Chicago Triad Valuation Report

Contents

Introduction

The reader is provided with a description of the process employed for the estimation of 2018 Tax Year values for the residential properties in the Chicago Triad. The following topics are presented for each Township in the Triad:

1. The sales used as the basis for the value estimates
2. The choice of Linear Additive vs Multiplicative Models
3. A discussion of the nature of and the role multiple regression analysis (MRA) plays in the valuation process
4. Comparable Sales Analysis process
5. Valuation Results
6. Sales Analysis

The next section applies generally to all the townships in the Chicago Triad.

Generally Available Data

CCAO

CCAO provides valuation and analysis data in the form of an SPSS .sav file for each township. The number of data fields is extensive and is summarized in Appendix A Variable Definitions.

GIS

Several of the techniques described herein make use of location coordinates in spatial analytical techniques. As such, Parcel Polygon shapefiles were downloaded from the Cook County Open Data Portal for use in the analysis.

Census Bureau Data

It was hypothesized that owner occupancy level may have an influence on the value of housing stock. As such digital and geographic data were obtained from Census.gov that provided owner occupancy data on the block group level.

Modeler's Comments

During the analysis for the Chicago Triad, it became clear that the database is a limiting factor on the quality of the valuation results. Five very important variables are cited and the limiting condition on each is provided:

1. Living area – in this list simply to note that it is clearly an important variable necessary for establishing value.
2. Quality of construction – this is normally the second or third most important variable in a valuation model. In Cook County, this variable is of no use because it exhibits virtually no variability. That is, approximately 99% of the properties are "average" quality. The other aspect of the quality variable is that it can take on only one of three values. Industry practice is to use five or more quality levels. Using twenty-one levels is not uncommon.

3. Condition – this also a very important variable. In Cook County, this variable is of no use because it exhibits virtually no variability. That is, approximately 99% of the properties are in “average” condition. Again, there are only three possible choices for condition. Typically, eight or more condition levels are used in other jurisdictions.
4. Effective Age vs. Age – It is easy to compute the age of a dwelling. What is more difficult is the concept of effective age. If a property has undergone modernization, its effective age is different from its actual age. Effective age is unavailable in Cook County. Upon asking about the use of building permit information as a means of keeping the housing inventory up to date, the answer was in the negative. (A “findings message was issued 04/01/2018). It is recommended that Effective age should be established and maintained.
5. Location – this variable was used in the analysis and valuation process. The publicly available GIS was used as a source for the location of each parcel. There are many parcels not in the GIS. Methods had to be devised to approximate the location of these parcels. Having the GIS and the parcel database in synch is most common in other jurisdictions.
6. A Location Influence factor was devised to overcome some of the deficiencies in the database. It proved to be a very significant variable that improved the accuracy of prediction by a considerable measure.

Comment: The condition rating of a property is with tied to its age or effective age. A brand-new home is defined to be in Average condition (what is expected for a new home). If the condition of a 60-year-old property is Average, it is with respect to its age. The “Average” rating is therefore what is typical for the age of the home and not invariant across age. There is more to this topic, but its discussion involves more detail than this report contemplates as its scope.

Determining the reasons for these deficiencies was beyond the scope of this effort. If asked for an opinion, this modeler leans toward lack of resources to keep the database up to date as opposed to inefficient use of existing resources. This is based on observing the size of the assessment staff in relation to the number of parcels in other jurisdictions in North America over a period spanning five decades.

Regarding the Quality Variable: a limited scope trial has been proposed in which publicly available images would be used to improve the quality (construction grade) data. It would involve a small number of sale properties. If, as hoped, valuation accuracy can be improved using the improved quality data, the question of moving forward with the approximately 1.5 million residential properties would be very interesting and important, but not a part of this scope of work.

The above discussion does not apply to condominiums. The reason is that <u>there is no condominium data available to support industry standard valuation methods</u>. At a minimum, location, living area, floor level and view are needed for mass appraisal of condominiums. CCAO has necessarily had to devise alternate methods of condominium valuation. Other than this statement, there is nothing further about condominium valuation in this report.

The remainder of the report presents:

- the detail for each of the eight townships in the Chicago Triad
- A set of appendices on methodology that is used commonly in each of the eight townships

Rogers Park

Summary

Key points about Rogers Park Modeling and Valuation

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- A holdout sample was used to validate that models were not being overfit
- Outlier identification was based on Statistically-based and defensible methods
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- The Location factor variable was significant and helped improve accuracy
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- These tests revealed no spatial pattern of over or under valuations
- Rogers Park was valued using the Comparable Sales direct market comparison method of valuation
- Performance measures of accuracy are superior to traditional modeling methods and well within IAAO standards

The Data

Sales Counts

The CCAO provided a file in SPSS. sav format for model building and valuation. The total number of records in the file was 7,660. Of those records 5,481 had a recorded sale amount. Certain procedural steps established by CCAO were taken to identify the candidate sales records for sales analysis.

Use if price>$100,000 and <$990,000 takes sales count from 5481 to 4816
select single family reduces count of sales to 4808
select if sale year>2012 reduces sales count to 1281
Using open market sales reduces sales count from 1281 to 926
926 is base count.

Data Fields

The initial list of data items available for analysis is provided in [Appendix A Variable Definitions.docx.](Appendix A Variable Definitions.docx) Certain additional data fields were created. Those that were relevant to Rogers Park are:

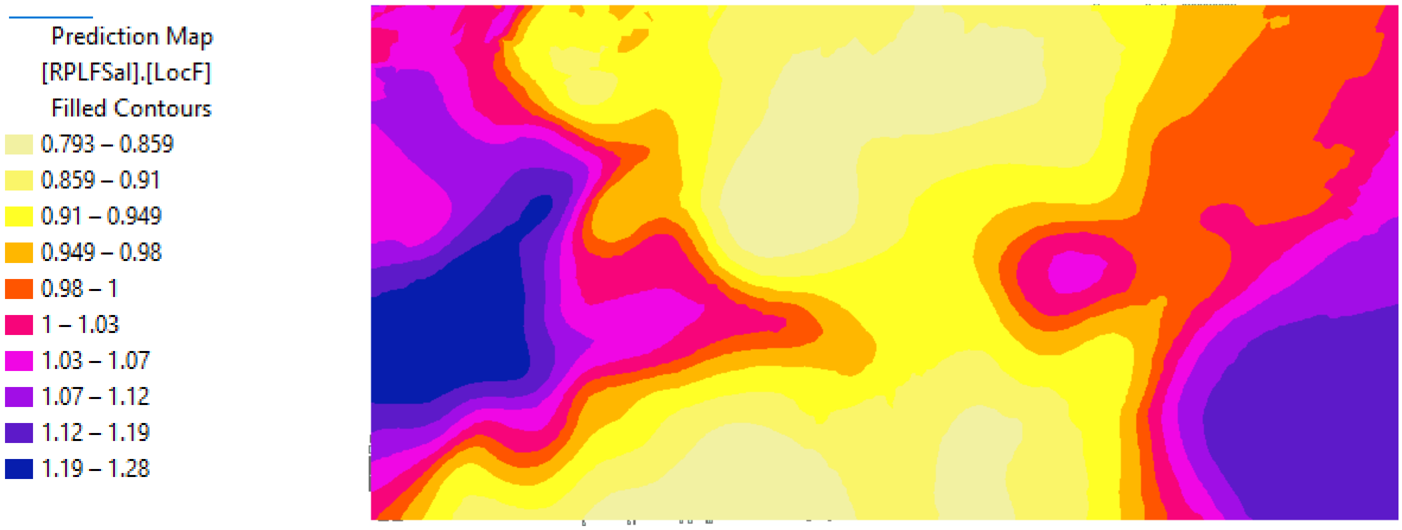
Location Factor

A location factor was derived by use of Geographically Weighted Regression (GWR). The process is one in which a model with a small number of variables not including spatial regime variables is calibrated. The resulting coefficient set is then used to value a "market basket home". The result is the value of the same home moved around the jurisdiction in question, called "market basket value". The actual value is arbitrary and depends on the chosen characteristics of the market basket home. The figure depicts the market basket value using proportional symbols.

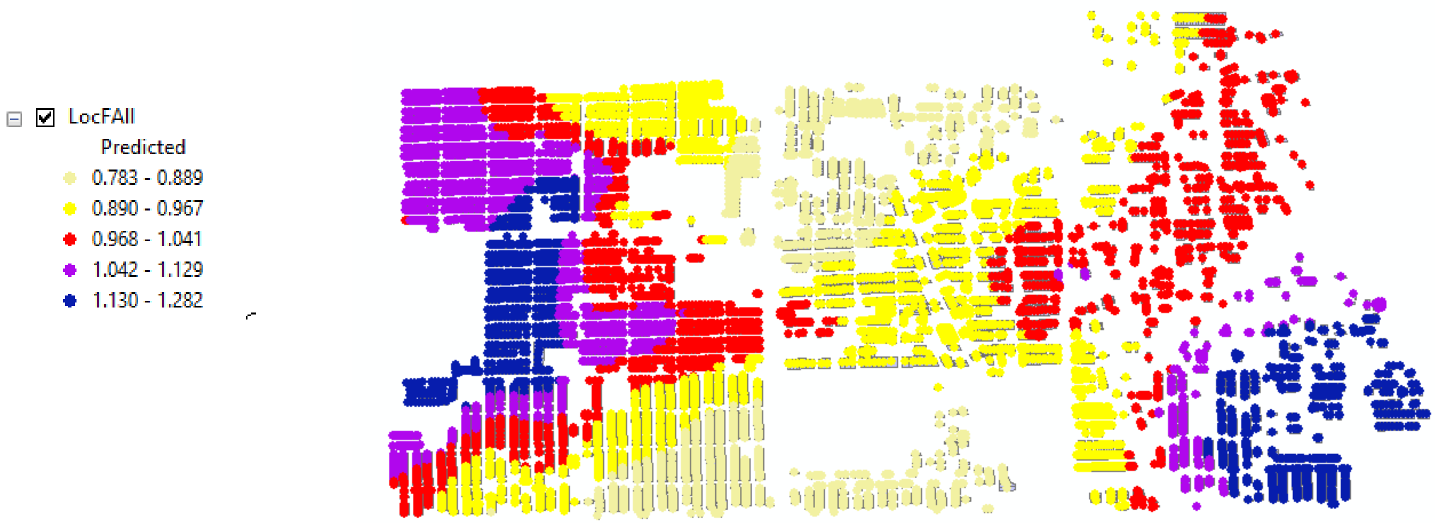
The Location factor is simply the market basket value divided by the average market basket value. The thematic map would look the same, but with a different scale.

The issue is applying the location factor derived from the sales to all properties needing to be valued. The solution is to develop a spatially averaged location factor surface and then to apply that to the inventory of properties to be valued. The method used to do this is called “Kriging” or in this case Universal Kriging.

The resultant surface and thematic legend are shown in the image below.

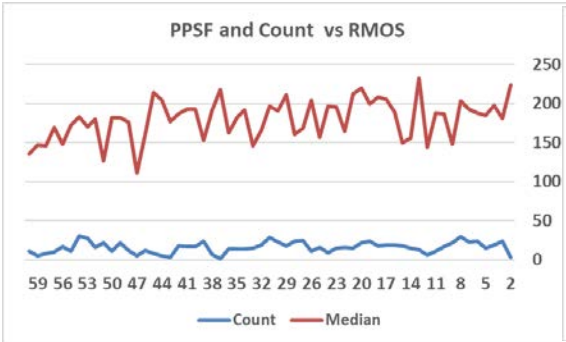


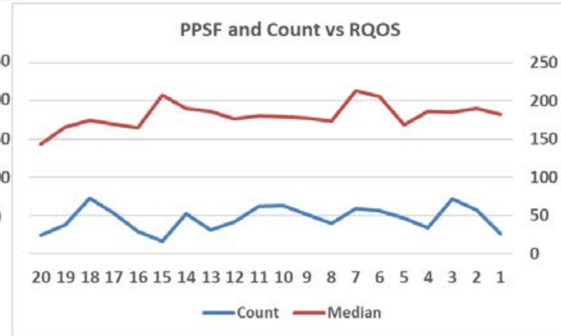
When applied to all properties the thematic map of Location Factor is given in the next image.



Reverse Quarter of Sale

The sales used in the analysis span a period of five years. To allow for time trending the sales to the valuation date, a reverse month of sale is computed. If the sale took place in December of 2017, the reverse month of sale (RMOS) is 1, November 2017, RMOS is 2, all the way back to January of 2013, RMOS is 60. In terms of using this variable directly in the model to be discussed, it is converted into a Reverse Quarter of Sale (RQOS). The rationale for this approach can be seen in the following two charts. The first shows Price per Square Foot (PPSF) and Count vs. RMOS. The second chart shows the same two variables vs RQOS. The RMOS variable is too granular and “noisy” as compared to the RQOS variable. Therefore, RQOS was used in the model building process as one of the independent variables.



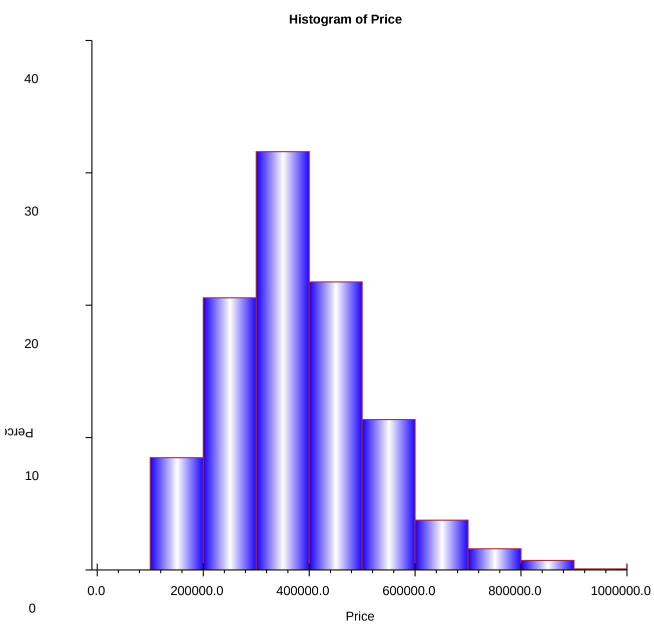


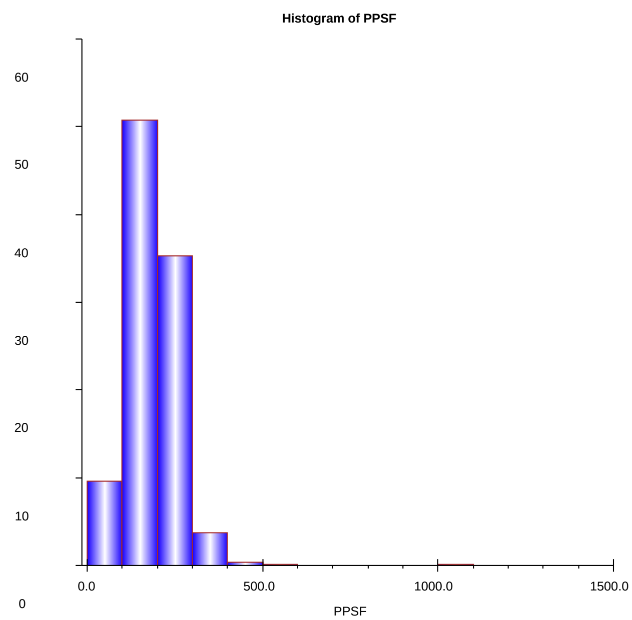
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. A sampling of the candidate variables follows.

Price and Price per Square Foot

The Price histogram is unremarkable, but the Price per Square Foot (PPSF) histogram looks as if it may have an outlier above $1,000/sq. ft. The property sold for $900,000 has 858 square feet of living area, three bedrooms, one and half baths on a 3,000 square foot lot. Note that PPSF is examined to gain an understanding of the market but is not a candidate variable in a regression model.

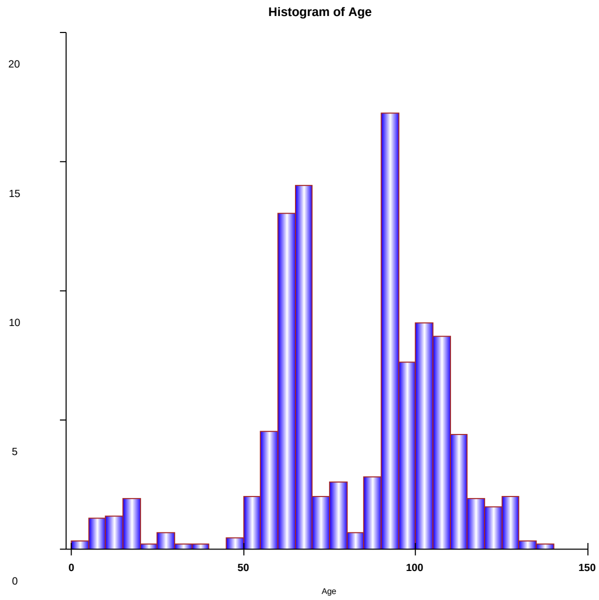


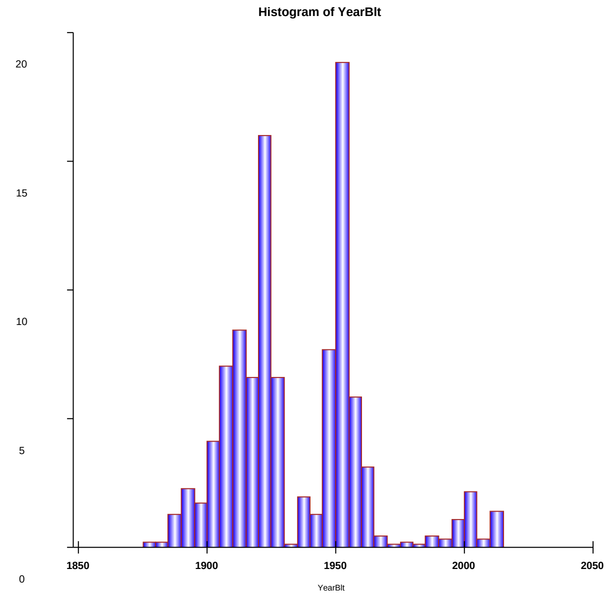


Year Built/Age

Consider the next two charts which are really representations of the same information. The first histogram is for Age. The second, year built. In this case, Age=2018-Year Built. The Age chart is informative, but the Year Built chart provides more a of a sense of the history of building in Rogers Park. There are three distinct building phases evident. They are:

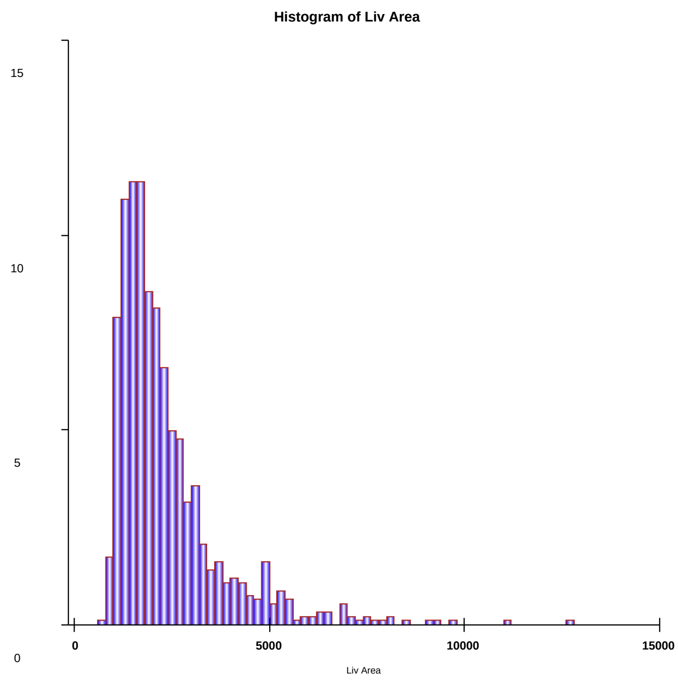
1. The run up to the depression years and the subsequent collapse in building starts
2. Post WW2 building boom
3. The market spurt and crash 2000-2010.





Building Size

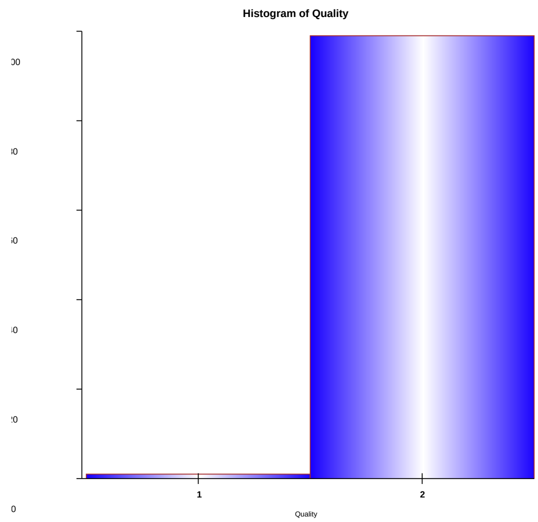
The size of a home is highly correlated with its selling price. The histogram of living area reveals the bulk of the properties are below 5,000 square feet. The table below the chart indicates the middle 50% of the properties range in size from 1.464 to 2,735. Clearly the two properties above 10,000 square feet are unusual. Those two sales were not used in the valuation model building process.



Variable	Median	25th Pctile	75th Pctile
Liv Area	1,950	1,464	2,735

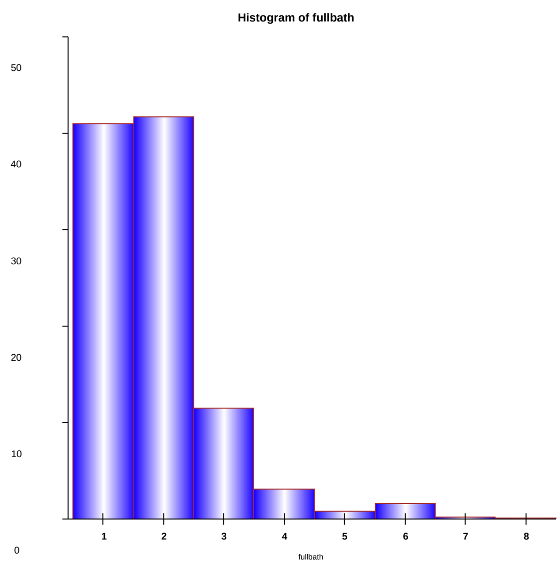
Construction. Quality

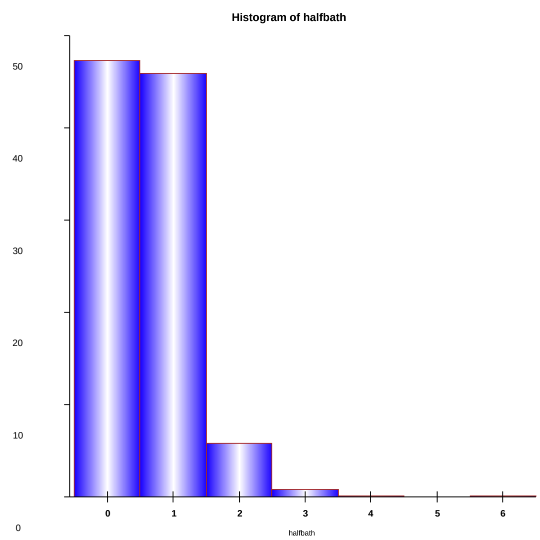
The quality of construction of a home is almost always an important factor in a model used to estimate value. The histogram of the quality variable reveals almost no variability. The table below the histogram shows the statistics for the three possible quality choices. This lack of variability means that construction quality is not a useful variable for the model building and valuation process.



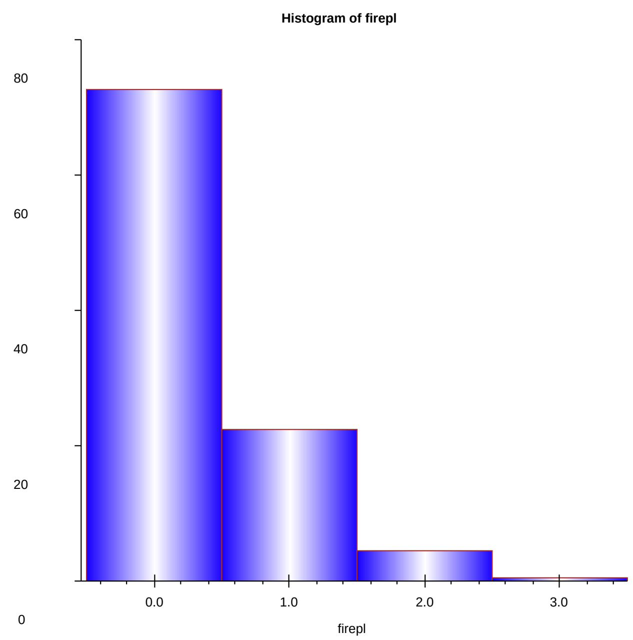
Living Area Stats			
Quality	**Count**	**Median**	**Percent**
Delux	10	3028	1.1%
Average	914	1942.5	98.9%
Poor	0	0	0.0%

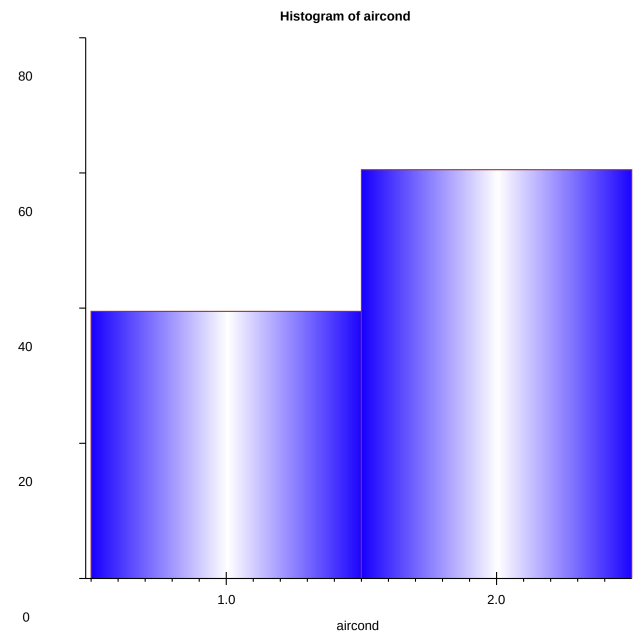
Baths
Most homes have either one or two full baths. The homes with one or more half baths are approximately equal to those with no half bath.





Fireplaces and Air Conditioning
The distribution of fireplace count and whether air conditioned or not look reasonable are likely candidate variables in the regression model.





Model Structure and Calibration

Structure

Two model structures were evaluated. They are referred to as "additive" and "multiplicative". The multiplicative model form is often referred to as a log-linear model. It turned out that for this dataset, the multiplicative form of the model had the superior performance.

Rather than using mathematical notation, an example of a portion of an additive and a multiplicative model are shown in the following figure.

Dependent	PRICE
Std Error for Estimate	60,658.0229
Constant:	61,682.3357
	Coeff
BSF	70.3645
LSF	21.6367
FULLBATH	15,214.2234
RQOS	
8	28,068.0924
9	39,113.1659
2	74,291.4178
3	43,555.8098
1	42,127.9684

Dependent	PRICE
Std Error for Estimate	0.1539
Constant:	954.9174
	Coeff
BSF	0.4956
LSF	0.2457
FULLBATH	0.0583
RQOS	
8	1.1280
9	1.1292
2	1.2355
3	1.1481
1	1.1540

The additive model is on the left. It says (as far as it is shown) that value is estimated as follows:

Add Est=$61,682+70.36*SFLA+21.64*LSF+15,2014*FULLBATH+28,068*RQOS8+ 39,113*RQOS9+…

The multiplicative Model is on the right. The interpretation is:

$$\text{Mult Est} = 954.92 * BSF^{0.496} * LSF^{0.246} * FULLBATH^{0.058} * 1.128^{RQOS8} * 1.128^{RQOS9} \ldots$$

Where:

SFLA is square foot of living area

LSF is lot size in square feet

FULLBATH is the number of full baths

RQOS8 is 1 when RQOS=8, 0 otherwise

RQOS9 is 1 when RQOS=9, 0 otherwise

The comparative statistics results for the baseline model (no outlier removal) is shown below. At this stage the multiplicative model shows more promise (better stats) than the additive.

Model	Count	Median	Mean	WgtMean	COD	PRD	PRB
Lin Add Base	926	0.999	1.067	1.000	20.078	1.067	-0.292
Mult Base	926	0.975	1.030	0.976	18.858	1.056	-0.205

The CLASS Variable

The CLASS variable combines several different aspects of a property into one of 13 categories, in this dataset. When allowed to enter the model, it is significant, but it causes other useful variables to be "masked" from consideration. When it is removed several additional variables enter the model which leads to improved performance of the model. The impact of the Class Variable is presented in Appendix F The Class Variable.

Model	Count	Median	Mean	WgtMean	COD	PRD	PRB
Mult Base	926	0.975	1.030	0.976	18.858	1.056	-0.205
Mult Base with Class	926	0.975	1.039	0.968	21.374	1.073	-0.292

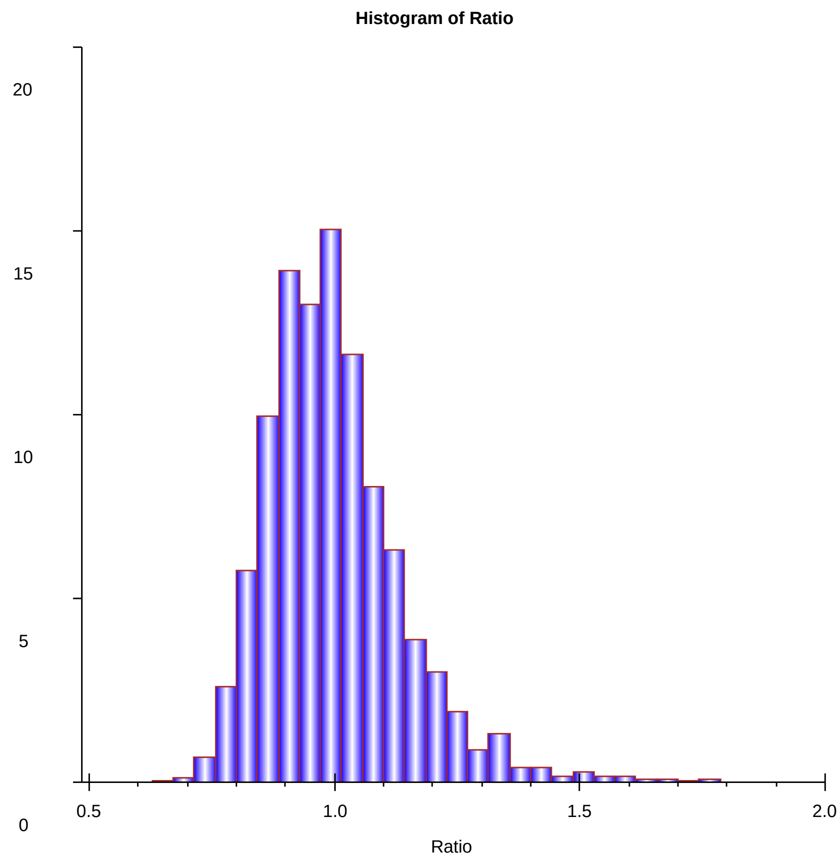
Outliers

When a model is first calibrated, it is often the case that some of the sales used in the modeling process are not representative of the group. Initially there are usually some extreme outliers. The traditional method for identifying outliers is to examine the ratio of estimated value to sale price for the sales in the sample. The method used is that described in the *IAAO Standard on Ratio Studies*. In brief, the process is:

1. Locate 25th percentile ratio
2. Locate the 75th percentile ratio
3. Compute Interquartile ratio or IQR (75th percentile-25th percentile)
4. Compute lower limit as 25th percentile - factor*IQR
5. Computer upper limit as 75th percentile + factor*IQR

The factor is typically chosen as 1.5 or 3.0 depending on whether the goal is to detect extreme outliers (3.0) factor or to take a deeper cut using a factor less than 3.0.

It is contended herein that the IAAO standard is faulty and needs to be modified to function as a reasonable tool in identifying outliers. First consider the distribution of ratios created by stochastic process used to simulate a sales sample along with the value estimates produced by a CAMA model. The figure below shows the histogram of the appraisal to sale ratios.



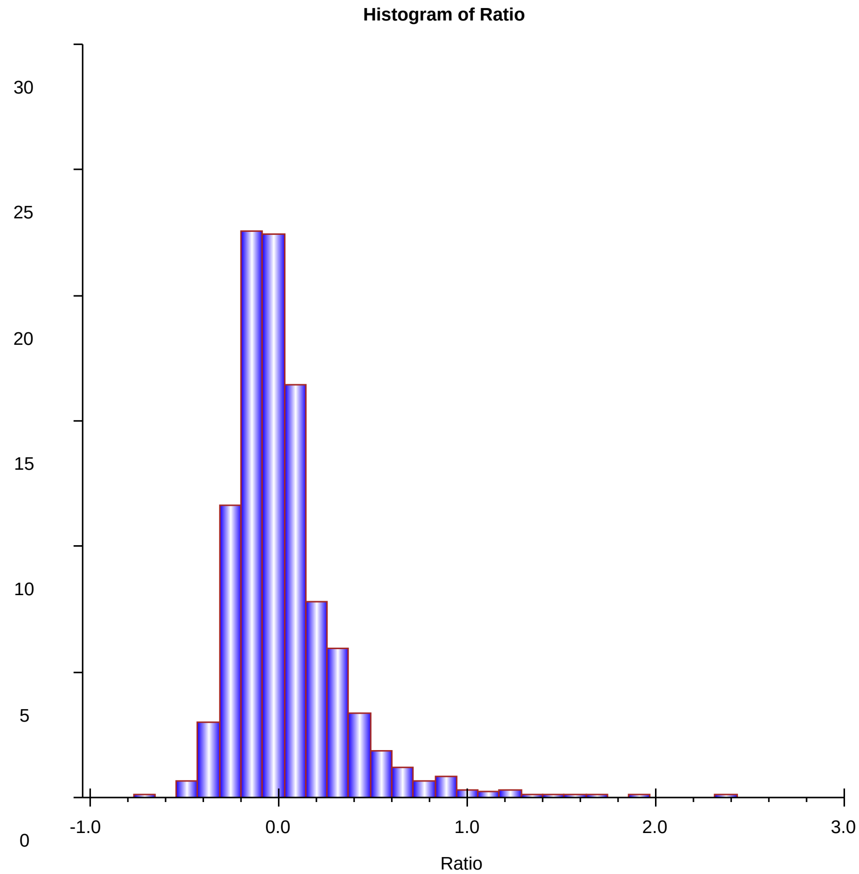
The sales ratio study for this distribution is as follows:

Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
2400	0.983	1.000	1.001	0.159	0.139	10.483	13.865	0.999	0.013

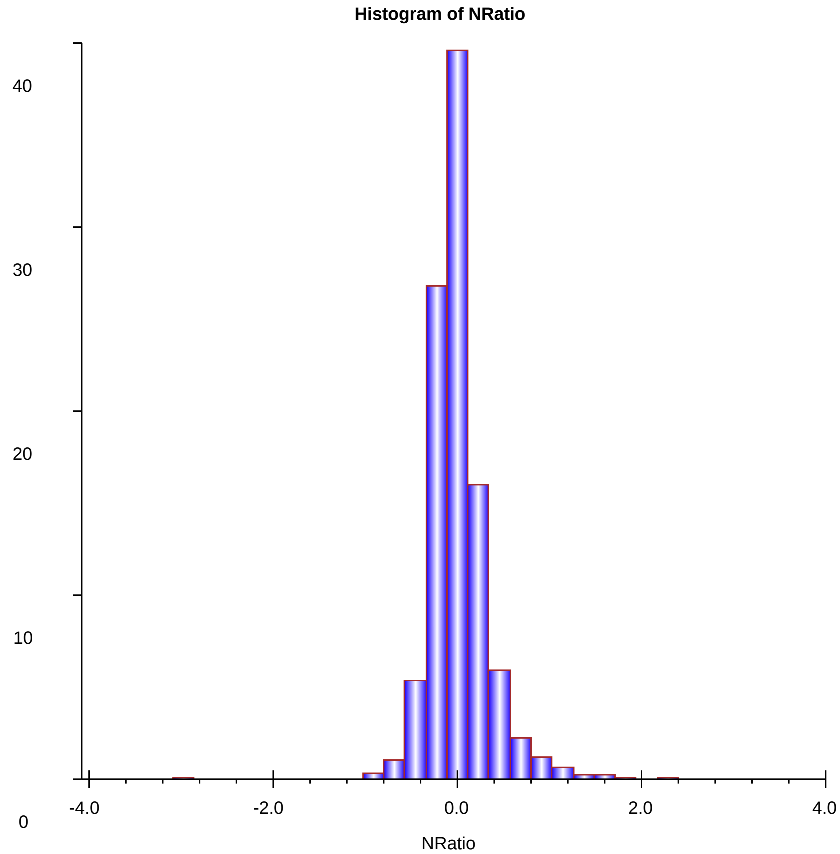
The corresponding Outlier detection parameters using various factors in the IQR detection process are shown below. The point being that for this simulation, an IQR factor of 0.75 produces 11.21% outliers while a factor of 1.0 produces 6.79% and so on down the table until a factor of 3.0 nets 14 outliers and 0.58% of the total sales.

IQR Factor	IQR	25th Pctile	75th Pctile	Low Lim	Upper Lim	Out Count	Out Pcnt
0.75	0.159	0.906	1.066	0.787	1.185	269	11.21%
1.00	0.159	0.906	1.066	0.747	1.225	163	6.79%
2.00	0.159	0.906	1.066	0.588	1.385	41	1.71%
3.00	0.159	0.906	1.066	0.428	1.544	14	0.58%

In the realistic case of Rogers Park, the histogram of ratio (centered on 0 and expressed as a decimal fraction) produced by the first model with no outliers removed is shown in the image below. It is evident that the histogram is not symmetric. The major reason for this is that although ratios above 0.0 are unbounded, ratios below 0.0 are bounded by a lower limit of -1.0. Another way of saying it is that the range of ratios where the estimate is below the price is compressed compared to those where the estimate exceeds the price.



A transformation on the ratios below 100% yields the far more symmetrical histogram below. The definition of the ratios below 100% is 1-price/estimate. Now, it is easily seen there is one extreme outlier at about -3.0. The same sale does not look so much an outlier in the original histogram.



The IQR calculations are revealing as well. The comparisons include using an IQR factor of 3.0 and one of 0.75. The outlier counts for the standard ratio (Ratio) and the normalized ratio NRatio both centered on 0 and expressed as a decimal fraction instead of a percent. What is telling is a comparison of the outliers removed from the low and high sides of the distribution. Using the standard ratio, the Low to High outlier ratio is much lower than that for the NRatio. In other words, the standard method is missing out on the outliers when the estimate is lower than the price.

IQR Factor	RatioType	IQR	25th Pctile	75th Pctile	Low Lim	High Lim	Out	Low	High	Pcnt Out	L/H%
3	Ratio	0.2775	-0.1493	0.1282	-0.9819	0.9608	14	0	14	1.51%	0.00%
3	NRatio	0.3037	-0.1755	0.1282	-1.0866	1.0394	12	1	11	1.30%	9.09%
0.75	Ratio	0.2775	-0.1493	0.1282	-0.3574	0.3364	118	15	103	12.74%	14.56%
0.75	NRatio	0.3037	-0.1755	0.1282	-0.4033	0.3560	139	45	94	15.01%	47.87%

It is the NRatio method of outlier detection that is used exclusively in the Chicago Triad.

The Multiple Regression Model – Holdout Sample

Statistical comparisons are presented for the major models evaluated in performing the valuation analysis in Rogers Park. Initial results were examined by use of a 20% holdout sample to validate the modeling process. Both the linear additive and multiplicative models were considered at this stage. Interestingly the holdout sample performed slightly better than the in-sample group. Also, the multiplicative model holds a slight advantage in these results. The point of the exercise is that there is no issue of “overfitting” the data to obtain favorable model performance statistics.

Model	HOLDOUTGROUP	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
Lin Add	HOut	182	0.990	1.027	0.983	0.274	0.209	16.186	20.311	1.044	-0.220
Lin Add	IN	632	1.010	1.033	0.985	0.259	0.221	16.622	21.384	1.048	-0.240
Lin Add	Combined	814	1.006	1.031	0.985	0.260	0.218	16.531	21.139	1.047	-0.236
Mult	HOut	182	0.969	0.986	0.963	0.207	0.156	12.790	15.832	1.024	-0.087
Mult	IN	632	0.970	0.989	0.964	0.235	0.164	13.787	16.597	1.026	-0.101
Mult	Combined	814	0.970	0.988	0.964	0.224	0.162	13.559	16.420	1.025	-0.098

The Actual Model

The image that follows is the actual MRA model after outliers had been removed. The number of outliers identified was 112 or 12.1%.

Dependent	PRICE
Std Error for Estimate	0.1539
Constant:	954.9174

Attribute		Coeff	Std. Error	bWeight	t value
BSF		0.4956	0.0352	0.6742	14.0762
LSF		0.2457	0.0152	0.3088	16.1554
FULLBATH		0.0583	0.0207	0.0816	2.8173
RQOS					
	8	1.128		0.074	
	9	1.1292		0.0825	
	2	1.2355		0.1506	
	3	1.1481		0.1026	
	1	1.154		0.0661	
	6	1.199		0.1269	
	7	1.2084		0.1371	
	4	1.2117		0.1001	
	5	1.1486		0.0872	
	19	0.9279		-0.043	
	13	1.05		0.025	
	12	1.0349		0.0208	
	11	1.1298		0.0892	
	10	1.1274		0.0877	
	17	1.0091		0.0062	
	16	0.9815		-0.0099	
	15	1.1242		0.0394	
	14	1.1257		0.0796	
	20	0.9265		-0.0352	
	18	1.0000		0.0000	
ROOMS					
	8	1.0157		0.0157	
	9	1.0501		0.0255	
	44	1.2429		0.0222	
	27	1.398		0.0342	
	14	1.0222		0.0067	
	7	1.0333		0.0377	
	30	1.2959		0.0528	
	5	0.9497		-0.0399	
	4	0.8641		-0.0332	
	36	1.2071		0.0332	
	24	1.2745		0.0494	
	25	1.206		0.0191	
	19	1.0129		0.0023	
	18	1.1806		0.0925	
	13	1.1931		0.0311	
	12	1.0456		0.0382	
	11	0.9391		-0.0262	
	10	1.0103		0.0065	
	16	1.0433		0.0129	
	15	1.0705		0.0218	
	28	1.378		0.0566	
	20	1.2019		0.056	
	21	1.0089		0.0027	
	22	1.4091		0.0922	
	23	0.8713		-0.014	
	6	1.0000		0.0000	

Attribute		Coeff	Std. Error	bWeight	t value
LOCF		0.7923	0.0565	0.246	14.0305
NUM					
	2	0.7377		-0.2499	
	3	0.6902		-0.1789	
	1	0.8461		-0.1682	
	4	0.6669		-0.0922	
	5	0.6233		-0.1659	
	6	1.0000		0.0000	
EXTCON					
	3	1.0186		0.0157	
	1	1.0803		0.0702	
	4	1.048		0.0321	
	2	1.0000		0.0000	
FIREPL					
	2	1.0194		0.0112	
	3	1.0778		0.0152	
	1	1.0892		0.1035	
	0	1.0000		0.0000	

Model Statistics	
Total Valued	814
R squared	0.8141
Adjusted R squared	0.7995
COD	12.1003
COV Median	15.1919
COV Mean	14.9265
Median	0.9971
Mean	1.0111
Weighted Mean Ratio	0.9899

Performance Statistics of MRA Model

Performance statistics by neighborhood:

NBHDcode	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
10	163	1.014	1.027	1.002	0.225	0.162	12.690	15.784	1.025	-0.145
21	127	1.008	1.015	0.999	0.180	0.142	11.258	14.019	1.016	-0.081
22	89	0.970	0.998	0.982	0.187	0.152	12.069	15.179	1.017	-0.043
23	39	0.991	1.008	0.987	0.248	0.174	13.990	17.291	1.021	-0.104
31	189	0.993	1.004	0.982	0.213	0.148	12.097	14.694	1.023	-0.082
32	34	1.019	1.056	1.038	0.199	0.138	11.010	13.031	1.017	-0.059
33	15	0.922	0.952	0.962	0.269	0.140	12.851	14.681	0.990	0.074
40	120	1.006	1.013	0.993	0.196	0.143	11.072	14.146	1.020	-0.094
60	38	0.939	0.970	0.950	0.184	0.151	12.003	15.539	1.021	-0.181
Total	814	0.997	1.011	0.990	0.208	0.150	12.100	14.824	1.021	-0.080

Comparable Sales Valuation

The focus to this point has been on developing a rational multiple regression analysis (MRA) model. However, MRA is not the valuation method employed in valuing Rogers Park. It is an important step in the valuation by comparable sales analysis.

Why, comparable sales analysis? There are two very important reasons for using comparable sales valuation. First, it is more transparent and defensible to the taxpayer than an MRA model where the focus is on structure, coefficient, multicollinearity and other techno-statistical terms. The second is that it is usually more accurate than MRA.

The relationship between MRA and Comp sales

The basic process is as follows:

Find the sales properties which are most comparable to the subject property to be valued
Adjust the sale price for each comparable to account for differences between it and the subject's characteristics and for the date of sale
Weight these adjusted comparable sales estimates according to their similarity to the subject
Sum the weighted comparable sales estimates to get the final estimate

The connection to MRA is explained by the following:

Comp Estimate (Subject) = Comp Sale Price + an adjustment for differences in property characteristics and date of sale

Which can be shown to be

Estimate= Comp Price+[MRA(Subject)-MRA(Comp)]

Rearranging

Estimate = MRA(Subject) + [Comp Price-MRA(Comp)]

Which is

Estimate- MRA(Subject)+Residual error of Comp MRA Estimate

Another way of putting it, a Comp Sale estimate of value is the MRA estimate corrected by the residual error of the MRA estimate of the comp.

Comps Sale Illustrative Computation

The process of computing a comp sales estimate resulted in this formula using a bit more of a mathematical form:

$$Est(i) = MRA(Subj) + \left(CompPrice(i) - CompMRA(i)\right)$$

or

$$Est(i) = MRA(Subj) + CompResidErr(i)$$

The table below shows the computation for the case of five comparable sales.

	1	2	3	4	5
i	MRASubj	CompPrice(i)	CompMRA(i)	CompResidError(i)	Est(i)
1	$269,881	$220,000	$225,106	-$5,106	$264,774
2	$269,881	$200,500	$239,921	-$39,421	$230,460
3	$269,881	$260,000	$239,586	$20,414	$290,295
4	$269,881	$290,000	$245,229	$44,771	$314,652
5	$269,881	$229,900	$223,365	$6,535	$276,416
				Subj Est	$275,320

Using the numbered columns from the table above:

$$5(i) = 1(i) + 4(i)$$

And

$$SubjEst = average\ of\ 5(i)$$

Comparable Sales Selection

The “find comps” portion of the comp sales process involves setting comp sales selection parameters by “iteration”. The first iteration has the tightest specification on the sales that will be consider for analysis. The second iteration loosens up a bit on which sales will be considered and so on until enough iterations have been defined to value all properties. The iteration rules used in Rogers Parks are shown below with explanation of each iteration’s restriction on comp sales.

Rules	Iteration 1:	Iteration 2	Iteration 3:	Note
Buffer Size	1000	3000	5000	Max Distance in feet used for the search
Comparables	5	5	3	Required Number of Comparabl Sales
Weighting	0.5	0.5	0.5	Distance/Similarity weighting
BSF	+/- 20%	+/- 40%	+/- 100%	Max % Diff in Sq Ft Liv Area
LANDVAL	+/- 20%	+/- 40%		Max % Diff in Lot Value
AGE	+/- 20	+/- 40		Max Diff in Age
BEDS	+/- 1			Max Diff in Bedrooms

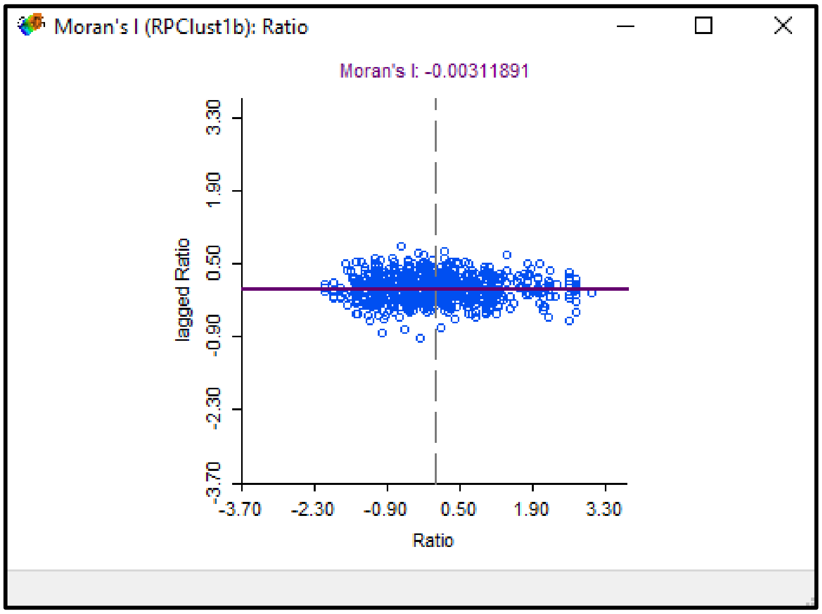
Comp Sale Results on the Sales Sample

As can be seen in the table below, this variant of the comp sale model has approximately the same performance statistics when considering overall accuracy (COD) and vertical equity as the corresponding MRA model use to adjust the sales. In most cases comp sales outperforms MRA. When it does not, it means the MRA model has accounted for location quite well.

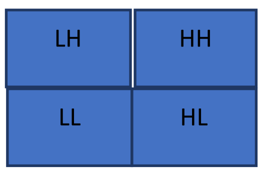
Iteration	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
1	678	0.999	1.019	1.009	0.179	0.148	11.451	14.550	1.010	0.005
2	133	1.001	1.021	0.988	0.327	0.183	15.443	17.890	1.034	-0.115
3	3	1.097	1.108	1.051	0.457	0.229	13.890	20.637	1.055	-0.216
Total	814	0.999	1.020	1.005	0.205	0.155	12.128	15.146	1.015	-0.022

Spatial Stability of the Value Estimates

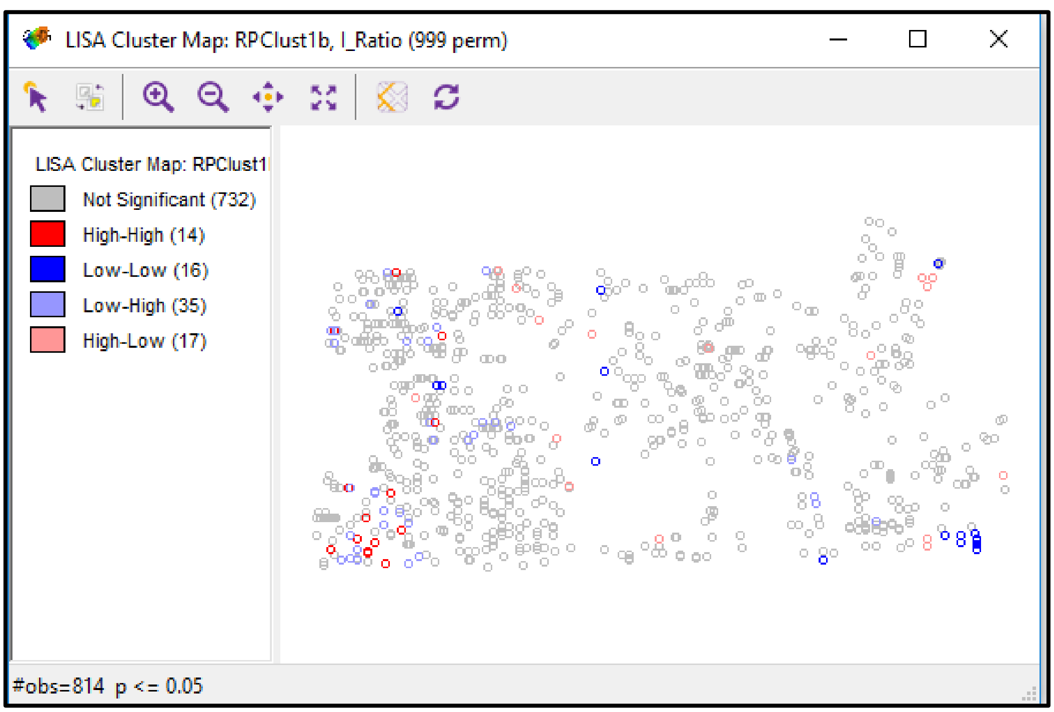
A means to verify the locational stability of the comparable sales estimates is provided computing what is termed “Local Indicators of Spatial Association” often referred to as LISA. Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. In mass appraisal it is customary to look for spatial clusters in the ratio of appraised value to sale price. The plot below has as its X Axis the z-transform of Ratio of the estimate to the Sales Price, defined as $z = (x - \mu)/\sigma$ where x is the individual ratio, μ is the mean ratio and σ is the standard deviation of the ratios in the sample. The Y axis is the average of the five nearest neighbors z scores not including the sale property of the X Axis. The fact that there is virtually no slope to the plot is a good indication that there are no spatial clusters of high or low ratios.



It is useful to be able to visualize the location of the significant clusters. The diagram below divides the Moran Scatterplot into four quadrants. Each quadrant is labelled to represent the type of association. Thus, **HH** represents high ratios near high ratios, **LL** for Low near Low**,** **LH** for Low near High and **HL** for Low near High.

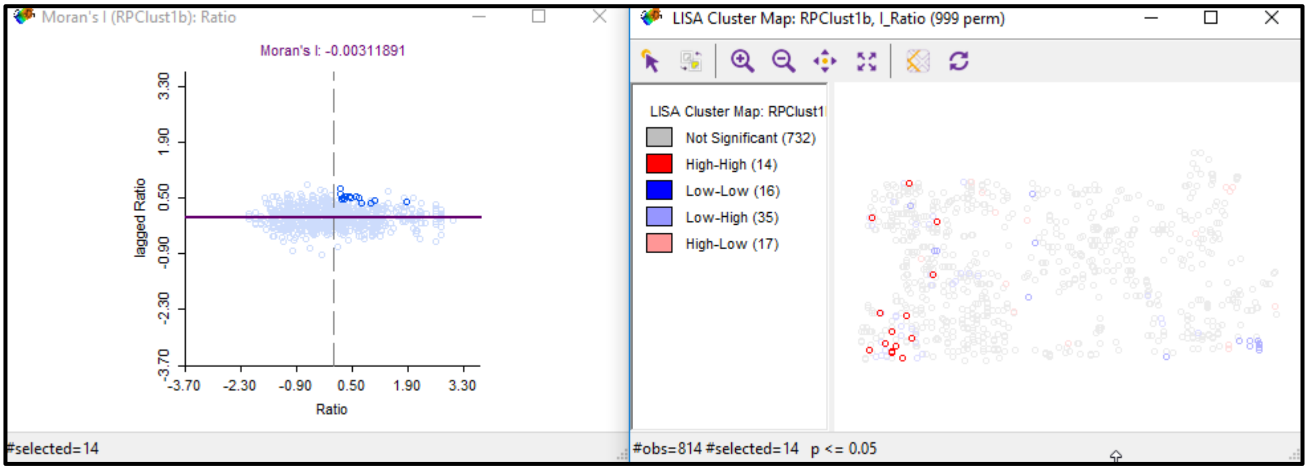


In addition to the scatterplot, consider the map below intended to highlight high and low ratio clusters. The map legend is particularly informative. There are only 14 statistically significant high ratios near high ratios and 16 low ratios near low ratios.

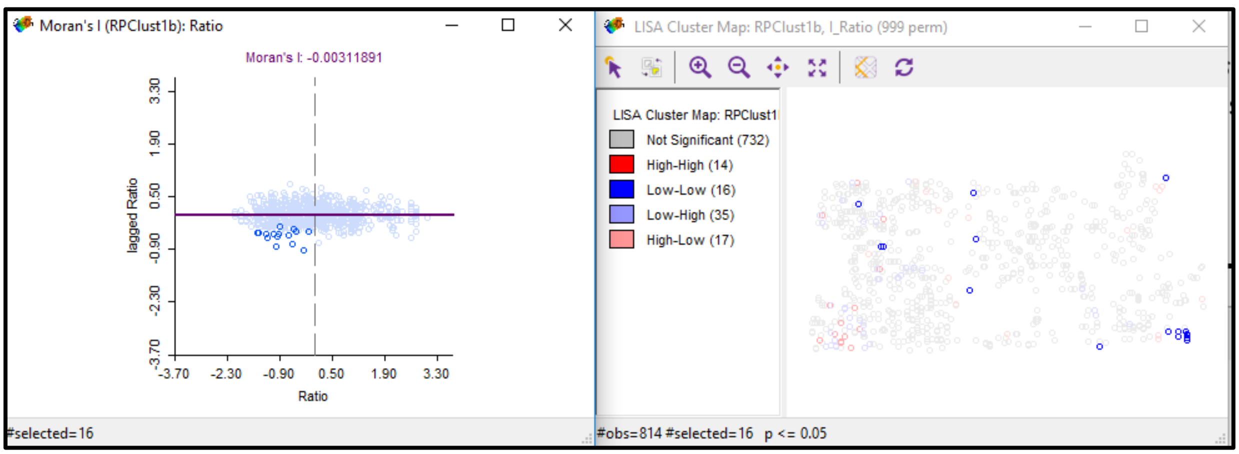


This is a particularly important finding. ***Namely, there is little indication of a pattern of spatial bias in the valuations.*** To further highlight the distribution of the four categories of statically significant associations, they are taken one at a time.

