

Predicting House Prices in Ames, Iowa Using Multiple Linear Regression

Authors:

Joel Laskow

Chris Johnson

Introduction:

Any homebuyer in the 21st century will know the struggle of finding affordable housing. Over the past 20 years, popular discourse has grown increasingly louder regarding the state of the housing market. What was once a market of modestly priced dwellings has propagated into an ocean of listings many consider wildly unaffordable.

The scenario has many consumers wondering if house ownership is worth the effort, or if buying a home today is even a safe investment. With the daunting prices of today's home deterring demand, there is a possibility of a collapse in the housing market; it's imperative, then, that housing agencies work to identify the true value of a house in today's market. If this market is to stabilize, consumers and agencies alike must be able to agree on a reasonable price for a new home.

We will attempt to tackle this issue by identifying key variables related to house prices in Ames, Iowa. From these findings we will make predictions on housing prices.

This report will serve as a guide to walk readers through our findings and conclusions.

Data Description:

All training and test datasets (Train and Test, respectively) were obtained through *Kaggle*. This data contains 79 explanatory variables for 1460 housing observation in the training set and 1459 in the test set. . These variables include numeric and categorical descriptors of a litany of features found in different houses. Full datasets with all 79 variables can be obtained from the links provided in the Appendix, along with key code and graphs from our study.

Data Cleaning:

Numeric

The Housing dataset obtained from Kaggle was assessed for missing variables. Respectively, 0.6% of raw numeric variables in the training set were missing; 0.6% of raw numeric variables were missing in the test set. This accounted for 18% of LotsFrontage values in the training set and 16% in the test dataset, 6% of GarageYrBlt values in the training set and 5% of GarageYrBlt values in the test set. 1% of MasVnrArea values were missing from the test set. Missing values were imputed with

the mean of the respective explanatory variable. In the train set, 94% of Alley, 47% of FireplaceQu, 100% PoolQC, 81% Fence, and 95% of MiscFeature variables were missing; respectively, 93%, 50%, 100%, 80%, and 95% were missing from the test set. Given the proportion of missing values, these variables were not included in later tests. Remaining variables with missing values were imputed with the respective explanatory variable mode. Full graphs of missing variables before and after cleaning are included in **Figure 6**. Less than 0.1% of the test set was still missing after cleaning; however, due to time constraints and technical difficulties we were unable to identify and fix the remaining missing values. We moved forward under the assumption that these missing datapoints would not influence predictions.

Only one instance of MSSubClass of level “150” appeared in our data (test and train). Given the level’s similarity to MSSubClass level “50”, this datapoint’s MSSubClass factor level was changed to Level “50”. For descriptions of each MSSubClass level, see **Appendix II**.

Lastly, an [RShiny](#) application was created to explore distribution of GrLivArea and the distribution of SalePrice, as well as scatterplots exploring the correlation of these variables, in all Neighborhoods within our training set.

Analysis:

Part 1

Our first endeavor sought to quantify the relationship between GrLivArea (Living Area Above Ground), and Sale Price in the 3 neighborhoods: Brookside (BrkSide), Edwards (Edwards), and North Ames (NAMES).

Normality Checks

We checked the distribution of GrLivArea and SalePrice using visual means. Both variables were logged to account for outliers within the dataset. QQ and Histograms plots, as well as scatterplots comparing SalePrice and GrLivArea, before and after log transformation are shown in **Figure 2**. Visual assessment of log-transformed variables provided little evidence against assumptions of normality within the dataset.

Outlier Check

Residual vs Fitted plots, as well as Scale Location plots and Residualvs Leverage plots were used to visually identify potential high-leverage outliers (**Figure 3a**). Potential high-leverage outliers were identified at row 339, 136, 131, 190, 104, and 186, 411 of the dataset. After removal of these points, the dataset appeared free of high-leverage points (**Figure 3b**).

K-Fold Cross Validation was implemented to check the validity of the full model. At $k=5$, the mean CV Press score for our regression model was 4.847129. Adjusted R Squared with and without outliers were identified as 0.4857 and 0.4899 respectively.

Final Model Assessment

Final assessment of regression model comparing $\log(\text{GrLivArea})$ to $\log(\text{SalePrice})$ generated with and without high-leverage outliers (**Figure 4**). With outliers, we identified the coefficients associated with each parameter of our model (**Figure 5**).

The outlier-included data suggest that a doubling of GrLivArea is associated with, at the 95% confidence interval, a multiplicative change in median SalePrice of (3.1076, 3.6247) for Brookside, (1.89092, 2.06258) for Edwards, and (2.11195, 2.31203) for NAmes.

The outlier-omitted data suggest that a doubling of GrLivArea is associated with, at the 95% confidence interval, a multiplicative change in median SalePrice of (3.13508, 3.6724) for Brookside, (1.88357, 2.04337) for Edwards, and (1.88357, 2.27093) for NAmes.

Part 2

Variable Selection:

We constructed correlation matrices to identify potential multicollinearity among the 74 remaining explanatory variables in the cleaned test and train datasets. The logged variables, such as SalePrice, MsVnRArea, TotalBsmntSF, X1stFlrSF, X2ndFlrSF, LotFrontage, LotArea, YearBuilt, YearRemodAdd, GarageArea, WoodDeckSF, and OpenPorchSF, revealed some evidence of multicollinearity: LotArea with LotFrontage, YearBuilt with YearRemodAdd, and TotalBsmntSF with X1stFlrSF. Consequently, we addressed multicollinearity by removing LotFrontage, YearBuilt, TotalBsmntSF, and MasVnrArea. Additionally, WoodDeckSF, OpenPorchSF, MasVnrArea, BsmntFullBath, BsmntHalfBath, and HalfBath were excluded due to a lack of clear correlation with SalePrice." (**Figure 7**)

We moved forward with the following numerical variables: FirePlaces, GarageYrBlt, FullBath, TotRmsAbvGrd, X1stFlrSF, and LotArea, YearRemodAdd. Due to time constraints we focused solely on the categorical variable MSSubClass, which identifies the type of dwelling involved in the

sale. This variable was chosen because it handled descriptors and characteristics that were explained by other categorical variables. This left us with 8 variables out of the initial 74.

