

Explore and summarize prosperLoanData

Simon Shen

01/21/2018

Prosper Marketplace is America's first peer-to-peer lending marketplace, with over \$7 billion in funded loans. Here, borrowers can request personal loans and investors can fund anywhere from \$2,000 to \$35,000 per loan request. I am interested in the features regarding to borrowers, lenders and Prosper.org. hence, the next we're going to explore and the prosper loan dataset.

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

There are 113937 observations and 81 variables in the data set.

```
## [1] "ListingKey"
Reference ## [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"
Reference ## [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"

```

All features of this dataset prosperLoanData

Due to many variables in this dataset, I only use 15 variables. Those are:

- Term
- BorrowerAPR
- BorrowerRate
- LenderYield
- ProsperRating..numeric.
- ProsperRating..Alpha.
- listingCategory..numeric.
- Occupation
- EmploymentStatus
- LoanOriginalAmount
- IncomeRange
- CurrentDelinquencies
- LoanOriginationDate
- LoanOriginationQuarter
- Investors names(prosper_raw)

```
## [1] 113937      15
```

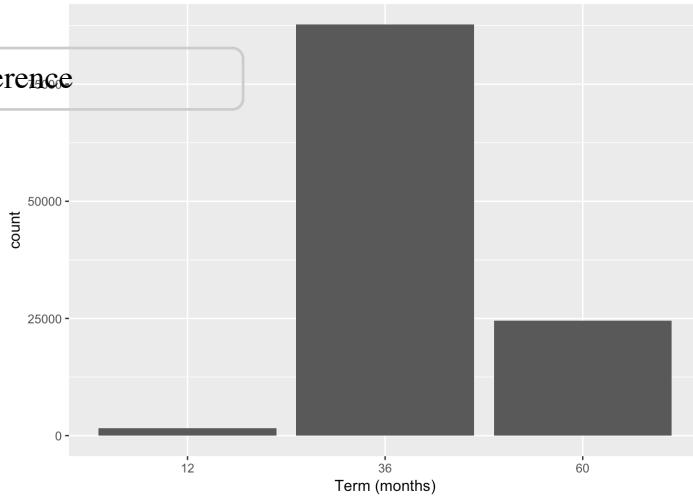
Reference

```
## 'data.frame': 113937 obs. of 15 variables:
##   $ Term                  : int 36 3
##   $ BorrowerAPR            : num 0.16
##   $ BorrowerRate            : num 0.15
##   $ LenderYield             : num 0.13
##   $ ProsperRating..numeric. : int NA 6
##   NA 6 3 5 2 4 7 7 ...
##   $ ProsperRating..Alpha.    : Factor w/
##     levels "", "A", "AA", "B", ...: 1 2 1 2 6 4 7
##     5 3 3 ...
##   $ ListingCategory..numeric.: int 0 2
##     0 16 2 1 1 2 7 7 ...
##   $ Occupation               : Factor w/
##     68 levels "", "Accountant/CPA", ...: 37 43 37
##     52 21 43 50 29 24 24 ...
##   $ EmploymentStatus          : Factor w/
##     9 levels "", "Employed", ...: 9 2 4 2 2 2 2 2
##     2 2 ...
##   $ LoanOriginalAmount        : int 9425
##     10000 3001 10000 15000 15000 3000 10000 10
##     000 10000 ...
##   $ IncomeRange                : Factor w/
##     8 levels "$0", "$1-24,999", ...: 4 5 7 4 3 3
##     4 4 4 4 ...
##   $ CurrentDelinquencies       : int 2 0
##     1 4 0 0 0 0 0 ...
##   $ LoanOriginationDate        : Date, for
##     mat: "2007-09-12" "2014-03-03" ...
##   $ LoanOriginationQuarter      : Factor w/
##     33 levels "Q1 2006", "Q1 2007", ...: 18 8 2 3
##     2 24 33 16 16 33 33 ...
##   $ Investors                  : int 258
##     1 41 158 20 1 1 1 1 1 ...
```

Univariate Plots Section

In this section, I want to explore the dataset regarding to borrowers created in prosper, the Loan Term they are more likely to select, the prosper rating, the borrowers' occupation and their annual incomes, listing category,borrower's interest rate, lender yield, loan amount, borrowers late payment and the investors distributions.

Reference

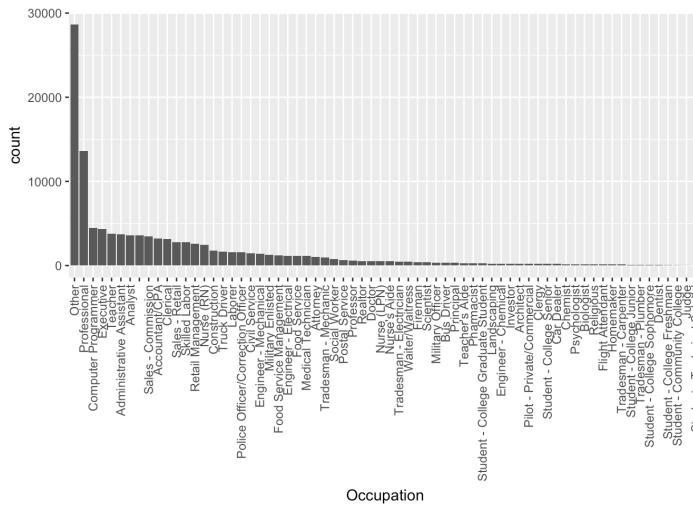


```
##    12     36     60
## 1614  87778 24545
```

There are almost 80% of the borrowers more likely to choose 3-year fixed loan term while they creating the listing.

```
## [1] "Q1 2011" "Q4 2012"
```

There are 1614 loan listing with fixed Term 1 year, those listing are created from 2010 to 2013, I am not sure if this part of data is correct or not? as from prosper.com, that only support Fixed term—3 or 5 years.

