

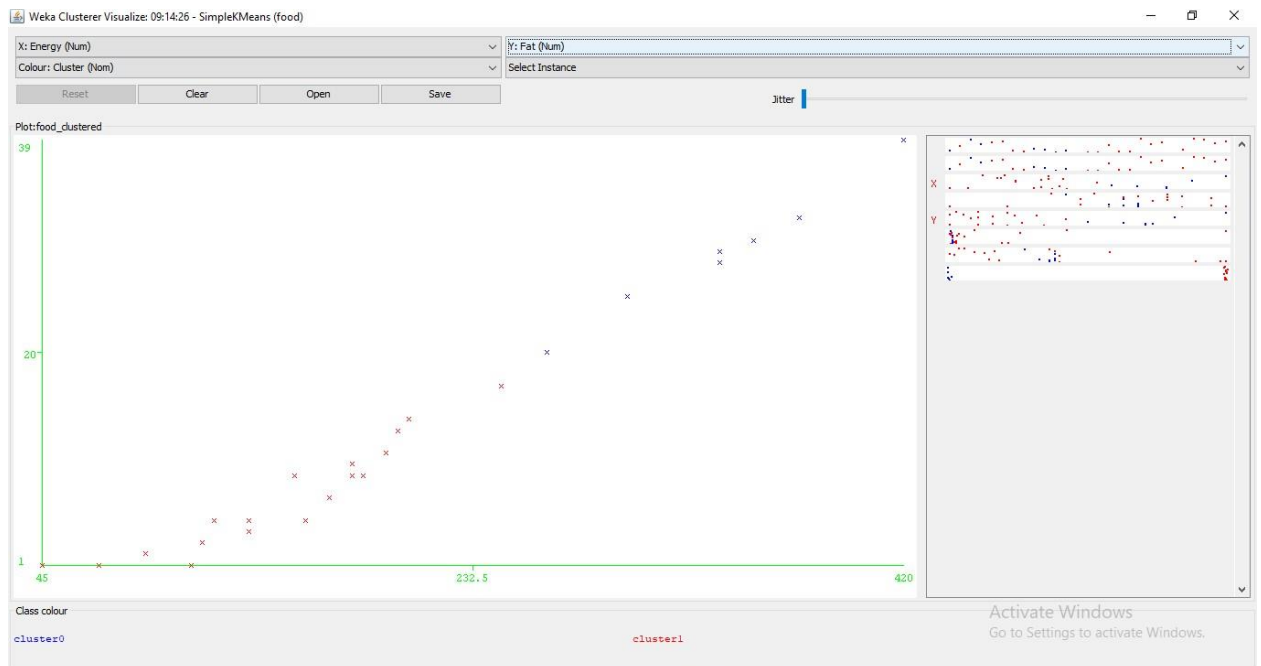
# 732A75: Clustering Lab

Liu ID: vinbe289 , muhha866

