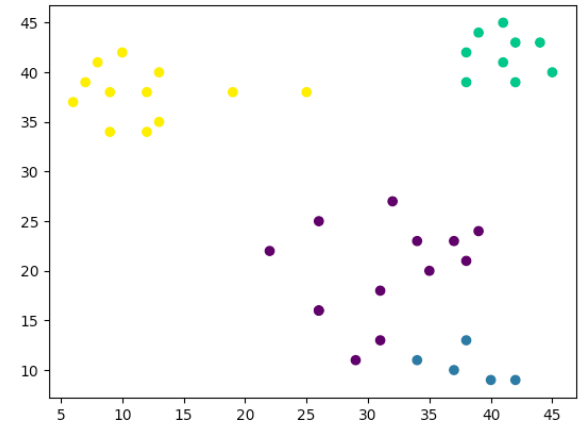




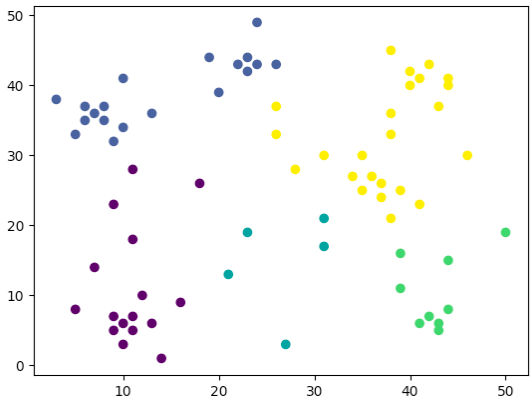


AGGLOMERATIVE HIERARCHICAL CLUSTERING

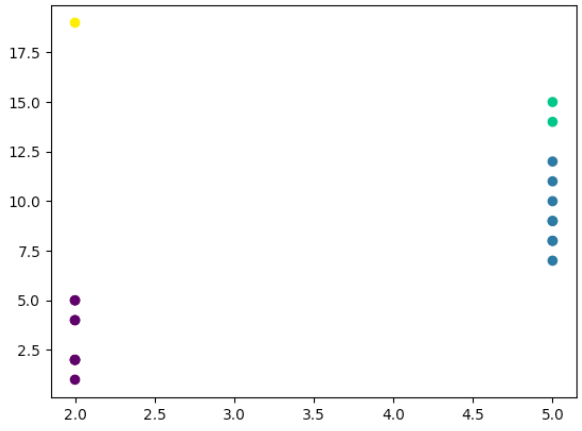
4clusters.csv

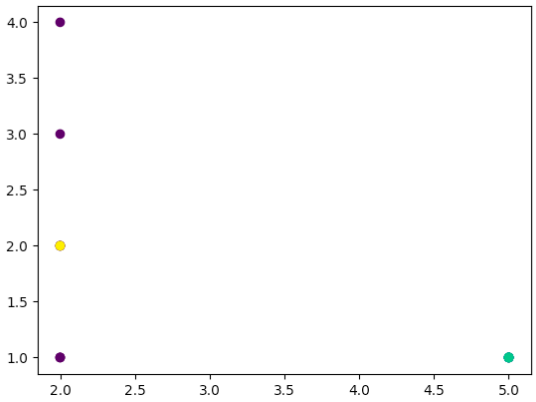


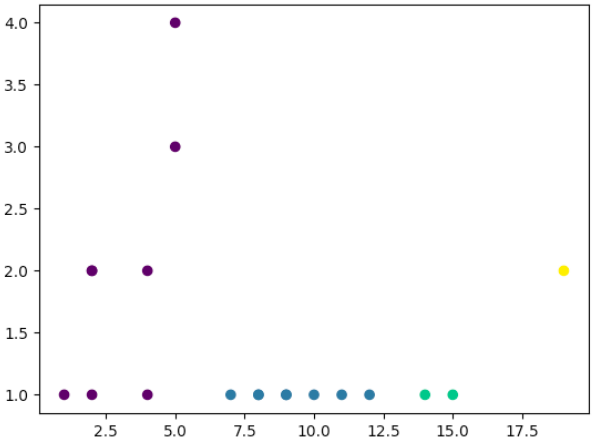
many_clusters.csv



AccidentsSet01.csv (images comparing in order: axis 1,2, axis 1,3, axis 2,3)

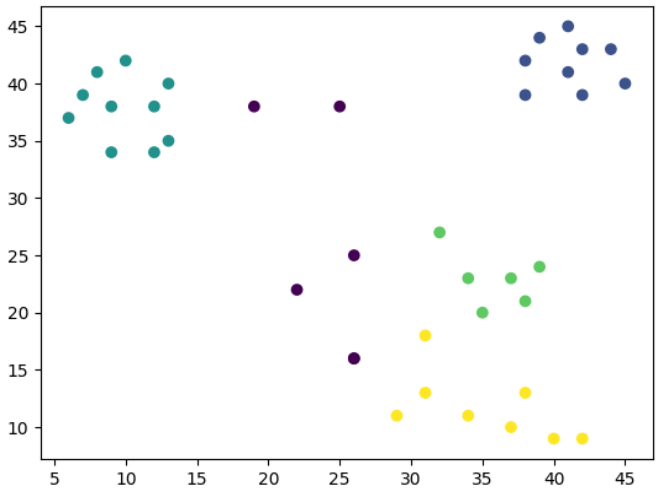






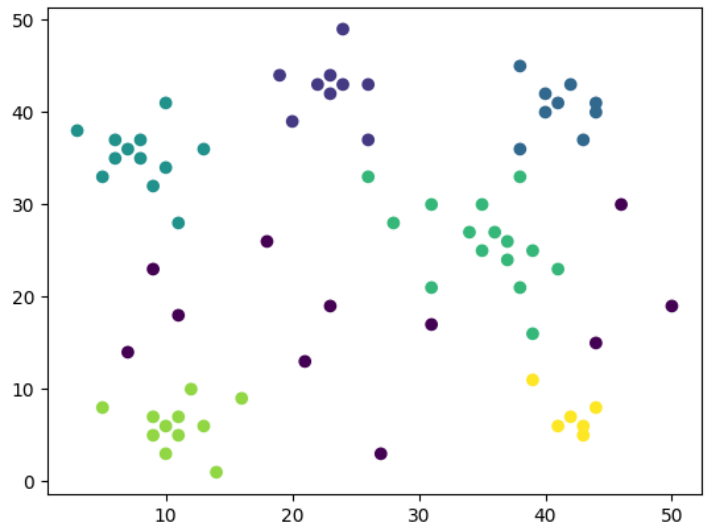
DBSCAN

4clusters.csv

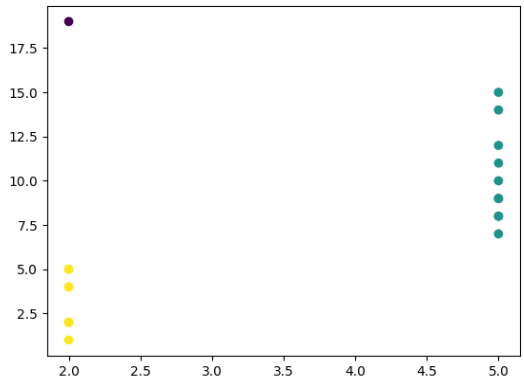


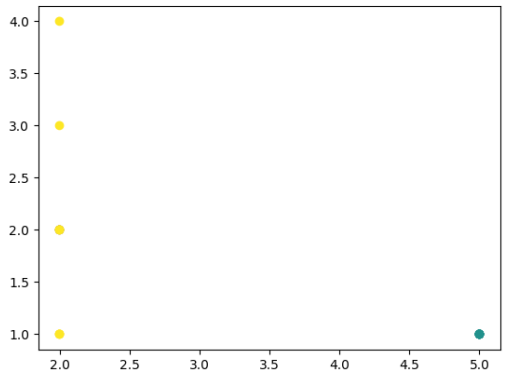
many_clusters.csv

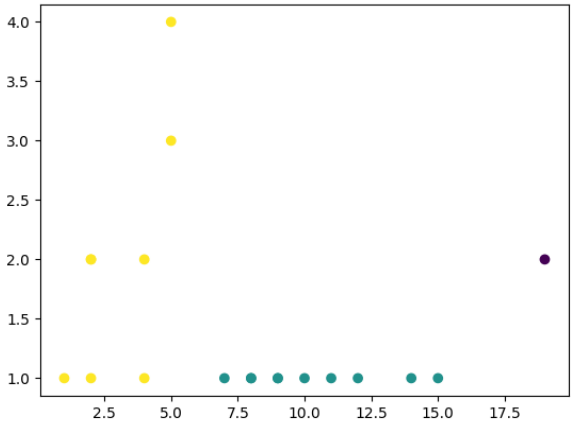
(NOTE: top center is a cluster but not colored very well)



AccidentsSet01.csv (images comparing in order: axis 1,2, axis 1,3, axis 2,3)







Analysis

K-Means was the most consistent and robust of the clustering methods. While DBSCAN was arguably the best given our evaluation parameter, K-Means most of the time built the same clusters but included the outliers which made its average SSE larger. For any of the data distributions that DBSCAN didn't work on, K-Means was then our best model in that case. If we had made it so that it handled outliers it may have been our strongest clustering algorithm.

