Incarceration Rate by Zip Code

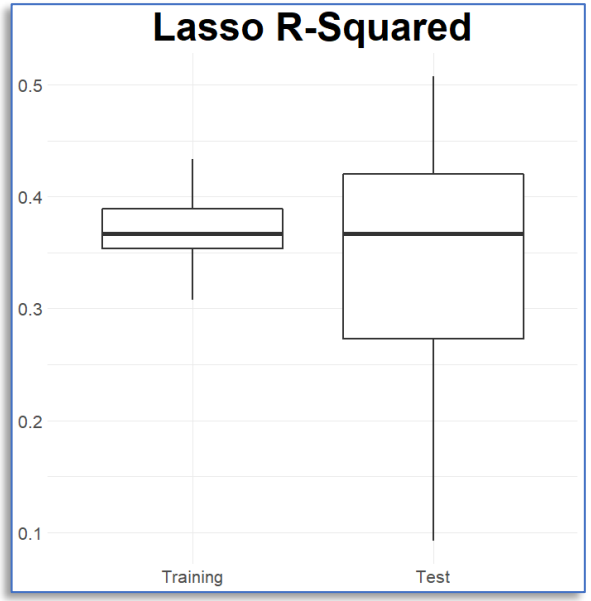
By Matthew Bellwoar

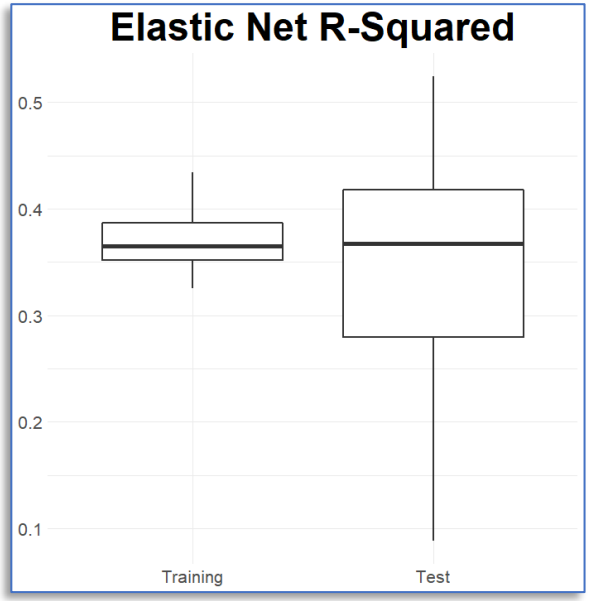
https://github.com/mbellwoar/9890PROJECT

