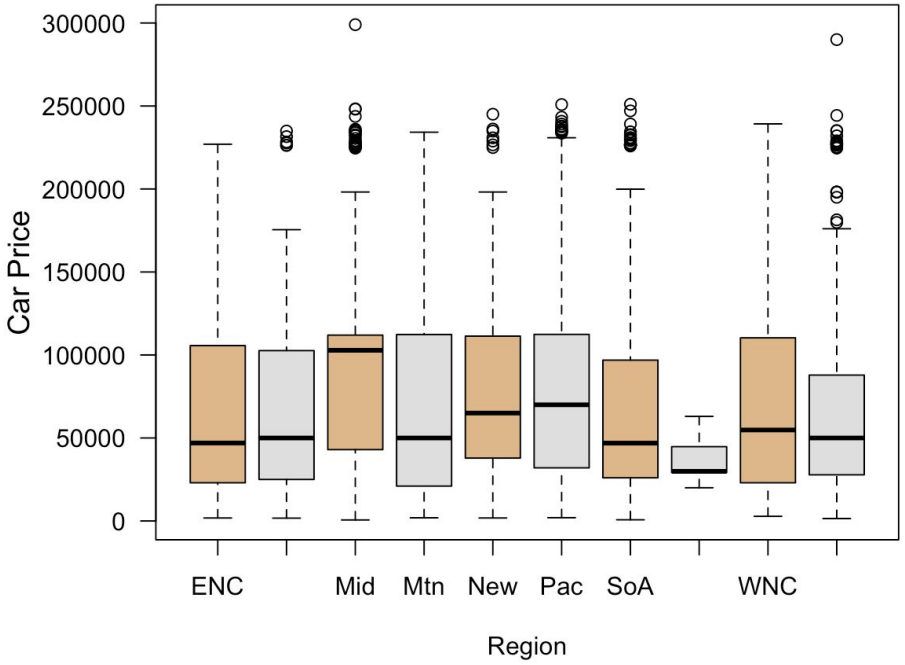
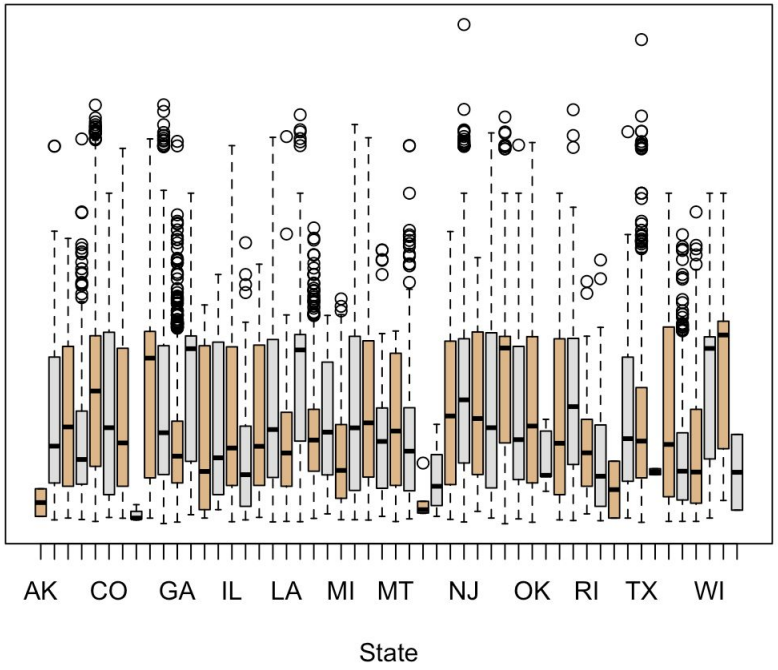
PREDICTING CAR PRICES

