Final Project

Zane Brotherton, Nate Adam, Feolu Kolawole

2023-08-14

Project Description

```
set.seed(123)
House_Data = read.table(url('https://raw.githubusercontent.com/zbrotherton2158/okadaStats/master/ames20
```

The study

questions to be answered are the following: • What field does the data come from? This data comes from the assessors office of Ames, Iowa • What are the goals of the study? Are there any effects of particular interest? The purpose of this data is to be able to predict the value of a house in Ames Iowa Our areas of interest are the effect of the neighborhoods, lot area, total rooms above grade, year built, and total square feet on the price of the house. • How might these goals be answered, i.e. tests / confidence intervals? This goal could be answered by a prediction model that, given information about a house, will output a 95% confidence interval predicting the price of the house.

The data

Predictor Variable Spread

```
require(tidyverse)
```

```
## Loading required package: tidyverse

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
Numerical_Count = length(select_if(House_Data, is.numeric))
Categorical_Count = length(select_if(House_Data, negate(is.numeric)))
```

Clean "none" values

```
# NOTE: Might have to revamp
# Turned Nones into 0's for respective rows

# j is the column of each selected box
for (j in 1:80){
```

```r
  # clear the haveNumbers in case it has values
  haveNumbers = c()
  # i is the row of each selected box
  for (i in 1:2000) {
    # if the selected box is empty, fill it with "None"
    if (House_Data[i,j] == "") {
      House_Data[i,j] = "None"
    }
    # record whether the selected box contained a number in a list called haveNumbers
    haveNumbers = c(haveNumbers, !grepl("\\D", House_Data[i,j]))
  }
  # if any of the selected boxes in the column contain a number,
  # replace all "None"s in the column with 0s
  if (any(haveNumbers) && (j != 59)) {
    House_Data[j] = replace(House_Data[j], House_Data[j] == "None", 0)
  }
}
```

Correlation Info

Numerical Correlation

```r
require(corrplot)
```

```
## Loading required package: corrplot
```

```
## corrplot 0.92 loaded
```

```r
require(dplyr)
```

Numerical_Correlations = cor(House_Data_Numerical, House_Data_Numerical$SalePrice)

Analysis

- How are the predictor variables spread out? Are there any noteworthy features to their spread that could be highly influential observations?

The variables are seemingly random between discrete, ordinal, nominal, and continuous types of variables. Of these, we found 26 to be numerical and 54 categorical. Within the discrete variables, variables with a spread like fireplaces where despite being numerical only have values of 1 - arbitrarily small number may have a higher effect on the price per unit increase.

- Are any of the predictor variables highly correlated?

The highest correlation numerical variables are:

- Overall Quality (~0.805)
- Gr. Living Area (~0.720)
- Year Built (~0.57)
- Full Bath (~0.561)
- Year Remodeled (~0.535)
- Rooms Above Ground (~0.504)
- Fireplace (~0.48)
- 1st Floor Square Ft. (~0.619)

- 2nd Floor Square Ft. (~0.295)*

The most statistically significant categorical variables are:

The models

- Which predictor variables, if any, should be included in the model a priority?

The highest correlation numerical variables and most statistically significant categorical listed under "The Data" are of the most interest and are the most prioritized in the model. (Neighborhoods, lot area, total rooms above grade, year built, and total square feet etc.)

- Are there any interactions that should be considered for inclusion in the model?
- Are there any three way interactions that should be considered?
- Are there any interactions that should NOT be considered?

Model Draft

```r
#factorize these?
model = lm(SalePrice ~
             #Numerical
             Overall.Qual +
             Gr.Liv.Area +
             Year.Built +
             Year.Remod.Add +
             Full.Bath +
             #Categorical
             Neighborhood +
             Condition.1 +
             Condition.2 +
             Bldg.Type +
             House.Style +
             Roof.Matl +
             Exter.Qual +
             BsmtFin.Type.1 +
             Kitchen.Qual +
             Fireplace.Qu +
             Garage.Type +
             Pool.QC
           , House_Data)
summary(model)
```

```
##
## Call:
## lm(formula = SalePrice ~ Overall.Qual + Gr.Liv.Area + Year.Built +
##     Year.Remod.Add + Full.Bath + Neighborhood + Condition.1 +
##     Condition.2 + Bldg.Type + House.Style + Roof.Matl + Exter.Qual +
##     BsmtFin.Type.1 + Kitchen.Qual + Fireplace.Qu + Garage.Type +
##     Pool.QC, data = House_Data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -175321  -11880     247   11314  168763
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
```

