AN INDICATOR ANALYSIS TO IDENTIFY POTENTIAL VACANT RESIDENTIAL PROPERTIES

Brian A. Mikelbank
Associate Professor of Urban Studies
Director, Center for Planning Research & Practice
Maxine Goodman Levin College of Urban Affairs
Cleveland State University

Research Conducted as
Visiting Scholar, Federal Reserve Bank of Cleveland,
Community Affairs Office

Abstract

It is hard to avoid the deeply disconcerting news about the housing market lately. Whether it be from a slowing market, foreclosures, predatory lending, sprawl, etc., the problem of vacant and abandoned properties is likely the result of a myriad of housing specific and broader economic factors. Accordingly, many cities are actively tracking the distribution of vacant and abandoned properties within their borders. But what about cities that are not? Where do they start in order to get a grasp on the forces operating in their housing market?

We present here the construction of a vacant and abandoned property indicator system. The goal was to use data that would likely be readily available, and methods that would be easily replicable to produce an indicator that would be useful in understanding the state of a city's housing. The process is demonstrated for two Ohio cities currently battling the spread of vacant and abandoned properties, Columbus and Cleveland.

Acknowledgements: This research is the first part of a larger project, *Emerging Threats to Residential and Community Stability: Identification and Analysis of Vacant, Abandoned and Foreclosed Properties.* The second part is a spatial econometric analysis to estimate the impacts of vacant, abandoned and foreclosed properties in the city of Columbus, OH. The project is funded through the Visiting Scholars Program, Community Affairs Office, Federal Reserve Bank of Cleveland.

Discussions with several individuals have aided this research considerably: Charlie Post, Ivan Maric, Tom Bier (all at The Center for Housing Research & Policy), Eben Dowell, Kerry Sullivan, Roberta Garber (all at Community Research Partners), Lisa Nelson, Francisca Richter, and Ruth Clevenger (all at the Federal Reserve Bank of Cleveland). Finally, feedback from my presentation of this material at the National Vacant Properties Campaign meeting in Pittsburgh, PA was also helpful.

Community Research Partners provided nearly all of the data that were necessary to conduct the Columbus analysis.

The findings reported here are the opinion of the author alone, and do not reflect the views at the Federal Reserve Bank of Cleveland or the Federal Reserve System. Any errors or omissions are the sole responsibility of the author.

Introduction

Housing has had more than its fair share of news coverage lately, and it is overwhelmingly negative. Newcomers to the headline arena are the collapse of the subprime lending industry, and related, the rapid rise in foreclosed, vacant and abandoned properties. These join longer-term recurring (and yes, also related) topics of suburbanization and sprawl.

This news has motivated several groups into action. The most well known is likely the National Vacant Properties Campaign, organized by Smart Growth America, Local Initiatives Support Coalition, and the Metropolitan Institute. The campaign provides a national framework for addressing issues related to the vacant and abandoned property problem.

At the state level, Rebuild Ohio is a consortium of individuals and organizations, whose mission is "to promote reclamation and redevelopment of vacant and abandoned property for economic vitality and enhanced quality of life in neighborhoods, towns and cities throughout the state (www.vacantproperties.org/rebuildohio.html)." As part of this effort, Rebuild Ohio brought together a set of experts from public, private, and academic institutions to discuss the research agenda that Rebuild Ohio was putting forward. Building off of these discussions, the author and the Office of Community Affairs at the Federal Reserve Bank of Cleveland developed the ideas leading to the research detailed in these pages.

The goal of the research, and this paper, is to detail the construction of an indicator database relating to vacant and abandoned properties. We wish to illuminate the process by which a city might go about turning available raw data into a set of "actionable" pieces of information. This set of data could then form the foundation of a vacant and abandoned property strategy. This process is demonstrated for the two Ohio cities of Columbus and Cleveland[1].

---

[1] Note that a companion piece to this paper, "Under the Hood: An Indicator Analysis to Identify Potential Vacant Residential Properties" provides a more hands-on approach to constructing

The remainder of this paper is organized as follows: The next section discusses the background and provides the context of this research. The third section details the data that were available for the Cleveland and Columbus demonstrations of the indicator analysis. With the data assembled, the following section describes the way in which these raw input variables are combined to produce a single V&A property indicator. The results are discussed next, followed by a discussion of possible extensions to the straight-forward approach presented here. The final section concludes.

Background

Rebuild Ohio commissioned a survey of select Ohio cities to determine their approach in battling the vacant and abandoned (V&A) property problem that has gripped many, if not all, Ohio cities. This survey was part of their effort to measure the economic cost of V&A properties borne by municipalities.

What was striking about the survey results was the variance in the manner in which cities were coping with the crisis. For example, Dayton OH conducts a complete census of its housing stock every two years, and does regular updates in between in order to provide current information to city officials and other interested parties. Springfield OH, on the other hand, deals with vacant properties only when they appear as problems through their code enforcement efforts. It was this contrast in approaches to the problem that motivated this research.

On the surface, this effort might appear to be only marginally useful for the cities that are already actively tracking properties within their borders. They are likely amassing rich time-and space specific databases of parcel level characteristics that are useful for action and planning. Further, these have often been used in detailed analyses of the V&A property phenomena. However, these cities might still yield considerable utility from a V&A indicator database in terms of better understanding the forces that led to

---

the indicator system. This paper deals more intensively with the data and methodological issues.

their current V&A distribution. In other words, knowing the distribution of V&A properties really tells you only about the end result of a complex process, not about the process itself. Further, developing a V&A property indicator score *and* knowing which properties are actually vacant and abandoned gives a city an additional edge by providing information on occupied properties that are judged likely to be vacant, and similarly, on vacant properties that are absent the traditional warning signs.

On the other hand, this research should have the greatest, most obvious, utility to cities that are just now beginning to deal proactively with the V&A property problem. The paper represents a roadmap. Its starting point is a blank slate; its final destination is a data-based snapshot of a city's housing stock potentially most vulnerable to vacancy and abandonment.

In discussion of indicator systems more broadly, Sawicki and Flynn (1996)[2] note the importance of both the participatory process and the capability of affecting citizen action and policy direction. While former plays a smaller role in an analysis intended for a housing/development audience, the later two ring true. Identification of parcels and neighborhoods of high V&A risk must represent a call-to-action among local residents and leaders. Certainly the advantage of the knowing "risk" in advance is so that active steps can be taken to minimize that risk.

The Federal Reserve Bank of Cleveland's interest in such a demonstration is in heading off the debilitating impact V&A properties can impart on the local community, in terms of its neighborhood quality, its housing, and its people. The Center for Housing Research & Policy has been studying the trends that impact the regional housing market for more than twenty-five years, and so certainly such an analysis of vacant and abandoned properties is a natural extension of the research the Center has been producing. Especially relevant for this paper is the Center's maintenance of a

---

[2] David S. Sawicki and Patrice Flynn, 1996. "Neighborhood Indicators: A Review of the Literature and an Assessment of Conceptual and Methodological Issues." *Journal of the American Planning Association* 62(Spring): 165-183.

parcel-level database for properties in Cuyahoga County (the home county of Cleveland, OH).