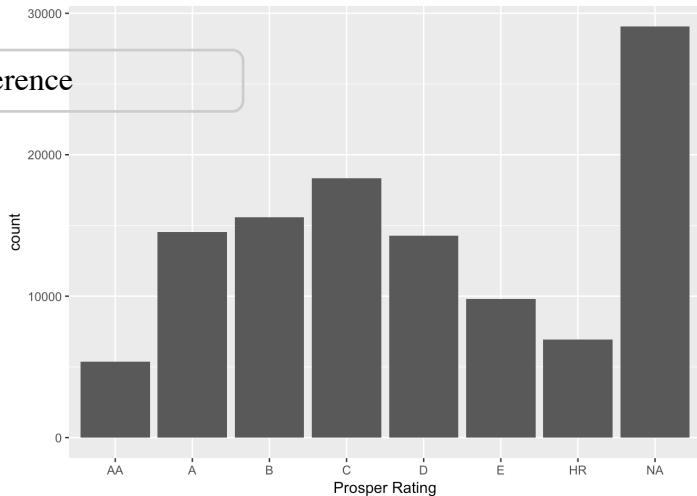


Reference	Other
##	
Professional	
##	28617
13628	
##	Computer Programmer
Executive	
##	4478
4311	
##	Teacher
Administrative Assistant	
##	3759
3688	
##	Analyst
##	3602
3588	
##	Sales - Commission
Accountant/CPA	
##	3446
3233	
##	Clerical
Sales - Retail	
##	3164
2797	
##	Skilled Labor
Retail Management	
##	2746
2602	
##	Nurse (RN)
Construction	
##	2489
1790	
##	Truck Driver
Laborer	
##	1675
1595	
## Police Officer/Correction Officer	
Civil Service	
##	1578
1457	
##	Engineer - Mechanical
Military Enlisted	
##	1406
1272	
##	Food Service Management
Engineer - Electrical	
##	1239
1125	
##	Food Service
Medical Technician	
##	1123
1117	
##	Attorney

Tradesman - Mechanic	
##	1046
Reference	
##	Social Worker
Postal Service	
##	741
627	
##	Professor
Realtor	
##	557
543	
##	Doctor
Nurse (LPN)	
##	494
492	
##	Nurse's Aide
Tradesman - Electrician	
##	491
477	
##	Waiter/Waitress
Fireman	
##	436
422	
##	Scientist
Military Officer	
##	372
346	
##	Bus Driver
Principal	
##	316
312	
##	Teacher's Aide
Pharmacist	
##	276
257	
## Student - College Graduate Student	
Landscaping	
##	245
236	
##	Engineer - Chemical
Investor	
##	225
214	
##	Architect
Pilot - Private/Commercial	
##	213
199	
##	Clergy
Student - College Senior	
##	196
188	
##	Car Dealer

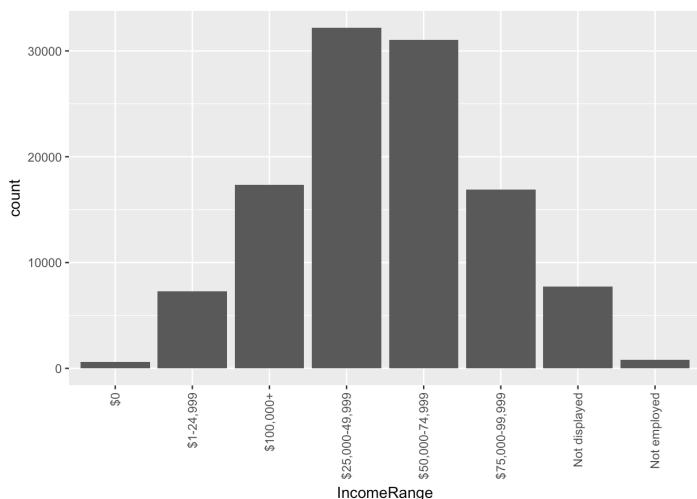
Chemist	
##	180
Reference	
##	
145	
##	Psychologist
Biologist	
##	145
125	
##	Religious
Flight Attendant	
##	124
123	
##	Homemaker
Tradesman - Carpenter	
##	120
120	
##	Student - College Junior
Tradesman - Plumber	
##	112
102	
##	Student - College Sophomore
Dentist	
##	69
68	
##	Student - College Freshman
Student - Community College	
##	41
28	
##	Judge
Student - Technical School	
##	22
16	

The largest number is Other, it is understandable, because all cases not in above can be Other. If we don't consider the "Other" case, Professional has the most top count and almost three times as many as the second. The count of borrowers on the top three is professional, Computer Programer, Executive, at the bottom of the list are the students, generally, the students don't need to borrow money, or they don't have strong credit rating to create loan request from prosper.



```
##          A      AA      B      C      D
E     HR
## 29084 14551  5372 15581 18345 14274  97
95   6935
```

Prosper Ratings, from lowest-risk to highest-risk, are labeled AA, A, B, C, D, E, and HR (“High Risk”), we can see 29084 prosper rating is labeled “NA”, this is understandable since the prosper rating is provided since 2009 after SEC registration, there is no rating labeled to the borrowers before 2009. If we do not consider the “NA”, the prosper rating appears Binomial distribution, 57% of borrowers are labeled middle risk B,C and D. Borrowers’s rating C are the highest frequency up to 22%.

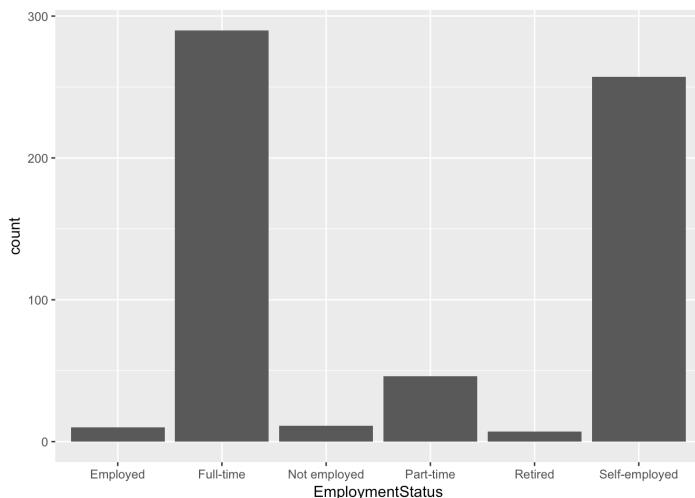


```

##          $0      $1-24,999      $100
Reference 0-24,999+ $25,000-49,999 $50,000-74,999
##          621        7274
17337      32192        31050
## $75,000-99,999 Not displayed Not emp
loyed
##          16916        7741
806

```

The borrower's income ranges appears Binomial distribution. most of borrowers' annual income are fall into \$25,000 - 75,000. and there are 621 borrowers with zero incomes they got loans, it's default value if prosper not get the source of borrower's incomes, or still the borrowers can get loans if they have good credit rating from other path.



