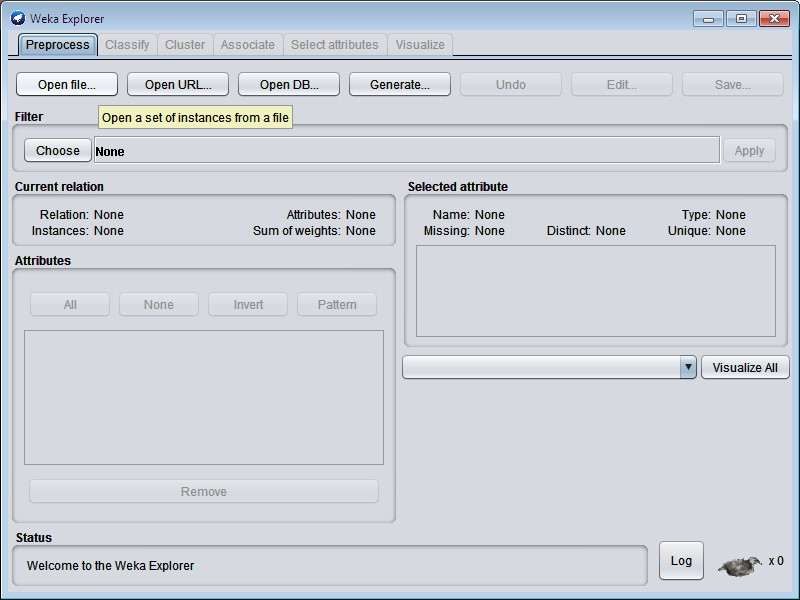
LAB: CLUSTERING DATA

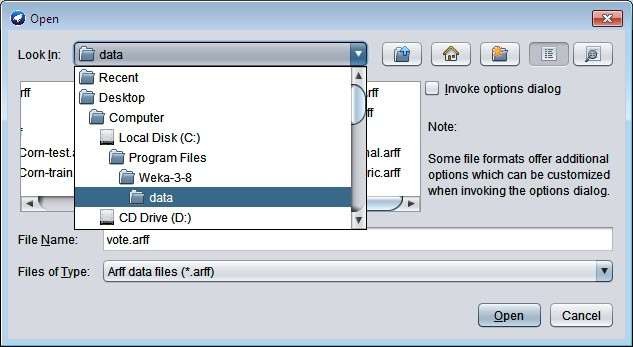
**Exercise 1:** This experiment illustrates the use of simple k-mean clustering with Weka. The sample data set used for this example is based on the vote.arff data set. This document assumes that appropriate pre-processing has been performed.

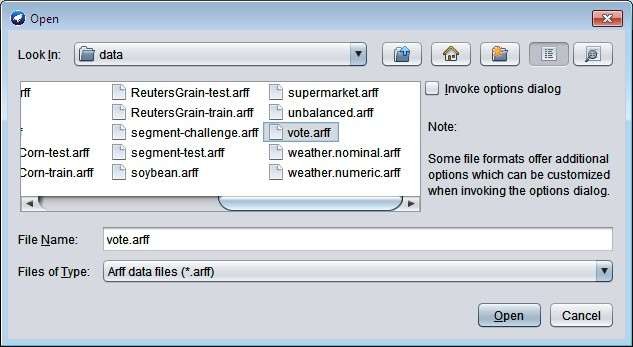
1. **Using Weka Explore**
   1. Open **weka** tool and click **Explorer.**

