LENDING RISK MANAGEMENT CASE STUDY
Gramener Case Study

SUBMISSION

Group Name:

1. Hoshedar Mody
2. Vivek Mohan Singh
3. Kishore Kumar Poosa
4. Krishna Teja (no contact and participation)

Abstract- For Lending Risk Management and Decision Support System

Why did we start? – To Minimize Risk of Lending

- Two **types of risks** are associated with the bank's decision:
- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

What did we do? – Analyses and Analytics – Helping to make risk free or minimal risk Decisions

Analyzed and cleaned data to formulate the edifice of Lending Risk Management

- The data contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- In this case study, we use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

Plotted Graphs and isolated the Potential Risk bearing prospective borrowers of the Loans

- In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
- In other words, the company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

For Lending Risk Management and Decision Support System

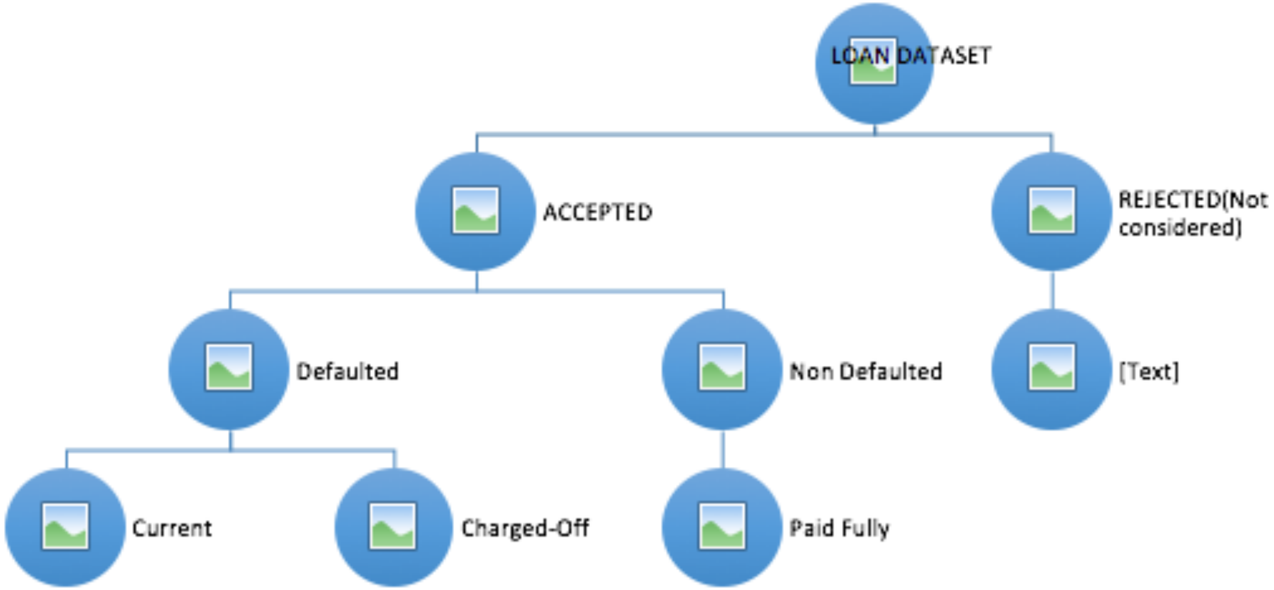
What does it mean?

When company receives a loan application company has to make a decision for loan approval based on profit.

For making business related decision we have to identify patterns which indicates if a person is likely to default majorly ,

Categorizing Loan Data Set into two main Data sets for analysis **DEFAULTERS & NON DEFAULTERS**

We made most of our analysis using this dataset tree model



Lending Risk Decision Support Methodology

Format Clean Up

- Ensure Valid Data format
- Clean data to enhance quality and format of Data

Extraction and Transformation

- Transformed raw Data into Analytical form
- Extracted valid accurate data for Analysis

Analysis and Decision making

- Performed Analysis on Data
- Transformed Data into Visual Plots to facilitate decision making

Data Cleaning and Enhancing

- Removed columns with more than 90% NAs
- 48 Columns used in the Study
- Annual Income bucket slot - annual_income_slot
- Interest Rate is a % which is converted to numeric absolute
- Blanks in Employee Title were populated with ‘other’
- Duplicates Checked (id and member_id)
- Removed Outliers based on Annual Income
- Segmented Data based on loan_status

STRATEGY :

- Identify largest source of loss
- Understand CREDIT LOSS :

1. Amount of money lost by lender when borrower refuses to pay or Runs away with the money.
2. Borrower who by default cause largest amount of loss to the lender hence is labelled as “Charged off” or “Defaulter.

- Thereof, Identifying Risky Applicants i.e. Reduced Defaulter -> Reduced Credit loss .
- Identify Strong Default indicators.
- Use such Default indicators to derive decision as to pass the loan or reject it.

Risk Analysis Table

	Total Data Set	Charged	Fully Paid	Current
1. Total number of (count)	39703	5626	32937	1140
2. Average (Mean LOAN)	11219			
4. Standard Deviation in LOAN Amount	55239			
5. 25 % (loan Amount)	5500			
6. 50% (loan Amount)	11219			
7. 75% (loan Amount)	15000			
8. 100% (loan Amount)	35000			

Univariate Risk Analysis

- For Univariate analysis we analyzed various categorical and statistical variables:
- Firstly we tried to generate a main Correlation Matrix on previously sorted dataset in R .
- For plotting we use corrgram package:

corrgram(loan_Copy, order=TRUE, lower.panel=panel.shade, upper.panel=panel.pie, text.panel=panel.txt, main="All variables correlation Matrix")

- From threre we had the Idea that strongly correlated variables are the one with dark **BLUE**

Shade i.e. that the one in the central region of the plots

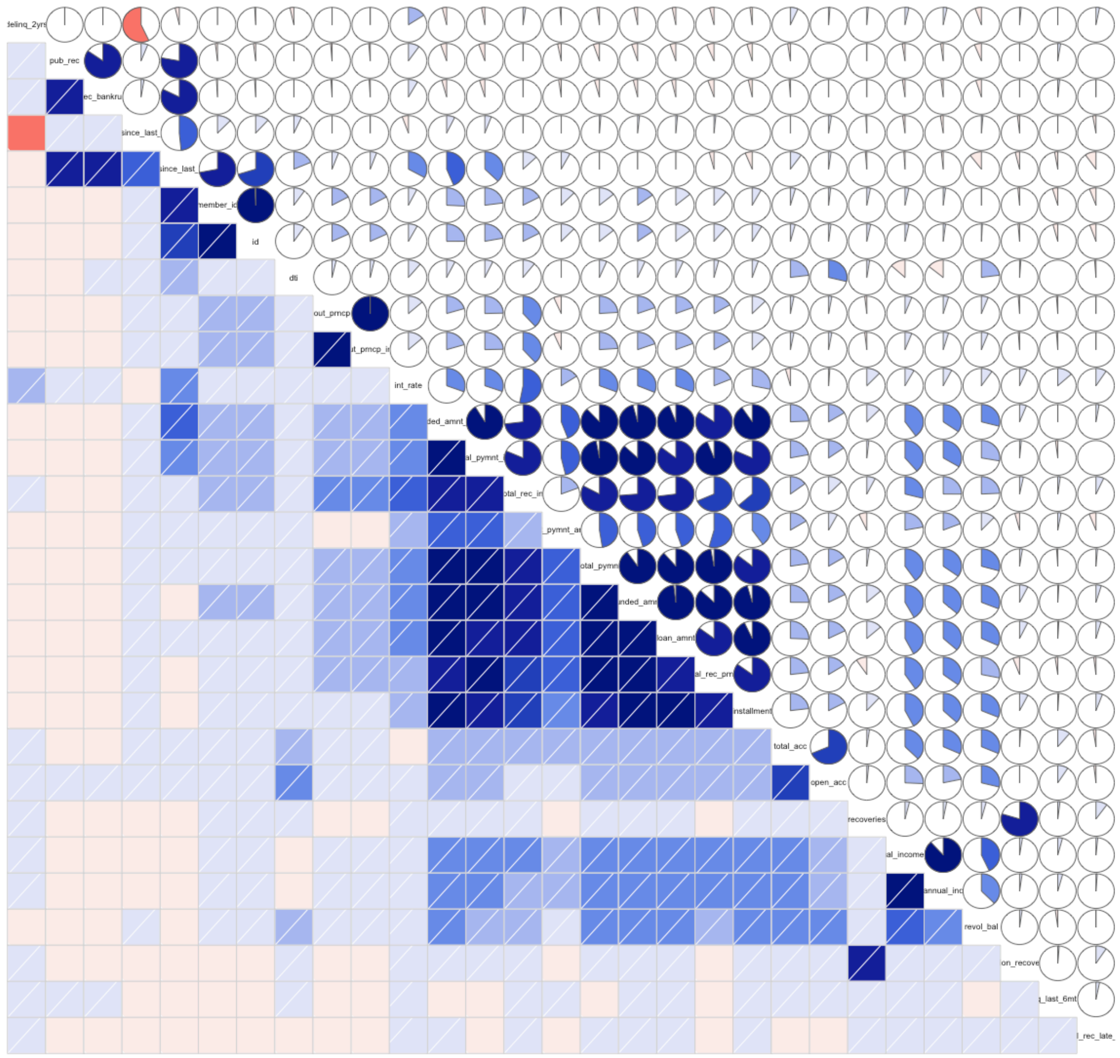
1. All the money related metric such as loan amount paid,