Agglomerative Hierarchical Clustering seemed to be our worst method. This could be due to us choosing complete linkage to calculate cluster distance, and in the future when using this algorithm we will try the other distance implementations.

DBSCAN was a strong algorithm when used with datasets with clear clusters such as 4_clusters and many_clusters. The algorithm does poorly with uniformly distributed data and data where clusters have drastically different densities. However, it does show lots of success when there are outliers in the data. It is the only method of our three that can actually deal with outliers by ignoring them, where all the others take the SSE hit by clustering them.

Overall each algorithm has strengths and weaknesses which reduce the effectiveness of each algorithm individually. As the lab report mentioned, different approaches can be used for distance calculations and outlier detection, leading to different results using the same base algorithm. But in certain datasets some algorithms just don't have the ability to consistently perform well. This highlights the importance of running a variety of algorithms when cluster in order to create a holistic overview of the true clusters.

Another observation we made during the comparisons of the different algorithms is just how difficult it can be. For the two and three dimension datasets, visual observation of clustering was the fastest and most relevant method. This unfortunately did not do us any good for a dataset with more than three dimensions. For those datasets the evaluation itself becomes much more subjective, we decided to compare based on Mean SSE for the clusters, but different choices for 'best' algorithm can vary based on evaluation criteria.

Appendix

Output From DBSCAN

4clusters.csv

```
Epsilon: 5, MinPoints: 4 - Created 4 clusters.
------------------------------------
----- DBSCAN CLUSTERING OUTPUT -----
------------------------------------
------------ Cluster: 1 ------------
Center: [41.11 41.78]
Max Distance to Center: 4.28
Min Distance to Center: 0.79
Avg Distance to Center: 2.91
Sum of Squared Errors: 86.44
9 Points:
[41 45]
[39 44]
[42 43]
[44 43]
[38 42]
[41 41]
[45 40]
[38 39]
[42 39]
------------------------------------
------------ Cluster: 2 ------------
Center: [9.90 37.80]
Max Distance to Center: 4.34
Min Distance to Center: 0.92
Avg Distance to Center: 3.43
Sum of Squared Errors: 128.5
10 Points:
[10 42]
[ 8 41]
[13 40]
[ 7 39]
[ 9 38]
[12 38]
[ 6 37]
[13 35]
[ 9 34]
[12 34]
------------------------------------
------------ Cluster: 3 ------------
Center: [35.83 23.00]
Max Distance to Center: 5.54
Min Distance to Center: 1.17
Avg Distance to Center: 2.99
Sum of Squared Errors: 64.83
6 Points:
[32 27]
[39 24]
[34 23]
[37 23]
[38 21]
[35 20]
------------------------------------
------------ Cluster: 4 ------------
Center: [35.25 11.75]
Max Distance to Center: 7.56
Min Distance to Center: 1.46
Avg Distance to Center: 4.75
Sum of Squared Errors: 217.0
8 Points:
[31 18]
[31 13]
[38 13]
[29 11]
[34 11]
[37 10]
[40  9]
[42  9]
------------------------------------
-------- OUTLIER STATISTICS --------
Percent of data as outliers: 15.0%
Total number of outliers: 6
[19 38]
[25 38]
[26 25]
[22 22]
[26 16]
[26 16]
```

AccidentsSet01.csv

```
Epsilon: 3, MinPoints: 3 - Created 2 clusters.
------------------------------------
----- DBSCAN CLUSTERING OUTPUT -----
------------------------------------
------------ Cluster: 1 ------------
```

Center: [5.00 10.30 1.00]
Max Distance to Center: 4.7
Min Distance to Center: 0.3
Avg Distance to Center: 2.16
Sum of Squared Errors: 64.1
10 Points:
[5 15 1]
[5 14 1]
[5 12 1]
[5 11 1]
[5 10 1]
[5 9 1]
[5 9 1]
[5 8 1]
[5 8 1]
[5 7 1]

---------- Cluster: 2 ----------
Center: [2.00 3.12 2.00]
Max Distance to Center: 2.74
Min Distance to Center: 0.88
Avg Distance to Center: 1.65
Sum of Squared Errors: 24.88
8 Points:
[2 1 1]
[2 2 2]
[2 2 2]
[2 2 1]
[2 4 2]
[2 4 1]
[2 5 3]
[2 5 4]

-------- OUTLIER STATISTICS --------
Percent of data as outliers: 5.0%
Total number of outliers: 1
[2 19 2]

AccidentsSet03.csv
Epsilon: 1, MinPoints: 5 - Created 4 clusters.

----- DBSCAN CLUSTERING OUTPUT -----

---------- Cluster: 1 ----------
Center: [3.40 0.20 2.00 1.00 0.20]
Max Distance to Center: 0.92
Min Distance to Center: 0.49
Avg Distance to Center: 0.73
Sum of Squared Errors: 2.8
5 Points:
[3.00 0.00 2.00 1.00 0.00]
[3.00 0.00 2.00 1.00 1.00]
[4.00 0.00 2.00 1.00 0.00]
[3.00 1.00 2.00 1.00 0.00]
[4.00 0.00 2.00 1.00 0.00]

---------- Cluster: 2 ----------
Center: [3.38 0.00 4.00 1.00 0.50]
Max Distance to Center: 1.7
Min Distance to Center: 0.62
Avg Distance to Center: 0.9
Sum of Squared Errors: 7.88
8 Points:
[5.00 0.00 4.00 1.00 0.00]
[4.00 0.00 4.00 1.00 0.00]
[3.00 0.00 4.00 1.00 1.00]
[3.00 0.00 4.00 1.00 0.00]
[3.00 0.00 4.00 1.00 0.00]
[3.00 0.00 4.00 1.00 1.00]
[3.00 0.00 4.00 1.00 0.00]
[3.00 0.00 4.00 1.00 2.00]

---------- Cluster: 3 ----------
Center: [1.27 0.36 2.00 1.09 0.82]
Max Distance to Center: 1.85
Min Distance to Center: 0.5
Avg Distance to Center: 0.88
Sum of Squared Errors: 20.55
22 Points:
[2.00 0.00 2.00 1.00 0.00]
[2.00 0.00 2.00 1.00 0.00]
[3.00 1.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 0.00]
[1.00 1.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]
[1.00 0.00 2.00 2.00 2.00]
[2.00 1.00 2.00 1.00 2.00]
[1.00 0.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]
[2.00 1.00 2.00 1.00 1.00]
[1.00 0.00 2.00 2.00 1.00]
[1.00 1.00 2.00 1.00 1.00]
[1.00 0.00 2.00 1.00 1.00]

[1.00 0.00 2.00 1.00 1.00]
[1.00 1.00 2.00 1.00 0.00]
[1.00 1.00 2.00 1.00 0.00]
[1.00 1.00 2.00 1.00 0.00]

---------- Cluster: 4 ----------
Center: [1.00 1.00 4.00 1.00 0.73]
Max Distance to Center: 1.62
Min Distance to Center: 0.27
Avg Distance to Center: 0.65
Sum of Squared Errors: 12.36
22 Points:
[1.00 0.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 2.00]
[1.00 1.00 4.00 1.00 2.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 2.00 4.00 1.00 2.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 1.00]
[1.00 1.00 4.00 1.00 0.00]
[1.00 1.00 4.00 1.00 1.00]

-------- OUTLIER STATISTICS --------
Percent of data as outliers: 8.0%
Total number of outliers: 5
[5.00 0.00 4.00 1.00 1.00]
[10.00 0.00 4.00 1.00 1.00]
[1.00 0.00 4.00 2.00 0.00]
[2.00 0.00 4.00 2.00 0.00]
[1.00 0.00 4.00 3.00 3.00]

Iris.csv

Epsilon: 0.5, MinPoints: 13 - Created 3 clusters.