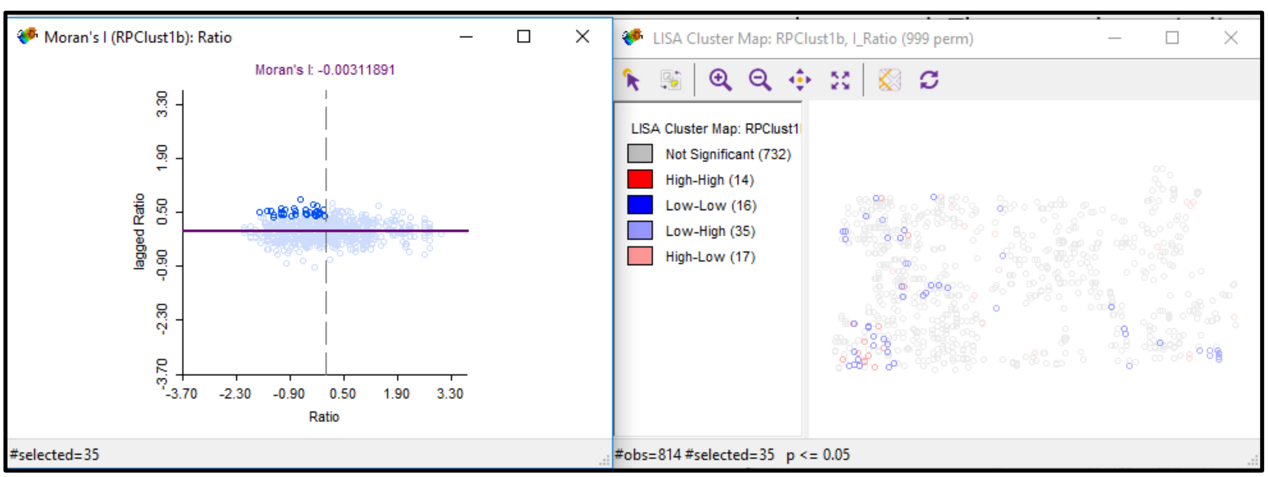
The first in the series is the High near High ratios. They are scattered and sparse.



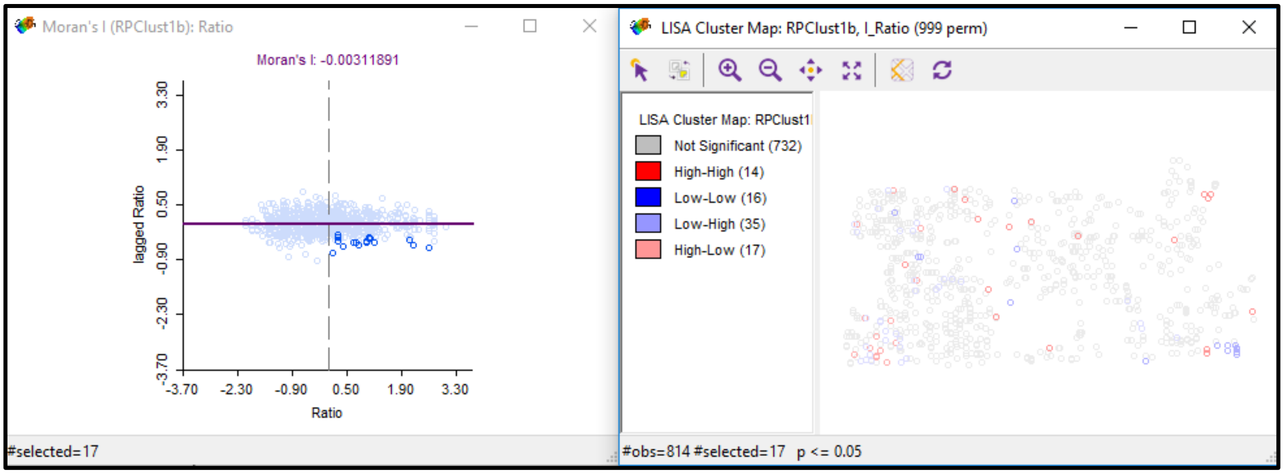
Next is the Low ratios near Low ratios. Again, the are sparse and scattered. There may be an indication of a small cluster in the southeast of the map.



Low near High ratios are shown next. Again, it is a small number compared to the total (about 4.3%) and they are widely scattered.



Finally, the High near Low ratios - again, few and scattered.



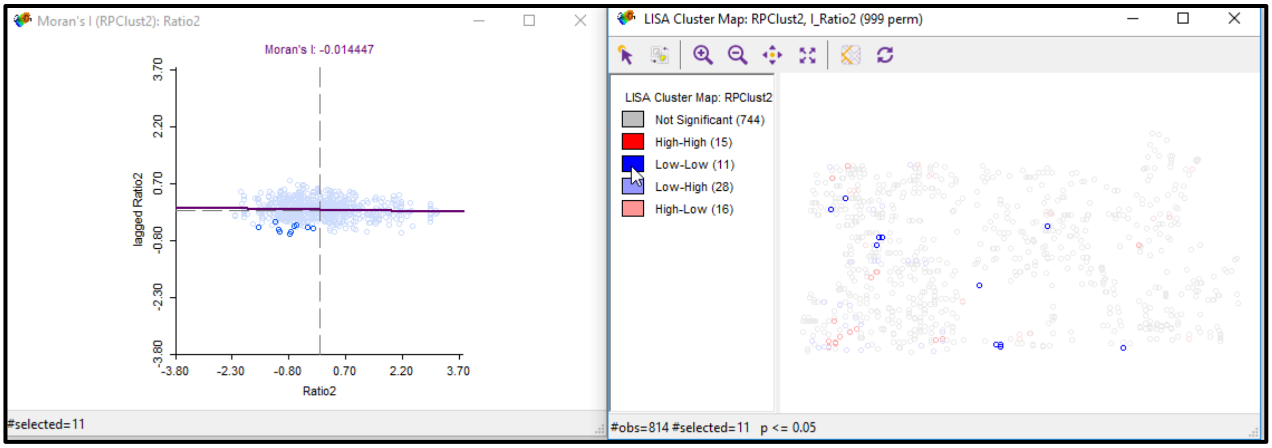
Further examination of Comp Sales by NBHD in the first table below indicates a level of value issue with NBHDs 33 and 60 in which the median ratios of 0.886 and 0.968 is noticeably different from the target ratio of 1.00. The comparable sale algorithm used in producing these value estimates has options that allow for different weight of the comp sales. At a slight loss in overall accuracy, the variability among NBHDs is reduced as shown in the second table below. The measurement of variability is the standard deviation of the median estimate by NBHD. This figure is shown at the far right of each portion of the tables of results. The second method reduces the variability by 43%.

NBHDcode	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB	STD DEV
10	163	1.009	1.029	1.017	0.245	0.174	13.481	16.901	1.012	0.006	**0.044**
21	127	1.014	1.036	1.025	0.202	0.147	11.448	14.164	1.011	-0.031	
22	89	0.983	1.009	0.997	0.211	0.148	11.894	14.663	1.012	-0.013	
23	39	1.039	1.048	1.032	0.235	0.197	14.867	18.744	1.016	0.013	
31	189	0.997	1.008	0.993	0.190	0.141	11.054	13.955	1.015	-0.027	
32	34	0.990	1.012	0.998	0.299	0.170	13.420	16.759	1.015	-0.025	
33	15	0.886	0.946	0.962	0.300	0.157	15.284	16.587	0.983	0.111	
40	120	1.015	1.028	1.014	0.176	0.138	10.447	13.380	1.014	-0.046	
60	38	0.968	0.995	0.973	0.193	0.158	12.080	15.859	1.023	-0.211	
Total	814	0.999	1.020	1.005	0.205	0.154	12.128	15.064	1.015	-0.022	

NBHDcode	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB	STD DEV
10	163	1.025	1.033	1.022	0.237	0.189	14.435	18.276	1.011	0.038	**0.025**
21	127	1.040	1.052	1.040	0.227	0.163	12.553	15.487	1.011	-0.016	
22	89	0.999	1.016	1.003	0.204	0.162	12.683	15.889	1.013	-0.001	
23	39	0.993	1.059	1.041	0.225	0.210	16.250	19.824	1.017	0.019	
31	189	0.997	1.011	0.999	0.188	0.157	12.159	15.510	1.012	0.009	
32	34	0.994	1.031	1.015	0.239	0.196	14.742	18.999	1.016	-0.016	
33	15	0.961	0.958	0.976	0.248	0.157	13.588	16.356	0.982	0.100	
40	120	1.036	1.038	1.022	0.169	0.146	10.879	14.087	1.015	-0.040	
60	38	0.994	1.028	1.006	0.180	0.155	11.719	15.091	1.022	-0.183	
Total	814	1.014	1.029	1.016	0.198	0.167	12.950	16.254	1.013	0.000	

One additional note about the performance of the second configuration of the comp sales method is the remarkable outcome for PRD and PRB. An overall PRB of 0.000 is a rare event.

Revisiting the LISA measure using the new comp sales values in the next image, the number of low ratios near low ratios dropped from 16 to 11. ***Any evidence of Low near Low clustering is gone!***


Moran's I (RPClust2): Ratio2
Moran's I: -0.014447
lagged Ratio2
Ratio2
#selected=11
LISA Cluster Map: RPClust2, I_Ratio2 (999 perm)
LISA Cluster Map: RPClust2
Not Significant (744)
High-High (15)
Low-Low (11)
Low-High (28)
High-Low (16)
#obs=814 #selected=11 p <= 0.05

Lake View

Summary

Key points about Lake View Modeling and Valuation

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- Owner Occupancy data was considered, but did not prove to be statistically significant
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Lake View was valued using the Comparable Sales direct market comparison method of valuation
- Performance statistics were well within IAAO standards

The Data

Sales Counts

Starting with the Lake View combined sales and subjects file of 23,031 records where Amount1 is the sales price

Filter		Count
Amount1		
	>0	16,474
	>250,000	12,725
	<5,000,000	12,710
multi<1		12,318
sqftb<9000.		12,244
Sale Year>2012		4,962
puremarket=1		4,079
Starting Count		4,079

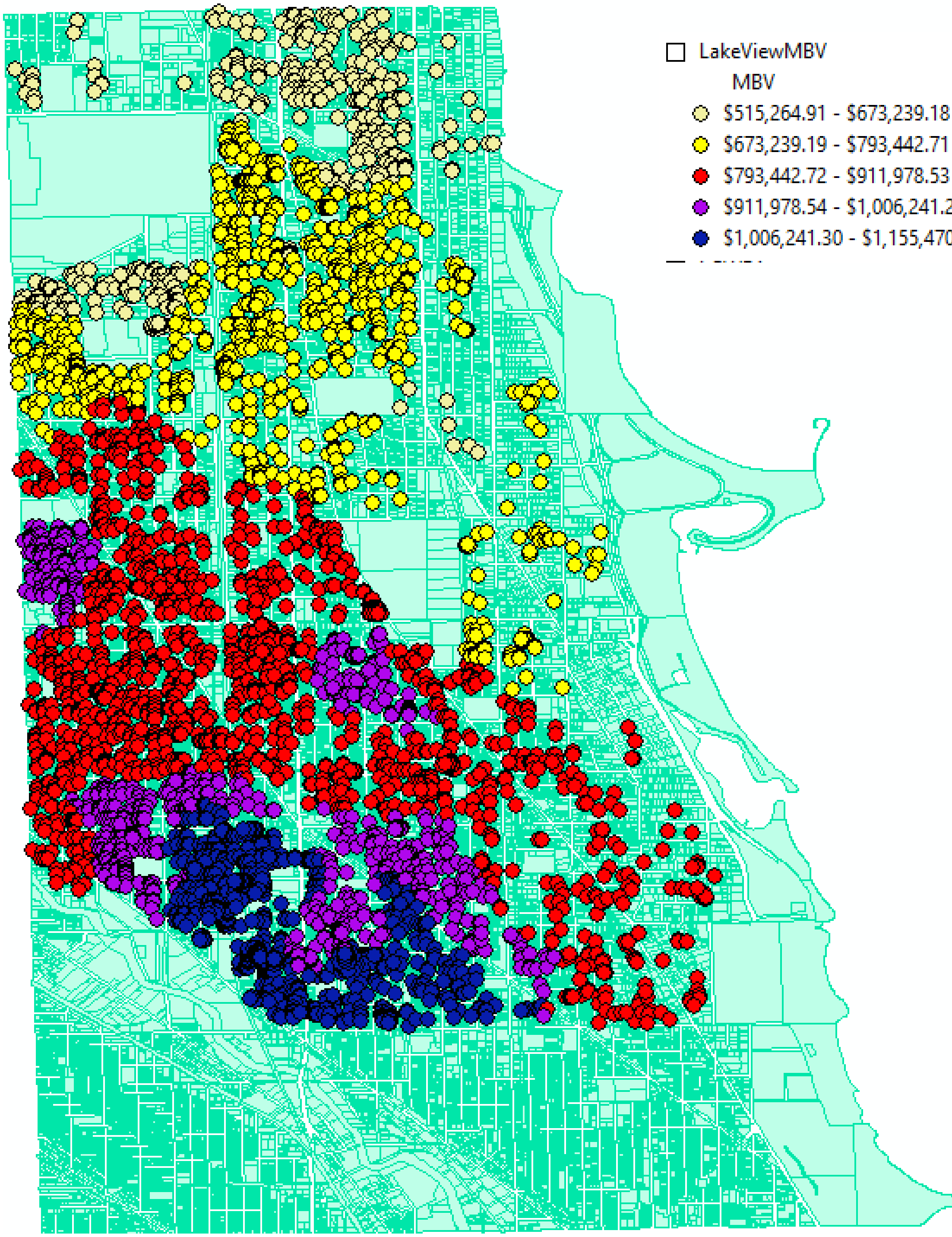
Data Fields

The initial list of data items available for analysis is provided in Appendix A Variable Definitions. Certain additional data fields were created. Those that were relevant to Lake View are:

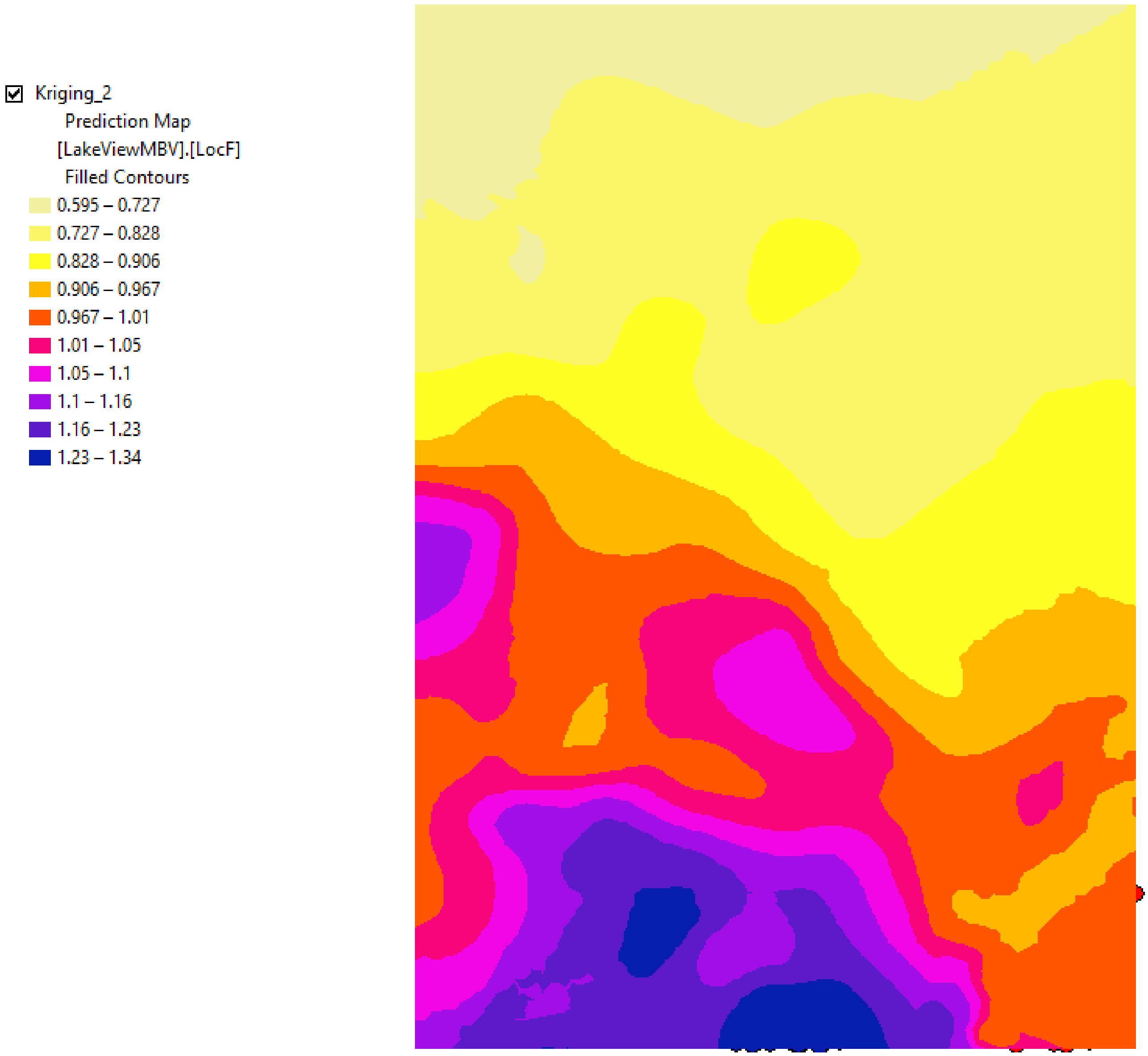
Location Factor

A location factor was derived by use of Geographically Weighted Regression (GWR). The process is described in Appendix B Location Factor.

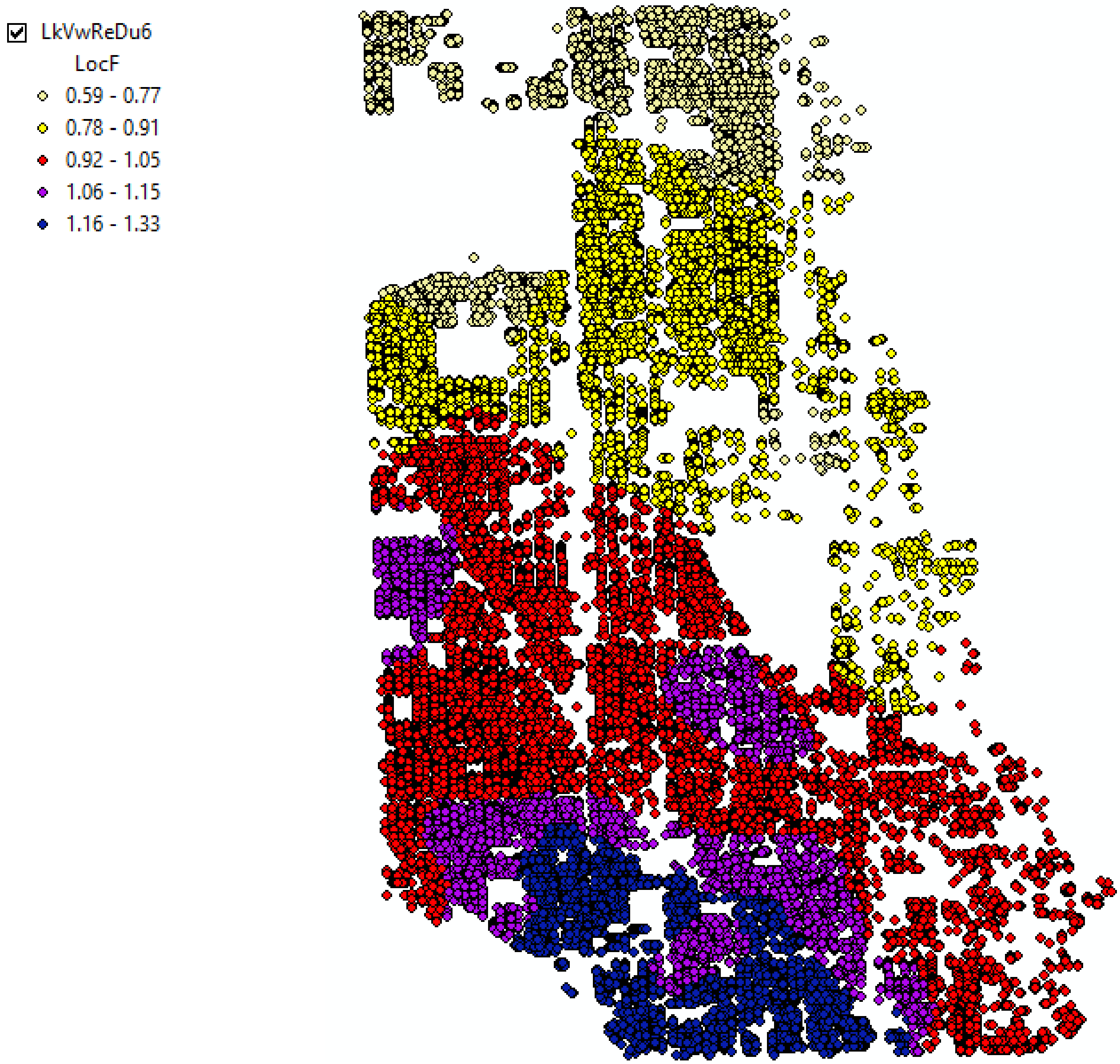
The first image below is that of the Market Basket Value or MBV.



The resultant continuous surface and thematic legend obtained from Kriging is shown in the image below.

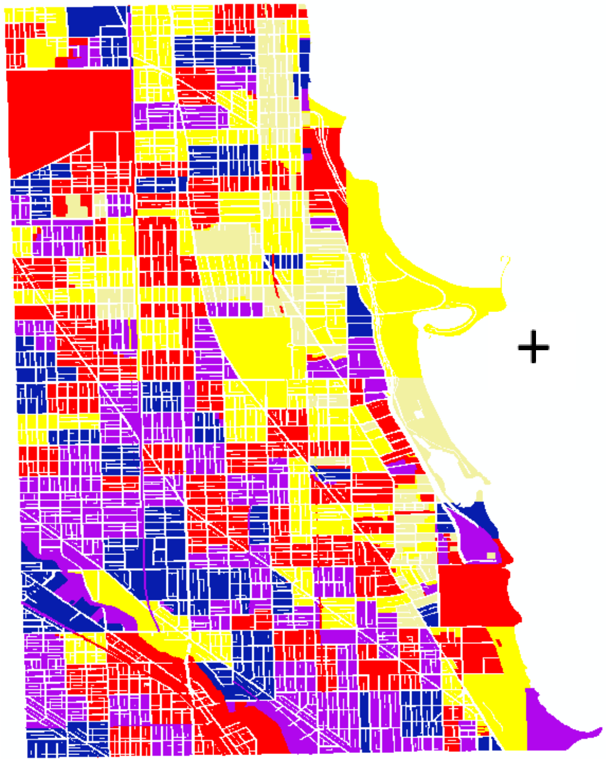


When applied to all properties the thematic map of Location Factor is given in the next image.

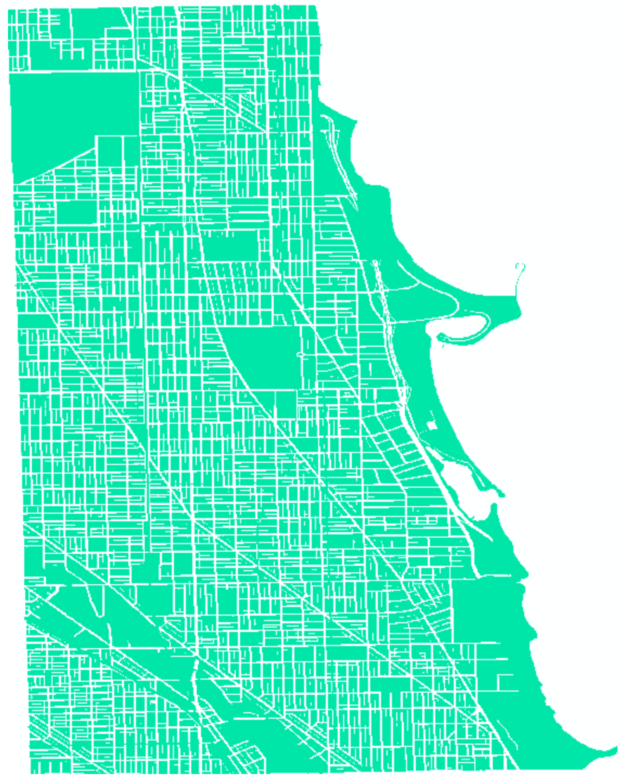


Owner Occupancy

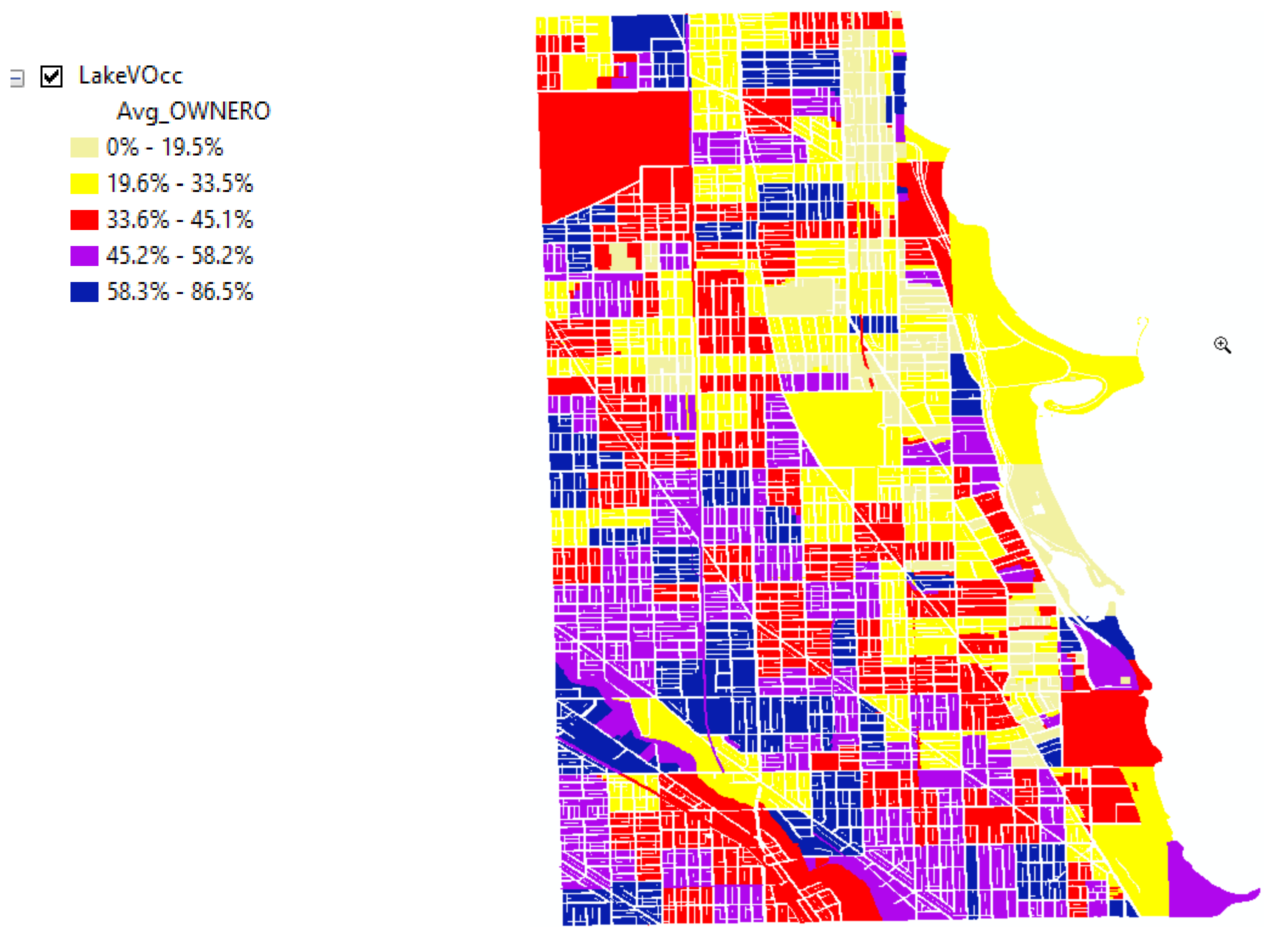
Owner Occupancy data is available at the Census Block Group level. County data is organized at several levels including parcel, block and neighborhood. Since the two geographies are organized differently, they were joined using what is called a “spatial join”. The image on the left below is of the owner occupancy level. The image on the right represents the parcel fabric.



+

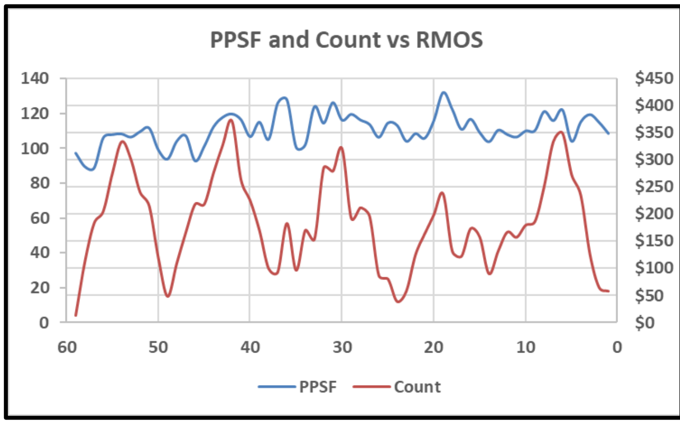


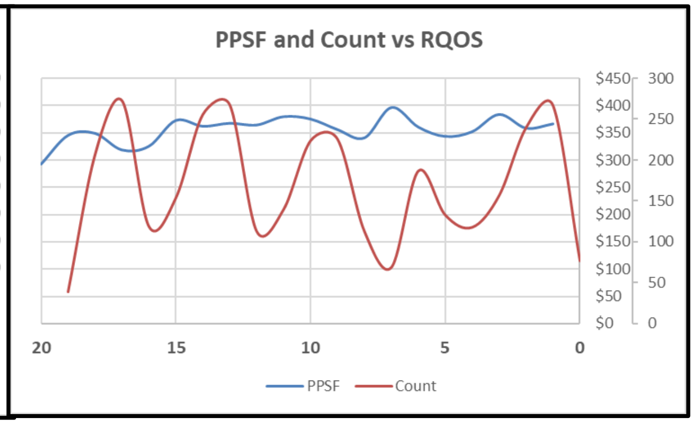
When joined the result becomes a parcel fabric with spatially interpolated owner occupancy data. The owner occupancy data is thus made available at the individual parcel level and becomes a candidate variable in an MRA Model.



Reverse Quarter of Sale

The sales used in the analysis span a period of five years. To allow for time trending the sales to the valuation date, first, a reverse month of sale is computed. If the sale took place in December of 2017, the reverse month of sale (RMOS) is 1, November 2017, RMOS is 2, all the w2ay back to January of 2013, RMOS is 60. In terms of using this variable directly in the model to be discussed, it is converted into a Reverse Quarter of Sale (RQOS). The rationale for this approach can be seen in the following two charts. The first shows Price per Square Foot (PPSF) and Count vs. RMOS. The second chart shows the same two variables vs RQOS. The RMOS variable is too granular and "noisy" as compared to the RQOS variable.



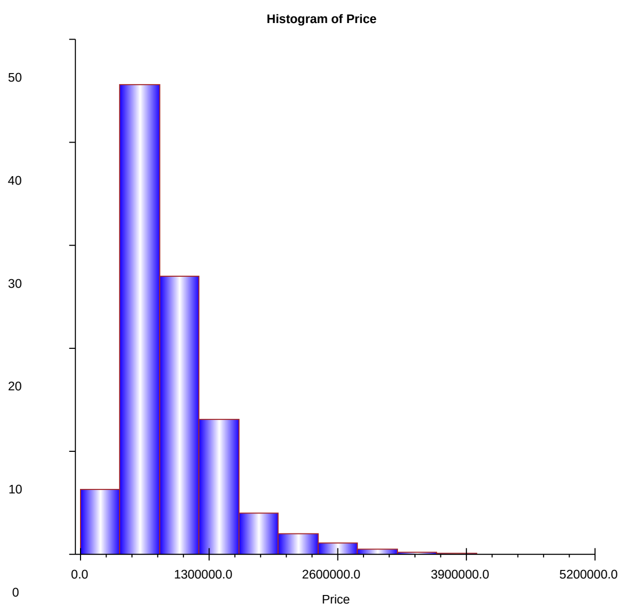


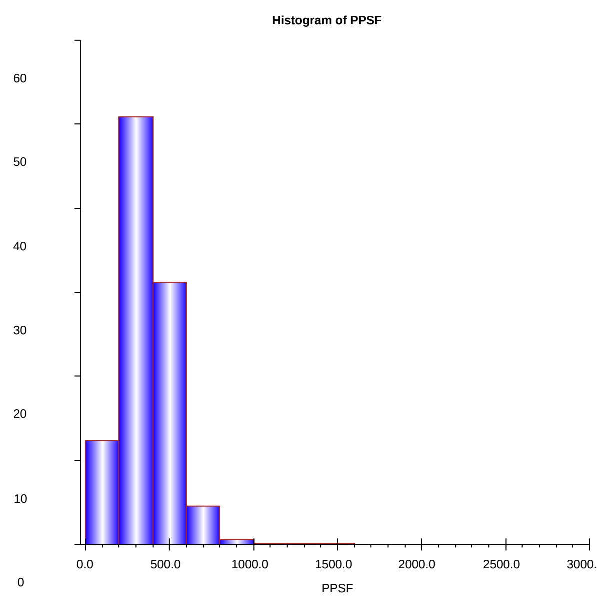
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.

Price and Price per Square foot

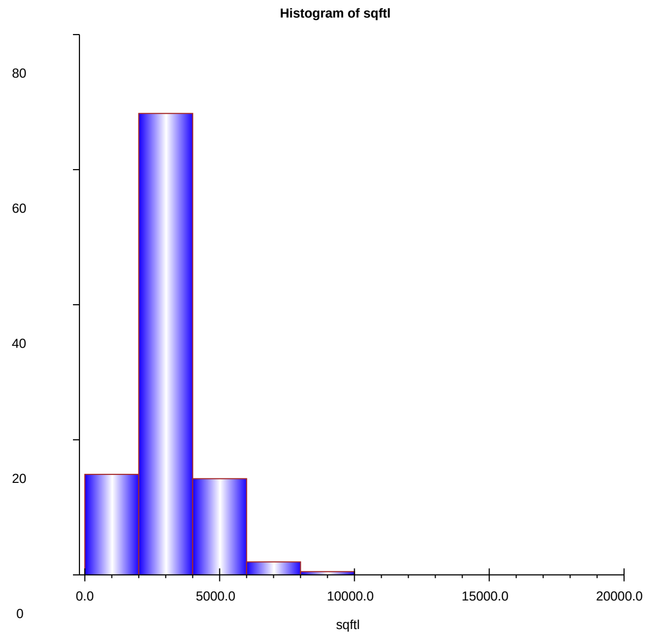
The price histogram fits the range of prices specified at the outset of modeling, namely a range of $250K-$5,000K. The price per square foot range as what are likely to be outlier situations. In other words, $2,000 per square foot is not likely to represent a true open market situation,

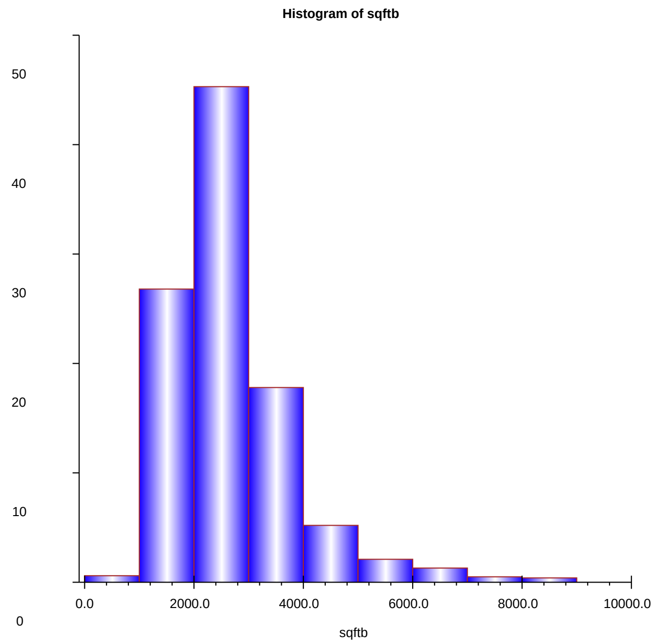




Square foot Land and Building

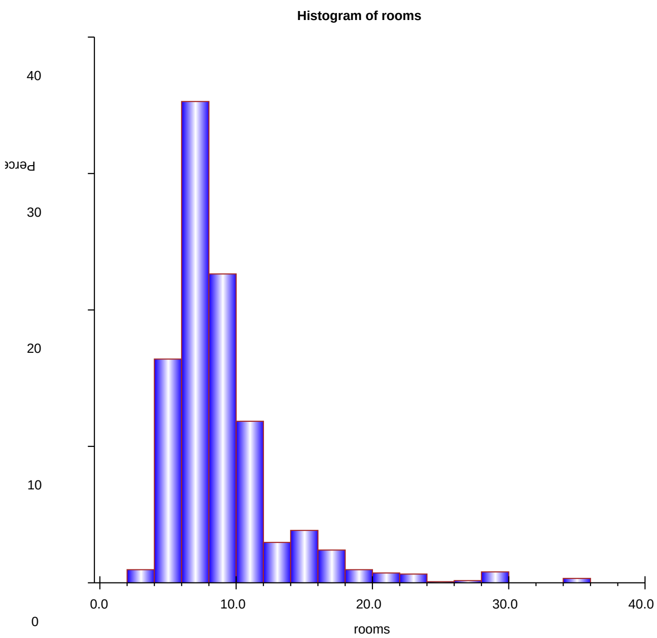
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not.

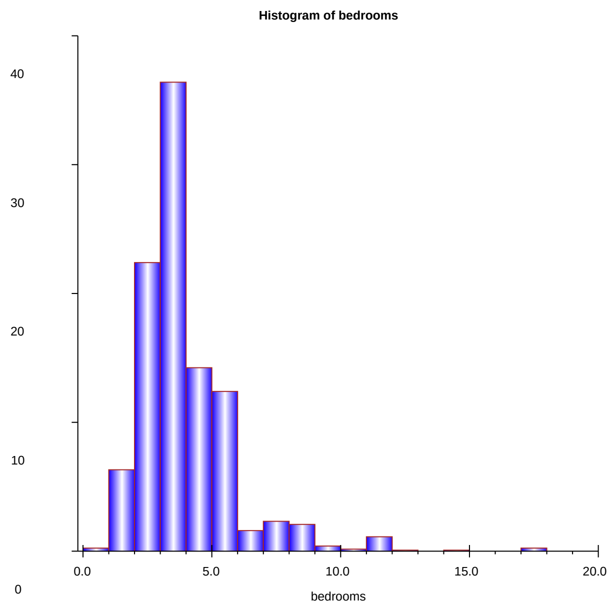




Rooms and Bedrooms

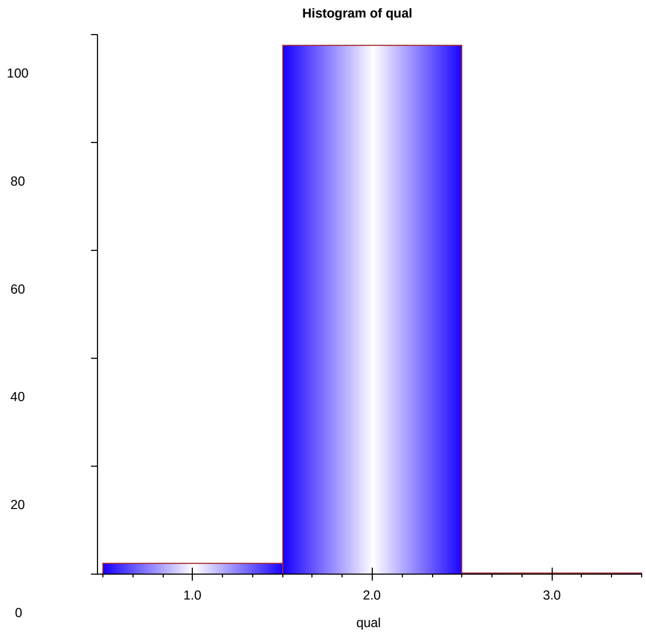
Looks like there are homes with 36 rooms and some with 18 bedrooms. For the specific case of 36/18, it looks as if they are six-unit apartment buildings with each unit having 6 rooms and 3 bedrooms.

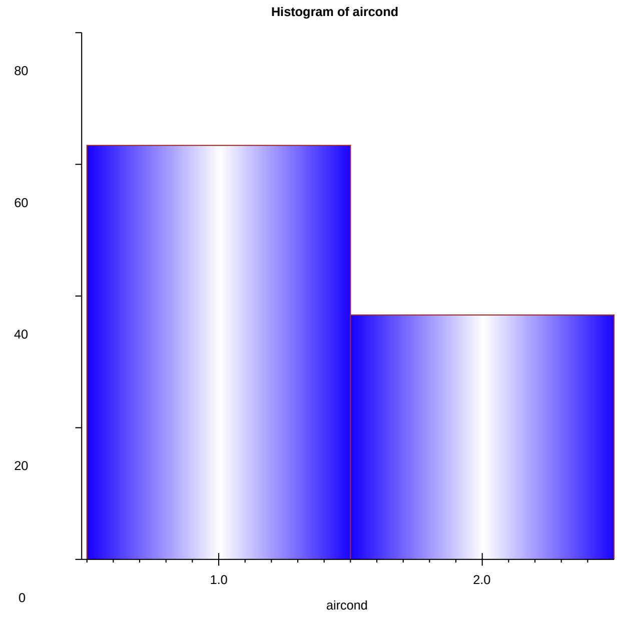




Quality and Air Conditioning

The quality of construction variable has little variability and almost certainly will not be a useful variable. On the other hand, air conditioning may well be useful in a model.





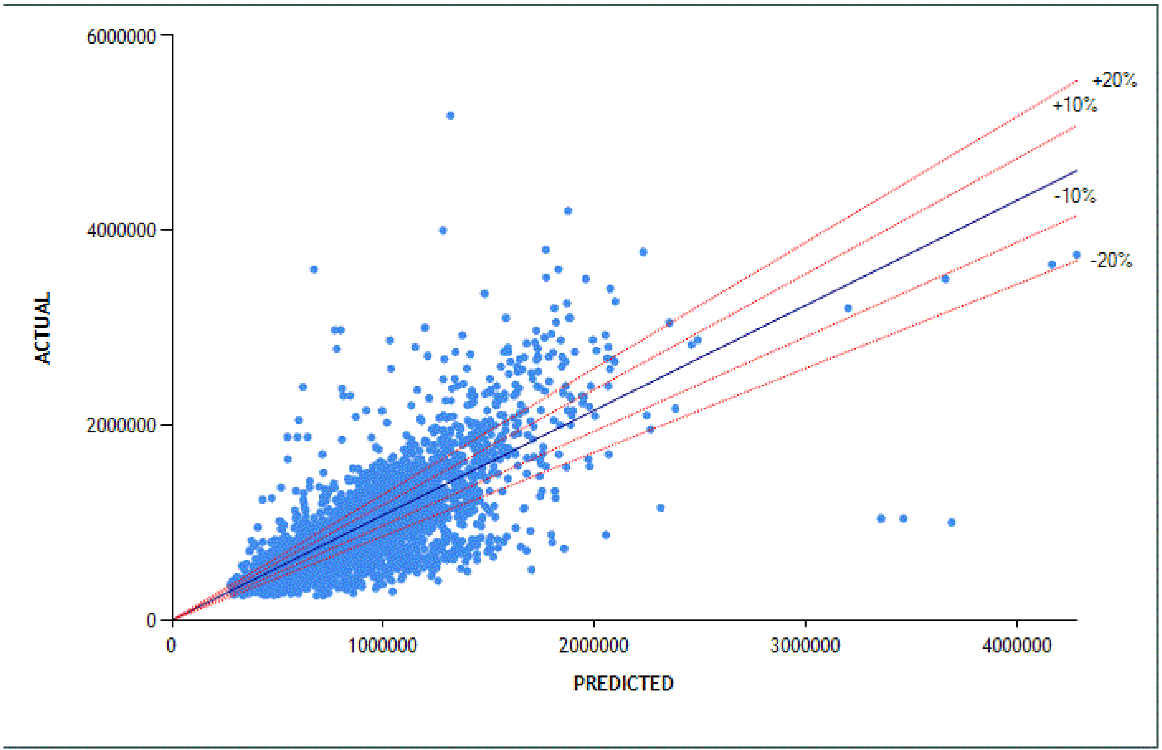
Model Structure and Calibration

Outlier Detection

The method for outlier detection is described in detail in [Appendix D Outlier Detection](). After following that process the sales used for the analysis became 3,390 of the original 4,078 (16.87%)

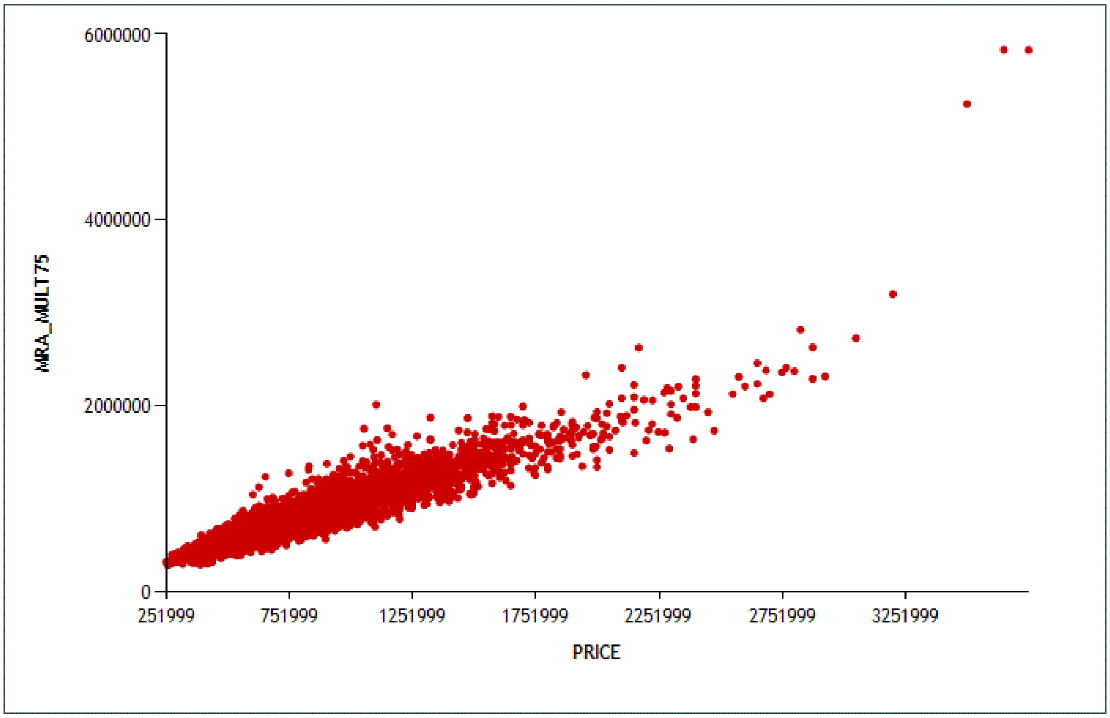
The estimation vs actual price for the baseline model and the model with 688 outliers removed follow.

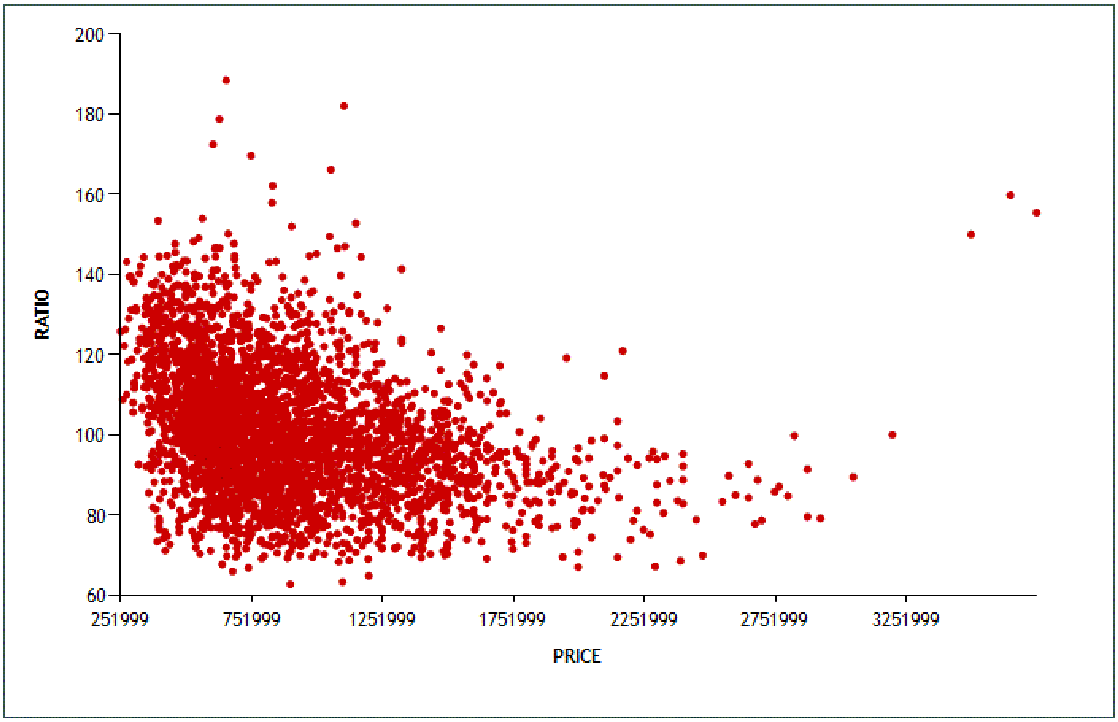
No Outlier Exclusion



Excluding 688 Outliers

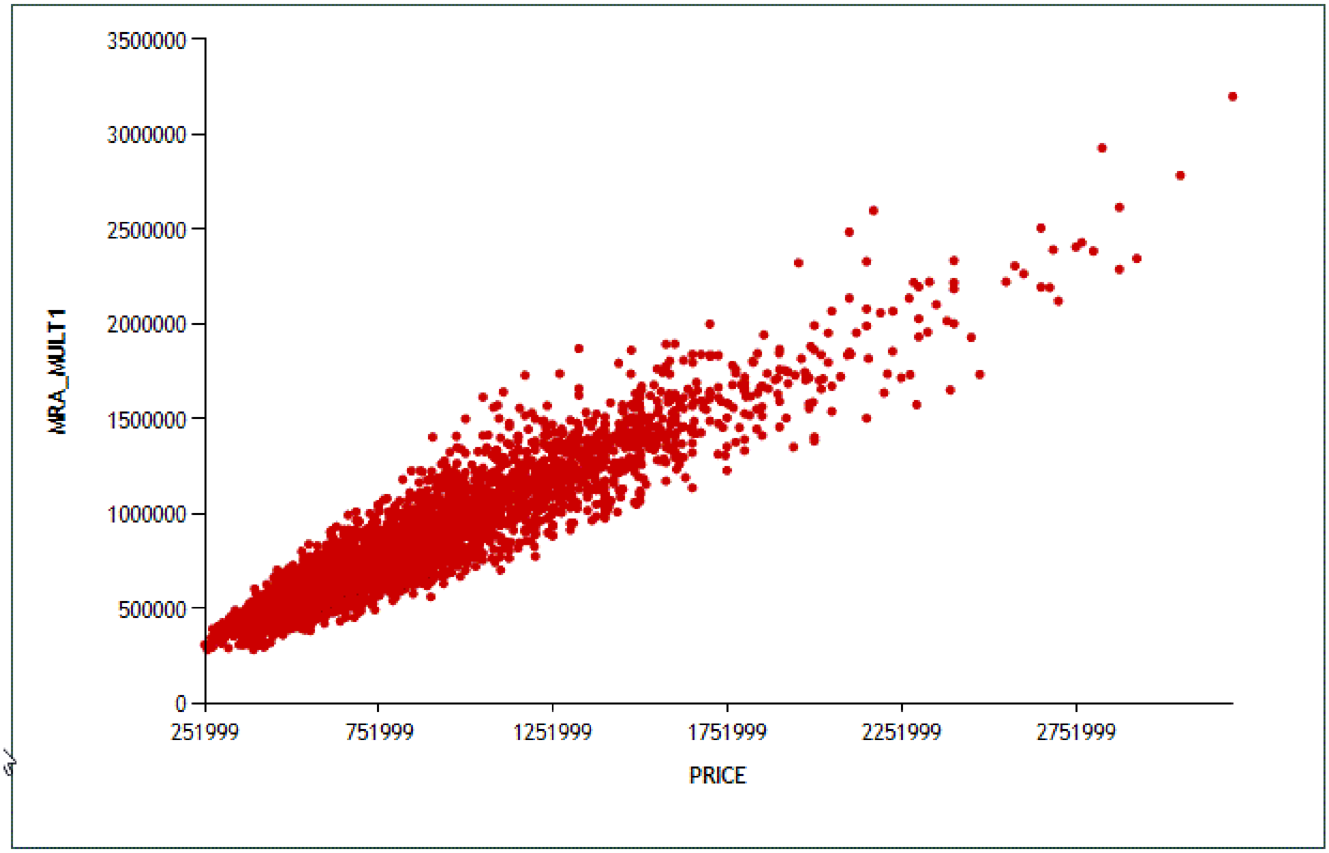
The two charts below were created after 688 outliers were removed by the IQR method. The first is a plot of the estimate vs. price. The second is a plot of the ratio of the estimate to price vs price. The second of the two shows a few points that are clearly outliers. Thirteen sales were removed based on these plots.





Second Pass Outlier Identification

Thirteen sales were removed and the model recalibrated. The plot of the estimate vs. price is repeated using 3077 sales.



Structure

Two model structures were evaluated. They are referred to as additive and multiplicative. The multiplicative model form is often referred to as a log-linear model. This confuses the model structure with the calibration process. It turned out that for this dataset, the multiplicative form of the model had the superior performance.

A portion of the entire model structure for both the additive and multiplicative model structures.is shown in the figure below. The main difference between the two is that the additive model is expressed in terms of dollar adjustments, whereas the multiplicative model is expressed in terms of percentage or fractional adjustment of a base value, for such variables as living area, the variable is raised to a power.

The performance statistics of the multiplicative model are superior to the additive model.

Multiplicative			Additive		
Dependent	PRICE		Dependent	PRICE	
Std Error for Estimate	0.1597		Std Error for Estimate	176,821.53	
Constant:	2,638.70		Constant:	203,015.50	
Attribute	Coeff	t value	Attribute	Coeff	t value
FIREPL			FIREPL		
2	1.0646		2	117,827.58	
3	1.099		3	137,853.19	
1	0.9978		1	-19,521.67	
6	0.9827		6	825,477.08	
4	1.0552		4	134,272.52	
5	0.8806		5	-36,350.49	
0	1		0	0	
AIRCOND			AIRCOND		
2	0.9458		2	-23,995.56	
1	1		1	0	
SQFTB	0.4606	32.847	SQFTB	160.2736	28.6371
SQRTAGE	-0.1413	-15.823	SQRTAGE	-25,451.42	-15.89
LOCF	0.453	9.6669	LOCF	300,058.47	5.7876
SQFTL	0.3168	28.856	SQFTL	85.3166	20.5985
BEDROOMS	0.0427	3.3097	BEDROOMS	7,629.81	2.5578
RENOV			RENOV		
1	1.1258		1	118,528.71	
0	1		0	0	
SITE			SITE		
3	0.8088		3	-152,735.70	
1	1.1155		1	106,623.93	
2	1		2	0	
Model Statistics			Model Statistics		
Total Valued	3377		Total Valued	3377	
R squared	0.860		R squared	0.812	
Adjusted R squared	0.857		Adjusted R squared	0.808	
COD	12.902		COD	16.099	
COV Median	16.169		COV Median	21.085	
COV Mean	15.858		COV Mean	20.701	
Median	0.997		Median	1.009	
Mean	1.013		Mean	1.025	
Weighted Mean Ratio	0.987		Weighted Mean Ratio	1.000	

The comparative performance statistics for the two model structures are [resented below. Clearly the Multiplicative form is superior.

