

# Loan Data Analysis by Jack Bae

## Introduction

---

**Prosper**(www.prosper.com) is a peer-to-peer marketplace lending platform. Here is an explanation of Prosper from their website: > Prosper allows people to invest in each other in a way that is financially and socially rewarding. On Prosper, borrowers list loan requests between \$2,000 and \$35,000 and individual investors invest as little as \$25 in each loan listing they select. Prosper handles the servicing of the loan on behalf of the matched borrowers and investors.

And the process of how peer-to-peer lending at Prosper works: > - Borrowers choose a loan amount, purpose and post a loan listing. - Investors review loan listings and invest in listings that meet their criteria. - Once the process is complete, borrowers make fixed monthly payments and investors receive a portion of those payments directly to their Prosper account.

The Prosper Loan data analyzed here contains 113,937 records last updated on 03/11/2014.

## Exploring the Dataset

The dataset contains 113937 records with 81 variables.

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

The names of the 81 variables:

```
## [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"

```

Overview of dataset variables:

```
##   ListingKey      ListingNumber      ListingCreationDate
## Length:113937      Min. :     4      Length:113937
## Class :character  1st Qu.: 400919  Class :character
## Mode  :character Median : 600554  Mode  :character
##                           Mean  : 627886
##                           3rd Qu.: 892634
##                           Max.  :1255725
##
##   CreditGrade        Term        LoanStatus        ClosedDate
## Length:113937      Min. :12.00    Length:113937    Length:113937
## Class :character  1st Qu.:36.00    Class :character  Class :character
## Mode  :character Median :36.00    Mode  :character  Mode  :character
##                           Mean  :40.83
##                           3rd Qu.:36.00
##                           Max.  :60.00
##
##   BorrowerAPR      BorrowerRate      LenderYield
## Min.  :0.00653    Min.  :0.0000    Min.  :-0.0100
## 1st Qu.:0.15629   1st Qu.:0.1340   1st Qu.: 0.1242
## Median :0.20976   Median :0.1840   Median : 0.1730
## Mean   :0.21883   Mean   :0.1928   Mean   : 0.1827
## 3rd Qu.:0.28381   3rd Qu.:0.2500   3rd Qu.: 0.2400
## Max.   :0.51229   Max.   :0.4975   Max.   : 0.4925
## NA's   :25
##
##   EstimatedEffectiveYield EstimatedLoss      EstimatedReturn
## Min.  :-0.183       Min.  :0.005      Min.  :-0.183
## 1st Qu.: 0.116       1st Qu.:0.042      1st Qu.: 0.074
## Median : 0.162       Median :0.072      Median : 0.092
## Mean   : 0.169       Mean   :0.080      Mean   : 0.096
## 3rd Qu.: 0.224       3rd Qu.:0.112      3rd Qu.: 0.117
## Max.   : 0.320       Max.   :0.366      Max.   : 0.284
## NA's   :29084        NA's   :29084      NA's   :29084
##
##   ProsperRating..numeric. ProsperRating..Alpha. ProsperScore
## Min.  :1.000          Length:113937      Min.  : 1.00
## 1st Qu.:3.000          Class :character  1st Qu.: 4.00
## Median :4.000          Mode  :character  Median : 6.00
## Mean   :4.072          Mean   :5.95
## 3rd Qu.:5.000          3rd Qu.:8.00
## Max.   :7.000          Max.   :11.00
## NA's   :29084          NA's   :29084
##
##   ListingCategory..numeric. BorrowerState          Occupation
## Min.  : 0.000          Length:113937      Length:113937
## 1st Qu.: 1.000          Class :character  Class :character
## Median : 1.000          Mode  :character  Mode  :character
## Mean   : 2.774
## 3rd Qu.: 3.000
## Max.   :20.000
##
##   EmploymentStatus EmploymentStatusDuration IsBorrowerHomeowner
## Length:113937      Min.  : 0.00      Length:113937
## Class :character  1st Qu.: 26.00      Class :character
## Mode  :character Median : 67.00      Mode  :character
```

```

##               Mean   : 96.07
##      3rd Qu.:137.00
##      Max.   :755.00
##      NA's    :7625
## CurrentlyInGroup     GroupKey      DateCreditPulled
## Length:113937       Length:113937    Length:113937
## Class :character    Class :character  Class :character
## Mode  :character    Mode  :character  Mode  :character
##
##
##
##
## CreditScoreRangeLower CreditScoreRangeUpper FirstRecordedCreditLine
## Min.   : 0.0          Min.   : 19.0        Length:113937
## 1st Qu.:660.0         1st Qu.:679.0      Class :character
## Median :680.0         Median :699.0      Mode   :character
## Mean   :685.6         Mean   :704.6
## 3rd Qu.:720.0         3rd Qu.:739.0
## Max.   :880.0         Max.   :899.0
## NA's   :591           NA's   :591
## CurrentCreditLines  OpenCreditLines TotalCreditLinespast7years
## Min.   : 0.00          Min.   : 0.00      Min.   : 2.00
## 1st Qu.: 7.00          1st Qu.: 6.00      1st Qu.: 17.00
## Median :10.00          Median : 9.00      Median : 25.00
## Mean   :10.32          Mean   : 9.26      Mean   : 26.75
## 3rd Qu.:13.00          3rd Qu.:12.00      3rd Qu.: 35.00
## Max.   :59.00          Max.   :54.00      Max.   :136.00
## NA's   :7604           NA's   :7604      NA's   :697
## OpenRevolvingAccounts OpenRevolvingMonthlyPayment InquiriesLast6Months
## Min.   : 0.00          Min.   : 0.0        Min.   : 0.000
## 1st Qu.: 4.00          1st Qu.: 114.0     1st Qu.: 0.000
## Median : 6.00          Median : 271.0     Median : 1.000
## Mean   : 6.97          Mean   : 398.3     Mean   : 1.435
## 3rd Qu.: 9.00          3rd Qu.: 525.0     3rd Qu.: 2.000
## Max.   :51.00          Max.   :14985.0    Max.   :105.000
## NA's   :697            NA's   :697      NA's   :697
## TotalInquiries   CurrentDelinquencies AmountDelinquent
## Min.   : 0.000          Min.   : 0.0000    Min.   : 0.0
## 1st Qu.: 2.000          1st Qu.: 0.0000    1st Qu.: 0.0
## Median : 4.000          Median : 0.0000    Median : 0.0
## Mean   : 5.584          Mean   : 0.5921    Mean   : 984.5
## 3rd Qu.: 7.000          3rd Qu.: 0.0000    3rd Qu.: 0.0
## Max.   :379.000         Max.   :83.0000    Max.   :463881.0
## NA's   :1159            NA's   :697      NA's   :7622
## DelinquenciesLast7Years PublicRecordsLast10Years
## Min.   : 0.000          Min.   : 0.0000
## 1st Qu.: 0.000          1st Qu.: 0.0000
## Median : 0.000          Median : 0.0000
## Mean   : 4.155          Mean   : 0.3126
## 3rd Qu.: 3.000          3rd Qu.: 0.0000
## Max.   :99.000           Max.   :38.0000
## NA's   :990              NA's   :697
## PublicRecordsLast12Months RevolvingCreditBalance BankcardUtilization
## Min.   : 0.000          Min.   : 0          Min.   :0.000

```

```

## 1st Qu.: 0.000      1st Qu.: 3121      1st Qu.:0.310
## Median : 0.000      Median : 8549      Median :0.600
## Mean   : 0.015      Mean   : 17599     Mean   :0.561
## 3rd Qu.: 0.000      3rd Qu.: 19521     3rd Qu.:0.840
## Max.   :20.000      Max.   :1435667    Max.   :5.950
## NA's   :7604        NA's   :7604      NA's   :7604
## AvailableBankcardCredit TotalTrades
## Min.   : 0           Min.   : 0.00
## 1st Qu.: 880         1st Qu.: 15.00
## Median : 4100        Median : 22.00
## Mean   : 11210       Mean   : 23.23
## 3rd Qu.: 13180       3rd Qu.: 30.00
## Max.   :646285       Max.   :126.00
## NA's   :7544         NA's   :7544
## TradesNeverDelinquent..percentage. TradesOpenedLast6Months
## Min.   :0.000          Min.   : 0.000
## 1st Qu.:0.820          1st Qu.: 0.000
## Median :0.940          Median : 0.000
## Mean   :0.886          Mean   : 0.802
## 3rd Qu.:1.000          3rd Qu.: 1.000
## Max.   :1.000          Max.   :20.000
## NA's   :7544           NA's   :7544
## DebtToIncomeRatio IncomeRange      IncomeVerifiable
## Min.   : 0.000          Length:113937    Length:113937
## 1st Qu.: 0.140          Class :character  Class :character
## Median : 0.220          Mode  :character  Mode  :character
## Mean   : 0.276
## 3rd Qu.: 0.320
## Max.   :10.010
## NA's   :8554
## StatedMonthlyIncome  LoanKey        TotalProsperLoans
## Min.   : 0             Length:113937    Min.   :0.00
## 1st Qu.: 3200          Class :character  1st Qu.:1.00
## Median : 4667          Mode  :character  Median :1.00
## Mean   : 5608          Mean   : 1.42
## 3rd Qu.: 6825          3rd Qu.:2.00
## Max.   :1750003         Max.   :8.00
## NA's   :91852          NA's   :91852
## TotalProsperPaymentsBilled OnTimeProsperPayments
## Min.   : 0.00          Min.   : 0.00
## 1st Qu.: 9.00          1st Qu.: 9.00
## Median : 16.00          Median : 15.00
## Mean   : 22.93          Mean   : 22.27
## 3rd Qu.: 33.00          3rd Qu.: 32.00
## Max.   :141.00          Max.   :141.00
## NA's   :91852           NA's   :91852
## ProsperPaymentsLessThanOneMonthLate ProsperPaymentsOneMonthPlusLate
## Min.   : 0.00          Min.   : 0.00
## 1st Qu.: 0.00          1st Qu.: 0.00
## Median : 0.00          Median : 0.00
## Mean   : 0.61          Mean   : 0.05
## 3rd Qu.: 0.00          3rd Qu.: 0.00
## Max.   :42.00          Max.   :21.00
## NA's   :91852           NA's   :91852

```

```

## ProsperPrincipalBorrowed ProsperPrincipalOutstanding
## Min. : 0 Min. : 0
## 1st Qu.: 3500 1st Qu.: 0
## Median : 6000 Median : 1627
## Mean : 8472 Mean : 2930
## 3rd Qu.:11000 3rd Qu.: 4127
## Max. :72499 Max. :23451
## NA's :91852 NA's :91852
## ScorexChangeAtTimeOfListing LoanCurrentDaysDelinquent
## Min. :-209.00 Min. : 0.0
## 1st Qu.: -35.00 1st Qu.: 0.0
## Median : -3.00 Median : 0.0
## Mean : -3.22 Mean : 152.8
## 3rd Qu.: 25.00 3rd Qu.: 0.0
## Max. : 286.00 Max. :2704.0
## NA's :95009
## LoanFirstDefaultedCycleNumber LoanMonthsSinceOrigination LoanNumber
## Min. : 0.00 Min. : 0.0 Min. : 1
## 1st Qu.: 9.00 1st Qu.: 6.0 1st Qu.: 37332
## Median :14.00 Median : 21.0 Median : 68599
## Mean :16.27 Mean : 31.9 Mean : 69444
## 3rd Qu.:22.00 3rd Qu.: 65.0 3rd Qu.:101901
## Max. :44.00 Max. :100.0 Max. :136486
## NA's :96985
## LoanOriginalAmount LoanOriginationDate LoanOriginationQuarter
## Min. : 1000 Length:113937 Length:113937
## 1st Qu.: 4000 Class :character Class :character
## Median : 6500 Mode :character Mode :character
## Mean : 8337
## 3rd Qu.:12000
## Max. :35000
##
## MemberKey MonthlyLoanPayment LP_CustomerPayments
## Length:113937 Min. : 0.0 Min. : -2.35
## Class :character 1st Qu.: 131.6 1st Qu.: 1005.76
## Mode :character Median : 217.7 Median : 2583.83
## Mean : 272.5 Mean : 4183.08
## 3rd Qu.: 371.6 3rd Qu.: 5548.40
## Max. :2251.5 Max. :40702.39
##
## LP_CustomerPrincipalPayments LP_InterestandFees LP_ServiceFees
## Min. : 0.0 Min. : -2.35 Min. : -664.87
## 1st Qu.: 500.9 1st Qu.: 274.87 1st Qu.: -73.18
## Median : 1587.5 Median : 700.84 Median : -34.44
## Mean : 3105.5 Mean : 1077.54 Mean : -54.73
## 3rd Qu.: 4000.0 3rd Qu.: 1458.54 3rd Qu.: -13.92
## Max. :35000.0 Max. :15617.03 Max. : 32.06
##
## LP_CollectionFees LP_GrossPrincipalLoss LP_NetPrincipalLoss
## Min. :-9274.75 Min. : -94.2 Min. : -954.5
## 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0
## Median : 0.00 Median : 0.0 Median : 0.0
## Mean : -14.24 Mean : 700.4 Mean : 681.4
## 3rd Qu.: 0.00 3rd Qu.: 0.0 3rd Qu.: 0.0

