Estimating Land Values using Residential Sales Data

Stanley D. Longhofer
W. Frank Barton School of Business
Wichita State University
stan.longhofer@wichita.edu

Christian L. Redfearn
Price School of Public Policy
University of Southern California
redfearn@usc.edu

Abstract

Land prices are at the heart of urban economics but are generally not observed directly. Though they are central to household and firm location choices, land-only sales in urban areas are rare and often outliers. Indeed, urban areas are in part defined by a largely contiguous area of high land-use intensity – those places in which developable land is scarce. In this paper, we make use of more-common market data to infer land prices: house sales. Using locally weighted regressions, we estimate the value of a standardized structure across two urban counties: Maricopa, Arizona and Sedgwick, Kansas. Because the value of the standardized structure should be invariant across different locations in a metropolitan area, any remaining variation in the value surface should reflect differences in land values. By pinning down this surface using vacant lot sales at the periphery, we are able to extract land values throughout the metropolitan area, even in locations where vacant land sales are rare.

Current Draft: August 2022

We gratefully acknowledge financial support from the Lincoln Institute for Land Policy through its David C. Lincoln Fellowships in Land Value Taxation. We also thank Chris Morlan with the Sedgwick County Appraiser's Office for assistance with our data as well as the editor and two anonymous referees for helpful suggestions. We have also benefited from participating in Lincoln's Urban Economics and Public Finance Conference, Lincoln's conference on the special issue of Journal of Housing Economics on land price measurement, and the Homer Hoyt Institute's Weimer School Session.

1. Introduction

Land prices are at the heart of urban economics but are generally not observed directly. Though they are central to household and firm location choices, land-only sales in urban areas are rare and often outliers. Indeed, urban areas are in part defined by a largely contiguous area of high land-use intensity – those places in which developable land is scarce. In this article, we make use of more-commonly available market data to infer land prices: single-family house sales. Each house sale contains an underlying land parcel that contributes to the property price. Importantly, house sales are far more common and offer significantly more spatial coverage than do land-only sales. Our contribution is a novel method for extracting the level of land prices by exploiting information about location value that is implicitly embedded into each house sale. Our method reveals the benefits of acknowledging spatial and temporal variations in the implicit prices of housing characteristics. Doing so allows us to uncover a much richer land price surface. Our method makes use of standard tools and freely available software, making it a feasible and generalizable alternative to traditional approaches. Moreover, it obviates the reliance on strong assumptions about uniform implicit pricing and static fixed effects that are somewhat in conflict with the dynamic housing and land markets we are attempting to study.

It is very common to use location fixed effects in hedonic house price regressions. These are problematic when housing submarkets within metropolitan areas evolve asymmetrically over time. But more to the specific issue of land prices: the coefficients on the location dummies do not reveal the *level* of land prices. Rather, these coefficients are estimates of the location premium of a *housing bundle* in one place relative to another. In this article, we develop a statistically rigorous, straightforward, and tractable method by which vacant land sales on the periphery of a metropolitan area can be used to estimate underlying implicit land values throughout the city, including areas in which there are few or no vacant lot sales.

A practical barrier to measuring land prices is the fact that real estate is a bundled good, containing both land and structures. Moreover, both of these bundled goods are themselves heterogeneous. While land prices are determined in large measure by local amenities and disamenities that are external to a site, the ability to develop on a parcel in order to capture local amenities is partially determined by the parcel's size and shape. While there are regularities in this regard by neighborhood and vintage, they are far from homogeneous. And, of course, the structures that comprise the second part of the bundle also vary widely, as seen by the diversity of single-family residences across metropolitan areas. Existing technology addresses this heterogeneity but does not recover land prices. Our innovation lies in how we use this technology and the resulting estimated coefficients in conjunction with raw land sales to infer land prices throughout a metropolitan area.

Our analysis involves a two-step process. In the first step, we use single-family residential sales data to estimate a constant-quality price surface using a locally weighted regression (LWR) model. The key advantage of using LWR is its inherent flexibility to recover *local* pricing of dwelling attributes. By estimating the price function independently at each sale location, the Law of One Price is imposed, but only very locally; the data thus reveal spatial patterns in implicit prices of housing characteristics across the metropolitan area. As a result, this LWR methodology avoids the significant omitted variable problems that would otherwise plague the location fixed effect coefficients estimated in a traditional hedonic model. Using the coefficient vectors from the set of local regressions, we estimate a property value surface for a "standardized housing unit" at each and every location in the city. Importantly, this standardized housing unit is hypothetical and does not reflect the actual structures on parcels across the city. Instead, it is used to ask what each parcel would be worth if it had the same physical structure on it.

Because the value of the standardized structure (its construction cost less accrued depreciation) will not vary across the metropolitan area, this standardized property value surface implicitly measures the relative premia for different locations in the city. It does not, however, "pin down" precise land values for each property. That is, the standardized surface is obtained by pricing an identical housing bundle at every location, where this bundle includes both the typical dwelling characteristics and the typical lot characteristics. We recover land prices by fitting this standardized surface to the few land sales we do observe. Although there are relatively few sales of vacant lots in the core, there are many vacant lot sales at the periphery. Moreover, in many of these neighborhoods on the edge of the city, both vacant lot and completed dwelling sales occur at roughly the same time.

Using this fact, the second step of our analysis involves using vacant land sales in neighborhoods on the periphery to pin down the absolute level of the standardized value surface estimated in the first step. With this calibration, we can theoretically use the standardized value surface to estimate land values throughout the city, including in those neighborhoods without vacant lot sales.

In order to "pin down" the relative value surface, we identify several neighborhoods with both improved parcel and vacant lot sales. In these neighborhoods, we use the improved-parcel regression results to estimate the value of the standardized dwelling in these neighborhoods. Subtracting off the value of the typical lot in these neighborhoods provides an estimate of the implied value of the standardized improvements.

These steps are justified by the standard urban economics assumption that local amenities and disamenities should be capitalized into land values, not structure values. The locally weighted regressions allow the standardized structure to be priced everywhere, which explicitly holds constant the physical characteristics of improvements to the property at each location. By subtracting off the value of this standardized dwelling structure – which once again should be invariant across the community – we are able to make use of amenity capitalization in dwelling sales to recover land prices throughout the metropolitan area.

It is worth noting that while we make use of the tautology that the overall property value can be separated into its land and structure components additively, this imposes no restrictions on the functional form of the locally weighted regressions used to estimate these value surfaces. In our case, we use a traditional log-log specification to estimate housing values, but our methodology for extracting land values would work equally well regardless of the specific regression model specification employed.

The land surfaces we estimate are consistent with our understanding of two markets with distinctly different urban dynamics: Phoenix, Arizona and Wichita, Kansas. Phoenix is a rapidly growing metropolitan area with varied barriers to growth at its periphery. In contrast, Wichita grows very slowly and approximates the classic flat featureless plain of urban economic theory. Nevertheless, the land-value surfaces we estimate for these two disparate markets show striking similarities. In particular, they tend to show higher land values toward the center of the cities, especially in comparison with the spatial distribution of sale prices of single-family homes, which tend to show the highest values at the periphery of the city. This is consistent with the notion that smaller, older homes are located closer to the center of city on higher valued lots, while larger, high-end structures are built at the periphery on relatively lower valued lots.

In the next section we discuss the challenges associated with estimating lot values using single-family residential sales data, as well as recent related research. We contrast this research with our own empirical strategy. Section 3 contains a discussion of our empirical results, while Section 4 concludes by discussing the broader implications and caveats of our analysis.

2. Estimating Land Values using Residential Sales Data

Given that land is an essential component in urban and housing markets, it is somewhat surprising that relatively little empirical work has been done to address the practical challenges in accurately decomposing land and building values from residential sales data. Bostic, Longhofer and Redfearn (2007) and Davis and Heathcote (2007) demonstrated the importance of estimating land values in understanding property value dynamics at both the micro and macro levels. In the wake of these papers, numerous authors have attempted to estimate aggregate land value indices for different metropolitan areas and similar large-scale geographies.[1] For the most part, these papers focus on estimating market-level land leverage (the land-to-total property value ratio) but do not provide a direct method to estimate the land values of individual parcels.

Relatively few authors have attempted to estimate parcel-level land values from improved property sales. Gloudemans (2000, 2002) and Gloudemans, Handel and Warwa (2002) each attempt to use non-linear regression (hedonic) techniques to estimate land values from improved parcel sales data. Specifically, these papers model total property value as additive in its land and building components but *multiplicative* within the characteristics of each of these components. Because land and building values are separable in this model, they argue it is possible to use the regression coefficients to separately estimate land and building values, concluding that their land value estimates perform quite well relative to standard computer assisted mass appraisal benchmarks (the average assessed value-to-sale price ratio and the coefficient of dispersion of this ratio). In the end, these articles argue it is feasible to use multiple regression coefficients to directly estimate land values even when there are few or no vacant sales. As we discuss further below, however, these articles do not consider the extent to which their land value estimates are biased by omitted physical structure characteristics that are correlated with geography.

Francke and van de Minne (2017) develop a different non-linear hedonic pricing model and use it on data from the Netherlands to disentangle the value of the land and the value of the structure. In their model, land values are modeled as a function of lot size, location dummies, time of sale and property type, and are estimated simultaneously with the structure characteristics. Their primary focus is on measuring housing depreciation, however, with their land value modeling intended as a control to better estimate their variables of interest. As a result, they do not attempt to test the robustness of their land value

[1] Bourassa, et al. (2011), Chang and Chen (2011), Diewert, de Haan and Hendriks (2015), Haughwout, Orr and Bedill (2008), Nichols, Oliner and Mulhall (2013), Rambaldi, McAllister and Fletcher (2016), and Wong, et al. (2018), among others.

estimates. In particular, they too do not discuss whether their location dummy variables may be correlated with the physical characteristics of the structures in those neighborhoods.

Ashley, Plassmann and Tideman (1999) in many respects is most closely associated with our analysis. They estimate total property values of commercial properties in downtown Portland using a simple hedonic specification. They then use a quadratic spatial smoothing technique to estimate land value. In contrast, our work uses locally weighted regressions to estimate a total relative property value surface over a city, which we then "pin down" to actual values based on actual land values at the periphery of the city.

The Challenges with Hedonic Land Value Estimates

In order to understand the motivation behind our analysis and how it compares with prior research, it is worth reviewing the challenges inherent in using hedonic property value regression coefficients to directly estimate land values.

In its most basic form, the hedonic pricing equation takes a form such as

$$V = \alpha + \beta X + \gamma I + \delta Z \tag{1}$$

where V is the value (sale price) of the property, X is a vector of land characteristics (including lot size, street type, lot amenities, etc.), I is a vector of neighborhood dummy variables and Z is a vector of structure characteristics (building size, number of bedrooms, construction-quality variables, etc.). It is important to note that Z may include a number of interaction effects between structure and lot characteristics. Moreover, Z may need to include a number of locational interactions to account for the fact that the shadow prices of the physical characteristics of a house will vary across a metropolitan area. For example, a very small lot size might affect the value of additional square feet of the structure. Similarly, the value of an added bathroom or more square feet may differ based on the neighborhood in which the property is located.

Conceptually, we can decompose the value of a parcel into its lot value, L, and structure value, S, with $V = L + S$. Using this construct, one might want to use the physical lot and structure characteristics from (1) above to estimate L and S separately, with

$$L = \alpha_L + \beta X + \gamma I \tag{2}$$

and

$$S = \alpha_S + \delta Z, \tag{3}$$

where $\alpha = \alpha_L + \alpha_S$. That is, it is tempting to believe that the hedonic estimates from (1) can be applied to (2) and (3) to decompose the total property value into its land and building components.

Upon inspection, a number of problems with this approach become evident. Most obviously is the challenge inherent in allocating the intercept term α into its land and building components, α_L and α_S. This is not just a theoretical problem; the constant term in direct regression using sale price as the dependent variable can often exceed any reasonable estimate of total land value, indicating that it captures both land and building value components.

A second problem is apparent from the use of value (sale price) as the dependent variable in (1). Typically, the natural log of price is used as the dependent variable in most hedonic regressions. Not only does this transformation address the heteroskedasticity concerns so prevalent with housing data, it also results in regression coefficients that have a more natural and intuitive interpretation. The valuation model that underlies the log-linear specification, however, is multiplicative, not additive. As a result, land and building values are not separable in a traditional log-linear regression model, even if the "constant term" problem could be addressed.[2]

The most serious concern with using traditional hedonic regression coefficients to infer land values, however, is the likelihood that the estimated neighborhood coefficients, γ, will be biased because of a failure to accurately model the structure value characteristics, Z, in (1) above. The physical housing structures in most residential neighborhoods are generally quite homogeneous. For example, homes within a given neighborhood are likely to have similar sizes, floor plans and construction materials, reflecting the vintage of when they were built. Unless an extensive number of neighborhood/building characteristic interaction terms are incorporated in to Z, these variables will inevitably be highly correlated with the neighborhood indicator variables, I, making it likely that the estimated γ will be biased.[3]

In many hedonic applications, this may not be an inordinate problem, as the neighborhood dummy variables are implicitly included to control for any location-related factors, including both vintage effects

[2] As discussed above, Gloudemans (2000, 2002), Gloudemans, Handel and Warwa (2002) and Francke and van de Minne (2017) attempt to address this problem using non-linear regression techniques. Even so, their work does not address the more fundamental problem we address next.

[3] It should also be noted that data limitations may make it virtually impossible to include the requisite interaction terms, as there may be many neighborhoods with insufficient sales to permit estimation of both location fixed effects and interaction effects.

and locational amenities. If the goal is to use location fixed effect coefficients estimates as a proxy for land values, however, potential omitted variable bias becomes a more salient concern.[4] It is worth noting that this omitted variable bias is just as likely to exist in the non-linear regression models employed by Gloudemans (2000, 2002), Gloudemans, Handel and Warwa (2002) and Francke and van de Minne (2017) as it is in the simple formulation given in (1) above.[5]

This need to include numerous interaction terms in hedonic regressions is really a reflection of the fact that global pricing of attributes does not hold in urban areas with housing submarkets. Goodman and Thibodeau (1998, 2003, 2007) have demonstrated the existence of submarkets and the significant differences in the pricing of housing characteristics with urban areas. Redfearn (2009) showed that the Law of One Price could be easily rejected using data from Los Angeles. McMillen and Redfearn (2010) demonstrate that not only do prices of attributes vary spatially, but they vary in a way that is consistent with rational microeconomic behavior. In light of these findings, we abandon standard hedonic analysis in favor of a more flexible approach to controlling for quality differences across dwellings.

Our approach is based on work by Cleveland and Devlin (1988) and is called locally weighted regression (LWR). LWR was first used in a real estate context by Stock (1991) and Meese and Wallace (1991). The basic notion is that the implicit pricing of housing characteristics occurs locally, within submarkets. That is, the Law of One Price holds where buyers pursue similar dwellings, forcing sellers to adjust pricing accordingly. Because housing is a bundled good, there is little in the way of market forces to impel prices to be constant across all dwellings. Indeed, if prices for swimming pools were “too high” in one neighborhood, there is no practical way in which owners of swimming pools where prices for them are low to trade them in other markets. Hence, only local competition exists, and only local pricing should be consistent.

There are now numerous applications of LWR in practice: Bindanset & Lombard (2014), Borst & McCluskey (2008), and Cohen, Coughlin and Zabel (2020) all support the use of a more local perspective on housing markets. Malone and Redfearn (2020) demonstrate how these local dynamics work to induce significant bias in repeat-sales aggregate indexes like the Case-Shiller metropolitan prices indexes. Even

[4] One way of addressing this would be to run separate hedonic regressions for various submarkets within an urban area. This would reduce, but not eliminate, potential omitted variable bias.
[5] Indeed, the fact that Gloudemans (2002) includes a “vacant land” factor and find that vacant parcels are worth 30 percent less than otherwise identical improved parcels suggests that an omitted variable bias is a significant problem for his land value estimates.

more recently, Agarwal, et al. (2021) use LWR to document a remarkable evolution of house prices by submarkets within Singapore. The common theme is the relaxation of common implicit property characteristics.

The LWR begins with the identification of an observation and selects “neighboring sales” to be included in a regression that will estimate implicit prices for that house. “Neighboring” sales needn’t be just those that are closest physically; in this work, “close” is defined not only across latitude and longitude, but also size and date. In this way, the typical buyer search process is captured: buyers look at a particular point in time, within neighborhoods, at dwellings of a particular size. Specifically, in our model distance is given by:

$$d_{ij} = f(x\text{-coordinate}, y\text{-coordinate}, \text{living area}, \text{date of sale}), \tag{4}$$

where f is the Euclidean distance between subject property i and property j across the four dimensions in the distance equation. All variables are standardized so that each has mean 0 and standard variance of 1. Dwellings that are “close” under this approach are physically near one another (x, y), are the same size (above-grade living area and number of bedrooms) and are sold close to each other temporally.

These N “closest” observations are then used to estimate a weighted least squares regression using the tri-cubic function as weights:

$$w_{ij} = \left[1 - \left(\frac{d_{ij}}{\max(d_{ij})}\right)^3\right]^3 \tag{5}$$

In this way, the dwellings most like the subject dwelling get the most weight, with the weights declining in an accelerating manner.[6]

The LWRs are run at every dwelling sale observation. At each point, a set of implicit prices for a home’s physical characteristics at that particular location is estimated. As a result, a *surface* of implicit price estimates is obtained, rather than just a single implicit price estimate that would apply to the entire metropolitan area. This is possible because of the overlapping samples and resulting smoothness of the local regression coefficients. As the window sizes expand and the number of observations in each local regression is expanded, the resulting surface gets smoother. Indeed, the traditional pooled OLS hedonic specification can be thought of a local regression with a uniform weighting kernel that uses all the

[6] In the analysis below we also considered other weighting functions, including inverse distances, with essentially similar results. Essentially, any weighting function that penalizes distance achieves the same qualitative outcome.

observations; at each point, the same parameter estimates would be recovered. As such, the LWR nests the hypothesis that implicit prices are uniform throughout the market – the data can reveal this regularity, it is not imposed.

