

Prosper Loan Analysis by Gregory Winkler

This report explores approximately 114,000 Propser Loan Data for loans made from November 2005 to March 2014. Prosper is a marketplace lending platform that allows for peer to peer loans.

Univariate Plots Section

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 113937 obs. of 83 variables:
## $ ListingKey : chr "1021339766868145413AB3B" "10273602499503308B223C1" "OE...
## $ ListingNumber : int 193129 1209647 81716 658116 909464 1074836 750899 76819...
## $ ListingCreationDate : POSIXct, format: "2007-08-26 19:09:29" "2014-02-27 08:28:07" ...
## $ CreditGrade : chr "C" NA "HR" ...
## $ Term : Factor w/ 3 levels "12","36","60": 2 2 2 2 2 3 2 2 ...
## $ LoanStatus : Factor w/ 12 levels "Cancelled","Chargedoff",...: 3 4 3 4 4 4 ...
## $ ClosedDate : POSIXct, format: "2009-08-14" NA ...
## $ BorrowerAPR : num 0.165 0.12 0.283 0.125 ...
## $ BorrowerRate : num 0.158 0.092 0.275 0.0974 ...
## $ LenderYield : num 0.138 0.082 0.24 0.0874 ...
## $ EstimatedEffectiveYield : num NA 0.0796 NA 0.0849 ...
## $ EstimatedLoss : num NA 0.0249 NA 0.0249 ...
## $ EstimatedReturn : num NA 0.0547 NA 0.06 ...
## $ ProsperRating (numeric) : int NA 6 NA 6 3 5 2 4 ...
## $ ProsperRating (Alpha) : chr NA "A" NA ...
## $ ProsperScore : num NA 7 NA 9 4 10 2 4 ...
## $ ListingCategory (numeric) : int 0 2 0 16 2 1 1 2 ...
## $ BorrowerState : chr "CO" "CO" "GA" ...
## $ Occupation : Factor w/ 67 levels "Accountant/CPA",...: 36 42 36 51 20 42 4...
## $ EmploymentStatus : Factor w/ 8 levels "Employed","Full-time",...: 8 1 3 1 1 1 1 ...
## $ EmploymentStatusDuration : int 2 44 NA 113 44 82 172 103 ...
## $ IsBorrowerHomeowner : Factor w/ 2 levels "False","True": 2 1 1 2 2 2 1 1 ...
## $ CurrentlyInGroup : Factor w/ 2 levels "False","True": 2 1 2 1 1 1 1 1 ...
## $ GroupKey : chr NA NA "783C3371218786870A73D20" ...
## $ DateCreditPulled : POSIXct, format: "2007-08-26 18:41:46" "2014-02-27 08:28:14" ...
## $ CreditScoreRangeLower : int 640 680 480 800 680 740 680 700 ...
## $ CreditScoreRangeUpper : int 659 699 499 819 699 759 699 719 ...
## $ FirstRecordedCreditLine : POSIXct, format: "2001-10-11" "1996-03-18" ...
## $ CurrentCreditLines : int 5 14 NA 5 19 21 10 6 ...
## $ OpenCreditLines : int 4 14 NA 5 19 17 7 6 ...
## $ TotalCreditLinespast7years : int 12 29 3 29 49 49 20 10 ...
## $ OpenRevolvingAccounts : int 1 13 0 7 6 13 6 5 ...
## $ OpenRevolvingMonthlyPayment : num 24 389 0 115 220 1410 214 101 ...
## $ InquiriesLast6Months : int 3 3 0 0 1 0 0 3 ...
## $ TotalInquiries : num 3 5 1 1 9 2 0 16 ...
## $ CurrentDelinquencies : int 2 0 1 4 0 0 0 0 ...
## $ AmountDelinquent : num 472 0 NA 10056 ...
## $ DelinquenciesLast7Years : int 4 0 0 14 0 0 0 0 ...
## $ PublicRecordsLast10Years : int 0 1 0 0 0 0 0 1 ...
## $ PublicRecordsLast12Months : int 0 0 NA 0 0 0 0 0 ...
## $ RevolvingCreditBalance : num 0 3989 NA 1444 ...
## $ BankcardUtilization : num 0 0.21 NA 0.04 0.81 0.39 0.72 0.13 ...
## $ AvailableBankcardCredit : num 1500 10266 NA 30754 ...
## $ TotalTrades : num 11 29 NA 26 39 47 16 10 ...
```

```

## $ TradesNeverDelinquent (percentage) : num  0.81 1 NA 0.76 0.95 1 0.68 0.8 ...
## $ TradesOpenedLast6Months           : num  0 2 NA 0 2 0 0 0 ...
## $ DebtToIncomeRatio               : num  0.17 0.18 0.06 0.15 0.26 0.36 0.27 0.24 ...
## $ IncomeRange                     : Ord.factor w/ 8 levels "Not displayed"<...: 5 6 1 5 8 8 5 5 ...
## $ IncomeVerifiable                : Factor w/ 2 levels "False","True": 2 2 2 2 2 2 2 ...
## $ StatedMonthlyIncome              : num  3083 6125 2083 2875 ...
## $ LoanKey                         : chr  "E33A3400205839220442E84" "9E3B37071505919926B1D82" "69...
## $ TotalProsperLoans               : int  NA NA NA NA 1 NA NA NA ...
## $ TotalProsperPaymentsBilled      : int  NA NA NA NA 11 NA NA NA ...
## $ OnTimeProsperPayments          : int  NA NA NA NA 11 NA NA NA ...
## $ ProsperPaymentsLessThanOneMonthLate: int  NA NA NA NA 0 NA NA NA ...
## $ ProsperPaymentsOneMonthPlusLate: int  NA NA NA NA 0 NA NA NA ...
## $ ProsperPrincipalBorrowed       : num  NA NA NA NA 11000 NA NA NA ...
## $ ProsperPrincipalOutstanding    : num  NA NA NA NA ...
## $ ScorexChangeAtTimeOfListing    : int  NA NA NA NA NA NA NA ...
## $ LoanCurrentDaysDelinquent     : int  0 0 0 0 0 0 0 ...
## $ LoanFirstDefaultedCycleNumber : int  NA NA NA NA NA NA NA ...
## $ LoanMonthsSinceOrigination    : int  78 0 86 16 6 3 11 10 ...
## $ LoanNumber                      : int  19141 134815 6466 77296 102670 123257 88353 90051 ...
## $ LoanOriginalAmount             : int  9425 10000 3001 10000 15000 15000 3000 10000 ...
## $ LoanOriginationDate            : POSIXct, format: "2007-09-12" "2014-03-03" ...
## $ LoanOriginationQuarter         : chr  "Q3 2007" "Q1 2014" "Q1 2007" ...
## $ MemberKey                       : chr  "1F3E3376408759268057EDA" "1D13370546739025387B2F4" "5F...
## $ MonthlyLoanPayment              : num  330 319 123 321 ...
## $ LP_CustomerPayments             : num  11396 0 4187 5143 ...
## $ LP_CustomerPrincipalPayments   : num  9425 0 3001 4091 ...
## $ LP_InterestandFees              : num  1971 0 1186 1052 ...
## $ LP_ServiceFees                 : num  -133.2 0 -24.2 -108 ...
## $ LP_CollectionFees              : num  0 0 0 0 0 0 0 ...
## $ LP_GrossPrincipalLoss           : num  0 0 0 0 0 0 0 ...
## $ LP_NetPrincipalLoss             : num  0 0 0 0 0 0 0 ...
## $ LP_NonPrincipalRecoverypayments: num  0 0 0 0 0 0 0 ...
## $ PercentFunded                  : num  1 1 1 1 1 1 1 ...
## $ Recommendations                 : int  0 0 0 0 0 0 0 ...
## $ InvestmentFromFriendsCount     : int  0 0 0 0 0 0 0 ...
## $ InvestmentFromFriendsAmount    : num  0 0 0 0 0 0 0 ...
## $ Investors                        : int  258 1 41 158 20 1 1 1 ...
## $ LoanOriginationDate.bucket     : Factor w/ 10 levels "(2005,2006]",...: 2 9 2 7 8 8 8 ...
## $ BankcardUtilization.bucket      : Factor w/ 5 levels "(0,0.25]","(0.25,0.5]",...: NA 1 NA 1 4 2 ...

## ListingKey          ListingNumber          ListingCreationDate
## Length:113937      Min.    : 4        Min.   :2005-11-09 20:44:28
## Class :character   1st Qu.: 400919   1st Qu.:2008-09-19 10:02:14
## Mode  :character   Median : 600554   Median  :2012-06-16 12:37:19
##                           Mean   : 627886   Mean    :2011-07-09 08:07:23
##                           3rd Qu.: 892634   3rd Qu.:2013-09-09 19:40:48
##                           Max.   :1255725   Max.   :2014-03-10 12:20:53
##
## CreditGrade        Term                  LoanStatus
## Length:113937      12: 1614    Current      :56576
## Class :character   36:87778   Completed    :38074
## Mode  :character   60:24545   Chargedoff   :11992
##                           Defaulted    : 5018
##                           Past Due (1-15 days) : 806

```

```

##                                     Past Due (31-60 days): 363
##                                     (Other) : 1108
##   ClosedDate           BorrowerAPR     BorrowerRate
##   Min.    :2005-11-25 00:00:00  Min.    :0.00653  Min.    :0.0000
##   1st Qu.:2009-07-14 00:00:00  1st Qu.:0.15629  1st Qu.:0.1340
##   Median  :2011-04-05 00:00:00  Median  :0.20976  Median  :0.1840
##   Mean    :2011-03-07 20:21:21  Mean    :0.21883  Mean    :0.1928
##   3rd Qu.:2013-01-30 00:00:00  3rd Qu.:0.28381  3rd Qu.:0.2500
##   Max.    :2014-03-10 00:00:00  Max.    :0.51229  Max.    :0.4975
##   NA's    :58848          NA's    :25
##   LenderYield   EstimatedEffectiveYield EstimatedLoss
##   Min.    :-0.0100  Min.    :-0.183      Min.    :0.005
##   1st Qu.: 0.1242  1st Qu.: 0.116      1st Qu.:0.042
##   Median  : 0.1730  Median  : 0.162      Median  :0.072
##   Mean    : 0.1827  Mean    : 0.169      Mean    :0.080
##   3rd Qu.: 0.2400  3rd Qu.: 0.224      3rd Qu.:0.112
##   Max.    : 0.4925  Max.    : 0.320      Max.    :0.366
##   NA's    :29084          NA's    :29084
##   EstimatedReturn ProsperRating (numeric) ProsperRating (Alpha)
##   Min.    :-0.183    Min.    :1.000      Length:113937
##   1st Qu.: 0.074    1st Qu.:3.000      Class  :character
##   Median  : 0.092    Median  :4.000      Mode   :character
##   Mean    : 0.096    Mean    :4.072
##   3rd Qu.: 0.117    3rd Qu.:5.000
##   Max.    : 0.284    Max.    :7.000
##   NA's    :29084          NA's    :29084
##   ProsperScore   ListingCategory (numeric) BorrowerState
##   Min.    : 1.00    Min.    : 0.000      Length:113937
##   1st Qu.: 4.00    1st Qu.: 1.000      Class  :character
##   Median  : 6.00    Median  : 1.000      Mode   :character
##   Mean    : 5.95    Mean    : 2.774
##   3rd Qu.: 8.00    3rd Qu.: 3.000
##   Max.    :11.00    Max.    :20.000
##   NA's    :29084
##   Occupation     EmploymentStatus
##   Other        :28617  Employed    :67322
##   Professional  :13628  Full-time   :26355
##   Computer Programmer: 4478 Self-employed: 6134
##   Executive    : 4311 Not available: 5347
##   Teacher      : 3759 Other       : 3806
##   (Other)      :55556  (Other)     : 2718
##   NA's         : 3588 NA's       : 2255
##   EmploymentStatusDuration IsBorrowerHomeowner CurrentlyInGroup
##   Min.    : 0.00      False:56459      False:101218
##   1st Qu.: 26.00      True :57478      True : 12719
##   Median  : 67.00
##   Mean    : 96.07
##   3rd Qu.:137.00
##   Max.    :755.00
##   NA's    :7625
##   GroupKey      DateCreditPulled CreditScoreRangeLower
##   Length:113937  Min.    :2005-11-09 00:30:04  Min.    : 0.0
##   Class  :character  1st Qu.:2008-09-16 22:25:27  1st Qu.:660.0
##   Mode   :character  Median :2012-06-17 07:52:34  Median :680.0

```

```

##          Mean    :2011-07-09 15:28:40   Mean    :685.6
##          3rd Qu.:2013-09-11 14:30:24   3rd Qu.:720.0
##          Max.   :2014-03-10 12:20:56   Max.   :880.0
##          NA's    :591
## CreditScoreRangeUpper FirstRecordedCreditLine      CurrentCreditLines
##          Min.   : 19.0        Min.   :1947-08-24 00:00:00   Min.   : 0.00
##          1st Qu.:679.0       1st Qu.:1990-06-01 00:00:00   1st Qu.: 7.00
##          Median :699.0       Median :1995-11-01 00:00:00   Median :10.00
##          Mean   :704.6        Mean   :1994-11-17 07:00:07   Mean   :10.32
##          3rd Qu.:739.0       3rd Qu.:2000-03-14 00:00:00   3rd Qu.:13.00
##          Max.   :899.0        Max.   :2012-12-22 00:00:00   Max.   :59.00
##          NA's    :591         NA's    :697                 NA's    :7604
## OpenCreditLines TotalCreditLinespast7years OpenRevolvingAccounts
##          Min.   : 0.00     Min.   : 2.00        Min.   : 0.00
##          1st Qu.: 6.00     1st Qu.: 17.00       1st Qu.: 4.00
##          Median : 9.00     Median : 25.00       Median : 6.00
##          Mean   : 9.26     Mean   : 26.75       Mean   : 6.97
##          3rd Qu.:12.00     3rd Qu.: 35.00       3rd Qu.: 9.00
##          Max.   :54.00     Max.   :136.00       Max.   :51.00
##          NA's    :7604     NA's    :697
## OpenRevolvingMonthlyPayment InquiriesLast6Months TotalInquiries
##          Min.   : 0.0       Min.   : 0.000     Min.   : 0.000
##          1st Qu.:114.0      1st Qu.: 0.000     1st Qu.: 2.000
##          Median :271.0      Median : 1.000     Median : 4.000
##          Mean   :398.3       Mean   : 1.435     Mean   : 5.584
##          3rd Qu.:525.0      3rd Qu.: 2.000     3rd Qu.: 7.000
##          Max.   :14985.0      Max.   :105.000    Max.   :379.000
##          NA's    :697        NA's    :1159
## CurrentDelinquencies AmountDelinquent DelinquenciesLast7Years
##          Min.   : 0.0000    Min.   : 0.0       Min.   : 0.000
##          1st Qu.: 0.0000    1st Qu.: 0.0       1st Qu.: 0.000
##          Median : 0.0000    Median : 0.0       Median : 0.000
##          Mean   : 0.5921    Mean   : 984.5     Mean   : 4.155
##          3rd Qu.: 0.0000    3rd Qu.: 0.0       3rd Qu.: 3.000
##          Max.   :83.0000     Max.   :463881.0    Max.   :99.000
##          NA's    :697        NA's    :7622      NA's    :990
## PublicRecordsLast10Years PublicRecordsLast12Months RevolvingCreditBalance
##          Min.   : 0.0000    Min.   : 0.000     Min.   : 0
##          1st Qu.: 0.0000    1st Qu.: 0.000     1st Qu.: 3121
##          Median : 0.0000    Median : 0.000     Median : 8549
##          Mean   : 0.3126    Mean   : 0.015     Mean   : 17599
##          3rd Qu.: 0.0000    3rd Qu.: 0.000     3rd Qu.: 19521
##          Max.   :38.0000     Max.   :20.000     Max.   :1435667
##          NA's    :697        NA's    :7604      NA's    :7604
## BankcardUtilization AvailableBankcardCredit TotalTrades
##          Min.   :0.000     Min.   : 0         Min.   : 0.00
##          1st Qu.:0.310     1st Qu.: 880       1st Qu.: 15.00
##          Median :0.600     Median : 4100      Median : 22.00
##          Mean   :0.561     Mean   : 11210     Mean   : 23.23
##          3rd Qu.:0.840     3rd Qu.: 13180     3rd Qu.: 30.00
##          Max.   :5.950     Max.   :646285     Max.   :126.00
##          NA's    :7604      NA's    :7544      NA's    :7544
## TradesNeverDelinquent (percentage) TradesOpenedLast6Months
##          Min.   :0.000     Min.   : 0.000

```

```

## 1st Qu.:0.820          1st Qu.: 0.000
## Median :0.940          Median : 0.000
## Mean   :0.886          Mean   : 0.802
## 3rd Qu.:1.000          3rd Qu.: 1.000
## Max.   :1.000          Max.   :20.000
## NA's   :7544           NA's   :7544
## DebtToIncomeRatio      IncomeRange    IncomeVerifiable
## Min.   : 0.000          $25,000-49,999:32192  False: 8669
## 1st Qu.: 0.140          $50,000-74,999:31050  True :105268
## Median : 0.220          $100,000+:17337
## Mean   : 0.276          $75,000-99,999:16916
## 3rd Qu.: 0.320          Not displayed : 7741
## Max.   :10.010          $1-24,999     : 7274
## NA's   :8554            (Other)       : 1427
## StatedMonthlyIncome     LoanKey       TotalProsperLoans
## Min.   : 0              Length:113937    Min.   :0.00
## 1st Qu.: 3200           Class :character  1st Qu.:1.00
## Median : 4667           Mode  :character  Median :1.00
## Mean   : 5608           NA's   :91852      Mean   :1.42
## 3rd Qu.: 6825           NA's   :91852      3rd Qu.:2.00
## Max.   :1750003         NA's   :91852      Max.   :8.00
## TotalProsperPaymentsBilled OnTimeProsperPayments
## Min.   : 0.00           Min.   : 0.00
## 1st Qu.: 9.00           1st Qu.: 9.00
## Median : 16.00          Median : 15.00
## Mean   : 22.93          Mean   : 22.27
## 3rd Qu.: 33.00          3rd Qu.: 32.00
## Max.   :141.00          Max.   :141.00
## NA's   :91852           NA's   :91852
## ProsperPaymentsLessThanOneMonthLate ProsperPaymentsOneMonthPlusLate
## Min.   : 0.00           Min.   : 0.00
## 1st Qu.: 0.00           1st Qu.: 0.00
## Median : 0.00           Median : 0.00
## Mean   : 0.61           Mean   : 0.05
## 3rd Qu.: 0.00           3rd Qu.: 0.00
## Max.   :42.00           Max.   :21.00
## NA's   :91852           NA's   :91852
## ProsperPrincipalBorrowed ProsperPrincipalOutstanding
## Min.   : 0              Min.   : 0
## 1st Qu.: 3500           1st Qu.: 0
## Median : 6000           Median : 1627
## Mean   : 8472           Mean   : 2930
## 3rd Qu.:11000            3rd Qu.: 4127
## Max.   :72499            Max.   :23451
## NA's   :91852           NA's   :91852
## ScorexChangeAtTimeOfListing LoanCurrentDaysDelinquent
## Min.   :-209.00          Min.   : 0.0
## 1st Qu.: -35.00          1st Qu.: 0.0
## Median : -3.00           Median : 0.0
## Mean   : -3.22           Mean   : 152.8
## 3rd Qu.: 25.00           3rd Qu.: 0.0
## Max.   : 286.00           Max.   :2704.0
## NA's   :95009

```

```

##  LoanFirstDefaultedCycleNumber  LoanMonthsSinceOrigination  LoanNumber
##  Min.   : 0.00                  Min.   : 0.0               Min.   : 1
##  1st Qu.: 9.00                  1st Qu.: 6.0               1st Qu.: 37332
##  Median :14.00                  Median : 21.0              Median : 68599
##  Mean   :16.27                  Mean   : 31.9              Mean   : 69444
##  3rd Qu.:22.00                  3rd Qu.: 65.0              3rd Qu.:101901
##  Max.   :44.00                  Max.   :100.0              Max.   :136486
##  NA's    :96985

##  LoanOriginalAmount  LoanOriginationDate      LoanOriginationQuarter
##  Min.   :1000          Min.   :2005-11-15 00:00:00  Length:113937
##  1st Qu.:4000          1st Qu.:2008-10-02 00:00:00  Class  :character
##  Median :6500          Median :2012-06-26 00:00:00  Mode   :character
##  Mean   :8337          Mean   :2011-07-21 03:18:19
##  3rd Qu.:12000         3rd Qu.:2013-09-18 00:00:00
##  Max.   :35000         Max.   :2014-03-12 00:00:00