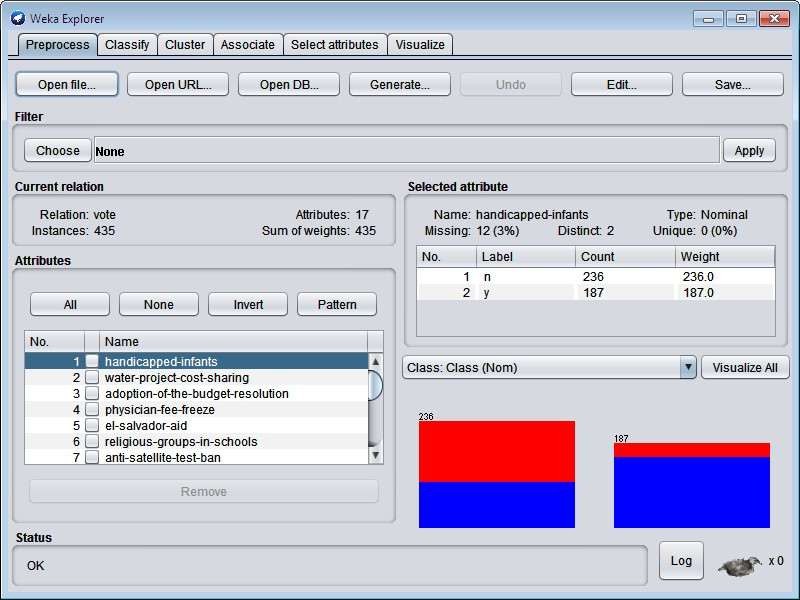
****

* 1. click**Open file**…in Preprocess tab.- choose **vote.arff**

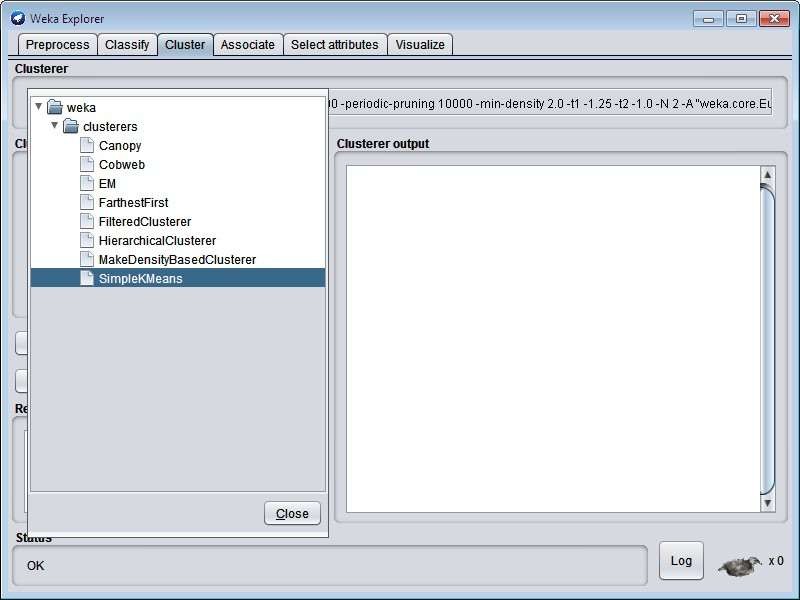
****

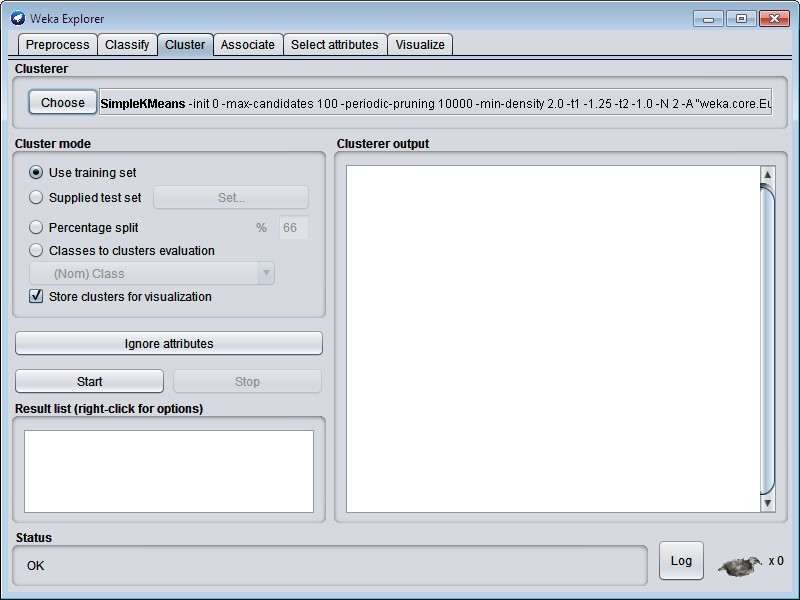




****

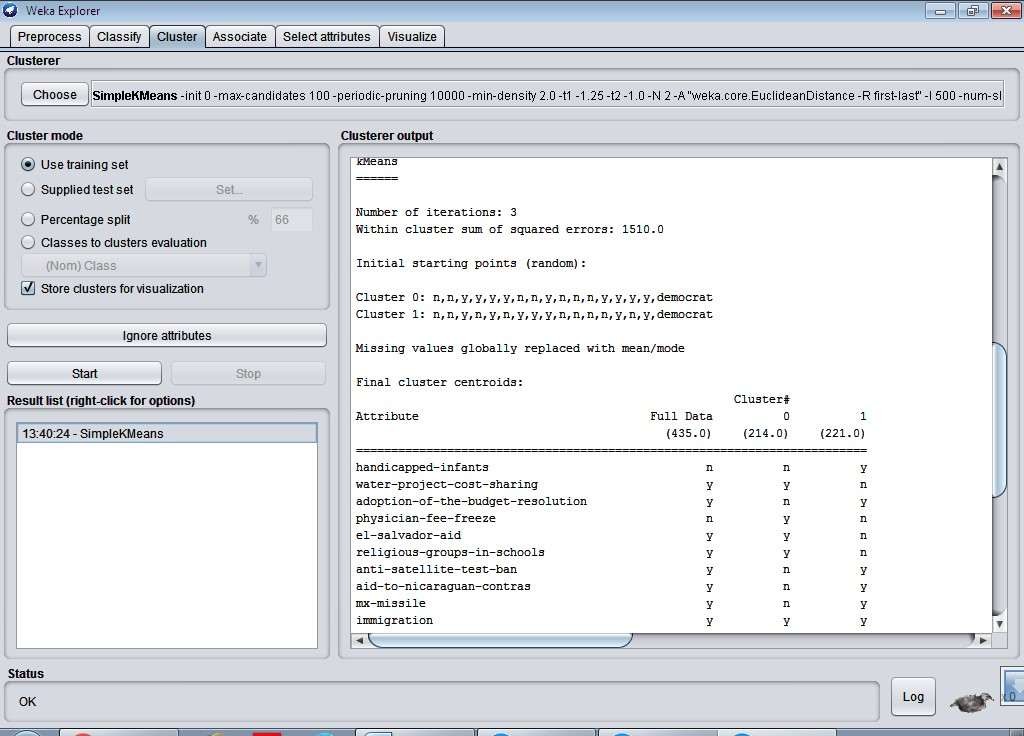
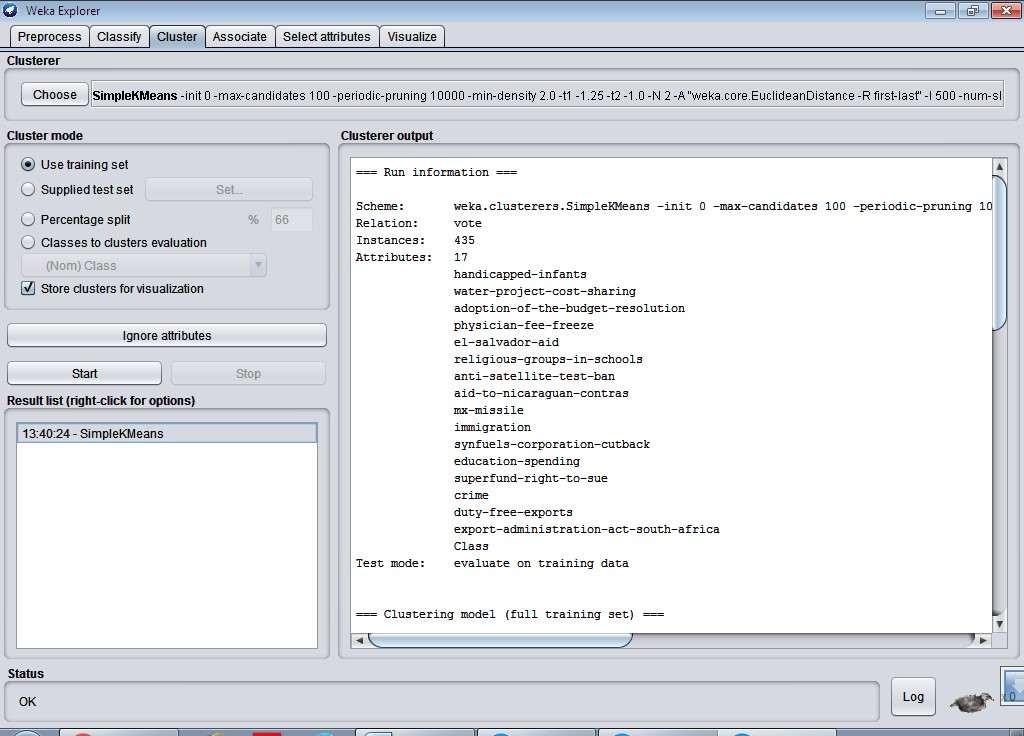
Choose **cluster** tab – click **choose** button – choose **SimpleKmeans.**

