Loan Borrower`s Case Study

Presented by:

Shayak Roy

This presentation deck contains the details of analysis of problem a Consumer Finance Company that wants to understand factors that causes borrowers to default.

Objective:

Use of **loan** data provided which has all information for all loans issued through the time period 2007 to 2011 and analyze the reasons behind a borrower **defaulting** the loan.

Input Data:

- Loan data information for borrowers who applied for loan including their details such as reason for loan, amount funded, their past records, verification status etc.
- This data has 4 years of information of borrowers categorized into 3, based on their loan paying status i.e. Fully Paid Off, Current and Defaulted.

Deliverables:

- Analysis through R programming to find the reasons for a borrower to Charge Off.
- R plots to visually analyze the affects and reason for Charged Off borrowers.
- PDF presentation depicting the important plots analyzing the data set.

Tools used:

- R- Studio for data cleaning and analysis
- Excel for data verification.

Data cleaning steps:

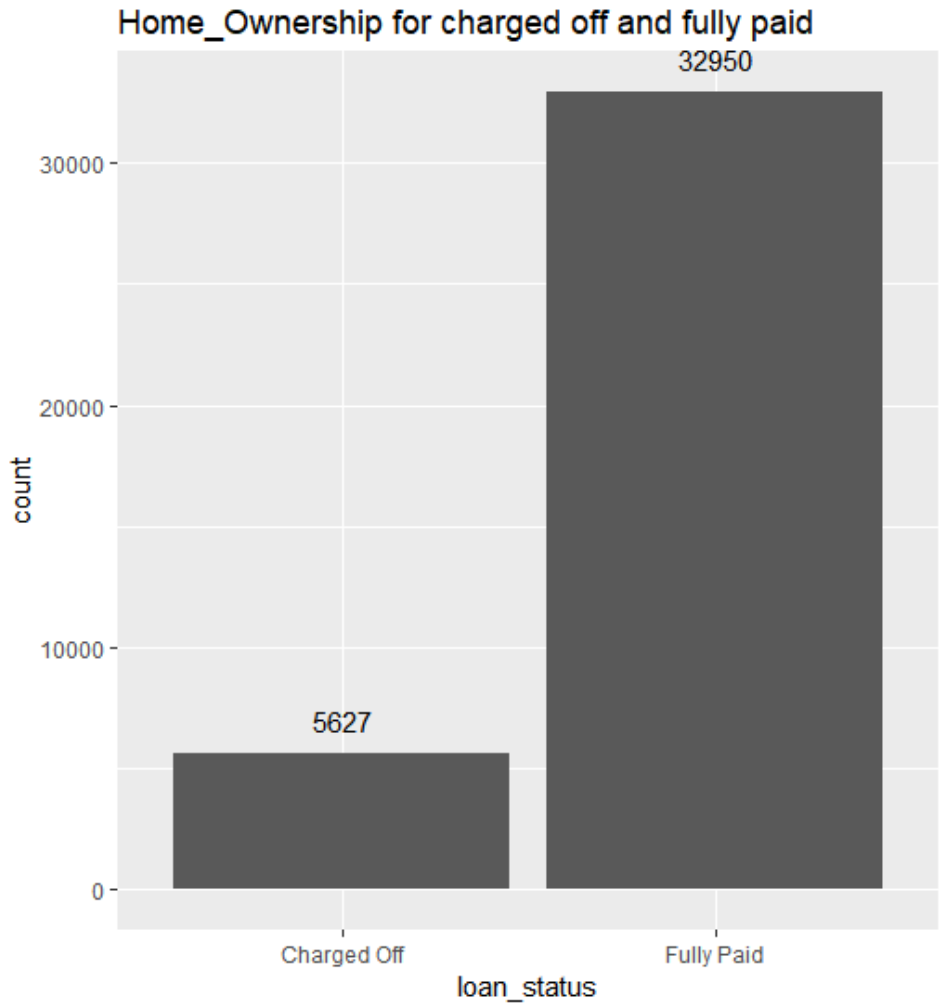
- Removing unwanted columns and hence reducing complexity of data set.
- Converting “n/a” by NA in columns where ever necessary.
- Removing special characters such as “%“ where ever necessary.
- Converting **issue_d** to a R standardized date format for better understanding, and later breaking it down to year month for analyzing.
- Converted few columns to factors for convenience during plotting.
- Created subset for only Charged Off borrowers as we need to focus more on them.

Assumptions:

- We are taking only 2 types, They are FULLY PAID and CHARGEDOFF in this analysis for some of the Plots.
- As the CURRENT Payer may pay or not pay the full loan. So neglecting it.
- The Months word has been removed from the term(i.e. 36 And 60).
- In the same group study, for some resources showing the **Issue_d** column of the loan dataset in the format of (MM-YY) AND for some others it is showing as (YY-MM).So, We are taking the format of MM-YY (i.e.DEC-11) and proceeding further.

Bivariate Analysis

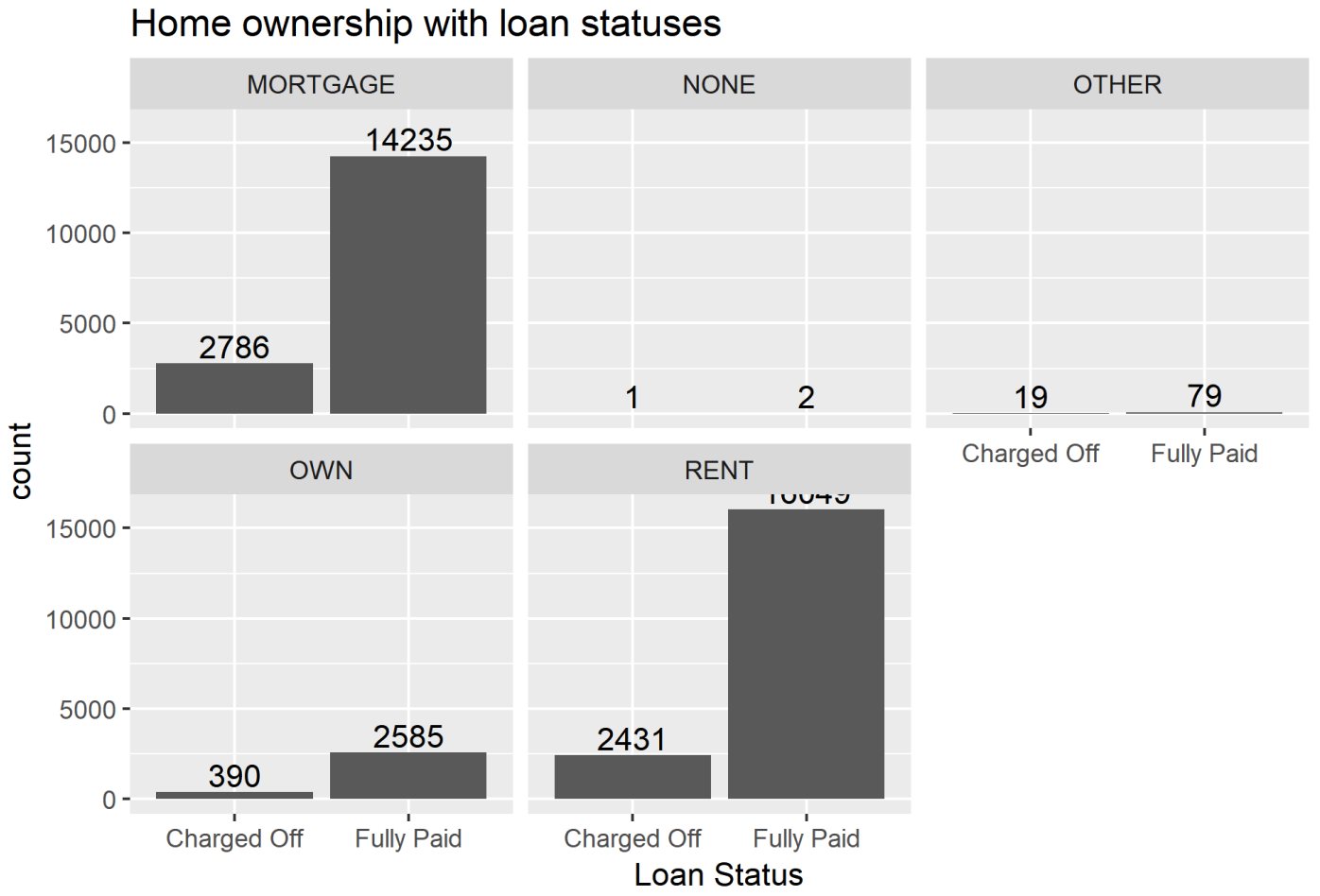
1. Overall Home_Ownership



- The **Home_Ownership** for charged off is 5627 while for Fully Paid is 32950, And as explained earlier That the CURRENT Payer may pay or not pay the full loan. So neglecting it.

