Variables that influence Credit Score

Occupation

4 CreditGrade 0

18 BorrowerState 0

19 Occupation 0

20 EmploymentStatus 0

22 IsBorrowerHomeowner 0

25 DateCreditPulled 0

28 FirstRecordedCreditLine 0

32 OpenRevolvingAccounts 0

33 OpenRevolvingMonthlyPayment 0

48 IncomeRange 0

49 IncomeVerifiable 0

50 StatedMonthlyIncome 0

68 MonthlyLoanPayment 0

77 PercentFunded 0

87 IncomeLevel 0

8 BorrowerAPR 7

26 CreditScoreRangeLower 113

27 CreditScoreRangeUpper 113

88 MeanCreditScore 113

31 TotalCreditLinespast7years 130

34 InquiriesLast6Months 130

36 CurrentDelinquencies 130

39 PublicRecordsLast10Years 130

83 FirstRecordedCreditYear 130

38 DelinquenciesLast7Years 192

35 TotalInquiries 218

43 AvailableBankcardCredit 1383

44 TotalTrades 1383

45 TradesNeverDelinquent..percentage. 1383

46 TradesOpenedLast6Months 1383

29 CurrentCreditLines 1396

30 OpenCreditLines 1396

40 PublicRecordsLast12Months 1396

41 RevolvingCreditBalance 1396

21 EmploymentStatusDuration 1400

37 AmountDelinquent 1400

47 DebtToIncomeRatio 1537

dfs (People who have been assigned a loan)

dfs\_1 (People who haven’t been assigned a CreditGrade)

dfs\_2 (People who have a CreditGrade and have a CreditScore)

People have been assigned a creditgrade even though they don’t have a creditscore. About 113 of them. First lets remove those people.

There are people who have a CreditScore but they don’t have any information or don’t have any OpenCreditLines.

People with “Unknown” CreditGrade do have CreditScore information available and can in principle be graded into creditGrades.

Look at the variables and see which have high NA counts and have less information Some are keys and ID’s and don’t convey much.

There are also a huge number of people who don’t have an assigned CreditGrade.

