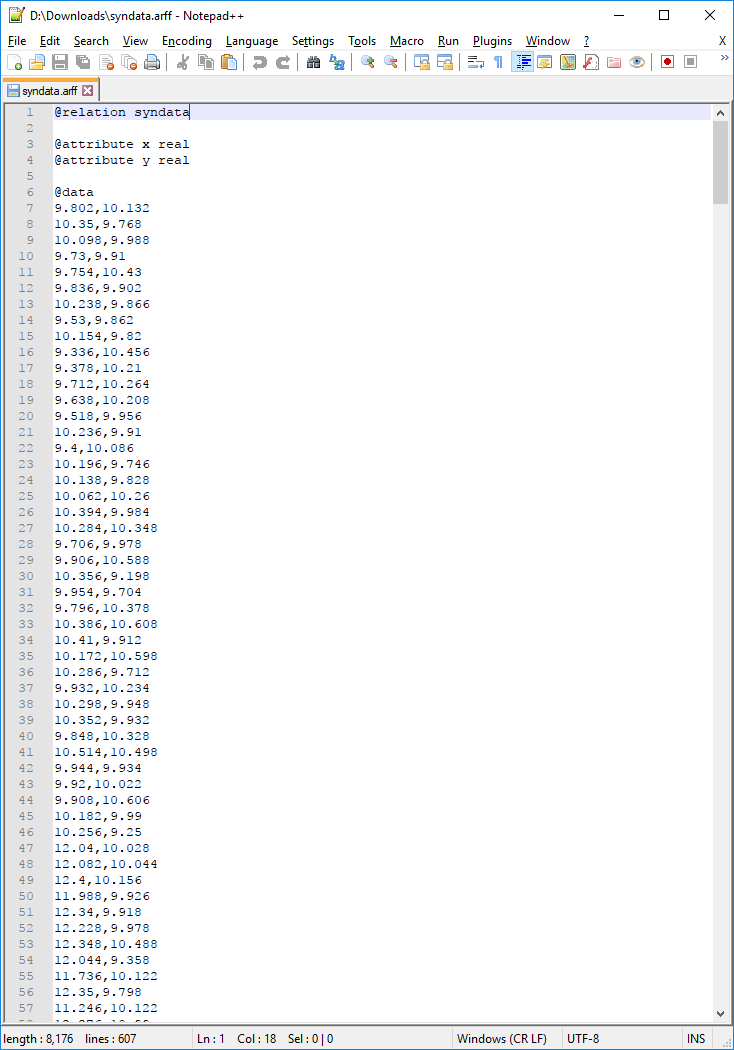
Patrick Austin  
CS 491: Data Mining

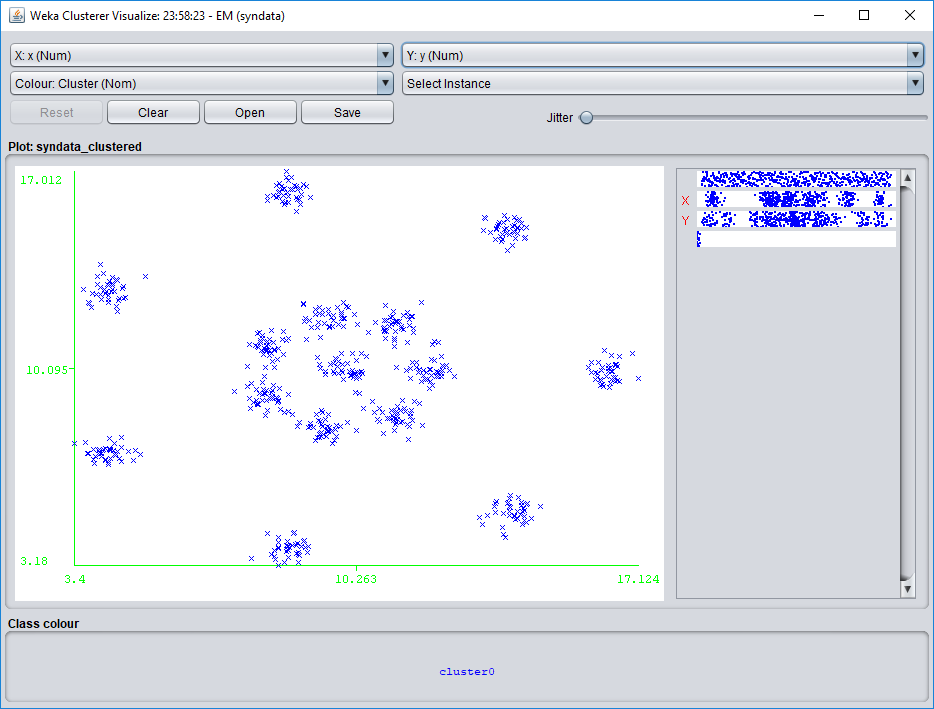
Homework # 4

11/22/2017

4-1 a. The full converted file synData1.arff is available on request. Here is a screenshot showing the format.



b. By visualizing the data, it becomes clear that there are 15 clusters, as shown in the plot below.



