Assignment 1

Osama Bin Zahir

2023-09-08

# Source of dataset: Kaggle - <https://www.kaggle.com/datasets/iamsouravbanerjee/airline-dataset?resource=download>

# Loading required packages

library(readxl)  
library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

# Loading the dataset

airline <- read.csv("C:\\Users\\Osama Zahir\\Desktop\\airline dataset.csv")

# Printing descriptive statistics

summary(airline)

## Passenger.ID First.Name Last.Name Gender   
## Min. :10000 Length:98619 Length:98619 Length:98619   
## 1st Qu.:32608 Class :character Class :character Class :character   
## Median :55338 Mode :character Mode :character Mode :character   
## Mean :55169   
## 3rd Qu.:77695   
## Max. :99999   
## Age Nationality Airport.Name Airport.Country.Code  
## Min. : 1.0 Length:98619 Length:98619 Length:98619   
## 1st Qu.:23.0 Class :character Class :character Class :character   
## Median :46.0 Mode :character Mode :character Mode :character   
## Mean :45.5   
## 3rd Qu.:68.0   
## Max. :90.0   
## Country.Name Airport.Continent Continents Departure.Date   
## Length:98619 Length:98619 Length:98619 Length:98619   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Arrival.Airport Pilot.Name Flight.Status   
## Length:98619 Length:98619 Length:98619   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

# Converting categorical variable into categorical variable

gender\_mapping = c("Female" = 0, "Male" = 1)  
airline$Gender = gender\_mapping[airline$Gender]  
unique(airline$Gender)

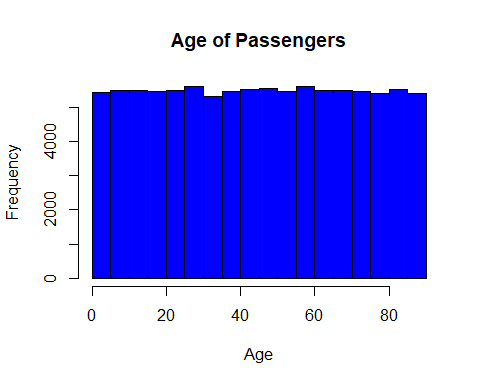
## [1] 0 1

head(airline)

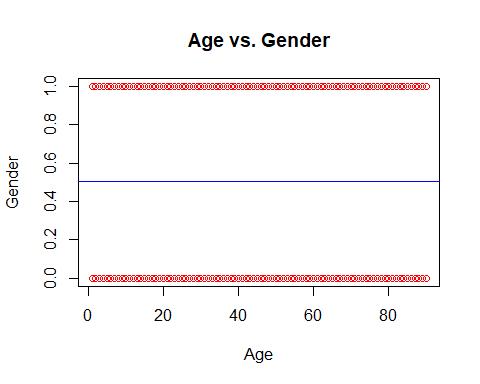
## Passenger.ID First.Name Last.Name Gender Age Nationality  
## 1 10856 Edithe Leggis 0 62 Japan  
## 2 43872 Elwood Catt 1 62 Nicaragua  
## 3 42633 Darby Felgate 1 67 Russia  
## 4 78493 Dominica Pyle 0 71 China  
## 5 82072 Bay Pencost 1 21 China  
## 6 39630 Lora Durbann 0 55 Brazil  
## Airport.Name Airport.Country.Code Country.Name  
## 1 Coldfoot Airport US United States  
## 2 Kugluktuk Airport CA Canada  
## 3 Grenoble-Isère Airport FR France  
## 4 Ottawa / Gatineau Airport CA Canada  
## 5 Gillespie Field US United States  
## 6 Coronel Horácio de Mattos Airport BR Brazil  
## Airport.Continent Continents Departure.Date Arrival.Airport Pilot.Name  
## 1 NAM North America 6/28/2022 CXF Edithe Leggis  
## 2 NAM North America 12/26/2022 YCO Elwood Catt  
## 3 EU Europe 1/18/2022 GNB Darby Felgate  
## 4 NAM North America 9/16/2022 YND Dominica Pyle  
## 5 NAM North America 2/25/2022 SEE Bay Pencost  
## 6 SAM South America 6/10/2022 LEC Lora Durbann  
## Flight.Status  
## 1 On Time  
## 2 On Time  
## 3 On Time  
## 4 Delayed  
## 5 On Time  
## 6 On Time

# Plotting Histogram and Scatterplot

hist(airline$Age, main = "Age of Passengers", xlab = "Age", ylab = "Frequency", col = "blue", border = "black")



plot(airline$Age, airline$Gender, main = "Age vs. Gender", xlab = "Age", ylab = "Gender", col = "red")  
abline(lm(airline$Gender~airline$Age), col = "blue")



# A Scatterplot of Age vs. Gender was created because age was the only usable numerical variable and since the gender variable was converted to numerical, it was used to plot Age vs. Gender.