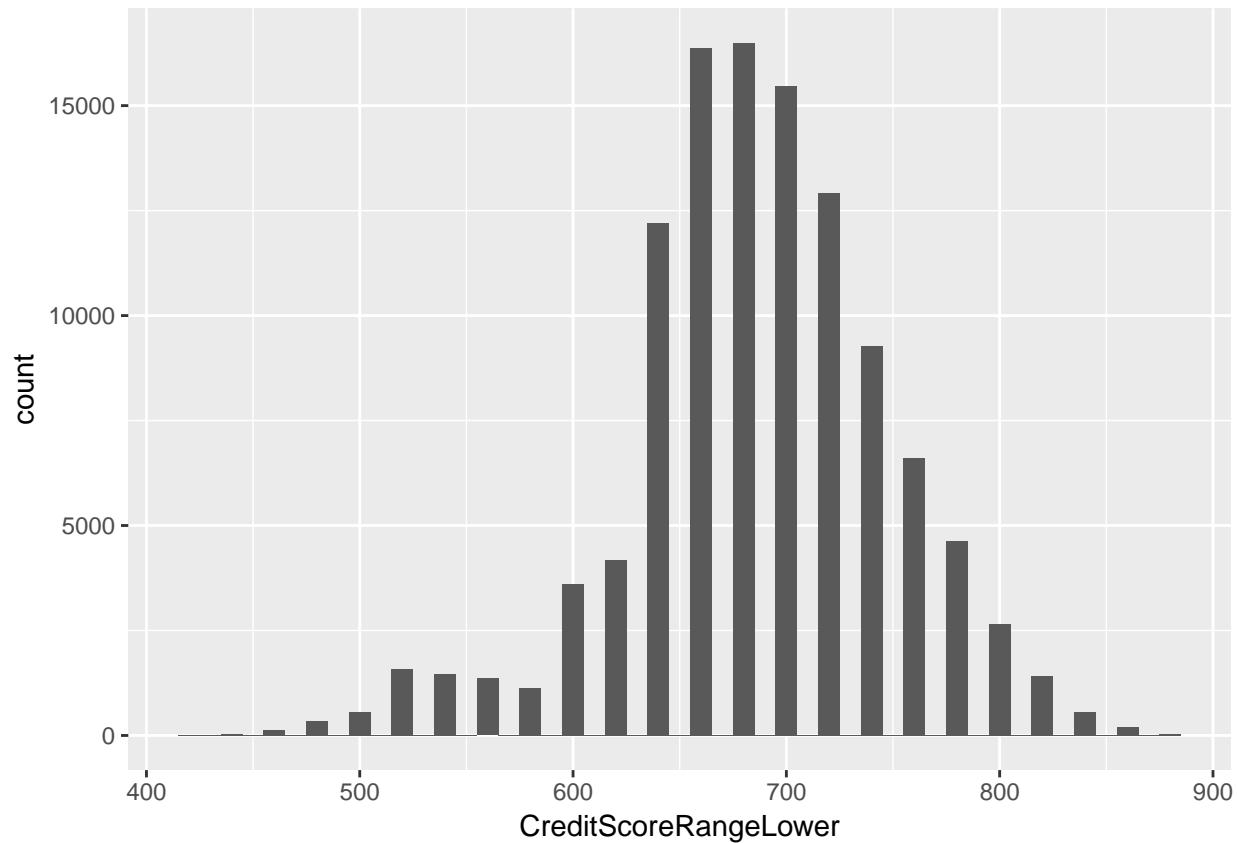
##
##  MemberKey           MonthlyLoanPayment  LP_CustomerPayments
##  Length:113937       Min.   : 0.0      Min.   : -2.35
##  Class  :character   1st Qu.: 131.6    1st Qu.: 1005.76
##  Mode   :character   Median  : 217.7    Median  : 2583.83
##                      Mean   : 272.5    Mean   : 4183.08
##                      3rd Qu.: 371.6    3rd Qu.: 5548.40
##                      Max.   :2251.5    Max.   :40702.39
##  LP_CustomerPrincipalPayments  LP_InterestandFees  LP_ServiceFees
##  Min.   : 0.0          Min.   : -2.35  Min.   :-664.87
##  1st Qu.: 500.9        1st Qu.: 274.87  1st Qu.: -73.18
##  Median : 1587.5       Median : 700.84  Median : -34.44
##  Mean   : 3105.5       Mean   : 1077.54  Mean   : -54.73
##  3rd Qu.: 4000.0       3rd Qu.: 1458.54  3rd Qu.: -13.92
##  Max.   :35000.0       Max.   :15617.03  Max.   : 32.06
##
##  LP_CollectionFees  LP_GrossPrincipalLoss  LP_NetPrincipalLoss
##  Min.   :-9274.75     Min.   : -94.2    Min.   : -954.5
##  1st Qu.:  0.00       1st Qu.:  0.0    1st Qu.:  0.0
##  Median :  0.00       Median :  0.0    Median :  0.0
##  Mean   : -14.24     Mean   : 700.4    Mean   : 681.4
##  3rd Qu.:  0.00       3rd Qu.:  0.0    3rd Qu.:  0.0
##  Max.   :  0.00       Max.   :25000.0   Max.   :25000.0
##
##  LP_NonPrincipalRecoverypayments  PercentFunded  Recommendations
##  Min.   : 0.00          Min.   :0.7000  Min.   : 0.00000
##  1st Qu.: 0.00          1st Qu.:1.0000  1st Qu.: 0.00000
##  Median : 0.00          Median :1.0000  Median : 0.00000
##  Mean   : 25.14          Mean   :0.9986  Mean   : 0.04803
##  3rd Qu.: 0.00          3rd Qu.:1.0000  3rd Qu.: 0.00000
##  Max.   :21117.90        Max.   :1.0125  Max.   :39.00000
##
##  InvestmentFromFriendsCount  InvestmentFromFriendsAmount  Investors
##  Min.   : 0.00000      Min.   : 0.00      Min.   : 1.00
##  1st Qu.: 0.00000      1st Qu.: 0.00      1st Qu.: 2.00
##  Median : 0.00000      Median : 0.00      Median : 44.00
##  Mean   : 0.02346      Mean   : 16.55      Mean   : 80.48
##  3rd Qu.: 0.00000      3rd Qu.: 0.00      3rd Qu.:115.00

```

```

##   Max.    :33.00000      Max.    :25000.00      Max.    :1189.00
## 
##   LoanOriginationDate.bucket BankcardUtilization.bucket
##   (2012,2013]:34345          (0,0.25]   :15842
##   (2011,2012]:19553          (0.25,0.5] :20866
##   (2013,2014]:12172          (0.5,0.75] :26578
##   (2007,2008]:11552          (0.75,1]   :34521
##   (2006,2007]:11460          (1,5]       : 1742
##   (Other)      :24833          NA's       :14388
##   NA's        :     22

```

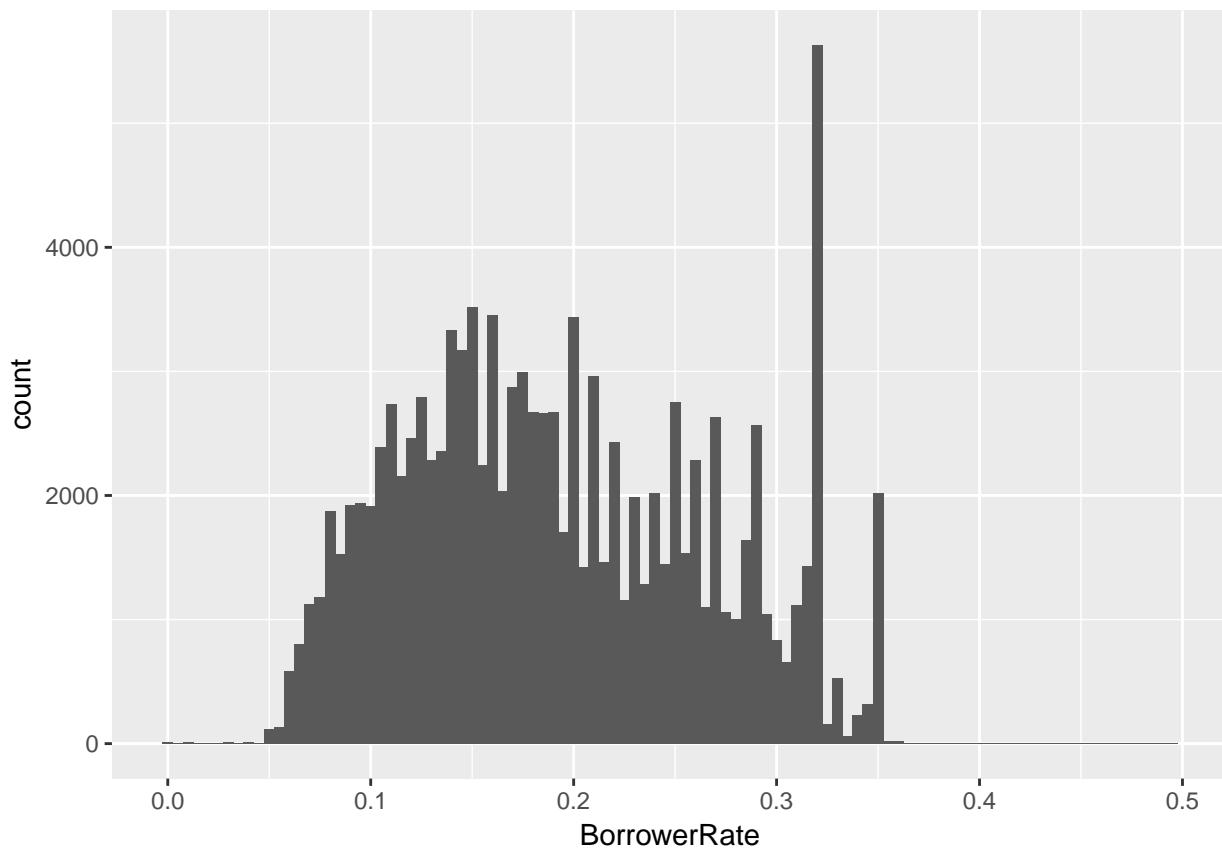


```

##   Min. 1st Qu. Median Mean 3rd Qu. Max. NA's
##   0.0  660.0 680.0 685.6 720.0 880.0 591

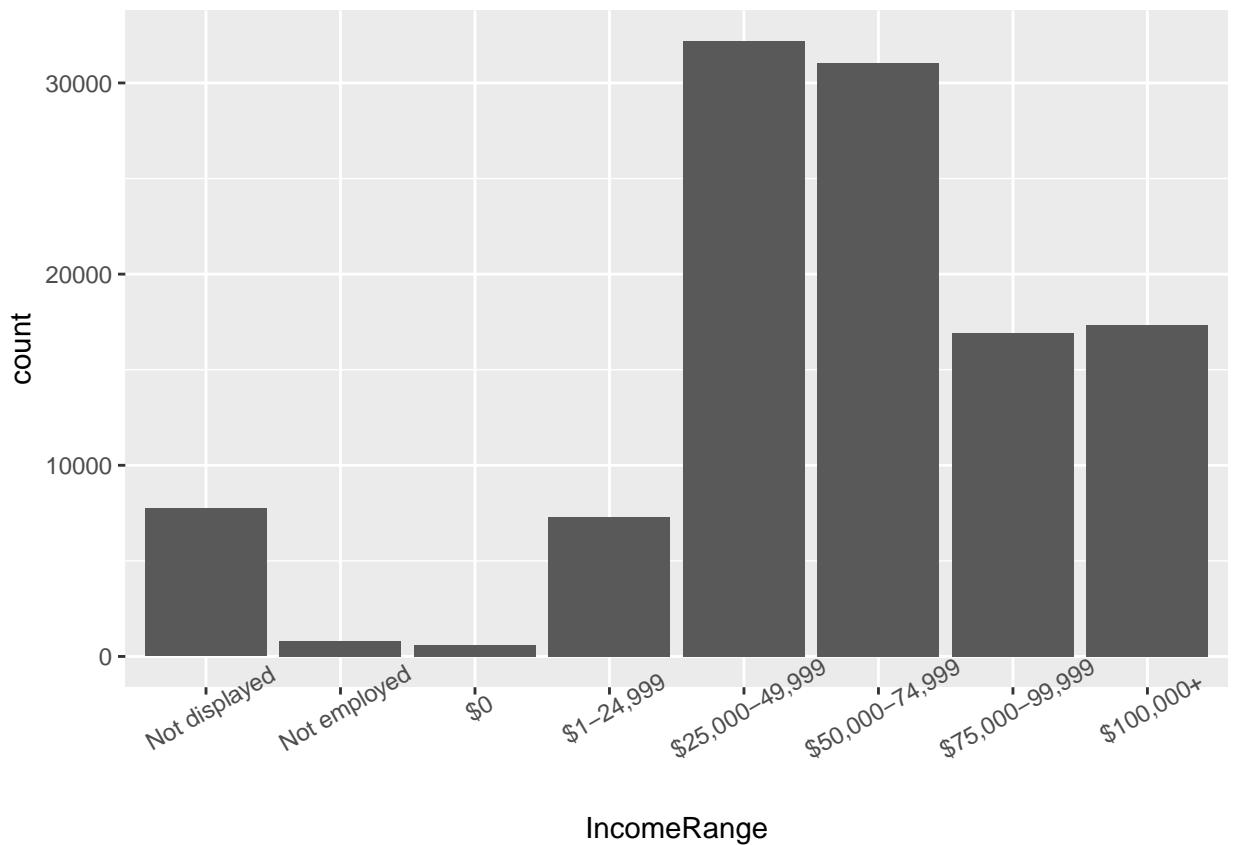
```

The credit scores have a median of 680.

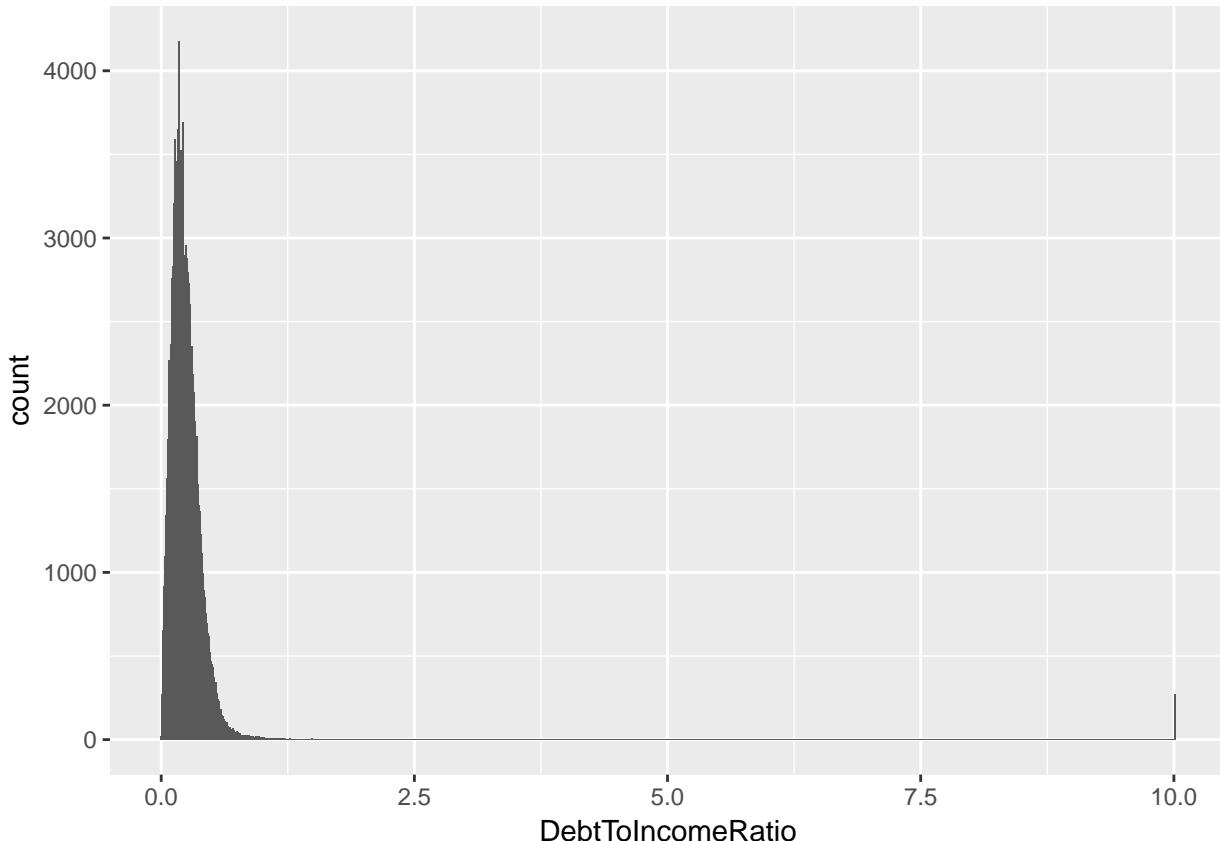


```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.  
## 0.0000 0.1340 0.1840 0.1928 0.2500 0.4975
```

The borrower's interest rates have a normal distribution with the exception of a significant number of loans around 0.33 (33%).



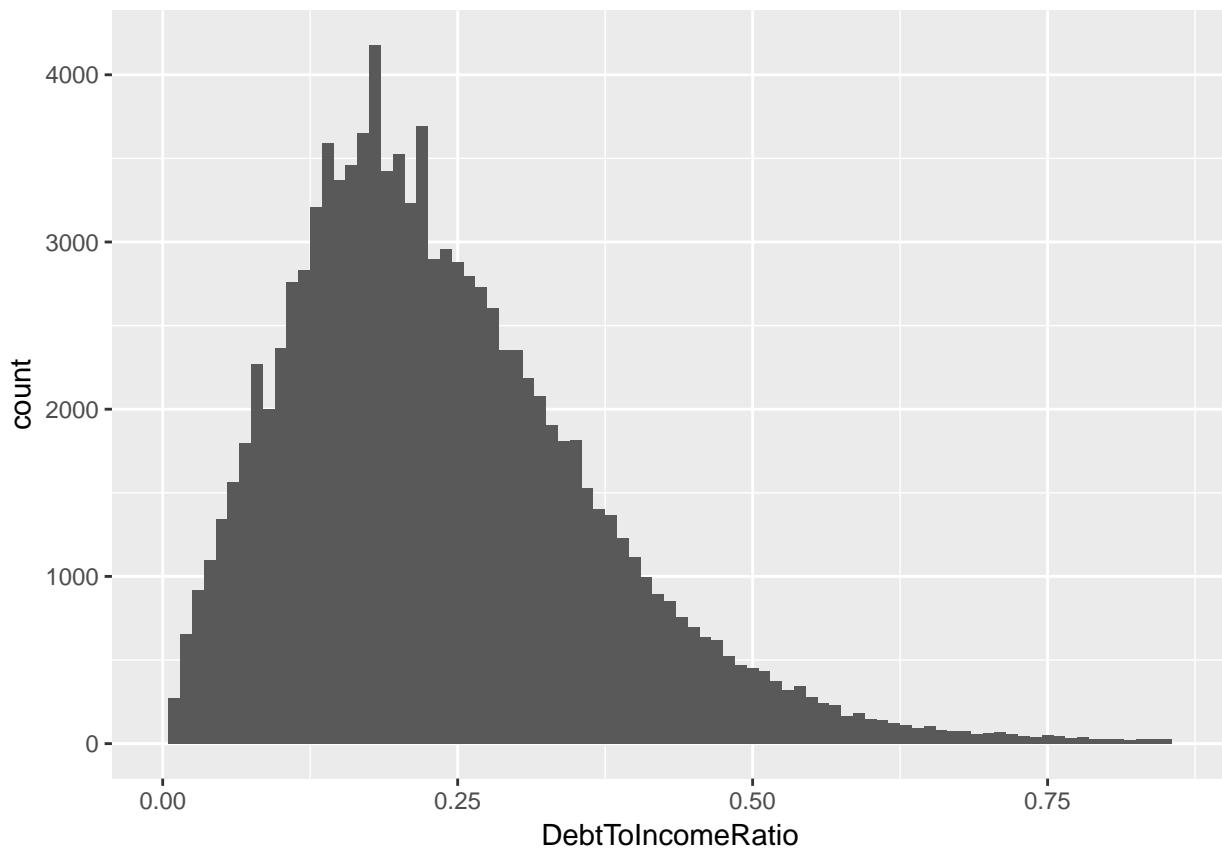
The largest group of borrowers have incomes in the \$25,000 to \$74,999 Income Ranges.



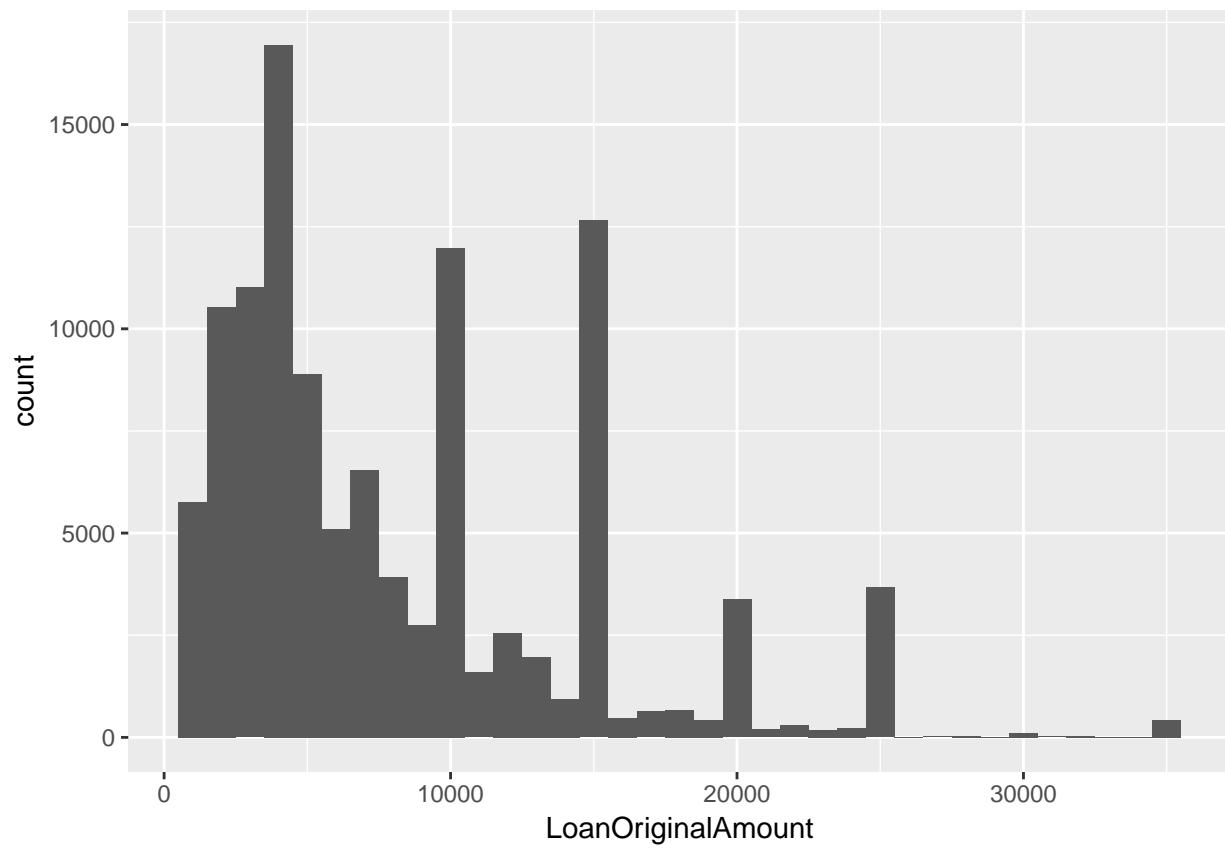
```
##      Min. 1st Qu. Median     Mean 3rd Qu.   Max.    NA's
##  0.000  0.140  0.220  0.276  0.320 10.010  8554
```

The Debt-to-Income Ratio for most borrowers is reasonable with a mean around 0.27. However, there is a sizable number of borrowers over 10. Such a high number would be difficult to ever pay back. Taking a quick look at the IncomeRanges and whether or not that income is verifiable shows that these borrowers are predominantly low income and possibly may not be disclosing their true income.