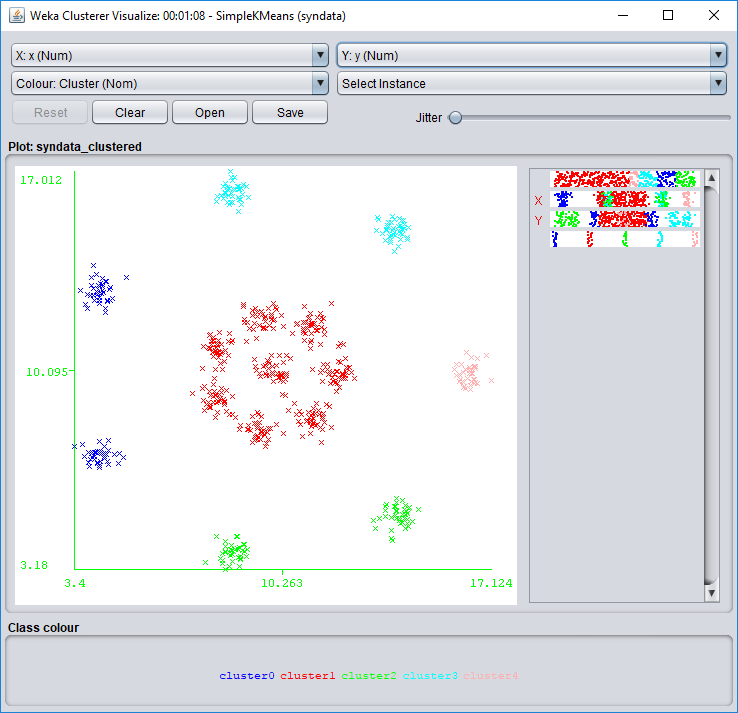
c.

For k = 5:

i. 10 iterations were needed.

ii. The within cluster SSE was approximately 16.64

iii. See below.

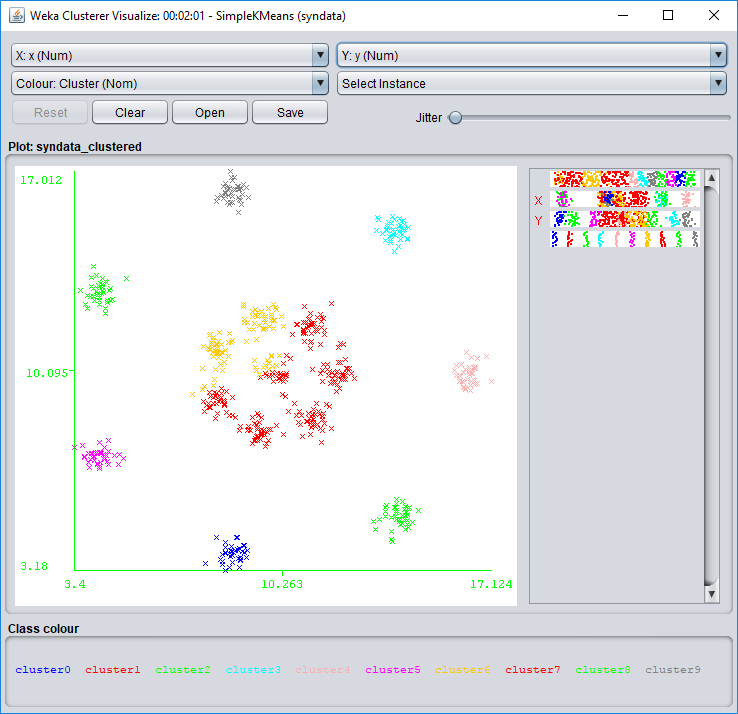


For k = 10:

i. 9 iterations were needed.

ii. The within cluster SSE was approximately 2.75

iii. See below.

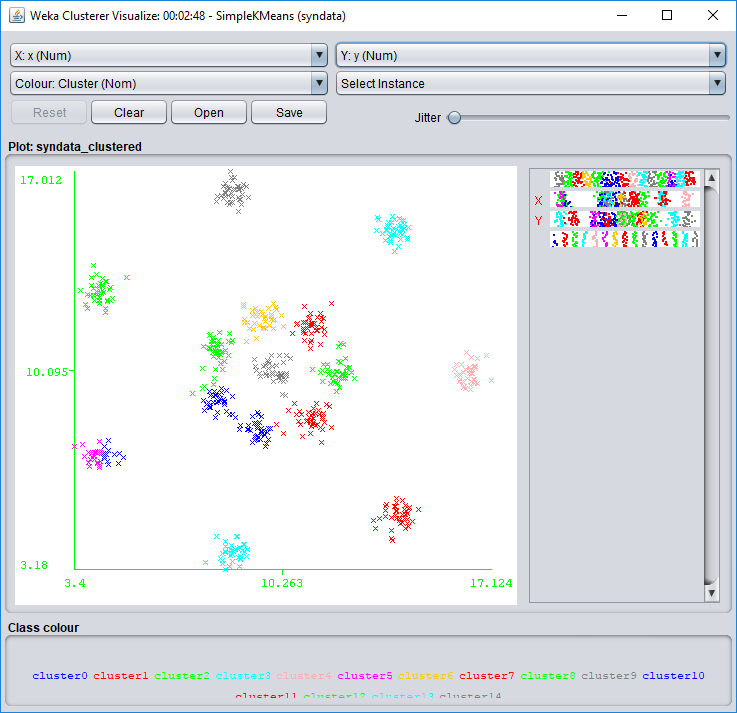


For k = 15:

i. 5 iterations were needed.

ii. The within cluster SSE was approximately 0.86

iii. See below.