It should be explicitly stated that the focus of the research is not in building an "early warning system". That term has negative connotations that are not consistent with the goals here. "Early warning system" evokes images of red flashing lights and wailing sirens, attempting to provide advance notice for the population to flee, take cover, or brace for something most unpleasant. An indicator system, as we call the effort outlined here, provides *information*. The end goal isn't to send residents a message to get out while you can, but to provide a tool so that municipalities can effectively direct the appropriate attention and resources where needed.

Data Sources and Samples

In thinking about the incidence of vacant and abandoned properties, we categorized and organized the relevant forces into three levels of influence. The first of these focuses on the characteristics of the property itself – we call these the parcel level indicators. The second category is also based on parcel level indicators, but this category reflects the characteristics of the parcels surrounding the parcel of interest. We call this category the nearby parcel level. The final category, the neighborhood level, allows for influences across a broader geography[3].

It is important to note that the indicator system could be constructed with one, two, or three of these levels of data. It might be that time, resources, or data dictate that all three levels not be implemented at once. Even in a scaled back configuration, the system would still yield useful information to a municipality. There is nothing that would restrict additional layers of data being added at a later date. Additionally, there might be an aggregation level that makes sense in a particular research context – formal neighborhood boundaries, political wards, or planning areas come to mind.

[3] It is tempting to think that city-level influences would have a place in an indicator system such as this. If the system is built for a single city, however, those city-level indicators would have no variation within the city. That is, if we wished to include the city's unemployment rate as an indicator, that same rate would apply to each parcel in the city. It would have no role in differentiating between parcels that are more or less likely to become vacant.

The data selected for the Columbus and Cleveland demonstrations reflect data availability for each city. It is unlikely that the data available, or even desirable, for any two locations will be the same. Thus, there will be variation across indicator systems constructed for this purpose, and that flexibility is a favorable characteristic of composite indicator systems. They can be customized to reflect the unique market characteristics of each city in which they are applied.

An additional advantage, and likely an advance response to criticism, is that there is no *single* indicator or predictor of which properties will or will not become vacant and abandoned. And thus, no indicator included herein is perfect. Surely there are houses, blocks, or whole neighborhoods with healthy local housing markets, but where one or more indicators fall in a range that would otherwise cause concern. Certainly the reverse is true also – we do find vacant and abandoned properties where indicators would lead us to believe there would be none. These are among the primary reasons we avoid the "early warning system" moniker. It is not *the* tool a municipality should be using, but we do argue it is *a* tool that would provide useful information.

Parcel level data

The various offices of county government are rich sources of parcel level data. The county auditor, within its responsibility of appraising and assessing real estate, maintains a host of property-specific information. In particular, we used the age, condition, and construction quality of the house, posing that older, poorer condition and quality housing would be more likely to be or become vacant. We used the appraised value of the house, thinking that lower value housing would be a likelier candidate for V&A.

We used a ratio the auditor's estimated market value to the most recent sales price (occurring after 2000) to reflect possible house quality changes. If the sale was prior to the market value assessment, we thought a high ratio was indicative of V&A. For example, if a house sold in 2002 for a price of $145,000 but was assessed at 100,000 in 2006, the ratio would be greater than one, and perhaps indicative of a decrease in quality since the last transaction. Conversely, if the house sold in 2007 after its 2006

assessment and the sales price-to-assessed value ratio was low that would similarly indicate a potential quality decline.

We also took into account the annualized appreciation rate between the last two sales of a house. Houses achieving slow or negative appreciation could be more susceptible to V&A. For Cleveland, we created an additional variable to highlight properties potentially involved in non-arms length transactions. The 'quick-resale' variable indicates back-to-back transactions within 90 days, with an annualized appreciation/depreciation rate of at least 50%.

The last two parcel level variables used indicate whether the property was recently tax delinquent, and whether the property had a recent interruption in its water service[4].

Nearby parcel level data

The exact same parcel level data as just discussed form the basis of the nearby parcel level data also. The first decision to make in constructing the nearby parcel indicators was what to consider "nearby". For this research, we consider parcels in the same census block to meet the nearby criteria. While there are certainly other more complex choices available, blocks are a universally available geography, and since parcel data are often available with their associated census geography, it makes aggregation possible without the use of Geographic Information Systems, which may not be available or skill-level accessible. The manner in which each indicator is aggregated to the block level is discussed in the methodology section to follow.

Neighborhood level data

Neighborhood level data are reported at the census tract level, and many of the indicators are in fact data from the decennial census. We used several indicators relating to income. The first was median household income. Second, we used housing burden for owners and renters, which measures the percentage of households spending more than 40% of their income on housing. Since single-earner households might lack the income safety net of a second income earner, and since households

[4] In the Columbus data, the water utility variable is reported at the neighborhood (census tract) level.

with children under the age of 18 likely incur a wider range of costs than do households without, the percentage of single person households with children under 18 was included in the analysis.

Another group of census variables is included to reflect local housing dynamics. We use the ratio of housing unit growth to household growth between 1990 and 2000. If the growth of housing units outpaces the growth of households, then V&A is more likely. We also include the median house price appreciation between the previous two sales for the tract. Housing tenure is included in two ways. First, the percent of the housing units that are renter occupied is included, as is the change in renter occupancy between 1990 and 2000. The final census variable we utilized reflects residential turnover in the five years prior to the census. It is the percent of households living in the same house as they were living in 1995. An additional demographic variable that is included is the percentage of the population over 60 years of age, to reflect potential near-term turnover.

We included a variety of other economic factors at this level also. We included the percentage of originated loans (purchase and refinance) that were high-priced, separating out those that went to owner-occupiers and those that went to non-owner occupiers (investors/landlords). Also included is the percentage of home purchase and refinance loans that were made by subprime lenders. By also including the incidence of foreclosure filings in the census tract, the combination of these indicators is meant to capture the incidence and impact of the recent lending crisis on the distribution of vacant and abandoned properties.

Data Sources

The search for the necessary data underlying an indicator system might be simplified if there exists a centralized community data system for your city or region. These can be maintained by any number of organizations, from a city itself, a local nonprofit or research organization, or a university. The benefit of such a system is that they generally bring together data from various sources, making them available in one location, ideally delivered electronically over the Internet, and in a consistent format. The ability to use one of these systems makes the data integration task much easier.

As is evident from the variety of data detailed above, organizing the data so that they are ready for analysis can be a formidable task. The current research benefited from NEO CANDO in Cleveland (at Case Western Reserve University) and DataSource in Columbus (at Community Research Partners).

