prosperLoanData

Ramin Mohammadi

May 9, 2017

true

Abstract

I explore a data set about Prospect loans p2p. My goals with the study are understanding how this practice works and find what seems to affect the borrower rate.

Introduction

Dataset has 113937 records with 81 variables (including Date, characters, and numeric)

Summary of data set

```
## 'data.frame':    113937 obs. of  81 variables:
##  $ ListingKey                       : chr  "1021339766868145413AB3B" "1027360
2499503308B223C1" "0EE9337825851032864889A" "0EF535600248271529901A" ...
##  $ ListingNumber                    : int  193129 1209647 81716 658116 909464
1074836 750899 768193 1023355 1023355 ...
##  $ ListingCreationDate              : chr  "2007-08-26 19:09:29.263000000" "2
014-02-27 08:28:07.900000000" "2007-01-05 15:00:47.090000000" "2012-10-22 11:02:35.
010000000" ...
##  $ CreditGrade                      : chr  "C" NA "HR" NA ...
##  $ Term                             : int  36 36 36 36 36 60 36 36 36 36 ...
##  $ LoanStatus                       : chr  "Completed" "Current" "Completed"
"Current" ...
##  $ ClosedDate                       : chr  "2009-08-14 00:00:00" NA "2009-12-
17 00:00:00" NA ...
##  $ BorrowerAPR                      : num  0.165 0.12 0.283 0.125 0.246 ...
##  $ BorrowerRate                     : num  0.158 0.092 0.275 0.0974 0.2085 ..
.
##  $ LenderYield                      : num  0.138 0.082 0.24 0.0874 0.1985 ...
##  $ EstimatedEffectiveYield          : num  NA 0.0796 NA 0.0849 0.1832 ...
##  $ EstimatedLoss                    : num  NA 0.0249 NA 0.0249 0.0925 ...
##  $ EstimatedReturn                  : num  NA 0.0547 NA 0.06 0.0907 ...
##  $ ProsperRating..numeric.          : int  NA 6 NA 6 3 5 2 4 7 7 ...
##  $ ProsperRating..Alpha.            : chr  NA "A" NA "A" ...
##  $ ProsperScore                     : num  NA 7 NA 9 4 10 2 4 9 11 ...
##  $ ListingCategory..numeric.        : int  0 2 0 16 2 1 1 2 7 7 ...
##  $ BorrowerState                    : chr  "CO" "CO" "GA" "GA" ...
##  $ Occupation                       : chr  "Other" "Professional" "Other" "Sk
illed Labor" ...
##  $ EmploymentStatus                 : chr  "Self-employed" "Employed" "Not av
ailable" "Employed" ...
##  $ EmploymentStatusDuration         : int  2 44 NA 113 44 82 172 103 269 269 .
..
##  $ IsBorrowerHomeowner              : chr  "True" "False" "False" "True" ...
##  $ CurrentlyInGroup                 : chr  "True" "False" "True" "False" ...
##  $ GroupKey                         : chr  NA NA "783C3371218786870A73D20" NA
...
```

```
...
##  $ DateCreditPulled                   : chr  "2007-08-26 18:41:46.780000000" "2
014-02-27 08:28:14" "2007-01-02 14:09:10.060000000" "2012-10-22 11:02:32" ...
##  $ CreditScoreRangeLower              : int  640 680 480 800 680 740 680 700 820
820 ...
##  $ CreditScoreRangeUpper              : int  659 699 499 819 699 759 699 719 839
839 ...
##  $ FirstRecordedCreditLine            : chr  "2001-10-11 00:00:00" "1996-03-18
00:00:00" "2002-07-27 00:00:00" "1983-02-28 00:00:00" ...
##  $ CurrentCreditLines                 : int  5 14 NA 5 19 21 10 6 17 17 ...
##  $ OpenCreditLines                    : int  4 14 NA 5 19 17 7 6 16 16 ...
##  $ TotalCreditLinespast7years         : int  12 29 3 29 49 49 20 10 32 32 ...
##  $ OpenRevolvingAccounts              : int  1 13 0 7 6 13 6 5 12 12 ...
##  $ OpenRevolvingMonthlyPayment        : num  24 389 0 115 220 1410 214 101 219 2
19 ...
##  $ InquiriesLast6Months               : int  3 3 0 0 1 0 0 3 1 1 ...
##  $ TotalInquiries                     : num  3 5 1 1 9 2 0 16 6 6 ...
##  $ CurrentDelinquencies               : int  2 0 1 4 0 0 0 0 0 0 ...
##  $ AmountDelinquent                   : num  472 0 NA 10056 0 ...
##  $ DelinquenciesLast7Years            : int  4 0 0 14 0 0 0 0 0 0 ...
##  $ PublicRecordsLast10Years           : int  0 1 0 0 0 0 0 1 0 0 ...
##  $ PublicRecordsLast12Months          : int  0 0 NA 0 0 0 0 0 0 0 ...
##  $ RevolvingCreditBalance             : num  0 3989 NA 1444 6193 ...
##  $ BankcardUtilization                : num  0 0.21 NA 0.04 0.81 0.39 0.72 0.13
0.11 0.11 ...
##  $ AvailableBankcardCredit            : num  1500 10266 NA 30754 695 ...
##  $ TotalTrades                        : num  11 29 NA 26 39 47 16 10 29 29 ...
##  $ TradesNeverDelinquent..percentage. : num  0.81 1 NA 0.76 0.95 1 0.68 0.8 1 1
...
##  $ TradesOpenedLast6Months            : num  0 2 NA 0 2 0 0 0 1 1 ...
##  $ DebtToIncomeRatio                  : num  0.17 0.18 0.06 0.15 0.26 0.36 0.27
0.24 0.25 0.25 ...
##  $ IncomeRange                        : chr  "$25,000-49,999" "$50,000-74,999"
"Not displayed" "$25,000-49,999" ...
##  $ IncomeVerifiable                   : chr  "True" "True" "True" "True" ...
##  $ StatedMonthlyIncome                : num  3083 6125 2083 2875 9583 ...
##  $ LoanKey                            : chr  "E33A3400205839220442E84" "9E3B370
71505919926B1D82" "6954337960046817851BCB2" "A0393664465886295619C51" ...
##  $ TotalProsperLoans                  : int  NA NA NA NA 1 NA NA NA NA NA ...
##  $ TotalProsperPaymentsBilled         : int  NA NA NA NA 11 NA NA NA NA NA ...
##  $ OnTimeProsperPayments              : int  NA NA NA NA 11 NA NA NA NA NA ...
##  $ ProsperPaymentsLessThanOneMonthLate: int  NA NA NA NA 0 NA NA NA NA NA ...
##  $ ProsperPaymentsOneMonthPlusLate    : int  NA NA NA NA 0 NA NA NA NA NA ...
##  $ ProsperPrincipalBorrowed           : num  NA NA NA NA 11000 NA NA NA NA NA ..
.
##  $ ProsperPrincipalOutstanding        : num  NA NA NA NA 9948 ...
##  $ ScorexChangeAtTimeOfListing        : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ LoanCurrentDaysDelinquent          : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ LoanFirstDefaultedCycleNumber      : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ LoanMonthsSinceOrigination         : int  78 0 86 16 6 3 11 10 3 3 ...
##  $ LoanNumber                         : int  19141 134815 6466 77296 102670 123
257 88353 90051 121268 121268 ...
##  $ LoanOriginalAmount                 : int  9425 10000 3001 10000 15000 15000
3000 10000 10000 10000 ...
##  $ LoanOriginationDate                : chr  "2007-09-12 00:00:00" "2014-03-03
00:00:00" "2007-01-17 00:00:00" "2012-11-01 00:00:00" ...
##  $ LoanOriginationQuarter             : chr  "Q3 2007" "Q1 2014" "Q1 2007" "Q4
2012"
```

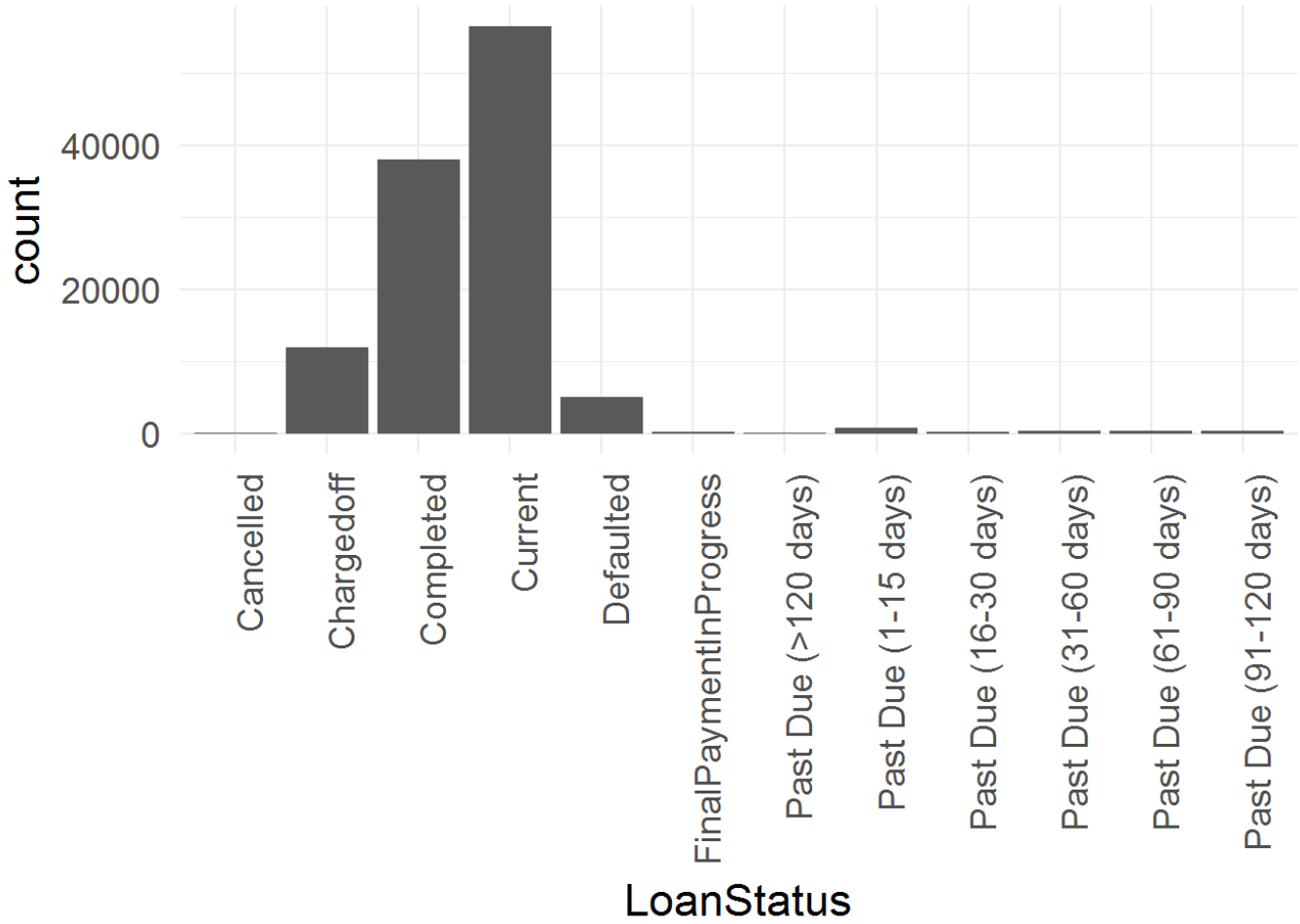
```
2012" ...
##  $ MemberKey                          : chr  "1F3E3376408759268057EDA" "1D13370
546739025387B2F4" "5F7033715035555618FA612" "9ADE356069835475068C6D2" ...
##  $ MonthlyLoanPayment                 : num  330 319 123 321 564 ...
##  $ LP_CustomerPayments                : num  11396 0 4187 5143 2820 ...
##  $ LP_CustomerPrincipalPayments       : num  9425 0 3001 4091 1563 ...
##  $ LP_InterestandFees                 : num  1971 0 1186 1052 1257 ...
##  $ LP_ServiceFees                     : num  -133.2 0 -24.2 -108 -60.3 ...
##  $ LP_CollectionFees                  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ LP_GrossPrincipalLoss              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ LP_NetPrincipalLoss                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ LP_NonPrincipalRecoverypayments    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ PercentFunded                      : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Recommendations                    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ InvestmentFromFriendsCount         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ InvestmentFromFriendsAmount        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Investors                          : int  258 1 41 158 20 1 1 1 1 1 ...
```

variables selection

Following variable have been dropped as in order to make the analysis easier to undrestand

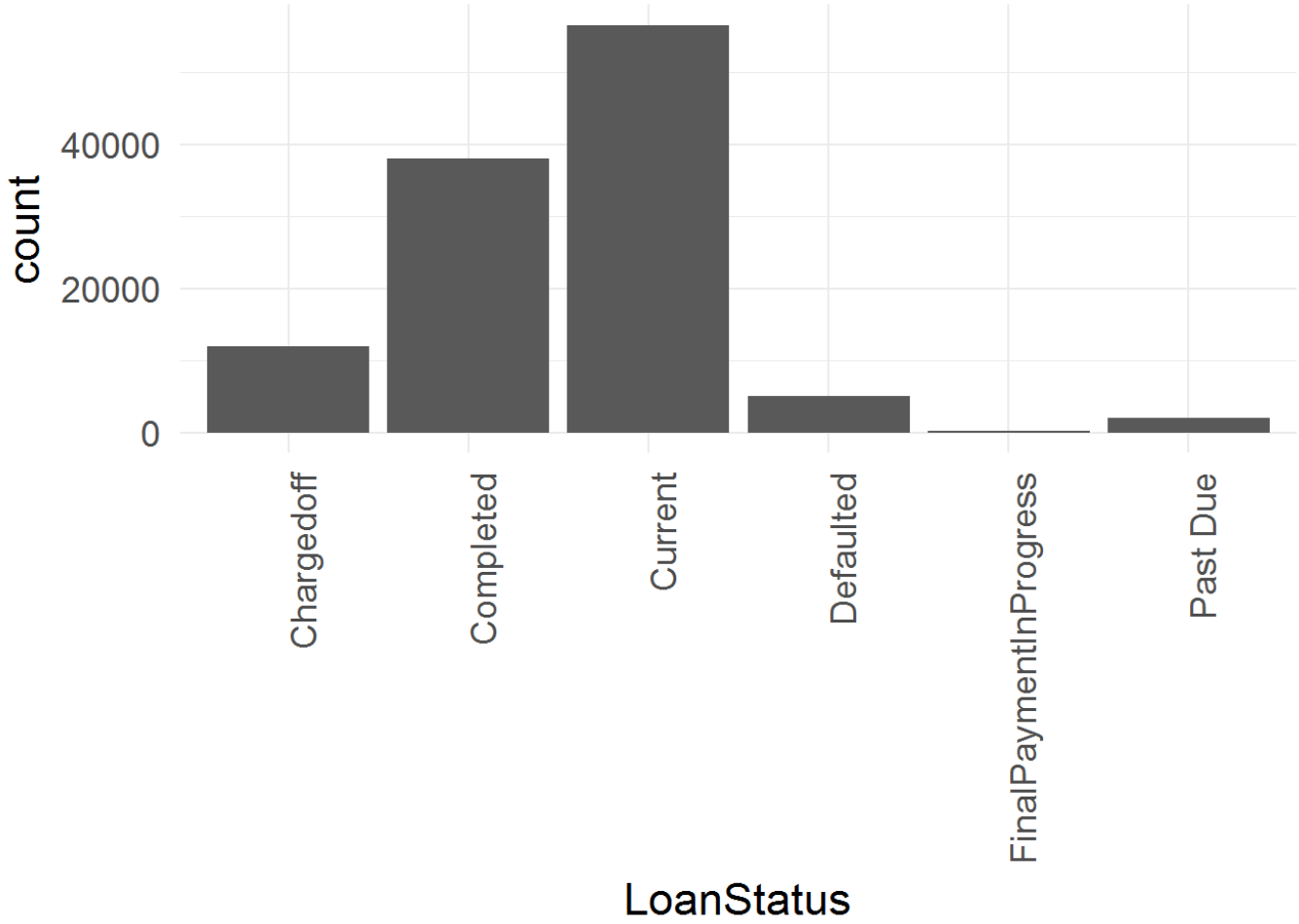
looking throug dataset and droping column which have more than 2000 records empty or NA

Univariate ploting



As we can see we are having mutiple ‘past due’ based on different day values , for simplicity and making our plot to be easier to undrestand i convert all values whihc start with “Past” to “Past Due”

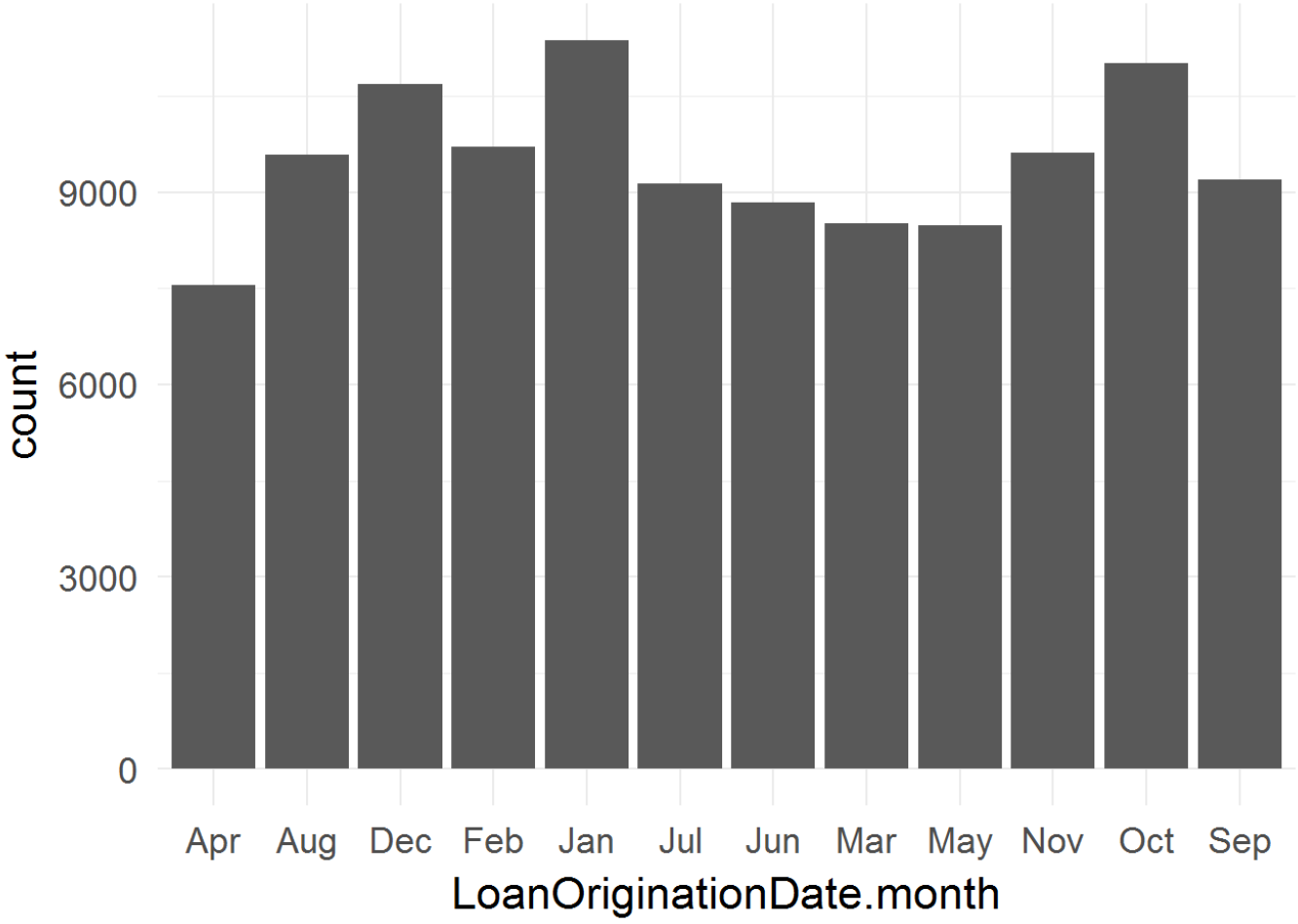
Also we can see Cancelled status is only happening 5 times out of 113932 records, so i remove those records

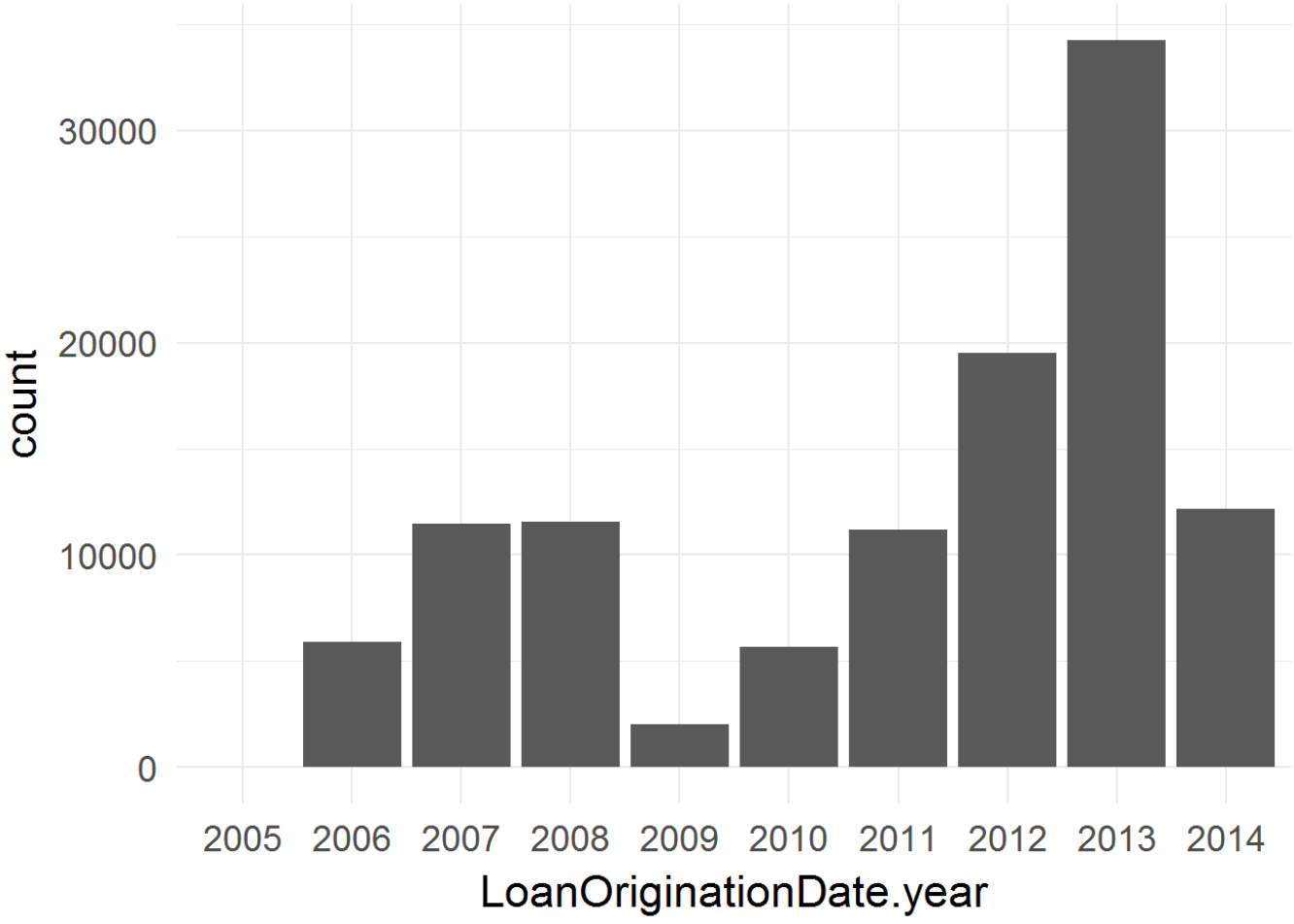


majority of loans in our datase are in current status then second majority is completed however we can see we have final payment in progress which are not much so i am removing the cases which are final payments