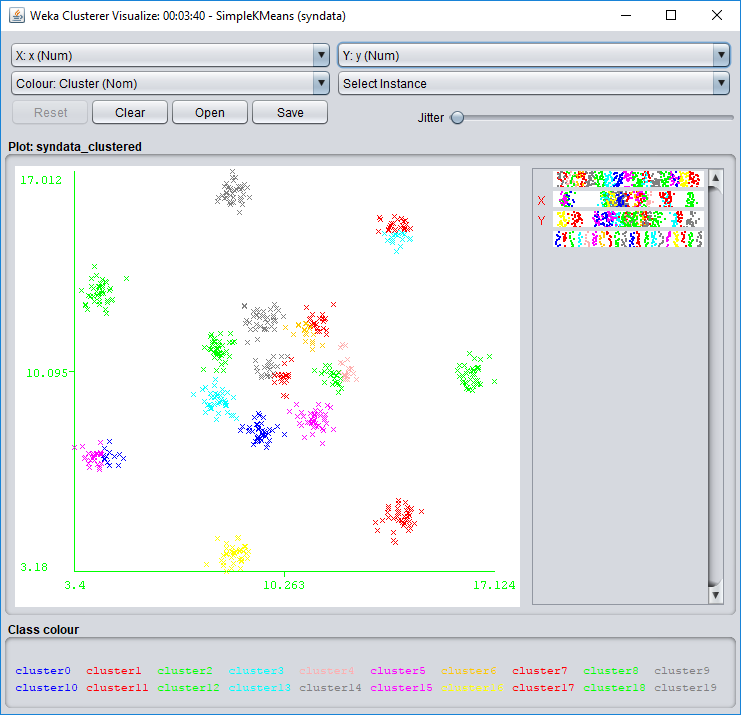


For k = 20:

i. 8 iterations were needed.

ii. The within cluster SSE was approximately 0.50

iii. See below.

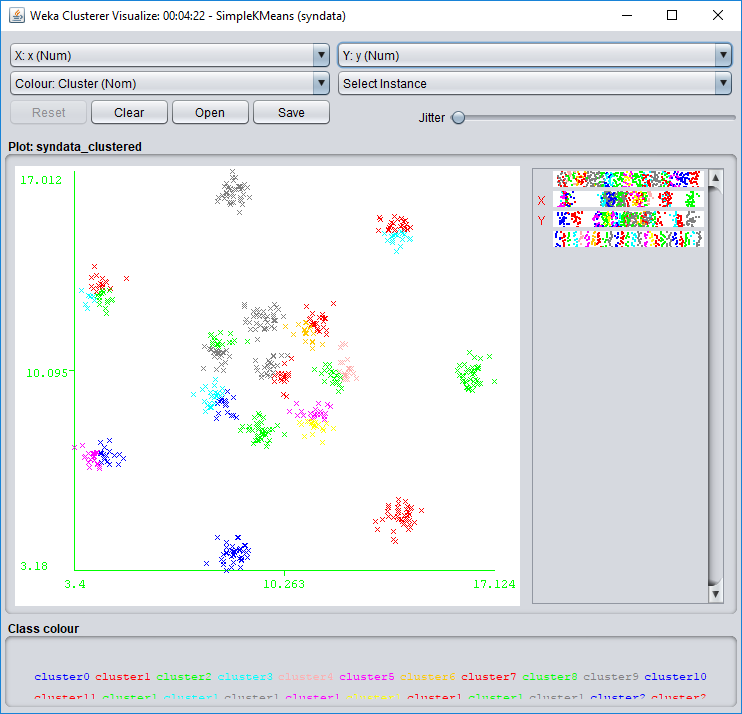


For k = 25:

i. 9 iterations were needed.