Further, it is worth restating that there is no magic menu from which to select data for an indicator system. Similarly, there are no "best" data. Data are typically imperfect, and every indicator effort will first and foremost be a function of the available data upon which it is based. This underscores the effort being labeled an indicator system, rather than an early warning system. There is no sure-fire method of predicting V&A.

Methods

With data collected at the parcel, nearby parcel and neighborhood level, the next issue is how they might be meaningfully combined. The data assembled are a mix of qualitative and quantitative data, with the latter measured at different scales. Some of the data are represented in percentages (percent renter-occupied, for example), while others are measured as dollars (such as median household income). The qualitative data are typically indicative of some condition (such as having had a water service interruption).

There were several goals to keep in mind when deciding how to integrate the thirty-plus individual indicators. Since these efforts are geared towards community-based action, primary among them was that the results are easy to interpret. Despite any other characteristics it may have, an indicator system has failed if its output can't be clearly communicated to its target audience.

Additionally, a method was needed that would produce results that were consistent for the entire city, so that comparisons between parcels or even neighborhoods, would be valid. This means that the final parcel indicator had to account for the fact that not all variables were available for every parcel. Thus, while some parcels have the full set of indicators, others might have fewer, so any raw tally of indicators would be troublesome.

Finally, we wanted for our integration strategy to be straight-forward. The idea was to present a simple integration path, so as not to construct a barrier to implementation for communities, offices, or individuals whose data experience or expertise might not lend itself to more complicated approaches. Certainly, there are numerous other, and more complex ways, to bring these data together into a single indicator, and the section of this paper entitled "Extensions" puts forth several such options.

Nonetheless, the approach presented here provides a consistent framework within which to examine parcel characteristics, it is easily interpretable, and is replicable without sophisticated or specialized software. In fact, the analysis presented herein relies entirely on functionality contained in Microsoft Excel[5]. GIS was used only for mapping, not for any spatial analytic functions.

For the parcel level and neighborhood level input data, the range of each variable was divided into categories ranging from the condition that would be least relevant to a property being V&A, to the condition that would be most troublesome. Then, observations that fell within each category were assigned a value, ranging from 0 (assigned to all observations in the least relevant category) to 1 (assigned to all observations in the most troublesome category). Thus, for each parcel, all indicators are translated into this uniform scale, where the closer the translated value is to 1, the more likely a candidate that property is to be V&A. This type of classification is similar to what is done when a thematic map or a histogram is constructed. The quantitative information present in the original data is traded for a broader representation of the trends or patterns in the data. Specifically, all of the observations in a single category receive the same "score", but at the same time, the distribution of that variable is easier to understand, since the complexity of that variable is reduced.

---

[5] The most relevant limitation present in Excel, is that its maximum number of rows is 65,536. Depending on the size of the city of interest this could be a limiting factor. Since both the Cleveland and Columbus datasets had more than 100,000 observations, analyses were conducted in SPSS.

There was not a standard procedure for determining how many categories would represent each indicator. For example, if a property was on the list indicating an interruption in their water service, that property was assigned a 1; if not, it was assigned a 0. For quantitative variables, the distribution across the city was examined via a histogram, and if natural breaks were present those served as the primary breakpoints between categories. If there existed no natural breaks in the data, quartiles and/or quintiles were used to determine the categories. In some distributions, one or two natural breaks were present, but the remainder of the distribution offered no logical breakpoints. In these cases, both natural breaks and quartile breaks were used.

Before this same process was applied to the nearby parcel data, an additional step was necessary. For each parcel, for each parcel-level indicator, we wanted an additional indicator that reflected the status of that indicator at nearby parcels. Recall that the concept of nearby parcels is operationalized here as parcels within the same census block. How to represent this set of indicators was another analysis decision. The path taken here was for each indicator, to calculate the proportion of surrounding parcels that had the level of that indicator that was most associated with V&A properties. Thus, since a score of 1 on a single indicator is most strongly associated with V&A, the nearby parcel indicators reflect the proportion of surrounding parcels that score a 1 on that particular indicator. This reveals a quantitative score, ranging from 0 to 1, that is then categorized by the same procedure as is described for the parcel and neighborhood level variables.

Table 1a and 1b show how the input variables were categorized into indicator values between 0 and 1 for Cleveland and Columbus, respectively.

Since the final goal is not to have 34 indicators per parcel, these indicators need be summarized so that each parcel has a single overall V&A indicator. The process followed here is to simply sum each of the indicators for each parcel, and then divide by the number of indicators available for that parcel. The result is a single indicator, ranging from 0 to 1, where the closer the value is to 1, the more likely a candidate that property is to be V&A.

Results: Cleveland, Columbus, and Beyond

The distribution of these final indicator scores is shown in Figures 1 and 2. The Columbus indicator is based on 182,630 parcels and its average is 0.38. The minimum indicator score is 0.13, and the maximum is 0.75. The Cleveland indicator is based on 124,007 parcels, ranges from 0.15 to 0.74, and has a mean of 0.41. It should be noted that the set of indicators for each city was developed independently, and so the results from each city are not directly comparable. This is because the value of each indicator (ranging from 0 to 1) was based on the distribution of the raw input data for each city individually. Therefore, the same input data can contribute differently to its indicator in each city. For example, a median household income equal to or below $20,000 comprised the lowest income category in Columbus, while in Cleveland the lowest category was equal to or below $18,000. And so a neighborhood with a median household income of $19,000 would contribute a value of 1 to the Columbus indicator, but a value of 0.75 to the Cleveland indicator. It is for this reason that the indicator set of each city is meant for comparisons within each city, but not between the two.

The spatial distribution of the indicators is shown in Figures 3 and 4. In these maps the darker orange and red colors are those closer to 1, and thus more associated with V&A. The darker blues are parcels with scores closer to 0, those we would anticipate being less associated with V&A.

The distribution of indicators in both cities shows pockets that are dominated by primarily one color. Darker reds appear on the near east and west sides of downtown Columbus, with an additional pocket to the immediate northeast of downtown. In Cleveland, the east side has scores decidedly more associated with V&A. Cleveland's west side shows the lighter and darker blues that are associated with lower indicator scores. The lowest scores in Columbus are typically found to the north, while moderately low scores are also found in the north, but more prevalently south of the city.

Their patterns are similar in that the locations farthest away from the downtown are associated with lower V&A scores. Even in Cleveland, where the east side generally has higher scores, there is a visible band of green and blue around much of the city's border, where V&A scores are decidedly lower. This could indicate that the market forces associated with V&A have not yet reached these locations to the same degree, and/or there could be positive externalities from neighboring suburban locations that are keeping these forces at bay, relative to their inner city neighborhood counterparts. Columbus' indicator scores, however, follow a more traditional concentric ring, or perhaps a sector model pattern, while Cleveland's has a more distinct east-west dichotomy.