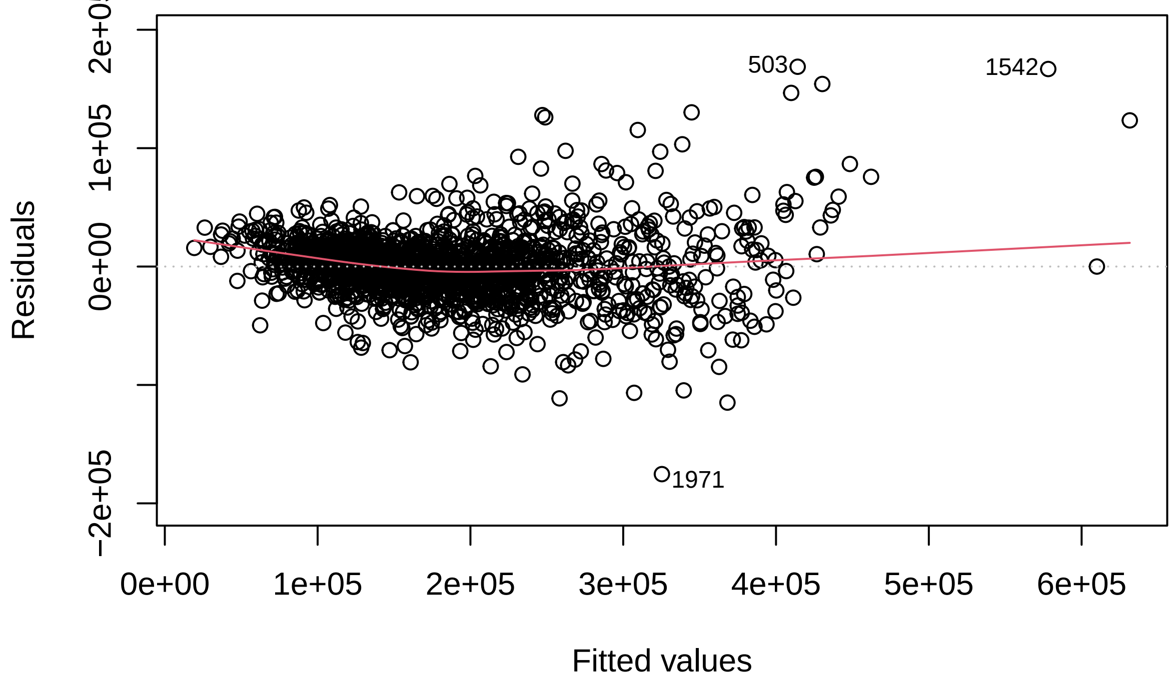
```
## (Intercept)         -1.563e+06  1.335e+05 -11.710  < 2e-16 ***
## Overall.Qual         1.112e+04  8.454e+02  13.157  < 2e-16 ***
## Gr.Liv.Area          7.224e+01  2.535e+00  28.496  < 2e-16 ***
## Year.Built           3.917e+02  5.012e+01   7.815 9.00e-15 ***
## Year.Remod.Add       2.151e+02  4.376e+01   4.914 9.67e-07 ***
## Full.Bath           -3.675e+02  1.710e+03  -0.215 0.829912
## NeighborhoodBlueste  1.828e+04  1.189e+04   1.537 0.124461
## NeighborhoodBrDale   1.143e+04  1.115e+04   1.025 0.305435
## NeighborhoodBrkSide  3.614e+03  8.636e+03   0.418 0.675639
## NeighborhoodClearCr  8.601e+03  8.941e+03   0.962 0.336186
## NeighborhoodCollgCr -1.040e+03  7.565e+03  -0.137 0.890723
## NeighborhoodCrawfor  2.429e+04  8.226e+03   2.953 0.003186 **
## NeighborhoodEdwards -6.826e+03  8.023e+03  -0.851 0.395019
## NeighborhoodGilbert -5.213e+03  7.926e+03  -0.658 0.510865
## NeighborhoodGreens   3.069e+04  1.390e+04   2.208 0.027372 *
## NeighborhoodGrnHill  1.349e+05  2.675e+04   5.043 5.01e-07 ***
## NeighborhoodIDOTRR  -4.556e+03  8.758e+03  -0.520 0.602974
## NeighborhoodMeadowV  4.262e+03  9.213e+03   0.463 0.643677
## NeighborhoodMitchel -1.726e+02  8.160e+03  -0.021 0.983122
## NeighborhoodNonemes -4.933e+03  7.817e+03  -0.631 0.528044
## NeighborhoodNoRidge  4.258e+04  8.275e+03   5.145 2.94e-07 ***
## NeighborhoodNPkVill  7.279e+03  1.035e+04   0.703 0.481867
## NeighborhoodNridgHt  2.902e+04  7.732e+03   3.754 0.000180 ***
## NeighborhoodNWAmes  -7.734e+03  8.056e+03  -0.960 0.337142
## NeighborhoodOldTown -4.345e+03  8.381e+03  -0.518 0.604218
## NeighborhoodSawyer  -2.401e+03  8.111e+03  -0.296 0.767243
## NeighborhoodSawyerW -5.168e+03  7.917e+03  -0.653 0.514036
## NeighborhoodSomerst  1.342e+04  7.525e+03   1.783 0.074778 .
## NeighborhoodStoneBr  4.447e+04  8.291e+03   5.363 9.16e-08 ***
## NeighborhoodSWISU   -6.571e+03  9.406e+03  -0.699 0.484919
## NeighborhoodTimber   1.022e+04  8.186e+03   1.248 0.212157
## NeighborhoodVeenker  7.593e+03  9.925e+03   0.765 0.444311
## Condition.1Feedr     5.204e+03  4.385e+03   1.187 0.235511
## Condition.1Norm      1.155e+04  3.584e+03   3.223 0.001290 **
## Condition.1PosA      1.742e+04  8.351e+03   2.086 0.037107 *
## Condition.1PosN      1.337e+04  6.768e+03   1.976 0.048323 *
## Condition.1RRAe     -5.366e+03  7.140e+03  -0.752 0.452413
## Condition.1RRAn      1.182e+04  6.178e+03   1.914 0.055823 .
## Condition.1RRNe      4.854e+03  1.226e+04   0.396 0.692291
## Condition.1RRNn      4.248e+03  1.008e+04   0.422 0.673398
## Condition.2Feedr     1.217e+04  1.775e+04   0.686 0.493015
## Condition.2Norm      1.639e+04  1.544e+04   1.061 0.288823
## Condition.2PosA      1.678e+05  3.118e+04   5.382 8.29e-08 ***
## Condition.2PosN      1.557e+04  2.246e+04   0.693 0.488230
## Condition.2RRAn      1.492e+04  3.055e+04   0.489 0.625222
## Condition.2RRNn      3.422e+04  2.412e+04   1.418 0.156228
## Bldg.Type2fmCon     -6.220e+03  4.261e+03  -1.460 0.144512
## Bldg.TypeDuplex     -2.404e+04  3.698e+03  -6.502 1.01e-10 ***
## Bldg.TypeTwnhs      -3.702e+04  4.662e+03  -7.942 3.37e-15 ***
## Bldg.TypeTwnhsE     -2.537e+04  2.883e+03  -8.800  < 2e-16 ***
## House.Style1.5Unf    1.735e+04  7.340e+03   2.364 0.018184 *
## House.Style1Story    1.516e+04  2.452e+03   6.184 7.61e-10 ***
## House.Style2.5Fin   -3.887e+04  1.617e+04  -2.404 0.016317 *
## House.Style2.5Unf   -1.400e+04  7.326e+03  -1.911 0.056180 .
```

```
## House.Style2Story   -5.745e+03  2.470e+03  -2.326 0.020144 *
## House.StyleSFoyer    2.236e+04  4.353e+03   5.136 3.09e-07 ***
## House.StyleSLvl      1.872e+03  3.841e+03   0.488 0.625944
## Roof.MatlCompShg     4.675e+05  3.163e+04  14.780  < 2e-16 ***
## Roof.MatlMembran     5.216e+05  4.123e+04  12.651  < 2e-16 ***
## Roof.MatlMetal       5.042e+05  4.166e+04  12.104  < 2e-16 ***
## Roof.MatlTar&Grv     4.679e+05  3.180e+04  14.714  < 2e-16 ***
## Roof.MatlWdShake     4.536e+05  3.312e+04  13.697  < 2e-16 ***
## Roof.MatlWdShngl     5.400e+05  3.346e+04  16.138  < 2e-16 ***
## Exter.QualFa        -3.895e+04  7.725e+03  -5.042 5.05e-07 ***
## Exter.QualGd        -3.860e+04  4.236e+03  -9.113  < 2e-16 ***
## Exter.QualTA        -4.044e+04  4.727e+03  -8.556  < 2e-16 ***
## BsmtFin.Type.1BLQ   -2.936e+03  2.483e+03  -1.182 0.237187
## BsmtFin.Type.1GLQ    3.455e+03  2.198e+03   1.572 0.116114
## BsmtFin.Type.1LwQ   -9.276e+03  3.093e+03  -2.999 0.002742 **
## BsmtFin.Type.1None  -1.951e+04  4.268e+03  -4.570 5.19e-06 ***
## BsmtFin.Type.1Rec   -8.032e+03  2.495e+03  -3.220 0.001305 **
## BsmtFin.Type.1Unf   -1.395e+04  2.086e+03  -6.688 2.96e-11 ***
## Kitchen.QualFa      -3.949e+04  5.175e+03  -7.631 3.65e-14 ***
## Kitchen.QualGd      -3.662e+04  3.096e+03 -11.828  < 2e-16 ***
## Kitchen.QualTA      -3.797e+04  3.460e+03 -10.974  < 2e-16 ***
## Fireplace.QuFa      -5.531e+03  6.758e+03  -0.818 0.413178
## Fireplace.QuGd      -3.105e+03  5.555e+03  -0.559 0.576219
## Fireplace.QuNone    -9.656e+03  5.722e+03  -1.688 0.091642 .
## Fireplace.QuPo      -6.238e+03  7.274e+03  -0.858 0.391209
## Fireplace.QuTA      -7.968e+03  5.671e+03  -1.405 0.160124
## Garage.TypeAttchd    7.678e+03  6.923e+03   1.109 0.267526
## Garage.TypeBasment   1.014e+03  8.647e+03   0.117 0.906682
## Garage.TypeBuiltIn   7.441e+03  7.417e+03   1.003 0.315854
## Garage.TypeCarPort   5.188e+03  1.010e+04   0.514 0.607637
## Garage.TypeDetchd    9.093e+03  6.977e+03   1.303 0.192633
## Garage.TypeNone      2.550e+03  7.335e+03   0.348 0.728089
## Pool.QCFa           -5.726e+04  2.421e+04  -2.365 0.018134 *
## Pool.QCGd           -3.024e+04  2.188e+04  -1.382 0.167134
## Pool.QCNone         -5.314e+04  1.547e+04  -3.435 0.000605 ***
## Pool.QCTA           -3.119e+04  2.419e+04  -1.289 0.197416
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25770 on 1910 degrees of freedom
## Multiple R-squared:  0.8969, Adjusted R-squared:  0.8921
## F-statistic: 186.8 on 89 and 1910 DF,  p-value: < 2.2e-16
```

```r
plot(model)
```

