Data Visualization of Lending Club Issued Loans

Introduction:

Lending Club ("LC") is the world's largest peer to peer online platform. It reduces the cost of lending and borrowing for individuals with advanced data analytics. Peer to peer companies' function is matching people who have money with people who want to borrow money. As a leader in the peer to peer lending industry, LC completed its initial public offering in December 2014, which definitely improved public trust in this fast growing company.

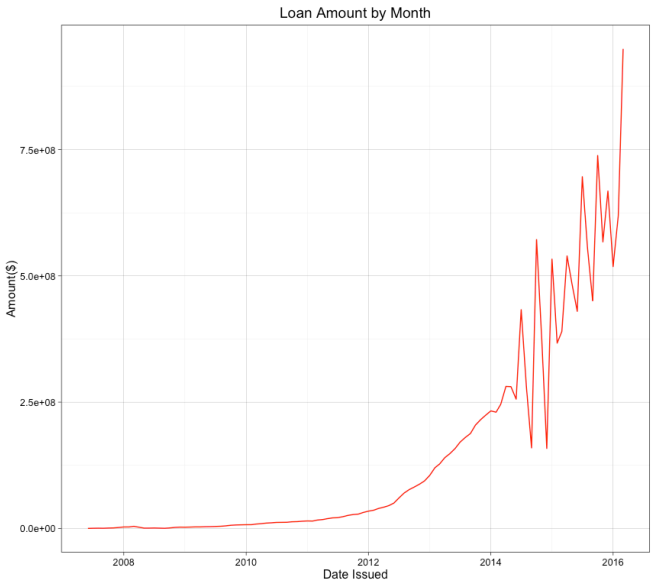
This analysis is based on data published on the website of Lending Club company. (You can find it here.) The dataset contains 1,021,286 observation and 111 variables collected from 2007 to 2016Q1. Forty variables were selected for this analysis. Four trends were investigated: growth of issued loans, both in terms of dollars and volume, geographical distribution of loans, purposes for loans, and interest rate changes over time.

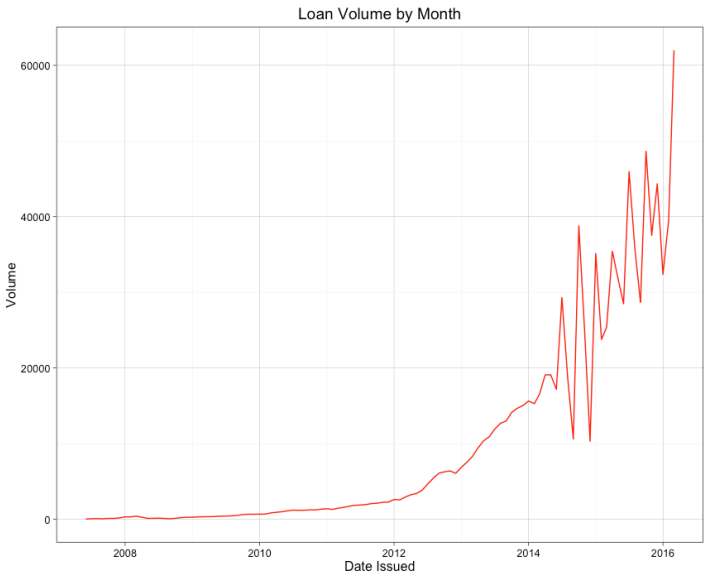
Data:

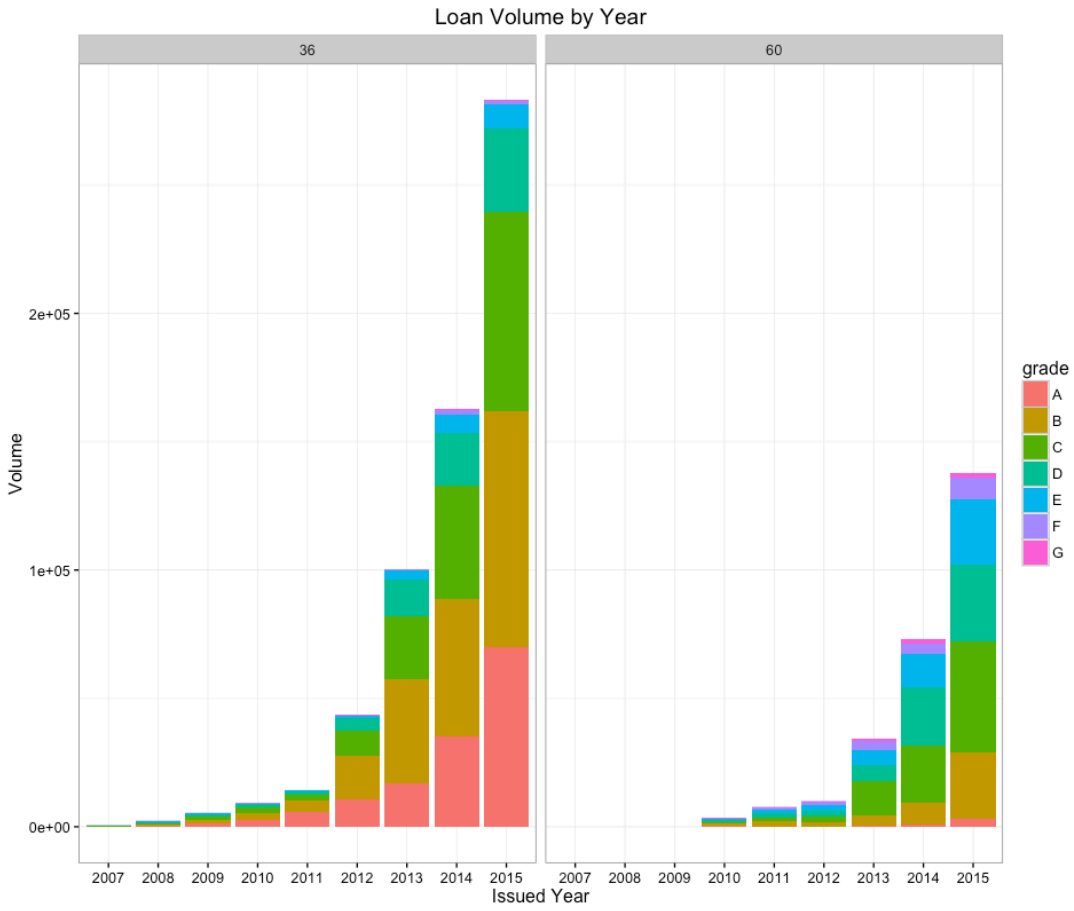
```
library(dplyr)
# Preparartion of valid data
loandf_07_11 <- read.csv('data/LoanStats07-11.csv', stringsAsFactors = FALSE, header = TRU
loandf_12_13 <- read.csv('./data/LoanStats12-13.csv', stringsAsFactors = FALSE, header = TR
loandf_14 <- read.csv('./data/LoanStats14.csv', stringsAsFactors = FALSE, header = TRUE)
loandf_15 <- read.csv('./data/LoanStats15.csv', stringsAsFactors = FALSE, header = TRUE)
loandf_16Q1 <- read.csv('./data/LoanStats16Q1.csv', stringsAsFactors = FALSE, header = TRU
loandf <- rbind(loandf_07_11, loandf_12_13, loandf_14, loandf_15, loandf_16Q1)

# 41 variables
var_name = c('id', 'member_id', 'loan_amnt', 'term', 'int_rate', 'installment', 'grade', 's
             'addr_state', 'dti', 'earliest_cr_line', 'open_acc', 'total_acc', 'total_pymnt
             'total_rev_hi_lim', 'mort_acc',    'mths_since_recent_bc', 'mths_since_recent_

loandf <- loandf %>% select(one_of(var_name))

# Preprocessing -- standardize date and term format
loandf$issue_d <- as.Date(gsub("^", "15-", loandf$issue_d), format="%d-%b-%y")
loandf <- mutate(loandf, term = ifelse(term == ' 36 months', 36, 60))

save(loandf, file = 'loandf.RData')
```

