

AMERICAN INTERNATIONAL UNIVERSITY–BANGLADESH

(AIUB)

FACULTY OF SCIENCE & TECHNOLOGY

# INTRODUCTION TO DATA SCIENCE

**Section: C**

**Group: 05**

**Supervised By**

## TOHEDUL ISLAM

**Submitted By**

|  |  |  |
| --- | --- | --- |
| **Name** |  | **Id** |
| **Nafisa Sharmin Sadia** |  | **21-45162-2** |

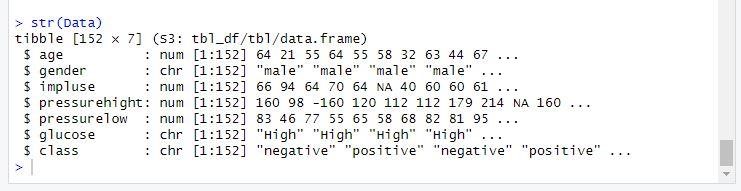
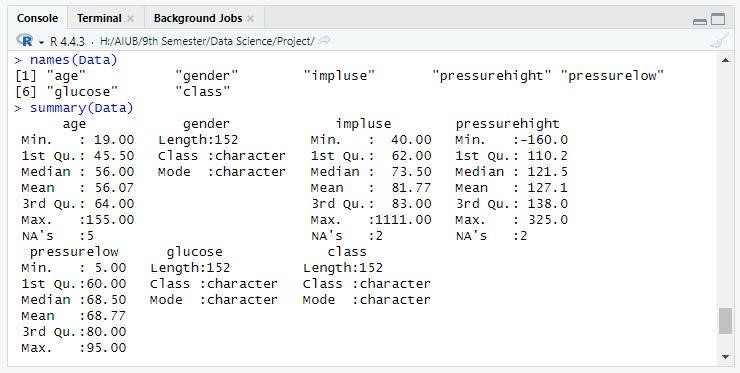
**Heart Disease Classification Dataset:**

The dataset examines the relationship between heart attack occurrences and various demographic and medical factors. Key attributes include age, gender, heart rate (impulse), systolic and diastolic blood pressure (high pressure and low pressure), and blood sugar level (glucose). The outcome variable indicates the presence or absence of a heart attack. This dataset facilitates analyzing how physiological and lifestyle-related factors influence cardiovascular health, identifies heart attack risk patterns, and informs preventive healthcare strategies. • **Description:** This is the summary of the dataset

**Code:**



**Output:**

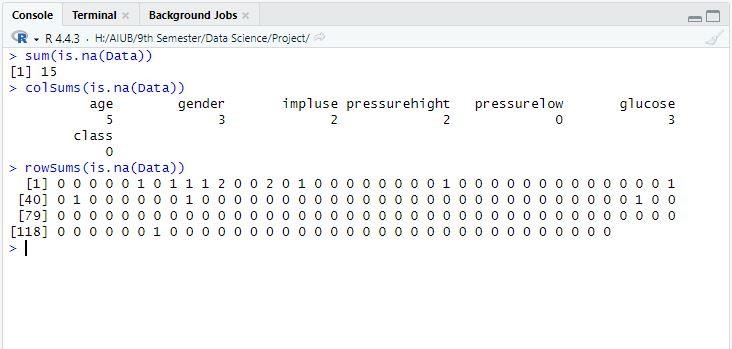
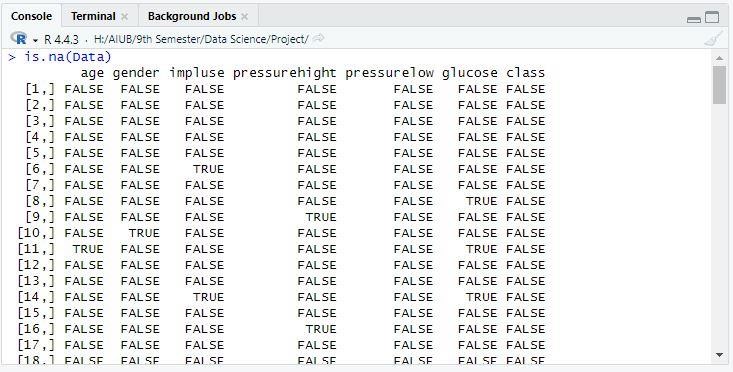