Full Regression Model:

$$\log(\text{SalePrice}) = \text{Fireplaces} + \text{GarageYrBlt} + \text{FullBath} + \text{TotRmsAbvGrd} + \log(\text{X1stFlrSF}) + \log(\text{LotArea}) + \text{YearRemodAdd} + \text{MSSubClass}$$

Forward, Backward, and Stepwise selection was conducted on candidate variables in SAS. Neither forward nor stepwise selection recommended any variables for removal. Backward selection elected for removal of TotRmsAbvGrd, however change in CV Press with removal was negligible compared to those of the stepwise and forward selection models. A summary of our forward, backward, and stepwise selection models are found in **Figure 8** and **Figure 9**. Final regression model was as follows:

$$\log(\text{SalePrice}) = \text{Fireplaces} + \text{GarageYrBlt} + \text{FullBath} + \text{TotRmsAbvGrd} + \log(\text{X1stFlrSF}) + \log(\text{LotArea}) + \text{YearRemodAdd} + \text{MSSubClass}$$

Assumption Checks

After constructing our final model, we examined residual distributions to assess normality, variance, linearity, and independence assumptions. The Residual Plot displayed random distribution without observable variance changes, while QQPlots and Histograms supported normality assumptions (**Figure 10**). Overall, assessments from scatterplots, QQ Plots, and Histograms found little evidence against the assumptions of normality, constant variance, or linear trend between $\log(\text{SalePrice})$ and our model's variables. Lastly, we assume independence among observations in our dataset.

Detecting High-Leverage Points

After checking assumptions, we checked for high-leverage datapoints within our training datasets. High leverage outliers were identified visually through Cook's D Bar Plot, and a Studentized Residuals vs Leverage Plot (**Figure 11**). Potential high leverage outliers were identified in rows 51, 513, and 859 of the cleaned training dataset.

To test the effects of these outliers on the final model, the training dataset was edited to remove these points. Final Adjusted R-Square changed from 0.7757 (with high-leverage outliers) to 0.7618 (without). Residual Standard Error changed from 0.1926 to 0.193. Given the minimal change in Adjusted R-Square and Residual Standard Error, the outliers were left in the dataset.

Predictions and Model Comparison

House SalePrice values for the test dataset were predicted using the final linear regression model (MLR). Predictions were also made with a simple linear regression model (SLR) using only

YearBuilt, and a second multiple linear regression model (Custom) using GrLivArea and FullBath. Respectively, Adjusted R2 values were 0.7691, 0.3436, and 0.93, while CV PRESS values were 55.51860, 153.34537, and 1149. **(Figure 12)**. Final Kaggle scores were, 0.20623 (MLR), 0.32043 (SLR), and 0.32034 (Custom).

Conclusion:

We are confident that our results reliably quantify the relationship between key housing variables and sales price in Ames, Iowa. It's important to iterate that the findings found in this report are limited to the population of Ames, Iowa. Additional information and comprehensive testing is required to extrapolate results to the larger housing population; nevertheless, we consider our findings invaluable for housing agencies and buyers seeking to identify optimal housing prices within Ames County. We further believe that similar testing methods might be applied to samples representative of other housing populations. Such testing could yield profound results that allow predictions of the wider housing population.

Appendix I

RShiny Link:

- https://cjohnson4510.shinyapps.io/Housing_App/

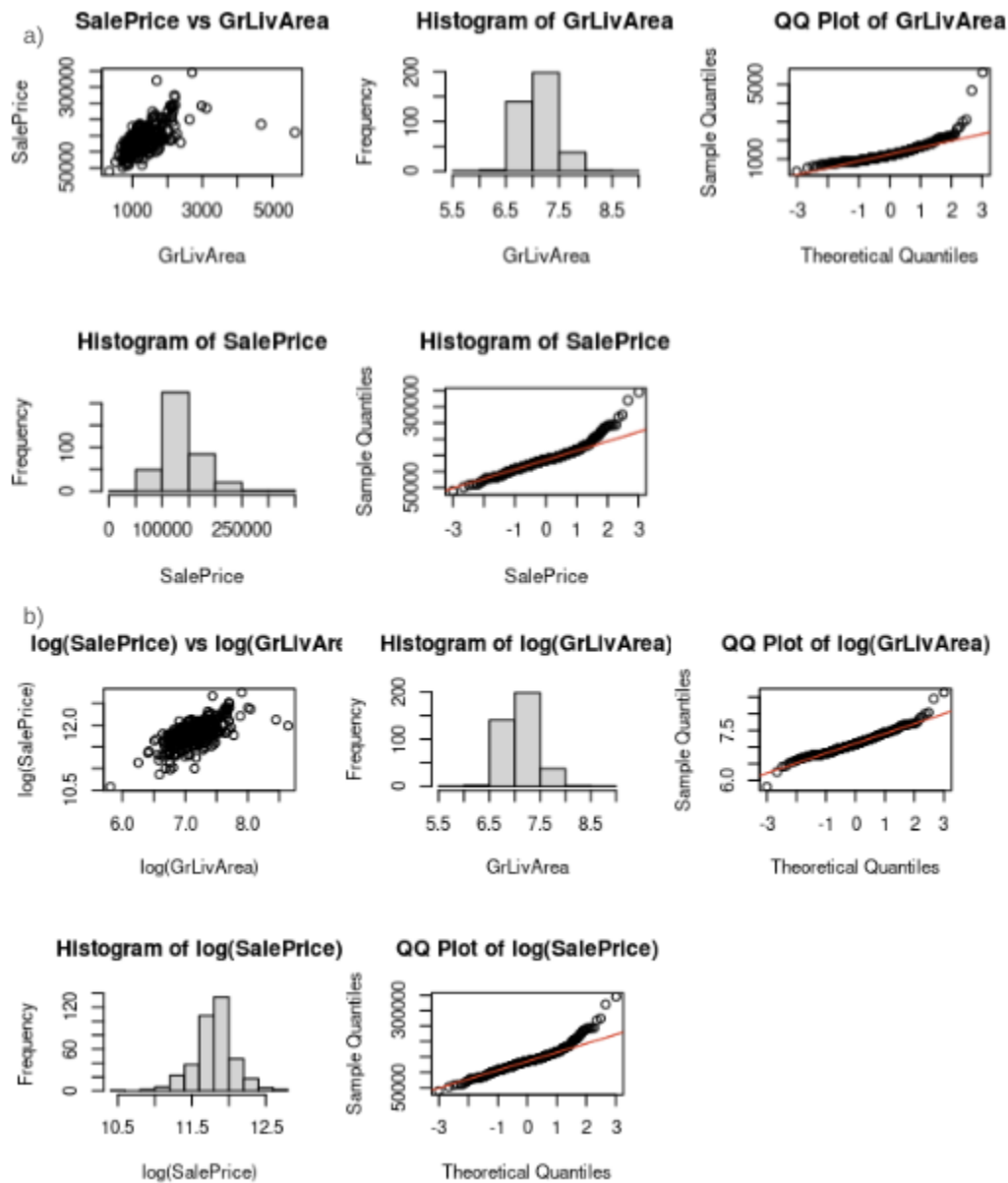


Figure 2. Normality graphs comparing GrLivArea to SalePrice. a) Scatterplot (top left) of SalePrice and GrLivArea, as well as histograms and QQ plots of each variable. b) Scatterplots, QQ Plots, and Histograms of variables after log transformation.

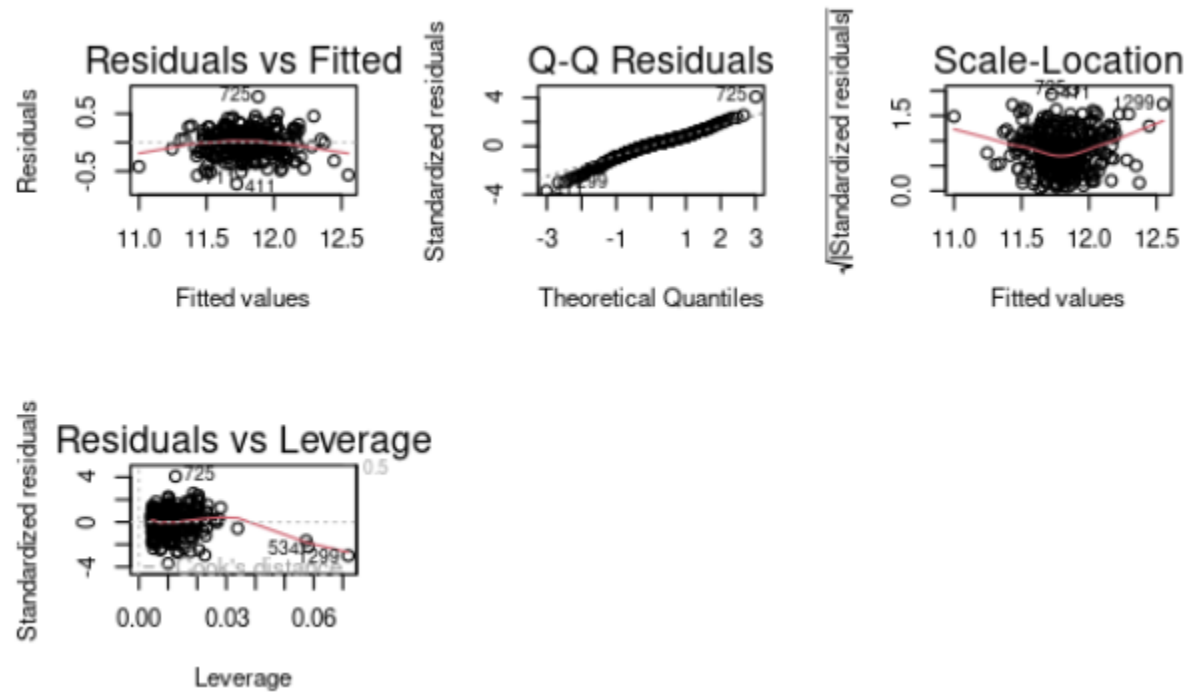


Figure 3. Visual assessment of high-leverage datapoints in Ames dataset. a) Residual and QQ plots with high-leverage datapoints. b) Residual and QQ Plots without high-leverage datapoints.