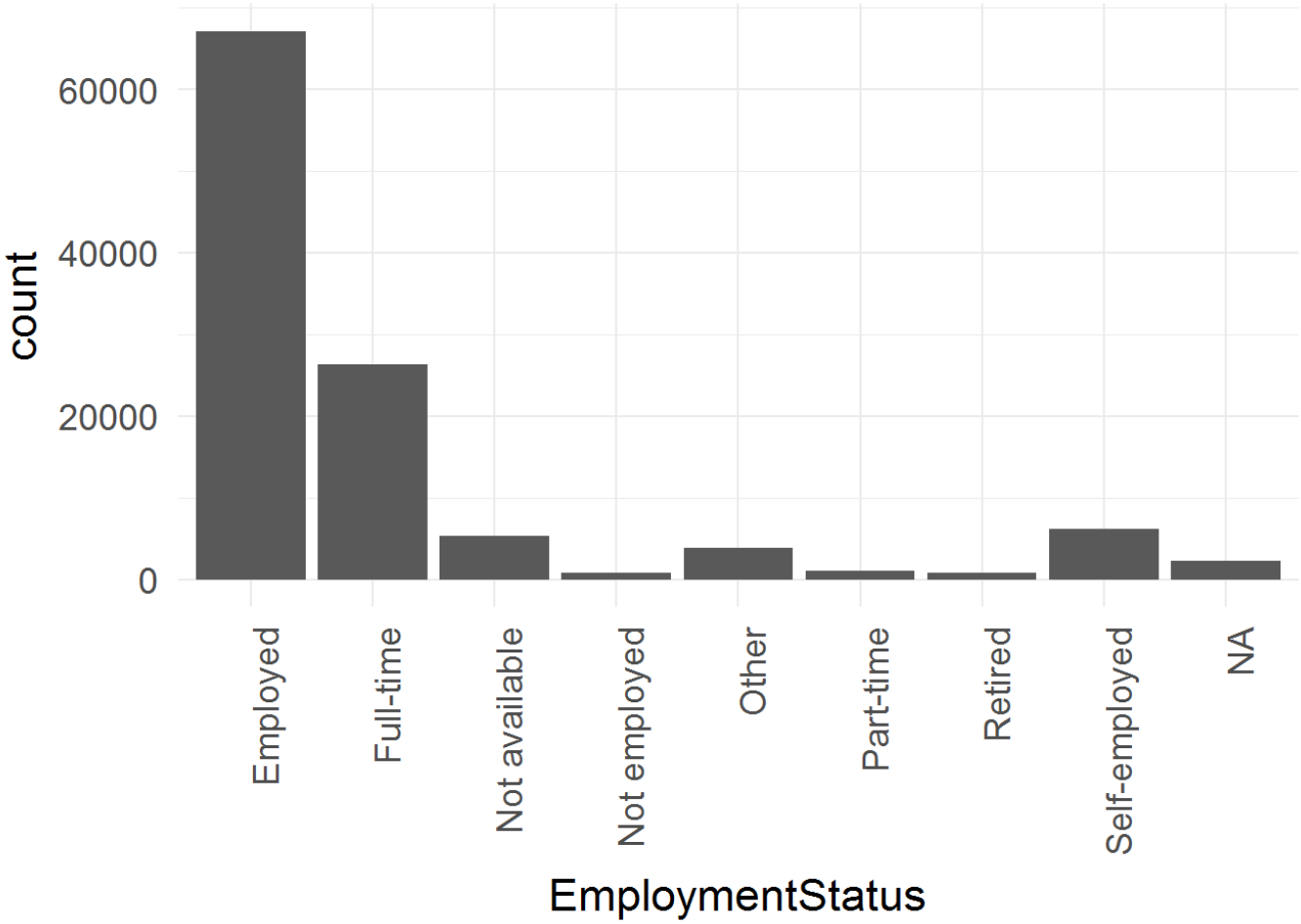
```
##
## Chargedoff  Completed    Current  Defaulted   Past Due
##      11992      38074      56576       5018       2067
```

We can see our plot is more undrestandable

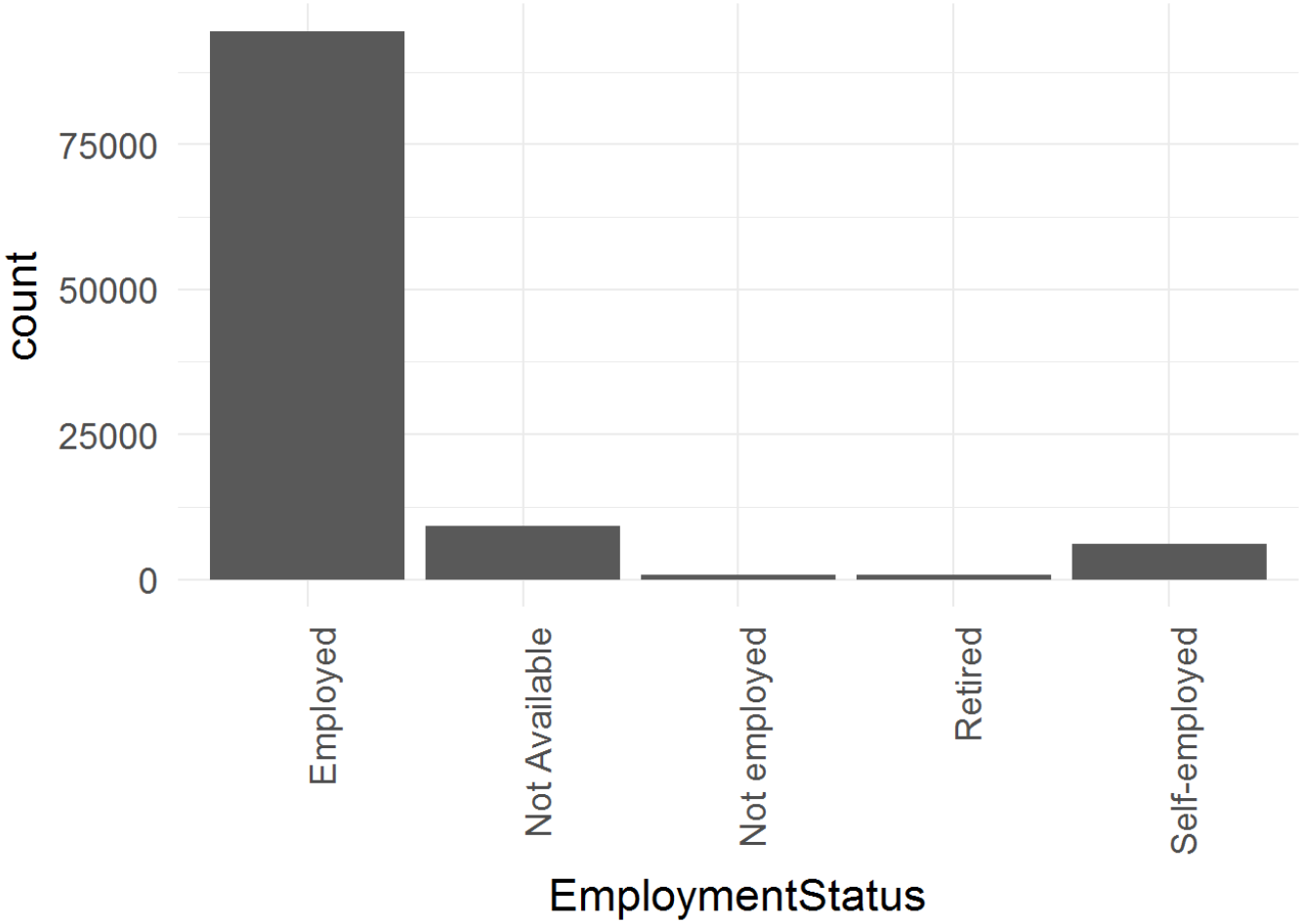




Most of the loans happend on 2013 and also highest amount of loan request are first month of the year and october , we can check for reason in multivariate analysis to see what is the reason behind



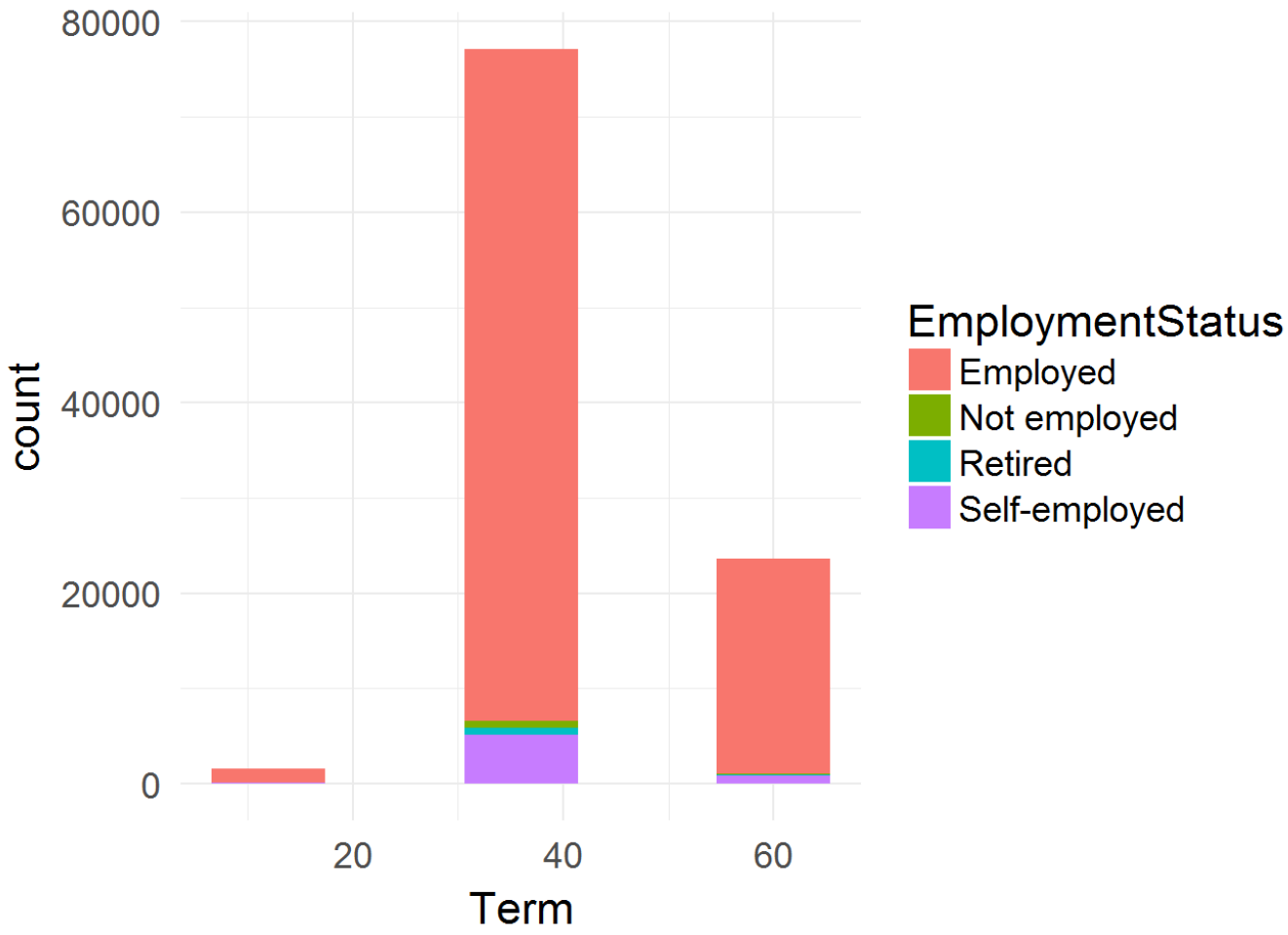
mixing “not available”, “other” and “NA” values for Employment Status and change them to “Not Available”, Also we can see we are having a value as Employed which can contain (Fulltime or PartTime) so for simplicity i mix them together



Majority of borrowers are Employed, however we have 835 case which dont have any information regarding their employment status and assumin that we are dealing with loan information means the information is missing so, i remove cases without employment status information

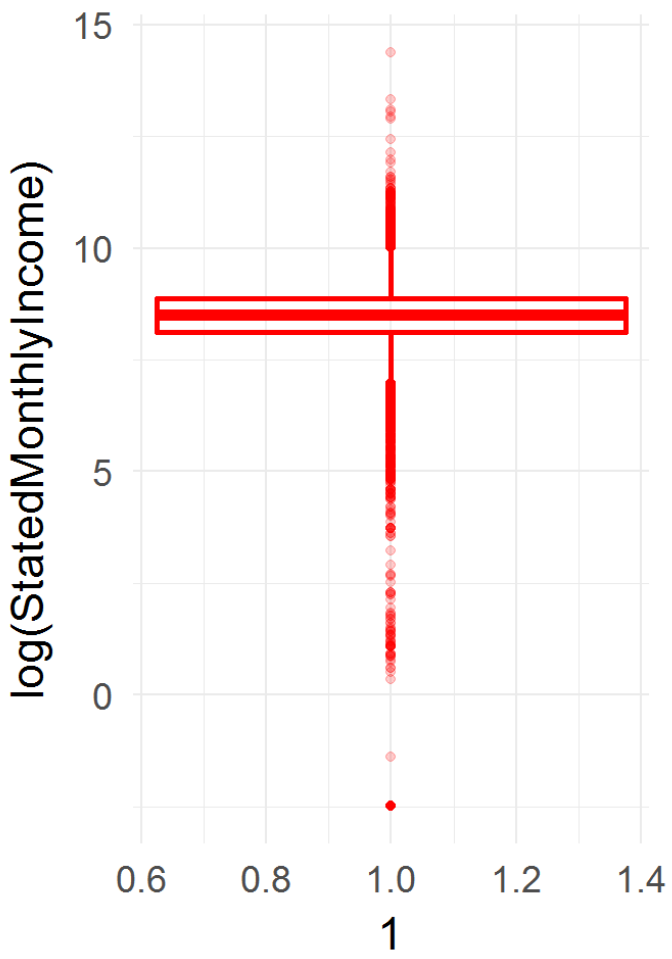
```
##
##      Employed  Not employed       Retired Self-employed
##         94577           835           795          6124
```

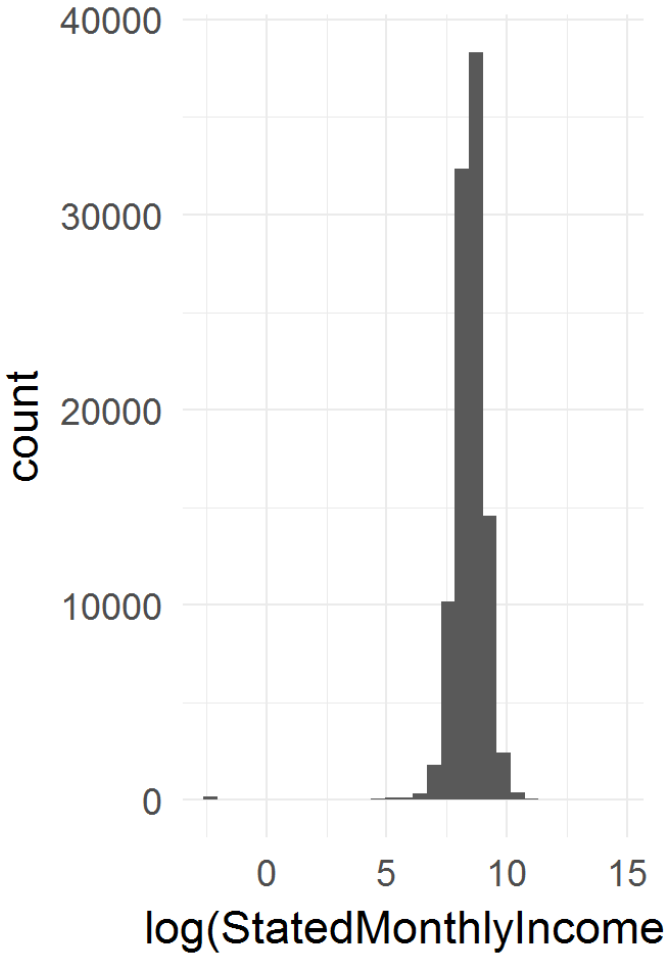
We can see we have more Employed than other values which is expecting, however Not available is not quite clear as for getting loan you need to provide that information




```
##
##    12    36    60
##  1555 77124 23652
```

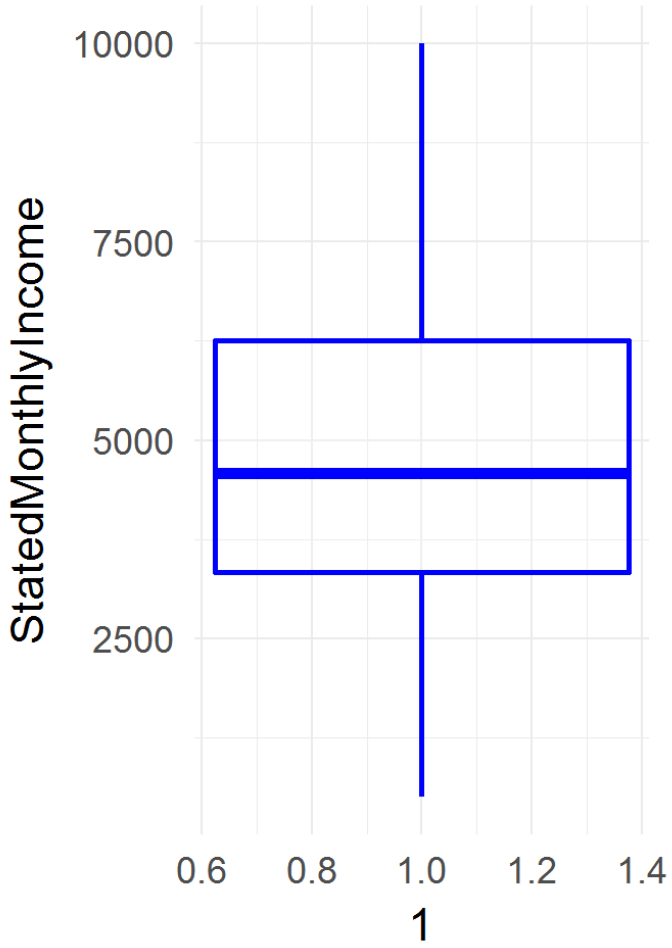
We can see most of the Loans having 36 months terms which also expecting as having higher terms will require paying more intrest and having lower term means monthly payment should be more , also we can see retired class are having 36 months loans mostly

