Data

The credit data used in this project includes both categorical and quantitative data.

Quantitative variables are

- Income
- Limit
- Rating
- Cards: Number of Credit Cards
- Age
- Education(years), and
- Balance

and Categorical variables include

- Gender: Female or Male
- Student: Yes or No
- Married: Yes or No, and
- Ethnicity: African, American, Asian, or Caucasian

To take a closer look at each variables included in the **Credit** dataset, we looked at the Distributions of quantitative variables are shown below in Figure 1.

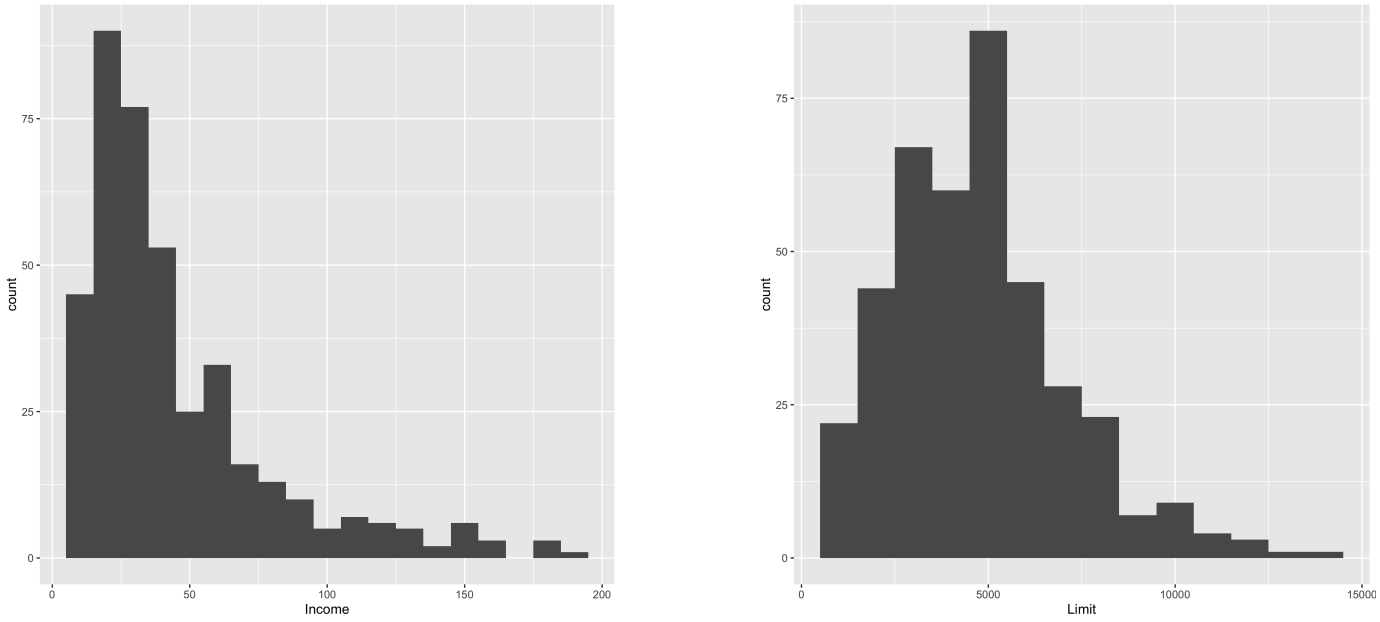
Below (in Figure 2) is the histogram showing the distribution of Balance:

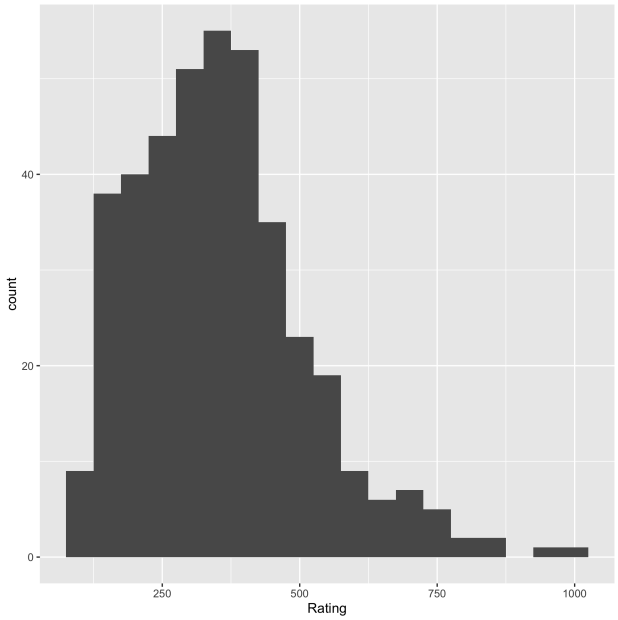
A scatterplot matrix of all the quantitative variables are show below in Figure 3:

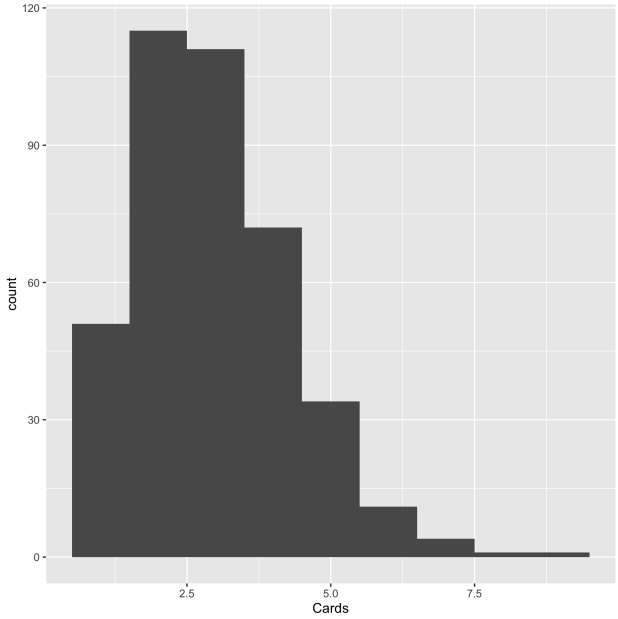
Distribution of categorical variables is shown below in Figure 4.

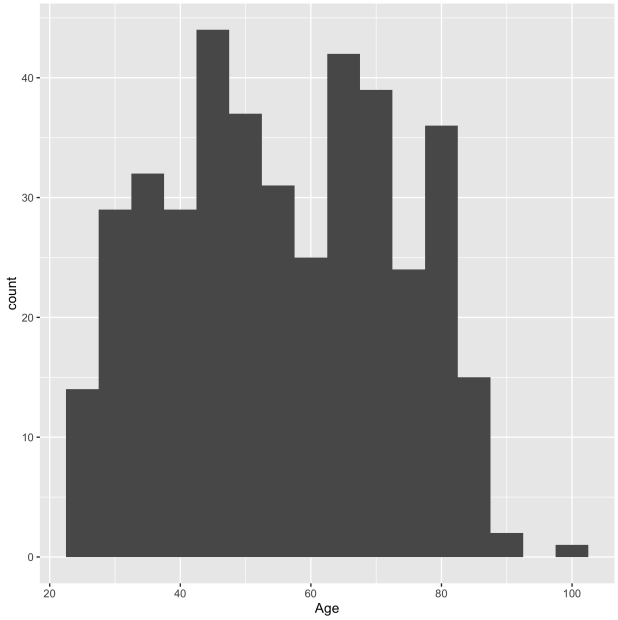
Appendix

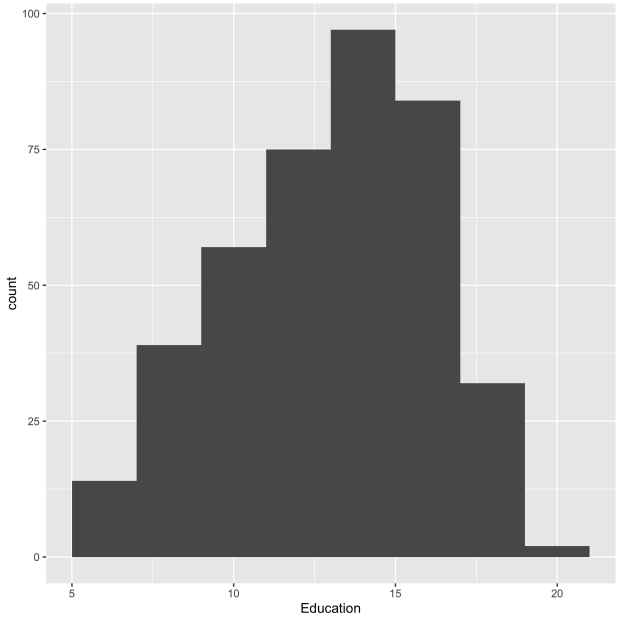
Figure 1: Histograms for quantitative variables in Credit

