CUNY MSDS Capstone Project

COMMERCIAL BUILDING ENERGY CONSUMPTION

ANALYSIS AND PREDICTION

April 9, 2019

John Grando
john.grando@spsmail.cuny.edu

Contents

Abstract

Commercial Building Energy Consumption accounts for approximately 25%[1] of the United States energy production profile. Many economical and sociological factors are pushing owners of these buildings to reduce energy consumption and optimize performance. However, it is difficult to say whether a building is operating efficiently or not. Using publicly available data, models can be constructed to predict major fuel consumption. Keywords: building energy consumption, predicted energy consumption, baseline energy model.

Introduction

Building owners, local governments, and utility providers are all looking for ways to reduce energy consumption. Reasons for doing so can vary all the way from social responsibility to economic gain. Some people want to show off an efficient building, others want to identify properties that are in need of improvement. However, this concept of evaluating building performance typically requires someone to measure the property's consumption and then compare it to a standard practice equivalent. However, comparing summary statistics between buildings, such as energy use per square foot, is not as simple as it seems because there are a multitude of factors that affect a building's energy consumption profile. Factors such as building type, number of employees, etc. can may have critical importance for some buildings and not others. This complexity of making similar comparisons creates a situation where it is difficult to determine whether a building is performing consistent with, or better than, standard practice buildings.

Commercially, using the most popular example, ENERGY STAR [2] has implemented a benchmarking algorithm that scores buildings on a scale from 1 – 100 using market-available data. The output of this benchmarking algorithm is a unit-less score, as well as a reference 'baseline' building. However, the methodology is not released and it is unclear what factors are important to influence the energy consumption of the building. These barriers make it difficult to provide custom comparisons and nearly impossible to make batch predictions from a set of buildings, or variations of buildings.

Every few years, the U.S. Energy Information Administration (EIA) conducts a survey attempting to record pertinent features of these buildings, known officially as the Commercial Buildings Energy Consumption Survey (CBECS)[3]. While the survey is expansive (i.e. more than 600 tracked features), it is essential to create a model that is usable and only requires predictors that can be attained by building operators. Therefore, in this study a series of models will be evaluated in order to determine the most important predictors which will then be used to train a final, more complex, model.

[1] EIA - `https://www.eia.gov/energyexplained/index.php?page=us_energy_commercial`
[2] ENERGY STAR - `https://www.energystar.gov/`
[3] EIA Microdata

Research

Related Work

The idea of determing building energy efficiency is not a novel concept in itself. As previously mentioned, ENERGY STAR has a building benchmarking tool[4]. Additionally, the United States Green Building Council has created the Arc Platform[5] which provides benchmarking and active monitoring features. While these platforms provide building comparisons in the form of an overall score, it is difficult to explore the space around the building attribute inputs themselves as well as compare consumption of a specific building to its equivalent standard practice building. With this functionality, a more direct comparison can be made and relative environmental impact can be measured.

Literature Review

There are a variety of texts that are dedicated to the analysis of building energy consumption, and determining operating efficiency. For example, ASHRAE Guildine 14ASHRAE Guideline 14[6] provides a standardized set of energy, demand, and water savings calculation procedures. Also, there are guidelines that must be followed for buildings undergoing new construction or major renovation, which have energy compliance sections (ASHRAE Guideline 90.1, 189.1, and International Energy Conservation Code)[7]. Particularly, there is a well thought out process for auditing commercial buildings, known as ASHRAE Audits, which start at the lowest level (I) and progress to the highest level (III) as the opportunity for energy and cost savings becomes more apparent [8].

Theory and Hypothesis

Commercial buildings are complex and encompass a wide variety of purposes. In order to be functional, they all must be powered, and require a considerable amount of energy to operate properly, which can be costly. In fact, there is a whole industry dedicated to ensuring the proper operation of a structure. The most direct example is the ASHRAE energy audit process. As part of the initial audit process, an assessment of the building's overall operational efficiency is gauged. Typically, an auditor will walk through the buidling, analyze utility bills, and make the closest energy consumption comparisons they can. This takes years of experience and sometimes requires highly tuned spreadsheets that have been developed over a long period of time. It can take a surprising amount of effort just to determine if a building is operating efficiently or not, which demonstrates how useful it could be to have a model at hand which predicts building consumption based on easily attainable features.

The CBECS data set provide some insight as to what building attributes most greatly affect building energy consumption. Over 400 survey questions are recorded and coupled with major

[4] https://www.energystar.gov/buildings/about-us/how-can-we-help-you/benchmark-energy-use/benchmarking

[5] https://arcskoru.com/

[6] -https://www.techstreet.com/standards/guideline-14-2014-measurement-of-energy-demand-and-water-savings?product_id=1888937

[7] https://www.energycodes.gov/status-state-energy-code-adoption

[8] http://aea.us.org/3143-2.html

fuel consumption. These fuel sources are Electricity, Natural Gas, District Heat, and Fuel Oil. However, it would not be useful to construct a model with a large number of predictors, as it would require a large amount of time and effort to compile the necessary information in order to provide a prediction. Therefore, one of the main focuses for this study will be to extract the fewest amount of predictors necessary in order to make accurate predictions.

Given the complex nature, it is unlikely a linear regression will provide the best prediction accuracy. This point is especially highlighted by the fact the the goal of this study is produce a parsimonious set of predictors, which means a small subset must be selected. Therefore, an investigation into more complex, nonlinear, algoriithms will be performed in order to keep the number of necessary predictors as low as possible while still capturing complex interactions.

Data and Methods

General Process

Due to the large number of features in the survey responses, it is not possible to analyze each one individually. Therefore, the first steps in the process will be centered around selecting a smaller subset. First, a principle component analysis will be performed. Second, a partial least squares model will be fit to the response. Third, a random forest regression will be used in order to try and extract any nonlinear relationships. Fourth, an attempt to construct a lasso regression model will be made. Fifth, a forward selection linear model will also be fit in order to see if an automated approach can be taken. Finally, a simple neural network model will be trained to gauge the possible effectiveness of using this model type. The magnitude and contribution percentage of each variable will be considered in selecting features from this model. Also, the various error rates from each preliminary model will be used as a benchmark for the final model performance.

After the prelminary set of models have been run and summarized, the extracted variables will be analyzed in order to verify their importance, gauge their potential predictive power, and to check whether they are easily attainable for a building operator/owner. This step is very important because it is essential worthwhile variables are used to predict the outcome. Selecting a variable that, for one reason or another, is erroneous may lead to reduced predictive power in the final model. If a variable did pass our initial analysis but doesn't actually have much predictive power (i.e. it only changes values by a slight amount) then it may not be worthwhile to select it at all. All selected variables increase the complexity of the model; therefore, we wish to only select those that will matter. Finally, the predictor must be usable, and 'knowable'. These three concepts will be used in the analyzation of the candidate variables from the the preliminary analysis.

Finally, a neural network model will be built to take the verified subset of features and make predictions for the selected major fuel use. A variety of hyperparameters will be tested, using cross-validation, and compared on a common error metric. This step will reveal the optimal hyperparameter combination to use for the model. The prospective model will then be retrained on the entired entire training and validation data. This model's selected error metrics will then be compared to the preliminary models, which should be considered a floor for performance. Once this is done, the model can then be re-assessed for feature selection as well as analyzed for the value/tradeoff of adding/removing certain features.

For each fuel end-use, two parrallel final models will be considered; one model will use a total consumption response variable in units of mmBTU, and another model that will be considered will have the response variable normailzed based on gross floor area in units of BTU per square foot. The final decision on which model will be chosen will come after the neural network models have been fully trained.

Data Pre-Processing

The raw data set consists of 6,720 samples and 1,119 features. However, multiple steps of preprocessing were required in order to prepare the data. Note, there are many columns which are being used as imputation flags and statistical weights (for aggregation) which, when removed, reduced the number of features down to approximately 400. While these columns are useful to indicate where values have been imputed into the dataset by the source's own methodologies, rather than try to change back the data to the original records it has been determined that the imputed values were sufficiently applied and the dataset will not be imputed any further.

After evaluating the reduced data set, some feature engineering efforts were taken. First, very specific cases which resulted in many null responses to follow-up survey questions (e.g. buildings less than 1,000 gross square feet), buildings open for less than a year, and features with a large amount of nulls were removed. Second, some null entries were converted to zero when logically appropriate. For example, if a building was indicated to not be cooled, then a follow up question asking what percentage of the building is cooled was not asked, resulting in an null. In this instance, the null value was replaced with a zero. Third, some values were removed as they simply did not apply to the study (e.g. expenditure for energy sources in USD). Fourth, nominal categorical values that had null responses were encoded to a special value. The thinking for this approach is that if, in fact, an null value for a feature ends up being a significant predictor, then it can be analzyed what factors make this situation occur. Fifth, the categorical features were then one-hot encoded to separate columns. The preprocessed data set was transformed to 6661 rows and 456 features (before one-hot encoding).

Electricity

General

The preprocessed data was passed to the aforementioned set of algorithms in order to determine the best possible set of candidate predictors with some additional adjustments. Only buildings that indicated electricity being used `ELUSED` were included in the samples for this major fuel use. Then, one of each pair of predictors with correlations above 0.75 were removed, to avoid model selection issues. Additionally, the other major fuel consumption values were removed from the set of possible predictors. Also, the numeric predictors were transformed via BoxCox methodology as well as centered and scaled due to the varying scales and skewness.

Two potential outlier was found in the analysis. A public assembly space reported an energy consumption of 1E09 BTUs whereas the next highest consumption for this budiding type, with similar area was 3E08 BTU (less than one third of the value), and the 3rd quartile value of this subset is 5.5E06. While it is noted that there were significantly higher indications of refrigeration use than other comparables, the inclusion of this data point still vastly skews most models due to its high leverage. Similarly, an 'Other' space type has a reported energy consumption of 7E08 BTU whereas the next highest value is 2E08 BTU and the third quartile value is 2E06. While this building is large (1.4 mmSF), and has a lot of server equipment (¿500), it is still greatly beyond the next closest category and seems to be causing instability in the models due to lack of similar data points. Therefore, these points have been removed and the caveat of instabilitity past a maximum limit will be instituted (¿5E08 BTU), due to lack of additional information.

Response Analysis

The response data appear to be unimodal and have a heavy right skew. After filtering for this model's end-use, there are 6499 samples in the data set. The energy use was convert to units mmBTU (1e6 BTU) and the log was taken in an attempt to maintain homoscedacity as the variance of the energy used also scales with the magnitude. *Appendix*

Variable Selection - PCA

RMSE: NA, Rsquared: NA
Top 5: `COOK.2[NO]`, `LAUNDR..1[NA]`, `ELCPLT..1[NA]`, `PBA.14[EDUCATION]`, `BLDPLT.2[NO]`

The principle component analysis indicates that only About 5.0% of the variance in the data can be explained in the first principle component, which then drops to about 2% for the second principle component. These results reveal that there does not appear to be a clear set of axes that can explain the variance of the data very well, which indicates there may be some very complex interactions taking place in the predictors. *Appendix*

Variable Selection - PLS

RMSE: 46880, Rsquared: 0.548
Top 5: `NWKERPerSf`, `RGSTRNPerSf`, `FDSEATPerSf`, `RFGWINPerSf`, `PCTERMNPerSf`

This model returned a promising result; however, it must be noted that all predictors were used in this process. Looking at the output thus far, it is appears that the number of workers, receptical equipment, and refrigeration equipment, influence electrical consumption. *Appendix*

Variable Selection - Random Forest

RMSE: 163434, Rsquared: 0.173
Top 5: `RFGWINPerSf`, `RGSTRNPerSf`, `NWKERPerSf`, `RFGICNPerSf`, `PCTERMNPerSf`

The resulting error metrics were much less promising. However, similarly selected variables are are picked for this model when compared to the PLS. *Appendix*

Variable Selection - Lasso

RMSE: NA, Rsquared: NA
Top 5: NA

This resulted in a very poor fit, which is not unexpected. Lasso models typically work when a few variables can be used to predict the response, which does not appear to be the case in this instance. Due to the lack of fit, this model will not be used in the variable selection process. Additionally, the actual model was poor enough that predictions could not be made on the data, which is the reason for the lack of reported metrics.

Variable Selection - Forward Selection

RMSE: 90503, Rsquared: 0.315
Top 5: `NWKERPerSf`, `RFGWINPerSf`, `RFGWI.1[YES]`, `RFGICNPerSf`, `PCTERMNPerSf`

This model was building using the leaps package which iteratively selected the best predictor variable up to a limit of 100. Unsurprisingly, the best model turned out to be the maximum setting. Large refrigeration equipment load and typical office space attributes dominated this analysis, as appears to be the case in previous models. It seems that in order to capture energy use for all buidling types, much more than 5 variables will be necessary. *Appendix*

Variable Selection - Recursive Feature Elimination

RMSE: 48022, Rsquared: 0.517
Top 5: `RGSTRNPerSf`, `RFGICNPerSf`, `NWKERPerSf`, `PBAPLUS.32[FAST FOOD]`, `RFGWINPerSf`

A more direct approach was taken with this model, which is specifically used to extract useful features from data sets. *Appendix*

Variable Selection - Simple Neural Network

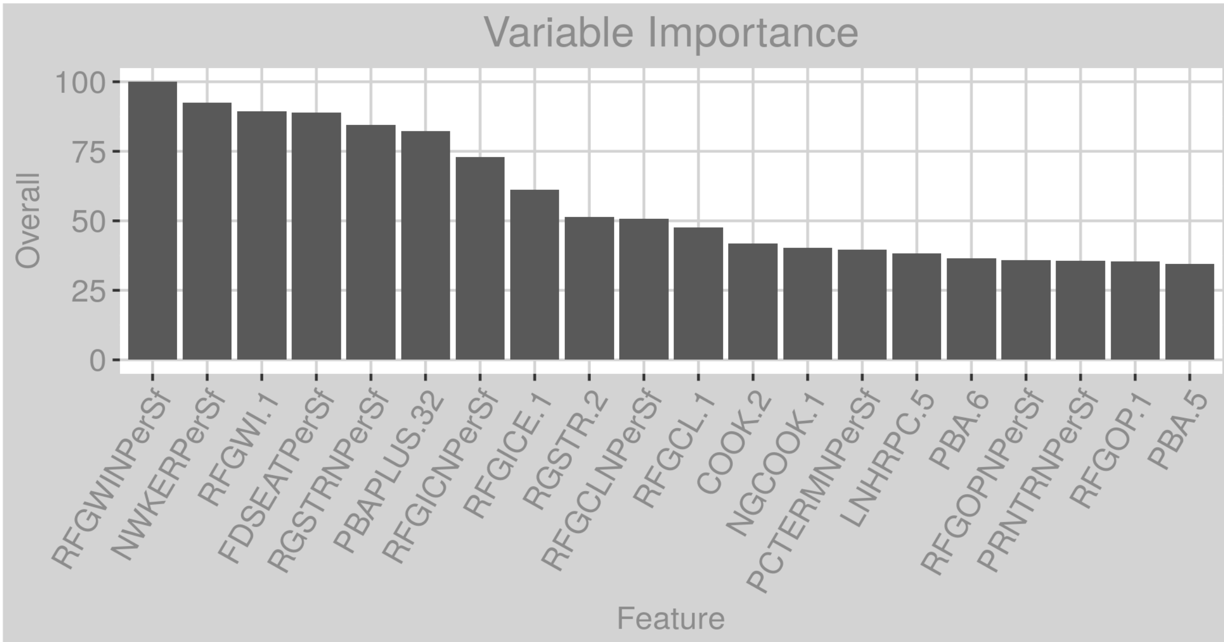
RMSE: 48644, Rsquared: 0.499
Top 5: `FDSeatPerSf`, `PBAPLUS.32[FAST FOOD]`, `RFGWINPerSf`, `RFGWI.1[YES]`, `RGSTRNPerSf`

Given that the final model will be a neural network, it made sense to try a simple out-of-the-box training model to see if any particular features work better with this process. As can be seen, there are some new attributes that surface which were not indicated to be of high importance previously. *Appendix*

Variable Selection - Selected Variable Analysis

Feature Extraction Model Results

Model	RMSE	R2	MAE
partialLeastSquares	46880.41	0.5475726	24461.50
recursiveFeatureExtraction	48022.21	0.5168333	29511.04
neuralNetwork	48644.23	0.4990216	28964.80
leaps	90503.65	0.3149572	59157.08
randomForest	163434.65	0.1731248	141674.66



In order to rank the most impactful features, the variable importance metrics from the selected models were all set to the same scale then summed up. As a preliminary check, the top 20 predictors are plotted in the appendix and are generally discussed here. It seems the attempts to create stratified random samples may have been beneficial in this case since there are some building type specific end-uses that are highly ranked. As previously noted, there are many attributes associated with refrigeration, office, and food sales equipment. Also, the attribute identifying one of the more atypical building types, speaking in an energy intensity sense, has made it into the top 20 (`PBA.5[NON-REFRIGERATED WAREHOUSE]`). Additionally, some occupancy features (`NWKERPerSf`, `FDSEATPerSf`) have been included which is expected given that they impact interior space cooling and ventilation loads. In an attempt to truly follow the imporant predictors, no variables have been removed from this set and the order of importance remains unchanged. *Appendix*

Natural Gas

General

As previously noted, only buildings that indicated natural gas being used `NGUSED` were included in the samples for this major fuel use. Then, one of each pair of predictors with correlations above 0.75 were removed, to avoid model selection issues. Numeric predictors were transformed via BoxCox methodology as well as centered and scaled due to the varying scales and skewness. Note, no further commentary will be made in the following sections unless it differs from previous sections.

Response Analysis

After filtering for this model's end-use, there are 6662 samples in the data set. The same transformations were applied to this response variable as electricity. *Appendix*

Variable Selection - PCA

RMSE: NA, Rsquared: NA
Top 5: `EDSEATPerSf`, `PBA.14[EDUCATION]`, `STRLZR.1[YES]`, `MCHEQP[NA]`, `ACT2PCT` *Appendix*

Variable Selection - PLS

RMSE: 73076, Rsquared: 0.293
Top 5: `FDSEATPerSf`, `HEATP`, `RFGWINPerSf`, `NWKERPerSf`, `RGSTRNPerSf`

The Rsquared and RMSE values are still on the poorer side compared to those in the electricity study, which suggests there is either a more complex relationship, or there is greater variance in response, given the available data. However, the presence of attribute which indicates the percent of the building that is heated (`HEATP`) is promising. *Appendix*

Variable Selection - Random Forest

RMSE: 175808, Rsquared: 0.193
Top 5: `DRYCL.1[YES]`, `FDSEATPerSf`, `RFGWINPerSf`, `LAUNDR.3[OFF-SITE]`, `STRLSZR.1[YES]`

Again, the error values are much lower; however, the presesence of notable fuel-using equipment have been indicated to be of high importance. *Appendix*

Variable Selection - Lasso

RMSE: NA, Rsquared: NA

Variable Selection - Forward Selection

RMSE: 100385, Rsquared: 0.0.246
Top 5: `FDSEATPerSf`, `TVVIDEONPerSf`, `NGWATR.2[NO]`, `RGSTRNPerSf`, `RFGWINPerSf` *Appendix*

Variable Selection - Recursive Feature Elimination

RMSE: 59212, Rsquared: 0.523
Top 5: `NWKERPerSf`, `RFGICNPerSf`, `RFGWINPerSf`, `FDSEATPerSf`, `RGSTRNPerSf` *Appendix*

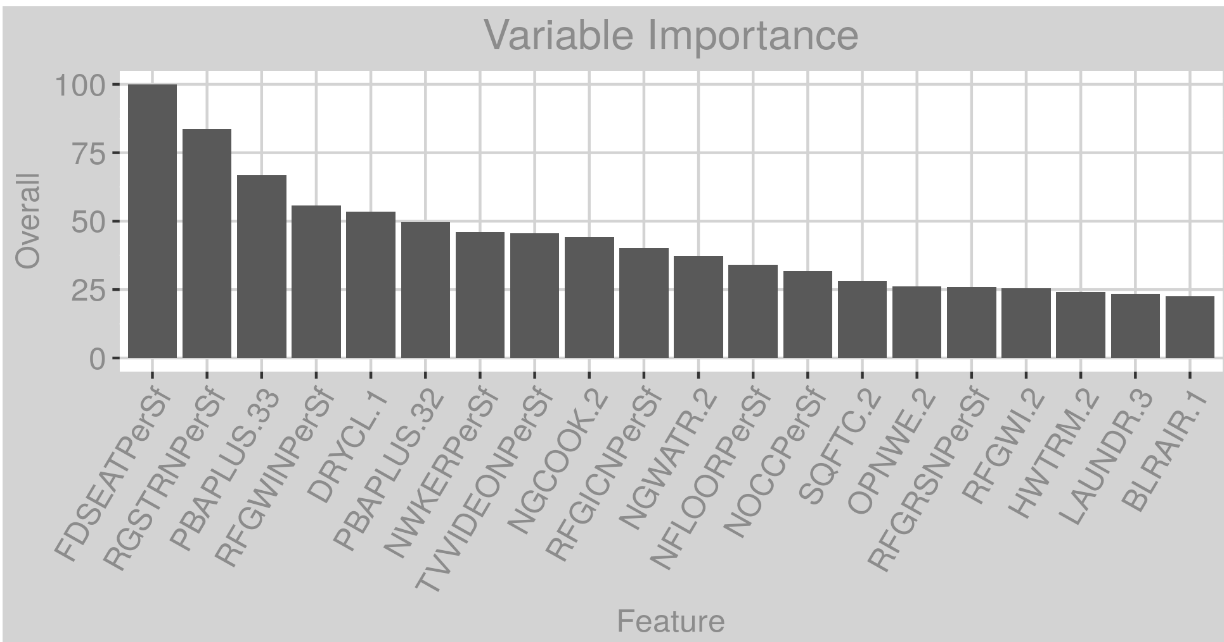
Variable Selection - Simple Neural Network

RMSE: 69784, Rsquared: 0.334
Top 5: `FDSEATPerSf`, `RGSTRNPerSf`, `DRYCL.1[YES]`, `PBAPLUS.33[RESTAURANT/CAFETERIA]`, `PBAPLUS.32[FAST FOOD]` *Appendix*

Variable Selection - Selected Variable Analysis

Feature Extraction Model Results

Model	RMSE	R2	MAE
recursiveFeatureExtraction	59212.72	0.5230241	34203.37
neuralNetwork	69784.67	0.3336745	35642.13
partialLeastSquares	73075.99	0.2934536	32068.61
leaps	100385.86	0.2462162	53658.33
randomForest	175808.23	0.1930417	147144.46



As with the electricity model, attributes related to occupancy seem to have made a large impact, possibly due to the need to heat ventilation air, especially given some of these occupancy types are associated with 24/7 operation. Also as expected, cooking and large heating equipment attributes are high on the list. Surprisingly, the number of floors per gross floor area has shown some importance, perhaps due to building shape and its relationship with heating needs (i.e. volume to area ratio). *Appendix*

Neural Network Models

General

The choice to use neural networks for the final model was multi-faceted. First, these types of models are very good at capturing complex non-linear interactions. This appears to be the case with the data set given the failure of lasso models as well as the low percentage of variance capture for the first few dimensions of the principal component and partial least squares analyses. Secondly, neural networks have the ability to select different loss functions. This is beneficial because it is important to highlight practicality of the results returned. As the estimated energy consumption grows, it is somewhat acceptable for the error rate to grow proportionally if it results in better fits for the low estimates. As an example, a large datacenter may use a lot of energy so a slightly higher relative error rate may not be a big issue since it could be a small portion of the overall consumption; however, if a non-heated warehouse with a moderate error rate, comparative to the rest of the data set, would be wildly innacurate. Therefore, the loss function for this set of models was chosen to be the mean squared logarithmic error in an effort to reflect the reasoning noted above.

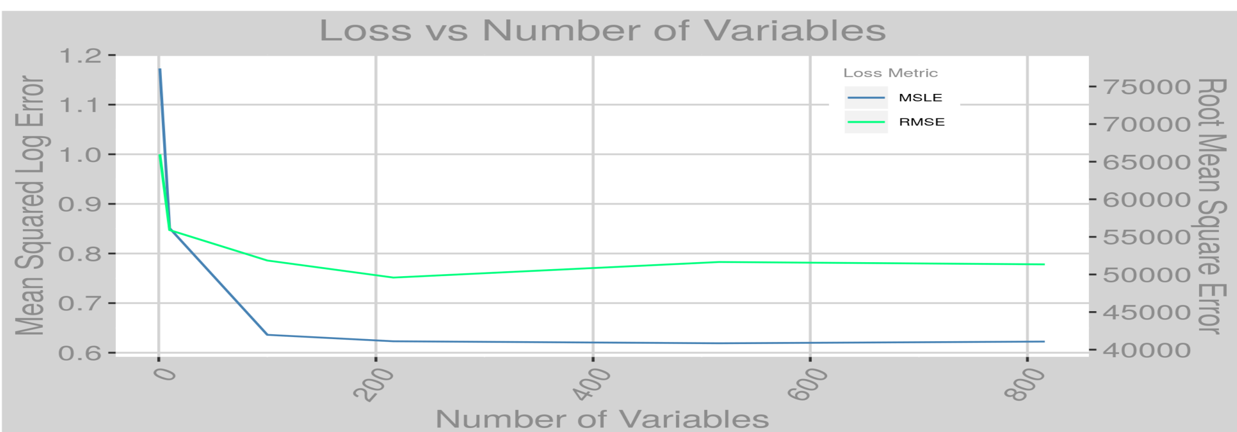
Hyperparameter Training

In order to select the most optimized set of parameters, some hyperparameter training was performed. Some standard searches were made, such as varying the dropout rate, regularization, learning rate, and batch size; however, one additional training set was incorporated to highlight the goals of this study. A series of models were tested which had an incrementally decreasing number of variables, by least importance, in order to test the loss of accuracy.

Electricity

Summary

The final selected model consisted of a 3 hidden layers, 200 hidden layer nodes, a dropout rate of 0.6, no regularization, batch sizes of 150, using the rmsprop() algorithm with a learning rate of 0.001, and 100 predictors. As can be seen in the graph below, the number of variables needed to obtain near-peak performance, is much less than the full set.



The final selected model, after re-training, has a MSLE of 0.875 and RMSE of 15357. Comparing this model ('Full Neural Network') to the previous feature extraction models, which used many more variables, the performance is competitive. Additionally, the results were then multiplied by their respect gross floor area and then compared to a set of feature extraction models that were trained on total consumption. Again, it can be seen that this neural network model has shown to be competetive in this manner and, in fact, has a better Rsquared value.

The resdiuals indicate that the variance scales with the response variable; however, since neural network models do not operate on a principle of homoscedacitity, only underlying patterns are of concern. Additionally, the noted error pattern is by design so that higher error in higher consumption projects are acceptable. *Appendix*

Per SF Model Comparison

Model	RMSE	R2	MAE
partialLeastSquares	46880.41	0.5475726	24461.50
recursiveFeatureExtraction	48022.21	0.5168333	29511.04
neuralNetwork	48644.23	0.4990216	28964.80
Full Neural Network	50149.82	0.5007321	26153.83
leaps	90503.65	0.3149572	59157.08
randomForest	163434.65	0.1731248	141674.66

Per SF Model Comparison To Total Feature Extraction Models

Model	RMSE	R2	MAE
recursiveFeatureExtraction	10462.67	0.8305013	4168.206
partialLeastSquares	12109.87	0.7636348	3292.720
neuralNetwork	12337.32	0.7681178	3617.357
Full Neural Network	13152.85	0.7130382	3363.973
randomForest	17284.70	0.8087479	5495.435
leaps	18490.39	0.4527306	4865.344

Variable Selection Summary

The final model chosen uses 100 variables. Many of the features within the set have to do with the amount of receptacle equipment within the building as well as major electrical devices (e.g. MRI machines) and essential equipment (e.g. data center servers, refrigeration). Also, building type identifiers have been included for various categories. *Appendix*

The automated selection process does seem to have included some highly correlated pairs, such as the number of workers per square foot as well as the categorical bin of workers. This does reduce the number of necessary questions, but it is unclear if both are necessary and/or if they are possibly detrimental. Also, there are a number of questions that may be automatically known just based on the usage type as some questions do not apply to all buidlings. It is possible take the steps used in asking questions in the survey in order to build a live form that can automatically parse out the meaningful questions based on building type, which could reduce the need to enter a value for all the selected variables.

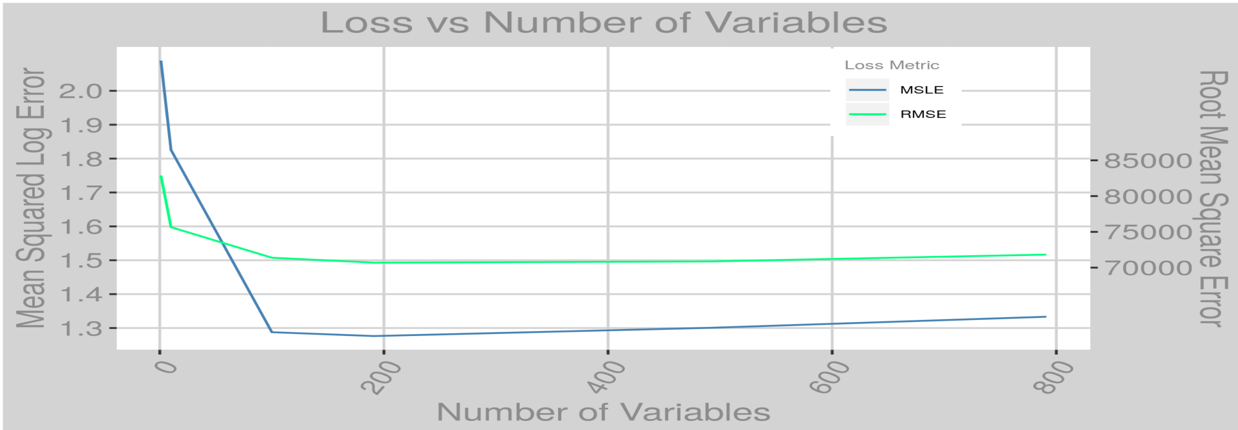
Natural Gas

Summary

The final selected model consisted of a 4 hidden layers, 400 hidden layer nodes, a dropout rate of 0.9, no regularization, batch sizes of 50, using the rmsprop() algorithm with a learning rate of 0.001, and 100 predictors. As can be seen in the graph below, the number of variables needed to obtain near-peak performance, is much less than the full set.

Variable Selection Summary

The final selected model, after re-training, has a MSLE of 0.875 and RMSE of 58528. Comparing this model ('Full Neural Network') to the previous feature extraction models, which used many



more variables, the performance is actuall better. Additionally, the results were then multiplied by their respect gross floor area and then compared to a set of feature extraction models that were trained on total consumption. Again, it can be seen that this neural network model is the best performing out of the set and has a better Rsquared value than the per SF model. Much like the electricity data set, there appears to be heteroscedacicity in the residuals, but may be due to the selected loss function.

