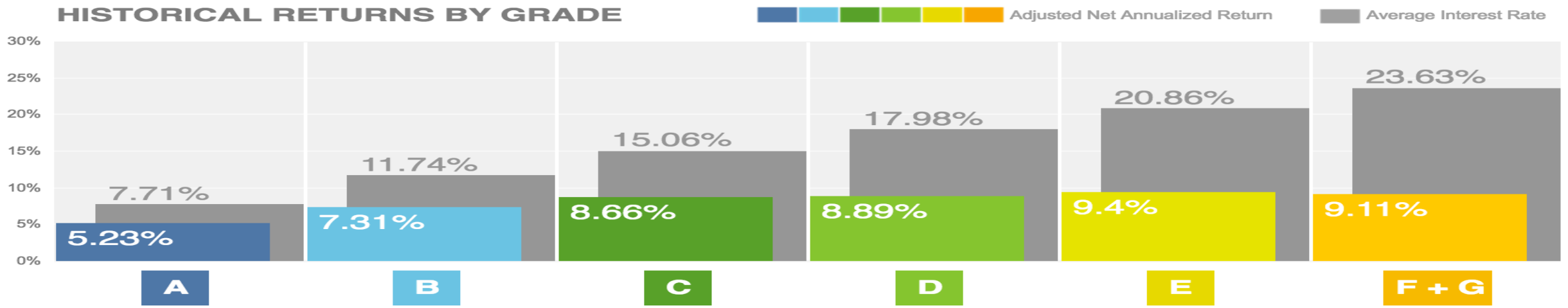
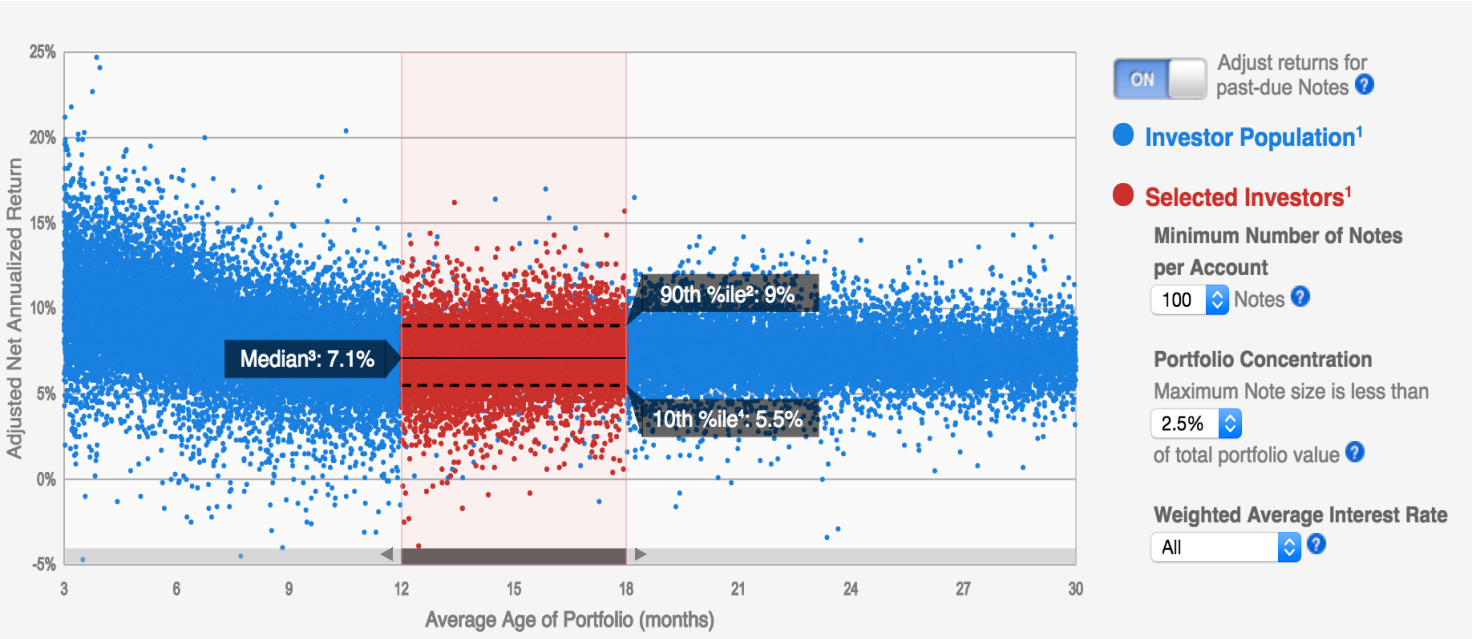
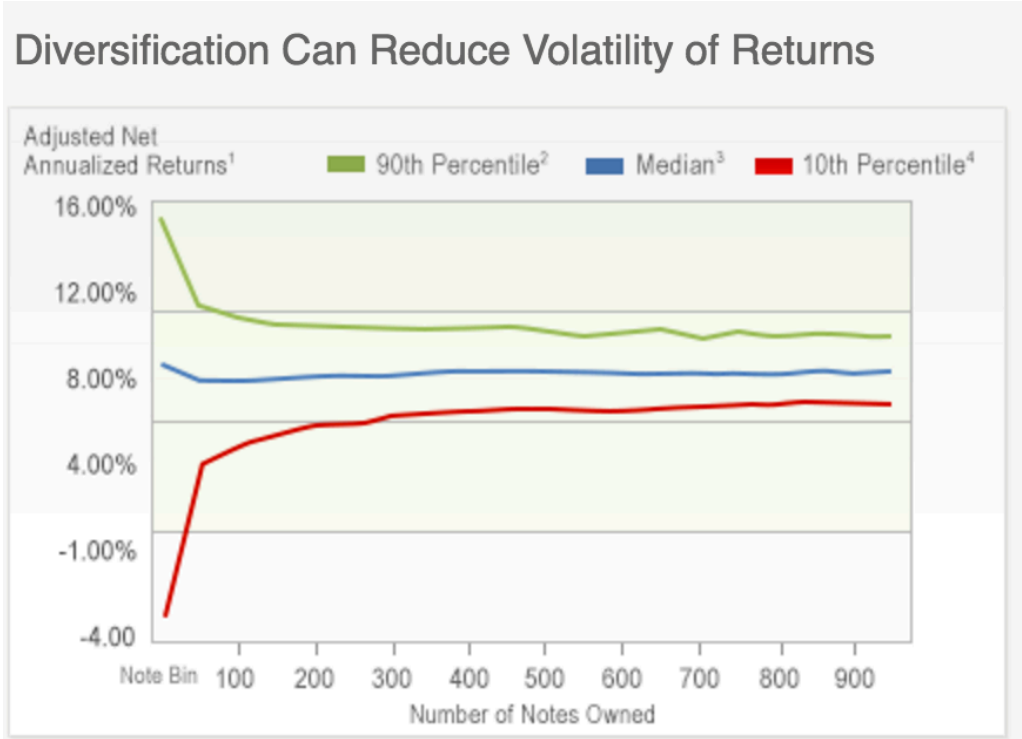
Invest with P2P Lending