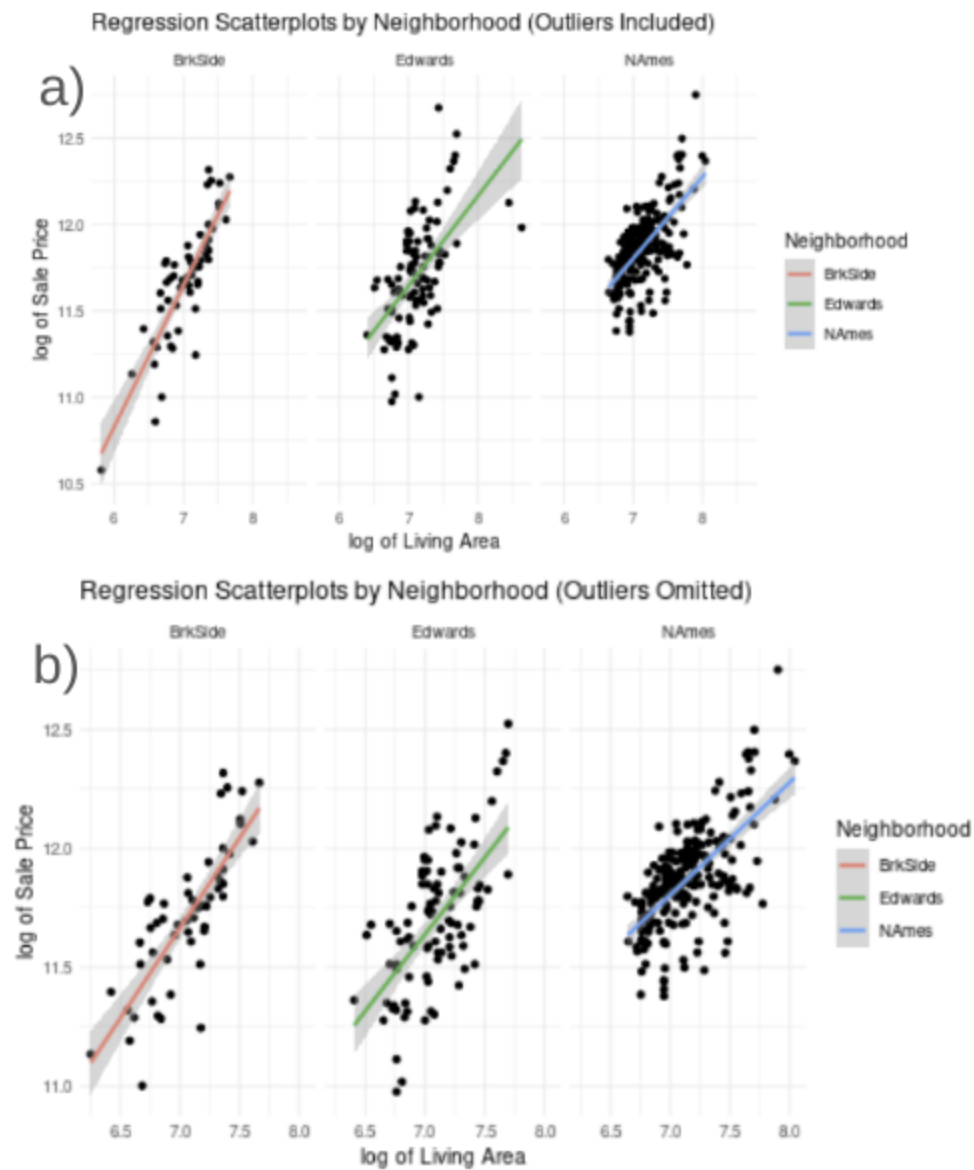


Figure 4. Scatterplot of $\log(\text{SalePrice})$ vs $\log(\text{GrLivArea})$. a) Outliers included. b) Outliers omitted.

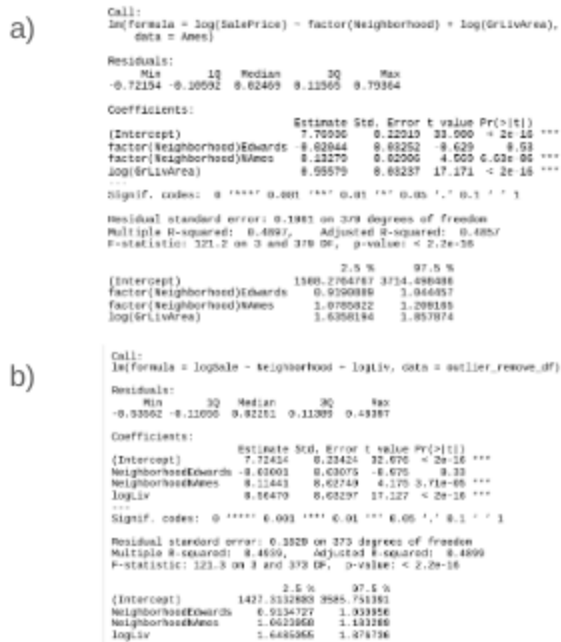


Figure 5. Coefficient summary and confidence intervals. a) Coefficients with high-leverage outliers. b) Coefficients without high-leverage outliers.

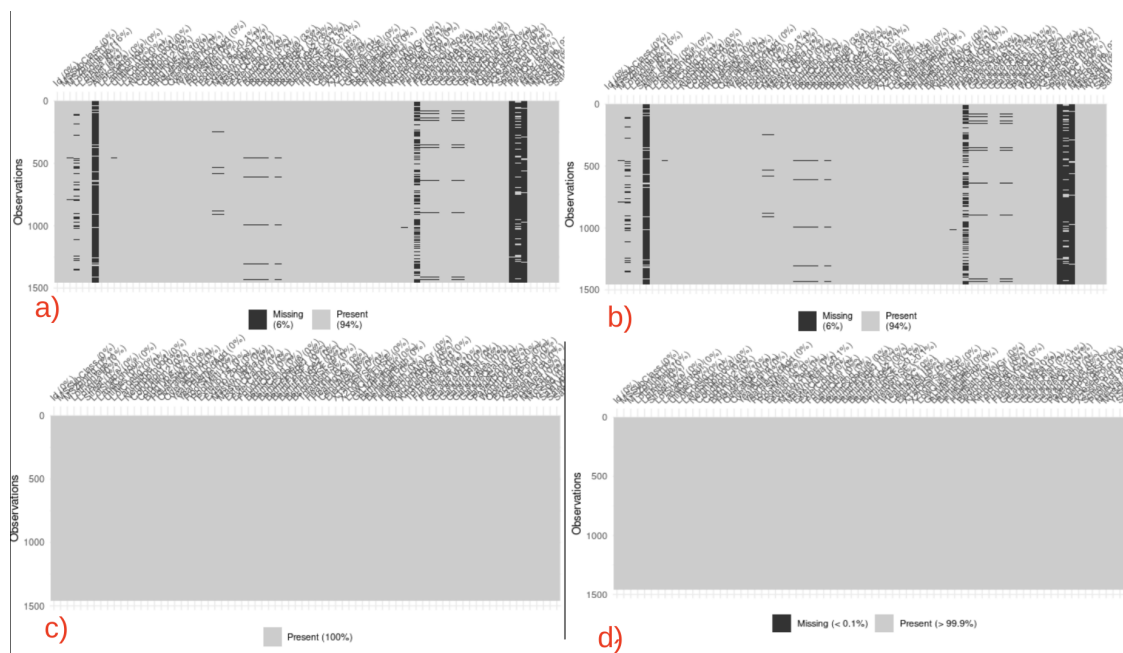


Figure 6: Test and Train datasets before and after cleaning. a) Train dataset before data cleaning. b) Test dataset before data cleaning. c) Train dataset after cleaning. d) Test dataset after cleaning.

