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.

Analysis

As noted above, we used several regression methods on the dataset Credit.

OLS

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0006	0.0121	-0.05	0.9579
Income	-0.6024	0.0201	-30.00	0.0000
Limit	0.9436	0.1886	5.00	0.0000
Rating	0.4023	0.1893	2.13	0.0344
Cards	0.0563	0.0150	3.76	0.0002
Age	-0.0177	0.0122	-1.45	0.1473
Education	0.0037	0.0123	0.30	0.7610
GenderFemale	-0.0112	0.0121	-0.92	0.3572
StudentYes	0.2841	0.0123	23.08	0.0000
MarriedYes	-0.0005	0.0122	-0.04	0.9671
EthnicityAsian	0.0206	0.0148	1.39	0.1664
EthnicityCaucasian	0.0188	0.0148	1.27	0.2049

Table 1: Multiple Ordinary Linear Regression (OLS)

```
## Loading required package: Matrix
```

12 x 1 sparse Matrix of class “dgCMatrix” s0 (Intercept) .
Income -0.568706769 Limit 0.718657903 Rating 0.593058827 Cards 0.044252756 Age -0.025384931 Education

-0.005879651 GenderFemale -0.010677773 StudentYes 0.273184181 MarriedYes -0.011027761 EthnicityAsian 0.016378729 EthnicityCaucasian 0.011011789

	s0
(Intercept)	0.00
Income	-0.57
Limit	0.72
Rating	0.59
Cards	0.04
Age	-0.03
Education	-0.01
GenderFemale	-0.01
StudentYes	0.27
MarriedYes	-0.01
EthnicityAsian	0.02
EthnicityCaucasian	0.01

Table 2: Final Coefficents for Ridge Regression

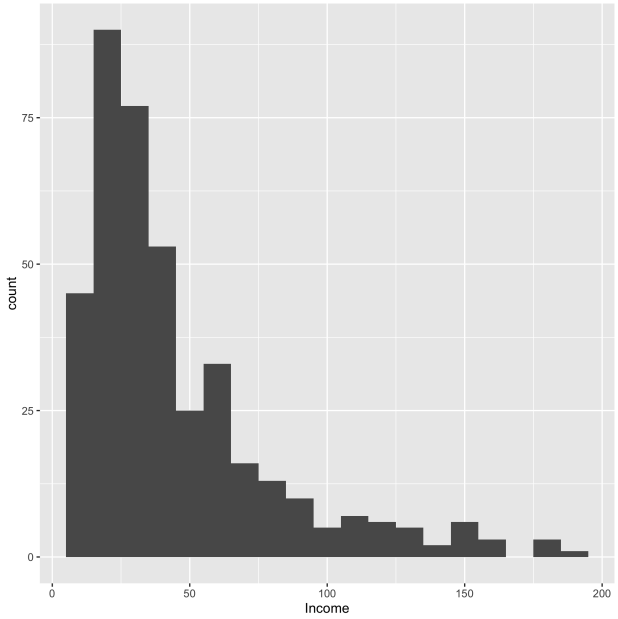
	s0
(Intercept)	0.00
Income	-0.55
Limit	0.93
Rating	0.37
Cards	0.04
Age	-0.02
Education	0.00
GenderFemale	0.00
StudentYes	0.27
MarriedYes	0.00
EthnicityAsian	0.00
EthnicityCaucasian	0.00

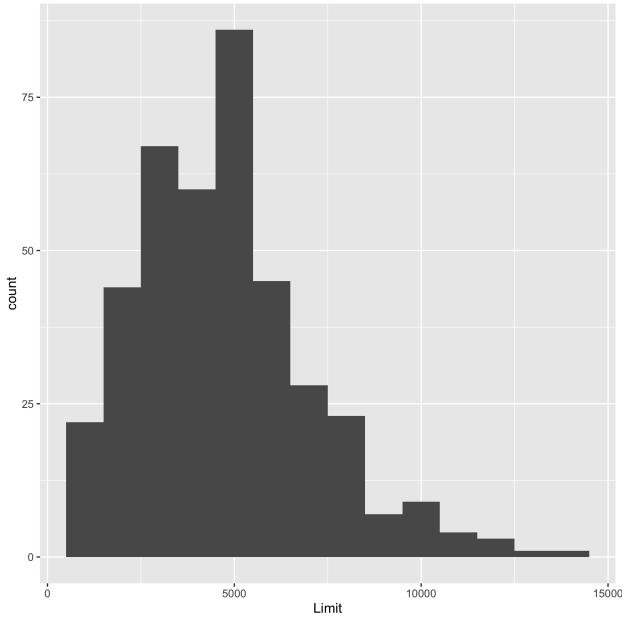
Table 3: Final Coefficents for Lasso Regression

