Questions to answer while doing your lab:

1) How many columns are there?

2) Which column is unnecessary? (Which column did you clean?)

3) Are the gene 1 or gene 2 clusters easier to separate?

4) On the heat map, how can you tell which features are highly correlated?

5) What is the shape of your training data set? What is the shape of your test data shape? What do the numbers correspond to?

6) Fill in the accuracies of the models including all the features.

Model	Accuracy
SVM	
Logistic Regression	
Decision Tree	
K-Nearest Neighbor	

7) Fill in the accuracies for the models for the separate genes

Model	Gene 1 Accuracy	Gene 2 Accuracy
SVM		
Logistic Regression		
Decision Tree		
K-Nearest Neighbor		

8) Which gene gives us the more accurate model? Is this what you expected from the heat map?