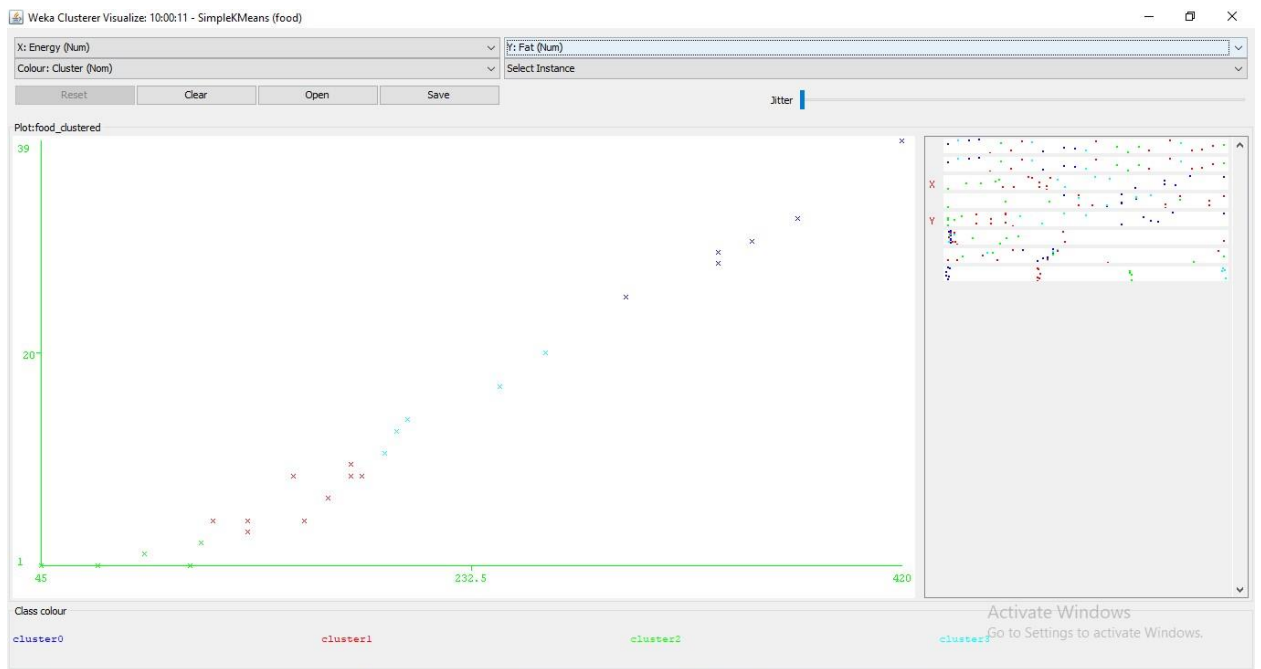
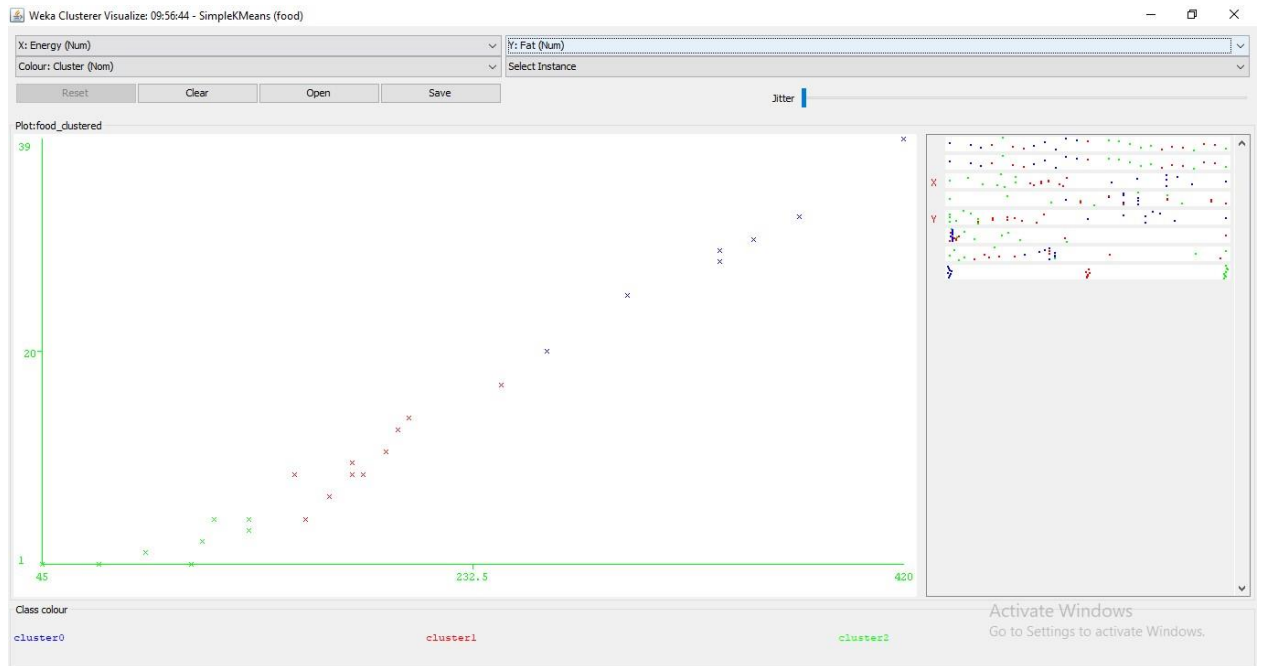
## Simple K means:

1. Attributes Energy and Fat is chosen since the values for these 2 attributes does not have a large range and the variation in values is neither too large nor too small, it is optimum. Name attribute can be ignored since it does not play a role in clustering and each observation has different a name. Calcium was not chosen since the range of values is huge. Protein was not chosen since there is no much of a variation in values. Iron was not chosen since the values are on a smaller scale compared to other attributes.

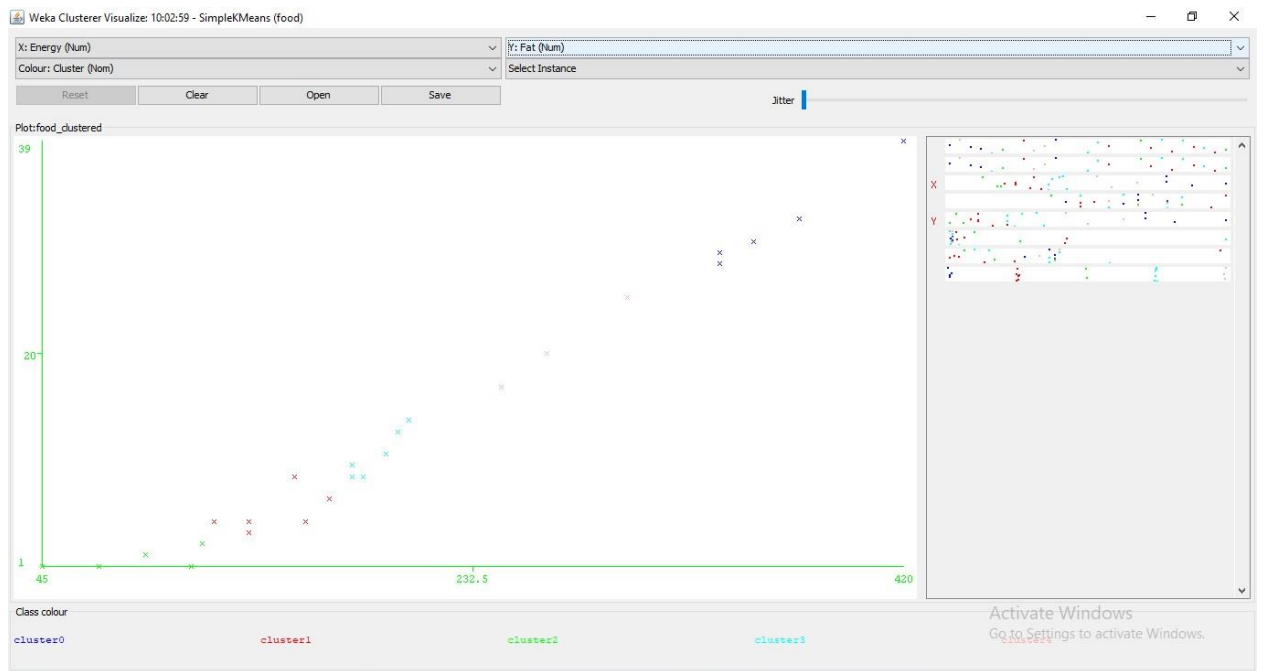
2.



