Washington D.C House Price Analysis and Prediction

May 2, 2019

Through our analysis of the DC Properties data set, we would like to understand how residential property variables affect the property price, which is a quantitative outcome. Our analytic goals are Interpretation and Prediction. Since our outcome variable is the price of properties in DC, we want to know which factors have an effect on the magnitude of the price. For instance, holding all variables constant, does having more bedrooms have a positive effect on the price. We would like to use our model to predict property prices in DC.

In an article written by the Washington Post in 2018 (Park, 2018), inventory of homes in the real estate market has been on the rise. With more homes in the market, buyers are taking more time evaluating available homes causing homes to be on the market 19 days longer when compared to the previous year. Home prices is one of the biggest concerns for both the home buyer and the home seller. We want to use our analysis results to help real estate agents in the Washington D.C. area better understand how these residential properties variables affect the price of the home. They can help home sellers maximize their profits and home buyers to identify if the selling price is a good deal. The real estate agents can provide reasonable prices for the house that is acceptable to both buyers and sellers, thus, speeding up the time needed to sell a home.

To explore our topic, we will be using the Washington D.C. Residential Properties data set from Kaggle for the years 2015-2018. There are approximately 65,000 observations all of which are properties listed or sold in the city. Our data set includes numerical variables with number of bathrooms, bedrooms, kitchens, fireplaces and gross building area. For categorical variable, the data set includes heater type, air-conditioning (Yes/ No), qualification (qualified/ under-qualified), structure type, condition, extwall type, roof type, intwall type, usecode (residential or commercial use), zip code and quadrant. We also have two continuous variables, AYB and YLRMDL which represents the earliest time the main portion of the building was built, and year structure was remodeled in respective order. (See Appendix A for Variable Description)

The outcome variable is “PRICE” which is the price of properties in DC between 2015 to 2018. The main predictors are the ZIPCODE, YLRMDL, and the number of rooms in the properties. For example, for ZIPCODE the price of properties should vary throughout the area in DC, the year of the last remodeling is also important to the price of the properties. Also, the number of rooms in properties such as bathrooms and bedrooms should affect the price, for instance, the more rooms in the house means the house is supposed to be bigger in size and price. The predictors in the model are both categorical and numeric.

From the first visual plot (ggpair1), the AC variables boxplot shows a difference in means of Y and N of AC, with the mean of Y being higher. From the 2nd plot (ggpair2), the correlation between PRICE and YLRMDL (0.263) is positive. From the 3rd plot (ggpair3), the

correlation of PRICE and ZIPCODE (-0.17) is negative. Also, from the 4th plot (ggpair4), the mean of QUALFIED boxplot overlap so there is little difference in prices for properties that are qualified or unqualified. (See Appendix B for ggpair)

The first step of data pre-processing is to omit the NA from the data set. Our final data set has approximately 58,000 observations. Next, STORIES, USECODE, ZIPCODE are turned into factor. Then we releveled the categorical variables. For example, releveling "Multi" as the reference level for STRUCT, "Average" for CNDTN, "Carpet" for INTWALL and so on. (See Appendix A for Variable's Reference Level) For the variable selection, we use the stepwise method for our model. We found that variables "HEAT", "EXTWALL", "USECODE" and "QUADRANT" are non-significant.We reduced our model from 16 variables to 20. After running the OLS, and test for the its assumption, we found that (1) from the residual and Breusch-Pegan test, there are heteroscedasticity in our residuals, leading to WLS model, (2) from the Durbin-Watson test, it suggested that the data don't have the problem of serial correlation/ auto correlation. (3) from the multicollinearity test, it suggested that there is no multicollinearity among the independent variables.

We ran OLS model, WLS model, non-Linear model, and decision tree model for our data set. For each of the model, we ran at least two specifications. For the OLS and WLS models, we run the plain model and log outcome variable model. For non-linear model, we ran quadratic and cubic polynomial models and interaction model. For decision tree model, we run the regression tree model, bagging tree, and random forest tree.

OLS

We ran plain OLS as well as OLS with log specification on PRICE. In plain OLS, the variables which include BATHRM, HF_BATHRM, BEDRM, YLRMDL, QUALIFIED, GBA, STRUCT, CNDTN and FIREPLACES are highly significant with p-value < 0.001. AYB and kitchen are significant with p-value between 0.001 and 0.05. In categorical variables such as STORIES, ROOF, INTWALL, USECODE and ZIPCODE, some categories of these categorical variables are significant while others are not.

The plain OLS model's coefficients have small dependent variables with both positive and negative impact on PRICE. For the significant variables that carry positive effect on PRICE are BATHRM, HF_BATHRM, YLRMDL, GBA, STRUCT, CNDTN and FIREPLACES while AYB, BEDRM, AYB, STORIES, QUALIFIED, and KITCHENS bring negative effect on PRICE. The plain OLS model has R-squared: 0.5745 which means around 57% of the sample variation in PRICE is explained by all the variables in our data set except PRICE.

Our OLS model with log on PRICE shows similar result as plain OLS with some minor differ. All the coefficients that bring either positive or negative effect on PRICE have no change; however, log on PRICE OLS does make variable STRUCT not significant and some categories of the categorical variables much less significant or not significant. The R-squared of log on PRICE OLS model is 0.5972 which is slightly better than the plain OLS.

A Cross Validation of the MSE of the plain OLS and OLS Log model resulted in a MSE of 139,860,214,097 for the plain OLS model and a MSE of 664,162,889,919 for the OLS Log model. We have noticed that log on PRICE OLS has significantly higher MSE compare to the

plain OLS and this could mean data values within log on PRICE model are spread widely around its mean more than the plain OLS model. As the smaller MSE helps us find a better model, and log on PRICE OLS does not present much difference on the results of coefficients and significance compare to the plain OLS, OLS as a better model. (See Appendix C and D for OLS and Appendix F for OLS Log)

WLS

Through the Breusch-Pagan test, we found that our model has Heteroscedasticity. This can be corrected with the Weighted Least Squares (WLS) regression. We weighted the model by 1 divided by the residuals of the OLS model squared. Because the WLS model has a lower for the larger residual errors, the model R-squared is inflated. We calculated a Real R-squared of 0.57417. The MSE of the WLS model is 139,678,604,484, which is lower than the MSE of the OLS Model 139,860,214,097. The WLS model performs better than the OLS model.

The reference level of our ZIPCODE variable is 20001 which is the Shaw neighborhood. The Shaw neighborhood in NW DC includes the U street corridor and Howard University. The largest coefficient (of the ZIPCODE categorical variable is 20007, which belongs to the Georgetown neighborhood. Located along the Potomac River, Georgetown Neighborhood is a large commercial and entertainment district and Georgetown University is located in this neighborhood. The coefficient of ZIPCODE20007 is 3.716e+05. This means that property prices in Georgetown are $371,600 high than property prices in Shaw.

Because the distribution of PRICE is not normal, we ran a WLS Log model by logging the PRICE. The Real R-Squared of the WLS Log model is 0.59719. The MSE of the WLS Log model is 664,162,600,009. When comparing the MSE of the plain WLS model and the WLS log model, the MSE of the WLS Log model is a lot larger. The WLS Log model did not perform as well as the WLS model. (See Appendix G for WLS and Appendix H for WLS Log)

Non-linear

The nonlinear model uses the specifications of polynomials and interaction. After trying different variables in polynomial specifications, we choose the YLRMDL (Recent year structure was remodeled or improved). We found the Quadratic model has a R-squared of 0.577 which means that the model can explain 57.7% of the dependent variable, PRICE. The coefficients for quadratics terms are highly significant. This model suggests that as the positive coefficients of the polynomial term of YLRMDL are higher, the higher the price of DC properties will rise. The cubic model is also significant with the R squared at 57.71%. Also, the coefficient has positive signs for all its cubic terms, so the higher YLRMDL the higher properties price in DC will rise. The adjusted R squared of the cubic model is 0.0001 higher than the quadratic model. The MSE for each model is 138,182,148,211, and 138,193,897,111 for quadratic and cubic, respectively. Thus, the Cubic model is the best non-linear model. (See Appendix I for Polynomial)

The non-linear model with an interaction term of Binary of QUALIFIED (when Q = 1 and U = 0) and Continuous of YLRMDL. The R-squared is at 57.63% and the interaction term has a

positive effect and is highly significant at p-value < 2e-16. The main effect of YLRMDL is negative. The negative main effect is offset when the property is Qualified because the interaction effect is positive. As the year in YLRMDL increases by 1, the price will change by 1410 (2.510e+3-1.100e+3=1410) which is stronger when the properties are qualified. The MSE for the interaction model is 138,371,192,960. (See Appendix J for Interaction)

Decision Tree Model

We decided to run the regression tree first. We put predictors that are significant to the model as independent variables and the PRICE in the data set as dependent variables. We set a loop to find the best minder which controls the size of the tree. After we ran the code, we find out that when the mindev is 0.0005 that the deviance is less than 0.5% of the deviance of the root, the model has the lowest MSE, which is 155,723,354,073.411. (See Appendix K for Regression Tree results)

Generally, the regression tree has high variance and cross validation error. The bagging tree is used by setting a random sample and run the trees many times and use the average of the samples. As the variance of the sample means is lower than the actual variance of the raw data, we run the bagging tree with number of trees=100 and we find out that the MSE error flattens after 30 trees. The ZIPCODE is the most important variables. The R-squared is 67.81% and the MSE after we run the random splitting cross validation is 113,971,066,657. The MSE is much lower than the regression tree but the bagging tree always use the strong predictor at the root thus generate the similar result and the reduction of variance is small. (See Appendix L for Bagging Tree results)

As a result, we run the Random Forest as we have 16 predictors in the model and we set the number of variables in each of the regression tree to 5. Therefore, it solves the problem that the contribution of variance reduction is small in the bagging tree. And we run 100 trees and find out that the MSE error flattens after 30 trees. We also check the importance of each variable in the model, and we find out the ZIPCODE is the most important variable. The R-squared is 68.44% and the MSE after we run the random splitting cross validation is 106,664,599,063. Therefore, the Random Forest model is the best model to predict the prices of properties in DC. (See Appendix M for Random Forest results)

Cross Validation

We did a cross validation of the MSE for the Plain OLS, OLS Log, Plain WLS, WLS Log, Quadratic polynomial , Cubic polynomial, Interaction, Regression Tree, Bagging Tree, Random Forest Tree models. The Random Forest models is our best model as it has the lowest MSE. The log model for OLS and WLS did not perform as well as the plain mode. Please refer to the results in the table below.

Model	MSE
Plain OLS	139,860,214,097
OLS - Log(PRICE)	664,162,889,919

Plain WLS	139,678,604,484
WLS - Log(PRICE)	664,162,600,009
Quadratic polynomial model	138,182,148,211
Cubic polynomial model	138,193,897,111
Interaction model (QUALIFIED and Continuous of YLRMDL)	138,371,192,960
Regression tree	155,723,354,073
Bagging tree	113,971,066,657
Random Forest tree	106,664,599,063

Learning & Challenges

In order to conduct the analysis, we had to be familiar with our data set and understand the definition of each variables so that we were able to clean the data. While working on data cleaning, we faced challenges such as dealing with none numerical variables, releveling the variables, and transforming variables. In terms of technique issue, we encountered variety of errors while building the models; therefore, we had to conduct research to resolve the issues. As we had a large data set, some of the models took hours to run resulting in R studio application crashing at times. Throughout the project, we have used what we have learned in class and applied the knowledge to our analysis. We have gain stronger capabilities of identifying the variable significance, interpreting results, and selecting the proper models with various specifications.

Conclusion

The final result of our analysis indicates that WLS performs only slightly better than OLS while Heteroscedasticity presents. Throughout the model selection process, we conclude that Regression Tree is comparatively the best model for our data set to predict Washington D.C. house pricing. Random Forest is the best specification to run the model and ZIPCODE is the most important variable. In terms of the significance of ZIPCODE variable, we have also noticed that the most valuable properties are within Georgetown neighborhood in NW Washington D.C. Lastly, our MSE for each model are really high because the data set includes a variety of property types resulting in a wide range for the PRICE. This results in high variance but less bias.

Appendix A: Data Set

```
##     Variables                                                Description
## 1      BATHRM                                          Number of bath room
```

```
## 2   HF_BATHRM                                        Number of half bathroom
## 3        HEAT                                                    Heater type
## 4          AC                                                  Have AC or not
## 5       BEDRM                                               Number of bedrooms
## 6         AYB The earliest time the main portion of the building was built
## 7      YLRMDL          The recent year structure was remodeled or improved
## 8     STORIES                          Number of stories in primary dwelling
## 9       PRICE                                                           Price
## 10  QUALIFIED                                             Qualified or not
## 11        GBA                             Gross building area in square feet
## 12     STRUCT                                                      Structure
## 13      CNDTN                                                      Condition
## 14    EXTWALL                                           Extrerior wall type
## 15       ROOF                                                     Roof type
## 16    INTWALL                                                Interior wall
## 17   KITCHENS                                             Number of kitchen
## 18 FIREPLACES                                           Number of fireplace
## 19    USECODE                                             Property use code
## 20    ZIPCODE                                                       Zipcode
## 21   QUADRANT                                                City quadrant
```

Levels

HEAT

```
## [1] "Reference Level: Air-Oil"

##  [1] "Air-Oil"        "Air Exchng"     "Elec Base Brd"  "Electric Rad"
##  [5] "Evp Cool"       "Forced Air"     "Gravity Furnac" "Hot Water Rad"
##  [9] "Ht Pump"        "Ind Unit"       "No Data"        "Wall Furnace"
## [13] "Warm Cool"      "Water Base Brd"
```

AC

```
## [1] "Reference Level: N"

## [1] "N" "Y"
```

STORIES

```
## [1] "Reference Level: 1"

## [1] "1" "2" "3" "4" "9"
```

QUALIFIED

```
## [1] "Reference Level: N"

## [1] "Q" "U"
```

STRUCT

```
## [1] "Reference Level: Multi"
```

```
## [1] "Multi"         ""              "Row End"       "Row Inside"
## [5] "Semi-Detached" "Single"        "Town End"      "Town Inside"
```

CNDTN

```
## [1] "Reference Level: Average"

## [1] "Average"   ""          "Excellent" "Fair"      "Good"      "Poor"
## [7] "Very Good"
```

EXTWALL

```
## [1] "Reference Level: Adobe"

##  [1] "Adobe"     ""          "Aluminum"  "Brick"     "Concrete"
##  [6] "Default"   "Hardboard" "Plywood"   "Shingle"   "Siding"
## [11] "SPlaster"  "Stone"     "Stucco"
```

ROOF

```
## [1] "Reference Level: Built Up"

##  [1] "Built Up"     ""             "Clay Tile"    "Comp Shingle"
##  [5] "Concrete"     "Metal"        "Neopren"      "Shake"
##  [9] "Shingle"      "Slate"        "Typical"
```

INT WALL

```
## [1] "Reference Level: Carpet"

## [1] "Carpet"       ""             "Ceramic Tile" "Default"
## [5] "Hardwood"     "Lt Concrete"  "Vinyl Sheet"
```

USECODE

```
## [1] "Reference Level: 11"

## [1] "11" "12" "13" "15" "23" "24"
```