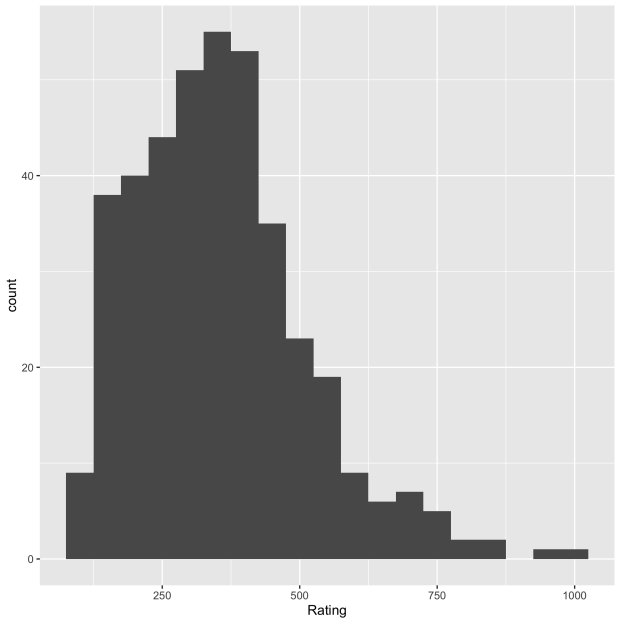
Appendix

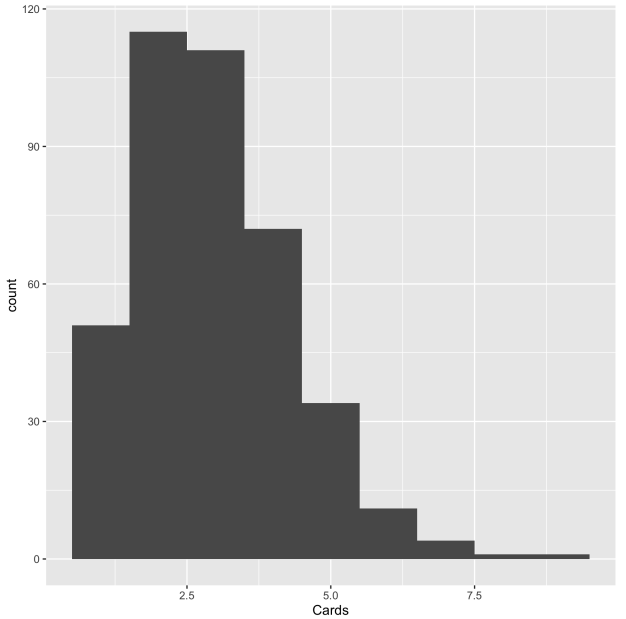
Figure 1: Histograms for quantitative variables in Credit