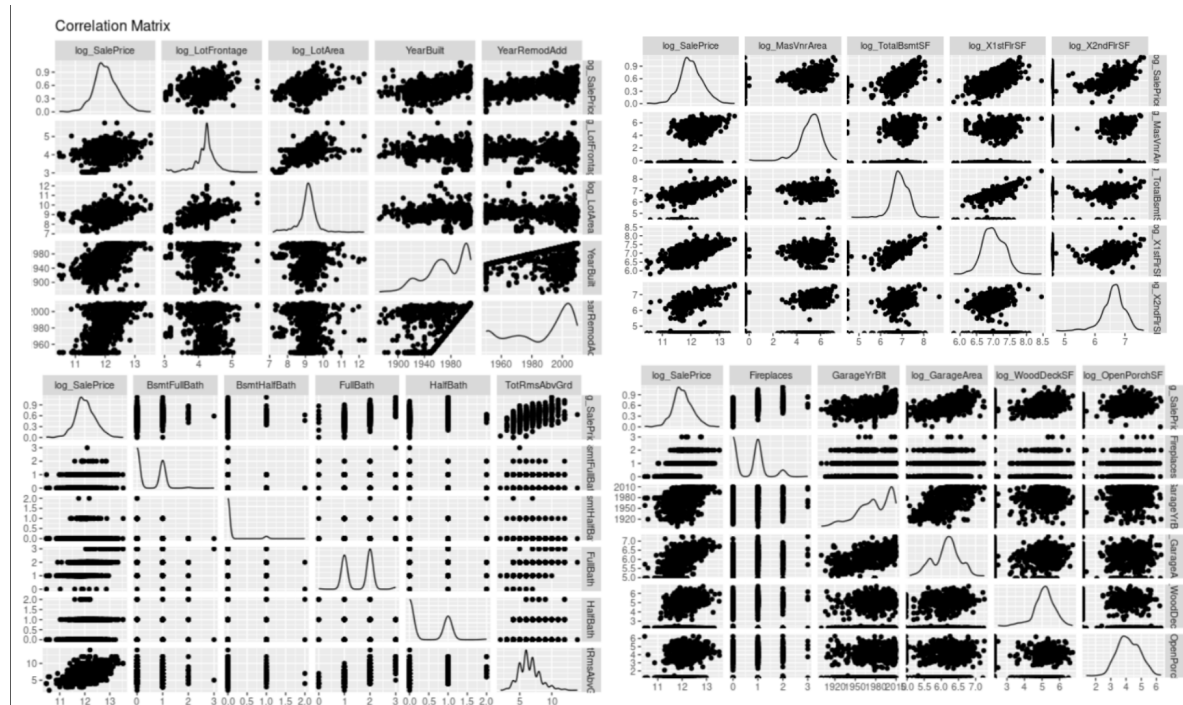


Figure 7: Correlation Matrices, Explanatory Variables vs $\log(\text{SalePrice})$

The GLMSELECT Procedure						Root MSE	0.19193
Backward Selection Summary						Dependent Mean	12.02405
Step	Effect Removed	Number Effects In	Number Parms In	Adjusted R-Square	SBC	CV PRESS	
0		9	22	0.7691*	-4681.6571*	55.0320*	
Selection stopped at a local minimum of the cross validation PRESS.							
Stop Details						Root MSE	0.19193
Candidate For	Effect	Candidate CV PRESS	Compare CV PRESS				
Removal	TotRmsAbvGrd	55.1482	> 55.0320				
The GLMSELECT Procedure						Root MSE	0.19193
Forward Selection Summary						Dependent Mean	12.02405
Step	Effect Entered	Number Effects In	Number Parms In	Adjusted R-Square	SBC	CV PRESS	
0	Intercept	1	1	0.0000	-2673.2872	233.1096	
1	log_X1stFlrSf	2	2	0.3704	-3342.4651	146.7944	
2	MSSubClass	3	16	0.6717	-4205.0769	77.4820	
3	YearRemodAdd	4	17	0.7324	-4497.4628	62.9954	
4	Fireplaces	5	18	0.7452	-4562.8544	59.9625	
5	GarageYrBlt	6	19	0.7554	-4616.0642	57.7340	
6	log_LotArea	7	20	0.7631	-4656.4895	56.0357	
7	FullBath	8	21	0.7680	-4681.1151	54.9612	
8	TotRmsAbvGrd	9	22	0.7691*	-4681.6571*	54.8469*	
* Optimal Value of Criterion						SBC	-4681.65712
						CV PRESS	55.03205
Selection stopped because all effects are in the final model.							
The GLMSELECT Procedure						Root MSE	0.19193
Stepwise Selection Summary						Dependent Mean	12.02405
Step	Effect Entered	Effect Removed	Number Effects In	Number Parms In	Adjusted R-Square	SBC	CV PRESS
0	Intercept		1	1	0.0000	-2673.2872	233.1248
1	log_X1stFlrSf		2	2	0.3704	-3342.4651	147.0949
2	MSSubClass		3	16	0.6717	-4205.0769	77.9763
3	YearRemodAdd		4	17	0.7324	-4497.4628	63.5033
4	Fireplaces		5	18	0.7452	-4562.8544	60.3709
5	GarageYrBlt		6	19	0.7554	-4616.0642	58.0424
6	log_LotArea		7	20	0.7631	-4656.4895	56.3586
7	FullBath		8	21	0.7680	-4681.1151	55.2389
8	TotRmsAbvGrd		9	22	0.7691*	-4681.6571*	55.0554*
* Optimal Value of Criterion						CV PRESS	55.05540
Selection stopped because all effects are in the final model.							

Figure 8. Summary Results for Backward (top), Forward (middle), and Stepwise (bottom).

	Adjusted R2	CV Press	Variables for Removal
Backward	0.7691	55.03205	TotRmsAbvGrd
Forward	0.7691	54.84692	None
Stepwise	0.7691	55.05540	None

Figure 9. Primary values used for evaluating selection methods.