ZIPCODE

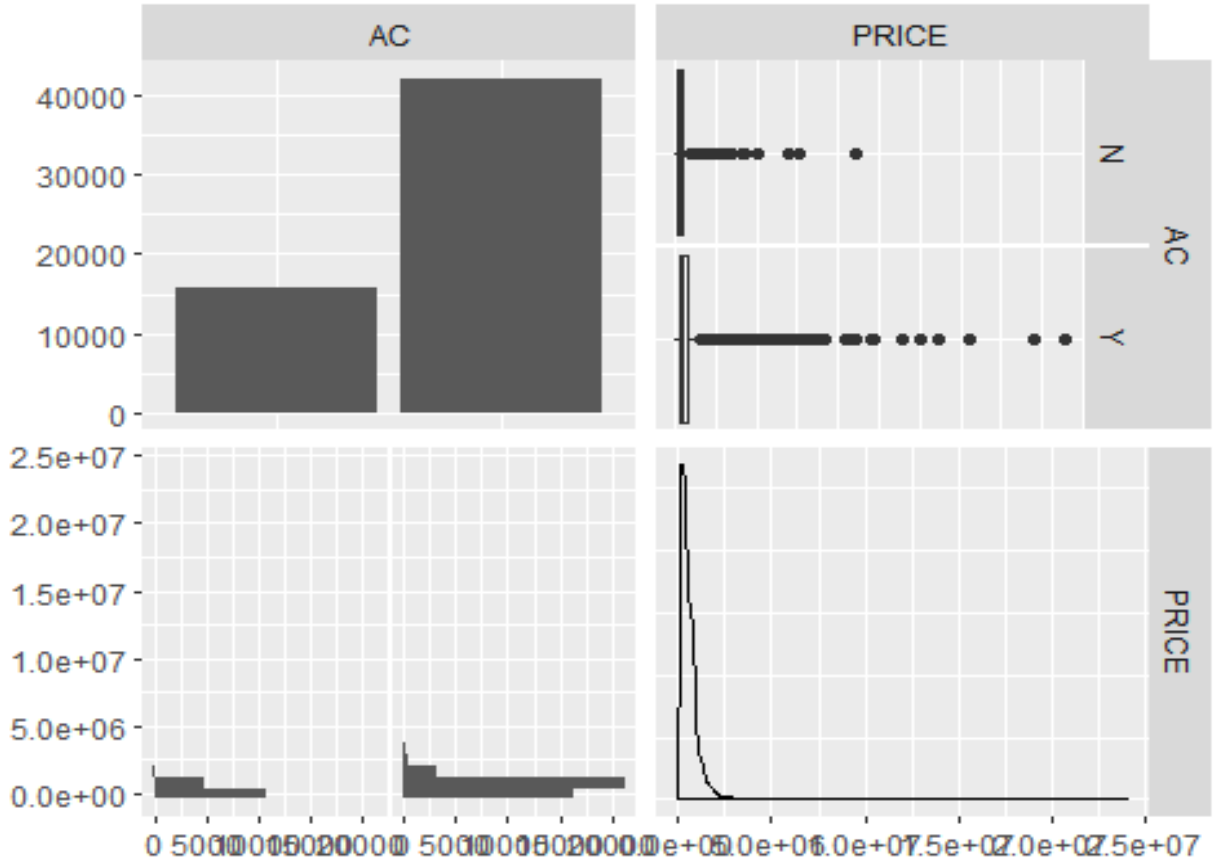
```
## [1] "Reference Level: 20001"

##  [1] "20001" "20002" "20003" "20005" "20007" "20008" "20009" "20010"
##  [9] "20011" "20012" "20015" "20016" "20017" "20018" "20019" "20020"
## [17] "20024" "20032" "20036" "20037"
```

QUADRANT

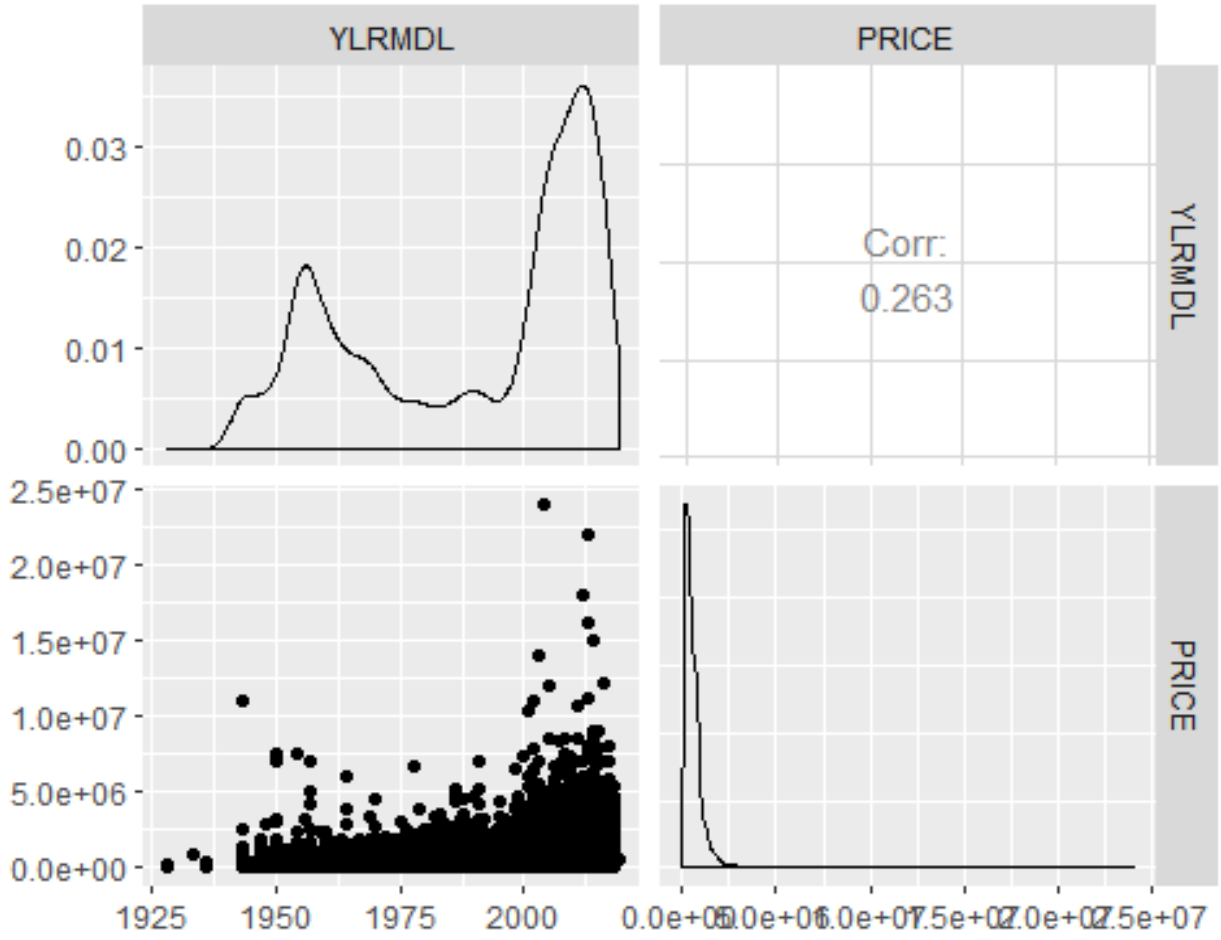
```
## [1] "Reference Level:NE"

## [1] "NE" ""   "NW" "SE" "SW"
```

Appendix B: Descriptive Stastics and Correlation Analysis

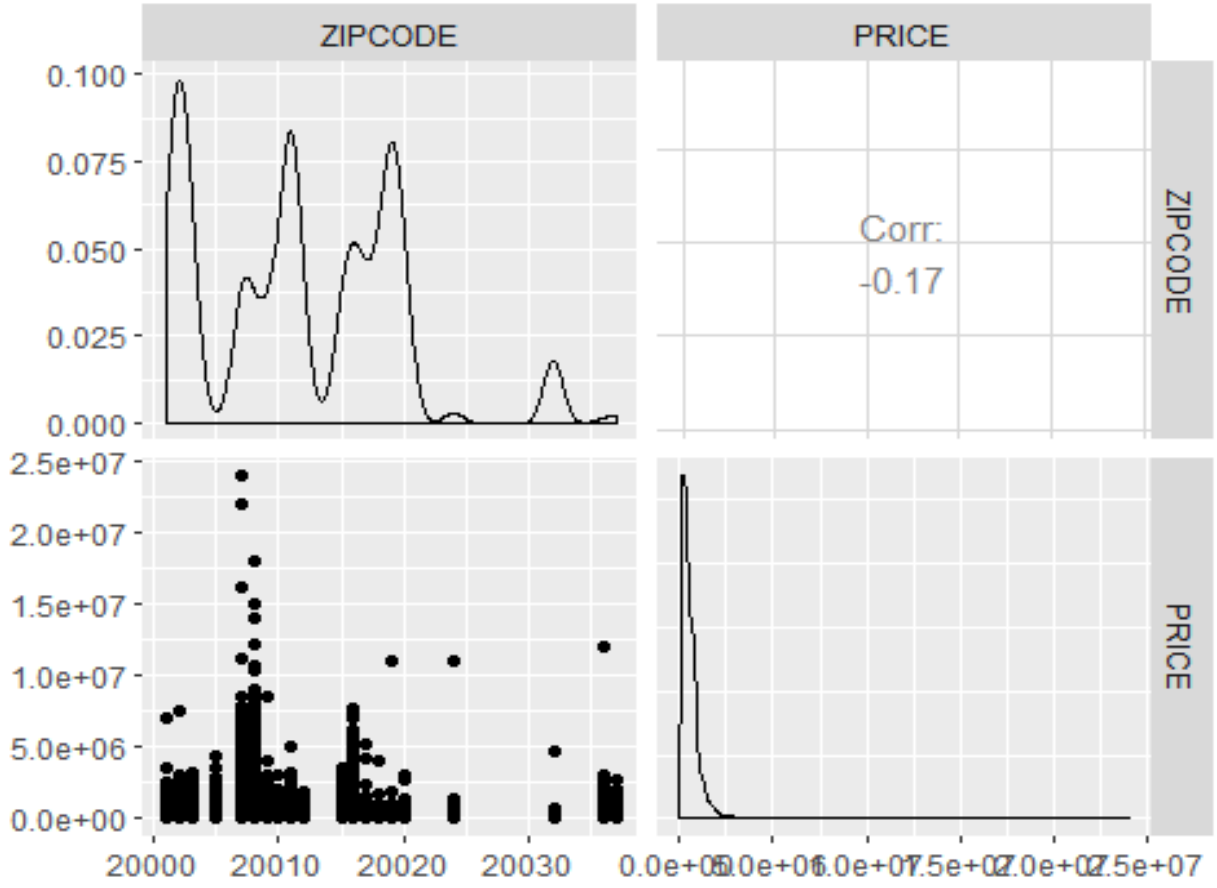
ggpair1



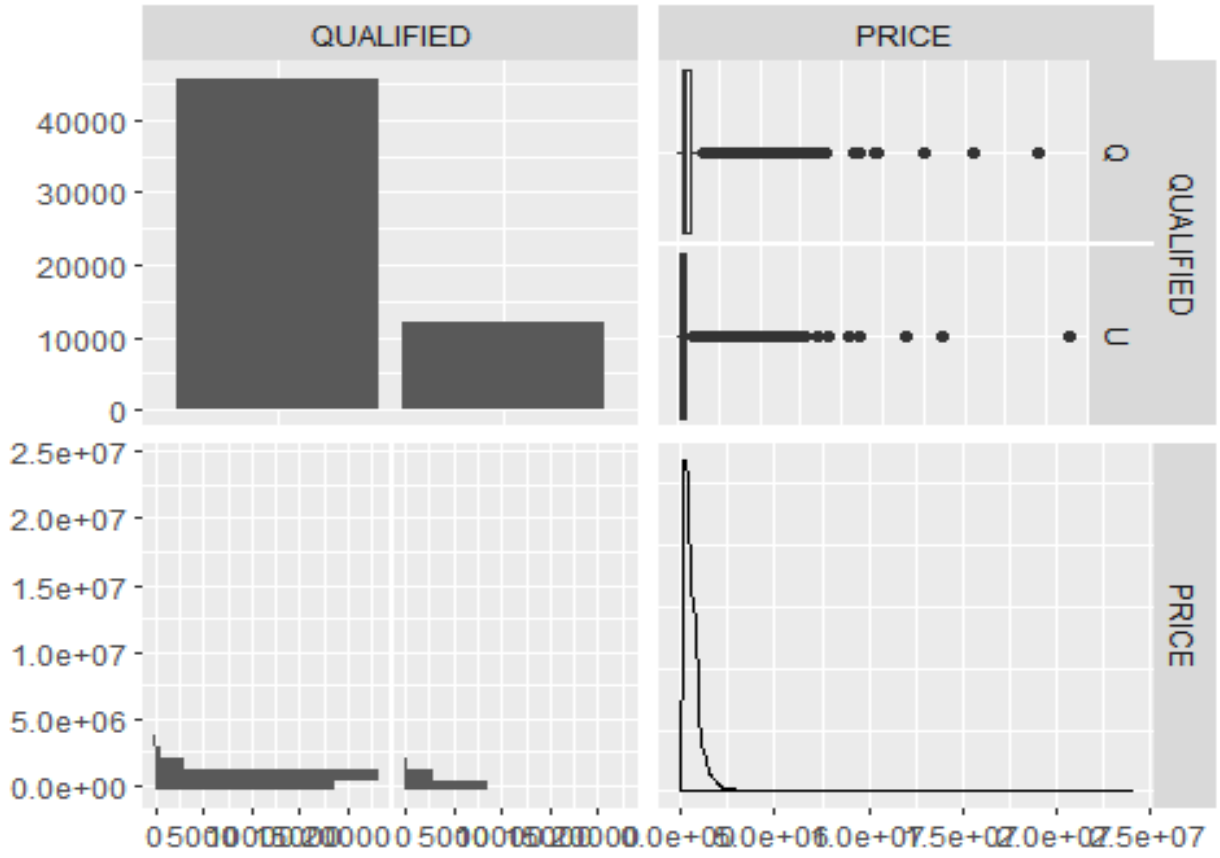
ggpair2


YLRMDL
PRICE
YLRMDL
PRICE
Corr:
0.263
0.03
0.02
0.01
0.00
2.5e+07
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
1925
1950
1975
2000
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07

ggpair3


ZIPCODE
PRICE
0.100
0.075
0.050
0.025
0.000
Corr:
-0.17
ZIPCODE
2.5e+07
2.0e+07
1.5e+07
1.0e+07
5.0e+06
0.0e+00
PRICE
20000
20010
20020
20030
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07

