A Case study on

LendingClub
Loans

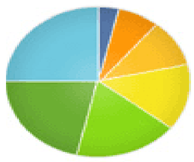
The Business Objective

About Lending Club

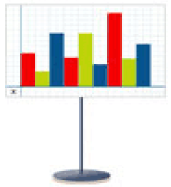
Lending Club (LC) is a peer-to-peer online lending platform connecting borrowers and investors, enabling borrowers to obtain a loan, and investors to purchase notes backed by payments made on loans.

https://en.wikipedia.org/wiki/Lending_Club

Borrowers apply for loans.
Investors open an account.



Borrowers get funded.
Investors build a portfolio.



Borrowers repay automatically.
Investors earn & reinvest.

Objective

To identify patterns which indicate if a person is likely to default or not using Exploratory Data Analysis. By identifying these risky loan applicants, such defaulting loans can be reduced thereby cutting down the amount of credit loss to the Lending Company

About the Data

Key Attributes for Analysis

Loan Attributes

- Loan amount
- Interest rate
- Installment
- Loan status
- Loan grade/sub-grade,
- Issue date
- Term
- Installment
- purpose

Consumer Attributes

- Home Ownership
- Annual Income
- Verification status
- Address State
- Employment Length

Consumer Financial Attributes

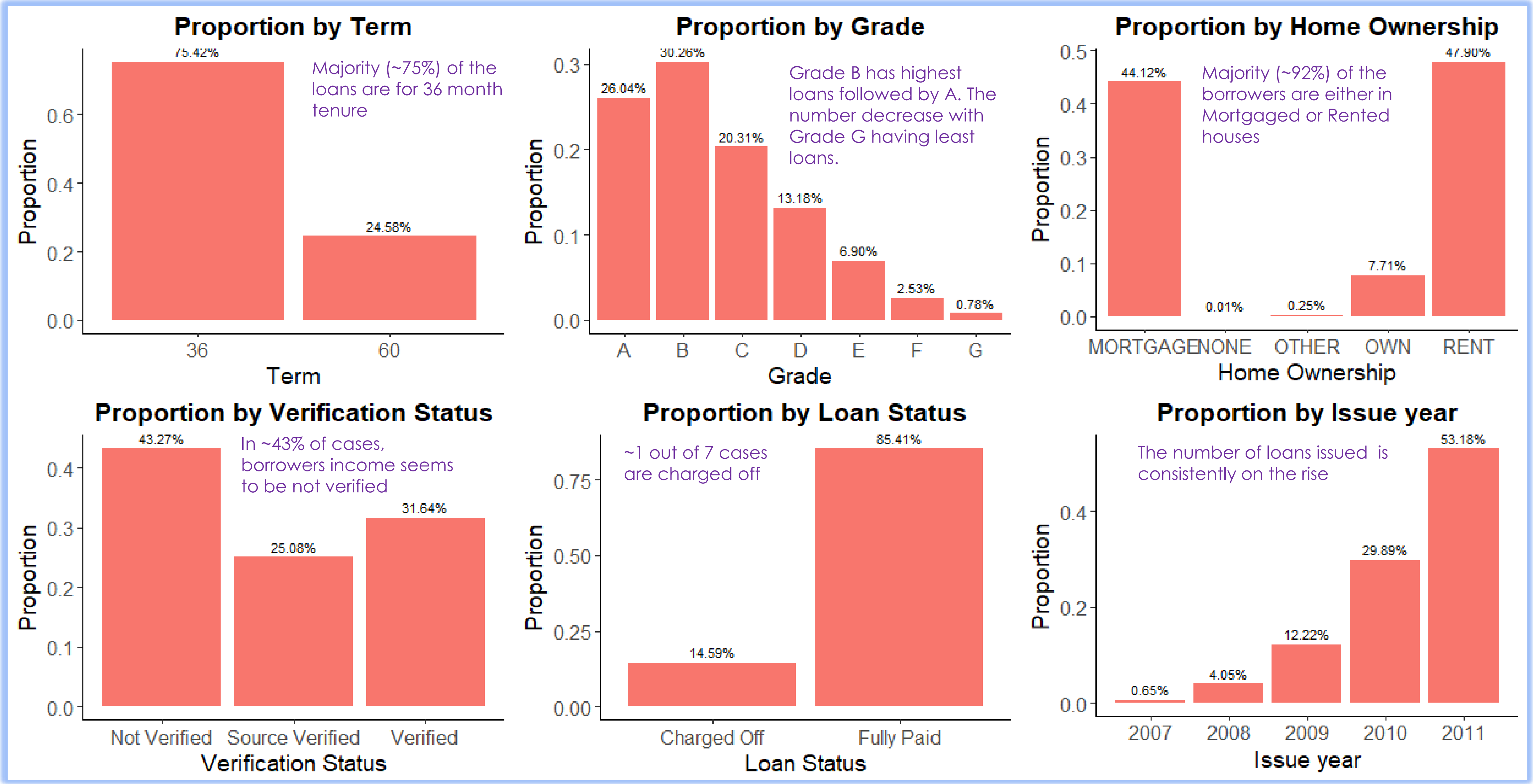
- Debt to income ratio
- Credit line age
- Total_Acc
- Revolving Util. rate
- Revolving Balance
- Pub_rec
- inq_last_6mths
- Earliest_Credit line
- delinq_2yrs

