Five years in: Assessing the impacts of Chicago's Large Lots Program

Measuring the effectiveness of Chicago's $1 residential lot sale program on crime and wealth-building

Matthew Stern
Master's Project – April 2019
Department of City and Regional Planning

Abstract

Since 2014, the City of Chicago has sold more than 1,200 city-owned vacant properties to same-block landowners in the South and West sides for $1 each. This *Large Lots Program* is part of a national trend in which cities encourage productive reuse of vacant land by heavily discounting the sale of empty lots in distressed neighborhoods to local buyers. These experimental programs present opportunities for marginalized communities such as wealth-building, crime reduction, and increased control of neighborhood change processes. They also present risks and opportunity costs. Despite these trade-offs vacant land disposition programs have been understudied and no published literature evaluates whether programs actually achieve their stated goals. Using data from a variety of public and private data sources I evaluate the impact of Chicago's Large Lots program in three economically distressed neighborhoods. I investigate *who* is buying Large Lots, *how many* parcels they are buying, and *how far* they live from the parcels they are purchasing. Additionally, I perform a block-level difference-in-differences analysis to explore whether the program *reduces crime*. Finally, I examine the potential *wealth generation* affects for residents of low-income communities using a parcel-based projection of land value changes.

Key Words

Shrinking cities; Legacy Cities; vacant land; side yards; Chicago; wealth-building;
difference-in-differences

Acknowledgements

Kyle Onda, PhD candidate in the Department of City and Regional Planning, helped me gain access to and initially interpret ZTRAX data.

Dr. Bill Lester, my research advisor, was fundamental in guiding my perspective and analysis, and taught me just about everything I know about statistics.

Table of Contents

Table of Figures

Table of Tables

I. Introduction

Urban land vacancy is both a problem and an opportunity. Decades of urban shrinkage and the recent foreclosure crisis have contributed to high levels of urban vacancy that decrease tax revenue, increase municipal costs, wreak havoc on real estate markets, and have been linked to increased crime. On the other hand, extreme vacancy also presents a canvas on which planners and communities can experiment with visionary new neighborhood designs (e.g. "greening" and urban agriculture), find space for the construction of affordable housing, and increase family wealth in low-income communities. In Chicago, with a booming real estate market overall but south- and west-side submarkets defined by extreme vacancy and disinvestment, this duality is clearly visible.

In Chicago and elsewhere, extreme vacancy is geographically bound up with race and class, and often dates back to red-lining and mid-century urban renewal. Therefore, questions about the consequences of vacancy and the actions taken to address it have impacts not only on municipal finances and neighborhood quality generally, but on our cities' most vulnerable communities specifically. The ideal policy solution to extreme urban vacancy would create safer, more vibrant communities; restore a functional market for land; and increase the tax base. It would also be co-designed with the communities that have experienced decades of disinvestment, so that they can both steer the process of neighborhood change and benefit from the development and improved property values that result.

These were, in brief, the goals of the City of Chicago's Large Lots program when it was founded in 2014. The program made vacant, city-owned, residentially-zoned parcels available for sale to non-delinquent same-block landowners for $1. According its organizers, the Large Lots program sought to "return property to the tax rolls…allow local residents to have greater control over vacant land in their neighborhood, create wealth… increase safety, build community, and raise home values" (ChiHackNight 2017). At the end of 2017, the City had sold 1,248 parcels through the program via five neighborhood-specific pilots and one city-wide expansion. Local journalists and bloggers have captured stories of lots converted into playgrounds and gardens (Koski 2015), and then-Mayor Rahm Emmanuel put his support behind the expansion of the program (Spielman 2016). However, no evaluation of the program has been published.

Chicago's Large Lots program is one of many similar programs across the country. New Orleans, Cleveland, Buffalo, St. Louis, and Detroit are among the cities that make vacant land available to citizens through *side yard* programs, usually heavily discounted and restricted in some way to the local community. The few studies that have analyzed these programs tend to focus on measures of how successful cities have been at implementing them (Dewar 2006; Ganning and Tighe 2015; Dewar 2015), rather than the impacts of the programs themselves. I have located no published studies that measure the impact any vacant land disposition program against the goals described in the first paragraph.[1]

This study seeks to rectify this gap in scholarship by answering three questions about Chicago's Large Lots program. First, who is purchasing parcels? I use data from the Zillow ZTRAX real estate transactions database that suggests that 13.5% of parcels are bought by purchasers who buy more than 2 parcels per

[1] Ongoing research at UIUC in partnership with the US Dept. of Agriculture is documenting what buyers of Large Lots parcels in Chicago are doing with their land (ChiHackNight 2017). Although they have not yet published their work, they have recently presented in conference proceedings (Gobster et al. 2018).

program round, in violation of city policy; up to 28% of buyers may not live on the block where they are purchasing, and at least 7% of buyers appear to be professional developers. Second, what is the impact of large lots sales on crime? I combine data from the Chicago Department of Planning and Development, the Zillow ZTRAX database, and the Chicago police department to demonstrate that the presence of at least one Large Lots sale on a block reduces crime by between 2.9% and 6.5%. Third, how much wealth might these property transfers be potentially creating? I run models based on multiple data sources to demonstrate that Large Lots program activity in three neighborhoods has generated between $16 and $83 million in wealth, and could generate as much as between $97 and $214 million by 2025. Based on these findings I suggest that the Large Lots program is quite effective but could be improved by placing some additional limitations on buyer eligibility and collecting more information from applicants to better inform the aldermen who approve and deny applications.

This paper begins with a review of the literature applicable to this study, vacancy management policies used in cities around the country, and a history of vacancy and vacancy management in Chicago in particular. Then, in section 3, I explain and motivate my research questions and document the data sources and dataset construction methods that underpin the analysis. In section 4 I detail the methodology used to answer the three research questions and, in section 5, I review general information about the study areas for context, and the findings for each research question. Finally I discuss the implications of critical findings and make concluding remarks.

II. Literature & Policy Review

Shrinking cities and the foreclosure crisis

Many American cities in the northeast and Midwest have been shrinking over the last half century. In the 1950s, 29 cities (including 8 of the country's largest 10) experienced peak population; another 66 peaked in the '60s and '70s (Mallach 2017). Scholarship on that shrinkage—and how planners should respond to it—was slow to develop because the American fixation on growth limited discourse on decline and because discourse that did exist was problematically intermingled with conversations about race, decay, and deindustrialization in the *urban crisis* dialogue (Dewar 2006; Mallach 2017).

In more recent years, scholars have more carefully inspected this *shrinking cities* trend, seeking to "measure changes, where they are, what has caused the decline, what they represent, and how planning efforts can resolve attendant problems" (Hackworth 2014, 6). Frequently cited causes for urban shrinkage include deindustrialization, suburbanization, and natural demographic change (Großmann et al. 2013); post World War II job losses in cities (Dewar 2006); "the decreasing ability of cities to annex, migration to the Southwest and West, and the spread of automobile ownership" (Beauregard 2001, 130); and white flight (Mallach 2017). Twentieth century federal government policy and programs such as HOLC red-lining maps, the suburban focus of new federal home mortgage insurance products, investment in highway construction, and urban renewal are accepted by many as partially or mostly responsible for this trend (e.g. Jackson 1985), although Beauregard (2001) refutes these claims.

Urban shrinkage has created issues of mass vacancy and property distress, which have been shown to have a variety of negative effects including declining property tax revenues, blight, decimation of real estate markets, decreased liquidity and equity for surviving nearby homeowners, and increased crime. More recently, home foreclosures following the sub-prime mortgage crisis in the late 2000's have provided a new source of urban vacancy (Ellen, Lacoe, and Sharygin 2013; Cui and Walsh 2015; Mallach 2018). These challenges disproportionately plague communities whose residents are predominantly poor and non-white (Immergluck and Smith 2006; Schilling and Logan 2008). The resulting concentration of vacancy is therefore both a "symptom and a disease" of poverty and *de facto* segregation (Schilling and Logan 2008, 452; Butler 2016).

Shrinkage and extreme vacancy are also widely recognized as creating a range of opportunities, and scholars and practitioners engaged with shrinking cities are exploring how to utilize them (e.g. Pagano and Bowman 2000; Schilling 2002). Vacant properties can be converted into community assets via land banking (Tappendorf and Denzin 2011), greening (Schilling and Logan 2008), or the construction of affordable housing (Kelly 2016); they can be used as wealth-building mechanisms in communities hard-hit by decades of disinvestment; and they present unique opportunities to experiment with innovative urban design. Related to this is an active wider conversation about what the *right size* of a city is and should be and how to go about achieving it (Hummel 2015).

