CLUSTERING - Activity 1 of 2

Normalized Jaccard Similarity and Distance for GAM data:

* avoid skewing result due to differences in #windows detected by 2 NPs
* let the **denominator** in the Jaccard equation = **min{|A|,|B|}**
  + **|A|**: represents the total number of attributes where the attribute of *A* is 1
  + **|B|**: represents the total number of attributes where the attribute of *B* is 1

Use k-means clustering to *form 3 clusters* from the NPs that captured the Hist1 region.

Before the next class (Thursday) implement the following steps of k-means clustering and perform the steps on the NPs that captured the Hist1 region.

1-Randomly select k = 3 distinct data points. These are the initial clusters.

2-Measure the distance between each point and each of the ‘k’ clusters.

3-Assign each point to the nearest cluster.