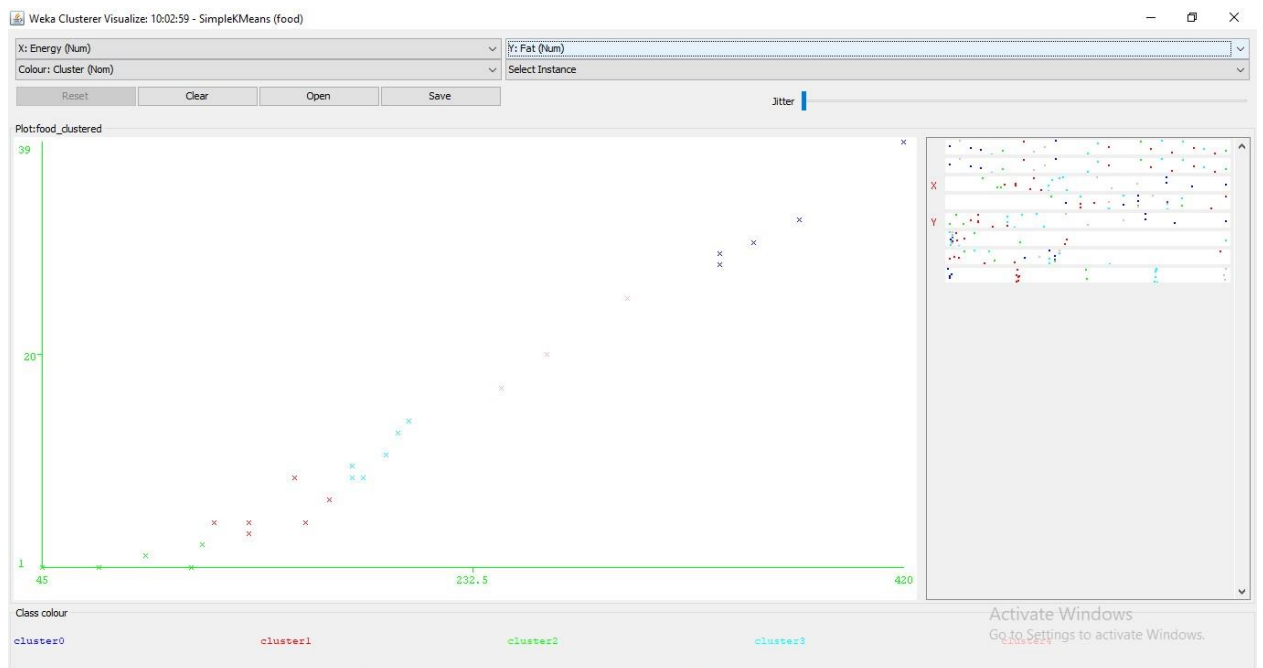
# 732A75: Clustering Lab



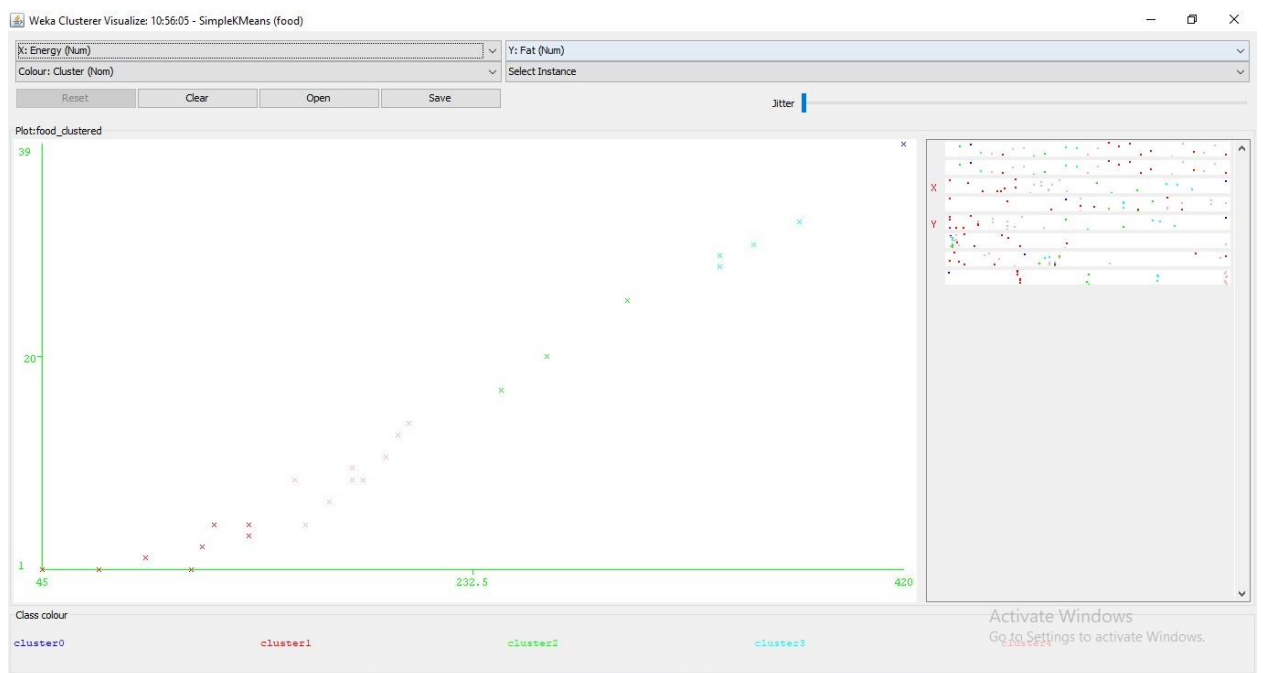
