Analysis

Exploratry Data Analysis

Quantitative Variables

There are 7 quantitative variables in the dataset, including Income, Limit, Rating, Cards, Age, Education, Balance. For each of the variable, we computed minimum, maximum, range, median, quantile, IQR, mean, and standard deviation for each variable.

The following table is the summary statistics

	min	max	range	median	first_quartile	third_quartile	IQR	mean	sd
Income	10.354	186.634	176.28	33.1155	21.00725	57.47075	36.4635	45.21889	35.244273
Limit	855.000	13913.000	13058.00	4622.5000	3088.00000	5872.75000	2784.7500	4735.60000	2308.198848
Rating	93.000	982.000	889.00	344.0000	247.25000	437.25000	190.0000	354.94000	154.724143
Cards	1.000	9.000	8.00	3.0000	2.00000	4.00000	2.0000	2.95750	1.371275
Age	23.000	98.000	75.00	56.0000	41.75000	70.00000	28.2500	55.66750	17.249807
Education	5.000	20.000	15.00	14.0000	11.00000	16.00000	5.0000	13.45000	3.125207
Balance	0.000	1999.000	1999.00	459.5000	68.75000	863.00000	794.2500	520.01500	459.758877

Then we make Histograms and Boxplots for each of the variables
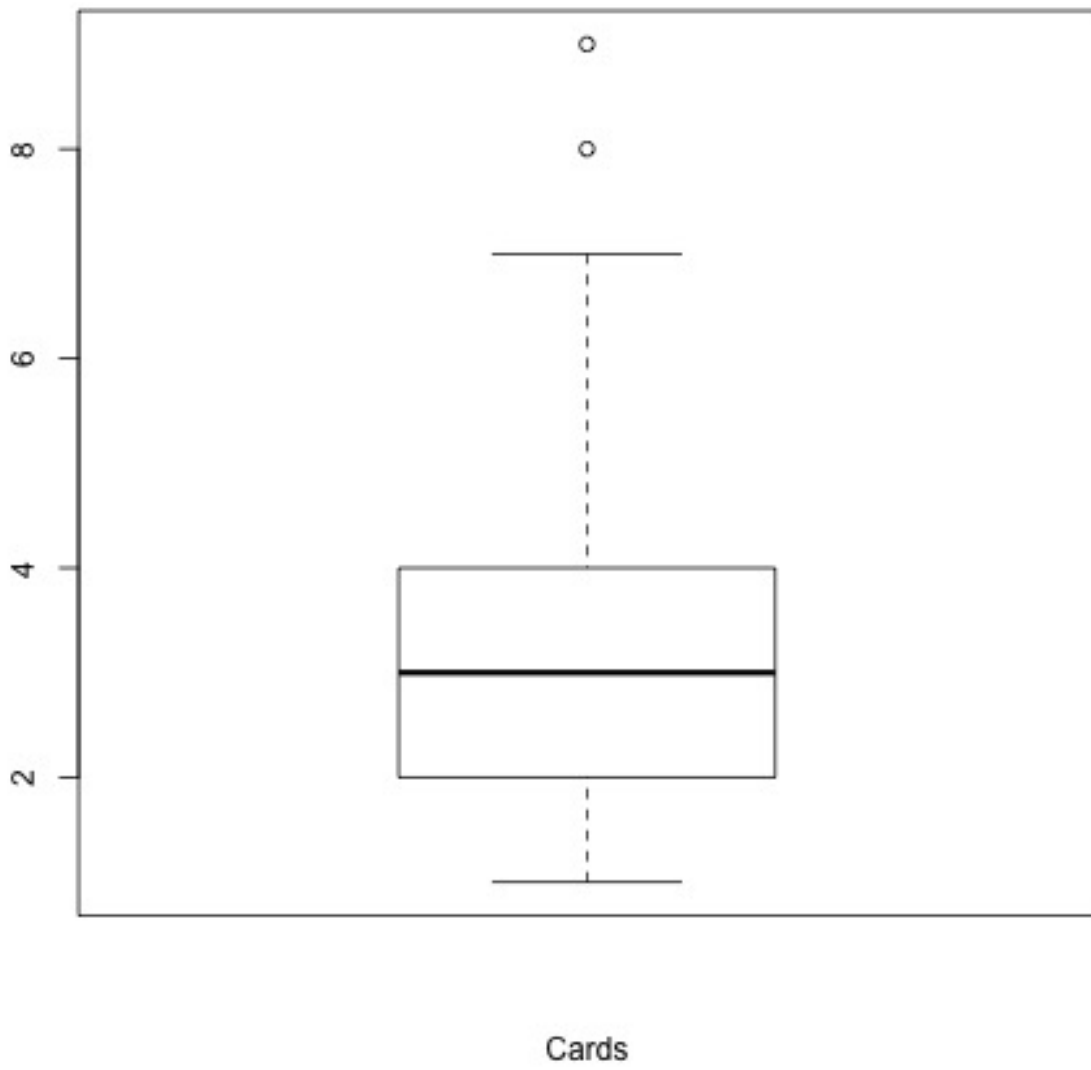
For Predictor Age:

Boxplot

Boxplot of Age

Age
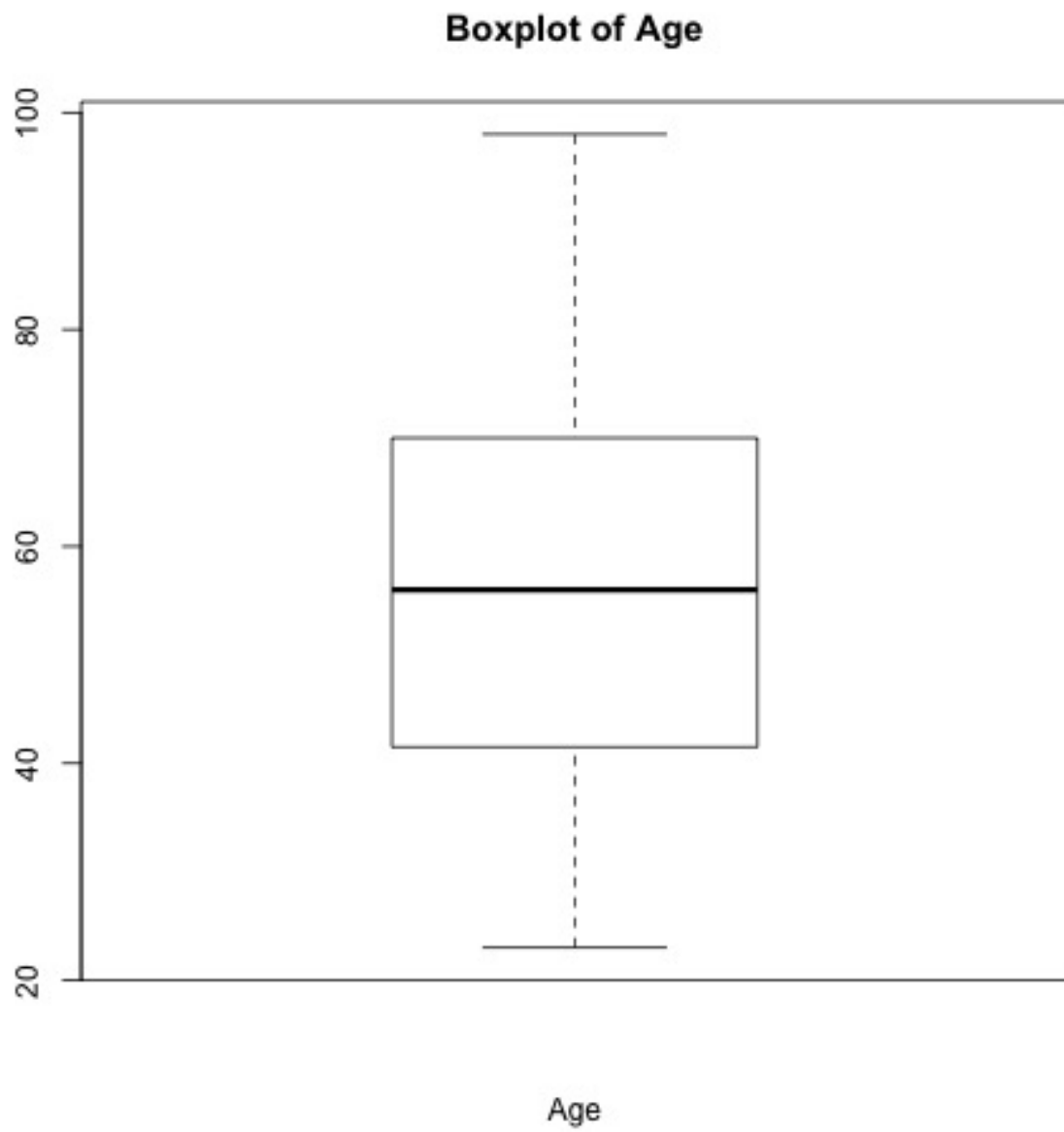
Histogram
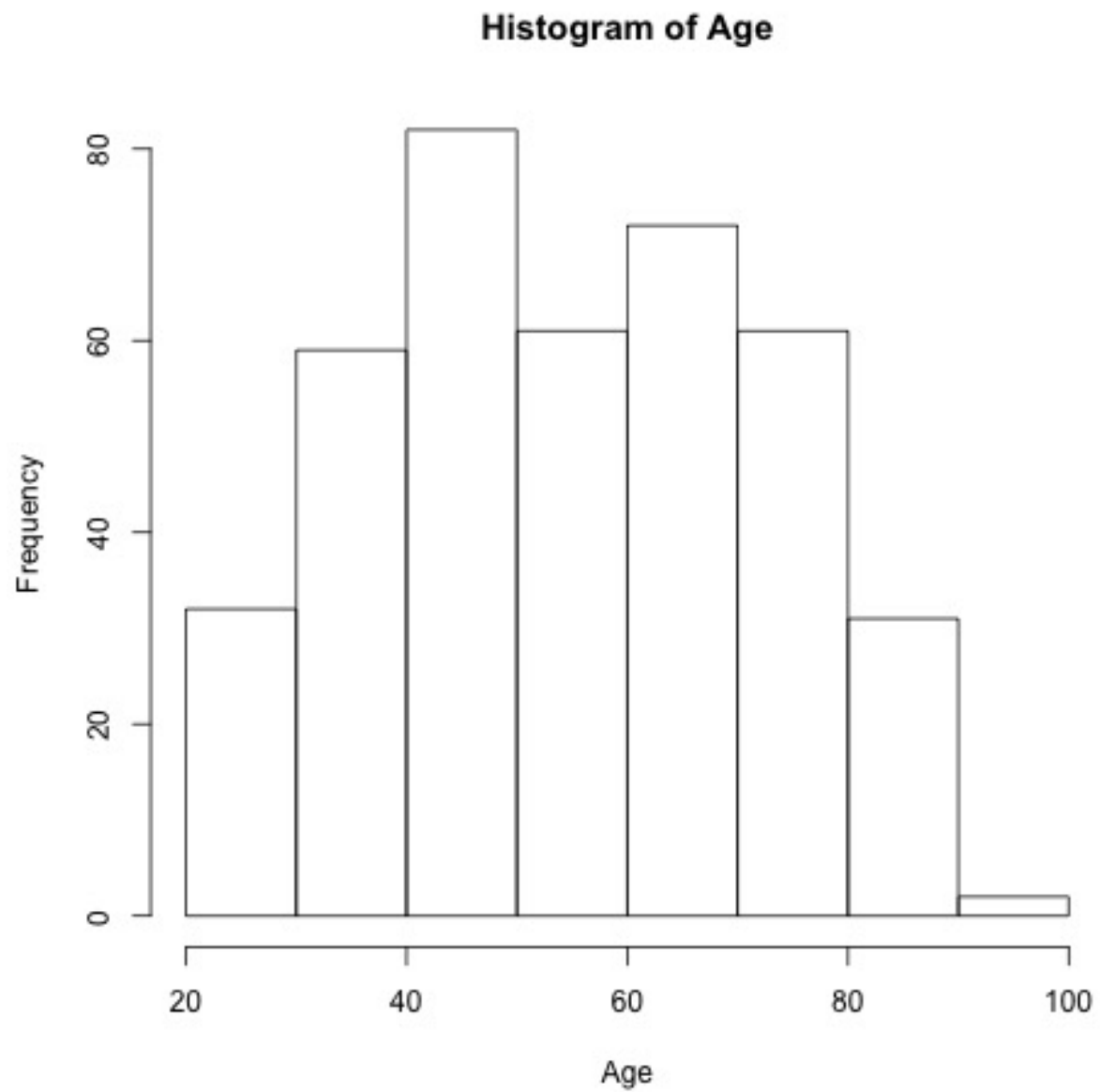
Histogram of Age

