ELECTRICITY DEMAND PREDICTION

Ankur Wadhwa
wadhwa5@purdue.edu

Contents

Total Site Electricity Usage Prediction

Ankur Wadhwa
wadhwa5@purdue.edu

---

1. Introduction

We all know that electricity plays an important role in everyone's life from individual level to national level. It is one of the most important factor in accelerating growth of a nation. But due to rising demand for electricity we are forced to produce more and in the process, we are harming nature.

This work aims to investigate the significant factors which affect the total site electricity usage in residential areas so that we can educate the residents to use the electricity effectively. To achieve this aim I first did the literature review to know which all factors are currently relevant for the analysis. Then I included some more factors which I found statistically significant. With the use of predictive analytics principal, I found which factors affect the most.

I have only addressed factors which are major drivers of electricity usage in West Region at residential level. Factors can be broadly classified as socio-economic, dwelling and appliance related. Socio-economic factors mainly consist of number of occupants, family composition like number of children below 16 etc, employment status, educational level. Dwelling mainly consist of age of house or apartment, number of rooms, number of floors, total area, do they have HVAC system, total cooling and heating area, presence of low energy lightings. Appliance includes all the gadgets like PC, DVD player, washer, cooking appliances.

2. Literature review

According to previous study in this area, factors are mainly classified as Socio-Economic, Dwelling and Appliance. Either Statistical or Econometric methods are commonly implemented to find out the major predictors which affect the electricity usage. Tso and Yau[3] applied machine learning techniques such as regression analysis, decision tree and neural network for the prediction of electric energy consumption on a dataset of1516 households in Hong Kong China. For Residential Energy Consumption Survey (RECS), Carlson et al. [2] analysed that how many domestic appliances contribute to the household electricity usage. RV Jones [5] have compared all the current studies which are done in different countries. There are many studies at state level which have shown us the effect of different factors.

3. Data Overview

3.1 Source

Our main data source is the dataset floated by U.S. Energy Information Administration (Dataset) and for the purpose of final exam I have only considered the West region of Residential Data (RECS) of year 2005. Total data set contain 4384 observations and 1075 variables. After filtering it to west region I was left with 967 observations.

3.2 Variable Selection

Studies outlined in the above section and in references has shown us a range of factors which have positive and negative effect on total site electricity usage. By utilizing the previous research and my knowledge following factors have been taken in account for this study:-

- Socio-Economic Factors: (i) number of occupants, (ii) family composition (number of children, number of teenagers, number of adults and number of old people), (iii)Age of household responsible person (HRP), (iv) Employment status and median income of HRP, (v) Disposable income.
- Dwelling Factors: (i) Dwelling type (type of home whether it is apartment or town house etc), (ii) Dwelling age (When this house was built), (iii) Number of rooms, (iv)Number of Bedrooms, (v)Total Floor Area (vi)Area used for heating and cooling, (vii) Presence of low energy lighting
- Appliance Factors : (i) How many PC house have, (ii) HVAC appliance, (iii) Cooking appliance (like microwave, oven, toaster), (iv) preservation and cooling appliance (like refrigerator, freezer), (v) Washing and Drying equipment, (vi) how many appliances are plugged in whole day.

3.4 Feature Engineering

After observing the data, I found that there were no predictors which could tell us the number of adults, teenagers, old age persons and children in a house. So I have created four variables namely “CHILDC”,”TEENC”,”ADULTC” and “OLDC” which mean number of children (below age of 13yr are considered children), number of teenager (age group of 13-19yr), number of adults (from 20-65yr) and number of old people (above 65yr) respectively. These counts were made using the already present variables which share the age of every person in a house.

Survey was designed in a questionnaire manner, so if a person does not have anything at a particular level then questions related to that will have a legitimate skip. But when we are doing our analysis it can cause problem. Also, our data would have some missing values. So at many places for missing data and skip answer I read the questionnaire and tried to replace extremely large value which was earlier assigned to zero. For example, if someone does not have a swimming pool how can she even have a filter system for the swimming pool. But according to a survey it was a skip so a large value was given like 99 was replaced with 0. Now our new coding will be that if a person has marked 0 for swimming pool then filter system is also 0.

In many fields of our data as in code sheet it was written that they are coded with 99 but when actually checking the data I found it to be either NA or 9. So I had to deal with those too and for that I checked the imputed column given in the data set whether they are imputed or not. Then I replaced the value 0 depending upon the previously explained concept. But in few cases I was not able to decide, to be precise 12 observations, so I decided to remove them.

I have also standardized the variable which could be correlated to factors such as total area, total heating area and total cooling area. In many cases they can be same while for others they can be different. But since we required all three so best way to protect their individuality was to standardize them.

3.6 Final dataset

Final variable used for analyzing the data.

1	NHSLDMEM	How many people normally live in this household
2	CHILDC	Total number of chidlren in house (below 13 yr)
3	TEENC	Total number of teenager in house (13-19yr)
4	ADULTC	Total number of adutls (20-65yr)
5	OLDC	Total number of old age people in house (above 65yr)
6	AGEHHMEM	Age of the householder
7	EMPLOYHH	How would you describe employment status
8	KOWNRENT	Dwelling owned or rented
9	HHINCOME	Mid-point of MONEYPY range
10	HOWPAYEL	How electricity is paid
11	AGEHHMEM	Age of the yountest household member
12	NUMPC	How many PCs do you use
13	TVCOLOR	How many TV sets do you use in your home
14	TVONWD	Hours TV on during weekdays
15	TVONWE	Hours TV on during week end days
16	VCR	How many VCR players in your home
17	DVD	How many DVD players in your home
18	NUMCFAN	How many ceiling fans does your household use
19	USENOTMO	Number of months dehumidifier used each year
20	OVENUSE	How often you use your oven
21	AMTMICRO	How often is microwave used
22	NUMFRIG	How many refrigerators do you use in your home
23	NUMFREEZ	How many separate freezers are used in your home
24	DISHWASH	Does your household use an automatic dishwasher
25	DWASHUSE	How often automatic dishwasher used
26	CWASHER	Use a clothes washer in home
27	WASHLOAD	How many loads of laundry are washed in a week
28	DRYER	Do you use a clothes dryer in home
29	DRYRUSE	How often use your clothes dryer
30	SWIMPOOL	Does your own swimming pool with filtering system
31	TYPEHUQ	Type of Home (Respondent answer)
32	YEARMADE	Year Home Built
33	TOTROOMS	Total number of bedrooms and other rooms
34	BEDROOMS	Number of bedrooms
35	TOTSQFT	Total square footage
36	EQUIPM	Type of heating equipment provides the heat
37	LGT12	Lights are turned on more than 12 hours per day
38	LGT12EE	Energy efficient bulbs on more than 12 hours per day
39	LGT4	Lights on between 4 hours and 12 hours per day
40	LGT4EE	Energy efficient bulbs on 4-12 hours per day
41	LGT1	Lights on between 1 hours and 4 hours per day
42	LGT1EE	Energy efficient bulbs on 1-4 hours per day
43	TOTHSQFT	Total heated square footage
44	TOTCSQFT	Total air-conditional square footage
45	URBRUR	Reported type of neighborhood
46	HD65	Heat Deg-Days: Base65 01 TO 12-2005 (Inoculated)
47	CD65	Cool Deg-Days: Base65 01 TO 12-2005 (Inoculated)
48	TEMPHOME	Setting During the day when someone is home
49	TEMPGONE	Setting During the day when no one is home
50	TEMPNITE	Setting During sleeping hours
51	WINDOWS	Number of windows
52	TYPEGLASS	Type of window glass
53	SPOUSE	Household living with a spouse/partner
54	WALLTYPE	Exterior Wall Construction Material
55	HHAGE	How old are you/is the householder
56	KWH	Kilowatt Hours Of Electricity Used
55	HHAGE	How old are you/is the householder
56	KWH	Kilowatt Hours Of Electricity Used

Figure 1: Final selected predictors and their meaning

4. Exploratory Data Analysis

I performed various exploratory data analysis to understand our dataset better and identify the relationships that exist between predictors and response variable. The analysis and insights drawn from them are mentioned below.

- The distribution of the response variable in our dataset is shown in the above graph. We can see that data is highly skewed and is not normal. To confirm the same, I also drew a QQ-plot which is attached in the appendix.

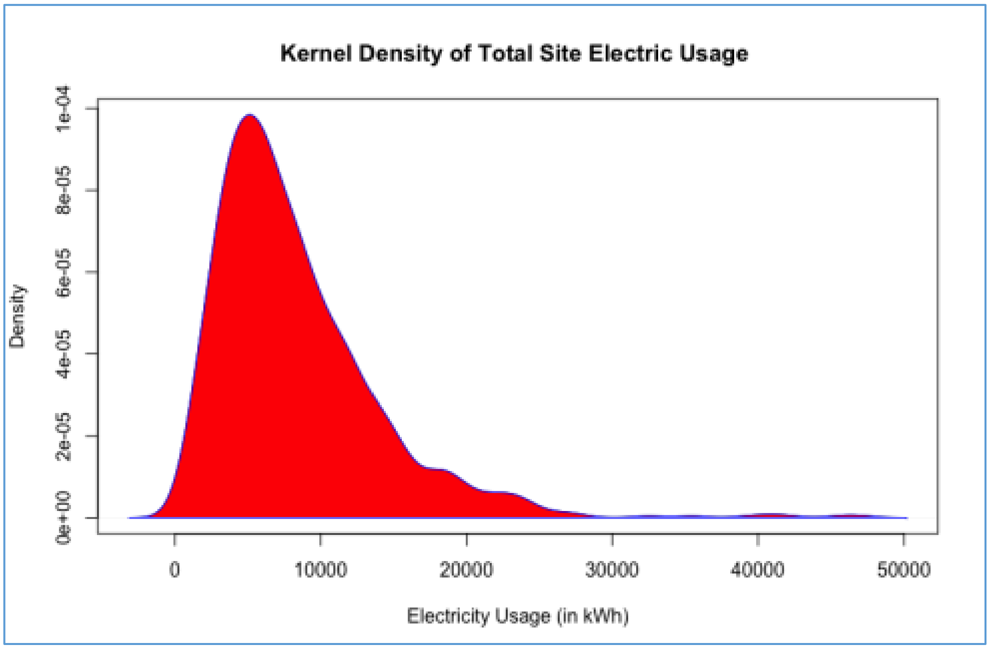


Figure 2: Density plot of our response variable i.e. Total Site Electric Usage

- This violin plot shows us the density and median value of Heading days and cooling days with 65F as reference point.

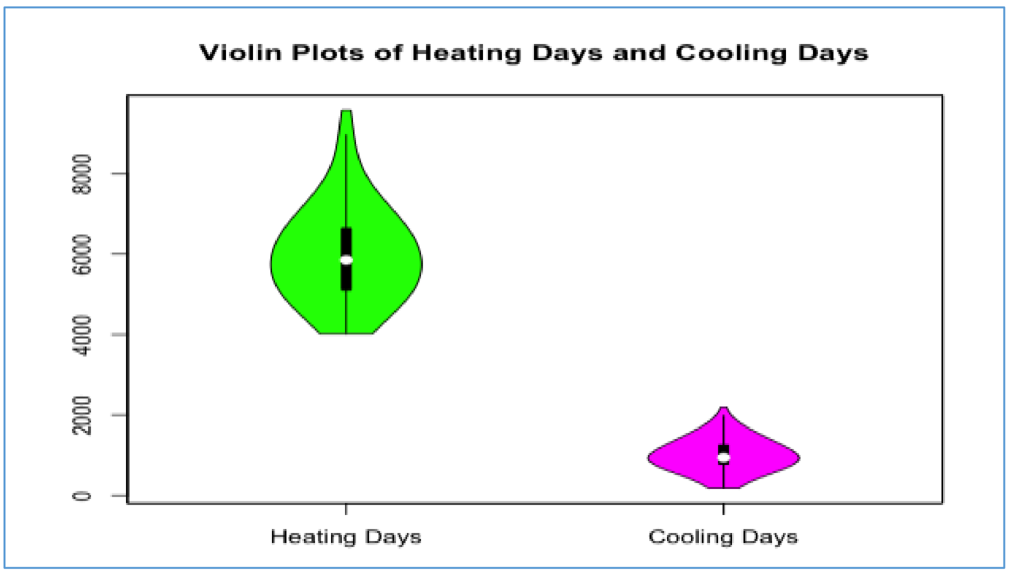


Figure 3: Violin plot of total heating and cooling days with reference of 65^{o}F

➢ Bar plots show the average usage of electricity (in kWhr) of the houses which were built in different years. We can observe that houses which were constructed till 1959 have almost the same usage whereas there was a modest increase in usage till the year 1990-1994 when energy usage shot up. After that it again become relatively stable. One of the reasons for the rise in electricity usage was increased use of personal computer and other home appliances. After a certain time these equipment started to become more efficient then again the usage of electricity dropped.