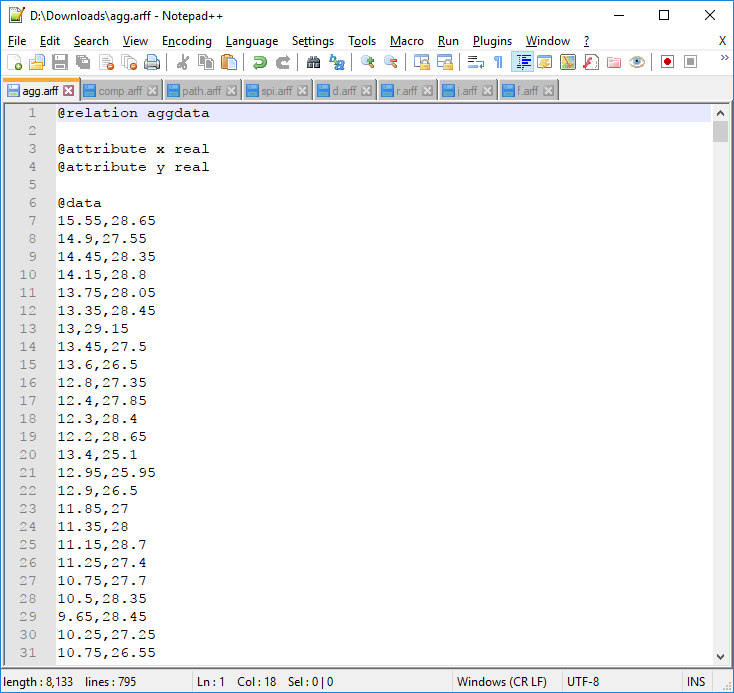
ii. The within cluster SSE was approximately 0.43

iii. See below.



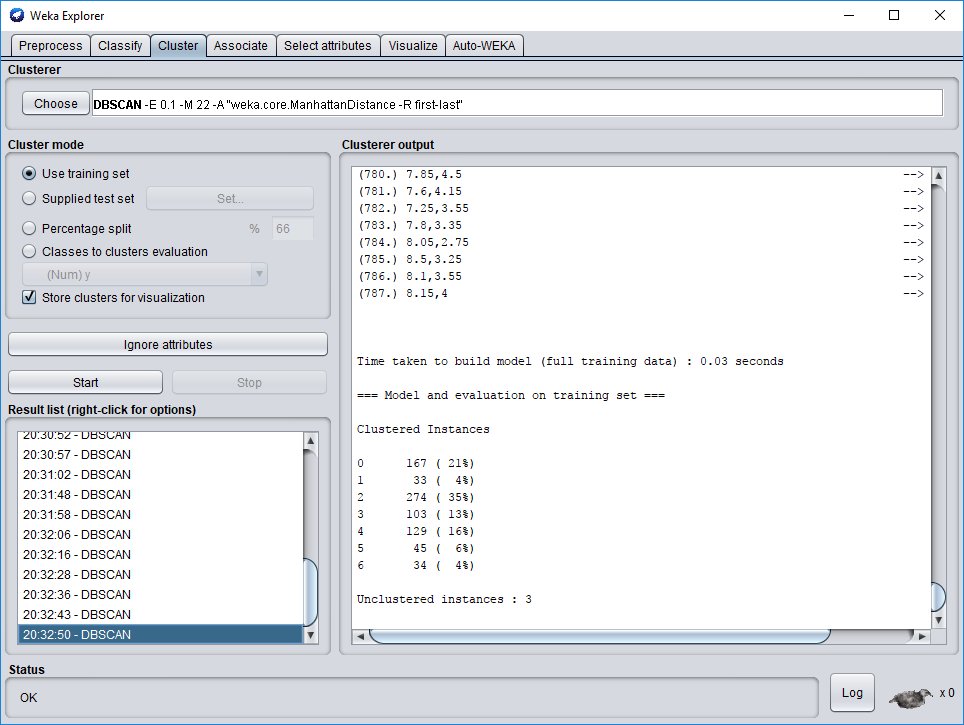
We note that over these iterations SSE became smaller as k became greater.

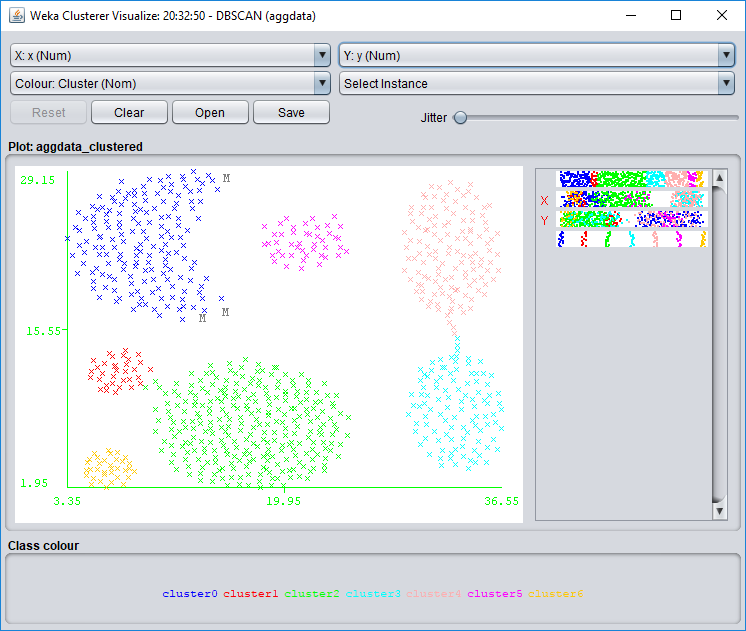
4-3 a. The full converted files are available on request. Here is a screenshot showing a sample of the format:



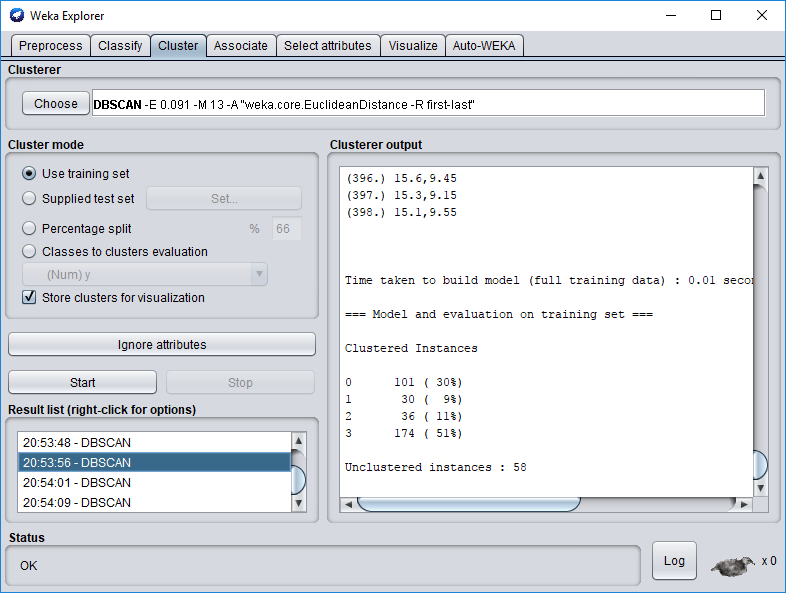
b.

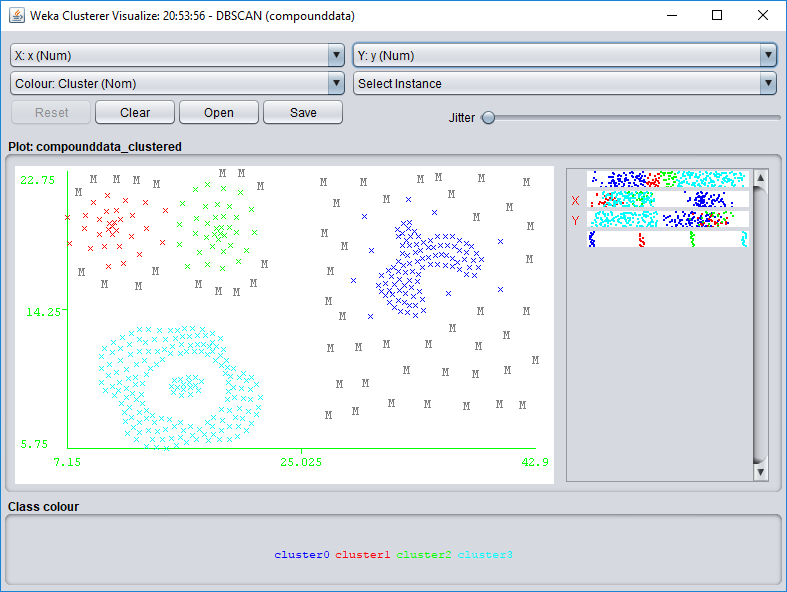
Aggregation: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. In this case the correct number of clusters could be found with only a few unclustered instances.