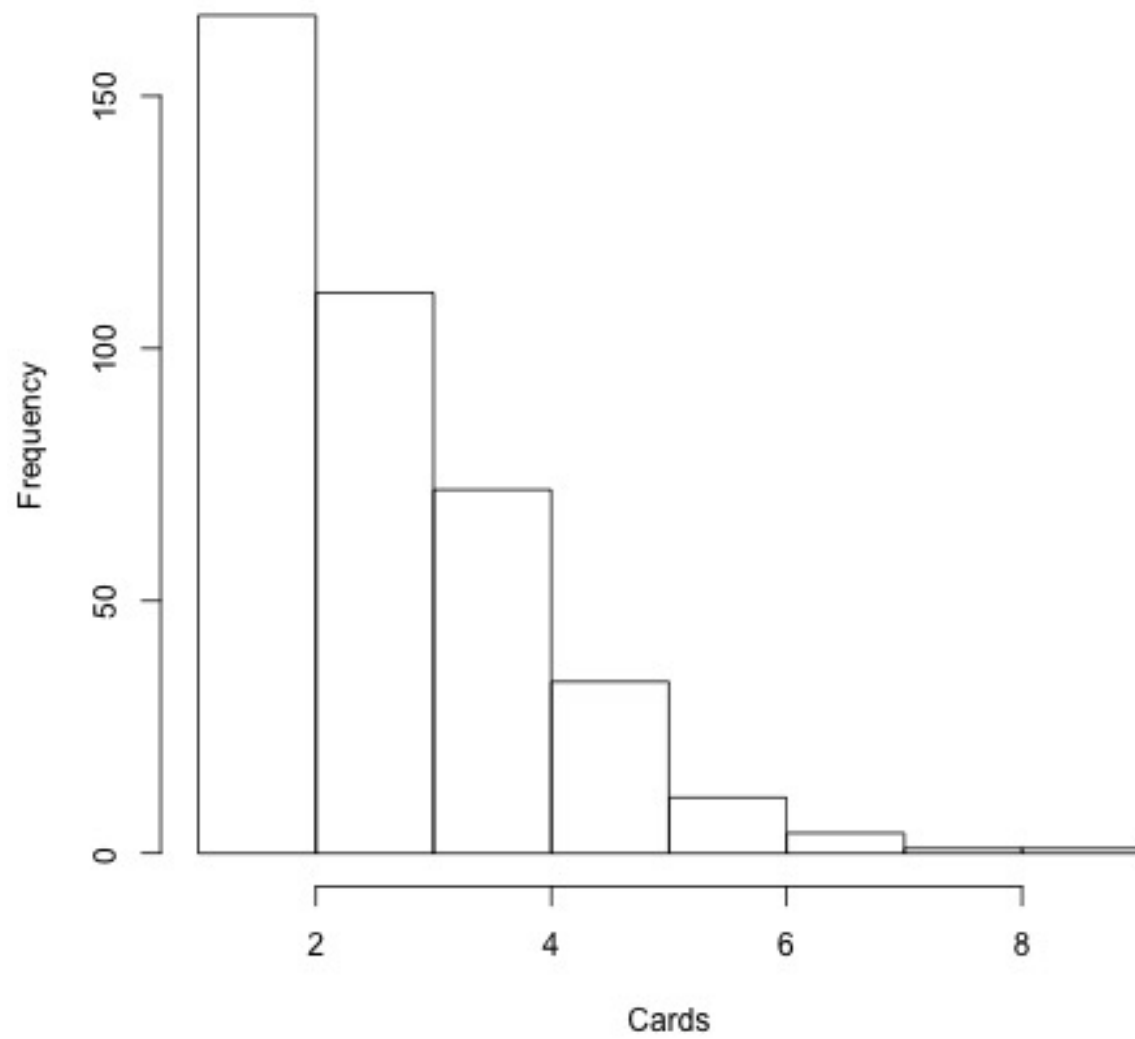For Predictor Cards

Boxplot

Boxplot of Cards

Cards
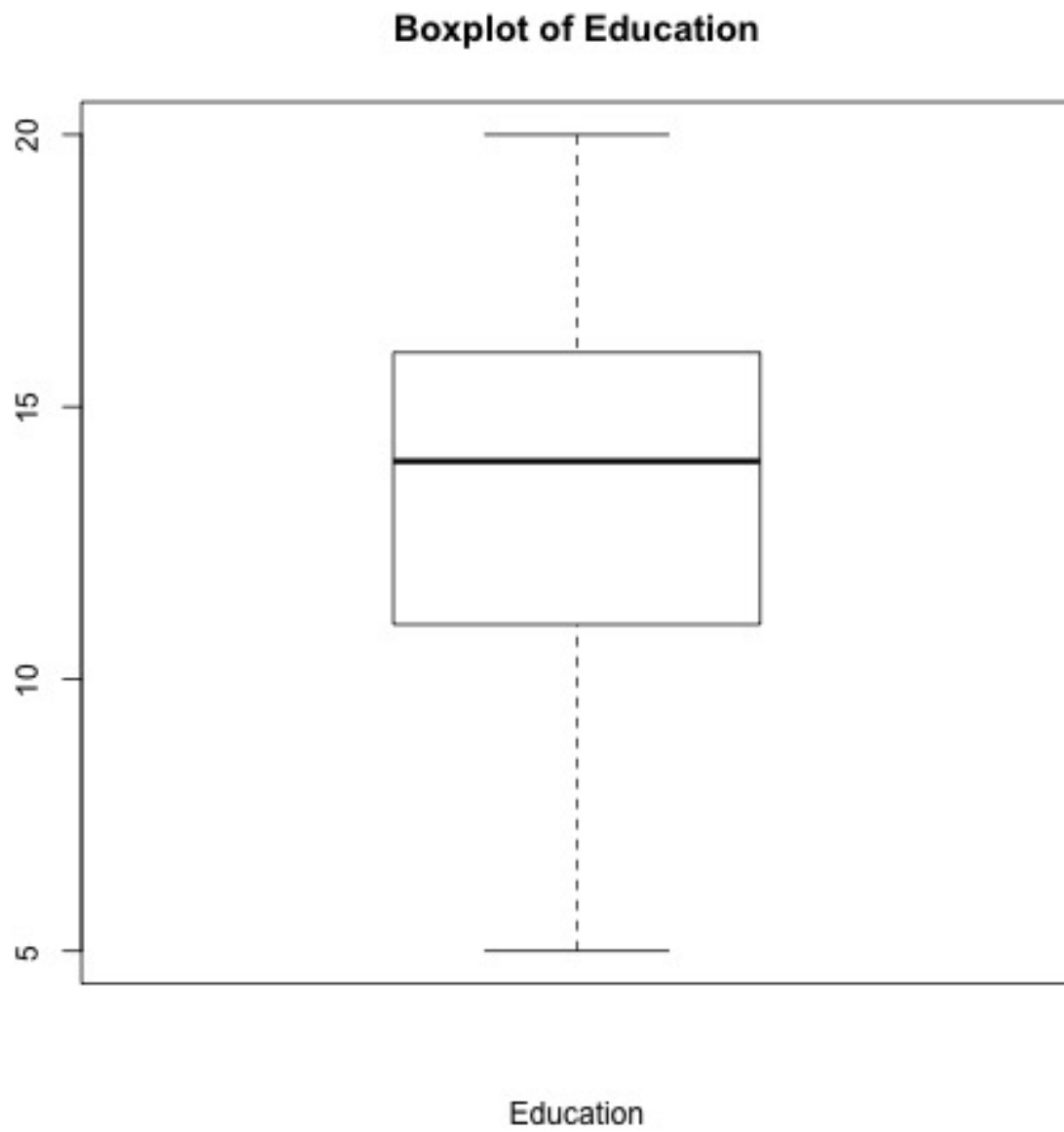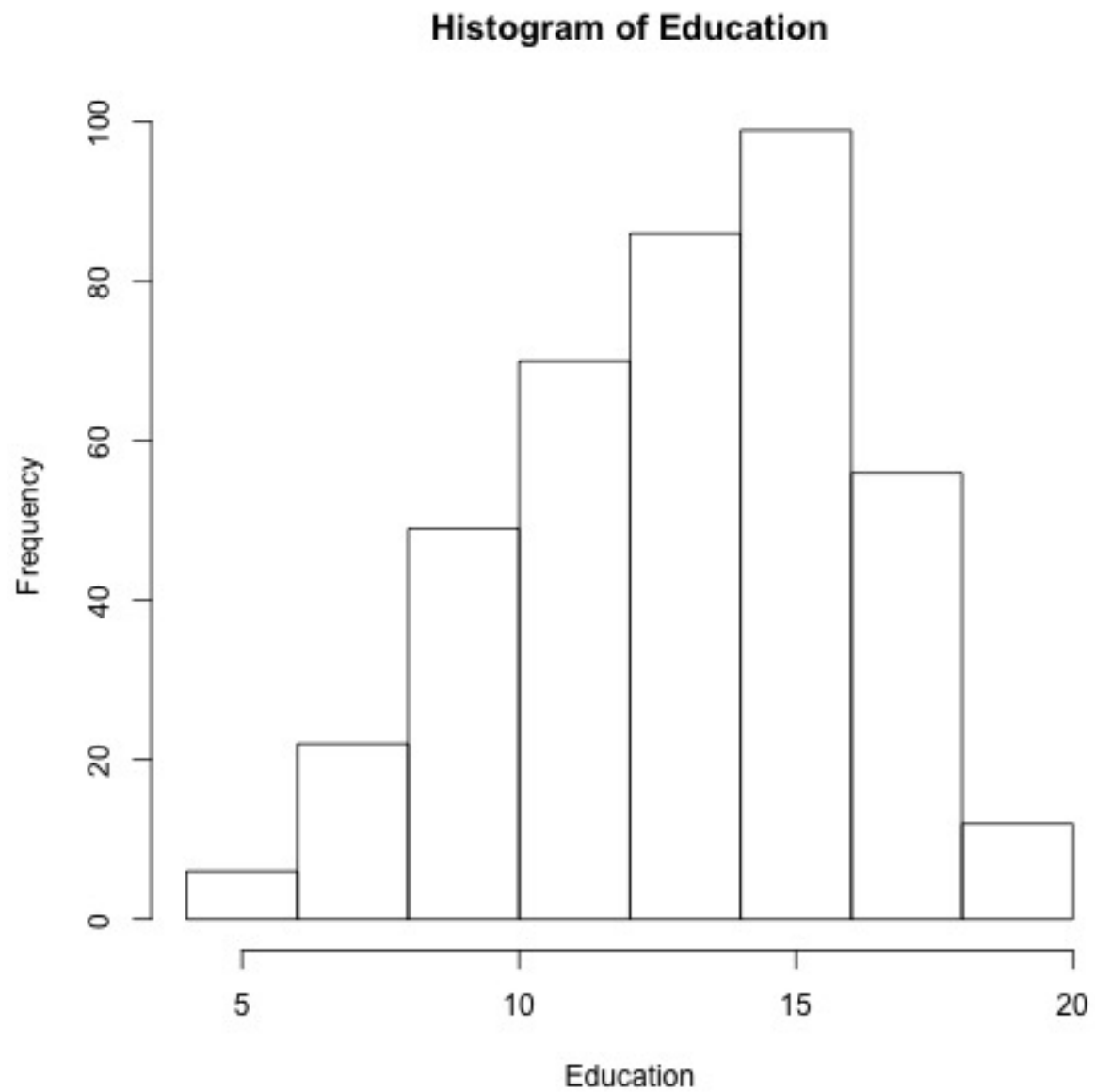
Histogram
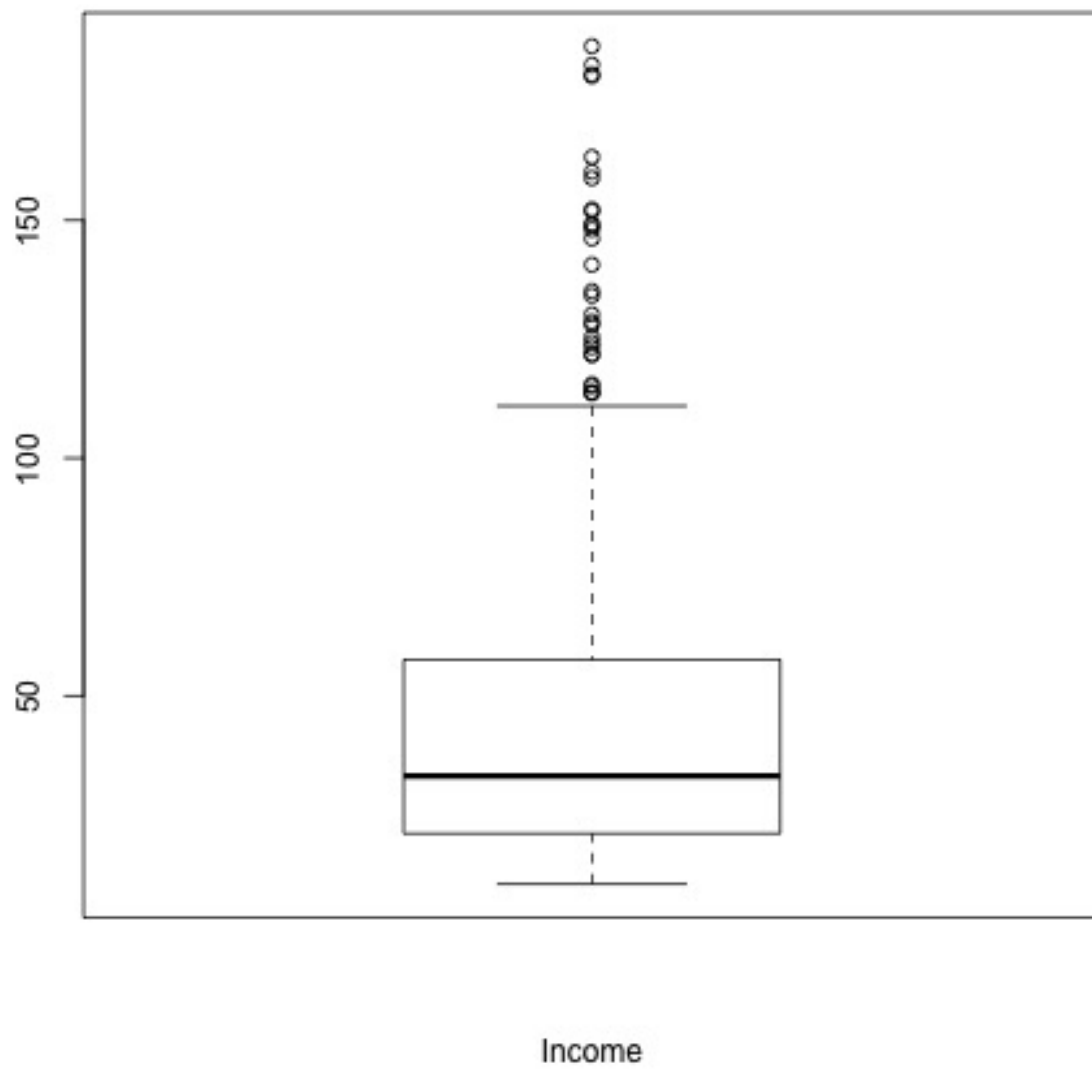
Histogram of Cards

