EDA CASE STUDY

SUBMISSION

Group Name:
1. Savita Upadhyay
2. Vishnu Poddar
3. Vividh Garg
4. Mohit Mamgain

Objective

❑ Our Finance company is the largest online loan marketplace which facilitates personal loans, business loans and financing for medical procedures. Lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Borrowers who **default** cause the largest amount of loss to the lenders.

❑ Our aim is to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

❑ Identification of these risky loan applicants helps us to reduce such loans thereby cutting down the amount of credit loss.

Business Understanding

❏ When our company receives a loan application, the decision of loan approval based on the applicant's profile.

❏ When loan application is received, two types of decision are been made:

    1. Loan accepted: If company approves loans, there are 3 possible scenarios:

        (a) Fully paid: Applicant has fully paid the loan with interest

        (b) Current: Applicant is in process of paying the loan

        (c) Charged off: Applicant has not paid loan in due time i.e. he/she has defaulted.

    2. Loan rejected: The company has rejected the loan because applicant does not meet the required requirement.

Data Cleaning and Preparation

❑ Identification of dates which are not in proper format. Cleaned into one standard format.

❑ Identification of all columns that don't have any other value other than NA and 0. Removed such columns.

❑ Identify all columns that are not useful in analysis for e.g. the url and description column etc. Removed such columns also.

❑ Few columns such as rate of interest etc. have been read as characters due to the presence of the "%" symbol. Convert all such instances to numeric.

❑ Identification of outliers from annual income column. Removed the outliers.

❑ New variables are also introduced for useful analysis like years are extracted from date column and interest rate slot column also introduced.

Univariate Analysis

For analysis we have removed **"current"** loan status as we don't know either such applicants will default or not.

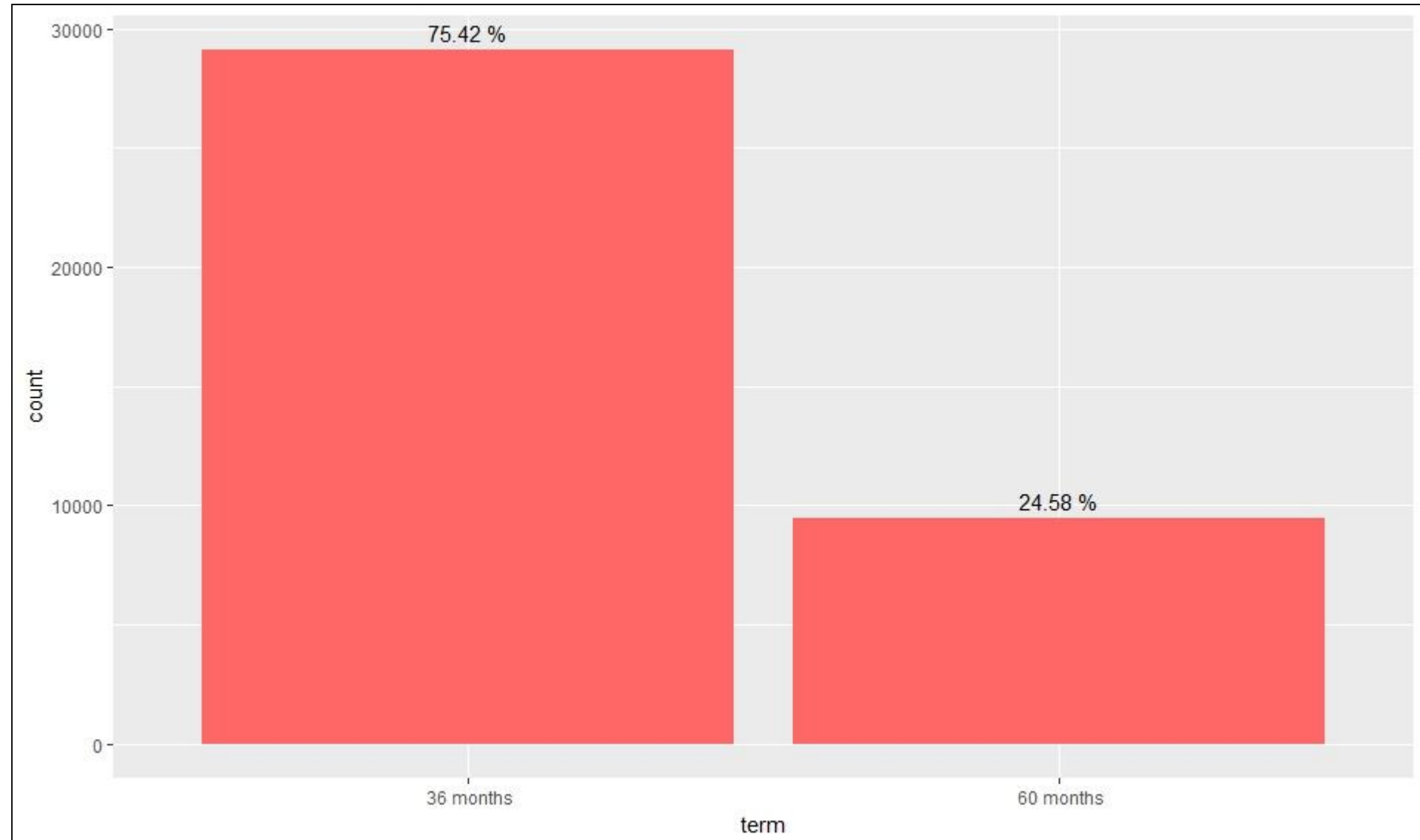Plot shows that **14.59%** of loan borrowers are defaulters.

frequency of people who default or not

Univariate Analysis
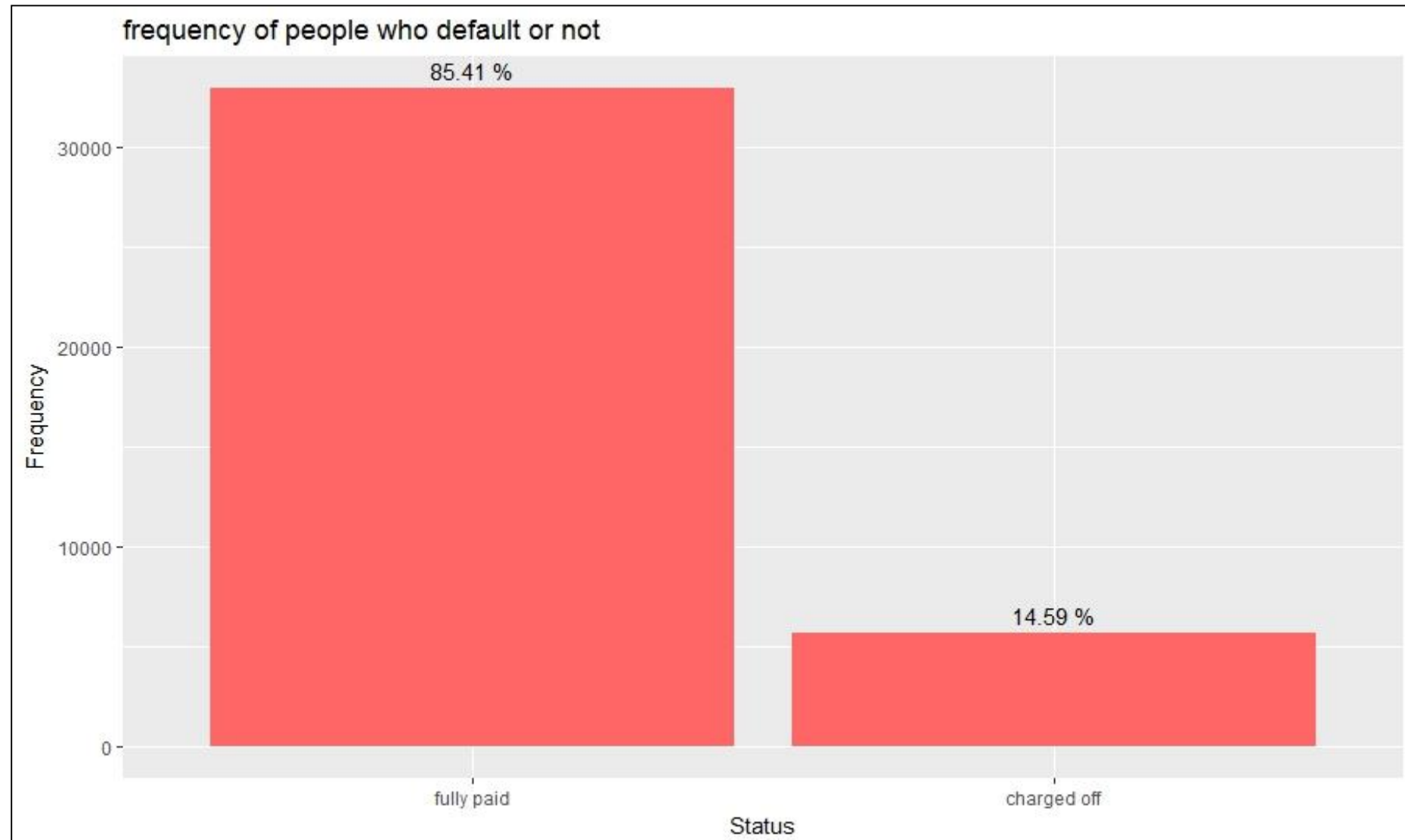
Plot shows loan applicants are increasing year by year with **53%** applicant in 2011.

Univariate Analysis
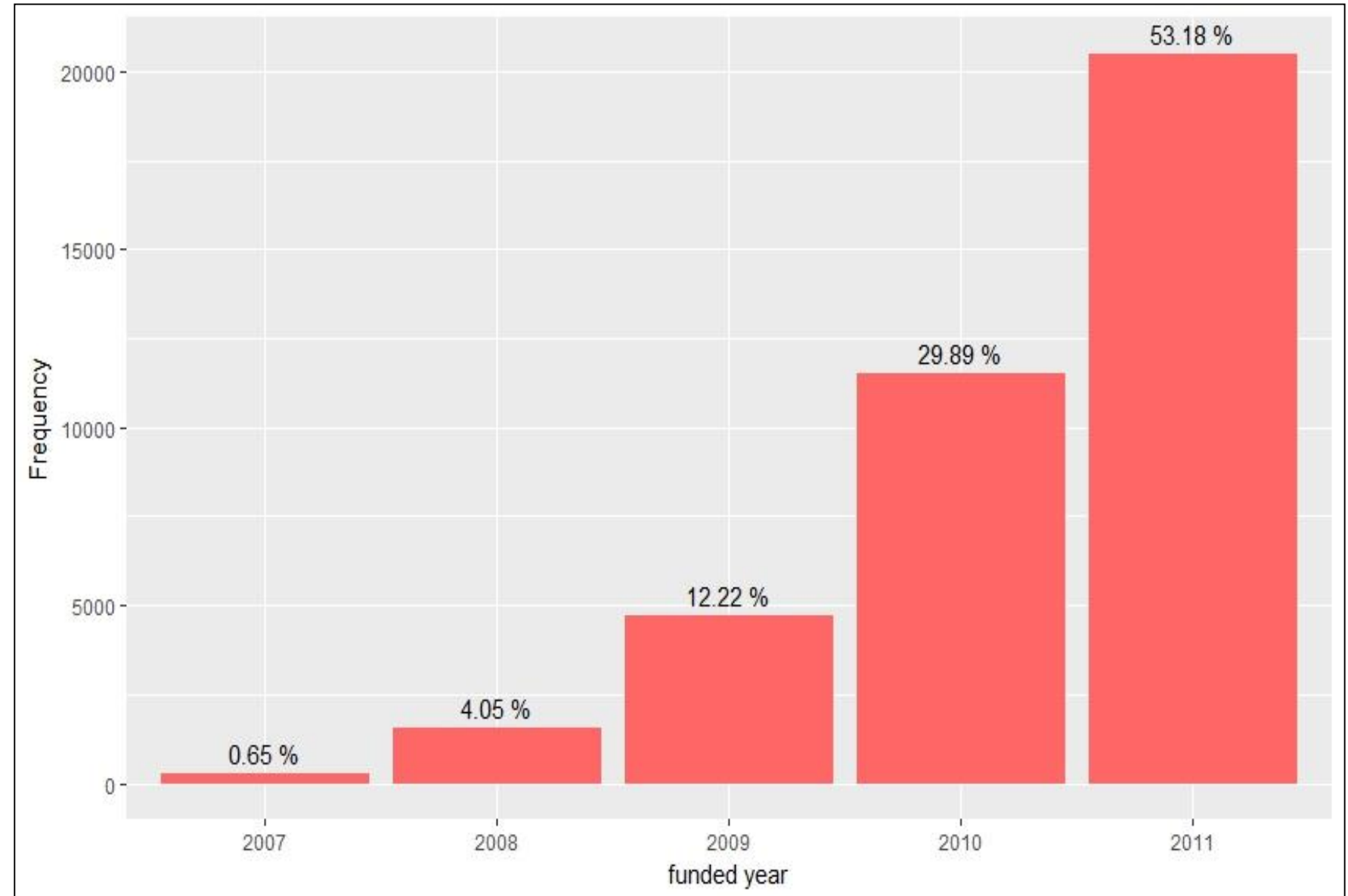
Plot shows **75%** of loan borrowers have applied loan for term of **36 months**.

