Undervalued or Overpriced? A Predictive Model of Calgary Property Assessments

Course: Data 603 Statistical Modelling with Data

Group Members:

1. Elise Beaupre, 30038889, elise.beaupre@ucalgary.ca
2. Michael Morgan, 30301873, michael.morgan@ucalgary.ca
3. Sundeep Parmar, 30301836, sundeep.parmar@ucalgary.ca
4. Yu Ling Wong, 30297454, yuling.wong@ucalgary.ca

---

1. Purpose

1.1 Domain:

This project examined the domain of municipal property assessment and urban valuation within the City of Calgary. Each year, the city evaluates all property parcels to establish fair market values for taxation and budgeting. These assessments reflect complex relationships between property characteristics, market trends, and location.

1.2 Investigation Goal:

By applying statistical modelling techniques, we aimed to better understand how these factors contribute to assessed property values and identify potential undervalued or overpriced areas. Our goal was to create a model that could predict the assessed value of farmland, residential, and retail properties based on the publicly available data from the City of Calgary.

We identified communities with below-average farmland or retail property values that may present attractive investment opportunities.

Additionally, we aimed to create a model that allowed us to examine how previous investments compare to similar properties. Are they overpriced or undervalued?

1.3 Motivation:

We considered ourselves in a position to consult clients on acquiring property. For example, we could work with a client looking to purchase a retail property to expand their business. Our predictive model helped the client make decisions with their investments.

Furthermore, we have found ourselves in the position of buying property for ourselves. Michael bought a condo last year and was interested in seeing how this purchase compared to similar properties.

1.4 Practical Implications:

For future home buyers, the model could help them determine how much they can expect to pay for their ideal home. For real estate investors, the model helps determine if a property is a good investment by comparing it to the predicted value of similar properties.

1.5 Population(s) of Interest:

We studied properties in Calgary; more specifically, we filtered all property parcels in Calgary down to three categories: farmland, residential, and retail.

1.6 Variables of Interest:

The target (response) variable chosen was *Assessed Property Value ($)*, which is continuous and quantitative. To investigate the effect of the location and type of property, the following qualitative variables were chosen: *City Quadrant, City Ward, Community, Property Type, and Property Use*. The quantitative predictors chosen were: *Age of Building, Land Size in Square Feet*, both of which are factors that are commonly considered when buying property.

1.7 Data Collection Method:

Yearly property assessments are conducted by the City of Calgary to allow for fair and equitable taxation. The Sales Comparison Approach was used to assess most properties, based on the property market value as of July 1 of the prior year, and the characteristics of the property as of December 31 (The City of Calgary, 2025c).

---

2. Data

2.1 Source of Data:

The dataset that we used was the "Current Year Property Assessments (Parcel)" from The City of Calgary's Open Data Portal (The City of Calgary, 2025a).

This dataset contains information on every residential, non-residential, and farm land property in Calgary for 2025, {see 2.2 breakdown}. There were 588,825 total rows in the set. Each row contains additional details that were also removed from modelling, such as: Roll Number {unique identifier}, Address {unique identifier}, and Multipolygon {Geospatial data}. The data is maintained for public interest and taxation purposes. The best efforts have been made to rule out any biases to favour one property or misrepresent areas within Calgary.

2.2 Variable of Interest:

Table 1

Variables of Property Assessments Dataset

Variable	Type	Domain
Assessed Property Value (ASSESSED_VALUE)	Integer	Market value used by the city of Calgary for taxation purposes [$250 - $1,939,810,000]
Property Address (ADDRESS)	String	Physical municipal address within Calgary [562,707 unique values]. {Note: used to construct Quadrant}
City Quadrant (QUADRANT)	String	Geographically defined quarter of Calgary [NE, NW, SE, SW]
Community Code (COMM_CODE)	String	Three-character abbreviation for a defined Calgary community [304 unique values]
Year of Construction (YEAR_OF_CONSTRUCTION)	Integer	Year when the building was completed and/or ready for occupancy [1886 - 2024]. {Note: used to construct Age}
Age of Building (AGE)	Integer	{Calculated} Value in years of a building since construction [1 - 139]

Property Assessment Class (ASSESSMENT_CLASS)	String	Two-character code designation of property assessment class [NR (Non-Residential), RE (Residential), FL (farmland)]
Property Type (PROPERTY_TYPE)	String	Two-character code designation of the property condition [LO (Land Only), IO (Improvement Only), LI (Land & Improvement)]
Sub-Property Type (SUB_PROPERTY_USE)	String	Classification to differentiate property usage and improvement type [185 unique values]
Property Land Size (LAND_SIZE_SF)	Integer	Total area or the size in square feet of an individual property. [1 sf - 27,643,993 sf]
Ward (WARD)	String	A geographical division within the City of Calgary made up of communities that elects a single counsellor [1 - 14] {Note: Calculated from Community Code}
Sector (SECTOR)	String	A geographical division within the City of Calgary to indicate the cardinal position of the property [North,NorthEast,Centre,South…] {Note: Calculated from Community Code}

2.3 Permission/Accessibility:

The Current Year Property Assessments dataset was available on the Open Calgary website, shared under the Open Government Licence - City of Calgary. According to the licence, users may "Copy, modify, publish, translate, adapt, distribute or otherwise use the Information in any medium, mode or format for any lawful purpose" (The City of Calgary, 2025b).

2.4 Time Frame of Data Collection:

The data represents the 2025 assessment year (based on July 1, 2024, market values and December 31, 2024, physical condition). There are adjustments made based on property sales and availability in the community in question.

---

3. Methodology

3.1 Focus of Statistical Investigation:

We aim to identify which property characteristics (e.g., community, property type, age) are the most significant predictors of assessed property value. Building on these insights, our final goal is to predict the assessed value of a property given the characteristics of the property.