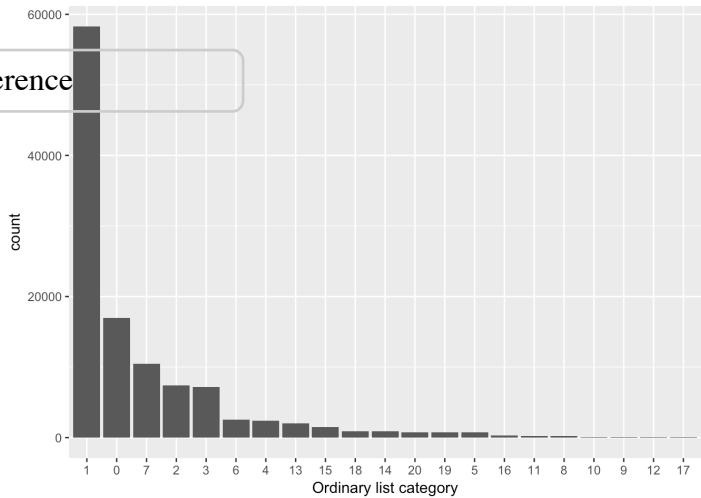
```

##          Employed      Full-ti
me Not available Not employed
##          2255        67322        263
55          5347        835
##          Other       Part-time     Retir
ed Self-employed
##          3806        1088         7
95          6134

```

```
## [1] 11 15
```

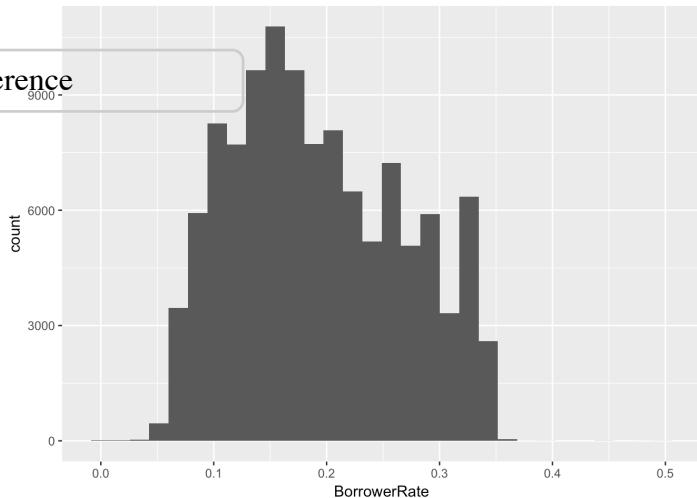
Further exploring the borrowers whose incomes are zero, find 11 borrowers are Not employed, they may not have incomes? all the loan requests were created on Q3 2007. I guess it should be data issue on the person with zero annual income.



##	1	0	7	2	3	6
4	13	15	18	14	20	
##	58308	16965	10494	7433	7189	2572
95	1996	1522	885	876	771	23
##	19	5	16	11	8	10
9	12	17				
##	768	756	304	217	199	91
85	59	52				

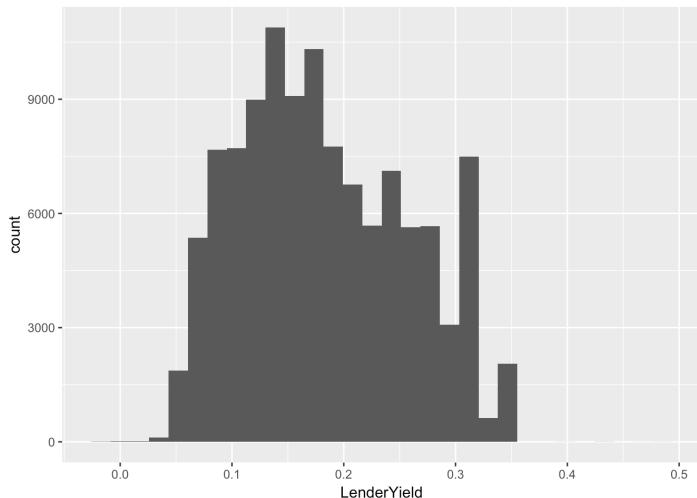
Category: The category of the listing that the borrower selected when posting their listing: 0 - Not Available, 1 - Debt Consolidation, 2 - Home Improvement, 3 - Business, 4 - Personal Loan, 5 - Student Use, 6 - Auto, 7- Other, 8 - Baby&Adoption, 9 - Boat, 10 - Cosmetic Procedure, 11 - Engagement Ring, 12 - Green Loans, 13 - Household Expenses, 14 - Large Purchases, 15 - Medical/Dental, 16 - Motorcycle, 17 - RV, 18 - Taxes, 19 - Vacation, 20 - Wedding Loans.

Out of the listing category, three categories comes out as the highest past 10000 loans. There's N/A and Other categories, so we can't know for sure the specific category. But one comes out highest which is category 1 - Debt Consolidation, where's one take out a loan to pay many others. This comes really high with 58308 loans,it could be that many Prosper visitors comes with already have loans, and want to search some loans to pay for it?