Model	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
Additive	3377	1.009	1.025	1.000	0.260	0.212	16.099	20.701	1.025	0.002
Multiplicative	3377	0.997	1.013	0.987	0.222	0.161	12.902	15.858	1.026	-0.057

Location Factor and Owner Occupancy

In Lake View, the owner occupancy variable was not significant, but the location factor variable entered the model with a strong significance.

Multiplicative		
Attribute	**Coeff**	**t value**
LOCF	0.453	9.6669

The Final MRA Model

Multiplicative								
Dependent	PRICE							
Std Error for Estimate	0.1597							
Constant:	2,638.70							
Attribute	**Coeff**	**t value**	**Attribute**	**Coeff**	**t value**	**Attribute**	**Coeff**	**t value**
RQOS			GAR			**SQFTB**	0.4606	32.847
8	0.9424		**2**	0.971		**SQRTAGE**	-0.141	-15.823
9	0.9471		**1**	0.941		**LOCF**	0.453	9.6669
2	0.9751		**6**	1.09		**SQFTL**	0.3168	28.856
1	0.9753		**7**	0.956		**BEDROOMS**	0.0427	3.3097
6	0.973		**4**	0.99		RENOV		
7	0.9944		**5**	1.046		**1**	1.1258	
4	0.991		**8**	1.329		**0**	1	
5	0.9627		**3**	1		SITE		
19	0.8923		NUM			**3**	0.8088	
18	0.8987		**2**	0.785		**1**	1.1155	
13	0.9197		**3**	0.787		**2**	1	
12	0.9388		**1**	0.788		**Model Statistics**		
11	0.9935		**4**	0.796		**Total Valued**	3377	
10	0.9699		**5**	0.797		**R squared**	0.860	
17	0.8581		**6**	1		**Adjusted R squared**	0.857	
16	0.8697		CEILING			**COD**	12.902	
15	0.9443		**0**	0.933		**COV Median**	16.169	
14	0.9222		**1**	0.986		**COV Mean**	15.858	
20	0.7529		**2**	1		**Median**	0.997	
3	1		BSFN			**Mean**	1.013	
NGHCDE			**2**	0.958		**Weighted Mean Ratio**	0.987	
92	1.3159		**1**	1.029				
44	1.0958		**3**	1				
70	1.0592		EXTCON					
42	0.9719		**3**	0.94				
41	1.0046		**1**	0.988				
60	1.0406		**4**	0.893				
31	0.9256		**2**	1				
63	1.5261		FIREPL					
32	0.9322		**2**	1.065				
34	1.0874		**3**	1.099				
120	1.1089		**1**	0.998				
200	0.9699		**6**	0.983				
12	0.8986		**4**	1.055				
50	0.8065		**5**	0.881				
150	1.1622		**0**	1				
110	0.9141		AIRCOND					
11	0.7072		**2**	0.946				
62	1.2887		**1**	1				
84	1.2326		BASMENT					
22	0.9271		**2**	0.859				
93	1.2336		**3**	0.973				
81	1		**4**	0.958				
			1	1				

MRA Stats by NBHD

NBHD	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
11	80	1.054	1.022	0.979	0.390	0.213	17.670	20.792	1.044	-0.150
12	223	1.002	1.015	0.988	0.266	0.179	14.622	17.599	1.028	-0.103
22	51	1.026	1.014	0.985	0.257	0.172	13.609	16.923	1.030	-0.228
31	265	0.984	1.013	0.991	0.232	0.165	13.535	16.261	1.023	-0.087
32	189	0.990	1.013	0.983	0.232	0.168	13.638	16.551	1.031	-0.097
34	47	1.008	1.012	0.986	0.172	0.155	11.625	15.354	1.026	-0.171
41	161	0.986	1.009	0.998	0.146	0.138	10.546	13.682	1.011	-0.028
42	154	0.986	1.012	0.988	0.222	0.158	12.703	15.567	1.024	-0.071
44	55	0.998	1.010	0.980	0.232	0.138	11.128	13.665	1.030	-0.124
50	29	0.981	1.019	1.004	0.229	0.207	16.314	20.285	1.015	0.026
60	24	1.036	1.021	0.986	0.392	0.208	17.487	20.332	1.035	-0.008
62	21	1.001	1.024	0.940	0.361	0.222	19.314	21.689	1.090	-0.228
63	59	0.986	1.018	0.982	0.247	0.200	15.532	19.605	1.037	-0.125
70	410	0.990	1.011	0.991	0.178	0.148	11.472	14.613	1.020	-0.060
81	431	1.001	1.012	0.991	0.223	0.153	12.464	15.168	1.021	-0.058
84	332	1.001	1.012	0.982	0.228	0.159	12.850	15.713	1.031	-0.074
92	26	1.036	1.015	0.984	0.253	0.177	13.350	17.438	1.032	-0.112
93	324	1.008	1.014	0.988	0.228	0.170	13.525	16.734	1.027	-0.039
110	96	1.008	1.012	0.988	0.198	0.159	12.465	15.733	1.025	-0.083
120	150	0.995	1.010	0.991	0.194	0.141	11.089	13.979	1.019	-0.038
150	110	1.006	1.010	0.995	0.197	0.143	11.395	14.180	1.016	-0.035
200	140	1.015	1.010	0.966	0.188	0.143	11.370	14.182	1.045	-0.151
Total	3377	0.997	1.013	0.987	0.222	0.161	12.902	15.856	1.026	-0.057

Comparable Sales Valuation

The comparable sales valuation method was described in the Comparable Sales Valuation subsection in the preceding section on Rogers Park.

Comp Sale Results on the Sales Sample

The methodology used for the comparable sale selection process involves setting selection parameters by "iteration". The first iteration has the tightest specification on the sales that will be consider for analysis. The second iteration loosens up a bit on which sales will be considered and so on until enough iterations have been defined to value all properties. The iteration rules used in Lake View are shown below with explanation of each iteration's restriction on comp sales.

Rules	Iteration 1:	Iteration 2:	Note
Buffer Size	1000	3000	Max Distance in feet used for the search
Comparables	5	5	Required Number of Comparabl Sales
Weighting	0.5	0.5	Distance/Similarity weighting
SQFTB	+/- 20%	+/- 40%	Max % Diff in Sq Ft Liv Area
SQFTL	+/- 20%	+/- 40%	Max % Diff in Lot Value
AGE	+/- 20	+/- 40	Max Diff in Age
NGHCDE	=		Exact Match of NBHD

Iteration	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
1	2712	1.006	1.023	1.002	0.201	0.156	12.170	15.271	1.021	-0.049
2	665	1.016	1.028	0.992	0.253	0.179	14.429	17.419	1.036	-0.059
Total	3377	1.007	1.024	1.000	0.213	0.161	12.641	15.718	1.024	-0.052

Method Comparison

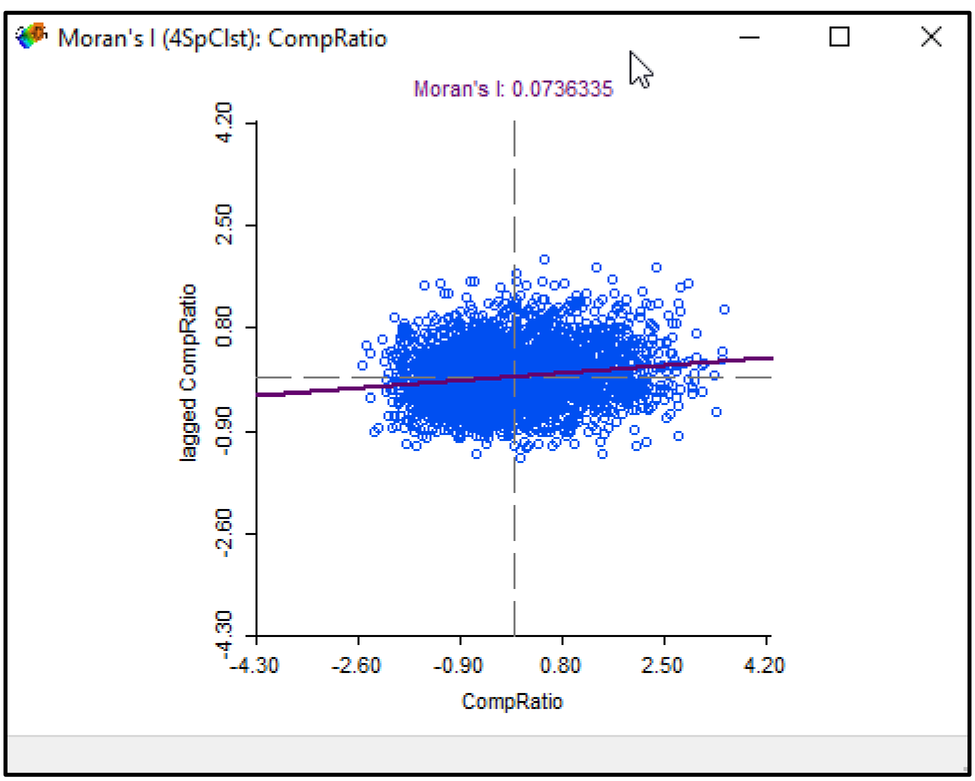
As can be seen, the comp sale model has better performance statistics when considering overall accuracy (COD) and vertical equity as the corresponding MRA model use to adjust the sales. In most cases comp sales outperforms MRA. When it does not, it means the MRA model has accounted for location quite well.

Model	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
Additive	3377	1.009	1.025	1.000	0.260	0.212	16.099	20.701	1.025	0.002
Multiplicative	3377	0.997	1.013	0.987	0.222	0.161	12.902	15.858	1.026	-0.057
Comp Sales	3377	1.007	1.024	1.000	0.213	0.161	12.641	15.718	1.024	-0.052

Spatial Dependency

A means to verify the locational stability of the estimates is provided computing Local Indicators of Spatial Association often referred to as LISA. Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. In mass appraisal it is customary to look for spatial clusters in the ration of appraised value to sale price. The plot below has as its X Axis the z-transform of Ratio, defined as $z = (x - \mu)/\sigma$ where x is the individual ratio, μ is the mean ratio and σ is the standard deviation of the ratios in the sample. The Y axis is the average of the five nearest transformed ratios not including the ratio of the X Axis. The fact that there very little slope to the plot is a good indication that there are no spatial clusters of high or low ratios.

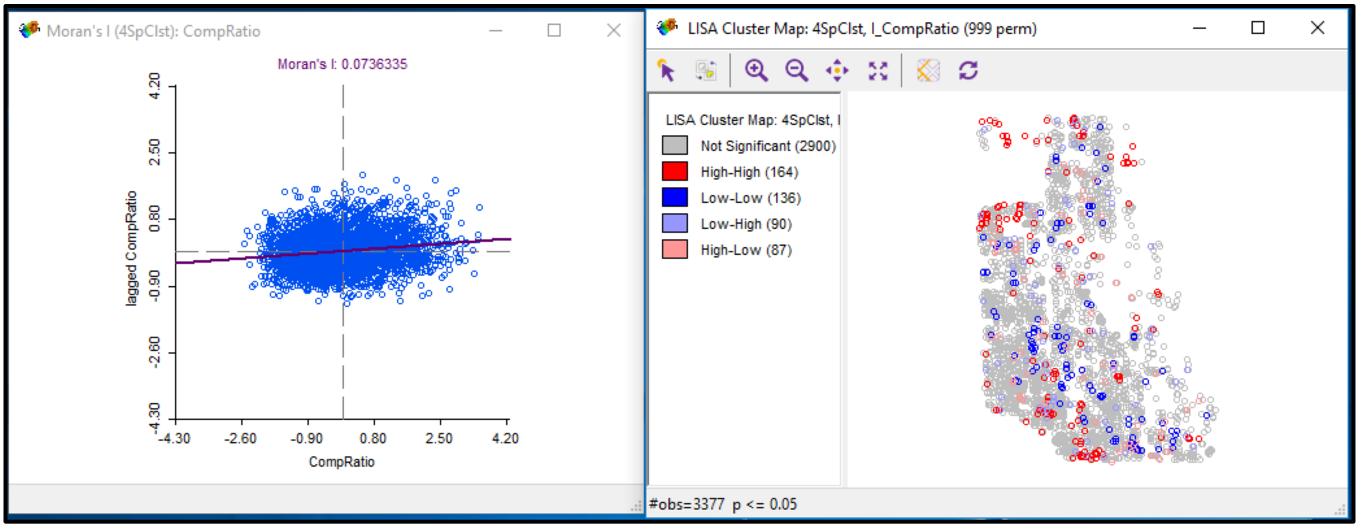
The Moran's I global statistic of 0.0736335 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values.



Consider the side-by-side paring of the scatterplot with the cluster plot. Five plots follow to indicate the location of specific clusters.

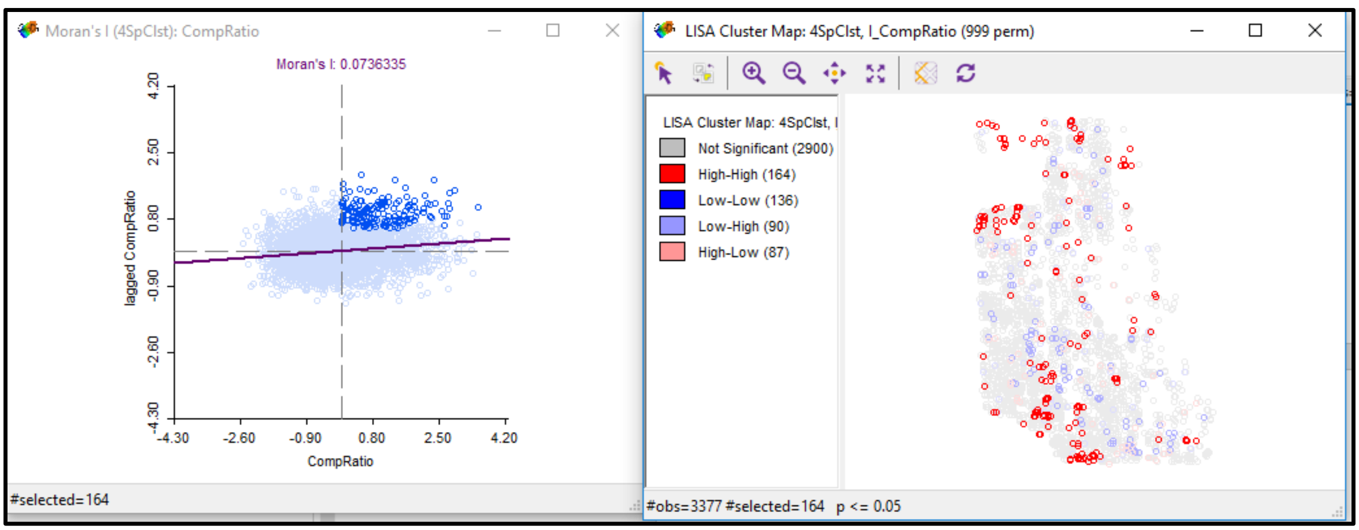
All Points

All points showing with indication of significant categories of clusters – high ratios near high ratios, low ratios near low ratios and then low ratios near high ratios and high ratios near low ratios. The next four images will focus on each category.



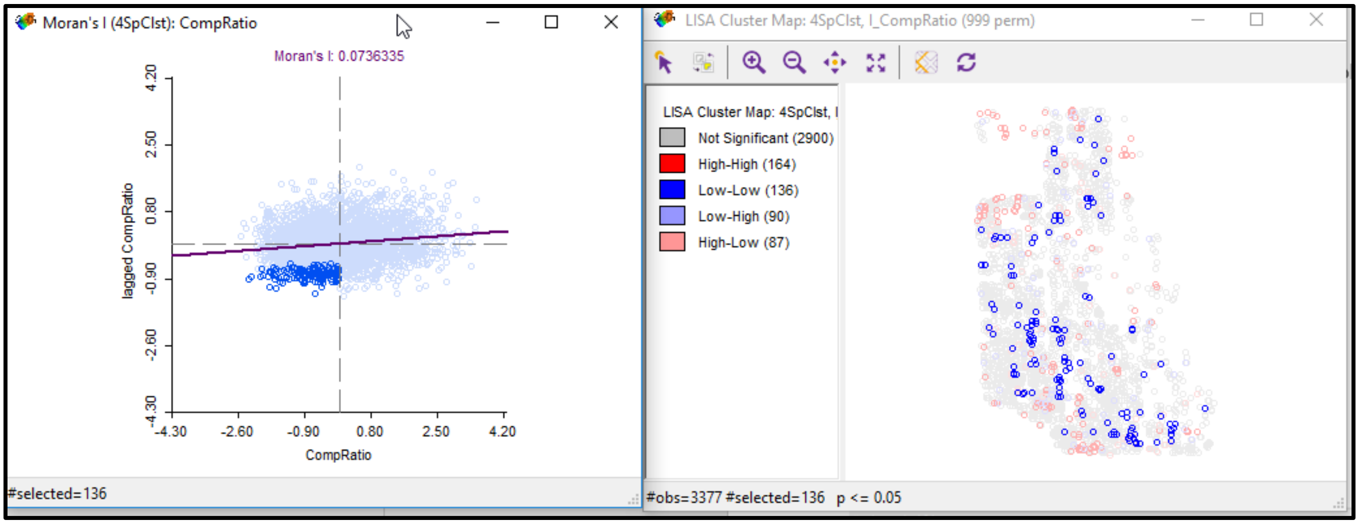
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



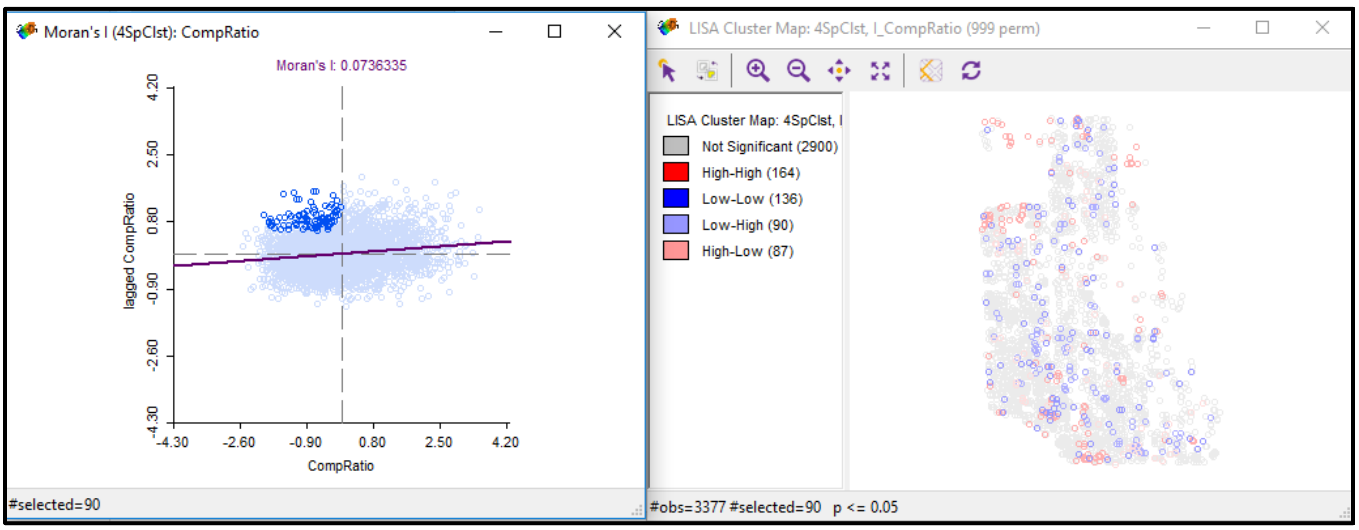
Low near Low Ratios

The low near low ratios are also spread uniformly around the town.



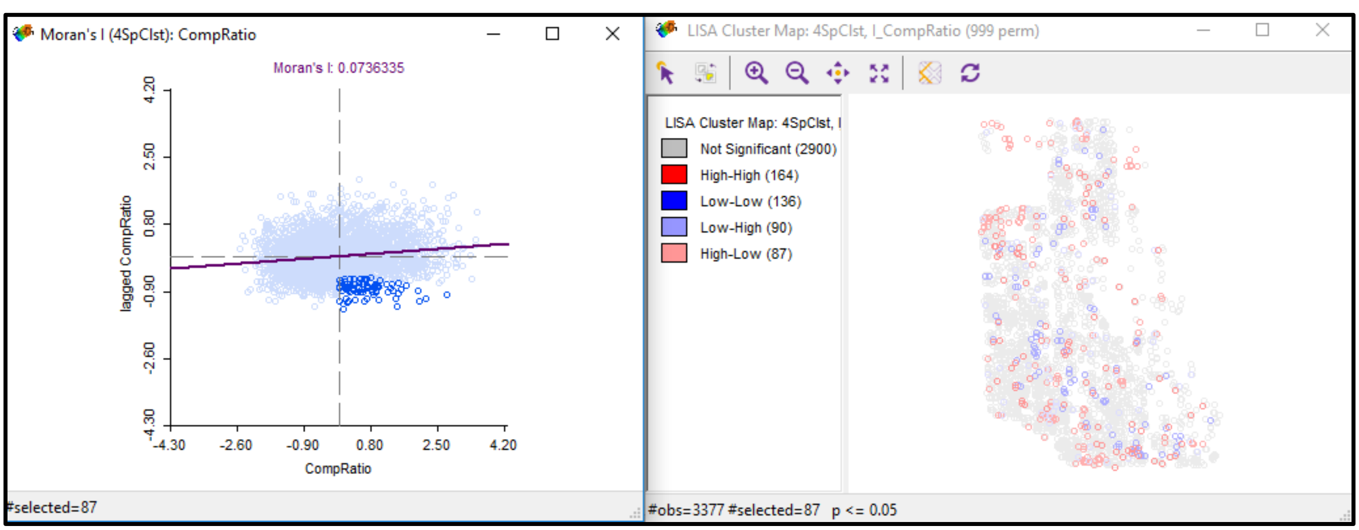
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.



Additional Support for Spatial Uniformity

The median ratios by NBHD are very close to one another. The two measures at the far right of the table are the standard deviation and standard error of the median ratio respectively. Both indicate little variability by NBHD.

NBHD	Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB	STD DEV
11	80	1.009	1.022	0.980	0.275	0.185	15.371	18.057	1.043	-0.207	0.020
12	223	1.023	1.037	1.007	0.251	0.181	14.374	17.409	1.030	-0.127	StdERR
22	51	1.046	1.037	1.007	0.247	0.169	12.897	16.258	1.030	-0.254	0.004
31	265	1.007	1.031	1.004	0.241	0.170	13.479	16.500	1.027	-0.123	
32	189	1.011	1.027	0.998	0.236	0.166	13.138	16.168	1.028	-0.085	
34	47	1.016	1.018	0.991	0.206	0.167	12.435	16.370	1.027	-0.176	
41	161	1.008	1.025	1.010	0.177	0.144	10.605	14.011	1.015	-0.056	
42	154	0.983	1.021	0.997	0.228	0.162	12.984	15.834	1.025	-0.074	
44	55	1.001	1.005	0.993	0.115	0.114	8.662	11.364	1.012	-0.028	
50	29	0.966	1.021	1.003	0.248	0.206	16.501	20.140	1.018	0.015	
60	24	1.004	1.017	0.975	0.392	0.239	19.940	23.501	1.043	-0.011	
62	21	0.965	0.994	0.922	0.327	0.194	17.112	19.537	1.078	-0.191	
63	59	1.001	1.038	0.993	0.231	0.207	15.597	19.934	1.046	-0.160	
70	410	0.997	1.015	0.995	0.178	0.146	11.180	14.393	1.020	-0.070	
81	431	1.010	1.023	1.001	0.220	0.155	12.465	15.124	1.022	-0.068	
84	332	1.019	1.030	1.008	0.211	0.158	12.339	15.302	1.022	-0.038	
92	26	1.040	1.024	0.992	0.254	0.179	13.357	17.488	1.032	-0.113	
93	324	1.004	1.022	1.001	0.223	0.172	13.532	16.811	1.021	-0.020	
110	96	0.979	1.010	0.983	0.196	0.165	13.257	16.292	1.027	-0.096	
120	150	1.014	1.029	1.009	0.205	0.151	11.620	14.616	1.020	-0.035	
150	110	1.011	1.022	1.003	0.222	0.153	12.429	15.005	1.019	-0.049	
200	140	1.006	1.014	0.997	0.135	0.119	8.750	11.703	1.017	-0.032	
Total	3377	1.007	1.024	1.000	0.213	0.161	12.641	15.702	1.024	-0.052	

Hyde Park

Summary

Key points about Hyde Park Modeling and Valuation

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- Owner Occupancy data was considered, but did not prove to be statistically significant
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Hyde Park was valued using the Multiple Regressions Analysis direct market comparison method of valuation
- Log linear models introduce what is called a retransformation bias
- The bias is corrected to ensure that the weighted mean ratio of estimated value to sale price is 1.000
- Performance statistics were well within IAAO standards
- Additional care was given to the model structure to ensure logical changes in value when comparing 2015 values to 2018 values by neighborhood and class

The Data

Sales Counts

Starting with the Hyde Park combined sales and subjects file of 63,540 records where Amount1 is the sales price

Filter	Count
AMOUNT1	
>0	37,641
>65,000	24,015
<790,000	23,676
multi<1	23,487
sqftb<9,000	23,312
Sale Year>2012	4,941
Puremarket=1	2,819
Starting Count	2,819

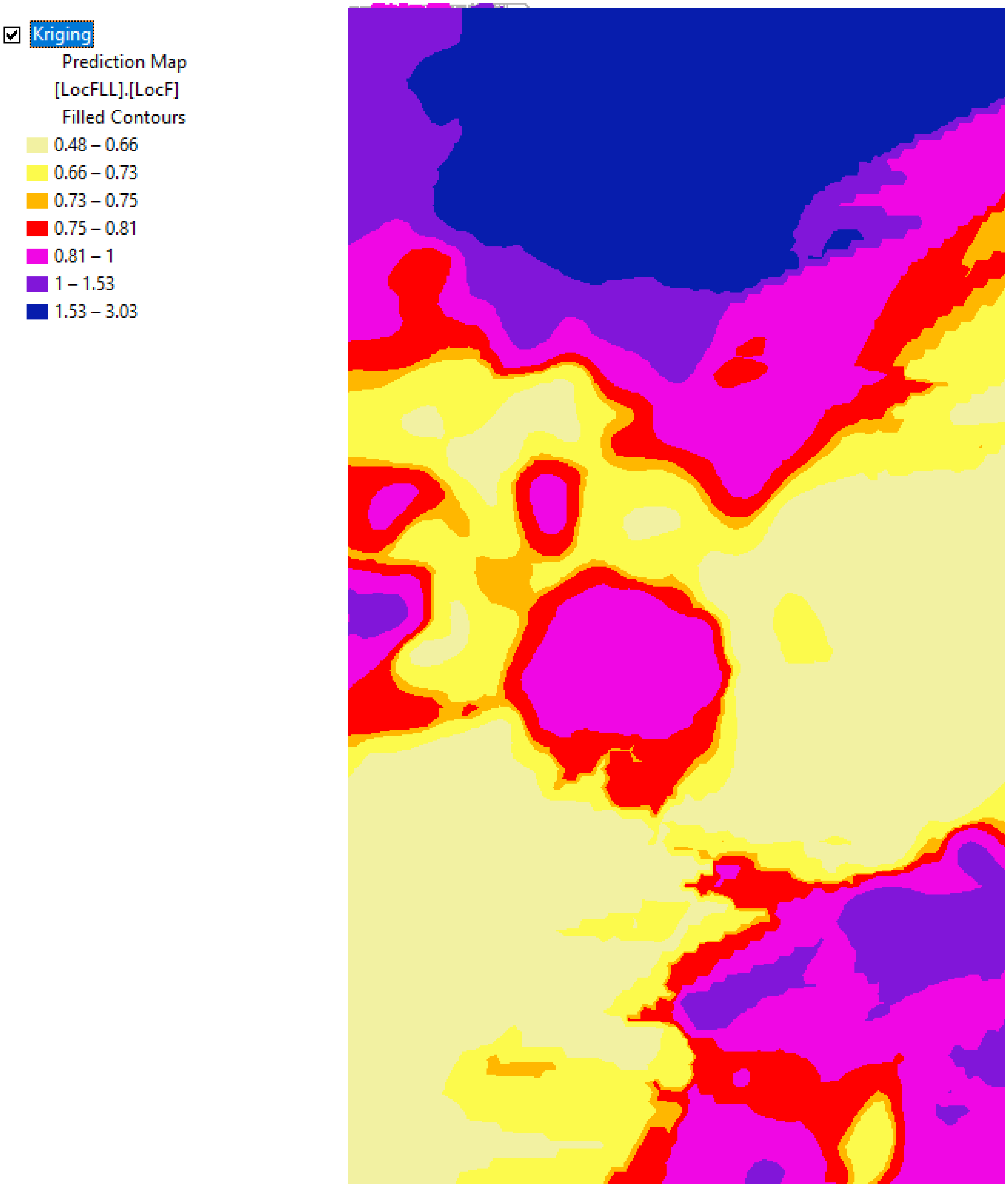
Data Fields

The initial list of data items available for analysis is provided in Appendix A Variable Definitions. Certain additional data fields were created. Those that were relevant to Hyde Park are:

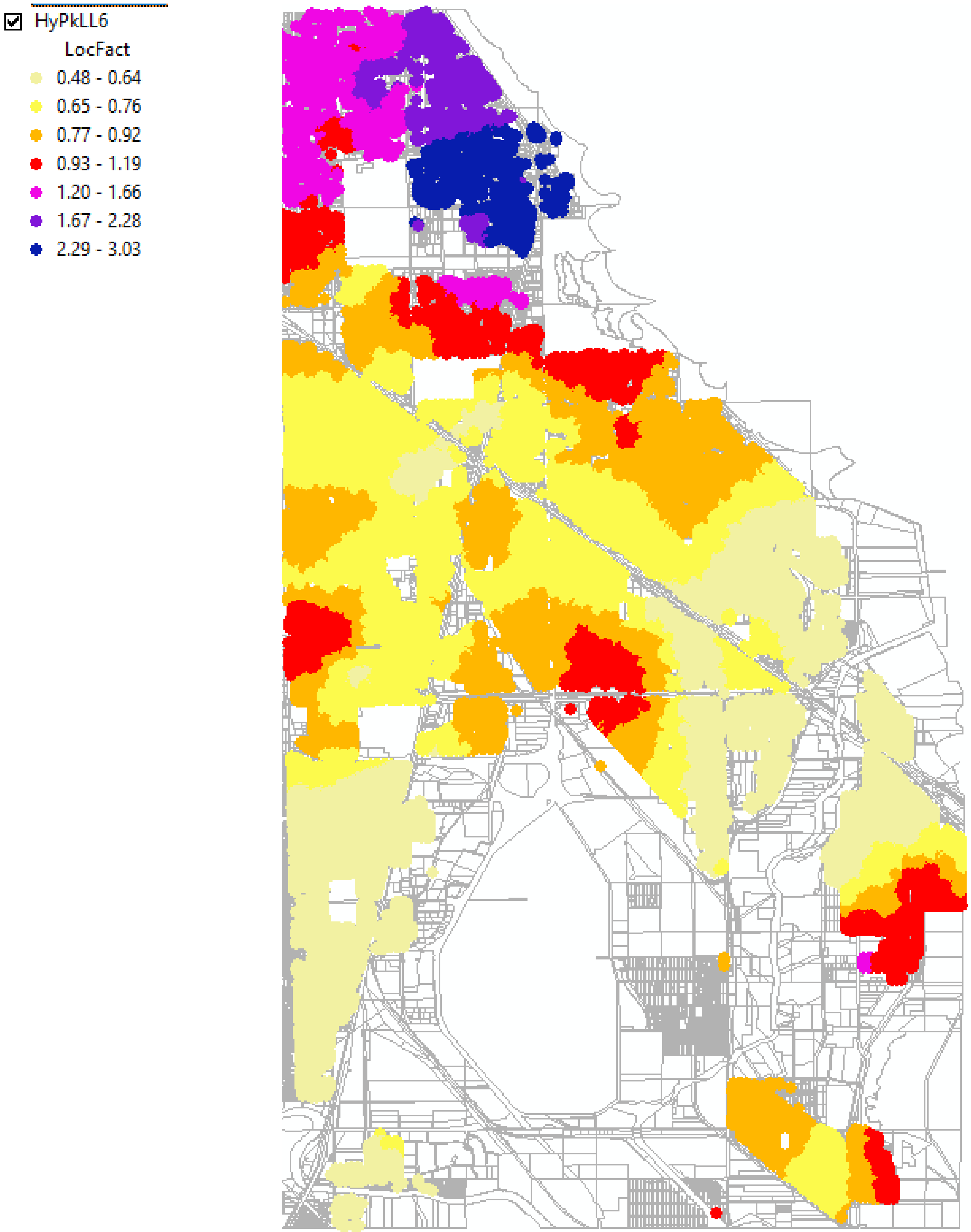
Location Factor

The process employed for determining the Location Factors for Lake Township is outlined in Appendix B Location Factor..

The resultant surface is shown in the image below.



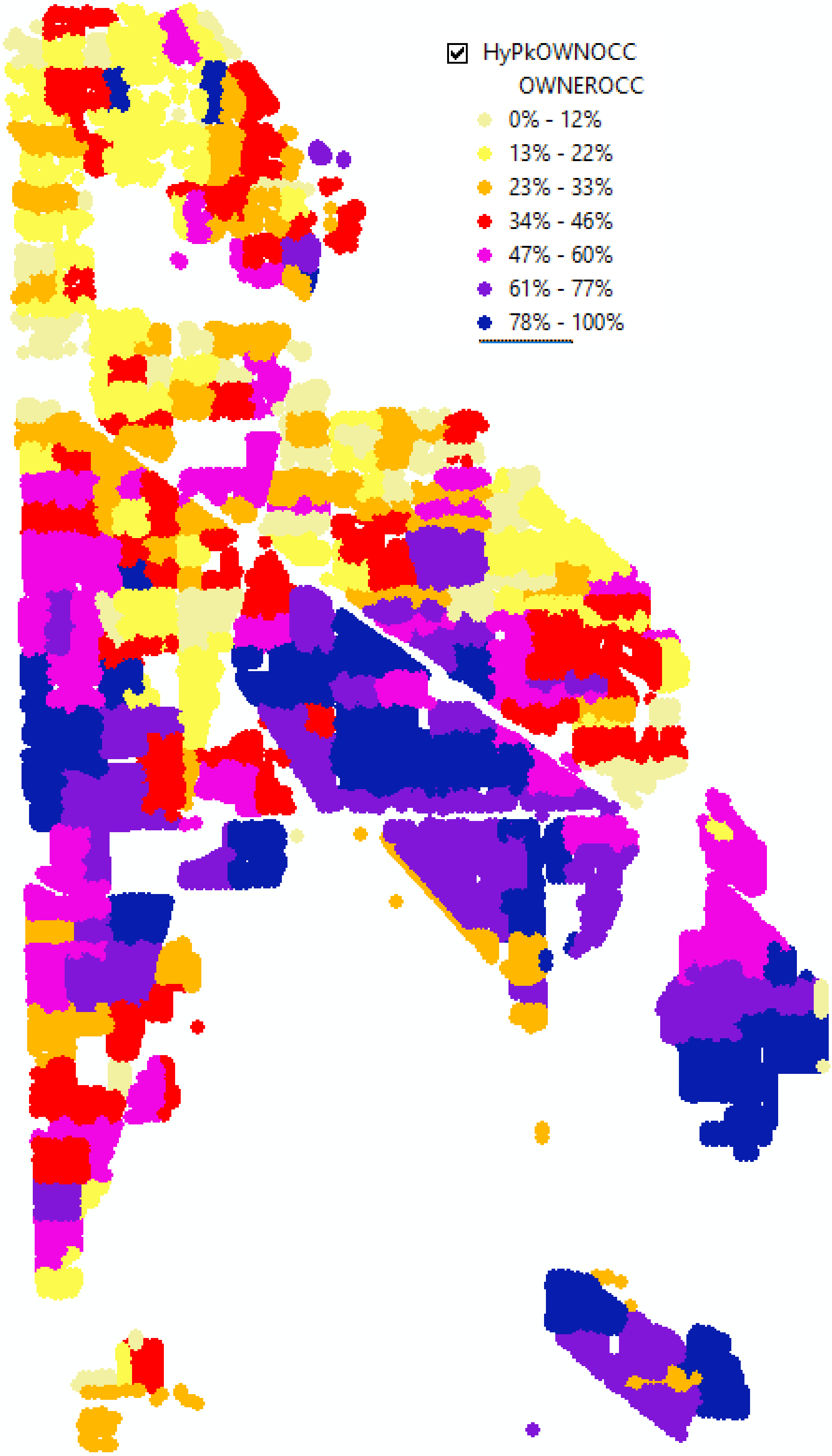
When applied to all properties the thematic map of Location Factor is given in the next image.



Owner Occupancy

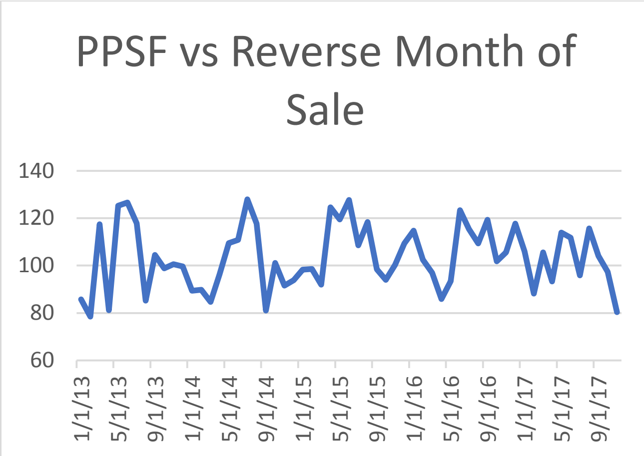
Owner Occupancy data is available at the Census Block Group level. County data is organized at several levels including parcel, block and neighborhood. Since the two geographies are organized differently, they were joined using what is called a "spatial join".

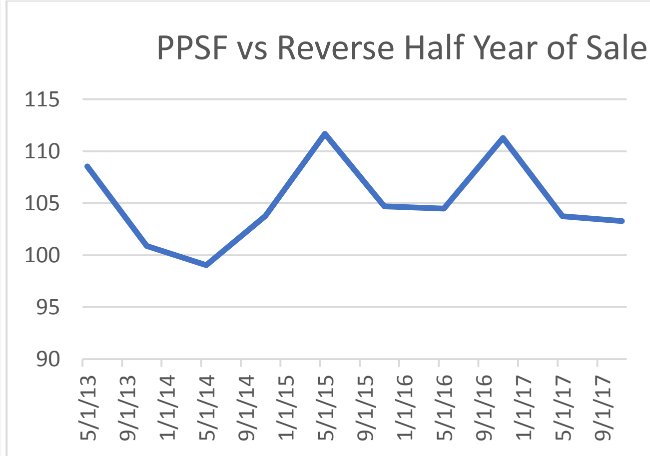
When joined the result becomes a parcel fabric with spatially interpolated owner occupancy data. The owner occupancy data is thus made available at the individual parcel level and becomes a candidate variable in an MRA Model.



Reverse Half of Sale

The sales used in the analysis span a period of five years. To allow for time trending the sales to the valuation date, first, a reverse month of sale is computed. If the sale took place in December of 2017, the reverse month of sale (RMOS) is 1, November 2017, RMOS is 2, all the way back to January of 2013, RMOS is 60. In terms of using this variable directly in the model to be discussed, it is converted into a Reverse Half Year of Sale (RHOS). The rationale for this approach can be seen in the following two charts. The first shows Price per Square Foot (PPSF) and Count vs. RMOS. The second chart shows the same two variables vs RHOS. The RMOS variable is too granular and "noisy" as compared to the RHOS variable.



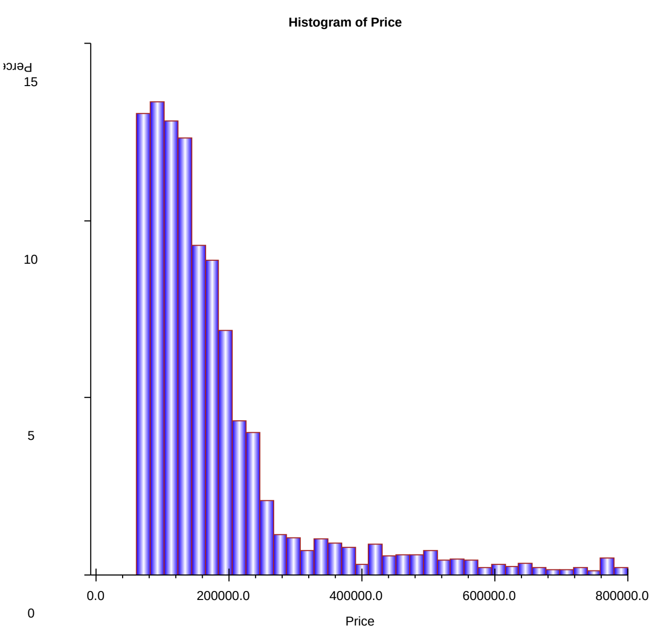


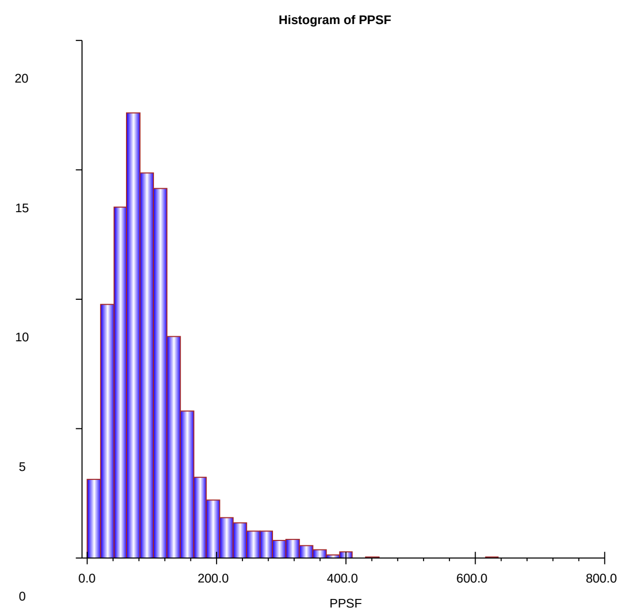
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.

Price and Price per Square foot

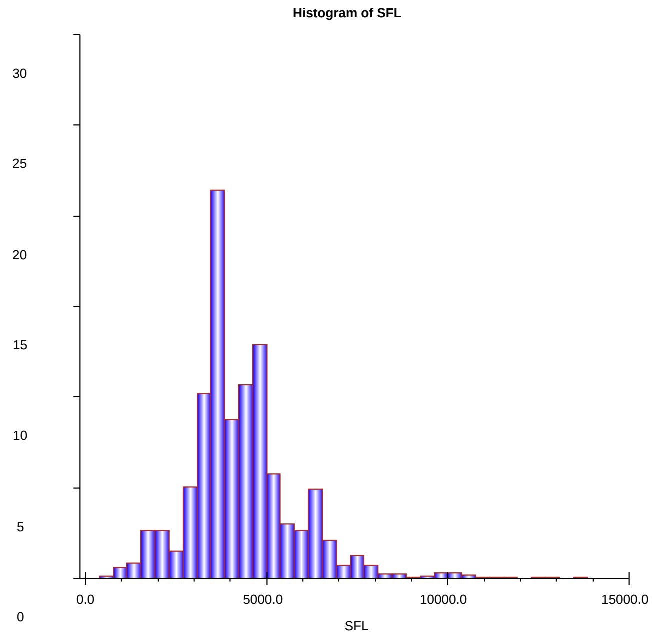
The price histogram fits the range of prices specified at the outset of modeling, namely a range of $65,000-$790,000. The price per square foot range indicates what are likely to be outlier situations. In other words, $400 per square foot and above is not likely to represent a true open market situation.

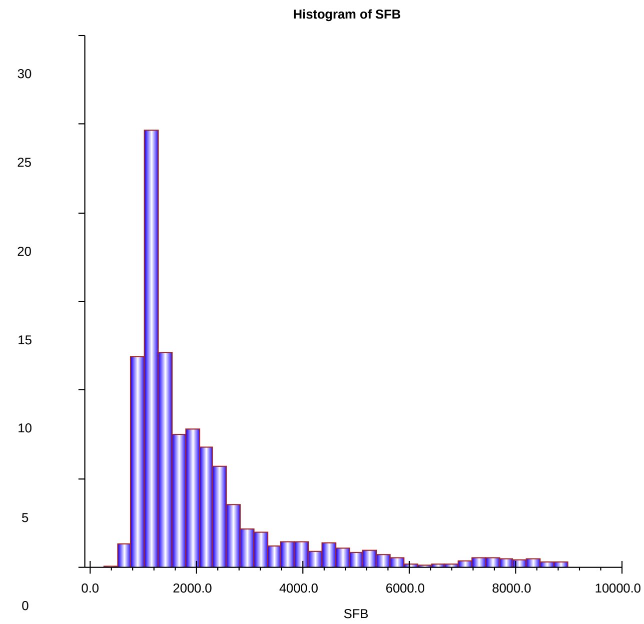




Square foot Land and Building

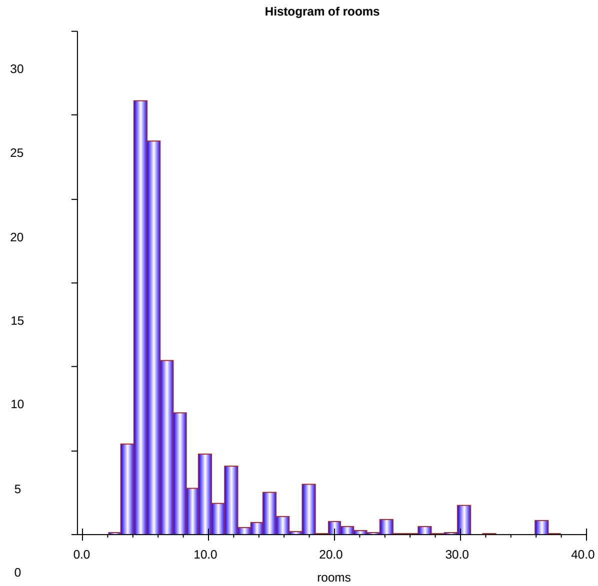
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not. It is interesting to note the group of land sizes at about 10,000 sq. ft. and the building sizes around 8,000 sq. ft. They seem to be separate categories of properties. The scatterplot of sq. ft. building vs. sq. ft. land shows they are not the same groups. Rather there is a distinct group of 10,00 sq. ft. land and another distinct group of 8,00 sq. ft. building.

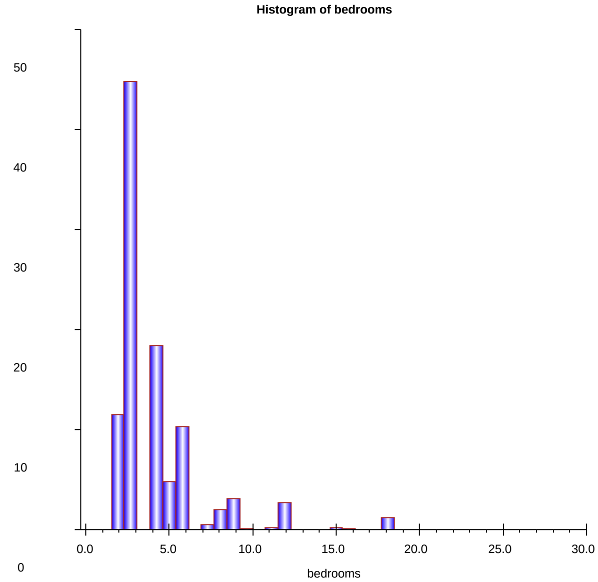




Rooms and Bedrooms

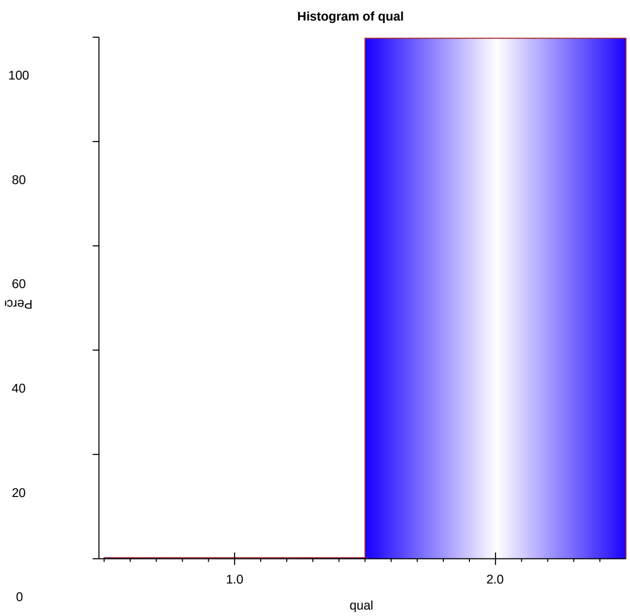
Looks like there are homes with 36 rooms and some with 18 bedrooms. For the specific case of 36/18, it looks as if they are six-unit apartment buildings with each unit having 6 rooms and 3 bedrooms.

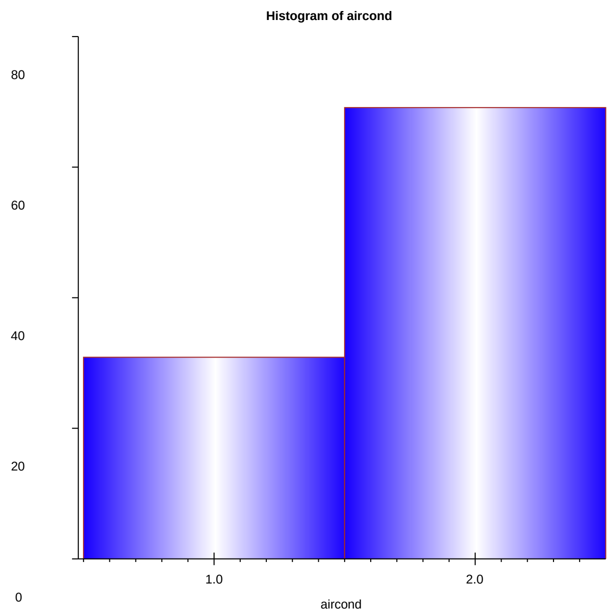




Quality and Air Conditioning

The quality of construction variable has little variability and almost certainly will not be a useful variable. On the other hand, air conditioning may well be useful in a model.

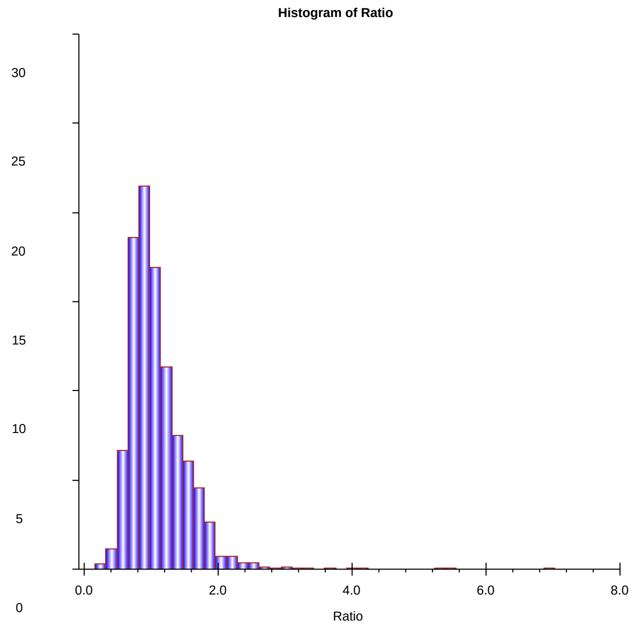


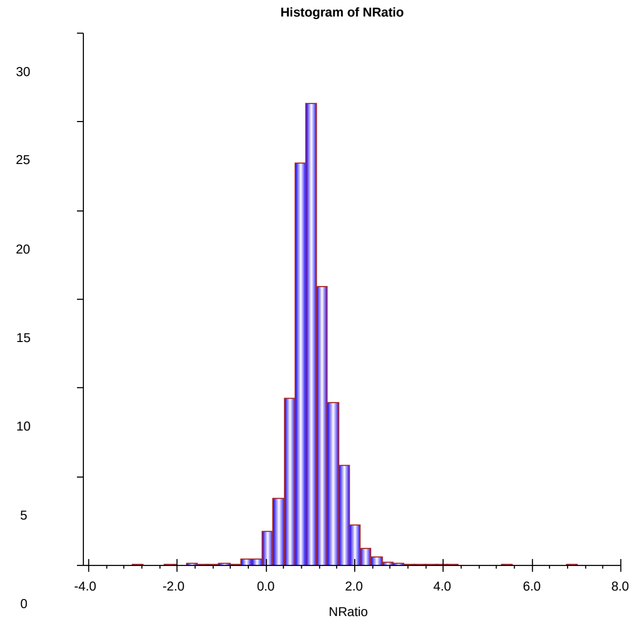


Model Structure and Calibration

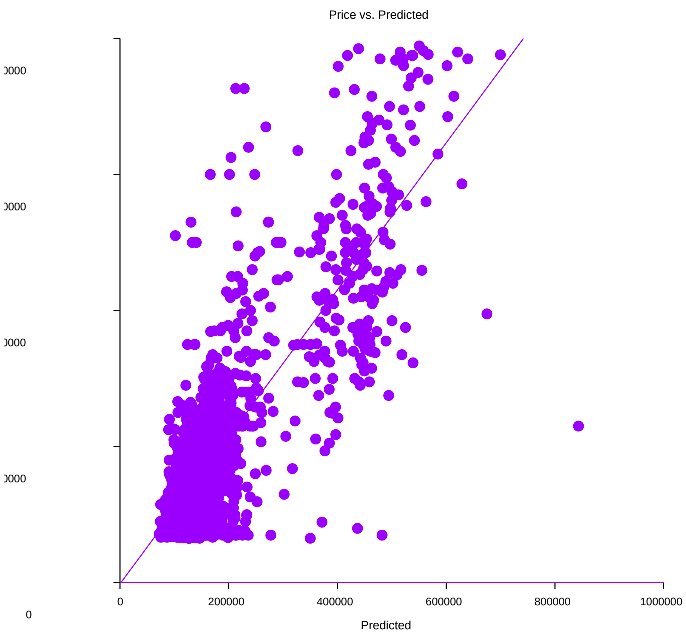
Outlier Detection

The method for outlier detection is described in detail in Appendix D Outlier Detection.. After following that process the sales used for the analysis became 2,157 of the original 2,767 (20.04%)



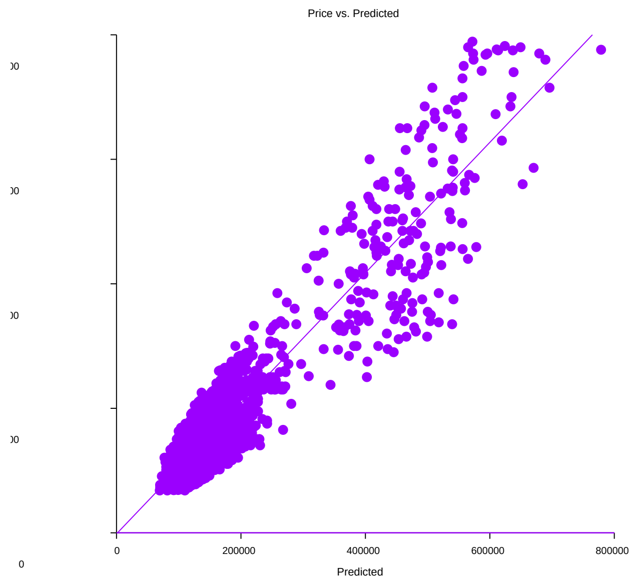


No Outlier Exclusion



Excluding Outliers

The chart below was created after 568 outliers (20.5%) were removed by the IQR method. It was applied in two passes because this was a very noisy dataset. This means that an initial outlier detection was made on the first modeling pass. When the new model was calibrated with outliers removed, a second outlier detection was performed. The model used in valuation was the one following the second outlier detection.



Structure
Once again, two model structures were evaluated, additive and multiplicative. The multiplicative model form is often referred to as a log-linear model. This confuses the model structure with the calibration process. It turned out that for this dataset, the multiplicative form of the model was chosen for further processing. The Horizontal Equity performance statistics of the multiplicative model are superior to the additive model. The Vertical Equity statistics are marginally better for the additive model, but not enough so to decide in its favor. As will be seen later in this report, the actual model used was a segmented model with the very high-end properties separated from the rest.

Model	Count	Median	WgtMean	COD	PRD	PRB
Multipicative	2168	0.981	0.971	20.337	1.057	-0.083
Additive	2168	1.009	1.000	22.800	1.053	-0.029

Retransformation Bias
The reader will note that the median and weighted mean for the multiplicative model are lower than that for the additive model. The topic is discussed in detail in Appendix E Retransformation Bias. Final value estimates produced by the Multiplicative model were corrected by dividing by 0.971.

Location Factor and Owner Occupancy
In Hyde Park, the owner occupancy variable was not significant, but the location factor variable entered the model with a strong significance.