Per SF Model Comparison

Model	RMSE	R2	MAE
Full Neural Network	58258.14	0.3653484	29697.14
recursiveFeatureExtraction	59212.72	0.5230241	34203.37
neuralNetwork	69784.67	0.3336745	35642.13
partialLeastSquares	73075.99	0.2934536	32068.61
leaps	100385.86	0.2462162	53658.33
randomForest	175808.23	0.1930417	147144.46

Per SF Model Comparison To Total Feature Extraction Models

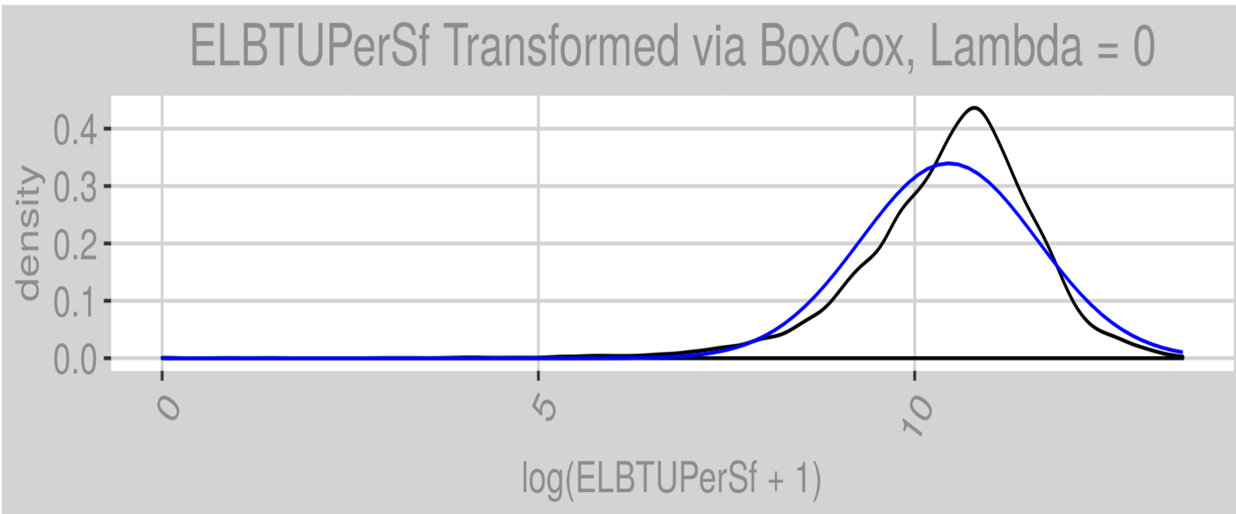
Model	RMSE	R2	MAE
Full Neural Network	10974.00	0.7863814	3773.057
recursiveFeatureExtraction	23757.53	0.6354168	6035.513
partialLeastSquares	25929.37	0.5476751	5022.284
randomForest	27500.11	0.7483615	5866.231
neuralNetwork	27866.35	0.4738025	5798.053
leaps	32968.18	0.2655183	6076.418

Future Work

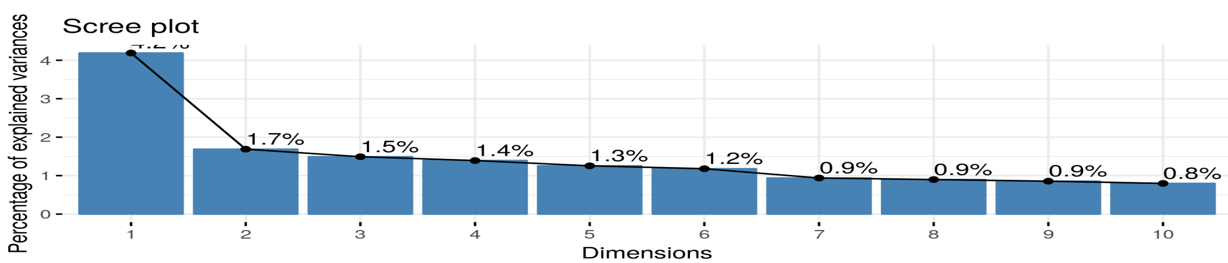
While it was determined that some heteroscedacitity would be acceptable, there does appear to be area for improvement. Additionally, as mentioned at the beginning of this report, the sampling of this data set was stratified to reflect the building population. However, it is noted that there are some building classes that have greater variance than others. Therefore, it may be useful to use this stratification as a weighted method, based on `PBAPLUS`, in order to try and emphasize accuracy on the most prevalent budiding types. Also, there was not a lot of attention paid to the actual transformations of the predictors given the large quantity of them. It is possible a better fit can be obtained with more intelligent transformations applied to the features after further analysis.

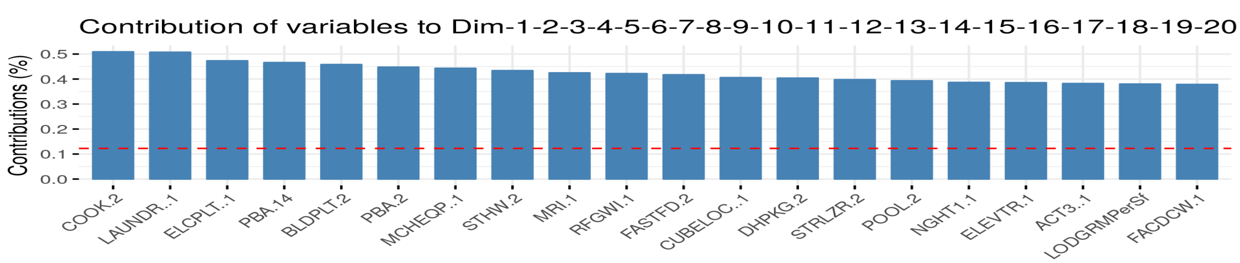
Appendix - Electricity

Response

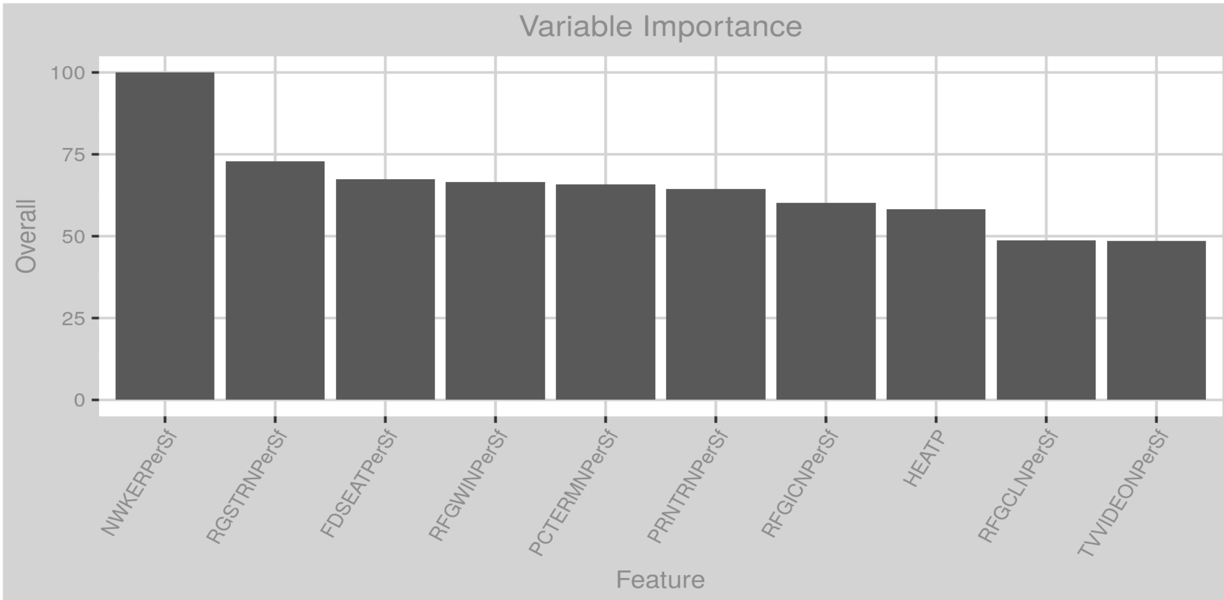


PCA

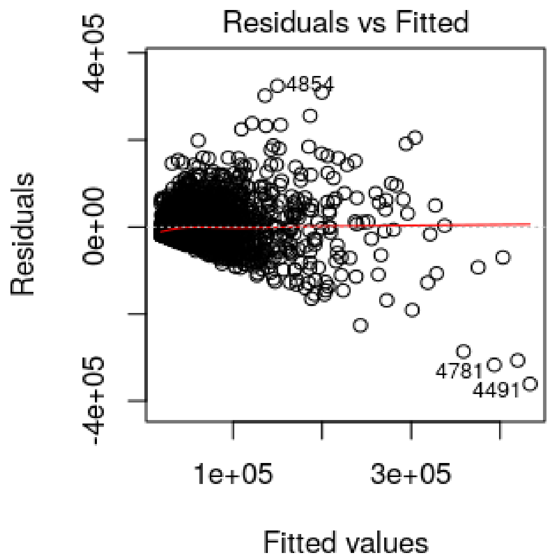




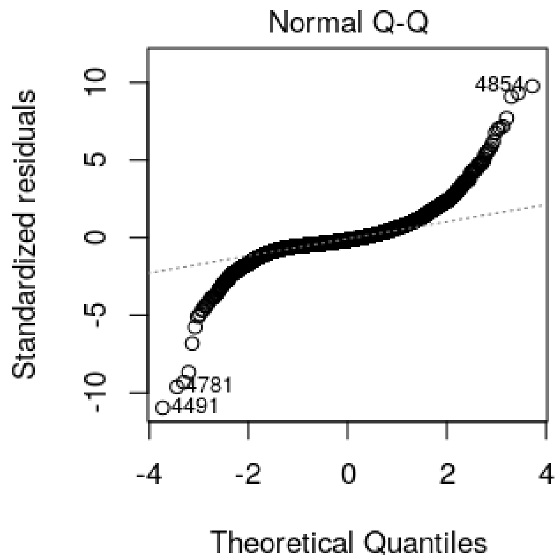
PLS



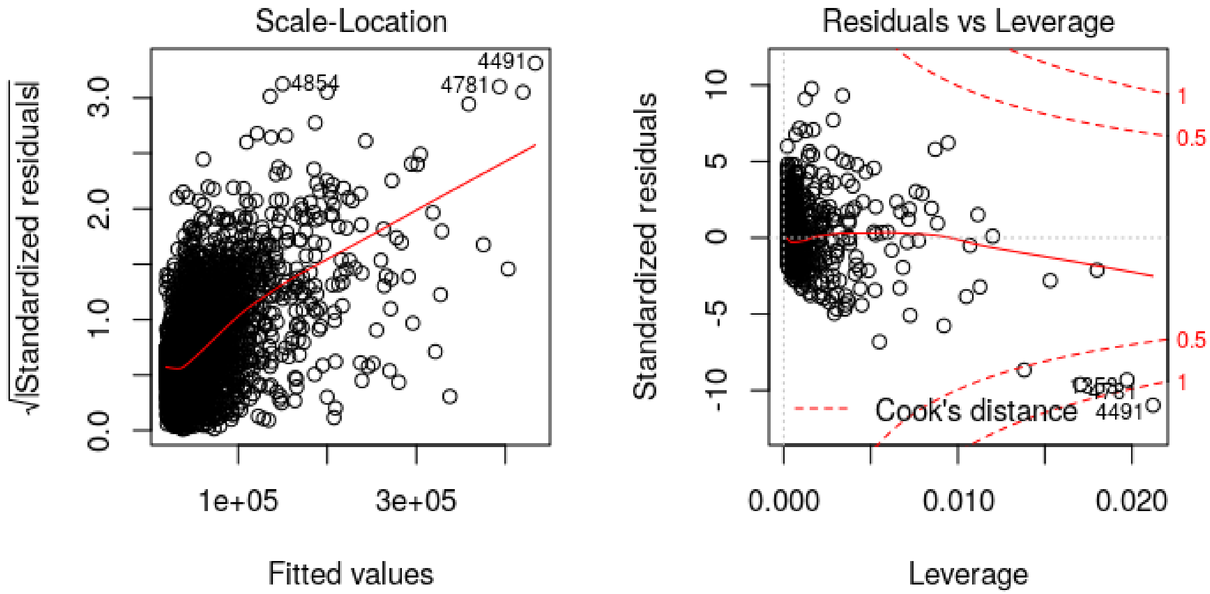
Variable Importance
Overall
100
75
50
25
0
NWKERPerSf
RGSTRNPerSf
FDSEATPerSf
RFGWINPerSf
PCTERMNPerSf
PRNTRNPerSf
RFGICNPerSf
HEATP
RFGCLNPerSf
TVVIDEONPerSf
Feature

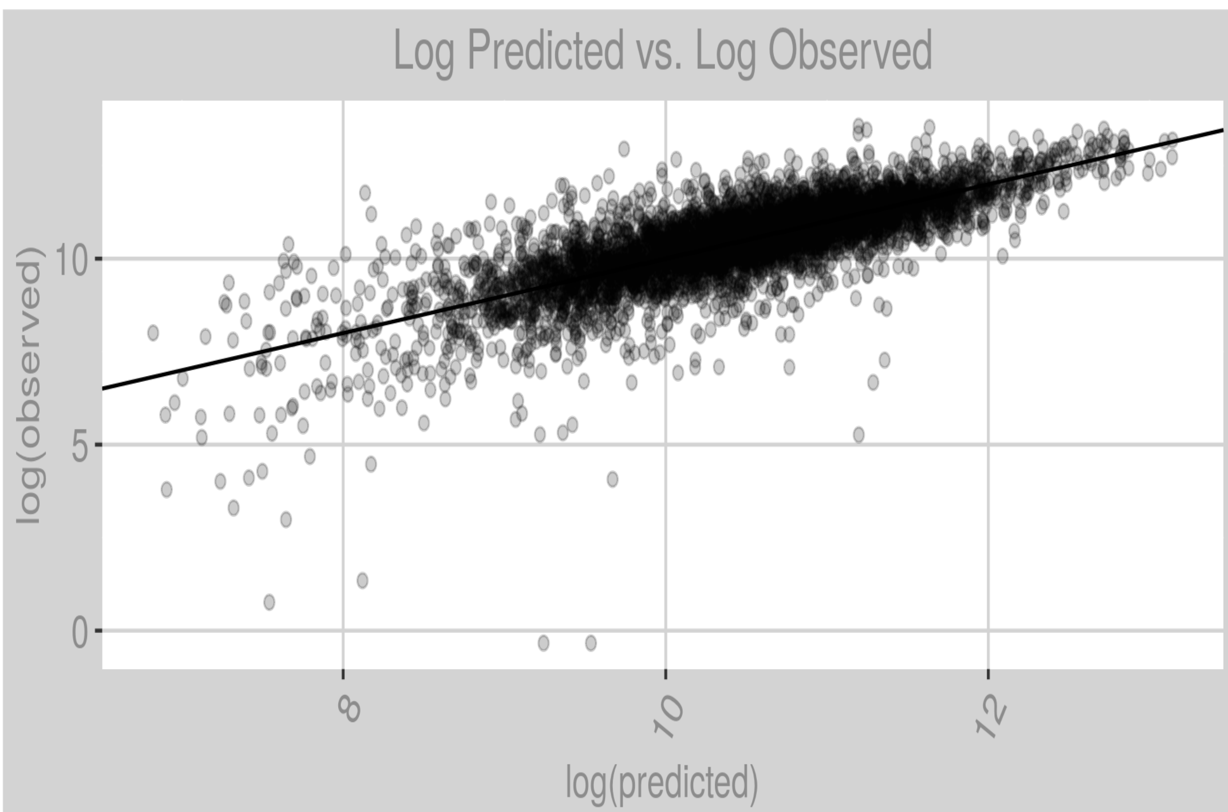


Residuals vs Fitted
Residuals
4e+05
0e+00
-4e+05
4854
4781
4491
1e+05
3e+05
Fitted values

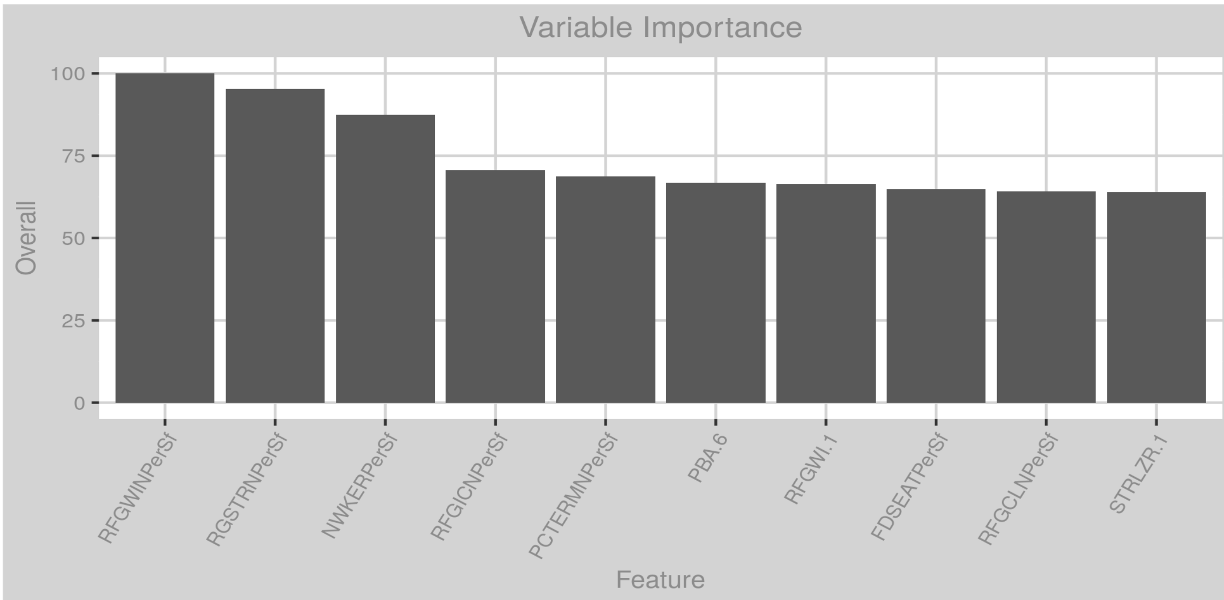


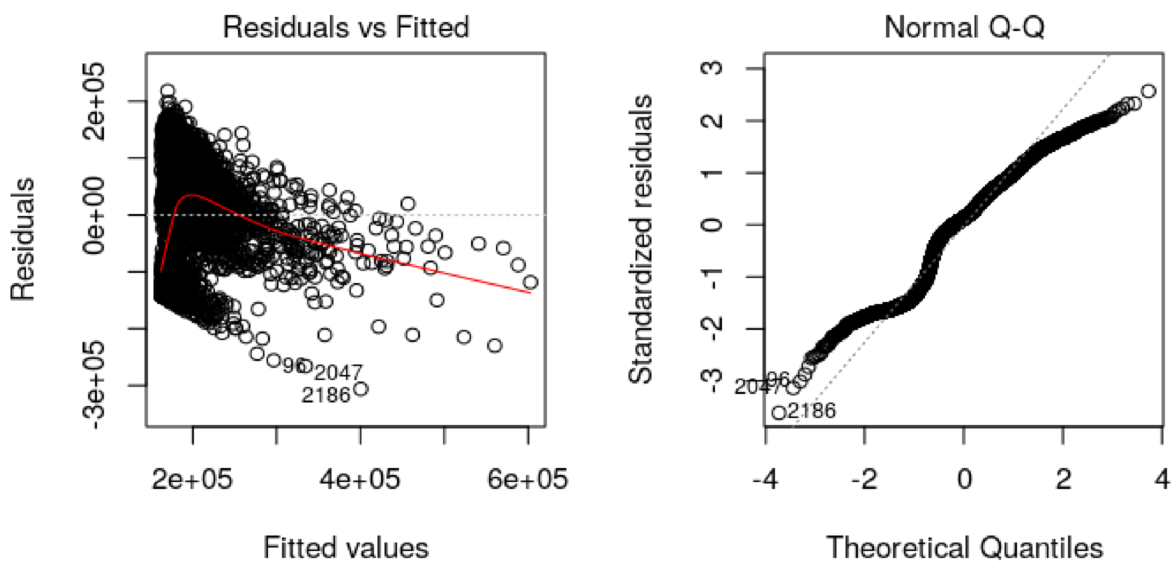
Normal Q-Q
Standardized residuals
10
5
0
-5
-10
4854
4781
4491
-4
-2
0
2
4
Theoretical Quantiles

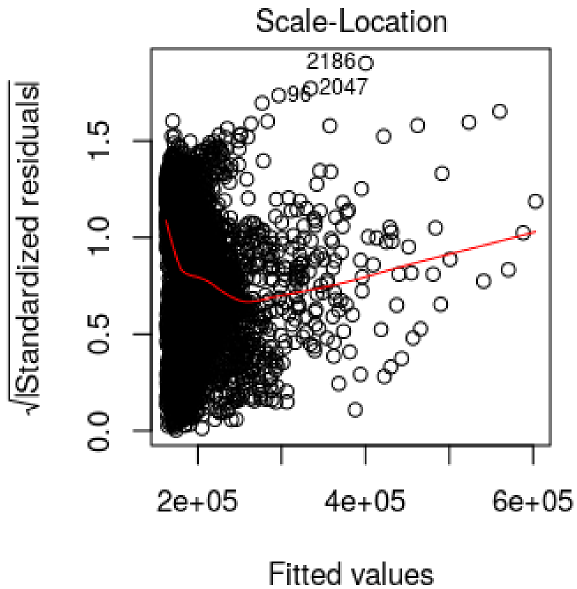

Scale-Location
√|Standardized residuals|
3.0
2.0
1.0
0.0
4491
4854
4781
1e+05
3e+05
Fitted values
Residuals vs Leverage
Standardized residuals
10
5
0
-5
-10
1
0.5
Cook's distance
4491
0.000
0.010
0.020
Leverage

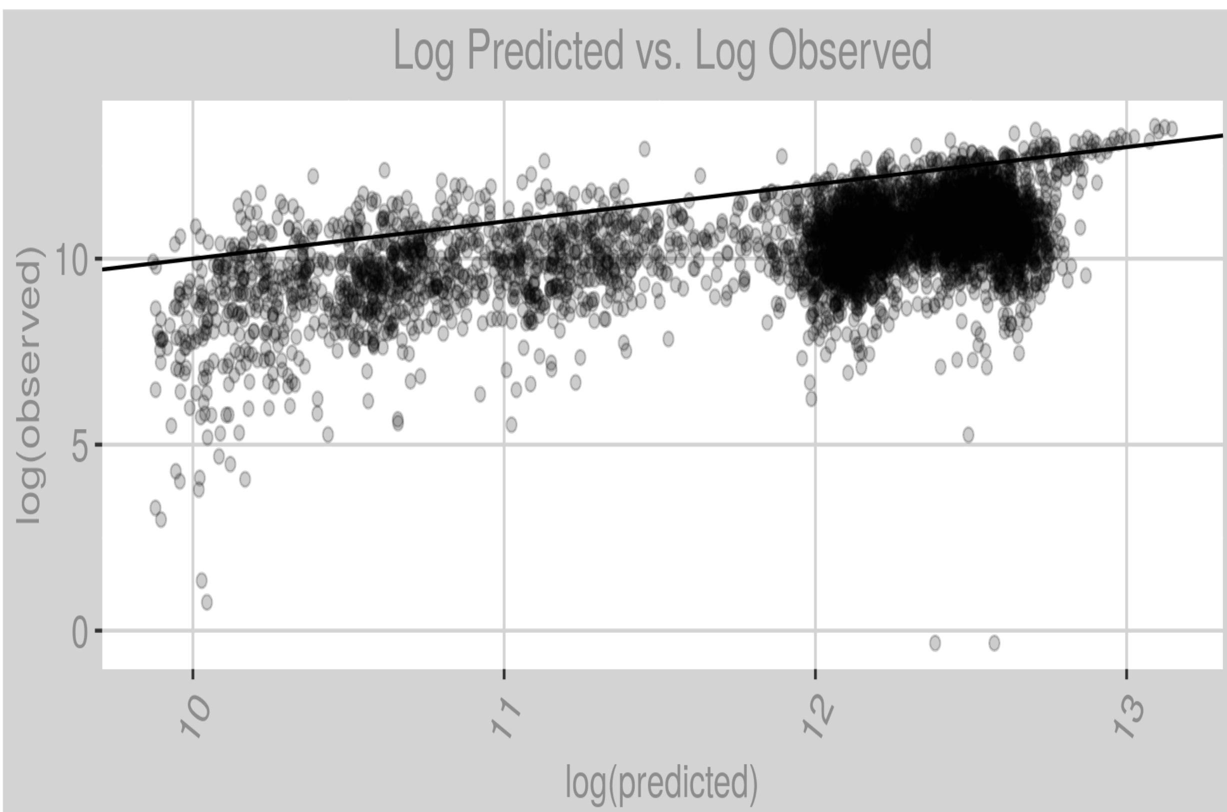

Log Predicted vs. Log Observed
log(observed)
10
5
0
8
10
12
log(predicted)

Random Forest

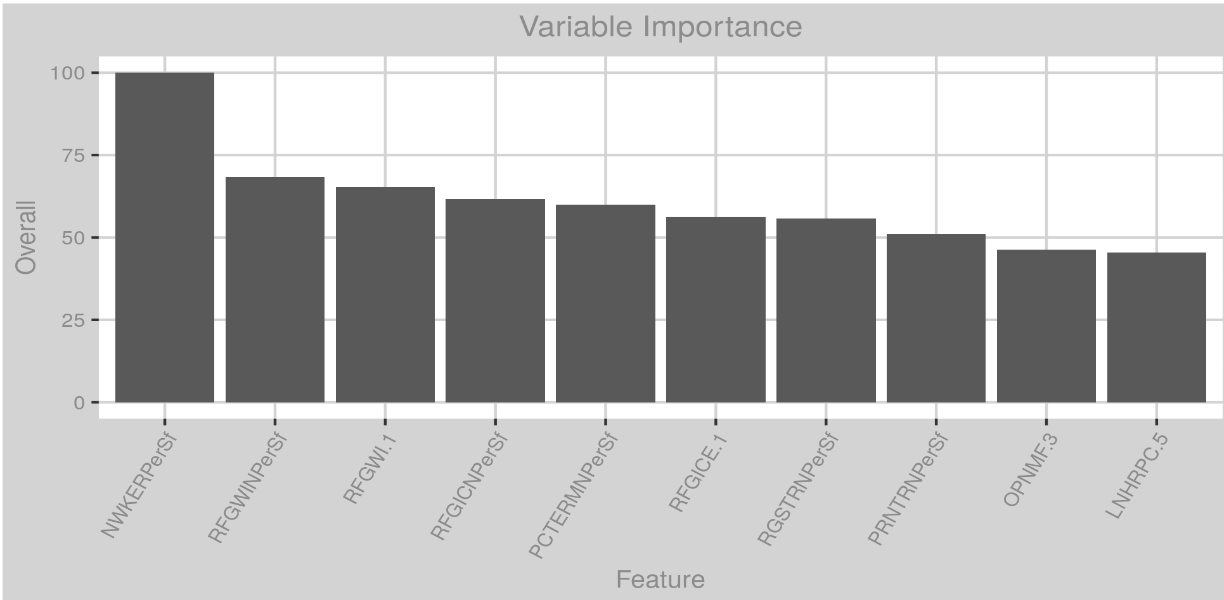


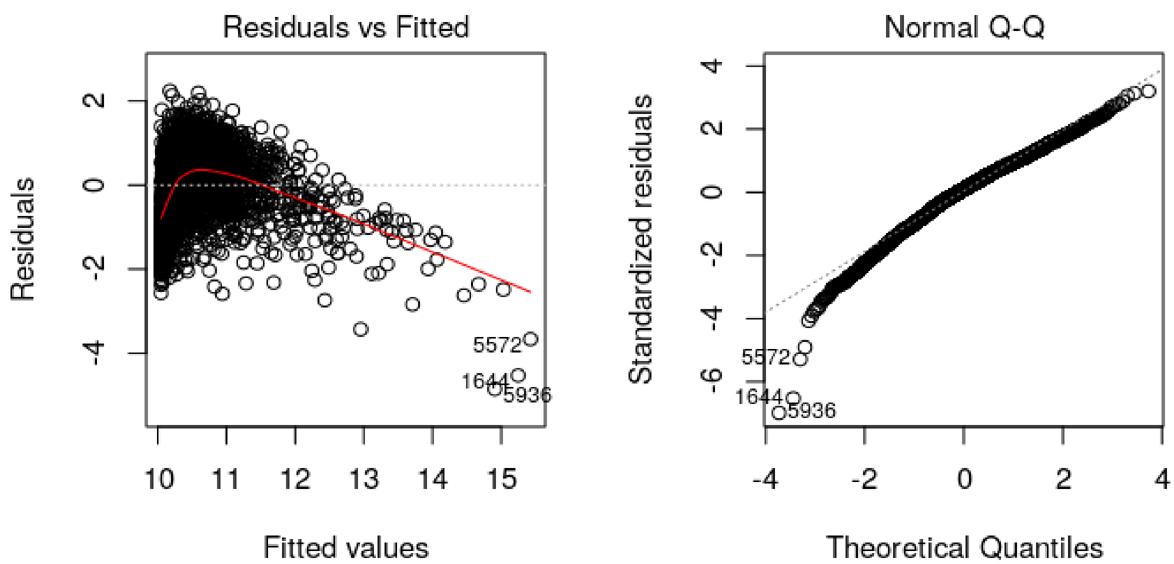


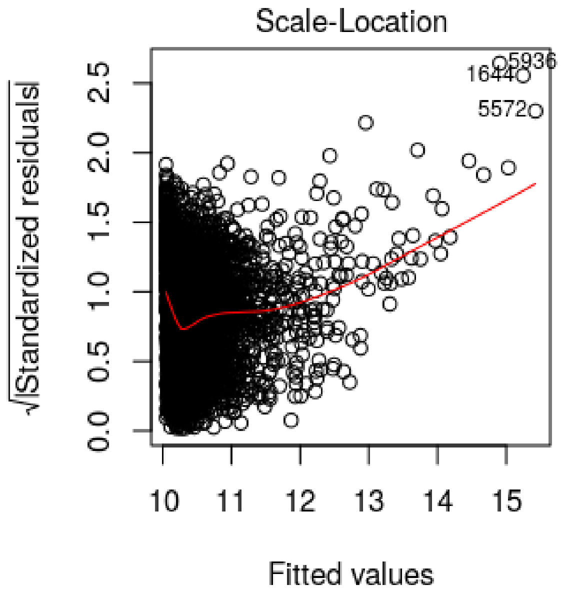

Scale-Location
2186
96
2047
√|Standardized residuals|
0.0
0.5
1.0
1.5
2e+05
4e+05
6e+05
Fitted values
Residuals vs Leverage
Standardized residuals
3
2
1
0
-2
-4
4404
5936
Cook's distance
0.000
0.010
0.020
Leverage