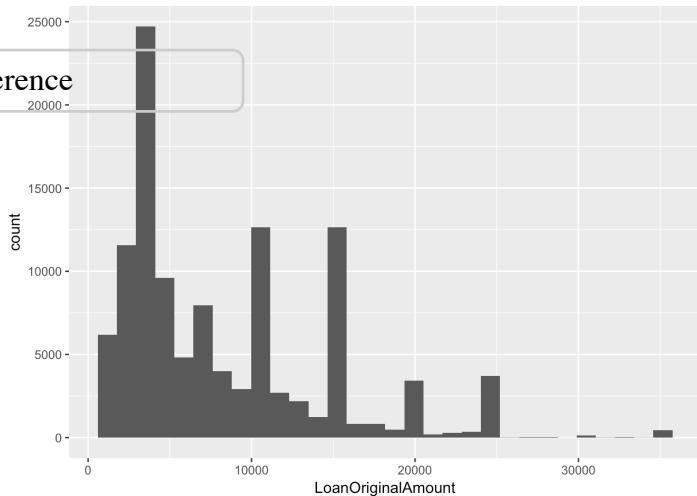
```
##      Min. 1st Qu. Median      Mean 3rd Qu.
Max.
##  0.0000  0.1340  0.1840  0.1928  0.2500
0.4975
```

The borrower's rate follow an almost unimodal distribution, with the peak around 0.16. There's small spike occurs around 0.3.



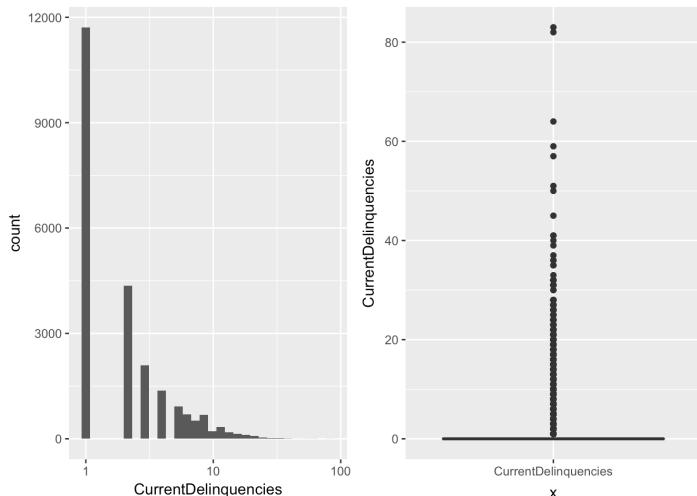
```
##      Min. 1st Qu. Median      Mean 3rd Qu.
Max.
## -0.0100  0.1242  0.1730  0.1827  0.2400
0.4925
```

It's very similar to the borrower's rate, the lender's yield also follow an almost unimodal distribution, with the peak around 0.14. There's small spike occurs around 0.3.



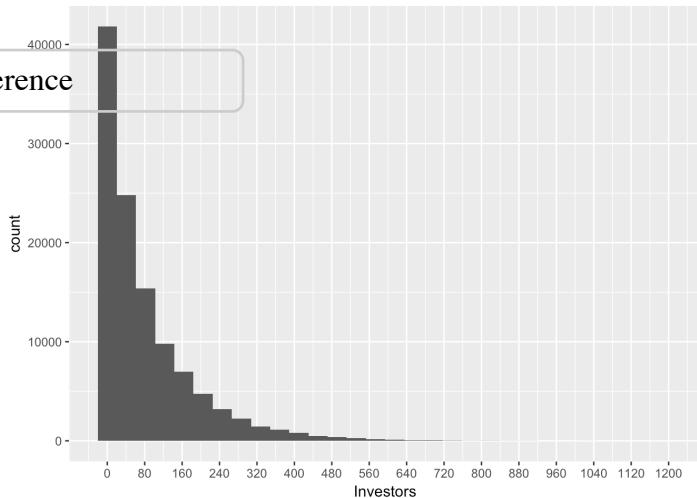
```
##      Min. 1st Qu. Median      Mean 3rd Qu.
Max.      1000     4000    6500    8337   12000
##      35000
```

This is interesting, here we see that the distribution is right skewed, with the peak USD 3500, then some small spikes occur around 10000, 20000, 30000 dollars.



```
##      Min. 1st Qu. Median      Mean 3rd Qu.
Max.      NA's
##      0.0000  0.0000  0.0000  0.5921  0.0000
83.0000    697
```

Here we can see that there is 75% of borrowers no delinquency record on prosper peer-to-peer lending marketplace, to my surprise, we found the maximum delinquency up to 83. It's really high.



```
##      Min. 1st Qu. Median      Mean 3rd Qu.
Max.      1.00    2.00   44.00  80.48 115.00
          1189.00
```

We see that 25% of listing are funded by 1 or 2 investors, most of listing are less than 40 investors.

Univariate Analysis

There are 113937 observations and 81 variables in the raw data set. Due to many variables in this dataset, I only select 15 features from dataset prosperLoanData.

Which Term loan the borrowers prefer to select?

There are almost 80% of the borrowers are more likely to select 3-year fixed loan while they creating the list. Why it's 3-year? is it related to BorrowerRate? Is longer term caused higher BorrowerRate? I will explore in the following section later.

what people need money?

If we don't consider the "Other" case, Professional has the most top count and almost three times as many as the second (I am thinking Professional is not exactly accurate, because many occupations can be Professional). The count of borrowers on the top three are Professional, Computer Programer and Executive, at the bottom of the list are the students.

How is the prosper rating ditribution?

The prosper rating appears Binomial distribution, 57% of borrowers are labeled middle risk B,C and D. Reference Borrowers with rating C is the highest frequency up to 22%.