The Final MRA Models

NBHD 10

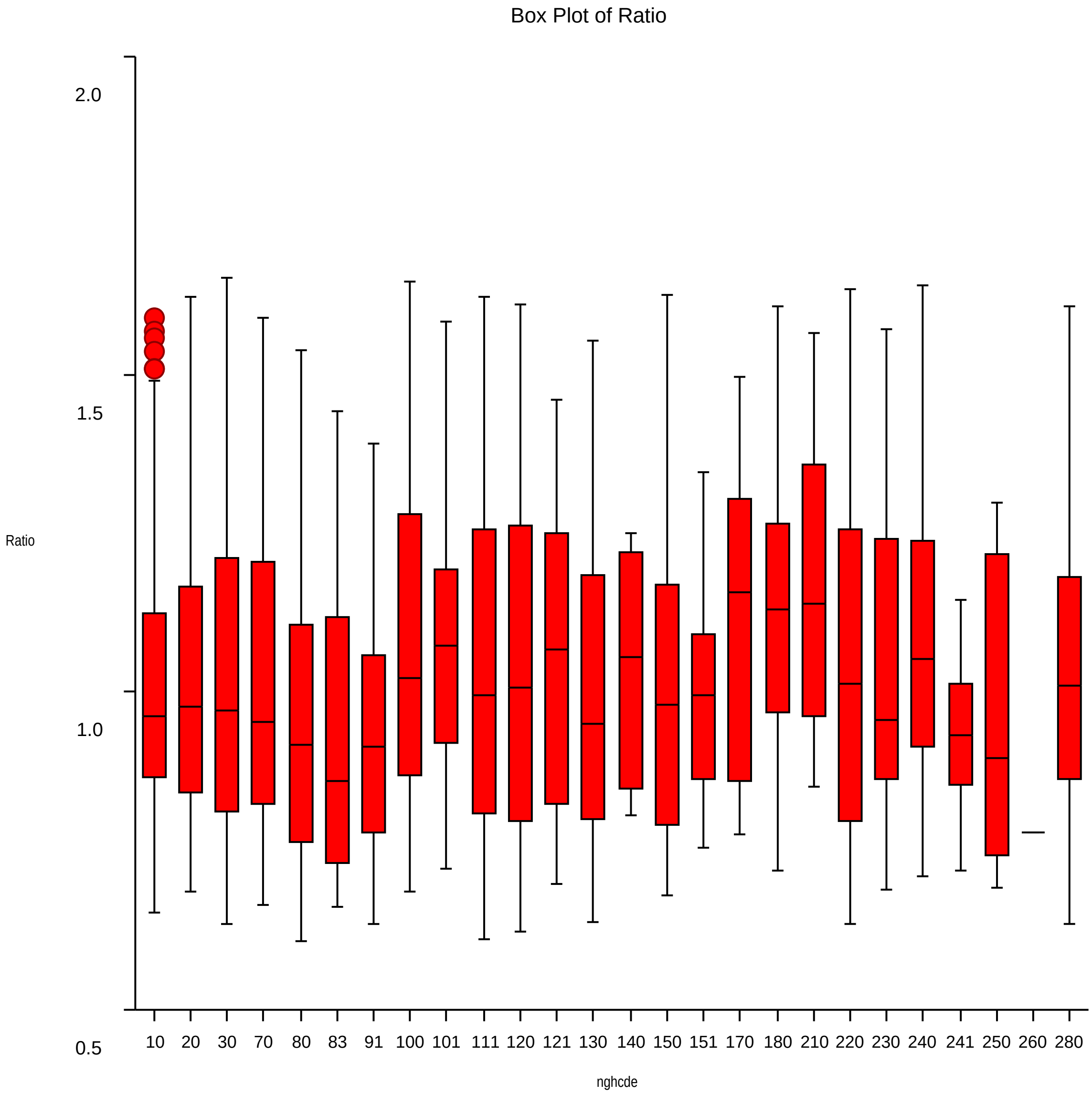
NBHD 10		
Variable	Coefficient	T Value
Intercept	8.004152	13.107
lnSFB	0.397646	4.696
lnSFL	0.1840567	4.176
lnAgeT	-0.1234992	-4.068
lnBEDS	0.2092469	2.842
lnFIXT	0.4475397	3.724
lnLocF	0.4912548	5.916
(CLASS_2_6	-0.2326369	-4.078
(CLASS_2_9	-0.579388	-3.347
(RH7=1)	-0.07615889	-1.186
(RH8=1)	-0.510699	-7.726
(RH9=1)	-0.2592988	-3.937
(RH10=1)	-0.2159355	-3.199
(NUM5=1)	-0.4258315	-4.675
(NUM6=1)	0.6119503	12.582

Remaining NBHDs

Remaining NBHDs		
Variable	Coefficient	T Value
Intercept	8.449196	32.249
lnSFB	0.5245894	15.84
lnSFL	0.1400339	6.479
lnAgeT	-0.1721225	-7.295
lnBEDS	0.08122066	3.017
lnFIXT	0.2146985	4.808
lnLocF	0.6199709	18.589
(CLASS_2_6=1)	0.07187945	1.977
(CLASS_2_9=1)	-0.6036162	-4.32
(CLASS_2_95=1)	-0.0850668	-2.492
(RH7=1)	-0.1205422	-6.903
(RH8=1)	-0.1383876	-6.513
(RH9=1)	-0.1252984	-6.07
(RH10=1)	-0.1212596	-5.423
(NUM5=1)	-0.3498941	-9.236
(NUM6=1)	0.2529937	10.567
(NBHD20=1)	0.5054465	10.345
(NBHD30=1)	0.1866431	8.093
(NBHD70=1)	0.1745059	6.305
(NBHD120=1)	0.106257	4.702
(NBHD150=1)	0.4117913	8.039
(NBHD220=1)	0.0803887	2.78
(NBHD230=1)	0.1182545	2.975
(GAR6=1)	-0.2407273	-2.842

Model Performance Stats by NBHD

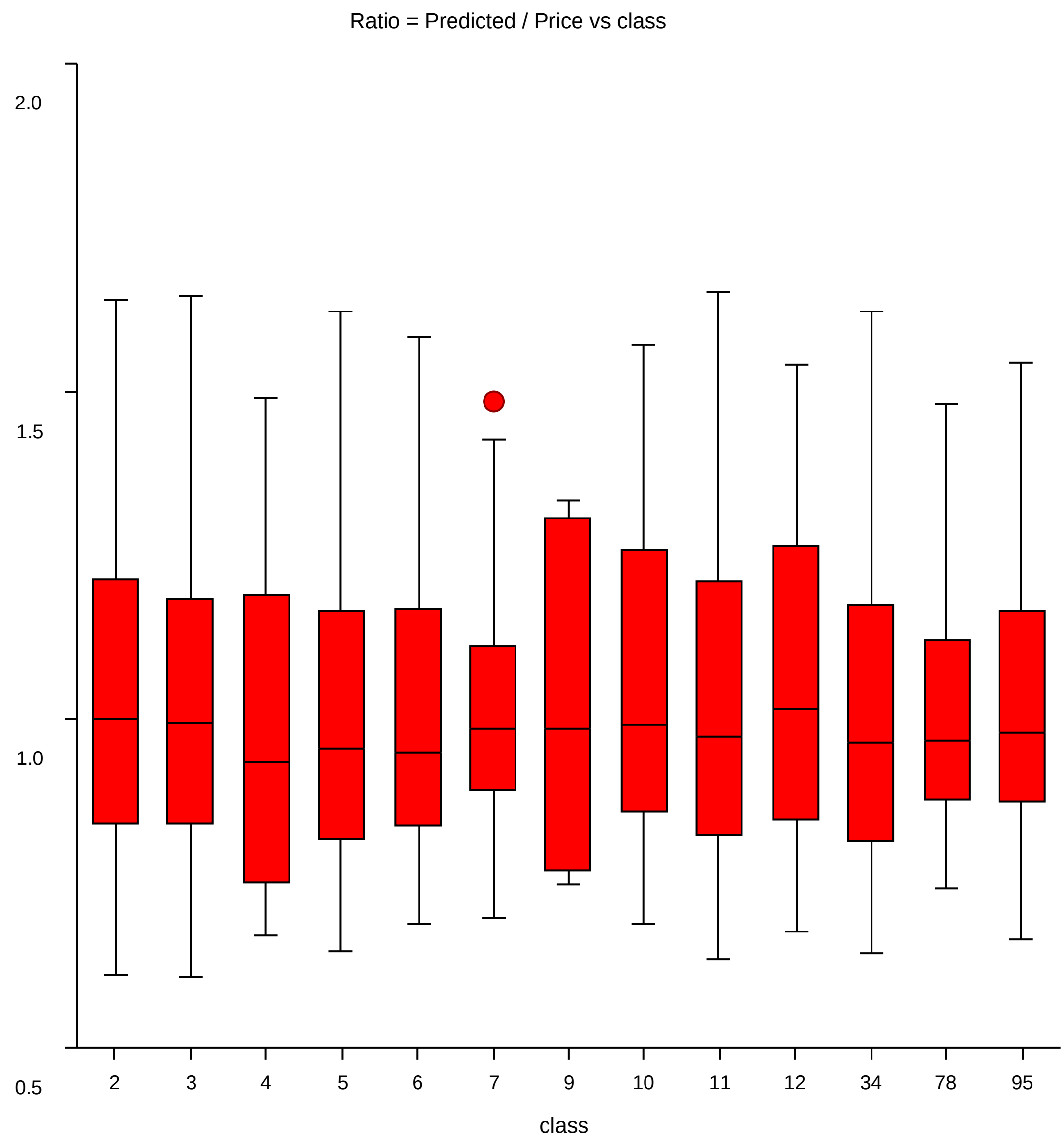
nghcde	Count	Median	Mean	WtMean	IQR	StDev	COD	COV	PRD	PRB
10	216	0.963	1.018	0.980	0.256	0.214	17.192	21.040	1.039	-0.146
20	174	0.977	1.023	0.977	0.325	0.223	18.238	21.763	1.047	-0.395
30	131	0.972	1.025	0.975	0.400	0.251	20.931	24.526	1.052	-0.398
70	88	0.953	1.027	0.984	0.383	0.244	21.111	23.703	1.044	-0.159
80	227	0.918	0.953	0.906	0.343	0.231	20.735	24.255	1.052	-0.315
83	39	0.860	0.938	0.891	0.386	0.226	21.083	24.104	1.053	-0.357
91	35	0.914	0.946	0.908	0.282	0.208	17.269	21.956	1.042	-0.429
100	56	1.023	1.064	1.014	0.411	0.246	20.117	23.155	1.049	-0.233
101	10	1.074	1.087	1.056	0.271	0.246	15.420	22.589	1.030	-0.016
111	187	0.995	1.045	0.981	0.446	0.262	22.364	25.110	1.065	-0.439
120	128	1.009	1.040	0.982	0.464	0.275	23.300	26.477	1.059	-0.279
121	14	1.068	1.037	0.986	0.427	0.239	18.655	23.020	1.052	-0.380
130	129	0.951	1.008	0.948	0.386	0.259	22.596	25.694	1.063	-0.462
140	4	1.054	1.041	0.998	0.372	0.193	14.370	18.537	1.044	-0.433
150	26	0.981	1.026	0.976	0.379	0.243	19.202	23.669	1.051	-0.494
151	25	0.996	0.988	0.971	0.229	0.152	11.834	15.437	1.017	-0.219
170	8	1.158	1.127	1.072	0.443	0.255	18.132	22.594	1.051	-0.944
180	29	1.131	1.135	1.107	0.296	0.222	15.249	19.536	1.026	-0.181
210	10	1.140	1.172	1.154	0.396	0.244	16.395	20.809	1.016	0.106
220	75	1.014	1.032	0.975	0.459	0.260	21.780	25.218	1.059	-0.577
230	44	0.956	1.025	0.985	0.378	0.237	19.340	23.090	1.041	-0.347
240	249	1.052	1.087	1.052	0.324	0.213	16.565	19.564	1.033	-0.177
241	45	0.933	0.936	0.923	0.158	0.102	8.965	10.939	1.014	-0.138
250	8	0.896	0.972	0.928	0.472	0.239	21.151	24.595	1.047	-0.666
260	1	0.780	0.780	0.780	0.000		0.000		1.000	0.000
280	199	1.009	1.035	0.991	0.317	0.216	17.148	20.898	1.044	-0.380
Total	2157	0.983	1.025	0.975	0.347	0.232	19.643	22.648	1.051	-0.079



Box Plot of Ratio
2.0
1.5
1.0
0.5
Ratio
10 20 30 70 80 83 91 100 101 111 120 121 130 140 150 151 170 180 210 220 230 240 241 250 260 280
nghcde

Model Performance by Class

Class	Count	Median	Mean	WtMean	IQR	StDev	COD	COV	PRD	PRB
2	208	1.002	1.028	0.981	0.373	0.230	18.990	22.364	1.048	-0.482
3	627	0.994	1.028	0.976	0.343	0.243	19.915	23.617	1.053	-0.382
4	47	0.934	0.984	0.933	0.437	0.244	21.962	24.740	1.055	-0.317
5	149	0.957	1.014	0.956	0.348	0.242	20.855	23.910	1.061	-0.140
6	82	0.951	1.025	0.965	0.331	0.237	19.437	23.141	1.062	-0.142
7	68	0.987	1.021	0.974	0.219	0.181	14.037	17.721	1.048	-0.095
9	5	0.986	1.028	0.984	0.536	0.270	21.734	26.245	1.046	-0.071
10	112	0.991	1.055	1.002	0.399	0.245	20.480	23.199	1.052	-0.036
11	542	0.975	1.026	0.973	0.388	0.249	20.996	24.245	1.055	-0.159
12	18	1.017	1.056	0.994	0.417	0.262	21.593	24.817	1.062	-0.293
34	86	0.966	1.008	0.946	0.360	0.232	19.682	23.021	1.065	-0.346
78	74	0.968	1.012	0.977	0.243	0.178	14.465	17.616	1.036	-0.073
95	139	0.981	1.020	0.986	0.291	0.200	16.459	19.581	1.035	-0.058
Total	2157	0.983	1.025	0.975	0.347	0.236	19.643	23.017	1.051	-0.079



Ratios by Neighborhood and Class

It is customary for CCAO to analyze new revaluation results in relation to previous values. Sharp increases or decreases by neighborhood and class are examined for reasonableness. Such was the case in Hyde Park. Initial values submitted to CCAO raised concerns about specific areas with unexpected increases or decreases. It turned out that models with similar summary performance statistics could have different change percentages at the neighborhood/class level.

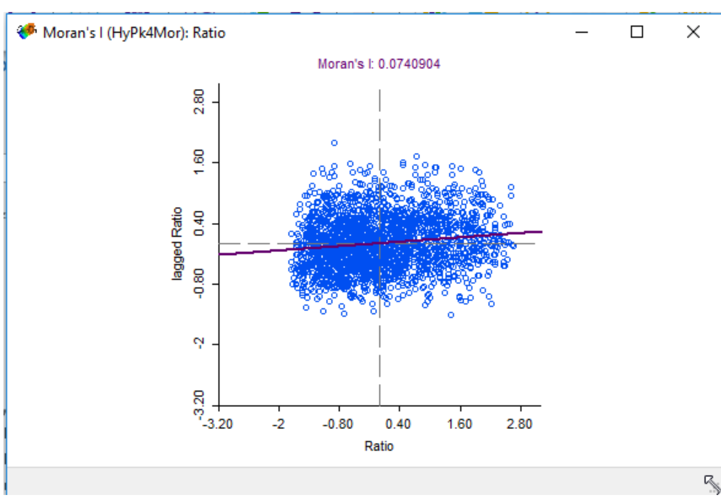
Nearly two weeks were invested in fine tuning the model structure such that increases and decreases in median value were more in line with expectations. A neighborhood/Class matrix of change was developed so that one could gain an overview of the value changes. The final modeled values are represented in the next image. The weighted average ratio of 2018 to 2015 values is presented first, followed by the count involved in that cell's computation. This way one can easily distinguish low count cells from other that are more representative. Conditional formatting was used to highlight low and high values. For example, it is easy to spot Class 9 as a group of properties needing appraisal review. This is in large part due the fact that there were only five useable sales over the five-year period of sales used in calibrating the models. Of these, two were in neighborhood 10, two in 20 and one in 180.

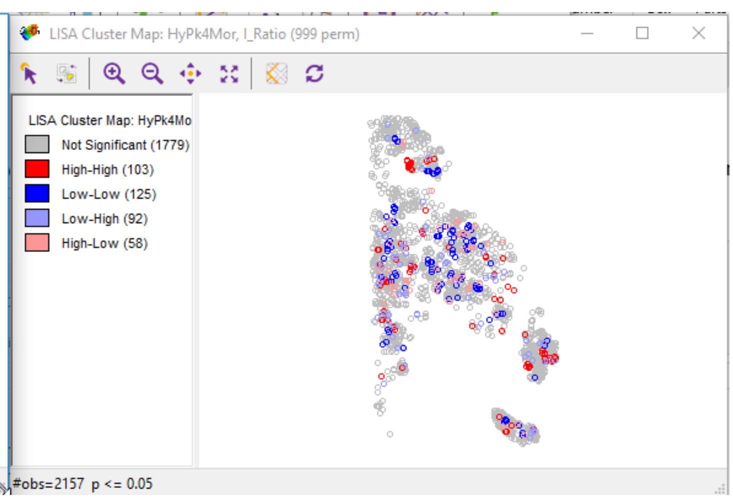
MBHD/Class	2		3		4		5		6		7		8		9		10		11		12		34		78		95		Total	
10	1.83	7	1.60	27	1.65	11	1.44	153	1.22	313	1.19	175	1.06	7	0.59	26	2.03	639	1.17	1,506	1.20	31	1.48	2	1.17	162	1.39	248	**1.31**	**3,307**
20	0.89	17	1.30	39	1.22	13	1.29	127	0.88	349	1.11	64	0.91	8	0.37	104	1.12	334	1.13	224	1.03	5	0.78	1	0.93	60	1.16	691	**0.98**	**2,036**
30	1.33	428	1.29	1,135	1.24	77	1.21	421	1.33	56	1.07	29		0		0	1.31	96	0.89	2,469	0.99	76	1.17	61	1.07	6	1.38	37	**1.03**	**4,891**
70	1.57	251	1.48	328	1.11	20	1.53	262	1.32	84	1.45	234	0.76	1		0	1.53	84	1.07	1,509	1.04	35	1.13	79	1.10	39	0.98	176	**1.19**	**3,102**
80	1.18	781	0.97	3,718	0.99	379	0.97	936	0.94	150	0.94	88		0	0.31	2	1.30	299	0.92	1,796	0.86	57	0.95	121	0.88	21	1.28	369	**0.98**	**8,717**
83	1.17	7	1.06	118	0.96	68	0.90	125	1.00	130	0.92	4		0		0	1.14	9	0.98	376	1.08	5	1.08	3	0.78	1	0.92	3	**0.98**	**849**
91	1.06	85	0.92	158	0.88	2	0.81	7		0	0.83	46		0		0		0		0		0	1.00	130	0.79	36	0.81	61	**0.91**	**525**
100	0.93	666	1.00	1,194	0.95	81	0.93	285	0.86	28	1.09	15		0	0.32	1	0.94	23	0.72	1,767	0.93	104	1.04	51	0.81	11	1.04	49	**0.84**	**4,275**
101	0.91	165	0.86	201	0.83	12	0.80	23	0.73	6	0.46	1		0		0	0.82	1	0.67	394	0.74	43	0.77	1		0		0	**0.75**	**847**
111	0.98	553	0.92	1,933	0.94	148	0.85	519	0.76	50	0.77	203	0.59	3	0.34	3	1.31	375	0.72	343	0.68	21	0.86	437	0.77	76	0.68	95	**0.88**	**4,759**
120	1.06	463	1.06	973	0.98	68	1.16	236	0.99	40	0.88	40		0		0	1.26	166	0.91	1,180	0.85	45	1.04	122	0.91	16	1.19	61	**1.00**	**3,410**
121	1.24	292	1.15	348	1.02	17	1.03	64	0.79	3	0.82	40		0		0		0	0.79	179	0.73	18	1.02	54	0.77	19	1.07	80	**1.02**	**1,114**
130	1.09	751	1.05	2,600	1.00	52	1.05	436	0.85	25	0.89	26		0		0	1.20	17	0.80	272	0.73	28	0.98	183	0.85	15	1.18	5	**1.02**	**4,410**
140	1.26	118	1.35	68	0.94	6	1.02	16	0.99	4	0.91	30		0		0	1.15	1	0.93	531	0.84	57	0.99	9	0.80	2		0	**0.96**	**842**
150		0	1.08	11	1.13	23	1.02	26	0.98	205		0		0	0.37	7		0	1.11	72	1.91	1	0.92	4		0		0	**0.99**	**349**
151	1.07	3	0.91	63	0.82	14		0	0.72	1	0.77	1	0.74	3	0.44	1		0	0.67	24		0	0.84	169	0.69	14		0	**0.82**	**293**
170	1.14	260	1.15	408	0.99	4	0.96	149	1.03	1	0.97	1		0		0	1.53	12	0.82	10	0.82	4	1.05	16	1.05	1		0	**1.10**	**866**
180	1.11	641	1.09	856	0.95	22	1.07	386	0.98	12	0.80	11		0		0	1.32	1532	0.80	225	0.85	32	0.92	74	0.89	2	1.00	93	**1.13**	**3,886**
210	1.10	531	1.08	573	1.11	37	1.43	113	0.95	26	0.91	32		0		0		0	0.74	727	1.03	127	1.04	23	0.67	3	1.09	2	**0.94**	**2,194**
220	1.04	445	0.98	918	0.90	36	1.01	861	0.91	5	0.96	14		0		0	0.90	34	0.69	373	0.78	10	0.95	64	0.87	4	0.84	6	**0.93**	**2,770**
230	1.23	46	1.09	106	1.12	1	0.93	10	1.11	1	0.87	3		0		0	1.04	726	0.82	213	0.73	9	1.03	16		0	1.10	7	**0.99**	**1,138**
240	1.00	1,383	0.98	2,110	0.95	88	1.02	249	0.72	6	1.01	17	0.79	1		0	1.27	130	0.72	436	0.66	47	0.81	152	0.75	79	0.98	2	**0.94**	**4,700**
241	1.08	7	0.99	301	1.04	11		0		0	0.96	4		0	0.36	1		0		0		0	1.04	57	0.88	96		0	**0.97**	**477**
250	1.06	123	0.87	309	0.86	6	1.03	15		0	0.85	1		0		0		0	0.69	19	0.66	5	0.79	55	0.85	4	1.31	5	**0.89**	**542**
260	1.69	220	1.48	64	1.57	7	1.28	7	1.19	1	1.14	2		0		0		0	0.95	17	1.25	4	1.58	66		0	1.46	113	**1.53**	**501**
280	1.03	701	1.00	1,312	1.03	25	0.92	67	0.91	6	0.86	75		0		0	0.50	1	0.79	298	0.77	53	0.78	163	0.88	39		0	**0.95**	**2,740**
TOTAL	**1.09**	**8,944**	**1.02**	**19,871**	**1.01**	**1,228**	**1.06**	**5,493**	**0.98**	**1,502**	**1.04**	**1,156**	**0.91**	**23**	**0.39**	**145**	**1.34**	**4,479**	**0.92**	**14,960**	**0.89**	**817**	**0.93**	**2,113**	**0.95**	**706**	**1.16**	**2,103**	**1.01**	**63,540**

Spatial Dependency

A means to verify the locational stability of the estimates is provided computing Local Indicators of Spatial Association often referred to as LISA. Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. In mass appraisal it is customary to look for spatial clusters in the ration of appraised value to sale price. The plot below has as its X Axis the z-transform of Ratio, defined as $z = (x - \mu)/\sigma$ where x is the individual ratio, μ is the mean ratio and σ is the standard deviation of the ratios in the sample. The Y axis is the average of the five nearest transformed ratios not including the ratio of the X Axis. The fact that there very little slope to the plot is a good indication that there are no spatial clusters of high or low ratios.

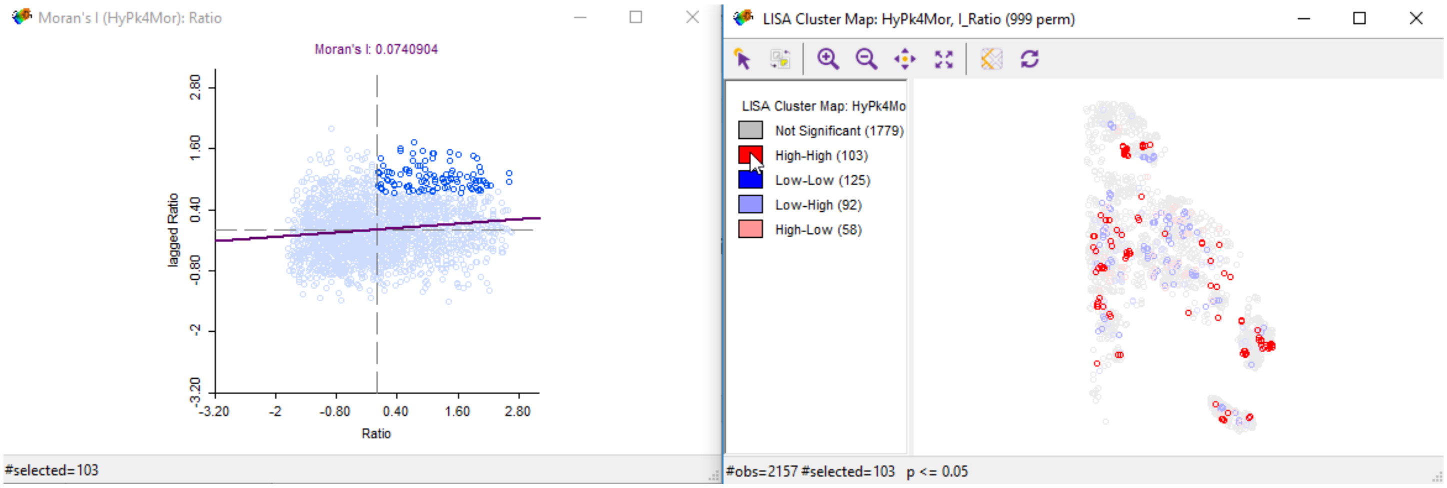
The Moran's I global statistic of 0.0740904 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of Hyde Park, the statistic is very close to 0.0.





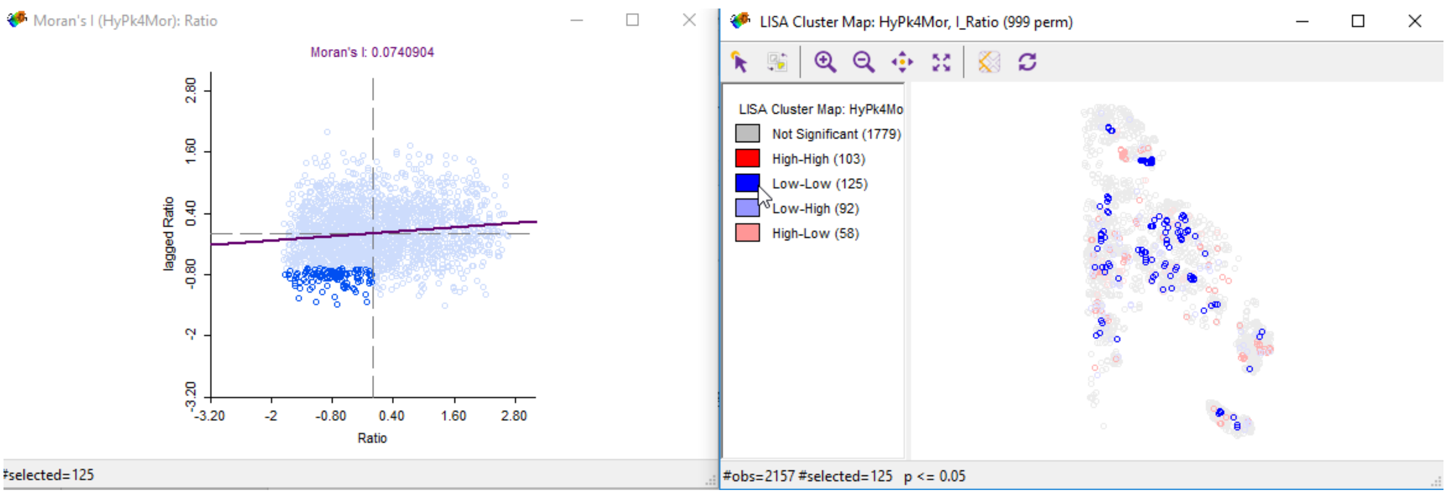
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



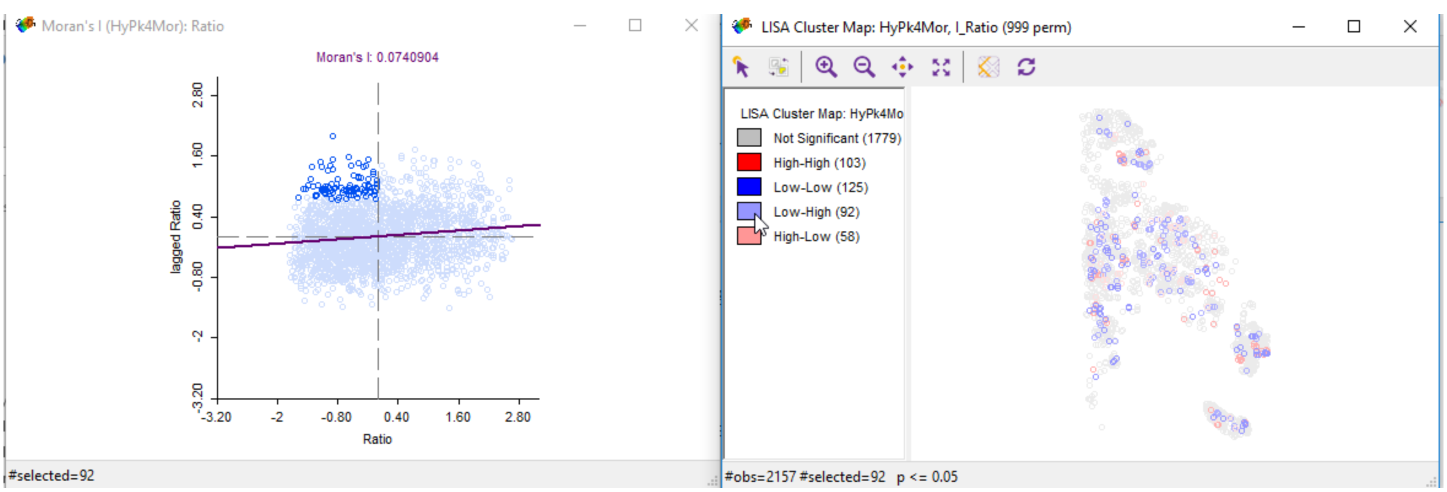
Low near Low Ratios

The low near low ratios are also spread uniformly around the town.



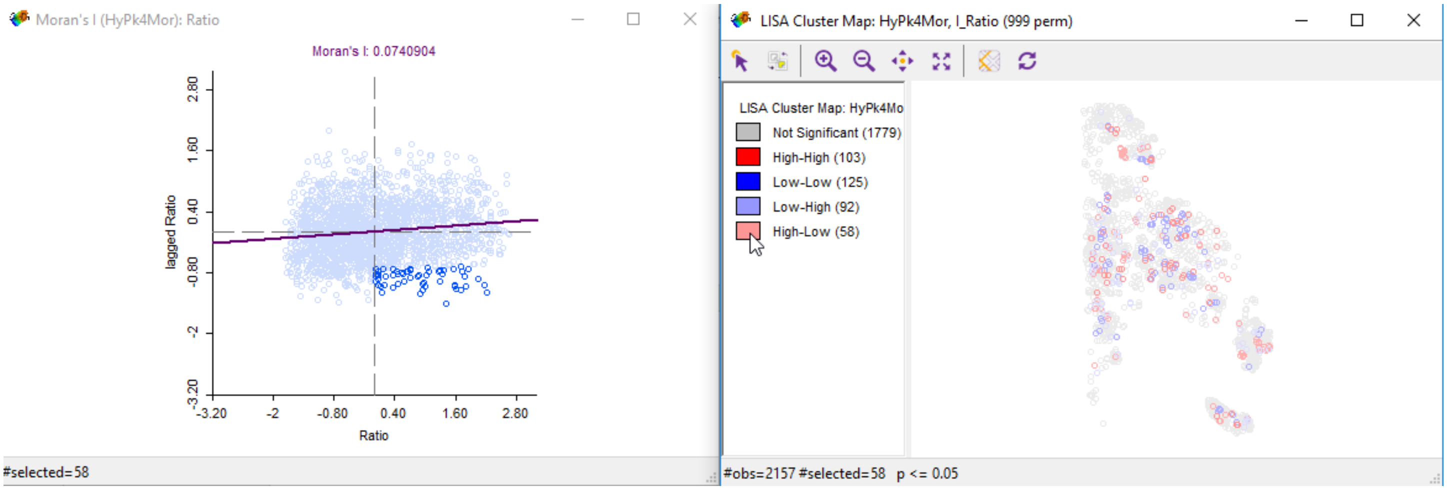
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.



Moran's I (HyPk4Mor): Ratio
Moran's I: 0.0740904
lagged Ratio
Ratio
#selected=58
LISA Cluster Map: HyPk4Mor, I_Ratio (999 perm)
LISA Cluster Map: HyPk4Mo
Not Significant (1779)
High-High (103)
Low-Low (125)
Low-High (92)
High-Low (58)
#obs=2157 #selected=58 p <= 0.05

Jefferson

Summary

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- Owner Occupancy data was considered, but did not prove to be statistically significant
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Jefferson was valued using the Multiple Regressions Analysis direct market comparison method of valuation
- Log linear models introduce what is called a retransformation bias
- The bias is corrected to ensure that the weighted mean ratio of estimated value to sale price is 1.000

The Data

Sales Counts

The rules used by CCAO to filter the sales were applied to the sales data. The rules are:

*select if (amount1>100000).
*select if (amount1<990000).
*select if (multi<1).
*select if sqftb<9000
*select if (year1>2012).
*select if puremarket=1
This yields 14,151 as the starting count of sales used in the analysis.

Data Fields

The initial list of data items available for analysis is provided in Appendix A Variable Definitions. Certain additional data fields were created. Those that were relevant to Jefferson are:

Location Factor

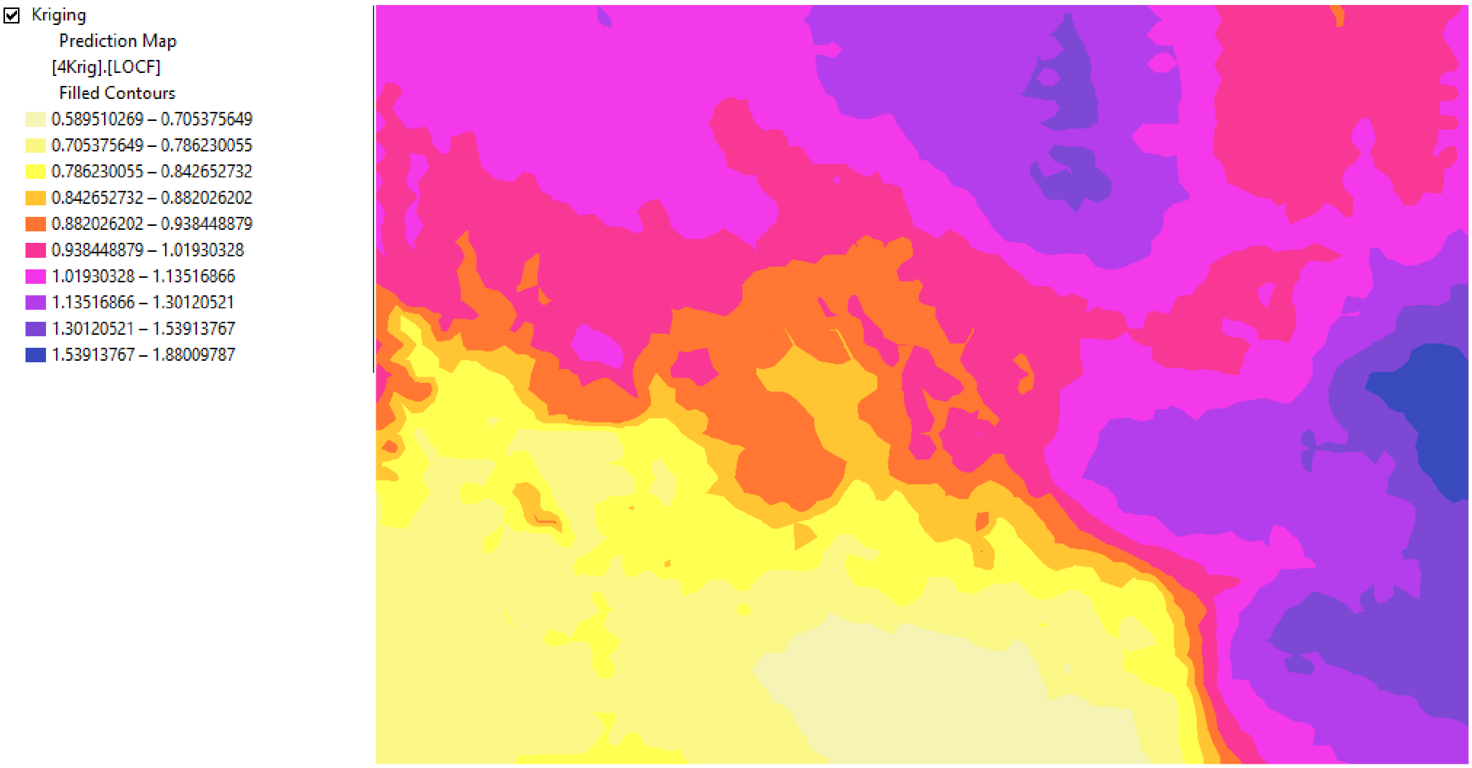
A location factor was derived by use of Geographically Weighted Regression (GWR). Created a simple multiplicative model and did a very light outlier detection using the previously described NRATIO.

Variable	IQR	25th Pctile	75th Pctile	LL	UL
NRATIO	0.338446	0.819351	1.157797	-0.19599	2.173135

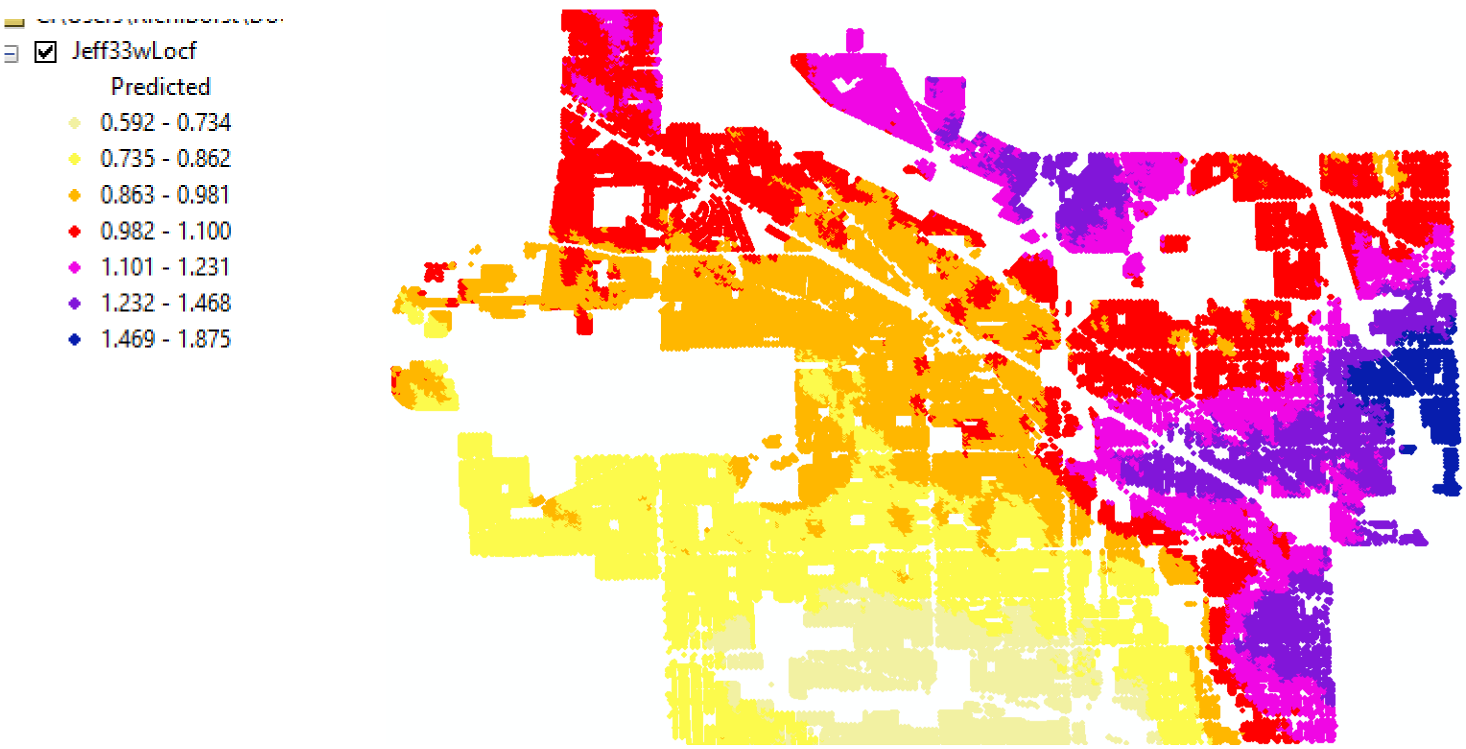
This was based on a 3.0 factor on the IQR. It resulted in identifying 205 outliers. Created a file for GWR and removed the outliers.

The rest of the process was as described in Appendix D Outlier Detection.

The resultant surface and thematic legend are shown in the image below.

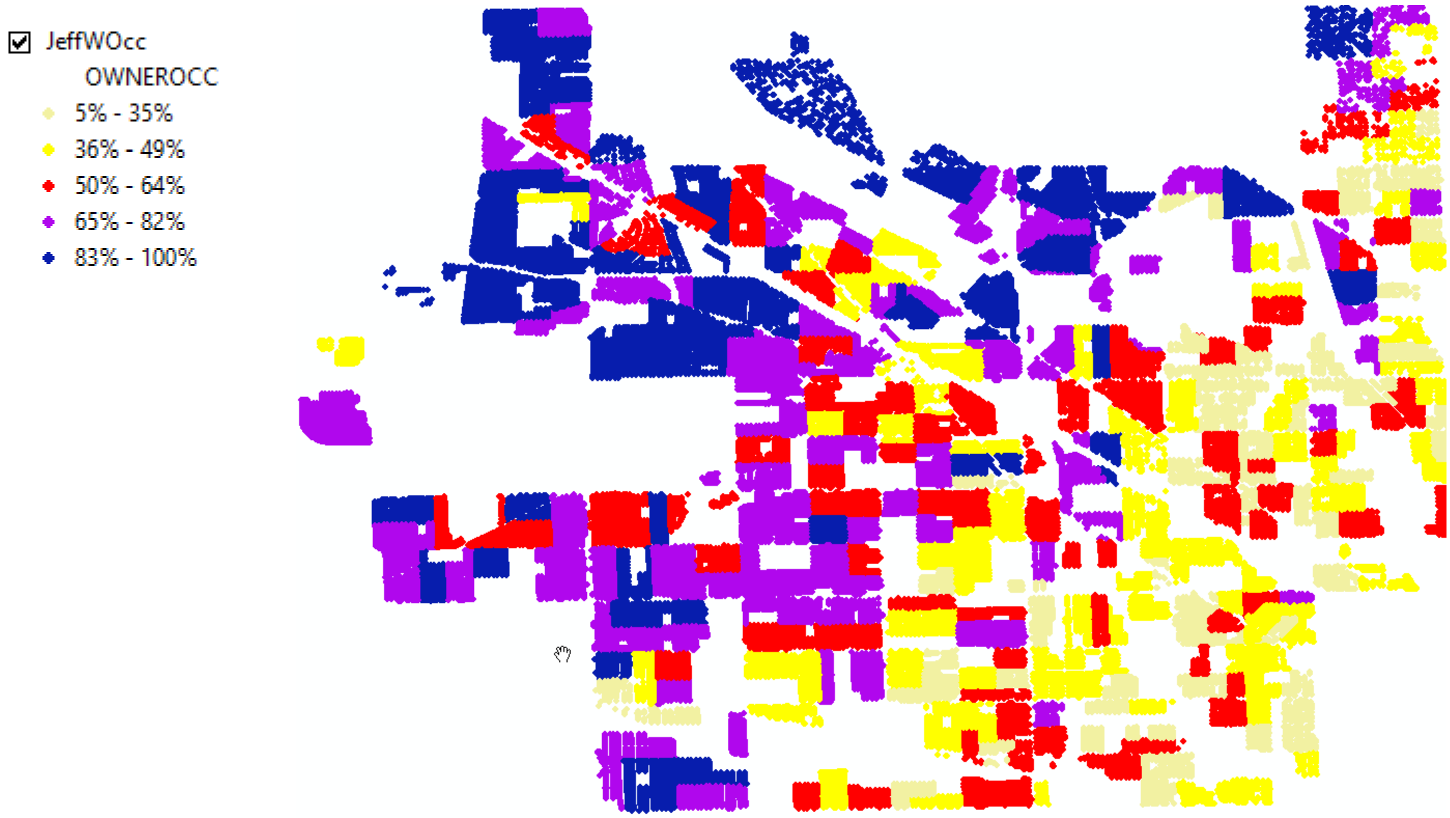


When applied to all properties the thematic map of Location Factor is given in the next image.



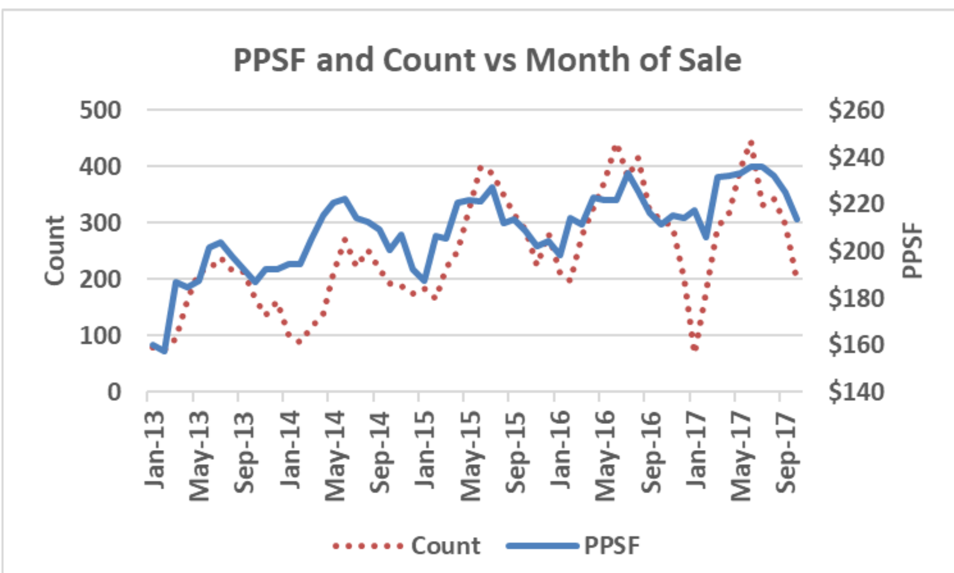
Owner Occupancy

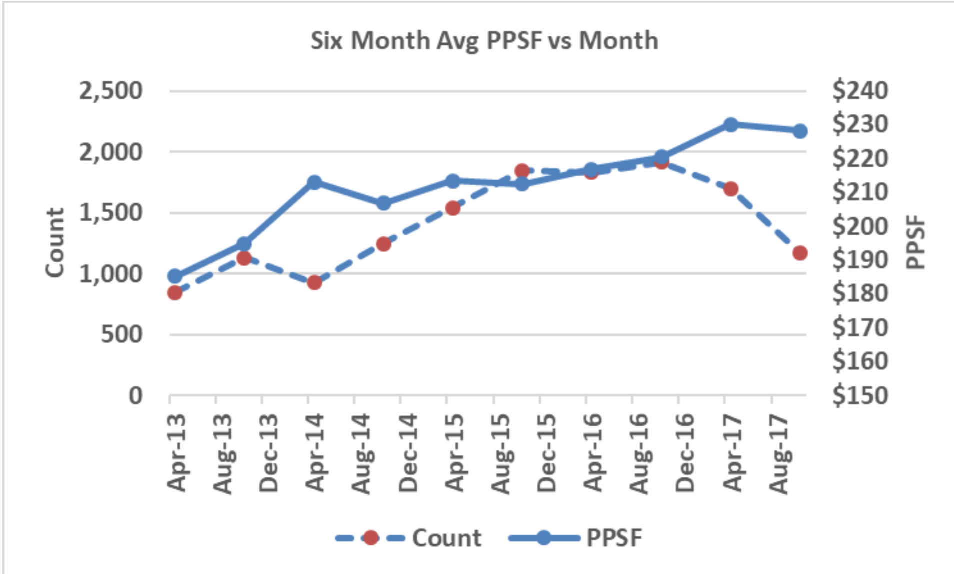
The process for obtaining the Owner Occupancy variable is described in Appendix C Owner Occupancy. The resultant variable is shown below.



Reverse Half of Sale

The sales used in the analysis span a period of five years. To allow for time trending the sales to the valuation date, first, a reverse month of sale is computed. If the sale took place in December of 2017, the reverse month of sale (RMOS) is 1, November 2017, RMOS is 2, all the way back to January of 2013, RMOS is 60. In terms of using this variable directly in the model to be discussed, it is converted into a Reverse Half Year of Sale (RHOS). The rationale for this approach can be seen in the following two charts. The first shows Price per Square Foot (PPSF) and Count vs. Month of Sale. The second chart shows the same two variables vs Six Month Average of PPSF. The RMOS variable is too granular and "noisy" as compared to the RHOS variable.



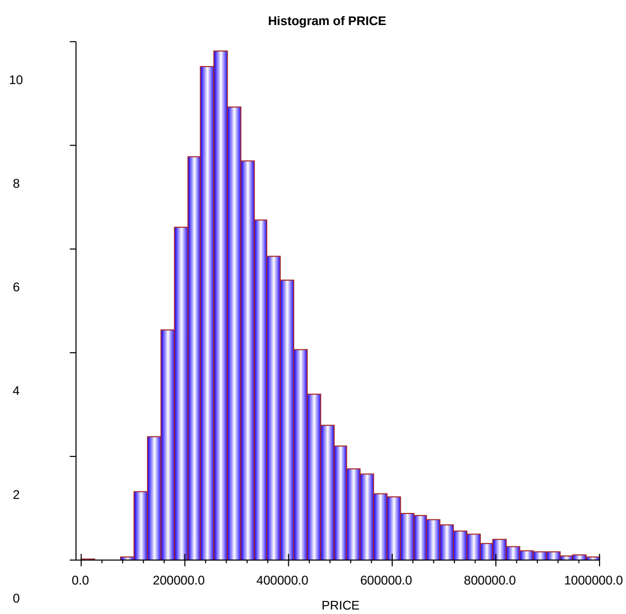


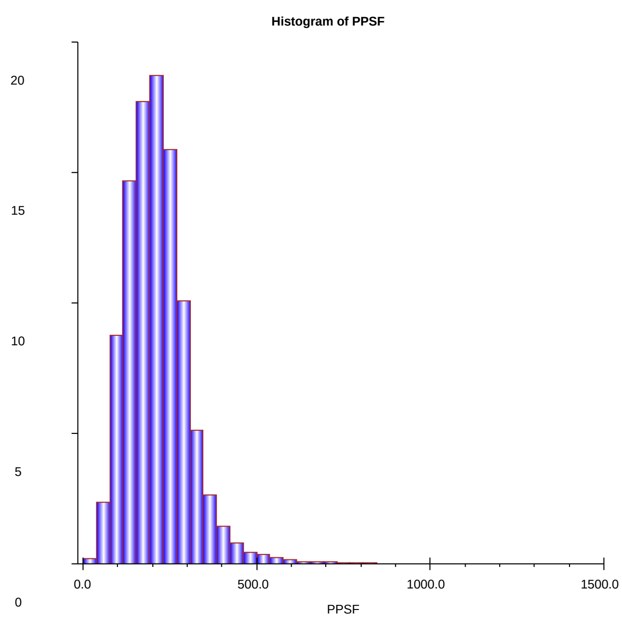
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.

Price and Price per Square foot

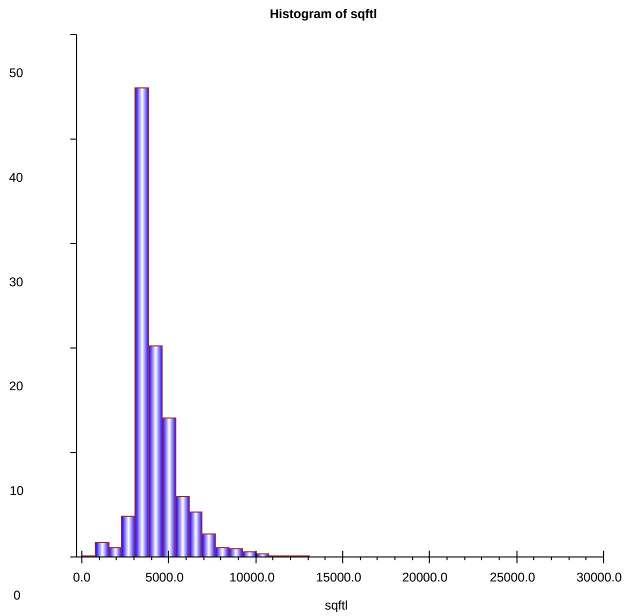
The price histogram fits the range of prices specified at the outset of modeling, namely a range of $100,000-$990,000. The price per square foot range indicates what are likely to be outlier situations. In other words, $600-$700 per square foot and above is not likely to represent a true open market situation.

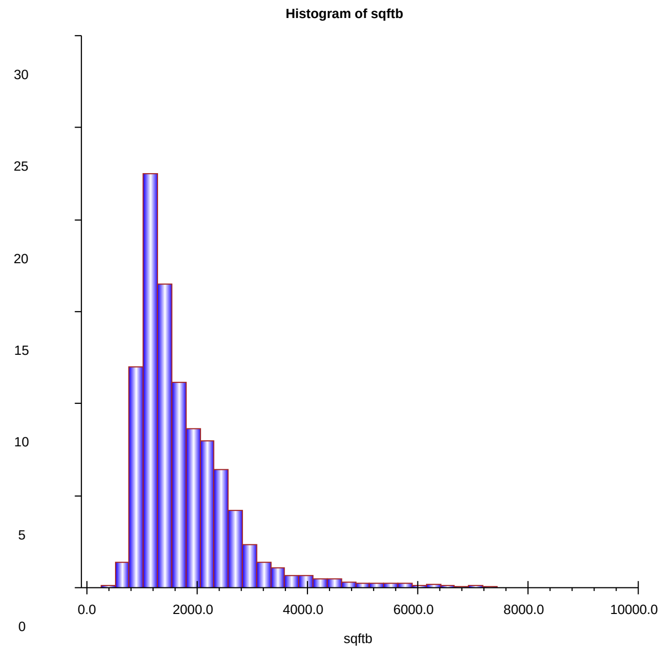




Square foot Land and Building

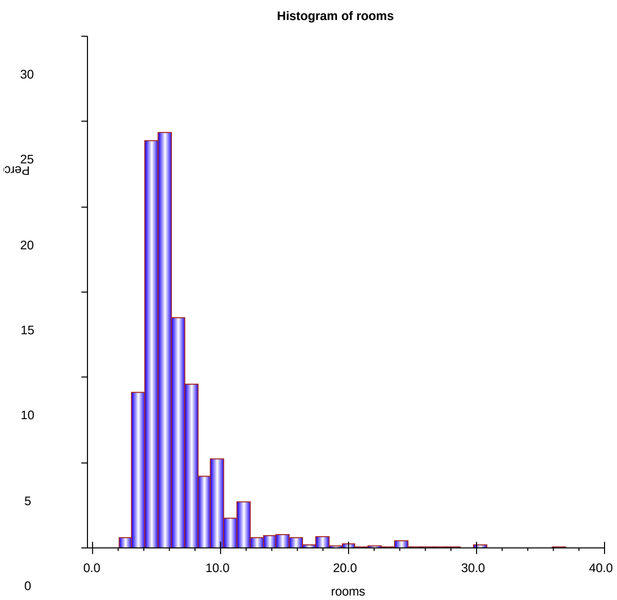
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not. It

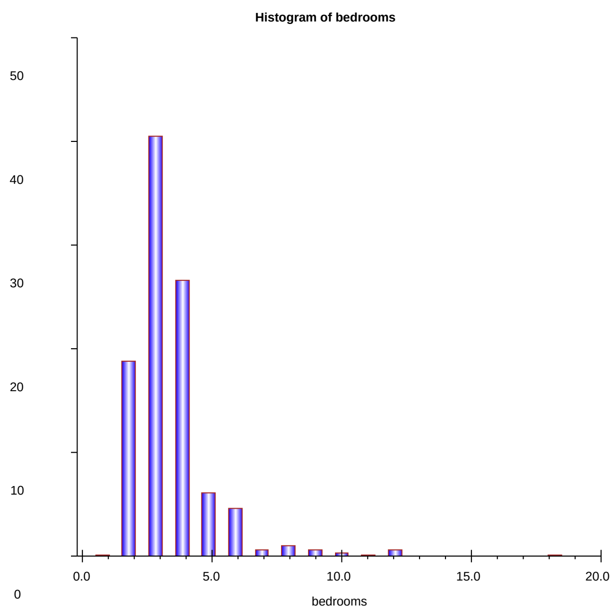




Rooms and Bedrooms

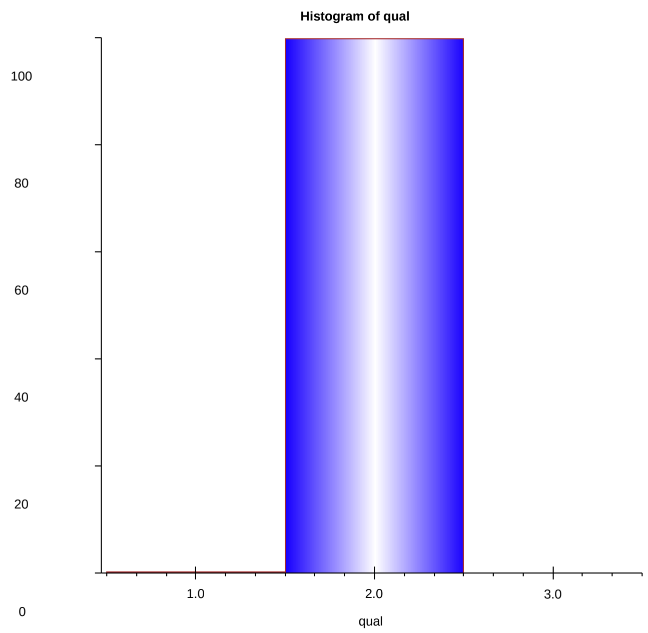
Looks like there are homes with 36 rooms and some with 18 bedrooms. For the specific case of 36/18, it looks as if they are six-unit apartment buildings with each unit having 6 rooms and 3 bedrooms.