```

```

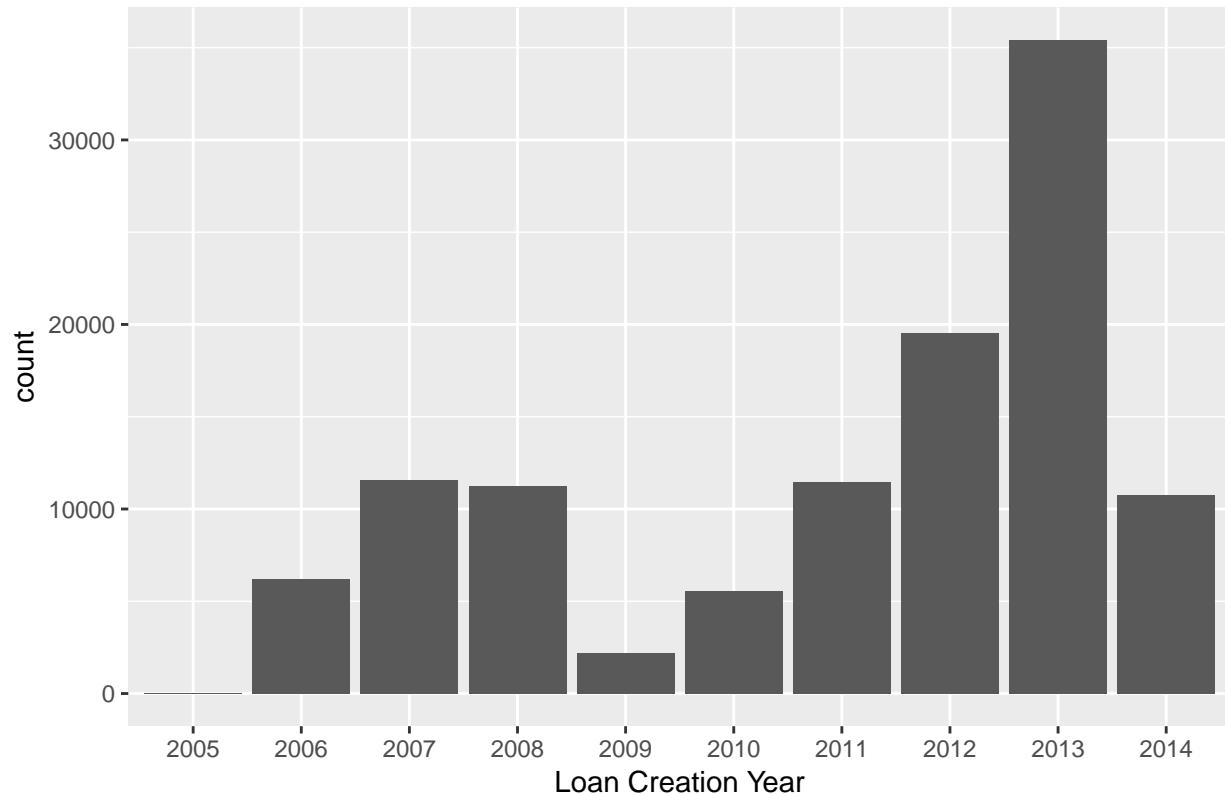
##  Max.    : 0.00   Max.    :25000.0      Max.    :25000.0
##
##  LP_NonPrincipalRecoverypayments PercentFunded     Recommendations
##  Min.    : 0.00       Min.    :0.7000      Min.    : 0.00000
##  1st Qu.: 0.00       1st Qu.:1.0000      1st Qu.: 0.00000
##  Median : 0.00       Median :1.0000      Median : 0.00000
##  Mean    : 25.14      Mean   :0.9986      Mean   : 0.04803
##  3rd Qu.: 0.00       3rd Qu.:1.0000      3rd Qu.: 0.00000
##  Max.    :21117.90     Max.   :1.0125      Max.   :39.00000
##
##  InvestmentFromFriendsCount InvestmentFromFriendsAmount   Investors
##  Min.    : 0.00000      Min.    : 0.00      Min.    : 1.00
##  1st Qu.: 0.00000      1st Qu.: 0.00      1st Qu.: 2.00
##  Median : 0.00000      Median : 0.00      Median : 44.00
##  Mean    : 0.02346      Mean   : 16.55      Mean   : 80.48
##  3rd Qu.: 0.00000      3rd Qu.: 0.00      3rd Qu.: 115.00
##  Max.    :33.00000      Max.   :25000.00      Max.   :1189.00
##

```

While exploring the dataset, I found the need to create new variables and data frames. - New variables: ListingCreationYear, ListingCategoryName, CreditScoreAverage - New data frames: loansFrom2009 (loan data after July 12, 2009), loansTo2009 (loan data before July 12, 2009)

## Univariate Plots Section

Number of Loans by Year



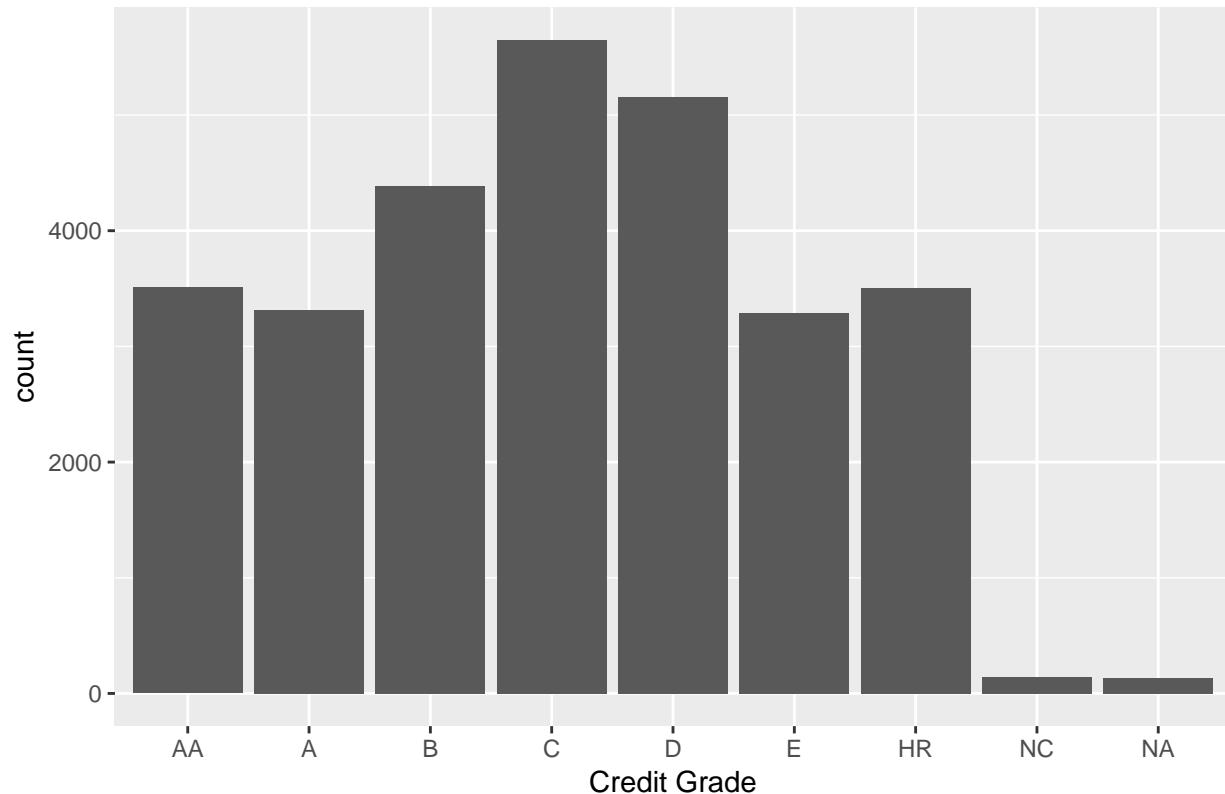
The number of loans by year graph shows that the number of loans increased from 2006 to 2009 but decreased

in 2009 due to the global financial crisis. Loans increased again starting in 2010 through 2013. The decrease in loans can be possibly explained by not having full year data for 2014.

```
## [1] "2005-11-09 20:44:28.847000000" "2014-03-10 12:20:53.760000000"
```

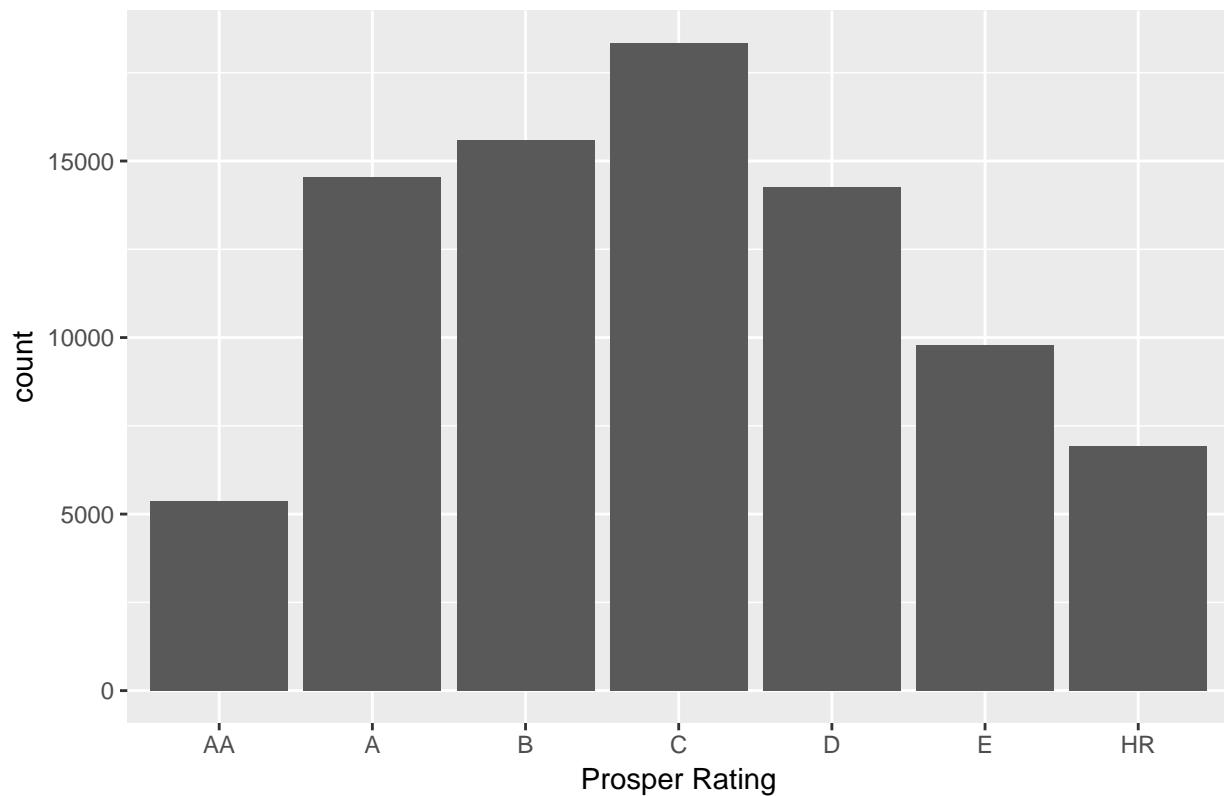
Checking the range of loan listing creation date, the data confirms that the most recent data for 2014 is March 10, confirming 2014 data does not contain full year data.