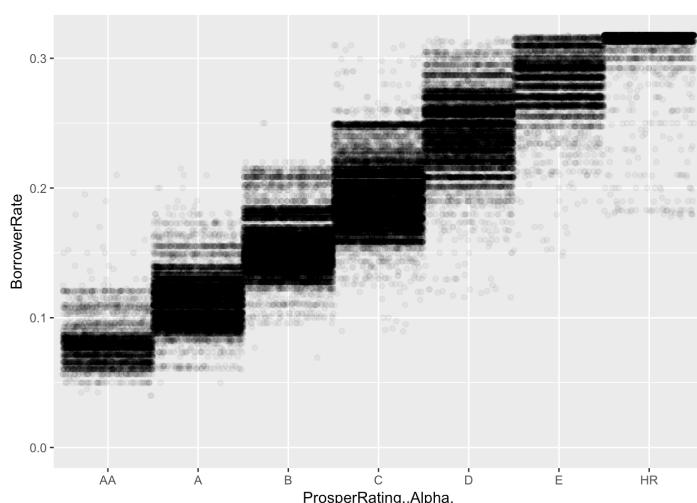
How is the borrowers' annual incomes?

The borrowers' income ranges appears Binomial distribution. 78% of borrowers' annual income are fall into \$25,000 - 75,000. and there are 621 borrowers with zero incomes they got loans, further exploring the borrowers whose incomes are zero, only find 11 borrowers are Not employed while they creating list on prosper, I guess it should be data issue on the person with zero annual income.

Why people need money?

Out of the listing category, three categories comes out as the highest past 10000 loans. There's N/A and Other categories, so we can't know for sure the specific category. But one comes out highest which is category 1 - Debt Consolidation, where's one take out a loan to pay many others. This comes really high with 58308 loans, it could be that many Prosper visitors comes with already have loans, and want to search some loans to pay for it?

Bivariate Plots Section



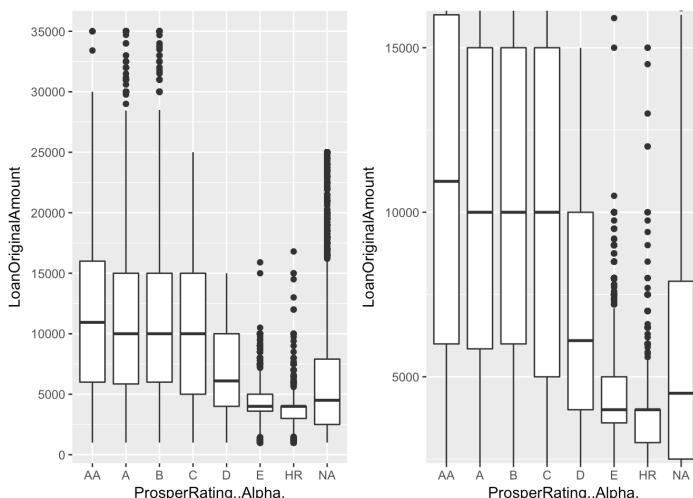
```

## Reference Pearson's product-moment correlation
##
## data: as.numeric(ProsperRating..Alpha.)
## and BorrowerRate
## t = 917.37, df = 84851, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9524846 0.9537172
## sample estimates:
## cor
## 0.9531049

```

AA LOWEST risk-> HR HIGHEST risk

We see that the higher risk labeled to borrowers, the higher borrower rate they would cost. And we see that it has a strong positive linear relationship. This is can be explained as the interest rate is the one where the loaner has benefit from the money that he/she loans.

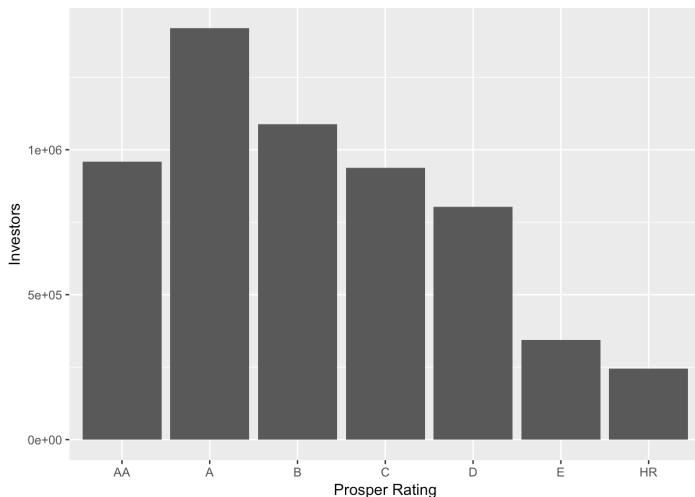


```

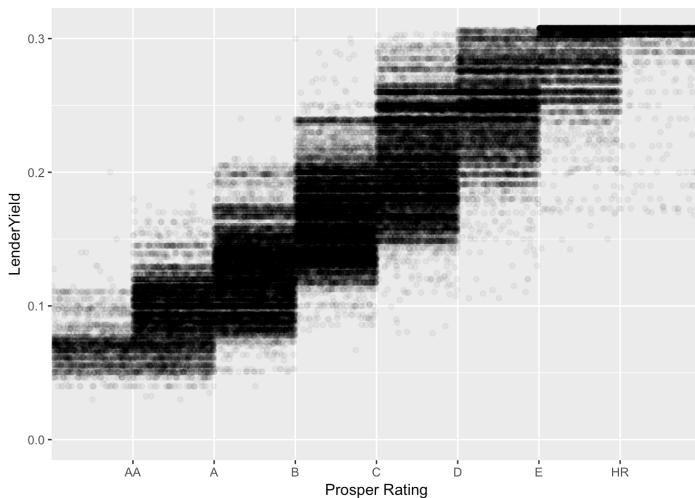
## Pearson's product-moment correlation
##
## data: as.numeric(ProsperRating..Alpha.)
## and LoanOriginalAmount
## t = -138.17, df = 84851, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.4340341 -0.4230486
## sample estimates:
## cor
## -0.4285572

```

It shows rating E and HR has smaller IQR compared to the rest of the rating, rating B and C has no outlier, rating A,B has a smaller outlier, rating AA has two outline only. With a correlation of -0.43, it shows a negative moderately strong linear relationship, as the risk rating increase, the lower amount would be borrowed from prosper.



The distribution has a bell shape and is right-skewed, we can see that investors prefer to loan to those borrowers holding lower risk rating. This is understandable since loan to those person with higher risk may be higher loss the money, even though the lenders would have higher yield by lending to higher risk case (see the next plot).