Bivariate Analysis

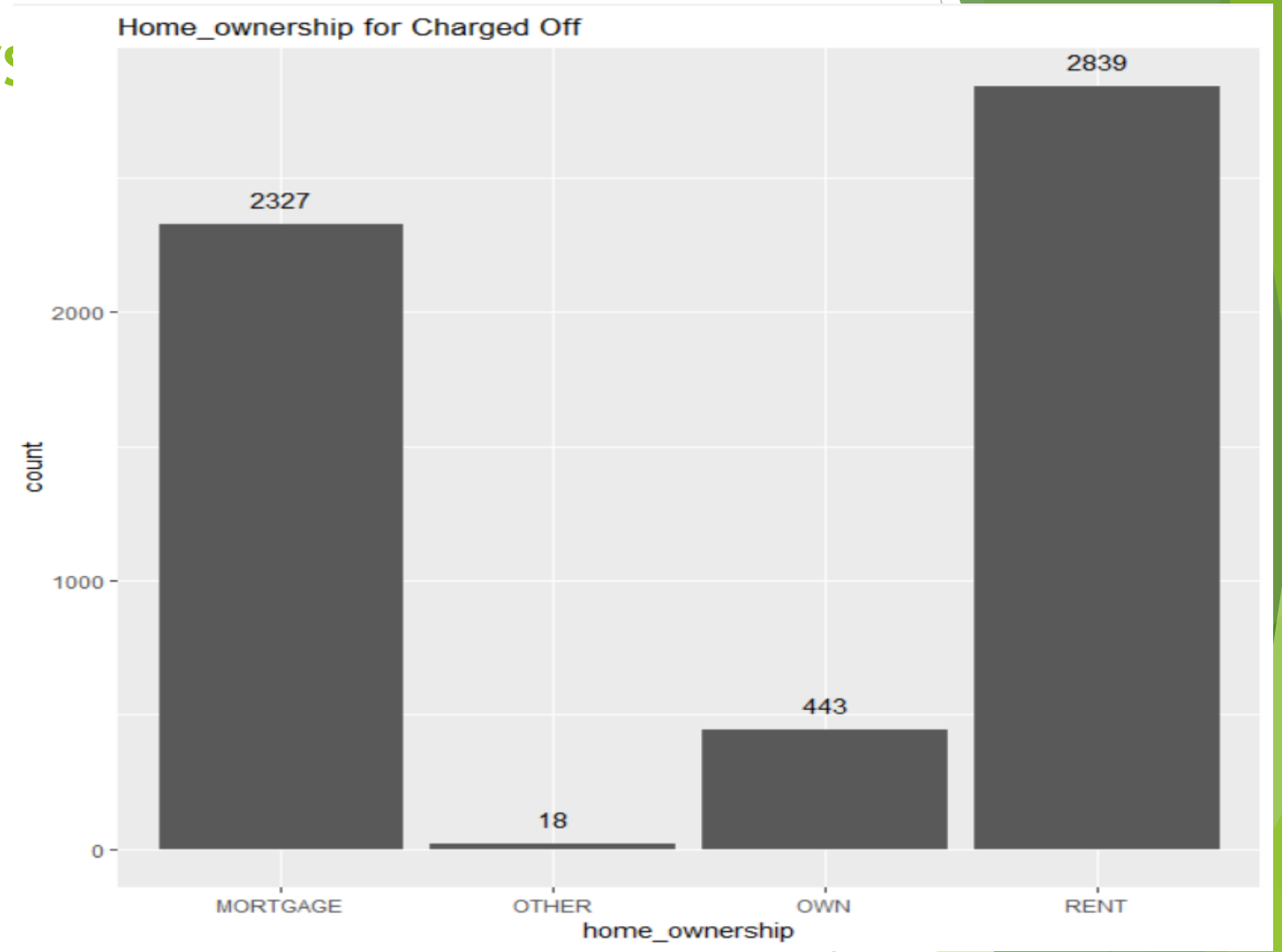
2. Home Ownership with Loan_Status I Charged Off and Fully paid



- The **Loan_Status** For Charged Off and Fully paid are more for Mortgage and Rent type of Home Ownership.

Univariate Analys

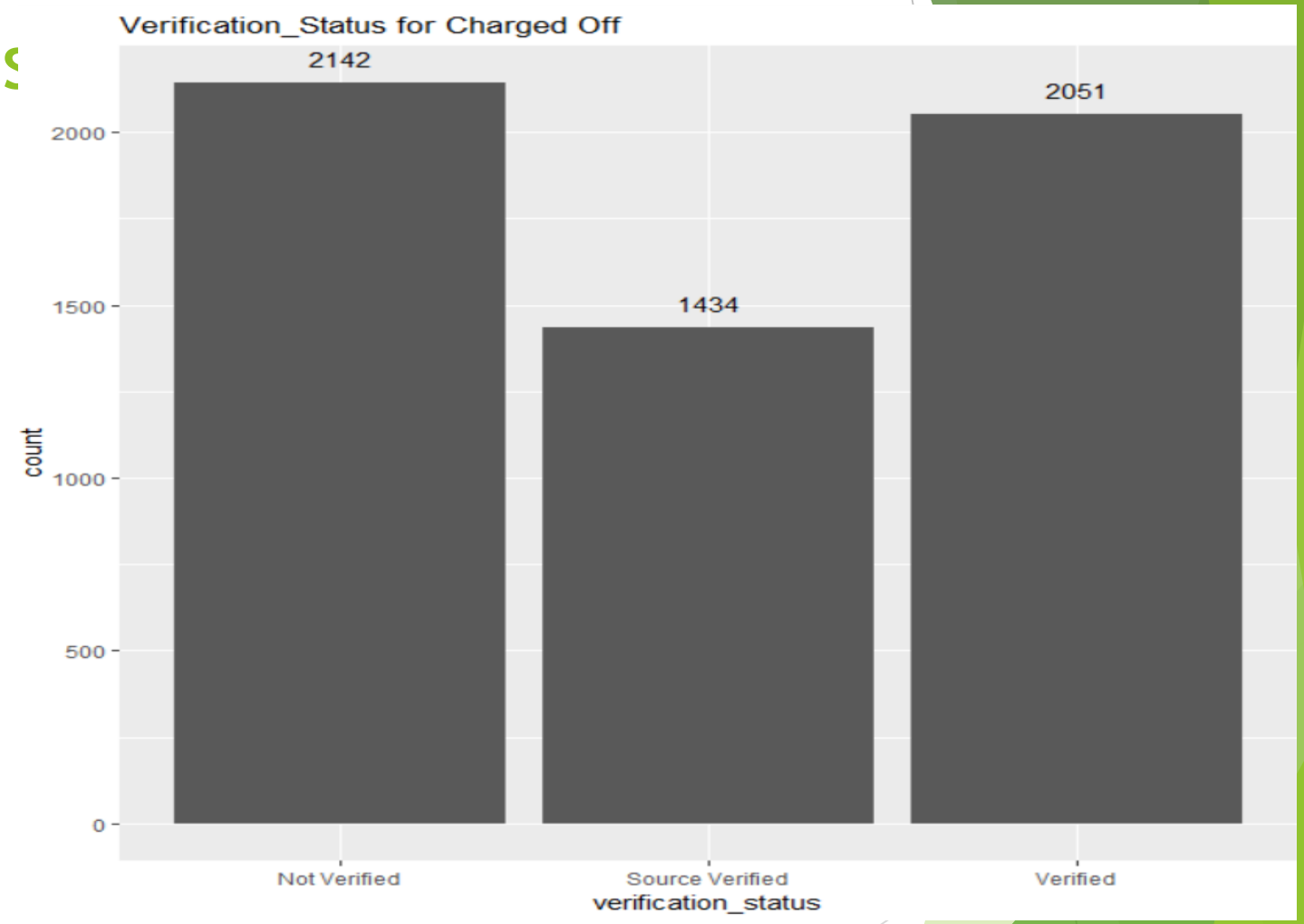
3. Charged Off-Home_Ownership



- Out of 39718, 5627 are charged off
- The Mortgage and Rent type of Home_Ownership has been more Charged off.

Univariate Analysis

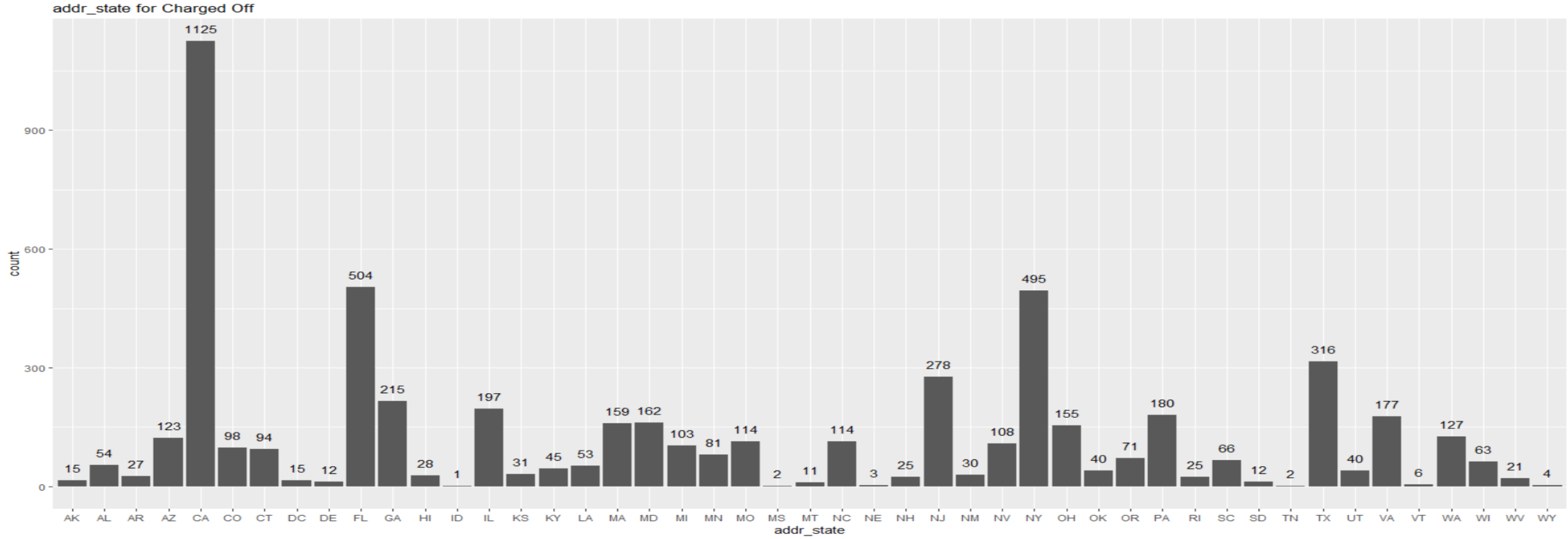
4. Charged Off – Verification Status



- The Verification Status for Charged off is more for “NOT VERIFIED”.
- This shows the Negligence of the Bank Superiors(With out proper verification) while providing the loan to the customers