### Number of Loans by Credit Grade (Pre–2009)



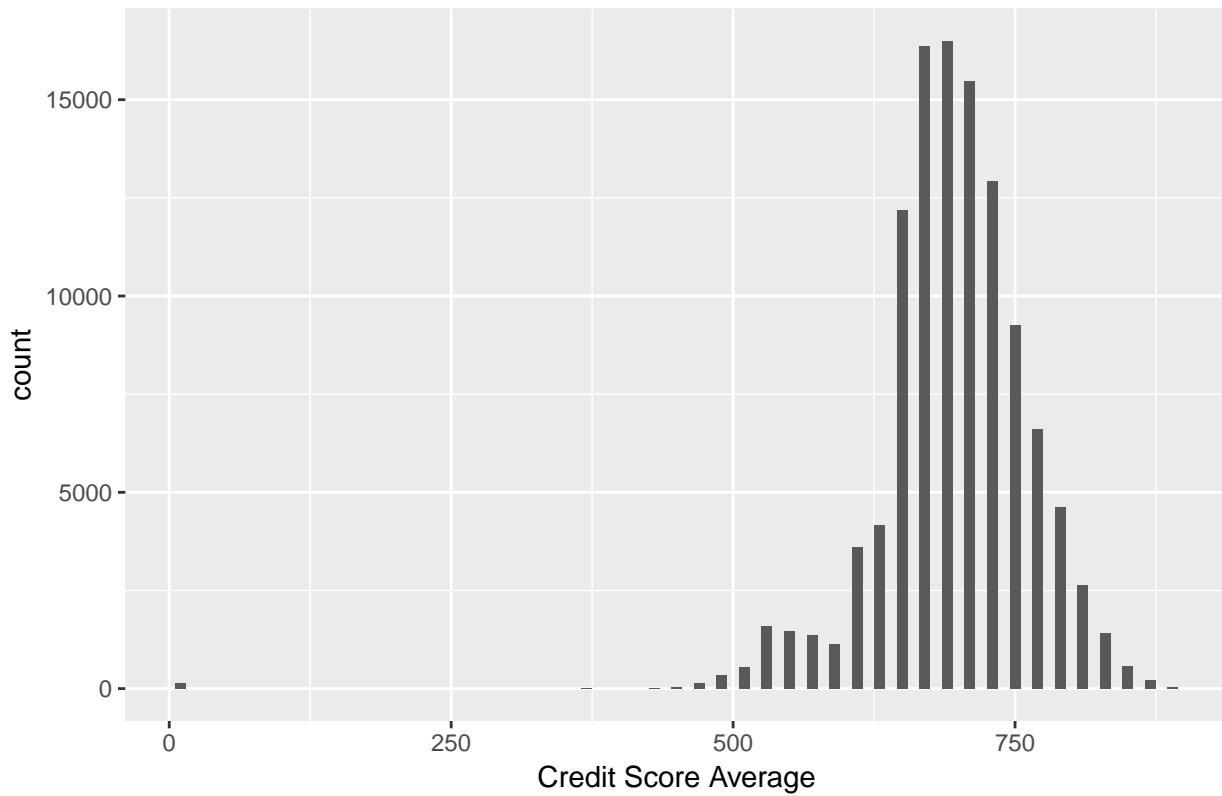
The credit grade data for data before July 12, 2009 shows a normal distribution of data. Prosper established a new rating system starting July 12, 2009.

### Number of Loans by Credit Grade (Post–2009)



The credit grade for post-2009 also shows a normal distribution.

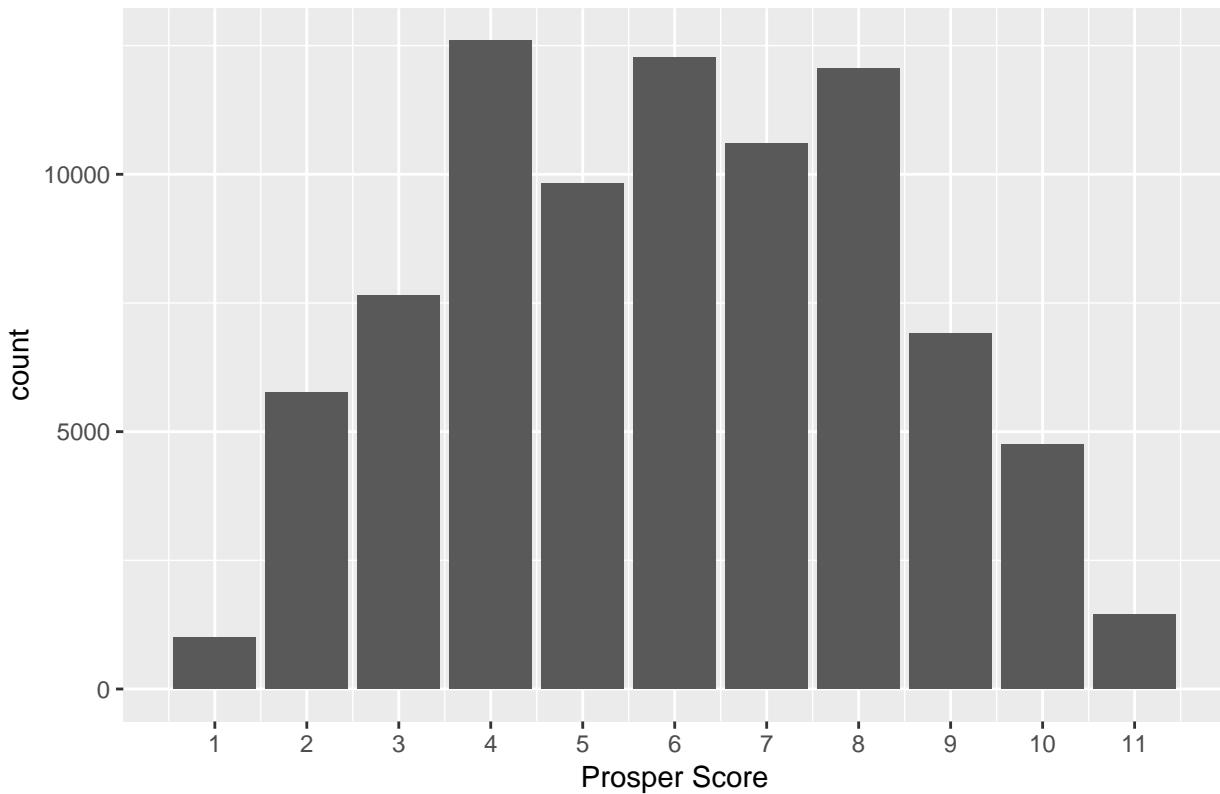
## Average Credit Score



```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max. NA's
##    10.0   670.0  690.0  695.6  730.0  890.0  591
```

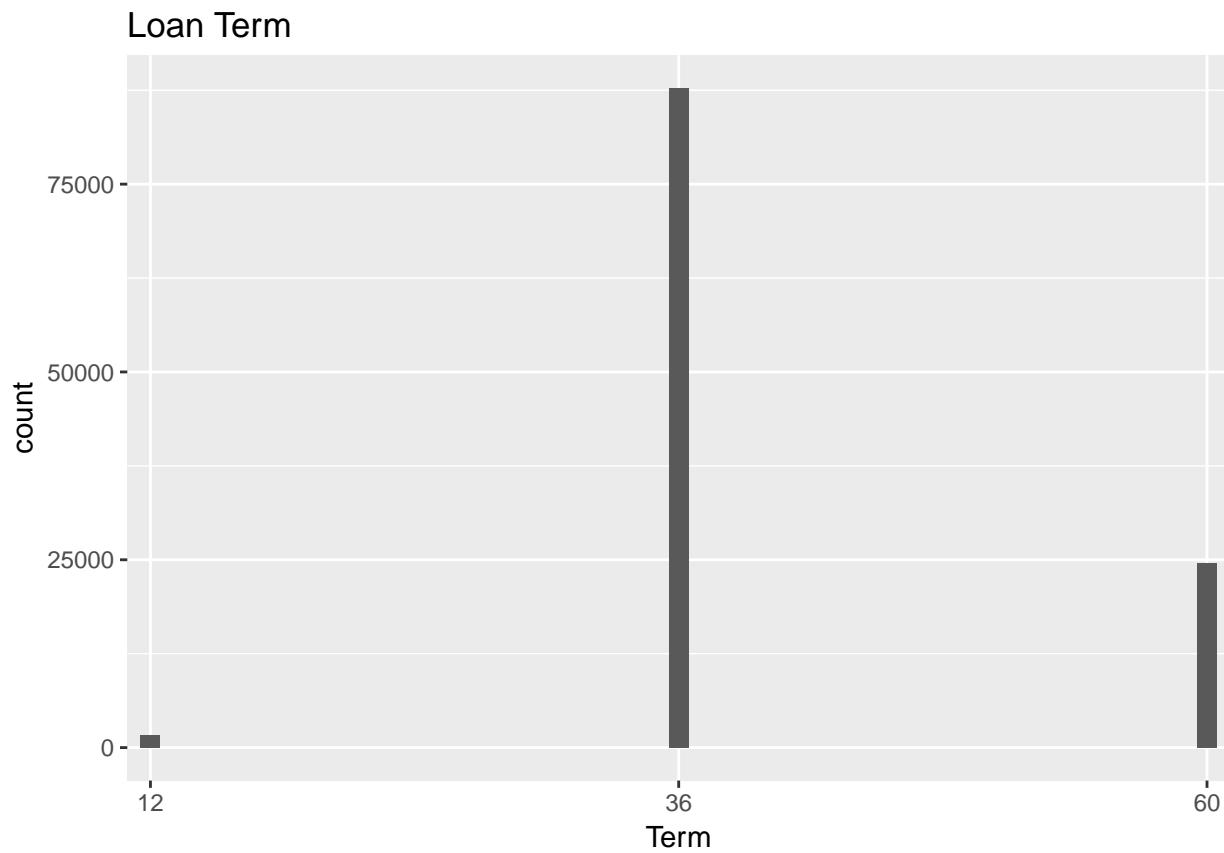
Plot above is the average credit score, which is the average of lower and upper range of credit score ranges. The credit score average plot shows a normal distribution with an average of 695.6.