```
## Source: local data frame [8 x 3]
## Groups: IncomeRange [?]
##
##      IncomeRange IncomeVerifiable     n
##                <ord>            <fctr> <int>
## 1 Not displayed           False     49
## 2 Not displayed           True      10
## 3 Not employed            False     21
## 4 Not employed            True      3
## 5 $1-24,999              False    115
## 6 $1-24,999              True     72
## 7 $50,000-74,999          True      1
## 8 $100,000+               True      1
```

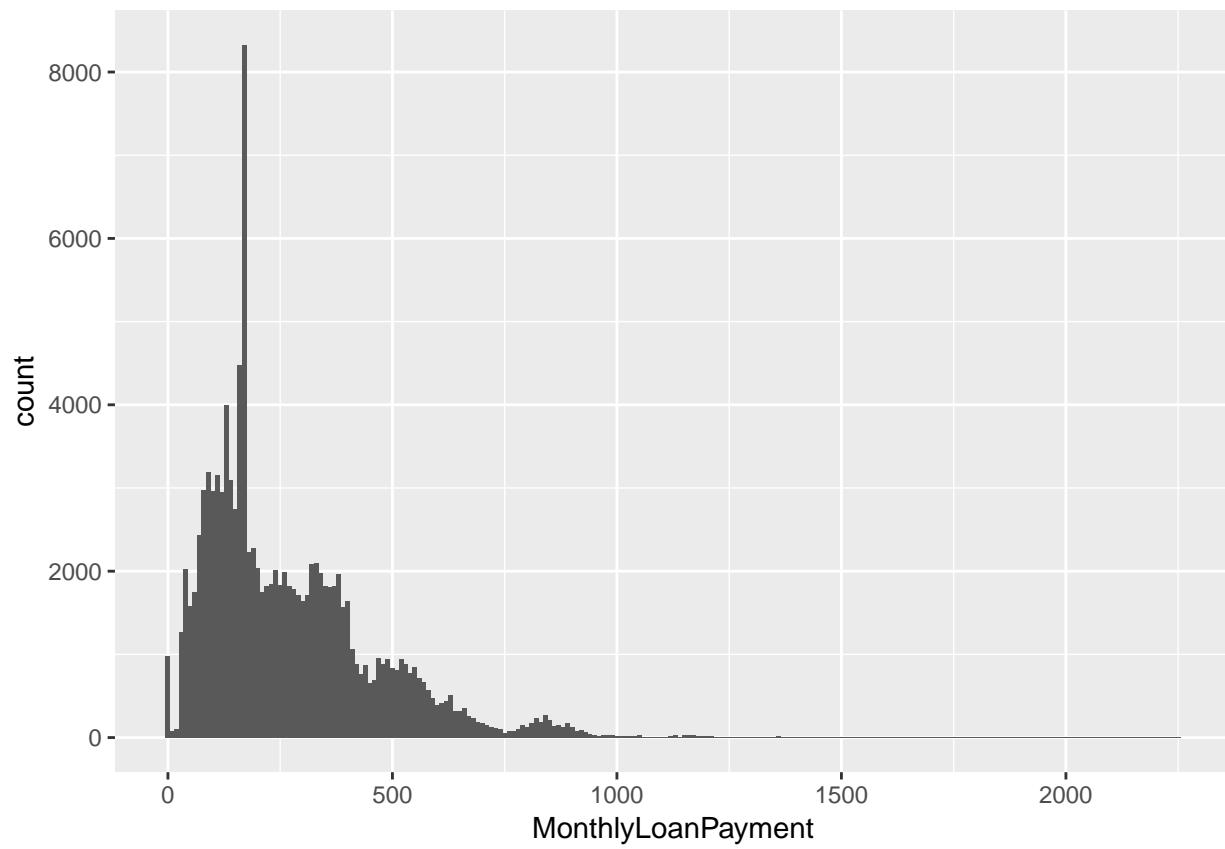


Excluding the highest 1% of debt-to-income ratios, we see that the majority of borrowers are centered around 0.2.



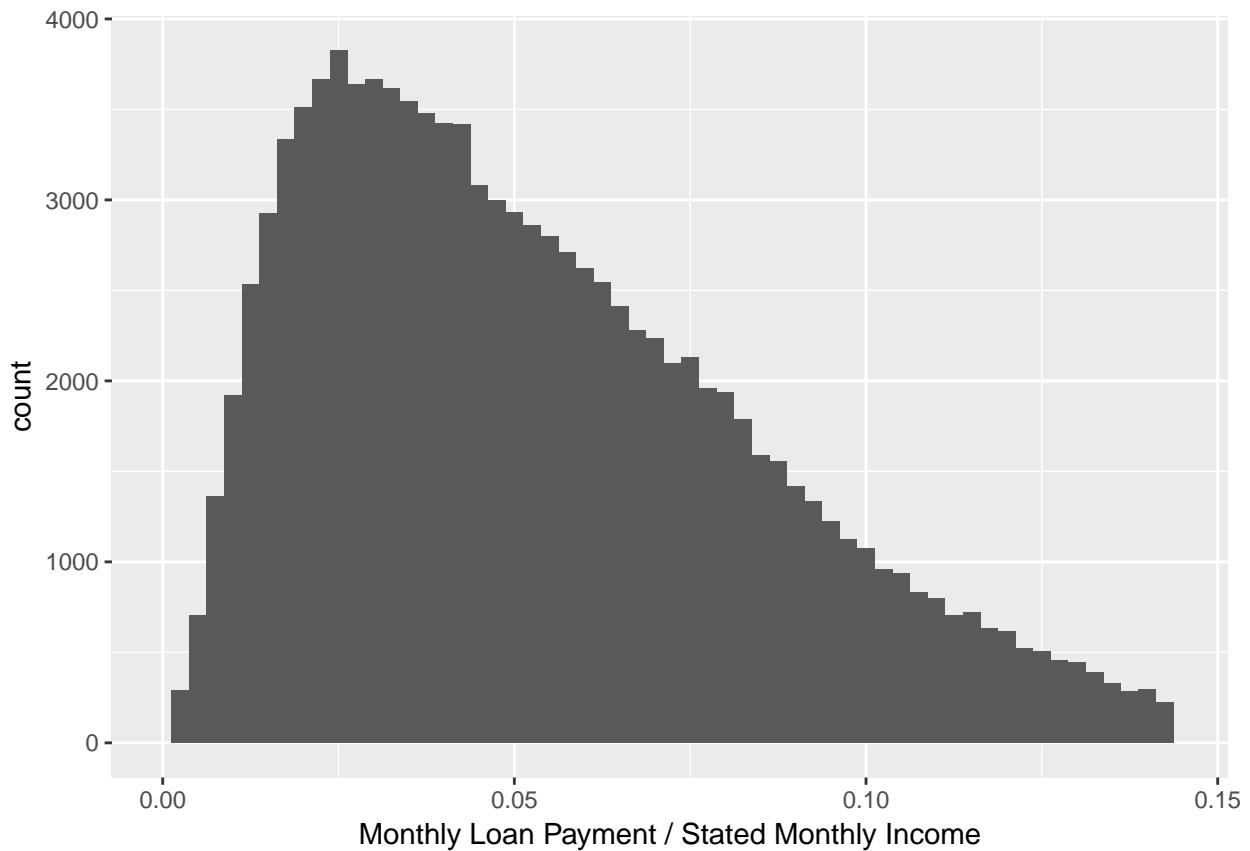
```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##    1000    4000   6500    8337   12000   35000
```

Median loan amount is \$6,500, but range from as low as \$1,000 up to \$35,000. Amounts are most frequent in \$5,000 intervals.

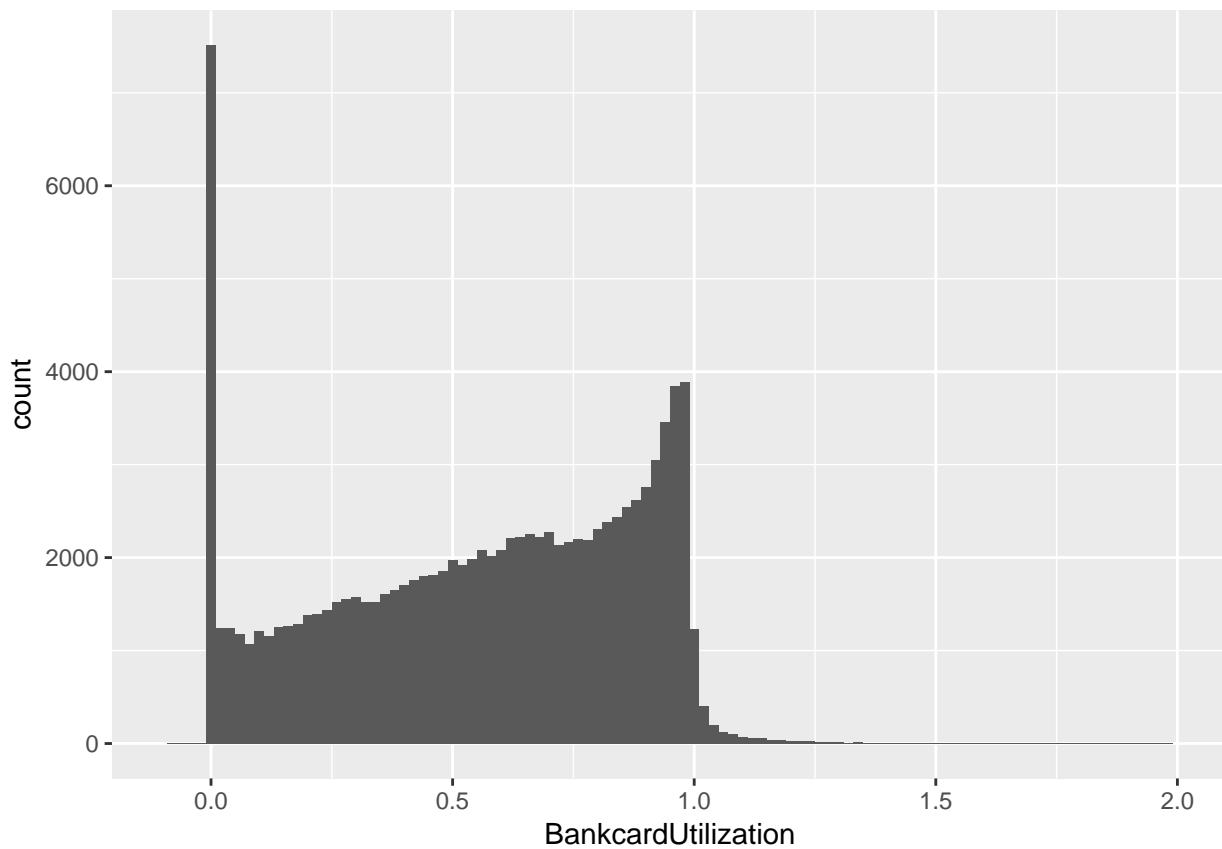


```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
##      0.0   131.6  217.7   272.5  371.6  2252.0
```

Monthly loan payments are for the most part below \$500.

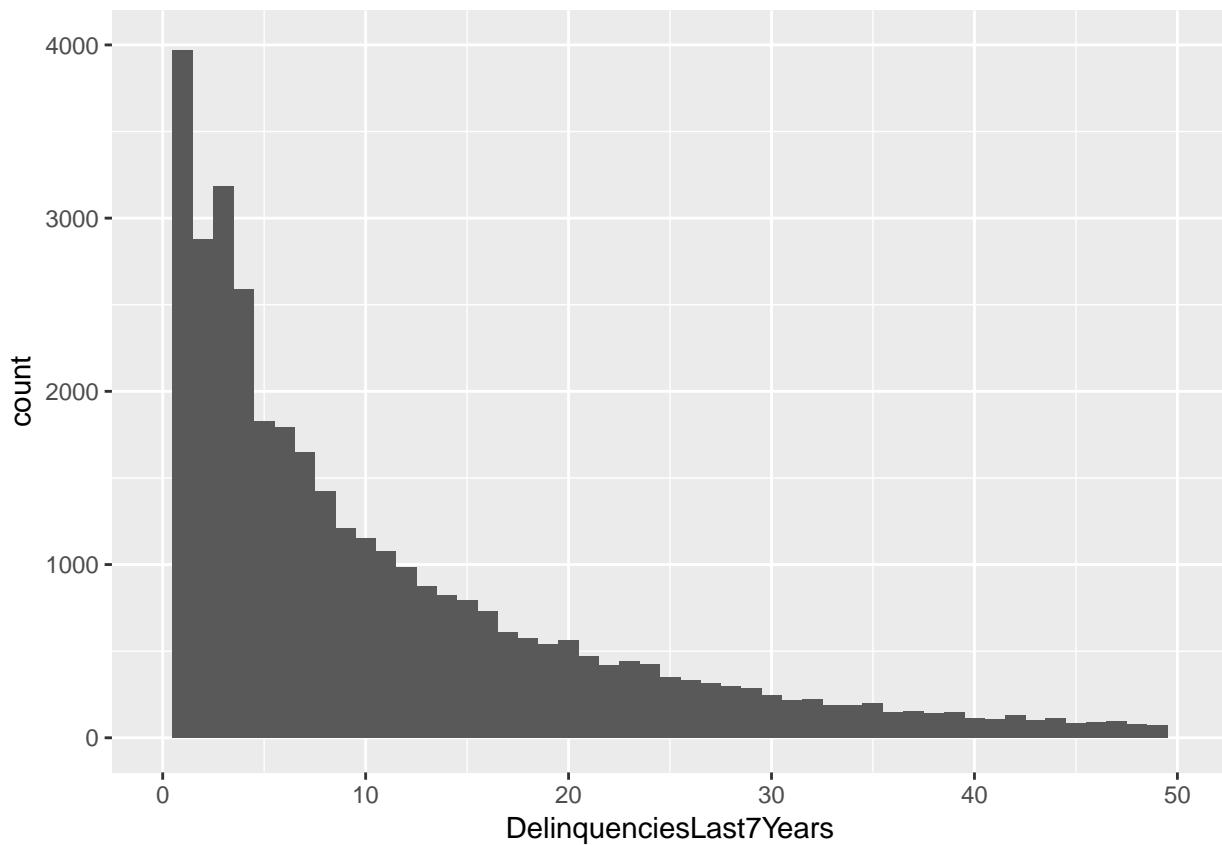


Dividing the monthly loan payment by the borrower's stated monthly income, we see that most borrowers have a monthly payment that is less than 5% of their income. However there is a significant right skew to this graph. I created this ratio because it stands to reason that borrowers paying a high amount of their monthly income will struggle to repay their loan.



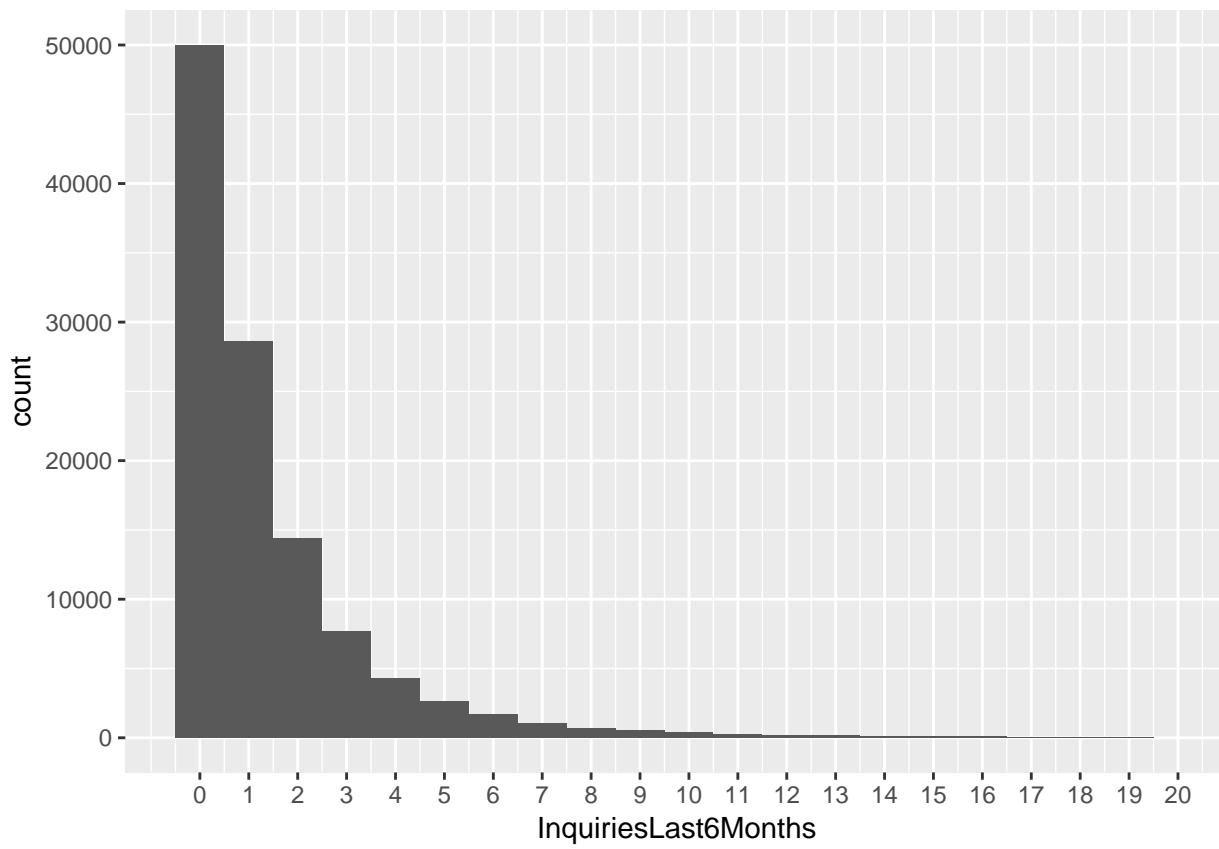
```
##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.    NA's
## 0.000   0.310   0.600   0.561   0.840   5.950  7604
```

Bankcard Utilization shows a large number of borrowers that have either used none of their available credit or all of it.



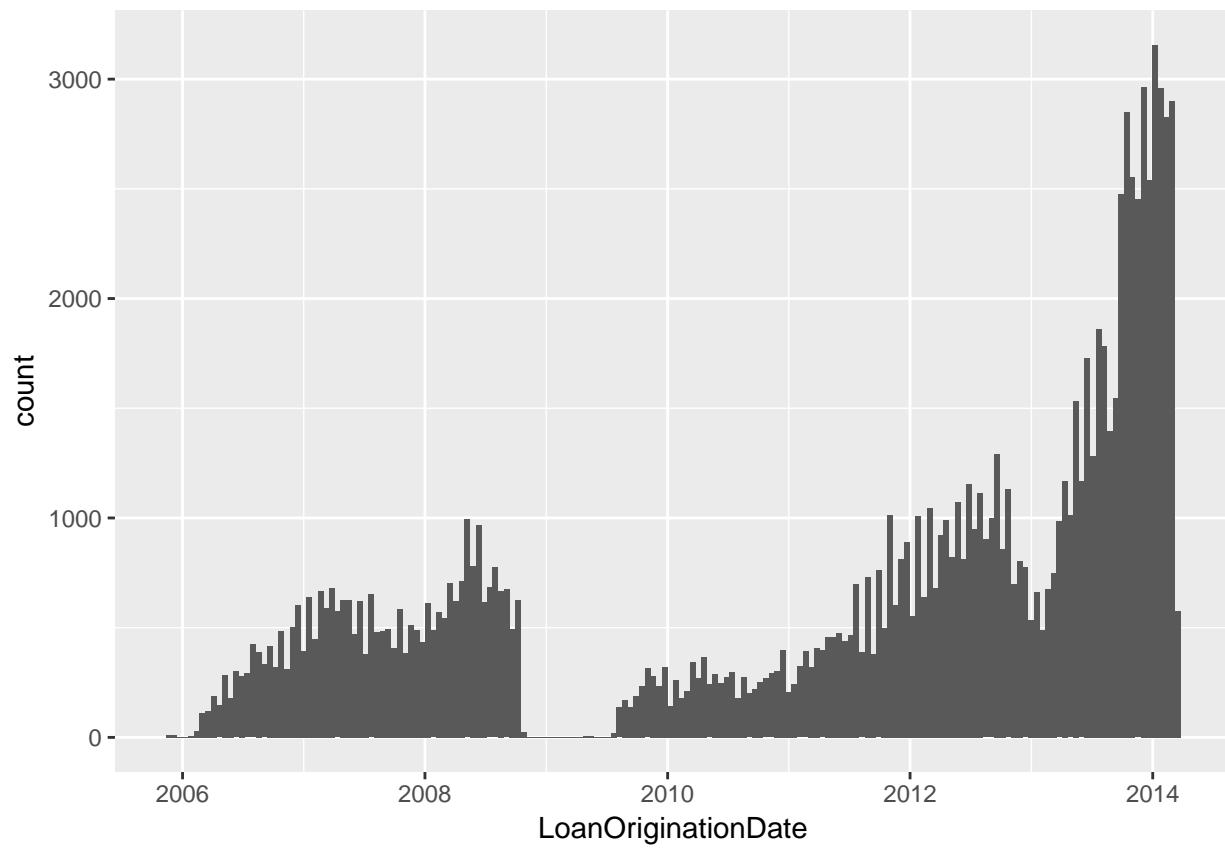
```
## # A tibble: 3 × 2
##   `loanData$DelinquenciesLast7Years > 0`     n
##   <lgl> <int>
## 1 FALSE 76439
## 2 TRUE  36508
## 3 NA    990
```

Of the more than 36,000 borrowers that had delinquent accounts within the last 7 years on their credit report, the majority had less than 5.

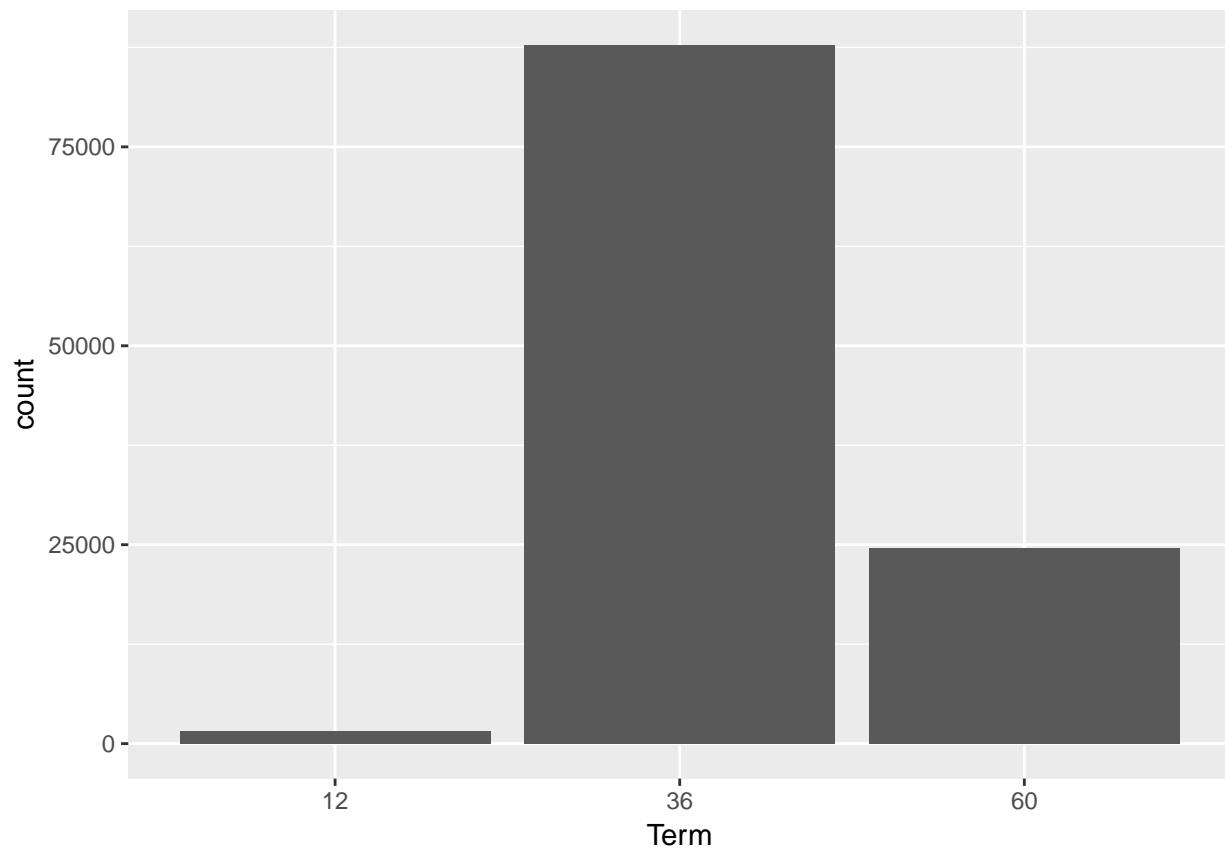