Reed Dalton, Gaby Lio, Ashwin Param, Bhavana Vijay, Reece Wooten

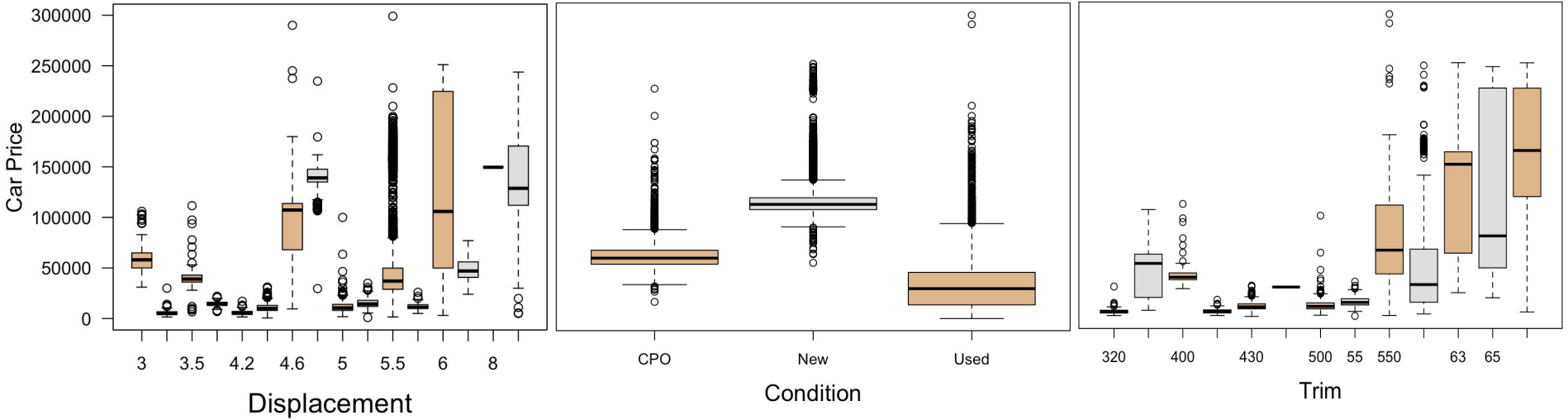
INITIAL DATA ANALYSIS



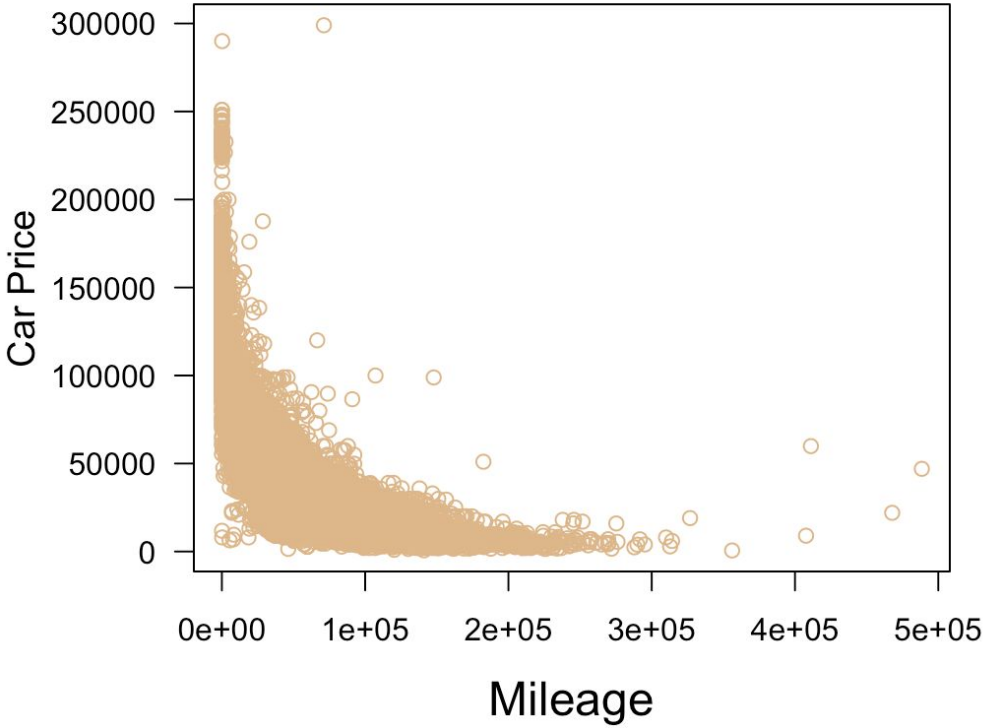


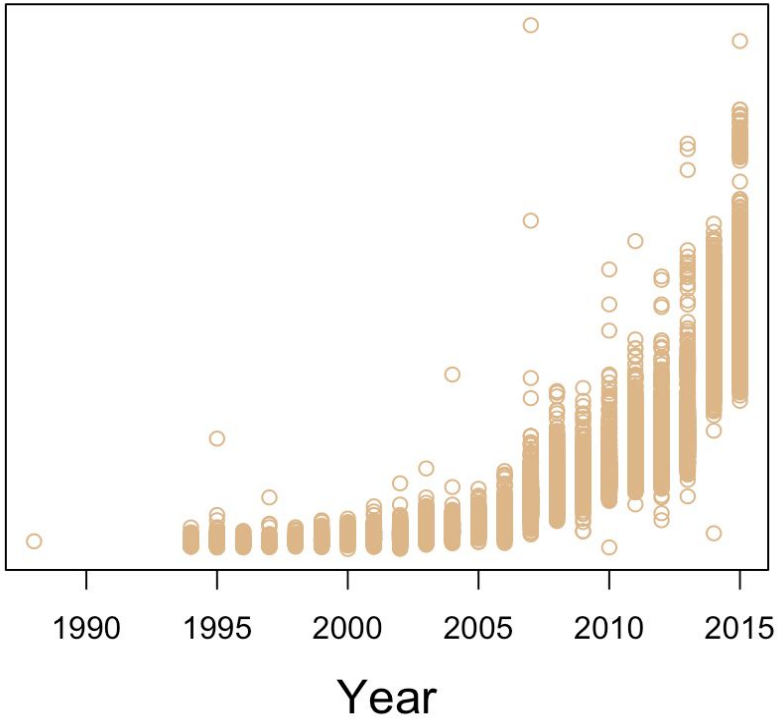
SubTrim was also dropped because of low variability

