

Prosper Loan by Michael Roker

This is an exploration of a dataset which approximately 113,937 records of personal loans taken from Prosper, a loan institution.

Univariate Plots Section

```
## 'data.frame': 113937 obs. of 81 variables:
## $ ListingKey : Factor w/ 113066 levels "00003546482094282EF90E5",...: 7180 7...
## $ ListingNumber : int 193129 1209647 81716 658116 909464 1074836 750899 76819...
## $ ListingCreationDate : Factor w/ 113064 levels "2005-11-09 20:44:28.847000000",...: 1...
## $ CreditGrade : Factor w/ 9 levels "", "A", "AA", "B", ...: 5 1 8 1 1 1 1 1 1 1 ...
## $ Term : int 36 36 36 36 36 60 36 36 36 36 ...
## $ LoanStatus : Factor w/ 12 levels "Cancelled", "Chargedoff", ...: 3 4 3 4 4 4 ...
## $ ClosedDate : Factor w/ 2803 levels "", "2005-11-25 00:00:00", ...: 1138 1 12...
## $ BorrowerAPR : num 0.165 0.12 0.283 0.125 0.246 ...
## $ BorrowerRate : num 0.158 0.092 0.275 0.0974 0.2085 ...
## $ LenderYield : num 0.138 0.082 0.24 0.0874 0.1985 ...
## $ EstimatedEffectiveYield : num NA 0.0796 NA 0.0849 0.1832 ...
## $ EstimatedLoss : num NA 0.0249 NA 0.0249 0.0925 ...
## $ EstimatedReturn : num NA 0.0547 NA 0.06 0.0907 ...
## $ ProsperRating..numeric. : int NA 6 NA 6 3 5 2 4 7 7 ...
## $ ProsperRating..Alpha. : Factor w/ 8 levels "", "A", "AA", "B", ...: 1 2 1 2 6 4 7 5 3 ...
## $ ProsperScore : num NA 7 NA 9 4 10 2 4 9 11 ...
## $ ListingCategory..numeric. : int 0 2 0 16 2 1 1 2 7 7 ...
## $ BorrowerState : Factor w/ 52 levels "", "AK", "AL", "AR", ...: 7 7 12 12 25 34 18...
## $ Occupation : Factor w/ 68 levels "", "Accountant/CPA", ...: 37 43 37 52 21 43...
## $ EmploymentStatus : Factor w/ 9 levels "", "Employed", ...: 9 2 4 2 2 2 2 2 2 ...
## $ EmploymentStatusDuration : int 2 44 NA 113 44 82 172 103 269 269 ...
## $ IsBorrowerHomeowner : Factor w/ 2 levels "False", "True": 2 1 1 2 2 2 1 1 2 2 ...
## $ CurrentlyInGroup : Factor w/ 2 levels "False", "True": 2 1 2 1 1 1 1 1 1 1 ...
## $ GroupKey : Factor w/ 707 levels "", "00343376901312423168731", ...: 1 1 33...
## $ DateCreditPulled : Factor w/ 112992 levels "2005-11-09 00:30:04.487000000", ...: 1...
## $ CreditScoreRangeLower : int 640 680 480 800 680 740 680 700 820 820 ...
## $ CreditScoreRangeUpper : int 659 699 499 819 699 759 699 719 839 839 ...
## $ FirstRecordedCreditLine : Factor w/ 11586 levels "", "1947-08-24 00:00:00", ...: 8639 661...
## $ CurrentCreditLines : int 5 14 NA 5 19 21 10 6 17 17 ...
## $ OpenCreditLines : int 4 14 NA 5 19 17 7 6 16 16 ...
## $ TotalCreditLinespast7years : int 12 29 3 29 49 49 20 10 32 32 ...
## $ OpenRevolvingAccounts : int 1 13 0 7 6 13 6 5 12 12 ...
## $ OpenRevolvingMonthlyPayment : num 24 389 0 115 220 1410 214 101 219 219 ...
## $ InquiriesLast6Months : int 3 3 0 0 1 0 0 3 1 1 ...
## $ TotalInquiries : num 3 5 1 1 9 2 0 16 6 6 ...
## $ CurrentDelinquencies : int 2 0 1 4 0 0 0 0 0 0 ...
## $ AmountDelinquent : num 472 0 NA 10056 0 ...
## $ DelinquenciesLast7Years : int 4 0 0 14 0 0 0 0 0 0 ...
## $ PublicRecordsLast10Years : int 0 1 0 0 0 0 0 1 0 0 ...
## $ PublicRecordsLast12Months : int 0 0 NA 0 0 0 0 0 0 0 ...
## $ RevolvingCreditBalance : num 0 3989 NA 1444 6193 ...
## $ BankcardUtilization : num 0 0.21 NA 0.04 0.81 0.39 0.72 0.13 0.11 0.11 ...
## $ AvailableBankcardCredit : num 1500 10266 NA 30754 695 ...
## $ TotalTrades : num 11 29 NA 26 39 47 16 10 29 29 ...
```

```

## $ TradesNeverDelinquent..percentage. : num  0.81 1 NA 0.76 0.95 1 0.68 0.8 1 1 ...
## $ TradesOpenedLast6Months           : num  0 2 NA 0 2 0 0 0 1 1 ...
## $ DebtToIncomeRatio               : num  0.17 0.18 0.06 0.15 0.26 0.36 0.27 0.24 0.25 0.25 ...
## $ IncomeRange                     : Factor w/ 8 levels "$0","$1-24,999",...: 4 5 7 4 3 3 4 4 4 4
## $ IncomeVerifiable               : Factor w/ 2 levels "False","True": 2 2 2 2 2 2 2 2 2 2 ...
## $ StatedMonthlyIncome             : num  3083 6125 2083 2875 9583 ...
## $ LoanKey                         : Factor w/ 113066 levels "00003683605746079487FF7",...: 100337
## $ TotalProsperLoans              : int  NA NA NA NA 1 NA NA NA NA ...
## $ TotalProsperPaymentsBilled     : int  NA NA NA NA 11 NA NA NA NA ...
## $ OnTimeProsperPayments          : int  NA NA NA NA 11 NA NA NA NA ...
## $ ProsperPaymentsLessThanOneMonthLate: int  NA NA NA NA 0 NA NA NA NA ...
## $ ProsperPaymentsOneMonthPlusLate: int  NA NA NA NA 0 NA NA NA NA ...
## $ ProsperPrincipalBorrowed       : num  NA NA NA NA 11000 NA NA NA NA ...
## $ ProsperPrincipalOutstanding    : num  NA NA NA NA 9948 ...
## $ ScorexChangeAtTimeOfListing   : int  NA NA NA NA NA NA NA NA ...
## $ LoanCurrentDaysDelinquent     : int  0 0 0 0 0 0 0 0 0 ...
## $ LoanFirstDefaultedCycleNumber: int  NA NA NA NA NA NA NA NA ...
## $ LoanMonthsSinceOrigination    : int  78 0 86 16 6 3 11 10 3 3 ...
## $ LoanNumber                      : int  19141 134815 6466 77296 102670 123257 88353 90051 12126 ...
## $ LoanOriginalAmount             : int  9425 10000 3001 10000 15000 15000 3000 10000 10000 ...
## $ LoanOriginationDate            : Factor w/ 1873 levels "2005-11-15 00:00:00",...: 426 1866 260 ...
## $ LoanOriginationQuarter         : Factor w/ 33 levels "Q1 2006","Q1 2007",...: 18 8 2 32 24 33 ...
## $ MemberKey                       : Factor w/ 90831 levels "00003397697413387CAF966",...: 11071 10 ...
## $ MonthlyLoanPayment              : num  330 319 123 321 564 ...
## $ LP_CustomerPayments             : num  11396 0 4187 5143 2820 ...
## $ LP_CustomerPrincipalPayments   : num  9425 0 3001 4091 1563 ...
## $ LP_InterestandFees              : num  1971 0 1186 1052 1257 ...
## $ LP_ServiceFees                 : num  -133.2 0 -24.2 -108 -60.3 ...
## $ LP_CollectionFees              : num  0 0 0 0 0 0 0 0 0 ...
## $ LP_GrossPrincipalLoss           : num  0 0 0 0 0 0 0 0 0 ...
## $ LP_NetPrincipalLoss             : num  0 0 0 0 0 0 0 0 0 ...
## $ LP_NonPrincipalRecoverypayments: num  0 0 0 0 0 0 0 0 0 ...
## $ PercentFunded                  : num  1 1 1 1 1 1 1 1 1 ...
## $ Recommendations                 : int  0 0 0 0 0 0 0 0 0 ...
## $ InvestmentFromFriendsCount    : int  0 0 0 0 0 0 0 0 0 ...
## $ InvestmentFromFriendsAmount    : num  0 0 0 0 0 0 0 0 0 ...
## $ Investors                        : int  258 1 41 158 20 1 1 1 1 ...

## [1] 113937      81

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

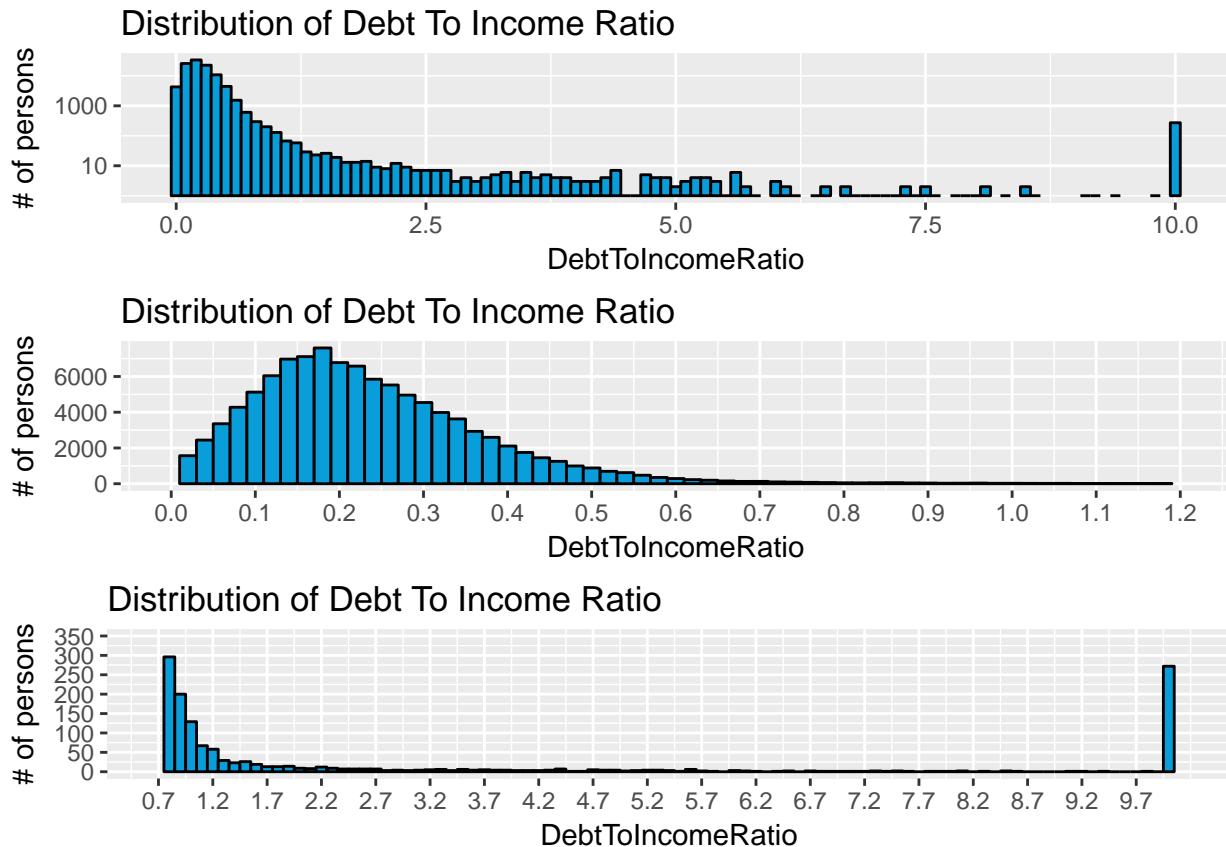
```

As seen above, the dataset has 113,937 observations and 81 variables. All of these variables will not be explored going on this set.

```

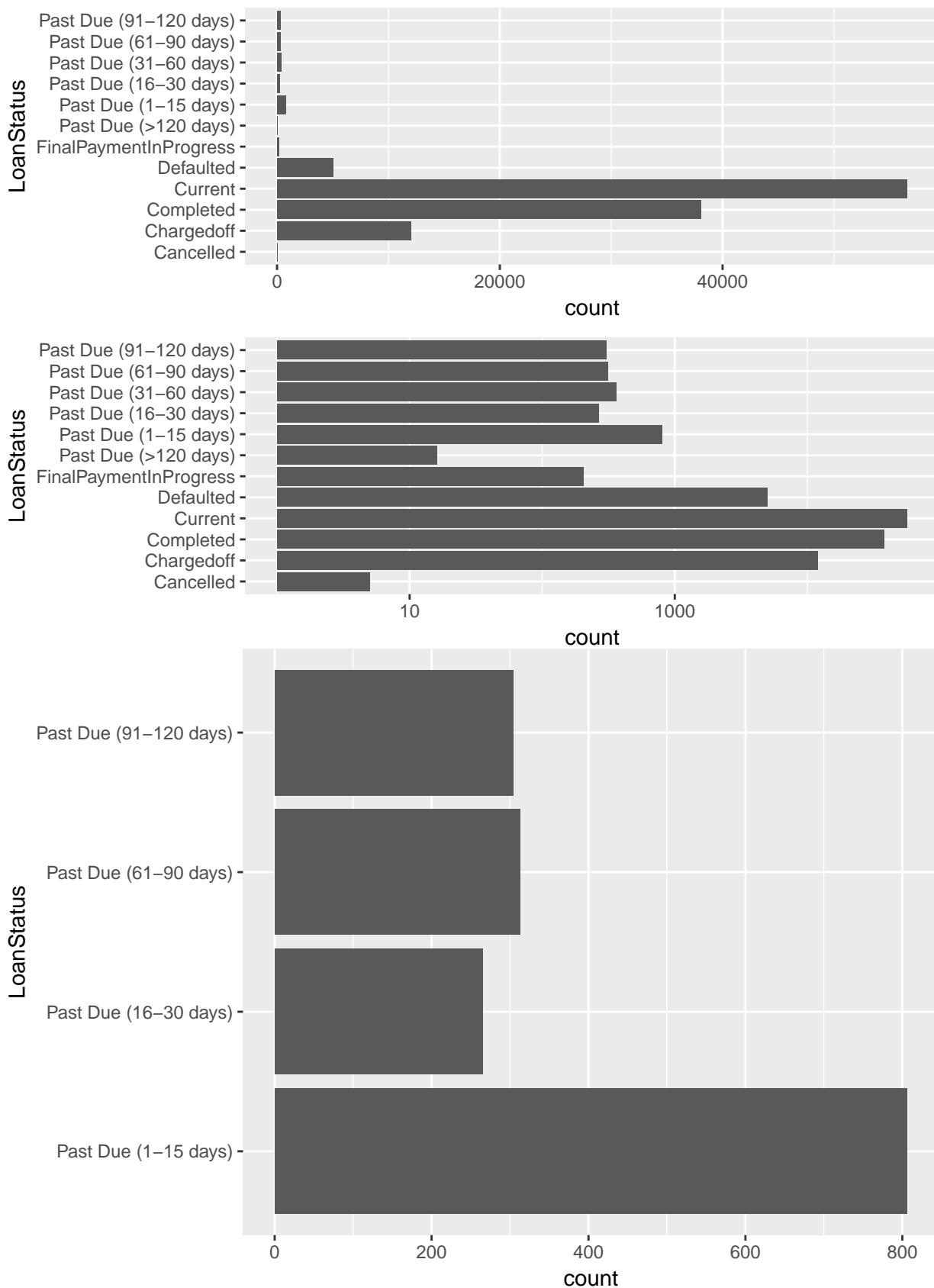
## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
## 0.000 0.140 0.220 0.276 0.320 10.010 8554

```



Plotting the variable `DebtToIncomeRatio`, as shown above for the summary, the mean is 0.320 and the max is 10.010. Also, there are 8554 NA's which in the plots have been omitted. The plots above give a better showing of the summary which shows that the majority of the population are about 0.18. The first plot shows the whole distribution and the second and third zoom in on 0 - 1.2 and 0.7 - 10.1.

