Loan Data EDA

Marissa Schmucker

February, 2018

2014 Prosper Loan Data Exploration

Prosper Website (www.prosper.com)

> Prosper is America's first marketplace lending platform, with over $10 billion in funded loans.

> 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.

> Prosper Funding LLC is a wholly-owned subsidiary of Prosper Marketplace, Inc.

> Prosper Marketplace is backed by leading investors including Sequoia Capital, Francisco Partners, Institutional Venture Partners, and Credit Suisse NEXT Fund.

Variables to Explore

There are 81 variables, of which we will be exploring 15.

- Term
- LoanStatus
- BorrowerRate

- ListingCategory
- BorrowerState
- Occupation
- EmploymentStatus
- CreditScoreRangeLower
- CreditScoreRangeUpper
- OpenCreditLines
- CurrentDelinquencies
- AmountDelinquent
- DebttoIncomeRatio
- StatedMonthlyIncome
- MonthlyLoanPayment

To clean up our data, we will extract NAs to form a new object called noNA. Also, the variable "ListingCategory..numeric." will be renamed "ListingCategory". "Term" and "ListingCategory" will be changed from numeric(int) to categorical(fctr).

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

```
##  [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"
```

Final Object Variable Names

```
## 'data.frame':    97903 obs. of  15 variables:
##  $ Term                 : Factor w/ 3 levels "12","36","60": 2 2 2 2 3 2 2 2 2 3 ...
##  $ LoanStatus           : Factor w/ 12 levels "Cancelled","Chargedoff",..: 3 4 4 4 4 4 4 4 4 4 ...
##  $ BorrowerRate         : num  0.158 0.092 0.0974 0.2085 0.1314 ...
##  $ ListingCategory      : Factor w/ 21 levels "0","1","2","3",..: 1 3 17 3 2 2 3 8 8 2 ...
##  $ BorrowerState        : Factor w/ 52 levels "","AK","AL","AR",..: 7 7 12 25 34 18 6 16 16 22 ...
##  $ Occupation           : Factor w/ 68 levels "","Accountant/CPA",..: 37 43 52 21 43 50 29 24 24 22 ...
##  $ EmploymentStatus     : Factor w/ 9 levels "","Employed",..: 9 2 2 2 2 2 2 2 2 2 ...
##  $ CreditScoreRangeLower: int  640 680 800 680 740 680 700 820 820 640 ...
##  $ CreditScoreRangeUpper: int  659 699 819 699 759 699 719 839 839 659 ...
##  $ OpenCreditLines      : int  4 14 5 19 17 7 6 16 16 2 ...
##  $ CurrentDelinquencies : int  2 0 4 0 0 0 0 0 0 1 ...
##  $ AmountDelinquent     : num  472 0 10056 0 0 ...
##  $ DebtToIncomeRatio    : num  0.17 0.18 0.15 0.26 0.36 0.27 0.24 0.25 0.25 0.12 ...
##  $ StatedMonthlyIncome  : num  3083 6125 2875 9583 8333 ...
##  $ MonthlyLoanPayment   : num  330 319 321 564 342 ...
##  - attr(*, "na.action")=Class 'exclude'  Named int [1:16034] 3 18 40 41 43 64 70 77 79 91 ...
##   .. ..- attr(*, "names")= chr [1:16034] "3" "18" "40" "41" ...
```

Number of Loans

There are just short of 114,000 loans in our original dataset. If we exclude loans with missing values (NAs), we have 97,903 loans. This will be our final set of loans to analyze.
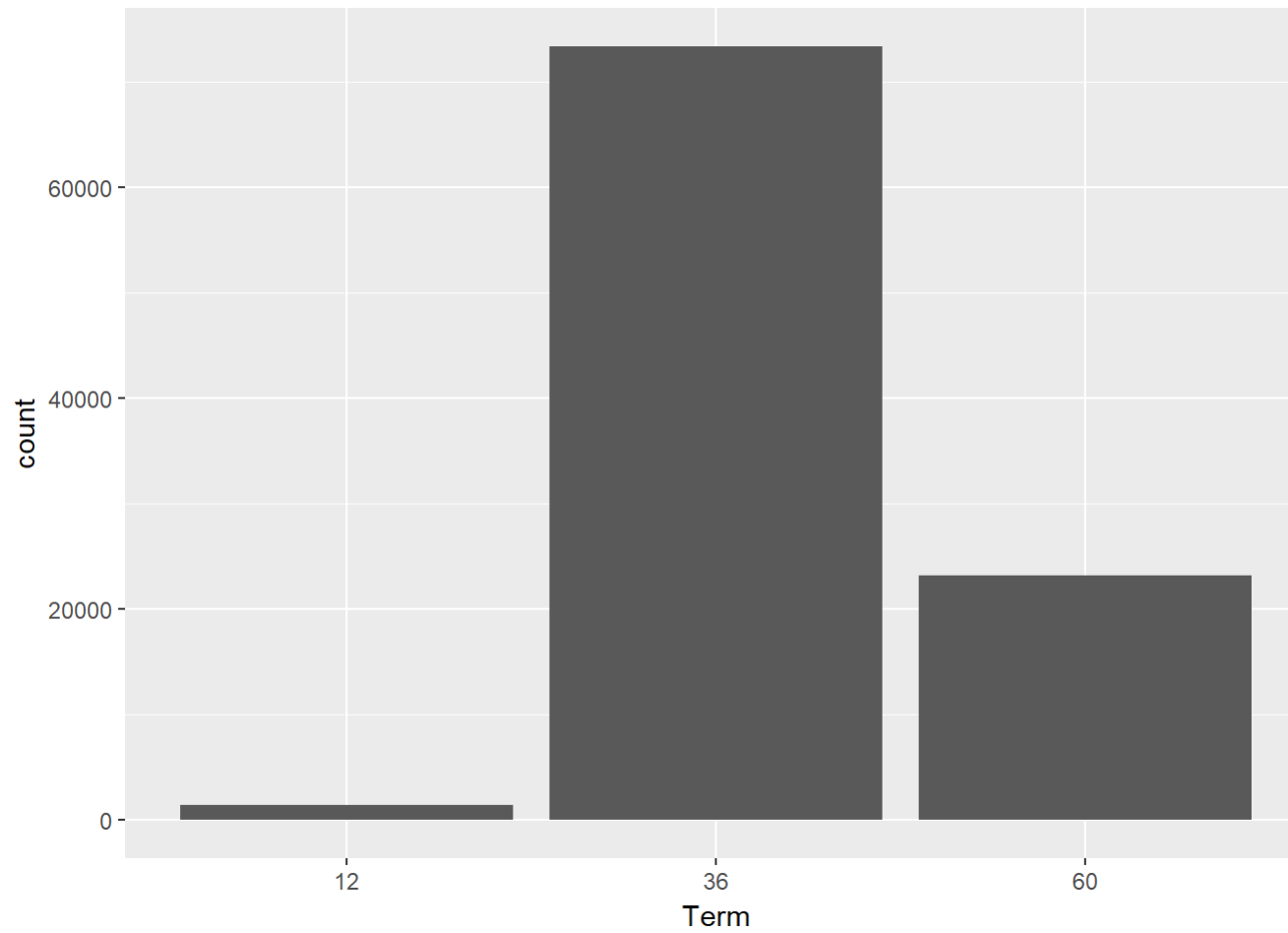
```
## [1] 113937
```

```
## [1] 97903
```

Univariate Plots

Term

Most of our loans have a 36 month term, followed by a 60 month term, and a 12 month term being the least popular.

```
##     12     36     60
##   1415  73345  23143
```


Status of Loan

Each loan is set to one status of 12 possible, shown below. Most loans are current or completed, with a little over 3000 in default.

```
##             Cancelled              Chargedoff              Completed
##                     1                    9423                  30880
##               Current               Defaulted FinalPaymentInProgress
##                 52478                    3075                    189
##    Past Due (>120 days)    Past Due (1-15 days)   Past Due (16-30 days)
##                    14                     722                    242
##   Past Due (31-60 days)   Past Due (61-90 days) Past Due (91-120 days)
##                   327                     275                    277
```

Loan Data EDA
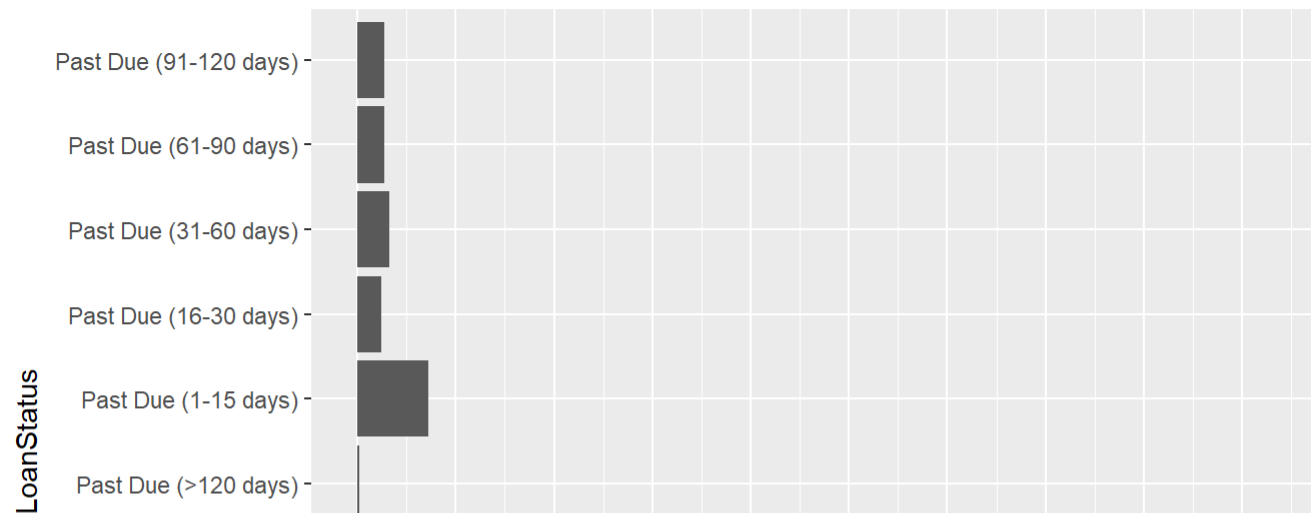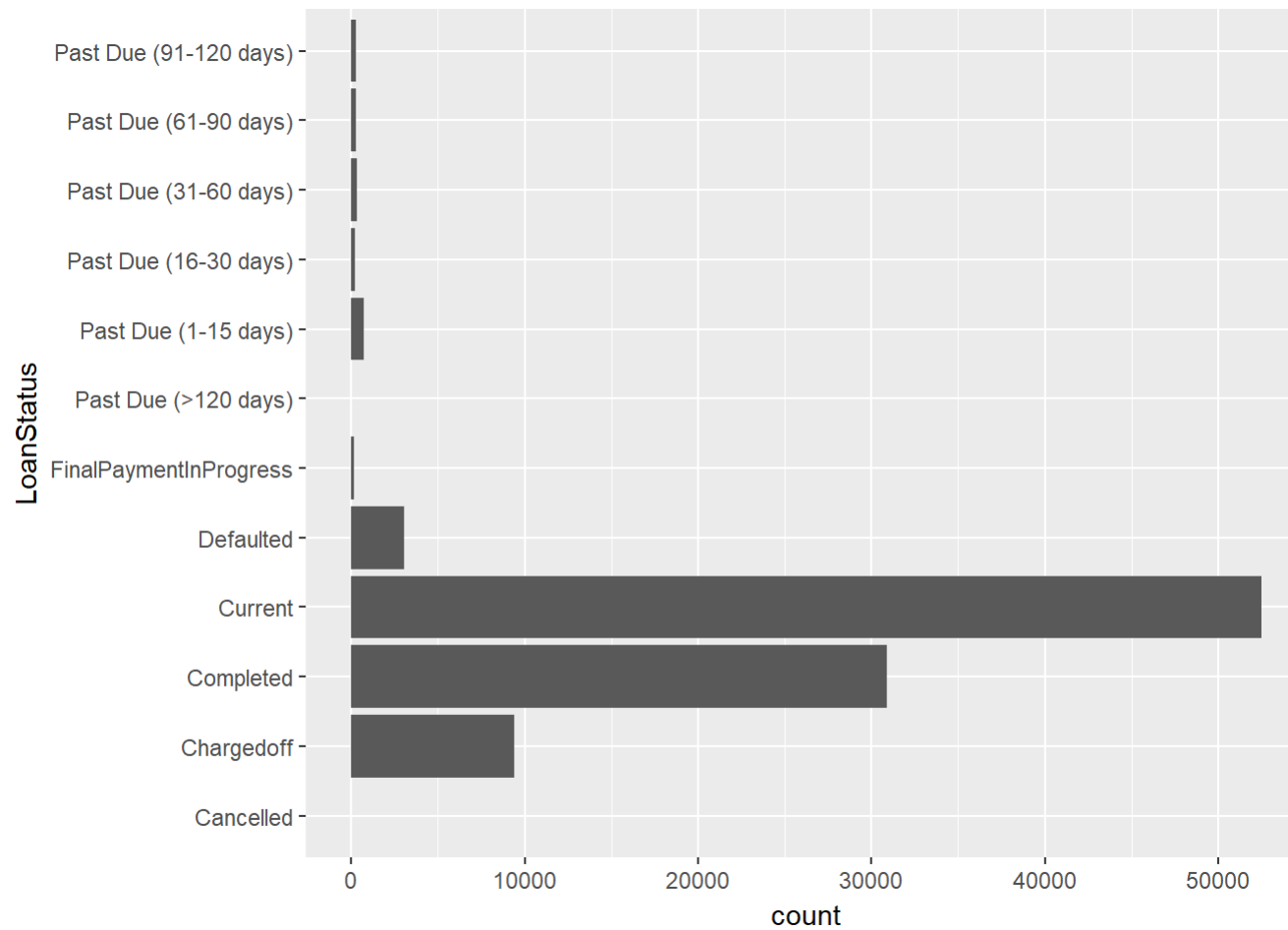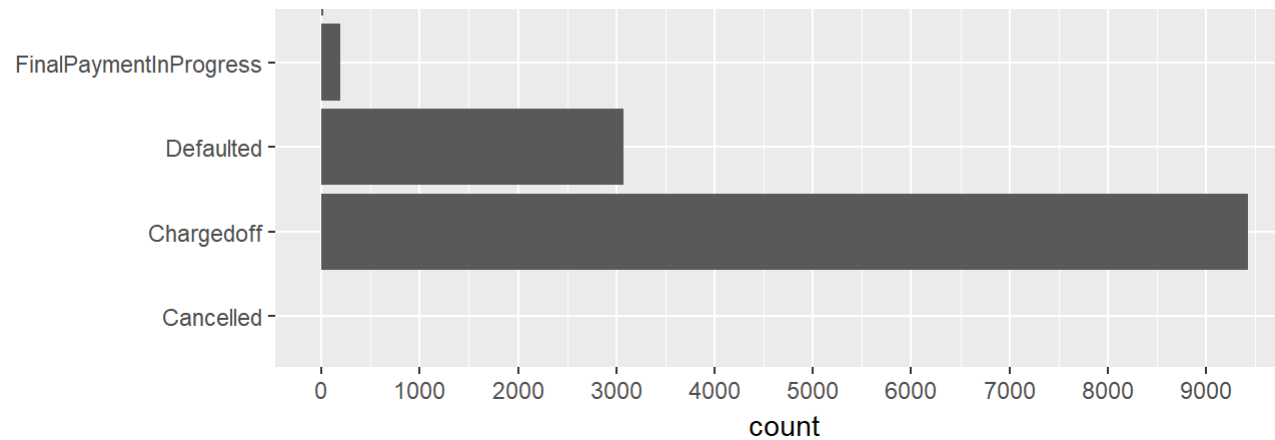
Interest Rate

The most popular interest rates are in the 10-20% range, yet there seems to be another rate with a high count of borrowers, 32%. According to Prosper.com, they offer loans with an APR as high as 35.99%. They state, "Annual percentage rates (APRs) through Prosper range from 5.99% APR (AA) to 35.99% APR (HR) for first-time borrowers, with the lowest rates for the most creditworthy borrowers."

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.1314  0.1800  0.1907  0.2492  0.3600
```
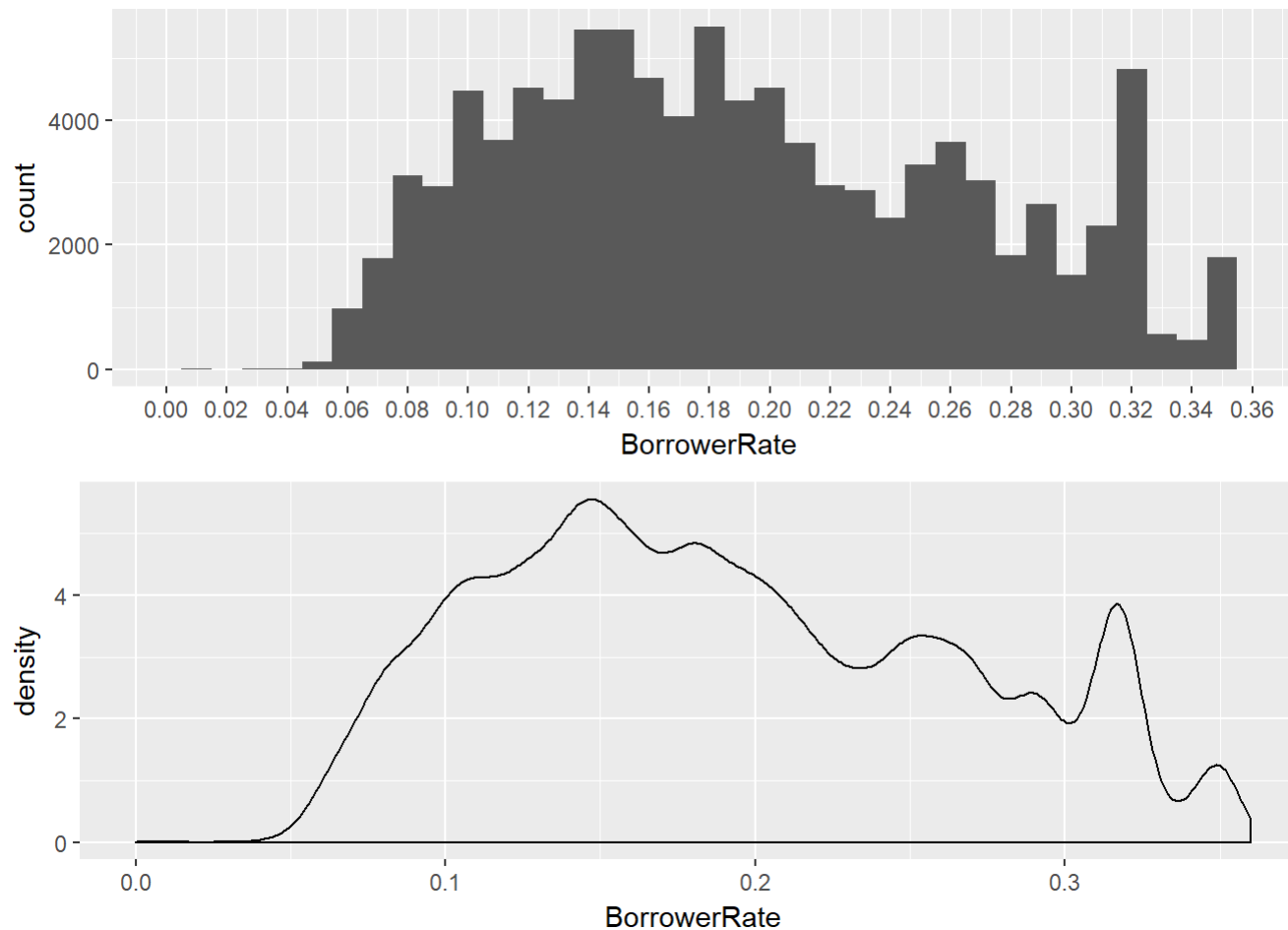
Loan Category

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 and 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

Debt Consolidation looks like the most popular loan category by far. In our second plot, we will exclude that category in order to zoom in on all of the other categories. By doing this, we can see that "Home Improvement" and "Business" make up about 12% of our loans.

```
##      0     1     2     3     4     5     6     7     8     9    10    11
##   9159 54628  6959  5205  2271   605  2363  9531   191    83    82   201
##     12    13    14    15    16    17    18    19    20
##     46  1788   806  1404   289    50   788   722   732
```
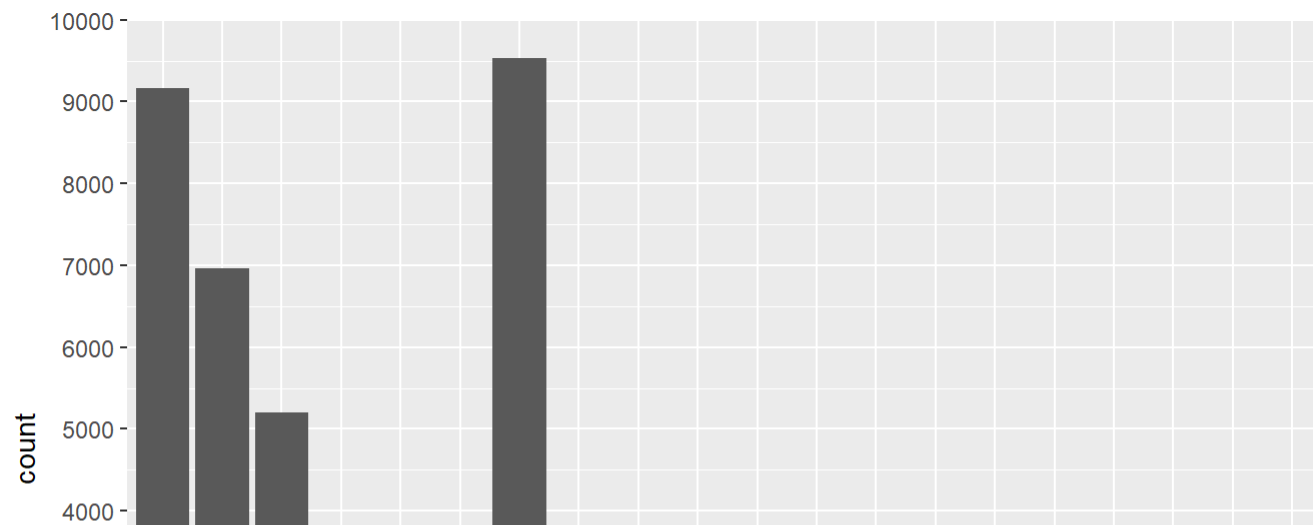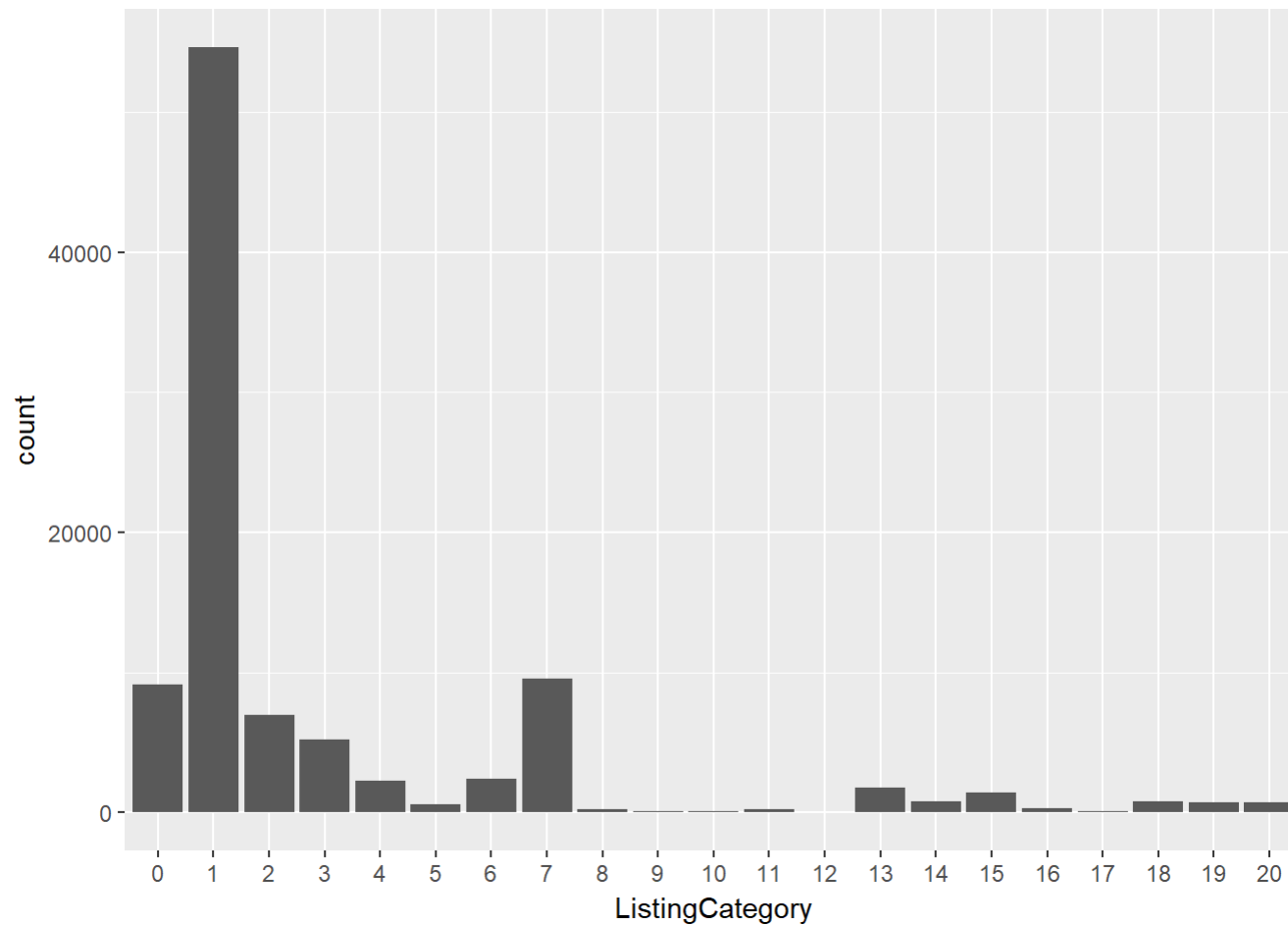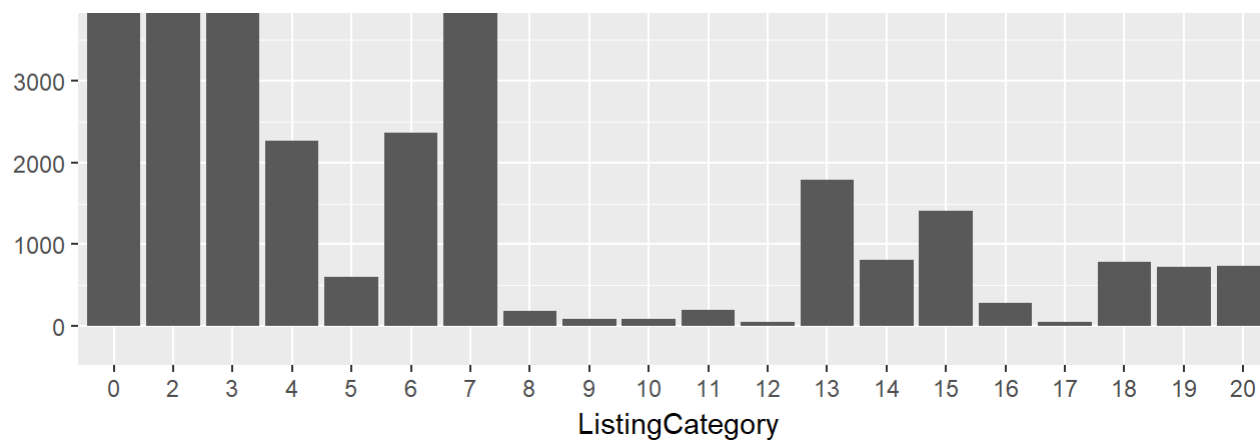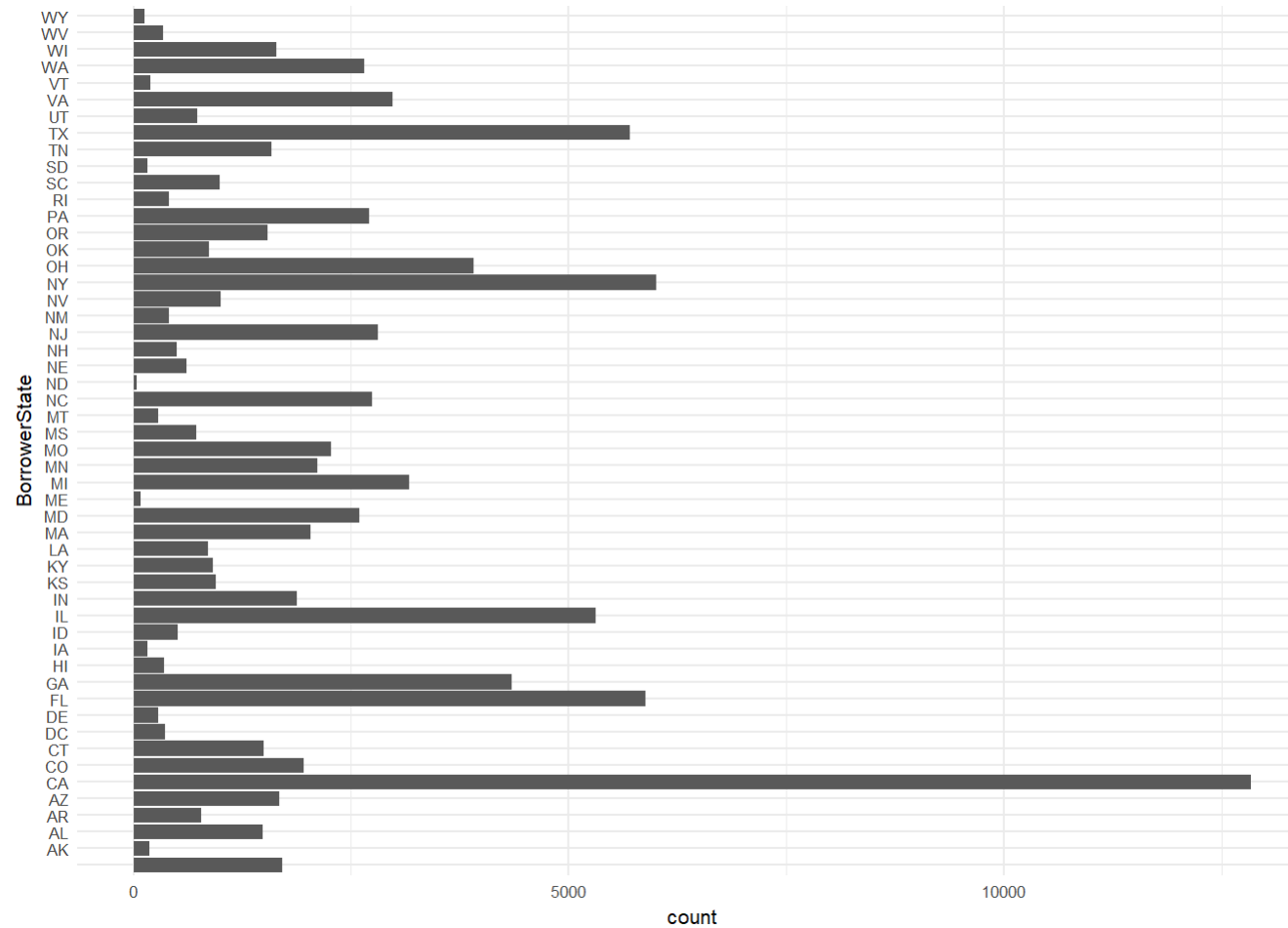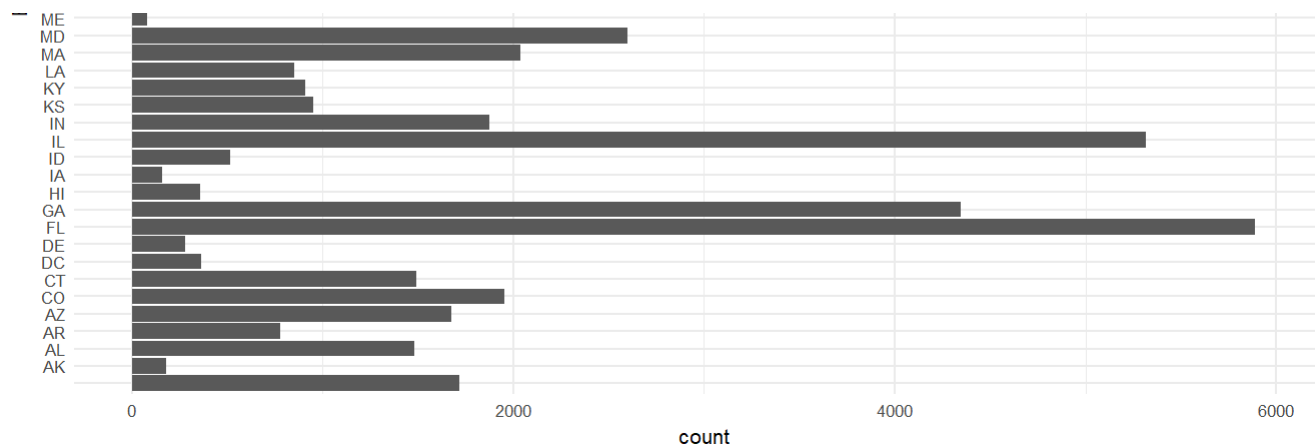
State

Looks like California has more loans than any other state. This makes sense since Prosper is located in CA, and CA is in the top 10 when it comes to cost of living (more people in need of loans.)