Univariate Analysis

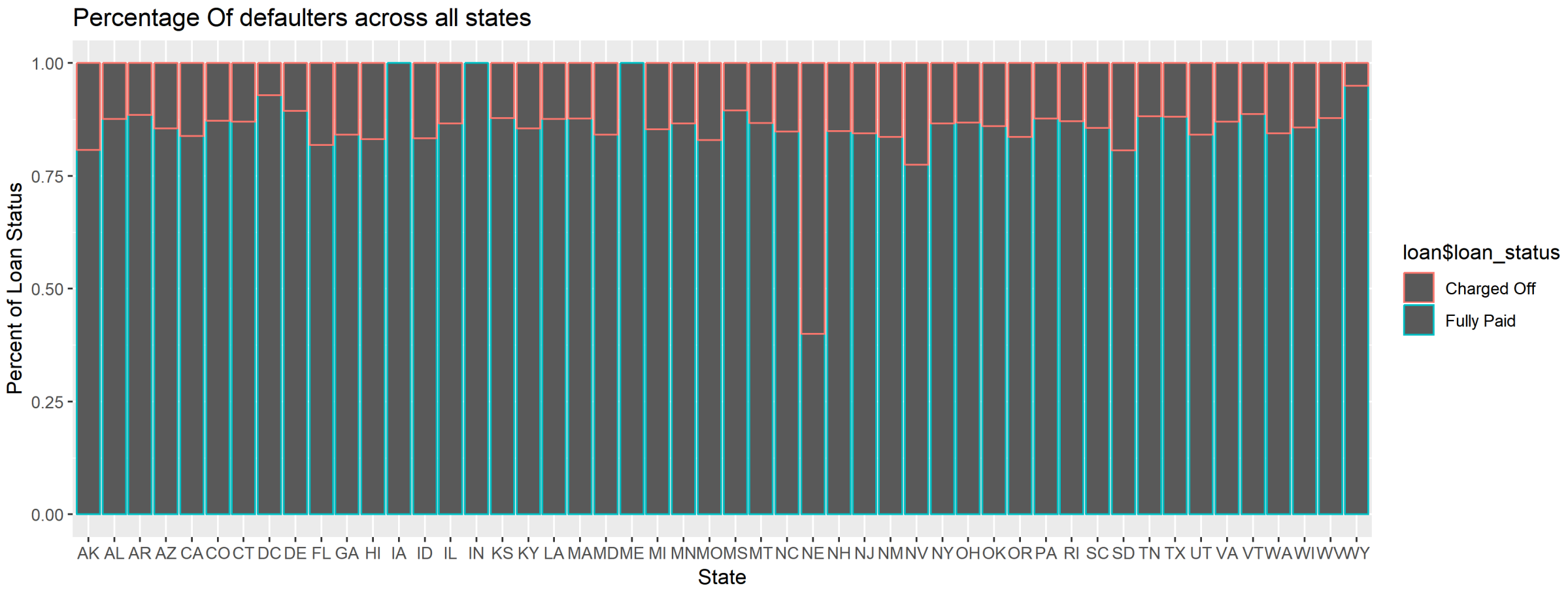
5.a Charged Off-STATE



- Based on above plot, we could see that there are high no of **Charged Off** is more in the CA State. But lets analyze it in terms of percentage in the below.

Bivariate Analysis

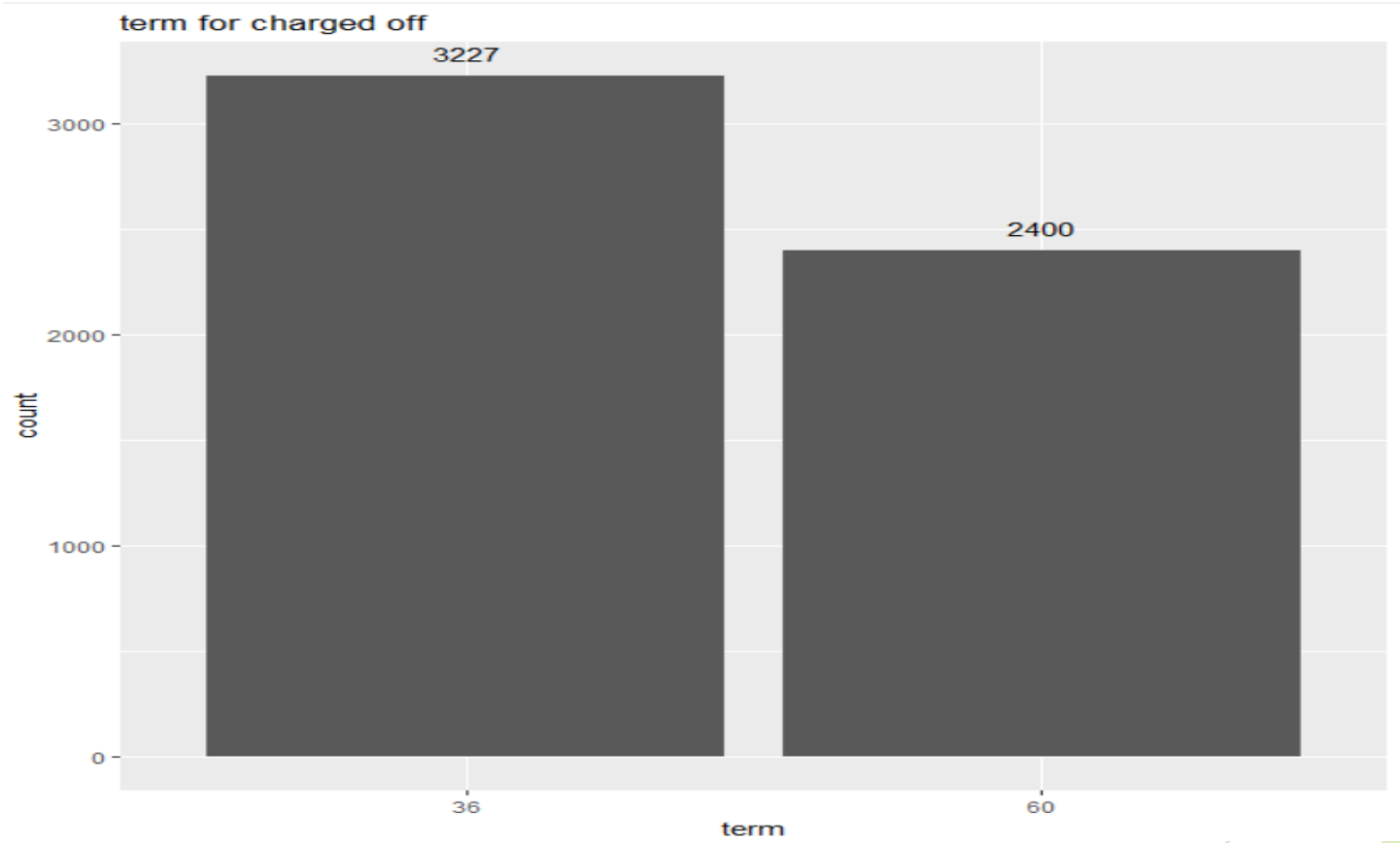
5.b Percentage of defaulters across all states



- The plot shows that NE has highest percentage of defaulters amongst all by excluding the Current Loan_Status.

Univariate Analysis

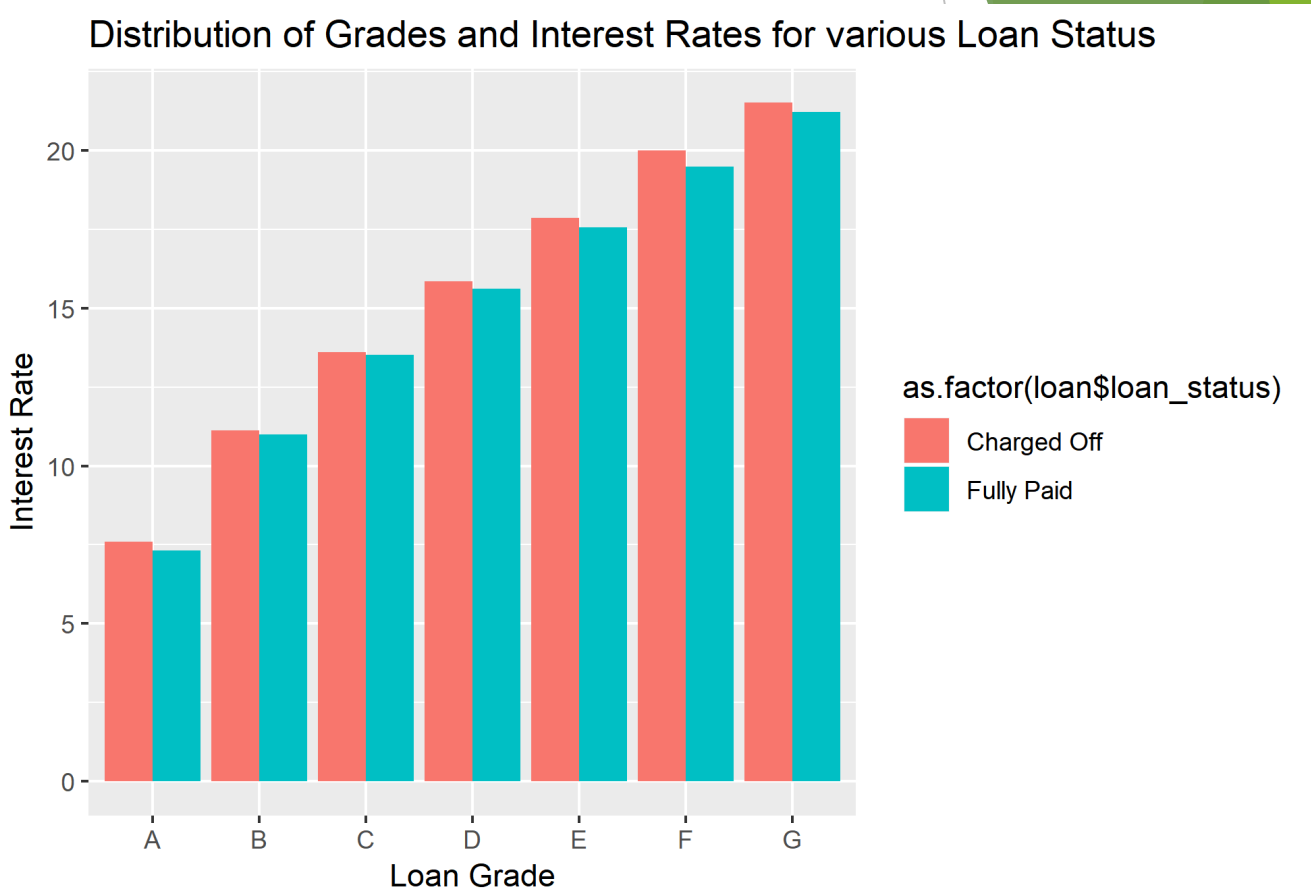
6. Plot of Charged Off borrowers based on term loans



- Based on above plot, we could see that there are high no of **Charged Off** borrowers who took term loan for 36 months.