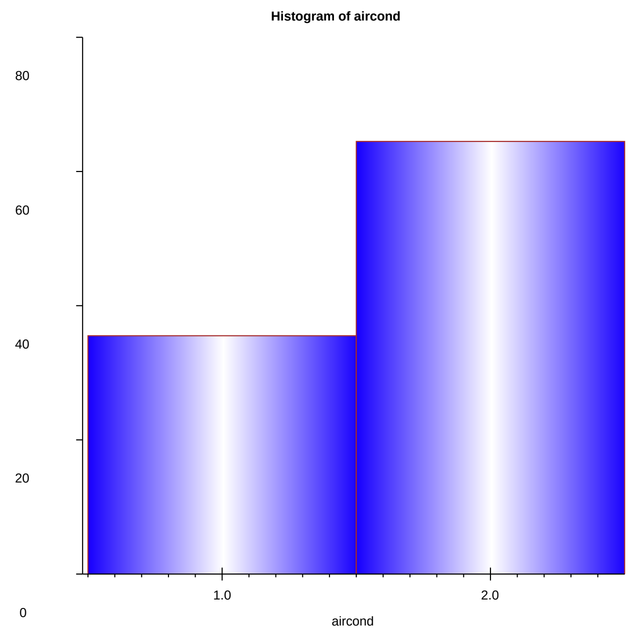




Quality and Air Conditioning

The quality of construction variable has little variability and almost certainly will not be a useful variable. On the other hand, air conditioning may well be useful in a model.





Model Structure and Calibration

Outlier Detection

Outliers were detected according to the process outlined in Appendix D Outlier Detection..

Structure

Once again, two model structures were evaluated, additive and multiplicative. The multiplicative model form is often referred to as a log-linear model. This confuses the model structure with the calibration process. It turned out that for this dataset, the multiplicative form of the model was chosen for further processing. The Horizontal Equity performance statistics of the multiplicative model are superior to the additive model. The Vertical Equity statistics are marginally better for the additive model, but not enough to decide in its favor. As will be seen later in this report, the actual model used was a segmented model with the very high-end properties separated from the rest.

Model	Count	Median	WgtMean	COD	PRD	PRB
Multiplicative	11,854	0.996	0.987	13.534	1.027	-0.087
Additive	11,854	1.010	1.000	14.383	1.028	-0.068

Retransformation Bias

The topic of Retransformation Bias is discussed in Appendix E Retransformation Bias. The approach used here is to correct the value estimates by the inverse of the weighted mean prediction. In the final model results the actual correction was Predicted/0.9743.

Location Factor and Owner Occupancy

In Jefferson, the owner occupancy variable was not significant, but the location factor variable entered the model with a strong significance.

The final MRA Model

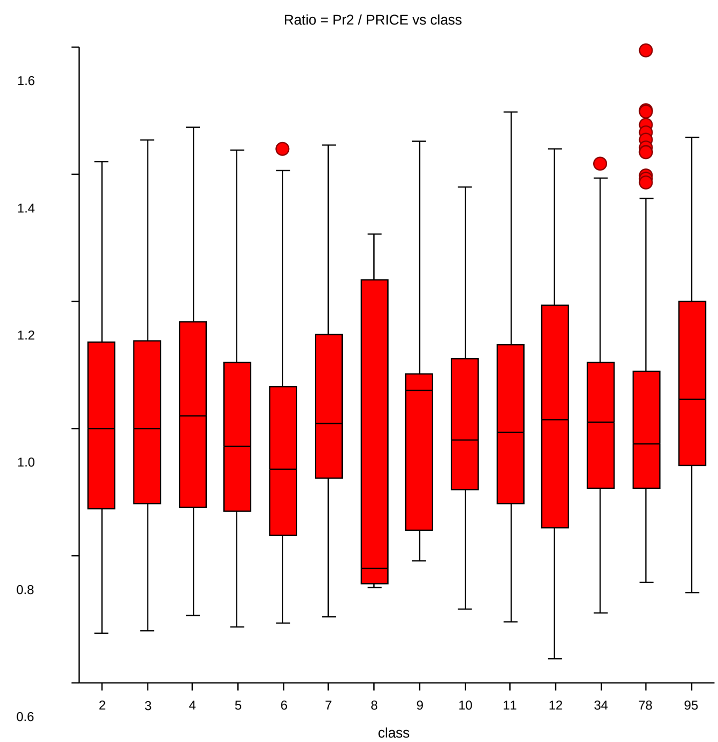
Variable	Coefficient	T Value
Intercept	8.034	144.042
lnSFL	0.195	35.655
lnSFB	0.412	57.815
lnFIXT	0.193	14.478
lnAGE	-0.047	-14.916
lnFIREPL	0.047	7.446
lnLocF	0.972	108.737
(SA_12=1)	-0.020	-2.350
(SA_23=1)	0.048	4.761
(SA_26=1)	0.057	3.774
(NB_70=1)	-0.086	-7.171
(NB_81=1)	-0.078	-7.885
(NB_120=1)	-0.051	-5.265
(NB_140=1)	0.070	5.191
(NB_580=1)	0.208	3.691
(CL_10=1)	-0.168	-8.786
(CL_11=1)	-0.157	-25.760
(CL_78=1)	0.068	6.416
(rs5=1)	-0.051	-12.099
(num5=1)	0.080	4.551
(rf3=1)	0.096	3.158
(rf5=1)	0.108	5.445
(bsfn2=1)	-0.098	-9.450
(AC_2=1)	-0.013	-3.583
(comm1=1)	-0.426	-26.239
(comm2=1)	-0.523	-13.500
(comm3=1)	-0.650	-5.434
(gar3=1)	0.023	7.355
(gar8=1)	0.295	6.506
(renov1=1)	0.362	12.627

Model Performance Stats by NBHD

NGHCDE	Count	Median	WgtMean	COD	PRD	PRB
10	376	0.999	0.994	13.142	1.025	-0.178
21	118	1.013	0.998	13.729	1.032	-0.144
22	147	1.018	0.985	13.049	1.033	-0.248
30	236	1.033	1.014	12.694	1.031	-0.193
41	221	1.060	1.046	13.112	1.022	-0.164
42	130	1.038	1.023	12.814	1.027	-0.191
50	233	1.024	1.019	13.564	1.020	-0.077
60	186	0.975	0.977	14.183	1.034	-0.276
70	685	0.990	0.983	16.058	1.035	-0.170
71	323	1.015	0.993	13.470	1.026	-0.179
74	213	0.990	0.962	13.800	1.034	-0.178
81	311	0.996	0.987	15.226	1.029	-0.116
82	551	1.000	0.991	15.130	1.033	-0.115
90	424	0.983	0.970	14.011	1.028	-0.262
101	233	0.976	0.974	14.814	1.032	-0.214
110	387	0.947	0.949	12.822	1.019	-0.094
120	325	0.983	0.981	14.233	1.034	-0.185
140	163	1.013	0.993	14.270	1.023	-0.188
150	1,048	0.991	0.990	14.723	1.026	-0.232
171	189	1.005	0.996	14.278	1.031	-0.325
180	332	1.047	1.012	15.303	1.037	-0.103
200	697	0.994	0.988	14.933	1.027	-0.228
210	124	1.015	1.016	14.703	1.026	-0.220
250	122	1.002	0.977	14.735	1.032	-0.407
260	299	0.961	0.959	13.813	1.027	-0.294
270	100	0.952	0.960	16.659	1.034	-0.299
280	80	1.022	0.999	16.999	1.040	-0.222
361	172	1.048	1.033	11.955	1.023	-0.217
362	264	0.967	0.962	12.818	1.029	-0.184
371	648	0.990	0.985	13.932	1.029	-0.231
390	190	0.952	0.965	12.539	1.017	-0.065
402	146	0.945	0.947	13.149	1.029	-0.225
410	448	0.994	0.974	13.980	1.024	-0.168
420	94	0.951	0.954	11.815	1.021	-0.095
430	667	1.000	0.991	12.707	1.025	-0.338
440	530	0.997	0.998	12.585	1.021	-0.125
461	307	0.972	0.972	13.360	1.029	-0.226
463	39	1.043	1.044	8.081	1.012	-0.198
520	49	1.048	1.053	12.715	1.010	0.031
560	81	0.958	0.969	16.359	1.022	-0.097
580	9	0.981	1.001	10.120	1.006	0.047
600	14	1.087	1.090	11.910	1.015	-0.192
Combined	11,911	0.995	0.986	14.157	1.029	-0.089

Model Performance Stats by Class

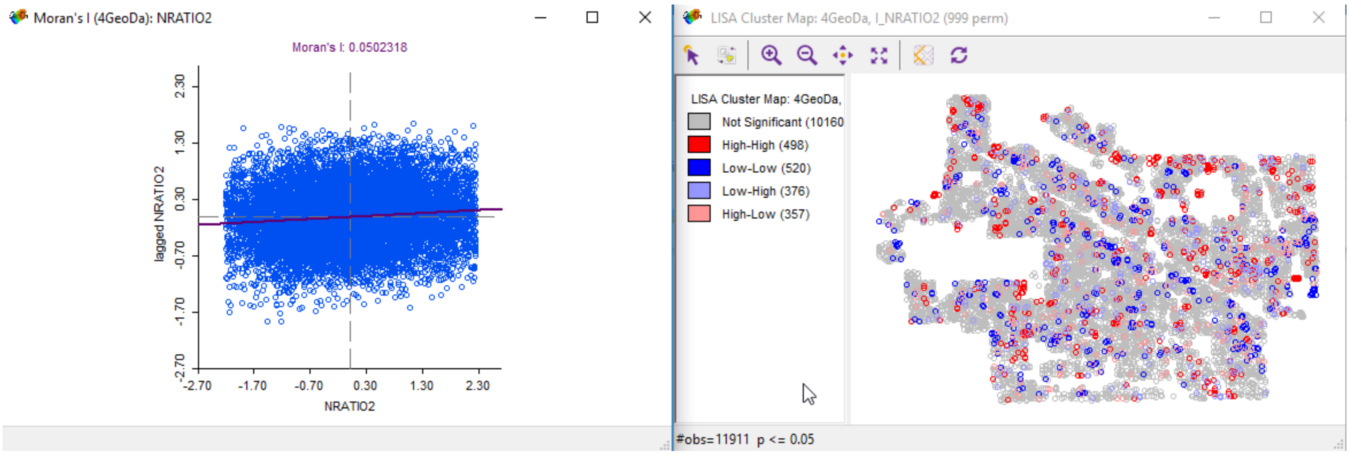
CLASS	Count	Median	WgtMean	COD	PRD	PRB
2	1,122	1.000	0.983	14.643	1.030	-0.225
3	4,792	1.000	0.991	14.266	1.027	-0.203
4	471	1.019	0.993	15.280	1.040	-0.214
5	1,302	0.973	0.965	13.476	1.030	-0.166
6	336	0.935	0.936	14.059	1.030	-0.163
7	197	1.009	1.007	12.423	1.023	-0.187
8	5	0.780	0.873	24.518	1.090	-0.606
9	7	1.060	1.008	14.307	1.022	-0.138
10	84	0.982	0.991	11.956	1.020	-0.222
11	2,514	0.993	0.990	14.294	1.024	-0.078
12	153	1.014	0.984	17.009	1.043	-0.132
34	182	1.010	1.000	11.895	1.015	-0.068
78	474	0.977	0.985	11.546	1.025	-0.154
95	272	1.046	1.054	12.957	1.019	-0.067
Combinec	11,911	0.995	0.986	14.157	1.029	-0.089



Spatial Dependency

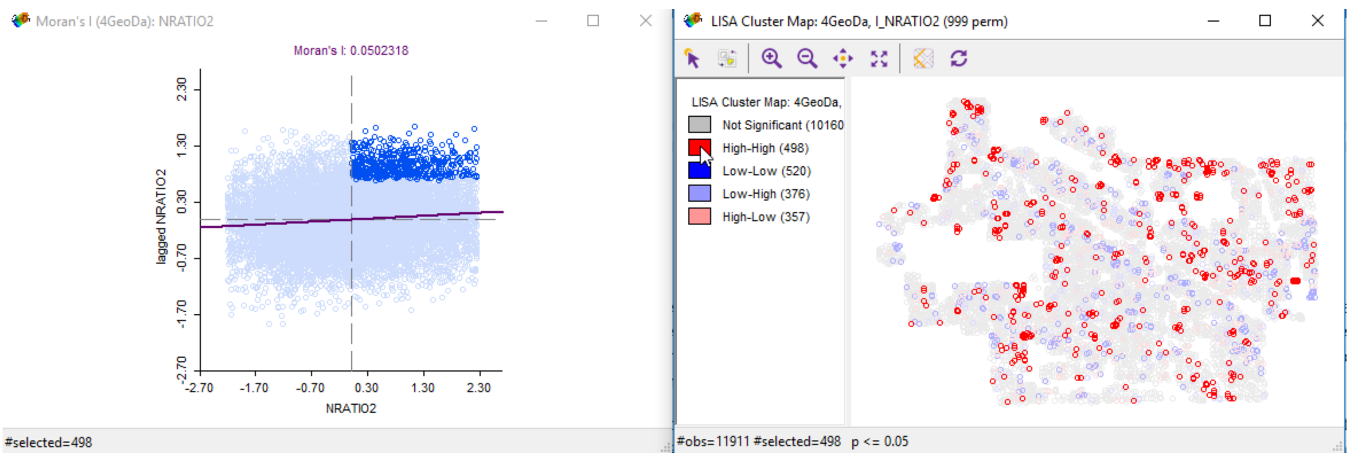
A means to verify the locational stability of the estimates is provided computing Local Indicators of Spatial Association often referred to as LISA. Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. In mass appraisal it is customary to look for spatial clusters in the ration of appraised value to sale price. The plot below has as its X Axis the z-transform of Ratio, defined as $z = (x - \mu)/\sigma$ where x is the individual ratio, μ is the mean ratio and σ is the standard deviation of the ratios in the sample. The Y axis is the average of the five nearest transformed ratios not including the ratio of the X Axis. The fact that there very little slope to the plot is a good indication that there are no spatial clusters of high or low ratios.

The Moran's I global statistic of 0.0502 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of Jefferson, the statistic is very close to 0.0.



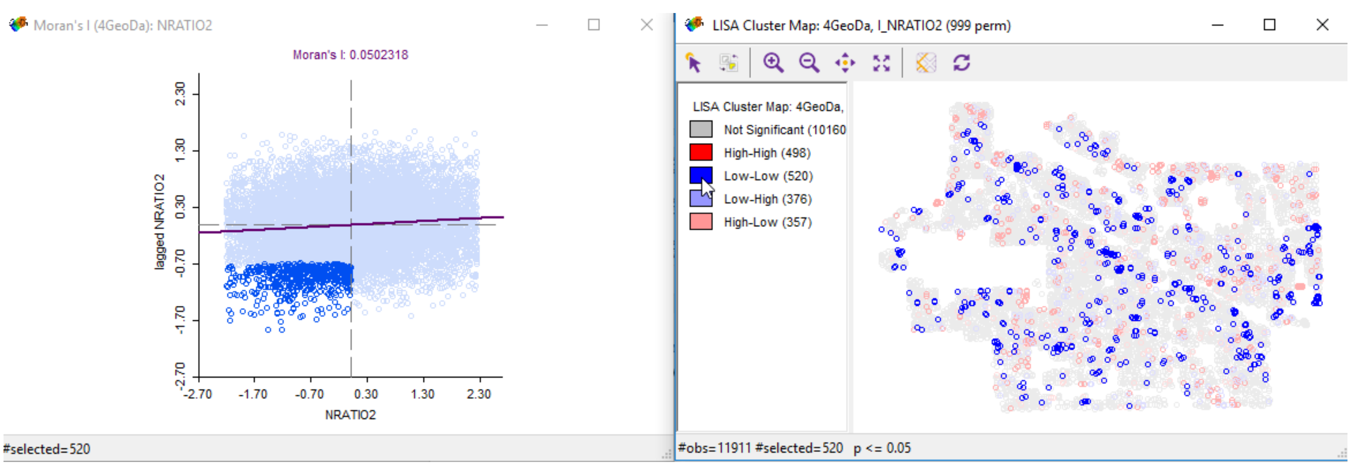
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



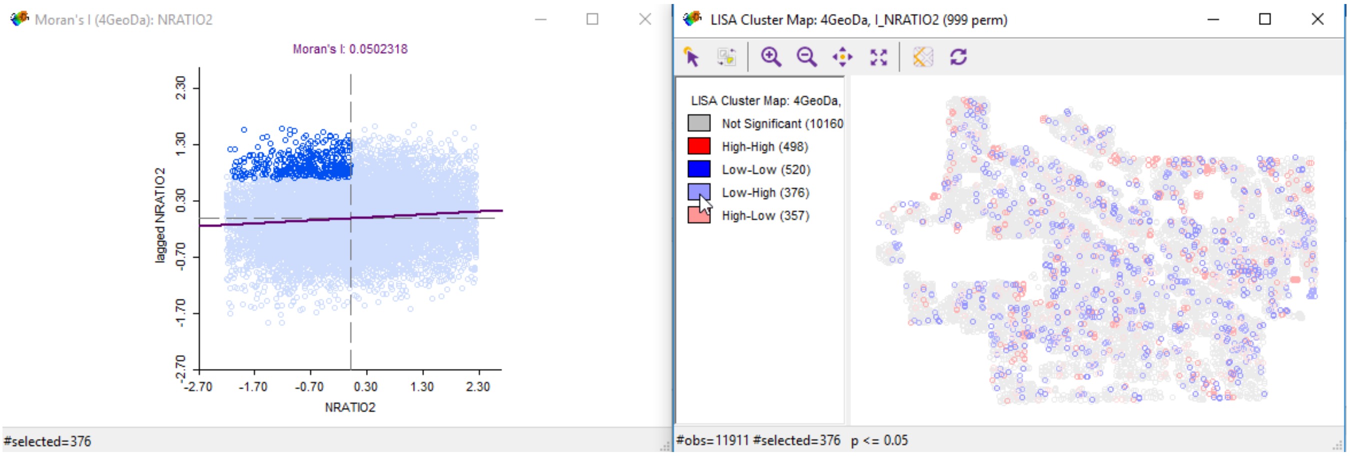
Low near Low Ratios

The low near low ratios are also spread uniformly around the town.



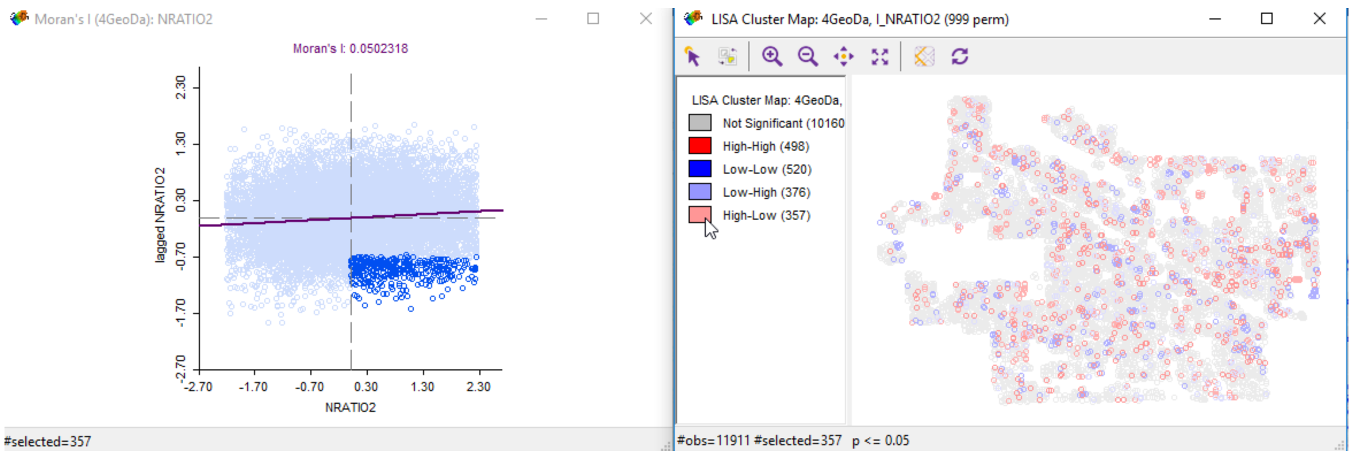
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.



Lake

Summary

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- Owner Occupancy data was considered, but did not prove to be statistically significant
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Hyde Park was valued using the Multiple Regressions Analysis direct market comparison method of valuation
- Log linear models introduce what is called a retransformation bias
- The bias is corrected to ensure that the weighted mean ratio of estimated value to sale price is 1.000

The Data

Sales Counts

As in previous models, the initial selection of sales for modeling followed the practices of CCAO. The filters employed were:

*select if (amount1>45000).
*select if (amount1<700000).
*select if (multi<1).
*select if sqftb<9000.
*select if puremarket=1
This yielded 10,160 as the starting point for sale used in the analysis.

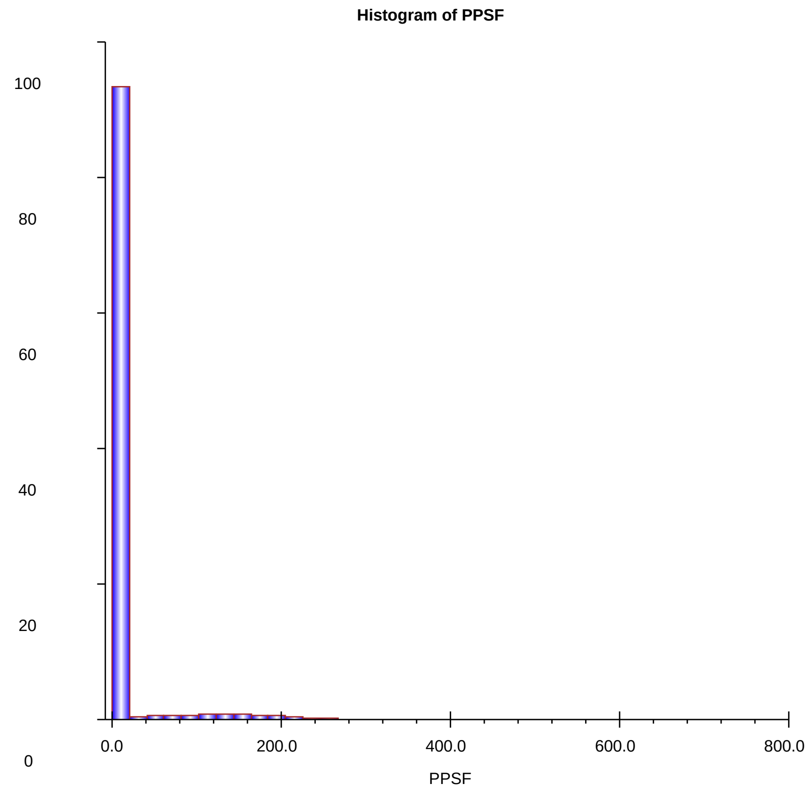
Data Fields

The initial list of data items available for analysis is provided in Appendix A Variable Definitions. Certain additional data fields were created. Those that were relevant to Lake are:

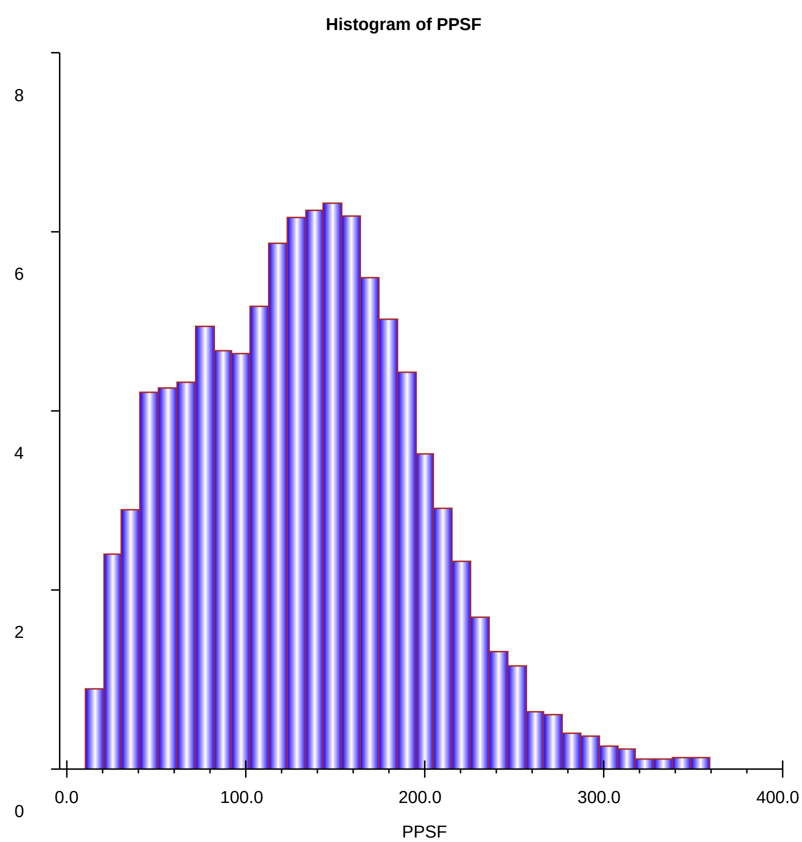
Location Factor

The process employed for determining the Location Factors for Lake Township is outlined in Appendix B Location Factor.

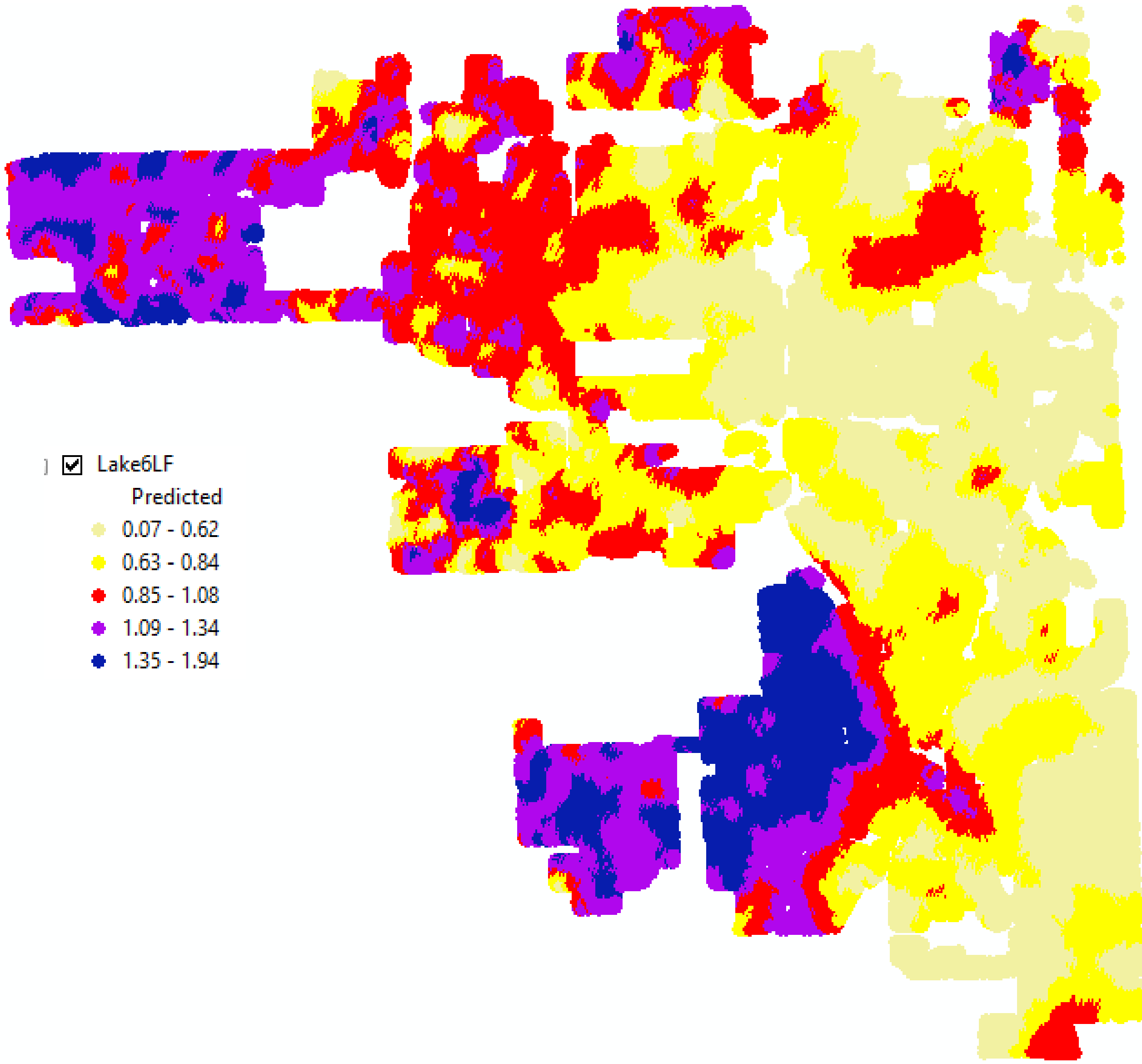
The initial histogram of price per square foot showed that there were some serious outliers. As can be seen, the higher values (say>$350/sqft cause the scale to mask the distribution of realistic values.



A Rational trimming of the highest and lowest values took the sales count from 10,160 to 10,071 with a histogram as in the next image.

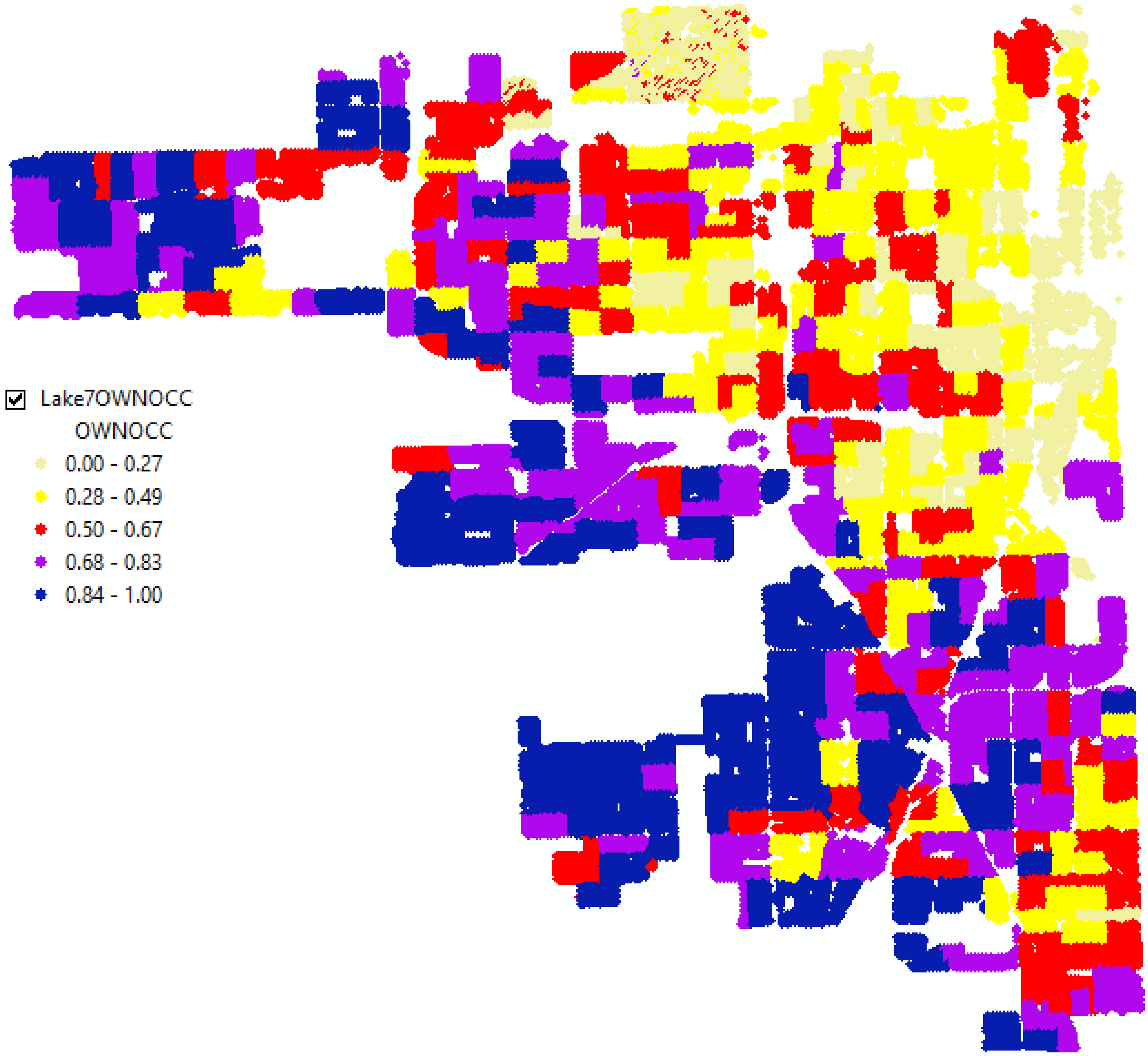


The location factor variable was developed according to the process described in Appendix B Location Factor. A thematic map of the location factor is presented in the next image.



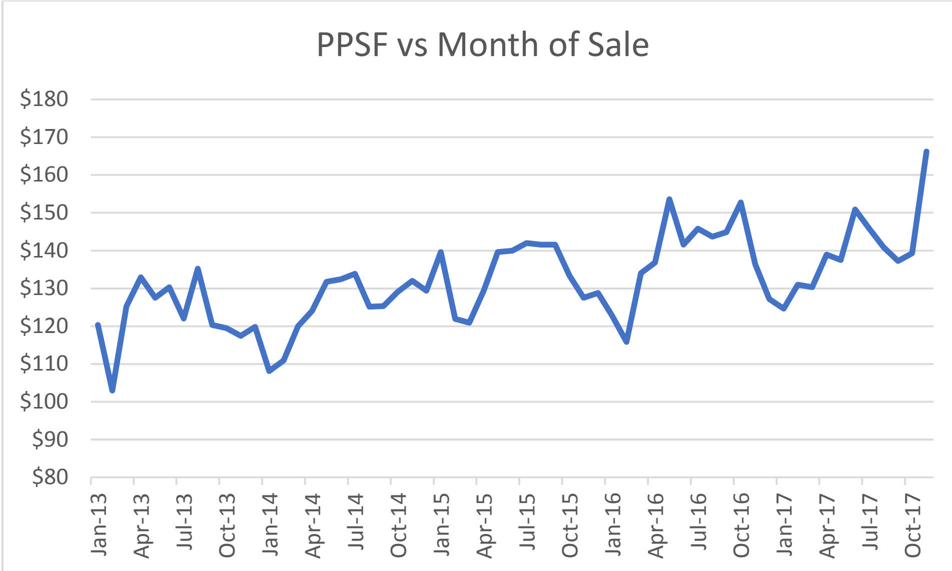
Owner Occupancy

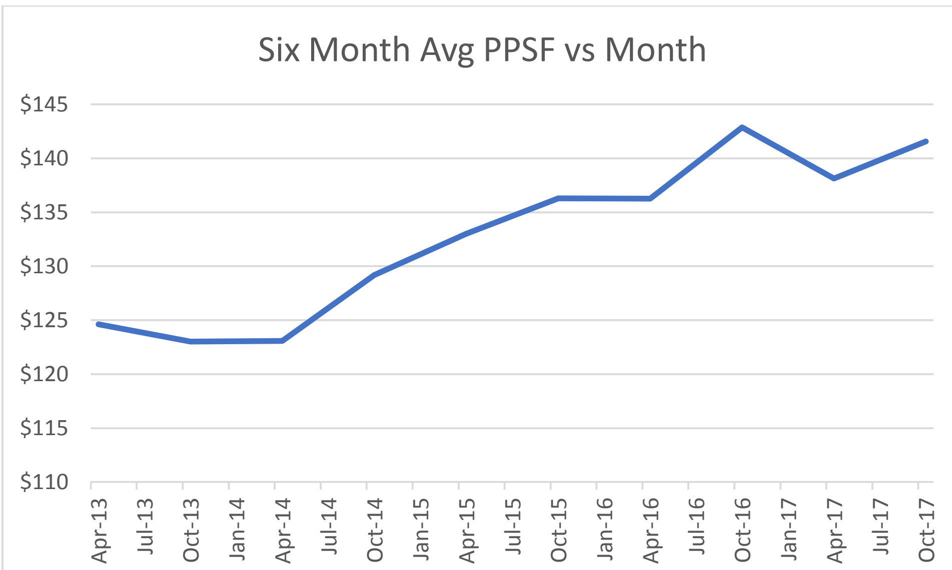
The process for obtaining the Owner Occupancy variable is described in Appendix C Owner Occupancy. The resultant variable is shown below.



Reverse Half of Sale

As was done for other townships in the Chicago Triad, the Reverse Half of Sale was chosen as the variable to express time dependency of value. The reason is that it is more stable than a monthly variable.



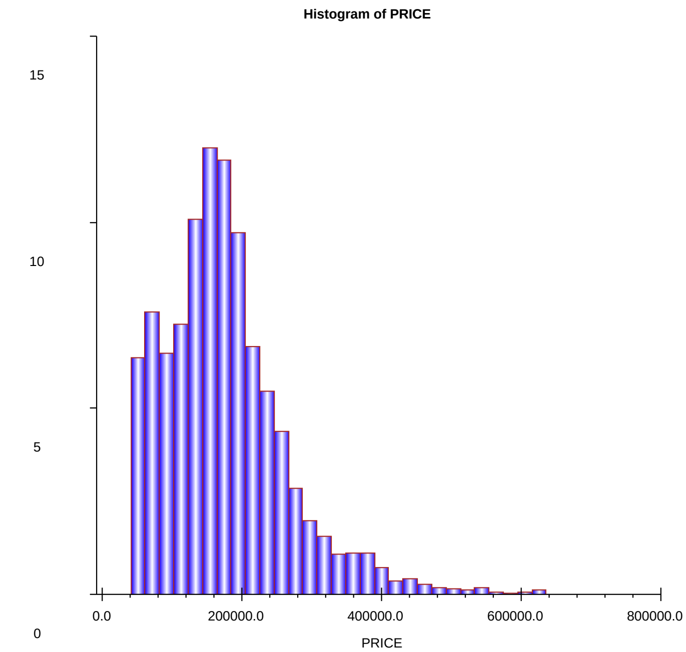


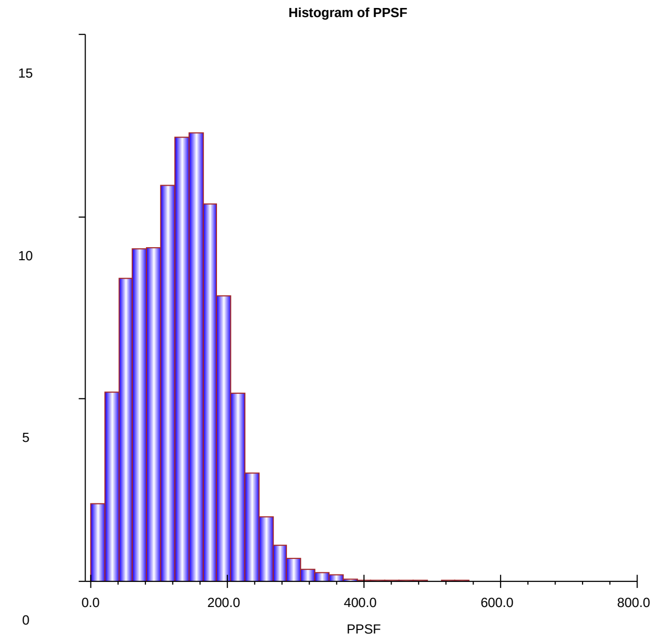
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.

Price and Price per Square foot

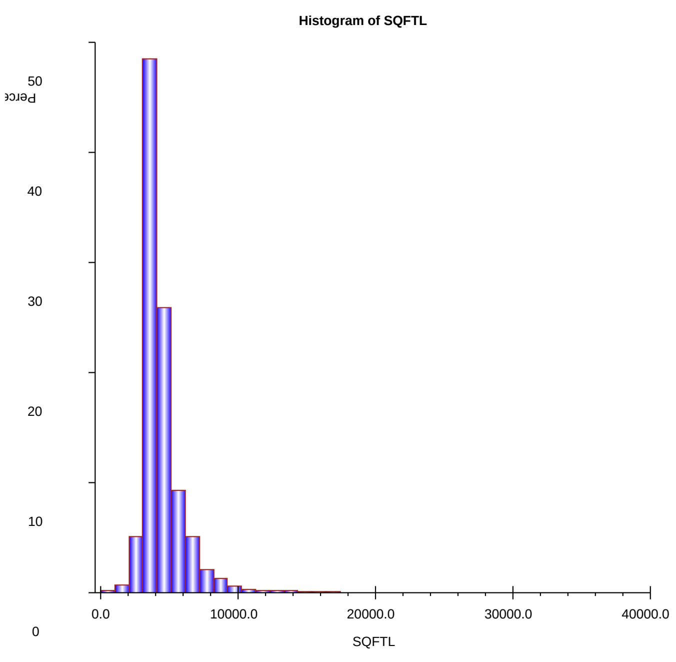
The price histogram fits the range of prices specified at the outset of modeling, namely a range of $100,000-$990,000. The price per square foot range indicates what are likely to be outlier situations. In other words, $600-$700 per square foot and above is not likely to represent a true open market situation.

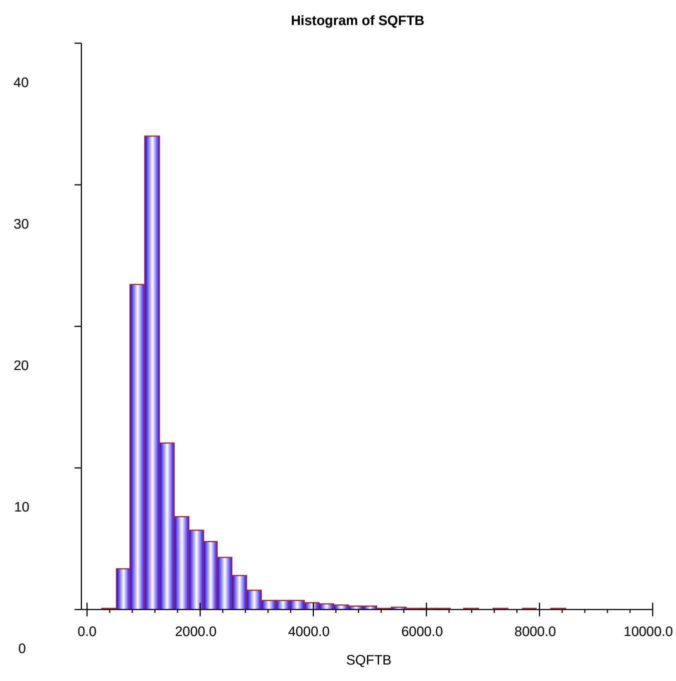




Square foot Land and Building

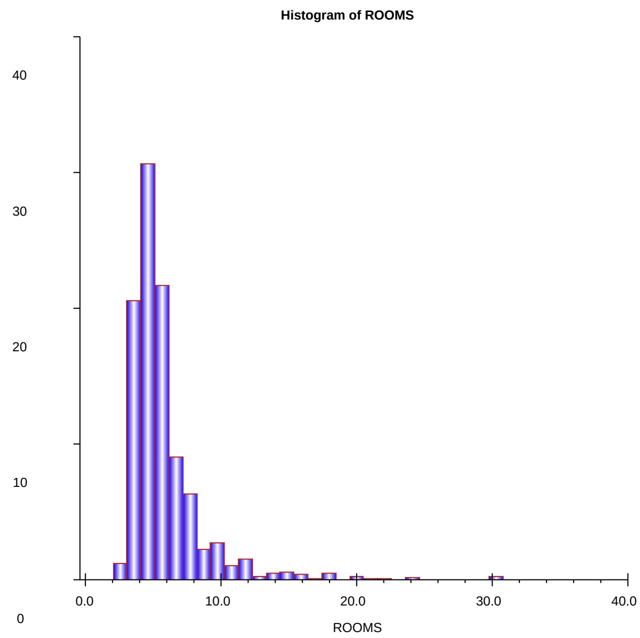
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not. It will be determined as part of the outlier detection process.

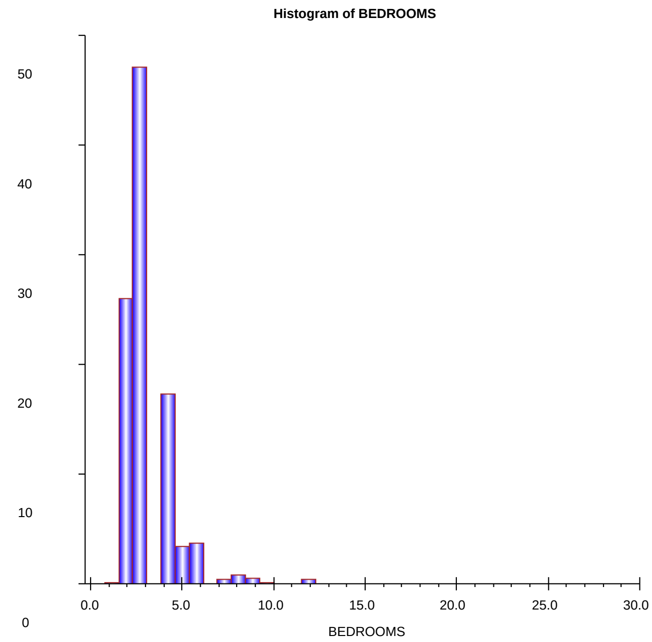




Rooms and Bedrooms

Looks like there are homes with 36 rooms and some with 18 bedrooms. For the specific case of 36/18, it looks as if they are six-unit apartment buildings with each unit having 6 rooms and 3 bedrooms.





Model Structure and Calibration

Outlier Detection

The method for outlier detection is described in detail in Appendix D Outlier Detection.. After following that process the sales used for the analysis became 8,154 of the original 10,160 (19.7%)

Structure

Note: the results below are from preliminary versions of the models. The final model has better performance statistics than shown here.

As in other townships, both the additive and multiplicative forms of the model were evaluated. Considering the results for COD, PRD and PRB, the multiplicative form of the model was used.

Model	Count	Median	WgtMean	COD	PRD	PRB
Additive	9354	1.006	0.997	22.581	1.065	-0.119
Multiplicative	9355	0.985	0.972	21.245	1.063	-0.159

Retransformation Bias

The reader will note that the median and weighted mean for the multiplicative model are lower than that for the additive model. This is a result of what is called the "retransformation bias". This topic is discussed in more detail in Appendix E Retransformation Bias. The final values are corrected for this characteristic of the multiplicative model calibration process.

Location Factor and Owner Occupancy

Both the Location Factor variable and the Owner Occupancy variable we statistically significant and important in Lake Township.

The final MRA Model

Variable	Coefficient	T- Value	Variable	Coefficient	T- Value	Variable	Coefficient	T- Value
Intercept	8.43912	66.552	(CL6=1)	0.09330	5.414	(SUBAREA=2)	0.11753	4.732
lnSFL	0.12159	11.359	(RS=2)	-0.02252	-2.959	(SUBAREA=3)	0.18977	3.710
lnSFB	0.38197	32.288	(RS=3)	-0.07322	-2.728	(SUBAREA=4)	-0.26617	-4.697
lnFIXT	0.11005	8.760	(RS=4)	-0.00238	-0.171	(SUBAREA=5)	0.44444	7.487
lnAGE	-0.13867	-22.948	(RS=5)	-0.03129	-3.833	(SUBAREA=6)	-0.03531	-0.581
lnFIREPL	0.06355	5.918	(NUM=1)	-0.02547	-1.364	(SUBAREA=7)	0.02995	0.506
lnLocF	0.36765	13.388	(NUM=2)	-0.13433	-5.557	(SUBAREA=8)	0.00672	0.115
lnOWNOC	0.11390	4.076	(NUM=3)	-0.22615	-8.446	(SUBAREA=9)	0.15658	2.865
(NGHCDE=40)	0.01712	0.348	(NUM=4)	-0.19122	-3.284	(SUBAREA=10)	0.16104	3.123
(NGHCDE=51)	-0.66764	-13.987	(NUM=5)	-0.23709	-5.518	(SUBAREA=11)	0.06794	1.372
(NGHCDE=52)	0.50432	10.676	(NUM=6)	0.03316	1.737	(SUBAREA=12)	0.04321	0.727
(NGHCDE=61)	-0.04213	-0.758	(GAR=2)	-0.00440	-0.461	(SUBAREA=13)	0.06975	1.278
(NGHCDE=70)	-0.17549	-6.643	(GAR=3)	0.01885	3.141	(SUBAREA=14)	0.01395	0.267
(NGHCDE=71)	-0.27640	-9.669	(GAR=4)	0.05664	4.271	(SUBAREA=15)	0.11567	2.229
(NGHCDE=80)	-0.52487	-11.116	(GAR=5)	0.03535	1.197	(SUBAREA=16)	-0.19627	-2.917
(NGHCDE=91)	-0.77671	-16.505	(GAR=6)	-0.07860	-1.227	(SUBAREA=17)	-0.06673	-1.132
(NGHCDE=92)	-0.24490	-4.731	(GAR=7)	-0.01993	-2.821	(SUBAREA=18)	-0.00335	-0.057
(NGHCDE=110)	-0.11042	-6.285	(GAR=8)	-0.09548	-1.746	(SUBAREA=19)	-0.05268	-0.873
(NGHCDE=120)	-0.55795	-10.656	(CL10=1)	-0.32284	-12.609	(SUBAREA=20)	-0.04065	-0.671
(NGHCDE=121)	-0.73786	-10.626	(CL95=1)	-0.14752	-5.507	(SUBAREA=21)	-0.14302	-2.361
(NGHCDE=130)	-0.83244	-14.614	(RHOS=2)	-0.01843	-1.972	(SUBAREA=22)	0.08989	1.713
(NGHCDE=150)	-0.22399	-5.492	(RHOS=3)	-0.03004	-3.292	(SUBAREA=23)	0.01604	0.307
(NGHCDE=151)	-0.08582	-1.545	(RHOS=4)	-0.06455	-6.912	(SUBAREA=24)	-0.02925	-0.533
(NGHCDE=170)	-1.01040	-16.014	(RHOS=5)	-0.07640	-8.341	(SUBAREA=25)	-0.14807	-2.126
(NGHCDE=171)	-0.74768	-11.745	(RHOS=6)	-0.11099	-11.311	(SUBAREA=26)	0.09685	1.610
(NGHCDE=191)	-0.19950	-4.840	(RHOS=7)	-0.17676	-18.052	(SUBAREA=27)	0.23074	3.085
(NGHCDE=192)	-0.36906	-5.753	(RHOS=8)	-0.19450	-18.180	(SUBAREA=28)	0.30570	4.512
(NGHCDE=193)	-0.18699	-2.887	(RHOS=9)	-0.24091	-24.375	(SUBAREA=29)	0.18991	2.363
(NGHCDE=194)	-0.27233	-4.440	(RHOS=10)	-0.26863	-24.920	(SUBAREA=30)	0.01819	0.197
(NGHCDE=200)	-0.20535	-5.663	(BSFN=2)	-0.02799	-1.227	(SUBAREA=31)	-0.19854	-1.971
(NGHCDE=212)	-0.27522	-3.112	(BSFN=3)	-0.03580	-6.454	(SUBAREA=32)	0.26581	3.120
(NGHCDE=221)	-0.50774	-7.001				(SUBAREA=33)	0.19224	2.792
(NGHCDE=222)	-1.16291	-13.176				(SUBAREA=34)	0.24691	3.036
(NGHCDE=223)	-0.21327	-2.252				(SUBAREA=35)	0.14344	2.233
(NGHCDE=230)	-0.13017	-2.352				(SUBAREA=36)	0.11223	1.711
(NGHCDE=251)	-0.42217	-6.447						
(NGHCDE=260)	-0.45531	-7.597						
(NGHCDE=271)	0.34315	7.899						
(NGHCDE=274)	0.07525	1.730						
(NGHCDE=281)	-0.64526	-14.311						
(NGHCDE=282)	-0.45263	-12.182						
(NGHCDE=293)	-0.24503	-2.770						
(NGHCDE=300)	0.28011	6.036						
(NGHCDE=310)	-0.76035	-13.939						
(NGHCDE=312)	-0.45017	-9.175						
(NGHCDE=321)	-0.59173	-13.518						
(NGHCDE=323)	-0.22463	-4.380						
(NGHCDE=330)	-0.75514	-15.666						
(NGHCDE=345)	-0.88015	-13.845						
(NGHCDE=350)	-0.02605	-0.883						
(NGHCDE=361)	-0.15524	-3.084						
(NGHCDE=380)	0.21848	5.371						
(NGHCDE=420)	0.42417	8.048						
(NGHCDE=422)	0.24601	14.677						
(NGHCDE=423)	0.19917	11.602						
(NGHCDE=431)	0.12816	3.467						
(NGHCDE=432)	0.35995	9.095						

Model Performance Stats by NBHD

NGHCDE	Count	Median	WgtMean	COD	PRD	PRB
30	1,240	0.989	0.979	13.764	1.029	-0.235
40	347	0.973	0.966	15.992	1.037	-0.258
51	35	1.013	0.983	19.158	1.070	-0.262
52	98	0.987	0.973	16.330	1.028	-0.115
61	67	0.987	0.966	10.507	1.017	-0.134
70	236	0.926	0.932	19.125	1.054	-0.389
71	91	0.924	0.929	20.162	1.053	-0.404
80	76	1.025	0.981	18.080	1.051	-0.330
91	59	1.063	0.971	23.810	1.078	-0.428
92	28	0.992	0.974	17.692	1.051	-0.278
110	176	0.955	0.972	16.128	1.037	-0.347
120	33	0.953	0.987	19.706	1.053	-0.378
121	31	1.031	1.016	14.752	1.015	0.048
130	23	1.008	1.011	17.520	1.028	-0.143
150	88	0.940	0.950	20.177	1.048	-0.364
151	43	0.983	0.952	17.801	1.050	-0.388
170	31	1.012	0.990	22.317	1.060	-0.253
171	16	0.948	0.969	17.043	1.024	-0.024
191	78	1.000	0.982	15.074	1.035	-0.484
192	105	0.954	0.947	14.223	1.031	-0.440
193	211	0.989	0.984	13.557	1.026	-0.332
194	104	0.982	0.980	13.478	1.032	-0.521
200	411	0.980	0.971	16.405	1.040	-0.384
212	103	0.997	0.963	22.054	1.072	-0.459
221	83	0.928	0.941	20.871	1.046	-0.307
222	15	1.057	0.990	17.920	1.068	-0.455
223	10	0.837	0.843	14.860	1.045	-1.076
230	26	0.951	0.959	15.493	1.039	-0.462
251	20	0.911	0.962	21.950	1.045	-0.389
260	40	0.968	0.964	19.276	1.039	-0.281
271	411	0.986	0.977	14.448	1.027	-0.183
274	148	1.008	0.984	17.288	1.037	-0.156
281	68	1.059	0.983	17.263	1.054	-0.246
282	235	1.021	0.980	18.573	1.050	-0.165
293	6	0.911	0.884	20.673	1.083	-0.551
300	97	0.957	0.972	18.698	1.041	-0.290
310	35	1.049	1.000	20.927	1.060	-0.491
312	53	1.021	0.959	19.313	1.057	-0.374
321	69	1.021	0.984	18.601	1.055	-0.333
323	39	1.026	0.984	18.014	1.042	-0.646
330	64	1.021	0.983	17.089	1.039	-0.125
345	29	0.984	0.980	18.312	1.036	-0.428
350	255	0.989	0.980	13.066	1.027	-0.339
361	45	0.942	0.939	19.550	1.039	-0.113
380	1,581	0.999	0.983	13.897	1.028	-0.166
420	15	0.974	1.006	9.669	1.001	0.605
422	341	0.986	0.974	14.089	1.030	-0.244
423	608	0.988	0.978	15.482	1.035	-0.207
431	98	0.993	0.965	13.078	1.030	-0.330
432	33	1.027	0.981	14.009	1.034	-0.201
Combined	8,154	0.988	0.976	15.586	1.034	-0.084

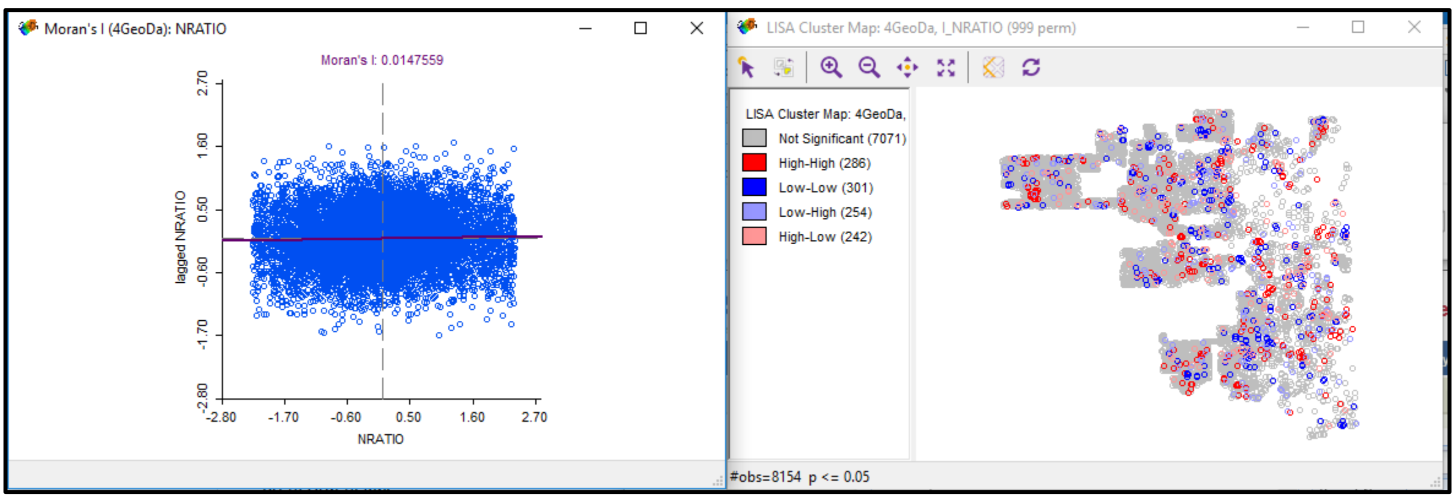
Model Performance Stats by Class

CLASS	Count	Median	WgtMean	COD	PRD	PRB
2	1,615	0.993	0.975	16.333	1.038	-0.146
3	3,637	0.987	0.978	15.151	1.031	-0.100
4	164	0.990	0.953	14.764	1.044	-0.108
5	793	0.983	0.965	14.994	1.040	-0.120
6	175	1.011	0.984	14.093	1.026	-0.100
7	127	1.007	1.003	13.754	1.021	-0.038
8	2	0.870	0.907	10.480	0.959	0.151
10	75	0.981	0.954	17.688	1.046	-0.487
11	1,007	0.966	0.956	17.238	1.047	-0.123
12	86	1.128	1.069	16.360	1.040	-0.062
34	240	0.998	0.991	15.046	1.032	-0.062
78	165	1.005	1.002	11.890	1.009	-0.006
95	68	1.005	1.000	14.406	1.009	0.026
Combined	8,154	0.988	0.976	15.586	1.034	-0.084

Spatial Dependency

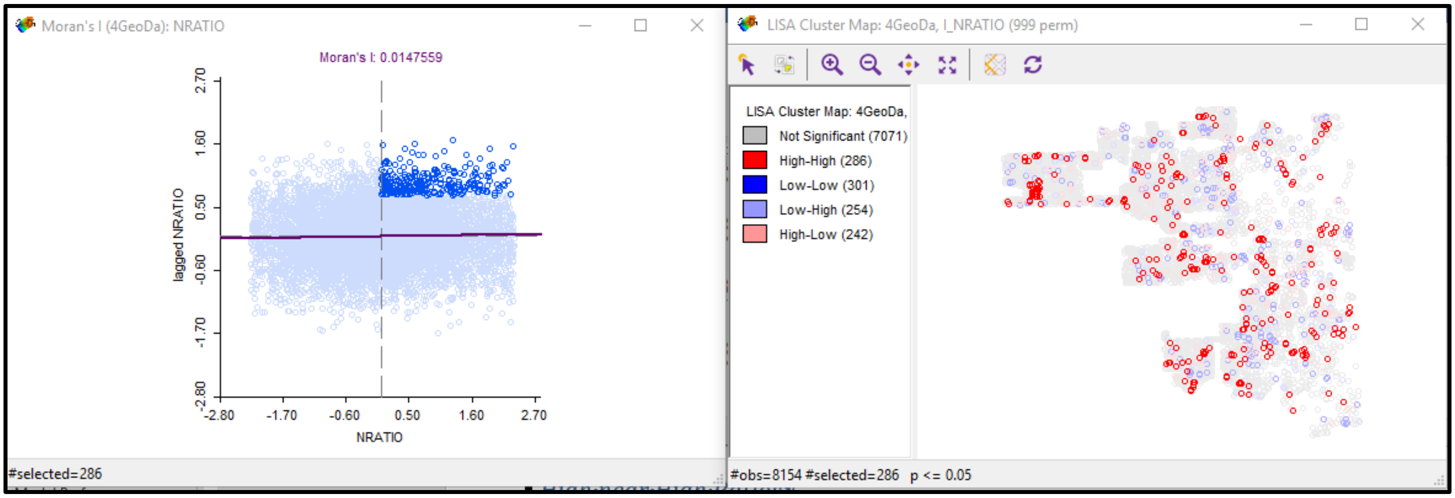
Additional Discussion of Spatial Dependency is provided in Appendix G Spatial Dependency.docx.