```
##
##     CA     NY     FL     TX     IL     GA     OH     MI     VA     NJ     NC     PA
## 12836   6002   5887   5710   5315   4344   3905   3172   2979   2812   2746   2713
##     WA     MD     MO     MN     MA     CO     IN            AZ     WI     TN     OR
##   2648   2598   2269   2118   2038   1955   1876   1716   1678   1644   1586   1538
##     CT     AL     NV     SC     KS     KY     OK     LA     AR     UT     MS     NE
##   1492   1484   1002    996    951    910    871    854    779    737    718    607
##     ID     NH     NM     RI     DC     HI     WV     MT     DE     VT     AK     SD
##    515    502    409    405    363    357    345    283    282    192    181    166
##     IA     WY     ME     ND
##    160    133     83     41
```

Loan Data EDA

Occupation

Since we know that most loans are taken out in CA, it makes sense that we'd see a greater number for common occupations in that state, such as Computer Programmers. But, since most of the occupations were classified as "Other" or "Professional", we have no way of knowing what occupation is truly the most common in loan applicants. Still, we can omit those categories in our second graph to get a better idea of some of the more popular occupations.

```
##
##                              Other                        Professional
##                              23782                              12341
##                 Computer Programmer                          Executive
##                               3994                               3859
##                            Teacher                            Analyst
##                               3480                               3390
##             Administrative Assistant                     Accountant/CPA
##                               3379                               2947
##                           Clerical                 Sales - Commission
##                               2796                               2763
##                      Skilled Labor                         Nurse (RN)
##                               2488                               2400
##                  Retail Management                      Sales - Retail
##                               2366                               2277
##   Police Officer/Correction Officer                        Truck Driver
##                               1526                               1464
##                            Laborer                      Civil Service
##                               1447                               1401
##                       Construction
##                               1383                               1333
##                Engineer - Mechanical                  Military Enlisted
##                               1315                               1151
##            Food Service Management                 Engineer - Electrical
##                               1124                               1056
##                  Medical Technician                       Food Service
##                               1044                                941
##                Tradesman - Mechanic                           Attorney
##                                868                                852
##                      Social Worker                     Postal Service
##                                692                                594
##                          Professor                        Nurse (LPN)
##                                518                                460
##               Tradesman - Electrician                      Nurse's Aide
##                                446                                431
##                             Doctor                            Fireman
##                                418                                398
##                    Waiter/Waitress                          Scientist
##                                364                                342
##                    Military Officer                         Bus Driver
```

```
##                               327                               297
##                         Principal                           Realtor
##                               287                               273
##                   Teacher's Aide                         Pharmacist
##                               249                               246
##               Engineer - Chemical                          Architect
##                               215                               186
##         Pilot - Private/Commercial                             Clergy
##                               184                               171
## Student - College Graduate Student                         Car Dealer
##                               169                               144
##                           Chemist                         Landscaping
##                               139                               126
##                         Biologist                    Flight Attendant
##                               120                               117
##           Student - College Senior                        Psychologist
##                               114                               108
##                         Religious                  Tradesman - Plumber
##                                95                                87
##             Tradesman - Carpenter                            Investor
##                                75                                66
##           Student - College Junior                            Dentist
##                                62                                56
##                         Homemaker     Student - College Sophomore
##                                43                                43
##         Student - College Freshman                              Judge
##                                29                                22
##        Student - Community College     Student - Technical School
##                                15                                 8
```
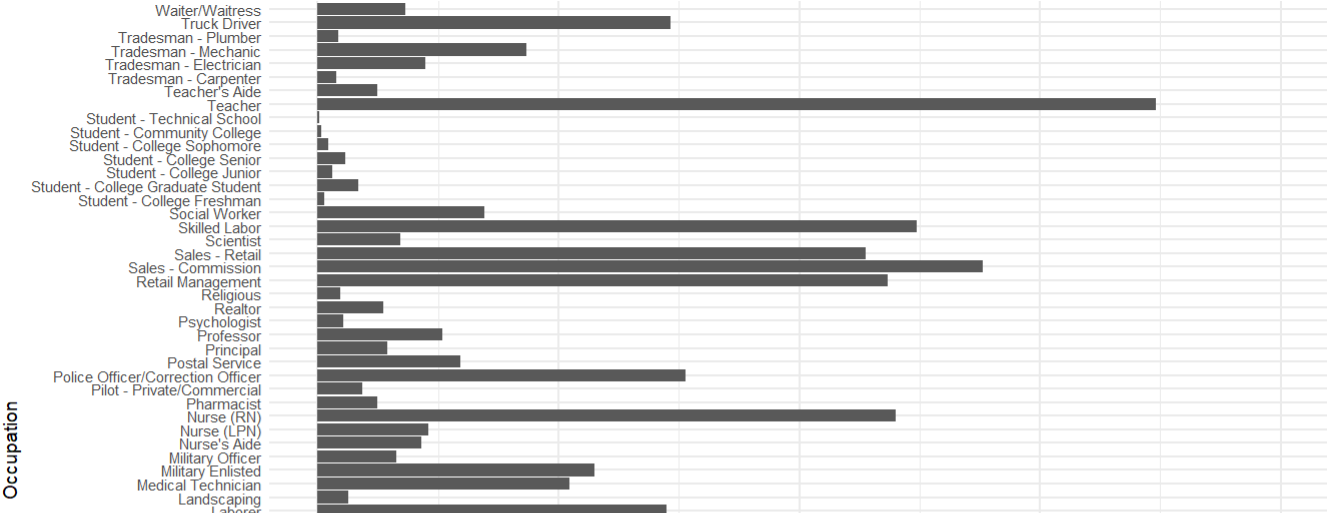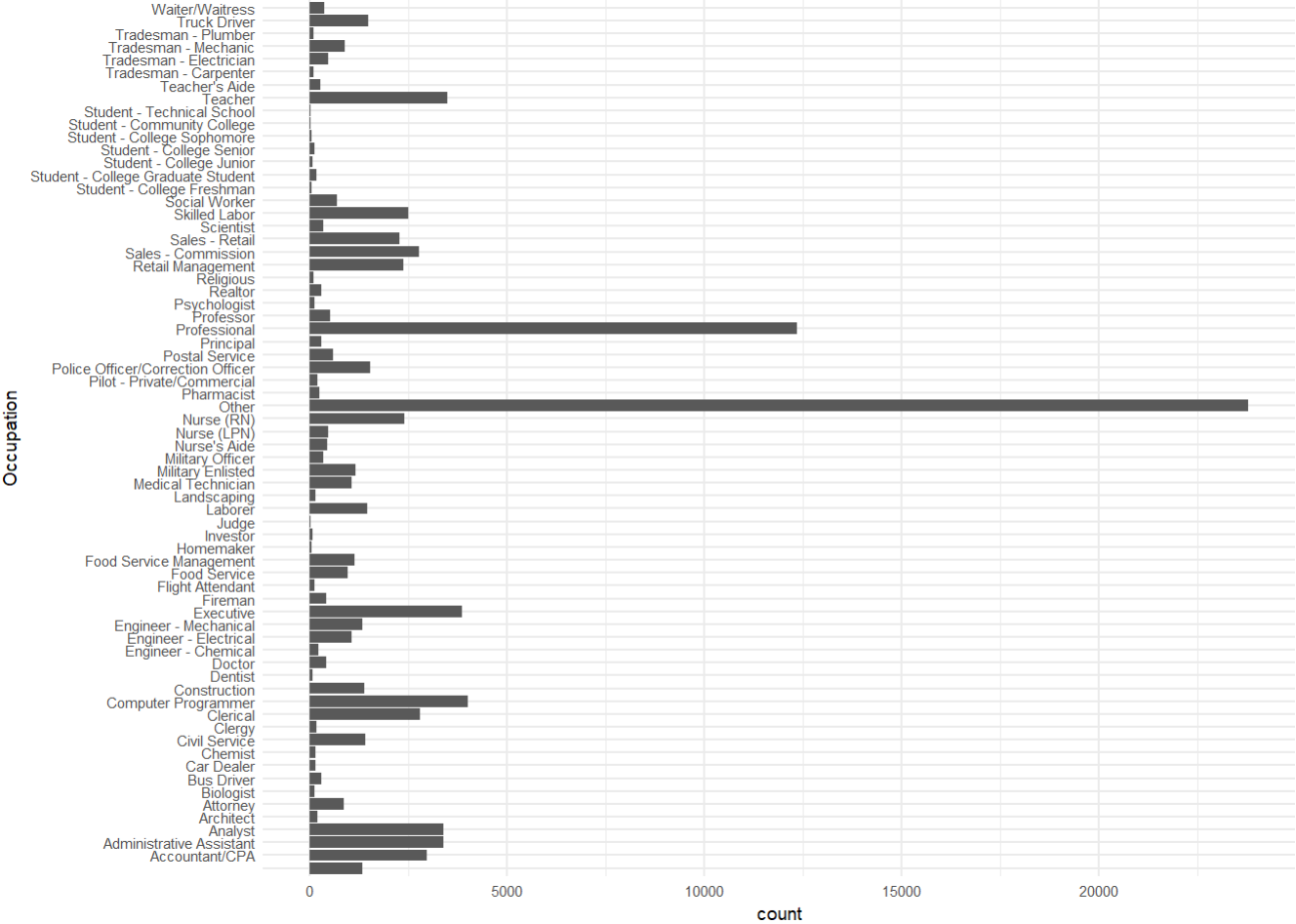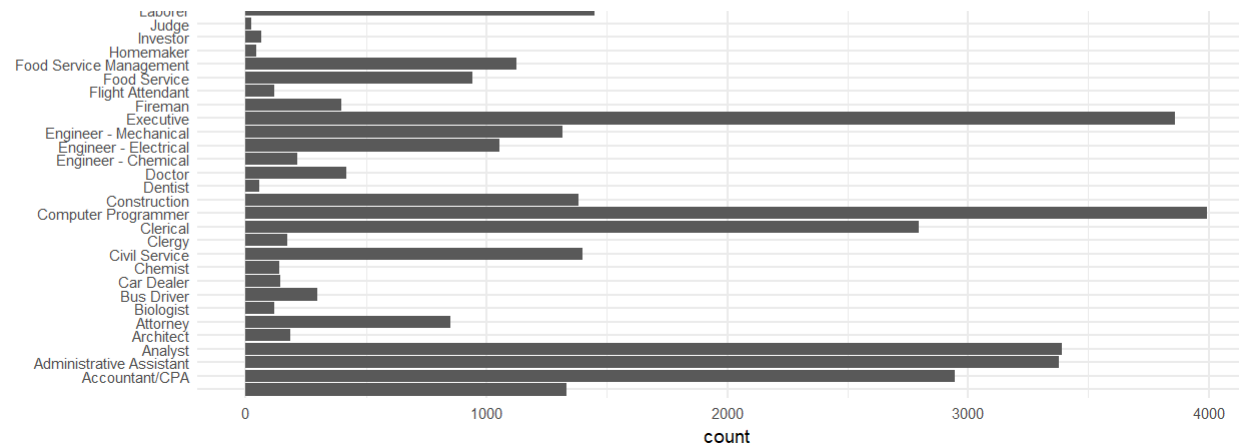
Employment

There are very few applicants that are not employed. It would be difficult to secure a loan without some kind of employment. The unemployed applicants could be students requesting a student loan of some kind that will have a deferred payment arrangement. This is exactly what we were able to show when taking the unique occupations for the unemployed borrowers. Most of them are students.

```
##
##      Employed      Full-time          Other Self-employed      Part-time
##         65896          25590           3526          1092            969
##       Retired   Not employed                  Not available
##           735             95              0             0
```

Loan Data EDA
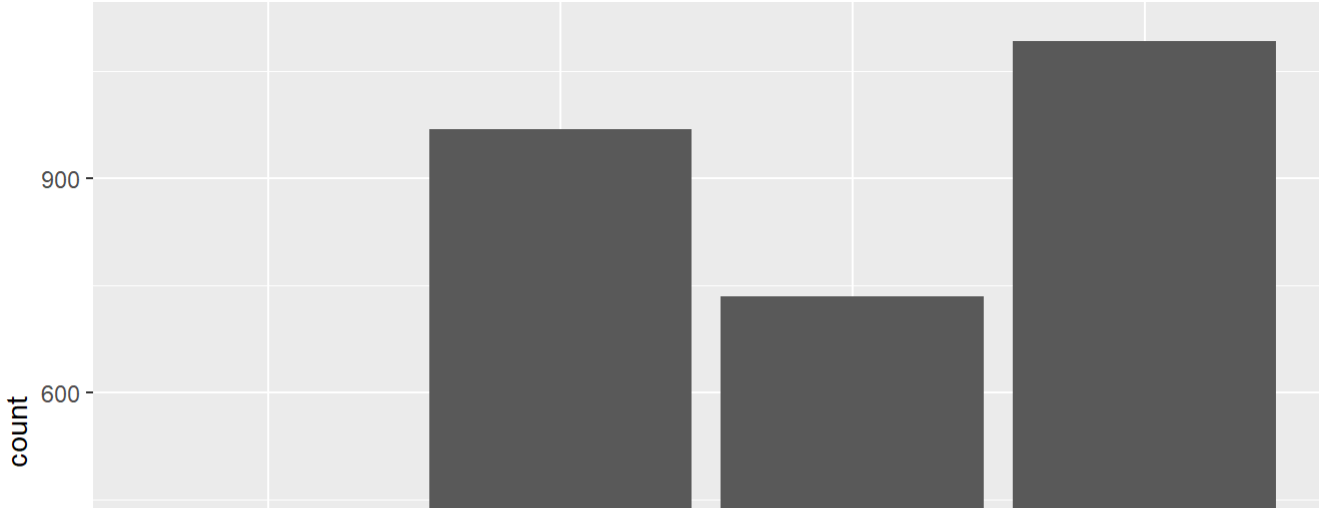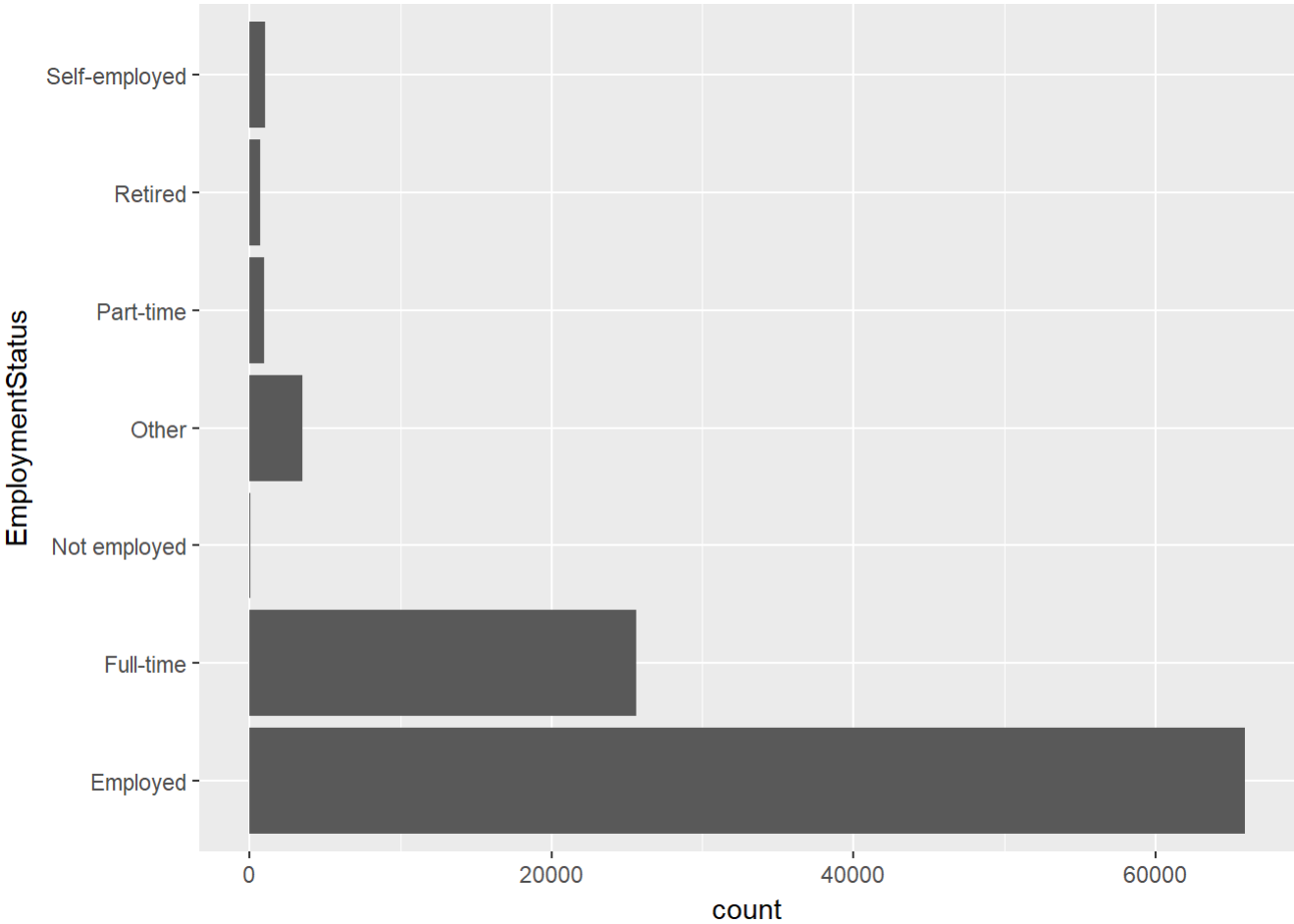
```
##  [1] Other                     Student - College Graduate Student
##  [3] Sales - Commission        Student - Community College
##  [5] Psychologist              Student - College Senior
##  [7] Student - College Junior  Professional
##  [9] Student - College Sophomore  Analyst
## [11] Teacher's Aide            Retail Management
## [13] Homemaker                 Sales - Retail
## [15] Nurse's Aide              Waiter/Waitress
## [17] Student - Technical School  Student - College Freshman
## [19] Skilled Labor
## 68 Levels:  Accountant/CPA Administrative Assistant Analyst ... Waiter/Waitress
```

Credit Score

Most loan applicants have a credit score between 650 and 750, as seen in our histograms and boxplot.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   520.0   660.0   680.0   690.4   720.0   880.0
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   539.0   679.0   699.0   709.4   739.0   899.0
```