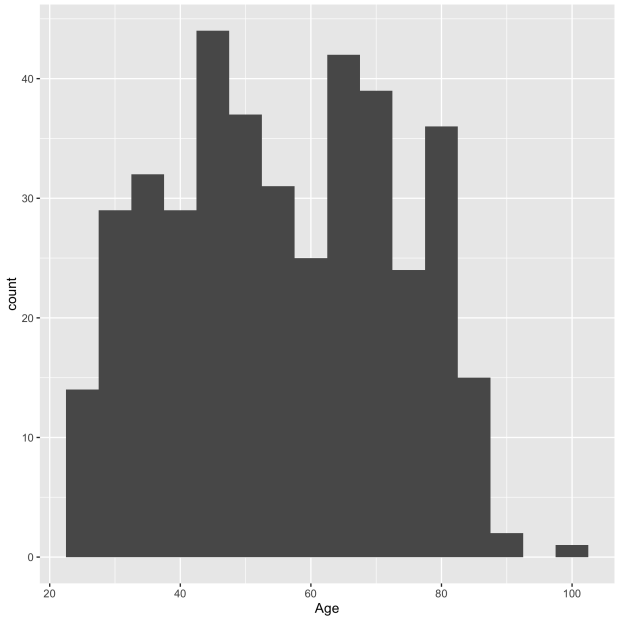
	OLS	Ridge	Lasso	PCR	PLS
(Intercept)	-0.00	0.00	0.00		
Income	-0.60	-0.57	-0.55	-0.60	-0.60
Limit	0.94	0.72	0.93	0.96	0.68
Rating	0.40	0.59	0.37	0.38	0.66
Cards	0.06	0.04	0.04	0.05	0.04
Age	-0.02	-0.03	-0.02	-0.02	-0.02
Education	0.00	-0.01	0.00	-0.01	-0.01
GenderFemale	-0.01	-0.01	0.00	-0.01	-0.01
StudentYes	0.28	0.27	0.27	0.28	0.28
MarriedYes	-0.00	-0.01	0.00	-0.01	-0.01
EthnicityAsian	0.02	0.02	0.00	0.02	0.01
EthnicityCaucasian	0.02	0.01	0.00	0.01	0.01











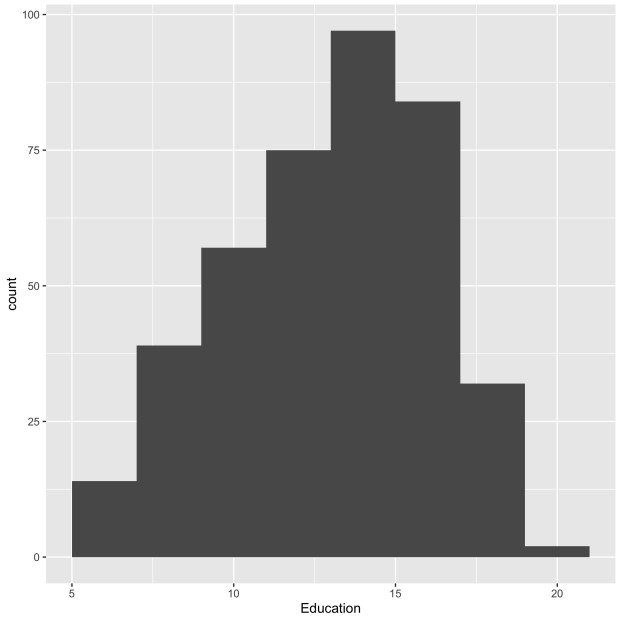
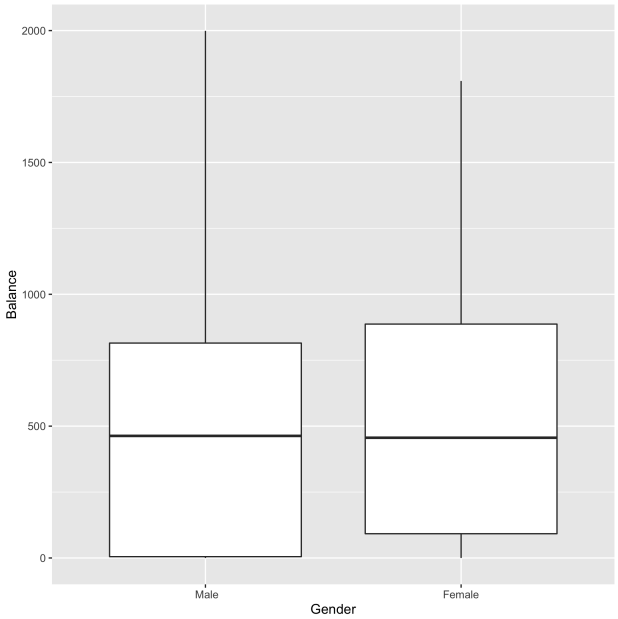
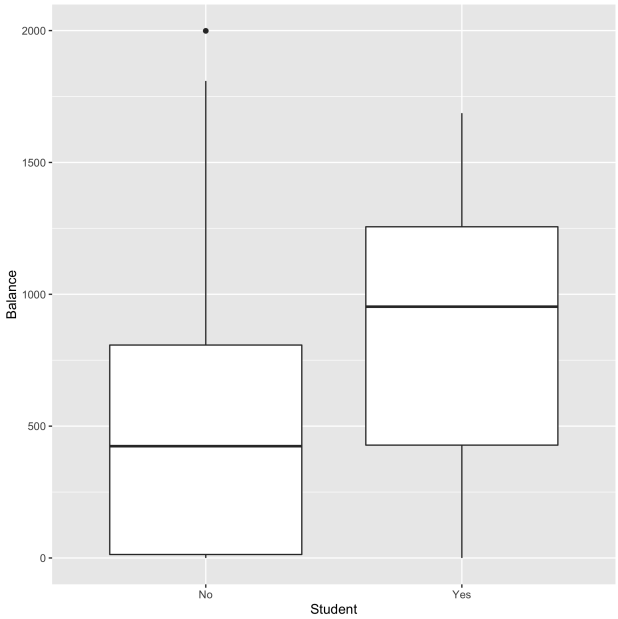