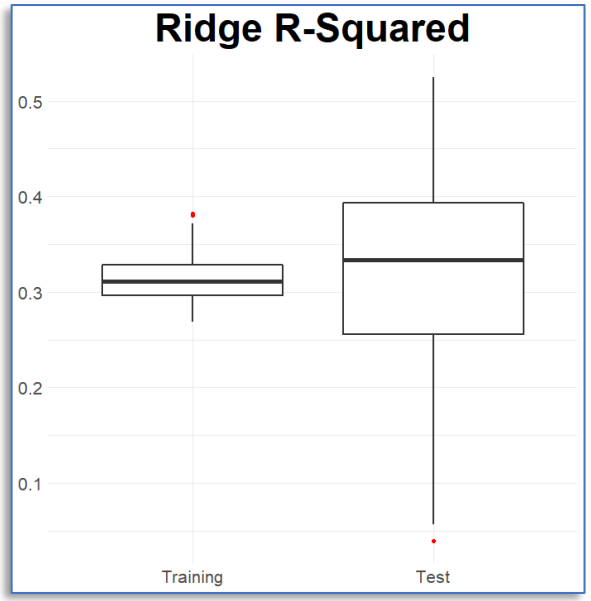
Data – Incarceration Rate by zip code

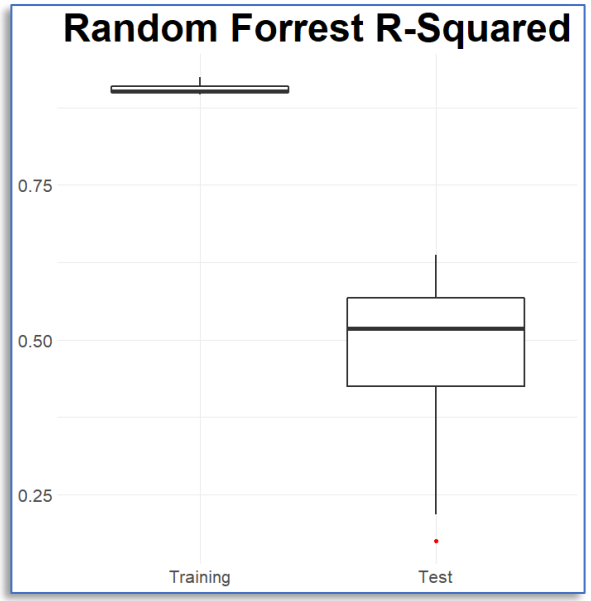
n = 1,894
p = 68

- In 2010, Maryland and New York included incarcerated populations in census data
- Rate - the number of imprisoned people divided by the total population and then multiplied by 100,000
- The census also provided the population data per zip code
- The IRS provides tax information for the same
- Square root transformation to offset skewing