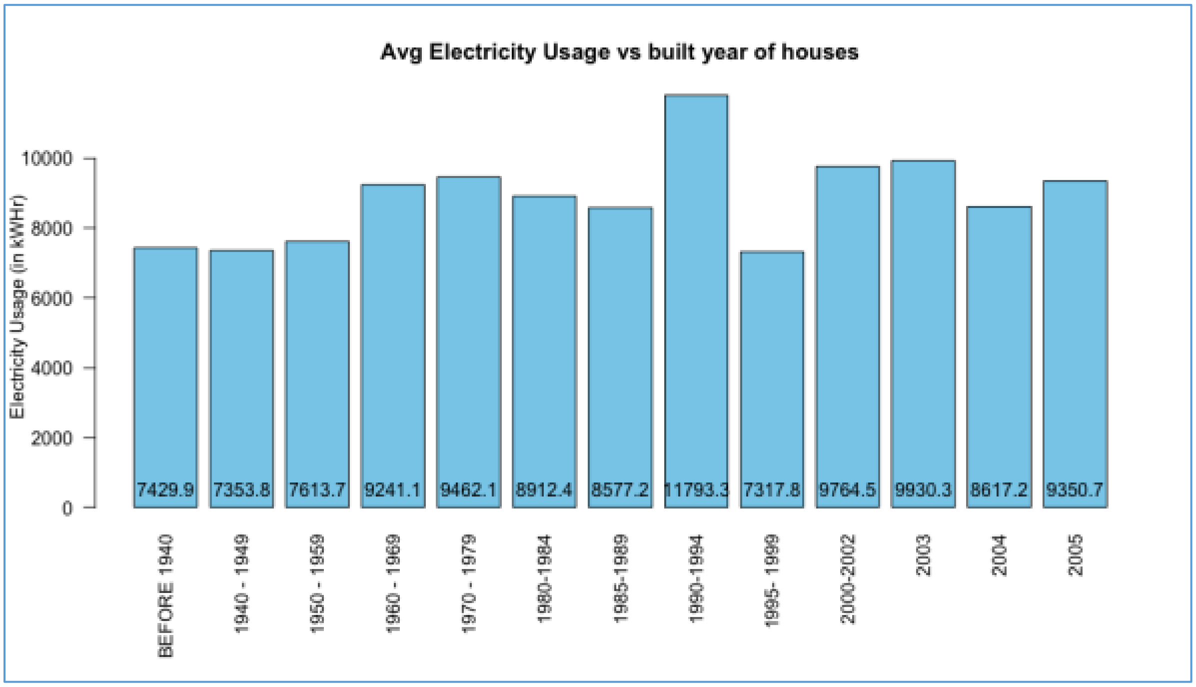


Figure 4: Bar plot of Total Site Electric Usage by houses built in different year

➢ Graph below show us the House hold income distribution in the west region.

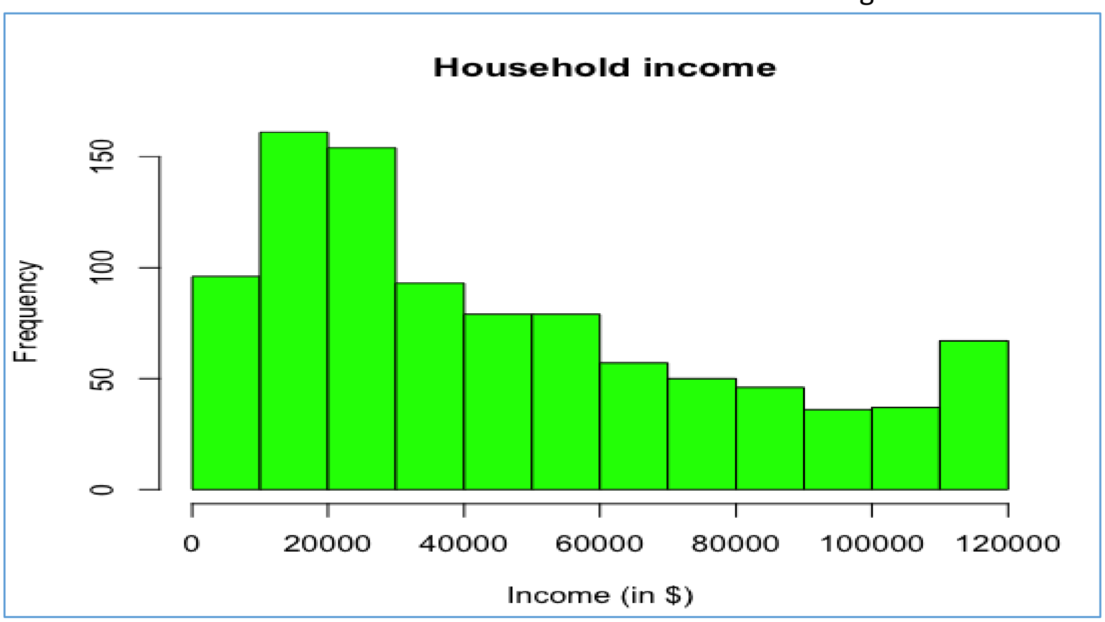


Figure 5: Histogram of Household income of families

➢ Graph show us that most of the family has 1-3 TV sets in their house, and we know that TV is one of the equipment which is used almost whole day, so I think It will impact the total site electricity usage.

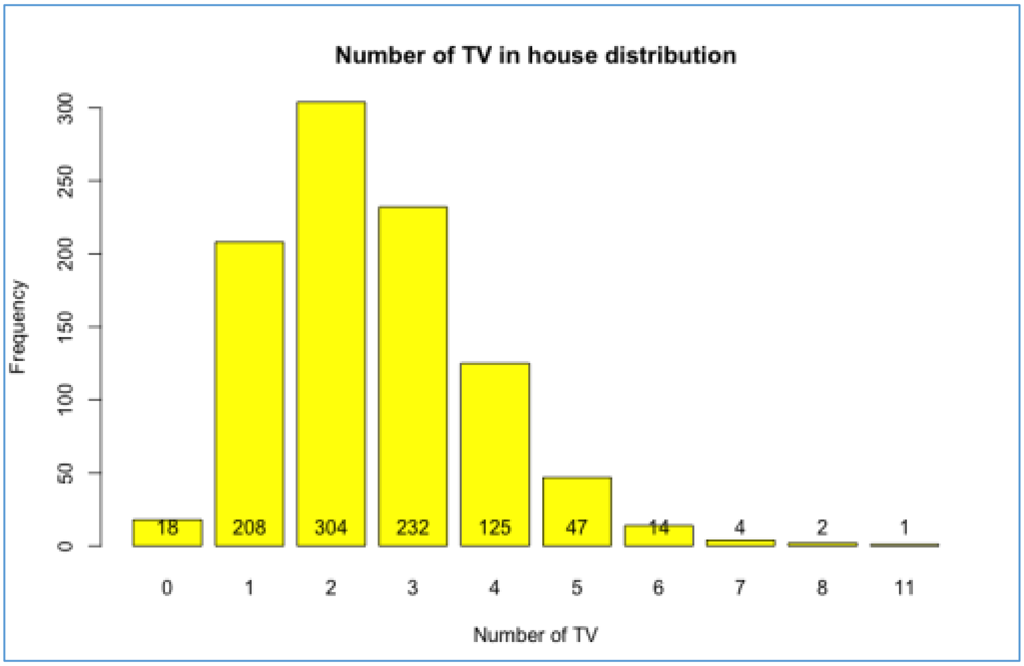


Figure 5: Distribution of Number of TV in families.

➢ From below graph we can see that in west region maximum people live in cities. We can infer that most people who gave survey lives in big cities or town.

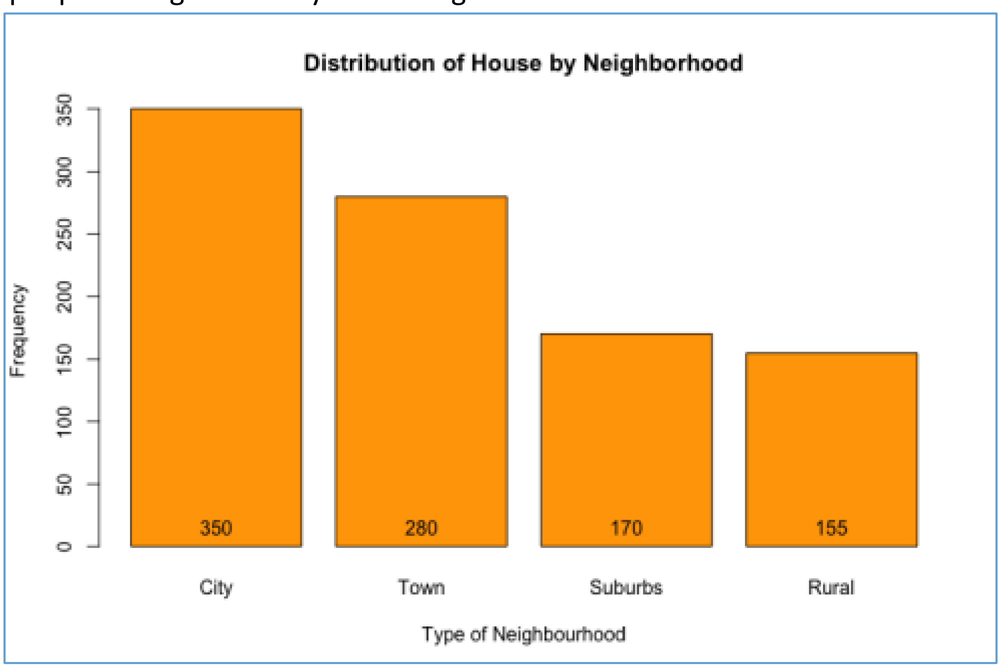


Figure 6: Bar plot showing number of families in different localities (like rural, suburbs, town, city)

- The violin plot shows the distribution of temperature setting of houses when someone is at home, when no one is at home and during sleep time. From this we can observe, that Temperature of home when no one is at home is usually less than when someone is at home but it is almost same as night temperature.

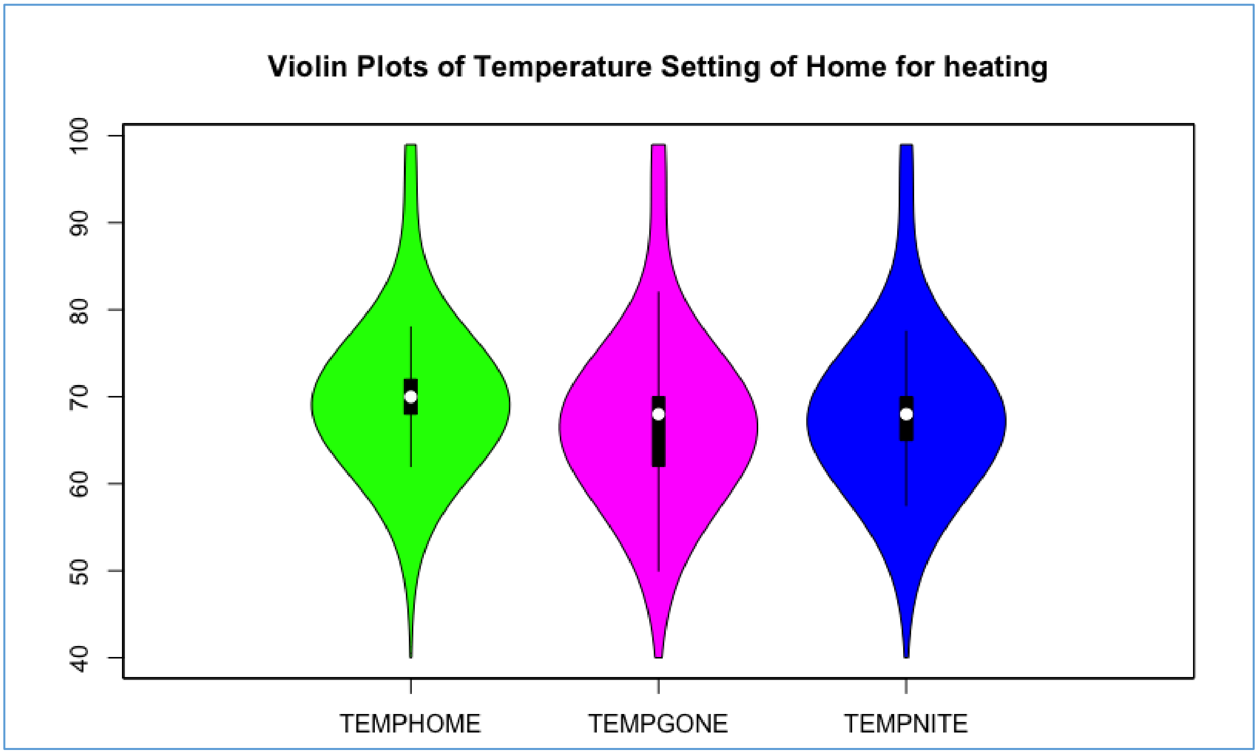


Figure 7: Violin plot of different temperature maintained at home at different time

5. Methodology

Since my objective is to predict the Total Site Electricity Energy Usage (in KWH), it is a regression problem. And while undergoing the literature review I came to know what all techniques have been used in the past. First of all I checked the normality assumption of dependent variable (KWH) by plotting the QQ-plot and came to know our KWH not linear at all. So, we could not go ahead with the linear regression model. Instead of Simple Linear Regression, I started with MARS.