Columbus Comparison

In the case of Columbus, we have the advantage of comparing our V&A scores to the actual distribution of vacant and abandoned properties, using the inventory compiled by the City of Columbus.

If the overall indicator is consistent with reality, we would expect, on average, that actual V&A properties would score higher on our indicator than those that are not V&A. That is in fact the case, as there exists a statistically significant difference between vacant and non-vacant properties. Figure 5, in fact, shows two useful pieces of information. They are box plots of the V&A indicator, separated by whether the house is vacant (on the right) or not (on the left). First, the 'box' part of the plot shows the spread of the middle 50% of all observations in each category, and these are located in distinct positions on the number line. The median for non-vacant properties is approximately 0.35, while the median vacant property score is above 0.55.

Secondly, there is substantial overlap between the two scores. For example, many non-vacant properties scored very highly on the V&A indicator, just as many vacant properties had very low scores. Certainly, the dynamics of a complex housing market, especially one in turmoil from a combination of factors (foreclosure, predatory lending, demographic changes) aren't going to be completely understood or accounted for by a indicator system as put forth here. However these anomalies represent opportunities

to learn more about the local housing market. High scores among non-vacant properties could represent properties on the brink of V&A, or equally likely, houses that simply avoided the V&A cycle, despite indicators that would lead us to believe otherwise. Do these properties have anything in common? Are they randomly distributed throughout the city, or is there some spatial pattern to their survival? Do these properties have any other common characteristics that would explain how they have been able to avoid V&A while other properties, subject to the same conditions, have succumb to the process?

Similarly, there are vacant properties with very low V&A indicator scores. The opportunity here is to learn why these properties fell vacant while showing so few traditional characteristics of V&A properties. Perhaps we can glean from the characteristics of these properties an input variable that is not currently used in the indicator effort. In either case, the instances where the indicator seems to "indicate" the wrong outcome represent chances to learn more about the process, and potentially where we have the process mischaracterized.

Observations on specific variables

Again, with information on which Columbus properties are actually vacant, we could compare the distribution of specific variables to the distribution of V&A. Of course, there is no way of knowing whether these comparisons are merely quirks of the Columbus housing market, or whether they might apply more broadly. Nonetheless, two variables that revealed interesting differences from our expectations are discussed here.

First, we included a parcel-level variable measuring the appreciation rate between the last two recent sales, thinking that houses that lost value, or gained most slowly would be more susceptible to being V&A. In the Columbus case, we learned that while this was true, there was an additional category that was also highly related to a property being V&A: the *fastest* appreciating. In retrospect, the extremely high appreciation rates related to V&A might be indicative of a quick-resale at a substantially higher price, or a "flip". A quick resale at a higher price might not be troublesome, so long as the final price reflects the actual market value. For example,

if buyer #1 purchases the house at a bargain price, and then quickly resells the property at its true market value, there is nothing untoward to cause concern – a buyer "bought low" and sold at a fair price. However, as is more common amongst today's predatory lending crisis, a house is purchased and then resold at a price that far exceeds its actual market value. This process typically involves some type of fraud, either in the appraisal or lending process, or both. In this case, the final purchaser is often left "upside-down" with a loan for an amount greater than the house is actually worth. This is a key ingredient in foreclosure and/or V&A.

At the neighborhood level, we included a variable measuring tenure, thinking that since renters move more often than do owners, that rental properties might fall V&A more often. Additionally, renters could move into home-ownership, leaving behind vacant rental property that would again be a V&A property candidate. In Columbus' reality, however, it wasn't the neighborhoods that were dominated by rental properties, but rather the neighborhoods that were a mix of rental and owner-occupied properties. It is possible that an influx of renters in previously owner-occupied properties was causing a churning in the owner-occupied market, leaving some units vacant. Similarly, these vacant houses would make the neighborhood less attractive to subsequent owners and renters, which could cause higher vacancy rates among all housing units.

Extensions

The analyst faces a variety of trade-offs at each decision point in constructing an indicator. These include conceptualization, data selection, and methodology[6]. In building this indicator system, when a decision was required we took the path that was simpler. Never was analytic validity sacrificed for simplicity, but if the more straight-forward approach was reasonable, it was chosen. Our overall method of taking quantitative input data and "binning" them is analogous to the process of calculating quartiles, creating a histogram or thematic map. In each of those

---

[6] Etzioni, Amitai and Edward Lehman. 1967. "Some Dangers in 'Valid' Social Measurements: Preliminary Notes." *The Annals of the American Academy of Political and Social Science* 373: 1-15.

applications, the input data detail is sacrificed for the purpose of summary. These decisions were in line with the overall goal of presenting a system that others could replicate without advanced spatial or statistical training. For those that are so inclined, however, several extensions to the approach presented here come to mind.

Most prominently, in this paper the input variables are not explicitly weighted. This means that each input variable is by default given equal weight, or equal importance, in determining the final indicator score for each parcel. It might be that some variables should count more heavily in that final calculation. For example, one might posit that the interruption of utility service is a more useful indicator than is neighborhood household income, and so the first variable might receive a higher weight, while the second receives a lower weight. One might also wish to give more influence to the different levels of data collected. For example, the data relating directly to the parcel might be weighted more heavily than nearby parcel data, or even neighborhood data.

Further, it is worth noting that an implicit weighting occurs simply through the selection of the input variables. For example, since there are 9 individual parcel indicators, 9 nearby-parcel indicators, and at least 16 neighborhood indicators, this unweighted approach gives heavier weight, *as a group*, to the neighborhood indicators. Table 2 shows how the input variables are distributed across the levels of data collected (parcel, nearby parcel and neighborhood). For equal weighting by group to be achieved, then an equal number of parcel, nearby parcel, and neighborhood variables would have had to be used in the analysis. Alternatively, it would be possible to sum the indicators within each group, divide by the sum by the number of indicators in each group, and then sum the resulting proportions. This would also result in each data level being weighted equally[7].

An implicit weighting also occurs if several input variables were included to measure a common underlying theme. For example, there are six lending variables included as economic factors in the model. While each is intended to capture different aspects of

[7] This indicator would range from zero to three. One could always divide by three to re-scale the indicator from zero to one.

neighborhood lending activity (including the local potential severity of the lending crisis), it gives the lending theme more influence in the overall score than, say, tax delinquency, an important variable in its own right, but one for which we have only a single input variable. Table 2 also lists the input variables by theme they are intended to represent.

One way to rid the analysis of the redundancy in some of the input variables would be to use factor analysis. From a large set of input variables, factor analysis can identify a smaller number of underlying themes, and it is possible to construct these "factors" such that they are orthogonal – that is, they are uncorrelated. One could even conduct factor analysis upon each level of data to identify the underlying themes in the data at the parcel, nearby parcel and neighborhood level.