Issues taken care in the given data

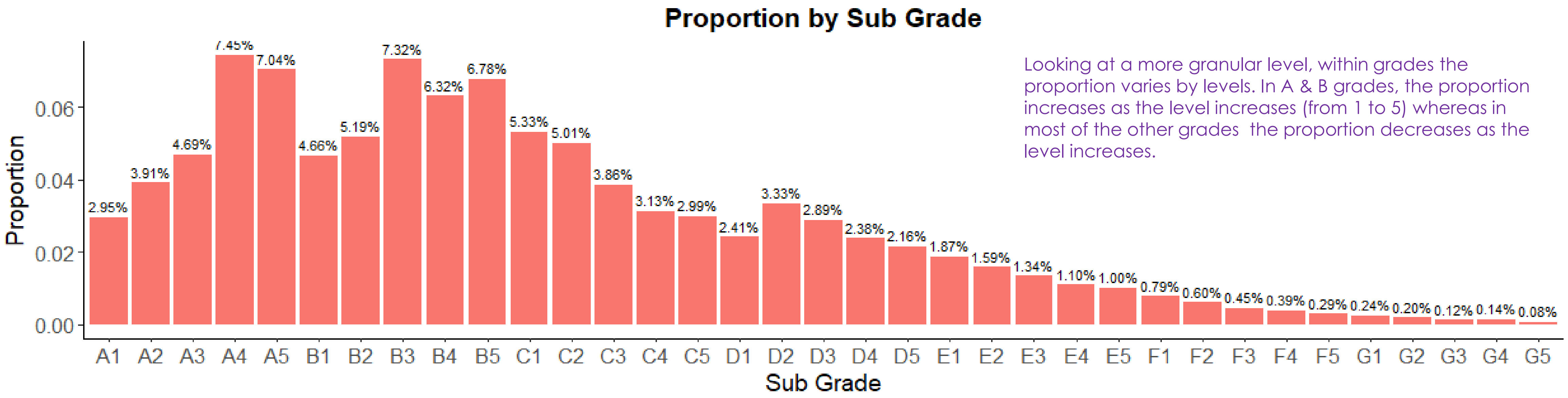
- NA values
- Unnecessary columns that are of not much use for Analysis)
- Dates as strings
- Loan term as string
- Percentages as strings
- Outliers

Univariate Analysis

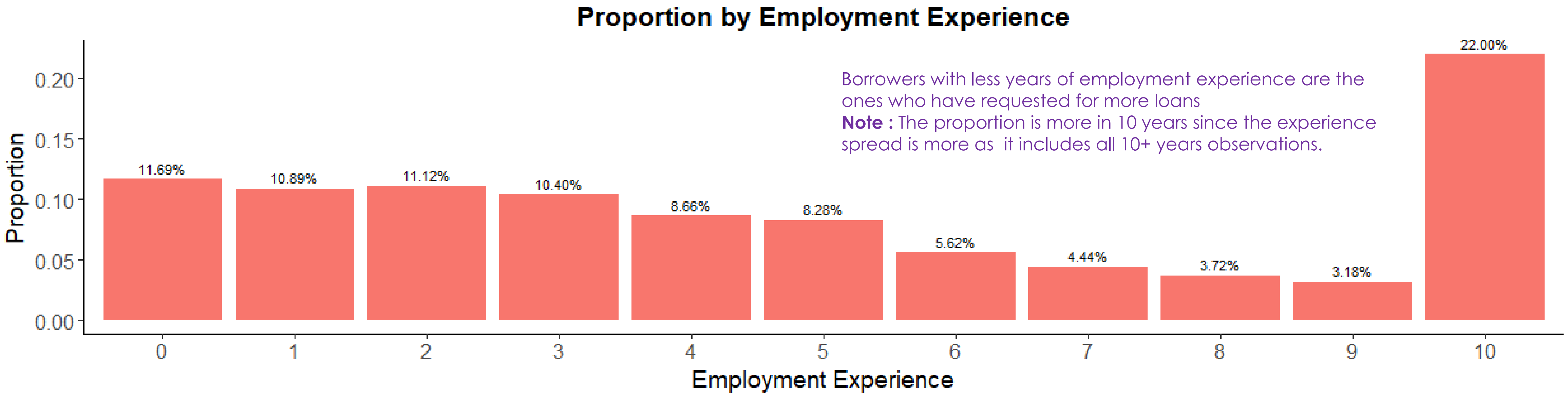
Categorical variables



Univariate Analysis
Categorical variables contd…



Looking at a more granular level, within grades the proportion varies by levels. In A & B grades, the proportion increases as the level increases (from 1 to 5) whereas in most of the other grades the proportion decreases as the level increases.

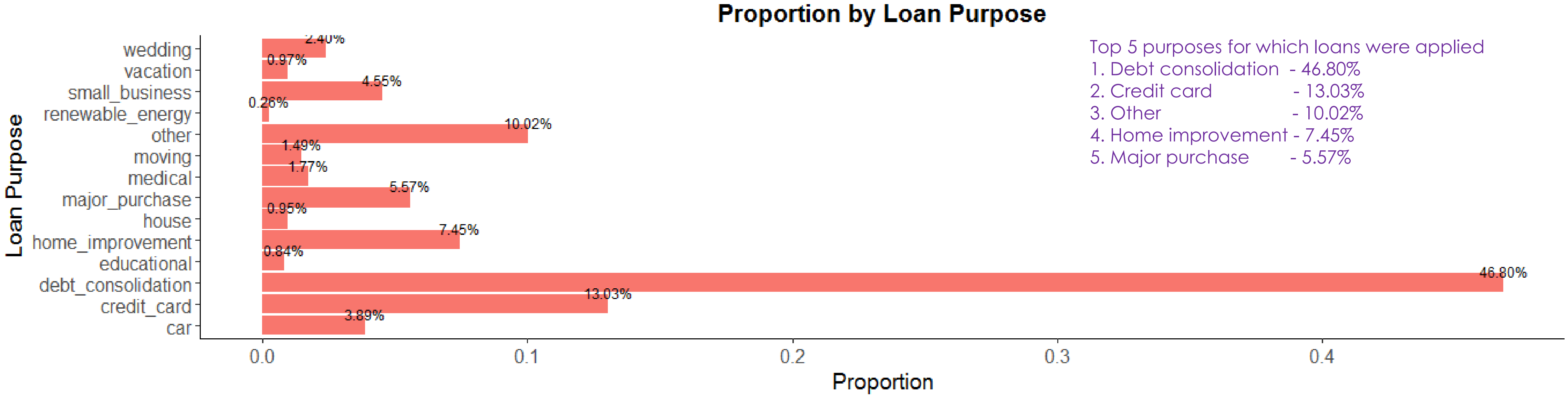


Borrowers with less years of employment experience are the ones who have requested for more loans
Note : The proportion is more in 10 years since the experience spread is more as it includes all 10+ years observations.

