

# Project 3

Group 3

4/4/2021

Group Leader: Brianna Johnson

Member Names: Marco Sousa, Ben Pfeffer, Nikita Seleznev, Brianna Johnson

## Introduction to the US Census Binary Income Dataset

Credited to Ronny Kohavi and Barry Becker, The 1994 United States Census Bureau data has several features describing an individual to predict in a binary manner whether that individual will earn more or less than \$50,000 per year.

## [1] 30162

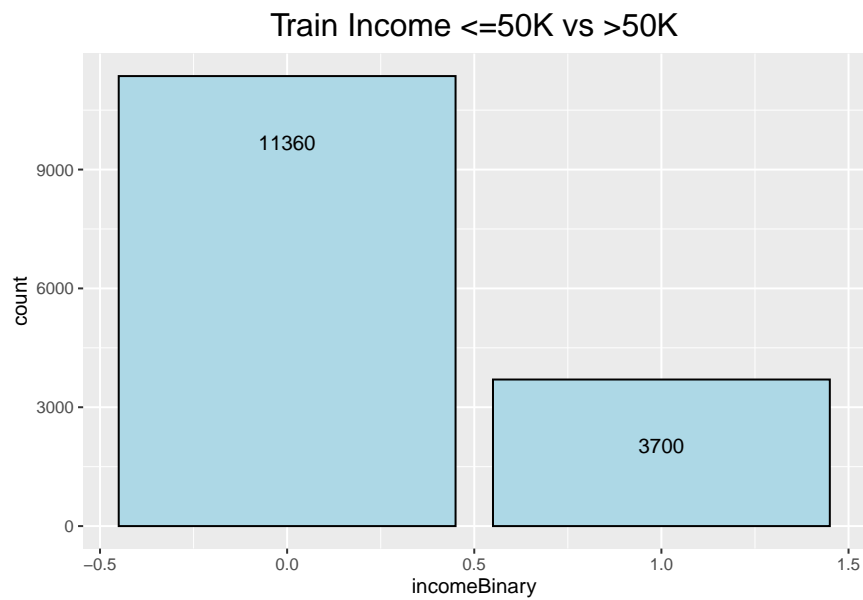
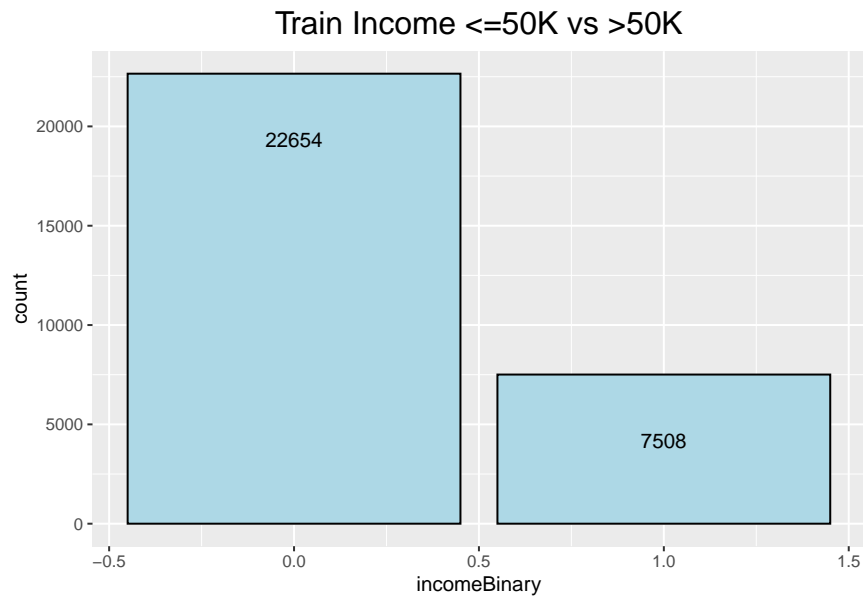
## [1] 15060

The training data is displayed as follows.