5.1 Multi-Adaptive Regression Splines (MARS)

MARS is a stepwise regression model consists of sum-of-splines that allow the response to vary non-linearly with input variables as shown in below equation below:

$$Y = \beta_o + \sum_{m=1}^{p} \beta_m h_m(X) + \epsilon$$

Where each *h(x)* represents the linear splines, $\beta_o$represents the intercept and $\beta_m$is vector coefficients.

5.2 Classification and regression trees (CART)

CART method use the concept of decision tree. We use Gini-coefficient to choose the splits and keep growing our tree until required depth is achieved. We also do pruning so that it does not over fit the model. But we have some disadvantage, it is a greedy algorithm while making splits, it only tries to minimise the error for next node instead of minimising overall error. It is known to over fit the data.

5.3 Bagged Classification and regression trees

In this we do bootstrap aggregation technique which is also known as bagging. It takes training sample from original data with replacement and train the model and in the end we take average of all the trees. Since it is combination of many trees, it offers a disadvantage in terms of difficulty in interpretation.

5.4 Random Forest

It is also a tree based ensemble method. In this we take *N* bootstrapped regression trees *(T_b) with m* predictors where $\approx \sqrt{p}$, where *p* is number of predictors*.* RF cover large amount of model space as compared to other tree based method and it also provides a best tradeoff between variance and bias. It has a limitation same as bagged tree, we cannot interpret it but from partial plots we can see that which variable has what kind of relationship with our independent variable.

5.5 Bayesian Additive Regression trees (BART)

BART is a tree based method which use the Bayesian concept in making the tree. Trees are trained in such a manner that each tree is constrained by a regularization prior to be a weak learner. It uses the back-fitting approach to generate samples from the posterior.

$$Y = \sum_{j=1}^{p} g\left(X; T_j M_j\right) + \in$$

5.6 Support-Vector Machine (SVM)

Earlier SVM was only used for classification problem, but now we can also use it for regression problem. In this we try to fix a plane which has maximum margin from the data point. With linear, polynomial or radial kernel option we maximise the margins.

6. Model Selection

First, I tried to optimise the each of different models by doing the parameter tuning using cross validation. Then taking all the best model, I obtained the RMSE on training data, RMSE on testing data, R-sq and 10-fold Cross Validation error. Later to select the best model I did the statistical test (Fisher Test), which compared the model using pairwise t-test and on that basis best model was selected.

Model Name	*RMSE of train data*	*CV Error*	*RMSE of test data*	*R-square value*
MARS_model	4318.29	5016.16	5066.89	0.41
CART_Model	4619.52	5339.24	5674.86	0.32
Bag_model	3586.32	4288.08	5082.53	0.59
RF_model	1958.37	3545.70	3827.20	0.89
BART_model	3233.42	4432.90	4832.56	0.67
SVM_model	4021.95	4562.66	5065.70	0.48

Table 1: Model Summary table

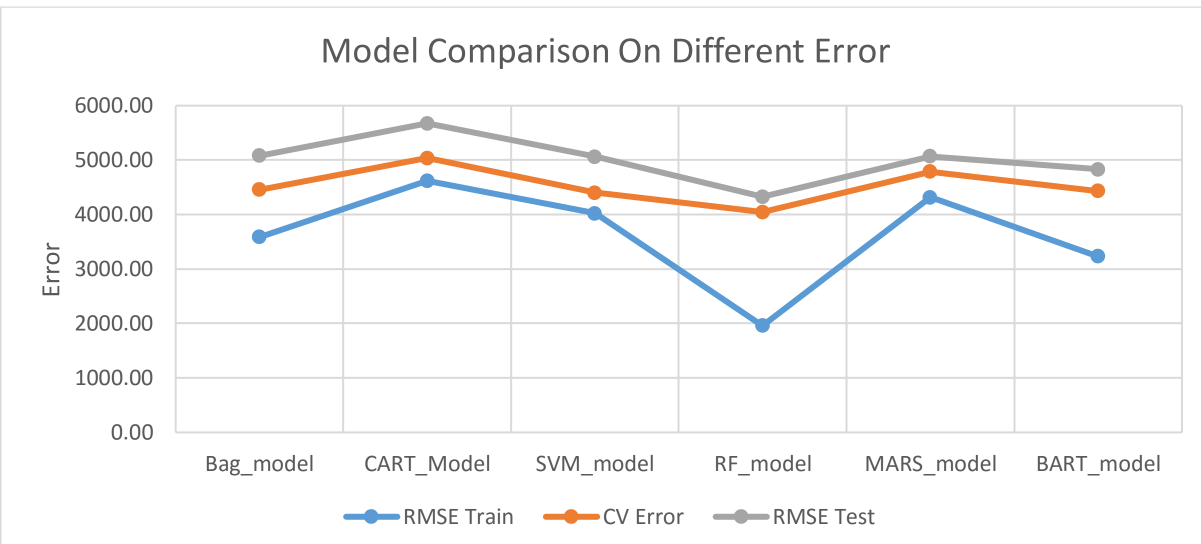


Figure 8: Model Summary of different model

To know which model is significantly different and best among this I performed the ANOVA and Fisher test to compare the mean obtained from 10-fold cross validation error. In this case our hypothesis is

$$Ho: \quad Both\ model\ has\ same\ cross\ validation\ error$$
$$H_1: \quad Model\ one\ is\ better\ than\ model\ two$$

The results were that Random forest model was significantly different we can see the P-value on next page. But when we see other models such as BART and SVM they have comparable results.

Factor	N	Mean	Grouping		
error_cart	10	5339	A		
error_mars	10	5016	A	B	
error_svm	10	4563	A	B	
error_bart	10	4433	A	B	C
error_bag	10	4288		B	C
error_rf	10	3546			C

Table 2: Fisher Test Summary

Above table represent the groups which are almost same, all those who have same Groups have almost same cv-error. Below table show us the p-vale of pairwise t-test

Model comparison	p-value
error_cart - error_mars	0.480
error_bag - error_mars	0.115
error_rf - error_mars	**0.002**
error_bart - error_mars	0.205
error_svm - error_mars	0.323
error_bag - error_cart	0.025
error_rf - error_cart	**0.000**
error_bart - error_cart	0.051
error_svm - error_cart	0.093
error_rf - error_bag	**0.108**
error_bart - error_bag	0.751
error_svm - error_bag	0.549
error_bart - error_rf	**0.056**
error_svm - error_rf	**0.030**
error_svm- error_bart	0.777

Table 3: p-value table from paired t-test in ANOVA

From Fisher's Test, we can see that random forest, bart and bagging model are significantly different from others. We can go with either of the model, and I choose to go with random forest as if we compare other parameter such as R-sq and mean cv error it is a better choice than other too. Also, it reduces variance as well as bias and provide a good trade-off.

Parameter of best model i.e for random forest is number of trees are 123 and it is selected by using cross validation and from the graph between number of tress and error.

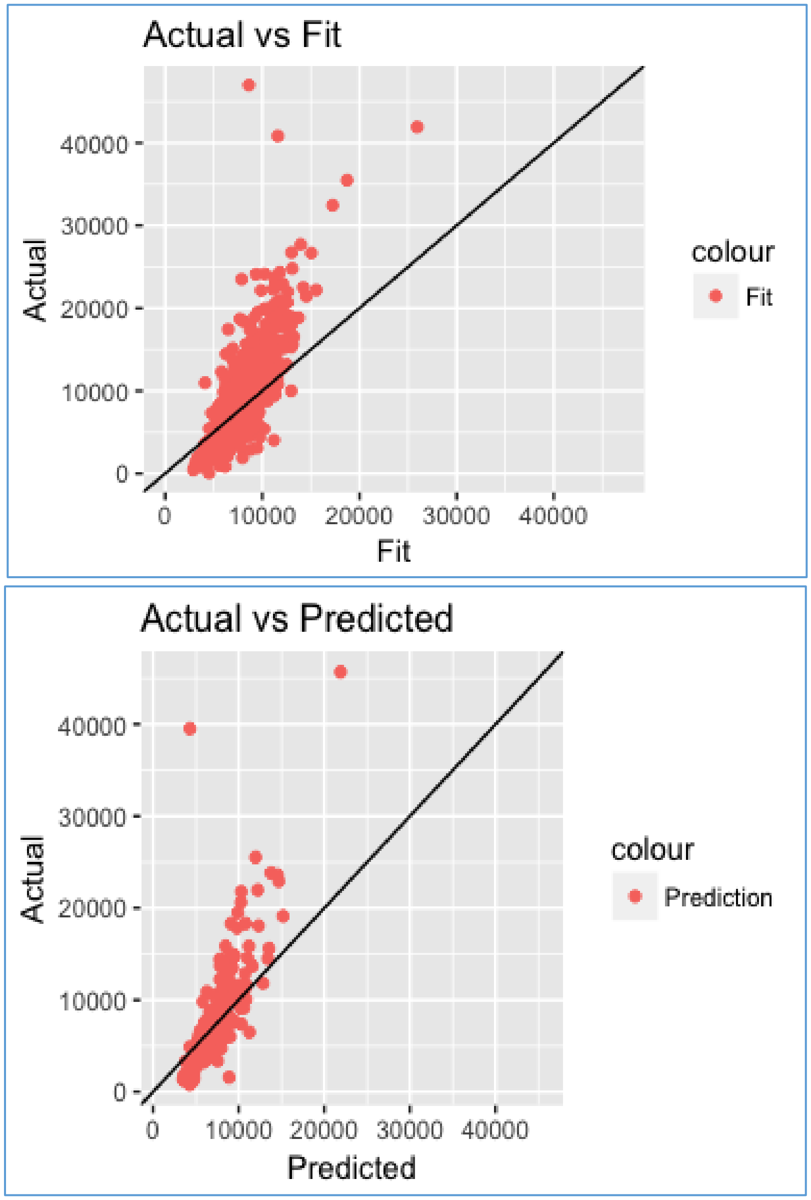


Figure 10: Actual vs Fit and Actual Vs Predicted for Random Forest Model

In above graph, Actual vs Fit we can see that points are lying on the straight line approximately. There are some outliers in the graph too. Even in Actual vs Predicted graph, we can see that they are almost on a line so our model is good.

9. Inferences

Variable Importance plot show us the which all predictors are explaining our model variability.

