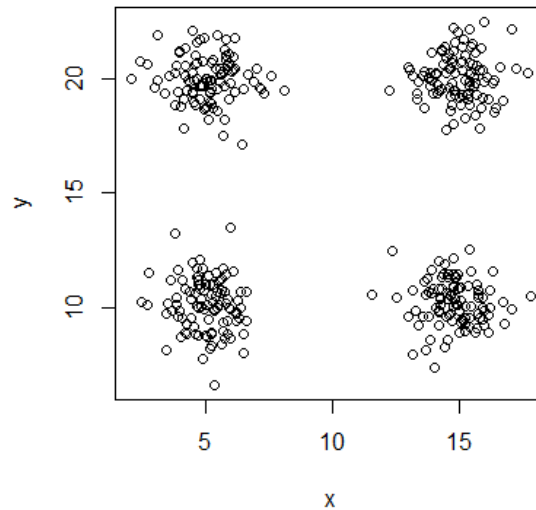# Math 5364 Homework 18

1. Consider the data set from problem 1 on Homework 17 (displayed below).

   (a) Find the number of clusters that maximizes the silhouette coefficient, and plot the silhouette coefficient vs K.

   (b) What is the maximum possible value of the silhouette coefficient?

   (c) Plot SSW vs K. Does the optimal value of K suggested by this plot agree with the one based on the silhouette coefficient?

   (d) Optional: Is the silhouette coefficient for this clustering statistically significant? (It may be a good idea to let R run while you're out of the office to do this problem.)

2. Repeat problem 1 for the `wdbc` data set. Also, find the weighted entropy and purity for the optimal clustering.