Customer Finance Company

CASE STUDY

SUBMISSION

Group Members Name:

1. Thanigachalam TamizhSelvam
2. Swati Bansal
3. Vishal Gunasekaran
4. Rajanish Mirajkar

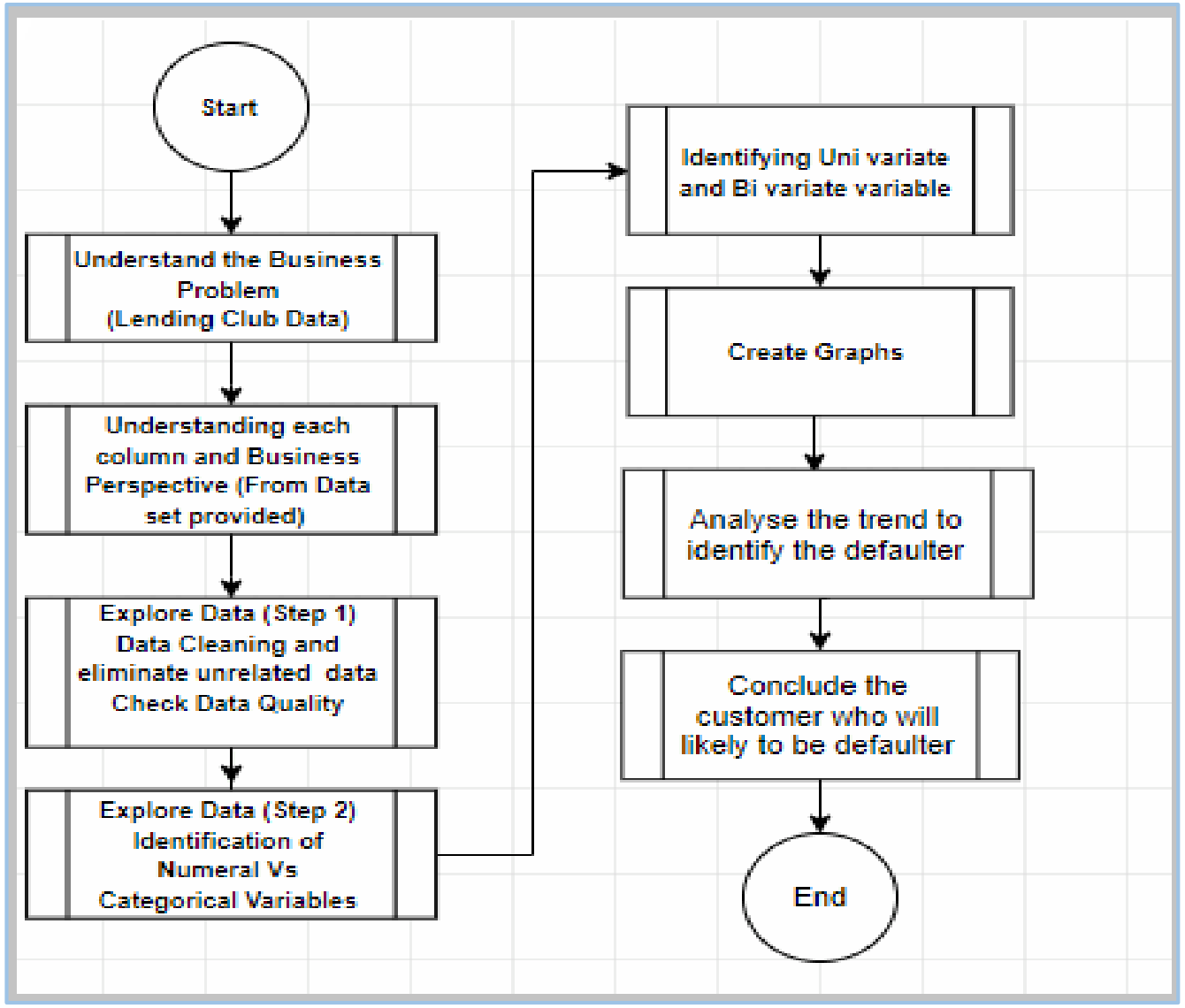
Business Objective

Consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. 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

The *Objective* of analysis 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

Flow chart


Start
Understand the Business Problem (Lending Club Data)
Understanding each column and Business Perspective (From Data set provided)
Explore Data (Step 1) Data Cleaning and eliminate unrelated data Check Data Quality
Explore Data (Step 2) Identification of Numeral Vs Categorical Variables
Identifying Uni variate and Bi variate variable
Create Graphs
Analyse the trend to identify the defaulter
Conclude the customer who will likely to be defaulter
End

Data Exploration

Data set contains 111 columns and below major columns considered for analysis

loan_amnt	verification_status
funded_amnt_inv	issue_d
term	loan_status
int_rate	purpose
grade	title
sub_grade	addr_state
emp_title	dti
emp_length	delinq_2yrs
home_ownership	inq_last_6mths
annual_inc	total_pymnt_inv
funded_amnt	

Problem Statement: Understand how **consumer attributes** and **loan attributes** influence the tendency of default.

Data Cleaning and Manipulation

1. Data Cleaning

- Removing % sign from "**int_rate"** column for using for better analysis.
- Handling Date columns: Derived new columns Month and Year from **"issued_d_month" and "issued_d_year" columns.**

2. Derived Columns

- A new numeric **emp_length_value** column is derived from emp_length_value field

  (By making assumption as < 1 year : considered as a 0.5 and 10+ years : 11 years)
- Derived new fields **region_value** and **state name** with the help of state code field from data set
- Derived new field **loan_amt_range** _bin with loan amount field(Between 50000-40000)
- Derived new field **annual_income_bin** with annual_inc field (Between 10000-100000)
- Derived new field **int_rate_bin** with int_rate field (Between 10000-100000)

Data Cleaning and Manipulation Continued

3. Removed Columns

There are around 56 columns are removed from given data set

mths_since_last_major_derog	mo_sin_old_il_acct	total_rev_hi_lim	num_rev_tl_bal_gt_0
annual_inc_joint	mo_sin_old_rev_tl_op	inq_fi	num_sats
dti_joint	mo_sin_rcnt_rev_tl_op	total_cu_tl	num_tl_120dpd_2m
verification_status_joint	mo_sin_rcnt_tl	inq_last_12m	num_tl_30dpd
tot_coll_amt	mort_acc	acc_open_past_24mths	num_tl_90g_dpd_24m
tot_cur_bal	mths_since_recent_bc	avg_cur_bal	num_tl_op_past_12m
open_acc_6m	mths_since_recent_bc_dlq	bc_open_to_buy	pct_tl_nvr_dlq
open_il_6m	mths_since_recent_inq	bc_util	percent_bc_gt_75
open_il_12m	mths_since_recent_revol_delinq	chargeoff_within_12_mths	tax_liens
open_il_24m	num_accts_ever_120_pd	delinq_amnt	tot_hi_cred_lim
mths_since_rcnt_il	num_actv_bc_tl	total_bc_limit	total_bal_ex_mort
total_bal_il	num_actv_rev_tl	total_il_high_credit_limit	
il_util	num_bc_sats	open_rv_24m	num_il_tl
open_rv_12m	num_bc_tl	max_bal_bc	num_op_rev_tl
all_util	num_rev_accts		

More insight into removed fields

1.Unique data so, this will not for any analysis.

policy_code (ex: In dataset its only policy code 1 values are shared)
application_type (ex: Individual)

2.There is no Joint application type data, so we can remove the below columns.

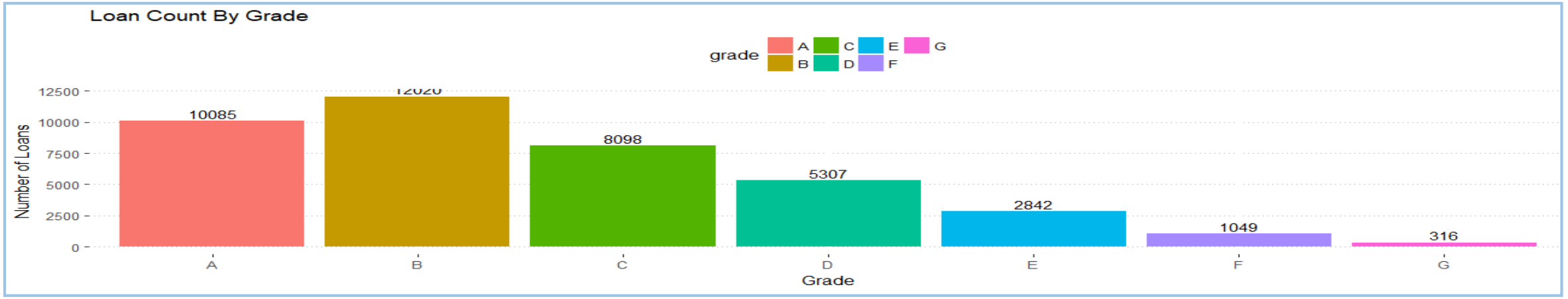
annual_inc_joint
dti_joint
verification_status_joint

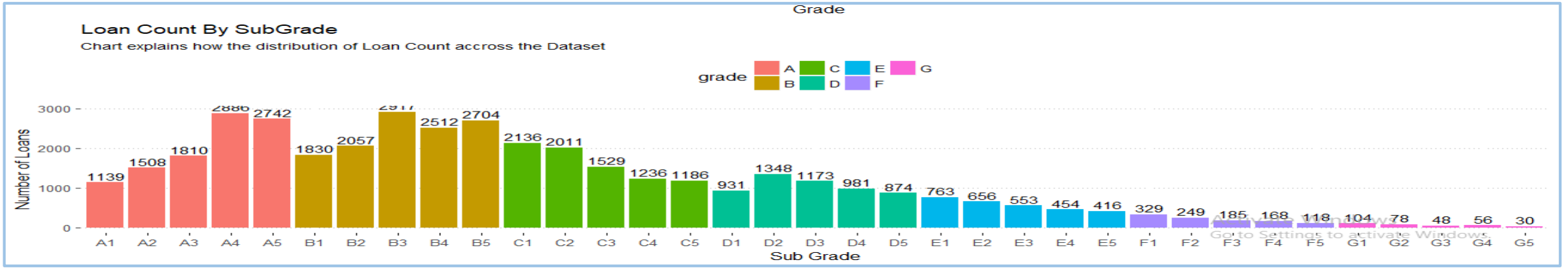
3.Column are contains only NA values, not making any reasons to add data.

acc_open_past_24mths	num_actv_rev_tl	total_bc_limit	tot_coll_amt
avg_cur_bal	num_bc_sats	total_il_high_credit_limit	tot_cur_bal
bc_open_to_buy	num_bc_tl	num_tl_90g_dpd_24m	open_acc_6m
bc_util	num_il_tl	num_tl_op_past_12m	open_rv_24m
chargeoff_within_12_mths	num_op_rev_tl	pct_tl_nvr_dlq	open_il_12m
delinq_amnt	num_rev_accts	percent_bc_gt_75	open_il_24m
mo_sin_old_il_acct	num_rev_tl_bal_gt_0	tax_liens	inq_fi
mo_sin_old_rev_tl_op	num_sats	tot_hi_cred_lim	max_bal_bc
mo_sin_rcnt_rev_tl_op	num_tl_120dpd_2m	total_bal_ex_mort	total_bal_il
mo_sin_rcnt_tl	num_tl_30dpd	num_actv_bc_tl	il_util
mort_acc	mths_since_recent_inq	mths_since_recent_revol_delinq	open_rv_12m
mths_since_recent_bc	num_accts_ever_120_pd	mths_since_recent_bc_dlq	all_util
acc_now_delinq	open_il_6m	mths_since_rcnt_il	

Analysis

Loan Count By Grade and Sub Grade

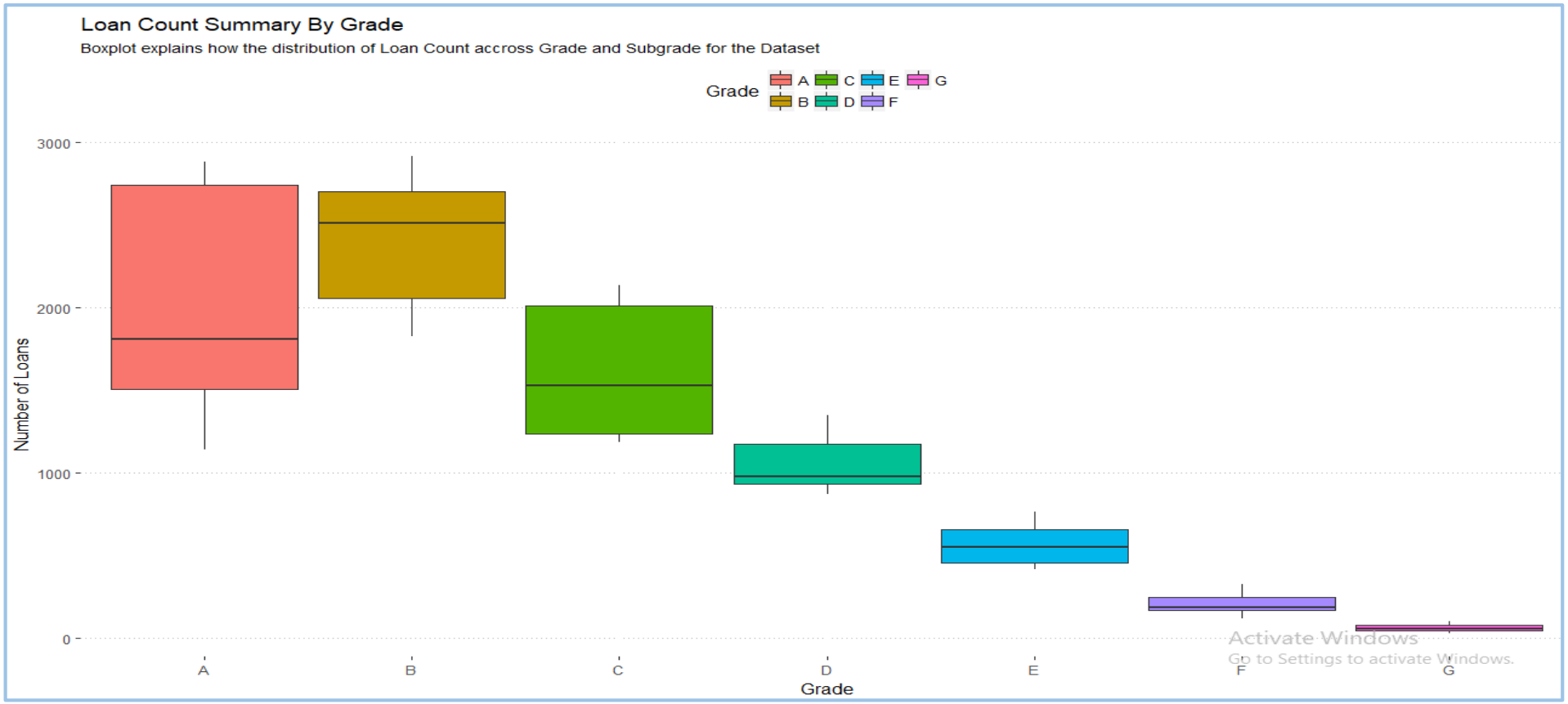




The above graph shows the loan count by grade ,i.e Distribution of loan with different grades. With the above graph it is clear that **Grade B** got highest amount of loan.