LC_Data Preparation hosted with ❤ by **GitHub** **view raw**

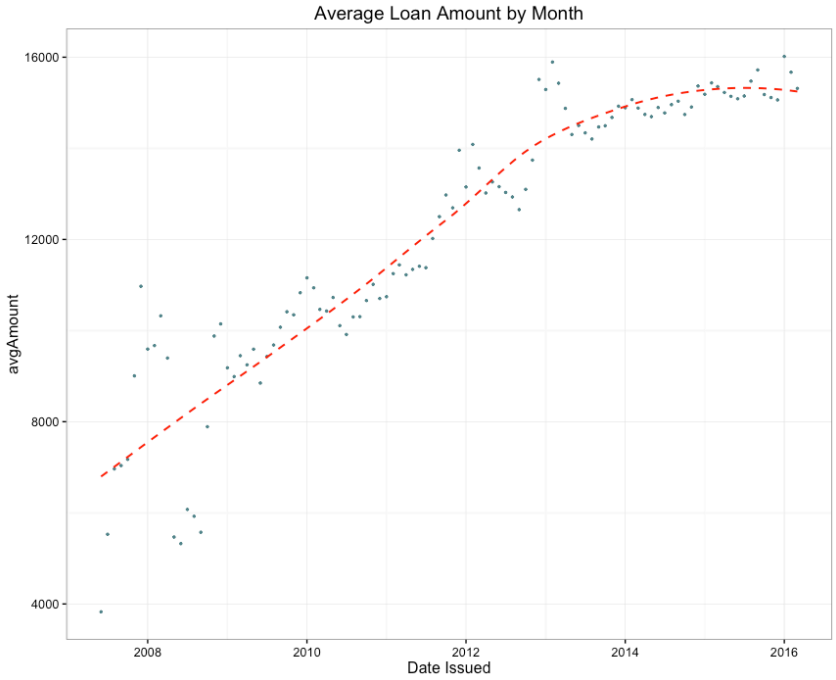
1.Growth of Issued Loans







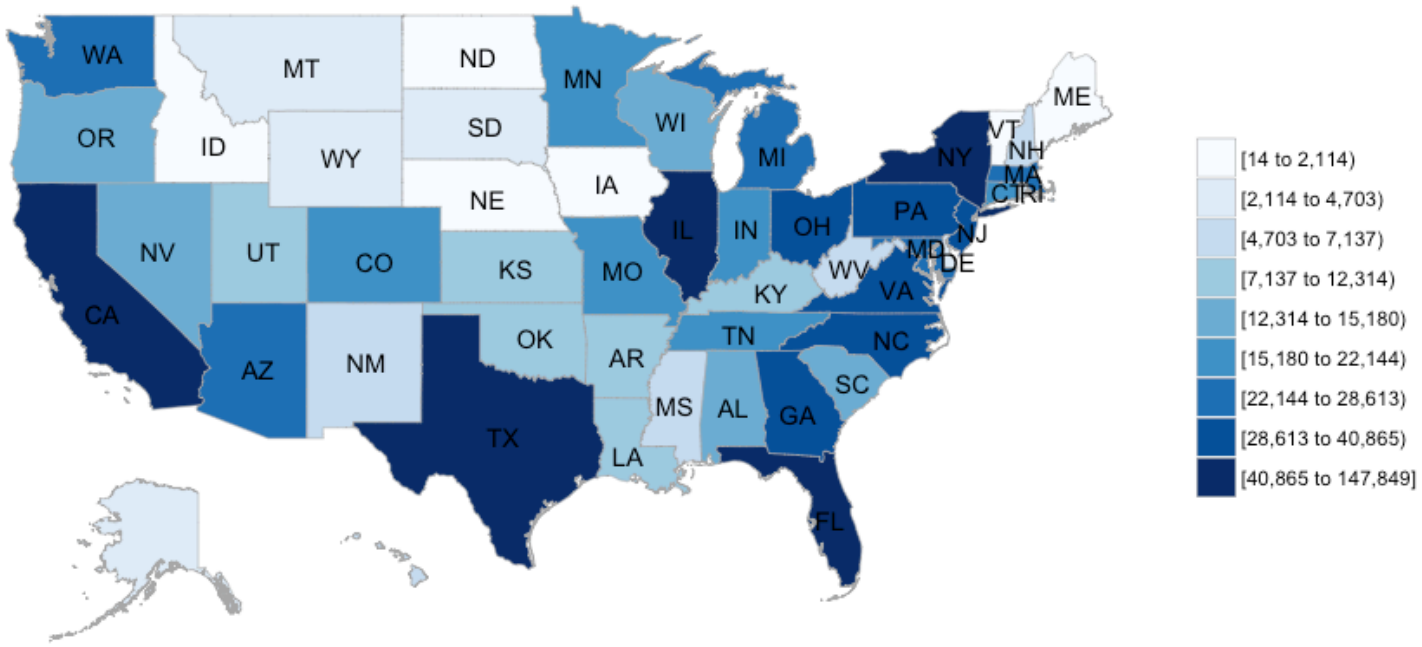
Lending Club got launched during 2007, but its business grew significantly since the start of 2012. From 2014 to 2016Q1, monthly loan amount and volume have been increasing drastically, but the increasing trend is unstable in the short term. Additionally, the increasing paths of the loan amounts and volume are similar. We can still expect a high growth rate in the following year based on past growth and the performance of first quarter in 2016.