Loan Data EDA

Loan Data EDA
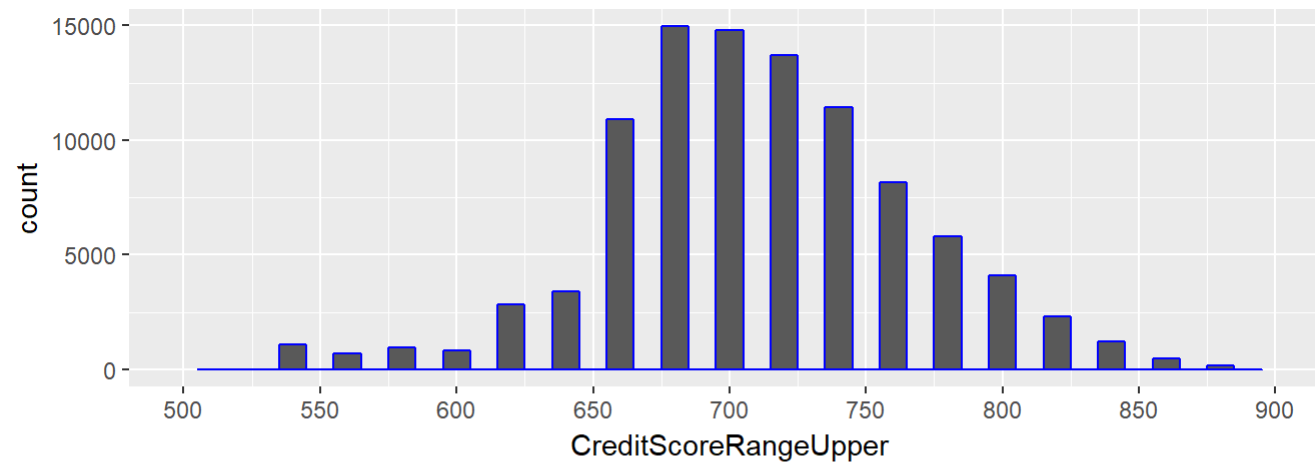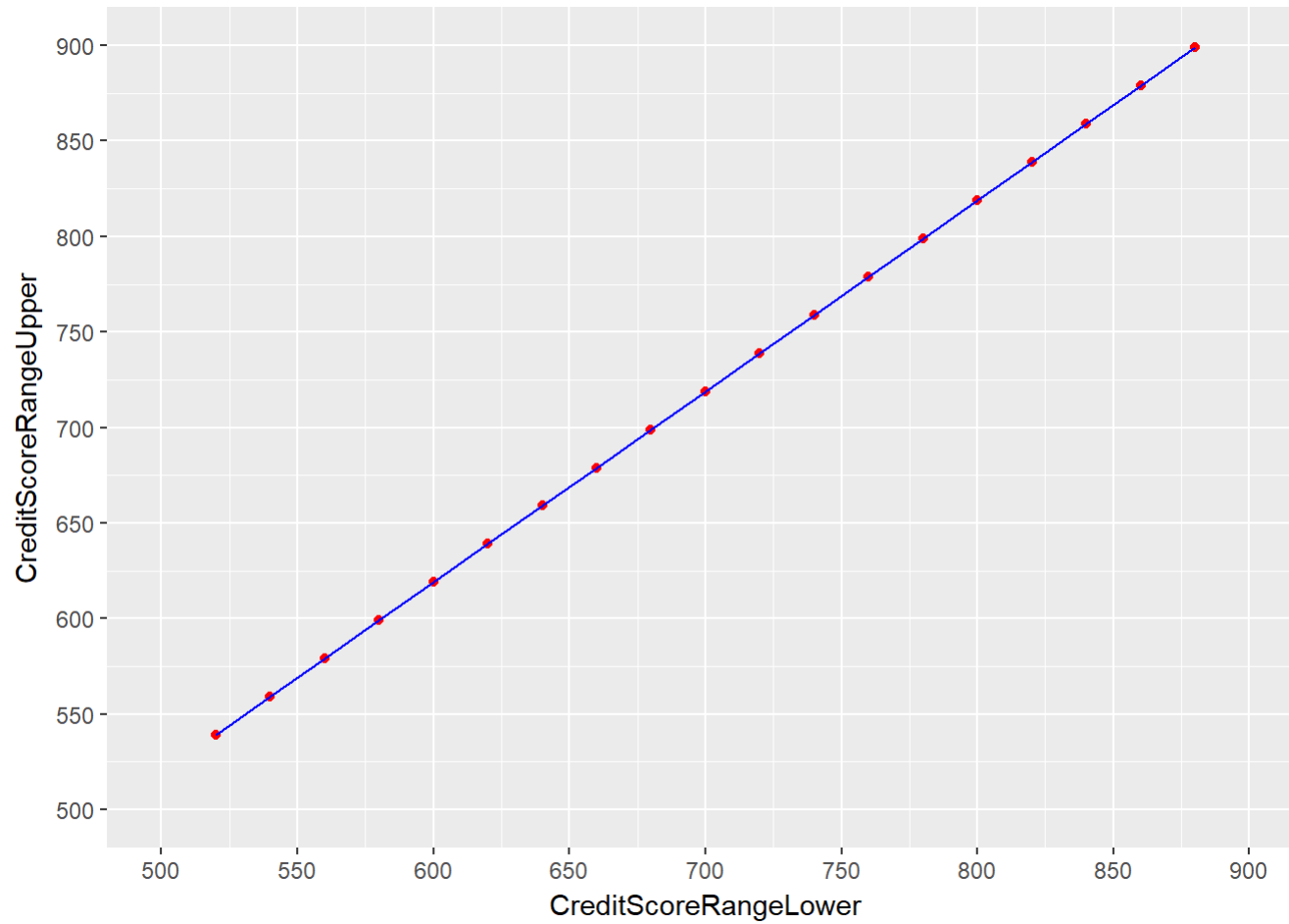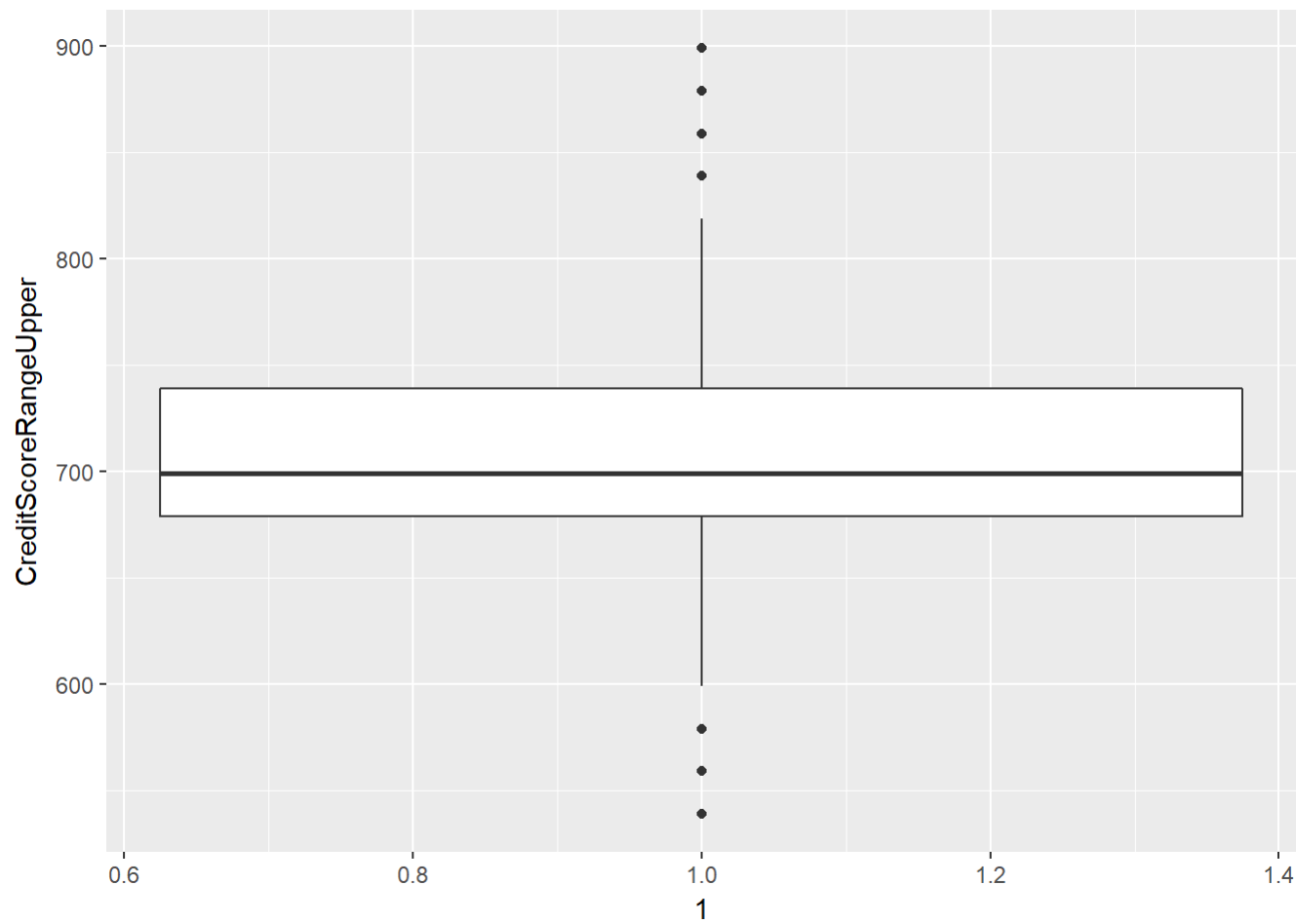
Credit Score Buckets

Here, we will put our upper range credit scores into buckets to simplify for future analysis. Investors/lenders normally put credit scores into 5 categories: Bad, Poor, Fair, Good, and Excellent.
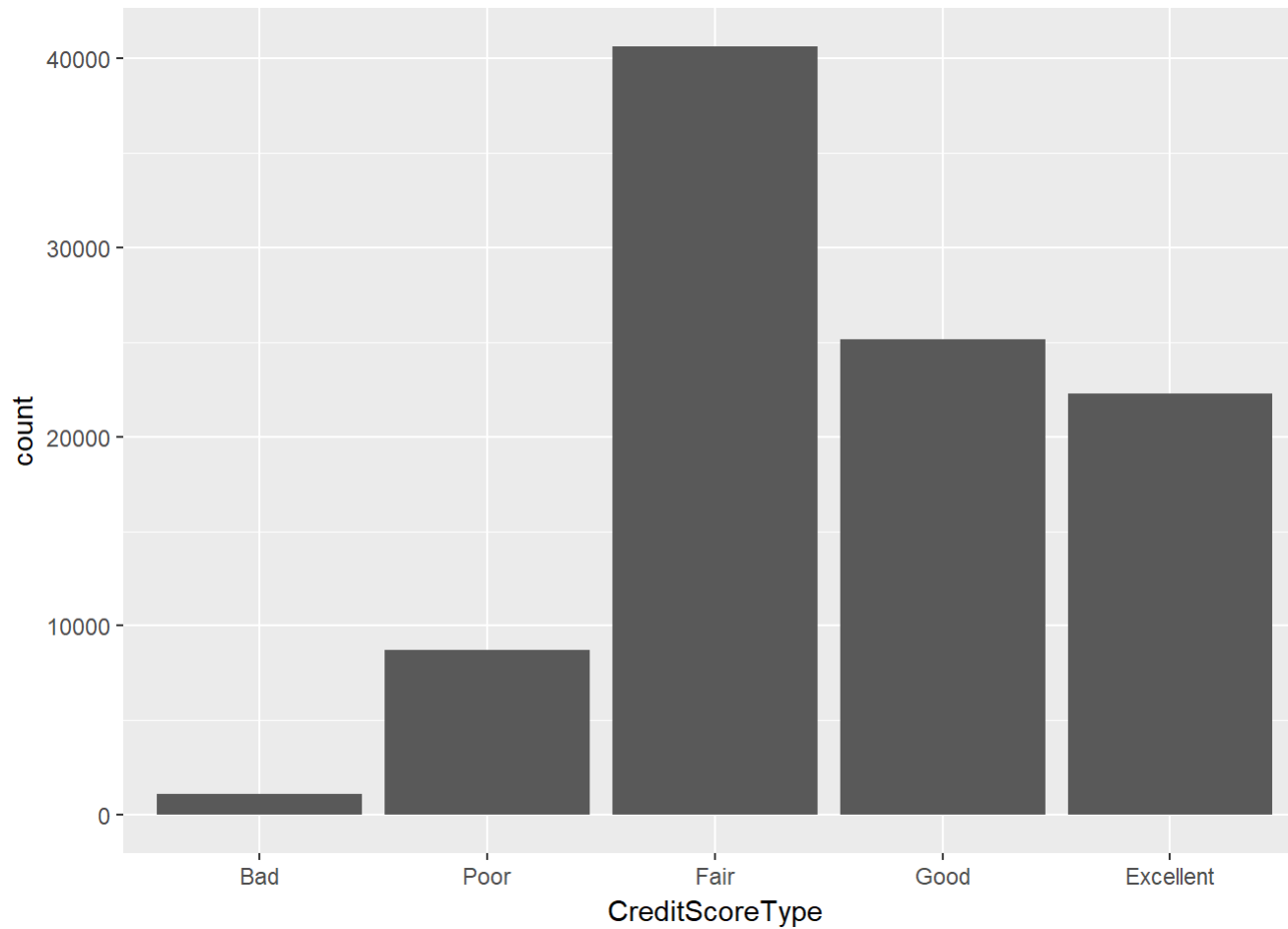
Credit Lines

Most people have around 7 open credit lines.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   6.000   9.000   9.312  12.000  54.000
```
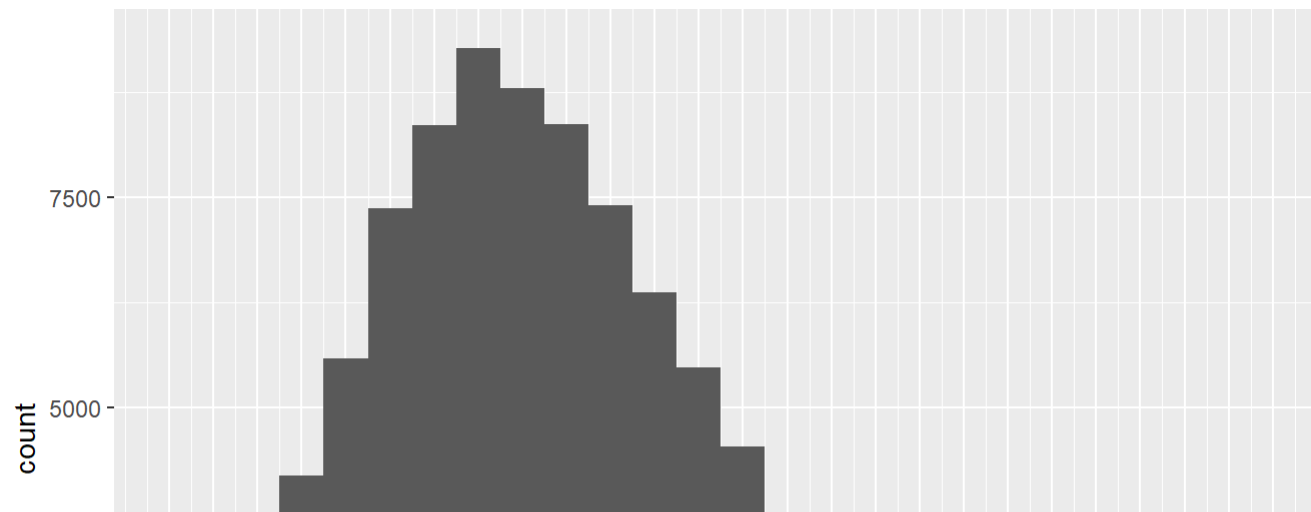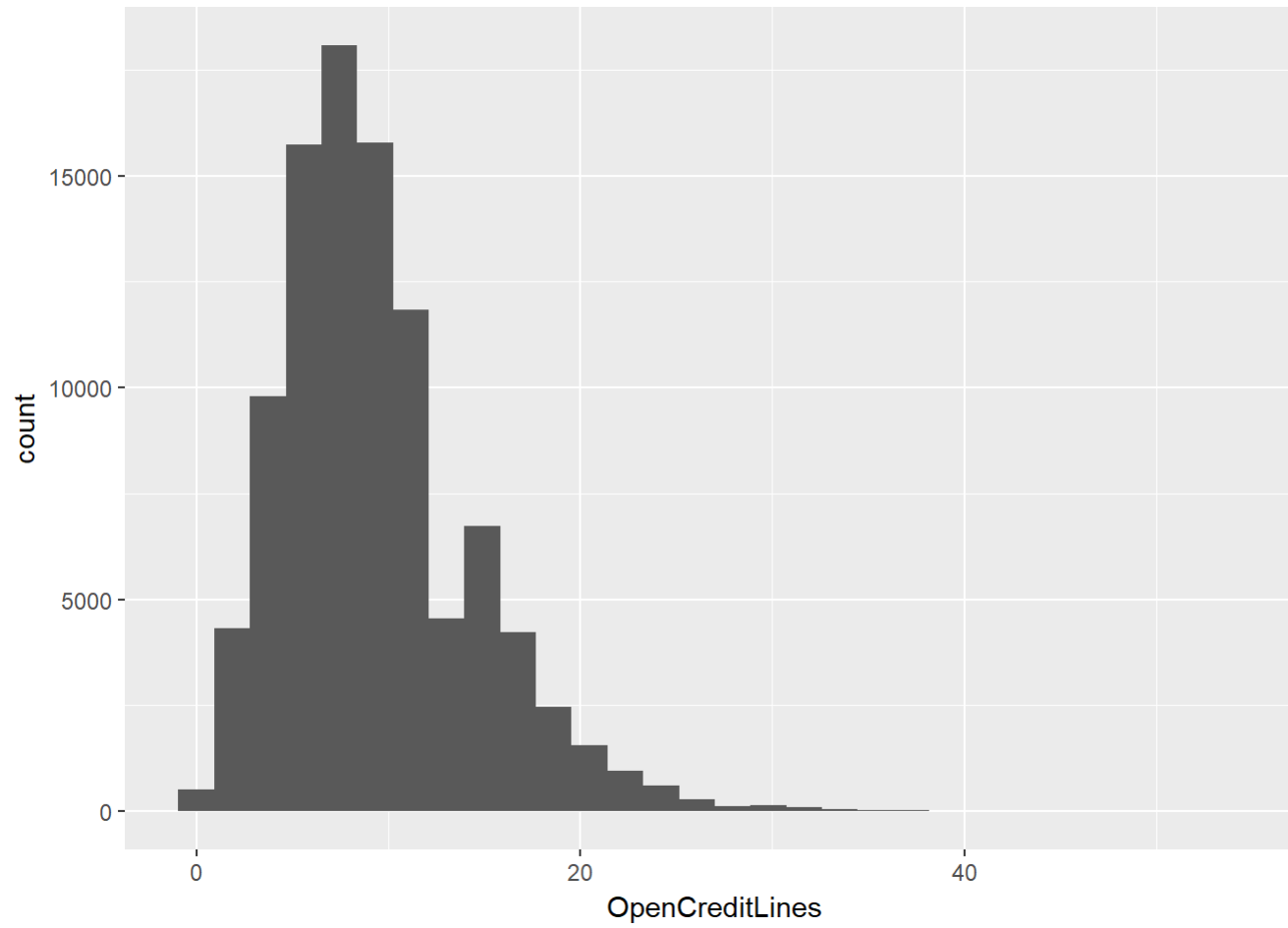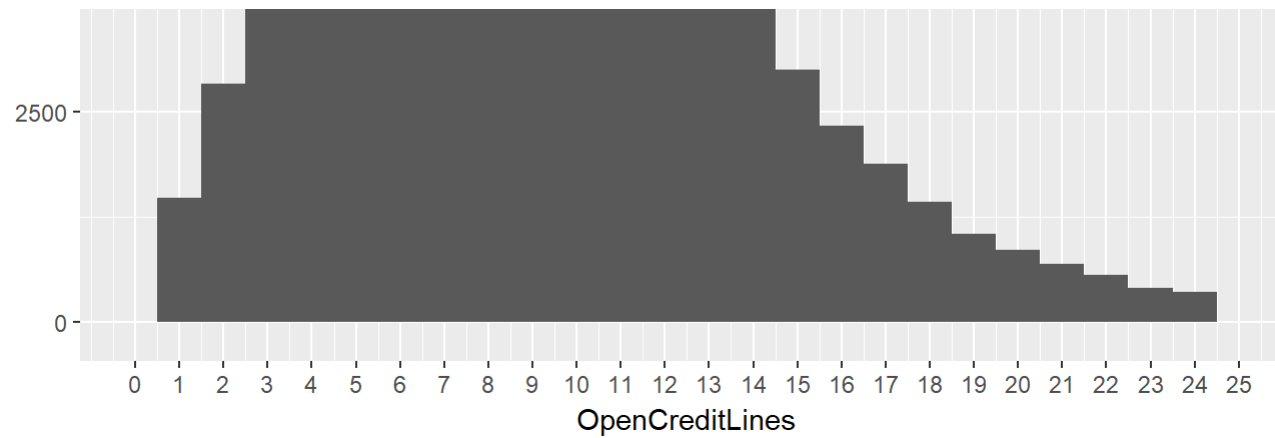
Delinquencies

In order to get a better idea of the number of delinquencies in our set, we'll set the number of delinquencies as factor for our summary. That will give us the number of loans with each particular number of delinquencies. We can also zoom in on a section of the tail with our second plot, and then zoom out again by transforming our data in our third plot and using a boxplot to see outliers in the fourth plot.

```
## 
##     0      1      2      3      4      5      6      7      8      9     10     11
## 79175  10125   3553   1606   1049    593    459    338    251    163    120    107
##    12     13     14     15     16     17     18     19     20     21     22     23
##    85     61     32     35     25     18     17     14     16     12      9      6
##    24     25     26     27     28     30     31     32     33     35     37     40
##     4      3      2      8      2      1      2      3      1      1      1      1
##    41     45     50     51     83
##     1      1      1      1      1
```
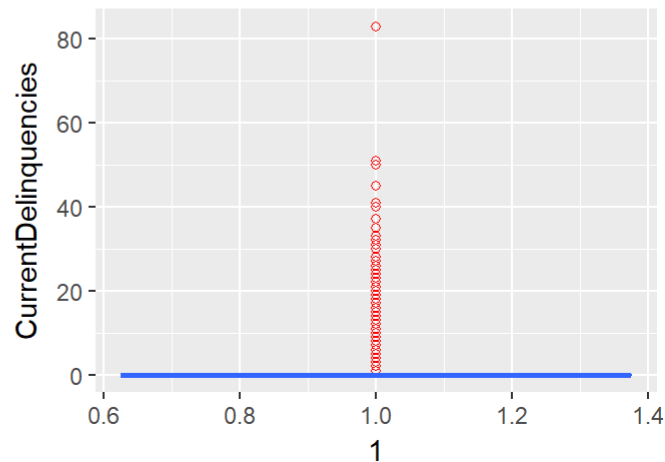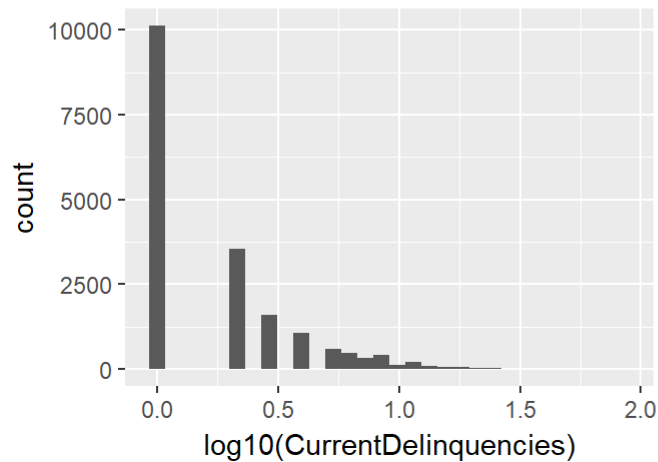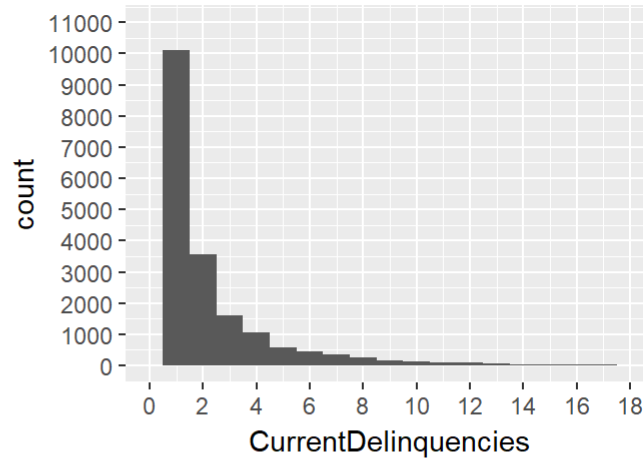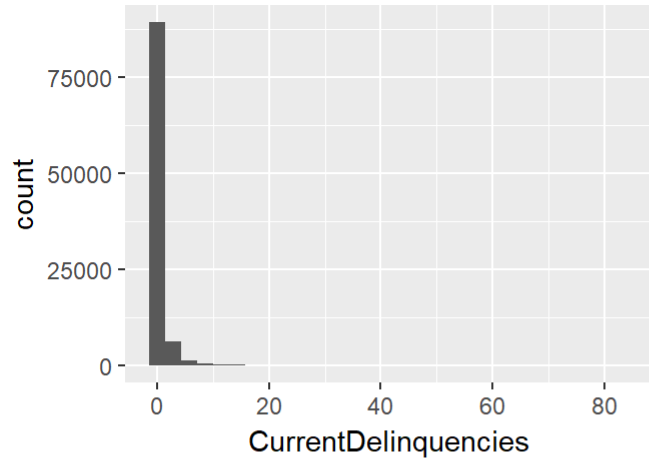
Amount Delinquent

Most accounts are $0 delinquent, so our second set of summary data uses log10 to transform our data, adding 1 to avoid an Inf error. When we filter out the accounts with $0 delinquent, we get 15,524 borrowers (around 15% of all borrowers) who have had a positive delinquent balance.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0       0    1003       0  463881
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4828  0.0000  5.6664
```

```
## [1] 15524
```