3.2 Background/Context:

In the City of Calgary, the Assessment Business Unit applies assessments to properties annually for the purpose of evenly distributing taxation. This unit is responsible for determining the market value of all properties within the city limits. Property tax is the primary source of revenue for the city, funding essential services like police, fire, public transit, and infrastructure. (The City of Calgary, n.d.-a, n.d.-b)

3.3 Modelling Workflow

Property assessment data was examined to determine whether a predictive model could be developed. The dataset contained three major property categories: farmland, retail, and residential. Because the dataset was large and the characteristics of each category differed, we chose to build a model for only one property type. In other words, it was more practical to focus on predicting values within a single, more homogeneous group. To decide which category to use, we first fit preliminary models for each property type.

The division of labour and roles in the project were assigned equitably among co-authors with respect to individual areas of interest and skillsets. Michael was interested in modelling residential data because he bought a condo last year during a high-demand time. He was interested in seeing how the assessed value of his condo compared to others of the same age, sub-property type, etc. Sundeep was our data wrangler and prepared the large data sets for use in R. He also spearheaded the creation of presentation slides. Elise was interested in modelling retail data as she previously worked as Sales Manager for restaurants that were looking into expanding. YuLing conducted diagnostics on the final selected model. All contributed to writing, figure creation, and presenting.

After fitting individual models for residential and retail data, it was found that the retail model had the most potential for predictability, a better adjusted R^2 value, so this model was employed for further analysis and eventual prediction.

All p-values of statistical tests were compared to an alpha of 5%. See <u>*Appendix Figure A1*</u> for a flowchart of the steps taken.

<u>Initial Full Additive Model</u>

A multiple linear regression model for predicting property assessed value was fit including all possible predictors: AGE, ASSESSMENT_CLASS, COMM_CODE, QUADRANT, LAND_SIZE_SF, and SUB_PROPERTY_USE. An overall F-test revealed that at least one of the predictors is significant to be in the model.

<u>Multicollinearity Assumption Check</u>

The initial attempt at modelling assessed property values included COMM_CODE and QUADRANT as potential predictors. However, a variance inflation factor (VIF) test revealed high levels of multicollinearity due to the hundreds of dummy variables for communities and the redundancy in the data. Therefore, WARD was used instead of COMM_CODE, which vastly reduced the number of dummy variables. WARD and QUADRANT still contained redundant information, so QUADRANT (which resulted in a lower adjusted coefficient of determination than WARD) was dropped to avoid multicollinearity entirely.

Once multicollinearity was resolved, the improved full additive model included the predictors AGE, ASSESSMENT_CLASS, WARD, LAND_SIZE_SF, and SUB_PROPERTY_USE. Individual t-tests were performed to determine the significance of each potential independent variable. The stepwise selection procedure was employed and resulted in identical findings. The resulting reduced additive model for predicting assessed value was found to include the predictors AGE, WARD, LAND_SIZE_SF, and SUB_PROPERTY_USE.

<u>Interaction Terms</u>

Interaction terms were then investigated to see if any would add value to the model. Individual t-tests showed that all interaction terms were significant except for AGE*SUB_PROPERTY_USE. The resulting model, which was to become our final model, now included the interaction terms AGE*LAND_SIZE_SF, AGE*WARD, LAND_SIZE_SF* SUB_PROPERTY_USE, LAND_SIZE_SF*WARD, and SUB_PROPERTY_USE*WARD.

<u>Higher Order</u>

Higher-order terms were also examined for the quantitative variables AGE and LAND_SIZE_SF. However, none were found to be effective at improving model predictive performance (adjusted R^2) more than 1%, so they were not employed.

The final model suggested for the prediction of ASSESSED_VALUE contained the independent variables AGE, WARD, LAND_SIZE_SF, SUB_PROPERTY_USE, and the interaction terms AGE*LAND_SIZE_SF, AGE*WARD, LAND_SIZE_SF*SUB_PROPERTY_USE, LAND_SIZE_SF*WARD, and SUB_PROPERTY_USE*WARD. This final model with substituted regression coefficient values can be found in *Appendix Figure A2* and the final model can be found in *Appendix Figure A3*.

---

4. Statistical Methods and Results

4.1 Model Chosen

We will be using Multiple Linear Regression (MLR) to predict the assessed value of a property using the other attributes of the property.

4.2 Rationale:

Multiple Linear Regression is used when there are multiple predictors and the response variable (the variable that is to be predicted) is numerical. We are attempting to use multiple variables to predict the assessed value of properties in Calgary, which is a numerical variable which fits the MLR modelling technique.

4.3 Results of Modelling and Interpretation

Results of Initial Global F-test

The global F-test determines whether at least one of the predictors in the full model is significant when compared to the null model.

Table 2
Hypothesis Statements of the Global F-Test

Global F-Test	
Null Hypothesis	$\beta_1 = \beta_2 = ... = \beta_p = 0$
Alternative Hypothesis	*at least one β is not zero ($i = 1,2,...,p$)*

The results showed an F-value of 547.5 with a p-value of 2.2e-16. With an α of 0.05, the null hypothesis was rejected, and the conclusion was made that the model adjusted for multicollinearity (full model) was an improvement on the null model. As such, we were able to proceed with improving the model. See *Appendix Figure A4* for full results.

Individual T-Test on Coefficients

The individual t-test on the full set of predictors tests which independent variables are significant in the model.

Table 3
Hypothesis Statements of Individual T-Test

Individual t-test of Independent Variables	
Null Hypothesis	$\beta_k = 0$
Alternative Hypothesis	$\beta_k \neq 0$ ($k = 1, 2, \ldots, p$)

The results of this test showed a t-statistic of 0.355 and a p-value of 0.722731 for ASSESSMENT_CLASS, so there was not enough evidence to reject the null hypothesis. Assessment_Class was not significant with $\alpha = 0.05$. All other coefficients resulted in p-values less than 0.05. See *Appendix Table A1* for full results.