ggpair4



Appendix C: OLS

```
##
## Call:
## lm(formula = PRICE ~ ., data = dc_completed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2988452  -154054     7011   147261 18366882
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.413e+06  4.476e+05  -3.157 0.001596 **
## BATHRM               5.949e+04  2.630e+03  22.614  < 2e-16 ***
## HF_BATHRM            2.127e+04  3.045e+03   6.986 2.86e-12 ***
## HEATAir Exchng      -1.124e+05  1.142e+05  -0.984 0.325150
## HEATElec Base Brd   -4.122e+03  7.587e+04  -0.054 0.956671
## HEATElectric Rad    -1.008e+05  8.621e+04  -1.170 0.242146
## HEATEvp Cool        -6.323e+04  1.192e+05  -0.530 0.595899
## HEATForced Air      -2.614e+04  6.043e+04  -0.433 0.665364
## HEATGravity Furnac   8.712e+03  8.732e+04   0.100 0.920529
```

```
## HEATHot Water Rad   -1.843e+04  6.037e+04  -0.305 0.760176
## HEATHt Pump         -8.143e+03  6.193e+04  -0.131 0.895387
## HEATInd Unit        -7.533e+04  1.682e+05  -0.448 0.654214
## HEATNo Data         -2.526e+04  1.057e+05  -0.239 0.811132
## HEATWall Furnace    -2.338e+04  7.333e+04  -0.319 0.749855
## HEATWarm Cool       -1.851e+04  6.049e+04  -0.306 0.759637
## HEATWater Base Brd   3.777e+04  7.221e+04   0.523 0.600901
## ACY                 -1.797e+03  5.293e+03  -0.340 0.734184
## BEDRM               -1.258e+04  2.093e+03  -6.010 1.86e-09 ***
## AYB                 -1.771e+02  6.462e+01  -2.740 0.006150 **
## YLRMDL               8.182e+02  9.334e+01   8.765  < 2e-16 ***
## STORIES2            -9.832e+04  9.376e+03 -10.486  < 2e-16 ***
## STORIES3            -1.519e+05  1.049e+04 -14.483  < 2e-16 ***
## STORIES4            -1.287e+05  2.435e+04  -5.285 1.26e-07 ***
## STORIES9            -3.653e+04  9.654e+04  -0.378 0.705147
## QUALIFIEDU          -9.643e+04  4.128e+03 -23.362  < 2e-16 ***
## GBA                  2.936e+02  3.574e+00  82.146  < 2e-16 ***
## STRUCTRow End        7.606e+04  1.953e+04   3.895 9.84e-05 ***
## STRUCTRow Inside     8.068e+04  1.922e+04   4.199 2.69e-05 ***
## STRUCTSemi-Detached  8.980e+04  2.430e+04   3.695 0.000220 ***
## STRUCTSingle         6.059e+04  2.605e+04   2.326 0.020023 *
## STRUCTTown End       1.155e+05  4.892e+04   2.360 0.018259 *
## STRUCTTown Inside    8.541e+04  3.589e+04   2.380 0.017320 *
## CNDTNExcellent       4.538e+05  1.440e+04  31.512  < 2e-16 ***
## CNDTNFair            7.805e+04  1.746e+04   4.470 7.84e-06 ***
## CNDTNGood            9.115e+04  4.352e+03  20.945  < 2e-16 ***
## CNDTNPoor            2.214e+05  4.143e+04   5.345 9.06e-08 ***
## CNDTNVery Good       2.966e+05  6.686e+03  44.352  < 2e-16 ***
## EXTWALLAluminum      2.050e+05  3.848e+05   0.533 0.594237
## EXTWALLBrick         1.163e+05  3.844e+05   0.302 0.762335
## EXTWALLConcrete      1.041e+05  3.870e+05   0.269 0.787875
## EXTWALLDefault       3.566e+04  4.052e+05   0.088 0.929880
## EXTWALLHardboard     1.519e+05  3.870e+05   0.392 0.694740
## EXTWALLPlywood       3.509e+05  4.305e+05   0.815 0.414965
## EXTWALLShingle       1.579e+05  3.848e+05   0.410 0.681579
## EXTWALLSiding        1.252e+05  3.845e+05   0.326 0.744681
## EXTWALLSPlaster      3.517e+05  5.474e+05   0.643 0.520544
## EXTWALLStone         1.871e+05  3.847e+05   0.487 0.626601
## EXTWALLStucco        1.820e+05  3.845e+05   0.473 0.636013
## ROOFClay Tile       -5.142e+04  2.271e+04  -2.264 0.023576 *
## ROOFComp Shingle    -1.082e+05  5.792e+03 -18.679  < 2e-16 ***
## ROOFConcrete         3.785e+05  1.114e+05   3.396 0.000683 ***
## ROOFMetal            5.109e+03  4.564e+03   1.119 0.263006
## ROOFNeopren          5.151e+04  1.358e+04   3.792 0.000150 ***
## ROOFShake           -1.273e+05  2.049e+04  -6.210 5.32e-10 ***
## ROOFShingle         -7.111e+04  2.500e+04  -2.845 0.004447 **
## ROOFSlate           -3.280e+04  7.939e+03  -4.131 3.62e-05 ***
## ROOFTypical         -1.033e+05  4.212e+04  -2.453 0.014183 *
## INTWALLCeramic Tile  1.254e+05  6.232e+04   2.012 0.044197 *
## INTWALLDefault       6.150e+04  6.490e+04   0.948 0.343336
```

```
## INTWALLHardwood      2.881e+04  9.118e+03   3.159 0.001581 **
## INTWALLLt Concrete  -1.080e+05  5.582e+04  -1.935 0.052986 .
## INTWALLVinyl Sheet   1.165e+06  7.399e+04  15.747  < 2e-16 ***
## KITCHENS            -1.232e+04  5.090e+03  -2.420 0.015507 *
## FIREPLACES           8.676e+04  2.339e+03  37.095  < 2e-16 ***
## USECODE12            2.452e+04  2.063e+04   1.188 0.234687
## USECODE13            2.641e+04  1.829e+04   1.443 0.148906
## USECODE15            1.277e+04  6.727e+04   0.190 0.849504
## USECODE23           -2.466e+05  2.145e+04 -11.497  < 2e-16 ***
## USECODE24           -3.224e+04  8.236e+03  -3.915 9.06e-05 ***
## ZIPCODE20002        -7.709e+04  1.434e+04  -5.374 7.72e-08 ***
## ZIPCODE20003        -7.004e+03  1.820e+04  -0.385 0.700377
## ZIPCODE20005         1.030e+05  3.468e+04   2.971 0.002967 **
## ZIPCODE20007         3.676e+05  1.027e+04  35.798  < 2e-16 ***
## ZIPCODE20008         2.834e+05  1.245e+04  22.759  < 2e-16 ***
## ZIPCODE20009         8.153e+04  1.138e+04   7.166 7.79e-13 ***
## ZIPCODE20010        -7.446e+03  1.090e+04  -0.683 0.494727
## ZIPCODE20011        -9.603e+04  8.951e+03 -10.728  < 2e-16 ***
## ZIPCODE20012        -1.438e+05  1.265e+04 -11.370  < 2e-16 ***
## ZIPCODE20015         3.619e+03  1.165e+04   0.311 0.756099
## ZIPCODE20016         1.637e+05  1.102e+04  14.853  < 2e-16 ***
## ZIPCODE20017        -1.161e+05  1.587e+04  -7.318 2.55e-13 ***
## ZIPCODE20018        -1.954e+05  1.586e+04 -12.322  < 2e-16 ***
## ZIPCODE20019        -1.828e+05  1.495e+04 -12.230  < 2e-16 ***
## ZIPCODE20020        -2.095e+05  1.785e+04 -11.734  < 2e-16 ***
## ZIPCODE20024        -1.010e+05  3.542e+04  -2.851 0.004356 **
## ZIPCODE20032        -2.574e+05  1.920e+04 -13.408  < 2e-16 ***
## ZIPCODE20036         2.994e+05  4.243e+04   7.055 1.74e-12 ***
## ZIPCODE20037         2.148e+05  2.840e+04   7.564 3.97e-14 ***
## QUADRANT            -3.052e+04  3.882e+04  -0.786 0.431662
## QUADRANTNW          -5.281e+04  1.176e+04  -4.491 7.11e-06 ***
## QUADRANTSE          -1.238e+04  9.806e+03  -1.262 0.206806
## QUADRANTSW           9.549e+04  2.467e+04   3.870 0.000109 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 384300 on 57580 degrees of freedom
## Multiple R-squared:  0.5756, Adjusted R-squared:  0.5749
## F-statistic: 858.2 on 91 and 57580 DF,  p-value: < 2.2e-16
```

Appendix D: Variable Selection (Stepwise Method)

```
## Start:  AIC=1483330
## PRICE ~ BATHRM + HF_BATHRM + HEAT + AC + BEDRM + AYB + YLRMDL +
##     STORIES + QUALIFIED + GBA + STRUCT + CNDTN + EXTWALL + ROOF +
##     INTWALL + KITCHENS + FIREPLACES + USECODE + ZIPCODE + QUADRANT
##
##              Df  Sum of Sq        RSS     AIC   F value    Pr(>F)
## - HEAT       13 1.5548e+12 8.5066e+15 1483314    0.8097  0.650447
```

```
## - AC          1 1.7032e+10 8.5051e+15 1483328    0.1153  0.734184
## <none>                     8.5050e+15 1483330
## - KITCHENS    1 8.6531e+11 8.5059e+15 1483333    5.8583  0.015507 *
## - AYB         1 1.1088e+12 8.5062e+15 1483335    7.5065  0.006150 **
## - STRUCT      6 3.0517e+12 8.5081e+15 1483338    3.4434  0.002114 **
## - BEDRM       1 5.3360e+12 8.5104e+15 1483364   36.1251 1.862e-09 ***
## - QUADRANT    4 6.5549e+12 8.5116e+15 1483366   11.0943 5.398e-09 ***
## - HF_BATHRM   1 7.2083e+12 8.5123e+15 1483376   48.8007 2.864e-12 ***
## - EXTWALL    11 1.4066e+13 8.5191e+15 1483403    8.6571 1.617e-15 ***
## - YLRMDL      1 1.1349e+13 8.5164e+15 1483404   76.8316 < 2.2e-16 ***
## - USECODE     5 2.1413e+13 8.5265e+15 1483465   28.9931 < 2.2e-16 ***
## - STORIES     4 3.4309e+13 8.5394e+15 1483554   58.0693 < 2.2e-16 ***
## - INTWALL     5 3.8229e+13 8.5433e+15 1483578   51.7622 < 2.2e-16 ***
## - ROOF        9 7.1398e+13 8.5764e+15 1483794   53.7078 < 2.2e-16 ***
## - BATHRM      1 7.5539e+13 8.5806e+15 1483838  511.4043 < 2.2e-16 ***
## - QUALIFIED   1 8.0614e+13 8.5857e+15 1483872  545.7676 < 2.2e-16 ***
## - FIREPLACES  1 2.0325e+14 8.7083e+15 1484690 1376.0327 < 2.2e-16 ***
## - CNDTN       5 3.6551e+14 8.8706e+15 1485746  494.9077 < 2.2e-16 ***
## - ZIPCODE    19 6.6424e+14 9.1693e+15 1487629  236.6827 < 2.2e-16 ***
## - GBA         1 9.9673e+14 9.5018e+15 1489719 6747.9786 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=1483314
## PRICE ~ BATHRM + HF_BATHRM + AC + BEDRM + AYB + YLRMDL + STORIES +
##     QUALIFIED + GBA + STRUCT + CNDTN + EXTWALL + ROOF + INTWALL +
##     KITCHENS + FIREPLACES + USECODE + ZIPCODE + QUADRANT
##
##              Df  Sum of Sq        RSS     AIC   F value    Pr(>F)
## - AC          1 8.1817e+10 8.5067e+15 1483313    0.5539  0.456718
## <none>                     8.5066e+15 1483314
## - KITCHENS    1 8.0339e+11 8.5074e+15 1483318    5.4393  0.019692 *
## - AYB         1 1.1038e+12 8.5077e+15 1483320    7.4731  0.006265 **
## - STRUCT      6 3.0989e+12 8.5097e+15 1483323    3.4968  0.001852 **
## + HEAT       13 1.5548e+12 8.5050e+15 1483330    0.8097  0.650447
## - BEDRM       1 5.3152e+12 8.5119e+15 1483348   35.9858 1.999e-09 ***
## - QUADRANT    4 6.5359e+12 8.5131e+15 1483350   11.0627 5.735e-09 ***
## - HF_BATHRM   1 7.1569e+12 8.5138e+15 1483361   48.4553 3.415e-12 ***
## - YLRMDL      1 1.1033e+13 8.5176e+15 1483387   74.7010 < 2.2e-16 ***
## - EXTWALL    11 1.4237e+13 8.5208e+15 1483389    8.7629 9.537e-16 ***
## - USECODE     5 2.1453e+13 8.5281e+15 1483449   29.0494 < 2.2e-16 ***
## - STORIES     4 3.4259e+13 8.5409e+15 1483538   57.9865 < 2.2e-16 ***
## - INTWALL     5 3.8212e+13 8.5448e+15 1483563   51.7420 < 2.2e-16 ***
## - ROOF        9 7.1619e+13 8.5782e+15 1483780   53.8767 < 2.2e-16 ***
## - BATHRM      1 7.5201e+13 8.5818e+15 1483820  509.1393 < 2.2e-16 ***
## - QUALIFIED   1 8.0690e+13 8.5873e+15 1483857  546.3003 < 2.2e-16 ***
## - FIREPLACES  1 2.0556e+14 8.7122e+15 1484689 1391.7374 < 2.2e-16 ***
## - CNDTN       5 3.6880e+14 8.8754e+15 1485752  499.3833 < 2.2e-16 ***
## - ZIPCODE    19 6.8800e+14 9.1946e+15 1487762  245.1601 < 2.2e-16 ***
## - GBA         1 1.0056e+15 9.5122e+15 1489756 6808.2876 < 2.2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Step:  AIC=1483313
## PRICE ~ BATHRM + HF_BATHRM + BEDRM + AYB + YLRMDL + STORIES +
##     QUALIFIED + GBA + STRUCT + CNDTN + EXTWALL + ROOF + INTWALL +
##     KITCHENS + FIREPLACES + USECODE + ZIPCODE + QUADRANT
##
##              Df  Sum of Sq        RSS     AIC   F value    Pr(>F)
## <none>                     8.5067e+15 1483313
## + AC          1 8.1817e+10 8.5066e+15 1483314    0.5539  0.456718
## - KITCHENS    1 7.8302e+11 8.5075e+15 1483316    5.3014  0.021312 *
## - AYB         1 1.1002e+12 8.5078e+15 1483318    7.4489  0.006350 **
## - STRUCT      6 3.0921e+12 8.5098e+15 1483322    3.4892  0.001887 **
## + HEAT       13 1.6196e+12 8.5051e+15 1483328    0.8435  0.613760
## - BEDRM       1 5.2792e+12 8.5120e+15 1483346   35.7424 2.265e-09 ***
## - QUADRANT    4 6.4695e+12 8.5132e+15 1483349   10.9503 7.109e-09 ***
## - HF_BATHRM   1 7.0809e+12 8.5138e+15 1483359   47.9410 4.438e-12 ***
## - YLRMDL      1 1.1327e+13 8.5180e+15 1483387   76.6863 < 2.2e-16 ***
## - EXTWALL    11 1.4331e+13 8.5210e+15 1483388    8.8206 7.153e-16 ***
## - USECODE     5 2.1415e+13 8.5281e+15 1483448   28.9980 < 2.2e-16 ***
## - STORIES     4 3.4191e+13 8.5409e+15 1483536   57.8713 < 2.2e-16 ***
## - INTWALL     5 3.8274e+13 8.5450e+15 1483562   51.8260 < 2.2e-16 ***
## - ROOF        9 7.1798e+13 8.5785e+15 1483779   54.0116 < 2.2e-16 ***
## - BATHRM      1 7.5793e+13 8.5825e+15 1483822  513.1525 < 2.2e-16 ***
## - QUALIFIED   1 8.0624e+13 8.5873e+15 1483855  545.8577 < 2.2e-16 ***
## - FIREPLACES  1 2.0564e+14 8.7123e+15 1484688 1392.2434 < 2.2e-16 ***
## - CNDTN       5 3.7038e+14 8.8771e+15 1485761  501.5310 < 2.2e-16 ***
## - ZIPCODE    19 6.9257e+14 9.1993e+15 1487789  246.7896 < 2.2e-16 ***
## - GBA         1 1.0109e+15 9.5176e+15 1489787 6844.4538 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## lm(formula = PRICE ~ BATHRM + HF_BATHRM + BEDRM + AYB + YLRMDL +
##     STORIES + QUALIFIED + GBA + STRUCT + CNDTN + EXTWALL + ROOF +
##     INTWALL + KITCHENS + FIREPLACES + USECODE + ZIPCODE + QUADRANT,
##     data = dc_completed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2987904  -154117     6938   146880 18372495
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.365e+06  4.407e+05  -3.098 0.001952 **
## BATHRM               5.890e+04  2.600e+03  22.653  < 2e-16 ***
## HF_BATHRM            2.087e+04  3.013e+03   6.924 4.44e-12 ***
## BEDRM               -1.249e+04  2.089e+03  -5.978 2.27e-09 ***
```

```
## AYB                  -1.763e+02  6.461e+01  -2.729 0.006350 **
## YLRMDL                7.818e+02  8.928e+01   8.757  < 2e-16 ***
## STORIES2             -9.793e+04  9.359e+03 -10.464  < 2e-16 ***
## STORIES3             -1.514e+05  1.047e+04 -14.459  < 2e-16 ***
## STORIES4             -1.279e+05  2.434e+04  -5.254 1.49e-07 ***
## STORIES9             -3.526e+04  9.653e+04  -0.365 0.714885
## QUALIFIEDU           -9.632e+04  4.123e+03 -23.364  < 2e-16 ***
## GBA                   2.941e+02  3.555e+00  82.731  < 2e-16 ***
## STRUCTRow End         7.657e+04  1.952e+04   3.924 8.73e-05 ***
## STRUCTRow Inside      8.137e+04  1.920e+04   4.237 2.27e-05 ***
## STRUCTSemi-Detached   8.990e+04  2.430e+04   3.700 0.000216 ***
## STRUCTSingle          6.131e+04  2.604e+04   2.355 0.018533 *
## STRUCTTown End        1.159e+05  4.892e+04   2.370 0.017805 *
## STRUCTTown Inside     8.690e+04  3.588e+04   2.422 0.015436 *
## CNDTNExcellent        4.513e+05  1.435e+04  31.456  < 2e-16 ***
## CNDTNFair             7.823e+04  1.743e+04   4.487 7.24e-06 ***
## CNDTNGood             9.018e+04  4.281e+03  21.067  < 2e-16 ***
## CNDTNPoor             2.201e+05  4.059e+04   5.423 5.90e-08 ***
## CNDTNVery Good        2.943e+05  6.585e+03  44.693  < 2e-16 ***
## EXTWALLAluminum       2.031e+05  3.848e+05   0.528 0.597615
## EXTWALLBrick          1.134e+05  3.844e+05   0.295 0.767954
## EXTWALLConcrete       1.006e+05  3.870e+05   0.260 0.794978
## EXTWALLDefault        3.399e+04  4.052e+05   0.084 0.933156
## EXTWALLHardboard      1.508e+05  3.870e+05   0.390 0.696725
## EXTWALLPlywood        3.260e+05  4.298e+05   0.759 0.448104
## EXTWALLShingle        1.560e+05  3.848e+05   0.405 0.685120
## EXTWALLSiding         1.226e+05  3.844e+05   0.319 0.749786
## EXTWALLSPlaster       3.692e+05  5.441e+05   0.679 0.497434
## EXTWALLStone          1.847e+05  3.846e+05   0.480 0.631030
## EXTWALLStucco         1.796e+05  3.845e+05   0.467 0.640492
## ROOFClay Tile        -5.119e+04  2.270e+04  -2.255 0.024147 *
## ROOFComp Shingle     -1.081e+05  5.782e+03 -18.691  < 2e-16 ***
## ROOFConcrete          3.808e+05  1.114e+05   3.418 0.000632 ***
## ROOFMetal             5.469e+03  4.546e+03   1.203 0.228972
## ROOFNeopren           5.103e+04  1.358e+04   3.758 0.000171 ***
## ROOFShake            -1.275e+05  2.047e+04  -6.232 4.64e-10 ***
## ROOFShingle          -7.169e+04  2.499e+04  -2.869 0.004124 **
## ROOFSlate            -3.258e+04  7.932e+03  -4.108 4.00e-05 ***
## ROOFTypical          -1.027e+05  4.211e+04  -2.439 0.014726 *
## INTWALLCeramic Tile   1.253e+05  6.231e+04   2.011 0.044336 *
## INTWALLDefault        6.214e+04  6.429e+04   0.966 0.333827
## INTWALLHardwood       2.868e+04  9.097e+03   3.152 0.001620 **
## INTWALLLt Concrete   -1.091e+05  5.572e+04  -1.958 0.050284 .
## INTWALLVinyl Sheet    1.165e+06  7.394e+04  15.756  < 2e-16 ***
## KITCHENS             -1.169e+04  5.077e+03  -2.302 0.021312 *
## FIREPLACES            8.697e+04  2.331e+03  37.313  < 2e-16 ***
## USECODE12             2.455e+04  2.063e+04   1.190 0.234015
## USECODE13             2.671e+04  1.829e+04   1.460 0.144230
## USECODE15             1.419e+04  6.726e+04   0.211 0.832926
## USECODE23            -2.464e+05  2.144e+04 -11.493  < 2e-16 ***
```

```
## USECODE24           -3.271e+04  8.224e+03  -3.978 6.96e-05 ***
## ZIPCODE20002        -7.472e+04  1.424e+04  -5.246 1.56e-07 ***
## ZIPCODE20003        -4.405e+03  1.810e+04  -0.243 0.807762
## ZIPCODE20005         1.041e+05  3.467e+04   3.003 0.002675 **
## ZIPCODE20007         3.709e+05  1.008e+04  36.810  < 2e-16 ***
## ZIPCODE20008         2.851e+05  1.240e+04  22.996  < 2e-16 ***
## ZIPCODE20009         8.337e+04  1.132e+04   7.368 1.76e-13 ***
## ZIPCODE20010        -5.553e+03  1.085e+04  -0.512 0.608937
## ZIPCODE20011        -9.448e+04  8.897e+03 -10.619  < 2e-16 ***
## ZIPCODE20012        -1.426e+05  1.262e+04 -11.302  < 2e-16 ***
## ZIPCODE20015         6.074e+03  1.156e+04   0.525 0.599414
## ZIPCODE20016         1.659e+05  1.089e+04  15.237  < 2e-16 ***
## ZIPCODE20017        -1.142e+05  1.580e+04  -7.230 4.90e-13 ***
## ZIPCODE20018        -1.936e+05  1.577e+04 -12.277  < 2e-16 ***
## ZIPCODE20019        -1.820e+05  1.491e+04 -12.209  < 2e-16 ***
## ZIPCODE20020        -2.067e+05  1.779e+04 -11.621  < 2e-16 ***
## ZIPCODE20024        -9.786e+04  3.539e+04  -2.765 0.005687 **
## ZIPCODE20032        -2.545e+05  1.914e+04 -13.299  < 2e-16 ***
## ZIPCODE20036         3.024e+05  4.238e+04   7.136 9.74e-13 ***
## ZIPCODE20037         2.173e+05  2.834e+04   7.668 1.77e-14 ***
## QUADRANT            -2.966e+04  3.881e+04  -0.764 0.444633
## QUADRANTNW          -5.129e+04  1.166e+04  -4.398 1.09e-05 ***
## QUADRANTSE          -1.247e+04  9.804e+03  -1.272 0.203373
## QUADRANTSW           9.568e+04  2.466e+04   3.880 0.000104 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 384300 on 57594 degrees of freedom
## Multiple R-squared:  0.5755, Adjusted R-squared:  0.5749
## F-statistic:  1014 on 77 and 57594 DF,  p-value: < 2.2e-16
```