* **Description:** To show the values that are missing from the dataset

**Code:**

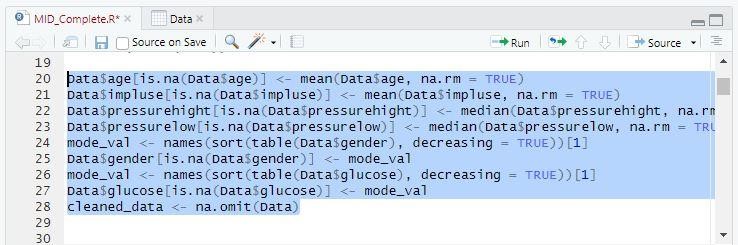


**Output:**

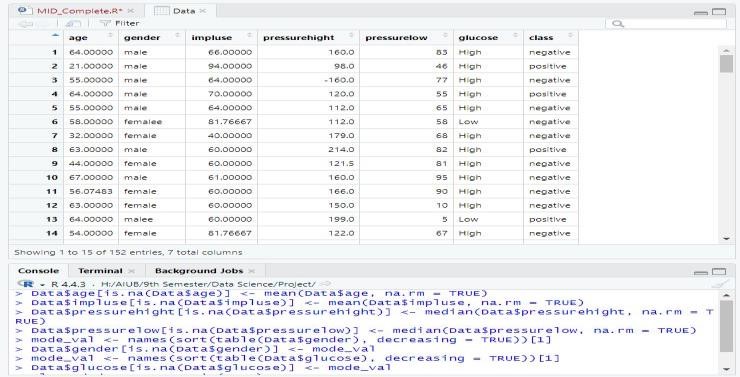


* **Description:** To handle the values that are missing

**Code:**



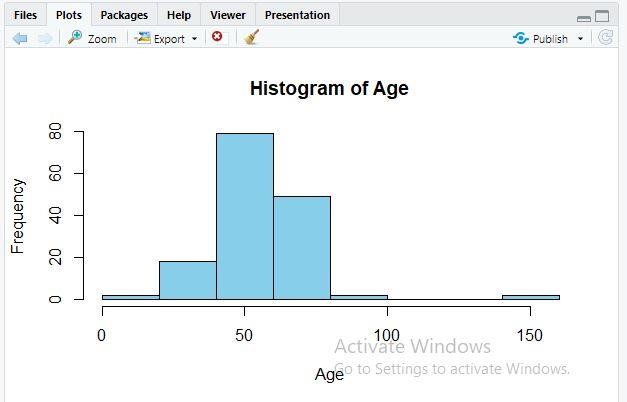
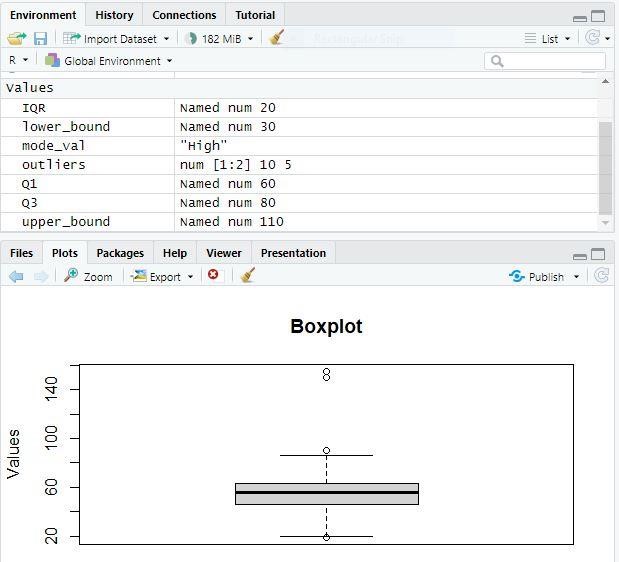
**Output:**



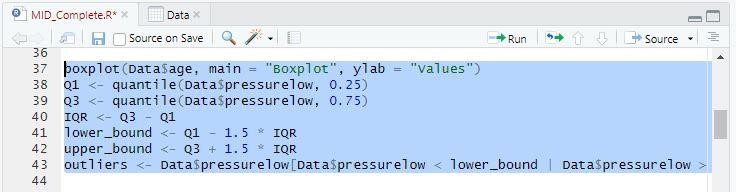
To show missing values on the graph **Code:**



**Output:**



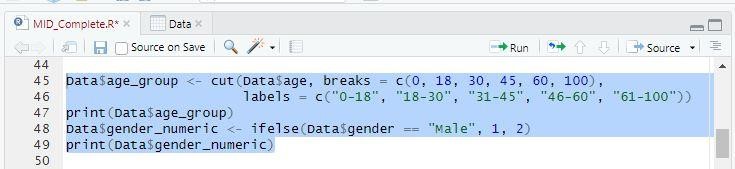
**Description:** To show outliers in the dataset and handle the outliers **Code:**