## Prosper Score (Post-2009)



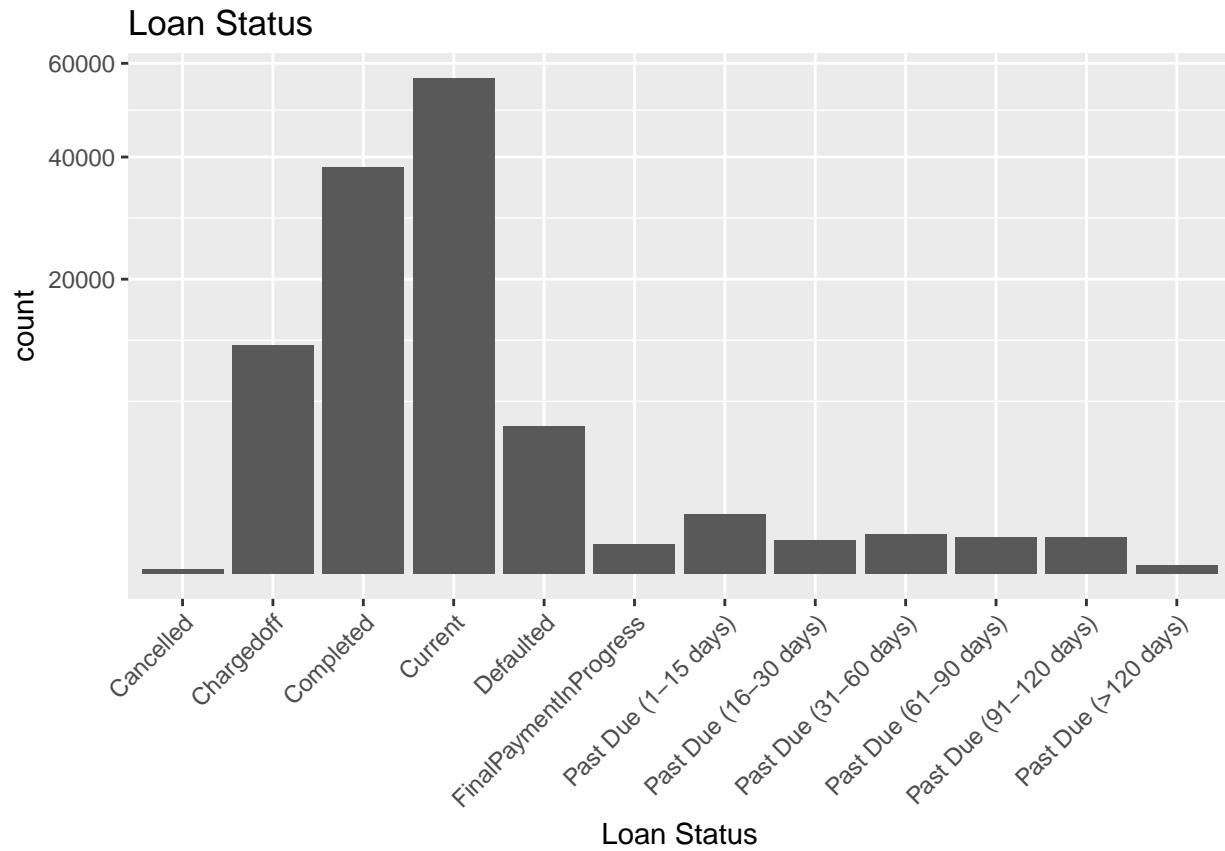
```
##      Min. 1st Qu. Median      Mean 3rd Qu.    Max.    NA's
##    1.00    4.00   6.00    5.95    8.00   11.00 29084
```

The Prosper Score is a custom risk score built using historical Prosper data. The higher Prosper Score indicates the lowest risk users. The Prosper Score data shows a normal distribution.

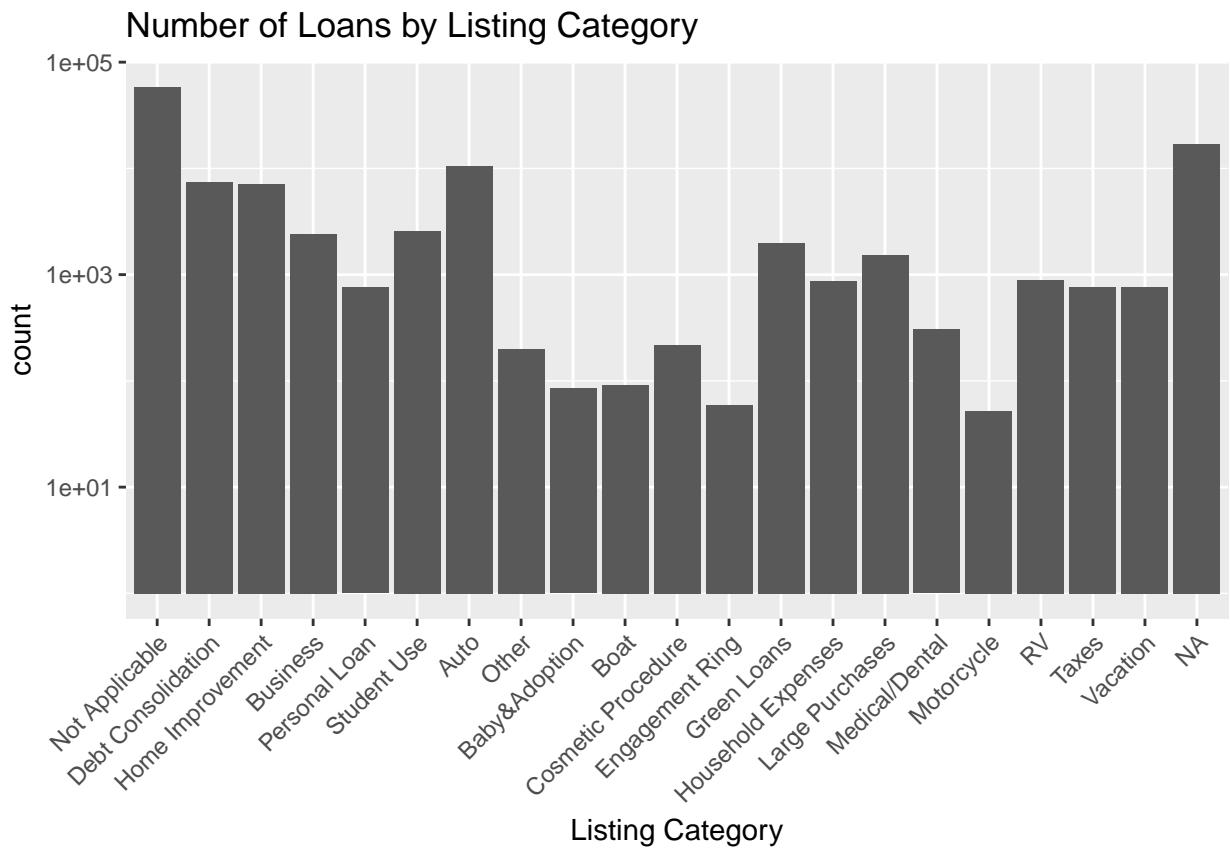


```
##  
##      12     36     60  
##  1614 87778 24545
```

Approximately 77% of the loans were 36 month loans. It is followed by 21.5% of 60 month loans and 1.5% of loans being 12 month loans.

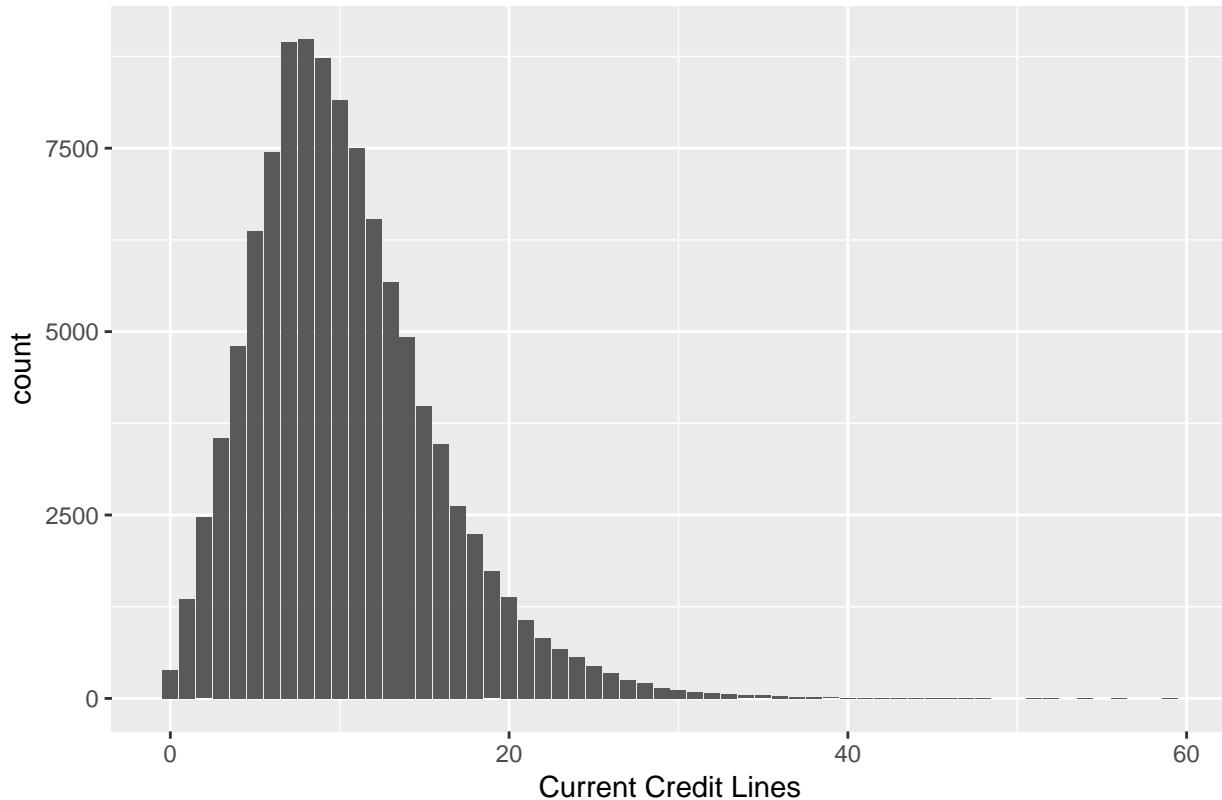


Most of the loan status is current. It is followed by completed, charged off, and defaulted loan status.



Observing the plot, the most common category the borrower selected when posting their listing is auto. It is followed by debt consolidation and home improvement.

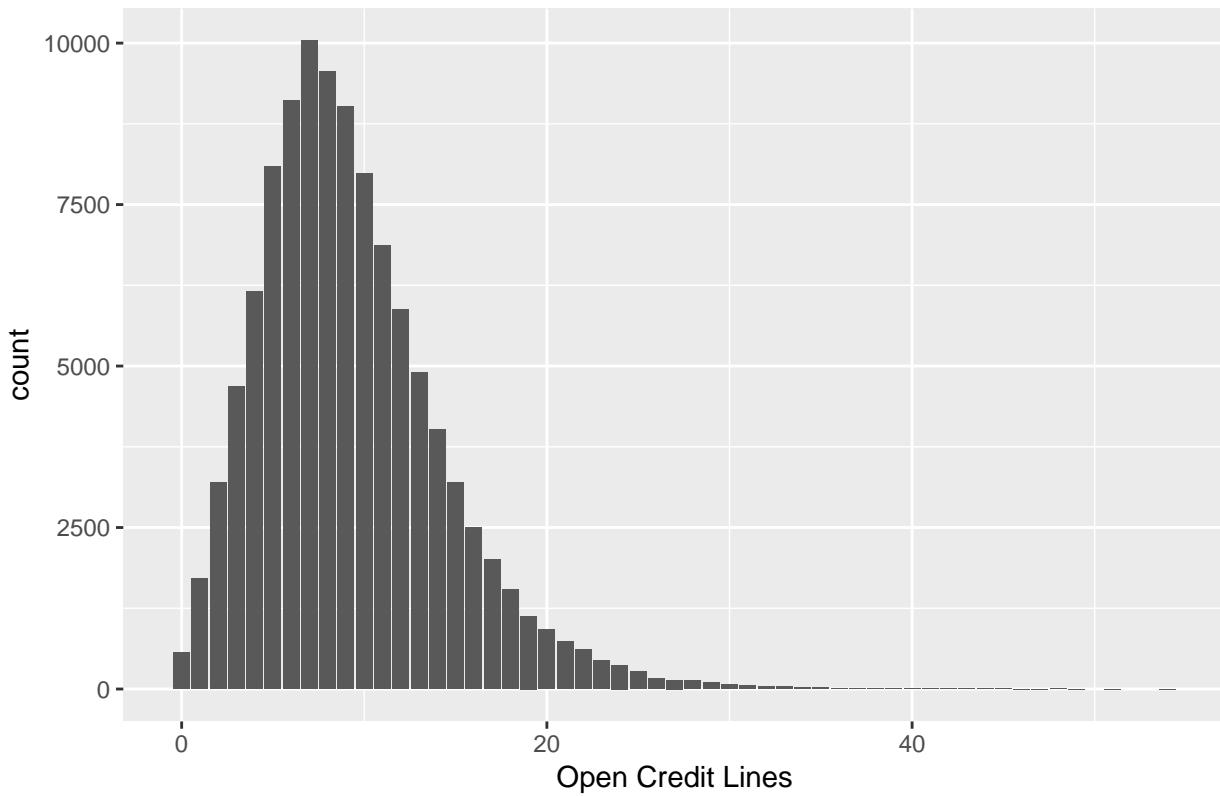
## Number of Current Credit Lines



```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.    NA's
##      0.00    7.00   10.00    10.32   13.00   59.00  7604
```

The number of current credit lines at the time when credit profile was pulled ranges from 0 to 59, with average of 10.32.