```
## # A tibble: 3 × 2
##   `loanData$InquiriesLast6Months > 0`     n
##   <lgl> <int>
## 1 FALSE 50005
## 2 TRUE  63235
## 3 NA    697
```

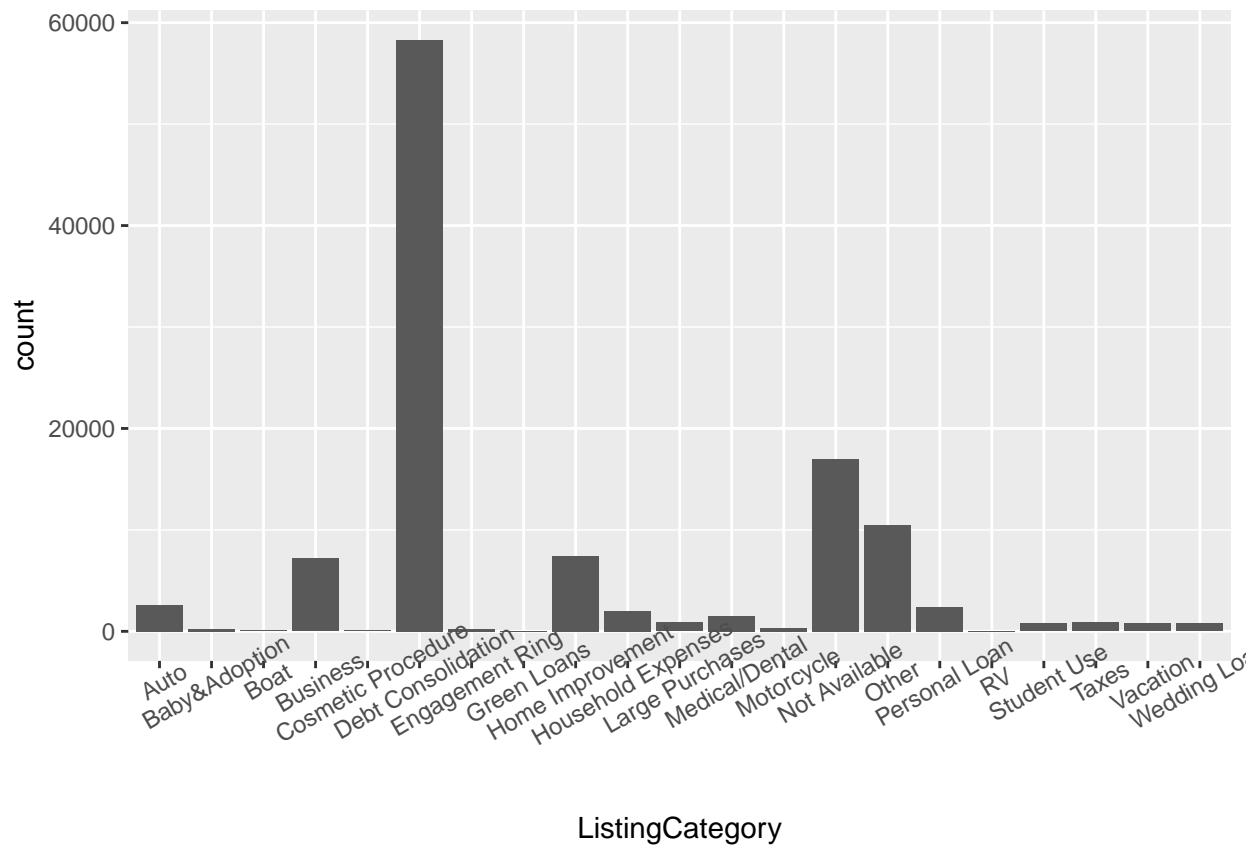
Many of the borrowers did not have any credit inquiries in the 6 months leading up to their loans.



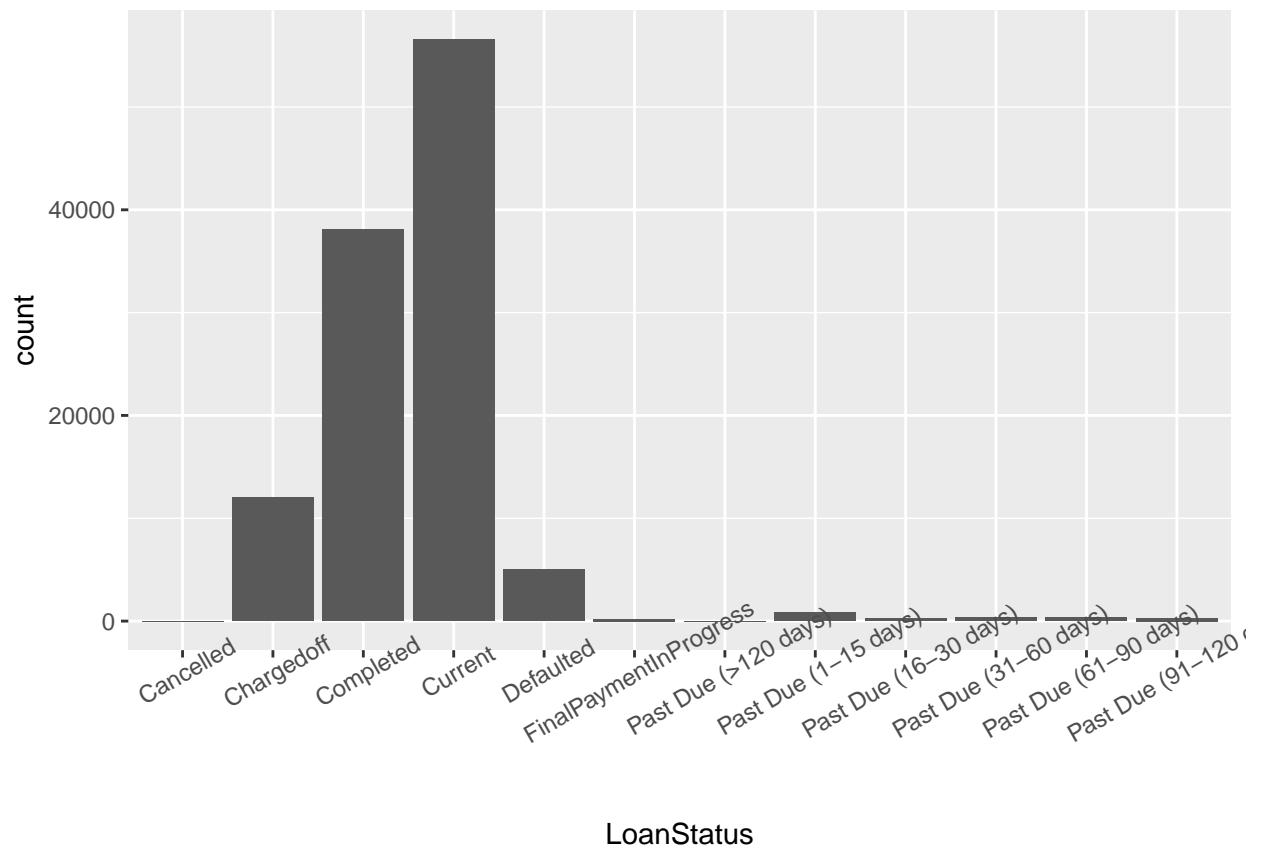
Most of the loans were made in 2013 and 2014, the last 2 years of the data. Note the gap with no loans in late 2008 and first half of 2009.



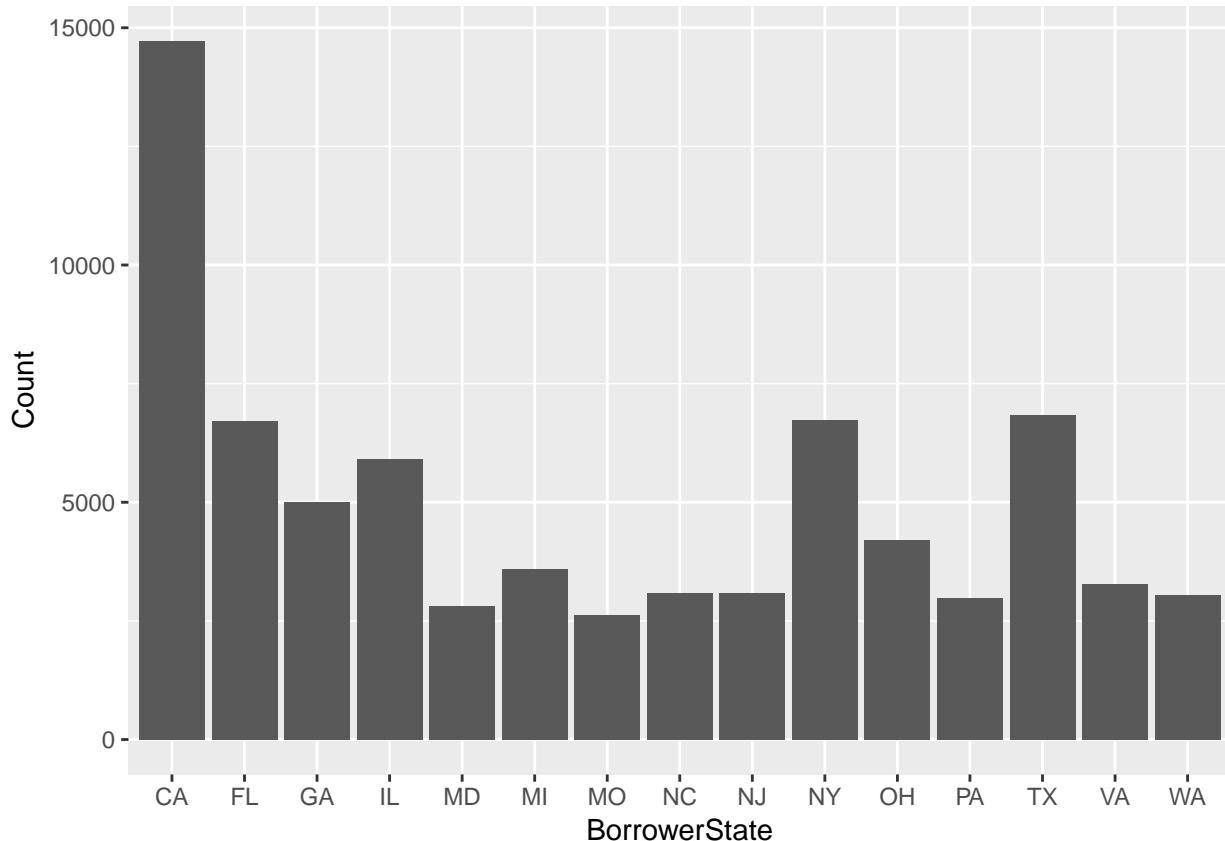
The majority of loans are short term at 36 months to maturity.



Debt Consolidation appears as the most frequent reason for borrowing.



The high charge-offs and defaulted loans in this dataset are worth taking a closer look at to see if there are any possible causes. Separately, the high amount of completed loans is expected given that the majority of loans are 36 months and the dataset spans just shy of 10 years.



California leads the way with the most borrowers, with some of the other more populated states trailing behind.

Univariate Analysis

Observations

There are 113,937 loan records in the dataset with 81 variables about the borrower and their loan. Some of the key variables are the borrower rate, credit score, debt-to-income ratio, bankcard utilization rate.

- Most of the borrowers have a credit score of approximately 680.
- Median borrower rate is 0.184, but exhibits a right skew with a spike around 0.33.
- Median debt-to-income ratio is 0.22.
- Median bankcard utilization is 0.60 with most borrowers at 0 or 1.

Further Exploration

I am interested in taking a closer look at the borrower's credit history to see what relationships there may be with a borrower's interest rate. Based on some of the plots we saw earlier, I would expect there to be a few variables that explain some of the rate differentials. Some variables that might contribute are the number of delinquencies, credit inquiries, debt-to-income ratio, and their bankcard utilization rate. Also, some of the qualitative variables will be interesting to look at to see if they contributed to any disparity. For example, does the borrower's geography or purpose of loan have a noticeable impact.

Also, this dataset provides an interesting time frame in that it includes loans made at the peak of the economic cycle (2006-2007) through one of the worst financial collapses in recent history. I would expect defaults to be

high in 2009 and 2010 as many people became unemployed.

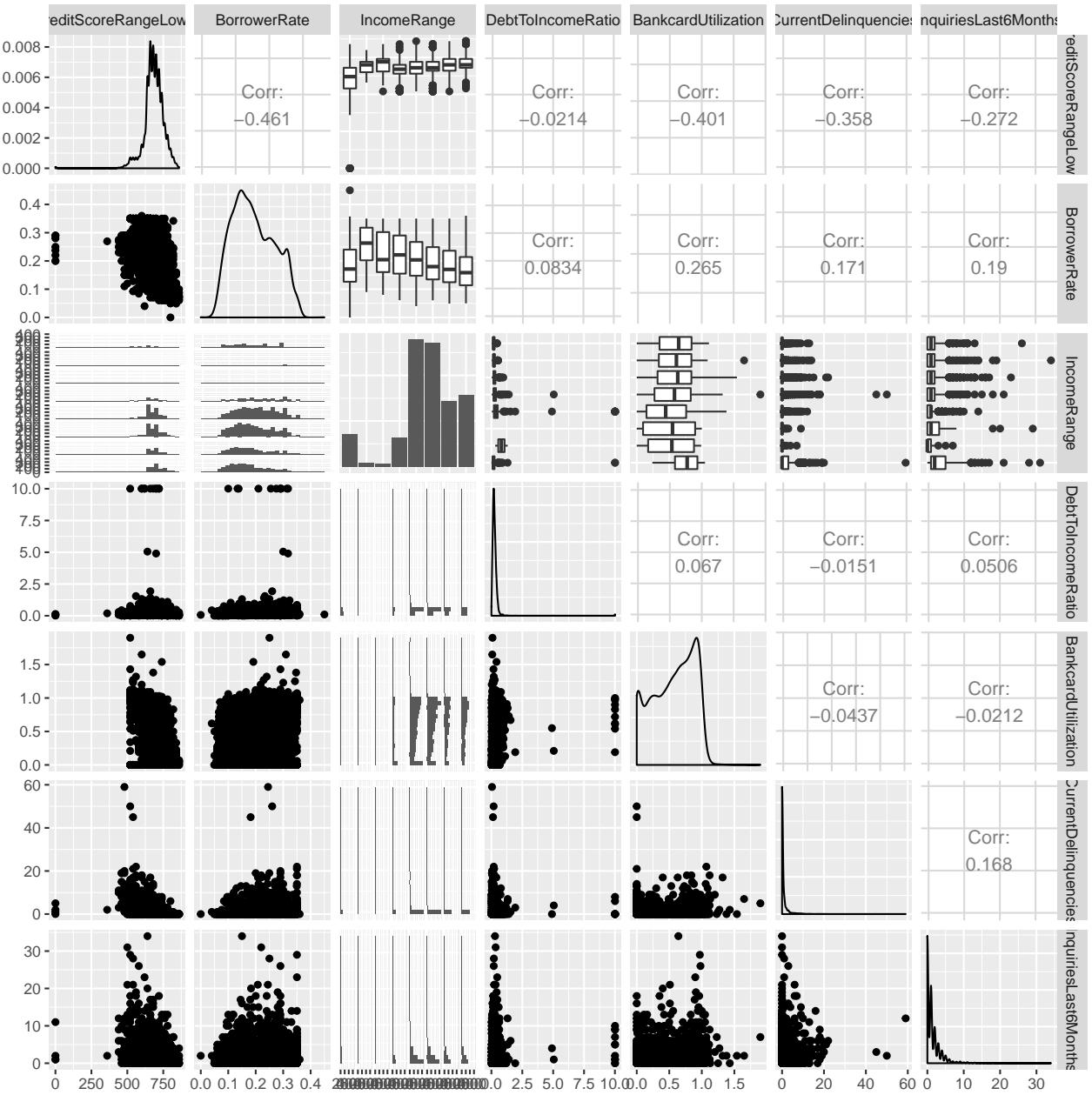
The time period of this data should be an interesting factor as the financial collapse in 2008-2009 resulted in many people losing their jobs and therefore income. I would expect defaults to be high in 2008 through 2010 as many people became unemployed.

Data Adjustments

I made some adjustment to the dataset. Specifically, I created a new variable for the Listing Category to show what it is rather than a vague numerical value. Also, I created Credit Score Buckets to group credit scores for my analysis.

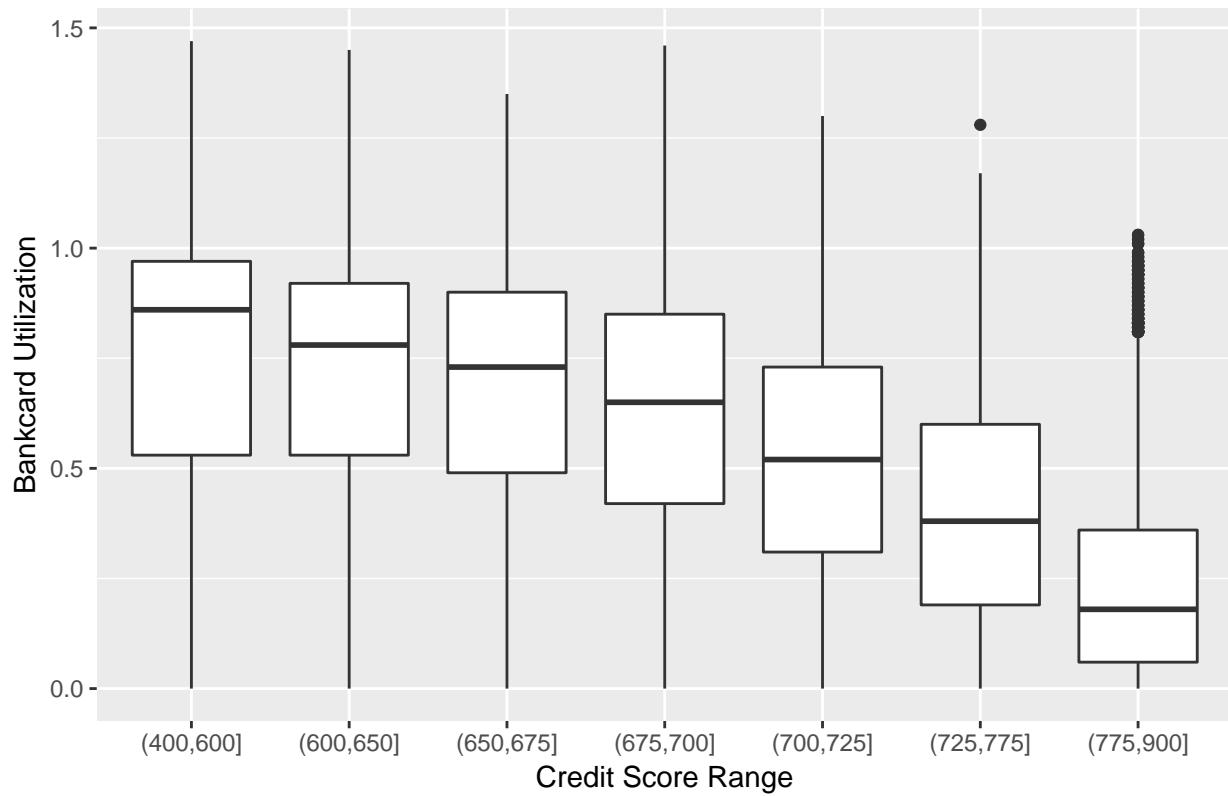