----- DBSCAN CLUSTERING OUTPUT -----

---------- Cluster: 1 ----------
Center: [4.97 3.39 1.47 0.24]
Max Distance to Center: 0.88
Min Distance to Center: 0.06
Avg Distance to Center: 0.43
Sum of Squared Errors: 10.25
46 Points:
[5.10 3.50 1.40 0.20]
[4.90 3.00 1.40 0.20]
[4.70 3.20 1.30 0.20]
[4.60 3.10 1.50 0.20]
[5.00 3.60 1.40 0.20]
[5.40 3.90 1.70 0.40]
[4.60 3.40 1.40 0.30]
[5.00 3.40 1.50 0.20]
[4.40 2.90 1.40 0.20]
[4.90 3.10 1.50 0.10]
[5.40 3.70 1.50 0.20]
[4.80 3.40 1.60 0.20]
[4.80 3.00 1.40 0.10]
[4.30 3.00 1.10 0.10]
[5.40 3.90 1.30 0.40]
[5.10 3.50 1.40 0.30]
[5.70 3.80 1.70 0.30]
[5.10 3.80 1.50 0.30]
[5.40 3.40 1.70 0.20]
[5.10 3.70 1.50 0.40]
[4.60 3.60 1.00 0.20]
[5.10 3.30 1.70 0.50]
[4.80 3.40 1.90 0.20]
[5.00 3.00 1.60 0.20]
[5.00 3.40 1.60 0.40]
[5.20 3.50 1.50 0.20]
[5.20 3.40 1.40 0.20]
[4.70 3.20 1.60 0.20]
[4.80 3.10 1.60 0.20]
[5.40 3.40 1.50 0.40]
[5.20 4.10 1.50 0.10]
[4.90 3.10 1.50 0.10]
[5.00 3.20 1.20 0.20]
[5.50 3.50 1.30 0.20]
[4.90 3.10 1.50 0.10]
[4.40 3.00 1.30 0.20]
[5.10 3.40 1.50 0.20]
[5.00 3.50 1.30 0.30]
[4.40 3.20 1.30 0.20]

[5.00 3.50 1.60 0.60]
[5.10 3.80 1.90 0.40]
[4.80 3.00 1.40 0.30]
[5.10 3.80 1.60 0.20]
[4.60 3.20 1.40 0.20]
[5.30 3.70 1.50 0.20]
[5.00 3.30 1.40 0.20]

---------- Cluster: 2 ----------
Center: [6.20 2.90 4.75 1.53]
Max Distance to Center: 0.5
Min Distance to Center: 0.17
Avg Distance to Center: 0.37
Sum of Squared Errors: 1.84
13 Points:
[6.40 3.20 4.50 1.50]
[6.50 2.80 4.60 1.50]
[6.30 3.30 4.70 1.60]
[6.10 2.90 4.70 1.40]
[6.30 2.50 4.90 1.50]
[6.10 2.80 4.70 1.20]
[6.00 2.70 5.10 1.60]
[6.10 3.00 4.60 1.40]
[6.20 2.90 4.30 1.30]
[6.20 2.80 4.80 1.80]
[6.10 3.00 4.90 1.80]
[6.30 2.80 5.10 1.50]
[6.00 3.00 4.80 1.80]

---------- Cluster: 3 ----------
Center: [5.69 2.77 4.18 1.29]
Max Distance to Center: 0.54
Min Distance to Center: 0.09
Avg Distance to Center: 0.34
Sum of Squared Errors: 2.4
18 Points:
[5.50 2.30 4.00 1.30]
[5.70 2.80 4.50 1.30]
[5.90 3.00 4.20 1.50]
[5.60 3.00 4.50 1.50]
[5.80 2.70 4.10 1.00]
[5.60 2.50 3.90 1.10]
[6.10 2.80 4.00 1.30]
[6.00 2.90 4.50 1.50]
[5.80 2.70 3.90 1.20]
[5.40 3.00 4.50 1.50]
[5.60 3.00 4.10 1.30]
[5.50 2.50 4.00 1.30]
[5.50 2.60 4.40 1.20]
[5.80 2.60 4.00 1.20]
[5.60 2.70 4.20 1.30]
[5.70 3.00 4.20 1.20]
[5.70 2.90 4.20 1.30]
[5.70 2.80 4.10 1.30]

-------- OUTLIER STATISTICS --------
Percent of data as outliers: 49.0%
Total number of outliers: 73
[5.80 4.00 1.20 0.20]
[5.70 4.40 1.50 0.40]
[5.50 4.20 1.40 0.20]
[4.50 2.30 1.30 0.30]
[7.00 3.20 4.70 1.40]
[6.90 3.10 4.90 1.50]
[4.90 2.40 3.30 1.00]
[6.60 2.90 4.60 1.30]
[5.20 2.70 3.90 1.40]
[5.00 2.00 3.50 1.00]
[6.00 2.20 4.00 1.00]
[5.60 2.90 3.60 1.30]
[6.70 3.10 4.40 1.40]
[6.20 2.20 4.50 1.50]
[5.90 3.20 4.80 1.80]
[6.40 2.90 4.30 1.30]
[6.60 3.00 4.40 1.40]
[6.80 2.80 4.80 1.40]
[6.70 3.00 5.00 1.70]
[5.70 2.60 3.50 1.00]
[5.50 2.40 3.80 1.10]
[5.50 2.40 3.70 1.00]
[6.00 3.40 4.50 1.60]
[6.70 3.10 4.70 1.50]
[6.30 2.30 4.40 1.30]
[5.00 2.30 3.30 1.00]
[5.10 2.50 3.00 1.10]
[6.30 3.30 6.00 2.50]
[5.80 2.70 5.10 1.90]
[7.10 3.00 5.90 2.10]
[6.30 2.90 5.60 1.80]
[6.50 3.00 5.80 2.20]
[7.60 3.00 6.60 2.10]
[4.90 2.50 4.50 1.70]
[7.30 2.90 6.30 1.80]
[6.70 2.50 5.80 1.80]
[7.20 3.60 6.10 2.50]
[6.50 3.20 5.10 2.00]
[6.40 2.70 5.30 1.90]
[6.80 3.00 5.50 2.10]

[5.70 2.50 5.00 2.00]
[5.80 2.80 5.10 2.40]
[6.40 3.20 5.30 2.30]
[6.50 3.00 5.50 1.80]
[7.70 3.80 6.70 2.20]
[7.70 2.60 6.90 2.30]
[6.00 2.20 5.00 1.50]
[6.90 3.20 5.70 2.30]
[5.60 2.80 4.90 2.00]
[7.70 2.80 6.70 2.00]
[6.30 2.70 4.90 1.80]
[6.70 3.30 5.70 2.10]
[7.20 3.20 6.00 1.80]
[6.40 2.80 5.60 2.10]
[7.20 3.00 5.80 1.60]
[7.40 2.80 6.10 1.90]
[7.90 3.80 6.40 2.00]
[6.40 2.80 5.60 2.20]
[6.10 2.60 5.60 1.40]
[7.70 3.00 6.10 2.30]
[6.30 3.40 5.60 2.40]
[6.40 3.10 5.50 1.80]
[6.90 3.10 5.40 2.10]
[6.70 3.10 5.60 2.40]
[6.90 3.10 5.10 2.30]
[5.80 2.70 5.10 1.90]
[6.80 3.20 5.90 2.30]
[6.70 3.30 5.70 2.50]
[6.70 3.00 5.20 2.30]
[6.30 2.50 5.00 1.90]
[6.50 3.00 5.20 2.00]
[6.20 3.40 5.40 2.30]
[5.90 3.00 5.10 1.80]

mammal_milk.csv
Epsilon: 5, MinPoints: 3 - Created 4 clusters.

----- DBSCAN CLUSTERING OUTPUT -----

---------- Cluster: 1 ----------
Center: [88.50 2.57 2.80 5.68 0.49]
Max Distance to Center: 3.49
Min Distance to Center: 0.88
Avg Distance to Center: 2.31
Sum of Squared Errors: 59.41
10 Points:
[90.10 2.60 1.00 6.90 0.35]
[88.50 1.40 3.50 6.00 0.24]
[88.40 2.20 2.70 6.40 0.18]
[90.30 1.70 1.40 6.20 0.40]
[90.40 0.60 4.50 4.40 0.10]
[87.70 3.50 3.40 4.80 0.71]
[86.90 4.80 1.70 5.70 0.90]
[86.50 3.90 3.20 5.60 0.80]
[90.00 2.00 1.80 5.50 0.47]
[86.20 3.00 4.80 5.30 0.70]

---------- Cluster: 2 ----------
Center: [82.00 7.12 6.47 4.18 0.89]
Max Distance to Center: 3.03
Min Distance to Center: 1.12
Avg Distance to Center: 1.84
Sum of Squared Errors: 22.43
6 Points:
[82.10 5.90 7.90 4.70 0.78]
[81.90 7.40 7.20 2.70 0.85]
[81.60 10.10 6.30 4.40 0.75]
[81.60 6.60 5.90 4.90 0.93]
[82.80 7.10 5.10 3.70 1.10]
[82.00 5.60 6.40 4.70 0.91]

---------- Cluster: 3 ----------
Center: [73.37 10.27 11.73 2.73 1.63]
Max Distance to Center: 3.85
Min Distance to Center: 1.74
Avg Distance to Center: 2.99
Sum of Squared Errors: 29.21
3 Points:
[76.30 9.30 9.50 3.00 1.20]
[71.30 12.30 13.10 1.90 2.30]
[72.50 9.20 12.60 3.30 1.40]

---------- Cluster: 4 ----------
Center: [65.17 10.73 20.40 2.23 1.22]
Max Distance to Center: 1.2
Min Distance to Center: 0.5
Avg Distance to Center: 0.95
Sum of Squared Errors: 3.0
3 Points:
[65.90 10.40 19.70 2.60 1.40]
[64.80 10.70 20.30 2.50 1.40]
[64.80 11.10 21.20 1.60 0.85]

-------- OUTLIER STATISTICS --------
Percent of data as outliers: 12.0%
Total number of outliers: 3

[70.70 3.60 17.60 5.60 0.63]
[46.40 9.70 42.00 0.00 0.85]
[44.90 10.60 34.90 0.90 0.53]

Many_clusters.csv
Epsilon: 6, MinPoints: 6 - Created 6 clusters.

