HW Bonus 03: k-Means EZ
See the associated Dropbox for due date.
Thomas Kinsman

This assignment is like HW05, without noise.

Homework is to be programmed only in one of the following languages. No other languages will be accepted. Please limit yourself exclusively to: Java, Python, Matlab, or R. The last three have good native graphics and plotting support.

Assume that the grader has no knowledge of the language or API calls, but can read comments.
Use prolific block comments before each section of code, or complicated function call to explain what the code does, and why you are using it. Put your name and date in the comments at the heading of the program.

Hand in IN A ZIP FILE:

1. Your write-up including your results, **HW05_LastName_FirstName_kMeans.pdf**
2. Your well commented code, **HW05_LastName_FirstName_kMeans…**

You are provided with a file of training data. This data has only three attributes to select from.

IMPLEMENT k-MEANS CLUSTERING.

1. The data file is provided.
2. The data points represents stars in space. The three attributes are (X,Y,Z) coordinates.
   The units are "Astronomical Units". Your job is to identify the galaxies in this star field using k-Means clustering, and your own smarts.

3. **Write-UP Should Include:**
   a. What distance metric did you use? **USE THE EUCLIDEAN**

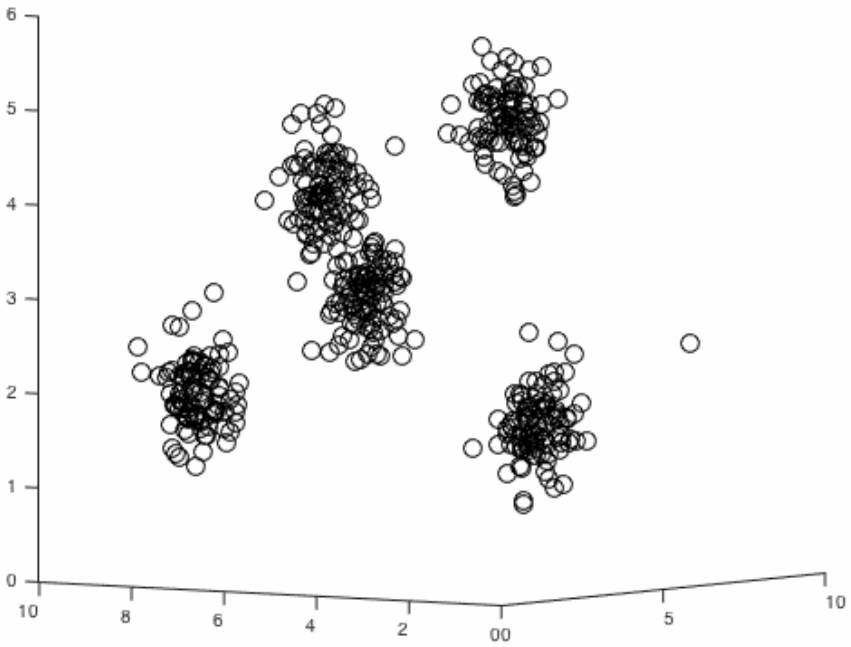
   b. What prototype did you use for a cluster? **USE A CENTER OF MASS AS THE CENTER.**

   c. Plot the SSE versus K **FROM K = 1 to 15.** (1) (Do not use Excel.)
   Identify the knee in the graph.

   d. What value of K did you select? What was the associated SSE? (1)

   e. Sort the clusters from smallest to largest. For each cluster, 1 to K report:
      i. **How many data points were in it?**
      ii. **What was the center of mass of this cluster?**

   f. Assuming K is under 11, plot each cluster in a different name space color, starting with: red, green, blue, yellow, magenta, black, gray, brown, orange, pink, and cyan. (Skip white, and use cyan last.)
   Plot this on an (X,Y) axis only, with Z coming out of the paper at us. (1) (Do not use Excel.)


6
5
4
3
2
1
0
10
8
6
4
2
00
5
10