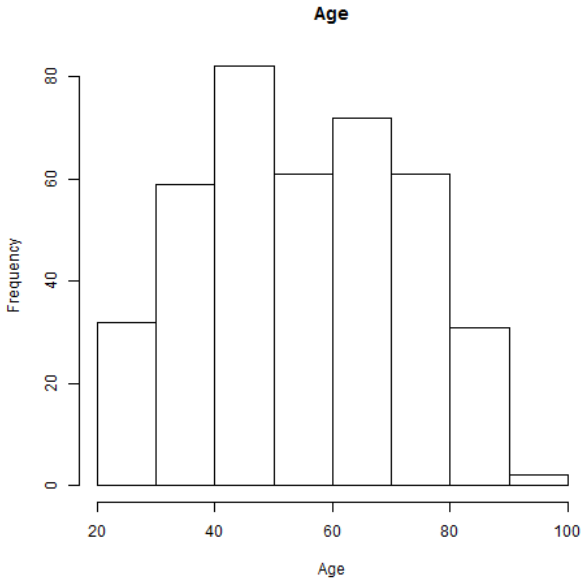
Data

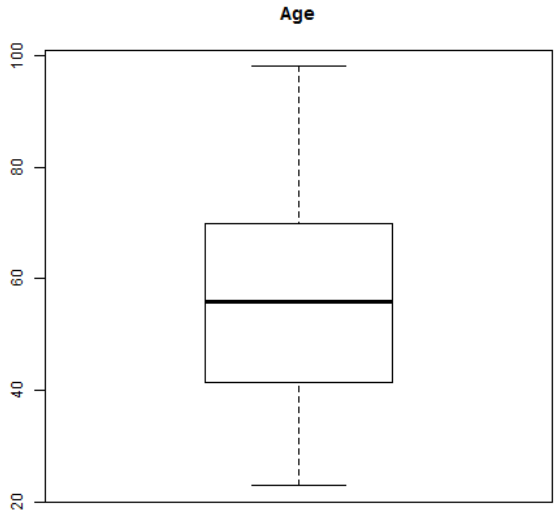
The `Credit.csv` dataset consists of `balance` observations (average credit card debt) for 400 different individuals, as well as observations for a number of quantitative and qualitative variables detailed below:

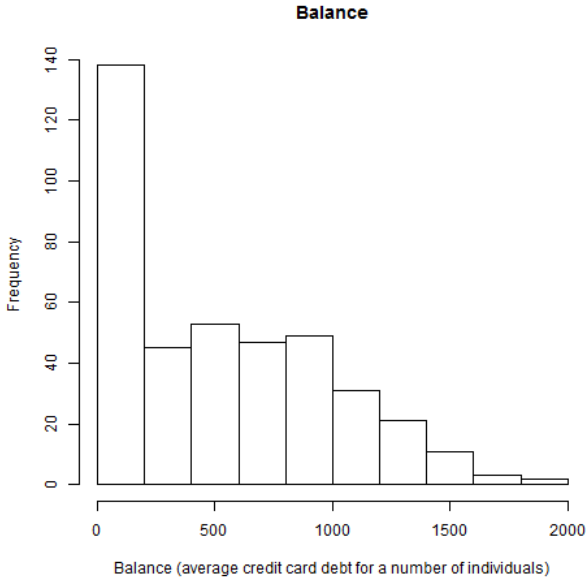
Quantitative Variables

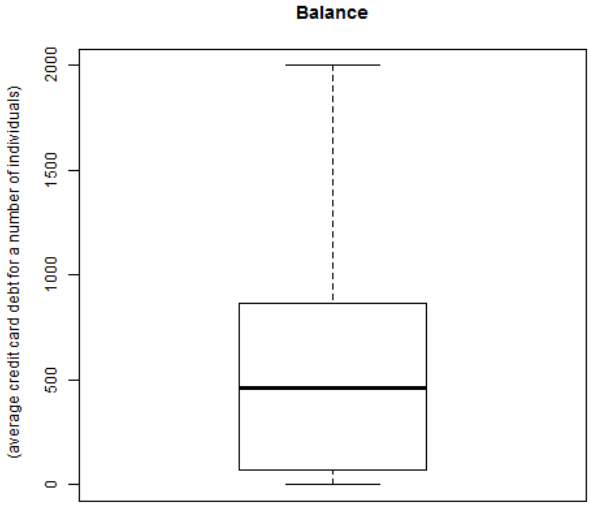
- `age`
- `cards` (number of credit cards)
- `education` (years of education)
- `income`(in thousands of dollars)
- `limit` (credit limit)
- `rating` (credit rating)