Annual income.

2.These have a strong correlation with variables such as term , emp_title, loan_status, purpose etc.

- So we further reduced our analysis spectrum and plotted separate Correlation Matrix for Deeply Correlated Metrics:

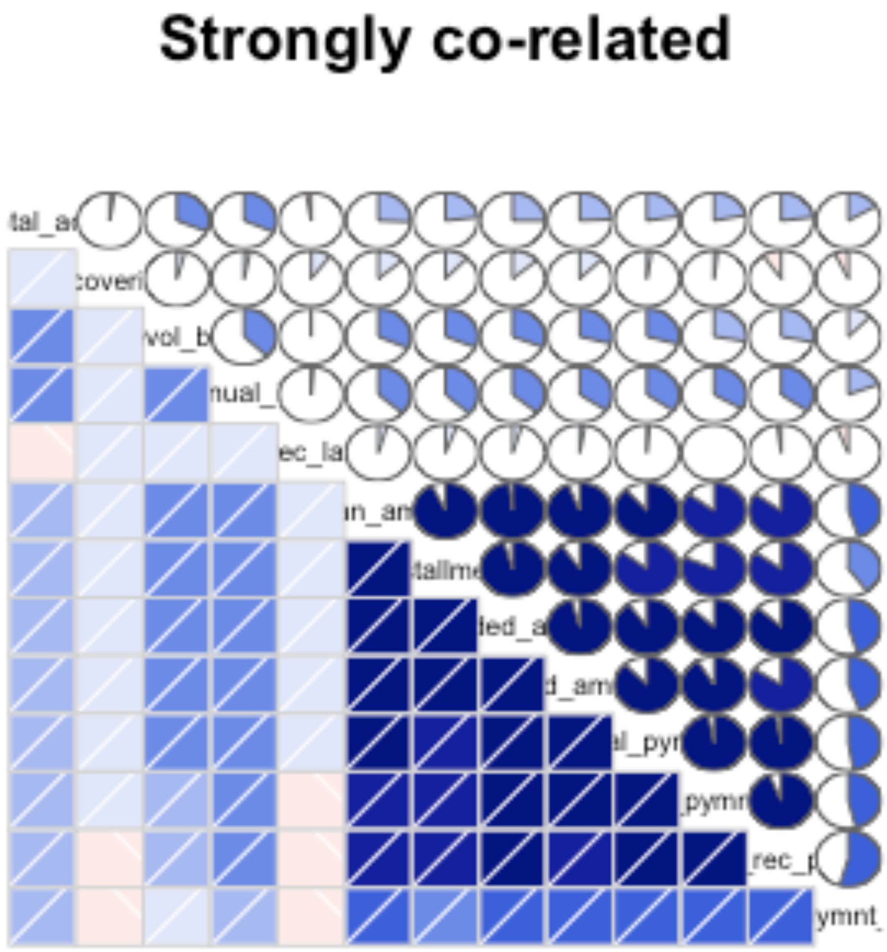
All variables correlation Matrix



- Here we used command:

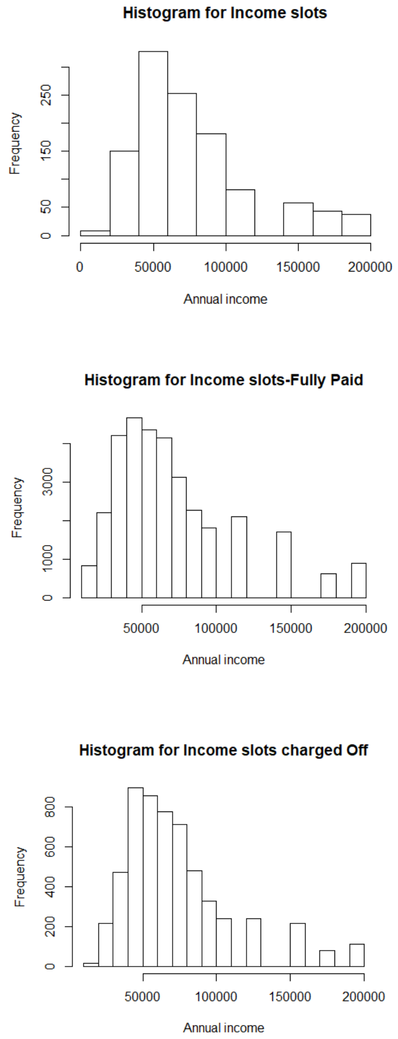
```
strong_correl_var <- c("loan_amnt","funded_amnt","funded_amnt_inv","term","emp_title","home_ownership","installment","annual_inc","verification_status","revol_bal","purpose","title","addr_state", "revol_util","total_acc","revol_bal","revol_util","total_pymnt","total_pymnt_inv","total_rec_prncp","total_rec_late_fee","recoveries","last_pymnt_amnt")

analysis_data<-loan_Copy[ , (names(loan_Data) %in% strong_correl_var)]

corrgram(analysis_data, order=TRUE, lower.panel=panel.shade, upper.panel=panel.pie, text.panel=panel.txt,       main="Strongly co-related")
```

- Now we plotted various plots to broaden our analysis.