The stages and impacts of property distress

Vacancy is a label that applies to certain conditions within an interrelated ecosystem of property distress. Tax and mortgage *delinquency* can lead to *foreclosure* and/or property *abandonment* as well as the creation of *zombie* properties that have no clear owner or title. In healthy real estate markets, private capital will see opportunity in these conditions and a new owner will restore the property to productive use (Mallach 2011; Kelly 2016). In weaker markets, however, these properties frequently remain *vacant* for extended periods of time, generating blight as well as dangerous conditions as structures begin to decompose (National Vacant Properties Campaign 2005; Butler 2016; Kelly 2016). Such properties are not only the result of weak markets but are also causes of them as their presence impacts surrounding land value through poor appearance and a surplus of supply (Schilling and Logan 2008). Property abandonment and declines in land valuation both depress property tax revenues and, where ownership and/or responsibility for upkeep fall to cities and counties, municipal expenditures are increased (Mallach 2018). These self-reinforcing effects together create a cycle of decreasing value and increasing costs difficult for cities, neighborhoods, and remaining residents to recover from. The result is often extreme levels of geographically clustered vacancy which have fundamental impacts on the physical, social, and economic fabric of neighborhoods. Alan Mallach calls this effect *hypervacancy*, a label he assigns to any geography experiencing a vacancy rate of over 20% (Mallach 2018, 5).

Additionally, a significant distinction exists between vacant structures (in which an unused structure exists on a lot) and vacant land (in which no structure exists). In cities and neighborhoods that have experienced large-scale population loss and a decimation of demand, land vacancy is often achieved through intentional demolition—the final step of the disinvestment cycle (Whitaker and Fitzpatrick IV 2013). This demolition is frequently financed with public dollars on properties that have fallen into local government control and is performed in an attempt to decrease the public costs and negative externalities of the vacant structures (Frazier, Bagchi-Sen, and Knight 2013; Mallach 2011). However, scholars have argued elsewhere that demolition simply "replace[s] one problem with another" (Goldstein, Jensen, and Reiskin 2001, 7), and that it is a neoliberal "spatio-temporal fix" for uneven disinvestment and the foreclosure crisis (Rosenman and Walker 2016275). Regardless, demolition is always the first step towards converting vacant properties to green space (Hummel 2015; Schilling and Logan 2008) and publically-funded demolition is frequently a mandatory ingredient for any positive reuse where structures have fallen into extreme disrepair in low value neighborhoods (Kelly 2016, 1028).

Property distress, crime, and wealth

The negative impacts of various stages of property distress on neighborhood health have been thoroughly studied. This section reviews existing scholarship measuring impacts on crime and wealth, the two outcome variables utilized in this paper. The methods and findings of these studies underpin the analysis herein.

The range of theoretical connections between property distress and crime are well reviewed in recent literature (see for example Arnio, Baumer, and Wolff 2012; Ellen, Lacoe, and Sharygin 2013; Cui and Walsh 2015; Spader, Schuetz, and Cortes 2016; Twinam 2017). These theories include links between crime and physical deterioration; the absence of *eyes on the street* (Jacobs 1961); *routine activity theory* (crime requires potential offenders, targets, and an absence of capable guardians; Cohen and Felson 1979); and *broken windows* (Wilson and Kelling 1982).

A growing body of research demonstrates these connections empirically. Ellen et al. (2013) provide a robust analysis of foreclosures in New York City, using both ordinary least squares and negative binomial difference-in-differences models to demonstrate significant causal linkages between the number of properties in active foreclosure and crime at the blockface (a single city block) level. They find that each additional foreclosure on a block leads to an uptick in crime between 0.7% and 1.4%. Building on Ellen et al.'s work, Cui and Walsh (2015) study foreclosures and vacancy in Pittsburgh, Pennsylvania, by comparing the number of crimes within a radius of each distressed property to a wider donut-shaped control area, again with a D-in-D methodology. They find that increases in crime are driven by foreclosure-associated vacancies (rather than foreclosures per se), and that longer vacancies lead to greater increases in crime.

Some studies in this vein use Chicago as a proving ground: Twinam (2017) looks broadly at land use and crime in the city, using an instrumental variable-informed circle and donut method. He finds that higher density mixed-use areas experience less crime than the average residential area and that increasing residential density is weakly correlated with decreased crime—both arguments suggest that repurposing vacant lots could have positive neighborhood effects. Spader et al. (2016) seek to demonstrate that demolishing or rehabbing vacant structures can reduce crime. Using data from Neighborhood Stabilization Program projects in Chicago, Cleveland, and Denver, they also construct a D-in-D circle-in-donut model; they find that demolitions reduce certain types of nearby crime in Cleveland but that this impact lasts only four quarters; they report no significant results for Chicago or Denver.

Research has also connected property distress to decreased home values. Immergluck and Smith (2006), again using a circle-and-ring methodology, find in a Chicago study that each foreclosure within 1/8 of a mile of a given home reduces the value of that home by at least 0.9%. They also note that foreclosures are more likely to occur in low-income census tracts and the effect of foreclosures in these tracts is higher (1.4% decrease in price per foreclosure). Campbell et al. (2011) find that recent foreclosures within a ¼ mile radius of a home decreases the natural log of its selling price by 1.7%, and for foreclosures even closer (within 1/10 of a mile), the effect increases to 8.7%. Whitaker and Fitzpatrick (2013) argue that foreclosures are actually a lagging indicator of property neglect and a non-comprehensive predictor of vacancy; they use property tax delinquency as well as vacancy data from

USPS in their analysis of distressed properties in Cuyahoga County, and find that within 500 feet of a property for sale in low-income neighborhoods, the presence of vacant properties reduces sale price by 1.1%, tax delinquent properties by 2%, and vacant and delinquent properties by 4.6%.

Overall, this body of work contains compelling theoretical and empirical arguments in support of the idea that concentrated property distress can have substantial impacts on neighborhood health. This work implies (or demonstrates, as in Twinam 2017) that making city blocks more vibrant can have a positive spillover effect on the lives of neighborhood residents. And yet, it is worth inspecting whether attempts to do so actually have these predicted impacts. After all, as explained by Spader et al. (2016), urban renewal, public housing development, and property demolition programs have all been motivated by similar intentions, with mixed results.

Review of vacancy management policies

Because of the opportunities inherent in urban vacancy as well as the problems theoretically solved by vacancy reduction, policies that encourage the productive reuse of vacant urban land have garnered support from many corners. Advocates include those interested in equity and social justice, land-use efficiency, and municipal fiscal health. In recent decades many cities have sought to encourage redevelopment of vacant urban land. In this section, I review a non-comprehensive selection of such programs to capture the diversity of extant policies.

In some cases, cities see municipally-owned vacant property primarily as an opportunity to expand their stock of affordable housing. For example, beginning in the late 1980s, Portland, Oregon and Richmond, Virginia both transferred ownership of city-owned vacant property to community development corporations with the express direction that they establish affordable housing on the sites (Schilling 2002). More recently, since 2014 New York City has sold 200 lots to housing developers for $1 in exchange for commitments to keep housing units affordable, often for between 20 and 60 years (596 Acres 2018).

Certain programs are designed to target specifically vacant and blighted *structures*. In the early 1990s, San Diego, California found success working with owners of privately held vacant buildings, tightening codes and changing enforcement practices while helping owners renovate, demolish, or sell (Schilling 2002). Buffalo, NY offers up derelict homes for $1 via their four-decade-old *Urban Homesteading Program*; buyers must renovate them within 18 months and live in them for at least 36 (City of Buffalo n.d.; Fahey 2015). During the recession, the federal Neighborhood Stabilization Program funded demolition and rehabilitation of vacant homes in low-income neighborhoods hard-hit by the foreclosure crisis in cities across the United States (Spader, Schuetz, and Cortes 2016).

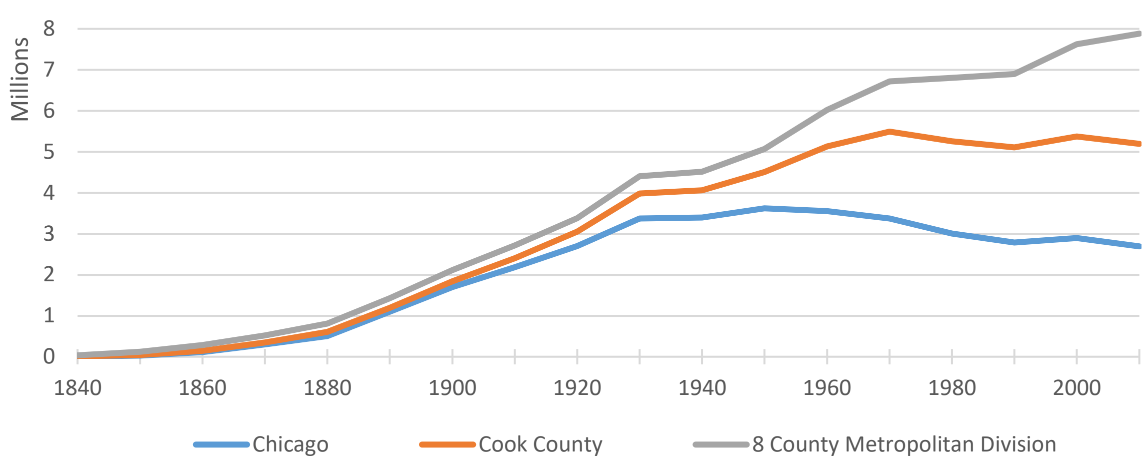
A select few programs target non-homeowners specifically. Detroit, for example, administers the unique *Occupied Buy-Back Program*, in which tenants of properties which are foreclosed upon by their owners get the opportunity to purchase the home for $1000 (Gallagher 2018). At least 400 people have become homeowners through the program, which includes financial counselling services (Hunter 2018).

