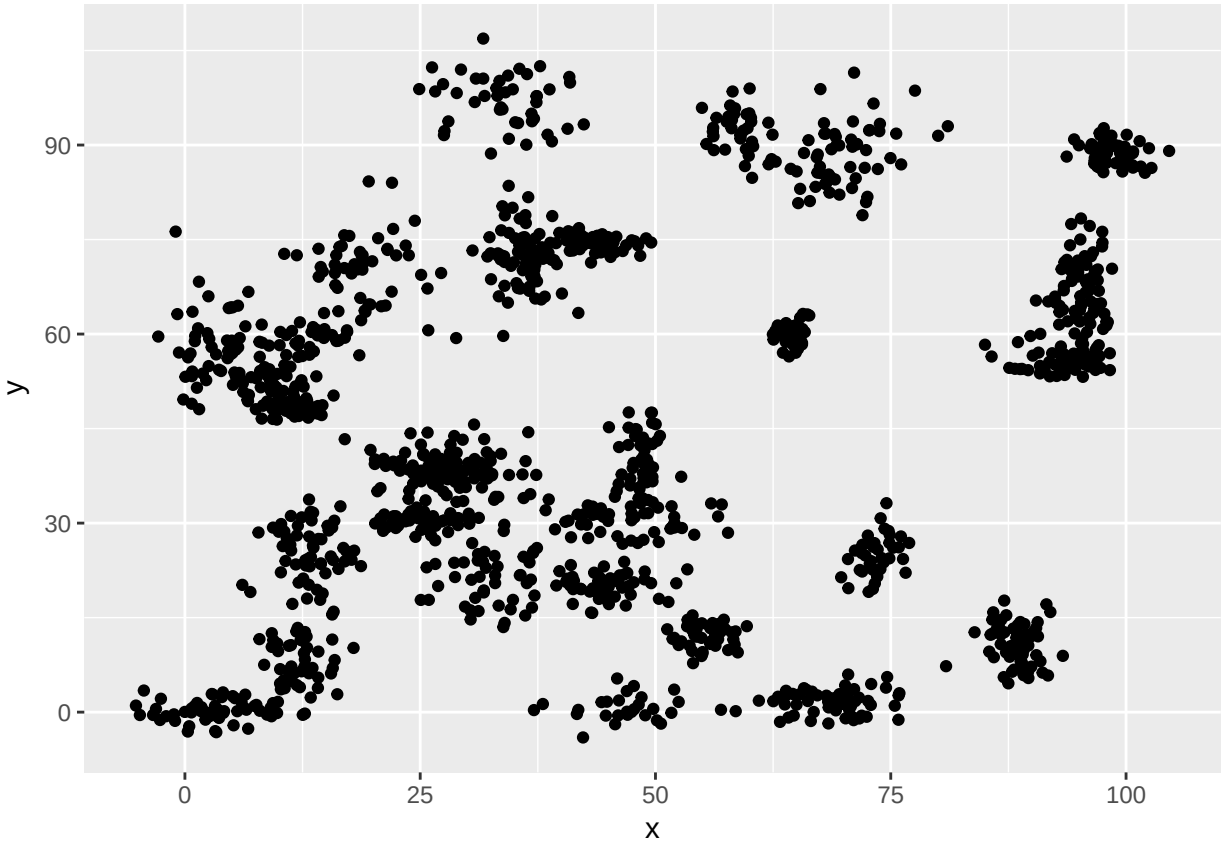
Classification Testing - KNN vs. Logistic Regression

Demond Love

Fitting a k-nearest neighbor and logistic regression model to the binary-classifier-data.csv dataset and compare the results.

```
library(readr)
library(ggplot2)
library(MLmetrics)
library(caTools)
library(ROCR)
library(formattable)
library(gmodels)
library(class)
options(scipen=999)
setwd('/Users/Love/Documents/Projects')
binary = read.csv('./binary-classifier-data.csv')
```

```
ggplot(binary, aes(x, y)) + geom_point()
```




```
binary$label = as.factor(binary$label)
split = sample.split(binary, SplitRatio = 0.8)
training = subset(binary, split =='TRUE')
testing = subset(binary, split =='FALSE')
```