Bivariate Analysis

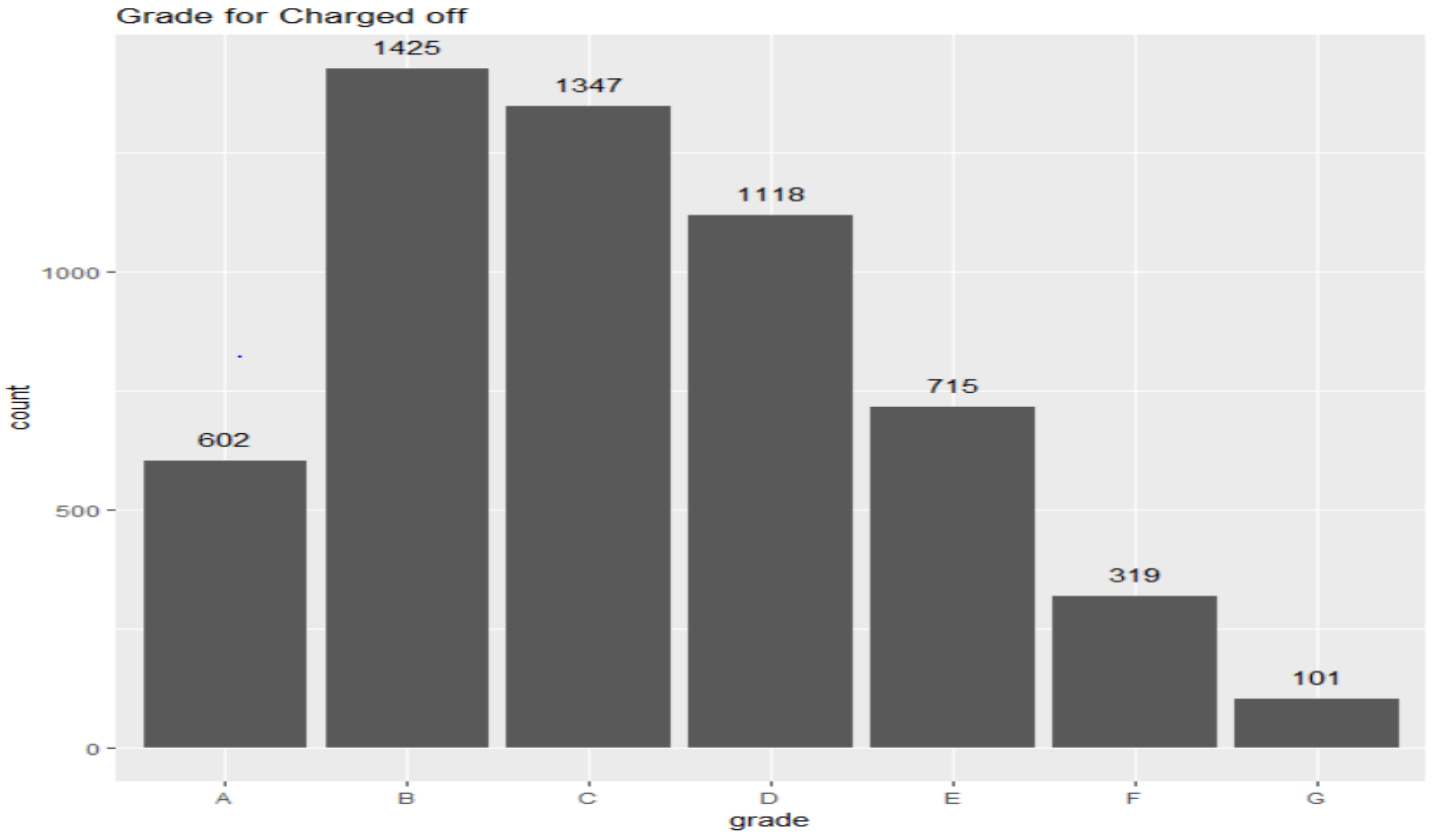
7. Grade Vs Interest Rates for Charged_off / Fully paid



- This analysis helps us to understand the interest rates at various grades. Also to identify the interest rate patterns

Univariate Analysis

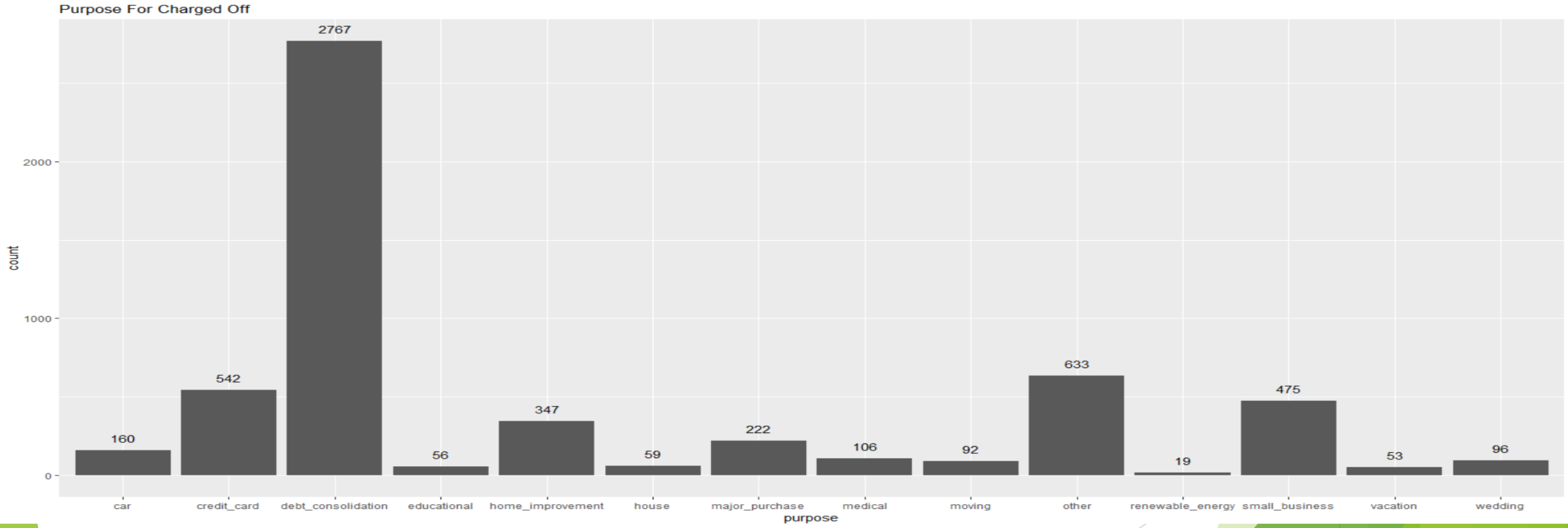
8. Charged off- - GRADE



- Based on this plot, we could analyze that charged Off borrowers are mostly graded between **B** and **D.**

Univariate Analysis

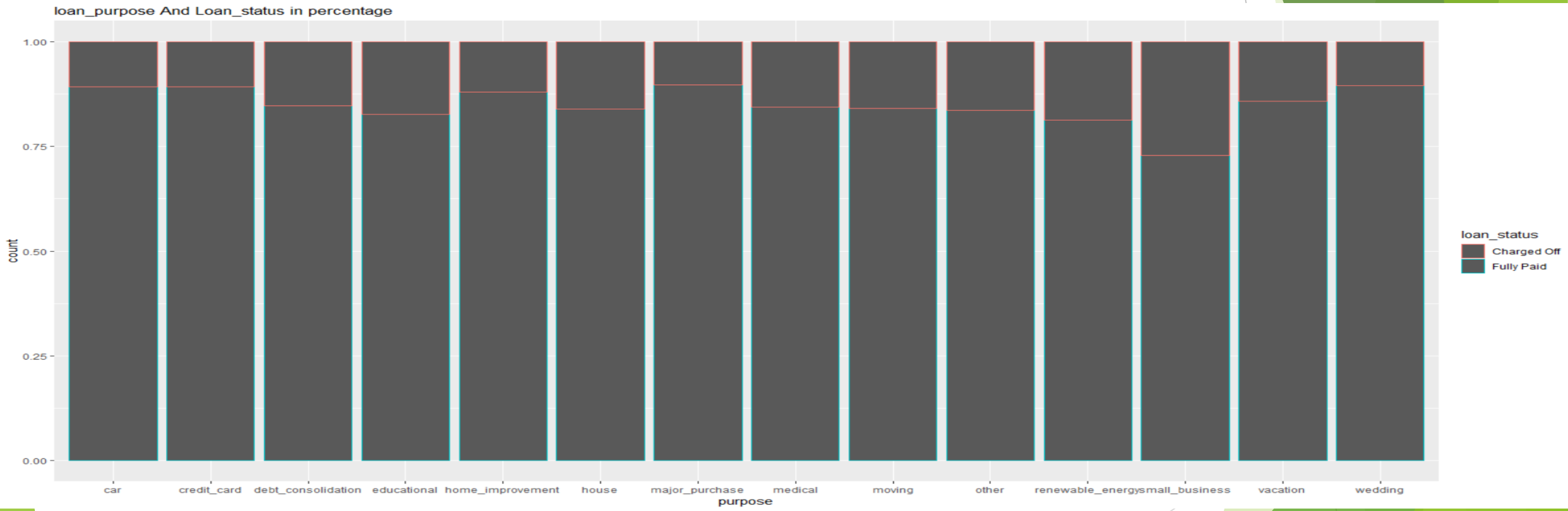
9. Charged Off - Purpose



- The Charged off happened more for the DEBT_CONSOLIDATION Purpose

Bivariate Analysis

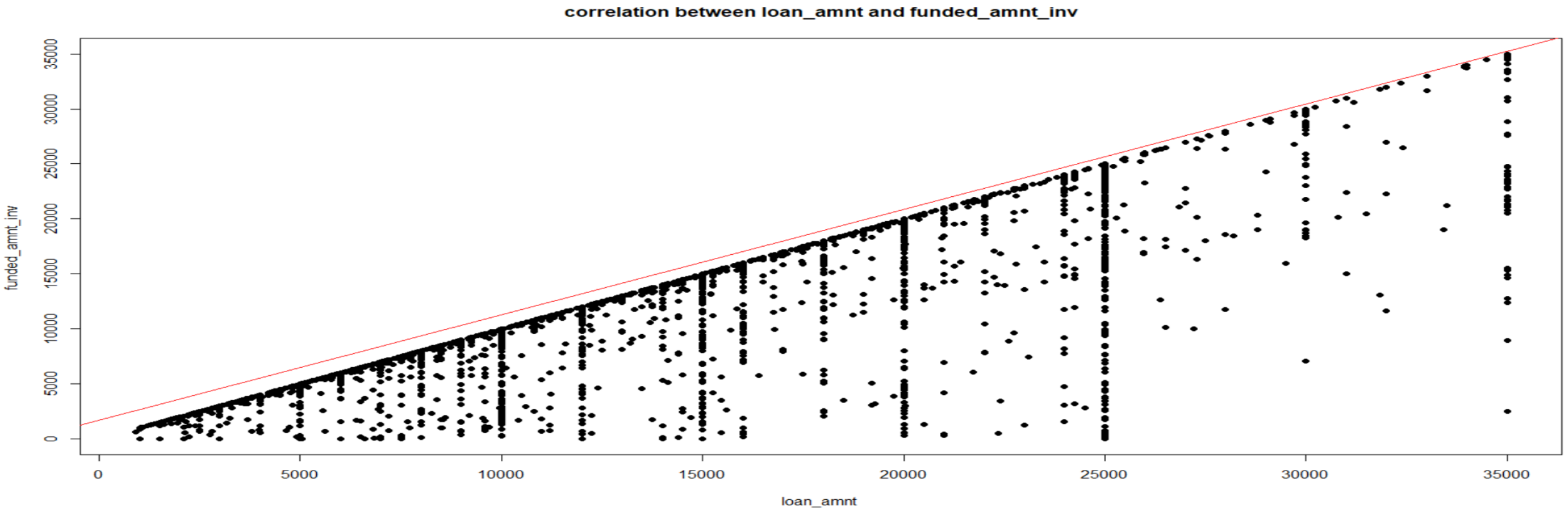
10. Loan_Status Vs Loan_Purpose In percentage



- For the Population (Considering charged off and Fully Paid)of loan we see that higher % of people with purpose as small business as charged off But we see that loan for Debt has higher %(around 2767 loans) among charged off.