Log10(AmountDelinquent)
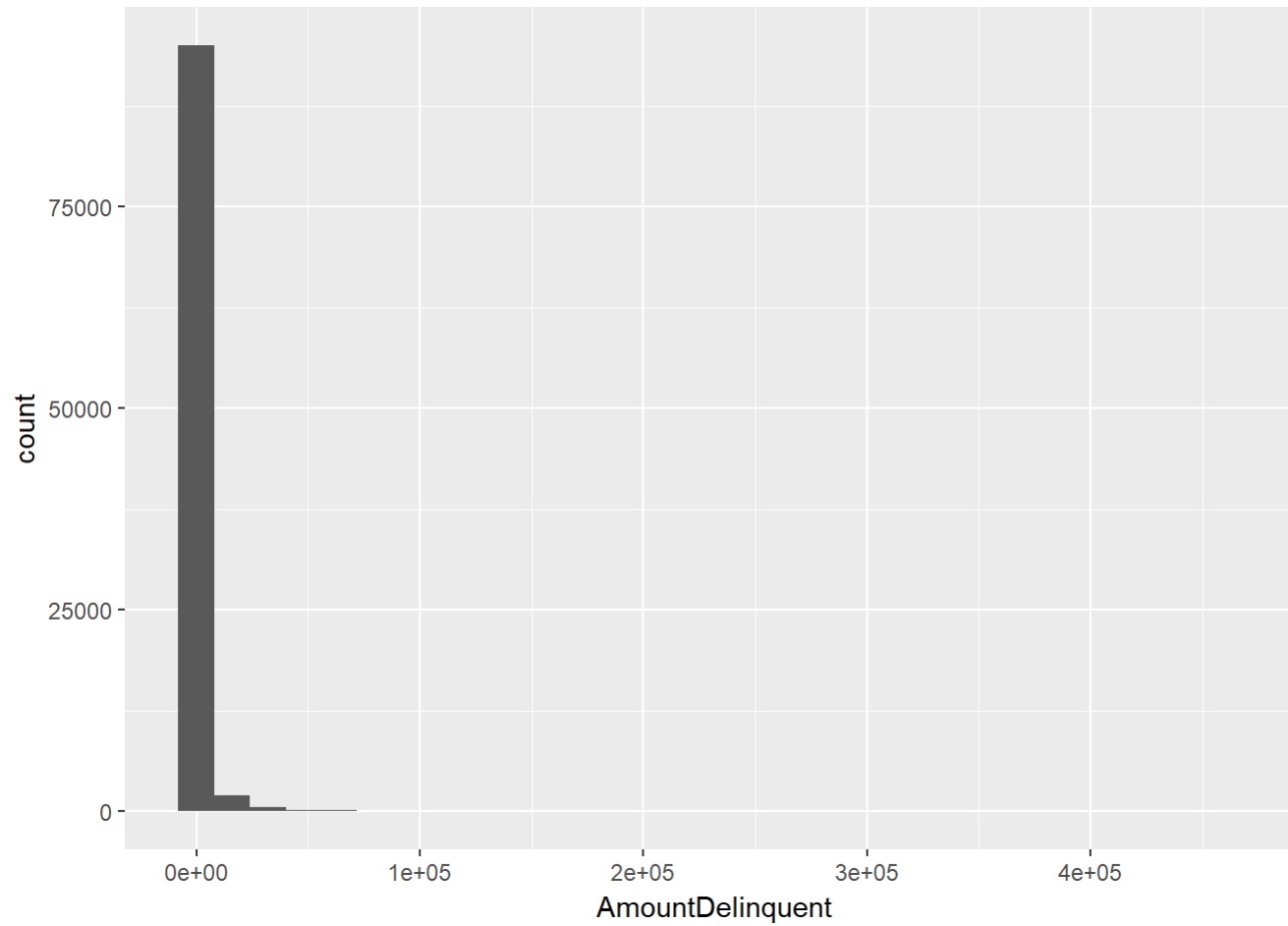
log10(AmountDelinquent)
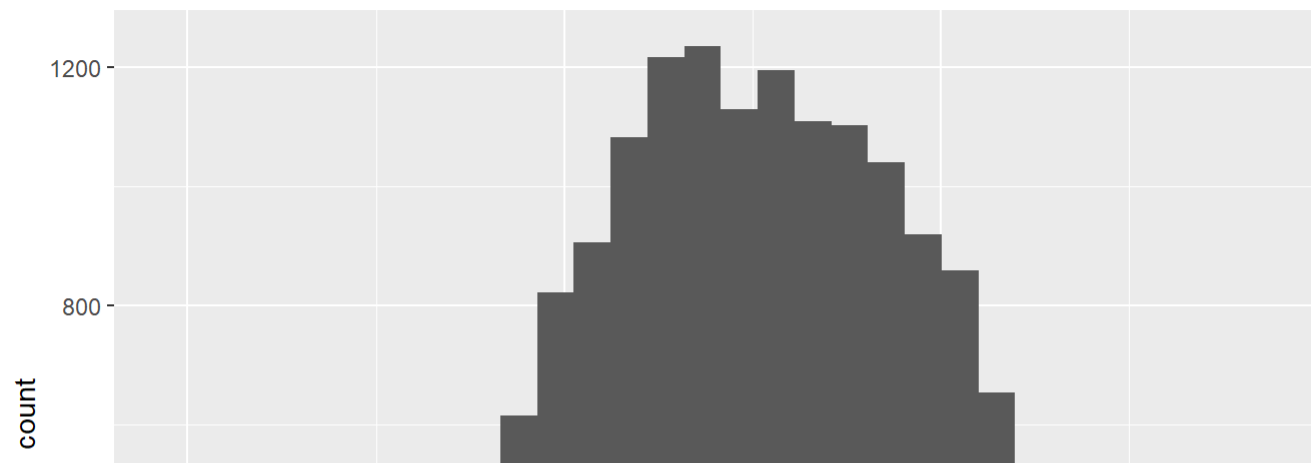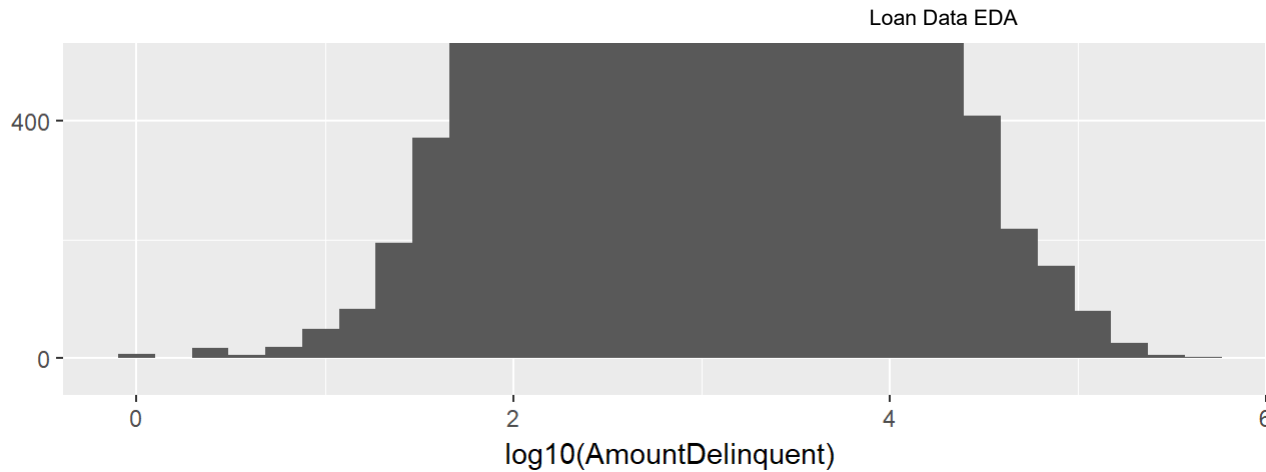
Debt to Income Ratio

In general, those with a higher debt to income ratio have a harder time qualifying for a loan. You can see below that most borrowers stay between a 10 and 30% debt to income ratio. It is very rare to see a DTIR above 50% since most lenders/investors do not give loans to people with DTIRs above 43%. The higher the DTIR, the higher the risk of default. There seems to be one major outlier, a whopping 1001% DTIR! What's going on there? Could be a mistake, or maybe there really is someone with 10x their income in debt. I suppose that's why they need to consolidate. Let's take a look at some of the outliers by filtering anything over 100%(1.0 DTIR.) Many of these accounts show a high number of open credit lines and low Stated Monthly Income (possibly unverified income.)

> To calculate your debt-to-income ratio, you add up all your monthly debt payments and divide them by your gross monthly income. Your gross monthly income is generally the amount of money you have earned before your taxes and other deductions are taken out.

Consumer Finance Website (www.ConsumerFinance.Gov)

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.150   0.220   0.276   0.320  10.010
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.06070 0.08636 0.09628 0.12057 1.04179
```
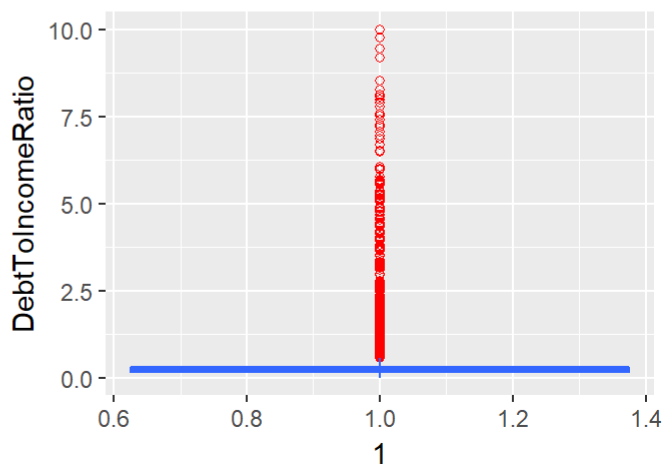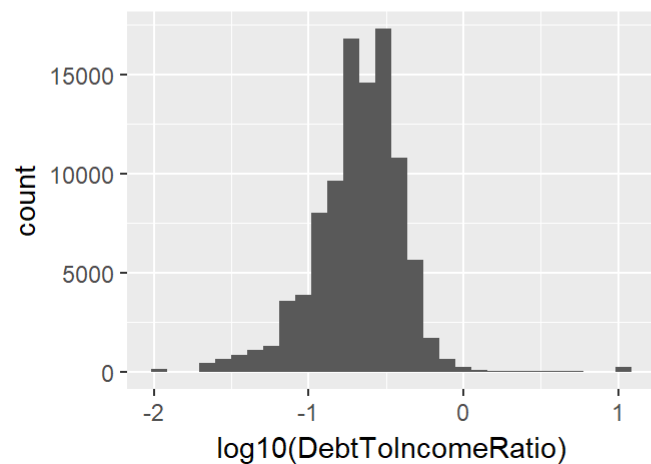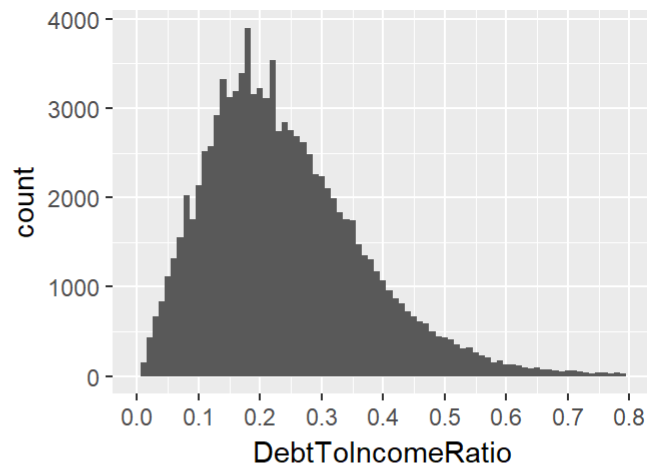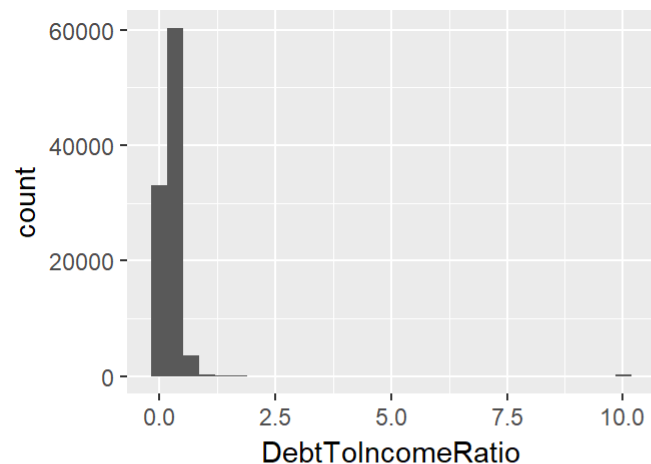
DTIR > 43% (0.43)

There are 8,819 out of 97,903 borrowers with DTIRs greater than 43%. This shows that Prosper is not a traditional lending institution, although the majority of DTIRs are below 43%. In order to spread out the risk of lending to applicants with high DTIRs, they have multiple investors that help give these borrowers a chance to qualify for a loan. 10.01 looks like the maximum DTIR and risk that investors are willing to take on. There are a little over 200 borrowers with a 10.01 DTIR. Later on in our analysis, we'll see what listing category is most popular for these borrowers. I'm guessing it's going to be Loan Consolidation, but I guess we'll have to see.

```
## [1] 8819
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4400  0.4700  0.5200  0.8803  0.6100 10.0100
```
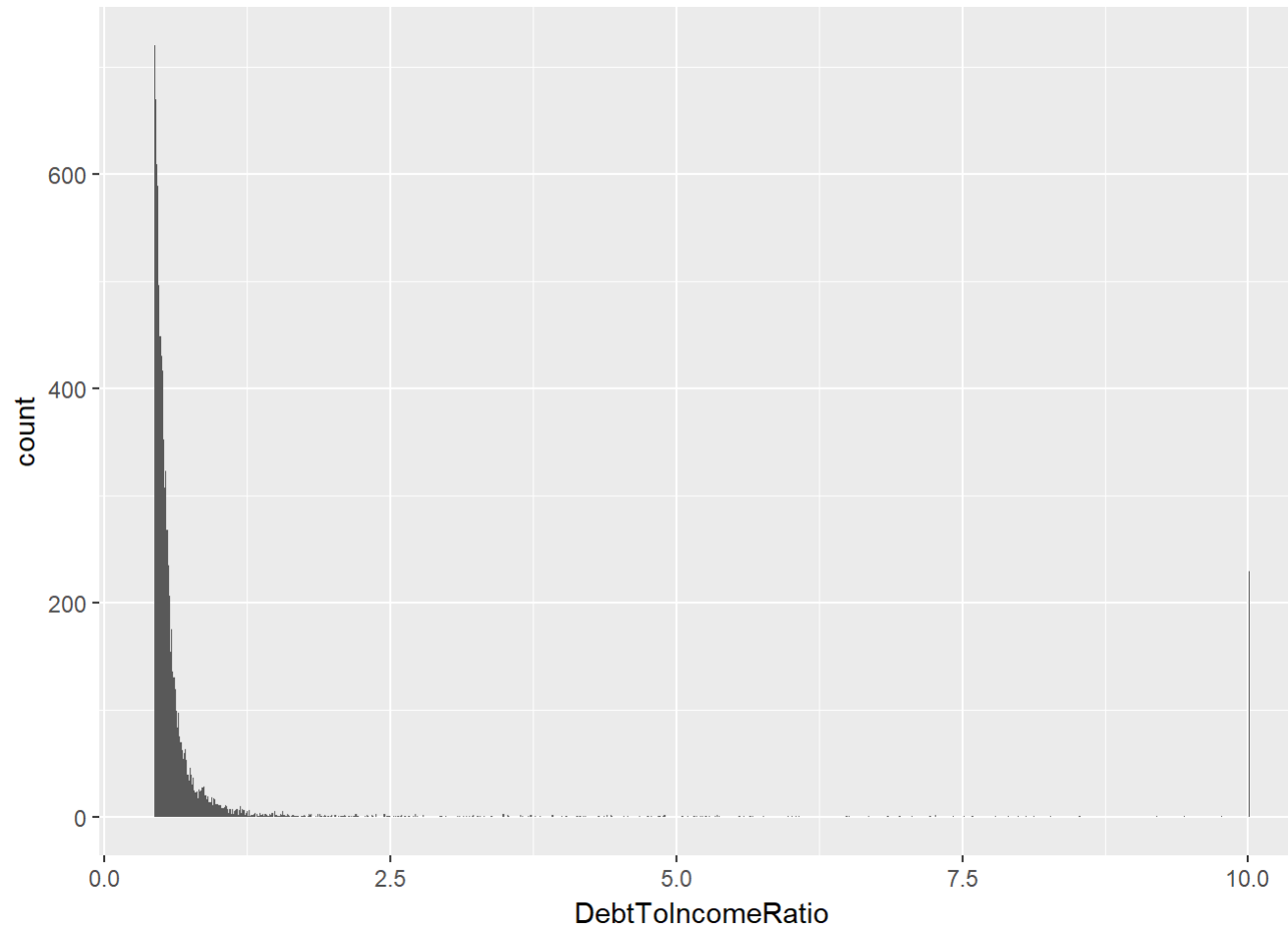
```
## Term                      LoanStatus   BorrowerRate    ListingCategory
## 12:  1    Completed          :99   Min.   :0.0100   0      :173
## 36:220    Chargedoff         :69   1st Qu.:0.1449   1      : 28
## 60:  8    Defaulted          :36   Median :0.1890   3      :  9
##           Current            :23   Mean   :0.2002   2      :  7
##           Past Due (16-30 days): 1  3rd Qu.:0.2700   7      :  6
##           Past Due (61-90 days): 1  Max.   :0.3500   15     :  2
##           (Other)            : 0                     (Other):  4
## BorrowerState                          Occupation
##        :44     Other                        :105
## CA     :31     Sales - Commission           : 11
## FL     :17     Homemaker                    : 10
## IL     :16     Student - College Graduate Student:  9
## GA     :13     Professional                 :  8
## NY     :10     Student - College Senior     :  8
## (Other):98     (Other)                      : 78
##        EmploymentStatus CreditScoreRangeLower CreditScoreRangeUpper
## Self-employed:74     Min.   :520.0      Min.   :539.0
## Full-time    :65     1st Qu.:620.0      1st Qu.:639.0
## Employed     :36     Median :680.0      Median :699.0
## Not employed :27     Mean   :677.2      Mean   :696.2
## Part-time    :17     3rd Qu.:720.0      3rd Qu.:739.0
## Other        : 5     Max.   :860.0      Max.   :879.0
## (Other)      : 5
## OpenCreditLines  CurrentDelinquencies AmountDelinquent  DebtToIncomeRatio
## Min.   : 0.000   Min.   : 0.0000    Min.   :    0.0   Min.   :10.01
## 1st Qu.: 5.000   1st Qu.: 0.0000    1st Qu.:    0.0   1st Qu.:10.01
## Median : 8.000   Median : 0.0000    Median :    0.0   Median :10.01
## Mean   : 8.694   Mean   : 0.5459    Mean   :  734.9   Mean   :10.01
## 3rd Qu.:12.000   3rd Qu.: 0.0000    3rd Qu.:    0.0   3rd Qu.:10.01
## Max.   :30.000   Max.   :20.0000    Max.   :37077.0   Max.   :10.01
##
## StatedMonthlyIncome MonthlyLoanPayment  CreditScoreType
## Min.   :    0.000  Min.   :    0.0    Bad      :10
## 1st Qu.:    0.083  1st Qu.: 108.5     Poor     :51
## Median :    0.083  Median : 209.5     Fair     :69
## Mean   :  110.478  Mean   : 298.7     Good     :45
## 3rd Qu.:    1.417  3rd Qu.: 385.5     Excellent:54
## Max.   :17083.333  Max.   :1047.6
##
```
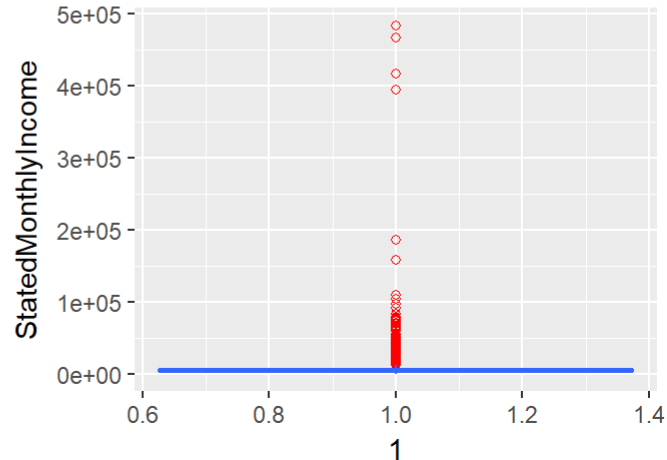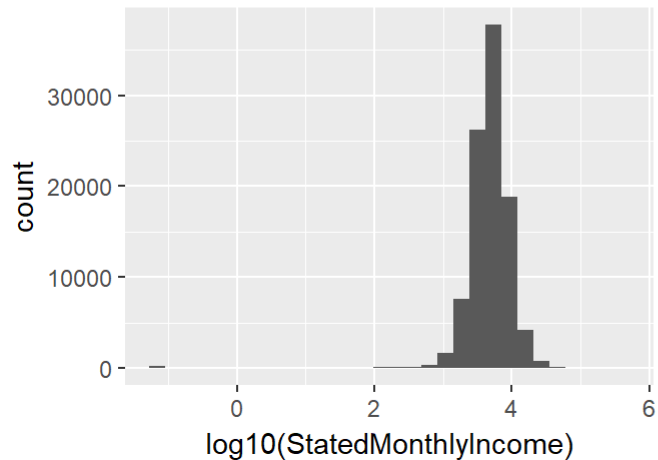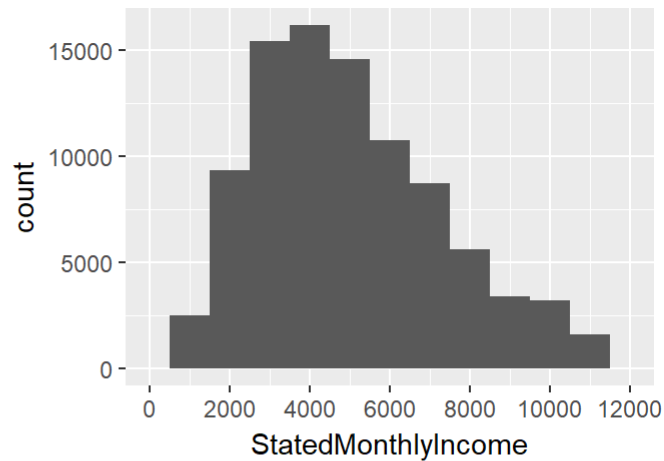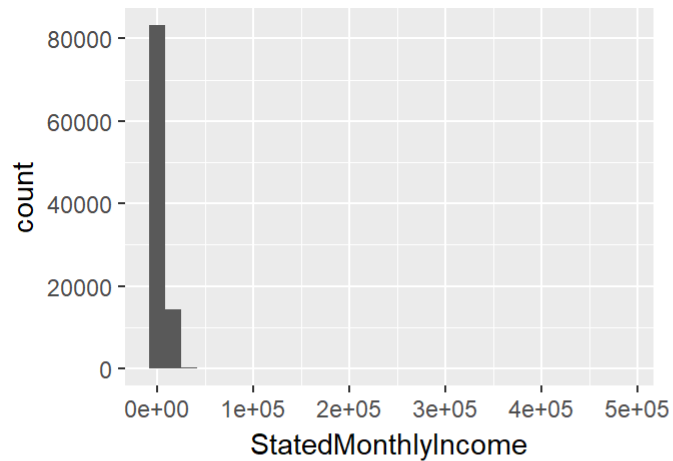
Stated Monthly Income

Most borrowers report having a monthly income somewhere between $2000 and $7000.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    3333    4833    5717    6970  483333
```
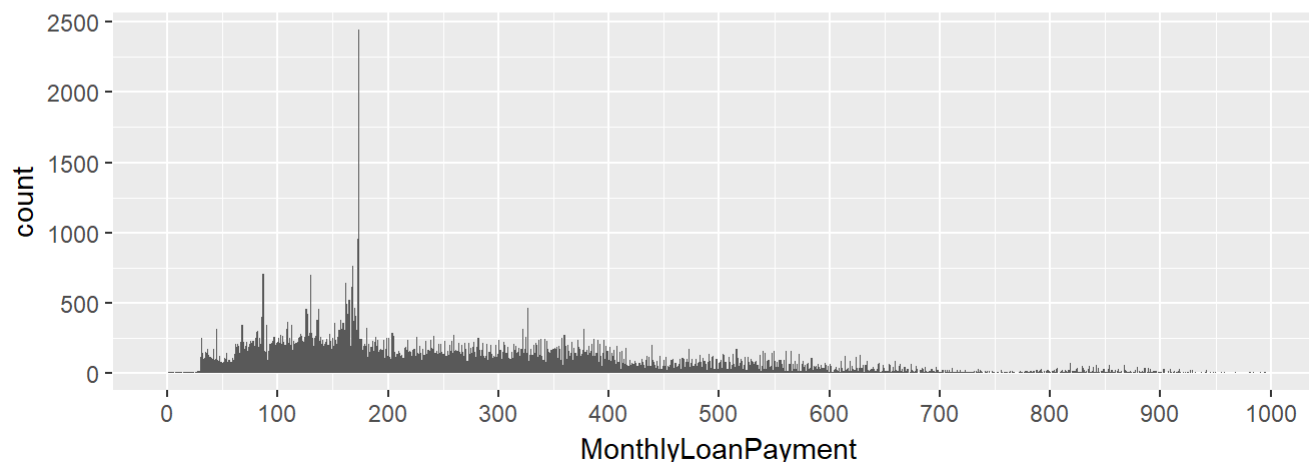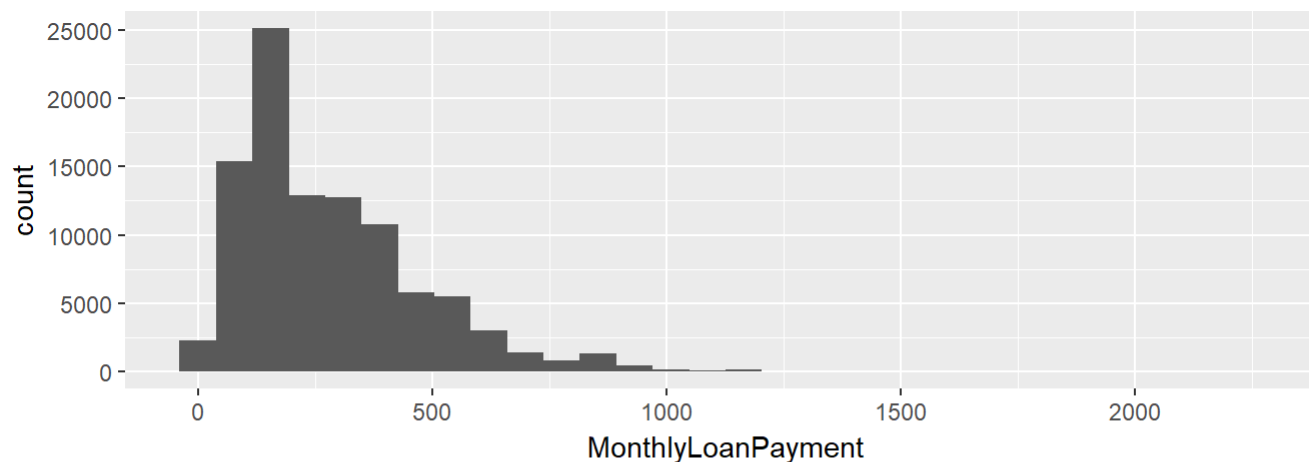
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.523   3.684   3.673   3.843   5.684
```

Loan Payments… Because Who Doesn't Love Loan Payments?

Looks like the majority of loan payments are below $1000/month, and most of those are between $50 and $400/month. When we adjust the scale and binwidth, we can see the spike in loan count around 175.

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   139.3   232.5   281.5   379.4  2251.5
```


Summary of Univariate Plots and Focus of Analysis

The Interest is in the Interest!

Our main focus will be on exploring the impact of any given variable on the interest rate that each borrower is given. For example, how does a person's credit score or debt-to-income ratio(DTIR) impact the rate? Will someone living in CA get the same interest rate as someone with the same stats living in TX? Will a person's occupation or income make a difference?