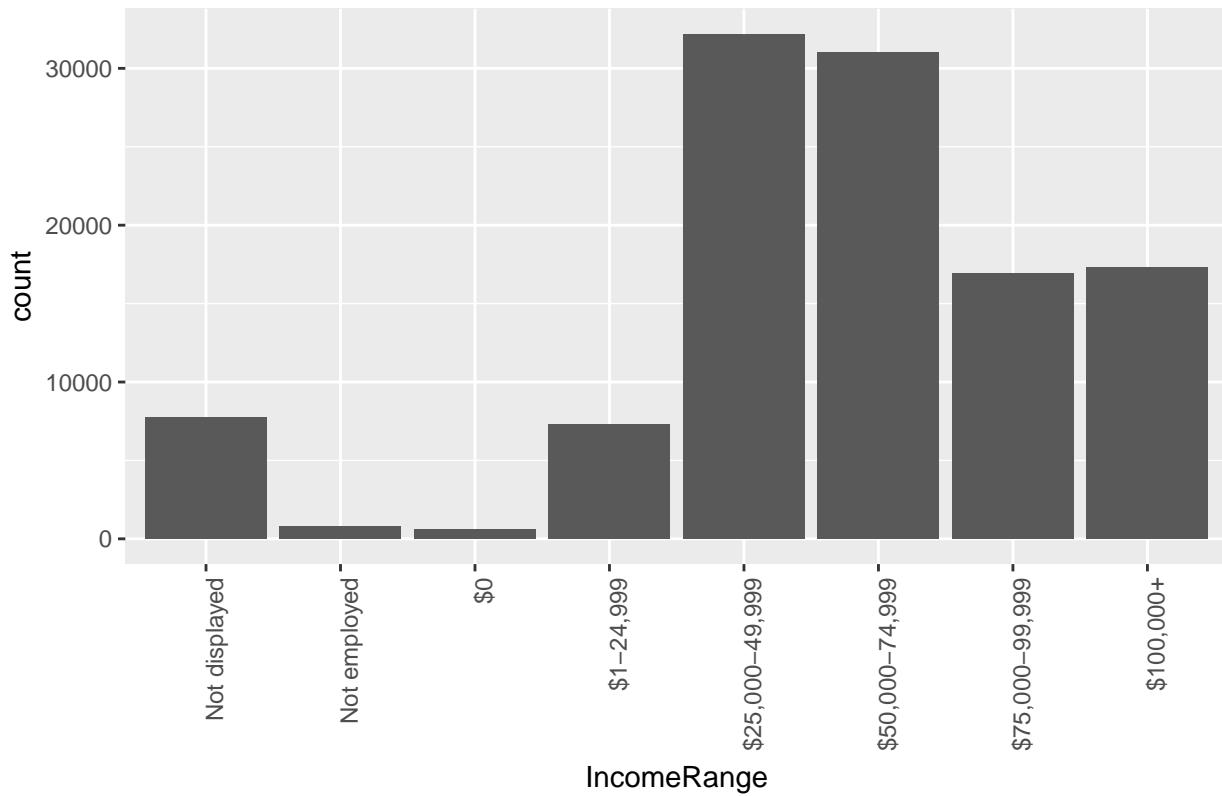
## Number of Loans by Open Credit Lines



```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.    NA's
##      0.00    6.00   9.00    9.26  12.00   54.00  7604
```

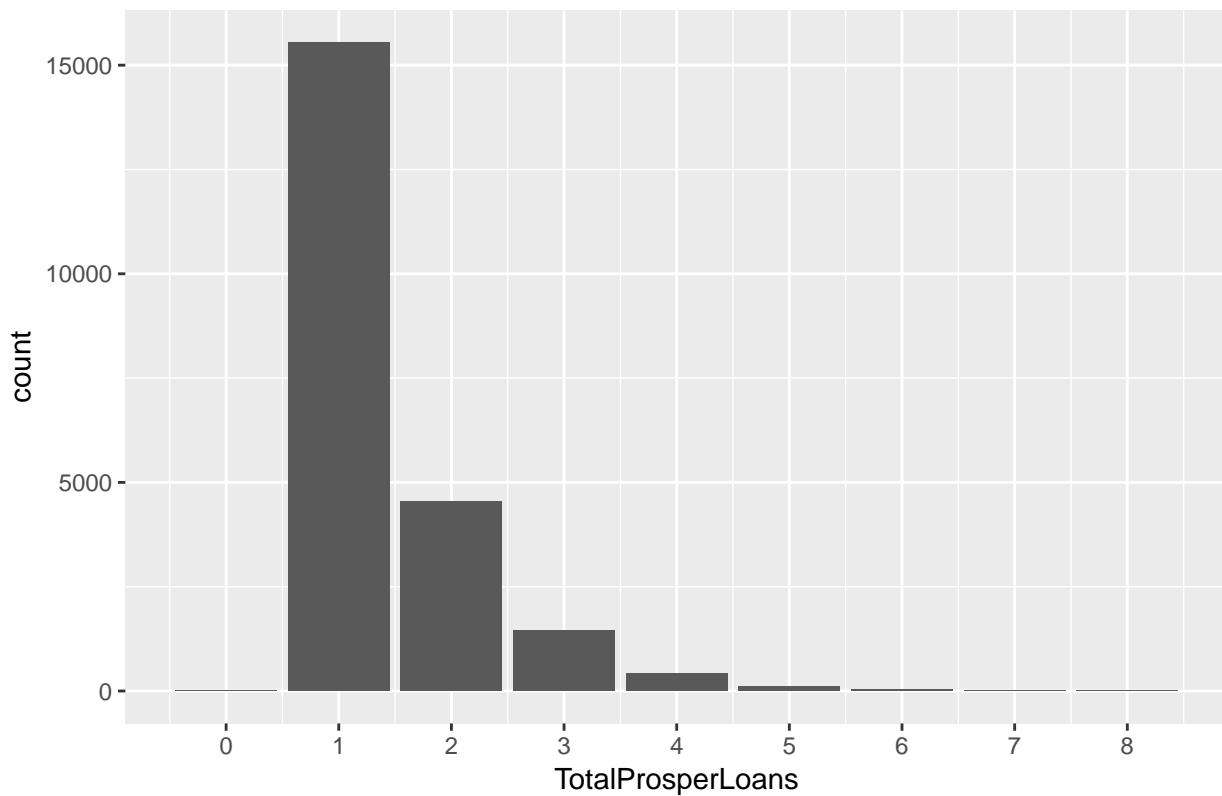
The number of open credit lines at the time when credit profile was pulled ranges from 0 to 54, with average of 9.26.

## Number of Loans by Income Range



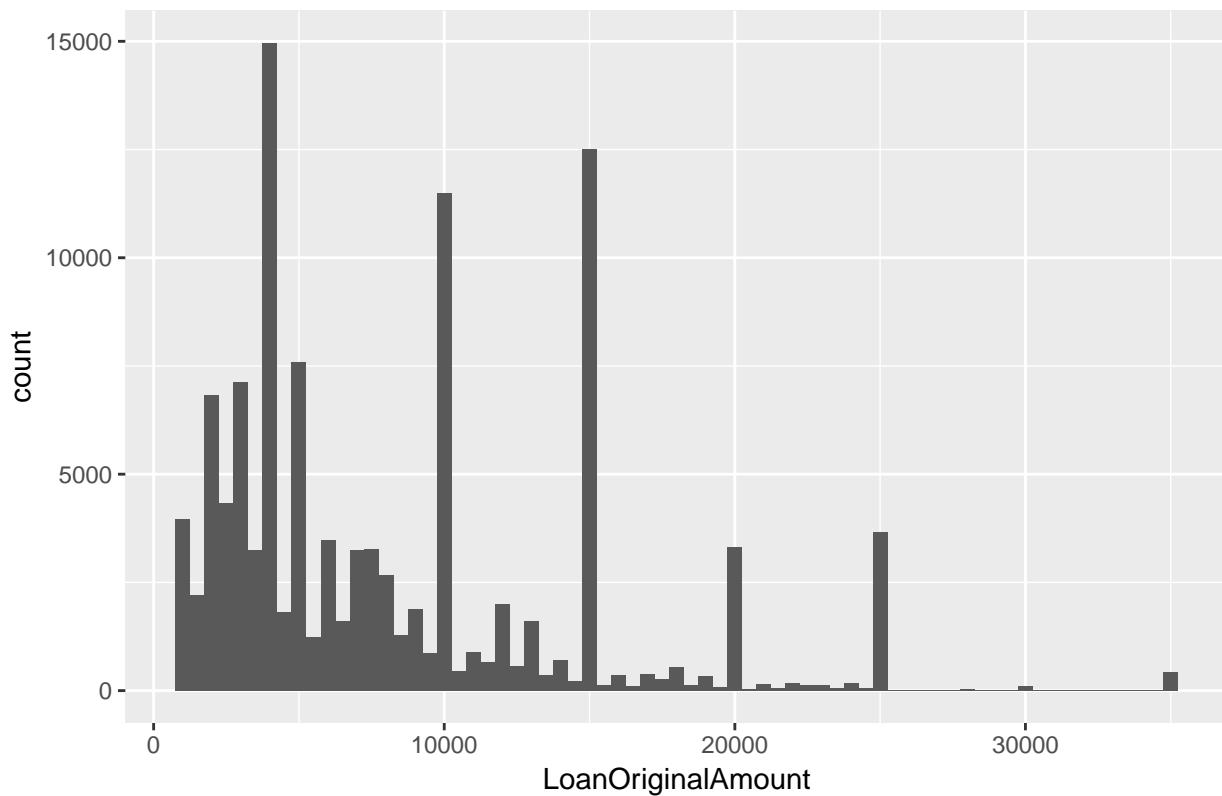
The income range data for borrowers show that the \$25,000-49,999 income bracket request most loans, followed by \$50,000-74,999 income range.

## Total Prosper Loans



Data shows that most of the Prosper applicants were first-time borrowers.