Given the local parameter estimates from the LWRs, a hypothetical standard dwelling can be priced at each location. In this way, quality is held constant. The resulting price surface is that of the standard dwelling everywhere within the geographic support of the data. Because this hypothetical standardized dwelling structure is the same at all locations by construction, the resulting values become a relative land value surface shifted by the value of the standard dwelling improvements. Of course, we never observe this structure value. Rather, we must use the sale prices of unimproved lots to calibrate the height of the value surface, backing into the constant value of the standard dwelling. If our assumption about amenity capitalization is correct – that capitalization of amenities is into land and not structures – then the pattern of land prices should be echoed in the standardized price surface. Subtracting off this constant value of the hypothetical standard improvements, therefore, provides an estimate of the land price surface.

To reiterate, our method for estimating a land value surface involves the following steps:

1. Estimate a LWR at the location of each single-family home sale within the county using a standard hedonic specification.
2. Define a standardized housing structure based on typical physical characteristics of housing within the county.
3. Predict the value of parcel with this standardized structure at each location using the LWR coefficients from Step 1.
4. Estimate the value of the standardized structure by subtracting the value of vacant land in locations where both improved parcel and vacant lot sales occur simultaneously.
5. Subtract the estimated standardized structure value from the values predicted in Step 3 to come up with a pure land value surface across the metropolitan area.

3. Data and Empirical Results

We conduct our analysis using data on residential sales from the central county in two vastly differing metropolitan areas: Phoenix, Arizona (Maricopa County) and Wichita, Kansas (Sedgwick County). Maricopa County's population at the 2020 Census was just over 4.42 million, an increase of 15.8 percent since the 2010 Census. In contrast, the population of Sedgwick County was just under 525,000 in 2020 and had grown by only 5.1 percent since the 2010 Census. While the Phoenix metro area is bounded by a number of mountains and American Indian reservations that may limit its growth, Wichita approximates the prototypical "flat featureless plain" of urban economic theory, with a perfectly elastic supply of land and no natural or legal barriers to new development. As such, we believe these two metropolitan areas bracket the range of dynamics experienced by cities across the U.S. and therefore demonstrate the applicability of our method to a variety of urban areas.

Data for both counties were provided by the respective county assessor offices, which each collect and maintain property characteristic and sales data for use in their computer assisted mass appraisal (CAMA) efforts following guidelines established by the International Association of Assessing Officers (IAAO). A description of the cleaning we did to prepare our data from each county can be found in the appendix.

We chose to study both counties through a window of relative housing market stability: 2014-2018. Our final Maricopa County samples include 309,012 improved parcel sales and 6,578 vacant lot sales, while our Sedgwick County sample consists of 28,309 improved parcel sales and 309 vacant lot sales. Table 1 provides an expositional description of the variables used in our analysis, while Tables 2 and 3 show summary statistics for each of these samples.[7] These summary statistics clearly demonstrate the dramatic differences between the two communities. Homes in Maricopa County were typically larger and newer and situated on smaller lots than those in Sedgwick County, and the median sale price was nearly 80 percent higher in Maricopa County. In contrast to improved parcels, vacant lot sales in Maricopa County were much larger than those in Sedgwick County, with a median price per square foot of land that was nearly two-and-a-half times higher.

The challenge facing this line of research can be clearly seen in Figures 1a and 1b, which show the spatial distribution of the improved property sales (gray dots) and the vacant lot sales we use to "tie-down" our property value surfaces (colored dots) in our cleaned data for both Maricopa (Figure 1a) and Sedgwick (Figure 1b) Counties. Both of these markets exhibit a pattern common to many, if not most, metropolitan

[7] It should be noted that a much richer set of property characteristics were available in each of our data sets that we chose not to use for two reasons. First, because the distribution of these characteristics tends to be highly correlated with geography, the exclusion or inclusion of each would depend on the data used in each local regression. More importantly, since our goal is to create a land value estimation procedure that is widely applicable to different communities, a parsimonious model specification is preferred.

areas: In many parts of the city, there are an insufficient number of improved parcel and vacant lot sales to allow estimating both land and improved property values in the same neighborhood, making it difficult to directly estimate the land values of the improved parcels.

Of course, there are differences between these two metropolitan areas as well. In Maricopa County, “tie-down” neighborhoods can be found throughout the metropolitan area, whereas in Sedgwick County these areas mostly exist on the periphery. While occasional vacant lot sales do occur in inner parts of the Wichita metropolitan area, they are typically associated with the activities non-profit organizations such as Habitat for Humanity, and it is worth questioning whether those lot sales really provide a good estimate of residential land values. In other instances, sales are to adjoining property owners or have been sold for purposes other than to build a single-family house. Importantly, in Wichita it is virtually unheard of to purchase a developed property with the intent to tear down and rebuild a new home on the lot; the periphery of the city is so close to the core that there is little benefit to doing inner-city tear downs, and it would be extremely rare to find a parcel (or adjoining parcels) where the underlying land value exceeds the value of the existing properties enough to justify the cost of tearing down the existing structures.

Figures 2 and 3 show the spatial variation in total finished living area and dwelling ages in our Maricopa and Sedgwick County improved property sale samples. As one would expect, the oldest and smallest homes are generally located in the core area of each of the cities, while the homes are newer and larger on the periphery. While the patterns are not identical – age is more consistently “in the center” in both cities – these characteristics exhibit strong spatial patterns in each of our metropolitan areas. Similar spatial variation could be seen across other housing characteristics as well.

This spatial variation in the characteristics could potentially be controlled for using standard approaches if their associated implicit prices were constant. Figures 4 and 5, however, demonstrate that this is not the case. These figures show the estimated LWR coefficients for living area and age across our two cities, demonstrating how sharply these implicit prices vary geographically.[8] Age, which is generally accepted as a small detriment to house prices due to depreciation in pooled regressions, has a significantly positive impact on sale price in some areas, while in nearby neighborhoods the marginal impact of age may seem unreasonably negative. In the latter, we think the age variable is picking up omitted

[8] In our LWR regressions, we chose to model age using a log instead of a quadratic specification. This choice has virtually no impact on the results and in particular does not explain the widely varying “age effects” we estimate across our two metropolitan areas. In the end, we chose to adopt the log specification because it tends to be more sedate with respect to forecast error.

characteristics. In the former, older homes may actually be more valuable in some historic neighborhoods. Given the wide variation in the shadow prices of these characteristics across different parts of the city – along with the paucity of vacant lot sales in the developed parts of the city – it becomes evident that attempts to use traditional hedonic methods to isolate the underlying land values of the improved parcel sales would be nearly impossible.

As a result, Figures 4 and 5 show that it is easy to reject the assumption that our data should be pooled and that a single set of metropolitan shadow prices should be imposed on dwelling characteristics. Consequently, uncovering true land values will require local pricing of a property's physical characteristics.

Table 4a shows the summary statistics from our Maricopa County LWRs for our sample period from 2014 to 2018. To reiterate, each of these regressions is centered around an individual sale using its 300 nearest neighbors as observations and uses the following hedonic specification:[9]

$$\begin{aligned} \ln(Price) = \beta_0 + \beta_1 \ln(Living\ Area) + \beta_2 \ln(Lot\ Size) + \beta_3 \ln(Age) \\ + \beta_4 Bath\ Fixtures + \beta_5 Quality + \beta_6 Gated + \beta_7 Golf + \beta_8 Cul\text{-}de\text{-}sac \\ + \beta_9 Arterial + \beta_{10} Mountain + \sum \gamma_j Quarter\ of\ Sale \end{aligned} \tag{6}$$

The entries in the table show the distribution of the regression coefficients across these 306,668 different regressions.[10] In the same way, Table 4b shows the summary statistics for the 27,859 Sedgwick County LWRs.

Focusing on the Maricopa results (Table 4a), the mean price elasticity of above-grade living area is 0.35. There is substantial variation across the city, however, ranging from 0.25 to 0.45 (10^{th} to 90^{th} percentiles). Similar variation can be found for each of the estimated coefficients in both of our markets. While there is considerable variation in these shadow prices across each of these counties, the average coefficients are all quite typical of what one would find in a traditional hedonic regression.

[9] Once again, the nearest neighbors are determined using the distance function defined in (4) above.
[10] In some of the Maricopa LWRs the coefficients for Quality and certain lot types could not be estimated because of a lack of variation in the data within that particular regression. For example, if none of the 300 observations in a particular regression included a mountain lot, that variable was omitted in that particular LWR. This explains why the count in the final column is smaller for some regressors. LWRs with fewer than 250 observations in the final estimation sample are excluded from the analysis. We also ran versions of our model using more and fewer "near neighbors" with similar results. Complete results from each of these regressions are available upon request.

4. Estimating a Standardized House Value Surface

As discussed above, our LWR results provide location-specific estimates of the shadow prices of a home's physical characteristics. In this section, we define a hypothetical standardized house and use the LWR coefficients to predict the value of this uniform structure at each location across the city. It is important to note that any particular standardized structure will not actually exist in every part of a city. Nevertheless, it is straightforward to predict its value at each location by simply applying the LWR coefficients to the standardized house's characteristics. As we will show below, using multiple standardized structures aids in estimating the final land value surface.

We define our preliminary standardized structures using the 50th percentile (median) values of each of the physical characteristics in our LWRs; these values are shown for each of our counties in the P50 columns in Tables 5a and 5b.[11] Thus, the Maricopa County P50 standardized dwelling is a 20-year-old, home with 1,980 square feet of finished living area on an 8,172 square foot lot with eight bathroom fixtures and a quality score of 3 ("Average"). Similarly, a Sedgwick County P50 standardized structure is a 25-year-old, 3-bedroom, 2-bath, 1,314 square foot home on a 9,496 square foot lot with an "Average" CDU grade (9).

Using these characteristics, we price this P50 standardized dwelling at the location of each of the LWRs, thereby creating spatial price distributions for dwellings that are identical by construction. It is important to remember that these standardized structures will not actually exist on the lots across our cities. Instead, they are used with each location's LWR coefficients to estimate the price at which this hypothetical dwelling *would* sell at each location if it were in fact there. Once again, our premise is that spatial variation in the prices of parcels that each have the same standard dwelling derives from local amenities and disamenities, which is ultimately attributable to the land value, not the structure.

Figures 6a and 6b contrast the spatial distribution of observed sales prices across Maricopa County and across Sedgwick County for the period from 2014-2018, while Figures 7a and 7b show the distribution of P50 standardized structure values for Maricopa and Sedgwick Counties over these same time frames. In order to understand these figures, note that the sale prices in Figures 6a and 6b are the actual market values of properties across each city. The variation in these market values reflects the combined effects of differences in structure characteristics and land values across the city. In contrast, Figures 7a and 7b show

[11] As we explain further below, we also create standardized structures based on the 20th, 35th, 65th and 80th percentile dwelling characteristics. We will refer to these as the P20, P35, P65 and P80 standard structures, respectively.

relative standardized single-family residential property value surfaces. Because the hypothetical standard structure is the same at each location (and hence has the same structure value across the city) differences in the heights of these surfaces reflect *only* relative differences in land values across the city.

Several points are worth noting in comparing these surfaces. First, as might be expected given the age and size of the homes found in these locations, the highest observed sale prices occur on the periphery of each city. For Maricopa County (Figure 6a), this is most notable in the East-Central and Northeast parts of the county, but it is also true in Southeast and Northwest Maricopa County as well. In Sedgwick County (Figure 6b) the highest observed sale prices occur on the far East and Northwest parts of the city. Second, for each of the two communities the standardized values (Figures 7a and 7b) tend to show much less variation across the metropolitan area. Moreover, while the same general locations tend show the highest values, high standardized values are more likely to be found toward the center of both communities. Upon reflection, this is not surprising. At the periphery very large homes are built on relatively cheap land. Closer in, smaller homes are generally found on more valuable land. With a standardized structure, however, the only remaining difference across locations is due to land value.

Calibrating the Relative Value Surface

Though we are interpreting the spatial variation in our standardized value surface as the manifestation of variation in local land premia, we cannot use the surfaces in Figure 7 to immediately recover underlying land values. This is because the standardized dwelling remains a bundle of both land and improvements. To recover land values, we need to use the information from our vacant land sales to "pin down" these surfaces. We do this by identifying neighborhoods with at least 275 improved parcel sales and 50 vacant lot sales.[12] For Maricopa County there were 27 such neighborhoods in the 2014-2018 era, while there were four such neighborhoods in Sedgwick County. Figures 1a and 1b identify these neighborhoods through their vacant lot sales (the colored dots). In these neighborhoods, we are able to directly estimate land values using the vacant lot sales.

For each sale in these neighborhoods, we used the LWR results to estimate the value of a standardized structure that had the median characteristics of all improved parcels our sample (the P50 structures shown

[12] Neighborhoods are defined using a grid of hexes. We went through many iterations of hex sizes, with little qualitative difference to the resulting land price estimates. The balance we struck was intended to provide sufficient improved and unimproved properties to measure our "tie-down" neighborhoods.

in Tables 5a and 5b). The median of these estimated standardized property values for each neighborhood are shown in the middle columns of Tables 6a and 6b for Maricopa and Sedgwick Counties, respectively. Once again, the figures in this column reflect the typical expected sale price of a parcel with the P50 structure in each of these neighborhoods. It is worth noting that these estimated standardized property values are always less than the median sale prices of improved parcels in the neighborhood. This reflects the fact that the P50 standardized structures are smaller and older homes than are typically present in neighborhoods where vacant lots are actively being sold.

To get the implied value of the standardized structure, we next subtract the median sale price of a vacant lot in each neighborhood (column 4) from the neighborhood's estimated standardized property value. The final standardized structure value estimates are shown in the final column of Tables 6a and 6b. In principle, this final estimated structure value should be the same across the entire community, and we interpret differences in these estimates across neighborhoods to reflect normal estimation error.[13]

Comparison of Tables 6a and 6b further highlights the dramatic ways our two communities differ. In Maricopa County, the typical house in "new development" neighborhoods in 2014-2018 sold for $337,500 and the typical lot cost $135,000. In contrast, the typical house in Sedgwick County neighborhoods with active new home development in 2014-2018 was $184,600 with a typical lot price of just $25,000.[14]

Turning to the implied standardized structure value estimates, although there are some extreme outliers, the Maricopa implied structure value is between $100,000 and $200,000 in nearly 56 percent of the neighborhoods, while all of the Sedgwick County estimates fall within a fairly tight range.[15] Our final estimated P50 standardized structure value is the median of the estimates across these tie-down neighborhoods. In Maricopa County, this value is $114,217, while in Sedgwick County it is $120,882. At

---

[13] Two obvious outliers are Neighborhoods 15 and 16 in Maricopa County, which each have negative implied standardized land values. Closer examination of Table 6a reveals that this is due to the extraordinarily high sale prices of vacant lots in these neighborhoods. It turns out these two neighborhoods are both in the very affluent area east of the Phoenix Mountains Reserve. Furthermore, it appeared that many of the vacant lot sales in these neighborhoods may have been wholesale transactions that would ultimately be subdivided into multiple parcels. Omitting these neighborhoods from the analysis would ultimately have virtually no impact on the results that follow; without these two neighborhoods the median standardized structure value would have been $114,605 rather than $114,217.

[14] One might question why the estimated standardized property values are nearly always less than the median property values in these neighborhoods. This is due to the fact that the P50 standardized structures reflect smaller and older homes than those that are typically present in the tie down neighborhoods.

[15] The outliers generally arise in cases where the structure characteristics used to predict the standardized property value are outliers for the neighborhood in question.

first glance, these implied standardized structure values might seem low. Remember, however, that they reflect depreciated structure values, not the cost of building them as new structures.

The final step in our analysis is to use these estimated standardized structure values to derive a land value surface across the metropolitan area. As discussed above, these estimated structure values should be spatially invariant, with local amenities capitalized into land and not structure prices. We therefore subtract this (uniform) standardized structure value ($114,217 in Maricopa County and $120,882 in Sedgwick County) from the standardized dwelling value surface derived from our LWRs to get the estimated value of land for each of the parcels across the entire county. The distribution of these estimated land values is shown by the green (P50) plot in Figures 8a and 8b for each of our counties.[16]

The distribution of estimated land values across Maricopa County seems quite reasonable, with typical values ranging from $0 to $40 per square foot. The Sedgwick County estimates, on the other hand, are more problematic. While the majority of parcels have very plausible land value estimates, a significant number of them have negative implied land values. Upon investigation, it turned out that the negative land value estimates often arose from LWRs where the characteristics of the P50 structure were out of sample, raising questions about the ability of these regressions to estimate the value of the P50 structure. Considered another way, however, these negative implied land values may be reflecting the fact that a P50 structure may be an over-improvement in many older neighborhoods, meaning that a structure with these characteristics would not be the highest and best use (in terms of size and features) for a parcel in that neighborhood. In either case, the P50 structure would not a good benchmark to use to back out underlying land values in these particular neighborhoods.

In the appraisal literature, it has long been understood that a parcel's land value is its value under its highest and best use as though vacant, irrespective of the parcel's current use. This derives from the premise that different users will bid for the land, with the one valuing it most determining its ultimate price, consistent with standard urban economic models. This concept remains true for developed parcels as well. Because improvements are typically long-lived, the value of the underlying land can vary from its value under its current use as market conditions and amenities around the parcel change over time. If the land's value under a competing land use is large enough, such a user may be willing to buy the parcel and tear down the existing improvements in order to convert the parcel to its highest and best use. In many instances, however, the land value may exceed its value under the current use but be insufficiently high to

[16] The reason for the other distributions in these figures will become clear in a moment.

justify a teardown.[17] Nevertheless, the value of the land is based on the highest and best use, not the current (suboptimal) use.