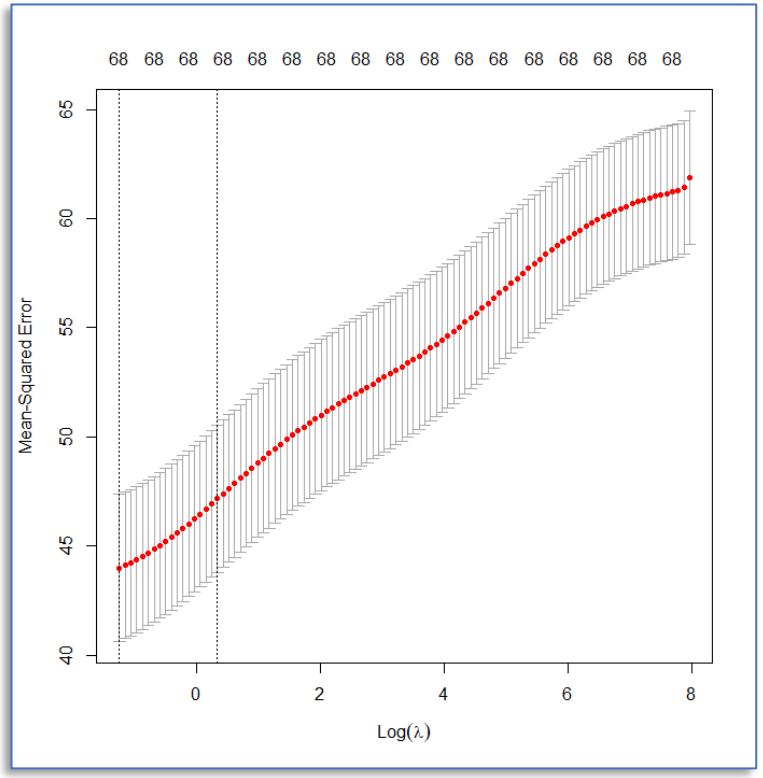

Lasso R-Squared
0.5
0.4
0.3
0.2
0.1
Training
Test


Elastic Net R-Squared
0.5
0.4
0.3
0.2
0.1
Training
Test

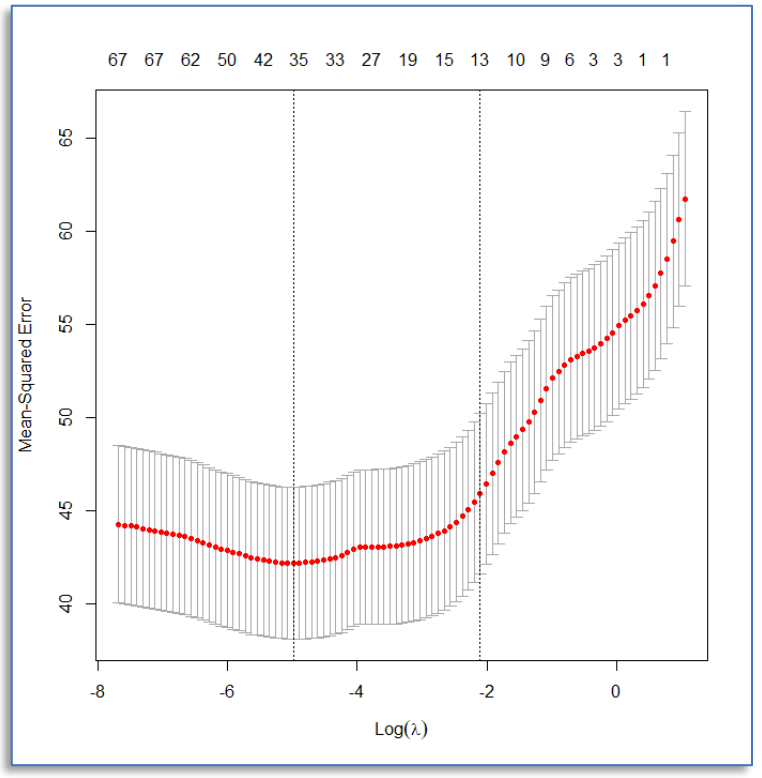

Ridge R-Squared
0.5
0.4
0.3
0.2
0.1
Training
Test


Random Forrest R-Squared
0.75
0.50
0.25
Training
Test