The Moran's I global statistic of 0.0147 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of Lake, the statistic is very close to 0.0.



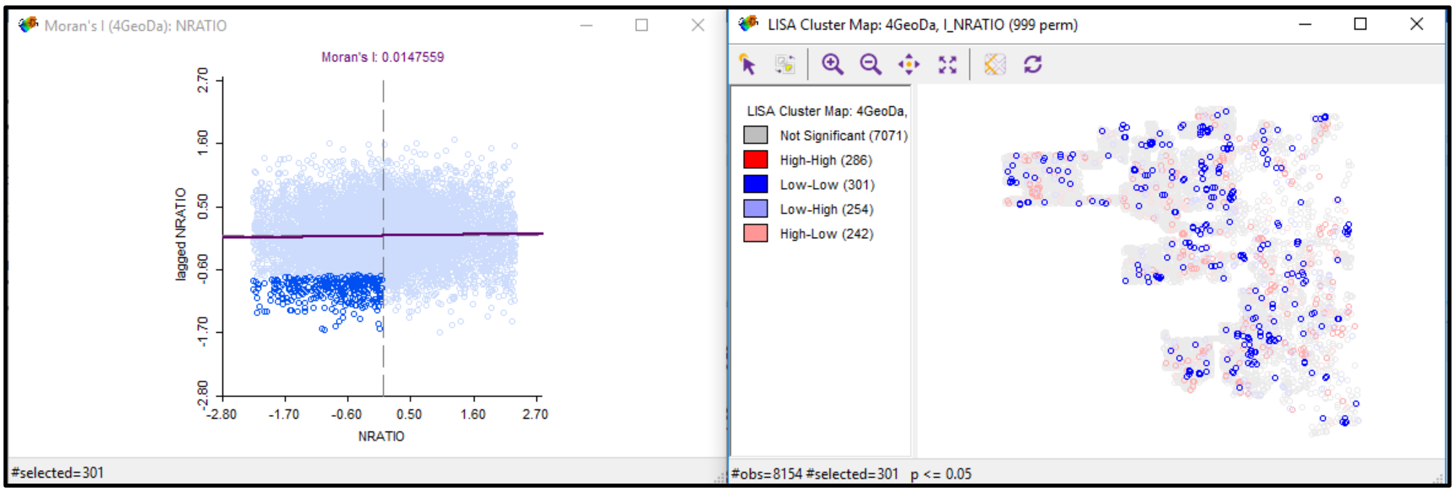
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



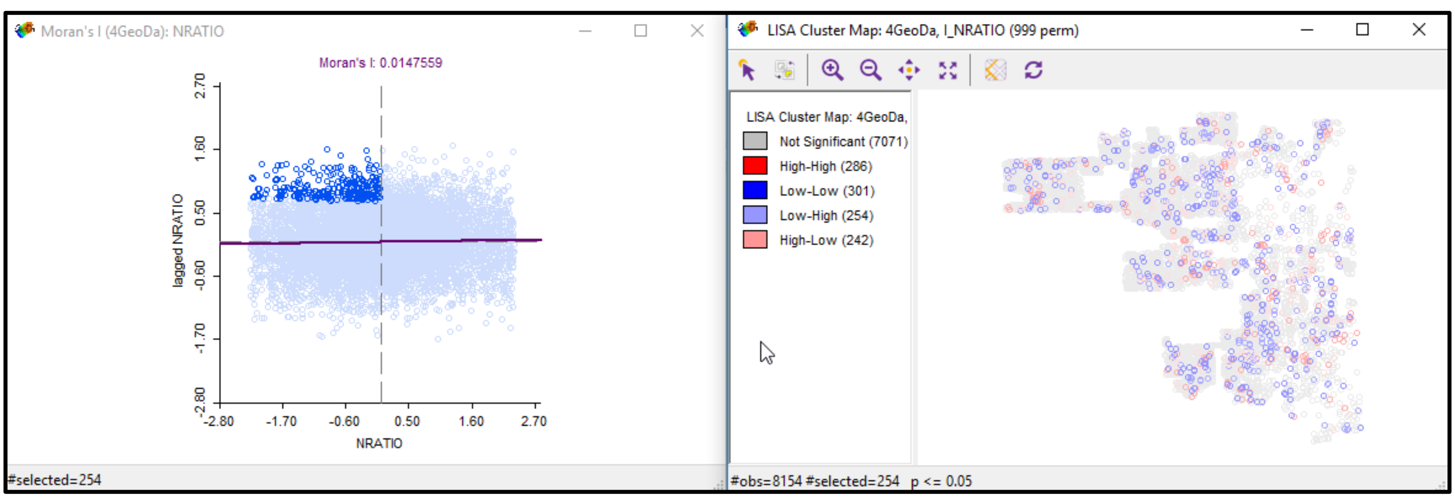
Low near Low Ratios

The low near low ratios are also spread uniformly around the town.



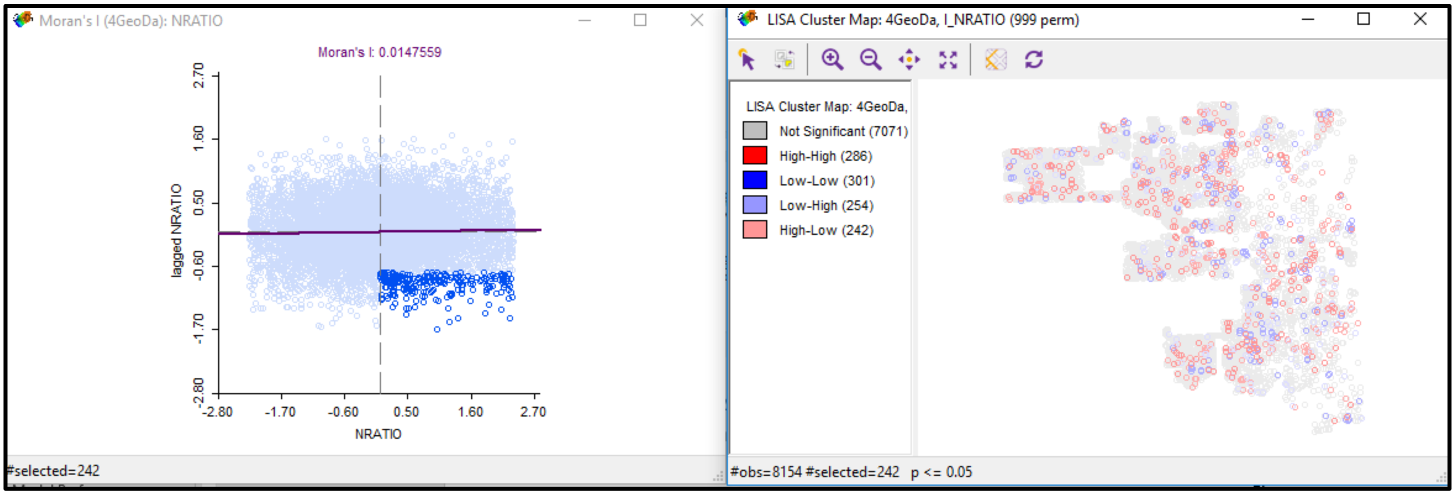
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.


Moran's I (4GeoDa): NRATIO
Moran's I: 0.0147559
lagged NRATIO
2.70
1.60
0.50
-0.60
-1.70
-2.80
-2.80
-1.70
-0.60
0.50
1.60
2.70
NRATIO
#selected=242
LISA Cluster Map: 4GeoDa, I_NRATIO (999 perm)
LISA Cluster Map: 4GeoDa,
Not Significant (7071)
High-High (286)
Low-Low (301)
Low-High (254)
High-Low (242)
#obs=8154 #selected=242 p <= 0.05

West

Summary

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- Owner Occupancy data was considered, but did not prove to be statistically significant
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Hyde Park was valued using the Multiple Regressions Analysis direct market comparison method of valuation
- Log linear models introduce what is called a retransformation bias
- The bias is corrected to ensure that the weighted mean ratio of estimated value to sale price is 1.000
- Performance statistics were well within IAAO standards

The Data

Sales Counts

After filtering according to practices of the CCAO lister below, the starting number of sales used in the analysis was 4,804.

*select if (amount1>75000).
*select if (amount1<790000).
*select if (multi<1).
*select if sqftb<9000. Did not use
*do not select if age<10 and (amount1<1600000 and (amount1/sqftb) <75 and class<95).

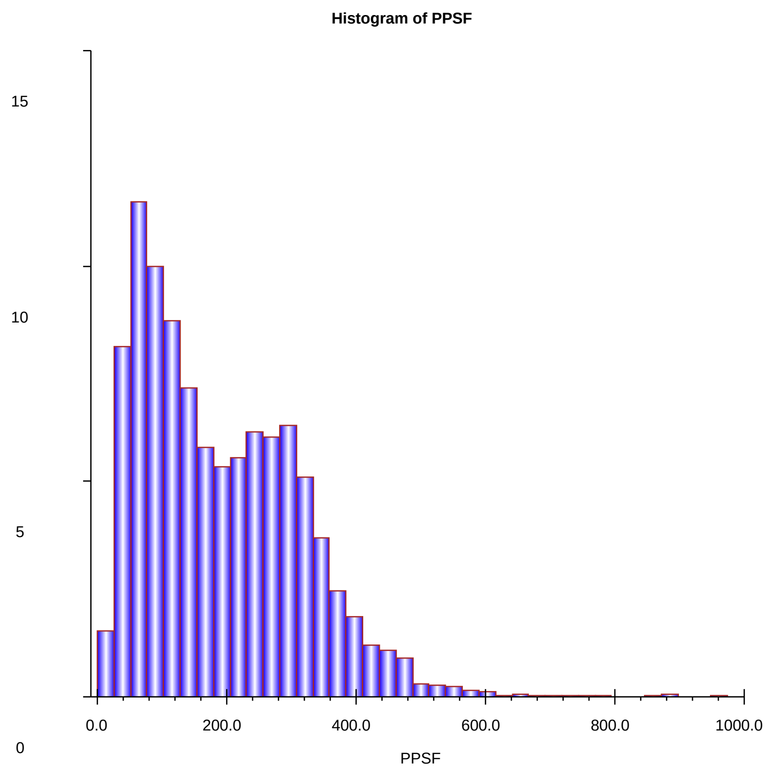
Data Fields

The initial list of data items available for analysis is provided in Appendix A Variable Definitions. Certain additional data fields were created. Those that were relevant to West Township are:

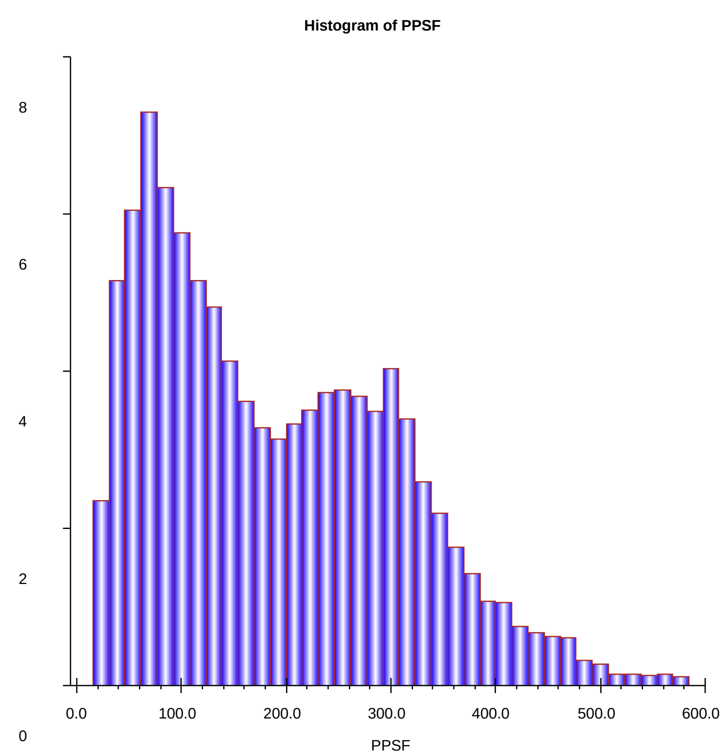
Location Factor

The process employed for determining the Location Factors for West Township is outlined in Appendix B Location Factor.

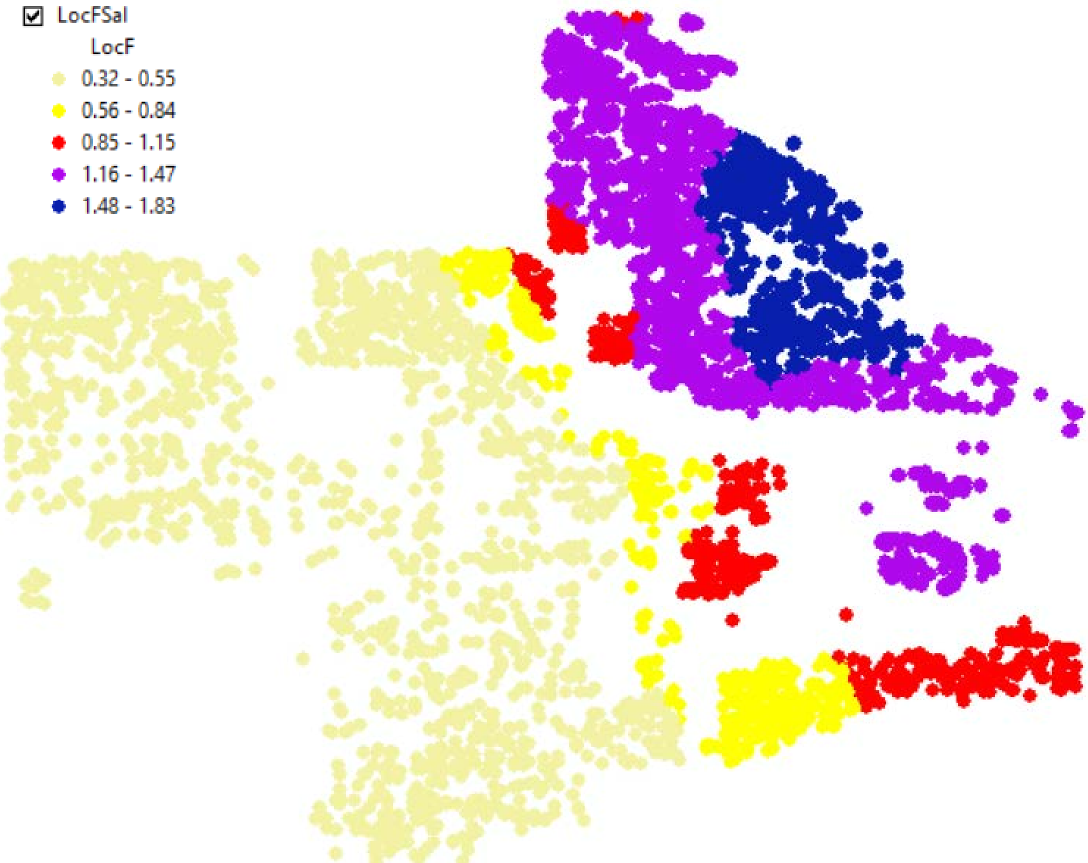
The histogram of price per square foot showed that there were some potential outliers.



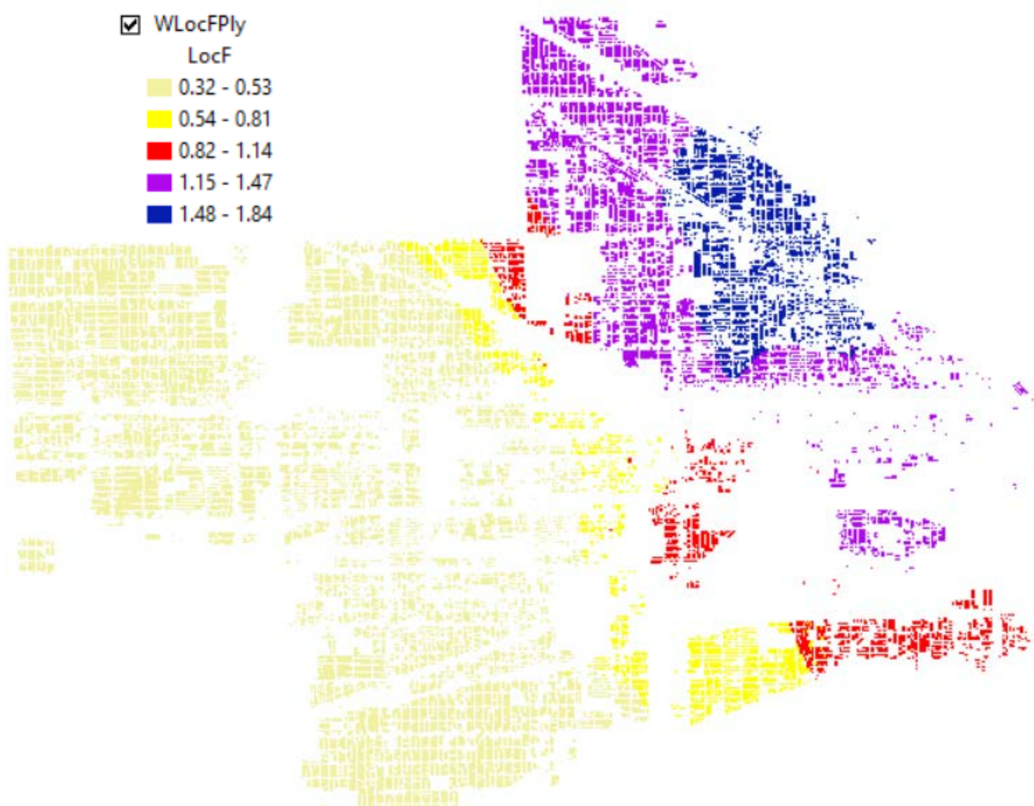
Strictly for purposes of avoiding extreme outliers, the lower half percent (0.5%) and the upper half percent (99.5%) sales were not considered. This amounts to cutoff points of $19.sqft and $584/sqft. The resulting histogram is presented below.



A thematic map of the location factor is presented in the next image. The location factor derived for the sale properties is shown below.

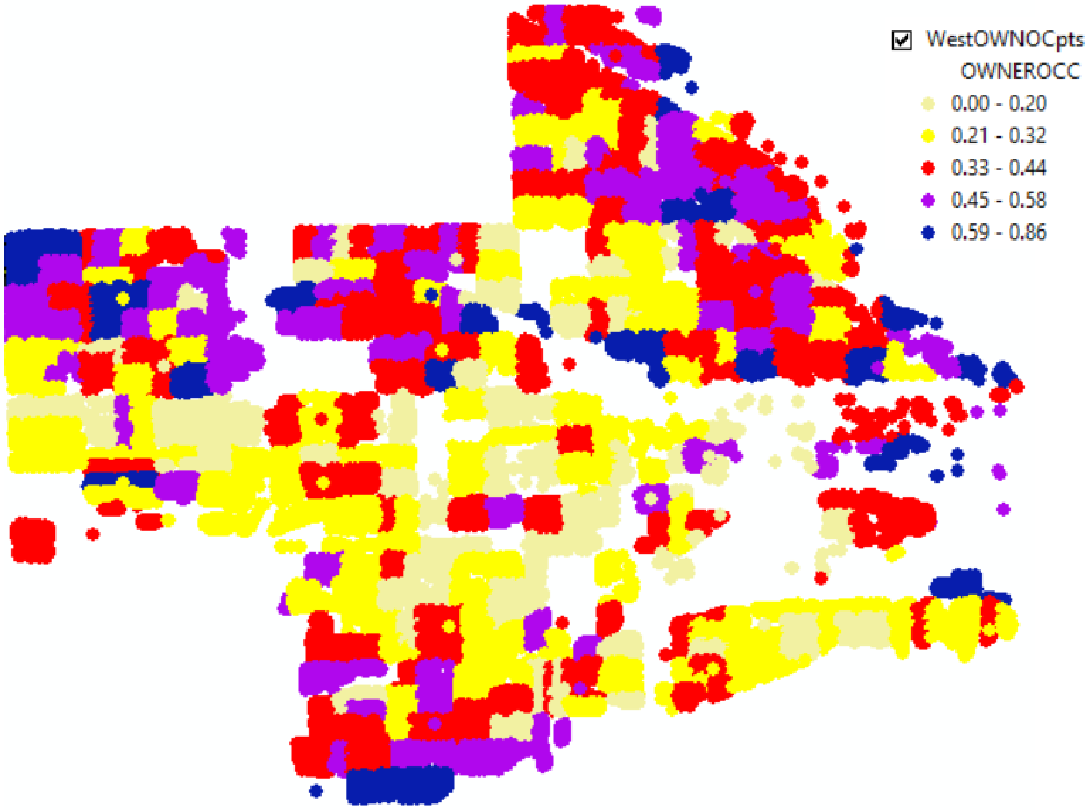


The location factor applied to the parcels in West Township follows below.



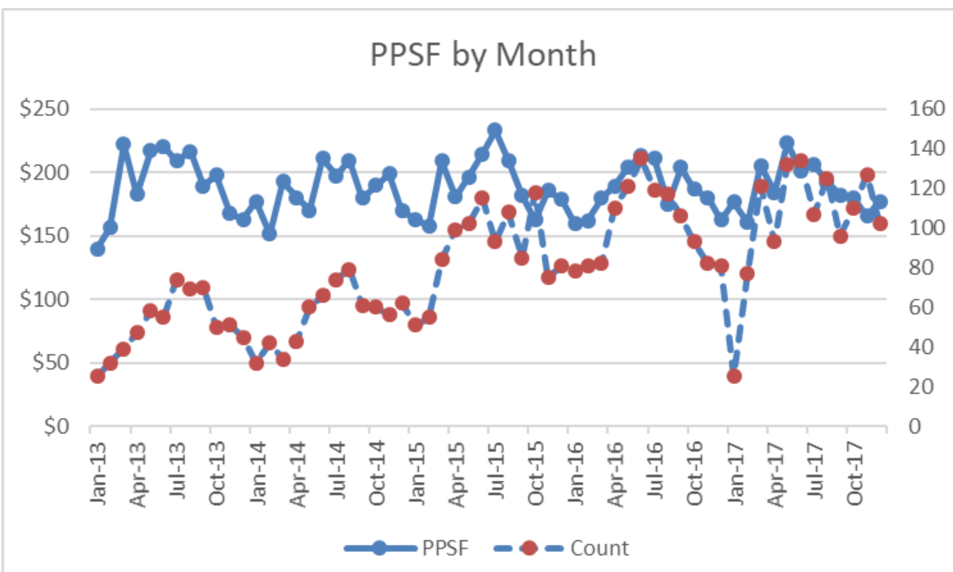
Owner Occupancy

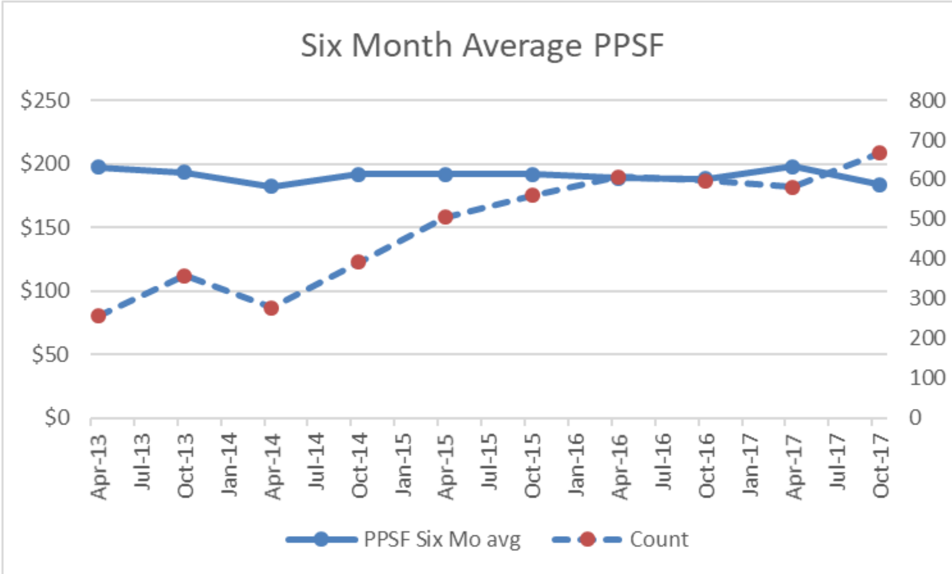
The process for obtaining the Owner Occupancy variable is described in Appendix C Owner Occupancy. The resultant variable is shown below.



Reverse Half of Sale

As was done for other townships in the Chicago Triad, the Reverse Half of Sale was chosen as the variable to express time dependency of value. The reason is that it is more stable than a monthly variable. It is unusual to see the sale count rising steadily over the five-year period, while the PPSF is somewhat flat.



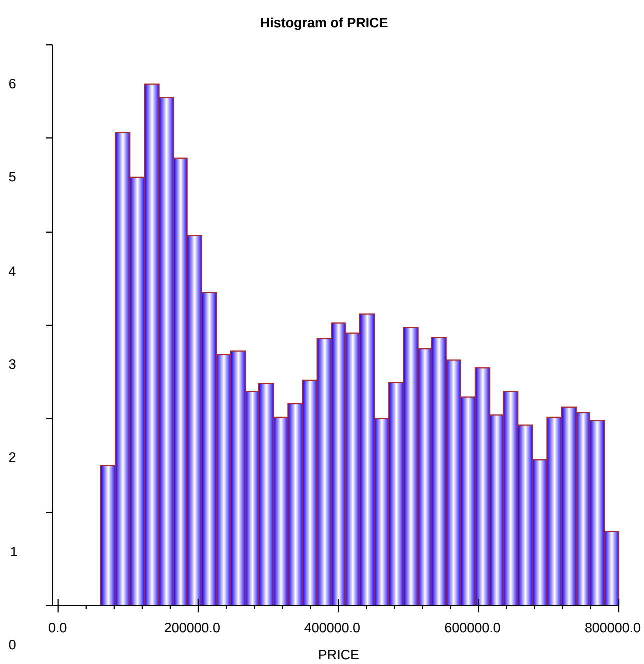


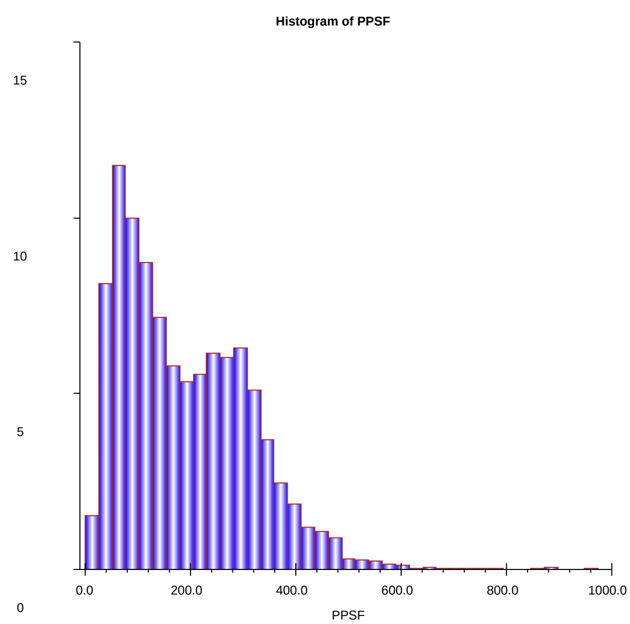
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.

Price and Price per Square foot

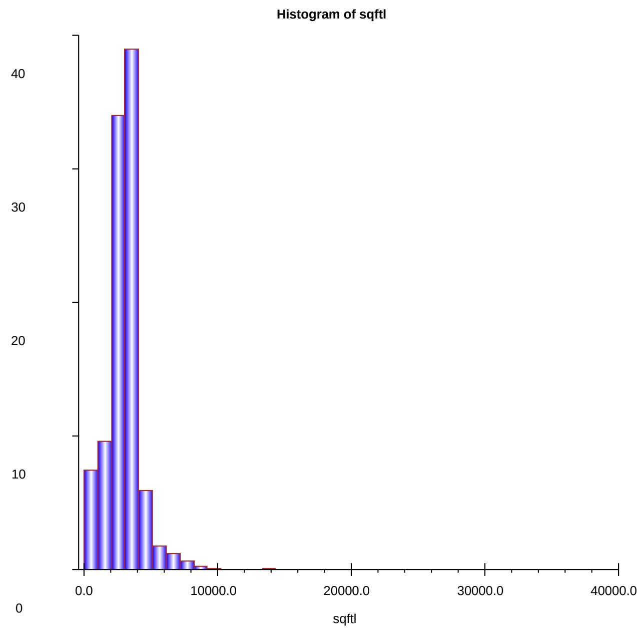
The price histogram fits the range of prices specified at the outset of modeling, namely a range of $75,000-$780,000. The price per square foot range indicates what are likely to be outlier situations. In other words, $600-$700 per square foot and above is not likely to represent a true open market situation

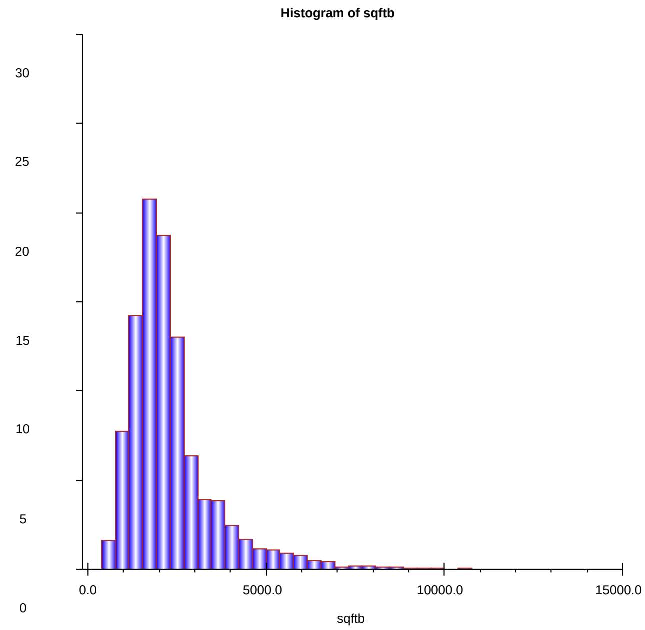




Square foot Land and Building

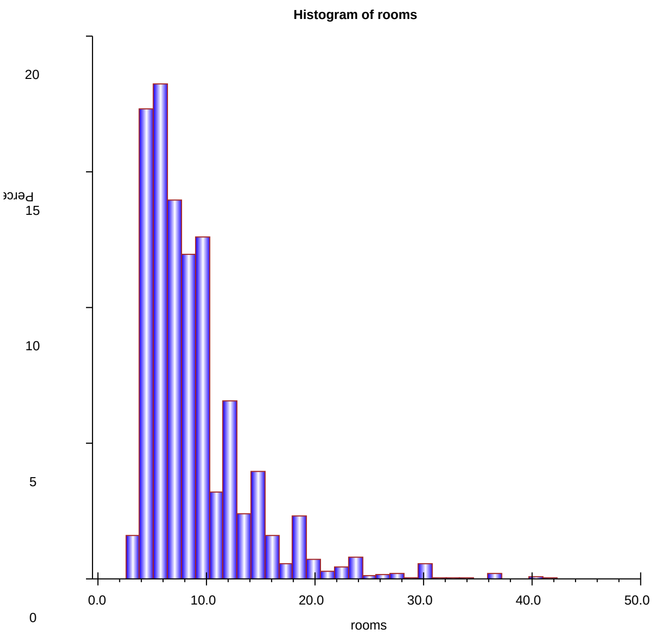
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not. It will be determined as part of the outlier detection process.

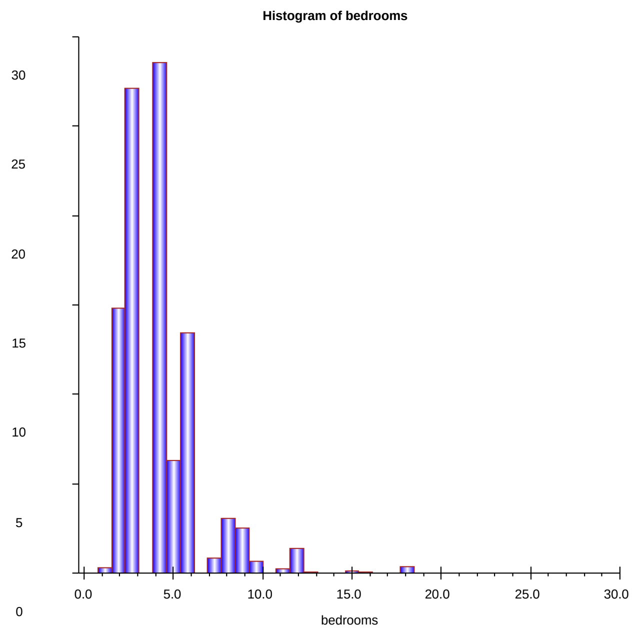




Rooms and Bedrooms

Looks like there are homes with 42 rooms and some with 24 bedrooms. For the specific case of 42/24, which averages 1.75 bedrooms per unit, implying a combination of 0, 1- and 2-bedroom apartments.





Model Structure and Calibration

Outlier Detection

The method for outlier detection is described in detail in Appendix D Outlier Detection. After following that process the sales used for the analysis became 3,677 of the original 4,804 (23.5%)

Structure

Note: the results below are from preliminary versions of the models. No outliers have been removed. The final model has better performance statistics than shown here.

As in other townships, both the additive and multiplicative forms of the model were evaluated. Considering the results for COD, PRD and PRB, the multiplicative form of the model was used.

Structure	Count	Median	WgtMean	COD	PRD	PRB
Multiplicative	4,804	0.962	0.958	27.698	1.105	-0.125
Additive	4,804	0.999	1.000	30.920	1.113	-0.093

Retransformation Bias

The reader will note that the median and weighted mean for the multiplicative model are lower than that for the additive model. This is a result of what is called the "retransformation bias". This topic is discussed in more detail in Appendix E Retransformation Bias. The final values are corrected for this characteristic of the multiplicative model calibration process.

Location Factor and Owner Occupancy

Both the Location Factor variable and the Owner Occupancy variable were statistically significant and important in West Township. This is seen in the next section, The final MRA Model where the location factor variable has a T-Value of 27.99.

The final MRA Model

Variable	Coefficient	T-Value	Variable	Coefficient	T-Value
Intercept	6.97800	57.25	(BSMT2=1)	-0.06031	-7.78
lnLocFp	1.47063	27.99	(BSMT4=1)	-0.05365	-3.06
lnOCCp	0.15635	5.07	(CL3=1)	-0.05703	-4.92
lnLSF	0.14692	12.16	(CL7=1)	-0.13148	-6.25
lnAGE	-0.09573	-13.75	(CL8=1)	-1.03480	-6.19
lnSFB	0.42562	34.86	(CL11=1)	-0.13201	-6.67
lnFIXT	0.16576	9.78	(CL12=1)	-0.32549	-13.03
lnFIREPL	0.03920	3.33	(CL78=1)	-0.12687	-5.40
(gar3=1)	0.05790	9.53	(CL95=1)	-0.15791	-7.67
(gar6=1)	-0.25911	-4.07	(RHOS1=1)	0.27531	19.53
(num1=1)	-0.08044	-4.31	(RHOS2=1)	0.27324	19.23
(num2=1)	-0.10115	-4.88	(RHOS3=1)	0.23884	16.85
(num3=1)	-0.11427	-4.53	(RHOS4=1)	0.22131	15.65
(num4=1)	-0.26508	-7.16	(RHOS5=1)	0.17726	12.34
(num5=1)	-0.17778	-6.07	(RHOS6=1)	0.15890	11.00
(extcon2=1)	0.04108	5.28	(RHOS7=1)	0.12896	8.59
(extcon3=1)	0.02426	2.21	(RHOS8=1)	0.07501	4.61
(extcon4=1)	0.11222	3.06	(RHOS9=1)	0.07673	5.03
(BSFN2=1)	-0.05429	-4.62			
(NB20=1)	0.19262	12.87			
(NB30=1)	0.39324	15.11			
(NB51=1)	0.44147	11.51			
(NB52=1)	0.49380	13.45			
(NB60=1)	0.45702	12.80			
(NB80=1)	-0.13937	-4.54			
(NB91=1)	-0.15671	-8.81			
(NB92=1)	-0.24531	-4.17			
(NB101=1)	-0.28181	-9.73			
(NB102=1)	-0.06572	-4.11			
(NB103=1)	0.24132	9.84			
(NB120=1)	0.47305	16.38			
(NB131=1)	0.65397	21.76			
(NB132=1)	0.51372	20.24			
(NB141=1)	0.29231	16.34			
(NB150=1)	0.34462	12.66			
(NB151=1)	0.50341	15.33			
(NB152=1)	0.52573	10.64			

Model Performance Stats by NBHD

nghcde	Count	Median	Mean	COD	COV	PRB
11	95	0.941	0.958	13.862	16.213	-0.125
13	190	1.021	0.990	15.222	18.513	-0.190
20	185	0.993	0.973	15.171	17.952	-0.198
30	588	0.995	0.990	12.589	15.431	-0.123
40	50	1.063	1.024	16.298	19.518	-0.284
51	63	0.980	0.992	14.155	16.889	-0.216
52	67	0.983	0.995	12.840	15.614	-0.215
60	116	0.987	0.992	11.651	13.695	-0.206
80	31	1.008	0.977	14.632	17.986	-0.363
85	8	0.933	0.886	8.713	12.861	0.060
91	111	0.987	0.988	14.904	17.591	-0.203
92	8	1.008	0.961	11.884	13.979	-0.227
101	35	1.004	0.976	17.559	20.417	-0.386
102	148	0.979	0.985	14.704	17.426	-0.146
103	76	0.975	0.985	10.322	13.773	-0.140
104	11	1.013	1.017	7.393	8.964	0.034
115	246	0.990	0.986	14.987	17.593	-0.147
120	391	0.994	0.994	12.236	14.955	-0.125
131	88	0.970	0.985	15.618	18.226	-0.426
132	233	0.998	0.989	11.660	14.269	-0.162
141	235	0.987	0.975	16.832	19.475	-0.142
150	342	0.983	0.983	14.672	17.106	-0.229
151	69	0.987	0.991	14.515	17.624	-0.187
152	20	0.974	1.008	10.520	13.898	0.254
170	271	0.993	1.000	12.220	14.921	-0.023
Combined	3,677	0.992	0.989	13.712	16.473	-0.030

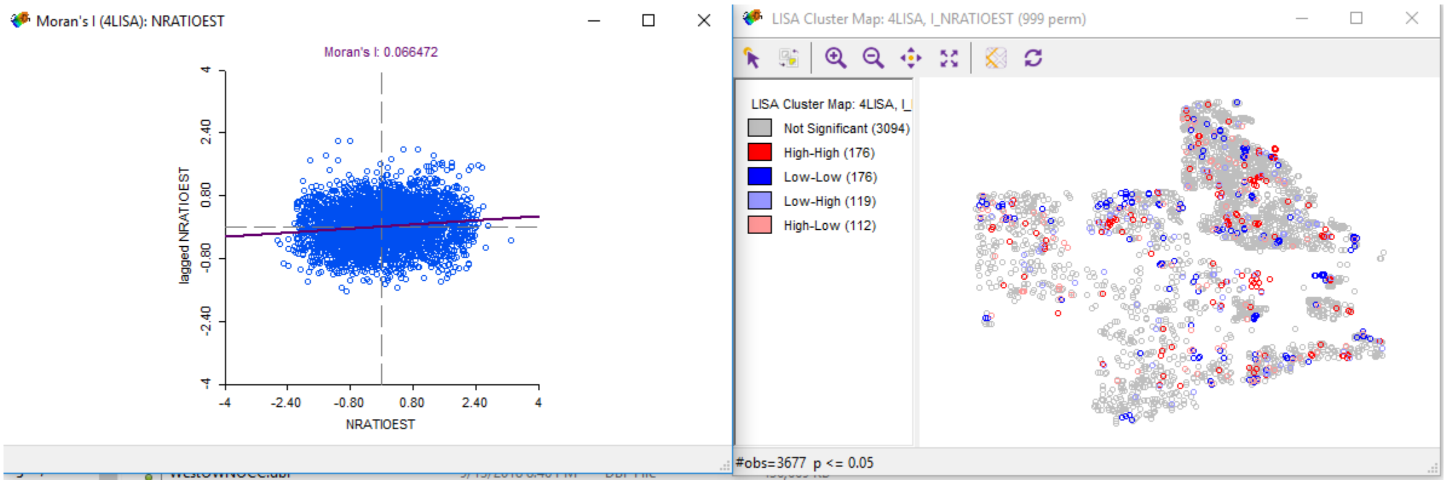
Model Performance Stats by Class

class	Count	Median	Mean	COD	PRD	PRB
2	177	0.973	0.950	16.165	1.037	-0.041
3	373	0.998	0.970	15.464	1.047	-0.062
4	41	1.048	1.048	12.676	1.010	0.004
5	159	0.997	0.993	15.708	1.039	-0.057
6	72	0.964	1.021	13.070	0.989	0.066
7	218	0.980	0.986	10.542	1.023	-0.044
8	1	1.000	1.000	0.000	1.000	0.000
10	37	1.000	1.014	11.140	1.002	0.007
11	1,618	0.992	0.987	14.627	1.028	-0.032
12	176	1.006	0.980	15.450	1.038	-0.071
34	6	1.258	1.216	15.071	1.052	-0.299
78	150	0.994	0.998	9.339	1.009	-0.065
95	649	0.988	0.996	10.954	1.013	-0.045
Combinec	3,677	0.992	0.989	13.712	1.024	-0.030

Spatial Dependency

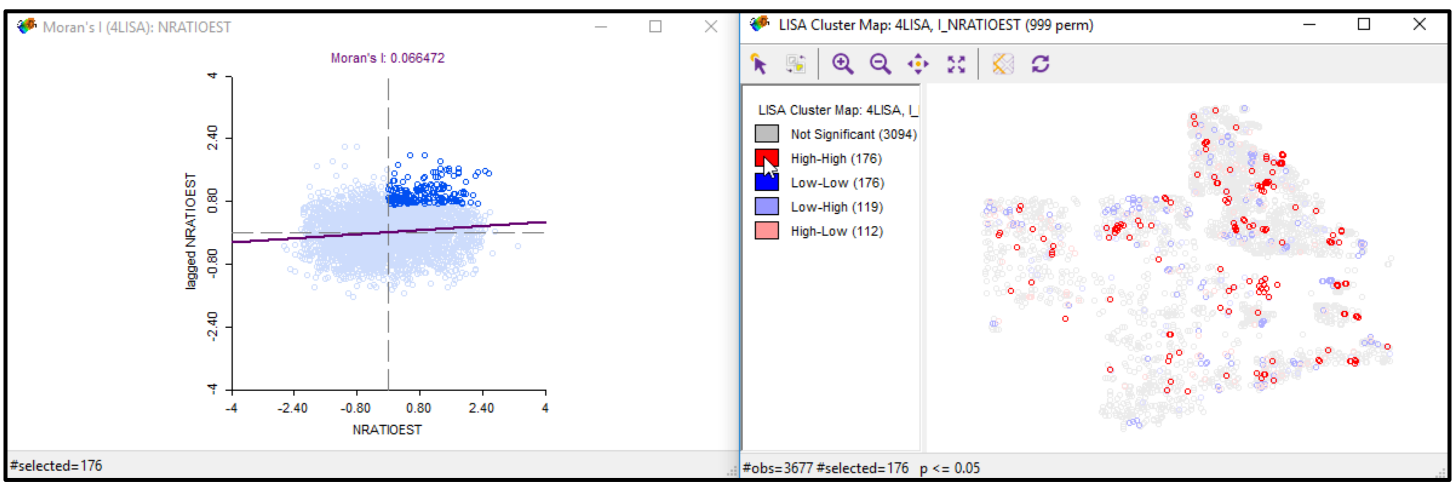
Additional Discussion of Spatial Dependency is provided in Appendix G Spatial Dependency.docx.

The Moran's I global statistic of 0.0665 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of West, the statistic is very close to 0.0.



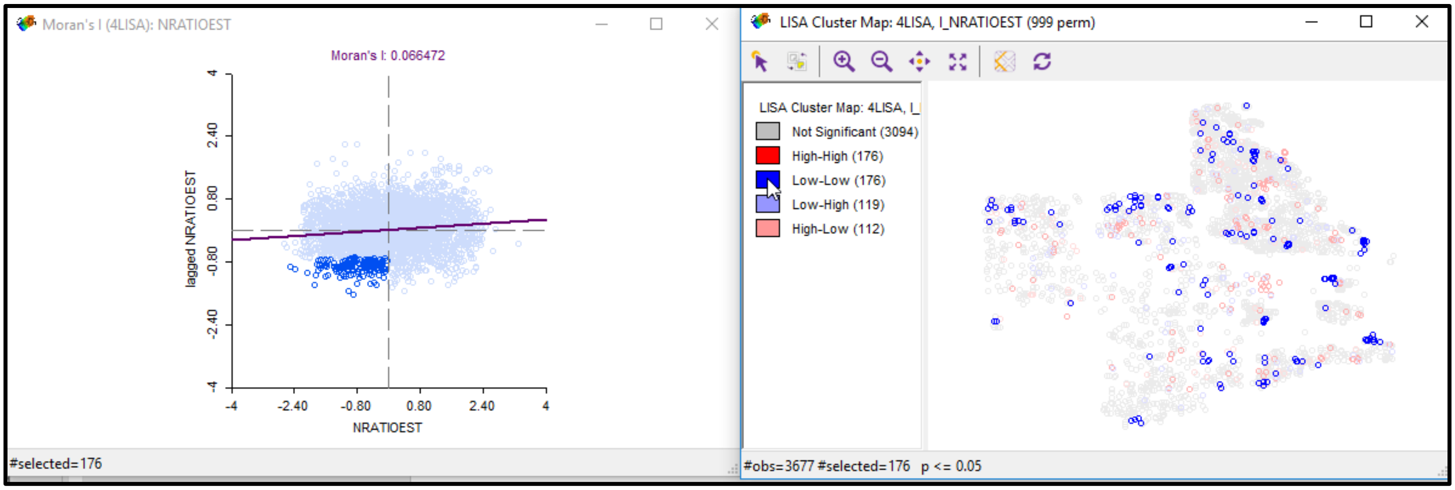
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



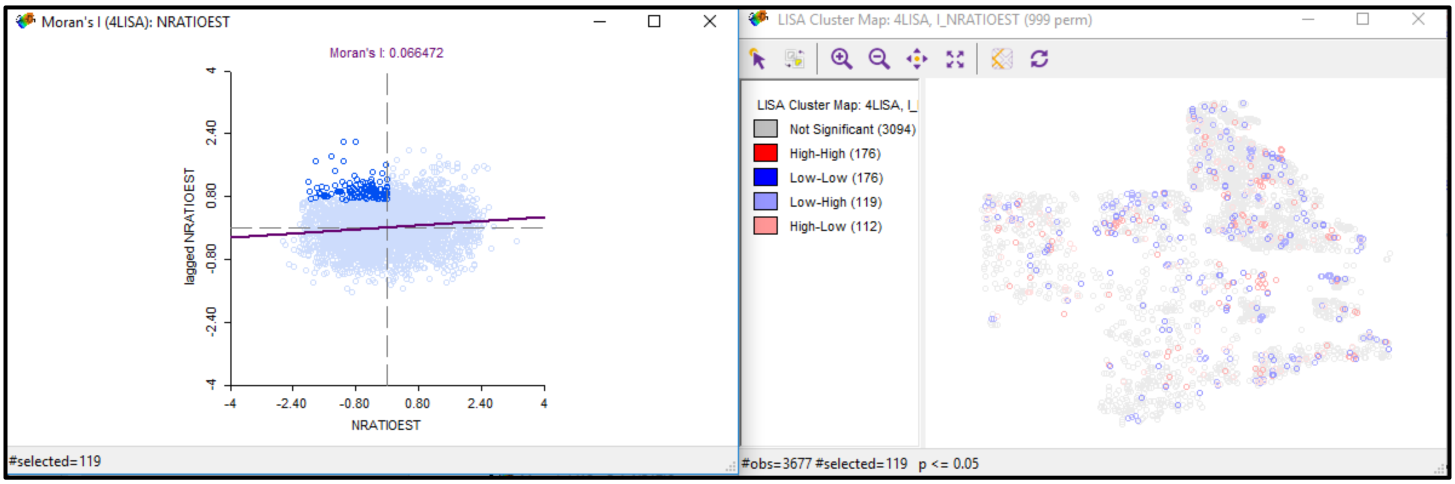
Low near Low Ratios

The low near low ratios are also spread uniformly around the town.



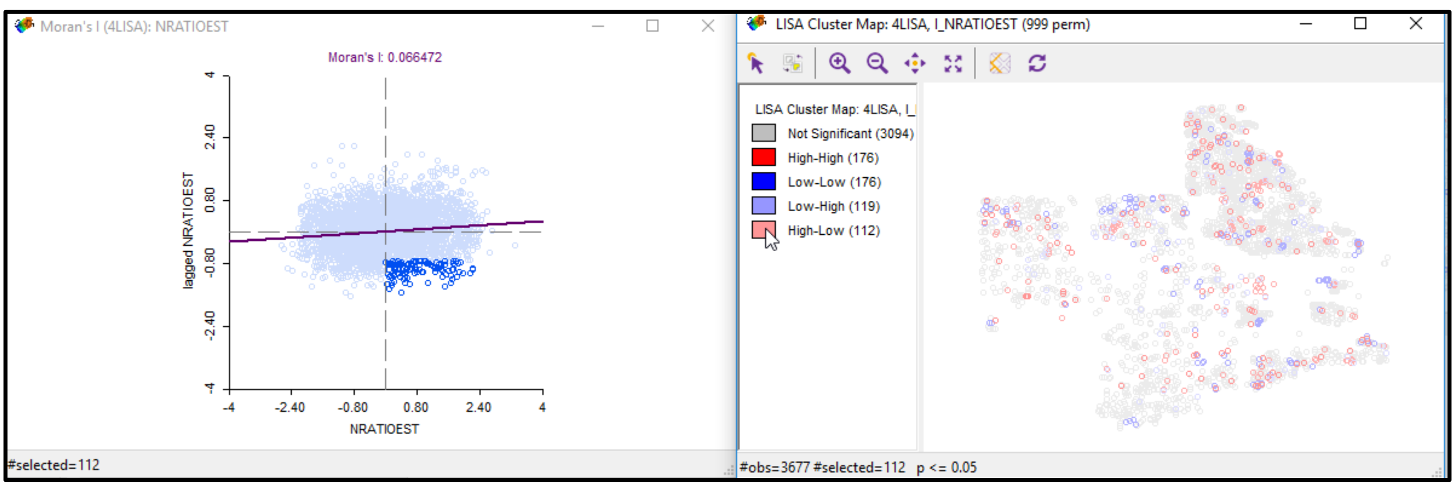
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.