Plot 1

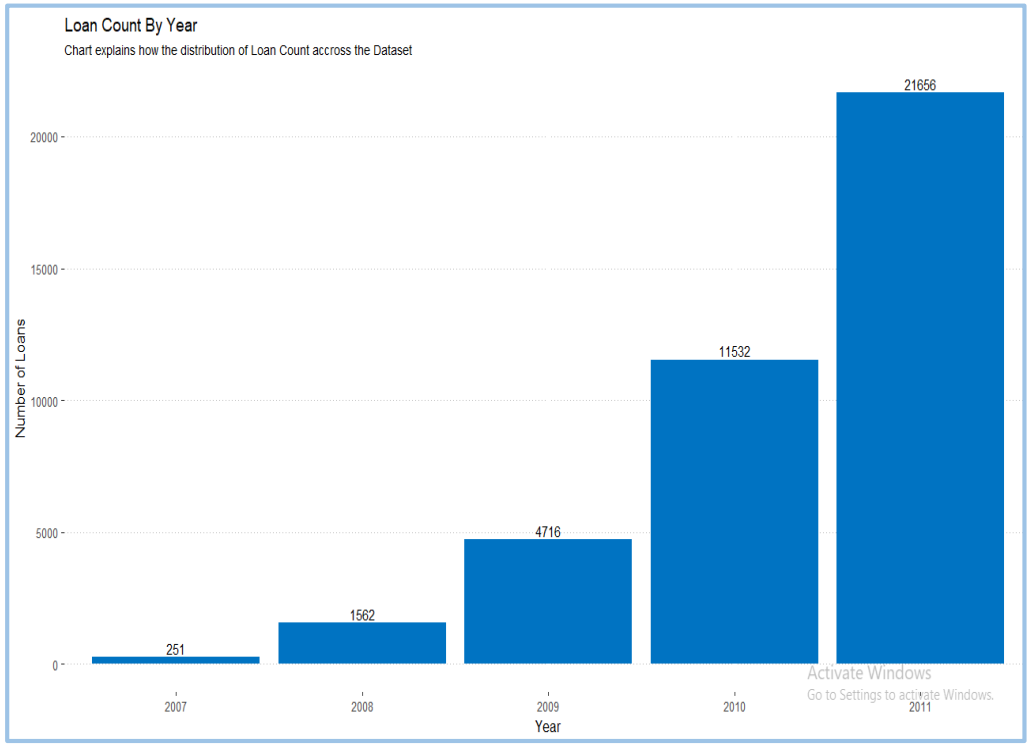
Loan Count By Grade and Sub Grade

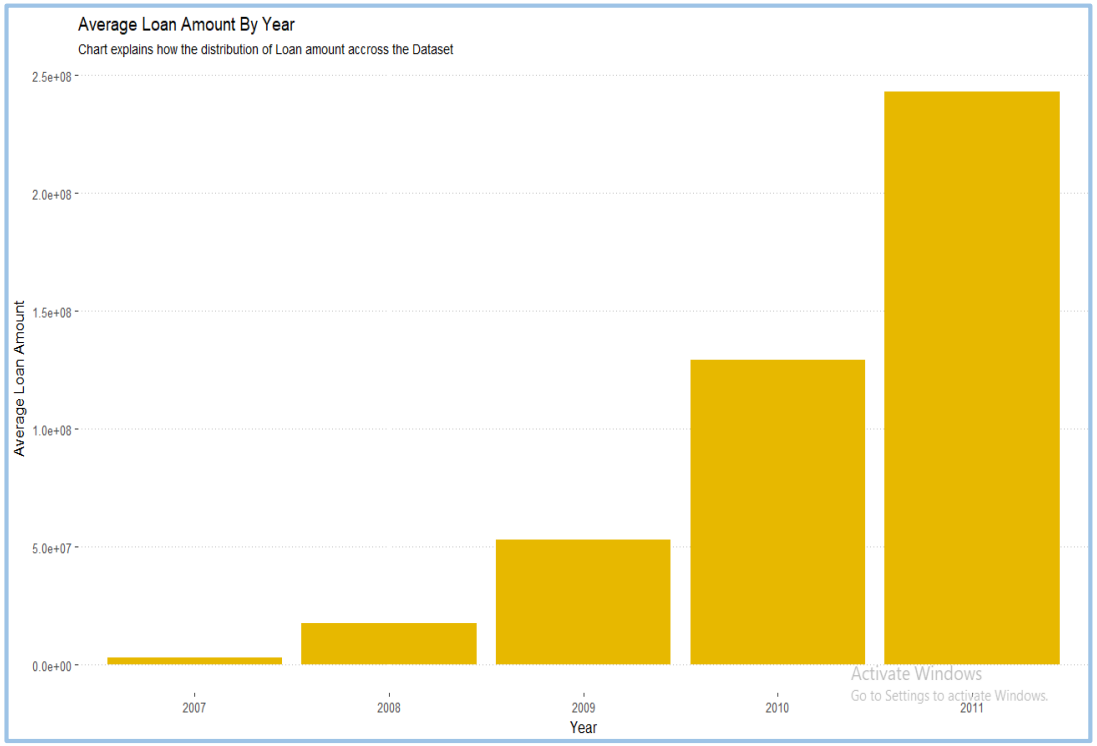


The distribution of loan presented using Box Plot

Plot2

Distribution of loan Yearly

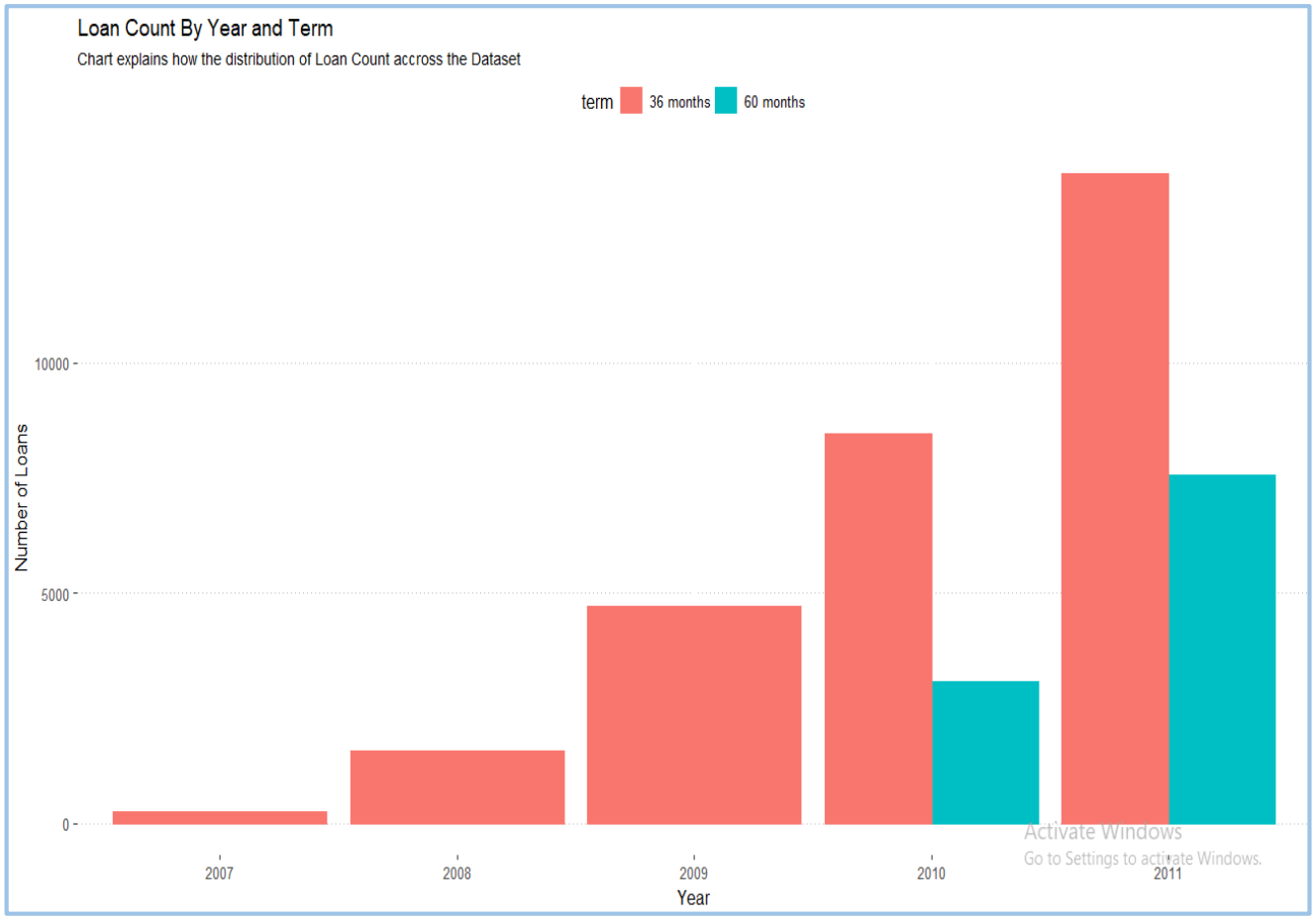




- The above graphs depicts that count of loan distribution and average of loan distribution.
- It is clear that in lend amount in 2011 is high.
- Also the graph shows the lending of loan is increasing across the years.

Plot3

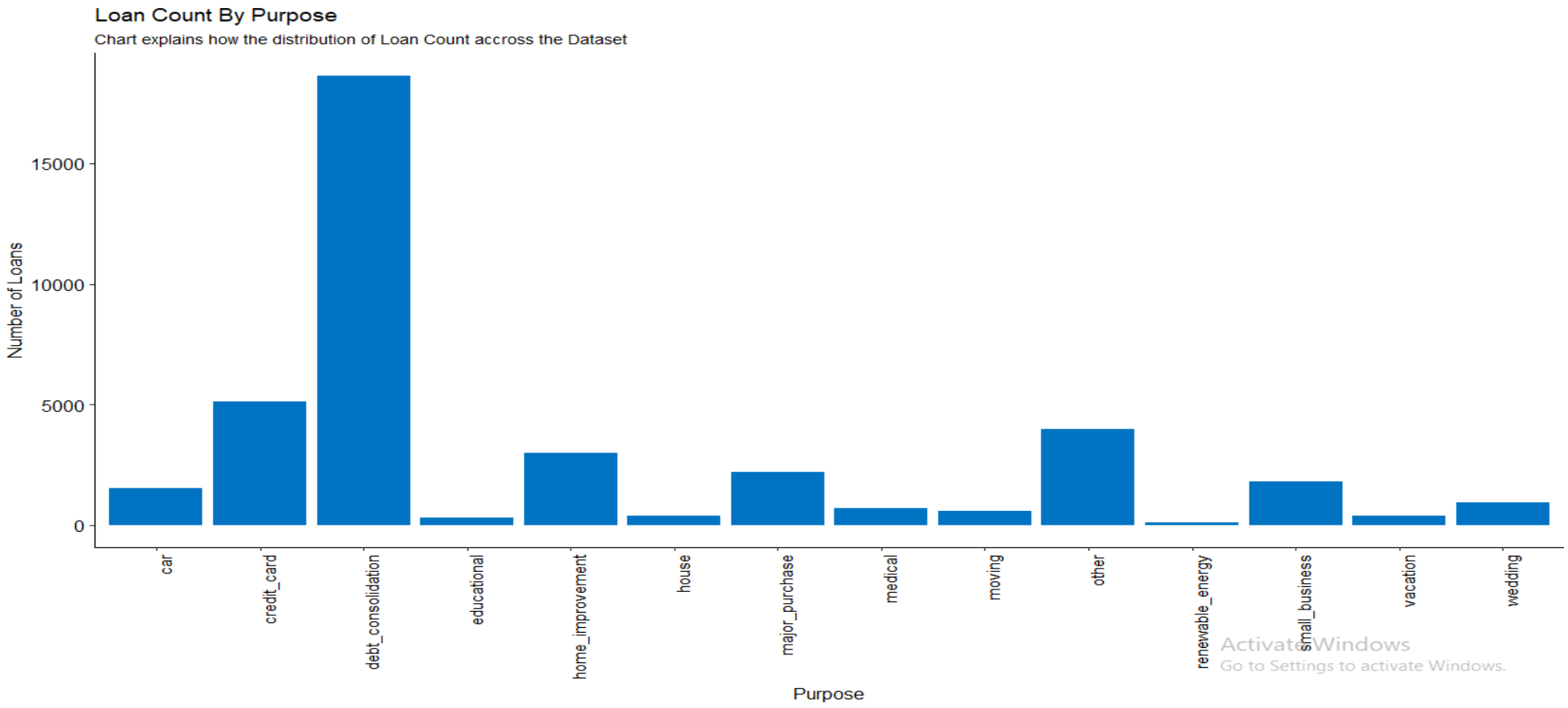
Loan Count By Year and Term



The graph shows data about the loan term for 36 and 60 months.

Plot4

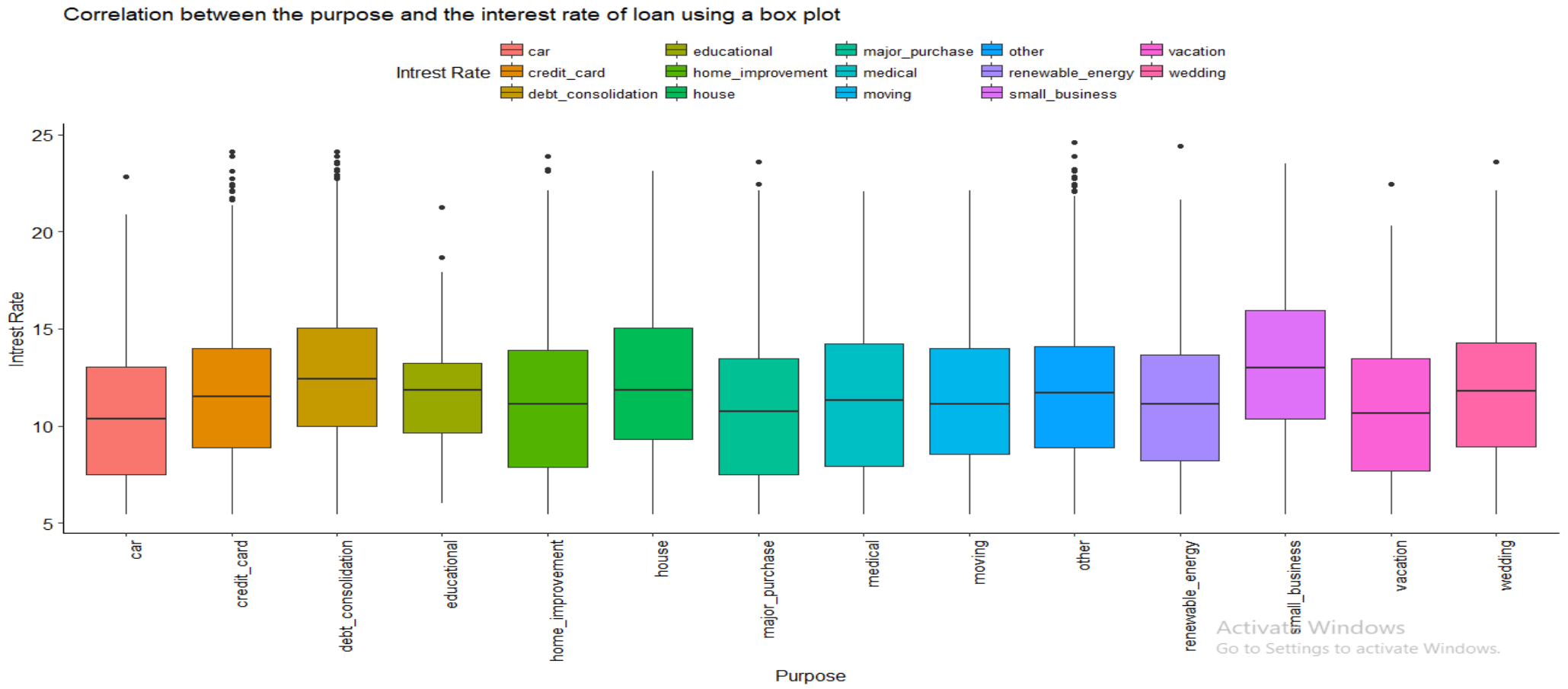
Loan count by Purpose



The above graph shows Debt consolidation is the most requested loan purpose type.

Plot5

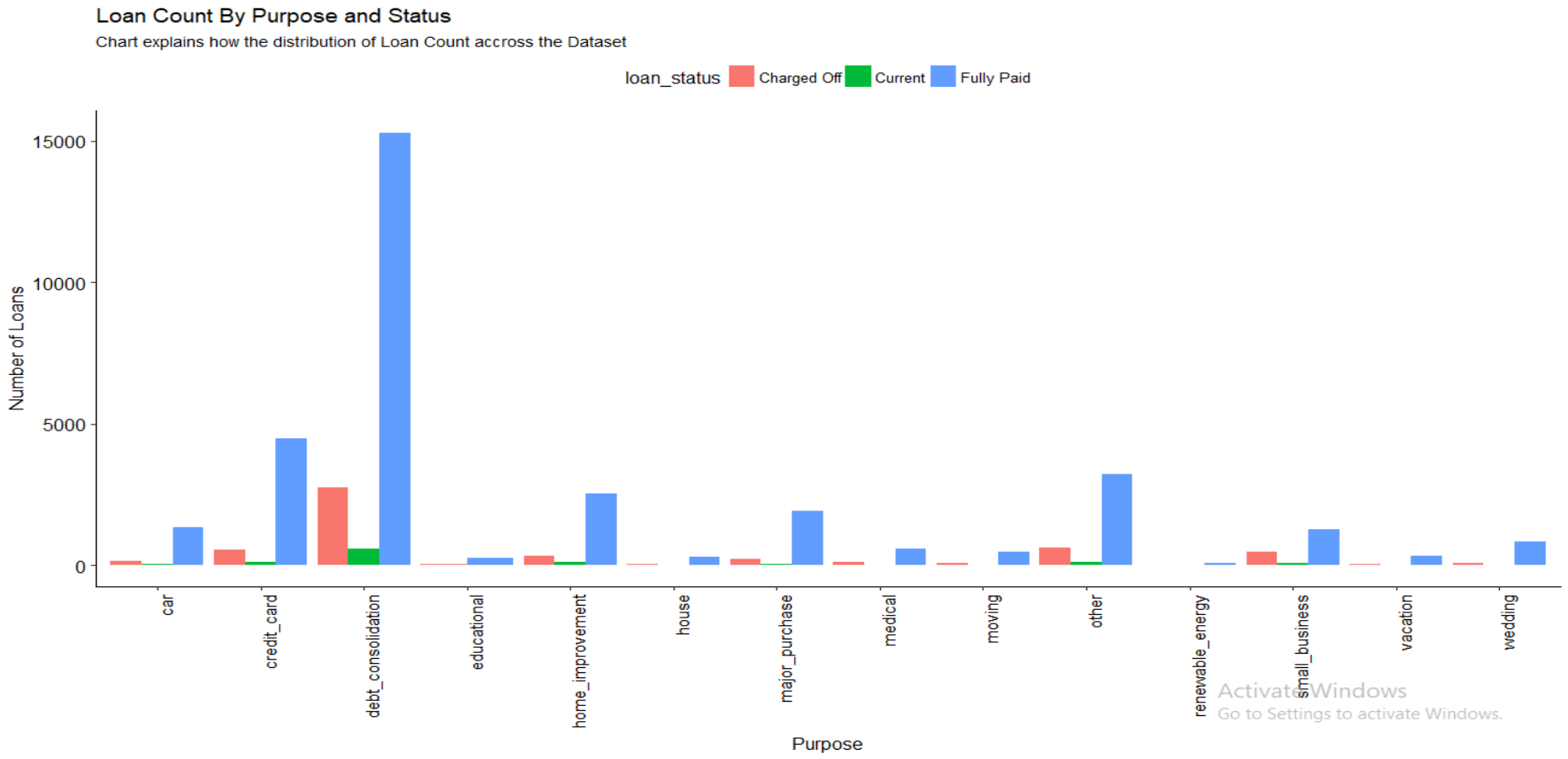
Correlation –Purpose and Interest rate



The graph shows correlation between the purpose and interest.