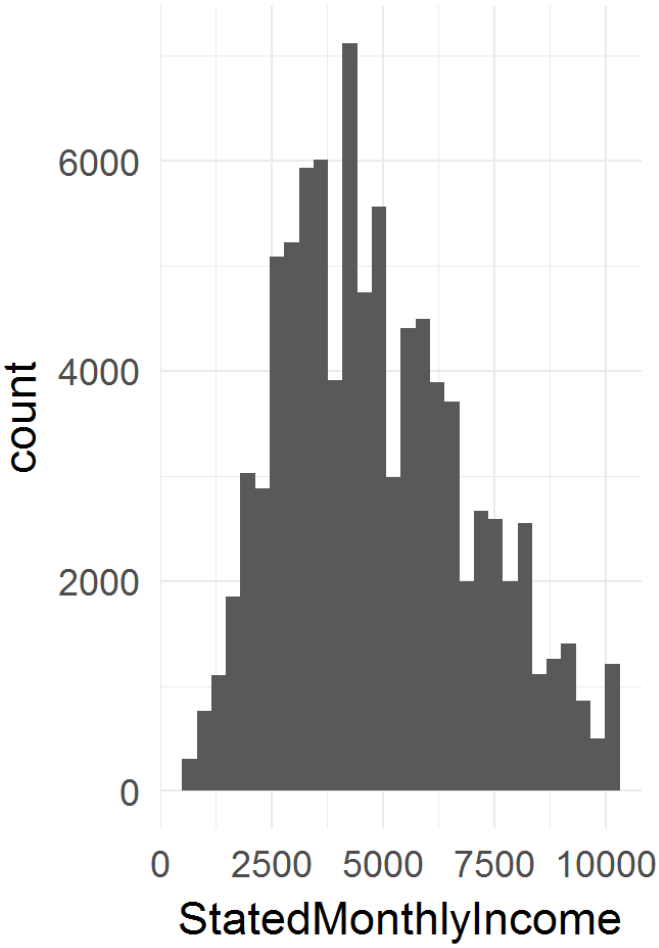




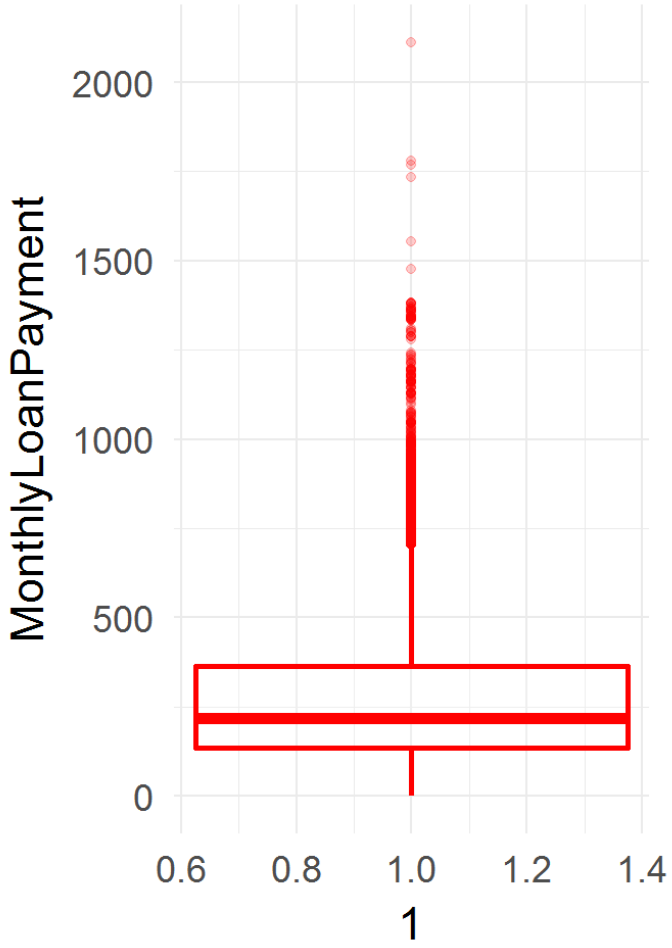

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    3333    4833    5747    7000 1750000
```

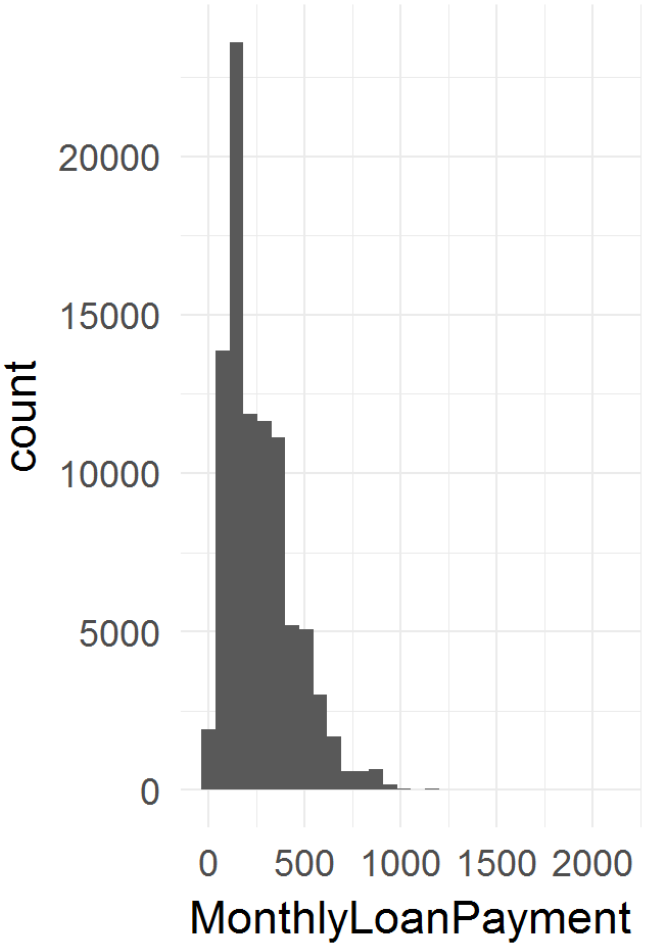
at this point there are 1393 case which have income of 0, my assumption is either there has been a mistake or input is wrong as bank will not provide you a loan with income of 0 , so for my work i get ride of records with income of 0 , ALSO i am intrested in monthly incomes less than equal to 10000 as more than 10000 is not likly to happen





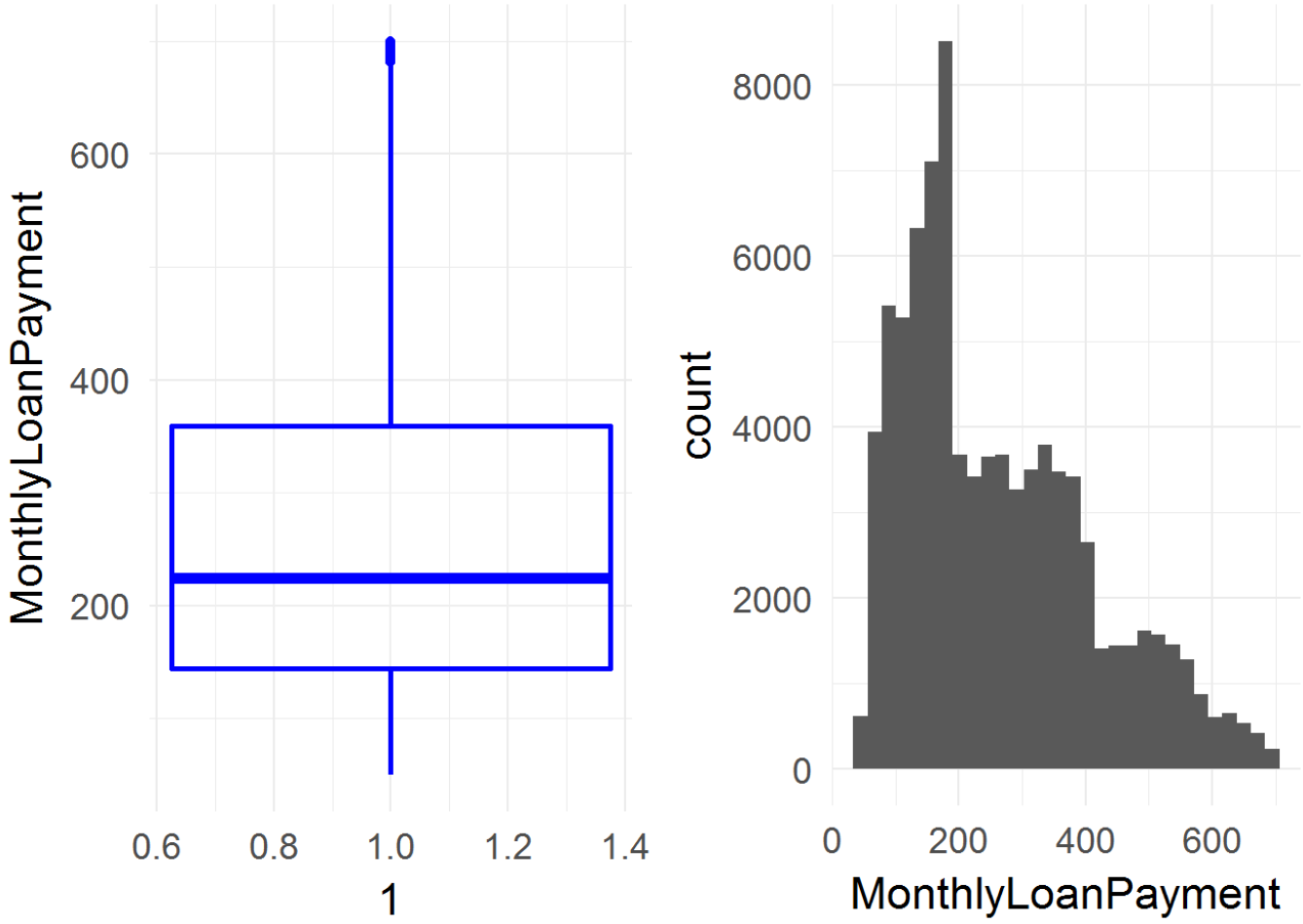
we can see median is 4538 and we have majority of data within first quartile and third quartile (3317,6250)



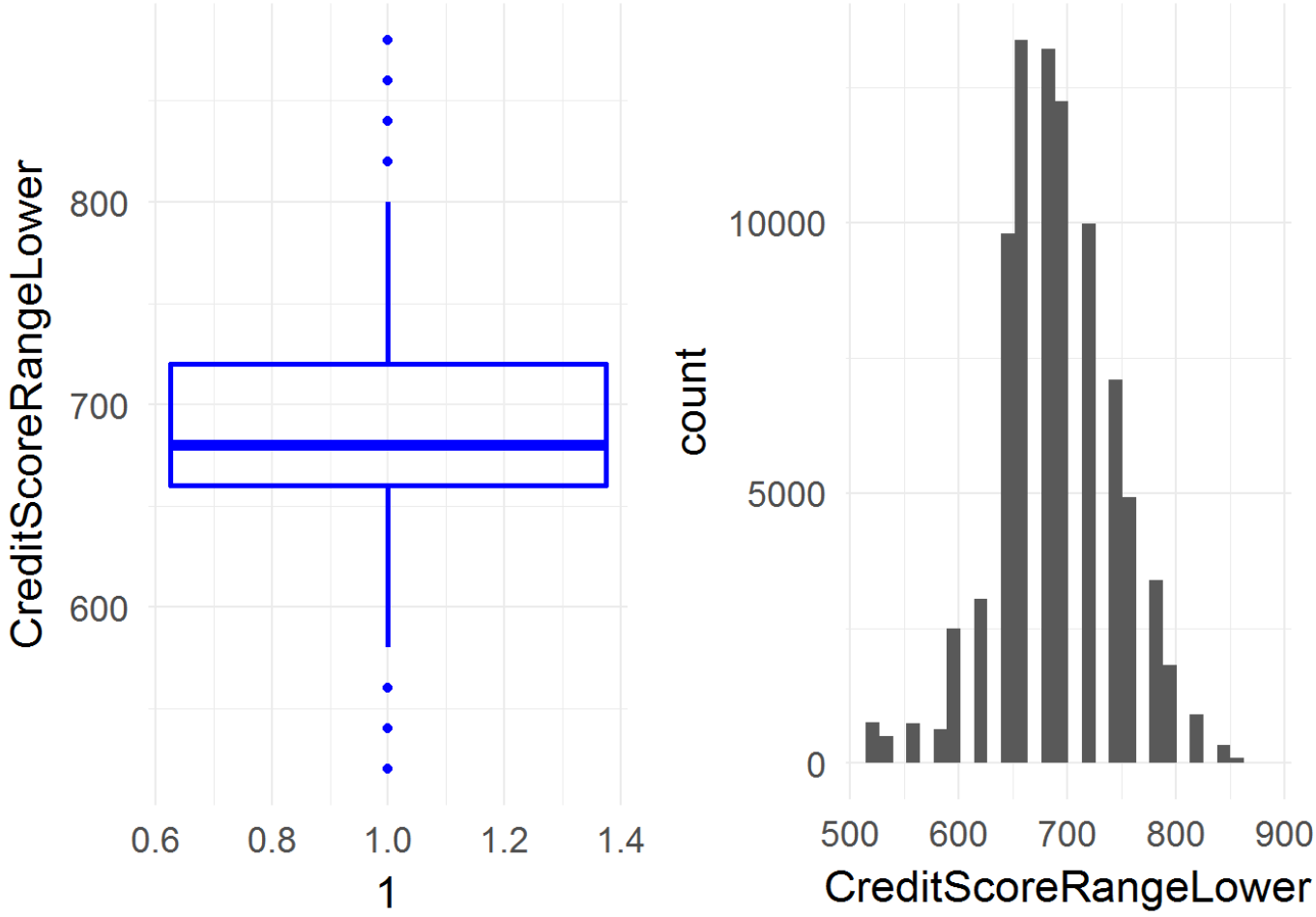


```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   135.7   219.0   265.5   363.1  2112.0
```

looking at summary table we can see monthly loan payment of 0 which indicates of wrong data, as we can not have monthly payment of 0, i will remove records with value of 0 , also we only have 473 case with monthly payment of greater than 700$, i removed those case as well as they are outlier to me.

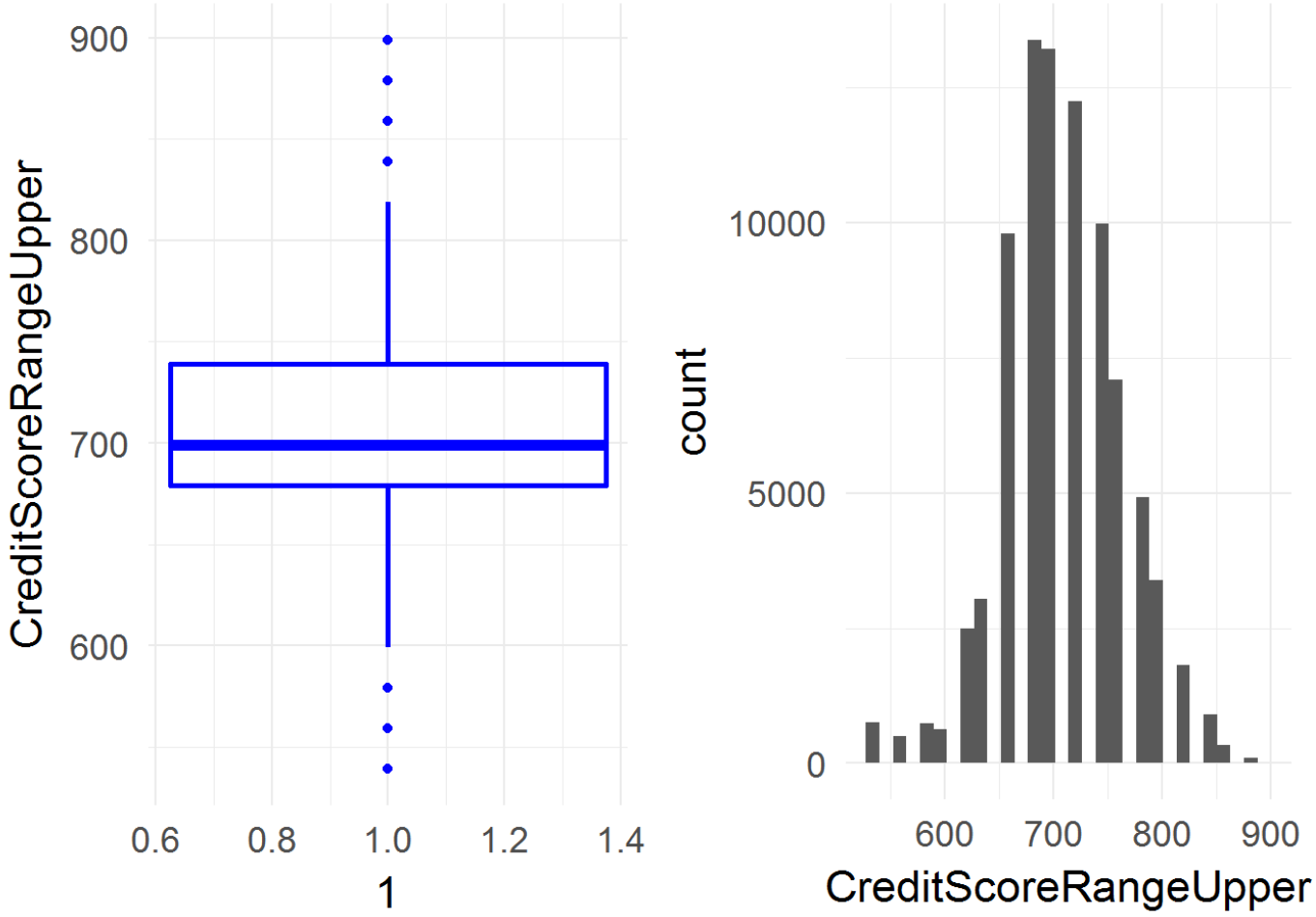


monthly loan payment has distribution which is right skewed



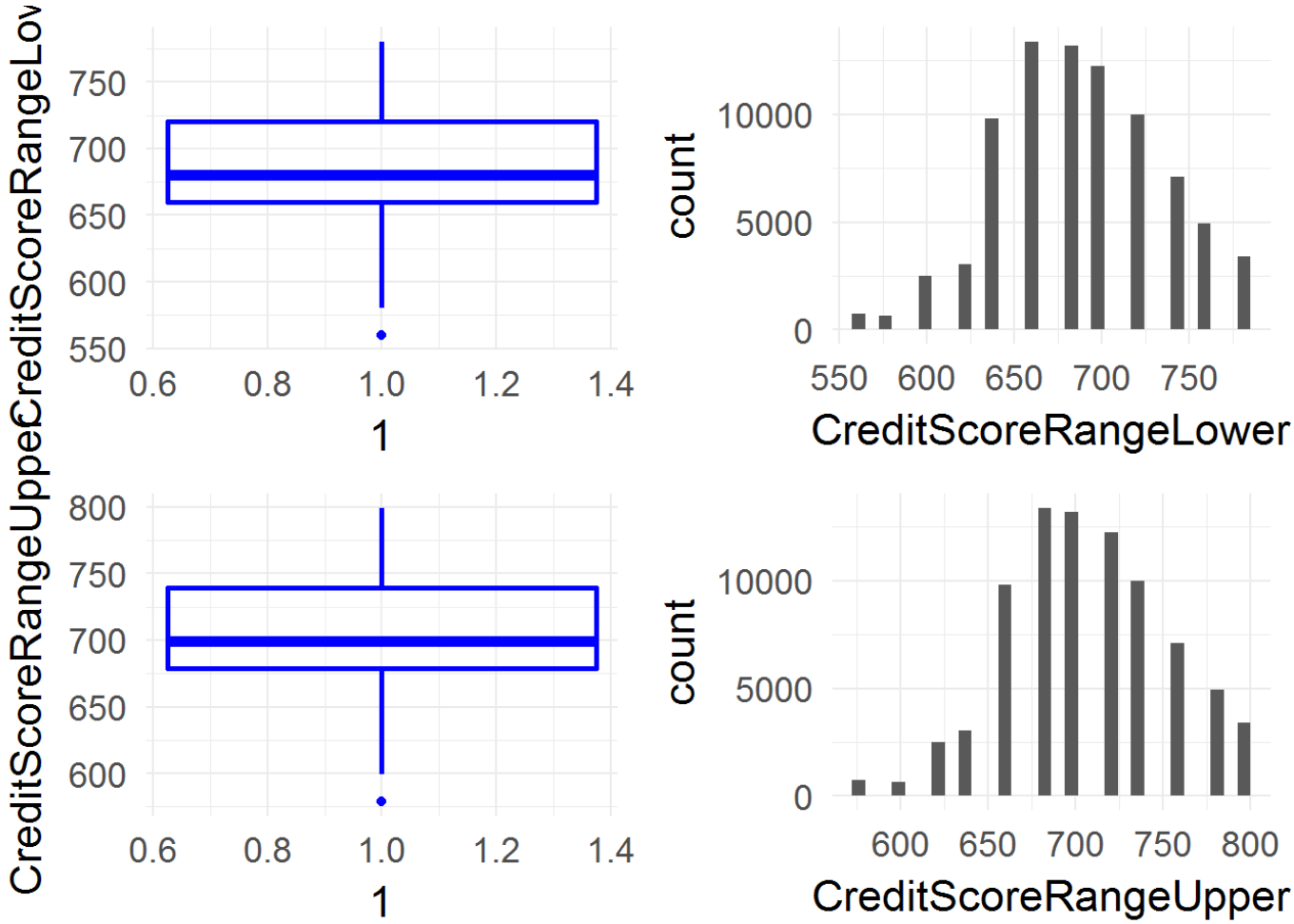

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   520.0   660.0   680.0   689.8   720.0   880.0
```

lower credite score range has ouliers based on box-plot , i consider to eliminate records with lower credite range less than 550 and higher than 800

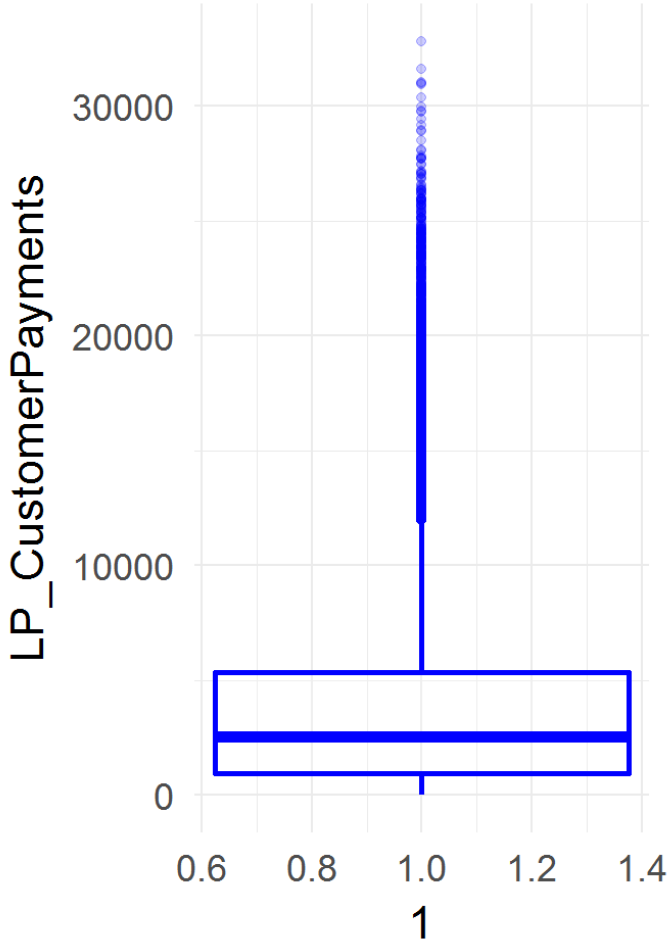


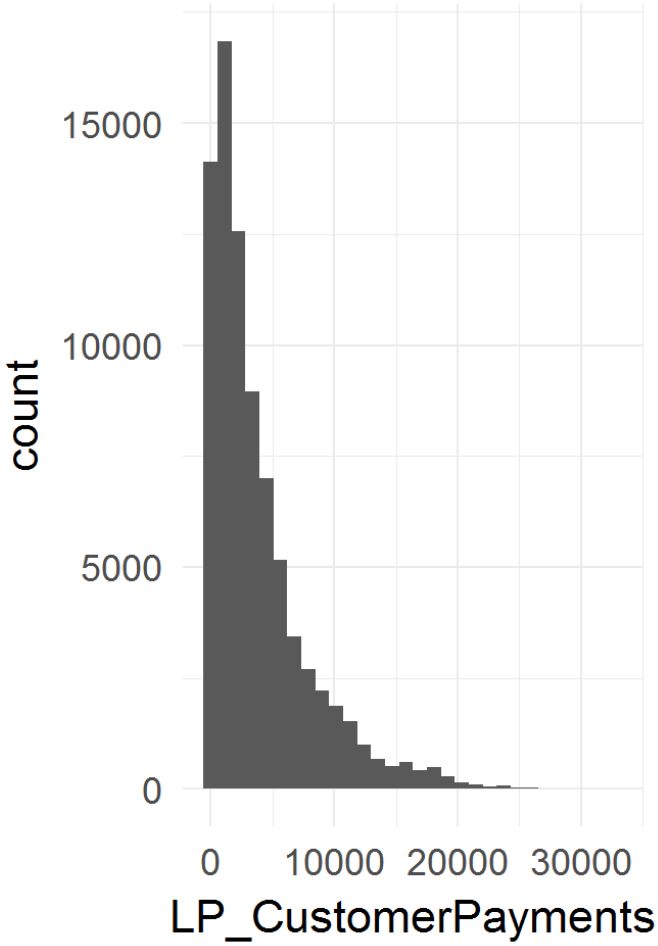

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   520.0   660.0   680.0   689.8   720.0   880.0
```

i have consider same condition for Upper range credit, so i rmoved credite range less than 550 and higher than 800



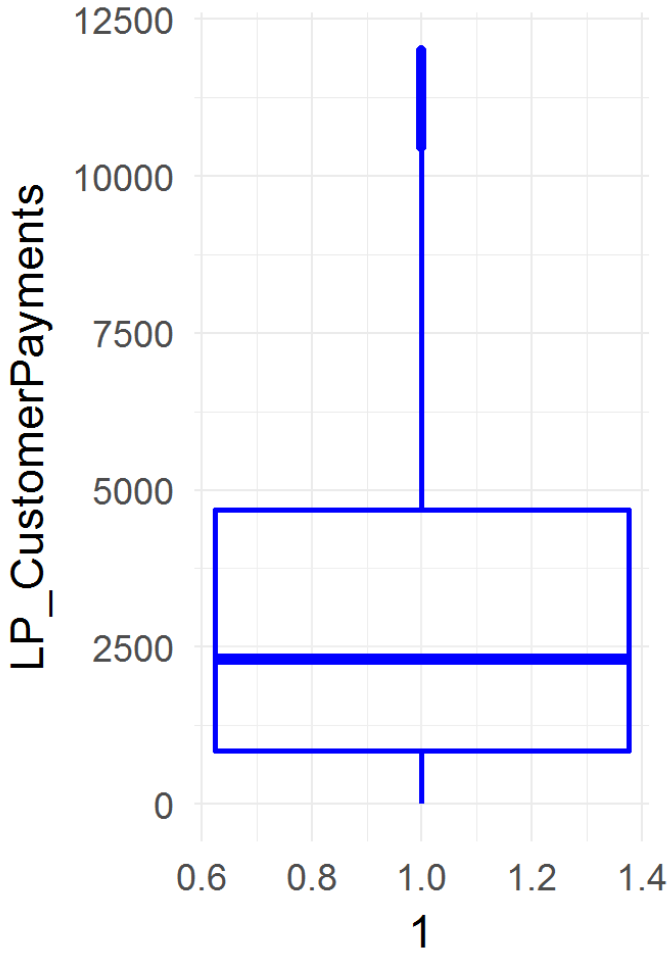
we can see both Upper and Lowe rcredit range having similar distribution however, they are having slightly different quartiles. later i am creating a new variable as average credit score and remove Upper and Lower scores.

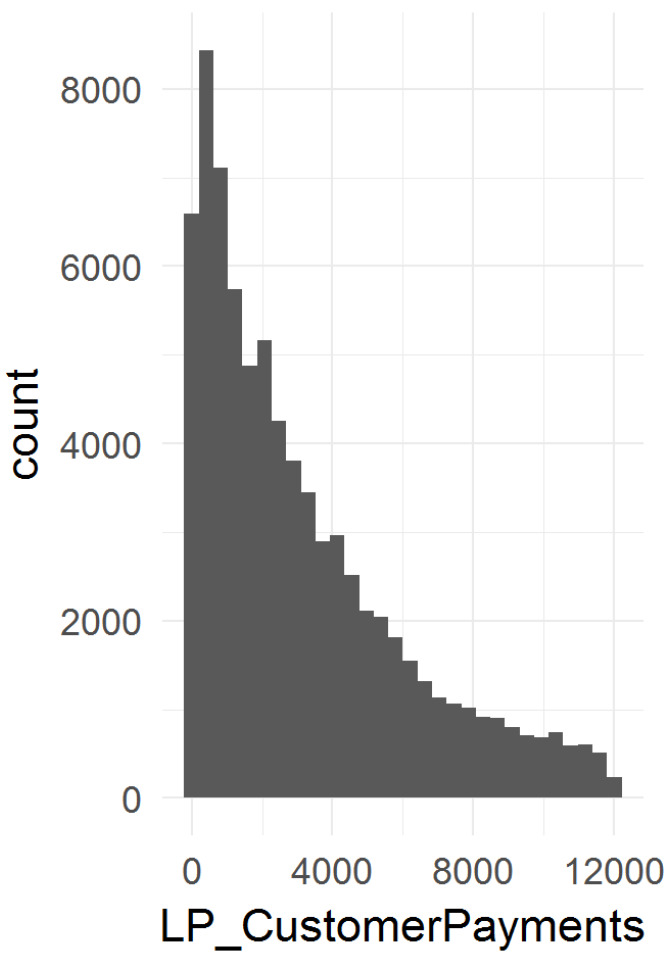