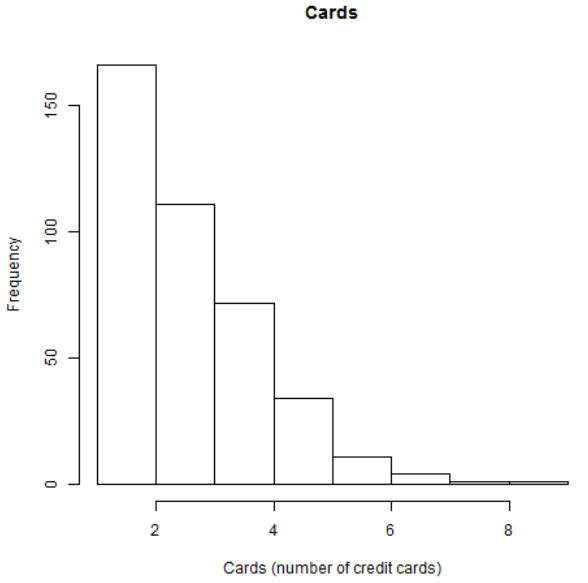
A very general overview of the distribution for each variable is provided in the histograms and boxplots below:

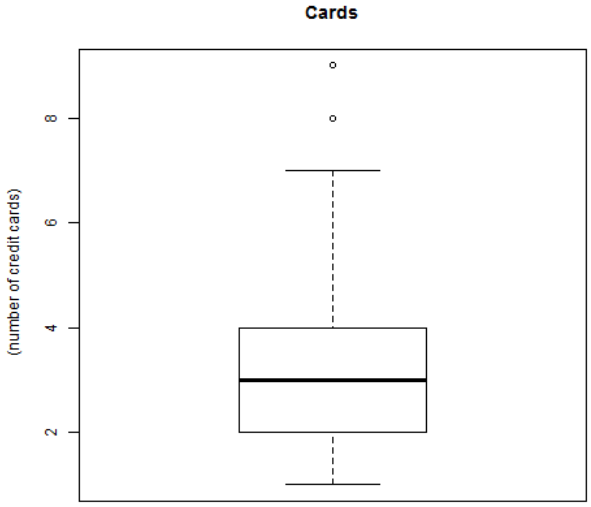


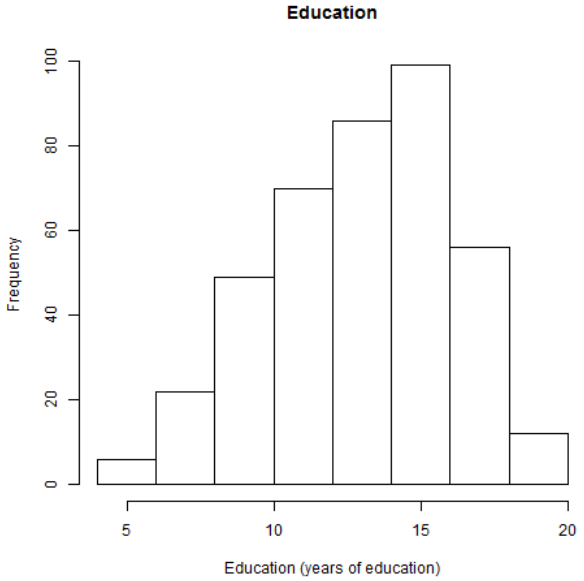


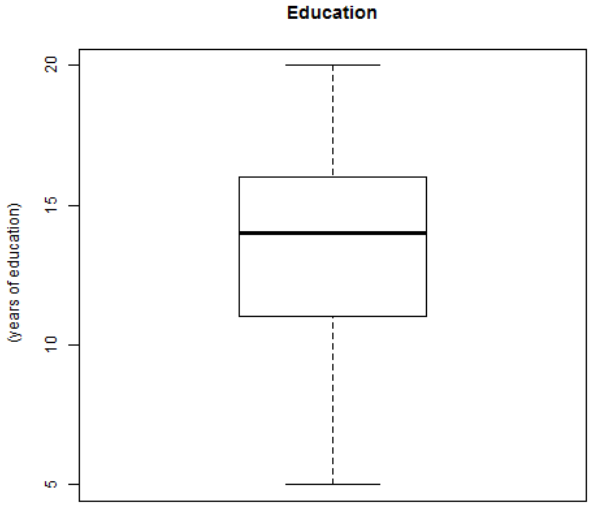

Balance
Frequency
0
20
40
60
80
100
120
140