Finally, policies frequently referred to as *side yard* programs have focused on transferring ownership of vacant land to nearby homeowners. Columbus, Ohio sells unused parcels to any interested buyer for market value but 50% rebates are available to owner-occupants adjacent to available parcels if they keep the land mowed for three years; another dollar-for-dollar rebate for improvements made to the land can cover the other 50% (City of Columbus 2016). In Saint Louis, Missouri, owners of occupied residential or commercial structures adjacent to vacant lots owned by the city can pay a $125 application fee and mow the land for two years in exchange for the deed (St. Louis Land Reutilization Authority n.d.). In New Orleans, any landowner sharing a common boundary with an available property can buy it at market value but the city will credit owners for improvements made (NORA n.d.). In Detroit, more than 10,000 city-owned vacant parcels have been sold; owners can buy lots next to their homes for $100 (Gallagher 2018). Cleveland, Pittsburgh, Buffalo, and Philadelphia have similar programs.

Vacancy and municipal sales of vacant lots in Chicago

Chicago, which had a 2010 population more than 25% lower than its 1950 peak, is one of America's shrinking cities (Mallach 2017; US Census Bureau and Steiner 2018). The effect of suburbanization on Chicago's growth patterns are clearly visible in Figure 1, which documents a shift of regional population growth away from the city, first into Cook County (1930-1970), and more recently into the surrounding metropolitan region. In the last fifty years, the city has also experienced substantial income bifurcation, with middle-class neighborhoods declining from 50% to 16% of census tracts since 1970, and requisite increases in both higher and lower income areas (22% and 62% in 2017, respectively) (Lutton 2019).

Figure 1: Population in Chicago and its metropolitan region



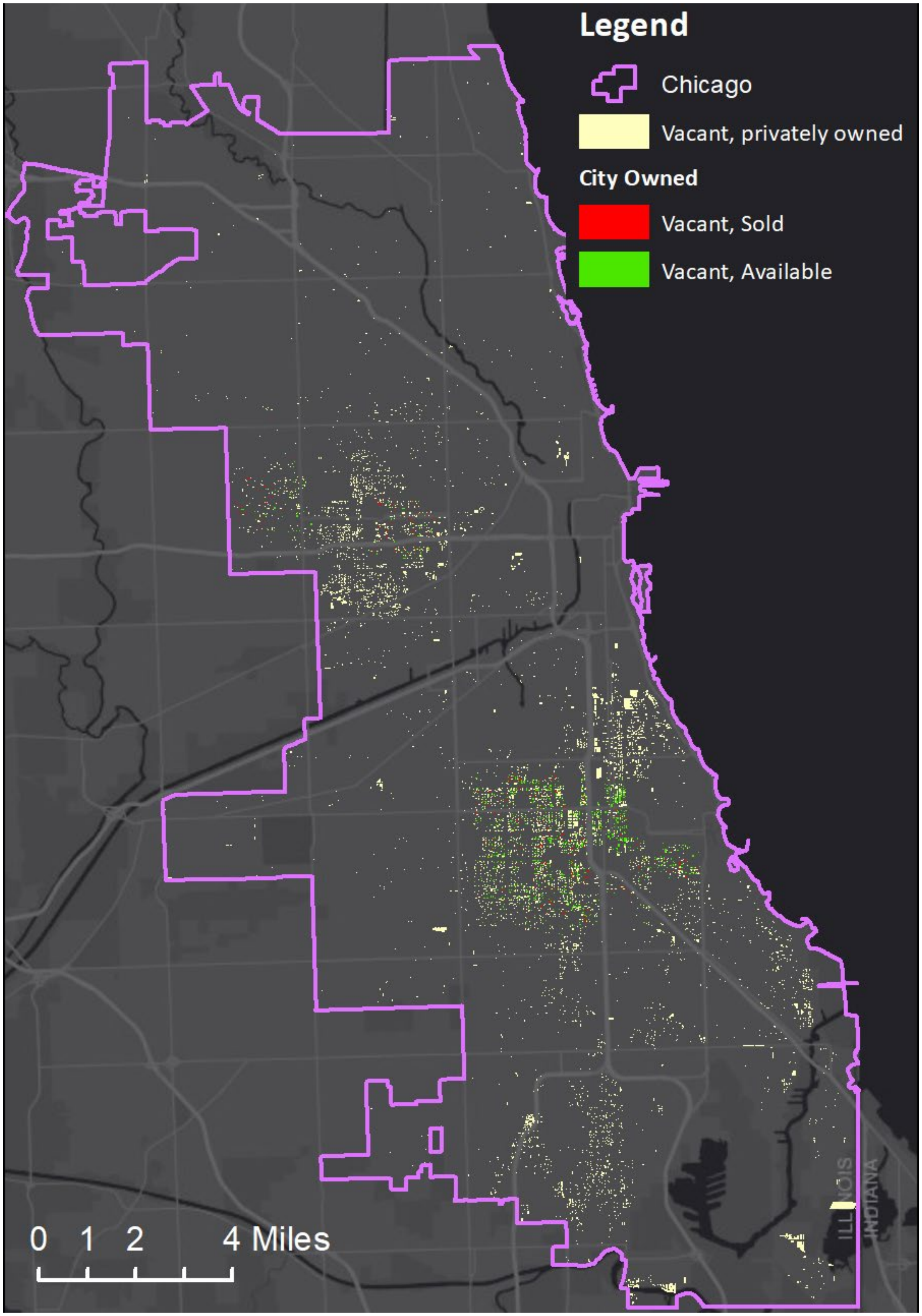
Source: US Census Bureau and Steiner 2018; Minnesota Population Center 2011

The city's south and west sides, home to the highest concentrations of poverty, also contain the substantial majority of the city's residential vacant land. In all, according to the Chicago Metropolitan Agency for Planning's 2013 land use inventory, there are 27,000 vacant residential parcels in the city, occupying over 2600 acres.[2] Figure 2 demonstrates the overwhelming concentration of these parcels to the south and west of downtown. By its own count, the City of Chicago as of February 2019 owns 11,600 of these parcels, and has sold an additional 4,200 in recent years.[3] These parcels are similarly concentrated in the south and west sides.

[2] To arrive at these numbers, I imposed CMAP's land use inventory, which captures congruous land use areas, on the Cook County Assessor's parcel map, and limited results to within Chicago city limits.

[3] Data from the Chicago Data Portal's City Owned Land Inventory, retrieved 2/20/19. It is believed that this data remains not completely accurate, although city staff are actively working to improve the city's land inventory system (Dickhut personal interview, 2018).

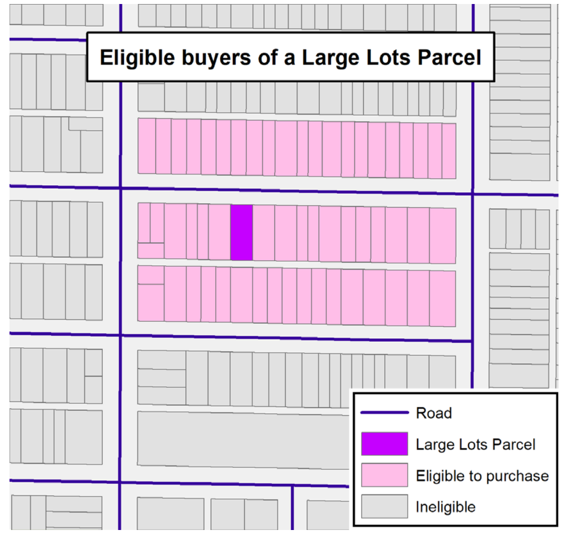
Figure 2: Residential Land vacancy in Chicago



Data from the Chicago Metropolitan Agency for Planning 2014 Land Use Inventory, and Chicago DPD

Like many of the cities described above, Chicago has offered vacant, city-owned parcels for sale for decades. Though the Adjacent Neighbors Land Acquisition Program (ANLAP), which has existed since the 1980s, owner-occupants adjacent to city owned property could buy a single lot at below-market-rate. However, this program saw limited success and some viewed ANLAP as "actually preventing," rather than facilitating, the purchase of vacant lots (Drummer, qtd at 2:05 in ChiHackNight 2017; Dickhut 2018).