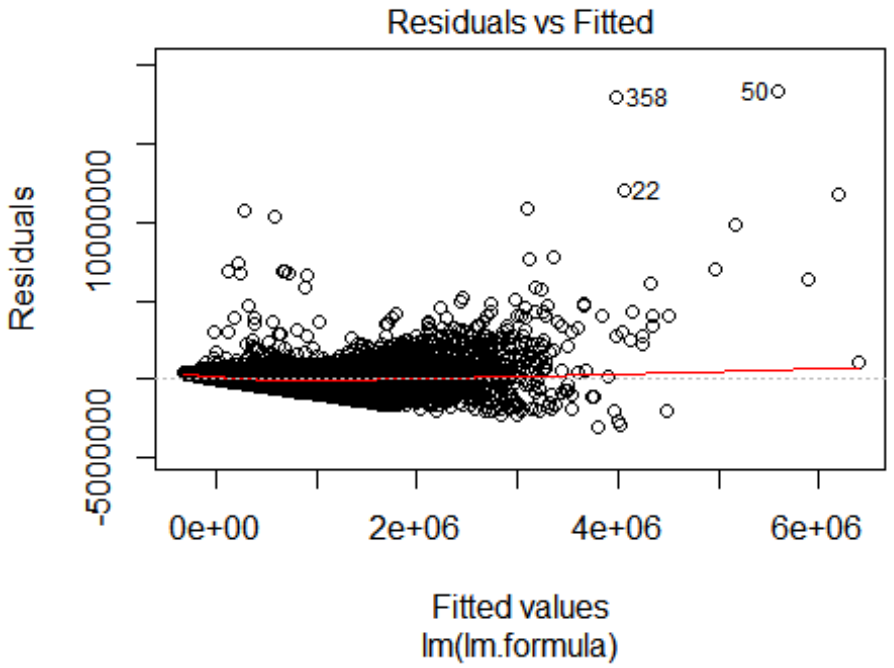
Appendix E: OLS Assumptions

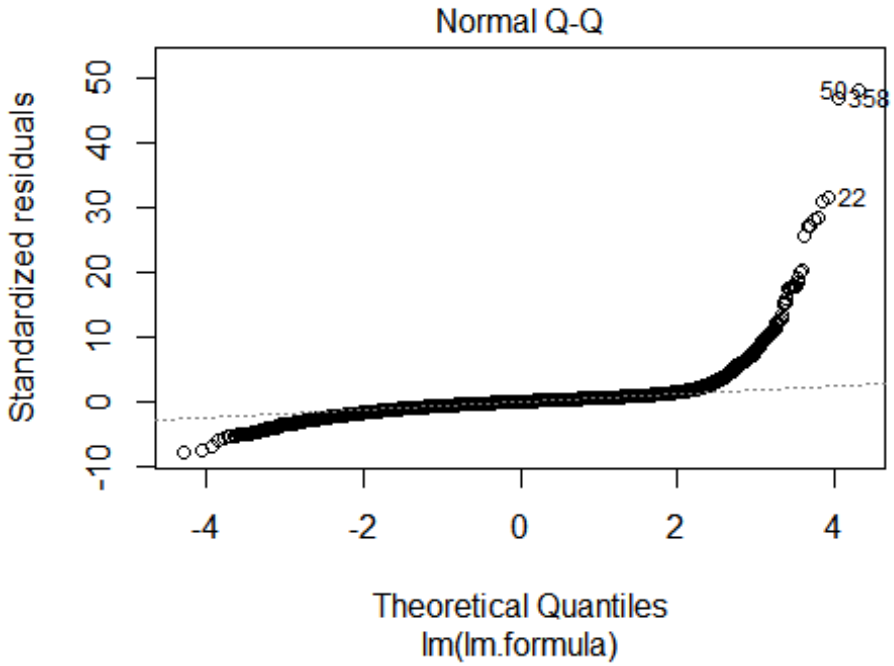
```
##
## Call:
## lm(formula = lm.formula, data = dc_completed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3001573  -154646     6552   147153 18364323
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -1.274e+06  2.167e+05  -5.877 4.19e-09 ***
## BATHRM               5.830e+04  2.597e+03  22.451  < 2e-16 ***
## HF_BATHRM            2.024e+04  3.008e+03   6.729 1.72e-11 ***
## BEDRM               -1.276e+04  2.090e+03  -6.108 1.02e-09 ***
## AYB                 -1.856e+02  6.465e+01  -2.871 0.004094 **
## YLRMDL               7.806e+02  8.923e+01   8.749  < 2e-16 ***
```

```
## STORIES2             -9.790e+04  9.357e+03 -10.463  < 2e-16 ***
## STORIES3             -1.497e+05  1.046e+04 -14.309  < 2e-16 ***
## STORIES4             -1.272e+05  2.436e+04  -5.219 1.80e-07 ***
## STORIES9             -3.477e+04  9.662e+04  -0.360 0.718926
## QUALIFIEDU           -9.563e+04  4.123e+03 -23.195  < 2e-16 ***
## GBA                   2.956e+02  3.540e+00  83.508  < 2e-16 ***
## STRUCTRow End         7.455e+04  1.953e+04   3.817 0.000135 ***
## STRUCTRow Inside      7.805e+04  1.922e+04   4.062 4.88e-05 ***
## STRUCTSemi-Detached   9.137e+04  2.432e+04   3.757 0.000172 ***
## STRUCTSingle          6.584e+04  2.605e+04   2.528 0.011481 *
## STRUCTTown End        1.373e+05  4.869e+04   2.819 0.004815 **
## STRUCTTown Inside     1.275e+05  3.494e+04   3.648 0.000264 ***
## CNDTNExcellent        4.517e+05  1.433e+04  31.528  < 2e-16 ***
## CNDTNFair             8.145e+04  1.744e+04   4.670 3.02e-06 ***
## CNDTNGood             9.027e+04  4.282e+03  21.078  < 2e-16 ***
## CNDTNPoor             2.225e+05  4.062e+04   5.477 4.35e-08 ***
## CNDTNVery Good        2.954e+05  6.586e+03  44.859  < 2e-16 ***
## ROOFClay Tile        -3.986e+04  2.262e+04  -1.762 0.078069 .
## ROOFComp Shingle     -1.068e+05  5.689e+03 -18.774  < 2e-16 ***
## ROOFConcrete          3.753e+05  1.115e+05   3.364 0.000767 ***
## ROOFMetal             3.948e+03  4.515e+03   0.875 0.381800
## ROOFNeopren           4.961e+04  1.359e+04   3.651 0.000262 ***
## ROOFShake            -1.311e+05  2.044e+04  -6.412 1.45e-10 ***
## ROOFShingle          -7.213e+04  2.500e+04  -2.885 0.003915 **
## ROOFSlate            -3.852e+04  7.885e+03  -4.885 1.03e-06 ***
## ROOFTypical          -1.026e+05  4.215e+04  -2.435 0.014895 *
## INTWALLCeramic Tile   1.178e+05  6.235e+04   1.890 0.058809 .
## INTWALLDefault        6.031e+04  6.433e+04   0.937 0.348510
## INTWALLHardwood       2.504e+04  9.016e+03   2.777 0.005488 **
## INTWALLLt Concrete   -1.085e+05  5.576e+04  -1.946 0.051679 .
## INTWALLVinyl Sheet    1.179e+06  7.389e+04  15.951  < 2e-16 ***
## KITCHENS             -1.232e+04  5.079e+03  -2.426 0.015257 *
## FIREPLACES            8.616e+04  2.327e+03  37.032  < 2e-16 ***
## USECODE12             2.871e+04  2.064e+04   1.391 0.164242
## USECODE13             2.864e+04  1.829e+04   1.566 0.117468
## USECODE15             8.075e+03  6.732e+04   0.120 0.904519
## USECODE23            -2.474e+05  2.145e+04 -11.534  < 2e-16 ***
## USECODE24            -3.268e+04  8.228e+03  -3.972 7.14e-05 ***
## ZIPCODE20002         -2.318e+04  8.228e+03  -2.818 0.004838 **
## ZIPCODE20003          3.591e+04  9.608e+03   3.737 0.000186 ***
## ZIPCODE20005          1.027e+05  3.469e+04   2.959 0.003086 **
## ZIPCODE20007          3.715e+05  1.008e+04  36.844  < 2e-16 ***
## ZIPCODE20008          2.864e+05  1.240e+04  23.086  < 2e-16 ***
## ZIPCODE20009          8.400e+04  1.132e+04   7.417 1.21e-13 ***
## ZIPCODE20010         -3.881e+03  1.086e+04  -0.357 0.720722
## ZIPCODE20011         -8.687e+04  8.698e+03  -9.987  < 2e-16 ***
## ZIPCODE20012         -1.440e+05  1.261e+04 -11.415  < 2e-16 ***
## ZIPCODE20015          4.132e+03  1.156e+04   0.358 0.720694
## ZIPCODE20016          1.656e+05  1.089e+04  15.211  < 2e-16 ***
## ZIPCODE20017         -6.567e+04  1.131e+04  -5.809 6.32e-09 ***
```

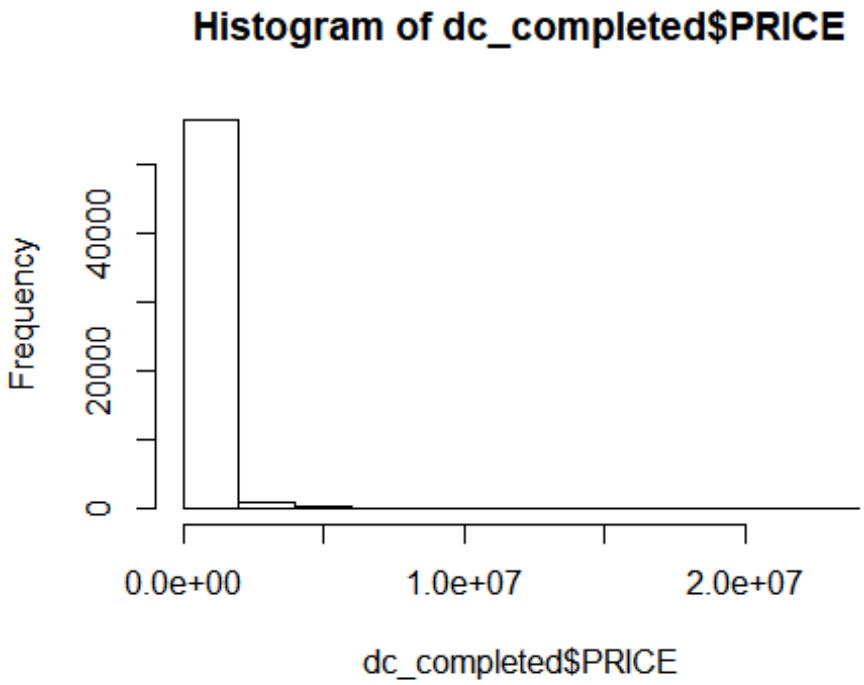
```
## ZIPCODE20018        -1.400e+05  1.110e+04 -12.605  < 2e-16 ***
## ZIPCODE20019        -1.399e+05  9.337e+03 -14.983  < 2e-16 ***
## ZIPCODE20020        -1.708e+05  9.904e+03 -17.243  < 2e-16 ***
## ZIPCODE20024         4.790e+04  2.252e+04   2.126 0.033471 *
## ZIPCODE20032        -1.992e+05  1.172e+04 -16.997  < 2e-16 ***
## ZIPCODE20036         3.023e+05  4.242e+04   7.128 1.03e-12 ***
## ZIPCODE20037         2.159e+05  2.837e+04   7.610 2.78e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 384700 on 57609 degrees of freedom
## Multiple R-squared:  0.5745, Adjusted R-squared:  0.574
## F-statistic:  1255 on 62 and 57609 DF,  p-value: < 2.2e-16
```