----- DBSCAN CLUSTERING OUTPUT -----

---------- Cluster: 1 ----------
Center: [23.00 42.67]
Max Distance to Center: 6.41
Min Distance to Center: 0.67
Avg Distance to Center: 3.21
Sum of Squared Errors: 136.0
9 Points:
[24 49]
[19 44]
[23 44]
[22 43]
[24 43]
[26 43]
[23 42]
[20 39]
[26 37]

---------- Cluster: 2 ----------
Center: [41.11 40.56]
Max Distance to Center: 5.52
Min Distance to Center: 0.46
Avg Distance to Center: 3.0
Sum of Squared Errors: 105.11
9 Points:
[38 45]
[42 43]
[40 42]
[41 41]
[44 41]
[40 40]
[44 40]
[43 37]
[38 36]

---------- Cluster: 3 ----------
Center: [8.00 35.17]
Max Distance to Center: 7.77
Min Distance to Center: 0.17
Avg Distance to Center: 3.51
Sum of Squared Errors: 203.67
12 Points:
[10 41]
[3 38]
[6 37]
[8 37]
[7 36]
[13 36]
[6 35]
[8 35]
[10 34]
[5 33]
[9 32]
[11 28]

---------- Cluster: 4 ----------
Center: [35.00 25.93]
Max Distance to Center: 11.44
Min Distance to Center: 0.93
Avg Distance to Center: 5.23
Sum of Squared Errors: 558.93
15 Points:
[26 33]
[38 33]
[31 30]
[35 30]
[28 28]
[34 27]
[36 27]
[37 26]
[35 25]
[39 25]
[37 24]
[41 23]
[31 21]
[38 21]
[39 16]

---------- Cluster: 5 ----------
Center: [10.91 6.09]
Max Distance to Center: 6.21
Min Distance to Center: 0.91
Avg Distance to Center: 3.15
Sum of Squared Errors: 151.82
11 Points:
[12 10]
[16 9]

[5 8]
[9 7]
[11 7]
[10 6]
[13 6]
[9 5]
[11 5]
[10 3]
[14 1]

---------- Cluster: 6 ----------
Center: [42.00 7.17]
Max Distance to Center: 4.87
Min Distance to Center: 0.17
Avg Distance to Center: 2.11
Sum of Squared Errors: 38.83
6 Points:
[39 11]
[44 8]
[42 7]
[41 6]
[43 6]
[43 5]

-------- OUTLIER STATISTICS --------
Percent of data as outliers: 15.0%
Total number of outliers: 11
[46 30]
[18 26]
[9 23]
[23 19]
[50 19]
[11 18]
[31 17]
[44 15]
[7 14]
[21 13]
[27 3]

Output From K-MEANS

4clusters.csv

Created 5 clusters.

----- KMEANS CLUSTERING OUTPUT -----

---------- Cluster: 2 ----------
Center: [41.111111111111114, 41.77777777777778]
Max Distance to Center: 4.28
Min Distance to Center: 0.79
Avg Distance to Center: 2.91
Sum of Squared Errors: 86.44
9 Points:
[41, 45]
[39, 44]
[42, 43]
[44, 43]
[38, 42]
[41, 41]
[45, 40]
[38, 39]
[42, 39]

---------- Cluster: 3 ----------
Center: [9.9, 37.8]
Max Distance to Center: 4.34
Min Distance to Center: 0.92
Avg Distance to Center: 3.43
Sum of Squared Errors: 128.5
10 Points:
[10, 42]
[8, 41]
[13, 40]
[7, 39]
[9, 38]
[12, 38]
[6, 37]
[13, 35]
[9, 34]
[12, 34]

---------- Cluster: 1 ----------
Center: [22.0, 38.0]
Max Distance to Center: 3.0
Min Distance to Center: 3.0
Avg Distance to Center: 3.0
Sum of Squared Errors: 18.0
2 Points:
[19, 38]
[25, 38]

---------- Cluster: 4 ----------
Center: [31.23076923076923, 19.923076923076923]
Max Distance to Center: 9.46
Min Distance to Center: 1.94
Avg Distance to Center: 6.54
Sum of Squared Errors: 613.23
13 Points:
[32, 27]
[26, 25]
[39, 24]
[34, 23]
[37, 23]
[22, 22]
[38, 21]
[35, 20]
[31, 18]
[26, 16]
[31, 13]
[26, 16]
[29, 11]

---------- Cluster: 0 ----------
Center: [38.2, 10.4]
Max Distance to Center: 4.24
Min Distance to Center: 1.26
Avg Distance to Center: 2.89
Sum of Squared Errors: 48.0
5 Points:
[38, 13]
[34, 11]
[37, 10]
[40, 9]
[42, 9]

AccidentsSet01.csv

Created 3 clusters.

----- KMEANS CLUSTERING OUTPUT -----

---------- Cluster: 2 ----------
Center: [5.0, 10.3, 1.0]
Max Distance to Center: 4.7

Min Distance to Center: 0.3
Avg Distance to Center: 2.16
Sum of Squared Errors: 64.1
10 Points:
[5, 15, 1]
[5, 14, 1]
[5, 12, 1]
[5, 11, 1]
[5, 10, 1]
[5, 9, 1]
[5, 9, 1]
[5, 8, 1]
[5, 8, 1]
[5, 7, 1]

---------- Cluster: 0 ----------
Center: [2.0, 19.0, 2.0]
Max Distance to Center: 0.0
Min Distance to Center: 0.0
Avg Distance to Center: 0.0
Sum of Squared Errors: 0.0
1 Points:
[2, 19, 2]

---------- Cluster: 1 ----------
Center: [2.0, 3.125, 2.0]
Max Distance to Center: 2.74
Min Distance to Center: 0.88
Avg Distance to Center: 1.65
Sum of Squared Errors: 24.88
8 Points:
[2, 1, 1]
[2, 2, 2]
[2, 2, 2]
[2, 2, 1]
[2, 4, 2]
[2, 4, 1]
[2, 5, 3]
[2, 5, 4]

AccidentsSet03.csv
Created 4 clusters.

----- KMEANS CLUSTERING OUTPUT -----

---------- Cluster: 2 ----------
Center: [2.1714285714285713, 0.2571428571428571, 2.5142857142857142, 1.0285714285714285, 0.6285714285714286]
Max Distance to Center: 3.27
Min Distance to Center: 0.87
Avg Distance to Center: 1.62
Sum of Squared Errors: 101.54
35 Points:
[3.0, 0.0, 2.0, 1.0, 0.0]
[5.0, 0.0, 4.0, 1.0, 1.0]
[5.0, 0.0, 4.0, 1.0, 0.0]
[2.0, 0.0, 2.0, 1.0, 0.0]
[3.0, 0.0, 2.0, 1.0, 1.0]
[4.0, 0.0, 4.0, 1.0, 0.0]
[4.0, 0.0, 2.0, 1.0, 0.0]
[2.0, 0.0, 2.0, 1.0, 0.0]
[3.0, 1.0, 2.0, 1.0, 0.0]
[3.0, 1.0, 2.0, 1.0, 1.0]
[4.0, 0.0, 2.0, 1.0, 0.0]
[3.0, 0.0, 4.0, 1.0, 1.0]
[3.0, 0.0, 4.0, 1.0, 0.0]
[1.0, 0.0, 2.0, 1.0, 0.0]
[3.0, 0.0, 4.0, 1.0, 0.0]
[3.0, 0.0, 4.0, 1.0, 1.0]
[3.0, 0.0, 4.0, 1.0, 0.0]
[3.0, 0.0, 4.0, 1.0, 2.0]
[1.0, 1.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[2.0, 1.0, 2.0, 1.0, 2.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[2.0, 1.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 2.0, 1.0]
[1.0, 1.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 1.0, 2.0, 1.0, 0.0]
[1.0, 1.0, 2.0, 1.0, 0.0]
[1.0, 1.0, 2.0, 1.0, 0.0]

---------- Cluster: 0 ----------
Center: [10.0, 0.0, 4.0, 1.0, 1.0]
Max Distance to Center: 0.0
Min Distance to Center: 0.0
Avg Distance to Center: 0.0

Sum of Squared Errors: 0.0
1 Points:
[10.0, 0.0, 4.0, 1.0, 1.0]

---------- Cluster: 3 ----------
Center: [1.0416666666666667, 0.9166666666666666, 4.0, 1.0833333333333333, 0.6666666666666666]
Max Distance to Center: 1.74
Min Distance to Center: 0.36
Avg Distance to Center: 0.75
Sum of Squared Errors: 17.96
24 Points:
[1.0, 0.0, 4.0, 2.0, 0.0]
[2.0, 0.0, 4.0, 2.0, 0.0]
[1.0, 0.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 2.0]
[1.0, 1.0, 4.0, 1.0, 2.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 2.0, 4.0, 1.0, 2.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 1.0]

---------- Cluster: 1 ----------
Center: [1.0, 0.0, 3.0, 2.5, 2.5]
Max Distance to Center: 1.22
Min Distance to Center: 1.22
Avg Distance to Center: 1.22
Sum of Squared Errors: 3.0
2 Points:
[1.0, 0.0, 4.0, 3.0, 3.0]
[1.0, 0.0, 2.0, 2.0, 2.0]

Iris.csv
Created 3 clusters.