```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##     -2.35    913.50   2524.00   3868.00   5323.00  32800.00
```

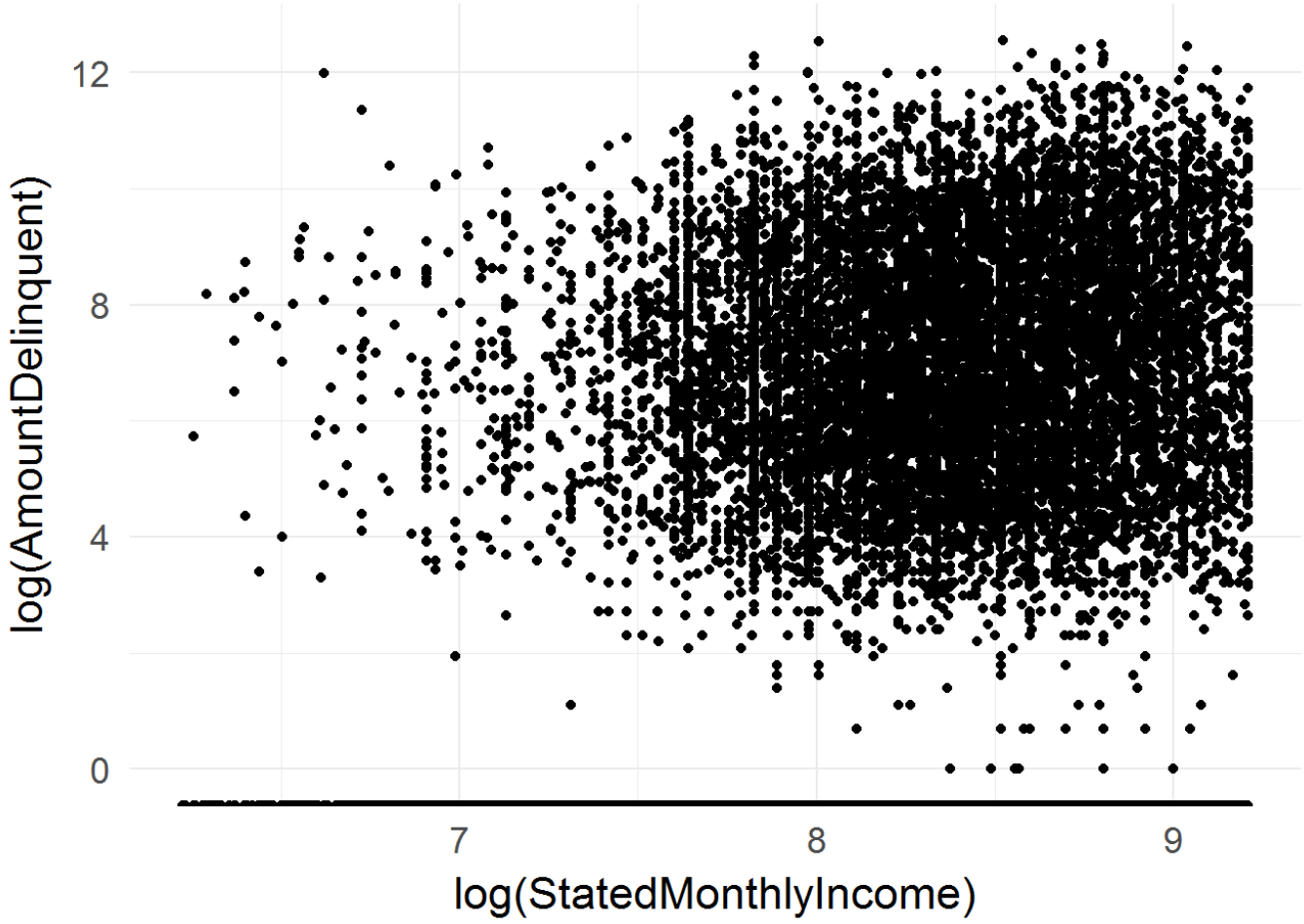
we can see here that on minimum value we are having negative value which indicate of either wrong input or wrong value, i will remove all records with values less than 0 and higher than 12000.



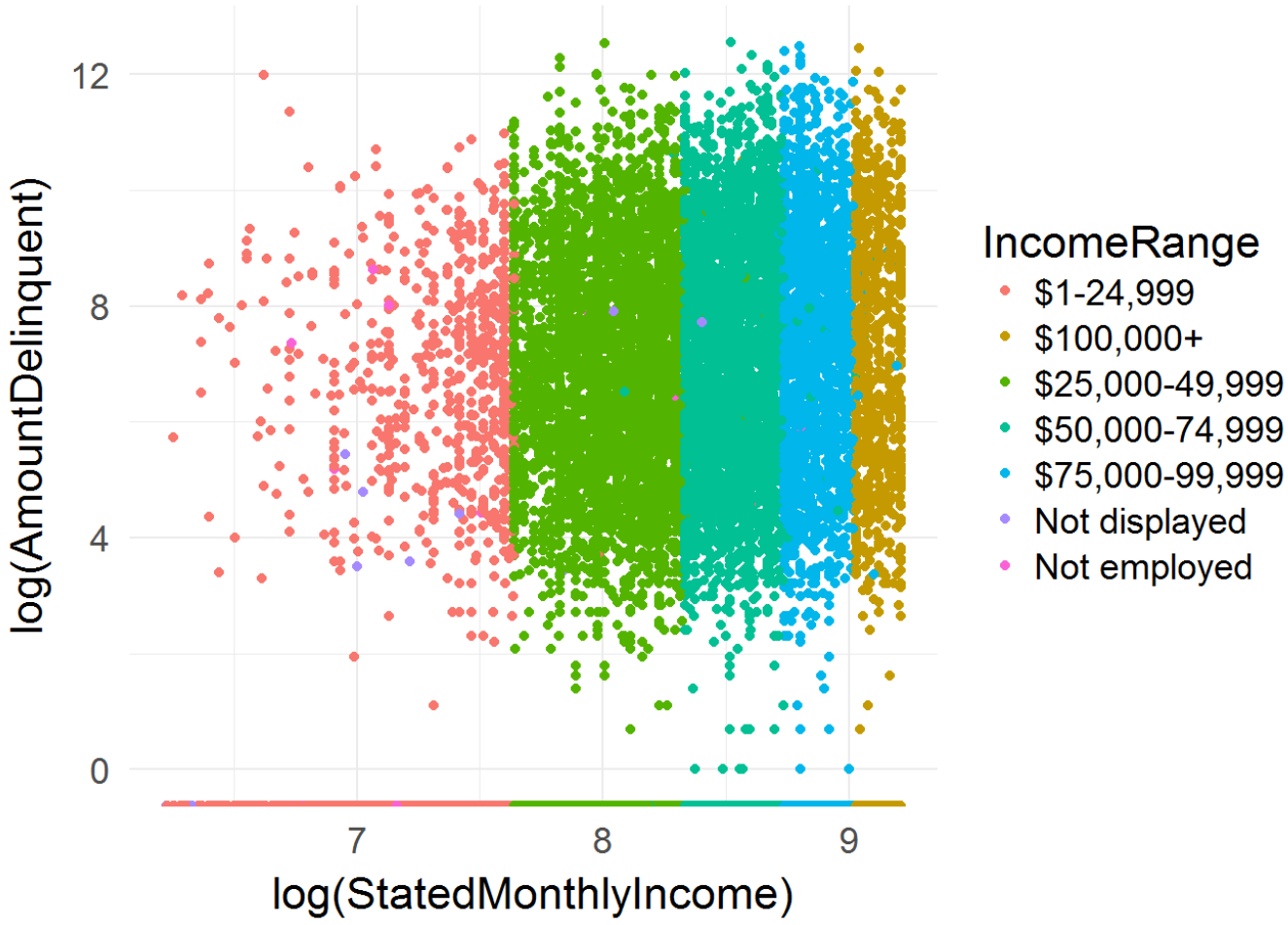


distribution is right skewed with median as 2338 , and we can see majority have payment less than 4702 which is 3rd quartil.

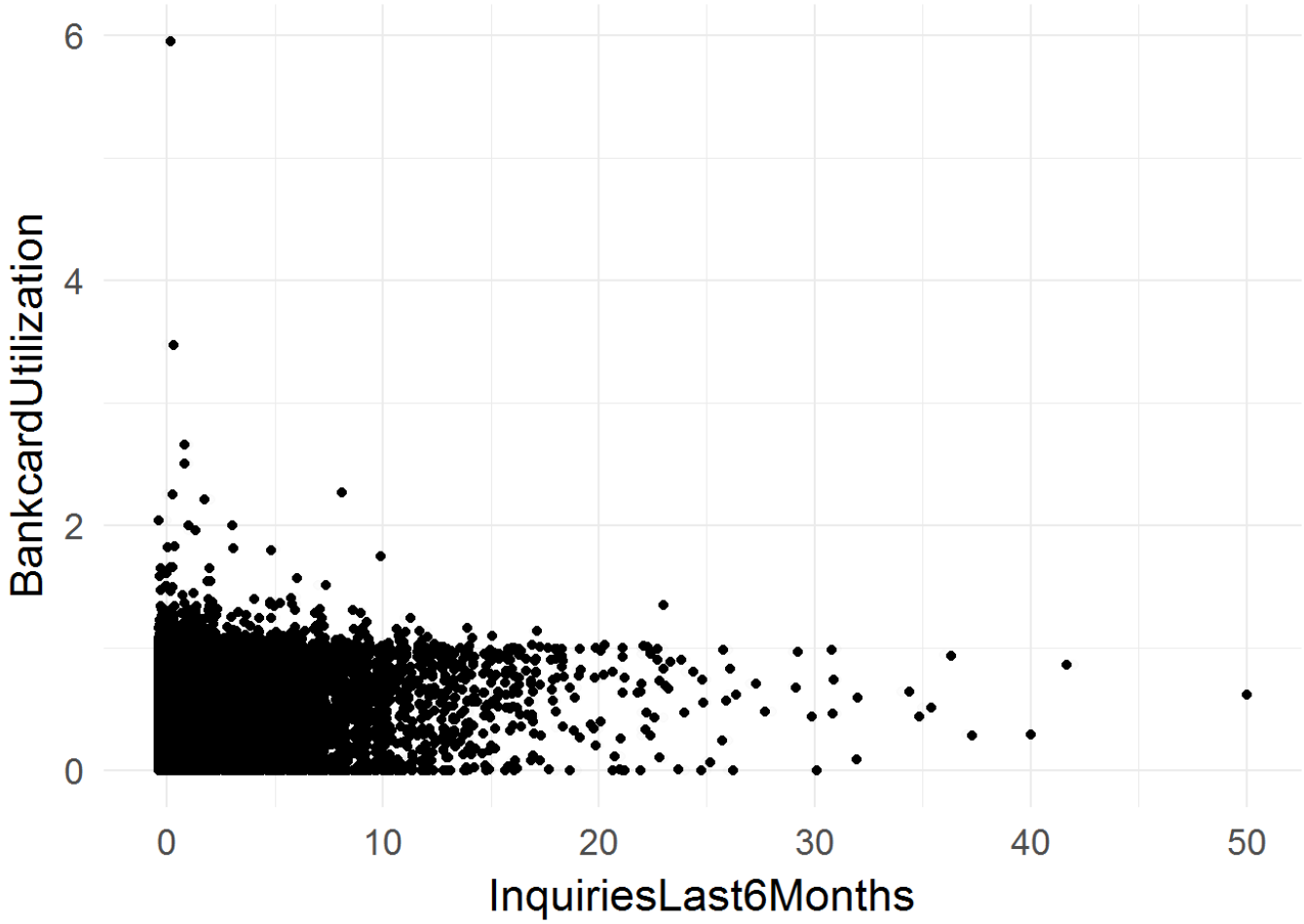
Bivariate Analysis



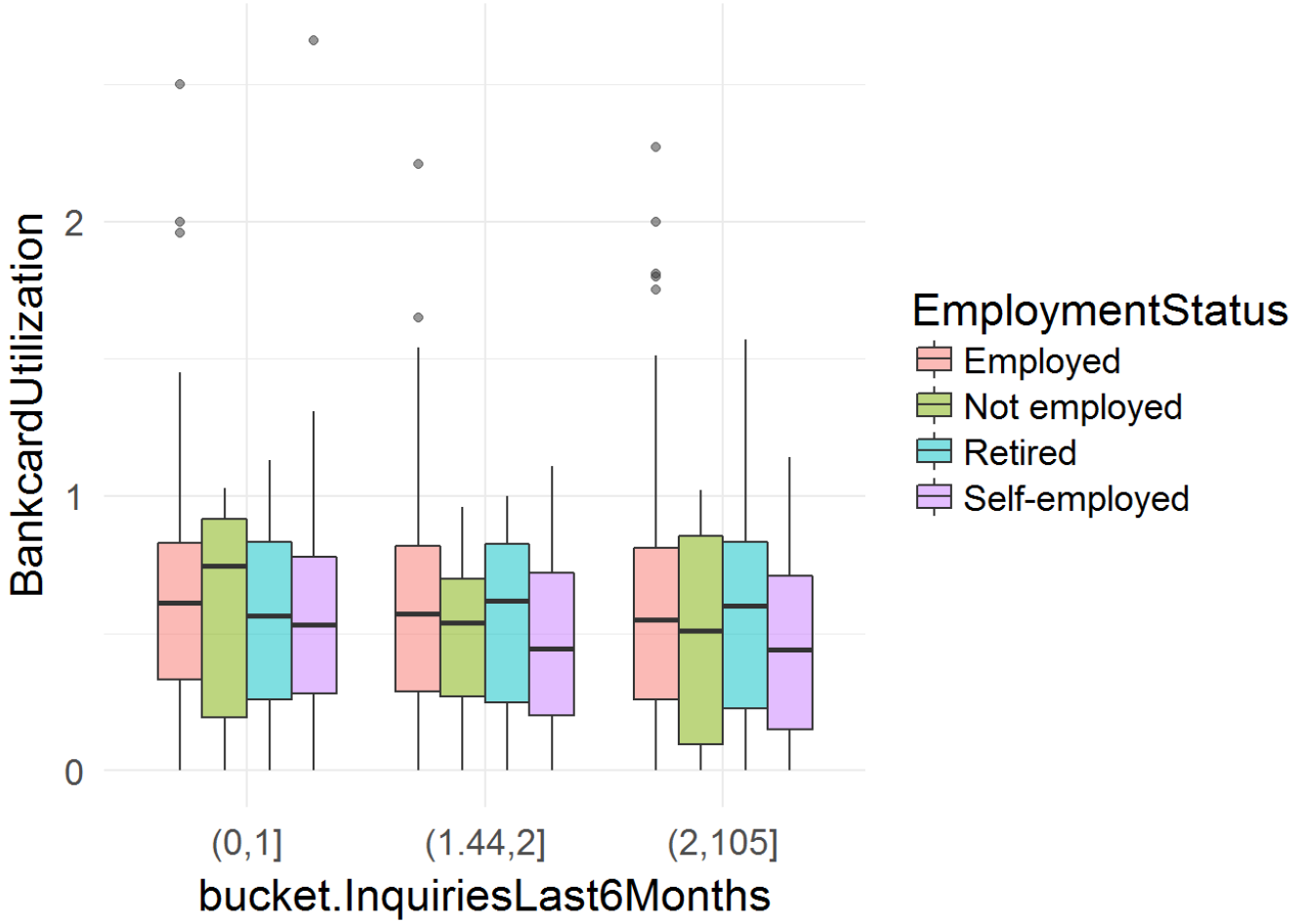
it seems that people with higher monthly incomes are more delinquent, i will add third variable to color the plot based on income Ranege



we can see chance of being delinquent incerases as income goes higher and when incomes reaches to 100K slightly decrase.



i have not seen much relation between bank card utilization and inquires within last 6 month, however it seems higher inquires tends to have abit less utilization, so i am looking in more details and create buckets for number of inquiries

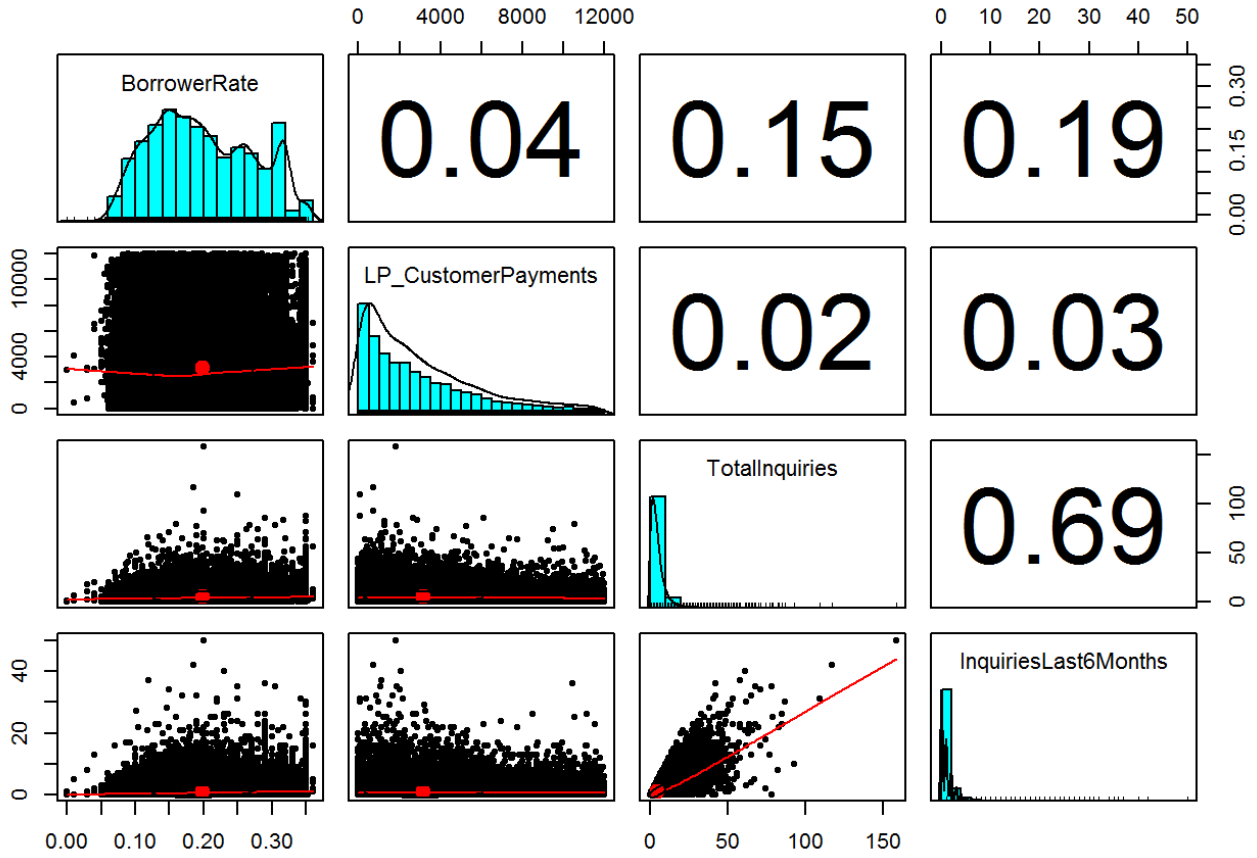


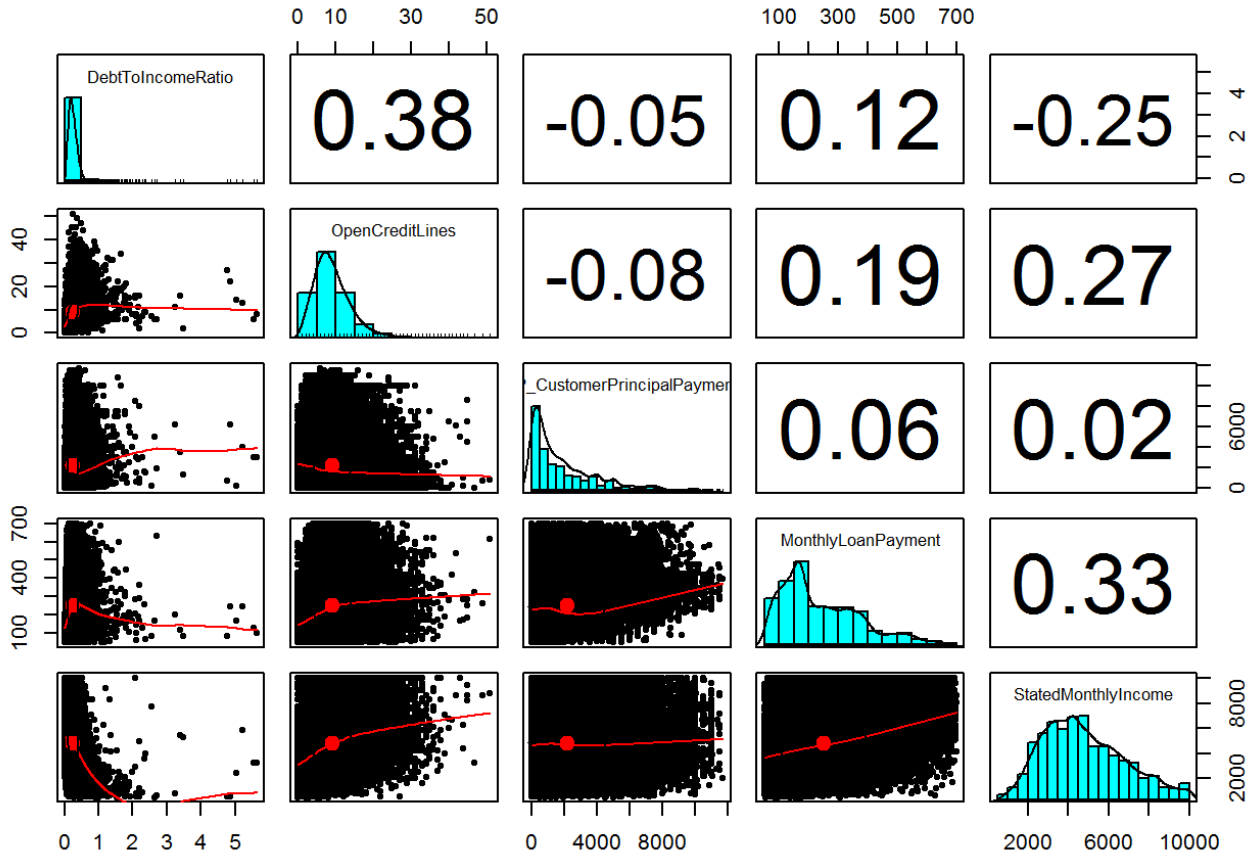
intresting fact i have found is self-employes tends to always having lower median than other people.

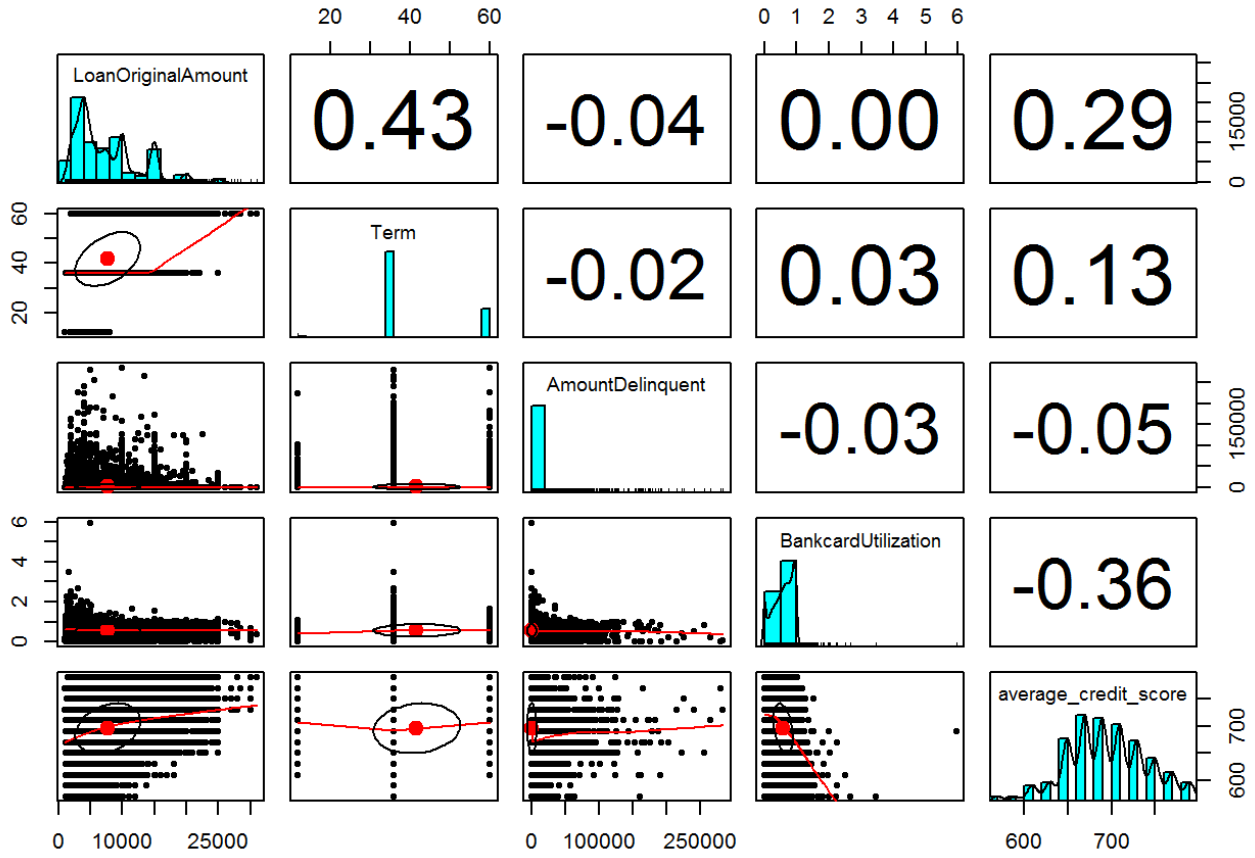
Multivariate analysis

i have created a new variable for average of credit scores (Lower and Upper) and removed upper and lower range variables

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   569.5   669.5   689.5   695.7   729.5   789.5
```