Longhofer (2021) shows that when the current use is not the parcel's highest and best use, any resulting value loss should be attributed to the structure in the form of "external obsolescence" and not to the land. As an extreme example, consider a Beverly Hills-style mansion being transported to a low-income neighborhood in South Los Angeles. The market value of the mansion in this neighborhood would be well below its structure cost. This would not cause the land value to become negative, as other potential land uses would still be willing to pay a positive price for the land if it were vacant. Instead, the resulting value loss should be attributed to the structure itself because such a building would be misplaced in this neighborhood. Now, this is intended to be an extreme hypothetical example and the fact that no one would build a mansion in such a neighborhood is beside the point. As noted above, it is often the case that existing structures differ from those that would be built today. The key point is that value loss that results from this "wrong" structure on the site is attributable to the structure, not the land.[18]

Our land value estimation technique implicitly assumes that the standardized structure is the highest and best use for every parcel across the city, an assumption that is clearly not accurate. As a result, our method inappropriately attributes the resulting value loss to the land when it should be attributed to the P50 structure. To address this problem, we defined four additional standardized structures based on the 20^{th}, 35^{th}, 65^{th} and 80^{th} percentiles of structure characteristics across our two counties. These characteristics are shown in Table 5a and 5b and are referred to as structures P20, P35, P65 and P80, respectively. Our intuition is that the land value at any given location will be determined by the structure that would willing to bid the most for the land (the parcel's "highest and best structure"). As a result, the true land value surface should be the upper envelope of the land value surfaces implied by each of these standardized structures. This is exactly what happens in the classic bid-rent model of urban land rents. The rent at a particular location – and hence the location's land value – is determined by the user that is willing to pay the highest price to be in that location. That is, the land value is determined by the upper envelope of its values under various potential uses.[19]

[17] The value of the option for future development may affect this choice as well.

[18] In fact, a similar thing would happen if a very small, non-confirming structure existed in Beverly Hills. In this instance, the value of the property as a whole would likely be its land value, and the structure, while still having some remaining economic life, would be virtually worthless because it would be the wrong structure for the site. In cases like this, a tear down would be expected.

[19] Because we use only five standardized structures in our analysis, we can only approximate the highest and best use of each

Following the methods outlined above, we estimated the cost of each of these standardized structures in our tie-down neighborhoods and used these estimates to derive five different land value surfaces for each county. Figure 8 shows the distribution of land values implied by each of these structures for each of our counties. For Maricopa County (Figure 8a), the distributions of estimated land value surfaces are relatively fairly similar across all standardized structure types.[20] Sedgwick County, on the other hand, exhibits much more variation across the different standardized structure types (Figure 8b).

As discussed above, because land values are determined by the highest and best use of a parcel, our final land value estimate in each county is the upper envelope of the values implied by the five standardized structures. The distribution of these land values per square foot for each county are depicted in Figures 9a and 9b. Once we account for each parcel's highest and best use structure, the resulting distributions of estimated land prices match well with our understanding of land values in each of the two cities. The vast majority of locations in Maricopa County have estimated land values falling between $0 and $50 per square foot, with a few extreme outliers on each end.[21] Estimated land values in Sedgwick County exhibit a very similar pattern, albeit with lower values (between $0 and $20 per square foot for most locations).

The spatial distribution of these land value estimates for the two counties are shown in Figures 10a and 10b. While there are clearly some outliers, many of the "peaks" shown in Figures 10a and 10b correlate well with various pockets of high-end neighborhoods throughout both of these cities.[22]

It is especially instructive to compare Figures 10a and 10b with Figures 6a and 6b, which showed the spatial distribution of actual sale prices, a distribution that was the result of the combined effects of land values and the structures actually found on the land. Figures 10a and 10b, in contrast, are true land value surfaces, reflecting the fact that we are controlling both for the spatial variation in the housing stock and the spatial variation in the implicit prices of the structures themselves. In each of our counties, the spatial distribution of land values is generally consistent with the distribution of actual sale prices. Careful

parcel, a tradeoff we accept for tractability.

[20] It is important to remember in this figure that a particular parcel's location within the distribution will change for each standardized structure type. It is conceivable that a parcel may be at the at the top end of the land value distribution for one structure type and in the opposite tail for another.

[21] Presumably these outliers could be further limited through the use of more standardized structure types to better reflect the highest and best use of each parcel.

[22] The most notable outlier of which we are aware is the peak in Southeast Wichita, which shows very high land values in a low end neighborhood. Further inspection of the LWRs in this area suggested that the uniform model specification performed particularly poorly in these areas, especially with respect to the coefficient on age. This issue could potentially be addressed by allowing the LWR specification to vary across the metropolitan area. We leave this as a topic for future research.

inspection, however, reveals that land values are relatively higher in more centrally located areas. These higher interior land values seem to correspond with the location of the more traditional affluent older neighborhoods in the city. While land values still remaining relatively high on the periphery (particularly in the Northeast of both communities), the relative “land premium” in these neighborhoods is clearly much lower than the overall “sale price premium.” In other words, in these periphery areas, very large, expensive structures tend to be built on relatively comparatively less expensive land, consistent with what one would expect.

5. Conclusions

The goal of this paper was to develop a rigorous yet tractable method for extracting urban land values in developed areas with few vacant land sales. Our intuition is straightforward: while dwelling sales cannot be directly used to estimate land values, they can be used to estimate location premia. These premia can be used in conjunction with land sales to recover underlying land prices even in neighborhoods with no vacant land sales. The statistical tool we use to estimate the location premia is locally weighted regression (LWR).

This is the first contribution of our project. Although the use of more flexible modeling statistical methods is becoming more common, “standard” hedonic regression remains the dominant tool of empiricists using housing data. In some applications, this is entirely appropriate, but the imposition of fixed implicit prices on housing characteristics is a strong assumption that is generally untested. Using LWR techniques, we are able to estimate unique shadow prices for various housing components at each location throughout the city. These LWRs confirm that, indeed, the implicit prices of a home’s physical characteristics vary considerably across the metropolitan area. In light of this result, it is clear that the use of fixed coefficients is a misspecification that can bias parameters used to inform policy.

Using the estimated LWR coefficients, we then proceed to predict the value of standardized dwelling units across the city. Because the physical characteristics of these structures are the same at each and every location (and hence should have the same structure values), the resulting value surface is, in fact, a measure of the relative value of land at each location. That is, we assume that local amenities are capitalized into land and not structures, meaning that spatial variation in the price surface derived from standard dwellings must be due to variation in land values.

Our final step is to "pin down" this relative value surface using vacant land sales (which primarily occur on the periphery of the city). We do this by estimating the value of the standard housing units in neighborhoods with vacant lot sales and subtract the observed selling prices of the lots to come up with the "value" of the standard structures. Subtracting this imputed structure value from our standardized price surfaces provides us with estimates of absolute land values at each location throughout the city.

What remains to do with this research is to apply to a set of taxation rules. We have applied an economist's perspective in minimizing errors when measuring land prices. To the extent that local taxing authorities want to minimize lawsuits about assessment errors, it might be possible to rethink some of our choices in assigning values at each parcel. This work is left for future research. We see the major contribution in this paper is the demonstration of the value of the rich land data bundled in improved property sales and the need to accommodate the reality of neighborhood variation – both in terms of the housing stock and the way these attributes are priced. We find that the results are robust to the particulars of the methods, and that local pricing and moving beyond vacant parcel sales are central to understand land prices in urban areas.

6. Appendix – Data Cleaning Procedures

Maricopa County Data

Improved parcel sales were identified as transactions classified with a "Single-Family Residential" property type containing a single living unit with at least 400 square feet of finished living area. We eliminated parcels with lots smaller than 6,000 square feet (the minimum developable lot size under the current zoning code) and those larger than 100,000 square feet. Finally, we eliminated a handful of parcels that were spatially disjoint from the rest of the data used in the analysis.

Vacant residential lot sales were identified in a similar manner, using "Vacant land" sales of parcels indicating a "Residential" land use. As with the improved parcel sales, we eliminated lots smaller than 6,000 square feet and larger than 100,000 square feet.

Sedgwick County Data

Improved parcel sales were identified as transactions classified as a "Land & Building" sale of a "Residential" parcel involving a "Single-Family Residence" in a "Neighborhood or Spot" location, to eliminate parcels that may have been intended for commercial uses. We further restricted the data to parcels with a single living unit with at least 400 square feet of finished living area and no more than 10 bedrooms or 10 full bathrooms. We also eliminated parcels with lots smaller than 4,000 square feet (the minimum developable lot size under the current zoning code) and those larger than 100,000 square feet. Finally, we eliminated a handful of parcels that were spatially disjoint from the rest of the data used in the analysis.

Vacant residential lot sales were identified in a similar manner, using "Land Only" sales for parcels with a property class equal to "Residential" or "Vacant Lot" in a "Neighborhood or Spot" location. As with the improved parcel sales, we eliminated lots smaller than 4,000 square feet and larger than 100,000 square feet, as well as observations coded as having one or more existing living units.

7. References

Agarwal, Sumit, Ying Fan, Daniel P. McMillen and Tien Foo Sing. 2021. “Tracking the pulse of a city—3D real estate price heat maps.” Journal of Regional Science 61(3), 543-569.

Ashley Richard, Forenz Plassmann and Nicholaus Tideman, 1999. “Improving the Accuracy of Downtown Land Assessment.” Lincoln Institute of Land Policy Working Paper WP99RA1.

Bidanset, P. E. and J.R. Lombard, 2014. Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. Cityscape, 16(3), 169–182

Borst, R. A. and W. J. McCluskey, 2008. Using geographically weighted regression to detect housing submarkets: Modeling large-scale spatial variations in value. Journal of Property Tax Assessment & Administration, 5(1), 21–21

Bostic, Raphael W., Stanley D. Longhofer and Christian L. Redfearn, 2007. “Land Leverage: Decomposing Home Price Dynamics.” *Real Estate Economics*, vol. 34, no. 2, pp. 183-208.

Bourassa, Steven C., Martin Hoesli, Donato Scognamiglio and Sumei Zhang, 2011. “Land Leverage and House Prices.” *Regional Science and Urban Economics*, vol 41, no. 2, 134-144.

Chang, Yan and Jian Chen, 2011. “A Consistent Estimate of Land Price, Structure Price, and Depreciation Factor.” Freddie Mac Working Paper.

Cohen, J. P., C. C. Coughlin and J. Zabel, 2020. Time-geographically weighted regressions and residential property value assessment. The Journal of Real Estate Finance and Economics, 60(1), 134– 154.

Cleveland, William S. and Susan J. Devlin, 1988. “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting.” *Journal of the American Statistical Association, vol. 83, pp. 596-610.*

Davis, Morris A. and Jonathan Heathcote, 2007. “The Price and Quantity of Residential Land in the United States.” *Journal of Monetary Economics*, vol. 54, no. 8, pp. 2595-2620.

Diewert, W. E., J. de Haan and R. Hendriks, 2015. “Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components.” *Econometric Reviews*, vol. 34, nos. 1–2, pp. 106–126.

Francke, Marc K. and Alex M. van de Minne, 2017. “Land, Structure and Depreciation.” *Real Estate Economics*, vol. 45, no. 2, pp. 415-451.

Goodman, Alan C. and Thomas G. Thibodeau, 1998. “Housing Market Segmentation.” Journal of Housing Economics vol. 7, pp. 121–143.

Goodman, Alan C. and Thomas G. Thibodeau, 2003. “Housing Market Segmentation and Hedonic Prediction Accuracy.” Journal of Housing Economics. Vol. 12, pp. 181–201.

Goodman, Alan C. and Thomas G. Thibodeau, 2007. “The Spatial Proximity of Metropolitan Area Housing Submarkets.” Real Estate Economics. Vol. 35, pp. 209–232.

Gloudemans, Robert J., 2000. “Implementing a Land Value Tax in Urban Residential Communities.” Lincoln Institute of Land Policy Working Paper WP00RG1.

Gloudemans, Robert J., 2002. “An Empirical Analysis of the Incidence of Location on Land and Building Values.” Lincoln Institute of Land Policy Working Paper WP02RG1.

Gloudemans, Robert J., Sheldon Handel and Michael Warwa, 2002. “An Empirical Evaluation of Alternative Land Valuation Models.” Lincoln Institute of Land Policy Working Paper WP02RG2.

Haughwout, Andrew, James Orr and David Bedoll, 2008. “The Price of Land in the New York Metropolitan Area.” *Current Issues in Economics and Finance*, vol. 14, no. 3.

Longhofer, Stanley D., 2021. “Land Values and External Obsolescence.” *The Appraisal Journal*, vol. 89, no. 2, pp. 95-103.

Malone, T. and C. L. Redfearn, 2020. To measure globally, aggregate locally: Urban land, submarkets, and biased estimates of the housing stock. Real Estate Economics, forthcoming.

McMillen, Daniel P. and Christian L. Redfearn, 2010. “Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions.” *Journal of Regional Science*, vol. 50, no. 3, pp. 712-733.

Meese, Richard and Nancy Wallace, 1991. “Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices.” *Real Estate Economics,* vol 19, no. 3, pp. 308-332.

Nichols, Joseph B, Stephen D. Oliner and Michael R. Mulhall, 2013. “Swings in Commercial and Residential Land Prices in the United States.” *Journal of Urban Economics*, vol. 73, no. 1, pp. 57-76.

Rambaldi, Alicia N., Ryan R. J. McAllister and Cameron S. Fletcher, 2016. “Decoupling Land Values in Residential Property Prices: Smoothing Methods for Hedonic Imputed Price Indices.” University of Queensland School of Economics Working Paper.

Redfearn, Christian L., 2009. “How Informative are Average Effects? Hedonic Regression and Amenity Capitalization in Complex Urban Housing Markets.” *Regional Science and Urban Economics,* vol. 39, no. 3, pp. 279-306.

Stock, James H., 1991. *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press.

Wong, Siu Kei, K.W. Chau, K. Karato and C. Shimizu, 2018. “Separating the Age Effect from a Repeat Sales Index: Land and Structure Decomposition.” *Journal of Real Estate Finance and Economics,* vol. 57, pp. 351-366.

8. Tables and Figures

Table 1 – Description of Variables

Variable	Description	Sample
Sale price	Sale price of the vacant lot or improved property sale	Both
Age of improvements	Year of sale minus the year the improvements were constructed; may be negative if the home was sold while the improvements were under construction	Both
Arterial fronting	A lot located on an arterial road	Maricopa only
Bathroom fixtures	Number of bathroom fixtures (bathrooms sinks, toilets, showers, tubs, etc.)	Maricopa only
Full bathrooms	Total number of full bathrooms	Sedgwick only
Half bathrooms	Total number of half bathrooms	Sedgwick only
Bedrooms	Total number of bedrooms in the home	Sedgwick only
CDU grade	A rating reflecting the physical condition, utility and desirability of the property (including location); values range from 0 (“Unsound/Undesirable) to 15 (“Excellent”); comparable to quality in the Maricopa data	Sedgwick only
Living area in SF	Total square feet of finished living area	Both
Lot size in SF	Total square feet of land area in the parcel	Both
Cul-de-sac lot	A lot located on a cul-de-sac	Both
Enhanced view lot	A lot with a view of a water feature or golf course, but not directly adjacent to one of these amenities	Sedgwick only
Golf course lot	A lot adjacent to the fairway or green of a golf course	Both
Mountain lot	A lot located on a mountain	Maricopa only
Quality	Residential quality class; values range from 0 to 7 with 3 being average and 7 the highest; comparable to CDU grade in the Sedgwick data	Maricopa only
Quarter of sale	The quarter in which the sale of the property occurred	Both
Walkout ranch lot	A lot with sufficient grading to allow for a ranch style home with a basement that has an exterior door (considered to be a premium lot in the market)	Sedgwick only
Waterfront lot	A lot adjacent to a lake or other water feature	Sedgwick only

Note: Variables available differ somewhat between the two samples; the third column indicates for which sample each variable is available.

Table 2a – Maricopa County Improved Parcel Sales Summary Statistics

Variable	Mean	Std.Dev.	Min.	25th Percentile	Median	75th Percentile	Max.
Sale price	$330,478	$269,308	$10	$197,000	$272,000	$380,000	$11,200,000
Living area square feet	2,189	900	400	1,569	1,997	2,609	15,200
Land square feet	11,270	10,183	6,000	7,022	8,187	10,302	99,840
Age of improvements	26.4	19.7	-1	12	21	40	118
Bathroom fixtures	8.5	3.1	0	6	8	11	48
Quality	3.5	0.7	0	3	3	4	7
Gated community	0.0	0.2	0	0	0	0	1
Golf course lot	0.0	0.2	0	0	0	0	1
Cul-de-sac lot	0.045	0.207	0	0	0	0	1
Arterial fronting	0.063	0.243	0	0	0	0	1
Mountain lot	0.007	0.085	0	0	0	0	1

Notes: Sample includes 309,012 improved parcel sales from 2014 through 2018. Age is calculated as the year of sale less the year the home was built. For new homes sold while under construction, the calculated age may be negative.

Table 2b – Sedgwick County Improved Parcel Sales Summary Statistics

Variable	Mean	Std.Dev.	Min.	25th Percentile	Median	75th Percentile	Max.
Sale price	$174,578	$113,956	$3,600	$107,000	$152,500	$212,500	$1,850,000
Living area in SF	1,459	534	420	1,104	1,352	1,674	8,411
Lot size in SF	11,705	7,719	4,005	7,892	9,705	12,586	99,840
Age of improvements (years)	36.4	27.9	0	13	30	60	146
Bedrooms	3.4	1.0	1	3	3	4	8
Full bathrooms	2.2	0.9	1	2	2	3	9
Half bathrooms	0.2	0.5	0	0	0	0	5
CDU grade	9.5	1.3	1	9	9	10	15
Walkout ranch lot	0.103	0.303	0	0	0	0	1
Waterfront lot	0.066	0.247	0	0	0	0	1
Golf course lot	0.011	0.102	0	0	0	0	1
Enhanced view lot	0.007	0.083	0	0	0	0	1
Cul-de-sac lot	0.069	0.253	0	0	0	0	1

Notes: Sample includes 28,309 improved parcel sales from 2014 through 2018. Age is calculated as the year of sale less the year the home was built. For new homes sold while under construction, the calculated age may be negative.

