

module1_part2

Andrew Estes

3/18/2022

```
#installing necessary libraries
```

```
library(pROC)
```

```
library(random)
```

```
library(tidyverse)
```

```
#creating random number generator function
```

```
random <- function(n){
```

```
  rando <- randomNumbers(n = n, min = -1000000000, max = 1000000000, col = 1)
```

```
}
```

```
#making the dataframe
```

```
df <- data.frame(random(1000))
```

```
colnames(df) = c('Y')
```

```
X <- seq(nrow(df))
```

```
#finalizing the dataframe
```

```
df <- df %>%
```

```
  mutate(Factor = factor(case_when(
```

```
    Y >= 0 ~ "Positive",
```

```
    Y < 0 ~ "Negative")) %>%
```

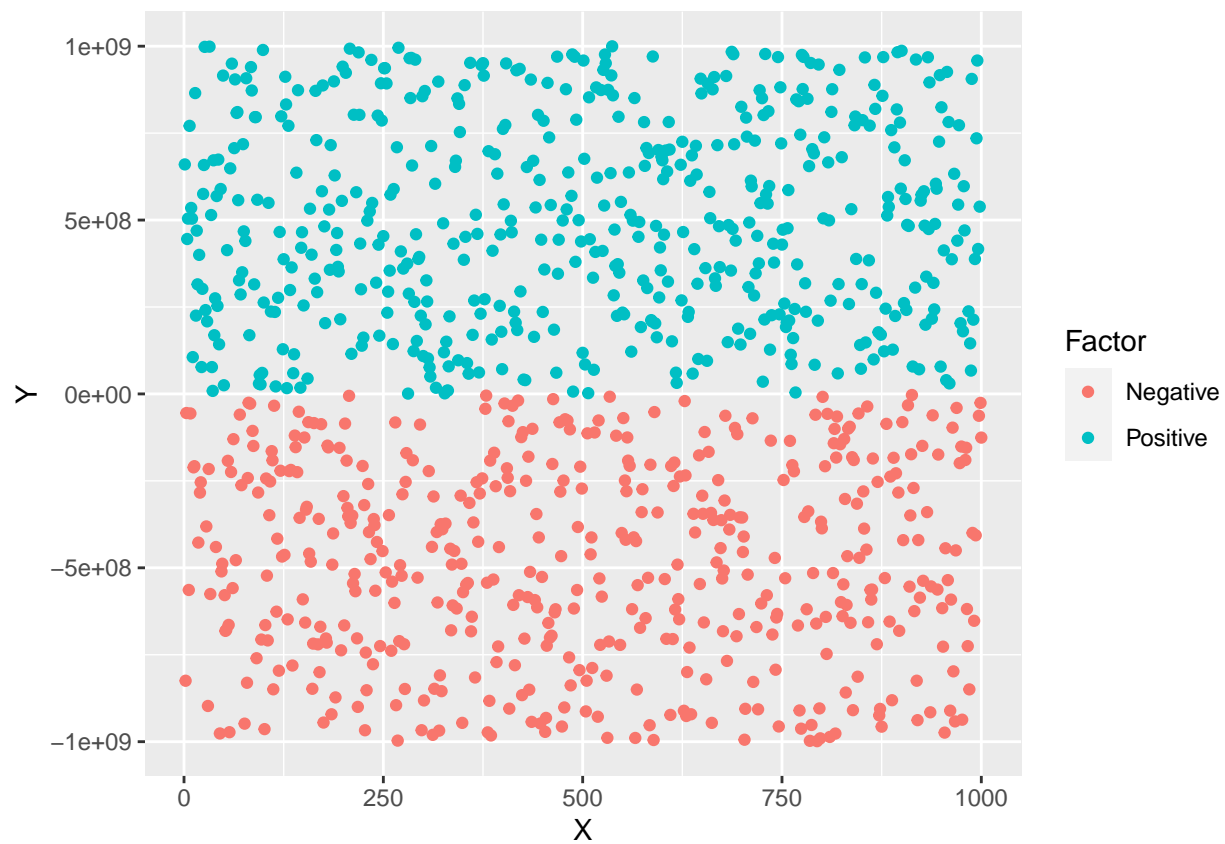
```
  mutate(X = X)
```

```
#viewing the data for understanding
```

```
head(df)
```

```
##           Y    Factor X
## 1 659968078 Positive 1
## 2 -824755452 Negative 2
## 3 -54903447  Negative 3
## 4 445474959 Positive 4
## 5 504205435 Positive 5
## 6 -563750352 Negative 6
```

```
#nice plot of output  
ggplot(df) +  
  geom_point(aes(x=X, y=Y, color=Factor))
```