INITIAL DATA ANALYSIS

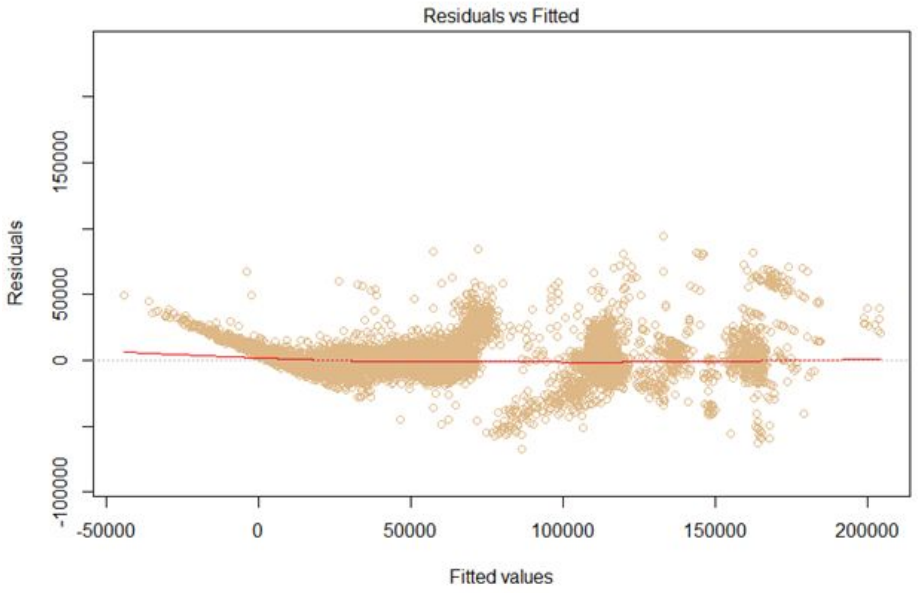


INITIAL DATA ANALYSIS

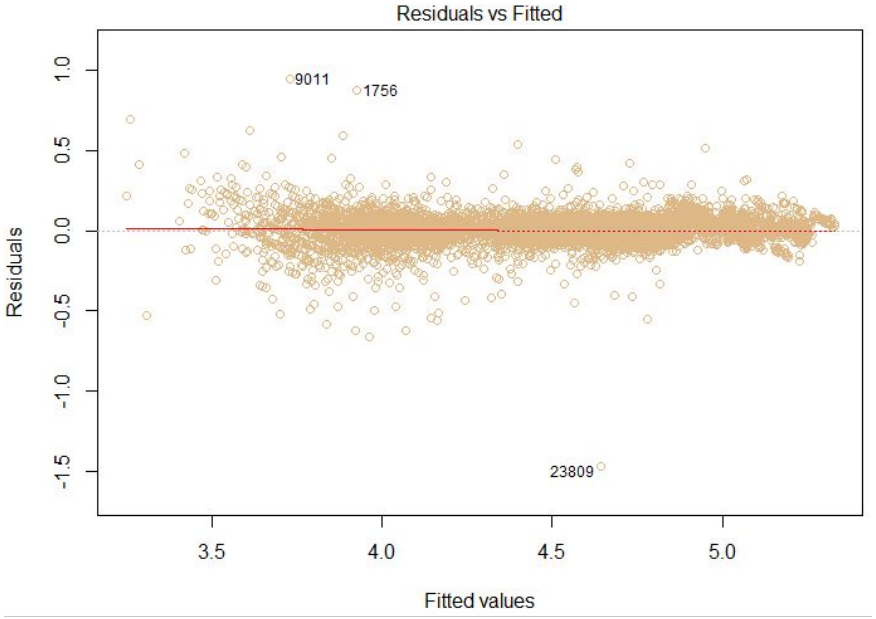




LINEAR MODELS



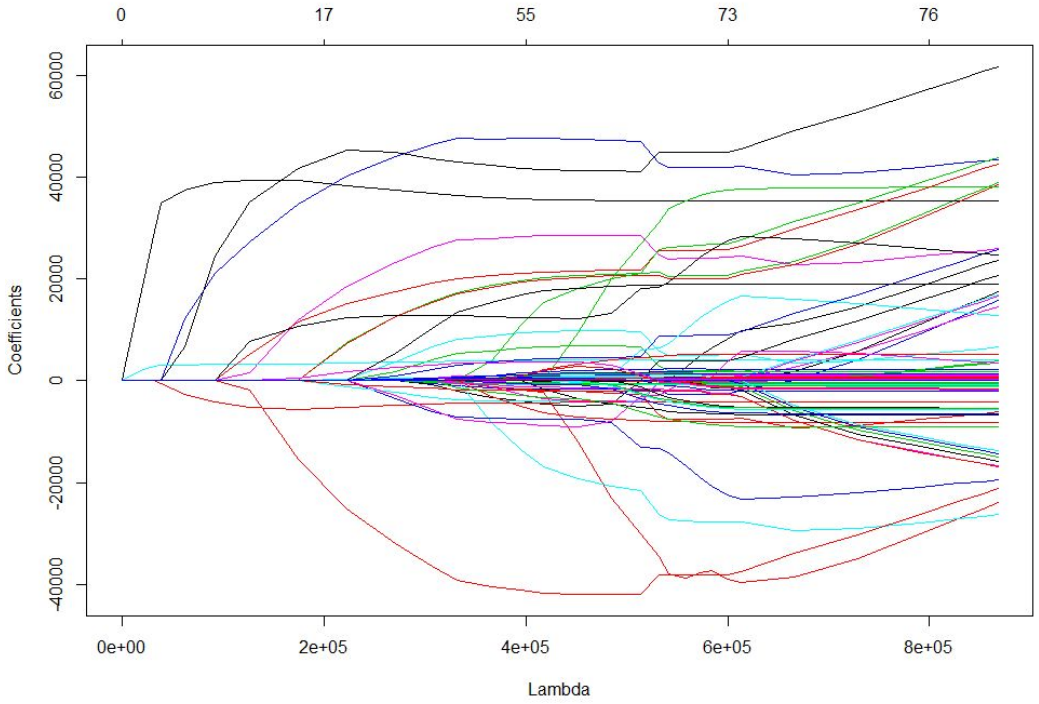
Price vs. All Variables



log*(Price) vs. All Variables*

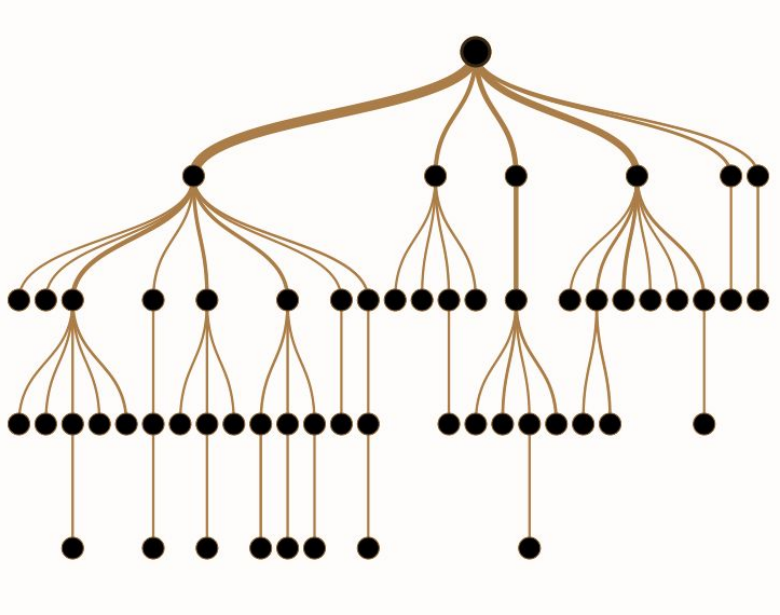
LINEAR MODELS

- Color
  - Black
- Displacement