Bivariate Plots Section

A quick look at a few of the variable relationships shows some meaningful correlations with the borrower's credit score. Also, the credit score appears to have a fairly strong negative correlation with the borrower's rate as would be expected.

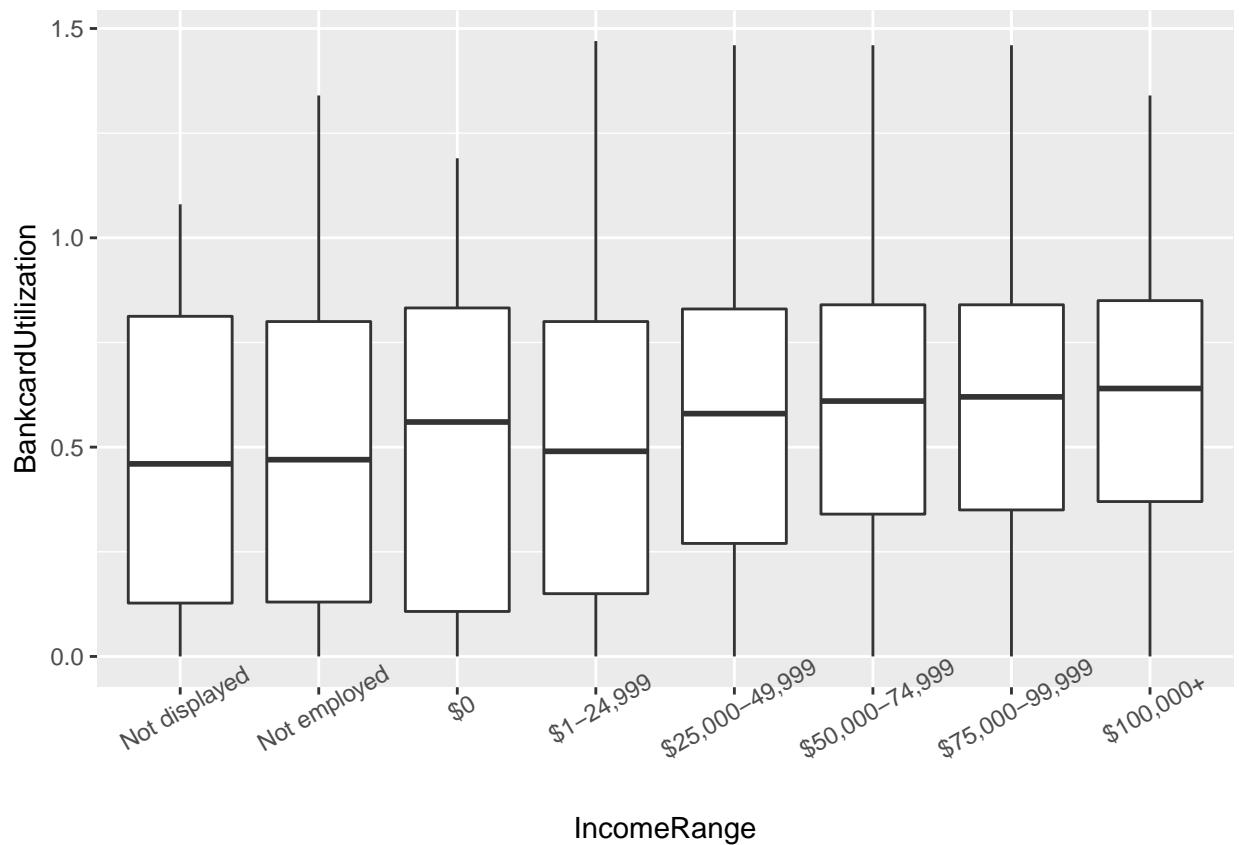


I quickly notice a couple of meaningful correlations for variables relative to the borrower's credit score.

Bankcard Utilization by Credit Score Ranges

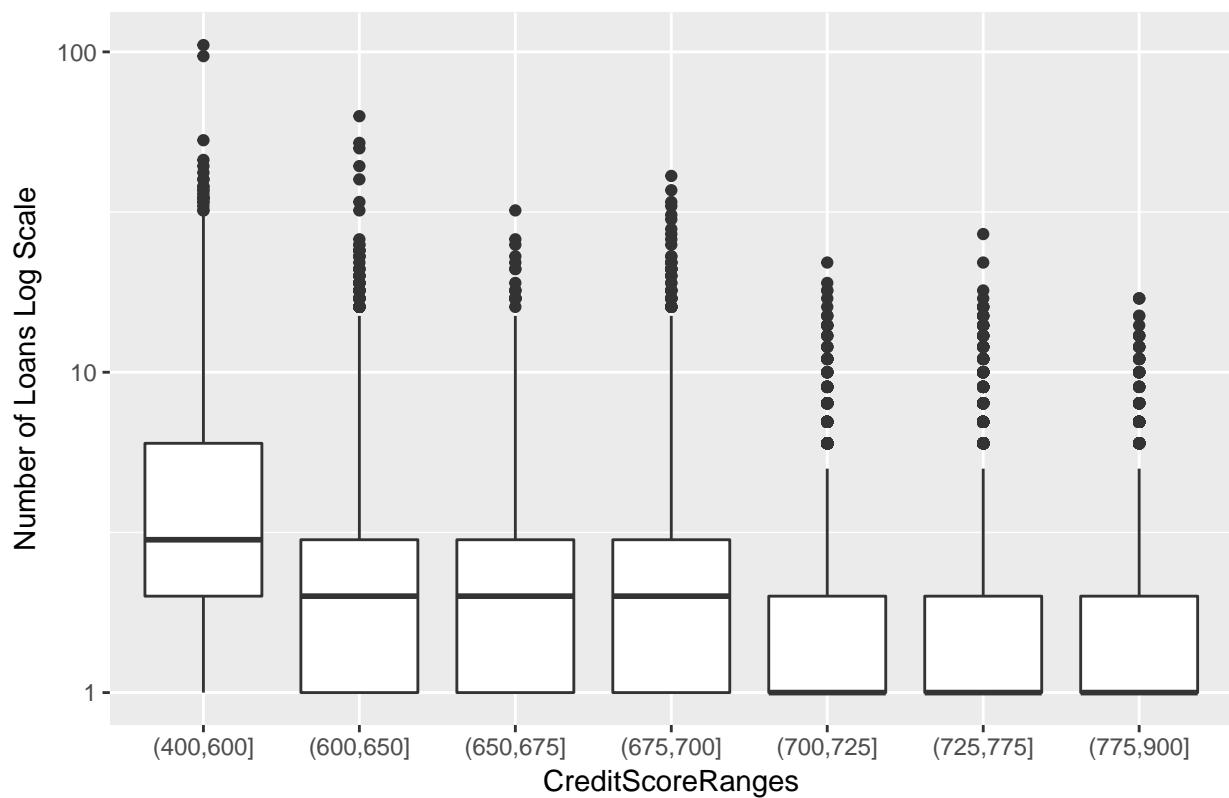


I took a subset of the data here to only include borrowers with a bankcard utilization below 1.5. We can see the correlation between the two variables well here as the range and median of bankcard utilization rises as credit scores decline.



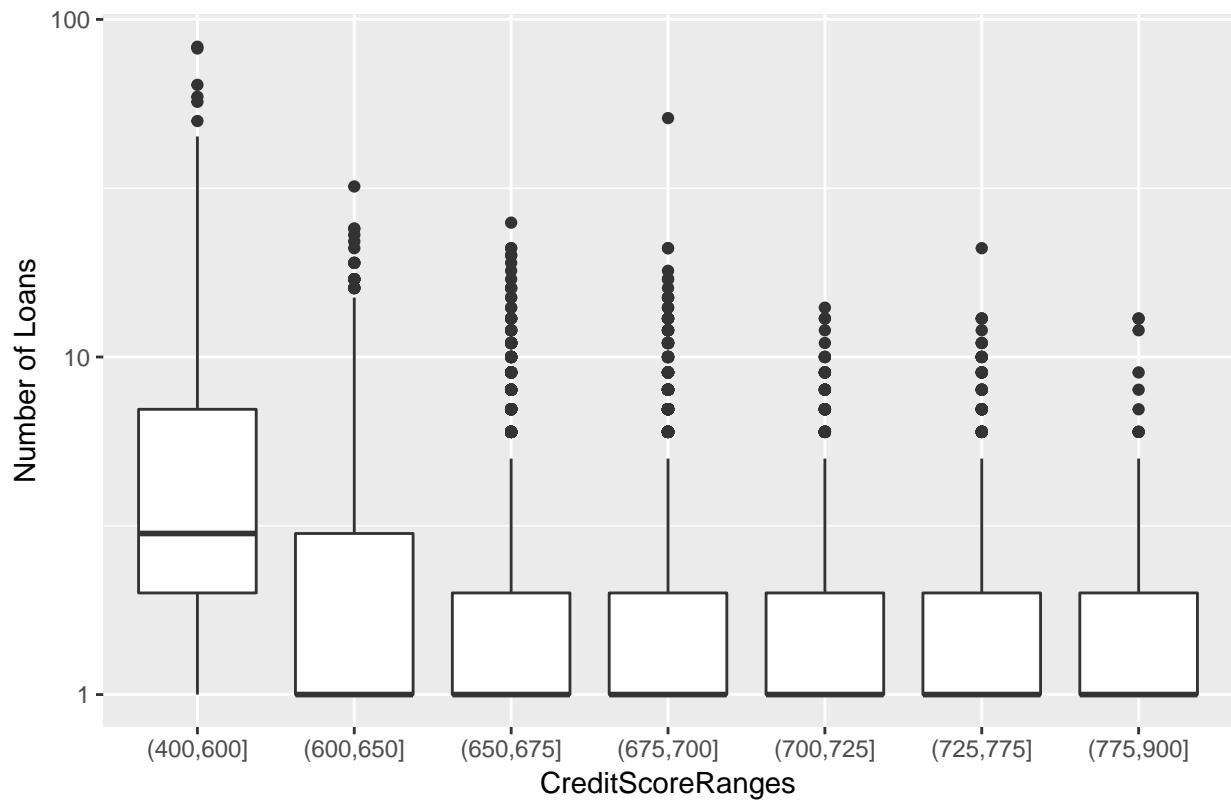
Bankcard Utilization rises slightly with income.

Credit Inquiries in the Last 6 Months – Log Scale



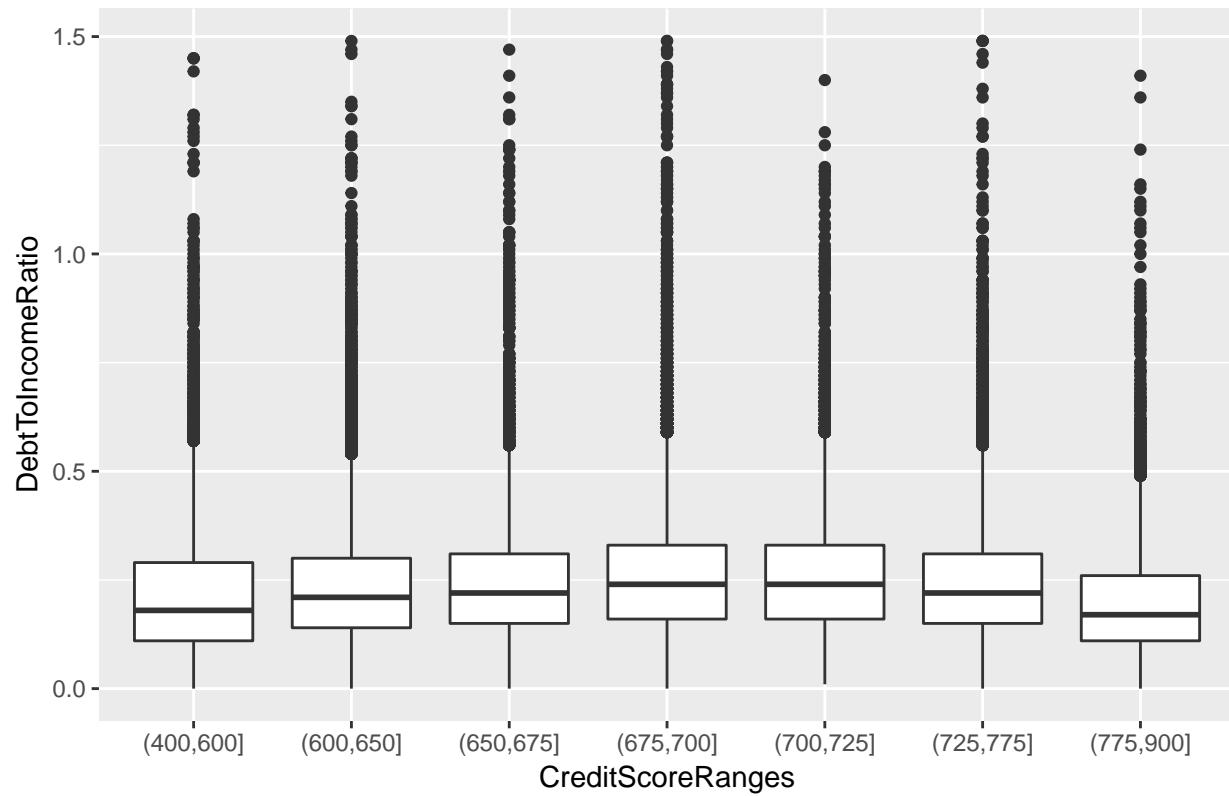
Inquiries in the last 6 months do not appear as meaningful as I was expecting. Typically 1 inquiry on a credit report will not cause a drop in a credit score. However, there does appear to be some significance to being above or below a credit score of 700.

Current Delinquent Accounts in Credit Report



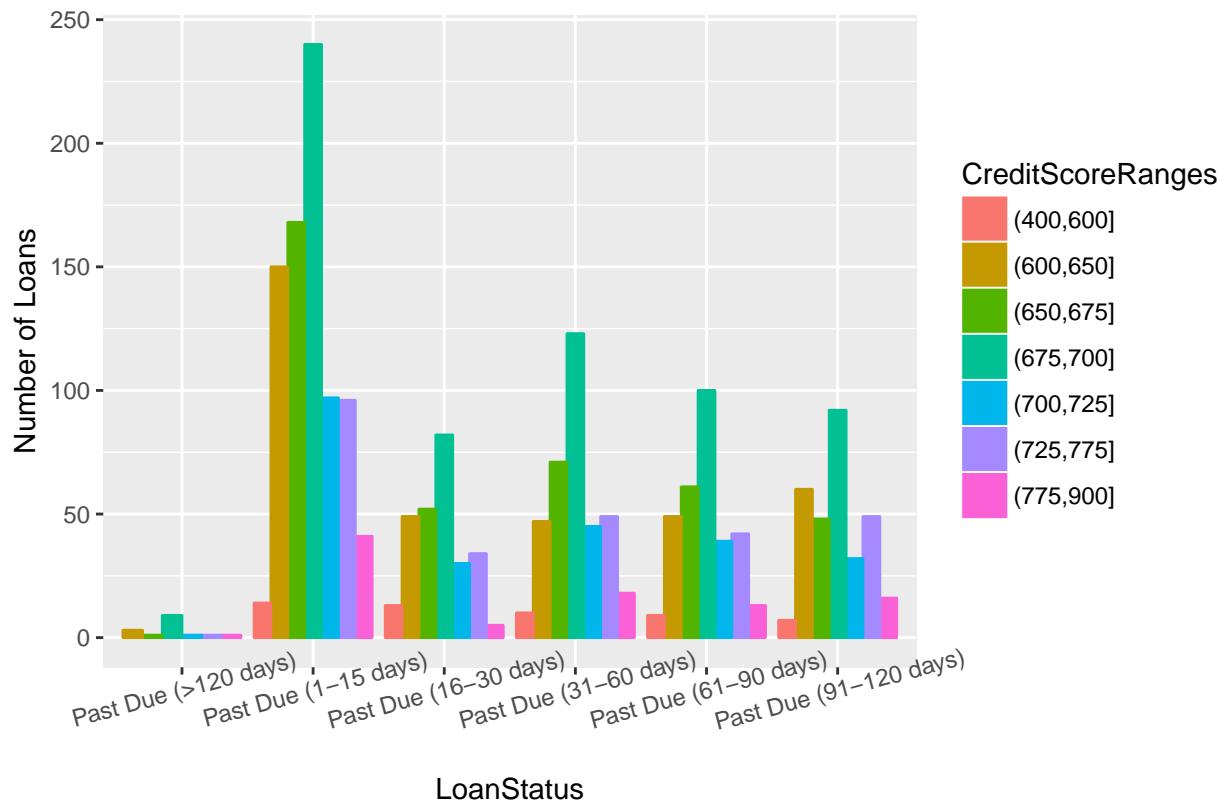
Higher current delinquencies tend to be associated with lower credit scores. The outliers for delinquencies begin to trend higher starting with borrowers below 700 credit scores.

Debt-to-Income Ratio by Credit Score



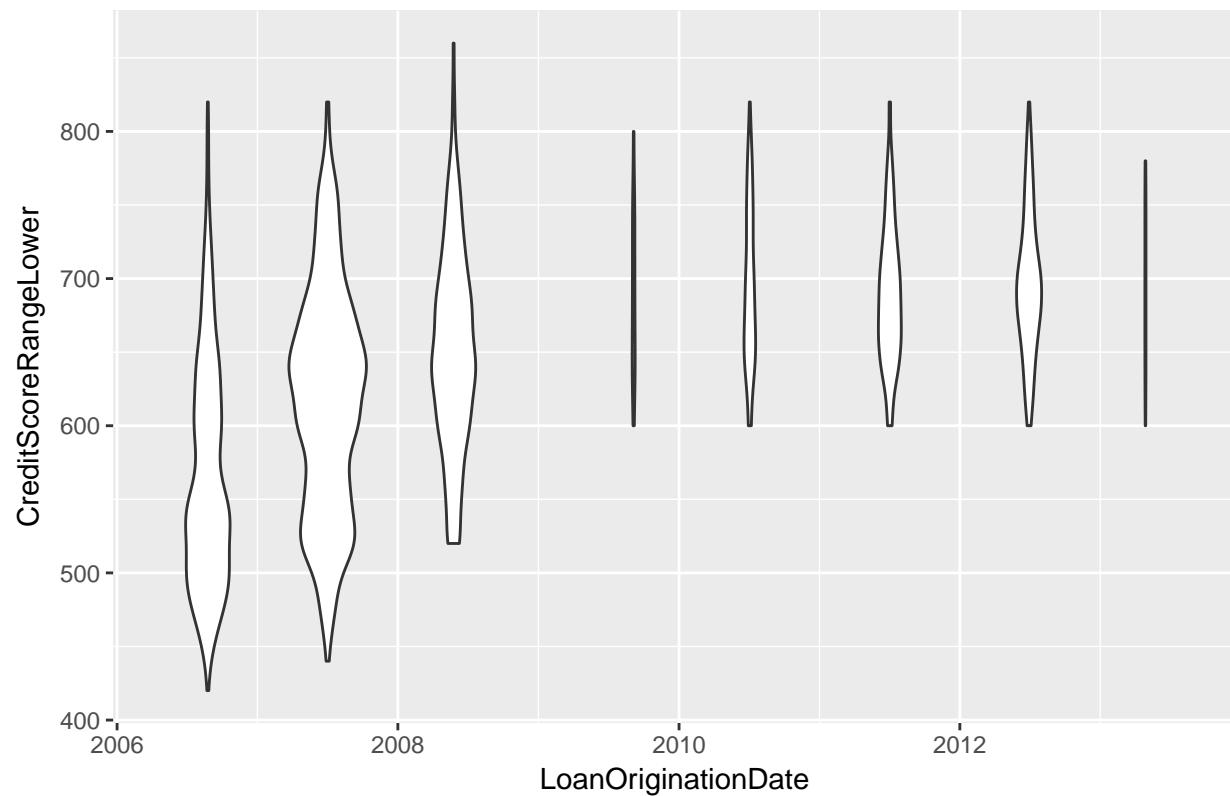
Debt-to-income shows normal distribution across credit scores and appears to have no correlation.

Loans Past Due by Credit Score Range



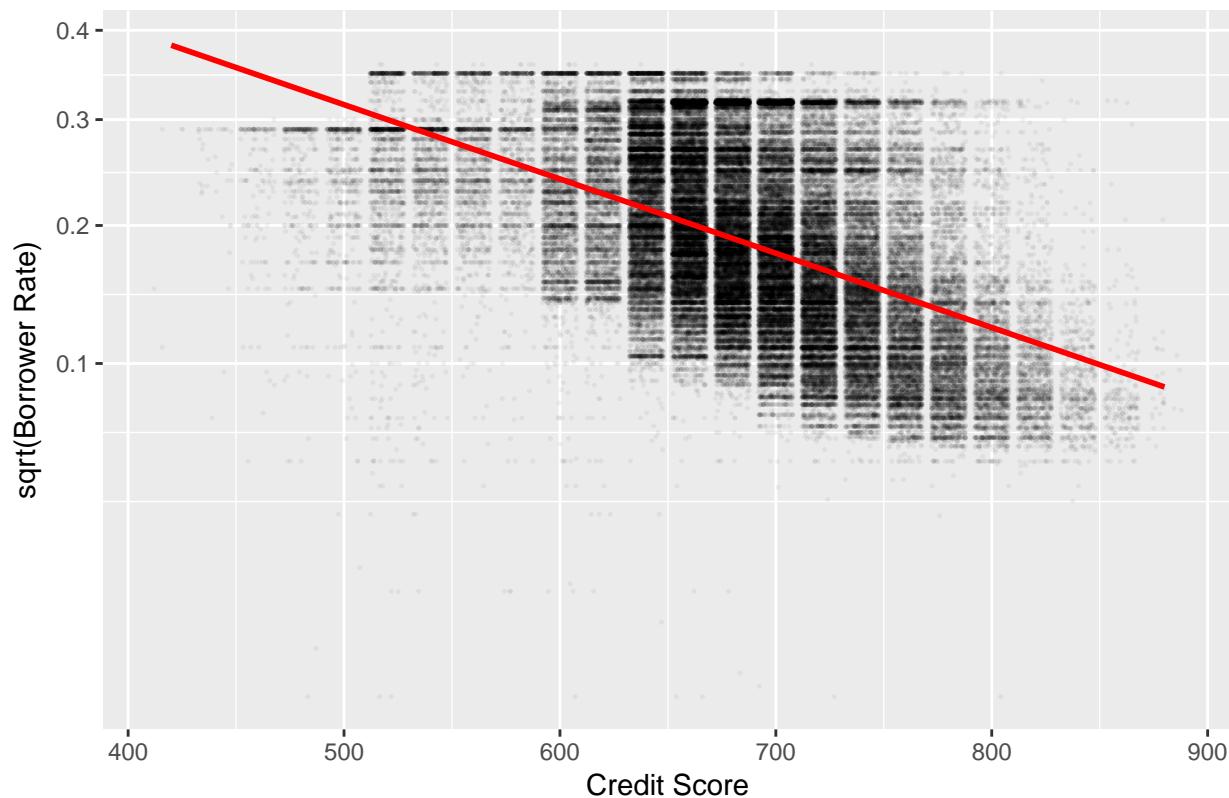
Lower credit quality borrowers appear more likely to fall behind on their payments.

Defaulted Loans by Origination Date and Credit Score



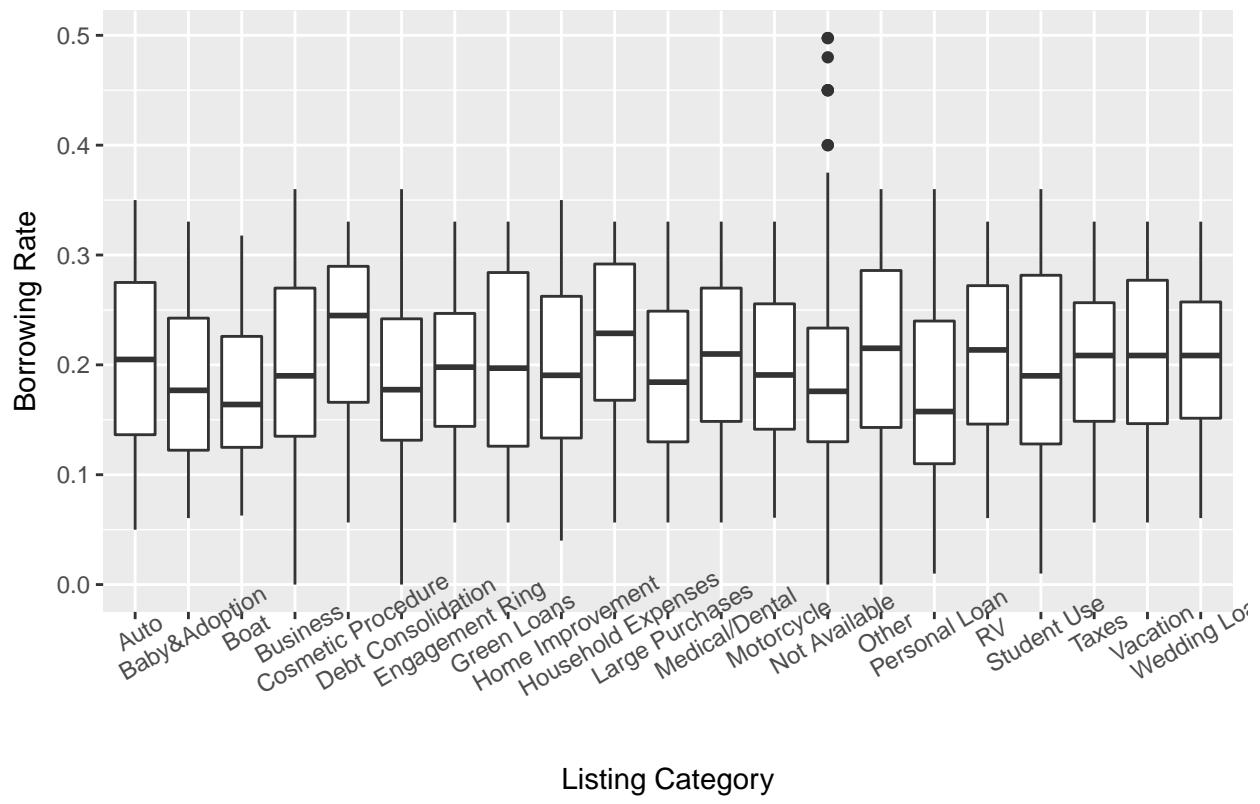
Loan defaults were much greater in the time period leading up to the financial crisis than afterwards. The lower credit score borrowers clearly were more likely to default prior to the crisis and again around 2011 and 2012.

Borrower Rate by Credit Score



Credit scores show some correlation with the borrower's interest rate with an R^2 of -0.46.

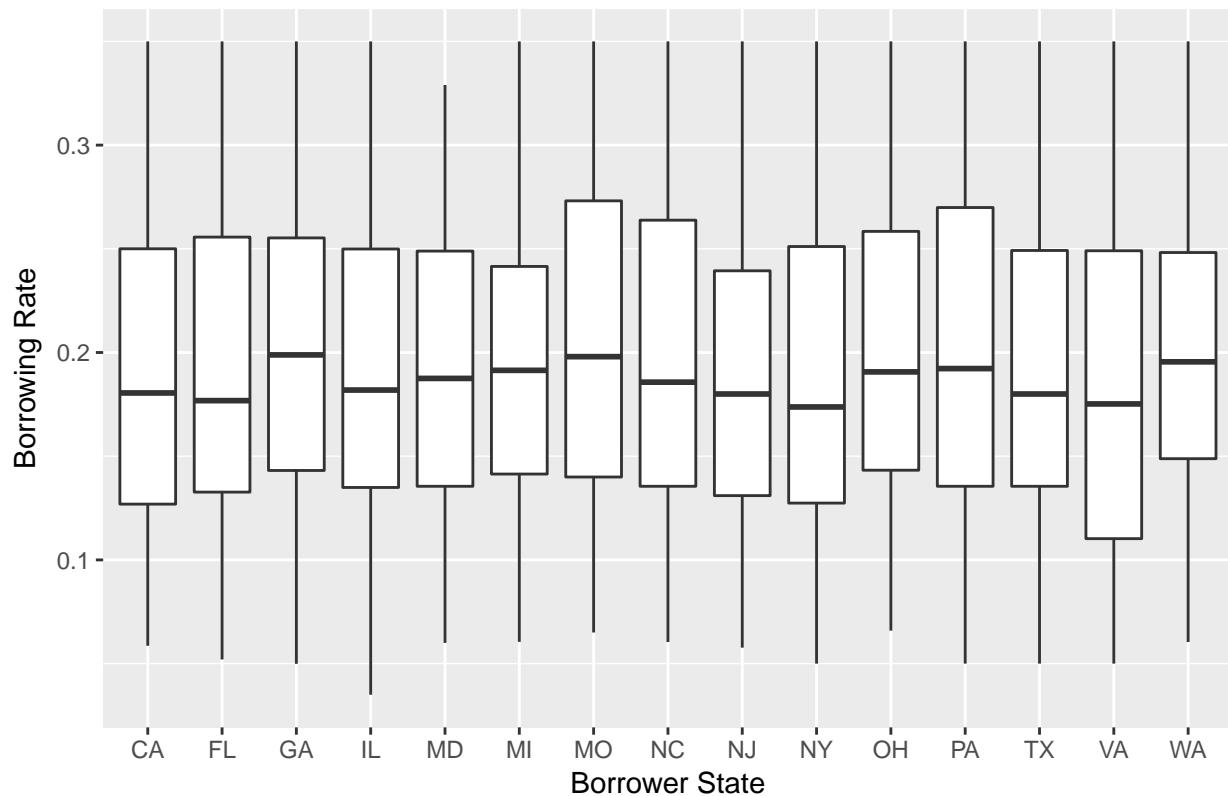
Borrowing Rates by Loan Purpose



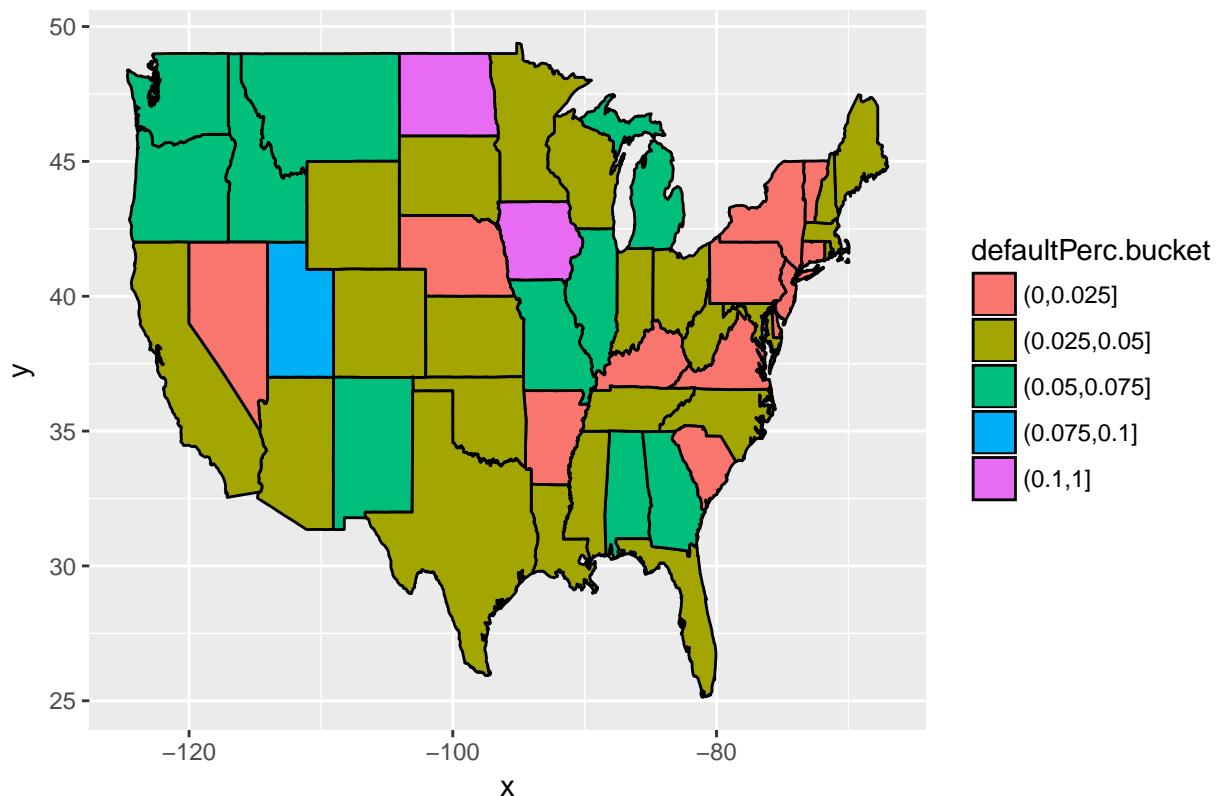
Listing Category

Loan purposes listed as Not Available have the greatest outliers. Cosmetic procedures and Household expenses have the highest rates on average.

Borrowing Rates by State



% of Loans in Default by State



As a percentage of the total loans in a state, the northeast region looks to be the least likely area to default, whereas the northwest and midwest show the greatest potential for default.

Bivariate Analysis

Observations