So far, we've observed a few things

- 75% of our loans have a 36 month term
- Around 3,000 accounts are in default
- Most borrowers have an interest rate between 10 and 20%, although there are also quite a few at 32%
- A little over 50% of loans are for debt consolidation
- CA has the highest number of borrowers, followed by FL, TX, and NY
- Aside from "Other" and "Professional", "Computer Programmer" and "Executive" are the top occupations
- Most borrowers are employed, around 1% are self-employed, 0.7% are retired, and less than 0.1% are unemployed
- Most credit scores are between 650 and 750
- Most borrowers have around 7 open credit lines
- A little over 5% of loans have been in default more than twice
- About 15% of our borrowers have had a positive "Amount Delinquent"
- Most borrowers have a DTIR somewhere between 10% and 30%
- Most reported income is between $2,000 to $7,000 per month
- The majority of monthly payments are under $1,000 with a count spike in loans with a $175/month payment.
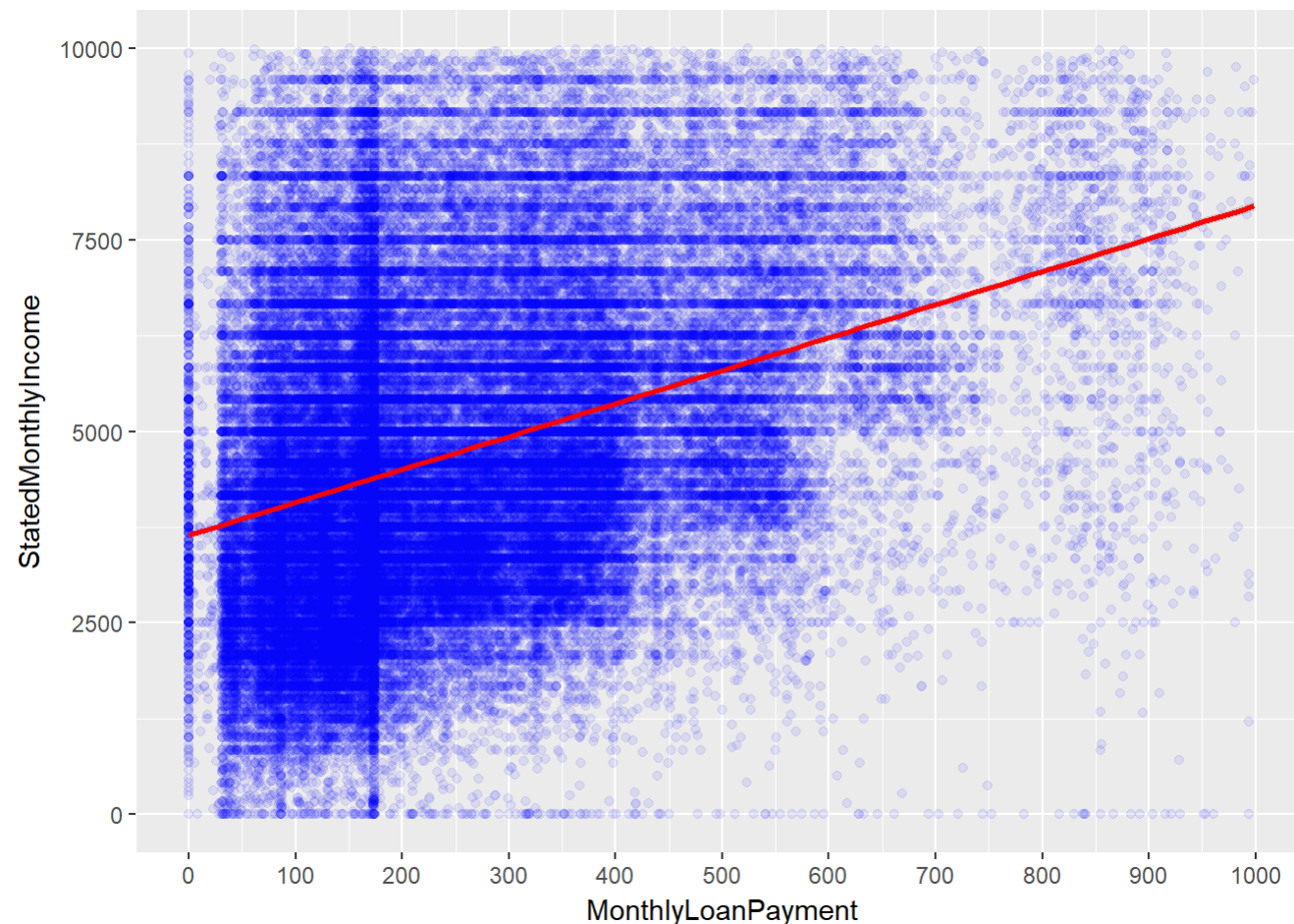
Bivariate Plots

## Loan Payments < $1200/month with Income < $20,000/month
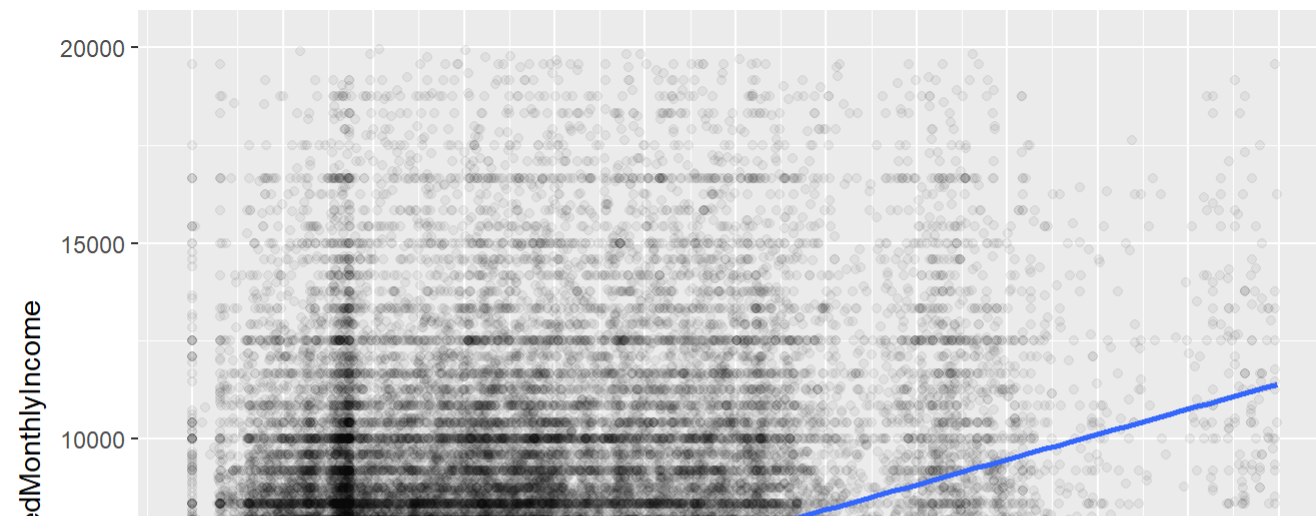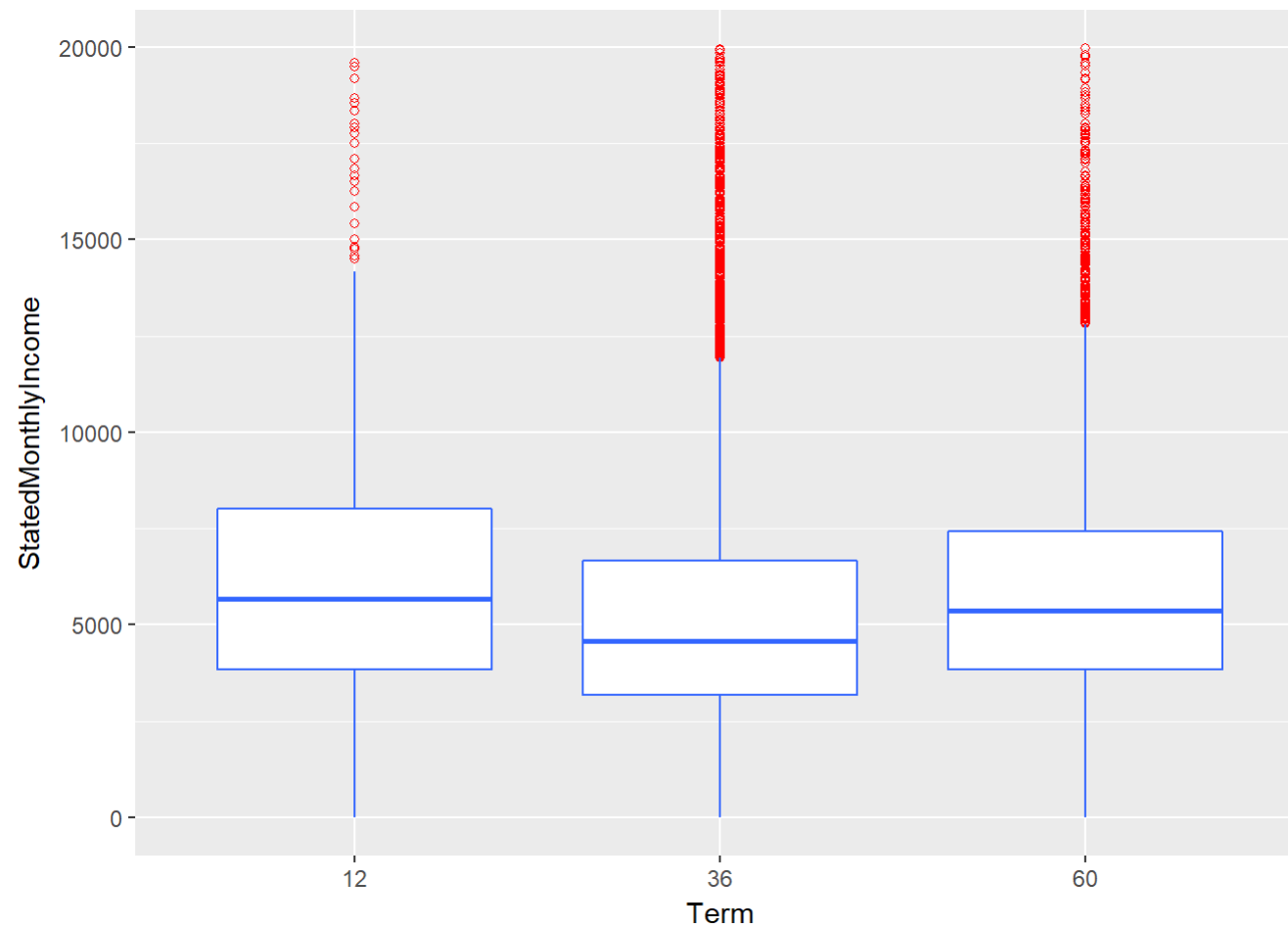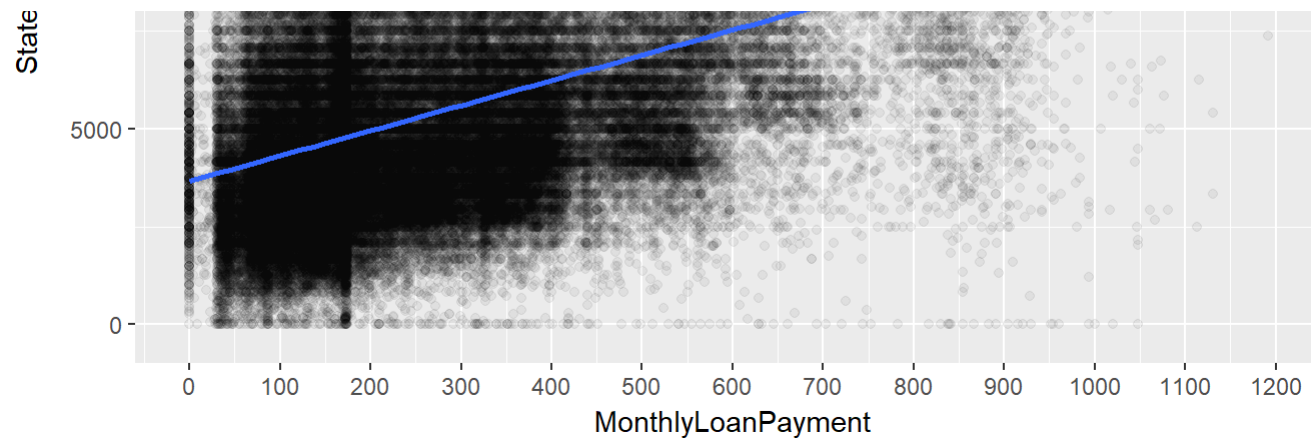
Loan Payments < $1000/month with Income < $10,000/month

A $175/month loan payment looks like the most common across all income levels. However, there is a clear trend upward since those with higher loan payments generally have a higher income.

Rate/Credit

You can see how credit score can make a big difference when it comes to the borrower rate in the plots below. However, there are still quite a few outliers we could examine. If we create a subset of borrowers with Credit Scores above 780 and a Rate above 0.25, we get 239 loans/borrowers. After summarising and plotting all of the variables looking for one that could give us some insight into what is going on here, I came up short. There really wasn't any one variable that stood out as being the reason for such high rates for what seems to be credit worthy borrowers. Maybe something would stand out in a further analysis of all 81 variables from our original dataset.
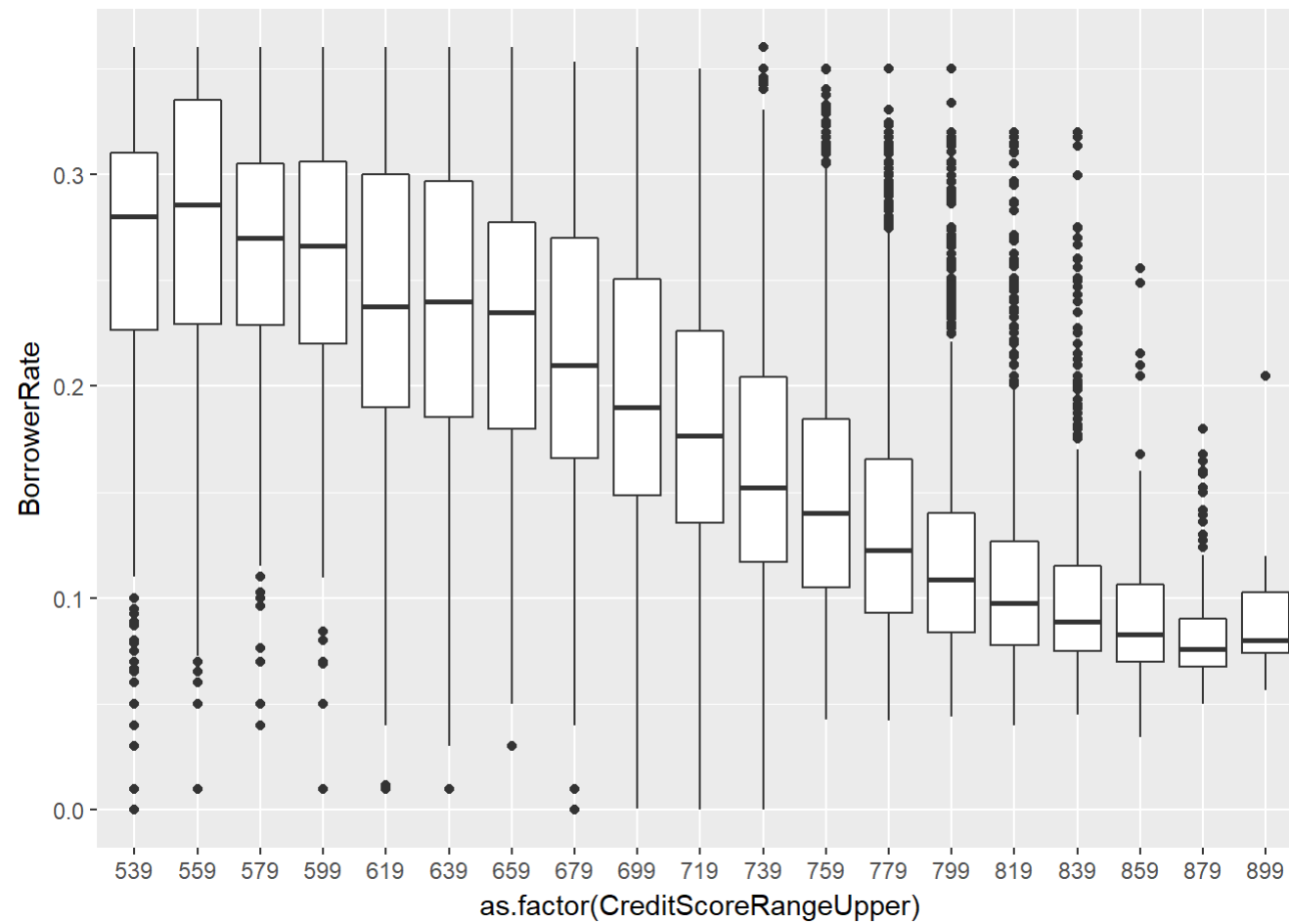
```
##    Term                    LoanStatus     BorrowerRate     ListingCategory
##    12:  2    Current               :100   Min.   :0.2506   1       :100
##    36:177    Completed             : 90   1st Qu.:0.2640   7       : 47
##    60: 60    Chargedoff            : 24   Median :0.2870   2       : 36
##              Defaulted             : 12   Mean   :0.2878   3       : 25
##              Past Due (1-15 days)  :  7   3rd Qu.:0.3149   13      :  6
##              Past Due (91-120 days):  3   Max.   :0.3500   15      :  6
##              (Other)               :  3                    (Other): 19
##    BorrowerState                 Occupation    EmploymentStatus
##    CA     : 29   Other                 : 71   Employed :195
##    TX     : 18   Professional          : 19   Other    : 23
##    FL     : 17   Administrative Assistant: 12  Full-time: 18
##    NY     : 16   Teacher               : 12   Retired  :  2
##    IL     : 13   Sales - Retail        : 11   Part-time:  1
##    MD     : 13   Computer Programmer   : 10            :  0
##    (Other):133   (Other)               :104   (Other)  :  0
##    CreditScoreRangeLower CreditScoreRangeUpper OpenCreditLines
##    Min.   :780.0         Min.   :799.0         Min.   : 1.000
##    1st Qu.:780.0         1st Qu.:799.0         1st Qu.: 5.000
##    Median :780.0         Median :799.0         Median : 8.000
##    Mean   :787.5         Mean   :806.5         Mean   : 8.828
##    3rd Qu.:800.0         3rd Qu.:819.0         3rd Qu.:11.000
##    Max.   :840.0         Max.   :859.0         Max.   :26.000
##
##    CurrentDelinquencies AmountDelinquent  DebtToIncomeRatio
##    Min.   :0.0000       Min.   :    0.0   Min.   : 0.0200
##    1st Qu.:0.0000       1st Qu.:    0.0   1st Qu.: 0.1800
##    Median :0.0000       Median :    0.0   Median : 0.2500
##    Mean   :0.2008       Mean   :  682.4   Mean   : 0.4333
##    3rd Qu.:0.0000       3rd Qu.:    0.0   3rd Qu.: 0.3950
##    Max.   :4.0000       Max.   :72302.0   Max.   :10.0100
##
##    StatedMonthlyIncome MonthlyLoanPayment  CreditScoreType
##    Min.   :    4.333   Min.   :  41.91    Bad      :  0
##    1st Qu.: 3166.667   1st Qu.: 171.92    Poor     :  0
##    Median : 4333.333   Median : 204.10    Fair     :  0
##    Mean   : 5091.442   Mean   : 255.99    Good     :  0
##    3rd Qu.: 6250.000   3rd Qu.: 323.94    Excellent:239
##    Max.   :30416.667   Max.   :1001.28
##
```
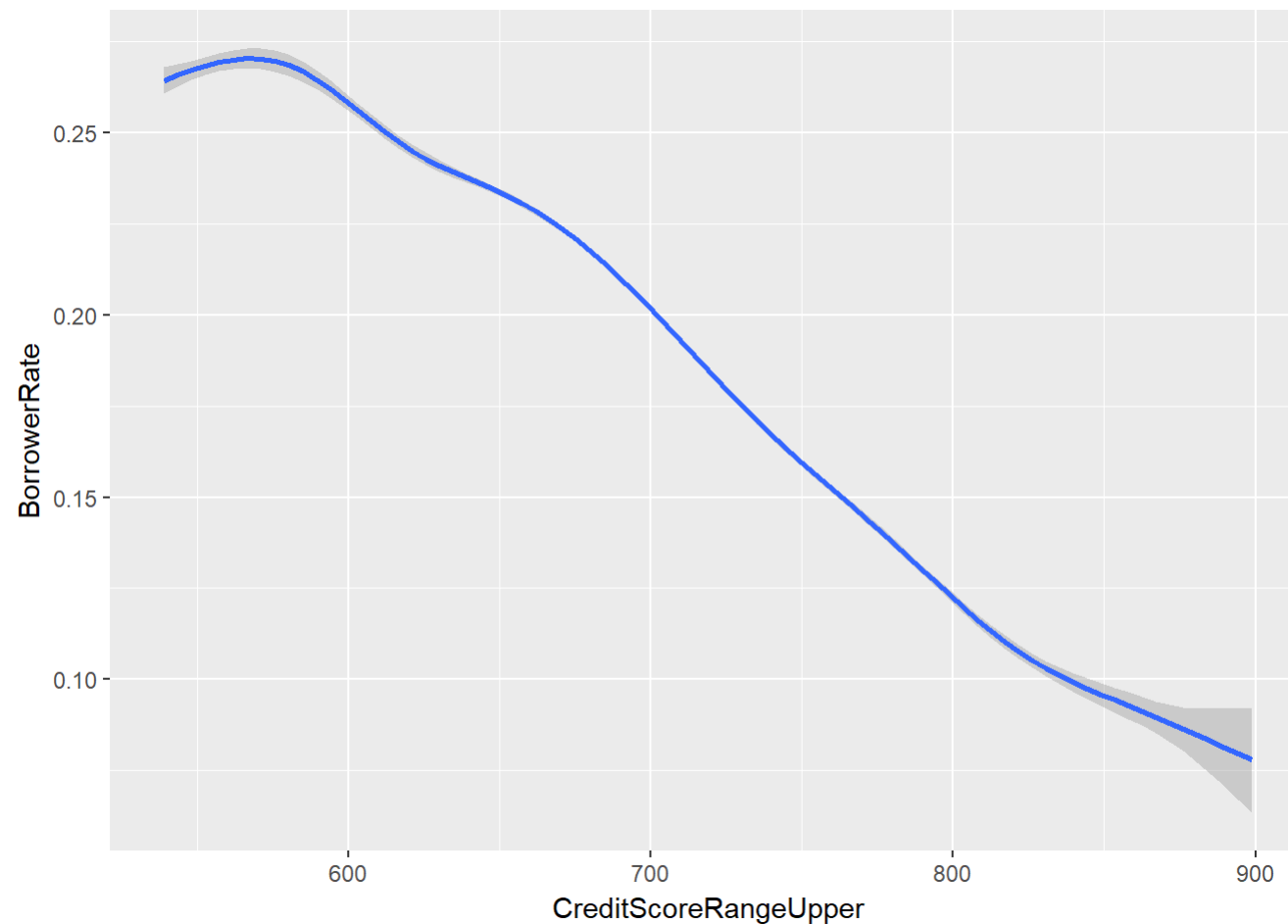
DTIR/Term

In the boxplots below, we can see the DTIR (less than 1%) summaries for each loan term. As expected, those with the least amount of debt tend to take out the short term loans, since they can afford the higher payments per month. Those that have higher DTIRs tend to try and keep their monthly payments as low as possible with a longer term.

DTIR/Term
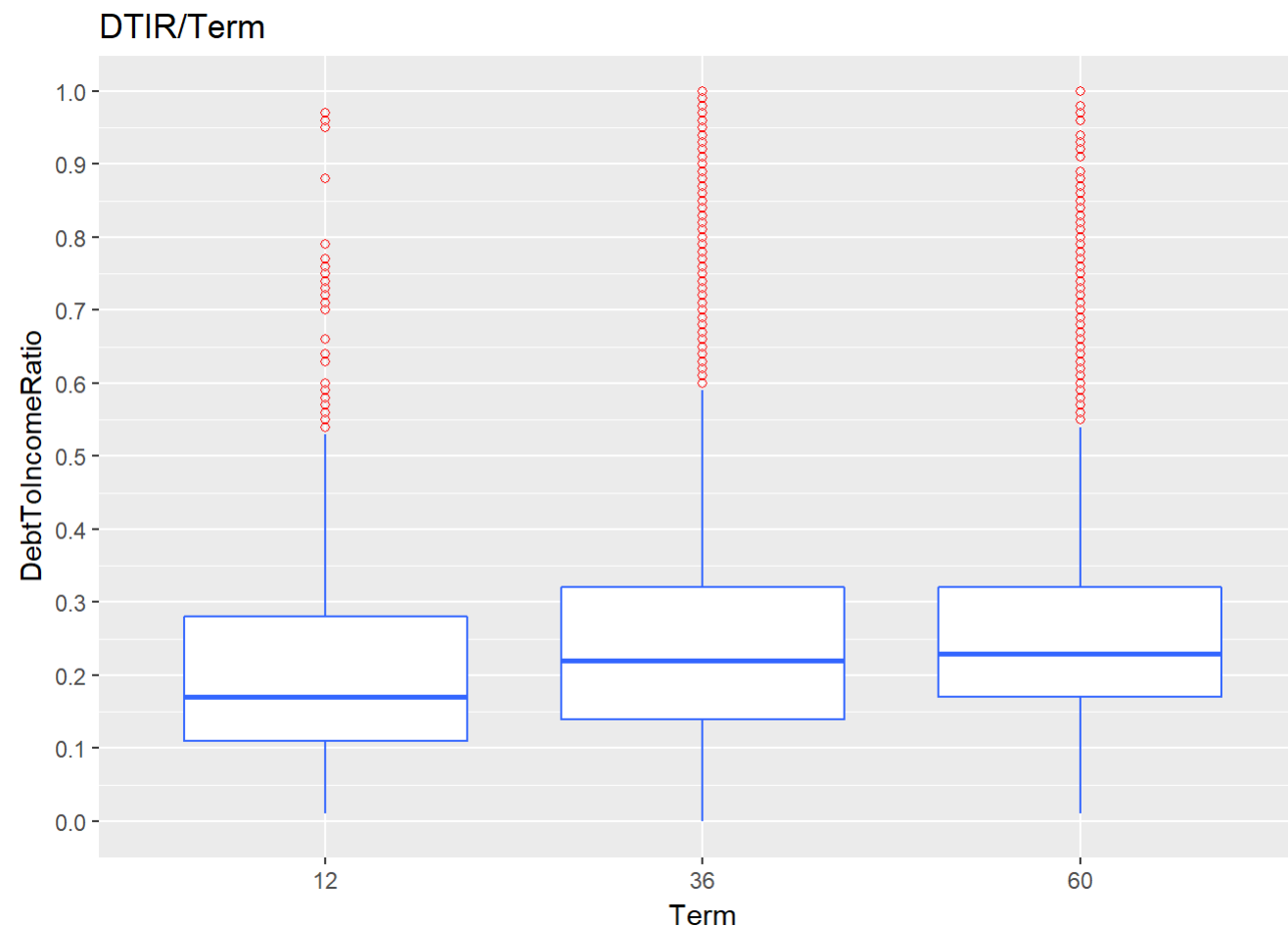
High DTIR

Let's take a further look at our high DTIR borrowers (almost 9,000!) Not surprisingly, "Debt Consolidation" is the most popular category. "Not Available" and "Other" are really just unknown categories, so the other popular categories are "Home Improvement" and "Business". This isn't really any different than our analysis of all DTIRs. About 56% of all loans are for debt consolidation and about 59% of high DTIR loans are for debt consolidation. That's only slightly higher. So, not really much to glean from this particular graph.