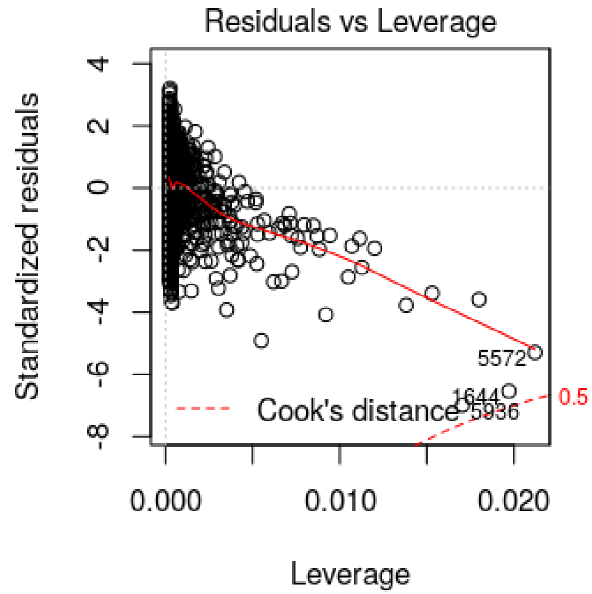

Log Predicted vs. Log Observed
log(observed)
10
5
0
10
11
12
13
log(predicted)

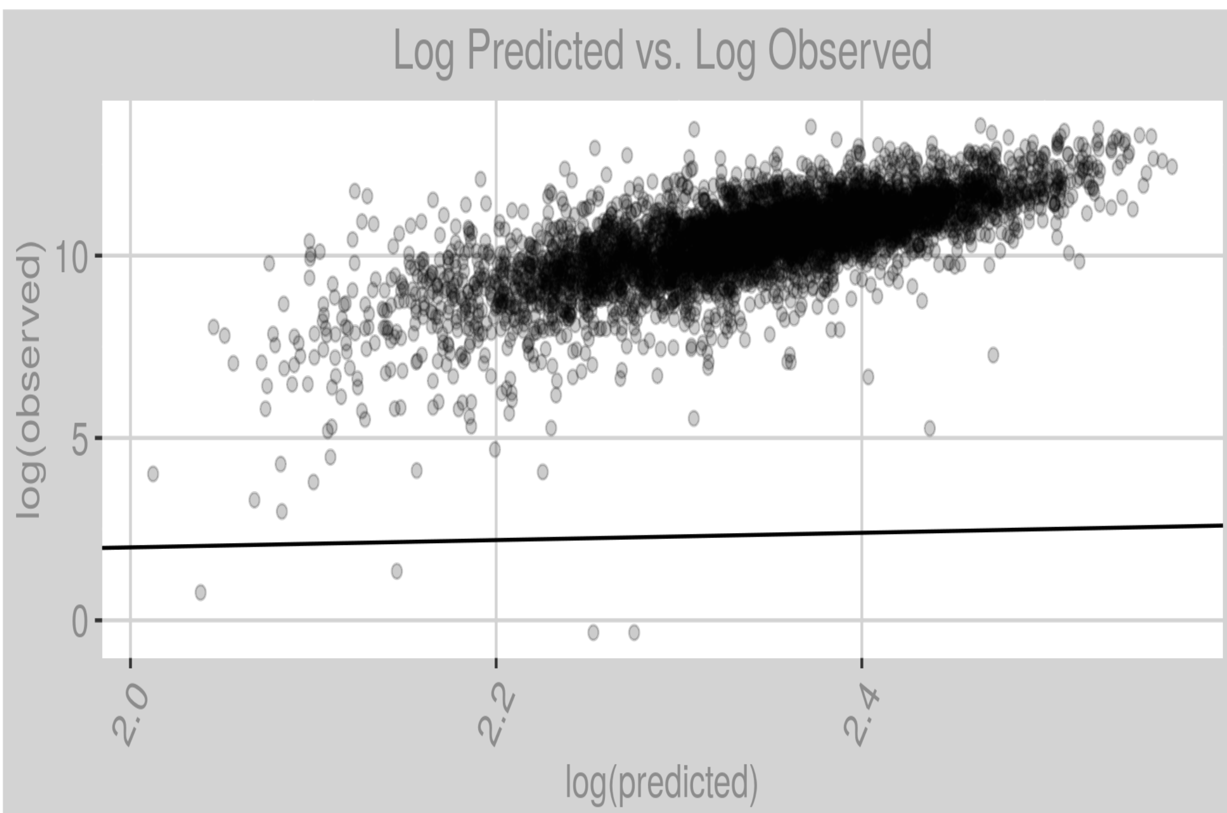
Forward Selection



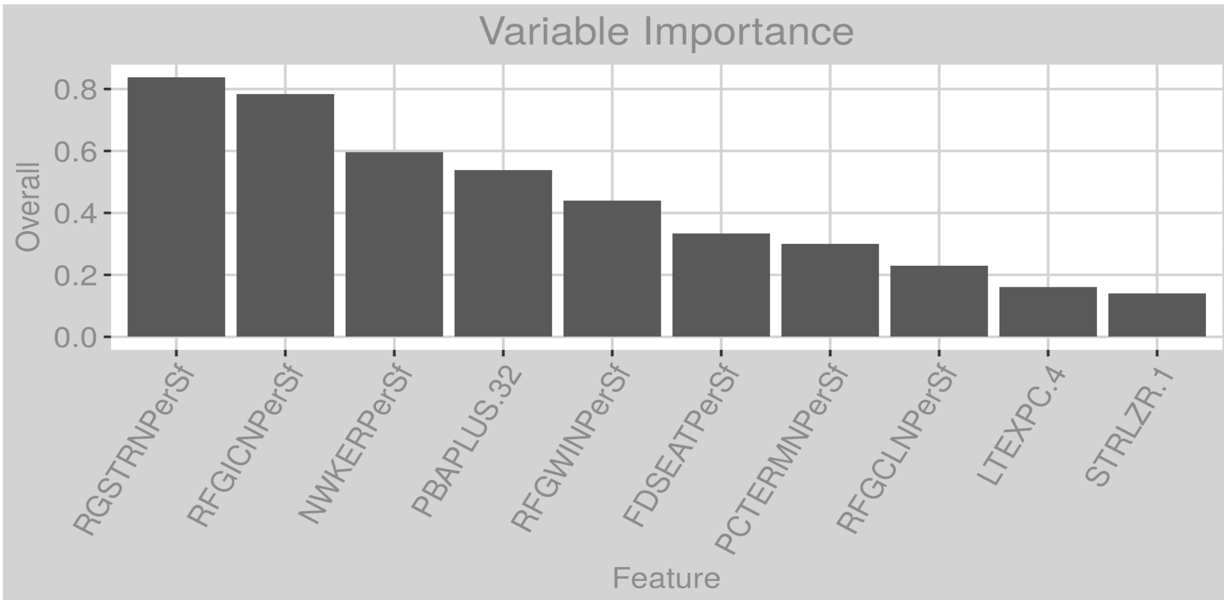


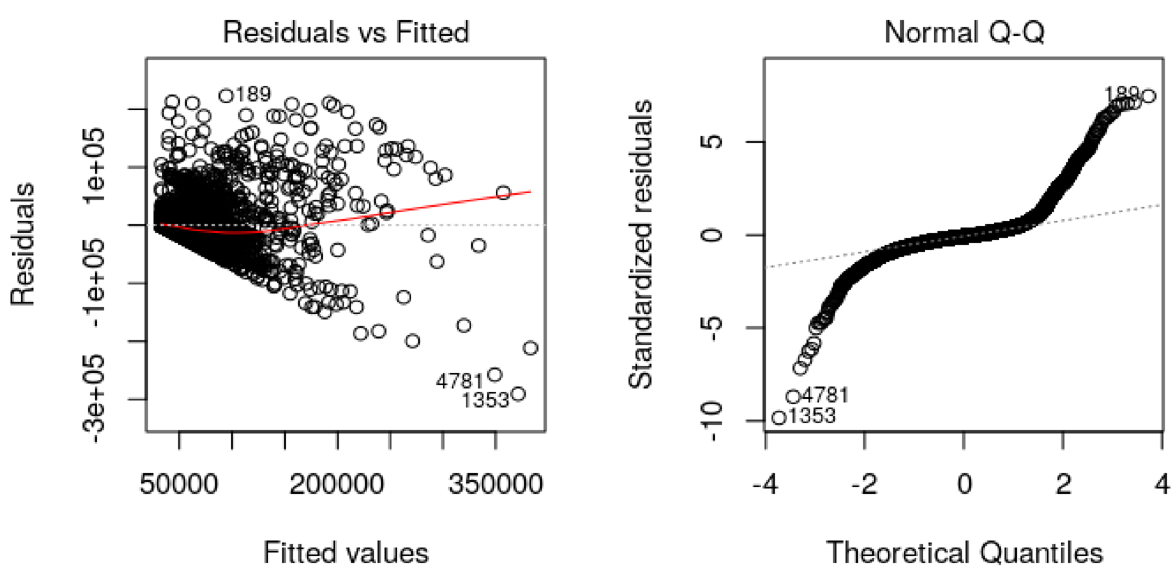

Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
10
11
12
13
14
15
Fitted values
5936
1644
5572

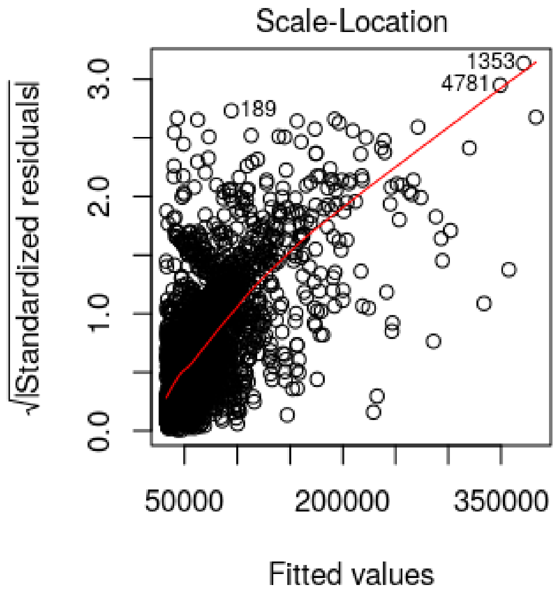

Residuals vs Leverage
Standardized residuals
4
2
0
-2
-4
-6
-8
0.000
0.010
0.020
Leverage
Cook's distance
0.5
5572
1644
5936


Log Predicted vs. Log Observed
log(observed)
10
5
0
2.0
2.2
2.4
log(predicted)

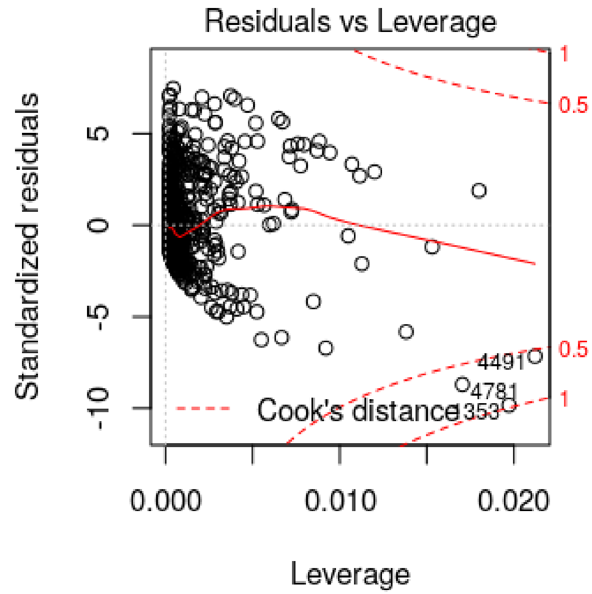
Recursive Feature Extraction



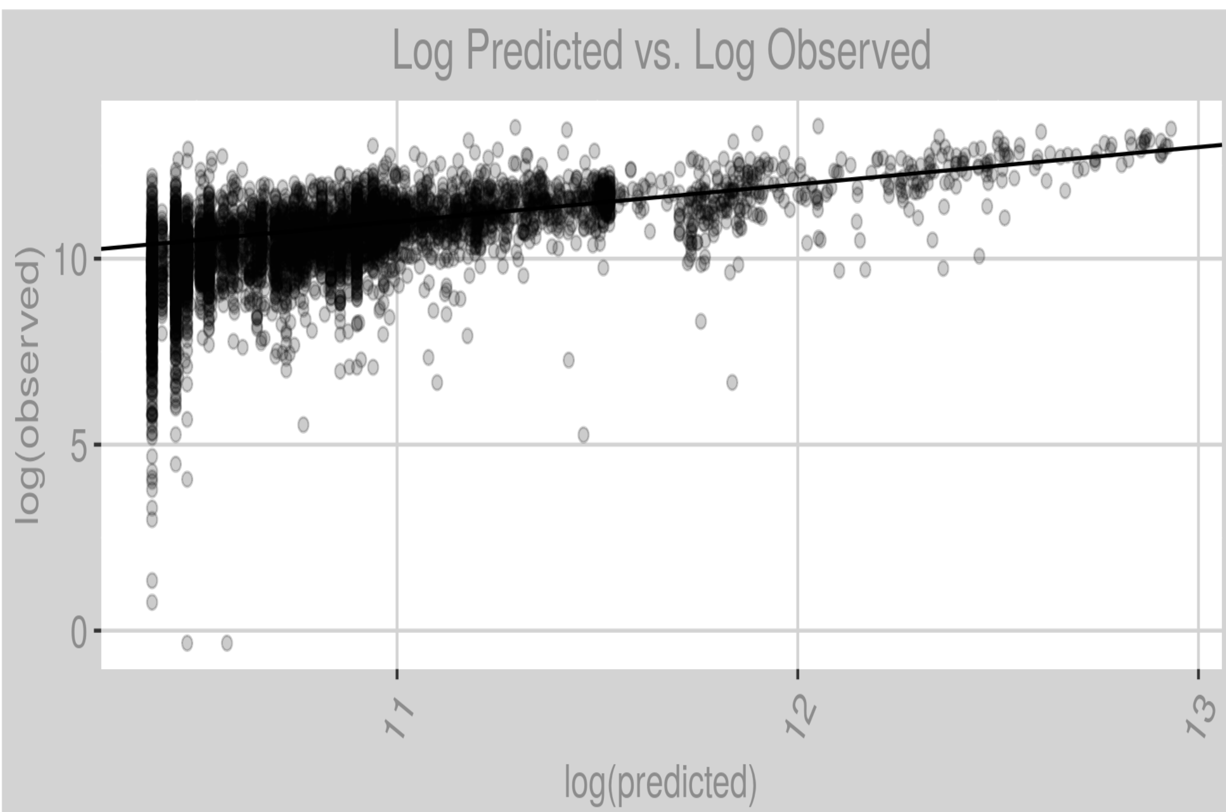




Scale-Location
√|Standardized residuals|
0.0
1.0
2.0
3.0
1353
4781
189
50000
200000
350000
Fitted values

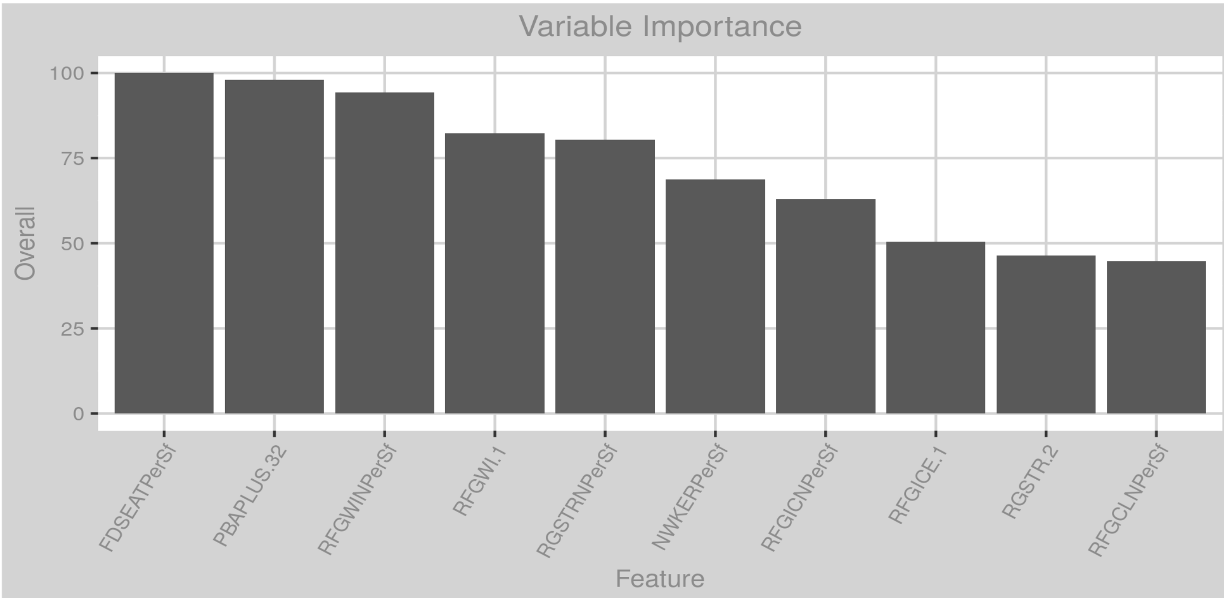


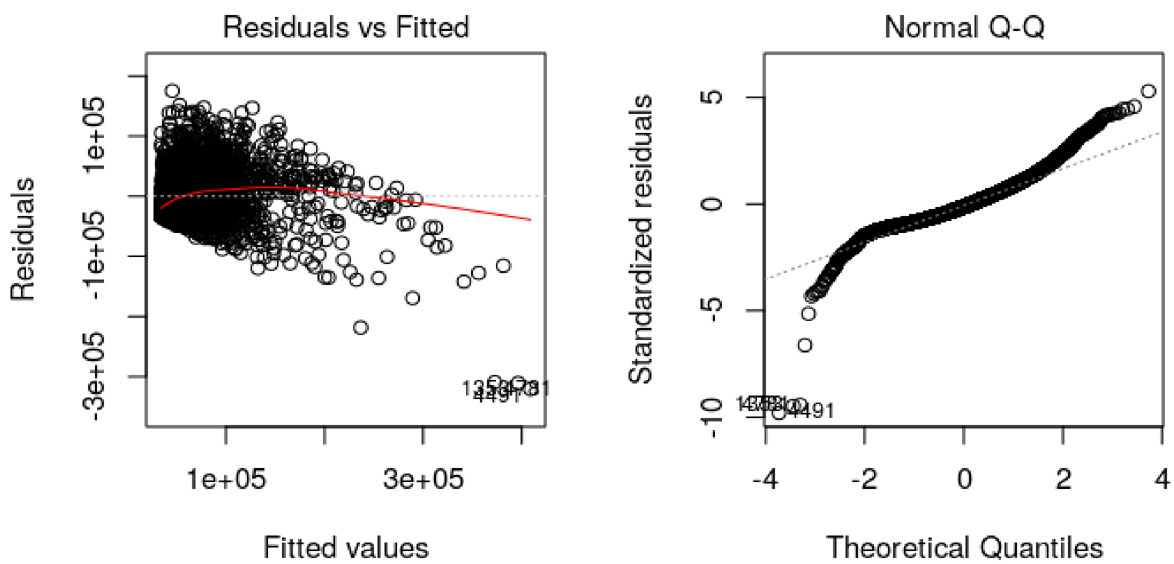
Residuals vs Leverage
Standardized residuals
-10
-5
0
5
1
0.5
0.5
1
4491
4781
1353
Cook's distance
0.000
0.010
0.020
Leverage

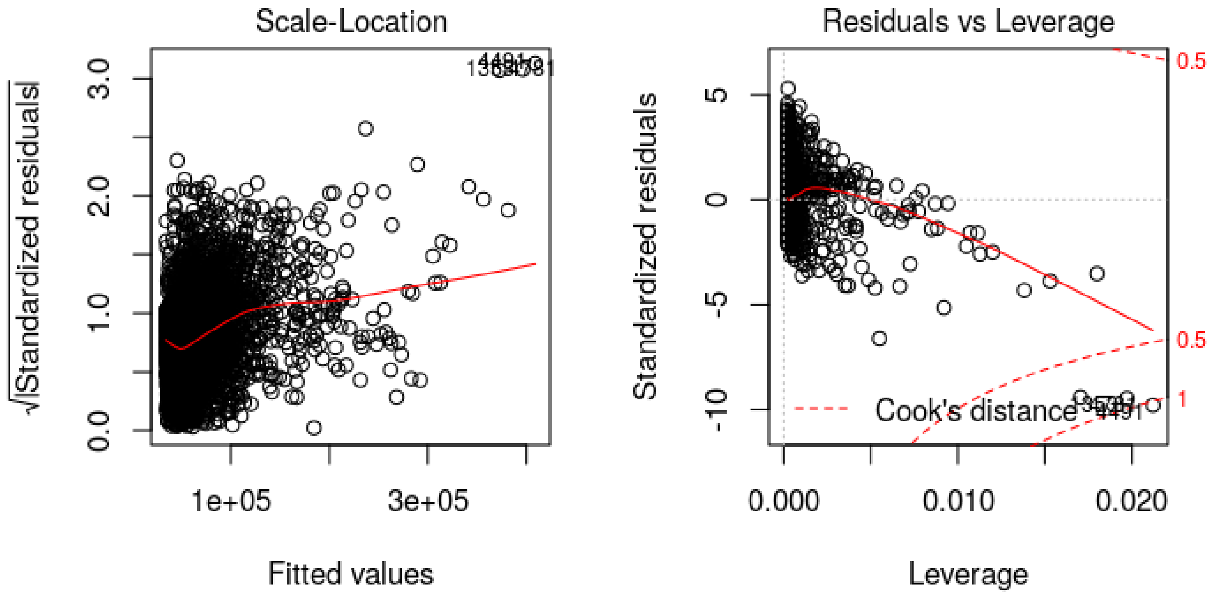


Log Predicted vs. Log Observed
log(observed)
0
5
10
11
12
13
log(predicted)

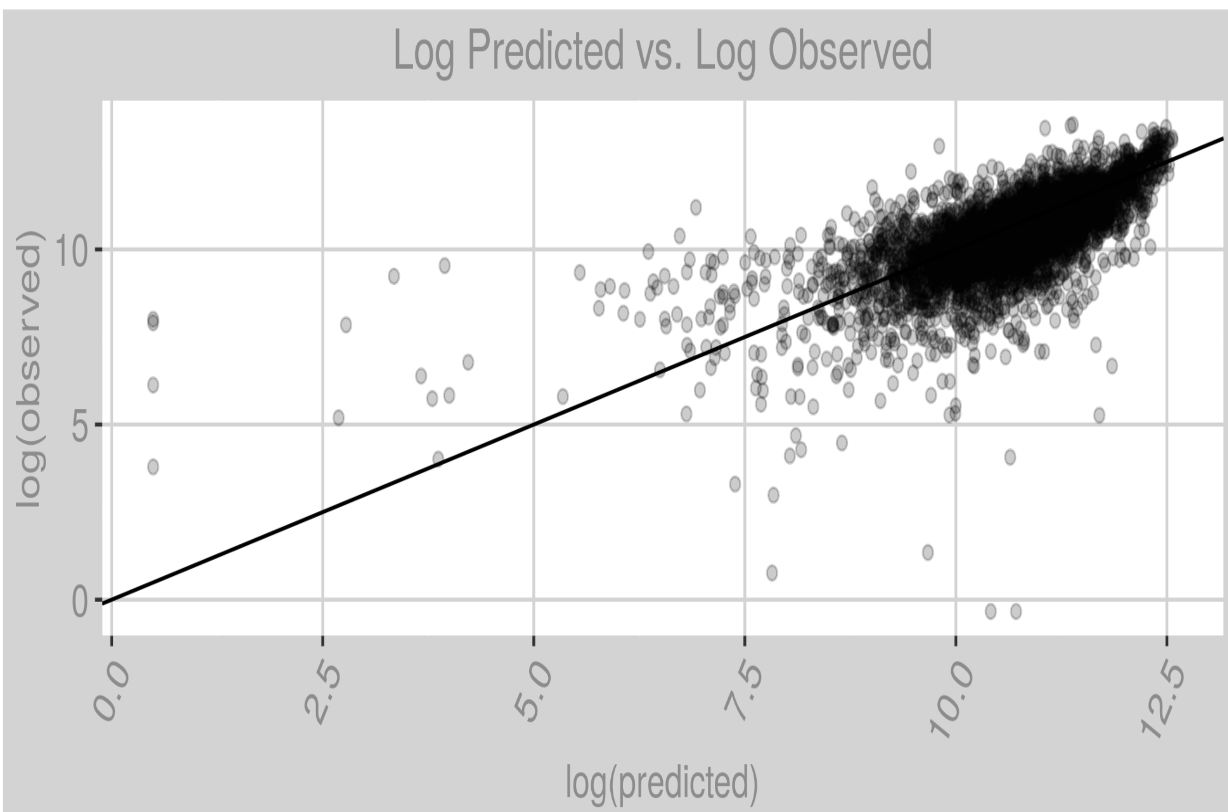
Simple Neural Network







Scale-Location
√|Standardized residuals|
0.0
1.0
2.0
3.0
1e+05
3e+05
Fitted values
Residuals vs Leverage
Standardized residuals
-10
-5
0
5
0.000
0.010
0.020
Leverage
Cook's distance
0.5
0.5
1

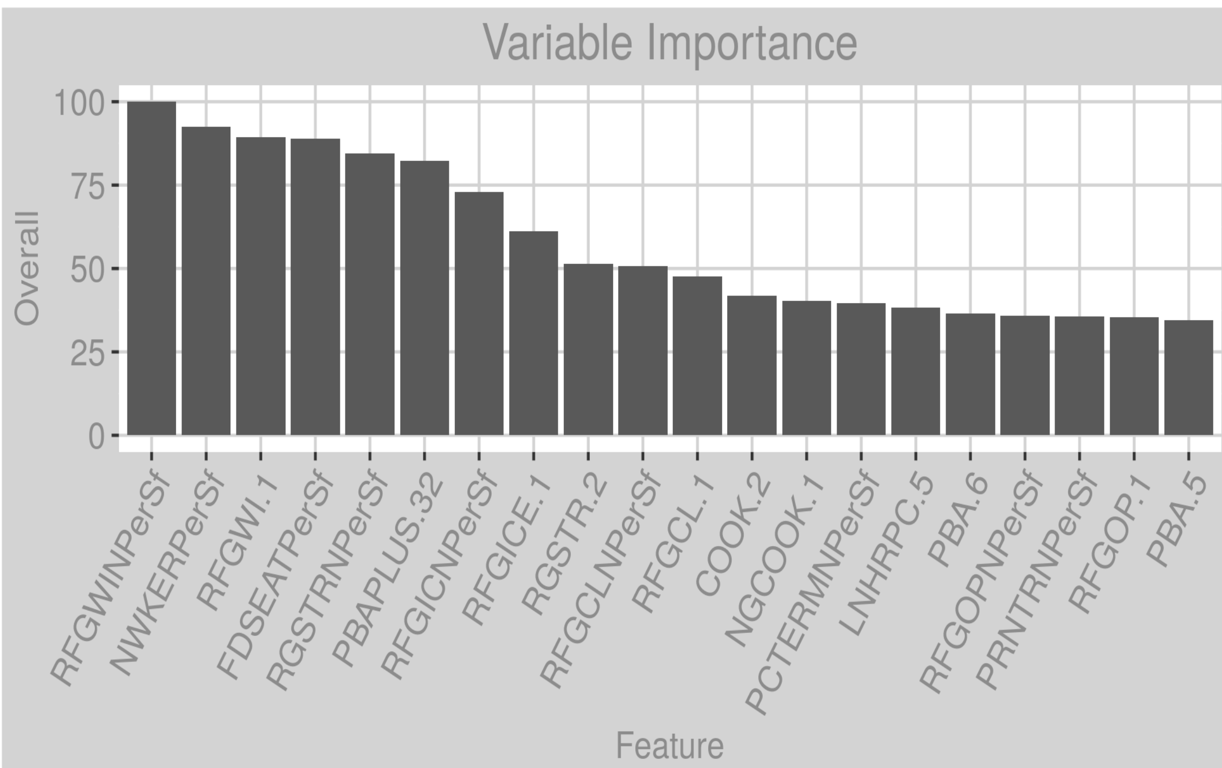


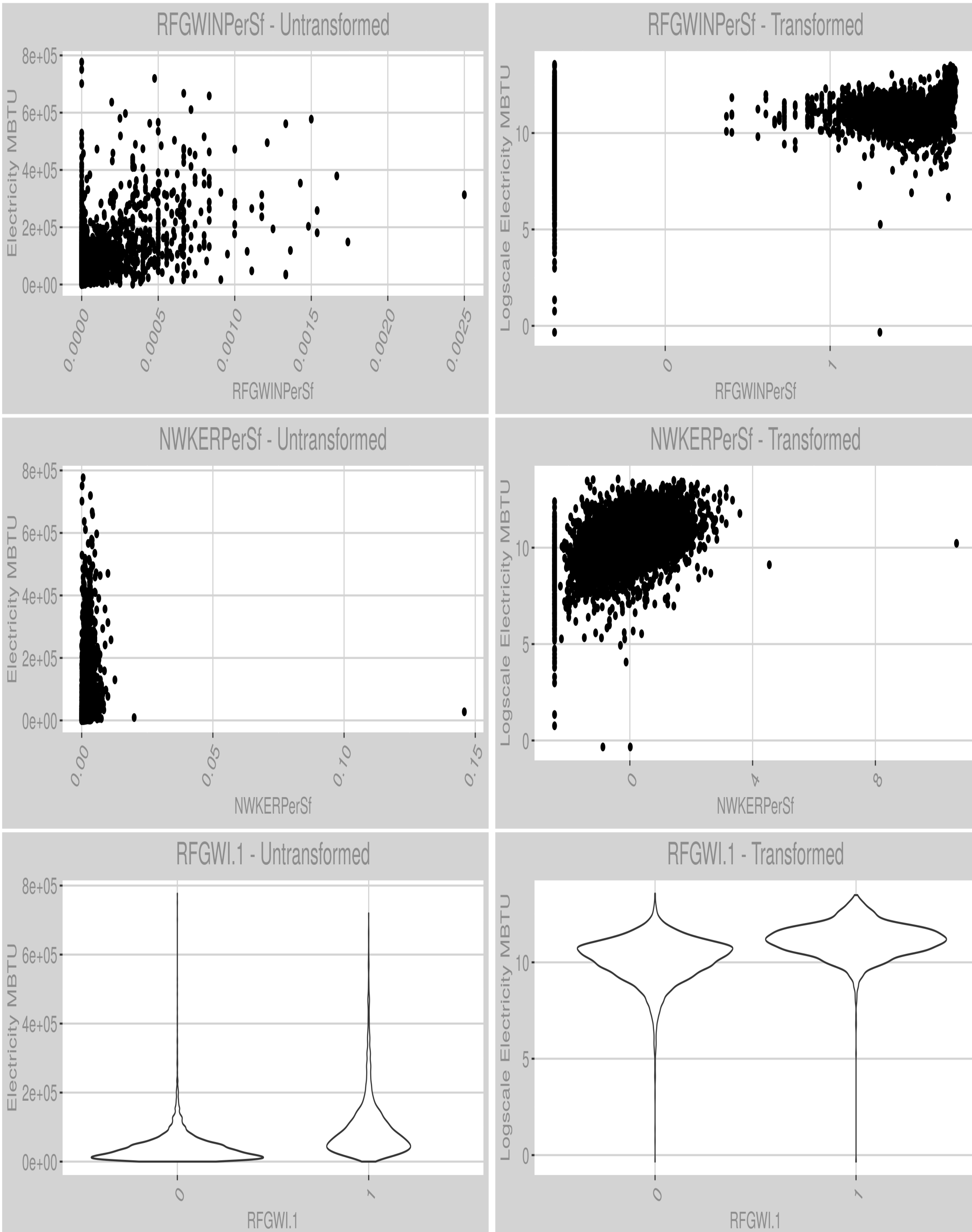
Log Predicted vs. Log Observed
log(observed)
0
5
10
0.0
2.5
5.0
7.5
10.0
12.5
log(predicted)

