Scam Ham Data

Mohammemd Amine Frakso

10/27/2019

##Loading the Data

scamData1 <- read.csv("C:/Users/FRAKMOH/Desktop/R working directory/scamdata1.csv", header = TRUE)  
str(scamData1)

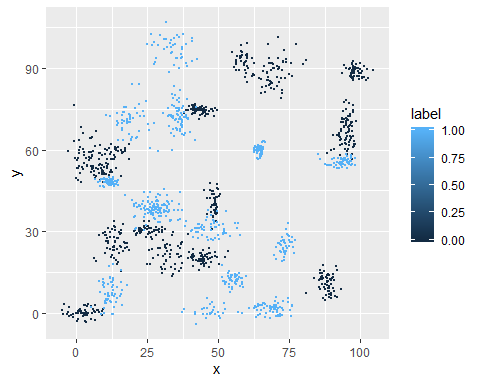
## 'data.frame': 1498 obs. of 3 variables:  
## $ label: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ x : num 70.9 75 73.8 66.4 69.1 ...  
## $ y : num 83.2 87.9 92.2 81.1 84.5 ...

scamData2 <- read.csv("C:/Users/FRAKMOH/Desktop/R working directory/scamdata2.csv", header = TRUE)  
str(scamData2)

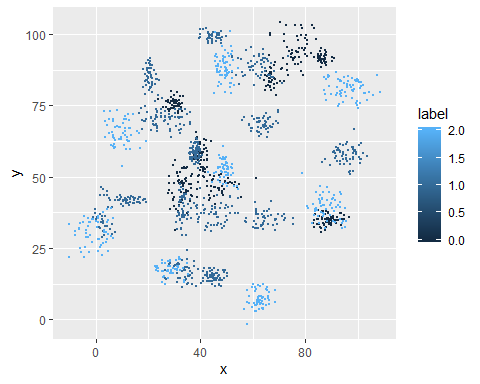
## 'data.frame': 1568 obs. of 3 variables:  
## $ label: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ x : num 30.1 31.3 34.1 32.6 34.7 ...  
## $ y : num 39.6 51.8 49.3 41.2 45.5 ...

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

library(ggplot2)  
ggplot(scamData1, aes(x=x, y=y, color = label)) + geom\_point(size=0.5)



ggplot(scamData2, aes(x=x, y=y, color = label)) +  
 geom\_point(size=0.5)

 ##Fit a k nearest neighbors model for each dataset

set.seed(20)  
clusters1 <- kmeans(scamData1[,2:3], 3)  
scamData1$Borough <- as.factor(clusters1$cluster)

library(class)  
clusterdata <- sample(1:nrow(scamData1),as.integer(0.5\*nrow(scamData1)))  
trainscamData1 <- scamData1[clusterdata,]  
testscamData1 <- scamData1[-clusterdata,]  
labelscamData1 <- scamData1[-clusterdata, 1]  
  
scamKNN <- knn(train = trainscamData1, test = testscamData1, cl = labelscamData1 , k=3)  
summary(scamKNN)

## 0 1   
## 369 380

Cant complet the assignement because I cant install some pckage: install.packages(“sklearn.datasets”) WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/> Installing package into ‘C:/Users/FRAKMOH/Documents/R/win-library/3.6’ (as ‘lib’ is unspecified) Warning in install.packages : package ‘sklearn.datasets’ is not available (for R version 3.6.1)