-- Yukai Zhong

The platform







How to invest

- Ideal Investment
  - Lend to borrowers with high interest rate
  - Target borrowers with less default possibility

- Automated investment with customized filters

Add Filters

☐ Interest Rate	☐ Loan Purpose	☐ Max Loan Amount Up to
☐ Review Status	☐ Verified Income	☐ Funding Progress
☐ Listing Expires in	☐ Exclude Relisted Loans	☐ CREDIT Score
☐ Max Debt-to-Income Ratio	☐ Earliest CREDIT line	☐ Open Credit Lines
☐ Total CREDIT Lines	☐ Revolving credit balance	☐ Revolving balance utilization
☐ Inquiries in the last 6 months	☐ Months Since Last Delinquency	☐ Months Since Last Record
☐ Home Ownership	☐ Min length of Employment	☐ Location State
☐ Keyword	☐ Public Records	☐ Delinquencies (Last 2 yrs)
☐ Exclude Loans already invested in	☐ Loan Term	☐ Initial Listing Status
☐ Monthly Income	☐ Loan ID	☐ Recent Listings
☐ Collections Excluding Medical	☐ Major Derogatory	☐ Joint Applications
☐ Total collection amount ever	☐ Accounts now delinquent	☐ Total current balance

Find right filters

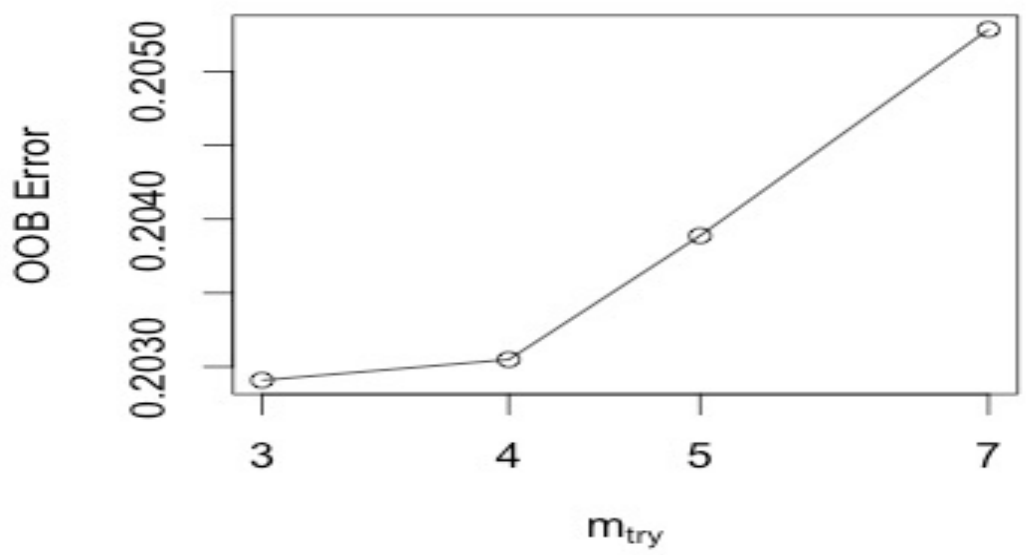
- Reasonable Expectations
  - With an ideal credit policy, default should be purely random
  - Even 1 -2 filters are sufficient for investment