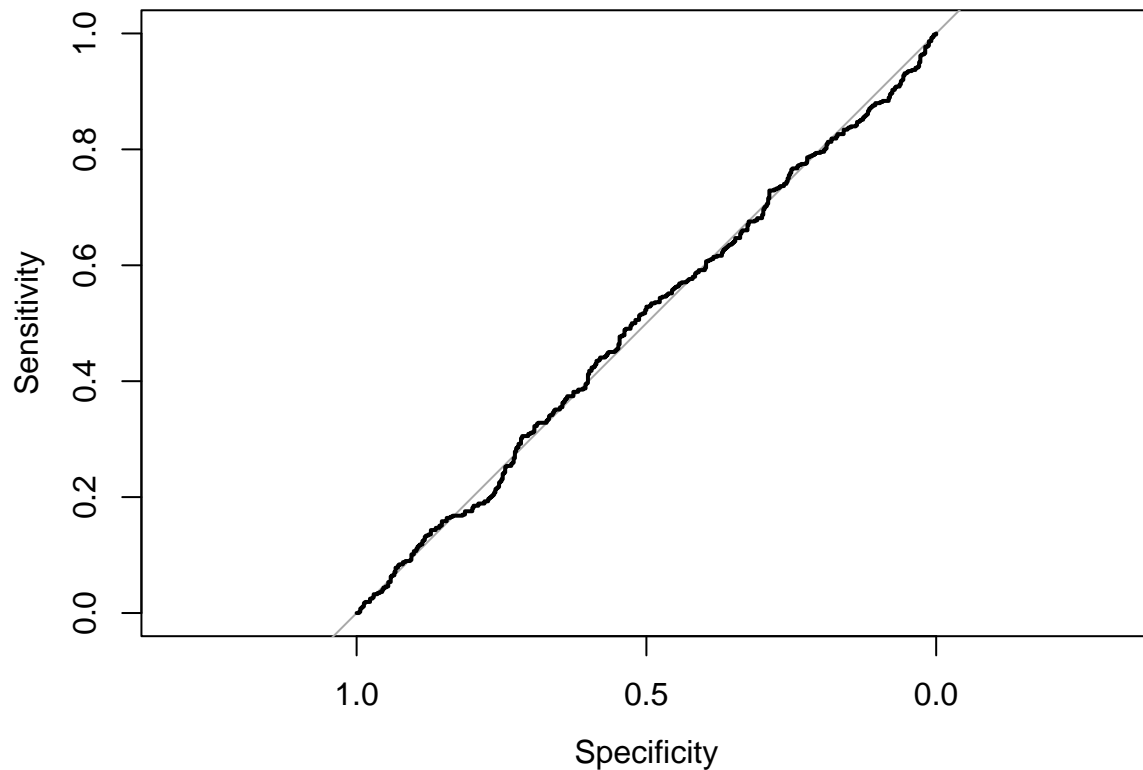


```

#logistic regression prediction
logit_predict <- predict(glm(Factor ~ X, family = binomial(link = "logit"), df))

#ROC curve analysis
roc <- roc(df$Factor, logit_predict)
plot(roc)

```



There is no relationship between the row number (as indicated in 1:n) and the positive/negative classification. This is akin to using my grades to predict the weather in Jakarta. A graph like this should be expected due to the randomness in the Response variable and the patterned change in the Predictor variable (1:n by 1).