

Summary of Data

```
## [1] 113937  
## [1] 81
```

Prosper which was founded in 2005 is a peer-to-peer lending platform that people can invest in each other. Prosper connects people who need money with those who have money to invest. Prosper loan data contain 113,937 rows with 81 columns.

```
## [1] "ListingKey"  
## [2] "ListingNumber"  
## [3] "ListingCreationDate"  
## [4] "CreditGrade"  
## [5] "Term"  
## [6] "LoanStatus"  
## [7] "ClosedDate"  
## [8] "BorrowerAPR"  
## [9] "BorrowerRate"  
## [10] "LenderYield"  
## [11] "EstimatedEffectiveYield"  
## [12] "EstimatedLoss"  
## [13] "EstimatedReturn"  
## [14] "ProsperRating..numeric."  
## [15] "ProsperRating..Alpha."  
## [16] "ProsperScore"  
## [17] "ListingCategory..numeric."  
## [18] "BorrowerState"  
## [19] "Occupation"  
## [20] "EmploymentStatus"  
## [21] "EmploymentStatusDuration"  
## [22] "IsBorrowerHomeowner"  
## [23] "CurrentlyInGroup"  
## [24] "GroupKey"  
## [25] "DateCreditPulled"  
## [26] "CreditScoreRangeLower"  
## [27] "CreditScoreRangeUpper"  
## [28] "FirstRecordedCreditLine"  
## [29] "CurrentCreditLines"  
## [30] "OpenCreditLines"  
## [31] "TotalCreditLinespast7years"  
## [32] "OpenRevolvingAccounts"  
## [33] "OpenRevolvingMonthlyPayment"  
## [34] "InquiriesLast6Months"  
## [35] "TotalInquiries"  
## [36] "CurrentDelinquencies"  
## [37] "AmountDelinquent"  
## [38] "DelinquenciesLast7Years"  
## [39] "PublicRecordsLast10Years"  
## [40] "PublicRecordsLast12Months"  
## [41] "RevolvingCreditBalance"  
## [42] "BankcardUtilization"
```

```

## [43] "AvailableBankcardCredit"
## [44] "TotalTrades"
## [45] "TradesNeverDelinquent..percentage."
## [46] "TradesOpenedLast6Months"
## [47] "DebtToIncomeRatio"
## [48] "IncomeRange"
## [49] "IncomeVerifiable"
## [50] "StatedMonthlyIncome"
## [51] "LoanKey"
## [52] "TotalProsperLoans"
## [53] "TotalProsperPaymentsBilled"
## [54] "OnTimeProsperPayments"
## [55] "ProsperPaymentsLessThanOneMonthLate"
## [56] "ProsperPaymentsOneMonthPlusLate"
## [57] "ProsperPrincipalBorrowed"
## [58] "ProsperPrincipalOutstanding"
## [59] "ScorexChangeAtTimeOfListing"
## [60] "LoanCurrentDaysDelinquent"
## [61] "LoanFirstDefaultedCycleNumber"
## [62] "LoanMonthsSinceOrigination"
## [63] "LoanNumber"
## [64] "LoanOriginalAmount"
## [65] "LoanOriginationDate"
## [66] "LoanOriginationQuarter"
## [67] "MemberKey"
## [68] "MonthlyLoanPayment"
## [69] "LP_CustomerPayments"
## [70] "LP_CustomerPrincipalPayments"
## [71] "LP_InterestandFees"
## [72] "LP_ServiceFees"
## [73] "LP_CollectionFees"
## [74] "LP_GrossPrincipalLoss"
## [75] "LP_NetPrincipalLoss"
## [76] "LP_NonPrincipalRecoverypayments"
## [77] "PercentFunded"
## [78] "Recommendations"
## [79] "InvestmentFromFriendsCount"
## [80] "InvestmentFromFriendsAmount"
## [81] "Investors"

```

Creating new dataframe based on prosper data

```

## 'data.frame': 113937 obs. of 13 variables:
##   $ DelinquenciesLast7Years : int 4 0 0 14 0 0 0 0 0 ...
##   $ PublicRecordsLast10Years: int 0 1 0 0 0 0 0 1 0 0 ...
##   $ DaysWithCreditLine     : num 6241 8274 5952 13041 5379 ...
##   $ InquiriesLast6Months  : int 3 3 0 0 1 0 0 3 1 1 ...
##   $ BorrowerRate           : num 0.158 0.092 0.275 0.0974 0.2085 ...
##   $ Term                  : Factor w/ 3 levels "12","36","60": 2 2 2 2 2 3 2 2 2 ...
##   $ ProsperRating          : Factor w/ 7 levels "AA","A","B","C",...: NA 2 NA 2 5 3 6 4 1 1 ...
##   $ ListingCreationDate    : Factor w/ 113064 levels "2005-11-09 20:44:28.847000000",...: 14184 11189 ...
##   $ LoanOriginalAmount     : int 9425 10000 3001 10000 15000 3000 10000 10000 ...

```

```

## $ ListingCategory      : Factor w/ 21 levels "Not available",...: 1 3 1 17 3 2 2 3 8 8 ...
## $ EmploymentStatus     : Factor w/ 9 levels "", "Employed",...: 9 2 4 2 2 2 2 2 2 ...
## $ AnnualIncome          : num  37000 73500 25000 34500 115000 ...
## $ RevolvingCreditBalance: num  0 3989 NA 1444 6193 ...

## DelinquenciesLast7Years PublicRecordsLast10Years DaysWithCreditLine
## Min.   : 0.000           Min.   : 0.0000           Min.   : 2151
## 1st Qu.: 0.000           1st Qu.: 0.0000           1st Qu.: 6817
## Median : 0.000           Median : 0.0000           Median : 8412
## Mean   : 4.155           Mean   : 0.3126           Mean   : 8761
## 3rd Qu.: 3.000           3rd Qu.: 0.0000           3rd Qu.: 10391
## Max.   :99.000            Max.   :38.0000           Max.   :26013
## NA's   :990              NA's   :697              NA's   :697
## InquiriesLast6Months   BorrowerRate    Term      ProsperRating
## Min.   : 0.000           Min.   :0.0000           12: 1614   C      :18345
## 1st Qu.: 0.000           1st Qu.:0.1340          36:87778  B      :15581
## Median : 1.000           Median :0.1840          60:24545  A      :14551
## Mean   : 1.435           Mean   :0.1928           D      :14274
## 3rd Qu.: 2.000           3rd Qu.:0.2500          E      : 9795
## Max.   :105.000           Max.   :0.4975          (Other):12307
## NA's   :697              NA's   :29084

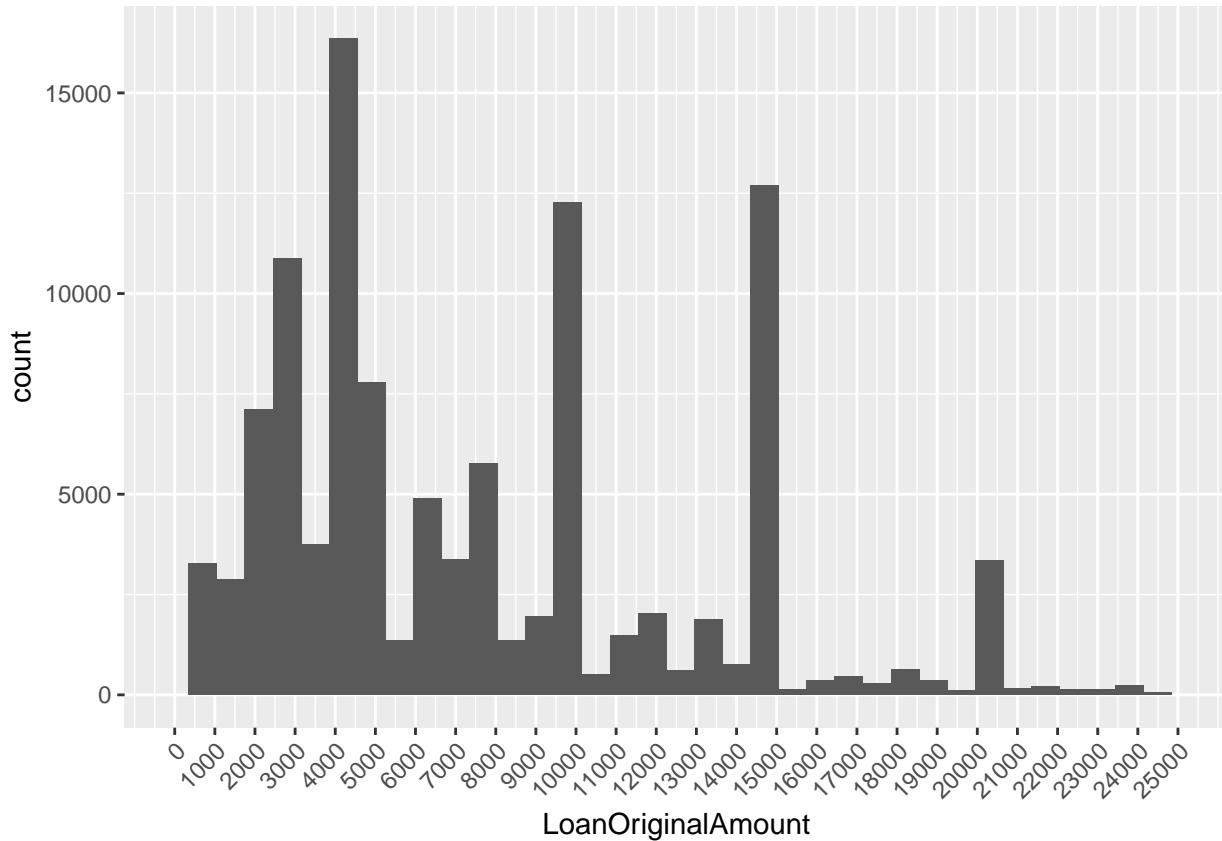
## ListingCreationDate   LoanOriginalAmount
## 2013-10-02 17:20:16.550000000:       6   Min.   : 1000
## 2013-08-28 20:31:41.107000000:       4   1st Qu.: 4000
## 2013-09-08 09:27:44.853000000:       4   Median : 6500
## 2013-12-06 05:43:13.830000000:       4   Mean   : 8337
## 2013-12-06 11:44:58.283000000:       4   3rd Qu.:12000
## 2013-08-21 07:25:22.360000000:       3   Max.   :35000
## (Other)                  :113912

## ListingCategory      EmploymentStatus AnnualIncome
## Debt consolidation:58308 Employed      :67322   Min.   : 0
## Not available     :16965 Full-time    :26355   1st Qu.: 38404
## Other             :10494 Self-employed: 6134   Median : 56000
## Home improvement : 7433 Not available: 5347   Mean   : 67296
## Business          : 7189 Other        : 3806   3rd Qu.: 81900
## Auto              : 2572                : 2255   Max.   :21000035
## (Other)           :10976 (Other)      : 2718

## RevolvingCreditBalance
## Min.   : 0
## 1st Qu.: 3121
## Median : 8549
## Mean   : 17599
## 3rd Qu.: 19521
## Max.   :1435667
## NA's   :7604