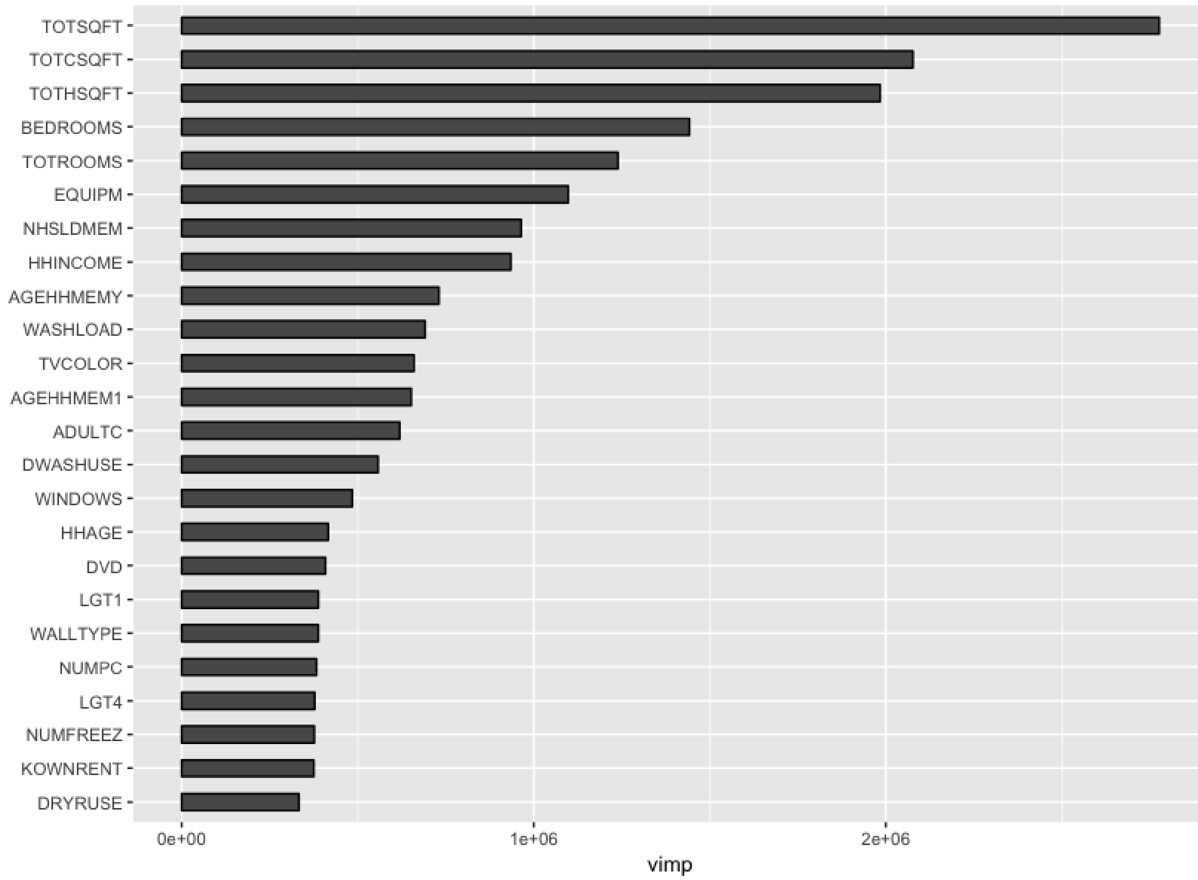
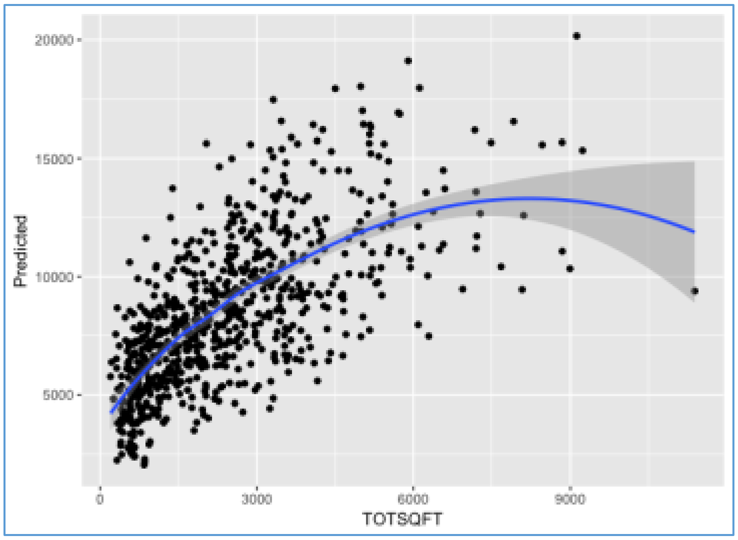


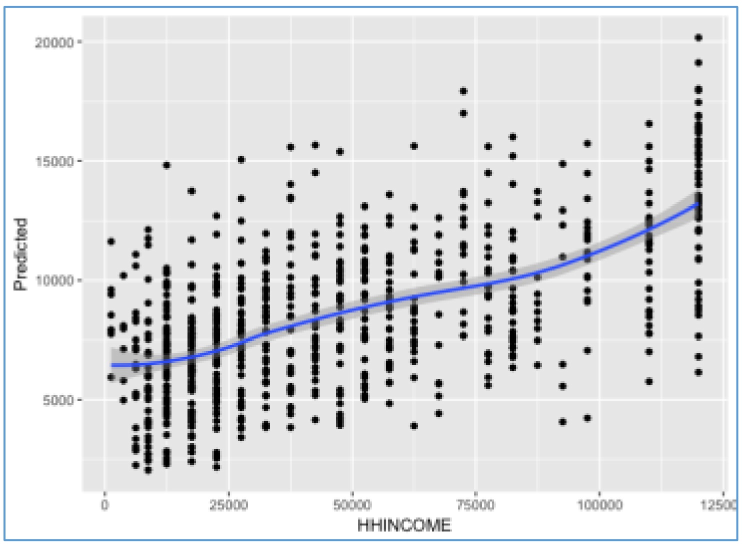
Figure 11: Variable Importance plot for top 24 predictors of Random Forest Model

From above figure we can see that, most important factor which affects electricity consumption is total area of the house. After that area in which we need cooling and heating. Followed by number of bedrooms, median income of house, age of youngest member, number of times you wash clothes, total number of adults living in house, number of windows, wall type. And all these factors seem to be common, but from partial plots we can go in depth whether they have positive or negative relationship with total site electricity usage. (I have also attached one more variable importance plot which have all the predictors in the appendix)

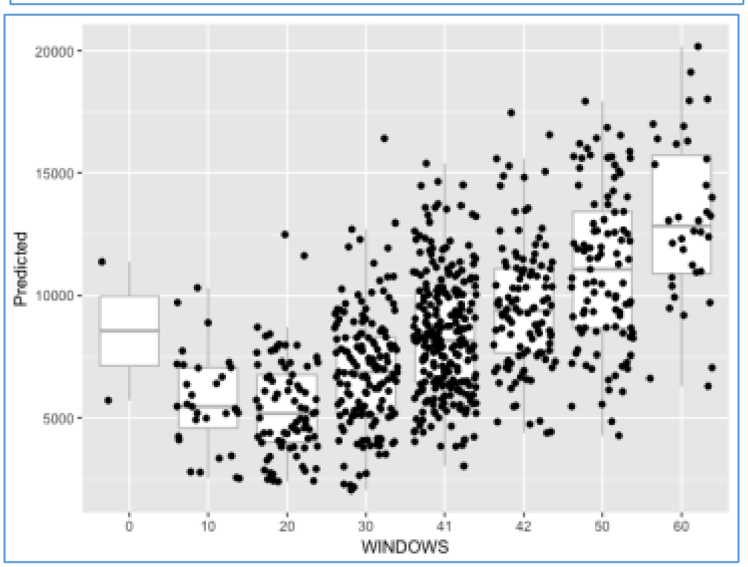
I have explained some of the partial plots here and rest are attached in the appendix.



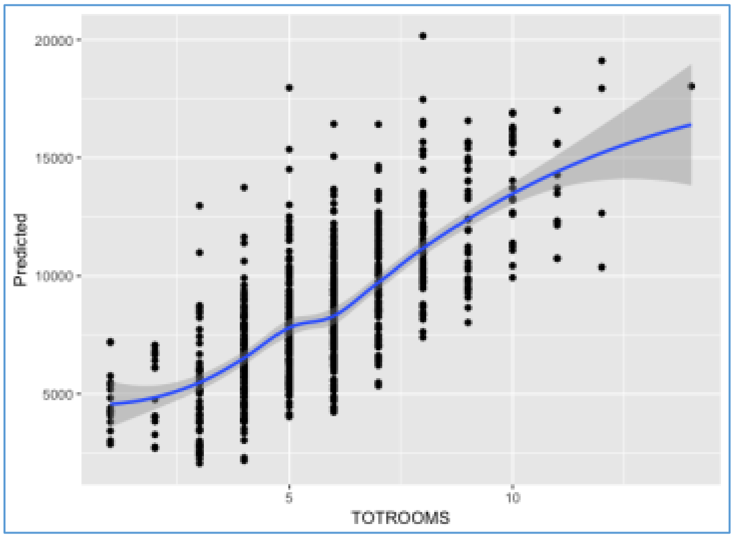
In this partial plot we can see that as we increase the total area of house electricity consumption is also increasing. Which is true in most of the cases but we can see a threshold point towards the end where consumption is going down even though it has a maximum area. This can be possible as that house maybe a vacation house.



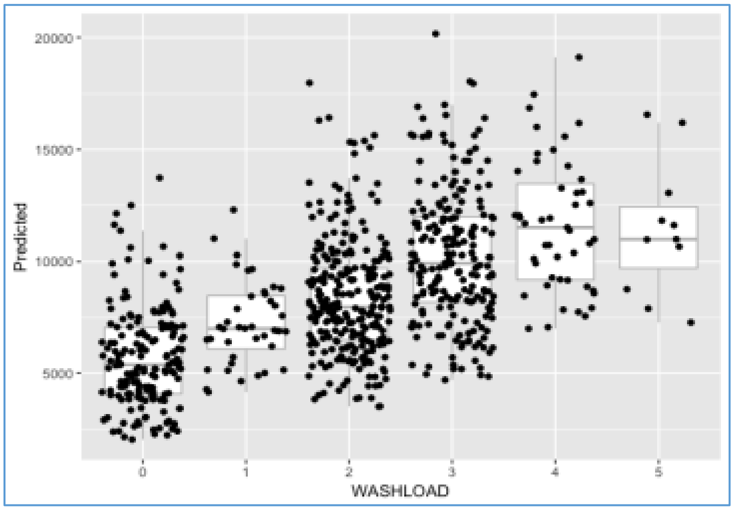
Plot show us that as the household income increases, house electricity consumption increases. This is true as we earn more, we become careless about our spending on small things, we tend to buy more appliances for our comfort. Most of the appliance these days use electricity so it is natural as we earn more we use more electrical energy.



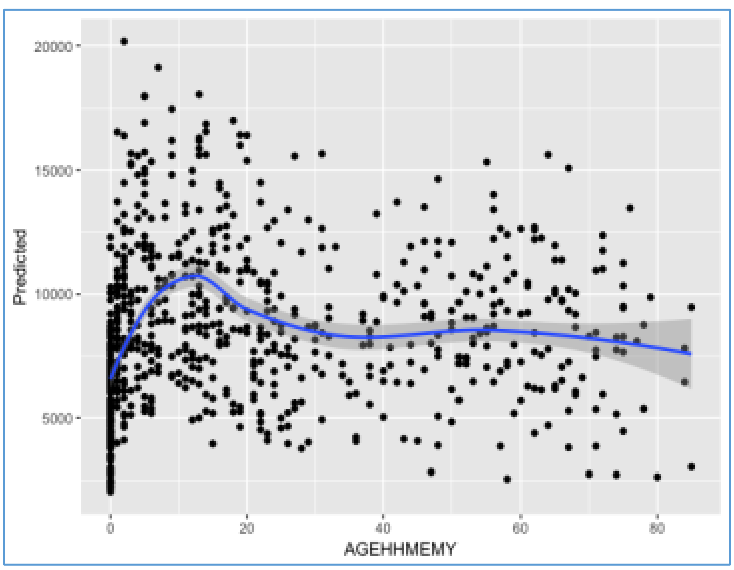
This one seem to be interesting, here 10 mean 1 or 2 window, 20 is 3-5, 30 is 6-9,41 is 10-15,42 is 16-19 etc. We see that when number of window in house is below 6 usage is same as window provides fresh air but as we increase windows, the consumption shoots very much as there is loss of energy (heating or cooling) from those windows. While building houses we can consider this as an important factor.