It was interesting to note some of the variables that had a correlation with the credit score. As expected, the bankcard utilization, current delinquencies and number of inquiries in the last 6 months all exhibited some correlation. However, I was a little surprised that the debt-to-income ratio showed no correlation. While income may not factor into the credit score, I would have expected an indirect relationship here where borrowers that have high debt-to-income are more likely to fall behind on payments and/or utilize a large portion of their available credit.

Near the beginning of the analysis it was noted that a spike in borrowing rates was observed around 33%. This appeared with some surprising results when graphed against credit scores. The data showed a negative correlation between credit scores and borrower rates, but for the subset of data with rates around 33% showed a high concentration corresponding to borrowers with credit scores above what might be expected. Based on the graph of the relationship, I would not have expected many loan rates to coincide with credit scores over 700.

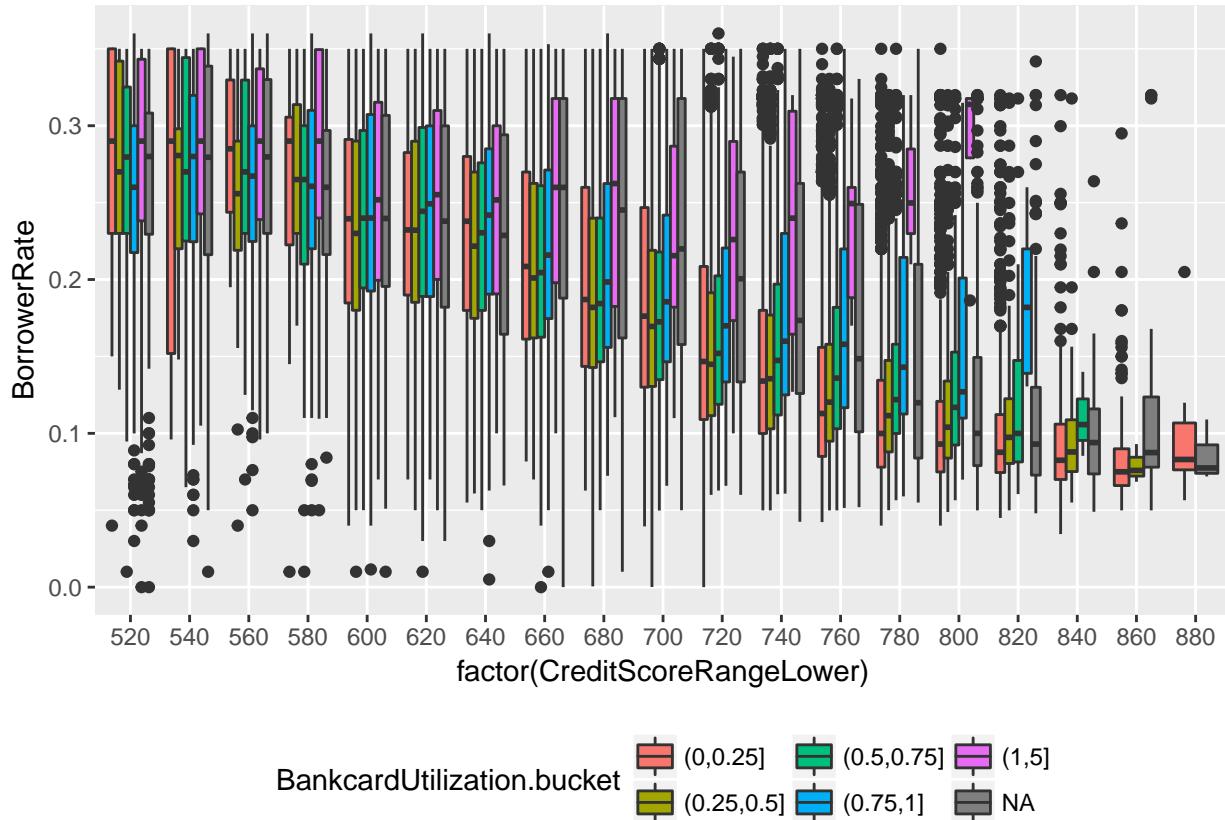
The number of credit inquiries and current delinquencies both showed that a higher occurrence in either was more likely to be associated with borrowers having lower credit scores.

Defaults were more frequent among low credit score borrowers which was to be expected. However, the more interesting aspect of the graph was the impact that exogenous factors have on the repayment of a loan. While a credit score can quantify a borrower's ability to repay their loan based on historical information, it is unable to anticipate future variants that can lead to unexpected results. This is evident in the high frequency of defaults preceding the 2008-2009 financial crisis when many people became unemployed.

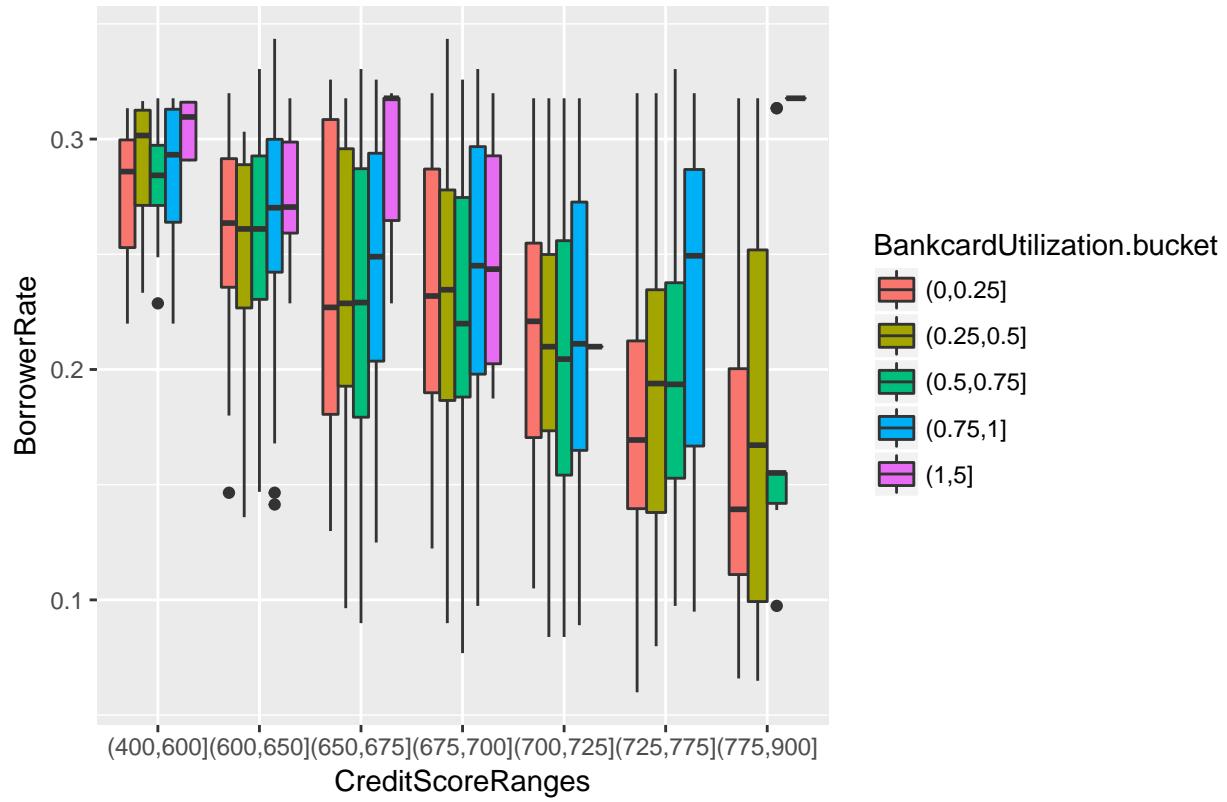
The rise in borrower rates in 2011 through 2012 was interesting to note because the FHLB Boston 3 year fully amortizing rate, a benchmark rate used by commercial banks, stayed at very low levels during this period. Since credit scores are representative of certain characteristic traits, it means that either lenders were demanding a higher risk premium or borrowers were taking loans out for longer terms.

The strongest relationship in the data I explored was the Credit Score and the Borrower's Rate with a R^2 of -0.46. The Bankcard Utilization rate and the Credit Scores was not significantly different though with a R^2 of -0.40. Also, Current Delinquencies and Credit Inquiries showed some correlation with Credit Scores.

Multivariate Plots Section

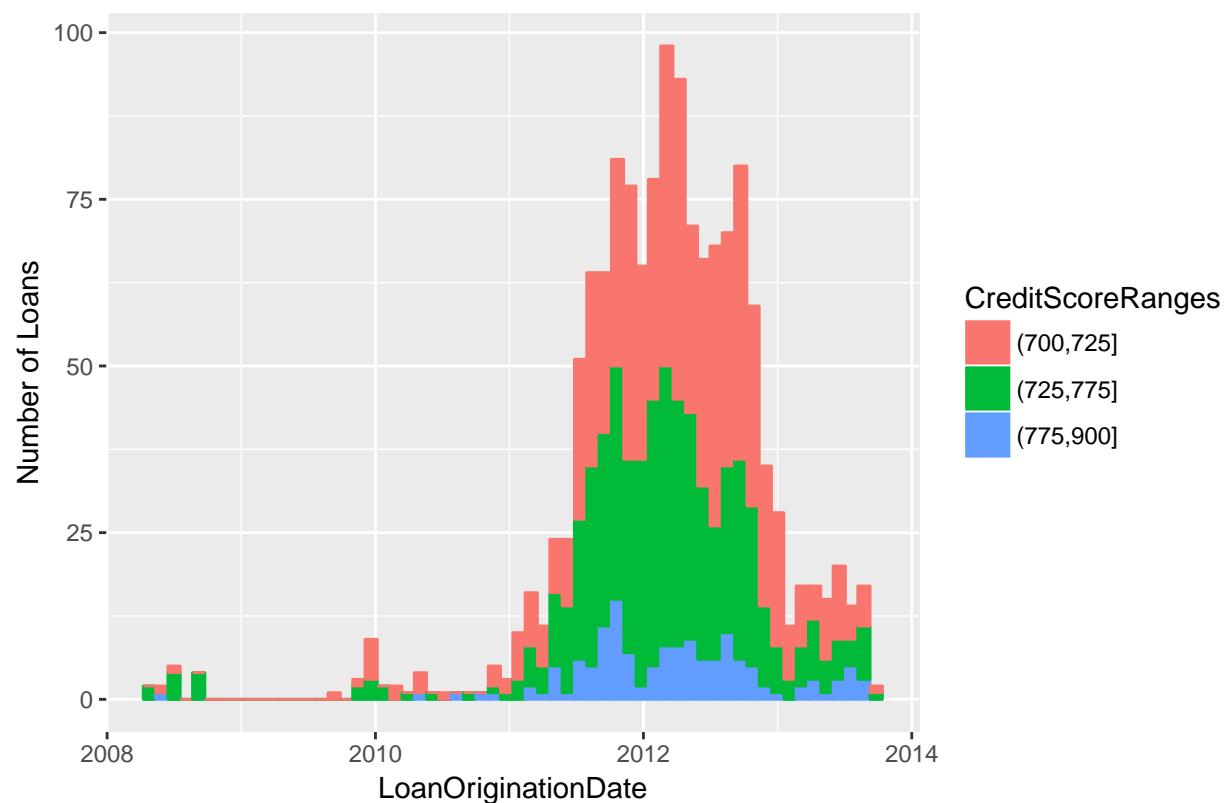


Loans Past Due

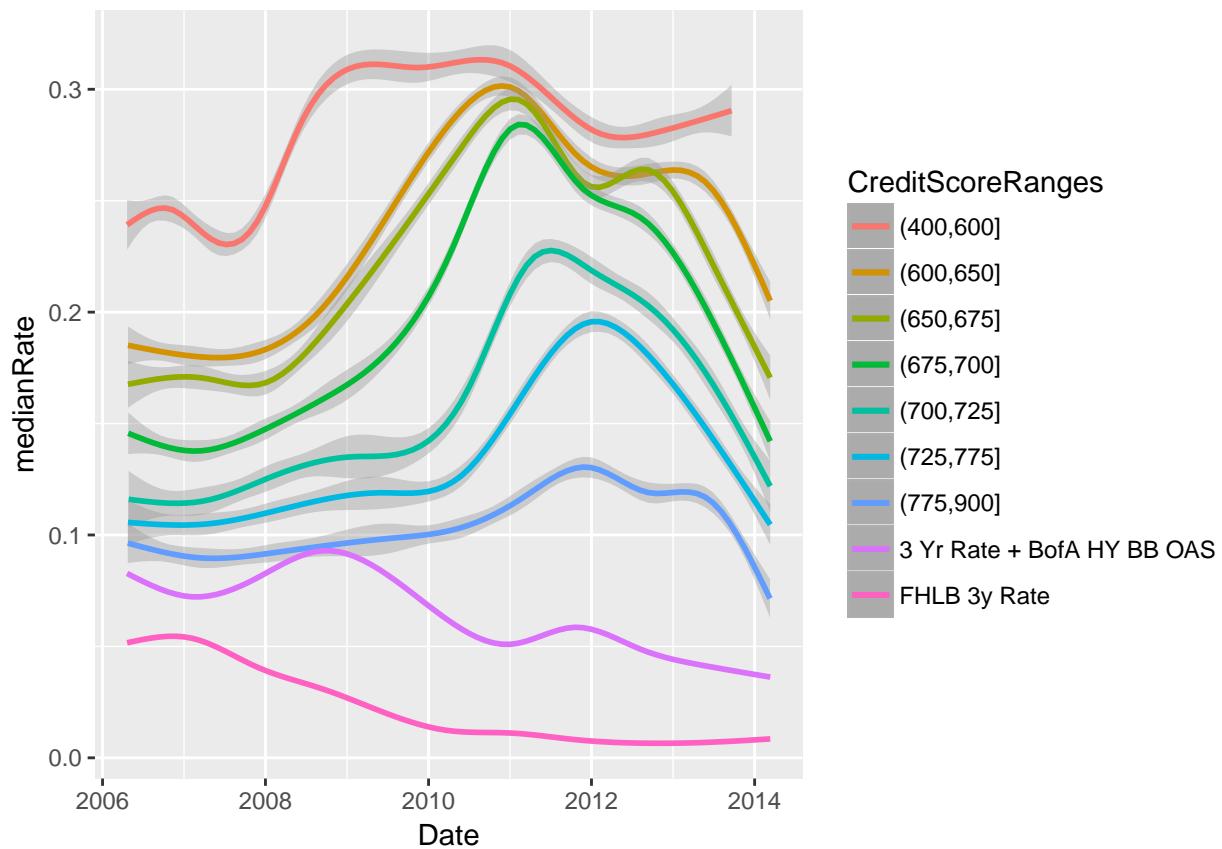


Loans that are past due tend to be higher rates and borrowers that have low credit scores and high bankcardutilization.

Loans with Borrower Rates between 31% and 34%



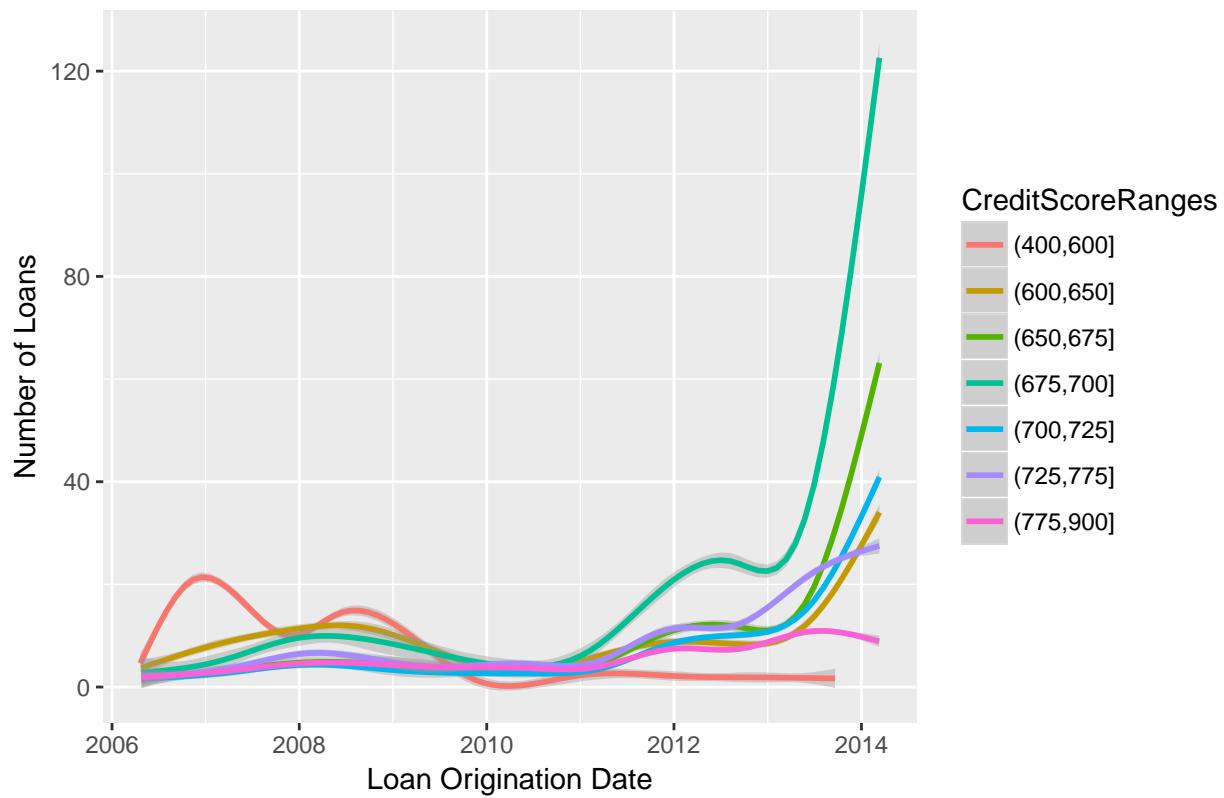
Graphing only the loans with interest rates between 31% and 34% having a credit score greater than 700, we see that these loans were almost entirely originated in 2011 and 2012.



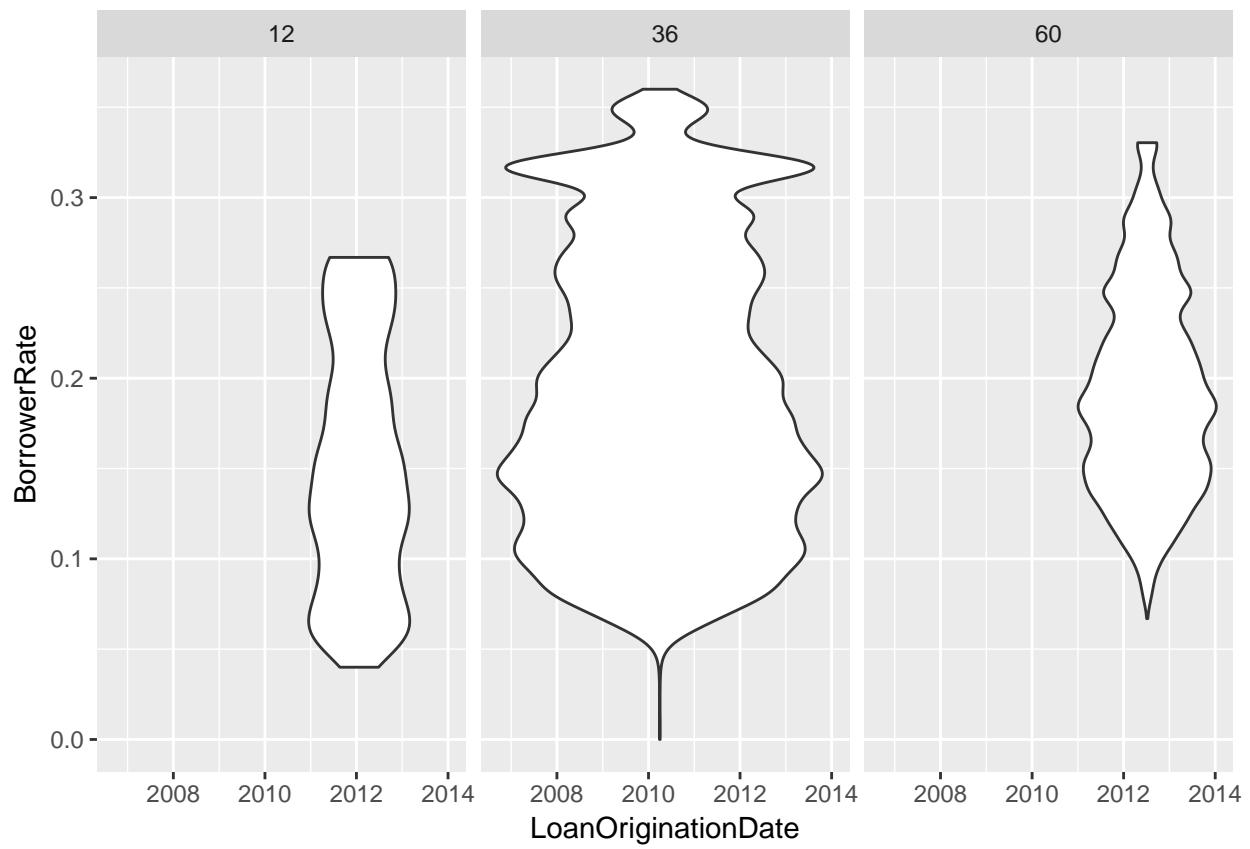