VariableList

Variable na\_count

1 ListingKey 0

2 ListingNumber 0

3 ListingCreationDate 0

4 CreditGrade 0

5 Term 0

6 LoanStatus 0

7 ClosedDate 0

9 BorrowerRate 0

10 LenderYield 0

15 ProsperRating..Alpha. 0

17 ListingCategory..numeric. 0

18 BorrowerState 0

19 Occupation 0

20 EmploymentStatus 0

22 IsBorrowerHomeowner 0

23 CurrentlyInGroup 0

24 GroupKey 0

25 DateCreditPulled 0

28 FirstRecordedCreditLine 0

32 OpenRevolvingAccounts 0

33 OpenRevolvingMonthlyPayment 0

48 IncomeRange 0

49 IncomeVerifiable 0

50 StatedMonthlyIncome 0

51 LoanKey 0

60 LoanCurrentDaysDelinquent 0

62 LoanMonthsSinceOrigination 0

63 LoanNumber 0

64 LoanOriginalAmount 0

65 LoanOriginationDate 0

66 LoanOriginationQuarter 0

67 MemberKey 0

68 MonthlyLoanPayment 0

69 LP\_CustomerPayments 0

70 LP\_CustomerPrincipalPayments 0

71 LP\_InterestandFees 0

72 LP\_ServiceFees 0

73 LP\_CollectionFees 0

74 LP\_GrossPrincipalLoss 0

75 LP\_NetPrincipalLoss 0

76 LP\_NonPrincipalRecoverypayments 0

77 PercentFunded 0

78 Recommendations 0

79 InvestmentFromFriendsCount 0

80 InvestmentFromFriendsAmount 0

81 Investors 0

82 Delinquent 0

84 ListCreationYear 0

85 LoanCreationYear 0

86 CreditPullYear 0

87 IncomeLevel 0

8 BorrowerAPR 7

26 CreditScoreRangeLower 113

27 CreditScoreRangeUpper 113

88 MeanCreditScore 113

31 TotalCreditLinespast7years 130

34 InquiriesLast6Months 130

36 CurrentDelinquencies 130

39 PublicRecordsLast10Years 130

83 FirstRecordedCreditYear 130

38 DelinquenciesLast7Years 192

35 TotalInquiries 218

43 AvailableBankcardCredit 1383

44 TotalTrades 1383

45 TradesNeverDelinquent..percentage. 1383

46 TradesOpenedLast6Months 1383

29 CurrentCreditLines 1396

30 OpenCreditLines 1396

40 PublicRecordsLast12Months 1396

41 RevolvingCreditBalance 1396

42 BankcardUtilization 1396

21 EmploymentStatusDuration 1400

37 AmountDelinquent 1400

47 DebtToIncomeRatio 1537

11 EstimatedEffectiveYield 5058

12 EstimatedLoss 5058

13 EstimatedReturn 5058

14 ProsperRating..numeric. 5058

16 ProsperScore 5058

52 TotalProsperLoans 16116

53 TotalProsperPaymentsBilled 16116

54 OnTimeProsperPayments 16116

55 ProsperPaymentsLessThanOneMonthLate 16116

56 ProsperPaymentsOneMonthPlusLate 16116

57 ProsperPrincipalBorrowed 16116

58 ProsperPrincipalOutstanding 16116

59 ScorexChangeAtTimeOfListing 16627

61 LoanFirstDefaultedCycleNumber 17033

The ProsperRating..Alpha is a character variable and is blank when the information is not availale. It shows up as “NA” in ProsperRating..Numeric.

Since ProsperLoanData had lot of NA’s I decided to leave them out of the study in my visualization, though in case of more detailed studies or machine learning studies to decide on an APR, those variables are no doubt useful.

Try a ScatterPlotMatrix – Really slow though

Following are the LoanStatus :

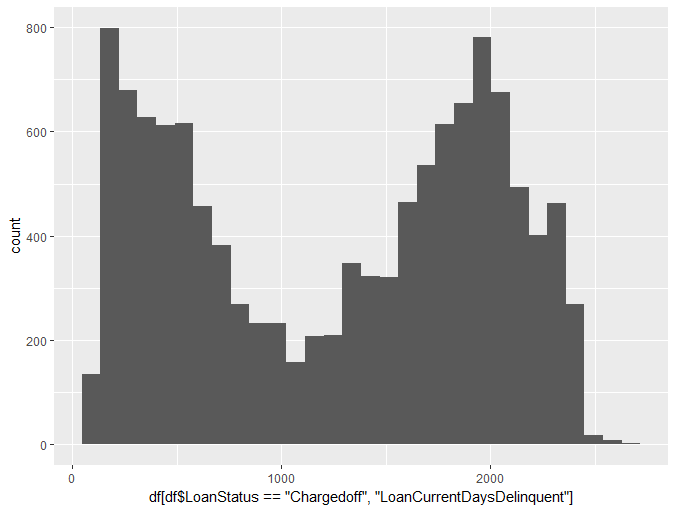
[1] Completed Current Past Due (1-15 days)

[4] Defaulted Chargedoff Past Due (16-30 days)

[7] Cancelled Past Due (61-90 days) Past Due (31-60 days)

[10] Past Due (91-120 days) FinalPaymentInProgress Past Due (>120 days)