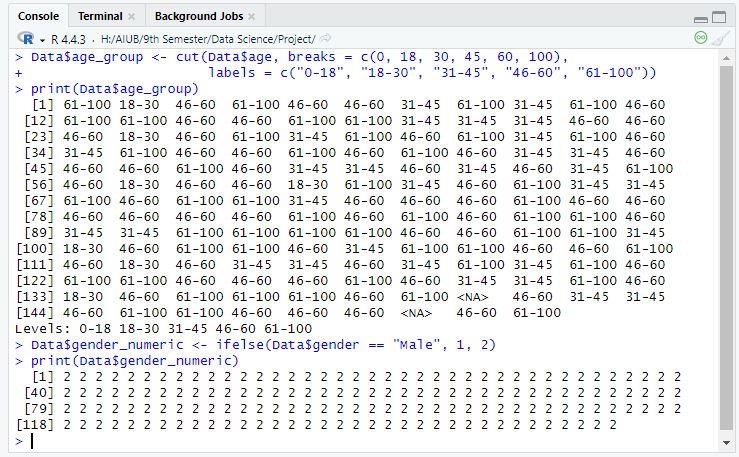
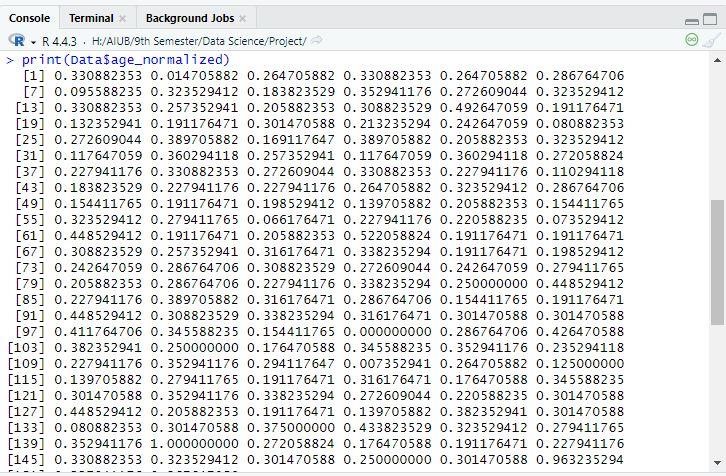
**Output:**

Convert attributes from numeric to categorical or categorical to numeric

**Code:**

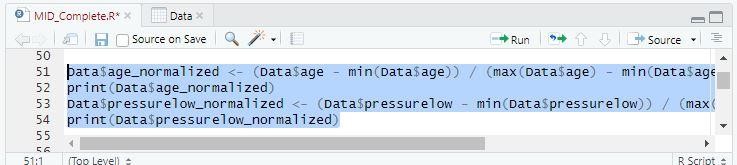


**Output:**



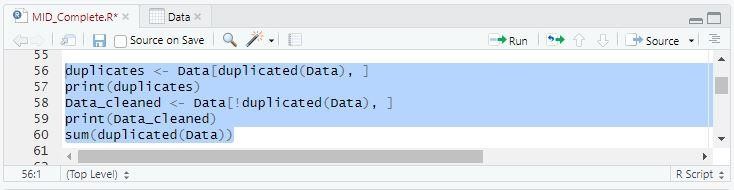
* **Description:** Apply the normalization method for any continuous attribute