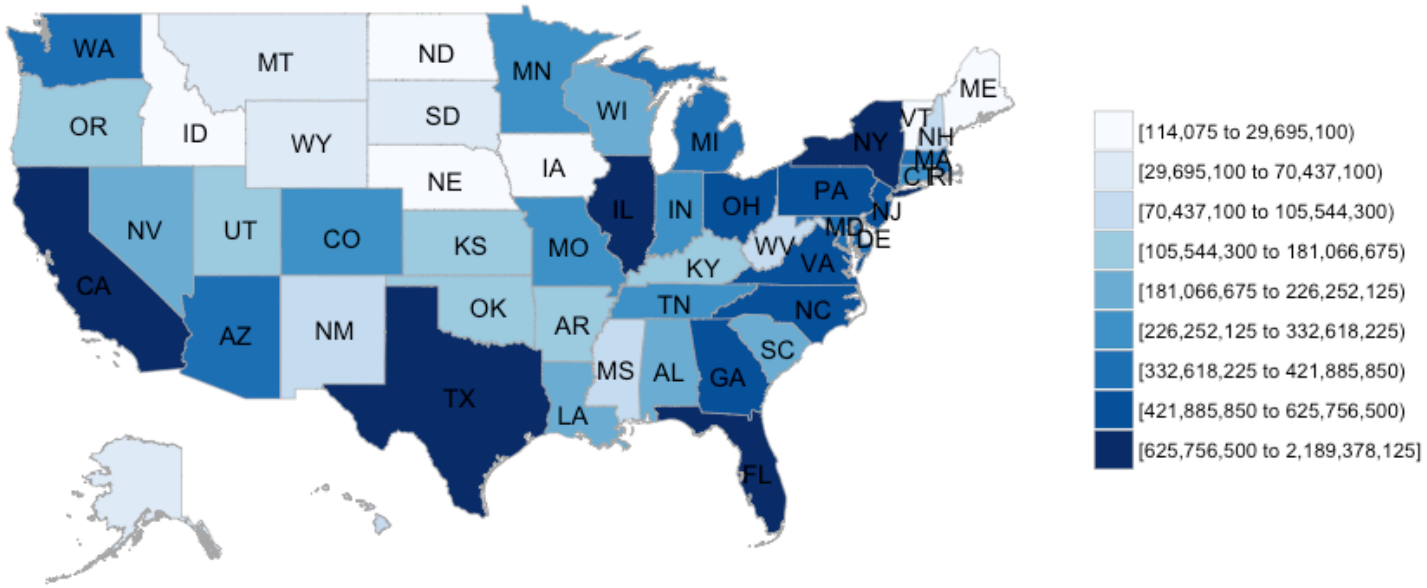
In order to learn more about the growth, we fit a basic trend line to figure out the average changes in loan amounts. Above line chart shows average loan amount increased at a constant rate from 2007 to 2012, but it grew with a descending rate from 2013 to 2014. The average loan amounts in 2015 and 2016Q1 roughly remained unchanged.

