

Fairfax County Real Estate Regression Analysis

Project 02

STAT 4214 (19604)

Shelby Neal

05/10/2022

1 Executive Summary

The following report further extends prior research on predicting price points for real estate sales. Since I currently plan to move to Fairfax County and have hopes of becoming a home buyer in the near future, I selected Fairfax County real estate data exclusively for analysis. The recent economic boom in the U.S. housing market over the past decade has caused property values to soar. As a result of these price jumps, data analysts and statisticians have strong incentive to discover the relationships that are causing the price of a home in the U.S. to go up.

Data was collected from Fairfax County's online real estate database and after pre-processing the full dataset consisted of 438 observations with 10 possible regressors (6 numerical and 1 categorical with 4 factor levels). Prior to generating the model, a hypothesis was made for the relationship of the response against each regressor. This resulted in 7 different hypotheses of which 5 were found to be true.

Both the postulated full and reduced models generated strong multiple R-squared values greater than 0.9, which indicated most of the variance could be explained by the models. The number of potential outliers was small relative to the full dataset and showed up to 8% of the observations may require consultation with a subject matter expert or Fairfax County database maintainer. One residual on the rambler home could not be calculated due to there being only one observation for that factor level in the dataset. As a result, calculating the PRESS statistic, which is a leave-one-out refitting technique, is impossible with only one observation. Due to this, it would be justifiable to remove this data point and factor level entirely in future analyses unless additional data for the rambler property type can be obtained.

Additionally, a look at the normal probability plot of the studentized deleted residuals show potentially greater-than-normal distributions for the dataset. Transformations of the response were tested but did not significantly aid the normality assumption required for OLS. The presence of extreme values about the tails could be due to any number of reasons; but, strong possibilities are measurement error, data corruption, and rare events. For the purpose of this analysis, the data is treated as Gaussian considering the Gaussian-like distribution about the centroid and lack of certainty on the subject matter and accuracy of reported data.

Finally, after obtaining the reduced models it was found that a property's type, gross leasing area, and recent sales price have the greatest predictive power when attempting to predict the home's value. For best results, prospective sellers should place higher focus on these metrics if planning to put their home on the market in 2022; and, prospective buyers should compare these metrics to the requested contract price prior to signing.

2 Problem Context

2.1 Reasons of Interest

Statisticians and data analysts have been using housing data for decades to better understand and predict perfect price points for real estate transactions. In 2021, the full U.S. housing stock gained \$6.9 trillion [2]. This increased the total market cap to more than \$43 trillion and marked the largest stock increase in a single year. Due to strong demand and limited supply, home values grew nearly 20% last year which marked another all-time high. These unprecedented rises in housing stock have pushed home values to more than twice their value during the Great Recession over a decade ago.

With the recent upswing in U.S. home values, prospective private sellers and property management firms have much greater incentive to capitalize on the volatility in the real estate market. Therefore, it is imperative sellers in the market are equipped with the best tools to prepare for upcoming 2022 real estate transactions. Regression analysis on historical sales data is one of these tools.

The price of a home is highly dependent upon location. For example, single-family residences in urban areas tend to get priced higher than their respective counterparts in rural areas. Therefore, comparing homes in similarly priced areas will likely result in a more accurate model. Other metrics, such as the available space and age of the home, also likely impact the total value of the property. To best analyze these metrics and their impact on house pricing, historical sales data on single-family homes built within the last decade were manually collected and recorded in csv format from Fairfax County's online database [1].

2.2 Data Collection

The data were indexed by parcel ID and include the following regressors: year built, type (end-unit town home, internal-unit town home, colonial home, residential condominium, or rambler), number of bedrooms, number of bathrooms, gross leasing area, most recent sales year, and sales price for that year. The response, total value, indicates the current estimated value of the property and is a fair expectation for a 2022 sales price should the owner be interested in selling. Only data on residential units were collected; commercial and industrial units were omitted.

Other fields available on Fairfax County's database were not included in the data collection, such as location (address), owner, lot size, land use code, description, subdivision, and book page. Location, by address, cannot improve the model unless this data is converted to geolocation coordinates (i.e., longitude and latitude); however, since all data was collected from the same county, it is expected most if not all of the homes lie within similarly priced areas and can be fairly compared. Subdivision is another field which could report a relatively more accurate location within the county; but, with over 40 different subdivisions, the inclusion of subdivision as a regressor would merely introduce unnecessary complexity to the model. Owner and book page, like parcel ID, are just other means of indexing the sales transactions. Land use code and description simply expand upon describing the use of the property, which is already explained by the property type. Lot size for residential condominiums is zero; thus, gross leasing area was the metric used to explain the available space within the property and lot size was unused.

There were a number of data points with missing values for gross leasing area, year sold, and sales price. As a result, these data were removed from the dataset. Since the data were manually recorded into csv format, the data were double checked to ensure no human error was introduced into the model. Lastly, some of the data recorded showed the property being sold years before it was built. Properties sold within 2 years of being built could be justified as having been new constructions and were left in the model; however, properties with larger gaps from year sold to year built were removed on grounds that the data must be inaccurately reported. For example, one property was said to have been sold in 1975 and then built in 2013. After pre-processing, the data remaining included 438 observations with 7 regressors.

2.3 Hypotheses

2.3.1 H1: Year Built is positively related to Total Value

It is expected that more recent constructions will have a higher total value due to reduced wear-and-tear on the property's interior, exterior, and appliances.

2.3.2 H2: Number of Bedrooms is positively related to Total Value

Properties with more bedrooms are expected to be priced higher than properties with less room for tenants.

2.3.3 H3: Number of Bathrooms is positively related to Total Value

Properties with more bathrooms are also expected to be priced higher on the same basis as H2.

2.3.4 H4: Gross Leasing Area is positively related to Total Value

Gross Leasing Area is the best indicator of the available living space for the property; it is expected that properties with higher GLA will be priced higher on the same basis as H2.

2.3.5 H5: Most Recent Year Sold is positively related to Total Value

Properties recently sold likely indicate stronger demand for those properties. It is expected that properties sold more recently will be priced higher than properties that have been off the market.

2.3.6 H6: Most Recent Sales Price is positively related to Total Value

Higher sales prices indicate higher property values; thus, properties that sold higher will likely sell higher in the future.

2.3.7 H7: Colonial and Rambler Property Types are positively related to Total Value

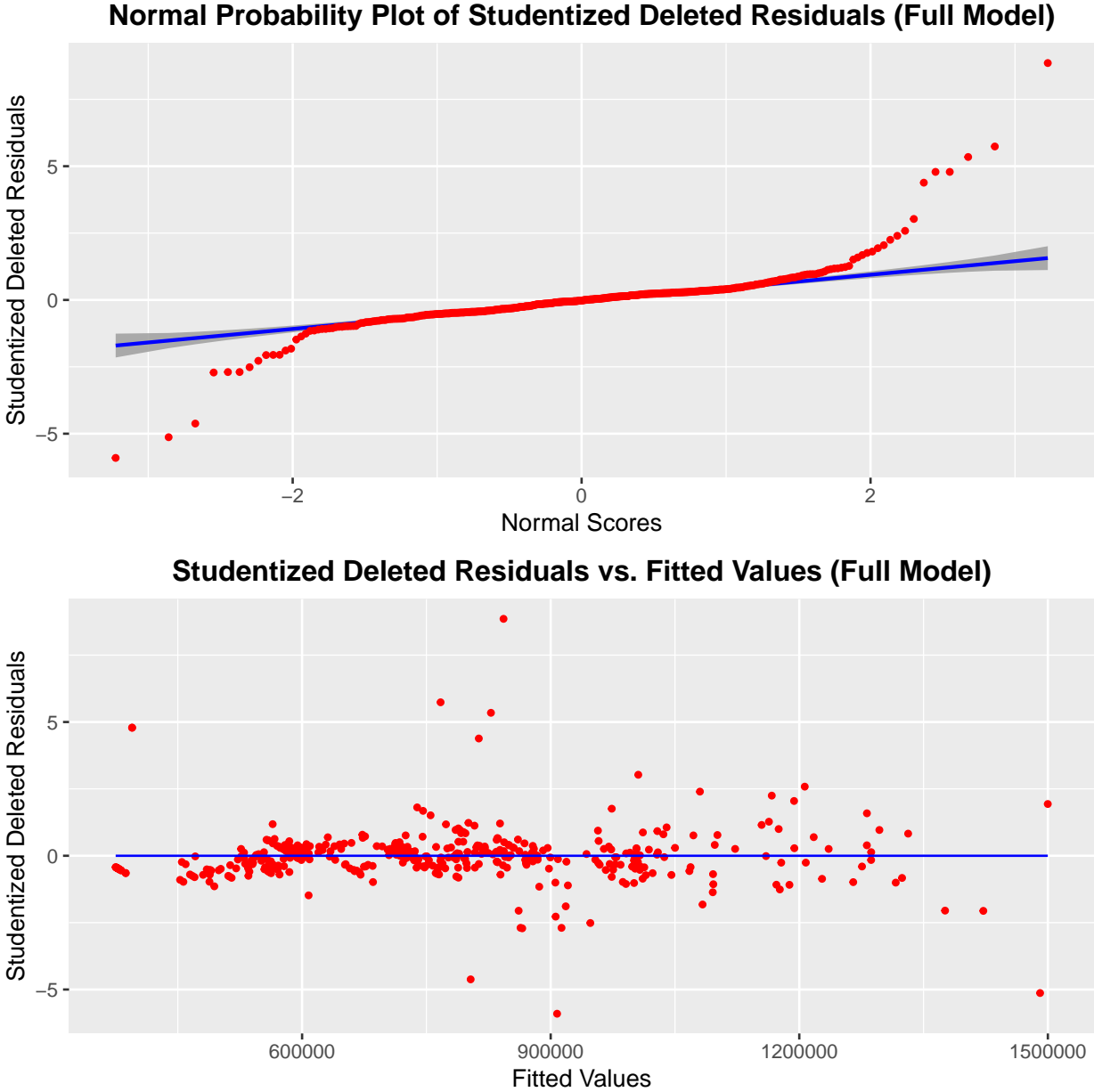
Colonial and Rambler property types have the luxury of not being attached to neighboring tenants. Due to this luxury, it is expected that these properties will be priced higher than town homes and condominiums.

3 Data Analysis

For the dataset used in this analysis, see the Appendix.

3.1 Full Model

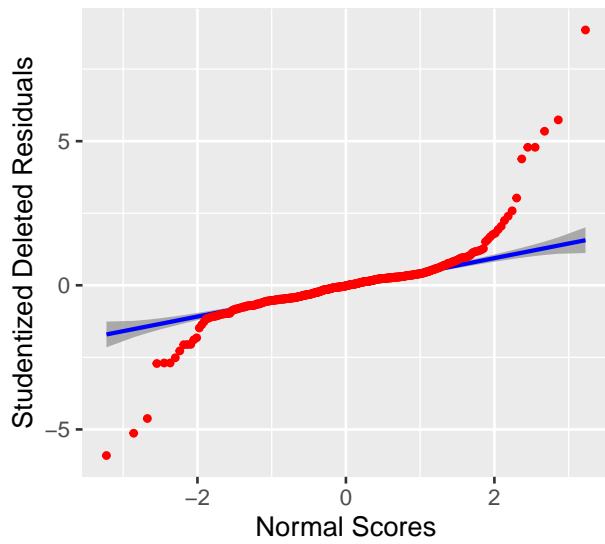
Using Ordinary Least Squares, a full model was generated with current value as the response and the following regressors: year built, property type, beds, baths, gross leasing area (GLA), year sold, and most recent sales price. A normal probability plot shows a potentially greater-than-normal distribution with no clear patterns in the plot of studentized deleted residuals versus fitted values.



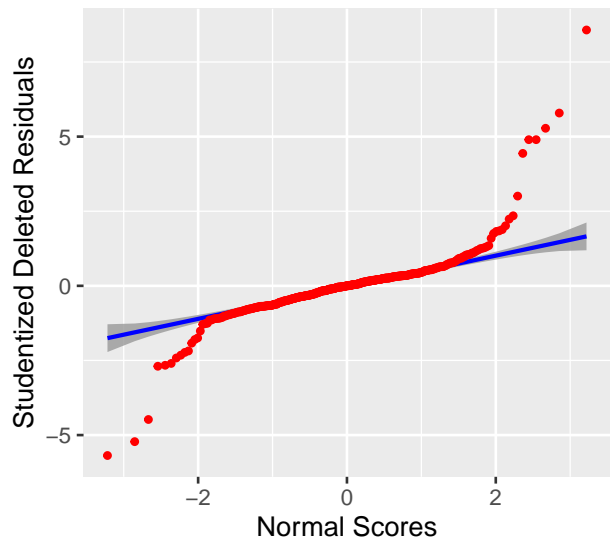
Accordingly, a transformation of the response may be appropriate. First, a boxcox transformation performed on the full model indicates an optimal transformation parameter of $\lambda = 0.869$. Other transformations, including taking the square root and log of the response, were also tested.

Normal Probability Plot of Studentized Deleted Residuals (Full Model)

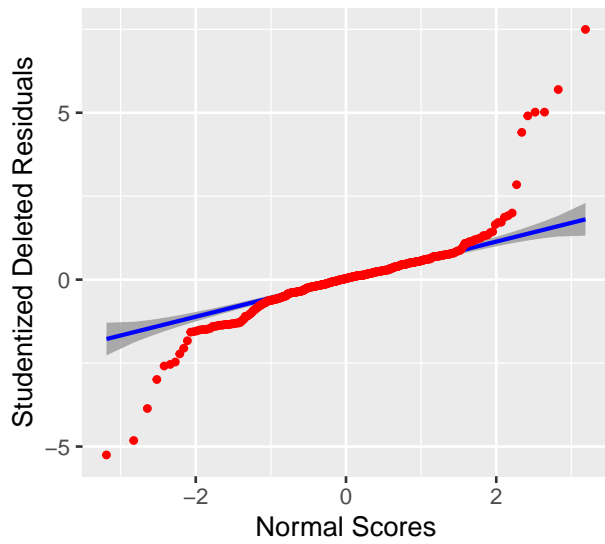
No Transformation



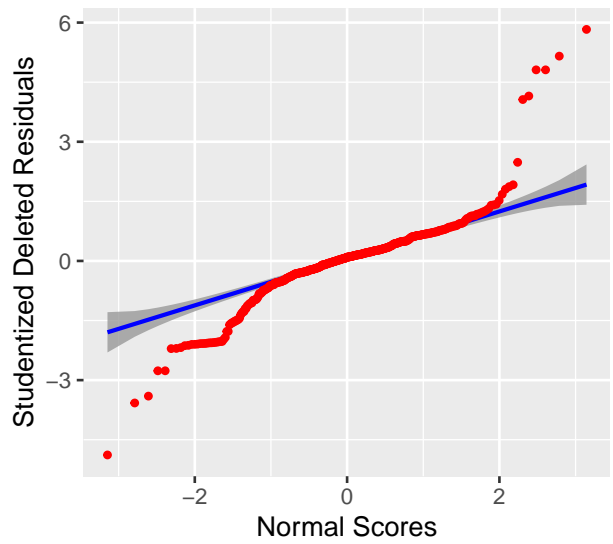
Box-Cox Transformation



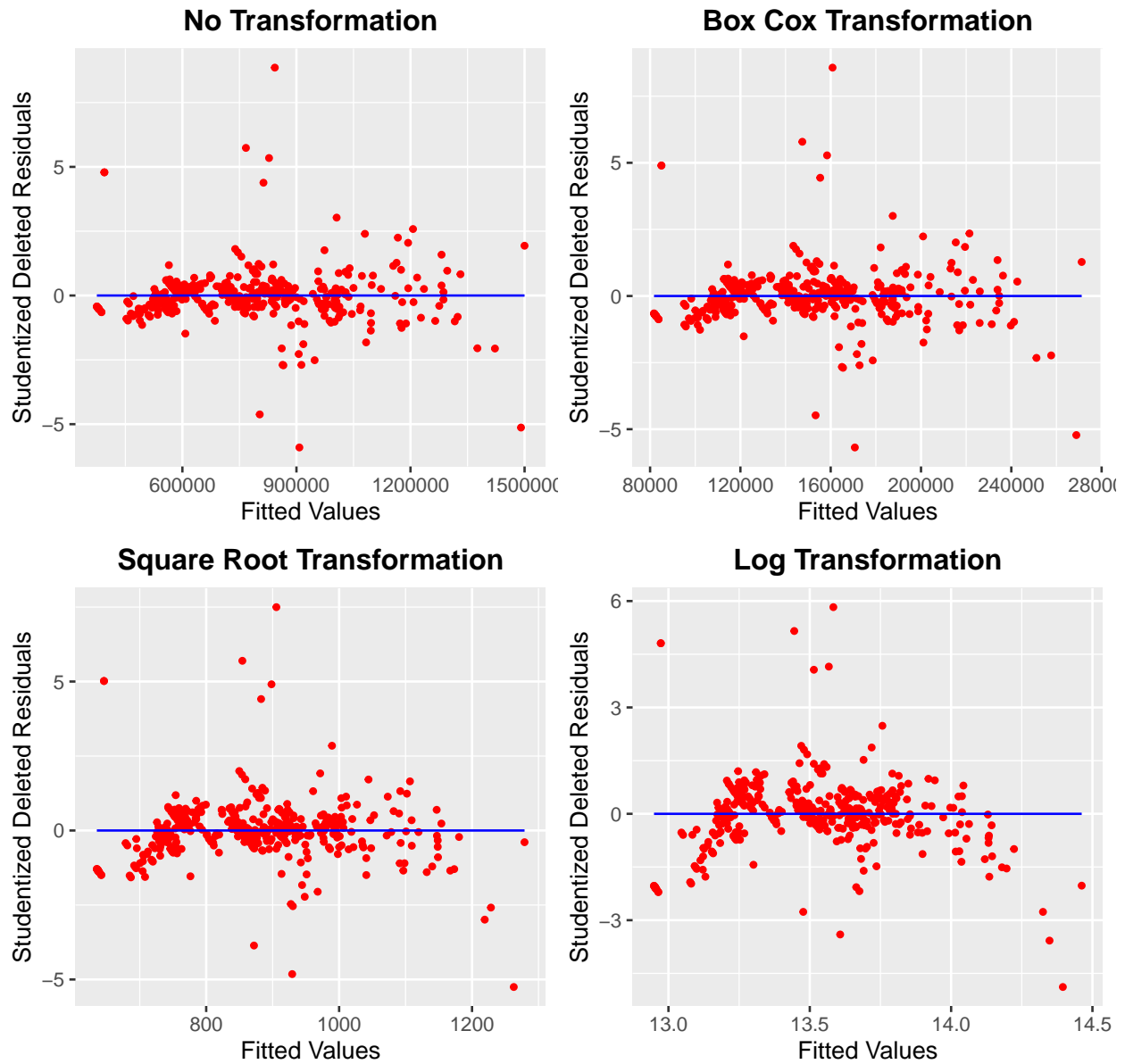
Square Root Transformation



Log Transformation



Studentized Deleted Residuals vs. Fitted Values (Full Model)



From the above results, it is clear the square root and log transformations are too extreme causing further skewness. The box cox transformation seems most appropriate bringing more symmetry about the mean; however, with $\lambda = 0.869$ the resulting transformation does not differ greatly from the original data. Thus, no transformation will be performed on the response for the full model. A breakdown of the full model shows the following:

Call:

```
lm(formula = value ~ built + factor(type) + beds + baths + GLA +  
    sold + price, data = housing_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-292332	-23807	-1057	15928	423097

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.628e+06	3.631e+06	-0.724	0.46966
built	4.861e+03	1.829e+03	2.658	0.00815 **
factor(type)condo_res	-1.143e+05	1.494e+04	-7.648	1.37e-13 ***
factor(type)end_gar_TH	-6.332e+04	1.082e+04	-5.851	9.74e-09 ***
factor(type)int_gar_TH	-8.007e+04	1.019e+04	-7.861	3.14e-14 ***
factor(type)rambler	2.847e+05	6.226e+04	4.573	6.31e-06 ***
beds	1.159e+04	5.175e+03	2.240	0.02562 *
baths	1.527e+04	5.733e+03	2.664	0.00801 **
GLA	5.139e+01	6.579e+00	7.811	4.44e-14 ***
sold	-3.420e+03	1.333e+03	-2.565	0.01065 *
price	5.000e-01	2.379e-02	21.020	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 55570 on 427 degrees of freedom
Multiple R-squared: 0.9414, Adjusted R-squared: 0.94
F-statistic: 685.8 on 10 and 427 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: value

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
built	1	3.0904e+10	3.0904e+10	10.0089	0.001669 **
factor(type)	4	1.7076e+13	4.2690e+12	1382.5809	< 2.2e-16 ***
beds	1	1.2413e+12	1.2413e+12	402.0292	< 2.2e-16 ***
baths	1	5.7916e+11	5.7916e+11	187.5720	< 2.2e-16 ***
GLA	1	8.6415e+11	8.6415e+11	279.8697	< 2.2e-16 ***
sold	1	1.9162e+10	1.9162e+10	6.2059	0.013111 *
price	1	1.3643e+12	1.3643e+12	441.8373	< 2.2e-16 ***
Residuals	427	1.3184e+12	3.0877e+09		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

With a multiple R-squared of 0.94, the full model indicates strong relationships between the regressors and response. This means that 94% of the variability in the projected sales price is explained by the relationship between the property's total value and the given regressors. Therefore, the regressors in the model have very large explanatory power and the variance not explained by the model is minimal.

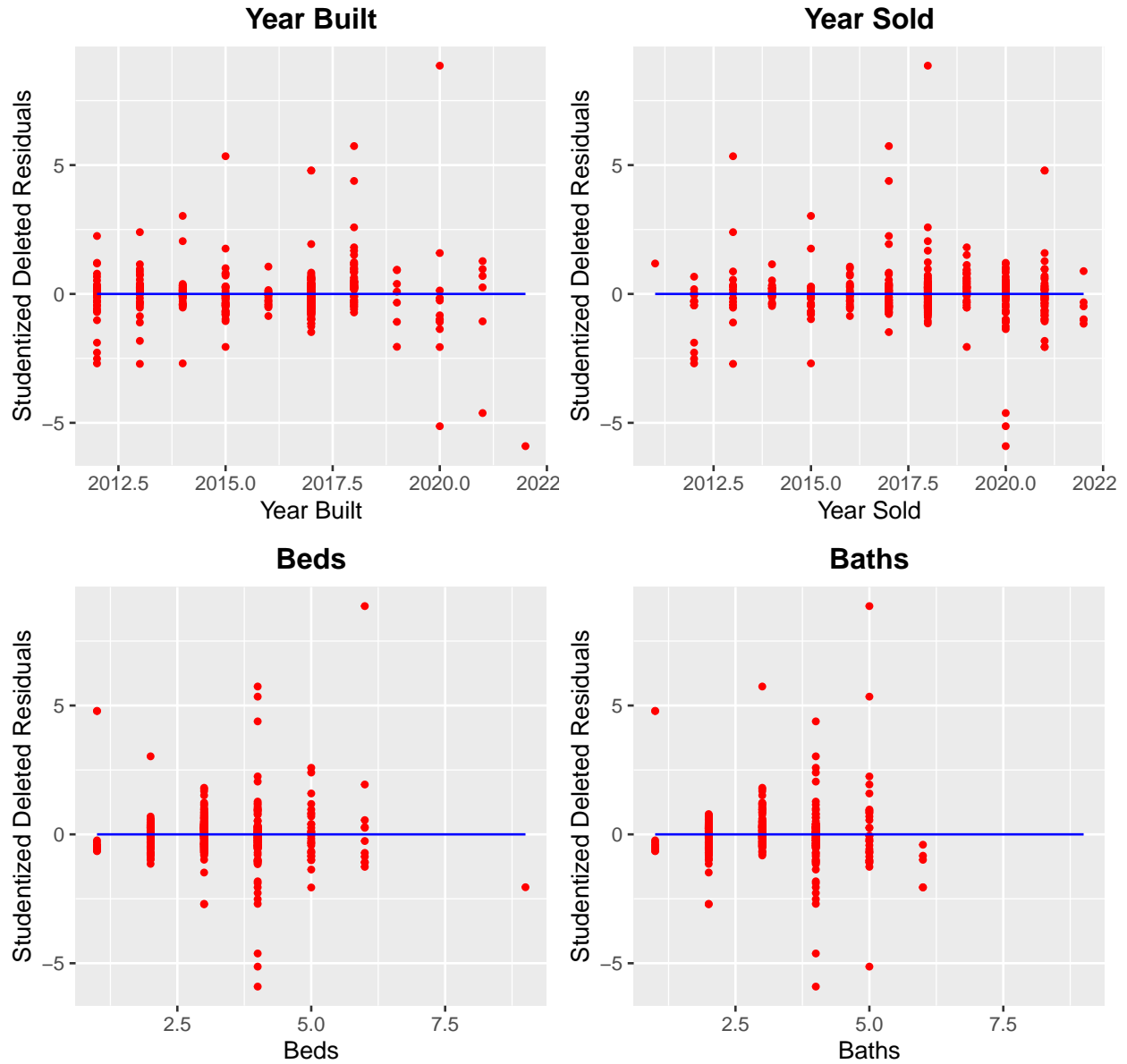
The overall F-statistic reported by the model is 658.8 on 10 and 427 degrees of freedom, resulting in a p-value extremely close to zero. With such a large F-statistic it can be determined that something within the model is significant when comparing the joint effects of all the regressors. Further analysis on the impact of each individual regressor against the full model are broken down in an ANOVA table.

In the ANOVA table, it is shown that the property type, number of bedrooms, number of bathrooms, gross leasing area, and most recent sales price all have very large F-statistics which result in p-values extremely close to zero. Therefore, the bulk of the explanatory power for the model lies within these regressors as they are significant at the $\alpha = 0.001$ level. The years built and sold have less impact on the model and yield p-values of 0.0017 and 0.013, respectively. The year sold is significant at the $\alpha = 0.05$ level, but is the weakest regressor in terms of explanatory power. Lastly, the year built is just shy of being significant at the $\alpha = 0.001$ level.

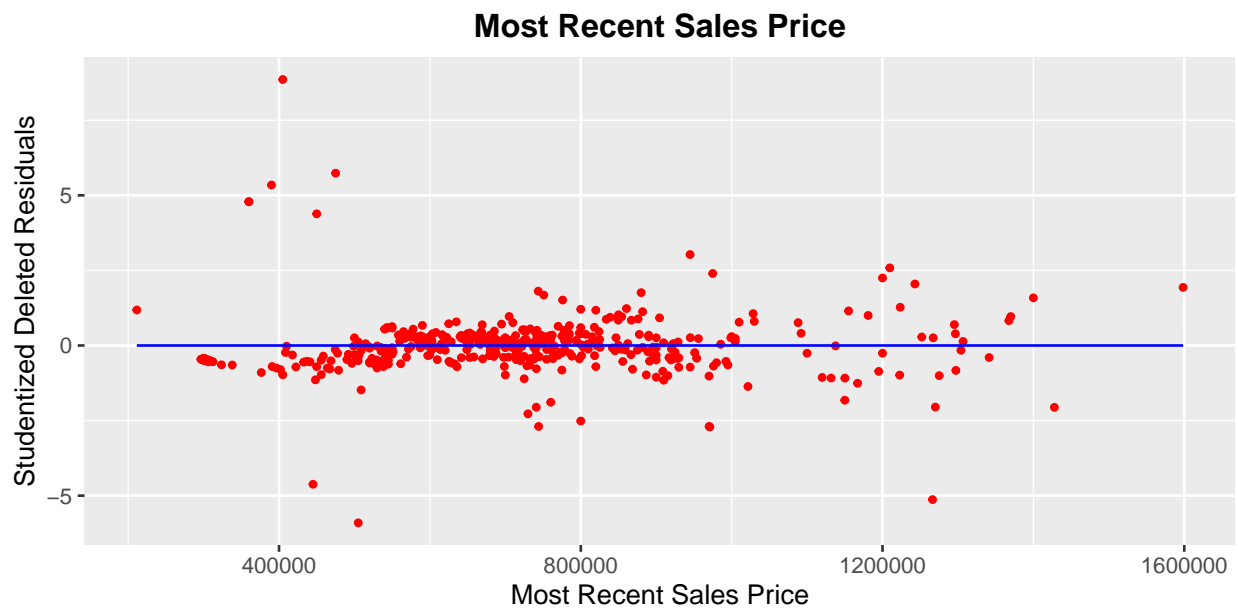
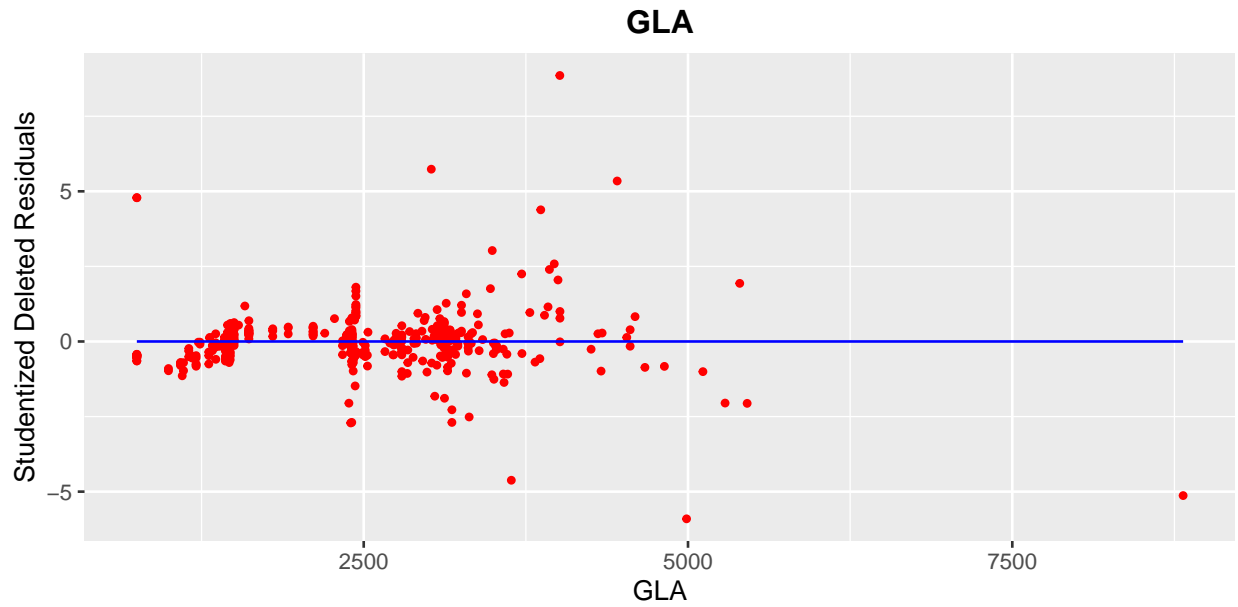
This was a very surprising realization. It was expected that the year the property sold would have large impact on the most recent sales price and consequently the total value. Since this regressor is marked as not significant at the $\alpha = 0.001$ and $\alpha = 0.01$ levels, it follows that the year sold and sales price for that year are likely correlated.

Examining the normal probability plot of the studentized deleted residuals shows a reasonably linear pattern near the centroid. Conversely, the tails (especially the higher tail), show departures from the line-of-best-fit. Since transforming the response showed minimal signs of improving the normality of the model, plots of each regressor against the studentized deleted residuals were generated to look for additional patterns.

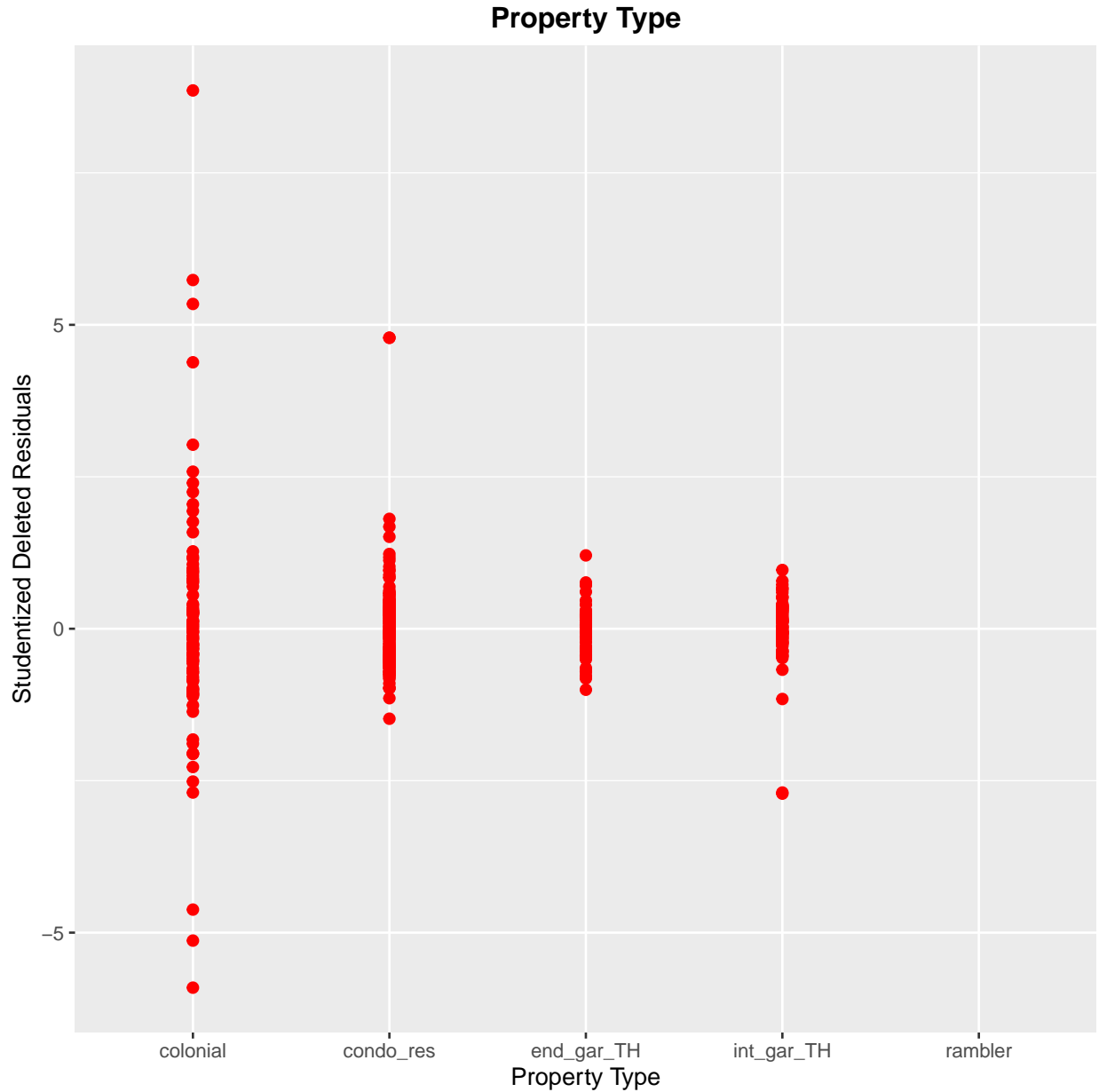
Studentized Deleted Residuals vs. Individual Regressors (Full Model)



The residuals for year built and year sold seem fairly distributed from the given plots with larger variance observed in recent years (especially 2020). Properties containing between 2 and 3 beds or baths show minimal variance; conversely, properties with 4 or more beds or baths have a much large residual range.



Smaller properties tend to receive more accurate predictions; on the other hand, cheaper properties (near \$400,000) show very high variance.



Lastly, a look into the property type indicates the observations with greatest variance occur when dealing with colonial homes; condominiums are predicted with better accuracy; town homes have the most accurate predictions; and, additional sample data for rambler homes is required to perform effective inference.

To confirm this deduction, Levene's test was performed which resulted in the following:

```
Levene's Test for Homogeneity of Variance (center = median)
  Df F value  Pr(>F)
group  4  12.817 7.202e-10 ***
433
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Levene's test was used over other analogous tests for homogeneity of variance due to suspicion the data may not be normal. The p -value is significant at the $\alpha = 0.001$ level which confirms the variance between property types is uneven.

Since a transformation on the response did not necessarily improve the normality assumption necessary for OLS, a closer look was taken on potential problem children introduced into the dataset.

We note that $p = 11$ on a model with 10 regressors and 438 observations; therefore, the cutoff for leverage is $\frac{2p}{n} = 0.05023$. Also, the cutoff for DFFITS is $2\sqrt{\frac{p}{n}} = 0.3169$. One observation was found with a Cook's distance greater than 1. Using these cutoffs, the following dataset indices were flagged:

	Leverage
3	1.00000000
26	0.12681720
28	0.06346666
30	0.05120455
31	0.05450066
32	0.06046404
37	0.06618422
58	0.10681701
185	0.09269680
186	0.06182024
206	0.06429308
211	0.05701622
213	0.13019705
348	0.05415703
349	0.05540425
351	0.06954637
353	0.05192556
354	0.06476778
432	0.14332443
433	0.11081273
434	0.34478147

	DFFITS
18	0.4090563
22	-0.4290812
23	-0.4684872
24	-0.3987750
25	-0.3259601
26	3.3751860
27	0.3909653
28	0.7882131
30	0.3686524
57	-0.3496392
58	0.4084742
185	-1.4769639
206	0.5072034
213	-0.7933473
268	-0.6111893
274	-0.3938737
286	-0.5505248
348	1.3727871
351	1.1984981
353	0.5262258
355	0.3819228
426	0.7076403
427	0.7076403
432	-2.4153457

433 1.8863424
 434 -3.7220491
 438 -0.4673406

Cook's D

434 1.188903

A subset of the data including observations yielding values greater than the above predetermined thresholds also show the following patterns:

[1] "Observations with Leverage"

[1] "3" "26" "28" "30" "31" "32" "37" "58" "185" "186" "206" "211"
 [13] "213" "348" "349" "351" "353" "354" "432" "433" "434"

[1] "Observations with DFFITS"

[1] "18" "22" "23" "24" "25" "26" "27" "28" "30" "57" "58" "185"
 [13] "206" "213" "268" "274" "286" "348" "351" "353" "355" "426" "427" "432"
 [25] "433" "434" "438"

[1] "Observations with Cook's D"

[1] "434"

	i..parcel_id	type	value	beds	baths	GLA	price	built	sold
3	47 3 07 092	rambler	1493800	8	9	5866	775000	2021	2019
18	47 4 15 003	colonial	1347500	5	4	3969	1209660	2018	2018
22	48 3 19 003	colonial	811000	4	4	3313	800045	2012	2012
23	48 3 19 007	colonial	766600	4	4	3180	744135	2012	2012
24	48 3 19 009	colonial	782100	4	4	3180	730000	2012	2012
25	48 3 19 015	colonial	815000	4	4	3123	760255	2012	2012
26	57 1 02 069	colonial	1266200	6	5	4012	405000	2020	2018
27	57 1 02 076	colonial	1210900	5	4	3932	974900	2013	2013
28	57 1 04 082	colonial	1167000	2	4	3491	944900	2014	2015
30	57 1 10 092	colonial	1367400	5	5	3291	1400000	2020	2021
31	57 1 14 069	colonial	987800	6	5	3384	732990	2013	2013
32	57 1 14 072	colonial	988000	5	5	2955	995000	2012	2021
37	57 1 15 D 005	colonial	945900	4	5	2988	970000	2012	2021
57	57 1 21 084 A	colonial	749200	4	4	2386	741000	2015	2019
58	57 1 23 048	colonial	626500	5	3	1584	211500	2012	2011
185	57 2 03 B 019 B	colonial	564500	4	4	3638	445000	2021	2020
186	57 2 45 005	colonial	1018600	6	4	4338	780000	2012	2015
206	57 2 49 001	colonial	1603600	6	5	5399	1598500	2017	2017
211	57 2 51 002	colonial	1255000	5	5	2966	1295234	2021	2021
213	57 3 02 004	colonial	1270100	9	6	5286	1270000	2019	2021
268	57 3 20 017	int_gar_TH	719700	3	2	2400	971069	2013	2013
274	57 3 20 023	colonial	984600	4	4	3048	1150000	2013	2021
286	57 3 21 004	int_gar_TH	718000	3	2	2408	970113	2014	2015
348	58 1 02 014	colonial	1066200	4	3	3021	475000	2018	2017
349	58 1 02 017 A	colonial	1180900	6	5	4668	1195000	2013	2021
351	58 1 07 002	colonial	1043400	4	4	3865	450000	2018	2017
353	58 1 27 013	colonial	1287900	4	5	3718	1200000	2012	2017
354	58 1 27 016	colonial	1254100	4	6	3720	1341336	2016	2016
355	58 1 27 050	colonial	1305200	4	4	3998	1243000	2014	2018
426	58 2 11 02 404	condo_res	651500	1	1	751	360000	2017	2021
427	58 2 11 02 405	condo_res	651500	1	1	751	360000	2017	2021
432	58 3 04 024	colonial	615400	4	4	4989	505000	2022	2020

433	58 3 04 035	colonial	1099300	4	5 4454 390000	2015 2013
434	58 3 21 001	colonial	1266400	4	5 8817 1266301	2020 2020
438	58 3 21 005	colonial	1310900	5	6 5456 1427841	2020 2021

The Rambler home has ridiculously high leverage relative to the other problem children; however, this is likely due to the nature of the property type when compared to other single-family residences. This Rambler has the highest value, beds, and baths out of all the problem children so this leverage is expected. It is also the only observation for the Rambler property type factor level; however, the corresponding DFFITS and Cook's D values for this Rambler are below the threshold.

In terms of DFFITS, observation #26 rose high above the rest with a DFFITS=3.375. This is due to the extreme price jump from most recent sales price to current value. According to the data provided by Fairfax County, the property was sold as a new construction in 2018 for \$405,000, then built in 2020, and within 2 years has tripled in value. From here, it may be prudent to refer to Fairfax County on the accuracy of reported data for this specific property.

For Cook's distance, observation #434 is the only property yielding a value above the threshold. The property also has a large negative DFFITS value. This is due to the extremely high reported GLA of 8,817 square feet. No other properties in this list seem to exceed a GLA of 6,000. This is another situation where referring to the origin of the data could be necessary.

In summary, 35/438 observations were flagged and should be brought before a subject matter expert or data maintainer for further review.