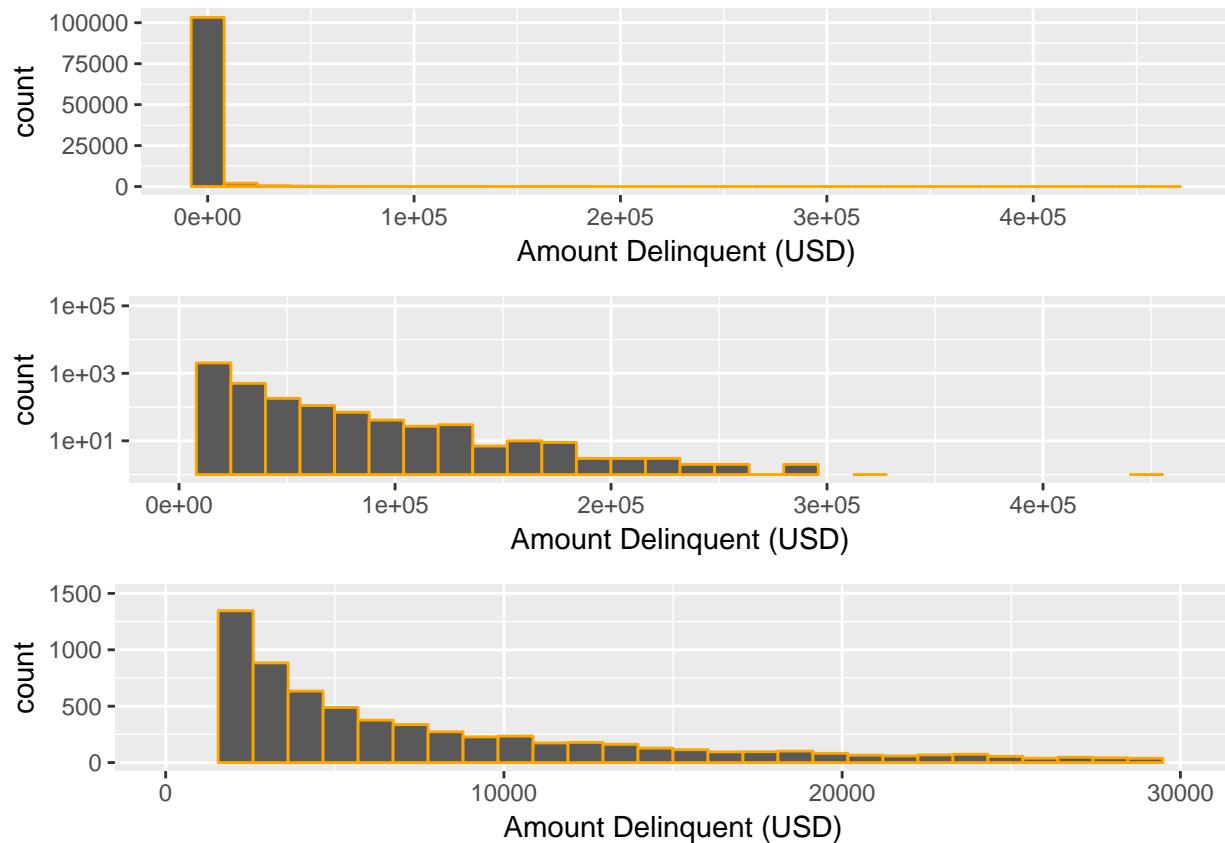
	Cancelled	Chargedoff	Completed
##	5	11992	38074
##	Current	Defaulted	FinalPaymentInProgress
##	56576	5018	205
##	Past Due (>120 days)	Past Due (1-15 days)	Past Due (16-30 days)
##	16	806	265
##	Past Due (31-60 days)	Past Due (61-90 days)	Past Due (91-120 days)
##	363	313	304



This is the loan status variable. Firstly, looking at the different statuses and how they look in the population.

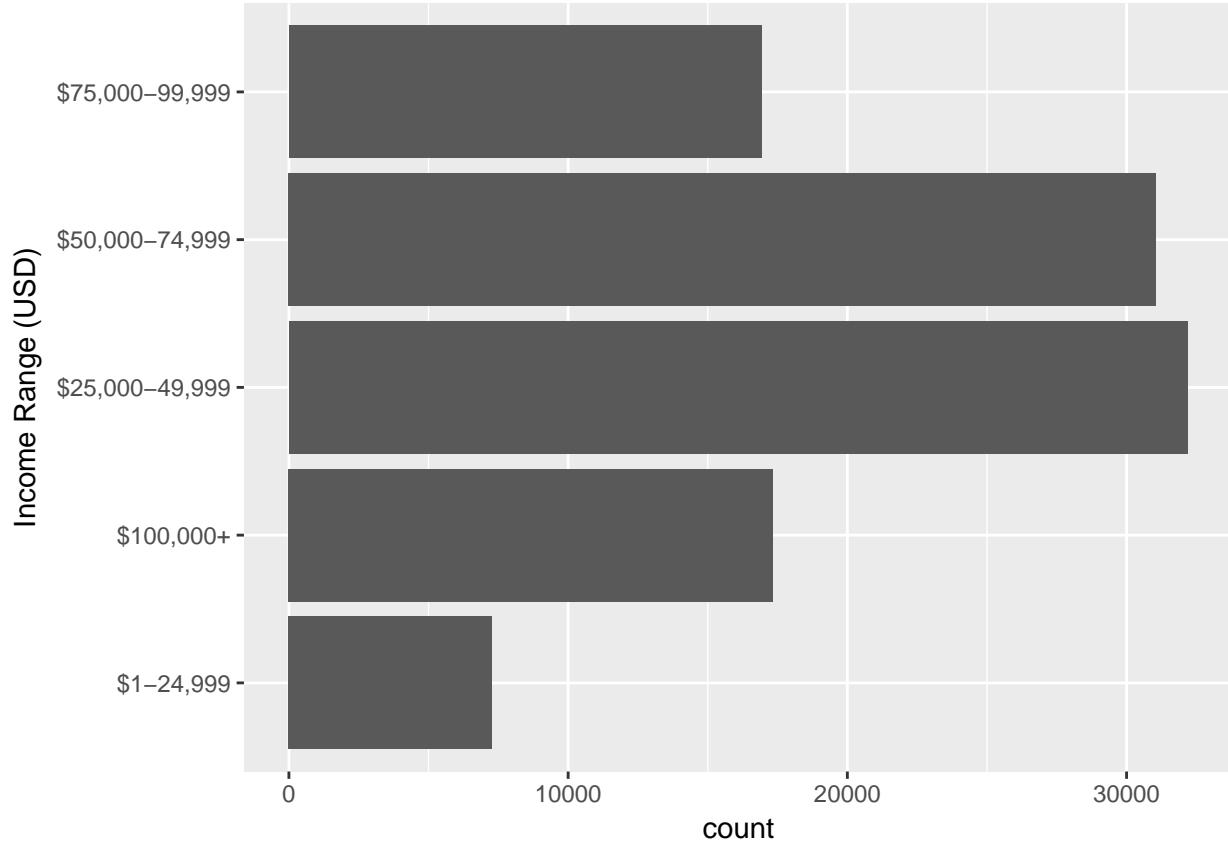
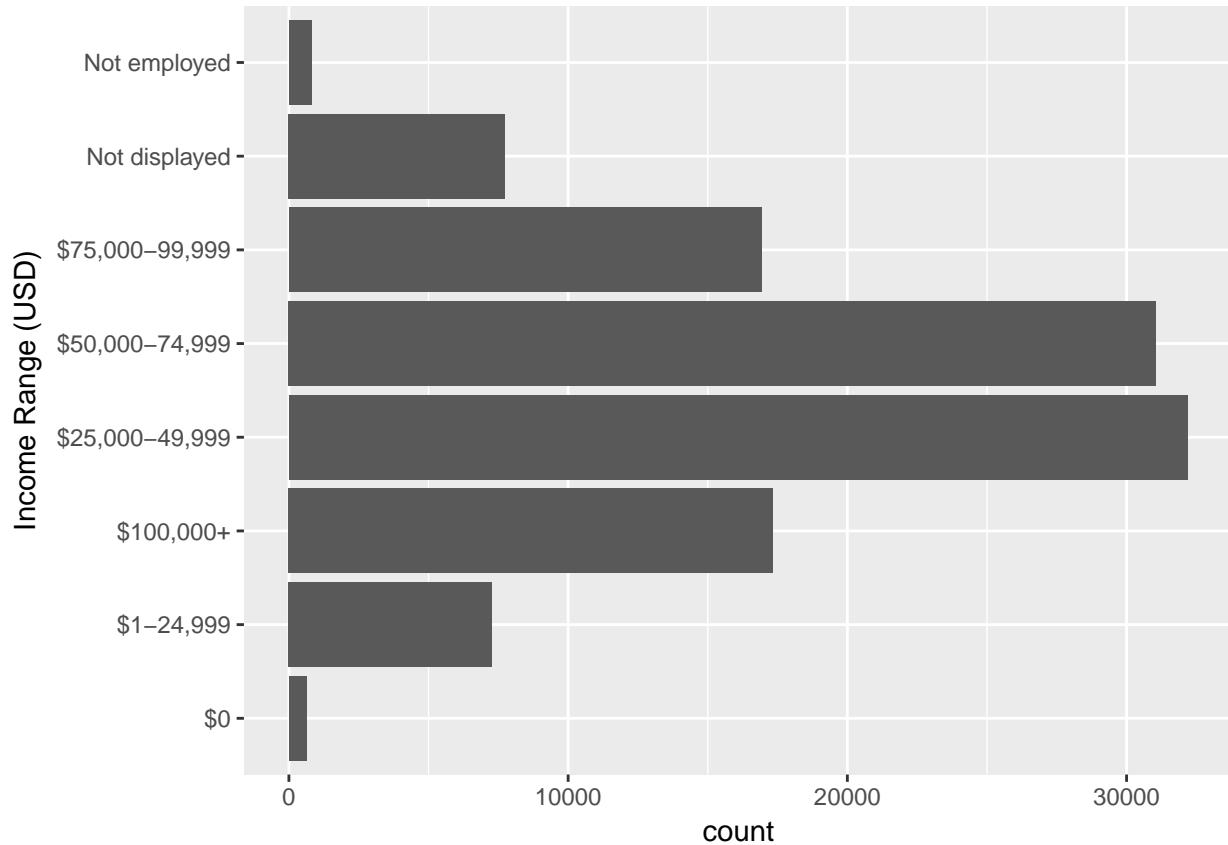
Afterward, because the past due payments were so minimal, I zoomed in to see how they were compared to the other statuses. Lastly, I took a view at only the past due payments variables to see how they compare with each other.