**Code:**

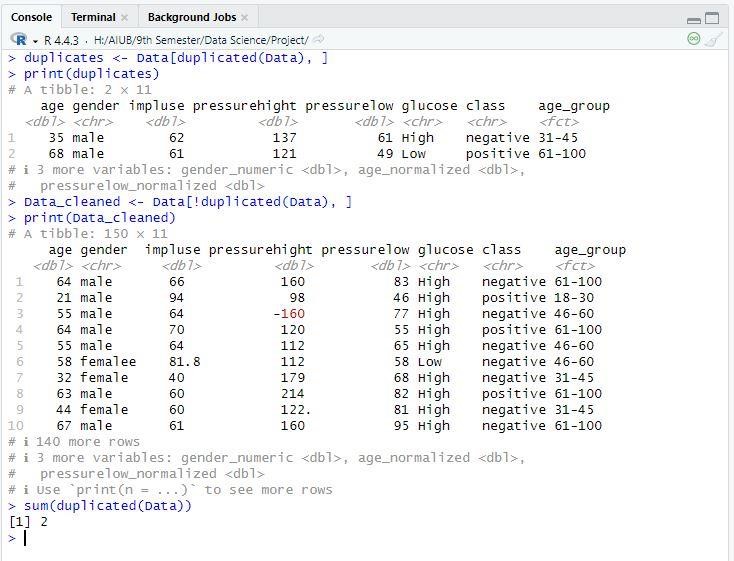


**Output:**

To find and remove duplicate values **Code:**

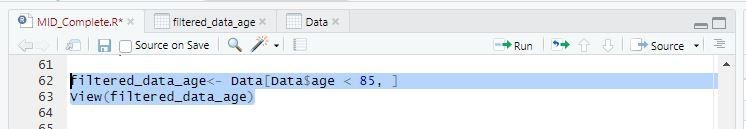


**Output:**

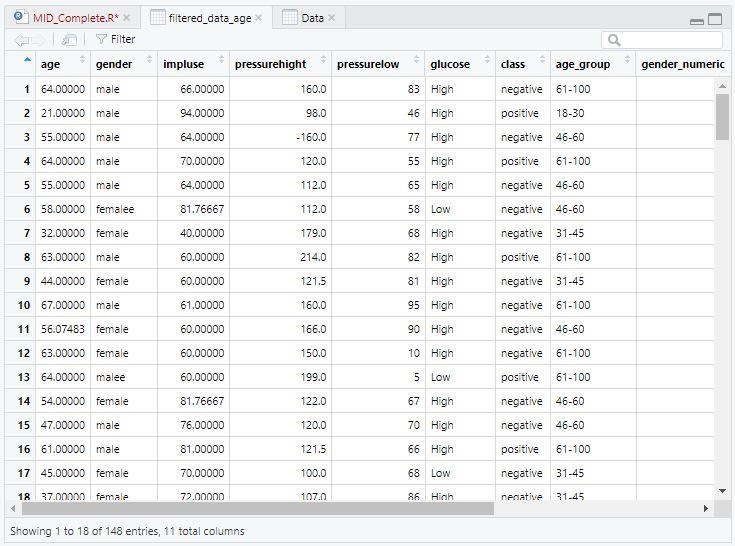


* **Description:** Apply filtering methods to filter the data.

**Code:**

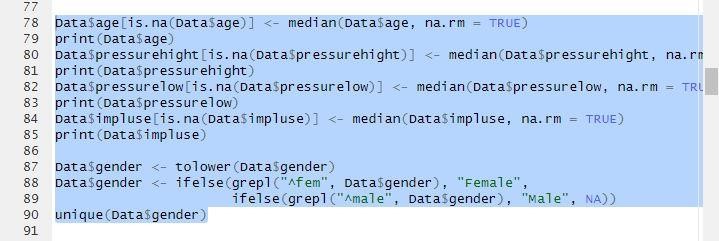
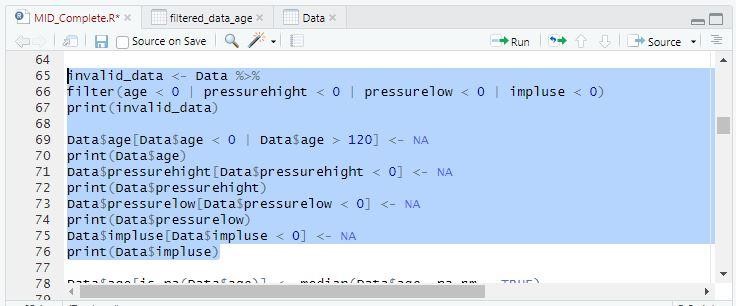