Univariate Analysis

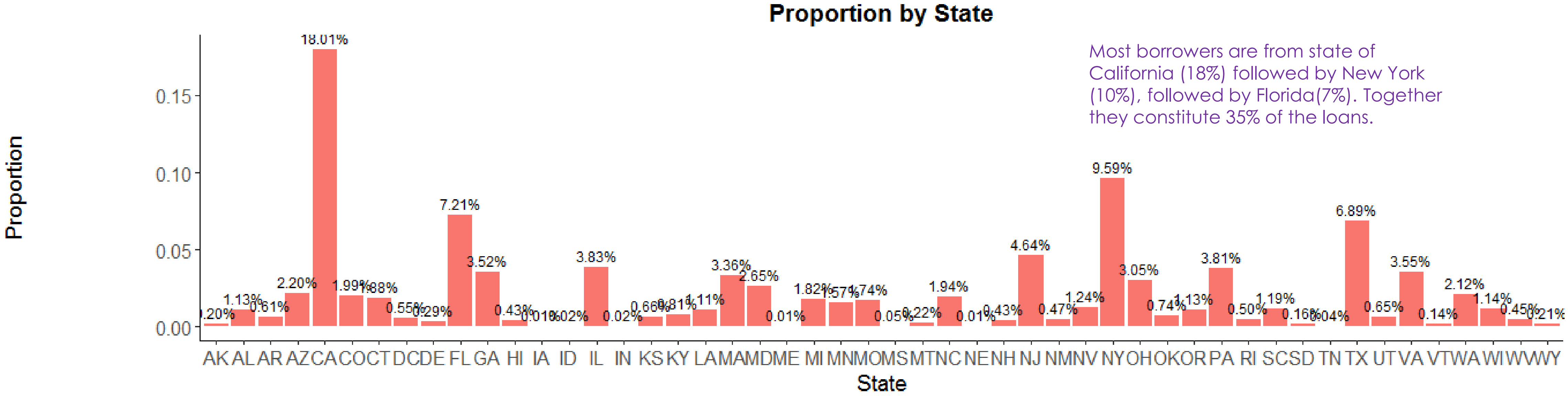
Categorical variables contd...

UpGrad



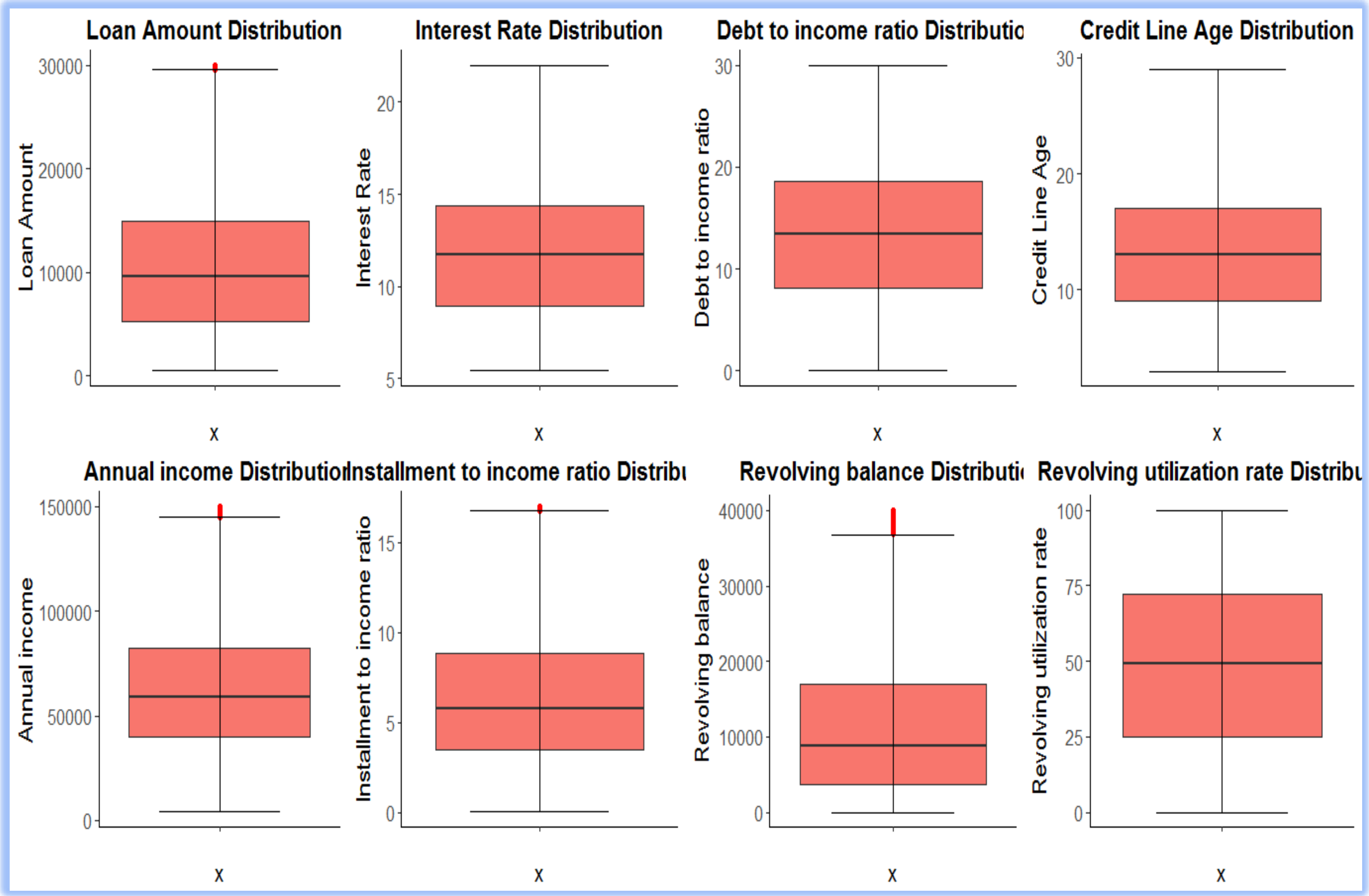
Top 5 purposes for which loans were applied