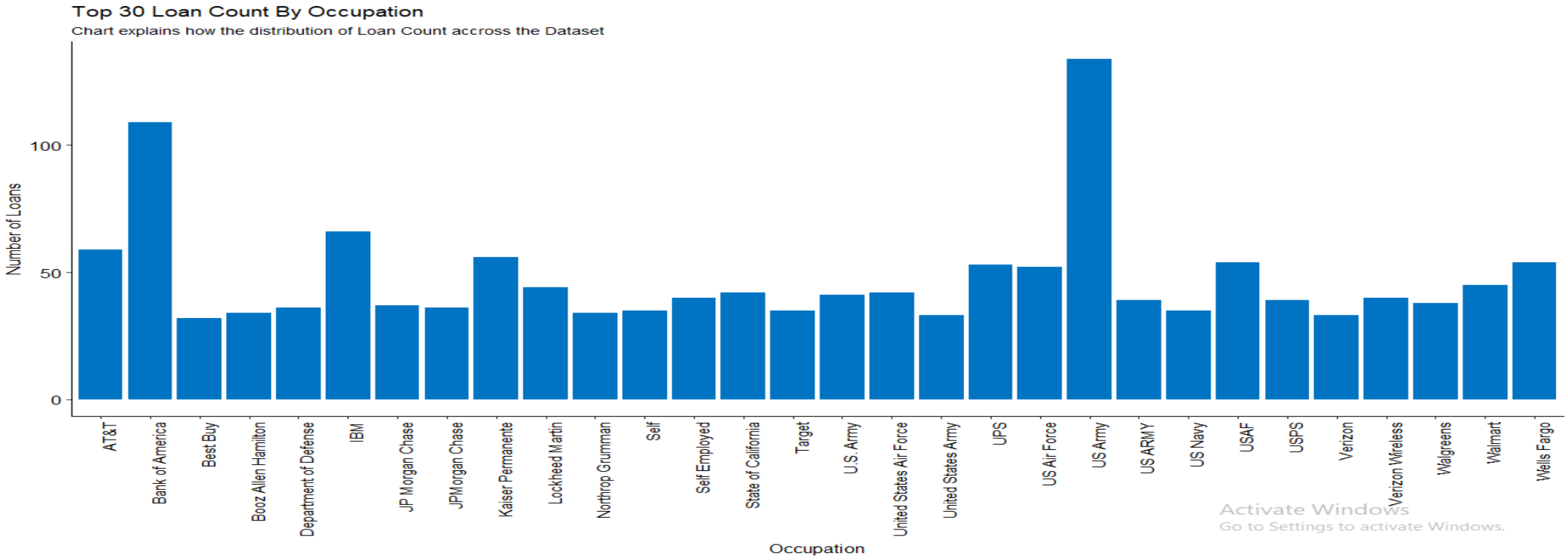
Plot6

Loan Count By Purpose and Status



Plot7

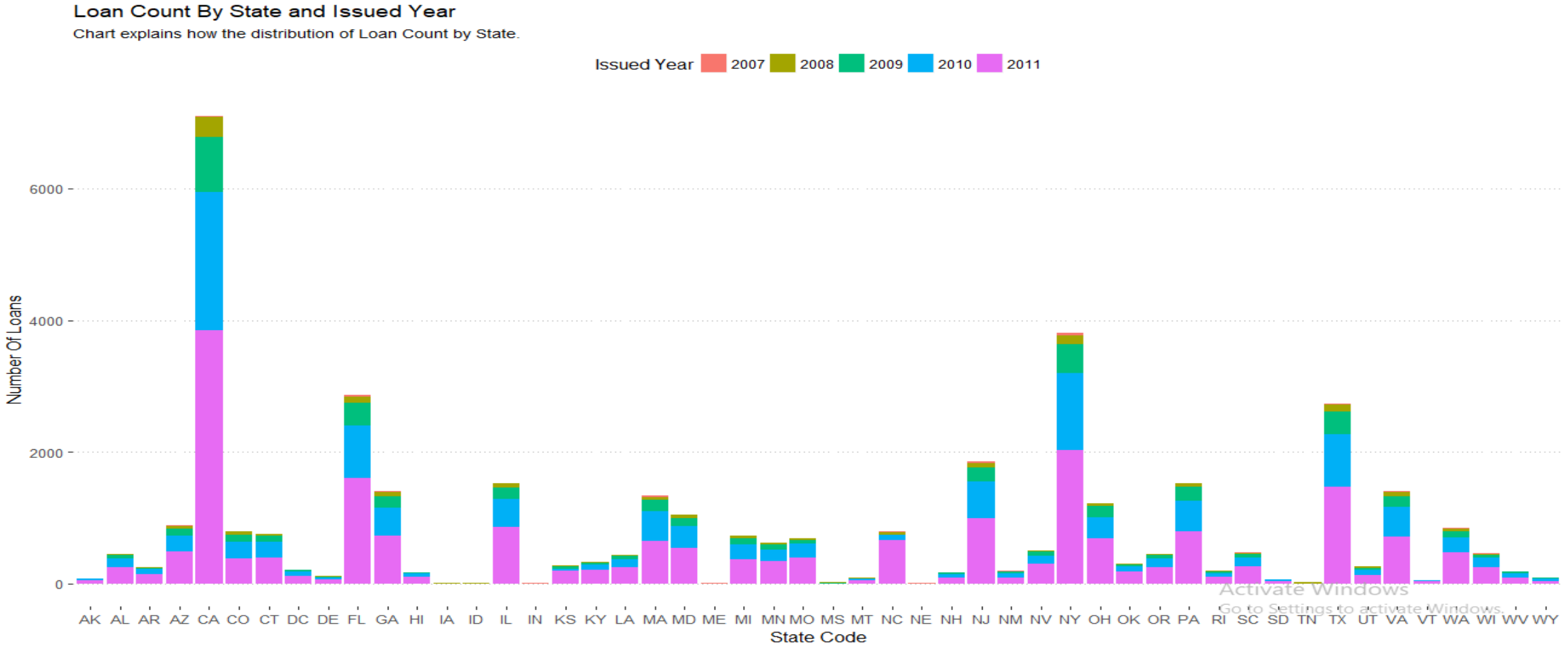
Top 30 Loan count by Occupation



The above graph shows distribution of loan by occupation, and US army and Bank occupation have more request of loan.

Plot 8

Loan Count By State and Issued Year

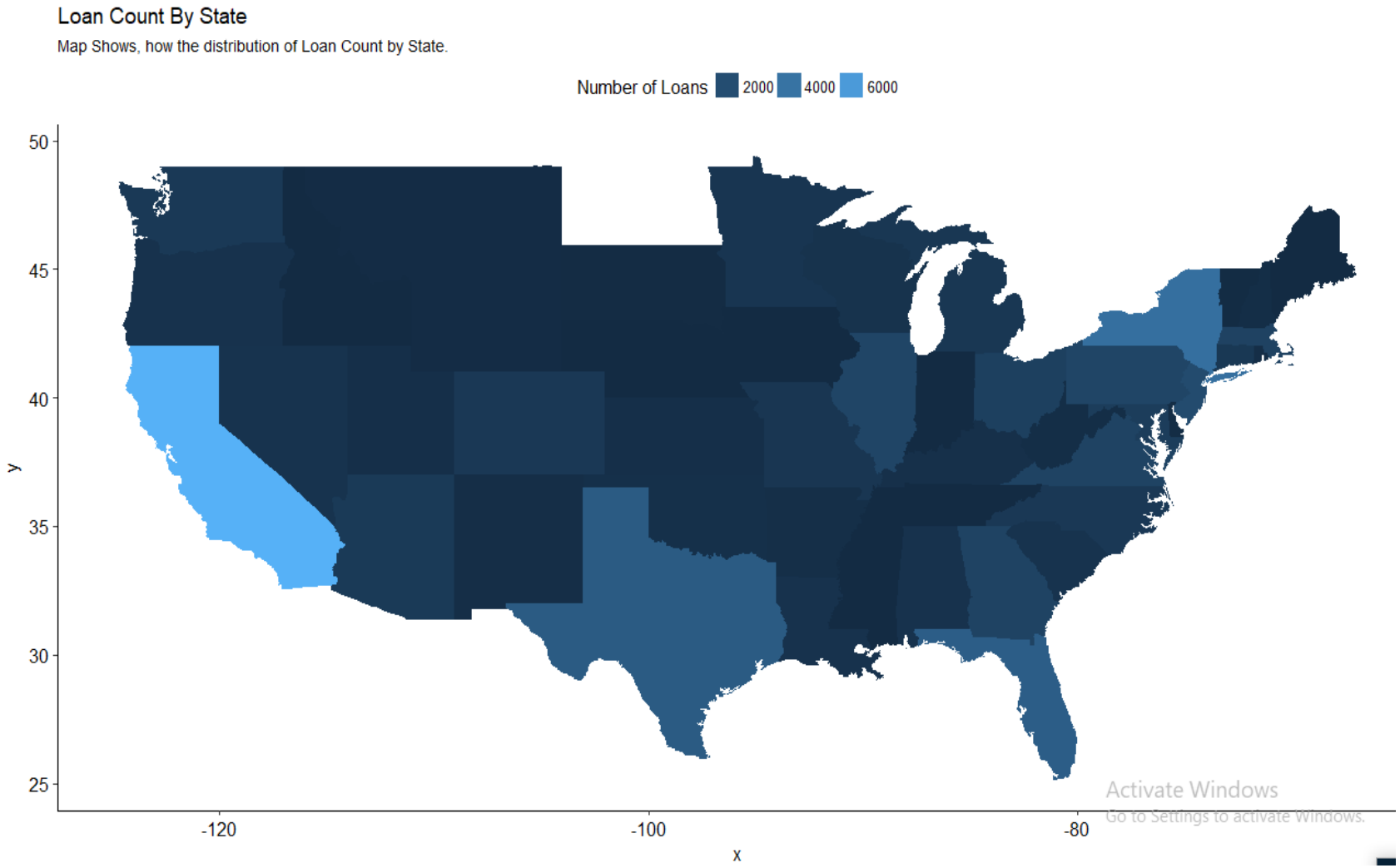


Plot 9

The Graph shows the loan count by state and issued year and the customer from CA state got more loan across the year

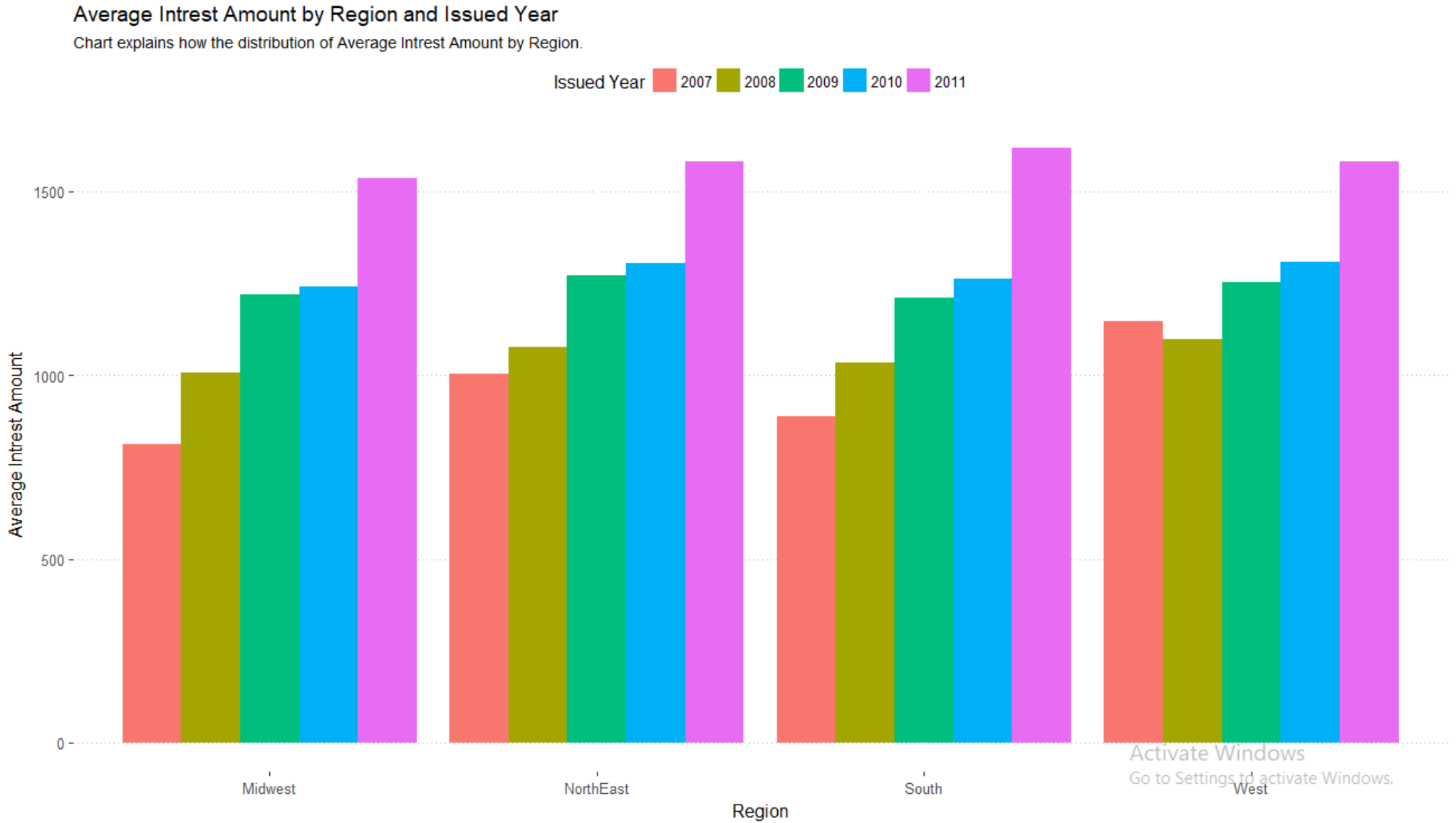
Loan count by state

The Map shows the loan count by state depicts as California state got more loan.