**Output:**

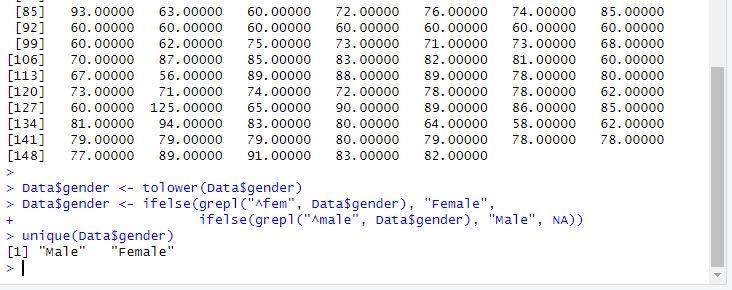
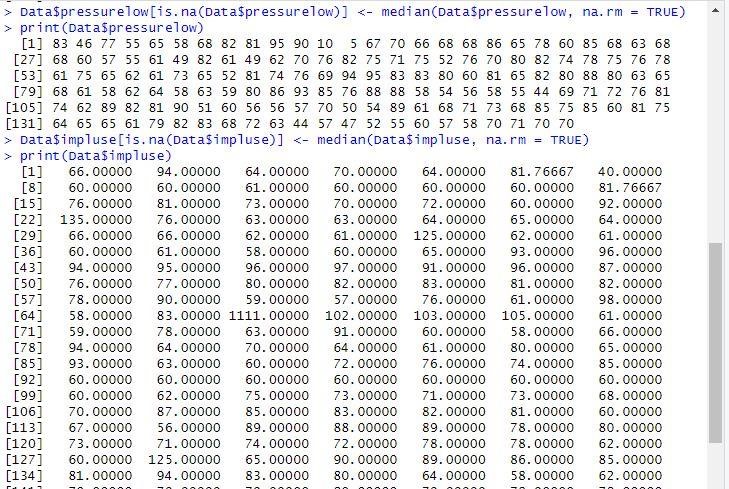
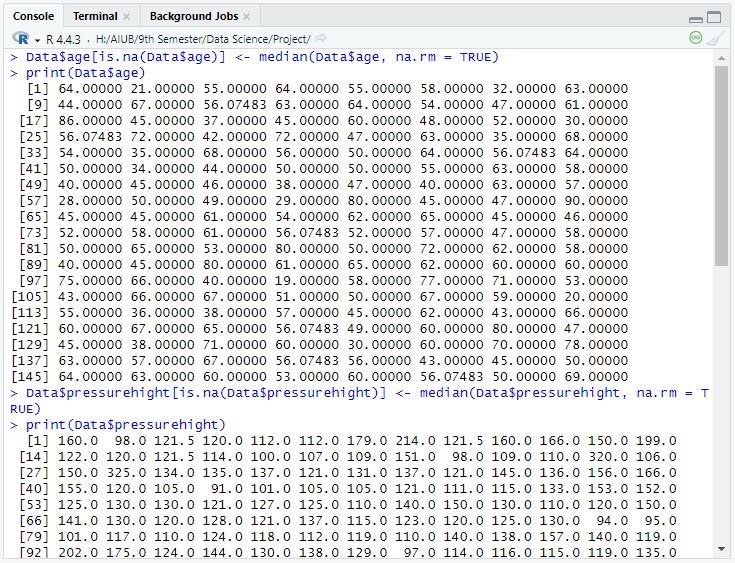
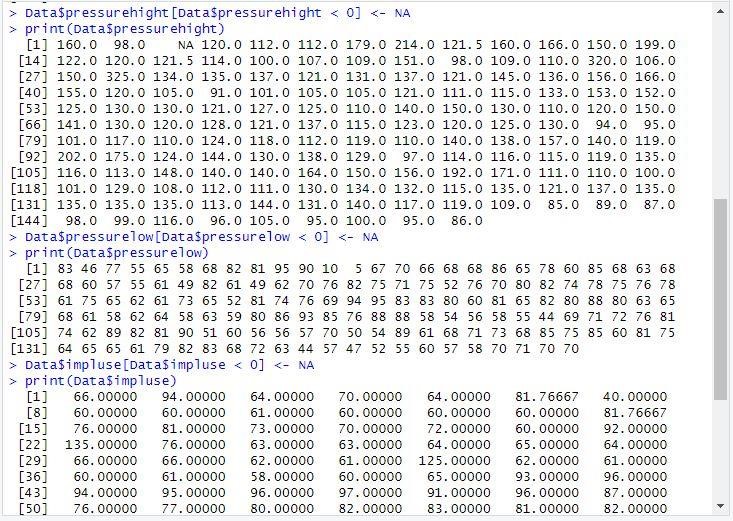
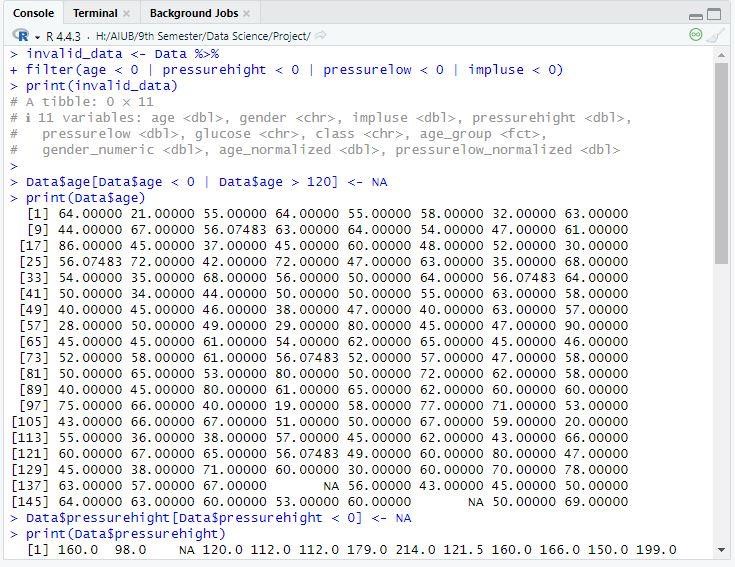