12 Levels: Cancelled Chargedoff Completed Current Defaulted ... Past Due (91-120 days)



Cancelled Chargedoff Completed

5 11992 38074

Current Defaulted FinalPaymentInProgress

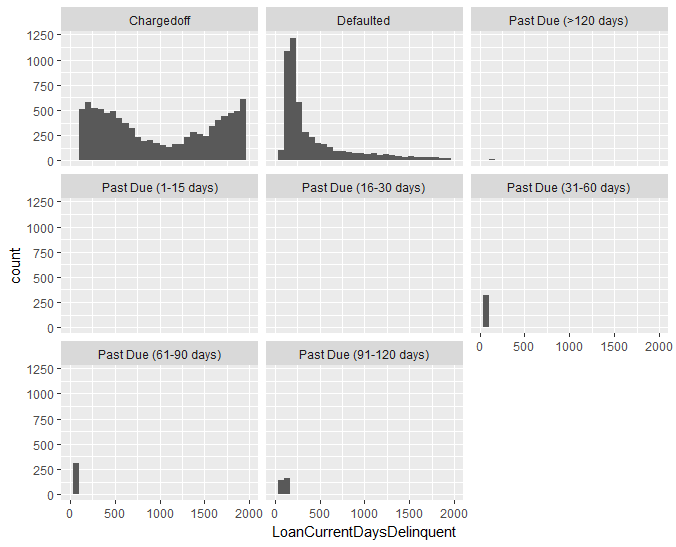
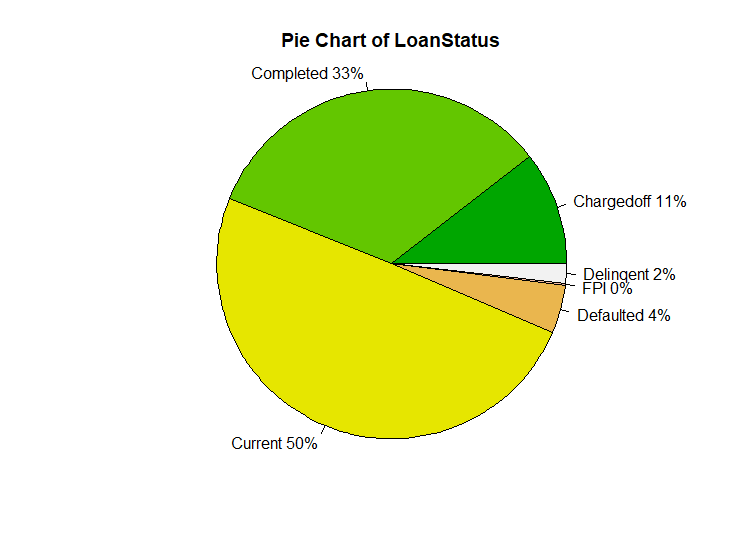
56576 5018 205

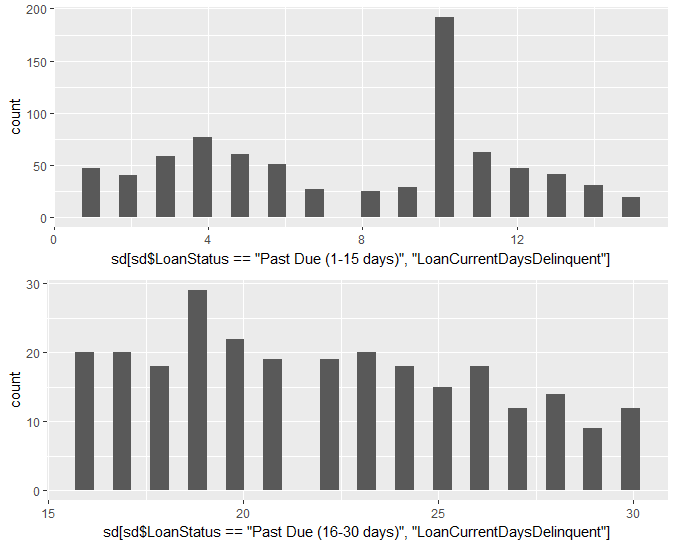
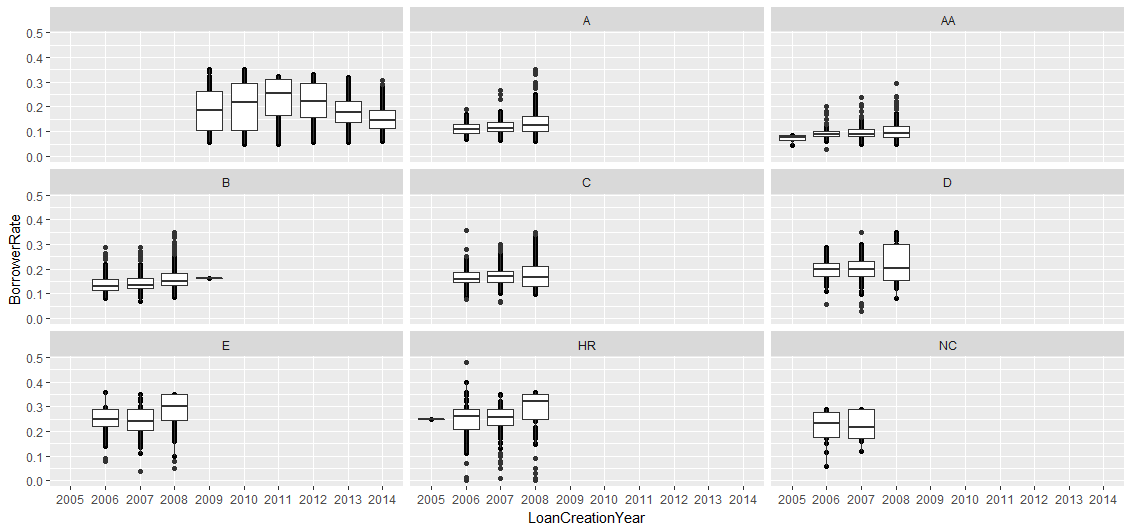
Past Due (>120 days) Past Due (1-15 days) Past Due (16-30 days)

