

The project will focus on explanatory data analytics. We have the data description as below

Loan\_ID: A unique loan number assigned to each loan customer. We have a total 500 Loan ID which means 500 loan contracts.

Loan\_status: Whether a loan is paid off, in collection (20%), new customer yet to payoff (20%), or paid off after the collection efforts (60%)

Principal: Basic principal loan amount at the origination i.e. the amount of money customers borrow. The principal amount is not high in this data with the average is $943.2

Terms: Can be weekly (7 days), biweekly, and monthly payoff schedule. In this data we can see the most of terms vary from 7 to 30 days (short-term)

Effective\_date: When the loan got originated and took effects ranging from 9/8/2016 to 9/14/2016

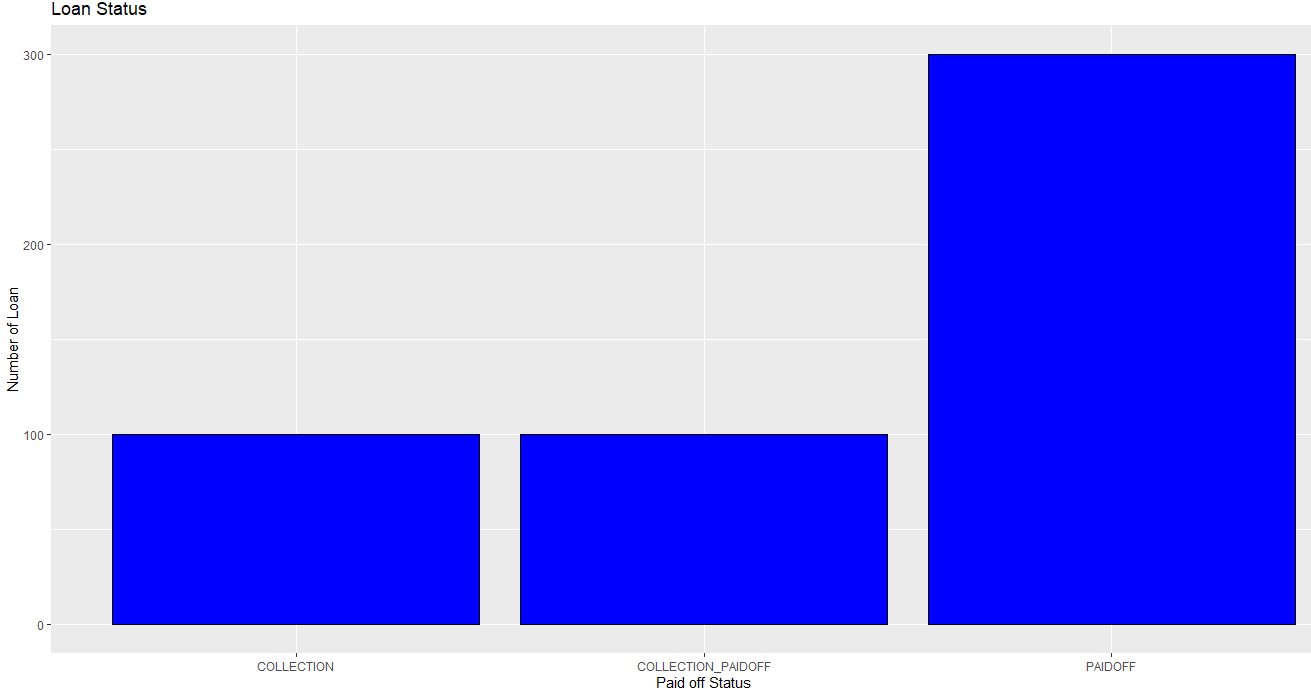
Due\_date Since it is one-time payoff schedule, each loan has one single due date