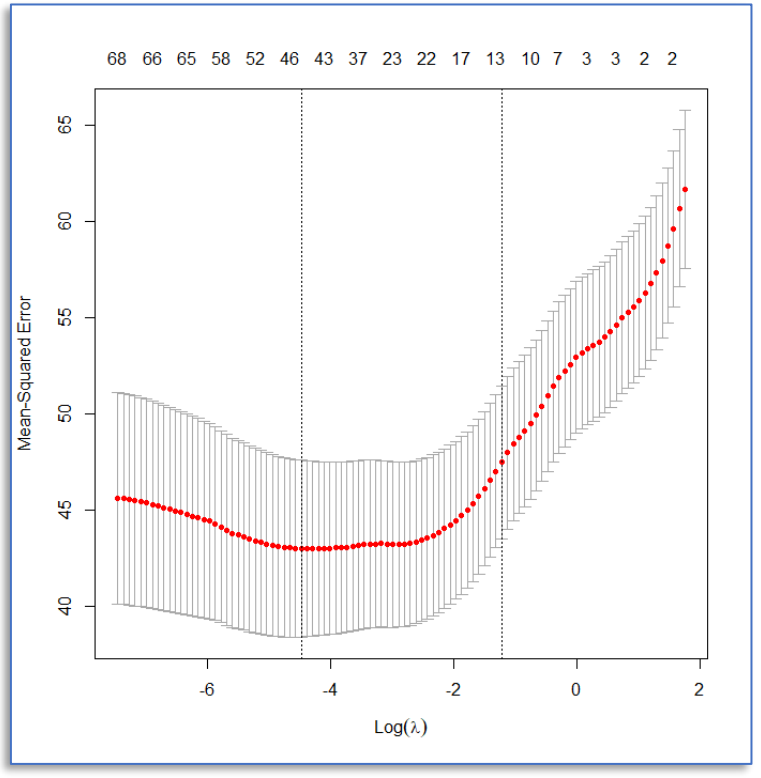
RIDGE

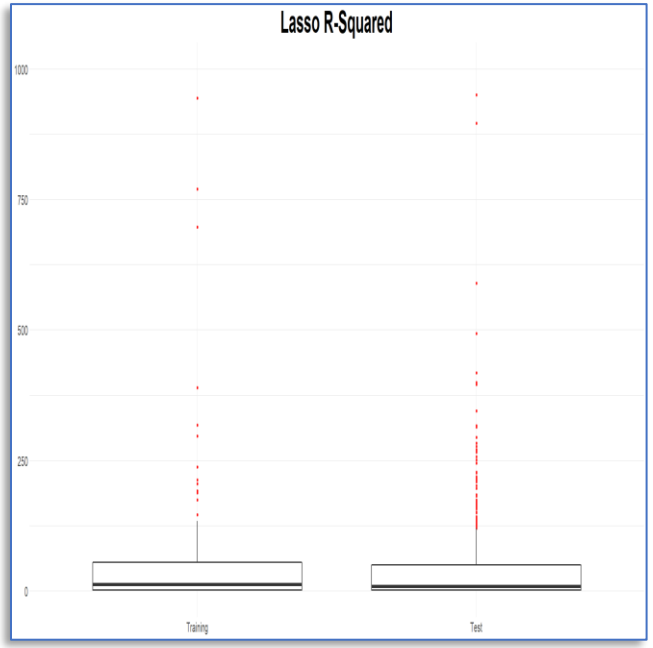


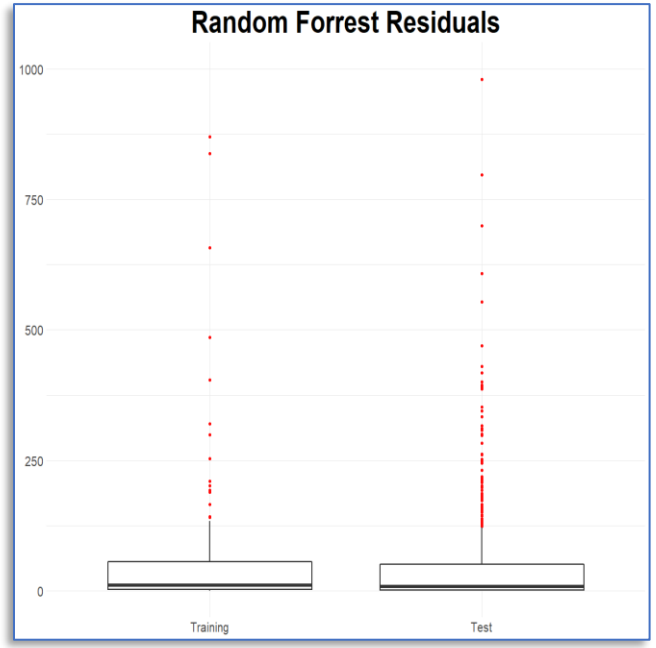
LASSO

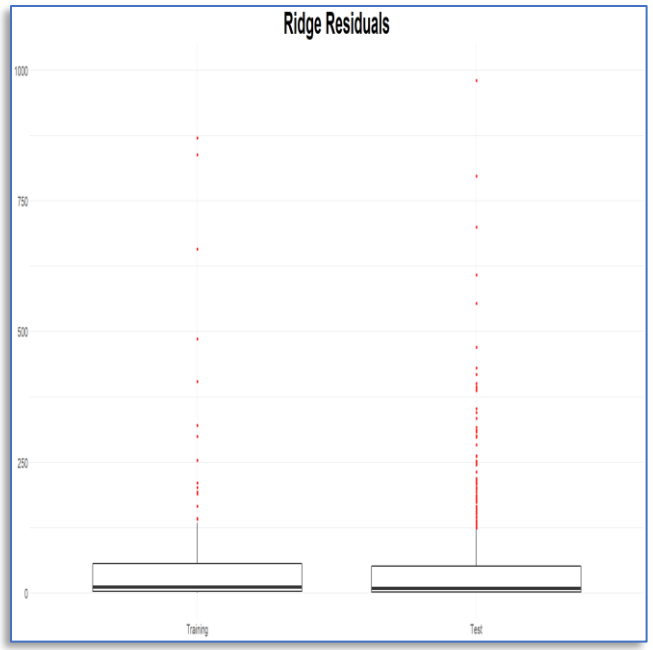


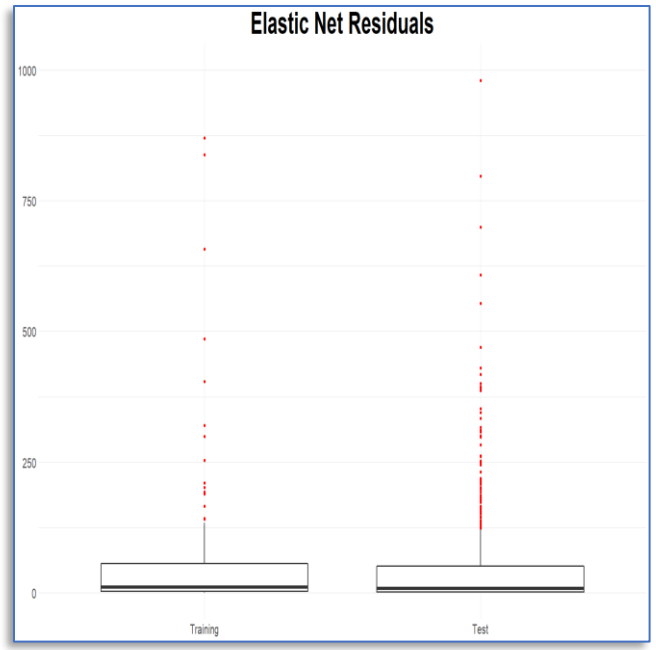
ELASTIC NET



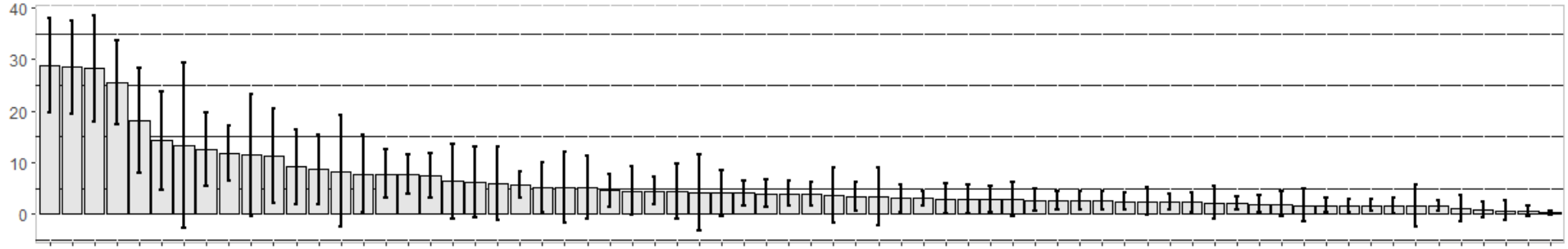

Lasso R-Squared
1000
750
500
250
0
Training
Test