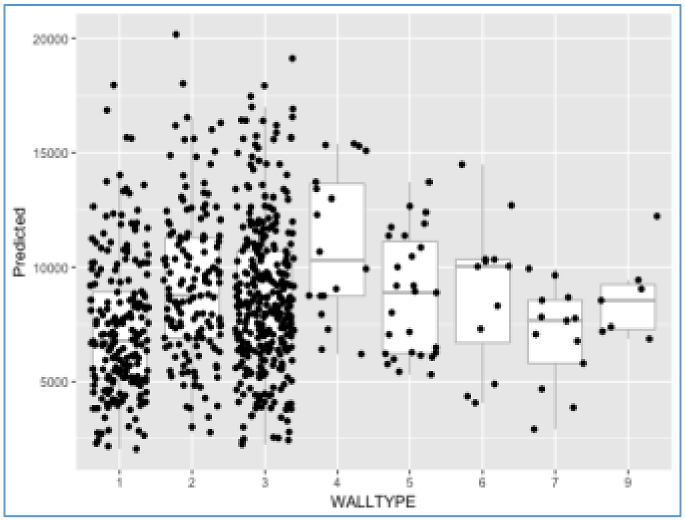
Graphs state that more the number of rooms in your house more usage you have. It is because as we have more room we will have more lighting, cooling, heating and other appliances in rooms and we will use more energy.



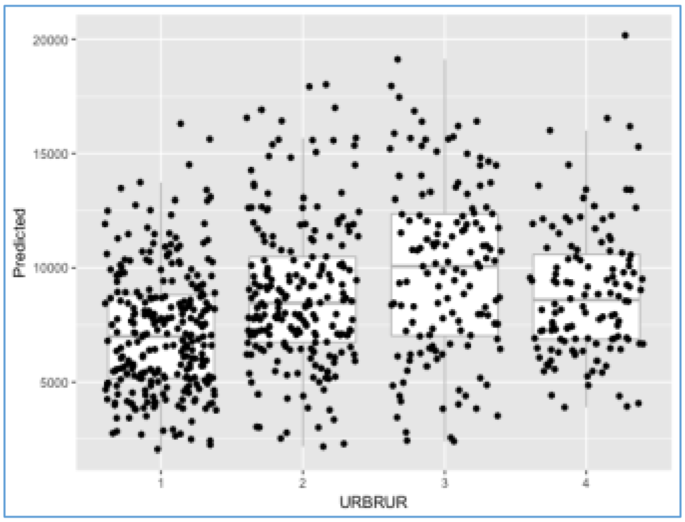
If we use washing machines more, we tend to use more energy. As washing machines are very heavy machine compared to other home appliances. More we use, more is the usage of electricity. From this plot, we can see that most of the family have chosen 2 (2-4 loads/week) -3 (5-9 loads/week) as density of points is very high in this area.



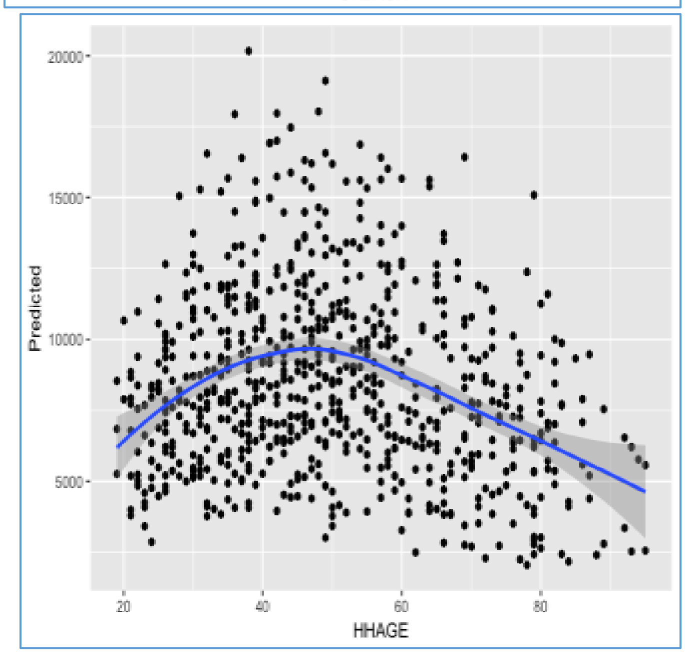
This graph is between age of youngest house member and usage of electricity. This graphs shows that when age of youngest person is child (below 12 yr) the consumption is still very less but as a person grows into a teenager (13-19 yr) there is a rise in usage, which is evident from the peak. As person grows into an adult and starts earning money, usage goes down as they have to manage all expenses on own. So a teenager tends to use energy lavishly in appliances like PlayStation, personal computer.



The construction of house also affects the total consumption and that can be seen in here as wall type was also among top predictors. Wall type 1 (bricks), wall type 2(wood), 3 (siding steel or alum) for rest you can see code book. But we can see that bricks wall are better than wood, as they provide better insulation I guess. We cannot comment on wall type 4-9 as we have very less data point for that.



This graph shows us which locality family is in. Region 1 is city, 2 is town, 3 is suburbs and 4 is rural. We can see that families living in suburbs tends to use more electricity as they have many heavy farm equipment or other reason could be families living in city and town use more energy efficient product. Also, in rural area we have high usage as compared to city, it would be mainly due to farm equipment.



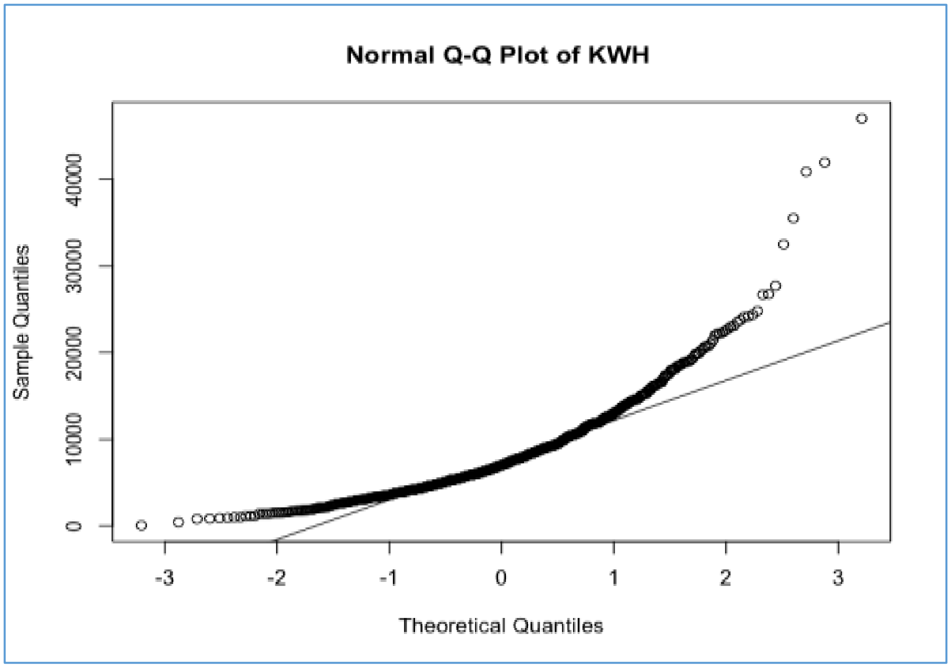
This graph is between HHAGE which mean an age of person who is responsible for household bills and total electricity usage. We can see that when person is young in early twenties, he/she earn less so they have less usage but as age grow they earn more and more till the age of 40-50 same goes their electricity usage, but after that when a person ages, he starts saving for his retirement and cuts down his expenses and becomes more conscious of electricity usage, hence saves it. When they are retired (around the age of 65) they consume very less energy as they have very less income to spend so they tend to use less.

10. References

1. Baker KJ,Rylatt RM.Improving the prediction of UK domestic energy-demand using annual consumption-data.ApplEnergy2008;85(6):475–82.
2. Carlson DR.Scott Matthews H,Bergés M. One size does not fit all: averaged data on house hold electricity is inadequate for residential energy policy and decisions. Energy Build 2013;64:132–44.
3. Parti M,Parti C. The total and appliance-specific conditional demand for electricity in the house hold sector. BellJ Econ1980;11(1):309–21.
4. Tso GKF, Yau KKW. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy2007;32 (9):1761–8. [9].
5. Jones RV, Fuertes A, The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings.
6. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth, Belmont
7. Hastie et al., 2009
8. Breiman 1996 Machine Learning, 24(2), 123-140.
9. Hugh Chipman - The Annals of Applied Statistics 2010, Vol. 4, No. 1, 266–29