Table 3a – Maricopa County Vacant Lot Sales Summary Statistics

Variable	Mean	Std.Dev.	Min	25th Percentile	Median	75th Percentile	Max
Price	$217,092	$294,430	$1,950	$65,000	$125,000	$240,000	$5,000,000
Lot SF	37,244	22,013	6,000	18,009	38,642	49,484	99,991
Price PSF	$7.32	$9.88	$0.04	$2.09	$4.37	$9.27	$212.91
Gated community	0.079	0.270	0	0	0	0	1
Golf course lot	0.048	0.213	0	0	0	0	1
Cul-de-sac lot	0.105	0.307	0	0	0	0	1
Arterial fronting	0.091	0.287	0	0	0	0	1
Mountain lot	0.039	0.194	0	0	0	0	1

Notes: Sample includes 9,740 vacant parcel sales from 2014 through 2018.

Table 3b – Sedgwick County Vacant Lot Sales Summary Statistics

Variable	Mean	Std.Dev.	Min	25th Percentile	Median	75th Percentile	Max
Price	$33,425	$35,565	$5,500	$21,000	$29,000	$36,500	$575,759
Lot SF	15,374	13,030	5,229	10,158	11,690	14,419	91,476
Price PSF	$2.62	$1.48	$0.37	$1.71	$2.18	$3.41	$7.93
Walkout ranch lot	0.003	0.057	0	0	0	0	1
Waterfront lot	0.275	0.447	0	0	0	1	1
Golf course lot	0.006	0.080	0	0	0	0	1
Enhanced view lot	0.003	0.057	0	0	0	0	1
Cul-de-sac lot	0.233	0.423	0	0	0	0	1

Notes: Sample includes 309 vacant parcel sales from 2014 through 2018.

Table 4a – Maricopa County Key LWR Summary Statistics

Variable	Mean	Std. Dev.	10th Percentile	Median	90th Percentile	Count
Ln(Living area)	0.35	0.16	0.25	0.35	0.45	306,668
Ln(Lot size)	0.12	0.10	0.06	0.12	0.18	306,668
Ln(Age)	-0.07	0.13	-0.12	-0.06	-0.02	306,668
Bathroom fixtures	0.01	0.02	0.00	0.01	0.01	306,622
Quality	0.06	0.12	0.01	0.05	0.10	241,344
Gated community	0.03	0.13	-0.03	0.03	0.09	88,169
Golf course lot	0.10	0.11	0.04	0.10	0.16	95,034
Cul-de-sac lot	0.00	0.08	-0.04	0.00	0.03	257,599
Arterial fronting	-0.04	0.07	-0.07	-0.03	0.00	296,672
Mountain lot	0.03	0.14	-0.03	0.02	0.09	19,882
Constant	8.68	1.56	7.72	8.61	9.60	306,668
R-square	0.58	0.15	0.47	0.59	0.69	306,668

Notes: Columns show the summary statistics for the estimated locally weighted regression (LWR) coefficients from 306,668 locally weighted regressions for sales from 2014 to 2018; output for quarterly indicator variables are omitted; the dependent variable in these regressions is the natural log of price; for some LWRs the coefficients for quality and certain lot types could not be estimated because of missing data.

Table 4b – Sedgwick County Key LWR Summary Statistics

Variable	Mean	Std. Dev.	10th Percentile	Median	90th Percentile	Count
Ln(Living area)	0.51	0.23	0.34	0.47	0.65	27,859
Ln(Lot size)	0.10	0.09	0.04	0.09	0.15	27,859
Ln(Age)	-0.21	0.16	-0.26	-0.17	-0.12	27,859
Bedrooms	0.01	0.03	0.00	0.01	0.03	27,859
Full bathrooms	0.08	0.05	0.05	0.08	0.11	27,859
Half bathrooms	0.04	0.07	0.00	0.04	0.08	27,859
CDU grade	0.07	0.04	0.05	0.07	0.09	27,859
Constant	7.07	1.90	5.81	7.27	8.45	27,859
R-square	0.71	0.09	0.65	0.71	0.78	27,859

Notes: Columns show the summary statistics for the estimated locally weighted regression (LWR) coefficients from 27,859 locally weighted regressions for sales from 2014 to 2018; output for quarterly indicator variables are omitted; the dependent variable in these regressions is the natural log of price.

Table 5a – Maricopa County Standardized Structure Characteristics

Variable	P20	P35	P50	P65	P80
Living area in SF	1,461	1,713	1,980	2,294	2,811
Lot size in SF	6,825	7,445	8,172	9,210	11,158
Age of improvements (years)	43	32	20	14	9
Bathroom fixtures	6	7	8	9	11
Quality	3	3	3	4	4

Note: P20 is a structure with the 20th percentile characteristics, P35 is a structure with the 35th percentile characteristics, and so forth.

Table 5b – Sedgwick County Standardized Structure Characteristics

Variable	P20	P35	P50	P65	P80
Living area in SF	1,025	1,172	1,314	1,482	1,722
Lot size in SF	7,402	8,396	9,496	10,811	13,273
Age of improvements (years)	58	46	25	13	5
Bedrooms	2	3	3	3	4
Full bathrooms	1	2	2	3	3
Half bathrooms	0	0	0	0	1
CDU grade	9	9	9	9	10

Note: P20 is a structure with the 20th percentile characteristics, P35 is a structure with the 35th percentile characteristics, and so forth.

Table 6a – Maricopa County Standardized P50 Structure Value Estimates

Neighborhood	Improved parcel sale price (median)	Estimated standardized property value	Vacant lot sale price (median)	Implied standardized structure value
Neighborhood 1	$350,000	$245,455	$170,000	$75,455
Neighborhood 2	$318,000	$225,742	$138,250	$87,492
Neighborhood 3	$379,005	$275,377	$256,000	$19,377
Neighborhood 4	$277,000	$232,782	$43,000	$189,782
Neighborhood 5	$275,000	$231,669	$75,000	$156,669
Neighborhood 6	$208,000	$157,461	$70,000	$87,461
Neighborhood 7	$237,000	$221,392	$90,000	$131,392
Neighborhood 8	$215,000	$160,885	$94,500	$66,385
Neighborhood 9	$246,000	$193,055	$78,450	$114,605
Neighborhood 10	$189,000	$180,493	$45,000	$135,493
Neighborhood 11	$242,700	$221,007	$135,000	$86,007
Neighborhood 12	$370,000	$240,669	$96,000	$144,669
Neighborhood 13	$156,000	$143,941	$85,000	$58,941
Neighborhood 14	$182,000	$174,828	$110,000	$64,828
Neighborhood 15	$400,000	$368,080	$525,000	($156,920)
Neighborhood 16	$455,000	$395,607	$925,000	($529,393)
Neighborhood 17	$250,000	$243,642	$101,500	$142,142
Neighborhood 18	$543,345	$365,998	$170,000	$195,998
Neighborhood 19	$279,900	$231,470	$198,937	$32,533
Neighborhood 20	$337,500	$263,160	$160,000	$103,160
Neighborhood 21	$390,000	$294,217	$180,000	$114,217
Neighborhood 22	$510,000	$353,752	$165,000	$188,752
Neighborhood 23	$774,950	$351,275	$250,000	$101,275
Neighborhood 24	$510,000	$316,563	$75,000	$241,563
Neighborhood 25	$350,000	$257,351	$90,000	$167,351
Neighborhood 26	$669,079	$388,895	$165,000	$223,895
Neighborhood 27	$1,150,000	$329,271	$160,000	$169,271
Medians	$337,500	$243,642	$135,000	$114,217

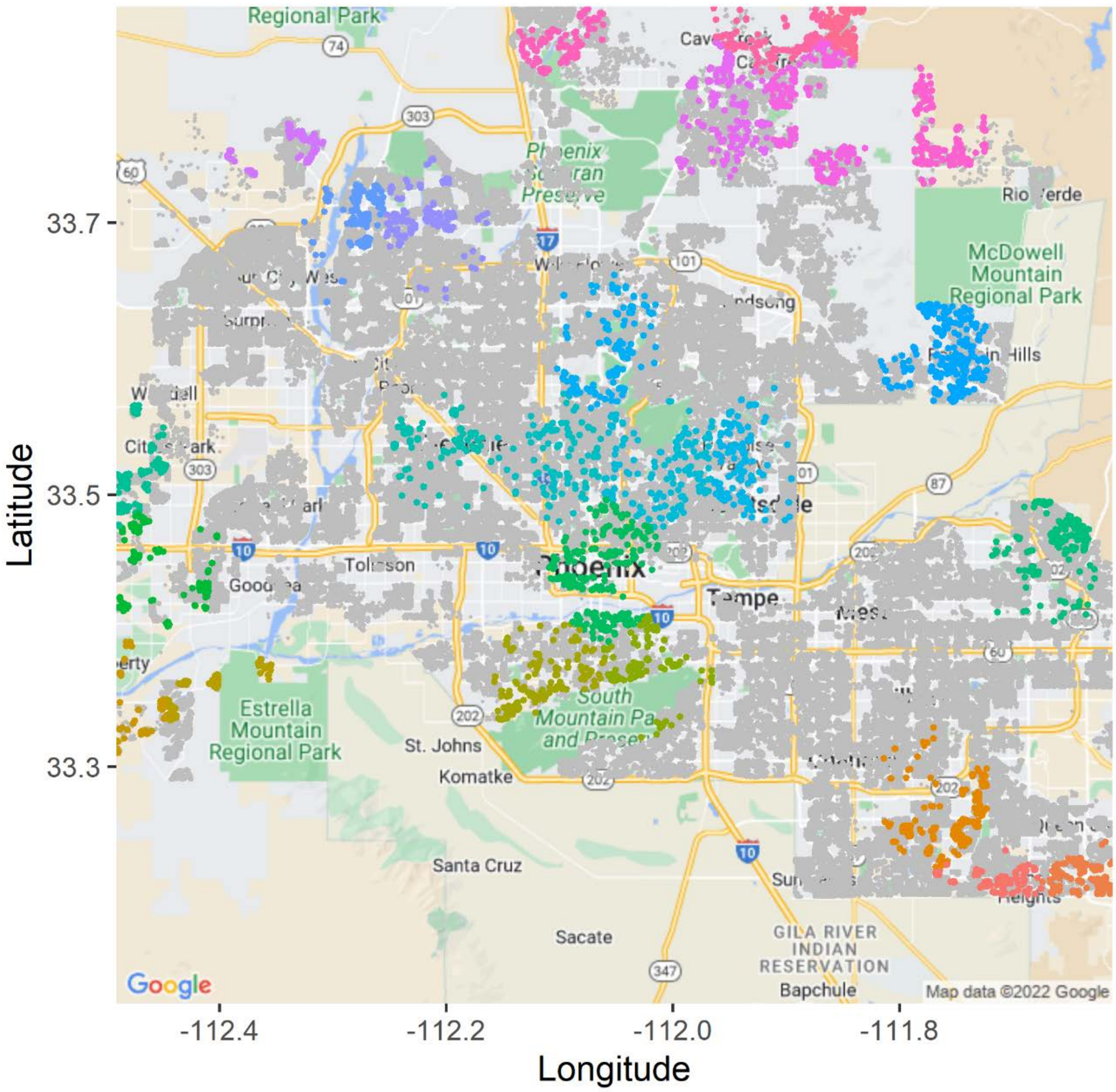
Notes: Cell entries are median values by neighborhood. Standardized property values are the predicted using neighborhood level property value regression results based on a parcel with the neighborhood's median lot size and the global P50 structure characteristics for all other characteristic (show in Table 6a). Implied structure value for each neighborhood is the implied structure value minus the median lot sale price for the neighborhood.

Table 6b – Sedgwick County Standardized P50 Structure Value Estimates

Neighborhood	Improved parcel sale price (median)	Estimated standardized property value	Vacant lot sale price (median)	Implied standardized structure value
Neighborhood 1	$272,500	$169,144	$29,500	$139,644
Neighborhood 2	$173,000	$139,666	$23,000	$116,666
Neighborhood 3	$247,000	$162,098	$37,000	$125,098
Neighborhood 4	$184,600	$132,873	$25,000	$107,873
Medians	$215,800	$150,882	$27,250	$120,882

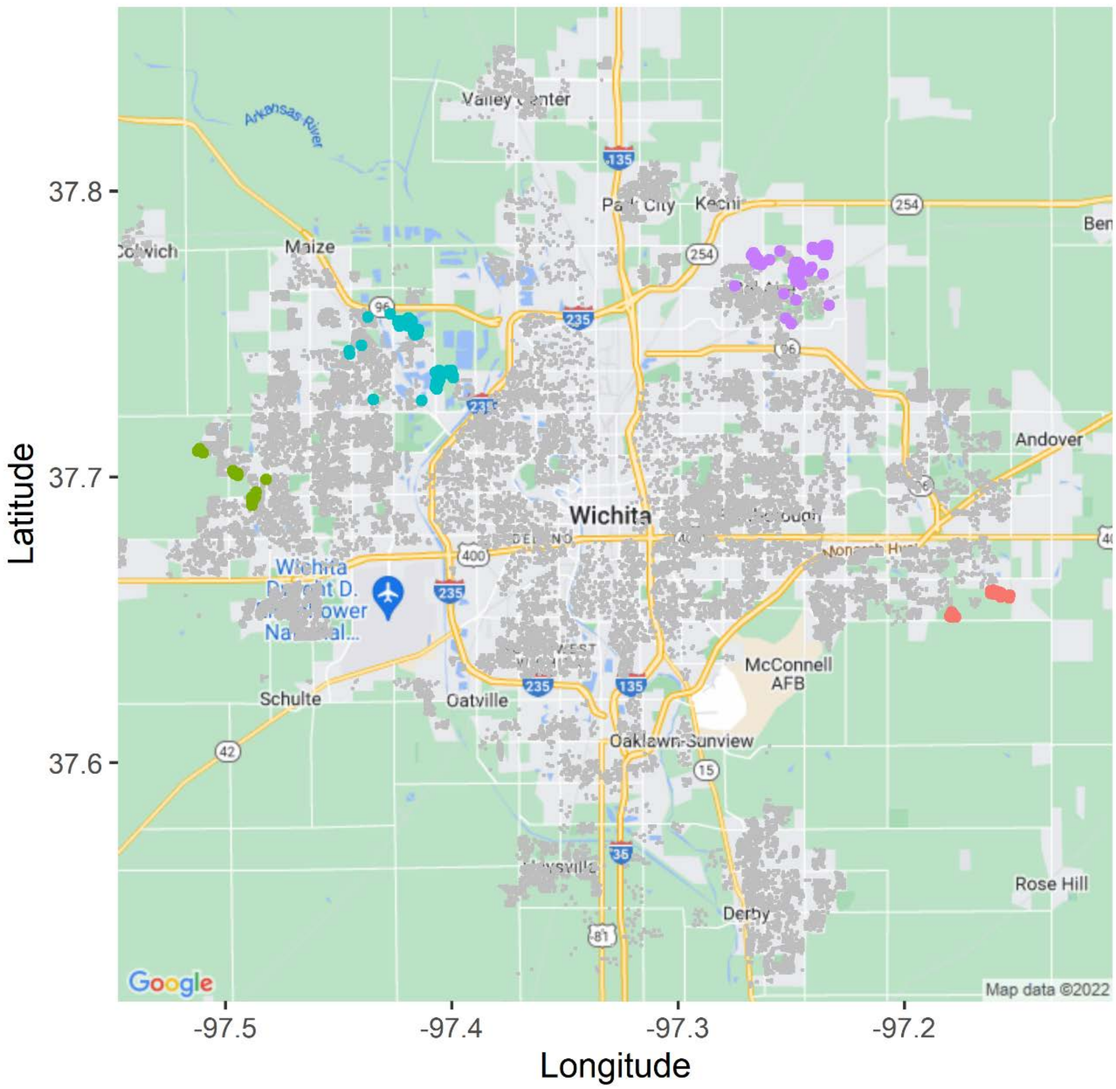
Notes: Cell entries are median values by neighborhood. Standardized property values are the predicted using neighborhood level property value regression results based on a parcel with the neighborhood's median lot size and the global P50 structure characteristics for all other characteristic (show in Table 6a). Implied structure value for each neighborhood is the implied structure value minus the median lot sale price for the neighborhood.

Figure 1a – Maricopa County Improved Parcel Sales and Lot Sales in the "Tie-down" Neighborhoods



Note: Neighborhoods are defined by applying a hex grid over the community. "Tie-down" neighborhoods have at least 275 improved parcel sales and 50 vacant lot sales. Distinct tie-down neighborhoods are identified by color in the figure.

Figure 1b – Sedgwick County Improved Parcel Sales and Lot Sales in the "Tie-down" Neighborhoods



Note: Neighborhoods are defined by applying a hex grid over the community. "Tie-down" neighborhoods have at least 275 improved parcel sales and 50 vacant lot sales. Distinct tie-down neighborhoods are identified by color in the figure.