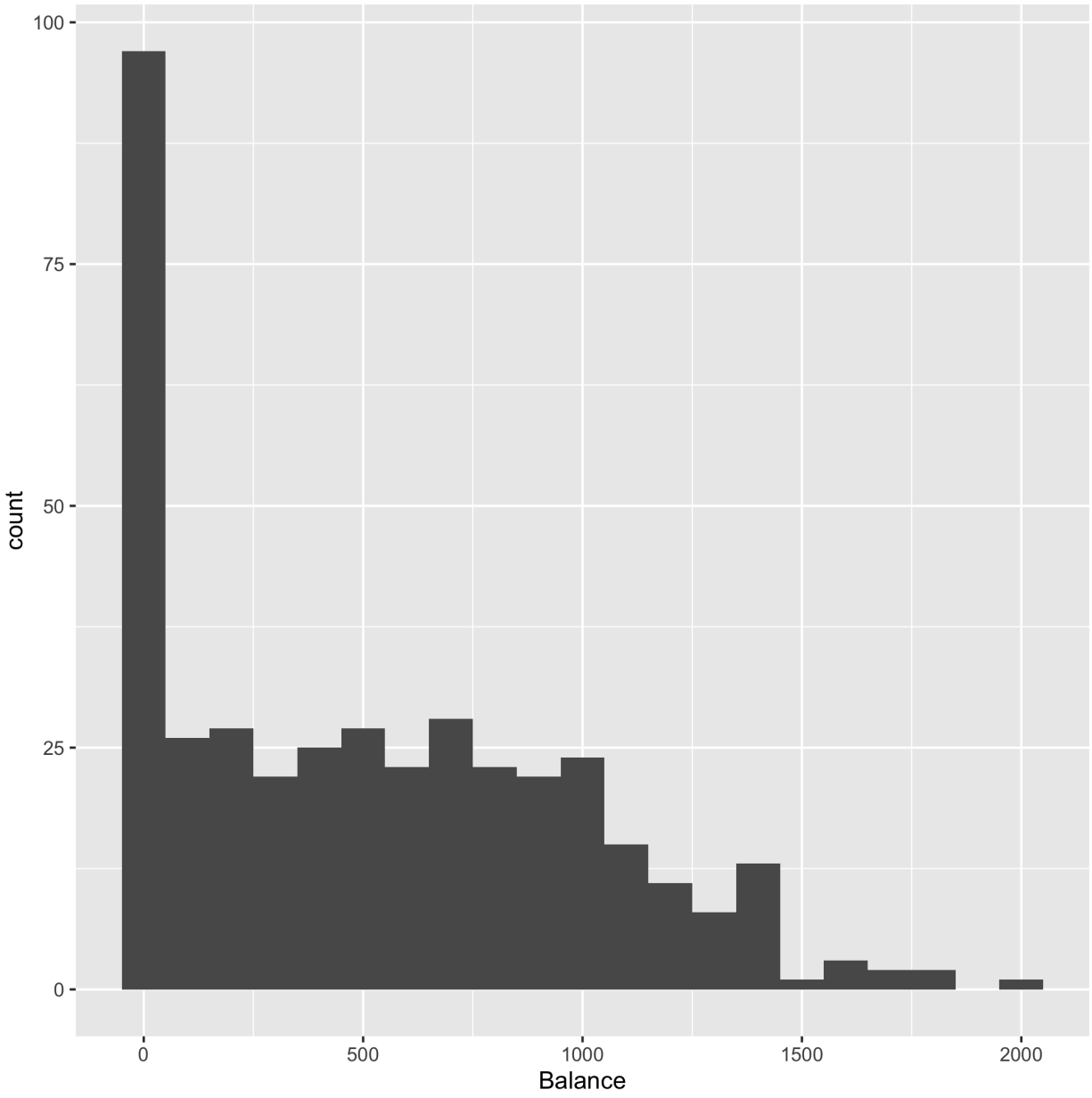



count
40
20
0
250
500
750
1000
Rating


count
120
90
60
30
0
2.5
5.0
7.5
Cards


count
40
30
20
10
0
20
40
60
80
100
Age


count
100
75
50
25
0
5
10
15
20
Education


100
75
50
25
0
count
0
500
1000
1500
2000