<u>Goodness-of-Fit Analysis on Additive Model</u>

The adjusted R^2 was examined to assess the proportion of variance explained by the predictors, adjusted for model complexity. The results of the procedure showed significant predictors and an adjusted R^2 of .8613 and RSE of 4,306,000 for the final first-order additive model for predicting ASSESSED_VALUE with the predictors AGE, LAND_SIZE_SF, SUB_PROPERTY_USE, and WARD.

<u>Results of Interaction Term Analysis and Final Goodness-of-Fit</u>

Interactions among independent variables were vetted and found to be significant except for AGE*SUB_PROPERTY_USE. After adding significant interaction terms, the performance of the model improved to an adjusted R^2 of 0.9155 and RSE of 3,361,000.

<u>Coefficient Interpretation</u>

The final model with interaction terms included many levels of categorical variables, so the coefficient interpretation in this report was limited to a selection of submodels.

Given a property in WARD 2 that had the SUB_PROPERTY_USE code "CM0203" for “retail - shopping centre neighbourhood”, the values were plugged in with the coefficients from the model output, as seen in the full formula located in *Appendix Figure A2* and referring to the final model found in *Appendix Figure A3*. The marginal effect for age was found to be 43,660.70 - 0.59*LAND_SIZE_SF. For every one year increase in age of building, the assessed property value increases by (43,660.70 - 0.59*LAND_SIZE_SF) dollars.

Now, consider a property in WARD 1 that had the SUB_PROPERTY_USE code "CM0201" for “retail - freestanding”. After reducing the formula, for every one-year increase in age of building, the assessed property value increases by (24,635.57 - 0.59*LAND_SIZE_SF) dollars. For every one foot increase of LAND_SIZE_SF, the assessed property value increases by (82.55 - 0.59*AGE).

<u>Vetting Assumptions and Diagnostics</u>

a) <u>Linearity</u>

The residual plot was utilized to investigate the linearity assumption between the predictors and the response variable. Based on the residual plot, linearity was not observed, as a noticeable curve was observed toward the tail of the line. To address this issue, an attempt was made to improve the model with a quadratic term for AGE. However, AGE became insignificant, and the adjusted R^2 did not improve; hence, higher order terms were not pursued. A log transformation on AGE was also attempted, but it led to the same conclusion. As a result, the linearity assumption was not met for this dataset.

b) <u>Independent Errors</u>

No time series data have been used in this model, but there are groups and geographical variables - WARD and SUB_PROPERTY_USE. Two boxplots were generated to check the independence assumption of these two variables (Residuals vs WARD and Residuals vs SUB_PROPERTY_USE), see *Appendix Figure A8*. It was found that the medians are close to zero for both plots, concluding no strong evidence of correlated errors. Additionally, since both variables are treated as categorical predictors with dummy variables, the independent errors condition is not violated.

c) <u>Equal Variance</u>

A residual plot and a scale-location plot were utilized to visualize the spread of the residuals, see *Appendix Figure A5 and Figure A6*. It was noticed that there appears to be some funnelling pattern in the residual plot, and the scale-location plot was not exactly horizontal, indicating heteroscedasticity may be present. To further verify this finding, a Breusch-Pagan Test was conducted. It was found that the p-value was 2.2e-16, which is less than 0.05, hence we reject the null hypothesis, concluding heteroscedasticity is indeed present.

To deal with the presence of heteroscedasticity, a log transform on the response variable (ASSESSED_VALUE) was performed for simplicity's sake, and the assumption was retested with the Breusch-Pagan Test. Another method to deal with heteroscedasticity is to utilize the Box-Cox transformation method. Since the response variable consists of only positive values, we can proceed with the Box-Cox method, as the transformation method can only be used on strictly positive response variables. The best lambda was found to be at 0.0707, and it was utilized to proceed with the transformation.

It seems that the p-values for the Breusch-Pagan Test remained at 2.2e-16, less than 0.05 for both transformation tests, which we fail to reject the null hypothesis, concluding heteroscedasticity is still present. The log transformation and Box-Cox Transformation were not helpful for this dataset.

Since both attempts to transform the response variable to reduce its variation failed, there is a strong signal that the underlying issue may not be the scale of the variable, but stems from something deeper in the model structure or the suitability of the chosen regression method.

d) <u>Normality</u>

The normality assumption was checked via plotting the normal probability plot (qqplot) and conducting the Shapiro-Wilk test on the residuals data. See *Appendix Figure A9* for the normal probability plot. The normal probability plot indicated that the residuals were not normally distributed, with the presence of heavy-tails that deviate significantly from the diagonal reference line, suggesting excessive kurtosis. Consistently, the Shapiro-Wilk test also concluded a p-value of less than 0.05, verifying the residuals are not normally distributed. Therefore, the dataset does not satisfy the normality assumption.

e) <u>Outliers</u>

Cook's distance was calculated to check for potential outliers within the dataset that might distort the residuals' distribution and variance. A Cook's distance plot was also computed, see *Appendix Figure A10*. It was found that 3 data points have a Cook's distance of value more than 0.5 (#738, 750, 927). These outliers were removed from the original dataset, and the model was subsequently rebuilt. The final valid model did not show an improvement in adjusted R^2 value; instead, it demonstrated a decrease in 0.02% adjusted R^2 value. Furthermore, the assumptions of homoscedasticity and normality were still not met. Given the absence of improvement in model performance and unresolved assumption violations, there is no strong evidence to justify the removal of the three outliers.

---

5. Result, Conclusion and Discussion

5.1 Results

<u>Final Model</u>
At the end of the model fitting and improvement process, an adjusted R^2 of 0.9155 and RSE of 3,361,000 were achieved, indicating the model explained over 91% of the variability in the assessment data.