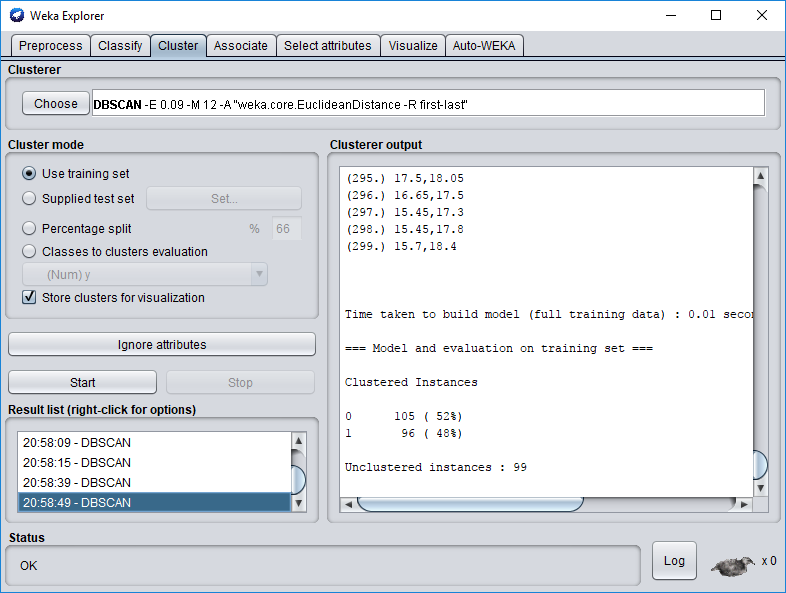


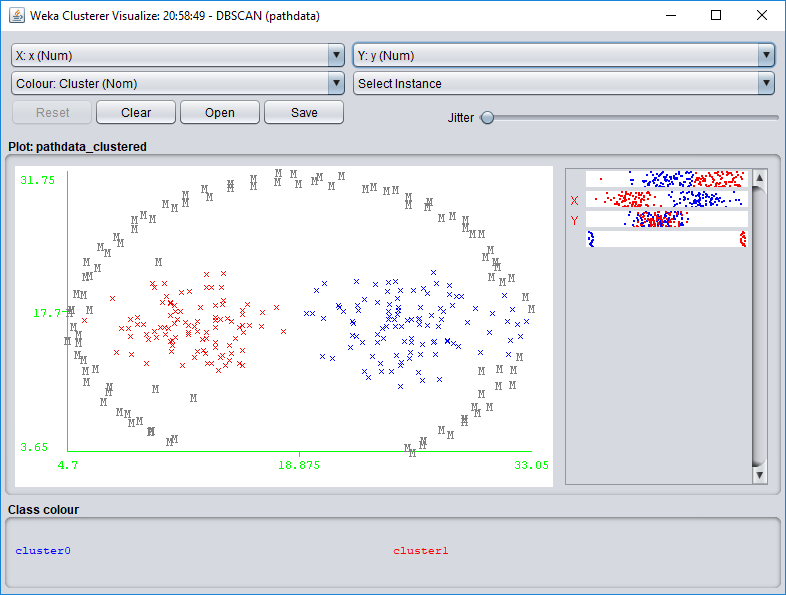
Compound: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. Note that in this case results were not very close.



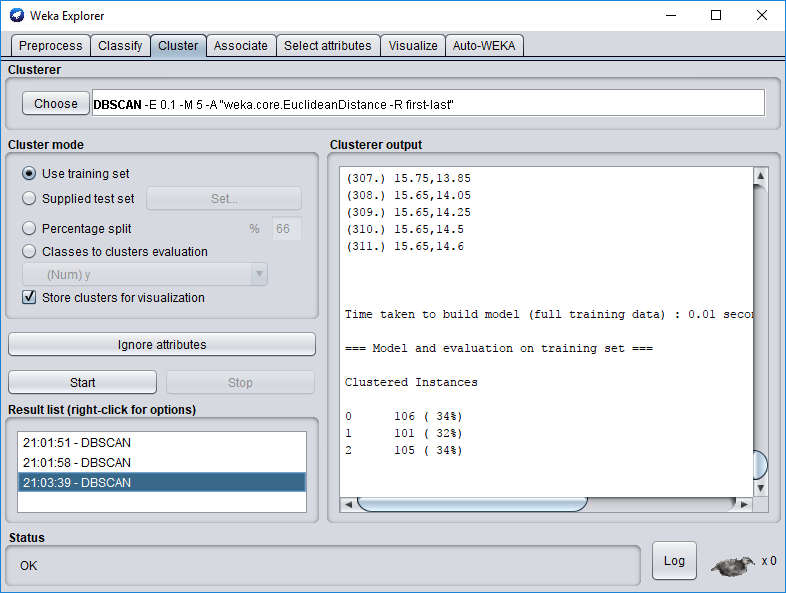


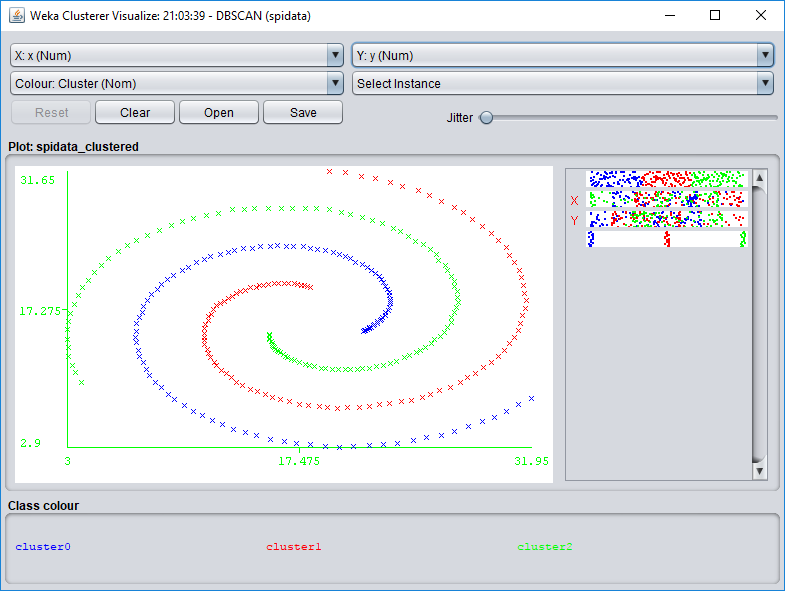
Pathbased: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. Note that in this case results were not very close, not capturing the exterior ring cluster.