For Predictor Education

Boxplot

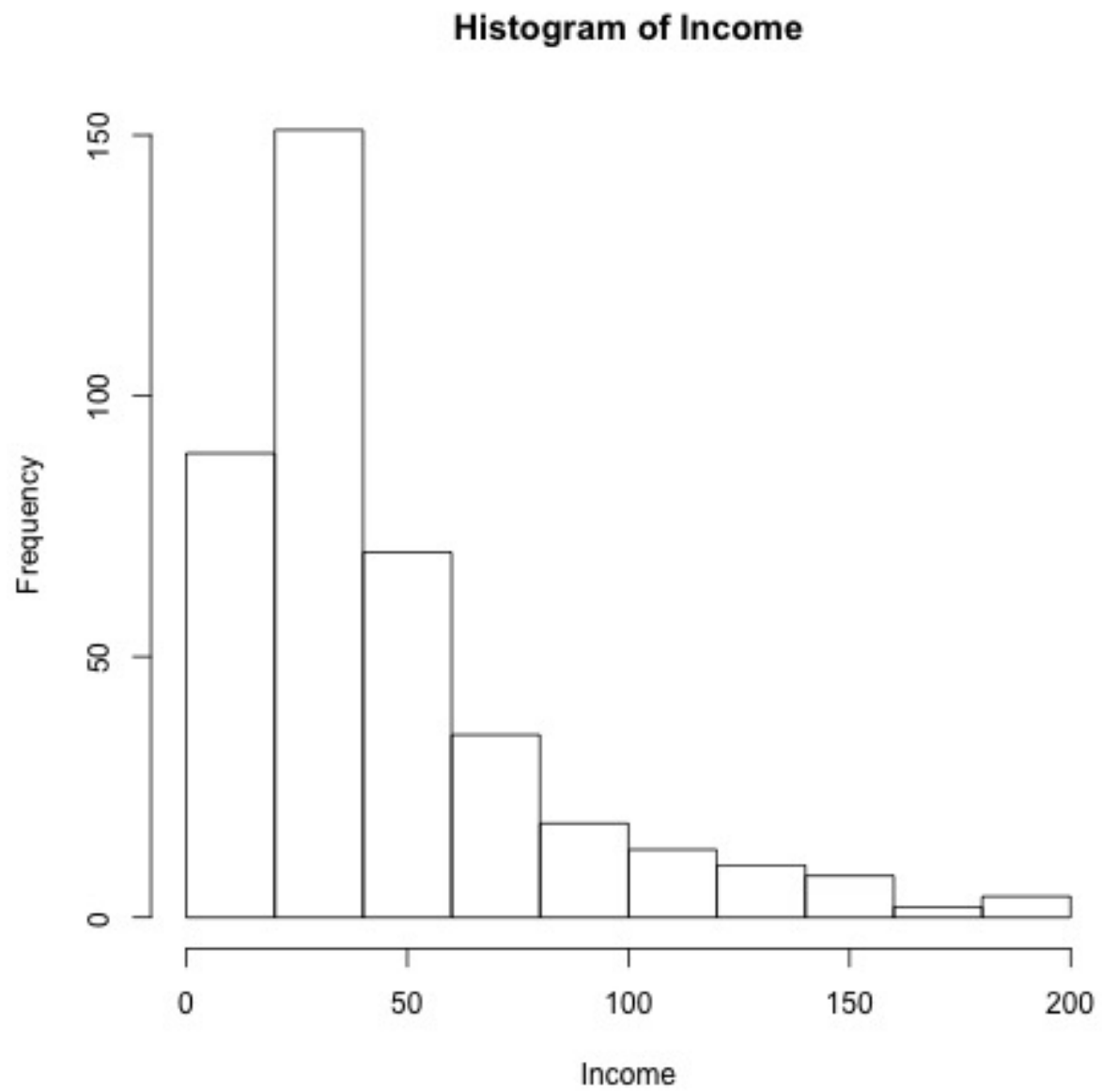# Boxplot of Education

Education

Histogram

Histogram of Education

For Predictor Income

Boxplot

Boxplot of Income

Income

Histogram

Histogram of Income

For Predictor Limit

Boxplot

Boxplot of Limit

Limit

Histogram

Histogram of Limit
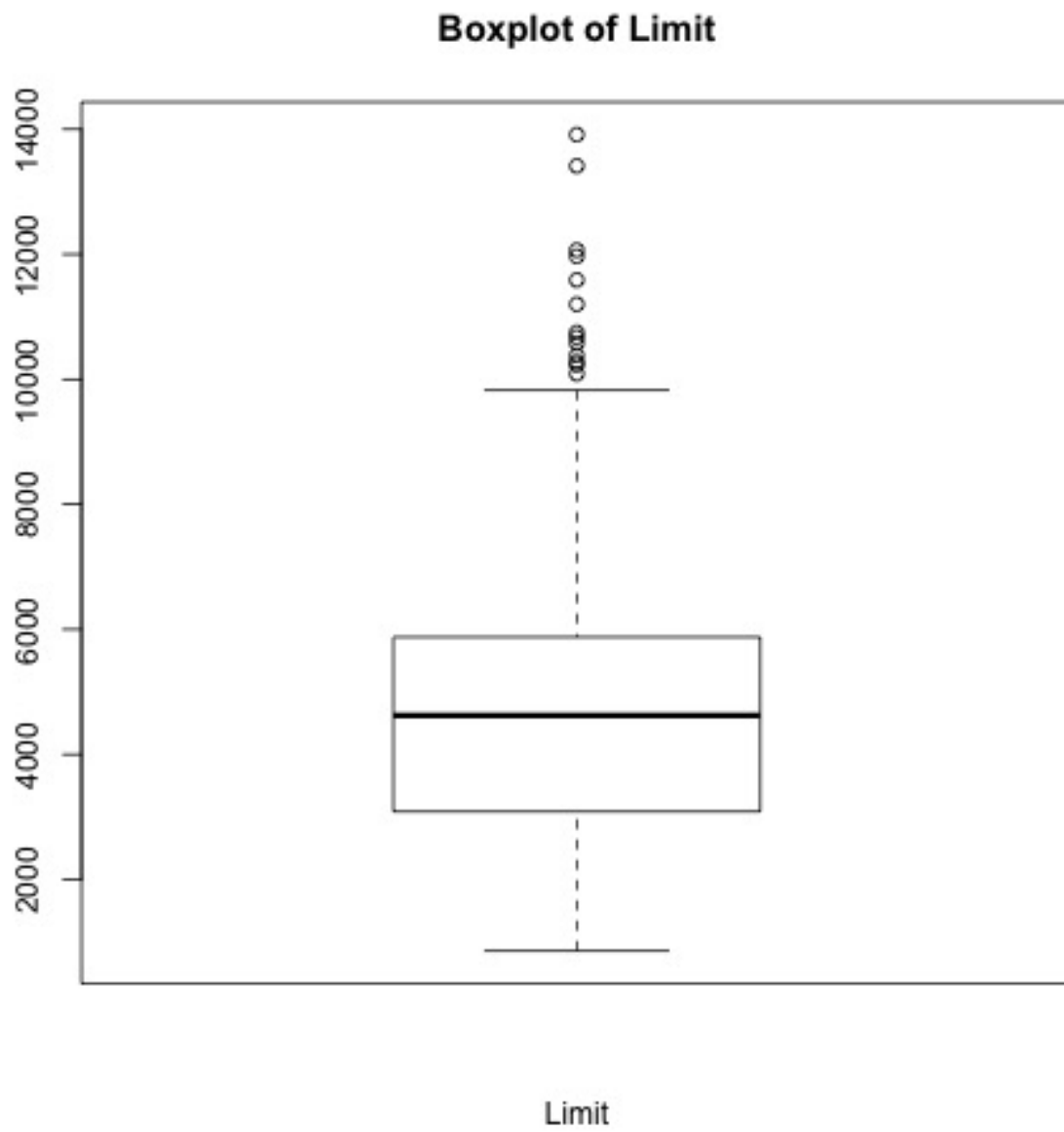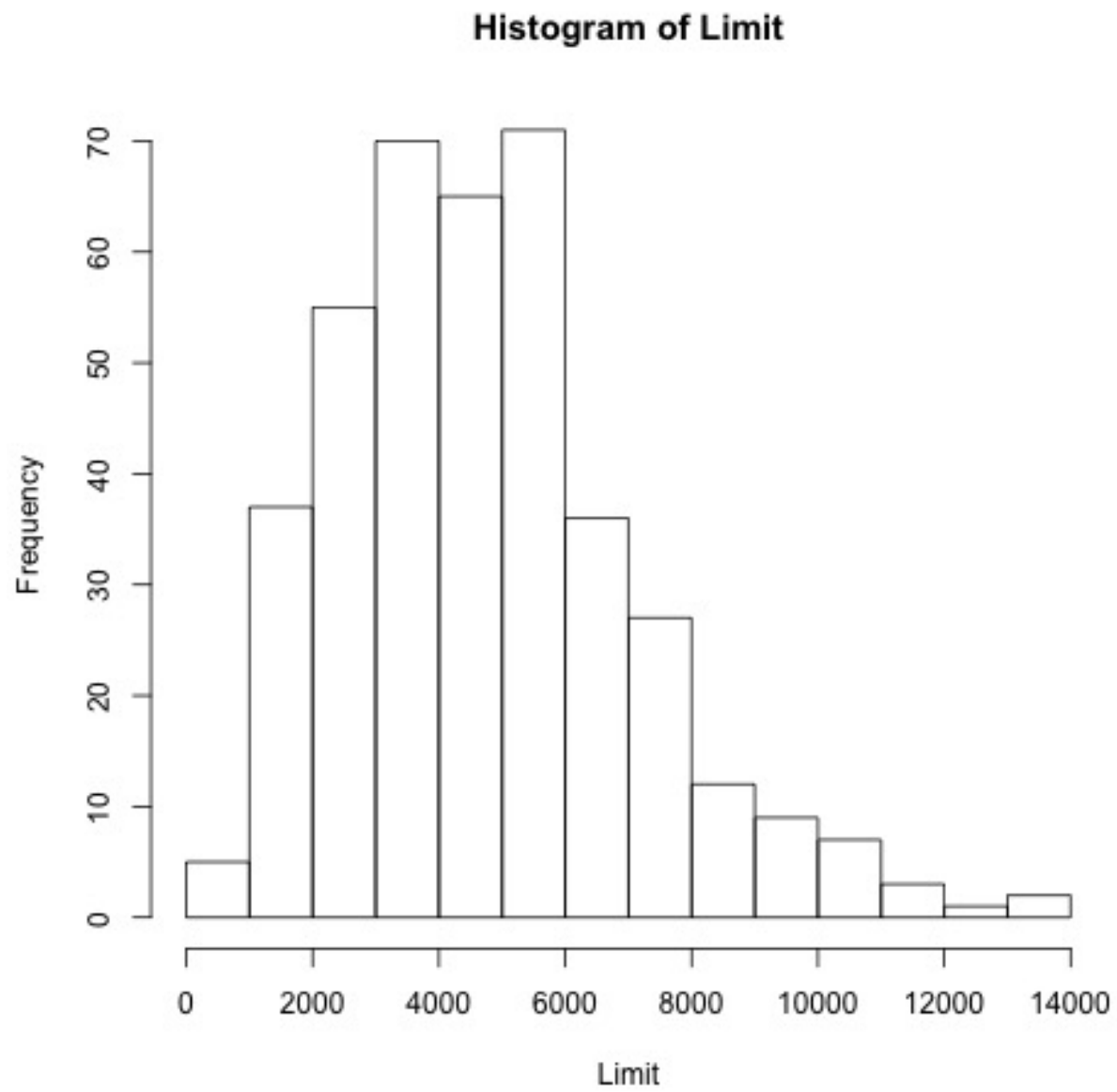
For Predictor Rating