1. Debt consolidation - 46.80%
2. Credit card - 13.03%
3. Other - 10.02%
4. Home improvement - 7.45%
5. Major purchase - 5.57%



Most borrowers are from state of California (18%) followed by New York (10%), followed by Florida(7%). Together they constitute 35% of the loans.

Univariate Analysis
Continuous variables



- Loan amounts range from $500 to $35000 with an average of $9600
- Outliers greater than $30000 (only 1% records) are capped to $30000

- Interest rates range from 5% to 24% with and average of 11%
- Outliers greater than 23% (0.5% records) are capped to 23%

- Lending Club borrowers has dti ranging from 0 to 30 with an average of 13.

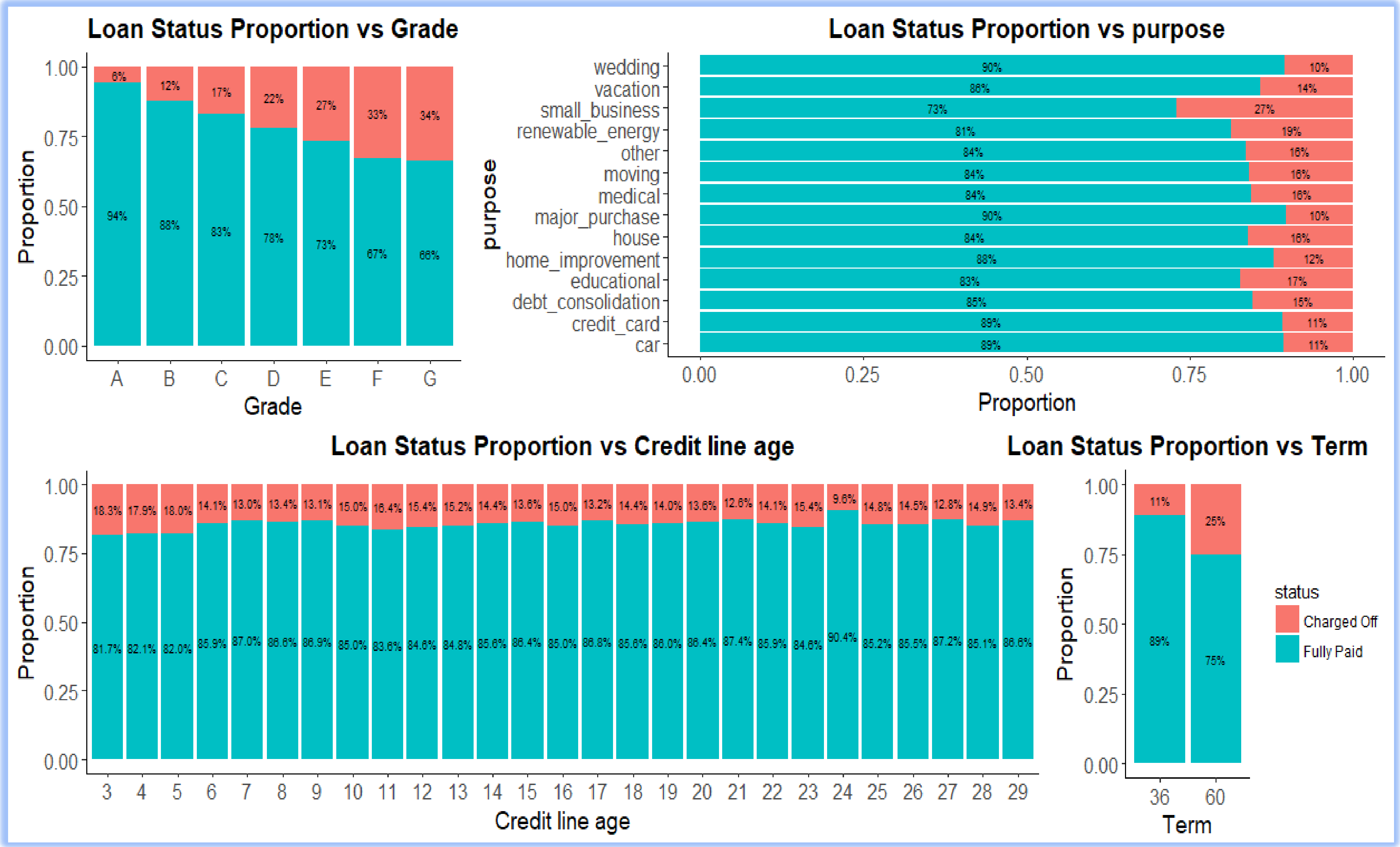
- Lending Club borrowers has credit history ranging from 3 to 51 years with an average of 13 years
- Outliers greater than 29 (only 2% records) are capped at 29

- Average of $58k of annual personal income
- Outliers greater than 150K (only 8% records) are capped at 150k.
- Should not be impacting the analysis on the target variable since we are just capping

- Lending Club borrowers has iti ranging from 0 to 32 with an average of 6.
- Outliers in iti > 17% capped to 17%.