Figure 2a: Maricopa County Spatial Distribution of Living Area

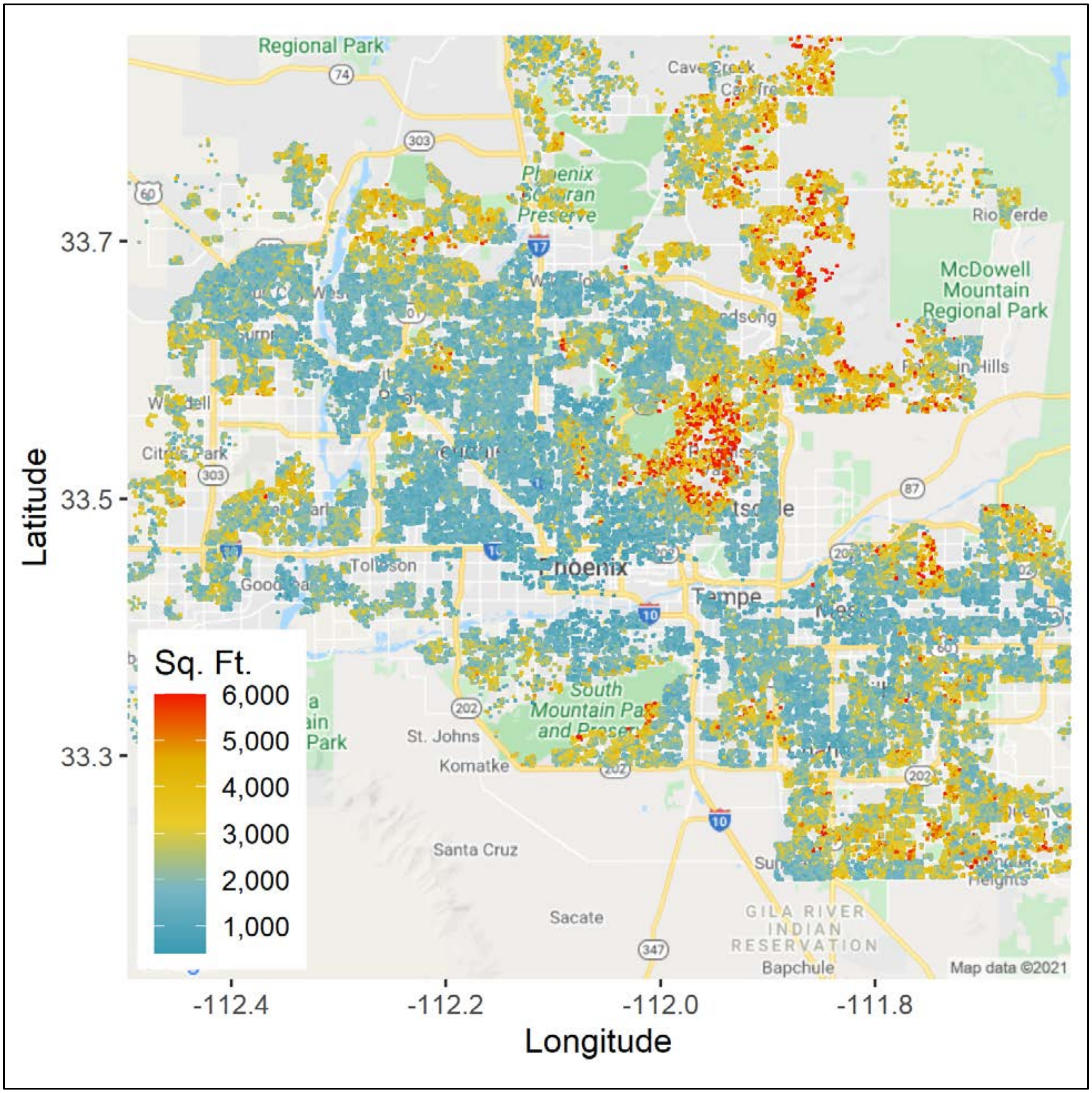


Figure 2b: Sedgwick County Spatial Distribution of Living Area

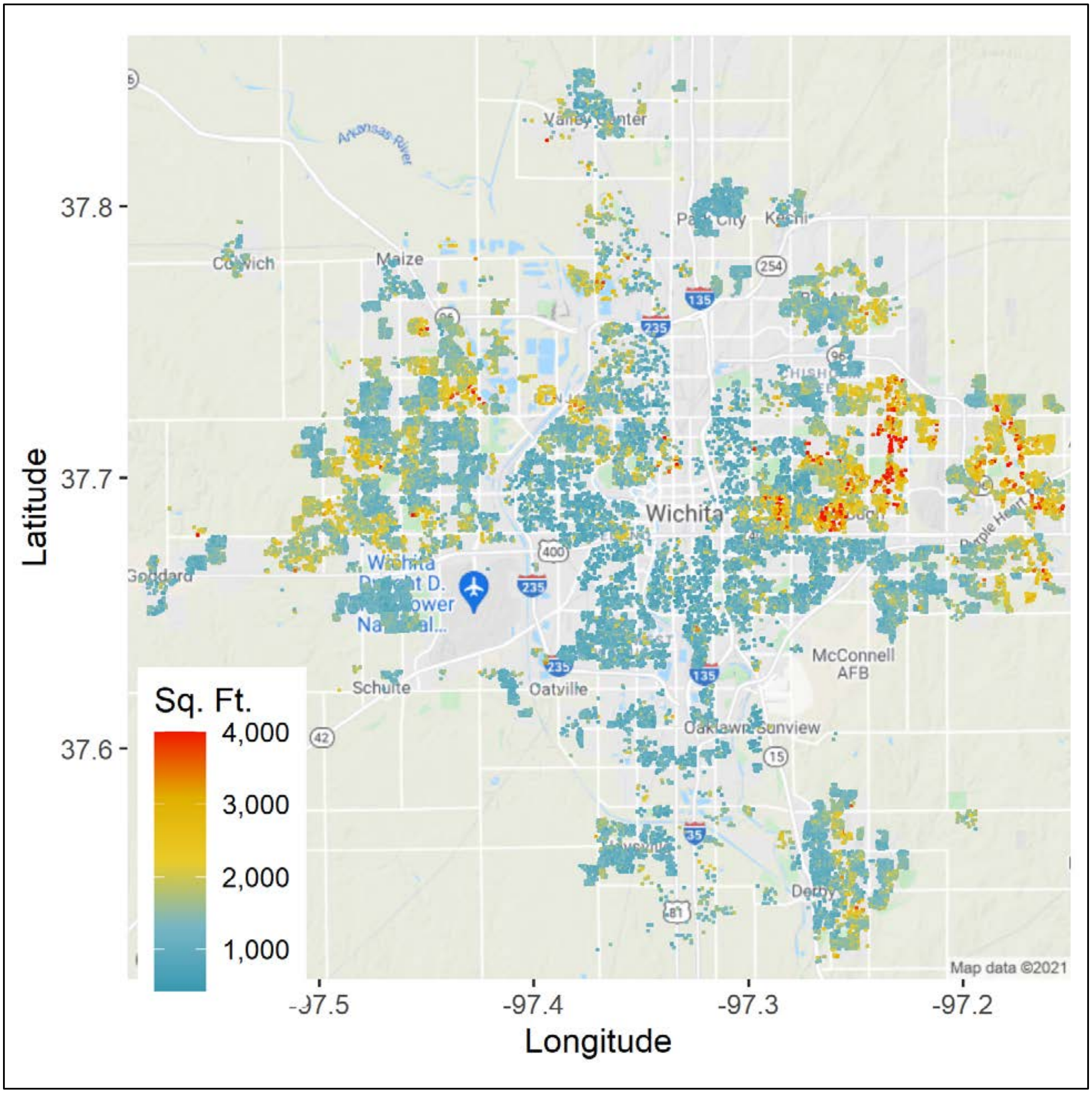


Figure 3a: Maricopa County Spatial Distribution of Dwelling Age

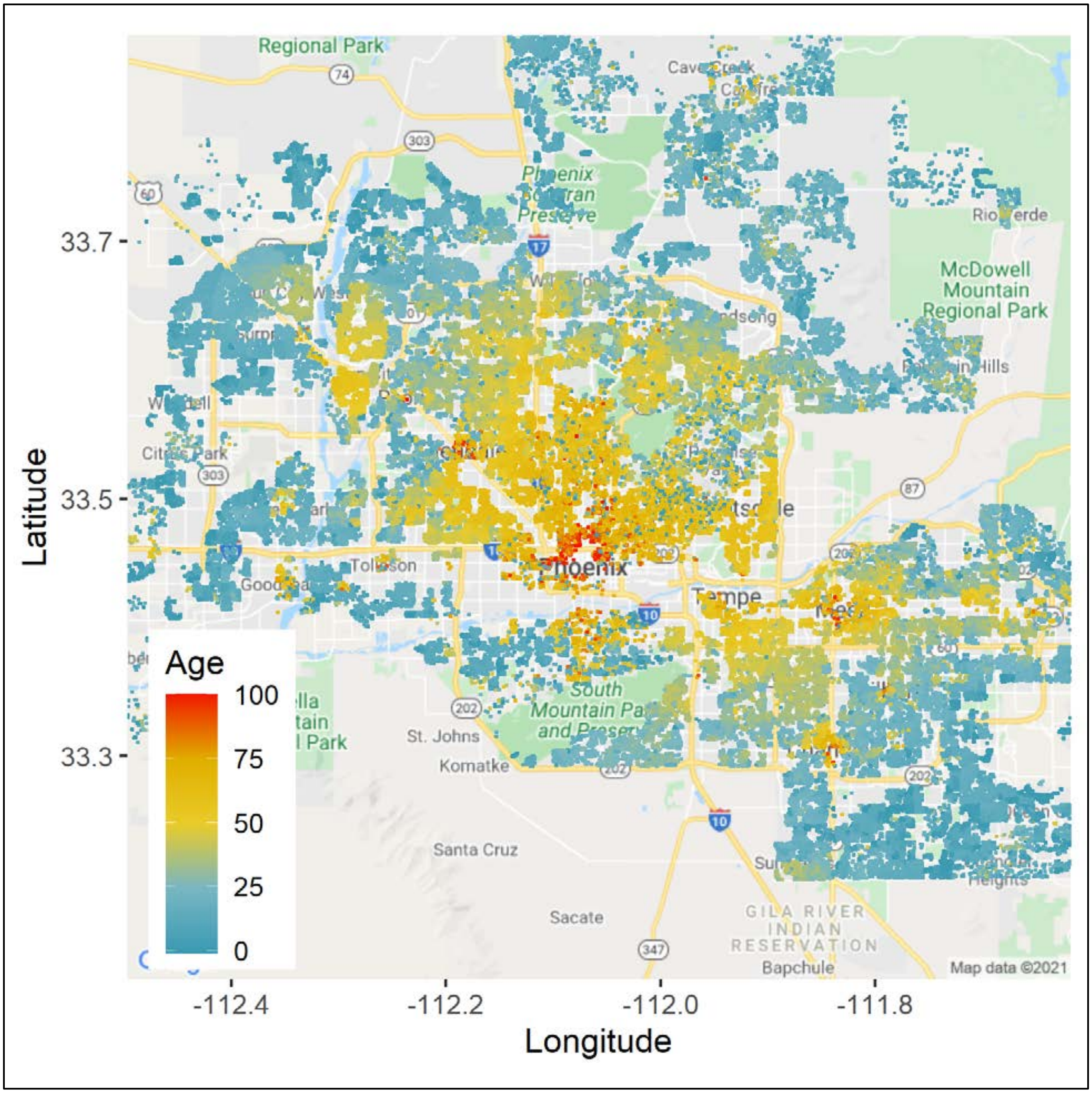


Figure 3b: Sedgwick County Spatial Distribution of Dwelling Age

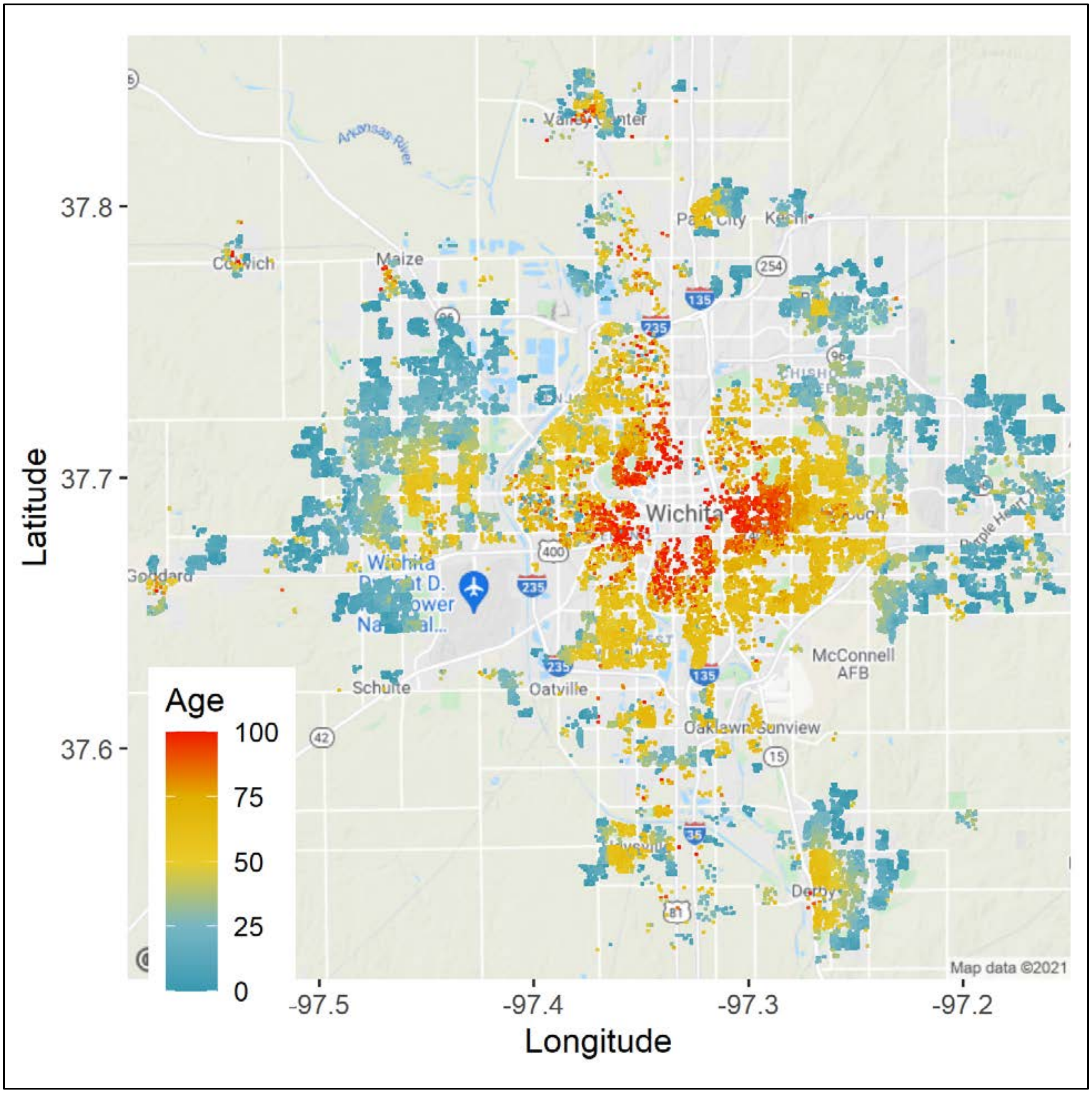
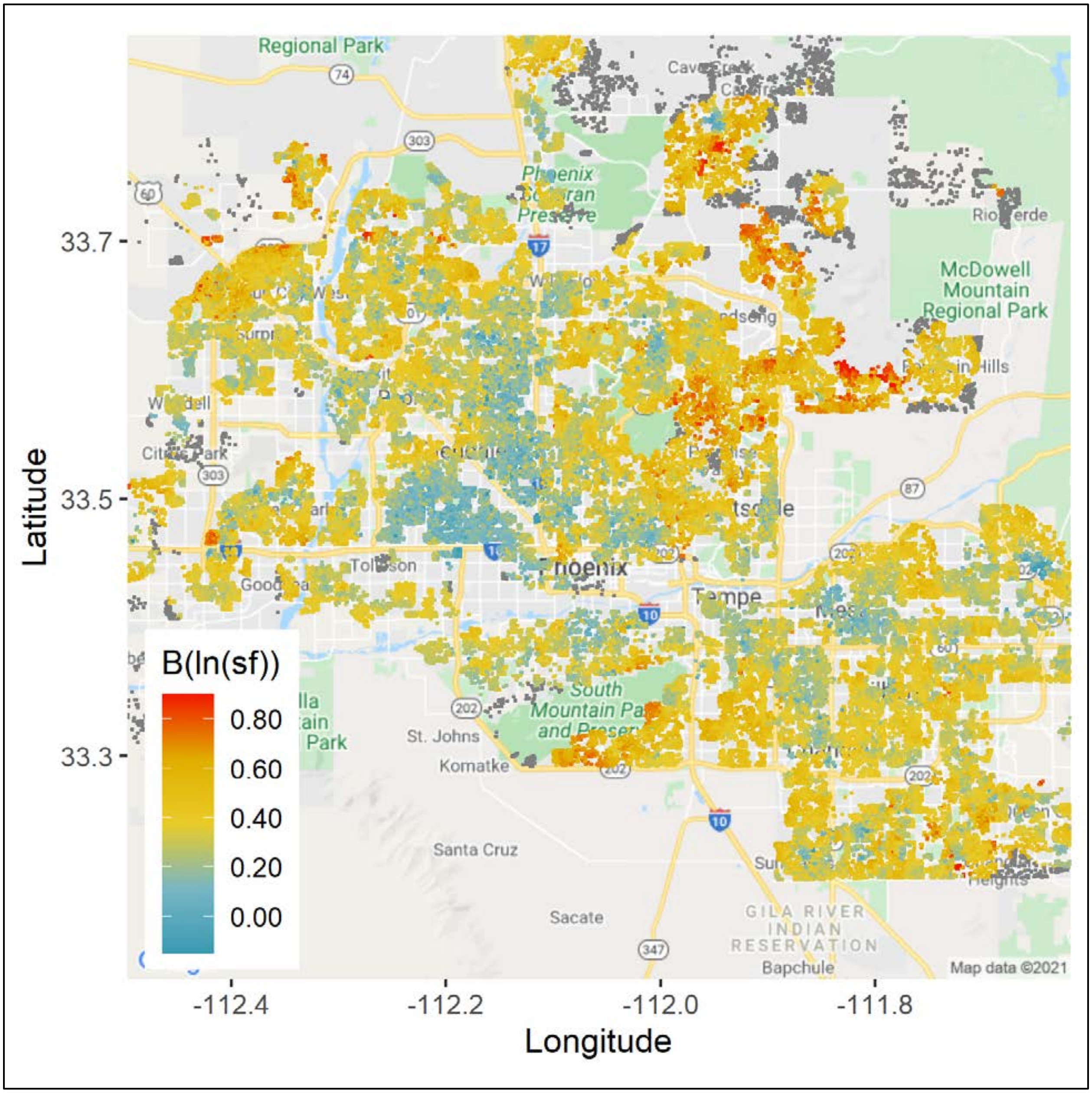
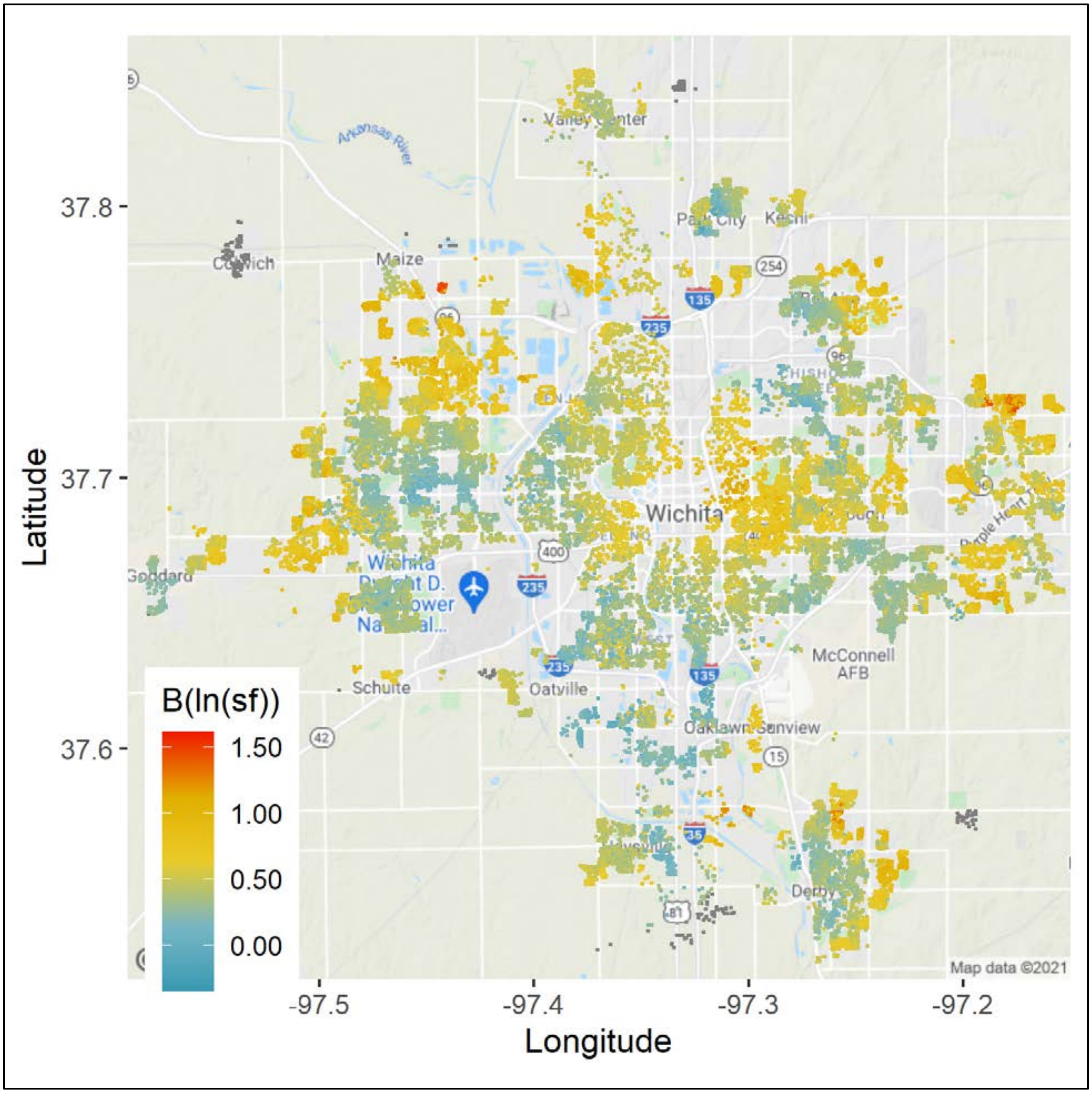