Bivariate Analysis

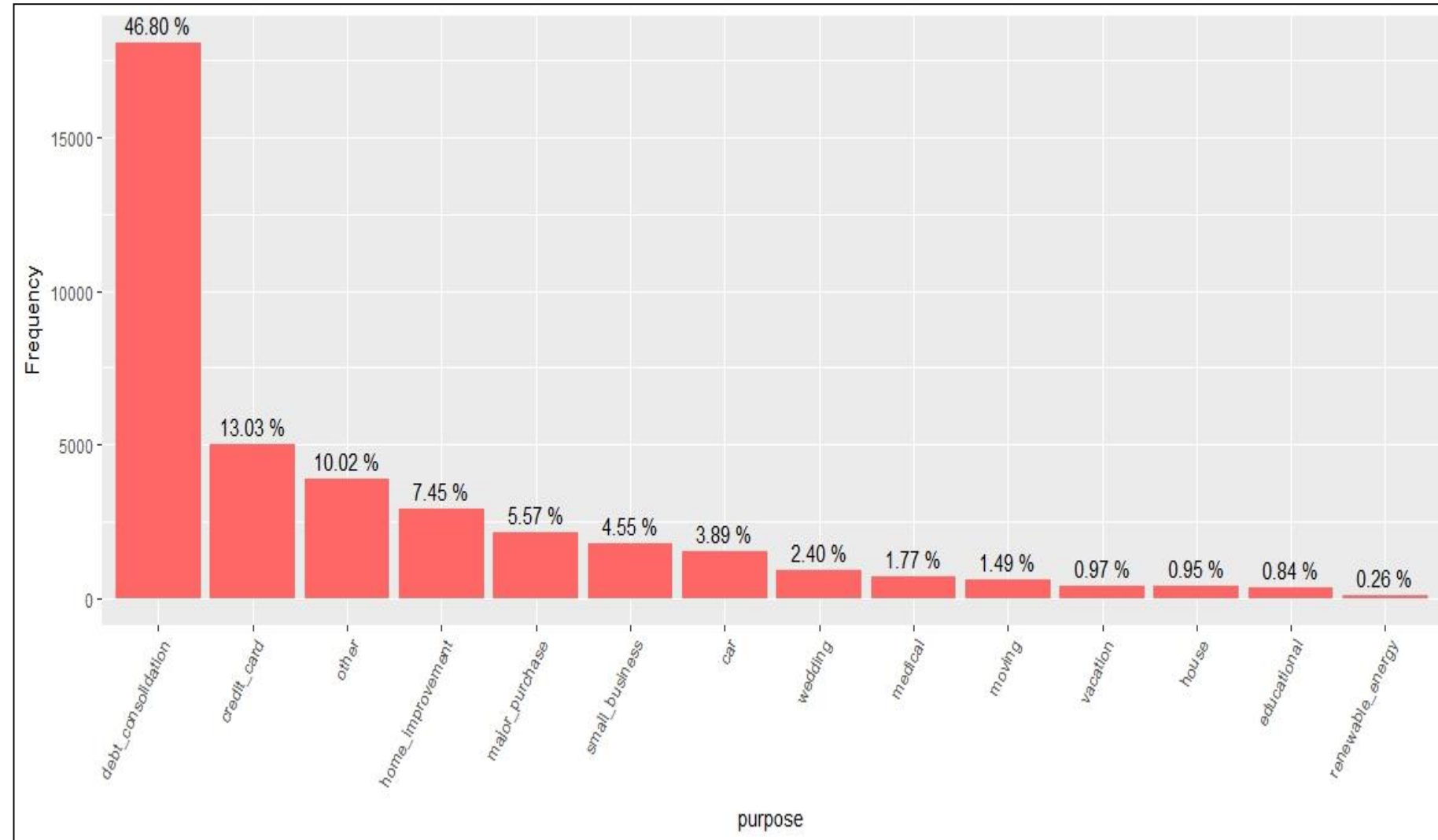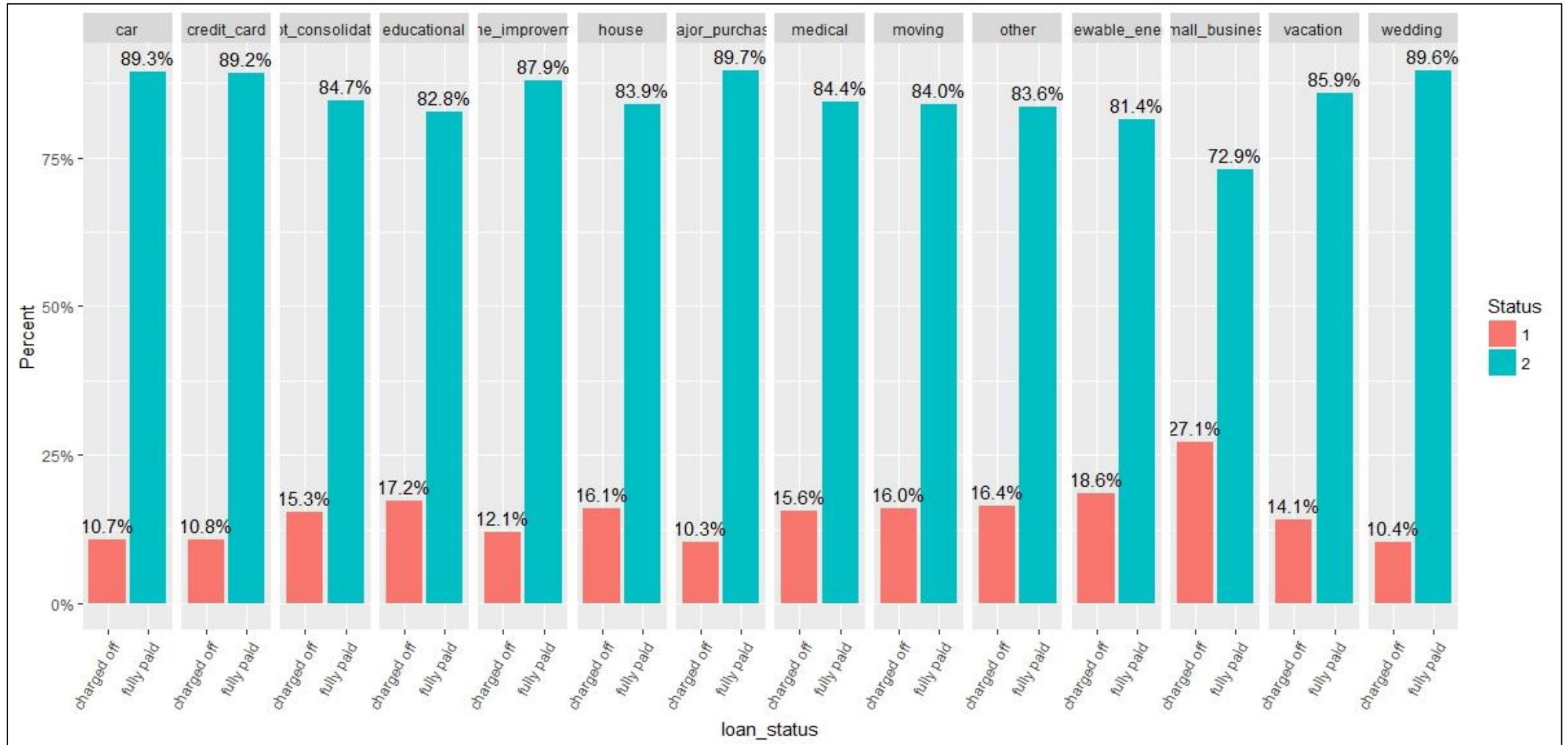Plot shows loan borrowers with **60 months** of term period are most likely to **default**.
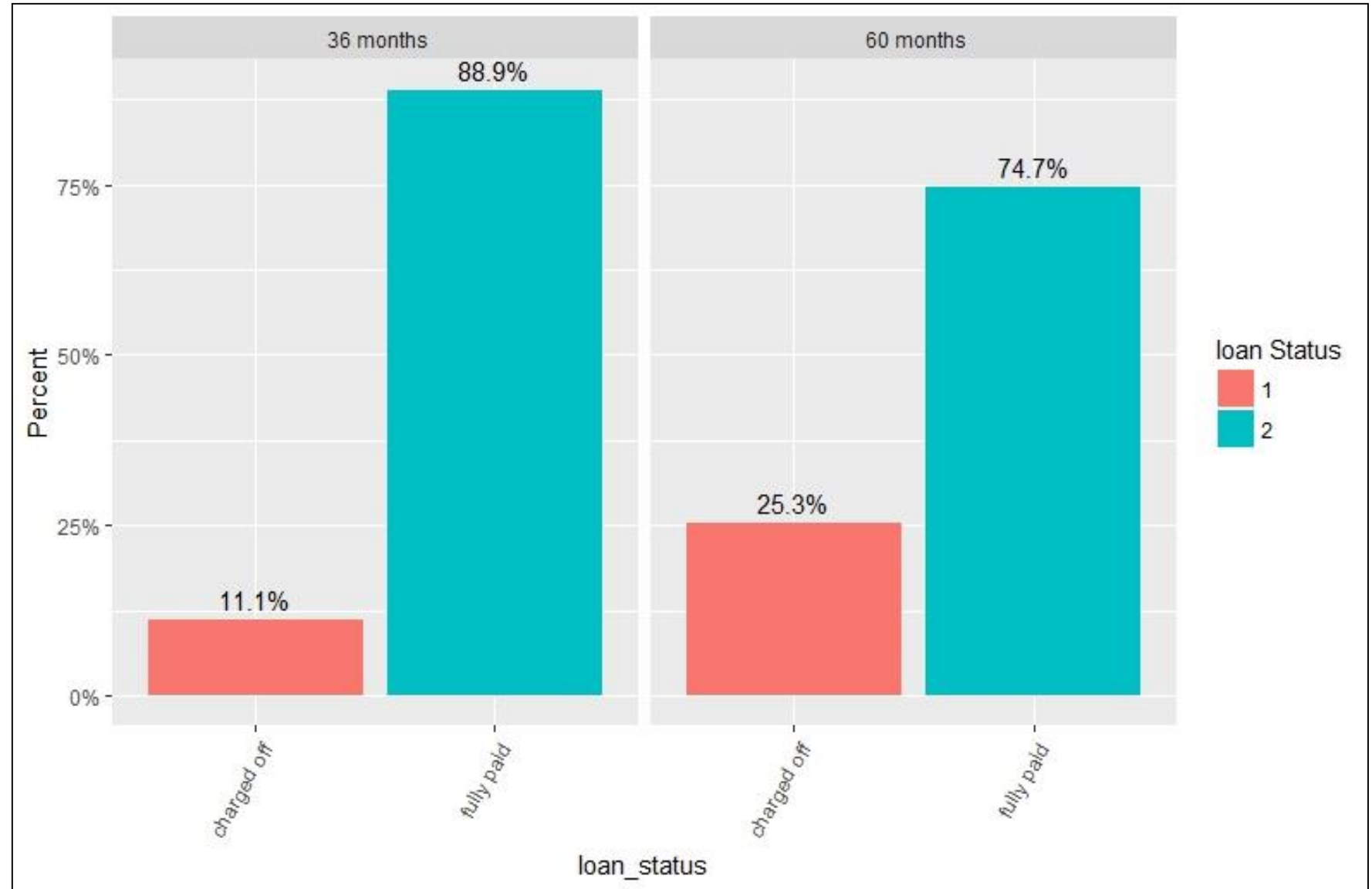
Plot shows **47%** of loan applicant applied loans for debt consolidation purpose.

Top 4 purpose of loan applicant:
(a) Debit consolidation
(b) Credit Card
(c) Home improvement
(d) Major purchase

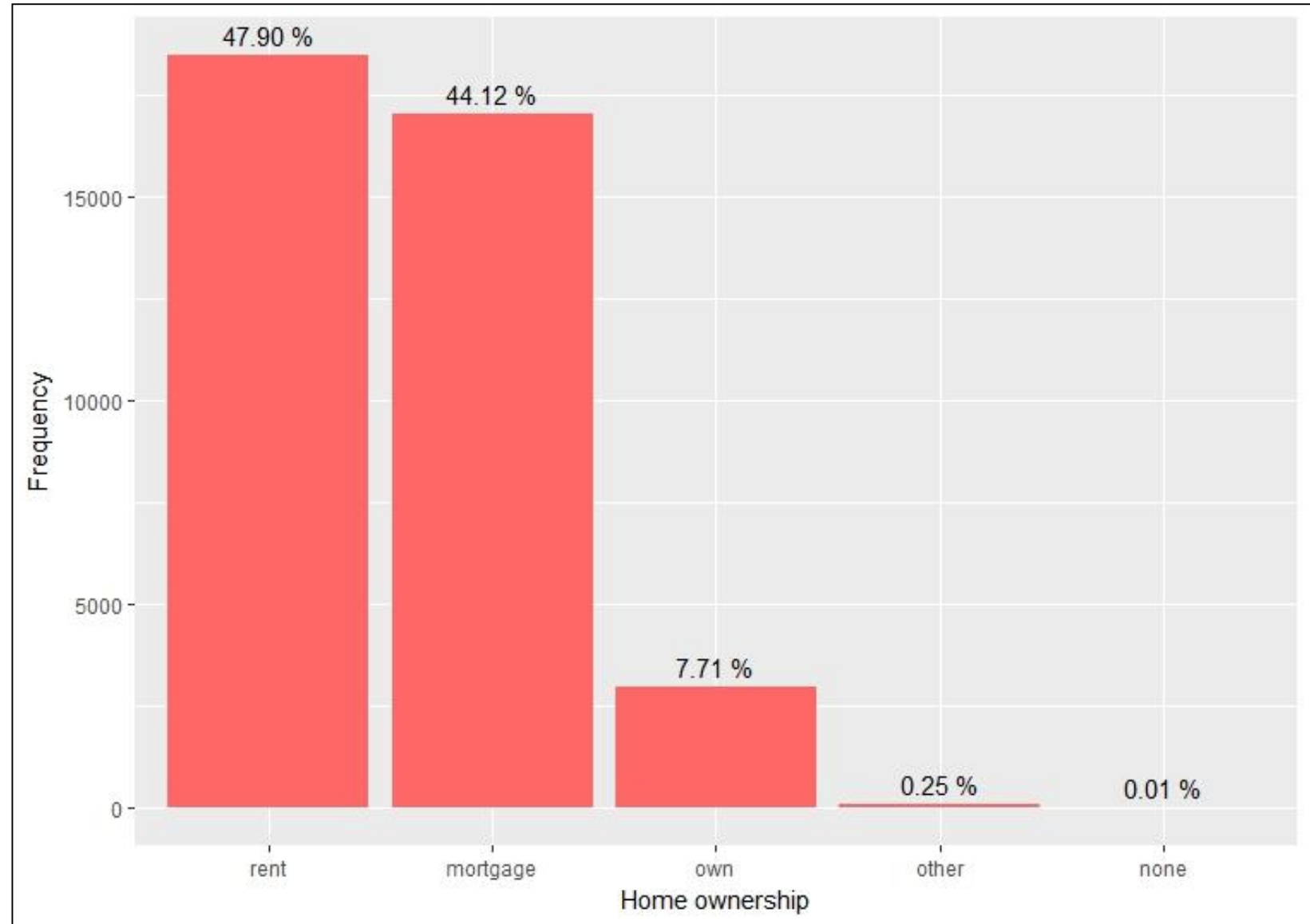We have ignore "other" as we don't know much about it.

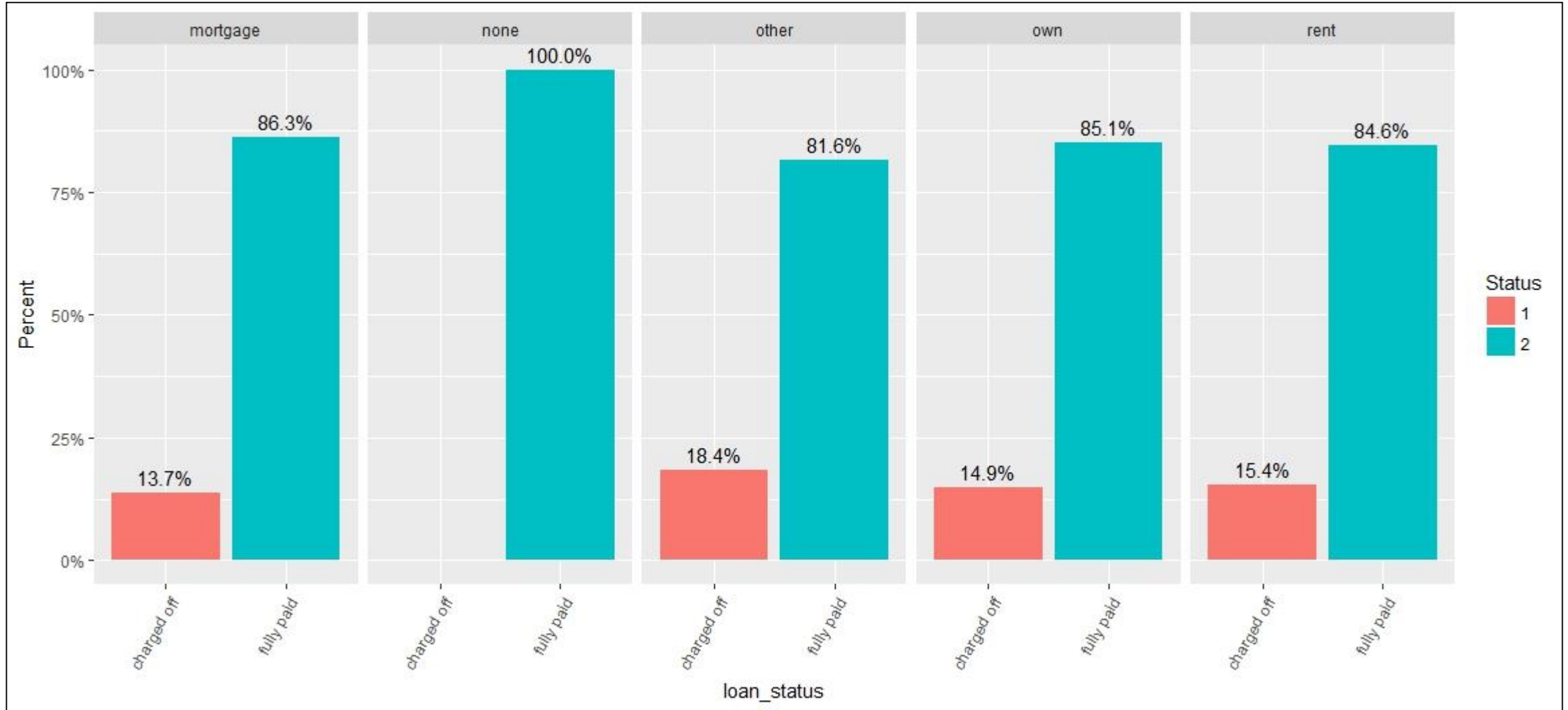# Bivariate Analysis

Loan borrowers with **small business** purpose are most likely to **default** in comparison to other purposes.