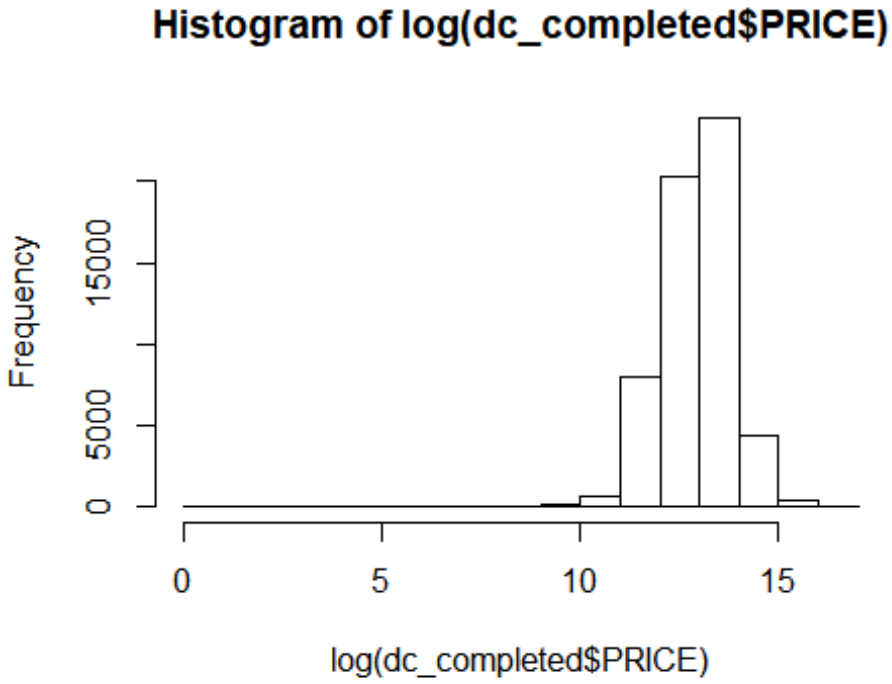
Residual Plots



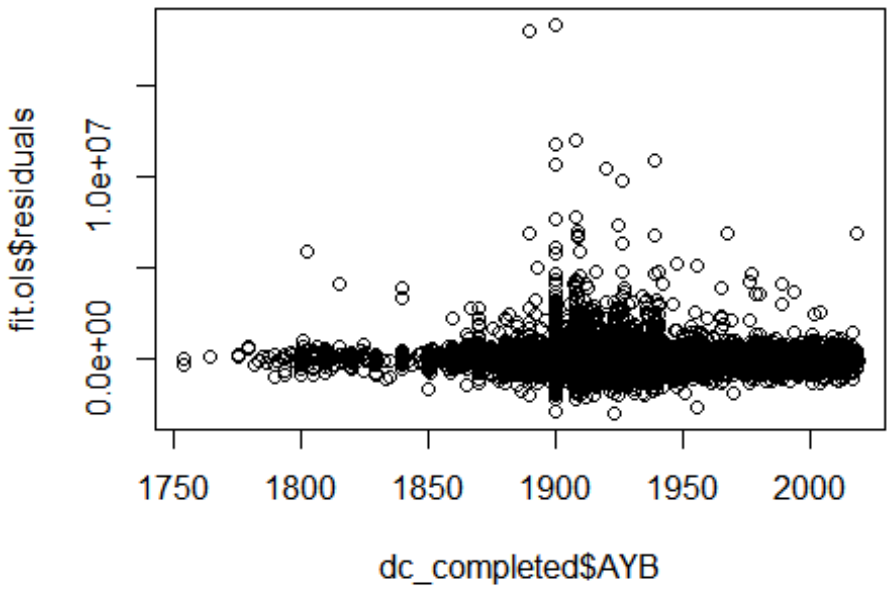


Y Distribution - Histogram





Serial Correlation




```
##
##  Durbin-Watson test
##
## data:  fit.ols
## DW = 1.7686, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

Heteroskedasticity

```
##
##  studentized Breusch-Pagan test
##
## data:  fit.ols
## BP = 3282.4, df = 62, p-value < 2.2e-16
```

Multicollinearity- Unstandarized

```
## Condition
## Index    Variance Decomposition Proportions
##                intercept BATHRM HF_BATHRM BEDRM AYB   YLRMDL STORIES
## 1        1.000 0.000     0.000  0.000     0.000 0.000 0.000  0.000
## 2        4.759 0.000     0.000  0.000     0.000 0.000 0.000  0.000
## 3      127.216 0.000     0.000  0.000     0.000 0.537 0.254  0.000
## 4      629.111 0.000     0.000  0.000     0.000 0.003 0.000  0.000
## 5     1189.642 0.000     0.000  0.000     0.000 0.000 0.000  0.000
## 6     1622.084 0.000     0.001  0.000     0.000 0.104 0.054  0.000
```

```
## 7     2066.300 0.000     0.005  0.000     0.007 0.001 0.001  0.000
## 8     2793.476 0.000     0.000  0.001     0.000 0.000 0.001  0.000
## 9     3774.154 0.000     0.115  0.001     0.563 0.000 0.000  0.000
## 10    4395.082 0.000     0.019  0.000     0.004 0.000 0.001  0.001
## 11    4598.686 0.000     0.018  0.009     0.000 0.007 0.000  0.001
## 12    5197.421 0.000     0.522  0.135     0.351 0.003 0.001  0.005
## 13    6141.400 0.000     0.217  0.754     0.064 0.009 0.005  0.010
## 14    7993.911 0.000     0.002  0.071     0.001 0.000 0.001  0.838
## 15    8420.575 0.000     0.013  0.011     0.001 0.001 0.000  0.027
## 16    9759.430 0.000     0.054  0.007     0.008 0.002 0.000  0.119
## 17  426644.302 1.000     0.033  0.011     0.000 0.331 0.680  0.001
##    QUALIFIED GBA   STRUCT CNDTN ROOF  INTWALL KITCHENS FIREPLACES USECODE
## 1  0.000     0.006 0.000  0.000 0.000 0.000   0.000    0.000      0.000
## 2  0.000     0.329 0.000  0.000 0.000 0.000   0.000    0.000      0.000
## 3  0.000     0.005 0.000  0.000 0.000 0.000   0.000    0.000      0.000
## 4  0.000     0.002 0.000  0.000 0.000 0.000   0.000    0.000      0.000
## 5  0.000     0.014 0.003  0.003 0.800 0.000   0.000    0.000      0.003
## 6  0.000     0.003 0.000  0.714 0.020 0.000   0.000    0.000      0.004
## 7  0.000     0.058 0.023  0.021 0.077 0.000   0.004    0.000      0.361
## 8  0.000     0.009 0.503  0.001 0.081 0.000   0.002    0.023      0.108
## 9  0.000     0.292 0.015  0.022 0.005 0.000   0.000    0.002      0.087
## 10 0.001     0.068 0.046  0.001 0.006 0.524   0.000    0.356      0.005
## 11 0.000     0.095 0.049  0.011 0.006 0.448   0.000    0.510      0.002
## 12 0.000     0.001 0.045  0.010 0.000 0.009   0.003    0.018      0.014
## 13 0.007     0.050 0.000  0.053 0.000 0.008   0.001    0.070      0.003
## 14 0.016     0.019 0.001  0.002 0.000 0.001   0.054    0.014      0.026
## 15 0.968     0.020 0.001  0.012 0.000 0.002   0.000    0.001      0.001
## 16 0.003     0.025 0.311  0.005 0.001 0.000   0.931    0.003      0.387
## 17 0.005     0.003 0.002  0.146 0.003 0.006   0.006    0.003      0.000
##    ZIPCODE
## 1  0.000
## 2  0.000
## 3  0.000
## 4  0.855
## 5  0.005
## 6  0.000
## 7  0.005
## 8  0.071
## 9  0.002
## 10 0.045
## 11 0.007
## 12 0.005
## 13 0.000
## 14 0.001
## 15 0.001
## 16 0.002
## 17 0.000

##     BATHRM  HF_BATHRM      BEDRM        AYB     YLRMDL    STORIES
##   2.796562   1.251070   2.031897   1.013554   1.682389   1.314558
```

```
##  QUALIFIED        GBA     STRUCT      CNDTN       ROOF    INTWALL
##   1.069179   2.989612   1.856916   1.637063   1.174135   1.023884
##   KITCHENS FIREPLACES    USECODE    ZIPCODE
##   3.140414   1.568100   1.952783   1.166259
```

Multicollinearity - Standarized

```
## Condition
## Index    Variance Decomposition Proportions
##            BATHRM HF_BATHRM BEDRM AYB   YLRMDL STORIES QUALIFIED GBA
## 1   1.000 0.021  0.007     0.023 0.000 0.008  0.019   0.001     0.020
## 2   1.226 0.001  0.034     0.002 0.000 0.013  0.001   0.012     0.000
## 3   1.507 0.000  0.001     0.007 0.005 0.109  0.001   0.048     0.009
## 4   1.699 0.004  0.001     0.007 0.036 0.013  0.023   0.015     0.000
## 5   1.852 0.001  0.020     0.003 0.576 0.001  0.016   0.067     0.001
## 6   1.878 0.001  0.057     0.001 0.298 0.000  0.082   0.104     0.003
## 7   1.987 0.001  0.000     0.000 0.011 0.025  0.001   0.562     0.001
## 8   2.046 0.008  0.191     0.000 0.000 0.020  0.113   0.123     0.001
## 9   2.141 0.016  0.391     0.019 0.021 0.002  0.157   0.043     0.006
## 10  2.235 0.010  0.037     0.014 0.006 0.001  0.198   0.002     0.010
## 11  2.324 0.013  0.183     0.110 0.039 0.009  0.274   0.010     0.013
## 12  2.582 0.001  0.002     0.111 0.006 0.005  0.042   0.001     0.005
## 13  2.937 0.007  0.000     0.022 0.000 0.686  0.000   0.001     0.000
## 14  3.256 0.264  0.008     0.668 0.002 0.047  0.000   0.000     0.119
## 15  3.868 0.617  0.063     0.010 0.000 0.057  0.027   0.009     0.741
## 16  4.183 0.034  0.004     0.002 0.000 0.004  0.043   0.000     0.073
##    STRUCT CNDTN ROOF  INTWALL KITCHENS FIREPLACES USECODE ZIPCODE
## 1  0.000  0.007 0.004 0.001   0.005    0.017      0.006   0.002
## 2  0.045  0.019 0.024 0.000   0.036    0.012      0.034   0.005
## 3  0.036  0.096 0.066 0.003   0.001    0.026      0.000   0.021
## 4  0.026  0.012 0.005 0.171   0.002    0.021      0.020   0.317
## 5  0.001  0.003 0.005 0.248   0.000    0.004      0.000   0.001
## 6  0.034  0.003 0.009 0.170   0.003    0.005      0.041   0.048
## 7  0.002  0.017 0.067 0.308   0.003    0.004      0.007   0.016
## 8  0.001  0.033 0.312 0.052   0.000    0.001      0.008   0.116
## 9  0.009  0.001 0.008 0.001   0.010    0.029      0.137   0.063
## 10 0.041  0.009 0.460 0.035   0.011    0.140      0.014   0.075
## 11 0.036  0.019 0.015 0.009   0.000    0.152      0.120   0.004
## 12 0.276  0.012 0.002 0.001   0.017    0.351      0.065   0.289
## 13 0.001  0.754 0.012 0.001   0.000    0.003      0.000   0.006
## 14 0.083  0.000 0.004 0.000   0.008    0.184      0.003   0.024
## 15 0.002  0.014 0.000 0.000   0.013    0.037      0.000   0.000
## 16 0.407  0.001 0.006 0.000   0.889    0.013      0.543   0.012

##     BATHRM  HF_BATHRM      BEDRM        AYB     YLRMDL    STORIES
##   2.796562   1.251070   2.031897   1.013554   1.682389   1.314558
##  QUALIFIED        GBA     STRUCT      CNDTN       ROOF    INTWALL
##   1.069179   2.989612   1.856916   1.637063   1.174135   1.023884
##   KITCHENS FIREPLACES    USECODE    ZIPCODE
##   3.140414   1.568100   1.952783   1.166259
```

OLS MSE

```
## [1] "mse.ols"          "139860214097.362"
```

Appendix F: OLS-Log(PRICE)

```
##
## Call:
## lm(formula = log.formula, data = dc_completed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.0695  -0.2530   0.0784   0.3208   4.1585
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.551e+00  3.090e-01  11.492  < 2e-16 ***
## BATHRM               1.114e-01  3.703e-03  30.079  < 2e-16 ***
## HF_BATHRM            6.566e-02  4.289e-03  15.308  < 2e-16 ***
## BEDRM                1.837e-02  2.979e-03   6.167 6.99e-10 ***
## AYB                 -3.019e-04  9.218e-05  -3.275 0.001056 **
## YLRMDL               4.636e-03  1.272e-04  36.439  < 2e-16 ***
## STORIES2            -5.356e-02  1.334e-02  -4.015 5.95e-05 ***
## STORIES3            -4.170e-02  1.492e-02  -2.795 0.005191 **
## STORIES4            -4.094e-02  3.474e-02  -1.179 0.238521
## STORIES9            -4.144e-02  1.378e-01  -0.301 0.763577
## QUALIFIEDU          -3.500e-01  5.878e-03 -59.542  < 2e-16 ***
## GBA                  1.033e-04  5.048e-06  20.463  < 2e-16 ***
## STRUCTRow End        2.719e-02  2.785e-02   0.977 0.328805
## STRUCTRow Inside     2.748e-02  2.740e-02   1.003 0.315843
## STRUCTSemi-Detached  5.254e-02  3.467e-02   1.515 0.129690
## STRUCTSingle         8.791e-02  3.714e-02   2.367 0.017930 *
## STRUCTTown End       1.407e-01  6.943e-02   2.026 0.042747 *
## STRUCTTown Inside    6.697e-02  4.981e-02   1.344 0.178819
## CNDTNExcellent       4.388e-01  2.043e-02  21.479  < 2e-16 ***
## CNDTNFair            1.184e-01  2.487e-02   4.761 1.93e-06 ***
## CNDTNGood            2.349e-01  6.106e-03  38.472  < 2e-16 ***
## CNDTNPoor            1.968e-01  5.792e-02   3.398 0.000678 ***
## CNDTNVery Good       4.439e-01  9.390e-03  47.274  < 2e-16 ***
## ROOFClay Tile        8.602e-03  3.225e-02   0.267 0.789687
## ROOFComp Shingle    -7.040e-02  8.111e-03  -8.679  < 2e-16 ***
## ROOFConcrete         3.283e-02  1.590e-01   0.206 0.836461
## ROOFMetal            2.665e-02  6.437e-03   4.141 3.47e-05 ***
## ROOFNeopren          1.438e-01  1.938e-02   7.421 1.18e-13 ***
## ROOFShake           -1.348e-01  2.914e-02  -4.625 3.76e-06 ***
## ROOFShingle         -7.192e-02  3.564e-02  -2.018 0.043613 *
## ROOFSlate           -2.395e-02  1.124e-02  -2.130 0.033170 *
## ROOFTypical         -6.589e-02  6.010e-02  -1.096 0.272984
## INTWALLCeramic Tile  3.227e-02  8.890e-02   0.363 0.716650
## INTWALLDefault       2.107e-01  9.172e-02   2.297 0.021645 *
```