Figure 3: Large Lots eligible buyers



In the early 2010's, community organizations in Englewood, one of the neighborhoods in Chicago's south side, began working with the city's Department of Planning and Development to develop a program to replace ANLAP. The result of this collaboration, called the Large Lots Program, was launched as a part of the city's Green Healthy Neighborhoods plan in 2014. The Large Lots program loosened many of ANLAP's requirements, including broadening eligibility from owner-occupants to owners (including nonprofit block clubs and corporations such as churches): buyers no longer need to reside on the block. According to DPD Deputy Commissioner Kathy Dickhut, this change was made because many homeowners in Englewood—often African American families who had moved to Chicago during the Great Migration— maintained close ties with their property and neighborhood despite having moved elsewhere since (Dickhut 2018). Eligible Large Lot purchasers may own anywhere on the same block as an available parcel (including across the street or across the back alleys common in the city; see Figure 3), can purchase two parcels per program round, and must retain ownership for five or more years. As long as they comply with city code, there is no restriction on what an owner does with their property. More thorough detailing of the requirements of both the ANLAP and Large Lots Program can be found in Table 1.

The Large Lots program was first implemented in the Englewood neighborhood in 2014, where over 300 parcels were sold (ChiHackNight 2017). Since that initial success, the program has been repeated in four other neighborhoods (East Garfield Park, Austin, Roseland/Pullman, and Auburn/Gresham) and then two rounds that included parcels from across the south and west sides. In total, 1,248 sales have been recorded through the program, although this figure does not yet include sales in the second (2018) expansion round, which as of writing are not believed to have been finalized. The program is expected to continue (Dickhut 2018).

Table 1: Regulations of Chicago's ANLAP and Large Lots Program

	ANLAP	Large Lots Program
Eligible parcels	City-owned, vacant, residentially zoned lots	
Cost	$1000 or more per lot, depending on assessed value	$1 per lot
Eligible buyers	Owner-residents immediately adjacent to eligible parcel	Owners of land on same block as eligible parcel, including nonprofit block clubs and corporations
Max purchase	1 lot	2 lots per program cycle
Sale restrictions	Cannot be sold separately from primary parcel within first 10 years	Cannot resell for 5 years
Other restrictions	In first 10 years, improvements limited to landscaping or construction of garage or integrated house extension	*None.*
Purchase process	Rolling applications	Interactive website; door-to-door canvassing about program; applications accepted in cycles
Aldermanic Approval	Letter of support from alderman required with application	Alderman approves list of properties to be offered during a cycle, then approves sales at end of cycle

(City of Chicago n.d.; ChiHackNight 2017; Dickhut 2018)

Extant evaluations of vacancy management

Although the study of urban shrinkage is not new, not enough scholarly attention has been focused on how cities have responded to shrinkage (Großmann et al. 2013). Few studies, for example, have evaluated the impacts of side yard programs. Margaret Dewar, one of the few authors who has studied this topic, suggested over a decade ago that "existing research on how to handle vacant and abandoned properties is strong on recommendations but weak on evaluation, especially for programs currently operating" (2006, 167); a review of the literature since then suggests that little has changed. More recently, Nassauer & Raskin (2014) posed a series of yet-to-be-answered questions about the impacts of "patchy" land vacancy—and informal and formal responses to it—on social capital and other "socio-ecological systems."

What studies do exist tend to focus on institutional hurdles to vacant land disposition and evaluate cities' relative successes at managing these challenges. Dewar, for example, has compared Cleveland's and Detroit's disposition efforts recognizing the importance of embedding sales in wider planning strategies, using land banks to clear titles before sale, and implementing programs to encourage positive reuse of sold parcels (2006). Ganning and Tighe, in a study of St. Louis, Missouri's side yard program, evaluate the policy constraints that prevent a large number of sales in that program. They find that local policy structure and state regulations both have substantial impact on success of lot sales but do not seek to measure the impacts of the sales on buyers or the surrounding community (2015). One notable exception is Schilling's case study of Richmond, Virginia's Neighborhoods In Bloom program, which found a 37% reduction of violent crime and a 19% reduction in property crime in the program's target neighborhoods after two years (Schilling 2002, 17).

Some analysis has been done on what owners of new parcels do with their land. Dewar has studied outcomes of managed versus auction sales in Flint and Detroit, finding that positive property reuse is much more likely following a managed sale such as through a side-yard program (2015). In an ongoing study of Chicago's Large Lots Program, a team led by Gobster and Stewart are performing visual assessments of improvements made by new parcel owners (Gobster et al. 2018).

There appears to be no formal evaluations of program success at a more fundamental level. The goals of policymakers who seek to dispose of municipally-owned vacant land are both direct (decreasing the city's financial liability and increasing tax revenue) and indirect (improving neighborhood health, reducing crime, encouraging reinvestment, and rebuilding wealth). No recent formal scholarship exists investigating how successful existing policies are at achieving this second set of goals.

Furthermore, critical scholars present arguments that complicate the narrative of urban blight management and vacancy management all together. Kelly suggests that vacancy management programs must grapple with a the dual realities that increasing development in distressed areas requires an influx of private capital and yet the market-based approaches that attract that capital are likely to reinforce segregation and inequality (2016, see pp. 1026, 1034). Sefransky, in study of the urban open space trend in Detroit, suggests that the greening of poor urban neighborhoods of color may be acts of urban colonialism that reclaim space from the poor for the benefit and profit of those better-off: "accumulation by green dispossession" (2014, 244). Rosenman and Walker, who review Cleveland's post-recession investment in intensive demolition, characterize extreme hypervacancy and that city's "demolition coalition" in a neoliberal framework, and characterize choices about how to manage high foreclosure rates as an "austerity machine" (2016). If demolition is "a necessary form of public investment to groom the urban landscape for future growth," (2016, 283) is the transfer back into the private market of vacant parcels that have been "groomed" both physically and legally (in terms of a clean title) simply the next step in a neoliberal process of uneven capital accumulation? These questions further motivate evaluations of the impacts of vacancy management policies.

III. Research Strategy & Data Methods

Research questions and analysis geography

To address the gap in scholarship described above, this study seeks to evaluate Chicago's Large Lots program and investigate what impacts lot sales have had. Considering the breadth of the program's desired direct and indirect outcomes, many angles of potential evaluation exist. This project seeks to opportunistically answer three distinct research questions based on available quantitative data. This section describes, motivates, and outlines a methodology to answer each research question.

First, who is buying Large Lots parcels? The Large Lots program seeks to transfer control of land to and generate wealth within the communities where the conjoined problems of land vacancy and suppressed property value are most severe. To keep these gains local, the program gives local alderman veto-power over specific sales and requires buyers to already own property on the same block. As long as they can demonstrate existing ownership, buyers can be individuals or incorporated groups such as (but not limited to) block groups, non-profits, or limited liability corporations. There is no mechanism for block- or neighborhood-residents who are not owners to participate in the program, and no requirement that owners actually live in the community to be eligible. I use the names and mailing addresses recorded on Large Lots deed transfers to explore who has purchased parcels, how many parcels they have purchased, and where they live.

Second, has the Large Lots program had an impact on crime? Originators of the Large Lots program argued that returning neighborhood land to local control, increasing community self-determination over land use, and decreasing the number of vacant lots would have a positive impact on crime levels. The literature suggests this argument has merit. I perform a block-level panel statistical analysis to compare the location of lots available and lots sold through the Large Lots program to the frequency and severity of crime.

Third, how much wealth creation is the Large Lots program expected to generate? One of the program's initial stated goals was to support the generation of wealth in communities that have been victims of historical disinvestment and whose residents have experienced institutional roadblocks that have limited access to wealth-generating homeownership. Transferring developable lots to community members at only a nominal cost at a time when property values across the city are rising—in some neighborhoods quite quickly—has the potential to achieve this goal. I use neighborhood-based property value trends and assessed values of sold lots to project the value of property transferred through the program, and therefore the expected generation of individual wealth, over time.

Due to data availability this analysis focuses on the 844 Large Lots sales conducted in three broadly construed sale 'neighborhoods', which I call *study areas*, between 2014 and 2017. Studied parcels include those sold during the first three neighborhood-specific pilot implementations of the program as well as those sales from the first all-neighborhood expansion round (LL6) that occurred in the neighborhoods previously treated in the first three pilots. Details of Large Lots sales considered and not considered in this analysis are reviewed in Table 2.

