

This graph makes it easier to determine the distribution and the density among disparate travelers’ ages that departed from the three cities (“S”,”C”,”Q). For those different cities, it is obvious to determine the median is between 25-30. Also, this graph shows us the normal distribution for city S is between age 10 -50, for city Q is from infant – age 45, and finally for city C is quite between from 20 to 40.

/\*

train <- read.csv("/Users/Abdulaziz/Desktop/lab4/train.csv")

train <- train[which(train$Age!= 'N/A'),] ## remove the null ages

library(vioplot) ## downloading the viplot library

## assign the ages who departed from 3 cities to three variables

x1 <-train$Age[train$Embarked =='S']

x2 <- train$Age[train$Embarked == 'Q']

x3 <- train$Age[train$Embarked == 'C']

## using vioplot show the distribution between the ages and the three cities

vioplot(x1, x2, x3, names=c("S", "Q", "C"),

col="purple")