Prediction Scenario

A scenario was considered to test the predictive capabilities of the model.

Suppose Sunny were interested in investing in a 2000 square-foot retail space near the University of Calgary to open a restaurant in a strip mall (SUB_PROPERTY_USE = CM0210) near the University of Calgary (WARD = 7). A building 5 years old would be sufficient. According to our model, the 95% confidence interval for the predicted assessed value of such a property would be as follows:

$-5,986,756 ≤ Assessed Value (est.) ≤ $7,469,687, with a fit estimate of $741,465.40.

The size of the confidence interval was surprising. Since the confidence interval is large and spans 0, it does not indicate a good prediction. One possible reason for this might be due to the violation of the normality assumption, as normality is crucial when it comes to computing an approximately correct prediction interval in a linear regression model (Chihara & Hesterberg, 2011).

Regression Assumptions Results

It appears that the majority of the assumptions were not met for the final model we have selected to predict the assessed value of retail properties in Calgary, specifically, the linearity, homoscedasticity, and normality assumptions. A linear relationship was not observed between the predictors and the response variable, and the residuals of the regression were not normally distributed. Heteroscedasticity was also found to be present for this model. Given the limitations of knowledge and time, future works such as exploring alternative regression techniques can be involved to better address these assumptions.

5.2 Evaluation of Approach

With a final adjusted R^2 of 0.9155, our approach appeared promising, but more could be done to improve upon our work. This was particularly evident in the predictive performance of our model. Perhaps this had something to do with our model not meeting several of the assumptions required for a Multiple Linear Regression Model.

In the context of residential data, while city property assessments provide several attributes about a property, they do not provide many qualities of a property that may come to mind. The assessments do not include the number of bedrooms or bathrooms, surely an important part of any home search. Additionally, when it comes to condos, the square footage in the data was the footprint of the condo building, not the square footage of each condo. So, for these units, square footage was likely not contributing much to the predictive performance of the model. This combination of omissions in the data likely contributed to poor model performance for residential properties.

5.3 Future Work

One avenue for future work would be to expand into residential data so that a potential home buyer could use the model to estimate how much their ideal home should cost. The reason we were not able to do that at this time was due to a lack of detail in the property assessment data. To properly model the value of residential properties, it would be necessary to obtain the number of bedrooms, the number of bathrooms, and more accurate square footage for each assessed property. Sources of this data could potentially come from home sale listings.

Additionally, to resolve assumptions that were not met for our model, alternative or more advanced regression methods can be explored.

---

6. References

Chihara, L. M., & Hesterberg, T. C. (2011). *Mathematical Statistics with Resampling and R*. John Wiley & Sons.

The City of Calgary. (n.d.-a). *Assessment & Tax*. City Departments and Business Unit Highlights. Retrieved November 3, 2025, from https://www.calgary.ca/our-services/assessment-tax.html

The City of Calgary. (n.d.-b). *Property tax breakdown*. Property Assessment and Tax. Retrieved November 3, 2025, from https://www.calgary.ca/property-owners/taxes/service-breakdown-calculator.html

The City of Calgary. (2025a). *Current Year Property Assessments (Parcel)* [Dataset]. https://data.calgary.ca/Government/Current-Year-Property-Assessments-Parcel-/4bsw-nn7w/about_data

The City of Calgary. (2025b). *Open Government Licence—City of Calgary*. https://data.calgary.ca/stories/s/Open-Calgary-Terms-of-Use/u45n-7awa

The City of Calgary. (2025c). *Residential property assessment*. How We Assess Properties. https://www.calgary.ca/property-owners/assessment/property-types.html