  - 5
  - 5.4
  - 5.5
- Trim
  - 55
  - 350
- Sound System
  - Boston Acoustic
  - Premium



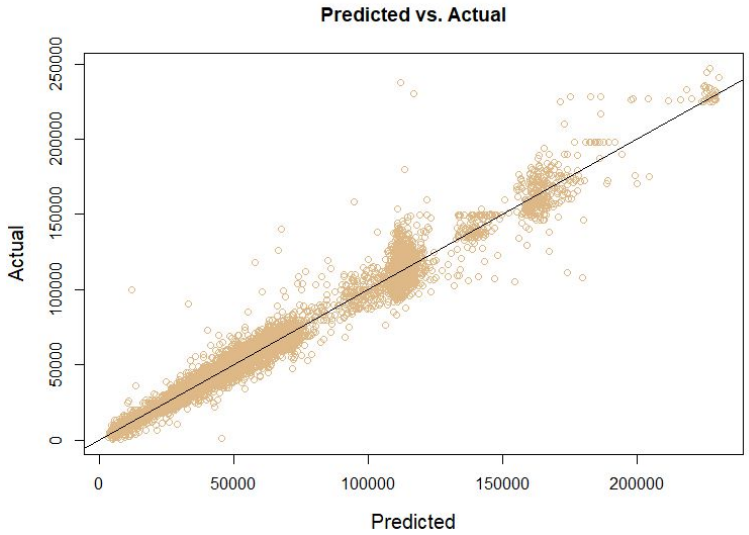
TREES

- Training Set vs. Test Set
  - 70/30
- Random Forest
  - *N* Number of Trees
  - *M* Number of Variables
  - Node Size
- Bagging
  - *N* Number of Trees
  - Node Size
- Boosting
  - *N* Number of Trees