BorrowerRate
LP_CustomerPayments
TotalInquiries
InquiriesLast6Months
0.04
0.15
0.19
0.02
0.03
0.69


DebtToIncomeRatio
OpenCreditLines
_CustomerPrincipalPaymer
MonthlyLoanPayment
StatedMonthlyIncome
0.38
-0.05
0.12
-0.25
-0.08
0.19
0.27
0.06
0.02
0.33



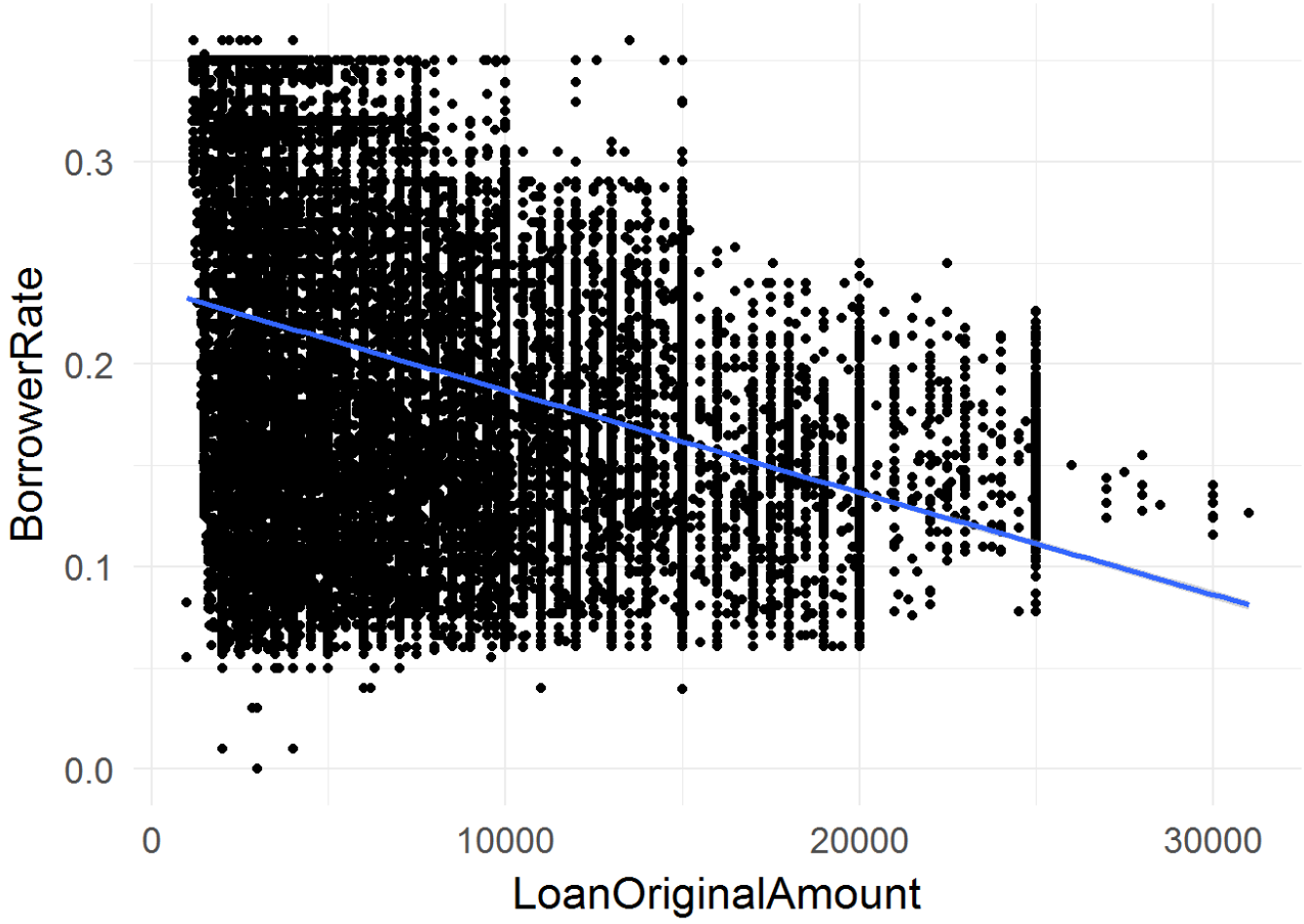
also for simplicity and due to having several columns it might be easier also to look at correlation values only

```
##                              BorrowerRate LP_CustomerPayments
## BorrowerRate                  1.000000000         0.026151565
## LP_CustomerPayments           0.026151565         1.000000000
## TotalInquiries                0.156490416         0.022570500
## InquiriesLast6Months          0.200695291         0.028414700
## DebtToIncomeRatio             0.148873661        -0.013999229
## OpenCreditLines              -0.070953477        -0.068313158
## LP_CustomerPrincipalPayments -0.083515302         0.951795513
## MonthlyLoanPayment           -0.246610229         0.139467865
## StatedMonthlyIncome          -0.143644517         0.032727448
## LoanOriginalAmount           -0.337905321         0.068347377
## Term                          0.004646364        -0.117841384
## AmountDelinquent              0.061881739        -0.001692444
## BankcardUtilization           0.219833111        -0.074049480
## average_credit_score         -0.460198628         0.087119691
##                              TotalInquiries InquiriesLast6Months
## BorrowerRate                    0.156490416           0.20069529
## LP_CustomerPayments             0.022570500           0.02841470
## TotalInquiries                  1.000000000           0.69705293
## InquiriesLast6Months            0.697052925           1.00000000
## DebtToIncomeRatio               0.009634313          -0.02746994
## OpenCreditLines                 0.120936535           0.03324081
## LP_CustomerPrincipalPayments    0.025935294           0.02830321
## MonthlyLoanPayment             -0.077040776          -0.10209987
## StatedMonthlyIncome             0.096477611           0.05001058
## LoanOriginalAmount             -0.109227254          -0.14140661
```

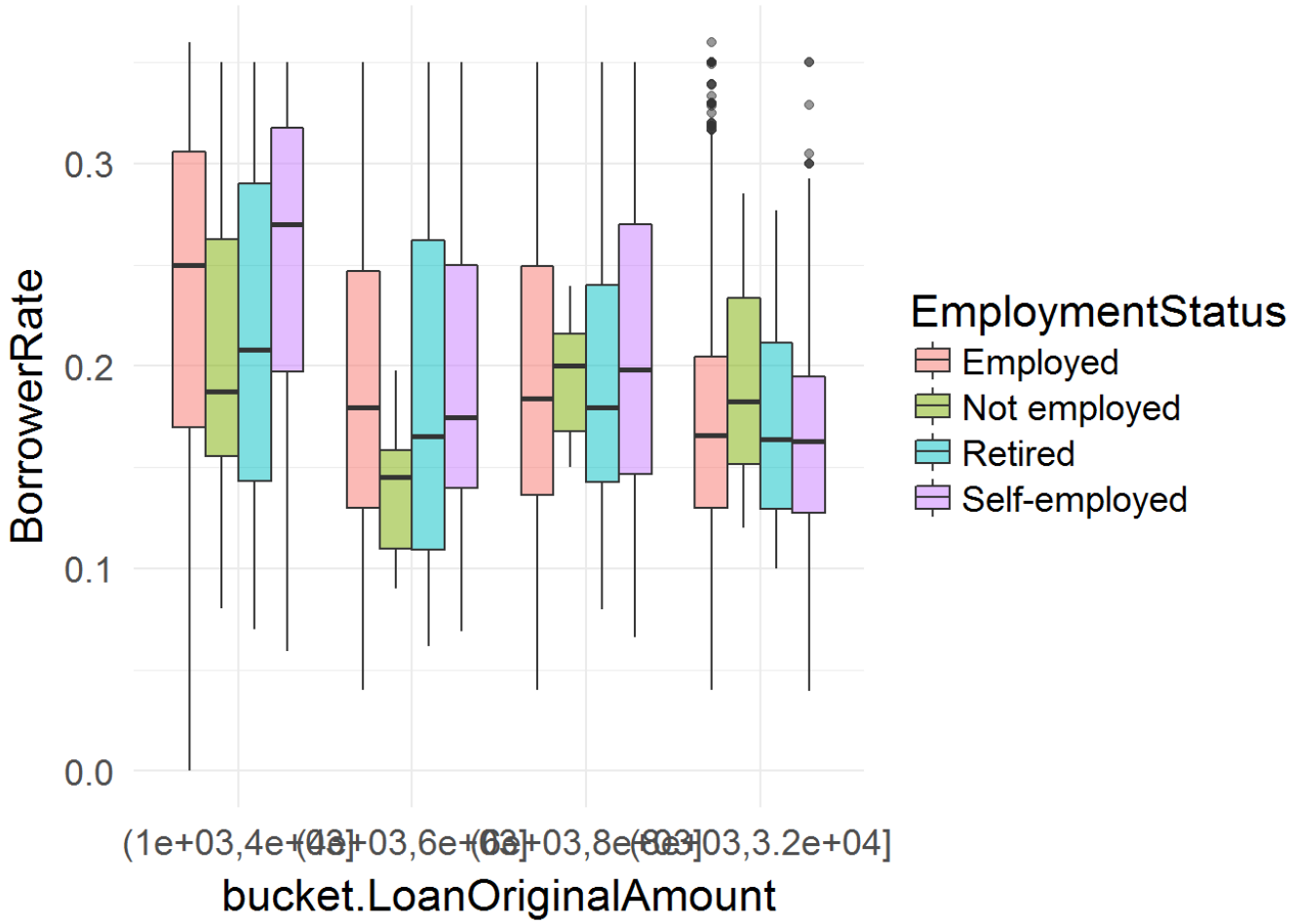
```
## Term                          -0.099237530          -0.11229828
## AmountDelinquent               0.026399073           0.02246410
## BankcardUtilization           -0.014369327          -0.06713653
## average_credit_score          -0.256119339          -0.20014655
##                            DebtToIncomeRatio OpenCreditLines
## BorrowerRate                      0.148873661     -0.07095348
## LP_CustomerPayments              -0.013999229     -0.06831316
## TotalInquiries                    0.009634313      0.12093654
## InquiriesLast6Months             -0.027469942      0.03324081
## DebtToIncomeRatio                 1.000000000      0.38339229
## OpenCreditLines                   0.383392292      1.00000000
## LP_CustomerPrincipalPayments     -0.049216514     -0.07520979
## MonthlyLoanPayment                0.124454076      0.18976134
## StatedMonthlyIncome              -0.249879206      0.26717027
## LoanOriginalAmount                0.086122916      0.20145927
## Term                              0.024816509      0.09719157
## AmountDelinquent                 -0.060286047     -0.06849887
## BankcardUtilization               0.160324103      0.12265506
## average_credit_score              0.005415324      0.12349143
##                            LP_CustomerPrincipalPayments
## BorrowerRate                                -0.083515302
## LP_CustomerPayments                          0.951795513
## TotalInquiries                               0.025935294
## InquiriesLast6Months                         0.028303207
## DebtToIncomeRatio                           -0.049216514
## OpenCreditLines                             -0.075209792
## LP_CustomerPrincipalPayments                 1.000000000
## MonthlyLoanPayment                           0.059784646
## StatedMonthlyIncome                          0.017978512
## LoanOriginalAmount                          -0.017903484
## Term                                        -0.228726640
## AmountDelinquent                            -0.006629417
## BankcardUtilization                         -0.102420859
## average_credit_score                         0.108915248
##                            MonthlyLoanPayment StatedMonthlyIncome
## BorrowerRate                       -0.24661023         -0.14364452
## LP_CustomerPayments                 0.13946786          0.03272745
## TotalInquiries                     -0.07704078          0.09647761
## InquiriesLast6Months               -0.10209987          0.05001058
## DebtToIncomeRatio                   0.12445408         -0.24987921
## OpenCreditLines                     0.18976134          0.26717027
## LP_CustomerPrincipalPayments        0.05978465          0.01797851
## MonthlyLoanPayment                  1.00000000          0.32666550
## StatedMonthlyIncome                 0.32666550          1.00000000
## LoanOriginalAmount                  0.93166234          0.34362711
## Term                                0.16836600          0.10734039
## AmountDelinquent                   -0.02843004          0.02911139
## BankcardUtilization                 0.01718271          0.10905134
## average_credit_score                0.22793169          0.13407573
##                            LoanOriginalAmount         Term
## BorrowerRate                       -0.337905321  0.004646364
## LP_CustomerPayments                 0.068347377 -0.117841384
## TotalInquiries                     -0.109227254 -0.099237530
## InquiriesLast6Months               -0.141406607 -0.112298284
## DebtToIncomeRatio                   0.086122916  0.024816509
## OpenCreditLines                     0.201459266  0.097191565
## LP_CustomerPrincipalPayments       -0.017903484 -0.228726640
```

```
## MonthlyLoanPayment                  0.931662339  0.168365997
## StatedMonthlyIncome                 0.343627107  0.107340390
## LoanOriginalAmount                  1.000000000  0.436408986
## Term                                0.436408986  1.000000000
## AmountDelinquent                   -0.039264342 -0.026019422
## BankcardUtilization                -0.004596677  0.028794494
## average_credit_score                0.295087684  0.140050067
##                              AmountDelinquent BankcardUtilization
## BorrowerRate                      0.061881739         0.219833111
## LP_CustomerPayments              -0.001692444        -0.074049480
## TotalInquiries                    0.026399073        -0.014369327
## InquiriesLast6Months              0.022464101        -0.067136527
## DebtToIncomeRatio                -0.060286047         0.160324103
## OpenCreditLines                  -0.068498871         0.122655057
## LP_CustomerPrincipalPayments     -0.006629417        -0.102420859
## MonthlyLoanPayment               -0.028430035         0.017182707
## StatedMonthlyIncome               0.029111392         0.109051336
## LoanOriginalAmount               -0.039264342        -0.004596677
## Term                             -0.026019422         0.028794494
## AmountDelinquent                  1.000000000        -0.030812783
## BankcardUtilization              -0.030812783         1.000000000
## average_credit_score             -0.050918863        -0.359444538
##                              average_credit_score
## BorrowerRate                         -0.460198628
## LP_CustomerPayments                   0.087119691
## TotalInquiries                       -0.256119339
## InquiriesLast6Months                 -0.200146549
## DebtToIncomeRatio                     0.005415324
## OpenCreditLines                       0.123491430
## LP_CustomerPrincipalPayments          0.108915248
## MonthlyLoanPayment                    0.227931689
## StatedMonthlyIncome                   0.134075730
## LoanOriginalAmount                    0.295087684
## Term                                  0.140050067
## AmountDelinquent                     -0.050918863
## BankcardUtilization                  -0.359444538
## average_credit_score                  1.000000000
```

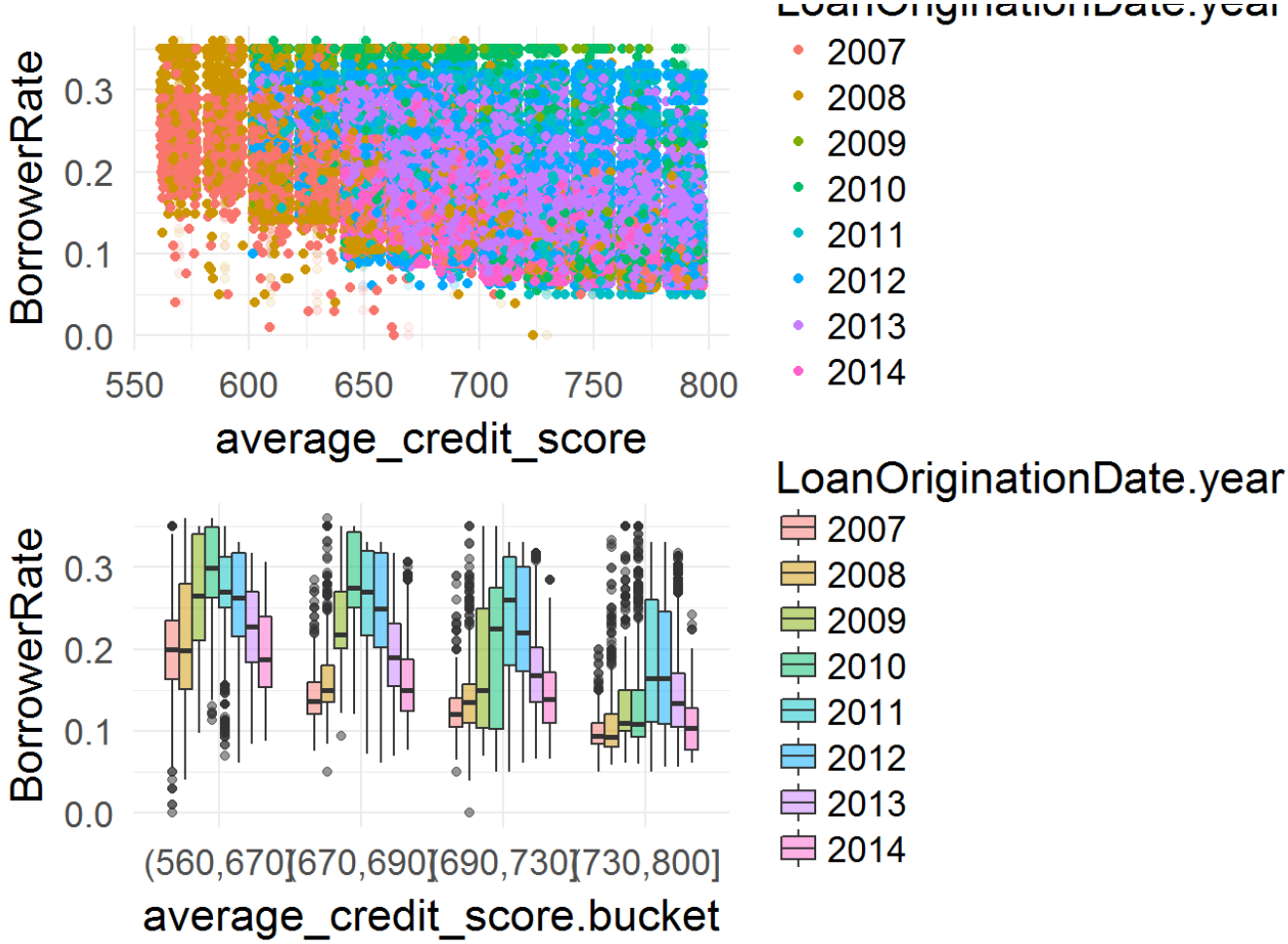
i am intrested in correlations higher than 0.3 as they seems good and as the correlation get closer to 1 it will be better, in this analysis i am looking at Borrower rate and find out variables which are correlated to it and can help, we can see Loan original amount, average credit score having highest correlation with borrower rate.