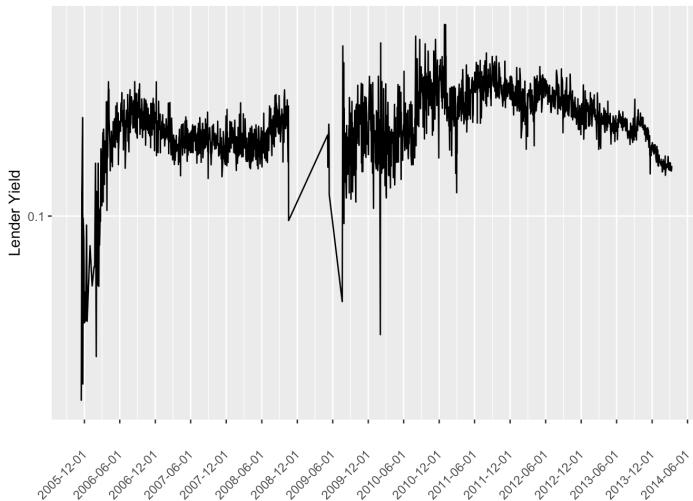


```

## Reference Pearson's product-moment correlation
##
## data: as.numeric(ProsperRating..Alpha.
## ) and LenderYield
## t = 917.52, df = 84851, p-value < 2.2e-
## 16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.9524993 0.9537315
## sample estimates:
## cor
## 0.9531194

```

We see that it has a strong positive linear relationship between risk rating and lenderYield, that means that if the lenders want to have higher yield, they would take on more risk of lose rate.



The LeanderYield quickly increase in the initial 1-year, then in a stable line after 2006. But the thing that surprised me is there was no LenderYield from Q3 2008 to Q2 2009. Then I searched the following from wikipedia.

Reference
On November 24, 2008, the SEC found Prosper to be in violation of the

Securities Act of 1933. As a result of these findings, the SEC imposed a cease and desist order on Prosper.

Due primarily to the novel nature of the peer-to-peer lending models, the SEC, after review, now treats all peer-to-peer lending transactions as sales of securities and requires that all platforms register with the SEC.

In July 2009, Prosper reopened their website for lending (“investing”) and borrowing after having obtained SEC registration for its loans. Hence, that can explain why there were a continuous gap period on lender yield.

Bivariate Analysis

There is a continuous gap period Prosper was not allowed to run their business until it had obtained SEC registration for its loans. The lender yield is stale in the years.

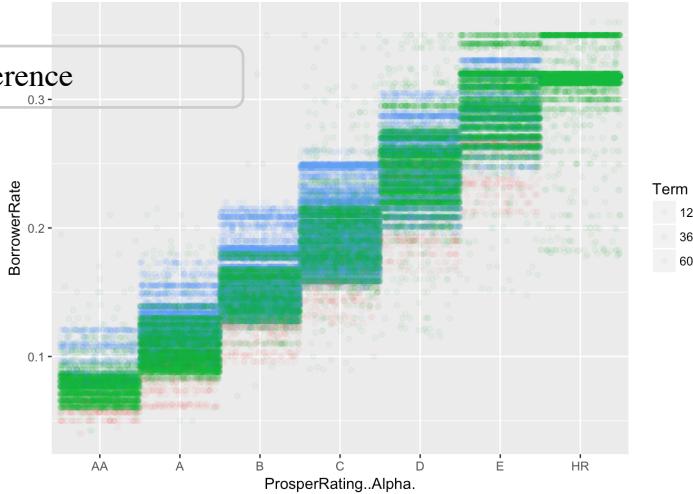
What's the impact after risk rating provided to borrowers?

We observed strong positive/negative linear correlation relationship between prosper risk rating and borrower rate, loan amount can request and lenderYield.

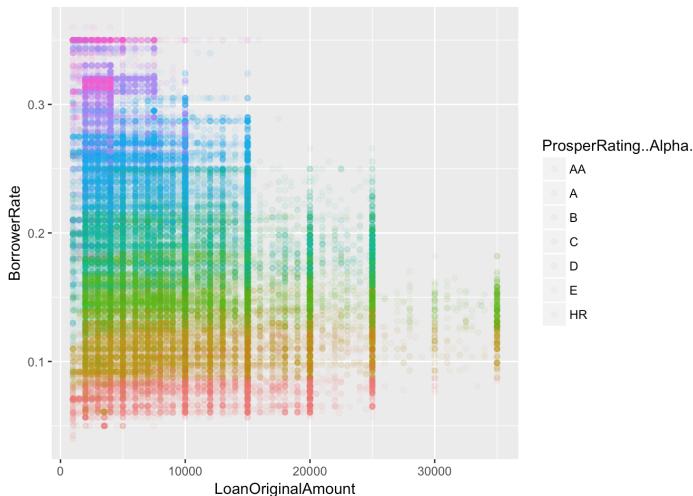
Borrowers holds lower risk rating (AA highest -> HR lowest) in prosper would 1. enjoy lower interest rates. 2. borrow more limit amount. 3. more popular with investors.

Lenders would have higher yield, if they are willing to take on higher risk of lose rate.

Multivariate Plots Section



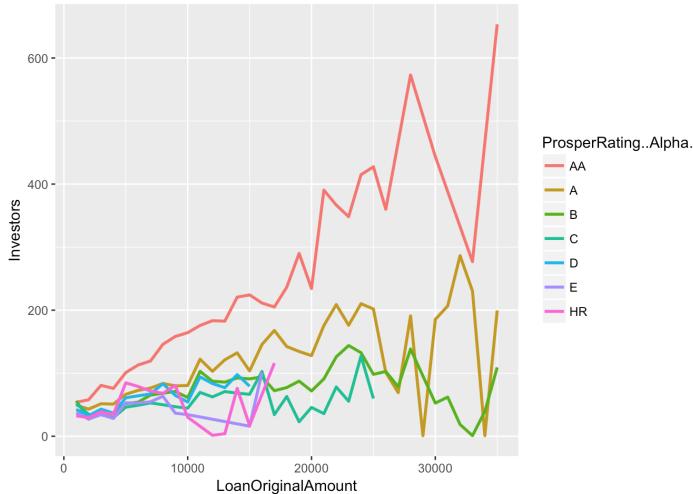
In the previous section we explored there is a strong positive linear correlation relationship between prosper rating and interest rate. here, by adding one more variable Term, we see that for each rating, AA to D, looks the longer loan Term would cause the higher interest rate.