North

Summary

- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- Owner Occupancy data was considered, but did not prove to be statistically significant
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Hyde Park was valued using the Multiple Regressions Analysis direct market comparison method of valuation
- Log linear models introduce what is called a retransformation bias
- The bias is corrected to ensure that the weighted mean ratio of estimated value to sale price is 1.000
- Performance statistics were within IAAO standards

The Data

Sales Counts

After filtering according to practices of the CCAO lister below, the starting number of sales used in the analysis was 1,110.

*select if (amount1>250,000).
*select if (amount1<5,000,000).
*select if (multi<1).
*select if sqftb<9000.
*select if (year1>2012).
*do not select if age<10 and (AMOUNT1<1600000 and (amount1/sqftb) <75 and class<95)

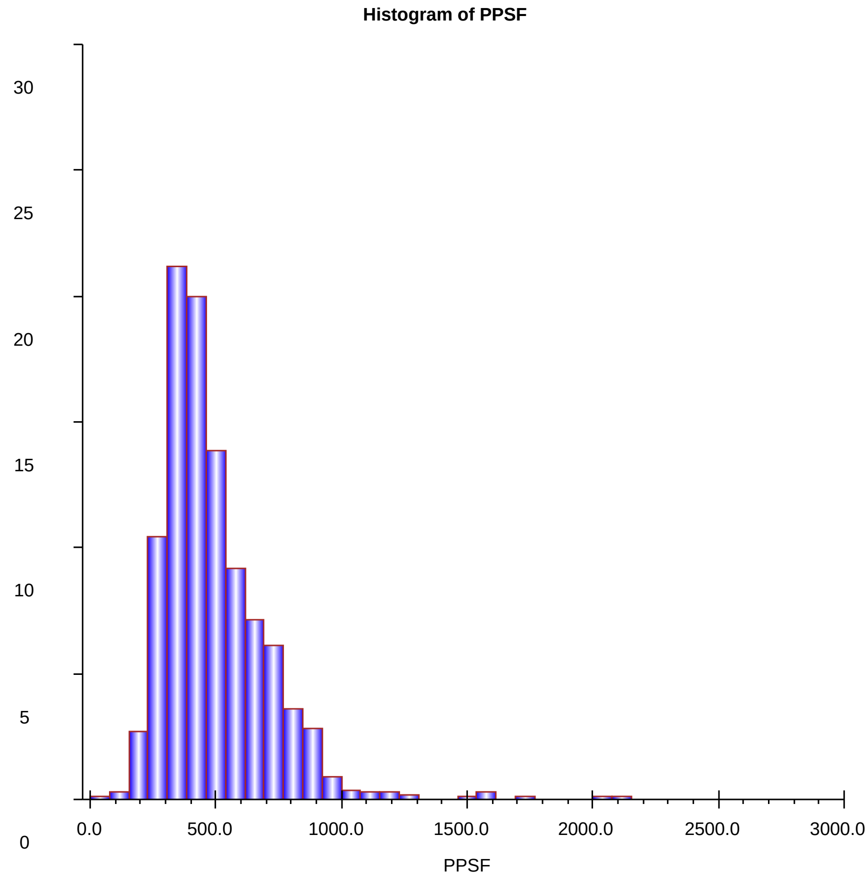
Data Fields

The initial list of data items available for analysis is provided in Appendix A Variable Definitions. Certain additional data fields were created. Those that were relevant to North Township are:

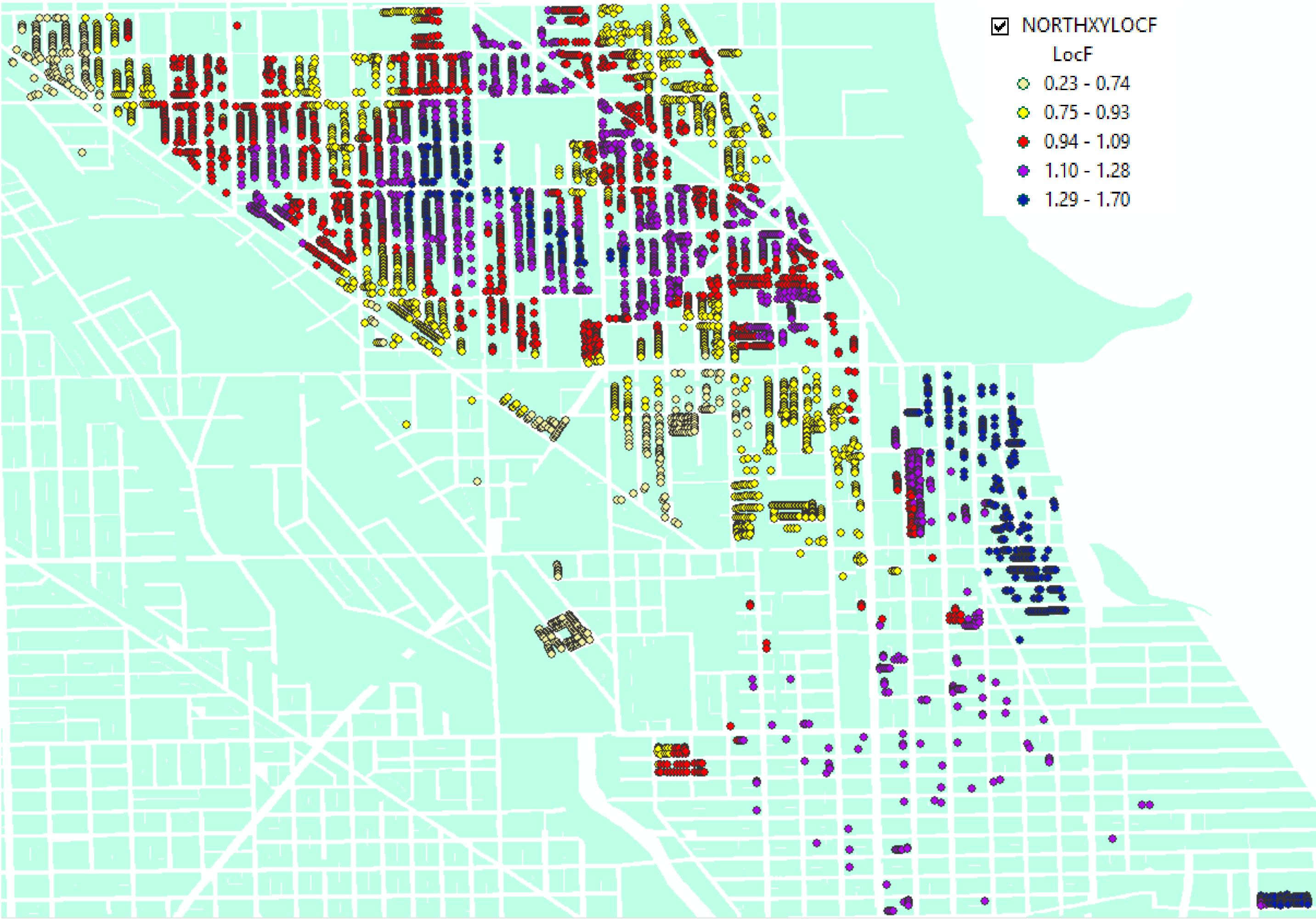
Location Factor

The process employed for determining the Location Factors for North Township is outlined in Appendix B Location Factorx

The histogram of price per square foot showed that there were some potential outliers. For purposes of developing the location factors and for subsequent analysis sevens sales with PPSF above $1,300/sqft were eliminated from further consideration.

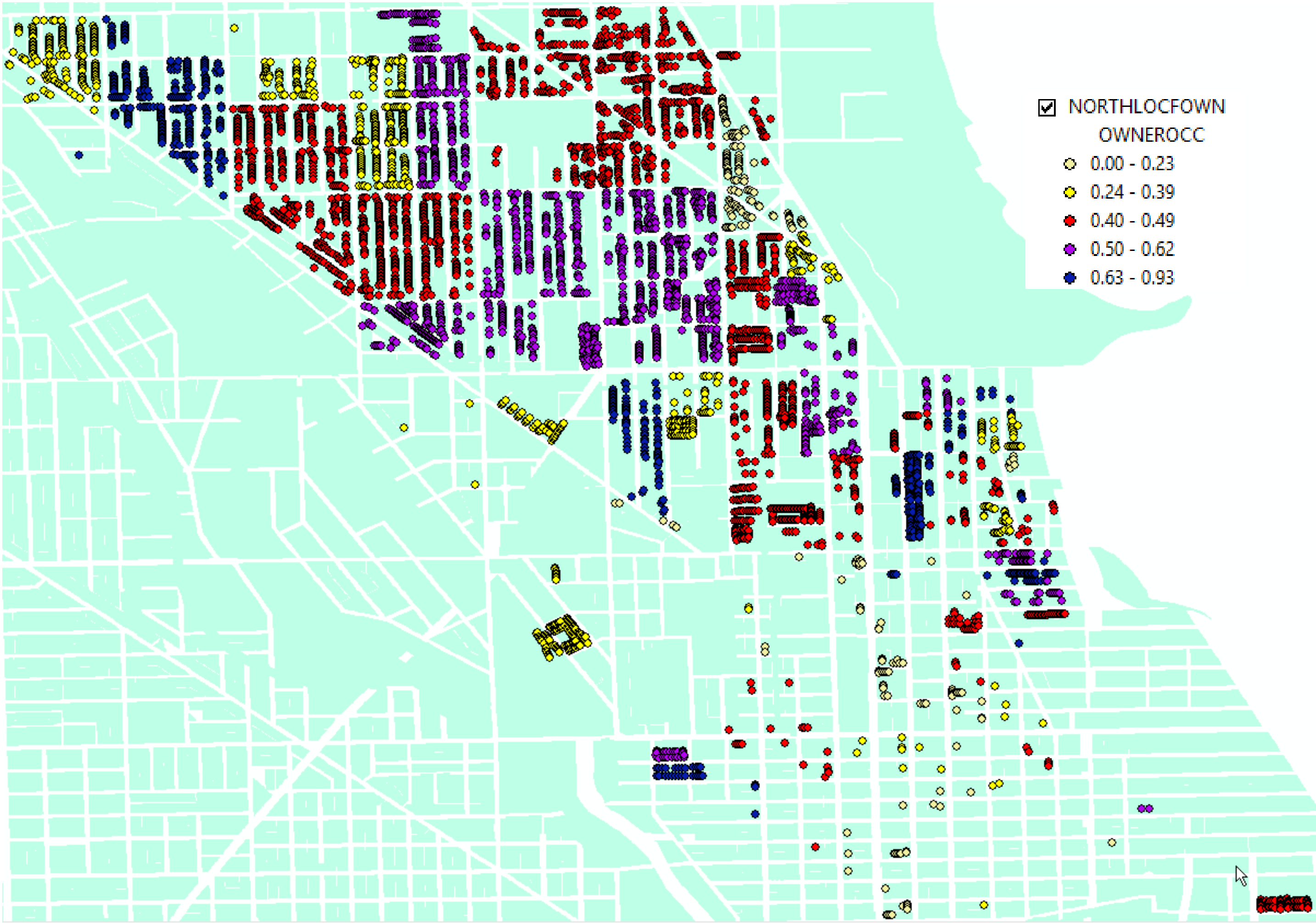


A thematic map of the location factor is presented in the next image. The location factor derived for the sale properties is shown below.



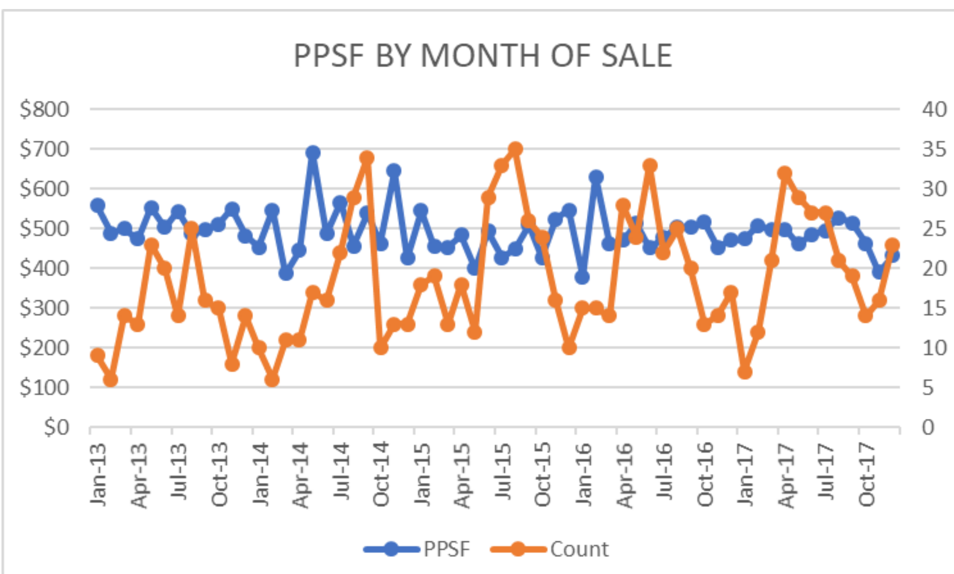
Owner Occupancy

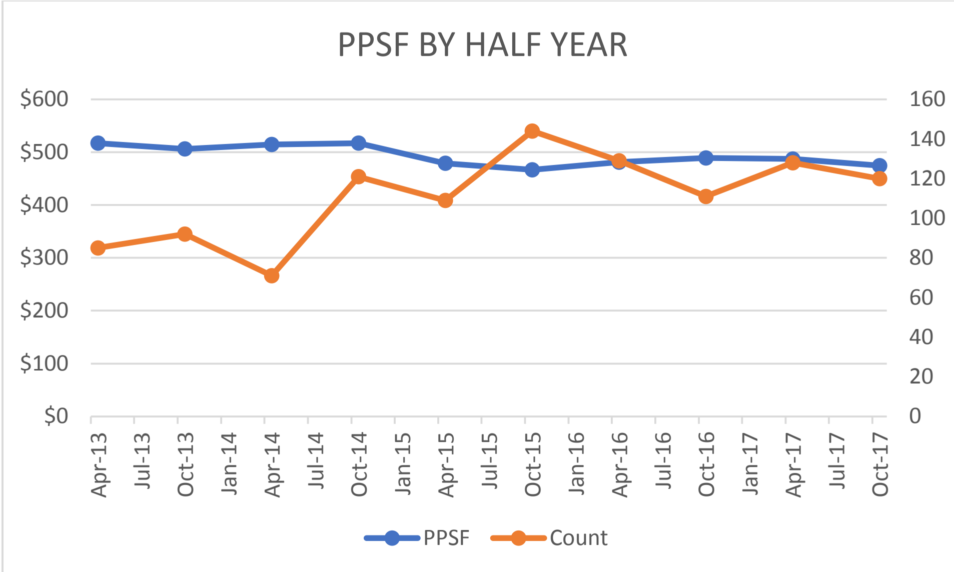
The process for obtaining the Owner Occupancy variable is described in Appendix C Owner Occupancy. The resultant variable is shown below.



Reverse Half of Sale

As was done for other townships in the Chicago Triad, the Reverse Half of Sale was chosen as the variable to express time dependency of value. The reason is that it is more stable than a monthly variable. The PPSF by Month chart is a bit too noisy to detect general trends. The PPSF by Half Year plot shows a slightly declining trend over the five-year period.



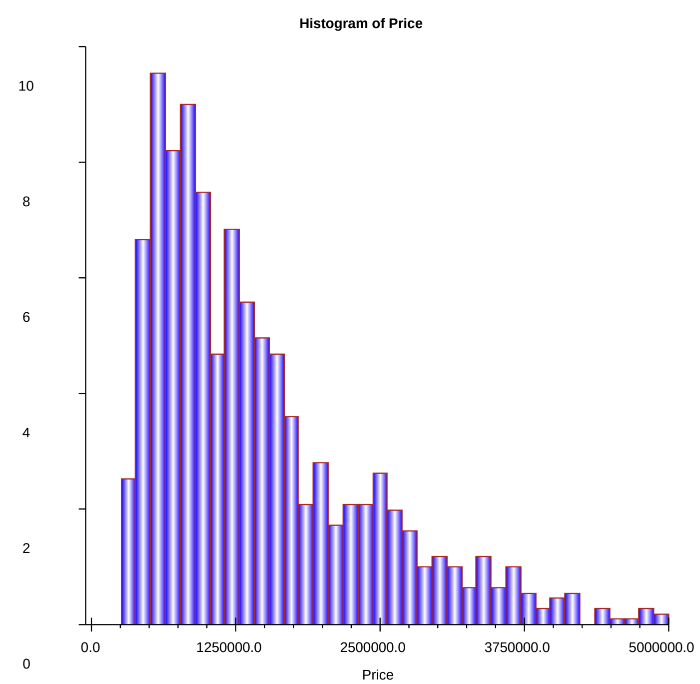


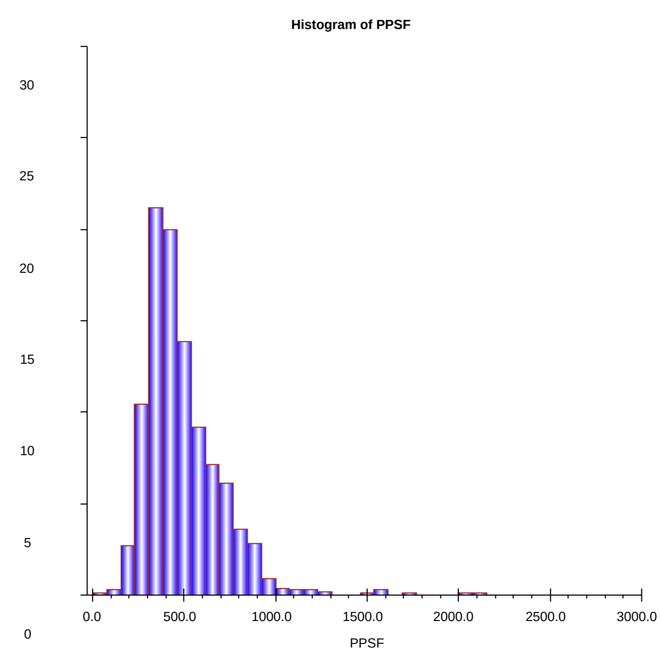
Exploratory Data Analysis

This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.

Price and Price per Square foot

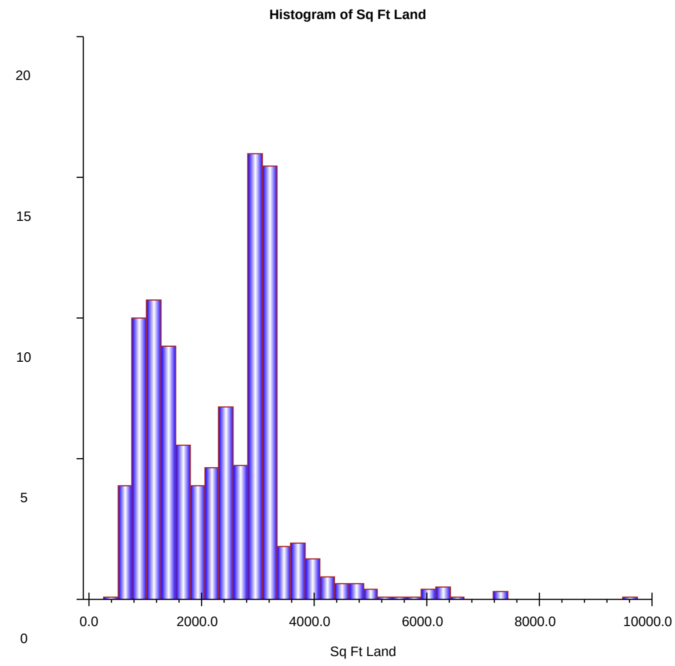
The price histogram fits the range of prices specified at the outset of modeling, namely a range of $250,000-$5,000,000. The price per square foot range indicates what are likely to be outlier situations. In other words, $1,400 per square foot and above is not likely to represent a true open market situation

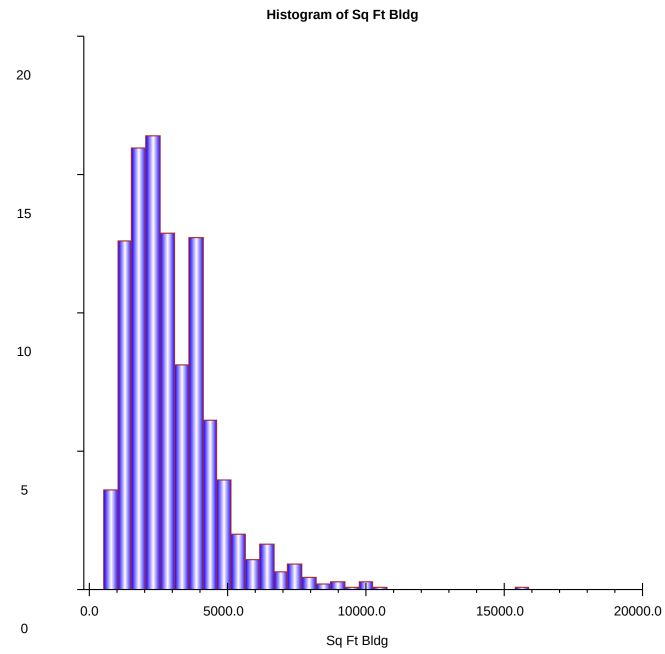




Square foot Land and Building

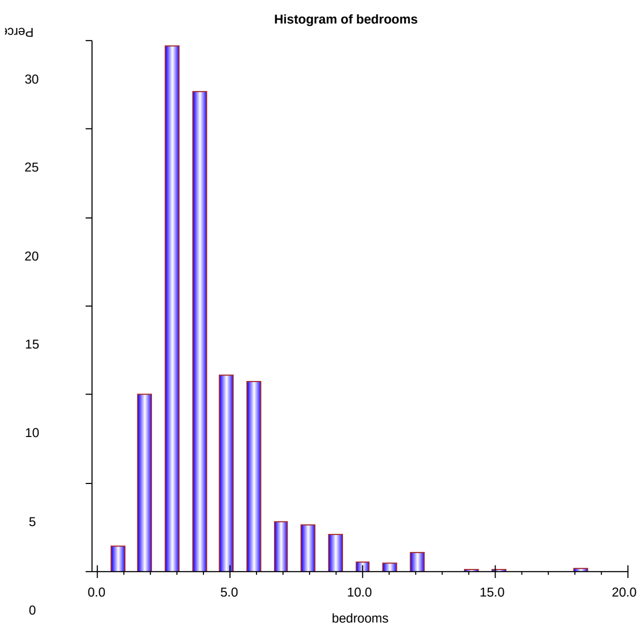
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not. It will be determined as part of the outlier detection process.

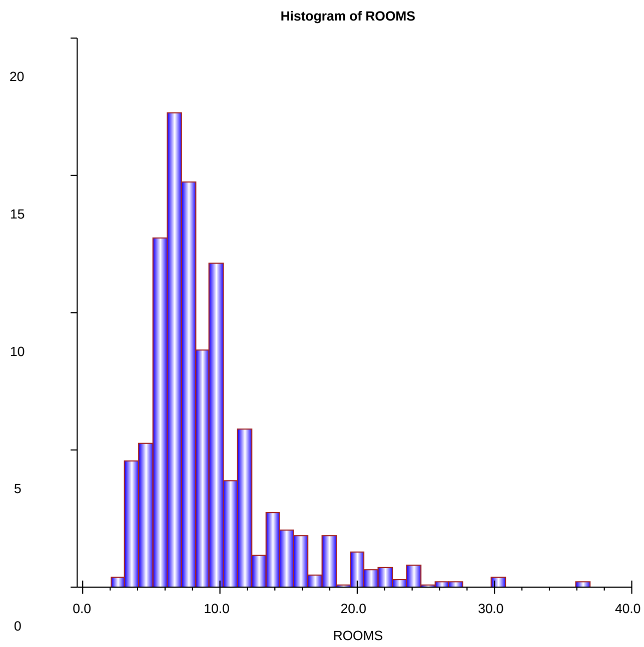




Rooms and Bedrooms

North Township sales are predominantly three- and four-bedroom dwellings, with a fair number below and above this range. The rooms count is reasonably consistent with the bedroom count the bedroom count and does not cause concern for significant numbers of outliers.





Model Structure and Calibration

Outlier Detection

The method for outlier detection is described in detail in Appendix D Outlier Detection. After following that process the sales used for the analysis became 964 of the original 1,103 (12.5%). This is one of the lowest percent outliers in the triad.

Structure

Note: the results below are from preliminary versions of the models. No outliers have been removed. The final model has better performance statistics than shown here.

As in other townships, both the additive and multiplicative forms of the model were evaluated. Considering the results for COD, PRD and PRB, the multiplicative form of the model was used.

Structure	Count	Median	WgtMean	COD	PRD	PRB
Additive	1,103	1.022	1.000	24.601	1.070	-0.009
Multiplicative	1,103	1.002	0.980	20.057	1.061	-0.033

Retransformation Bias

The reader will note that the median and weighted mean for the multiplicative model are lower than that for the additive model. This is a result of what is called the "retransformation bias". This topic is discussed in more detail in Appendix E Retransformation Bias. The final values are corrected for this characteristic of the multiplicative model calibration process.

Location Factor and Owner Occupancy

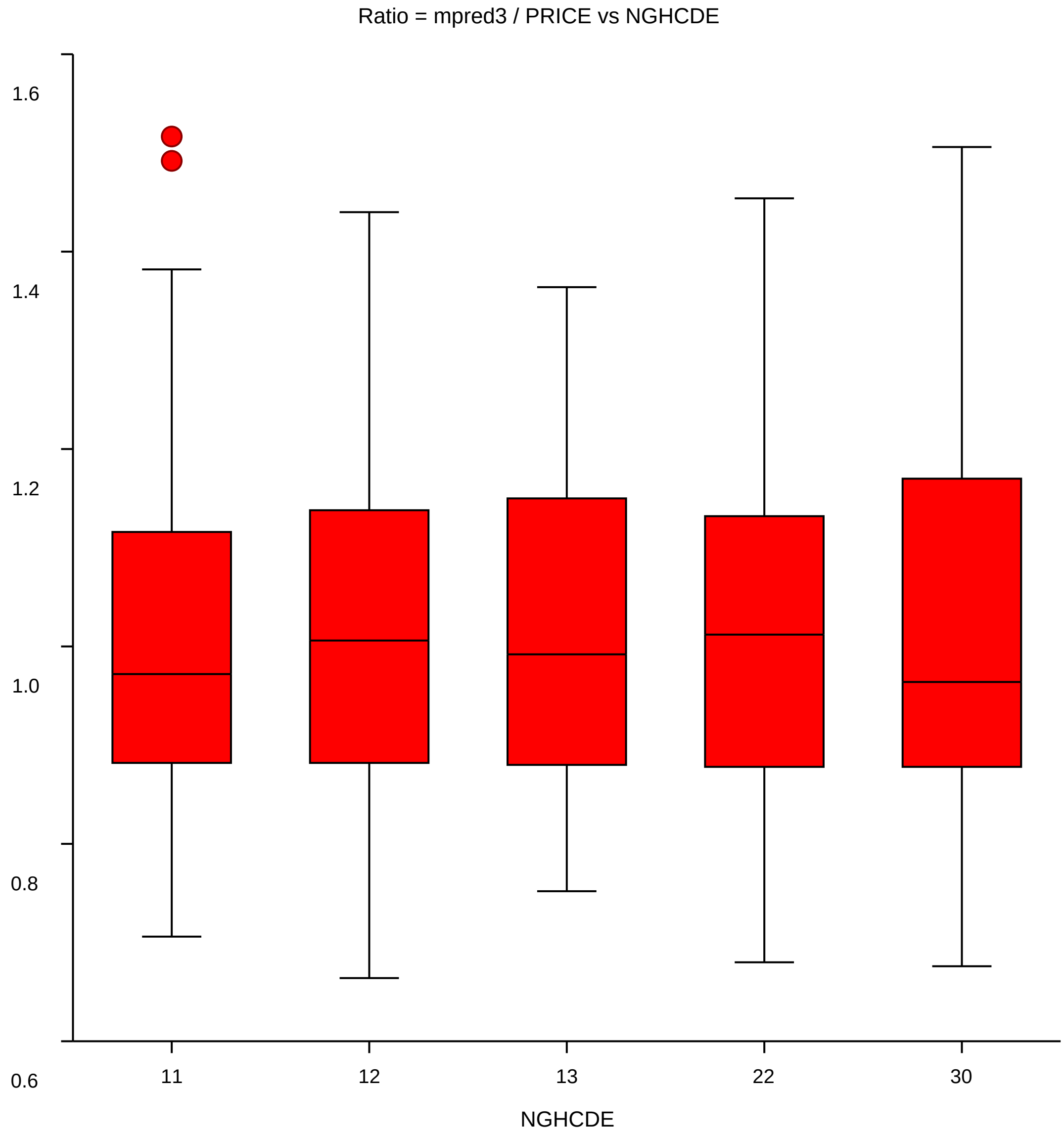
Only the Location Factor variable was statistically significant and important in North Township. This is seen in the next section, The final MRA Model.

The final MRA Model

Variable	Coefficient	T-Value
Intercept	7.30153	35.276
lnSFL	0.29906	13.958
lnSFB	0.57877	21.391
lnAGE	-0.14924	-10.606
lnLOCF	0.46196	14.377
lnFIXT	0.16753	5.287
(RHOS2=1)	0.07567	4.050
(RHOS3=1)	0.06667	3.279
(RHOS7=1)	-0.04224	-2.220
(RHOS8=1)	-0.04587	-1.758
(RHOS9=1)	-0.04413	-2.039
(RHOS10=1)	-0.10571	-4.658
(NB12=1)	-0.04953	-2.565
(NB13=1)	0.12194	3.506
(NB22=1)	0.07006	2.562
(num2=1)	-0.10505	-3.473
(num3=1)	-0.23232	-6.460
(num5=1)	-0.26091	-5.411
(num6=1)	0.45260	5.857
(extcon2=1)	0.11598	4.660
(extcon3=1)	0.04681	1.470
(bsfn3=1)	-0.04949	-3.506
(airc2=1)	-0.04015	-2.190
(renov1=1)	0.15352	3.332
(CL3=1)	-0.31795	-3.617
(CL4=1)	-0.74481	-5.102
(CL5=1)	-0.22224	-2.716
(CL6=1)	-0.16673	-2.168
(CL7=1)	-0.31255	-2.646
(CL8=1)	-0.31614	-3.766
(CL9=1)	-0.36693	-4.265
(CL10=1)	-0.18647	-2.400
(CL34=1)	-0.41884	-2.188
(CL78=1)	-0.42552	-5.278
(CL95=1)	-0.62189	-7.876

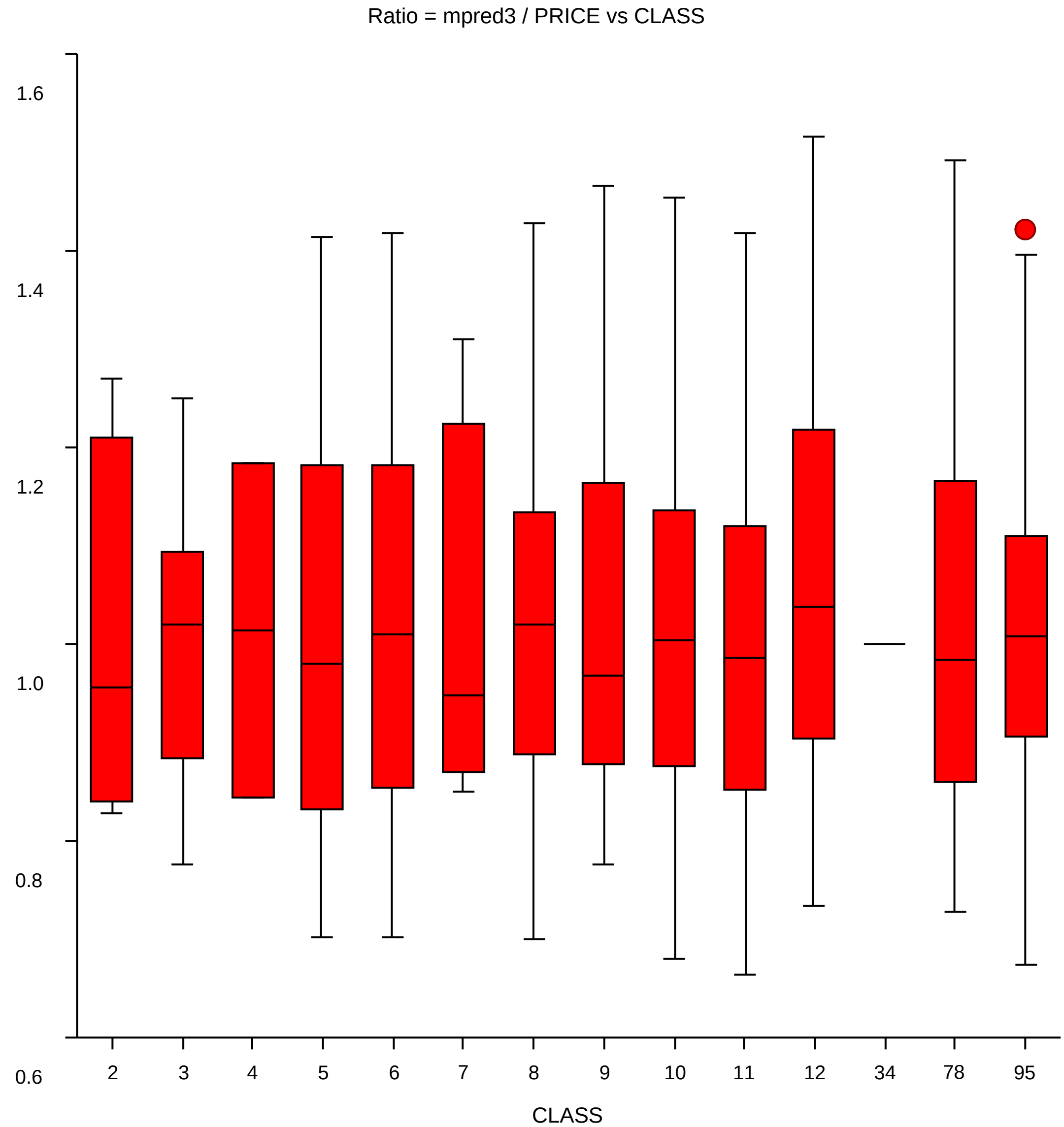
Model Performance Stats by NBHD

NGHCDE	Count	Median	WgtMean	COD	PRD	PRB
11	106	0.971	1.012	14.811	1.005	0.042
12	675	1.005	0.981	14.090	1.035	-0.033
13	44	0.993	1.002	13.431	1.011	0.017
22	101	1.012	0.978	14.099	1.038	-0.078
30	38	0.964	0.995	16.038	1.018	0.002
Combined	964	1.003	0.984	14.191	1.031	-0.026


Ratio = mpred3 / PRICE vs NGHCDE
1.6
1.4
1.2
1.0
0.8
0.6
11
12
13
22
30
NGHCDE

Model Performance Stats by Class

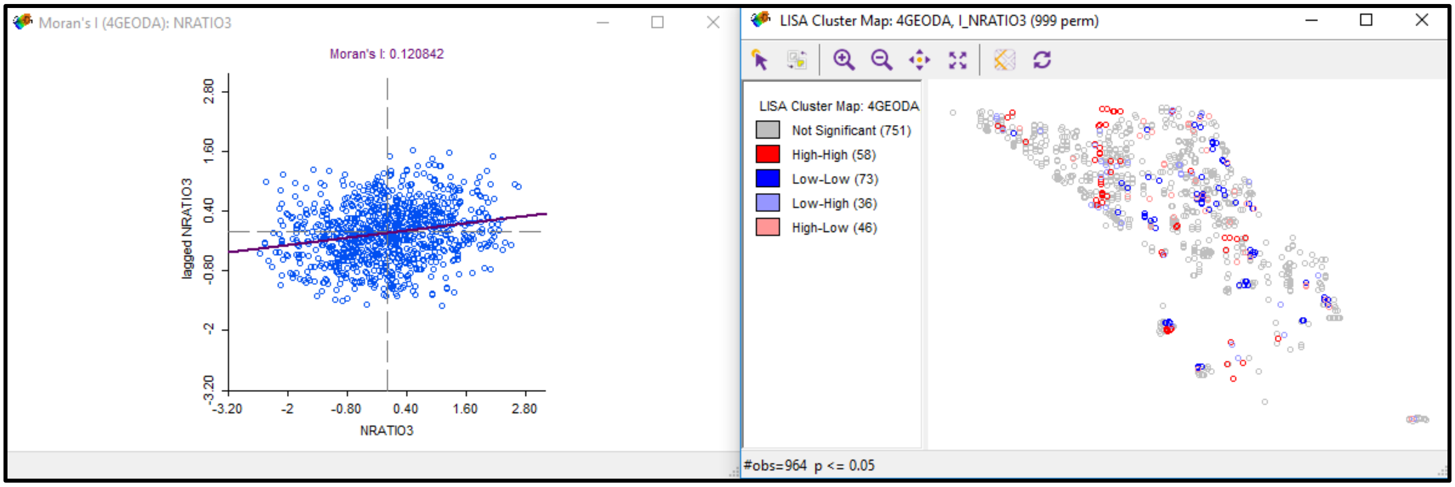
CLASS	Count	Median	WgtMean	COD	PRD	PRB
2	4	0.956	0.988	15.486	1.015	-0.010
3	13	1.020	0.990	9.637	1.019	-0.077
4	2	1.014	0.943	16.710	1.076	-0.318
5	24	0.979	0.974	18.646	1.050	-0.412
6	100	1.011	0.985	15.817	1.033	-0.124
7	4	0.948	0.975	13.013	1.040	-0.479
8	71	1.020	0.989	13.455	1.025	-0.199
9	27	0.967	0.977	14.513	1.039	-0.277
10	80	1.004	0.988	16.205	1.033	-0.059
11	164	0.985	0.969	15.608	1.035	-0.100
12	35	1.037	1.041	16.765	1.044	-0.112
34	1	1.000	1.000	0.000	1.000	0.000
78	94	0.983	0.968	15.672	1.049	-0.228
95	345	1.008	0.988	11.802	1.023	-0.040
Combined	964	1.003	0.984	14.191	1.031	-0.026



Spatial Dependency

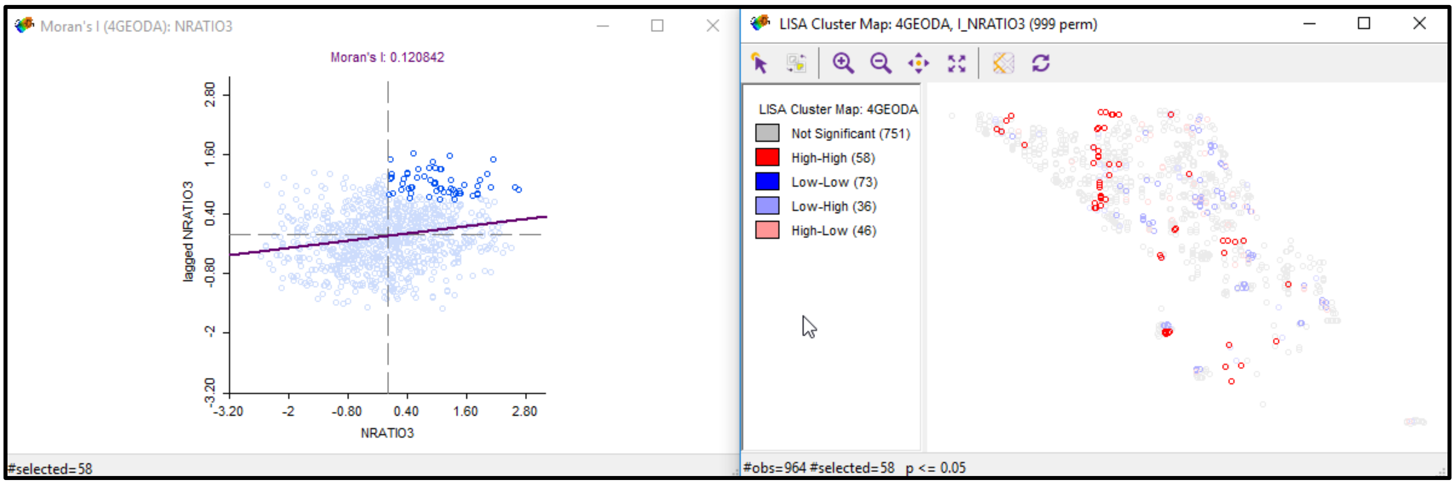
Additional Discussion of Spatial Dependency is provided in Appendix G Spatial Dependency.docx.

The Moran's I global statistic of 0.121 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of North, the statistic is close to 0.0. It is however, a bit higher than for other townships in the triad.



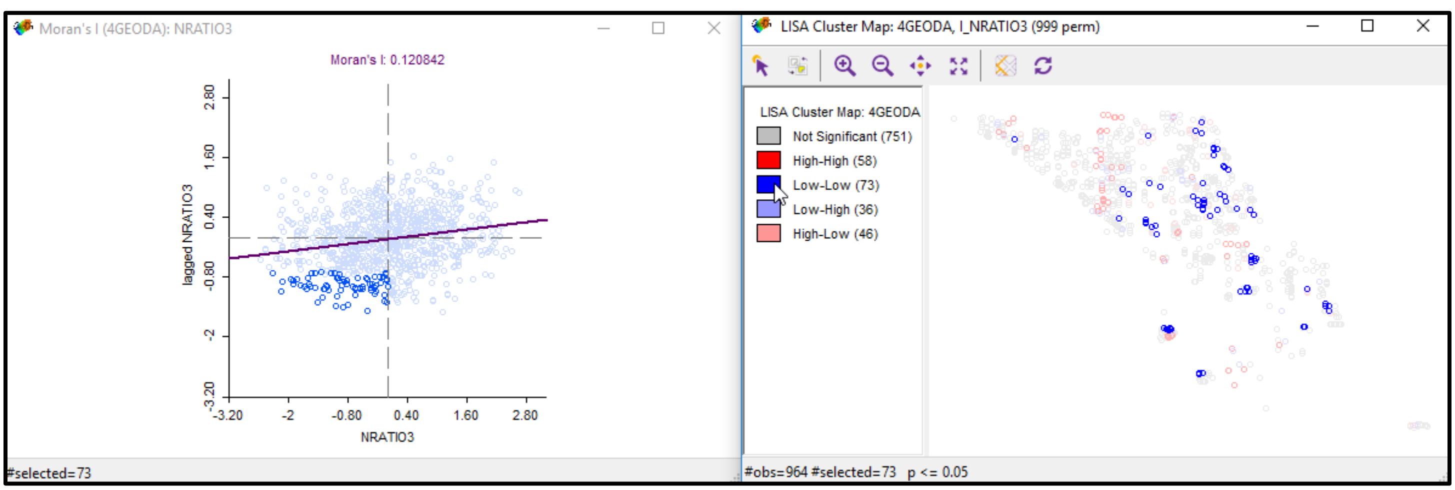
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



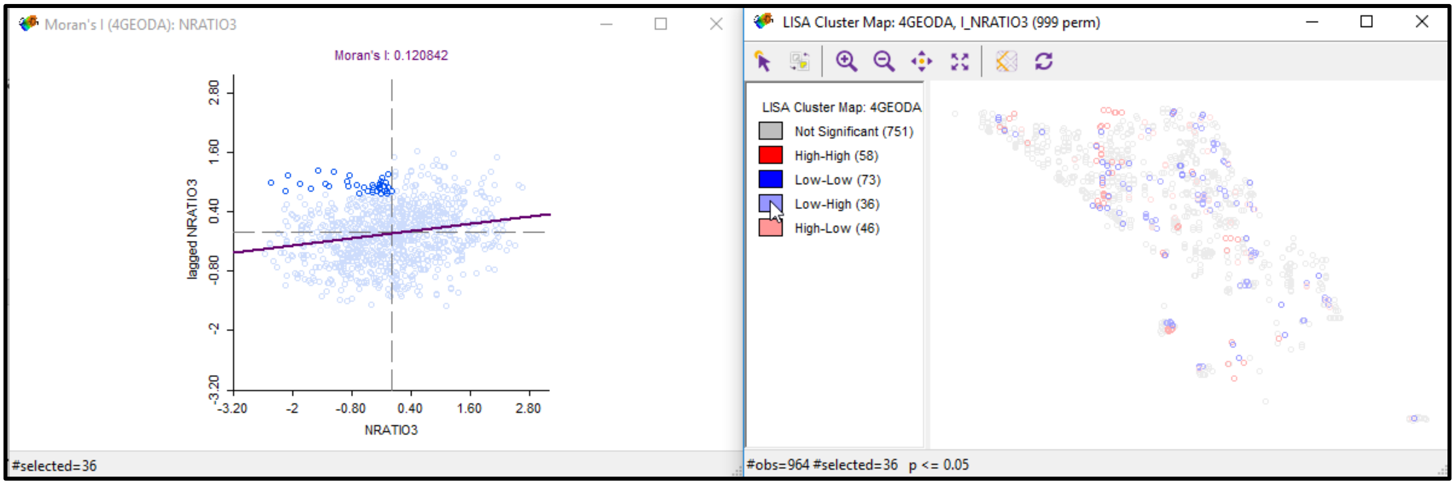
Low near Low Ratios

The low near low ratios are also spread uniformly around the town.



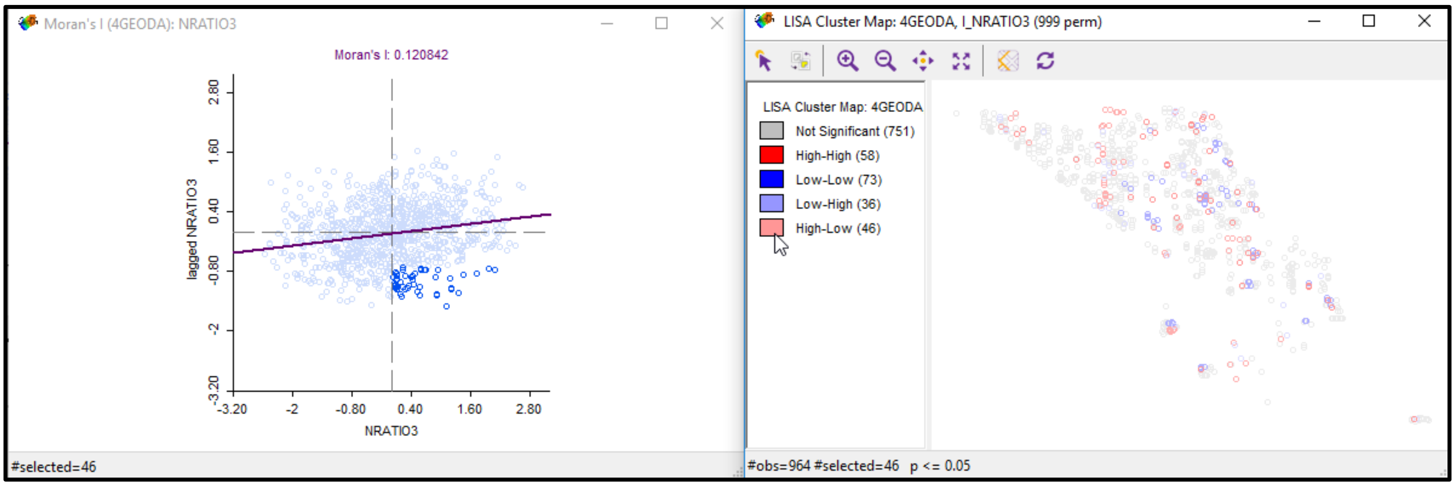
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.



South

Summary

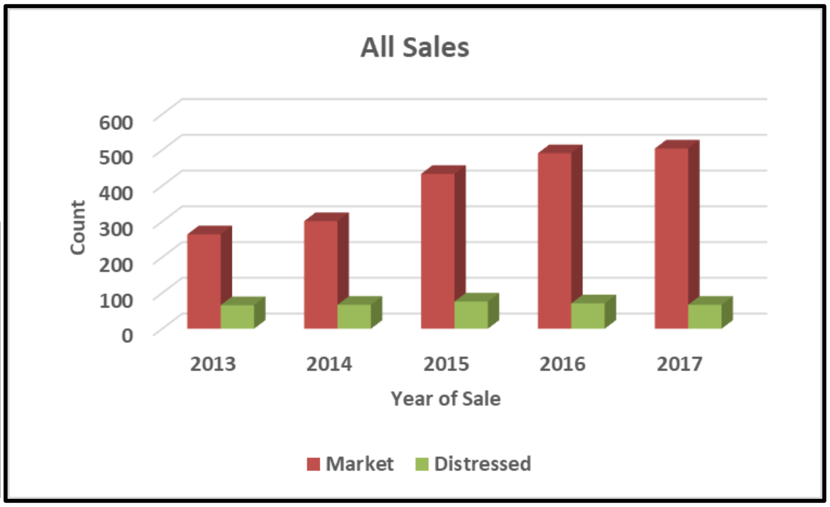
- Both linear additive and multiplicative (aka log-linear) model structures were evaluated
- The multiplicative model structure was chosen because of its superior performance measures
- Statistically-based methods of outlier removal were employed
- Geostatistical methods were used to derive a location influence factor used to improve model performance
- The Location Factor variable was statistically significant and contributed to an improved set of performance statistics for the final multiple regression model
- Geospatial analytic methods were used to ensure that there was no spatial bias in the valuation model
- The measures of potential spatial bias showed no clusters of overassessment or underassessment
- Hyde Park was valued using the Multiple Regressions Analysis direct market comparison method of valuation
- Log linear models introduce what is called a retransformation bias
- The bias is corrected to ensure that the weighted mean ratio of estimated value to sale price is 1.000

The Data

Sales Counts

The analysis and valuation of South Township was conducted in a slightly different way than the other seven townships. The analysis started with no filtering of sales using the CCAO filtering methodology. Instead all sales including open market and distressed were used at the start of the analysis. The five-year total was 2,340, with 1,993 open market and 347 distressed. The details are present in tabular and graphic form below.

Year	Market	Distressed	Grand Total
2013	264	66	330
2014	301	67	368
2015	433	76	509
2016	491	71	562
2017	504	67	571
Grand Total	1,993	347	2,340



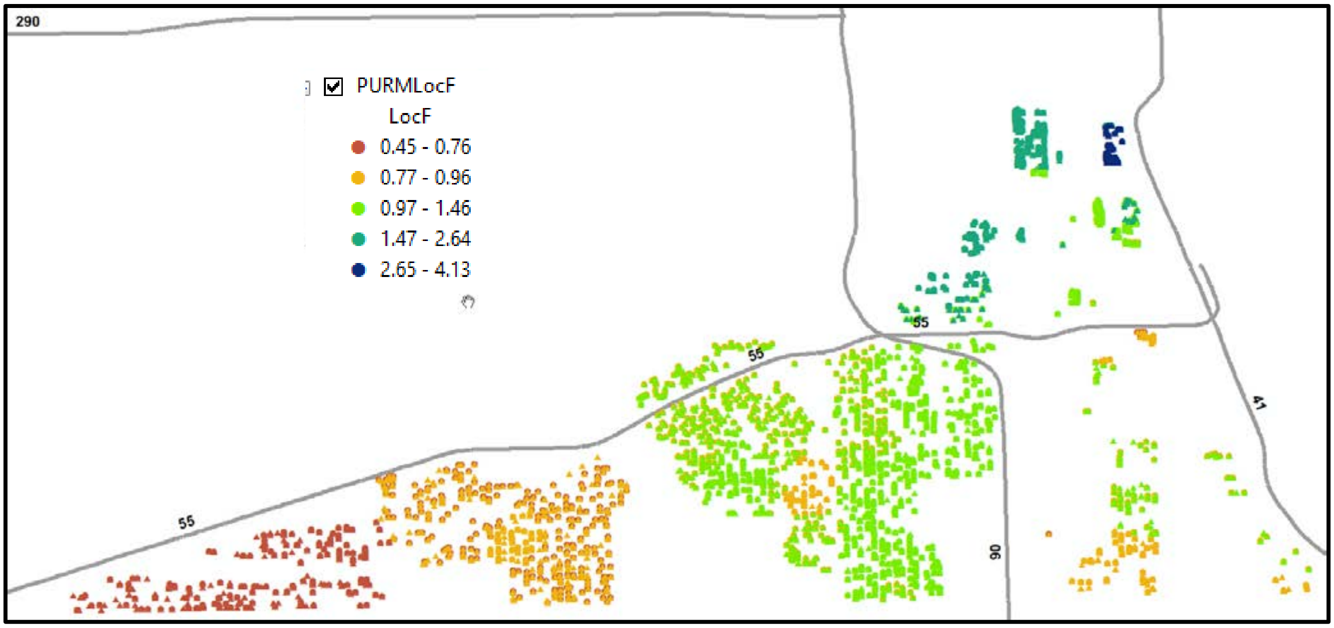
Data Fields

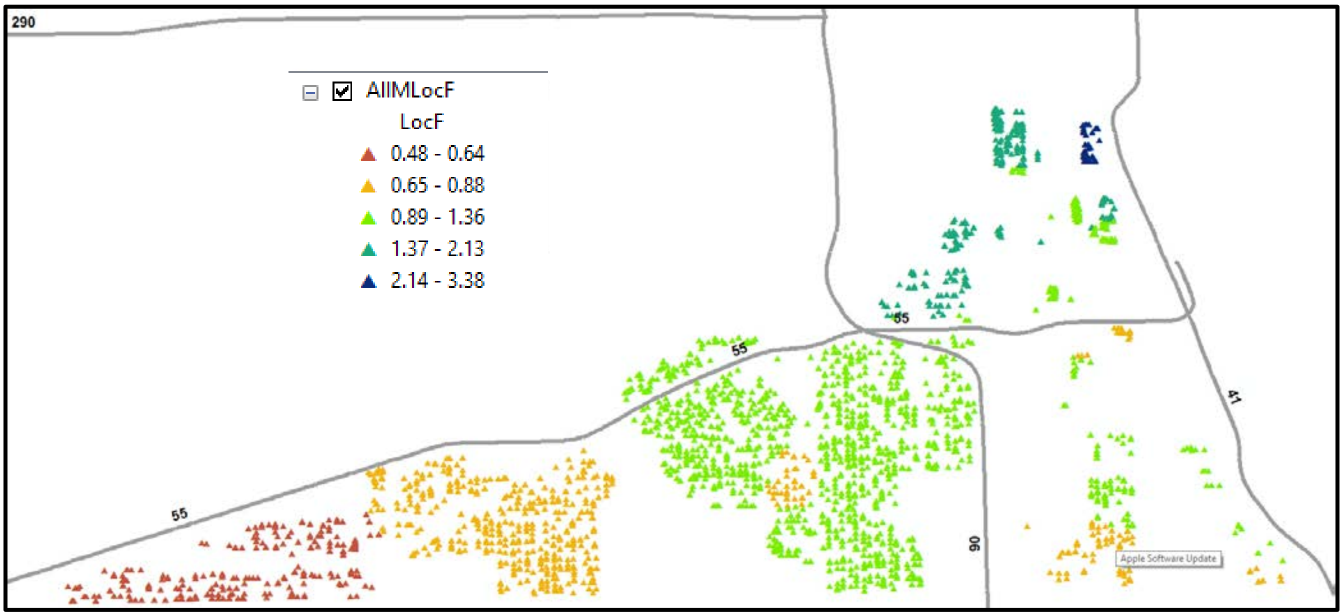
The initial list of data items available for analysis is provided in [Appendix A Variable Definitions](). Certain additional data fields were created. Those that were relevant to South Township are:

Location Factor

The process employed for determining the Location Factors for South Township is outlined in [Appendix B Location Factor.]()

The difference in the process for South is that two location factor variables were developed, one for pure market sales only and one for all sales. The first, of the two images below, is for the pure market transactions. The second includes all sales. The pure market theme shows a wider range of the location factor than does the theme for all sales. The likely reason for the difference is that the inclusion of the distressed sales provides a lower average sale price in certain areas.