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.     Max.   NA's
##      0.0     0.0     0.0    984.5    0.0 463881.0    7622
```



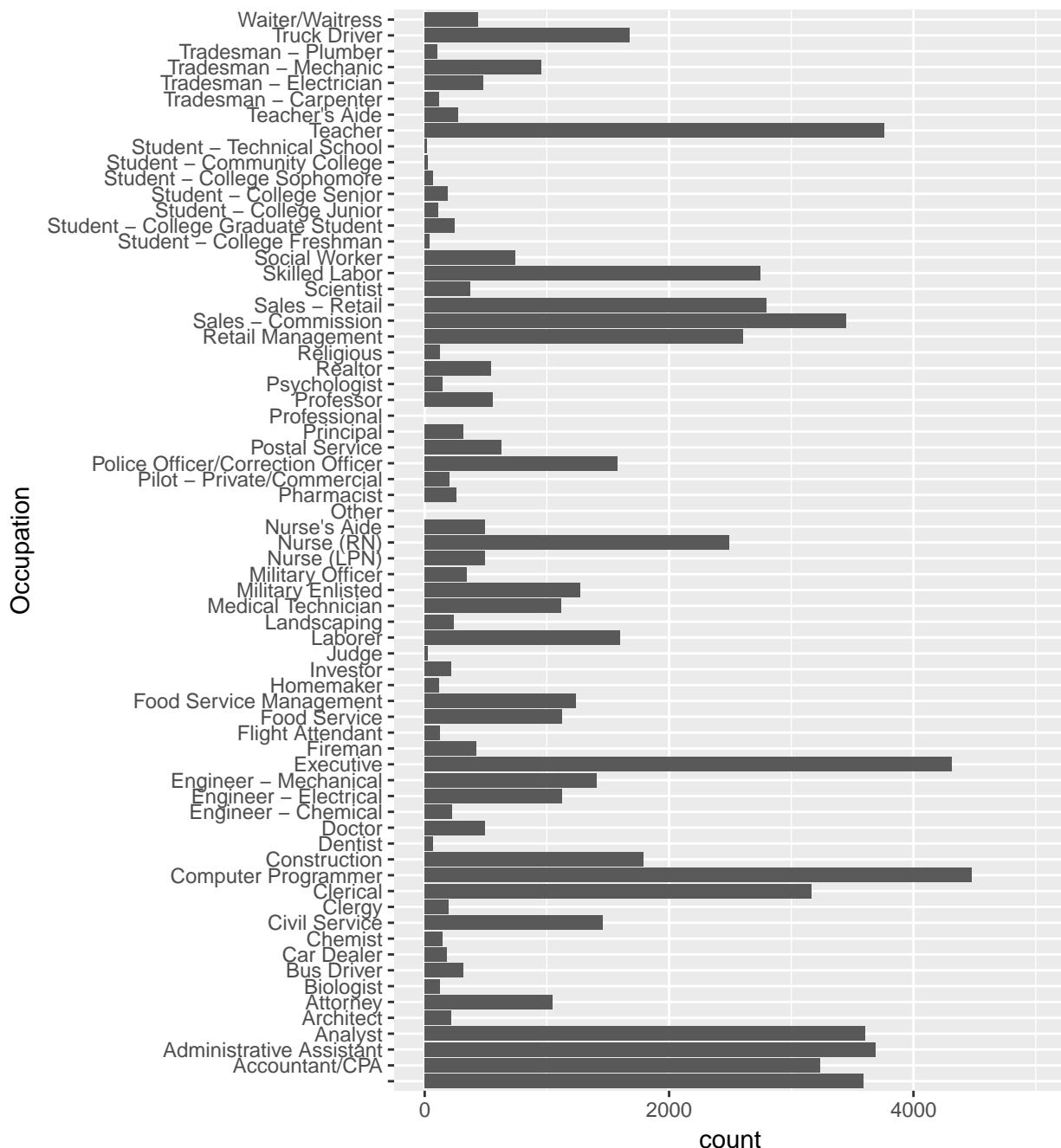
Exploring the Amount Delinquent variable, the mean shows at 984.5. There are a number of NA's so those were omitted in the above plots. For the first plot in the grid, it shows that the majority is nearer to 0 and that the max is an outlier. For the second plot, the y axis was scaled up to take a closer look at the counts. For the third, the Amounts were scaled to remove the 0's (to look at those that were actually delinquent) and limited x to see the distribution closer.

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



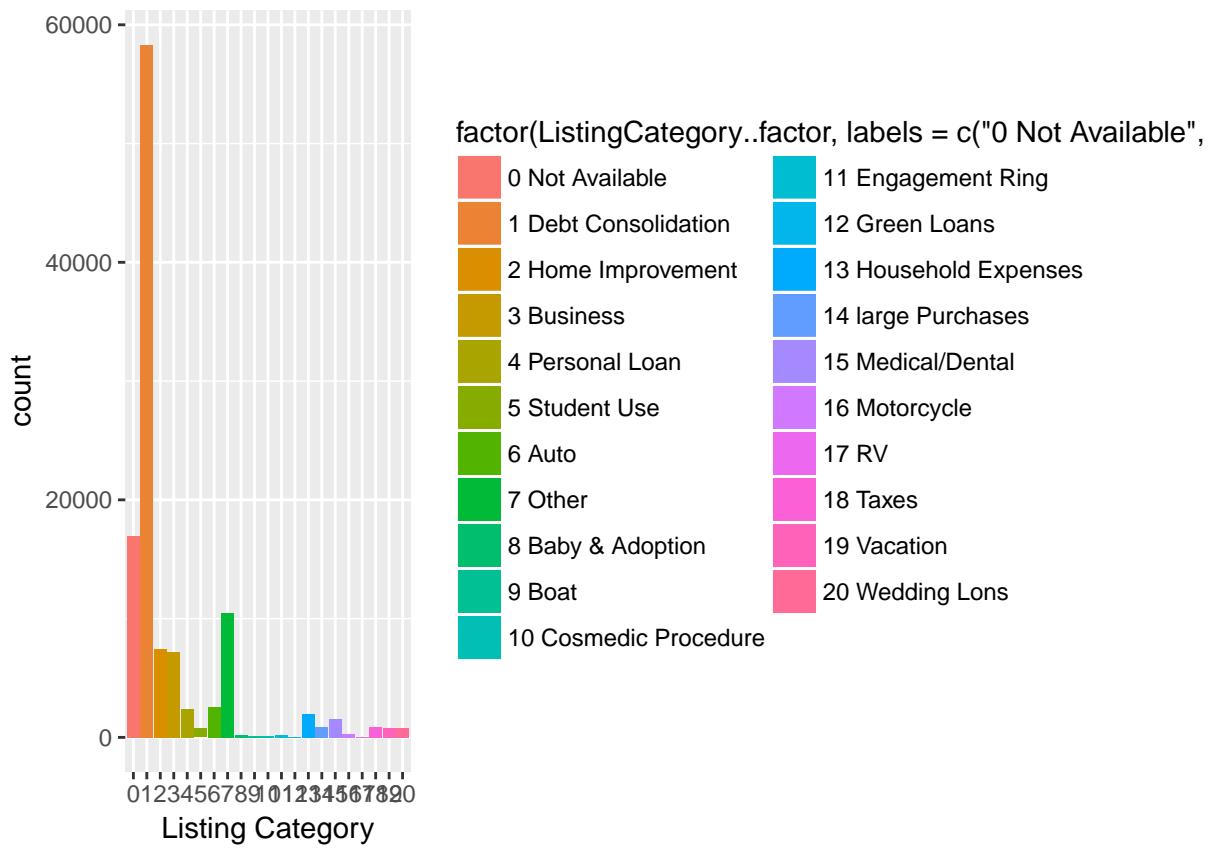
As we can see for the summary, these are the amounts for the Income Range variable. The first plot shows

all of the income ranges. The second removes not displayed, not employeed, and 0.



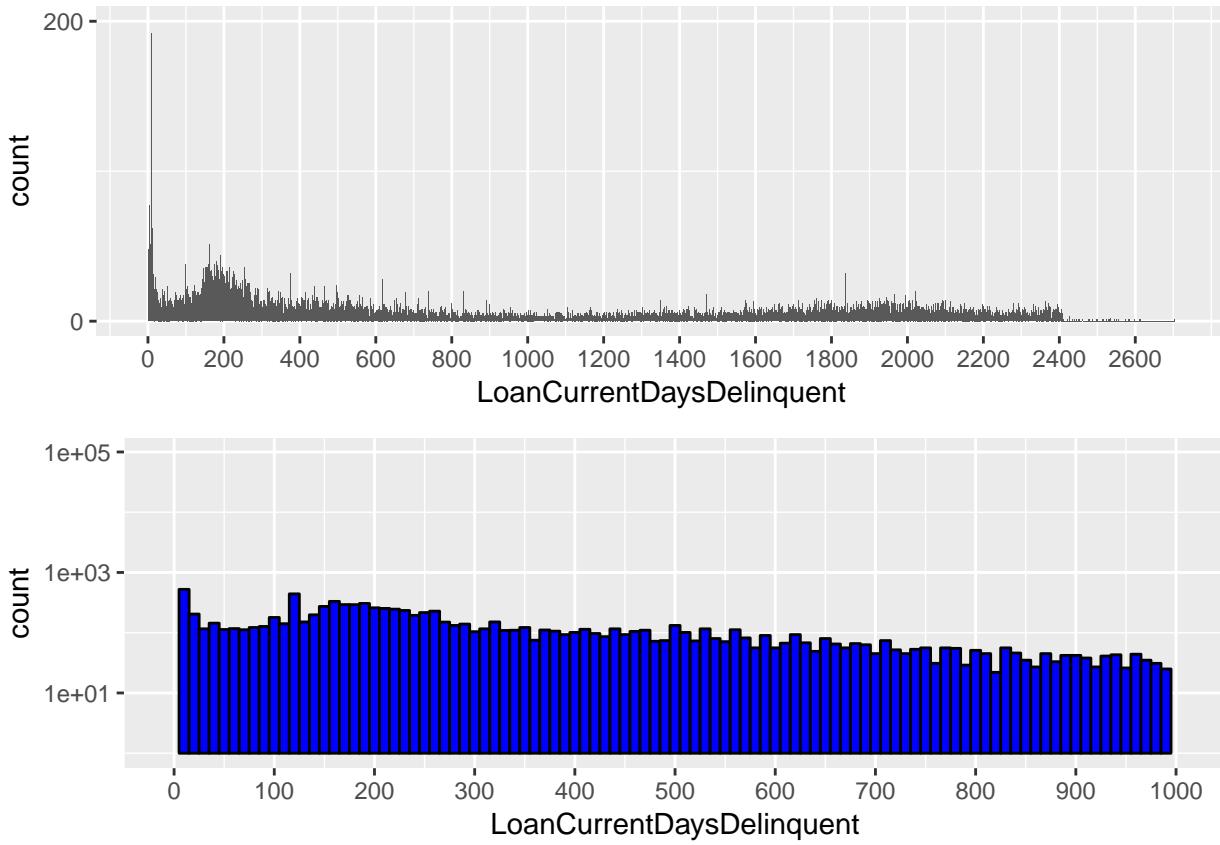
Among the top persons in the dataset are Teacher, Sales Executive, and Computer Programmer.

##	0	1	2	3	4	5	6	7	8	9	10	11
##	16965	58308	7433	7189	2395	756	2572	10494	199	85	91	217
##	12	13	14	15	16	17	18	19	20			
##	59	1996	876	1522	304	52	885	768	771			



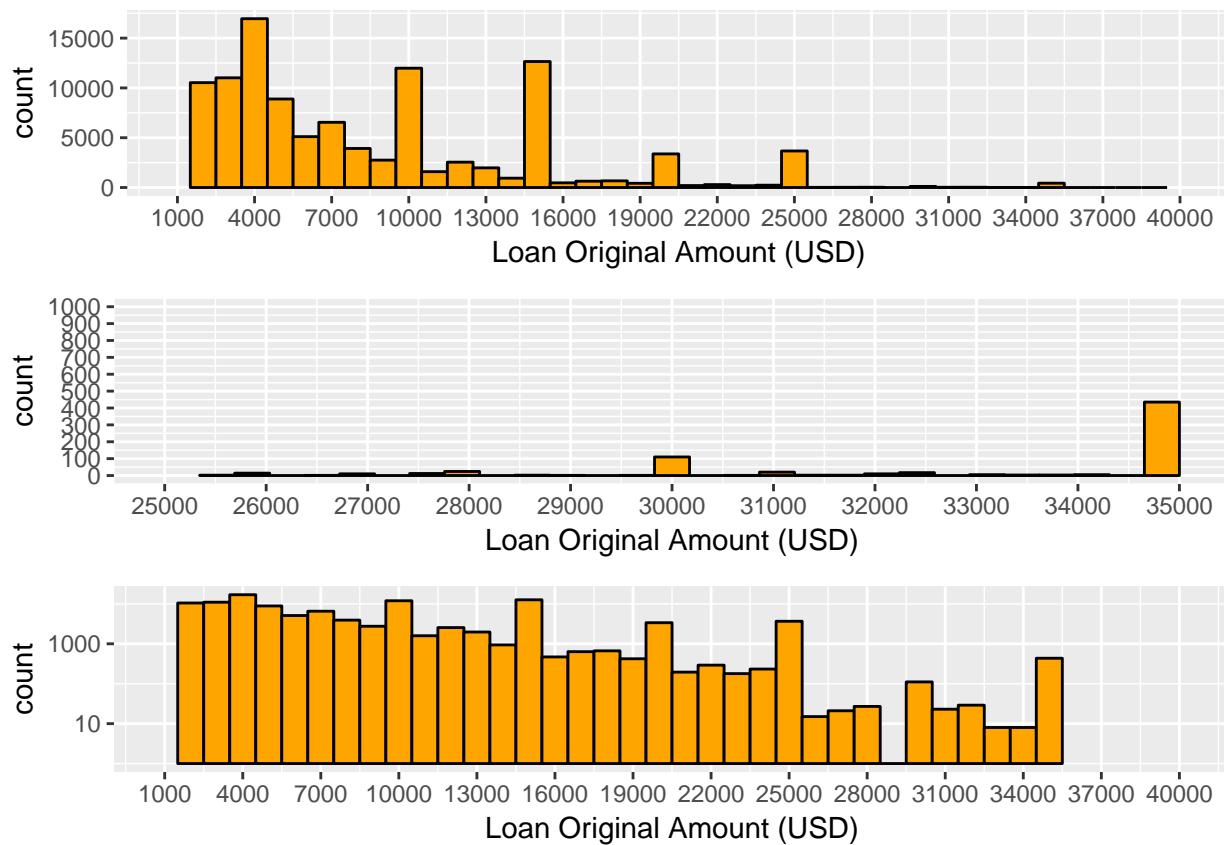
Changed this variable from numeric to category so that it can be used. Created a bar plot to show the amounts for each category. As it can be seen Debt Consolidation is the dominant reason for getting a loan with Prosper.

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##      0.0    0.0    0.0   152.8    0.0  2704.0
```



As seen above for the summary of Days Delinquent for payment, the mean is 152 days. The first plot is the whole population where 2704 is the max. The second plot shows the plot scaled by 10 from 0 to 1000 days delinquent.

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



As shown in the summary above the mean for the amount Lent is 8337. The max lent is 35000. For all plots, changed the binwidth to 1000 because loan amounts are measured by these amounts. In the third plot, scaled by 10. As is shown above most people lent 4000.

Univariate Analysis

Dataset Structure

This dataset has 113,937 observations along with 81 columns. The variables being observed are:

- Debt to Income Ratio
- Loan Status
- Amount Delinquent
- Income Range
- Occupation
- Listing Category
- Loan Current Days Delinquent
- Loan Original Amount

Features

The most interesting features in this dataset are the Loan Current Days Delinquent and the Loan Current Days Delinquent and if there is any commonalities between other variables and these two.