Detect invalid data in the data set and handle those values

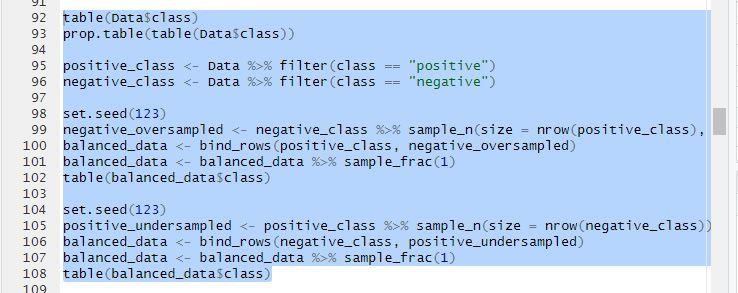
**Code:**



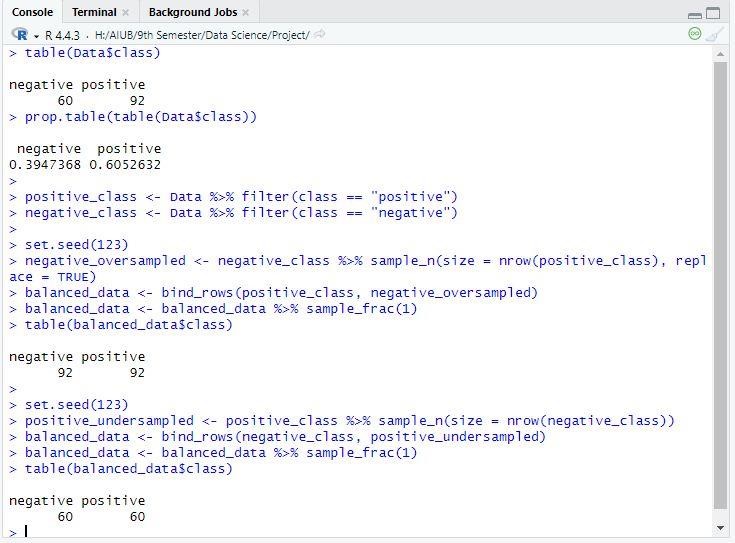
**Output:**



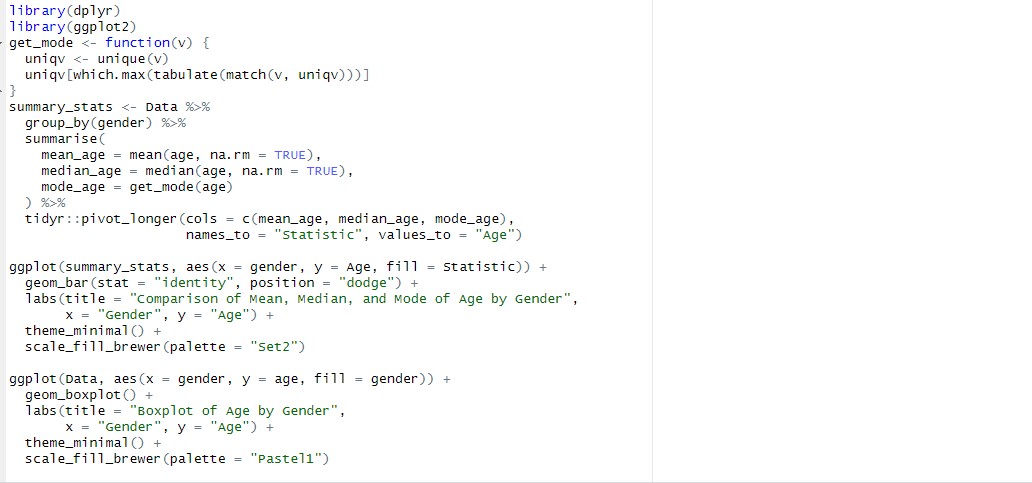
* **Description:** Convert the imbalanced data set into the balanced data set **Code:**