```
## INTWALLHardwood      1.184e-01  1.285e-02   9.209  < 2e-16 ***
## INTWALLLt Concrete  -7.179e-02  7.950e-02  -0.903 0.366524
## INTWALLVinyl Sheet   1.069e+00  1.053e-01  10.145  < 2e-16 ***
## KITCHENS             3.306e-02  7.242e-03   4.565 5.00e-06 ***
## FIREPLACES           8.967e-02  3.317e-03  27.031  < 2e-16 ***
## USECODE12            1.725e-02  2.943e-02   0.586 0.557775
## USECODE13           -5.309e-02  2.608e-02  -2.035 0.041812 *
## USECODE15            1.536e-01  9.598e-02   1.600 0.109618
## USECODE23           -2.921e-01  3.058e-02  -9.552  < 2e-16 ***
## USECODE24           -2.153e-02  1.173e-02  -1.836 0.066420 .
## ZIPCODE20002        -7.165e-02  1.173e-02  -6.108 1.02e-09 ***
## ZIPCODE20003         1.153e-01  1.370e-02   8.418  < 2e-16 ***
## ZIPCODE20005         1.910e-01  4.946e-02   3.861 0.000113 ***
## ZIPCODE20007         4.968e-01  1.438e-02  34.552  < 2e-16 ***
## ZIPCODE20008         4.118e-01  1.769e-02  23.287  < 2e-16 ***
## ZIPCODE20009         2.198e-01  1.615e-02  13.613  < 2e-16 ***
## ZIPCODE20010         4.765e-02  1.548e-02   3.078 0.002082 **
## ZIPCODE20011        -9.514e-02  1.240e-02  -7.671 1.73e-14 ***
## ZIPCODE20012        -9.189e-02  1.798e-02  -5.110 3.24e-07 ***
## ZIPCODE20015         2.163e-01  1.648e-02  13.129  < 2e-16 ***
## ZIPCODE20016         3.695e-01  1.552e-02  23.806  < 2e-16 ***
## ZIPCODE20017        -9.477e-02  1.612e-02  -5.879 4.14e-09 ***
## ZIPCODE20018        -2.783e-01  1.583e-02 -17.580  < 2e-16 ***
## ZIPCODE20019        -5.818e-01  1.331e-02 -43.704  < 2e-16 ***
## ZIPCODE20020        -5.322e-01  1.412e-02 -37.693  < 2e-16 ***
## ZIPCODE20024         1.728e-02  3.211e-02   0.538 0.590414
## ZIPCODE20032        -6.614e-01  1.671e-02 -39.582  < 2e-16 ***
## ZIPCODE20036         3.770e-01  6.048e-02   6.233 4.60e-10 ***
## ZIPCODE20037         3.853e-01  4.045e-02   9.526  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5485 on 57609 degrees of freedom
## Multiple R-squared:  0.5972, Adjusted R-squared:  0.5968
## F-statistic:  1378 on 62 and 57609 DF,  p-value: < 2.2e-16
```

OLS Log MSE

```
## [1] "mse.ols.log"     "664162889918.94"
```

Appendix G: WLS

```
##
## Call:
## lm(formula = lm.formula, data = dc_completed, weights = 1/(fit.ols$residua
ls^2))
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2240 -1.0001  0.9851  0.9998  2.2528
```

```
##
## Coefficients:
##                         Estimate Std. Error   t value Pr(>|t|)
## (Intercept)         -1.275e+06  2.790e+03  -457.104  < 2e-16 ***
## BATHRM               5.831e+04  3.652e+01  1596.597  < 2e-16 ***
## HF_BATHRM            2.022e+04  4.501e+01   449.262  < 2e-16 ***
## BEDRM               -1.271e+04  3.253e+01  -390.657  < 2e-16 ***
## AYB                 -1.844e+02  6.909e-01  -266.846  < 2e-16 ***
## YLRMDL               7.803e+02  1.055e+00   739.820  < 2e-16 ***
## STORIES2            -9.775e+04  8.071e+01 -1211.085  < 2e-16 ***
## STORIES3            -1.496e+05  1.240e+02 -1206.240  < 2e-16 ***
## STORIES4            -1.269e+05  9.115e+02  -139.208  < 2e-16 ***
## STORIES9            -3.450e+04  3.592e+03    -9.605  < 2e-16 ***
## QUALIFIEDU          -9.563e+04  5.230e+01 -1828.385  < 2e-16 ***
## GBA                  2.956e+02  4.587e-02  6443.397  < 2e-16 ***
## STRUCTRow End        7.451e+04  1.068e+02   697.776  < 2e-16 ***
## STRUCTRow Inside     7.796e+04  9.163e+01   850.745  < 2e-16 ***
## STRUCTSemi-Detached  9.135e+04  4.041e+02   226.083  < 2e-16 ***
## STRUCTSingle         6.590e+04  1.525e+02   432.193  < 2e-16 ***
## STRUCTTown End       1.371e+05  1.982e+02   691.851  < 2e-16 ***
## STRUCTTown Inside    1.276e+05  1.578e+02   808.435  < 2e-16 ***
## CNDTNExcellent       4.519e+05  3.036e+02  1488.350  < 2e-16 ***
## CNDTNFair            8.101e+04  7.118e+02   113.821  < 2e-16 ***
## CNDTNGood            9.028e+04  5.508e+01  1639.198  < 2e-16 ***
## CNDTNPoor            2.073e+05  5.735e+03    36.141  < 2e-16 ***
## CNDTNVery Good       2.955e+05  7.980e+01  3702.432  < 2e-16 ***
## ROOFClay Tile       -3.996e+04  1.442e+02  -277.108  < 2e-16 ***
## ROOFComp Shingle    -1.068e+05  6.037e+01 -1769.835  < 2e-16 ***
## ROOFConcrete        -7.007e+04  1.437e+05    -0.488  0.62584
## ROOFMetal            4.056e+03  7.019e+01    57.785  < 2e-16 ***
## ROOFNeopren          5.016e+04  2.657e+02   188.796  < 2e-16 ***
## ROOFShake           -1.314e+05  3.028e+02  -433.856  < 2e-16 ***
## ROOFShingle         -7.321e+04  2.624e+03   -27.905  < 2e-16 ***
## ROOFSlate           -3.874e+04  1.513e+02  -256.091  < 2e-16 ***
## ROOFTypical         -1.047e+05  2.326e+03   -45.018  < 2e-16 ***
## INTWALLCeramic Tile  1.140e+05  2.252e+03    50.621  < 2e-16 ***
## INTWALLDefault       6.330e+04  7.062e+03     8.963  < 2e-16 ***
## INTWALLHardwood      2.483e+04  7.157e+01   347.009  < 2e-16 ***
## INTWALLLt Concrete  -1.164e+05  1.080e+04   -10.776  < 2e-16 ***
## INTWALLVinyl Sheet   7.862e+05  2.525e+05     3.114  0.00185 **
## KITCHENS            -1.231e+04  6.096e+01  -201.966  < 2e-16 ***
## FIREPLACES           8.610e+04  4.459e+01  1930.804  < 2e-16 ***
## USECODE12            2.874e+04  1.364e+02   210.658  < 2e-16 ***
## USECODE13            2.870e+04  3.962e+02    72.440  < 2e-16 ***
## USECODE15            2.212e+04  9.137e+03     2.421  0.01550 *
## USECODE23           -2.475e+05  1.856e+02 -1333.803  < 2e-16 ***
## USECODE24           -3.274e+04  1.046e+02  -313.106  < 2e-16 ***
## ZIPCODE20002        -2.317e+04  8.200e+01  -282.548  < 2e-16 ***
## ZIPCODE20003         3.605e+04  1.528e+02   235.951  < 2e-16 ***
## ZIPCODE20005         1.087e+05  4.055e+03    26.813  < 2e-16 ***
```

```
## ZIPCODE20007          3.716e+05  1.169e+02  3178.353  < 2e-16 ***
## ZIPCODE20008          2.864e+05  2.758e+02  1038.315  < 2e-16 ***
## ZIPCODE20009          8.406e+04  1.313e+02   640.109  < 2e-16 ***
## ZIPCODE20010         -4.176e+03  1.640e+02   -25.461  < 2e-16 ***
## ZIPCODE20011         -8.695e+04  1.078e+02  -806.672  < 2e-16 ***
## ZIPCODE20012         -1.440e+05  1.058e+02 -1360.941  < 2e-16 ***
## ZIPCODE20015          4.337e+03  1.732e+02    25.044  < 2e-16 ***
## ZIPCODE20016          1.658e+05  1.938e+02   855.152  < 2e-16 ***
## ZIPCODE20017         -6.575e+04  9.785e+01  -671.919  < 2e-16 ***
## ZIPCODE20018         -1.400e+05  1.209e+02 -1158.364  < 2e-16 ***
## ZIPCODE20019         -1.400e+05  9.626e+01 -1454.272  < 2e-16 ***
## ZIPCODE20020         -1.707e+05  1.016e+02 -1681.075  < 2e-16 ***
## ZIPCODE20024          4.914e+04  1.129e+03    43.540  < 2e-16 ***
## ZIPCODE20032         -1.991e+05  1.888e+02 -1054.450  < 2e-16 ***
## ZIPCODE20036          3.027e+05  2.676e+03   113.109  < 2e-16 ***
## ZIPCODE20037          2.162e+05  1.715e+02  1261.008  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9999 on 57609 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 4.218e+08 on 62 and 57609 DF,  p-value: < 2.2e-16
```

Real R-Sq

```
## [1] "Real R-Sq" "0.57417"
```

WLS MSE

```
## [1] "mse.wls"          "139678604484.187"
```

Appendix H: WLS-Log(PRICE)

```
##
## Call:
## lm(formula = log.formula, data = dc_completed, weights = 1/(fit.ols.log$re
siduals^2))
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -1.6476 -1.0002  0.9967  0.9997  1.2911
##
## Coefficients:
##                       Estimate Std. Error   t value Pr(>|t|)
## (Intercept)          3.548e+00  6.156e-03   576.403  < 2e-16 ***
## BATHRM               1.113e-01  9.499e-05  1172.226  < 2e-16 ***
## HF_BATHRM            6.555e-02  1.084e-04   604.800  < 2e-16 ***
## BEDRM                1.838e-02  6.456e-05   284.619  < 2e-16 ***
## AYB                 -2.979e-04  2.472e-06  -120.540  < 2e-16 ***
## YLRMDL               4.633e-03  2.297e-06  2017.294  < 2e-16 ***
## STORIES2            -5.345e-02  2.161e-04  -247.306  < 2e-16 ***
```

```
## STORIES3             -4.131e-02  2.621e-04  -157.605  < 2e-16 ***
## STORIES4             -3.663e-02  2.721e-03   -13.466  < 2e-16 ***
## STORIES9             -2.672e-02  1.223e-02    -2.184  0.02894 *
## QUALIFIEDU           -3.501e-01  1.900e-04 -1842.561  < 2e-16 ***
## GBA                   1.033e-04  1.404e-07   735.413  < 2e-16 ***
## STRUCTRow End         2.659e-02  6.294e-04    42.249  < 2e-16 ***
## STRUCTRow Inside      2.704e-02  6.120e-04    44.183  < 2e-16 ***
## STRUCTSemi-Detached   5.205e-02  6.278e-04    82.906  < 2e-16 ***
## STRUCTSingle          8.998e-02  1.602e-03    56.159  < 2e-16 ***
## STRUCTTown End        1.371e-01  1.565e-03    87.603  < 2e-16 ***
## STRUCTTown Inside     6.557e-02  1.054e-03    62.204  < 2e-16 ***
## CNDTNExcellent        4.384e-01  3.144e-04  1394.480  < 2e-16 ***
## CNDTNFair             1.187e-01  3.347e-04   354.585  < 2e-16 ***
## CNDTNGood             2.350e-01  1.420e-04  1655.705  < 2e-16 ***
## CNDTNPoor             2.062e-01  7.310e-03    28.203  < 2e-16 ***
## CNDTNVery Good        4.440e-01  1.839e-04  2414.429  < 2e-16 ***
## ROOFClay Tile         8.545e-03  4.104e-04    20.818  < 2e-16 ***
## ROOFComp Shingle     -7.015e-02  1.645e-04  -426.319  < 2e-16 ***
## ROOFConcrete          7.398e-02  7.995e-02     0.925  0.35478
## ROOFMetal             2.667e-02  1.474e-04   180.915  < 2e-16 ***
## ROOFNeopren           1.442e-01  3.791e-04   380.532  < 2e-16 ***
## ROOFShake            -1.346e-01  5.997e-04  -224.432  < 2e-16 ***
## ROOFShingle          -7.234e-02  6.145e-04  -117.733  < 2e-16 ***
## ROOFSlate            -2.367e-02  2.113e-04  -112.025  < 2e-16 ***
## ROOFTypical          -6.562e-02  5.080e-04  -129.160  < 2e-16 ***
## INTWALLCeramic Tile   3.661e-02  4.036e-03     9.073  < 2e-16 ***
## INTWALLDefault        2.003e-01  1.037e-02    19.315  < 2e-16 ***
## INTWALLHardwood       1.183e-01  5.108e-04   231.637  < 2e-16 ***
## INTWALLLt Concrete   -8.084e-02  2.725e-02    -2.966  0.00302 **
## INTWALLVinyl Sheet    7.762e-01  1.559e-01     4.978 6.45e-07 ***
## KITCHENS              3.304e-02  1.473e-04   224.300  < 2e-16 ***
## FIREPLACES            8.965e-02  5.844e-05  1533.903  < 2e-16 ***
## USECODE12             1.447e-02  1.511e-03     9.576  < 2e-16 ***
## USECODE13            -5.300e-02  3.309e-04  -160.161  < 2e-16 ***
## USECODE15             1.675e-01  9.993e-03    16.758  < 2e-16 ***
## USECODE23            -2.917e-01  8.133e-04  -358.643  < 2e-16 ***
## USECODE24            -2.191e-02  3.256e-04   -67.278  < 2e-16 ***
## ZIPCODE20002         -7.176e-02  2.070e-04  -346.744  < 2e-16 ***
## ZIPCODE20003          1.156e-01  2.712e-04   426.189  < 2e-16 ***
## ZIPCODE20005          2.124e-01  9.916e-03    21.420  < 2e-16 ***
## ZIPCODE20007          4.966e-01  2.413e-04  2057.471  < 2e-16 ***
## ZIPCODE20008          4.121e-01  3.744e-04  1100.669  < 2e-16 ***
## ZIPCODE20009          2.195e-01  4.784e-04   458.836  < 2e-16 ***
## ZIPCODE20010          4.740e-02  5.547e-04    85.455  < 2e-16 ***
## ZIPCODE20011         -9.527e-02  2.117e-04  -450.109  < 2e-16 ***
## ZIPCODE20012         -9.231e-02  2.878e-04  -320.788  < 2e-16 ***
## ZIPCODE20015          2.164e-01  2.799e-04   773.162  < 2e-16 ***
## ZIPCODE20016          3.692e-01  2.817e-04  1310.626  < 2e-16 ***
## ZIPCODE20017         -9.496e-02  2.443e-04  -388.775  < 2e-16 ***
## ZIPCODE20018         -2.785e-01  2.618e-04 -1063.726  < 2e-16 ***
```

```
## ZIPCODE20019        -5.817e-01  2.446e-04 -2378.131  < 2e-16 ***
## ZIPCODE20020        -5.324e-01  3.031e-04 -1756.653  < 2e-16 ***
## ZIPCODE20024         1.537e-02  2.015e-03     7.627 2.44e-14 ***
## ZIPCODE20032        -6.615e-01  3.723e-04 -1776.903  < 2e-16 ***
## ZIPCODE20036         3.717e-01  2.627e-03   141.503  < 2e-16 ***
## ZIPCODE20037         3.867e-01  9.471e-04   408.325  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9998 on 57609 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 6.068e+06 on 62 and 57609 DF,  p-value: < 2.2e-16
```

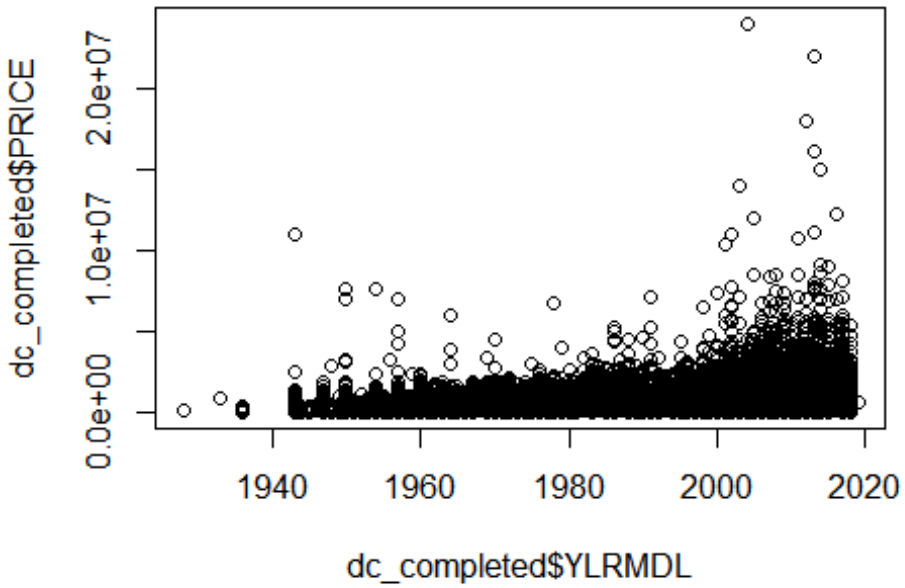
Real R-Sq