first look we can see ther seems to be a relation and having higher loan amount tends to getting lower rate, however i will go forward and create buckets for original amount and re create the plot

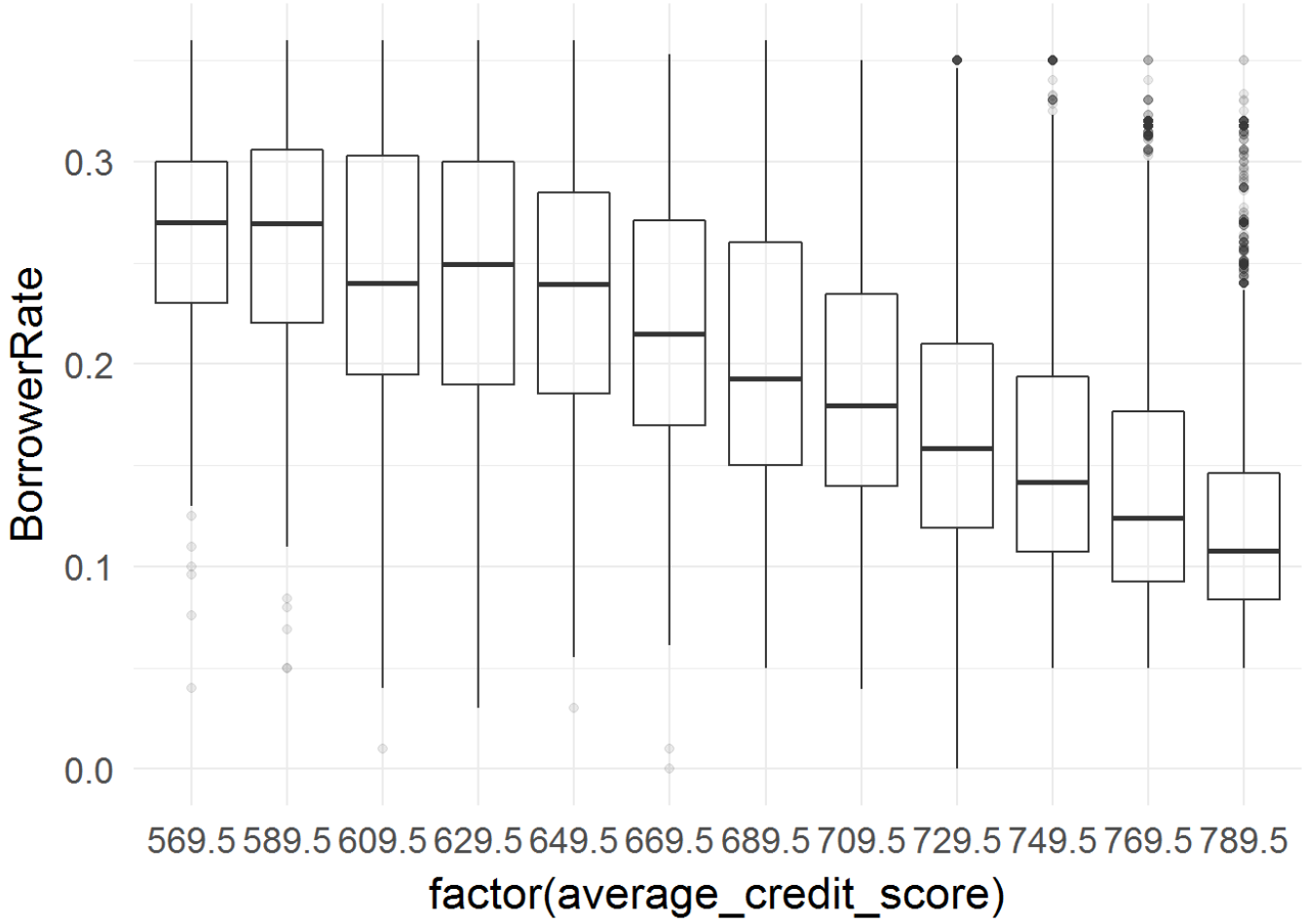


we can see as the loan amount increases, not employed customer tends to have higher borrower rate in compate to others, also it seems Retired customer seems to get lower rate in compare to other groups

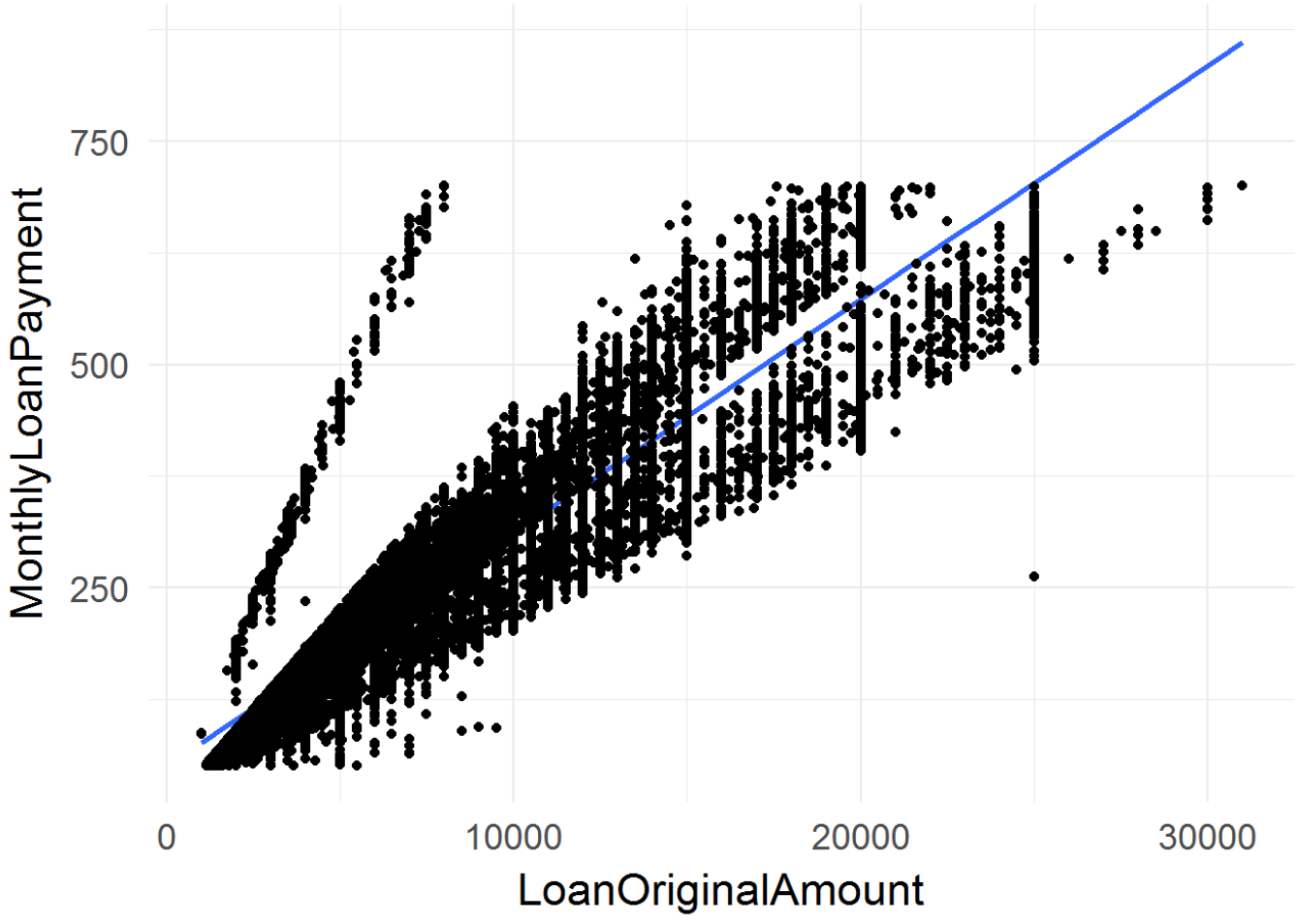


looking at the plot first i was thinking there is nothing intresting about this plot , however when adding parameter loan origin year for colour we can see lighter colours which are belonging to years 2005-2008 have les borrowerRate and they had lower average credit score in compare to recent years (2008 - 2014) which we can see avergae credit scores are above 600 and borrower rate also increases.

also defining a variable as average credit card buckets which has 5 buckets as, first quartile, median, 3rd quartile and maximum and ploting the box plots we can see during years 2007-2014 there is a curve which means we have an increase in borrower rate. we cans see year 2010 for each credit bucket is the pick and then we have a drastical decrease in borrower rate which 2014 is the lowest, and having higher credit score will gives lowest borrower rate.

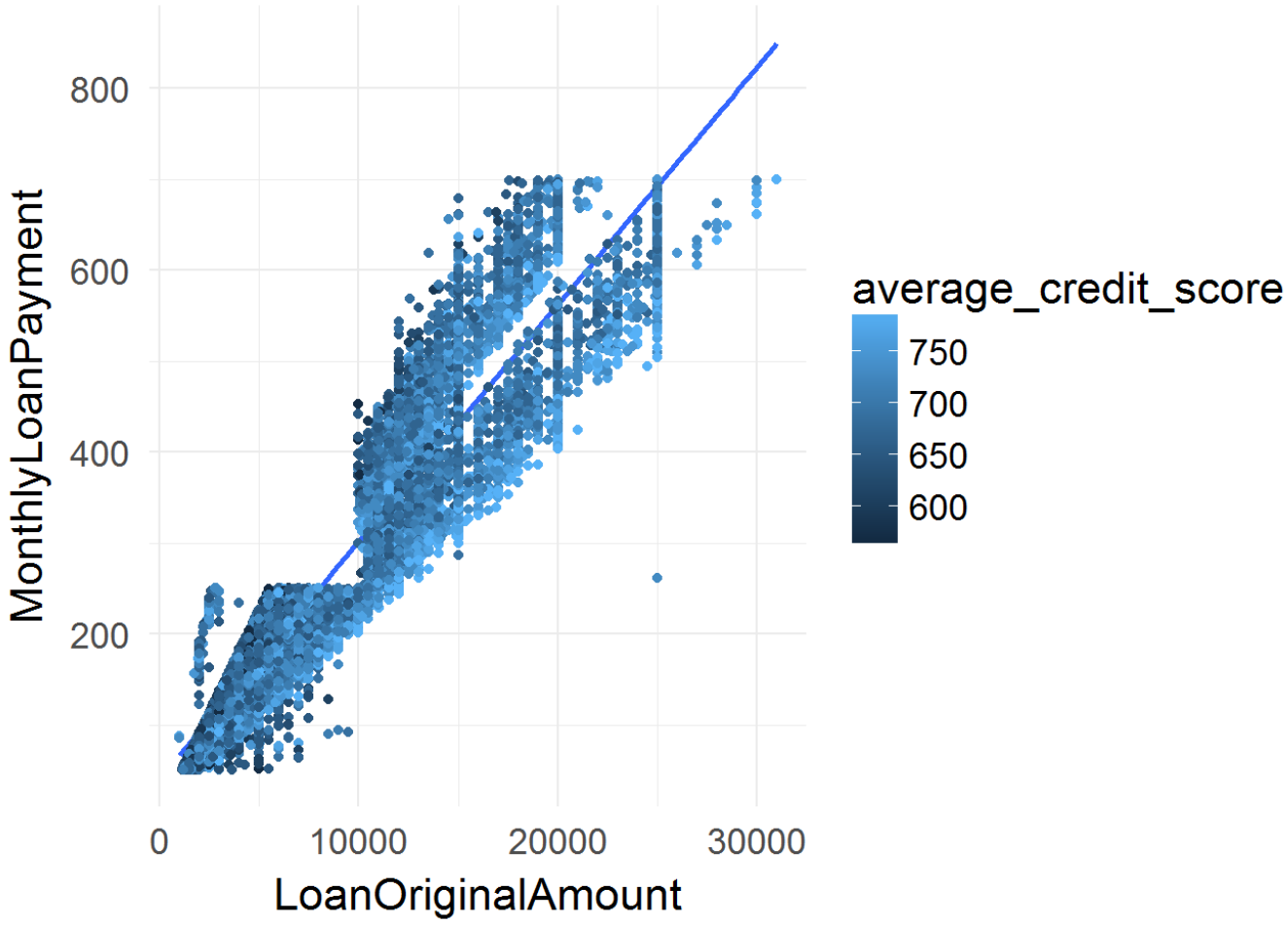


by looking at the boxes and medians we can see having higher credit score cause to get lower borrower rate

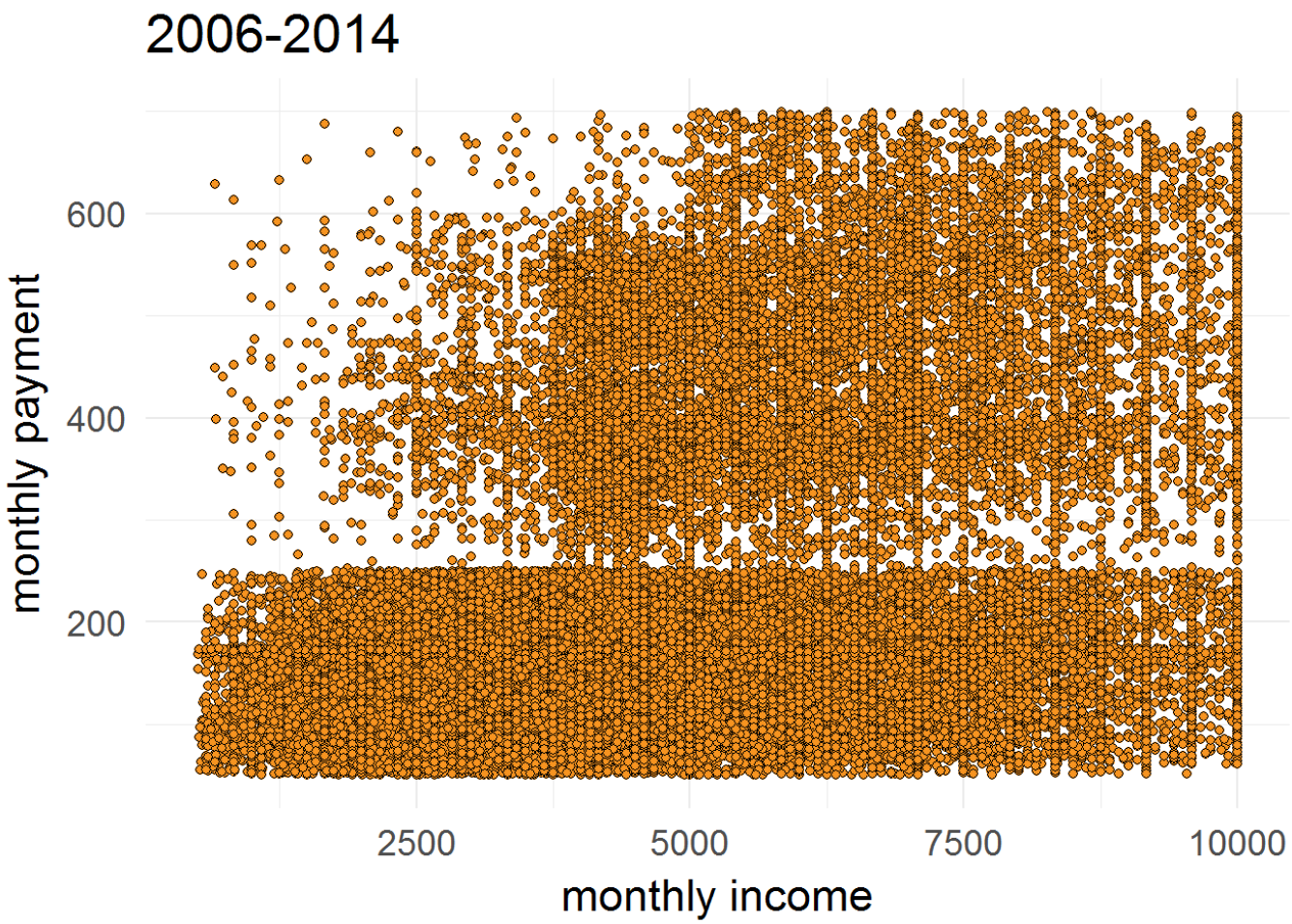


looking at the plot there seems to be intresting linear relation in upper portion of the data, we can see the

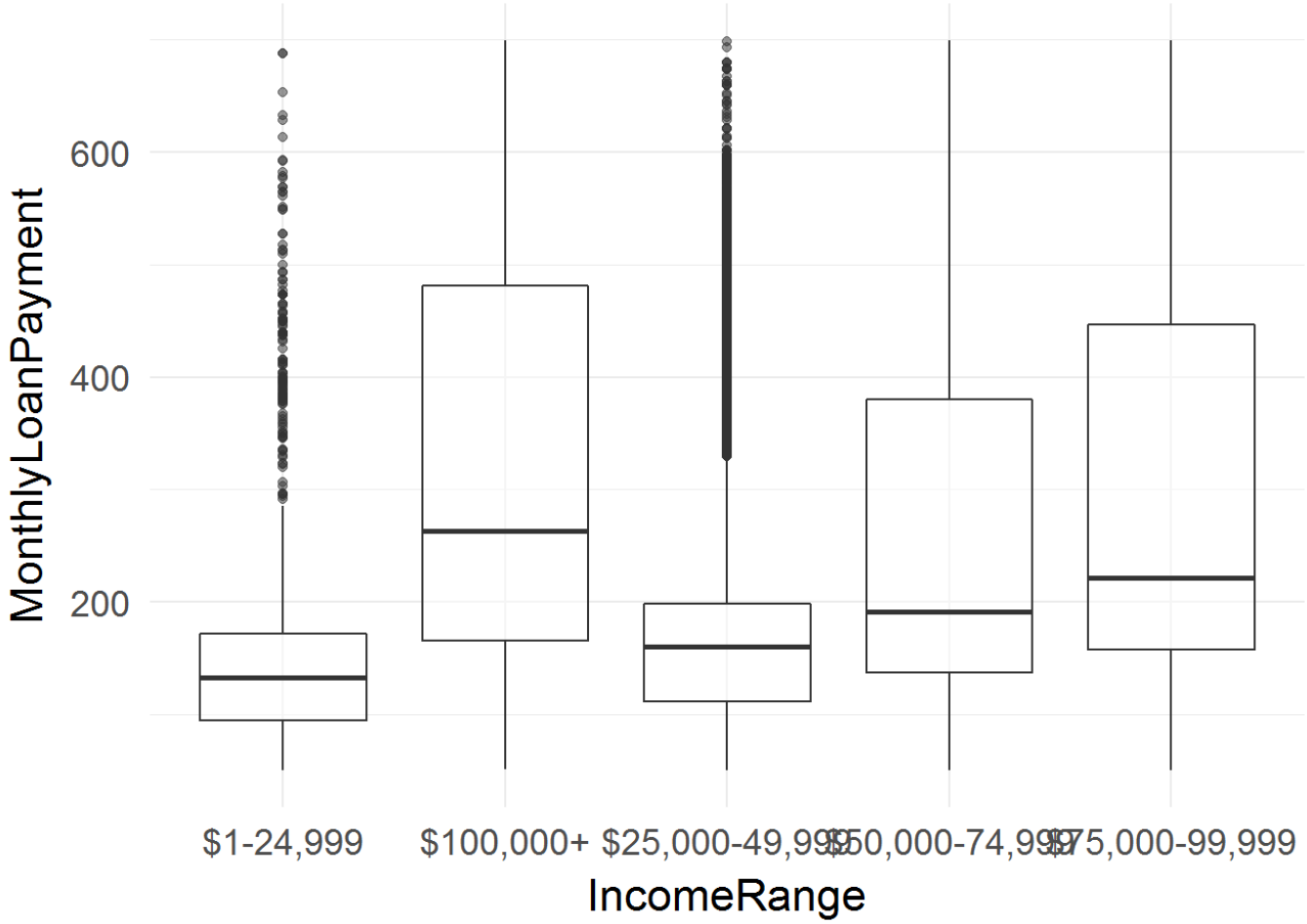
upper line has a monthly loan payment > 250 and loan original amount less than 10000 so i am subseting the data