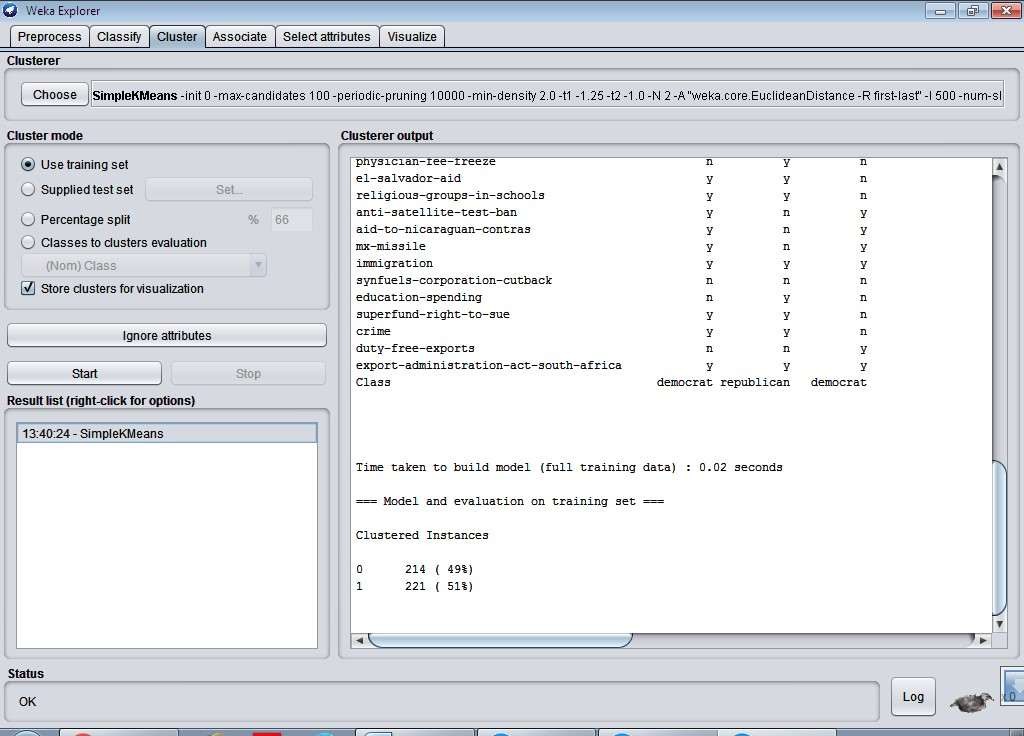


****

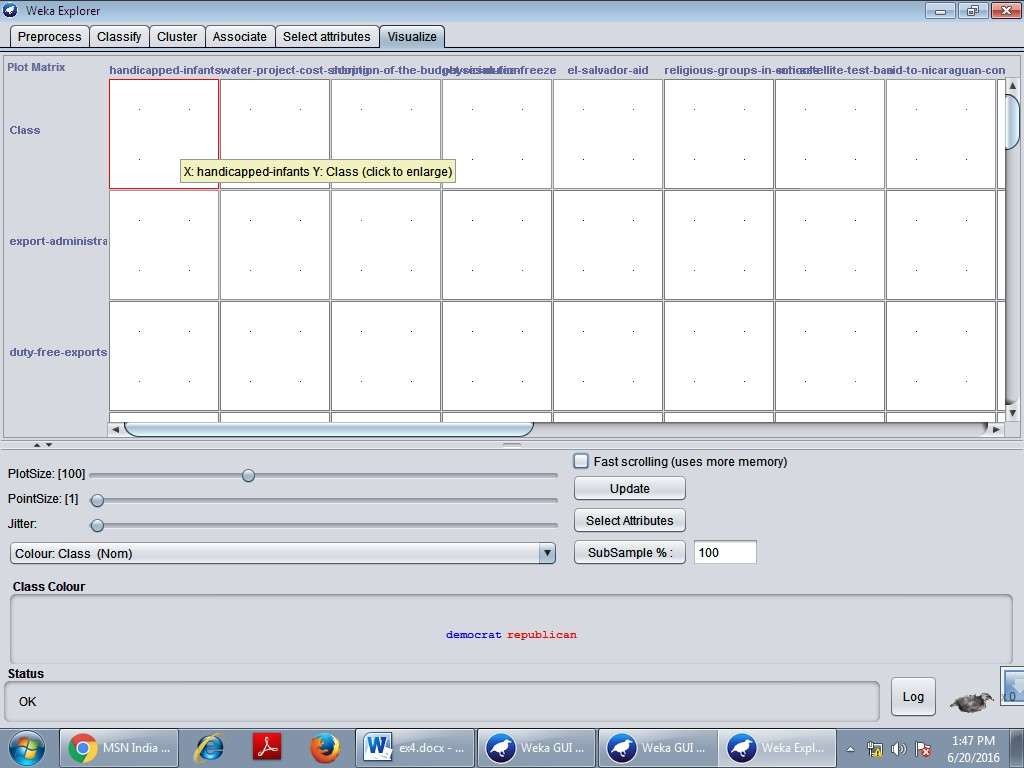
Click **Start** button

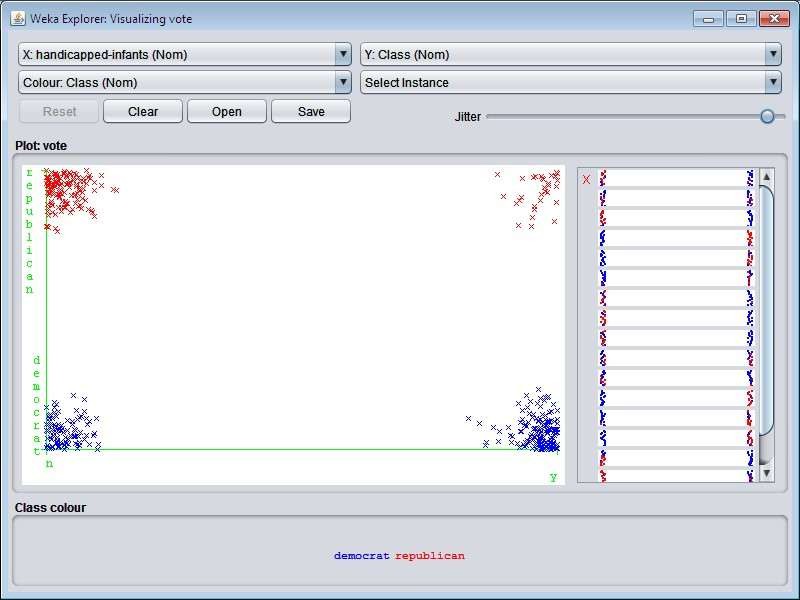
# OUTPUT:

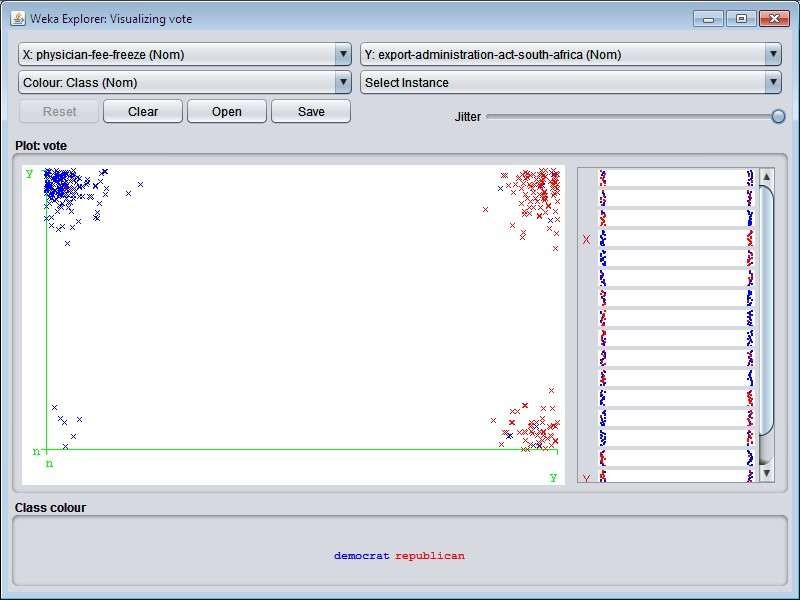
****



Goto - **Visualize** tab- click one box any visualize.