```
## [1] "Real R-Sq" "0.59719"
```

WLS Log MSE

```
## [1] "mse.wls.log"      "664162600008.756"
```

Appendix I: Polynomial



ANOVA

```
## Analysis of Variance Table
##
## Model 1: PRICE ~ BATHRM + HF_BATHRM + BEDRM + AYB + YLRMDL + STORIES +
```

```
##     QUALIFIED + GBA + STRUCT + CNDTN + ROOF + INTWALL + KITCHENS +
##     FIREPLACES + USECODE + ZIPCODE
## Model 2: PRICE ~ poly(YLRMDL, 2) + BATHRM + HF_BATHRM + BEDRM + AYB +
##     STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF + INTWALL +
##     KITCHENS + FIREPLACES + USECODE + ZIPCODE
## Model 3: PRICE ~ poly(YLRMDL, 3) + BATHRM + HF_BATHRM + BEDRM + AYB +
##     STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF + INTWALL +
##     KITCHENS + FIREPLACES + USECODE + ZIPCODE
## Model 4: PRICE ~ poly(YLRMDL, 4) + BATHRM + HF_BATHRM + BEDRM + AYB +
##     STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF + INTWALL +
##     KITCHENS + FIREPLACES + USECODE + ZIPCODE
##   Res.Df        RSS Df  Sum of Sq       F    Pr(>F)
## 1  57609 8.5270e+15
## 2  57608 8.4763e+15  1 5.0704e+13 344.686 < 2.2e-16 ***
## 3  57607 8.4740e+15  1 2.2848e+12  15.532 8.121e-05 ***
## 4  57606 8.4740e+15  1 6.4821e+06   0.000    0.9947
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Poly2

```
##
## Call:
## lm(formula = PRICE ~ poly(YLRMDL, 2) + BATHRM + HF_BATHRM + BEDRM +
##     AYB + STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF +
##     INTWALL + KITCHENS + FIREPLACES + USECODE + ZIPCODE, data = dc_complet
ed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3037354  -155240     6415   147554 18353651
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)          334766.68  126738.43   2.641 0.008259 **
## poly(YLRMDL, 2)1    4889462.24  519277.69   9.416  < 2e-16 ***
## poly(YLRMDL, 2)2    7790727.72  419678.78  18.564  < 2e-16 ***
## BATHRM                55906.57    2592.42  21.565  < 2e-16 ***
## HF_BATHRM             23495.11    3004.46   7.820 5.37e-15 ***
## BEDRM                -12699.76    2083.35  -6.096 1.10e-09 ***
## AYB                    -204.81      64.47  -3.177 0.001489 **
## STORIES2            -103747.33    9334.15 -11.115  < 2e-16 ***
## STORIES3            -153714.55   10435.66 -14.730  < 2e-16 ***
## STORIES4            -132348.21   24292.24  -5.448 5.11e-08 ***
## STORIES9             -37966.20   96331.44  -0.394 0.693493
## QUALIFIEDU           -95251.20    4110.68 -23.172  < 2e-16 ***
## GBA                    295.86       3.53  83.819  < 2e-16 ***
## STRUCTRow End         76054.04   19473.83   3.905 9.42e-05 ***
## STRUCTRow Inside      78936.60   19159.20   4.120 3.79e-05 ***
## STRUCTSemi-Detached   93833.74   24246.63   3.870 0.000109 ***
## STRUCTSingle          66937.91   25970.14   2.577 0.009954 **
```

```
## STRUCTTown End       127045.40   48552.90   2.617 0.008882 **
## STRUCTTown Inside    116623.35   34837.98   3.348 0.000816 ***
## CNDTNExcellent       420382.97   14383.85  29.226  < 2e-16 ***
## CNDTNFair             62210.14   17420.49   3.571 0.000356 ***
## CNDTNGood             86942.72    4273.50  20.345  < 2e-16 ***
## CNDTNPoor            167646.29   40607.76   4.128 3.66e-05 ***
## CNDTNVery Good       274579.30    6661.56  41.218  < 2e-16 ***
## ROOFClay Tile        -36439.60   22553.73  -1.616 0.106169
## ROOFComp Shingle    -101788.85    5678.69 -17.925  < 2e-16 ***
## ROOFConcrete         360705.31  111216.97   3.243 0.001182 **
## ROOFMetal              5936.24    4502.52   1.318 0.187366
## ROOFNeopren           45989.28   13550.23   3.394 0.000689 ***
## ROOFShake           -117572.86   20392.83  -5.765 8.19e-09 ***
## ROOFShingle          -71744.74   24925.53  -2.878 0.003999 **
## ROOFSlate            -35120.87    7863.86  -4.466 7.98e-06 ***
## ROOFTypical          -93358.48   42032.08  -2.221 0.026346 *
## INTWALLCeramic Tile  115568.75   62169.46   1.859 0.063042 .
## INTWALLDefault        48082.76   64144.73   0.750 0.453500
## INTWALLHardwood       14895.89    9005.67   1.654 0.098121 .
## INTWALLLt Concrete  -117968.86   55597.98  -2.122 0.033857 *
## INTWALLVinyl Sheet  1191080.67   73669.59  16.168  < 2e-16 ***
## KITCHENS             -11674.13    5064.48  -2.305 0.021165 *
## FIREPLACES            90108.26    2329.47  38.682  < 2e-16 ***
## USECODE12             29653.91   20577.66   1.441 0.149569
## USECODE13             29282.20   18239.33   1.605 0.108402
## USECODE15             12952.24   67121.10   0.193 0.846985
## USECODE23           -246293.73   21384.41 -11.517  < 2e-16 ***
## USECODE24            -26476.57    8210.33  -3.225 0.001261 **
## ZIPCODE20002         -23331.75    8203.78  -2.844 0.004456 **
## ZIPCODE20003          39214.10    9580.96   4.093 4.27e-05 ***
## ZIPCODE20005         110019.15   34589.92   3.181 0.001470 **
## ZIPCODE20007         381304.59   10067.61  37.874  < 2e-16 ***
## ZIPCODE20008         292426.75   12371.46  23.637  < 2e-16 ***
## ZIPCODE20009          88755.58   11293.97   7.859 3.95e-15 ***
## ZIPCODE20010         -10932.30   10830.16  -1.009 0.312772
## ZIPCODE20011        -102596.59    8713.81 -11.774  < 2e-16 ***
## ZIPCODE20012        -154627.14   12588.57 -12.283  < 2e-16 ***
## ZIPCODE20015           9184.70   11525.91   0.797 0.425527
## ZIPCODE20016         174387.48   10864.69  16.051  < 2e-16 ***
## ZIPCODE20017         -75099.70   11283.72  -6.656 2.85e-11 ***
## ZIPCODE20018        -149635.16   11082.91 -13.501  < 2e-16 ***
## ZIPCODE20019        -153307.35    9337.55 -16.418  < 2e-16 ***
## ZIPCODE20020        -179134.42    9884.62 -18.123  < 2e-16 ***
## ZIPCODE20024          52513.59   22458.47   2.338 0.019378 *
## ZIPCODE20032        -204953.49   11688.50 -17.535  < 2e-16 ***
## ZIPCODE20036         303962.25   42291.40   7.187 6.69e-13 ***
## ZIPCODE20037         224436.19   28290.10   7.933 2.17e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 383600 on 57608 degrees of freedom
## Multiple R-squared:  0.577,  Adjusted R-squared:  0.5766
## F-statistic:  1247 on 63 and 57608 DF,  p-value: < 2.2e-16
```

Poly3

```
##
## Call:
## lm(formula = PRICE ~ poly(YLRMDL, 3) + BATHRM + HF_BATHRM + BEDRM +
##     AYB + STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF +
##     INTWALL + KITCHENS + FIREPLACES + USECODE + ZIPCODE, data = dc_complet
ed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3043711  -155351     6324   147707 18353671
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          3.436e+05  1.267e+05   2.711 0.006715 **
## poly(YLRMDL, 3)1     5.025e+06  5.203e+05   9.656  < 2e-16 ***
## poly(YLRMDL, 3)2     7.873e+06  4.201e+05  18.738  < 2e-16 ***
## poly(YLRMDL, 3)3     1.566e+06  3.974e+05   3.941 8.12e-05 ***
## BATHRM               5.539e+04  2.595e+03  21.342  < 2e-16 ***
## HF_BATHRM            2.320e+04  3.005e+03   7.719 1.19e-14 ***
## BEDRM               -1.298e+04  2.084e+03  -6.229 4.73e-10 ***
## AYB                 -2.087e+02  6.447e+01  -3.237 0.001208 **
## STORIES2            -1.032e+05  9.334e+03 -11.061  < 2e-16 ***
## STORIES3            -1.531e+05  1.044e+04 -14.669  < 2e-16 ***
## STORIES4            -1.318e+05  2.429e+04  -5.427 5.76e-08 ***
## STORIES9            -3.691e+04  9.632e+04  -0.383 0.701587
## QUALIFIEDU          -9.473e+04  4.112e+03 -23.034  < 2e-16 ***
## GBA                  2.965e+02  3.533e+00  83.922  < 2e-16 ***
## STRUCTRow End        7.610e+04  1.947e+04   3.908 9.31e-05 ***
## STRUCTRow Inside     7.907e+04  1.916e+04   4.128 3.67e-05 ***
## STRUCTSemi-Detached  9.344e+04  2.424e+04   3.854 0.000116 ***
## STRUCTSingle         6.680e+04  2.597e+04   2.573 0.010098 *
## STRUCTTown End       1.265e+05  4.855e+04   2.606 0.009155 **
## STRUCTTown Inside    1.157e+05  3.483e+04   3.321 0.000899 ***
## CNDTNExcellent       4.154e+05  1.444e+04  28.772  < 2e-16 ***
## CNDTNFair            6.487e+04  1.743e+04   3.722 0.000198 ***
## CNDTNGood            8.678e+04  4.273e+03  20.309  < 2e-16 ***
## CNDTNPoor            1.814e+05  4.075e+04   4.450 8.59e-06 ***
## CNDTNVery Good       2.716e+05  6.703e+03  40.519  < 2e-16 ***
## ROOFClay Tile       -3.633e+04  2.255e+04  -1.611 0.107152
## ROOFComp Shingle    -1.013e+05  5.679e+03 -17.838  < 2e-16 ***
## ROOFConcrete         3.601e+05  1.112e+05   3.239 0.001202 **
## ROOFMetal            6.244e+03  4.503e+03   1.387 0.165519
## ROOFNeopren          4.685e+04  1.355e+04   3.457 0.000546 ***
## ROOFShake           -1.165e+05  2.039e+04  -5.711 1.13e-08 ***
## ROOFShingle         -7.051e+04  2.492e+04  -2.829 0.004671 **
```

```
## ROOFSlate           -3.517e+04  7.863e+03  -4.473 7.74e-06 ***
## ROOFTypical         -9.370e+04  4.203e+04  -2.229 0.025788 *
## INTWALLCeramic Tile  1.149e+05  6.216e+04   1.848 0.064585 .
## INTWALLDefault       4.539e+04  6.414e+04   0.708 0.479108
## INTWALLHardwood      1.334e+04  9.013e+03   1.480 0.138892
## INTWALLLt Concrete  -1.174e+05  5.559e+04  -2.112 0.034698 *
## INTWALLVinyl Sheet   1.185e+06  7.368e+04  16.082  < 2e-16 ***
## KITCHENS            -1.141e+04  5.064e+03  -2.254 0.024217 *
## FIREPLACES           9.063e+04  2.333e+03  38.848  < 2e-16 ***
## USECODE12            2.880e+04  2.058e+04   1.399 0.161673
## USECODE13            2.814e+04  1.824e+04   1.543 0.122861
## USECODE15            1.386e+04  6.711e+04   0.207 0.836375
## USECODE23           -2.457e+05  2.138e+04 -11.491  < 2e-16 ***
## USECODE24           -2.604e+04  8.210e+03  -3.171 0.001519 **
## ZIPCODE20002        -2.255e+04  8.205e+03  -2.748 0.005997 **
## ZIPCODE20003         3.928e+04  9.580e+03   4.100 4.13e-05 ***
## ZIPCODE20005         1.103e+05  3.459e+04   3.189 0.001431 **
## ZIPCODE20007         3.830e+05  1.008e+04  38.013  < 2e-16 ***
## ZIPCODE20008         2.940e+05  1.238e+04  23.752  < 2e-16 ***
## ZIPCODE20009         8.940e+04  1.129e+04   7.916 2.49e-15 ***
## ZIPCODE20010        -1.224e+04  1.083e+04  -1.130 0.258434
## ZIPCODE20011        -1.013e+05  8.719e+03 -11.617  < 2e-16 ***
## ZIPCODE20012        -1.554e+05  1.259e+04 -12.346  < 2e-16 ***
## ZIPCODE20015         1.018e+04  1.153e+04   0.883 0.377050
## ZIPCODE20016         1.760e+05  1.087e+04  16.187  < 2e-16 ***
## ZIPCODE20017        -7.520e+04  1.128e+04  -6.665 2.67e-11 ***
## ZIPCODE20018        -1.503e+05  1.108e+04 -13.558  < 2e-16 ***
## ZIPCODE20019        -1.528e+05  9.337e+03 -16.369  < 2e-16 ***
## ZIPCODE20020        -1.790e+05  9.883e+03 -18.107  < 2e-16 ***
## ZIPCODE20024         5.311e+04  2.246e+04   2.365 0.018035 *
## ZIPCODE20032        -2.045e+05  1.169e+04 -17.499  < 2e-16 ***
## ZIPCODE20036         3.051e+05  4.229e+04   7.215 5.45e-13 ***
## ZIPCODE20037         2.266e+05  2.829e+04   8.009 1.18e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 383500 on 57607 degrees of freedom
## Multiple R-squared:  0.5771, Adjusted R-squared:  0.5767
## F-statistic:  1229 on 64 and 57607 DF,  p-value: < 2.2e-16
```

Polynomial - MSE

```
##                            MSE Poly 2         MSE Poly 3
##    "mse.nonlinear" "138182148211.206"  "138193897110.64"
```