7. Appendix

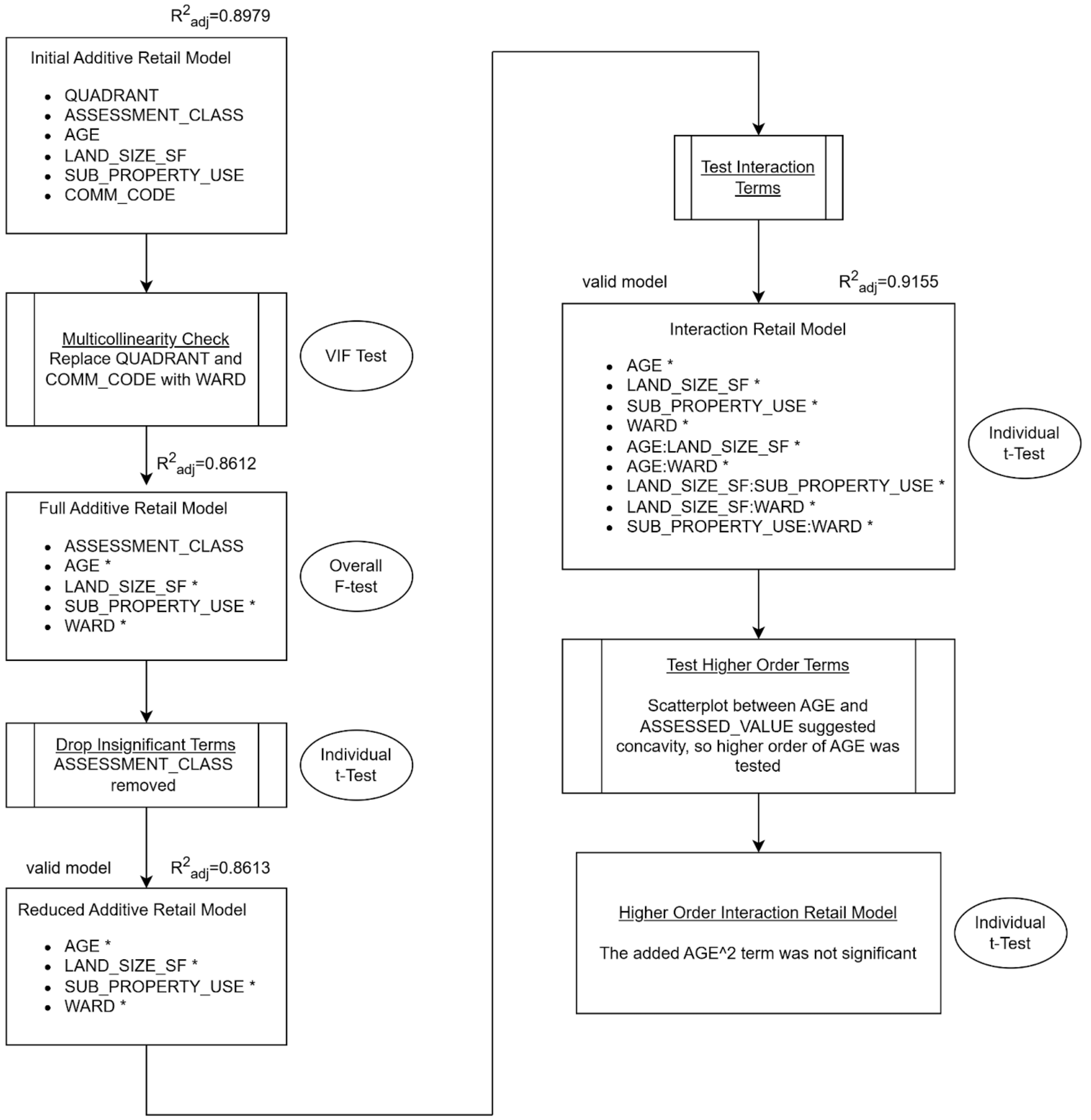
Table A1

Results of ANOVA. Full Model vs. Null Model

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-Statistic	P-Value
Regression	18	1.8282e+17	1.0156e+16	547.53	2.2e-16
Residual	1567	2.9067e+16	1.8443e+13		
Total	1585	2.1188e+17			

Figure A1

Flowchart of Model Fitting Workflow



Note. Flowchat was created using drawio (https://www.drawio.com/)

Figure A2

Final Retail Model with Regression Coefficients

$$
\begin{aligned}
\widehat{ASSESSED\_VALUE} = & -659862.81 + 24635.57\, AGE + 82.55\, LAND\_SIZE\_SF \\
& + 4289961.96\, SUB\_PROPERTY\_USE_{CM0203} + 1429833.90\, SUB\_PROPERTY\_USE_{CM0210} \\
& + 1266968.33\, WARD2 + 389144.29\, WARD3 + 1314591.50\, WARD4 + 3631597.85\, WARD5 \\
& + 4781321.76\, WARD6 + 1332096.45\, WARD7 + 1611623.48\, WARD8 + 1521600.51\, WARD9 \\
& + 849678.60\, WARD10 - 1229828.54\, WARD11 + 2414506.26\, WARD12 - 3174028.20\, WARD13 \\
& + 3599560.09\, WARD14 - 0.59\, AGE * LAND\_SIZE\_SF + 19025.19\, AGE * WARD2 \\
& + 300673.21\, AGE * WARD3 - 17962.50\, AGE * WARD4 - 97720.39\, AGE * WARD5 \\
& - 88895.7208\, AGE * WARD6 - 26584.17\, AGE * WARD7 - 28624.46\, AGE * WARD8 \\
& - 22608.52\, AGE * WARD9 + 40479.83\, AGE * WARD10 + 28235.58\, AGE * WARD11 \\
& - 123030.49\, AGE * WARD12 + 88615.60\, AGE * WARD13 - 103145.93\, AGE * WARD14 \\
& + 18.74\, LAND\_SIZE\_SF * SUB\_PROPERTY\_USECM0203 \\
& + 24.16\, LAND\_SIZE\_SF * SUB\_PROPERTY\_USECM0210 - 18.20\, LAND\_SIZE\_SF * WARD2 \\
& - 50.52\, LAND\_SIZE\_SF * WARD3 - 3.19\, LAND\_SIZE\_SF * WARD4 \\
& - 16.01\, LAND\_SIZE\_SF * WARD5 + 24.42\, LAND\_SIZE\_SF * WARD6 \\
& + 45.39\, LAND\_SIZE\_SF * WARD7 + 39.70\, LAND\_SIZE\_SF * WARD8 \\
& - 7.71\, LAND\_SIZE\_SF * WARD9 - 24.92\, LAND\_SIZE\_SF * WARD10 \\
& + 20.73\, LAND\_SIZE\_SF * WARD11 - 9.38\, LAND\_SIZE\_SF * WARD12 \\
& - 14.06\, LAND\_SIZE\_SF * WARD13 + 5.05\, LAND\_SIZE\_SF * WARD14 \\
& + 6127344.88\, SUB\_PROPERTY\_USE_{CM0203} * WARD2 \\
& - 809634.52\, SUB\_PROPERTY\_USE_{CM0210} * WARD2 + 9226911.93\, SUB\_PROPERTY\_USE_{CM0203} * WARD3 \\
& - 6441240.77\, SUB\_PROPERTY\_USE_{CM0210} * WARD3 - 4046462.30\, SUB\_PROPERTY\_USE_{CM0203} * WARD4 \\
& - 942109.43\, SUB\_PROPERTY\_USE_{CM0210} * WARD4 - 3561478.96\, SUB\_PROPERTY\_USE_{CM0203} * WARD5 \\
& - 2029385.54\, SUB\_PROPERTY\_USE_{CM0210} * WARD5 - 3169180.94\, SUB\_PROPERTY\_USE_{CM0203} * WARD6 \\
& - 1435336.77\, SUB\_PROPERTY\_USE_{CM0210} * WARD6 + 1293961.30\, SUB\_PROPERTY\_USE_{CM0203} * WARD7 \\
& - 1649105.02\, SUB\_PROPERTY\_USE_{CM0210} * WARD7 - 999201.31\, SUB\_PROPERTY\_USE_{CM0210} * WARD8 \\
& - 3773041.08\, SUB\_PROPERTY\_USE_{CM0203} * WARD9 - 1742033.19\, SUB\_PROPERTY\_USE_{CM0210} * WARD9 \\
& - 4563501.71\, SUB\_PROPERTY\_USE_{CM0203} * WARD10 - 2956622.88\, SUB\_PROPERTY\_USE_{CM0210} * WARD10 \\
& - 3197816.59\, SUB\_PROPERTY\_USE_{CM0203} * WARD11 - 1136865.72\, SUB\_PROPERTY\_USE_{CM0210} * WARD11 \\
& - 389667.37\, SUB\_PROPERTY\_USE_{CM0203} * WARD12 - 753189.58\, SUB\_PROPERTY\_USE_{CM0210} * WARD12 \\
& + 3129949.30\, SUB\_PROPERTY\_USE_{CM0203} * WARD13 - 3746320.85\, SUB\_PROPERTY\_USE_{CM0203} * WARD14 \\
& - 2139408.05\, SUB\_PROPERTY\_USE_{CM0210} * WARD14
\end{aligned}
$$