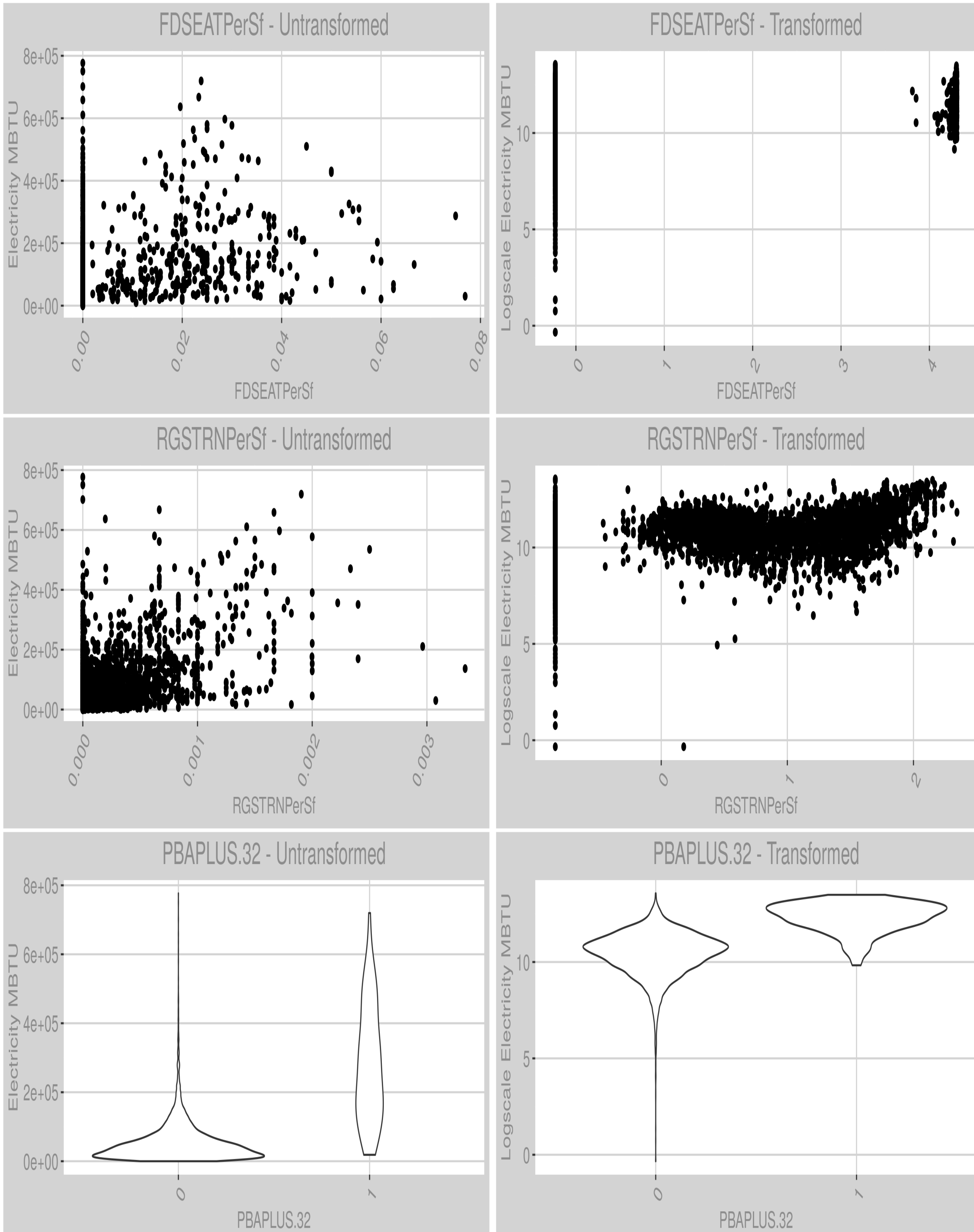
Select Variable Analysis

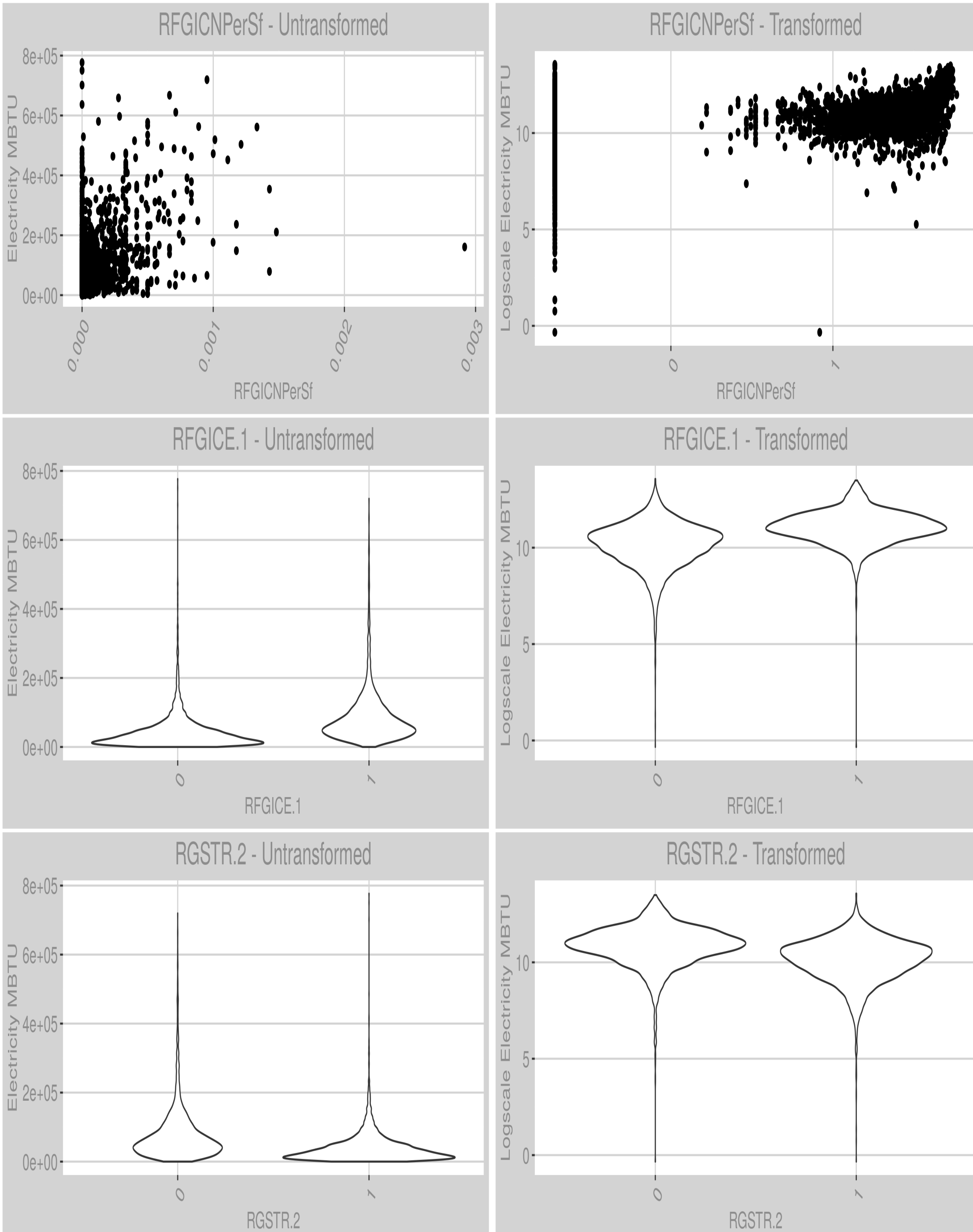
Feature Extraction Model Results

Model	RMSE	R2	MAE
partialLeastSquares	46880.41	0.5475726	24461.50
recursiveFeatureExtraction	48022.21	0.5168333	29511.04
neuralNetwork	48644.23	0.4990216	28964.80
leaps	90503.65	0.3149572	59157.08
randomForest	163434.65	0.1731248	141674.66

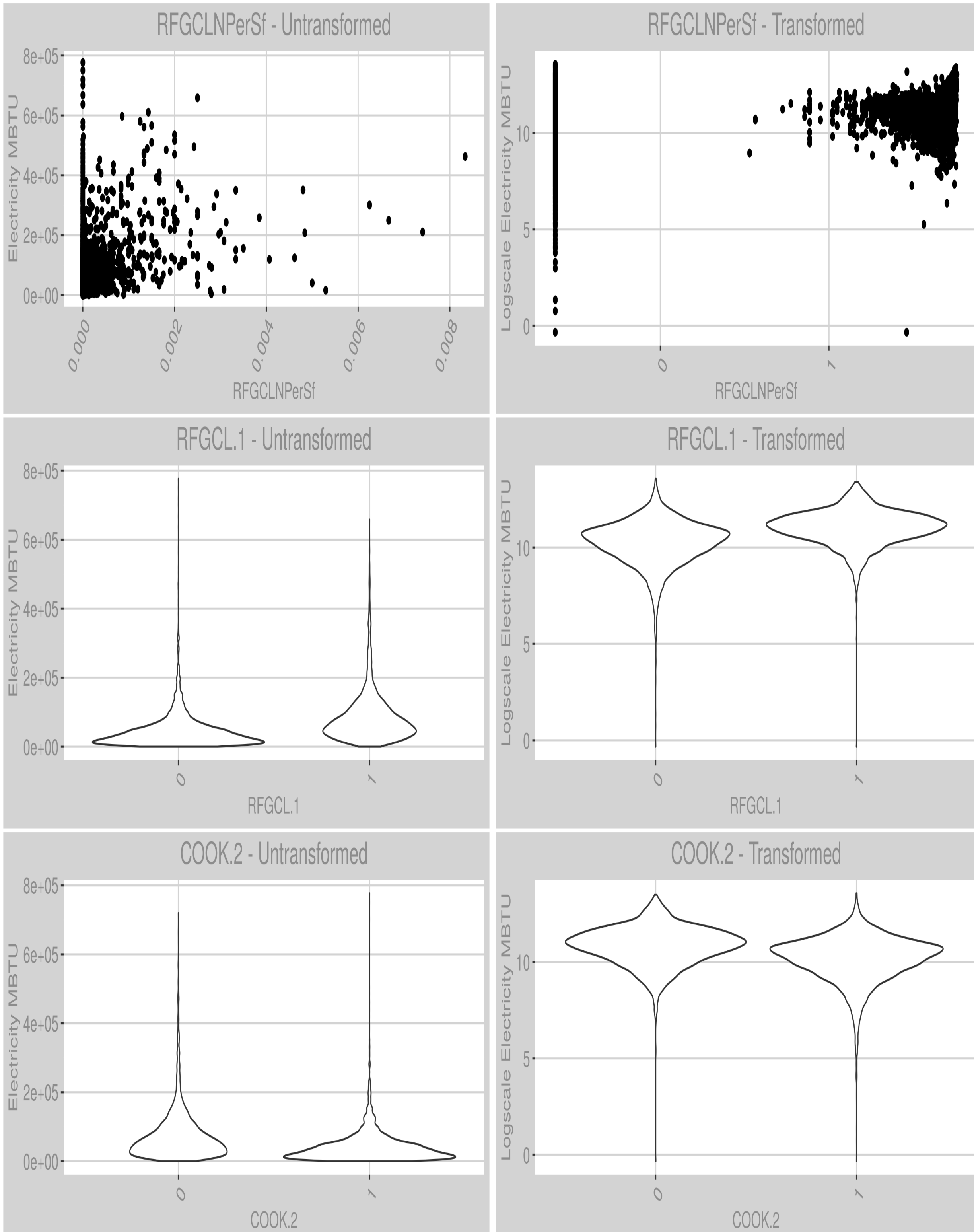



RFGWINPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
RFGWINPerSf
RFGWINPerSf - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RFGWINPerSf
NWKERPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.00
0.05
0.10
0.15
NWKERPerSf
NWKERPerSf - Transformed
Logscale Electricity MBTU
10
5
0
0
4
8
NWKERPerSf
RFGWI.1 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
RFGWI.1
RFGWI.1 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RFGWI.1

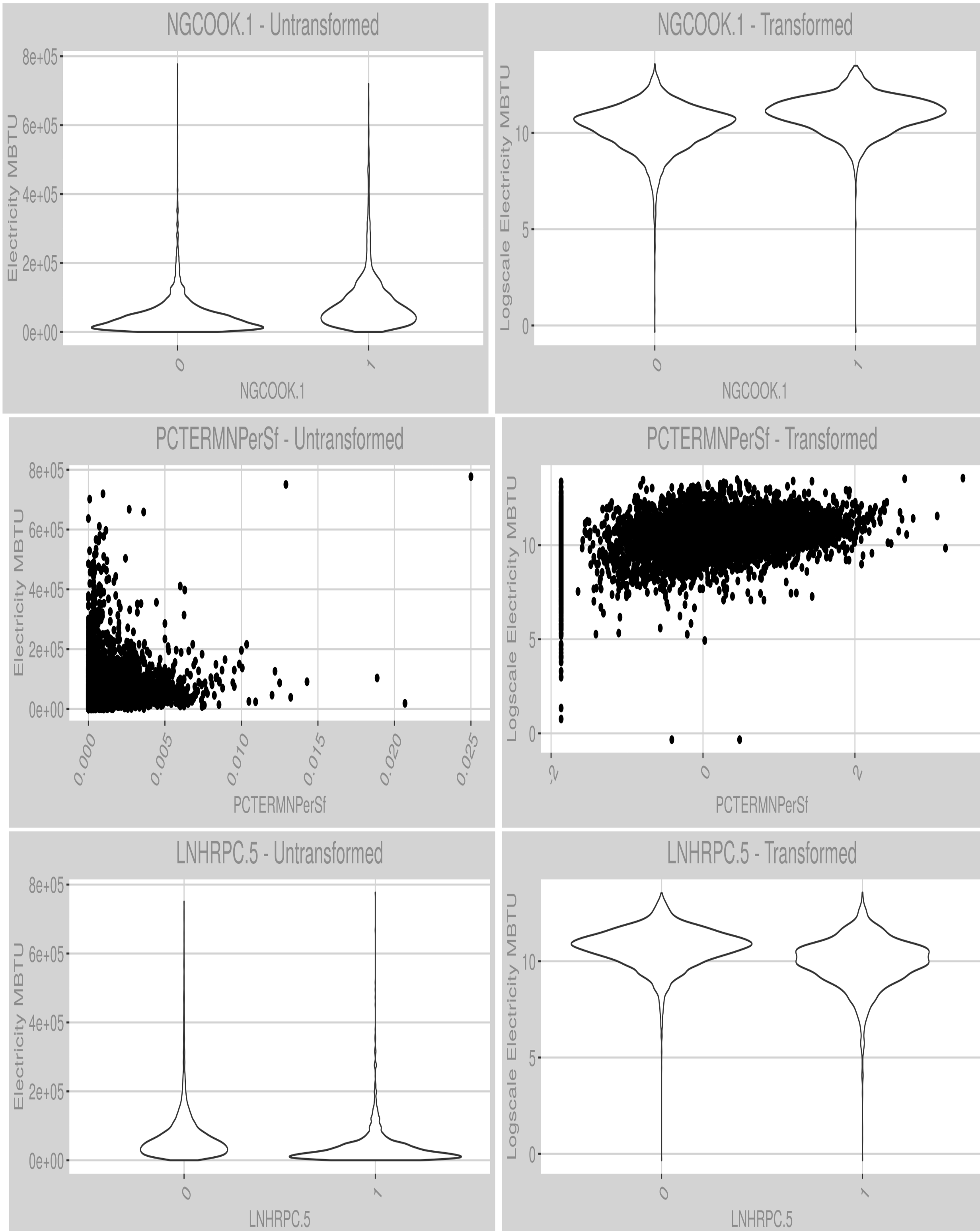

FDSEATPerSf - Untransformed
Electricity MBTU
FDSEATPerSf
FDSEATPerSf - Transformed
Logscale Electricity MBTU
FDSEATPerSf
RGSTRNPerSf - Untransformed
Electricity MBTU
RGSTRNPerSf
RGSTRNPerSf - Transformed
Logscale Electricity MBTU
RGSTRNPerSf
PBAPLUS.32 - Untransformed
Electricity MBTU
PBAPLUS.32
PBAPLUS.32 - Transformed
Logscale Electricity MBTU
PBAPLUS.32

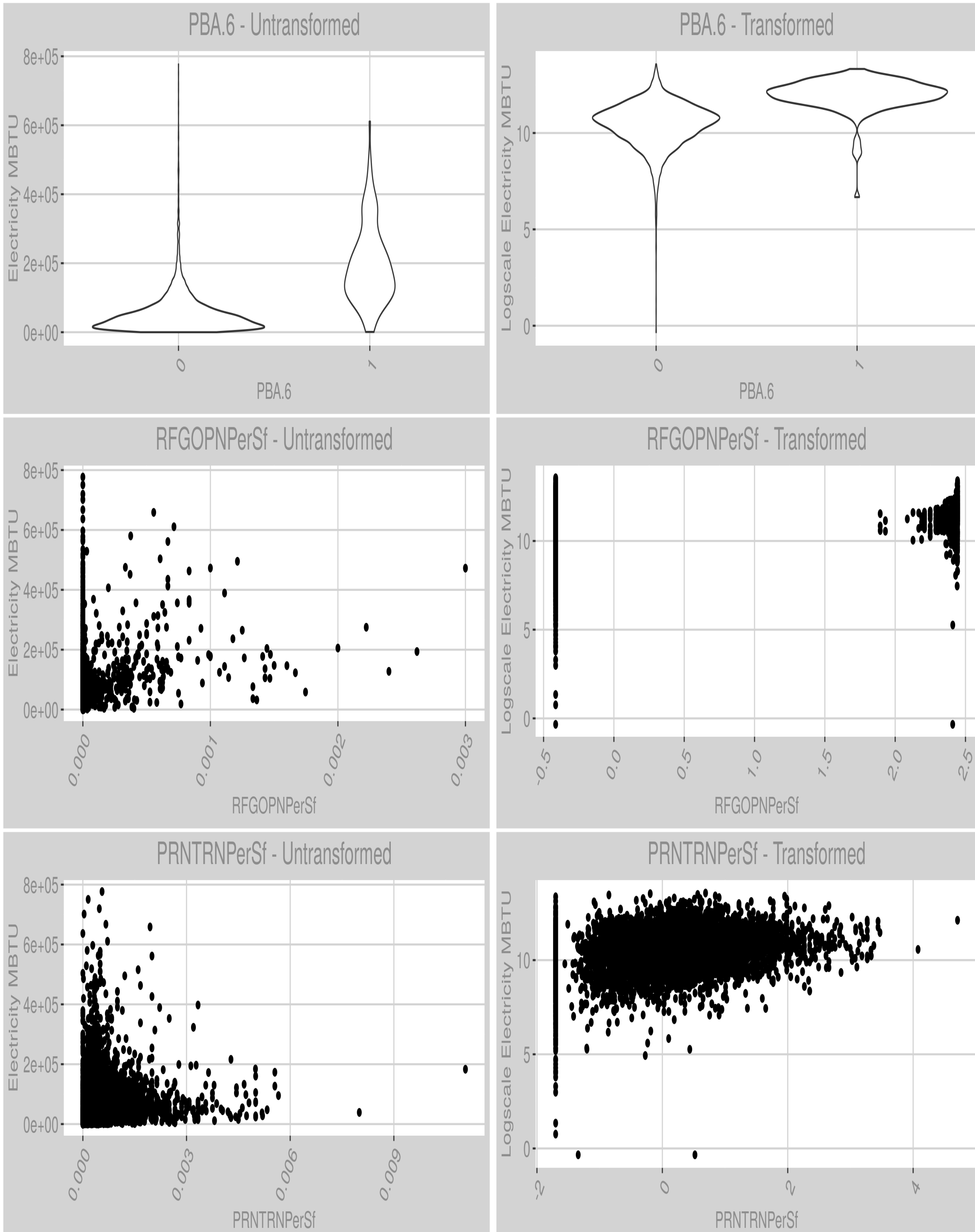


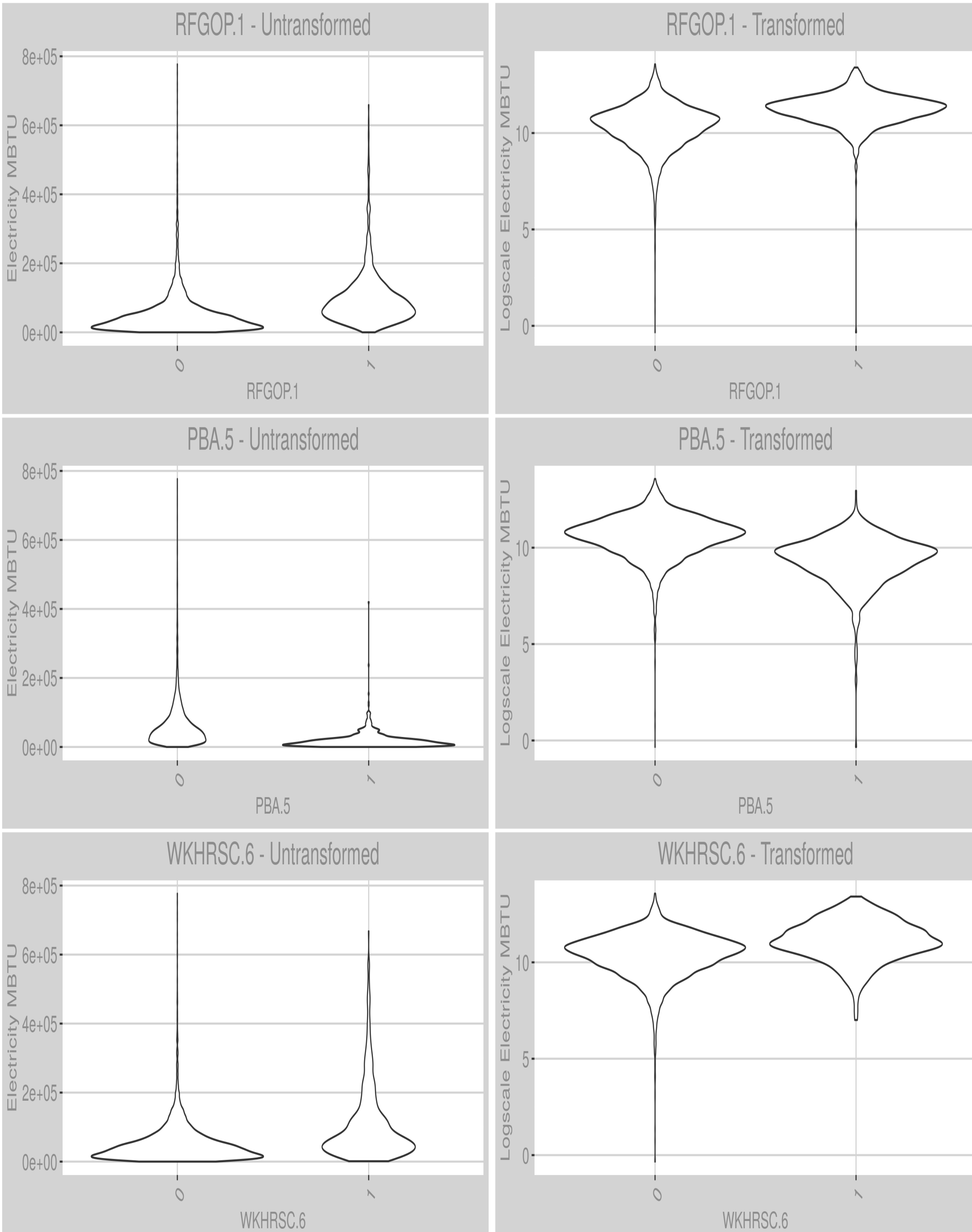
RFGICNPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.000
0.001
0.002
0.003
RFGICNPerSf
RFGICNPerSf - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RFGICNPerSf
RFGICE.1 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
RFGICE.1
RFGICE.1 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RFGICE.1
RGSTR.2 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
RGSTR.2
RGSTR.2 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RGSTR.2



RFGCLNPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.000
0.002
0.004
0.006
0.008
RFGCLNPerSf
RFGCLNPerSf - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RFGCLNPerSf
RFGCL.1 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
RFGCL.1
RFGCL.1 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
RFGCL.1
COOK.2 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
COOK.2
COOK.2 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
COOK.2

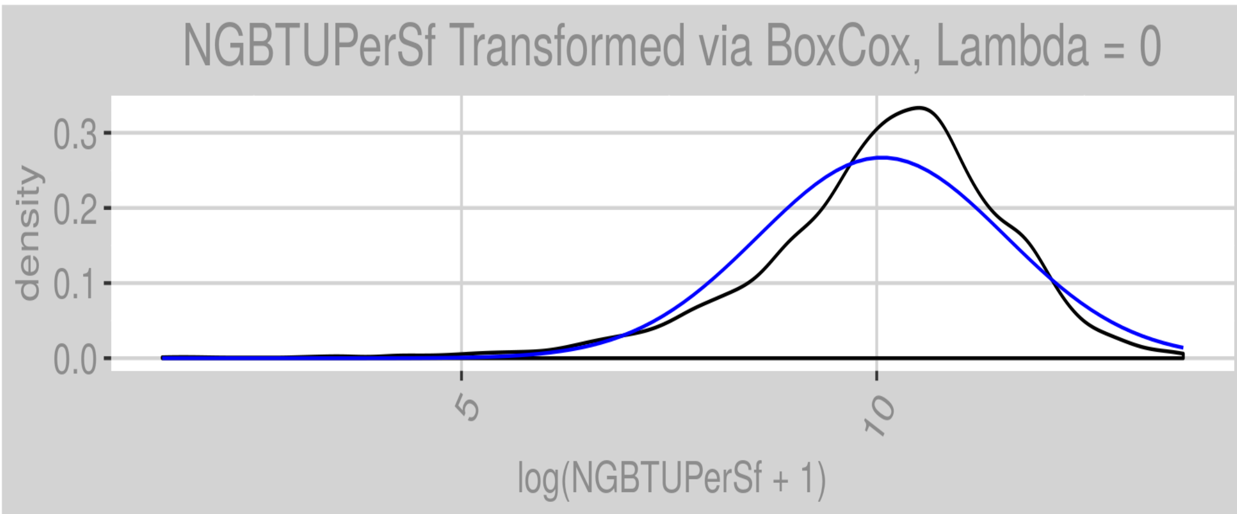

NGCOOK.1 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
NGCOOK.1
NGCOOK.1 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
NGCOOK.1
PCTERMNPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.000
0.005
0.010
0.015
0.020
0.025
PCTERMNPerSf
PCTERMNPerSf - Transformed
Logscale Electricity MBTU
10
5
0
-2
0
2
PCTERMNPerSf
LNHRPC.5 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
LNHRPC.5
LNHRPC.5 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
LNHRPC.5


PBA.6 - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0
1
PBA.6
PBA.6 - Transformed
Logscale Electricity MBTU
10
5
0
0
1
PBA.6
RFGOPNPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.000
0.001
0.002
0.003
RFGOPNPerSf
RFGOPNPerSf - Transformed
Logscale Electricity MBTU
10
5
0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
RFGOPNPerSf
PRNTRNPerSf - Untransformed
Electricity MBTU
8e+05
6e+05
4e+05
2e+05
0e+00
0.000
0.003
0.006
0.009
PRNTRNPerSf
PRNTRNPerSf - Transformed
Logscale Electricity MBTU
10
5
0
-2
0
2
4
PRNTRNPerSf