Boxplot

Boxplot of Rating

Rating

Histogram

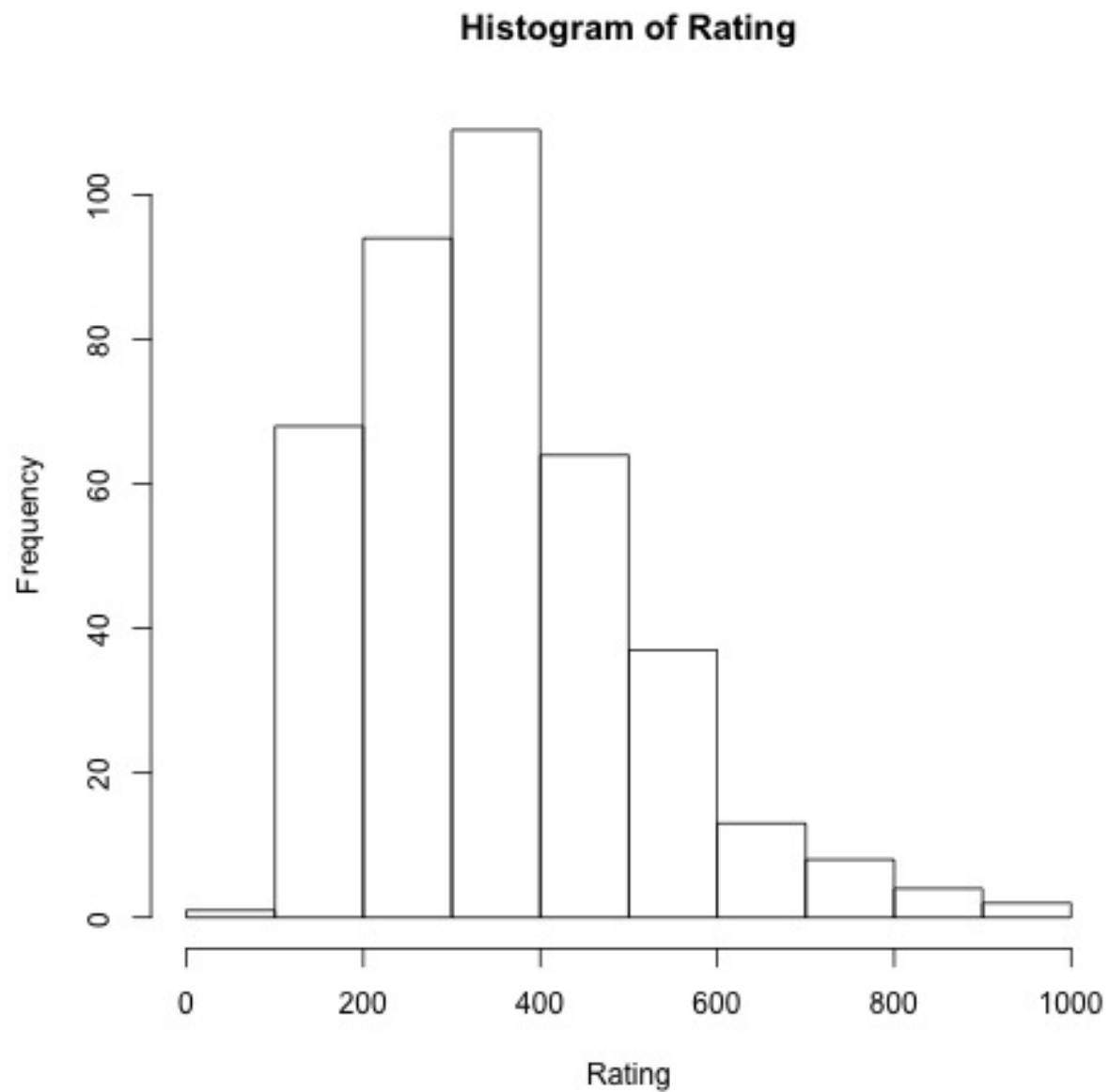## Histogram of Rating
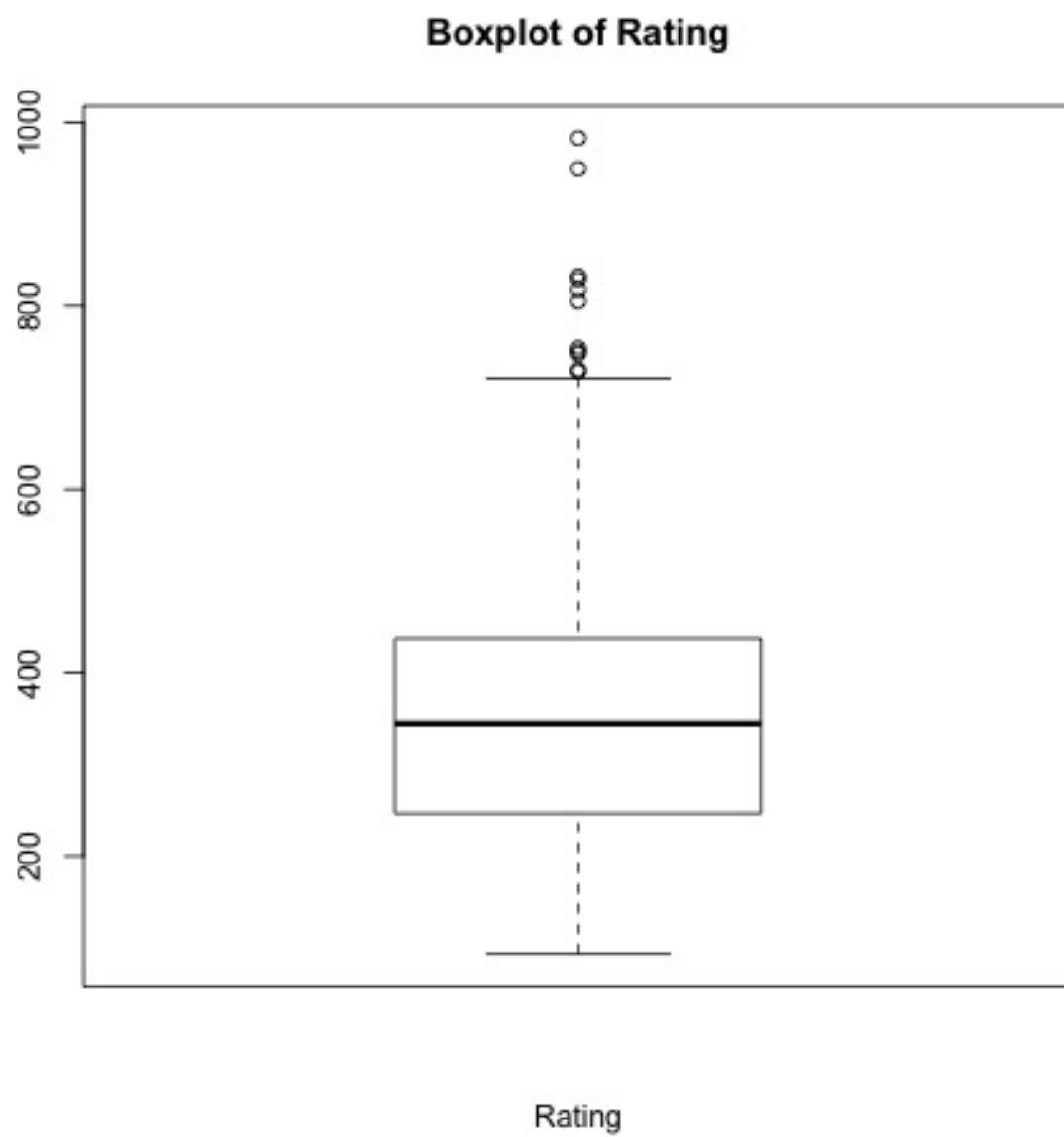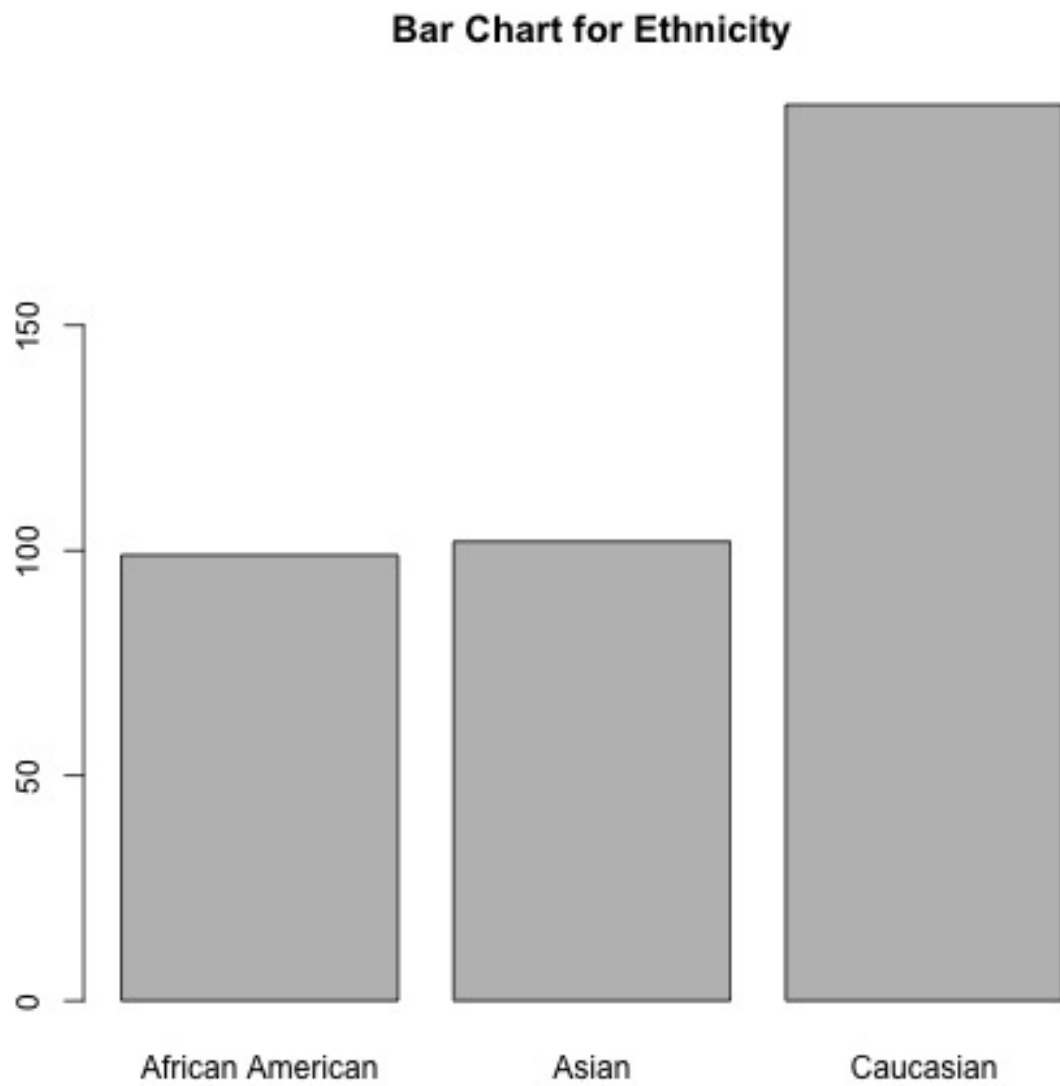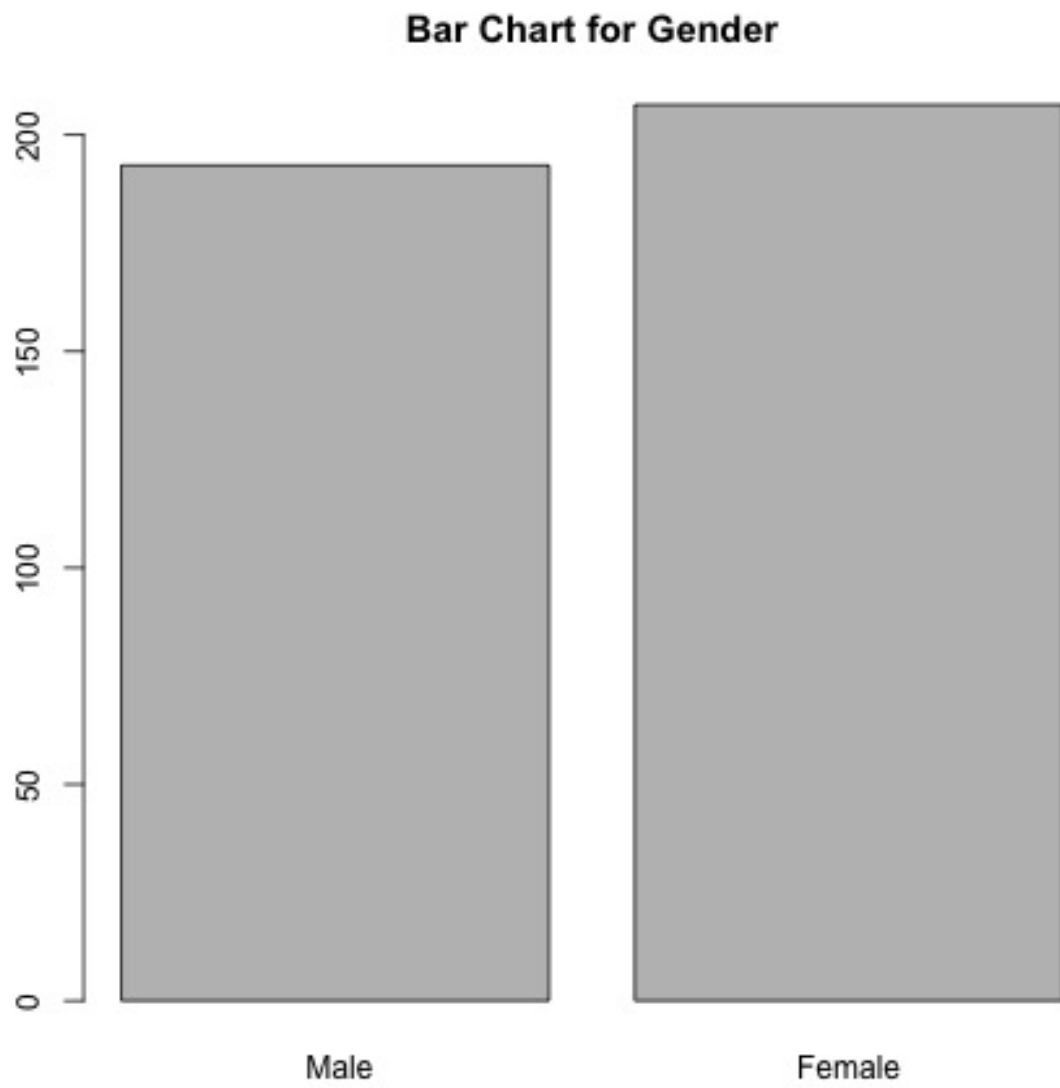