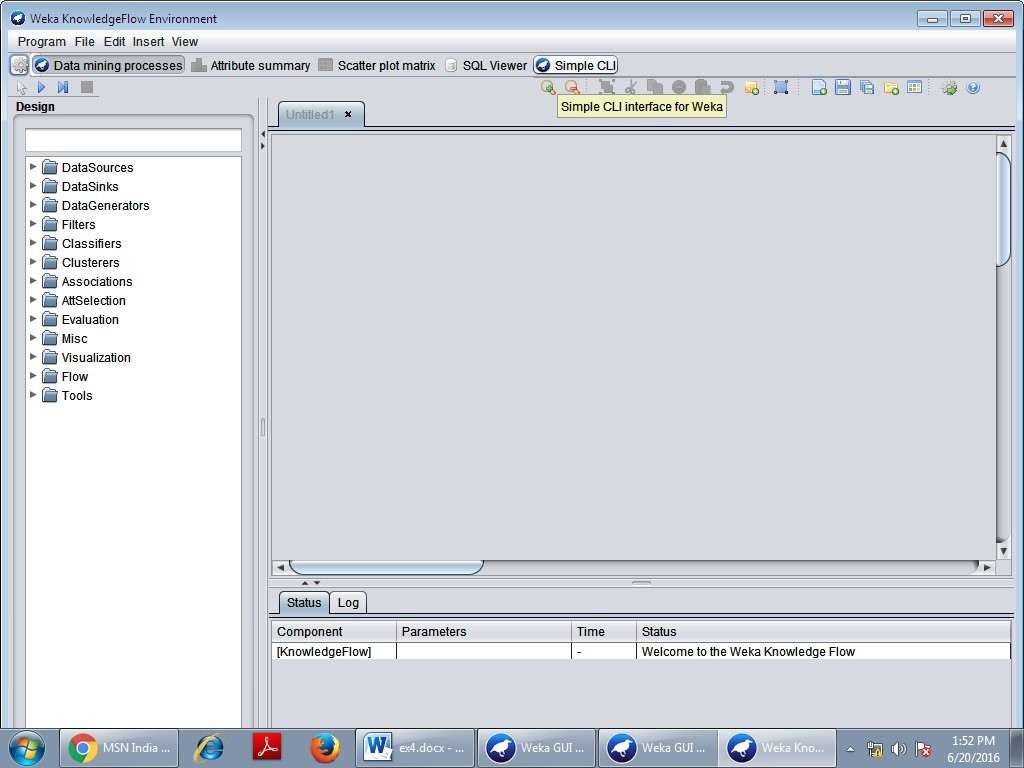


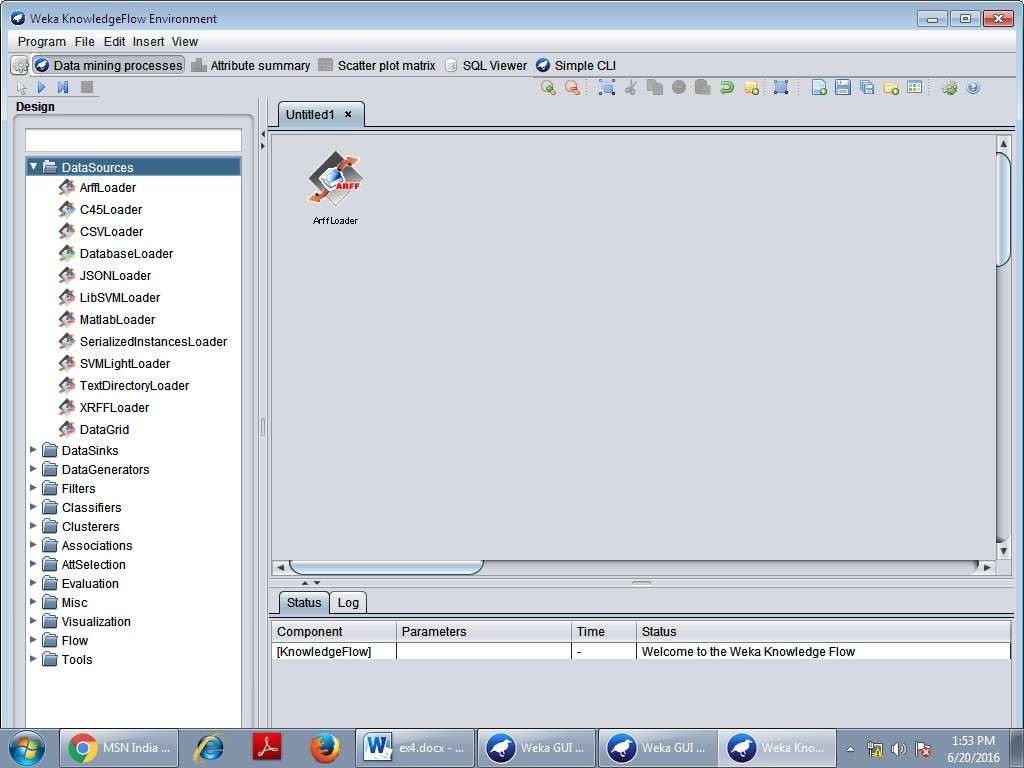




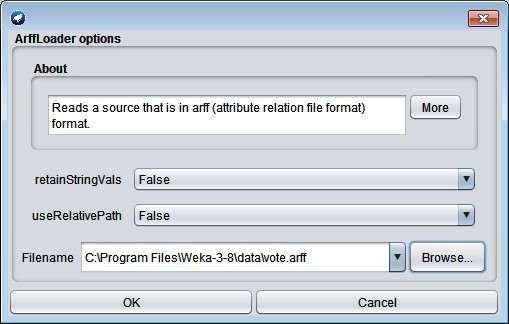
In the above move the jitter to last and to view the results of clustering.

1. **Using Weka KnowledgeFlow**
   1. Open **weka** tool and click **KnowledgeFlow**.

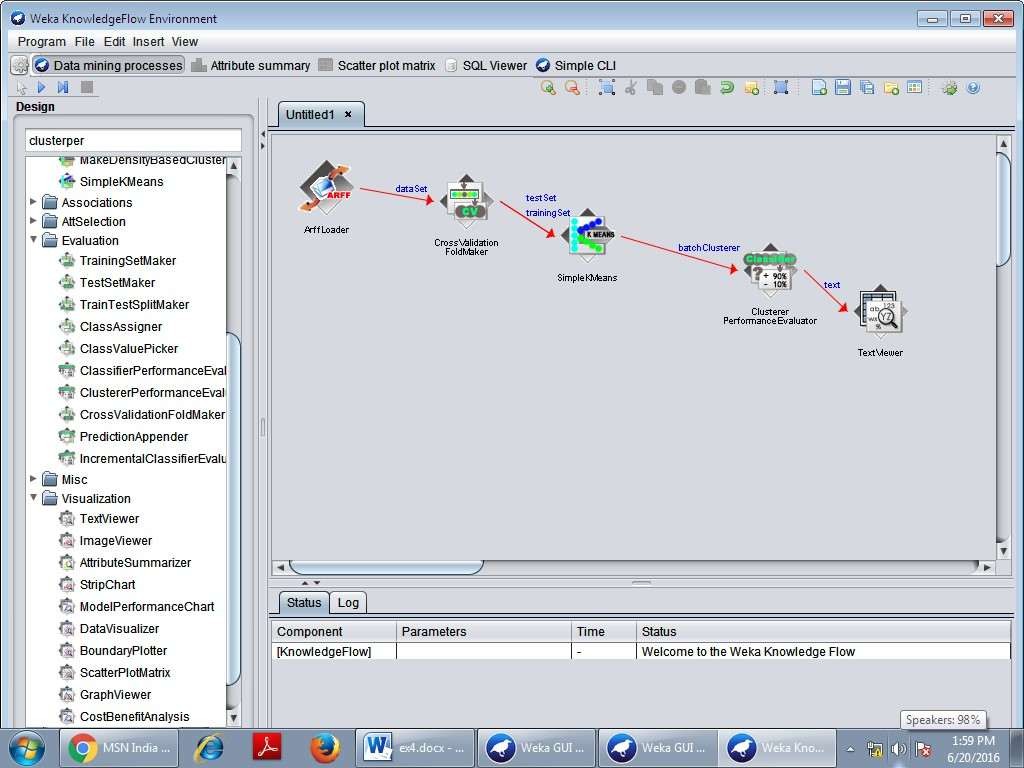


Click **DataSources** in the left side window and choose **ArffLoader** and draw in right side window

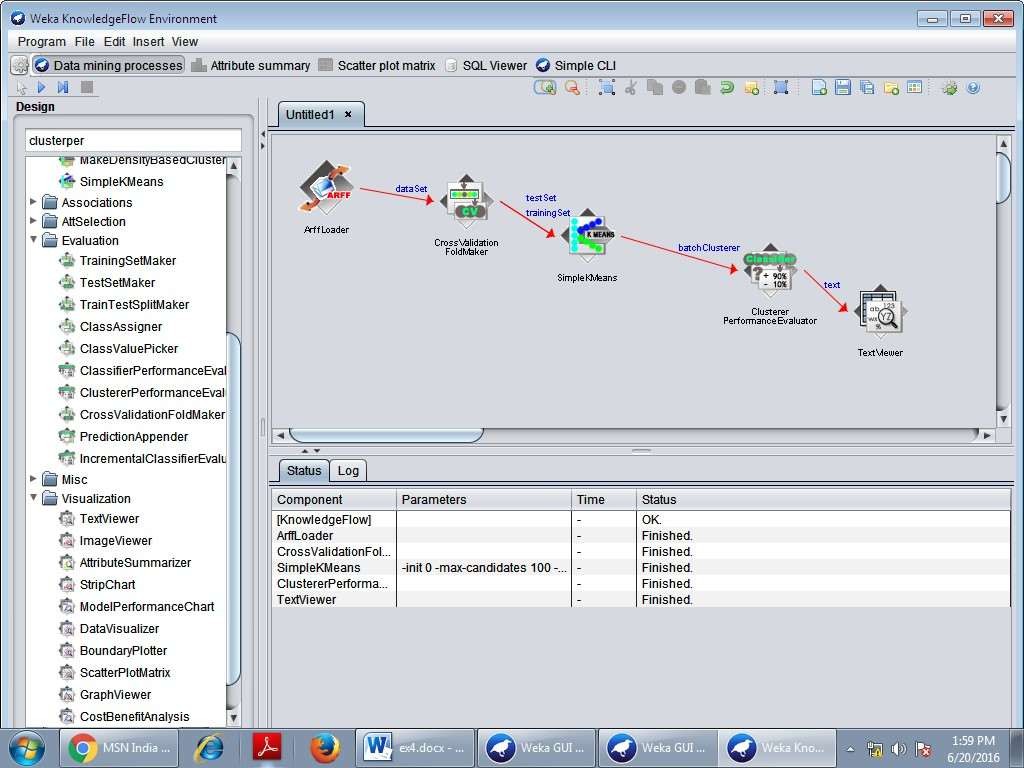
Double click **ArffLoader** and choose the file name.