Multivariate Analysis

Categorical variables



Grades

There is a clear relationship between the grade assigned by Lending Club and the loan status as follows.

94 % of A-grade loans are fully paid. **This percentage gradually lowers down to 66% for G-grade loans.**

Loan purpose

The less risky loan purpose are wedding loans/major purchase purpose loans with a 90% repayment rate. And **the most risky is small businesses funding, with a 73% repayment rate**

Credit line age

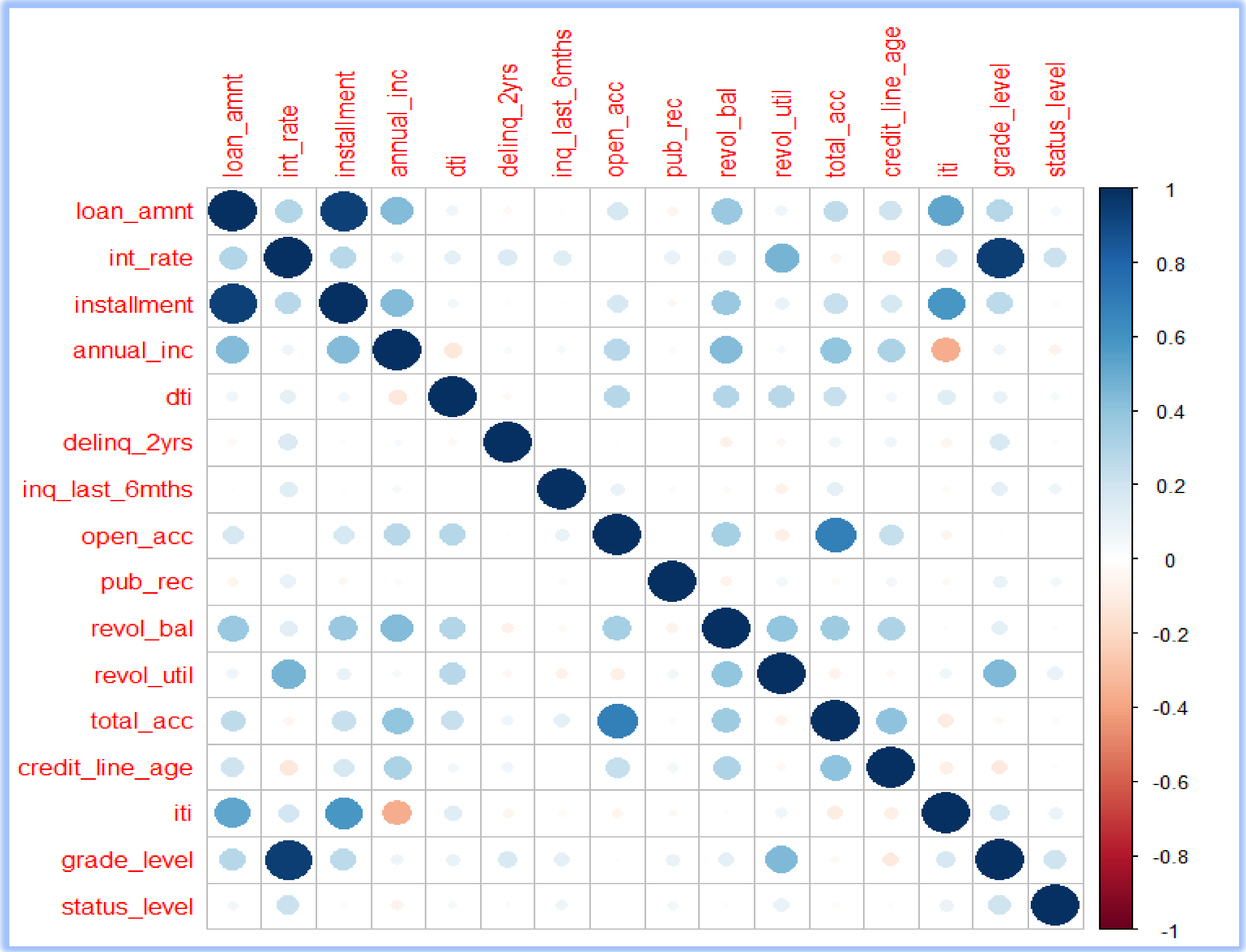
More risk of defaulting with **borrowers of Credit line age <=5**

Term

More Charge offs in 60 month tenures than 36 months tenure.