Balance

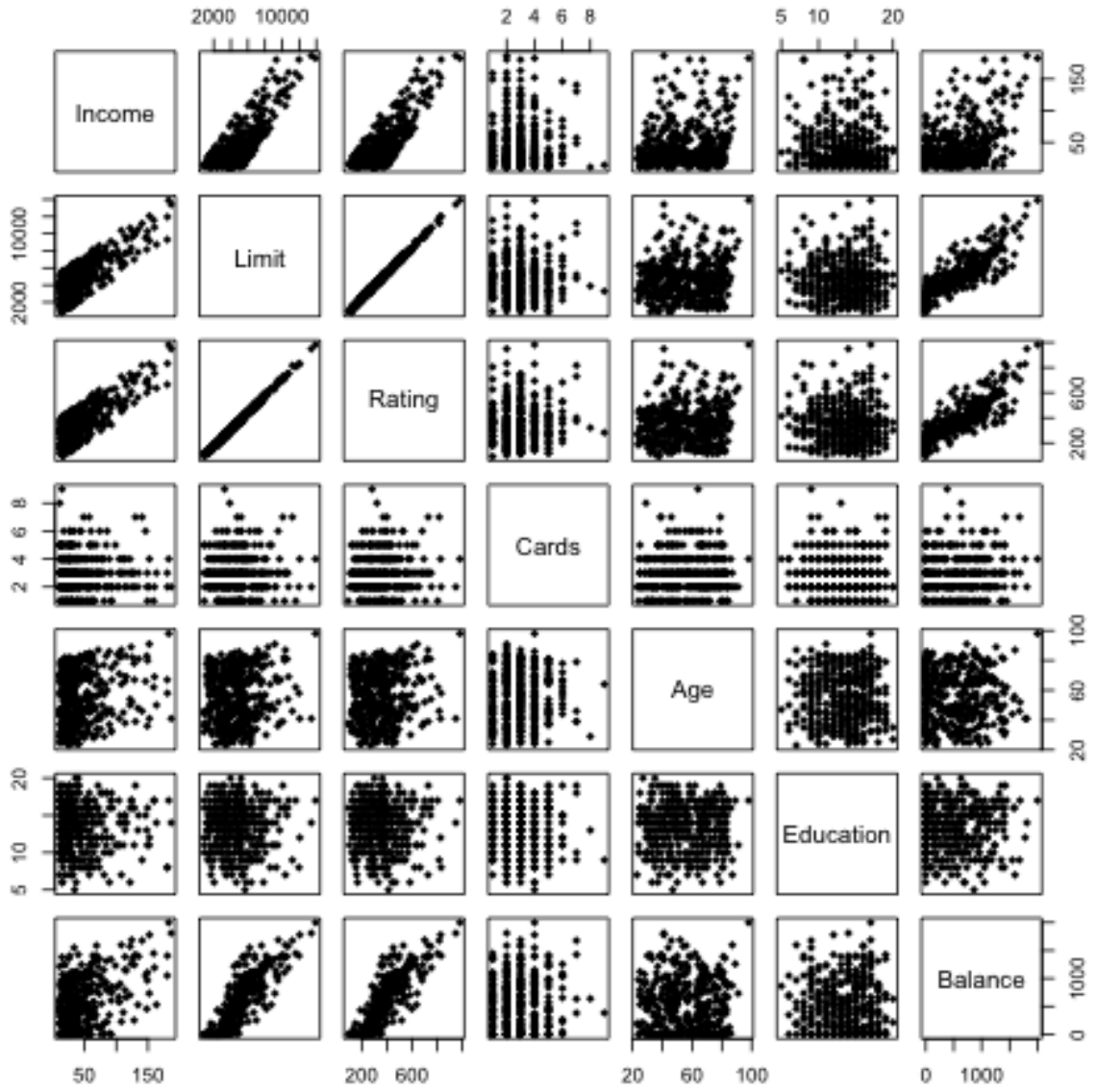
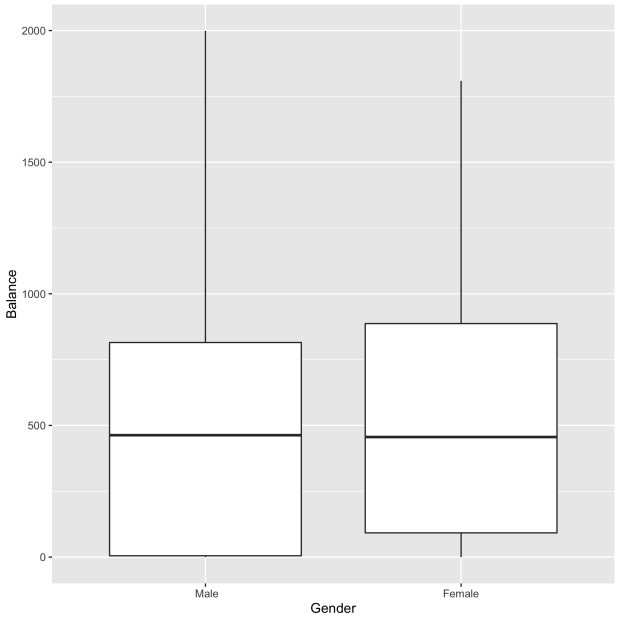
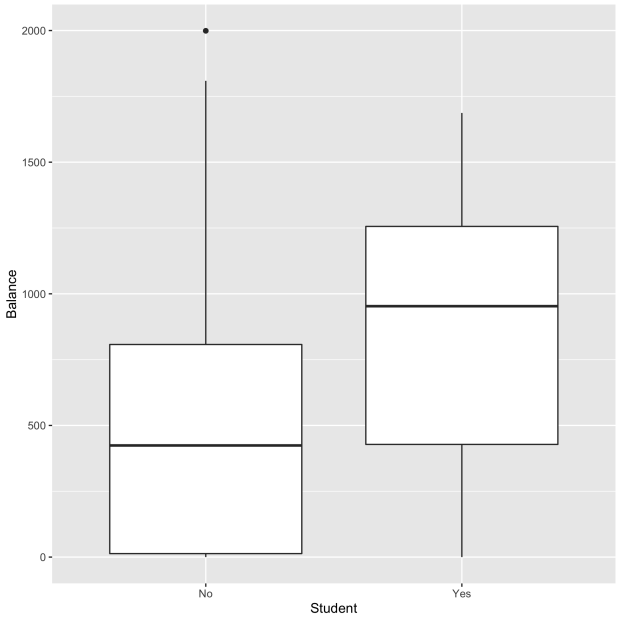
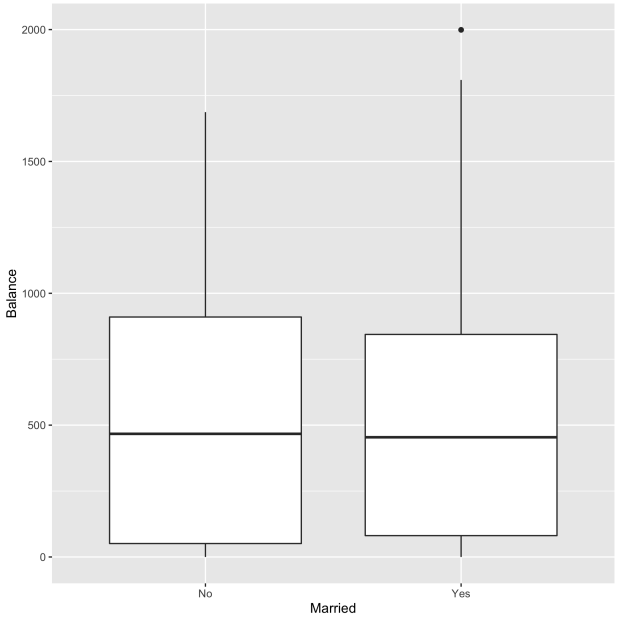
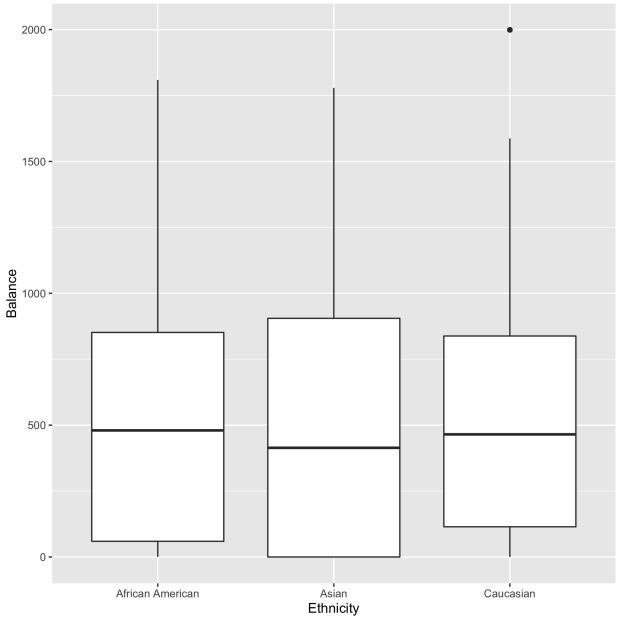


Figure 4: Conditional boxplots for categorical variables in Credit and Balance


Balance
2000
1500
1000
500
0
Male
Female
Gender


Balance
2000
1500
1000
500
0
No
Yes
Student


Balance
2000
1500
1000
500
0
No
Yes
Married


Balance
2000
1500
1000
500
0
African American
Asian
Caucasian
Ethnicity