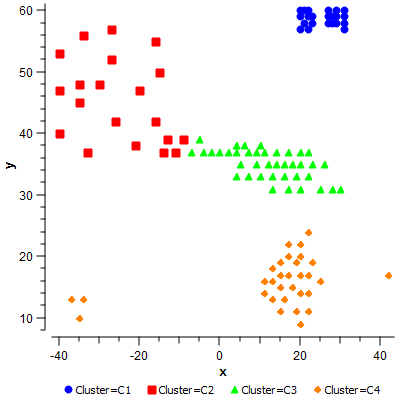
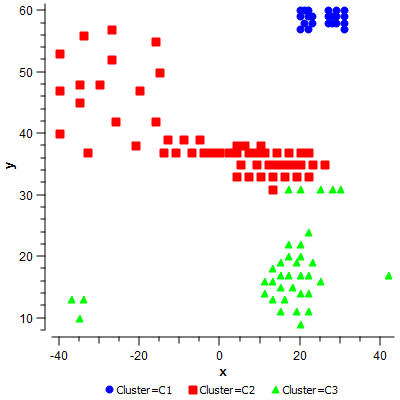
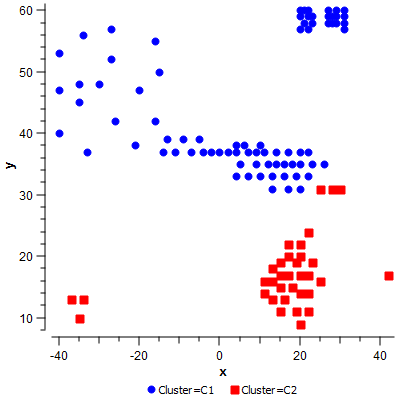
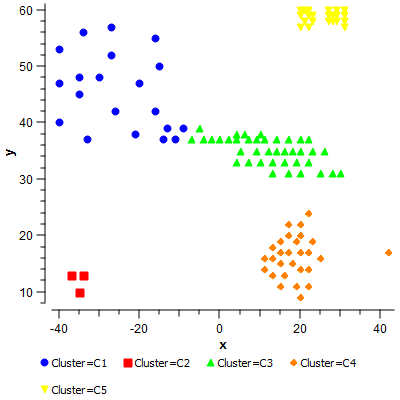
Problem Set 3

Consider the dataset shown on a grid in the attached sheet and convert it into a data file of 2-D points. Then do the following three problems (#1-#4). Feel free to use Orange or any other tool or your own written programs in Matlab or any programming language. These problems will require work over and above that is performed by standard packages and toolboxes.

1. Run k-means algorithm with this dataset for values of k equal to 2, 3, 4, and 5. Choose initial k cluster centers randomly.
   1. For each k value display the clusters formed with a different color for points belonging to each cluster.





* 1. Find the SSE for each value of k. Which value of k gives the minimum SSE?

|  |  |
| --- | --- |
| **k** | **SSE** |
| 2 | 2022.32 |
| 3 | 1464.95 |
| 4 | 971.56 |
| 5 | 795.41 |

The best value for K is 5, since this gives us the lowest sum of squared errors. See my work in the attached excel sheet.



* 1. Which of the k values gives you the clusters closest to your intuitive perception of clusters in this dataset?

Visually, 5 clusters appears to fit the data the best. I would also experiment with K values of 6, 7, 8, and 9.

* 1. Explain why the differences between the algorithms’s output and your intuitive clusters.

There are no differences between my favorite version and the algorithm’s output.

1. Run hierarchical clustering algorithm with the given dataset. Use single-link method.
   1. Display how clusters will be formed if we want to generate 2, 3, 4, and 5 clusters.
   2. For each number of clusters display all the points on a grid with different colors for points in each cluster.
   3. Explain any discrepancies between the clusters formed and your intuitive expectation of clusters.
2. Repeat #2 above for complete link method of hierarchical clustering.
3. Compute Rand Index for 4 clusters computed in #1 and #3 above. How would you interpret the value of the Rand index you obtain?
4. Do Q#1 from chapter 15 of the textbook, given on page 391.