0 - Not Available

1 - Debt Consolidation

2 - Home Improvement

3 - Business

4 - Personal Loan

5 - Student Use

6 - Auto

7 - Other

8 - Baby and 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

```
## [1] 8819
```
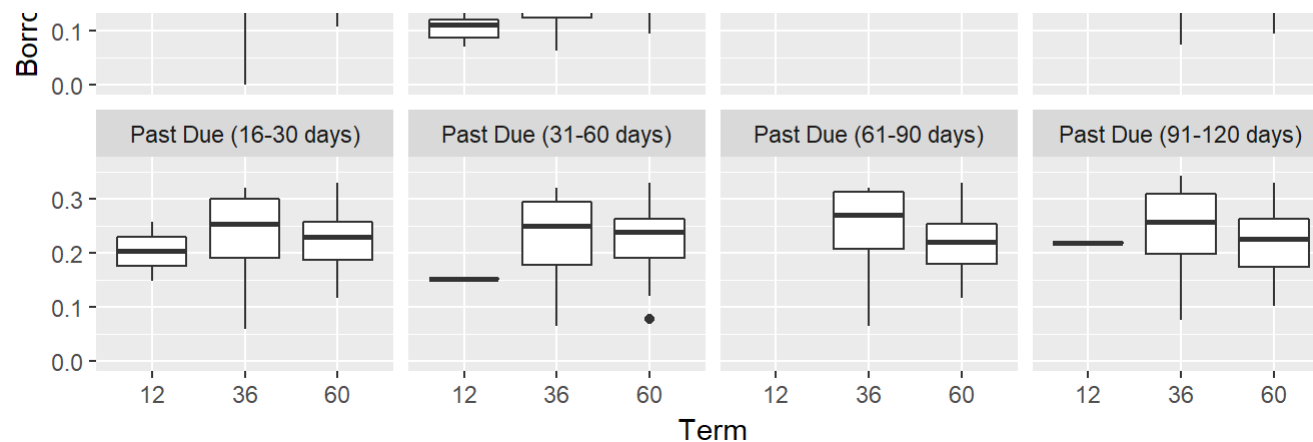
DTIRs Above 43%
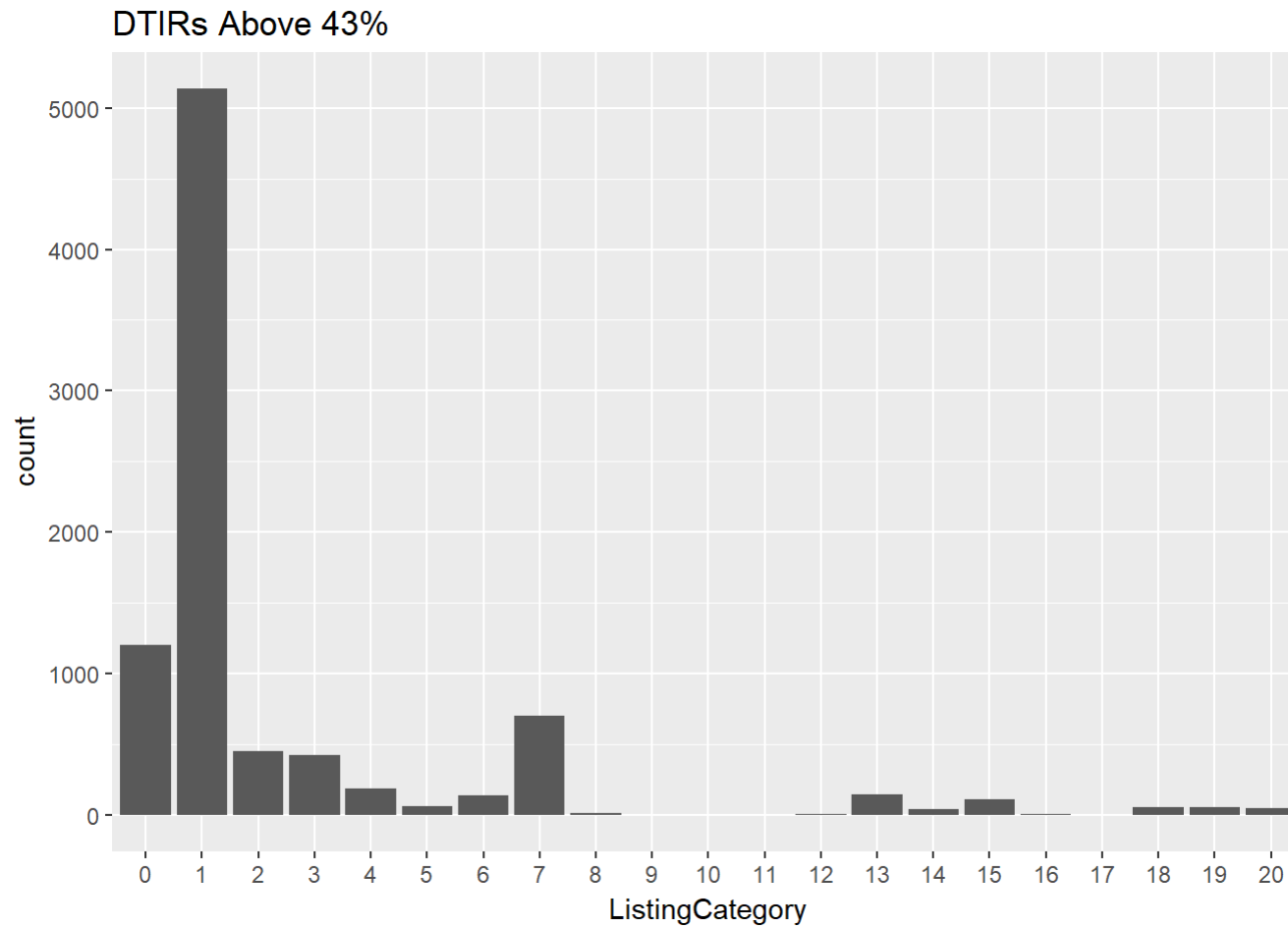

```
## [1] 0.589636
```

Multivariate Plots

## Boxplots
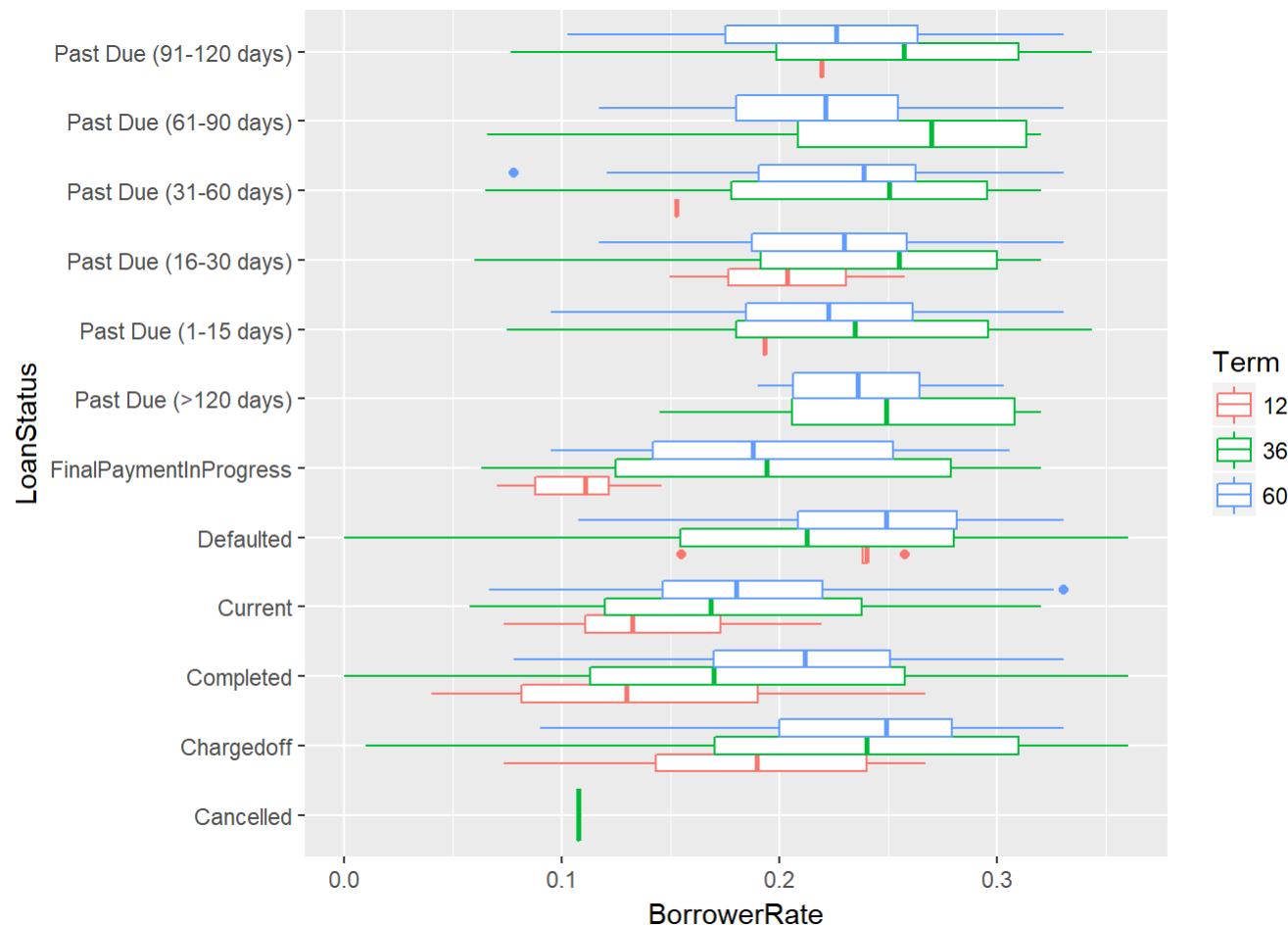
## Loan Status Rates per Term
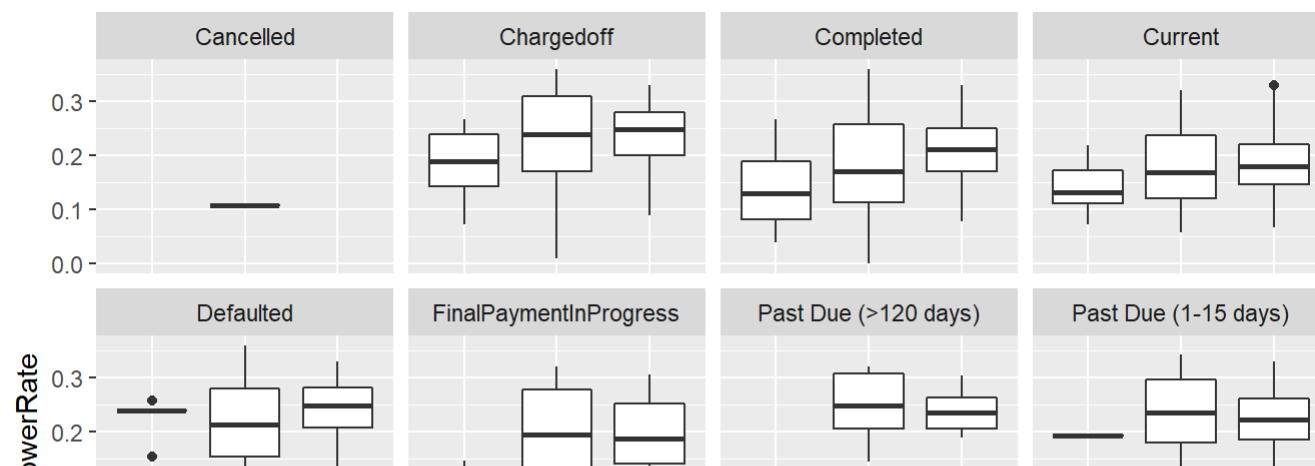
Rates/Category

Our boxplots below show the maximum rates in categories 0-7 (more common loans), which seems to also be the categories where bad/poor credit is more readily accepted. We can easily see in the second visualization that categories 4 and 5 only have 36 month terms. Also, in most categories you can see the longer the term, the higher the rate, although in some cases the 60 month term has similar or even lower rates than a 36 month term (categories 7 and 8 for example.)
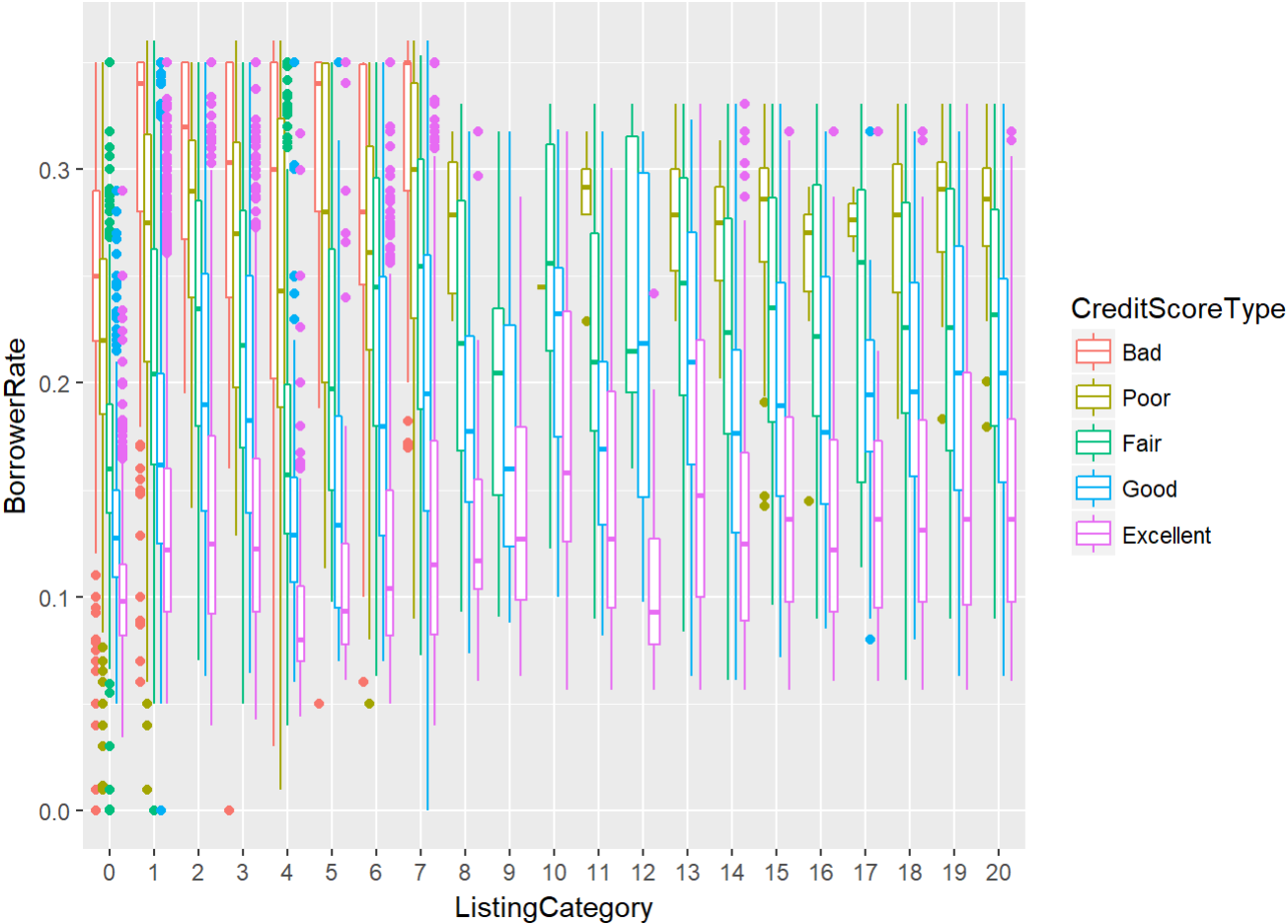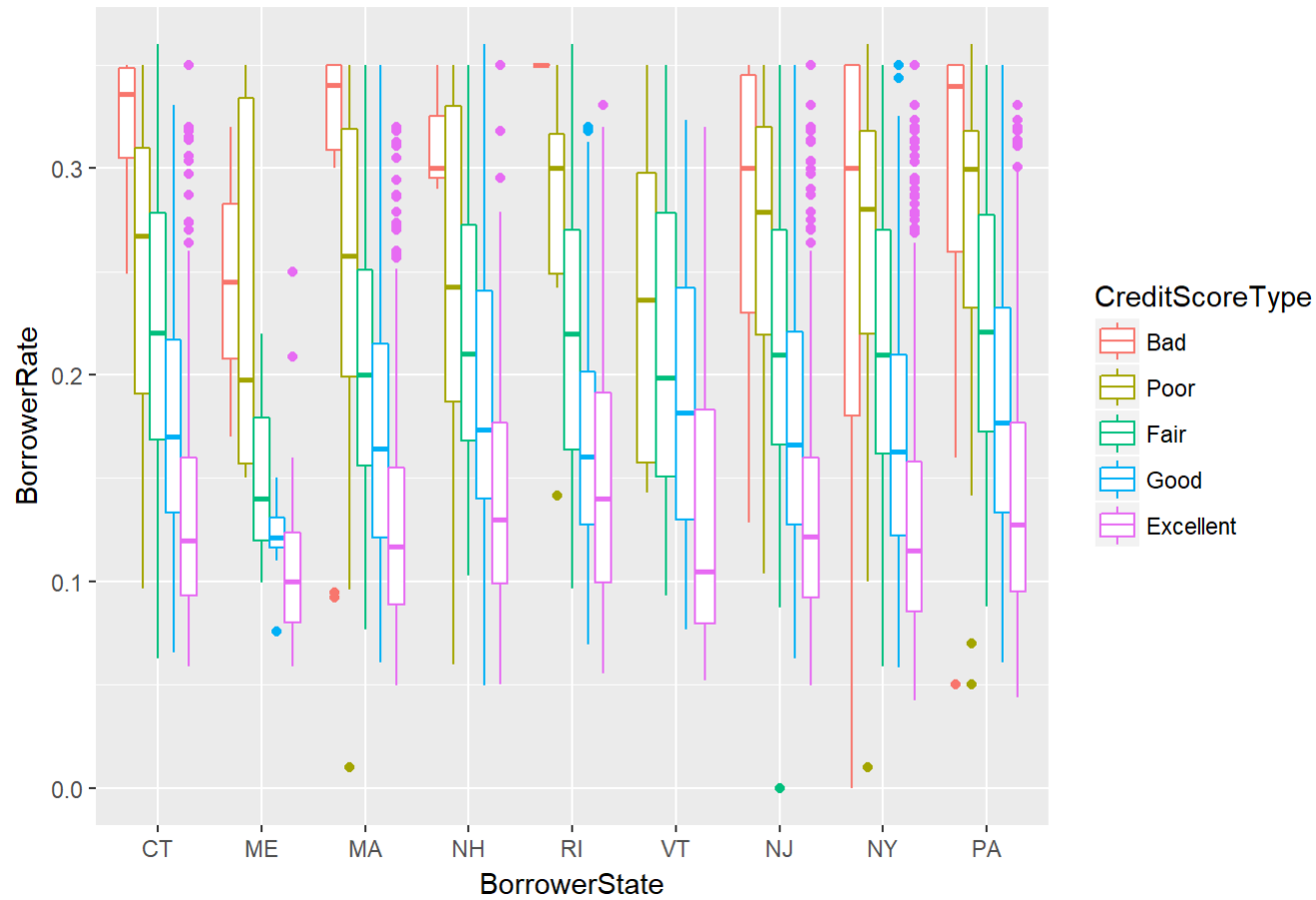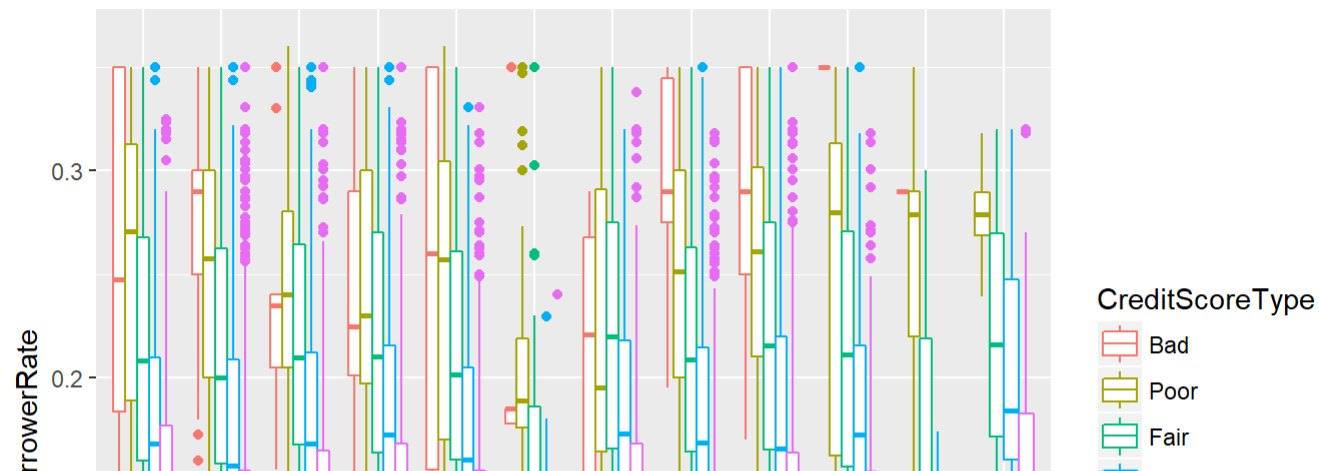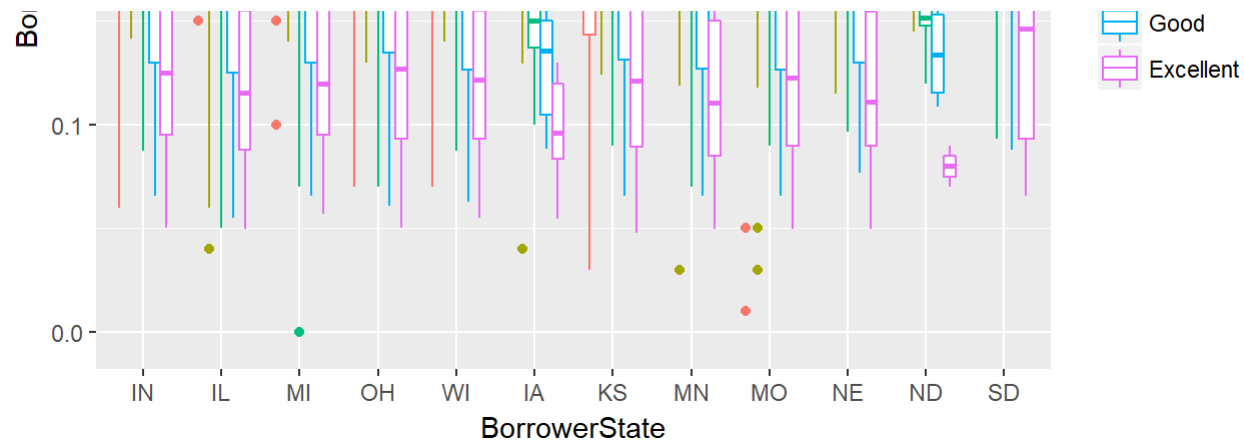
Category Rates per Term
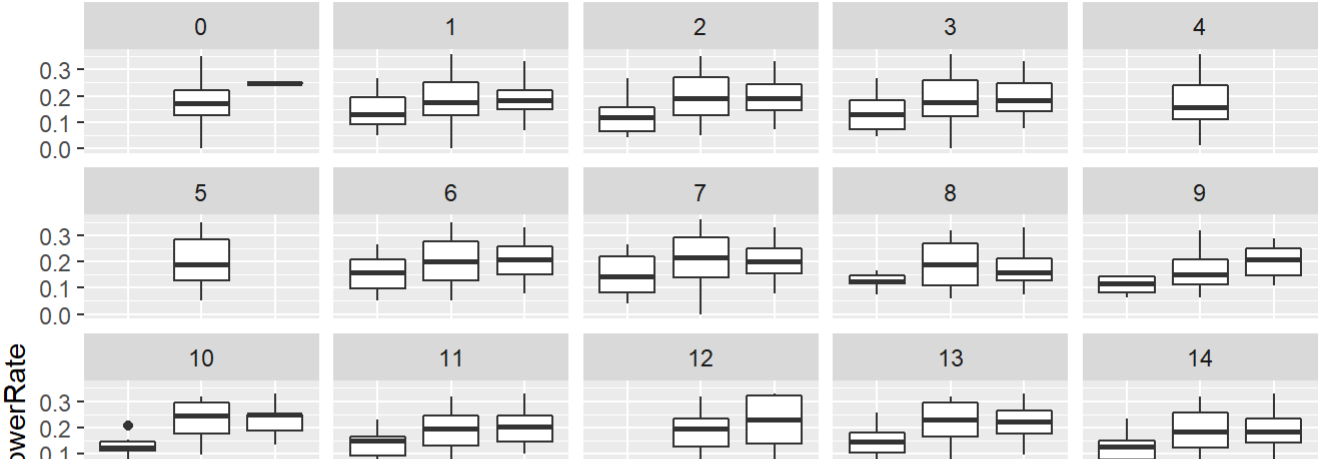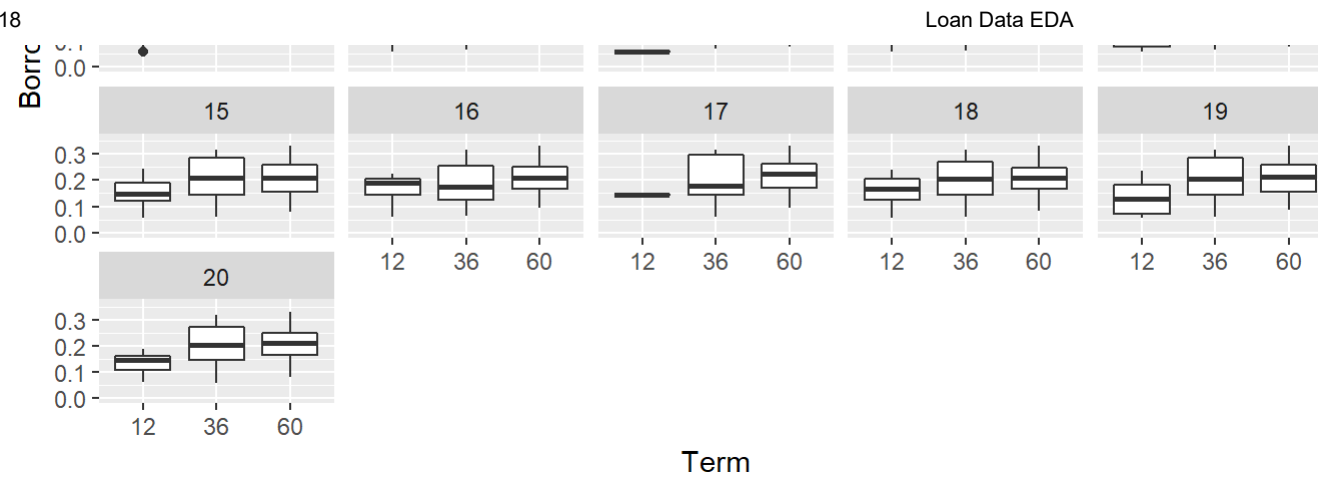
State Interest Rate Trends
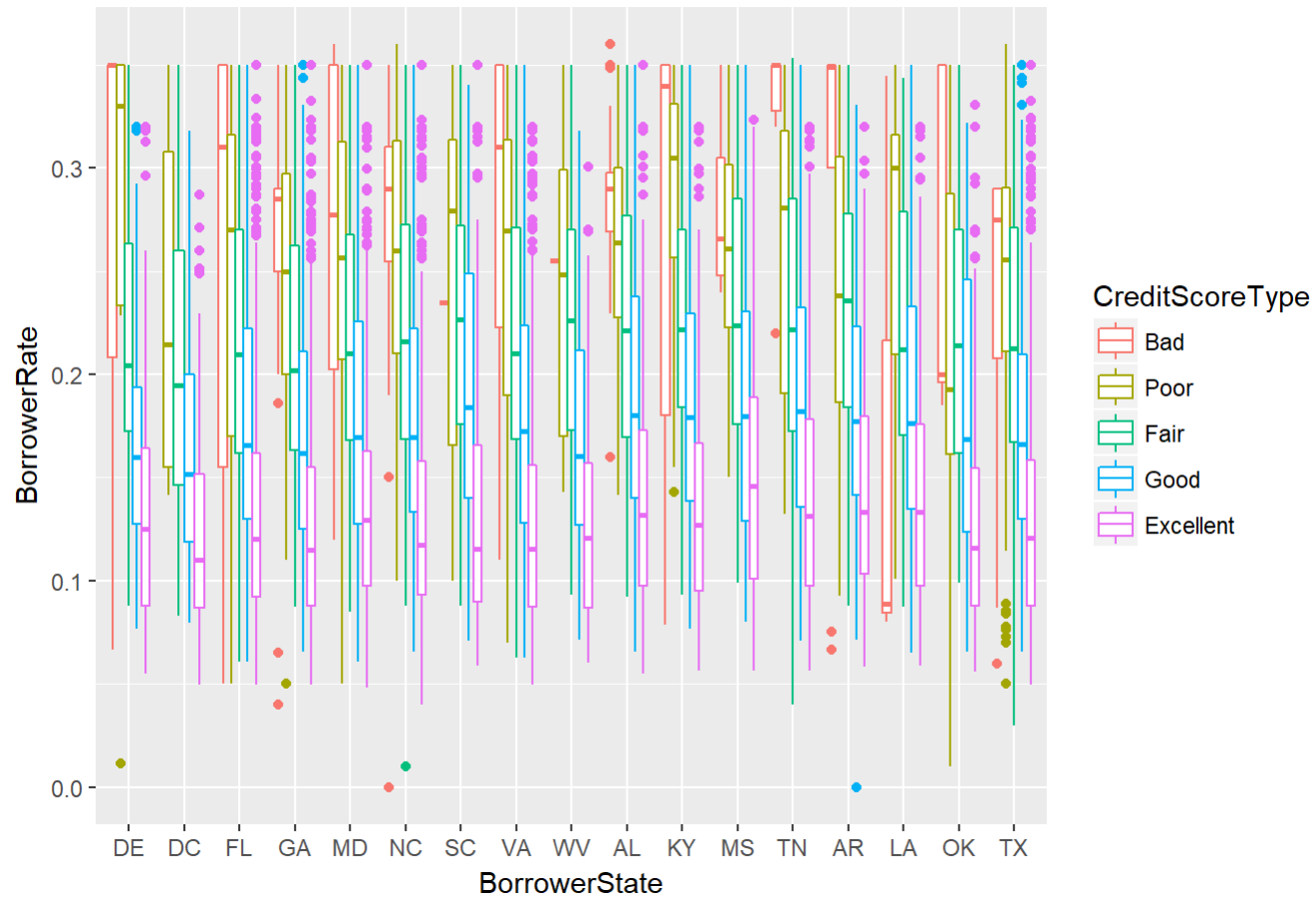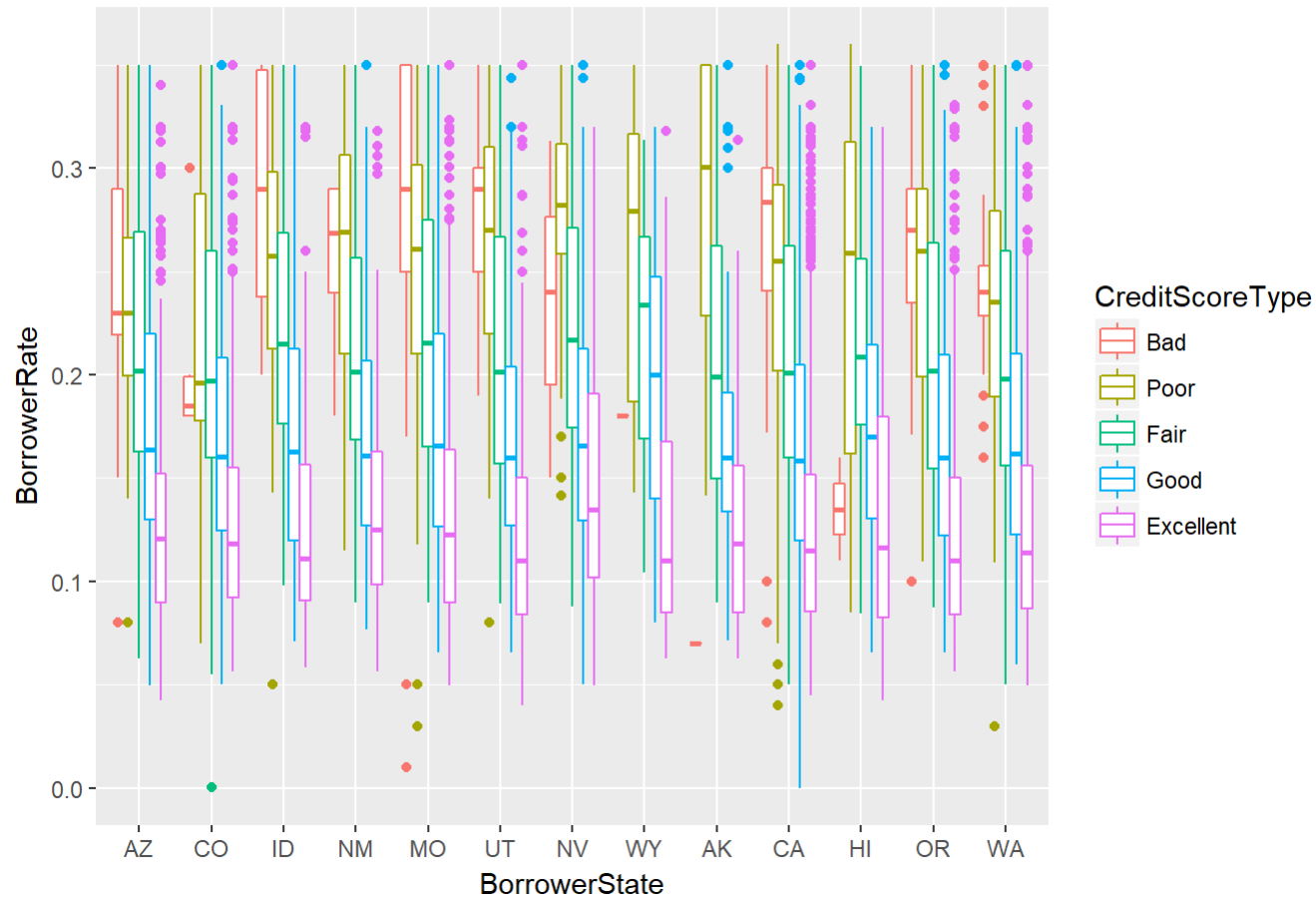
NorthEastern States