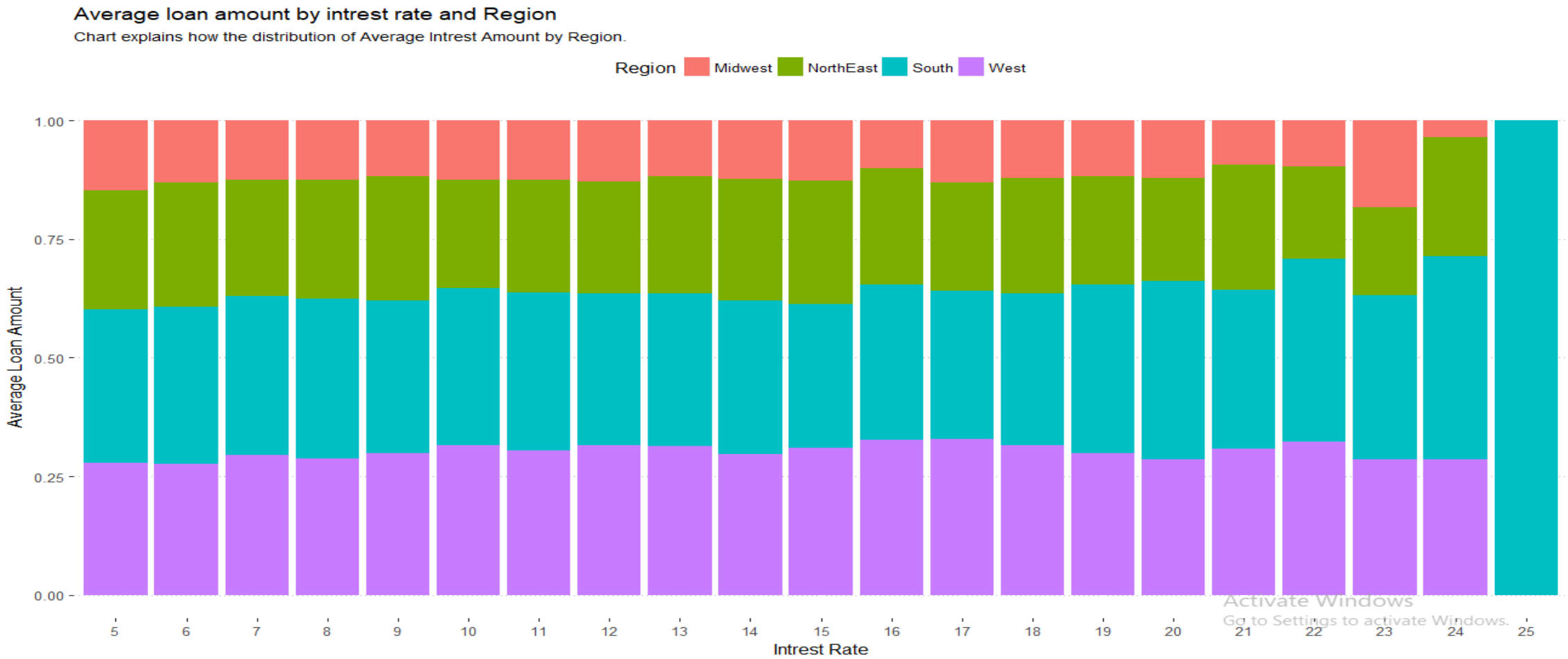
Plot 10

Average Interest Amount by Region and Issued Year



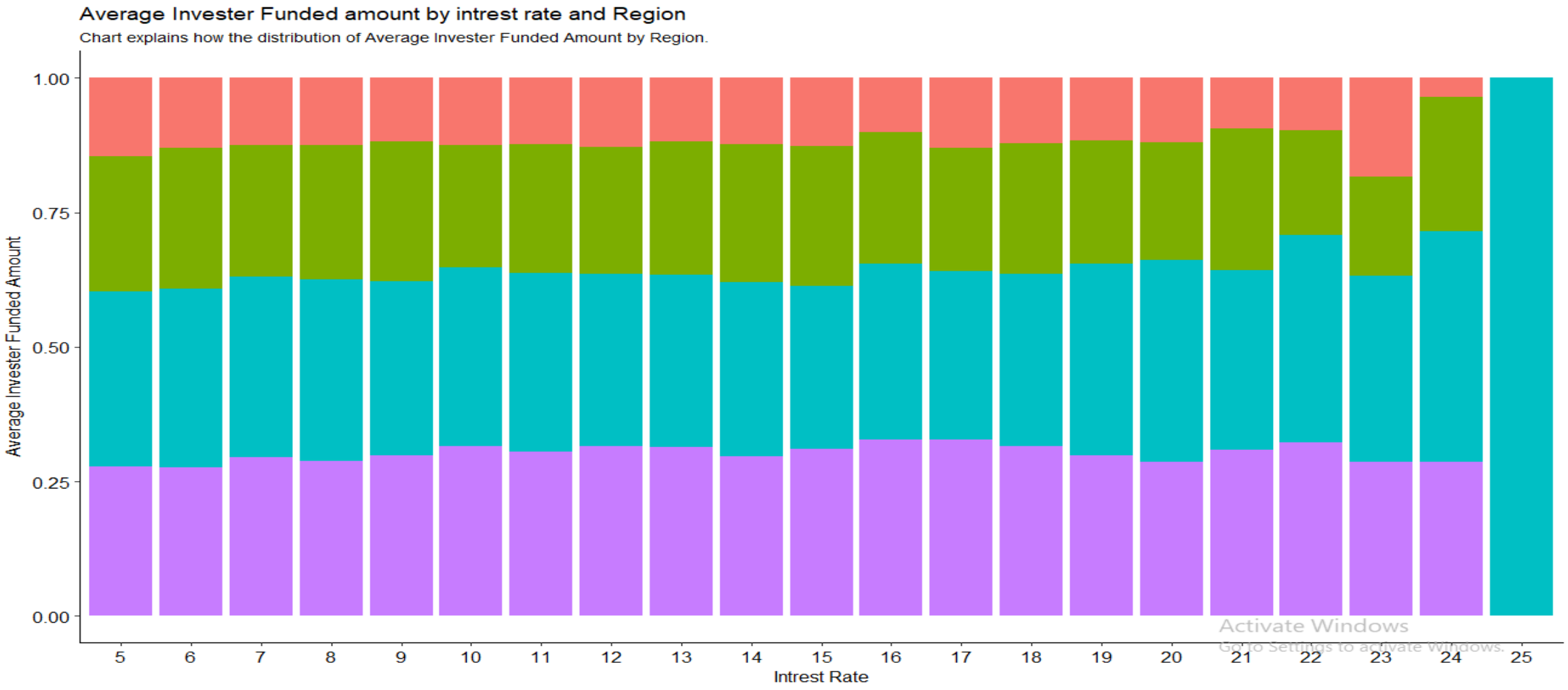
Plot 11

Average loan amount by Intrest rate and region



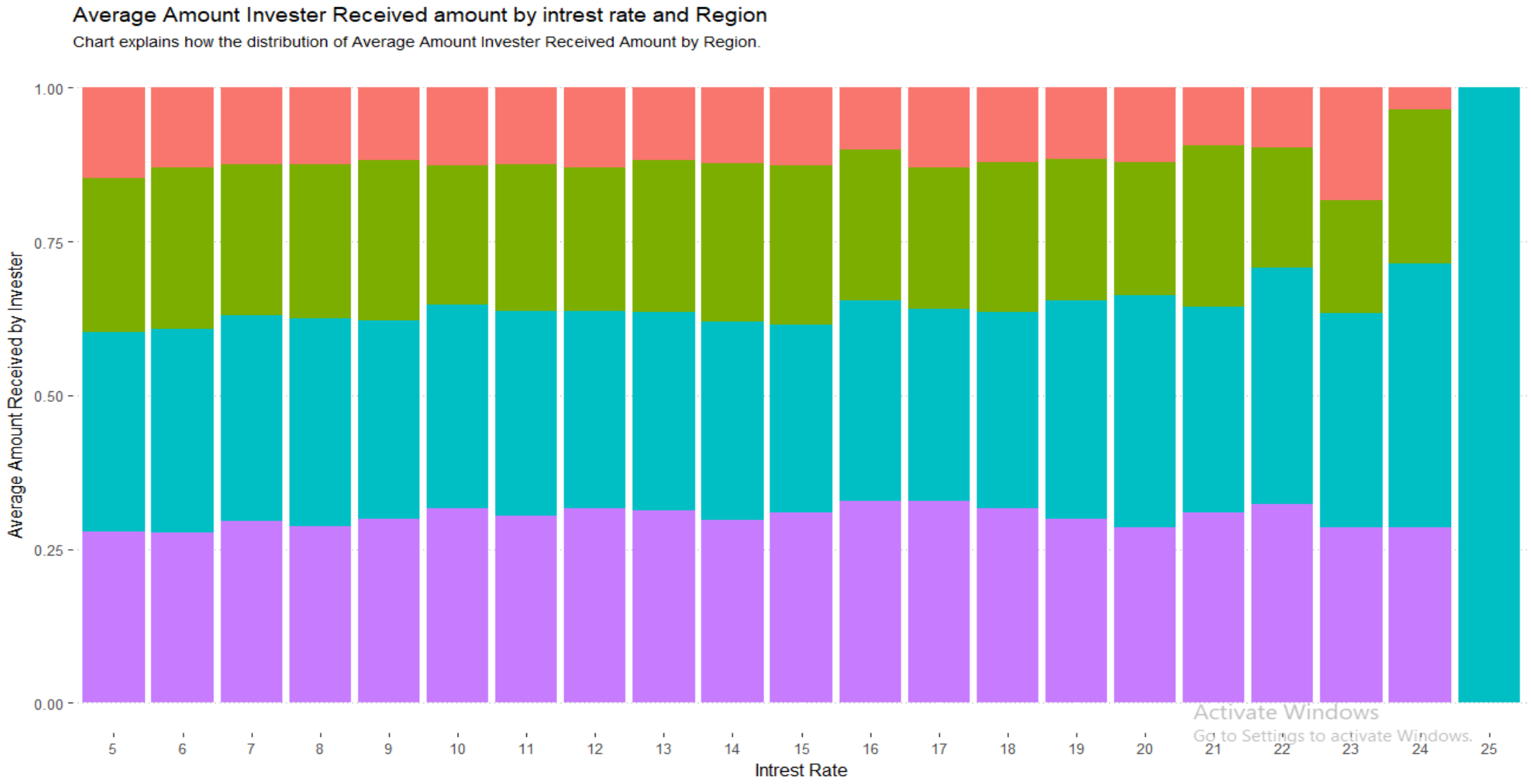
Plot 12

Average Investor Funded amount by Interest rate and Region



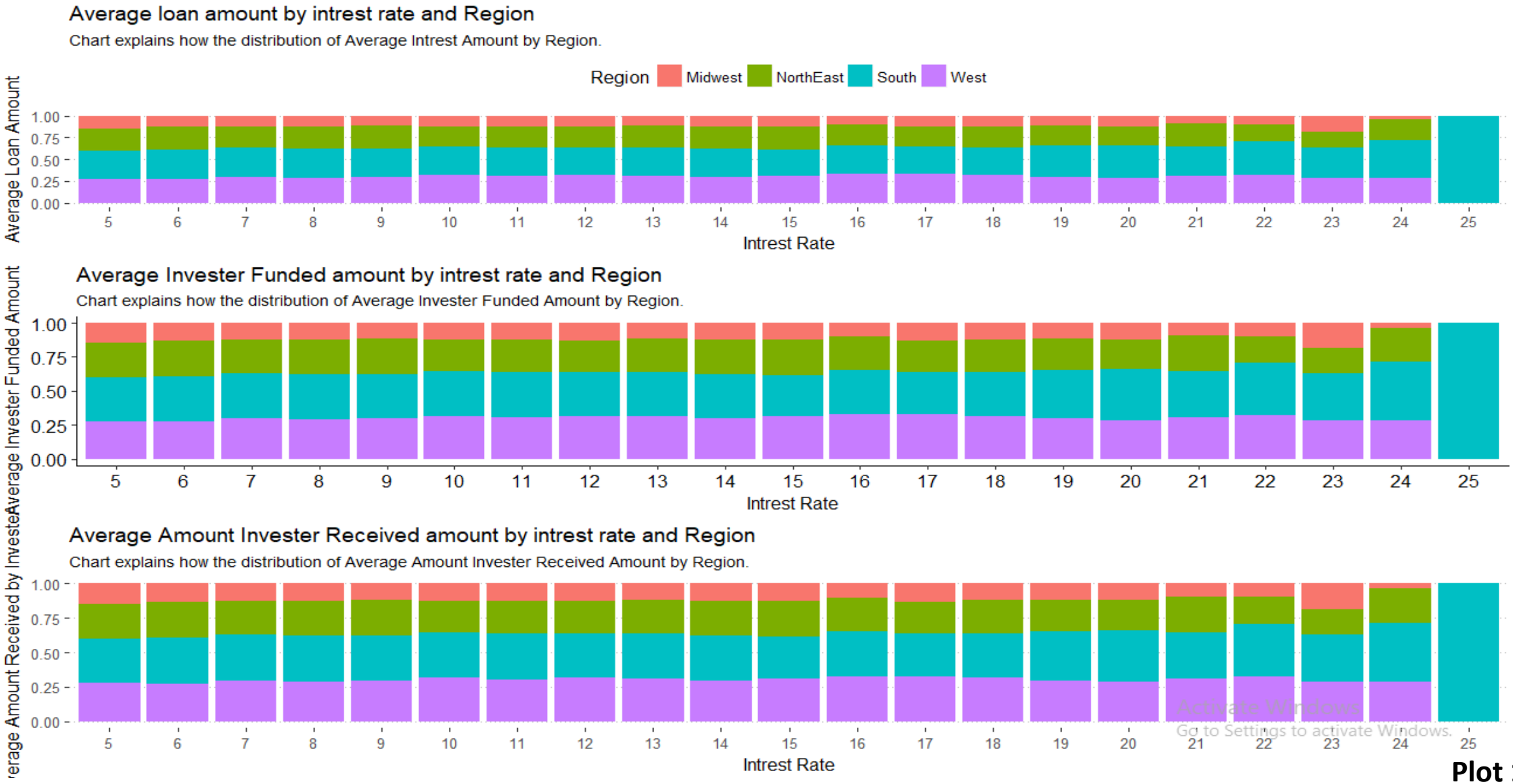
Plot 13

Average Amount Investor Received amount by Interest and Region



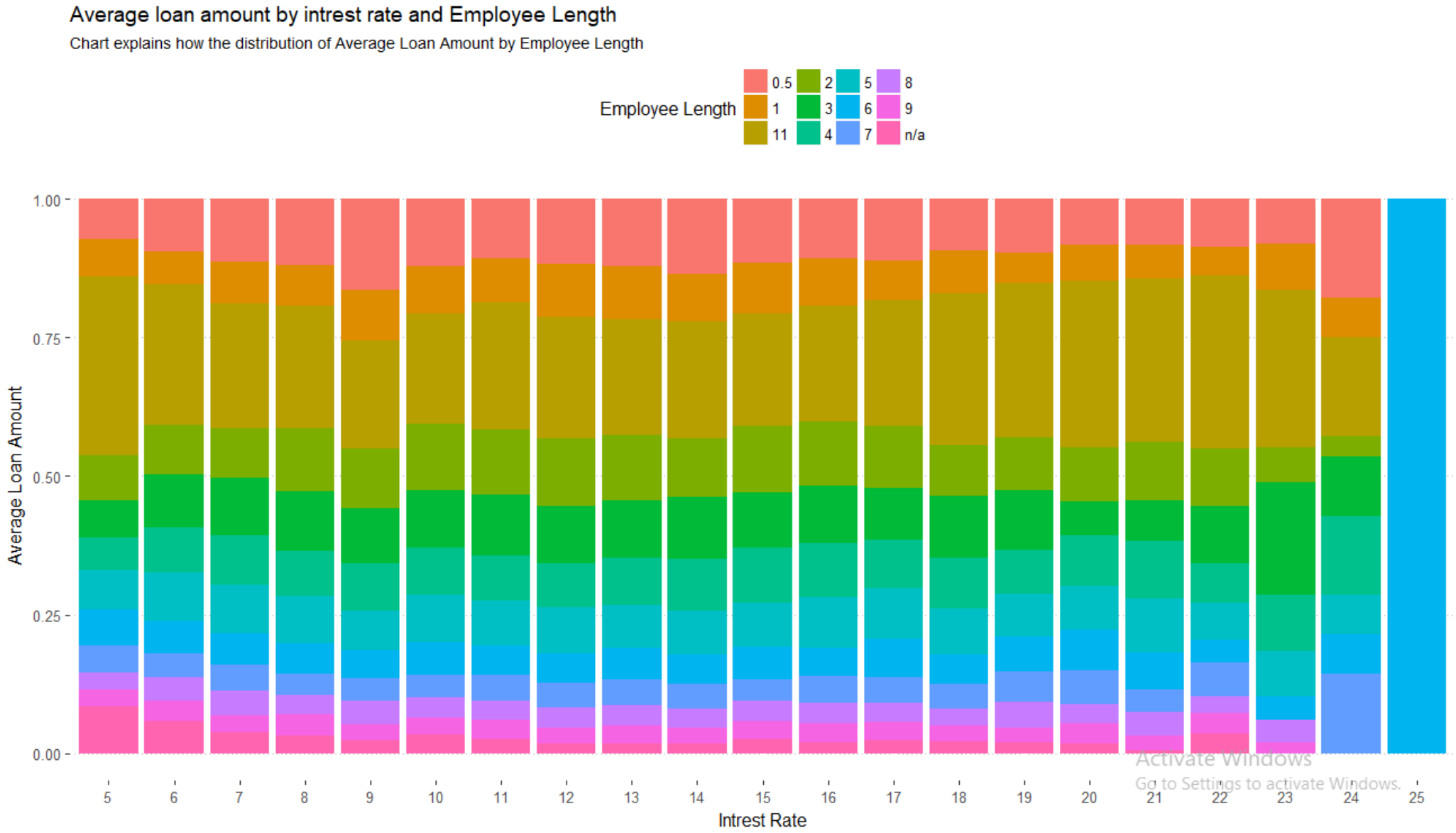
Plot 14

Average loan amount by interest rate and Region



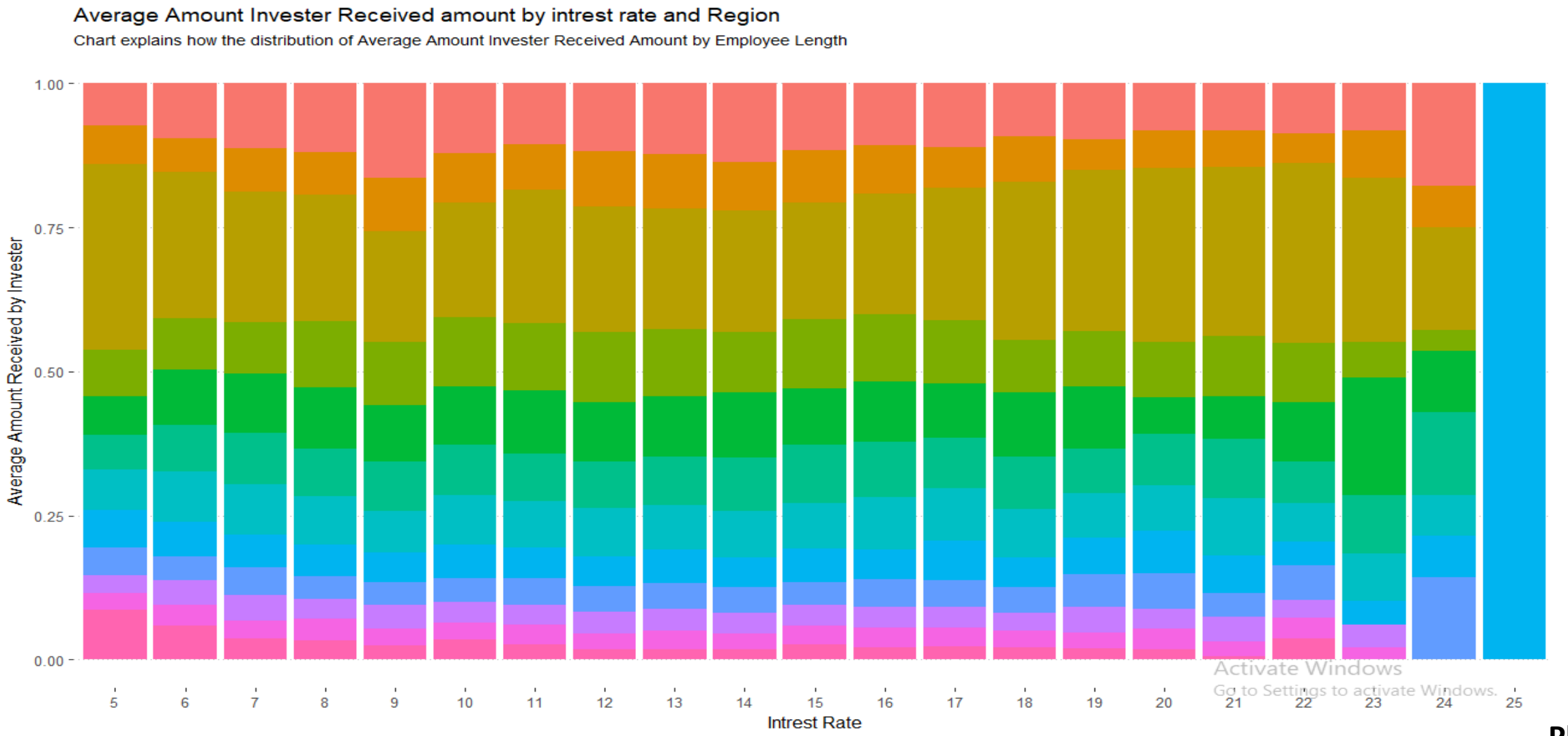
Plot 15

Average loan amount by Interest rate and Employee length