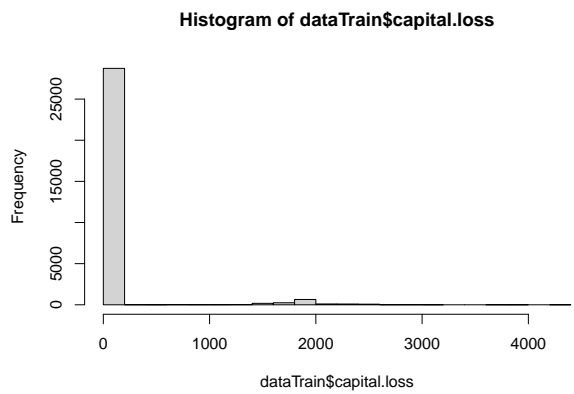
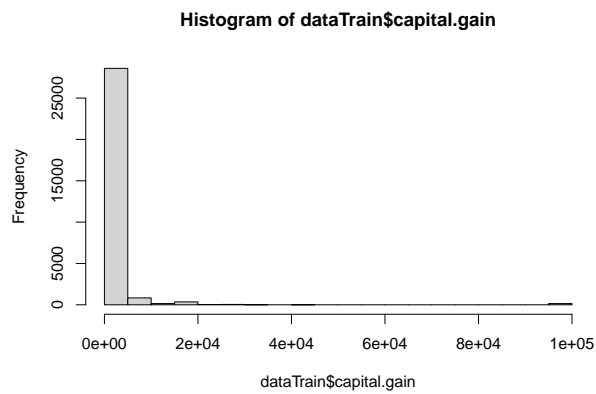
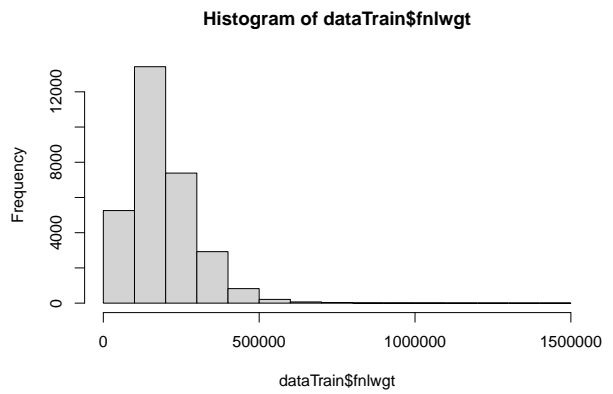
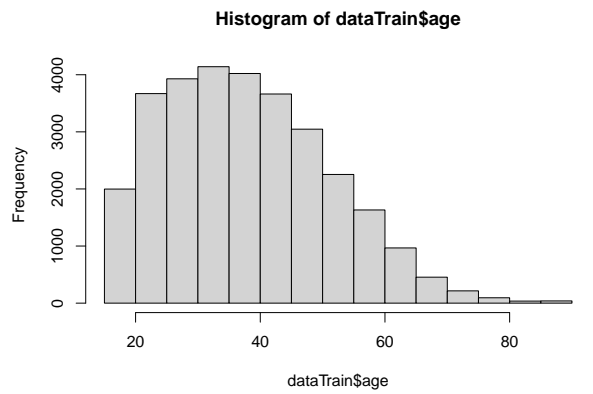
```
##   age      workclass fnlwgt education education.num      marital.status
## 1  39      State-gov  77516 Bachelors           13      Never-married
## 2  50 Self-emp-not-inc 83311 Bachelors           13 Married-civ-spouse
## 3  38      Private 215646   HS-grad            9      Divorced
## 4  53      Private 234721    11th              7 Married-civ-spouse
## 5  28      Private 338409 Bachelors           13 Married-civ-spouse
## 6  37      Private 284582  Masters            14 Married-civ-spouse
##      occupation relationship race      sex capital.gain capital.loss
## 1   Adm-clerical Not-in-family White   Male      2174          0
## 2   Exec-managerial      Husband White   Male          0          0
## 3 Handlers-cleaners Not-in-family White   Male          0          0
## 4 Handlers-cleaners      Husband Black   Male          0          0
## 5   Prof-specialty      Wife Black Female          0          0
## 6   Exec-managerial      Wife White Female          0          0
##  hours.per.week native.country income incomeBinary
## 1           40 United-States <=50K          0
## 2           13 United-States <=50K          0
## 3           40 United-States <=50K          0
## 4           40 United-States <=50K          0
## 5           40      Cuba <=50K          0
## 6           40 United-States <=50K          0
```

# Exploratory Data Analysis

## Binary Income Count



## Quick Histograms



EDA Next

Logistic Regression

Naive Bayes