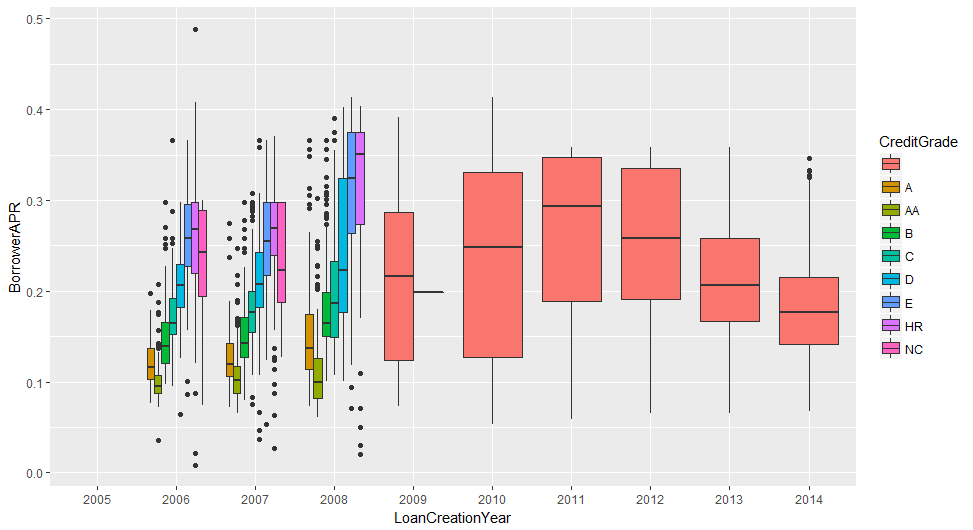
16 806 265

Past Due (31-60 days) Past Due (61-90 days) Past Due (91-120 days)