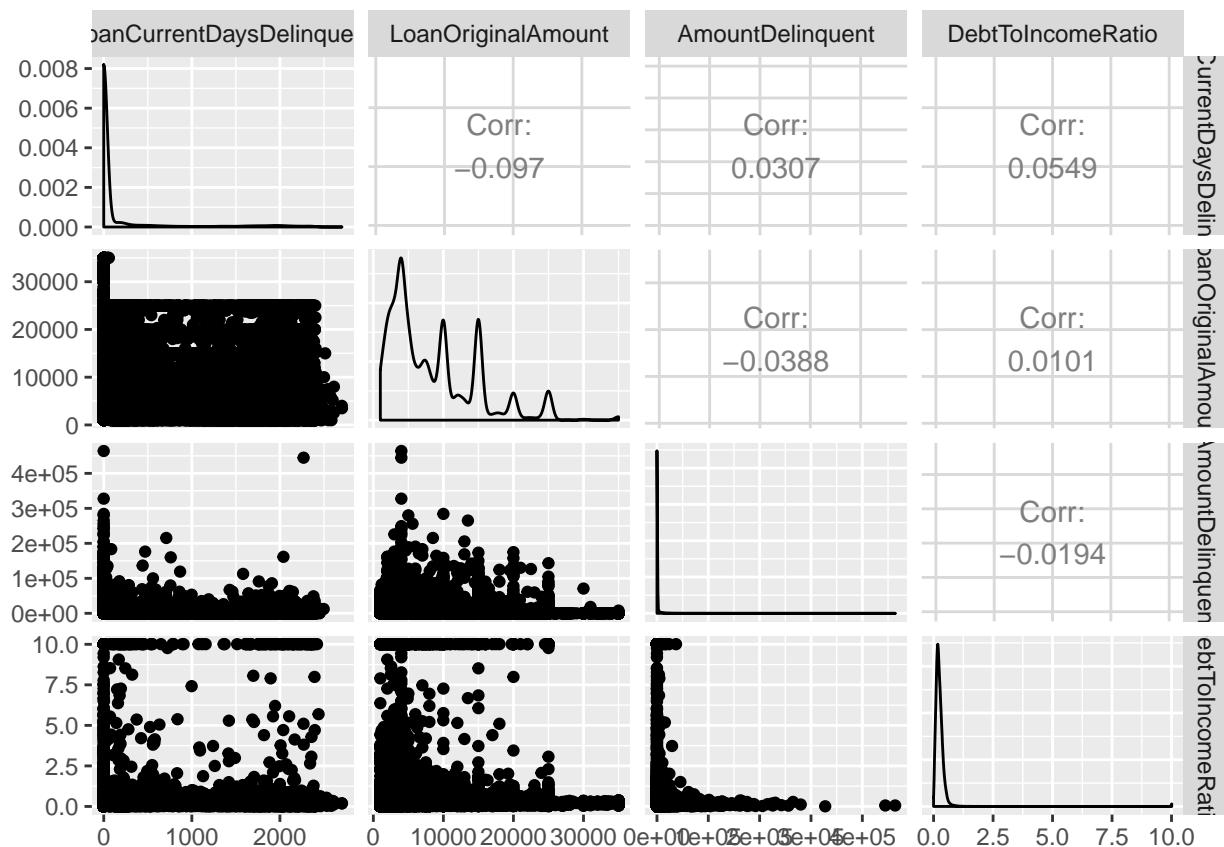
Additional Features

I think that displaying the reasons for their loan, and possibly their loan amounts may give understanding as to who mainly has problems paying their loan, or if there is a trend.

Changes made

The only change that was made was to the Listing Category, because this was in numeric mode and this was changed to a factor so it can be displayed better using a histogram. Also, so that it can be summarized with better clarity.

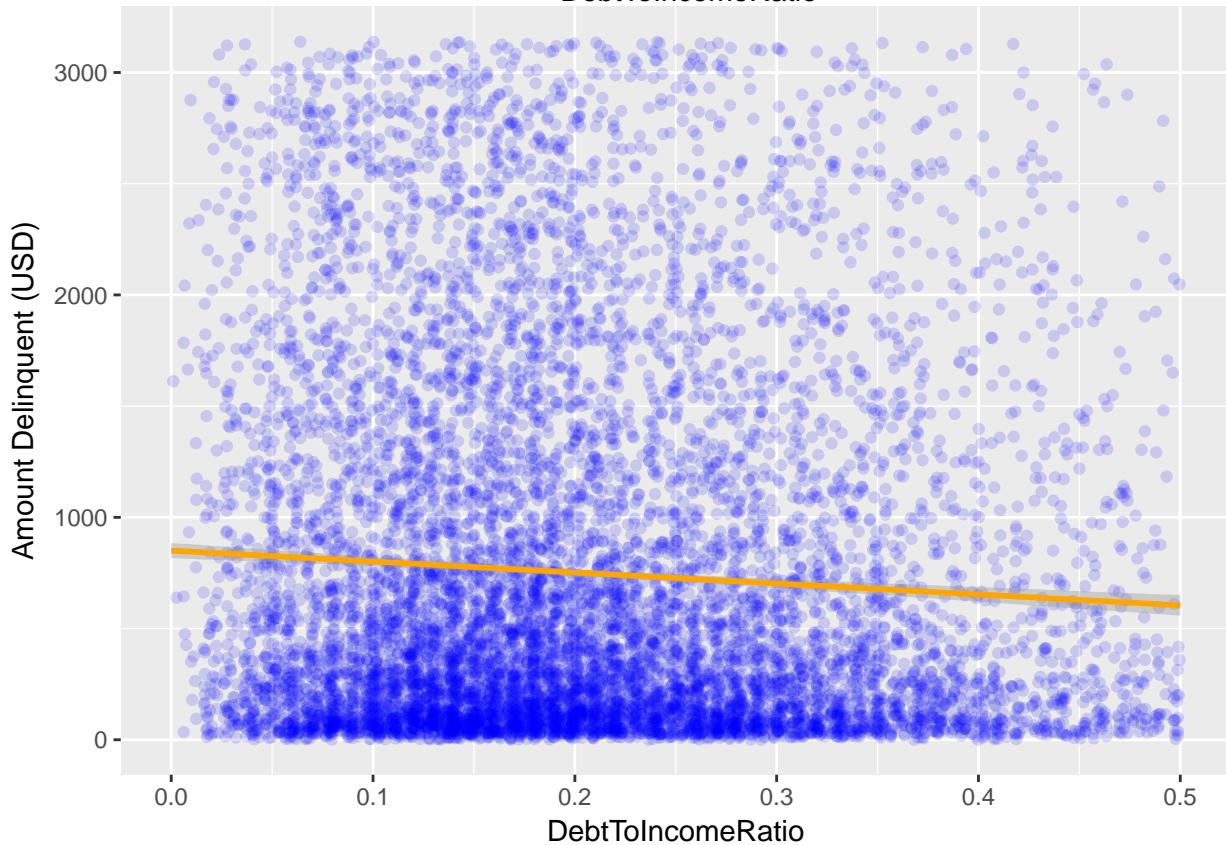
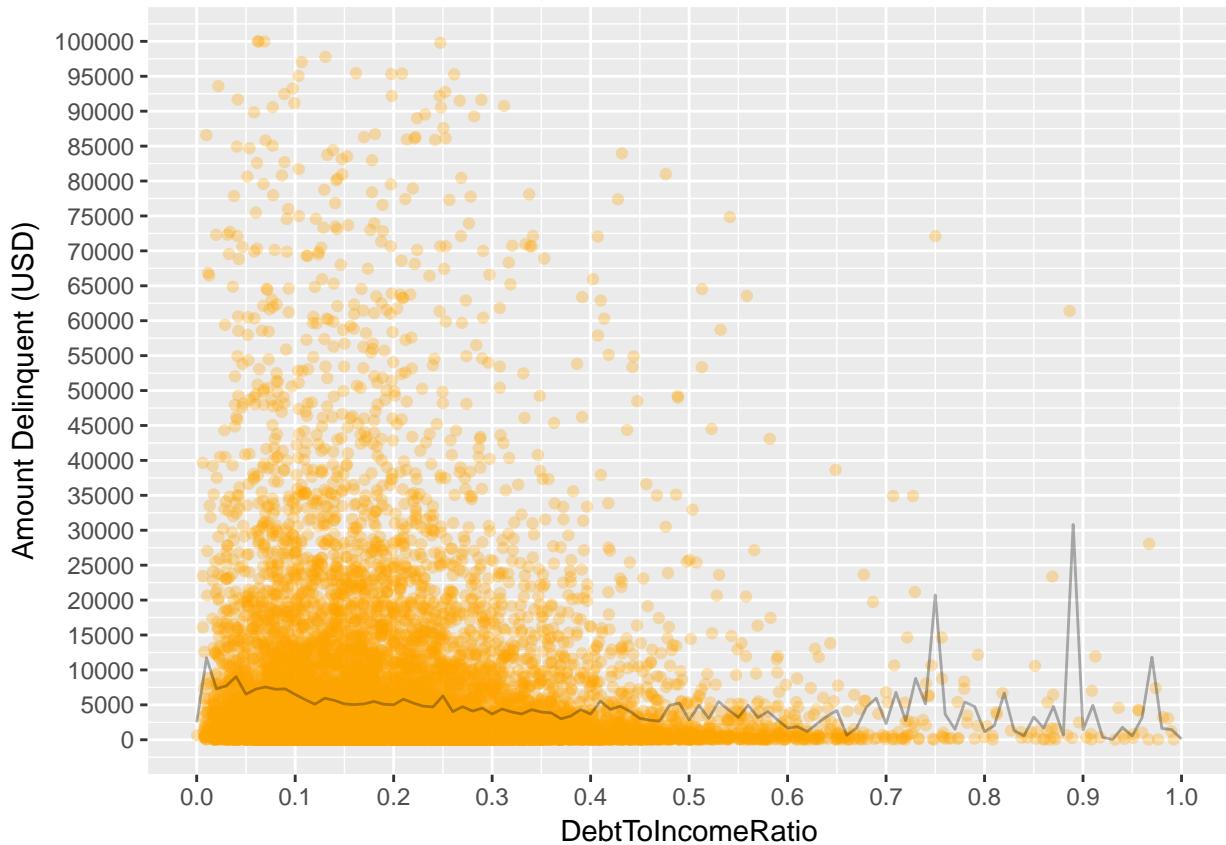
Bivariate Plots Section



This gives a birds eye view of all of the different plots together along with some of the correlations.

```
##  
## Pearson's product-moment correlation  
##  
## data: loandata.filtered$AmountDelinquent and loandata.filtered$DebtToIncomeRatio  
## t = -3.6555, df = 15523, p-value = 0.0002576  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.04503658 -0.01360293  
## sample estimates:  
## cor
```

-0.029327

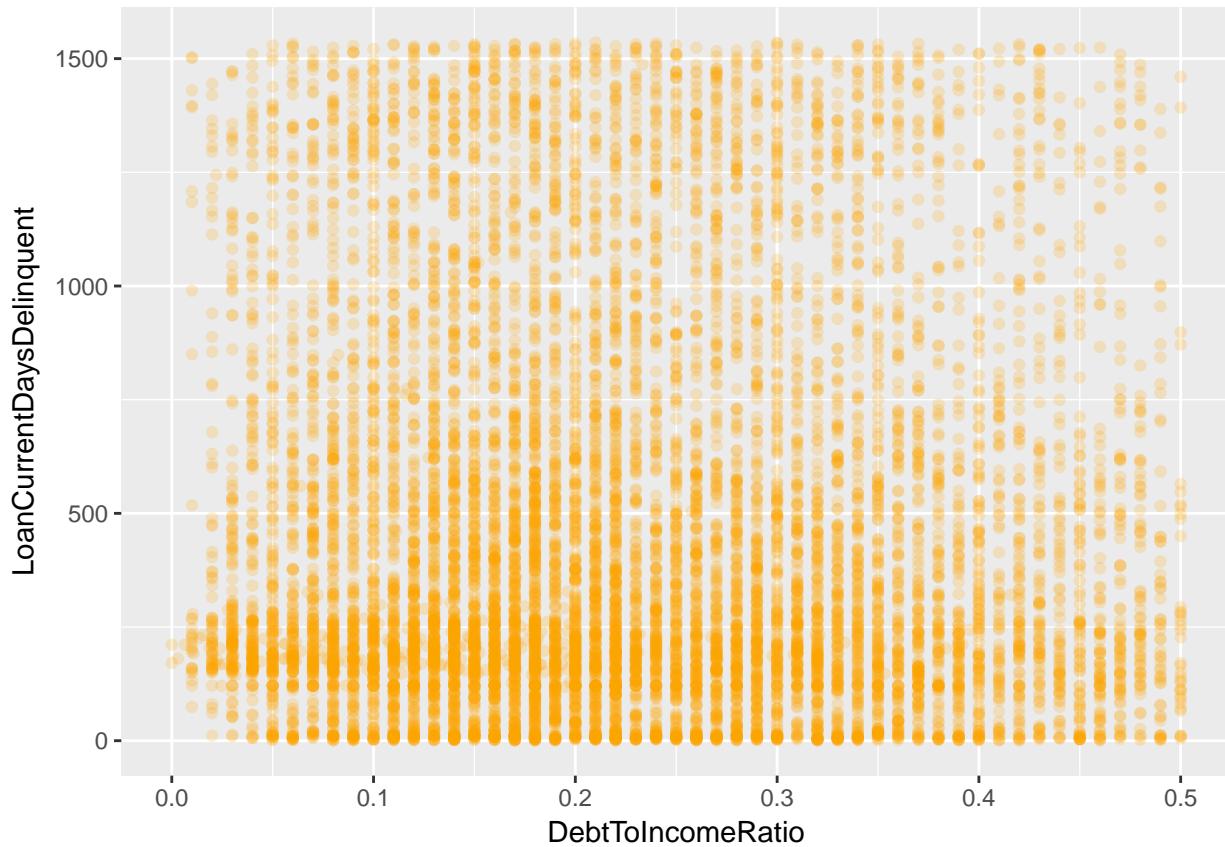


As seen above the correlation test is shown to be very low (-0.02) which can indicate there is little pattern between the two variables.

For the first plot above, this was scaled between 0 and 1.0 seeing how the majority are between this amount. Added a summarized geom line on the plot.