Univariate Analysis

Plot shows 48% of loan applicant applied loans have rented home and 44% are mortgaged.

Bivariate Analysis
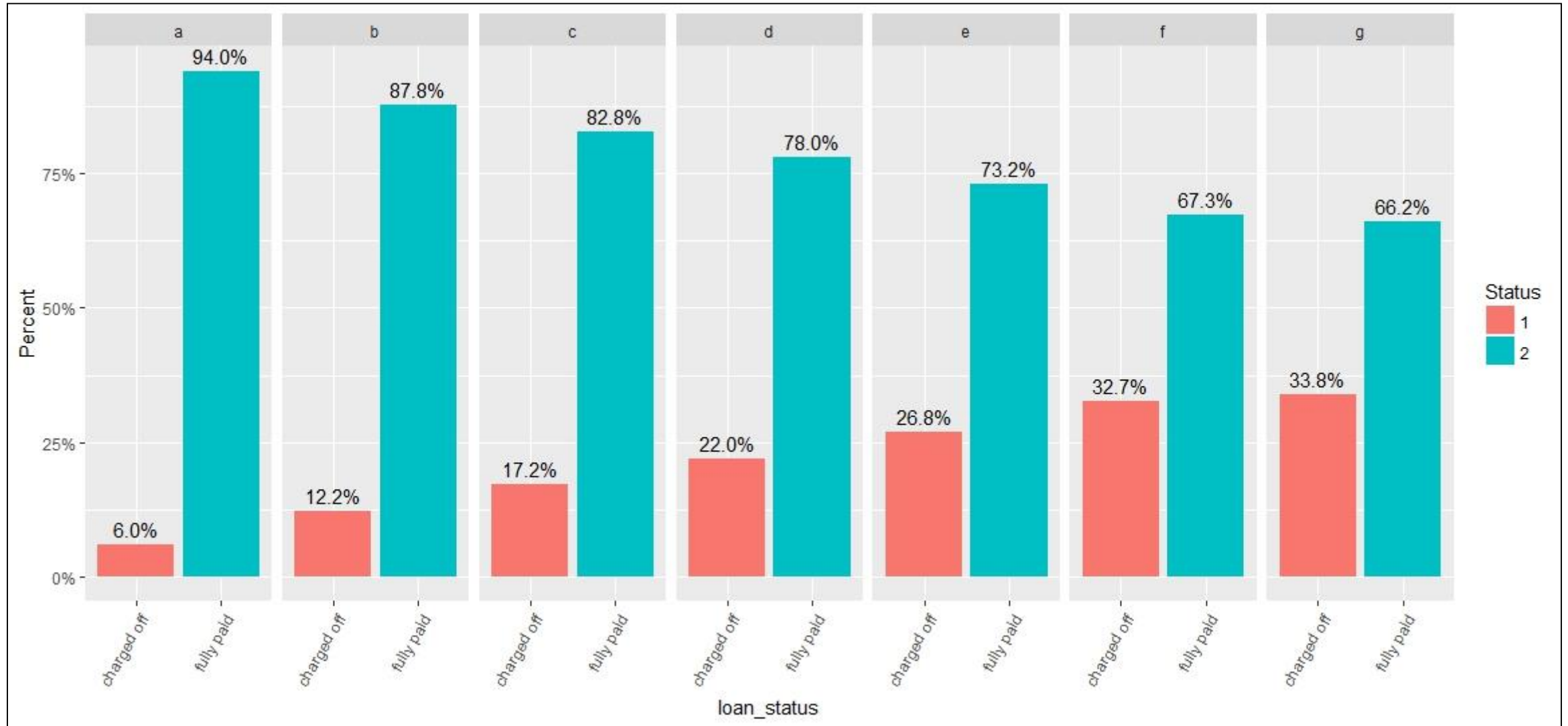

loan borrowers with **others as home ownership** are more likely to **default** in comparison to more home ownership

Univariate Analysis
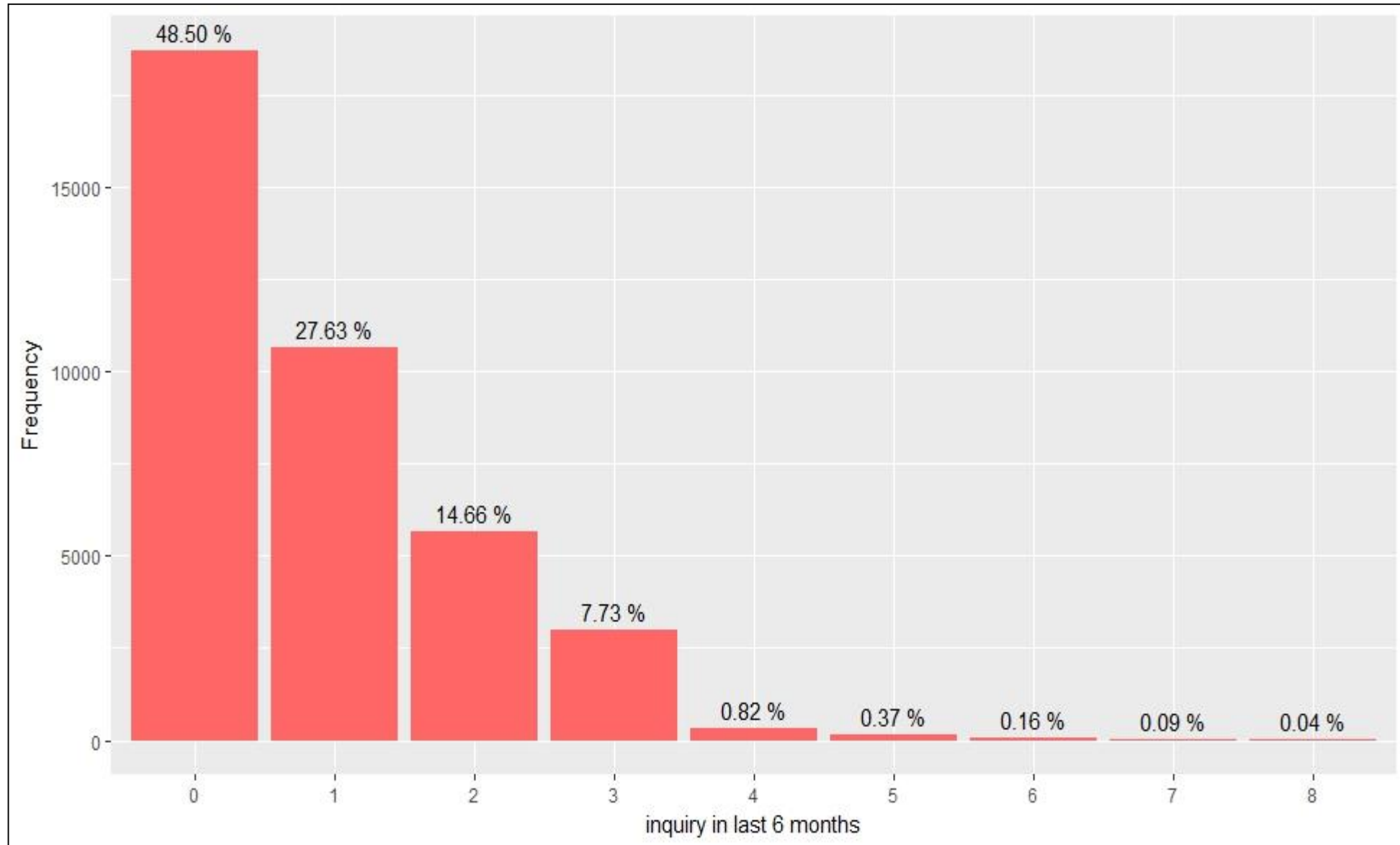
Plot shows **30%** of loan applicant applied loans have comes under **grade b**.
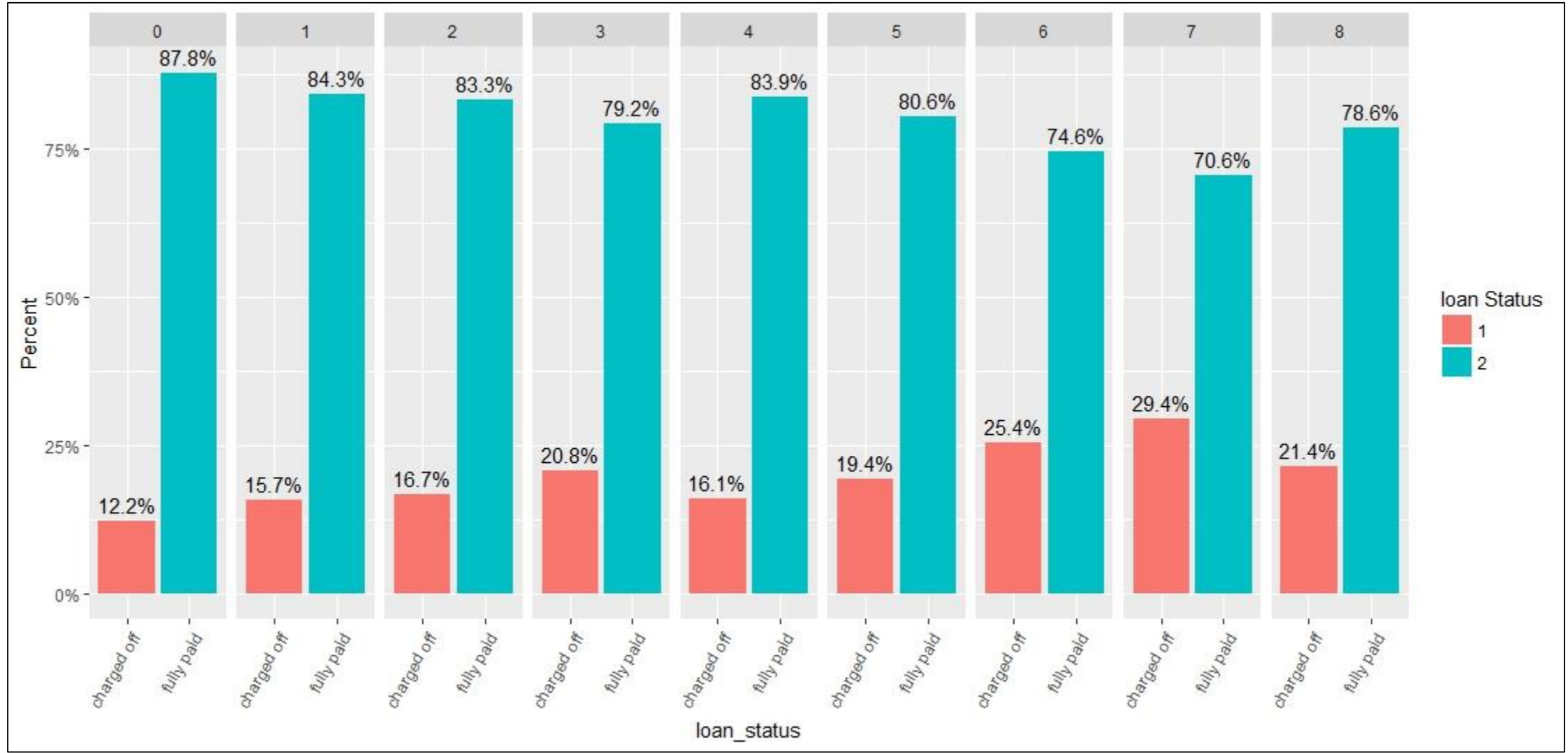
Bivariate Analysis
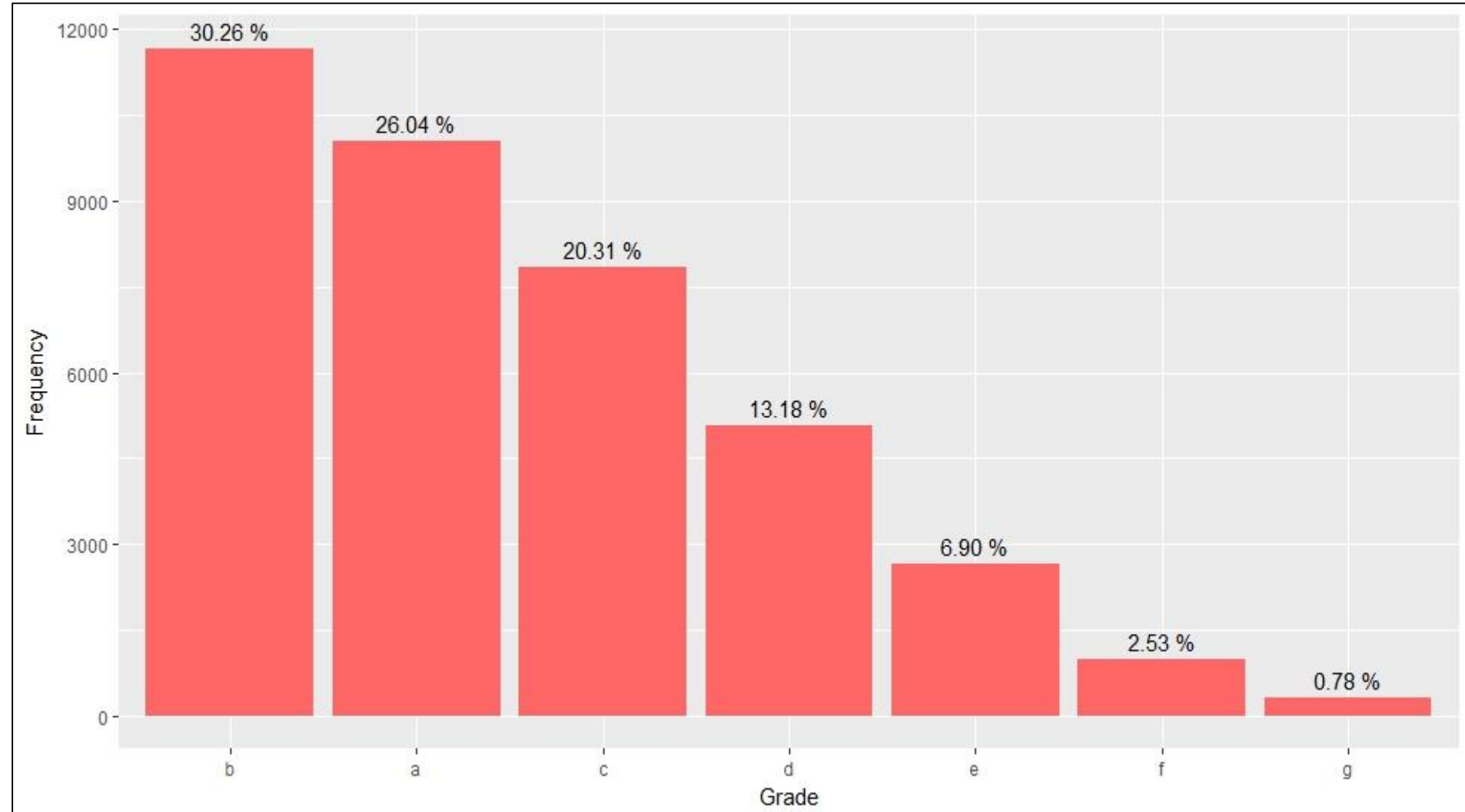
loan borrowers with **grade G** are more likely to **default** in comparison to other grades

Univariate Analysis

Plot shows **48.50%** of loan applicant applied loans have comes **nill** inquiry in past 6 months.