```

Univariate Plots Section

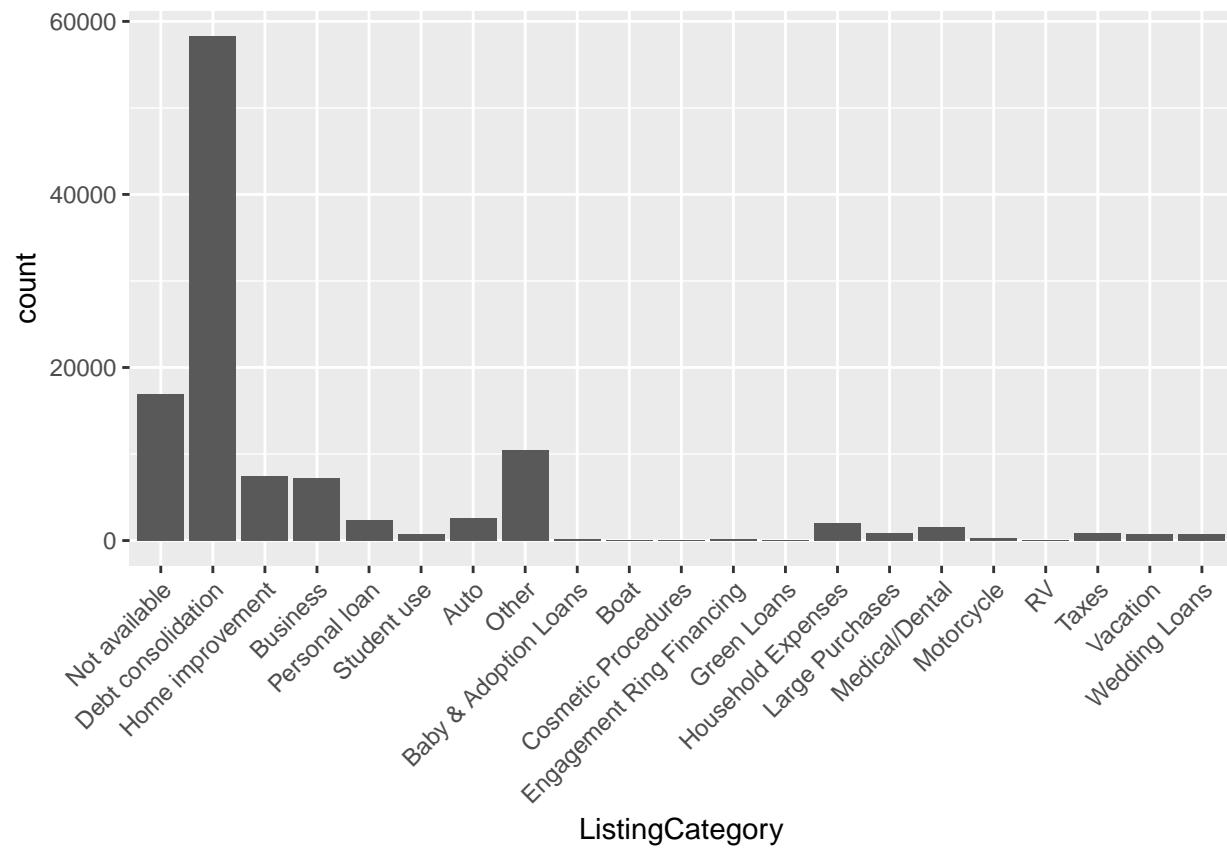


From the figure above we can see the amount of money that people borrow and as we can see people tend to borrow in whole numbers.

```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##     1000    4000   6500    8337   12000   35000
```

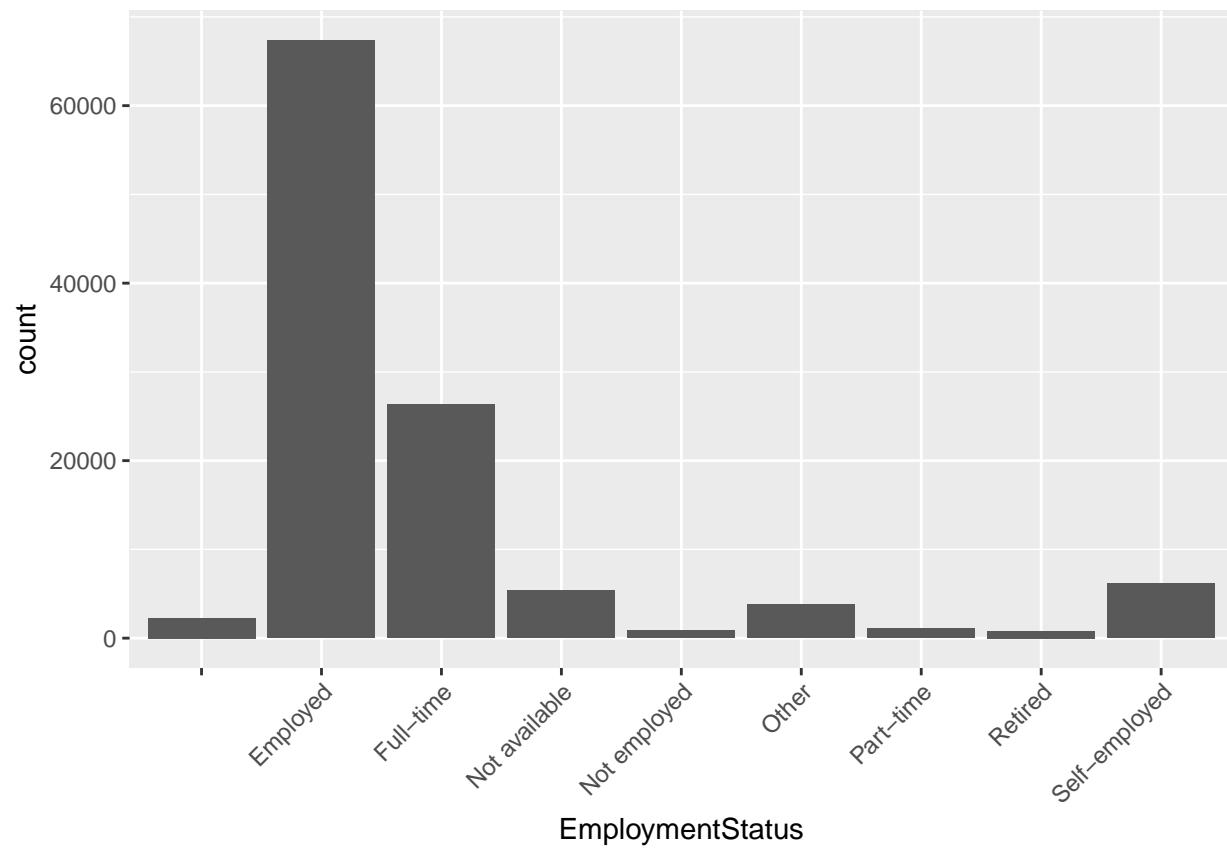
The minimum money that people borrowed is 1000 as opposed to maximum, which is 35000. the median and mean are 6500 and 8337 respectively.

Loan category



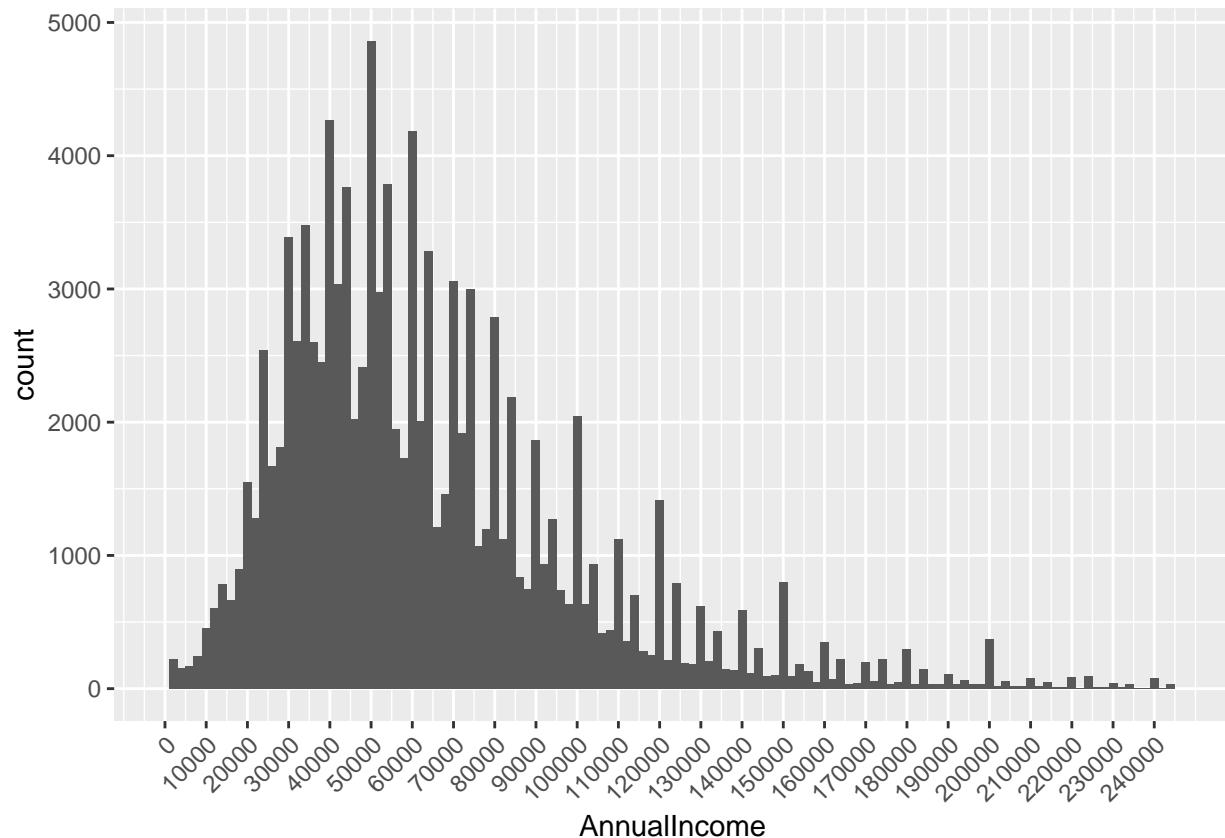
Sounds like most people borrow to cover their debts.

Employment status



As we can see most of the borrowers are employed.

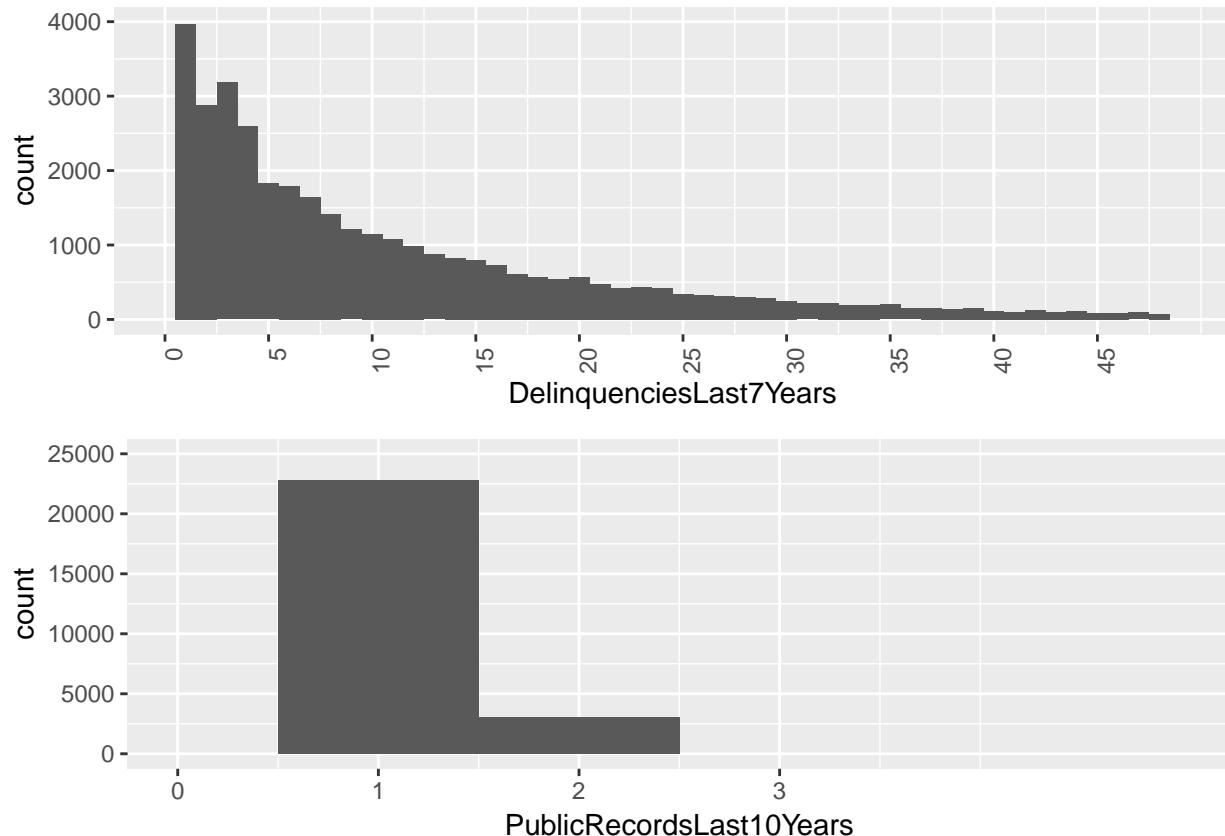
Income status



```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
##      0       38404     56000    67296    81900   21000035
```

The majority of income are between 20000 and 90000 annually. The median is 56000 and the mean is 67296.