3.2 Reduced Model

To improve upon the full model, an all possible regression model selection technique yielded the following results:

	n	predictors	adjr	cp
	7	price	0.8393525930	733.613715
	5	GLA	0.8105543225	942.924629
	4	baths	0.7542150120	1352.408684
	2	factor(type)	0.7161195971	1615.095915
	3	beds	0.7163153355	1627.870223
	6	sold	0.0024398544	6816.443249
	1	built	-0.0009164862	6840.837724
	18	factor(type) price	0.9187663568	153.003759
	27	GLA price	0.9103661158	217.980686
	25	baths price	0.9008751562	286.804441
	22	beds price	0.8937824241	338.237432
	16	factor(type) GLA	0.8531383909	625.623293
	23	baths GLA	0.8531841396	632.636153
	28	sold price	0.8513837850	645.691436
	20	beds GLA	0.8433599167	703.876568
	13	built price	0.8398860931	729.067023
	15	factor(type) baths	0.8199587602	864.566273
	11	built GLA	0.8143093218	914.537143
	26	GLA sold	0.8119140843	931.906224
	19	beds baths	0.7959220919	1047.872259
	14	factor(type) beds	0.7906435888	1075.679362
	8	built factor(type)	0.7577567225	1312.513999
	10	built baths	0.7587367582	1317.521944
	24	baths sold	0.7543276949	1349.494296
	17	factor(type) sold	0.7315395039	1501.317195
	9	built beds	0.7173675805	1617.510802
	21	beds sold	0.7156654226	1629.854011
	12	built sold	0.0001696732	6818.276025
	52	factor(type) GLA price	0.9361024228	29.092967
	50	factor(type) baths price	0.9277201820	89.317908
	47	factor(type) beds price	0.9246433373	111.424501
	61	baths GLA price	0.9241543253	118.731039
	33	built factor(type) price	0.9212608493	135.727088
	58	beds GLA price	0.9208899035	142.348596
	53	factor(type) sold price	0.9191645322	150.788762
	63	GLA sold price	0.9116903452	208.905895
	42	built GLA price	0.9105231350	217.350469
	56	beds baths price	0.9103149612	218.856573
	62	baths sold price	0.9067997864	244.288287
	40	built baths price	0.9029621363	272.053057
	59	beds sold price	0.8984761019	304.508782
	37	built beds price	0.8935385622	340.231073
	48	factor(type) baths GLA	0.8717287137	491.606859
	45	factor(type) beds GLA	0.8643896225	544.336952
	31	built factor(type) GLA	0.8607842024	570.241267
	54	beds baths GLA	0.8607330142	577.573841
	51	factor(type) GLA sold	0.8582501833	588.447753
	60	baths GLA sold	0.8533167822	631.229066
	38	built baths GLA	0.8530351013	633.266981

43	3	built sold price	0.8517473274	642.583814
35	3	built beds GLA	0.8471691658	675.706064
57	3	beds GLA sold	0.8443196332	696.321966
44	3	factor(type) beds baths	0.8319929903	777.101113
30	3	built factor(type) baths	0.8263932119	817.334546
49	3	factor(type) baths sold	0.8232143416	840.174176
29	3	built factor(type) beds	0.8131499820	912.484847
41	3	built GLA sold	0.8140635211	915.219983
46	3	factor(type) beds sold	0.7966612882	1030.953241
34	3	built beds baths	0.7959770882	1046.072365
55	3	beds baths sold	0.7955522149	1049.146254
32	3	built factor(type) sold	0.7588366206	1302.716892
39	3	built baths sold	0.7582701441	1318.875932
36	3	built beds sold	0.7171389915	1616.453087
91	4	factor(type) baths GLA price	0.9385165707	12.722798
88	4	factor(type) beds GLA price	0.9375167379	19.889755
72	4	built factor(type) GLA price	0.9369039317	24.282445
93	4	factor(type) GLA sold price	0.9362026710	29.309191
86	4	factor(type) beds baths price	0.9292524287	79.129611
92	4	factor(type) baths sold price	0.9283324539	85.724133
70	4	built factor(type) baths price	0.9282752297	86.134326
67	4	built factor(type) beds price	0.9266892870	97.502610
95	4	beds baths GLA price	0.9269384736	99.370237
98	4	baths GLA sold price	0.9257702956	107.802340
89	4	factor(type) beds sold price	0.9251612420	108.455874
81	4	built baths GLA price	0.9240242288	120.405740
73	4	built factor(type) sold price	0.9232396465	122.230170
97	4	beds GLA sold price	0.9219733245	135.209508
78	4	built beds GLA price	0.9212391336	140.509020
96	4	beds baths sold price	0.9145797352	188.577664
83	4	built GLA sold price	0.9131965618	198.561640
76	4	built beds baths price	0.9107887447	215.941663
82	4	built baths sold price	0.9066998110	245.456266
79	4	built beds sold price	0.8997009453	295.975242
84	4	factor(type) beds baths GLA	0.8747306648	469.950107
68	4	built factor(type) baths GLA	0.8742618169	473.310882
90	4	factor(type) baths GLA sold	0.8742526527	473.376572
65	4	built factor(type) beds GLA	0.8709541811	497.020530
87	4	factor(type) beds GLA sold	0.8678649685	519.164488
71	4	built factor(type) GLA sold	0.8620670521	560.724855
74	4	built beds baths GLA	0.8611662680	574.124945
94	4	beds baths GLA sold	0.8609676301	575.558746
80	4	built baths GLA sold	0.8529956284	633.101979
77	4	built beds GLA sold	0.8468613475	677.380238
64	4	built factor(type) beds baths	0.8388828444	726.912867
85	4	factor(type) beds baths sold	0.8344279466	758.846269
69	4	built factor(type) baths sold	0.8269040893	812.778450
66	4	built factor(type) beds sold	0.8130715667	911.932127
75	4	built beds baths sold	0.7955070227	1048.064287
115	5	factor(type) beds baths GLA price	0.9389104448	10.881076
106	5	built factor(type) baths GLA price	0.9387608682	11.950771
118	5	factor(type) baths GLA sold price	0.9387423689	12.083069
103	5	built factor(type) beds GLA price	0.9382911416	15.310015
117	5	factor(type) beds GLA sold price	0.9376855096	19.641183

108	5	built factor(type) GLA sold price	0.9376497950	19.896595
101	5	built factor(type) beds baths price	0.9299582096	74.902858
116	5	factor(type) beds baths sold price	0.9298962783	75.345758
107	5	built factor(type) baths sold price	0.9297761704	76.204709
104	5	built factor(type) beds sold price	0.9287147282	83.795597
119	5	beds baths GLA sold price	0.9283304022	90.128324
110	5	built beds baths GLA price	0.9267740005	101.336744
113	5	built baths GLA sold price	0.9258034976	108.325817
112	5	built beds GLA sold price	0.9237280005	123.272501
111	5	built beds baths sold price	0.9144319985	190.217623
99	5	built factor(type) beds baths GLA	0.8776417641	449.043166
114	5	factor(type) beds baths GLA sold	0.8769090401	454.283231
105	5	built factor(type) baths GLA sold	0.8751883301	466.588864
102	5	built factor(type) beds GLA sold	0.8715411644	492.671517
109	5	built beds baths GLA sold	0.8609943206	575.048847
100	5	built factor(type) beds baths sold	0.8389399092	725.818926
124	6	built factor(type) baths GLA sold price	0.9394495105	8.016572
121	6	built factor(type) beds baths GLA price	0.9392303201	9.580454
126	6	factor(type) beds baths GLA sold price	0.9391621973	10.066498
123	6	built factor(type) beds GLA sold price	0.9391576521	10.098927
122	6	built factor(type) beds baths sold price	0.9316009149	64.014853
125	6	built beds baths GLA sold price	0.9286814503	88.411362
120	6	built factor(type) beds baths GLA sold	0.8782256865	444.837260
127	7	built factor(type) beds baths GLA sold price	0.9400124643	5.000000

This list was further reduced under the criterion that Mallows's $C_p < 30$:

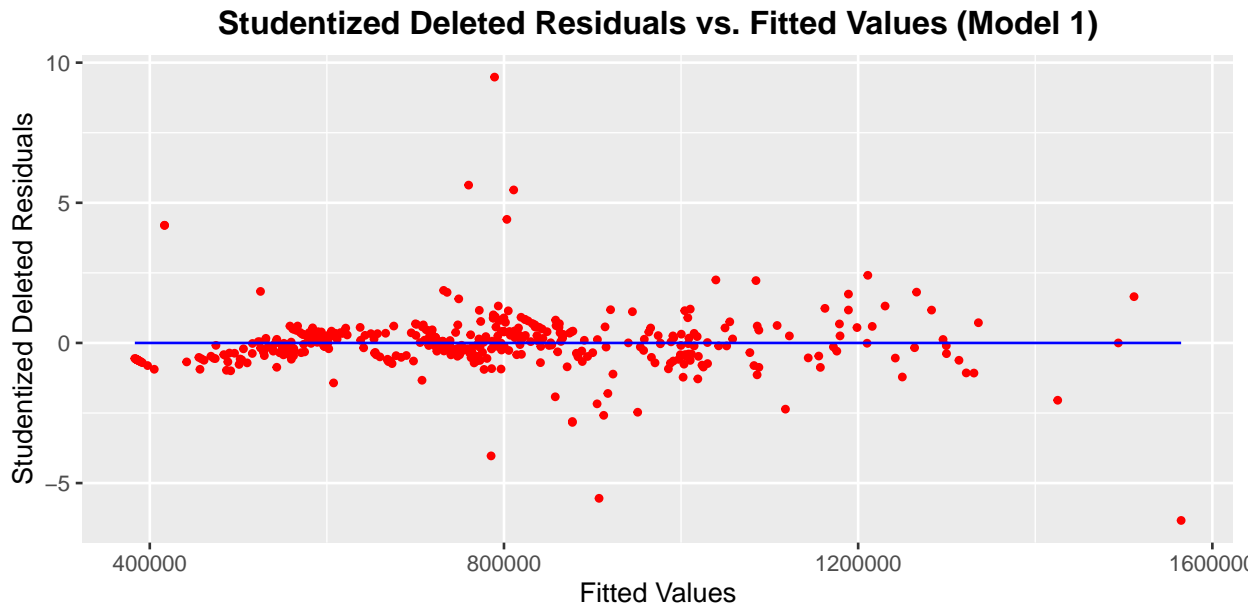
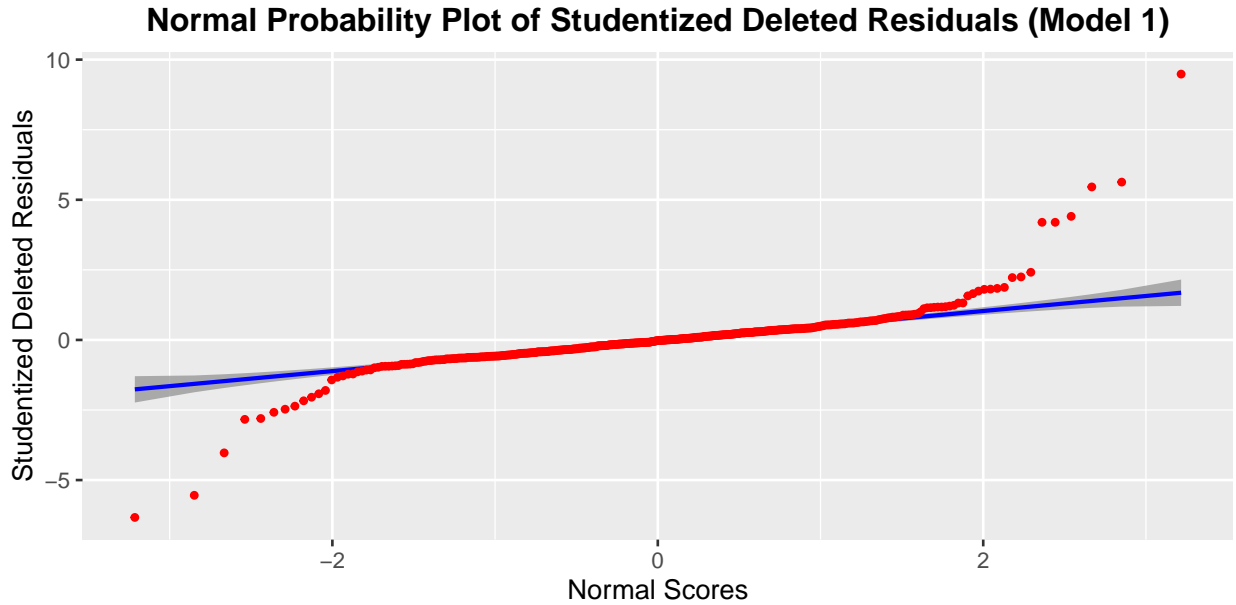
n		predictors	adjr	cp
1	3	factor(type) GLA price	0.9361024	29.092967
2	4	factor(type) baths GLA price	0.9385166	12.722798
3	4	factor(type) beds GLA price	0.9375167	19.889755
4	4	built factor(type) GLA price	0.9369039	24.282445
5	4	factor(type) GLA sold price	0.9362027	29.309191
6	5	factor(type) beds baths GLA price	0.9389104	10.881076
7	5	built factor(type) baths GLA price	0.9387609	11.950771
8	5	factor(type) baths GLA sold price	0.9387424	12.083069
9	5	built factor(type) beds GLA price	0.9382911	15.310015
10	5	factor(type) beds GLA sold price	0.9376855	19.641183
11	5	built factor(type) GLA sold price	0.9376498	19.896595
12	6	built factor(type) baths GLA sold price	0.9394495	8.016572
13	6	built factor(type) beds baths GLA price	0.9392303	9.580454
14	6	factor(type) beds baths GLA sold price	0.9391622	10.066498
15	6	built factor(type) beds GLA sold price	0.9391577	10.098927
16	7	built factor(type) beds baths GLA sold price	0.9400125	5.000000

From this reduced list, the full model yields the lowest Mallows's C_p value and is recommended as a highly accurate model; however, due to potential overfitting other models with the following regressors should also be considered:

1. property type, GLA, most recent sales price
2. property type, beds, baths, GLA, most recent sales price
3. year built, property type, baths, GLA, year sold, most recent sales price

3.2.1 Reduced Model #1

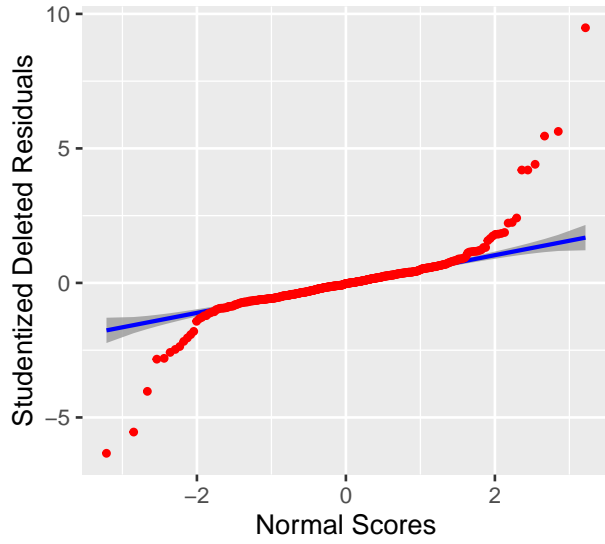
The reduced model with only property type, GLA, and most recent sales price used for regressors is the cheapest model with minimal loss of precision. This model is recommended due to ease of data collection and maintenance. With less regressors there is less overhead. A normal probability plot continues to show a greater-than-normal distribution with no clear patterns in the plot of studentized deleted residuals versus fitted values.



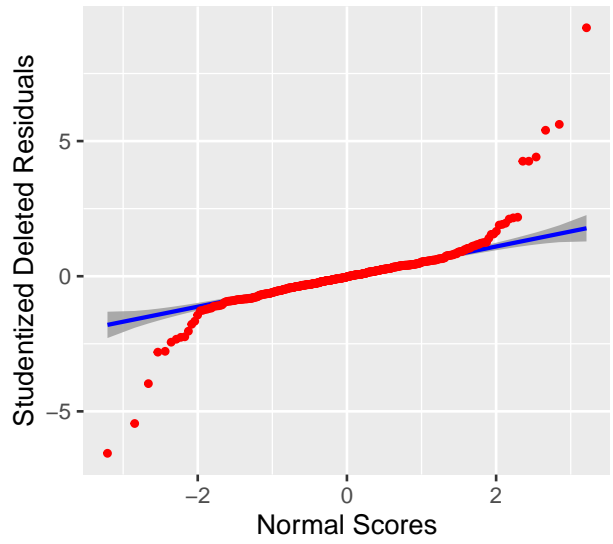
Applicable transformations were reexamined and the same conclusions drawn as with the full model.

Normal Probability Plot of Studentized Deleted Residuals (Model 1)

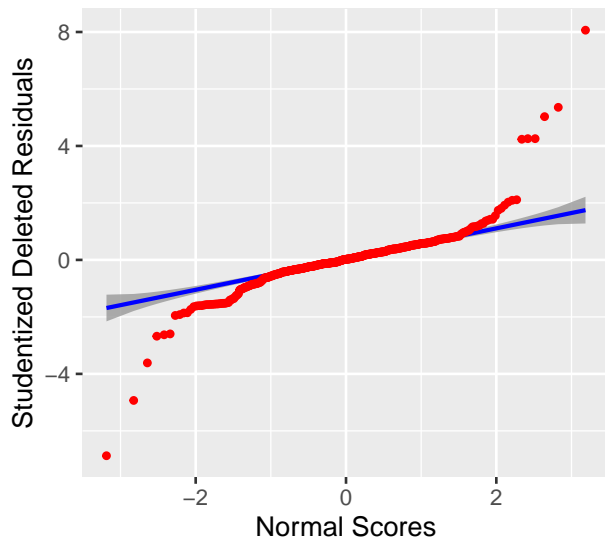
No Transformation



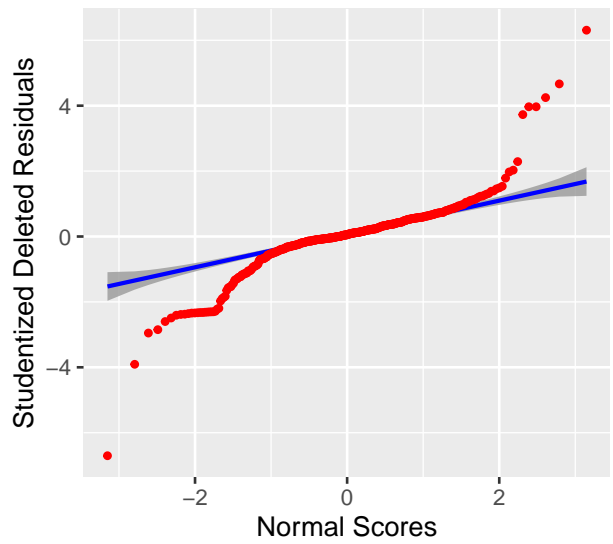
Box-Cox Transformation



Square Root Transformation



Log Transformation



Studentized Deleted Residuals vs. Fitted Values (Model 1)



Call:

```
lm(formula = value ~ factor(type) + GLA + price, data = housing_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-298359	-24019	-1461	19487	476833

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.076e+05	2.066e+04	14.890	< 2e-16 ***
factor(type)condo_res	-1.311e+05	1.247e+04	-10.510	< 2e-16 ***
factor(type)end_gar_TH	-8.871e+04	1.010e+04	-8.781	< 2e-16 ***
factor(type)int_gar_TH	-1.030e+05	9.522e+03	-10.815	< 2e-16 ***
factor(type)rambler	3.854e+05	6.010e+04	6.412	3.77e-10 ***

```

GLA                6.683e+01  6.146e+00  10.872  < 2e-16 ***
price              5.274e-01  2.225e-02  23.705  < 2e-16 ***

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 57350 on 431 degrees of freedom

Multiple R-squared: 0.937, Adjusted R-squared: 0.9361

F-statistic: 1068 on 6 and 431 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: value

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(type)	4	1.6166e+13	4.0416e+12	1228.8	< 2.2e-16 ***
GLA	1	3.0613e+12	3.0613e+12	930.8	< 2.2e-16 ***
price	1	1.8481e+12	1.8481e+12	561.9	< 2.2e-16 ***
Residuals	431	1.4175e+12	3.2889e+09		

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

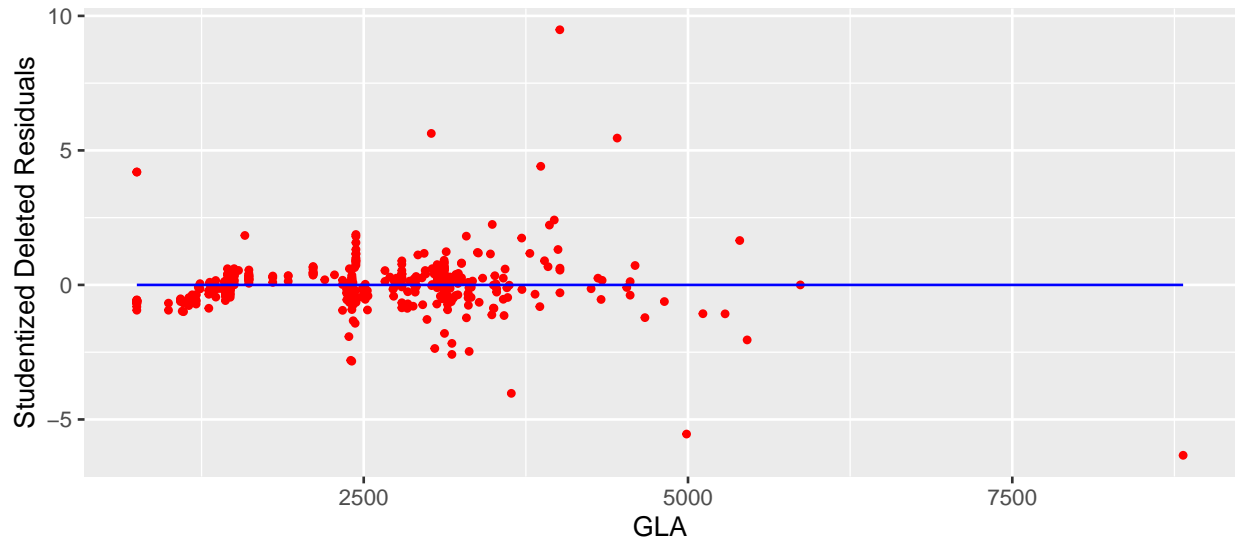
The regressors in the model continue to have large explanatory power and the variance not explained by the model is minimal. The adjusted R^2 value slightly dropped; however, this is expected behavior whenever regressors are removed from the model.

The overall F-statistic reported by the model is 1068 on 6 and 431 degrees of freedom yielding a zero-like p-value. Further analysis on the impact of each individual regressor against the full model are broken down in an ANOVA table.