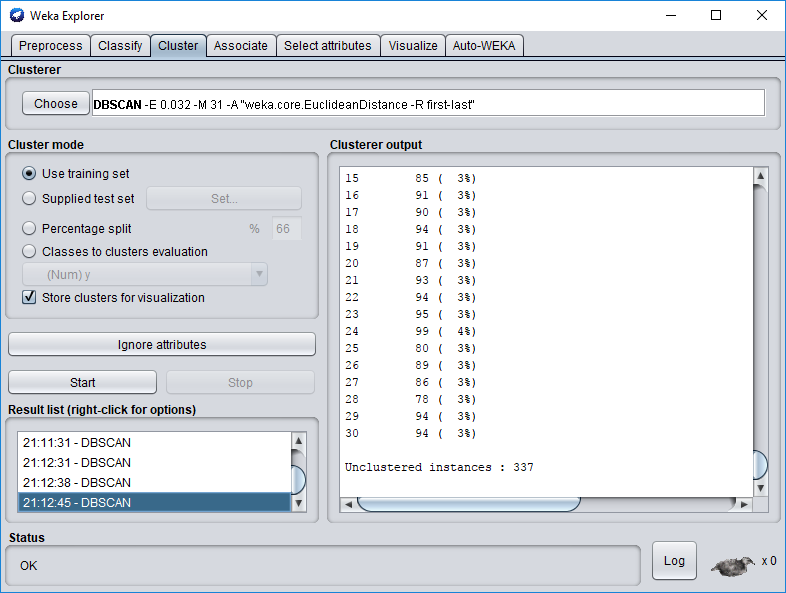


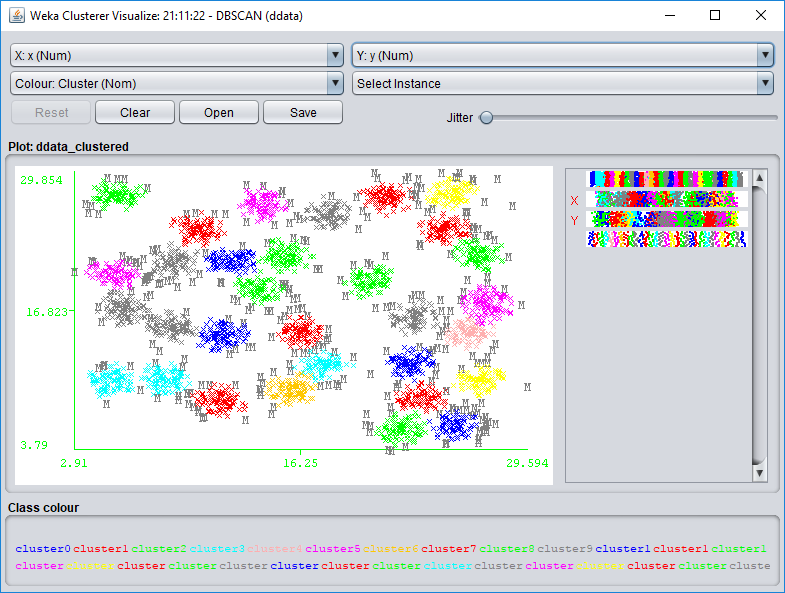
Spiral: shown are the parameters for DBSCAN that gave correct clustering, and a visualization of the clustering. In this case precise results could be found.