Appendix J: Interaction

```
##
## Call:
## lm(formula = PRICE ~ QUALIFIED * YLRMDL + BATHRM + HF_BATHRM +
```

```
##     BEDRM + AYB + STORIES + GBA + STRUCT + CNDTN + ROOF + INTWALL +
##     KITCHENS + FIREPLACES + USECODE + ZIPCODE, data = dc.ols_completed)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3012569  -153819     6411   146449 18397053
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          2.367e+06  3.220e+05   7.353 1.97e-13 ***
## QUALIFIED           -4.883e+06  3.181e+05 -15.352  < 2e-16 ***
## YLRMDL              -1.100e+03  1.495e+02  -7.356 1.92e-13 ***
## BATHRM               5.681e+04  2.593e+03  21.906  < 2e-16 ***
## HF_BATHRM            1.992e+04  3.002e+03   6.636 3.26e-11 ***
## BEDRM               -1.246e+04  2.085e+03  -5.977 2.29e-09 ***
## AYB                 -1.884e+02  6.451e+01  -2.921 0.003490 **
## STORIES2            -9.790e+04  9.337e+03 -10.485  < 2e-16 ***
## STORIES3            -1.491e+05  1.044e+04 -14.279  < 2e-16 ***
## STORIES4            -1.252e+05  2.431e+04  -5.151 2.61e-07 ***
## STORIES9            -3.155e+04  9.641e+04  -0.327 0.743455
## GBA                  2.970e+02  3.534e+00  84.051  < 2e-16 ***
## STRUCTRow End        7.602e+04  1.949e+04   3.901 9.61e-05 ***
## STRUCTRow Inside     7.942e+04  1.918e+04   4.142 3.45e-05 ***
## STRUCTSemi-Detached  9.163e+04  2.427e+04   3.776 0.000160 ***
## STRUCTSingle         6.972e+04  2.599e+04   2.682 0.007320 **
## STRUCTTown End       1.359e+05  4.859e+04   2.796 0.005178 **
## STRUCTTown Inside    1.276e+05  3.486e+04   3.661 0.000252 ***
## CNDTNExcellent       4.461e+05  1.430e+04  31.192  < 2e-16 ***
## CNDTNFair            6.428e+04  1.744e+04   3.686 0.000228 ***
## CNDTNGood            8.847e+04  4.275e+03  20.694  < 2e-16 ***
## CNDTNPoor            1.898e+05  4.059e+04   4.675 2.94e-06 ***
## CNDTNVery Good       2.873e+05  6.592e+03  43.581  < 2e-16 ***
## ROOFClay Tile       -3.914e+04  2.257e+04  -1.734 0.082949 .
## ROOFComp Shingle    -1.060e+05  5.677e+03 -18.677  < 2e-16 ***
## ROOFConcrete         4.084e+05  1.113e+05   3.668 0.000244 ***
## ROOFMetal            3.729e+03  4.505e+03   0.828 0.407842
## ROOFNeopren          4.777e+04  1.356e+04   3.522 0.000428 ***
## ROOFShake           -1.293e+05  2.040e+04  -6.339 2.33e-10 ***
## ROOFShingle         -7.205e+04  2.495e+04  -2.888 0.003876 **
## ROOFSlate           -3.814e+04  7.869e+03  -4.847 1.26e-06 ***
## ROOFTypical         -1.016e+05  4.207e+04  -2.414 0.015777 *
## INTWALLCeramic Tile  1.091e+05  6.223e+04   1.754 0.079419 .
## INTWALLDefault       5.442e+04  6.420e+04   0.848 0.396599
## INTWALLHardwood      1.839e+04  9.007e+03   2.042 0.041194 *
## INTWALLLt Concrete  -1.181e+05  5.565e+04  -2.122 0.033861 *
## INTWALLVinyl Sheet   1.162e+06  7.374e+04  15.754  < 2e-16 ***
## KITCHENS            -1.227e+04  5.069e+03  -2.420 0.015518 *
## FIREPLACES           8.601e+04  2.322e+03  37.044  < 2e-16 ***
## USECODE12            2.654e+04  2.060e+04   1.289 0.197517
## USECODE13            2.937e+04  1.826e+04   1.609 0.107685
```

```
## USECODE15            1.035e+04  6.718e+04   0.154 0.877607
## USECODE23           -2.440e+05  2.140e+04 -11.398  < 2e-16 ***
## USECODE24           -2.955e+04  8.213e+03  -3.598 0.000321 ***
## ZIPCODE20002        -2.292e+04  8.211e+03  -2.791 0.005254 **
## ZIPCODE20003         3.575e+04  9.588e+03   3.729 0.000193 ***
## ZIPCODE20005         1.026e+05  3.462e+04   2.963 0.003046 **
## ZIPCODE20007         3.731e+05  1.006e+04  37.076  < 2e-16 ***
## ZIPCODE20008         2.880e+05  1.238e+04  23.265  < 2e-16 ***
## ZIPCODE20009         8.265e+04  1.130e+04   7.314 2.63e-13 ***
## ZIPCODE20010        -6.949e+03  1.083e+04  -0.641 0.521262
## ZIPCODE20011        -8.609e+04  8.680e+03  -9.918  < 2e-16 ***
## ZIPCODE20012        -1.438e+05  1.259e+04 -11.422  < 2e-16 ***
## ZIPCODE20015         5.961e+03  1.153e+04   0.517 0.605229
## ZIPCODE20016         1.682e+05  1.086e+04  15.480  < 2e-16 ***
## ZIPCODE20017        -6.475e+04  1.128e+04  -5.739 9.57e-09 ***
## ZIPCODE20018        -1.400e+05  1.108e+04 -12.632  < 2e-16 ***
## ZIPCODE20019        -1.403e+05  9.318e+03 -15.056  < 2e-16 ***
## ZIPCODE20020        -1.687e+05  9.884e+03 -17.073  < 2e-16 ***
## ZIPCODE20024         4.568e+04  2.248e+04   2.032 0.042115 *
## ZIPCODE20032        -1.985e+05  1.169e+04 -16.977  < 2e-16 ***
## ZIPCODE20036         2.965e+05  4.233e+04   7.003 2.52e-12 ***
## ZIPCODE20037         2.198e+05  2.831e+04   7.764 8.36e-15 ***
## QUALIFIED:YLRMDL     2.510e+03  1.603e+02  15.654  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 383900 on 57608 degrees of freedom
## Multiple R-squared:  0.5763, Adjusted R-squared:  0.5758
## F-statistic:  1244 on 63 and 57608 DF,  p-value: < 2.2e-16
```

Interaction - MSE

```
## [1] "mse.interaction"  "138371192960.315"
```

Appendix K: Regression Tree (Normal)

```
## [1] "the optimal value of mindev is 0.001 with a mse 155723354073.411"
```

Appendix L: Bagging

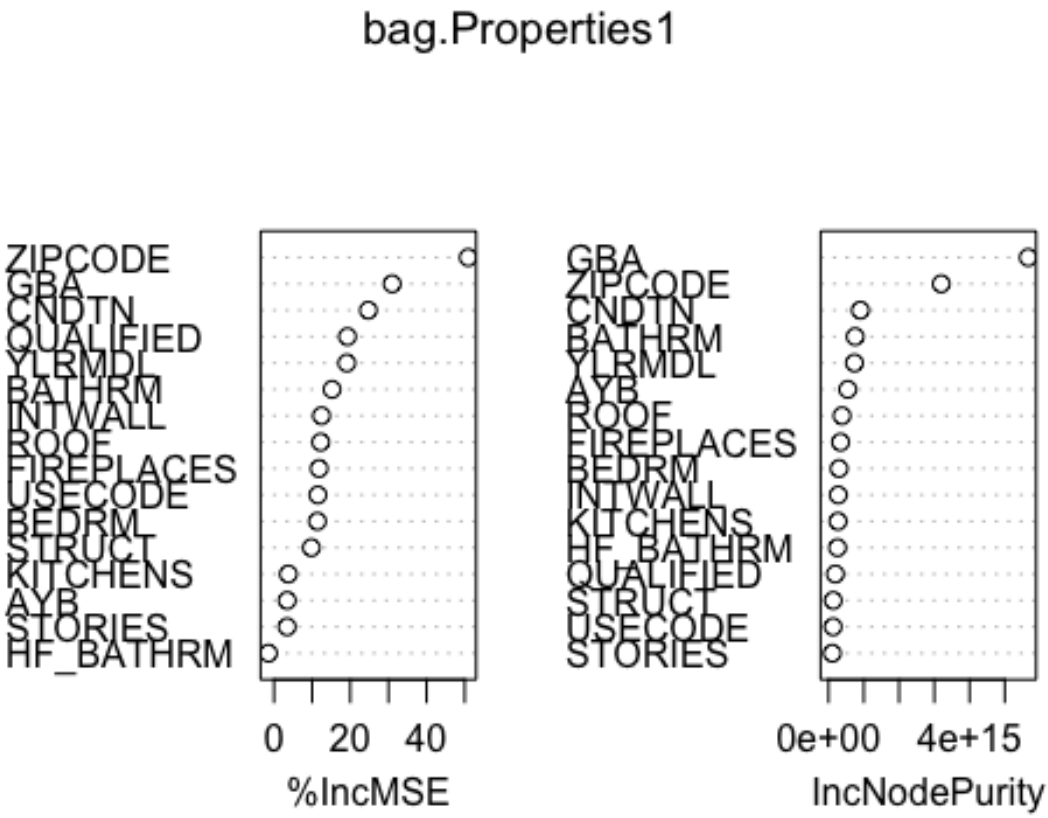
Summary

```
##
## Call:
##  randomForest(formula = PRICE ~ BATHRM + HF_BATHRM + BEDRM + AYB +      YL
RMDL + STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF +      INTWALL + KIT
CHENS + FIREPLACES + USECODE + ZIPCODE, data = DCproperties_complete[train,
], mtry = 16, ntree = 100, importance = T)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 16
##
##           Mean of squared residuals: 1.13932e+11
##                     % Var explained: 67.81
```

Errors


bag.Properties1
Error
1.2e+11
1.5e+11
1.8e+11
0
20
40
60
80
100
trees

Variable Importance Plot




```
##                %IncMSE IncNodePurity
## BATHRM      15.222499  7.563229e+14
## HF_BATHRM   -1.435953  2.556824e+14
## BEDRM       11.379668  2.995424e+14
## AYB          3.460575  5.463599e+14
## YLRMDL      19.100227  7.352670e+14
## STORIES      3.436576  1.070027e+14
## QUALIFIED   19.256921  1.935951e+14
## GBA         30.981056  5.664113e+15
## STRUCT       9.804628  1.335163e+14
## CNDTN       24.726642  9.027483e+14
## ROOF        12.240554  3.820945e+14
## INTWALL     12.457460  2.762263e+14
## KITCHENS     3.748280  2.676762e+14
## FIREPLACES  11.831467  3.413634e+14
## USECODE     11.506013  1.290894e+14
## ZIPCODE     50.912539  3.197357e+15
```

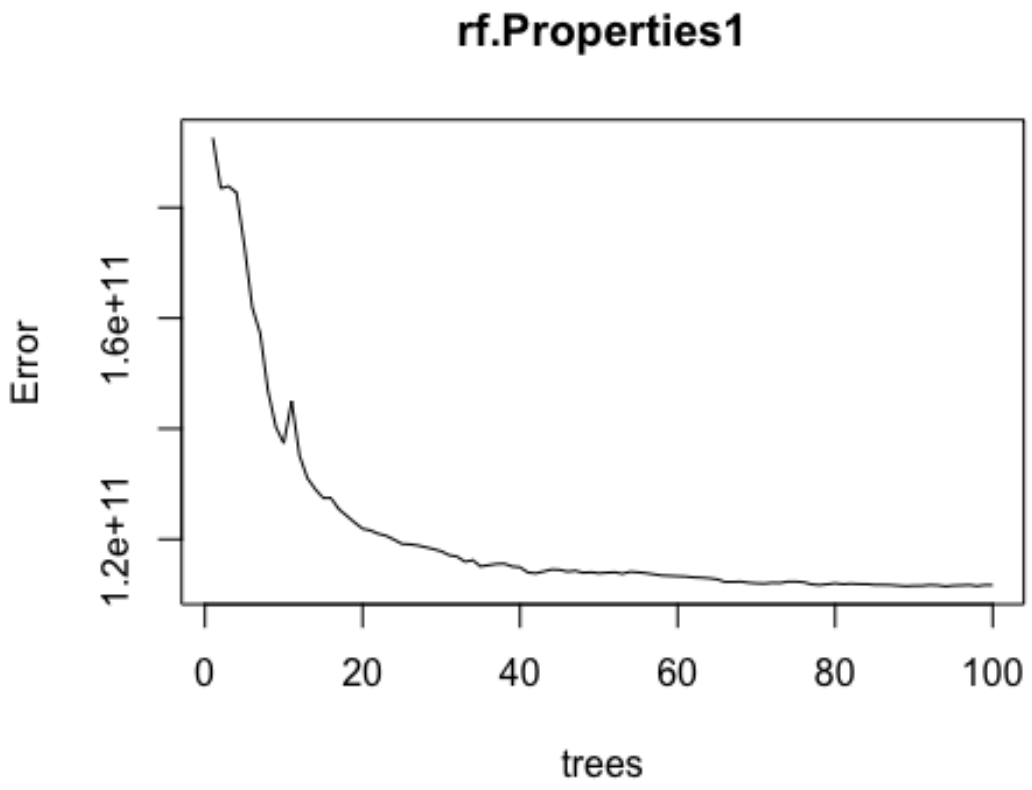
```
## [1] "mse.bagging" 113971066657
```

Appendix M: Random Forest

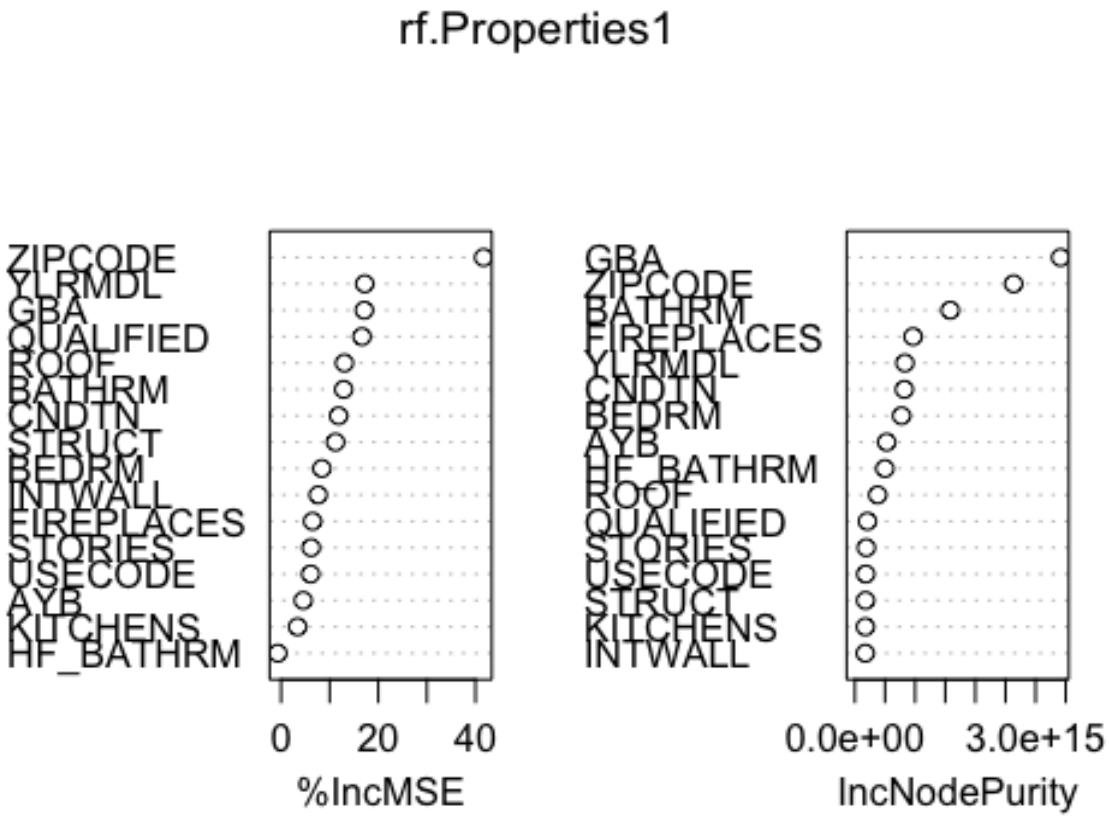
Summary

```
##
## Call:
##  randomForest(formula = PRICE ~ BATHRM + HF_BATHRM + BEDRM + AYB +      YL
RMDL + STORIES + QUALIFIED + GBA + STRUCT + CNDTN + ROOF +      INTWALL + KIT
CHENS + FIREPLACES + USECODE + ZIPCODE, data = DCproperties_complete[train,
], mtry = 5, ntree = 100, importance = T)
##                Type of random forest: regression
##                      Number of trees: 100
## No. of variables tried at each split: 5
##
##           Mean of squared residuals: 111712410065
##                     % Var explained: 68.44
```

Errors



Random Forest - Variable Importance Plot




```
##                %IncMSE IncNodePurity
## BATHRM      12.8808674  1.582576e+15
## HF_BATHRM   -0.6509726  4.993823e+14
## BEDRM        8.3983716  7.758771e+14
## AYB          4.5631896  5.291825e+14
## YLRMDL      17.2126655  8.301885e+14
## STORIES      6.1914950  1.896017e+14
## QUALIFIED   16.6521042  2.089199e+14
## GBA         17.1325366  3.414919e+15
## STRUCT      11.2039826  1.777269e+14
## CNDTN       11.8387884  8.206588e+14
## ROOF        13.0358556  3.783277e+14
## INTWALL      7.6335509  1.724120e+14
## KITCHENS     3.4094745  1.752854e+14
## FIREPLACES   6.4927379  9.662453e+14
## USECODE      6.0870371  1.807949e+14
## ZIPCODE     41.6648785  2.635157e+15
```

```
## [1] "mse.randomforest" 106664599063
```

Appendix O: Reference

Park. M. (2018, August 28). The D.C.-area real estate market is slowing. Here's what you need to know. The Washington post. Retrieved from: https://www.washingtonpost.com/business/2018/08/28/dc-area-real-estate-market-is-slowing-heres-what-you-need-know/?noredirect=on&utm_term=.18e6912bb1fd

Correa, C. (2018, July 31). D.C. Residential Properties. Retrieved May 2, 2019, from https://www.kaggle.com/christophercorrea/dc-residential-properties