Bivariate Analysis

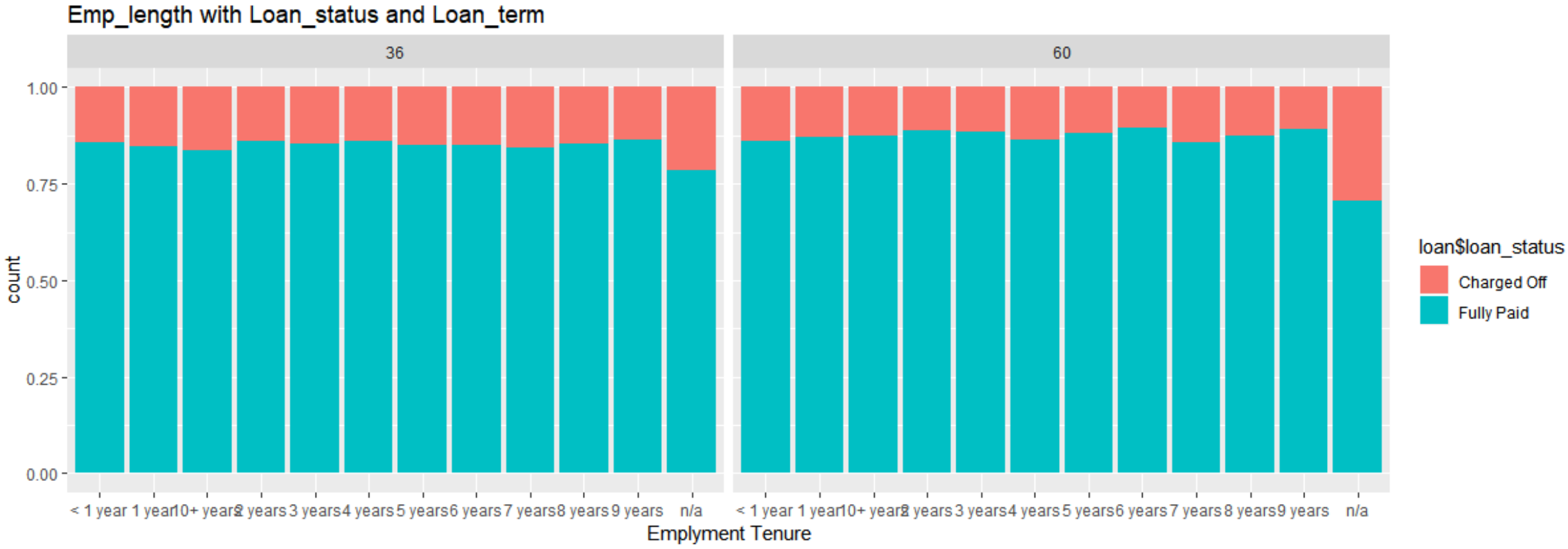
11. Corelation between Loan_amt And Funded_Amt



- The correlation between Loan_Amt And Funded_Amt is 0.937.

Bivariate Analysis

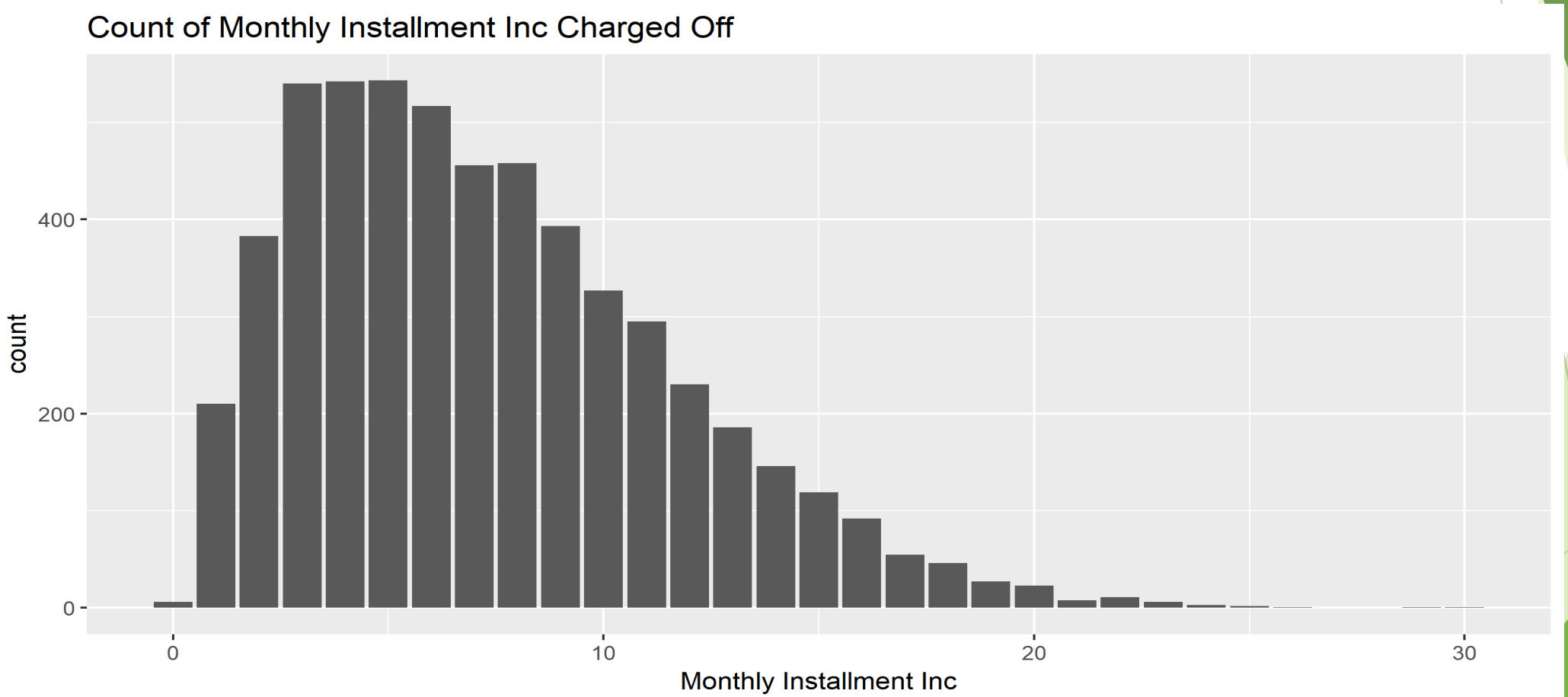
12. Employee Length With Loan_Status and Loan_Term :



- The Emp_Length with 10+ Years has been more charged off for 36 Months Tenure and
- The Emp_Length with 7 Years has been more charged off for 60 Months Tenure

Univariate Analysis

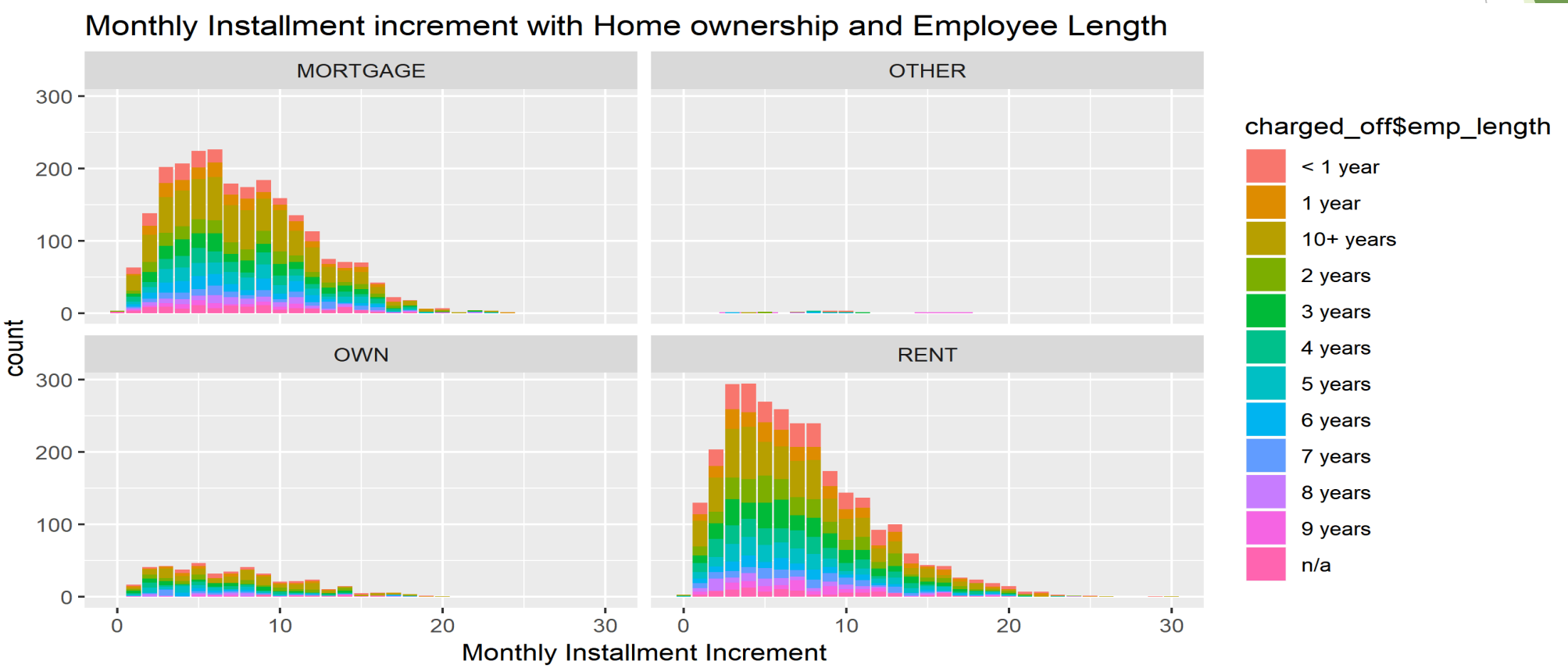
13. Monthly Installment of 5-10%



- The Monthly Installment is calculated based on the installment/monthly income(Annual income/12).
- From the above graph, We can see the Loan where monthly installment is around 5-10% of monthly