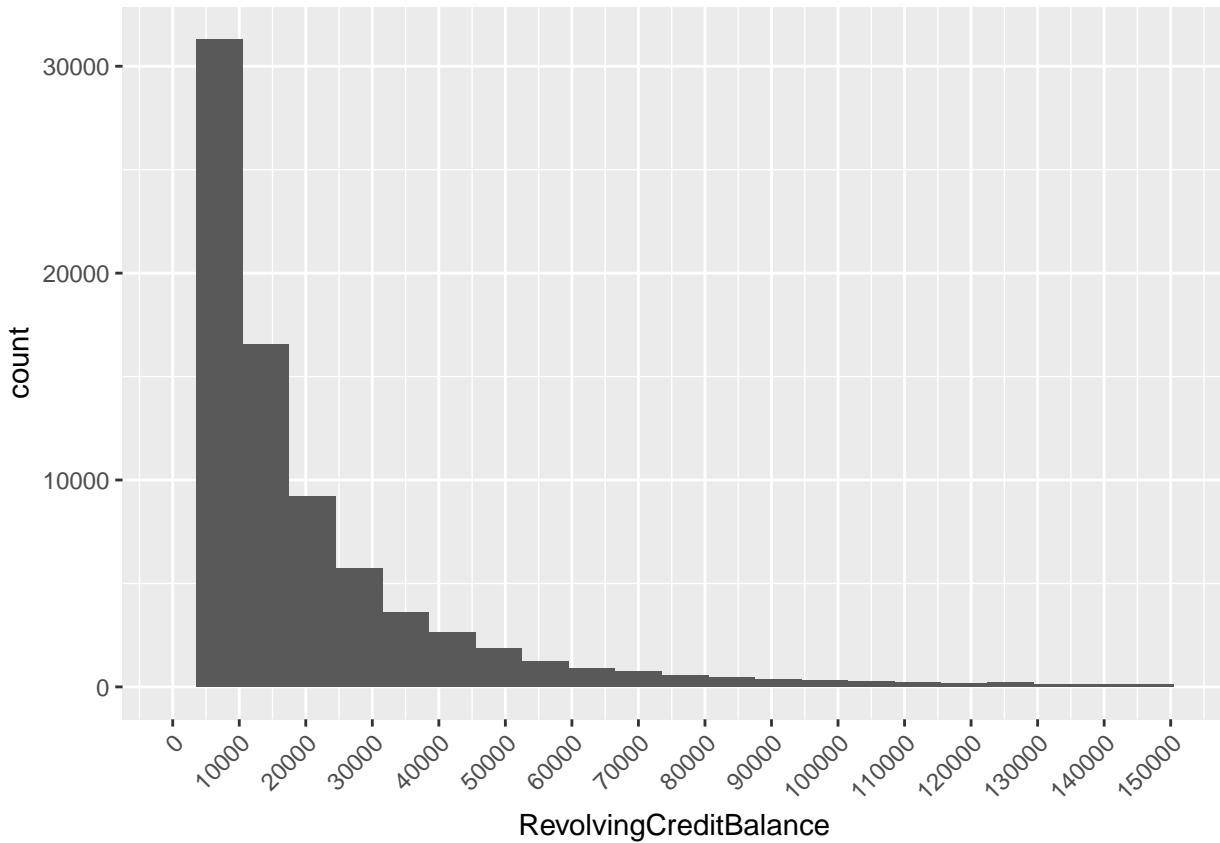
Payment history



```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
## 0.000 0.000 0.000 4.155 3.000 99.000    990
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
## 0.0000 0.0000 0.0000 0.3126 0.0000 38.0000    697
```

It is obvious that most of the borrower have zero or one delinquencies in the last 7 years. Similarly they have zero or one public records in the last 10 years.

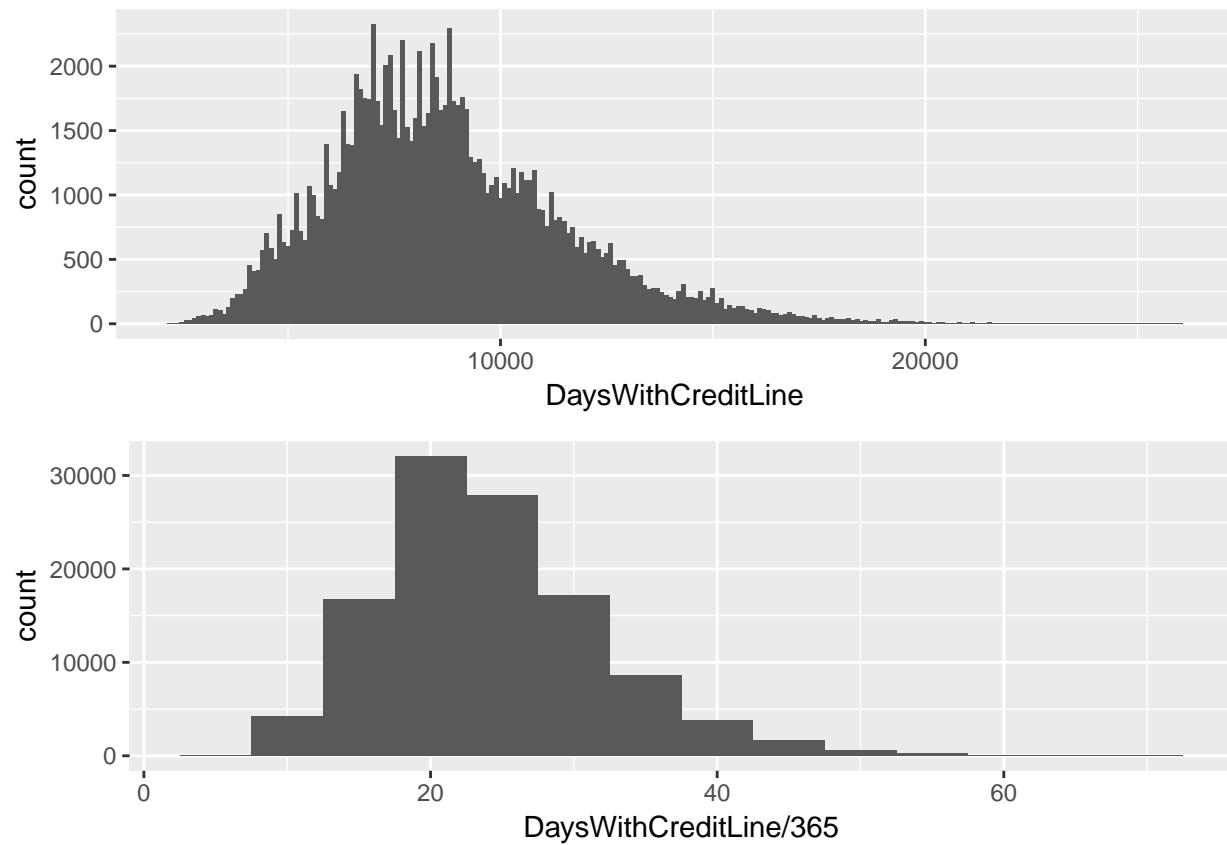
Revolving Credit Balance



```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max. NA's
##      0    3121  8549  17599  19521 1435667 7604
```

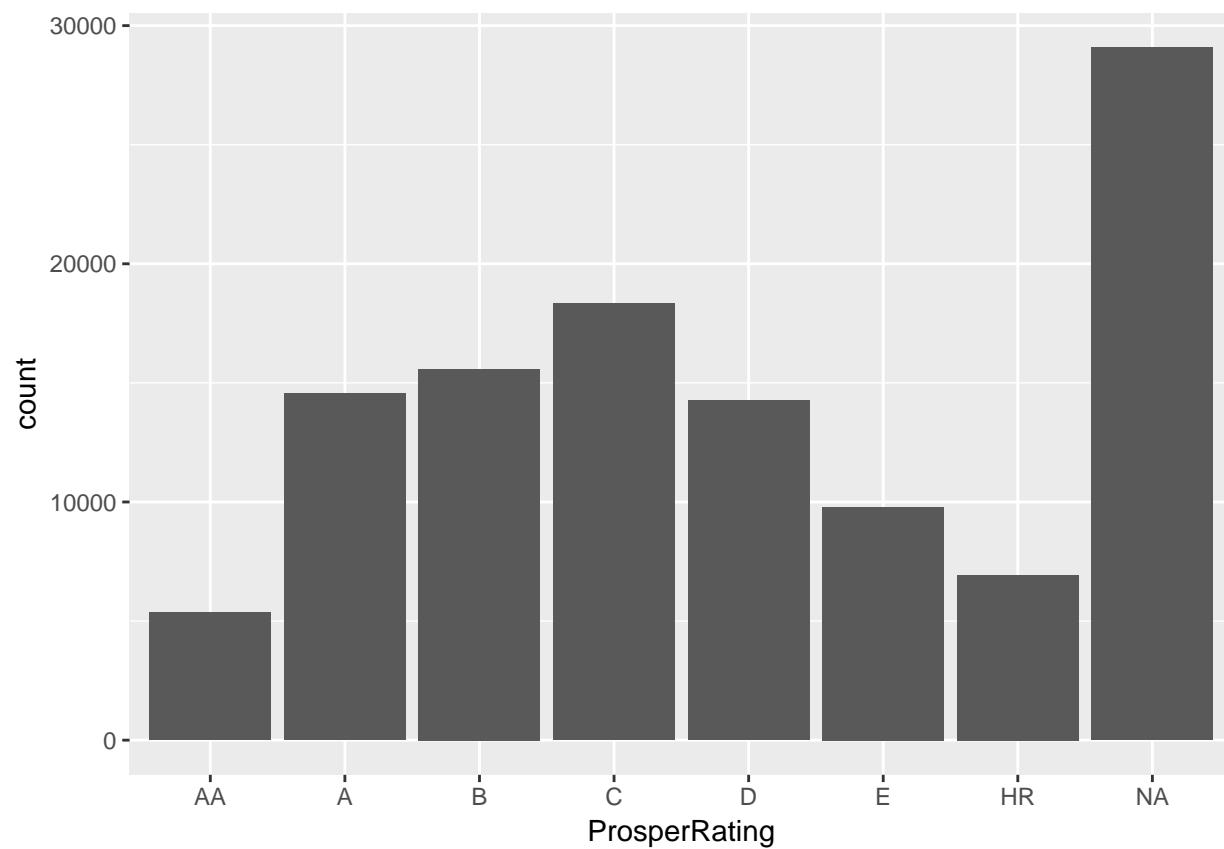
Revolving Credit Balance is the total outstanding balance that the borrower owes on his/her credit accounts. The median and mean are 8549 and 17600 respectively and the most common amount is 0.