MidWestern States

Southern States

Western States

Are there any differences between California, Texas, Florida, and New York? Any differences may be due to usury laws in each state. Looking at the Texas plots makes me think that there could be a program or law protecting those with credit scores lower than 600. According to Debt.org (https://www.debt.org/faqs/americans-in-debt/consumer-texas/), Texas consumers have some of the lowest credit scores in the country.

California

Texas

Florida

New York
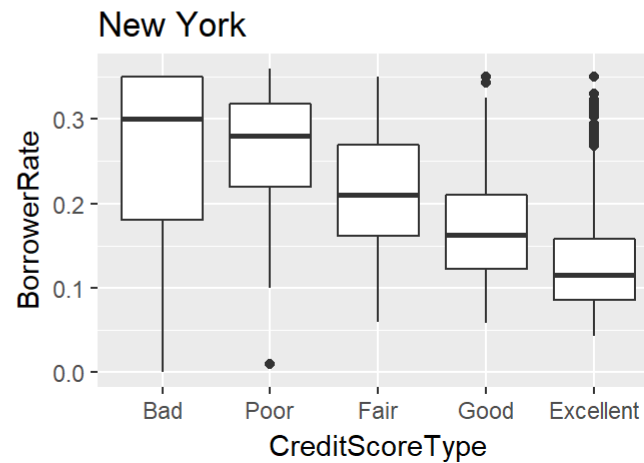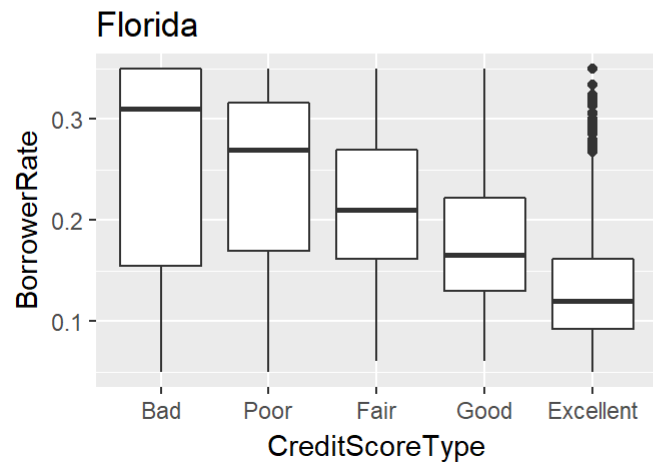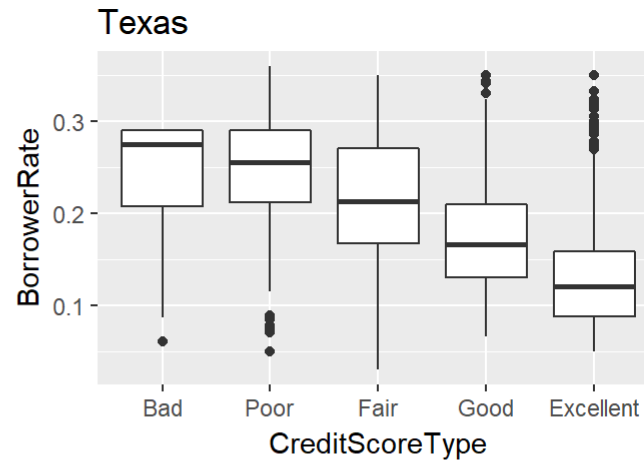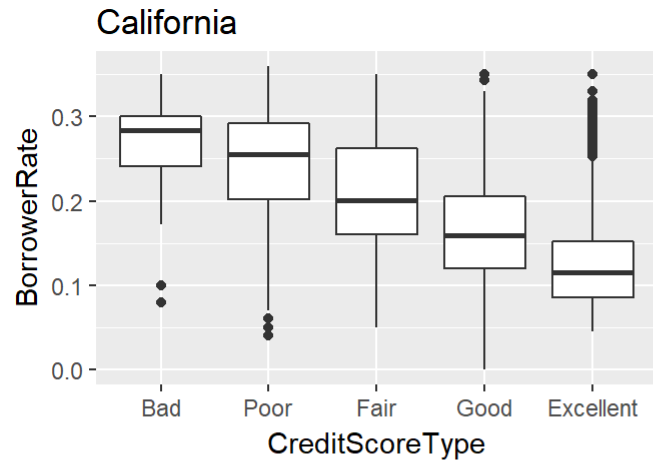
Let's take a closer look at that Texas plot…

It looks like borrowers with a credit score of 600 or lower in TX all have the same term, 36 months. Most of these borrowers selected "Not Available" (0) as their Listing Category. Many of them have defaulted or have been charged off, other than the couple of car loans with lower rates and a few others here and there. This is all very interesting, but it doesn't really tell us for sure if these borrowers were given special treatment.

```
## Term                    LoanStatus   BorrowerRate    ListingCategory
## 12: 0   Chargedoff          :20   Min.   :0.0600   0       :42
## 36:45   Defaulted           :15   1st Qu.:0.2075   1       : 1
## 60: 0   Completed           :10   Median :0.2750   3       : 1
##         Cancelled           : 0   Mean   :0.2461   6       : 1
##         Current             : 0   3rd Qu.:0.2900   2       : 0
##         FinalPaymentInProgress: 0  Max.   :0.2900   4       : 0
##         (Other)             : 0                     (Other): 0
## BorrowerState           Occupation      EmploymentStatus
## TX     :45   Other           :18   Full-time    :37
##        : 0   Professional    : 4   Self-employed: 4
## AK     : 0   Analyst         : 2   Part-time    : 3
## AL     : 0   Clerical        : 2   Retired      : 1
## AR     : 0   Executive       : 2                : 0
## AZ     : 0   Sales - Commission: 2  Employed     : 0
## (Other): 0   (Other)         :15   (Other)      : 0
## CreditScoreRangeLower CreditScoreRangeUpper OpenCreditLines
## Min.   :520       Min.   :539       Min.   : 0.000
## 1st Qu.:520       1st Qu.:539       1st Qu.: 2.000
## Median :520       Median :539       Median : 3.000
## Mean   :520       Mean   :539       Mean   : 4.911
## 3rd Qu.:520       3rd Qu.:539       3rd Qu.: 8.000
## Max.   :520       Max.   :539       Max.   :24.000
##
## CurrentDelinquencies AmountDelinquent DebtToIncomeRatio
## Min.   : 0.000    Min.   :     0   Min.   : 0.0200
## 1st Qu.: 2.000    1st Qu.:  1047   1st Qu.: 0.0800
## Median : 7.000    Median :  3868   Median : 0.1400
## Mean   : 8.622    Mean   : 15939   Mean   : 0.6044
## 3rd Qu.:11.000    3rd Qu.:  7984   3rd Qu.: 0.2100
## Max.   :31.000    Max.   :444745   Max.   :10.0100
##
## StatedMonthlyIncome MonthlyLoanPayment  CreditScoreType
## Min.   :   0.083   Min.   :  0.00   Bad      :45
## 1st Qu.:2367.000   1st Qu.: 71.31   Poor     : 0
## Median :3000.000   Median : 99.69   Fair     : 0
## Mean   :3270.200   Mean   :112.86   Good     : 0
## 3rd Qu.:3987.000   3rd Qu.:148.67   Excellent: 0
## Max.   :6666.667   Max.   :385.53
##
```

Texas Categories/Status
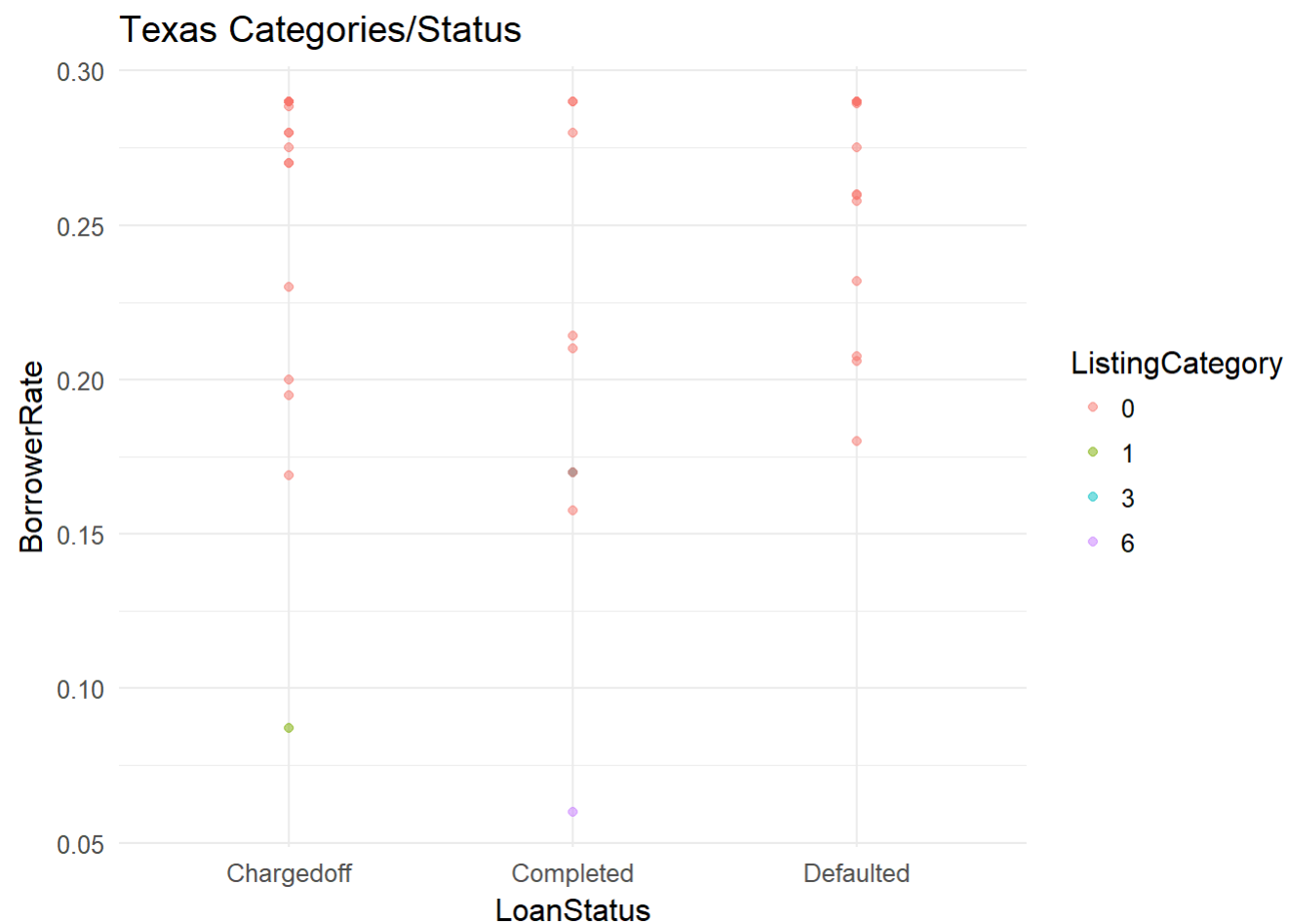
Occupation Rates and DTIR

Occupation is a difficult variable to visualize with it being a categorical variable with multiple characters. However, if we simply plot the mean and median rates using columns on a flipped coordinate, we can zoom in a bit on our data. We can see some interesting highs and lows, like our judges' low interest rates and our teacher's aides' high interest rates.

Loan Data EDA
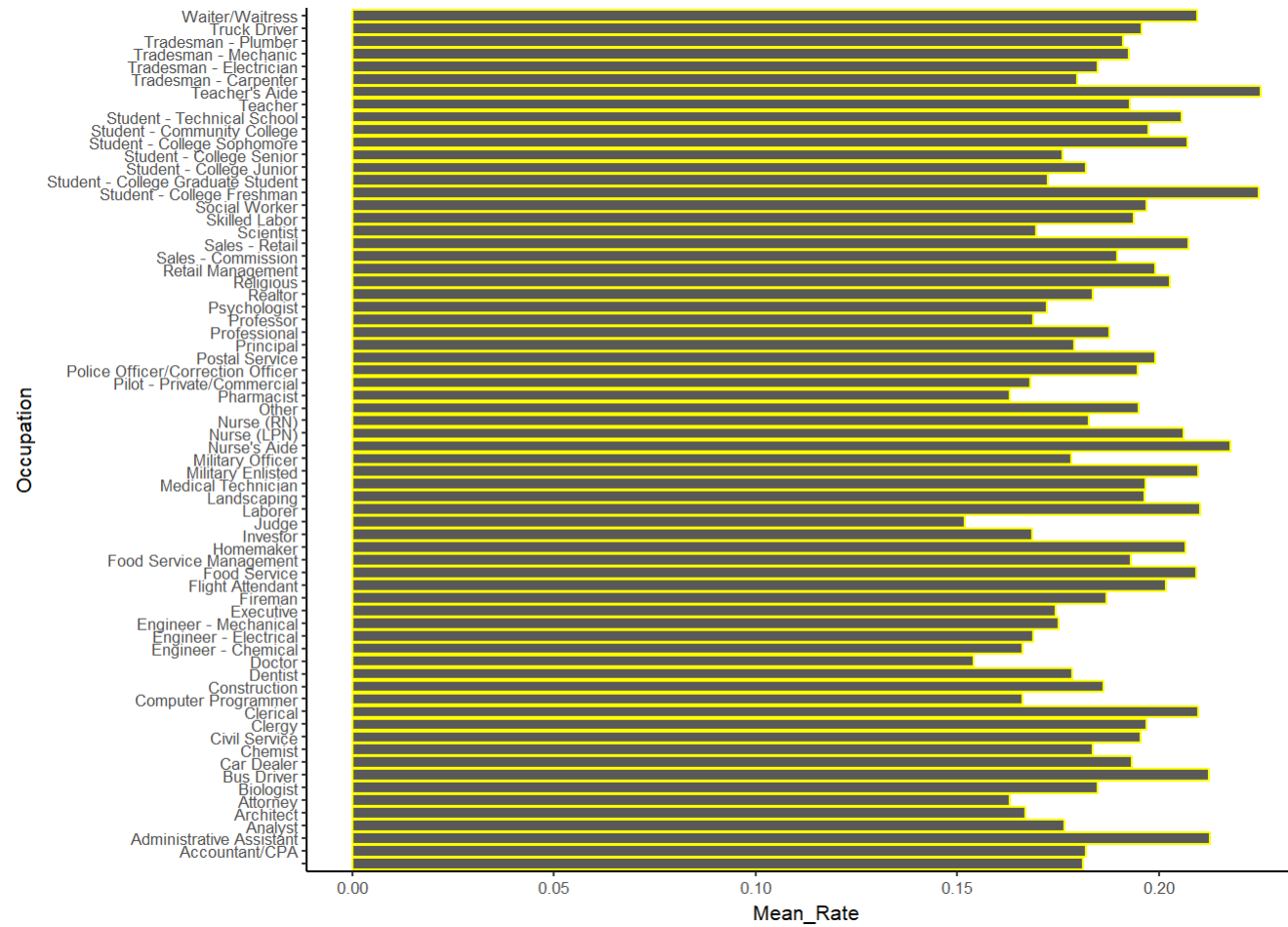
Occupation - Top 10 Mean Interest Rates

```
## # A tibble: 10 x 4
##                     Occupation Mean_DTIR Mean_Rate Mean_CreditScore
##                         <fctr>     <dbl>     <dbl>            <dbl>
##  1           Teacher's Aide 0.4805622 0.2253450         695.4659
##  2 Student - College Freshman 0.2434483 0.2246931         658.3103
##  3             Nurse's Aide 0.3519258 0.2177012         697.7007
##  4   Administrative Assistant 0.3038887 0.2125797         698.0648
##  5               Bus Driver 0.3010438 0.2124030         694.6902
##  6                   Laborer 0.2954941 0.2101951         696.6227
##  7                 Clerical 0.3256760 0.2097594         691.5465
##  8         Military Enlisted 0.2922155 0.2096598         695.1251
##  9           Waiter/Waitress 0.4320330 0.2095415         689.6044
## 10             Food Service 0.3394261 0.2092761         695.3018
```