On the second plot, used 95% quantile and added a geom smooth line to show the relationship between Amount Delinquent and Debt To Income Ratio.

```
##  
## Pearson's product-moment correlation  
##  
## data: loandata$DebtToIncomeRatio and loandata$LoanCurrentDaysDelinquent  
## t = 17.856, df = 105380, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.04890099 0.06093975  
## sample estimates:  
## cor  
## 0.05492236
```



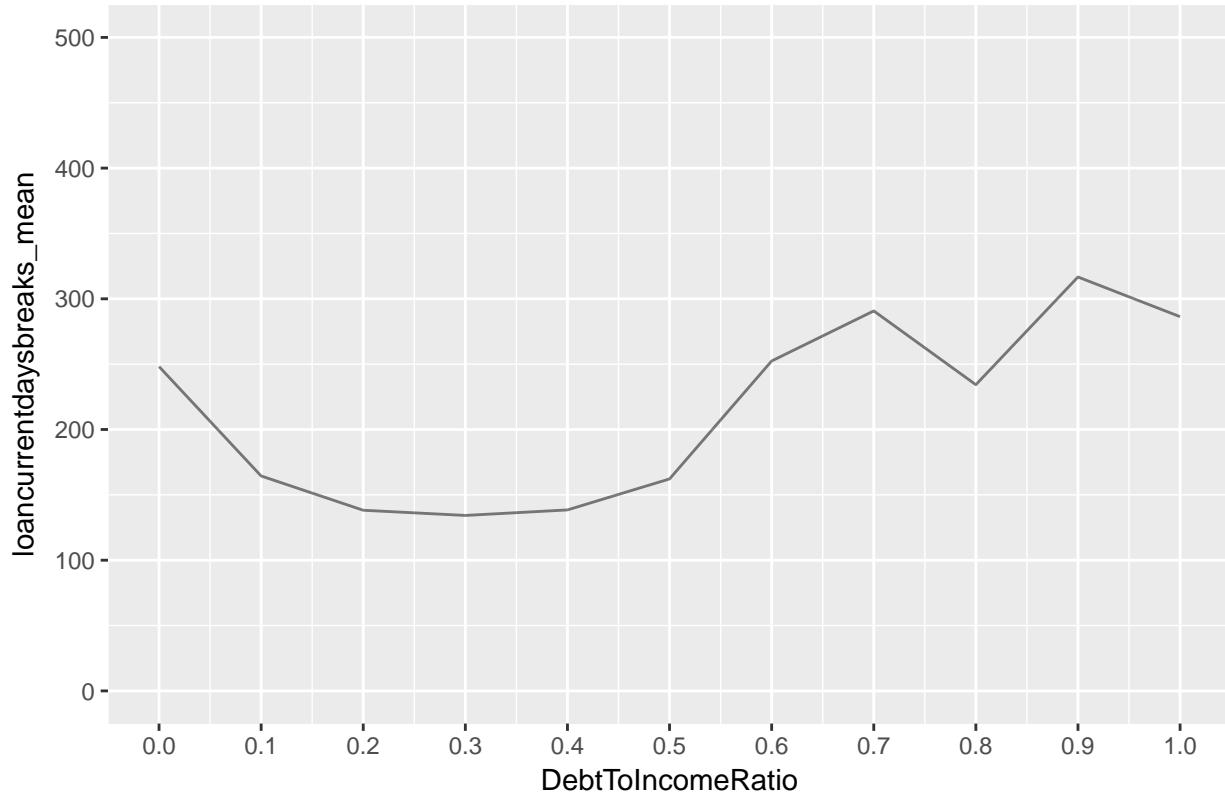
```
## loandata$IncomeRange: $0  
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's  
##      0.0     0.0    0.0   472.9    0.0 33134.0      3  
## -----  
## loandata$IncomeRange: $1-24,999  
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's  
##      0       0     0    658       0 160039      7  
## -----
```

```

## loandata$IncomeRange: $100,000+
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
##     0      0      0    1242      0  463881      1
## -----
## loandata$IncomeRange: $25,000-49,999
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
##     0.0    0.0    0.0   897.3    0.0 444745.0      6
## -----
## loandata$IncomeRange: $50,000-74,999
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
##     0.0    0.0    0.0   908.8    0.0 284169.0      1
## -----
## loandata$IncomeRange: $75,000-99,999
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##     0      0      0    1204      0  265084
## -----
## loandata$IncomeRange: Not displayed
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
##     0.0    0.0    0.0   419.1    0.0 10574.0    7602
## -----
## loandata$IncomeRange: Not employed
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max. NA's
##     0.0    0.0    0.0   656.6    0.0 91554.0      2
## # A tibble: 86 x 3
##   DebtToIncomeRatio loancurrentdaysbreaks_mean     n
##   <dbl> <dbl> <int>
## 1 0.100 0.100 248 4283
## 2 0.200 0.200 164 22430
## 3 0.300 0.300 138 37246
## 4 0.400 0.400 134 20819
## 5 0.500 0.500 162 4162
## 6 0.600 0.600 252 1816
## 7 0.700 0.700 291 557
## 8 0.800 0.800 234 347
## 9 0.900 0.900 317 189
## # ... with 76 more rows

```

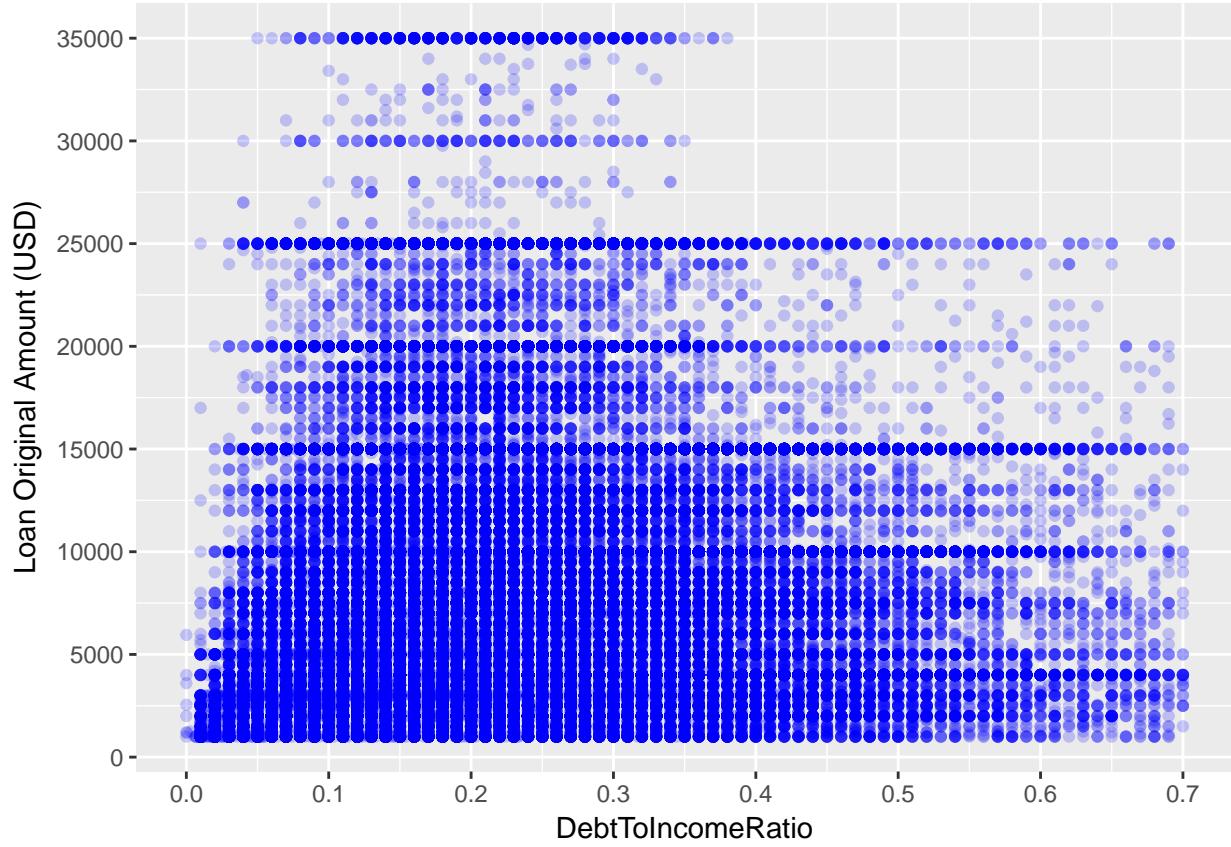
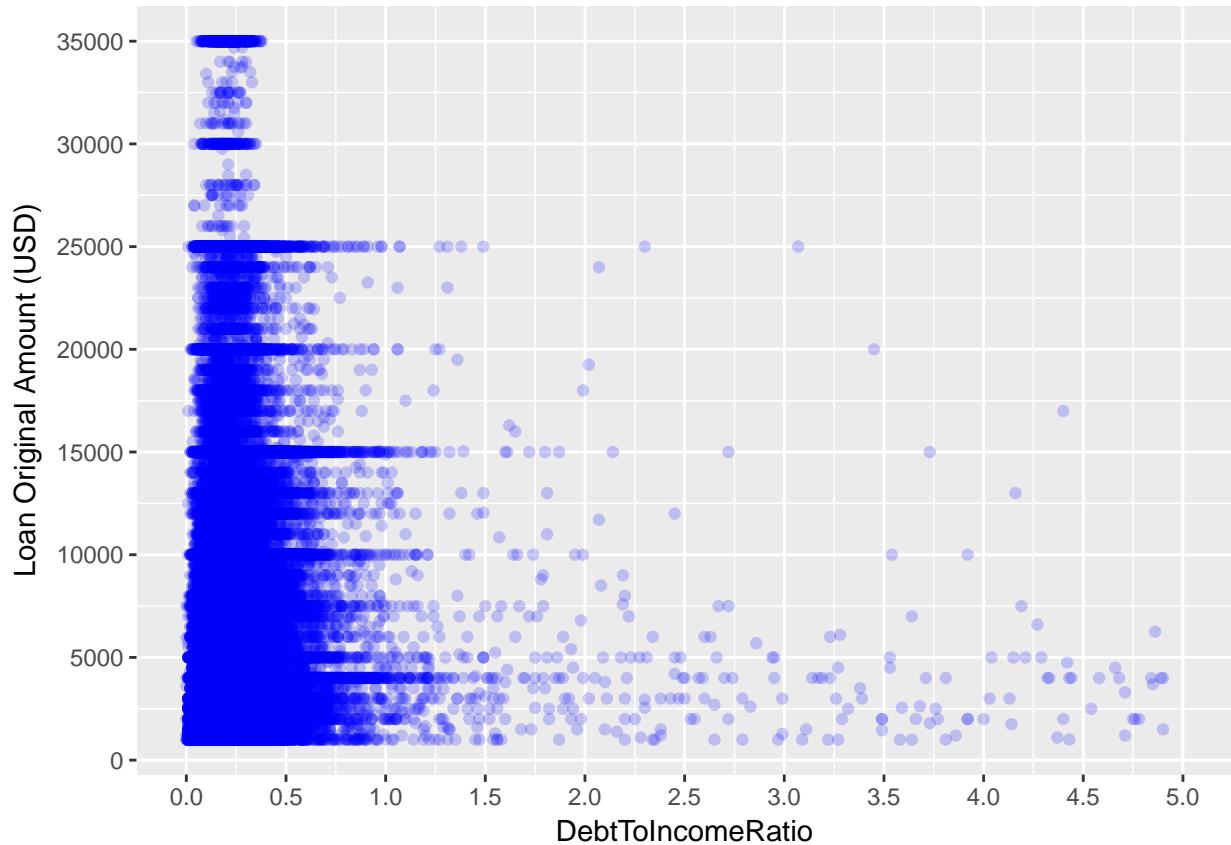
Average Loan Days Delinquent vs Debt To Income Ratio



For the first plot, a 95% confidence interval was given, with a scatterplot that was set to jitter.

For the second plot above, firstly, debttoincomeratio was rounded to 1/10's place. This would help with grouping the same. Secondly debttoincomeratio was grouped by its values and the mean was taken for the loan days. This was then plotted with a line plot.

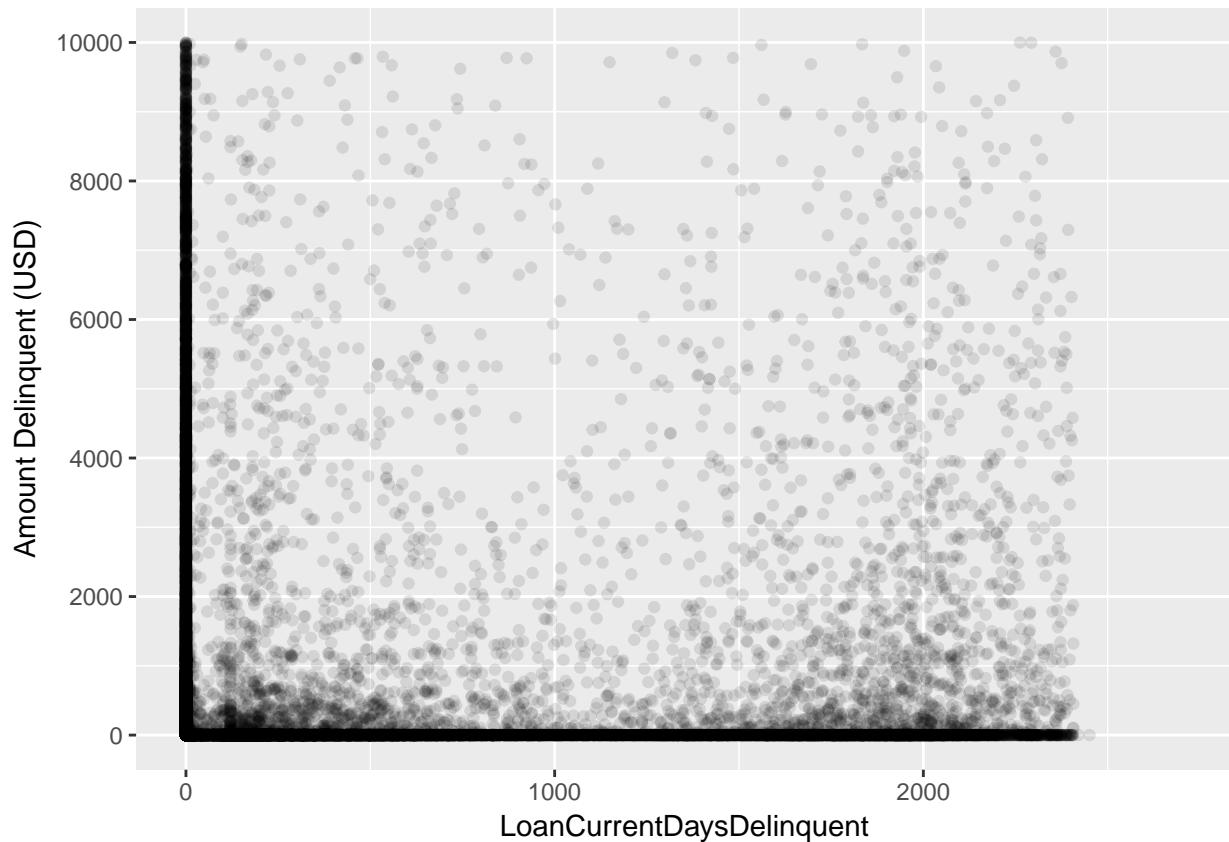
```
##  
## Pearson's product-moment correlation  
##  
## data: loandata$DebtToIncomeRatio and loandata$LoanOriginalAmount  
## t = 3.2828, df = 105380, p-value = 0.001028  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.004074882 0.016148830  
## sample estimates:  
## cor  
## 0.01011222
```