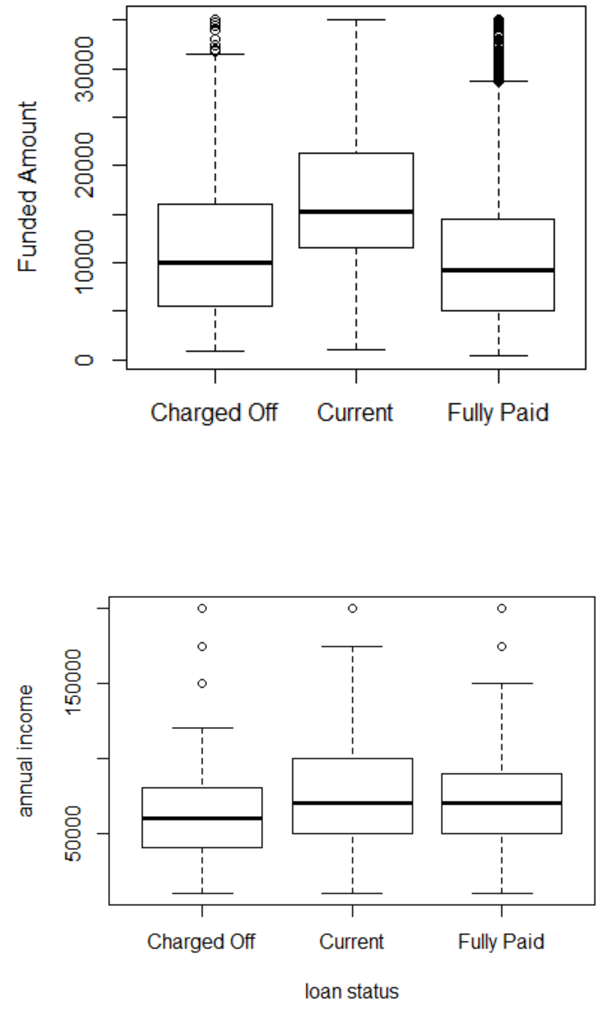


UpGrad

Annual Income range for each loan status



- Annual income 30000 to 80000 is more frequent for both Charged-Off and fully-paid Loans
- 63% of the Charged-Off and 75% fully paid customers are concentrated in this region.
- As the both are in the same region , we need to focus on other factors for risk
- Current loan status customers are concentrated between 40000 to 100000 (Annual Income Range)
- This indicates investor showing interest to invest on customers in this annual income range

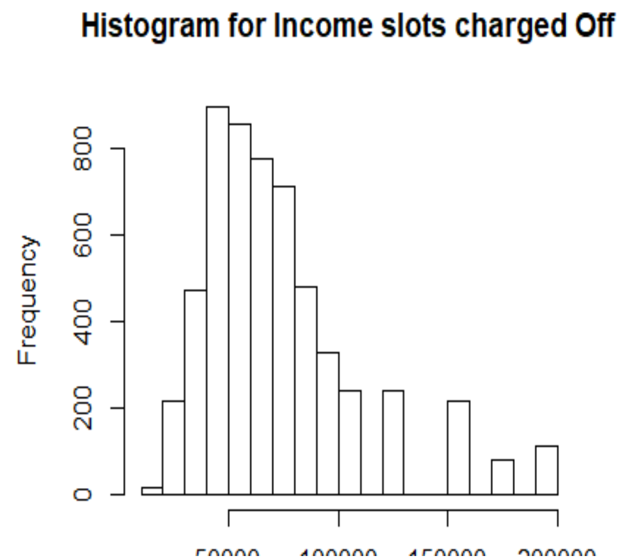


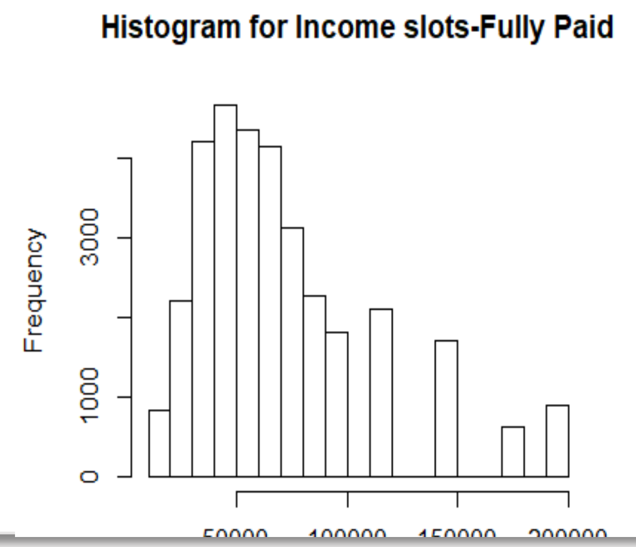
Funded Amount and Annual Income

- The Median for Loan Status - Charged-Off and Fully-Paid is 10,000
- But the Median for Current Loan Status is considerably higher(Bank has aggressive standards for approving Loans – which seems safe)
- The Median Annual Income is higher for Current and Fully Paid Loans and Bank is trying to Mitigate Risk of Lending.