Lowest/Highest Rates

Top 4/Bottom 4

Top 4 Occupations
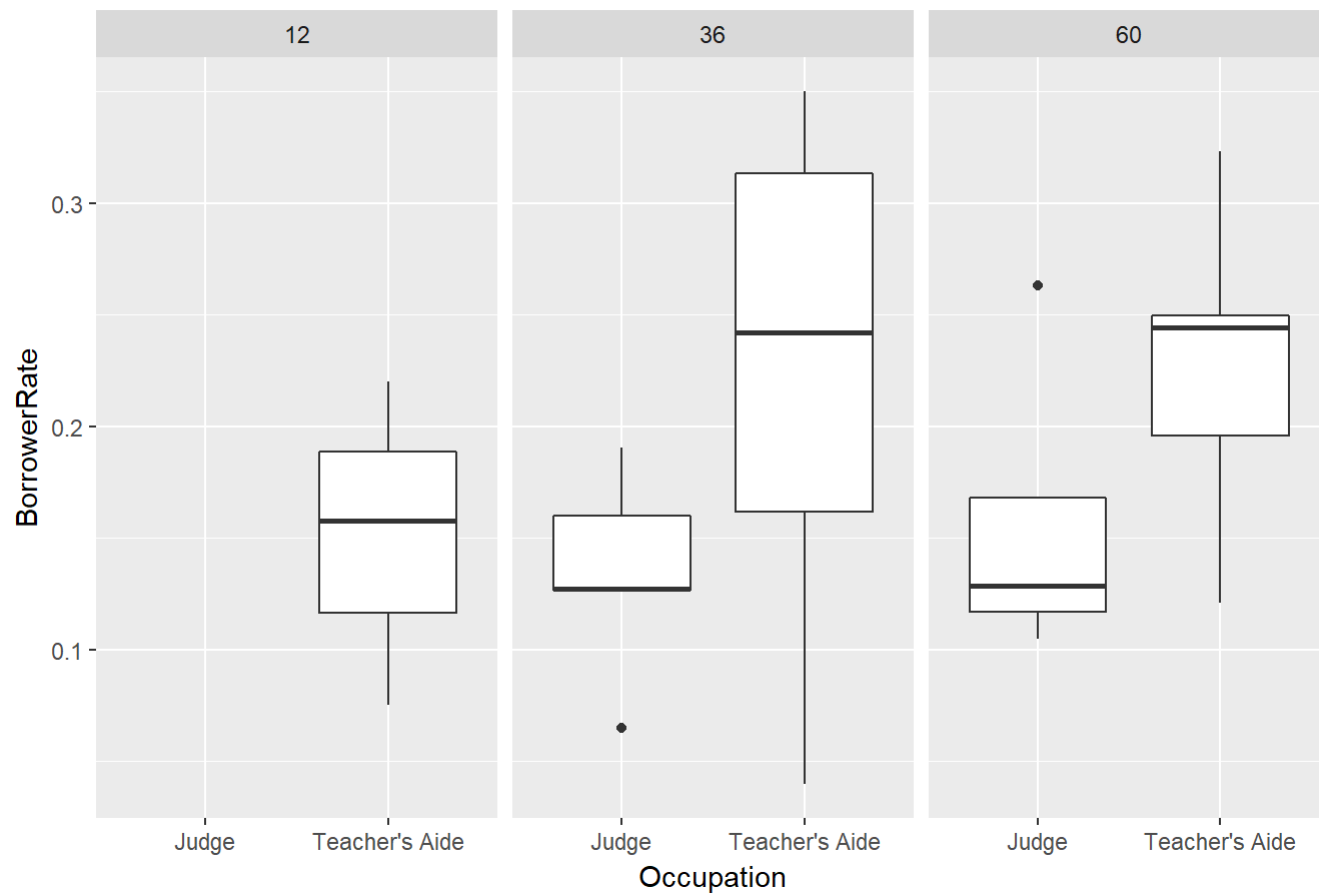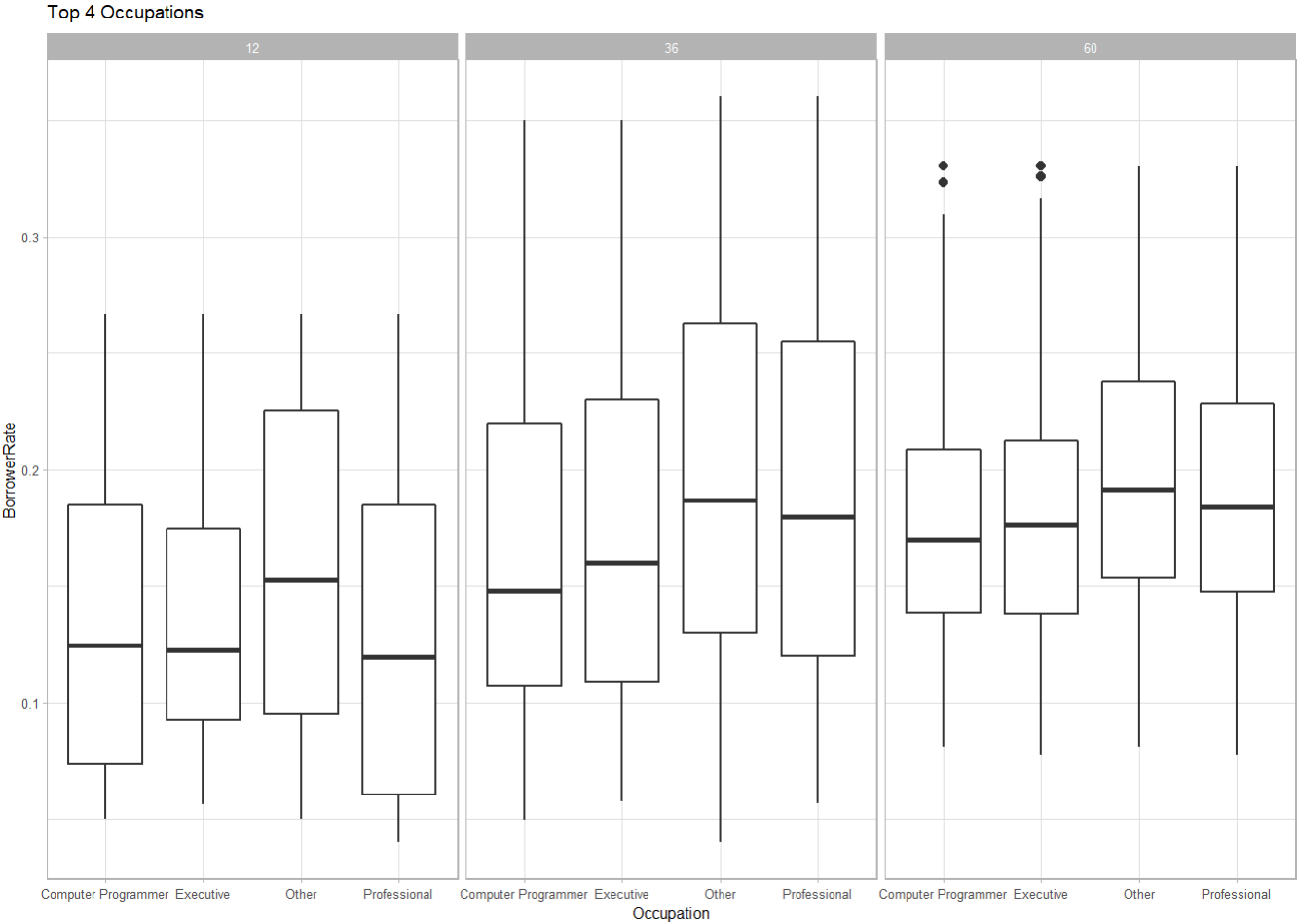

```
## 
##          Other       Professional Computer Programmer
##          23782              12341                 3994
##       Executive
##           3859
```
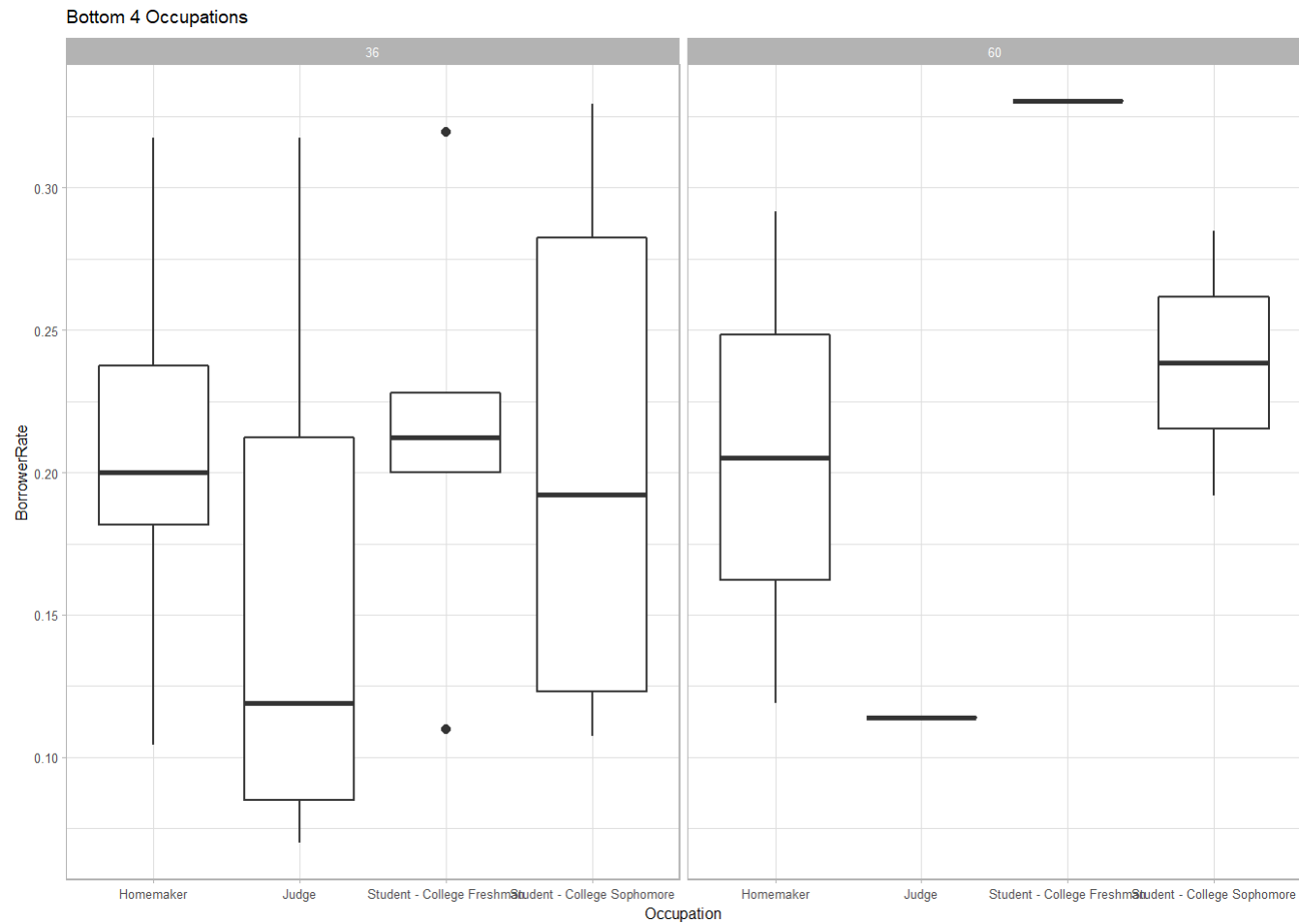
Bottom 4 Occupations


```
## 
##   Student - College Freshman                       Judge 
##                          29                          22 
## Student - Community College   Student - Technical School 
##                          15                           8 
```

Mean and Median Rate/DTIR/Credit Score per Term

```
## # A tibble: 3 x 4
##     Term Mean_DTIR Mean_Rate Mean_CreditScore
##   <fctr>     <dbl>     <dbl>            <dbl>
## 1     12 0.2202473 0.1438962         732.5689
## 2     36 0.2832932 0.1912774         704.6023
## 3     60 0.2564529 0.1916467         723.0617
```

```
## # A tibble: 3 x 4
##     Term Median_DTIR Median_Rate Median_CreditScore
##   <fctr>       <dbl>       <dbl>              <int>
## 1     12        0.17      0.1323                719
## 2     36        0.22      0.1795                699
## 3     60        0.23      0.1845                719
```

Scatterplot Matrix

Using a sample size of 20,000, we can construct a scatterplot matrix showing correlation coefficients for 5 of our quantitative variables. We can also see any significant differences in Term, whether having a 12 month, 36 month, or 60 month term makes a difference in correlation.
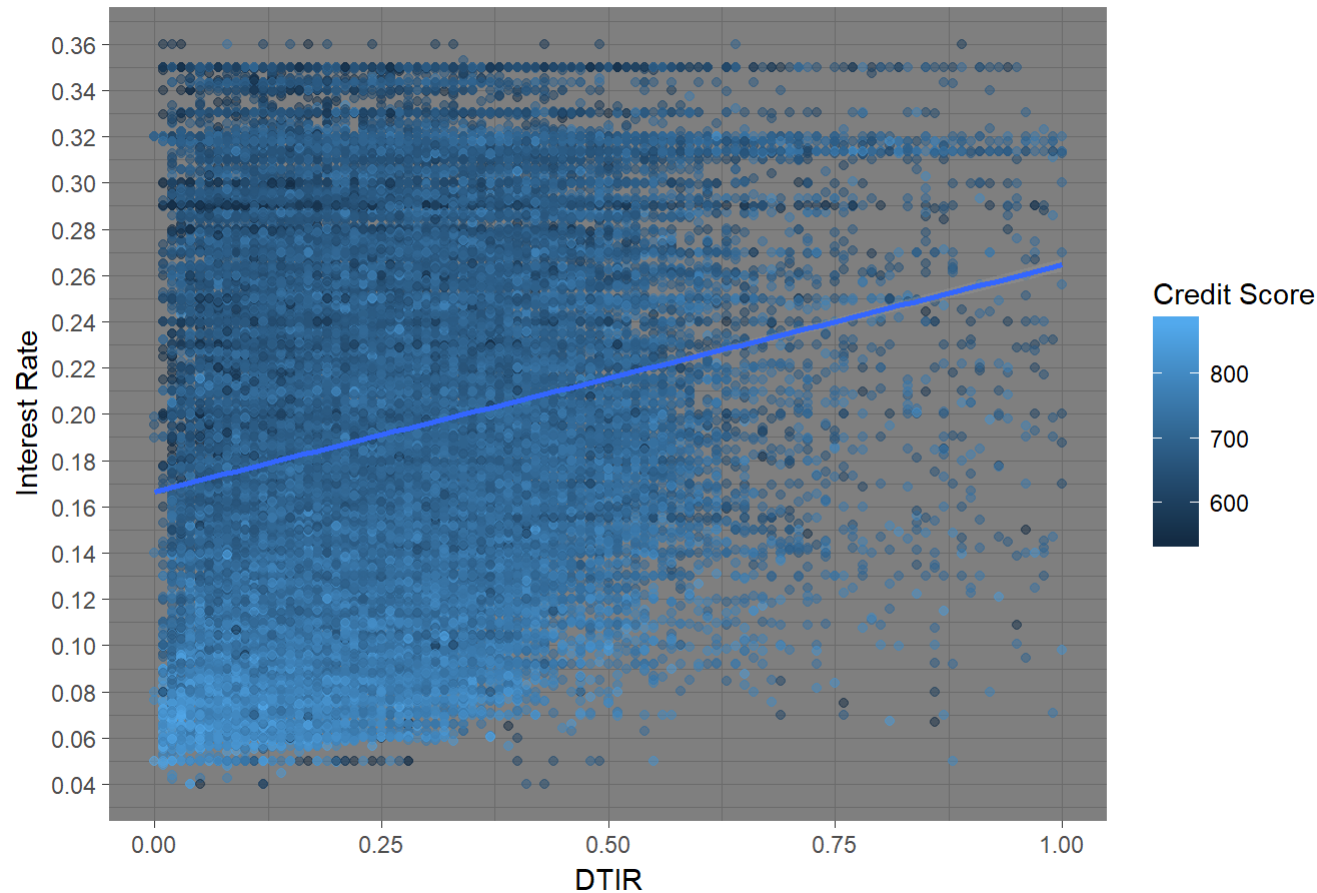
- Borrower Rate
- Credit Score Range Upper
- Debt to Income Ratio
- Stated Monthly Income
- Monthly Loan Payment
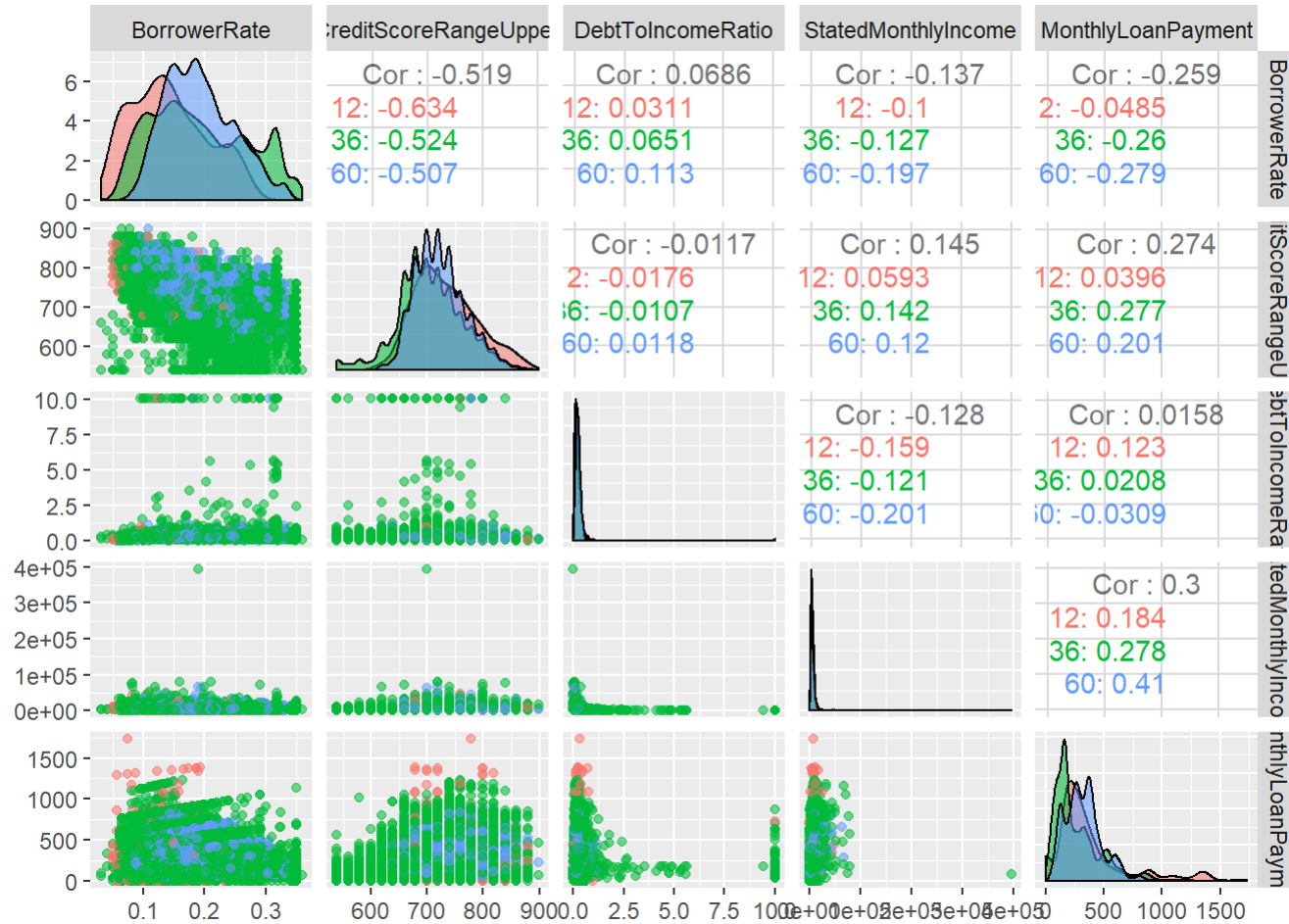
FINAL PLOT #1

Impact of DTIR and Credit Score on Interest Rates

Impact of DTIR and Credit Score on Interest Rates

Our plot above shows lighter colored dots mostly at the bottom and darker colors at the top. This makes it obvious how credit score impacts interest rate in most cases. We can also see a clear (multicolored) line at the 32% interest rate and a rather dark line at 35%. There is only a slight correlation between DTIR and interest rate. If you have a DTIR above 30%, you may end up with a slightly higher interest rate. Most investors probably just want to see that you can be trusted, thus the correlation between rate and credit score. This plot shows the general findings that emerged from our exploration of interest rate, DTIR, and Credit Scores.
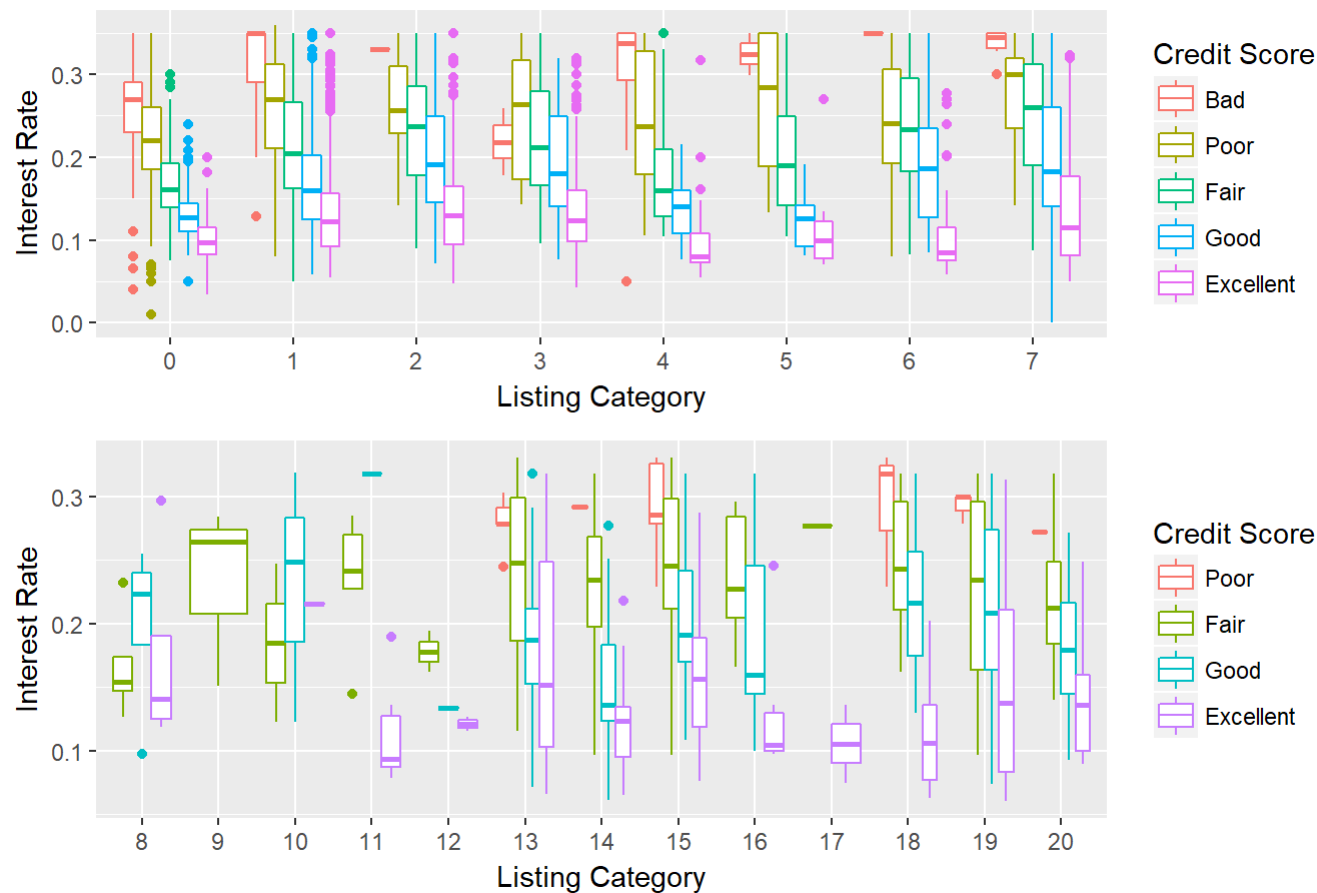
FINAL PLOT #2

Impact of Listing Category on Interest Rate

Number	Category

4/19/2018Loan Data EDA

| Number	Category
0	Not Available
1	Debt Consolidation
2	Home Improvement
3	Business
4	Personal Loan
5	Student Use
6	Auto
7	Other
8	Baby and 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

Impact of Listing Category on Interest Rate

The plot above is from our analysis of Interest Rates in each Listing Category. For the final plot, I split the categories up, with our more popular categories at the top of our grid and less popular ones at the bottom. Among our popular categories, "Debt Consolidation" looks to have some of the highest rates and "Not Available", some of the lowest. Among our less popular group, there doesn't seem to be any borrowers with bad credit and very little with poor credit even. This may be why we see lower rates in our plot at the bottom. I'm sure in order to get a loan for an "RV" or "Vacation", you'd have to have pretty good credit. These types of loans may also have shorter terms, which also means lower interest rates. We did see in our original exploration of Listing Categories that categories 0, 4, and 5 contained no 12 month loan terms at all.

FINAL PLOT #3

MidWest Interest Rates

MidWestern States
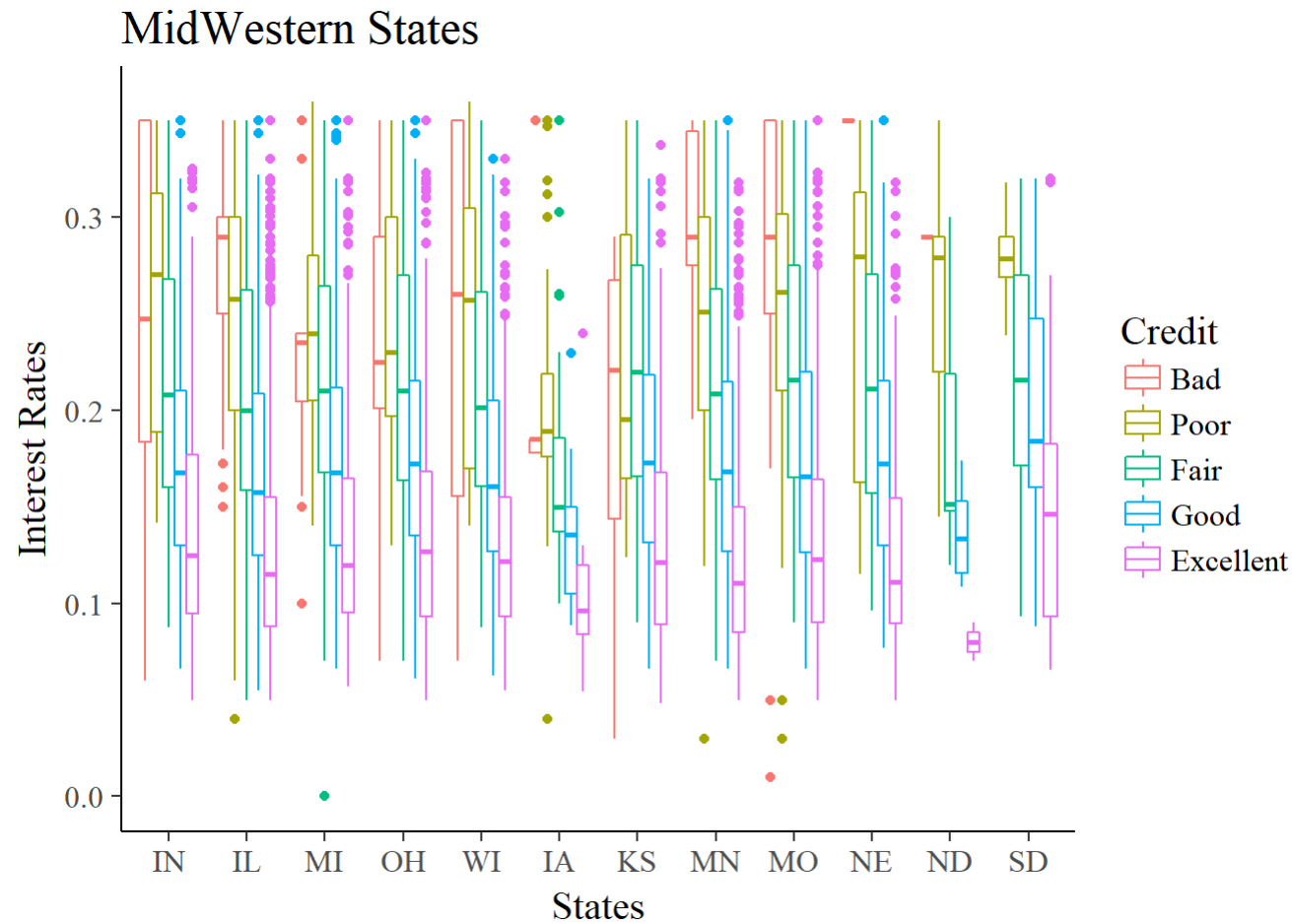
Since I live in the MidWest, specifically Wisconsin, I decided to make my final plot the MidWest plot. We established in our analysis of state interest rates that some rates may be affected by usury laws. When looking at the Midwest plot, SD stands out as a possibility of having restrictions on interest rates. Or, maybe they just tend to have better credit in South Dakota? When I look at Wisconsin, I notice the wide spread for bad credit in particular. Overall, though, we're looking at lower interest rates for Good/Excellent credit. That much remains clear.

```
##  [1] "Term"                "LoanStatus"
##  [3] "BorrowerRate"        "ListingCategory"
##  [5] "BorrowerState"       "Occupation"
##  [7] "EmploymentStatus"    "CreditScoreRangeLower"
##  [9] "CreditScoreRangeUpper" "OpenCreditLines"
## [11] "CurrentDelinquencies" "AmountDelinquent"
## [13] "DebtToIncomeRatio"   "StatedMonthlyIncome"
## [15] "MonthlyLoanPayment"  "CreditScoreType"
```

Closing Thoughts

In our analysis of Prosper Loans, we started out with 81 variables and settled on 15 to explore, adding one more along the way. After cleaning up our data (excluding NAs), we were able to keep 97,903 loans in our set. At times, it was difficult plotting the categorical data using just one plot. With the lengthy characters in a few of our variables, we solved this problem by breaking them up into categories. For our "ListingCategory", we grouped them by most popular and least popular in our final plots section. For our "BorrowerState" variable, we grouped them by region. And, for our "Occupation" variable, we plotted those with highest/lowest count and interest rates.

Our plots showed a moderate negative correlation between interest rate and credit score, but surprisingly showed a weak correlation between interest rate and all other variables, including DTIR. Two other variables that showed a low/moderate correlation are reported income and monthly loan payment. This isn't too surprising, as I'm sure most high income borrowers are hoping to pay down their loans quickly with "extra" income.

We've seen a spike in our data with the popular $175/month loan payment. The loan payments are determined by the original loan balance, interest rate, and term. So, in the future we could explore our "LoanOriginalAmount" variable from our original dataset.