363 313 304

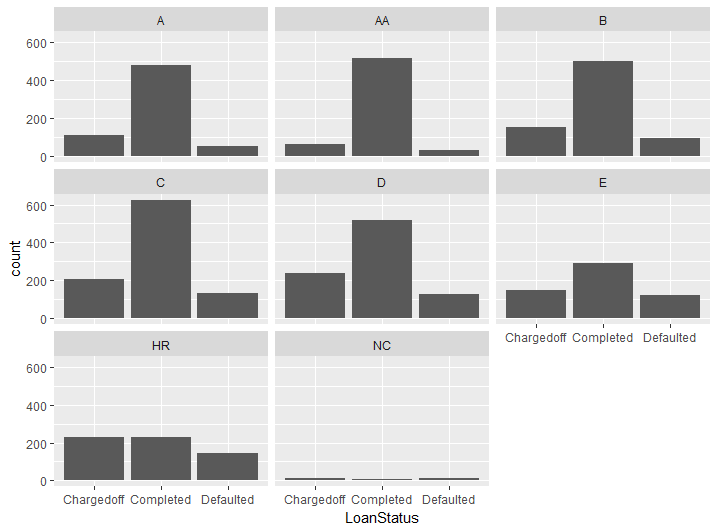


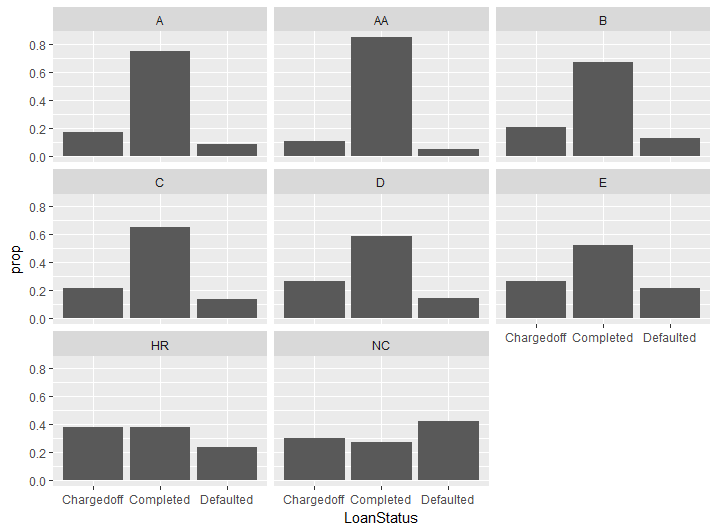




I find this unexpected that Credit Grades, A – NC were approved for loans between 2006 – 2008, but from 2009 and above the approvals are done for people without credit or the credit information was not used, as the CreditGrade is blank for those years. Also Its seen that the APR increases as the credit grade worsens. The mean APR rate was highest in 2011, while it went down considerably in 2014.

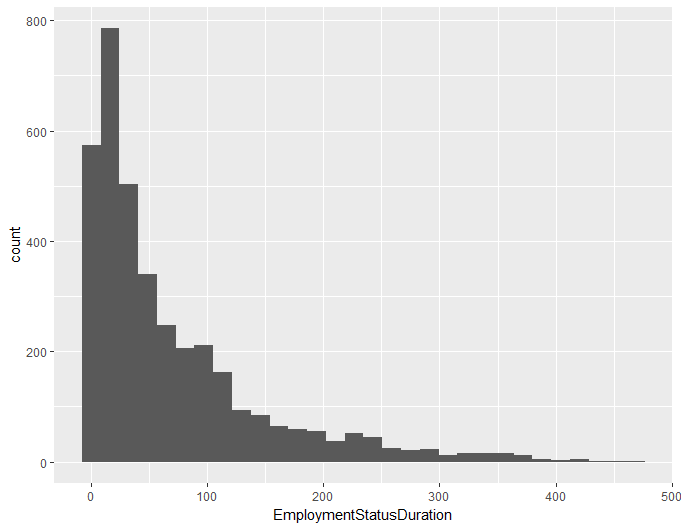
Following Distributions distributions of Loan Status for the different CreditGrades. And below it in proportions.

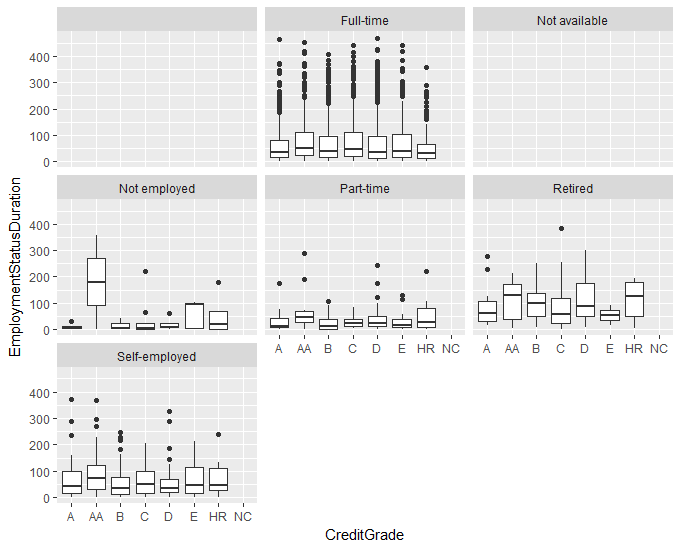


As seen here for the bad Credit Grades HR, and NC, the proportions of Loans Chargedoff vs completed continue to rise, becoming almost equal for Charged off, defaulted, and completed. Hence, there is a clear indication that the people with higher CreditGrades are less likely to default on their loans.