Bivariate Analysis

loan borrowers with **7 inquiry in last 6 months are more likely to default**.
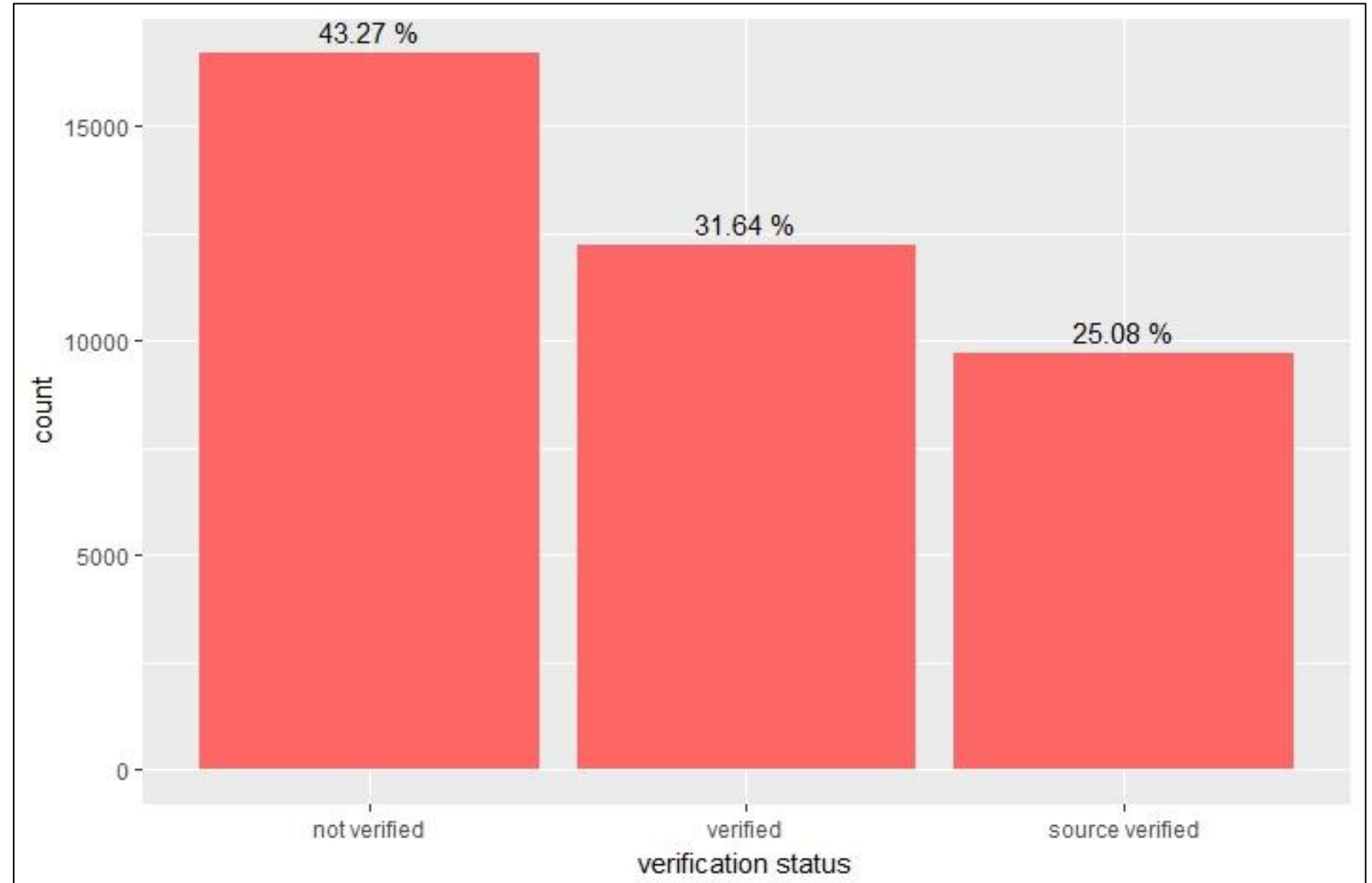
Univariate Analysis

Plot shows nearly **95%** of loan applicant have **nill** derogatory public records.

Bivariate Analysis
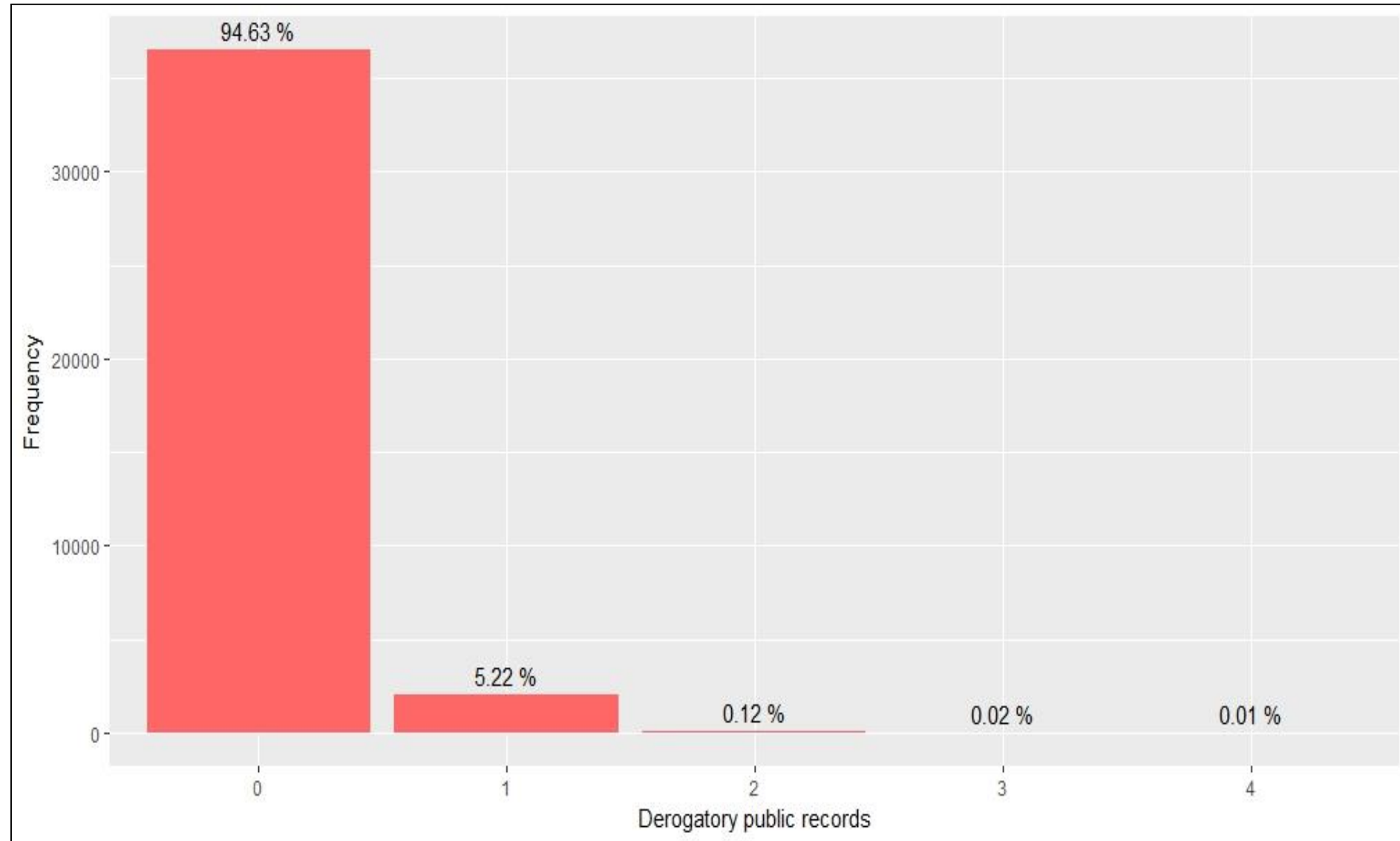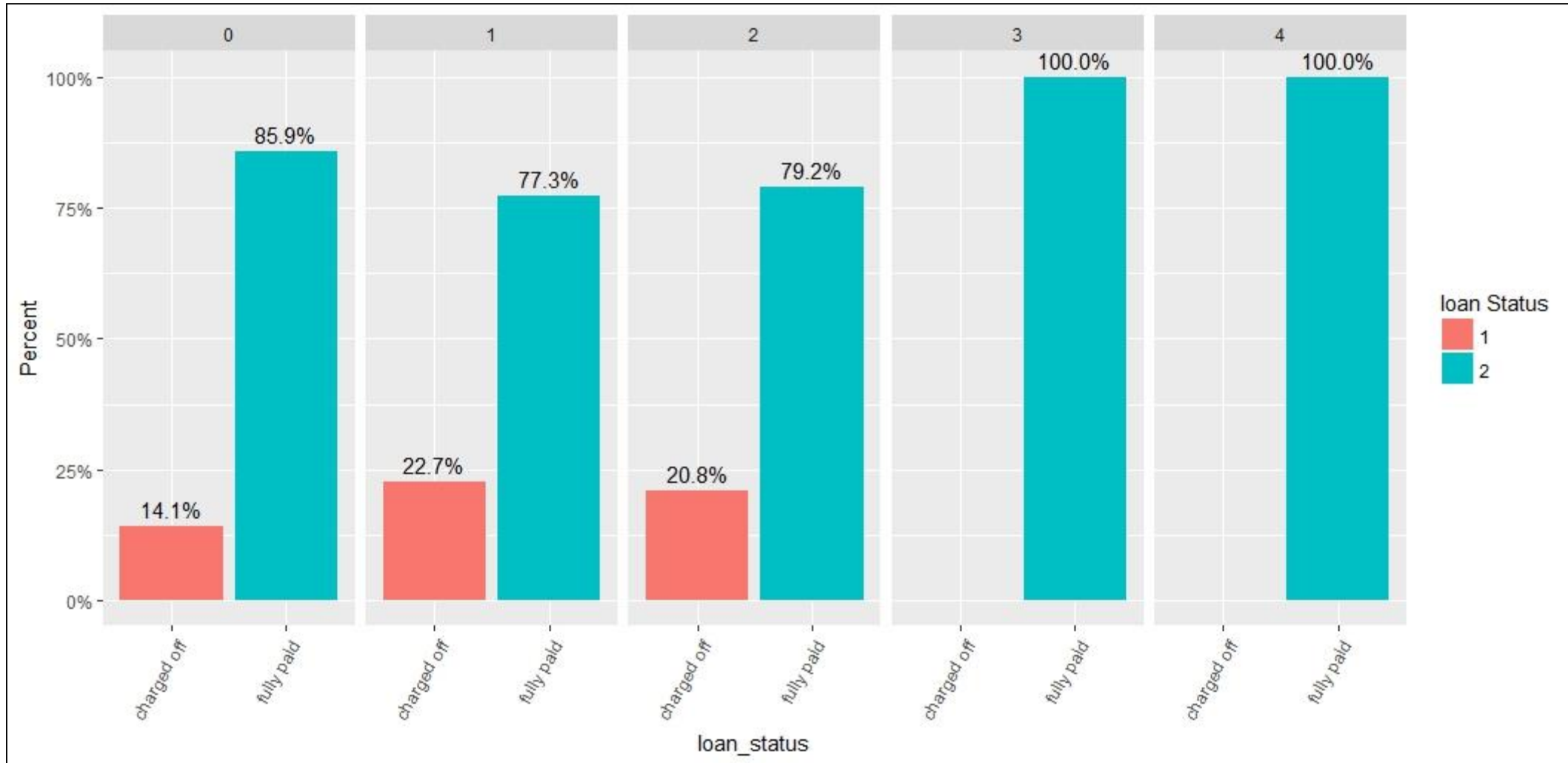
loan borrowers with **one derogatory public records are more likely to default**.

Univariate Analysis

Plot shows nearly **43%** of loan applicant are not verified income.

Bivariate Analysis
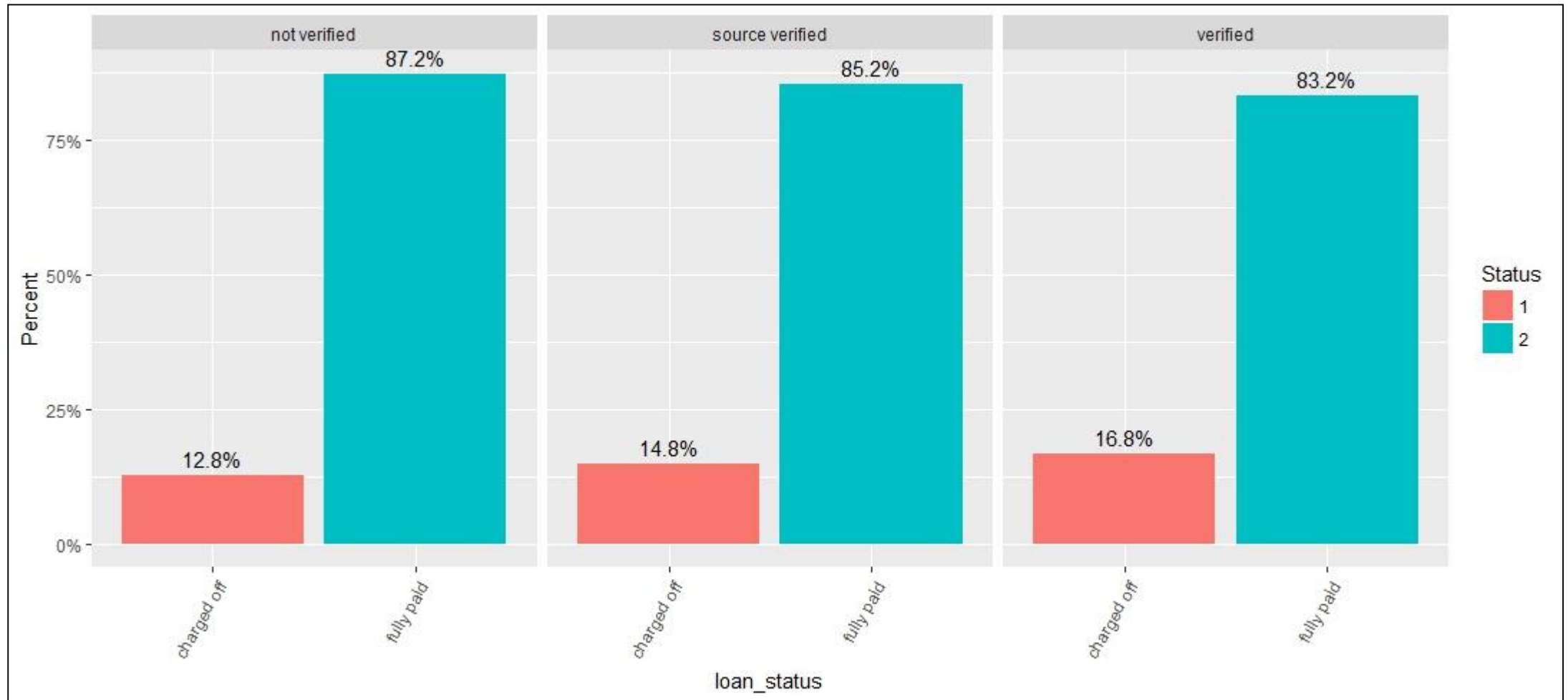
loan borrowers with **verified status** are more likely to **default**.