Figure 4a: Maricopa County LWR Coefficients on Living Area



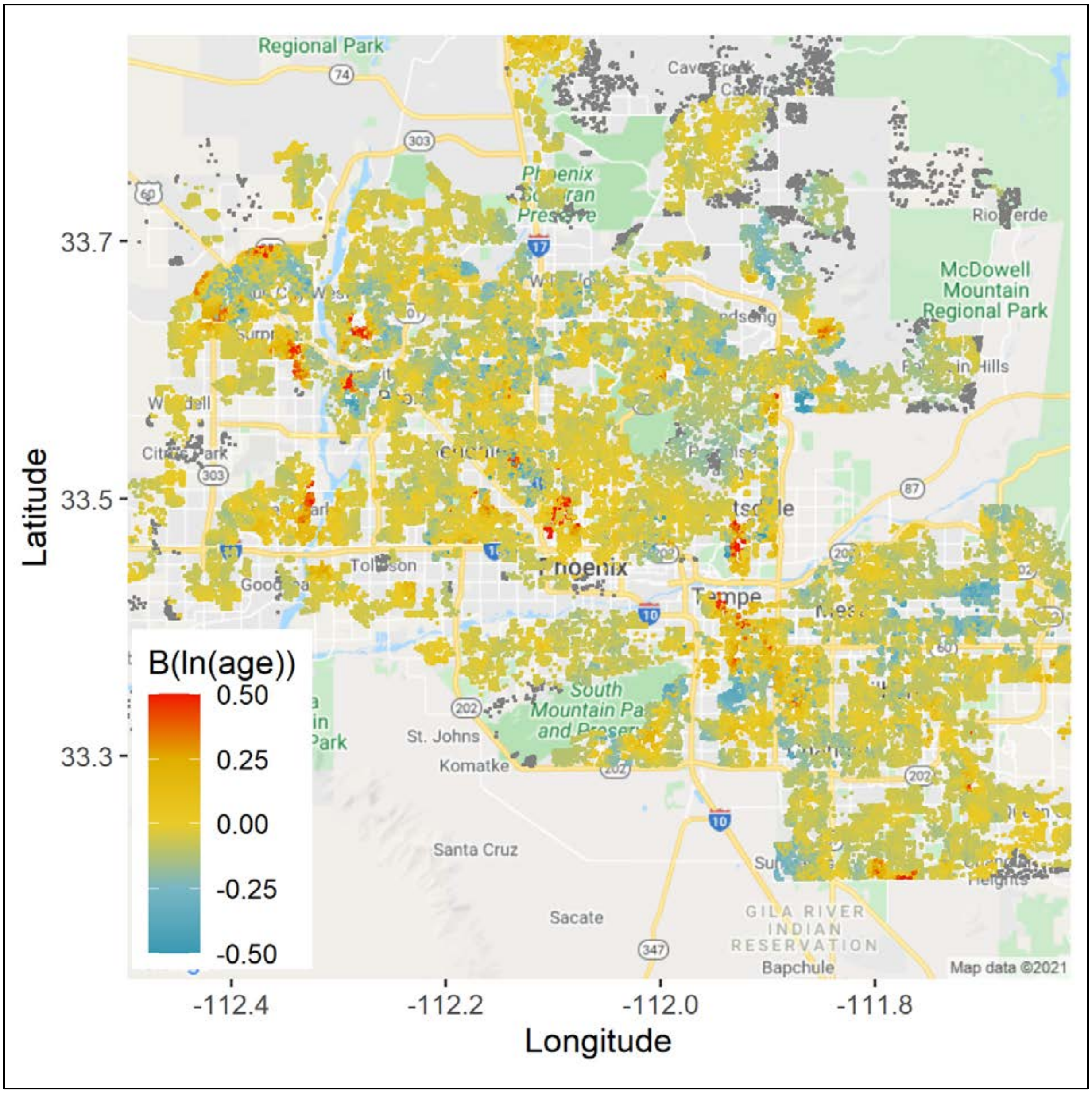
Note: Colors show the distribution of the elasticity of sale price with respect to square feet of living area from the locally weighted regressions across the community. Specifically, these regressions used the natural log of sale price as the dependent variable and colored dots show the magnitude of the coefficients on the natural log of living area in square feet from these regressions.

Figure 4b: Sedgwick County LWR Coefficients on Living Area



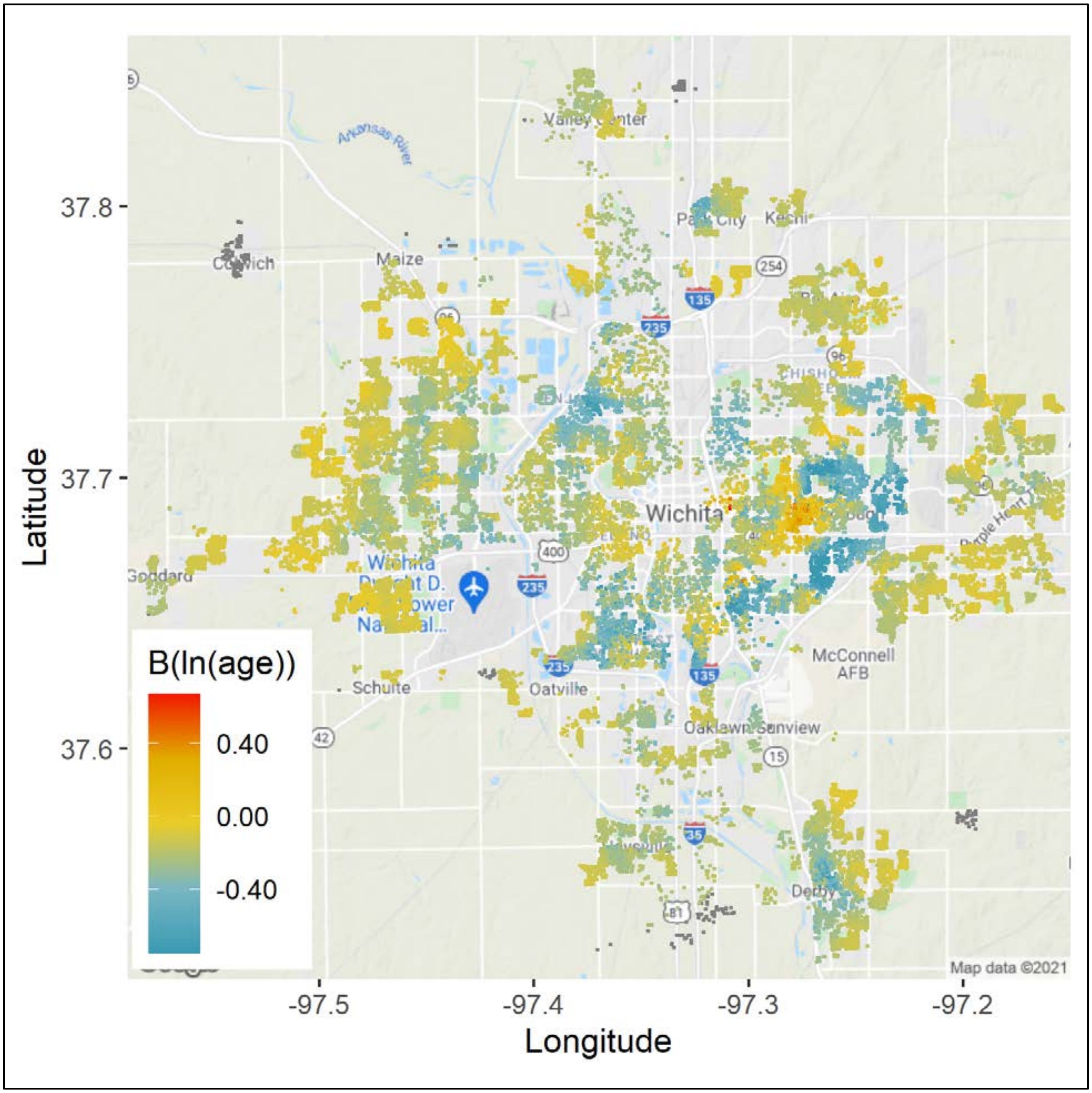
Note: Colors show the distribution of the elasticity of sale price with respect to square feet of living area from the locally weighted regressions across the community. Specifically, these regressions used the natural log of sale price as the dependent variable and colored dots show the magnitude of the coefficients on the natural log of living area in square feet from these regressions.

Figure 5a: Maricopa County LWR Coefficients on Dwelling Age



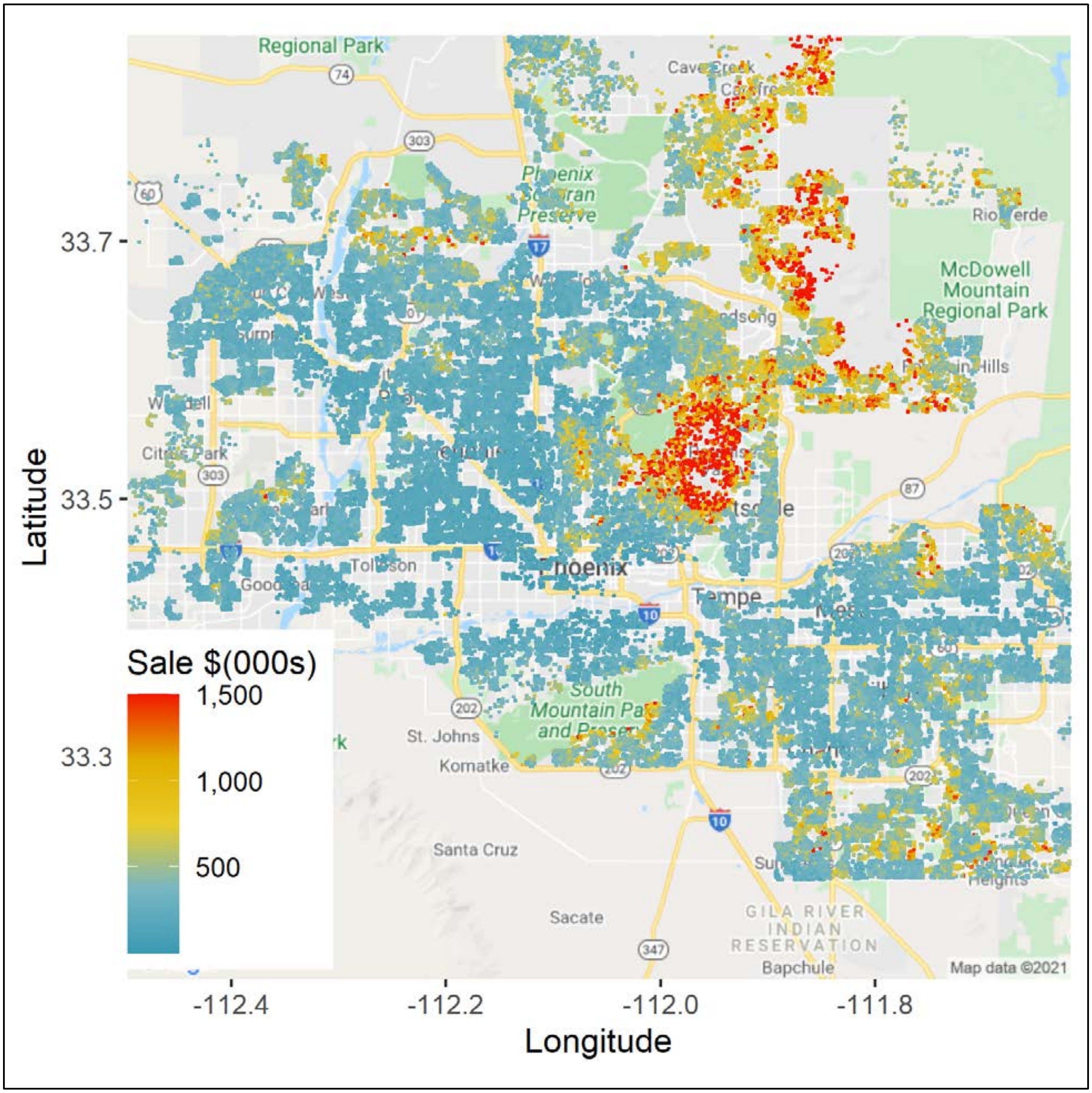
Note: Colors show the distribution of the elasticity of sale price with respect to dwelling age from the locally weighted regressions across the community. Specifically, these regressions used the natural log of sale price as the dependent variable and colored dots show the magnitude of the coefficients on the natural log of dwelling age from these regressions.

Figure 5b: Sedgwick County LWR Coefficients on Dwelling Age



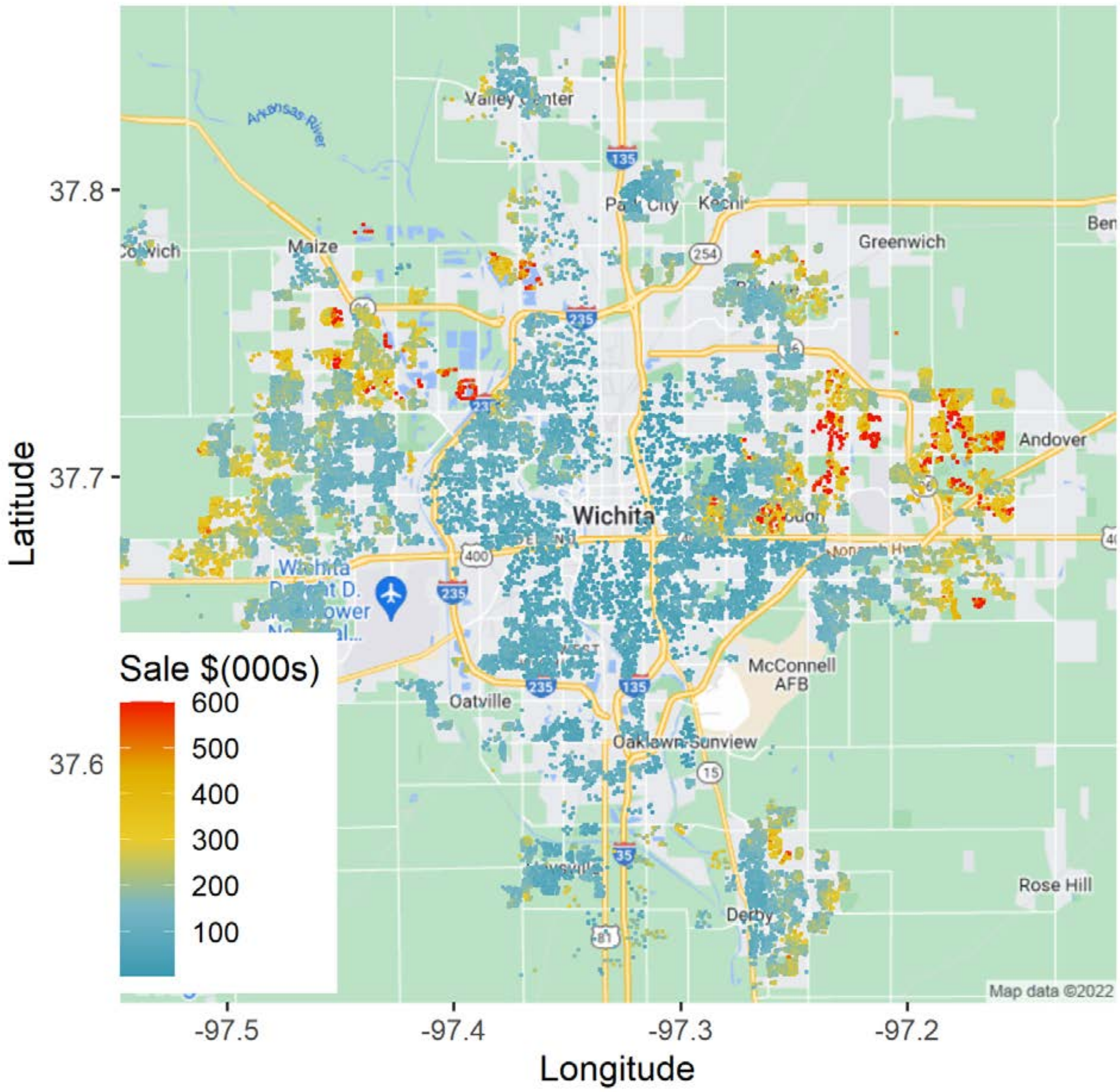
Note: Colors show the distribution of the elasticity of sale price with respect to dwelling age from the locally weighted regressions across the community. Specifically, these regressions used the natural log of sale price as the dependent variable and colored dots show the magnitude of the coefficients on the natural log of dwelling age from these regressions.