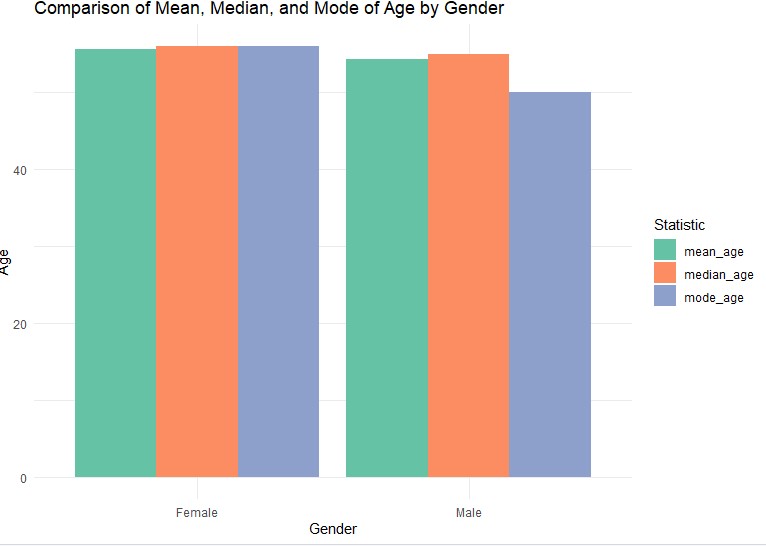
**Output:**



* **Description:** Split the dataset for Training and Testing **Code:**

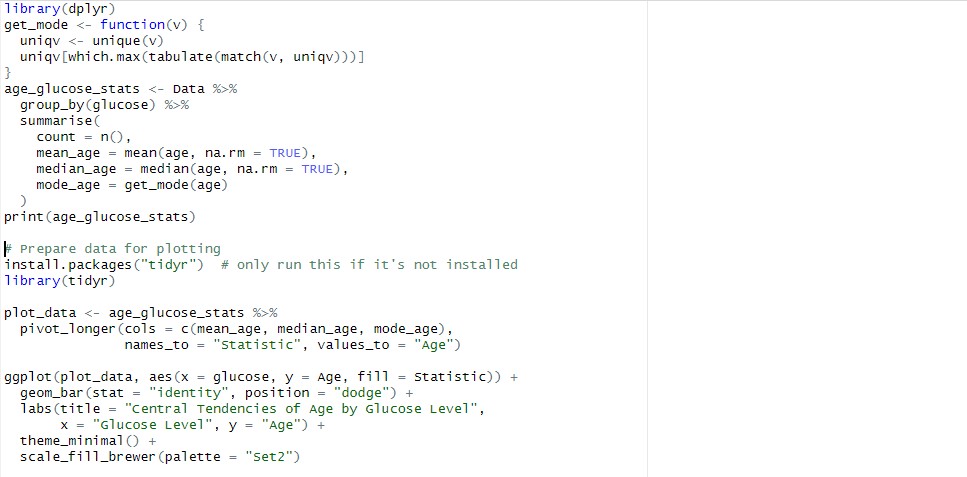


**Output:**

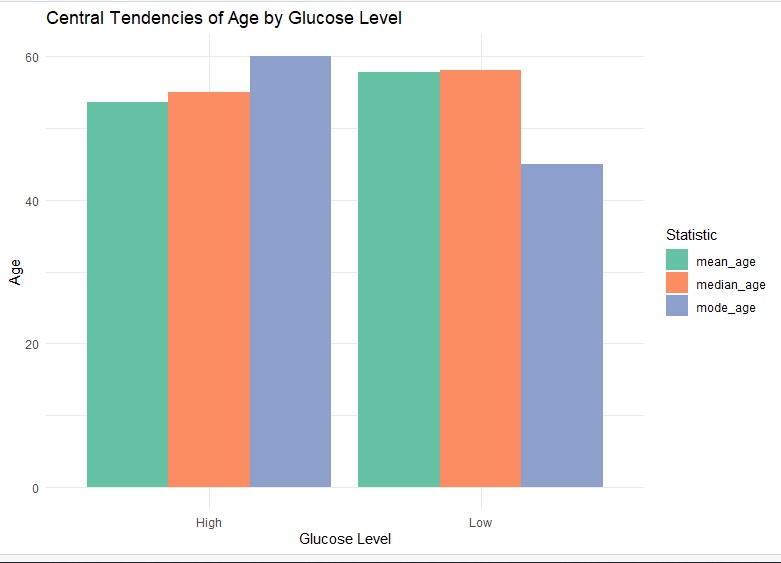


* **Description:** Comparing the central tendency of age across different gender groups using the mean, median, and mode.