Paidoff\_time The actual time a customer pays off the loan

Pastdue\_days How many days a loan has been past due

Age, education, gender A customer’s basic demographic information

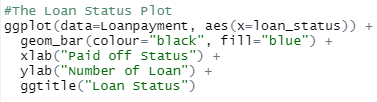
## The Loan Status Plot



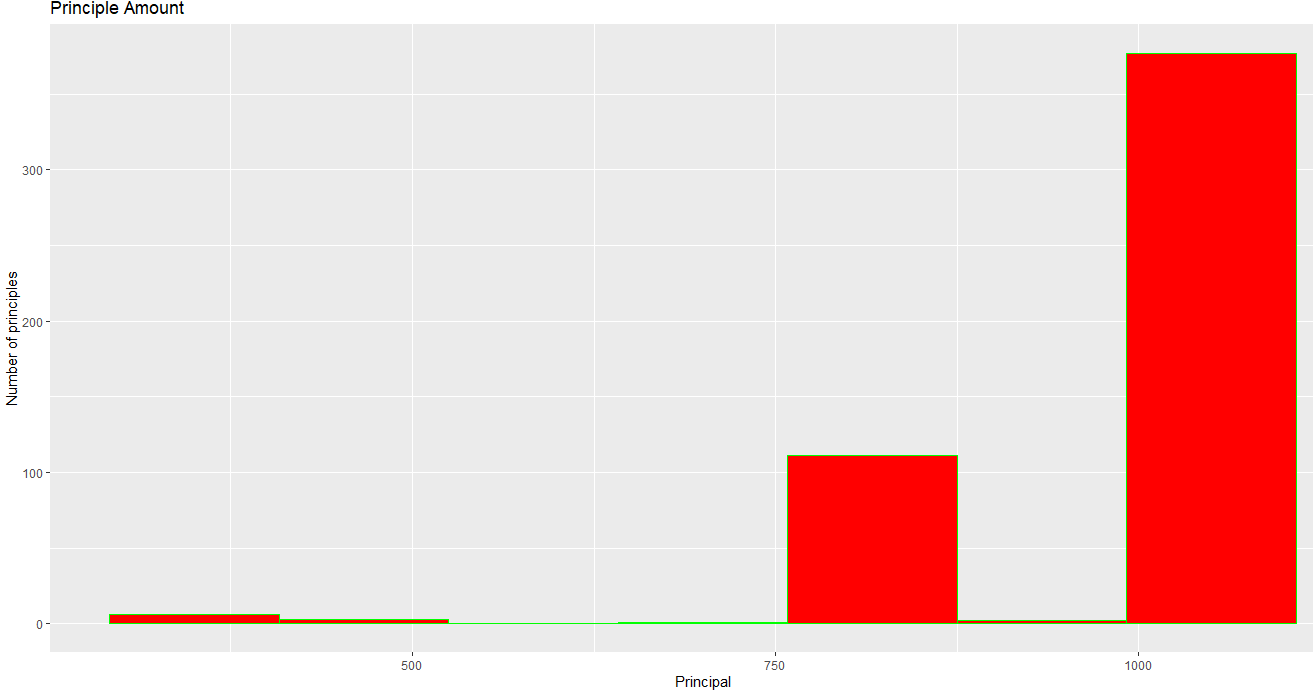
Loan Status indicates that whether a loan is paid off, in collection, new customer yet to pay off, or paid off after the collection efforts.

The collection and collection paid off status account for 20% each and 60% for paid off status.

R code:

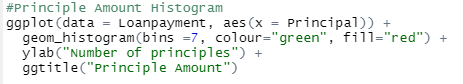


## Principle Amount Histogram

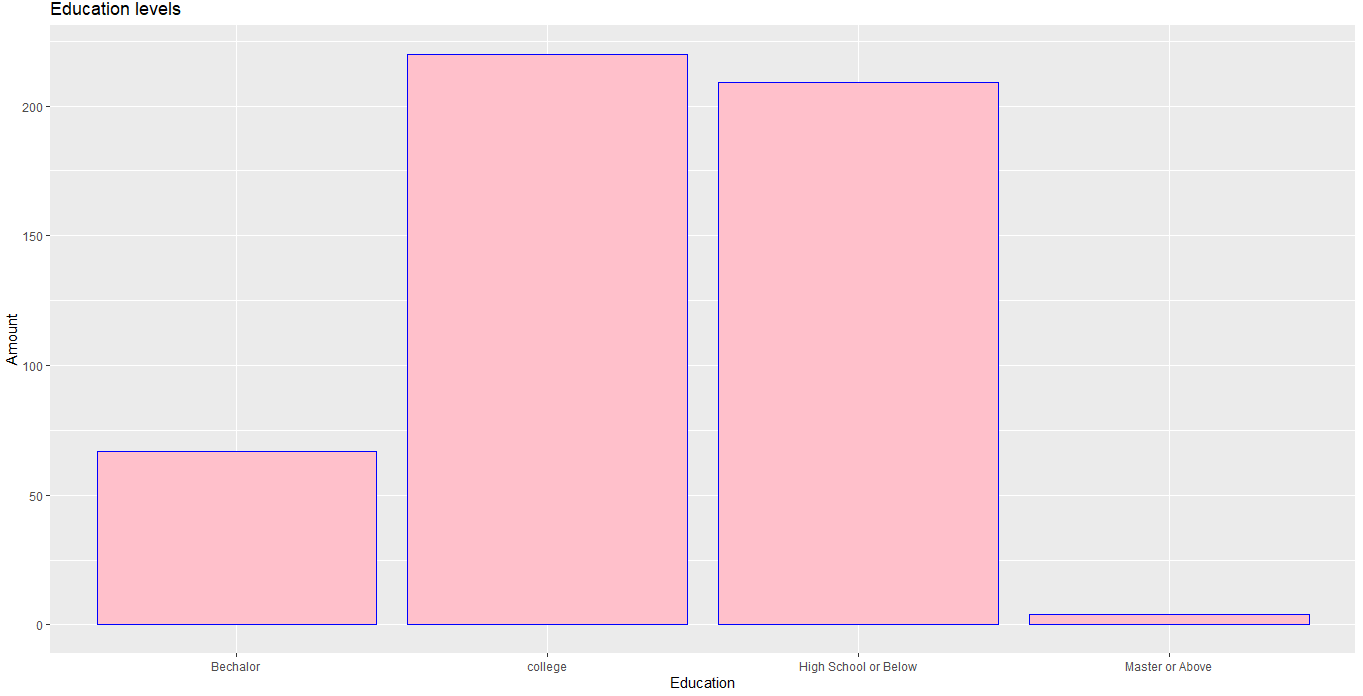


As you can see, the two most principal amounts that customers are likely to borrow are $750-$800 and over $1000. About 375 over 500 customers borrow the amount over $1000 while about 110 customers borrow $750-$800.

R code:

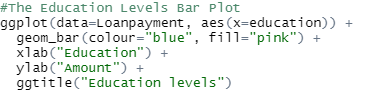


## The Education Levels Bar Plot

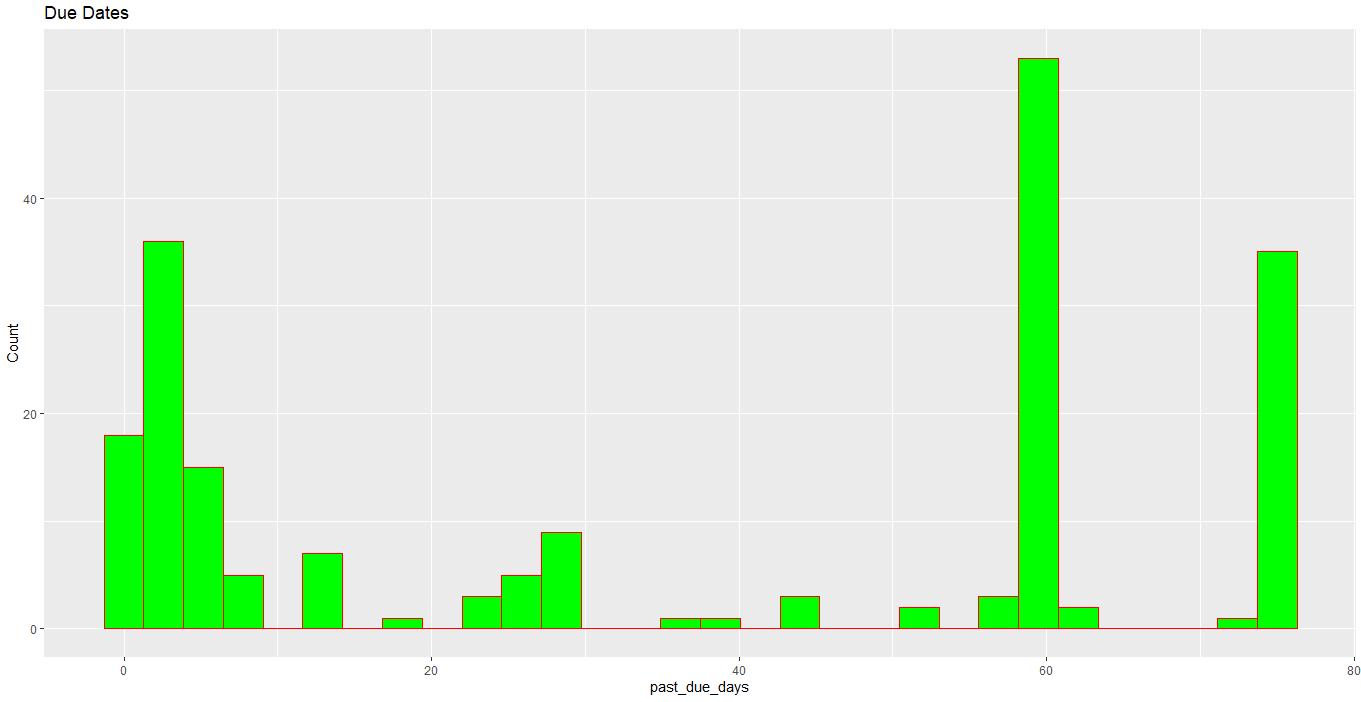


About 291 customers have the education levels are over high school while 209 ones are high are or below education levels. The rate is equivalent.

R code:



## Past due dates histogram

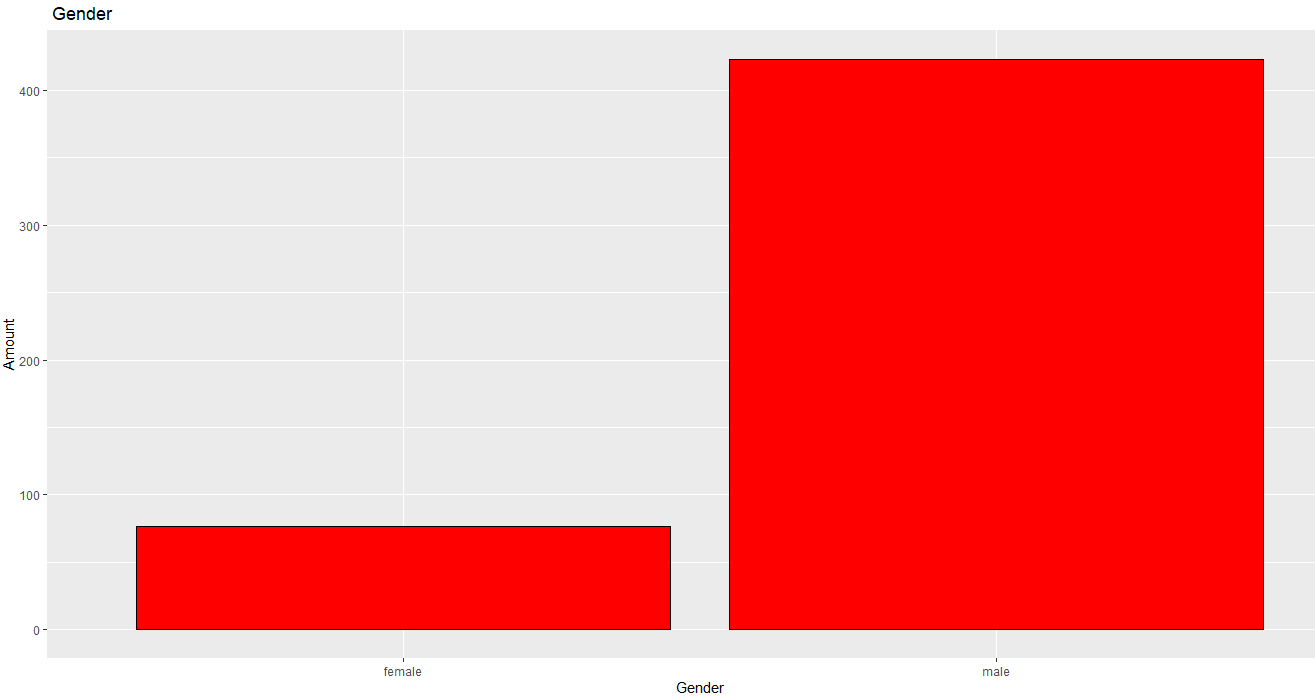
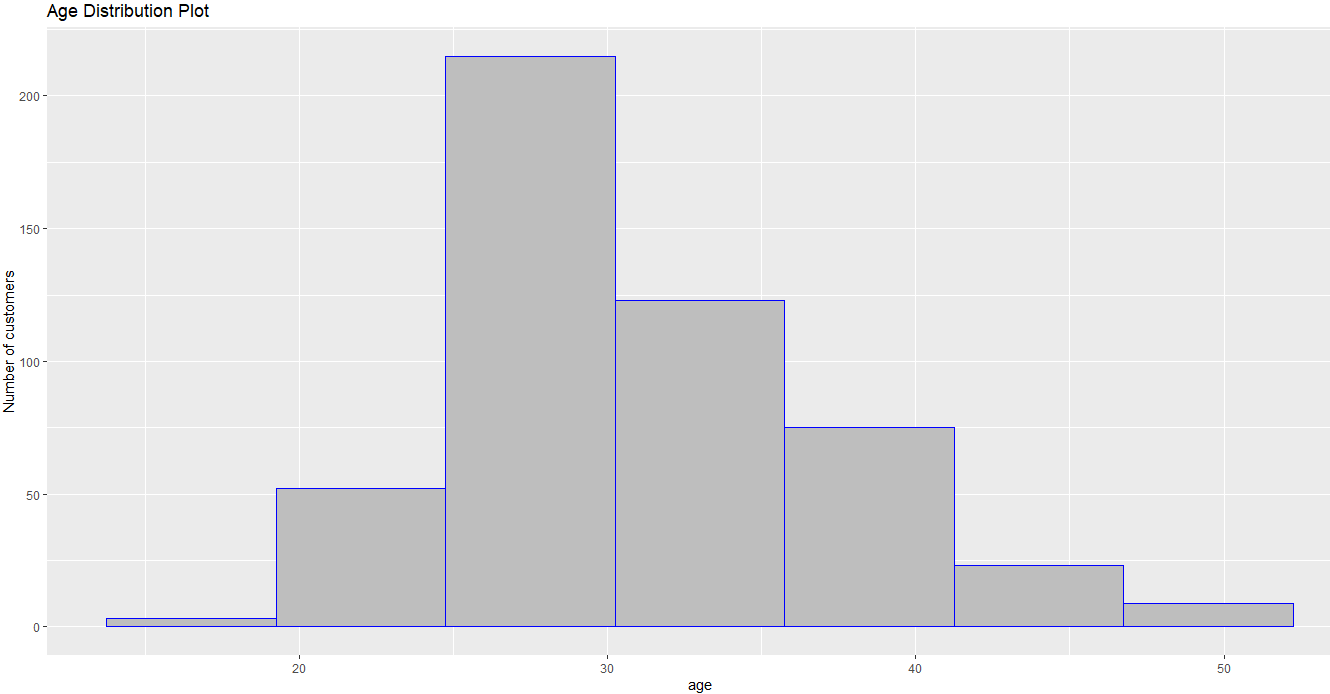


The average number of days that customers past due are 36 days. Most customers have past due dates of 60 days (57 people). Many customers will pay when the due date is about 30 days. However, there are quite a lot of customers who have the past due date up to 75 days (about 35 customers).

R code:



## The Gender Levels Bar Plot

There are 77 female and 421 male customers. The female customers are about 18% compared to male customers.

Moreover, the age distribution is like a bell-shaped plot. The labored age customer tends to borrow much more than the other age range especially from 25 to 40 age range.