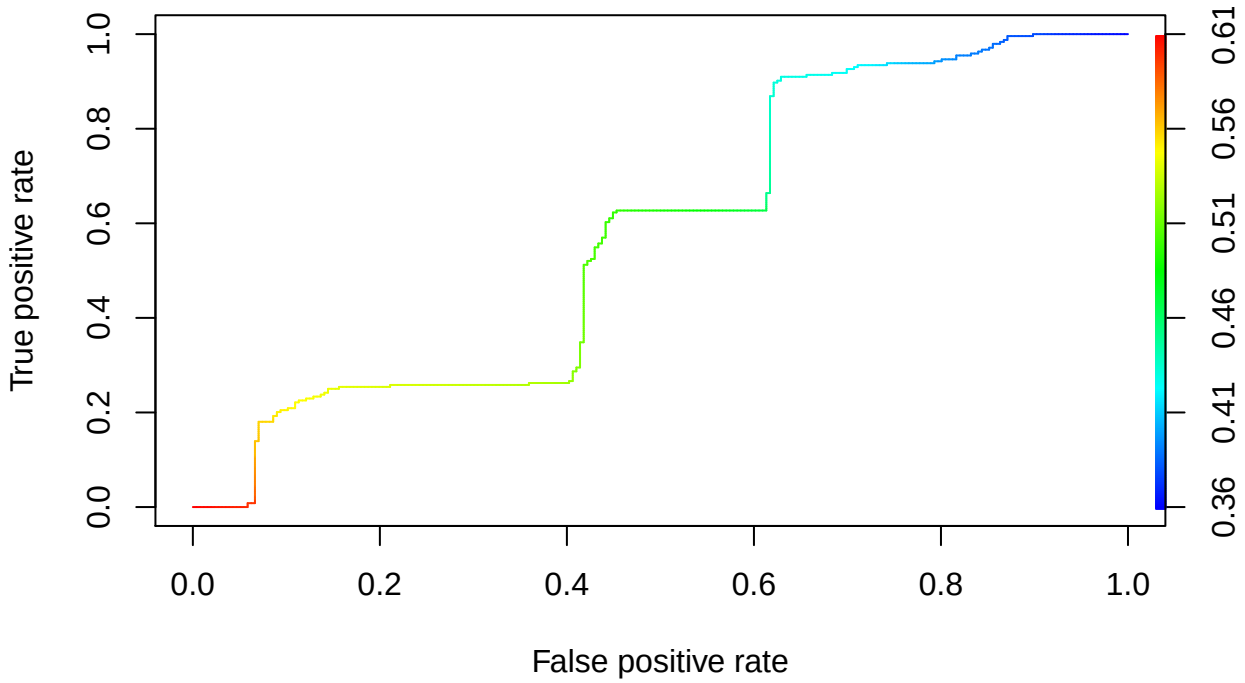
```r
logmodel = glm(label ~ x + y, data = training, family = binomial())
testingprediction = predict(logmodel, testing, type = 'response')
table(ActualValue=testing$label, PredictedValue=testingprediction>0.5)
```

```
##            PredictedValue
## ActualValue FALSE TRUE
##           0   143  113
##           1    98  146
```

```r
logmodelpercent = ((142+148)/(142+148+96+113))
percent(logmodelpercent)
```

```
## [1] 58.12%
```

```
ROCRPred = prediction(testingprediction, testing$label)
ROCRPref = performance(ROCRPred, "tpr", "fpr")
plot(ROCRPref, colorize = TRUE)
```




```
pred = ifelse(logmodel$fitted.values < 0.5, 0, 1)
F1_Score(y_pred = pred, y_true = training$label)
```

```
## [1] 0.5768463
```

```
Precision(y_pred = pred, y_true = training$label)
```

```
## [1] 0.5885947
```

```
Recall(y_pred = pred, y_true = training$label)
```

```
## [1] 0.5655577
```

```
knnmodel = knn(train = training, test = testing, cl = training$label, k = 8)
CrossTable(x = testing$label, y = knnmodel, prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  500
##
##
##                | knnmodel
## testing$label |         0 |         1 | Row Total |
## --------------|-----------|-----------|-----------|
##             0 |       252 |         4 |       256 |
##               |     0.984 |     0.016 |     0.512 |
##               |     0.981 |     0.016 |           |
##               |     0.504 |     0.008 |           |
## --------------|-----------|-----------|-----------|
##             1 |         5 |       239 |       244 |
##               |     0.020 |     0.980 |     0.488 |
##               |     0.019 |     0.984 |           |
##               |     0.010 |     0.478 |           |
## --------------|-----------|-----------|-----------|
##  Column Total |       257 |       243 |       500 |
##               |     0.514 |     0.486 |           |
## --------------|-----------|-----------|-----------|
##
##
```

```
knnpercent = (248+238)/499
percent(knnpercent)
```

```
## [1] 97.39%
```