0
500
1000
1500
2000
Balance (average credit card debt for a number of individuals)

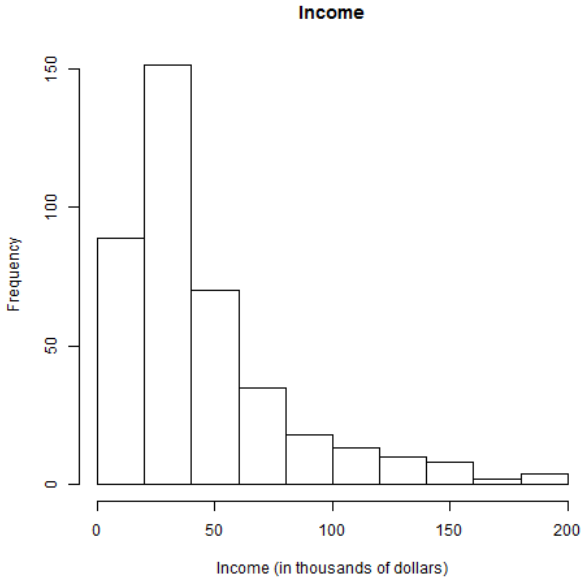

Balance
(average credit card debt for a number of individuals)
0
500
1000
1500
2000

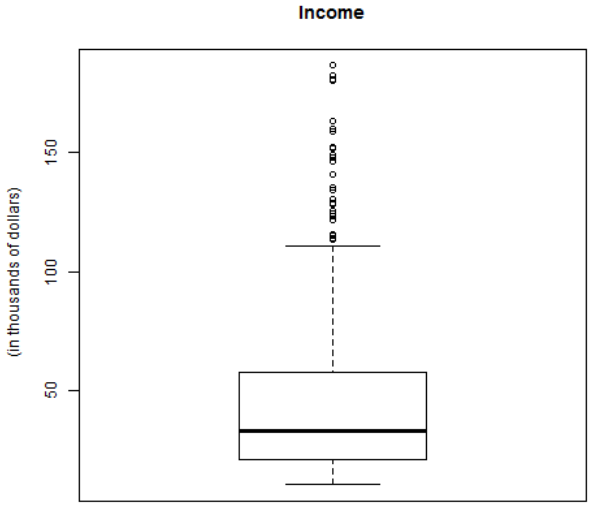

Cards
Frequency
0
50
100
150
2
4
6
8
Cards (number of credit cards)