Univariate Analysis
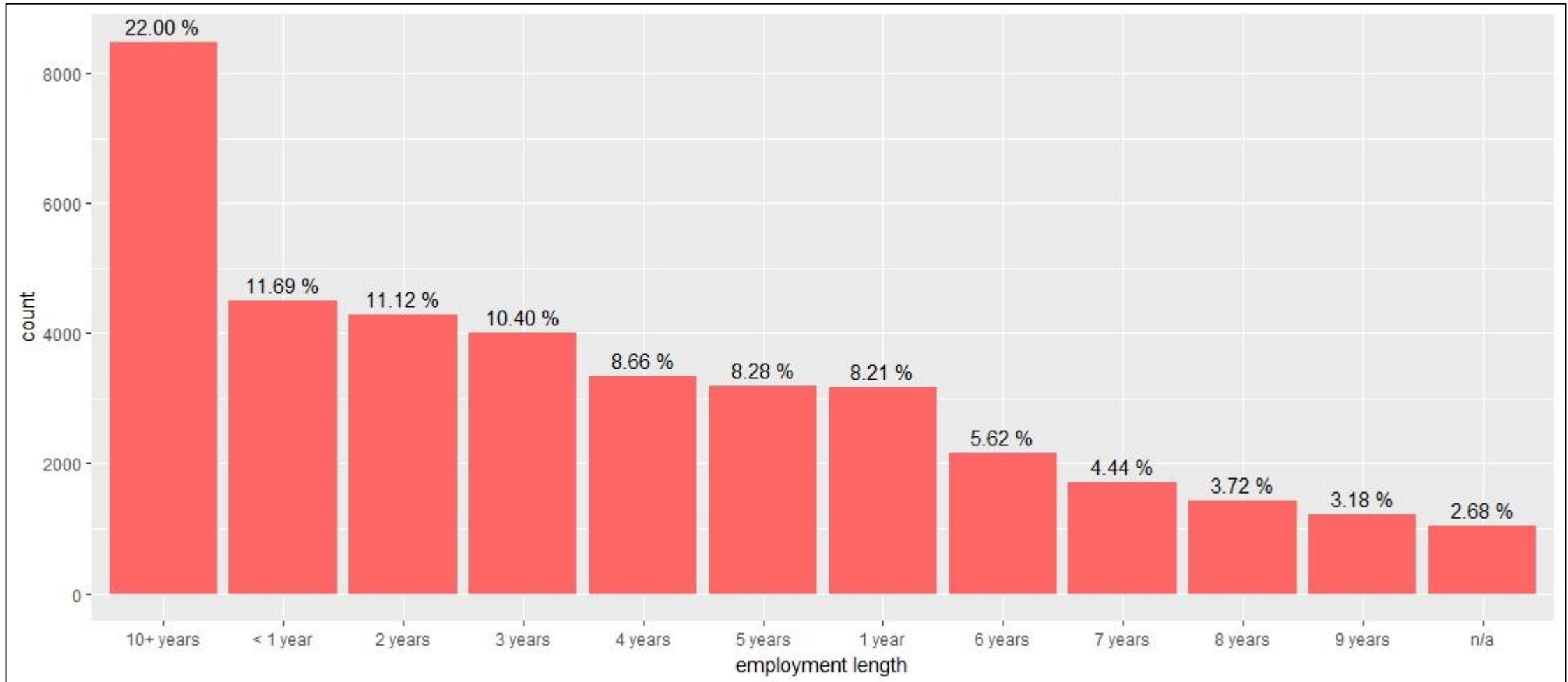
Plot shows 22% loan applicant have 10+ years of employment length.

n/a values are not been removed for analysis as we have used n/a values for loan applicant without job.

Bivariate Analysis

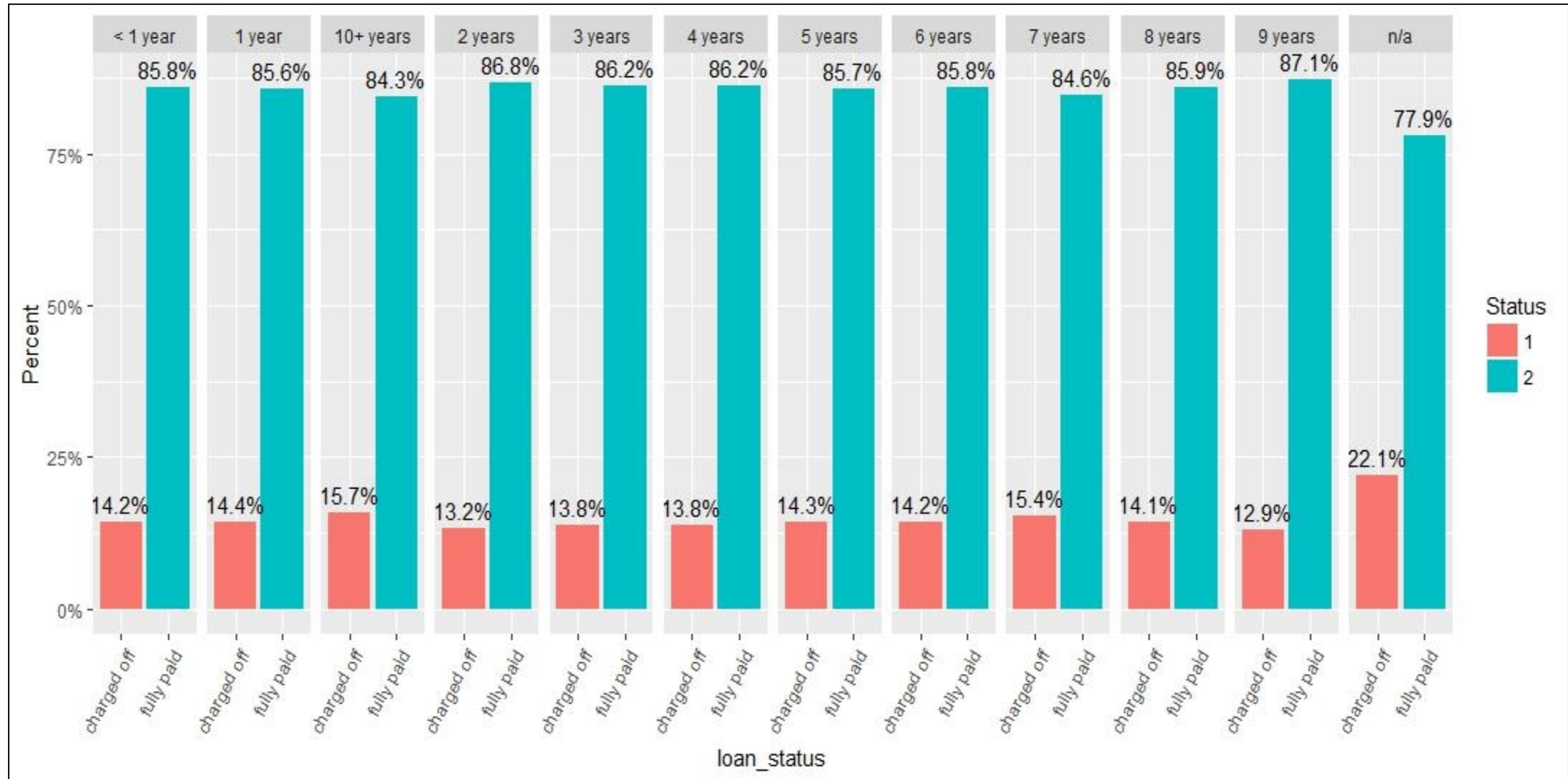Loan borrowers with **10+ years of employment length are more likely to default.**

Univariate Analysis of annual income
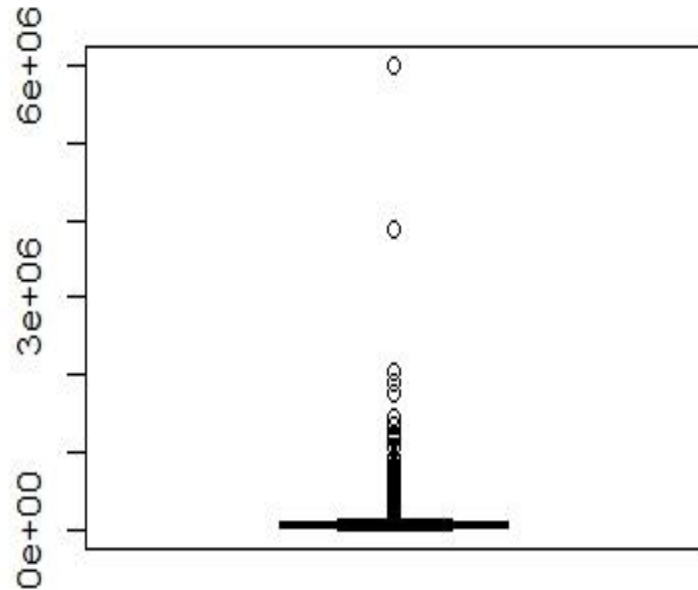
Min. = 4000
1st Qu. = 40000
Median = 58868
Mean = 68778
3rd Qu. = 82000
Max. = 6000000

Remove outlier above 150000.
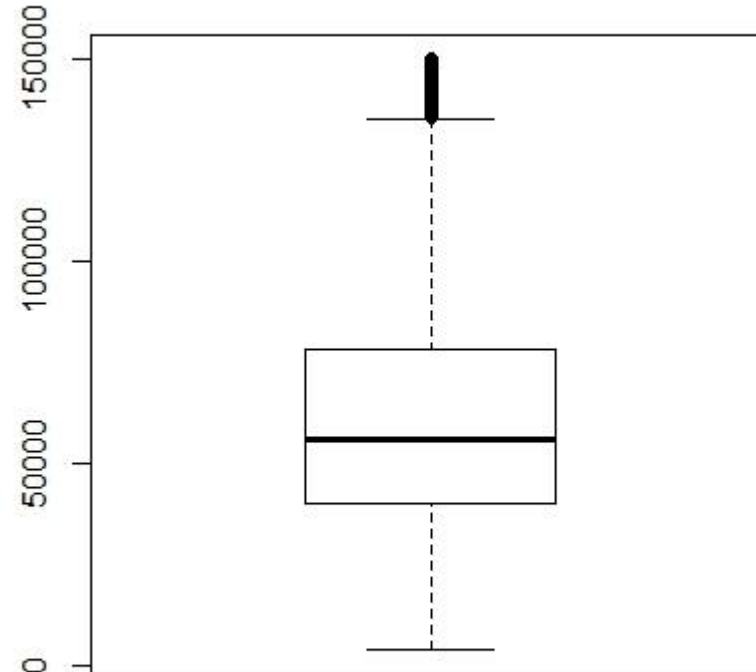
Min. = 4000
1st Qu. = 40000
Median = 56000
Mean = 61342
3rd Qu. = 78000
Max. = 150000

Bivariate Analysis

Annual income is divided into different slots:
0-50000 = low
50001-100000 = medium
100001-150000 = high

Loan borrower with low annual income are most likely to default in comparison to other.

Univariate Analysis of installment

Min. = 15.69
1st Qu. = 165.74
Median = 277.86
Mean = 322.47
3rd Qu. = 425.77
Max. = 1305.19
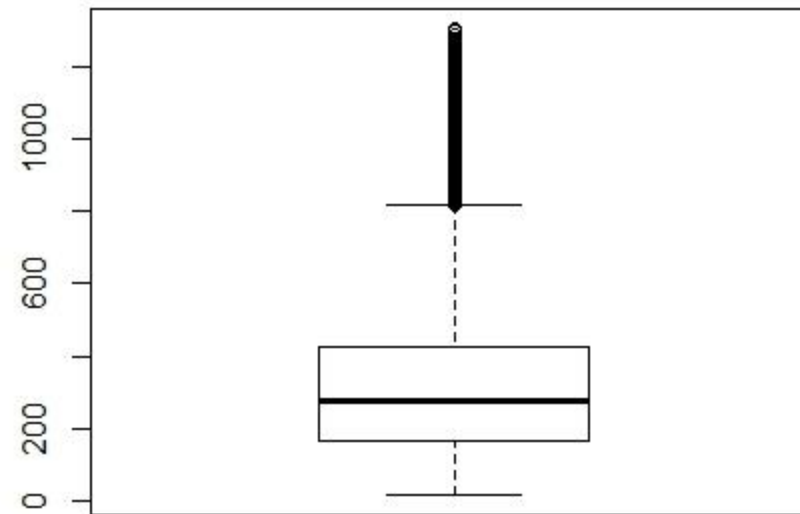
There is considerable amount of outliers which needed to removed for analysis

Bivariate Analysis

Installment is divided
into different slots:
0-200 = low
201-400 = medium
> 401 = high

Loan borrower with
high installment are
most likely to default
in comparison to
other.

Bivariate Analysis
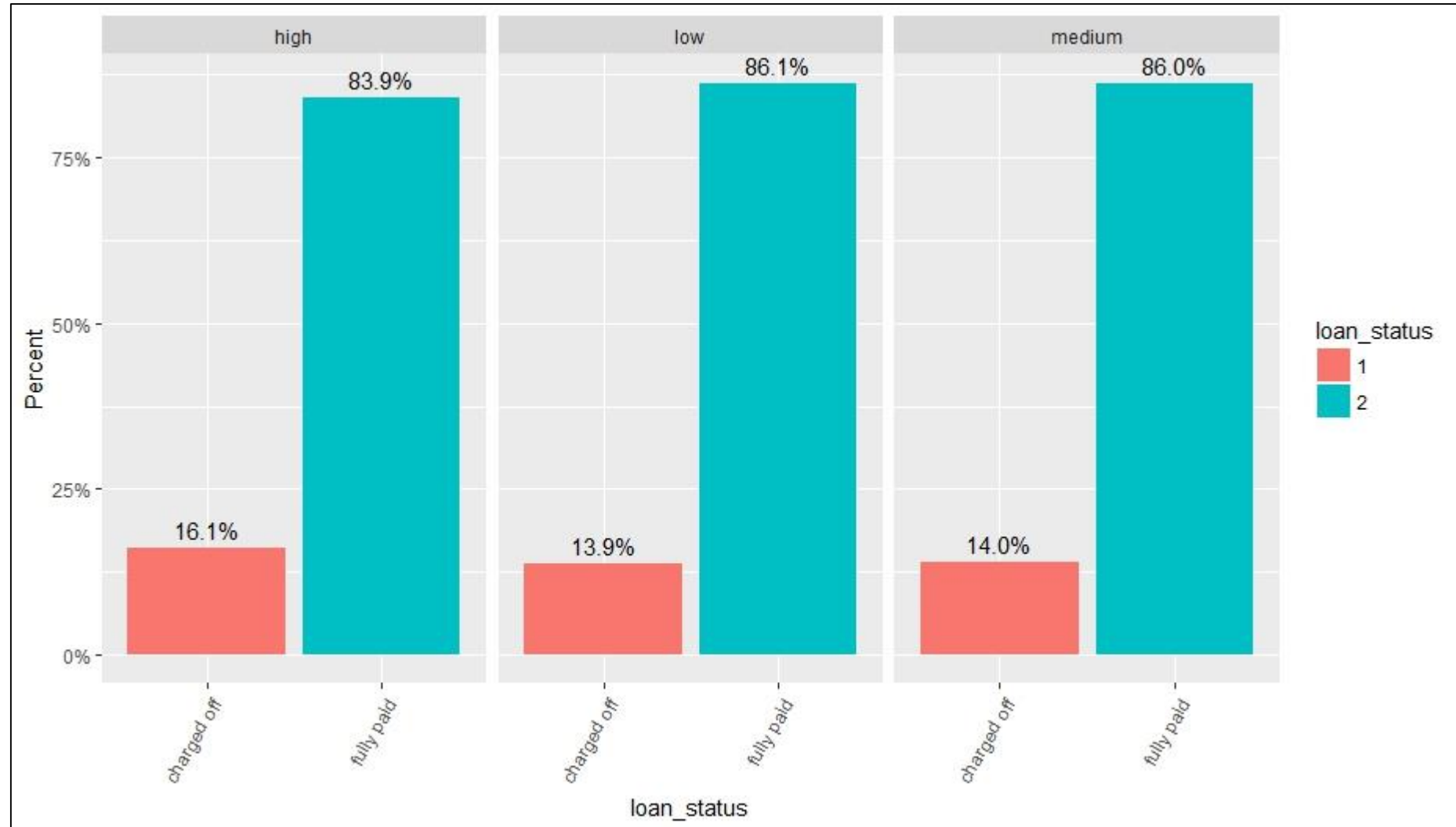
Installment is divided into different slots:
0-200 = low
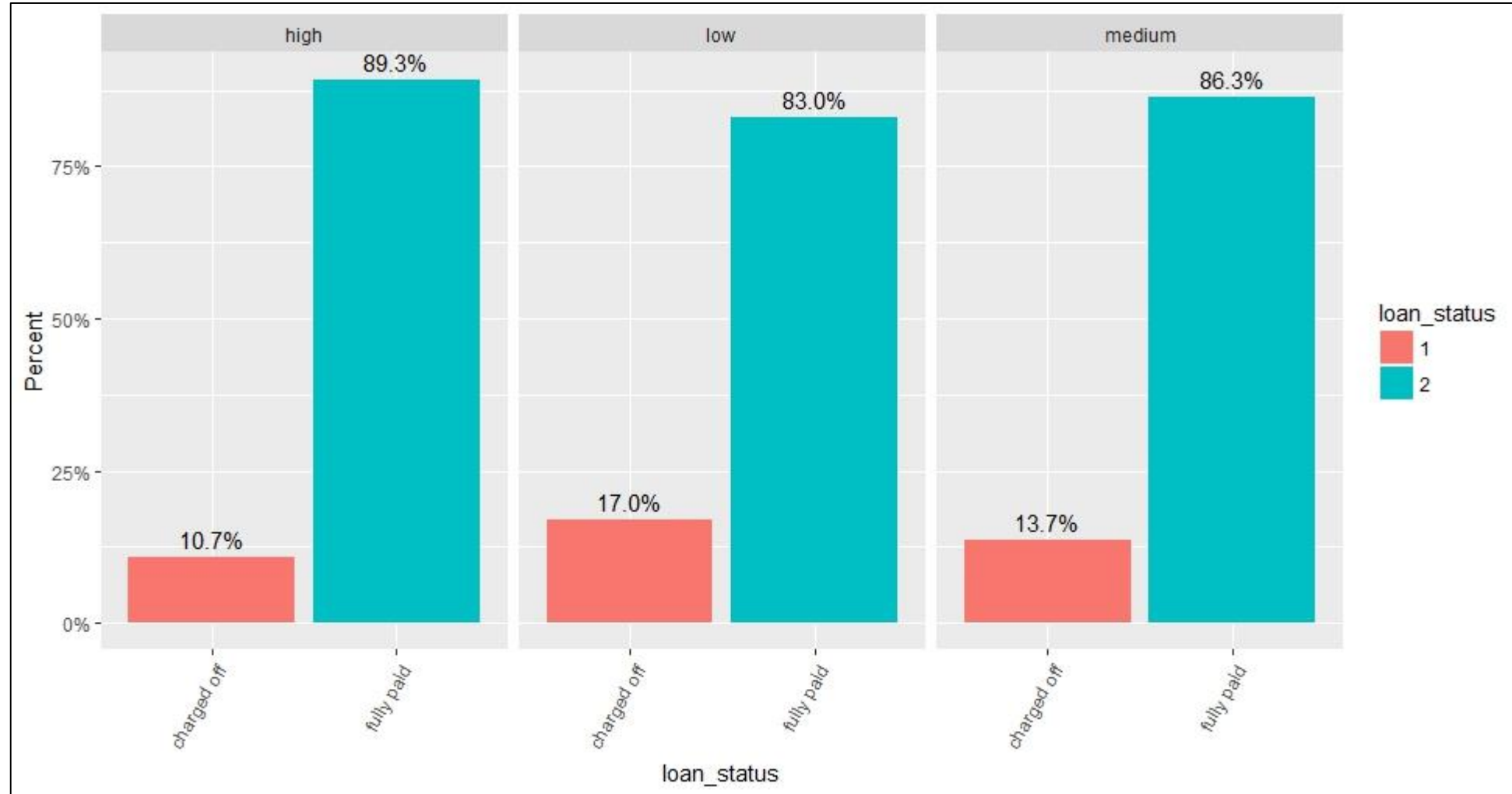201-400 = medium
> 401 = high