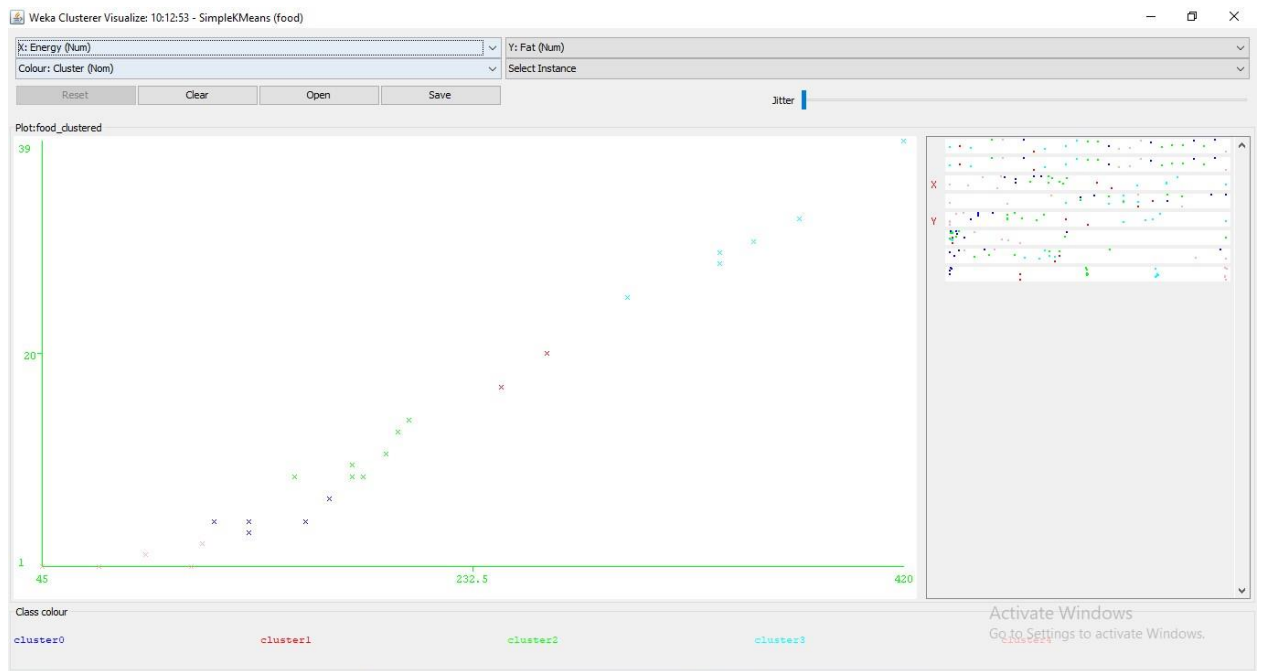
# 732A75: Clustering Lab



