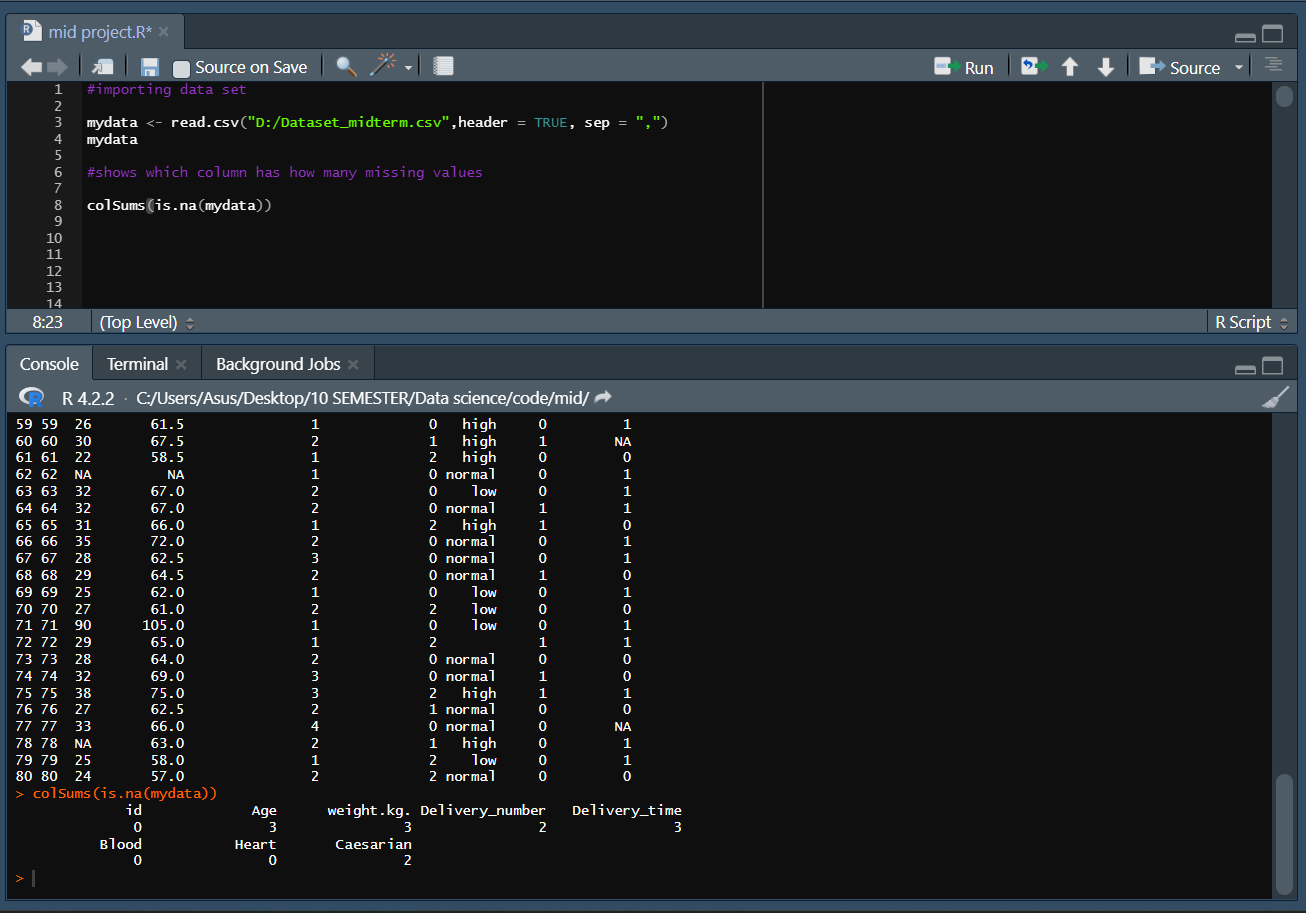
**Import dataset:**

Here I imported the dataset. After importing the dataset, I saw that there were some missing values, abnormal values, categorical and numerical values.