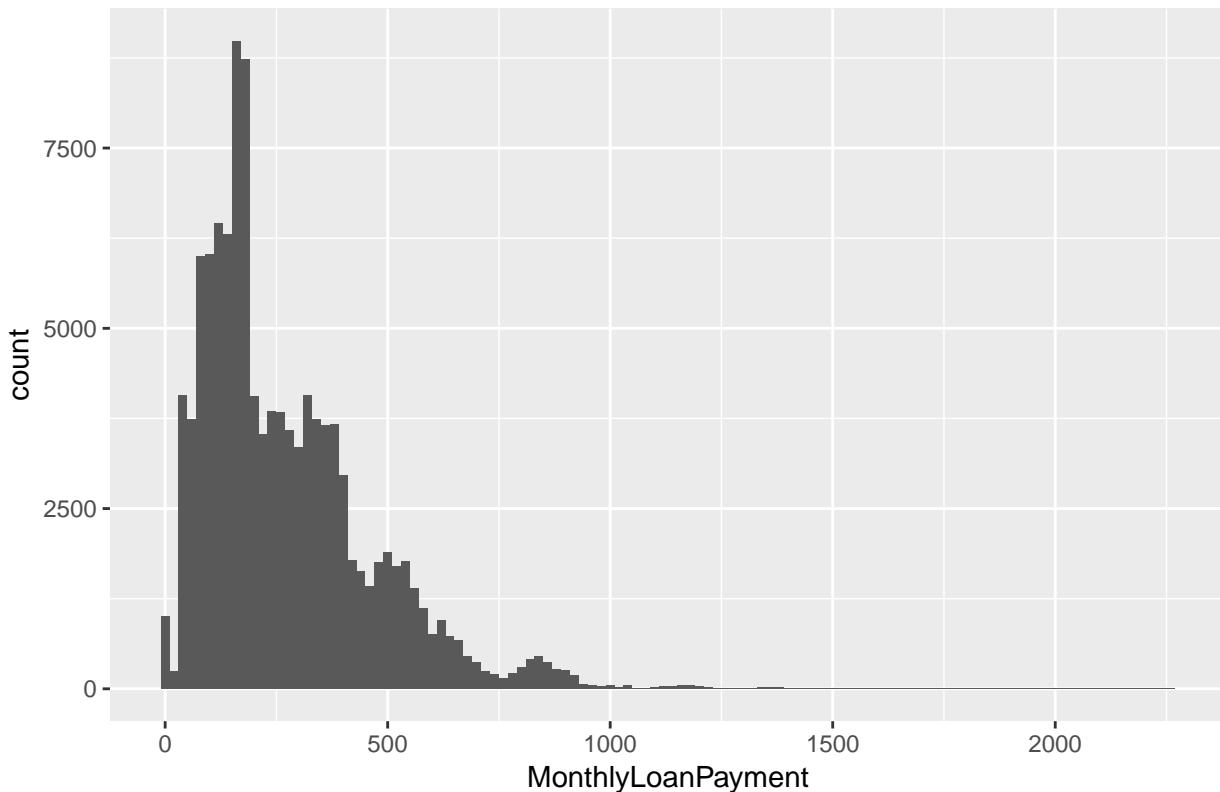
## Number of Loans by Original Loan Amount



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

The original loan amount data shows average of \$8,337. The plot shows significant amount of loans in \$5,000, \$10,000, \$15,000, \$20,000.

## Number of Loans by Original Loan Amount



```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##      0.0   131.6  217.7  272.5  371.6 2251.5
```

The monthly loan payment ranges from \$0 to \$2251.5 with average of \$272.5.

## Univariate Analysis

**What is the structure of your dataset?**

The Prosper loan dataset contains 113,937 loan data across 81 variables.

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

My main interest is how Prosper Score and how it correlates with late payments of loans.

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

In analysis of the dataset, I found the need to create a listing name for ListingCategory, where it originally had numerical values. Also, I found that the ListingCatoryYear would be helpful in seeing trends in listings year to year and CreditScoreAverage variable to have a single reference point for credit score instead of having a lower and upper range.

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

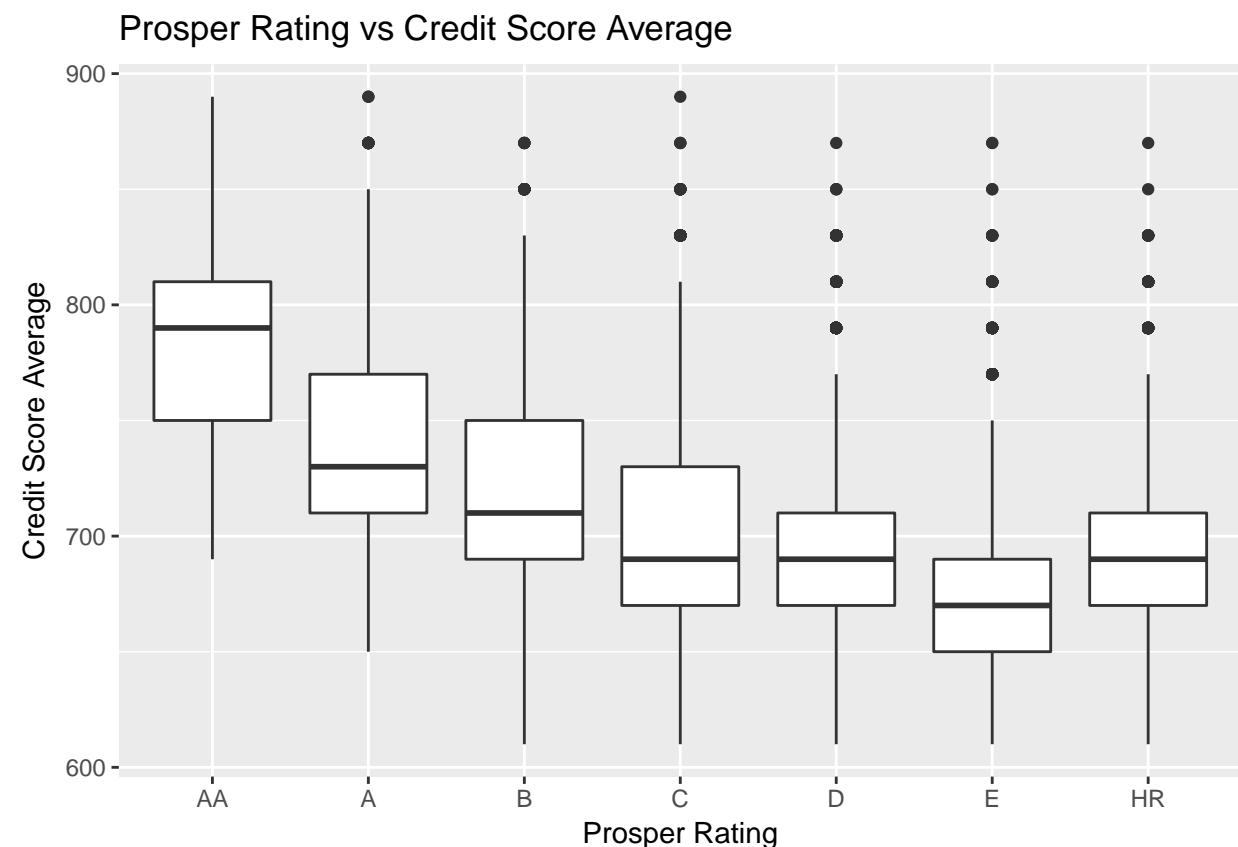
I created ListingCreationYear, ListingCategoryName, and CreditScoreAverage variables to the dataset in assisting analyzing Prosper loan dataset.

**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?**

It was interesting to see how much the 2009 Global financial crisis had impacted the number of Prosper loans initiated that year. I created ListingCreationYear to show prosper loan by year instead of showing continuous data variable.

## Bivariate Plots Section



```
## loans$ProsperRating..Alpha.:
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.    NA's
##      10.0   610.0  650.0   654.4  710.0   890.0    591
## -----
## loans$ProsperRating..Alpha.: A
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.
##      650.0   710.0  730.0   739.9  770.0   890.0
## -----
## loans$ProsperRating..Alpha.: AA
##      Min. 1st Qu. Median     Mean 3rd Qu.     Max.
```

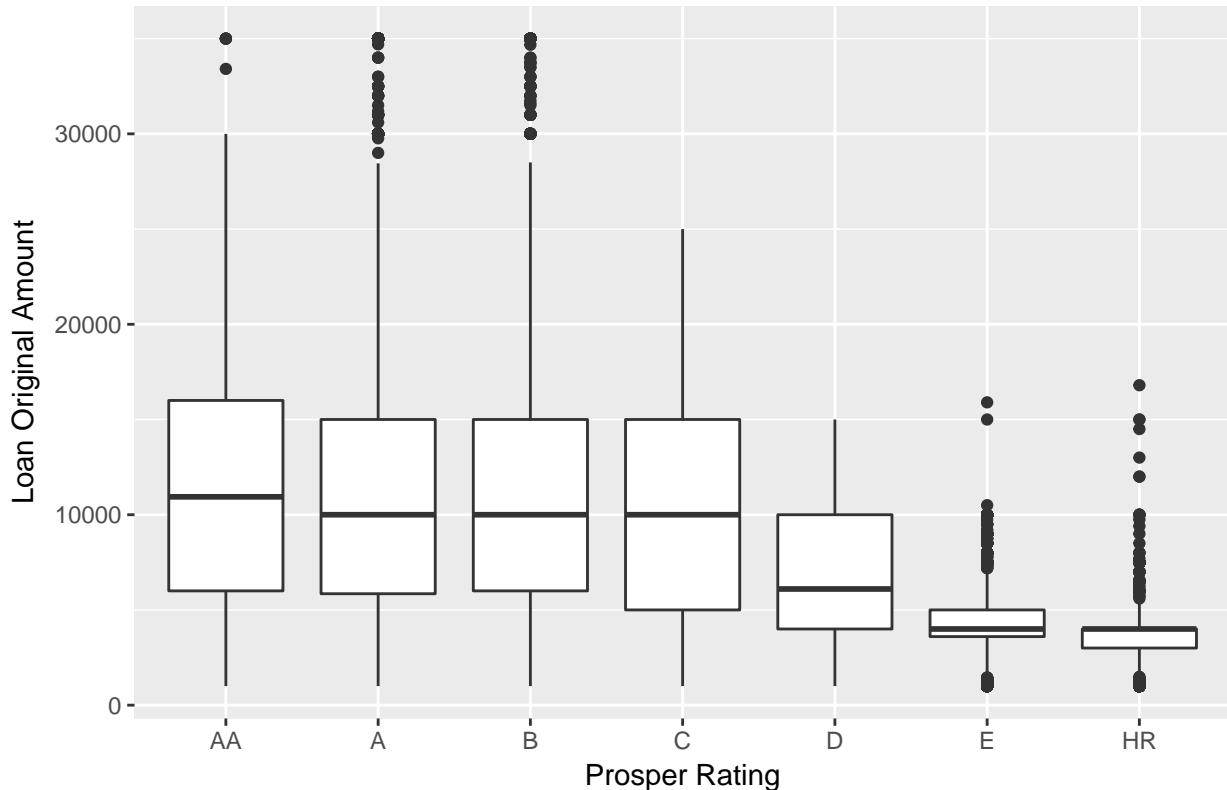
```

##   690.0    750.0    790.0    784.1    810.0    890.0
## -----
## loans$ProsperRating..Alpha.: B
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   610.0    690.0    710.0    716.9    750.0    870.0
## -----
## loans$ProsperRating..Alpha.: C
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   610.0    670.0    690.0    699.9    730.0    890.0
## -----
## loans$ProsperRating..Alpha.: D
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   610.0    670.0    690.0    690.3    710.0    870.0
## -----
## loans$ProsperRating..Alpha.: E
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   610.0    650.0    670.0    672.5    690.0    870.0
## -----
## loans$ProsperRating..Alpha.: HR
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   610     670      690     687      710     870

```

As expected, the higher Prosper Rating is associated with a higher average credit score. However, it was surprising to learn that even the lowest Prosper Ratings of D and E had applicants with average credit score of 870. This fact shows that Prosper Rating is based on multiple factors.

### Prosper Rating vs Loan Original Amount



```

## loans$ProsperRating..Alpha.:
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.