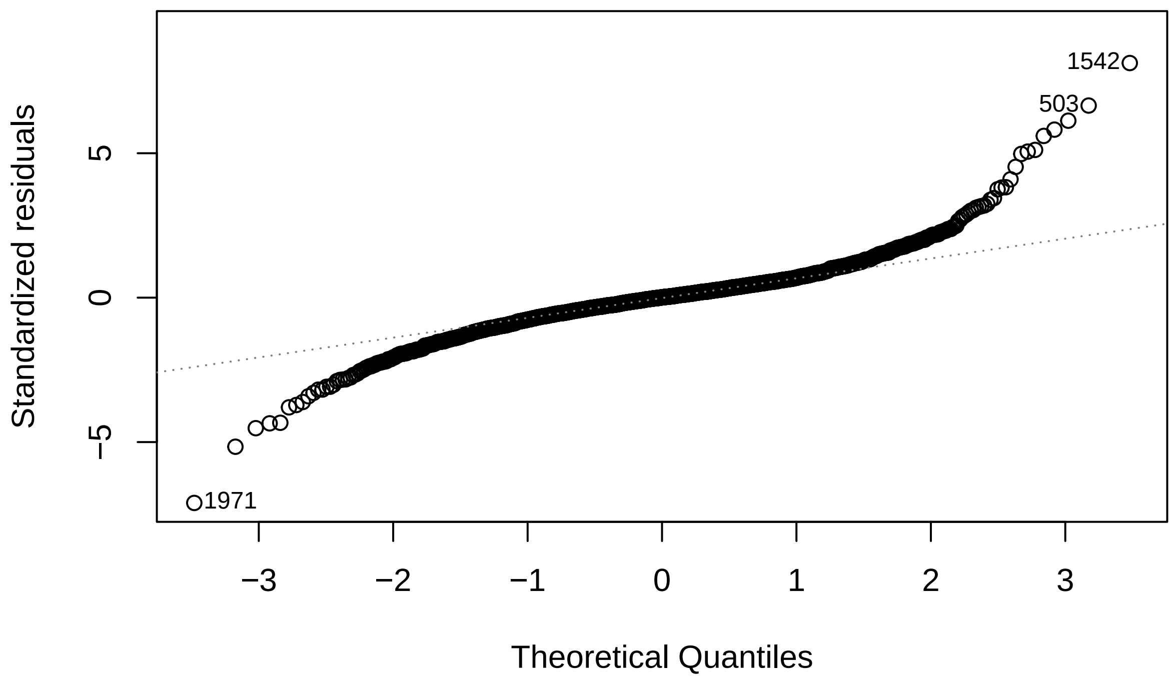
```
## Warning: not plotting observations with leverage one:
##   142, 664, 1234, 1302, 1546, 1729
```