Qualitative Variables

There are 4 qualitative variables in the dataset, including Ethinicity, Gender, Married, Student. For each of the variable, we computed the frequency of each category within the variable and visualize the results by Bar Chart.
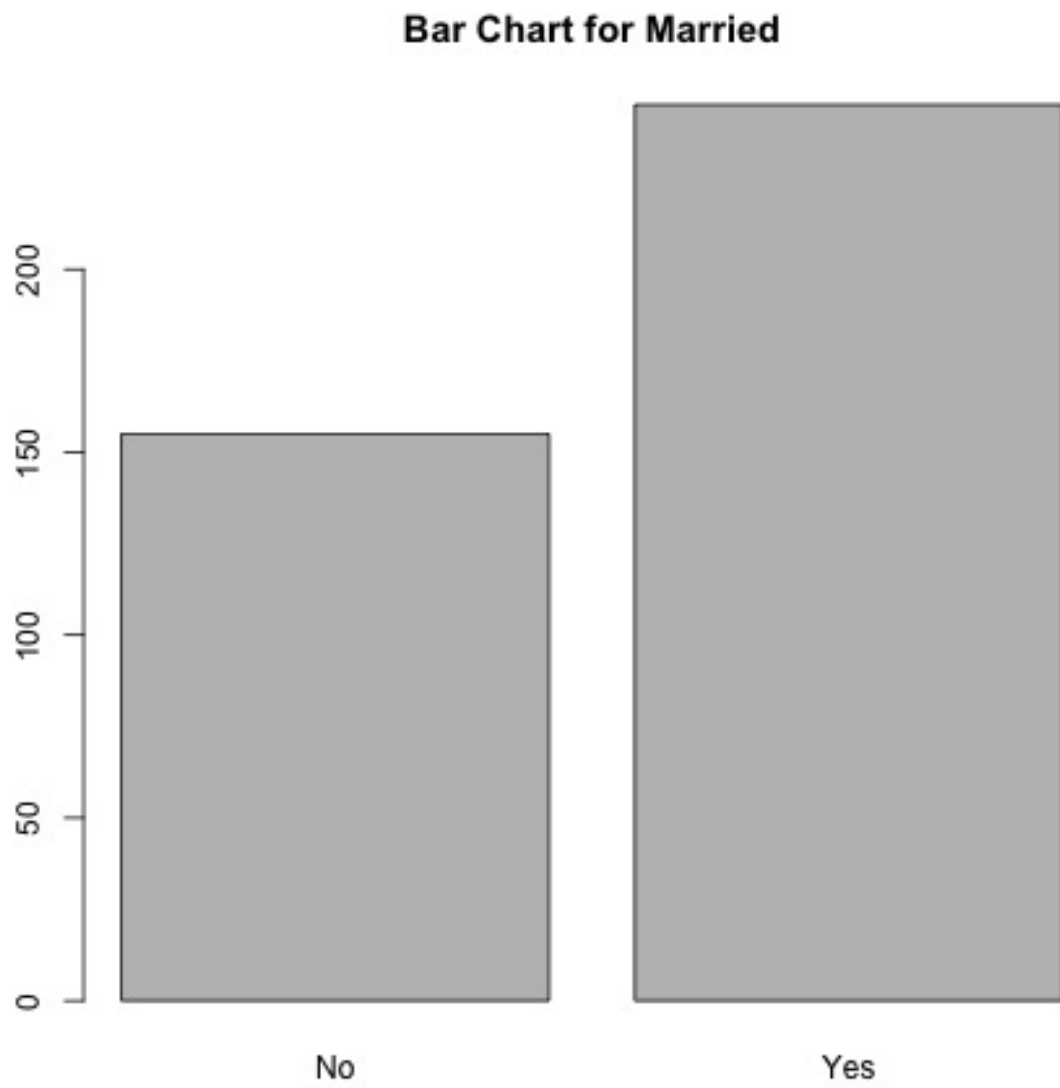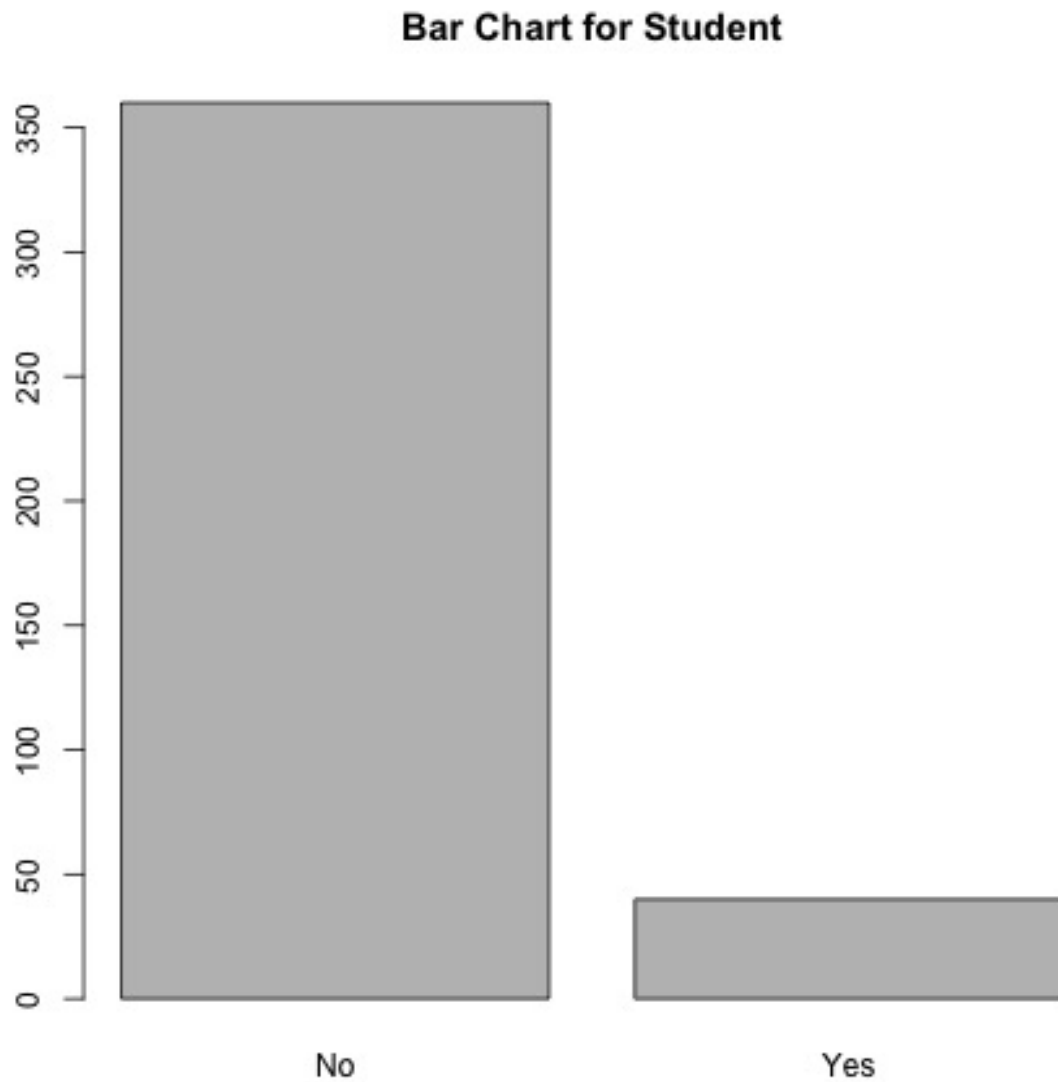
Ethinicity