Multivariate Analysis

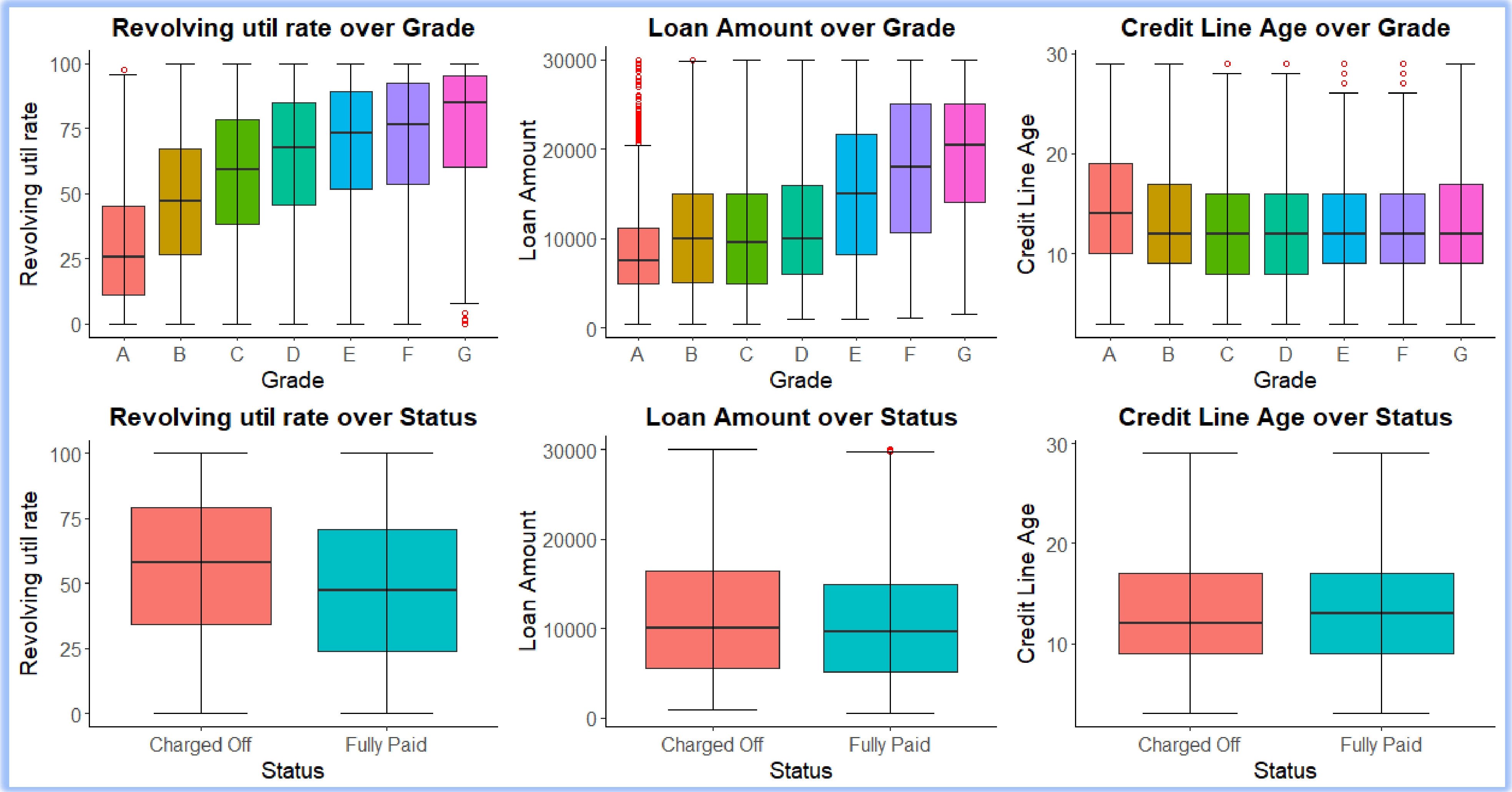
Continuous variables



Categorical variables - **Grades & Status** shows +ve correlation on the below:

1) Revolving utilization rate
2) Loan Amount
3) Credit Line age
4) Annual income
5) Interest rate
6) Installment to income ratio