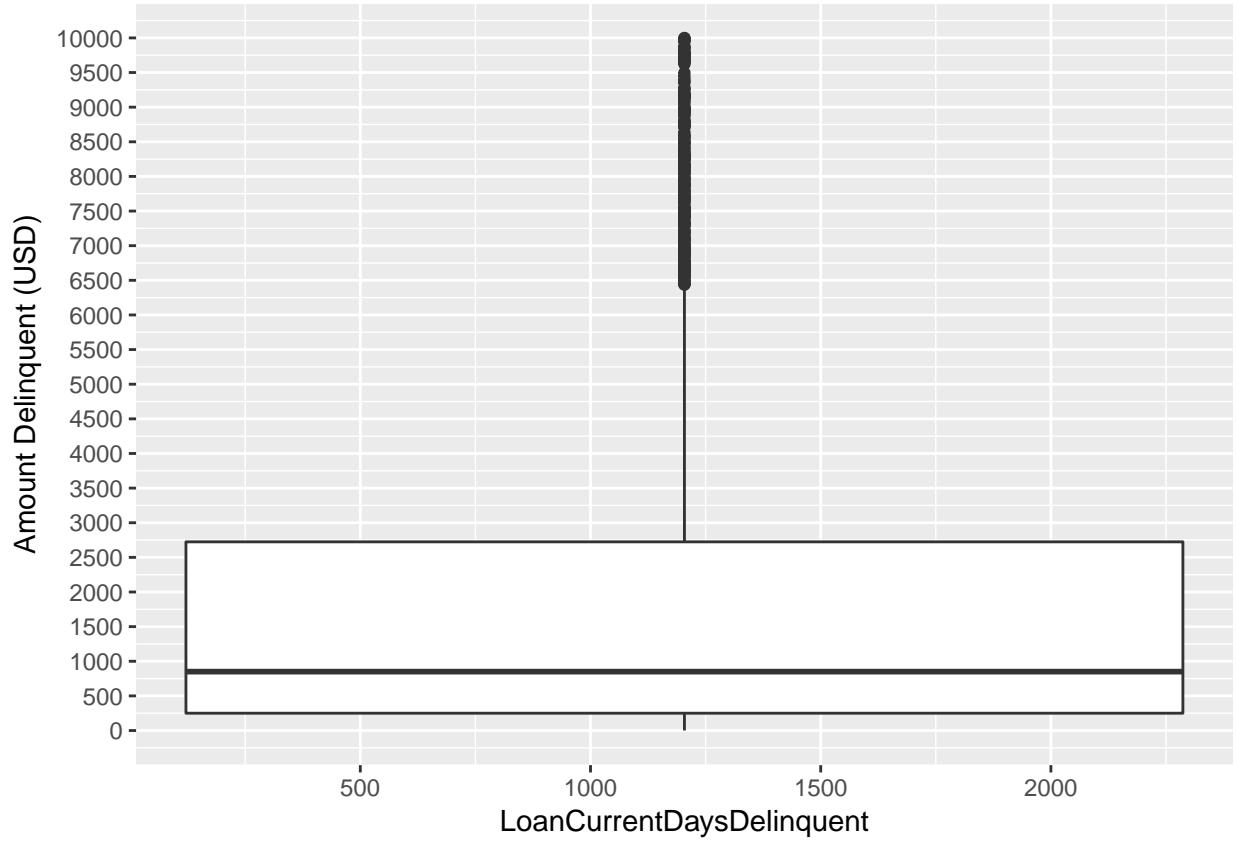
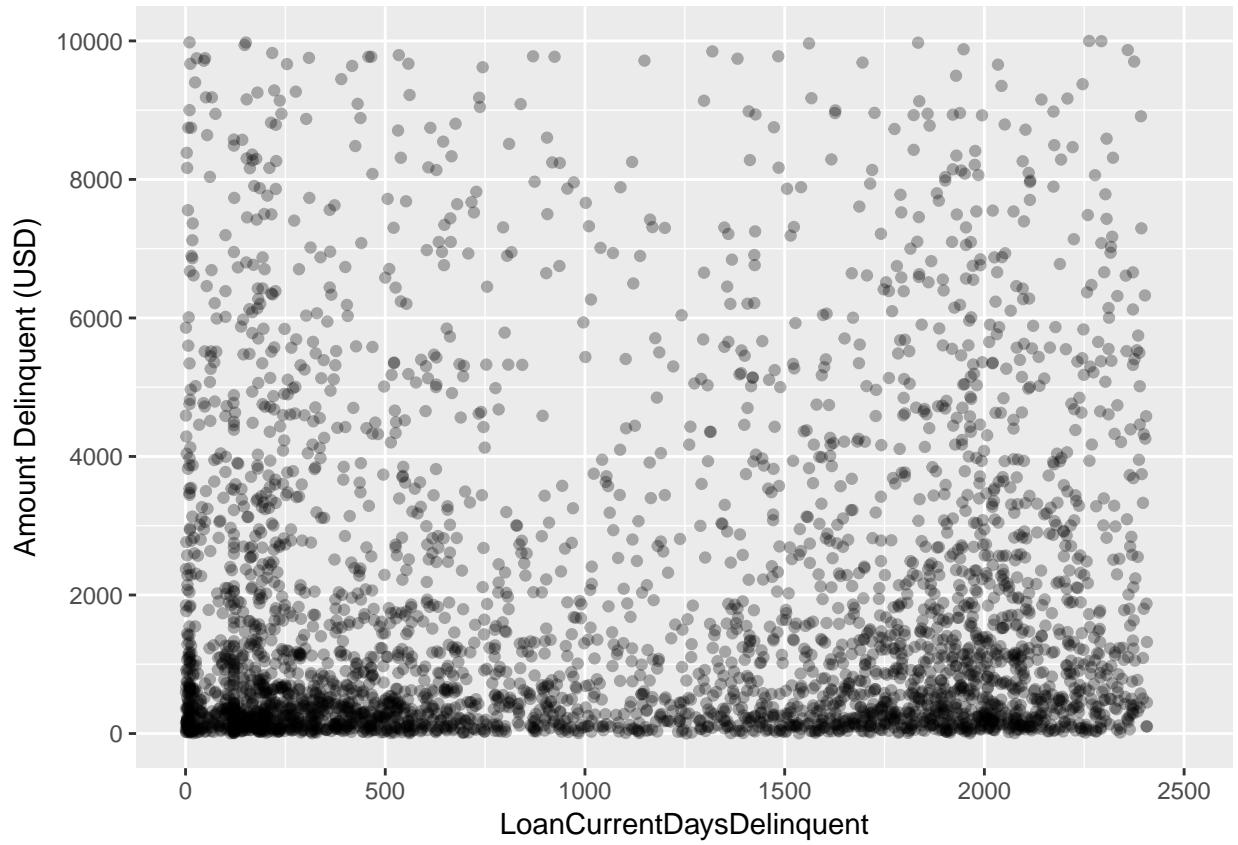


For the first plot, it can be seen that the majority of the population is up to 1.0.

As seen above a plot was made comparing DebtToIncomeRatio to LoanOriginalAmount. This was scaled for DebtToIncomeRatio to 0.7 being the maximum. As can be seen there are amounts borrowed for every \$5,000. However the trend can be seen that as the ratio goes higher, the amounts borrowed goes lower.

```
##  
## Pearson's product-moment correlation  
##  
## data: loandata$AmountDelinquent and loandata$LoanCurrentDaysDelinquent  
## t = 10.015, df = 106310, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.02469398 0.03670479  
## sample estimates:  
## cor  
## 0.03070049
```





The first plot above shows all values in the population. Alpha was reduced to 10% to show that the majority

persons in this dataset is not delinquent on both accounts.

For the second and third plot, the data was subsetted to omit 0 to take a closer look at those who were delinquent (days and amounts).

Bivariate Analysis

Investigation

For Debt To Income Ratio vs Amount Delinquent, generally as the ratio got higher, the amount persons owed decreased. This was not a very strong correlation.

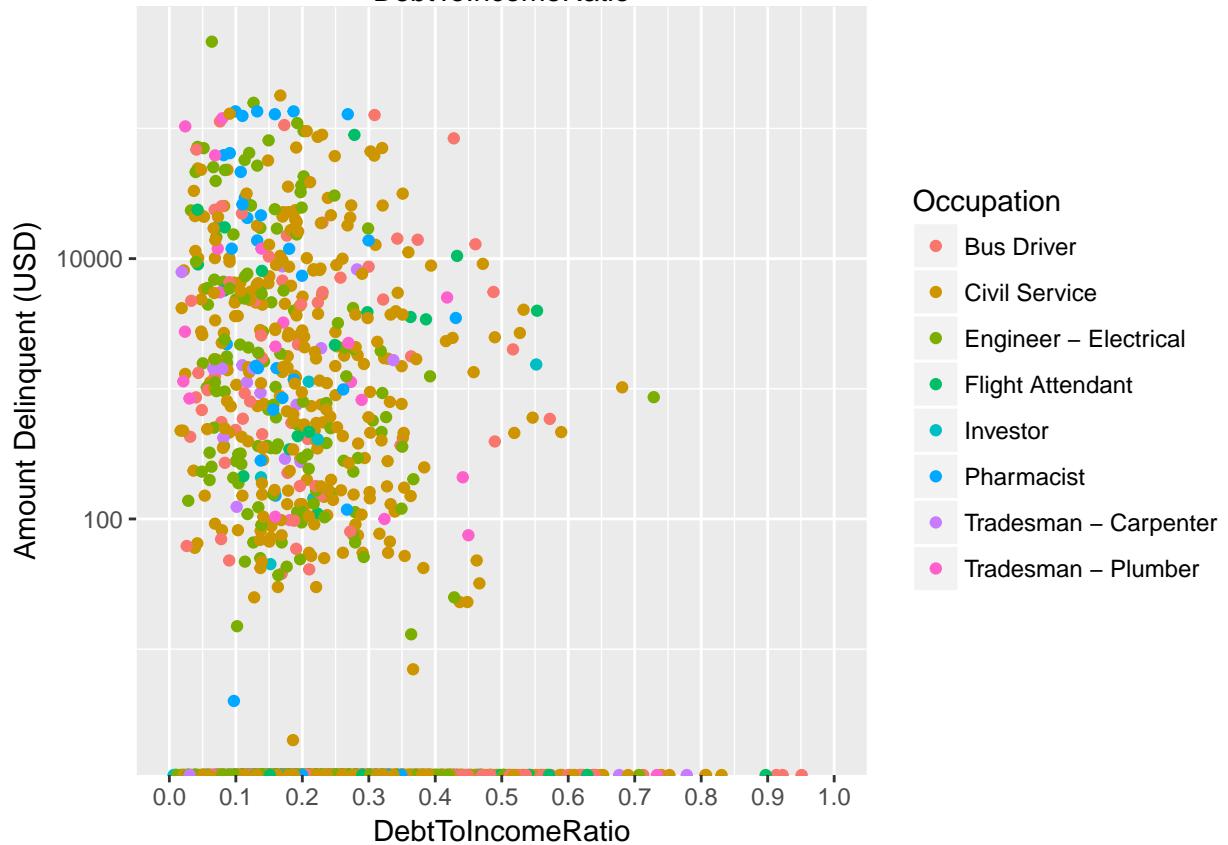
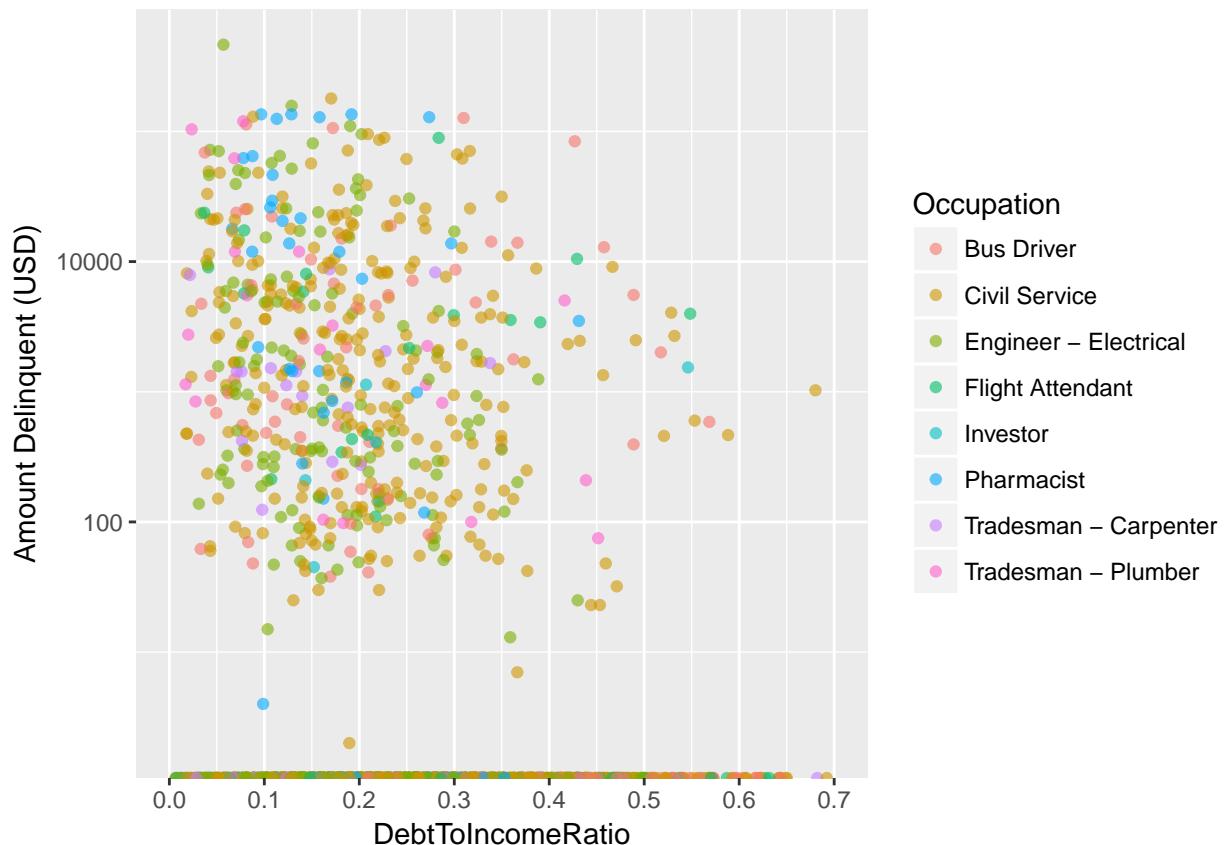
For Debt To Income Ratio vs the loan current days delinquent, this as well had no real pattern but was interesting to see that generally persons who had a higher debt ratio as expected they had more days delinquent in their payments.