Bar Chart for Ethnicity

Gender

Bar Chart for Gender

Married

## Bar Chart for Married

Student

Bar Chart for Student

Association

To study the association between predictors and response variable, we first calculate the correlation matrix among all predictors.

```
            Income       Limit      Rating       Cards         Age    Education      Balance
Income    1.00000000  0.79208834  0.79137763 -0.01827261 0.175338403 -0.027691982  0.463656457
Limit     0.79208834  1.00000000  0.99687974  0.01023133 0.100887922 -0.023548534  0.861697267
Rating    0.79137763  0.99687974  1.00000000  0.05323903 0.103164996 -0.030135627  0.863625161
Cards    -0.01827261  0.01023133  0.05323903  1.00000000 0.042948288 -0.051084217  0.086456347
Age       0.17533840  0.10088792  0.10316500  0.04294829 1.000000000  0.003619285  0.001835119
Education -0.02769198 -0.02354853 -0.03013563 -0.05108422 0.003619285  1.000000000 -0.008061576
Balance   0.46365646  0.86169727  0.86362516  0.08645635 0.001835119 -0.008061576  1.000000000
```

Then we visualize the scatterplot mattrix

17
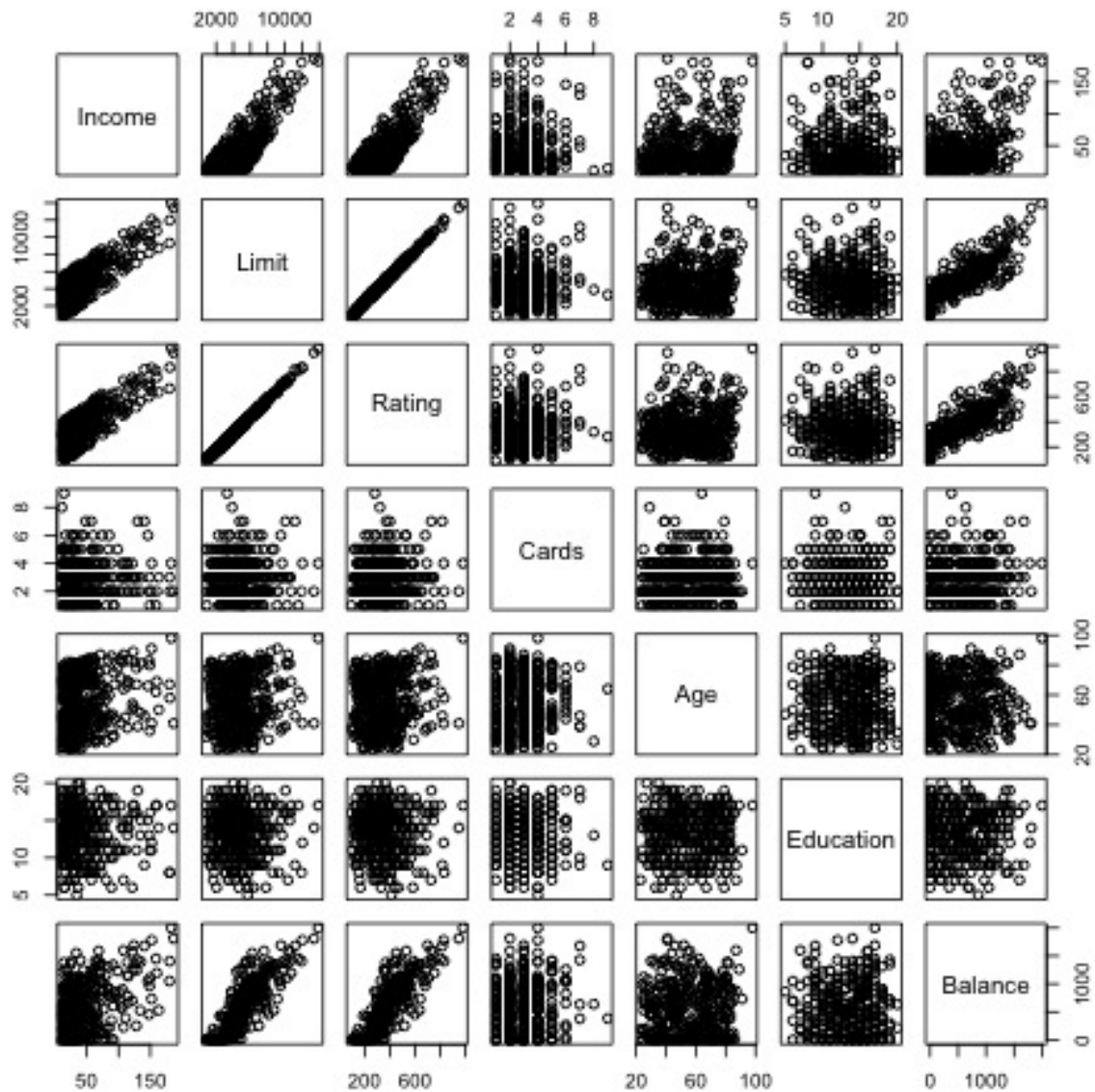
Analysis of variance (ANOVA) is a collection of statistical models used to analyze the differences among group means and their associated procedures (such as "variation" among and between groups)

```
Terms:
                credit$Income credit$Limit credit$Rating credit$Cards credit$Age credit$Education Residuals
Sum of Squares      18131167     55337912       432836        63557        90221           15437  10268781
Deg. of Freedom            1            1            1            1            1               1       393

Residual standard error: 161.6453
Estimated effects may be unbalanced
```
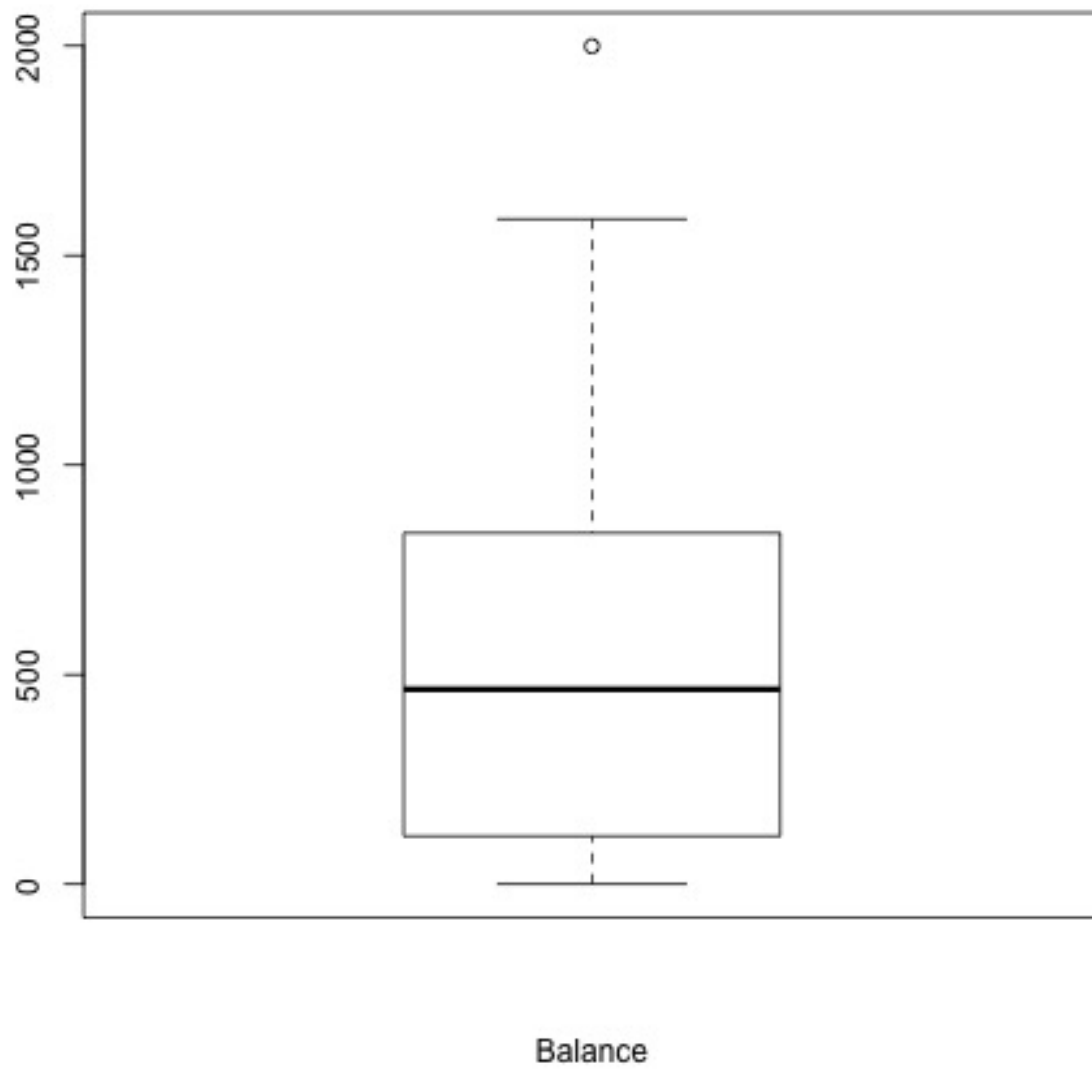
Finally we visualize the conditional boxplots of Balance conditioned to each category of Gender, Ethnicity, Student and Married.
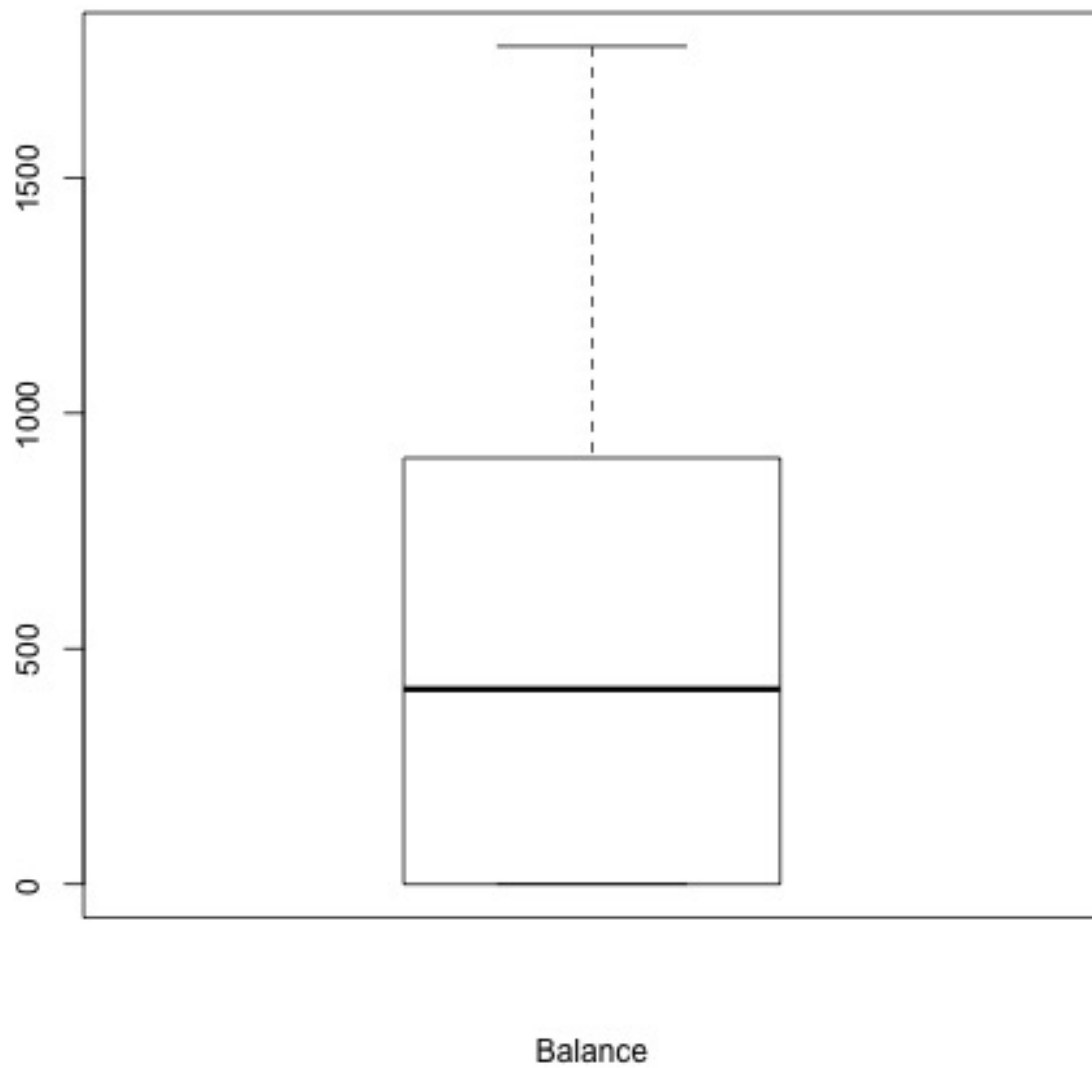
Ethinicity