A more complex approach to calculating the parcel level indicator would be to define each indicator as a percentage of its most extreme value. This would maintain the advantage of each indicator being scaled to one, and would also maintain the numeric detail of the quantitative input variables[8].

The analysis here is geared toward a cross-section, or snap-shot view. This is consistent with the goal of measuring the current level of V&A risk throughout the study areas. However, if these efforts were to be repeated, analysts could choose their input variables with regular updates in mind. In this fashion, the indicator could be used to monitor the changing trends in the market and to review relevant policies and goals[9].

With the availability of Geographic Information Systems, the nearby-parcel data calculations are a rich area for further development. Specifically, there are a variety

---

[8] Stuart Schwartz suggested this approach. Because it maintained many of the qualities of the current approach, this path was investigated for the analysis. However, many of the variables used here had outliers at the upper end of the distribution, making otherwise high values appear only moderate. It would have been possible to consider indicators as a percentage of the 95th percentile of the input variable to minimize the impact of these extreme values, but the concern was that this would have moved away the goal of a system that was most easily replicated.

[9] G. Thomas Kingsley, 1998. "Neighborhood Indicators: Taking Advantage of the New Potential." Working Paper. Chicago, IL: American Planning Association, October 1998

of ways to define the "nearby" concept. For example, one might wish to consider nearby parcels those within some distance buffer, or those that are immediate (contiguous) neighbors to the parcel of interest. Further, one might employ some type of pattern analysis techniques to identify statistical clusters of input variables of interest. For example, V&A might be more likely if the property is located in a foreclosure "hotspot".

The analysis presented here, is directed at the excel-level data user. The system is replicable in excel, and the user is free to enhance the system as the data evolve, as their skills develop, or as time permits.

Discussion & Conclusions

The analysis presented here is a demonstration of how raw data can be integrated to create useful and actionable information. This particular effort was built at three data levels, although subsequent efforts could utilize any subset of the three to begin the relevant database.

There are several reasons a municipality might want to undertake an analysis such as the once described herein. V&A properties are costly to the city, and costly to the neighborhoods in which they occur. They drain value and they drain morale. Left unattended, one V&A property can easily breed others, as uncertain or fearful neighbors re-evaluate their housing plans in light of these obvious signs of decay and market failure. Crime in and around these empty structures becomes a concern, as does arson.

This analysis is clearly geared towards cities that do not know the distribution of V&A properties within their borders, such that they can get a sense of what might be vacant, or at a minimum identify locations that might be especially susceptible to the forces of V&A. The data compilation might also prove useful to other offices of government, such as community/economic development, code enforcement, etc.

Even cities that are tracking V&A properties in their city, and are actively involved in the fight against V&A, might find such an analysis for their city informative. First of all, they can compare the indicator against its known distribution of V&A properties, as we did here for Columbus. Locations where the V&A indicator varies substantially from the known distribution of V&A properties might give the city a window into the near future, or shed new light on the process of V&A, identifying locations that have weathered the storm without becoming vacant, or have become vacant with no, or few, signs of housing despair. For all cities, it can help direct outreach and resources in the appropriate directions.

Since an indicator system is only as good as its input data, users should keep in mind potential drawbacks. For example, the neighborhood data are census variables, which are collected but once every 10 years. The process of urban change, especially when driven by distressed markets, can be substantial over that period of time. The degree to which local authorities make available parcel level data will vary by location, and even when available, the quality of those data may be of concern.

Beyond the issues of data quantity and quality, there is no magic formula for combining data into a single indicator, or even set of indicators. If the number of indicators is small, it might be possible to examine them individually. However, when the number of input variables gets beyond a dozen or so, data-overload sets in, and our ability to glean useful information and patterns across variables declines, necessitating a strategy for data reduction.

Obviously, the research presented here is only a starting point for communities. It is a path to demonstrate the breadth of the problem in our communities. How widespread is this crisis? How pervasive is it in a city's many neighborhoods? Further research by the author, supported by the Office of Community Affairs, will analyze their impact on the local housing market. The focus will be on the depth of the problem. Knowing V&A's spatial footprint, how big of a hole in the market is this problem gouging at each location? These critical pieces of the puzzle, measurements of the incidence and impact of the crisis, are necessary if local action and policy are to be appropriately formulated and prioritized.

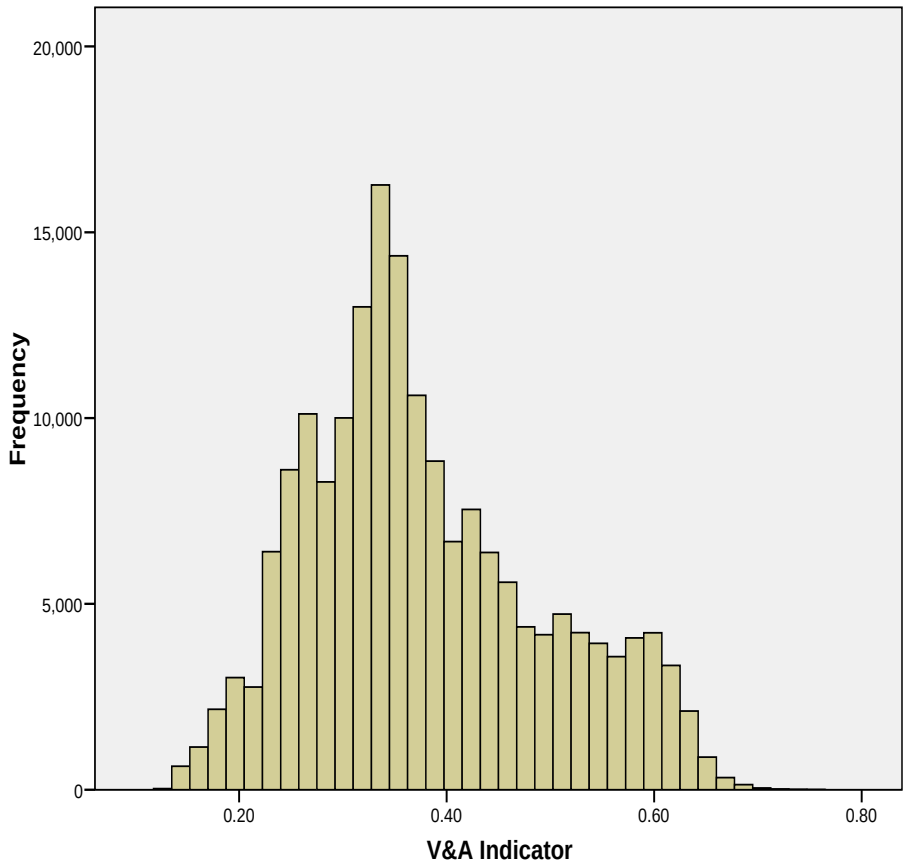

Figure 1. The distribution of the final V&A indicator score for Columbus.