Length of credit history



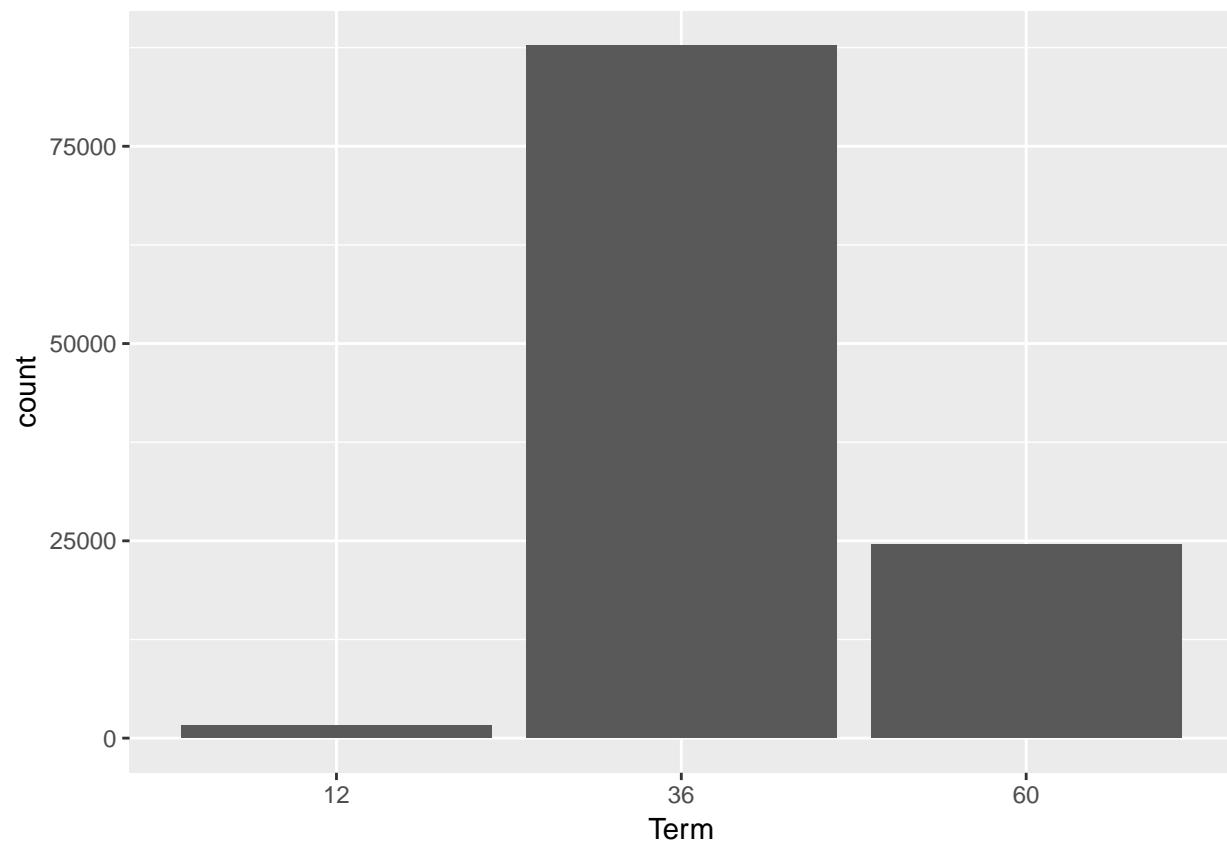
here is a credit line stores the credit story of 60 years.

Rating



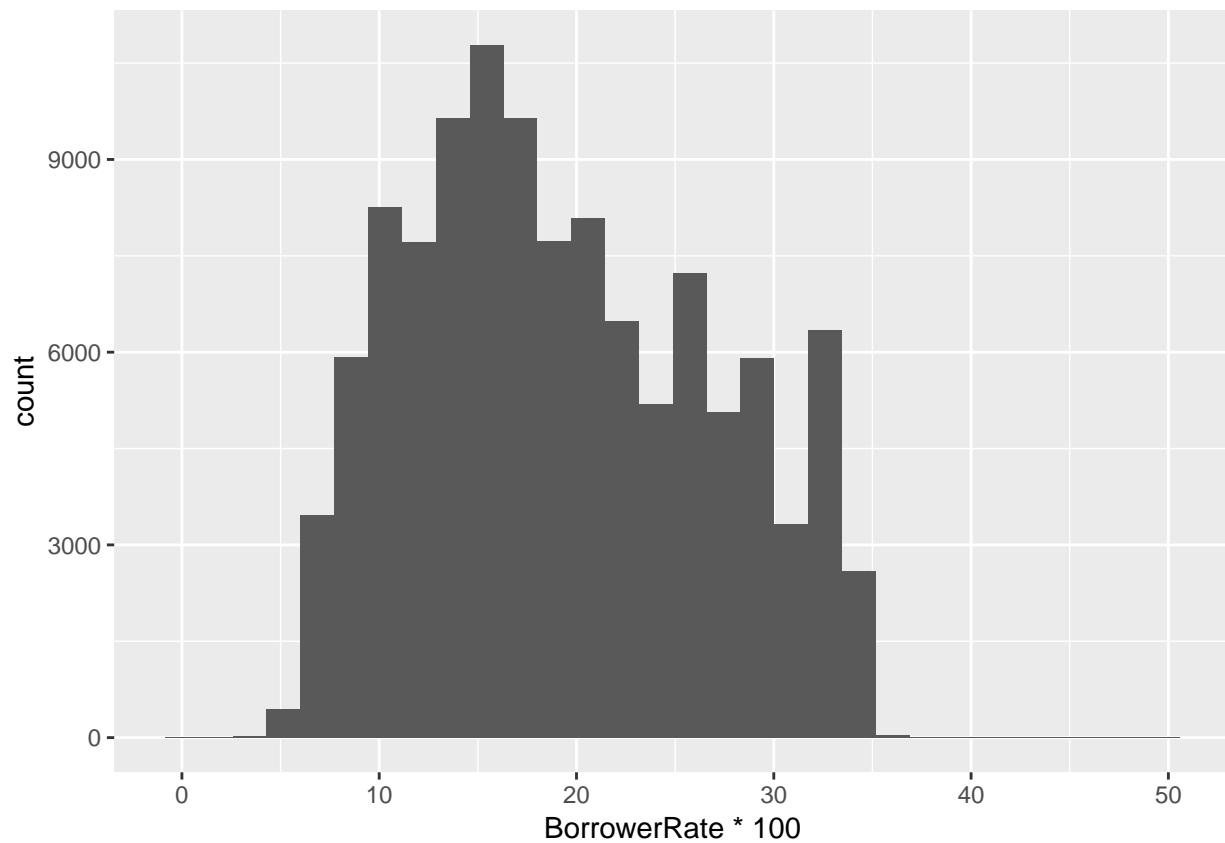
The most common rating is C follows by B. A and D are at the next steps (excluding the NA).