----- KMEANS CLUSTERING OUTPUT -----

---------- Cluster: 1 ----------
Center: [5.006, 3.418, 1.464, 0.244]
Max Distance to Center: 1.24
Min Distance to Center: 0.06
Avg Distance to Center: 0.48
Sum of Squared Errors: 15.24
50 Points:
[5.1, 3.5, 1.4, 0.2]
[4.9, 3.0, 1.4, 0.2]
[4.7, 3.2, 1.3, 0.2]
[4.6, 3.1, 1.5, 0.2]
[5.0, 3.6, 1.4, 0.2]
[5.4, 3.9, 1.7, 0.4]
[4.6, 3.4, 1.4, 0.3]
[5.0, 3.4, 1.5, 0.2]
[4.4, 2.9, 1.4, 0.2]
[4.9, 3.1, 1.5, 0.1]
[5.4, 3.7, 1.5, 0.2]
[4.8, 3.4, 1.6, 0.2]
[4.8, 3.0, 1.4, 0.1]
[4.3, 3.0, 1.1, 0.1]
[5.8, 4.0, 1.2, 0.2]
[5.7, 4.4, 1.5, 0.4]
[5.4, 3.9, 1.3, 0.4]
[5.1, 3.5, 1.4, 0.3]
[5.7, 3.8, 1.7, 0.3]
[5.1, 3.8, 1.5, 0.3]
[5.4, 3.4, 1.7, 0.2]
[5.1, 3.7, 1.5, 0.4]
[4.6, 3.6, 1.0, 0.2]
[5.1, 3.3, 1.7, 0.5]
[4.8, 3.4, 1.9, 0.2]
[5.0, 3.0, 1.6, 0.2]
[5.0, 3.4, 1.6, 0.4]
[5.2, 3.5, 1.5, 0.2]
[5.2, 3.4, 1.4, 0.2]
[4.7, 3.2, 1.6, 0.2]
[4.8, 3.1, 1.6, 0.2]
[5.4, 3.4, 1.5, 0.4]
[5.2, 4.1, 1.5, 0.1]
[5.5, 4.2, 1.4, 0.2]
[4.9, 3.1, 1.5, 0.1]
[5.0, 3.2, 1.2, 0.2]

[5.5, 3.5, 1.3, 0.2]
[4.9, 3.1, 1.5, 0.1]
[4.4, 3.0, 1.3, 0.2]
[5.1, 3.4, 1.5, 0.2]
[5.0, 3.5, 1.3, 0.3]
[4.5, 2.3, 1.3, 0.3]
[4.4, 3.2, 1.3, 0.2]
[5.0, 3.5, 1.6, 0.6]
[5.1, 3.8, 1.9, 0.4]
[4.8, 3.0, 1.4, 0.3]
[5.1, 3.8, 1.6, 0.2]
[4.6, 3.2, 1.4, 0.2]
[5.3, 3.7, 1.5, 0.2]
[5.0, 3.3, 1.4, 0.2]

---------- Cluster: 0 ----------
Center: [6.853846153846154, 3.076923076923077, 5.715384615384615, 2.0538461538461537]
Max Distance to Center: 1.55
Min Distance to Center: 0.24
Avg Distance to Center: 0.73
Sum of Squared Errors: 25.41
39 Points:
[7.0, 3.2, 4.7, 1.4]
[6.9, 3.1, 4.9, 1.5]
[6.7, 3.0, 5.0, 1.7]
[6.3, 3.3, 6.0, 2.5]
[7.1, 3.0, 5.9, 2.1]
[6.3, 2.9, 5.6, 1.8]
[6.5, 3.0, 5.8, 2.2]
[7.6, 3.0, 6.6, 2.1]
[7.3, 2.9, 6.3, 1.8]
[6.7, 2.5, 5.8, 1.8]
[7.2, 3.6, 6.1, 2.5]
[6.5, 3.2, 5.1, 2.0]
[6.4, 2.7, 5.3, 1.9]
[6.8, 3.0, 5.5, 2.1]
[6.4, 3.2, 5.3, 2.3]
[6.5, 3.0, 5.5, 1.8]
[7.7, 3.8, 6.7, 2.2]
[7.7, 2.6, 6.9, 2.3]
[6.9, 3.2, 5.7, 2.3]
[7.7, 2.8, 6.7, 2.0]
[6.7, 3.3, 5.7, 2.1]
[7.2, 3.2, 6.0, 1.8]
[6.4, 2.8, 5.6, 2.1]
[7.2, 3.0, 5.8, 1.6]
[7.4, 2.8, 6.1, 1.9]
[7.9, 3.8, 6.4, 2.0]
[6.4, 2.8, 5.6, 2.2]
[6.1, 2.6, 5.6, 1.4]
[7.7, 3.0, 6.1, 2.3]
[6.3, 3.4, 5.6, 2.4]
[6.4, 3.1, 5.5, 1.8]
[6.9, 3.1, 5.4, 2.1]
[6.7, 3.1, 5.6, 2.4]
[6.9, 3.1, 5.1, 2.3]
[6.8, 3.2, 5.9, 2.3]
[6.7, 3.3, 5.7, 2.5]
[6.7, 3.0, 5.2, 2.3]
[6.5, 3.0, 5.2, 2.0]
[6.2, 3.4, 5.4, 2.3]

---------- Cluster: 2 ----------
Center: [5.88360655737705, 2.7409836065573767, 4.388524590163935, 1.4344262295081966]
Max Distance to Center: 1.65
Min Distance to Center: 0.24
Avg Distance to Center: 0.73
Sum of Squared Errors: 38.29
61 Points:
[6.4, 3.2, 4.5, 1.5]
[5.5, 2.3, 4.0, 1.3]
[6.5, 2.8, 4.6, 1.5]
[5.7, 2.8, 4.5, 1.3]
[6.3, 3.3, 4.7, 1.6]
[4.9, 2.4, 3.3, 1.0]
[6.6, 2.9, 4.6, 1.3]
[5.2, 2.7, 3.9, 1.4]
[5.0, 2.0, 3.5, 1.0]
[5.9, 3.0, 4.2, 1.5]
[6.0, 2.2, 4.0, 1.0]
[6.1, 2.9, 4.7, 1.4]
[5.6, 2.9, 3.6, 1.3]
[6.7, 3.1, 4.4, 1.4]
[5.6, 3.0, 4.5, 1.5]
[5.8, 2.7, 4.1, 1.0]
[6.2, 2.2, 4.5, 1.5]
[5.6, 2.5, 3.9, 1.1]
[5.9, 3.2, 4.8, 1.8]
[6.1, 2.8, 4.0, 1.3]
[6.3, 2.5, 4.9, 1.5]
[6.1, 2.8, 4.7, 1.2]
[6.4, 2.9, 4.3, 1.3]
[6.6, 3.0, 4.4, 1.4]
[6.8, 2.8, 4.8, 1.4]

[6.0, 2.9, 4.5, 1.5]
[5.7, 2.6, 3.5, 1.0]
[5.5, 2.4, 3.8, 1.1]
[5.5, 2.4, 3.7, 1.0]
[5.8, 2.7, 3.9, 1.2]
[6.0, 2.7, 5.1, 1.6]
[5.4, 3.0, 4.5, 1.5]
[6.0, 3.4, 4.5, 1.6]
[6.7, 3.1, 4.7, 1.5]
[6.3, 2.3, 4.4, 1.3]
[5.6, 3.0, 4.1, 1.3]
[5.5, 2.5, 4.0, 1.3]
[5.5, 2.6, 4.4, 1.2]
[6.1, 3.0, 4.6, 1.4]
[5.8, 2.6, 4.0, 1.2]
[5.0, 2.3, 3.3, 1.0]
[5.6, 2.7, 4.2, 1.3]
[5.7, 3.0, 4.2, 1.2]
[5.7, 2.9, 4.2, 1.3]
[6.2, 2.9, 4.3, 1.3]
[5.1, 2.5, 3.0, 1.1]
[5.7, 2.8, 4.1, 1.3]
[5.8, 2.7, 5.1, 1.9]
[4.9, 2.5, 4.5, 1.7]
[5.7, 2.5, 5.0, 2.0]
[5.8, 2.8, 5.1, 2.4]
[6.0, 2.2, 5.0, 1.5]
[5.6, 2.8, 4.9, 2.0]
[6.3, 2.7, 4.9, 1.8]
[6.2, 2.8, 4.8, 1.8]
[6.1, 3.0, 4.9, 1.8]
[6.3, 2.8, 5.1, 1.5]
[6.0, 3.0, 4.8, 1.8]
[5.8, 2.7, 5.1, 1.9]
[6.3, 2.5, 5.0, 1.9]
[5.9, 3.0, 5.1, 1.8]
---

mammal_milk.csv
Created 4 clusters.