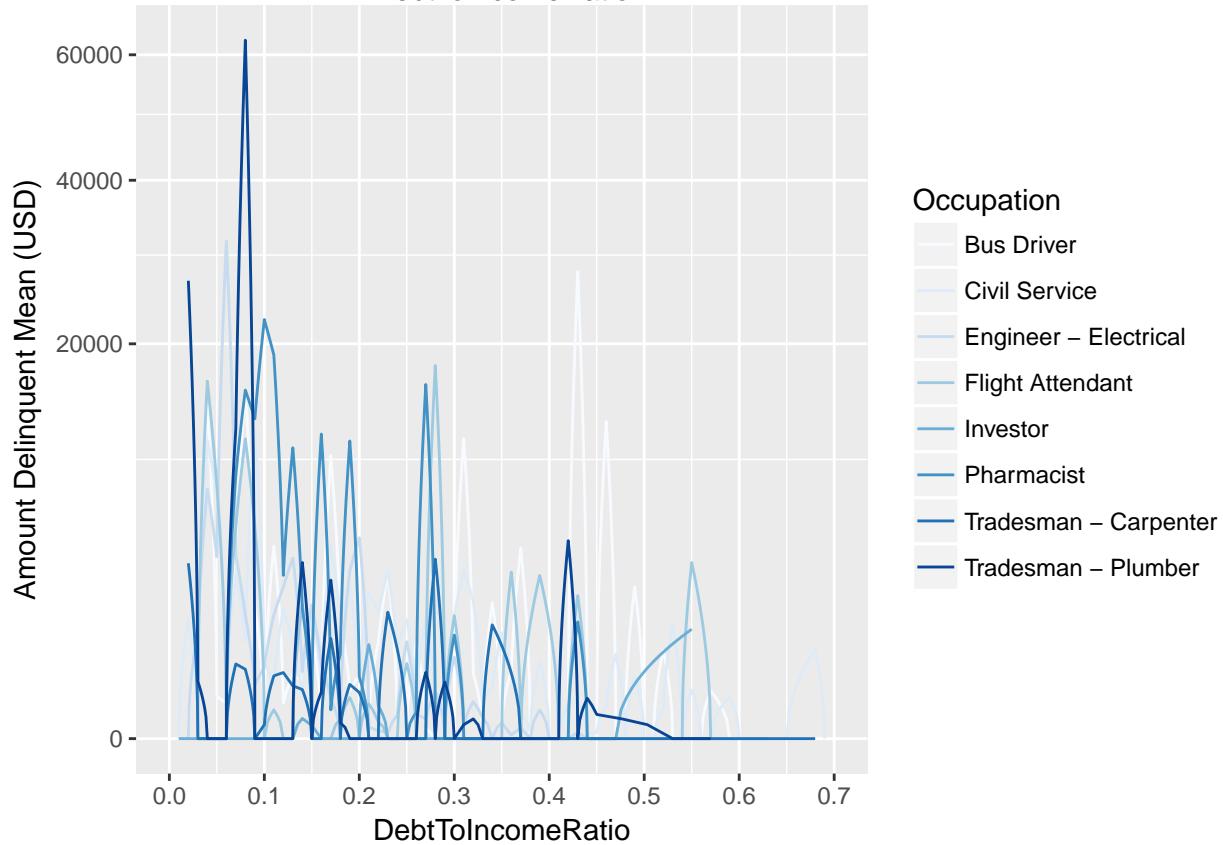
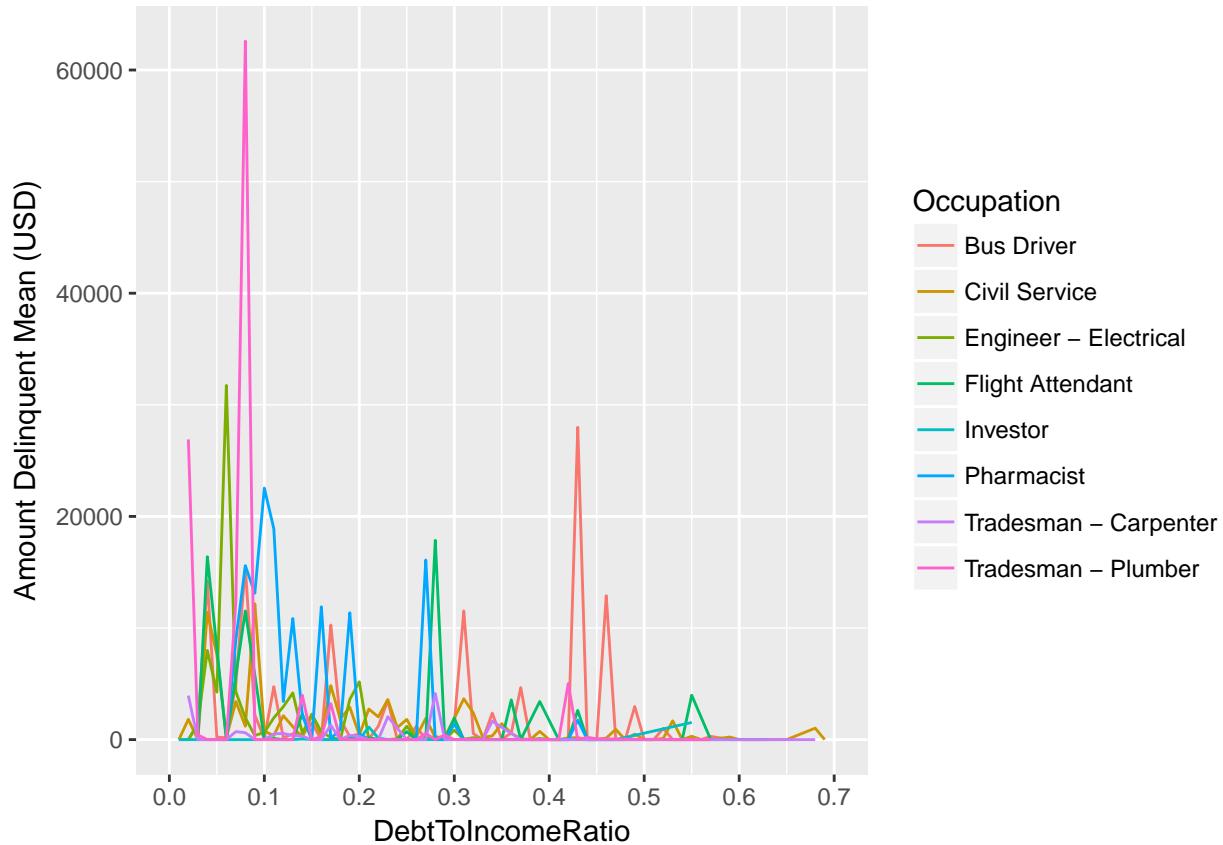
For the debt to income ratio vs amount loaned, as expected the amounts borrowed became lower as the ratio was higher.

Features

The most interesting feature found here was that as the ratio got higher the Amount that a person was delinquent lowered. This was also the strongest relationship found and most clear found.

Multivariate Plots Section





Above, looking at the relationships between debt to income ratio and amount delinquent, no real relationship

within Occupations. Above are the highest occupations in the for Amount Delinquent.

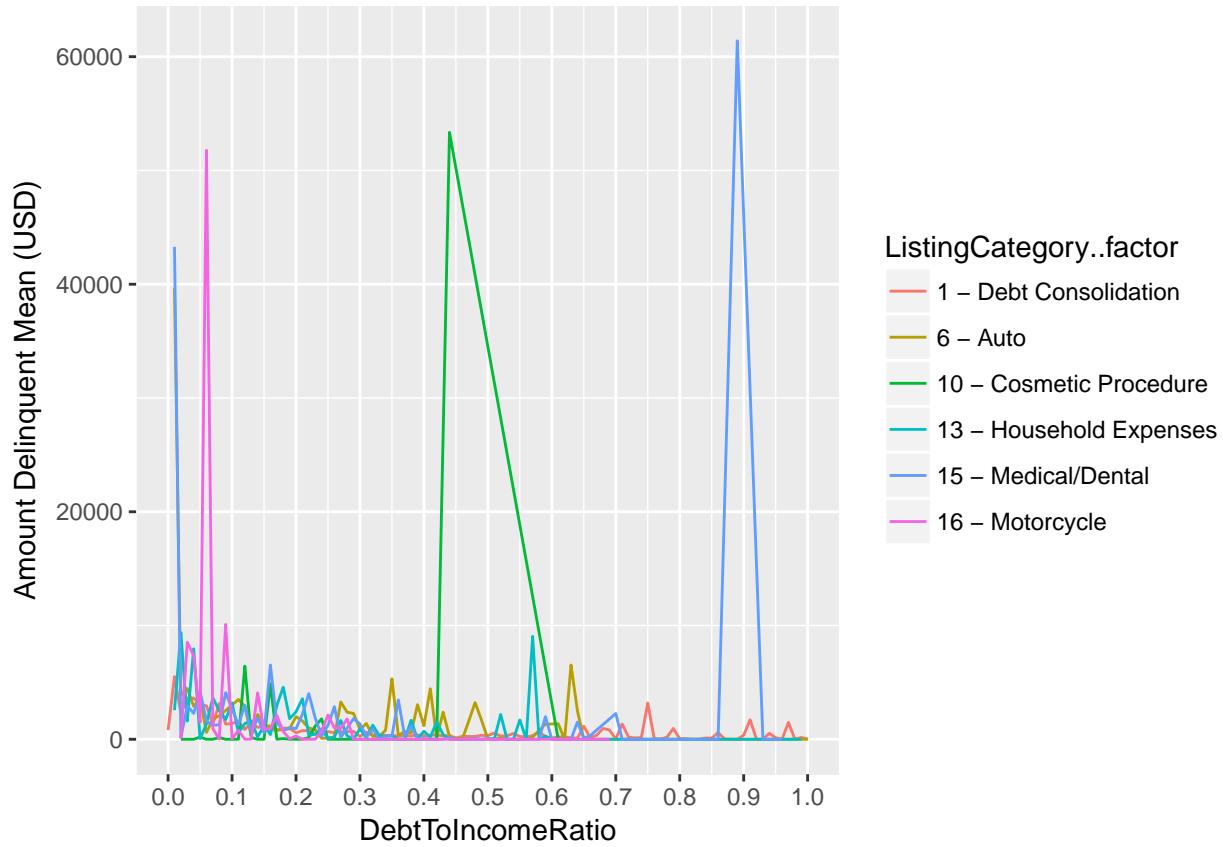
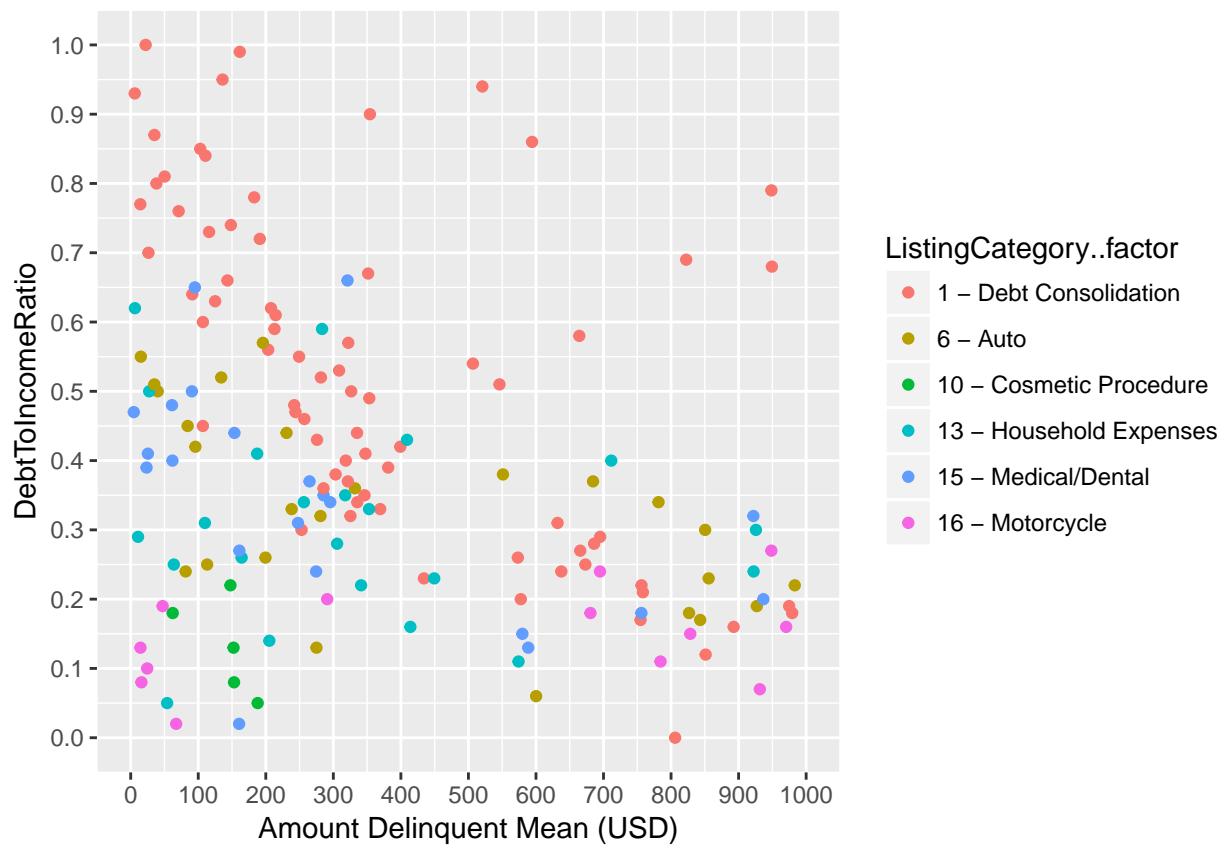
Looking at the different occupations in a geom line, we can see that the bus drivers has some high spikes for amount delinquent.