  - Shrinkage Parameter
  - Interaction Depth

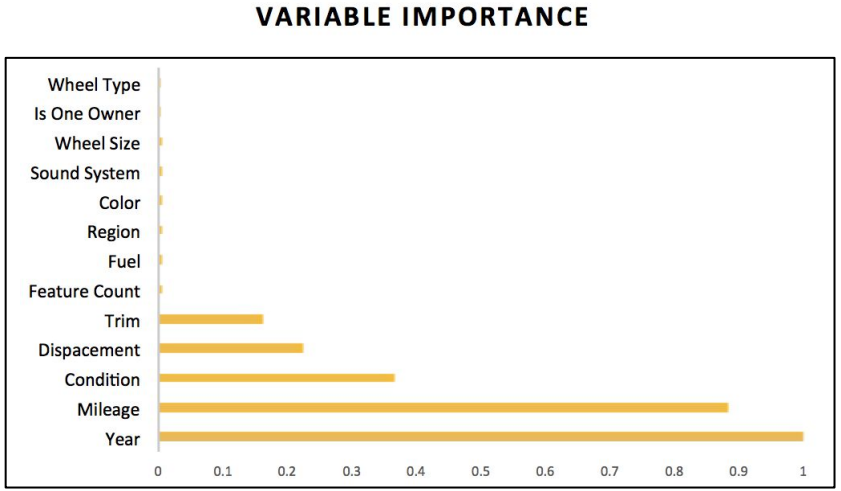


RESULTS

- RandomForest
  - Number of *n* trees = 200
  - Number of *m* variables = 6
  - Node size = 20



- RMSE
  - $6,688.96



INSIGHTS

- Random Forest vs. OLS
- Why RMSE?
- Limitations
  - Heteroskedasticity
  - Not randomly distributed