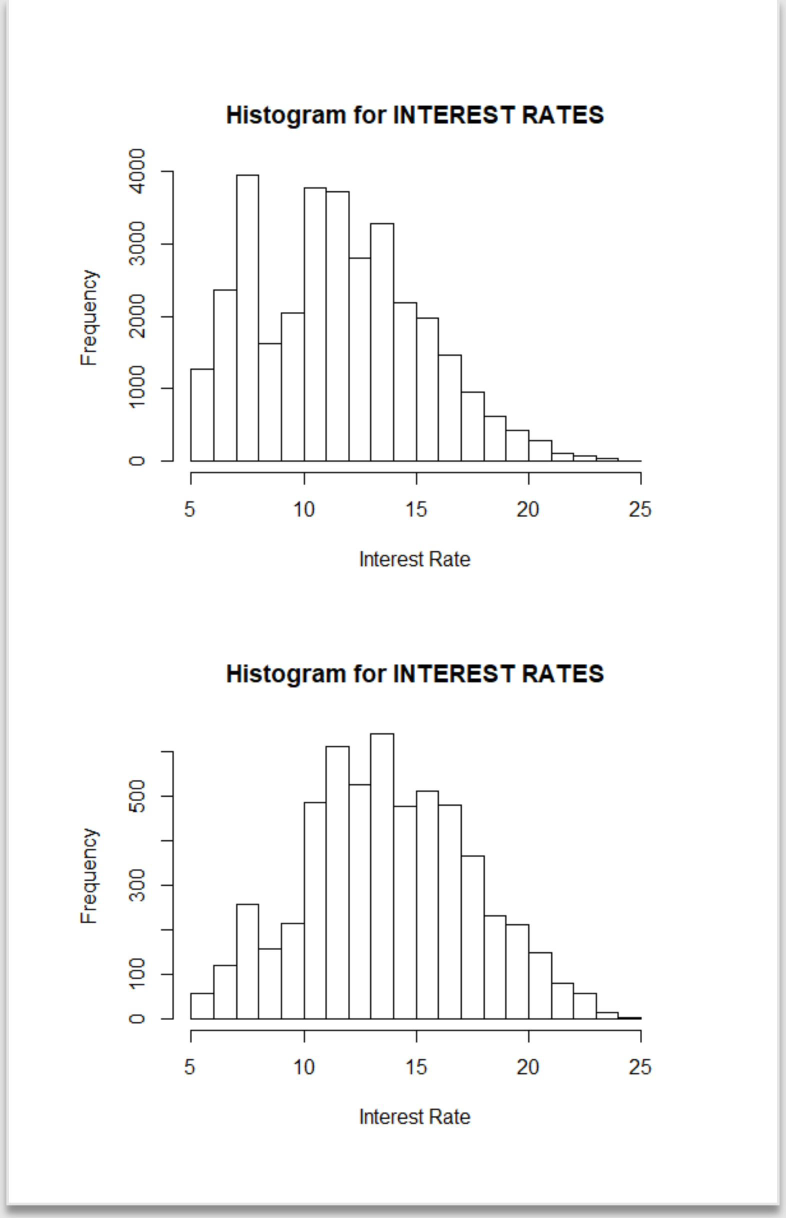
UpGrad

Histogram-Income Slot Analysis





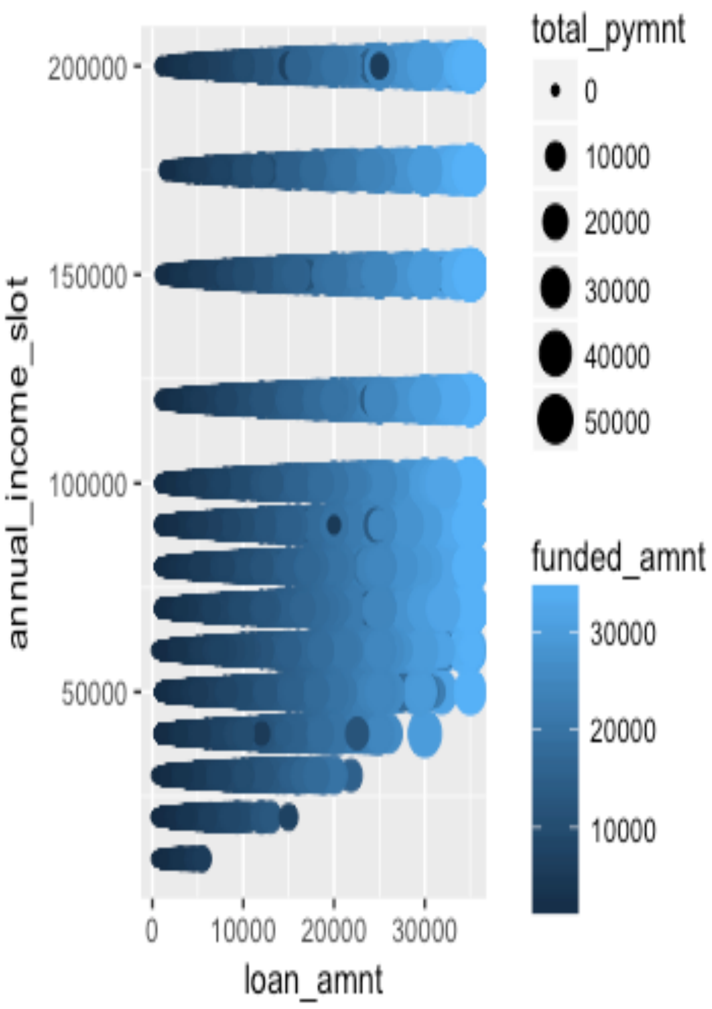
- Fully Paid Loans have higher income to loan ratio which shows high income borrowers are low business risk
- Max income range for Full paid is from 0 to 100,000
- Max income range for Charged-Off is 0 to 750,000
- 3000 + borrowers have fully paid Loans v/s 800+ borrowers who defaulted or loans were Charged-off
- Income was not a factor which influenced decisions

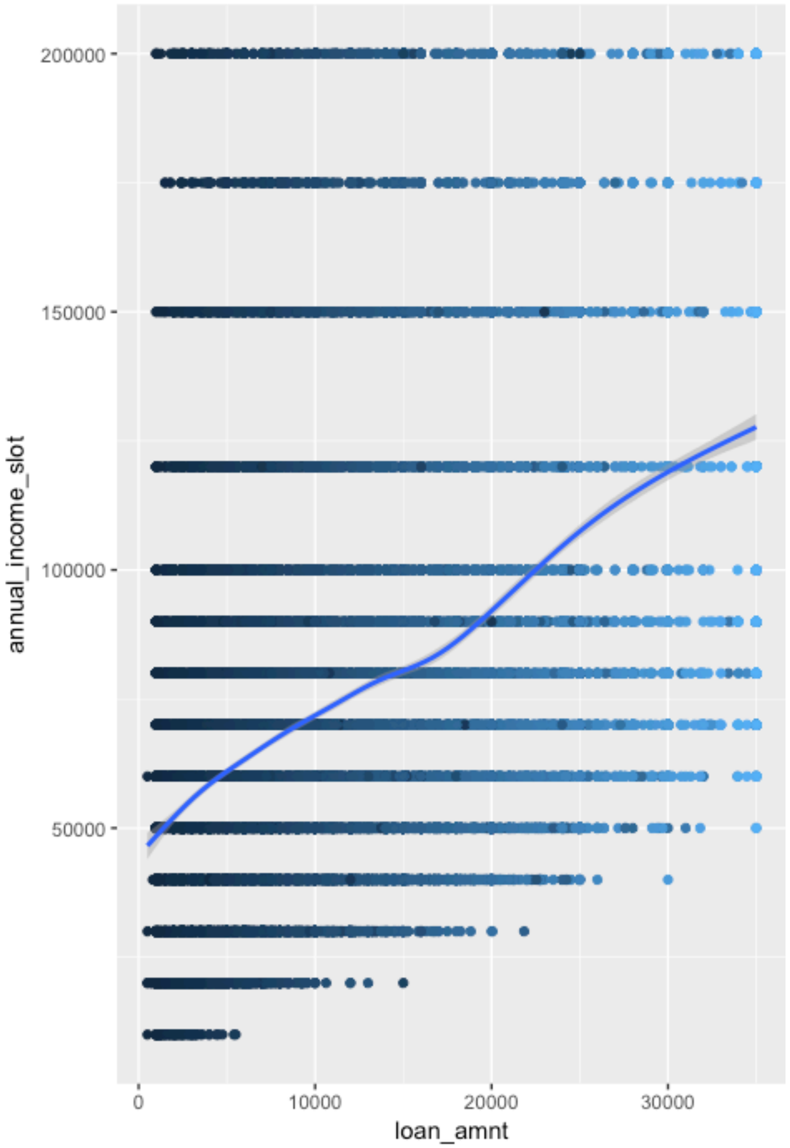


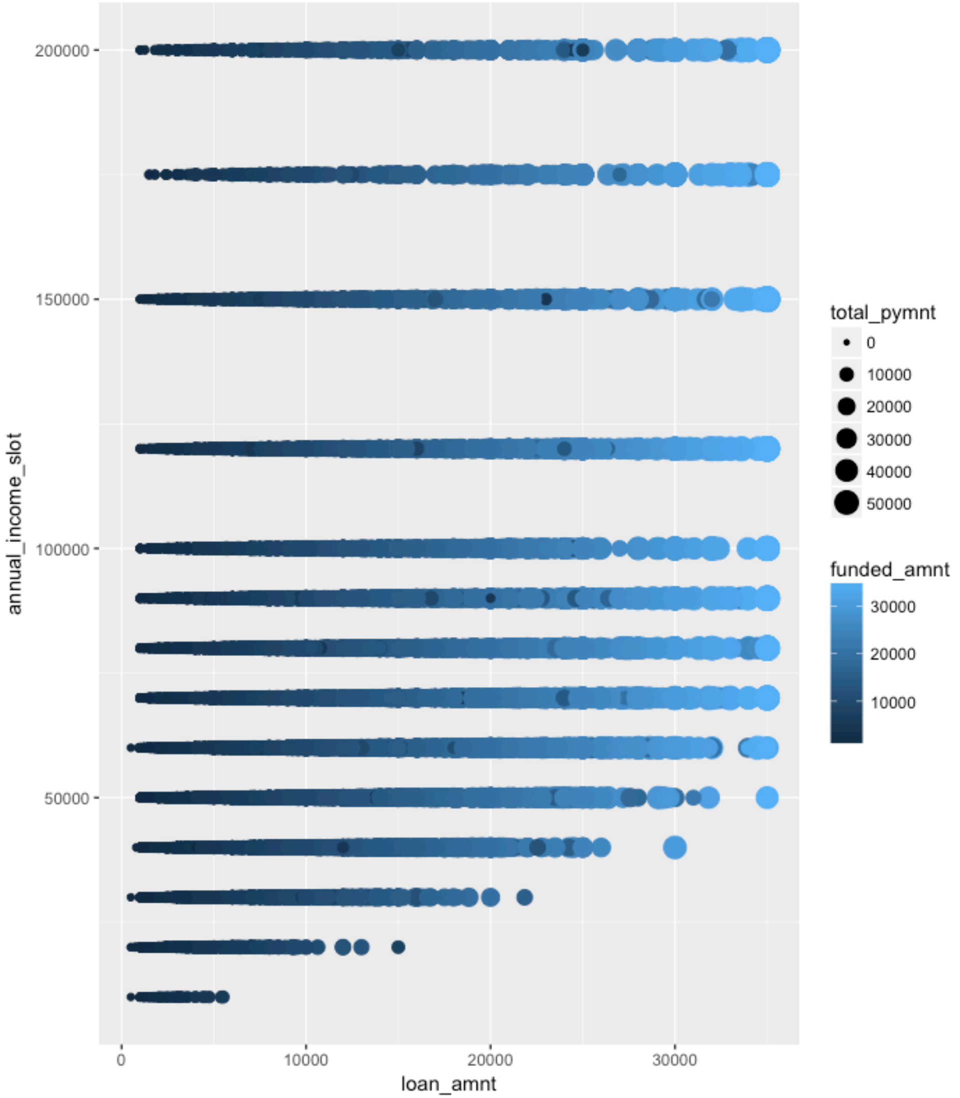
Interest Rates:

