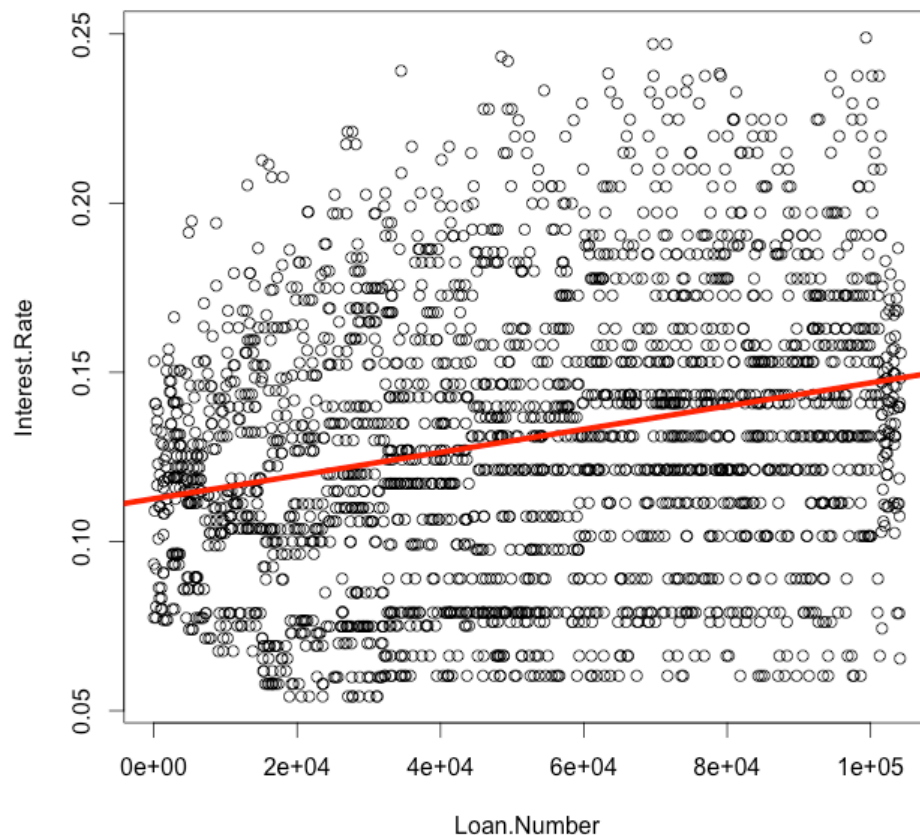


Example data-analysis plot for IST380, assignment 4, pr3

Overview: The line in red is a linear fit of the LendingClub interest rate (vertical axis) against the loan number (horizontal axis).



