# Introduction

We know that FICO score plays an important role in determining the interest rate of a loan, but what other characteristics of a borrower might affect the interest rate? Understanding the other factors that can effect a loan’s interest rate can help set a borrower’s expectations before they apply for a loan.

# Methods

## Data Collection

For this analysis we downloaded a dataset of 2,500 peer to peer loans issued through the Lending Club [1]. The data was downloaded from the Coursera Data Analysis course website [2] using the R Language [3].

## Exploratory Analysis

Exploratory analysis was performed by exploring tables and plots of the data. Transforms where identified and performed on the data based on the exploratory results and knowledge of the data. Terms were identified to be used in a regression model relating interest rate to features beyond only a FICO score. The additional features identified where debt to income ratio and loan length.

## Statistical Modelling

A standard multivariate linear regression model [4] was used. Model selection was performed on the basis of exploratory analysis and prior knowledge of the loan application process. Coefficients where estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [5].

## Reproducibility

To reproduce the results presented in this manuscript please request the author for a copy of the cached data and R code.

# Results

The loan data used in this analysis contains information about each application for a loan. Properties used in the analysis included the interest rate of the loan (Interest.Rate), FICO range of the borrower (FICO.range), debt to income ratio of the borrower (Debt.To.Income.Ratio) and the length of the loan (Loan.Length). There were no missing values in the data for these properties.

The goal of the analysis was to see what other features about a loan could affect the interest rate. First we fit a regression model relating the interest rate to FICO to set a baseline for improvement, and the fit additional features of the loan to reduce the error rate in the estimate. The final regression model was:

Interest.Rate = bo + b1(Debt.To.Income.Ratio) + f(FICO.Range) + g(Loan.Length) + e

Where b0 is an intercept term, b1 represents the change in interest rate associated with a change in the debt to income ratio of the borrower. The terms f(FICO.Range) and g(Loan.Length) represent factor models with 38 levels and 2 levels respectively. The error term e represents all source of unmeasured and unmodeled random variation in the interest rate.

# Conclusion

There is a significant positive association between FICO score and the interest rate one receives on a loan, but is lacking. By including the loan length and debt to income ratio, one can better estimate the rate a borrower will receive on a loan. Including these variables improved the regression model’s fit and did not remove the positive relationship between the variables.

# References

1. Lending Club URL: <https://www.lendingclub.com/home.action>
2. Coursera Data Analysis Course URL: <https://class.coursera.org/dataanalysis-001/class/index>
3. R Core Team (2012). ”R: A language and environment for statistical computing.” URL: [http://www.R-project.org](http://www.r-project.org)
4. Seber, George AF, and Alan J. Lee. *Linear regression analysis*. Vol. 936. Wiley, 2012.
5. Ferguson, Thomas S. *A Course in Large Sample Theory: Texts in Statistical Science*. Vol. 38. Chapman & Hall/CRC, 1996.