Bivariate Analysis

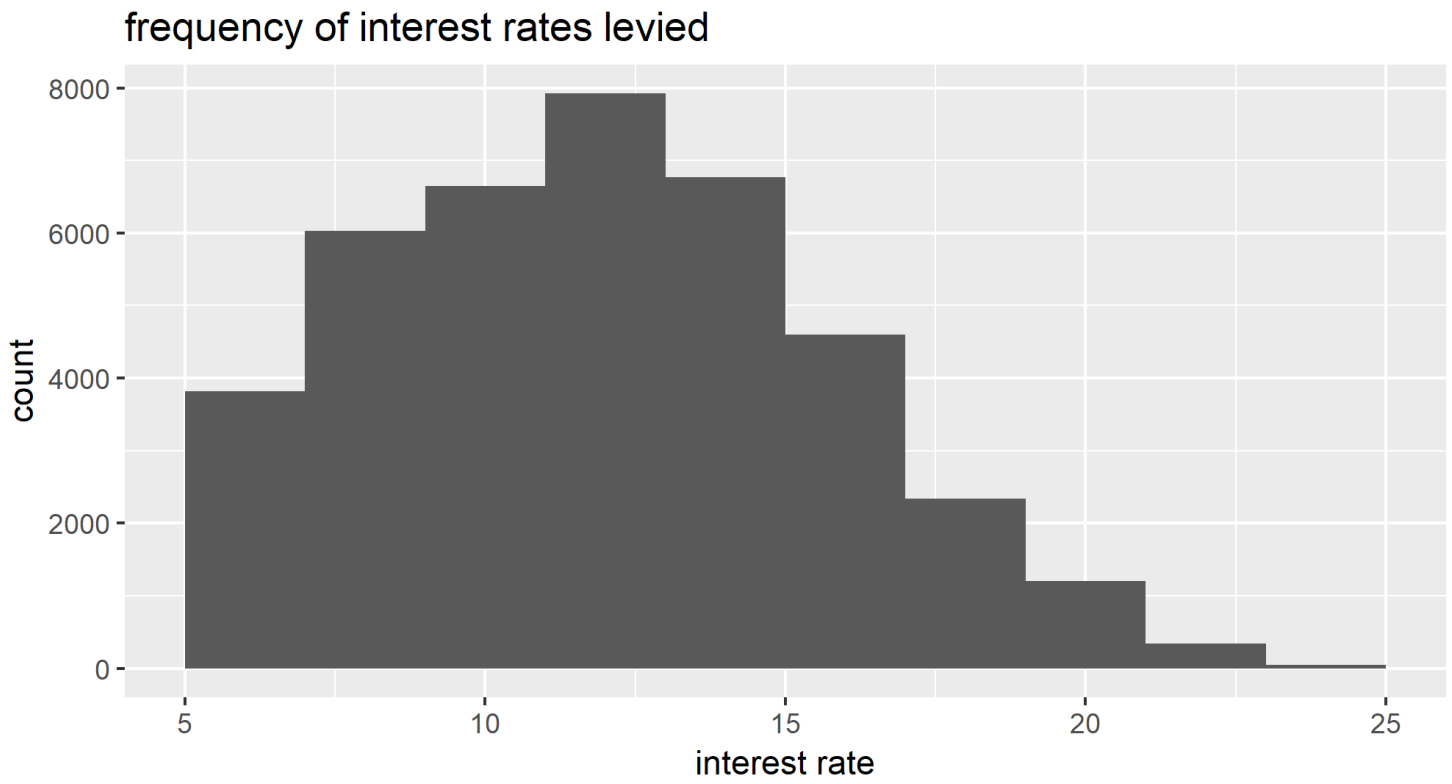
14. Monthly Installment increment with Home_ownership and Emp_Length



- From the Above graph, Loan where monthly installment is around 5-10% of monthly income on Mortgage/Rent are high defaulters

Univariate Analysis

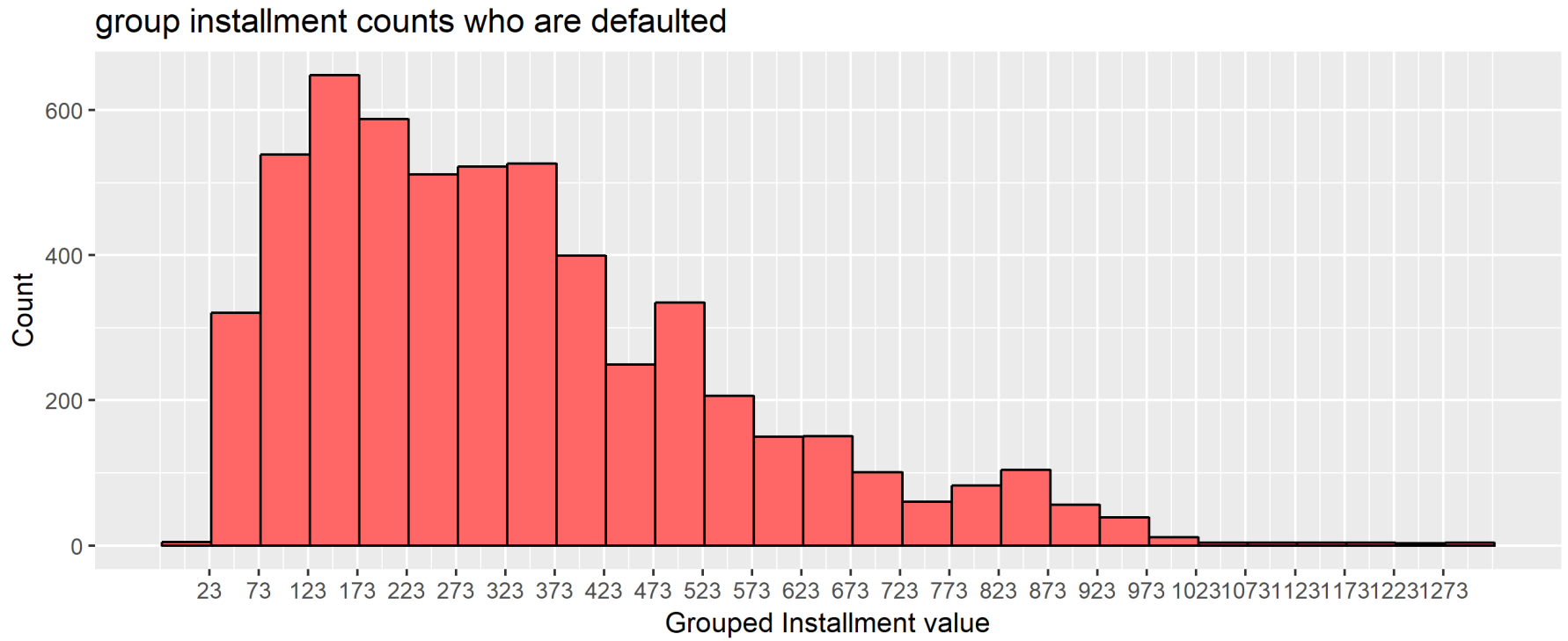
15. Observation with interest rates levied



- Based on above plot, we can figure out that company levied interest rates mostly between 7% and 15%

Univariate Analysis

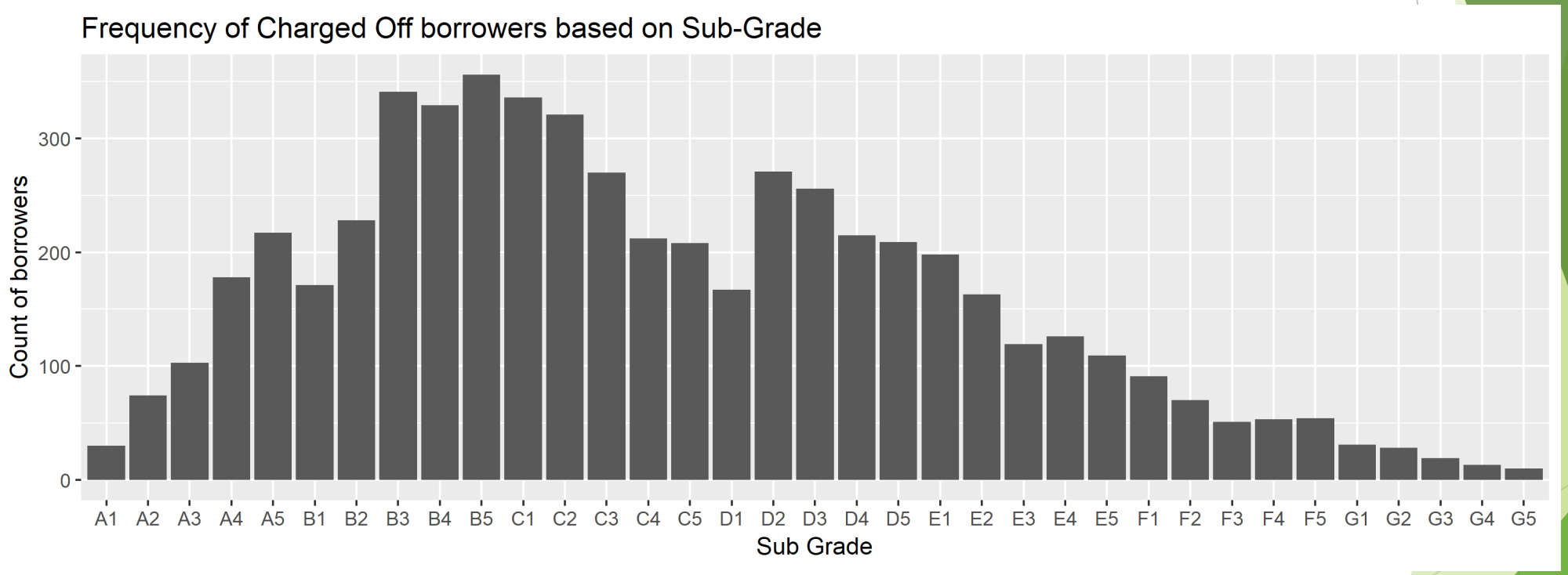
16. Monthly Payment owed by borrower Analysis



- Observation from above is that installment values vary mostly between 73 and 373 for Charged Off borrowers.