Summary: Here is the summary of the fit:

```
> summary( ir_vs_ln )
```

Call:

```
lm(formula = Interest.Rate ~ Loan.Number)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.086149	-0.029302	-0.000661	0.026710	0.114627

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.127e-01	1.630e-03	69.10	<2e-16 ***
Loan.Number	3.427e-07	2.731e-08	12.54	<2e-16 ***

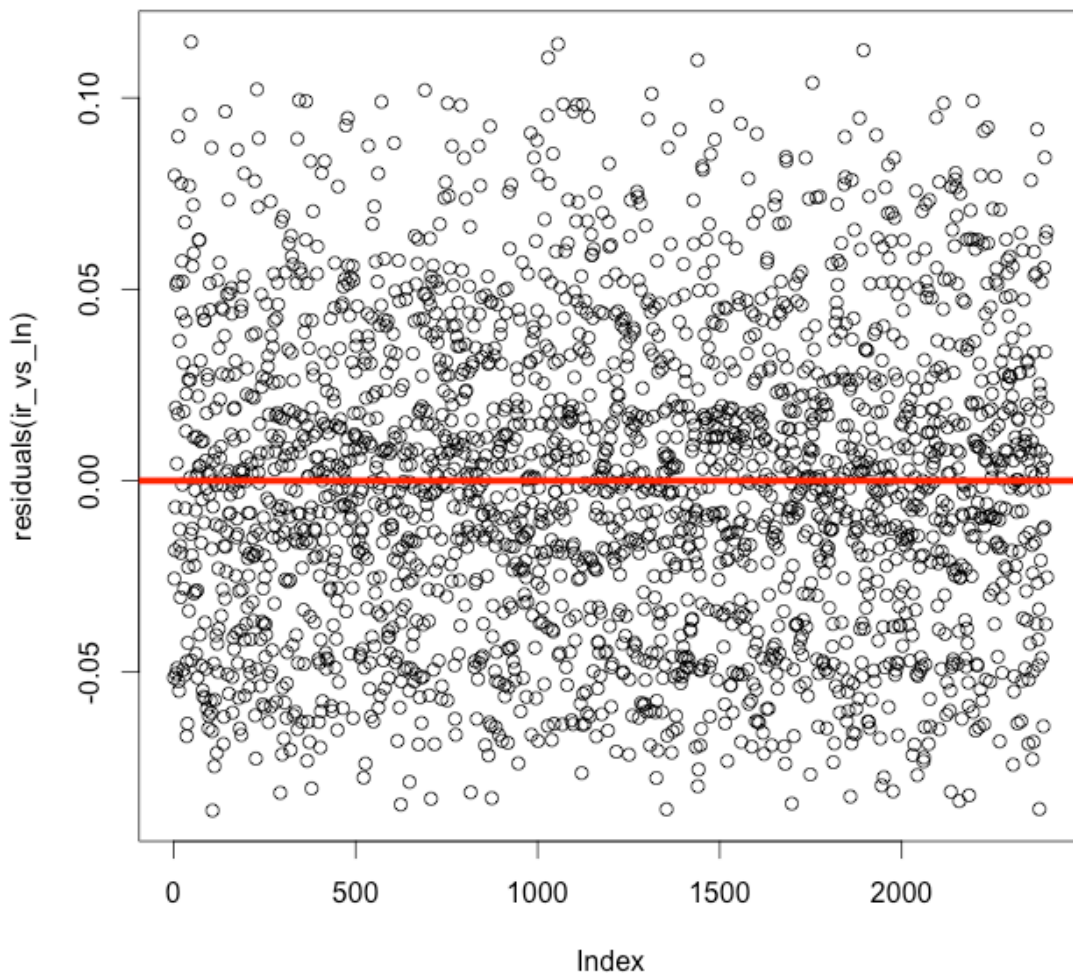
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04019 on 2398 degrees of freedom

Multiple R-squared: 0.06158, Adjusted R-squared: 0.06119

F-statistic: 157.4 on 1 and 2398 DF, p-value: < 2.2e-16

Conclusion: Indeed, there is a small and statistically very significant relationship between loan number and interest rate. In particular, the average interest rate increases very slightly (3.4×10^{-7}) with loan number. A possible explanation is that loan number correlates with time and early loans were discounted to build market share... ? Regardless, there is enough residual variance to make it clear that there are more factors involved than simply loan number (not surprisingly!) Here is a plot of the residuals:



Code: Below is the R code that generated these two plots and a function for predicting new interest rates based on this relationship:

```

# set the working directory - I like to work from the Desktop...
setwd("/Users/zdodds/Desktop")

# read in the data...
ld <- read.csv("loansData.csv")

# attach the dataframe, ld
# this way, you can use the columns without the $ symbol...
# (also, predict does not work with the $ symbol!)
attach(ld)

# plot the interest rate vs. the loan number
plot( Loan.Number, Interest.Rate )

# build a linear model named ir_vs_ln for the rate vs. the loan#
ir_vs_ln <- lm( Interest.Rate ~ Loan.Number )

# plot the best-fit line in red with width 5
abline( ir_vs_ln, col="red", lwd=5 )

# plot the residuals and the zero-intercept, zero-slope line
# you will need to "scroll" through the plots to see the
# previous one...
plot( residuals( ir_vs_ln ) )
abline( a=0, b=0, col="red", lwd=5 )

# here is a prediction function based on the loan number
# it's not very good!
# and, it presumes that ir_vs_ln is the model above...
# so, it's not the most general-purpose function!
pred_from_loan_number <- function( new_num )
{
  prediction <- predict( ir_vs_ln,
                        newdata=data.frame(Loan.Number=new_num) )
  # prediction is an indexed list of results
  # we will return only one result:
  result <- prediction[[1]]
  return(result)
}

# to run this predictor, call pred_from_loan_number( new_num )
# Here is an example:
p <- pred_from_loan_number(42000)

cat("\n\n") # some blank lines...
cat("The model predicts that loan #42000 is at an interest rate of ",p)

# for me, the above line prints the following:
# The model predicts that loan #42000 is at an interest rate of
0.1270426

```