Title: Loan Interest Rate increases with FICO score and loan length

Like any lending institution, Lending club has to set the interest rates on their loans in a way that allows them to be compensated for the general level of (default free) interest rates and the risk of default of the borrower while at the same time being able to compete with other credit providers such as banks.
One established measure for the credit riskiness of a borrower is the FICO score (see [2]). The higher the FICO score the less risky the borrower is deemed to be. This is why typically loans given to borrowers with higher FICO scores have lower interest rates.
The details of the computation of FICO scores are proprietary but it is known that they take into account among others (see [2]): Payment history, credit utilization, length of credit history, types of credit used and recent searches for credit.
It is well known that yield curves (that is interest rates as a function of maturity of a debt instrument) are not constant and in many situations upward sloping with increasing time to maturity (see [3]). For credit risky borrowers this can be often explained by the increase in default risk for longer term loans. This effect can be observed in the Lending club data: The 60 months loans have a higher interest rate than the 36 months ones after taking into account differences in FICO score.

Methods:
Data Collection
The data was downloaded from https://spark-public.s3.amazonaws.com/dataanalysis/loansData.csv on 2/11/2013.
As per the instructions of this assignment [1] "The data above consist of a sample of 2,500 peer-to-peer loans issued through the Lending Club (https://www.lendingclub.com/home.action). The interest rate of these loans is determined by the Lending Club on the basis of characteristics of the person asking for the loan such as their employment history, credit history, and creditworthiness scores."

Exploratory Analysis
Exploratory analysis was performed to clean the data. FICO ranges were replaced by the midpoint of the range. 2 Loans with missing values were removed. We also removed 3 loans with monthly income attribute bigger than 30K USD that were deemed possibly erroneous outliers (neither of the removals has an effect on the conclusions described herein). The remaining dataset consisted of 2495 loans. Various scatter plots between Interest rates and the other data attributes helped to show that there might be a relationship between interest rates and loan length after accounting for FICO scores while most of the other attributes seem to be either well accounted for by the FICO scores or not relevant to the interest rates. In the data sample the loan length of all loans is either 36 months or 60 months.

Statistical Modeling
To relate loan interest rates to FICO scores and loan length a standard multivariate linear regression model [4] was examined. Model selection was inspired by our exploratory analysis and the general knowledge about credit curves and interest rates.
Coefficients were estimated with ordinary least squares and standard errors were calculated using standard asymptotic approximations [5].

Results:

Our linear regression model for the interest rate of a loan in percent is
LoanRate=b_0+b_1*FICO+b2*1_(Loan.Length=60 months), where 1_(Loan.Length=60 months)

is the indicator variable that is 1 if Loan.Length=60 months and 0 otherwise. The estimates are b_0=72.59, b_1=-0.085, b_2=4.37 and all of these estimates are highly significant (P<2e-16). The interpretation is that on average a loan with 60 months maturity has an interest rate that is 4.37% higher than a loan with 36 months maturity. As expected, the loan rate gets lower the better the level of credit worthiness as measured by the FICO score. In particular every point in FICO score reduces the interest rate on a loan by on average 0.085%.

Conclusions:

Our work confirms our hypothesis that the level of loan rates is higher for longer maturity loans after accounting for FICO scores.

While this result is interesting and in general consistent with the theory about interest rates, it is clear that this relationship could change over time as the general level of interest rate changes. More work could be done for a dynamic model of interest rates over time once more data becomes available.

References

1. Coursera Class Data Analysis Data Assignement Page URL: https://class.coursera.org/dataanalysis-001/human_grading/view/#courses/294/assessments/4/submissions
2. Wikipedia page URL:http://en.wikipedia.org/wiki/FICO_score#FICO_score
3. Wikipedia page http://en.wikipedia.org/wiki/Yield_curve
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.