Now lets look at what factors lead to a high or low Credit Grade

Do a 2D plot of all factors vs. Credit Grade first. Is the

 Is the Employment status measured in days or months? It varies from 0 – 400



ScatterPlot of EmpoymentStatusDuration and Monthly – Study this (2/25/2018)

Final Plot1:

The mean Borrower APR becomes higher as the CreditGrade worsens, but there are outliers and some people with low credit still have low APR's and there is some overlap of APR's in each category. For the overlaps,factors other than the CreditGrade contributed to lowering the APR.

Also, APR’s were lowest in 2006, 90% for 2006-2007 the APR is very similar, but it is higher for 2008. From 2009 onwards, we don't have CreditGrades.

From Wiki:

“… the bottom and top of the box are always the 25th and 75th [percentile](https://en.wikipedia.org/wiki/Percentile) (the lower and upper [quartiles](https://en.wikipedia.org/wiki/Quartile), respectively), and the band near the middle of the box is always the 50th [percentile](https://en.wikipedia.org/wiki/Percentile) (the [median](https://en.wikipedia.org/wiki/Median)). But the ends of the whiskers can represent several possible alternative values…”

In R’s default boxplot{graphics} code,

upper whisker = min(max(x), Q\_3 + 1.5 \* IQR)   
lower whisker = max(min(x), Q\_1 – 1.5 \* IQR)

where IQR = Q\_3 – Q\_1, the box length.  
So the upper whisker is located at the \*smaller\* of the maximum x value and Q\_3 + 1.5 IQR,   
whereas the lower whisker is located at the \*larger\* of the smallest x value and Q\_1 – 1.5 IQR.

Here is a full discussion about the whisker of boxplot in default R code:  
http://r.789695.n4.nabble.com/Whiskers-on-the-default-boxplot-graphics-td2195503.html

Final Plot1:

The boxplot line in the middle is the median value. So one can see that the Borrower APR becomes higher as the CreditGrade worsens, but there are outliers (indicated by the points) which lie beyond the whiskers (above 95% C.L.). So, some people with low credit still have low APR's and there is some overlap of APR's in each category. For the overlaps,factors other than the CreditGrade contributed to lowering the APR. Also, interestingly, the APR’s are broader as they rise for different years, which means that there is more variability in the APR in the years after 2008. It could be because more factors were considered in deciding the APR, or there was a change in algorithm, which could have caused this trend.

One can see that among the different years, 2006 had the minimum APR’s, and it increased until 2009, 2010. APR’s then show a decreasing trend from 2011-2014. These happens for all creditgrade categories, except maybe HR and NC.

One can see that the C category lens towards the left and more so for E. So

The poorer creditgrades tend to have less available bankcredit as compared to the higher grades, while the bank card utilization is more for the poorer grades as compared to the higher grades. Bank Card utilization as it tends more towards the 1.0 value (100% utilization) while the

Higher grades tend towards 0. For thr Delinquent, Defaulted, Chargedoff loans, the higher credit grad distribution begins to look similar to the poorer grades.

<https://www.r-bloggers.com/whisker-of-boxplot/>

Notes on Resubmission:

* Implemented 80 words cut off for each Line
* Added a section for Variable choice and variables used.
* Removed a reduntant “BorrowerRate” Plot in Chunk 12
* use different color palette for BorrowerAPR and LoanStatus bar charts. (Chunks 12 and 13)
* Fixed the title size on Non-zero delinquent days. (Chunk 18)
* Changed the plots in Chunk 19, and 20. See before and after resubmission. Combined plots instead of facet\_wrap
* Pie chart is improved and categories which contribute close to 0% are removed.
* Chunk 28 : A box plot is used instead of geom\_point for CreditScore vs CreditGrade
* Chunk40: Reduced the DebtToIncomeRatio plot. Plot geom\_denisty and fill with CreditGrade instead of facet\_wrap
* Chunk 47: Use violin plot for Income Level instead of geom\_count
* Chunk 48 : Fixed overlapping labels on Trades plots
* Chunk 52 : Final Plot : Changed labels to non-scientific notation for AvailableBankCardCredit

theme(axis.text.x=element\_text(angle=60,vjust=0.8))

I am not just a cropped up photoshopped face. I am the whole “ME” 😊 ☹