- Higher the interest rate, higher the number of Charged-Off loans but we cannot generalize this conclusion.
- Interest rate between 10% to 14 % more charged off's exist than other interest rates making these interest rate for Loans and Borrowings Risky
- There is a remarkable trend where lower the interest rates charged to loans make people pay the loan with minimum risk and default
- Univariate analysis is giving us an idea on data but it is not driving us to conclusion.
- Lets draw plots with multiple variables to understand the pattern

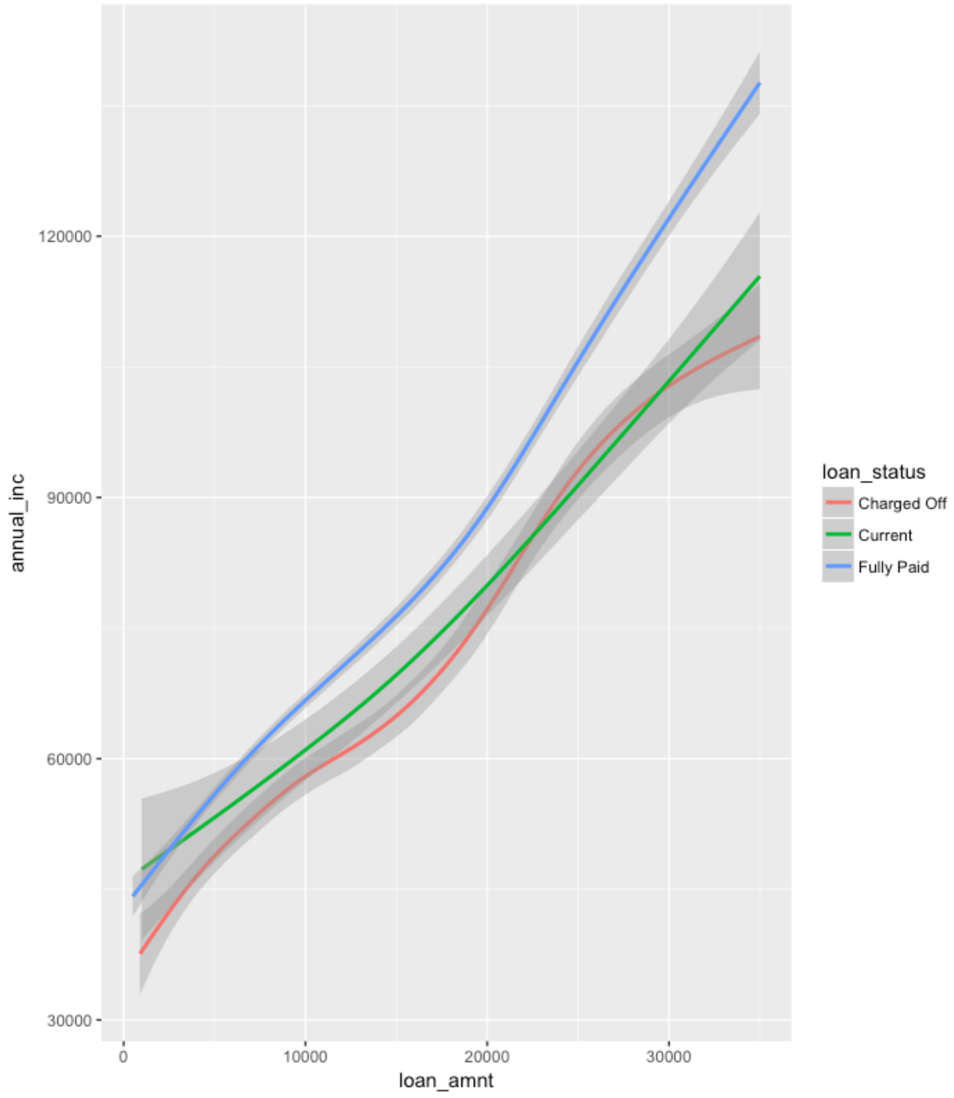
Annual Income vs Loan Amount Investment

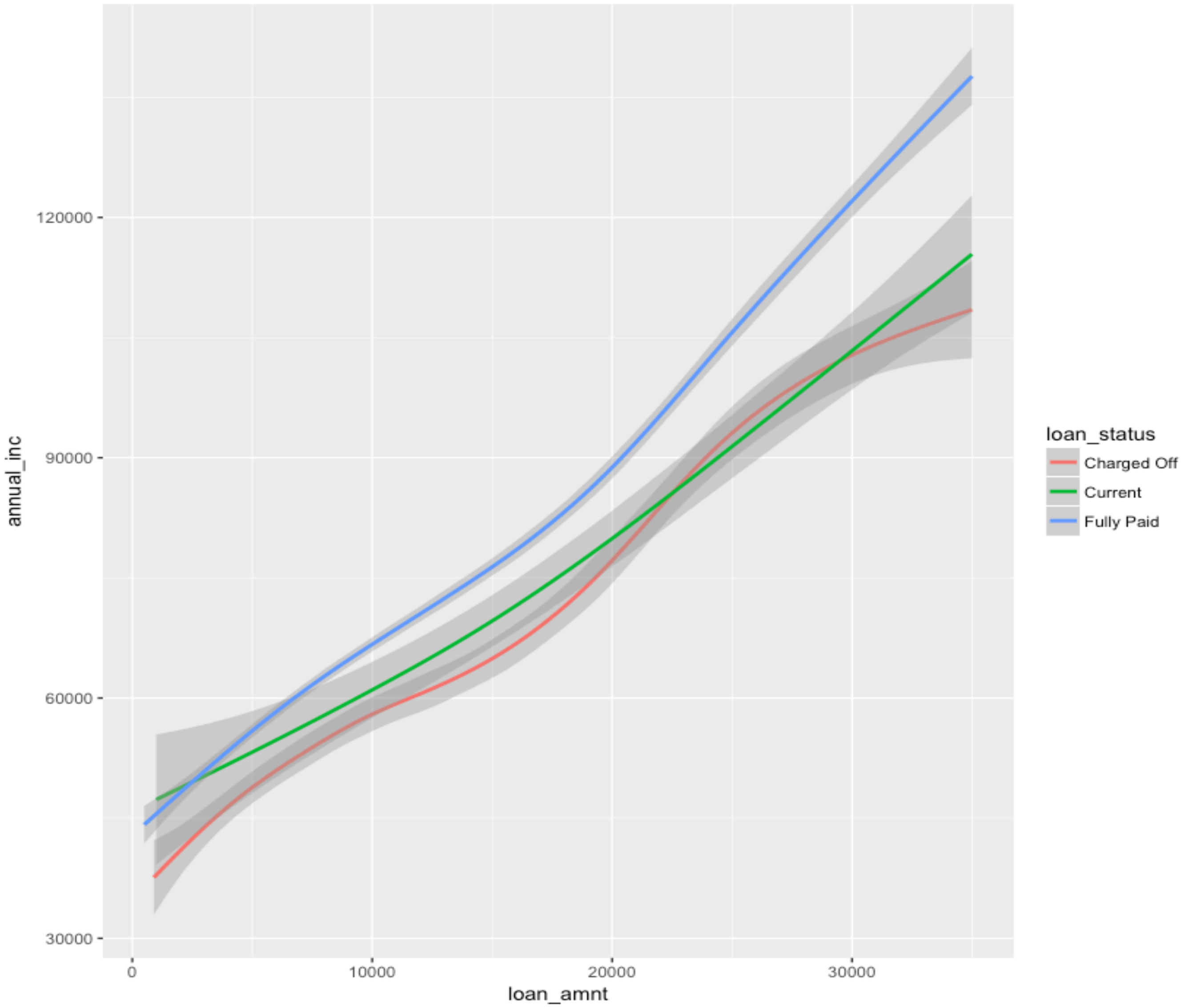




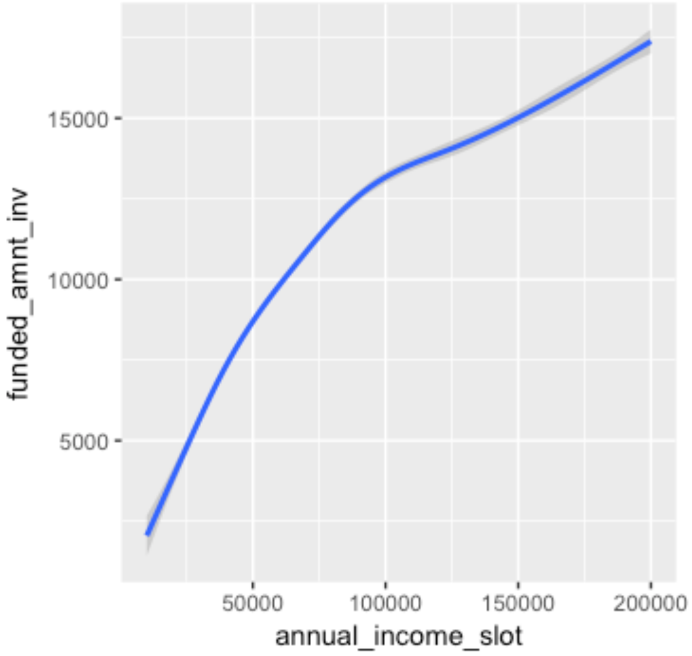


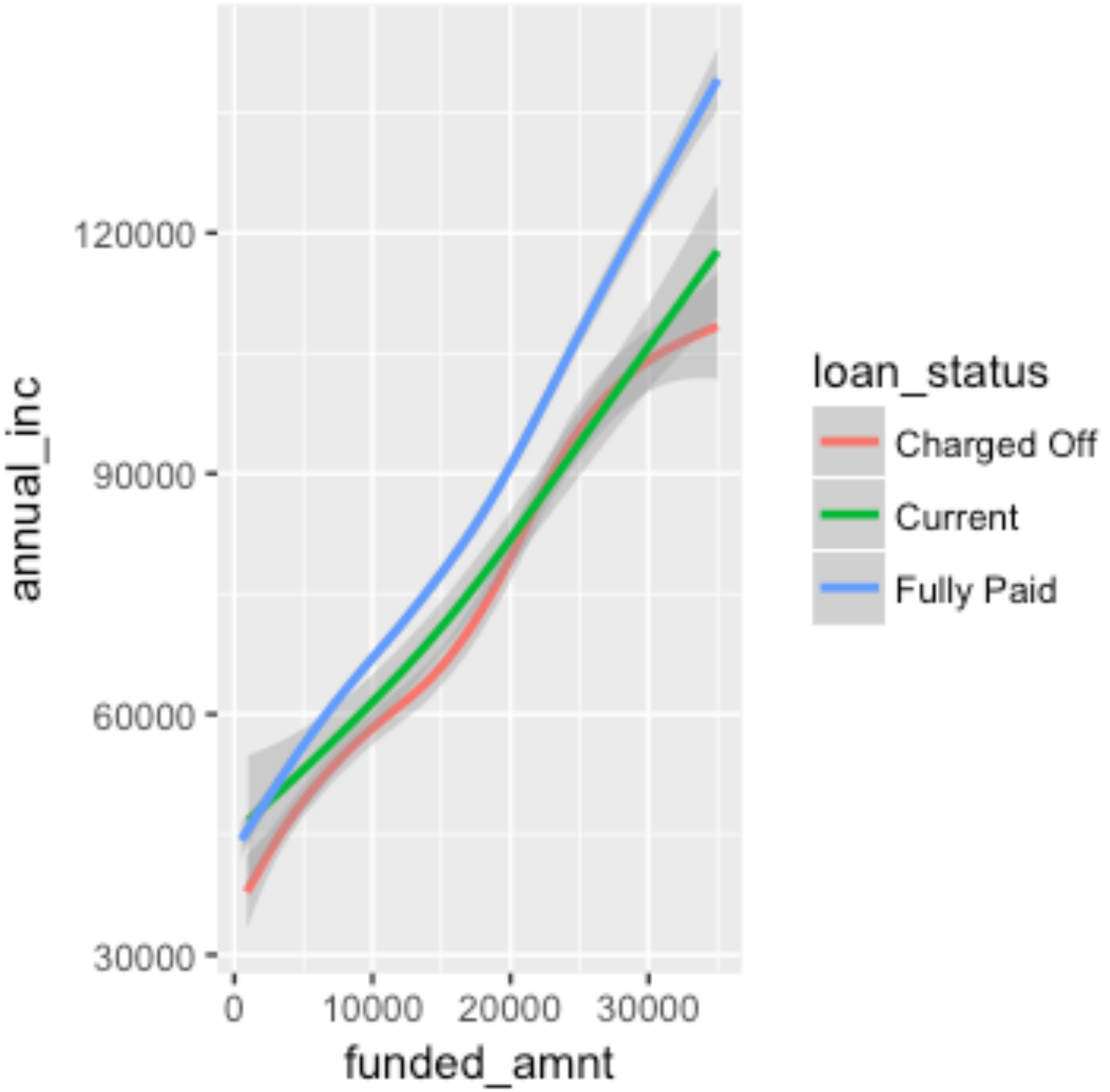
Trends in loan Type Analysis:





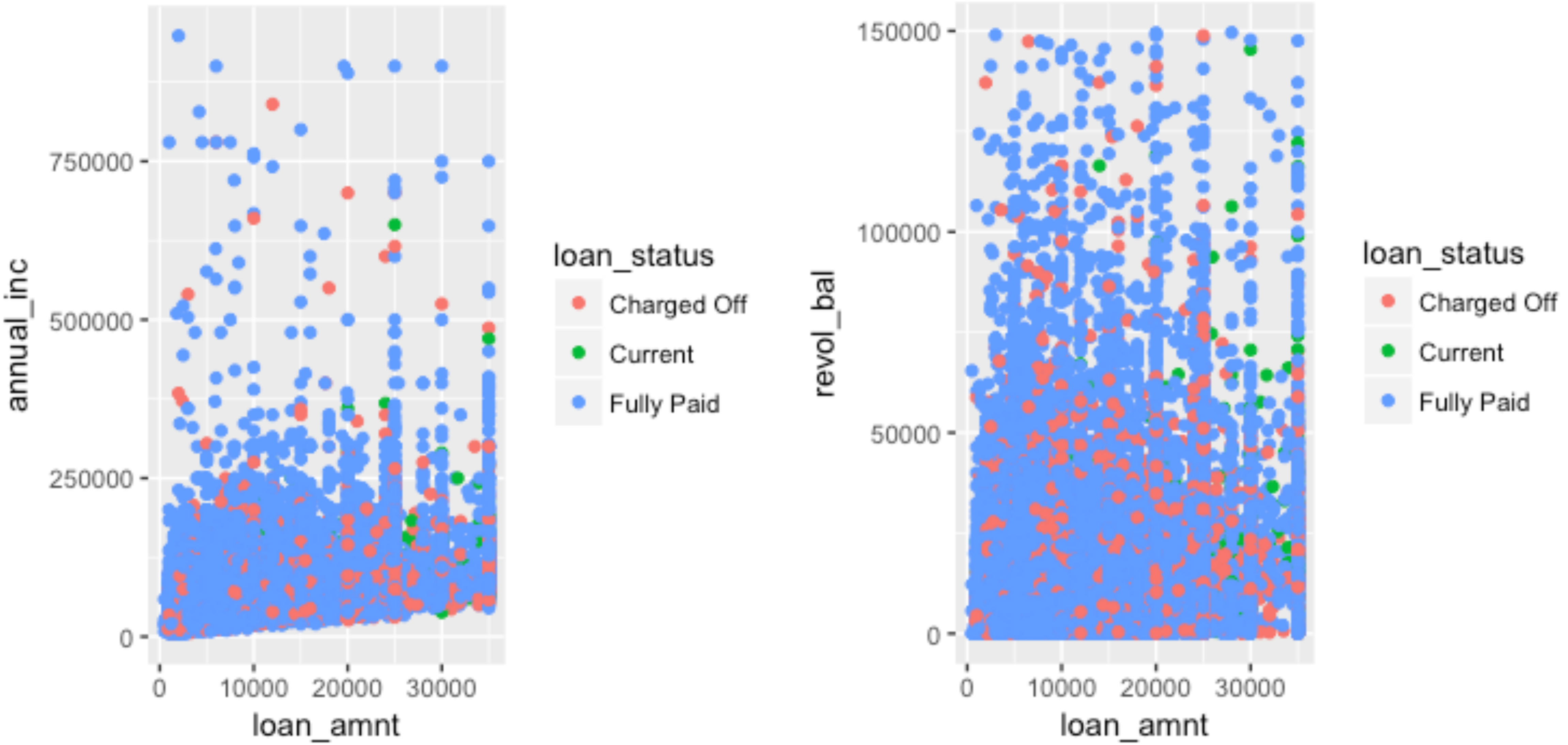
Final Results from this:





We derive from these plotted analysis that people with high annual income tend to linearly gain advantage of high funding but the trade off is that the charged off variable very steadily increase on high funded types.

Analyzing Loan Amount with other categorical variables (Bivariate Analysis)



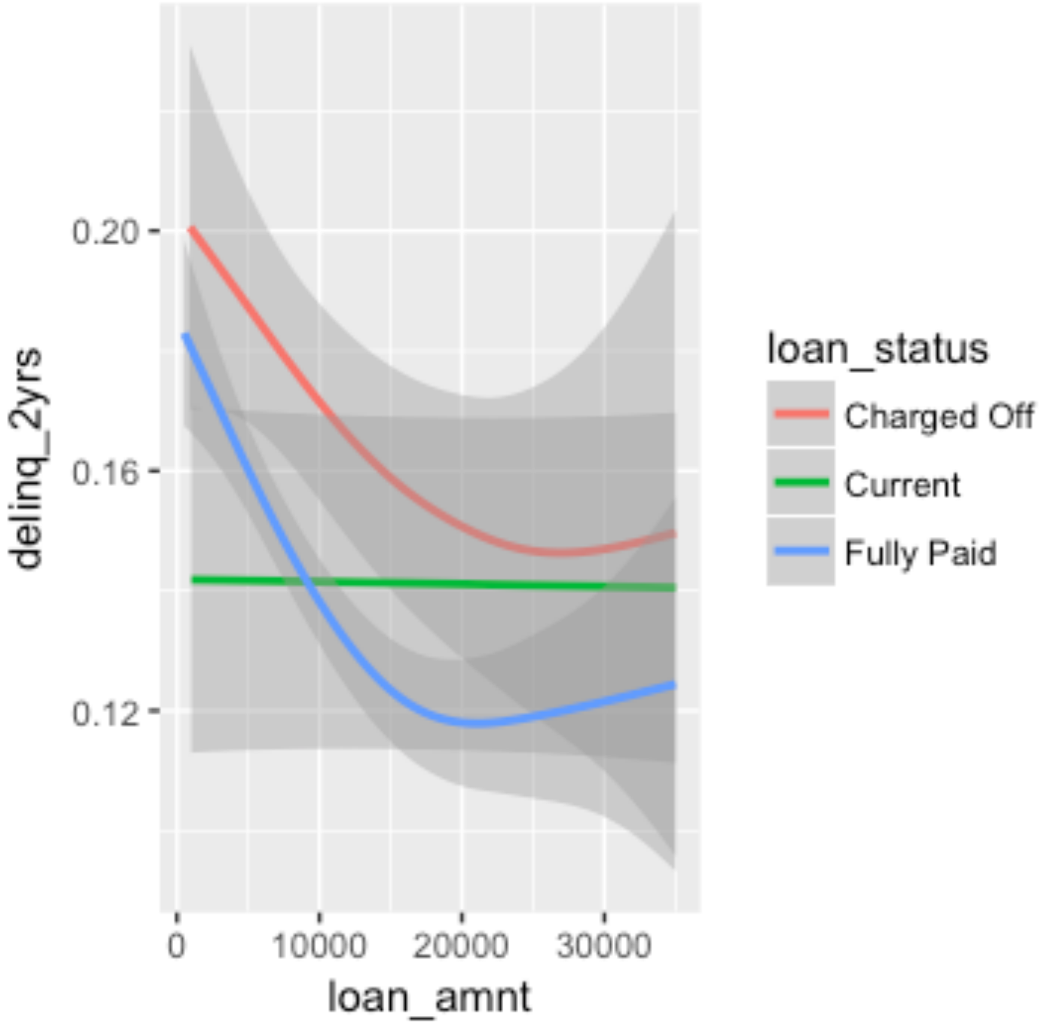
We Analyzed that :

We see that the delinquency rate is more for loan with lesser amount and a steady negative slope comes after loan amount >20000 this gives us a insight that :

For current and fully paid the delinquency rate varies as people who paid rarely tend to not pay or delay later,

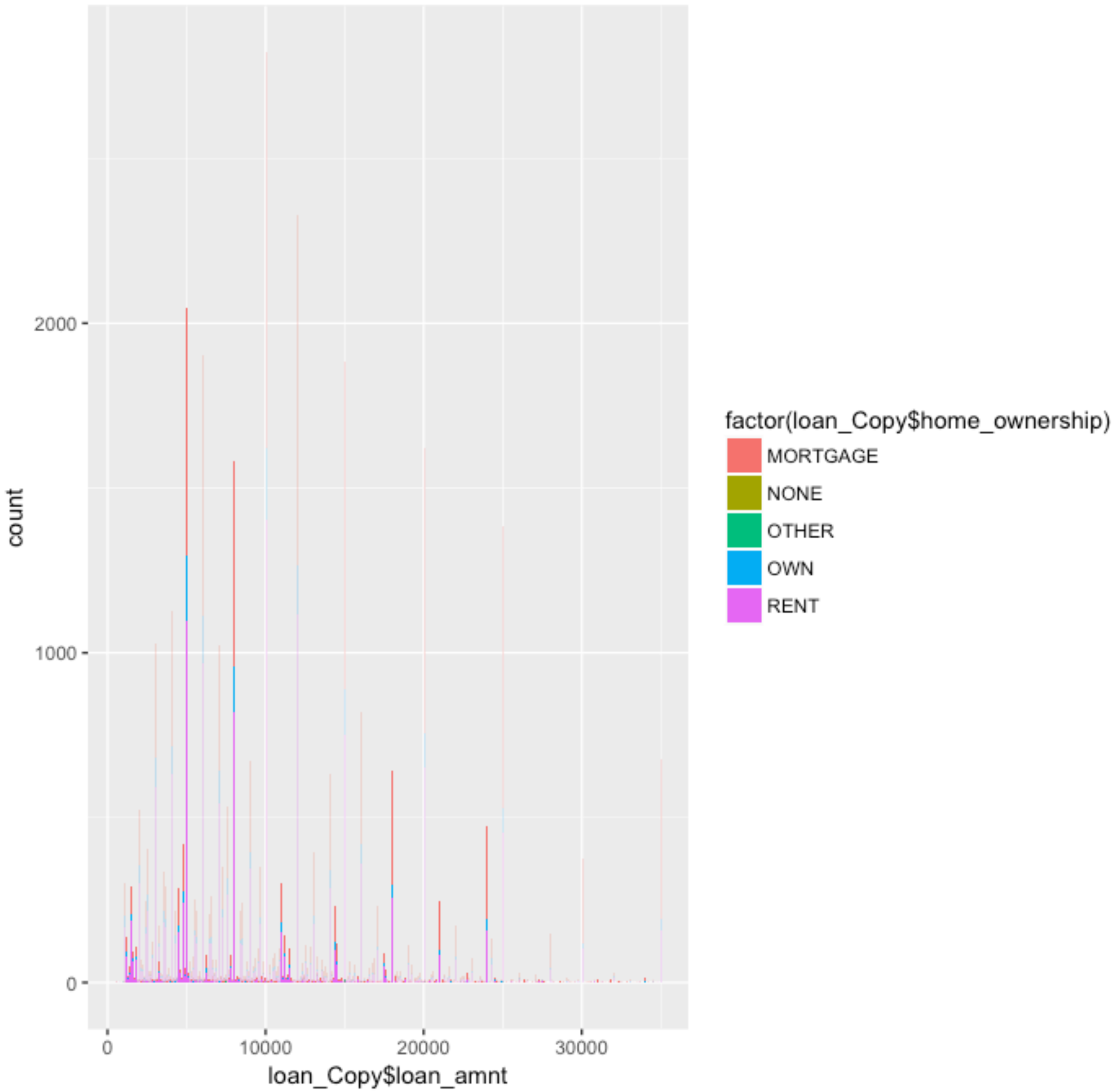
For people who are charged off at lesser loan amount tend to show great delinquency as the loan amount is less so they can afford but for higher loan the bank are taking stringent action to control their activity an insure a steady pay of loans