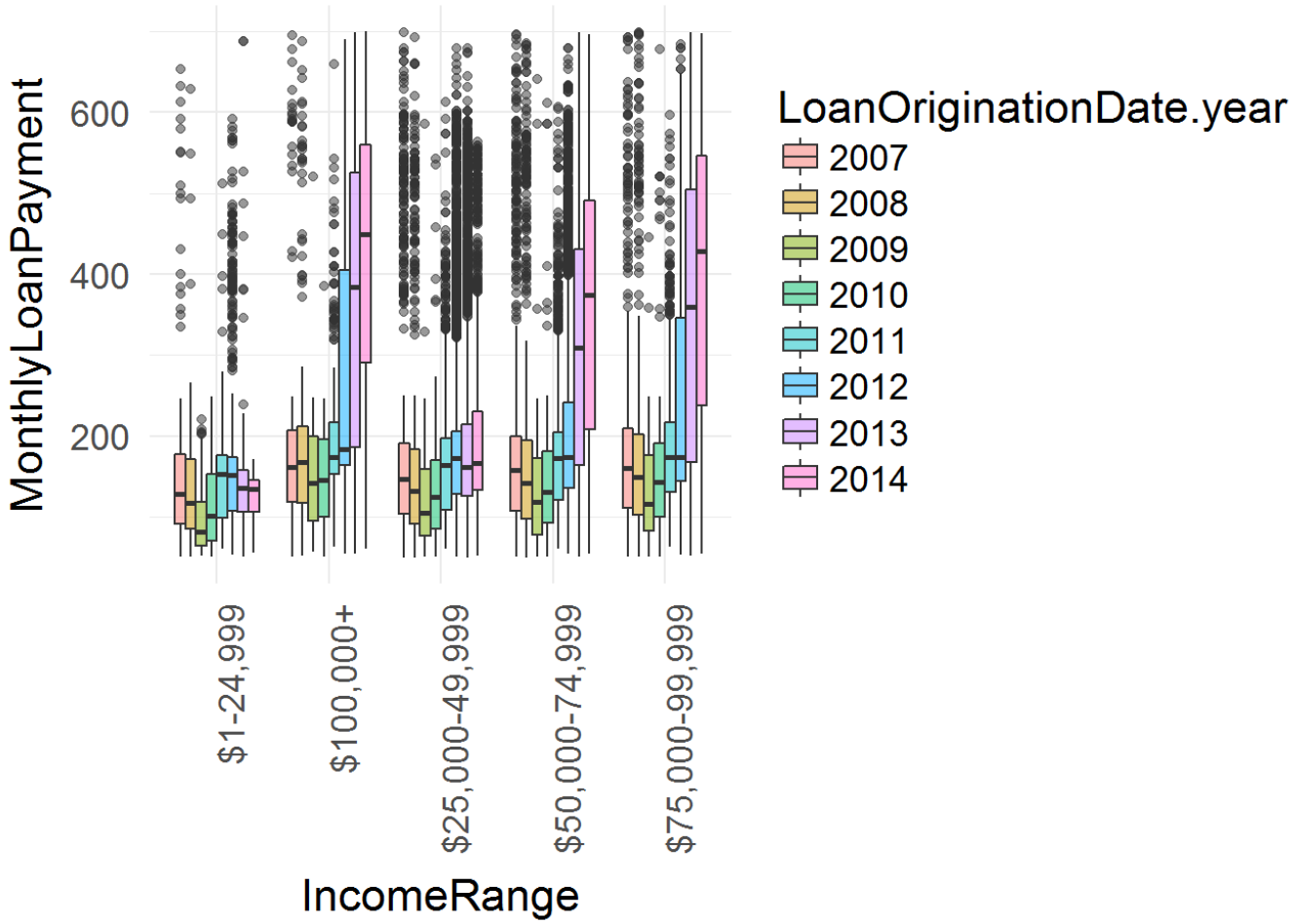
i have tried average_credit_score , Employment Status and year as character to fill the plot but i did not find any intresting relation for that line. however looking at the line can see ther is nice linear relation between as line passed through points (400,400) and (600,600) then slowly converge to lower slope



plot here does not show much as we have high density in the middle , i can see a quadratic curve in the data. however i am not using log as the range of monthtly income and monthly payment is not that different. i will use income range and look at the monthly payment based on the range



we can see as income range increases the monthly payment also increase, now i use loan year as a factor and re plot to see if we can find any important information



we can see within last few years(2011-2014) having higher income range will be ending in more monthly loan payments, and intresting fact for me was that in 2009 and 2010 we are having the lowest monthly payment

developing a Linear model

```
##
## Calls:
## m1: lm(formula = MonthlyLoanPayment ~ StatedMonthlyIncome, data = data)
## m2: lm(formula = MonthlyLoanPayment ~ StatedMonthlyIncome + BorrowerRate +
##     LP_CustomerPayments + InquiriesLast6Months, data = data)
## m3: lm(formula = MonthlyLoanPayment ~ StatedMonthlyIncome + BorrowerRate +
##     LP_CustomerPayments + InquiriesLast6Months + LoanOriginationDate.month +
##     LoanOriginationDate.year + DebtToIncomeRatio + OpenCreditLines,
##     data = data)
## m4: lm(formula = MonthlyLoanPayment ~ StatedMonthlyIncome + BorrowerRate +
##     LP_CustomerPayments + InquiriesLast6Months + LoanOriginationDate.month +
##     LoanOriginationDate.year + DebtToIncomeRatio + OpenCreditLines +
##     LP_CustomerPrincipalPayments + LoanOriginalAmount + Term +
##     average_credit_score, data = data)
##
## ================================================================================
============
##                                             m1            m2           m3
m4
## --------------------------------------------------------------------------------
--------------
##   (Intercept)                            101.748***    191.017***    -31.704***
144.676***
##                                           (1.450)       (2.263)       (3.329)
(2.495)
##   StatedMonthlyIncome                      0.027***      0.025***      0.019***
-0.000
##                                           (0.000)       (0.000)       (0.000)
(0.000)
##   BorrowerRate                                        -451.328***   -284.872***
180.012***
##                                                        (7.602)       (7.467)
(2.419)
##   LP_CustomerPayments                                  0.008***      0.029***
0.008***
##                                                        (0.000)       (0.000)
(0.000)
##   InquiriesLast6Months                                -6.373***      1.801***
0.422***
##                                                        (0.286)       (0.264)
(0.066)
##   LoanOriginationDate.month: Aug/Apr                                14.106***
0.893
##                                                                     (2.492)
(0.621)
##   LoanOriginationDate.month: Dec/Apr                                56.854***
2.866***
##                                                                     (2.487)
(0.624)
```

```
##   LoanOriginationDate.month: Feb/Apr                                -9.669***
-1.744**
##                                                                      (2.704)
(0.673)
##   LoanOriginationDate.month: Jan/Apr                               -19.348***
-1.933**
##                                                                      (2.693)
(0.671)
##   LoanOriginationDate.month: Jul/Apr                                 7.439**
0.238
##                                                                      (2.484)
(0.619)
##   LoanOriginationDate.month: Jun/Apr                                 5.556*
1.369*
##                                                                      (2.513)
(0.626)
##   LoanOriginationDate.month: Mar/Apr                                -4.322
-1.563*
##                                                                      (2.575)
(0.641)
##   LoanOriginationDate.month: May/Apr                                 0.764
-0.312
##                                                                      (2.549)
(0.635)
##   LoanOriginationDate.month: Nov/Apr                                45.676***
2.318***
##                                                                      (2.531)
(0.633)
##   LoanOriginationDate.month: Oct/Apr                                35.825***
1.995**
##                                                                      (2.440)
(0.609)
##   LoanOriginationDate.month: Sep/Apr                                24.952***
0.884
##                                                                      (2.520)
(0.628)
##   LoanOriginationDate.year: 2008/2007                                5.270*
-2.491***
##                                                                      (2.340)
(0.584)
##   LoanOriginationDate.year: 2009/2007                               -50.819***
-7.761***
##                                                                      (3.975)
(1.008)
##   LoanOriginationDate.year: 2010/2007                               -17.228***
-5.298***
##                                                                      (2.846)
(0.738)
##   LoanOriginationDate.year: 2011/2007                                2.226
-0.272
##                                                                      (2.467)
(0.647)
##   LoanOriginationDate.year: 2012/2007                               53.755***
12.639***
##                                                                      (2.180)
(0.592)
##   LoanOriginationDate.year: 2013/2007                              160.030***
```

```
21.660***
##                                                                                 (2.150)
(0.596)
##   LoanOriginationDate.year: 2014/2007                                         275.524***
29.713***
##                                                                                 (3.028)
(0.826)
##   DebtToIncomeRatio                                                           178.445***
14.831***
##                                                                                 (3.681)
(0.936)
##   OpenCreditLines                                                              -1.410***
-0.067*
##                                                                                 (0.118)
(0.029)
##   LP_CustomerPrincipalPayments
-0.006***
##
(0.000)
##   LoanOriginalAmount
0.029***
##
(0.000)
##   Term
-3.999***
##
(0.013)
##   average_credit_score
-0.031***
##
(0.003)
## ------------------------------------------------------------------------------
--------------
##   R-squared                                      0.1           0.2           0.5
1.0
##   adj. R-squared                                 0.1           0.2           0.5
1.0
##   sigma                                        142.2         135.8         113.7
28.3
##   F                                           9887.9        4162.8        1990.0
58290.9
##   p                                              0.0           0.0           0.0
0.0
##   Log-likelihood                          -388104.1     -385287.0     -350124.7
-270980.7
##   Deviance                              1230530808.7  1123034597.3  735494535.7
45564069.0
##   AIC                                      776214.1      770586.1      700301.4
542021.4
##   BIC                                      776241.2      770640.2      700534.1
542289.9
##   N                                         60869         60863         56909
56909
## ==============================================================================
============
```

last few plots shows that we can use some of the variables inside the model and add other variables and check the results, we can see we have good R-squared and adjusted R-squared, however there is a chance of over fitting which i have not consider for this project. also to be sure of chosen variables are rubust i am using an automatic variable selection based on forward and backeward regression using AIC metric

```
## Start:  AIC=197768.4
## MonthlyLoanPayment ~ (BorrowerRate + LoanStatus + LP_CustomerPayments +
##     TotalInquiries + InquiriesLast6Months + LoanOriginationDate.month +
##     LoanOriginationDate.year + DebtToIncomeRatio + OpenCreditLines +
##     LP_CustomerPrincipalPayments + StatedMonthlyIncome + LoanOriginalAmount +
##     Term + EmploymentStatus + AmountDelinquent + bucket.InquiriesLast6Months +
##     BankcardUtilization + IncomeRange + average_credit_score +
##     bucket.LoanOriginalAmount + average_credit_score.bucket) -
##     LoanStatus
##
##                                Df Sum of Sq      RSS    AIC
## - EmploymentStatus              3      2615 19328445 197767
## - InquiriesLast6Months          1       155 19325985 197767
## - BankcardUtilization           1       272 19326102 197767
## - StatedMonthlyIncome           1       400 19326230 197767
## <none>                                      19325830 197768
## - OpenCreditLines               1      1824 19327654 197769
## - TotalInquiries                1      1989 19327819 197770
## - bucket.InquiriesLast6Months   2      5795 19331625 197774
## - AmountDelinquent              1      5915 19331745 197776
## - LoanOriginationDate.month    11     23203 19349034 197783
## - IncomeRange                   4     18615 19344446 197790
## - DebtToIncomeRatio             1     51189 19377019 197847
## - average_credit_score          1     94099 19419930 197915
## - LP_CustomerPrincipalPayments  1    143080 19468910 197993
## - average_credit_score.bucket   3    175898 19501728 198040
## - LP_CustomerPayments           1    225559 19551389 198122
## - LoanOriginationDate.year      7   1005856 20331686 199310
## - bucket.LoanOriginalAmount     3   3213329 22539159 202479
## - BorrowerRate                  1   3885813 23211643 203385
## - Term                          1  34103528 53429358 228951
## - LoanOriginalAmount            1  71403430 90729260 245189
##
## Step:  AIC=197766.5
## MonthlyLoanPayment ~ BorrowerRate + LP_CustomerPayments + TotalInquiries +
##     InquiriesLast6Months + LoanOriginationDate.month + LoanOriginationDate.year
+
##     DebtToIncomeRatio + OpenCreditLines + LP_CustomerPrincipalPayments +
##     StatedMonthlyIncome + LoanOriginalAmount + Term + AmountDelinquent +
##     bucket.InquiriesLast6Months + BankcardUtilization + IncomeRange +
##     average_credit_score + bucket.LoanOriginalAmount + average_credit_score.buck
et
##
##                                Df Sum of Sq      RSS    AIC
## - InquiriesLast6Months          1       141 19328586 197765
## - BankcardUtilization           1       285 19328730 197765
## - StatedMonthlyIncome           1       425 19328870 197765
## <none>                                      19328445 197767
## - OpenCreditLines               1      1959 19330404 197768
## - TotalInquiries                1      1971 19330416 197768
## + EmploymentStatus              3      2615 19325830 197768
```

```
## - bucket.InquiriesLast6Months   2      5734 19334179 197772
## - AmountDelinquent              1      5944 19334389 197774
## - LoanOriginationDate.month    11     23127 19351572 197781
## - IncomeRange                   4     18875 19347320 197788
## - DebtToIncomeRatio             1     52127 19380572 197847
## - average_credit_score          1     94889 19423334 197915
## - LP_CustomerPrincipalPayments  1    143009 19471454 197991
## - average_credit_score.bucket   3    176656 19505101 198040
## - LP_CustomerPayments           1    225514 19553959 198120
## - LoanOriginationDate.year      7   1011468 20339913 199317
## - bucket.LoanOriginalAmount     3   3214092 22542537 202478
## - BorrowerRate                  1   3888253 23216698 203385
## - Term                          1  34110162 53438607 228950
## - LoanOriginalAmount            1  71412801 90741246 245187
##
## Step:  AIC=197764.7
## MonthlyLoanPayment ~ BorrowerRate + LP_CustomerPayments + TotalInquiries +
##     LoanOriginationDate.month + LoanOriginationDate.year + DebtToIncomeRatio +
##     OpenCreditLines + LP_CustomerPrincipalPayments + StatedMonthlyIncome +
##     LoanOriginalAmount + Term + AmountDelinquent + bucket.InquiriesLast6Months +

##     BankcardUtilization + IncomeRange + average_credit_score +
##     bucket.LoanOriginalAmount + average_credit_score.bucket
##
##                                Df Sum of Sq      RSS    AIC
## - BankcardUtilization           1       283 19328869 197763
## - StatedMonthlyIncome           1       426 19329012 197763
## <none>                                      19328586 197765
## - OpenCreditLines               1      1922 19330509 197766
## - TotalInquiries                1      2017 19330603 197766
## + InquiriesLast6Months          1       141 19328445 197767
## + EmploymentStatus              3      2601 19325985 197767
## - bucket.InquiriesLast6Months   2      6166 19334752 197771
## - AmountDelinquent              1      5927 19334513 197772
## - LoanOriginationDate.month    11     23126 19351712 197779
## - IncomeRange                   4     18874 19347461 197787
## - DebtToIncomeRatio             1     52112 19380698 197845
## - average_credit_score          1     95023 19423609 197913
## - LP_CustomerPrincipalPayments  1    142947 19471534 197989
## - average_credit_score.bucket   3    177039 19505625 198038
## - LP_CustomerPayments           1    225627 19554213 198119
## - LoanOriginationDate.year      7   1011575 20340161 199315
## - bucket.LoanOriginalAmount     3   3213997 22542583 202476
## - BorrowerRate                  1   3888666 23217252 203384
## - Term                          1  34112768 53441354 228950
## - LoanOriginalAmount            1  71412817 90741403 245186
##
## Step:  AIC=197763.2
## MonthlyLoanPayment ~ BorrowerRate + LP_CustomerPayments + TotalInquiries +
##     LoanOriginationDate.month + LoanOriginationDate.year + DebtToIncomeRatio +
##     OpenCreditLines + LP_CustomerPrincipalPayments + StatedMonthlyIncome +
##     LoanOriginalAmount + Term + AmountDelinquent + bucket.InquiriesLast6Months +

##     IncomeRange + average_credit_score + bucket.LoanOriginalAmount +
##     average_credit_score.bucket
##
##                                Df Sum of Sq      RSS    AIC
```

```
## - StatedMonthlyIncome           1       481 19329349 197762
## <none>                                      19328869 197763
## - OpenCreditLines               1      2042 19330911 197764
## - TotalInquiries                1      2124 19330993 197765
## + BankcardUtilization           1       283 19328586 197765
## + InquiriesLast6Months          1       138 19328730 197765
## + EmploymentStatus              3      2615 19326254 197765
## - bucket.InquiriesLast6Months   2      6454 19335323 197769
## - AmountDelinquent              1      5821 19334690 197770
## - LoanOriginationDate.month    11     23191 19352059 197778
## - IncomeRange                   4     18642 19347511 197785
## - DebtToIncomeRatio             1     52094 19380963 197844
## - average_credit_score          1     98556 19427425 197917
## - LP_CustomerPrincipalPayments  1    142874 19471743 197987
## - average_credit_score.bucket   3    176778 19505646 198036
## - LP_CustomerPayments           1    225469 19554338 198117
## - LoanOriginationDate.year      7   1011543 20340412 199313
## - bucket.LoanOriginalAmount     3   3215244 22544113 202476
## - BorrowerRate                  1   3912124 23240993 203413
## - Term                          1  34112928 53441796 228948
## - LoanOriginalAmount            1  71422755 90751624 245187
##
## Step:  AIC=197762
## MonthlyLoanPayment ~ BorrowerRate + LP_CustomerPayments + TotalInquiries +
##     LoanOriginationDate.month + LoanOriginationDate.year + DebtToIncomeRatio +
##     OpenCreditLines + LP_CustomerPrincipalPayments + LoanOriginalAmount +
##     Term + AmountDelinquent + bucket.InquiriesLast6Months + IncomeRange +
##     average_credit_score + bucket.LoanOriginalAmount + average_credit_score.buck
et
##
##                                Df Sum of Sq      RSS    AIC
## <none>                                      19329349 197762
## + StatedMonthlyIncome           1       481 19328869 197763
## - TotalInquiries                1      2062 19331411 197763
## + BankcardUtilization           1       337 19329012 197763
## - OpenCreditLines               1      2301 19331650 197764
## + InquiriesLast6Months          1       139 19329210 197764
## + EmploymentStatus              3      2643 19326707 197764
## - bucket.InquiriesLast6Months   2      6419 19335768 197768
## - AmountDelinquent              1      5873 19335223 197769
## - LoanOriginationDate.month    11     23178 19352528 197777
## - IncomeRange                   4     18546 19347895 197783
## - DebtToIncomeRatio             1     54326 19383676 197846
## - average_credit_score          1     98728 19428077 197916
## - LP_CustomerPrincipalPayments  1    142969 19472318 197986
## - average_credit_score.bucket   3    176852 19506201 198035
## - LP_CustomerPayments           1    225438 19554787 198116
## - LoanOriginationDate.year      7   1011168 20340518 199312
## - bucket.LoanOriginalAmount     3   3214769 22544118 202474
## - BorrowerRate                  1   3911650 23241000 203411
## - Term                          1  34117345 53446694 228949
## - LoanOriginalAmount            1  71535609 90864958 245223
```

```
##
## Call:
## lm(formula = MonthlyLoanPayment ~ BorrowerRate + LP_CustomerPayments +
```

```
##     TotalInquiries + LoanOriginationDate.month + LoanOriginationDate.year +
##     DebtToIncomeRatio + OpenCreditLines + LP_CustomerPrincipalPayments +
##     LoanOriginalAmount + Term + AmountDelinquent + bucket.InquiriesLast6Months +

##     IncomeRange + average_credit_score + bucket.LoanOriginalAmount +
##     average_credit_score.bucket, data = data)
##
## Coefficients:
##                                (Intercept)
##                                  5.897e+01
##                                BorrowerRate
##                                  2.362e+02
##                       LP_CustomerPayments
##                                  4.454e-03
##                             TotalInquiries
##                                  5.153e-02
##             LoanOriginationDate.monthAug
##                                 -1.295e+00
##             LoanOriginationDate.monthDec
##                                 -2.747e+00
##             LoanOriginationDate.monthFeb
##                                 -2.907e+00
##             LoanOriginationDate.monthJan
##                                 -1.828e+00
##             LoanOriginationDate.monthJul
##                                 -1.060e+00
##             LoanOriginationDate.monthJun
##                                 -6.625e-02
##             LoanOriginationDate.monthMar
##                                 -1.478e+00
##             LoanOriginationDate.monthMay
##                                 -7.905e-01
##             LoanOriginationDate.monthNov
##                                 -2.888e+00
##             LoanOriginationDate.monthOct
##                                 -1.461e+00
##             LoanOriginationDate.monthSep
##                                 -1.949e+00
##             LoanOriginationDate.year2008
##                                 -4.586e+00
##             LoanOriginationDate.year2009
##                                 -9.006e+00
##             LoanOriginationDate.year2010
##                                 -9.787e+00
##             LoanOriginationDate.year2011
##                                 -4.277e+00
##             LoanOriginationDate.year2012
##                                  7.935e+00
##             LoanOriginationDate.year2013
##                                  1.233e+01
##             LoanOriginationDate.year2014
##                                  1.206e+01
##                          DebtToIncomeRatio
##                                  9.752e+00
##                            OpenCreditLines
##                                 -6.581e-02
##             LP_CustomerPrincipalPayments
```

```
##                                -4.269e-03
##                        LoanOriginalAmount
##                                 2.629e-02
##                                      Term
##                                -4.072e+00
##                           AmountDelinquent
##                                -6.313e-05
##      bucket.InquiriesLast6Months(1.44,2]
##                                 1.011e+00
##       bucket.InquiriesLast6Months(2,105]
##                                 9.990e-01
##                      IncomeRange$100,000+
##                                 1.092e+00
##                 IncomeRange$25,000-49,999
##                                 2.215e+00
##                 IncomeRange$50,000-74,999
##                                 3.010e+00
##                 IncomeRange$75,000-99,999
##                                 2.030e+00
##                      average_credit_score
##                                 1.039e-01
##   bucket.LoanOriginalAmount(4e+03,6e+03]
##                                 2.345e+01
##   bucket.LoanOriginalAmount(6e+03,8e+03]
##                                 3.880e+01
## bucket.LoanOriginalAmount(8e+03,3.2e+04]
##                                 6.404e+01
##     average_credit_score.bucket(670,690]
##                                -2.889e+00
##     average_credit_score.bucket(690,730]
##                                -6.555e+00
##     average_credit_score.bucket(730,800]
##                                -1.549e+01
```

```
##
## Call:
## lm(formula = MonthlyLoanPayment ~ . - LoanStatus, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -488.63  -12.20   -2.53    9.26  139.34
##
## Coefficients:
##                                          Estimate Std. Error  t value
## (Intercept)                             6.019e+01  5.611e+00   10.727
## BorrowerRate                            2.364e+02  3.013e+00   78.463
## LP_CustomerPayments                     4.456e-03  2.357e-04   18.904
## TotalInquiries                          5.980e-02  3.369e-02    1.775
## InquiriesLast6Months                   -5.529e-02  1.115e-01   -0.496
## LoanOriginationDate.monthAug           -1.282e+00  7.405e-01   -1.732
## LoanOriginationDate.monthDec           -2.738e+00  7.537e-01   -3.633
## LoanOriginationDate.monthFeb           -2.915e+00  8.038e-01   -3.627
## LoanOriginationDate.monthJan           -1.840e+00  8.028e-01   -2.292
## LoanOriginationDate.monthJul           -1.045e+00  7.355e-01   -1.420
## LoanOriginationDate.monthJun           -5.658e-02  7.424e-01   -0.076
## LoanOriginationDate.monthMar           -1.492e+00  7.622e-01   -1.957
```

```
## LoanOriginationDate.monthMay             -7.962e-01  7.494e-01   -1.062
## LoanOriginationDate.monthNov             -2.881e+00  7.632e-01   -3.775
## LoanOriginationDate.monthOct             -1.448e+00  7.265e-01   -1.994
## LoanOriginationDate.monthSep             -1.940e+00  7.441e-01   -2.607
## LoanOriginationDate.year2008             -4.549e+00  6.173e-01   -7.369
## LoanOriginationDate.year2009             -8.882e+00  1.198e+00   -7.416
## LoanOriginationDate.year2010             -9.637e+00  8.768e-01  -10.992
## LoanOriginationDate.year2011             -4.142e+00  7.566e-01   -5.475
## LoanOriginationDate.year2012              8.069e+00  6.939e-01   11.628
## LoanOriginationDate.year2013              1.246e+01  6.964e-01   17.887
## LoanOriginationDate.year2014              1.221e+01  9.921e-01   12.304
## DebtToIncomeRatio                         9.655e+00  1.072e+00    9.006
## OpenCreditLines                          -5.920e-02  3.482e-02   -1.700
## LP_CustomerPrincipalPayments             -4.272e-03  2.837e-04  -15.056
## StatedMonthlyIncome                      -1.985e-04  2.493e-04   -0.796
## LoanOriginalAmount                        2.630e-02  7.818e-05  336.346
## Term                                     -4.072e+00  1.752e-02 -232.448
## EmploymentStatusNot employed             -4.654e+00  8.908e+00   -0.522
## EmploymentStatusRetired                  -8.216e-01  1.622e+00   -0.507
## EmploymentStatusSelf-employed             2.687e+00  1.426e+00    1.884
## AmountDelinquent                         -6.343e-05  2.072e-05   -3.061
## bucket.InquiriesLast6Months(1.44,2]       1.040e+00  3.755e-01    2.771
## bucket.InquiriesLast6Months(2,105]        1.121e+00  4.909e-01    2.283
## BankcardUtilization                      -3.417e-01  5.209e-01   -0.656
## IncomeRange$100,000+                      2.553e+00  1.985e+00    1.286
## IncomeRange$25,000-49,999                 2.511e+00  7.536e-01    3.331
## IncomeRange$50,000-74,999                 3.689e+00  1.068e+00    3.456
## IncomeRange$75,000-99,999                 3.100e+00  1.500e+00    2.067
## average_credit_score                      1.026e-01  8.404e-03   12.210
## bucket.LoanOriginalAmount(4e+03,6e+03]    2.344e+01  4.589e-01   51.086
## bucket.LoanOriginalAmount(6e+03,8e+03]    3.881e+01  7.308e-01   53.115
## bucket.LoanOriginalAmount(8e+03,3.2e+04]  6.405e+01  9.975e-01   64.215
## average_credit_score.bucket(670,690]     -2.875e+00  5.353e-01   -5.370
## average_credit_score.bucket(690,730]     -6.528e+00  6.612e-01   -9.873
## average_credit_score.bucket(730,800]     -1.545e+01  1.008e+00  -15.321
##                                          Pr(>|t|)
## (Intercept)                               < 2e-16 ***
## BorrowerRate                              < 2e-16 ***
## LP_CustomerPayments                       < 2e-16 ***
## TotalInquiries                           0.075891 .
## InquiriesLast6Months                     0.620045
## LoanOriginationDate.monthAug             0.083340 .
## LoanOriginationDate.monthDec             0.000281 ***
## LoanOriginationDate.monthFeb             0.000287 ***
## LoanOriginationDate.monthJan             0.021889 *
## LoanOriginationDate.monthJul             0.155473
## LoanOriginationDate.monthJun             0.939257
## LoanOriginationDate.monthMar             0.050347 .
## LoanOriginationDate.monthMay             0.288058
## LoanOriginationDate.monthNov             0.000160 ***
## LoanOriginationDate.monthOct             0.046186 *
## LoanOriginationDate.monthSep             0.009151 **
## LoanOriginationDate.year2008             1.77e-13 ***
## LoanOriginationDate.year2009             1.24e-13 ***
## LoanOriginationDate.year2010              < 2e-16 ***
## LoanOriginationDate.year2011             4.42e-08 ***
## LoanOriginationDate.year2012              < 2e-16 ***
```