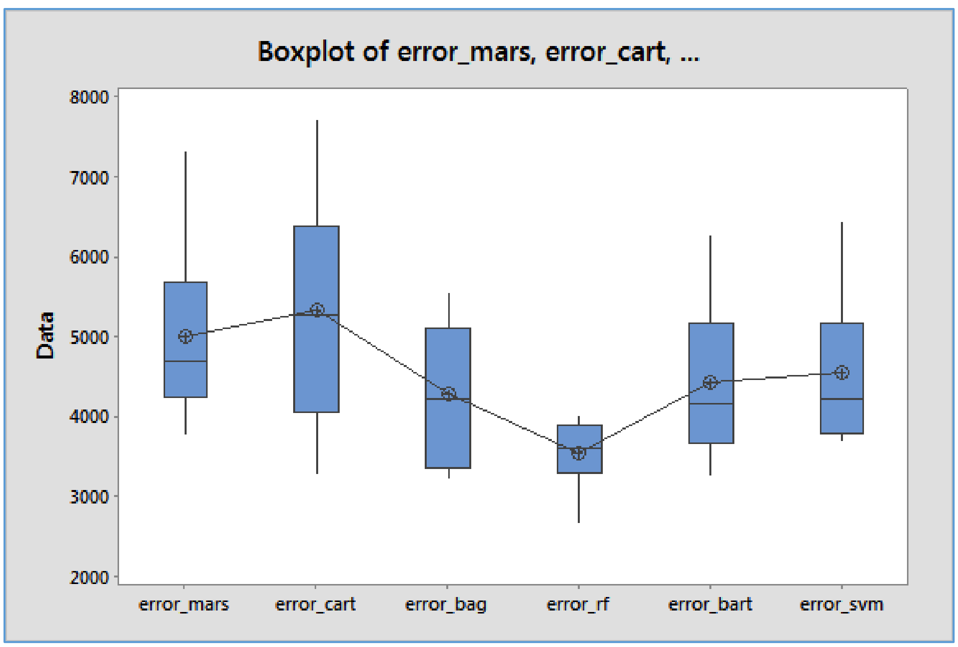
11. Appendix

11.1 QQ-Plot

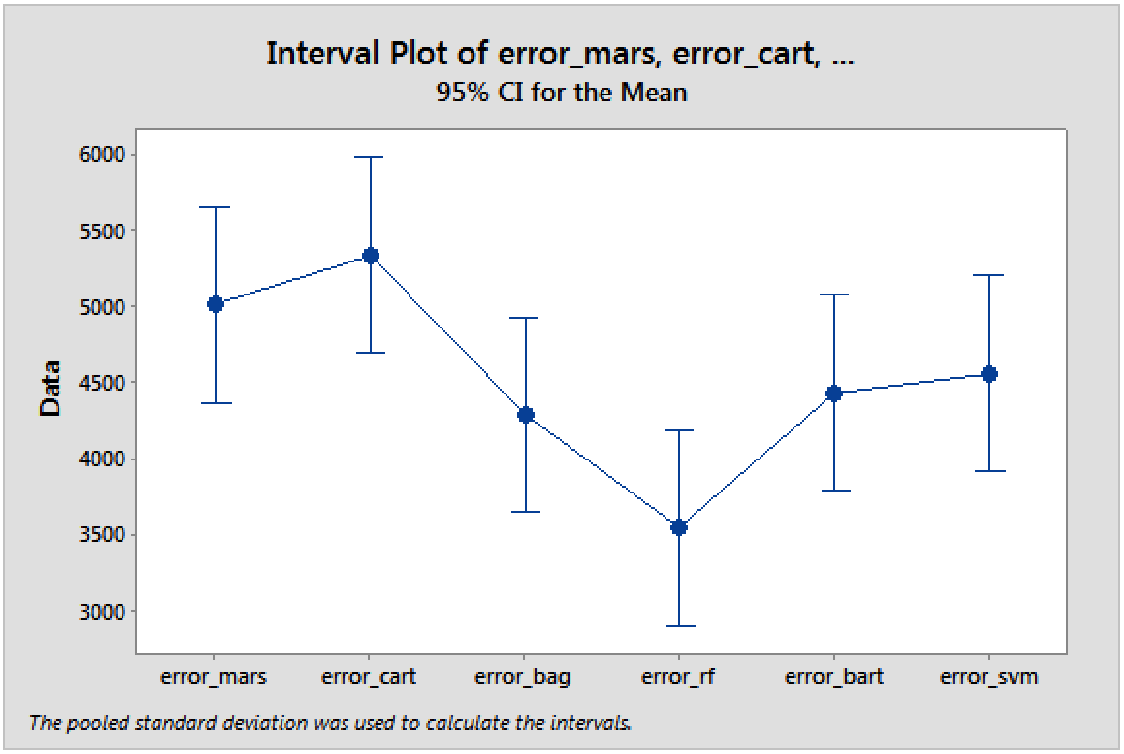


QQ-plot of response variable

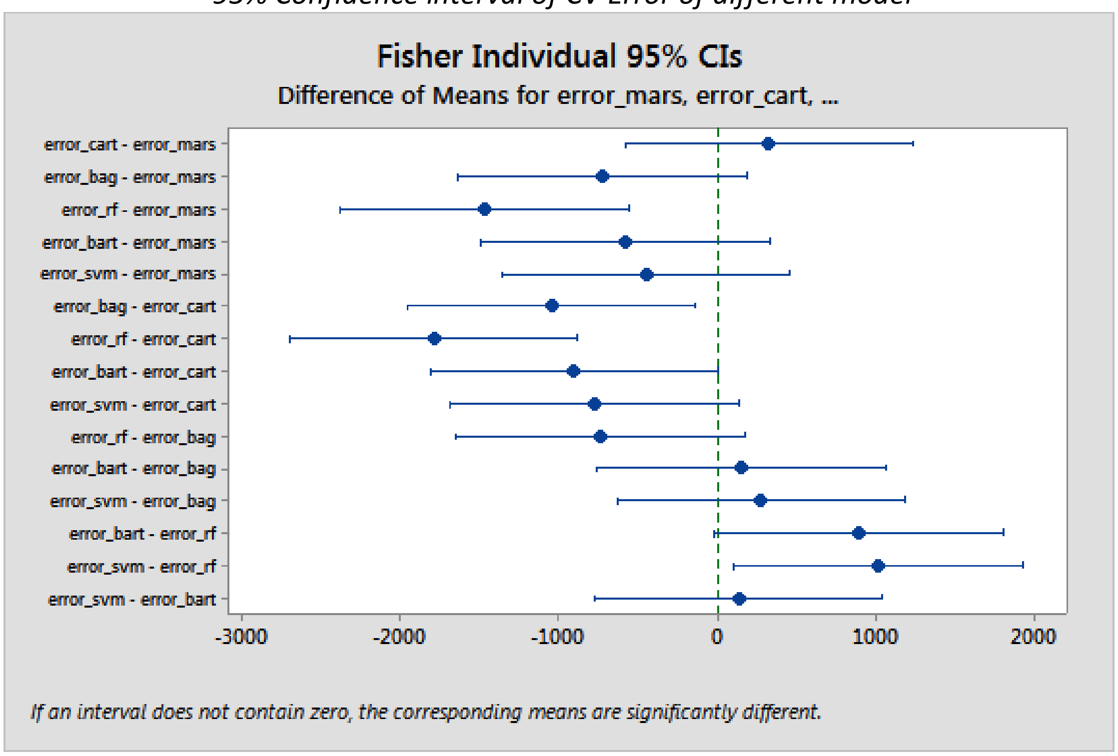
11.2 Statistical test results and plots



Box plot of cv error of different models



95% Confidence interval of CV Error of different model



Fisher test for comparing different model

One-way ANOVA: error_mars, error_cart, error_bag, error_rf, error_bart, error_svm

```
Method
Null hypothesis         All means are equal
Alternative hypothesis  At least one mean is different
Significance level      α = 0.05

Equal variances were assumed for the analysis.

Factor Information
Factor  Levels  Values
Factor       6  error_mars, error_cart, error_bag, error_rf, error_bart, error_svm

Analysis of Variance
Source  DF    Adj SS   Adj MS  F-Value  P-Value
Factor   5  19290834  3858167     3.73    0.006
Error   54  55856561  1034381
Total   59  75147395

Means
Factor       N  Mean  StDev     95% CI
error_mars  10  5016   1085  (4371, 5661)
error_cart  10  5339   1499  (4694, 5984)
error_bag   10  4288    868  (3643, 4933)
error_rf    10  3546    420  (2901, 4191)
error_bart  10  4433    971  (3788, 5078)
error_svm   10  4563    954  (3918, 5207)

Pooled StDev = 1017.05

Fisher Pairwise Comparisons
Grouping Information Using the Fisher LSD Method and 95% Confidence
Factor       N  Mean  Grouping
error_cart  10  5339  A
error_mars  10  5016  A B
error_svm   10  4563  A B
error_bart  10  4433  A B C
error_bag   10  4288    B C
error_rf    10  3546      C

Means that do not share a letter are significantly different.

Fisher Individual Tests for Differences of Means
                         Difference       SE of                                Adjusted
Difference of Levels       of Means  Difference        95% CI      T-Value     P-Value
error_cart - error_mars         323         455   ( -589, 1235)       0.71       0.481
error_bag - error_mars         -728         455   (-1640,  184)      -1.60       0.115
error_rf - error_mars         -1470         455   (-2382, -559)      -3.23       0.002
error_bart - error_mars        -583         455   (-1495,  329)      -1.28       0.205
error_svm - error_mars         -454         455   (-1365,  458)      -1.00       0.323
error_bag - error_cart        -1051         455   (-1963, -139)      -2.31       0.025
error_rf - error_cart         -1794         455   (-2705, -882)      -3.94       0.000
error_bart - error_cart        -906         455   (-1818,    6)      -1.99       0.051
error_svm - error_cart         -777         455   (-1688,  135)      -1.71       0.093
error_rf - error_bag           -742         455   (-1654,  170)      -1.63       0.108
error_bart - error_bag          145         455   ( -767, 1057)       0.32       0.751
error_svm - error_bag           275         455   ( -637, 1186)       0.60       0.549
error_bart - error_rf           887         455   (  -25, 1799)       1.95       0.056
error_svm - error_rf           1017         455   (  105, 1929)       2.24       0.030
error_svm - error_bart          130         455   ( -782, 1042)       0.29       0.777
Simultaneous confidence level = 64.68%
```

11.3 Final model Selection and other graphs

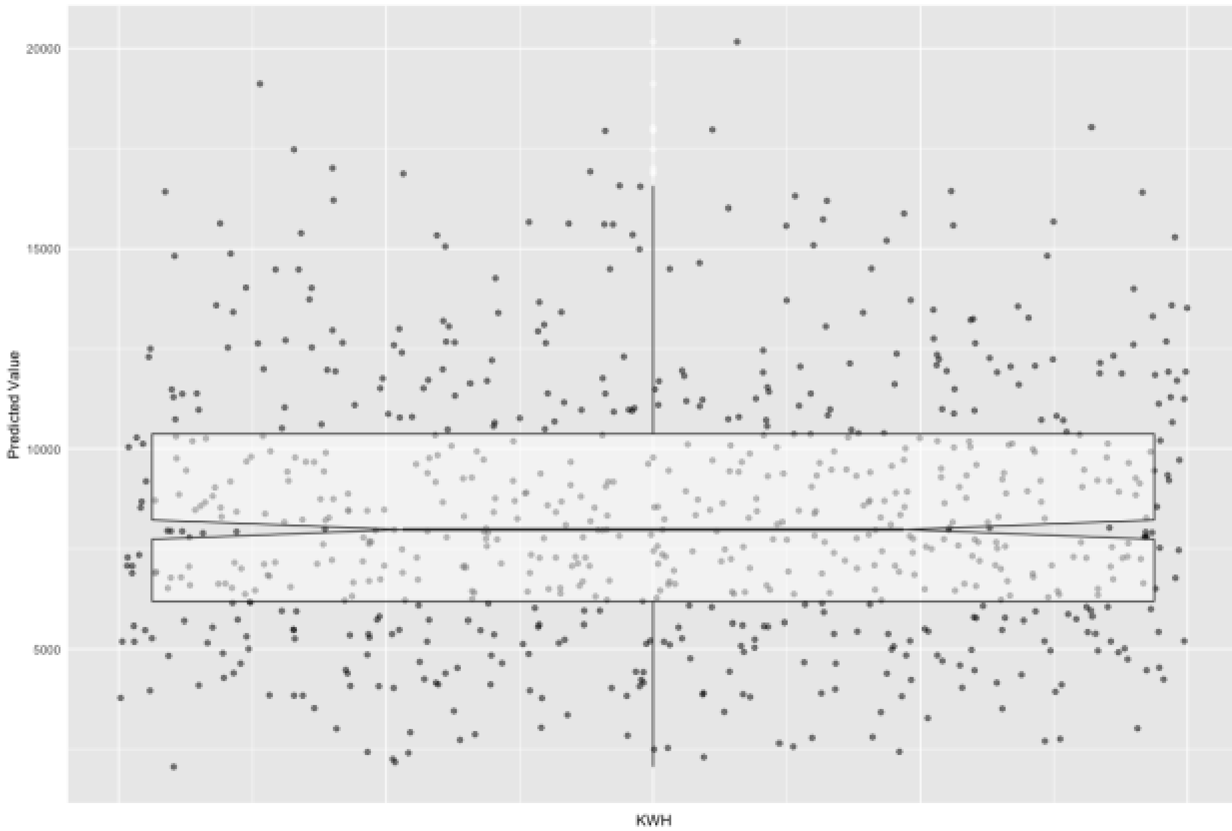
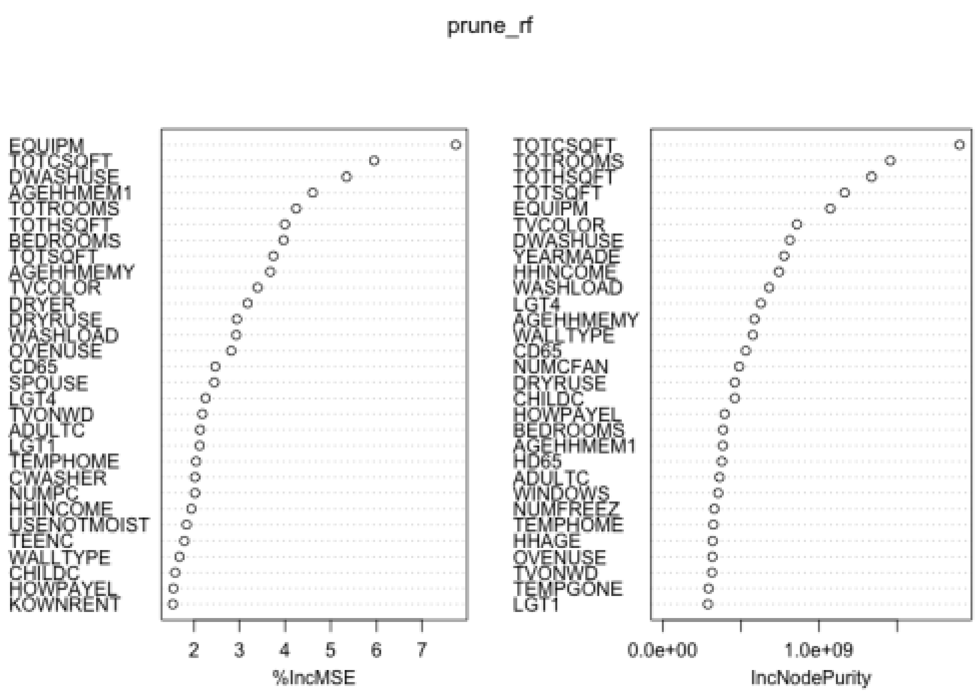


Figure: predicted median KWH usage. Points are jittered to help visualize predictions for each observation. Boxplot indicates the distribution of the predicted values