Figure 2: Distribution of Balance

Figure 3: Scatterplot Matrix for all quantiative variables

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





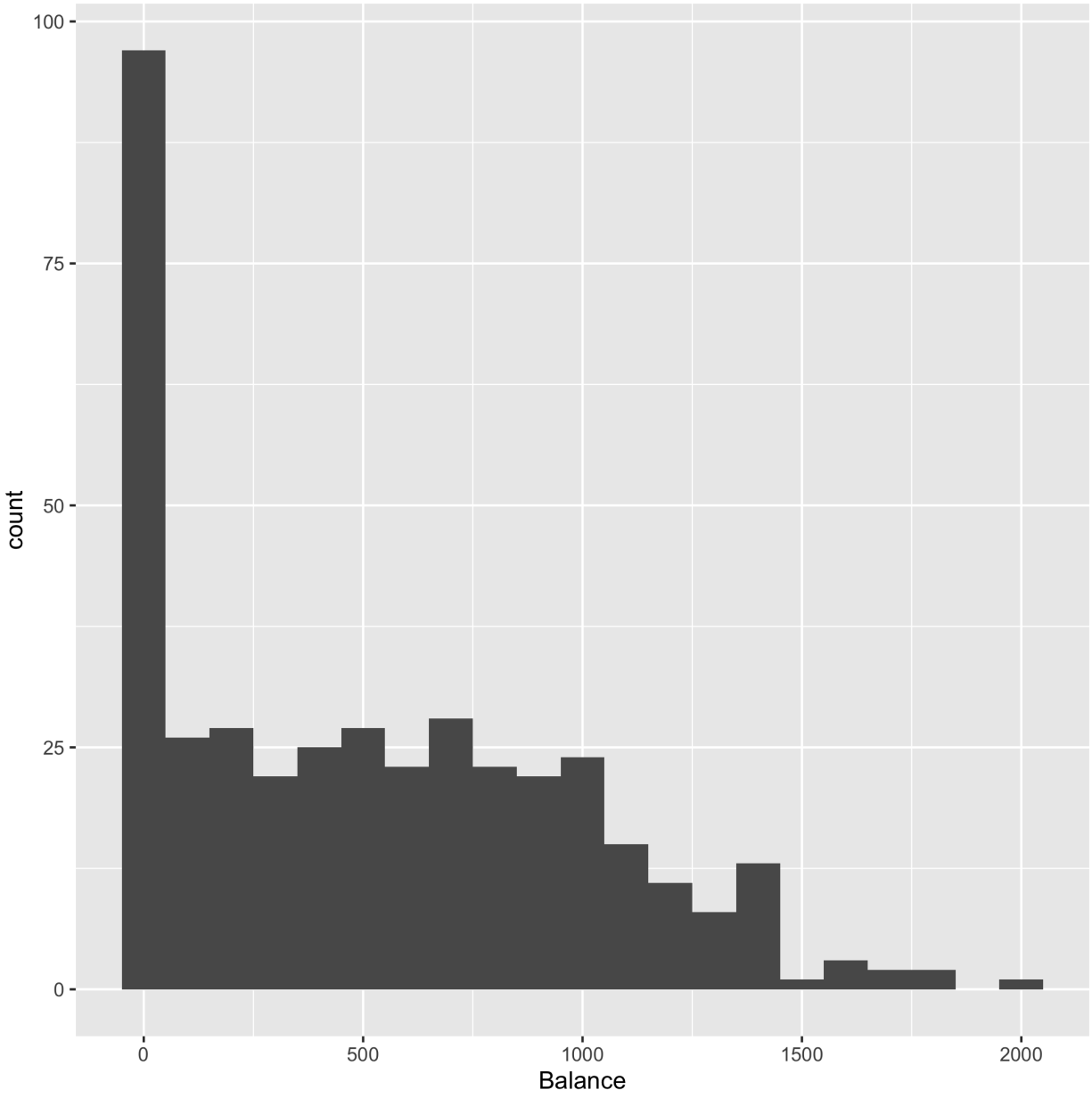


Figure 1: Fig 2: Balance

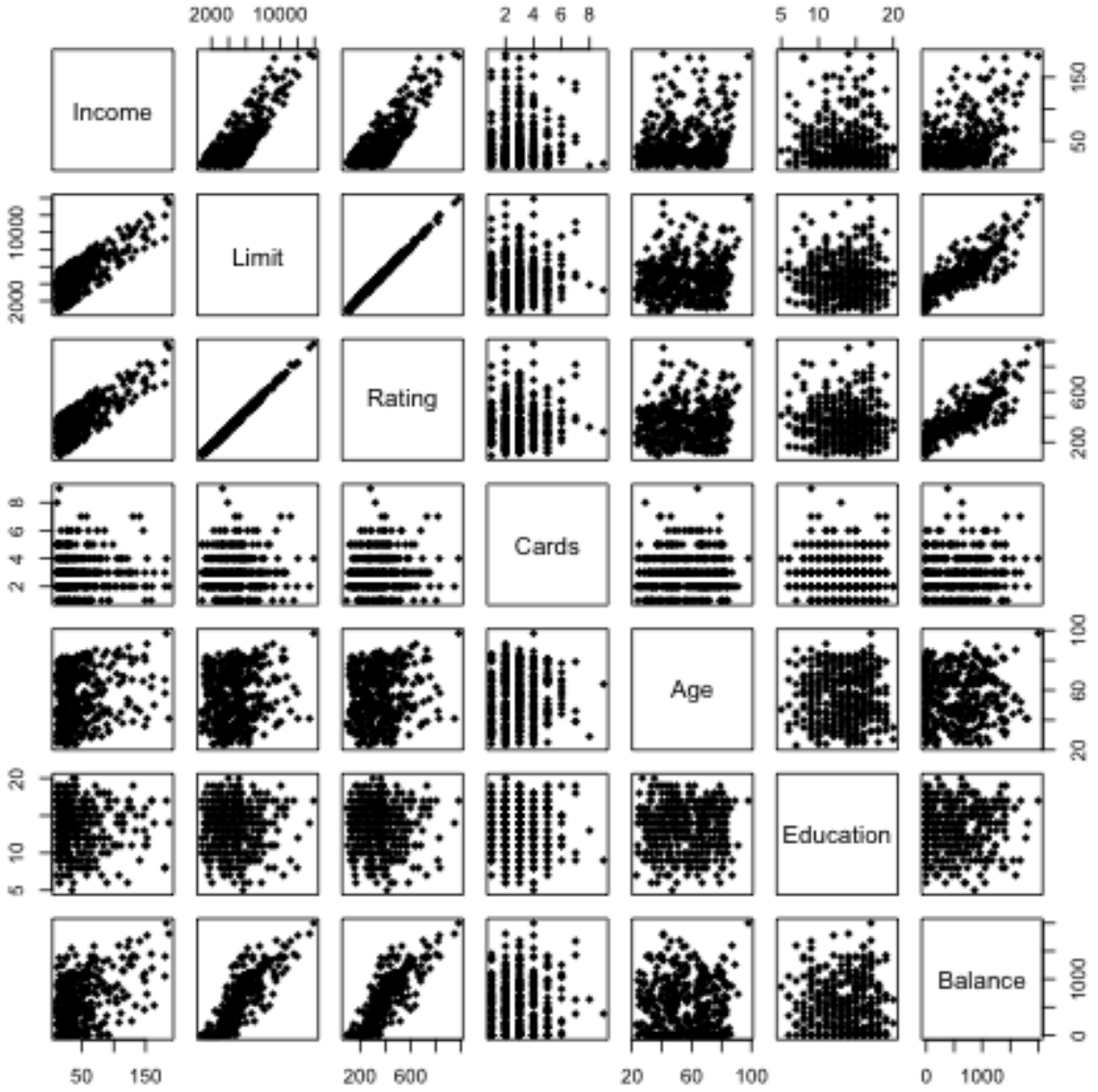
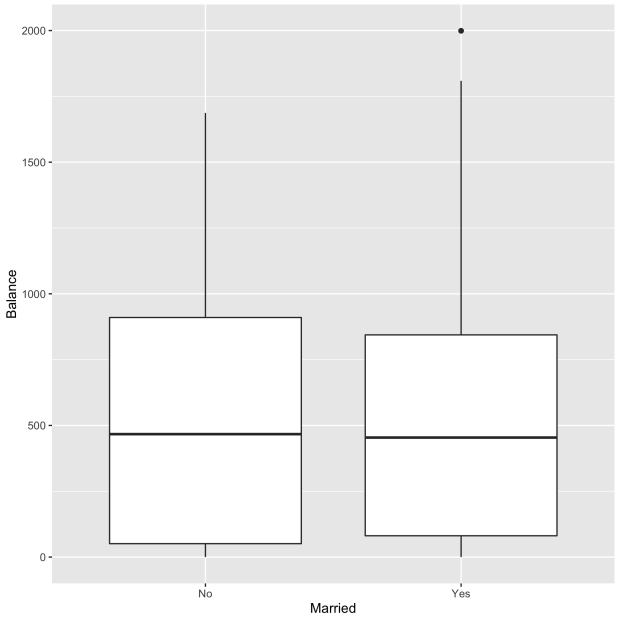
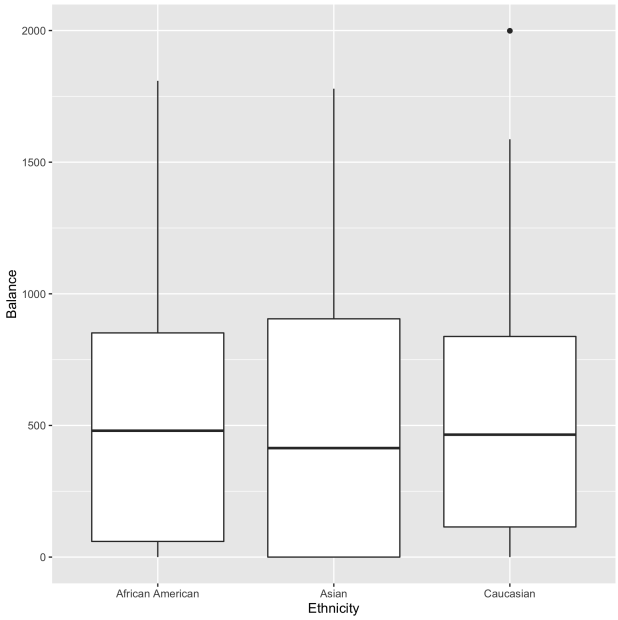


Figure 2: Fig 3: Scatterplot Matrix


2000
1500
1000
500
0
Balance
No
Yes
Married


2000
1500
1000
500
0
Balance
African American
Asian
Caucasian
Ethnicity