Table 2: Large Lots Program and Analysis Subset

PROGRAM	DATE	SALES INCLUDED IN ANALYSIS	SALES NOT INCLUDED
LL1: Greater Englewood	Spring 2014	279	
LL2: East Garfield Park	Summer 2014	155	
LL3: Austin	Fall 2014	50	
LL4: Roseland Pullman	Fall 2015		15
LL5: Auburn Gresham	Spring 2016		32
LL6: Expansion 1	Fall 2016	360	346
LL7: Expansion 2	Spring 2018		All (data not yet published)
TOTAL		**844**	**404**

Data: author's analysis of multiple data sources. Sums are not always perfect due to data source inconsistency.

Data sources & dataset construction

To answer these questions, data from seven distinct data sources were combined. These sources are described below and summarized in Table 3.

- The Large Lots website (www.largelots.org) makes certain data available for download (LISC Chicago and DataMade n.d.). Data available was found to include parcel eligibility for LL1, LL2, and LL3, and sales for LL1 and LL2.
- Staff at the City of Chicago Department of Planning and Development (DPD) directly provided additional Large Lots data. This data included Large Lots sales for all programs through the first expansion round (LL6). Where this data overlapped with largelots.org data, data was relatively congruent; where sources disagreed, DPD data was preferred.
- Cook County's 2014 tax assessment parcel data was collected from the County's open data portal (https://datacatalog.cookcountyil.gov/). This data included polygons for all parcels in the county and their 14 digit parcel identification number, crucial in establishing the location of Large Lots parcels (Cook County Clerk 2016).

- Parcel-level land use data was obtained from the Chicago Metropolitan Agency for Planning's 2013 Land Use inventory, available from CMAP's data hub (https://datahub.cmap.illinois.gov) (Chicago Metropolitan Agency for Planning 2017).
- Crime, street grid, city-owned parcels, and Chicago community areas (neighborhoods) data were obtained from the City of Chicago's open data portal (https://data.cityofchicago.org). This data was used to generate the unit of analysis and dependent variables for research question 2 (City of Chicago 2019).
- Transaction and land valuation data was obtained from the Zillow ZTRAX database (Zillow 2019).[4] ZTRAX, a national dataset of real estate assessment and transaction information available to academic researchers, provided crucial data that underpins analysis for all three research questions. This data includes transaction dates, transaction values, detailed information on buyers and sellers, and parcel size and assessment information from the county assessor.
- Finally, neighborhood price trends were obtained from the Cook County House Price Index produced by Institute for Housing Studies at DePaul University (https://price-index.housingstudies.org/) to inform the forecasting component of research question 3 (Institute for Housing Studies 2018a, 2018b).

Table 3: Data and Sources

DATA	SOURCE	ACCESSIBILITY
Crime	Chicago Data Portal	Public
Roads	Chicago Data Portal	Public
City-owned parcels	Chicago Data Portal	Public
Chicago Community Areas (neighborhoods)	Chicago Data Portal	Public
Parcels	Cook County Data Portal	Public
Land Use	CMAP Data Hub	Public
Program Lot Status	LargeLots.org	Public
Program Lot Status	Chicago DPD	Private
Transaction dates	Zillow ZTRAX	Private
Transaction values	Zillow ZTRAX	Private
Buyer and seller data	Zillow ZTRAX	Private
Tax assessment data	Zillow ZTRAX	Private
Neighborhood price trends	DuPaul Inst. Housing Studies	Public

The core analysis work addressing all three research questions was performed using a dataset constructed by the author that combines the individual data sources described above. Dataset construction can be understood as two distinct but overlapping processes: ZTRAX data construction and data assembly.

ZTRAX data construction: Zillow ZTRAX is a nationwide dataset containing terabytes of data on decades of real estate transactions and assessments, itself assembled from third party data sources. The data available to researchers is stored in state-based tables that represent exports from a relational database. ZTRAX's Illinois transaction and assessment data was combined and subsetted to produce a record of pertinent data on all properties in the three study areas. This was performed by writing a STATA program capable of extracting specified fields from certain ZTRAX transaction tables and merging

[4] Access to the Zillow ZTRAX database was made possible by Dr. Bill Lester, who is using the data for ongoing research. This project contributes analysis to that ongoing work.

them into a single dataset. The exact methodology for this stage is detailed in the Stata log in Appendix A, and is summarized here.

The program amalgamates data from five ZTRAX transactions tables (main [transaction details], property info, buyer name, buyer address, and seller name) and two ZTRAX assessment tables (main and value). At times, multiple buyers and sellers are recorded for each transaction. I kept data for the first three buyers and the first three sellers for each transaction. Additionally, all records in Cook County contained only one single buyer mailing address, which was kept. The ZTRAX data contains up to three date fields per transaction: the date the deed was produced (document date), signed (signature date), and recorded (recording date). Per Zillow, the document date was used preferentially as the transaction date, with the recording date substituted only where the document date did not exist (Zillow n.d.). This occurred only 3 times.

The program tags transactions that are likely to be Large Lots sales in the 3 neighborhoods studied by searching for quit claim transactions (per DPD, Large Lots are transacted almost universally with Quit Claim deeds (Dickhut 2018)) on parcels found to be inside the boundaries of the three study neighborhoods sold by the City of Chicago since January 1st, 2014.

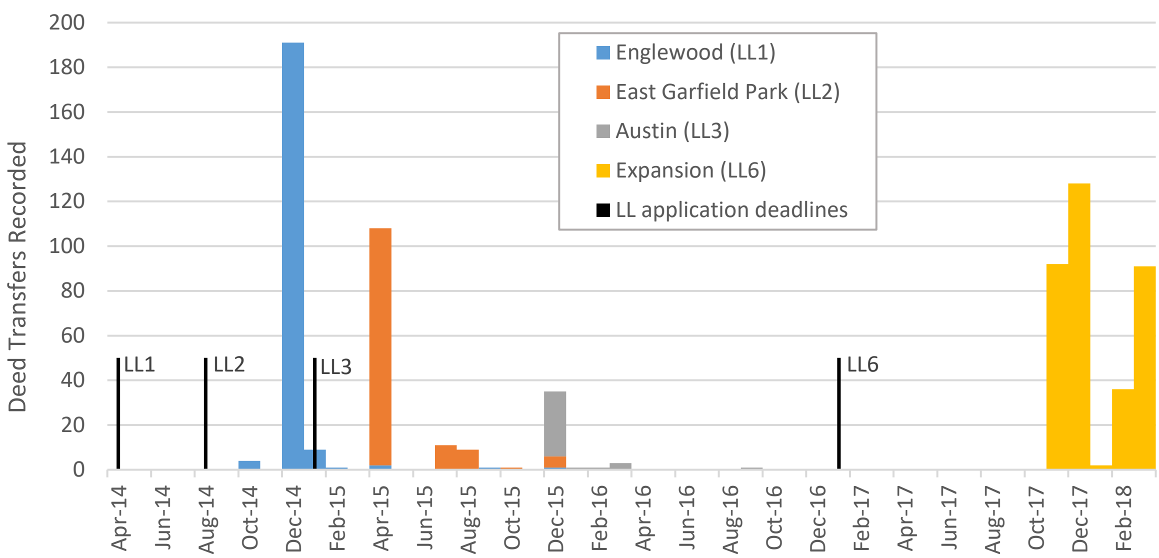
<u>Data assembly</u>: All data sources were assembled in ArcGIS. This process is summarized and contextualized here; a detailed step-by-step methodology can be found in Appendix B.

First, shapefiles from the Large Lots program website containing parcel eligibility and limited sale data for the first three program pilots were combined with parcel data from the Cook County Assessor to produce a map of all parcels in the three study neighborhoods. Land use data from the Chicago Metropolitan Area for Planning was joined with the parcel data to assign every parcel a land use as of 2014. Tabular data on all Large Lots program sales between 2014 and 2018 (LL1 through LL6; see Table 4) received from Chicago DPD was paired with parcels via PIN and the distinct sources of sale data were compared for accuracy. The much-more-comprehensive DPD data source was used for sale information while the parcel eligibility data from the shapefiles was maintained to create the comprehensive parcel status field described in Table 4. When this data is combined with the transaction dates from Zillow we can visualize the actual closing dates of parcels transacted in each Large Lots phase in relationship to the closing of the application window for that phase (Figure 4).