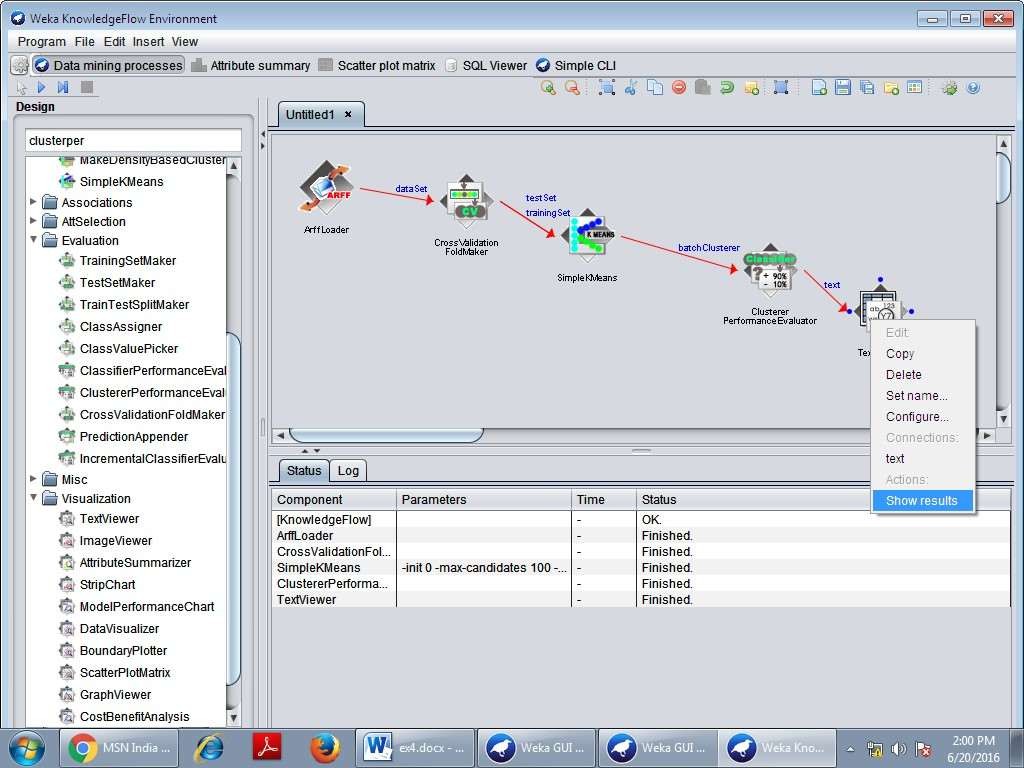
Similarly do the following.



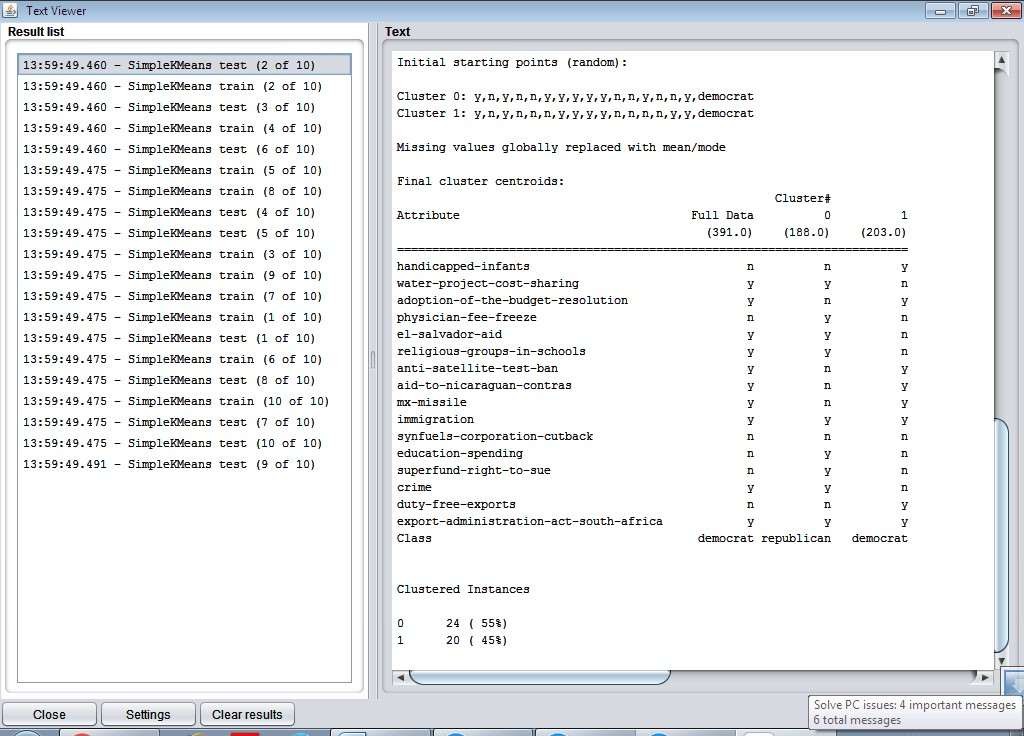
Click **Run** button.



Right click the **TextViwe**r choose **Show results** option.



# OUTPUT:



**Exercise 2:** This experiment illustrates the use of one hierarchical clustering with Weka. The sample data set used for this example is based on the vote.arff data set. This document assumes that appropriate pre-processing has been performed.