Residuals vs Fitted

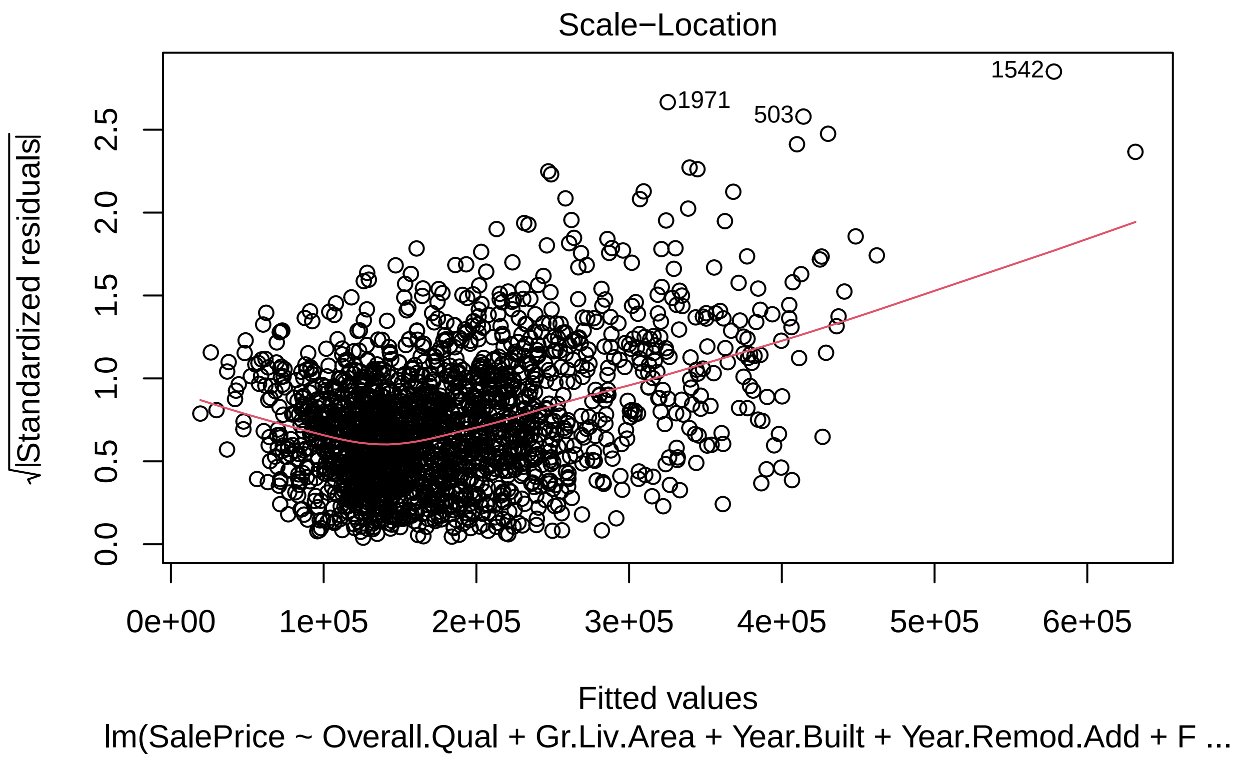


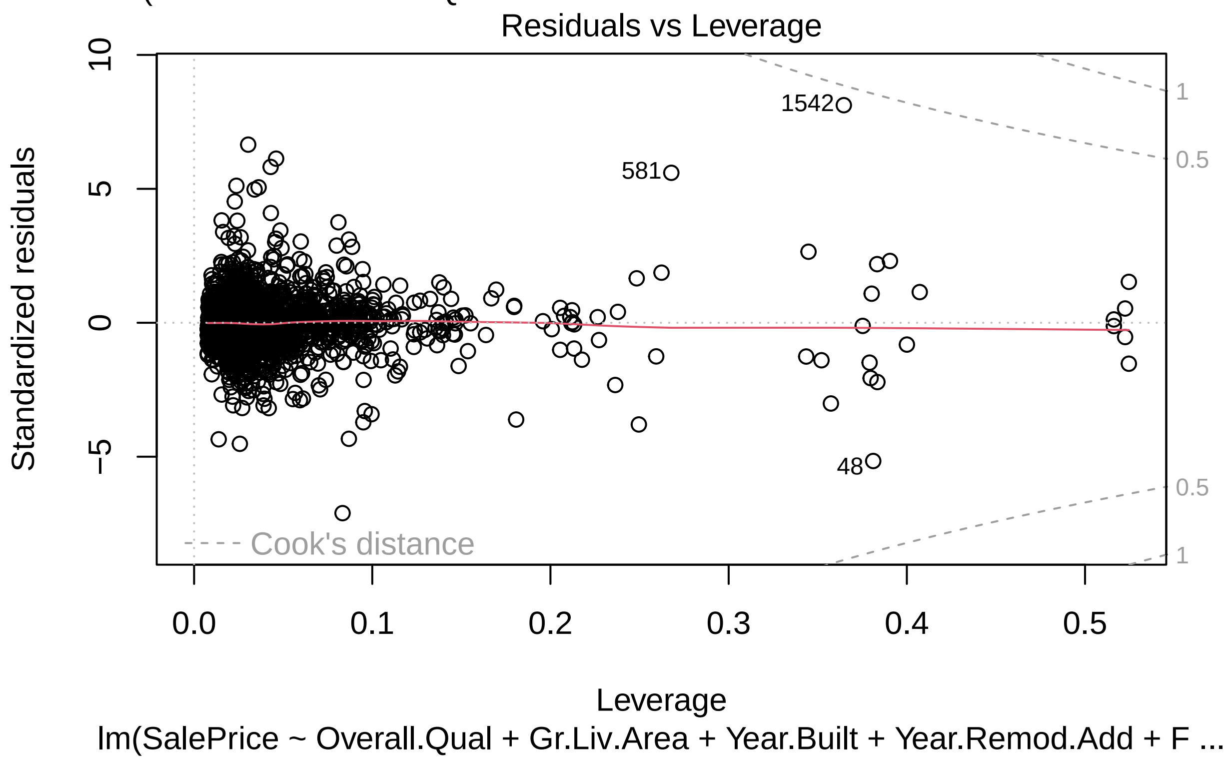
lm(SalePrice ~ Overall.Qual + Gr.Liv.Area + Year.Built + Year.Remod.Add + F ...

Q−Q Residuals



lm(SalePrice ~ Overall.Qual + Gr.Liv.Area + Year.Built + Year.Remod.Add + F ...