Conditional Boxplot of Balance of African American
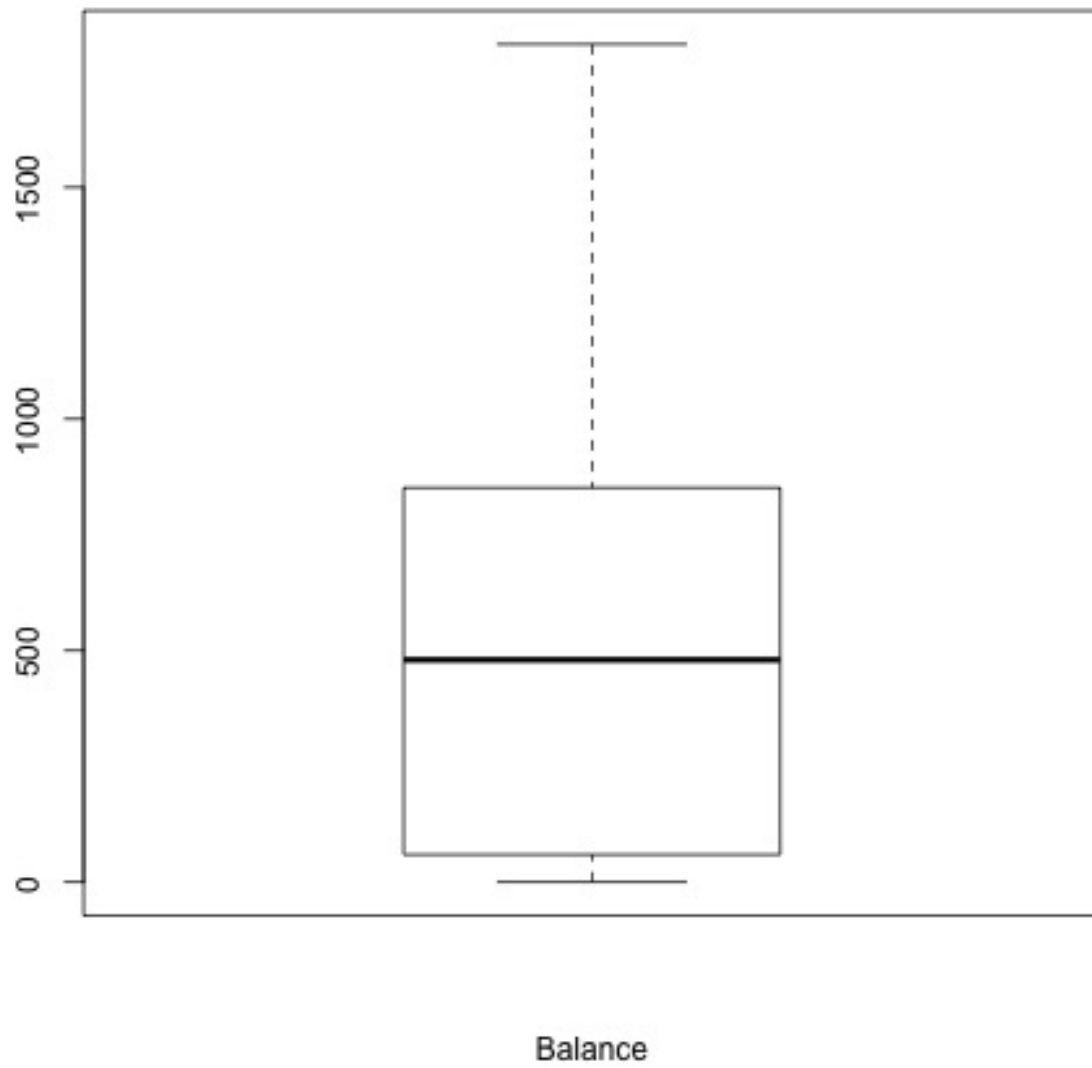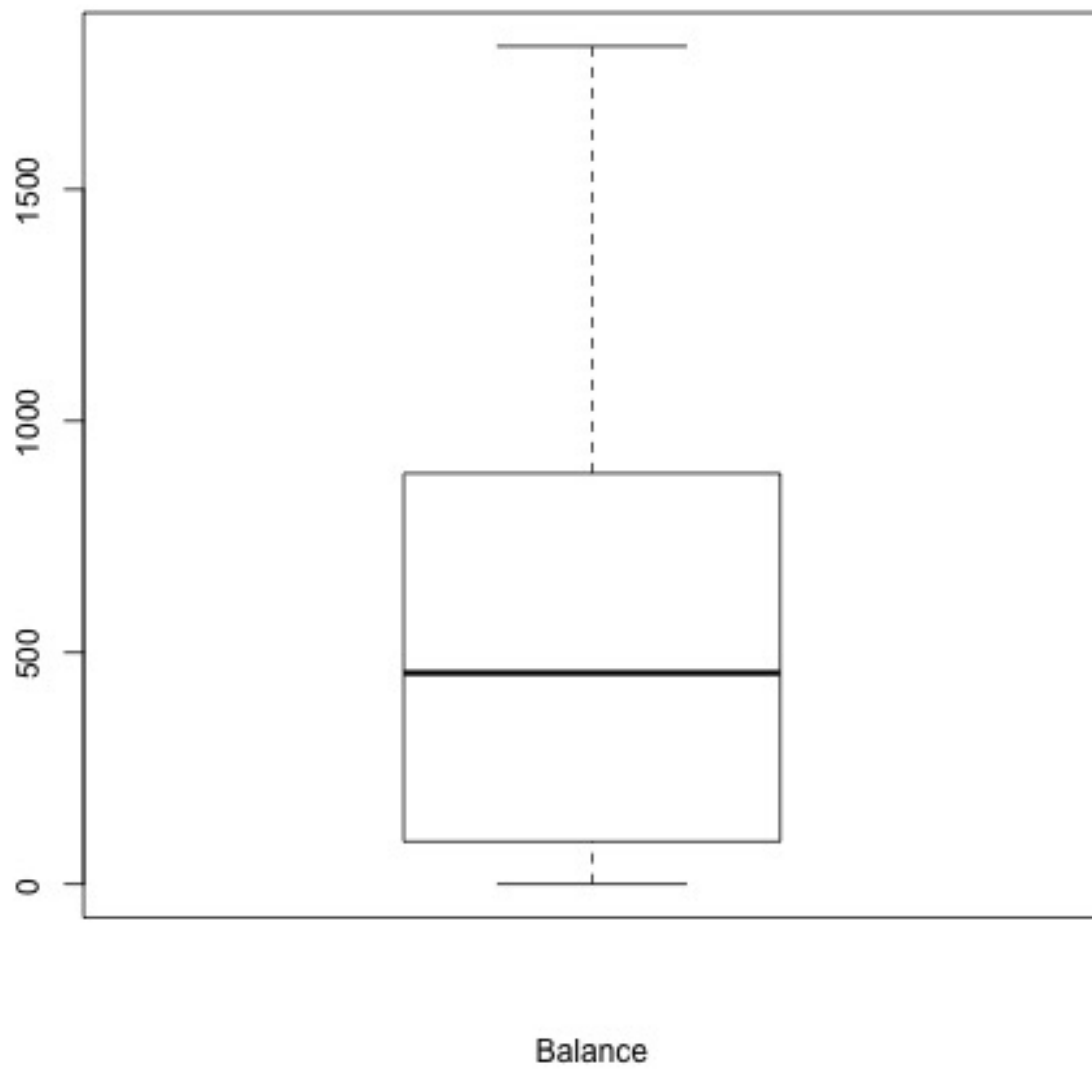