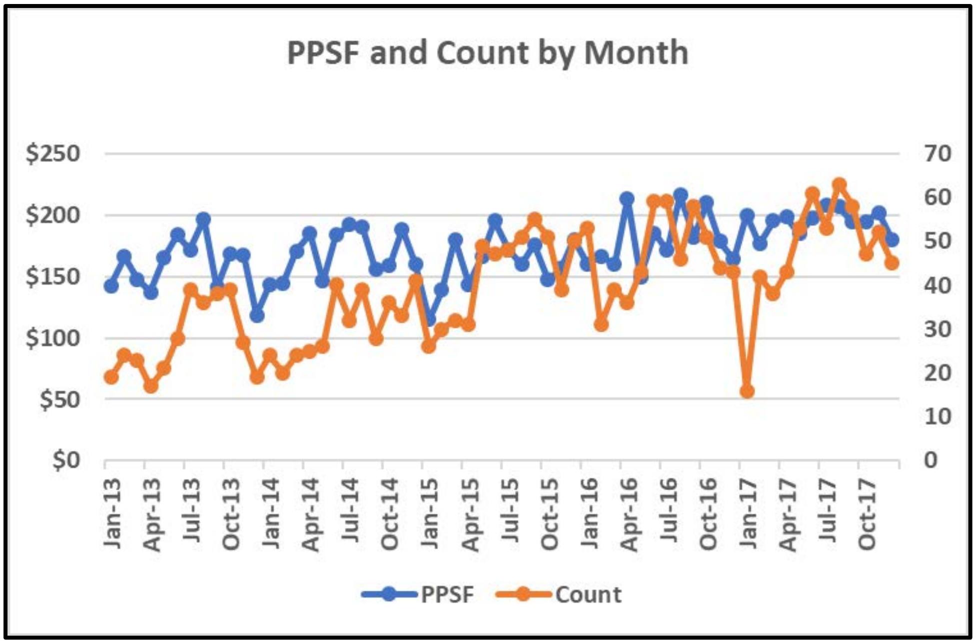


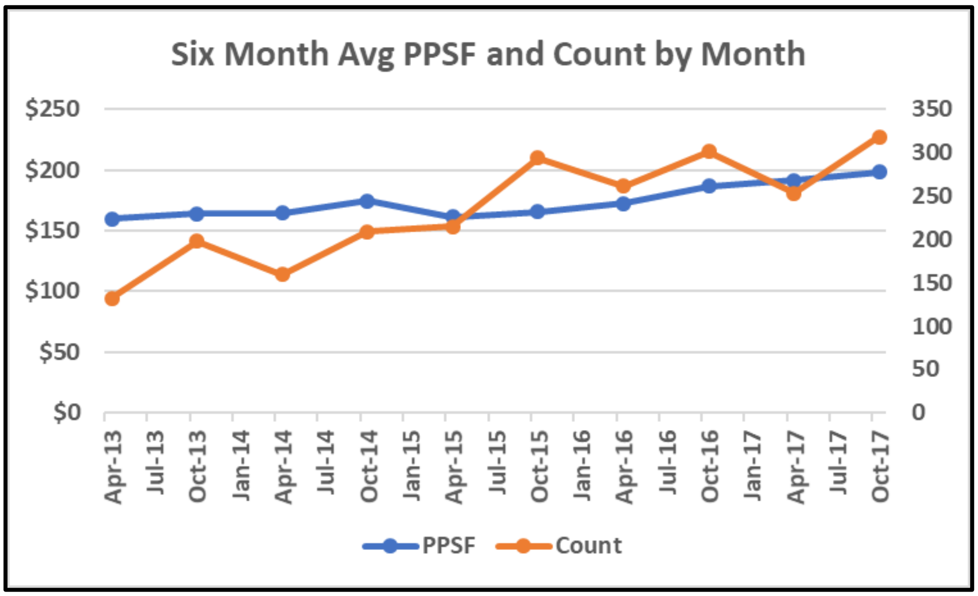
Owner Occupancy

Owner Occupancy was not pursued in the analysis for this township.

Reverse Half of Sale

The price per square foot (PPSF) and count by month in the first chart below are quite noisy. The second chart is a six-month average of PPSF and count which has a much easier to discern pattern, namely both count and PPSF are increasing over time. The six-month increment in sale month variable was chosen for the analysis.





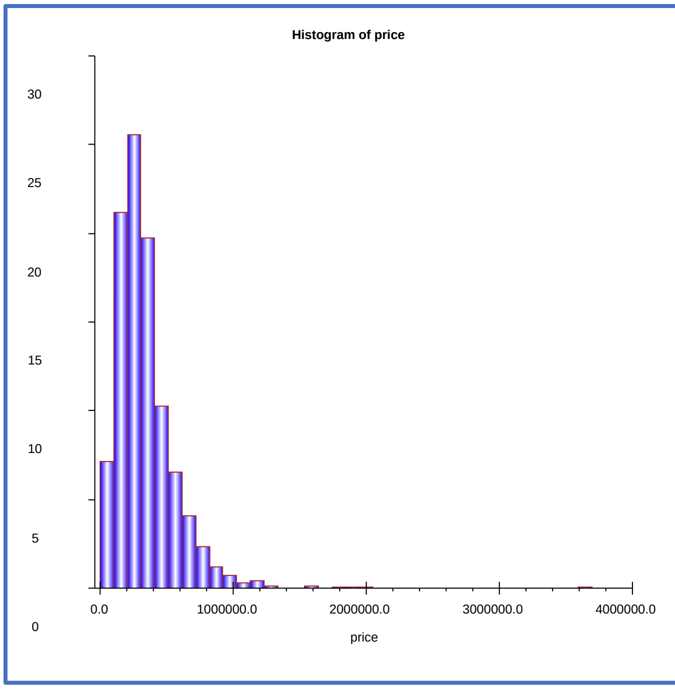
Exploratory Data Analysis

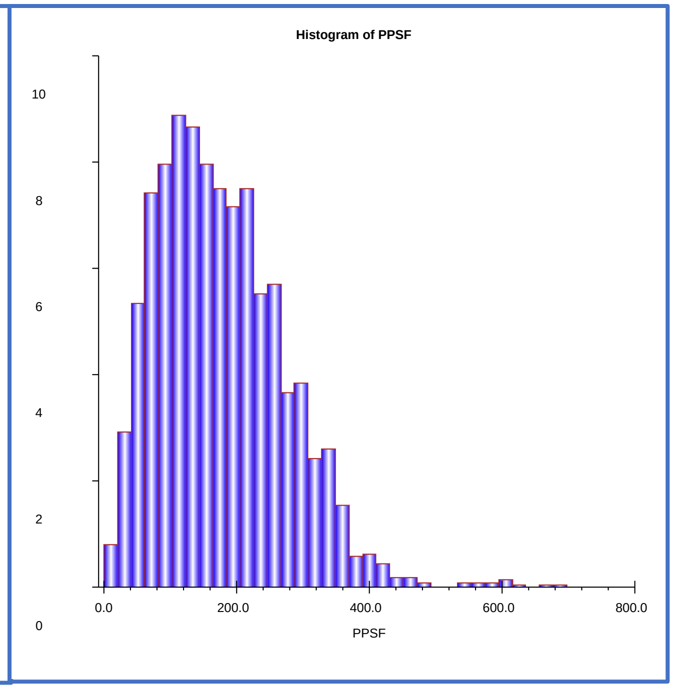
This phase of developing a mass appraisal model is called exploratory data analysis (EDA). One of the better methods of EDA is the histogram. The histogram helps isolate issues, if any, that may hamper the model calibration process. The data shown is before outliers are removed. Selected variables are examined herein.[1]

Price and Price per Square foot

The price per square foot range indicates what are likely to be outlier situations. It is possible that sales at over $500/sq. ft. may not represent a true open market situation.

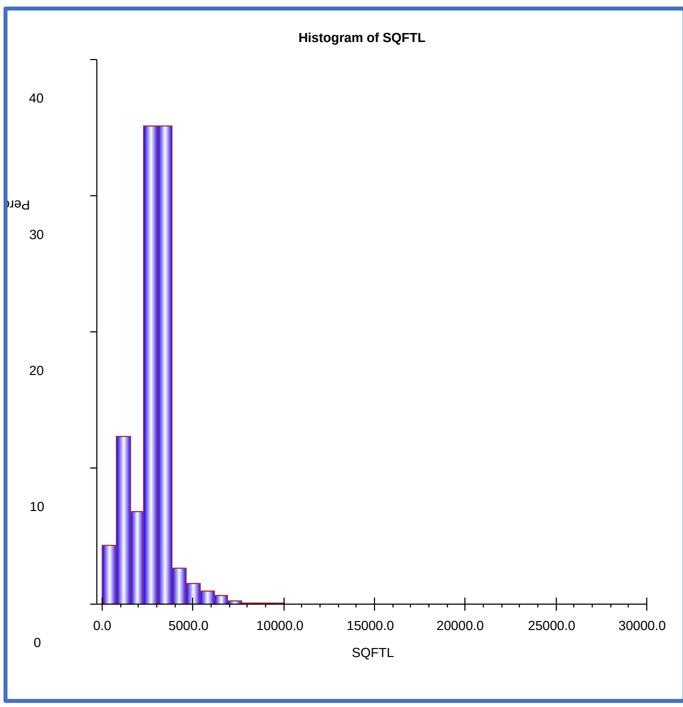
[1] The initial pass creating the histograms showed one very large lot, which caused the land size histogram to be nearly meaningless. What is presented has been filtered to shows lots up to 100,000 sq. ft, reducing the count by 1.

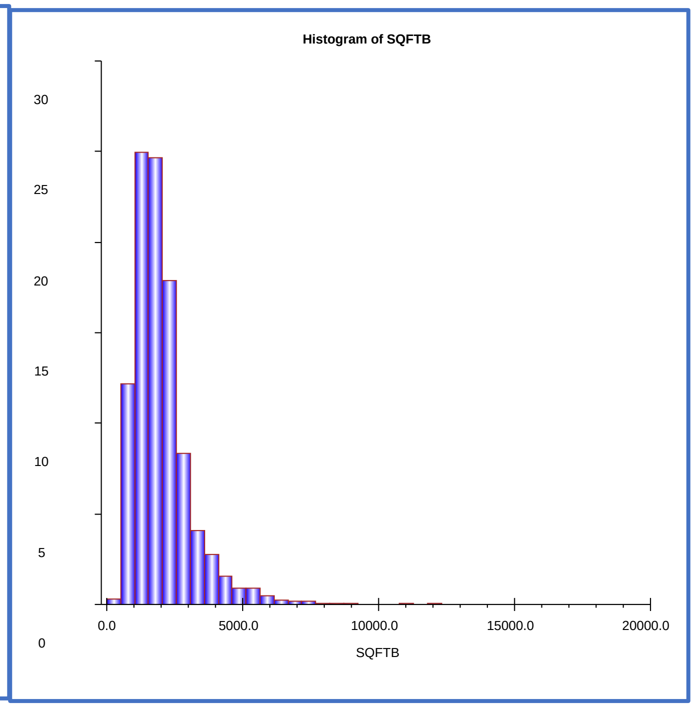




Square foot Land and Building

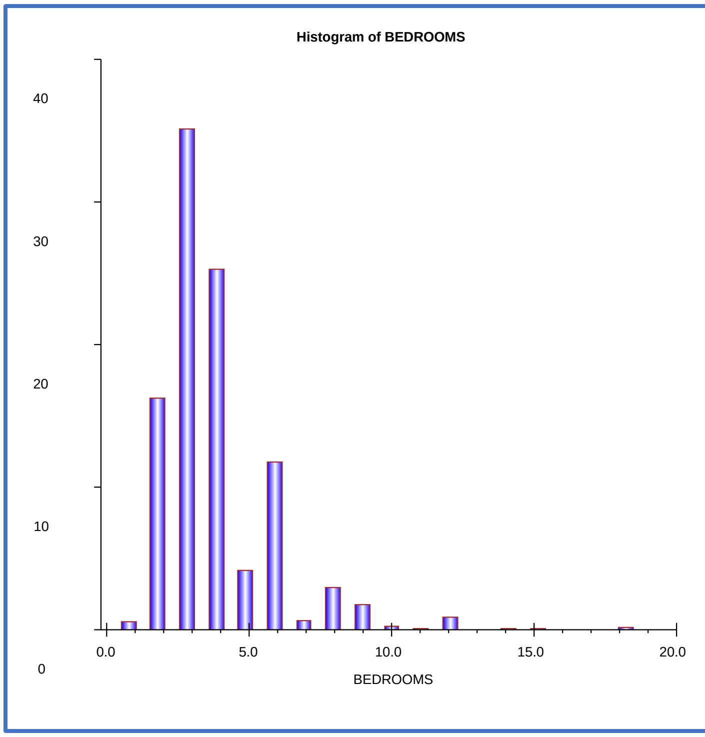
The high ends of both histograms are noted. At this stage of the investigation, it is too soon to know if these are outliers or not. It will be determined as part of the outlier detection process.

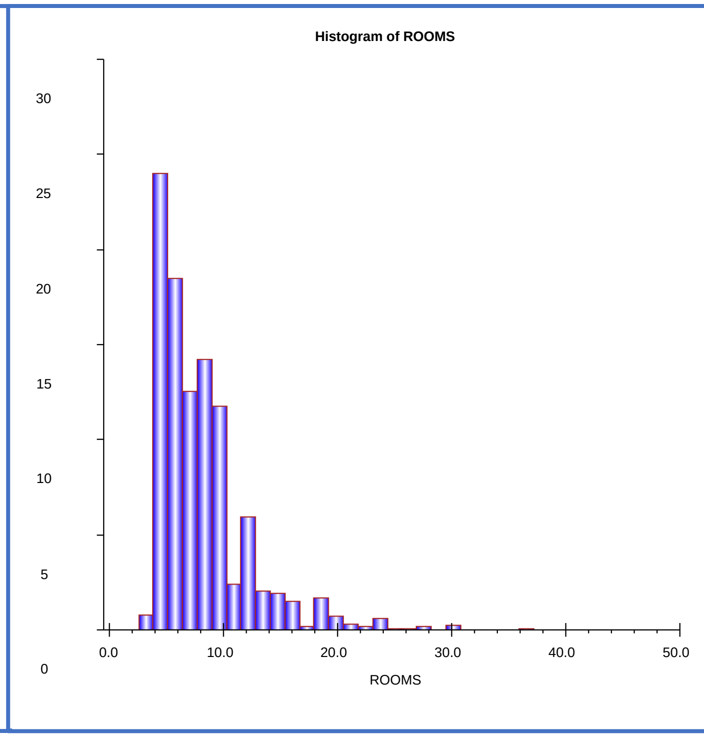




Rooms and Bedrooms

South Township sales are predominantly three- and four-bedroom dwellings, with a fair number below and above this range. The rooms count is reasonably consistent with the bedroom count the bedroom count and does not cause concern for significant numbers of outliers.





Model Structure and Calibration

Initially two versions of the analysis were carried forward, one including all sales and the other pure market sales only. The valuation results were very close. After reviewing with CCAO it was decided to work with pure market sales only.

Outlier Detection

The method for outlier detection is described in detail in Appendix D Outlier Detection. After following that process the sales used for the analysis became 1,600 of the original 1,993 (19.7%).

Structure

Note: the results below are from preliminary versions of the models. The counts are slightly different because the outliers are determined for each model structure. The final model has better performance statistics than shown here.

As in other townships, both the additive and multiplicative forms of the model were evaluated. Considering the results for COD, PRD and PRB, the multiplicative form of the model was used.

Structure	Count	Median	WgtAvg	COD	PRD	PRB
Additive	2,071	1.022	1.005	21.564	1.056	-0.048
Multiplicative	2,045	0.982	0.977	20.070	1.057	-0.107

Retransformation Bias

The reader will note that the median and weighted mean for the multiplicative model are lower than that for the additive model. This is a result of what is called the “retransformation bias”. This topic is discussed in more detail in Appendix E Retransformation Bias. The final values are corrected for this characteristic of the multiplicative model calibration process.

Location Factor and Owner Occupancy

Only the Location Factor variable was considered. It is statistically significant and important in South Township. This is seen in the next section, The final MRA Model.

The final MRA Model

Variable	Coefficient	T-Value
Intercept	8.52000	59.840
lnSFL	0.11624	7.767
lnAge	-0.16804	-28.833
lnSFB	0.48223	24.172
lnFIXT	0.12321	4.512
lnPMLocF	0.34164	11.976
(basment:	-0.14743	-8.913
(basment:	-0.04009	-2.311
(basment:	-0.15677	-4.936
(RH2=1)	-0.05590	-2.870
(RH3=1)	-0.06164	-3.411
(RH4=1)	-0.09819	-5.131
(RH5=1)	-0.14748	-7.922
(RH6=1)	-0.12861	-6.272
(RH7=1)	-0.16527	-8.119
(RH8=1)	-0.21041	-9.412
(RH9=1)	-0.24366	-11.459
(RH10=1)	-0.29520	-11.300
(RS2=1)	-0.10366	-9.151
(NB11=1)	0.66154	21.874
(NB12=1)	0.58896	10.089
(NB30=1)	0.24630	13.598
(NB40=1)	0.17463	4.477
(NB50=1)	0.15173	9.227
(gar3=1)	0.08128	7.586
(gar4=1)	0.14121	3.044
(gar6=1)	-1.99923	-10.300
(CL5=1)	0.19486	5.205
(CL6=1)	0.16252	2.962
(CL12=1)	-0.09606	-3.122
(num2=1)	-0.07930	-3.777
(num3=1)	-0.16067	-5.208

Model Performance Stats by NBHD

NGHCDE	Count	Median	WgtMean	COD	PRD	PRB
11	186	0.969	0.979	14.008	1.035	-0.296
12	30	0.996	0.985	8.555	1.021	-0.260
30	545	0.989	0.991	14.016	1.025	-0.075
40	30	0.997	0.994	9.124	1.013	-0.217
41	40	0.919	0.915	16.371	1.026	-0.014
42	22	0.814	0.836	18.053	1.040	-0.254
50	376	0.990	0.986	16.040	1.034	-0.093
60	371	1.007	0.997	18.195	1.048	-0.196
Combined	1600	0.988	0.984	15.492	1.035	-0.068

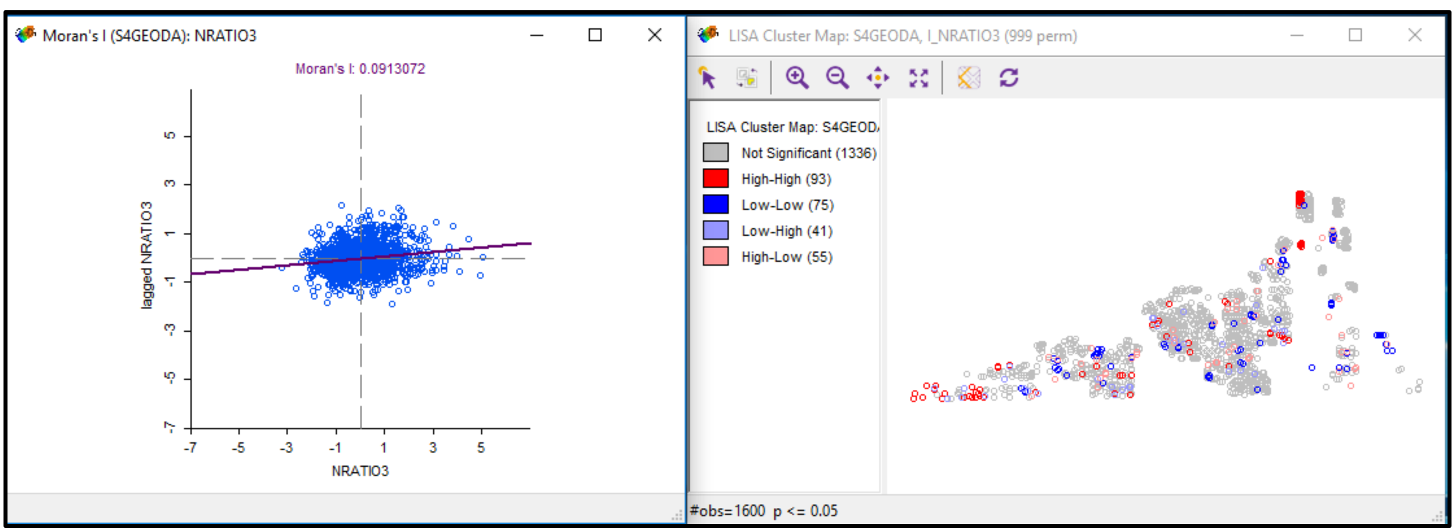
Model Performance Stats by Class

CLASS	Count	Median	WgtMean	COD	PRD	PRB
2	102	0.992	0.964	16.367	1.042	-0.246
3	231	0.951	0.957	13.913	1.022	-0.109
4	7	0.917	1.118	35.731	1.076	-1.014
5	28	0.994	0.969	15.211	1.053	-0.359
6	13	0.948	0.970	19.164	1.062	-0.360
7	71	0.964	0.970	10.563	0.995	0.135
8	4	0.822	0.804	10.061	1.061	-0.086
9	2	0.650	0.649	0.463	1.001	-0.011
10	10	0.789	0.775	18.297	1.042	-0.147
11	490	1.033	1.011	18.716	1.054	-0.181
12	54	0.983	0.993	19.526	1.033	-0.080
34	10	0.943	0.959	15.817	1.066	-0.371
78	166	0.978	0.977	11.885	1.027	-0.130
95	412	0.983	0.988	12.417	1.021	-0.029
Combined	1600	0.988	0.984	15.492	1.035	-0.068

Spatial Dependency

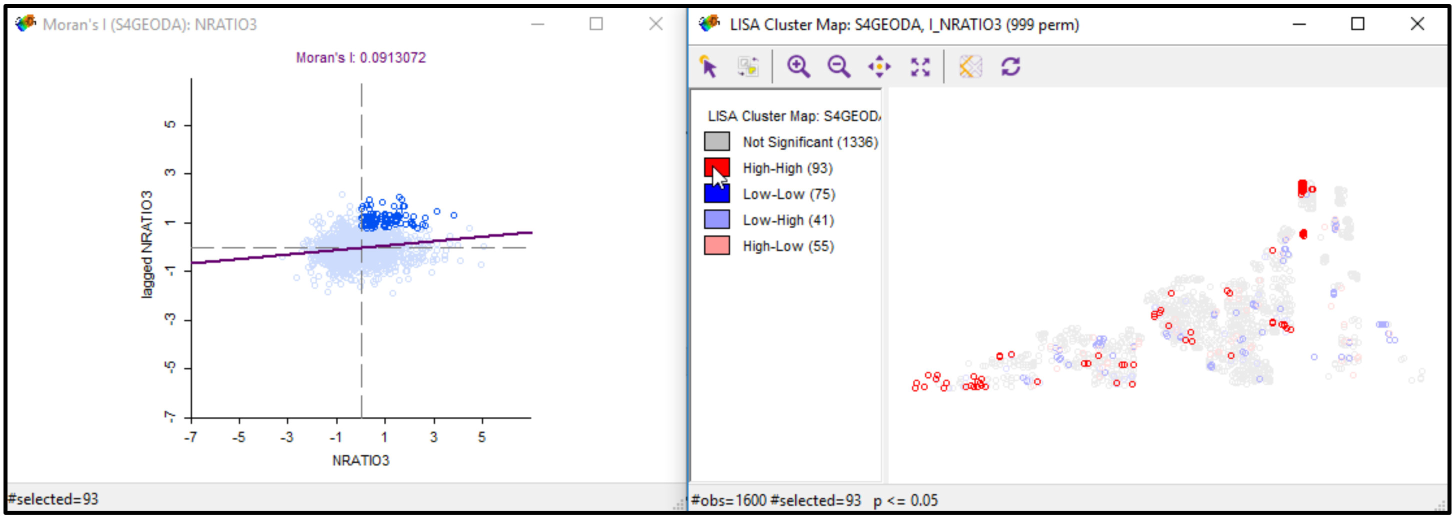
Additional Discussion of Spatial Dependency is provided in Appendix G Spatial Dependency.docx.

The Moran's I global statistic of 0.913 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of South Township, the statistic is close to 0.0. It is however, a bit higher than for other townships in the triad.



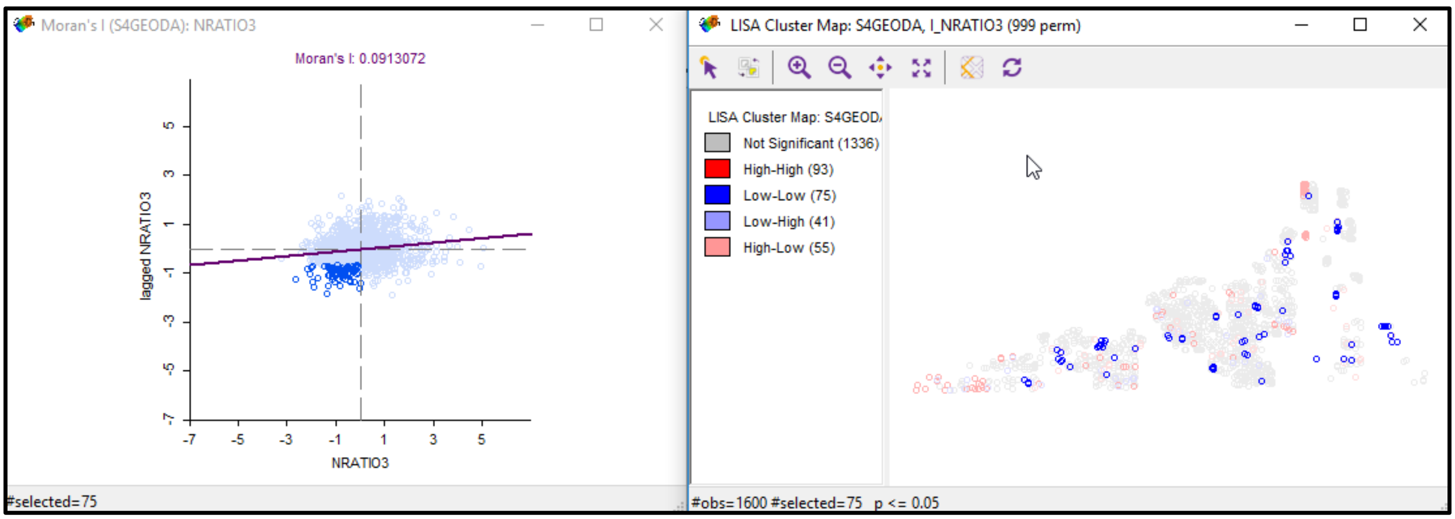
High near High Ratios

The pockets of high ratios near high ratios are geographically spread with no major bunching of points.



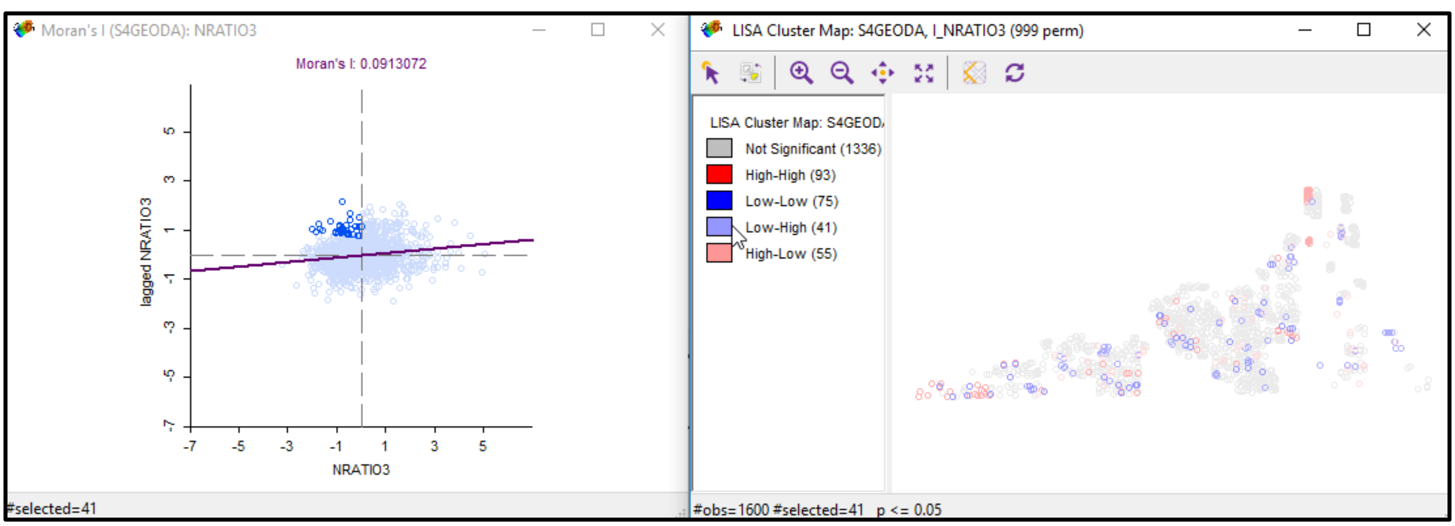
Low near Low Ratios

The low near low ratios are also spread uniformly around the township.



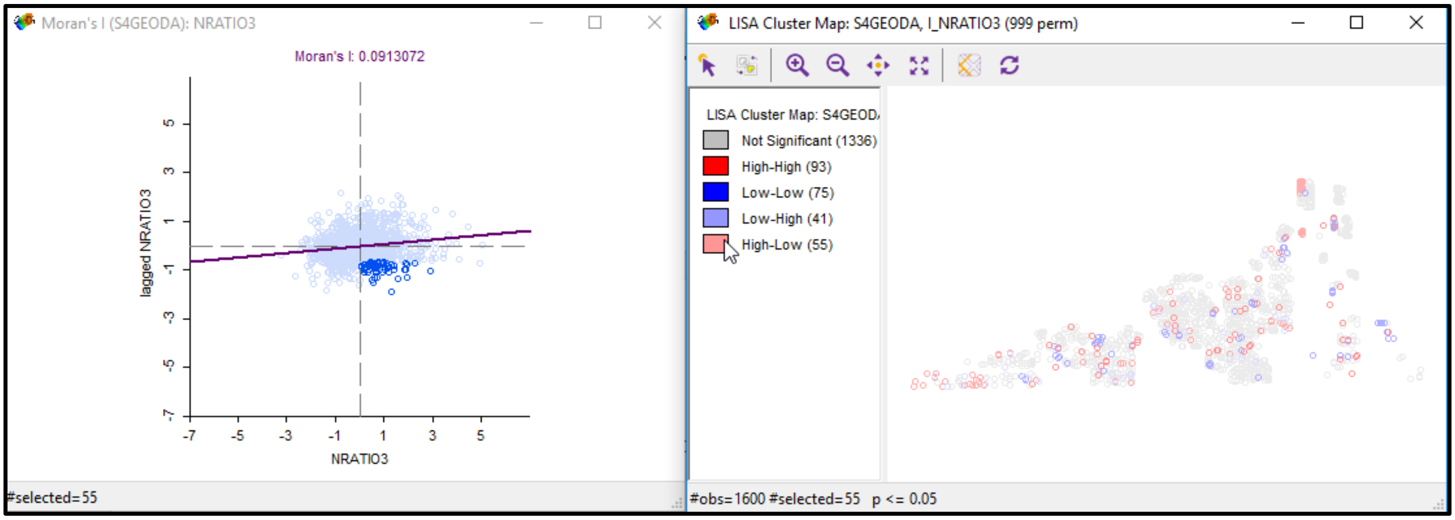
Low near High Ratios

Again, the low near high ratios are geographically dispersed.



High near Low Ratios

The last plot of this series also shows geographic dispersal of the high ratios near low ratios.


Moran's I (S4GEODA): NRATIO3
Moran's I: 0.0913072
lagged NRATIO3
NRATIO3
#selected=55
LISA Cluster Map: S4GEODA, I_NRATIO3 (999 perm)
LISA Cluster Map: S4GEOD
Not Significant (1336)
High-High (93)
Low-Low (75)
Low-High (41)
High-Low (55)
#obs=1600 #selected=55 p <= 0.05

Appendix A - Variable Definitions

Field Name	Description as Needed
pin	unique parcel identifier
pure market	open market 0,1 indicator
block	map block
town	township code
nghcde	neighborhood code
sqftl	lot square foot
landval	land value
class	categorica variable combines many property factors into logical groups
age	relative to 2018
sqftb	building square foot area
mos	Month of sale
yr	year of sale
mktval	previous market value
rs	type of residence
use	single family 1, multi-family 2
num	categorical variable relating to number of living units
extcon	exterior construction code
rf	roof construction code
rooms	rooms excluding baths
bedrooms	bedroom count
basement	basement type code
bsfn	basement finish code
heat	heating system code
flfurn	floor furn 0,1
unitht	unit heater 0,1
stove	Stove 0,1
solar	Solar 0,1
aircond	Yes=1, No=0
firepl	fireplace count
comm	no. commercial units
attc	attic type code
atfn	attic finish code
fullbath	coujnt
halfbath	count
plan	architectural code 1 architect, 2 stock plan
ceiling	cathedral ceiling code
qual	quality of construction code
renov	renovation, yes=1, no=0
site	site desirability code
gar	garage size code
prch	enclosed porch code
rep	state of repair code
amount1	Price

Appendix B - The Location Factor Process

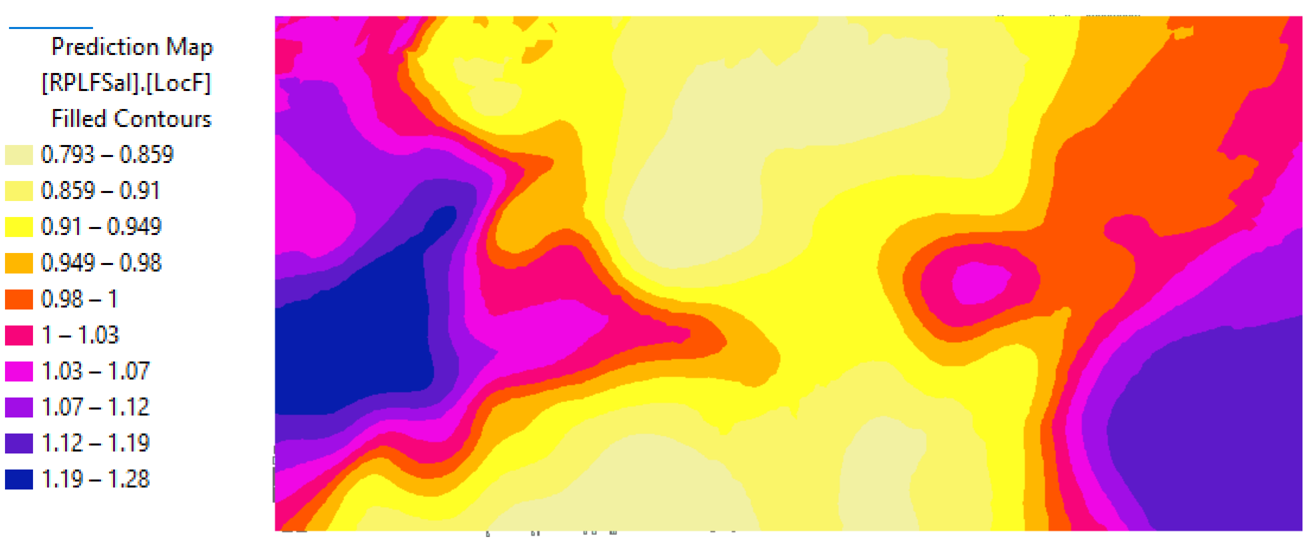
Note: The example used is Rogers Park.

A location factor was derived by use of Geographically Weighted Regression (GWR). The process is one in which a small number of variables not including spatial regime variables is calibrated. The resulting coefficient set is then used to value a “market basket home”. The result is the value of the same home moved around the jurisdiction in question, called “market basket value”. The actual value is arbitrary and depends on the chosen characteristics of the market basket home. The figure depicts the value using proportional symbols.

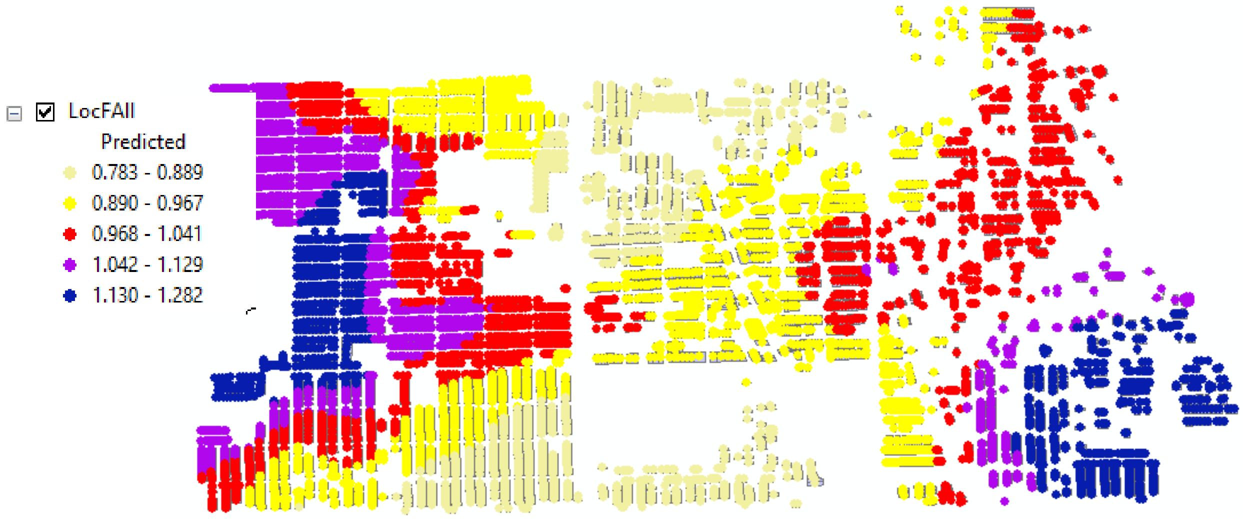
The Location factor is simply the market basket value divided by the average market basket value. The thematic map would look the same, but with a different scale.

The issue is applying the location factor derived from the sales to all properties needing to be valued. The solution is to develop a spatially averaged location factor surface and then to apply that to the inventory of properties to be valued. The method used to do this is called “Kriging” or in this case Universal Kriging.

The resultant surface and thematic legend are shown in the image below.



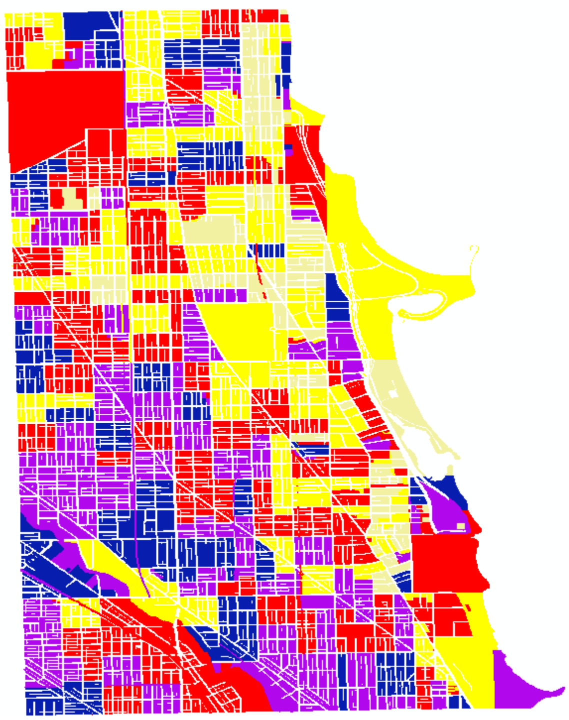
When applied to all properties the thematic map of Location Factor is given in the next image.



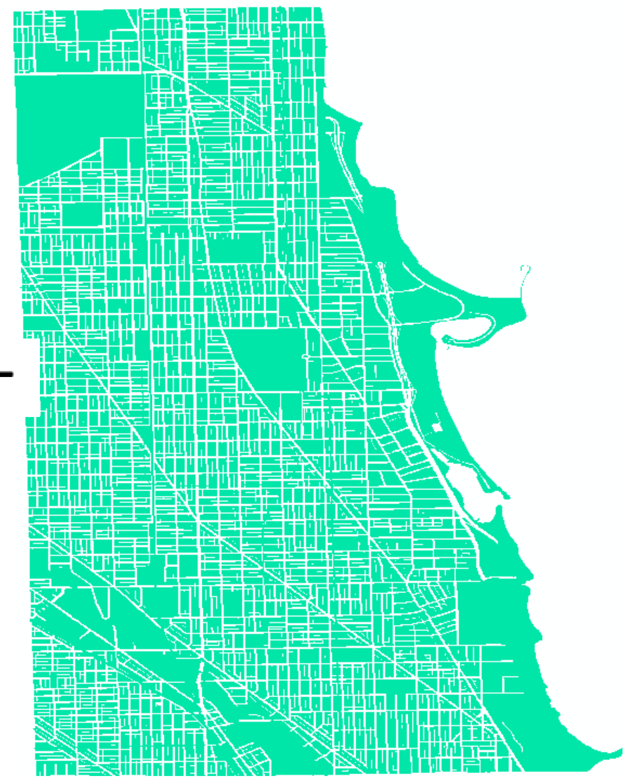
Appendix C - Owner Occupancy

Note: The example is taken from Lake View.

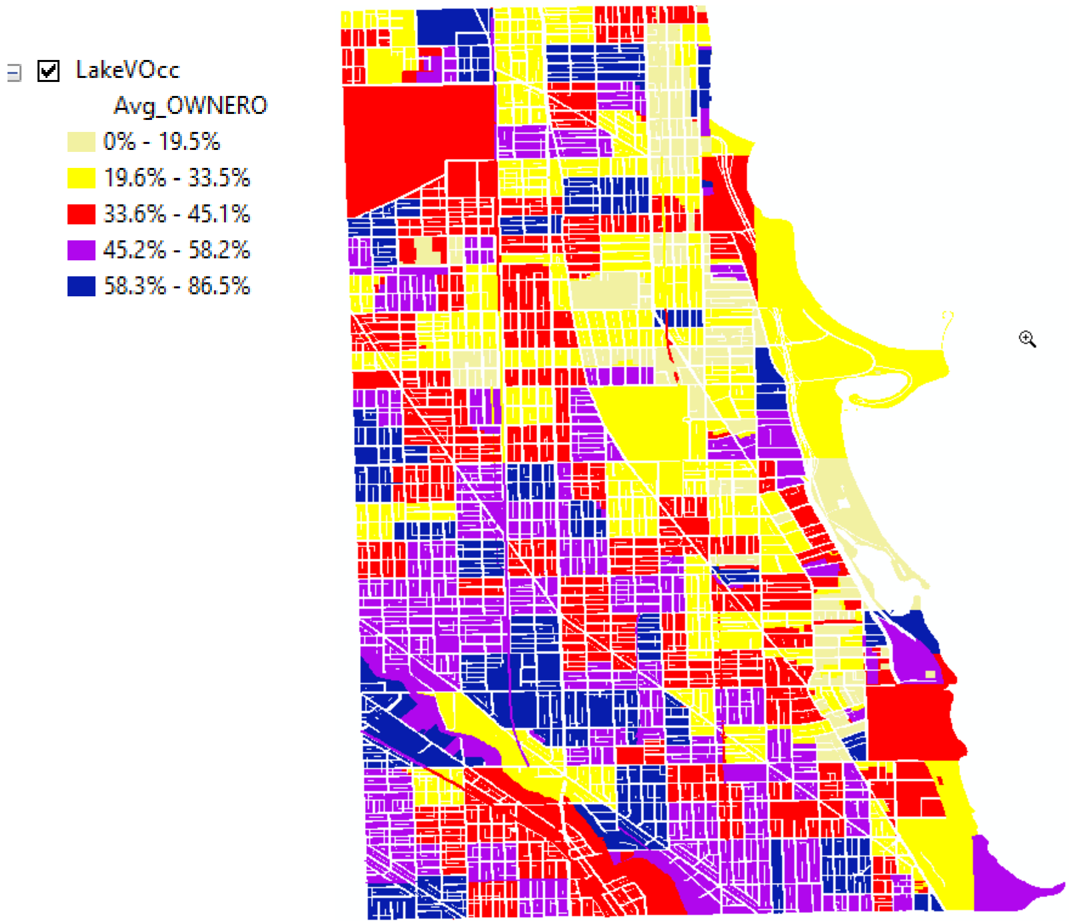
Owner Occupancy data is available at the Census Block Group level. County data is organized at several levels including parcel, block and neighborhood. Since the two geographies are organized differently, they were joined using what is called a “spatial join”. The image on the left below is of the owner occupancy level. The image on the right represents the parcel fabric.



+



When joined the result becomes a parcel fabric with spatially interpolated owner occupancy data. The owner occupancy data is thus made available at the individual parcel level and becomes a candidate variable in an MRA Model.


LakeVOcc
Avg_OWNERO
0% - 19.5%
19.6% - 33.5%
33.6% - 45.1%
45.2% - 58.2%
58.3% - 86.5%

Appendix D – Outlier Detection

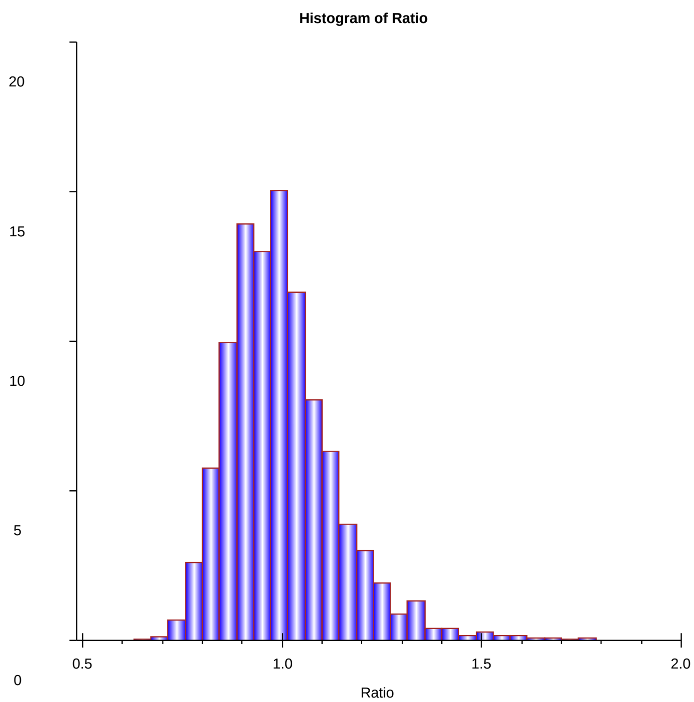
Note: This example is taken from Rogers Park

When a model is first calibrated, it is often the case that some of the sales used in the modeling process are not representative of the group. Initially there are usually some extreme outliers. The traditional method for identifying outliers is to examine the ratio of estimated value to sale price for the sales in the sample. The method used is that described in the IAAO Standard on Ratio Studies. In brief, the process is:

1. Locate 25th percentile ratio
2. Locate the 75th percentile ratio
3. Compute Interquartile ratio or IQR (75th percentile-25th percentile)
4. Compute lower limit as 25th percentile - factor*IQR
5. Computer upper limit as 75th percentile + factor*IQR

The factor is typically chosen as 1.5 or 3.0 depending on whether the goal is to detect extreme outliers (3.0) factor or to take a deeper cut a factor less than 3.0.

It is contended herein that the IAAO standard is faulty and needs to be modified to function as a reasonable tool in identifying outliers. First consider the distribution of ratios created by stochastic process used to simulate a sales sample along with the value estimates produced by a CAMA model. The figure below shows the histogram of the appraisal to sale ratios.



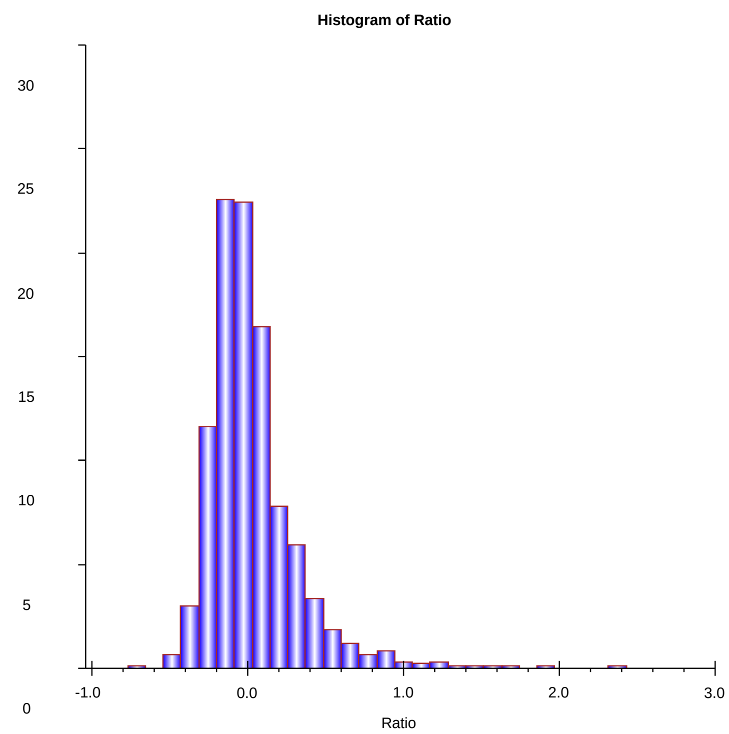
The sales ratio study for this distribution is as follows:

Count	Median	Mean	WgtMean	IQR	SD	COD	COV	PRD	PRB
2400	0.983	1.000	1.001	0.159	0.139	10.483	13.865	0.999	0.013

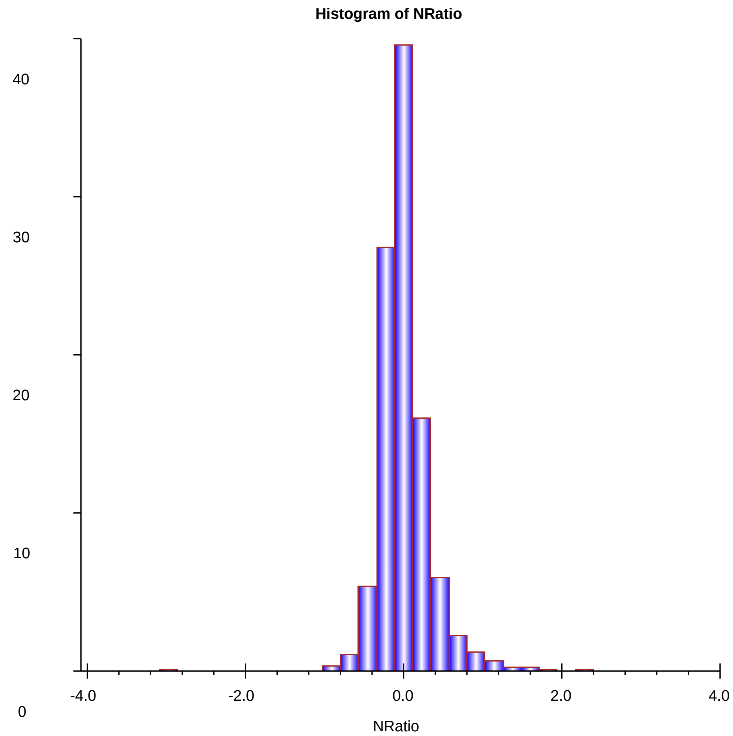
The corresponding Outlier detection parameters using various factors in the IQR detection process are shown below. The point being that for this simulation, an IQR factor of 0.75 produces 11.21% outliers while a factor of 1.0 produces 6.79% and so on down the table until a factor of 3.0 nets 14 outliers and 0.58%.

IQR Factor	IQR	25th Pctile	75th Pctile	Low Lim	Upper Lim	Out Count	Out Pcnt
0.75	0.159	0.906	1.066	0.787	1.185	269	11.21%
1.00	0.159	0.906	1.066	0.747	1.225	163	6.79%
2.00	0.159	0.906	1.066	0.588	1.385	41	1.71%
3.00	0.159	0.906	1.066	0.428	1.544	14	0.58%

In the realistic case of Rogers Park, the histogram of ratio (centered on 0 and expressed as a decimal fraction) produced by the first model with no outliers removed is shown in the image below. It is evident that the histogram is not symmetric. The major reason for this is that although ratios above 0.0 are unbounded, ratios below 0.0 are bounded by a lower limit of -1.0. Another way of saying it is that the range of ratios where the estimate is below the price is compressed compared to those where the estimate exceeds the price.



A transformation on the ratios below 100% yields the far more symmetrical histogram below. The definition of the ratios below 100% is 1-price/estimate. Now, it is easily seen there is one extreme outlier at about -3.0. The same sale does not look so much an outlier in the original histogram.



The IQR calculations are revealing as well. The comparisons include using an IQR factor of 3.0 and one of 0.75. The outlier counts for the standard ratio (Ratio) and the normalized ratio NRatio both centered on 0 and expressed as a decimal fraction instead of a percent. What is telling is a comparison of the outliers removed from the low and high sides of the distribution. Using the standard ratio, the Low to High outlier ratio is much lower than that for the NRatio. In other words, the standard method is missing out on the outliers when the estimate is lower than the price.

IQR Factor	RatioType	IQR	25th Pctile	75th Pctile	Low Lim	High Lim	Out	Low	High	Pcnt Out	L/H%
3	Ratio	0.2775	-0.1493	0.1282	-0.9819	0.9608	14	0	14	1.51%	0.00%
3	NRatio	0.3037	-0.1755	0.1282	-1.0866	1.0394	12	1	11	1.30%	9.09%
0.75	Ratio	0.2775	-0.1493	0.1282	-0.3574	0.3364	118	15	103	12.74%	14.56%
0.75	NRatio	0.3037	-0.1755	0.1282	-0.4033	0.3560	139	45	94	15.01%	47.87%

It is the NRatio method of outlier detection that is used most frequently in the Chicago Triad.

Appendix E – Retransformation Bias

Note: This example is from Jefferson Township.

The reader will note that the median and weighted mean for the multiplicative model are lower than that for the additive model. This is a result of what is called the “retransformation bias”. In the case of Jefferson, the multiplicative model was accomplished by taking natural logs of sales price and the continuous variables used in the model. The resultant prediction is in the natural log scale. To get an estimate of value, the logged value must be transformed back to the original scale – retransformed. It turns out that this transformation/retransformation process can introduce a bias in the final estimated value. The easiest way to see it is in the weighted mean of the additive vs multiplicative models. The additive model has a weighted mean of 1.000, while the weighted mean of the multiplicative model is 0.987.

The literature on this topic is mathematically complex and beyond the scope of this report. A glimpse at the topic in a journal article is shown below.

Short Communication

REGRESSION ANALYSIS OF LOG-TRANSFORMED DATA: STATISTICAL BIAS AND ITS CORRECTION

MICHAEL C. NEWMAN
University of Georgia, Savannah River Ecology Laboratory,
P.O. Drawer E, Aiken, South Carolina 29801

(*Received* 4 *June* 1992; *Accepted* 29 *September* 1992)

Abstract – Power and exponential models are used frequently in environmental chemistry and toxicology. Such models can generate biased predictions if derived with least-squares, linear regression of log-transformed variables. An easily calculated but seldom used estimate of bias can enhance the accuracy of subsequent predictions. This prediction bias and means of correcting it are presented, along with several examples.

Keywords – Statistics Regression Bias Log-transformed variables

THE PROBLEM

Power relationships

Conforming to the notation of Neter et al. [7], the regression model used to describe power relationships is

$$\log Y = \beta_0 + \beta_1 \log X + \epsilon \qquad (1)$$

where

β_0 = the regression intercept estimated by b_0
β_1 = the regression slope estimated by b_1
ϵ = the random error term.

Let ϵ_i represent the error term associated with the ith data pair (X_i, Y_i). Then the mean expected value of ϵ for any data pair, $E(\epsilon_i)$ is zero with a variance of σ_i^2. Variances of the error terms asso-

The approach used here is to correct the value estimates by the inverse of the weighted mean prediction. In the final model results the actual correction was Predicted/0.9743.

Appendix F - The Class Variable

Model with Class Variable

Wiz Mult Base with Class

Dependent	PRICE
Std Error for Estimate	0.2699
Constant:	1,332.7808

Attribute	Coeff
BSF	0.6786
CLASS	
10	1.1896
3	1.6721
6	1.8536
7	1.6584
4	1.6335
5	1.7216
8	1.2539
78	1.8431
95	1.2201
2	1.6974
34	1.8269
12	0.6631
11	1.0000
NUM	
2	1.0967
3	1.0742
1	1.2794
4	1.2681
5	1.0000 **
6	1.0000

** Problem Distinct Values Treated as Base

Model Statistics	
Total Valued	926
R squared	0.5114
Adjusted R squared	0.5022
COD	21.3741
COV Median	33.2495
COV Mean	30.5983
Median	0.9746
Mean	1.0387
Weighted Mean Ratio	0.9677

Model Without the Class Variable

Wiz MultBase No Class	
Dependent	PRICE
Std Error for Estimate	0.2405
Constant:	876.2168

Attribute	Coeff
BSF	0.4695
LSF	0.2543
ROOMS	0.1958
RQOS	
8	1.0416
9	0.9994
2	1.1008
1	0.9701
6	1.0622
7	1.0934
4	1.0526
5	1.0211
19	0.8289
18	0.8825
13	0.9102
12	0.8848
11	0.9967
10	1.0095
17	0.9277
16	0.9081
15	0.9368
14	0.9911
20	0.8064
3	1.0000
LOCF	0.9190
NUM	
2	0.6952
3	0.6937
1	0.7699
4	0.7190
5	0.6574
6	1.0000
REP	
3	0.6768
1	1.1234
2	1.0000

Model Statistics	
Total Valued	926
R squared	0.6177
Adjusted R squared	0.6049
COD	18.8576
COV Median	28.7837
COV Mean	26.7443
Median	0.9753
Mean	1.0301
Weighted Mean Ratio	0.9756

Appendix G – Spatial Dependency

A means to verify the locational stability of the estimates is provided computing Local Indicators of Spatial Association often referred to as LISA. Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. In mass appraisal it is customary to look for spatial clusters in the ration of appraised value to sale price. The plot below has as its X Axis the z-transform of Ratio, defined as $z = (x - \mu)/\sigma$ where x is the individual ratio, μ is the mean ratio and σ is the standard deviation of the ratios in the sample. The Y axis is the average of the five nearest transformed ratios not including the ratio of the X Axis. The fact that there is very little slope to the plot is a good indication that there are no spatial clusters of high or low ratios.

The Moran's I global statistic of 0.0147 at the top of the scatterplot is an indication of low spatial autocorrelation. It is a statistic that ranges from -1.0 to 1.0. A positive value for Moran's I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier. A value close to 0.0 means there is not much in the way of spatial patterns in the set of values. In the case of Jefferson, the statistic is very close to 0.0.