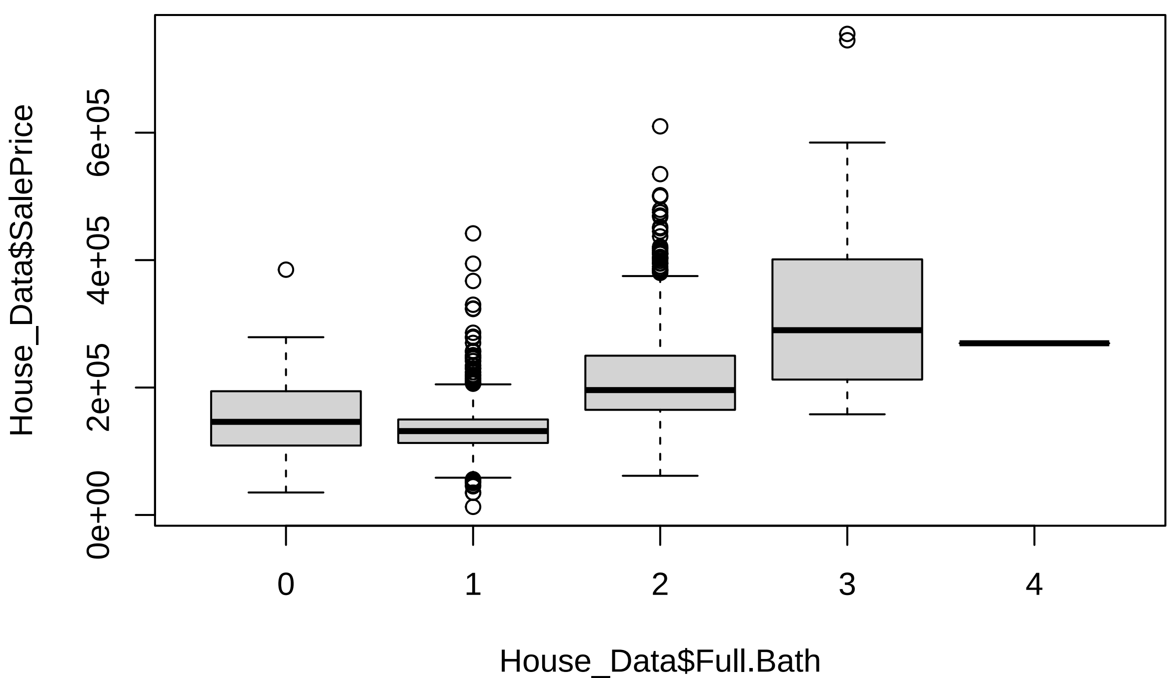



```r
#to investigate numerical variables
House_Data_Numerical = data.frame(select_if(House_Data, is.numeric))

print(cor(House_Data_Numerical, House_Data_Numerical$SalePrice))
```

```
##                        [,1]
## MS.SubClass      -0.07520213
```

```
## Lot.Area         0.24399035
## Overall.Qual      0.80454673
## Overall.Cond     -0.12019006
## Year.Built        0.57171649
## Year.Remod.Add    0.53496165
## X1st.Flr.SF       0.61912063
## X2nd.Flr.SF       0.29593043
## Low.Qual.Fin.SF -0.05105615
## Gr.Liv.Area       0.72092212
## Full.Bath         0.56122804
## Half.Bath         0.28746758
## Bedroom.AbvGr     0.15392465
## Kitchen.AbvGr    -0.12968211
## TotRms.AbvGrd     0.50427974
## Fireplaces        0.48079444
## Wood.Deck.SF      0.32936851
## Open.Porch.SF     0.31620588
## Enclosed.Porch   -0.12835243
## X3Ssn.Porch       0.03014930
## Screen.Porch      0.09847254
## Pool.Area         0.08364937
## Misc.Val         -0.01175391
## Mo.Sold           0.03268045
## Yr.Sold          -0.02830382
## SalePrice         1.00000000
```

```r
#to investigate categories
library(tidyverse)
House_Data_Categorical = data.frame(select_if(House_Data, negate(is.numeric)))
House_Data_Categorical$SalePrice = House_Data$SalePrice

boxplot(House_Data$SalePrice ~ House_Data$Full.Bath)
```



```r
sapply(lapply(House_Data_Categorical, unique), length)
```

```
##      MS.Zoning   Lot.Frontage         Street          Alley      Lot.Shape
##              7            121              2              3              4
##   Land.Contour      Utilities     Lot.Config     Land.Slope   Neighborhood
##              4              1              5              3             27
##    Condition.1    Condition.2      Bldg.Type    House.Style     Roof.Style
##              9              7              5              8              6
##      Roof.Matl   Exterior.1st   Exterior.2nd   Mas.Vnr.Type   Mas.Vnr.Area
##              7             14             16              5            362
##     Exter.Qual     Exter.Cond     Foundation      Bsmt.Qual      Bsmt.Cond
##              4              4              6              6              6
##  Bsmt.Exposure BsmtFin.Type.1   BsmtFin.SF.1 BsmtFin.Type.2   BsmtFin.SF.2
##              5              7            816              7            211
##    Bsmt.Unf.SF  Total.Bsmt.SF        Heating     Heating.QC    Central.Air
##            927            870              6              5              2
##     Electrical Bsmt.Full.Bath Bsmt.Half.Bath   Kitchen.Qual     Functional
##              4              4              3              4              7
##   Fireplace.Qu    Garage.Type  Garage.Yr.Blt  Garage.Finish    Garage.Cars
##              6              7            101              4              6
##    Garage.Area    Garage.Qual    Garage.Cond    Paved.Drive        Pool.QC
##            530              6              6              3              5
##          Fence   Misc.Feature      Sale.Type Sale.Condition      SalePrice
##              5              5             10              6            812
```

```r
investigative_lm = lm(SalePrice ~ ., data = subset(House_Data_Categorical,
select = -c(Utilities,Lot.Frontage,Mas.Vnr.Area,BsmtFin.SF.1,BsmtFin.SF.2,Bsmt.Unf.SF,Total.Bsmt.SF,Gara
summary(investigative_lm)
```

```
##
## Call:
## lm(formula = SalePrice ~ ., data = subset(House_Data_Categorical,
##     select = -c(Utilities, Lot.Frontage, Mas.Vnr.Area, BsmtFin.SF.1,
##         BsmtFin.SF.2, Bsmt.Unf.SF, Total.Bsmt.SF, Garage.Yr.Blt,
##         Garage.Area)))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -144541  -14322    -562   12866  194373
##
## Coefficients: (4 not defined because of singularities)
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           102212.56  101049.52   1.012 0.311911
## MS.ZoningC (all)      -70026.08   53994.08  -1.297 0.194828
## MS.ZoningFV           -38154.37   52605.88  -0.725 0.468372
## MS.ZoningI (all)      -61524.07   62500.84  -0.984 0.325067
## MS.ZoningRH           -17745.71   53589.37  -0.331 0.740576
## MS.ZoningRL           -32848.29   52297.62  -0.628 0.530018
## MS.ZoningRM           -39135.55   52732.20  -0.742 0.458091
## StreetPave             30577.58   12334.38   2.479 0.013266 *
## AlleyNone               2242.23    4129.22   0.543 0.587188
## AlleyPave               9613.18    6480.05   1.484 0.138119
## Lot.ShapeIR2            7184.25    4428.17   1.622 0.104897
## Lot.ShapeIR3           26411.54    8961.83   2.947 0.003250 **
```

```