- Data: 2007-2011 Issued Loan Data
  - https://www.lendingclub.com/info/download-data.action
  - *Look at loans rated as D & E*

- Define y
  - Loan_Status (Late (16-30 days)", "Late (31-120 days)", "Default", "Charged Off”)
- Define X
  - Remove variables with over 20% missing values
  - Remove categorical variables with only one factor
  - Remove missing values
  - 19 variables with 8939 records

Random Forest Model

- Split Data 80/20 as test and training
- Tune Max_Features
  - bestmtry <- tuneRF(trainingData[-9],trainingData$loan_status, ntreeTry=5000, stepFactor=1.5,improve=0.001, trace=TRUE, plot=TRUE, dobest=FALSE)



Random Forest

- Run the Model
  - rf.Master = randomForest(loan_status ~.,data = trainingData,mtry=3,importance = TRUE,na.action = na.omit,n.trees=20000,type="classification")
- Plot ROC

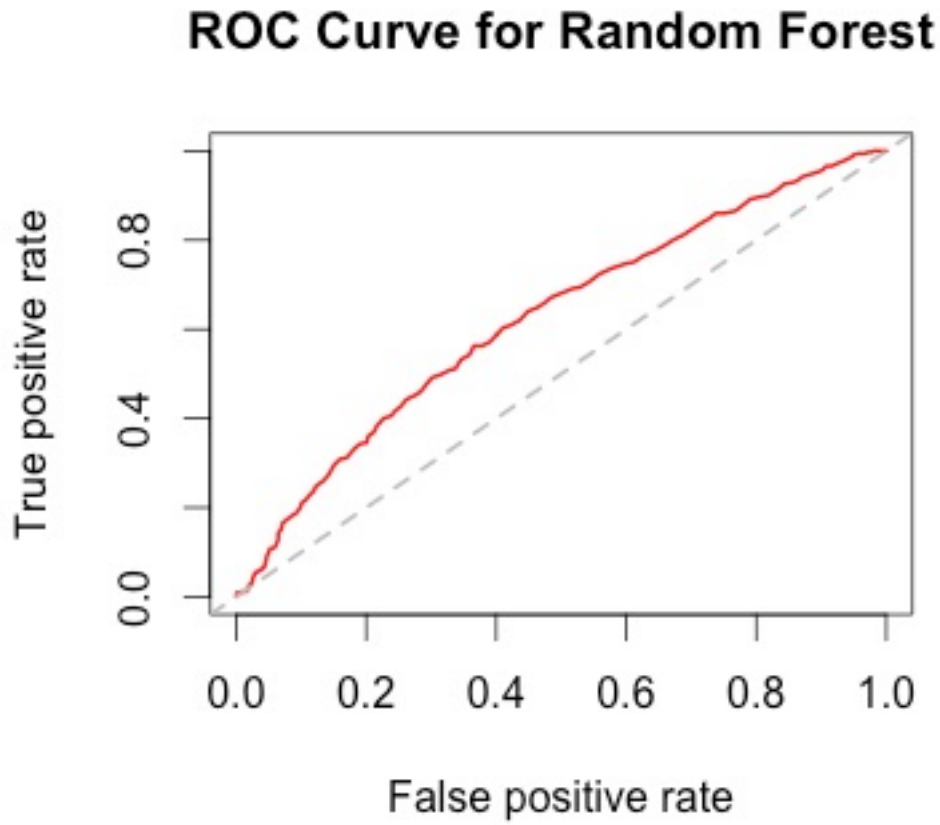
default.rf.pr = predict(rf.Master,type="prob",newdata=testingData)[,2]

default.rf.pred = prediction(default.rf.pr, testingData$loan_status)