# PROCEDURE:

1. Open the data file in Weka Explorer. It is presumed that the required data fields have been discretized.
2. Clicking on the cluster tab will bring up the interface for cluster algorithm.
3. We will use hierarchical clustering algorithm.
4. Visualization of the graph

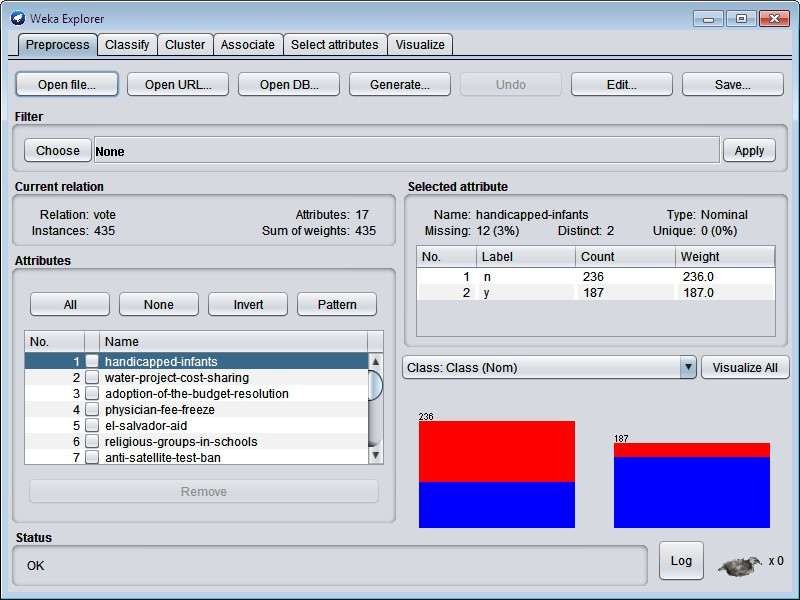
# STEPS:

The following screenshot shows the clustering rules that were generated when hierarchical clustering algorithm is applied on the given dataset.

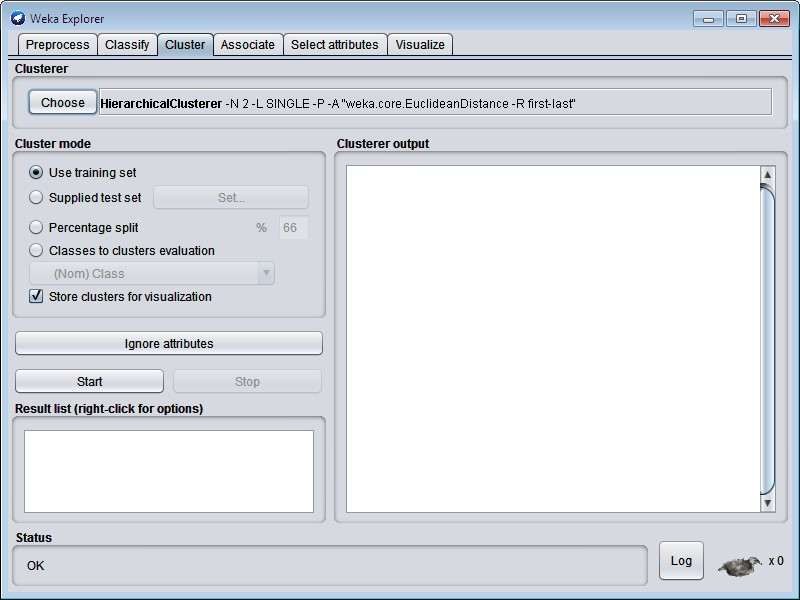
1. Open**Weka** tool and choose **Explorer.**