## Lot.ShapeReg           -1532.60    1701.81  -0.901 0.367938    
## Land.ContourHLS        -2867.03    5439.59  -0.527 0.598213    
## Land.ContourLow        -4397.28    7080.96  -0.621 0.534680    
## Land.ContourLvl         3408.24    4013.35   0.849 0.395870    
## Lot.ConfigCulDSac       4155.25    3394.35   1.224 0.221053    
## Lot.ConfigFR2         -12429.56    4402.24  -2.823 0.004804 ** 
## Lot.ConfigFR3           1975.85    8475.34   0.233 0.815688    
## Lot.ConfigInside        -742.31    1886.27  -0.394 0.693972    
## Land.SlopeMod           9290.98    4249.42   2.186 0.028916 *  
## Land.SlopeSev           5658.60   10604.63   0.534 0.593687    
## NeighborhoodBlueste     2788.41   14594.74   0.191 0.848504    
## NeighborhoodBrDale     17977.72   13610.67   1.321 0.186721    
## NeighborhoodBrkSide     3185.40   10724.36   0.297 0.766482    
## NeighborhoodClearCr    23704.93   10970.60   2.161 0.030847 *  
## NeighborhoodCollgCr     6020.35    8966.76   0.671 0.502049    
## NeighborhoodCrawfor    37284.29    9872.93   3.776 0.000164 ***
## NeighborhoodEdwards    -5932.40    9529.90  -0.623 0.533691    
## NeighborhoodGilbert    -5951.83    9427.24  -0.631 0.527898    
## NeighborhoodGreens     25179.35   16724.32   1.506 0.132361    
## NeighborhoodGrnHill   178580.75   30844.06   5.790 8.32e-09 ***
## NeighborhoodIDOTRR      8013.96   11418.60   0.702 0.482876    
## NeighborhoodMeadowV   -17518.38   12082.35  -1.450 0.147260    
## NeighborhoodMitchel     -147.48    9818.11  -0.015 0.988017    
## NeighborhoodNonemes    -3076.94    9421.15  -0.327 0.744010    
## NeighborhoodNoRidge    68087.42    9706.60   7.015 3.27e-12 ***
## NeighborhoodNPkVill    35445.77   19543.50   1.814 0.069896 .  
## NeighborhoodNridgHt    33772.89    9162.80   3.686 0.000235 ***
## NeighborhoodNWAmes     12253.53    9743.10   1.258 0.208680    
## NeighborhoodOldTown     -586.23   10656.03  -0.055 0.956133    
## NeighborhoodSawyer     -3339.80    9711.03  -0.344 0.730948    
## NeighborhoodSawyerW     5637.89    9465.84   0.596 0.551516    
## NeighborhoodSomerst    30007.56   10285.40   2.917 0.003573 ** 
## NeighborhoodStoneBr    60176.28   10012.79   6.010 2.25e-09 ***
## NeighborhoodSWISU      -2593.34   11225.24  -0.231 0.817320    
## NeighborhoodTimber     11275.52    9620.83   1.172 0.241359    
## NeighborhoodVeenker    27901.75   12226.09   2.282 0.022599 *  
## Condition.1Feedr        1482.41    5211.84   0.284 0.776113    
## Condition.1Norm        13234.59    4268.58   3.100 0.001963 ** 
## Condition.1PosA        33193.38    9634.87   3.445 0.000584 ***
## Condition.1PosN        16952.90    7916.63   2.141 0.032376 *  
## Condition.1RRAe         2961.98    8404.43   0.352 0.724557    
## Condition.1RRAn         3440.72    7302.91   0.471 0.637597    
## Condition.1RRNe         2978.83   14480.67   0.206 0.837040    
## Condition.1RRNn         -861.61   11820.56  -0.073 0.941901    
## Condition.2Feedr      -10258.31   20496.05  -0.501 0.616784    
## Condition.2Norm       -20107.93   18055.32  -1.114 0.265566    
## Condition.2PosA       148180.65   36261.41   4.086 4.58e-05 ***
## Condition.2PosN       -15973.62   25935.24  -0.616 0.538037    
## Condition.2RRAn        -7814.35   35020.46  -0.223 0.823455    
## Condition.2RRNn       -13907.84   28129.12  -0.494 0.621065    
## Bldg.Type2fmCon        -6532.26    5176.24  -1.262 0.207126    
## Bldg.TypeDuplex         3706.68    4522.89   0.820 0.412589    
## Bldg.TypeTwnhs        -42601.30    5801.08  -7.344 3.16e-13 ***
## Bldg.TypeTwnhsE       -34080.03    3731.01  -9.134  < 2e-16 ***
```

```
## House.Style1.5Unf       -9375.22    9065.75  -1.034 0.301214    
## House.Style1Story      -14093.61    2781.17  -5.068 4.45e-07 ***
## House.Style2.5Fin       65513.41   18233.80   3.593 0.000336 ***
## House.Style2.5Unf       12540.77    9230.79   1.359 0.174453    
## House.Style2Story        1522.01    2942.51   0.517 0.605048    
## House.StyleSFoyer      -32238.74    5376.80  -5.996 2.45e-09 ***
## House.StyleSLvl        -21085.77    4604.82  -4.579 5.00e-06 ***
## Roof.StyleGable         31046.78   15874.58   1.956 0.050652 .  
## Roof.StyleGambrel       42793.53   17871.68   2.394 0.016747 *  
## Roof.StyleHip           39773.70   15972.45   2.490 0.012860 *  
## Roof.StyleMansard       24593.05   23406.18   1.051 0.293537    
## Roof.StyleShed         -47819.07   38691.74  -1.236 0.216661    
## Roof.MatlCompShg       291617.12   38284.46   7.617 4.20e-14 ***
## Roof.MatlMembran       323216.58   53671.20   6.022 2.09e-09 ***
## Roof.MatlMetal         286483.50   52932.46   5.412 7.08e-08 ***
## Roof.MatlTar&Grv       308147.75   41379.16   7.447 1.49e-13 ***
## Roof.MatlWdShake       304350.03   40755.51   7.468 1.28e-13 ***
## Roof.MatlWdShngl       398305.77   40968.01   9.722  < 2e-16 ***
## Exterior.1stAsphShn      2122.09   33814.75   0.063 0.949968    
## Exterior.1stBrkComm     30124.37   20436.64   1.474 0.140650    
## Exterior.1stBrkFace     36058.92   11786.10   3.059 0.002251 ** 
## Exterior.1stCBlock     -25020.94   41390.84  -0.605 0.545586    
## Exterior.1stCemntBd     64254.75   24479.00   2.625 0.008742 ** 
## Exterior.1stHdBoard      4824.93   11614.07   0.415 0.677871    
## Exterior.1stMetalSd     -2535.91   14809.80  -0.171 0.864061    
## Exterior.1stPlywood     17232.67   11331.43   1.521 0.128493    
## Exterior.1stStone       22893.52   36986.71   0.619 0.536018    
## Exterior.1stStucco       8300.40   13531.61   0.613 0.539686    
## Exterior.1stVinylSd       282.73   13111.39   0.022 0.982798    
## Exterior.1stWd Sdng      4227.83   11172.45   0.378 0.705168    
## Exterior.1stWdShing     -6596.56   12537.85  -0.526 0.598863    
## Exterior.2ndAsphShn      8496.93   26396.47   0.322 