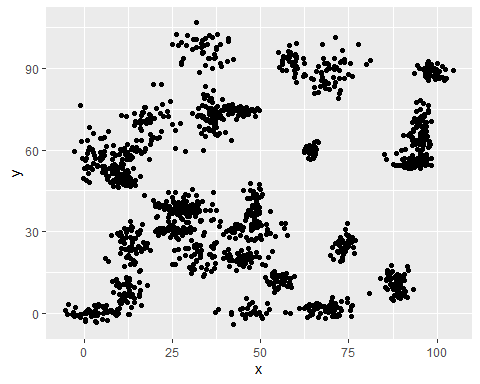
8.2 Assignment

Anna Harvey

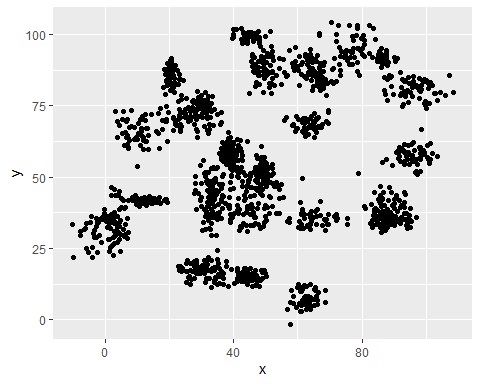
7/25/2020

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

ggplot(binary, aes(x = x, y = y)) + geom\_point()



ggplot(trinary, aes(x = x, y = y)) + geom\_point()



## 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] "If k = 3 , Accuracy = 96.6666666666667 %"

## [1] "If k = 5 , Accuracy = 98.6666666666667 %"

## [1] "If k = 10 , Accuracy = 98.6666666666667 %"

## [1] "If k = 15 , Accuracy = 96.6666666666667 %"

## [1] "If k = 20 , Accuracy = 98 %"

## [1] "If k = 25 , Accuracy = 97.3333333333333 %"

## [1] "For trinary dataset:"

## [1] "If k = 3 , Accuracy = 95.5414012738854 %"

## [1] "If k = 5 , Accuracy = 97.4522292993631 %"

## [1] "If k = 10 , Accuracy = 94.2675159235669 %"

## [1] "If k = 15 , Accuracy = 92.3566878980892 %"

## [1] "If k = 20 , Accuracy = 92.3566878980892 %"

## [1] "If k = 25 , Accuracy = 84.0764331210191 %"

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