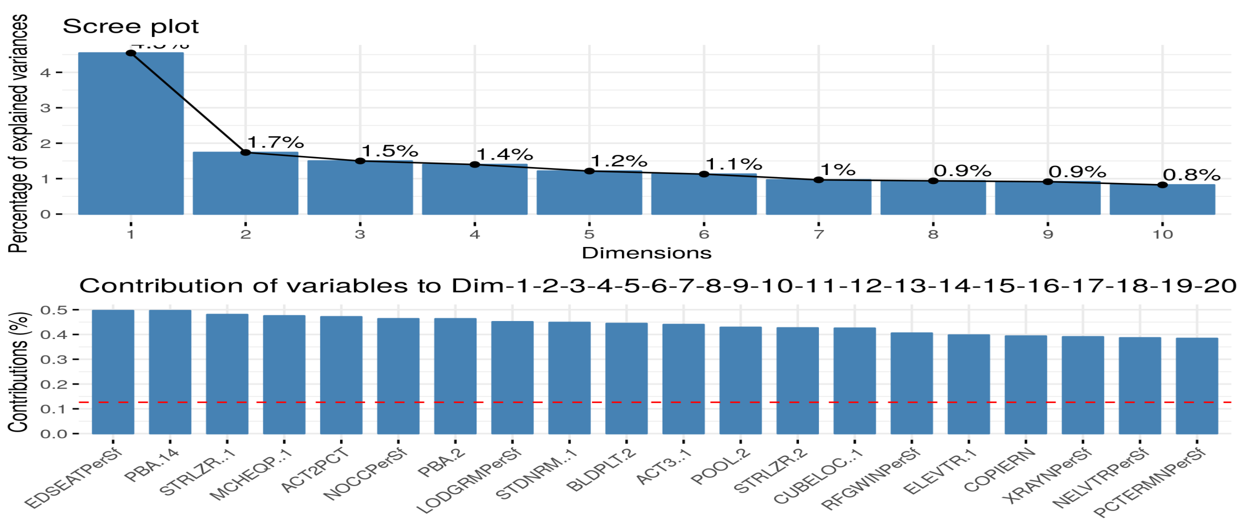

RFGOP.1 - Untransformed
RFGOP.1 - Transformed
PBA.5 - Untransformed
PBA.5 - Transformed
WKHRSC.6 - Untransformed
WKHRSC.6 - Transformed
Electricity MBTU
Logscale Electricity MBTU
0e+00
2e+05
4e+05
6e+05
8e+05
0
5
10
0
1
RFGOP.1
PBA.5
WKHRSC.6

Appendix - Natural Gas

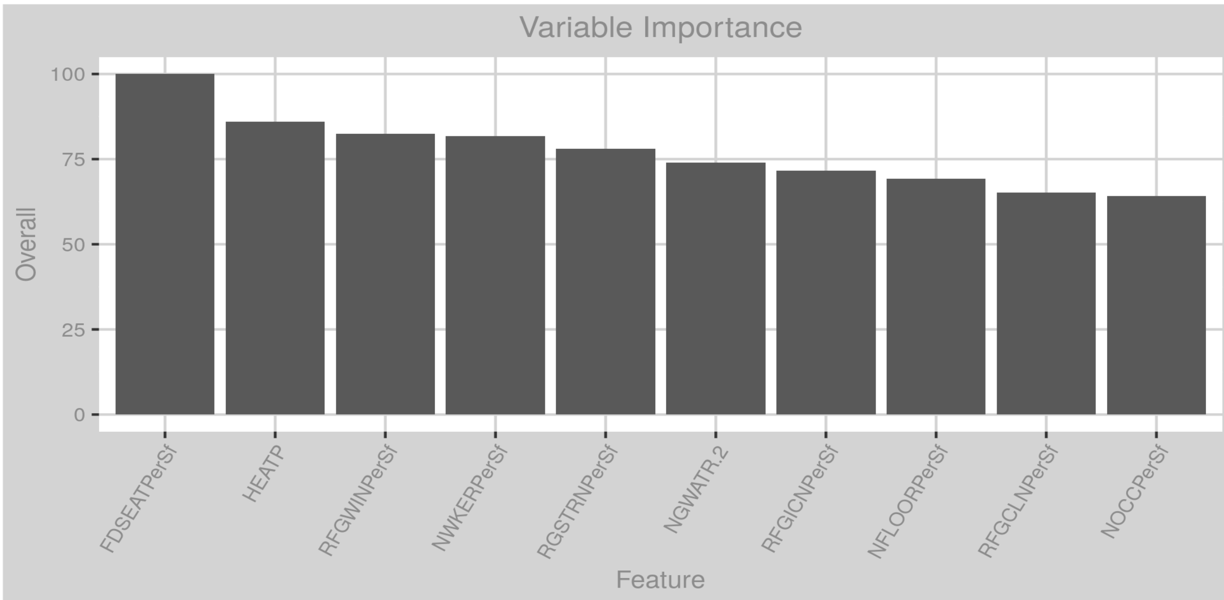
Response



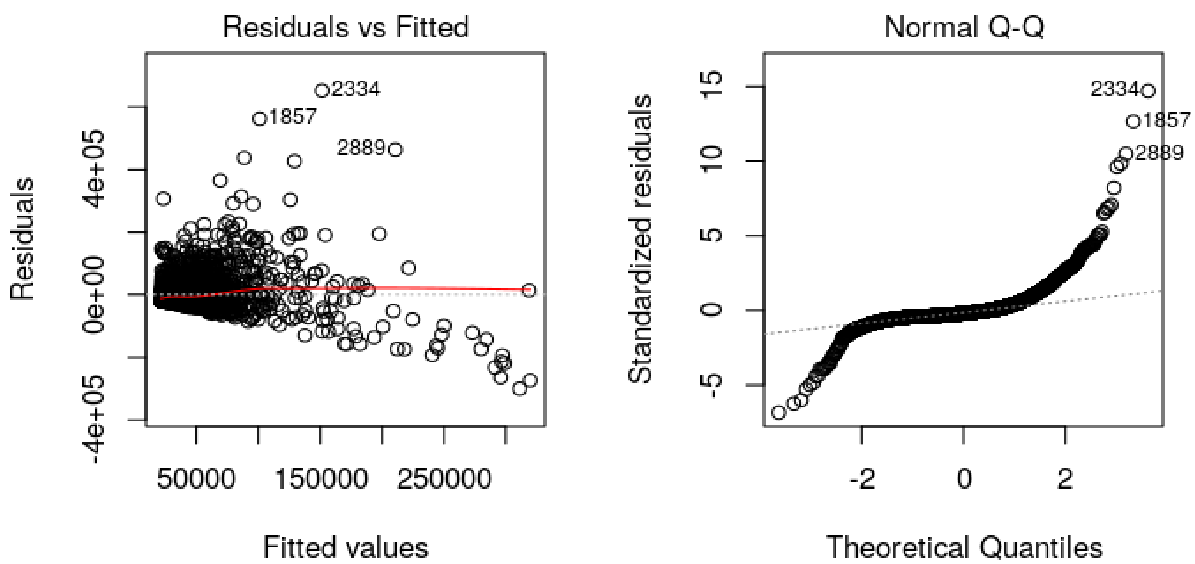
PCA



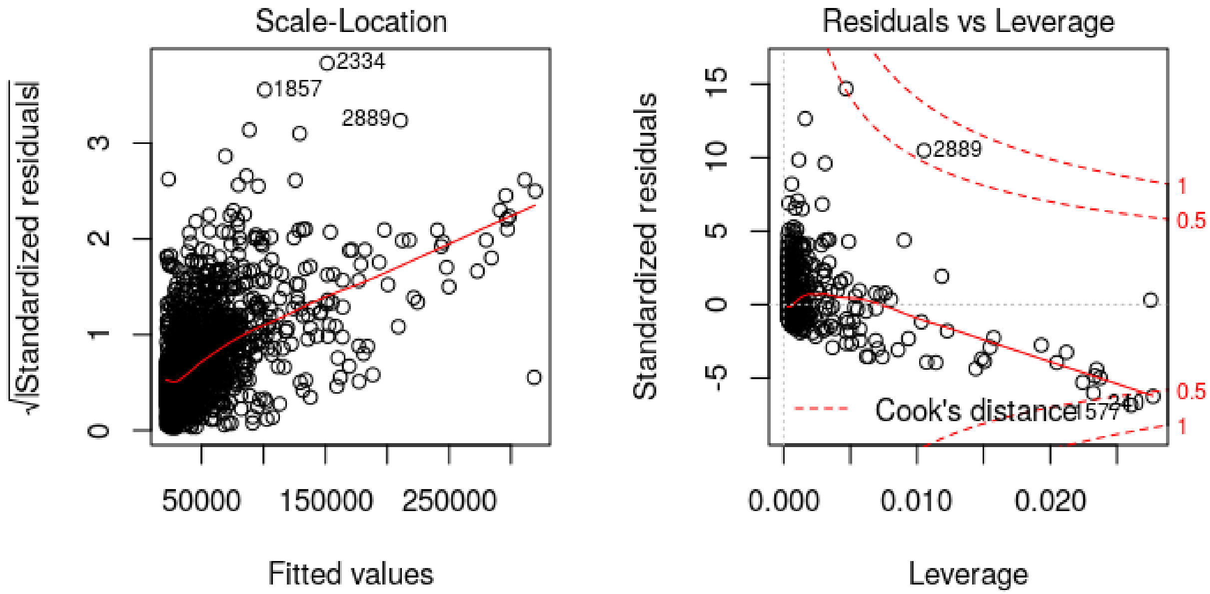
PLS

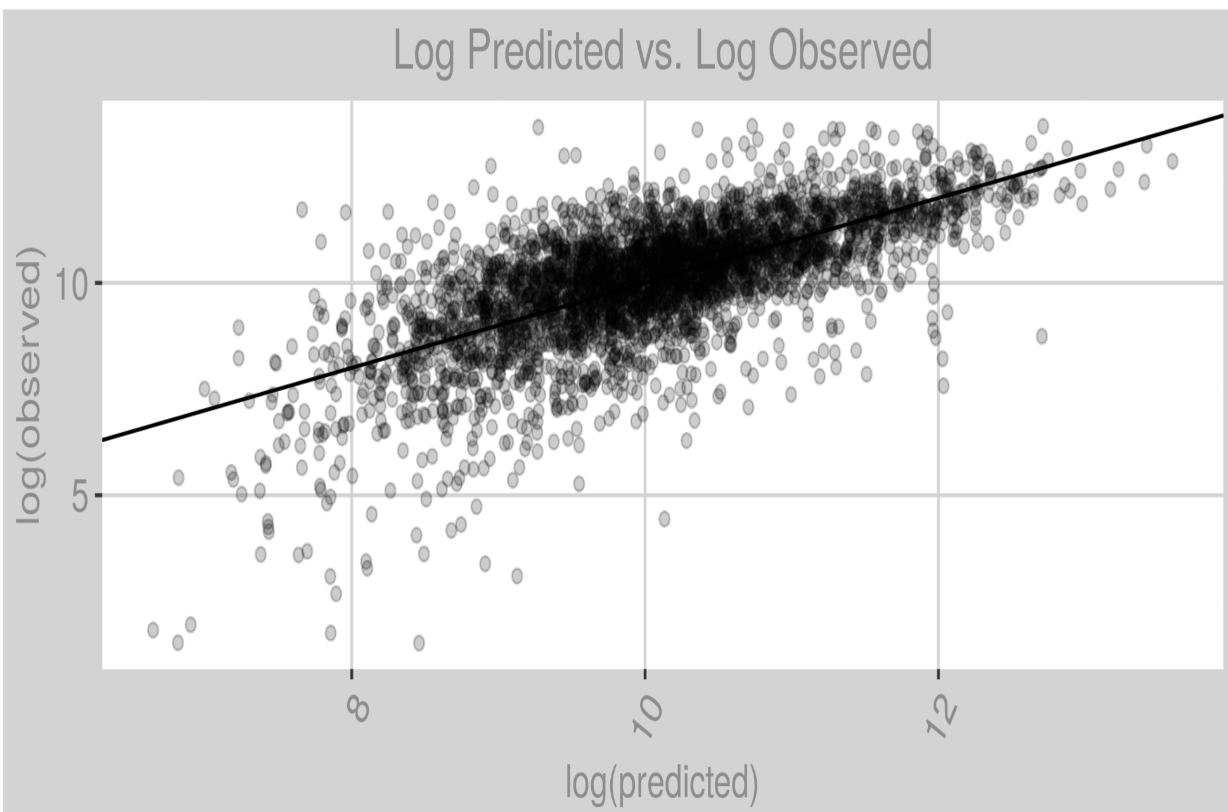


Variable Importance
Overall
100
75
50
25
0
FDSEATPerSf
HEATP
RFGWINPerSf
NWKERPerSf
RGSTRNPerSf
NGWATR.2
RFGICNPerSf
NFLOORPerSf
RFGCLNPerSf
NOCCPerSf
Feature

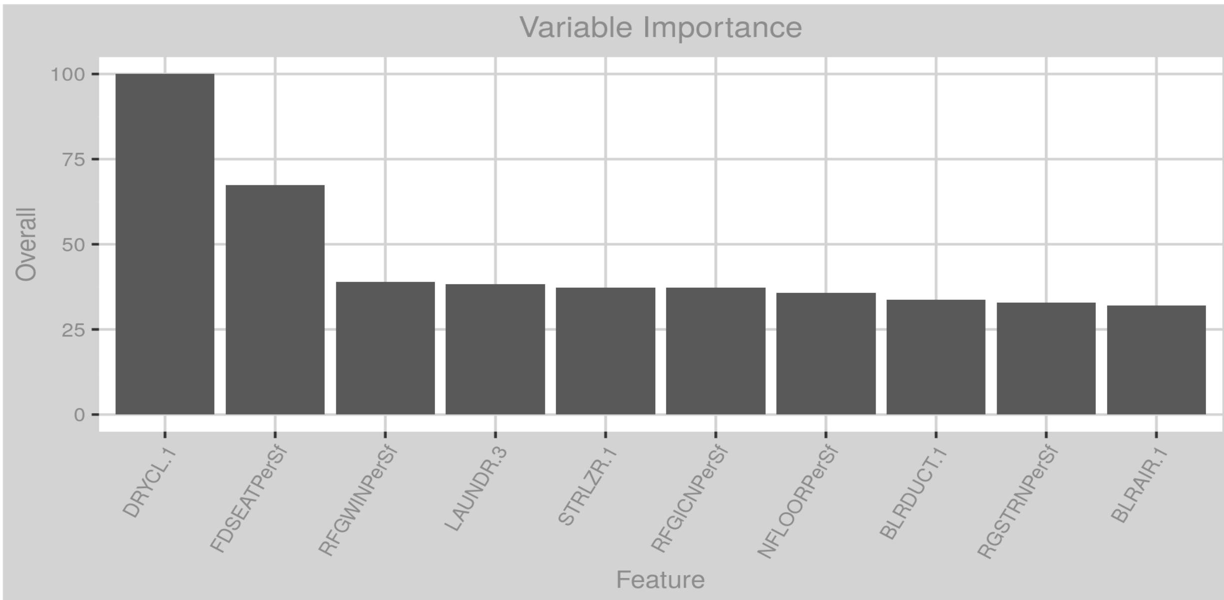


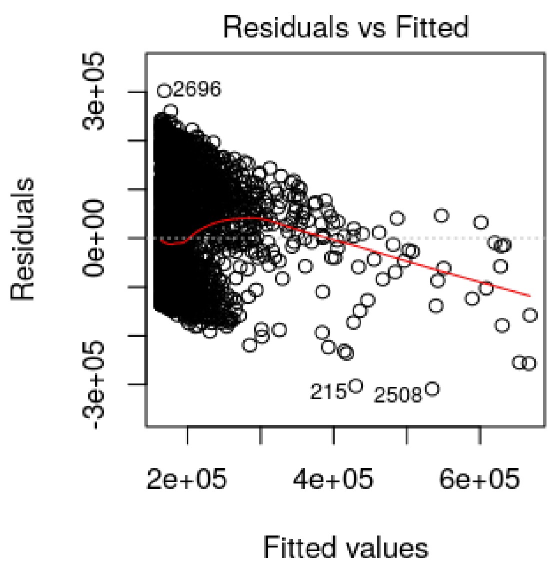
Residuals vs Fitted
Residuals
-4e+05
0e+00
4e+05
2334
1857
2889
50000
150000
250000
Fitted values
Normal Q-Q
Standardized residuals
-5
0
5
10
15
2334
1857
2889
-2
0
2
Theoretical Quantiles

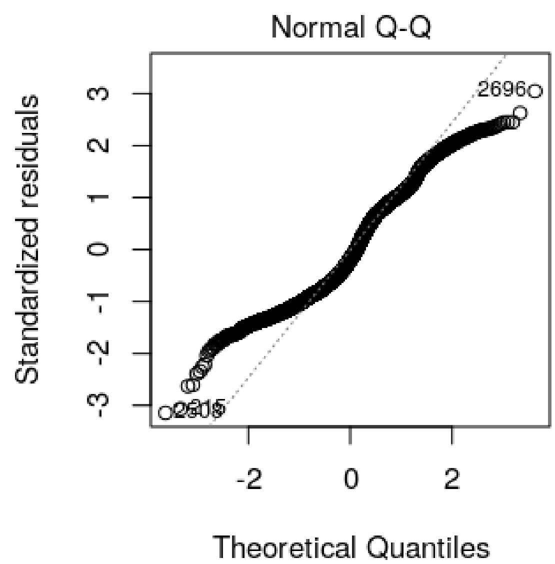

Scale-Location
√|Standardized residuals|
Fitted values
2334
1857
2889
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
2889
0.5
1

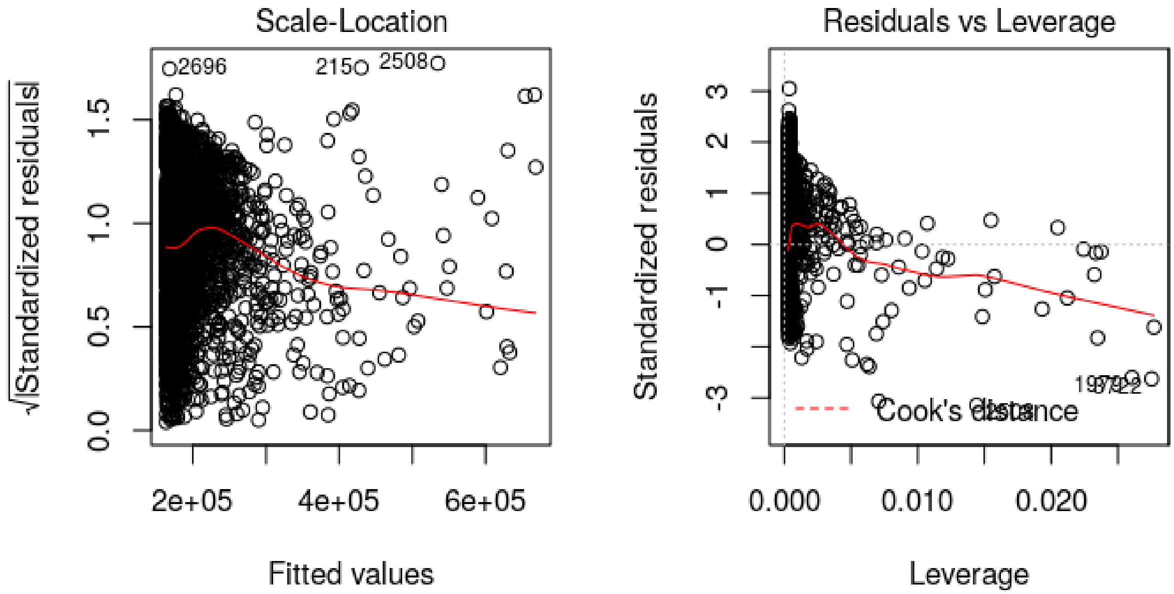

Log Predicted vs. Log Observed
log(observed)
log(predicted)

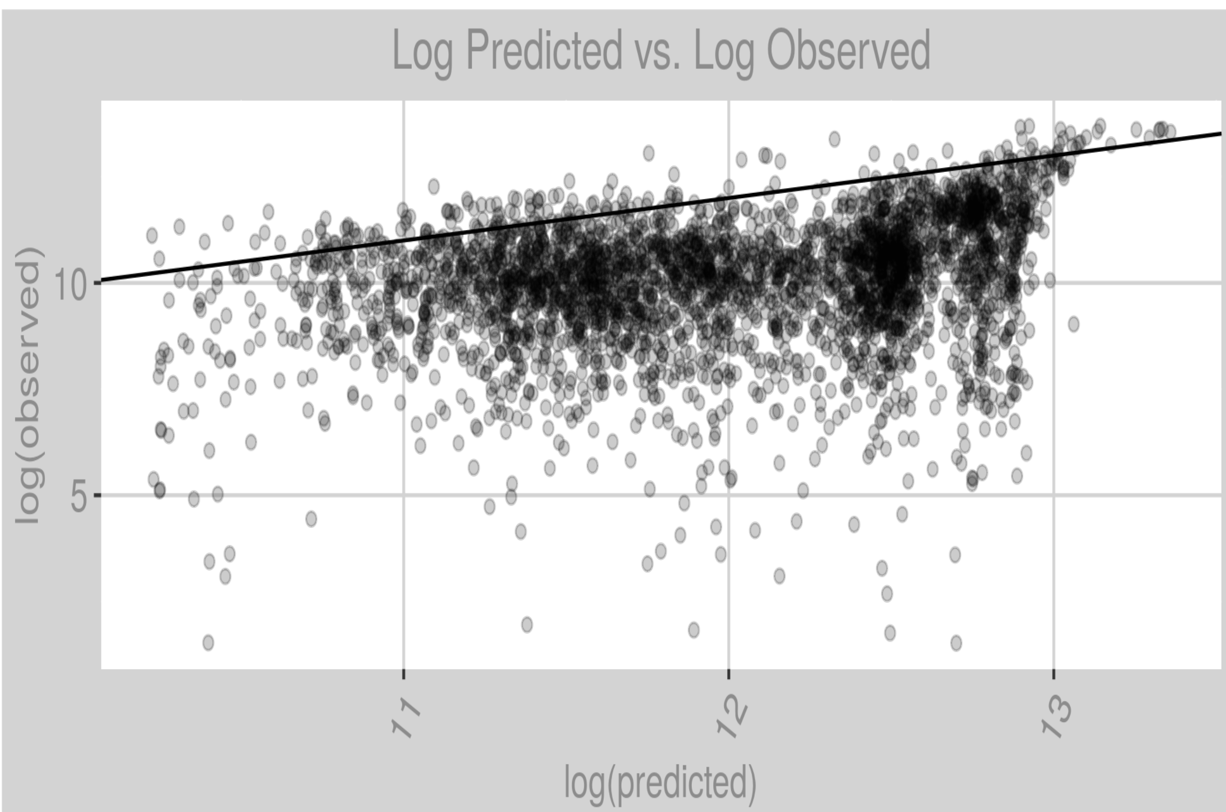
Random Forest



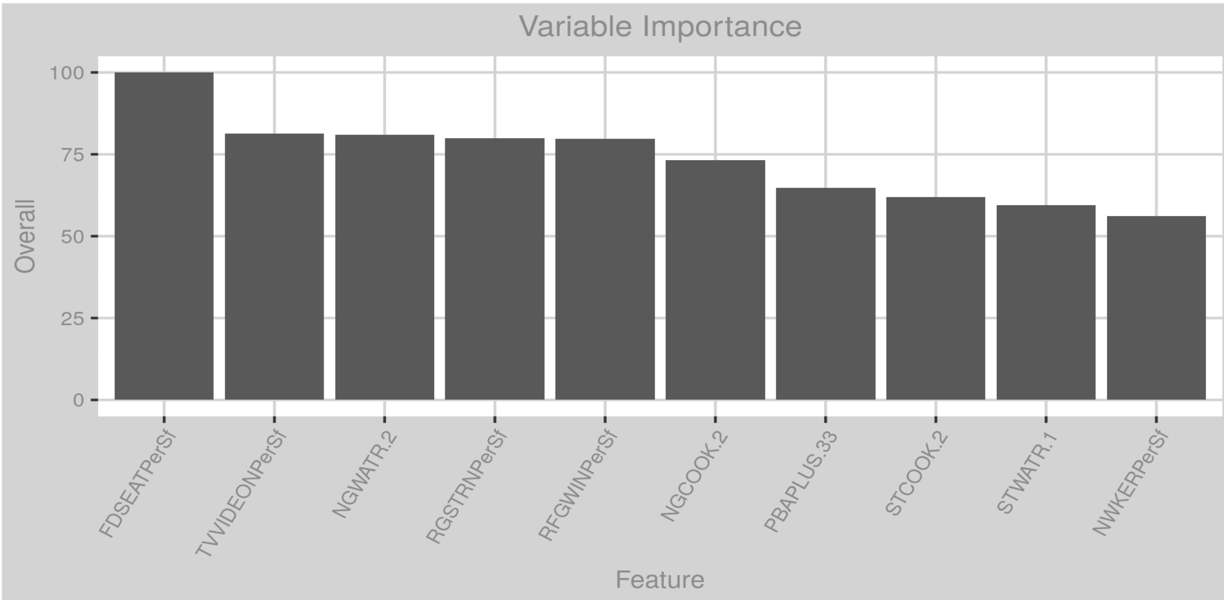


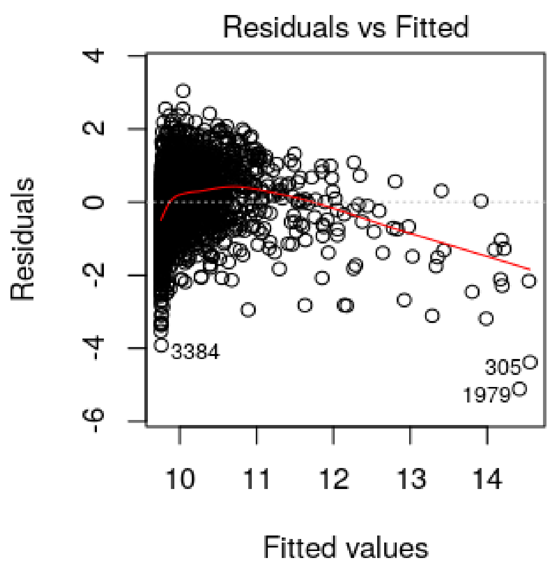


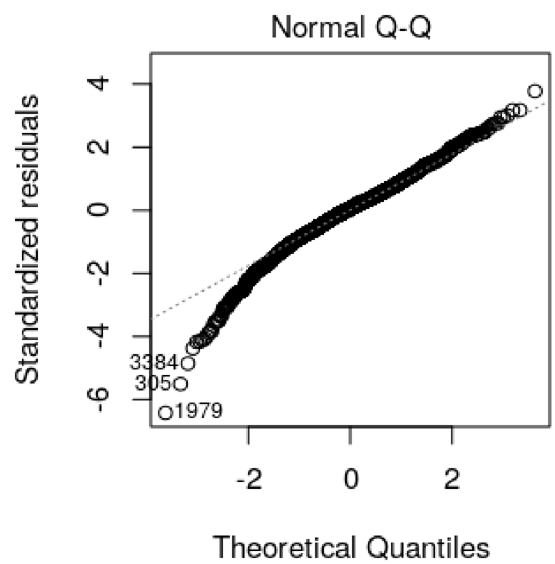

Scale-Location
Residuals vs Leverage
√|Standardized residuals|
Fitted values
Standardized residuals
Leverage
Cook's distance
2696
215
2508

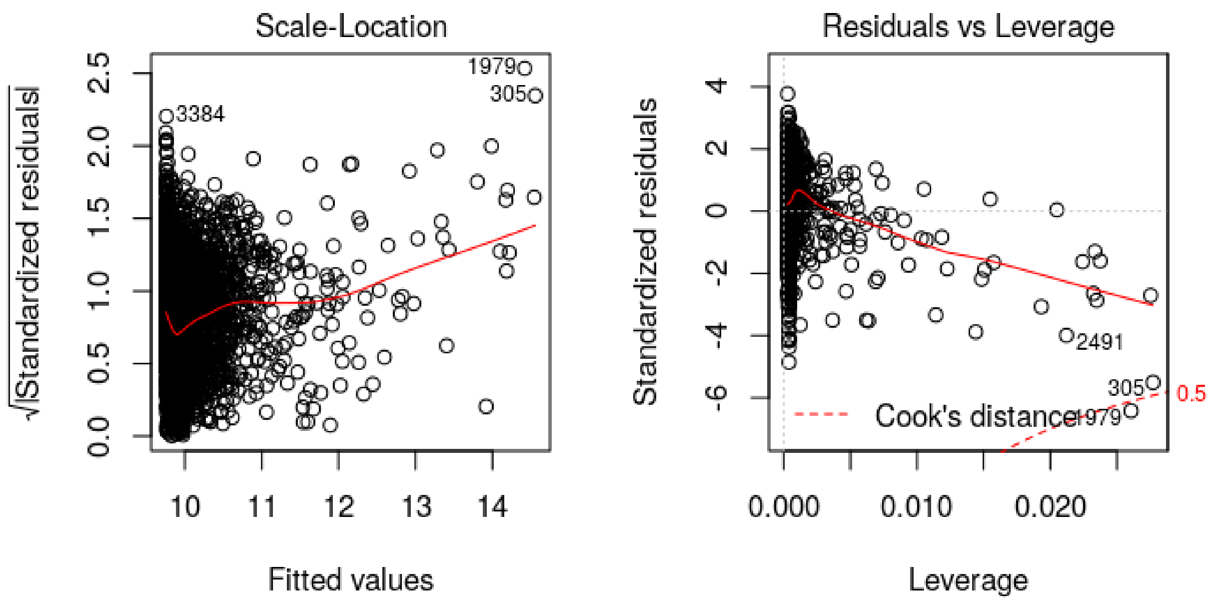

Log Predicted vs. Log Observed
log(observed)
log(predicted)