0.747569    
## Exterior.2ndBrk Cmn    -15276.71   20792.90  -0.735 0.462615    
## Exterior.2ndBrkFace     -6805.11   12995.67  -0.524 0.600591    
## Exterior.2ndCBlock      21606.10   26876.81   0.804 0.421567    
## Exterior.2ndCmentBd    -39244.16   24796.76  -1.583 0.113684    
## Exterior.2ndHdBoard      4763.94   11905.14   0.400 0.689088    
## Exterior.2ndImStucc     13405.43   15601.93   0.859 0.390338    
## Exterior.2ndMetalSd     14940.04   15084.16   0.990 0.322092    
## Exterior.2ndOther       -6972.15   32652.27  -0.214 0.830940    
## Exterior.2ndPlywood     -2752.58   11340.08  -0.243 0.808243    
## Exterior.2ndStone      -23969.45   21154.72  -1.133 0.257345    
## Exterior.2ndStucco      -6111.62   13500.36  -0.453 0.650820    
## Exterior.2ndVinylSd     11065.56   13179.19   0.840 0.401233    
## Exterior.2ndWd Sdng      5408.17   11452.92   0.472 0.636836    
## Exterior.2ndWd Shng      9885.27   12412.42   0.796 0.425906    
## Mas.Vnr.TypeBrkFace     -1830.69    8189.07  -0.224 0.823131    
## Mas.Vnr.TypeCBlock      51131.23   48170.98   1.061 0.288629    
## Mas.Vnr.TypeNone        -7119.98    8108.83  -0.878 0.380034    
## Mas.Vnr.TypeStone        1673.16    8522.55   0.196 0.844381    
## Exter.QualFa           -40456.73    9910.51  -4.082 4.66e-05 ***
## Exter.QualGd           -31067.19    4982.65  -6.235 5.64e-10 ***
## Exter.QualTA           -40208.55    5535.56  -7.264 5.62e-13 ***
## Exter.CondFa            -2944.14   12352.40  -0.238 0.811641    
```

```
## Exter.CondGd          5917.84   11157.30   0.530 0.595901
## Exter.CondTA          2136.43   10990.99   0.194 0.845900
## FoundationCBlock      2419.57    3185.69   0.760 0.447648
## FoundationPConc       5378.59    3461.78   1.554 0.120433
## FoundationSlab        7301.32   10045.83   0.727 0.467444
## FoundationStone       2660.00   12666.27   0.210 0.833687
## FoundationWood         467.26   15489.64   0.030 0.975938
## Bsmt.QualFa         -45219.32    6305.02  -7.172 1.08e-12 ***
## Bsmt.QualGd         -28962.66    3564.80  -8.125 8.34e-16 ***
## Bsmt.QualNone       -75661.72   41329.46  -1.831 0.067314 .
## Bsmt.QualPo         -19943.89   34000.29  -0.587 0.557561
## Bsmt.QualTA         -36519.97    4343.06  -8.409  < 2e-16 ***
## Bsmt.CondFa          -8836.65   18456.14  -0.479 0.632146
## Bsmt.CondGd          -6137.91   18339.06  -0.335 0.737898
## Bsmt.CondNone              NA         NA      NA       NA
## Bsmt.CondPo         -23919.49   29010.61  -0.825 0.409762
## Bsmt.CondTA          -5310.28   18034.00  -0.294 0.768441
## Bsmt.ExposureGd      22083.62    3074.91   7.182 1.01e-12 ***
## Bsmt.ExposureMn      -4716.21    3182.27  -1.482 0.138512
## Bsmt.ExposureNo      -7811.00    2379.86  -3.282 0.001050 **
## Bsmt.ExposureNone   -16393.13   20709.61  -0.792 0.428717
## BsmtFin.Type.1BLQ    -1501.57    2960.84  -0.507 0.612117
## BsmtFin.Type.1GLQ     4677.70    2625.31   1.782 0.074959 .
## BsmtFin.Type.1LwQ      850.35    3806.73   0.223 0.823264
## BsmtFin.Type.1None         NA         NA      NA       NA
## BsmtFin.Type.1Rec    -1782.76    3004.65  -0.593 0.553033
## BsmtFin.Type.1Unf    -6296.59    2678.74  -2.351 0.018854 *
## BsmtFin.Type.2BLQ    -6565.09    6641.78  -0.988 0.323066
## BsmtFin.Type.2GLQ     5646.08    8204.56   0.688 0.491440
## BsmtFin.Type.2LwQ    -3126.87    6410.16  -0.488 0.625753
## BsmtFin.Type.2None   19222.90   29660.80   0.648 0.517010
## BsmtFin.Type.2Rec    -4176.83    6370.39  -0.656 0.512126
## BsmtFin.Type.2Unf     -692.52    5244.37  -0.132 0.894960
## HeatingGasA           4908.42   31241.48   0.157 0.875174
## HeatingGasW          30402.98   32112.62   0.947 0.343890
## HeatingGrav           8378.25   35263.26   0.238 0.812225
## HeatingOthW          11152.35   38330.80   0.291 0.771124
## HeatingWall           8735.62   36573.95   0.239 0.811251
## Heating.QCFa        -10389.25    4794.07  -2.167 0.030360 *
## Heating.QCGd         -4437.86    2193.17  -2.023 0.043173 *
## Heating.QCPo        -33260.43   32261.48  -1.031 0.302699
## Heating.QCTA         -5320.87    2118.90  -2.511 0.012122 *
## Central.AirY          5992.27    3857.07   1.554 0.120464
## ElectricalFuseF      -5531.13    6140.91  -0.901 0.367870
## ElectricalFuseP      10225.67   16374.92   0.624 0.532399
## ElectricalSBrkr       1278.93    3082.43   0.415 0.678260
## Bsmt.Full.Bath1       5493.12    1926.78   2.851 0.004410 **
## Bsmt.Full.Bath2      22576.13    7913.88   2.853 0.004385 **
## Bsmt.Full.Bath3      17777.72   30832.87   0.577 0.564294
## Bsmt.Half.Bath1        736.33    3099.27   0.238 0.812233
## Bsmt.Half.Bath2     -22425.58   17653.38  -1.270 0.204135
## Kitchen.QualFa      -53747.49    6173.63  -8.706  < 2e-16 ***
## Kitchen.QualGd      -37760.70    3752.71 -10.062  < 2e-16 ***
## Kitchen.QualTA      -47255.24    4086.65 -11.563  < 2e-16 ***
```

```