Table 4: Parcel Status

CODE	MEANING	ENGLEWOOD	EAST GARFIELD PARK	AUSTIN
-1	Ineligible parcel	39,991	5,751	21,274
0	Available for sale in LL1, LL2, or LL3, but not purchased in pilot or expansion	3,484	235	166
1	Purchased in LL1 (Englewood)	279		
2	Purchased in LL2 (East Garfield Park)		155	
3	Purchased in LL3 (Austin)			50
4	Purchased in LL6 (First Expansion)	304	24	32

Figure 4: When Large Lots sales occurred



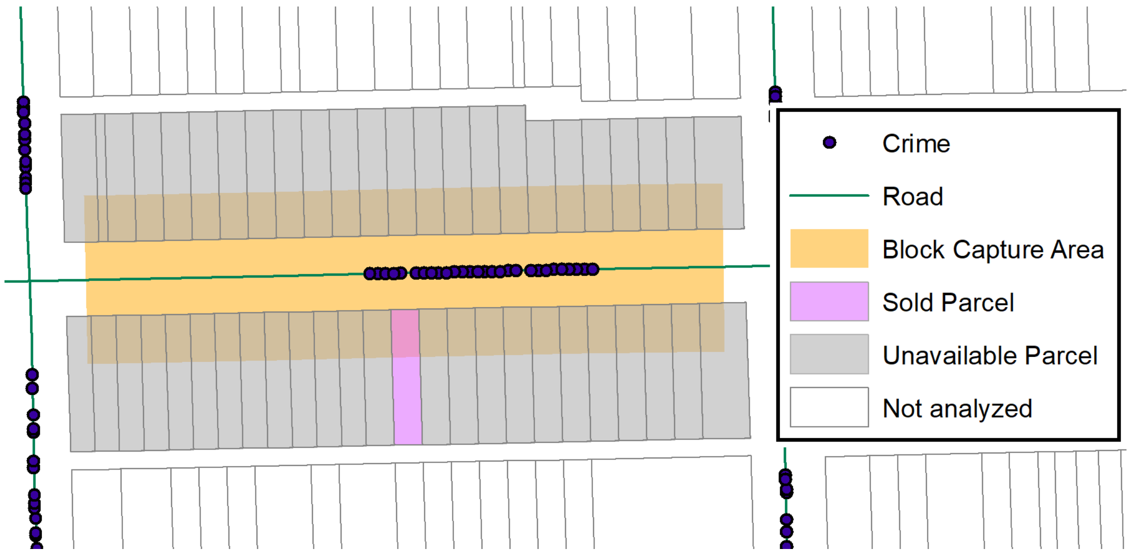
Separately, the City of Chicago's roads shapefile was used to construct blockfaces (single city blocks) and assign parcels to them. Every block in the three neighborhoods was manually inspected to evaluate whether the block was (1) relatively straight, (2) not a major commercial thoroughfare, (3) occupied by primarily residential land uses, and (4) subdivided into primarily single-family-house-sized parcels that (5) faced the street. Blocks that fit all of these standards were determined to be good for analysis in the panel dataset used for research question 2. Additionally, where a single blockface was divided into separate street polylines, these polylines were combined. 24% of blocks in the three neighborhoods, containing 72% of Large Lots sales, are included in the crime panel (See Table 5). All 844 Large Lots sales in the study neighborhoods are included in research questions 1 and 3.

Table 5: Blockfaces included in panel for Research Question 2

CATEGORY	BLOCKFACES	SOLD PARCELS
Included in crime panel (RQ 2)	1,345	611
Not included in crime panel	4,221	233
Total	5,566	844
% Included in crime panel	24.2%	72.4%

The polylines representing single blockfaces included in the crime panel were shortened by 50 feet from either end to avoid accidentally capturing parcels across streets, then inflated width-wise by 75' on either side to create block capture areas. These capture area polygons were then joined to the parcel data to assign each parcel to the correct blockface. Any parcel on a street corner where two analysis blocks meet was duplicated for the sake of the crime panel as it is reasonably expected that the impact of that sale would accrue to both blocks. However, this occurred for only 212 parcels (0.3%), of which only 10 were eligible for Large Lots sales and only 3 of which were sold (0.4%). Furthermore, crime data, which was geocoded based on city-provided coordinates that are anonymized to the blockface-level (it is not known where exactly on the blockface each crime occurred), was assigned to a blockface using the block capture polygons. This process is visually depicted in Figure 5.

Figure 5: Blockface capture design



IV. Methodology

RQ1: Who is buying Large Lots parcels?

To evaluate who is purchasing Large Lots parcels, I inspect Zillow ZTRAX transaction data about those parcels' buyers. Of the 844 Large Lots sales in the study areas, 725 (86%) matching transactions were located in ZTRAX. I first ask who buyers are and how many parcels they buy. More than 99% of records had only one or two buyers listed so this analysis was performed on the first two buyer fields. Using clues from the organization names and basic internet searches,[5] I categorize non-individual buyers into four types: community-based organizations (including block groups, churches, and registered non-profits); developers (LLCs with names that connote property development goals or associated with individuals described as professional developers online); trusts, estates, and other family investment vehicles; and other entities (including where the type is unknown)[6]. To tally parcels purchased by the same buyer, I count multiple instances of the same buyer name *or* buyer address due to typographical errors in the data. Where these two sums differed, data was manually inspected to determine whether the purchaser is likely to be the same entity.

This evaluation also includes an investigation, to the degree possible with available data, of how *local* buyers of Large Lots parcels are. In order to evaluate the proximity of Large Lot buyers to their purchased parcels, I geocoded the mailing addresses of the buyers associated with the transaction

[5] Non-individual buyers with names that included words like "development", "properties", "property management", "capital", and "real estate" were categorized as developers. Additionally, organizations were categorized as "developers" where google searches yielded news articles or other web sources confirming the entity or an associated individual was labeled as a "developer"

[6] It is the author's expectation that these organizations are either individuals who have placed their properties in LLCs for tax and liability reasons or developers.

record in ZTRAX. Mailing addresses are an imperfect correlate for how proximate a buyer might be to their new parcel as they may choose to have the deed sent to a lawyer, PO box, family member, second home or some other location. For this reason the results of this analysis are likely to be imperfect and conservative estimates of proximity.

Geocoding was completed using the DM10 address locator database for ArcMap available through UNC Libraries. All but 3 of the 725 Large Lots sales with associated ZTRAX data contained buyer addresses. Of these, 670 geocoded successfully, for a total sample of 79% of the 844 sales in the study area. Using ArcMap tools in the Illinois State Plane coordinate system, geocoded buyer mailing addresses were linked to the purchased parcel(s) and the distance between the two was measured. Because of the coordinate system used the distances associated with non-Illinois mailing addresses will be inaccurate but this is irrelevant considering the significant finding in this situation is only that the distance is very large. *Buyer-parcel distances* were calculated in feet and then categorized as follows:

1. buyer within 1000 feet (i.e. on the same block)
2. buyer in neighborhood (both parcel and buyer address fall into same named neighborhood, as defined by the City of Chicago) (all are within 10,000 feet)
3. buyer within 10,000 feet, but not same neighborhood
4. buyer within 24,000 feet
5. buyer within Chicago metropolitan area
6. Buyer outside Chicago metropolitan area

RQ2: Do Large Lots sales reduce crime?

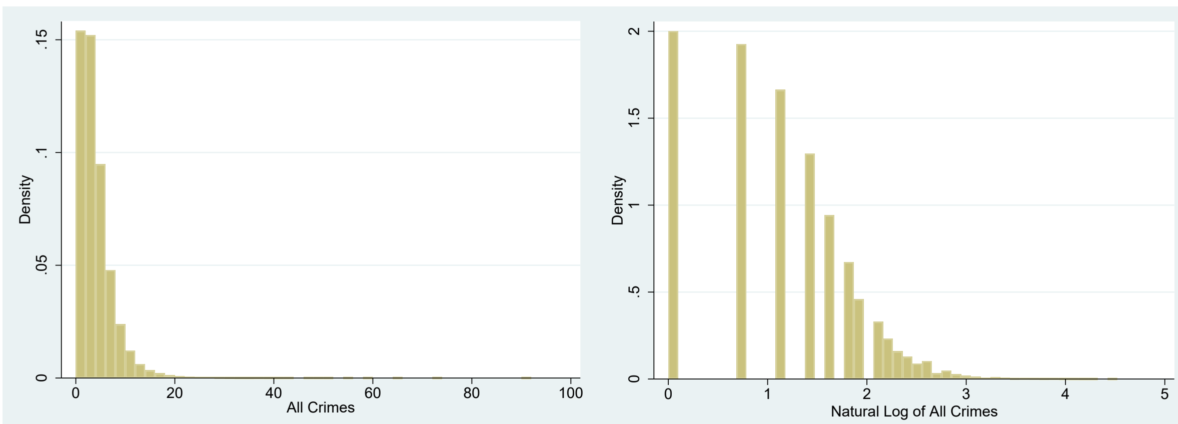
In order to move beyond measuring correlation to determine whether Large Lots sales have a causal impact on crime levels I follow the work of Ellen et. al (2013), Cui and Walsh (2015), Spader et. al (2016), and others and address this question using a difference in differences panel regression. I conduct the analysis using the blockface—a single residential block—as my unit of analysis. This follows the methodology and theoretical framework of Ellen et. al, and is the only empirically feasible choice because the City of Chicago's publically available crime data is randomized to the blockface level. As discussed in the dataset construction section only regularly shaped and primarily single-family residential blocks were used to address this question. In all, 1332 blocks were analyzed containing a total of 611 Large Lots sales. 346 blocks received the treatment of one or more lot sale providing a large control group of 986 blocks not impacted by the program.