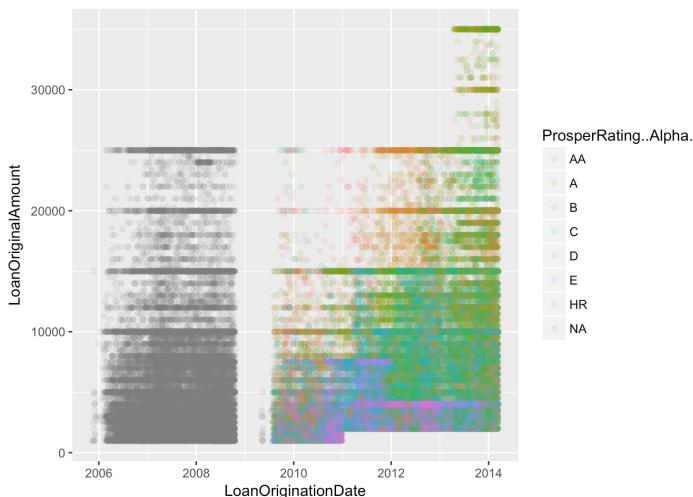
```
## 
## Pearson's product-moment correlation
## 
## data: LoanOriginalAmount and BorrowerRate
## t = -117.58, df = 113940, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.3341283 -0.3237719
## sample estimates:
##       cor
## -0.3289599
```

This scatter plot shows the relationship of Loan amount, borrow rate and prosper rating. Rating AA,A and B looks

can loan amount up to \$35000, rating C less than \$25000, rating D less than \$15000, rating E less than \$10000, and the rating highest risk rating HR less than \$5000, seeing the plot on y-axis, the interest rate is increasing from risk rating AA to HR.



We can see investors are more likely to invest in low risk request, not the amount of the loan, the investors on AA rating looks obviously higher than others.



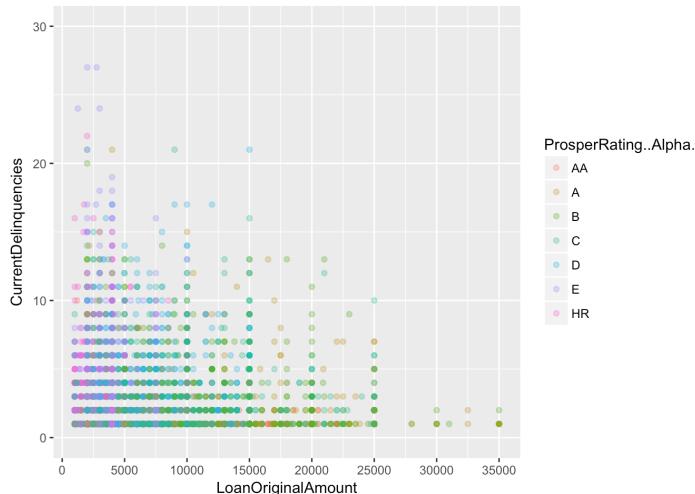
```
## substr(prosper$LoanOriginationQuarter,  
Reference4): Q1 2  
## [1] 35000  
## -----  
-----  
## substr(prosper$LoanOriginationQuarter,  
1, 4): Q2 2  
## [1] 35000  
## -----  
-----  
## substr(prosper$LoanOriginationQuarter,  
1, 4): Q3 2  
## [1] 35000  
## -----  
-----  
## substr(prosper$LoanOriginationQuarter,  
1, 4): Q4 2  
## [1] 35000
```

```
## prosper$ProsperRating..Alpha.: AA  
## [1] 35000  
## -----  
-----  
## prosper$ProsperRating..Alpha.: A  
## [1] 35000  
## -----  
-----  
## prosper$ProsperRating..Alpha.: B  
## [1] 35000  
## -----  
-----  
## prosper$ProsperRating..Alpha.: C  
## [1] 25000  
## -----  
-----  
## prosper$ProsperRating..Alpha.: D  
## [1] 15000  
## -----  
-----  
## prosper$ProsperRating..Alpha.: E  
## [1] 15900  
## -----  
-----  
## prosper$ProsperRating..Alpha.: HR  
## [1] 16800
```

This plot shows that the maximum loan amount in prosper website is \$15000 in initial year 2015, \$25000 before Q2 2013 and \$35000 after that. There should be no business run from Q3 2008 to Q2 2009 on prosper, after colored by risk rating, we also found there is no

risk rating provided to borrowers, all of them are NA.

The risk rating is implemented after Prosper reopened Reference from Q2 2009. The rating D, E, HR can borrow maximum amount around \$15000, rating C is up to \$25000, all rating AA, A, B are up to maximum loan amount \$35000.



We can see plot are left skewed. it shows that rating E,HR borrowed less amount, but they have more delinquencies. while the credit rating increasing, the fewer number of delinquencies come out, even though the borrowers labeled lower risk borrowed more money.