2.Geographical Distribution of Issued Loans

Loan Volume by State



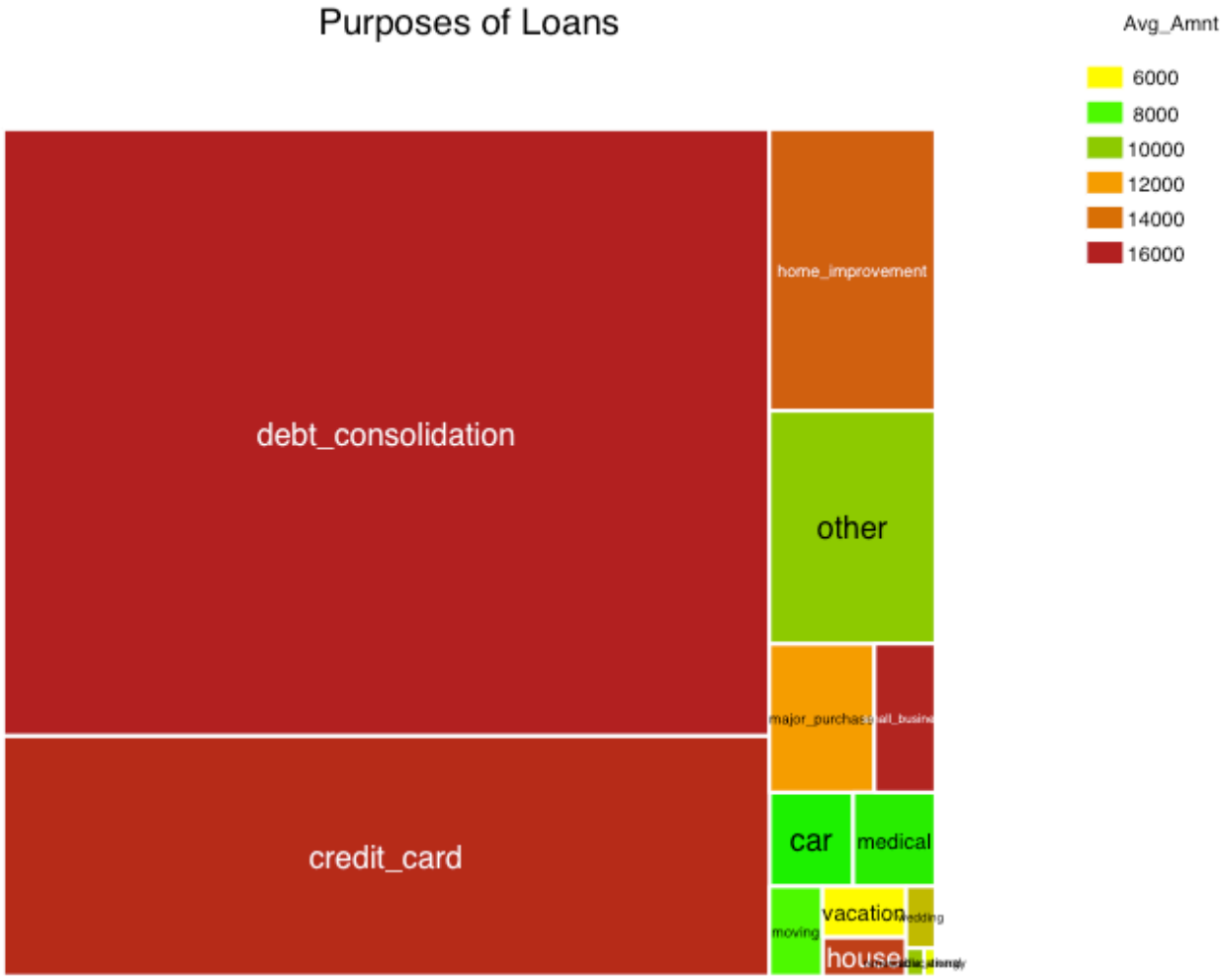
Loan Amount by State



From a geographical perspective, California, Texas, New York, Florida and Illinois have the largest dollar amount and volume of loans.

3.Why and How Much do People Borrow?

Peer to peer lending focuses on individual lending and borrowing, so it's important to understand why consumers decide to borrow money.