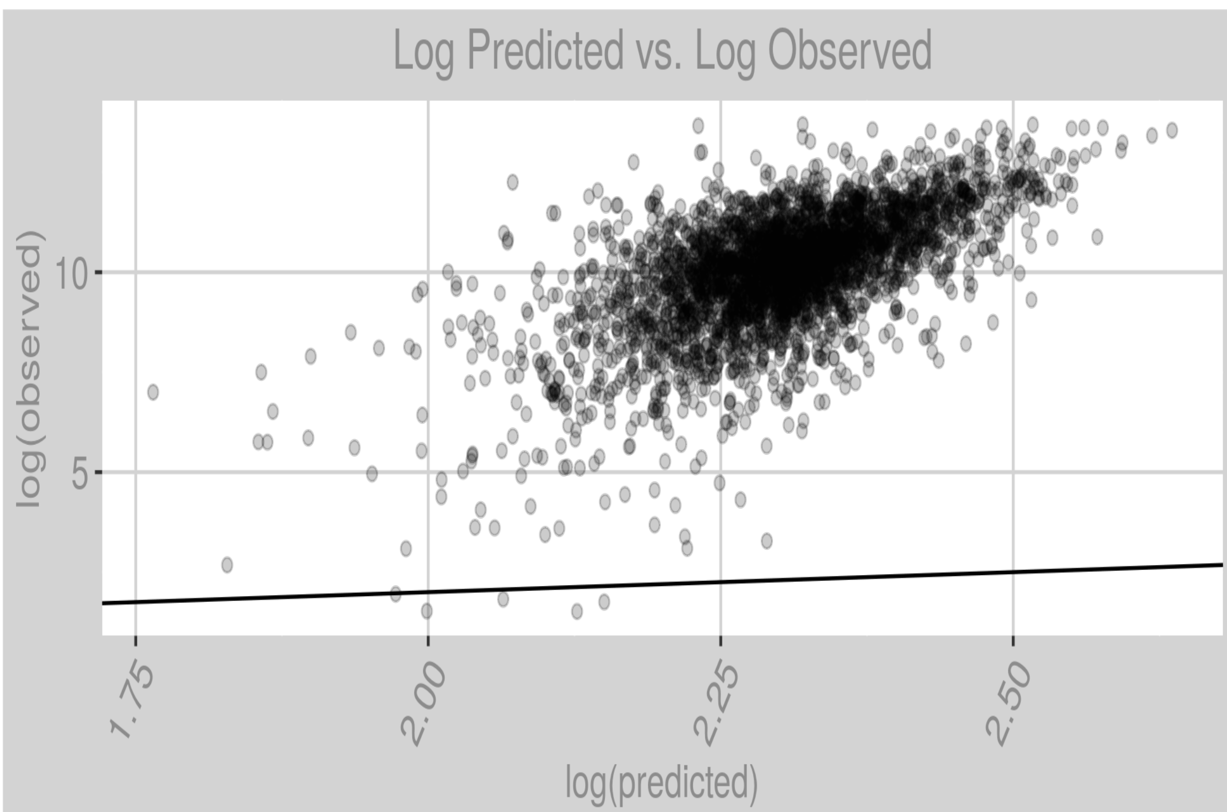
Forward Selection



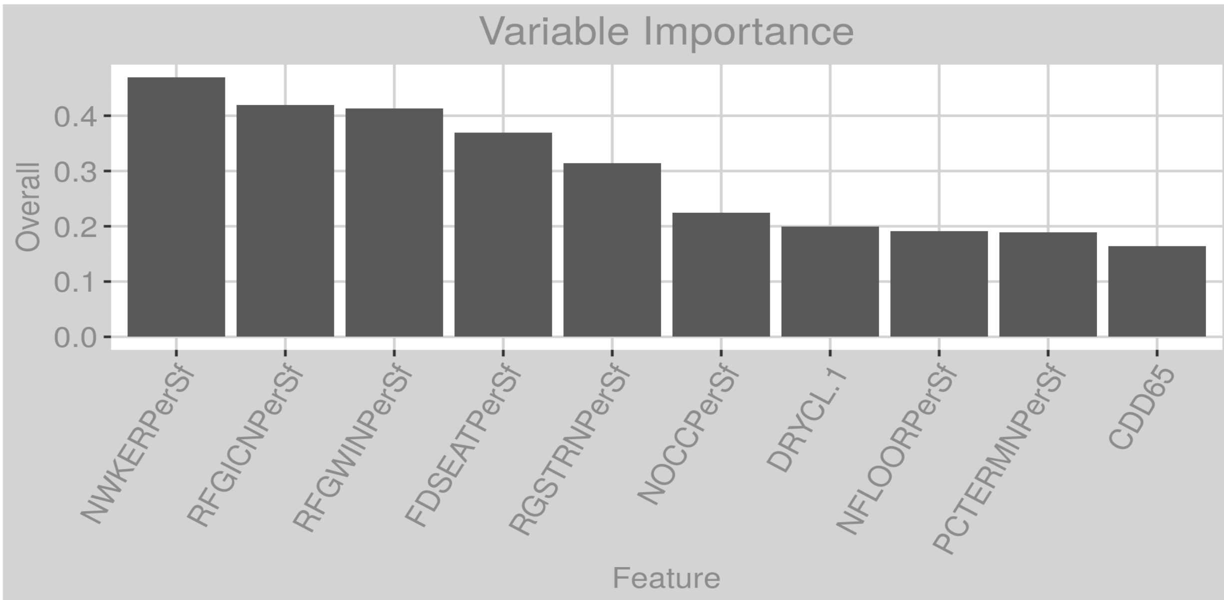


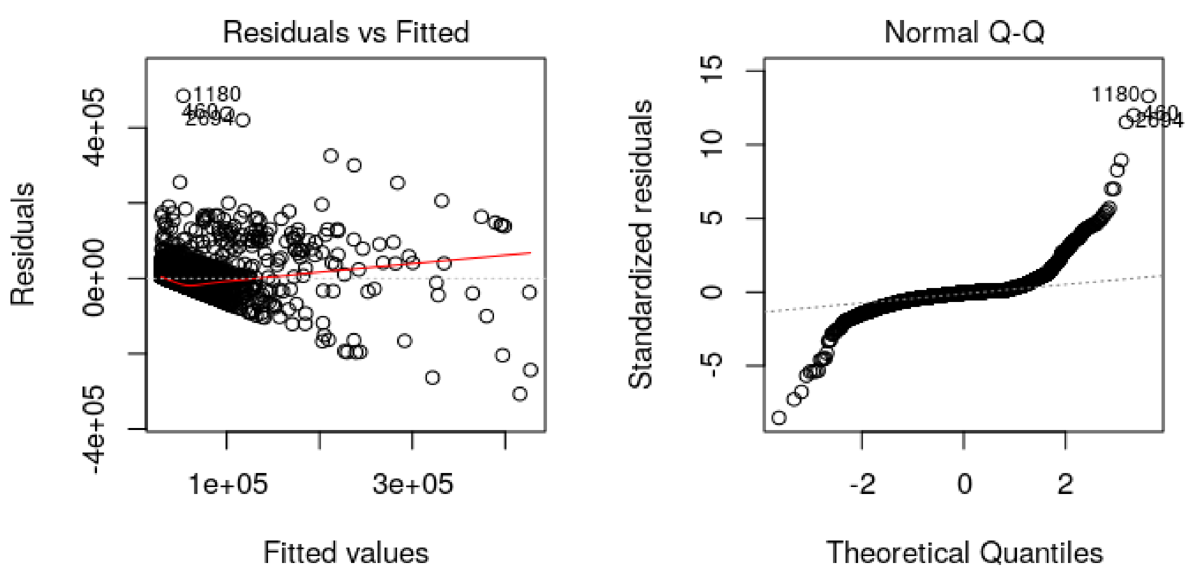


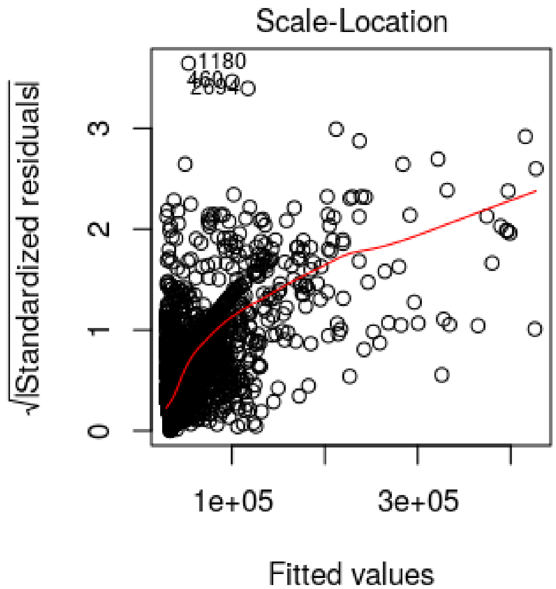

Scale-Location
√|Standardized residuals|
Fitted values
0.0
0.5
1.0
1.5
2.0
2.5
10
11
12
13
14
3384
1979
305
Residuals vs Leverage
Standardized residuals
Leverage
-6
-4
-2
0
2
4
0.000
0.010
0.020
Cook's distance
0.5
2491
305
1979

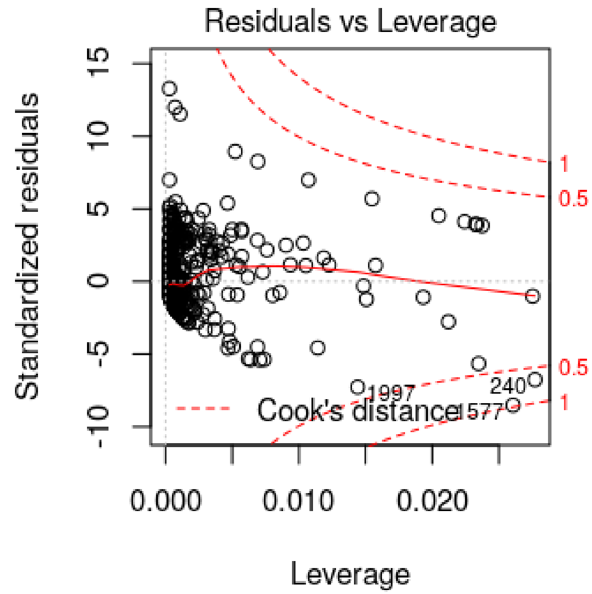

Log Predicted vs. Log Observed
log(observed)
log(predicted)
5
10
1.75
2.00
2.25
2.50

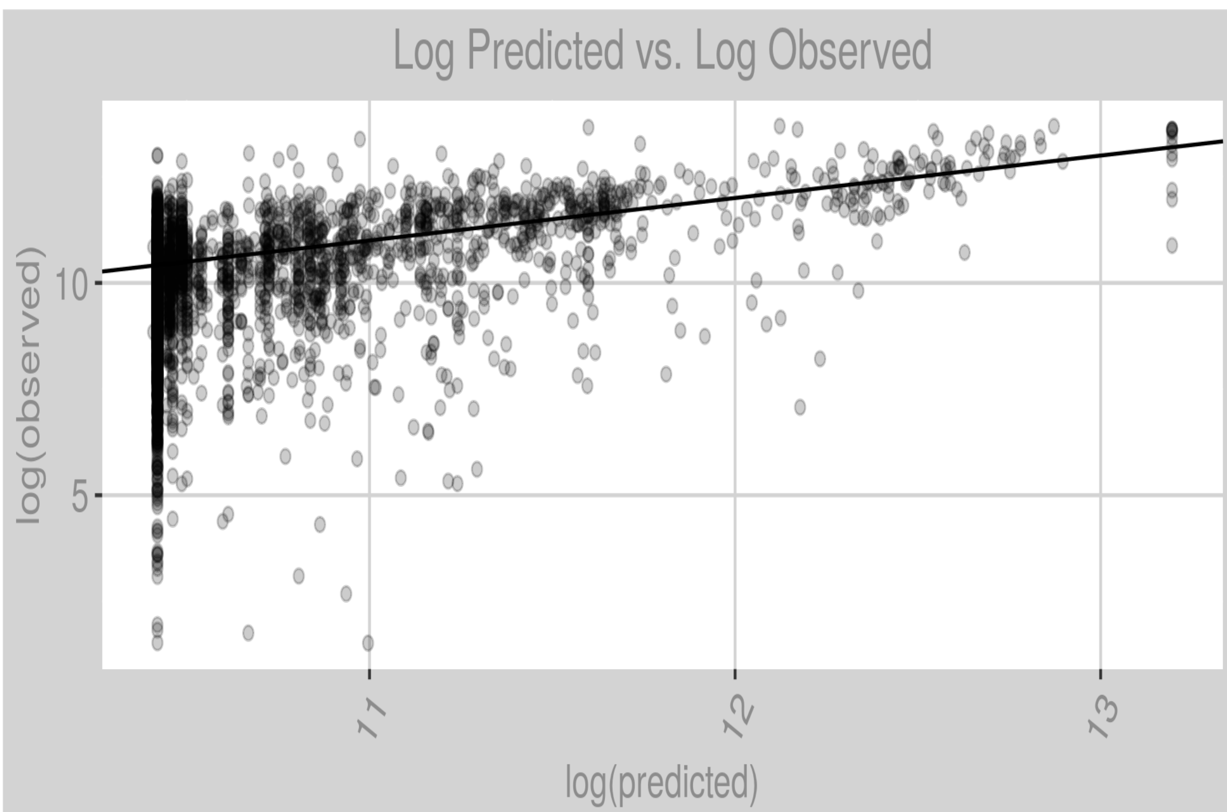
Recursive Feature Extraction



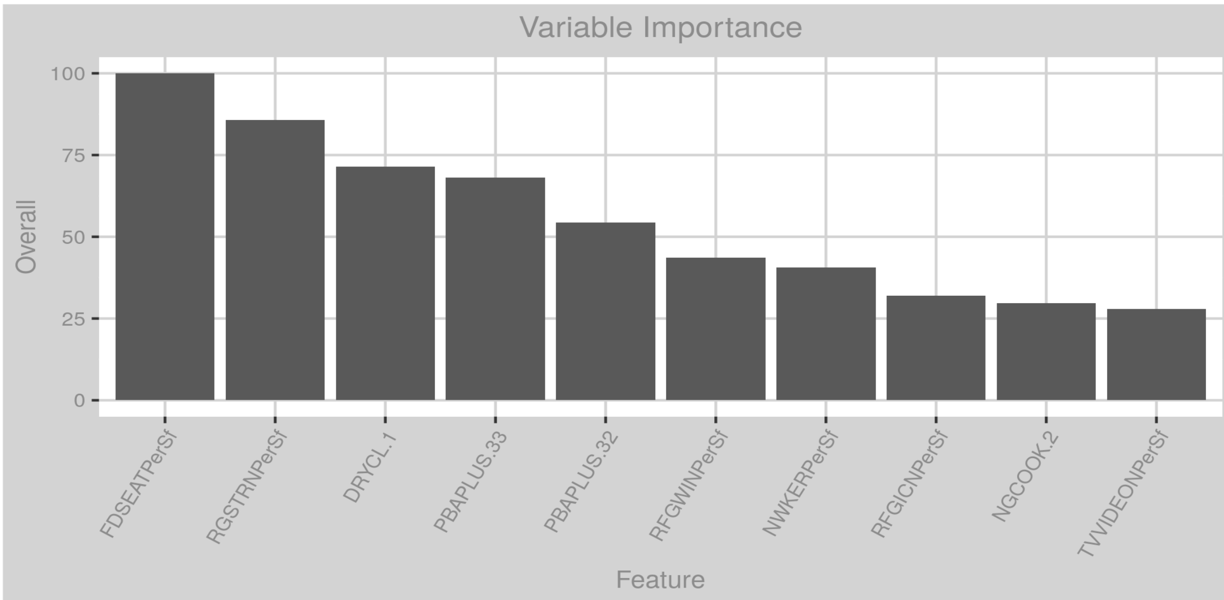


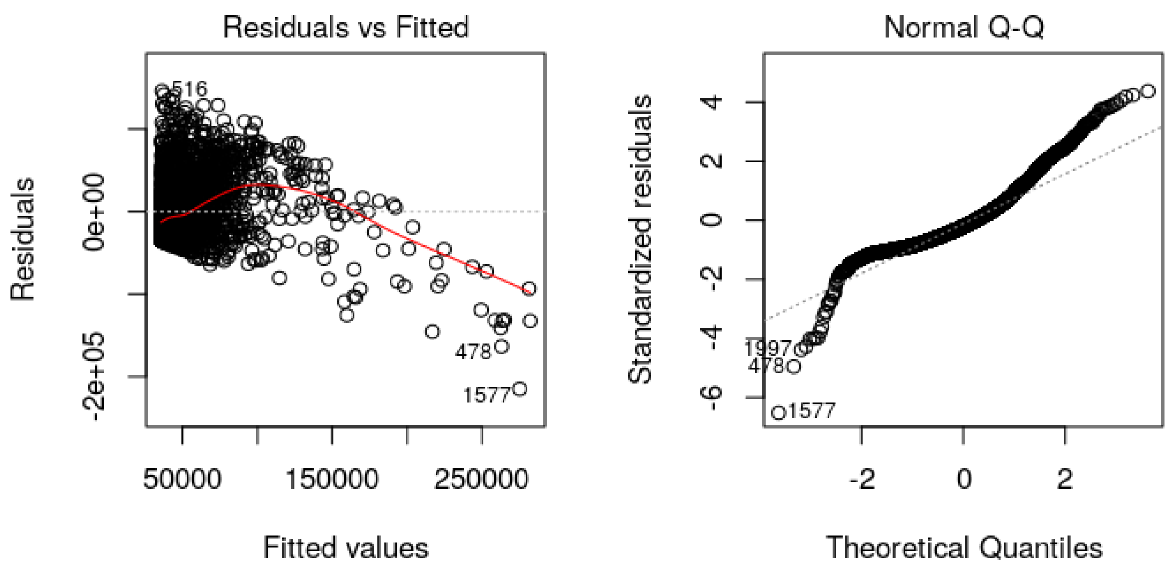

Scale-Location
√|Standardized residuals|
Fitted values
1180
460
2694
0
1
2
3
1e+05
3e+05

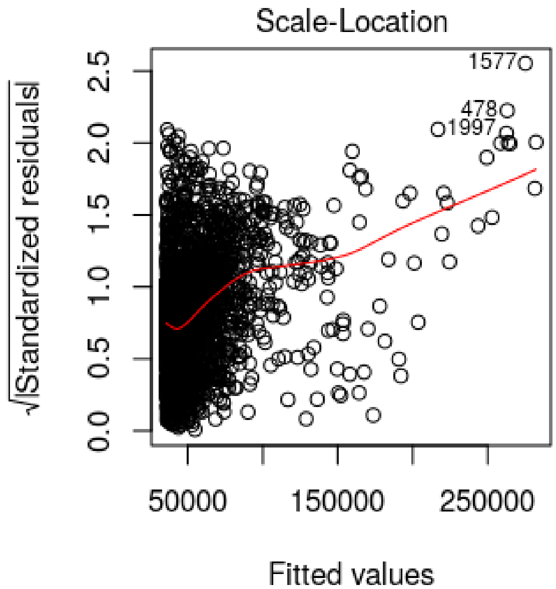

Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
1997
240
1577
1
0.5
0.5
1
15
10
5
0
-5
-10
0.000
0.010
0.020

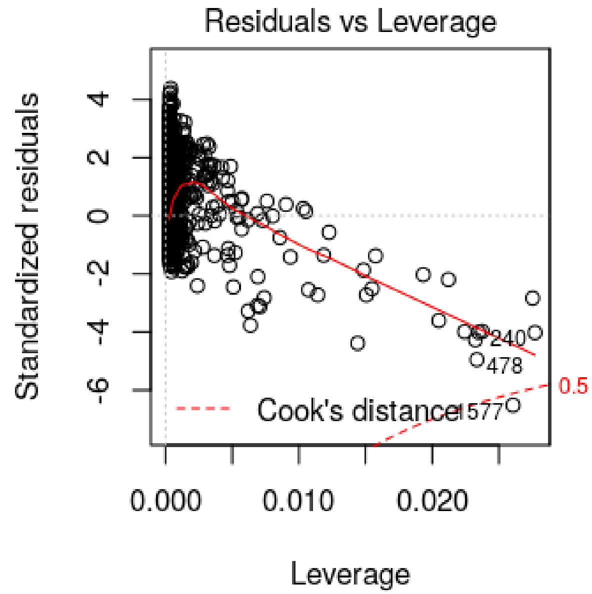

Log Predicted vs. Log Observed
log(observed)
log(predicted)
10
5
11
12
13

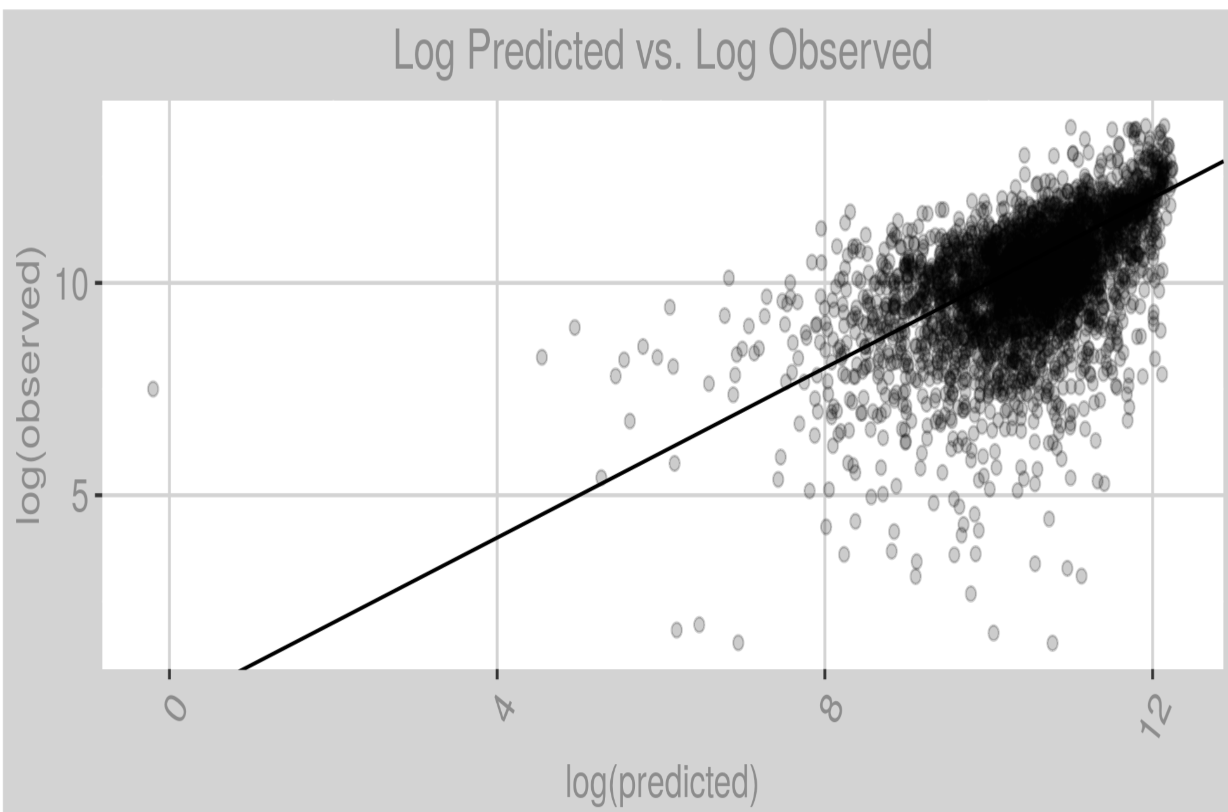
Simple Neural Network






Scale-Location
√|Standardized residuals|
Fitted values
50000
150000
250000
0.0
0.5
1.0
1.5
2.0
2.5
1577
478
1997

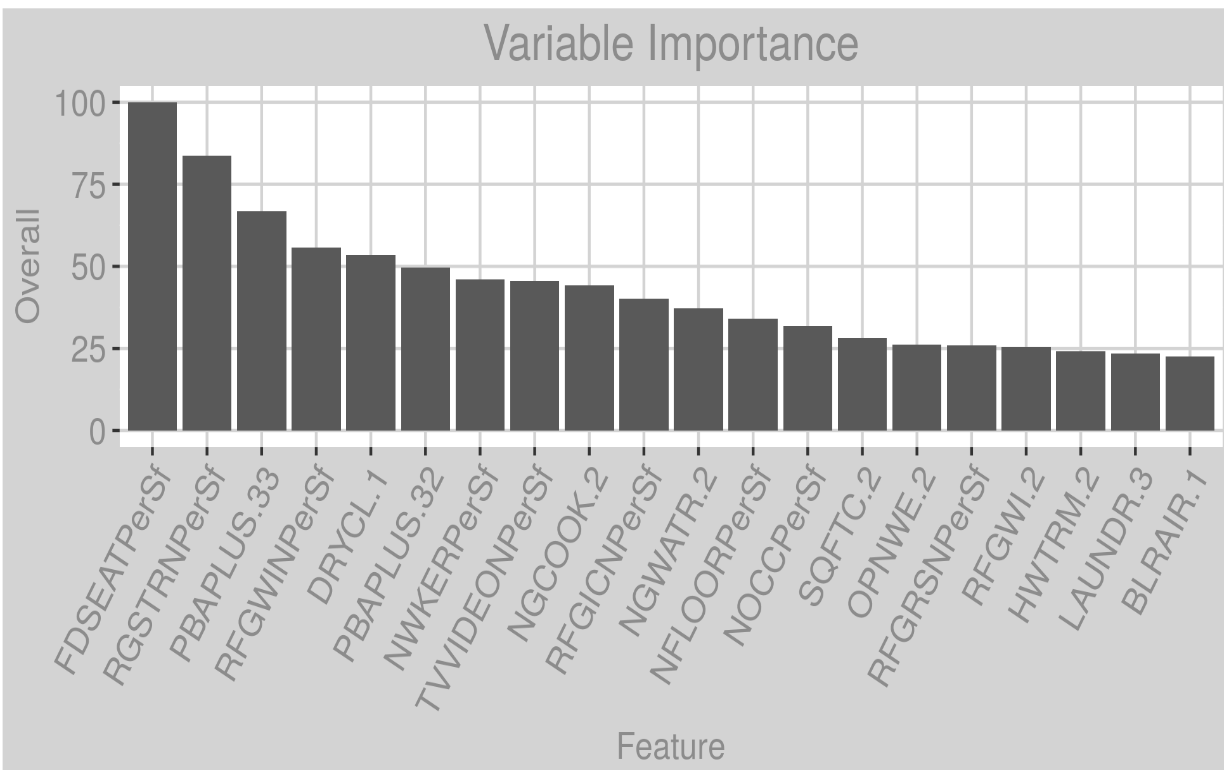

Residuals vs Leverage
Standardized residuals
Leverage
0.000
0.010
0.020
-6
-4
-2
0
2
4
Cook's distance
0.5
240
478
1577