Multivariate Analysis

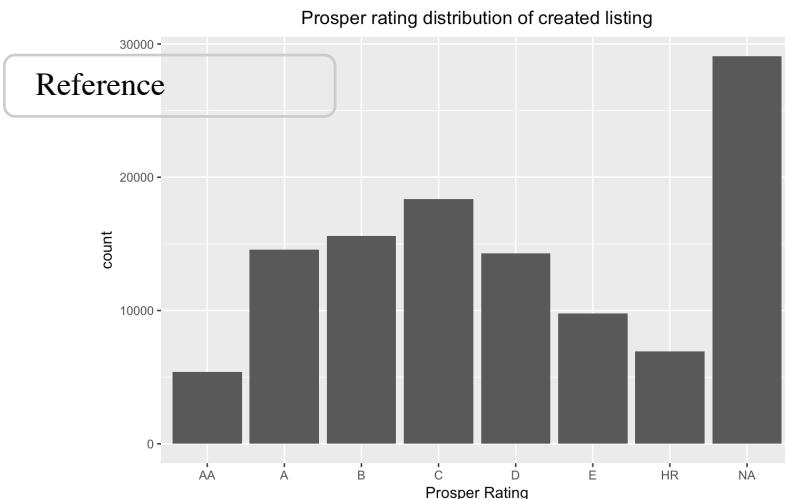
As the loan term and risk rating increase, the interest rate will increase also, there is a strong positive linear correlation relationship between prosper rating and interest rate. if we see the loan amount and risk rating, we found the higher risk rating lead borrowers have to pay more interest rate, and borrowers can have less loan amount also.

Investors are also more likely to invest their money to those borrowers they have lower risk and good credit rating, because they don't want to lost their money, but still can have yield from this.

Further study the current delinquencies of borrowers , that also explains that high-risk borrowers generally have a higher default rate.

Final Plots and Summary

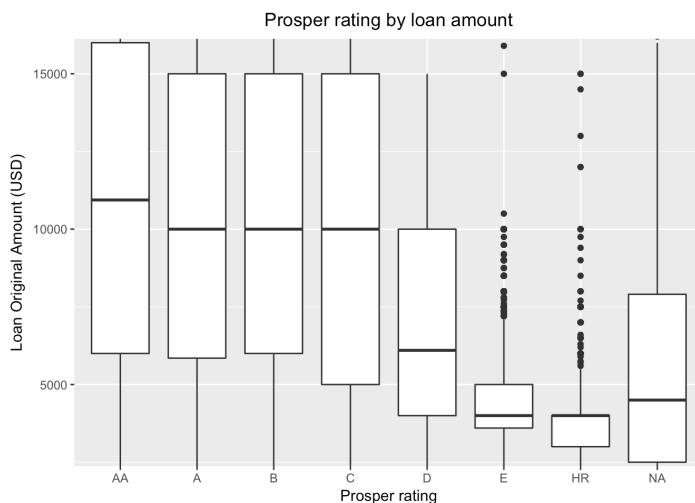
Plot One



Description One

There are almost 30000 loan listing created before Q2 2009 are not have rating provided. The risk rating from the highest AA to the lowers HR appears Binomial distribution, 57% of borrowers are labeled middle risk B,C and D. Borrowers with rating C with 22% the highest frequency.

Plot Two



Description Two

The rating E and HR has smaller IQR compared to the rest of the rating, rating B and C has no outlier, rating A,B has a smaller outlier, rating AA has two outline only. With a correlation of -0.43, it shows a negative moderately strong linear relationship, as the risk rating increase, the lower amount would be borrowed from prosper.

Plot Three



Description Three

There is a strong positive linear correlation relationship 0.95 between prosper rating and interest rate. The fixed (3 Or 5 years) loan Term which borrowers select actually causes different borrower rate on each risk rating.

Reflection

The dataset is from the Prosper loans, it's America's first peer-to-peer lending marketplace, where people could loans money by listing it in the website. There are 113937 observations and 81 variables in the dataset.

I extract 15 features and explore loan term distribution, the borrowers' Occupation, rating, annual incomes, employmentStatus, and what purpose they loan from Prosper. There almost 80% of the borrowers selected 3 years loan term. I am surprise that there are 1614 listing was created from 2010 to 2013 with 1 year loan term, prosper.com doesn't provide such term for loan. Not sure if prosper opened for this term or what problems caused this data issue. While we seeing Occupation of borrowers, the top 1 is Other, it's understandable because all not in the website selecting list can be other, the second to third are professional and computer programer. The prosper rating are not provided before Q2 2009, Ratings, from lowest-risk to highest-risk, are labeled AA, A, B, C, D, E, and HR ("High Risk"), the distributions of rating follow an almost unimodal distribution, with the peak of 18345. The borrowers incomes range distributions also looks follow an almost unimodal distribution, we noticed that there are 621 loan list with zero annual income, then I further explored their employment status when they creating list, most of them are in employed status, hence, those zero incomes

should be data issue. Finally, I explore the purpose of
borrowers loan from prosper, found the highest category
Reference is 1 - Debt Consolidation, where's one take out a loan to
pay many others. we may guess it could be that many
Prosper visitors comes with already have loans, and
want to search some loans to pay for it.

Prosper rating is a very important feature, looking at the statistics and plot, Prosper Rating is strongly correlated with borrower's interest rate, the number of investors, and the LenderYield. It can be explained that this lower risk rating the borrowers labeled, the lower borrower's interest rate would have, and will attract more investors also. For lenders, if they would like to fund for those borrowers labeled higher risk rating, the higher yield will have, but on other hand, it will increase the money lost rate.

As I only select 15 features for this dataset, there must be additional features also important to explore the prosper loan data, we can involve more features regarding to borrowers, lenders or prosper.org to explore more interesting results.

Reference

Udacity: <https://classroom.udacity.com/nanodegrees/nd002-cn-advanced/parts/7f46cd58-8041-4d9d-88a5-4b7c6f7be63e> (<https://classroom.udacity.com/nanodegrees/nd002-cn-advanced/parts/7f46cd58-8041-4d9d-88a5-4b7c6f7be63e>)
wikipedia: https://en.wikipedia.org/wiki/Prosper_Marketplace (https://en.wikipedia.org/wiki/Prosper_Marketplace)
ggplot: <http://ggplot2.org/> (<http://ggplot2.org/>)
Variable Definitions: <https://docs.google.com/document/d/1qEcwlBMIRYZT-l699-71TzlnWfk4W9q5rTCSvDVMpc/pub?embedded=true> (<https://docs.google.com/document/d/1qEcwlBMIRYZT-l699-71TzlnWfk4W9q5rTCSvDVMpc/pub?embedded=true>)

Reference