3.



# 732A75: Clustering Lab



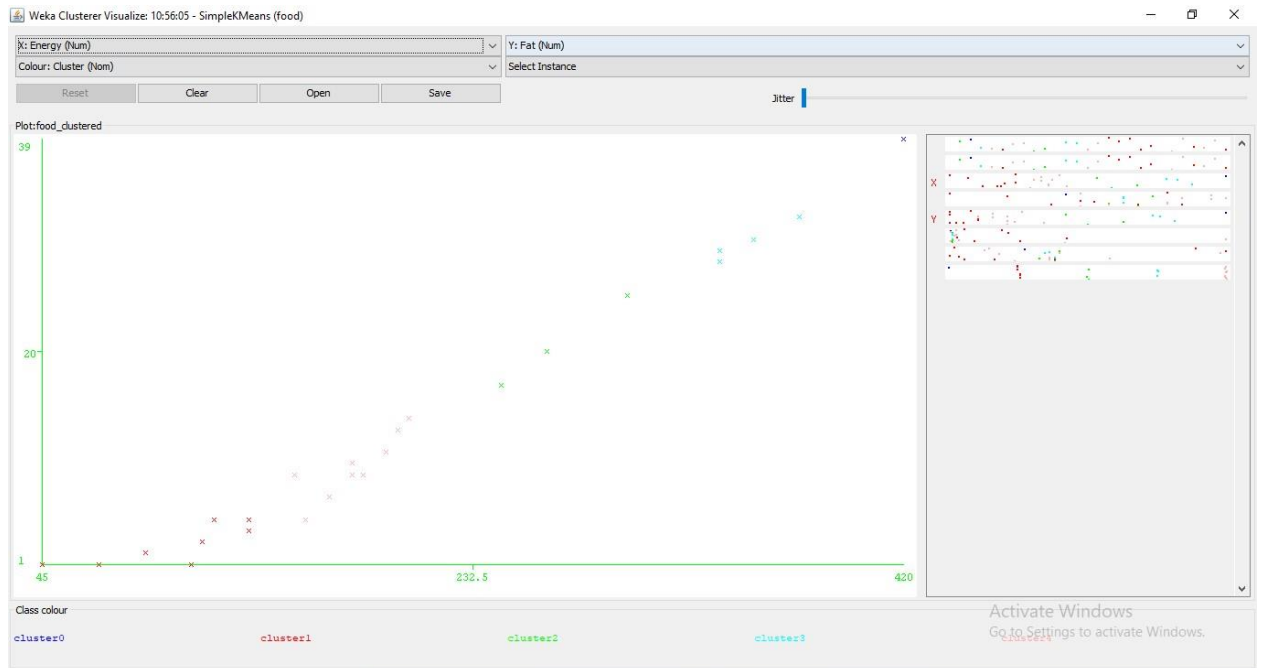
Seed value seems to change the position of initial cluster center. This results in the shifting of some observations from one cluster to the other adjacent cluster as it can be seen by comparing the plots in this step with the plots from Step 2.

4. The clusters are not the best clusters but decent clusters. Not all the members in the same clusters very close to each other even though in the majority of the cases it is similar. Not all members of

# 732A75: Clustering Lab

different clusters dissimilar. In some cases though a particular member is similar to a particular cluster it is assigned to a different cluster.