So the risky defaulters are loan borrowers with lesser amount and Current with intial less loan amount .



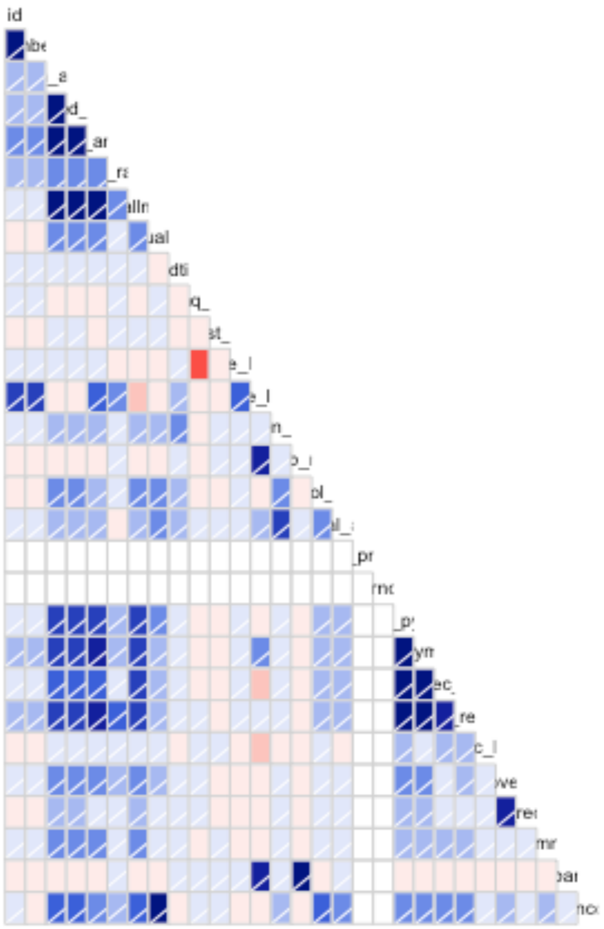
Loan amount vs Ownership as a factor variable

People with mortgage and loan amount between 0- 10000 are most in number, which again is a Risk Indicator of less required and unwanted category of less amounted loan seekers .

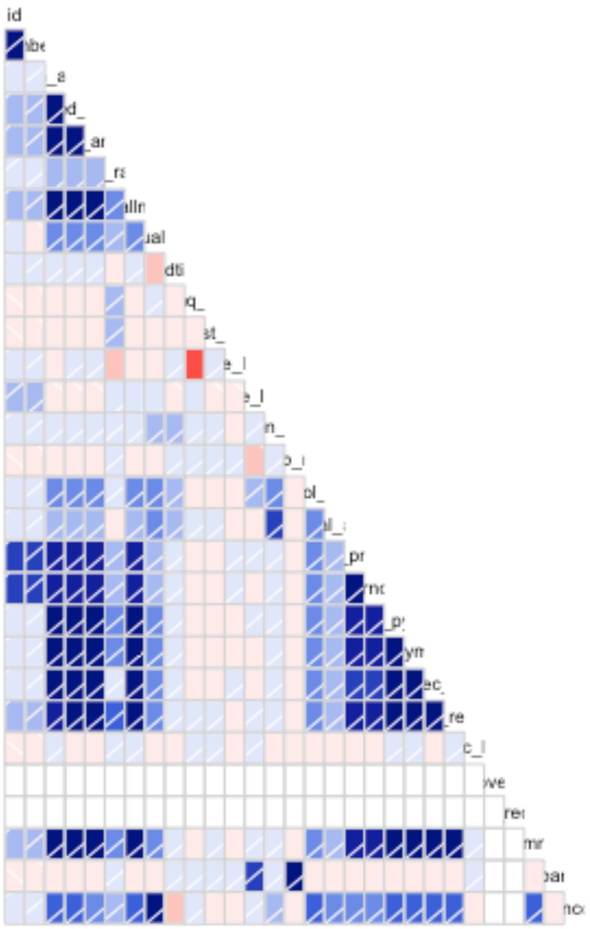


Final Correlation Spread Matrix of Segmented Dataset

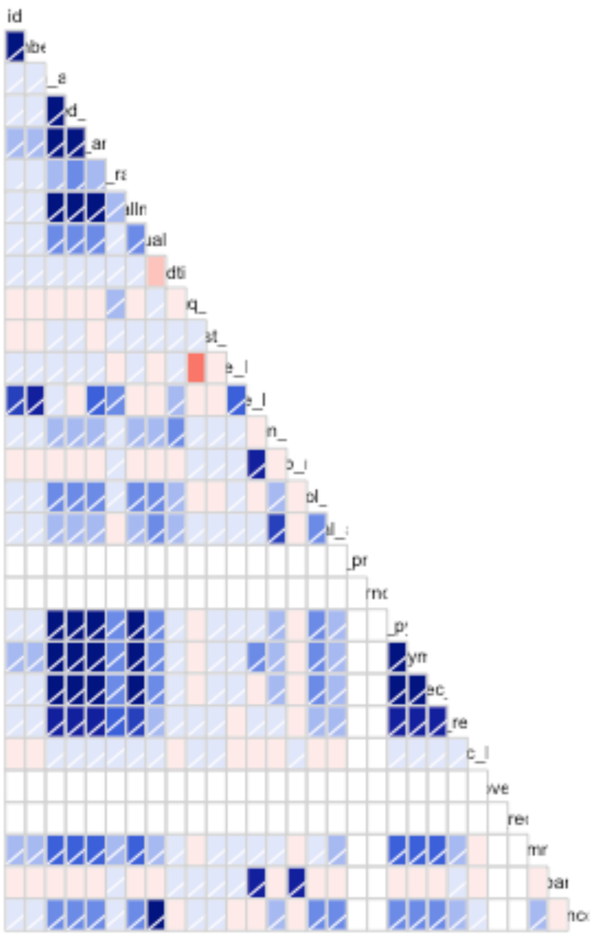
Charged off spread over variables



Current spread over variables



Fully Paid spread over variables



FINAL CONCLUDING RESULTS:

- We saw that the charged off category is the major risk blocking the business to company and generating credit loss, (RED)
- 10000-20000 is the danger region where high defaulters are generally found.
- The count for defaulters less than 10000 and greater than 25000 is nearly the same and hence show a regualar pattern .
- Loan funded above 25000 are taken great care of bank as the ratio of current and charged off is equally maintained and is less in count.