Note. Figure illustrates model equation that was computed in R Markdown

Figure A3

Final Model

$$
\begin{aligned}
\widehat{\mathrm{ASSESSED\_VALUE}} = & \beta_1 + \beta_2\, \mathrm{AGE} + \beta_3\, \mathrm{LAND\_SIZE\_SF} + \beta_4\, \mathrm{SUB\_PROPERTY\_USE} + \beta_5\, \mathrm{WARD} + \\
& \beta_6\, \mathrm{AGE{*}LAND\_SIZE\_SF} + \beta_7\, \mathrm{AGE{*}WARD} + \beta_8\, \mathrm{LAND\_SIZE\_SF{*}SUB\_PROPERTY\_USE} + \\
& \beta_9\, \mathrm{LAND\_SIZE\_SF{*}WARD} + \beta_{10}\, \mathrm{SUB\_PROPERTY\_USE{*}WARD}
\end{aligned}
$$

Note. Figure illustrates model equation that was computed in R Markdown

Figure A4

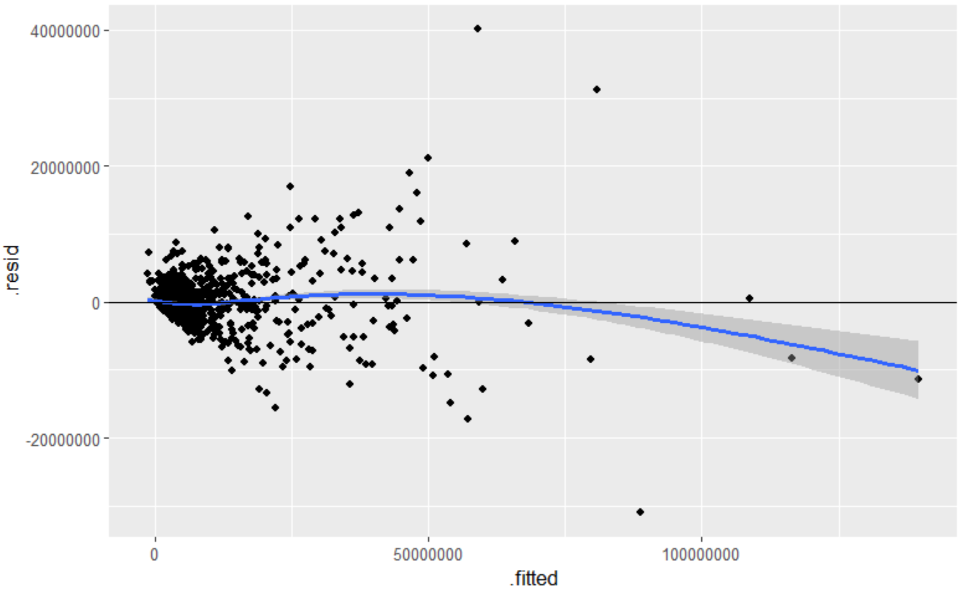
Results of Individual T-Test

```
Call:
lm(formula = ASSESSED_VALUE ~ ASSESSMENT_CLASS + AGE + LAND_SIZE_SF +
    SUB_PROPERTY_USE + WARD, data = retail3)

Residuals:
      Min        1Q    Median        3Q       Max
-34981367  -1219429   -156442    943773  46939057

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             2.172e+06  5.467e+05   3.972 7.44e-05 ***
ASSESSMENT_CLASSRE      6.891e+05  1.942e+06   0.355 0.722731
AGE                    -2.217e+04  6.155e+03  -3.602 0.000326 ***
LAND_SIZE_SF            7.499e+01  9.966e-01  75.250  < 2e-16 ***
SUB_PROPERTY_USECM0203  3.266e+06  3.991e+05   8.182 5.71e-16 ***
SUB_PROPERTY_USECM0210  8.305e+05  2.504e+05   3.317 0.000932 ***
WARD2                   2.020e+06  1.073e+06   1.883 0.059842 .
WARD3                  -6.780e+06  7.580e+05  -8.944  < 2e-16 ***
WARD4                  -1.121e+06  6.059e+05  -1.849 0.064590 .
WARD5                  -2.216e+06  7.361e+05  -3.011 0.002646 **
WARD6                   2.821e+06  7.196e+05   3.920 9.23e-05 ***
WARD7                  -2.342e+03  5.262e+05  -0.004 0.996449
WARD8                   1.981e+05  6.154e+05   0.322 0.747616
WARD9                  -1.278e+06  5.134e+05  -2.489 0.012904 *
WARD10                 -4.090e+06  5.833e+05  -7.013 3.46e-12 ***
WARD11                  2.465e+05  5.760e+05   0.428 0.668753
WARD12                 -2.091e+04  8.080e+05  -0.026 0.979360
WARD13                 -9.899e+05  1.003e+06  -0.987 0.323633
WARD14                 -3.826e+05  7.478e+05  -0.512 0.608960
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4307000 on 1567 degrees of freedom
Multiple R-squared:  0.8628,    Adjusted R-squared:  0.8612
F-statistic: 547.5 on 18 and 1567 DF,  p-value: < 2.2e-16
```