Loan length



Most loans have 36 months terms

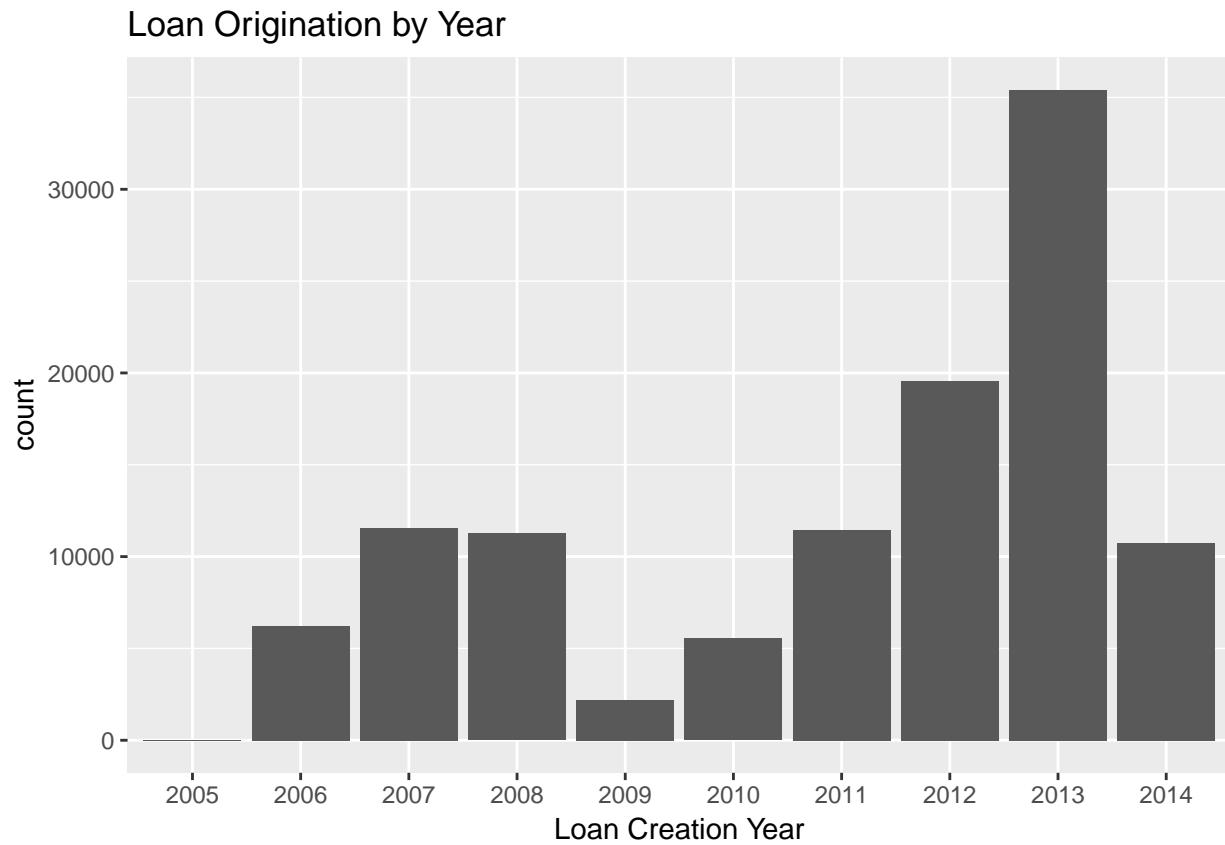
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.  
## 0.0000 0.1340 0.1840 0.1928 0.2500 0.4975
```

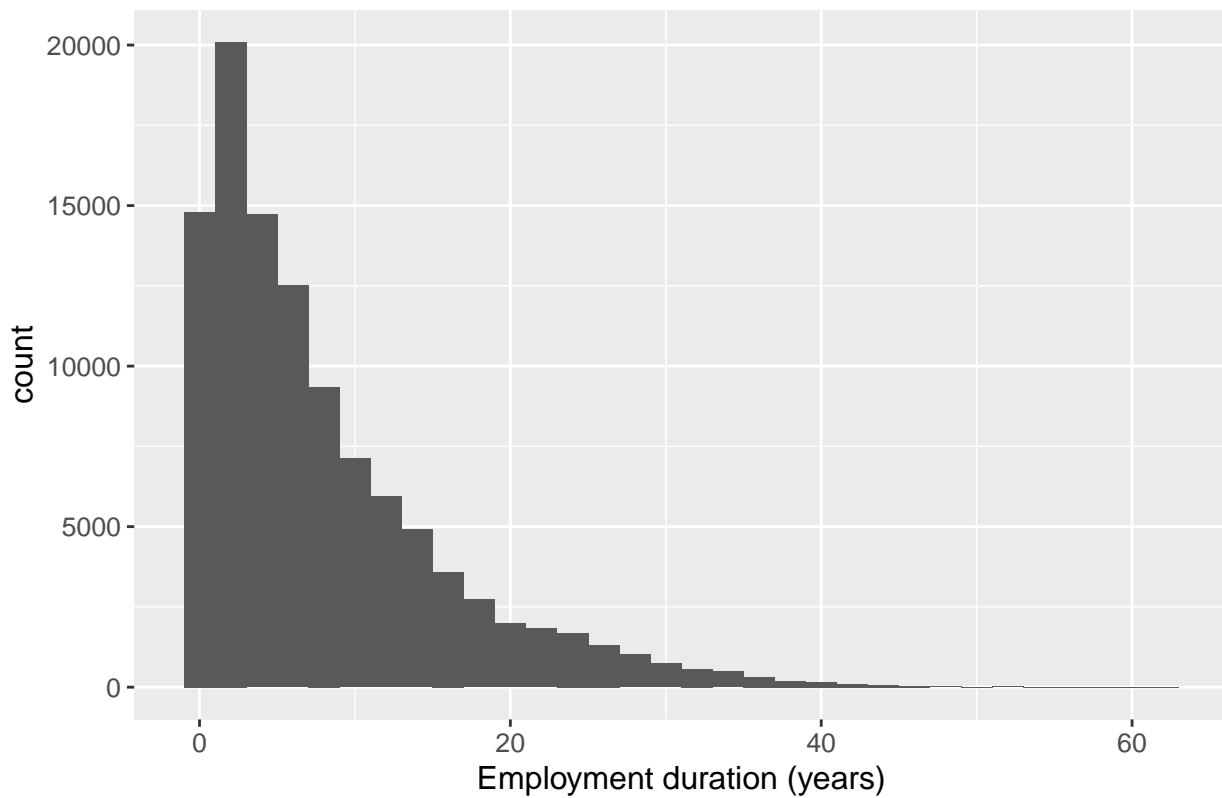
The median and mean for the borrower rate are 18.4% and 19.28% respectively, and The maximum borrower rate is 0.4975 or 49.75%.

Years borrowing

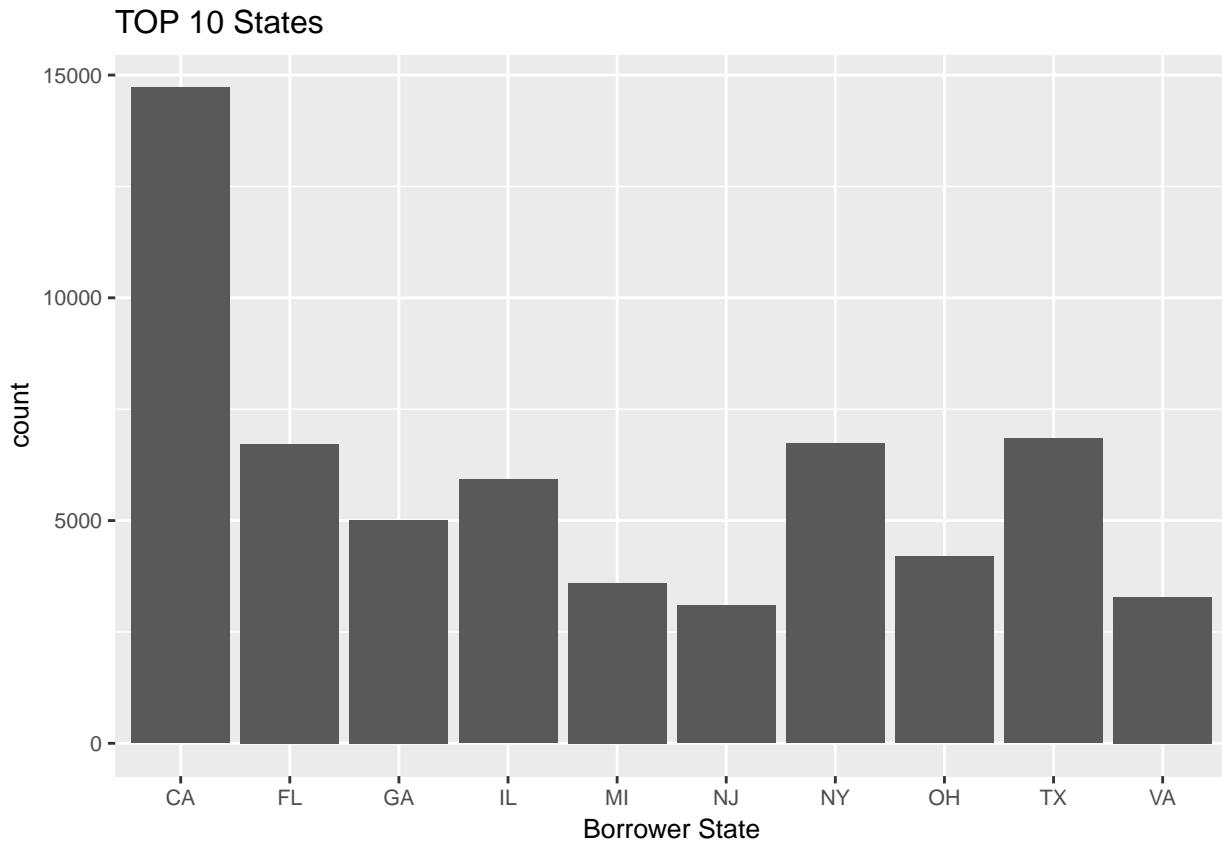


2013 is the year that people borrowed money more than any other years and 2009 is the minimum borrowing year. There can be many reasons behind that like economic crises or elections.

Employment Duration of the borrowers



With the increase of the length of employment there is a decrease in the number of people who borrow loans.



As we can see California is the state that people were more likely to loan and Florida, Illinois, New York and Texas are at the next steps.

Univariate Analysis

What is the structure of your dataset?

For the purpose of this project I am using the Prosper data set, which contains all Prosper loans created until March 11th, 2014. There are discrete and continuous variables in this dataset. Each variable is a column and each row is an observation.

What is/are the main feature(s) of interest in your dataset?

- DelinquenciesLast7Years
- PublicRecordsLast10Years
- DebtToIncomeRatio

- RevolvingCreditBalance
- DaysWithCreditLine
- LoanOriginalAmount
- ListingCategory

- EmploymentStatus
- AnnualIncome
- BorrowerRate
- Term
- ProsperRating
- Listing Creation Date

What other features in the dataset do you think will help support your investigation into your feature(s) of interest?

Other variables that help me in my investigation are Employment Duration, Debt To Income Ratio, Prosper Rating and Occupation.

Did you create any new variables from existing variables in the dataset?

I created following new variables during the analysis: ListingCretionYear and Days with credit line.

Of the features you investigated, were there any unusual distributions?

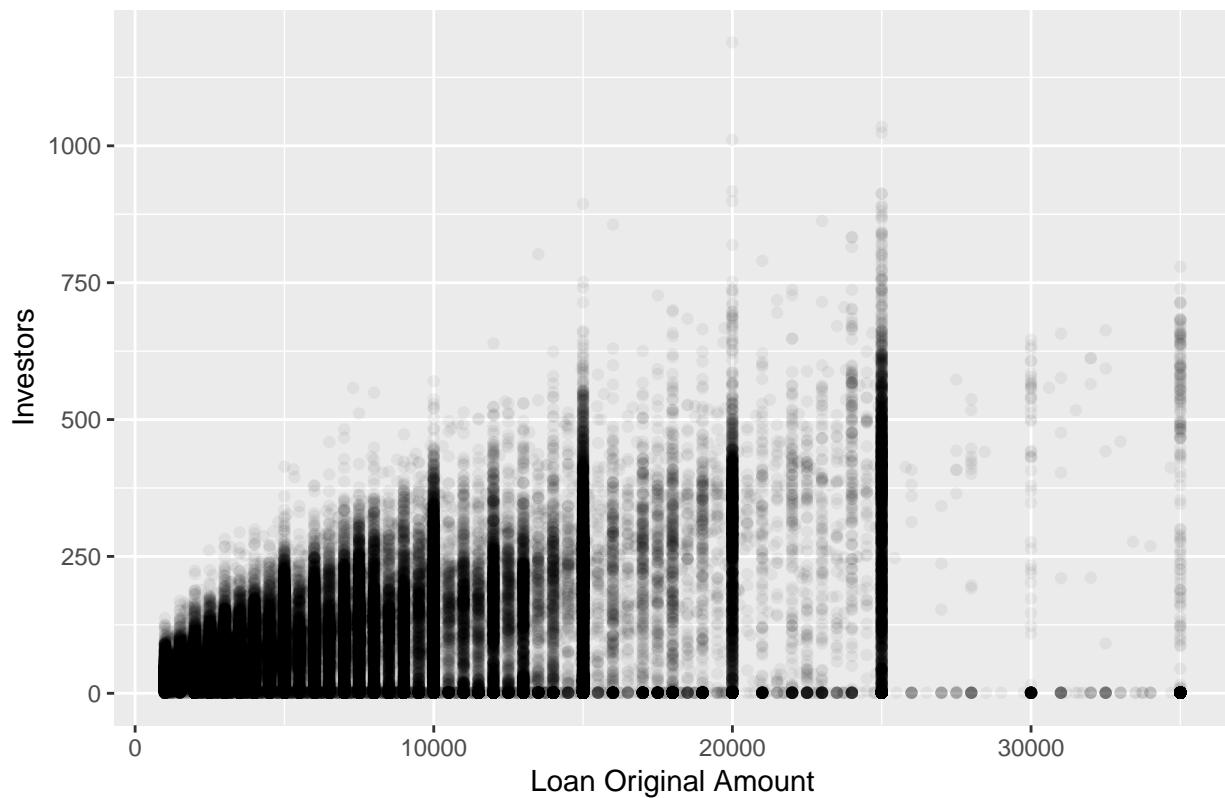
Did you perform any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

I set ListingCreationYear variable as a factor so when I plotted it would look discrete. I've alose taked care of ranked variables order in top 10 loan states.

Bivariate Plots Section

The relationship between Loan Original Amount and Number of Investors

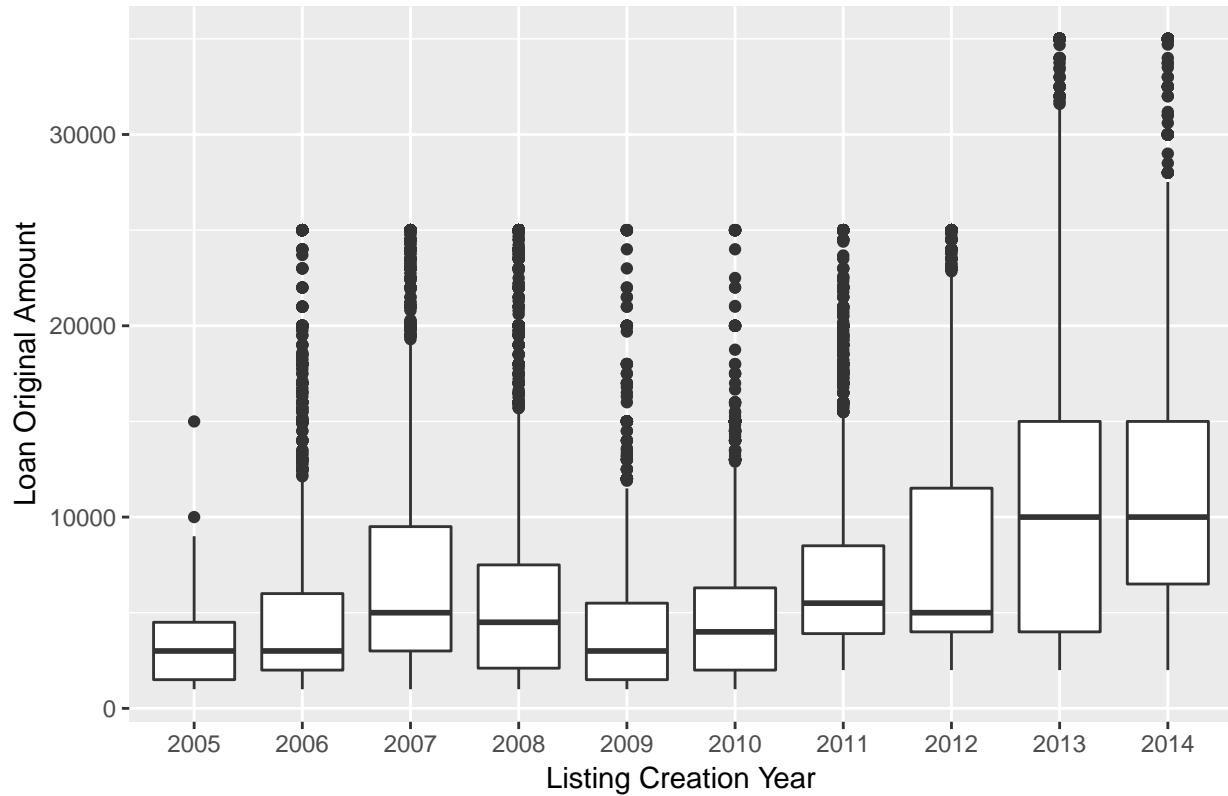
Loan Original Amount vs Number of Investors



As we can see larger loans have more investors.

Borrowed loan amount vary throughout the years

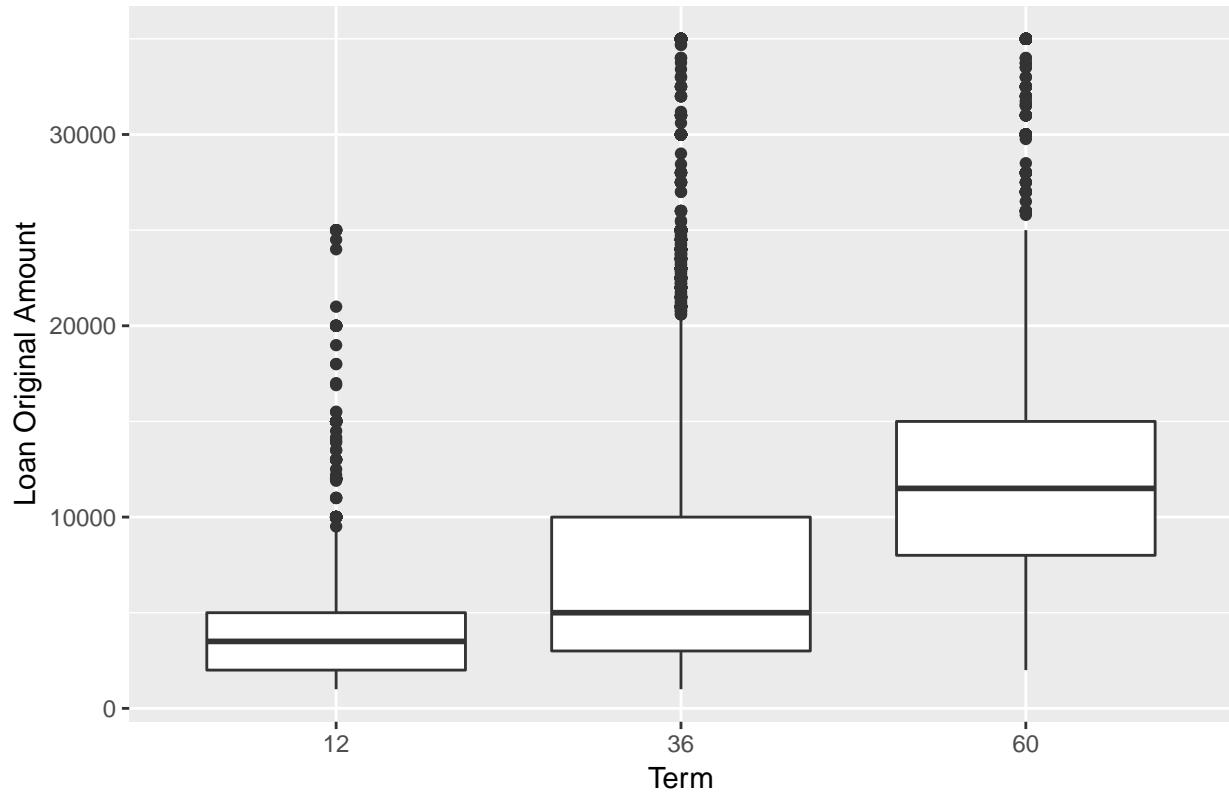
Listing Creation Year vs Loan Original Amount



Mean loan amounts went up slowly from 2005 to 2007, then it decreased at 2008 and went down to its minimum value at 2009. After that it recovered and increased and peaked at 2013 and 2014.

Relationship between loan amount and terms.

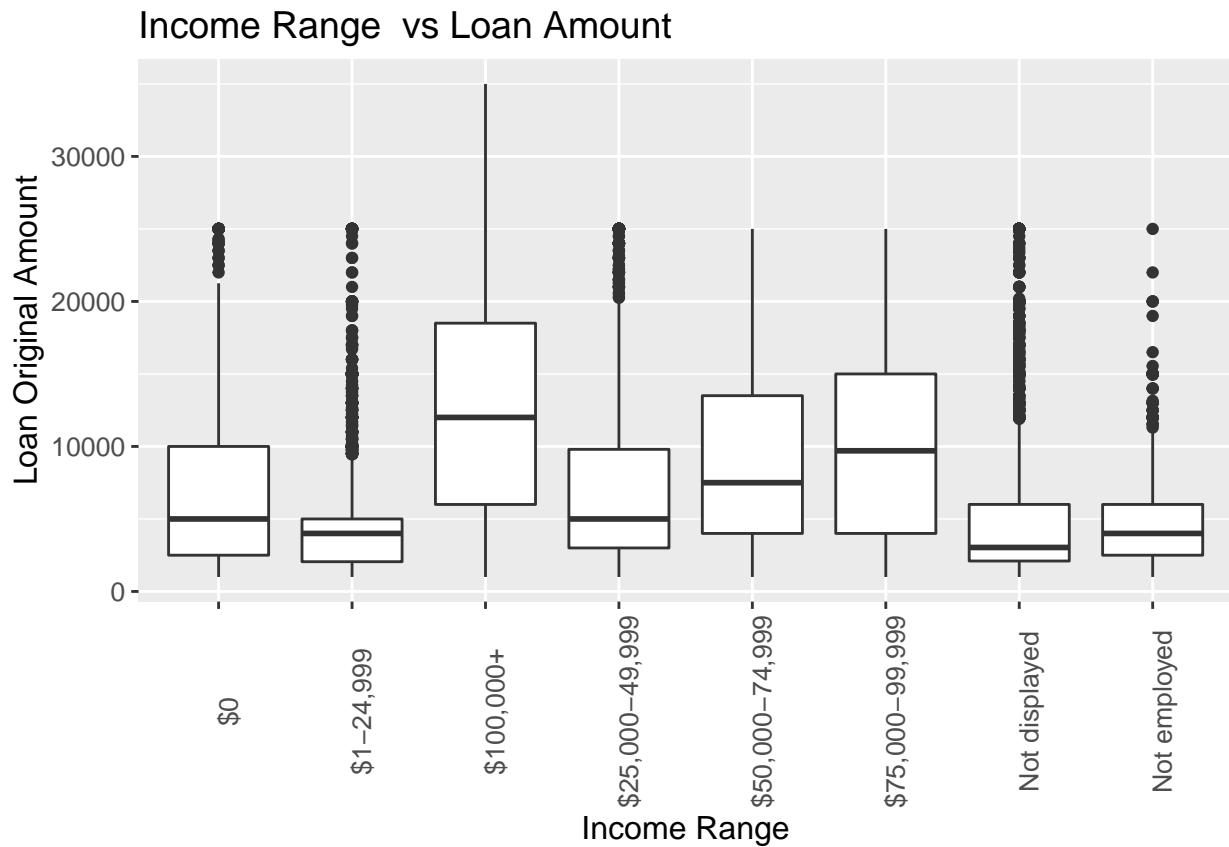
Term vs Loan Original Amount