Graphing the median borrowing rates over time, we see that interest rates started rising in 2010 and remained elevated through most of 2012. Meanwhile the FHLB Boston 3 year amortizing rate, a lending benchmark rate, remained at very low levels.

There is a slight increase in the credit risk premium in 2011 that may explain some of the rise in borrower rates observed.

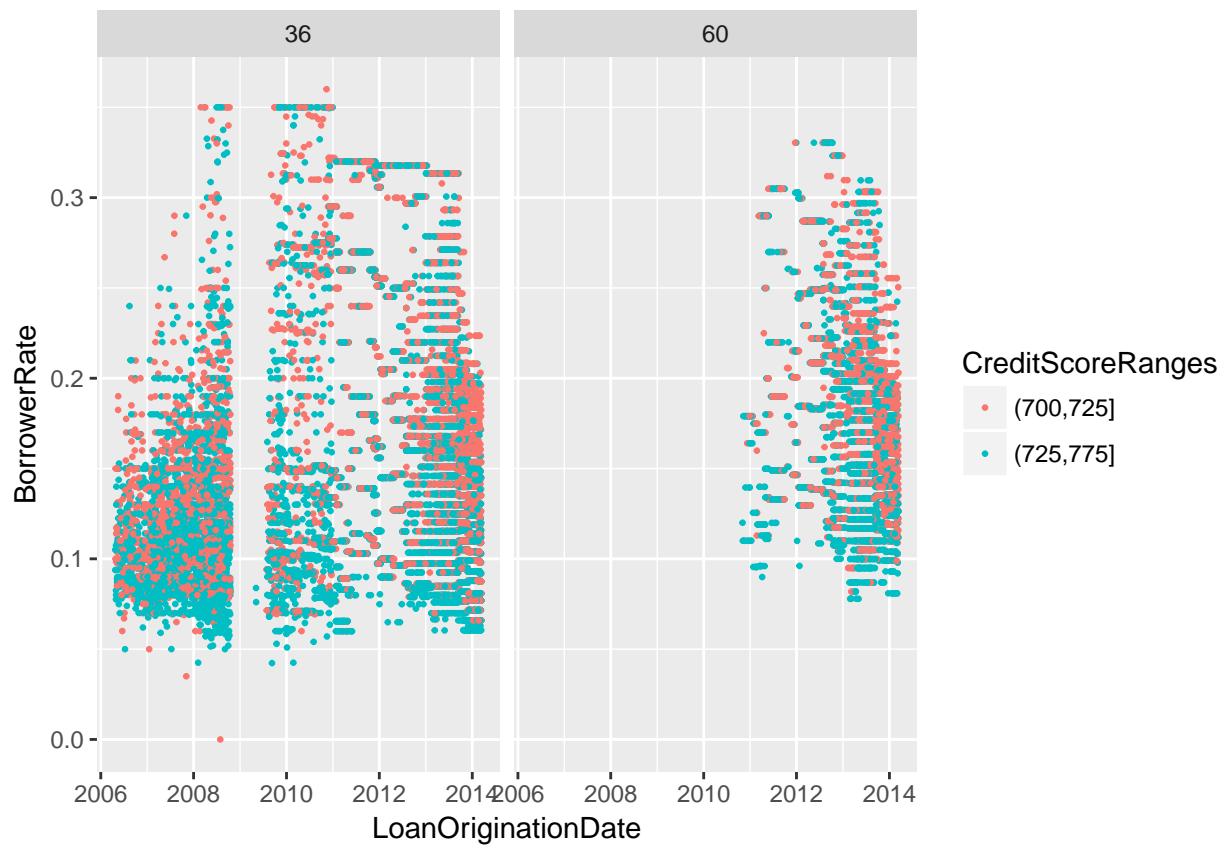
Number of Loans per day by Credit Score



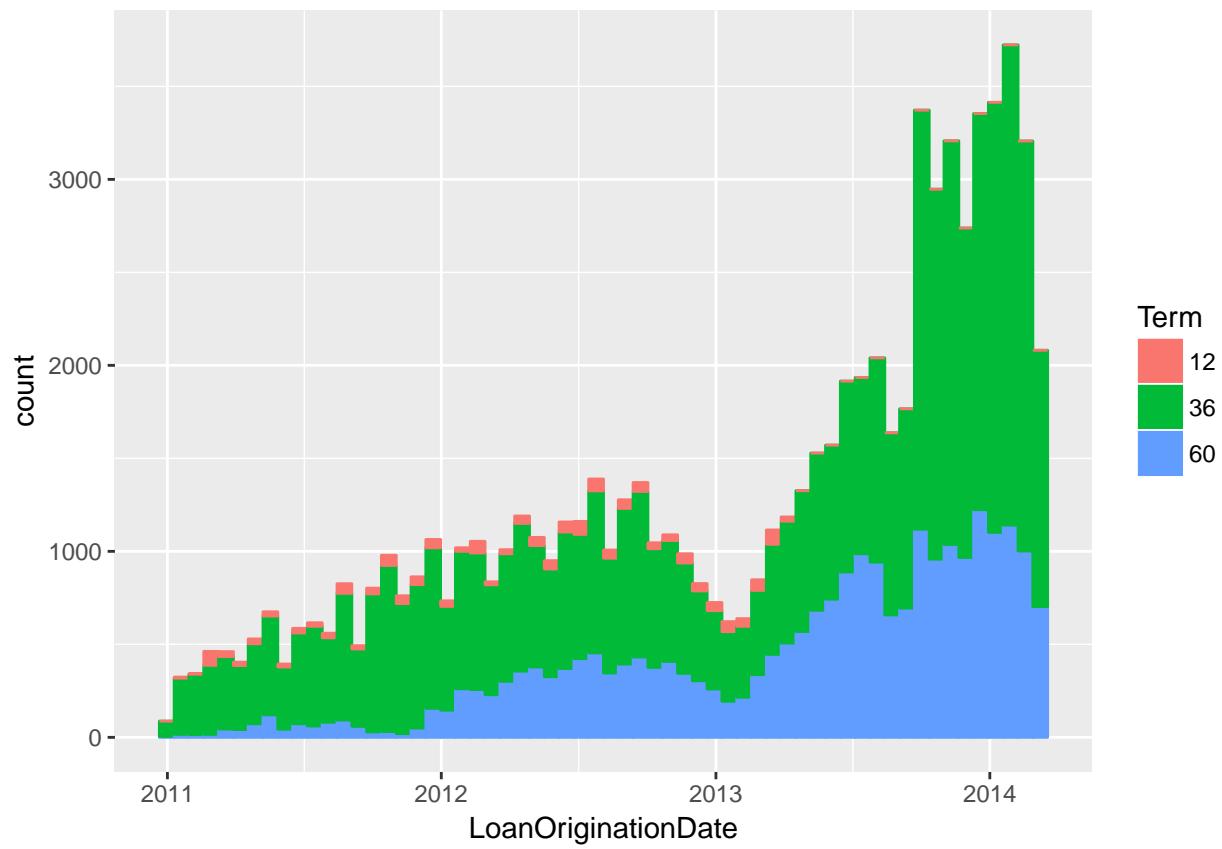
Separately we see a significant increase in loans to borrowers with 675 to 700 credit scores starting in 2011 and increasing at a significantly faster pace in 2013.



It appears that 12 and 60 month loans are new to Prosper starting around 2011.

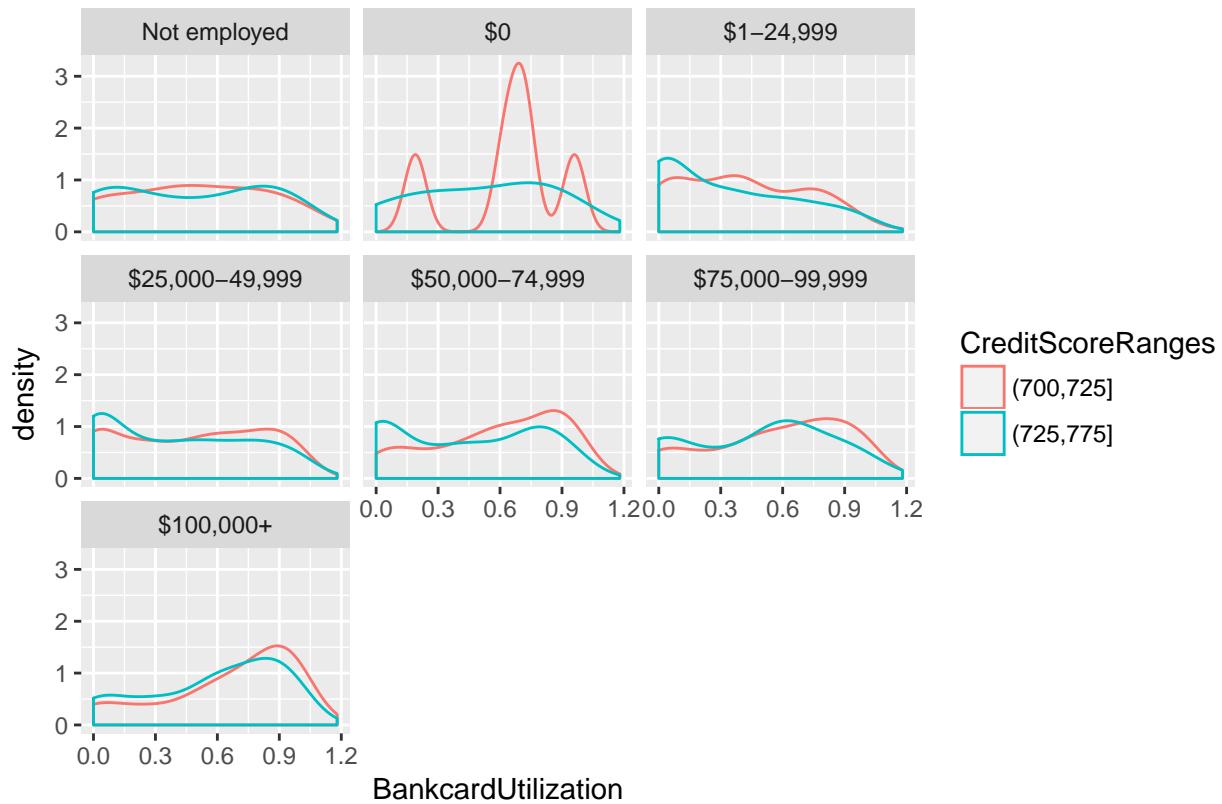


There has been an increase in 60 month term loans for borrowers with credit scores ranging from 700 to 775. However, the concentration around the 33% interest rate looks to be 36 month loans.



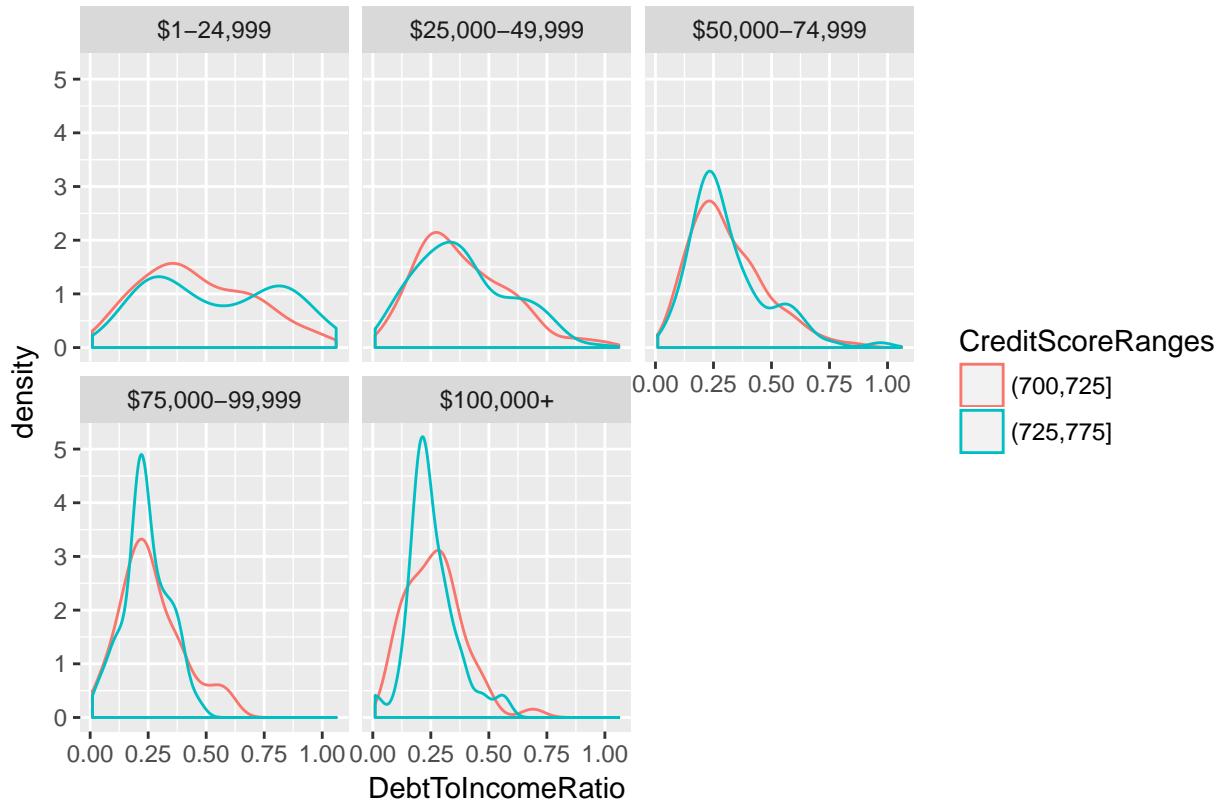
The 60 month loans do not make a noticeable increase until 2012.

Bankcard Utilization of Borrowers with Rates between 31% and 34%



For the selected range of credit scores, it does appear that bankcard utilization for higher income borrowers may be a contributing factor for the higher rates.

Debt-to-Income of Borrowers with Rates between 31% and 34%



Whereas higher debt-to-income may be more likely a contributing factor for lower income borrowers.

Multivariate Analysis

Observations

The median interest rates of the credit score ranges over the relevant period was insightful to provide a possible reason for why the data did not necessarily fit as well with other variables as I might have expected. Considering many of these borrowers are high risk, I would not be surprised if the increase in overall borrower rates during 2011 was from lenders being more conservative and expecting a higher return.

While bankcard utilization varied greatly, in general it was noticeably higher for loans that had a low credit score and high interest rate. Also, it looks to be a contributing factor to loans that are currently past due. This was inline with my expectations as these borrowers were most likely already struggling with repaying their debt.

As interest rate declined in 2013 and 2014, borrowers took advantage of the 60 month term option in increasing numbers.

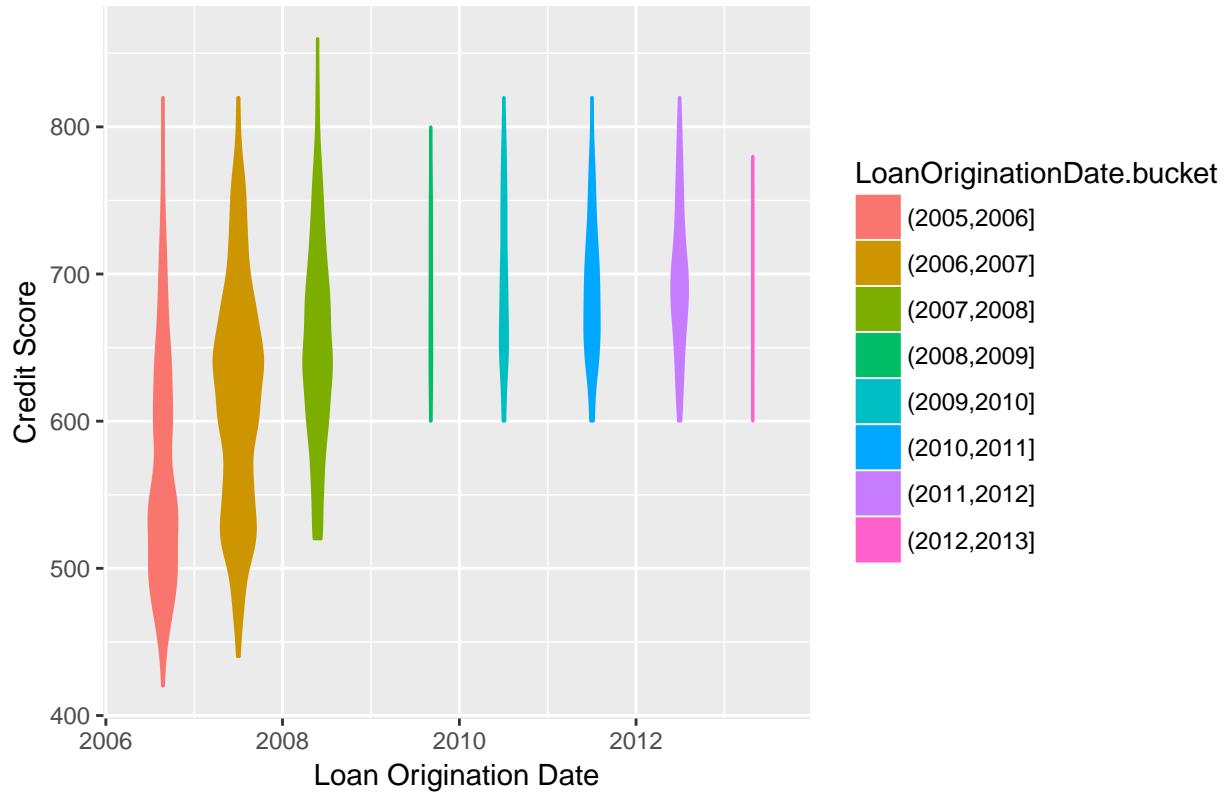
Interesting Interactions between Features

It was interesting to see that for the interest rates between 31% and 34% of higher credit score borrowers, bankcard utilization had a bigger factor for higher income borrowers. This was in contrast to debt-to-income that appeared to be more influential for lower income borrowers.

Final Plots and Summary

Plot One

Defaulted Loans by Origination Date and Credit Score



Description One

The difference in number of defaulted loans before and after the financial crisis was very interesting. This shows how our models and research can only provide insight on what we can expect based on historical examples. Also, we can see that these borrowers are of lower quality than the overall dataset with a median credit score of 640 and borrowing rate of 22.7%.