Random Forrest Residuals
1000
750
500
250
0
Training
Test

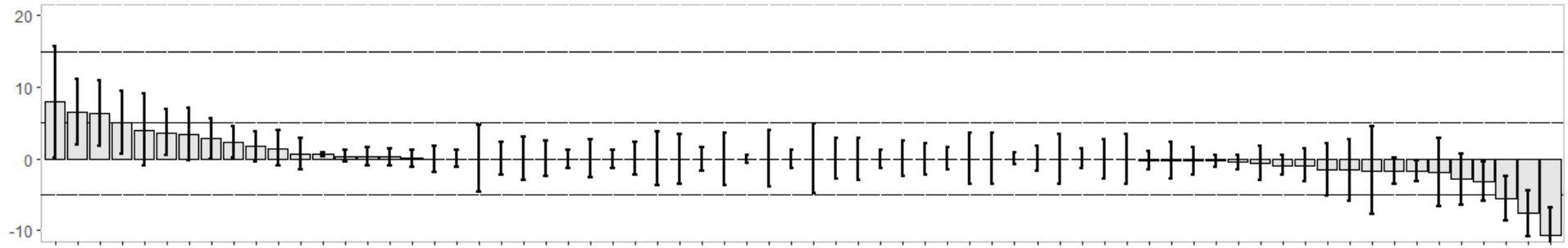

Ridge Residuals
1000
750
500
250
0
Training
Test


Elastic Net Residuals
1000
750
500
250
0
Training
Test

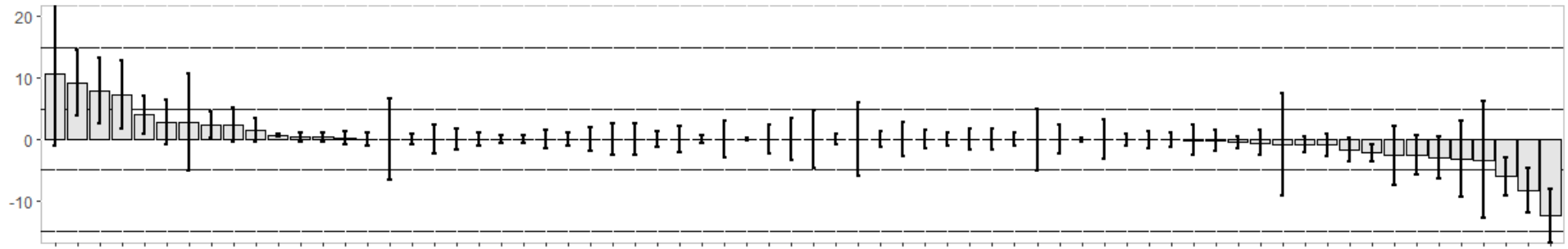
Random Forrest Betas



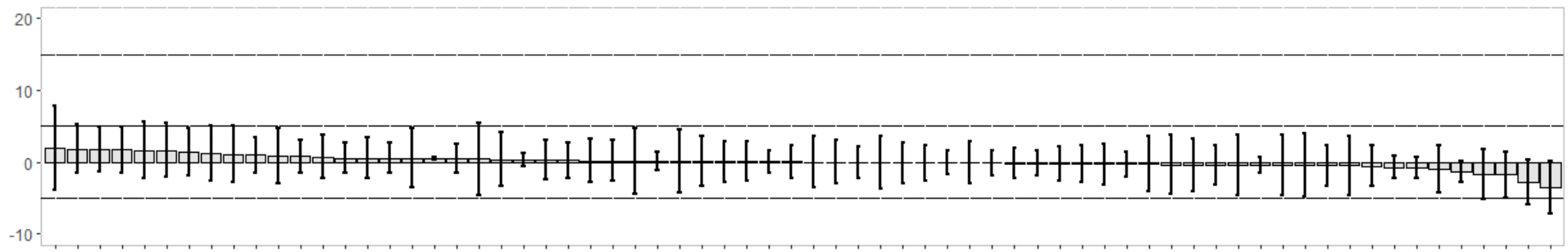
Elastic Net Betas



Lasso Betas



Ridge Betas



Performance and Conclusions

- Beta factors met theoretical assumptions of the model
- More comprehensive data could lead to some reliable conclusions
  - Demographic and education data
- The variables were not independent
- Performance: ~ 90 minutes to run the 450 lines of code