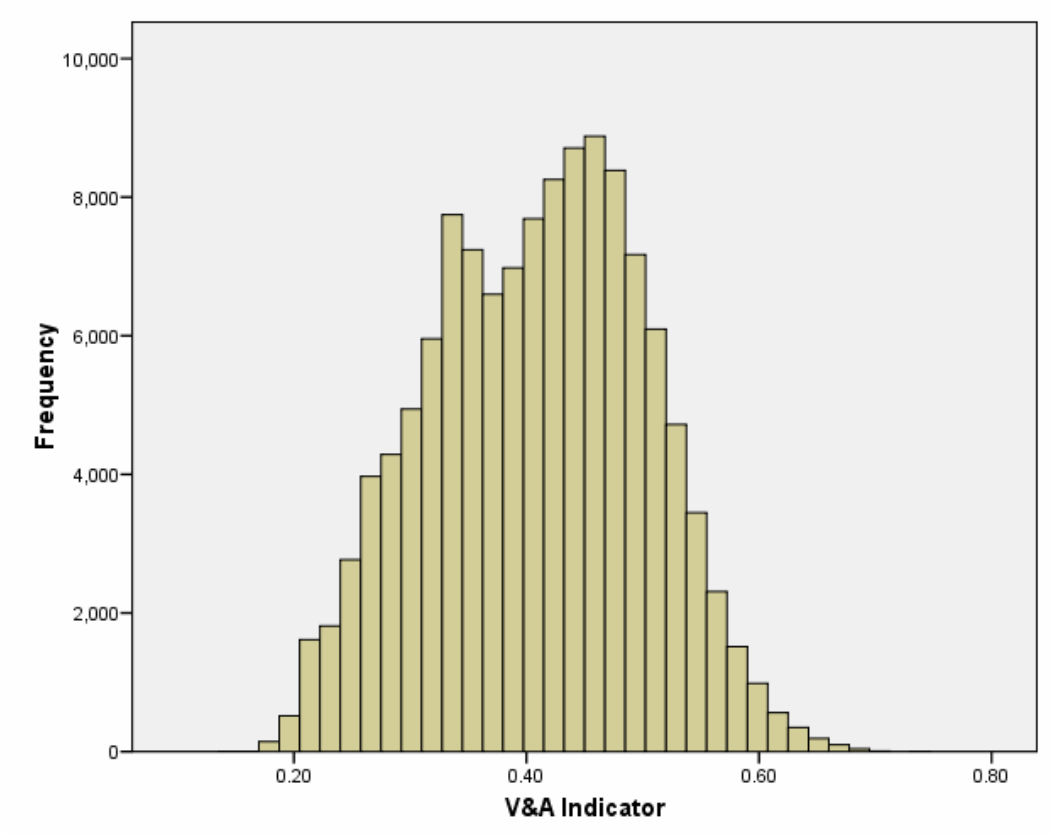

Figure 2. The distribution of the final V&A indicator score for Cleveland.

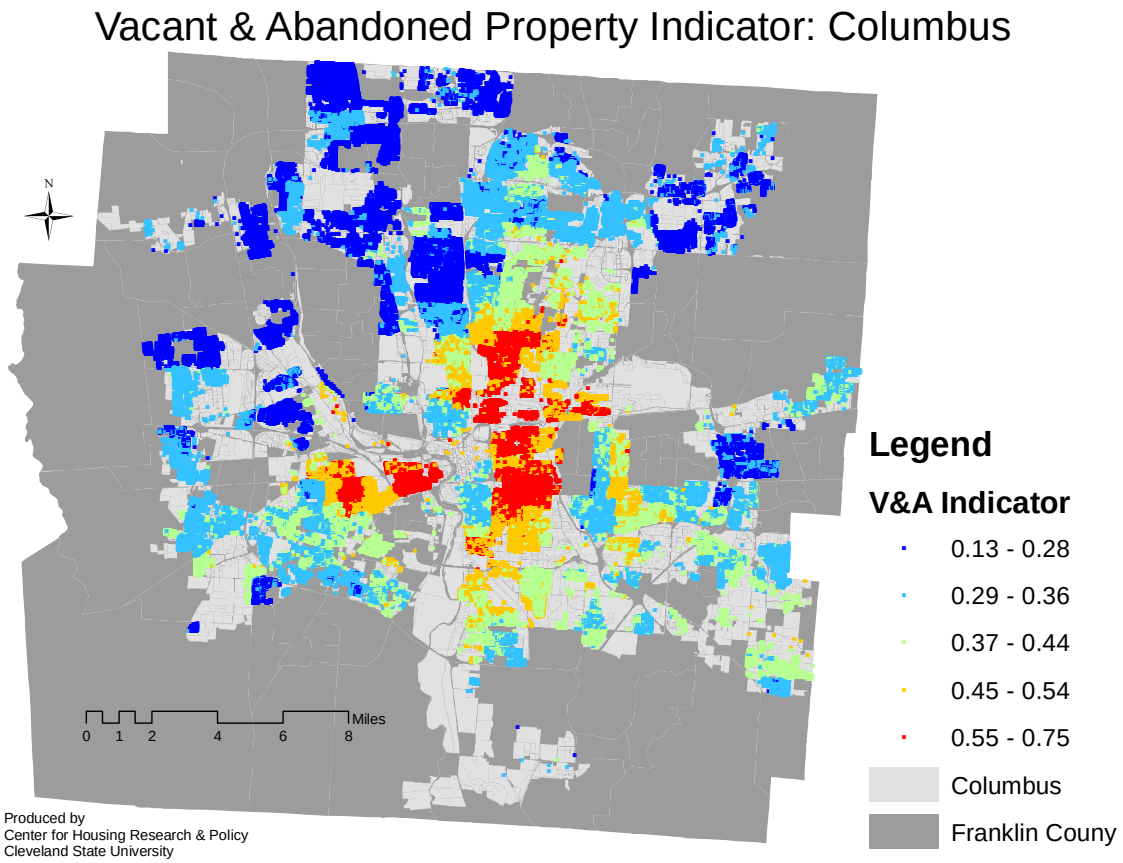

Figure 3. The spatial distribution of V&A indicator scores for Columbus.