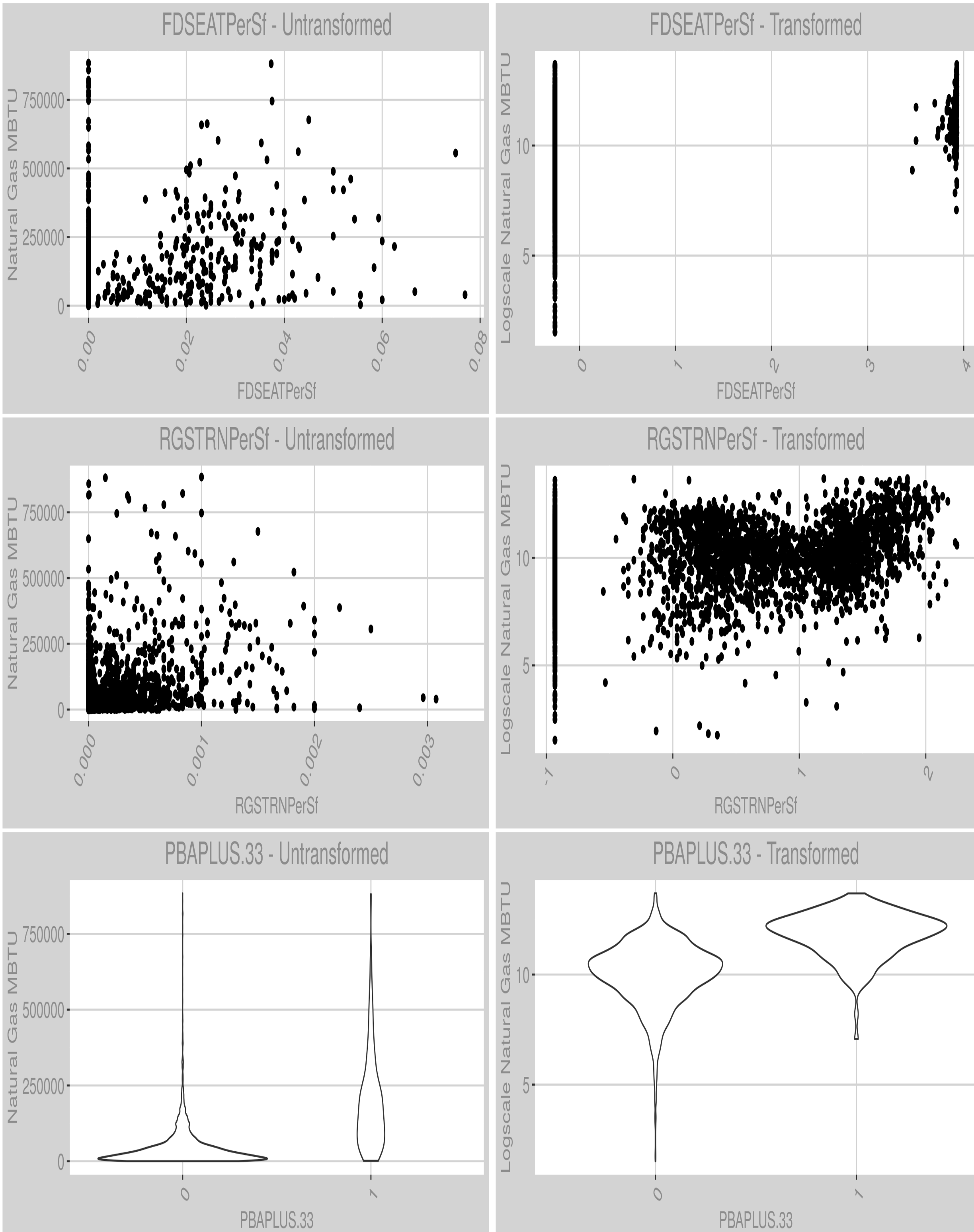

Log Predicted vs. Log Observed
log(observed)
log(predicted)
5
10
0
4
8
12

Select Variable Analysis

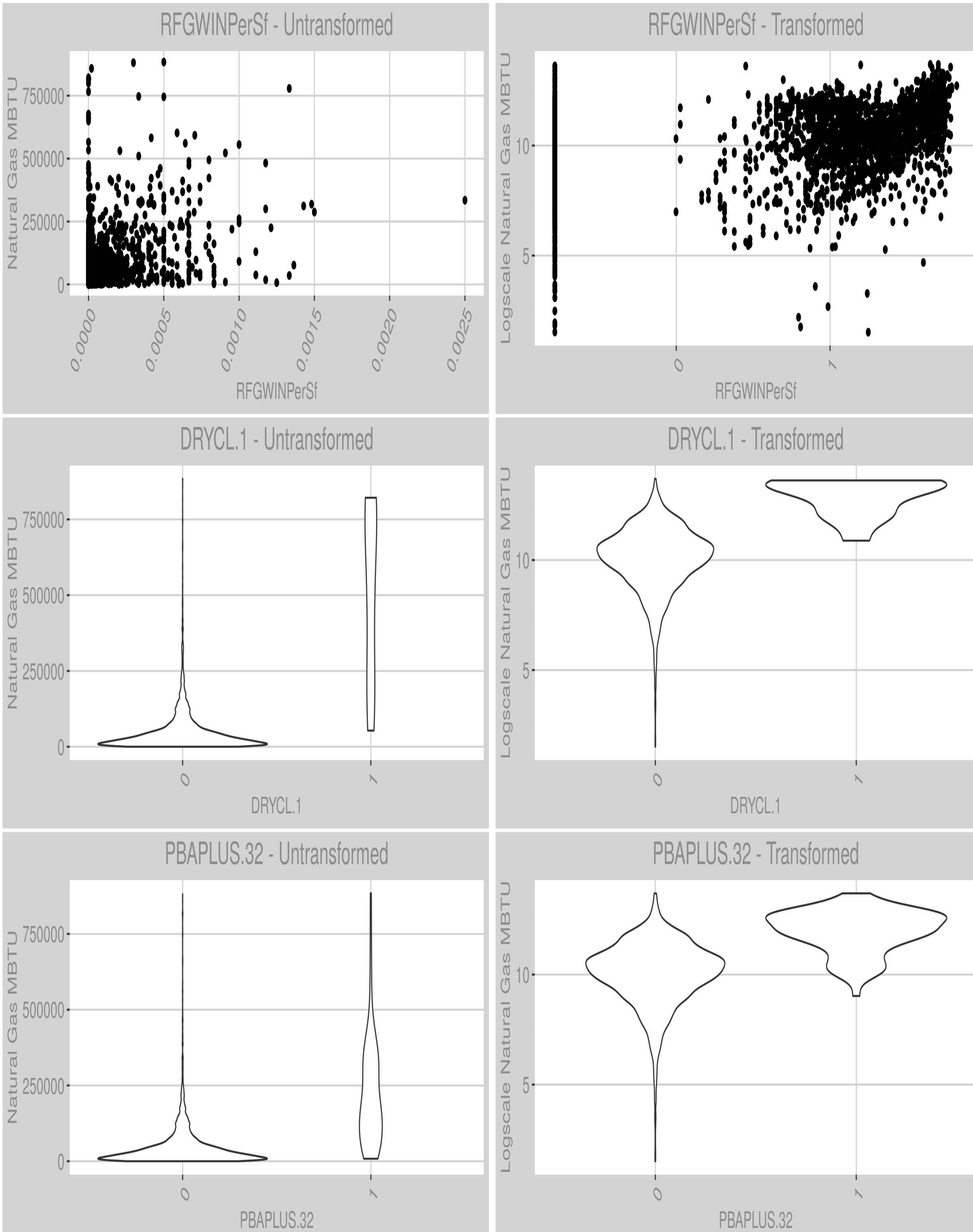
Feature Extraction Model Results

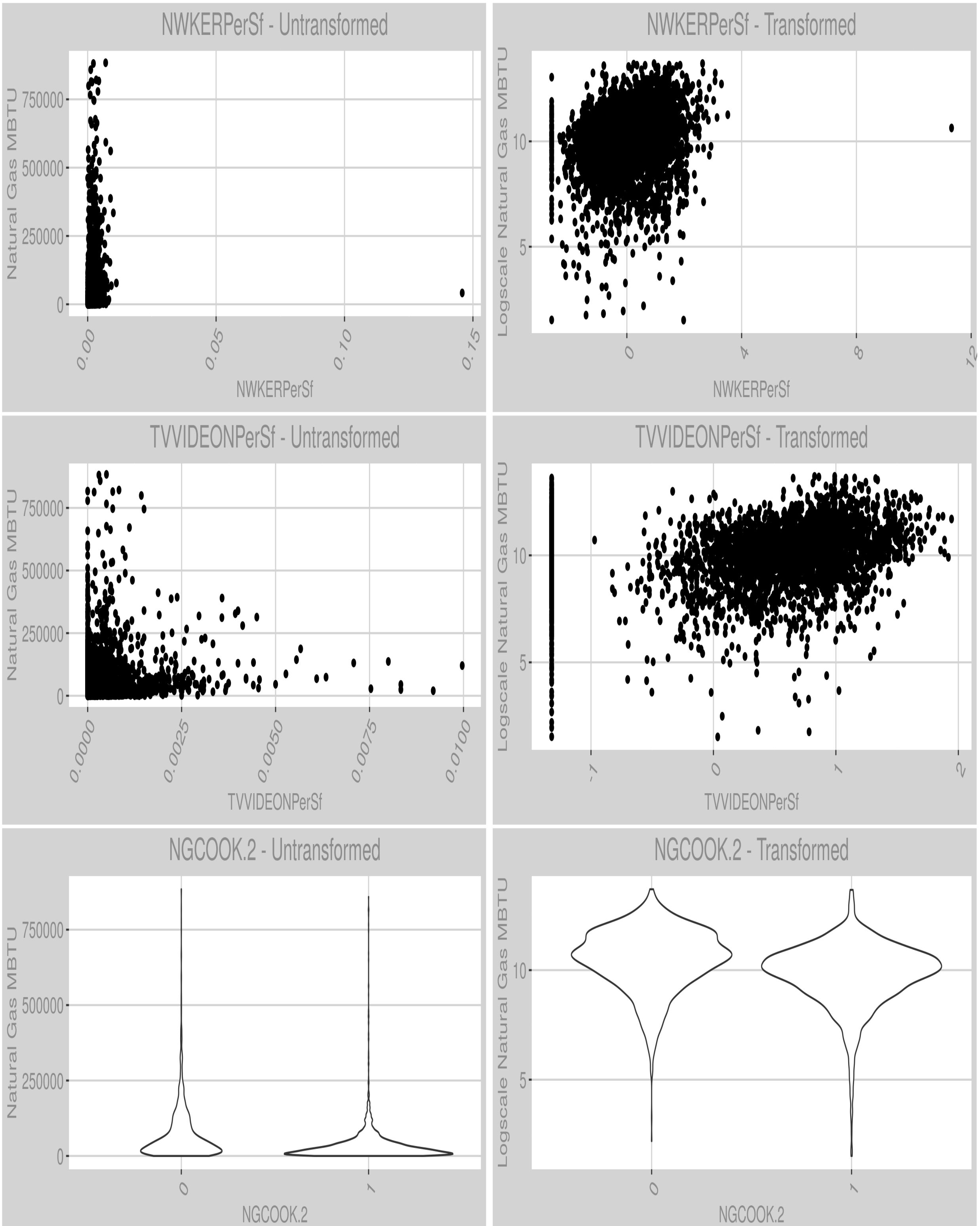
Model	RMSE	R2	MAE
recursiveFeatureExtraction	59212.72	0.5230241	34203.37
neuralNetwork	69784.67	0.3336745	35642.13
partialLeastSquares	73075.99	0.2934536	32068.61
leaps	100385.86	0.2462162	53658.33
randomForest	175808.23	0.1930417	147144.46

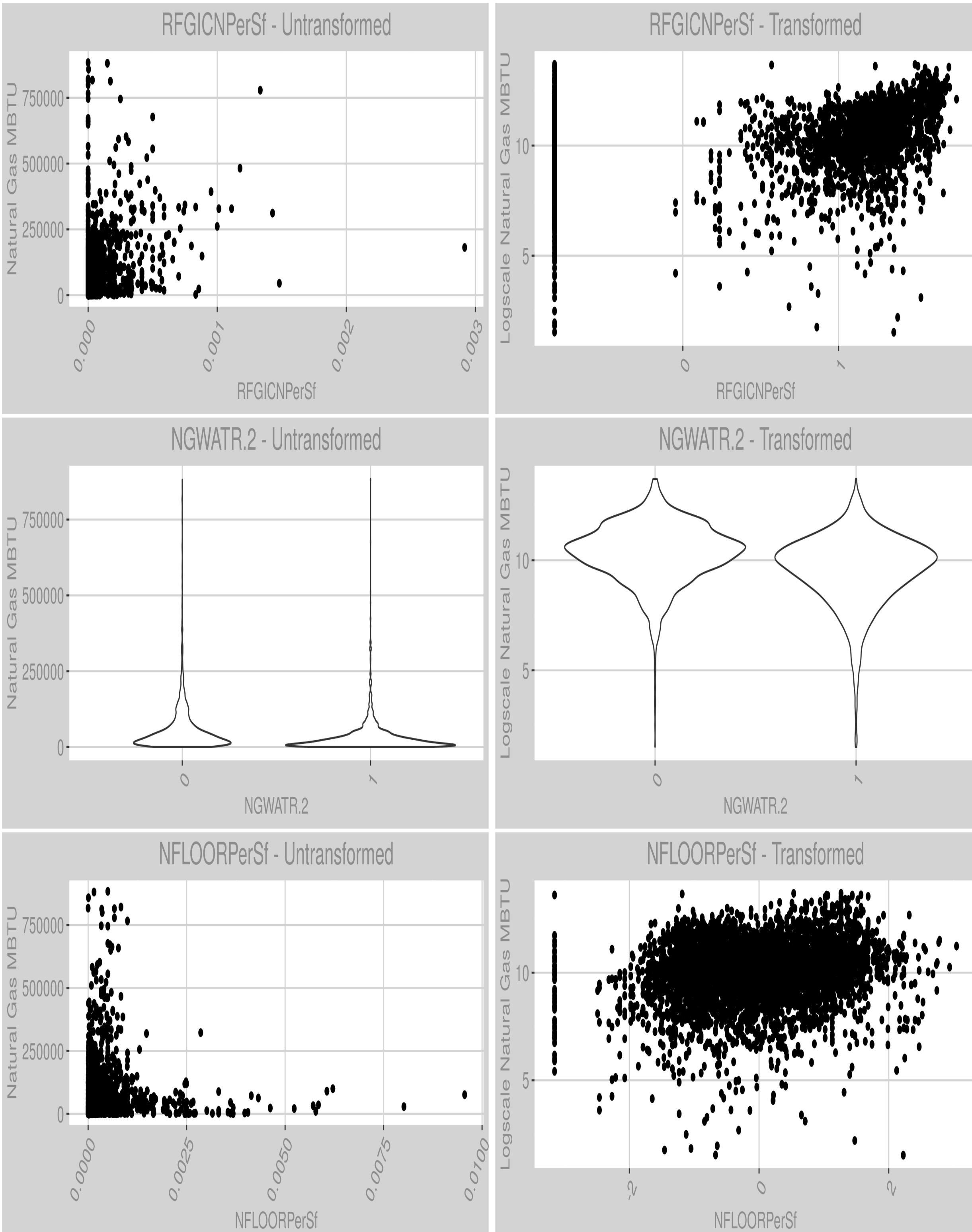




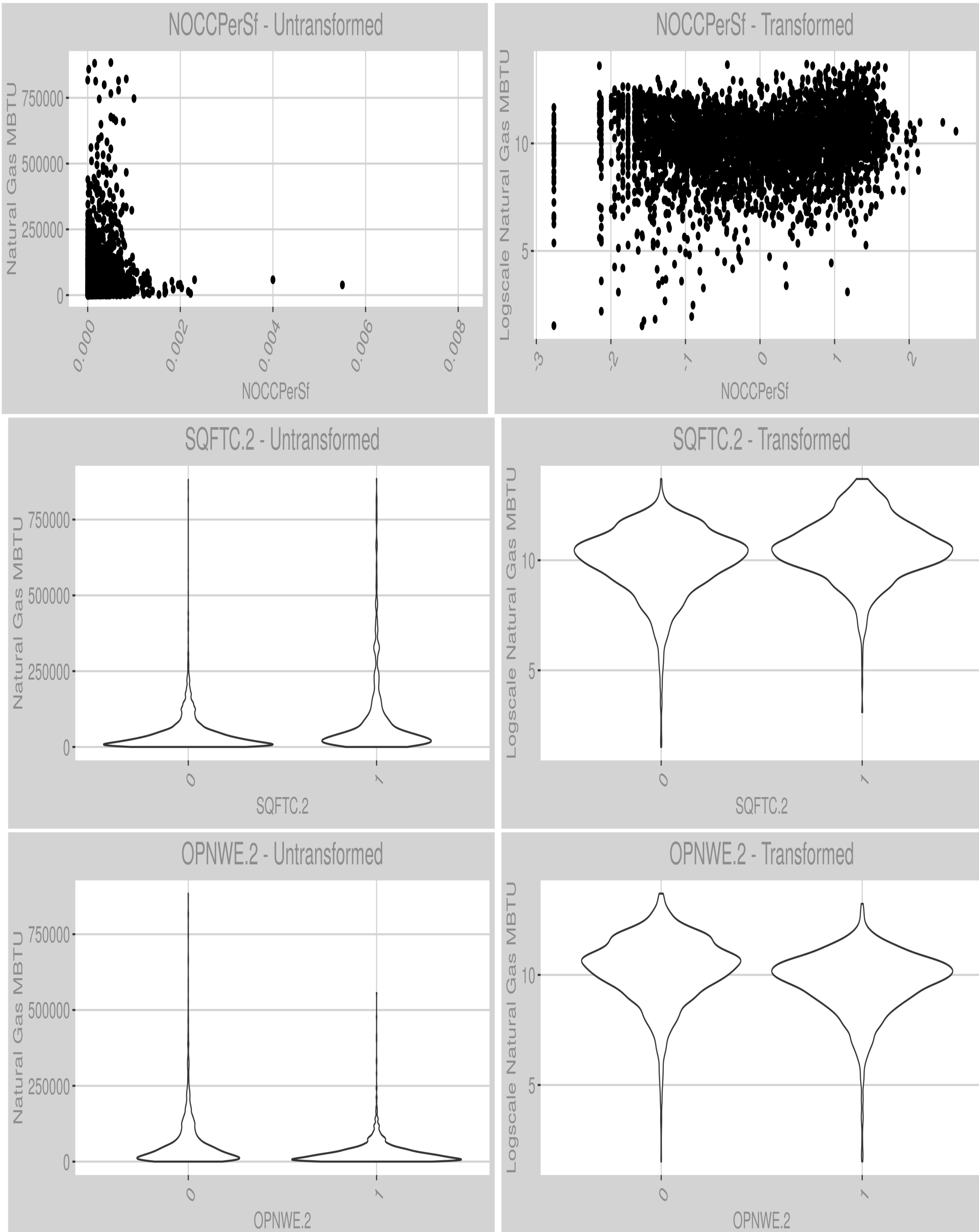
FDSEATPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.00
0.02
0.04
0.06
0.08
FDSEATPerSf
FDSEATPerSf - Transformed
Logscale Natural Gas MBTU
10
5
0
1
2
3
4
FDSEATPerSf
RGSTRNPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.000
0.001
0.002
0.003
RGSTRNPerSf
RGSTRNPerSf - Transformed
Logscale Natural Gas MBTU
10
5
-1
0
1
2
RGSTRNPerSf
PBAPLUS.33 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
PBAPLUS.33
PBAPLUS.33 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
PBAPLUS.33

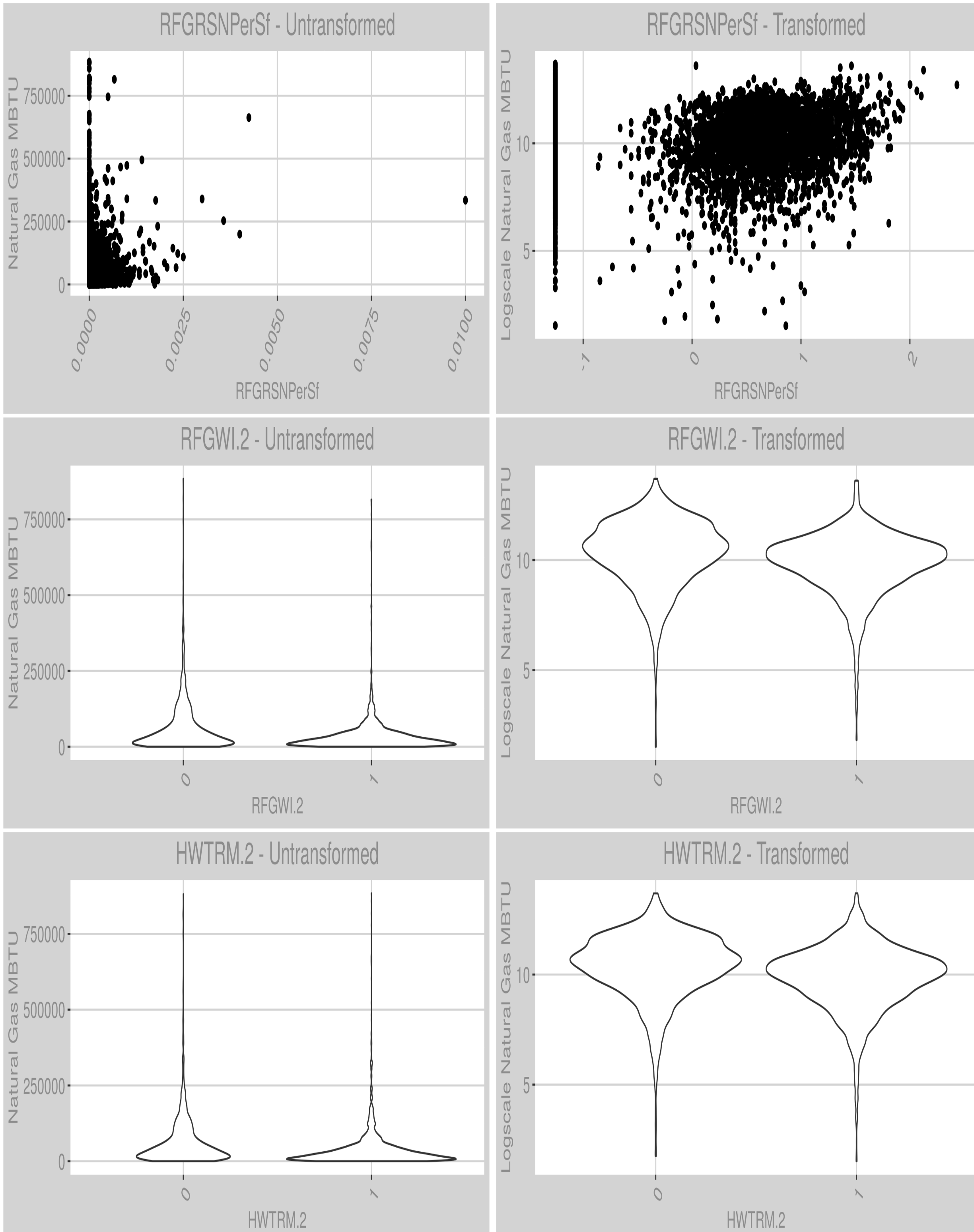

RFGWINPerSf - Untransformed
Natural Gas MBTU
0
250000
500000
750000
0.0000
0.0005
0.0010
0.0015
0.0020
0.0025
RFGWINPerSf
RFGWINPerSf - Transformed
Logscale Natural Gas MBTU
5
10
0
1
RFGWINPerSf
DRYCL.1 - Untransformed
Natural Gas MBTU
0
250000
500000
750000
0
1
DRYCL.1
DRYCL.1 - Transformed
Logscale Natural Gas MBTU
5
10
0
1
DRYCL.1
PBAPLUS.32 - Untransformed
Natural Gas MBTU
0
250000
500000
750000
0
1
PBAPLUS.32
PBAPLUS.32 - Transformed
Logscale Natural Gas MBTU
5
10
0
1
PBAPLUS.32

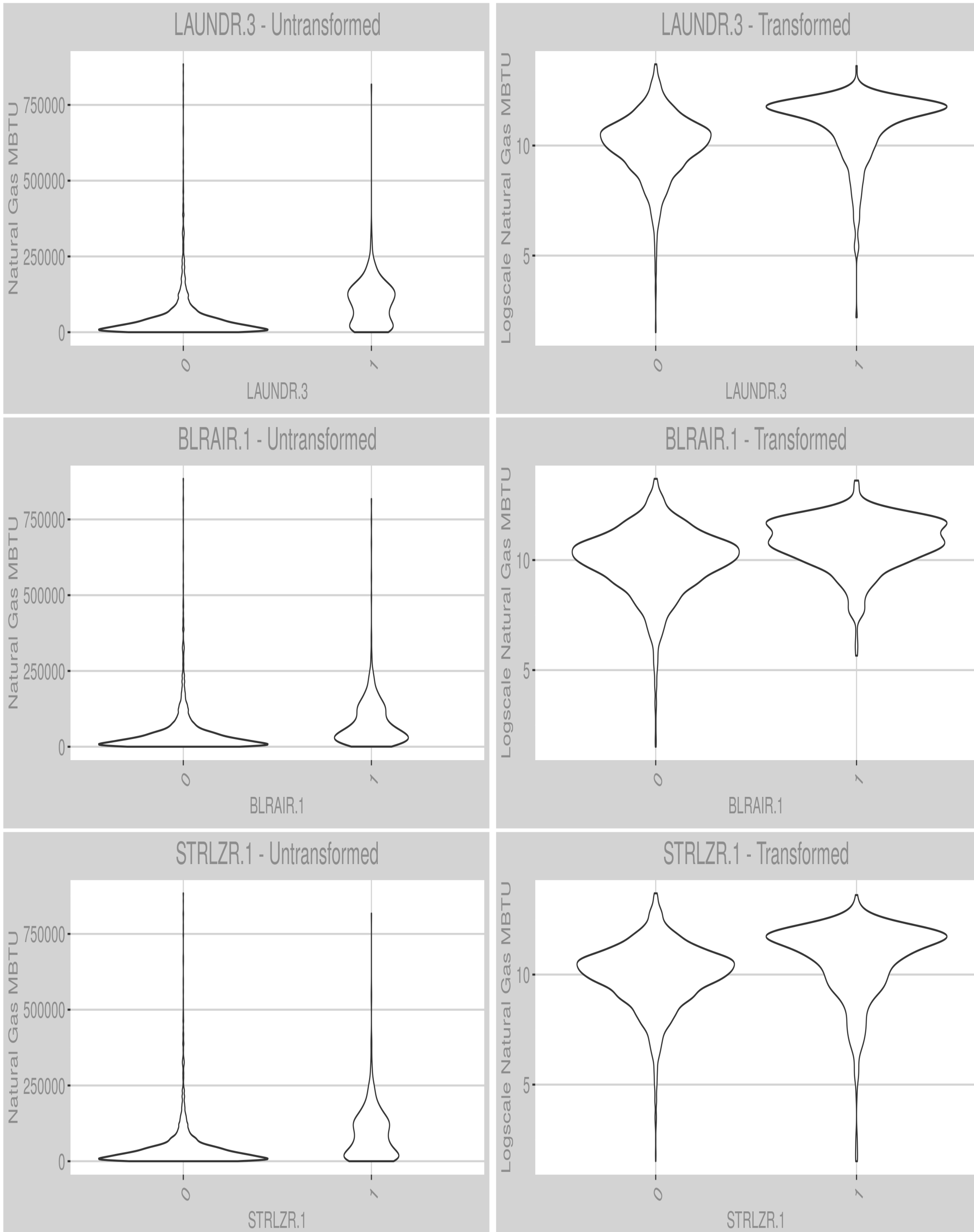

NWKERPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.00
0.05
0.10
0.15
NWKERPerSf
NWKERPerSf - Transformed
Logscale Natural Gas MBTU
10
5
0
4
8
12
NWKERPerSf
TVVIDEONPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.0000
0.0025
0.0050
0.0075
0.0100
TVVIDEONPerSf
TVVIDEONPerSf - Transformed
Logscale Natural Gas MBTU
10
5
-1
0
1
2
TVVIDEONPerSf
NGCOOK.2 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
NGCOOK.2
NGCOOK.2 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
NGCOOK.2



RFGICNPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.000
0.001
0.002
0.003
RFGICNPerSf
RFGICNPerSf - Transformed
Logscale Natural Gas MBTU
10
5
0
1
RFGICNPerSf
NGWATR.2 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
NGWATR.2
NGWATR.2 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
NGWATR.2
NFLOORPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.0000
0.0025
0.0050
0.0075
0.0100
NFLOORPerSf
NFLOORPerSf - Transformed
Logscale Natural Gas MBTU
10
5
-2
0
2
NFLOORPerSf

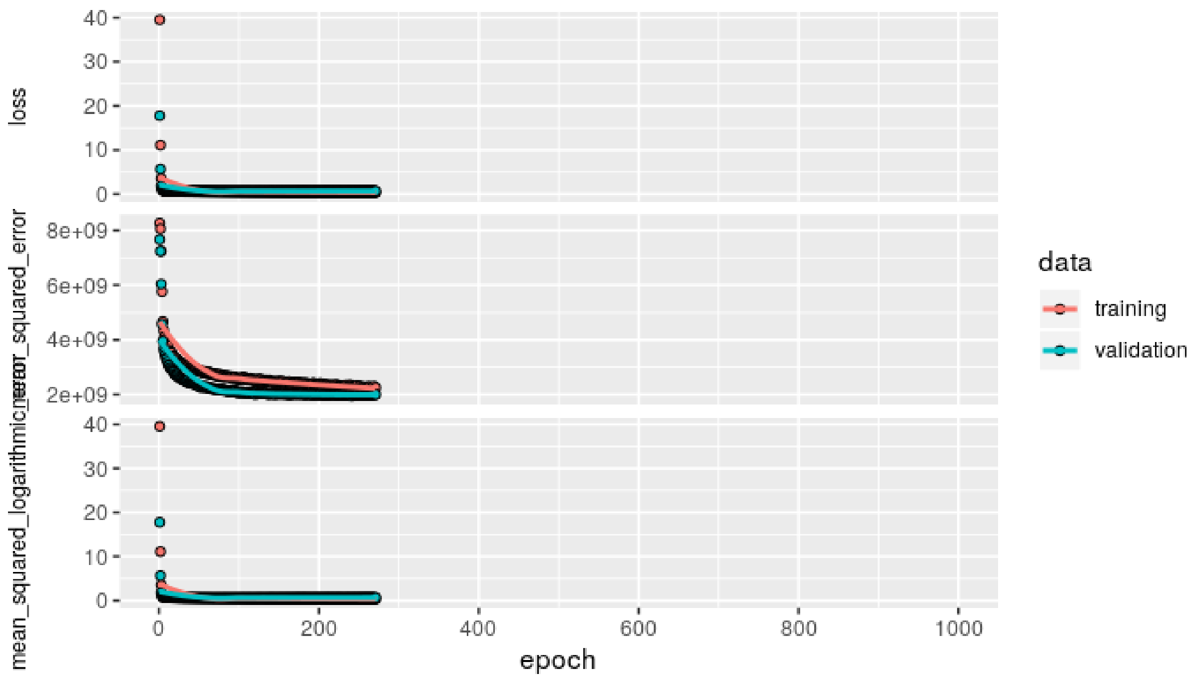

NOCCPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.000
0.002
0.004
0.006
0.008
NOCCPerSf
NOCCPerSf - Transformed
Logscale Natural Gas MBTU
10
5
-3
-2
-1
0
1
2
NOCCPerSf
SQFTC.2 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
SQFTC.2
SQFTC.2 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
SQFTC.2
OPNWE.2 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
OPNWE.2
OPNWE.2 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
OPNWE.2

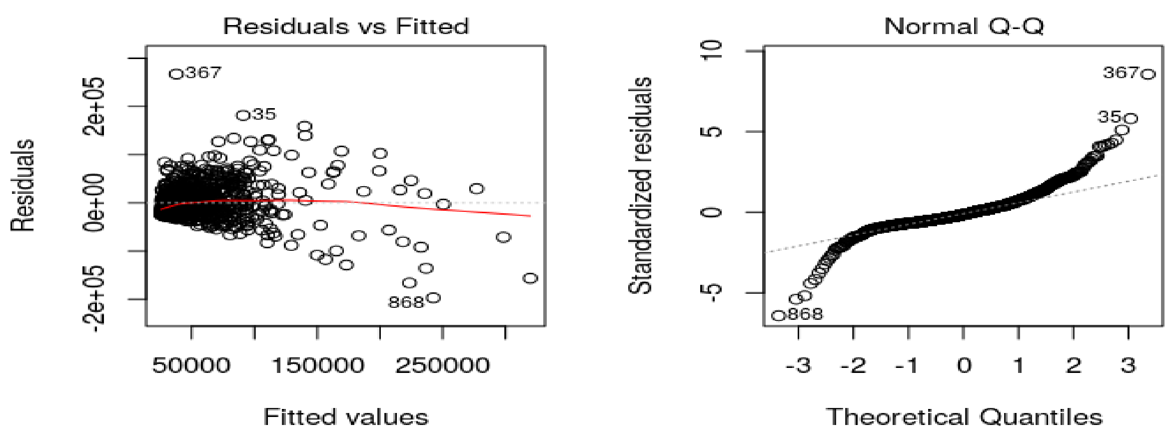

RFGRSNPerSf - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0.0000
0.0025
0.0050
0.0075
0.0100
RFGRSNPerSf
RFGRSNPerSf - Transformed
Logscale Natural Gas MBTU
10
5
-1
0
1
2
RFGRSNPerSf
RFGWI.2 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
RFGWI.2
RFGWI.2 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
RFGWI.2
HWTRM.2 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
HWTRM.2
HWTRM.2 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
HWTRM.2

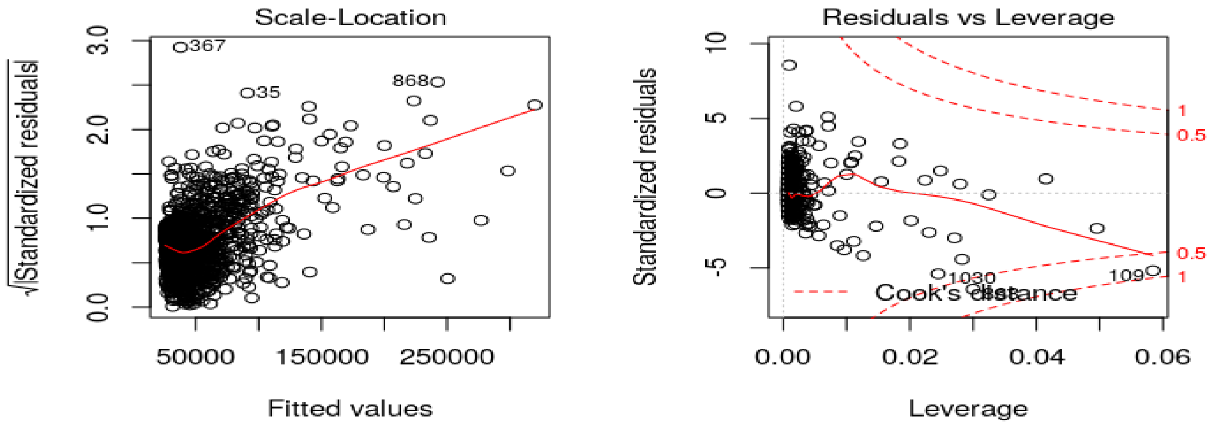

LAUNDR.3 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
LAUNDR.3
LAUNDR.3 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
LAUNDR.3
BLRAIR.1 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
BLRAIR.1
BLRAIR.1 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
BLRAIR.1
STRLZR.1 - Untransformed
Natural Gas MBTU
750000
500000
250000
0
0
1
STRLZR.1
STRLZR.1 - Transformed
Logscale Natural Gas MBTU
10
5
0
1
STRLZR.1