----- KMEANS CLUSTERING OUTPUT -----

---------- Cluster: 3 ----------
Center: [88.5, 2.57, 2.8, 5.68, 0.48500000000000004]
Max Distance to Center: 3.49
Min Distance to Center: 0.88
Avg Distance to Center: 2.31
Sum of Squared Errors: 59.41
10 Points:
[90.1, 2.6, 1.0, 6.9, 0.35]
[88.5, 1.4, 3.5, 6.0, 0.24]
[88.4, 2.2, 2.7, 6.4, 0.18]
[90.3, 1.7, 1.4, 6.2, 0.4]
[90.4, 0.6, 4.5, 4.4, 0.1]
[87.7, 3.5, 3.4, 4.8, 0.71]
[86.9, 4.8, 1.7, 5.7, 0.9]
[86.5, 3.9, 3.2, 5.6, 0.8]
[90.0, 2.0, 1.8, 5.5, 0.47]
[86.2, 3.0, 4.8, 5.3, 0.7]

---------- Cluster: 1 ----------
Center: [81.18571428571428, 7.428571428571429, 6.9, 4.014285714285714, 0.9314285714285715]
Max Distance to Center: 5.94
Min Distance to Center: 1.53
Avg Distance to Center: 2.67
Sum of Squared Errors: 63.53
7 Points:
[82.1, 5.9, 7.9, 4.7, 0.78]
[81.9, 7.4, 7.2, 2.7, 0.85]
[81.6, 10.1, 6.3, 4.4, 0.75]
[81.6, 6.6, 5.9, 4.9, 0.93]
[82.8, 7.1, 5.1, 3.7, 1.1]
[82.0, 5.6, 6.4, 4.7, 0.91]
[76.3, 9.3, 9.5, 3.0, 1.2]

---------- Cluster: 2 ----------
Center: [68.33333333333333, 9.55, 17.416666666666668, 2.9166666666666665, 1.33]
Max Distance to Center: 6.98
Min Distance to Center: 3.46
Avg Distance to Center: 5.54
Sum of Squared Errors: 191.96
6 Points:
[70.7, 3.6, 17.6, 5.6, 0.63]
[71.3, 12.3, 13.1, 1.9, 2.3]
[72.5, 9.2, 12.6, 3.3, 1.4]
[65.9, 10.4, 19.7, 2.6, 1.4]
[64.8, 10.7, 20.3, 2.5, 1.4]
[64.8, 11.1, 21.2, 1.6, 0.85]

---------- Cluster: 0 ----------
Center: [45.65, 10.149999999999999, 38.45, 0.45, 0.69]
Max Distance to Center: 3.69
Min Distance to Center: 3.69
Avg Distance to Center: 3.69
Sum of Squared Errors: 27.19
2 Points:
[46.4, 9.7, 42.0, 0.0, 0.85]
[44.9, 10.6, 34.9, 0.9, 0.53]

Many_clusters.csv
Created 6 clusters.

----- KMEANS CLUSTERING OUTPUT -----

---------- Cluster: 4 ----------
Center: [23.3, 41.7]
Max Distance to Center: 9.11
Min Distance to Center: 0.42
Avg Distance to Center: 4.01
Sum of Squared Errors: 228.2
10 Points:
[24, 49]
[19, 44]
[23, 44]
[22, 43]
[24, 43]
[26, 43]
[23, 42]
[20, 39]
[26, 37]
[26, 33]

---------- Cluster: 1 ----------
Center: [41.27272727272727, 38.90909090909091]
Max Distance to Center: 10.09
Min Distance to Center: 1.68
Avg Distance to Center: 4.4
Sum of Squared Errors: 277.09
11 Points:
[38, 45]
[42, 43]
[40, 42]
[41, 41]
[44, 41]
[40, 40]
[44, 40]
[43, 37]
[38, 36]
[38, 33]
[46, 30]

---------- Cluster: 2 ----------
Center: [8.785714285714286, 33.642857142857146]
Max Distance to Center: 11.97
Min Distance to Center: 1.27
Avg Distance to Center: 5.03
Sum of Squared Errors: 495.57
14 Points:
[10, 41]
[3, 38]
[6, 37]
[8, 37]
[7, 36]
[13, 36]
[6, 35]
[8, 35]
[10, 34]
[5, 33]
[9, 32]
[11, 28]
[18, 26]
[9, 23]

---------- Cluster: 5 ----------
Center: [34.0, 24.5]
Max Distance to Center: 12.3
Min Distance to Center: 1.12
Avg Distance to Center: 5.32
Sum of Squared Errors: 499.5
14 Points:
[31, 30]
[35, 30]
[28, 28]
[34, 27]
[36, 27]
[37, 26]
[35, 25]
[39, 25]
[37, 24]
[41, 23]
[31, 21]

[38, 21]
[23, 19]
[31, 17]

---------- Cluster: 3 ----------
Center: [41.2, 9.6]
Max Distance to Center: 15.66
Min Distance to Center: 2.61
Avg Distance to Center: 6.25
Sum of Squared Errors: 572.0
10 Points:
[50, 19]
[39, 16]
[44, 15]
[39, 11]
[44, 8]
[42, 7]
[41, 6]
[43, 6]
[43, 5]
[27, 3]

---------- Cluster: 0 ----------
Center: [11.357142857142858, 8.0]
Max Distance to Center: 10.86
Min Distance to Center: 1.06
Avg Distance to Center: 4.97
Sum of Squared Errors: 467.21
14 Points:
[11, 18]
[7, 14]
[21, 13]
[12, 10]
[16, 9]
[5, 8]
[9, 7]
[11, 7]
[10, 6]
[13, 6]
[9, 5]
[11, 5]
[10, 3]
[14, 1]

Output From Hierarchical Clustering

4clusters.csv

Threshold: 20 - Created 4 Clusters

-- HIERARCHICAL CLUSTERING OUTPUT --

---------- Cluster: 0 ----------

Center: [31.23076923076923, 19.923076923076923]

Max Distance to Center: 9.46

Min Distance to Center: 1.94

Avg Distance to Center: 6.54

Sum of Squared Errors: 613.23

13 Points:

[26, 16]

[26, 16]

[31, 18]

[29, 11]

[31, 13]

[22, 22]

[26, 25]

[37, 23]

[39, 24]

[38, 21]

[35, 20]

[34, 23]

[32, 27]

---------- Cluster: 1 ----------

Center: [38.2, 10.4]

Max Distance to Center: 4.24

Min Distance to Center: 1.26

Avg Distance to Center: 2.89

Sum of Squared Errors: 48.0

5 Points:

[37, 10]

[38, 13]

[34, 11]

[42, 9]

[40, 9]

---------- Cluster: 2 ----------

Center: [41.111111111111114, 41.77777777777778]

Max Distance to Center: 4.28

Min Distance to Center: 0.79

Avg Distance to Center: 2.91

Sum of Squared Errors: 86.44

9 Points:

[38, 39]

[38, 42]

[39, 44]

[41, 45]

[42, 39]

[41, 41]

[45, 40]

[44, 43]

[42, 43]

---------- Cluster: 3 ----------

Center: [11.916666666666666, 37.833333333333336]

Max Distance to Center: 13.08

Min Distance to Center: 0.19

Avg Distance to Center: 4.84

Sum of Squared Errors: 390.58

12 Points:

[12, 34]

[13, 35]

[9, 34]

[9, 38]

[7, 39]

[6, 37]

[12, 38]

[13, 40]

[8, 41]

[10, 42]

[25, 38]

[19, 38]

AccidentsSet01.csv

Threshold: 5.5 - Created 4 clusters

-- HIERARCHICAL CLUSTERING OUTPUT --

---------- Cluster: 0 ----------

Center: [2.0, 3.125, 2.0]

Max Distance to Center: 2.74

Min Distance to Center: 0.88

Avg Distance to Center: 1.65

Sum of Squared Errors: 24.88

8 Points:

[2, 5, 4]

[2, 5, 3]

[2, 4, 1]
[2, 4, 2]
[2, 2, 1]
[2, 1, 1]
[2, 2, 2]
[2, 2, 2]

---------- Cluster: 1 ----------
Center: [5.0, 9.25, 1.0]
Max Distance to Center: 2.75
Min Distance to Center: 0.25
Avg Distance to Center: 1.31
Sum of Squared Errors: 19.5
8 Points:
[5, 9, 1]
[5, 9, 1]
[5, 10, 1]
[5, 11, 1]
[5, 12, 1]
[5, 8, 1]
[5, 8, 1]
[5, 7, 1]

---------- Cluster: 2 ----------
Center: [5.0, 14.5, 1.0]
Max Distance to Center: 0.5
Min Distance to Center: 0.5
Avg Distance to Center: 0.5
Sum of Squared Errors: 0.5
2 Points:
[5, 14, 1]
[5, 15, 1]

---------- Cluster: 3 ----------
Center: [2.0, 19.0, 2.0]
Max Distance to Center: 0.0
Min Distance to Center: 0.0
Avg Distance to Center: 0.0
Sum of Squared Errors: 0.0
1 Points:
[2, 19, 2]

AccidentsSet03.csv
Threshold: 3.5 - Created 4 clusters.