```

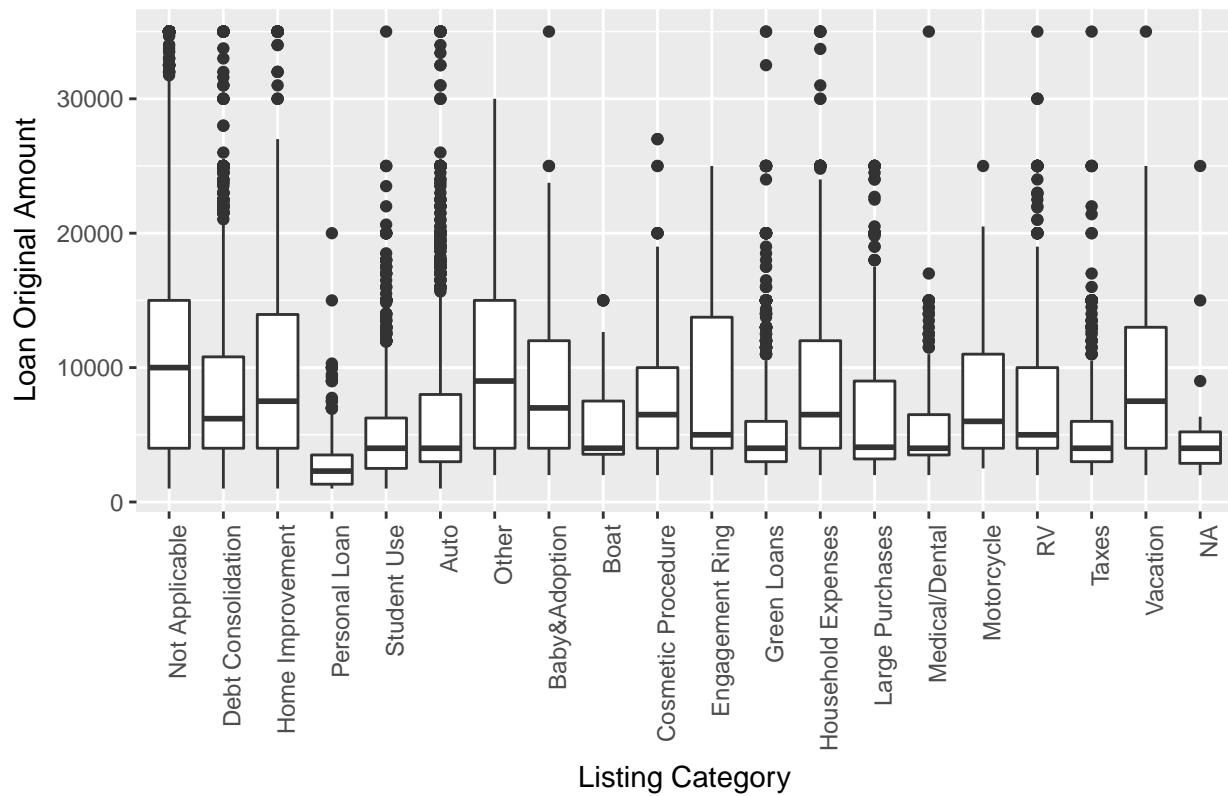
```

##      1000     2500     4500    6159     7904    25000
## -----
## loans$ProsperRating..Alpha.: A
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    5850   10000   11460   15000   35000
## -----
## loans$ProsperRating..Alpha.: AA
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    6000   10940   11584   16000   35000
## -----
## loans$ProsperRating..Alpha.: B
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    6000   10000   11622   15000   35000
## -----
## loans$ProsperRating..Alpha.: C
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    5000   10000   10392   15000   25000
## -----
## loans$ProsperRating..Alpha.: D
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    4000    6100    7083   10000   15000
## -----
## loans$ProsperRating..Alpha.: E
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    3600    4000    4586    5000   15900
## -----
## loans$ProsperRating..Alpha.: HR
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##   1000    3000    4000    3463    4000   16800

```

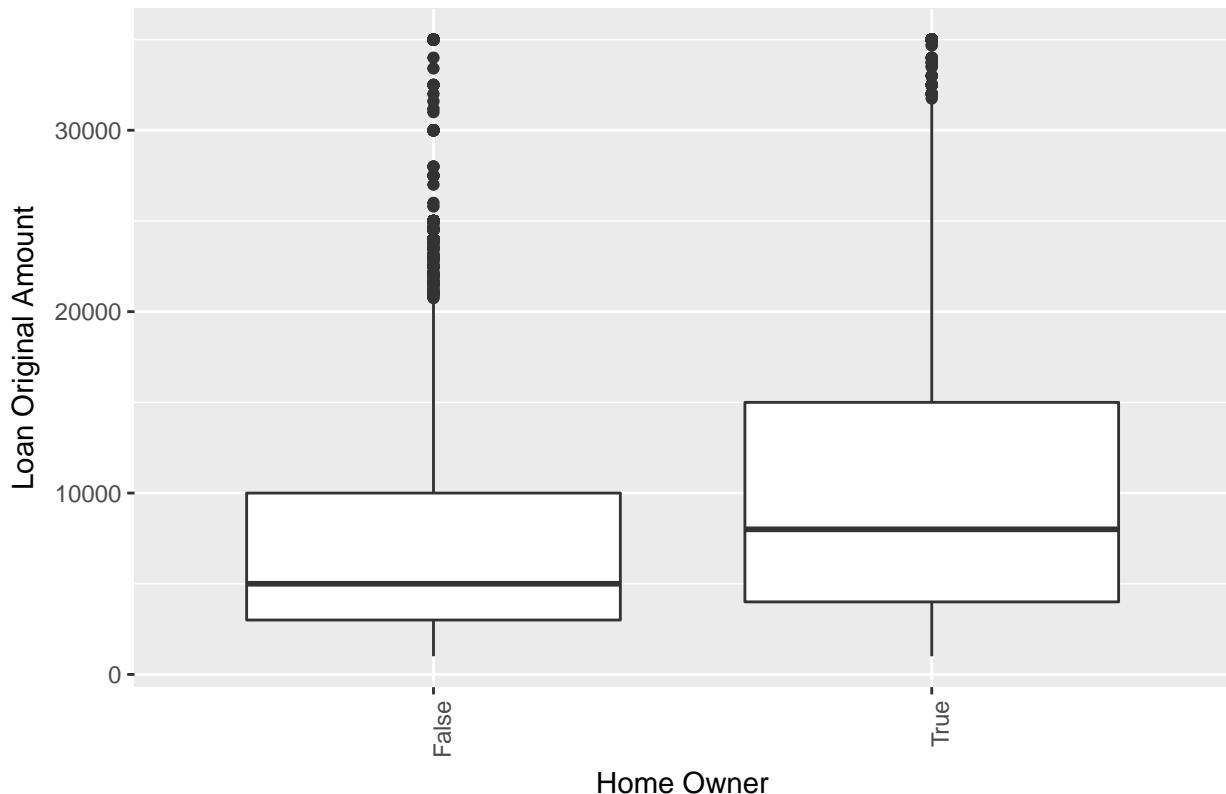
The plot shows correlation between Prosper Rating and loan amount of applicants. Data shows that applicants with higher rating tend to have higher loans on Prosper.

## Listing Category vs Loan Original Amount



The Prosper Rating vs. Loan Original Amount plot shows home improvement category with the highest average loan amount and personal loan category with the lowest average loan amount.

## Home Owner vs Loan Original Amount



The data shows that homeowners have, on average, higher loan amount from Prosper.

## Bivariate Analysis

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?

In investigating listing category vs. loan original amount data, I was surprised to see listing category did not have a significant correlation with loan original amount. I expected greater variance with some category than others.

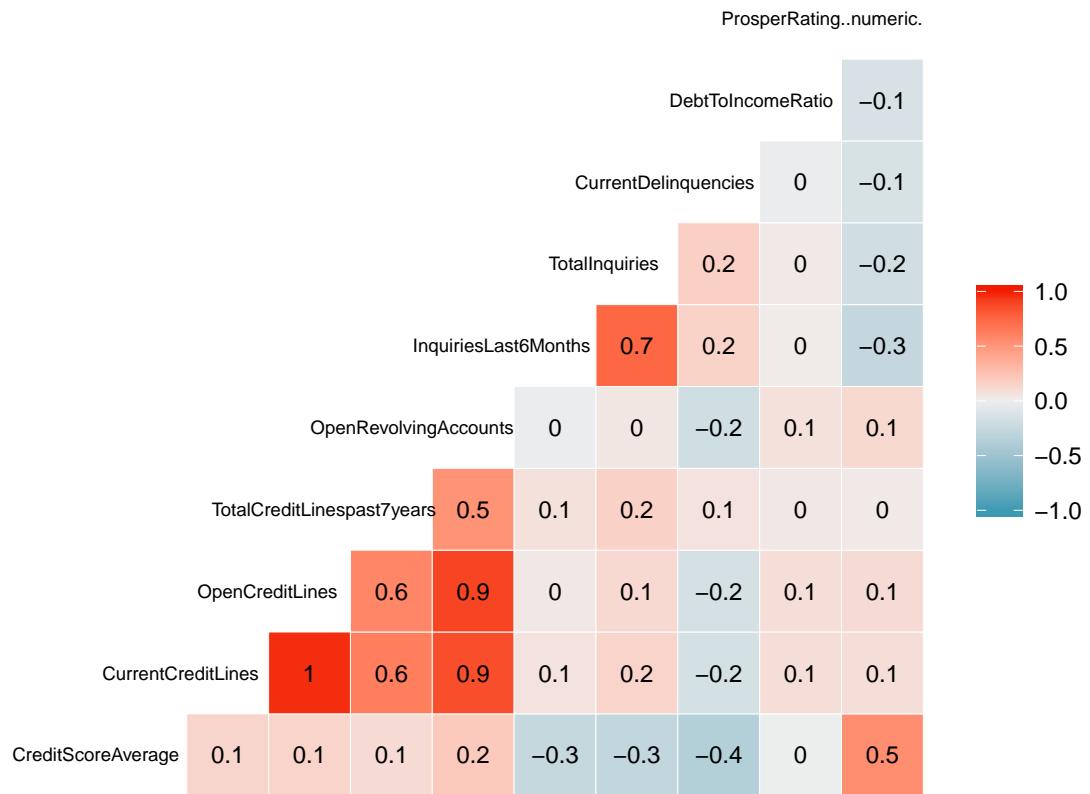
Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

I observed that homeowner status has strong correlation with loan original amount.

What was the strongest relationship you found?

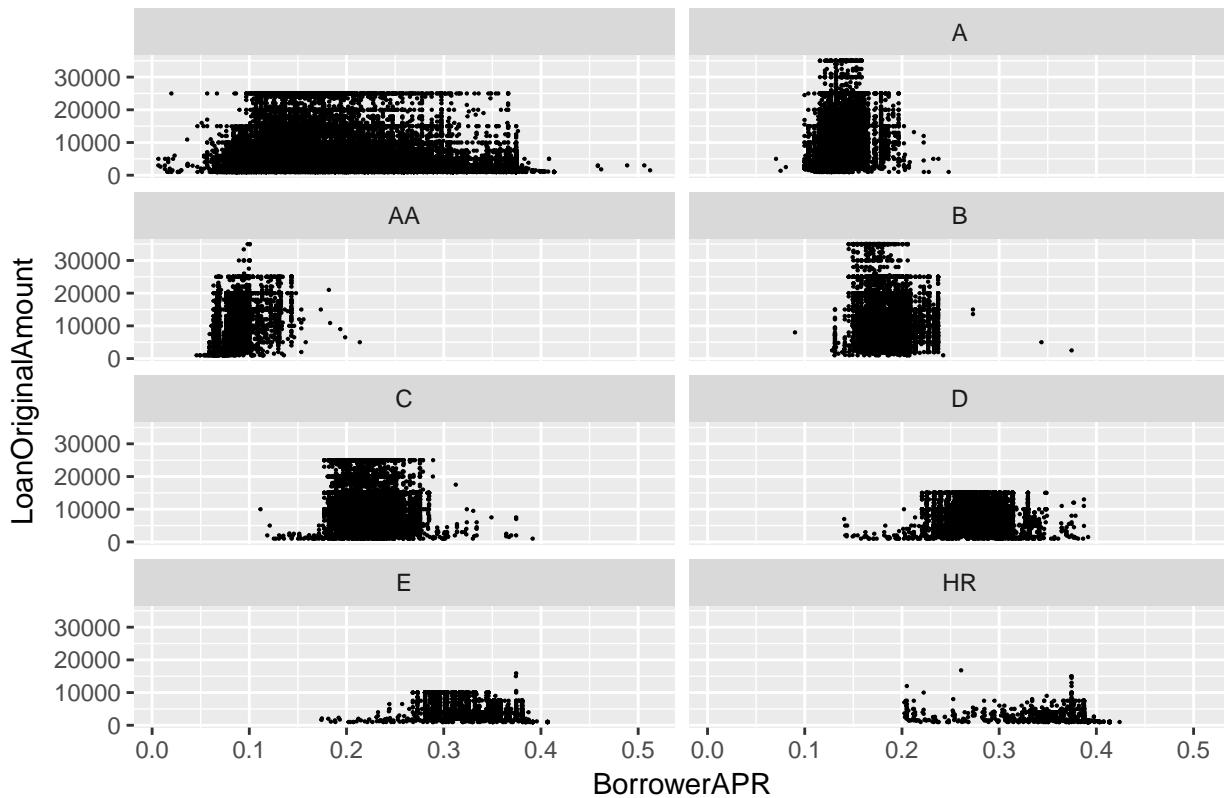
The strongest relationship was prosper rating vs. credit score average. It was still surprising to learn that some people with high credit score still rated low on Prosper rating.