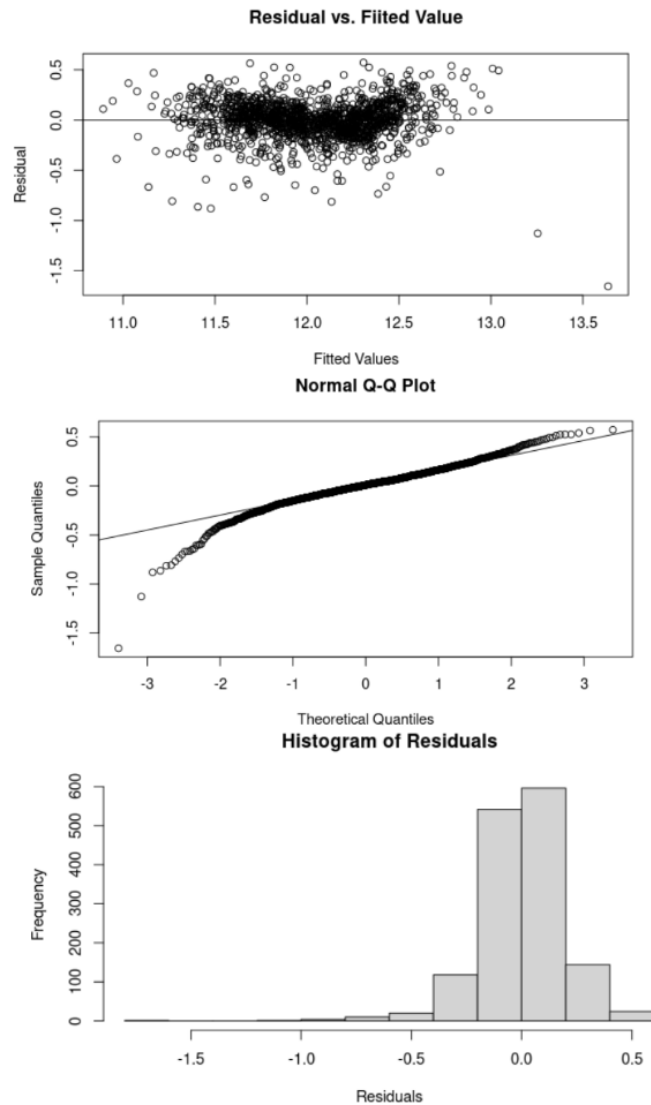


Figure 10. Residual plots for final model. Top) Residual plot; Middle) QQ plot of residuals; Bottom) Histogram of residuals.

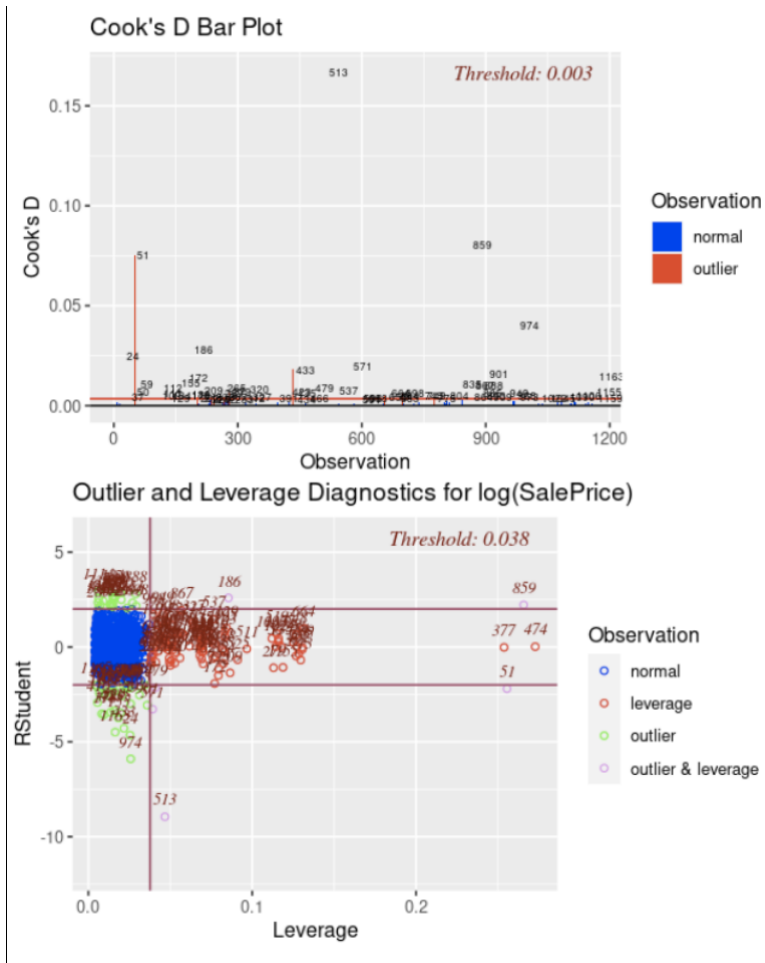


Figure 11. Graphical visualization of potential high-leverage outliers. Top) Cook's D per Observation. Bottom) RStudent values vs Leverage.

Predictive Models	Adjusted R2	CV PRESS	Kaggle Score
Multiple Linear Regression Model (MLR)	0.7691	55.51860	0.20623

<i>Simple Linear Regression (YearBuilt)</i>	0.3436	153.34537	0.32043
<i>Custom (GrLivArea + FullBath)</i>	0.93	1149	0.32034

Figure 12. Adjusted R2, CV PRESS, and Kaggle Score. Top) Final multiple linear regression model. Middle) Simple linear regression with YearBuilt . Bottom) Custom model, GrLivArea and FullBath.

Appendix II

Variables Descriptors

MSSubClass: Identifies the type of dwelling involved in the sale.

20	1-STORY 1946 & NEWER ALL STYLES
30	1-STORY 1945 & OLDER
40	1-STORY W/FINISHED ATTIC ALL AGES
45	1-1/2 STORY - UNFINISHED ALL AGES
50	1-1/2 STORY FINISHED ALL AGES
60	2-STORY 1946 & NEWER
70	2-STORY 1945 & OLDER
75	2-1/2 STORY ALL AGES
80	SPLIT OR MULTI-LEVEL
85	SPLIT FOYER
90	DUPLEX - ALL STYLES AND AGES
120	1-STORY PUD (Planned Unit Development) - 1946 & NEWER
150	1-1/2 STORY PUD - ALL AGES
160	2-STORY PUD - 1946 & NEWER
180	PUD - MULTILEVEL - INCL SPLIT LEV/FOYER
190	2 FAMILY CONVERSION - ALL STYLES AND AGES

MSZoning: Identifies the general zoning classification of the sale.