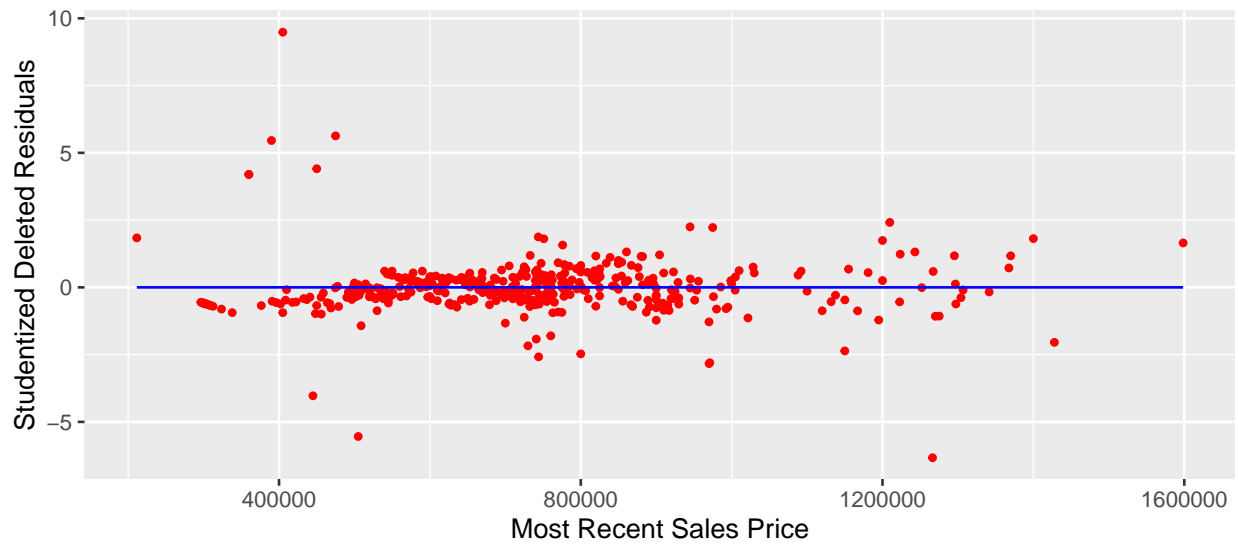
The ANOVA table reports each regressor as highly significant at the $\alpha = 0.001$ level. In order of significance, the regressors are property type, GLA, and most recent sales price. Plots of each regressor against the studentized deleted residuals were generated to look for new patterns but trends from the full model persisted in the reduced model.

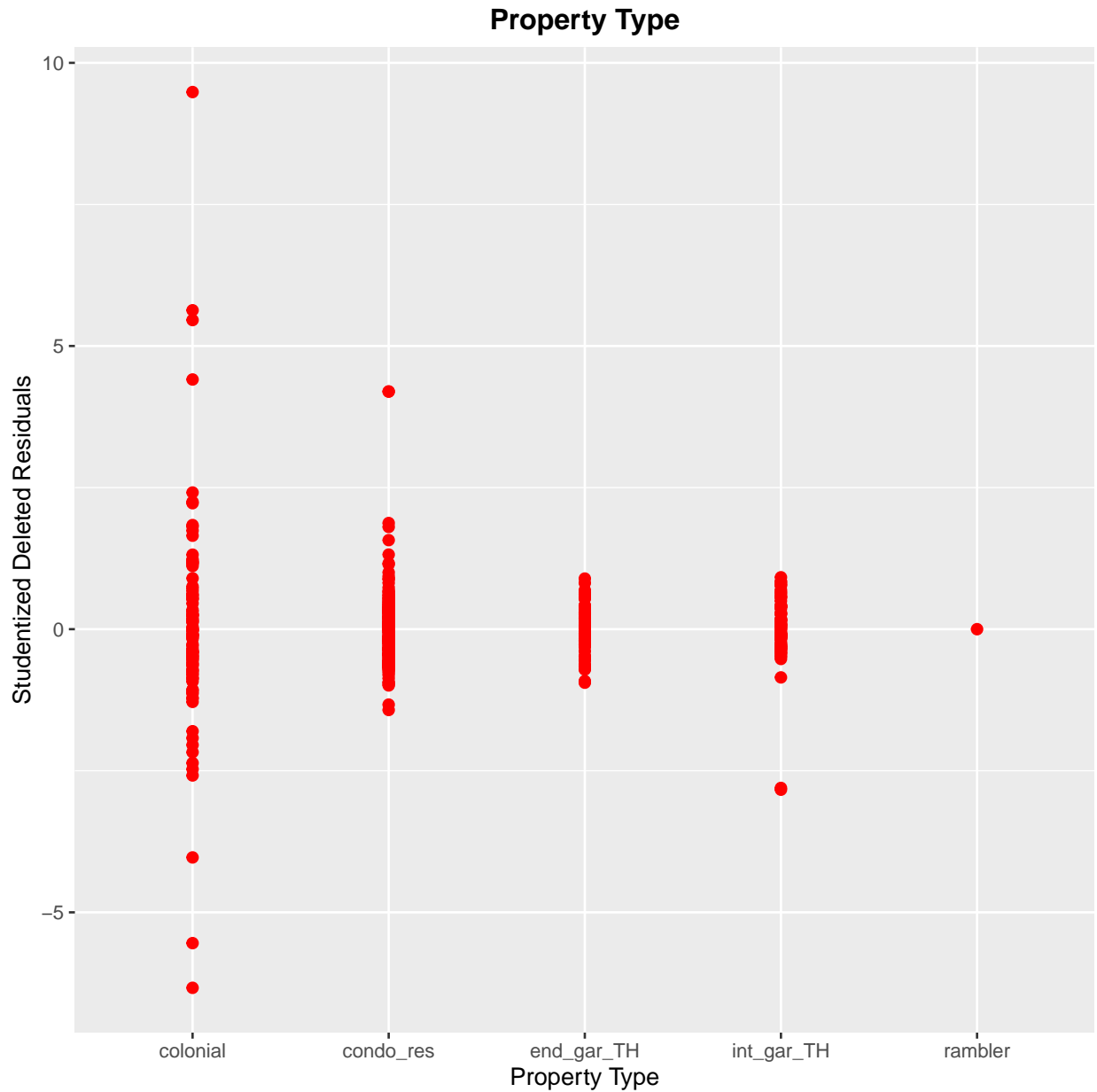
Studentized Deleted Residuals vs. Individual Regressors (Model 1)

GLA



Most Recent Sales Price





For the problem children, we note that $p = 7$ on a model with 6 regressors and 438 observations; therefore, the cutoff for leverage is $\frac{2p}{n} = 0.03196$. Also, the cutoff for DFFITS is $2\sqrt{\frac{p}{n}} = 0.2528$. Observation #434 continues to show a large Cook's distance. The following dataset indices were flagged with these thresholds:

	Leverage
3	1.00000000
26	0.07279979
30	0.04363700
58	0.06884443
185	0.05353023
206	0.05108823
211	0.03882104
348	0.03724880
351	0.05944380
432	0.09815537

```

433 0.09499510
434 0.26381163
438 0.03973992

```

```

      DFFITS
18  0.2999610
22 -0.2810603
23 -0.3183890
24 -0.2744867
26  2.6574855
30  0.3872171
57 -0.2781761
58  0.4998286
185 -0.9584047
206 0.3830054
268 -0.4419230
274 -0.3536297
286 -0.4446464
348 1.1075722
351 1.1083228
426 0.4455879
427 0.4455879
432 -1.8287975
433 1.7683589
434 -3.7907641
438 -0.4158740

```

```

      Cook's D
434 1.882098

```

A subset of the data including observations yielding values greater than the above predetermined thresholds also show the following patterns:

```
[1] "Observations with Leverage"
```

```
[1] "3" "26" "30" "58" "185" "206" "211" "348" "351" "432" "433" "434"
[13] "438"
```

```
[1] "Observations with DFFITS"
```

```
[1] "18" "22" "23" "24" "26" "30" "57" "58" "185" "206" "268" "274"
[13] "286" "348" "351" "426" "427" "432" "433" "434" "438"
```

```
[1] "Observations with Cook's D"
```

```
[1] "434"
```

	i..parcel_id	type	value	beds	baths	GLA	price	built	sold
3	47 3 07 092	rambler	1493800	8	9	5866	775000	2021	2019
18	47 4 15 003	colonial	1347500	5	4	3969	1209660	2018	2018
22	48 3 19 003	colonial	811000	4	4	3313	800045	2012	2012
23	48 3 19 007	colonial	766600	4	4	3180	744135	2012	2012
24	48 3 19 009	colonial	782100	4	4	3180	730000	2012	2012
26	57 1 02 069	colonial	1266200	6	5	4012	405000	2020	2018
30	57 1 10 092	colonial	1367400	5	5	3291	1400000	2020	2021
57	57 1 21 084 A	colonial	749200	4	4	2386	741000	2015	2019
58	57 1 23 048	colonial	626500	5	3	1584	211500	2012	2011
185	57 2 03 B 019 B	colonial	564500	4	4	3638	445000	2021	2020
206	57 2 49 001	colonial	1603600	6	5	5399	1598500	2017	2017

211	57	2	51	002	colonial	1255000	5	5	2966	1295234	2021	2021	
268	57	3	20	017	int_gar_TH	719700	3	2	2400	971069	2013	2013	
274	57	3	20	023	colonial	984600	4	4	3048	1150000	2013	2021	
286	57	3	21	004	int_gar_TH	718000	3	2	2408	970113	2014	2015	
348	58	1	02	014	colonial	1066200	4	3	3021	475000	2018	2017	
351	58	1	07	002	colonial	1043400	4	4	3865	450000	2018	2017	
426	58	2	11	02	404	condo_res	651500	1	1	751	360000	2017	2021
427	58	2	11	02	405	condo_res	651500	1	1	751	360000	2017	2021
432	58	3	04	024	colonial	615400	4	4	4989	505000	2022	2020	
433	58	3	04	035	colonial	1099300	4	5	4454	390000	2015	2013	
434	58	3	21	001	colonial	1266400	4	5	8817	1266301	2020	2020	
438	58	3	21	005	colonial	1310900	5	6	5456	1427841	2020	2021	

Some problem children persisted in the reduced model. Observations #3, #26, and #434 have the greatest potential to negatively impact the model. The rambler home has high leverage due to being the only observation for its factor level. The two colonial homes stray far from the mean in terms of change in price and GLA.

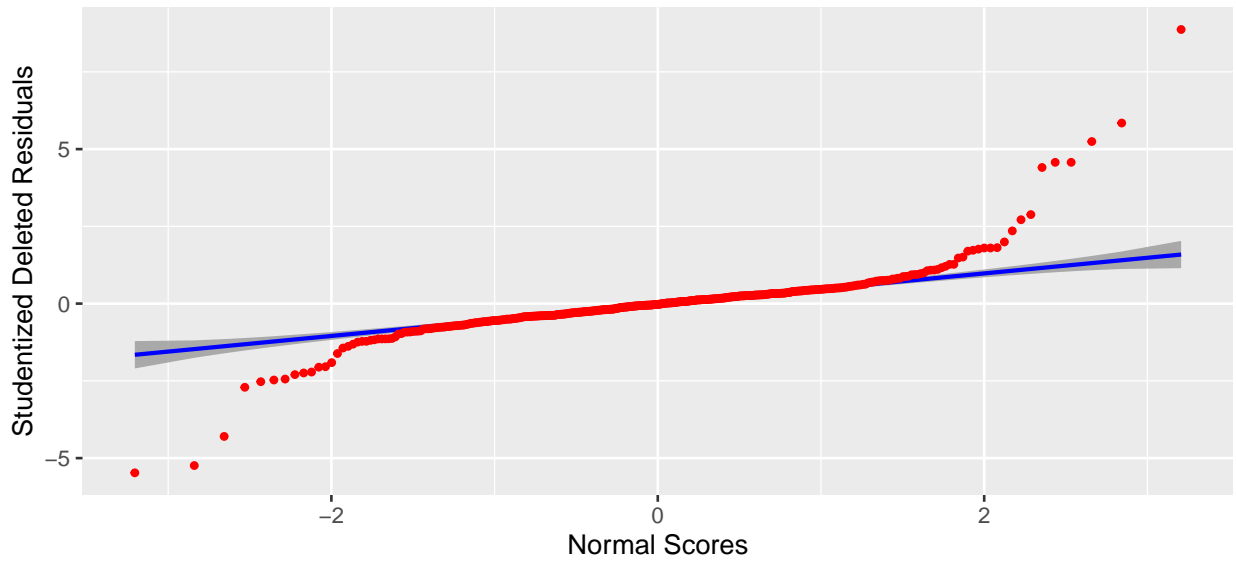
In summary, 23/438 observations were found to be above the predetermined thresholds. This means that 12 problem children are eliminated when using Model #1.

3.2.2 Reduced Model #2

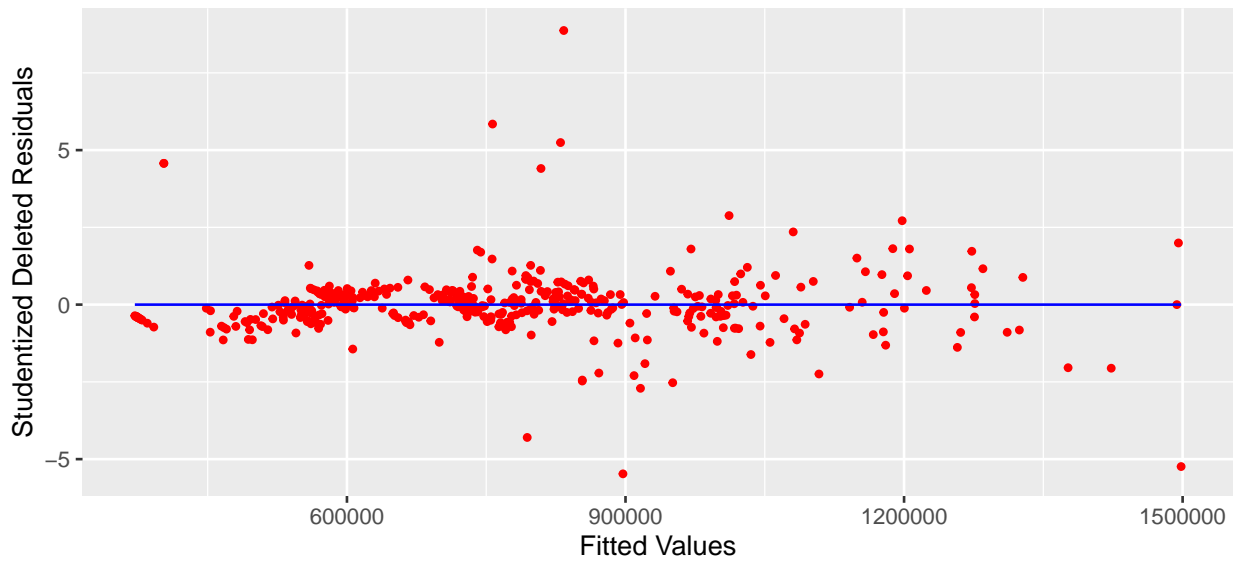
The reduced model with only property type, GLA, most recent sales price, beds, and baths used for regressors is similar to Model #1 with the addition of room counts. This model is still relatively cheap; however, additional overhead is required to ensure the number of beds and baths are reported accurately. A normal probability plot continues to show a potentially greater-than-normal distribution with no clear patterns in the plot of studentized deleted residuals versus fitted values.

Applicable transformations were reexamined and the same conclusions drawn as with the full model.

Normal Probability Plot of Studentized Deleted Residuals (Model 2)

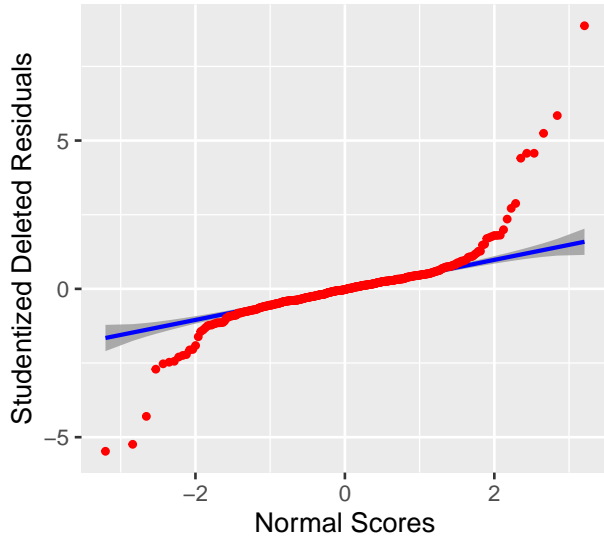


Studentized Deleted Residuals vs. Fitted Values (Model 2)

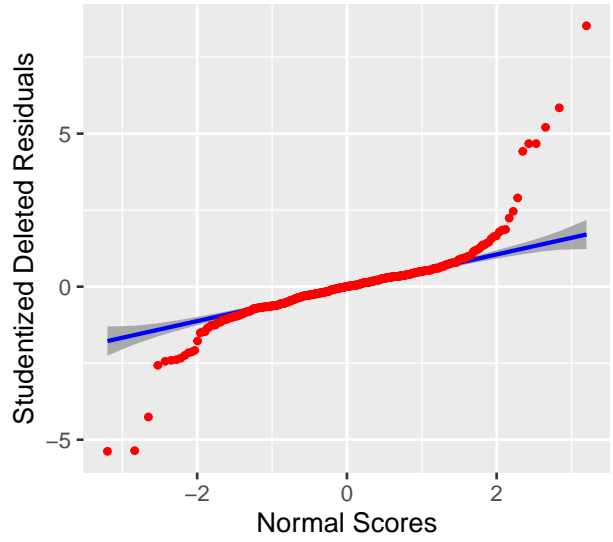


Normal Probability Plot of Studentized Deleted Residuals (Model 2)

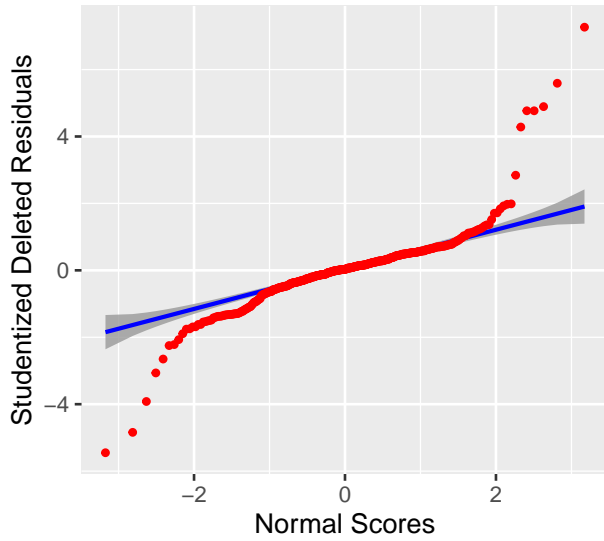
No Transformation



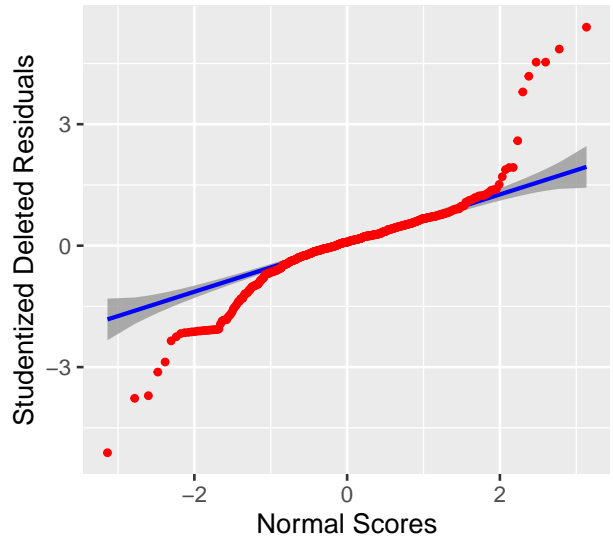
Box-Cox Transformation



Square Root Transformation

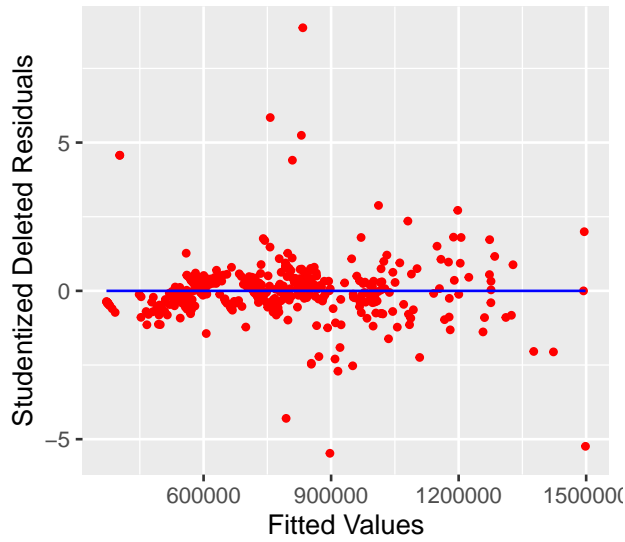


Log Transformation

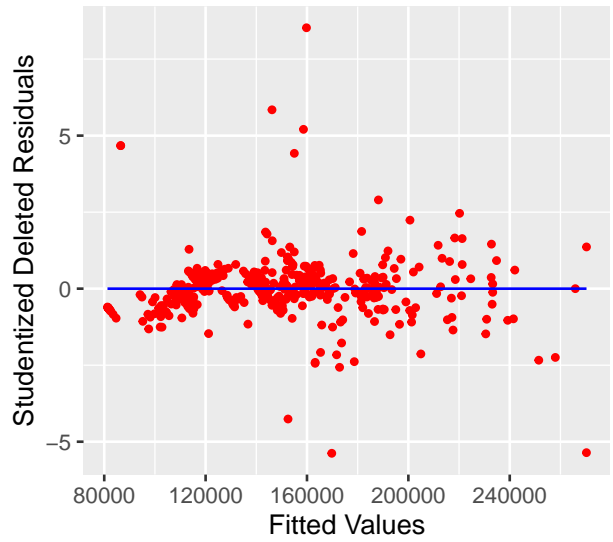


Studentized Deleted Residuals vs. Fitted Values (Model 2)