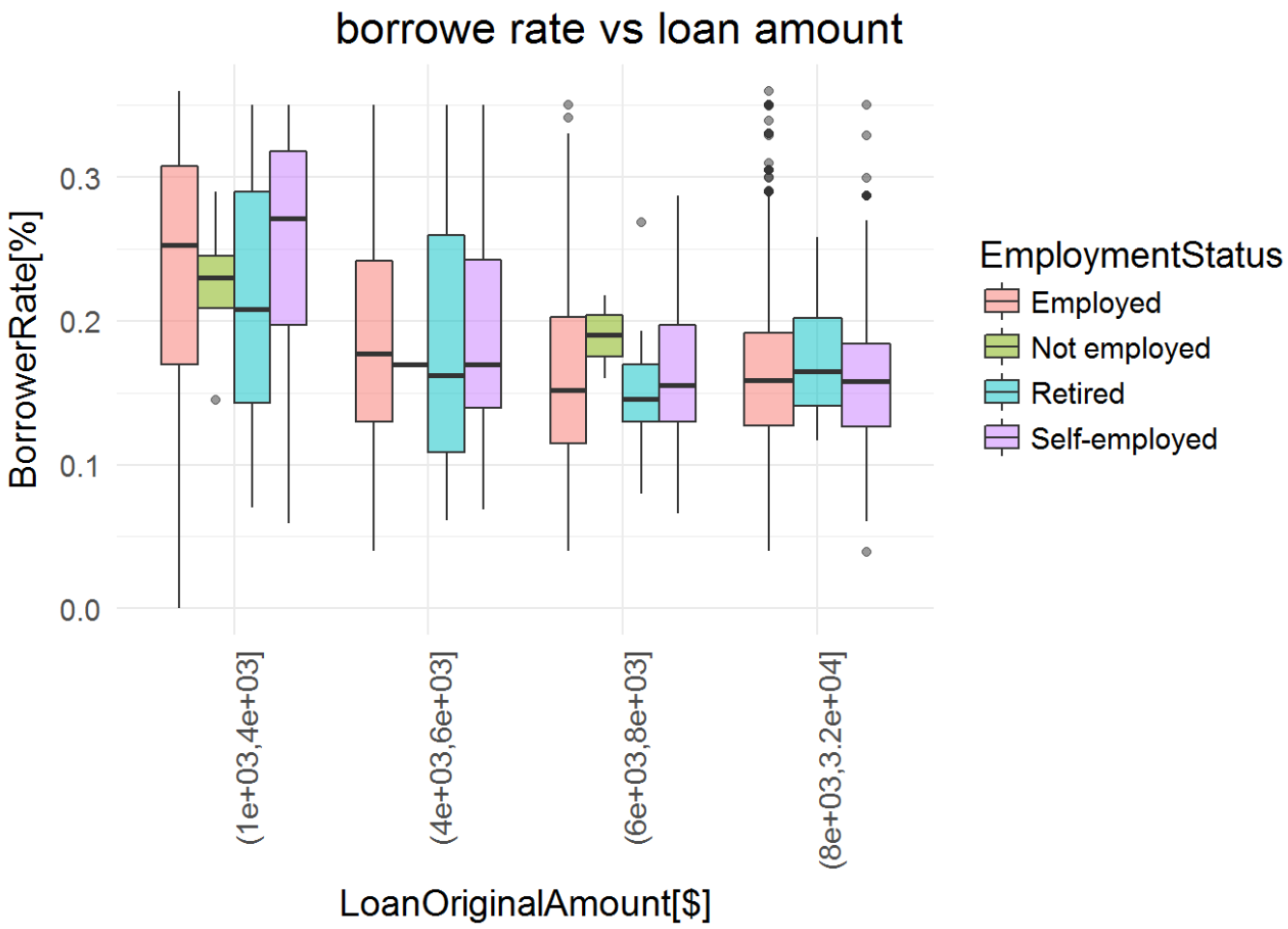
```
## LoanOriginationDate.year2013              < 2e-16 ***
## LoanOriginationDate.year2014              < 2e-16 ***
## DebtToIncomeRatio                         < 2e-16 ***
## OpenCreditLines                          0.089168 .
## LP_CustomerPrincipalPayments              < 2e-16 ***
## StatedMonthlyIncome                      0.425893
## LoanOriginalAmount                        < 2e-16 ***
## Term                                      < 2e-16 ***
## EmploymentStatusNot employed             0.601347
## EmploymentStatusRetired                  0.612407
## EmploymentStatusSelf-employed            0.059603 .
## AmountDelinquent                         0.002206 **
## bucket.InquiriesLast6Months(1.44,2]      0.005596 **
## bucket.InquiriesLast6Months(2,105]       0.022428 *
## BankcardUtilization                      0.511864
## IncomeRange$100,000+                     0.198461
## IncomeRange$25,000-49,999                0.000865 ***
## IncomeRange$50,000-74,999                0.000549 ***
## IncomeRange$75,000-99,999                0.038747 *
## average_credit_score                      < 2e-16 ***
## bucket.LoanOriginalAmount(4e+03,6e+03]    < 2e-16 ***
## bucket.LoanOriginalAmount(6e+03,8e+03]    < 2e-16 ***
## bucket.LoanOriginalAmount(8e+03,3.2e+04]  < 2e-16 ***
## average_credit_score.bucket(670,690]     7.92e-08 ***
## average_credit_score.bucket(690,730]      < 2e-16 ***
## average_credit_score.bucket(730,800]      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.12 on 30619 degrees of freedom
##   (30203 observations deleted due to missingness)
## Multiple R-squared:  0.9713, Adjusted R-squared:  0.9713
## F-statistic: 2.257e+04 on 46 and 30619 DF,  p-value: < 2.2e-16
```

we can see our model is really similar to automated model however we have not consider BankcardUtilization and EmploymentStatus which also are not statistically significant, also we can see variables like open credits have a negative effect of -8.332 which kinda seems off to me as i assume having more credit lines will increase your monthly payment as it shows you are frequenct a borrower.

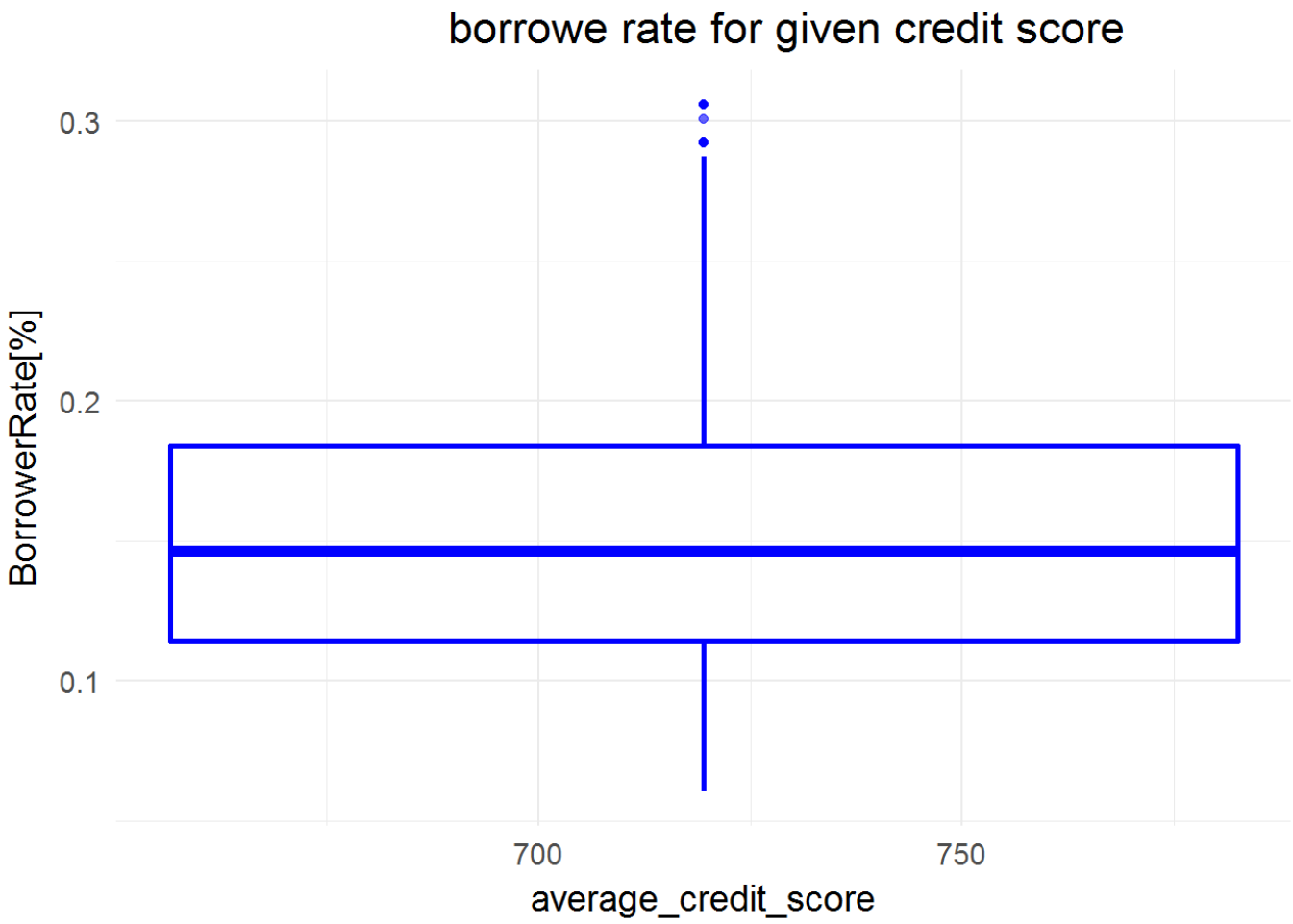
Final Plots and Summary

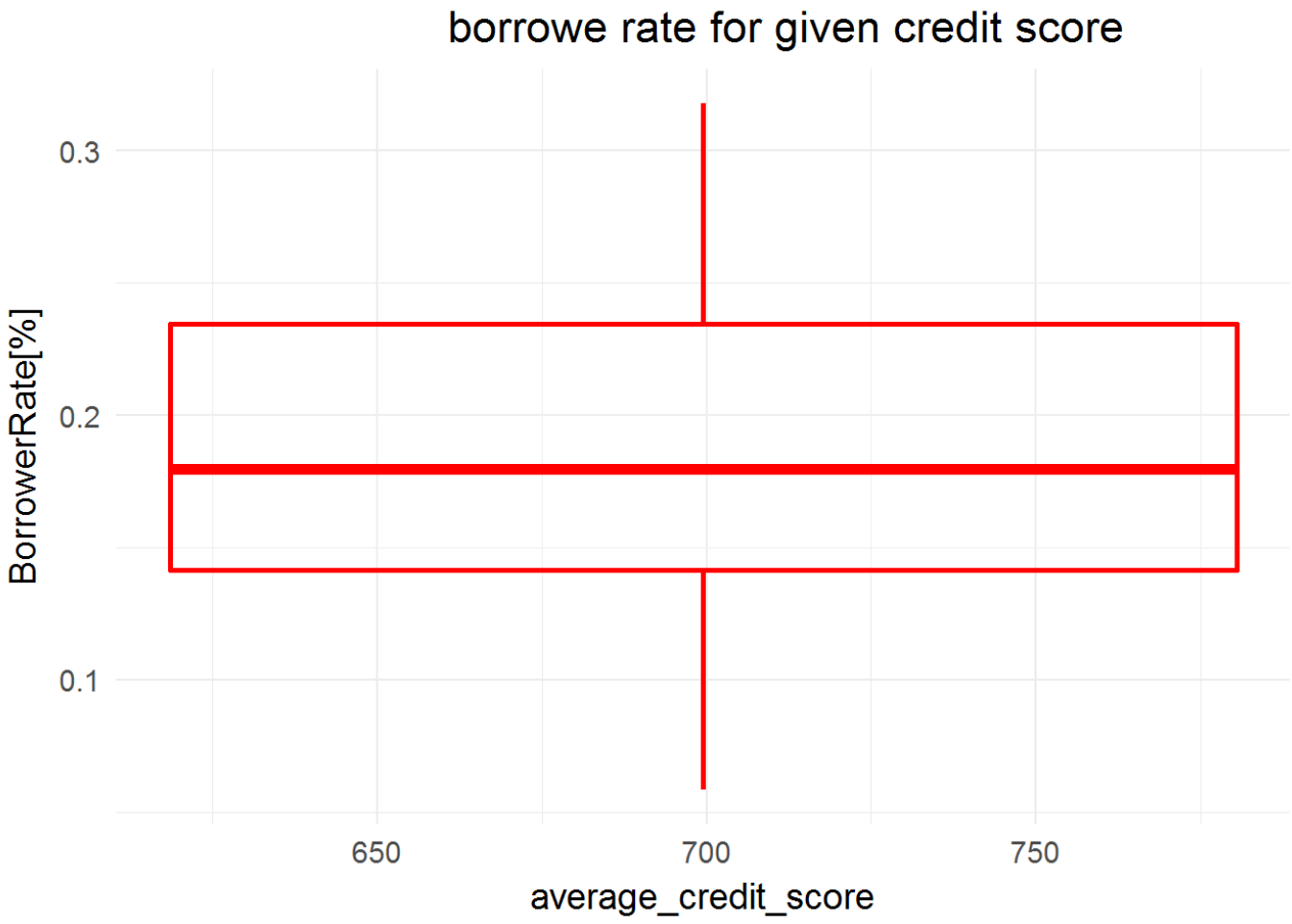
Plot One and Two

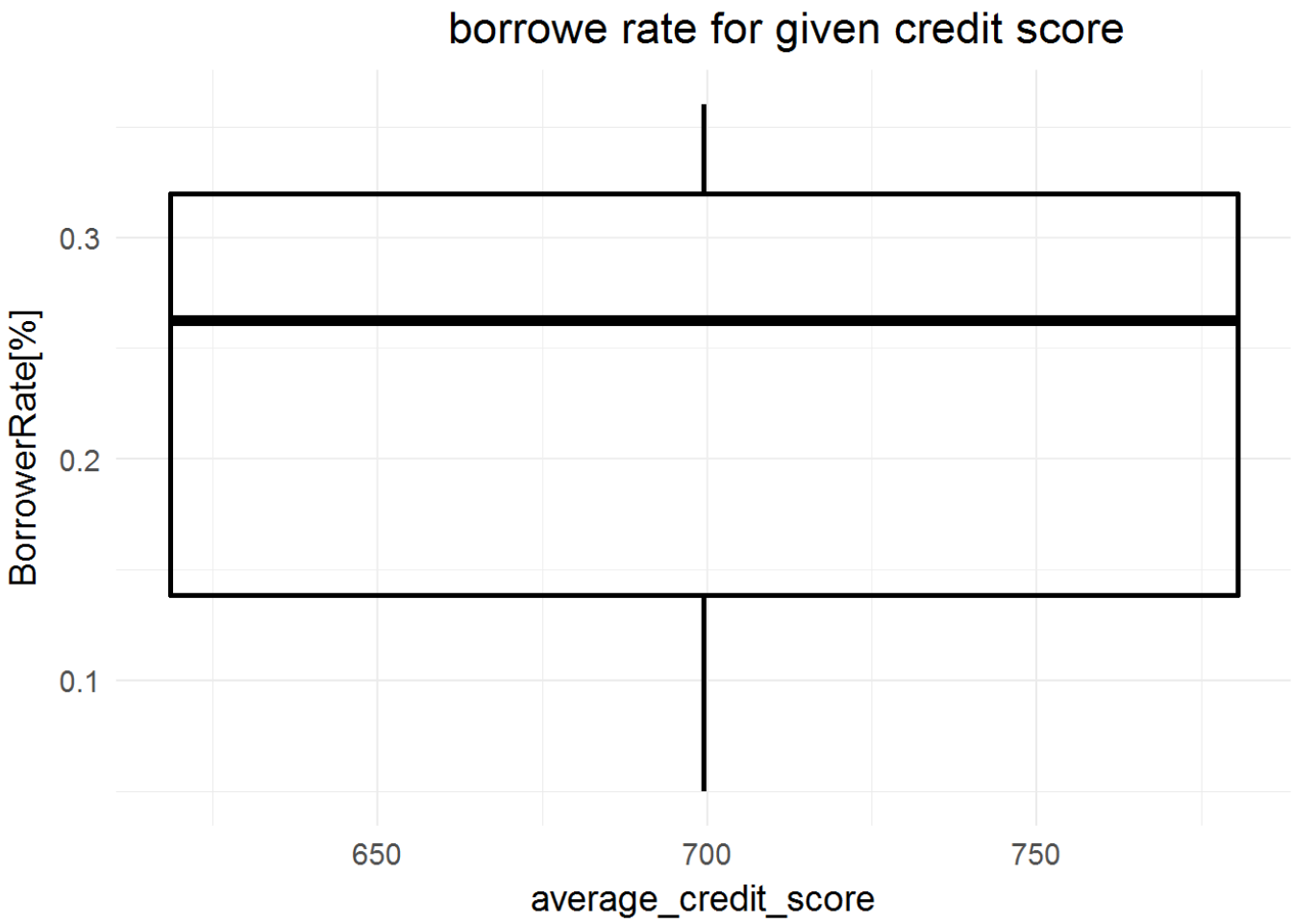
we have seen Loan Origin amount has siginificant correlation with the borrower rate, i have used employment status also to look deeper into the data and see the patterns, which we have seen for employment status and different loan amount we have differnet borrower rates.



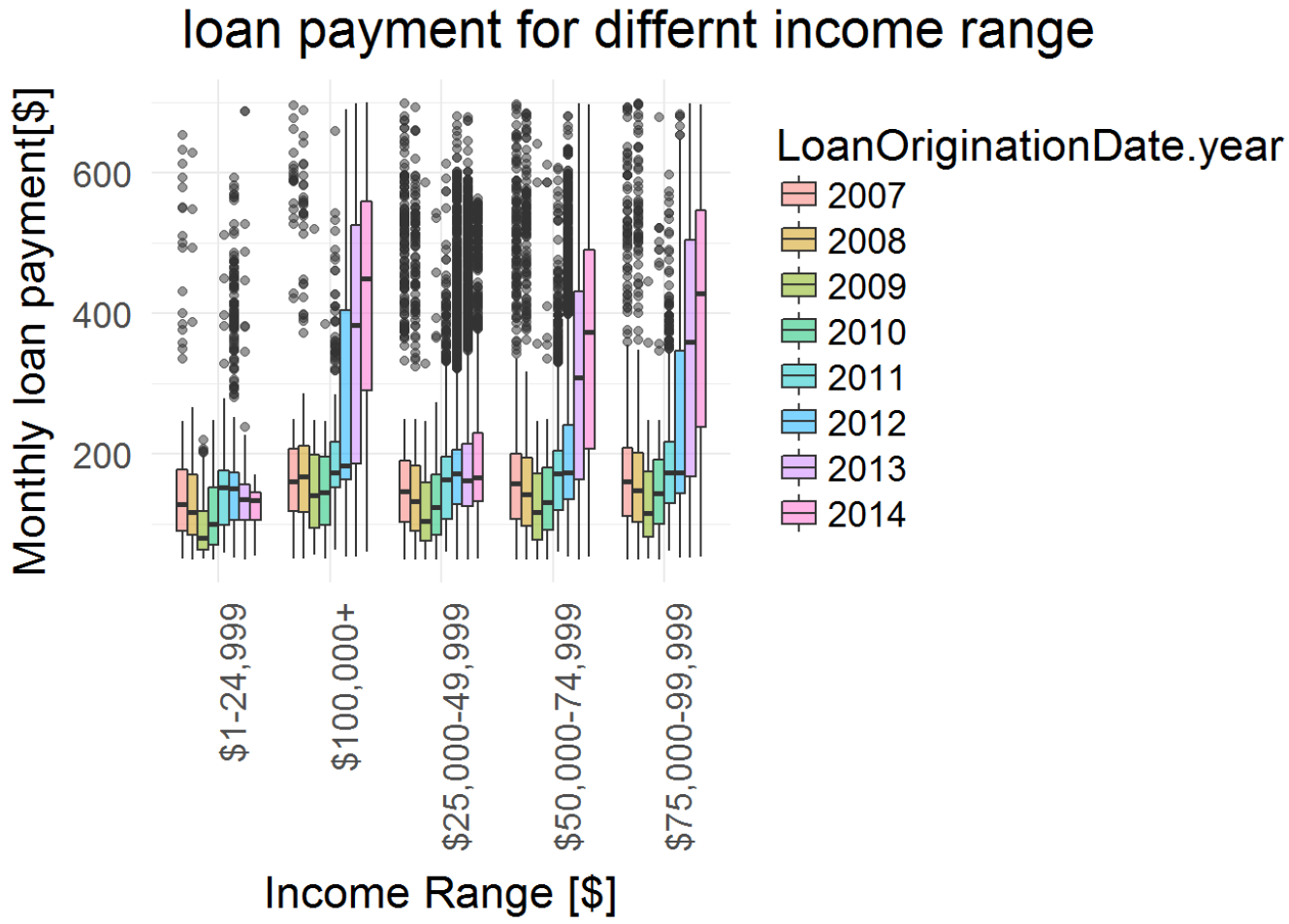
in last plot we can see within years the average credit score has an intresting impact to borrower rate as higher credit score will end in lower borrower rate , i have ploted for 2010 , 2013 and 2014 , we can see borrower rate based on credit score has decrease in recent years


borrowe rate for given credit score
BorrowerRate[%]
0.3
0.2
0.1
700
750
average_credit_score


borrowe rate for given credit score
BorrowerRate[%]
0.3
0.2
0.1
650
700
750
average_credit_score



i have found withing last few years(2011-2014) being in higher income range class will cause to pay more payment monthly.



Reflections

Issues

Are the following conclusions certain? No. There seems we need more data wrangling and cleaning on this data set.Also we need to have more records with usable values. Regarding this project I struggled with finding the relationship between the variables initially. It was hard for me to decide which variables are dependent, which are independent and which I should keep in my analysis.Through exploring, I found a borrower rate as my dependent variable and I designed my project around that.

Conclusion

My conclusions regarding this project are the following; The likelihood of having low/high Borrow rate depends on variables such as credit score, amount of the loan, customer payment and other variables which model showed us, but I did not consider them in my conclusion.

It seems that credit score has more impact on borrow rate within last few years, after 2010 we can see that having higher credit score gives us lower rates.

Intrestingly I found that having higher loan amount seems to decrease the borrow rate, however we need more data to approve this hypothesis.