Credit Scores

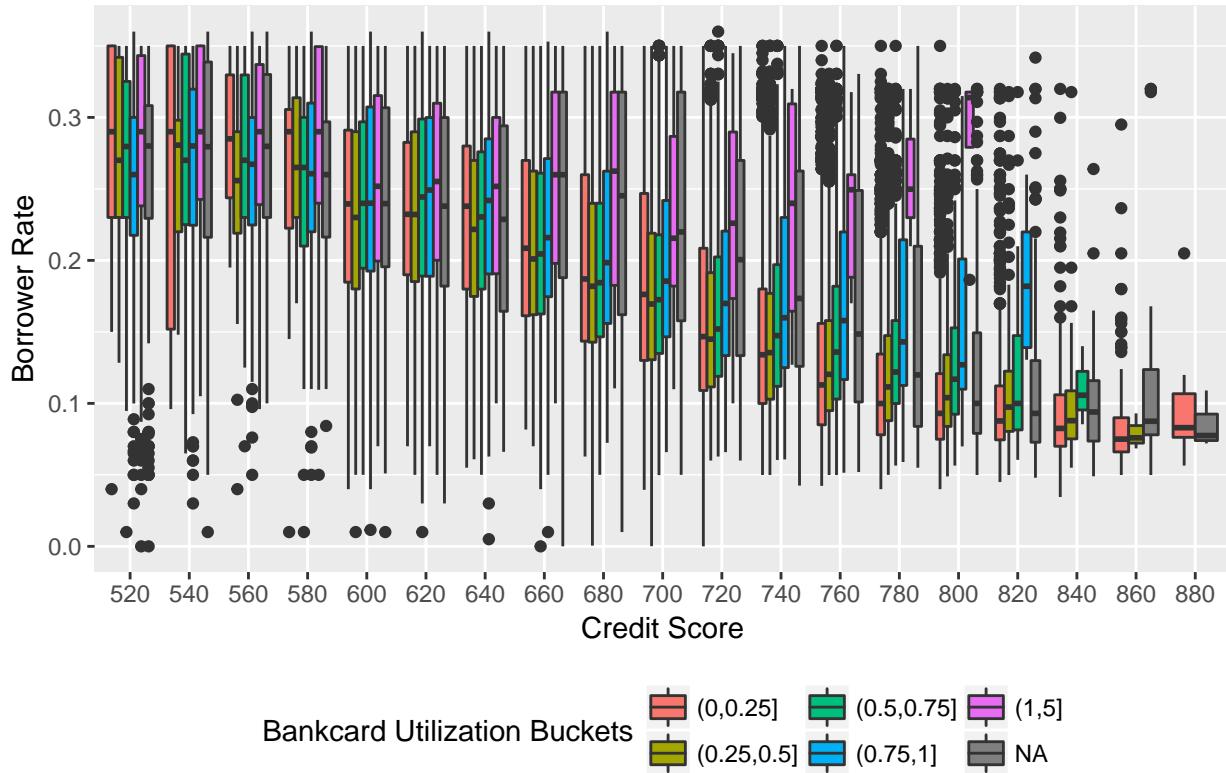
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
420.0	560.0	640.0	625.3	680.0	860.0

Borrower Rates

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00%	16.50%	22.70%	22.15%	28.50%	36.00%

Plot Two

Bankcard Utilization by Credit Score and Borrower Rate



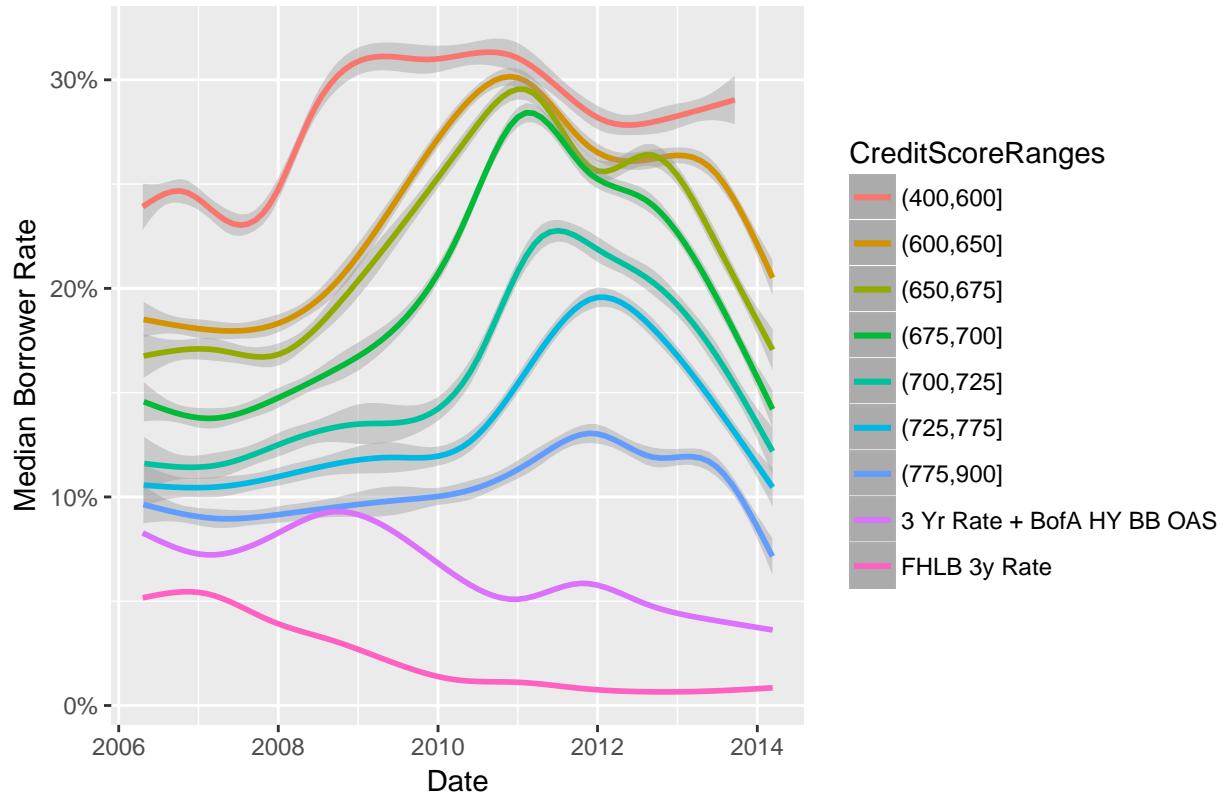
Description Two

The relationship between credit scores and borrower rates shows that there is a negative correlation (R^2 of -0.46). This makes sense as we would expect that the reason for the borrower's low credit score is that they would be more likely to not pay back their loan in full.

Credit Score Bucket	Median Bankcard Utilization
(400,600]	0.86
(600,650]	0.78
(650,675]	0.73
(675,700]	0.65
(700,725]	0.52
(725,775]	0.38
(775,900]	0.18

Plot Three

Median Rate by Credit Score over the Sample Period



Description Three

The median interest rates by credit score over time provide a couple of useful pieces of information. They show that lower credit borrowers are likely to have higher interest rates at any given point in time. Additionally, they show that during time periods of higher risk, these low credit borrowers may see larger differentials in pricing from other borrowers.

Credit Score Bucket	Median Rate
(400,600]	25.00%
(600,650]	23.20%
(650,675]	20.99%
(675,700]	18.25%
(700,725]	15.20%
(725,775]	13.50%
(775,900]	9.99%

Reflection

The Prosper dataset was interesting to explore as there was information about the borrower's historical experience with debt and also their current loan and its repayment performance. It was relatively easy to

work with the data as most of it was already in a workable format. However, there were a few variables that I found helpful to factor first to make them easier to work with.

It was interesting to bring in external interest rate data into the dataset. The Prosper data provided a lot of interesting information but there were a couple of times that I thought a stronger relationship should have existed but did not. Understanding the general interest rate market conditions helped to guide further exploration. This was most helpful with exploring the large spike in loan rates around 33%.

In the future, I think the analysis could be expanded to try to predict interest rates on new loans. I think there is enough information from the propser data that when combined with general market yield curves, it could come close to providing some prediction on where new loans would be given a set of borrower characteristics. However, Prosper's business model likely makes this a little more difficult in that at any point in time there could be an imbalance of borrowers and lenders since the business is not universally known as a typical bank would be.

Additional data sources

1. FHLB Boston Rates: <http://www.fhlbboston.com/rates/historicalrates/index.jsp>
2. Bank of America Merrill Lynch OAS: <https://fred.stlouisfed.org/series/BAMLH0A1HYBB>