Balance

Conditional Boxplot of Balance of Caucasian

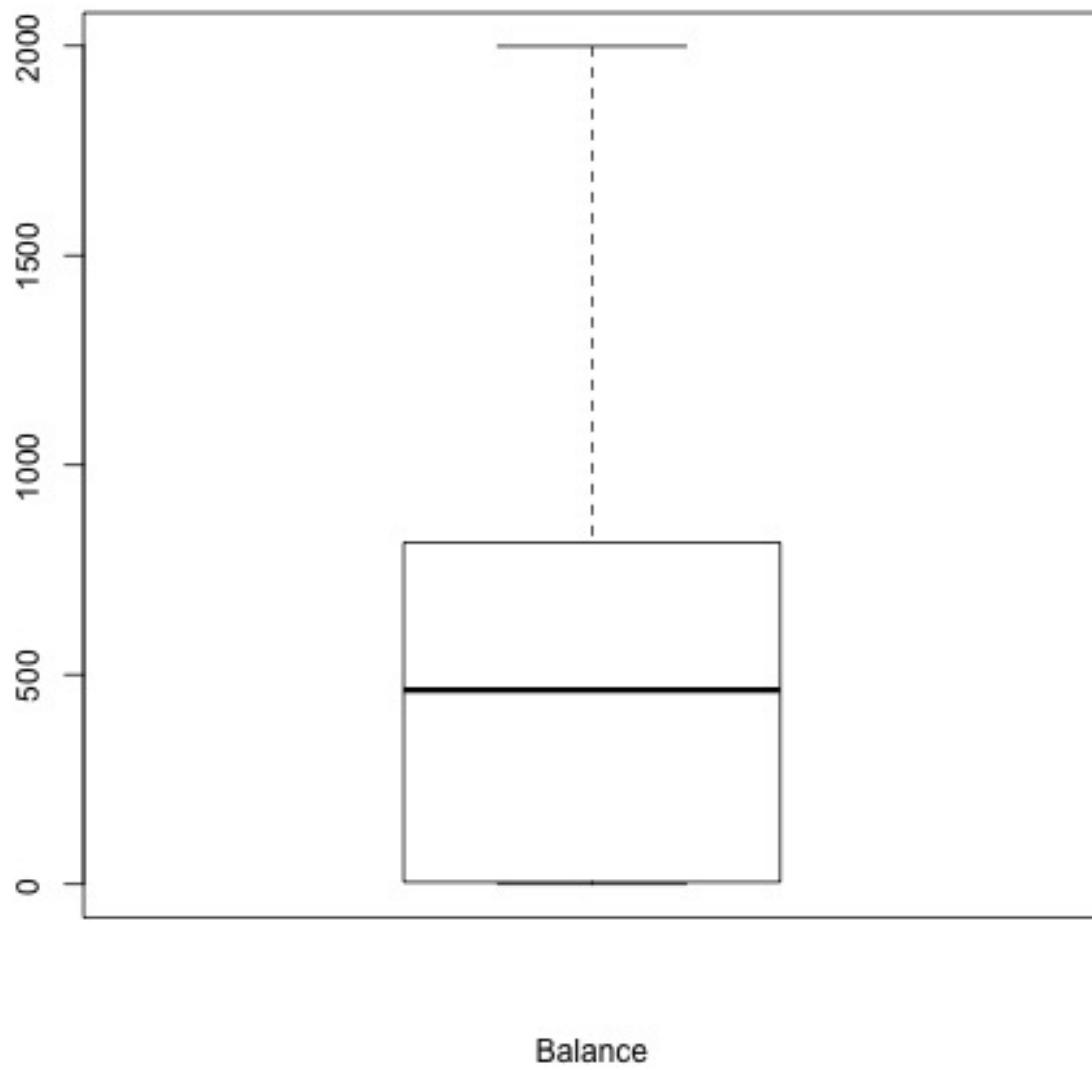

Balance

Conditional Boxplot of Balance of Asian

Balance

Gender

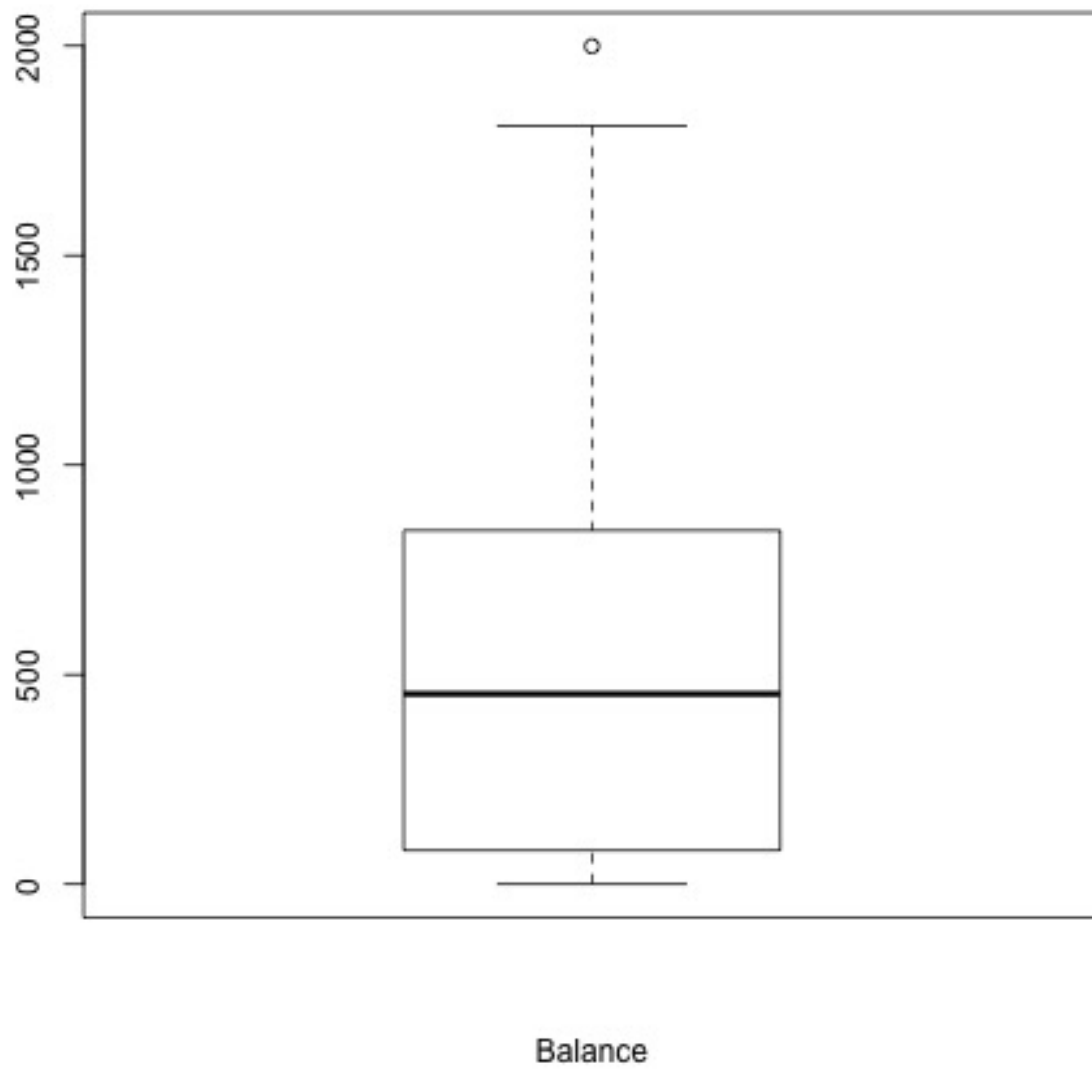## Conditional Boxplot of Balance of Female

Balance

Conditional Boxplot of Balance of Male
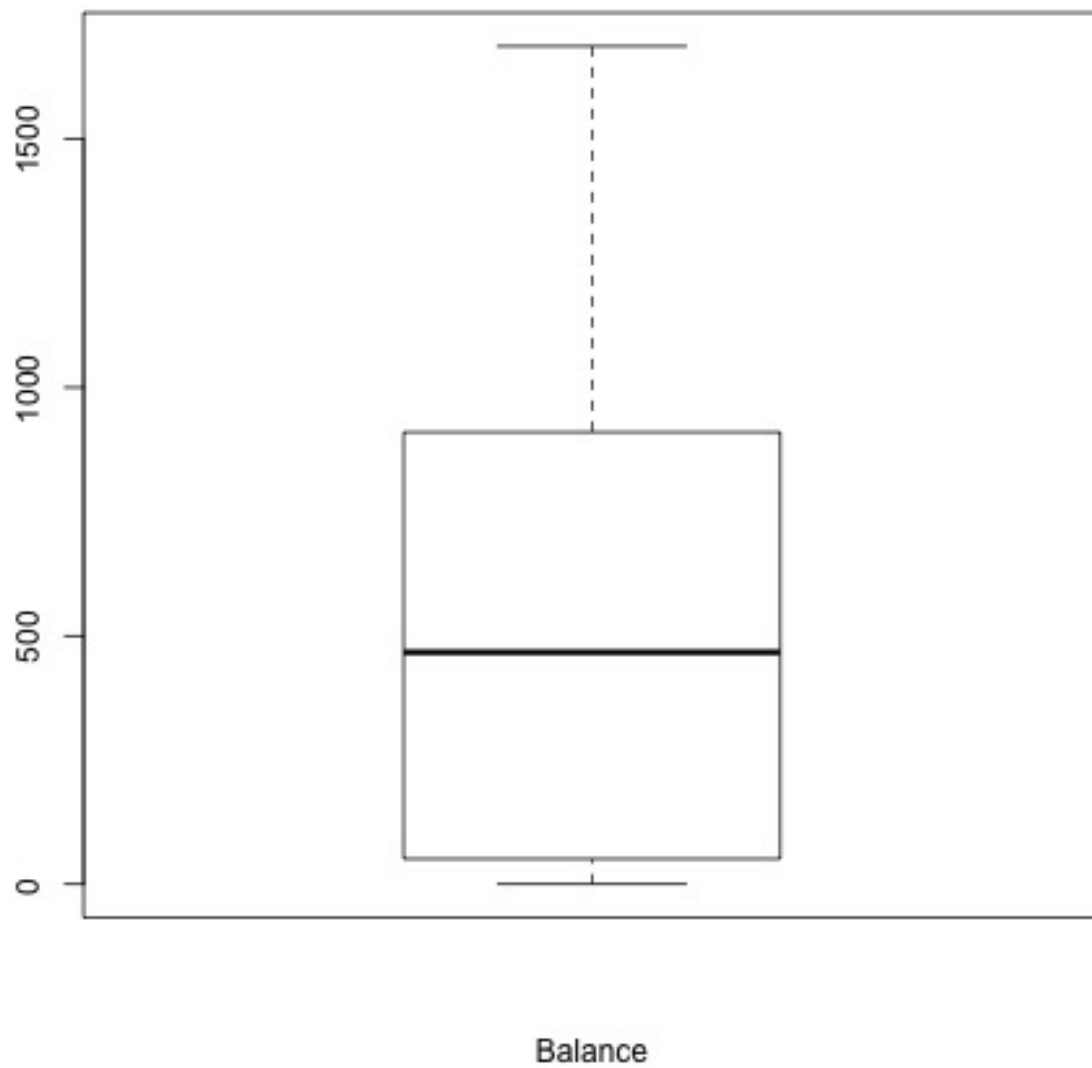

Balance

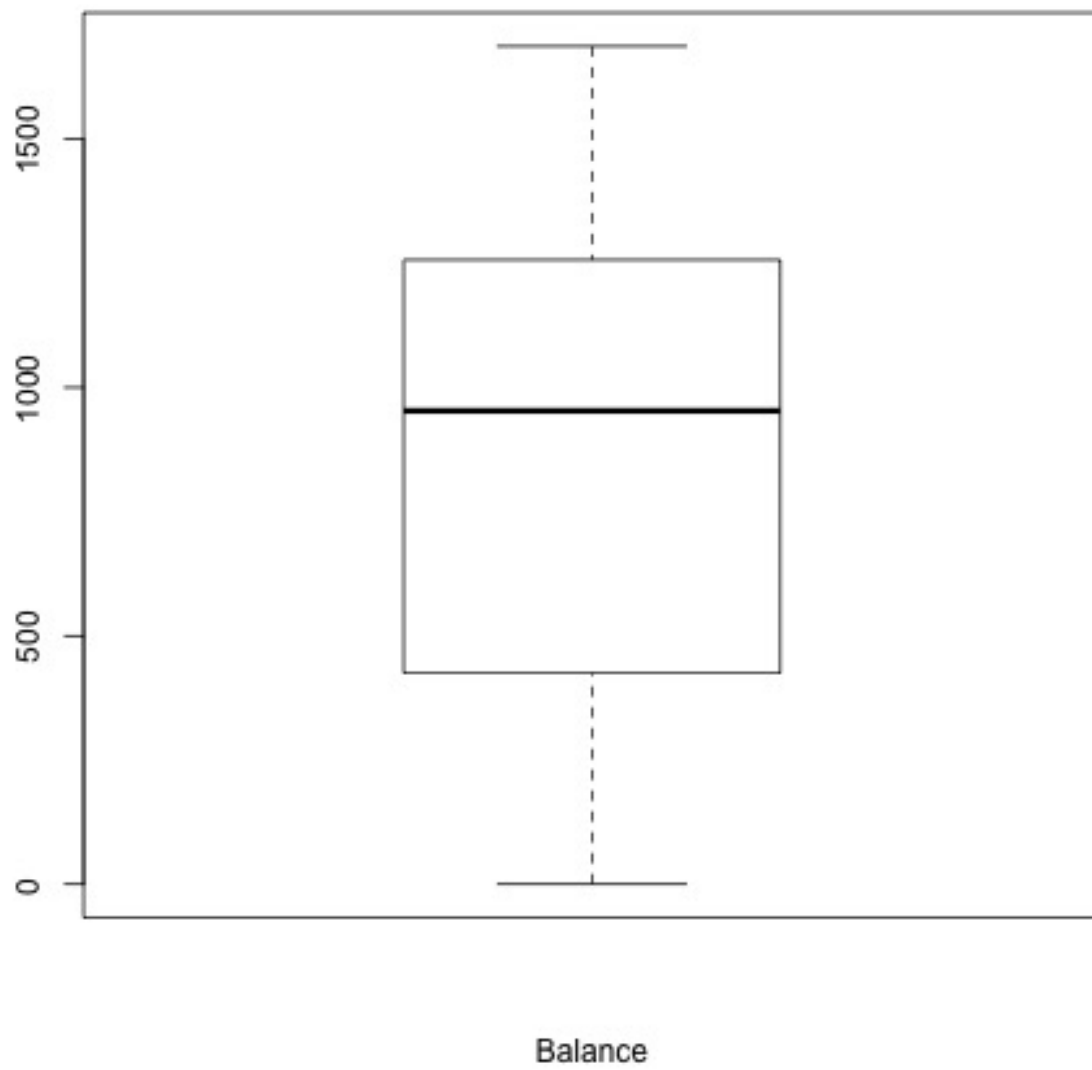Married

Conditional Boxplot of Balance of Married

Balance

Conditional Boxplot of Balance of Unmarried

Balance

Student

Conditional Boxplot of Balance of Student

Balance

Conditional Boxplot of Balance of Nonstudent

Balance