Annual income is divided into different slots:
0-50000 = low = green
50001-100000 = medium = blue
100001-150000 = high = pink
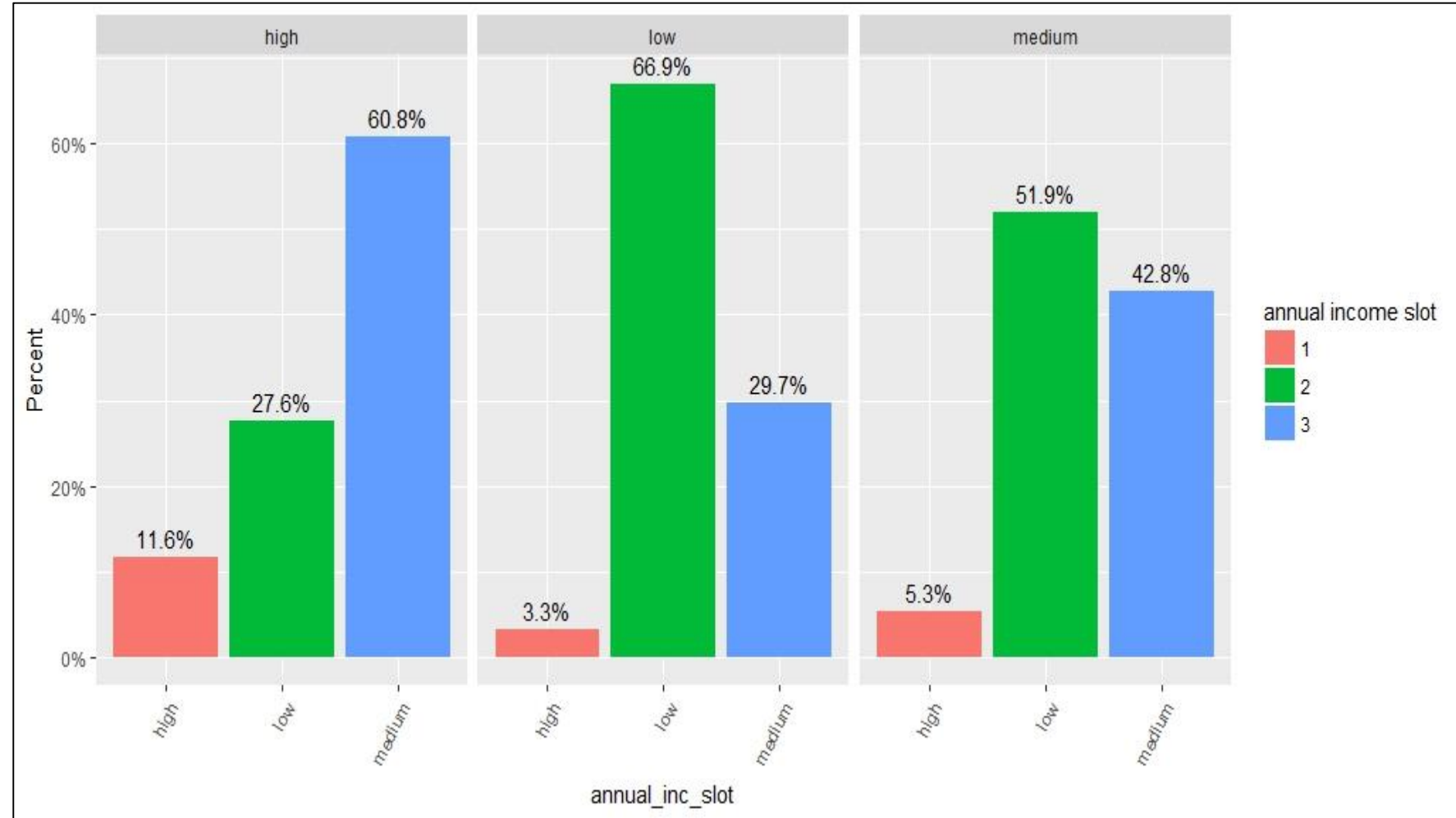
Loan borrower with low annual income and low installment has defaulted the most.

- ❑ Till now we have univariate and bivariate analysis and we have find out the percentage of defaulters in different variables like term, purpose, employment length etc.

- ❑ Now for more depth analysis and in order to identify important driver variables which are strong indicators of default, we have narrow down our analysis to "Charged Off" status only.

- ❑ As we have got 4 major purpose for which applicant are borrowing loan, therefore we have also narrow down our analysis to these 4 purposes (debt consolidation, credit card, home improvement and major purchase)

- ❑ We will doing analyzing the default percentage of different variable i.e. term, grade, verification status etc. for these 4 above said purposes.

Plot shows default percentage of 4 purposes for different term period.

We are taking maximum percentage for analysis

Debt consolidation = **39.43% for 36 month term**
Credit card = **8.45% for 36 month term**
Home improvement = **5.32% for 36 month term.**
Major purchase = **3.65% for 36 month term**.
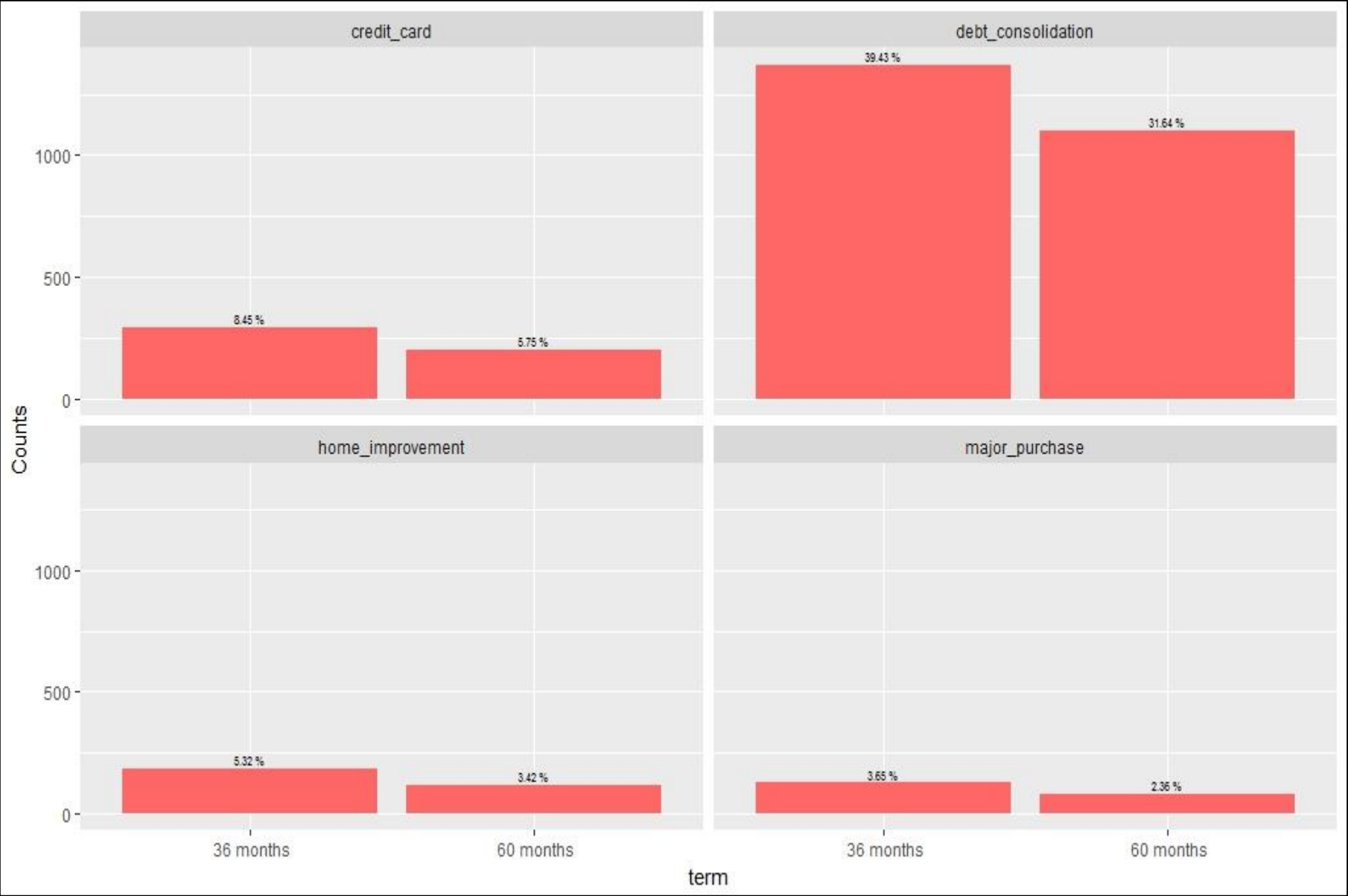
Plot shows default percentage in 4 purposes w.r.t grade.

We are taking maximum percentage for analysis.

Debt consolidation = **17.41% for b grade**
Credit card = **3.97% for b grade**
Home improvement = **2.53% for b grade.**
Major purchase = **1.61% for b grade.**
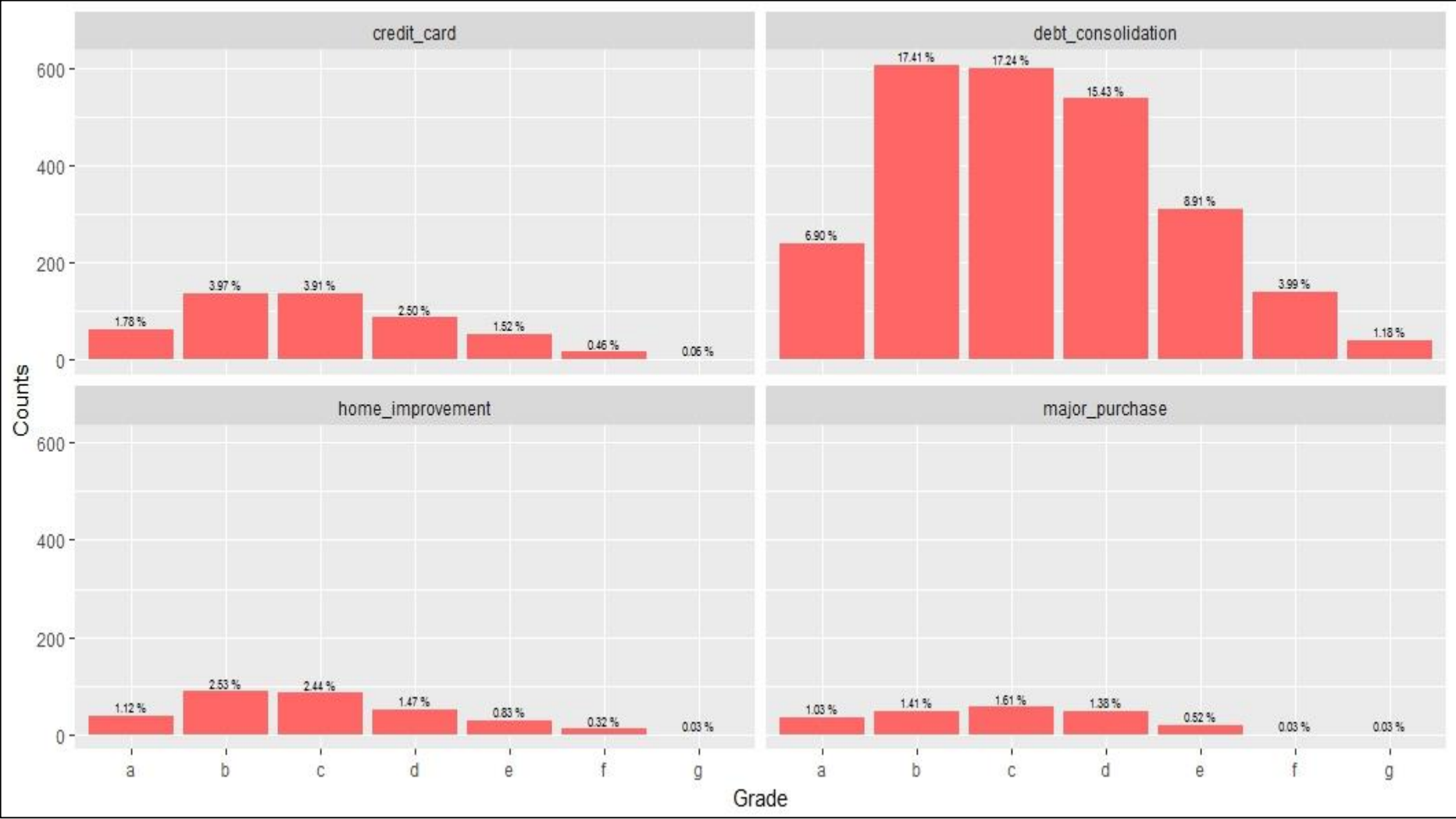
Plot shows default percentage of 4 purposes for different Interest rate slot.