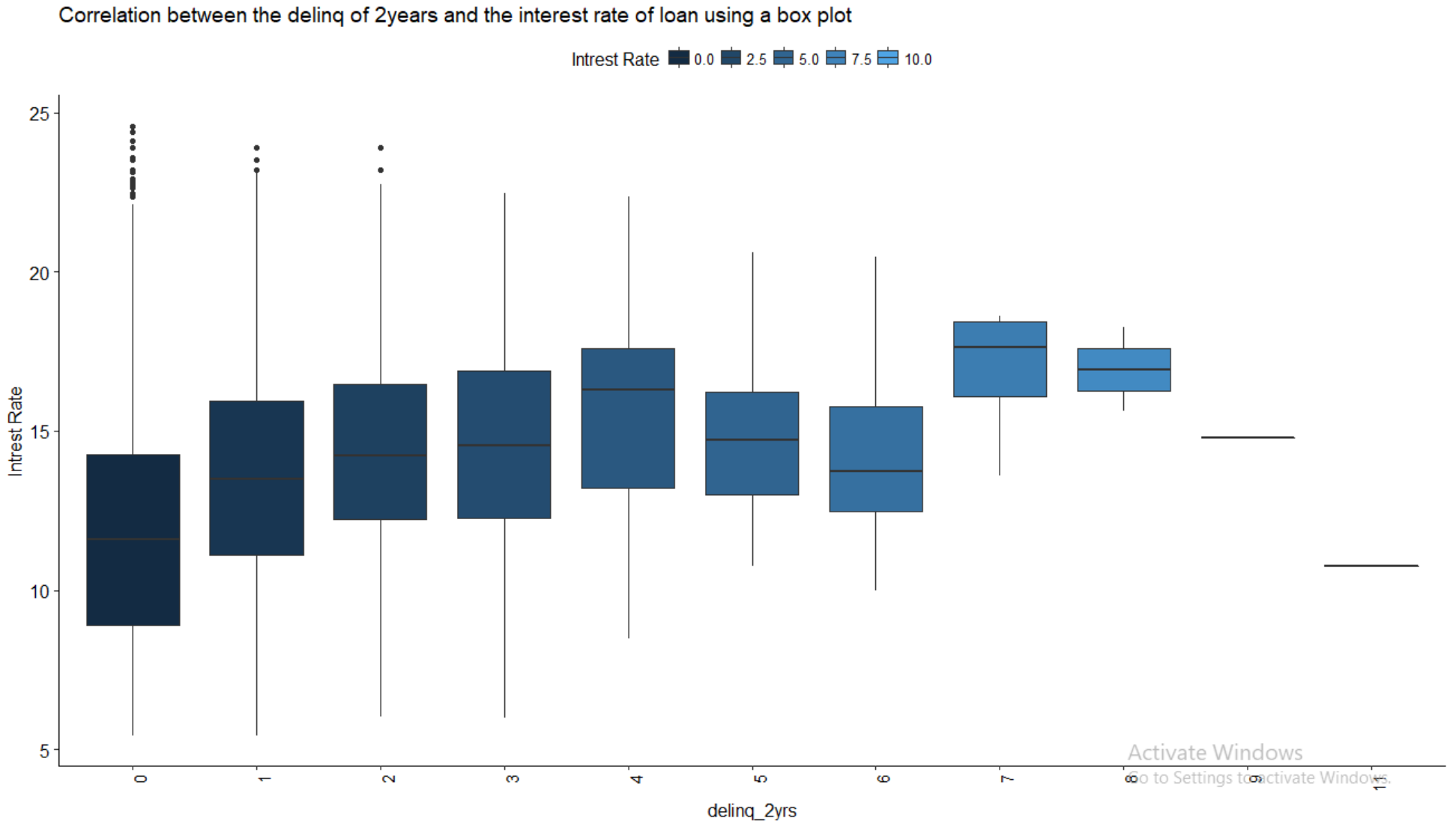
Plot 16

Average Amount Investor Received amount by Interest rate and Region



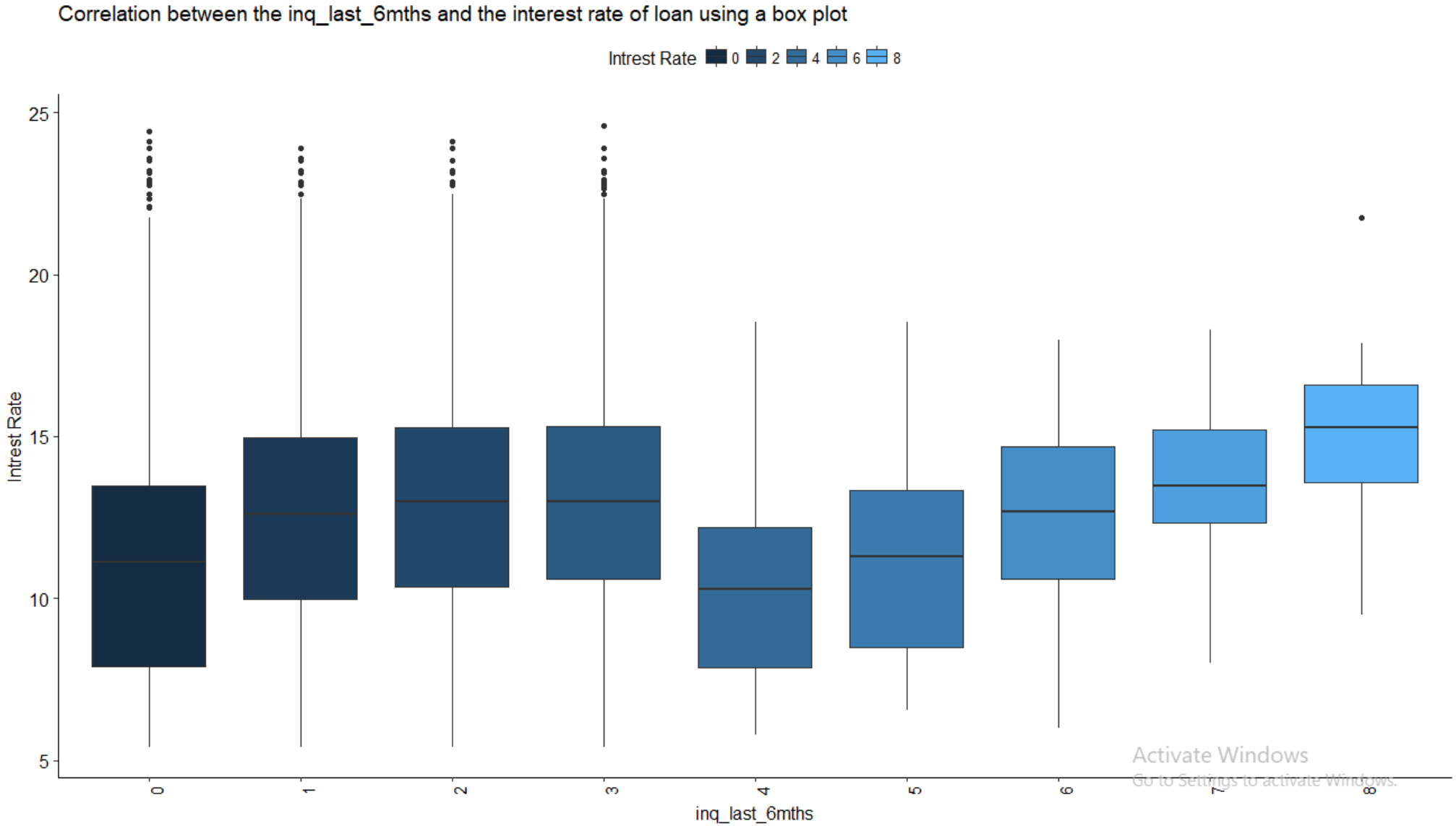
Plot 17

Correlation between the delinq of 2 years and the interest rate of loan

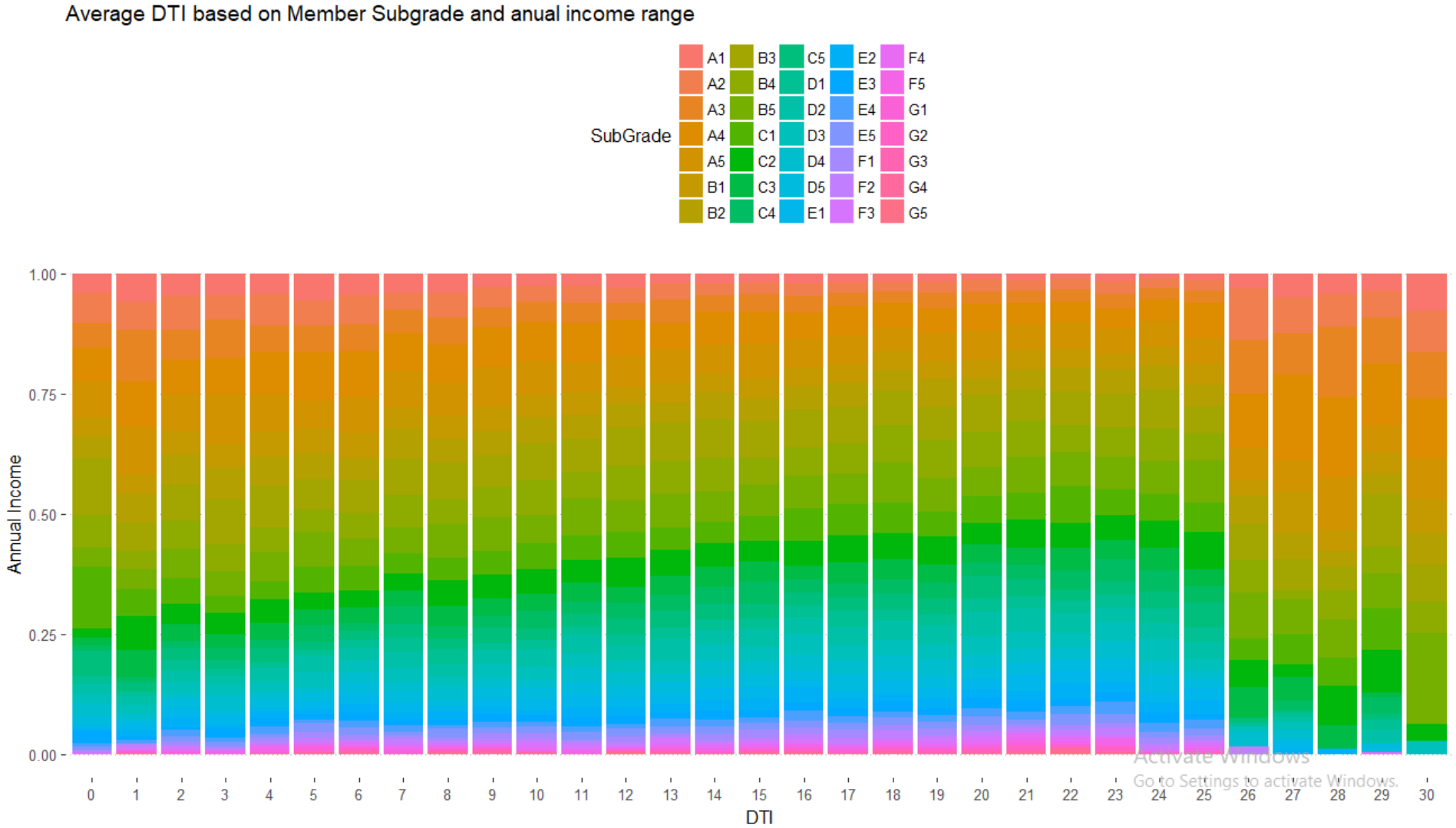


Plot 18

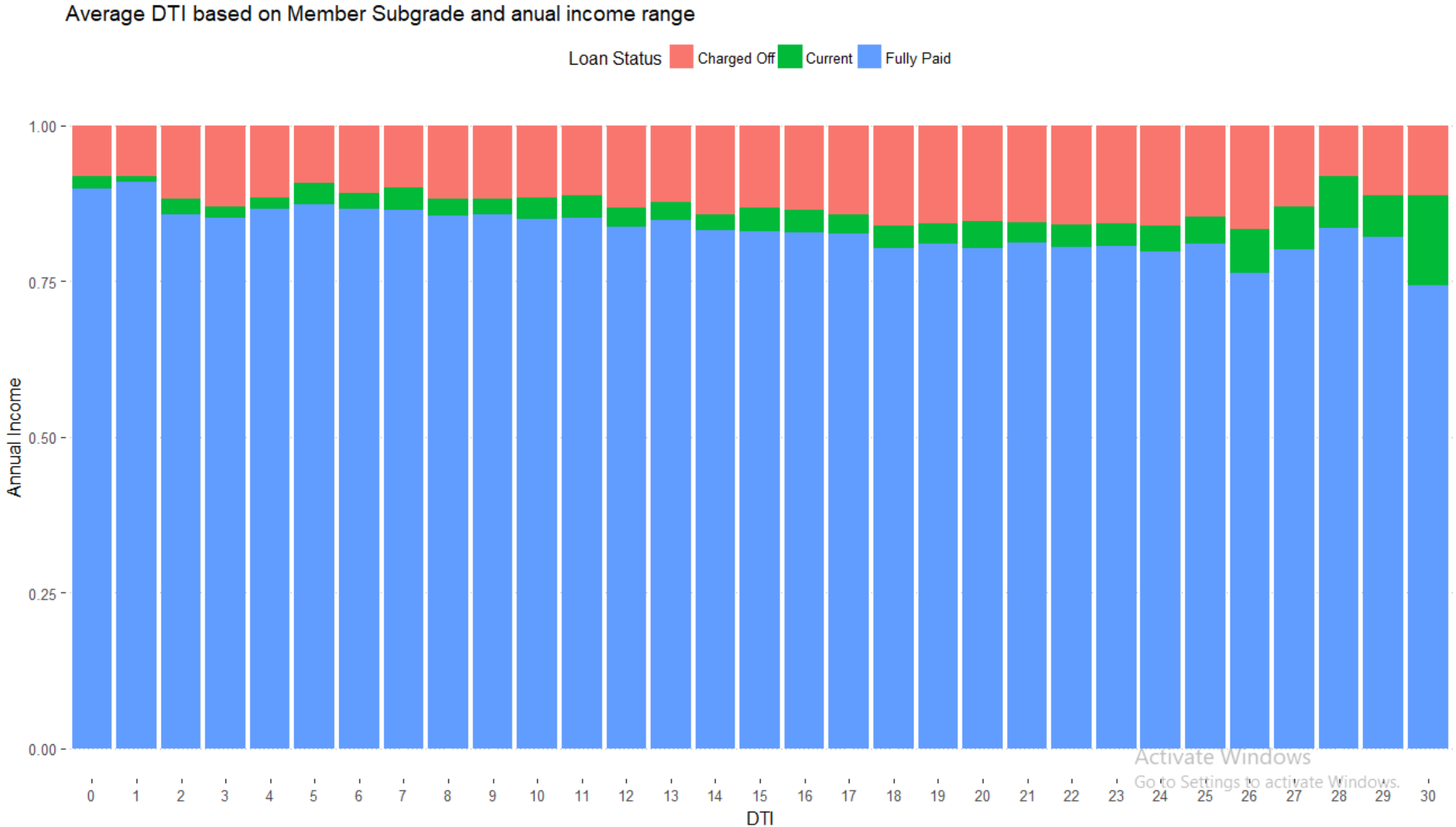
Correlation between the Inq_last_6 months And Interest rate



Plot 19


Average DTI based on Member Subgrade and anual income range
SubGrade
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
Annual Income
1.00
0.75
0.50
0.25
0.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
DTI

