Title: **Interest Rate Charged by LendingClub is Associated with the Loan Amount Requested, the Length of the Loan and the Amount Invested.**

Introduction:

Lending Club provides loans to individuals with investor members providing the funds for the loans. The company determines the interest rate of a loan by weighing credit risk and market conditions in addition to its base rate. Of all the variables considered by Lending Club, three are more strongly associated with the final interest rate, after an individual's FICO score is considered. The Loan Amount, Amount Funded by Investors and Length of the Loan seem to be positively correlated with the interest rate, where higher values of these variables correlate with higher interest rates and lower values correlating with lower interest rates.

Methods

Data Collection

For the data analysis, I used 2500 rows of lending data that was provided by Lending Club. The data was downloaded on 11/10/2013 from Amazon's S3 servers (via a link in the Assignment 1 Instruction page).

Exploratory Analysis

Exploratory analysis was performed by examining tables and plots of the observed data. I identified transformations to perform on the raw data on the basis of plots. Exploratory analysis was used to identify missing values and cleanup the data; identify the relationships of the different variables to the Interest Rate; and to see how each of the variables is distributed.

Statistical Modeling

In identifying associations between each of the variables in Lending Club's data and the final interest rate, I chose to apply the correlation function in R to Interest Rate and each of the rest of the variables in the data. I created four subsets of the data by dividing them into four quartiles, according to the midpoint of the FICO score range. That is, I added a new column into my data frame which is the midpoint of the FICO score range and created subsets using quartiles of the new column.

Results

To account for the FICO scores, I found it useful to create four subsets of the loans data. After adding a new variable to the loans data, FICO.MidPoint, the midpoint of the FICO score range, I calculated the quartiles in the new variable. I then created four subsets of the data, with the first subset containing rows with the FICO.MidPoint in the first quartile, and a second one with FICO

Mid Point in the second quartile, and the same for the third and fourth subsets. This has enabled me to measure the relationship of Interest Rate with the other variables for those rows whose FICO scores are similar.

Once the subsets are created, I ran the correlation function with the default parameters against the Interest Rate and Amount Requested for each subset of the loans data. The following correlations resulted,

Loan Subset 1: 0.55
Loan Subset 2: 0.64
Loan Subset 3: 0.62
Loan Subset 4: 0.48

The correlation resulted in positive values that are significantly above 0 and close to 1, which strongly indicate the positive association between Loan Amount and Interest Rate.

I also applied the correlation function for all other numeric variables in the data and found that the Amount Funded by Investors has a strong positive association with the Interest Rate. The correlation between Amount Funded and Interest Rate is as follows,

Loan Subset 1: 0.58
Loan Subset 2: 0.63
Loan Subset 3: 0.60
Loan Subset 4: 0.43

Conclusion

From this data analysis, it is apparent that the Interest Rate is strongly associated with the length of the loan, the Amount Requested and the Amount Funded by Investors. A higher Amount Requested, a higher Amount Funded or a longer Loan Length is likely to indicate a higher Interest Rate.

The length of the loan seems to be positively associated with the interest rate as can be seen in Fig. 1. The figure shows that the median interest rate for a 6 month loan is about 18% while it is about 13% for all loans in the data.

References

R-Bloggers. “Measuring associations between non-numeric variables” URL: http://www.r-bloggers.com/measuring-associations-between-non-numeric-variables/. Accessed: 11/16/2013.

Coghlan, Avril. “Using R for Multivariate Analysis”. URL: http://little-book-of-r-for-multivariate-analysis.readthedocs.org/en/latest/src/multivariateanalysis.html. Accessed: 11/16/2013.

UCLA: Statistical Consulting Group. “Scatter Plots”. URL: http://www.ats.ucla.edu/STAT/r/gbe/scatter.htm. Accessed: 11/17/2013.

Lending Club. “Interest Rates and How We Set Them”. URL: https://www.lendingclub.com/public/how-we-set-interest-rates.action. Accessed: 11/17/2013.