Note. Result computed with R.

Figure A5

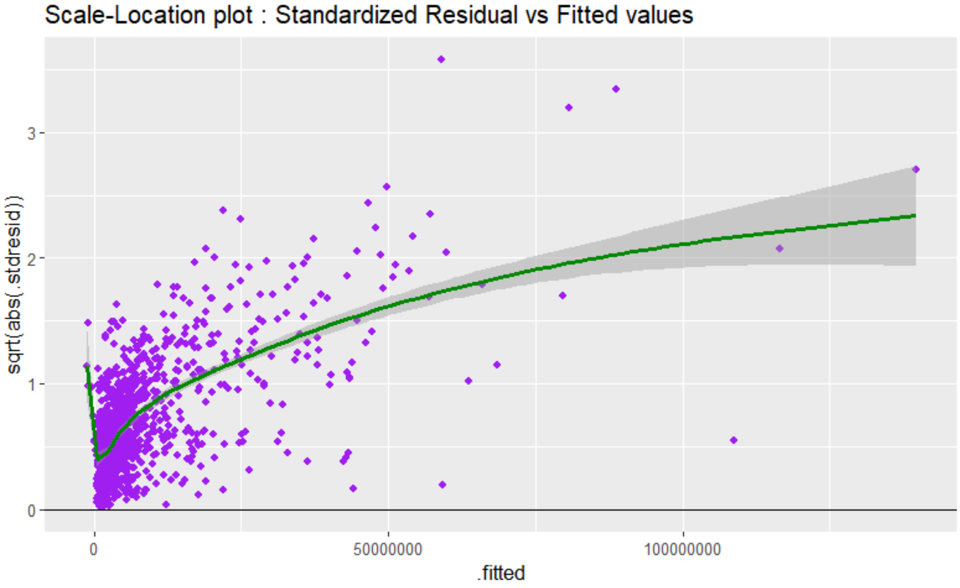
Residuals Plot of Final Model



Note. Plot computed with R.

Figure A6

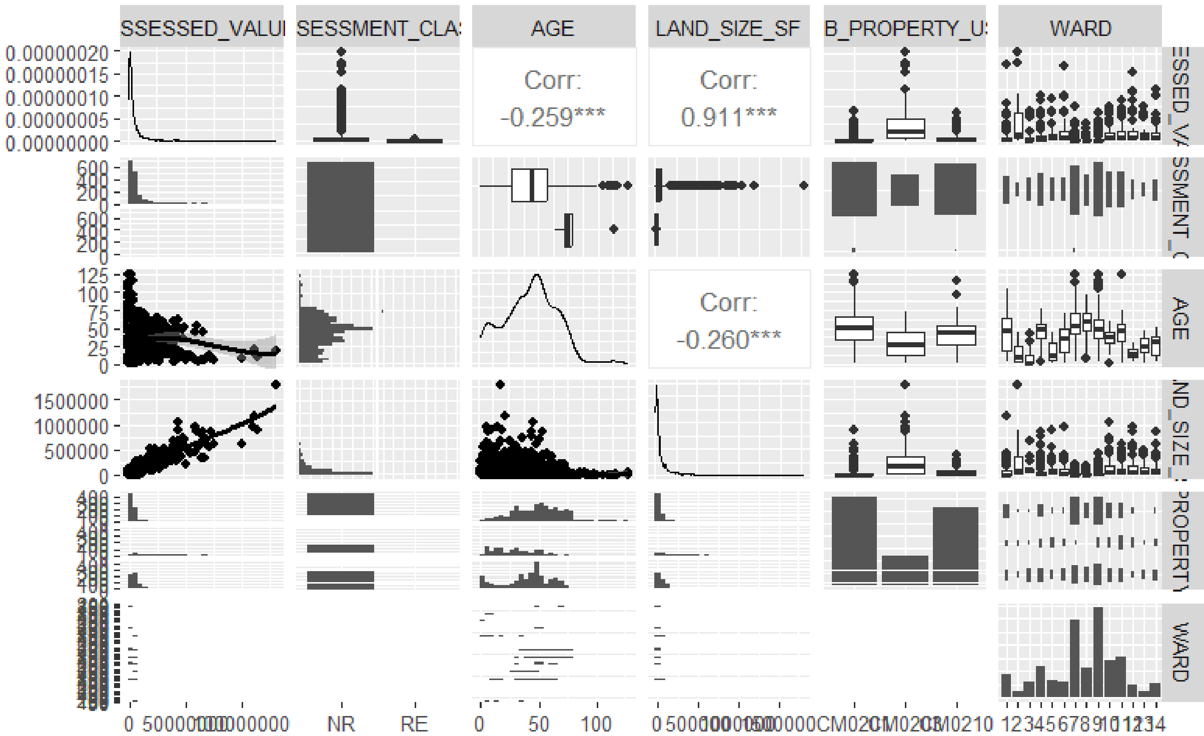
Scaled-Location Plot of Final Model



Note. Plot computed with R.