We are taking maximum percentage for analysis.

Debt consolidation = **34.91% for medium(10-15%) interest rate.**
Credit card = **7.79% for medium(10-15%) interest rate.**
Home improvement = **4.66% for medium(10-15%) interest rate.**
Major purchase = **3.22% for medium(10-15%) interest rate.**
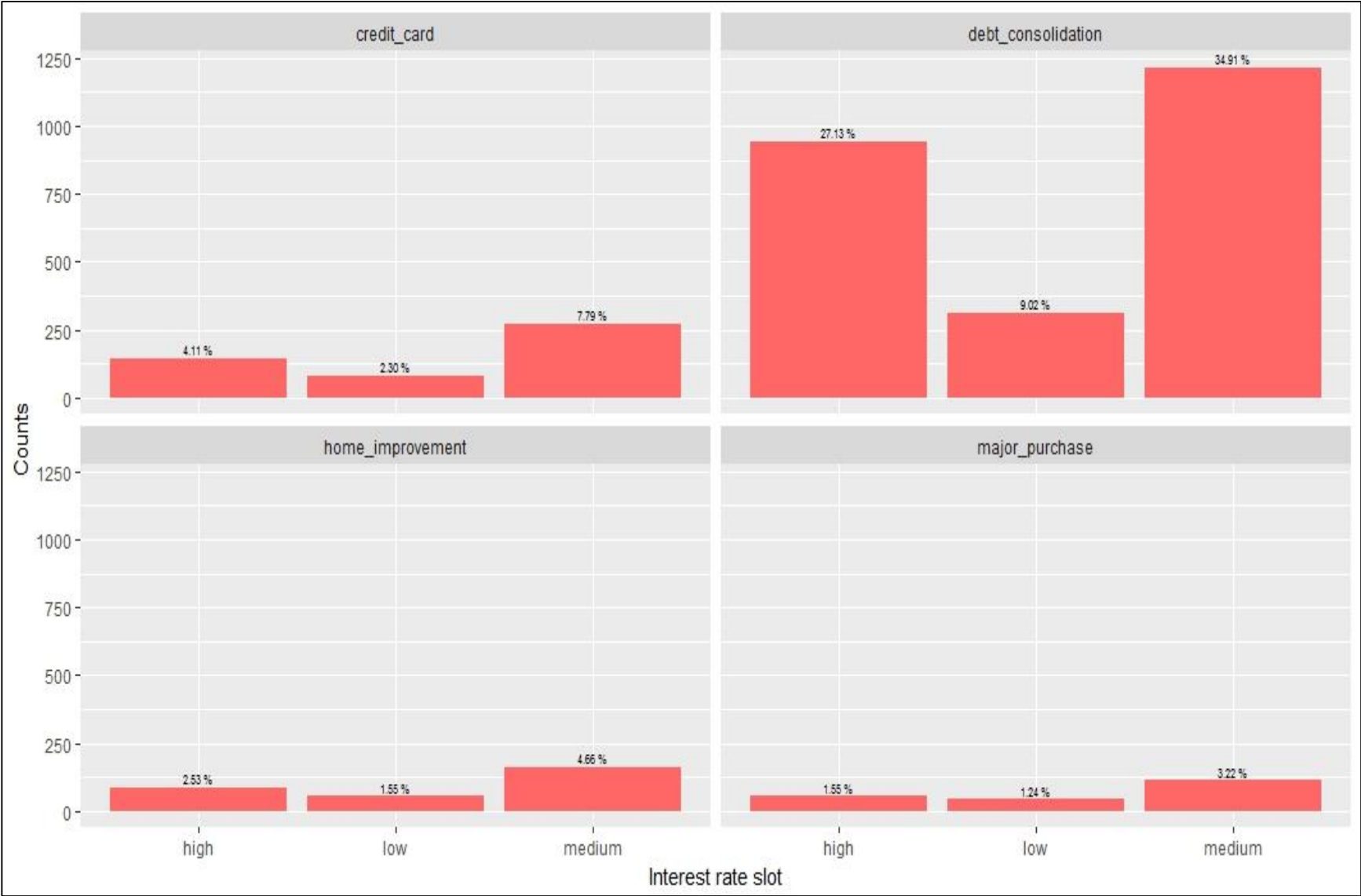
Plot shows default percentage of 4 purposes for different home ownership.

We are taking maximum percentage for analysis.

Debt consolidation = **37.76% for rented home.**
Credit card = **7.44% for rented home.**
Home improvement = **6.58% for mortgaged home.**
Major purchase = **3.28% for rented home.**
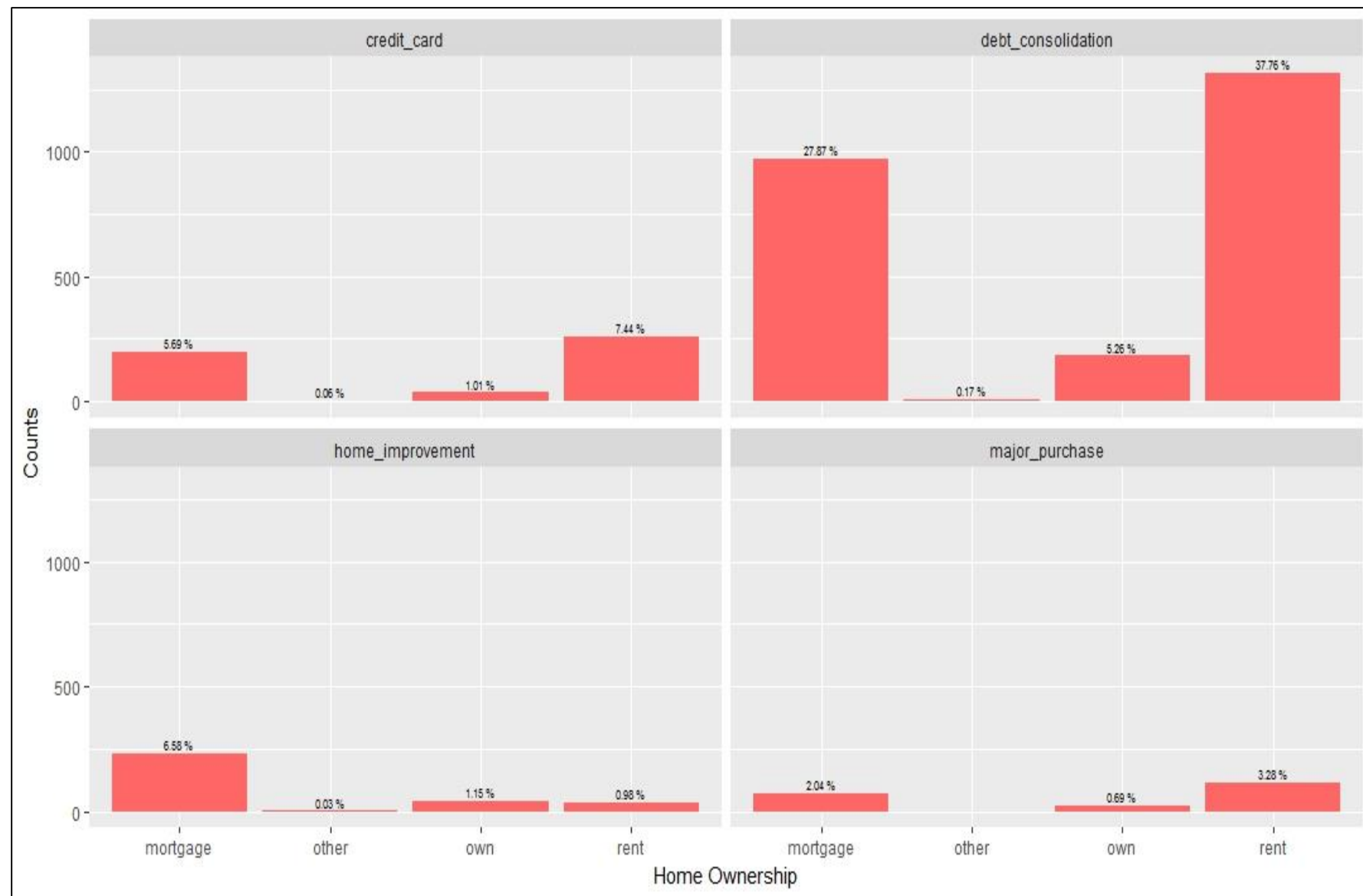
Plot shows default percentage of 4 purposes for Annual income slot.

We are taking maximum percentage for analysis.

Debt consolidation = **35.78% for low (0-50000) annual income.**
Credit card = **6.98% for low (0-50000) annual income.**
Home improvement = **4.57% medium (50000-100000) annual income.**
Major purchase = **3.65% for low (0-50000) annual income.**
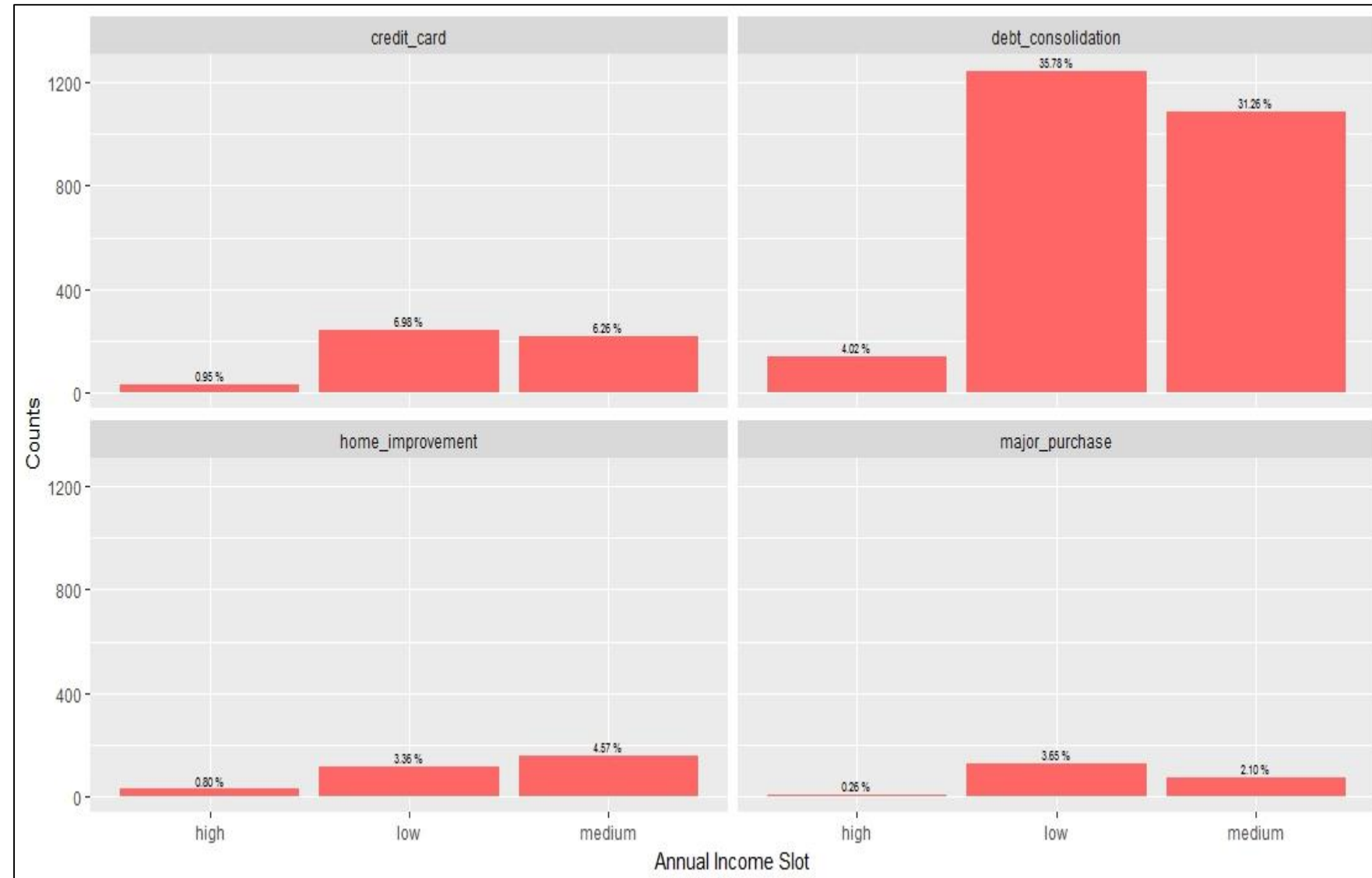
Plot shows default percentage of 4 purposes for loan amount slot.

We are taking maximum percentage for analysis.

Debt consolidation = **38.76% for medium (5000-15000) loan amount.**
Credit card = **8.42% for medium (5000-15000) loan amount.**
Home improvement = **4.71% medium (5000-15000) loan amount.**
Major purchase = **2.84% for medium (5000-15000) loan amount.**

Plot shows default percentage of 4 purposes for employment length.

We are taking maximum percentage for analysis.

Debt consolidation = **17.01% for 10+ years of employment length.**
Credit card = **2.96% for 10+ years of employment length.**
Home improvement = **2.50% for 10+ years of employment length.**
Major purchase = **1.09% for 10+ years of employment length.**
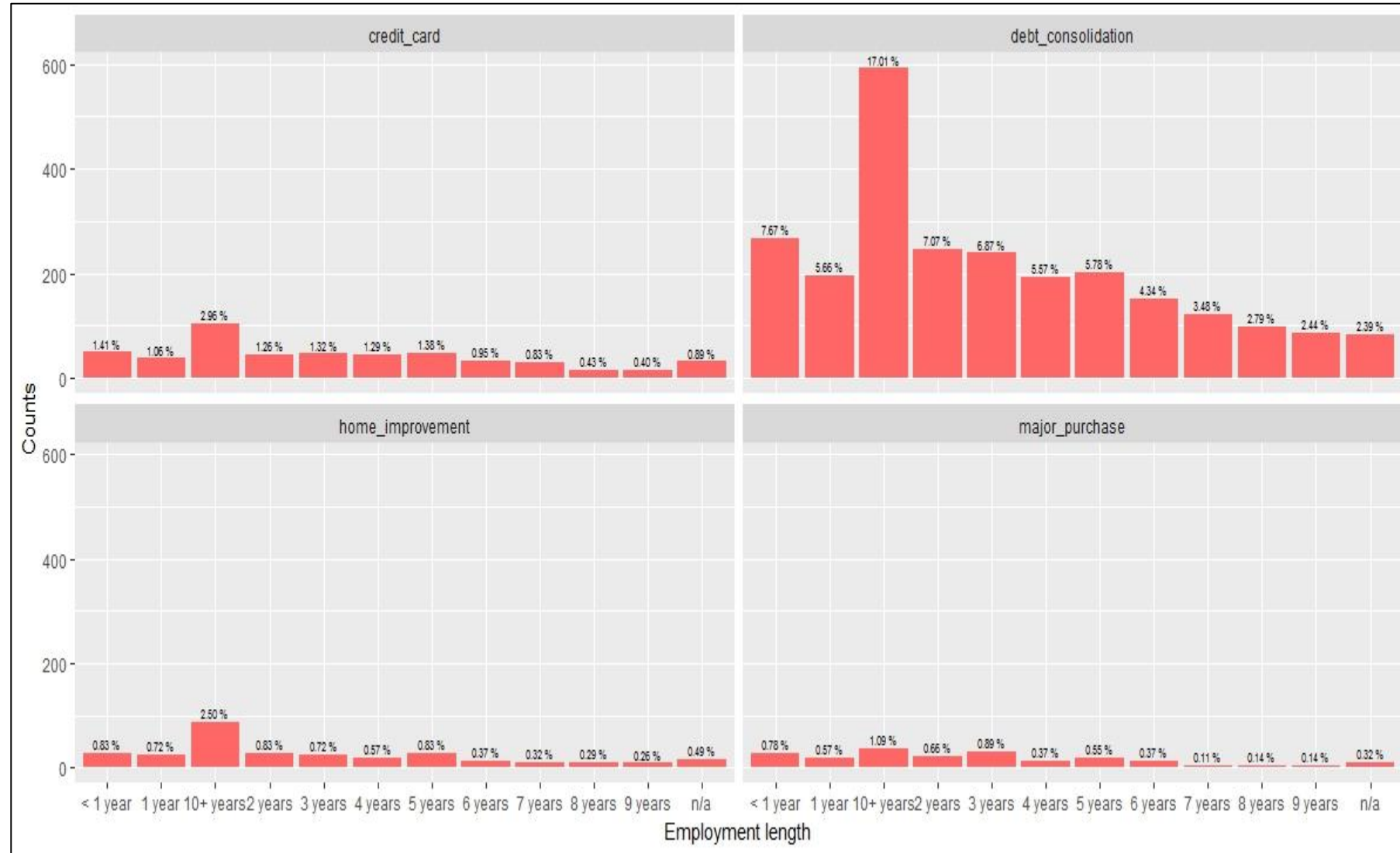
Plot shows default percentage of 4 purposes for Installment slots.