-- HIERARCHICAL CLUSTERING OUTPUT --

---------- Cluster: 0 ----------
Center: [1.119047619047619,
0.5952380952380952, 3.0,
1.0952380952380953, 0.6428571428571429]
Max Distance to Center: 2.01
Min Distance to Center: 1.15
Avg Distance to Center: 1.31
Sum of Squared Errors: 73.79
42 Points:
[2.0, 1.0, 2.0, 1.0, 1.0]
[2.0, 1.0, 2.0, 1.0, 2.0]
[1.0, 0.0, 2.0, 2.0, 1.0]
[1.0, 0.0, 2.0, 2.0, 2.0]
[2.0, 0.0, 2.0, 1.0, 0.0]
[2.0, 0.0, 2.0, 1.0, 0.0]
[1.0, 0.0, 2.0, 1.0, 0.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 0.0, 2.0, 1.0, 1.0]
[1.0, 1.0, 2.0, 1.0, 0.0]
[1.0, 1.0, 2.0, 1.0, 0.0]
[1.0, 1.0, 2.0, 1.0, 0.0]
[1.0, 1.0, 2.0, 1.0, 1.0]
[1.0, 1.0, 2.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 1.0]
[1.0, 0.0, 4.0, 1.0, 1.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]
[1.0, 1.0, 4.0, 1.0, 0.0]

[1.0, 1.0, 4.0, 1.0, 0.0]
[2.0, 0.0, 4.0, 2.0, 0.0]
[1.0, 0.0, 4.0, 2.0, 0.0]

---------- Cluster: 1 ----------
Center: [1.0, 1.0, 4.0, 1.5, 2.25]
Max Distance to Center: 1.95
Min Distance to Center: 0.56
Avg Distance to Center: 1.05
Sum of Squared Errors: 5.75
4 Points:
[1.0, 1.0, 4.0, 1.0, 2.0]
[1.0, 1.0, 4.0, 1.0, 2.0]
[1.0, 2.0, 4.0, 1.0, 2.0]
[1.0, 0.0, 4.0, 3.0, 3.0]

---------- Cluster: 2 ----------
Center: [3.466666666666667, 0.13333333333333333, 3.2, 1.0, 0.4666666666666667]
Max Distance to Center: 1.81
Min Distance to Center: 1.05
Avg Distance to Center: 1.37
Sum of Squared Errors: 29.6
15 Points:
[3.0, 0.0, 4.0, 1.0, 0.0]
[3.0, 0.0, 4.0, 1.0, 0.0]
[3.0, 0.0, 4.0, 1.0, 0.0]
[4.0, 0.0, 4.0, 1.0, 0.0]
[5.0, 0.0, 4.0, 1.0, 0.0]
[5.0, 0.0, 4.0, 1.0, 1.0]
[3.0, 0.0, 4.0, 1.0, 1.0]
[3.0, 0.0, 4.0, 1.0, 1.0]
[3.0, 0.0, 4.0, 1.0, 2.0]
[3.0, 0.0, 2.0, 1.0, 1.0]
[3.0, 0.0, 2.0, 1.0, 0.0]
[4.0, 0.0, 2.0, 1.0, 0.0]
[4.0, 0.0, 2.0, 1.0, 0.0]
[3.0, 1.0, 2.0, 1.0, 1.0]
[3.0, 1.0, 2.0, 1.0, 0.0]

---------- Cluster: 3 ----------
Center: [10.0, 0.0, 4.0, 1.0, 1.0]
Max Distance to Center: 0.0
Min Distance to Center: 0.0
Avg Distance to Center: 0.0
Sum of Squared Errors: 0.0
1 Points:
[10.0, 0.0, 4.0, 1.0, 1.0]

Iris.csv

Threshold: 4 - Created 3 clusters.

-- HIERARCHICAL CLUSTERING OUTPUT --

---------- Cluster: 0 ----------
Center: [5.006, 3.418, 1.4639999999999997, 0.244]
Max Distance to Center: 1.24
Min Distance to Center: 0.06
Avg Distance to Center: 0.48
Sum of Squared Errors: 15.24
50 Points:
[4.8, 3.4, 1.9, 0.2]
[4.8, 3.4, 1.6, 0.2]
[5.0, 3.4, 1.6, 0.4]
[5.1, 3.3, 1.7, 0.5]
[5.0, 3.5, 1.6, 0.6]
[5.1, 3.7, 1.5, 0.4]
[5.1, 3.8, 1.5, 0.3]
[5.1, 3.8, 1.6, 0.2]
[5.1, 3.8, 1.9, 0.4]
[5.3, 3.7, 1.5, 0.2]
[5.4, 3.7, 1.5, 0.2]
[5.5, 3.5, 1.3, 0.2]
[5.4, 3.4, 1.5, 0.4]
[5.4, 3.4, 1.7, 0.2]
[5.7, 3.8, 1.7, 0.3]
[5.4, 3.9, 1.7, 0.4]
[5.7, 4.4, 1.5, 0.4]
[5.8, 4.0, 1.2, 0.2]
[5.5, 4.2, 1.4, 0.2]
[5.2, 4.1, 1.5, 0.1]
[5.4, 3.9, 1.3, 0.4]
[4.4, 3.0, 1.3, 0.2]
[4.4, 2.9, 1.4, 0.2]
[4.3, 3.0, 1.1, 0.1]
[4.6, 3.2, 1.4, 0.2]
[4.6, 3.1, 1.5, 0.2]
[4.7, 3.2, 1.3, 0.2]
[4.4, 3.2, 1.3, 0.2]
[4.6, 3.4, 1.4, 0.3]
[4.6, 3.6, 1.0, 0.2]
[5.1, 3.5, 1.4, 0.3]

[5.1, 3.5, 1.4, 0.2]
[5.0, 3.5, 1.3, 0.3]
[5.0, 3.6, 1.4, 0.2]
[5.2, 3.4, 1.4, 0.2]
[5.2, 3.5, 1.5, 0.2]
[5.1, 3.4, 1.5, 0.2]
[5.0, 3.4, 1.5, 0.2]
[5.0, 3.3, 1.4, 0.2]
[5.0, 3.2, 1.2, 0.2]
[4.9, 3.1, 1.5, 0.1]
[4.9, 3.1, 1.5, 0.1]
[4.9, 3.1, 1.5, 0.1]
[4.8, 3.0, 1.4, 0.1]
[4.8, 3.0, 1.4, 0.3]
[4.9, 3.0, 1.4, 0.2]
[5.0, 3.0, 1.6, 0.2]
[4.8, 3.1, 1.6, 0.2]
[4.7, 3.2, 1.6, 0.2]
[4.5, 2.3, 1.3, 0.3]

---------- Cluster: 1 ----------
Center: [5.5321428571428575, 2.6357142857142857, 3.960714285714286, 1.2285714285714284]
Max Distance to Center: 1.07
Min Distance to Center: 0.16
Avg Distance to Center: 0.53
Sum of Squared Errors: 9.75
28 Points:
[5.5, 2.4, 3.7, 1.0]
[5.5, 2.4, 3.8, 1.1]
[5.6, 2.5, 3.9, 1.1]
[5.5, 2.5, 4.0, 1.3]
[5.5, 2.3, 4.0, 1.3]
[5.2, 2.7, 3.9, 1.4]
[5.7, 2.6, 3.5, 1.0]
[5.6, 2.9, 3.6, 1.3]
[6.0, 2.2, 4.0, 1.0]
[5.7, 2.9, 4.2, 1.3]
[5.7, 3.0, 4.2, 1.2]
[5.6, 3.0, 4.1, 1.3]
[5.7, 2.8, 4.1, 1.3]
[5.6, 2.7, 4.2, 1.3]
[5.8, 2.6, 4.0, 1.2]
[5.8, 2.7, 3.9, 1.2]
[5.8, 2.7, 4.1, 1.0]
[6.1, 2.8, 4.0, 1.3]
[5.9, 3.0, 4.2, 1.5]
[5.5, 2.6, 4.4, 1.2]
[5.7, 2.8, 4.5, 1.3]
[5.4, 3.0, 4.5, 1.5]
[5.6, 3.0, 4.5, 1.5]
[4.9, 2.5, 4.5, 1.7]
[5.0, 2.3, 3.3, 1.0]
[4.9, 2.4, 3.3, 1.0]
[5.1, 2.5, 3.0, 1.1]
[5.0, 2.0, 3.5, 1.0]