Multivariate Analysis
Continuous variables Vs Categorical

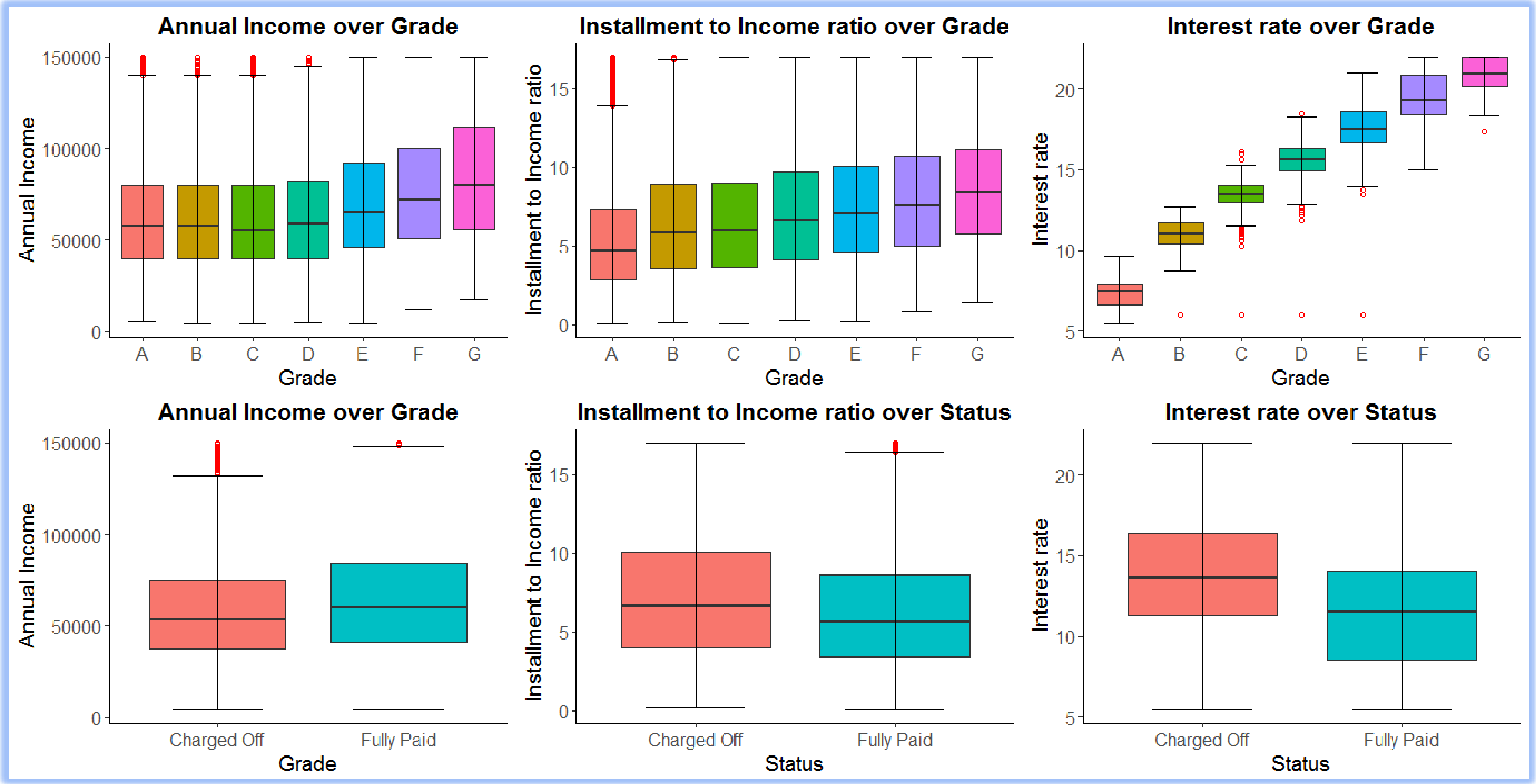


Defaulting rate is **more** for borrowers with **higher revolving utilization rates**

- Grades assigned based on Loan amount, Revolving line utilization rate, dti, earliest_cr_line (credit line age),
- Charged off loan seems to be more for higher revolving util rates
- Loan amounts and credit line age on an average shows some correlation on the defaulting rate

Multivariate Analysis
Continuous variables Vs Categorical contd…

UpGrad



Defaulting rate is more when the
- **Annual income is low**
- **Installment to income ratio is high** and
- **Interest rate is high**

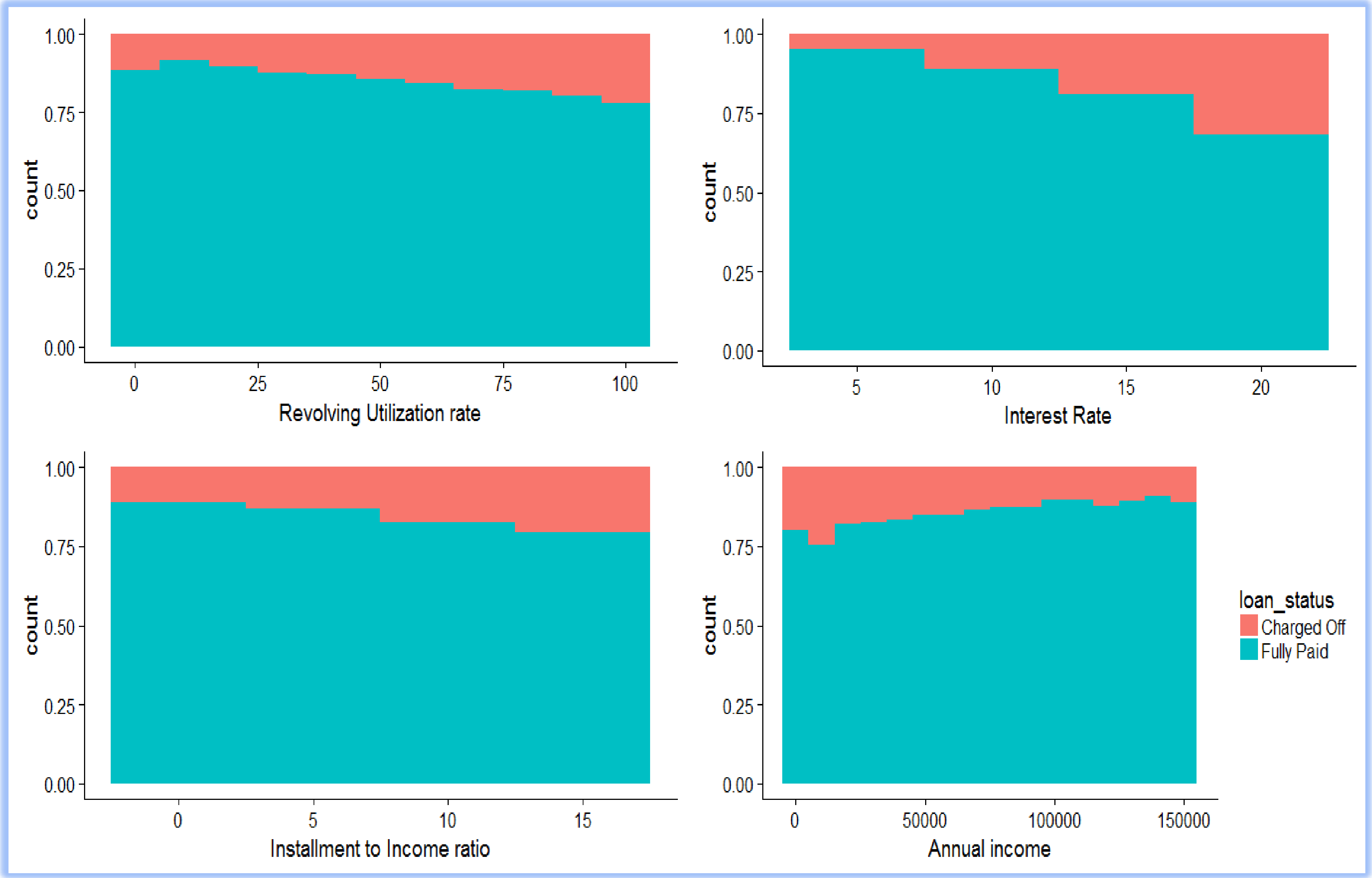
Annual income
- Interestingly annual income is bit higher for Grades E through G
- On an average Borrowers with lower annual income tend to default more.