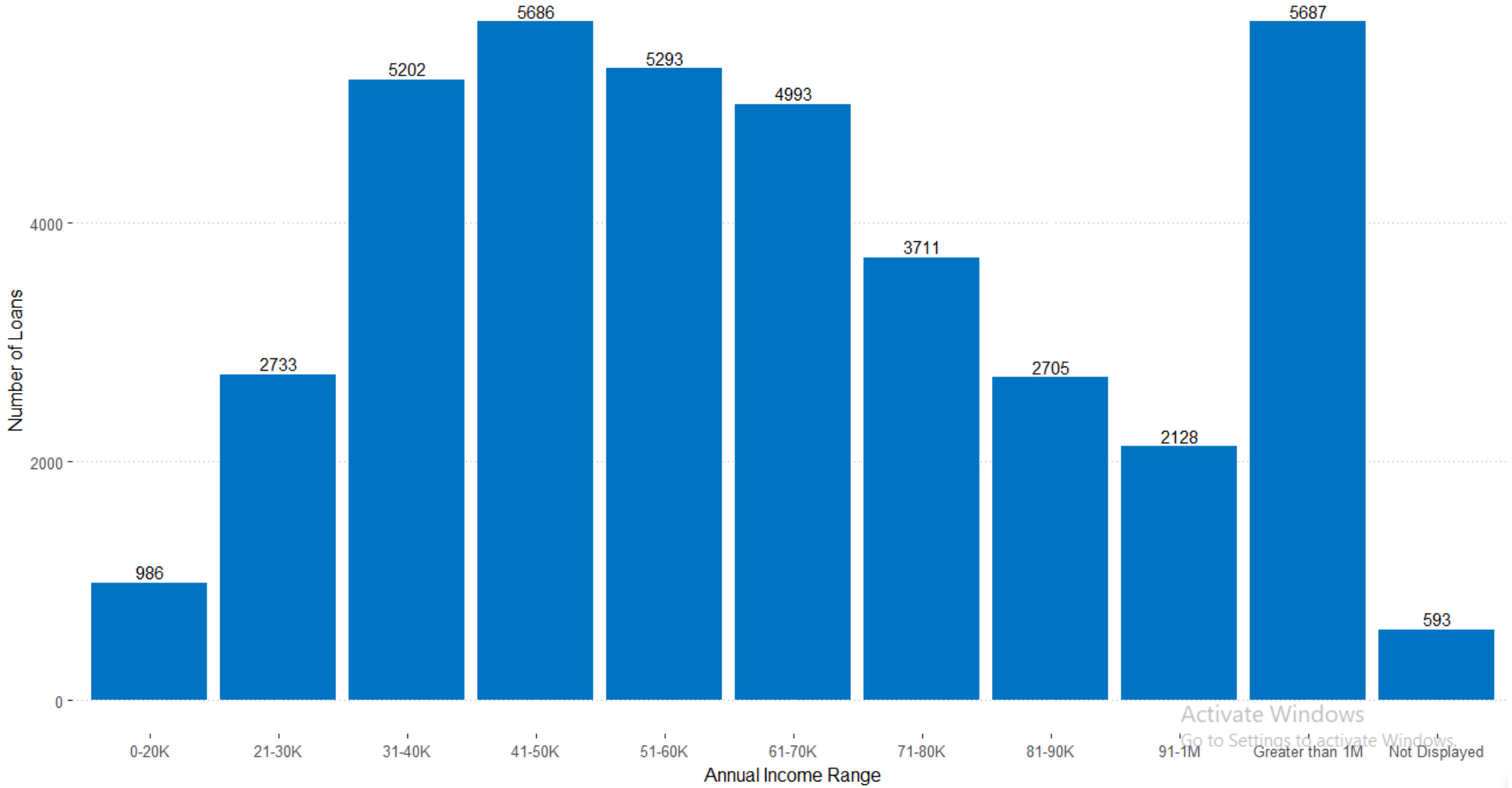
Plot 20



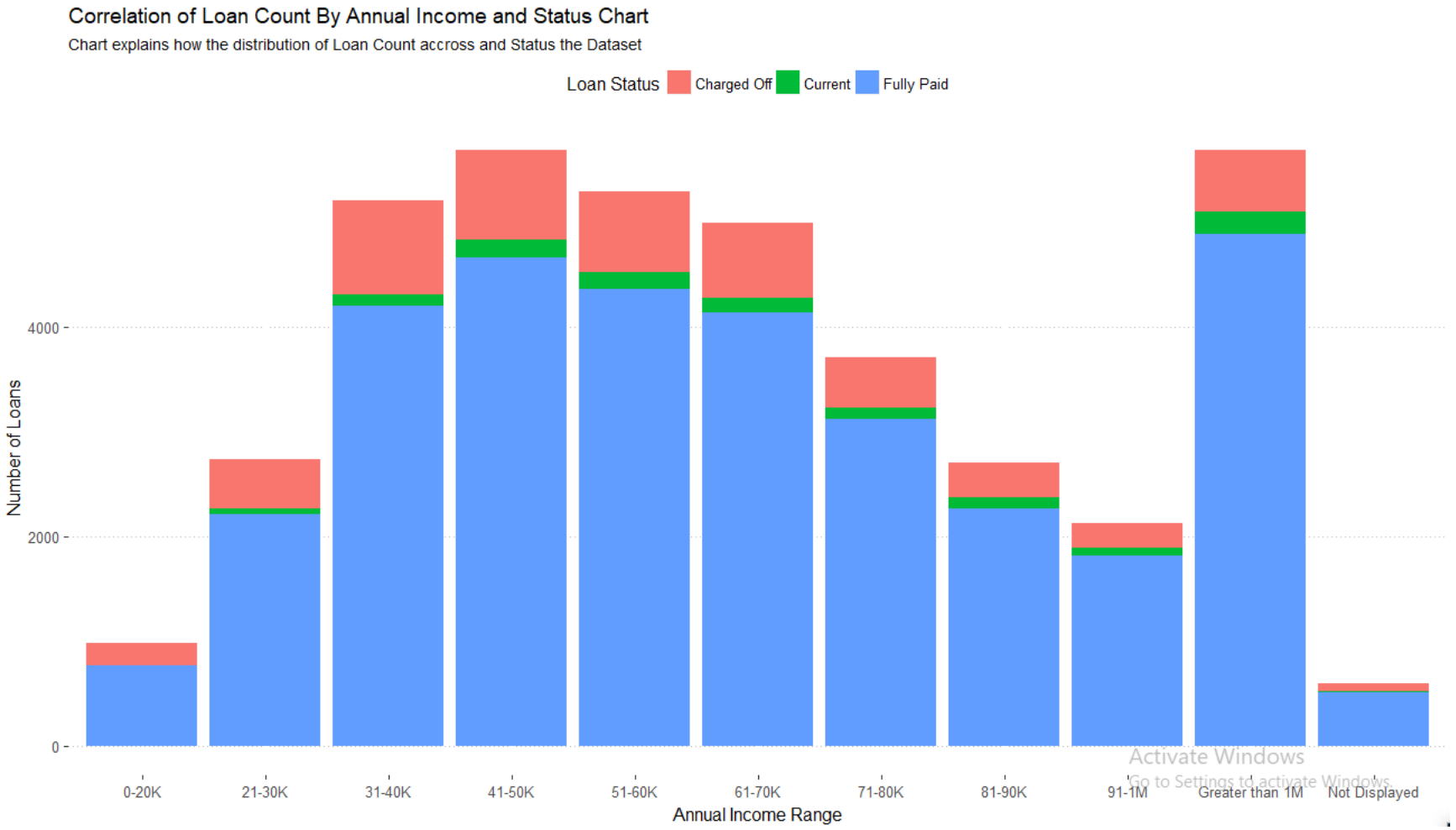
Plot 21

Loan Count By Annual Income

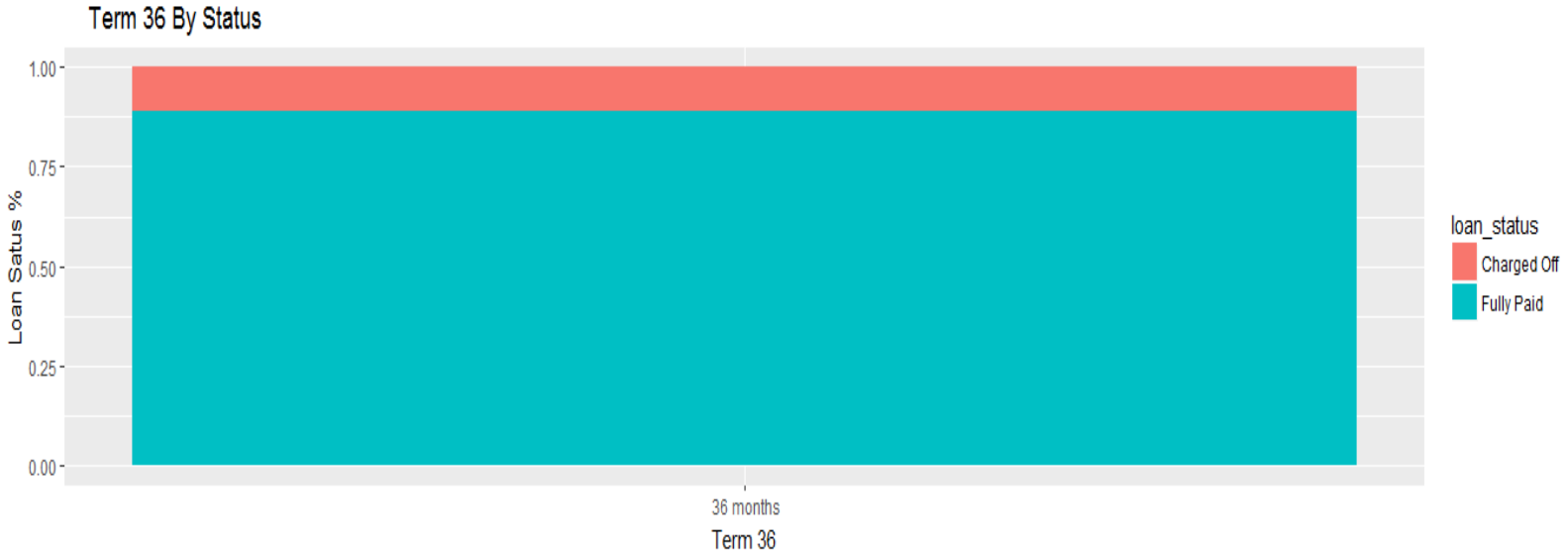
Chart explains how the distribution of Loan Count accross the Dataset

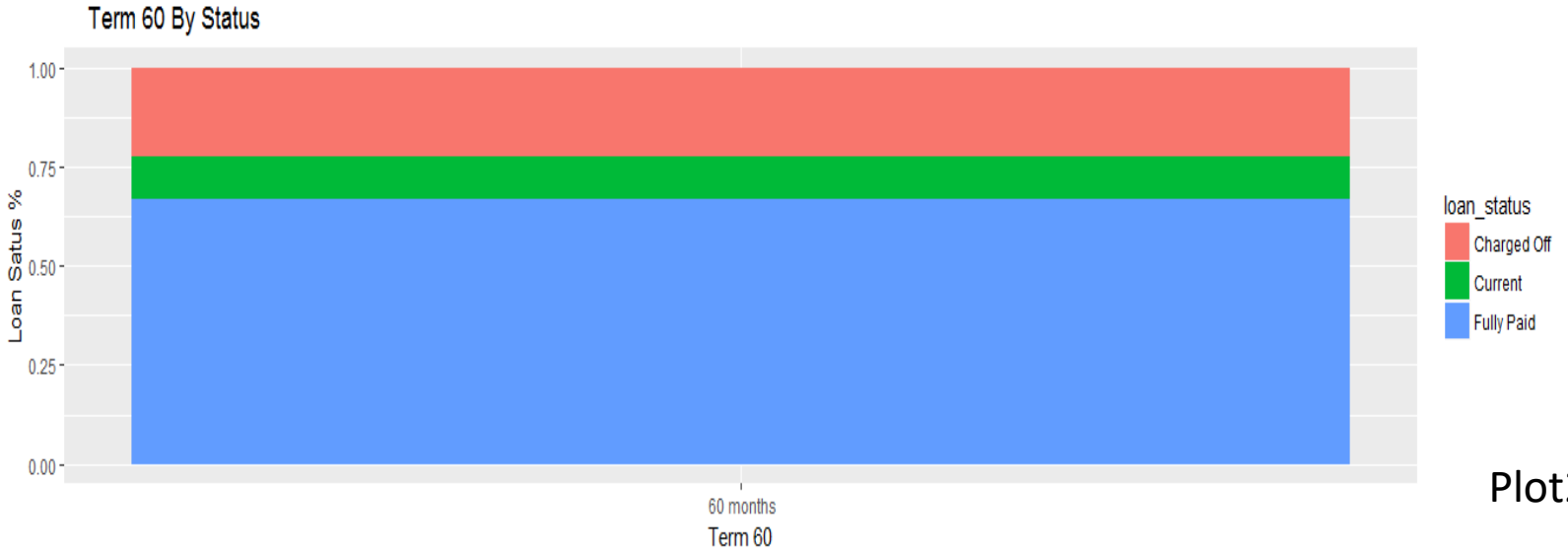


Plot 22


Correlation of Loan Count By Annual Income and Status Chart
Chart explains how the distribution of Loan Count accross and Status the Dataset
Loan Status
Charged Off
Current
Fully Paid
Number of Loans
0
2000
4000
0-20K
21-30K
31-40K
41-50K
51-60K
61-70K
71-80K
81-90K
91-1M
Greater than 1M
Not Displayed
Annual Income Range
Activate Windows
Go to Settings to activate Windows.

Plot 23





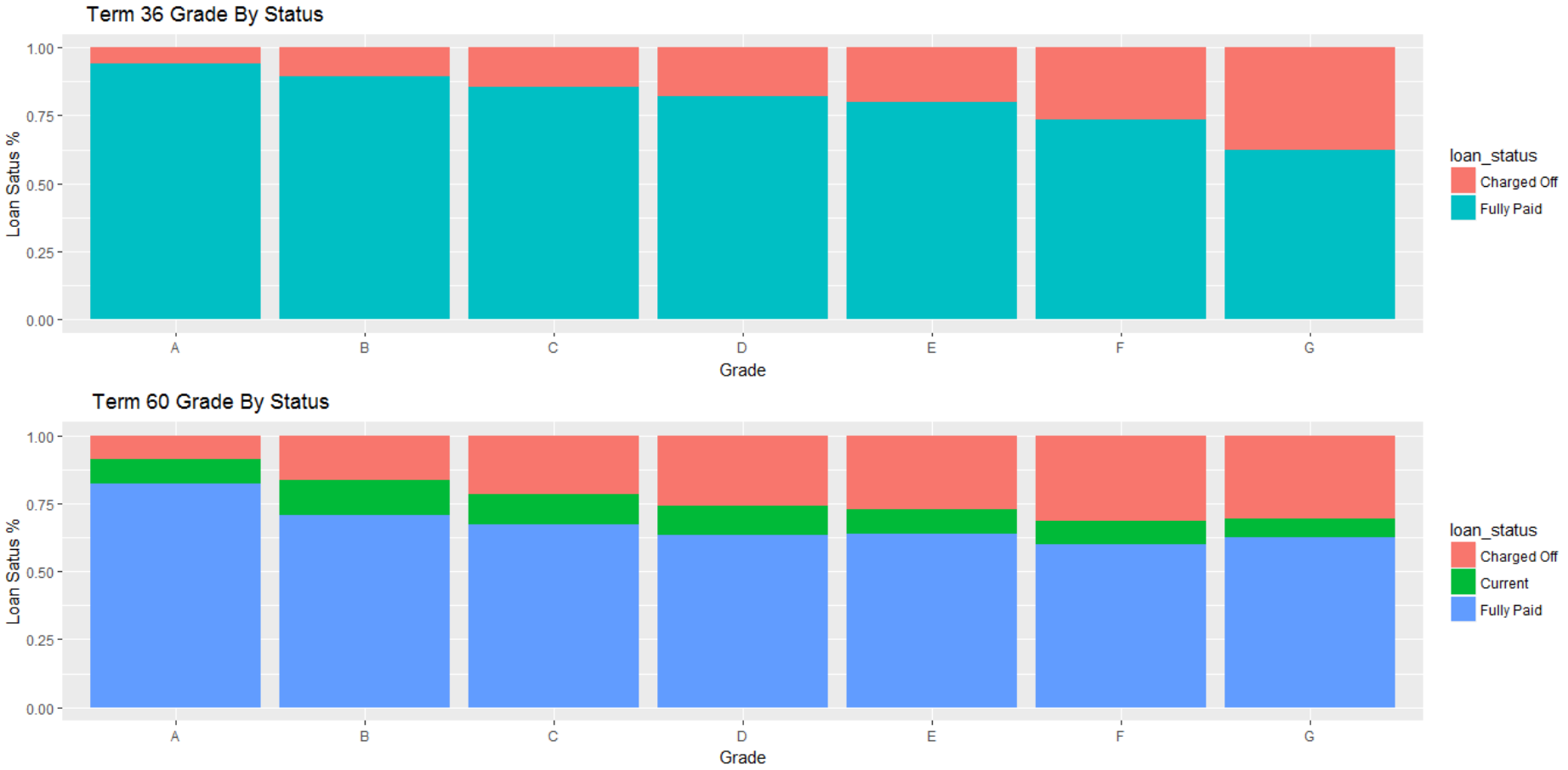
Graph Description:

The graph is plot against term and the percentage of loan status

Outcome:

- Term 60 has more probability of defaulters than Term 36
- Term 60 has approx 25 % Charged Off borrower

Plot24

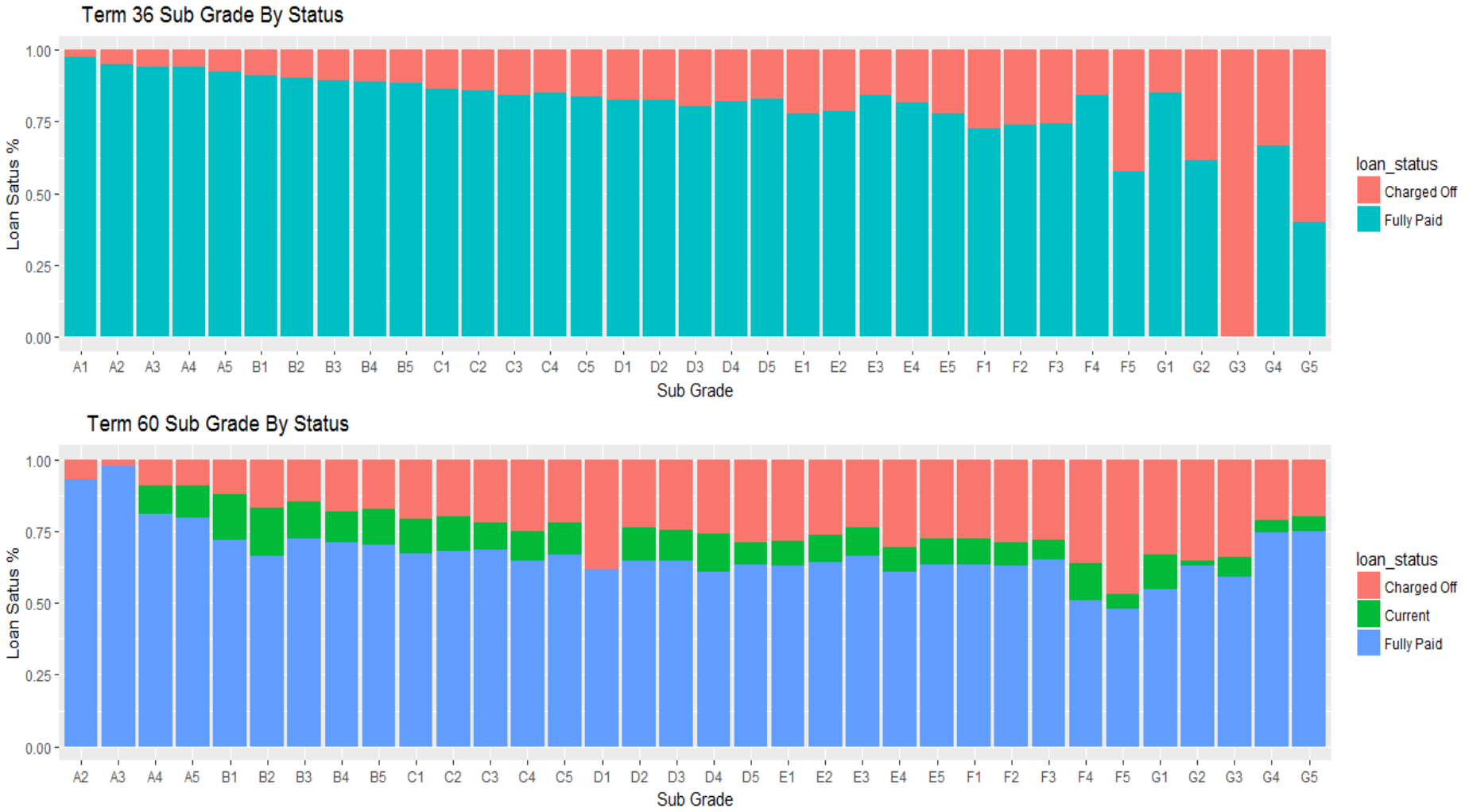


Graph Description:
The graph is plot against Grade and percentage of Loan status

Outcome:

- As the grade increases aphetically the percentage of defaulters also increases

Plot25



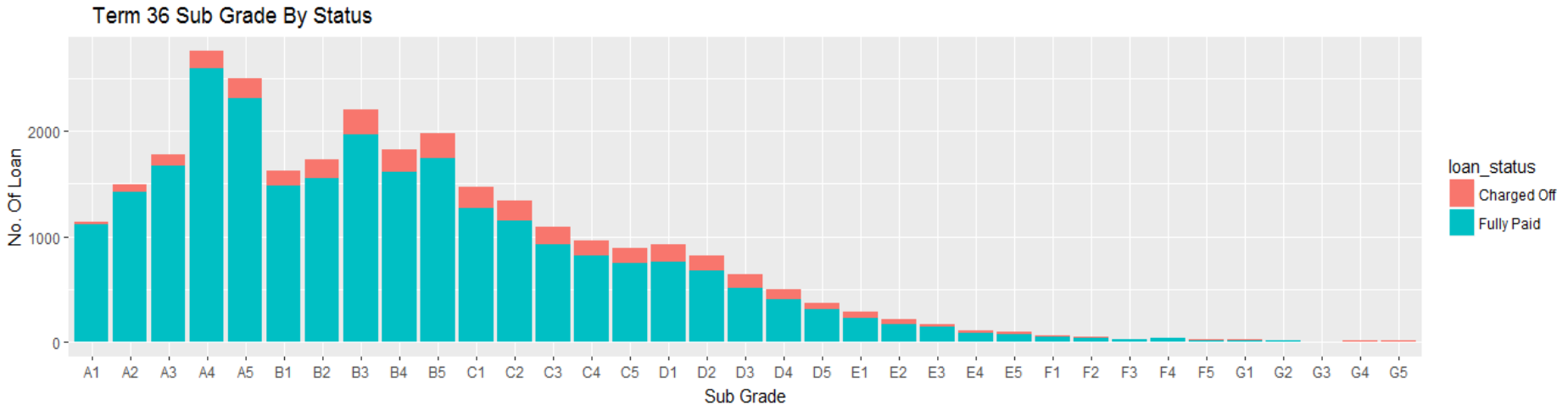
Plot26

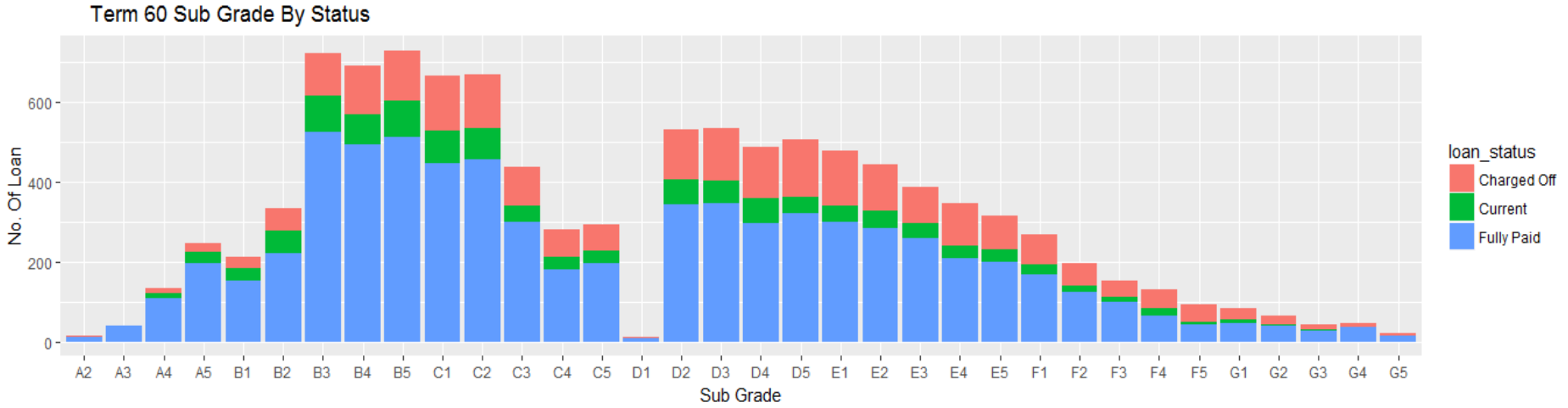
Graph Description:

The graph is plot against Sub Grade and percentage of Loan status

Outcome:

- Majorly the Grade increase alphabetically the Charged off percentage also increases
- Term 60 has more defaulters





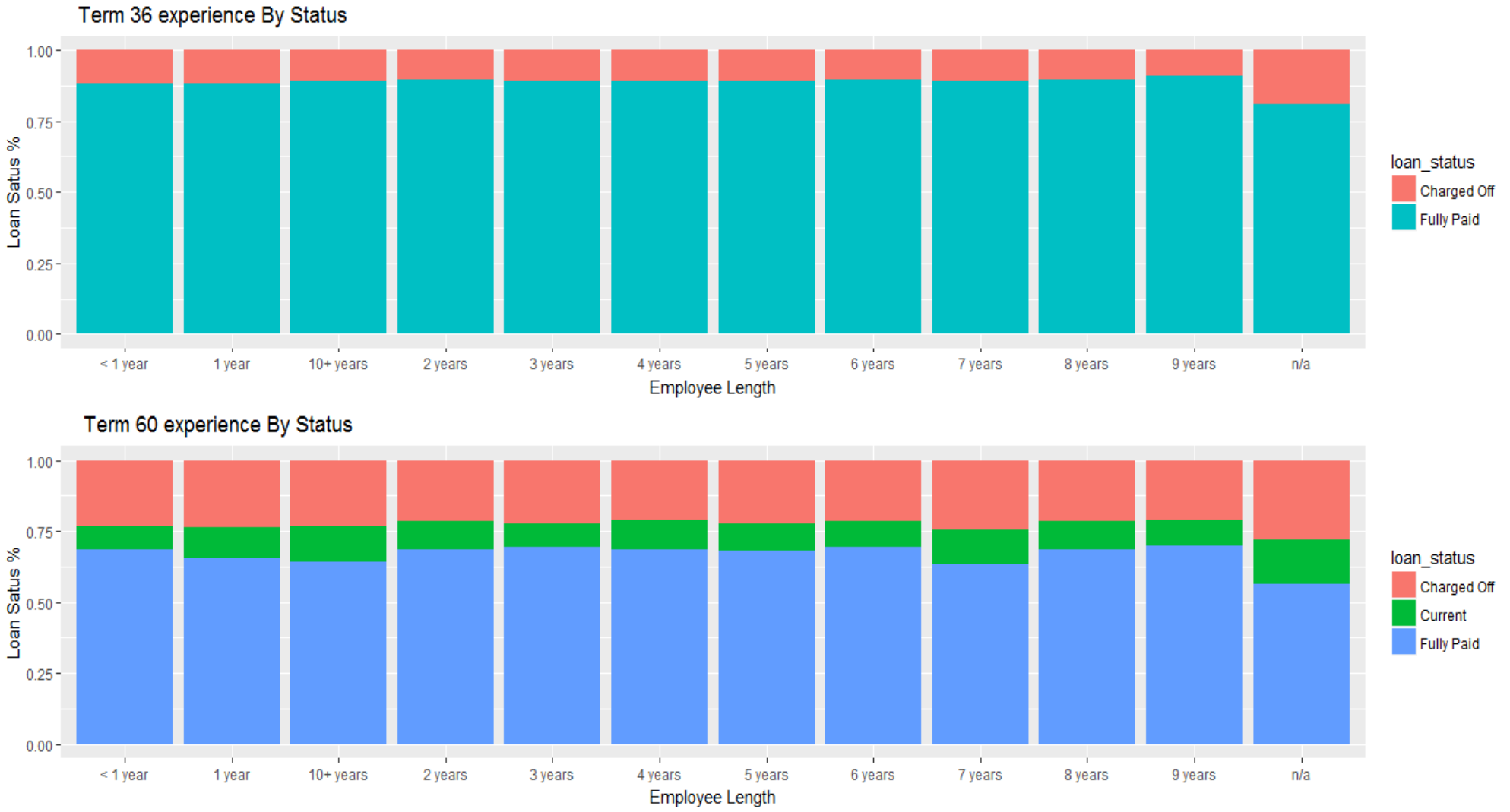
Plot27

Graph Description:

The graph is plot against Sub Grade and Number of Loan status

Outcome:

- Majorly the Grade increase alphabetically the Charged off percentage also increases
- Term 60 has more defaulters
- Term 36 has mainly borrowers from sub grade A3 to B5
- Term 60 Has mainly borrowers from sub grade B3 to D5



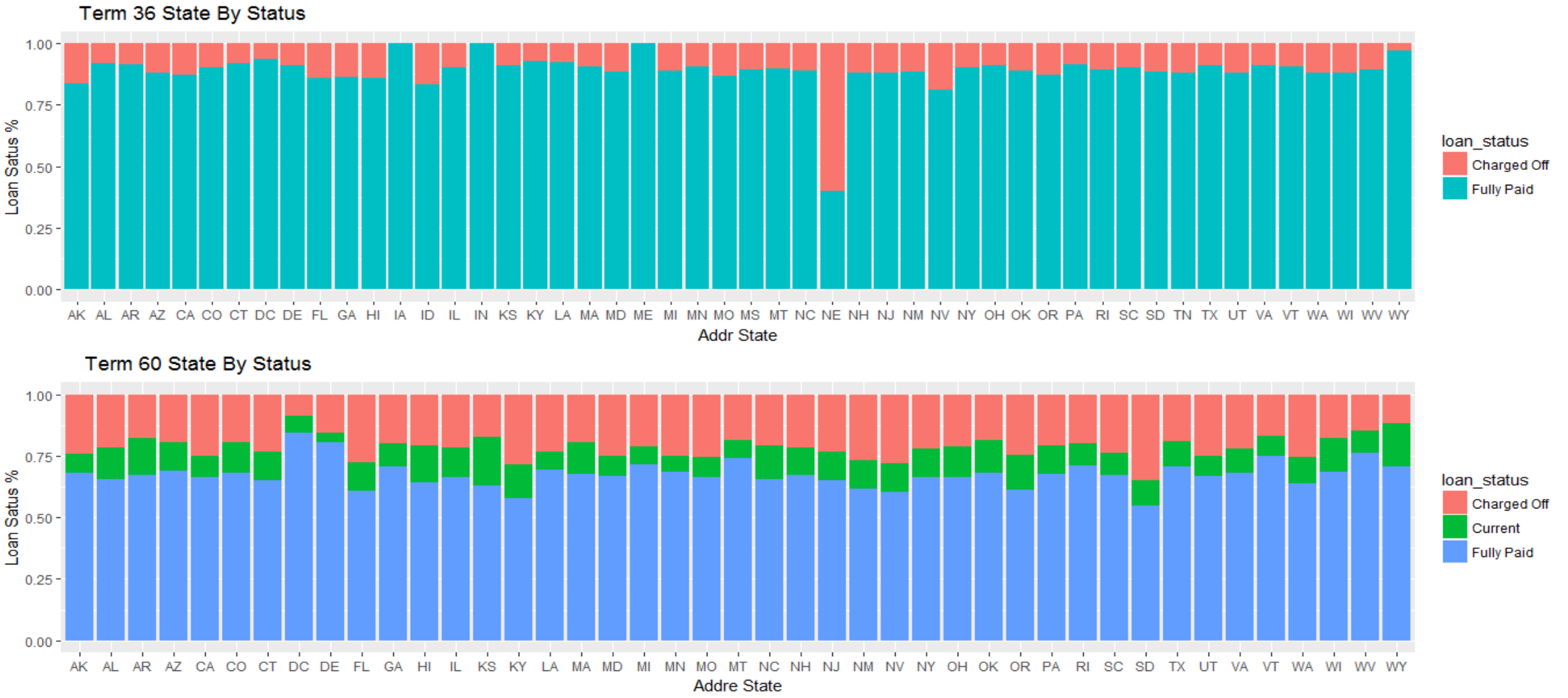
Graph Description:

The graph is plot against Experience Length and percentage of Loan status

Outcome:

- For Both Term 36 and 60 the borrower having ***no experience*** has more probability of charged Off