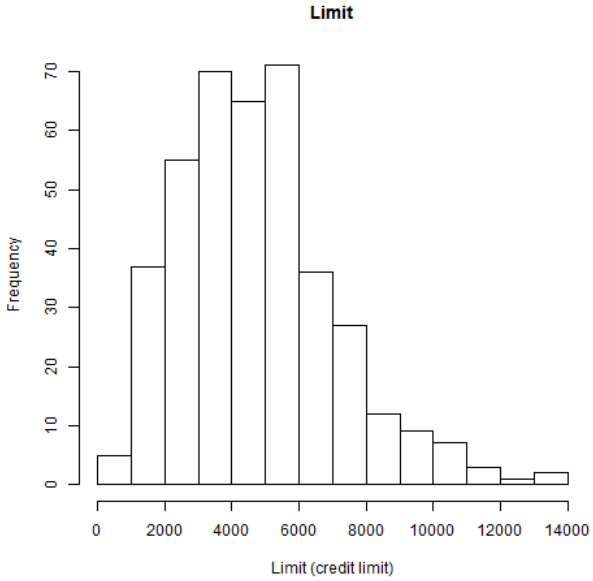

Cards
(number of credit cards)
2
4
6
8

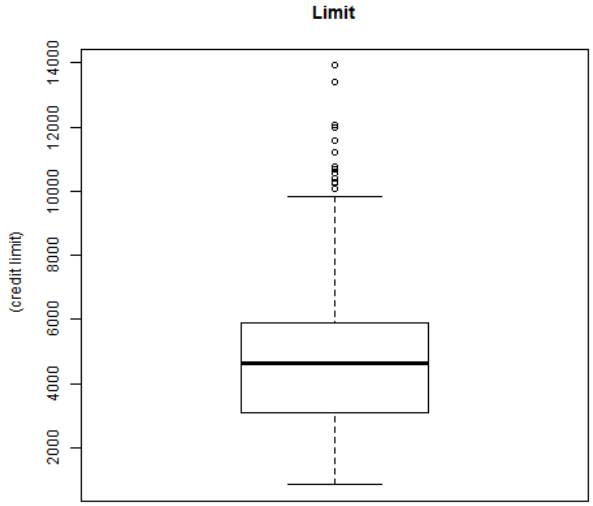

Education
Frequency
0
20
40
60
80
100
5
10
15
20
Education (years of education)

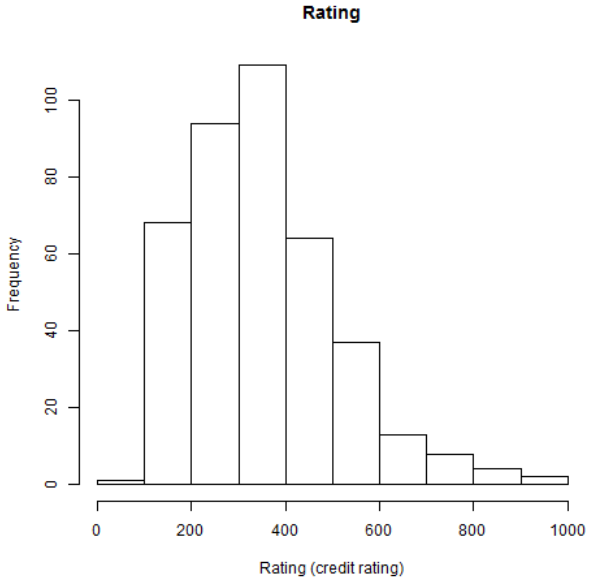

Education
(years of education)
5
10
15
20

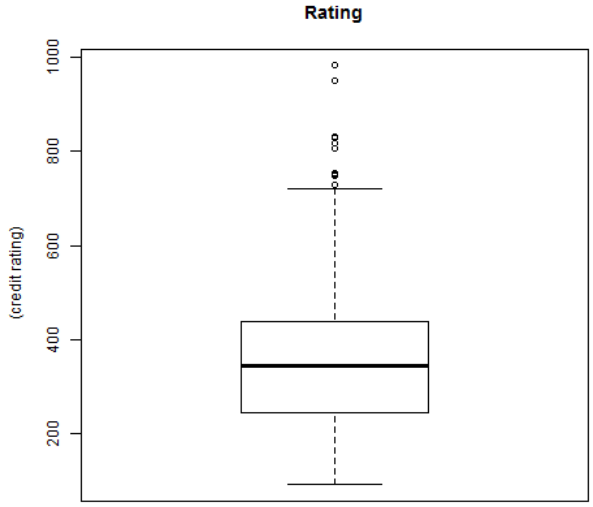

Income
Frequency
0
50
100
150
0
50
100
150
200
Income (in thousands of dollars)


Income
(in thousands of dollars)
50
100
150









A scatterplot matrix for all quantitative variables and a matrix of correlations are displayed below:

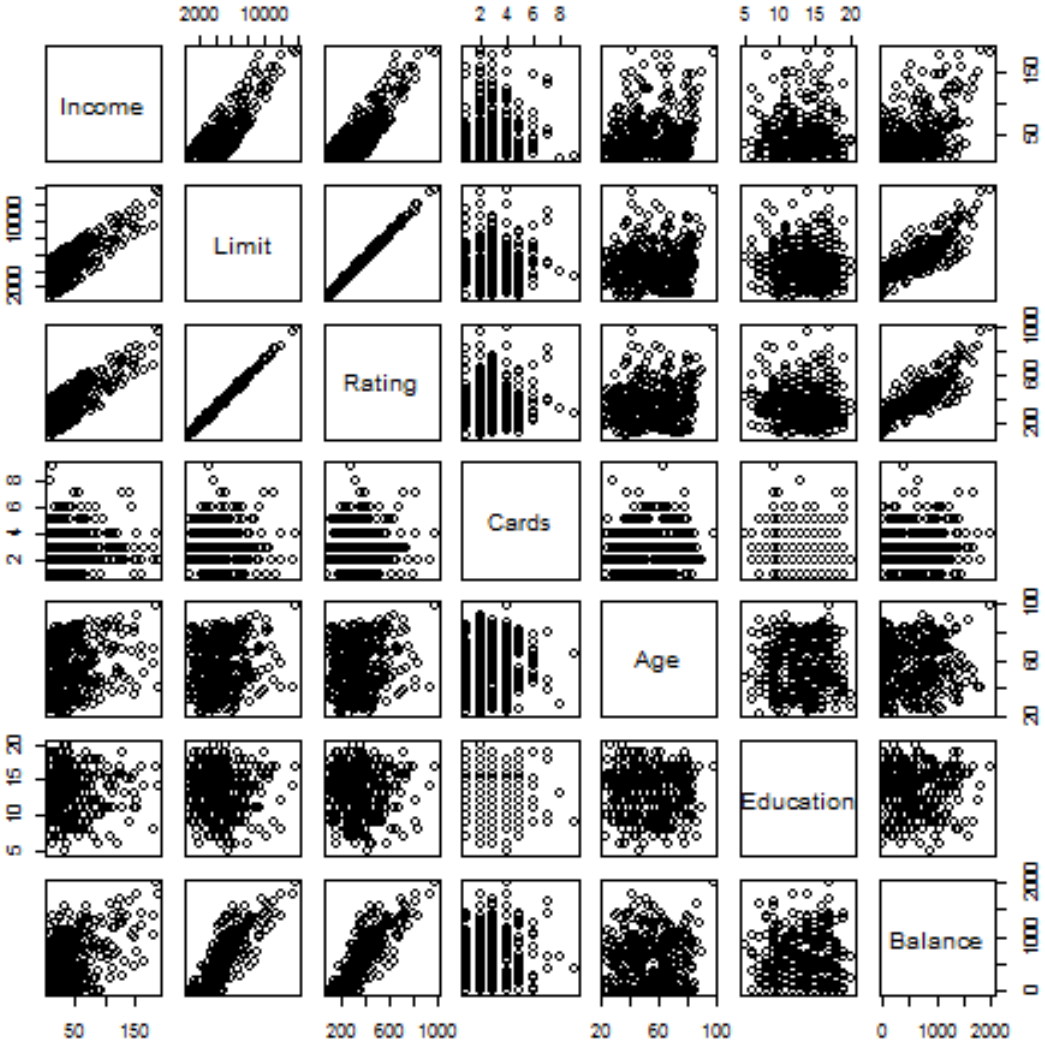


Table 1: Matrix of Correlations for all Quantitative Variables

	Income	Limit	Rating	Cards	Age	Education	Balance
Income	1.0000	0.7921	0.7914	-0.0183	0.1753	-0.0277	0.4637
Limit	0.7921	1.0000	0.9969	0.0102	0.1009	-0.0235	0.8617
Rating	0.7914	0.9969	1.0000	0.0532	0.1032	-0.0301	0.8636
Cards	-0.0183	0.0102	0.0532	1.0000	0.0429	-0.0511	0.0865
Age	0.1753	0.1009	0.1032	0.0429	1.0000	0.0036	0.0018
Education	-0.0277	-0.0235	-0.0301	-0.0511	0.0036	1.0000	-0.0081
Balance	0.4637	0.8617	0.8636	0.0865	0.0018	-0.0081	1.0000