Crime data from 2010 through 2018 was summarized by blockface and by quarter, defining a 36 quarter panel for the 1332 blocks—a total of 47,952 observations. Following the work of Immergluck and Smith (2006) and Frazier et. al (2013), details available in the crime database permitted the calculation of a number of crime sub-categories that may be particularly susceptible to changes in land vacancy and activation. These categories include narcotic arrests, violent crimes, violent crimes committed in public spaces, property crimes, and the Chicago police department's unique *public violence* crime categorization. Unfortunately, at the blockface-quarter level of detail these subcategorizations yield too few non-zero observations for useful analysis (Table 6). Thus, total crime was used. Due to the non-linear distribution of crime at the blockface-period level (Figure 6), two analyses were conducted: an OLS regression on the natural log of crime, and a negative binomial regression of the crime count.

Table 6: Incidence of non-zero crime observations by type

	Blockface-Period Observations	
Crime categories	**Non-zero**	**Zero**
All crimes	41,321	6,631
Narcotic arrests	14,487	33,465
Violent crimes	13,541	34,411
Violent in public	11,453	36,499
Property crimes	27,264	20,688
Public violence (per CPD)	7,334	40,618

Figure 6: Distribution of the incidence of crime and log crime in blockface-periods



Before running the panel regression I perform basic gut-check regressions correlating crime and vacancy in the study areas before the Large Lots program went into effect. To do this I keep only crime data from 2013 (the year before the program was announced) and compare pre-program vacancy rates to the natural log of crime. These tests are run to confirm or refute the underlying assumption that vacancy and crime are correlated.

I then turn to the full causation-demonstrating difference-in-differences panel regression. The independent (treatment) variable is the occurrence of a sale on a blockface in a period. Because the desired impact of a Large Lots sale is a permanent reduction of local crime the independent variable was recorded as a dummy variable that turns from 0 to 1 in the quarter in which the first sale on the block occurs and remains 1 throughout the remainder of the sample.

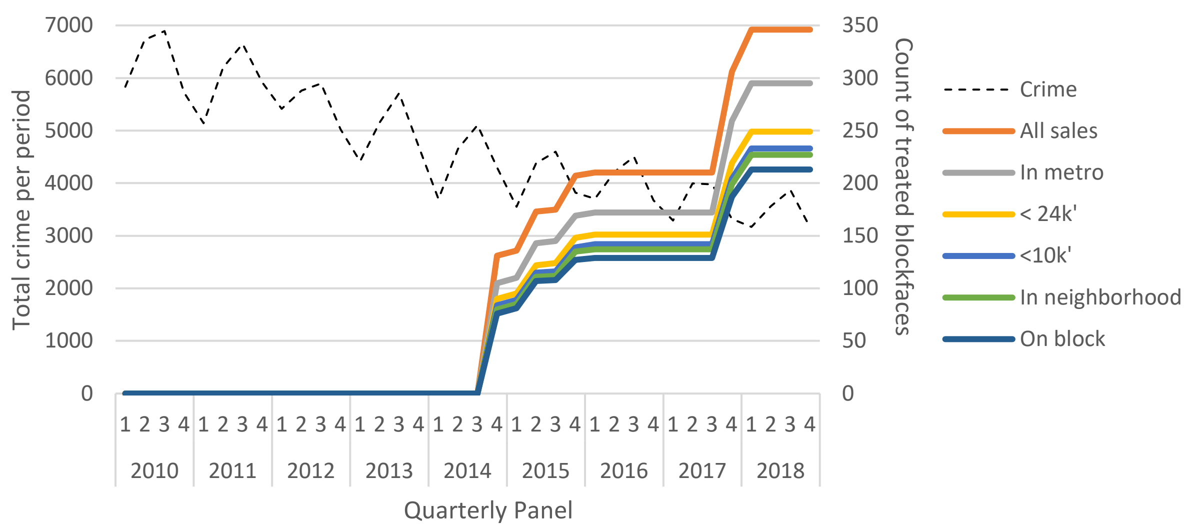
Buyers who live on the block may be more likely to quickly make improvements to the parcels and visit their new property regularly. Thus, I hypothesize that the more sensitive the treatment variable is to truly local buyers the stronger the causal crime reduction effect will be. For this reason, I use the buyer-parcel distance categories from research question 1 to define the following list of treatment variables:

- Treatment 0: All sales
- Treatment 1: Buyer in Chicago metro region
- Treatment 2: Buyer within 24,000 feet (approximately 4.5 miles)
- Treatment 3: Buyer within 10,000 feet

- Treatment 4: Buyer within neighborhood (always <10,000 feet)
- Treatment 5: Buyer on same block (within 1,000 feet)

To specify when treatments first occur, I utilize the signature or recording dates from ZTRAX as discussed in the dataset construction section above. 728 of 844 sales include sale date data; for the remaining 116 sales (13.7%), the median sale date for that sale's Large Lots phase was used. These exact dates were generalized to their quarter. The distribution of treatment over time for each treatment variable, as well as the total quarterly crime trend on the study blocks, can be found in Figure 7.

Figure 7: Distribution of independent and dependent variables over entire crime panel



Because crime levels (y_{bt}) are affected by period effects, and because each blockface in the sample is impacted by a variety of other block-specific factors that may impact incidence of crime, both period fixed effects [T_t] and blockface fixed effects [B_b] are included in the regression. Thus, the specified OLS panel regression is:

$$\ln(y_{bt}) = \alpha + \beta X_{bt} + B_b + T_t + \varepsilon$$

The negative binomial regression uses the same terms, but the outcome variable is y_{bt}. Because it is likely to take time for new owners of the Large Lots to activate their land, and because it may take further time for that activation's impact on crime to take effect, this analysis was repeated twice—seeking contemporaneous effect and again seeking a 4 quarter (1 year) lagged effect.

RQ3: How much wealth is generated?

In order to estimate how much wealth is generated by the Large Lots program, the value of the parcels transferred through the program and the appreciation trends present in the study areas are needed. Directly assessing the value of sold parcels is impossible because, by regulation, owners cannot sell the plots for five years: the owners of the Large Lots parcels transacted in the first pilot will not legally be able to sell until October 2019 at the earliest (See Figure 4). Instead, I indirectly assess the value of sold parcels via three mechanisms before measuring forecasted appreciation. First, I use 2016 tax assessment data for all residential vacant land in each neighborhood; second, I use tax assessments of the sold Large Lots parcels themselves; third, I use actual sale data from other vacant parcels in the same neighborhoods.

- All vacants assessment method (Model 1): In this model, I use average per square foot valuations from 2016 county tax assessments to estimate wealth generated by Large Lots sales. To obtain county tax assessment information, I use the Zillow ZTRAX Assessment tables to extract parcel assessment amounts, parcel size, and years of assessment. Within the three study areas a conservative selection process was used to identify residential vacant parcels: parcels must be either available through the Large Lots program or both be categorized as *vacant residential* by the 2013 CMAP land use inventory and contain less than $1,000 of property improvements in 2016 per ZTRAX. In all, over 11,000 parcels meet these qualifications. Because parcel size was available directly interpreted from the Cook Assessor's parcel shapefile (for all parcels), and from ZTRAX (for 93% of the 11,000 parcels), these two data were compared. It was found that where parcels had ZTRAX assessments but not parcel sizes, utilizing shapefile parcel size generated estimates of value per square foot with similar ranges. Thus, for valuation estimation purposes, shapefile parcel size was used when no ZTRAX parcel size data existed. As described in Figure 8, this was crucial for East Garfield Park where fewer than 10 parcels with the 2016 assessment values had size data in ZTRAX. Assessment per square foot of lot size was calculated and multiplied by ten because Cook uses an assessment level of 10% (Cook County Assessor n.d.). Total wealth generation is then modeled by multiplying the average assessed value per square foot of vacant land in each study area by the total area of lots sold in that area.