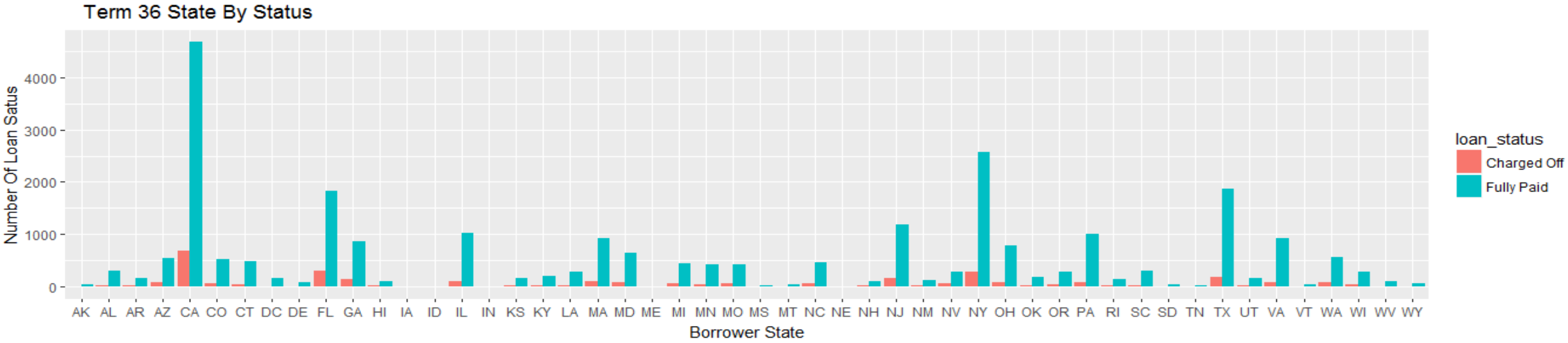
Plot28

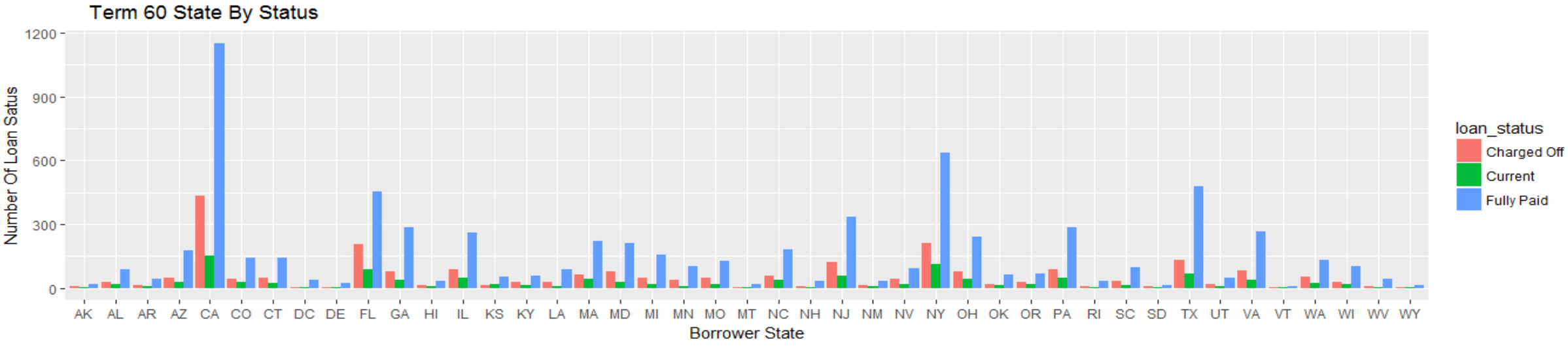


Graph Description:

The graph is plot against Borrower State and percentage of Loan status

Plot29

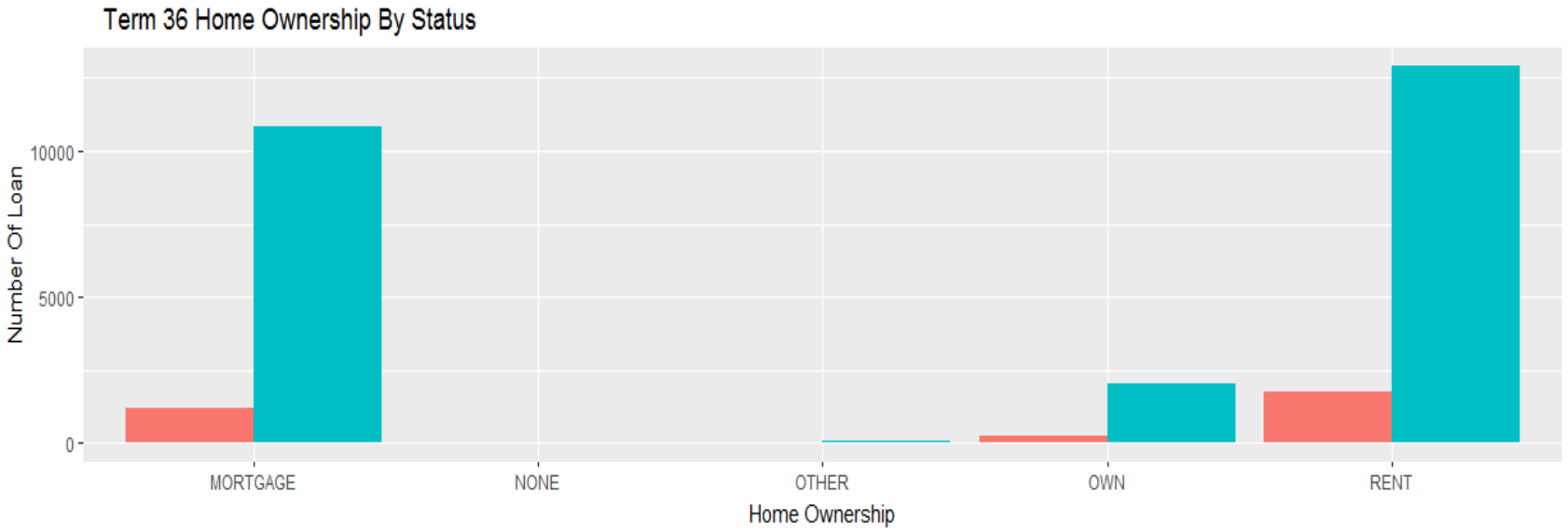


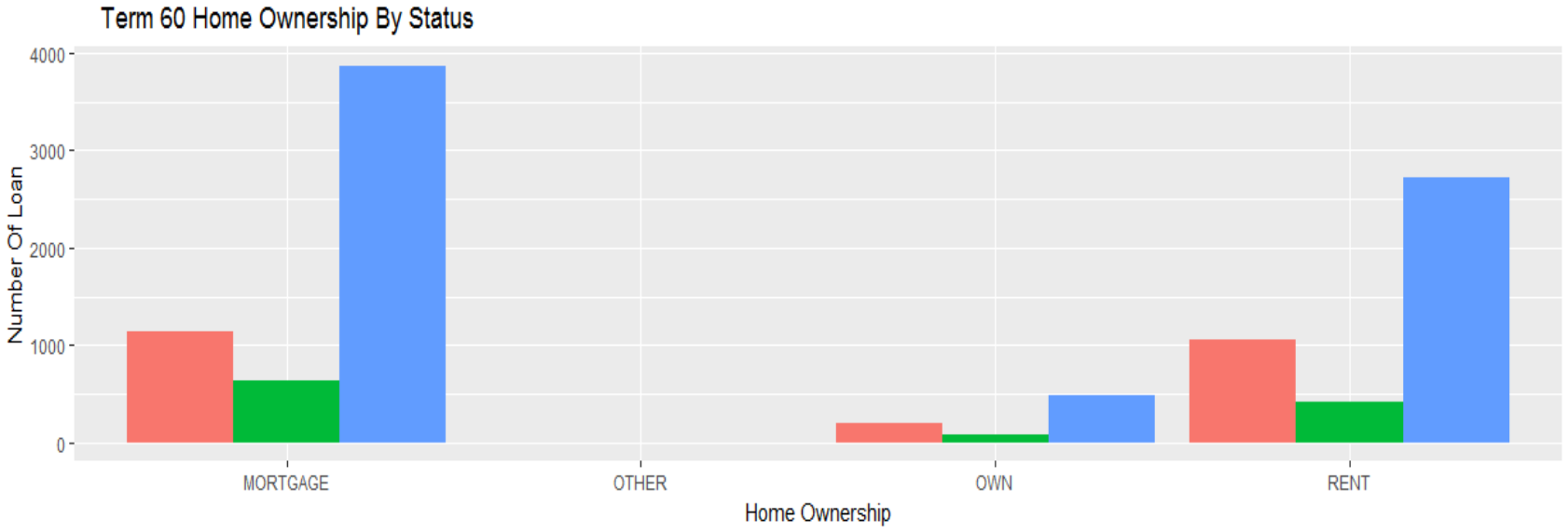


<u>Graph Description</u>:

The graph is plot against Borrower State and percentage of Loan status

Plot: 30





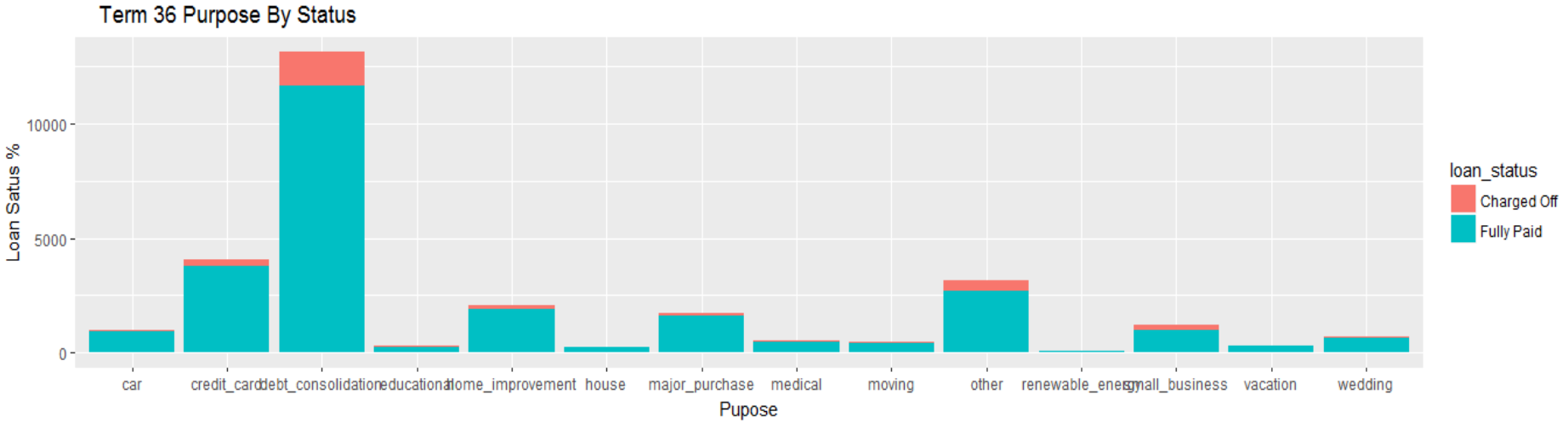
Graph description :

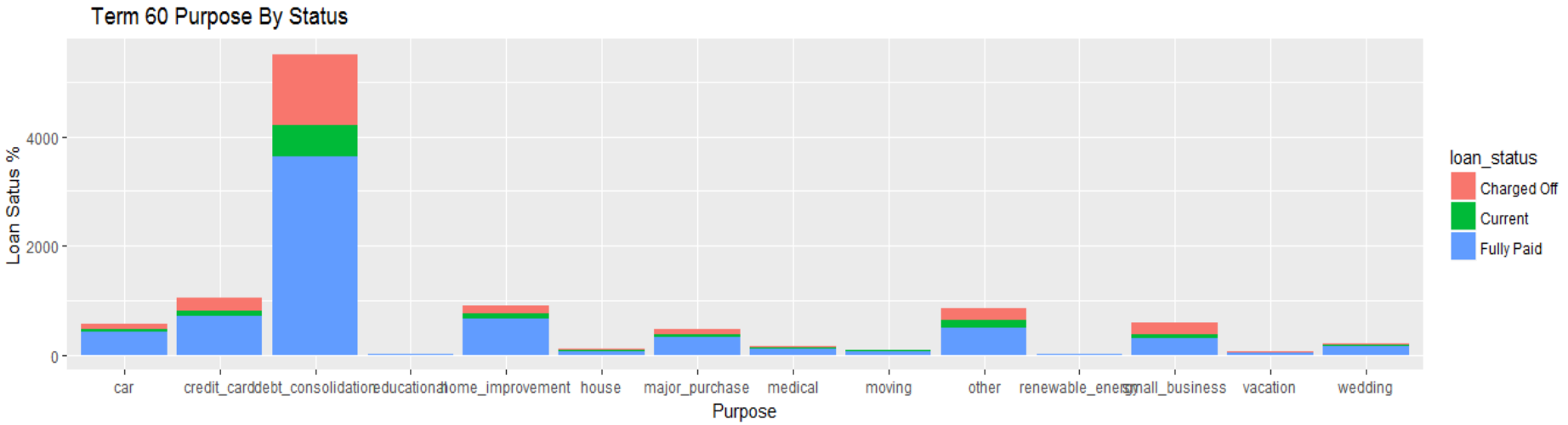
The graph is plotted against Home ownership and Number of Loan

Outcome:

- In both Term Home Ownership as Mortgage and Rent have more probability of being defaulters

Plot31





Graph Description:

Graph is plotted against purpose of loan abd percentage of loan status

Outcome:

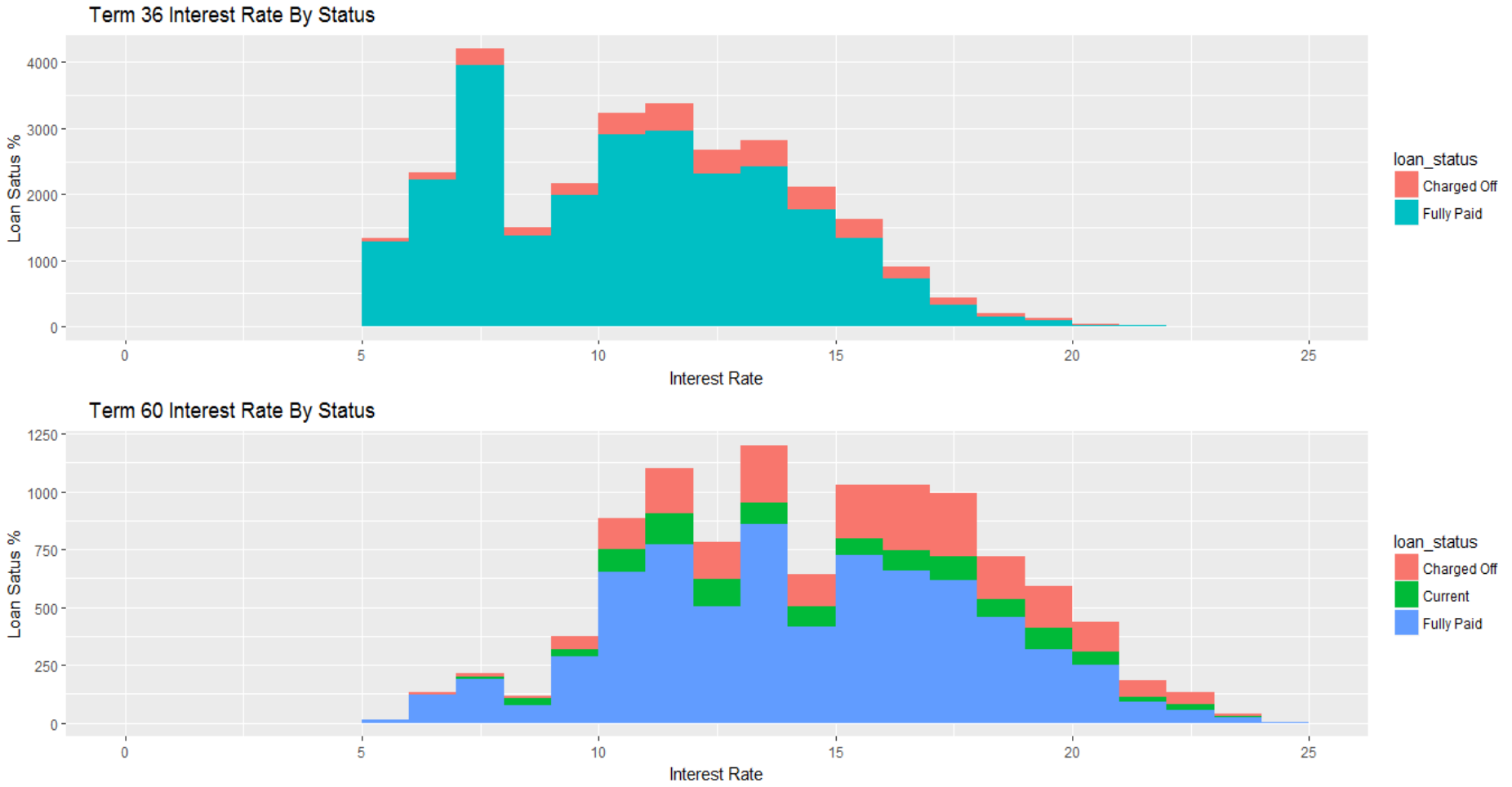
- For both terms Small business has got more probability of defaulters

Plot32

Graph Description:

The graph is plotted against Verification status and percentage of Loan Status

Plot33



Output:

- For both Term it is depicted from graph that as the interest rate increases the probability of being defaulter also increases

Plot34

Insights

1. Grade B has got the maximum number of borrowers. Refer Plot 1
2. The lending of loan is increasing across the years. Refer Plot 3
3. Term 60 has started from 2010 year. Refer Plot 4
4. Debt consolidation is the most requested loan purpose type. Refer Plot 5
5. US army and Bank occupation have more request of loan. Refer Plot 8
6. Customer from CA state got more loan across the year. Refer Plot 9
7. California state got more loan applicants. Refer Plot 10
8. Term 60 has got more defaulters than Term 36. Refer Plot 24
9. Small business has got more Interest Rate. Refer Plot 6

Recommendations

1. Term 60 applicant are more likely to be defaulter. Refer Plot 24
2. As the grade increases aphetically the percentage of defaulters also increases. Refer Plot 25,26,27
3. Applicant not having any employment experience are more likely to be a defaulter. Refer Pot 28
4. Home Ownership as Mortgage and Rent are more likely to be defaulter. Refer plot 31
5. Annual Income range 30k-40k are more likely to be defaulter, Refer Plot 23