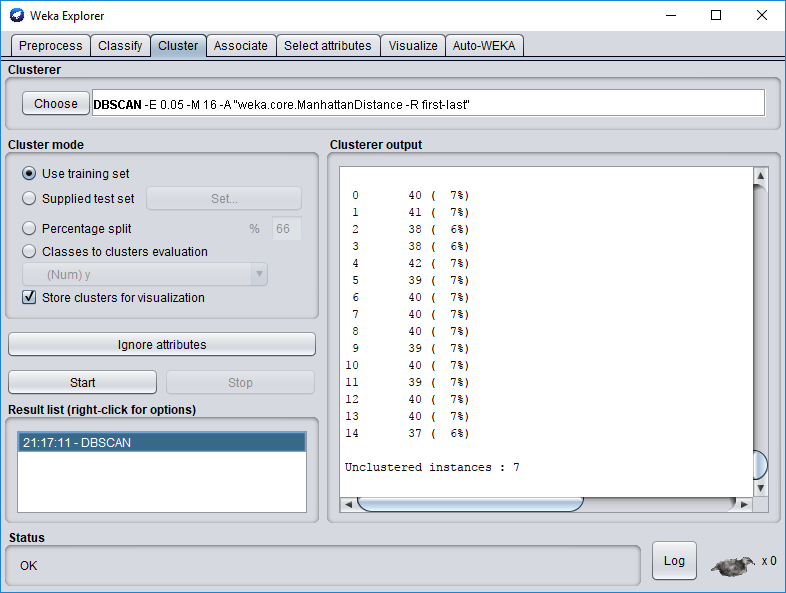


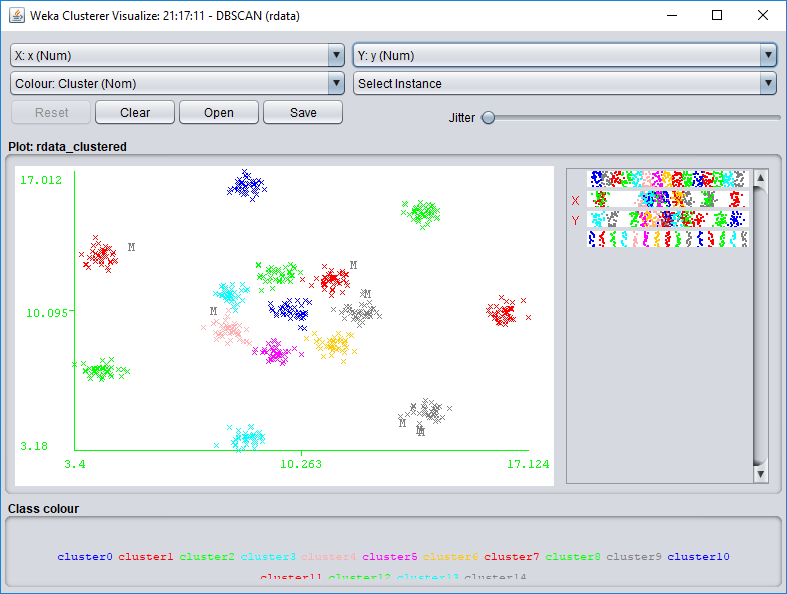
D31: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. In this case the correct number of clusters could be found, though with many unclustered instances.



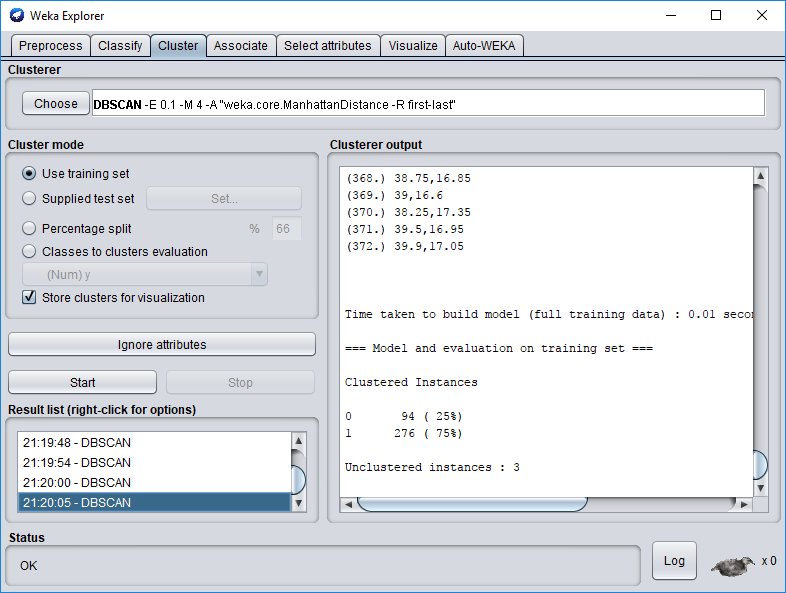


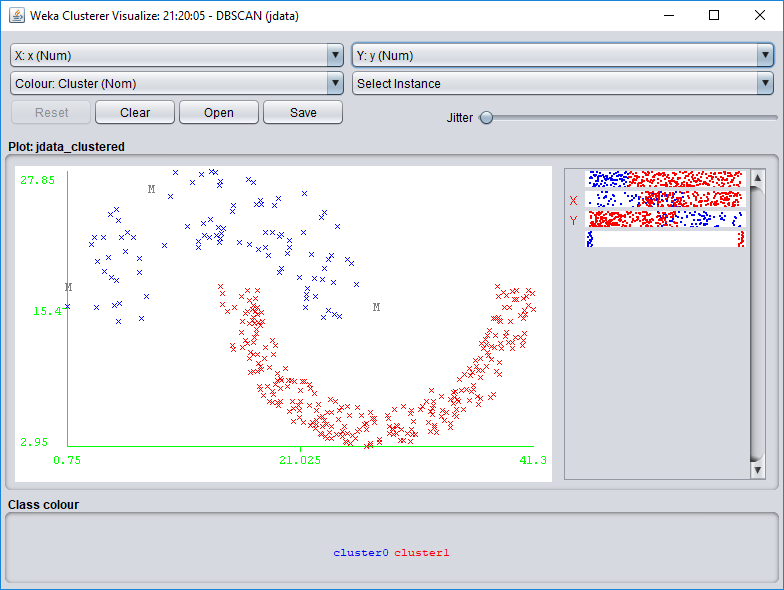
R15: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. In this case the correct number of clusters could be found with only a few unclustered instances.





Jain: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. In this case the correct number of clusters could be found with only a few unclustered instances.





Flame: shown are the parameters for DBSCAN that brought me closest to correct clustering, and a visualization of the clustering. In this case the correct number of clusters could be found with only a few unclustered instances.