****

1. Click - **Open file**… in **preprocess** tab –choose **vote.arff.**

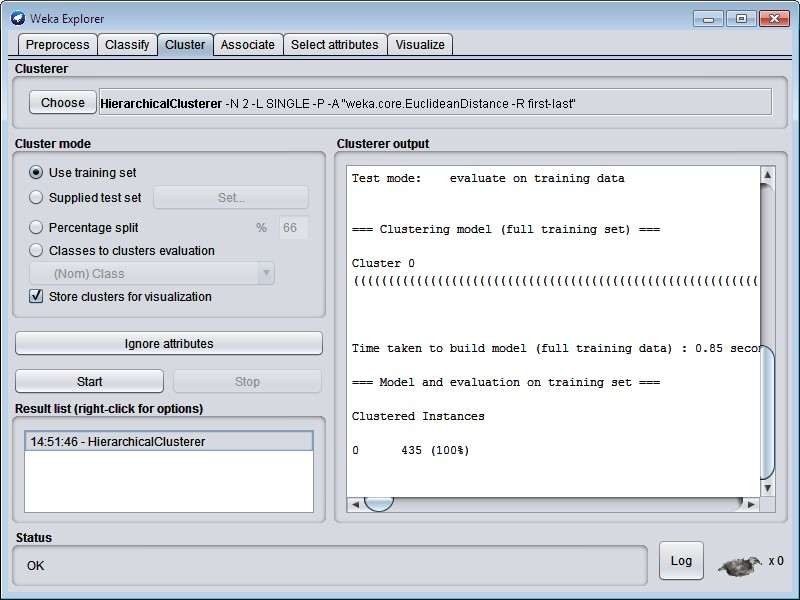
****

1. Goto **Cluster** tab – click **choose** button - select **HierarchicalClusterer.**

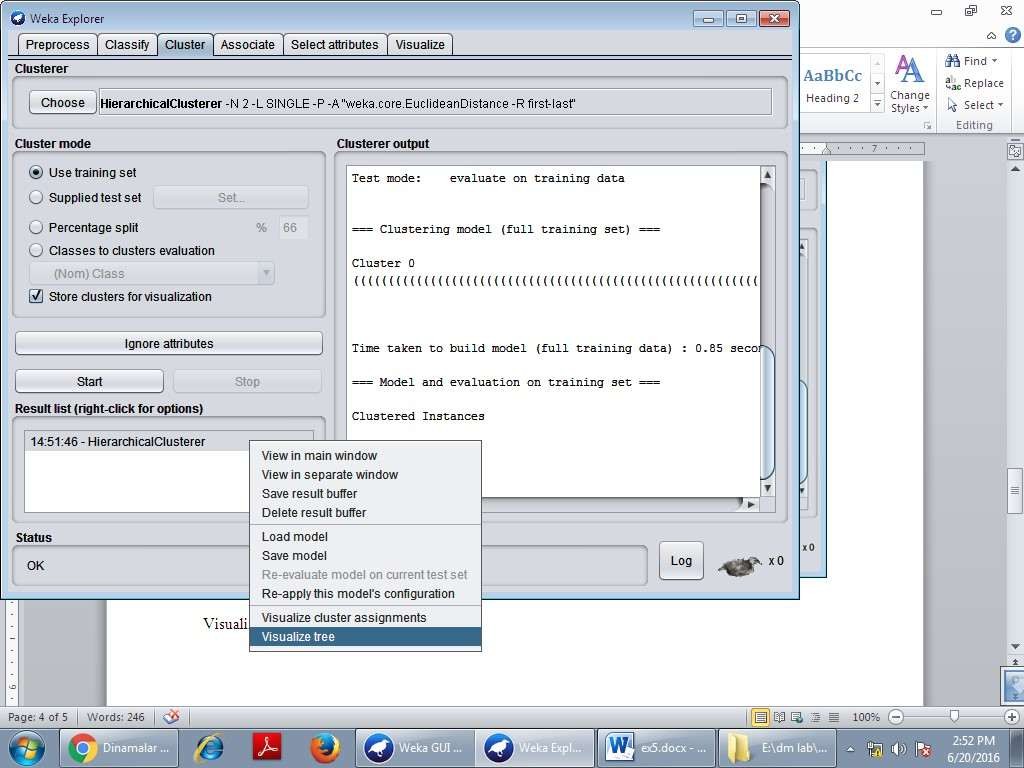
****

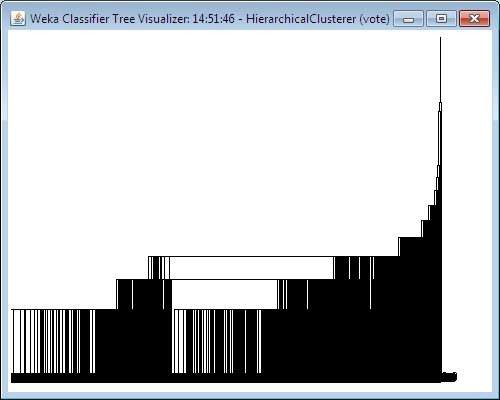
1. Click **Start** button.

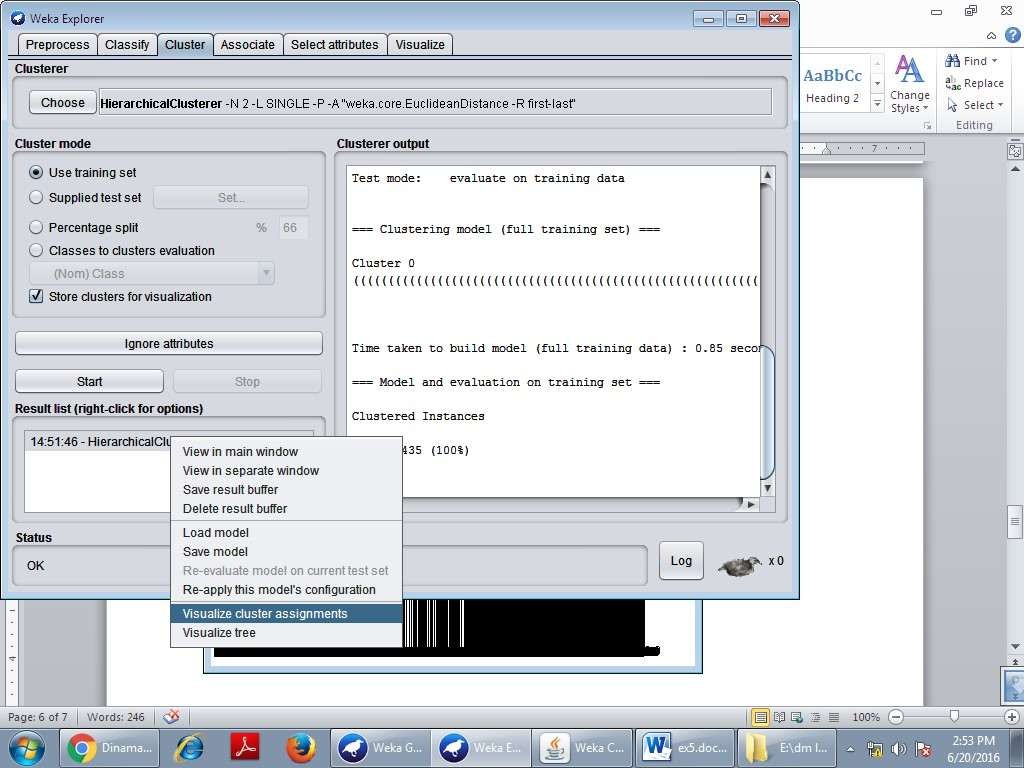
# OUTPUT:

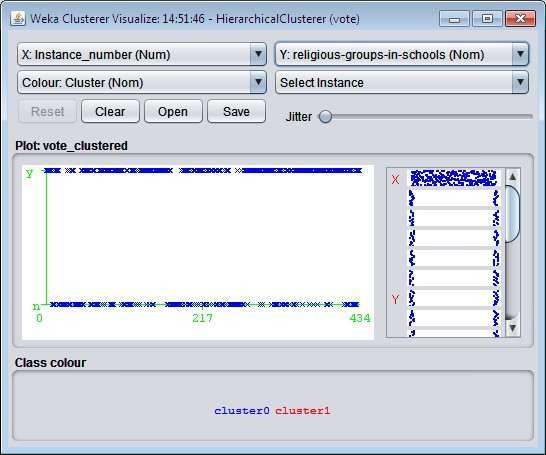


1. Visualize the tree by **right click**ing and choose **Visualize Tree** option.









**Exercise 3:** The goal of this data mining study is to find groups of animals in the ***zoo*** dataset, and to check whether these groups correspond to the real animal types in the dataset.

* 1. What types of variables are in this dataset?
  2. How many rows / cases are there?
  3. How many animal types are represented in this dataset? List them here.
  4. After removing the ***type*** attribute, go to the ***Cluster*** tab. How many clustering algorithms are available in **Weka**?
  5. List the clustering algorithms seen in class, and map these to the ones provided in **Weka**.
  6. Start using the ***SimpleKMeans*** clusterer choosing 7 clusters. Do the clusters learnt and their centroids seem to match the animal types?
  7. Start using the ***HierarchicalClusterer*** with Single Link.
  8. Start using the ***HierarchicalClusterer*** with Complete Link.
  9. Start using the ***HierarchicalClusterer*** with Average Link.
  10. Start using the ***HierarchicalClusterer*** with Mean Link.
  11. Start using the ***HierarchicalClusterer*** with Centroid Link.

**Exercise 4: Using Weka Knowledge Flow to cluster data for the above exercises.**

**Exercise 5: Using Java programming language to cluster data for the above exercises**