```
## df$Term: 12
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##     1000    2000   3500     4694    5000    25000
## -----
## df$Term: 36
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##     1000    3000   5000     7276    10000   35000
## -----
## df$Term: 60
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##     2000    8000  11500    12370   15000   35000
```

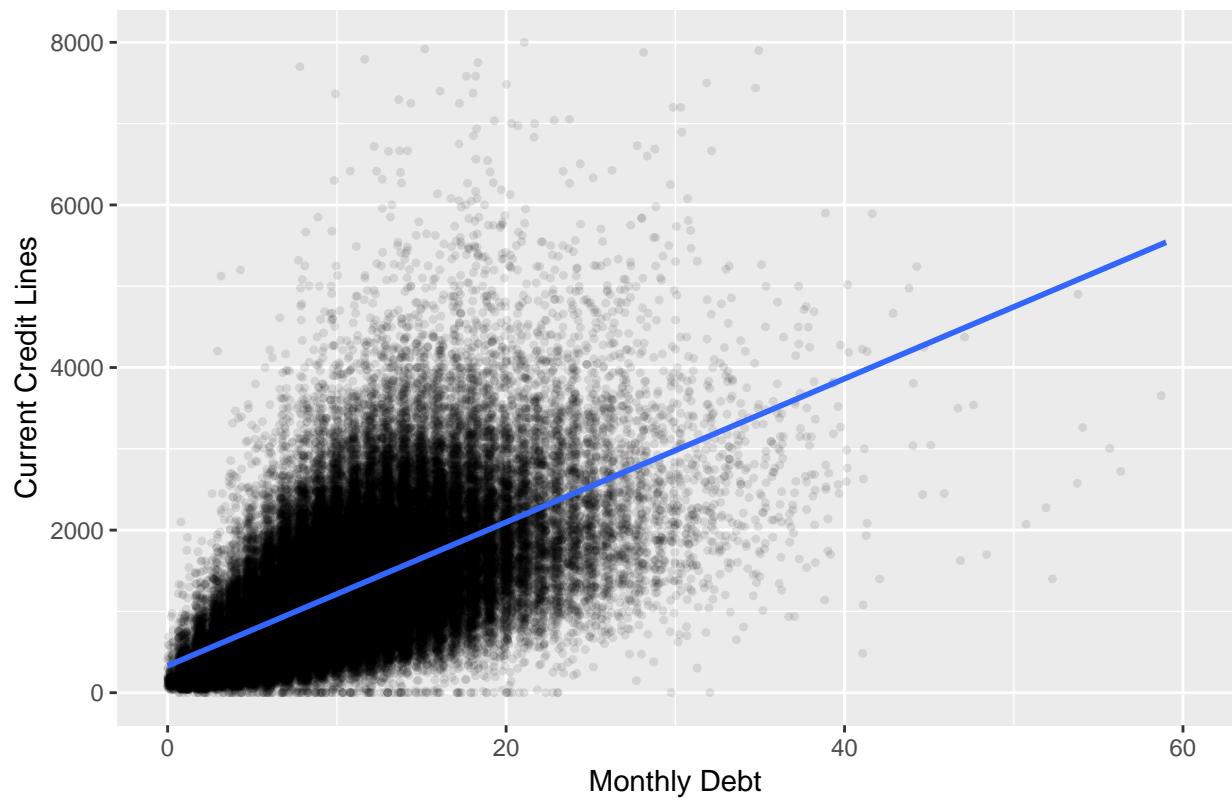
The loan amounts are getting larger with longer terms. As we can see from the Table median and mean are increasing by longer terms.

Relationships between loan amount and the IncomeRange



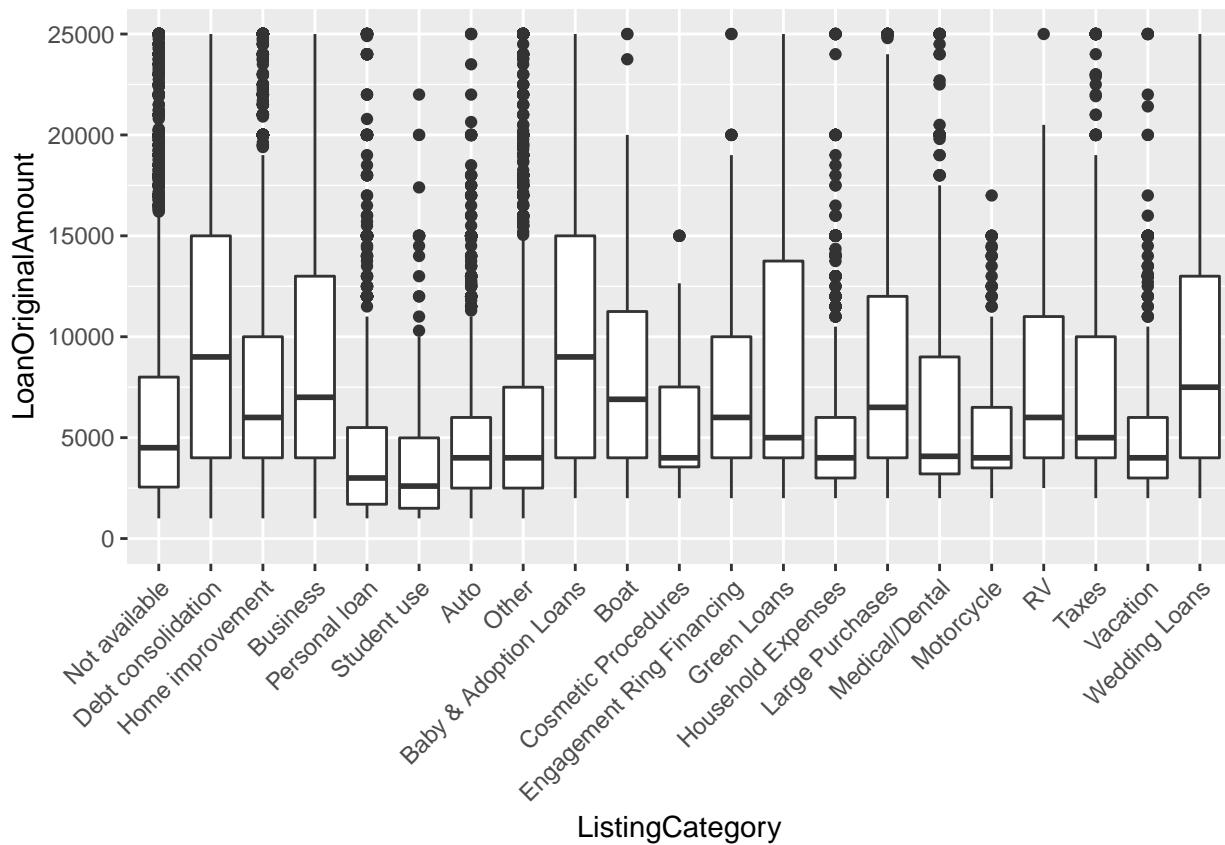
It is obvious from the box plots that larger loans on average are related to larger incomes.

Monthly Debt vs Current Credit Lines



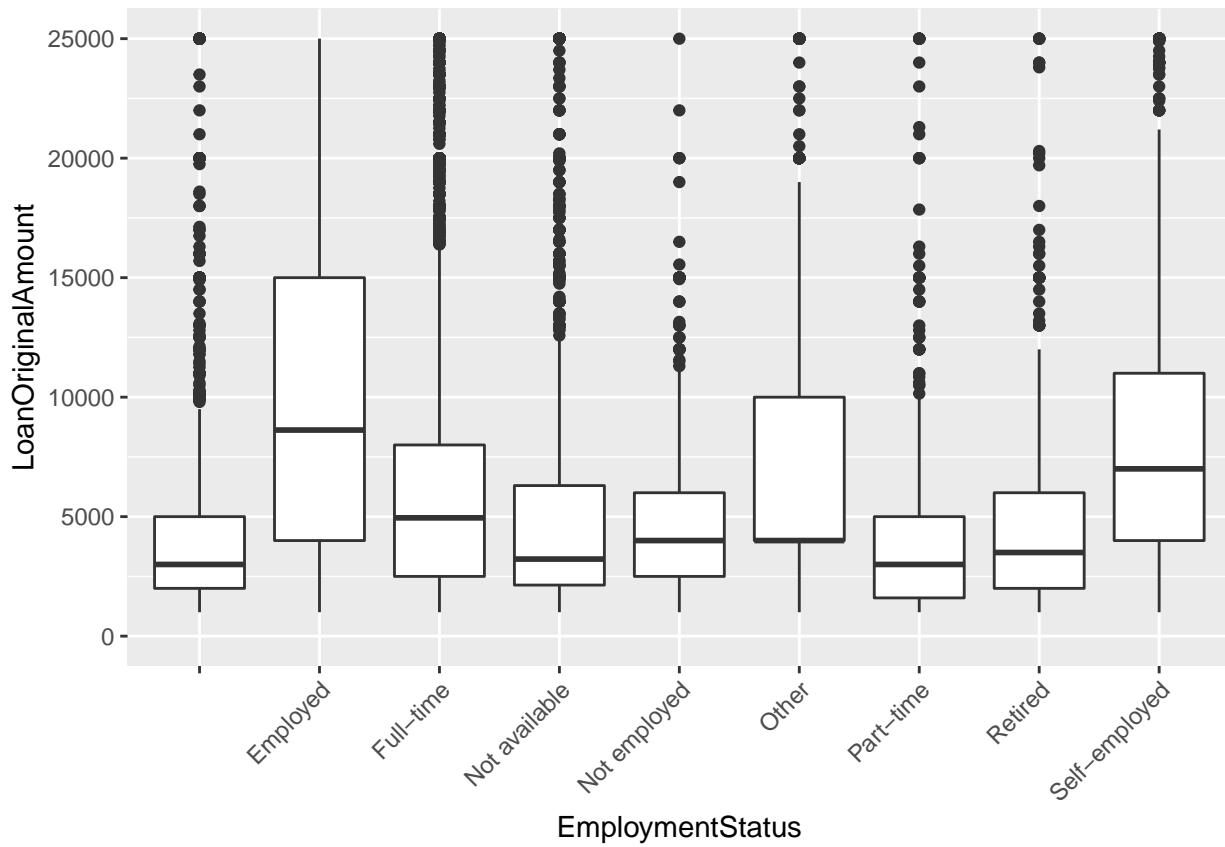
As we can see the amount of debt is growing by increasing the number of credit lines.

Relationship between LoanOriginalAmount with ListingCategory.

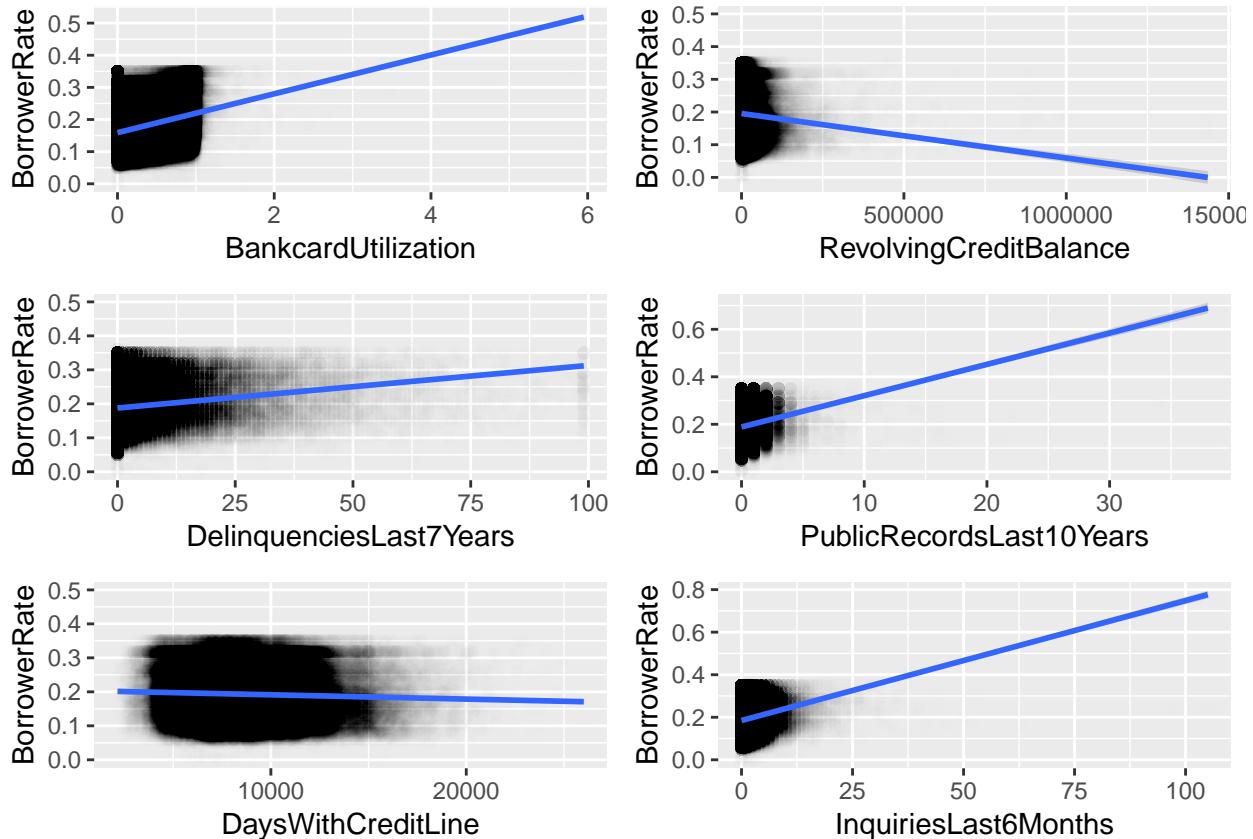


If we look at the mean values we can see that the mean of Baby & Adoption and Dept Consolidation are the highest as opposed to Student and Vacation that are the minimums.

Employment status and loan amount



As we can see Employed people requested more loans. Interestingly, Not employed people wanted more loan than part-time employed people.



As BankcardUtilization, DelinquenciesLast7Years, PublicRecordsLast10Years and InquiriesLast6Months increases with the borrower rate, as opposed to RevolvingCreditBalance, which decreased. DaysWithCreditLine has no significant change with BorrowerRate.

Correlation between BorrowerRate with BankcardUtilization:

```
## [1] 0.255482
```

Correlation between BorrowerRate with RevolvingCreditBalance:

```
## [1] -0.05960823
```

Correlation between BorrowerRate with DelinquenciesLast7Years:

```
## [1] 0.1702787
```

Correlation between BorrowerRate with PublicRecordsLast10Years:

```
## [1] 0.1283138
```

Correlation between BorrowerRate with DaysWithCreditLine:

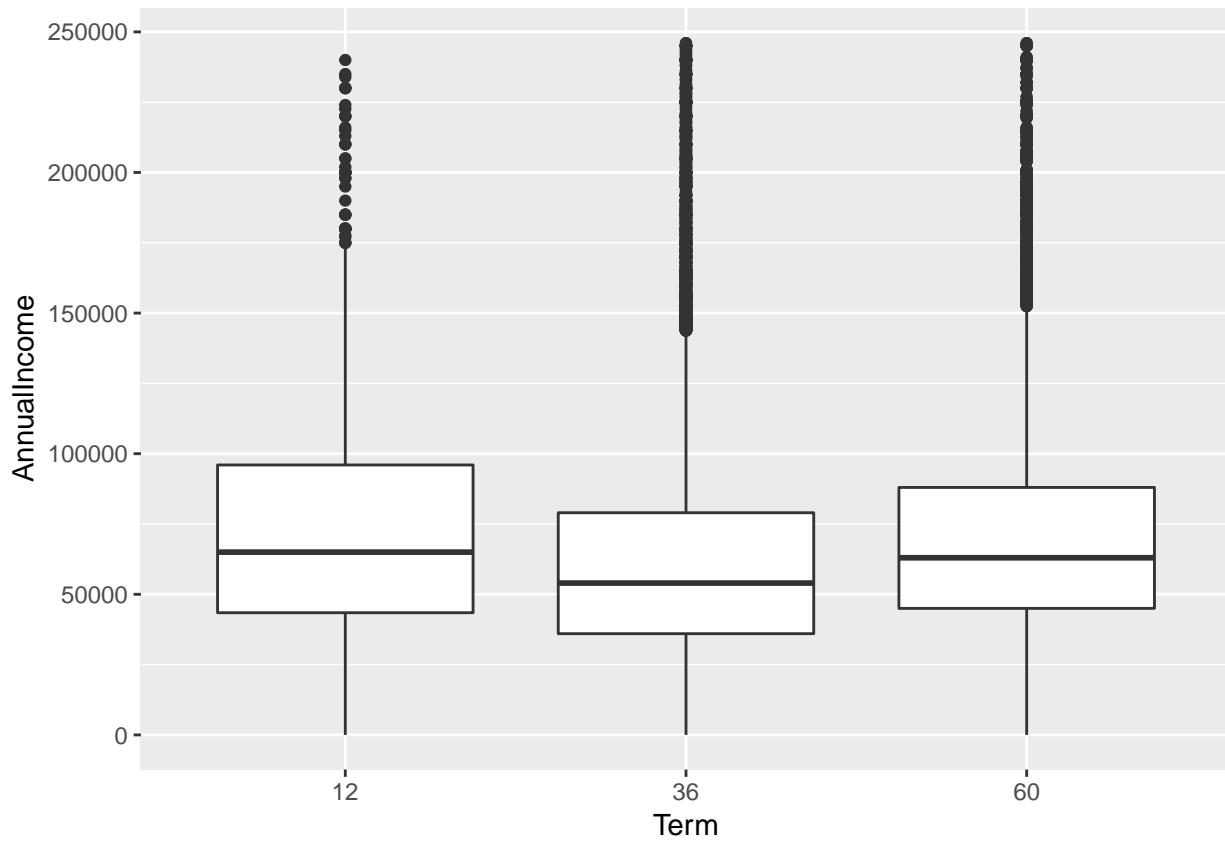
```
## [1] -0.0474466
```

Correlation between BorrowerRate with InquiriesLast6Months:

```
## [1] 0.18381
```

None of that variebles has strong relatipnship with each other. The strongest relationship is between BorrowerRate with BankcardUtilization with 0.25.

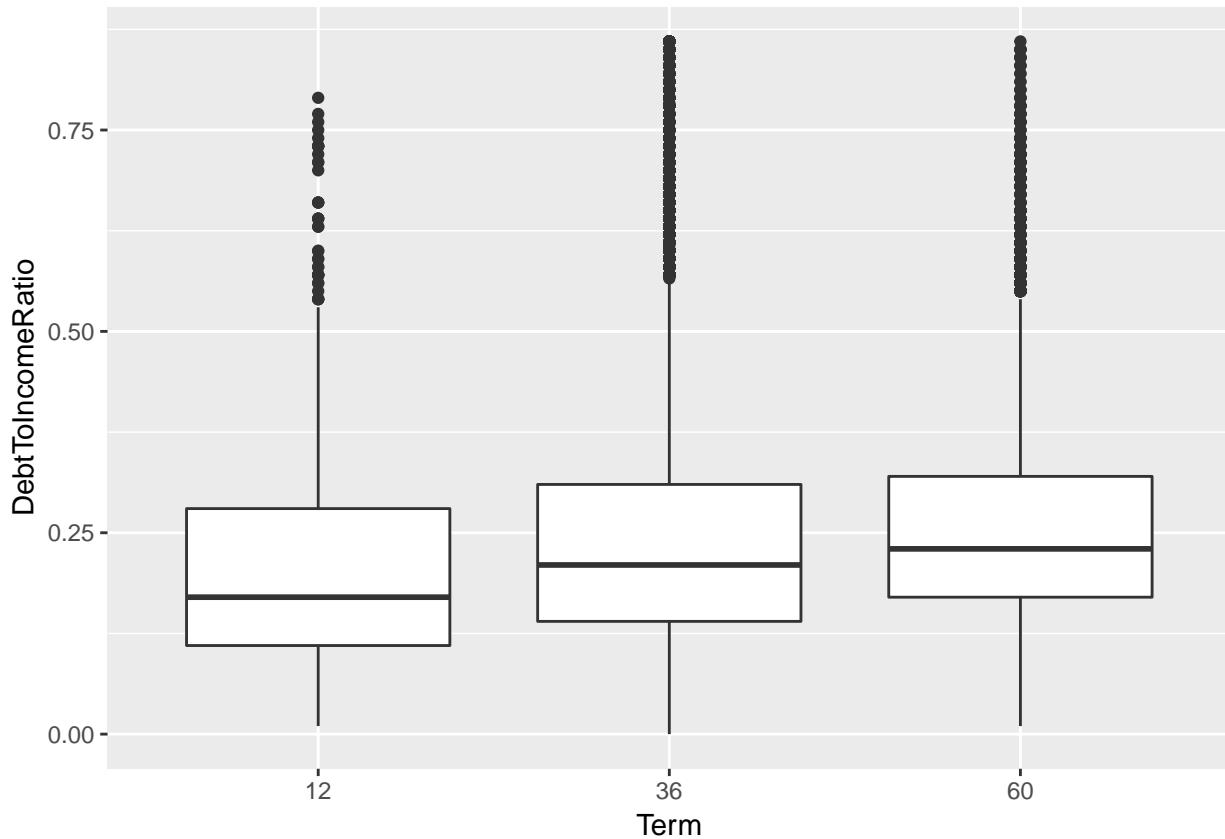
Relation of Term and AnnualIncome



```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##      0    44200   67000    82658   97925  7422574
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##      0    36000   54000    65289   80000 21000035
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##      0    45000   64000    73465   90000 1305000
```

The median and mean for 12 months term are the highest between term of 12,36 and 60.

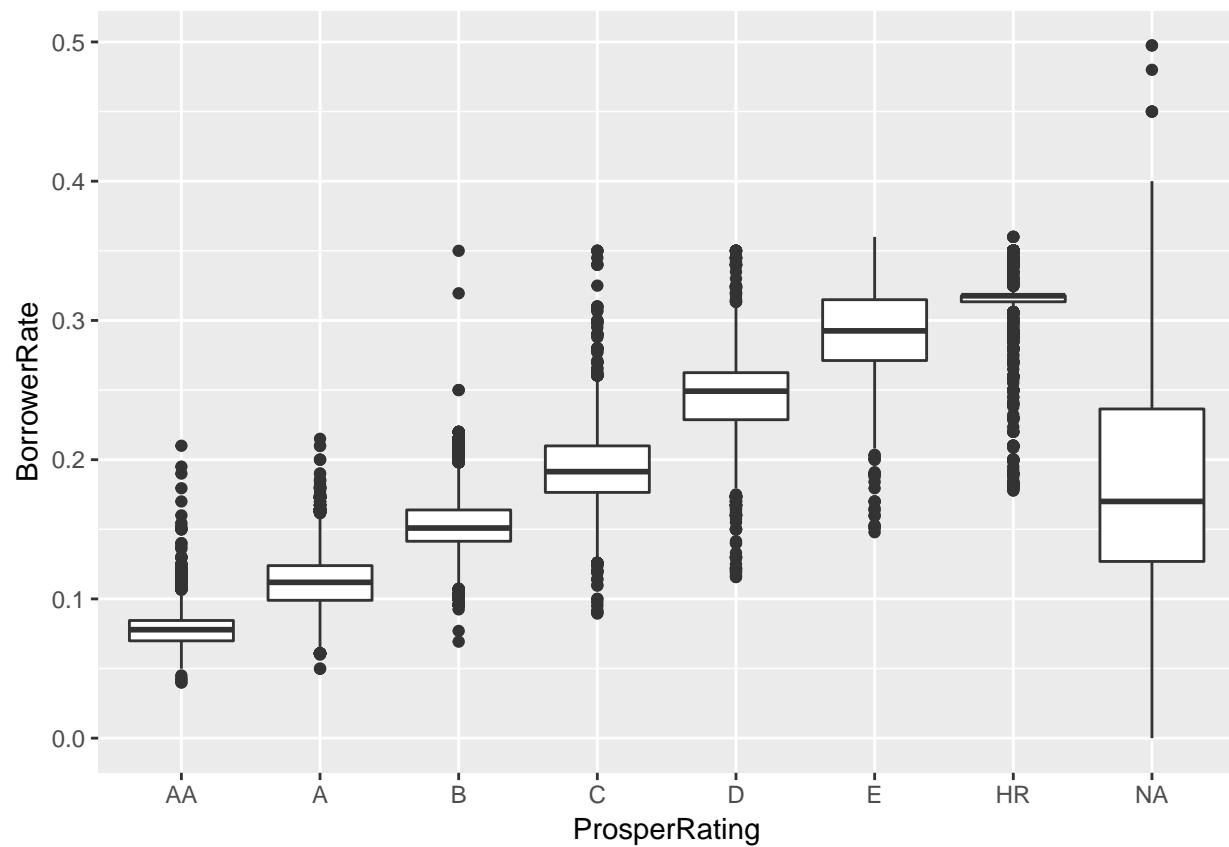
Relation between Term and DeptToIncomeRatio



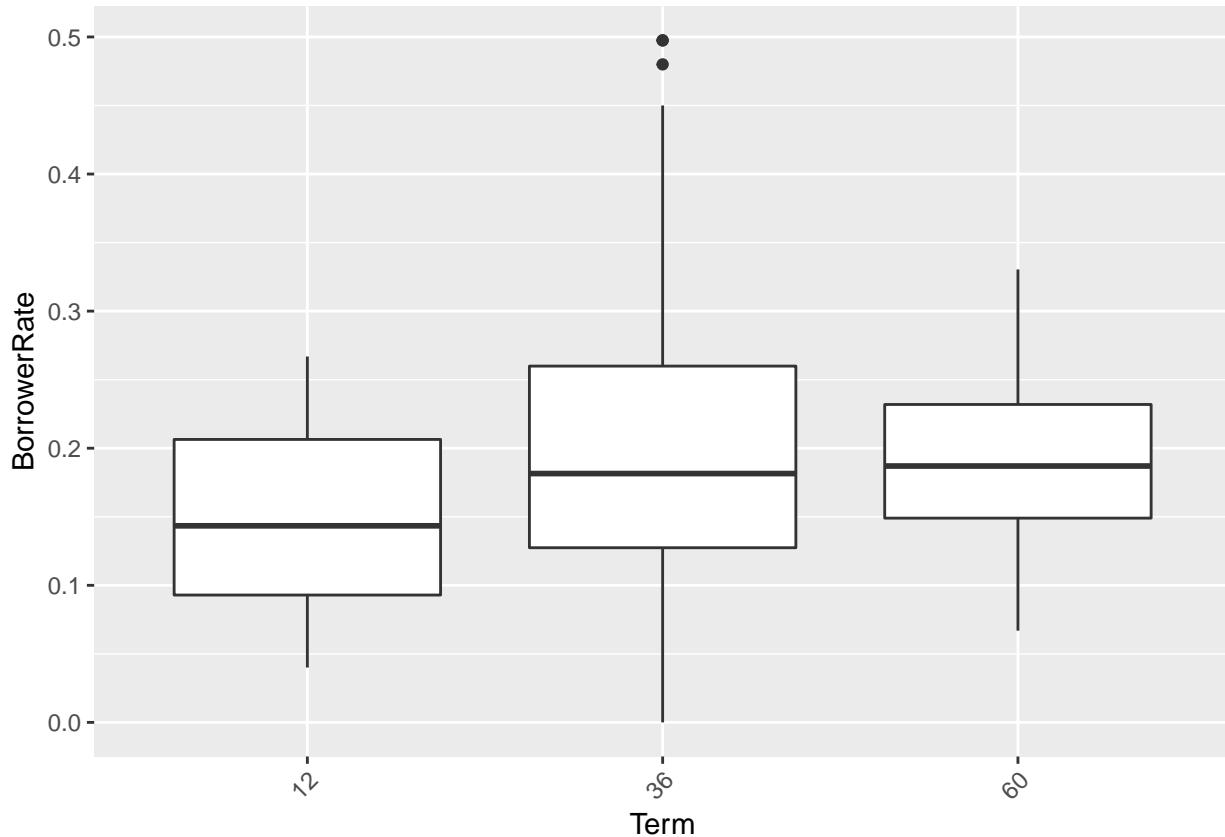
```
##      Min. 1st Qu. Median     Mean 3rd Qu.   Max. NA's
## 0.0100 0.1100 0.1700 0.2203 0.2800 10.0100    199
##      Min. 1st Qu. Median     Mean 3rd Qu.   Max. NA's
## 0.0000 0.1400 0.2100 0.2830 0.3100 10.0100 6953
##      Min. 1st Qu. Median     Mean 3rd Qu.   Max. NA's
## 0.0100 0.1700 0.2300 0.2565 0.3200 10.0100 1402
```

As we can see the median of the DebtToIncomeRatio increases as the terms goes up.

ProsperRating against BorrowerRate



As we can see The better rating belongs to lower borrower rate.



The median of 60 month term is the highest one, while the 12 month has the minimum one.

Bivariate Analysis

Tip: As before, summarize what you found in your bivariate explorations here. Use the questions below to guide your discussion.

Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

The number of investors is increasing with higher prosper score, loan amount is bigger, borrowers have less existing prosper loans, estimated loss is lower. The mean loan amount vary trough years. The minimum mean in 2009 and the maximum one is in 2013 and 2014. We also can see for instance the borrower rate increases as debt to income ratio increases. Moreover, we can see that with bigger amonth of loan the term is also longer.

Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

I noticed that employed people are more likely to loan than others. Also the debt to income ratio for rating AA is the lowest one.

What was the strongest relationship you found?

The strongest relationship that I found was between BorrowerRate with BankcardUtilization with value of 0.255482. # Multivariate Plots Section

Tip: Now it's time to put everything together. Based on what you found in the bivariate plots section, create a few multivariate plots to investigate more complex interactions between variables. Make sure that the plots that you create here are justified by the plots you explored in the previous section. If you plan on creating any mathematical models, this is the section where you will do that.

Multivariate Analysis

Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

Were there any interesting or surprising interactions between features?

OPTIONAL: Did you create any models with your dataset? Discuss the strengths and limitations of your model.

Final Plots and Summary

Tip: You've done a lot of exploration and have built up an understanding of the structure of and relationships between the variables in your dataset. Here, you will select three plots from all of your previous exploration to present here as a summary of some of your most interesting findings. Make sure that you have refined your selected plots for good titling, axis labels (with units), and good aesthetic choices (e.g. color, transparency). After each plot, make sure you justify why you chose each plot by describing what it shows.

Plot One

Description One

Plot Two

Description Two

Plot Three

Description Three

Reflection

Tip: Here's the final step! Reflect on the exploration you performed and the insights you found. What were some of the struggles that you went through? What went well? What was surprising? Make sure you include an insight into future work that could be done with the dataset.

Tip: Don't forget to remove this, and the other **Tip** sections before saving your final work and knitting the final report!