## Multivariate Plots Section



I was interested in learning which factors in the dataset has greatest correlation with each other and ultimately with the Prosper Rating. The heat map shows that the Pearson correlation between chosen variables for investigation. The Prosper Rating has greatest correlation with average credit score. It has little correlation with number of current credit lines, number of open credit lines, and number of revolving accounts.

## Borrower APR vs Loan Original Amount by Prosper Rating



In this plot, I wanted to investigate how much loans people take depending on Prosper rating. As expected, higher credit raters enjoy lower APR. The plot shows that higher credit raters tends to take higher loans, while lower credit raters has more variance in APR even within Prosper rating bracket.

## 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?**

In investigating subset of Prosper loan data, I was able to find that the average credit score had the strongest correlation.

In investigating Borrower APR vs Loan Original Amount by Prosper Rating I found that people with higher Prosper Rating are likely to have larger loans. I also found that the lower Prosper Rating users see more variation in APR even within their Prosper Rating bracket.

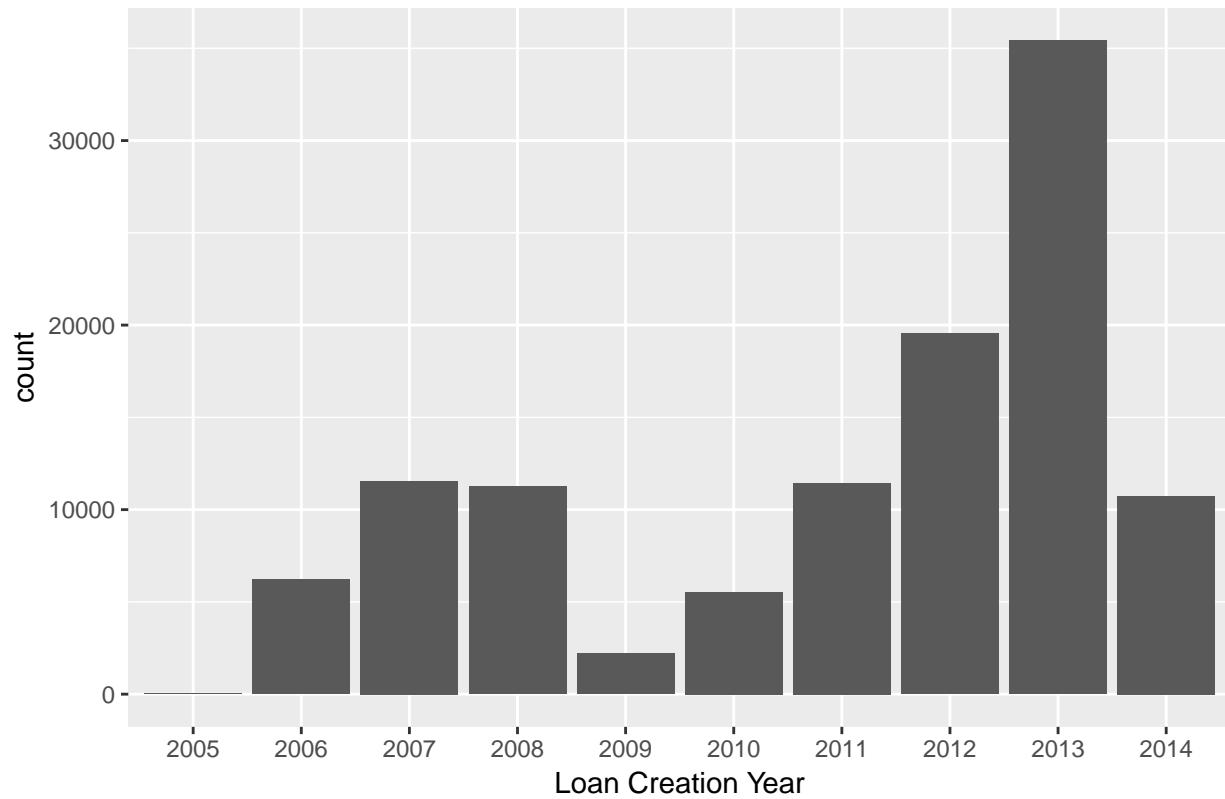
**Were there any interesting or surprising interactions between features?**

I was surprised to see little correlation betewen number of credit accounts opened and Prosper Rating.

## Final Plots and Summary

### Plot One

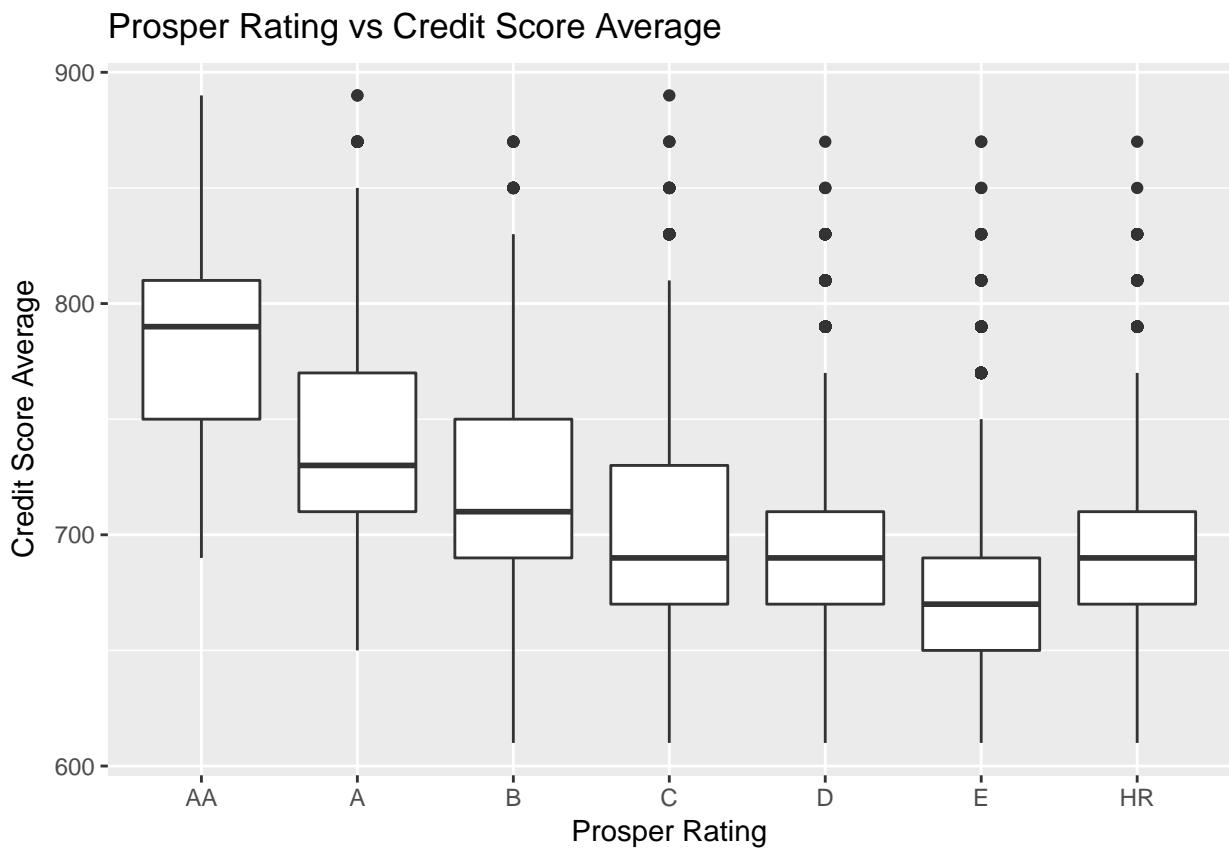
Number of Loans by Year



### Description One

This plot shows not only shows the growth of Prosper's business but also the growth of peer-to-peer lending service. It is showing steady growth since 2005, with the exception of Global Finanical Crisis in 2009.

## Plot Two

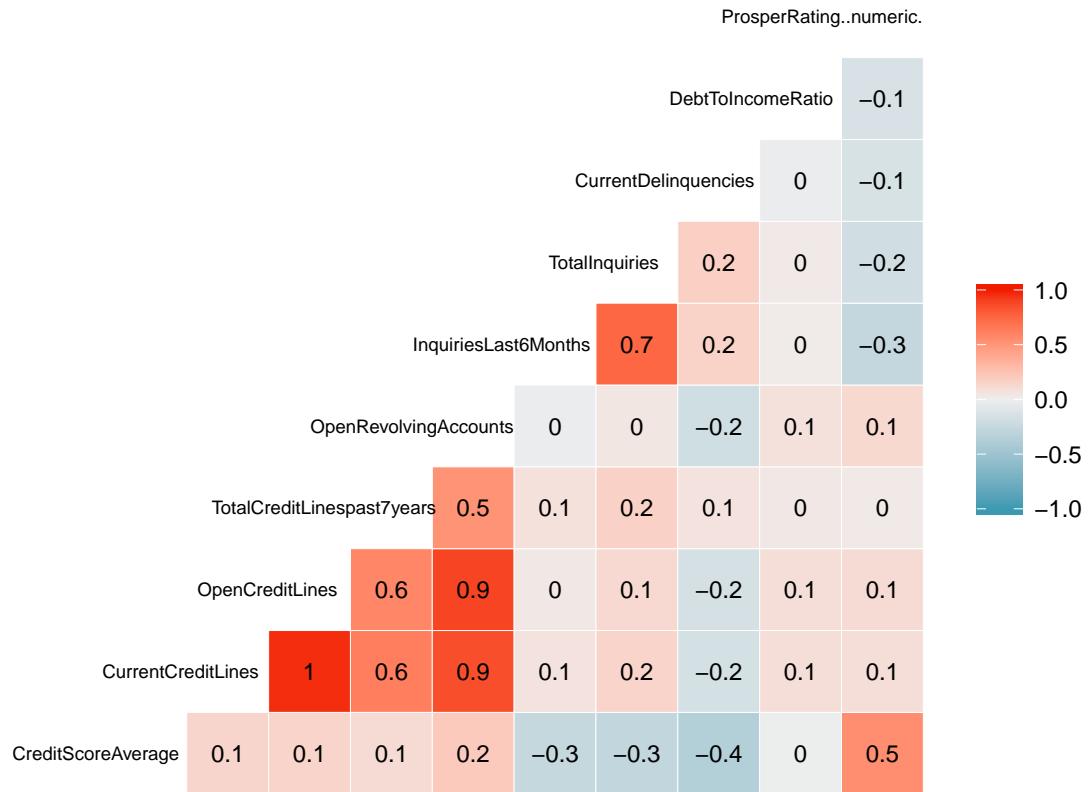


## Description Two

I was interested to see how Prosper rating is calculated. Although the Prosper Rating is strongly correlated with credit scores, I was also surprised to see lower Prosper Rating associated with high credit scores. This shows that Prosper Rating is calculated based on multiple factors.

## Plot Three

```
## Warning in ggcorm(loancredit, label = TRUE, label_size = 3, hjust = 0.8, :  
## data in column(s) 'IncomeRange' are not numeric and were ignored
```



### Description Three

In continuation of investigating how Prosper Rating is calculated, I created a subset of data with the factors that could potentially have correlation with Prosper Rating. The credit score has the strongest correlation to Prosper Rating.

---

## Reflection

Investigating the Prosper Loan dataset, I gained insight into the peer-to-peer lending service. I had no idea on who used this service and the APR the users enjoyed. I was mainly interested in who borrowed most for what purpose. Most of my struggles were with the ggplot syntax and organizing plots so that it displayed a trend that user to easily interpret.

Future work could include investigating Prosper loans by state. The number of loans by state could reveal insight into where Prosper could market further of its service. The listing category by state could show which market segment Prosper could target by each state.