Univariate Analysis

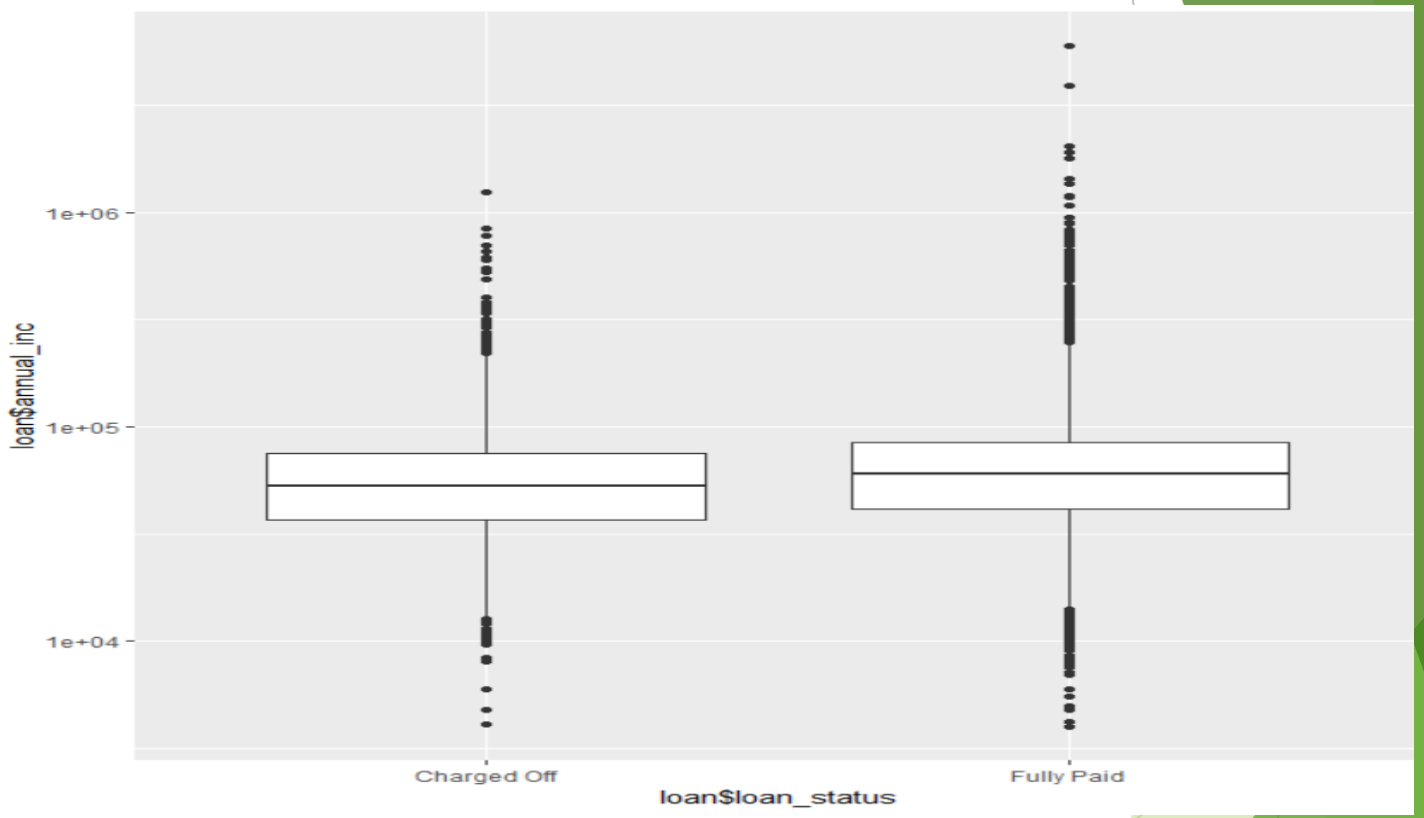
17. Analysis of Sub-grades for Charged Off borrowers



- Based on above plot, we could see that there are high no of **Charged Off** borrowers for sub-grades **B3 to C3** and later from **D2 to E2.**

Univariate Analysis

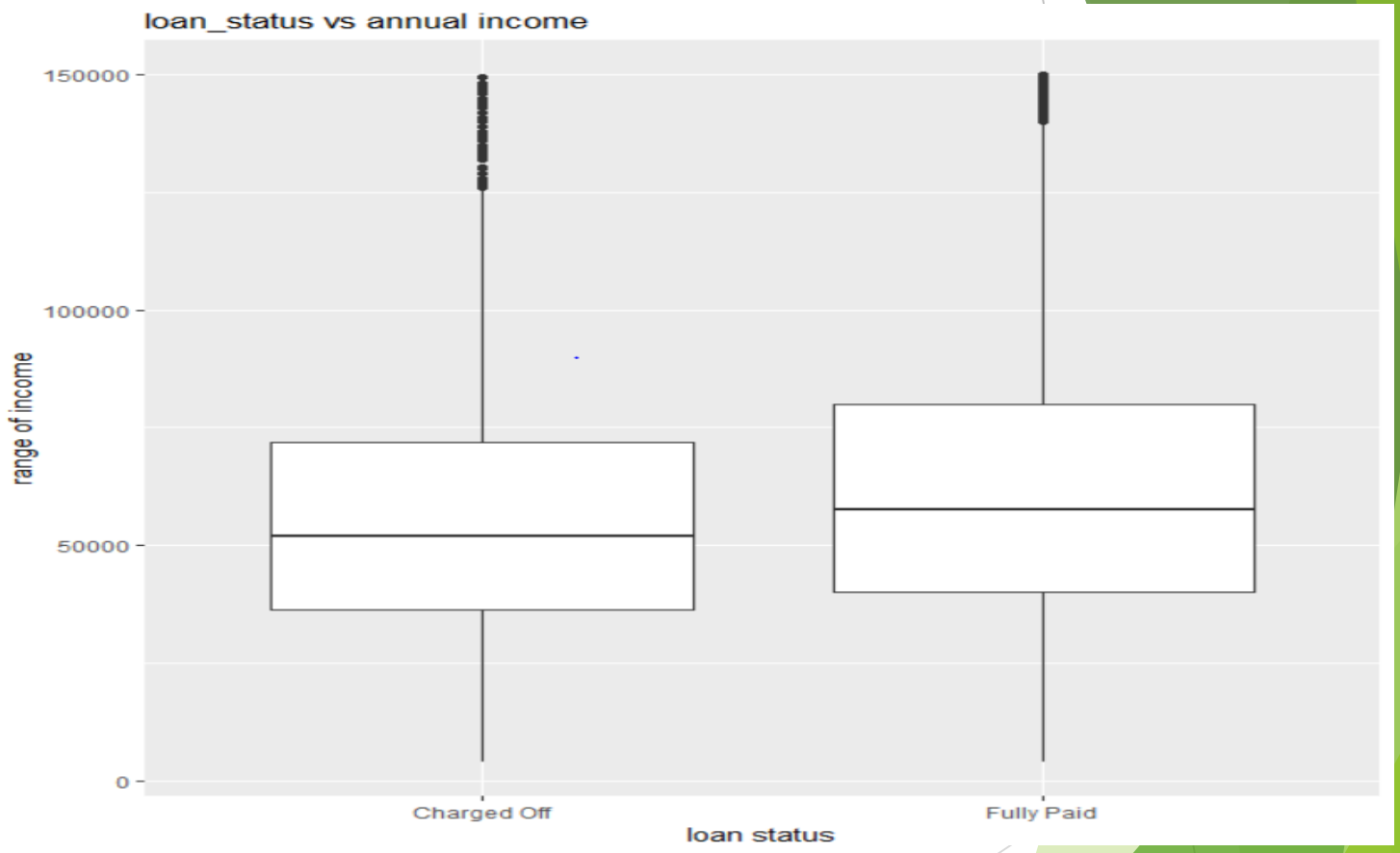
18.a Loan_Status With Annual_inc



- On plotting a log scale for annual income, we can see many outliers, hence removing these outliers and considering data upto 200000

Univariate Analysis

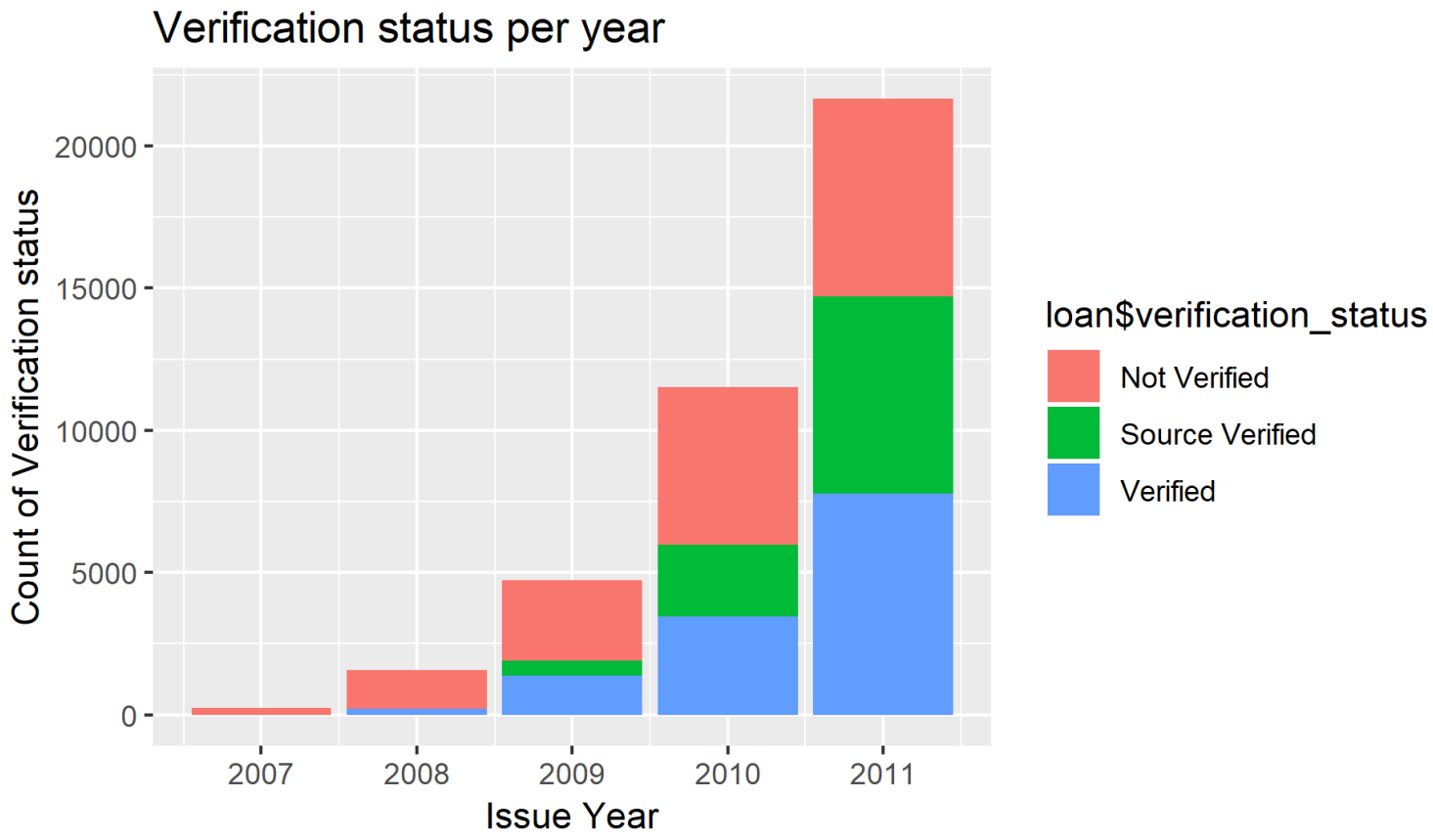
18.b Loan_Status With Annual_inc after removing Outliers