Figure A7

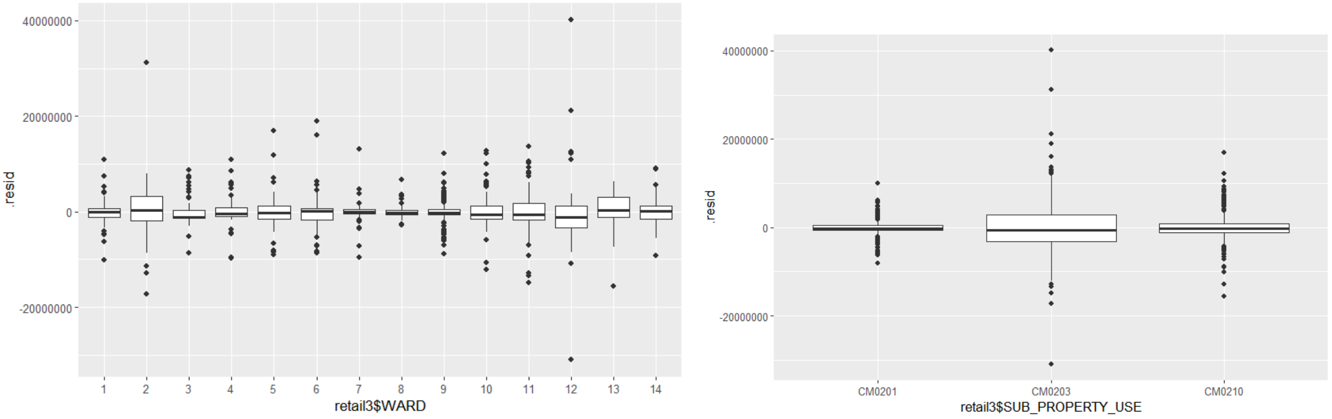
Pair Plots of Property Assessment Dataset



Note. Plots computed with R.

Figure A8

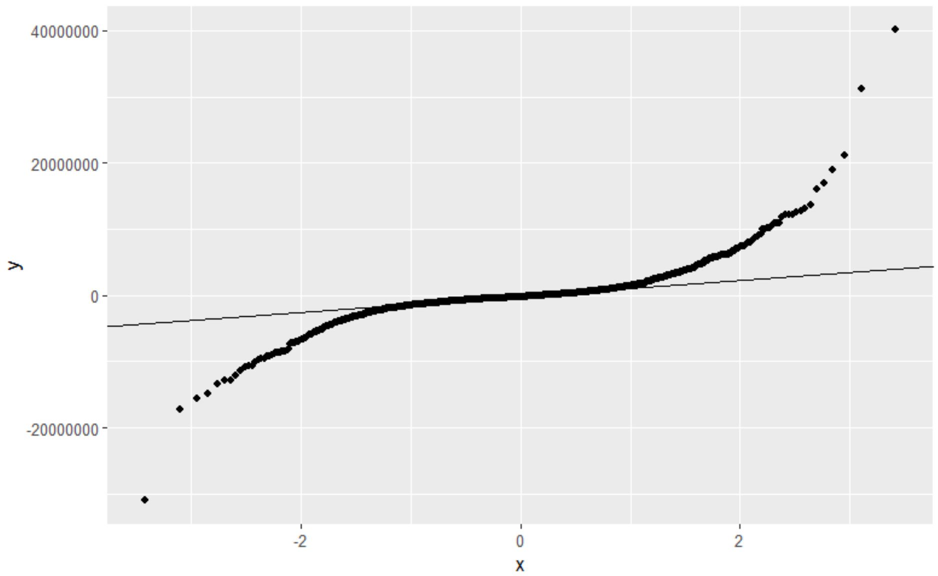
Boxplots of residuals vs individual predictor (WARD, SUB_PROPERTY_USE)



Note. Plots of residuals vs WARD (left), residuals vs SUB_PROPERTY_USE (right). Both computed with R.

Figure A9

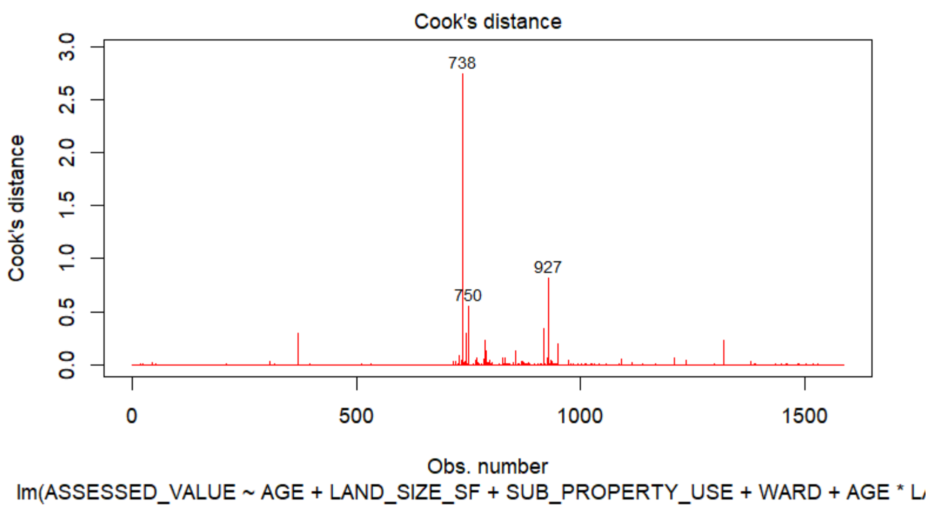
Normal Probability Plot (qqplot) of Residuals for Final Model



Note. Plot computed with R.

Figure A10

Cook's Distance Plot for Final Model



Note. Plot computed with R.