

# Used Pickup Truck Prices

Brian Contreras and Troy Whittemore

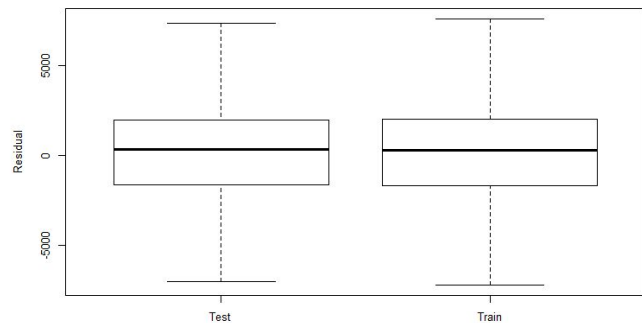
Data can be found on <https://github.com/BrianContreras317/9890-Project>

## Brief Description

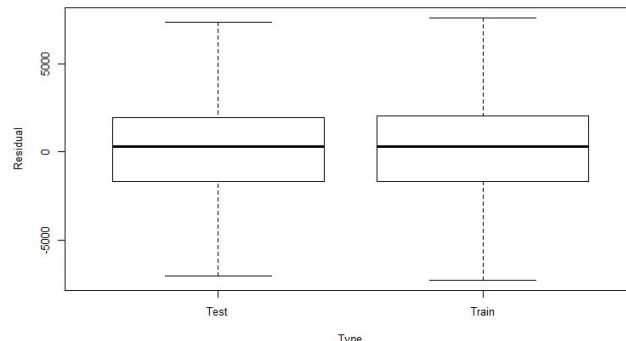
- We have 70,000 rows, we randomly selected around 10% (7,879) for efficiency
- We have 40 predictors for price. These predictors describe:
  - The size of the car
  - Engine Information
  - Market Information
  - Type of Car
  - History of the Car
- Our data includes 20 numeric variables and 20 categorical variables