**Code:**

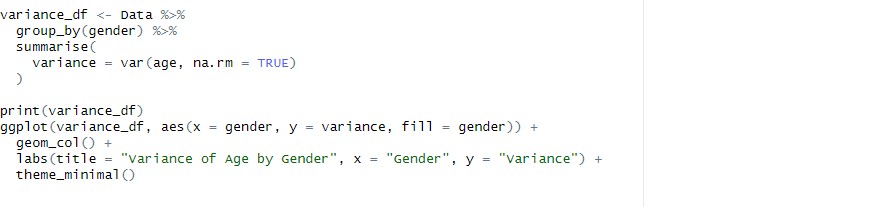
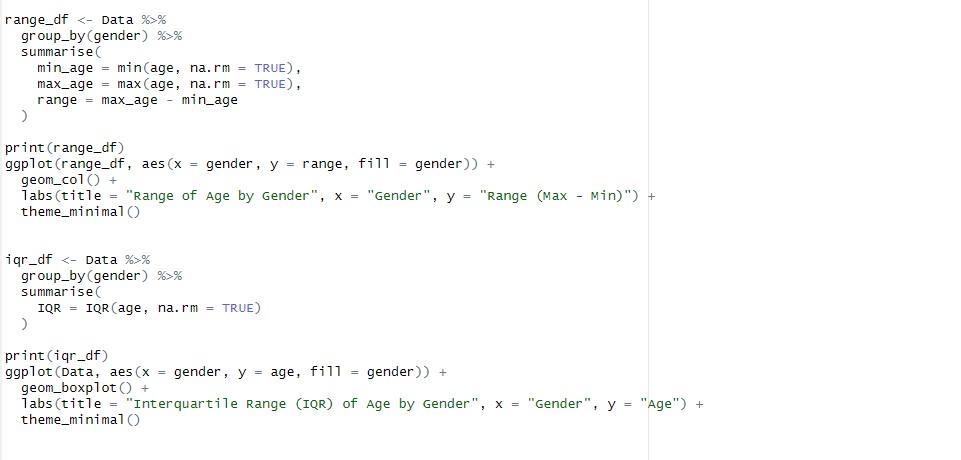


**Output:**

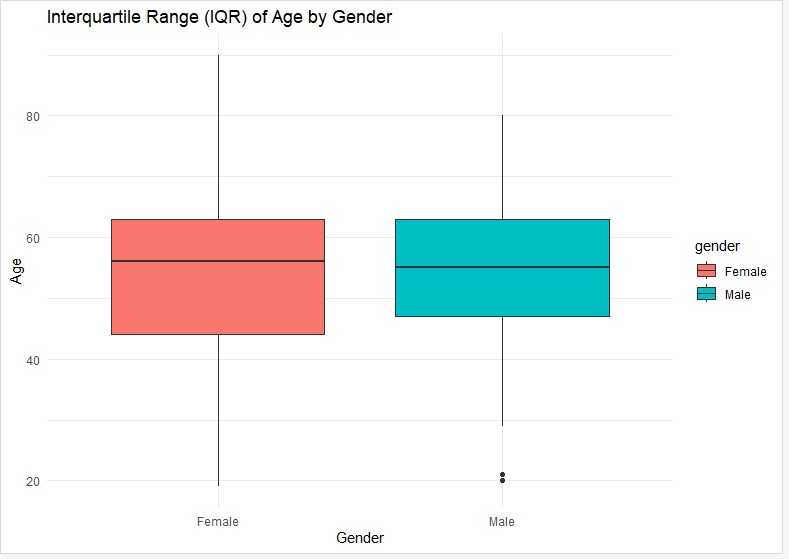
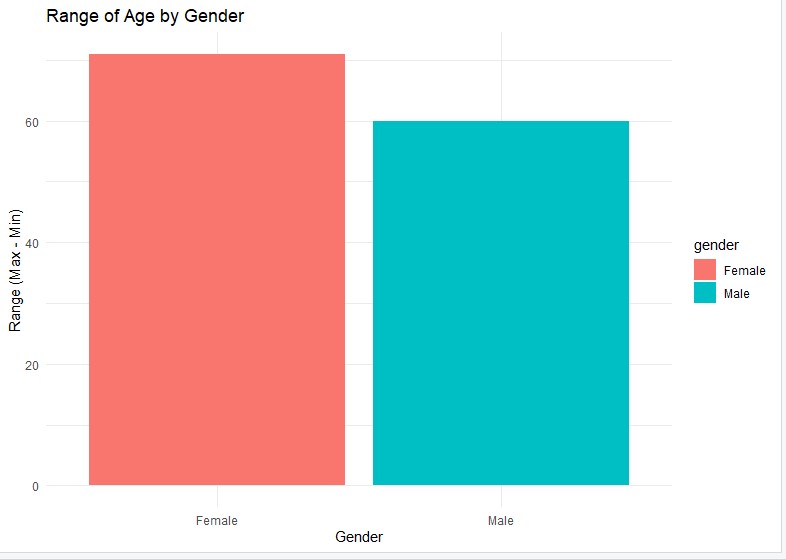
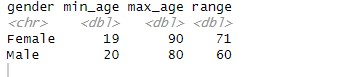


**Description:** Comparing the age’s central tendency across glucose levels using the mean, median, and mode

**Code:**



**Output:**



**Description:** Comparing the spread of Age across different groups of gender using the Range,

IQR, and Variance