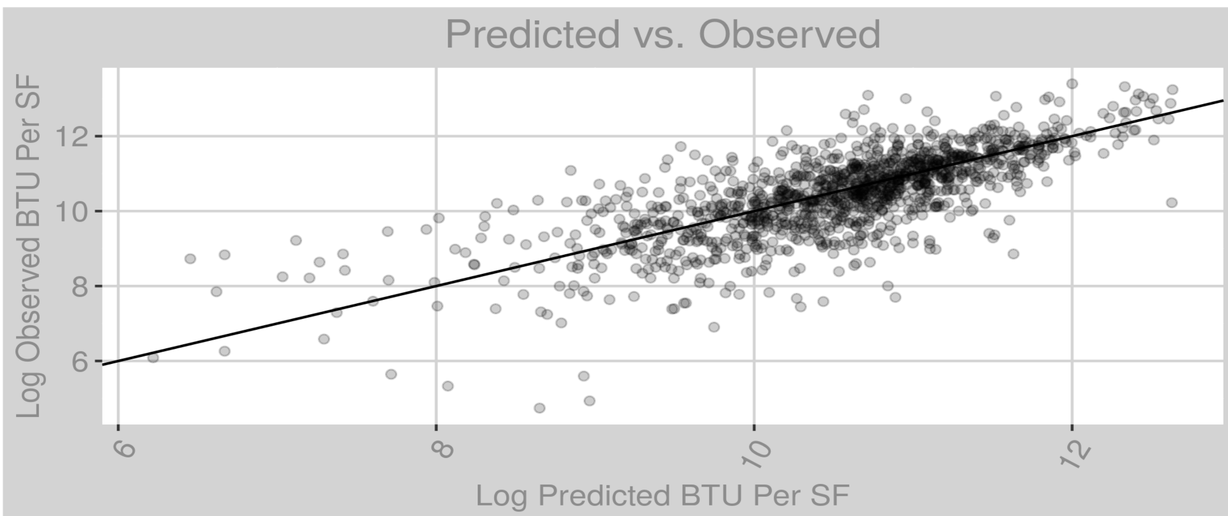
Appendix - Neural Networks

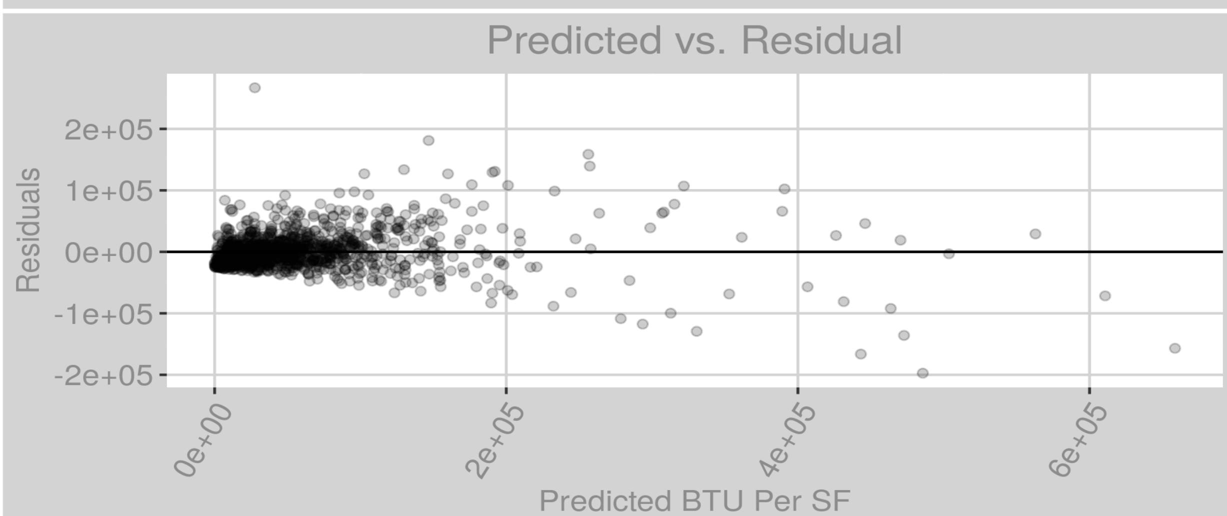
Electricity

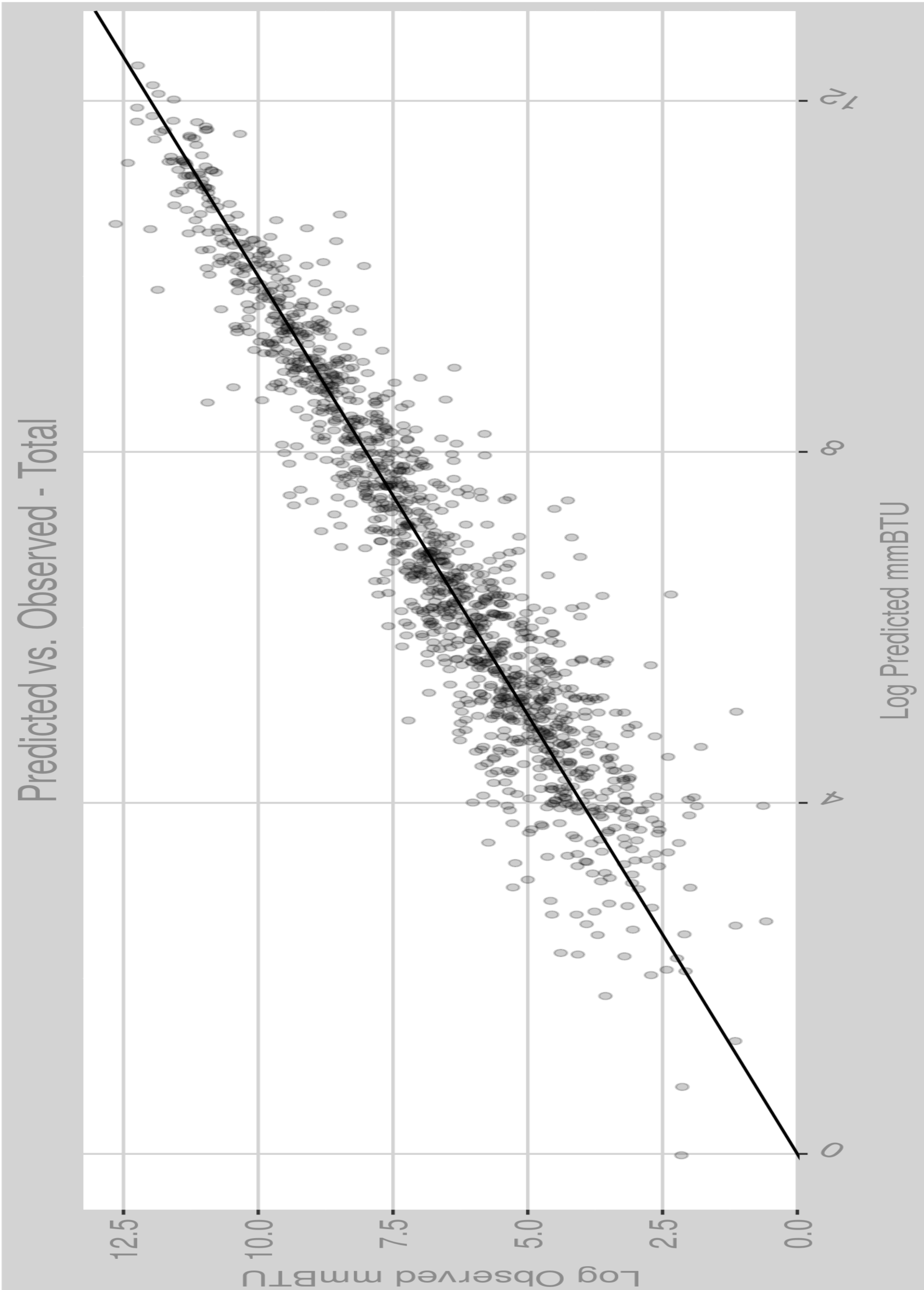





Scale-Location
√|Standardized residuals|
0.0
1.0
2.0
3.0
367
35
868
50000
150000
250000
Fitted values
Residuals vs Leverage
Standardized residuals
-5
0
5
10
1
0.5
0.5
1
1030
109
Cook's distance
0.00
0.02
0.04
0.06
Leverage


Predicted vs. Observed
Log Observed BTU Per SF
6
8
10
12
Log Predicted BTU Per SF


Predicted vs. Residual
Residuals
2e+05
1e+05
0e+00
-1e+05
-2e+05
0e+00
2e+05
4e+05
6e+05
Predicted BTU Per SF


Predicted vs. Observed - Total
Log Predicted mmBTU
0
4
8
12
Log Observed mmBTU
0.0
2.5
5.0
7.5
10.0
12.5

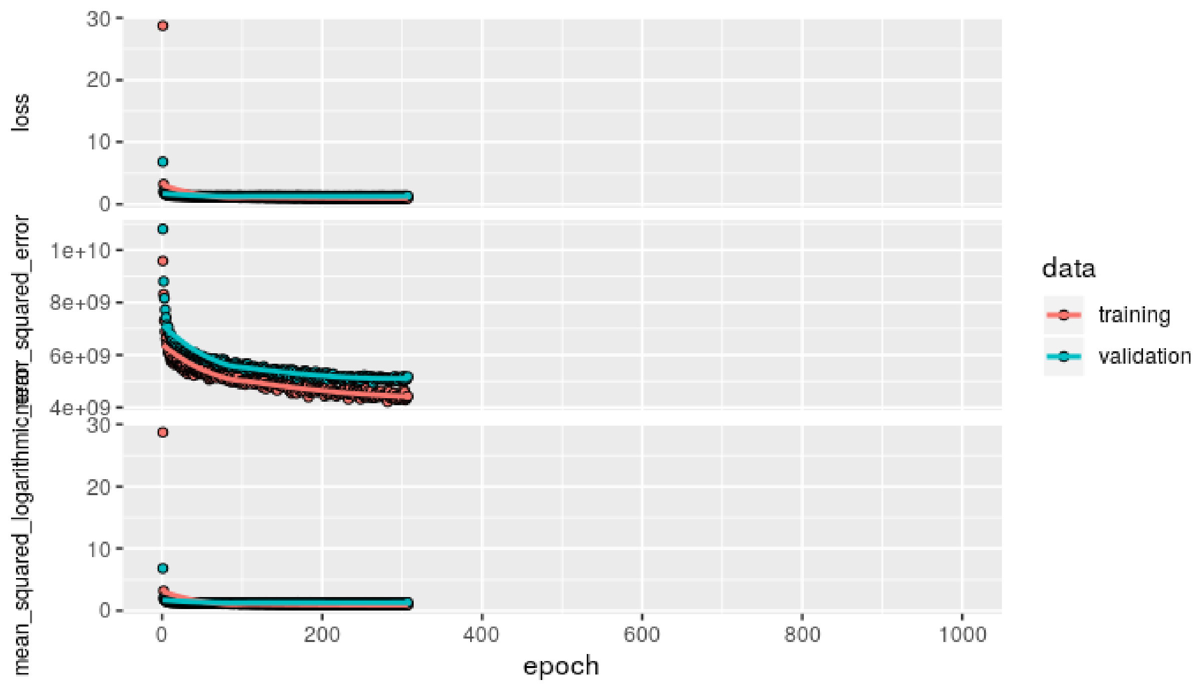
Selected Variables

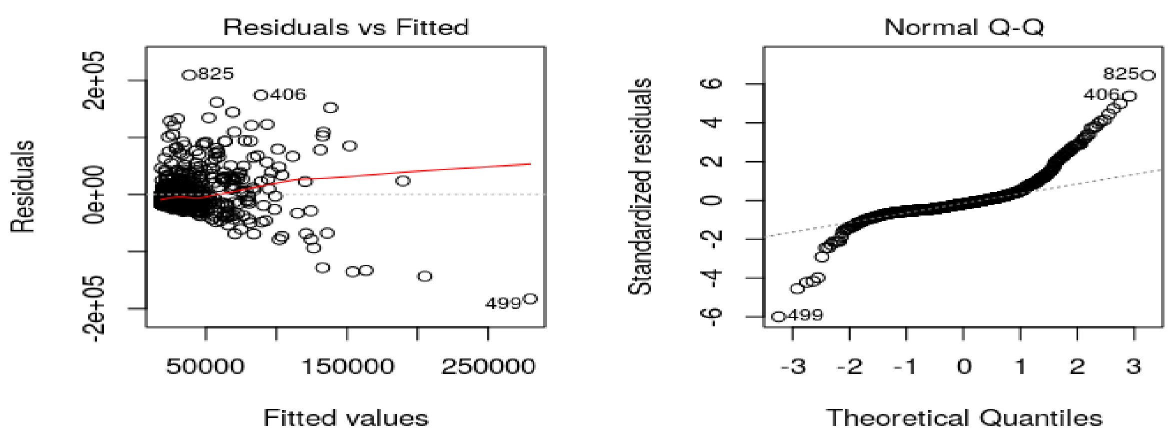
RFGWIN	Number of walk-in units	TVVIDEON	Number of TV or video displays	WKHRSC	Weekly hours category	PKLT	Lighted parking area
NWKER	Number of employees	HWTRM	Large amounts of hot water	GLSSPC	Percent exterior glass	FAX	FAX machines
RFGWI	Walk-in refrigeration units	RFGWI	Walk-in refrigeration units	CLVOAS	Cooling ventilation: Dedicated outside air system	OPNMF	Open during week
FDSEAT	Food service seating capacity	OPNMF	Open during week	LAUNDR	Laundry onsite	NGCOOK	Natural gas used for cooking
RGSTRN	Number of cash registers	EVAPCL	Evaporative or swamp coolers	CFLRP	Percent lit by compact fluorescent	RFGCOMPN	Number of compact refrigerators
PBAPLUS	More specific building activity	NWKERC	Number of employees category	RFGVEN	Refrigerated vending machines	FKHT2	Fuel oil used for secondary heating
RFGICN	Number of ice makers	ELCOOK	Electricity used for cooking	PBAPLUS	More specific building activity	COPIERN	Number of photocopiers
RFGICE	Commercial ice makers	STRLZR	Sterilizers or autoclaves	FACACT	Type of complex	FKTYPE	Specify fuel oil, diesel, or kerosene
RGSTR	Cash registers	MAINT	Regular HVAC maintenance	PCTERM	Computers used	LNHRPC	Lit off hours category
RFGCLN	Number of closed case refrigeration units	WKHRSC	Weekly hours category	VACANT	Completely vacant	NWKERC	Number of employees category

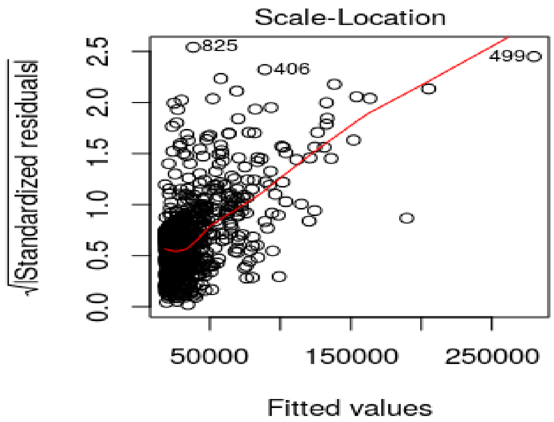
RFGCL	Closed case refrigeration units	TVVIDEO	TV or video displays	OTLT	Other type of bulbs	HCBED	Licensed bed capacity
COOK	Energy used for cooking	RFGRES	Full-size residential-type refrigerator	FASTFD	Fast food or small restaurant	CFLR	Compact fluorescent bulbs
NGCOOK	Natural gas used for cooking	GENUSE	Use of generated electricity	DCNTRSFC	Data center or server farm sqft category	ELEVTR	Elevators
PCTERMN	Number of computers	HEATP	Percent heated	RFTILT	Roof tilt	CHLPKG	Chiller system: Packaged unit
LNHRPC	Lit off hours category	MRI	MRI machines	TVVIDEON	Number of TV or video displays	CHLAIRCL	Chiller type: Air-cooled
PBA	Principal building activity	BOOSTWT	Booster water heaters	PBAPLUS	More specific building activity	LTEXPC	Percent of exterior lighted
RFGOPN	Number of open case refrigeration units	LTNR24	Lights off during 24 hours	LINACC	Linear accelerators	RFGCOMP	Half-size or compact refrigerators
PRNTRN	Number of printers	SERVERN	Number of servers	LABEQP	Laboratory equipment	KITCHN	Small kitchen area
RFGOP	Open case refrigeration units	XRAYN	Number of X-ray machines	WHRECOV	Waste heat recovery	TRIM	High-end trimming or light-level tuning
PBA	Principal building activity	ANYEGY	Any energy used	RFGICE	Commercial ice makers	GLSSPC	Percent exterior glass

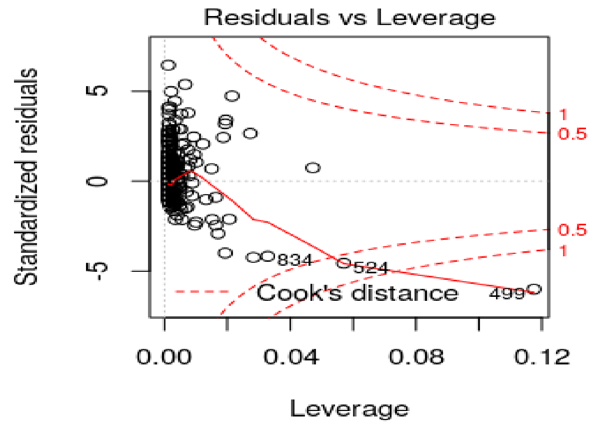
WKHRSC	Weekly hours category	LEDP	Percent lit by LED	MCHEQP	Machine equipment	RFGSTO	Large cold storage areas
OPNWE	Open on weekend	RFGVNN	Number of refrigerated vending machines	MONUSE	Months in use	CHLFNCL	Chiller system: Fan coil units in rooms
LOHRPC	Lit when open category	PBAPLUS	More specific building activity	OWNTYPE	Building owner	HTVVAV	Heating ventilation: Central air handling with VAV
PBAPLUS	More specific building activity	RWSEAT	Religious worship seating capacity	RFGSTO	Large cold storage areas	NELVTR	Number of elevators
MCHEQP	Machine equipment	PBAPLUS	More specific building activity	LAUNDR	Laundry onsite	NGHT1	Natural gas used for main heating

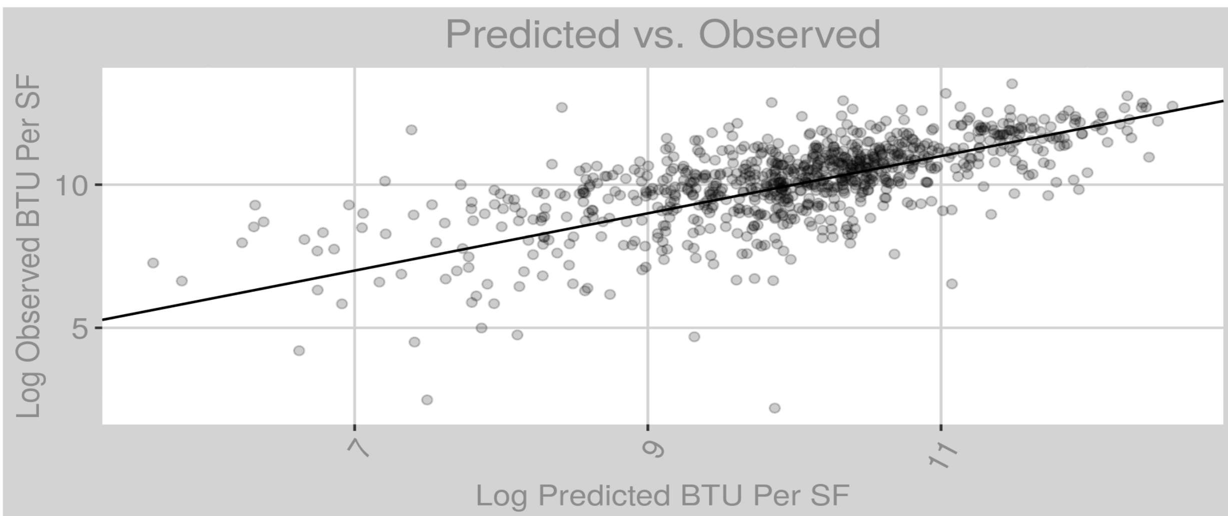
Natural Gas

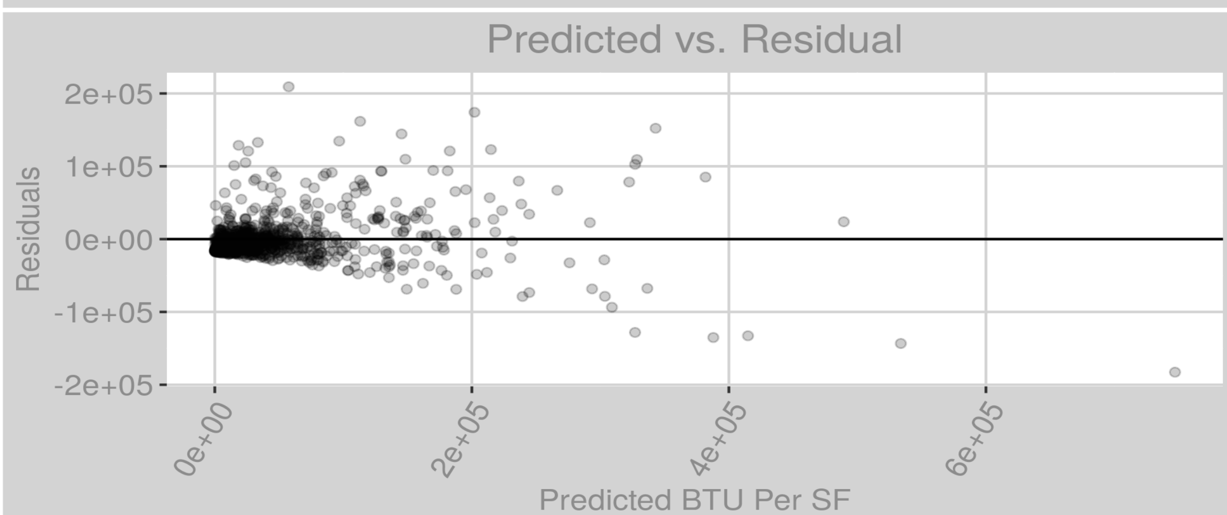

loss
mean_squared_error
mean_squared_logarithmic_error
epoch
data
training
validation

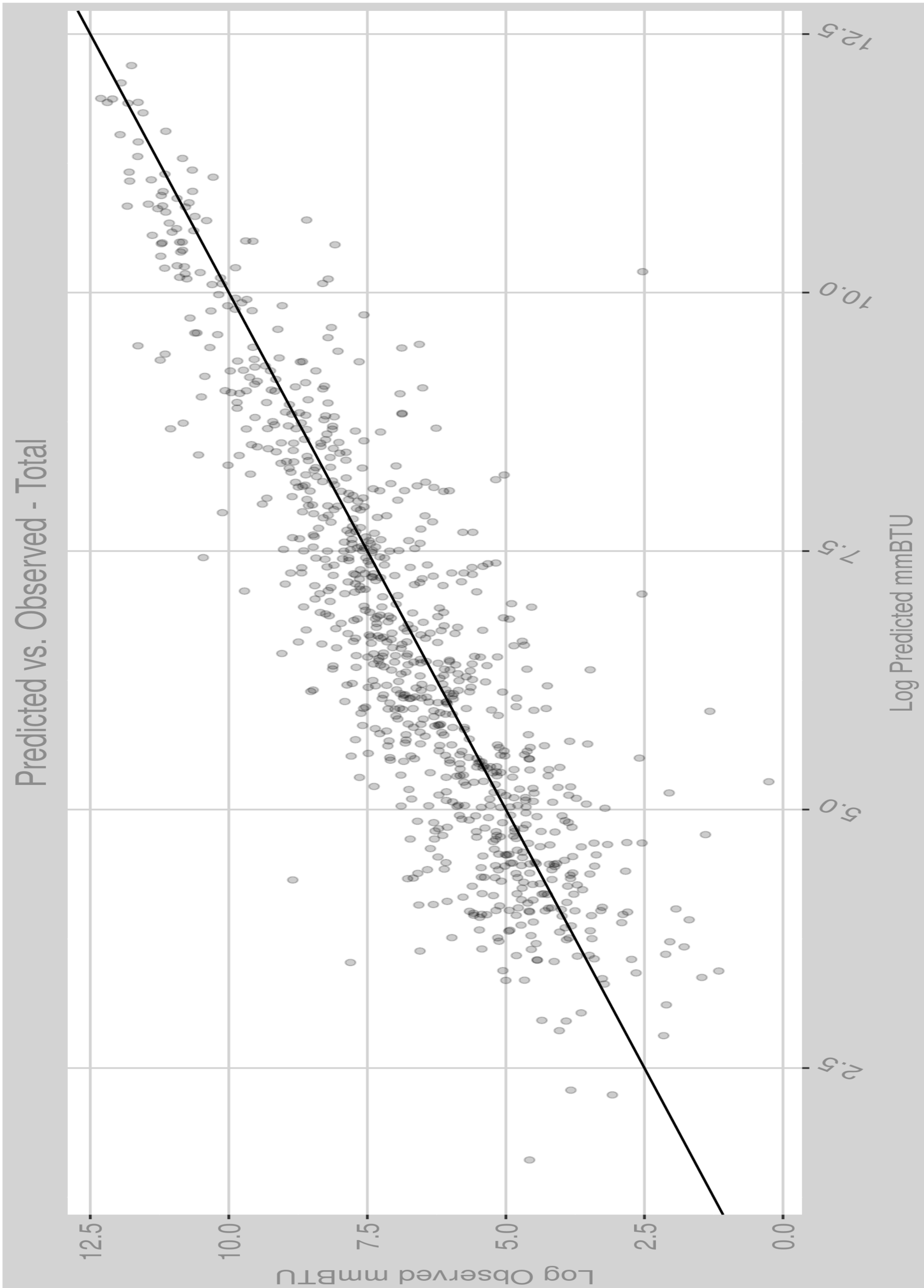

Residuals vs Fitted
Residuals
Fitted values
Normal Q-Q
Standardized residuals
Theoretical Quantiles


Scale-Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.5
825
406
499
50000
150000
250000
Fitted values


Residuals vs Leverage
Standardized residuals
5
0
-5
1
0.5
834
524
499
Cook's distance
0.00
0.04
0.08
0.12
Leverage


Predicted vs. Observed
Log Observed BTU Per SF
10
5
7
9
11
Log Predicted BTU Per SF


Predicted vs. Residual
Residuals
2e+05
1e+05
0e+00
-1e+05
-2e+05
0e+00
2e+05
4e+05
6e+05
Predicted BTU Per SF


Predicted vs. Observed - Total
Log Observed mmBTU
Log Predicted mmBTU
0.0
2.5
5.0
7.5
10.0
12.5
2.5
5.0
7.5
10.0
12.5

Selected Variables

FDSEAT	Food service seating capacity	FACACT	Type of complex	HCBED	Licensed bed capacity	CDD65	Cooling degree days (base 65)
RGSTRN	Number of cash registers	RGSTR	Cash registers	DHPKG	District heat system: Packaged unit	WTHTEQ	Water heating equipment
PBAPLUS	More specific building activity	PBA	Principal building activity	BLRRAD	Boiler system: Radiators	LNHRPC	Lit off hours category
RFGWIN	Number of walk-in units	XRAYN	Number of X-ray machines	STRLZR	Sterilizers or autoclaves	DHFNCL	District heat system: Fan coil units in rooms
DRYCL	Dry cleaning onsite	BLRFNCL	Boiler system: Fan coil units in rooms	PBAPLUS	More specific building activity	WTHTEQ	Water heating equipment
PBAPLUS	More specific building activity	BLRDUCT	Boiler system: Duct reheat	WKHRSC	Weekly hours category	NGCOOL	Natural gas used for cooling
NWKER	Number of employees	PRNTRN	Number of printers	PCTRMC	Number of computers category	LINACC	Linear accelerators
TVVIDEON	Number of TV or video displays	RFGCLN	Number of closed case refrigeration units	DHDUCT	District heat system: Duct reheat	BLRRAD	Boiler system: Radiators
NGCOOK	Natural gas used for cooking	LTNR24	Lights off during 24 hours	BLDPLT	Central plant in building	NGOTH	Natural gas for some other use
RFGICN	Number of ice makers	COPIER	Photocopiers	MCHEQP	Machine equipment	WKHRSC	Weekly hours category

NGWATR	Natural gas used for water heating	HCBED	Licensed bed capacity	NGSRC	How purchase natural gas	ELHT1	Electricity used for main heating
NFLOOR	Number of floors	BLRPKG	Boiler system: Packaged unit	DATACNTR	Data center or server farm	BLRINDC	Boiler system: Induction units
NOCC	Number of businesses	HEATP	Percent heated	PCTERMN	Number of computers	RFGOPN	Number of open case refrigeration units
SQFTC	Square footage category	DHRAD	District heat system: Radiators	STDNRM	Student or public computer center	TRNGRM	Computer-based training room
OPNWE	Open on weekend	WKHRSC	Weekly hours category	BOOSTWT	Booster water heaters	PBAPLUS	More specific building activity
RFGRSN	Number of residential refrigerators	LINACC	Linear accelerators	RFGCOMPN	Number of compact refrigerators	ELCOOK	Electricity used for cooking
RFGWI	Walk-in refrigeration units	BLRLOOP	Boiler system: Water loop heat pump	WBOARDS	Interactive whiteboards	RFGOP	Open case refrigeration units
HWTRM	Large amounts of hot water	ELWATR	Electricity used for water heating	PUBCLIM	Building America climate region	CUBELOC	Location of open plan
LAUNDR	Laundry onsite	CUBELOC	Location of open plan	RFGRES	Full-size residential-type refrigerator	GLSSPC	Percent exterior glass
BLRAIR	Boiler system: Central air handler	PBAPLUS	More specific building activity	DHFNCL	District heat system: Fan coil units in rooms	FDPREP	Commercial or large kitchen

STRLZR	Sterilizers or autoclaves	BOOSTWT	Booster water heaters	SERVERN	Number of servers	STDNRM	Student or public computer center
STCOOK	District steam used for cooking	RFGICE	Commercial ice makers	LABEQP	Laboratory equipment	DHDUCT	District heat system: Duct reheat
PBA	Principal building activity	LAPTPN	Number of laptops	PBAPLUS	More specific building activity	LAPTPC	Number of laptops category
BOOSTWT	Booster water heaters	TVVIDEON	Number of TV or video displays	CHLDUCT	Chiller system: Duct reheat	CHLFNCL	Chiller system: Fan coil units in rooms
STWATR	District steam used for water heating	LABEQP	Laboratory equipment	LNHRPC	Lit off hours category	CUBEC	Percent open plan