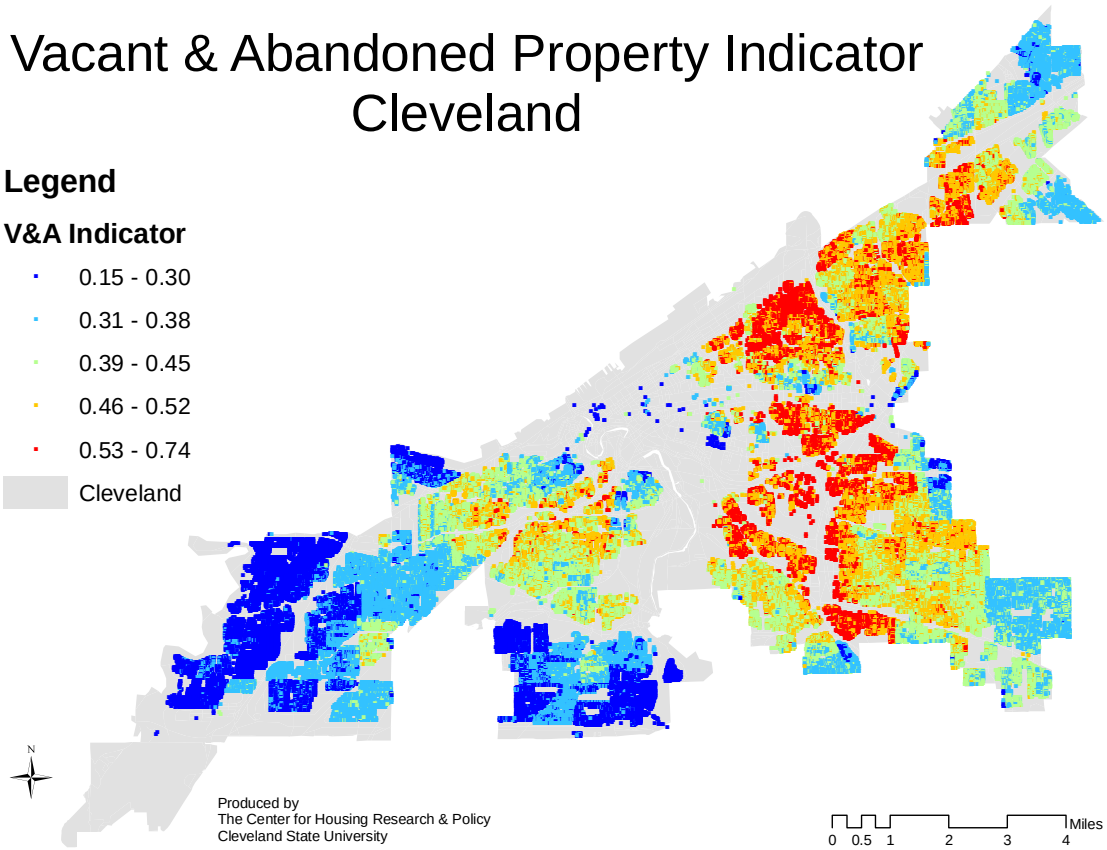

Figure 4. The spatial distribution of V&A indicator scores for Cleveland.

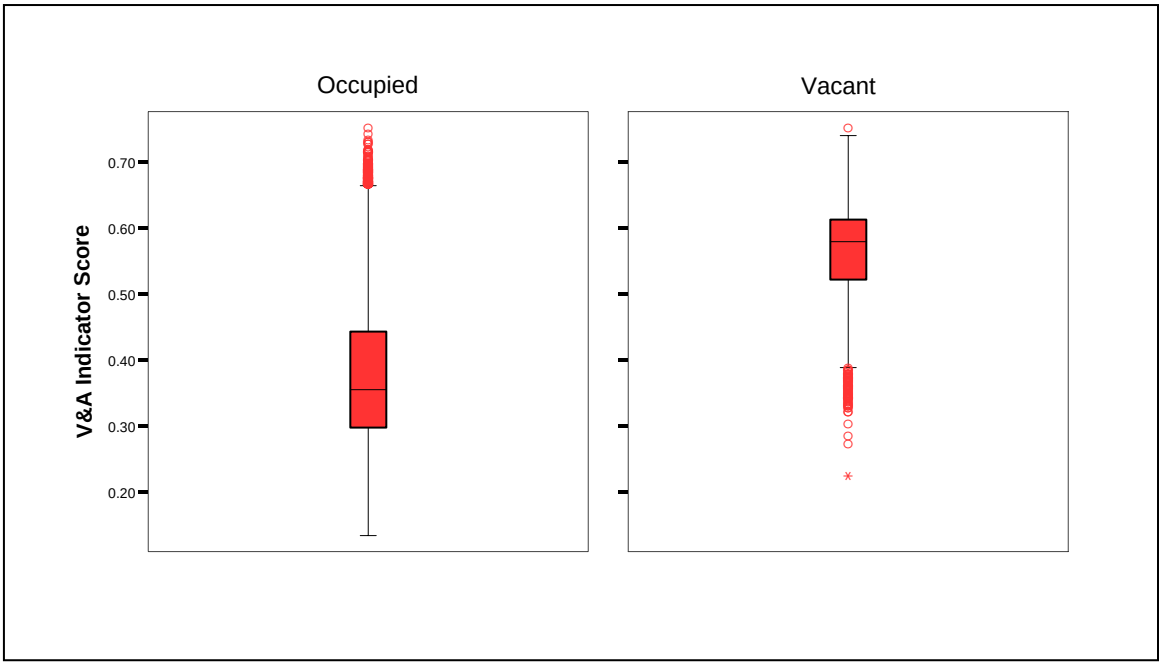

Figure 5. Box plots showing the distribution of the V&A Indicator score for occupied properties, and properties known to be vacant in Columbus.

Table 1a: Variable categorization for Cleveland

Input variable	If the value is GT	and LTET	Indicator Value
Age	110		1
	80	110	0.75
	50	80	0.5
	15	50	0.25
		15	0
Assessed value		25000	1
	25000	50000	0.75
	50000	100000	0.5
	100000	175000	0.25
	175,000		0
Price/Value ratio	3.75		1
	1.5	3.75	0.67
	0.5	1.5	0.33
		0.5	0
Appreciation		-15	1
	-15	0	0.5
	0	11	0.25
	11		0
Growth Ratio (change in housing units/change in households, 1990-2000)	4		1
	2	4	0.75
	-2	2	0.5
	-4	-2	0.25
		-4	0
Change in Percent Renter Occupied, 1990-2000	6		1
	0	6	0.67
	-6	0	0.33
		-6	0
Percent Renter Occupied, 2000	85		1
	70	85	0.75
	50	70	0.5
	25	50	0.25
		25	0

Table 1a: Variable categorization for Cleveland

Input variable	If the value is GT	and LTET	Indicator Value
Percent Housing Burden, Renters, 2000	65		1
	50	65	0.75
	35	50	0.5
	20	35	0.25
		20	0
Percent Housing Burden, Owner Occupiers, 2000	39		1
	27	39	0.67
	15	27	0.33
		15	0
Percent Single Parent Families, 2000	50		1
	5	50	0.5
		5	0
Percent Different house in 1995	70		1
	55	70	0.75
	40	55	0.5
	30	40	0.25
		30	0
Median Household Income		18000	1
	18000	27000	0.67
	27000	39000	0.33
	39000		0
Percent Population over 60	24		1
	18	24	0.67
	6	18	0.33
		6	0
Percent high priced home purchase loans (home owners)	65		1
	35	65	0.67
	10	35	0.33
		10	0
Percent high priced home purchase loans (non home owners)	65		1
	50	65	0.75
	20	50	0.5
	5	20	0.25

