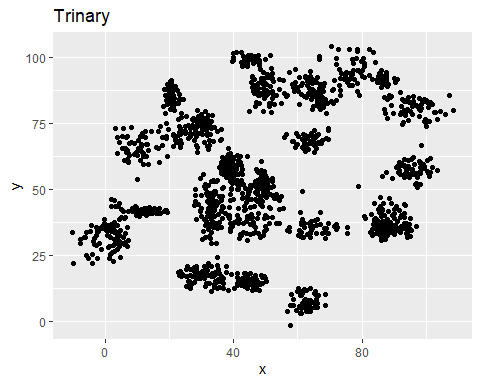
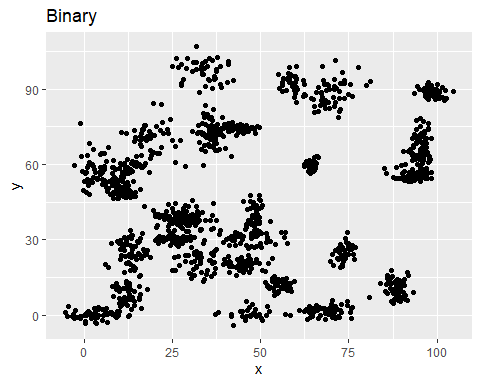
8.2 Assignment

Anna Harvey

7/25/2020

## a. Plot the data from each dataset using a scatter plot.



## b. Fit a k nearest neighbors model for each dataset for k=3, k=5, k=10, k=15, k=20, and k=25. Compute the accuracy of the resulting models for each value of k. Plot the results in a graph where the x-axis is the different values of k and the y-axis is the accuracy of the model.

## [1] "For binary dataset:"

## [1] "K-val = 3, Accuracy = 95.3333333333333 %"

## [1] "K-val = 5, Accuracy = 97.3333333333333 %"

## [1] "K-val = 10, Accuracy = 98.6666666666667 %"

## [1] "K-val = 15, Accuracy = 99.3333333333333 %"

## [1] "K-val = 20, Accuracy = 98 %"

## [1] "K-val = 25, Accuracy = 97.3333333333333 %"

## [1] "For trinary dataset:"

## [1] "K-val = 3, Accuracy = 94.2675159235669 %"

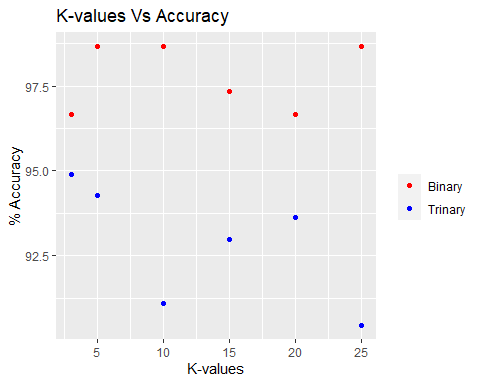
## [1] "K-val = 5, Accuracy = 94.2675159235669 %"

## [1] "K-val = 10, Accuracy = 91.7197452229299 %"

## [1] "K-val = 15, Accuracy = 89.171974522293 %"

## [1] "K-val = 20, Accuracy = 91.7197452229299 %"

## [1] "K-val = 25, Accuracy = 90.4458598726115 %"



## c. Looking back at the plots of the data, do you think a linear classifier would work well on these datasets?

This data could possibly use a linear classifier to define a decision boundary because it does look like the two sets of data are mostly separate from each other in their accuracy results.