5.



Each cluster represents a group of food similar to each other with the respect to the nutrient levels that is being considered i.e. Energy and Fat in this case. The plot with 5 number of clusters and seed value of 1234 is considered here. Each cluster can be characterized with respect to the considered attributes.

Cluster 1: Very low level energy

Cluster 4: Low level energy

Cluster 2: Medium level energy

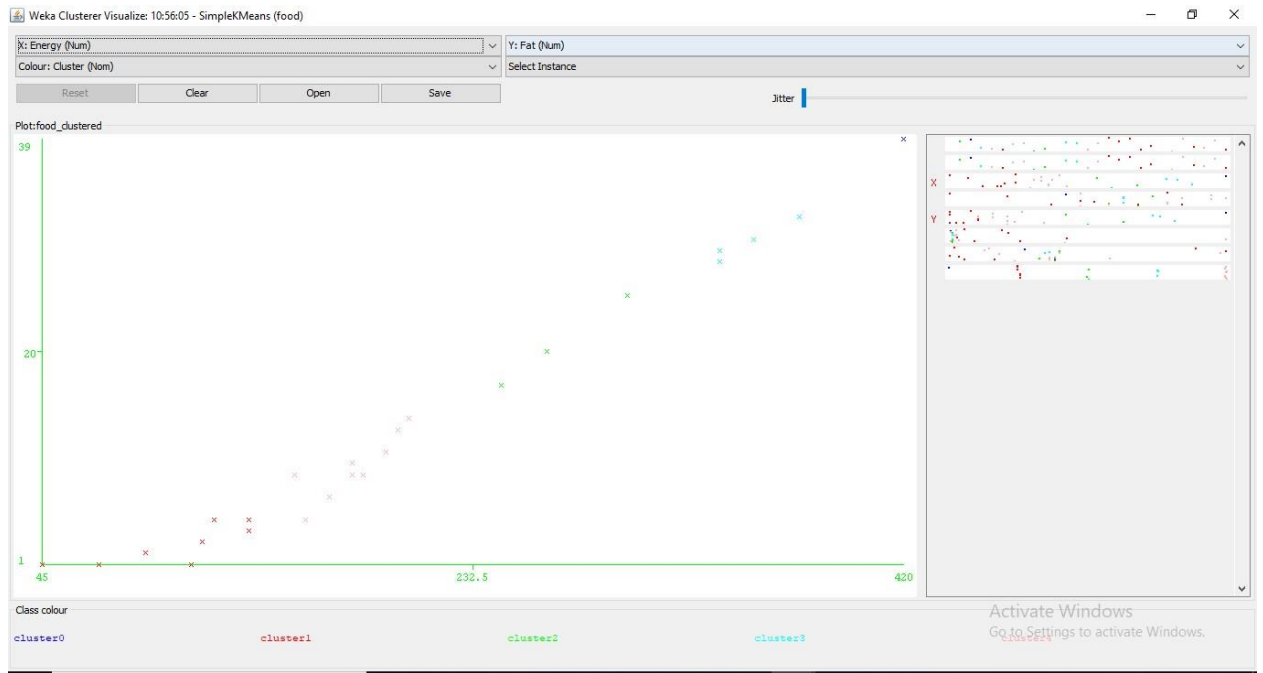
Cluster 3: High level energy

Cluster 0: Very high level energy

# 732A75: Clustering Lab

## Make Density Based Clusters:

1.



2.



## 732A75: Clustering Lab

The image above shows a series of plots (in order from right to left and from top to bottom) in the increasing order of Standard Deviation. The increase in Standard deviation aids the increase in the room for variability in a cluster, between observations (its values). When there is room for increase in large variability, all observations tend to fall into a common cluster. With a very large Standard Deviation all the observations fall into one cluster as shown above.