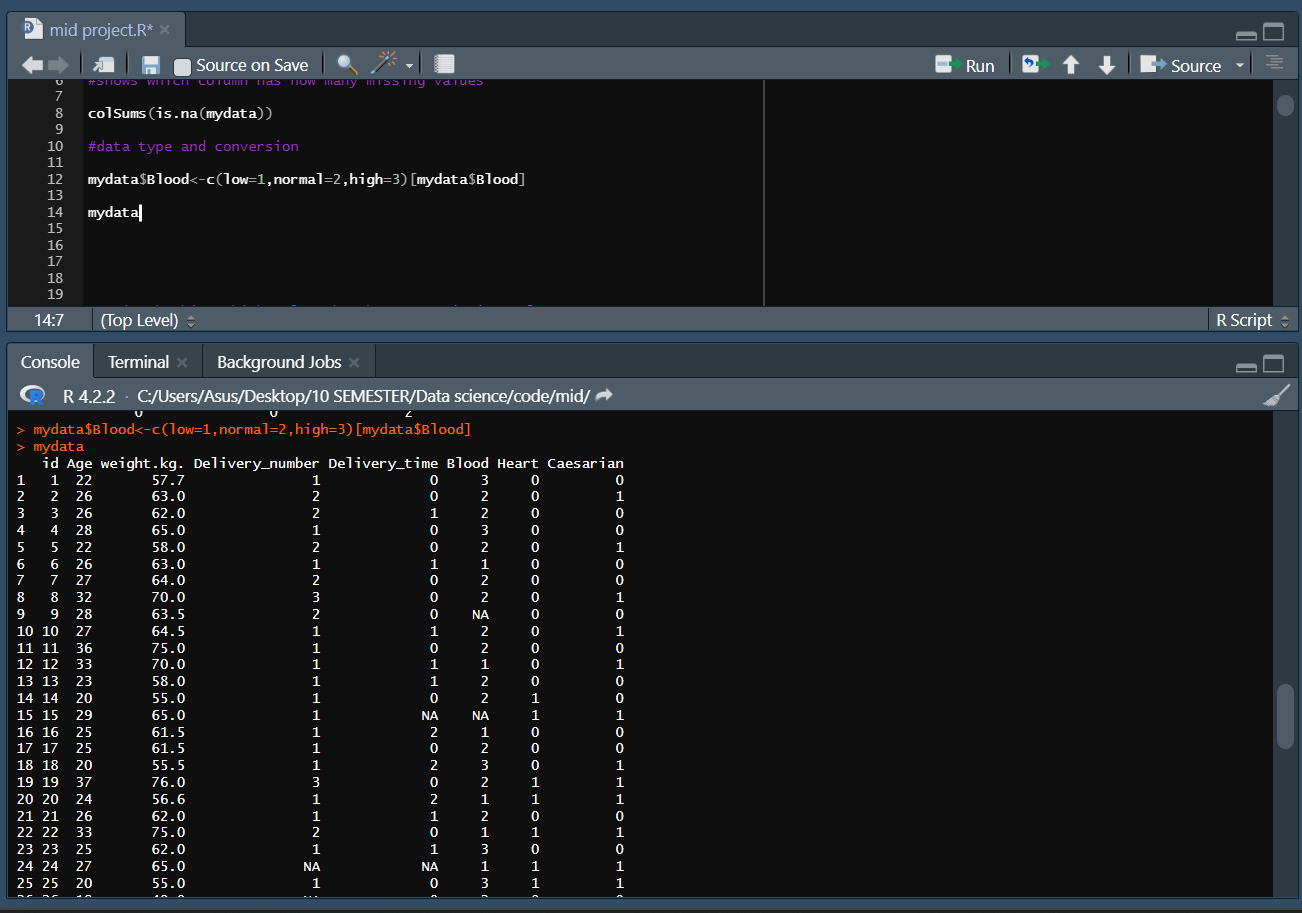
****

**DATA PREPARATION**

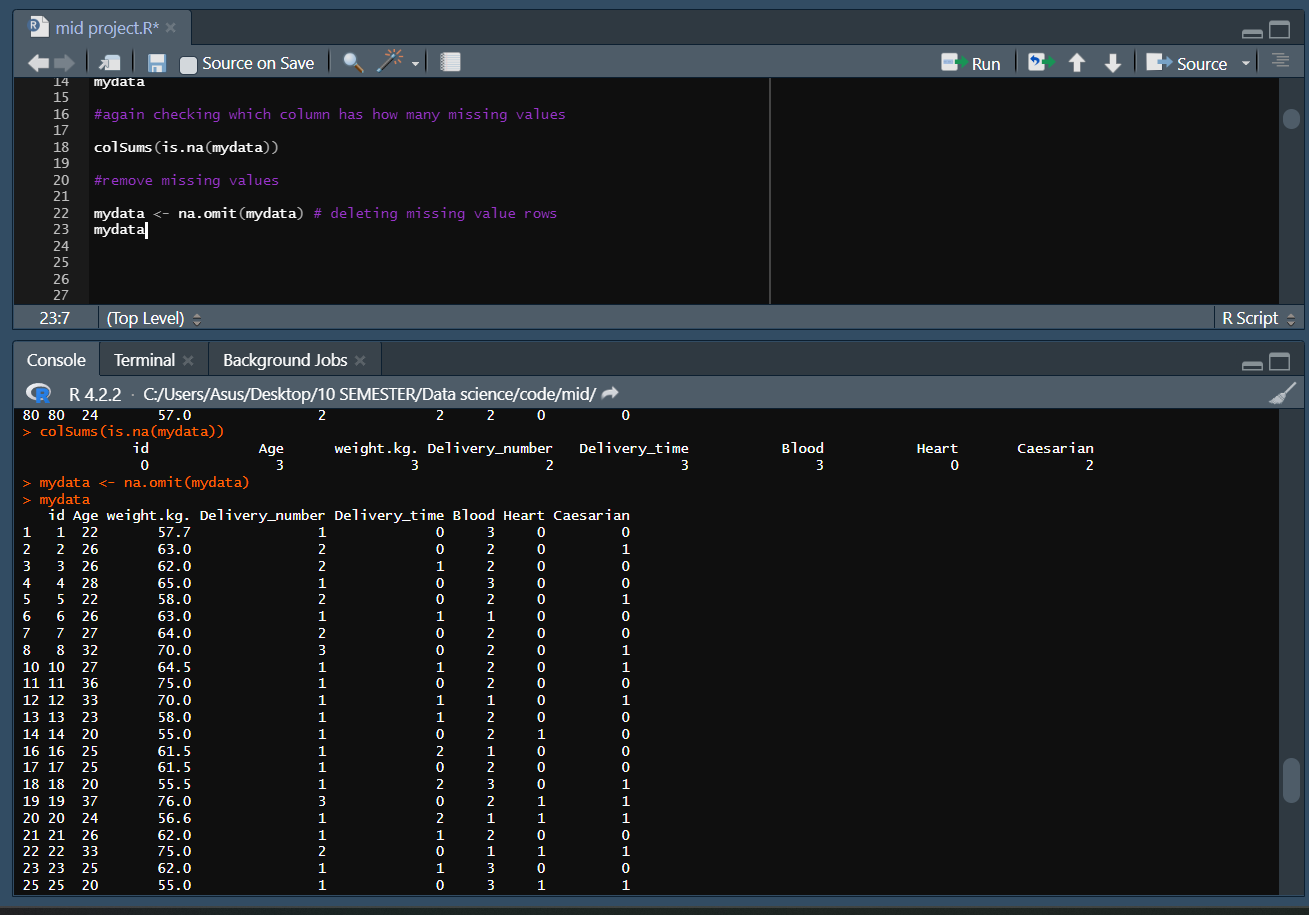
For this given dataset I have applied 3 steps. Those are given bellow:

1. Data type conversion
2. Missing value
3. Outlier

**Data type Conversion:**

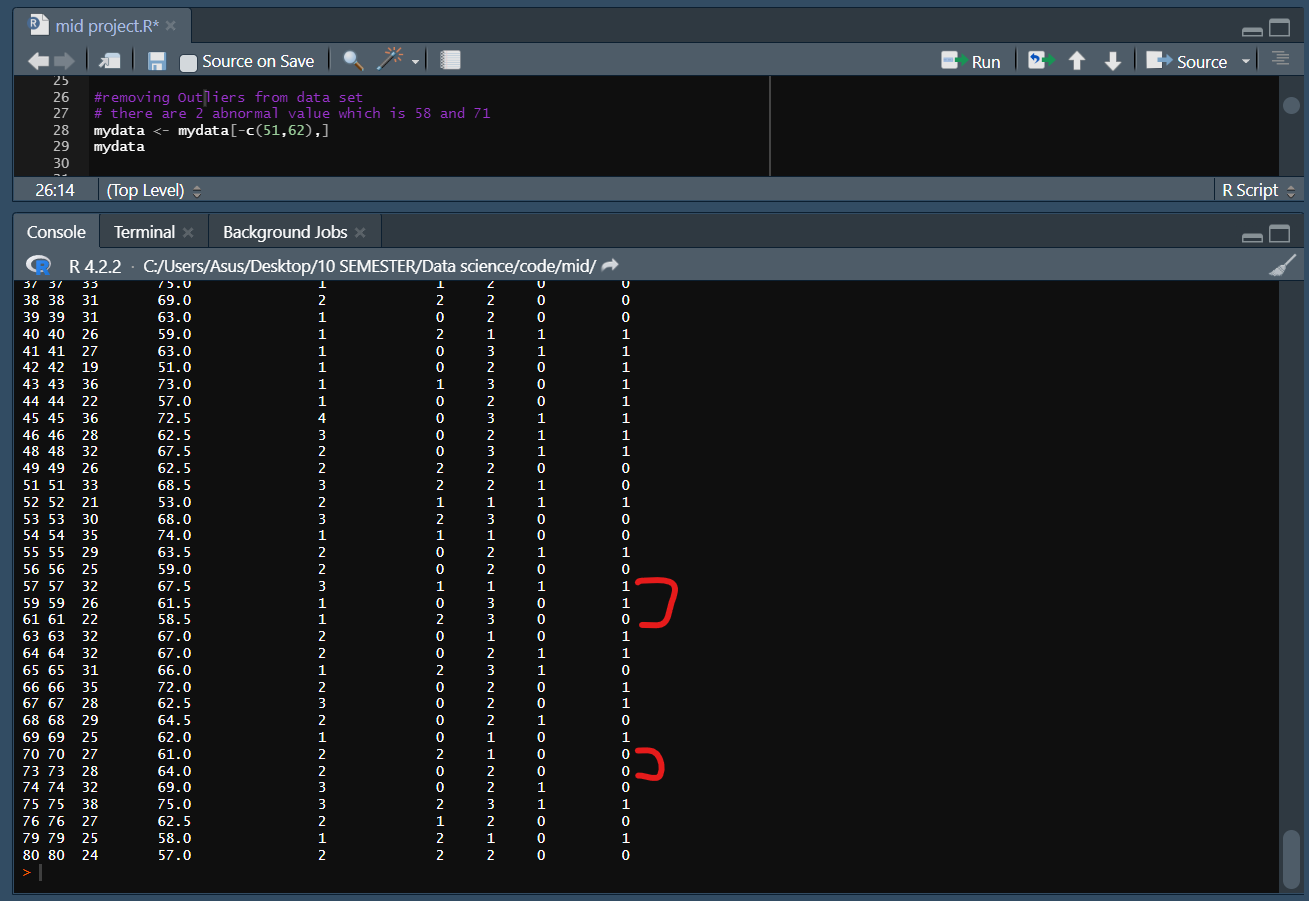
****Here blood attribute entries were categorical, and I converted them into numerical entries. For low blood pressure I applied 1, similarly for normal blood pressure = 2 and high blood pressure = 3. For this conversion the data set converted into full numerical. We can also convert a dataset numeric to categoric.

**Missing values:**

****After converting blood attribute categoric to numeric, I checked where was the missing values and saw that age entries had 3 missing values, weight entries had 3, delivery number entries had 2, delivery time entries had 3, blood entries had 3, caesarian entries had 2 missing values and id and heart entries had no missing values. So, I discarded those missing instances from the dataset.

**Outliers:**

Outliers were found in the 58th and 71st number rows, which had 95 and 90 for age and 110 and 105 for weight, respectively. So, I discarded those rows. Here I deleted 51 and 62 number rows because there were missing rows in between as I removed them earlier. After reducing outliers output was showing without 58 and 71 number rows which is marked with red on the screenshot.



**Univariate Exploration**

A screenshot of a computer

Description automatically generatedAfter preparing the dataset, I moved toward Univariate Exploration**.** I applied functions to estimate min, max, mean, and standard deviation. The dataset was evaluated by estimating the minimum, maximum, average value, and standard deviation for each of the dataset attributes. Here, for age min = 19, max = 40, deviation = 4.812 and mean = 28.21. Similarly for weight min = 51, max = 82, deviation = 6.34 and mean = 64.33. For delivery number min = 1, max = 4, deviation = 0.78 and mean = 1.7. For delivery time min = 0, max = 2, deviation = 0.83 and mean = 0.67. For blood min = 1, max = 3, deviation = 0.69 and mean = 2.05. For heart min = 0, max = 1, deviation = 0.48 and mean = 0.36 and lastly for caesarian min = 0, max = 1, deviation = 0.503 and mean = 0.53.

**Text

Description automatically generated**

**Summary of the data set**

At last, I applied summary function to summaries my data set and check mean, median, min, max and quadrate values each of the attributes of the dataset.