Installment to Income ratio
- Lowest Installment to income ratio in Grade A and highest iti in Grade G
- Charged off loan has higher installment to income ratio on an average than Fully paid loans

Interest rates
- Grade A loans has lowest interest rate and Grade E with highest interest rate.
- Charged off loan has higher interest rates on an average than Fully paid loans

Segmented Analysis

Continuous variables using bin buckets

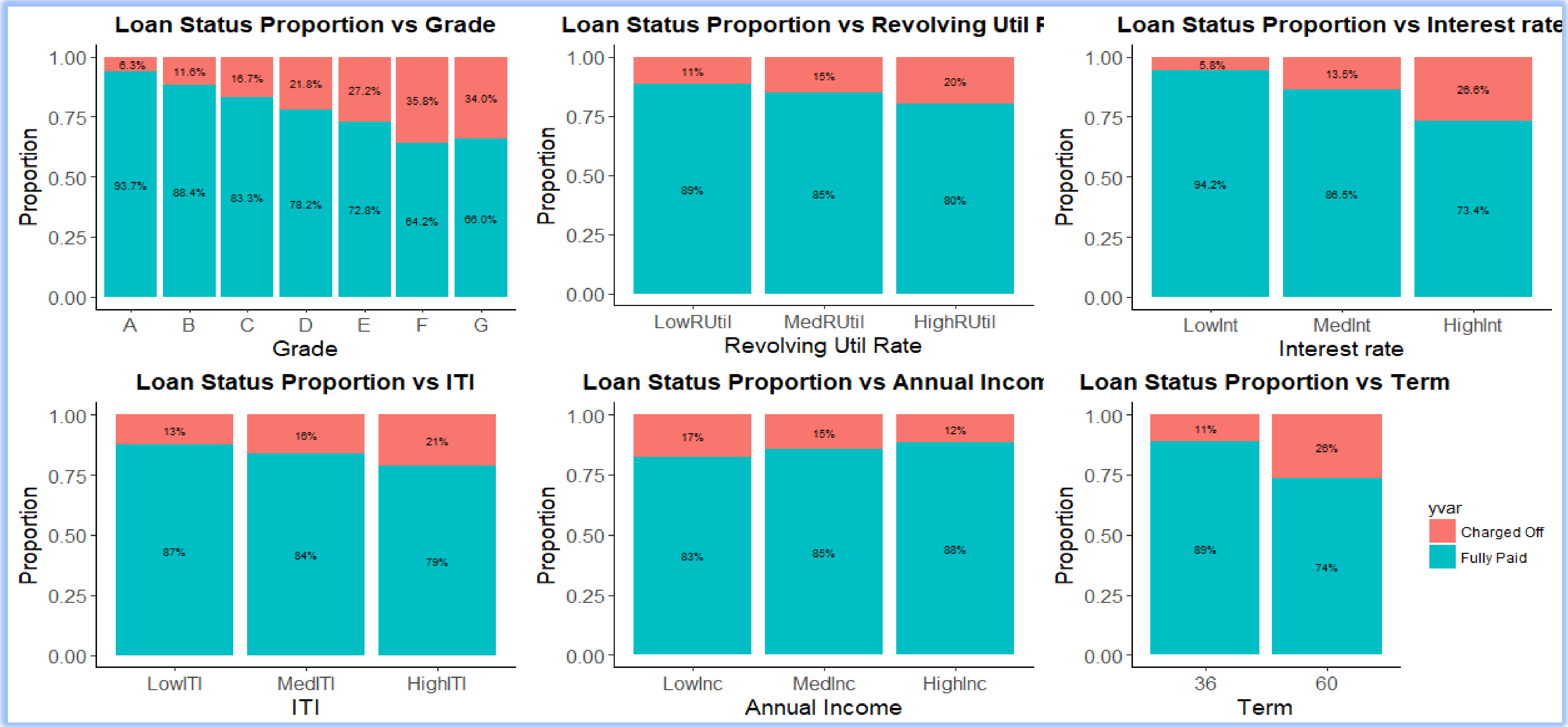


Defaulting rate depends on the following variables. Strong dependency in the below order :

1. Interest rate
2. Installment to income ratio
3. Revolving utilization rate
4. Annual income

Segmented Analysis

Categorical & Continuous (Segmented) variables



The above plots are for Debt Consolidation. Similar are the plots for other loan purposes. Though the impact of the above variables on the defaulting rate may vary based on the purpose of the loan, those are the common most strongly influencing variables

Segmented Analysis

Summary

↓	Grades	E F G
↑	Higher interest rates	>15%
↑	Higher loan tenure	60 Months
↑	Higher installment to income ratio	>15%
↓	Lower Income (< 50 K)	<50K

Defaulting rate increases to 80%
(~4 out of 5 borrowers default)
for Small Business, Renewable energy, Educational loans:

Defaulting rate increases to 50%
(~1 out of 2 borrowers default)
for Debt Consolidation, Credit Card types of loans

Conclusion

The study results show that below are the strong variables impacting defaulting rate

Grade :

- There is a clear relationship between the grade assigned by Lending Club and the probability of default.
- of **94.4%** of **A-grade loans** were reimbursed. This percentage gradually decreases to **61.8%** for **G-grade loans**.

Loan Purpose:

Loans for Debt consolidation and Credit card constitute 60% of total loans. Loan purpose is also a factor explaining default: **Wedding** is the less risky loan purpose and **Small business** is the riskiest.

Loan Tenure:

Defaulting rate depends on Loan tenure. Higher the loan tenure **(60 months),** Higher is the defaulting rate.

Interest rate:

- On an average **Charged off loans** paid higher interest rates **(13%)** than the **fully paid loans (11%)**. i.e Higher the interest rate, Higher the probability of default is.
- Over years interest rates for Grade D through E has been seen increasing especially post 2011 there is a steep rise.

Borrower Characteristics: Annual income, Installment to Income ratio, Revolving utilization rate

- Fully paid borrowers have less Installment to income ratio compared to the defaulting ones.
- **Fully paid borrowers** have an average of **$60K annual income** against **defaulting borrowers** with only an average of **$53K annual income**
- Revolving utilization rate is more (average **58.20%**) for **charged off loans** than that of **fully paid loans** (average **47.50%**)