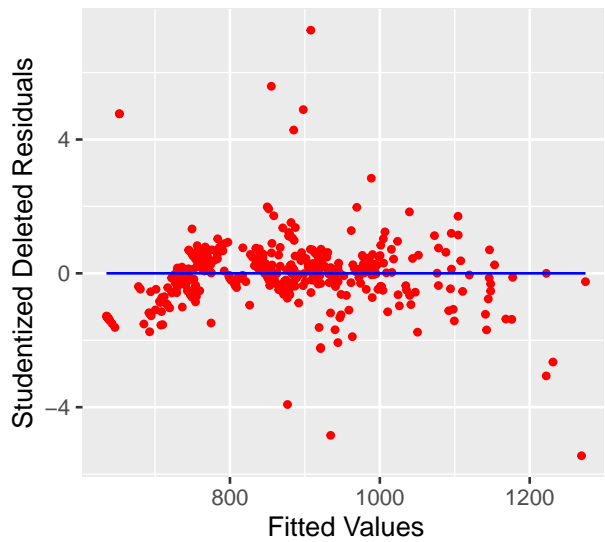
No Transformation



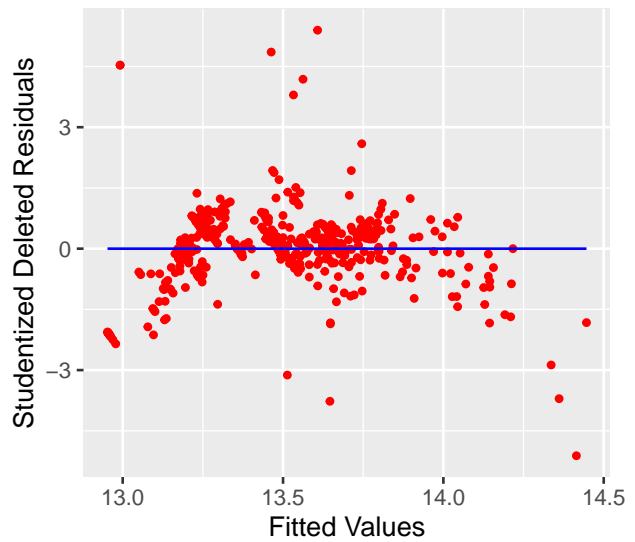
Box Cox Transformation



Square Root Transformation



Log Transformation



Call:

```
lm(formula = value ~ factor(type) + beds + baths + GLA + price,
    data = housing_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-281865	-21826	-1413	18022	432692

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.630e+05	2.275e+04	11.563	< 2e-16 ***
factor(type)condo_res	-1.066e+05	1.337e+04	-7.971	1.44e-14 ***
factor(type)end_gar_TH	-6.884e+04	1.076e+04	-6.399	4.10e-10 ***
factor(type)int_gar_TH	-8.468e+04	1.011e+04	-8.373	8.02e-16 ***

```

factor(type)rambler      2.835e+05  6.274e+04  4.519 8.05e-06 ***
beds                    1.010e+04  5.199e+03  1.942 0.05276 .
baths                  1.844e+04  5.608e+03  3.288 0.00109 **
GLA                    5.451e+01  6.563e+00  8.305 1.31e-15 ***
price                  4.913e-01  2.309e-02 21.278 < 2e-16 ***

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 56070 on 429 degrees of freedom

Multiple R-squared: 0.94, Adjusted R-squared: 0.9389

F-statistic: 840.6 on 8 and 429 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: value

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
factor(type)	4	1.6166e+13	4.0416e+12	1285.33	< 2.2e-16	***
beds	1	1.6717e+12	1.6717e+12	531.65	< 2.2e-16	***
baths	1	9.2809e+11	9.2809e+11	295.16	< 2.2e-16	***
GLA	1	9.5456e+11	9.5456e+11	303.57	< 2.2e-16	***
price	1	1.4236e+12	1.4236e+12	452.75	< 2.2e-16	***
Residuals	429	1.3489e+12	3.1444e+09			

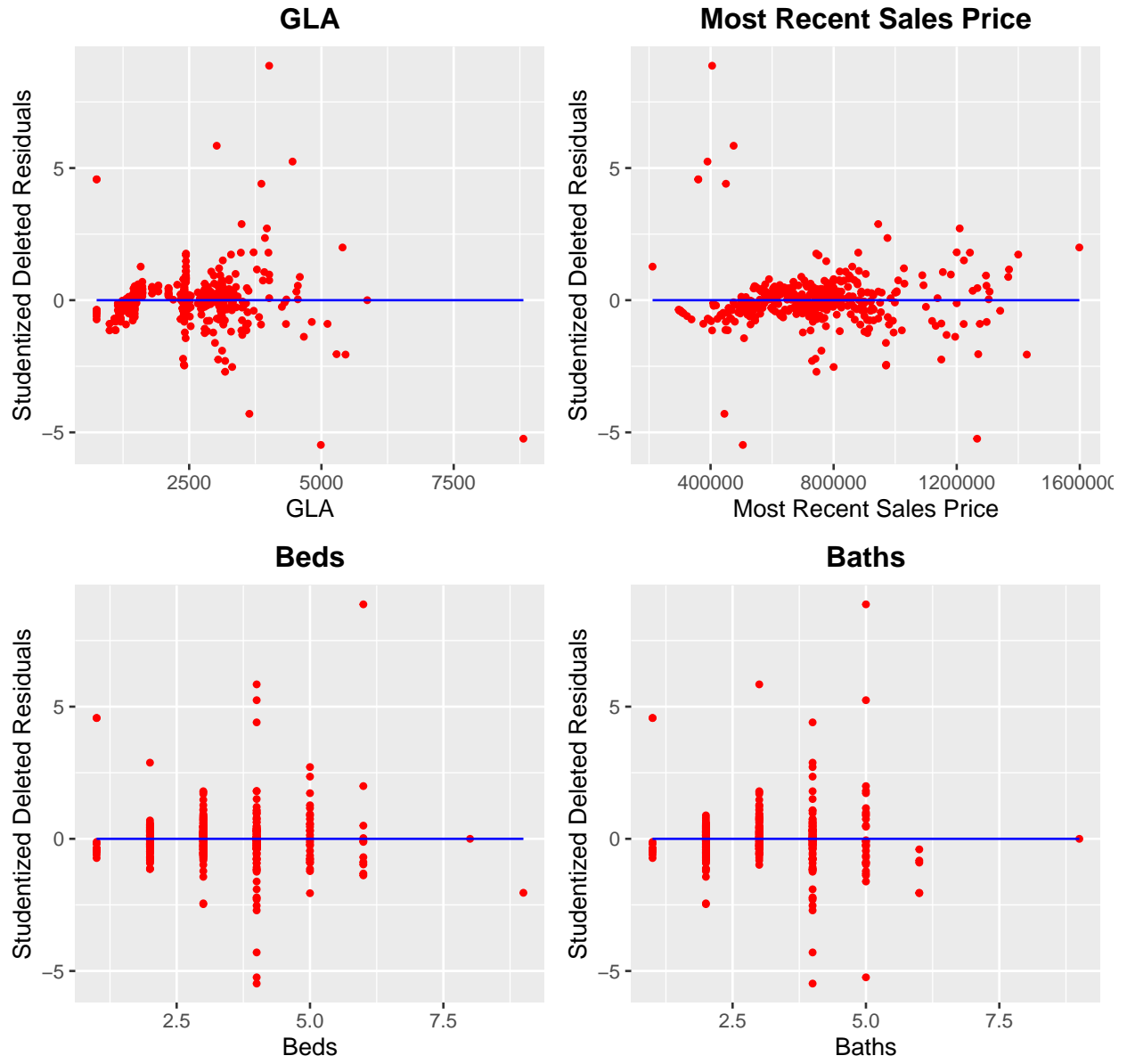
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

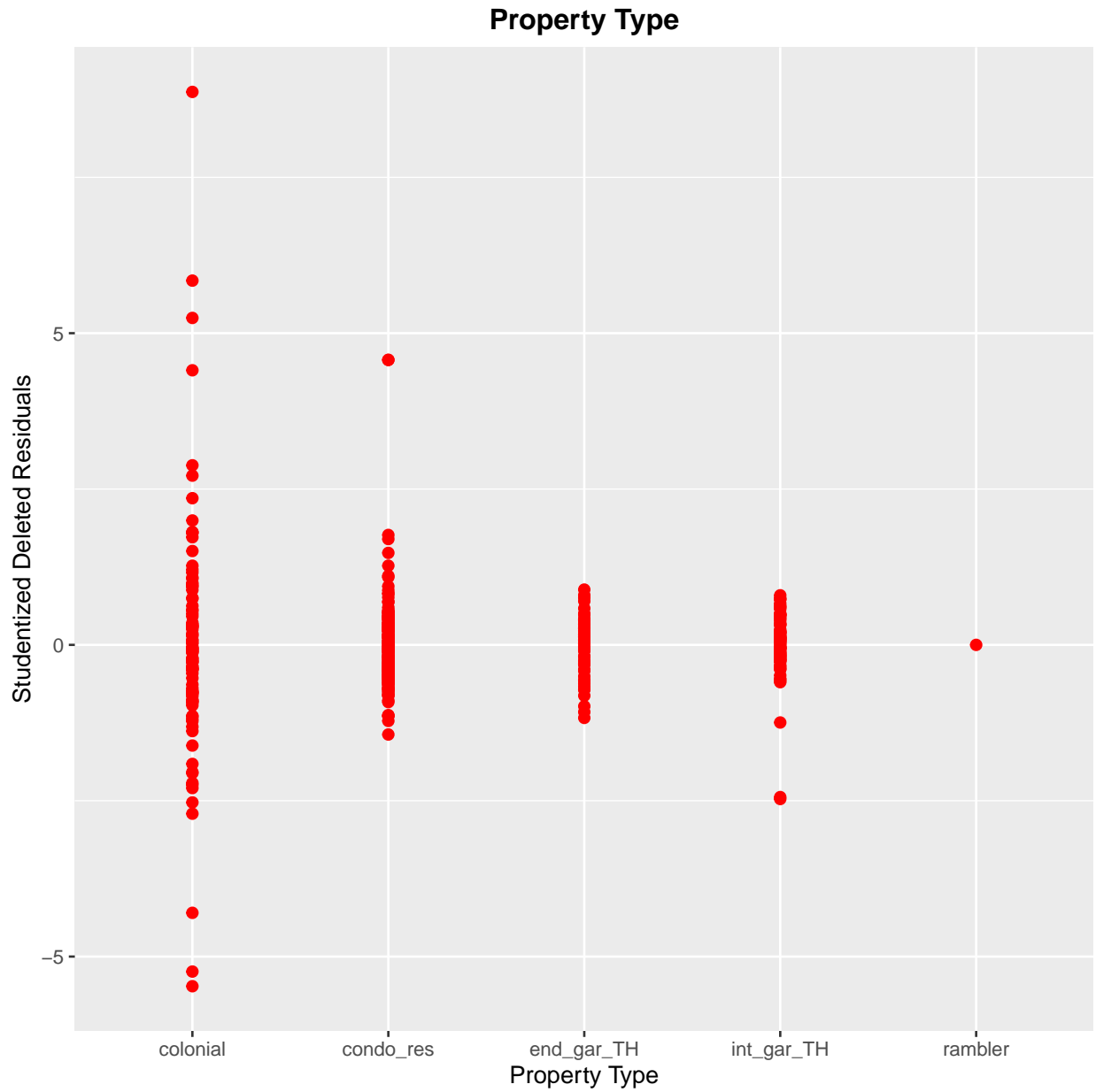
The regressors in the model continue to have large explanatory power and the variance not explained by the model is minimal. Property type, GLA, and most recent sales price are still significant at the $\alpha = 0.001$ level; however, beds and baths are less significant at the $\alpha = 0.1$ and $\alpha = 0.01$ levels, respectively.

The overall F-statistic reported by the model is 840.6 on 8 and 429 degrees of freedom yielding a zero-like p-value. Further analysis on the impact of each individual regressor against the full model are broken down in an ANOVA table.

The ANOVA table reports each regressor as highly significant at the $\alpha = 0.001$ level. In order of significance, the regressors are property type, beds, most recent sales price, GLA, and baths. Plots of each regressor against the studentized deleted residuals were generated to look for new patterns but trends from the full model persisted in the reduced model.

Studentized Deleted Residuals vs. Individual Regressors (Model 2)





For the problem children, we note that $p = 9$ on a model with 8 regressors and 438 observations; therefore, the cutoff for leverage is $\frac{2p}{n} = 0.04110$. Also, the cutoff for DFFITS is $2\sqrt{\frac{p}{n}} = 0.2867$. Observation #434 continues to show a large Cook's distance. The following dataset indices were flagged with these thresholds:

	Leverage
3	1.00000000
26	0.10626724
28	0.05820663
30	0.04444007
31	0.04523449
58	0.10295953
185	0.05529856
186	0.05061873
206	0.06014246
211	0.04209175

```
213 0.12707721
351 0.06060204
354 0.05708886
432 0.10040127
433 0.10990035
434 0.33812842
438 0.04671726
```

```
          DFFITS
2      0.3084117
18     0.4001916
22    -0.2941323
23    -0.3425882
24    -0.2979877
26     3.0587397
27     0.2889830
28     0.7161683
30     0.3723121
57    -0.3449834
58     0.4301905
185   -1.0399520
206     0.5046749
213   -0.7790985
268   -0.4607949
274   -0.3502825
286   -0.4652646
348     1.1850992
351     1.1188090
353     0.2897643
355     0.2964186
426     0.5645736
427     0.5645736
432   -1.8293179
433     1.8428377
434   -3.7456222
438   -0.4554443
```

```
      Cook's D
434 1.468284
```

A subset of the data including observations yielding values greater than the above predetermined thresholds also show the following patterns:

```
[1] "Observations with Leverage"
[1] "3"  "26" "28" "30" "31" "58" "185" "186" "206" "211" "213" "351"
[13] "354" "432" "433" "434" "438"
[1] "Observations with DFFITS"
[1] "2"  "18" "22" "23" "24" "26" "27" "28" "30" "57" "58" "185"
[13] "206" "213" "268" "274" "286" "348" "351" "353" "355" "426" "427" "432"
[25] "433" "434" "438"
[1] "Observations with Cook's D"
[1] "434"
```

```
i..parcel_id      type  value beds baths  GLA  price built sold
```


2	47	3	07	017	colonial	1069800	3	3	3477	880000	2015	2015
3	47	3	07	092	rambler	1493800	8	9	5866	775000	2021	2019
18	47	4	15	003	colonial	1347500	5	4	3969	1209660	2018	2018
22	48	3	19	003	colonial	811000	4	4	3313	800045	2012	2012
23	48	3	19	007	colonial	766600	4	4	3180	744135	2012	2012
24	48	3	19	009	colonial	782100	4	4	3180	730000	2012	2012
26	57	1	02	069	colonial	1266200	6	5	4012	405000	2020	2018
27	57	1	02	076	colonial	1210900	5	4	3932	974900	2013	2013
28	57	1	04	082	colonial	1167000	2	4	3491	944900	2014	2015
30	57	1	10	092	colonial	1367400	5	5	3291	1400000	2020	2021
31	57	1	14	069	colonial	987800	6	5	3384	732990	2013	2013
57	57	1	21	084 A	colonial	749200	4	4	2386	741000	2015	2019
58	57	1	23	048	colonial	626500	5	3	1584	211500	2012	2011
185	57	2	03	B 019 B	colonial	564500	4	4	3638	445000	2021	2020
186	57	2	45	005	colonial	1018600	6	4	4338	780000	2012	2015
206	57	2	49	001	colonial	1603600	6	5	5399	1598500	2017	2017
211	57	2	51	002	colonial	1255000	5	5	2966	1295234	2021	2021
213	57	3	02	004	colonial	1270100	9	6	5286	1270000	2019	2021
268	57	3	20	017	int_gar_TH	719700	3	2	2400	971069	2013	2013
274	57	3	20	023	colonial	984600	4	4	3048	1150000	2013	2021
286	57	3	21	004	int_gar_TH	718000	3	2	2408	970113	2014	2015
348	58	1	02	014	colonial	1066200	4	3	3021	475000	2018	2017
351	58	1	07	002	colonial	1043400	4	4	3865	450000	2018	2017
353	58	1	27	013	colonial	1287900	4	5	3718	1200000	2012	2017
354	58	1	27	016	colonial	1254100	4	6	3720	1341336	2016	2016
355	58	1	27	050	colonial	1305200	4	4	3998	1243000	2014	2018
426	58	2	11	02 404	condo_res	651500	1	1	751	360000	2017	2021
432	58	3	04	024	colonial	615400	4	4	4989	505000	2022	2020
433	58	3	04	035	colonial	1099300	4	5	4454	390000	2015	2013
434	58	3	21	001	colonial	1266400	4	5	8817	1266301	2020	2020
438	58	3	21	005	colonial	1310900	5	6	5456	1427841	2020	2021

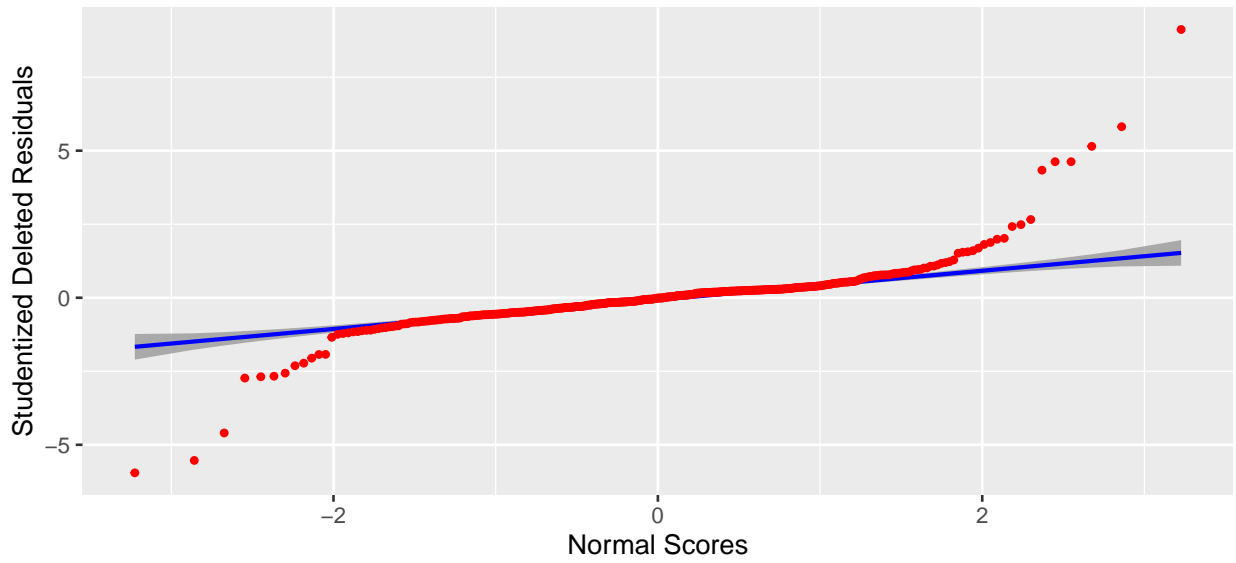
Observations #3, #26, and #434 are still the biggest red flags for the model. Additional sample data on rambler home is required for accurate predictions on other rambles. The price jump and GLA for observations #26 and #434 also require closer examination from a subject matter expert or Fairfax County official.

In summary, 31/438 observations were found to be above the predetermined thresholds. There is a greater count of problem children for this model when compared to Model #1; conversely, there are less problem children when compared to the full model.

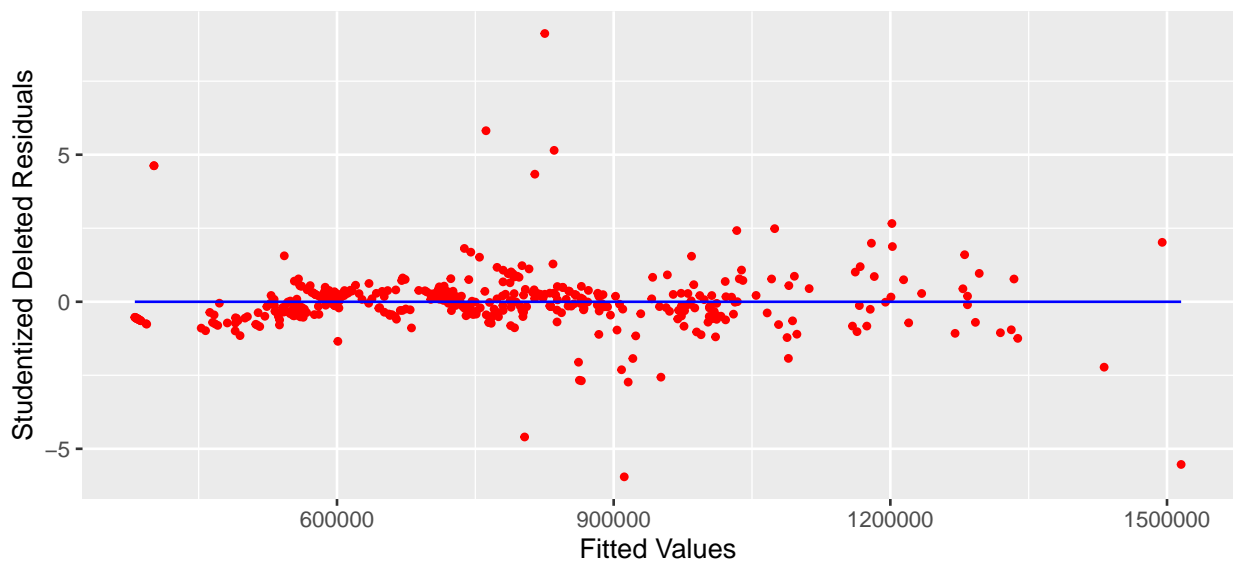
3.2.3 Reduced Model #3

The reduced model without beds as a regressor is another excellent choice for model selection due to it having the second lowest Mallows' Cp score of 8.0166 (losing only to the full model). Subsequently, there is additional overhead required when reporting the years the property was sold and built. From data collection, a number of observations had to be removed due to inconsistencies between the year built and year sold. When including these regressors in the model, it is crucial to validate these data when flags are raised. A normal probability plot continues to show a greater-than-normal distribution with no clear patterns in the plot of studentized deleted residuals versus fitted values.