Variable importance plot

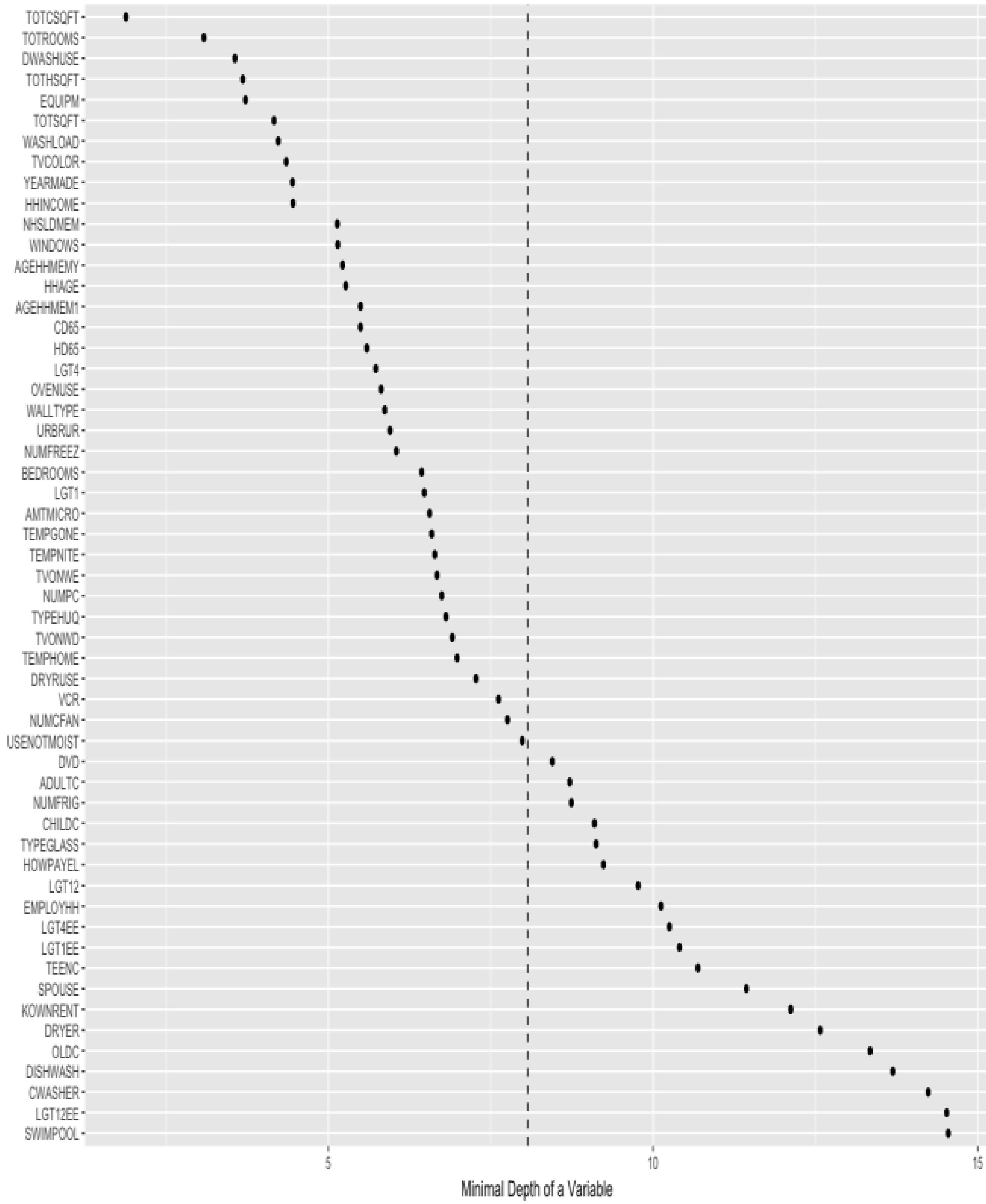
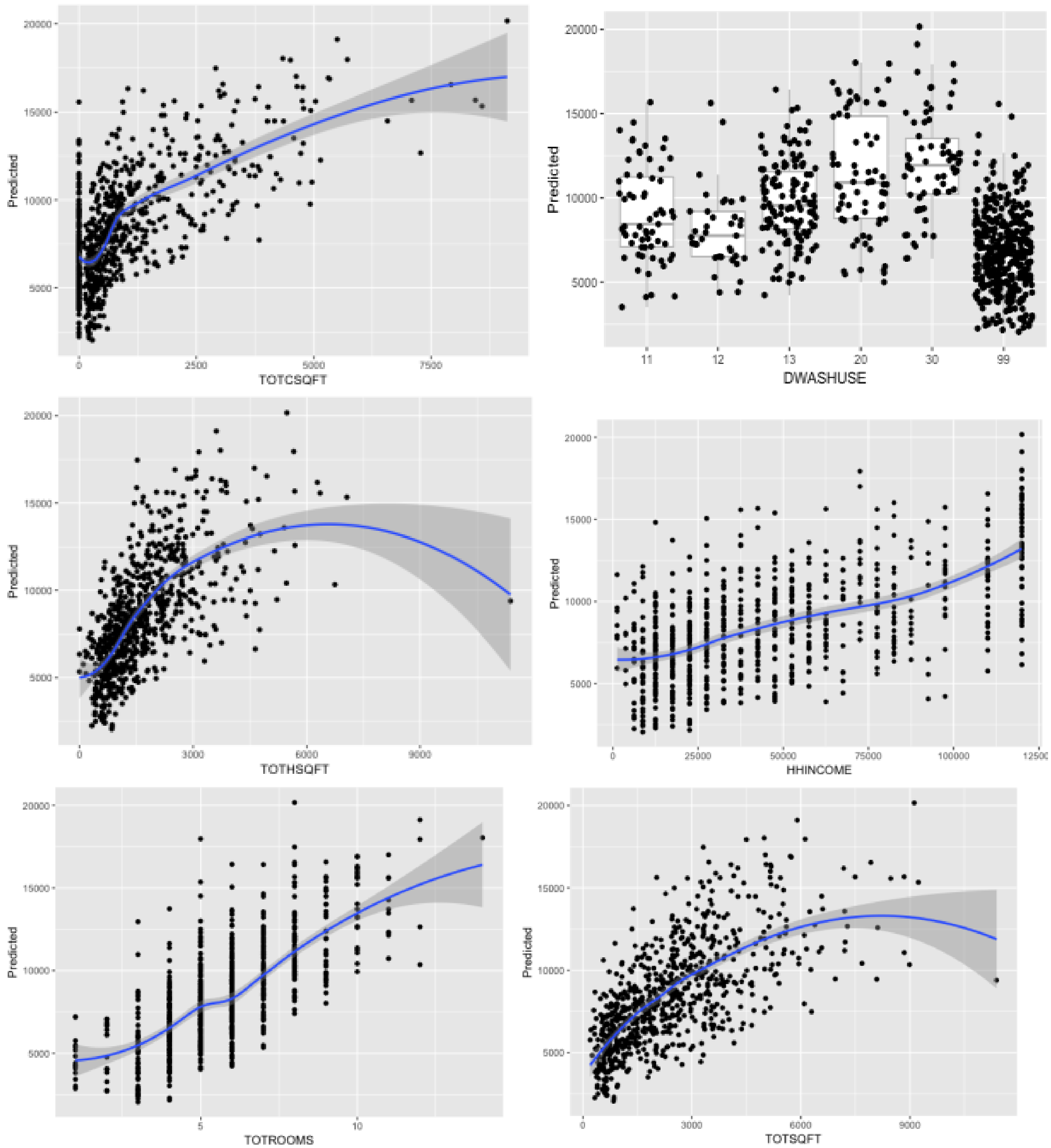
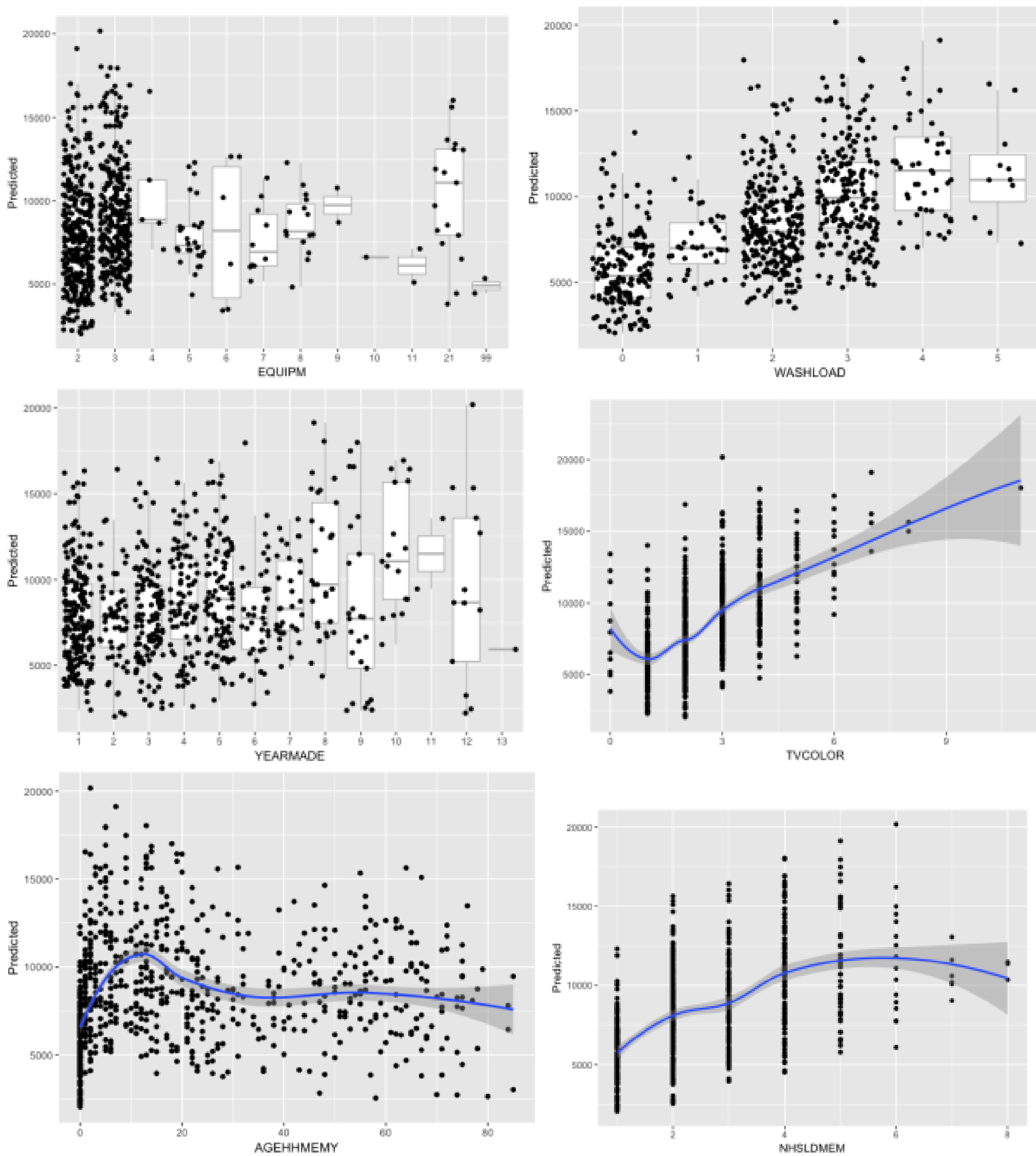
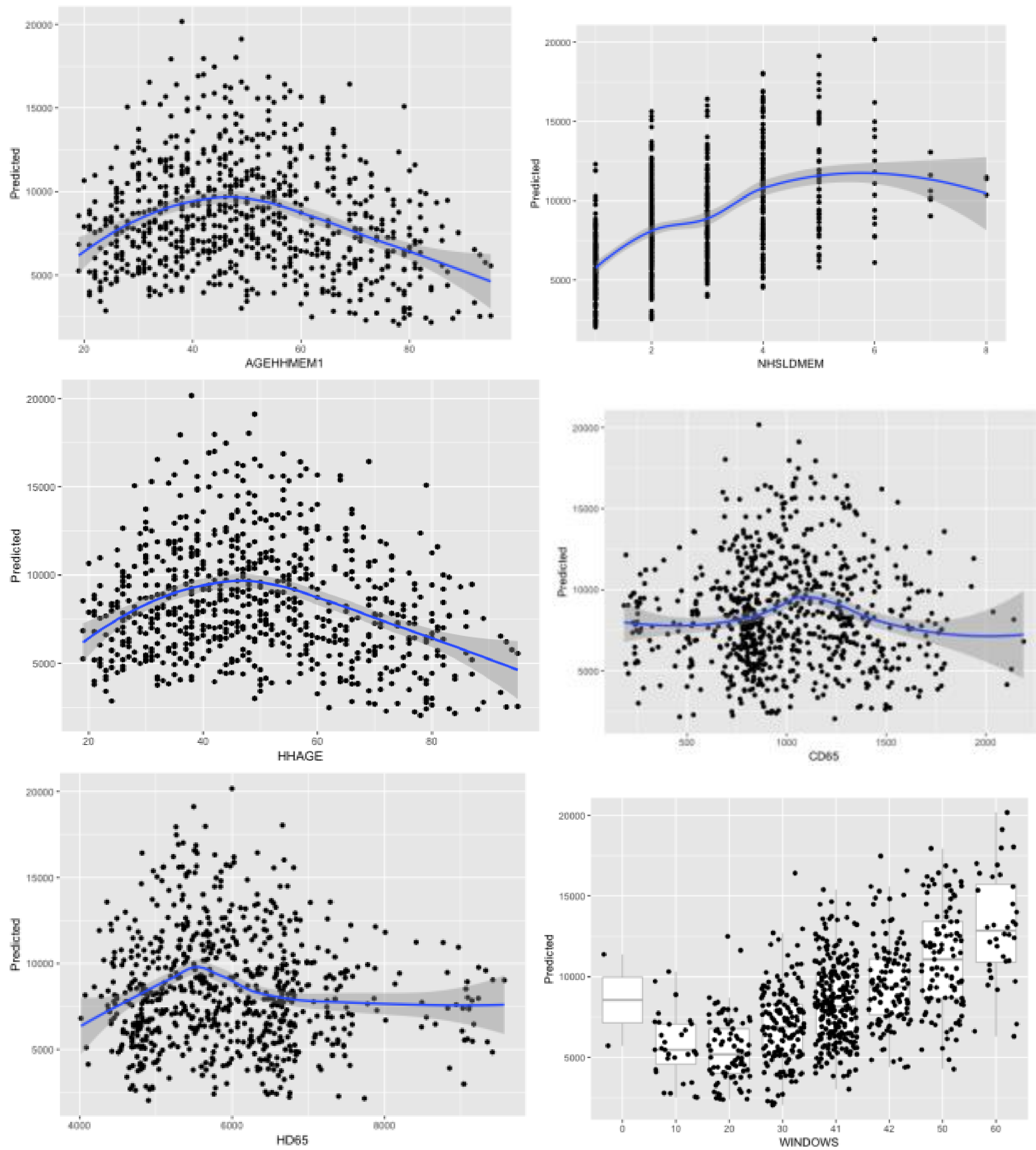