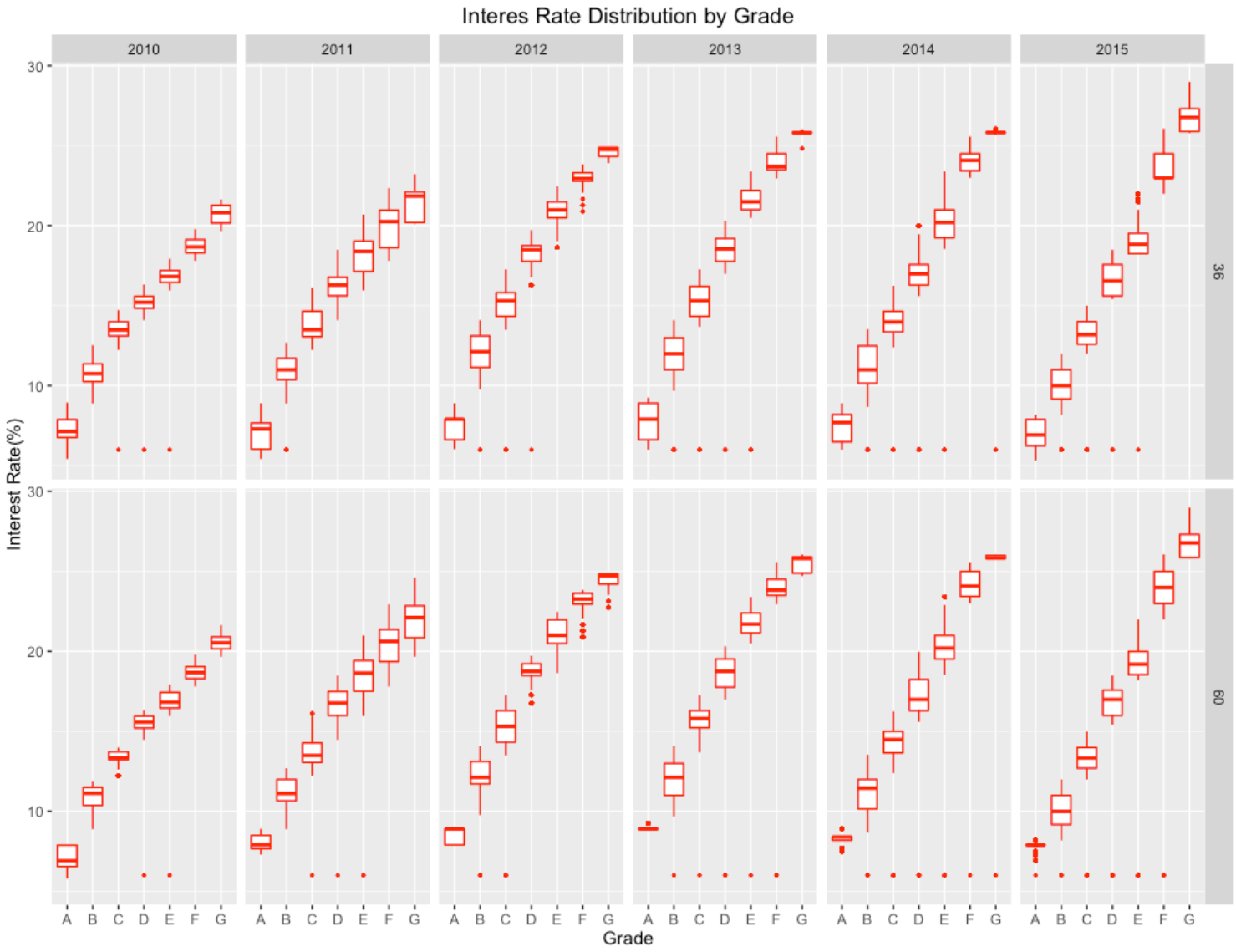


Unsurprisingly, debt consolidation is the most common reason for borrowing. The greatest advantage of peer to peer lending is the low cost, so loans issued by LC usually charge lower interest rates compared with money provided by traditional banks. Most consumers choose to consolidate debt to enjoy lower borrowing costs.

The color in the above tree map stands for different average amounts of loans for different purposes. Debt consolidation, credit card, house and small business are the most popular reasons for borrowing.

4.Return Rate Changes and Time

Two loan terms were explored, 36 months and 60 months. (Note in the plot below, the analysis begins in 2010 because LC did not introduce 60-month loans until 2009.) Interest rates vs. loan terms vs. loan grade are represented in the graph below.



According to the risk theory, longer loans own higher risk as well as higher interest rate than shorter ones. However, we cannot obtain the corresponding conclusion from above graph. Interest rate distributions with different terms have no significant difference.

However, it's really interesting that interest rates of different grades behave in a significant trend through years. From 2010 to 2015, interest rates for loans with grade D, E, F, G obviously increased. We still see there are larger disparities of interest rates among grades, possibly due in

part to LC's credit policy updates. This change is a great symbol that LC has become more and more proficient in the evaluation of loans' risk.

Summary:

In this project, we have a brief and clear introduction to the growth of LC's main business. It's obvious that LC is currently a fast-growing but the volume and amount of loans are erratic. This is surprising given that we expect public companies to be more stable. The drastic fluctuation of LC's stock price also proves the conclusion.

However, LC's business model still brings a great advantage over traditional banks. LC is improving its risk and credit evaluation technology and trying to extend its market from individuals to businesses. From the latest growth data, we still believe LC issued loans will continue to grow rapidly.

Next steps:

Geographical distribution of issued loans may be influenced by several factors such as state financial standing, state culture, LC's advertising strategies, etc. We need more data for the deeper exploration, so combining current dataset with other datasets is a good choice.

FICO is widely accepted by traditional banks as a credit index, but LC claims its algorithm considers more than 2000 variables in credit evolution and risk management. There may be some interesting correlation between grade and factors such as homeownership, number of inquires, debt to income ratio, etc. We can explore this more with more advanced data analytics.