Normal Probability Plot of Studentized Deleted Residuals (Model 3)



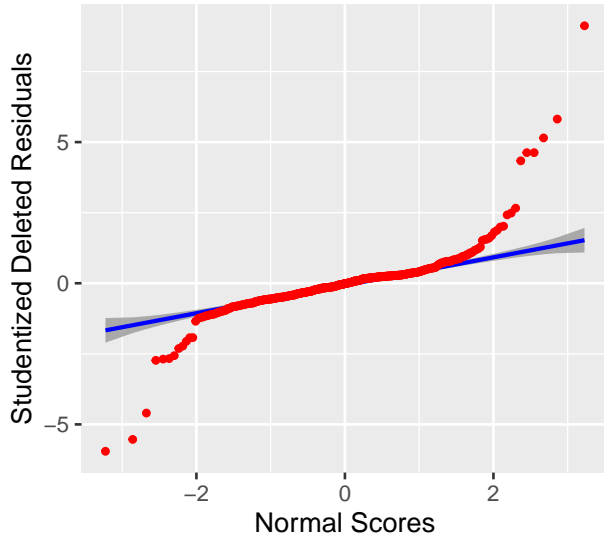
Studentized Deleted Residuals vs. Fitted Values (Model 3)



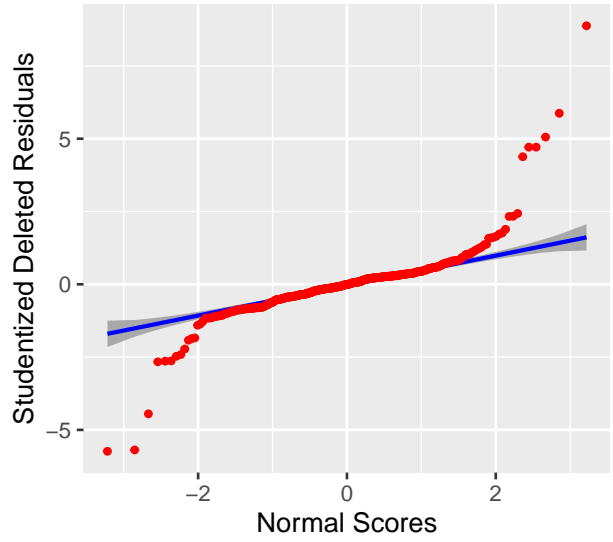
Applicable transformations were reexamined and the same conclusions drawn as with the full model.

Normal Probability Plot of Studentized Deleted Residuals (Model 3)

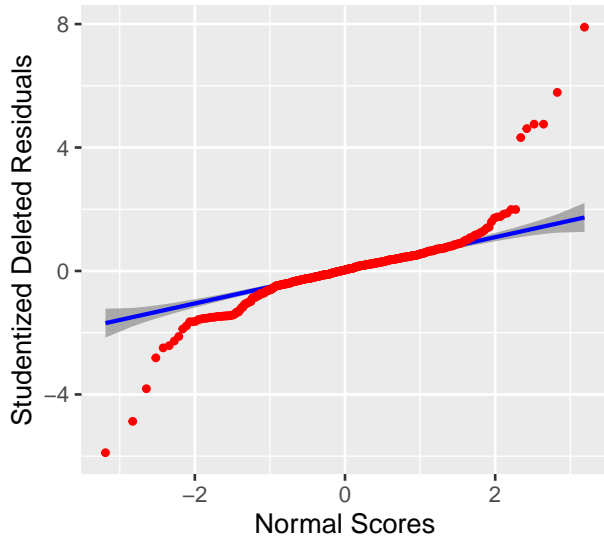
No Transformation



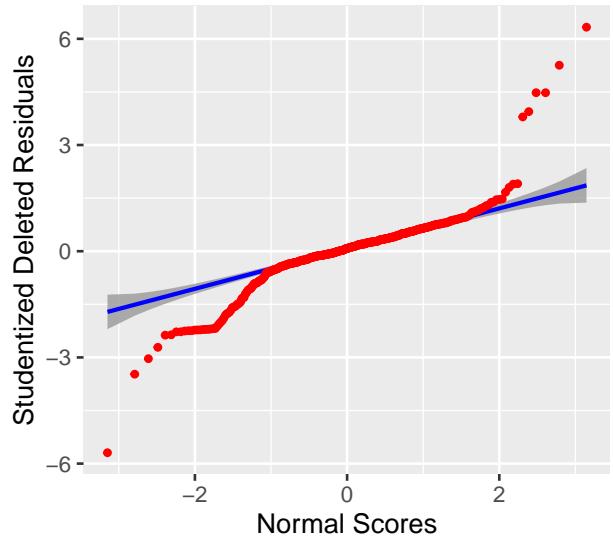
Box-Cox Transformation



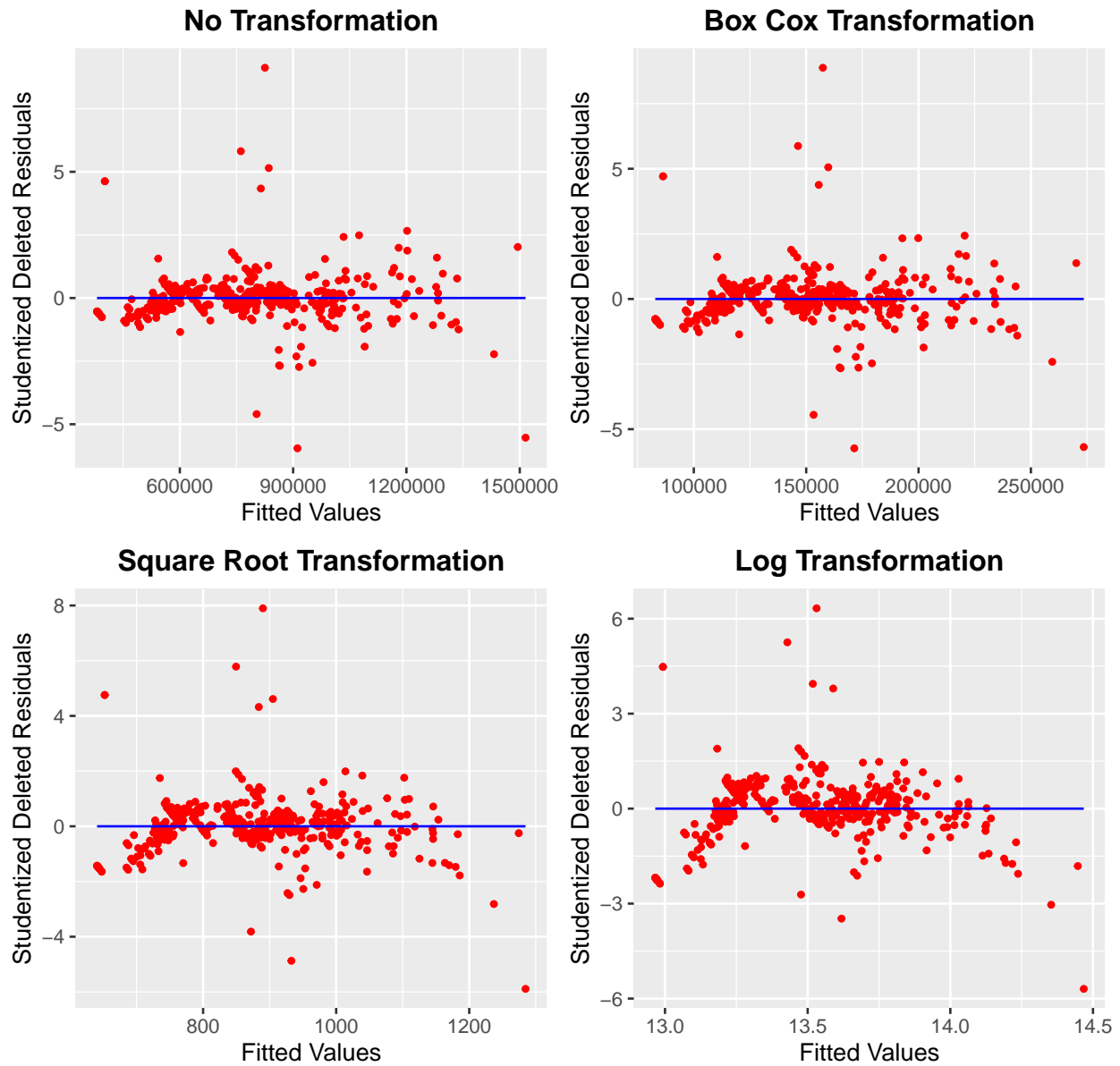
Square Root Transformation



Log Transformation



Studentized Deleted Residuals vs. Fitted Values (Model 3)



Call:

```
lm(formula = value ~ built + factor(type) + baths + GLA + price +
    sold, data = housing_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-296068	-23345	-625	15545	440859

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.213e+06	3.643e+06	-0.607	0.543973
built	4.485e+03	1.829e+03	2.452	0.014623 *
factor(type)condo_res	-1.216e+05	1.465e+04	-8.299	1.39e-15 ***
factor(type)end_gar_TH	-6.795e+04	1.067e+04	-6.368	4.94e-10 ***

```

factor(type)int_gar_TH -8.491e+04  1.000e+04  -8.490  3.43e-16  ***
factor(type)rambler    2.998e+05  6.219e+04   4.820  2.00e-06  ***
baths                  1.991e+04  5.370e+03   3.708  0.000236  ***
GLA                    5.415e+01  6.493e+00   8.339  1.04e-15  ***
price                  5.065e-01  2.371e-02  21.361  < 2e-16  ***
sold                   -3.242e+03  1.337e+03  -2.425  0.015735  *

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 55830 on 428 degrees of freedom
Multiple R-squared: 0.9407, Adjusted R-squared: 0.9394
F-statistic: 754.3 on 9 and 428 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: value

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
built	1	3.0904e+10	3.0904e+10	9.9159	0.001753	**
factor(type)	4	1.7076e+13	4.2690e+12	1369.7267	< 2.2e-16	***
baths	1	1.5351e+12	1.5351e+12	492.5570	< 2.2e-16	***
GLA	1	1.0684e+12	1.0684e+12	342.8066	< 2.2e-16	***
price	1	1.4307e+12	1.4307e+12	459.0521	< 2.2e-16	***
sold	1	1.8323e+10	1.8323e+10	5.8790	0.015735	*
Residuals	428	1.3339e+12	3.1167e+09			

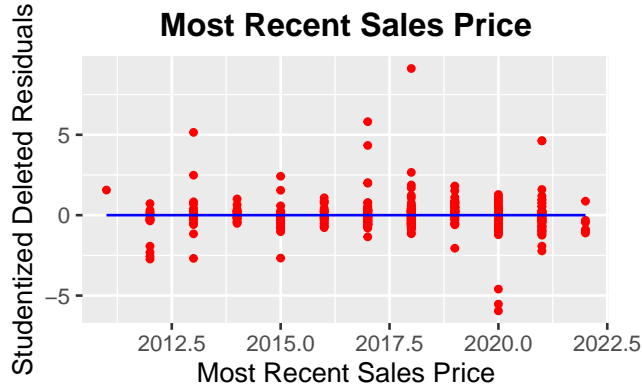
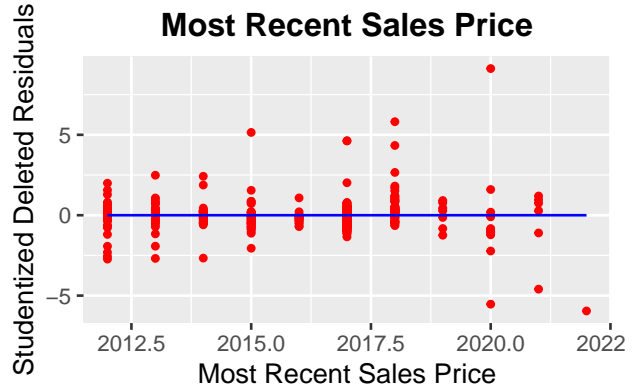
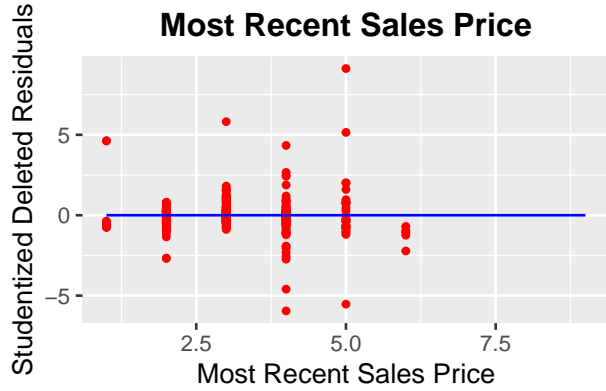
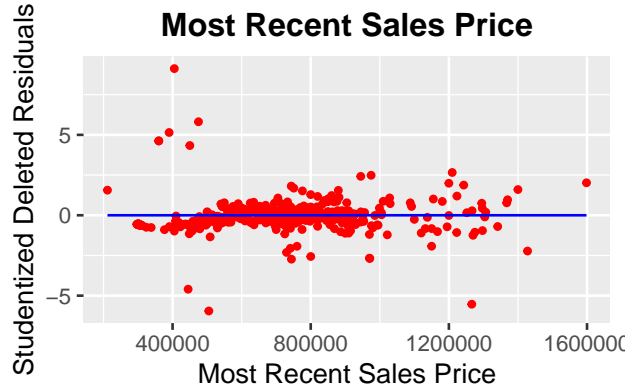
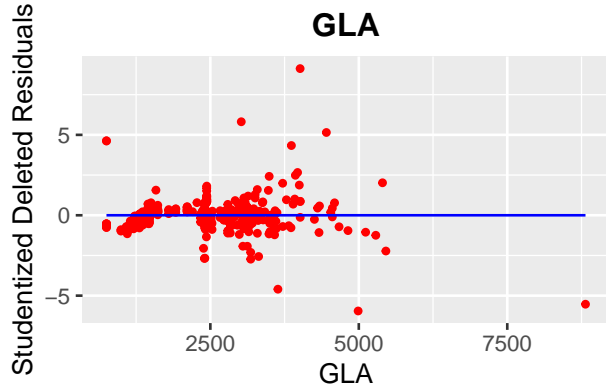
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

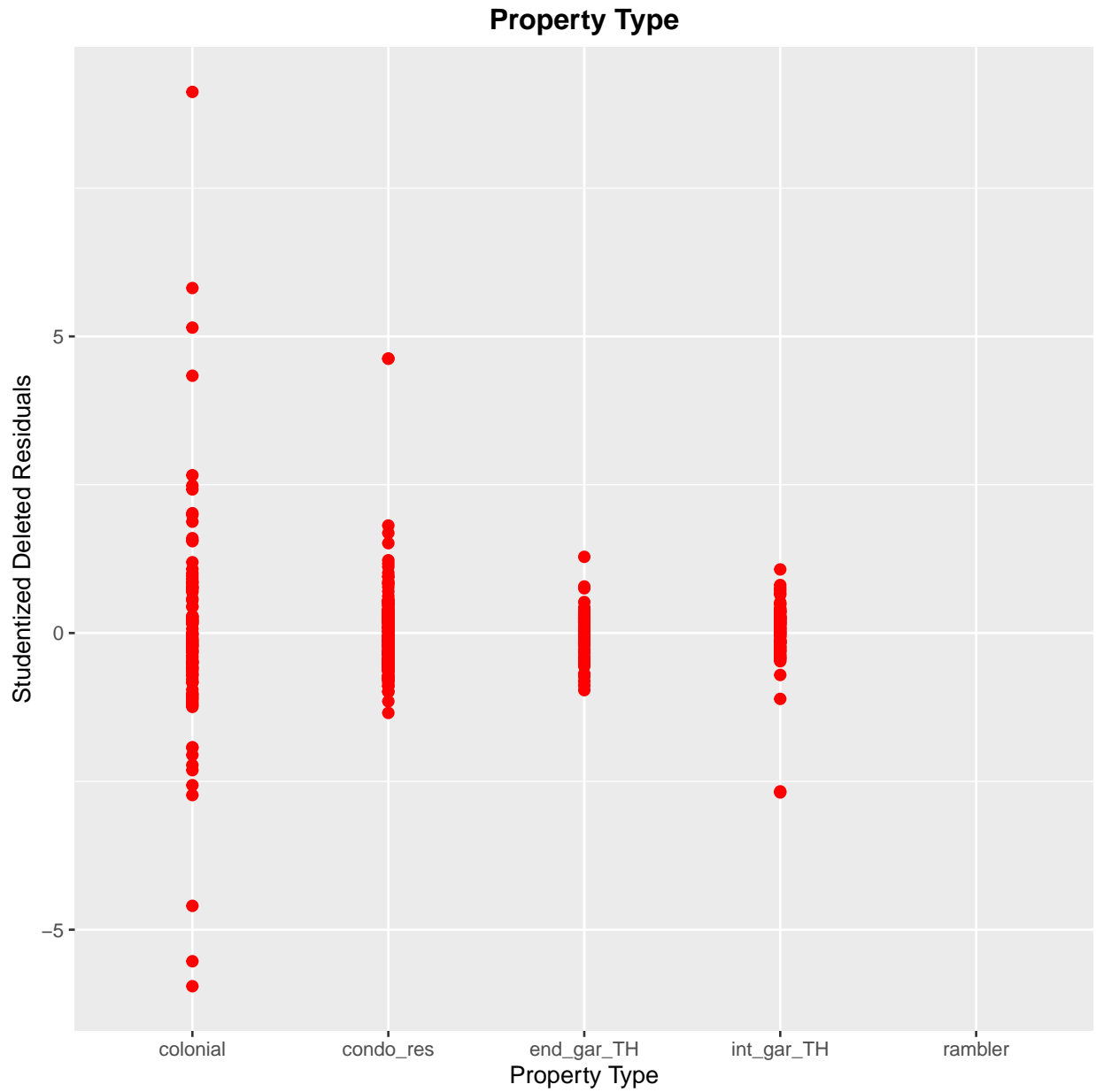
The regressors in the model continue to have large explanatory power and the variance not explained by the model is minimal. Property type, GLA, most recent sales price, and baths are significant at the $\alpha = 0.001$ level; however, year built and year sold are less significant at the $\alpha = 0.05$ level.

The overall F-statistic reported by the model is 754.3 on 9 and 428 degrees of freedom yielding a zero-like p-value. Further analysis on the impact of each individual regressor against the full model are broken down in an ANOVA table.

The ANOVA table reports property type, baths, GLA, and most recent sales price as significant at the $\alpha = 0.001$ level; year built at the $\alpha = 0.01$ level; and, year sold at the $\alpha = 0.05$ level. Plots of each regressor against the studentized deleted residuals were generated to look for new patterns but trends from the full model persisted in the reduced model.

Studentized Deleted Residuals vs. Individual Regressors (Model 3)





For the problem children, we note that $p = 10$ on a model with 9 regressors and 438 observations; therefore, the cutoff for leverage is $\frac{2p}{n} = 0.04566$. Also, the cutoff for DFFITS is $2\sqrt{\frac{p}{n}} = 0.3022$. Observation #434 continues to show a large Cook's distance. The following dataset indices were flagged with these thresholds:

	Leverage
3	1.00000000
16	0.04880388
26	0.10644996
30	0.05112993
32	0.06028667
37	0.05990708
38	0.04782647
58	0.07654997
185	0.09269680
187	0.04747732

```
206 0.06259034
211 0.05642110
268 0.04826906
348 0.05216499
349 0.05168637
351 0.06944606
354 0.04704895
432 0.14242343
433 0.10715135
434 0.30543747
```

DFFITS

```
18 0.4073251
22 -0.4316734
23 -0.4709371
24 -0.4021350
25 -0.3300069
26 3.1481409
27 0.3892090
30 0.3712511
57 -0.3496397
58 0.4497054
185 -1.4696205
206 0.5220358
268 -0.6050082
274 -0.4040535
286 -0.5448947
348 1.3642949
351 1.1848725
353 0.4144547
355 0.3239177
426 0.6345073
427 0.6345073
432 -2.4258711
433 1.7835296
434 -3.6684883
438 -0.4706342
```

Cook's D

```
434 1.25872
```

A subset of the data including observations yielding values greater than the above predetermined thresholds also show the following patterns:

```
[1] "Observations with Leverage"
```

```
[1] "3" "16" "26" "30" "32" "37" "38" "58" "185" "187" "206" "211"
[13] "268" "348" "349" "351" "354" "432" "433" "434"
```

```
[1] "Observations with DFFITS"
```

```
[1] "18" "22" "23" "24" "25" "26" "27" "30" "57" "58" "185" "206"
[13] "268" "274" "286" "348" "351" "353" "355" "426" "427" "432" "433" "434"
[25] "438"
```

```
[1] "Observations with Cook's D"
```

```
[1] "434"
```


	i..parcel_id	type	value	beds	baths	GLA	price	built	sold
3	47 3 07 092	rambler	1493800	8	9	5866	775000	2021	2019
16	47 4 03 052 A	colonial	1232300	4	4	3136	1223427	2021	2021
18	47 4 15 003	colonial	1347500	5	4	3969	1209660	2018	2018
22	48 3 19 003	colonial	811000	4	4	3313	800045	2012	2012
23	48 3 19 007	colonial	766600	4	4	3180	744135	2012	2012
24	48 3 19 009	colonial	782100	4	4	3180	730000	2012	2012
25	48 3 19 015	colonial	815000	4	4	3123	760255	2012	2012
26	57 1 02 069	colonial	1266200	6	5	4012	405000	2020	2018
27	57 1 02 076	colonial	1210900	5	4	3932	974900	2013	2013
30	57 1 10 092	colonial	1367400	5	5	3291	1400000	2020	2021
32	57 1 14 072	colonial	988000	5	5	2955	995000	2012	2021
37	57 1 15 D 005	colonial	945900	4	5	2988	970000	2012	2021
38	57 1 15 D 013	colonial	1038600	4	4	2836	1120000	2021	2021
57	57 1 21 084 A	colonial	749200	4	4	2386	741000	2015	2019
58	57 1 23 048	colonial	626500	5	3	1584	211500	2012	2011
185	57 2 03 B 019 B	colonial	564500	4	4	3638	445000	2021	2020
187	57 2 46 009	end_gar_TH	801200	4	4	2840	820000	2012	2021
206	57 2 49 001	colonial	1603600	6	5	5399	1598500	2017	2017
211	57 2 51 002	colonial	1255000	5	5	2966	1295234	2021	2021
268	57 3 20 017	int_gar_TH	719700	3	2	2400	971069	2013	2013
274	57 3 20 023	colonial	984600	4	4	3048	1150000	2013	2021
286	57 3 21 004	int_gar_TH	718000	3	2	2408	970113	2014	2015
348	58 1 02 014	colonial	1066200	4	3	3021	475000	2018	2017
349	58 1 02 017 A	colonial	1180900	6	5	4668	1195000	2013	2021
351	58 1 07 002	colonial	1043400	4	4	3865	450000	2018	2017
354	58 1 27 016	colonial	1254100	4	6	3720	1341336	2016	2016
355	58 1 27 050	colonial	1305200	4	4	3998	1243000	2014	2018
426	58 2 11 02 404	condo_res	651500	1	1	751	360000	2017	2021
432	58 3 04 024	colonial	615400	4	4	4989	505000	2022	2020
433	58 3 04 035	colonial	1099300	4	5	4454	390000	2015	2013
434	58 3 21 001	colonial	1266400	4	5	8817	1266301	2020	2020
438	58 3 21 005	colonial	1310900	5	6	5456	1427841	2020	2021

As with the other postulated models, observations #3, #26, and #434 are still the biggest potential problems for reasons previously addressed.