A	Agriculture
C	Commercial
FV	Floating Village Residential
I	Industrial
RH	Residential High Density
RL	Residential Low Density
RP	Residential Low Density Park
RM	Residential Medium Density

LotFrontage: Linear feet of street connected to property

LotArea: Lot size in square feet

Street: Type of road access to property

Grvl	Gravel
Pave	Paved

Alley: Type of alley access to property

Grvl	Gravel
Pave	Paved
NA	No alley access

LotShape: General shape of property

Reg	Regular
IR1	Slightly irregular
IR2	Moderately Irregular
IR3	Irregular

LandContour: Flatness of the property

Lvl	Near Flat/Level
Bnk	Banked - Quick and significant rise from street grade to building
HLS	Hillside - Significant slope from side to side
Low	Depression

Utilities: Type of utilities available

AllPub	All public Utilities (E,G,W,& S)
NoSewr	Electricity, Gas, and Water (Septic Tank)
NoSeWa	Electricity and Gas Only
ELO	Electricity only

LotConfig: Lot configuration

Inside	Inside lot
Corner	Corner lot
CulDSac	Cul-de-sac
FR2	Frontage on 2 sides of property
FR3	Frontage on 3 sides of property

LandSlope: Slope of property

Gtl	Gentle slope
Mod	Moderate Slope
Sev	Severe Slope

Neighborhood: Physical locations within Ames city limits

Blmngtn	Bloomington Heights
Blueste	Bluestem
BrDale	Briardale
BrkSide	Brookside
ClearCr	Clear Creek
CollgCr	College Creek
Crawfor	Crawford
Edwards	Edwards
Gilbert	Gilbert
IDOTRR	Iowa DOT and Rail Road
MeadowV	Meadow Village
Mitchel	Mitchell
Names	North Ames
NoRidge	Northridge
NPkVill	Northpark Villa
NridgHt	Northridge Heights
NWAmes	Northwest Ames
OldTown	Old Town
SWISU	South & West of Iowa State University
Sawyer	Sawyer
SawyerW	Sawyer West
Somerst	Somerset
StoneBr	Stone Brook
Timber	Timberland
Veenker	Veenker

Condition1: Proximity to various conditions

Artery	Adjacent to arterial street
Feedr	Adjacent to feeder street
Norm	Normal
RRNn	Within 200' of North-South Railroad
RRAn	Adjacent to North-South Railroad
PosN	Near positive off-site feature--park, greenbelt, etc.
PosA	Adjacent to postive off-site feature
RRNe	Within 200' of East-West Railroad
RR Ae	Adjacent to East-West Railroad

Condition2: Proximity to various conditions (if more than one is present)

Artery	Adjacent to arterial street
Feedr	Adjacent to feeder street
Norm	Normal
RRNn	Within 200' of North-South Railroad
RRAn	Adjacent to North-South Railroad
PosN	Near positive off-site feature--park, greenbelt, etc.
PosA	Adjacent to postive off-site feature
RRNe	Within 200' of East-West Railroad
RR Ae	Adjacent to East-West Railroad

BldgType: Type of dwelling

1Fam	Single-family Detached
2FmCon	Two-family Conversion; originally built as one-family dwelling
Duplx	Duplex
TwtnhsE	Townhouse End Unit
TwtnhsI	Townhouse Inside Unit

HouseStyle: Style of dwelling

1Story	One story
1.5Fin	One and one-half story: 2nd level finished
1.5Unf	One and one-half story: 2nd level unfinished
2Story	Two story
2.5Fin	Two and one-half story: 2nd level finished
2.5Unf	Two and one-half story: 2nd level unfinished
SFoyer	Split Foyer
SLvl	Split Level

OverallQual: Rates the overall material and finish of the house

10	Very Excellent
9	Excellent

8	Very Good
7	Good
6	Above Average
5	Average
4	Below Average
3	Fair
2	Poor
1	Very Poor

OverallCond: Rates the overall condition of the house

10	Very Excellent
9	Excellent
8	Very Good
7	Good
6	Above Average
5	Average
4	Below Average
3	Fair
2	Poor
1	Very Poor

YearBuilt: Original construction date

YearRemodAdd: Remodel date (same as construction date if no remodeling or additions)

RoofStyle: Type of roof

Flat	Flat	
Gable	Gable	
Gambrel		Gabrel (Barn)
Hip	Hip	
Mansard		Mansard
Shed	Shed	

RoofMatl: Roof material

ClyTile	Clay or Tile	
CompShg		Standard (Composite) Shingle
Membran		Membrane
Metal	Metal	
Roll	Roll	
Tar&Grv		Gravel & Tar
WdShake		Wood Shakes
WdShngl		Wood Shingles

Exterior1st: Exterior covering on house

AsbShng		Asbestos Shingles
AsphShn		Asphalt Shingles
BrkComm		Brick Common
BrkFaceBrick	Face	
CBlock	Cinder Block	
CemntBd		Cement Board
HdBoard		Hard Board
ImStucc	Imitation Stucco	
MetalSd	Metal Siding	
Other	Other	
Plywood		Plywood
PreCast	PreCast	
Stone	Stone	
Stucco	Stucco	
VinylSd	Vinyl Siding	
Wd Sdng		Wood Siding
WdShing		Wood Shingles

Exterior2nd: Exterior covering on house (if more than one material)

AsbShng		Asbestos Shingles
AsphShn		Asphalt Shingles
BrkComm		Brick Common
BrkFaceBrick	Face	
CBlock	Cinder Block	
CemntBd		Cement Board
HdBoard		Hard Board
ImStucc	Imitation Stucco	
MetalSd	Metal Siding	
Other	Other	
Plywood		Plywood
PreCast	PreCast	

Stone	Stone
Stucco	Stucco
VinylSd	Vinyl Siding
Wd Sdng	Wood Siding
WdShing	Wood Shingles

MasVnrType: Masonry veneer type

BrkCmn	Brick Common
BrkFace	Brick Face
CBlock	Cinder Block
None	None
Stone	Stone

MasVnrArea: Masonry veneer area in square feet

ExterQual: Evaluates the quality of the material on the exterior

Ex	Excellent
Gd	Good
TA	Average/Typical
Fa	Fair
Po	Poor

ExterCond: Evaluates the present condition of the material on the exterior

Ex	Excellent
Gd	Good
TA	Average/Typical
Fa	Fair
Po	Poor

Foundation: Type of foundation

BrkTil	Brick & Tile
CBlock	Cinder Block
PConc	Poured Contrete
Slab	Slab
Stone	Stone
Wood	Wood