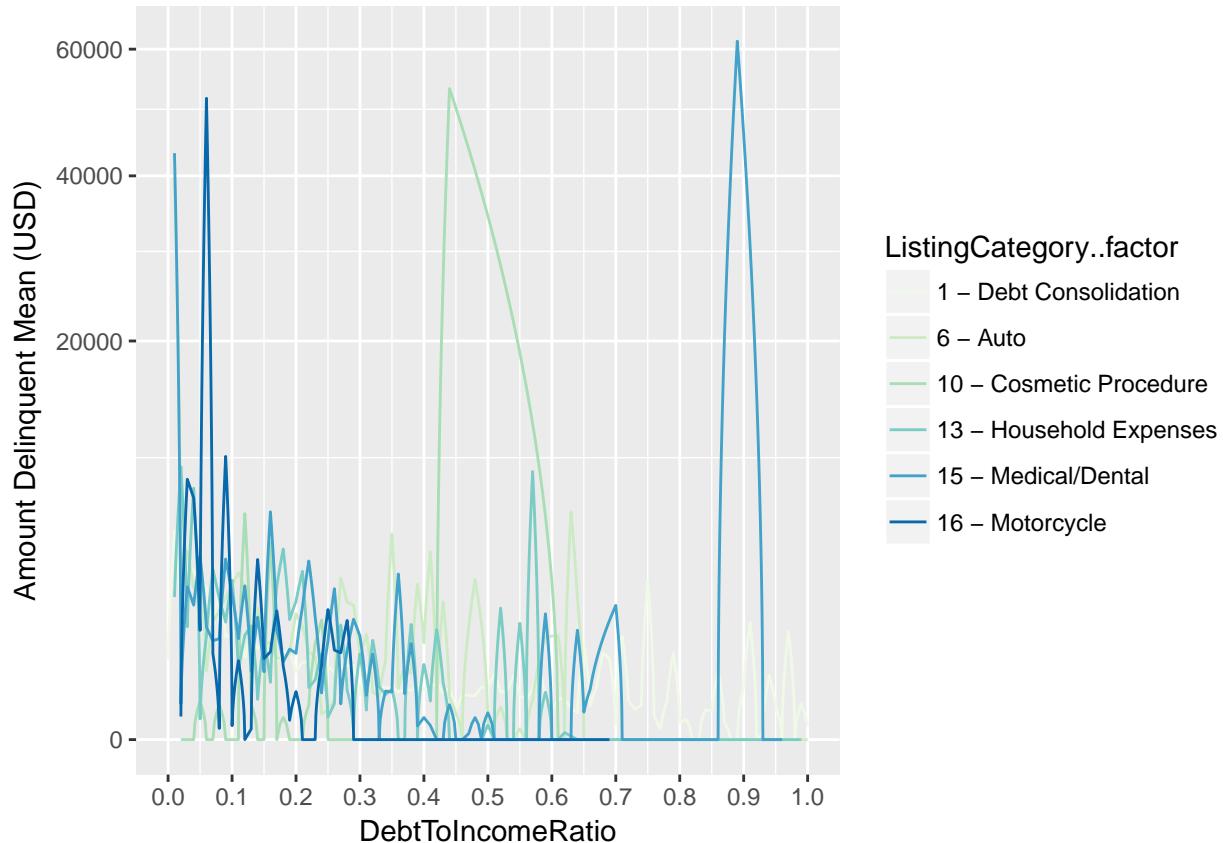
```
## loandata$ListingCategory..factor: 0
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##   0       0      0    1468      318  444745    7616
## -----
## loandata$ListingCategory..factor: 1
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0.0     0.0      0.0    758.3      0.0 265084.0
## -----
## loandata$ListingCategory..factor: 2
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##   0       0      0    1020      0  279970      1
## -----
## loandata$ListingCategory..factor: 3
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0       0      0    1128      0  284169
## -----
## loandata$ListingCategory..factor: 4
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##   0       0      0    1140      0  161344      3
## -----
## loandata$ListingCategory..factor: 5
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0.0     0.0      0.0    545.9      0.0 60616.0
## -----
## loandata$ListingCategory..factor: 6
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##   0       0      0    1631      0  164607      1
## -----
## loandata$ListingCategory..factor: 7
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.    NA's
##   0       0      0    1187      0  255963      1
## -----
## loandata$ListingCategory..factor: 8
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0.0     0.0      0.0    792.1      0.0 60997.0
## -----
## loandata$ListingCategory..factor: 9
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0       0      0    1212      0  58858
## -----
## loandata$ListingCategory..factor: 10
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0       0      0    2092      0  84834
## -----
## loandata$ListingCategory..factor: 11
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##   0       0      0    1155      0  43200
## -----
## loandata$ListingCategory..factor: 12
##   Min. 1st Qu. Median     Mean 3rd Qu.    Max.
```

```

##      0.0      0.0      0.0   260.3      0.0  9324.0
## -----
## loandata$ListingCategory..factor: 13
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    1430       0  179158
## -----
## loandata$ListingCategory..factor: 14
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    1363       0  223738
## -----
## loandata$ListingCategory..factor: 15
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    1606       0  111690
## -----
## loandata$ListingCategory..factor: 16
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    2871       0  463881
## -----
## loandata$ListingCategory..factor: 17
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    1346       0  50688
## -----
## loandata$ListingCategory..factor: 18
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    1748       0  327677
## -----
## loandata$ListingCategory..factor: 19
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0    1065       0  72705
## -----
## loandata$ListingCategory..factor: 20
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
##      0       0       0     960       0  105340

```





First, looked at all of the listing categories and the amount delinquent for each one. Next, grouped all of the debt to income ratio and calculated the means for each amount delinquent. This was done by summarising. Created a scatter plot of all of the amounts accross the debt to income ratio.

The second plot created is a line plot which shows the same averages in comparison.

Finally the third plot which is a line plot zooms in from 0 to 1.0 of debt to income ratio.

Multivariate Analysis

Features

For the Amount Delinquent Mean plot, this would clearly show the different Occupations and which ones have high means compared to their initial debt to income ratio.

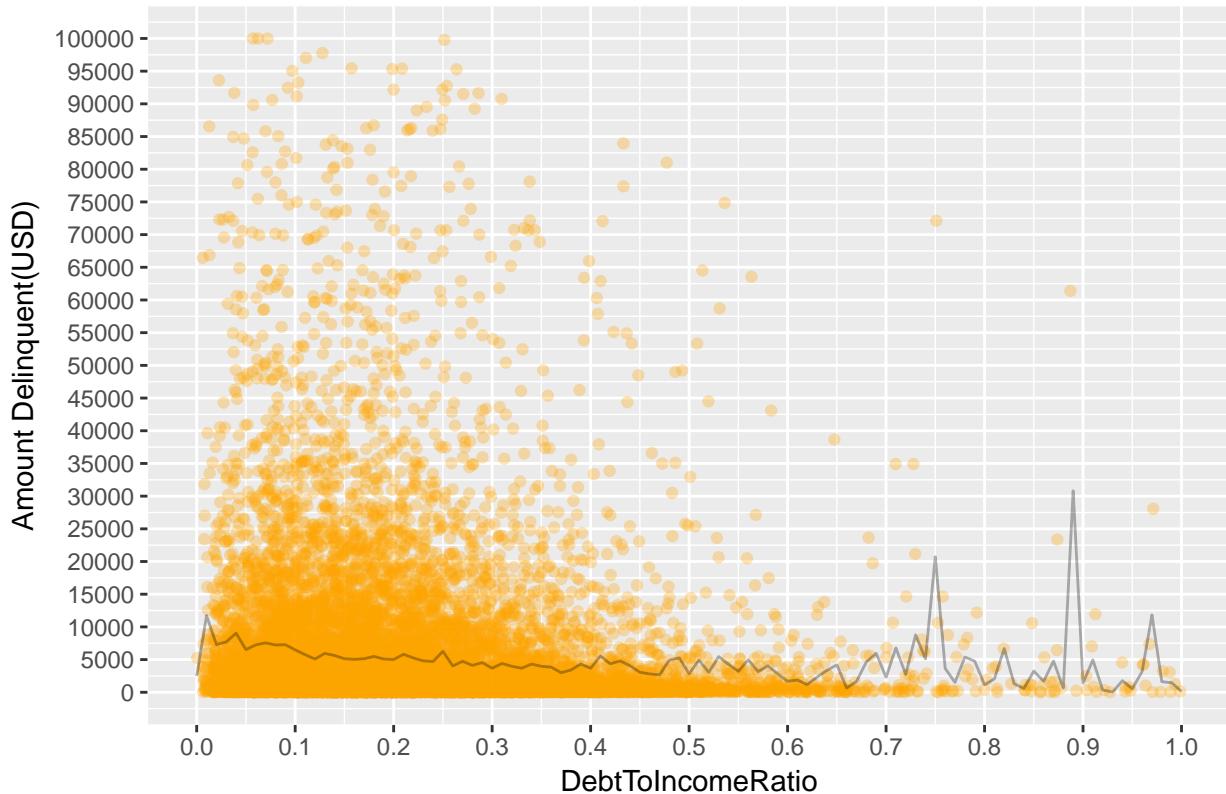
Additional Features

The interesting factor is that cosmetic procedure seems to be a top factor of persons who are in a delinquent status.

Final Plots and Summary

Plot One

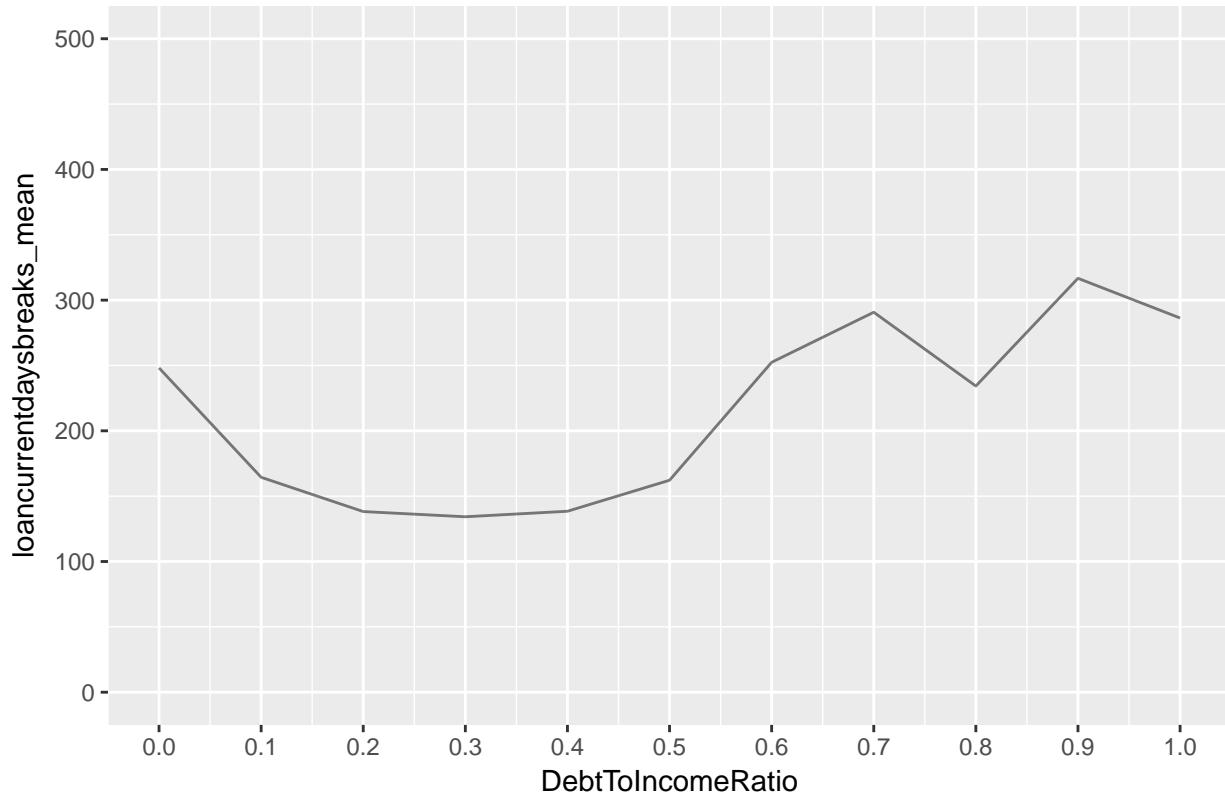
Amount Delinquent per Debt to Income Ratio



This plot gives a good idea of the dataset and the fact that generally as the debttoincomeratio increases the amountdelinquent gradually decreases. In other words, those who have good debt to income ratios will owe more, most likely because they borrowed more in the first place. If those who have a really bad ratio owe less, its probably because they borrowed less.

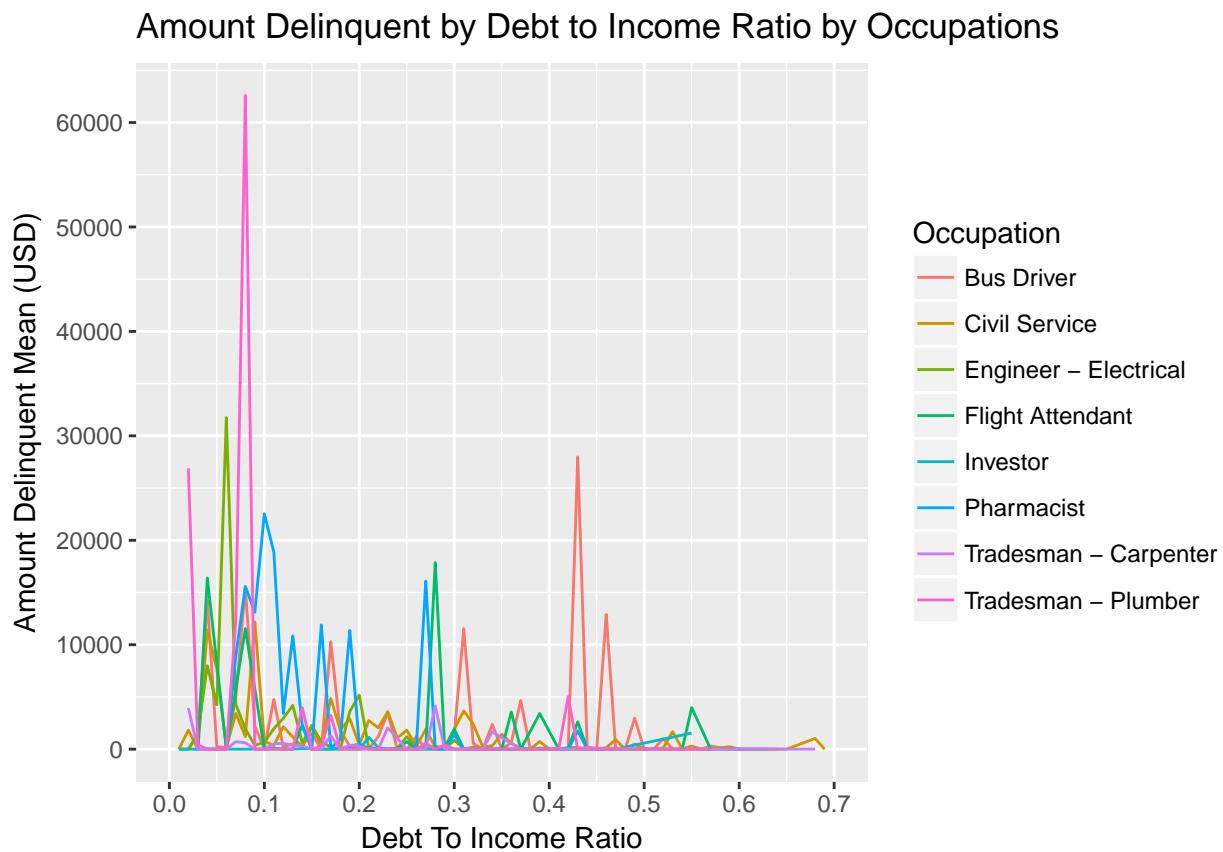
Plot Two

Average Loan Days Delinquent vs Debt To Income Ratio



This plot showed the days that a person had a loan for based on their debt to income ratio. This initially decreased but then gradually got higher as their ratio went higher. This would show that those with higher debt to income ratio has more days delinquent. This would make sense because those who have a worse reputation more likely would have more days delinquent.

Plot Three



This plot shows on the amount delinquent averages by the debt to income ratios and this is separated by occupations. As we can see above, the tradesman plumber has some high amounts owed as the ratio goes higher while Pharmacists and Investors owe less as the ratio's are higher. —

Reflections

The Prosper Loan dataset contains 81 variables with 113,937 observations. I began by exploring the individual variables such as Debt to Income Ratio and Amount Delinquent. A lot of reductions had to be made to variables like Occupation because there were many Occupations found within the dataset.

One trend found was Amount Delinquent and Debt To Income Ratio in that the higher the Debt To Income Ratio the less Amounts were Delinquent.

In the future, it would be great to explore the time that a person has to pay back their loan compared with their loan status to determine if these times are feasible or not.