Table 1a: Variable categorization for Cleveland

Input variable	If the value is GT	and LTET	Indicator Value
		5	0
Percent high priced refinance loans (home owners)	80		1
	60	80	0.75
	40	60	0.5
	15	40	0.25
		5	0
Percent high priced refinance loans (non home owners)	40		1
	25	40	0.67
	5	25	0.33
		5	0
Percent refinance loans made by subprime lenders	60		1
	30	60	0.67
	10	30	0.33
		10	0
Percent home purchase loans make by subprime lenders	75		1
	50	75	0.67
	20	50	0.33
		20	0
Appreciation, tract level		0	1
	0	5	0.67
	5	15	0.33
	15		0

Table 1a, continued: Qualitative Input Variables

Qualitative Input Variables:	Input Value	Indicator Value
Condition	Unsound, Very Poor, Poor	1
	Good, Average, Fair	0.5
	Very Good, Excellent	0
Construction quality	D, D+, E, E+	1
	B, B+, C, C+	0.5
	A, A+, AA	0
Flipped	Yes	1
	No	0
Tax Delinquent	Yes, 2006	1
	Yes, 2000-2005	0.5
	No	0
Water Shut-Off	Shut-off or abandoned	1
	Previous shut-off	0.5
	Otherwise	0

Table 1b: Variable categorization for Columbus

Input variable	If the value is GT	and LTET	Indicator Value
Age	70		1
	30	70	0.5
		30	0
Assessed value		20000	1
	20000	110000	0.75
	110000	250000	0.5
	250000	600,000	0.25
	600,000		0
Price/Value ratio (pre-sale)	6.5		1
	1.5	6.5	0.67
	0.5	1.5	0.33
		0.5	0
Price/Value ratio (post-sale)	3.5		1
	1.45	3.5	0.67
	0.5	1.45	0.33
		0.5	0
Appreciation	30		1
	10	30	0.75
	5	10	0.5
	0	5	0.25
		0	0
Growth Ratio (change in housing units/change in households, 1990-2000)	3.5		1
	1.5	3.5	0.75
	1	1.5	0.5
	-2	1	0.25
		-2	0
Neighborhood vacancies	25		1
	15	25	0.75
	5	15	0.5
	0	5	0.25
		0	0

Table 1b: Variable categorization for Columbus

Input variable	If the value is GT	and LTET	Indicator Value
Neighborhood water shut-offs, 2004	30		1
	15	30	0.67
	0	15	0.33
		0	0
Neighborhood Foreclosures, 2006	60		1
	30	60	0.67
	1	30	0.33
		1	0
Change in Percent Renter Occupied, 1990-2000	2.4		1
	0	2.4	0.67
	-2.3	0	0.33
		-2.3	0
Percent Renter Occupied, 2000	85		1
	65	85	0.75
	36	65	0.5
	19	36	0.25
		19	0
Percent Housing Burden, Renters, 2000	40		1
	32	40	0.67
	26	32	0.33
		26	0
Percent Housing Burden, Owner Occupiers, 2000	25		1
	21	25	0.67
	17	21	0.33
		17	0
Percent Single Parent Families, 2000	50		1
	30	50	0.67
	10	30	0.33
		10	0
Percent Different house in 1995	75		1
	60	75	0.67
	30	60	0.33

Table 1b: Variable categorization for Columbus

Input variable	If the value is GT	and LTET	Indicator Value
		30	0
Median Household Income		20000	1
	20000	60000	0.67
	60000	90000	0.33
	90000		0
Percent Population over 60	30		1
	20	30	0.67
	5	20	0.33
		5	0
Percent high priced home purchase loans (home owners)	60		1
	40	60	0.67
	5	40	0.33
		5	0
Percent high priced home purchase loans (non home owners)	35		1
	15	35	0.67
	5	15	0.33
		5	0
Percent high priced refinance loans (home owners)	65		1
	55	65	0.75
	26	55	0.5
	5	26	0.25
		5	0
Percent high priced refinance loans (non home owners)	35		1
	15	35	0.67
	5	15	0.33
		5	0
Percent refinance loans made by subprime lenders	40		1
	20	40	0.67
	5	20	0.33
		5	0
Percent home purchase loans make by subprime lenders	35		1
	15	35	0.67

Table 1b: Variable categorization for Columbus

Input variable	If the value is GT	and LTET	Indicator Value
	3	15	0.33
		3	0
Appreciation, tract level		10	1
	10	30	0.67
	30	70	0.33
	70		0

Table 1b, continued: Qualitative Input Variables, Columbus

Qualitative Input Variables:	Input Value	Indicator Value
Condition	Poor	1
	Fair	0.75
	Average	0.5
	Good	0.25
	Very Good	0
Tax Delinquent	Yes, 2006	1
	Yes, 1982-2005	0.5
	No	0

Table 2: Categorized input variables

Variable	Data Level			Location	
Structure-specific Indicators	**Parcel**	**Nearby Parcel**	**Neighborhood**	**Columbus**	**Cleveland**
Age	✓	✓		✓	✓
Assessed value	✓	✓		✓	✓
Price/Value ratio	✓	✓		✓	✓
Condition	✓	✓		✓	✓
Construction quality	✓	✓			✓
Tax Delinquent	✓	✓		✓	✓
Water Shut-Off	✓	✓			✓
Appreciation	✓	✓		✓	✓
Flipped	✓	✓			✓
Local Housing Dynamics					
Growth Ratio			✓	✓	✓
Change in Percent Renter Occupied, 1990-2000			✓	✓	✓
Percent Renter Occupied, 2000			✓	✓	✓
Percent Different house in 1995			✓	✓	✓
Percent Population over 60			✓	✓	✓
Appreciation			✓	✓	✓
Vacancies			✓	✓	
Water Shut-Offs			✓	✓	
Foreclosures, 2006			✓	✓	
Economic Factors					
Percent high priced home purchase loans (home owners)			✓	✓	✓
Percent high priced home purchase loans (non home owners)			✓	✓	✓
Percent high priced refinance loans (home owners)			✓	✓	✓
Percent high priced refinance loans (non home owners)			✓	✓	✓
Percent refinance loans made by subprime lenders			✓	✓	✓
Percent home purchase loans make by subprime lenders			✓	✓	✓
Income					
Median Household Income			✓	✓	✓
Percent Housing Burden, Renters, 2000			✓	✓	✓
Percent Housing Burden, Owner Occupiers, 2000			✓	✓	✓
Percent Single Parent Families, 2000			✓	✓	✓