BsmtQual: Evaluates the height of the basement

Ex	Excellent (100+ inches)
Gd	Good (90-99 inches)
TA	Typical (80-89 inches)
Fa	Fair (70-79 inches)
Po	Poor (<70 inches)
NA	No Basement

BsmtCond: Evaluates the general condition of the basement

Ex	Excellent
Gd	Good
TA	Typical - slight dampness allowed
Fa	Fair - dampness or some cracking or settling
Po	Poor - Severe cracking, settling, or wetness
NA	No Basement

BsmtExposure: Refers to walkout or garden level walls

Gd	Good Exposure
Av	Average Exposure (split levels or foyers typically score average or above)
Mn	Mimimum Exposure
No	No Exposure
NA	No Basement

BsmtFinType1: Rating of basement finished area

GLQ	Good Living Quarters
ALQ	Average Living Quarters
BLQ	Below Average Living Quarters
Rec	Average Rec Room
LwQ	Low Quality
Unf	Unfinished
NA	No Basement

BsmtFinSF1: Type 1 finished square feet

BsmtFinType2: Rating of basement finished area (if multiple types)

GLQ	Good Living Quarters
ALQ	Average Living Quarters
BLQ	Below Average Living Quarters
Rec	Average Rec Room
LwQ	Low Quality
Unf	Unfinished
NA	No Basement

BsmtFinSF2: Type 2 finished square feet

BsmtUnfSF: Unfinished square feet of basement area

TotalBsmtSF: Total square feet of basement area

Heating: Type of heating

Floor	Floor Furnace
GasA	Gas forced warm air furnace
GasW	Gas hot water or steam heat
Grav	Gravity furnace
OthW	Hot water or steam heat other than gas
Wall	Wall furnace

HeatingQC: Heating quality and condition

Ex	Excellent
Gd	Good
TA	Average/Typical
Fa	Fair
Po	Poor

CentralAir: Central air conditioning

N	No
Y	Yes

Electrical: Electrical system

SBrkr	Standard Circuit Breakers & Romex
FuseA	Fuse Box over 60 AMP and all Romex wiring (Average)
FuseF	60 AMP Fuse Box and mostly Romex wiring (Fair)
FuseP	60 AMP Fuse Box and mostly knob & tube wiring (poor)
Mix	Mixed

1stFlrSF: First Floor square feet

2ndFlrSF: Second floor square feet

LowQualFinSF: Low quality finished square feet (all floors)

GrLivArea: Above grade (ground) living area square feet

BsmtFullBath: Basement full bathrooms

BsmtHalfBath: Basement half bathrooms

FullBath: Full bathrooms above grade

HalfBath: Half baths above grade

Bedroom: Bedrooms above grade (does NOT include basement bedrooms)

Kitchen: Kitchens above grade

KitchenQual: Kitchen quality

Ex	Excellent
Gd	Good
TA	Typical/Average
Fa	Fair
Po	Poor

TotRmsAbvGrd: Total rooms above grade (does not include bathrooms)

Functional: Home functionality (Assume typical unless deductions are warranted)

Typ	Typical Functionality
Min1	Minor Deductions 1
Min2	Minor Deductions 2

Mod	Moderate Deductions
Maj1	Major Deductions 1
Maj2	Major Deductions 2
Sev	Severely Damaged
Sal	Salvage only

Fireplaces: Number of fireplaces

FireplaceQu: Fireplace quality

Ex	Excellent - Exceptional Masonry Fireplace
Gd	Good - Masonry Fireplace in main level
TA	Average - Prefabricated Fireplace in main living area or Masonry Fireplace in basement
Fa	Fair - Prefabricated Fireplace in basement
Po	Poor - Ben Franklin Stove
NA	No Fireplace

GarageType: Garage location

2Types	More than one type of garage
Attchd	Attached to home
Basment	Basement Garage
BuiltIn	Built-In (Garage part of house - typically has room above garage)
CarPort	Car Port
Detchd	Detached from home
NA	No Garage

GarageYrBlt: Year garage was built

GarageFinish: Interior finish of the garage

Fin	Finished
RFn	Rough Finished
Unf	Unfinished
NA	No Garage

GarageCars: Size of garage in car capacity

GarageArea: Size of garage in square feet

GarageQual: Garage quality

Ex	Excellent
Gd	Good
TA	Typical/Average
Fa	Fair
Po	Poor
NA	No Garage

GarageCond: Garage condition

Ex	Excellent
Gd	Good
TA	Typical/Average
Fa	Fair
Po	Poor
NA	No Garage

PavedDrive: Paved driveway

Y	Paved
P	Partial Pavement
N	Dirt/Gravel

WoodDeckSF: Wood deck area in square feet

OpenPorchSF: Open porch area in square feet

EnclosedPorch: Enclosed porch area in square feet

3SsnPorch: Three season porch area in square feet

ScreenPorch: Screen porch area in square feet

PoolArea: Pool area in square feet

PoolQC: Pool quality

Ex	Excellent
Gd	Good

TA	Average/Typical
Fa	Fair
NA	No Pool

Fence: Fence quality

GdPrv	Good Privacy
MnPrv	Minimum Privacy
GdWo	Good Wood
MnWw	Minimum Wood/Wire
NA	No Fence

MiscFeature: Miscellaneous feature not covered in other categories

Elev	Elevator
Gar2	2nd Garage (if not described in garage section)
Othr	Other
Shed	Shed (over 100 SF)
TenC	Tennis Court
NA	None

MiscVal: \$Value of miscellaneous feature

MoSold: Month Sold (MM)

YrSold: Year Sold (YYYY)

SaleType: Type of sale

WD	Warranty Deed - Conventional
CWD	Warranty Deed - Cash
VWD	Warranty Deed - VA Loan
New	Home just constructed and sold
COD	Court Officer Deed/Estate
Con	Contract 15% Down payment regular terms
ConLw	Contract Low Down payment and low interest
ConLI	Contract Low Interest
ConLD	Contract Low Down
Oth	Other

SaleCondition: Condition of sale

Normal	Normal Sale
Abnorml	Abnormal Sale - trade, foreclosure, short sale
AdjLand	Adjoining Land Purchase
Alloca	Allocation - two linked properties with separate deeds, typically condo with a garage unit
Family	Sale between family members
Partial	Home was not completed when last assessed (associated with New Homes)