In summary, 32/438 observations were found to be above the predetermined thresholds. This result is similar to the count of problem children Model #2 and the full model.

4 Conclusions

After fully analyzing both the full and reduced models as well as the individual impact of each regressor on the response, it was found that some hypotheses were confirmed while others were not.

Analysis of the individual regressors show a positive relationship between total value and the following regressors: number of bedrooms, number of bathrooms, gross leasing area, and most recent sales price. This confirms hypotheses 2, 3, 4, and 6. A close look at the values reported for the different property types show that colonial and rambler property types have higher value than their counterparts (town homes and condominiums). It follows that hypothesis 7 is also confirmed.

Surprisingly, the years built and sold appeared to have negative relationships with the total value of the property; however, this relationship was minimal with a slope near zero indicating this relationship could become easily inverted given additional data.

For Model #1, the only regressors left standing were the property type, gross leasing area, and most recent

sales price. This is not surprising as the other regressors are likely at least somewhat correlated to these three regressors. The gross leasing area is a more accurate description of the property's size, so including the number of bedrooms and bathrooms is a bit redundant. The years built and sold can also be somewhat explained by the most recent sales price which seems to have a strong impact on total value. The proportion of problem children for this model was also significantly lower when compared to the other models.

Model #2 is a good middle ground when trying to balance accuracy and overhead. Including the number of bedrooms and bathrooms slightly increases precision but additional data is required.

Model #3 is the most accurate choice out of the reduced models losing only to the full model. Years built and sold help paint a clearer picture for the sales data; however, this data can be easily misreported or outdated which could potentially lead to more extreme values at the edges of the data distribution.

Throughout the regression analysis, the normal probability plots show potentially greater-than-normal distributions which can lead to invalid assumptions for OLS (namely that the dataset distribution is assumed normal). Common reasons for the presence of these extreme values include: measurement error, data corruption, or rare events. These extreme values require consultation with a domain expert who can determine the accuracy of the reported data and either confirm, fix, or remove the observations.

For future studies and analysis, including geolocation (i.e., longitude and latitude) as a regressor could potentially improve model accuracy; however, it is uncertain how much of an impact this information could have considering the dataset is county-exclusive. Also, removing the rambler property from the dataset seems justifiable considering there is only one observation for that factor level. Furthermore, it should be noted that calculating the PRESS statistic on any model with less than 2 observations at a given factor level is impossible. The PRESS statistic is a leave-one-out refitting and prediction method; it follows that the PRESS statistic cannot be calculated if an observation is removed and there are no observations left for refitting. Another strategy would be to perform separate regression analyses on subsets of the original dataset which separate the observations based on property type. The test for homogeneity of variance showed an uneven variance distribution which could lead to less precision; so, separating the observations by factor level is appropriate.

Ultimately, it was found that the strongest indicators of a prospective sales price are the property type, gross leasing area, and recent sales price. Thus, homeowners and property management firms planning to sell their Fairfax properties in 2022 should look closely at these metrics when drafting sales contracts. Finally, consultation with a real estate domain expert to validate extreme values present in the dataset may be required.

References

- [1] “City of Fairfax, VA.” Real Estate Assessment Database | City of Fairfax, VA, <https://www.fairfaxva.gov/government/real-estate/online-services/real-estate-assessment-database>.
- [2] Zillow. U.S. Housing Market Has Doubled in Value since the Great Recession after Gaining \$6.9 Trillion in 2021, 27 Jan. 2022, <https://www.prnewswire.com/news-releases/us-housing-market-has-doubled-in-value-since-the-great-recession-after-gaining-6-9-trillion-in-2021--301469460.html>.

Appendix (Dataset)

	i..parcel_id	type	value	beds	baths	GLA	price	built	sold
1	47 3 07 015 B	colonial	1193900	6	5	3576	1200000	2020	2021
2	47 3 07 017	colonial	1069800	3	3	3477	880000	2015	2015
3	47 3 07 092	rambler	1493800	8	9	5866	775000	2021	2019
4	47 3 13 001	colonial	946700	4	3	3147	891179	2016	2016
5	47 3 13 002	colonial	981000	4	4	3147	922885	2015	2015
6	47 3 13 003	colonial	930100	4	4	3064	868711	2015	2015
7	47 3 13 005	colonial	932600	4	4	3292	900000	2015	2018
8	47 3 13 006	colonial	992300	4	3	3147	951054	2016	2016
9	47 3 13 007	colonial	975300	4	4	3091	919622	2016	2016
10	47 3 13 008	colonial	933000	4	4	3147	886831	2015	2015
11	47 3 13 009	colonial	994700	4	4	3095	920240	2015	2015
12	47 3 13 010	colonial	963800	5	4	3147	908794	2016	2016
13	47 3 13 011	colonial	940600	5	4	3147	801020	2015	2014
14	47 3 13 012	colonial	974800	5	4	3175	930000	2015	2017
15	47 3 13 013	colonial	985900	4	4	3147	930000	2015	2015
16	47 4 03 052 A	colonial	1232300	4	4	3136	1223427	2021	2021
17	47 4 03 052 B	colonial	1349100	5	5	3781	1370000	2021	2021
18	47 4 15 003	colonial	1347500	5	4	3969	1209660	2018	2018
19	47 4 24 001	colonial	1294100	5	4	4528	1306839	2020	2020
20	47 4 24 002	colonial	1278100	5	4	4554	1304018	2020	2020
21	47 4 24 003	colonial	1302900	5	4	4554	1296637	2019	2019
22	48 3 19 003	colonial	811000	4	4	3313	800045	2012	2012
23	48 3 19 007	colonial	766600	4	4	3180	744135	2012	2012
24	48 3 19 009	colonial	782100	4	4	3180	730000	2012	2012
25	48 3 19 015	colonial	815000	4	4	3123	760255	2012	2012
26	57 1 02 069	colonial	1266200	6	5	4012	405000	2020	2018
27	57 1 02 076	colonial	1210900	5	4	3932	974900	2013	2013
28	57 1 04 082	colonial	1167000	2	4	3491	944900	2014	2015
29	57 1 09 046	colonial	796700	3	2	2533	650000	2014	2018
30	57 1 10 092	colonial	1367400	5	5	3291	1400000	2020	2021
31	57 1 14 069	colonial	987800	6	5	3384	732990	2013	2013
32	57 1 14 072	colonial	988000	5	5	2955	995000	2012	2021
33	57 1 15 B 017	colonial	1079200	4	5	3377	904450	2019	2019
34	57 1 15 B 018	colonial	1008600	4	4	2918	838835	2019	2019
35	57 1 15 D 001	colonial	860000	4	4	3488	725000	2013	2013
36	57 1 15 D 002	colonial	1058800	4	5	3894	833790	2013	2013
37	57 1 15 D 005	colonial	945900	4	5	2988	970000	2012	2021
38	57 1 15 D 013	colonial	1038600	4	4	2836	1120000	2021	2021
39	57 1 20 002 A	int_gar_TH	866900	3	3	3305	889000	2017	2021
40	57 1 20 003 A	int_gar_TH	880400	3	3	3305	900000	2017	2021
41	57 1 20 005 A	int_gar_TH	842600	3	3	3305	825000	2016	2020
42	57 1 20 006 A	int_gar_TH	845800	3	3	3329	803500	2016	2016
43	57 1 20 007 A	int_gar_TH	842600	3	3	3305	783000	2016	2017
44	57 1 20 008 A	end_gar_TH	883800	3	4	3211	815000	2016	2017
45	57 1 20 009 A	int_gar_TH	896100	3	4	3305	881960	2014	2014
46	57 1 20 014 B	int_gar_TH	836200	4	3	3230	741978	2013	2013
47	57 1 20 015 B	end_gar_TH	904700	4	3	3254	800000	2012	2020
48	57 1 20 018 B	int_gar_TH	839400	4	3	3254	705062	2013	2018
49	57 1 20 019 B	int_gar_TH	839400	3	3	3254	762265	2013	2013
50	57 1 20 021 A	int_gar_TH	867400	4	3	3226	841913	2014	2014
51	57 1 20 022 A	int_gar_TH	860800	4	3	3226	875000	2015	2021

52	57	1	20	023	A	end_gar_TH	784600	3	2	2430	695000	2015	2016	
53		57	1	20	037	end_gar_TH	766800	3	2	2275	710000	2013	2019	
54		57	1	20	038	end_gar_TH	883800	3	4	3211	822000	2017	2018	
55		57	1	20	039	end_gar_TH	883800	3	4	3211	821000	2016	2016	
56		57	1	20	040	end_gar_TH	985800	3	4	3211	1005000	2014	2014	
57	57	1	21	084	A	colonial	749200	4	4	2386	741000	2015	2019	
58		57	1	23	048	colonial	626500	5	3	1584	211500	2012	2011	
59		57	1	23	057	colonial	855400	4	3	3190	687845	2012	2012	
60		57	1	37	001	colonial	974900	4	4	2907	900000	2015	2016	
61		57	1	37	002	colonial	989100	4	4	2949	890000	2014	2016	
62		57	1	37	003	colonial	1002100	4	4	2904	929000	2015	2015	
63		57	1	37	004	colonial	1018100	4	4	2904	945215	2015	2015	
64		57	1	37	005	colonial	1008700	4	4	3312	920990	2014	2015	
65		57	1	37	006	colonial	978600	4	4	2881	992500	2014	2019	
66		57	1	38	001	end_gar_TH	850800	4	4	2794	747468	2017	2017	
67		57	1	38	002	int_gar_TH	822900	4	4	2794	788370	2017	2017	
68		57	1	38	003	int_gar_TH	817500	4	3	2794	725000	2017	2019	
69		57	1	38	004	int_gar_TH	822900	4	4	2794	910000	2017	2022	
70		57	1	38	005	end_gar_TH	850800	4	4	2794	915416	2017	2022	
71		57	1	38	007	int_gar_TH	822900	4	4	2794	775000	2017	2017	
72		57	1	38	008	int_gar_TH	822900	4	4	2794	773000	2017	2017	
73		57	1	38	009	end_gar_TH	846900	4	3	2794	778774	2017	2017	
74	57	1	38	01	039	condo_res	551100	3	2	1471	561391	2017	2017	
75		57	1	38	01	040	condo_res	661000	3	2	2441	679000	2017	2021
76		57	1	38	01	041	condo_res	526600	3	2	2434	508788	2017	2017
77		57	1	38	01	042	condo_res	635300	3	2	2419	632874	2017	2017
78		57	1	38	01	043	condo_res	526600	3	2	1434	545000	2017	2021
79		57	1	38	01	044	condo_res	631600	3	2	2419	700000	2017	2022
80		57	1	38	01	045	condo_res	526600	3	2	1434	520000	2017	2018
81		57	1	38	01	046	condo_res	635300	3	2	2419	626221	2017	2017
82		57	1	38	01	048	condo_res	633200	3	2	2427	625306	2017	2018
83		57	1	38	01	049	condo_res	526600	3	2	1434	528000	2017	2018
84		57	1	38	01	050	condo_res	631600	3	2	2419	602000	2017	2021
85		57	1	38	01	051	condo_res	664200	3	2	2483	658314	2017	2017
86	57	1	38	01	052	condo_res	534000	3	2	1463	538506	2017	2017	
87		57	1	38	010	end_gar_TH	906900	3	4	3120	925000	2017	2021	
88		57	1	38	011	int_gar_TH	872000	3	4	3120	785000	2017	2020	
89		57	1	38	012	int_gar_TH	867600	3	3	3120	799990	2017	2018	
90		57	1	38	013	int_gar_TH	871500	3	4	3120	799990	2017	2018	
91		57	1	38	014	int_gar_TH	871500	3	4	3120	798000	2017	2018	
92		57	1	38	015	int_gar_TH	871500	4	4	3120	793979	2017	2017	
93		57	1	38	016	end_gar_TH	893900	3	3	3120	817990	2017	2018	
94		57	1	38	017	end_gar_TH	912400	3	3	3250	890000	2017	2020	
95		57	1	38	018	int_gar_TH	871500	4	4	3120	778659	2017	2017	
96	57	1	38	02	053	condo_res	536100	3	2	1471	545000	2017	2021	
97		57	1	38	02	054	condo_res	652300	3	2	2441	641731	2017	2017
98		57	1	38	02	055	condo_res	631600	3	2	2419	628221	2017	2018
99		57	1	38	02	056	condo_res	526600	3	2	1434	477590	2017	2017
100		57	1	38	02	057	condo_res	633200	3	2	2427	597239	2017	2017
101		57	1	38	02	059	condo_res	631600	3	2	2419	609839	2017	2017
102		57	1	38	02	061	condo_res	635300	3	2	2419	599839	2017	2017
103		57	1	38	02	062	condo_res	526600	3	2	1434	504918	2017	2017
104		57	1	38	02	063	condo_res	631600	3	2	2419	635839	2017	2017
105		57	1	38	02	064	condo_res	541600	3	2	1434	536811	2017	2017

106	57	1	38	02	065	condo_res	631600	3	2	2419	575000	2017	2019
107	57	1	38	02	067	condo_res	654600	3	2	2441	652631	2017	2018
108	57	1	38	02	068	condo_res	551100	3	2	1471	541511	2017	2018
109		57	1	38	020	int_gar_TH	876200	4	4	3120	859000	2017	2020
110		57	1	38	021	int_gar_TH	871500	4	4	3120	808670	2017	2017
111		57	1	38	022	end_gar_TH	901800	4	4	3120	853937	2017	2017
112		57	1	38	023	end_gar_TH	901800	4	4	3120	825000	2017	2018
113		57	1	38	024	int_gar_TH	871500	4	4	3120	825729	2017	2018
114		57	1	38	025	int_gar_TH	871500	4	4	3120	900000	2017	2022
115		57	1	38	026	end_gar_TH	946800	4	4	3120	922855	2017	2017
116		57	1	38	027	end_gar_TH	893900	3	3	3120	824196	2018	2018
117		57	1	38	028	int_gar_TH	871500	3	4	3120	777910	2018	2018
118		57	1	38	029	int_gar_TH	871500	3	4	3120	815000	2018	2020
119	57	1	38	03	101	condo_res	522600	2	2	1434	506154	2017	2018
120	57	1	38	03	102	condo_res	469300	2	2	1208	464076	2017	2017
121	57	1	38	03	103	condo_res	469300	2	2	1208	467302	2017	2017
122	57	1	38	03	104	condo_res	591200	2	2	1480	592904	2017	2018
123	57	1	38	03	105	condo_res	604300	2	2	1615	582122	2017	2017
124	57	1	38	03	106	condo_res	550800	2	2	1427	551167	2017	2018
125	57	1	38	03	107	condo_res	550800	2	2	1427	514729	2017	2017
126	57	1	38	03	108	condo_res	604300	2	2	1615	584692	2017	2018
127	57	1	38	03	201	condo_res	546200	2	2	1434	515000	2017	2018
128	57	1	38	03	202	condo_res	494300	2	2	1208	490000	2017	2018
129	57	1	38	03	203	condo_res	493800	2	2	1306	458717	2017	2017
130	57	1	38	03	204	condo_res	591200	2	2	1480	558160	2017	2017
131	57	1	38	03	205	condo_res	604300	2	2	1615	599692	2017	2017
132	57	1	38	03	206	condo_res	550800	2	2	1427	534667	2017	2018
133	57	1	38	03	207	condo_res	550800	2	2	1427	550000	2017	2020
134	57	1	38	03	208	condo_res	604300	2	2	1615	567122	2017	2017
135	57	1	38	03	301	condo_res	546200	2	2	1434	550000	2017	2018
136	57	1	38	03	302	condo_res	469300	2	2	1208	441007	2017	2017
137	57	1	38	03	303	condo_res	508800	2	2	1306	502030	2017	2017
138	57	1	38	03	306	condo_res	550800	2	2	1427	530000	2017	2018
139	57	1	38	03	307	condo_res	550800	2	2	1427	535000	2017	2017
140	57	1	38	03	308	condo_res	604300	2	2	1615	570000	2017	2017
141	57	1	38	03	401	condo_res	546200	2	2	1434	546000	2017	2018
142	57	1	38	03	402	condo_res	469300	2	2	1208	479007	2017	2018
143	57	1	38	03	403	condo_res	493800	2	2	1306	530000	2017	2021
144	57	1	38	03	404	condo_res	616200	2	2	1480	603459	2017	2017
145	57	1	38	03	405	condo_res	639300	2	2	1615	641657	2017	2017
146	57	1	38	03	407	condo_res	595800	2	2	1427	602732	2017	2018
147	57	1	38	03	408	condo_res	669300	2	2	1615	669402	2017	2018
148		57	1	38	030	int_gar_TH	871500	3	4	3120	769990	2018	2019
149		57	1	38	031	int_gar_TH	898700	3	3	3120	877625	2018	2018
150		57	1	38	032	end_gar_TH	893900	3	3	3120	846701	2018	2018
151		57	1	38	033	end_gar_TH	849800	4	2	2794	850000	2017	2020
152		57	1	38	034	int_gar_TH	822900	4	4	2794	798065	2017	2017
153		57	1	38	035	int_gar_TH	822900	4	4	2794	775000	2017	2017
154		57	1	38	036	int_gar_TH	822900	4	4	2794	774894	2017	2017
155		57	1	38	038	end_gar_TH	852300	4	4	2794	818358	2017	2018
156	57	1	38	04	014	condo_res	528600	3	2	1442	520000	2017	2018
157	57	1	38	04	101	condo_res	546200	2	2	1434	540000	2018	2021
158	57	1	38	04	102	condo_res	503500	2	2	1360	496551	2018	2018
159	57	1	38	04	104	condo_res	591200	2	2	1480	550000	2018	2021

160	57	1	38	04	105	condo_res	604300	2	2	1615	581232	2018	2019	
161	57	1	38	04	106	condo_res	531600	2	2	1469	529161	2018	2018	
162	57	1	38	04	202	condo_res	643600	2	2	1799	647217	2018	2018	
163	57	1	38	04	203	condo_res	442000	1	1	1152	408615	2018	2018	
164	57	1	38	04	204	condo_res	591200	2	2	1480	567000	2018	2018	
165	57	1	38	04	205	condo_res	604300	2	2	1615	575000	2018	2018	
166	57	1	38	04	206	condo_res	546600	2	2	1469	542536	2018	2018	
167	57	1	38	04	207	condo_res	661100	2	2	1918	660725	2018	2018	
168	57	1	38	04	208	condo_res	604300	2	2	1615	586289	2018	2019	
169	57	1	38	04	301	condo_res	571200	2	2	1434	571526	2018	2018	
170	57	1	38	04	302	condo_res	658600	2	2	1799	656513	2018	2018	
171	57	1	38	04	303	condo_res	442000	1	1	1152	417785	2018	2018	
172	57	1	38	04	305	condo_res	619300	2	2	1615	620000	2018	2020	
173	57	1	38	04	306	condo_res	531600	2	2	1469	523305	2018	2018	
174	57	1	38	04	307	condo_res	676100	2	2	1918	668113	2018	2018	
175	57	1	38	04	308	condo_res	604300	2	2	1615	584500	2018	2018	
176	57	1	38	04	401	condo_res	581200	2	2	1434	587185	2018	2018	
177	57	1	38	04	402	condo_res	668600	2	2	1799	669013	2018	2018	
178	57	1	38	04	403	condo_res	457000	1	1	1152	468840	2018	2018	
179	57	1	38	04	404	condo_res	591200	2	2	1480	590000	2018	2021	
180	57	1	38	04	405	condo_res	629300	2	2	1615	607989	2018	2018	
181	57	1	38	04	406	condo_res	556600	2	2	1469	568951	2018	2018	
182	57	1	38	04	407	condo_res	686100	2	2	1918	685478	2018	2018	
183	57	1	38	04	408	condo_res	629300	2	2	1615	616329	2018	2018	
184		57	2	02	102	colonial	960100	4	3	3307	900000	2016	2019	
185	57	2	03	B	019	B	colonial	564500	4	4	3638	445000	2021	2020
186		57	2	45	005	colonial	1018600	6	4	4338	780000	2012	2015	
187		57	2	46	009	end_gar_TH	801200	4	4	2840	820000	2012	2021	
188		57	2	46	010	int_gar_TH	782800	3	3	2840	763579	2012	2012	
189		57	2	46	012	end_gar_TH	784200	4	3	2731	738368	2012	2012	
190		57	2	46	013	int_gar_TH	790400	4	3	2731	810000	2012	2021	
191		57	2	46	015	end_gar_TH	796400	4	3	2796	741235	2012	2012	
192		57	2	46	016	int_gar_TH	763900	4	3	2726	731961	2012	2012	
193		57	2	46	017	end_gar_TH	779000	4	3	2702	688000	2012	2012	
194		57	2	46	018	end_gar_TH	829500	4	3	3043	727973	2012	2012	
195		57	2	46	019	int_gar_TH	823300	4	3	3067	723000	2012	2014	
196		57	2	46	020	int_gar_TH	827500	4	3	3067	775000	2012	2017	
197		57	2	46	021	int_gar_TH	732900	3	3	2394	755000	2012	2021	
198		57	2	46	022	int_gar_TH	830300	4	4	3067	780000	2012	2018	
199		57	2	46	023	end_gar_TH	833000	4	3	3067	747083	2012	2012	
200		57	2	46	024	end_gar_TH	820400	4	4	2978	719000	2012	2015	
201		57	2	46	025	int_gar_TH	814800	4	3	2905	795000	2012	2018	
202		57	2	46	026	end_gar_TH	726300	3	3	2200	662500	2012	2018	
203		57	2	48	001	colonial	1107700	6	5	3504	1167000	2017	2020	
204		57	2	48	002	colonial	1058000	5	5	3821	975892	2017	2017	
205		57	2	48	003	colonial	1045400	5	5	3604	953649	2016	2017	
206		57	2	49	001	colonial	1603600	6	5	5399	1598500	2017	2017	
207		57	2	49	002	colonial	1376700	5	5	4592	1367555	2017	2017	
208		57	2	50	001	colonial	1128900	6	5	3610	1150129	2020	2020	
209		57	2	50	002	colonial	1021000	5	4	3582	1021685	2020	2020	
210		57	2	50	003	colonial	1113200	6	5	3576	1131854	2019	2020	
211		57	2	51	002	colonial	1255000	5	5	2966	1295234	2021	2021	
212		57	2	51	003	colonial	1249400	5	5	3590	1267164	2021	2021	
213		57	3	02	004	colonial	1270100	9	6	5286	1270000	2019	2021	