Figure 6a: Maricopa County Actual Single-family Residential Sale Prices



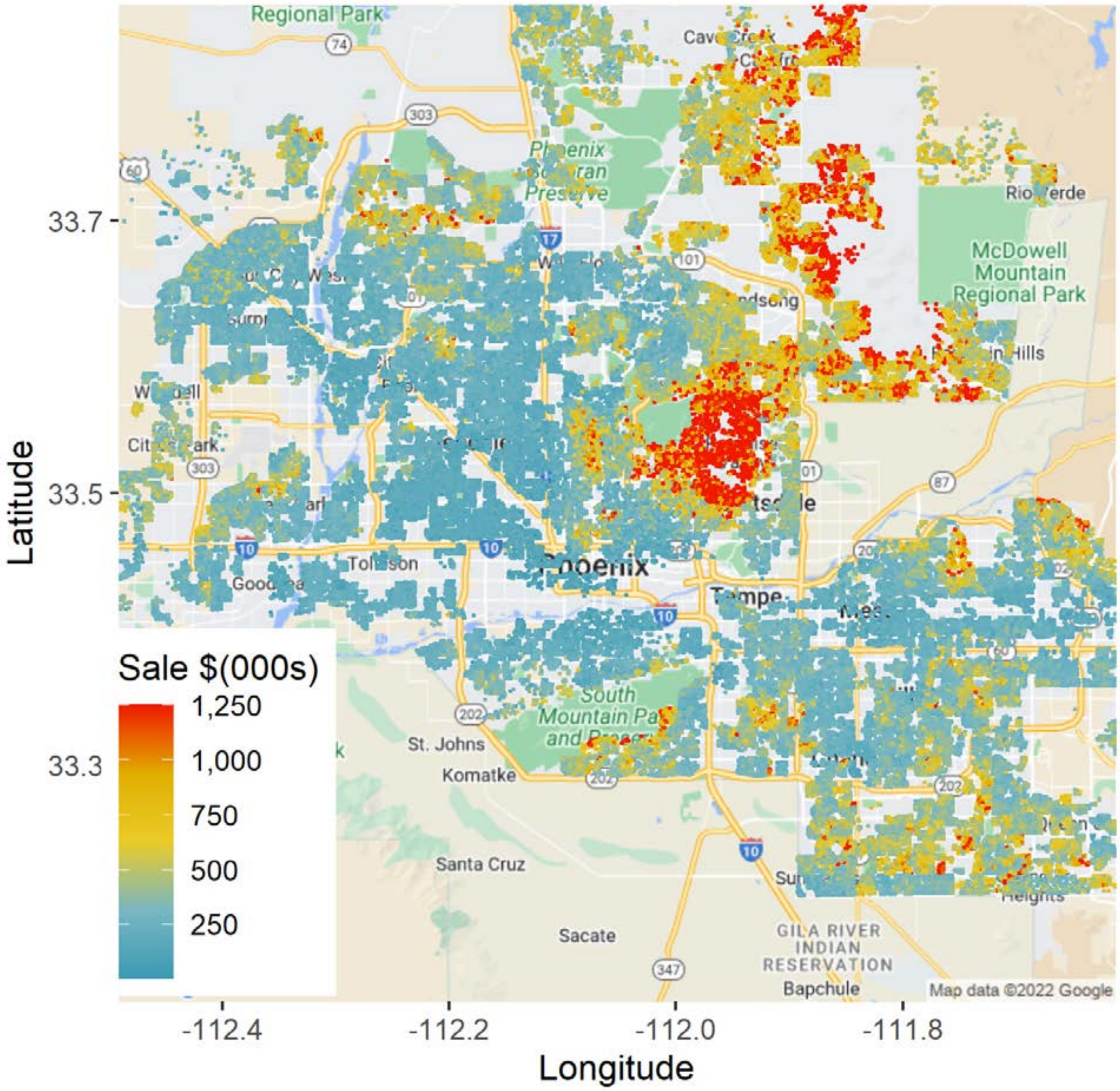
Note: This figure shows the geographic distribution of actual sale prices of single-family residential homes in Maricopa County.

Figure 6b: Sedgwick County Actual Single-family Residential Sale Prices



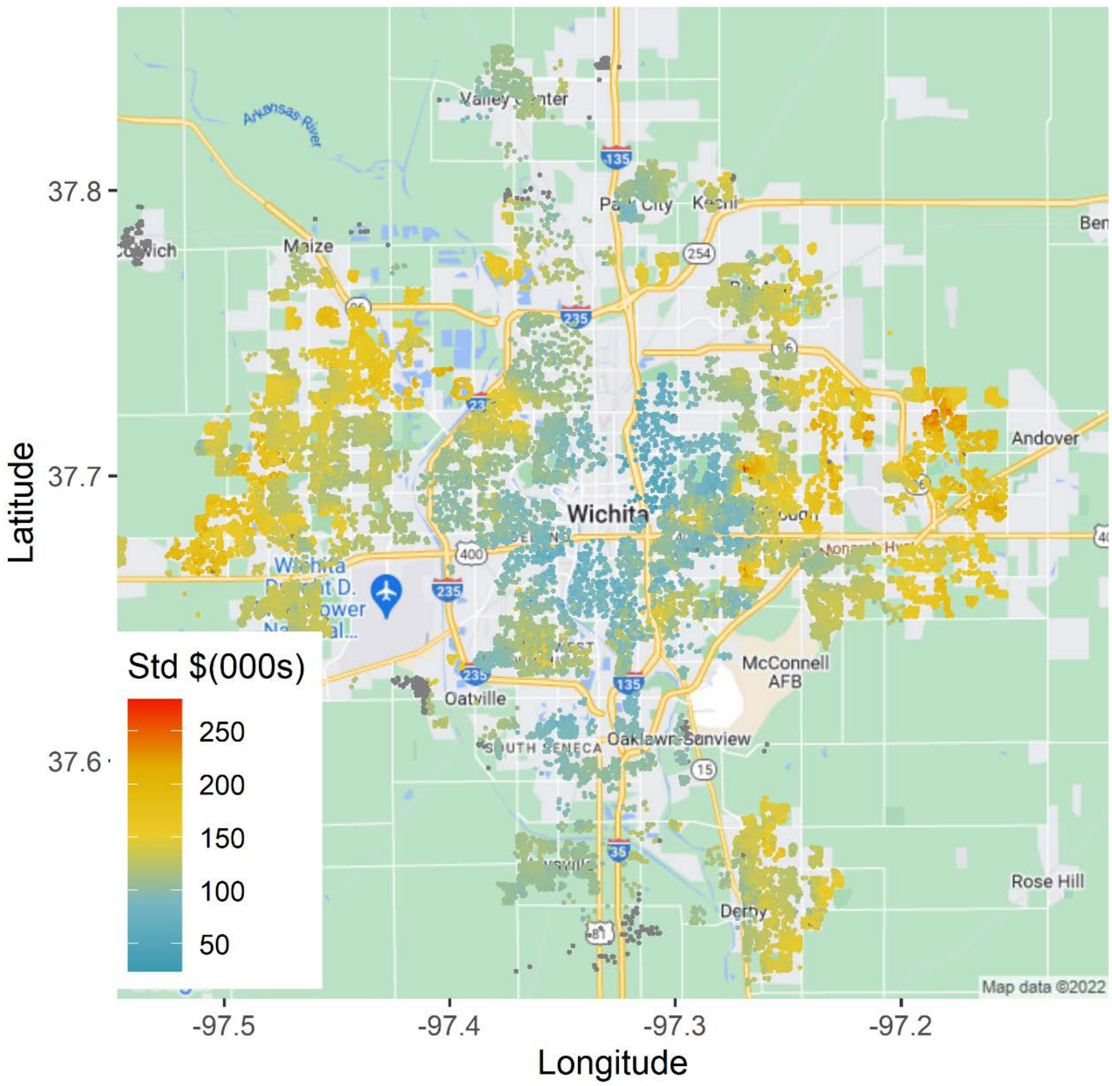
Note: This figure shows the geographic distribution of actual sale prices of single-family residential homes in Sedgwick County.

Figure 7a: Maricopa County Standardized Single-family Structure Values (Median Attributes)



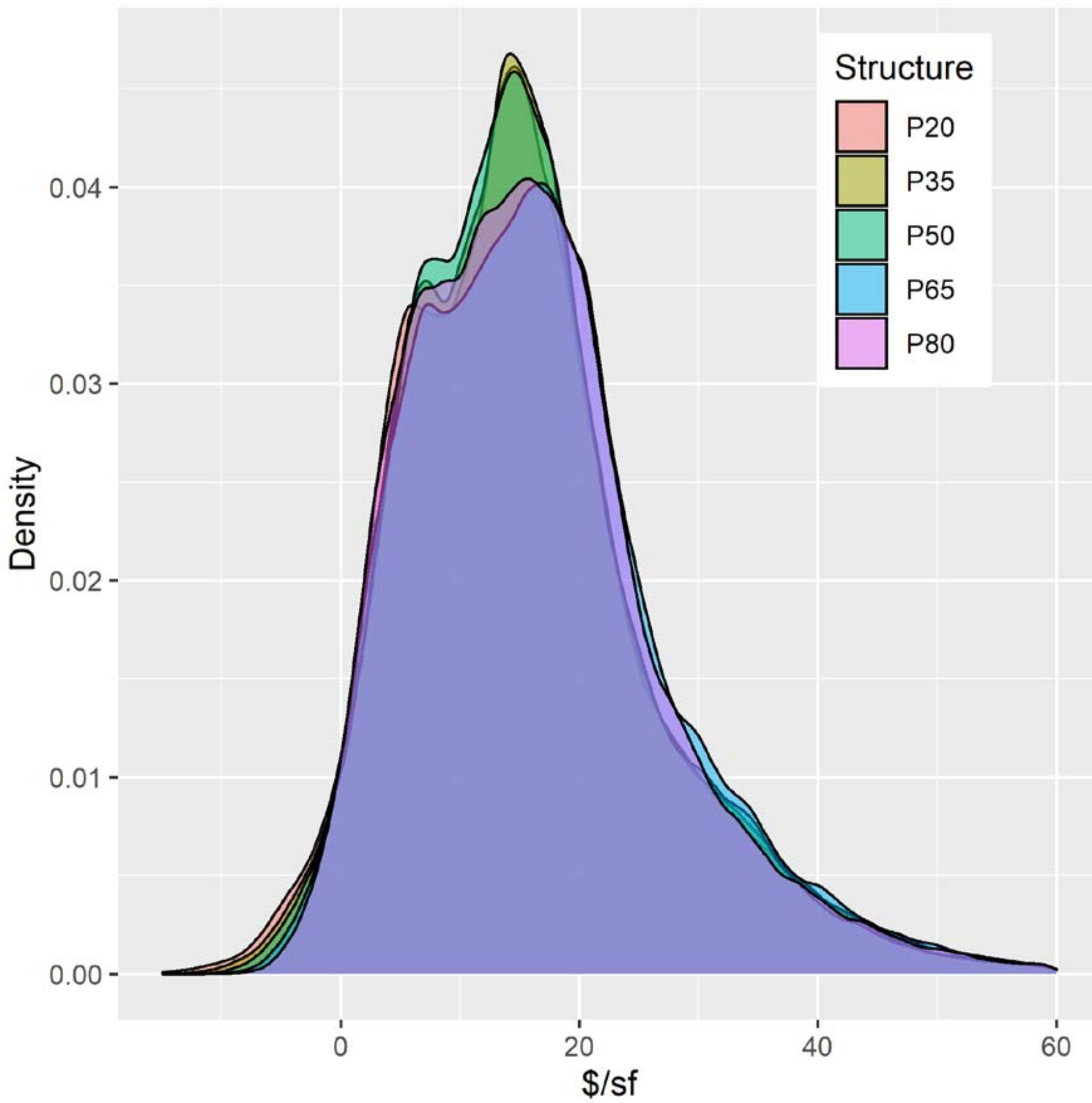
Note: This figure shows the geographic distribution of the predicted value of the standardized P50 structure (median attributes) at the location of each of the LWRs for Maricopa County. That is, these are the sale prices one would expect to observe if the hypothetical P50 structure were located on that parcel at the time of sale.

Figure 7b: Sedgwick County Standardized Single-family Structure Values (Median Attributes)



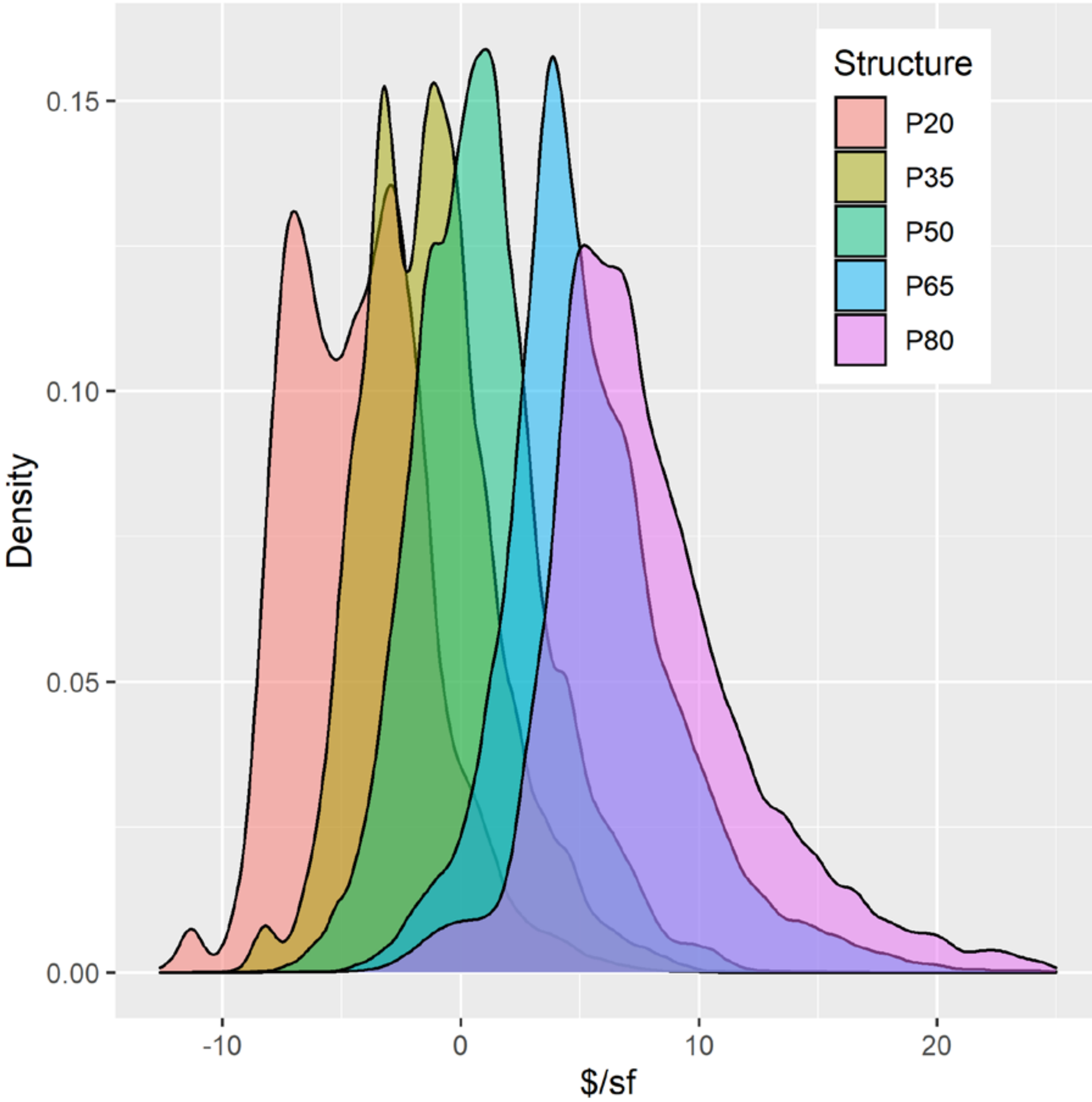
Note: This figure shows the geographic distribution of the predicted value of the standardized P50 structure (median attributes) at the location of each of the LWRs for Sedgwick County. That is, these are the sale prices one would expect to observe if the hypothetical P50 structure were located on that parcel at the time of sale.