## FunctionalMaj2          -28224.18   16579.36  -1.702 0.088862 .
## FunctionalMin1           -7843.98   10194.70  -0.769 0.441749
## FunctionalMin2           -2637.98   10268.87  -0.257 0.797293
## FunctionalMod              301.42   11431.09   0.026 0.978967
## FunctionalSev           -18326.27   24471.73  -0.749 0.454032
## FunctionalTyp            -2727.07    9135.36  -0.299 0.765343
## Fireplace.QuFa          -28035.62    7726.83  -3.628 0.000293 ***
## Fireplace.QuGd          -15810.49    6312.48  -2.505 0.012347 *
## Fireplace.QuNone        -30216.46    6452.70  -4.683 3.05e-06 ***
## Fireplace.QuPo          -26809.06    8303.08  -3.229 0.001266 **
## Fireplace.QuTA          -13804.89    6445.39  -2.142 0.032344 *
## Garage.TypeAttchd        23008.59    8472.20   2.716 0.006677 **
## Garage.TypeBasment       16944.07   10481.80   1.617 0.106160
## Garage.TypeBuiltIn       30115.79    8968.24   3.358 0.000802 ***
## Garage.TypeCarPort       -1321.76   12190.44  -0.108 0.913670
## Garage.TypeDetchd        12280.31    8488.10   1.447 0.148139
## Garage.TypeNone           9824.28   32306.31   0.304 0.761089
## Garage.FinishNone        21191.83   42481.12   0.499 0.617945
## Garage.FinishRFn         -2201.73    2074.14  -1.062 0.288600
## Garage.FinishUnf         -2339.46    2501.70  -0.935 0.349840
## Garage.Cars1             27389.57   44198.94   0.620 0.535543
## Garage.Cars2             39732.53   44256.96   0.898 0.369431
## Garage.Cars3             69102.26   44350.79   1.558 0.119392
## Garage.Cars4             62022.63   45428.57   1.365 0.172339
## Garage.Cars5             45451.15   53755.20   0.846 0.397934
## Garage.QualFa           -92662.47   42663.76  -2.172 0.029994 *
## Garage.QualGd           -65367.92   41670.26  -1.569 0.116898
## Garage.QualNone                NA         NA      NA       NA
## Garage.QualPo          -139571.20   49501.87  -2.820 0.004863 **
## Garage.QualTA           -89749.50   42488.10  -2.112 0.034797 *
## Garage.CondFa            75179.53   31123.19   2.416 0.015812 *
## Garage.CondGd            66318.36   31852.17   2.082 0.037480 *
## Garage.CondNone                NA         NA      NA       NA
## Garage.CondPo            89491.77   33343.91   2.684 0.007345 **
## Garage.CondTA            83576.90   30709.42   2.722 0.006562 **
## Paved.DriveP              3243.95    5547.38   0.585 0.558776
## Paved.DriveY              6509.84    3364.37   1.935 0.053158 .
## Pool.QCFa              -165945.95   35308.16  -4.700 2.80e-06 ***
## Pool.QCGd              -111285.06   26114.70  -4.261 2.14e-05 ***
## Pool.QCNone            -177794.63   18563.66  -9.578  < 2e-16 ***
## Pool.QCTA              -122124.69   28331.05  -4.311 1.72e-05 ***
## FenceGdWo                  665.44    5062.48   0.131 0.895437
## FenceMnPrv                2111.26    4168.61   0.506 0.612592
## FenceMnWw                 2566.99   12100.27   0.212 0.832020
## FenceNone                   94.93    3798.33   0.025 0.980064
## Misc.FeatureNone         -7091.01   20770.55  -0.341 0.732845
## Misc.FeatureOthr         26954.62   30428.45   0.886 0.375826
## Misc.FeatureShed         -5100.57   21098.60  -0.242 0.809003
## Misc.FeatureTenC         30924.24   47106.90   0.656 0.511608
## Sale.TypeCon             40974.32   21980.29   1.864 0.062468 .
## Sale.TypeConLD            9530.50    9340.65   1.020 0.307714
## Sale.TypeConLI           -4847.61   10918.71  -0.444 0.657117
## Sale.TypeConLw           -2557.13   15947.47  -0.160 0.872626
## Sale.TypeCWD             12531.18   11343.76   1.105 0.269450
```

```
## Sale.TypeNew            52337.82   18855.41   2.776 0.005566 **
## Sale.TypeOth            30715.80   17521.08   1.753 0.079762 .
## Sale.TypeVWD             4124.24   30083.03   0.137 0.890971
## Sale.TypeWD              -876.34    4198.03  -0.209 0.834666
## Sale.ConditionAdjLand   26481.01   10896.05   2.430 0.015184 *
## Sale.ConditionAlloca    17239.00    8646.77   1.994 0.046339 *
## Sale.ConditionFamily    -1464.56    6054.18  -0.242 0.808878
## Sale.ConditionNormal     9944.26    3054.32   3.256 0.001152 **
## Sale.ConditionPartial  -30822.10   18425.70  -1.673 0.094548 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28760 on 1767 degrees of freedom
## Multiple R-squared:  0.8812, Adjusted R-squared:  0.8657
## F-statistic: 56.52 on 232 and 1767 DF,  p-value: < 2.2e-16
```

Results

- What is the final regression model for the data?
- Using the standard diagnostic tests, does the model appear to fit the data well?
- What are the final confidence intervals for the effects of interest mentioned in the study section?

Use your validation data to construct these intervals. Do these intervals seem very sensitive to the choice of model (i.e. do they vary widely for different choices of variables in the model)? • What is your estimated prediction accuracy for your model? (Evaluated on the validation set). • Compare the intervals constructed using your final selected model fit to the validation set to the same intervals constructed on the training set. Are they very different? Which do you believe more?

Appendix

Ideally, there will be comments in the file, i.e. lines beginning with "#" to clarify what each part of the code is doing.

Acknowledgements : If you consult outside sources that refer to this data set, you should cite these

as references, and describe what you used from each source. Sources include material found on the internet, journal articles and books.