- On plotting a log scale for annual income, we can see many outliers, hence removing these outliers and considering data upto 150000

Bivariate Analysis

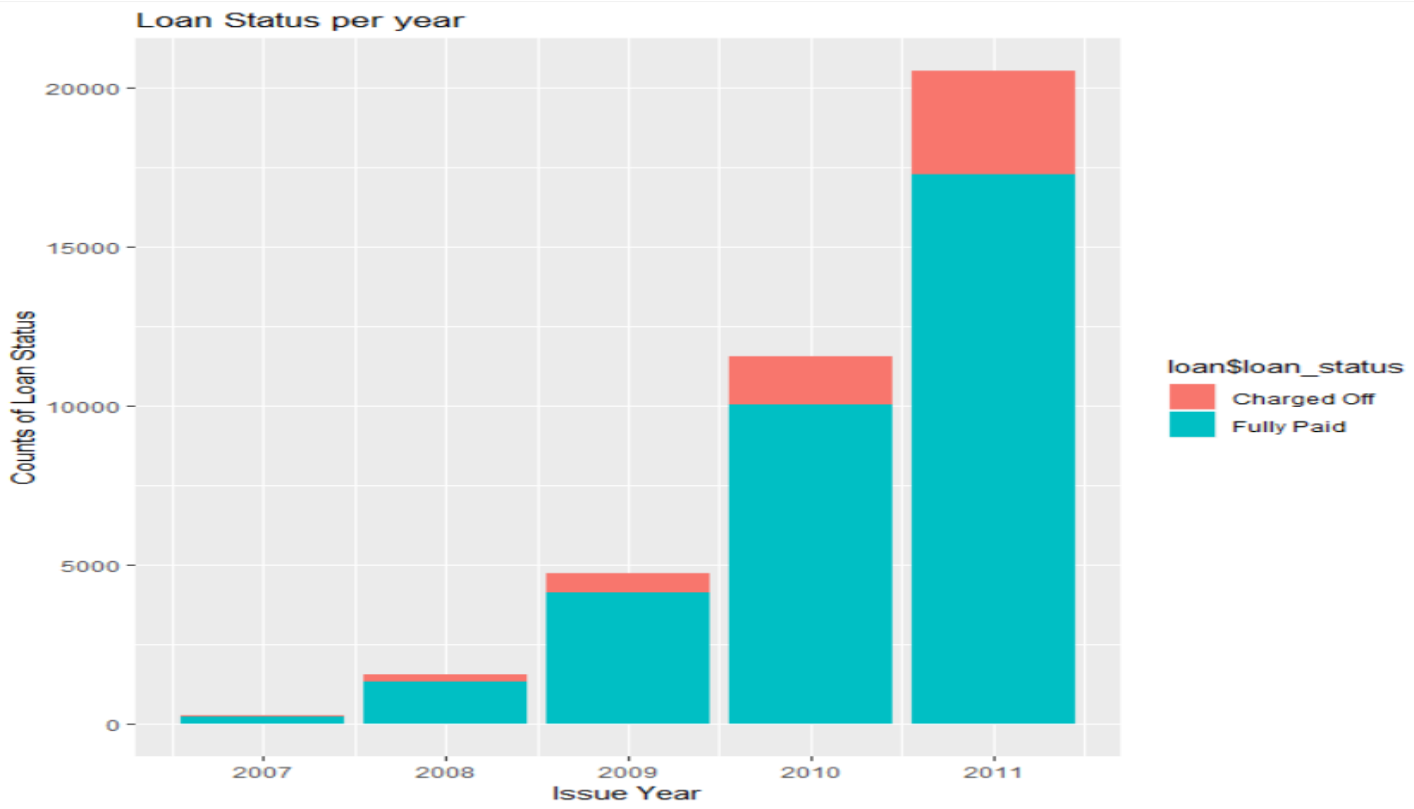
19. Status of verification based on loan issued year



- From this plot, its clear that in year 2010 and 2011, most borrowers were not verified

Bivariate Analysis

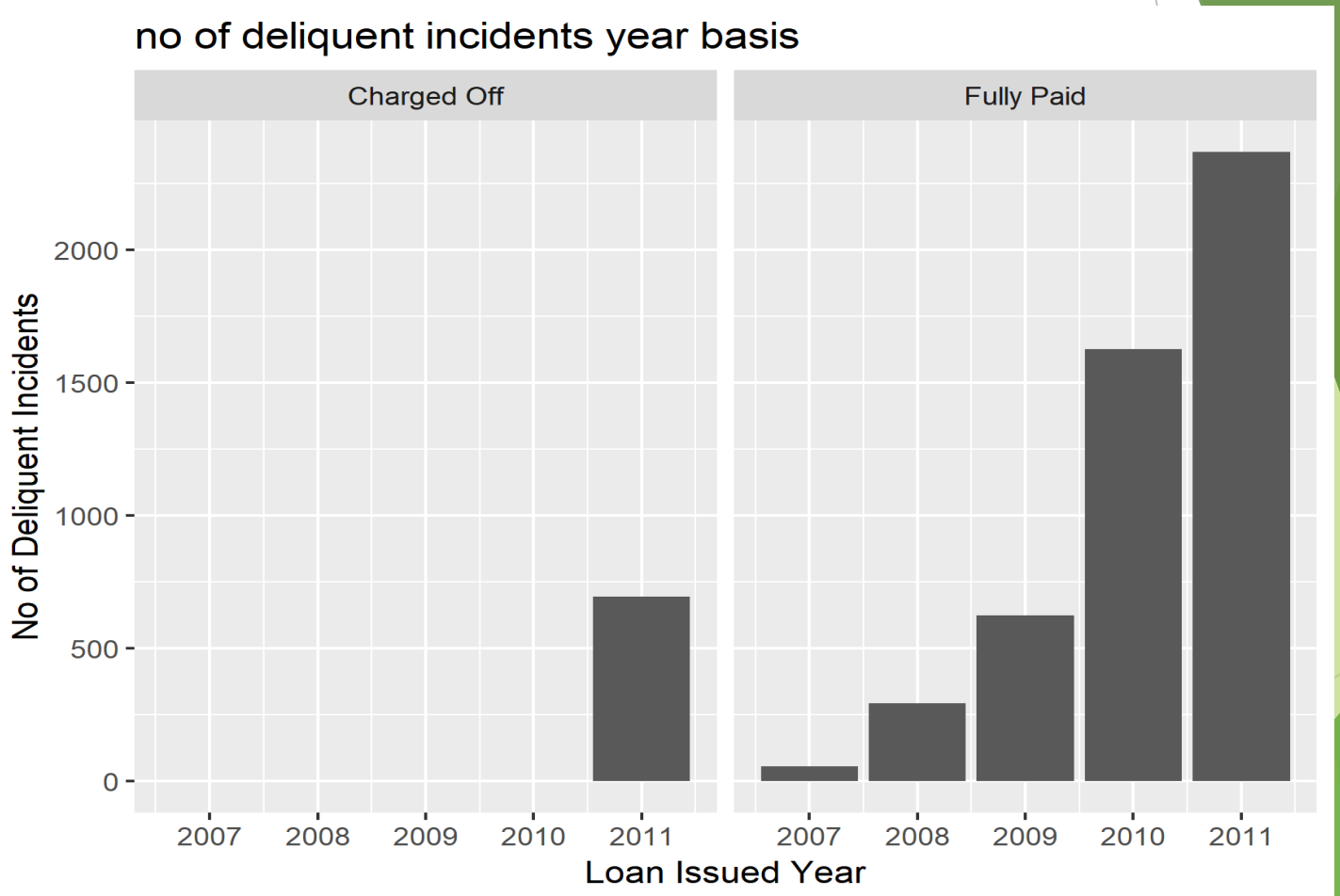
20. Loan Status issued per year



- From this plot, its evident that there were many Charged Off borrowers in year 2011 when they issued for loan.

Bivariate Analysis

21. Delinquent events per yea



- If we were to analyze delinquent incidents for Charged Off borrowers based on issue years, all Charge Offs occurred in year 2011.

Observations

- A large number of borrowers are **not Verified**.
- **Charged Off** borrowers are basically located in **CA,** also the same state where maximum non-verified borrowers are present. **NY** comes next to this list.
- However, **NE** has highest percentage of borrowers who are Charged Off.
- Both non-verified and Charged Off borrowers are having employment length of 10+ years.
- In year 2010 and 2011, most borrowers were not verified and all Charge Off incidents happened on issue year of 2011.
- Majority of Charge Off borrowers took loan with a purpose for their debt settlements.
- Charged Off scenarios are more frequent for sub grades **B3 to C3** and **D2 to E2** but with a gradual decrease.
- Grades have been decided base on interest rates or vice-versa.

Recommendations

- Company should take steps to do verification for each before providing loans.
- Company should take measures to do proper verification for 10+ experienced employees as majority of them defaulted.
- Company should also consider reason for loan and make judgments on weather to provide or not to provide loan, or to change the interest rates.
- Company should grade their borrowers carefully based on their payment history, earlier settlements etc.