Figure 8a: Maricopa County Standardized Land Value Distributions by Standardized Structure Type



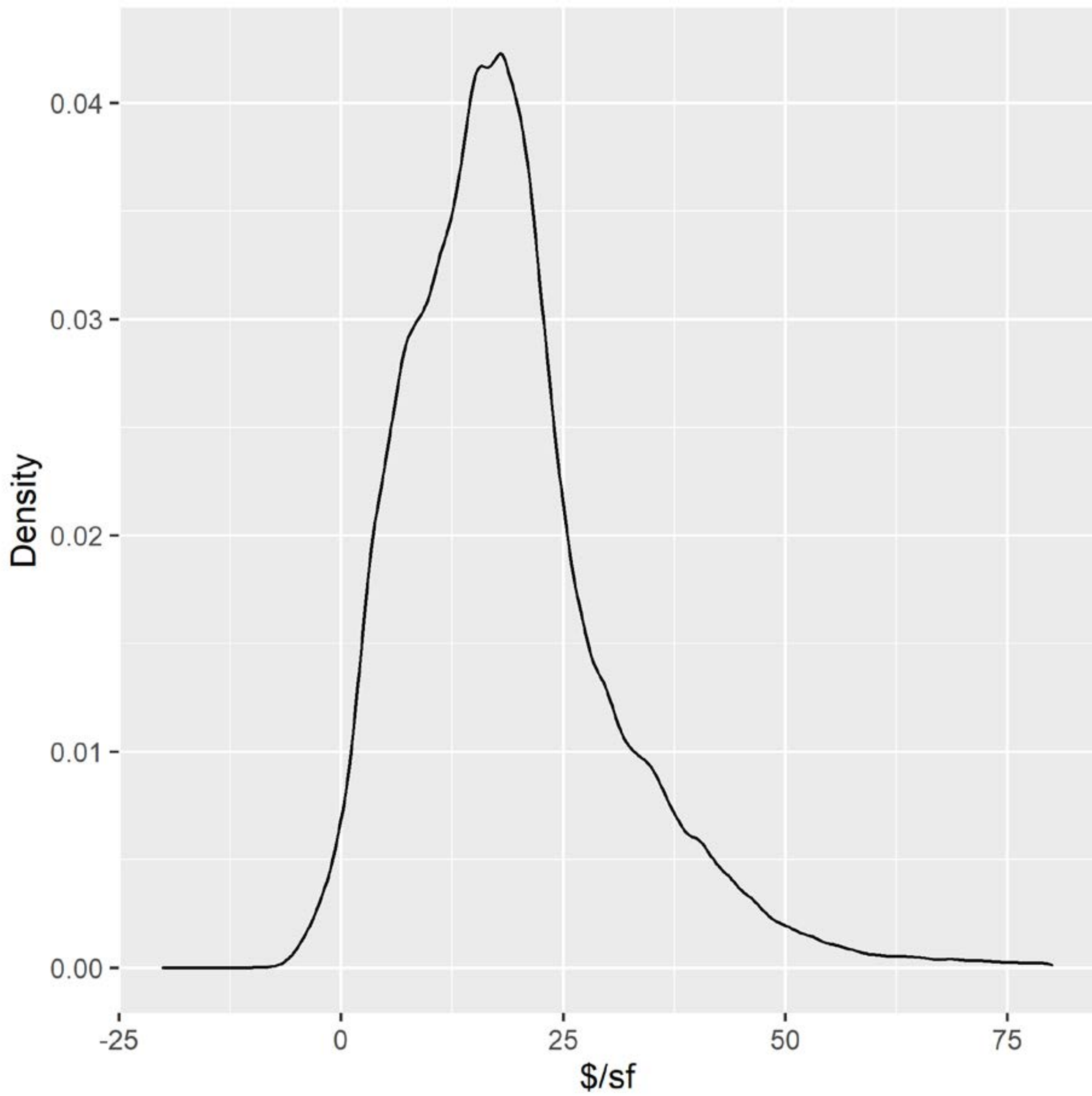
Note: This figure shows the estimated distribution of land values per square foot based on five different standardized structures. Each standardized structure is developed using a given percentile of each physical characteristics in the overall community. Thus, the P20 structure has the 20th percentile living area, number of plumbing fixtures, quality grade, etc. Other structures are defined similarly. Land values are estimated by first pricing standardized structures at each location based on that location's locally weighted regression results, subtracting the (uniform) cost of the standardized structure from the estimate and then dividing by the standardized lot size in square feet.

Figure 8b: Sedgwick County Distributions of Standardized Land Values per Square Foot by Standardized Structure Type



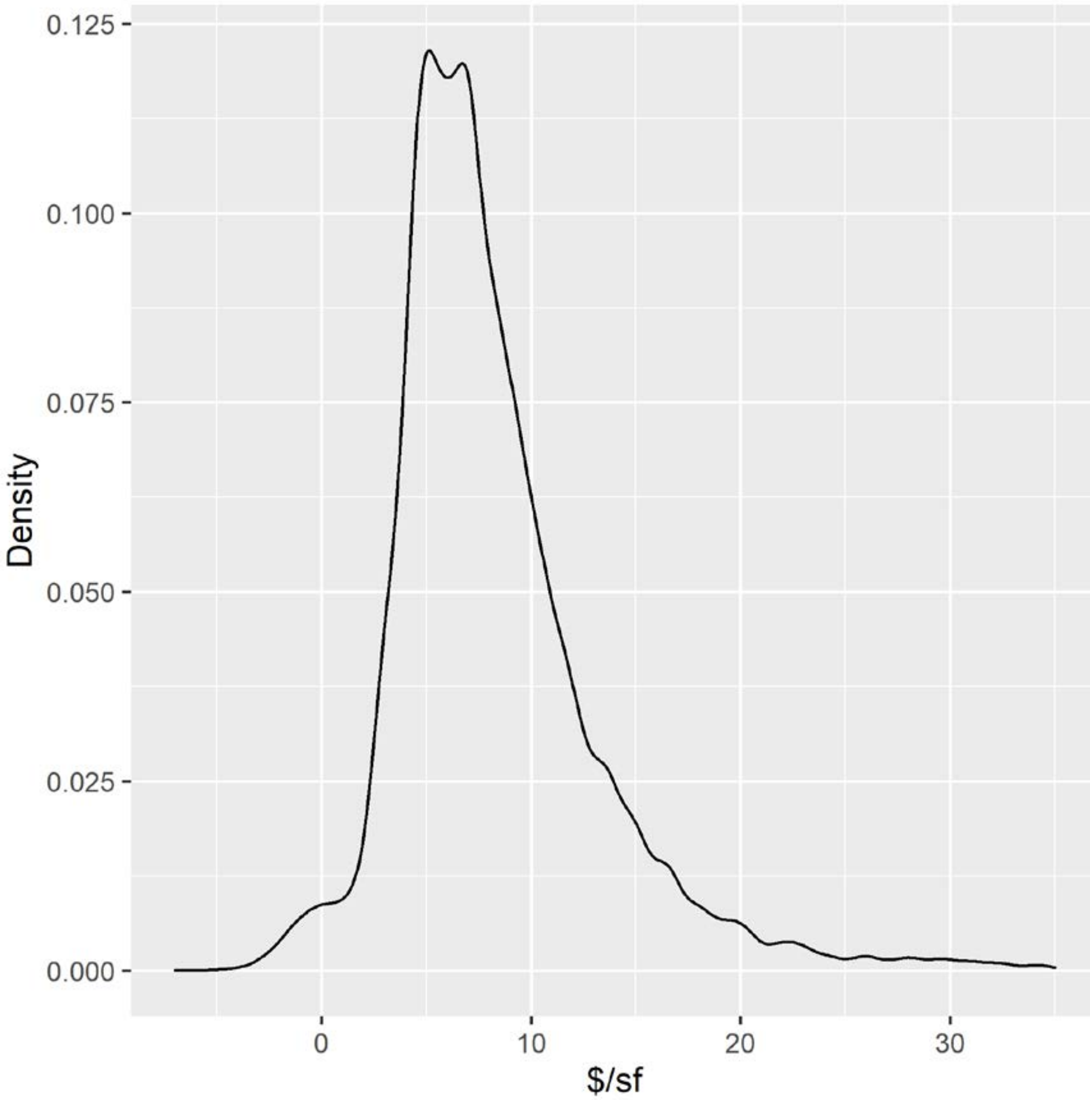
Note: This figure shows the estimated distribution of land values per square foot based on five different standardized structures. Each standardized structure is developed using a given percentile of each physical characteristics in the overall community. Thus, the P20 structure has the 20th percentile living area, number of bedrooms, number of bathrooms, CDU grade, etc. Other structures are defined similarly. Land values are estimated by first pricing standardized structures at each location based on that location's locally weighted regression results, subtracting the (uniform) cost of the standardized structure from the estimate and then dividing by the standardized lot size in square feet.

Figure 9a: Maricopa County Distribution of Land Prices per Square Foot using Upper Envelope of Standardized Structure Estimates



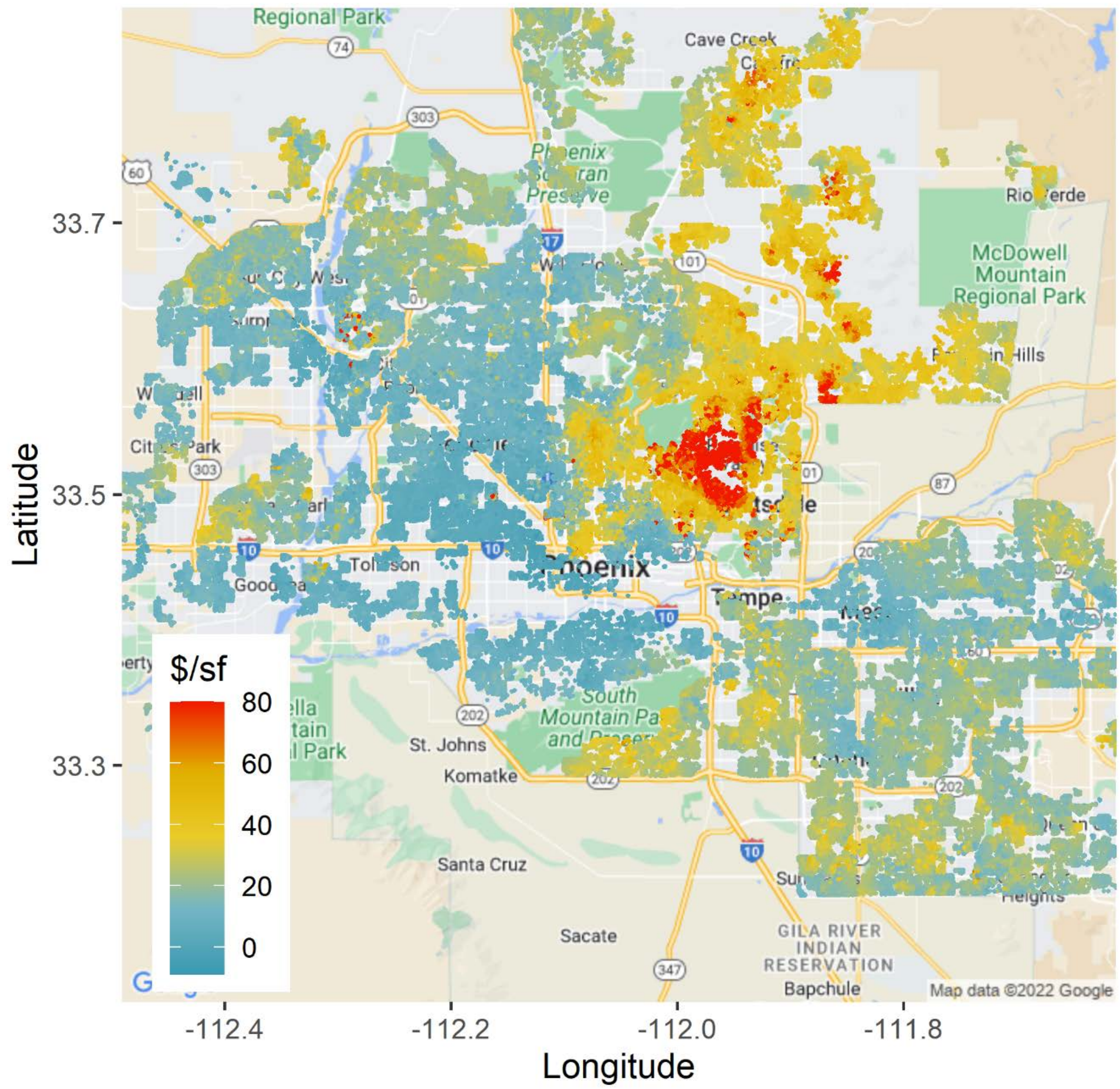
Note: This figure shows the distribution of land prices per square foot based on the upper envelope of five land value estimates from the five standardized structures. Note that for a given parcel, the estimated land value may have been in the upper tail of the P20 distribution in Figure 8a but in the lower tail of the P80 distribution of that figure. As a result, the distribution shown here is *not* the upper envelopes of those shown in Figure 8a.

Figure 9b: Sedgwick County Distribution of Land Prices per Square Foot using Upper Envelope of Standardized Structure Estimates



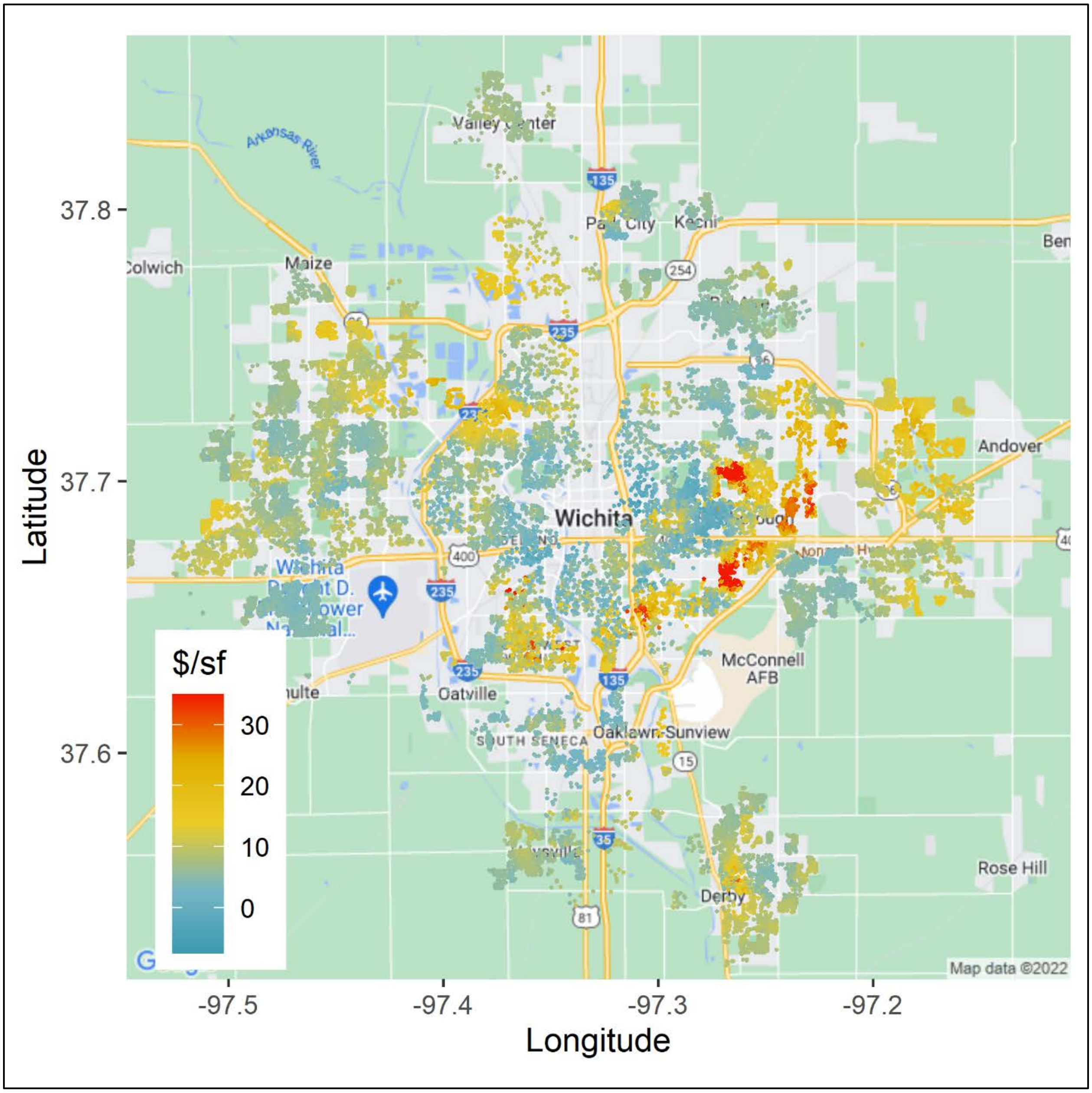
Note: This figure shows the distribution of land prices per square foot based on the upper envelope of five land value estimates from the five standardized structures. Note that for a given parcel, the estimated land value may have been in the upper tail of the P20 distribution in Figure 8b but in the lower tail of the P80 distribution of that figure. As a result, the distribution shown here is *not* the upper envelopes of those shown in Figure 8b.

Figure 10a: Maricopa County Spatial Distribution of Land Prices per Square Foot using Upper Envelope of Standardized Structure Estimates



Note: This figure shows the implied land values per square foot for parcels across Maricopa County as calculated by subtracting the value of the standardized structure from the predicted value of the parcel with that structure, and then taking the upper envelope of the values obtained from this exercise using the P20, P35, P50, P65 and P80 standardized structures.

Figure 10b: Sedgwick County Spatial Distribution of Land Prices per Square Foot using Upper Envelope of Standardized Structure Estimates



Note: This figure shows the implied land values per square foot for parcels across Sedgwick County as calculated by subtracting the value of the standardized structure from the predicted value of the parcel with that structure, and then taking the upper envelope of the values obtained from this exercise using the P20, P35, P50, P65 and P80 standardized structures.