---------- Cluster: 2 ----------
Center: [6.545833333333333, 2.963888888888889, 5.273611111111112, 1.8499999999999999]
Max Distance to Center: 2.08
Min Distance to Center: 0.18
Avg Distance to Center: 0.86
Sum of Squared Errors: 64.62
72 Points:
[6.1, 3.0, 4.6, 1.4]
[6.1, 2.9, 4.7, 1.4]
[6.0, 2.9, 4.5, 1.5]
[6.1, 2.8, 4.7, 1.2]
[6.2, 2.9, 4.3, 1.3]
[6.4, 2.9, 4.3, 1.3]
[6.3, 3.3, 4.7, 1.6]
[6.4, 3.2, 4.5, 1.5]
[6.0, 3.4, 4.5, 1.6]
[6.9, 3.1, 4.9, 1.5]
[7.0, 3.2, 4.7, 1.4]
[6.7, 3.1, 4.7, 1.5]
[6.7, 3.0, 5.0, 1.7]
[6.6, 2.9, 4.6, 1.3]
[6.5, 2.8, 4.6, 1.5]
[6.8, 2.8, 4.8, 1.4]
[6.6, 3.0, 4.4, 1.4]
[6.7, 3.1, 4.4, 1.4]
[6.3, 2.3, 4.4, 1.3]
[6.2, 2.2, 4.5, 1.5]
[6.0, 2.2, 5.0, 1.5]
[6.3, 2.8, 5.1, 1.5]
[6.0, 2.7, 5.1, 1.6]
[6.3, 2.5, 4.9, 1.5]
[6.2, 2.8, 4.8, 1.8]
[6.3, 2.7, 4.9, 1.8]
[6.3, 2.5, 5.0, 1.9]
[6.4, 2.7, 5.3, 1.9]
[6.1, 2.6, 5.6, 1.4]
[5.8, 2.7, 5.1, 1.9]
[5.8, 2.7, 5.1, 1.9]

[5.7, 2.5, 5.0, 2.0]
[5.6, 2.8, 4.9, 2.0]
[5.8, 2.8, 5.1, 2.4]
[6.0, 3.0, 4.8, 1.8]
[6.1, 3.0, 4.9, 1.8]
[5.9, 3.2, 4.8, 1.8]
[5.9, 3.0, 5.1, 1.8]
[6.4, 2.8, 5.6, 2.2]
[6.4, 2.8, 5.6, 2.1]
[6.5, 3.0, 5.8, 2.2]
[6.4, 3.1, 5.5, 1.8]
[6.5, 3.0, 5.5, 1.8]
[6.3, 2.9, 5.6, 1.8]
[6.7, 2.5, 5.8, 1.8]
[6.7, 3.0, 5.2, 2.3]
[6.9, 3.1, 5.1, 2.3]
[6.9, 3.1, 5.4, 2.1]
[6.8, 3.0, 5.5, 2.1]
[6.5, 3.0, 5.2, 2.0]
[6.5, 3.2, 5.1, 2.0]
[6.4, 3.2, 5.3, 2.3]
[6.2, 3.4, 5.4, 2.3]
[6.3, 3.4, 5.6, 2.4]
[6.3, 3.3, 6.0, 2.5]
[6.8, 3.2, 5.9, 2.3]
[6.9, 3.2, 5.7, 2.3]
[6.7, 3.3, 5.7, 2.1]
[6.7, 3.3, 5.7, 2.5]
[6.7, 3.1, 5.6, 2.4]
[7.7, 3.0, 6.1, 2.3]
[7.2, 3.6, 6.1, 2.5]
[7.9, 3.8, 6.4, 2.0]
[7.7, 3.8, 6.7, 2.2]
[7.7, 2.8, 6.7, 2.0]
[7.6, 3.0, 6.6, 2.1]
[7.7, 2.6, 6.9, 2.3]
[7.2, 3.0, 5.8, 1.6]
[7.2, 3.2, 6.0, 1.8]
[7.1, 3.0, 5.9, 2.1]
[7.4, 2.8, 6.1, 1.9]
[7.3, 2.9, 6.3, 1.8]

mammal_milk.csv
Threshold: 20 - Created 3 clusters.

-- HIERARCHICAL CLUSTERING OUTPUT --

---------- Cluster: 0 ----------
Center: [69.47142857142858, 9.514285714285714, 16.285714285714285, 2.928571428571428, 1.3114285714285714]
Max Distance to Center: 9.63
Min Distance to Center: 4.8
Avg Distance to Center: 6.35
Sum of Squared Errors: 300.16
7 Points:
[64.8, 10.7, 20.3, 2.5, 1.4]
[65.9, 10.4, 19.7, 2.6, 1.4]
[64.8, 11.1, 21.2, 1.6, 0.85]
[70.7, 3.6, 17.6, 5.6, 0.63]
[72.5, 9.2, 12.6, 3.3, 1.4]
[71.3, 12.3, 13.1, 1.9, 2.3]
[76.3, 9.3, 9.5, 3.0, 1.2]

---------- Cluster: 1 ----------
Center: [86.0625, 4.2749999999999995, 4.175, 5.11875, 0.635625]
Max Distance to Center: 7.67
Min Distance to Center: 1.24
Avg Distance to Center: 4.51
Sum of Squared Errors: 377.22
16 Points:
[90.0, 2.0, 1.8, 5.5, 0.47]
[90.3, 1.7, 1.4, 6.2, 0.4]
[90.1, 2.6, 1.0, 6.9, 0.35]
[88.4, 2.2, 2.7, 6.4, 0.18]
[88.5, 1.4, 3.5, 6.0, 0.24]
[90.4, 0.6, 4.5, 4.4, 0.1]
[86.5, 3.9, 3.2, 5.6, 0.8]
[87.7, 3.5, 3.4, 4.8, 0.71]
[86.2, 3.0, 4.8, 5.3, 0.7]
[86.9, 4.8, 1.7, 5.7, 0.9]
[82.8, 7.1, 5.1, 3.7, 1.1]
[81.9, 7.4, 7.2, 2.7, 0.85]
[82.0, 5.6, 6.4, 4.7, 0.91]
[81.6, 6.6, 5.9, 4.9, 0.93]
[82.1, 5.9, 7.9, 4.7, 0.78]
[81.6, 10.1, 6.3, 4.4, 0.75]

---------- Cluster: 2 ----------
Center: [45.65, 10.149999999999999, 38.45, 0.45, 0.69]
Max Distance to Center: 3.69
Min Distance to Center: 3.69
Avg Distance to Center: 3.69
Sum of Squared Errors: 27.19

2 Points:
[44.9, 10.6, 34.9, 0.9, 0.53]
[46.4, 9.7, 42.0, 0.0, 0.85]

Many_clusters.csv
Threshold: 28 - Created 5 clusters.

-- HIERARCHICAL CLUSTERING OUTPUT --

---------- Cluster: 0 ----------
Center: [11.0, 11.0]
Max Distance to Center: 17.0
Min Distance to Center: 1.41
Avg Distance to Center: 7.56
Sum of Squared Errors: 1202.0
16 Points:
[9, 23]
[11, 28]
[18, 26]
[7, 14]
[11, 18]
[10, 6]
[9, 7]
[9, 5]
[10, 3]
[11, 5]
[11, 7]
[13, 6]
[14, 1]
[16, 9]
[12, 10]
[5, 8]

---------- Cluster: 1 ----------
Center: [14.0, 39.0]
Max Distance to Center: 14.14
Min Distance to Center: 3.16
Avg Distance to Center: 8.54
Sum of Squared Errors: 1520.0
19 Points:
[23, 42]
[24, 43]
[22, 43]
[23, 44]
[26, 43]
[24, 49]
[20, 39]
[19, 44]
[9, 32]
[10, 34]
[13, 36]
[10, 41]
[7, 36]
[6, 37]
[6, 35]
[8, 35]
[8, 37]
[5, 33]
[3, 38]

---------- Cluster: 2 ----------
Center: [26.6, 14.6]
Max Distance to Center: 11.61
Min Distance to Center: 5.01
Avg Distance to Center: 7.18
Sum of Squared Errors: 286.4
5 Points:
[21, 13]
[23, 19]
[31, 17]
[31, 21]
[27, 3]

---------- Cluster: 3 ----------
Center: [42.77777777777778,
10.333333333333334]
Max Distance to Center: 11.28
Min Distance to Center: 2.63
Avg Distance to Center: 5.24
Sum of Squared Errors: 299.56
9 Points:
[39, 11]
[39, 16]
[44, 15]
[41, 6]
[42, 7]
[43, 5]
[43, 6]
[44, 8]
[50, 19]

---------- Cluster: 4 ----------
Center: [37.375, 32.666666666666664]
Max Distance to Center: 12.35

Min Distance to Center: 0.71
Avg Distance to Center: 8.38
Sum of Squared Errors: 1886.96
24 Points:
[44, 40]
[44, 41]
[42, 43]
[41, 41]
[40, 42]
[40, 40]
[43, 37]
[38, 45]
[26, 33]
[26, 37]
[28, 28]
[31, 30]
[38, 33]
[38, 36]
[46, 30]
[35, 25]
[34, 27]
[37, 26]
[36, 27]
[35, 30]
[38, 21]
[41, 23]
[37, 24]
[39, 25]