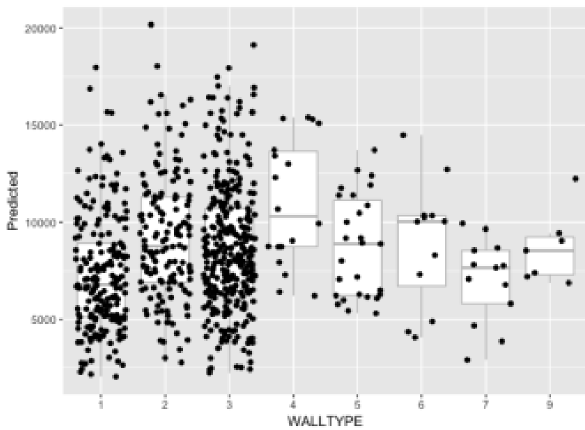
Figure: Minimal Depth variables in rank order, most important at the top. Vertical dashed line indicates the maximal minimal depth for important variables.

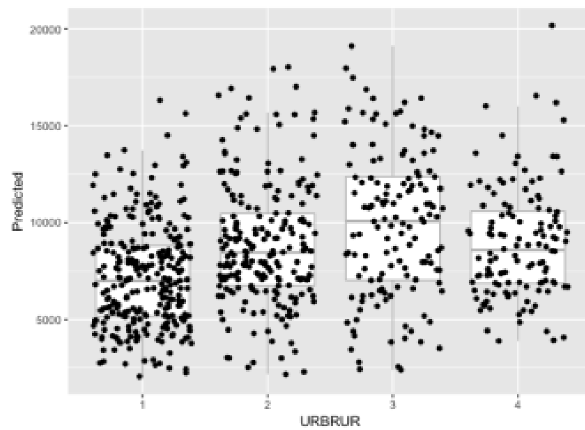
11.4 Partial Plots

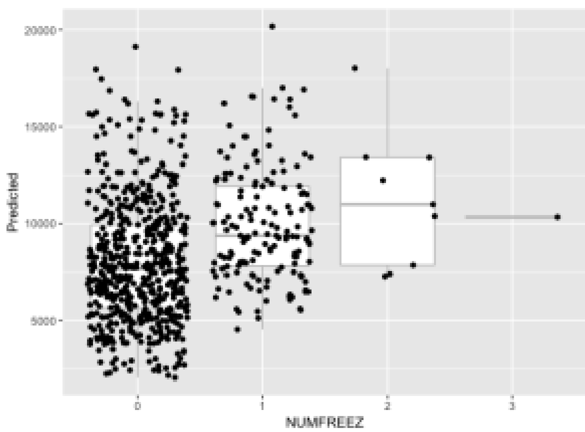


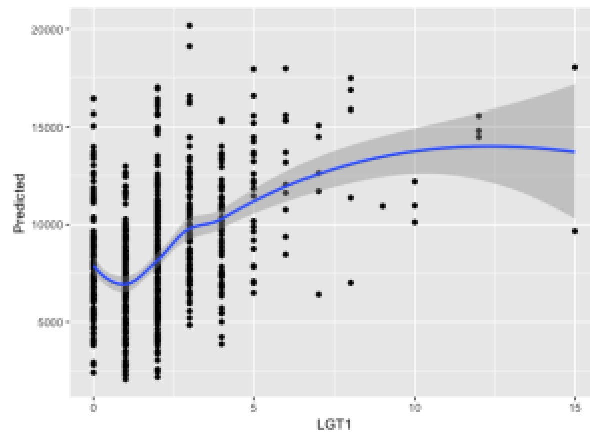

Predicted
EQUIPM
WASHLOAD
YEARMADE
TVCOLOR
AGEHHMEMY
NHSLDMEM

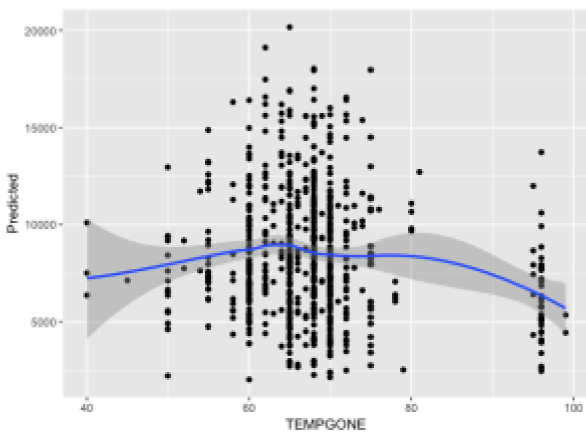

Predicted
AGEHHMEM1
NHSLDMEM
HHAGE
CD65
HD65
WINDOWS
20000
15000
10000
5000
20
40
60
80
2
4
6
8
500
1000
1500
2000
4000
6000
8000
0
10
20
30
41
42
50
60

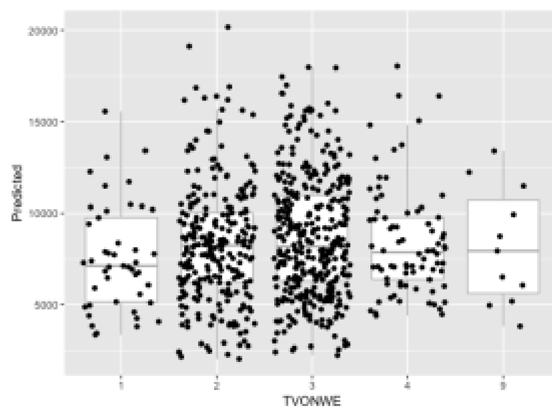

Predicted
20000
15000
10000
5000
1
2
3
4
5
6
7
9
WALLTYPE

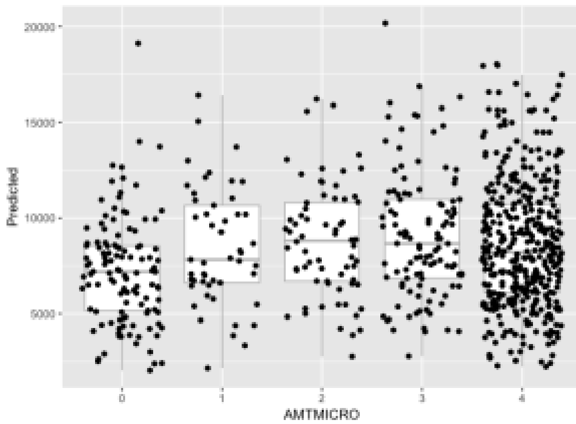

Predicted
20000
15000
10000
5000
1
2
3
4
URBRUR

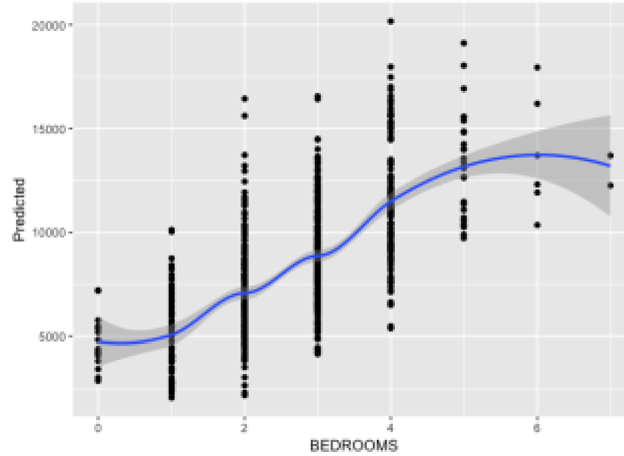

Predicted
20000
15000
10000
5000
0
1
2
3
NUMFREEZ

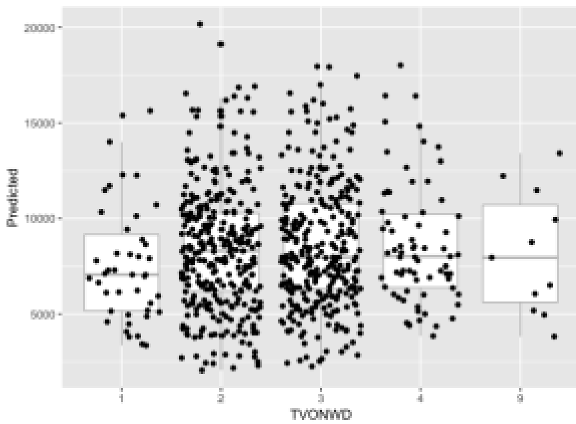

Predicted
20000
15000
10000
5000
0
5
10
15
LGT1

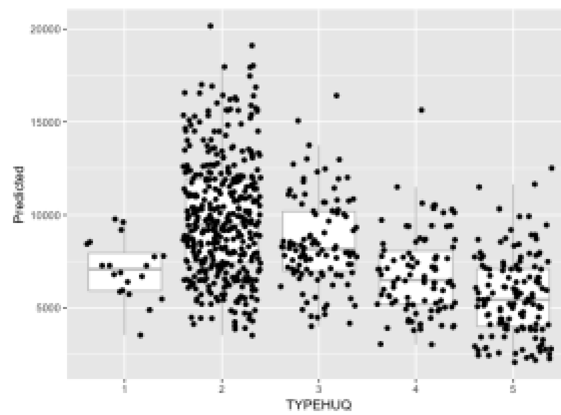

Predicted
20000
15000
10000
5000
40
60
80
100
TEMPGONE

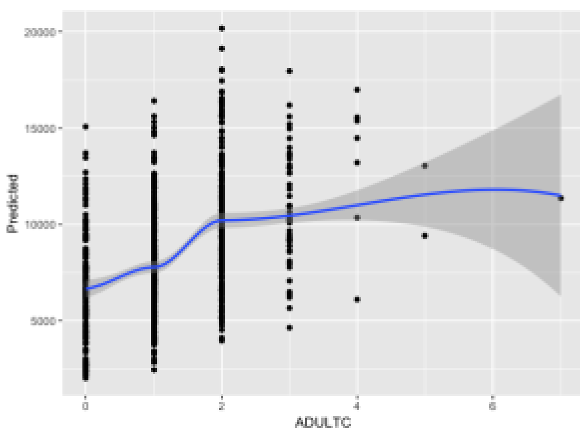

Predicted
20000
15000
10000
5000
1
2
3
4
9
TVONWE


Predicted
20000
15000
10000
5000
0
1
2
3
4
AMTMICRO


Predicted
20000
15000
10000
5000
0
2
4
6
BEDROOMS


Predicted
20000
15000
10000
5000
1
2
3
4
9
TVONWD


Predicted
20000
15000
10000
5000
1
2
3
4
5
TYPEHUQ


Predicted
20000
15000
10000
5000
0
2
4
6
ADULTC

“I have obeyed all rules for this exam and have not received any unauthorized aid or advice.”

Ankur Wadhwa
16th April,2017