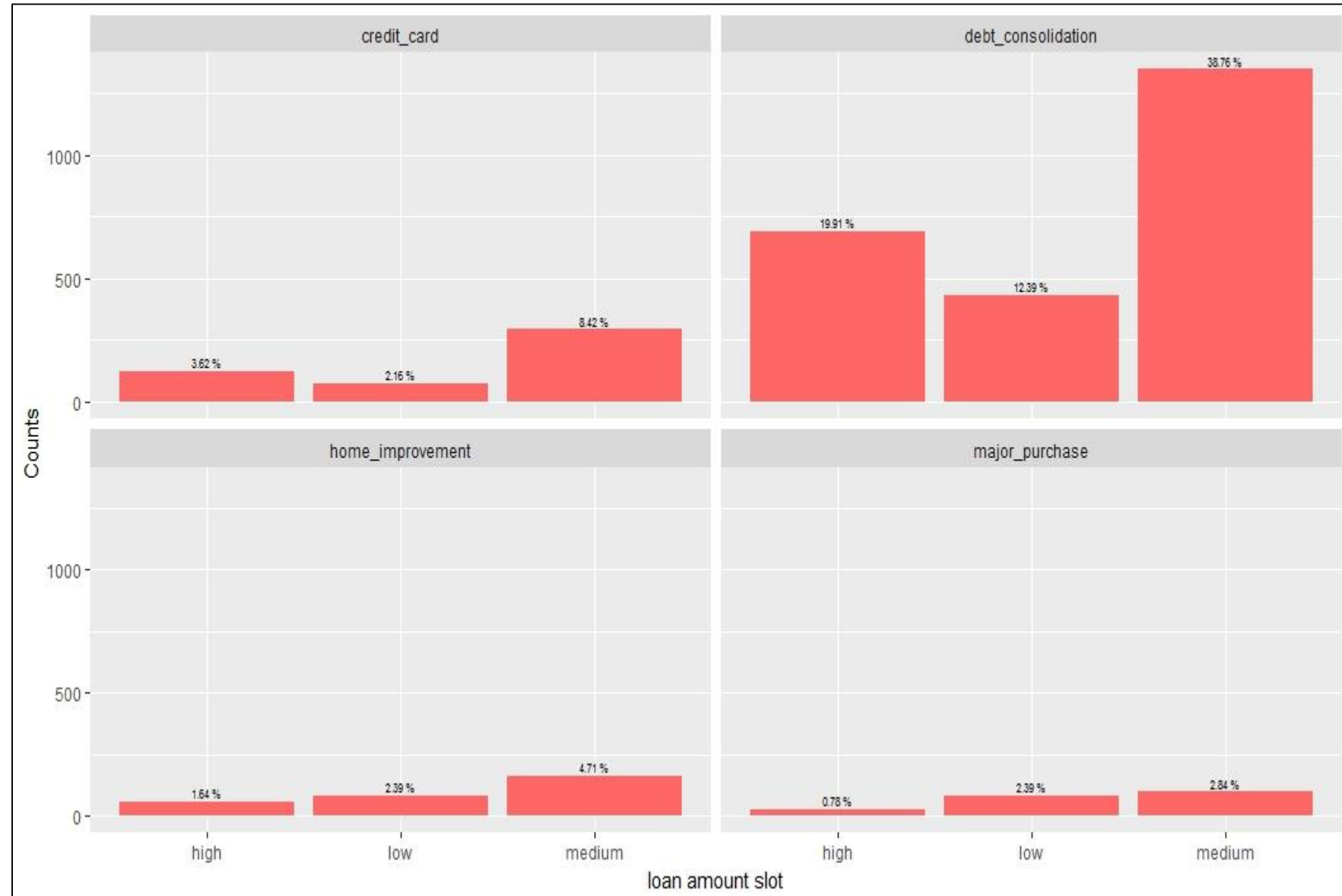We are taking maximum percentage for analysis.

Debt consolidation = **31.03% for medium(200-400) installment.**
Credit card = **6.38% for medium(200-400) installment.**
Home improvement = **3.42% for medium(200-400) installment.**
Major purchase = **3.13% for medium(200-400) installment.**
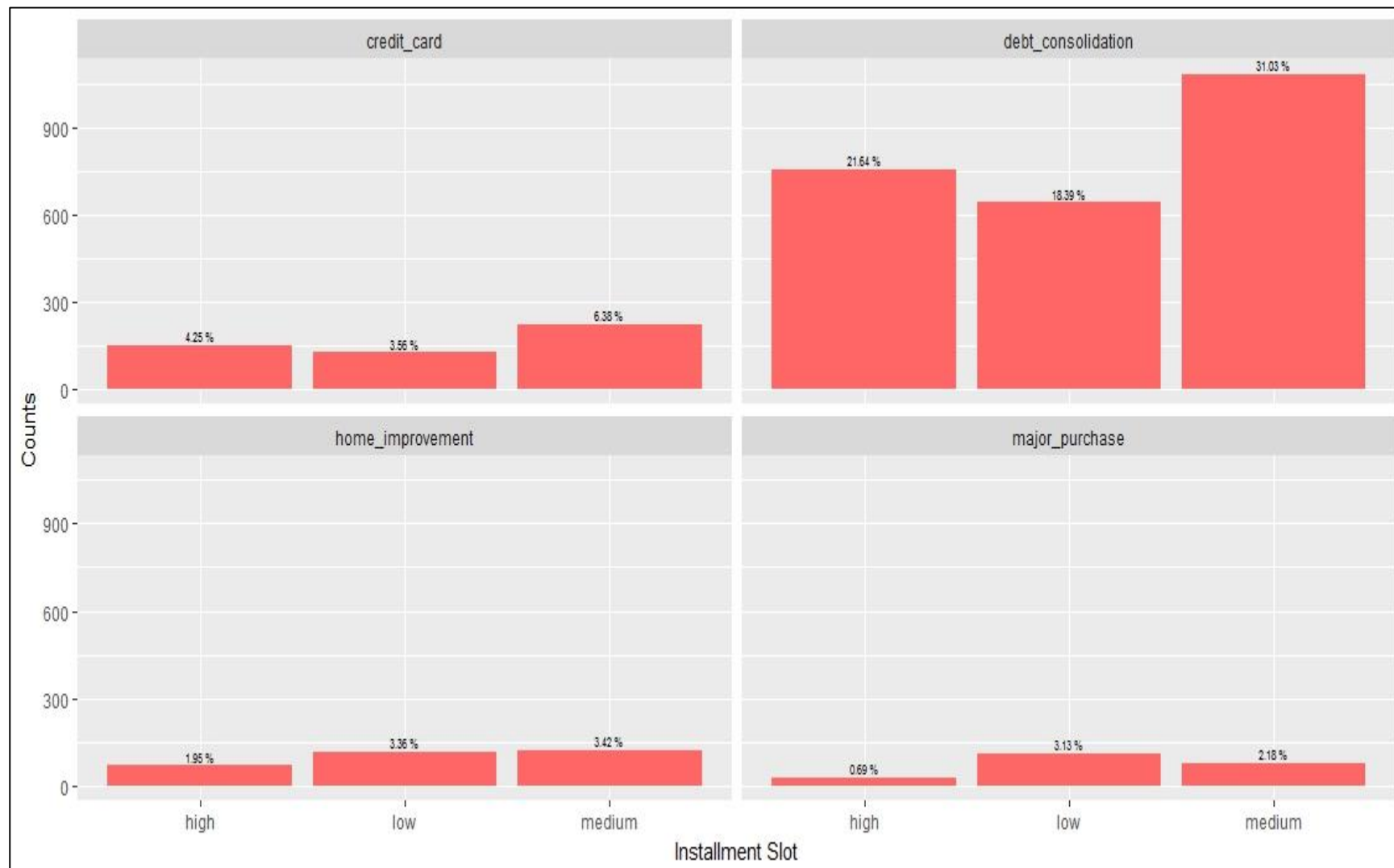
Plot shows default percentage of 4 purposes for verification status.
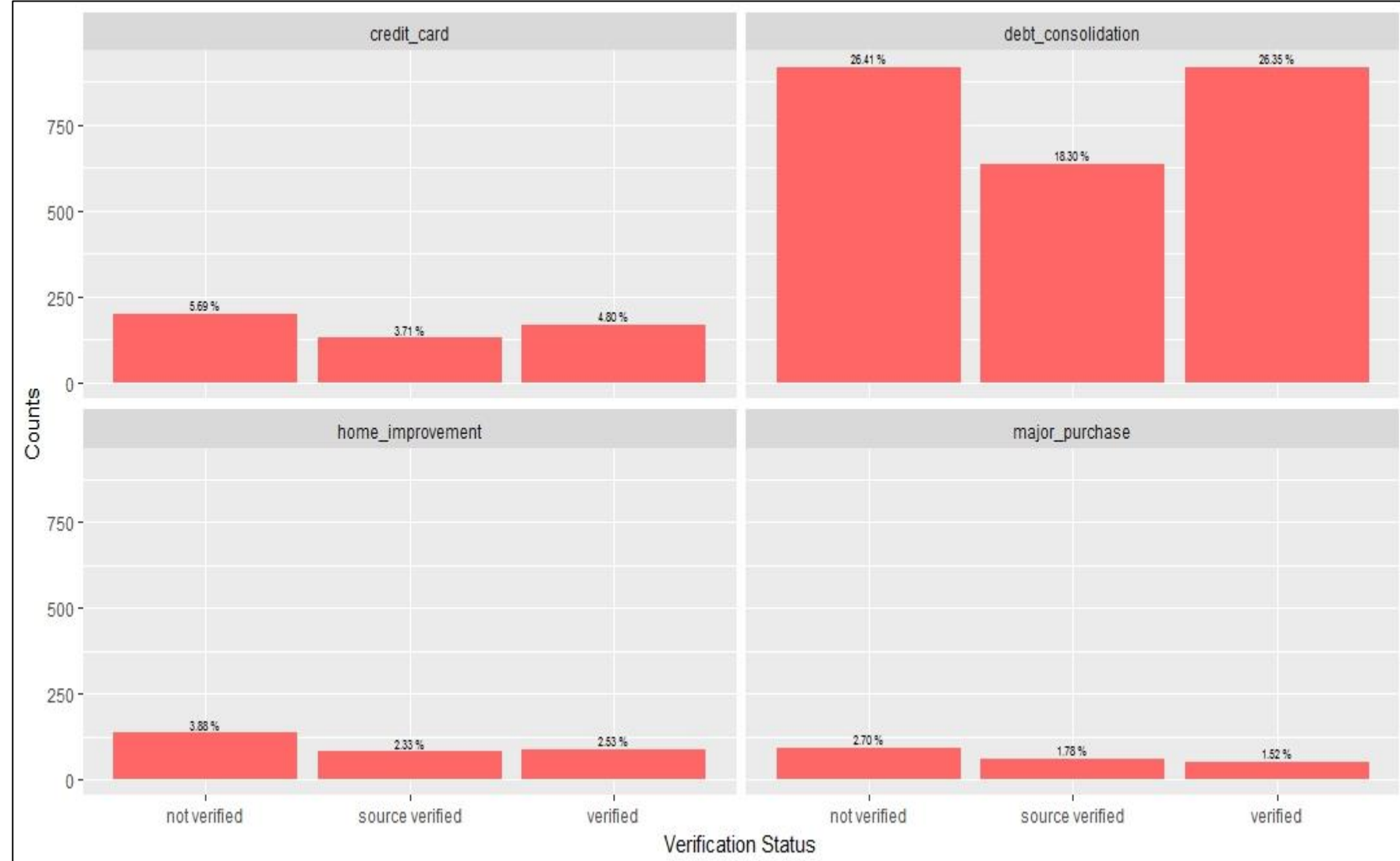
We are taking maximum percentage for analysis.

Debt consolidation = **26.41% for not verified status.**
Credit card = **5.69% for not verified status.**
Home improvement = **3.88% for not verified status.**
Major purchase = **2.70% for not verified status.**

S.No	variables	Debt_consolidation	Credit card	Home_improvement	Major_purpose
1	TERM	**36 months = 39.43%**	**36months = 8.45%**	**36months = 5.32%**	**36months = 3.65%**
2	GRADE	B = 17.41%	B = 3.97%	B = 2.53%	B = 1.61%
3	INTESREST RATE SLOT	**Medium = 34.91%**	**Medium = 7.79%**	**Medium = 4.66%**	**Medium = 3.22%**
4	HOME OWNERSHIP	**Rent = 37.76%**	**Rent = 7.44%**	**Mortgage = 6.58%**	**Rent = 3.28%**
5	ANNUAL INCOME SLOT	**Low = 35.78%**	**Low = 6.98%**	**Medium = 4.57%**	**Low = 3.65%**
6	LOAN AMOUNT SLOT	**Medium = 38.76%**	**Medium = 8.42%**	**Medium = 4.71%**	Medium = 2.84%
7	EMP_LENGTH	10+ Years= 17.01%	10+ Years =2.96%	10+ Years = 2.50 %	10+ Years = 1.09%
8	INSTALLMENT SLOT	Medium = 31.03%	Medium = 6.38%	Medium = 3.42%	**Low = 3.13%**
9	VERIFICATION STATUS	Not Verified = 26.41%	Not verified = 5.69%	Not verified = 3.88%	Not verified = 2.70%

Result

Important variable for debt consolidation:	Important variable for Credit card purpose:	Important variable for Home improvement purpose:	Important variable for major purchase purpose:
• Term	• Term	• Term	• Term
• Interest rate slot	• Interest rate slot	• Interest rate slot	• Interest rate slot
• Home ownership	• Home ownership	• Home ownership	• Home ownership
• Annual income slot	• Annual income slot	• Annual income slot	• Annual income slot
• Loan amount slot	• Loan amount slot	• Loan amount slot	• Installment slot

CONCLUSION

- Data quality issue are correctly identified and cleaned for analysis.

- New variables are introduced for useful analysis.

- Univariate analysis and Bivariate analysis is being done to analyse default percentage of loan borrower for different variables.

- Important driver variables are identified for four different major purpose for which loan application were received,

- Term, Interest rate slot, Home ownership, Annual income slot, Loan amount slot are driver variables for debt consolidation, credit card and home improvement purposed.

- Term, Interest rate slot, Home ownership, Annual income slot, Installment slot are driver variables for major purchase purpose.

THANKS