# Residuals boxplots

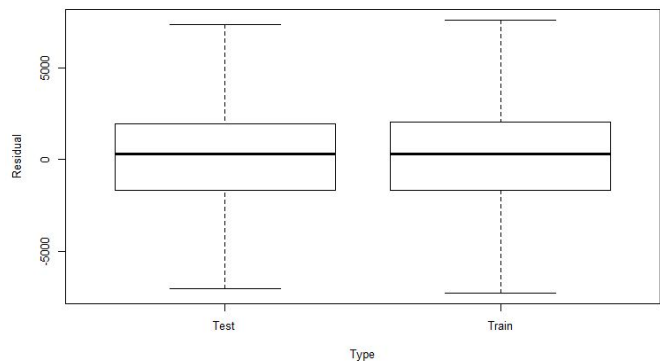
Elastic Residuals



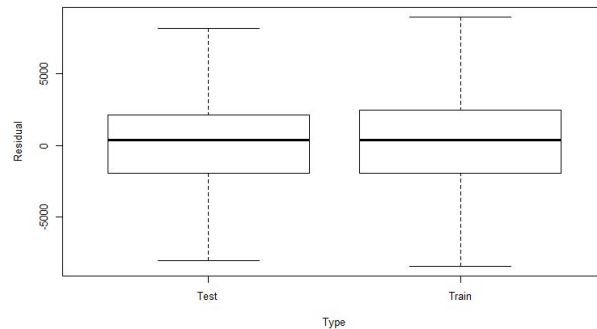
Lasso Residuals



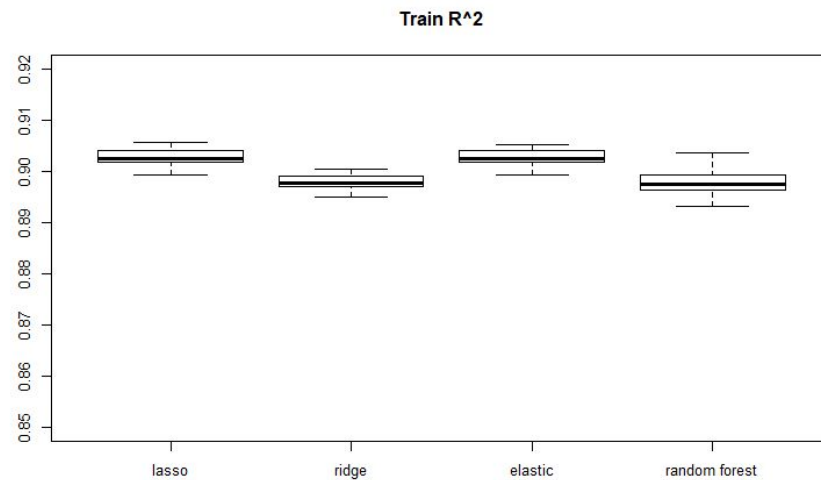
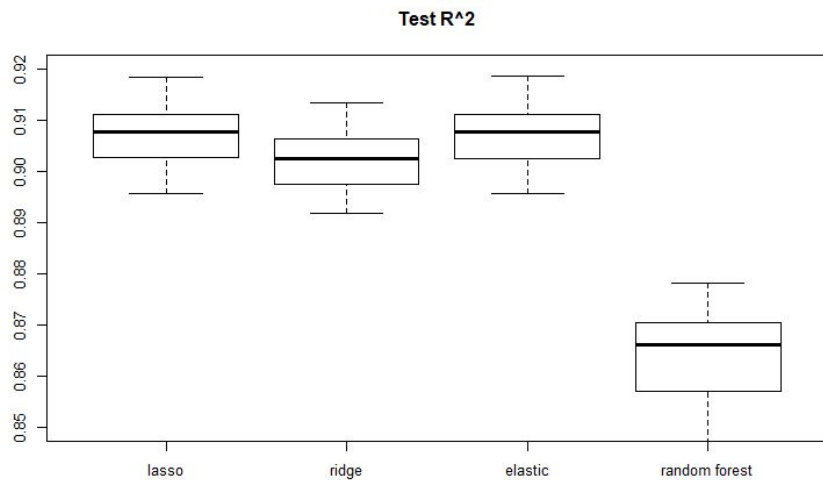
Ridge Residuals



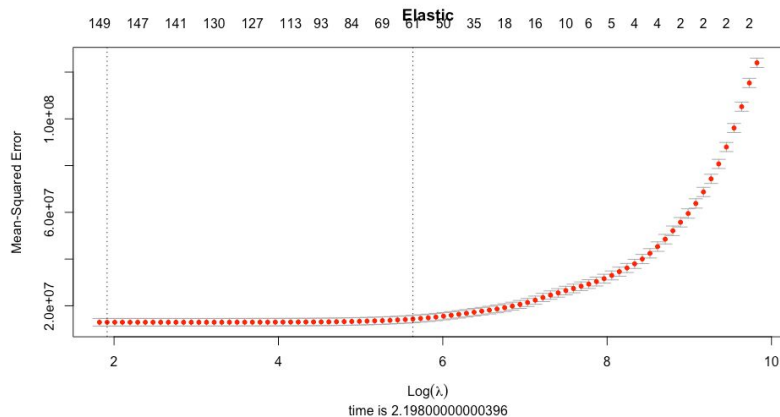
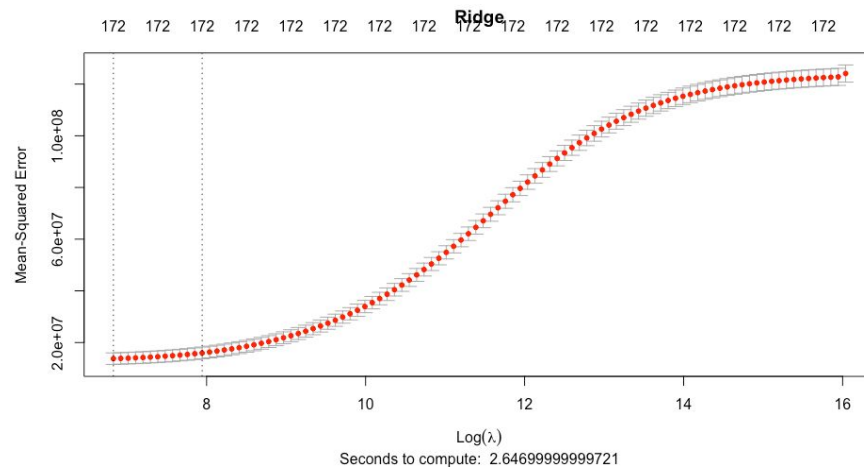
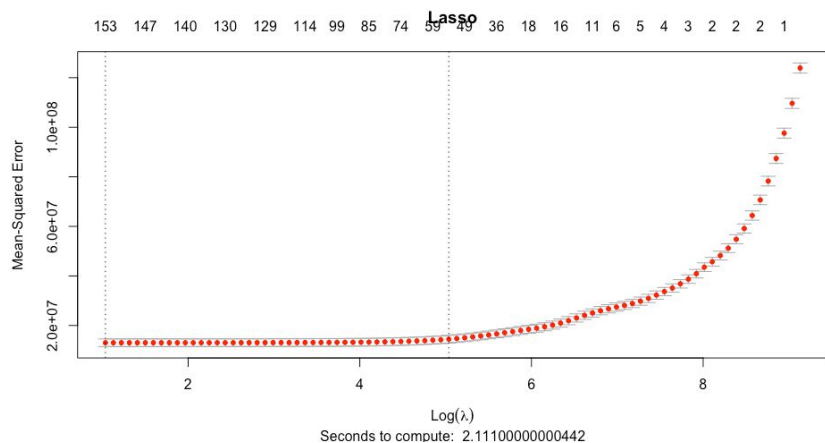
Random Forest Residuals