214	57	3	03	001	end_gar_TH	733900	3	2	2409	647990	2014	2014
215	57	3	03	002	int_gar_TH	740900	3	2	2409	749900	2014	2021
216	57	3	03	003	int_gar_TH	715900	3	2	2409	695000	2014	2020
217	57	3	03	004	end_gar_TH	733900	3	2	2409	665000	2014	2017
218	57	3	03	005	end_gar_TH	733900	3	2	2409	698000	2013	2020
219	57	3	03	006	int_gar_TH	715900	3	2	2409	635000	2012	2019
220	57	3	03	007	end_gar_TH	734300	3	2	2370	669900	2012	2014
221	57	3	03	008	end_gar_TH	733900	3	2	2409	650000	2014	2014
222	57	3	03	009	end_gar_TH	733900	3	2	2409	705000	2014	2020
223	57	3	03	010	end_gar_TH	727300	3	3	2337	648990	2014	2014
224	57	3	03	011	int_gar_TH	706100	3	2	2337	653065	2014	2014
225	57	3	03	012	end_gar_TH	723800	3	2	2337	763000	2013	2021
226	57	3	03	013	end_gar_TH	749300	3	2	2370	765000	2013	2021
227	57	3	03	015	int_gar_TH	715900	4	2	2409	677500	2013	2019
228	57	3	03	016	int_gar_TH	715900	3	2	2409	676500	2013	2016
229	57	3	03	017	int_gar_TH	715900	3	2	2409	653410	2013	2014
230	57	3	03	021	int_gar_TH	745500	3	2	2370	760000	2014	2021
231	57	3	03	022	A end_gar_TH	733900	3	2	2409	680000	2013	2018
232	57	3	03	023	end_gar_TH	733900	3	2	2409	650000	2013	2016
233	57	3	03	024	end_gar_TH	733900	3	2	2409	675000	2014	2017
234	57	3	03	025	end_gar_TH	733900	3	2	2409	639990	2013	2013
235	57	3	03	027	A end_gar_TH	729700	3	2	2409	685000	2013	2020
236	57	3	03	029	end_gar_TH	730100	3	2	2370	687500	2013	2019
237	57	3	03	030	A end_gar_TH	733200	3	3	2409	700000	2012	2019
238	57	3	03	031	end_gar_TH	733900	3	2	2409	770000	2013	2022
239	57	3	03	032	int_gar_TH	715900	3	2	2409	625000	2013	2018
240	57	3	03	033	int_gar_TH	715900	3	2	2409	680000	2013	2020
241	57	3	03	034	end_gar_TH	733900	3	2	2409	665000	2013	2016
242	57	3	03	035	int_gar_TH	725900	3	2	2409	725000	2013	2021
243	57	3	03	038	end_gar_TH	730100	3	2	2370	667500	2012	2015
244	57	3	03	039	int_gar_TH	709900	3	2	2390	589990	2012	2012
245	57	3	03	040	int_gar_TH	709900	3	2	2390	664900	2012	2018
246	57	3	03	041	end_gar_TH	732000	3	2	2390	740000	2012	2021
247	57	3	03	042	A end_gar_TH	734300	3	2	2370	706523	2013	2020
248	57	3	03	043	int_gar_TH	715900	3	2	2409	666795	2013	2013
249	57	3	03	044	end_gar_TH	733900	3	2	2409	669000	2013	2017
250	57	3	03	046	int_gar_TH	715900	3	2	2409	679500	2013	2018
251	57	3	03	047	int_gar_TH	715900	3	2	2409	710000	2013	2021
252	57	3	03	048	end_gar_TH	737400	3	3	2409	745000	2013	2021
253	57	3	20	001	end_gar_TH	807500	4	3	2752	740000	2014	2018
254	57	3	20	002	int_gar_TH	790300	3	3	2750	755000	2014	2019
255	57	3	20	003	int_gar_TH	779900	3	2	2854	713615	2014	2014
256	57	3	20	004	int_gar_TH	788000	3	3	2734	742805	2014	2014
257	57	3	20	005	int_gar_TH	790300	3	3	2750	761120	2014	2014
258	57	3	20	007	end_gar_TH	746900	3	3	2437	714468	2014	2014
259	57	3	20	008	int_gar_TH	772900	3	3	2888	709162	2014	2014
260	57	3	20	009	int_gar_TH	727600	3	3	2400	704059	2014	2014
261	57	3	20	010	end_gar_TH	792400	3	3	2888	696951	2014	2014
262	57	3	20	011	end_gar_TH	793800	3	3	2897	735000	2014	2019
263	57	3	20	012	int_gar_TH	724400	3	3	2408	697743	2014	2014
264	57	3	20	012	int_gar_TH	719700	3	2	2400	681154	2014	2014
265	57	3	20	014	int_gar_TH	720800	3	2	2408	664445	2014	2014
266	57	3	20	015	end_gar_TH	764100	3	2	2408	737785	2014	2014
267	57	3	20	016	end_gar_TH	744000	3	3	2417	679999	2013	2013

268	57	3	20	017	int_gar_TH	719700	3	2	2400	971069	2013	2013	
269	57	3	20	018	int_gar_TH	720800	3	2	2408	699900	2013	2020	
270	57	3	20	019	end_gar_TH	790200	3	2	2897	744999	2013	2015	
271	57	3	20	020	colonial	978300	4	3	3328	900000	2013	2019	
272	57	3	20	021	colonial	954400	4	3	3390	866549	2013	2013	
273	57	3	20	022	colonial	937200	4	4	3106	855736	2013	2013	
274	57	3	20	023	colonial	984600	4	4	3048	1150000	2013	2021	
275	57	3	20	024	colonial	989200	4	4	3208	928466	2014	2013	
276	57	3	20	025	end_gar_TH	853100	3	3	3165	800788	2013	2013	
277	57	3	20	026	int_gar_TH	834500	3	3	3166	806500	2013	2020	
278	57	3	20	027	end_gar_TH	853100	3	3	3165	784010	2013	2013	
279	57	3	20	028	end_gar_TH	778500	3	3	2512	739900	2014	2015	
280	57	3	20	029	int_gar_TH	833000	3	3	3166	758267	2014	2014	
281	57	3	20	030	int_gar_TH	832900	3	3	3165	785000	2014	2019	
282	57	3	20	031	end_gar_TH	753500	3	3	2512	714000	2014	2016	
283	57	3	21	001	end_gar_TH	741100	3	3	2417	735000	2014	2019	
284	57	3	21	002	int_gar_TH	725900	3	3	2408	717000	2014	2020	
285	57	3	21	003	int_gar_TH	718000	3	2	2408	678999	2014	2015	
286	57	3	21	004	int_gar_TH	718000	3	2	2408	970113	2014	2015	
287	57	3	21	005	int_gar_TH	720400	3	3	2400	687500	2014	2020	
288	57	3	21	006	int_gar_TH	721600	3	3	2408	685000	2014	2018	
289	57	3	21	008	end_gar_TH	736200	3	2	2408	697000	2015	2019	
290	57	3	21	009	int_gar_TH	725900	3	3	2408	728670	2015	2015	
291	57	3	21	010	int_gar_TH	720400	3	3	2400	684676	2015	2015	
292	57	3	21	011	int_gar_TH	718000	3	2	2408	680804	2015	2015	
293	57	3	21	012	int_gar_TH	756600	3	3	2408	777700	2015	2021	
294	57	3	21	013	int_gar_TH	720400	3	3	2400	725000	2015	2021	
295	57	3	21	014	end_gar_TH	743500	3	3	2409	741195	2015	2015	
296	57	3	22	001	end_gar_TH	725700	3	3	2409	732500	2015	2021	
297	57	3	22	002	int_gar_TH	707500	3	3	2408	680000	2015	2020	
298	57	3	22	003	end_gar_TH	726800	3	3	2417	698694	2015	2015	
299	57	3	22	004	end_gar_TH	773800	3	3	2528	754739	2015	2015	
300	57	3	22	005	int_gar_TH	747400	3	2	2494	721979	2015	2015	
301	57	3	22	006	end_gar_TH	736000	3	3	2512	720000	2015	2020	
302	57	3	22	007	end_gar_TH	743700	3	3	2529	775000	2015	2021	
303	57	3	22	008	int_gar_TH	713600	3	3	2494	670000	2015	2015	
304	57	4	05	01	034	colonial	1006700	6	5	3024	945000	2018	2020
305	57	4	05	02	010	colonial	1005200	5	4	3417	862350	2016	2016
306	57	4	05	04	015	colonial	996000	5	4	2664	910000	2019	2019
307	57	4	05	04	016	colonial	966000	5	4	2664	896430	2019	2019
308	57	4	05	04	021	colonial	945900	4	4	3525	778790	2016	2016
309	57	4	05	04	022	colonial	942500	4	4	3525	785893	2016	2016
310	57	4	05	04	023	colonial	1034300	5	5	3511	895806	2015	2015
311	57	4	05	04	024	colonial	990900	4	5	3519	845375	2015	2015
312	57	4	05	06	003	colonial	976300	4	5	3500	825000	2015	2018
313	57	4	05	06	004	colonial	976100	4	4	3500	917000	2014	2020
314	57	4	37	003	colonial	1036300	4	5	3858	980000	2012	2016	
315	57	4	38	001	colonial	1036500	4	3	3066	1004700	2017	2017	
316	57	4	38	002	colonial	1030600	4	4	3028	985537	2016	2017	
317	57	4	38	003	colonial	1031200	4	4	3087	956000	2015	2015	
318	57	4	38	004	colonial	1113800	4	3	3087	1088140	2017	2017	
319	57	4	38	005	colonial	1098000	4	3	3066	1028507	2016	2016	
320	57	4	38	006	colonial	1120300	5	4	3028	1092070	2017	2017	
321	57	4	38	037	colonial	1080200	4	4	2974	1030000	2015	2019	

322	57	4	38	038	colonial	1144000	5	4	4013	1009900	2015	2016	
323	57	4	38	039	colonial	1230000	4	4	4013	1180949	2015	2016	
324	57	4	38	040	colonial	1159200	4	4	4013	1137712	2017	2017	
325	57	4	38	041	colonial	1066400	4	4	3339	1000000	2015	2016	
326	57	4	39	01	001	condo_res	845000	3	3	2440	850000	2018	2020
327	57	4	39	01	002	condo_res	838400	3	3	2440	751000	2018	2018
328	57	4	39	01	003	condo_res	838400	3	3	2440	820000	2018	2020
329	57	4	39	01	004	condo_res	838400	3	3	2440	776000	2018	2019
330	57	4	39	01	005	condo_res	838400	3	3	2440	743607	2018	2019
331	57	4	39	01	006	condo_res	843500	3	3	2440	876000	2018	2022
332	57	4	39	02	013	condo_res	739100	3	3	2110	759750	2018	2020
333	57	4	39	02	014	condo_res	739100	3	3	2110	753033	2018	2020
334	57	4	39	02	015	condo_res	739100	3	3	2110	758056	2018	2019
335	57	4	39	02	016	condo_res	745700	3	3	2110	760987	2018	2020
336	57	4	39	02	017	condo_res	739100	3	3	2110	752189	2018	2019
337	57	4	39	03	007	condo_res	745700	3	3	2110	741900	2018	2020
338	57	4	39	03	008	condo_res	740600	3	3	2110	739900	2018	2020
339	57	4	39	03	009	condo_res	739100	3	3	2110	725170	2018	2019
340	57	4	39	03	010	condo_res	739100	3	3	2110	727300	2018	2019
341	57	4	39	03	012	condo_res	739100	3	3	2110	747789	2018	2020
342	57	4	39	04	019	condo_res	870000	3	3	2440	882000	2018	2019
343	57	4	39	04	020	condo_res	843500	3	3	2440	854296	2018	2020
344	57	4	39	04	021	condo_res	843500	3	3	2440	867040	2018	2020
345	57	4	39	04	022	condo_res	838400	3	3	2440	849676	2018	2019
346	57	4	39	04	023	condo_res	868500	3	3	2440	860598	2018	2018
347	57	4	39	04	024	condo_res	838400	3	3	2440	850000	2018	2021
348	58	1	02	014	colonial	1066200	4	3	3021	475000	2018	2017	
349	58	1	02	017	A	colonial	1180900	6	5	4668	1195000	2013	2021
350	58	1	03	038	colonial	1164000	5	5	4253	1100000	2018	2019	
351	58	1	07	002	colonial	1043400	4	4	3865	450000	2018	2017	
352	58	1	27	012	colonial	1209500	4	4	3622	1252065	2015	2015	
353	58	1	27	013	colonial	1287900	4	5	3718	1200000	2012	2017	
354	58	1	27	016	colonial	1254100	4	6	3720	1341336	2016	2016	
355	58	1	27	050	colonial	1305200	4	4	3998	1243000	2014	2018	
356	58	1	27	067	colonial	1217500	4	4	3921	1155000	2013	2014	
357	58	2	11	01	101	condo_res	525200	2	2	1238	520000	2017	2020
358	58	2	11	01	102	condo_res	601500	3	2	1501	595000	2017	2018
359	58	2	11	01	103	condo_res	590300	2	2	1472	583800	2017	2018
360	58	2	11	01	104	condo_res	351500	1	1	751	298500	2017	2017
361	58	2	11	01	105	condo_res	351500	1	1	751	303400	2017	2018
362	58	2	11	01	108	condo_res	469800	2	2	1225	410000	2017	2021
363	58	2	11	01	109	condo_res	590300	2	2	1472	609800	2017	2018
364	58	2	11	01	110	condo_res	592800	3	2	1450	540000	2017	2020
365	58	2	11	01	111	condo_res	525200	2	2	1238	499000	2017	2018
366	58	2	11	01	201	condo_res	540500	2	2	1359	500000	2017	2020
367	58	2	11	01	202	condo_res	601500	3	2	1501	603500	2017	2018
368	58	2	11	01	203	condo_res	590300	2	2	1472	590000	2017	2021
369	58	2	11	01	204	condo_res	351500	1	1	751	305960	2017	2017
370	58	2	11	01	205	condo_res	351500	1	1	751	296400	2017	2017
371	58	2	11	01	206	condo_res	434600	2	2	1111	456000	2017	2021
372	58	2	11	01	208	condo_res	351500	1	1	751	301600	2017	2018
373	58	2	11	01	209	condo_res	590300	2	2	1472	543000	2017	2019
374	58	2	11	01	210	condo_res	612800	3	2	1450	600000	2017	2019
375	58	2	11	01	211	condo_res	514900	2	2	1333	499905	2017	2018

376	58	2	11	01	301	condo_res	540500	2	2	1359	535300	2017	2018
377	58	2	11	01	302	condo_res	626500	3	2	1501	628250	2017	2017
378	58	2	11	01	303	condo_res	590300	2	2	1472	563200	2017	2017
379	58	2	11	01	304	condo_res	351500	1	1	751	298760	2017	2017
380	58	2	11	01	305	condo_res	351500	1	1	751	299050	2017	2017
381	58	2	11	01	306	condo_res	449600	2	2	1111	450000	2017	2020
382	58	2	11	01	307	condo_res	351500	1	1	751	302370	2017	2018
383	58	2	11	01	308	condo_res	351500	1	1	751	301500	2017	2017
384	58	2	11	01	309	condo_res	590300	2	2	1472	595445	2017	2017
385	58	2	11	01	310	condo_res	612800	3	2	1450	601720	2017	2018
386	58	2	11	01	311	condo_res	514900	2	2	1333	474540	2017	2017
387	58	2	11	01	401	condo_res	540500	2	2	1359	519580	2017	2018
388	58	2	11	01	402	condo_res	601500	3	2	1501	597400	2017	2018
389	58	2	11	01	403	condo_res	615300	2	2	1472	622245	2017	2018
390	58	2	11	01	404	condo_res	351500	1	1	751	323725	2017	2018
391	58	2	11	01	405	condo_res	351500	1	1	751	309425	2017	2018
392	58	2	11	01	408	condo_res	614700	2	2	1535	577783	2017	2017
393	58	2	11	01	410	condo_res	612800	3	2	1450	617095	2017	2018
394	58	2	11	01	411	condo_res	514900	2	2	1333	491915	2017	2017
395	58	2	11	02	101	condo_res	430900	2	2	1101	448300	2017	2018
396	58	2	11	02	102	condo_res	590300	2	2	1472	575000	2017	2020
397	58	2	11	02	103	condo_res	601500	3	2	1501	549900	2017	2021
398	58	2	11	02	104	condo_res	351500	1	1	751	299795	2017	2018
399	58	2	11	02	105	condo_res	351500	1	1	751	303900	2017	2018
400	58	2	11	02	107	condo_res	474700	2	2	1178	455815	2017	2018
401	58	2	11	02	108	condo_res	601500	3	2	1501	574900	2017	2021
402	58	2	11	02	109	condo_res	590300	2	2	1472	565000	2017	2020
403	58	2	11	02	110	condo_res	426000	2	2	1088	396360	2017	2018
404	58	2	11	02	201	condo_res	530300	2	2	1388	509000	2017	2018
405	58	2	11	02	202	condo_res	590300	2	2	1472	545000	2017	2019
406	58	2	11	02	203	condo_res	636500	3	2	1501	641800	2017	2018
407	58	2	11	02	204	condo_res	351500	1	1	751	312495	2017	2018
408	58	2	11	02	205	condo_res	351500	1	1	751	338000	2017	2020
409	58	2	11	02	206	condo_res	403000	2	2	995	405000	2017	2021
410	58	2	11	02	207	condo_res	459700	2	2	1178	432745	2017	2018
411	58	2	11	02	208	condo_res	601500	3	2	1501	615000	2017	2021
412	58	2	11	02	209	condo_res	590300	2	2	1472	587500	2017	2021
413	58	2	11	02	210	condo_res	426000	2	2	1088	390910	2017	2018
414	58	2	11	02	301	condo_res	536400	2	2	1313	505000	2017	2019
415	58	2	11	02	302	condo_res	615300	2	2	1472	617180	2017	2018
416	58	2	11	02	304	condo_res	351500	1	1	751	300400	2017	2018
417	58	2	11	02	305	condo_res	351500	1	1	751	308315	2017	2018
418	58	2	11	02	306	condo_res	403000	2	2	995	376500	2017	2018
419	58	2	11	02	307	condo_res	459700	2	2	1178	436800	2017	2019
420	58	2	11	02	308	condo_res	636500	3	2	1501	645100	2017	2018
421	58	2	11	02	309	condo_res	590300	2	2	1472	610845	2017	2018
422	58	2	11	02	310	condo_res	426000	2	2	1088	401900	2017	2018
423	58	2	11	02	401	condo_res	561400	2	2	1313	560000	2017	2020
424	58	2	11	02	402	condo_res	590300	2	2	1472	612270	2017	2018
425	58	2	11	02	403	condo_res	601500	3	2	1501	613705	2017	2018
426	58	2	11	02	404	condo_res	651500	1	1	751	360000	2017	2021
427	58	2	11	02	405	condo_res	651500	1	1	751	360000	2017	2021
428	58	2	11	02	408	condo_res	651500	3	2	1501	652900	2017	2018
429	58	2	11	02	409	condo_res	590300	2	2	1472	620250	2017	2018

430	58	2	11	02	410	condo_res	441000	2	2	1088	422435	2017	2018
431	58	3	04	003	A	colonial	1136600	6	5	4306	999000	2014	2019
432		58	3	04	024	colonial	615400	4	4	4989	505000	2022	2020
433		58	3	04	035	colonial	1099300	4	5	4454	390000	2015	2013
434		58	3	21	001	colonial	1266400	4	5	8817	1266301	2020	2020
435		58	3	21	002	colonial	1211200	5	6	4330	1222750	2020	2020
436		58	3	21	003	colonial	1278800	5	6	4818	1297291	2020	2021
437		58	3	21	004	colonial	1261600	5	5	5115	1275198	2020	2020
438		58	3	21	005	colonial	1310900	5	6	5456	1427841	2020	2021