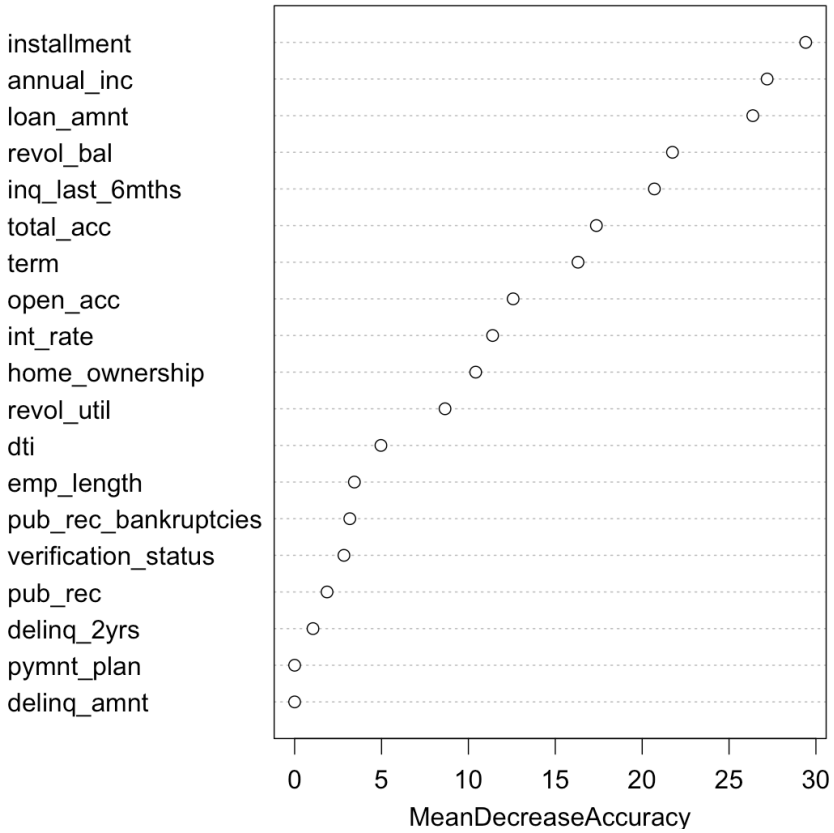
default.rf.perf = performance(default.rf.pred,"tpr","fpr")

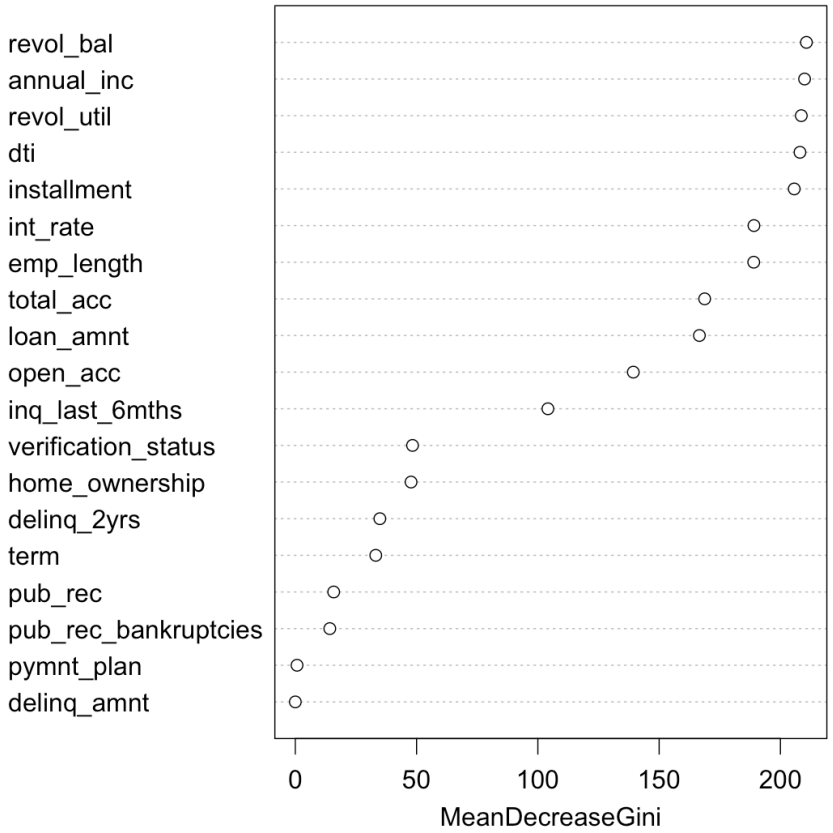
plot(default.rf.perf,main="ROC Curve for Random Forest",col=2,lwd=2)

abline(a=0,b=1,lwd=2,lty=2,col="gray")



Important Variables





- Deploy filters:
  - D & E gives 20% default rate
  - Income >=50000 & loan_amount<=5000 returns with 16% default rate
  - Only 22% of the notes meet criteria. Slower investment