# R<sup>2</sup> results



# Time to cross-validate

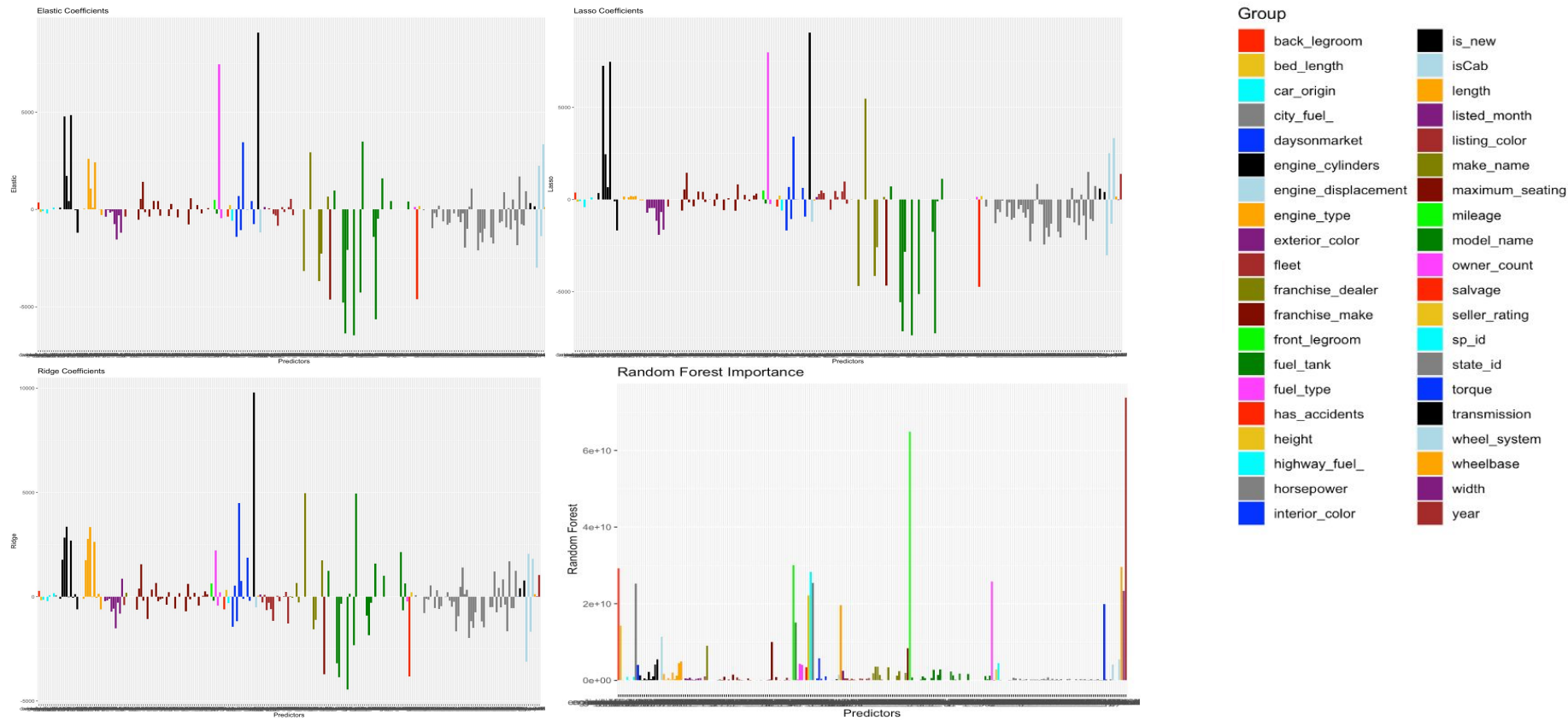


Model	Time for Cross Validation
Lasso	2.111
Ridge	2.647
Elastic	2.198

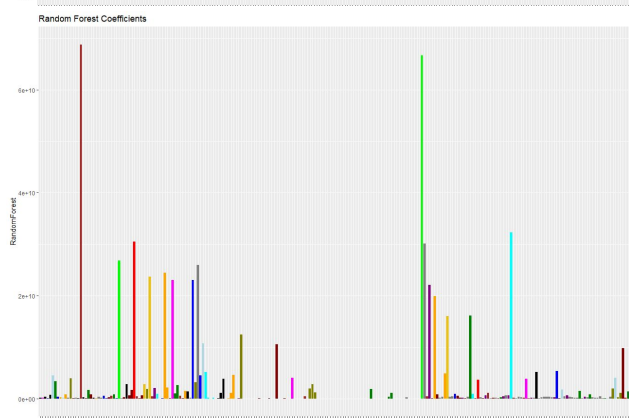
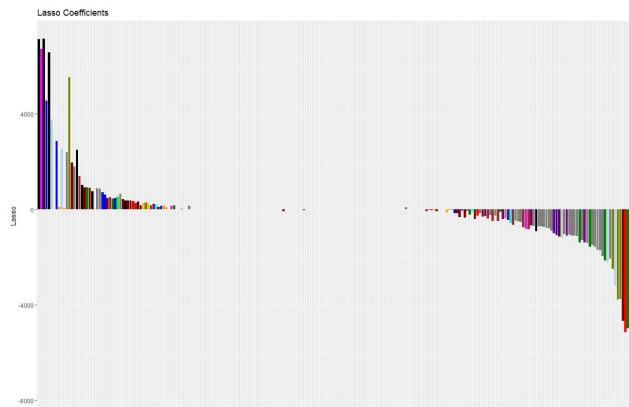
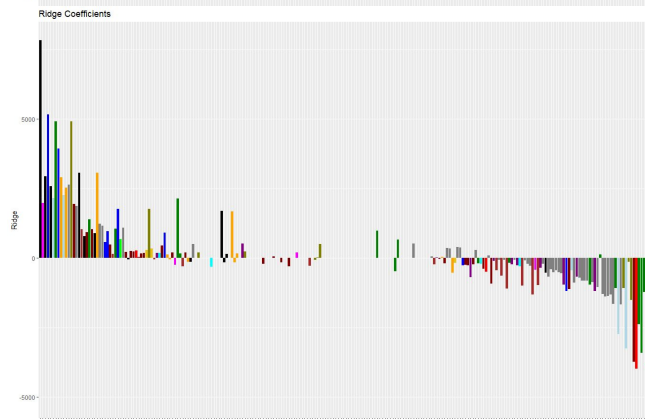
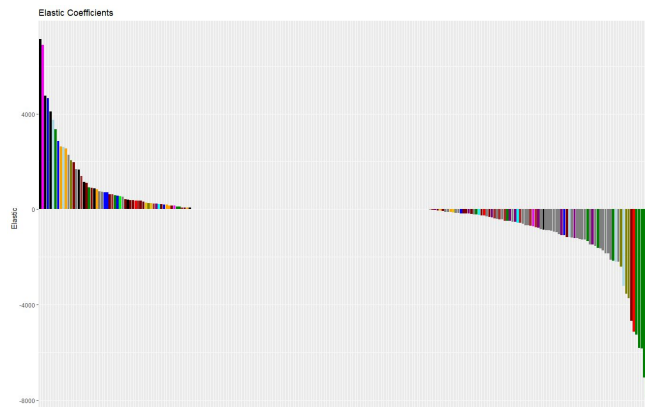
# Time vs 90% $R^2$ interval

Model	Time	90% $R^2$ interval
Lasso	3.376 sec	(0.859, 0.915)
Ridge	4.637 sec	(0.856, 0.910)
Elastic	3.418 sec	(0.859, 0.916)
Random Forest	87.217 sec	(0.820, 0.874)

# Estimated Coefficients (1 / 2), unordered



# Estimated Coefficients - by Elastic descending order



## Group





# Concluding Remarks

- Methods mostly agree which coefficients are most important
- Regression models performed best for us
- The models show us that we can account for ~90% of the variation of price using the predictors within the dataset.