Figure 8: Source of lot size data by neighborhood for residentially zoned vacant parcels

	ENGLEWOOD	EAST GARFIELD PARK	AUSTIN	ALL AREAS
USED ZTRAX SIZE	5,163	8	595	5,766
USED GIS SIZE	232	158	43	433

- Large Lots direct assessment method (Models 2a and 2b): Next, I subset the above findings to directly assess properties that have sold through the Large Lots program and calculate both per parcel and per square foot valuations. Matching on parcel identification number, I find that 352 of the 844 sold study parcels contained the necessary data allowing a direct assessment of wealth generation for 42% of Large Lots sales. Total wealth generation is modeled first by using this more narrow average value per square foot multiplied by total area of lots sold (Model 2a) and then by using an average value per sold property with known value multiplied by the total number of parcels sold (Model 2b). Both models are computed for the three study areas independently.
- All vacants sale method (Model 3): In this model I use the same standard for defining vacant land and value per square foot as the all vacants assessment method (Model 1). Here I use actual non-zero-dollar sales transactions that do not involve the City of Chicago or Cook County as buyers or sellers from ZTRAX. I use by-year-and-neighborhood counts and five-year weighted running averages to analyze trends in sale prices over time and use the 2016 running average (2014-2018) purchase price per square foot multiplied by the total area of lots sold for each neighborhood to estimate total value.

I also use the results from research question 1 to allocate wealth generation into proximity and buyer type classes. To do this I calculate a percentage of square feet sold (for models 1, 2a, and 3) and parcels sold (for model 2b) by study area for each class. For each model I then use that percentage to derive an estimate of the wealth generated for each class of buyer.

Finally, I use neighborhood housing price trends from DePaul University's Cook County House Price Index to forecast total wealth generation in 2020 and 2025 (Institute for Housing Studies 2018a, 2018b). The index provides multiple metrics of single family home valuation for a variety of Chicago submarkets, including three that overlap closely with each of the study neighborhoods. From each I use percent change since 2000 (annualized over 18 years), percent change since the value floor during the 2008 recession (approximately 2012 for all Chicago markets, so annualized over 6 years), and percent change since 2017. For each neighborhood I use the minimum and maximum growth rates to define the upper and lower limits of my projection.

V. Findings

Study area context

First, I present basic statistics by study area to define the context in which the Large Lots program is operating in each neighborhood. The site of the first Large Lots pilot, referred to by the program as Greater Englewood, is the largest of the three study areas, with 211 million square feet of non-right-of-way land. It also has the largest vacancy rate by square footage (56 million square feet, or 26.5%). It is comprised of the neighborhoods of Englewood and Woodlawn, and portions of New City, Fuller Park, Washington Park, and Grand Crossing. East Garfield Park is a much smaller area that is congruent with the Chicago neighborhood of the same name; it has a vacancy rate that is slightly greater than Englewood by parcel count but smaller by size. Austin, which is larger than East Garfield Park but still about half the size of Englewood, has a much smaller vacancy rate than the others. It is comprised of the Austin and Galewood neighborhoods. Size and vacancy figures are described in Table 7.

Table 7: Size and vacancy rates by study area

	Englewood		East Garfield Park		Austin	
	Count	Size (sq. ft.)	Count	Size	Count	Size
Total Parcels	44,058	211,618,154	6,165	33,176,188	21,522	136,492,466
Vacant, sold through LL	583	2,442,797	179	630,018	82	368,330
Vacant, LL eligible	3,484	14,526,577	235	713,404	166	708,396
Vacant, other	8,614	39,054,599	1,487	6,100,009	1,160	6,733,649
Not vacant	31,377	155,594,181	4,264	25,732,757	20,114	128,682,092
% Vacant land	28.8%	26.5%	30.8%	22.4%	6.5%	5.7%
% Available through LL	9.2%	8.0%	6.7%	4.0%	1.2%	0.8%
% Sold through LL	1.3%	1.2%	2.9%	1.9%	0.4%	0.3%

RQ1: Who is buying Large Lots parcels?

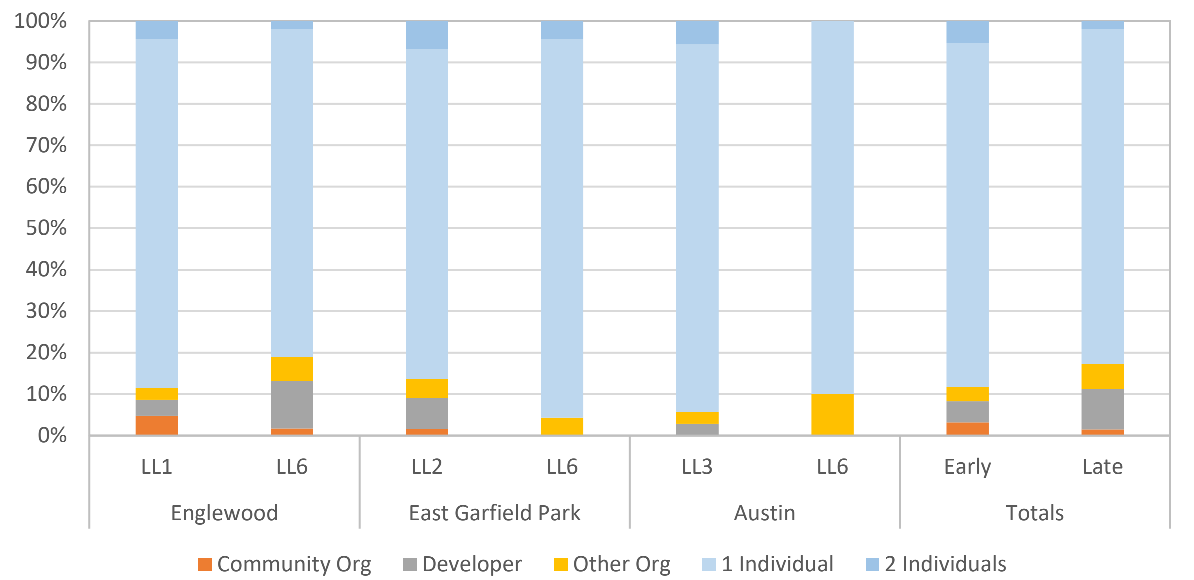
What types of entities buy parcels? I broadly categorize buyers into individuals and non-individuals—some type of organization or incorporation. A substantial majority (nearly 86%) of buyers were individuals, while the remaining buyers were recorded in ZTRAX as non-individuals. Because Large Lots 6 included sales in all three study neighborhoods, I have two samples for each neighborhood: the initial pilots in 2014-2016, and the expansion in 2017-2018. Overall, purchases by organizations and incorporations increased from 12% in the initial pilots to 17.2% in the expansion round. This activity is driven by growth in non-individual purchases in Englewood (+7.4%) and Austin (+4.3%), while non-individual purchases decreased in East Garfield Park (-9.3%).

Further, I find that developers have purchased at least 7% of Large Lots properties, while community-based organizations have purchased only about 2%. In Englewood, which has seen more than half of Large Lots sales, developer activity increased dramatically from 8 parcels (3.8%) in the first round to 34 (11.5%) in the second round; community organization decreased across these rounds from 10 parcels (4.8%) to 5 (1.7%). Conversely, developer purchases dropped to zero between 2014 and 2016 in both East Garfield Park and Austin. Considering that alderman have the ability to veto any individual sale, it may be that local leaders in Englewood continued to see value in developer purchases while leaders in the other neighborhoods did not see value in subsidizing further development. These findings are described in Table 8 and Figure 9.

Table 8: Typology of buyers per Large Lots phase

Neighborhood & Phase / Buyer Type	Englewood		East Garfield Park		Austin		Early (LL1-3)	Late (LL6)	All Sales	% of Total
	LL1	LL6	LL2	LL6	LL3	LL6				
One Individual	176	234	105	21	31	27	312	282	594	81.9%
Two Individuals	9	6	9	1	2	0	20	7	27	3.7%
Individual buyers	185	240	114	22	33	27	332	289	621	85.7%
Community Org.	10	5	2	0	0	0	12	5	17	2.3%
Developer	8	34	10	0	1	0	19	34	53	7.3%
Trusts, Estates, etc.	0	6	1	0	0	0	1	6	7	1.0%
Other/Unknown	6	11	5	1	1	3	12	15	27	3.7%
Non-Indiv. Buyers	24	56	18	1	2	3	44	60	104	14.3%
All Sales	209	296	132	23	35	30	376	349	725	
% Non-Individual	11.5%	18.9%	13.6%	4.3%	5.7%	10.0%	11.7%	17.2%	14.3%	
% Community Org.	4.8%	1.7%	1.5%	0.0%	0.0%	0.0%	3.2%	1.4%	2.3%	
% Developer	3.8%	11.5%	7.6%	0.0%	2.9%	0.0%	5.1%	9.7%	7.3%	

Figure 9: Parcels purchased by buyer type



<u>How many parcels do buyers purchase?</u> I look at quantity of parcels purchased by each buyer. By policy, the same individual or organization is eligible to purchase a maximum of two parcels per program round (Dickhut 2018). Very few buyers were active across phases despite being eligible to purchase parcels during each phase: only five buyers in Englewood and 3 buyers in East Garfield Park purchased in both the neighborhood-specific early program phase (LL1/LL2) and LL6. Additionally, one buyer purchased across neighborhoods, buying in East Garfield Park in LL2 and again in Austin in LL3.

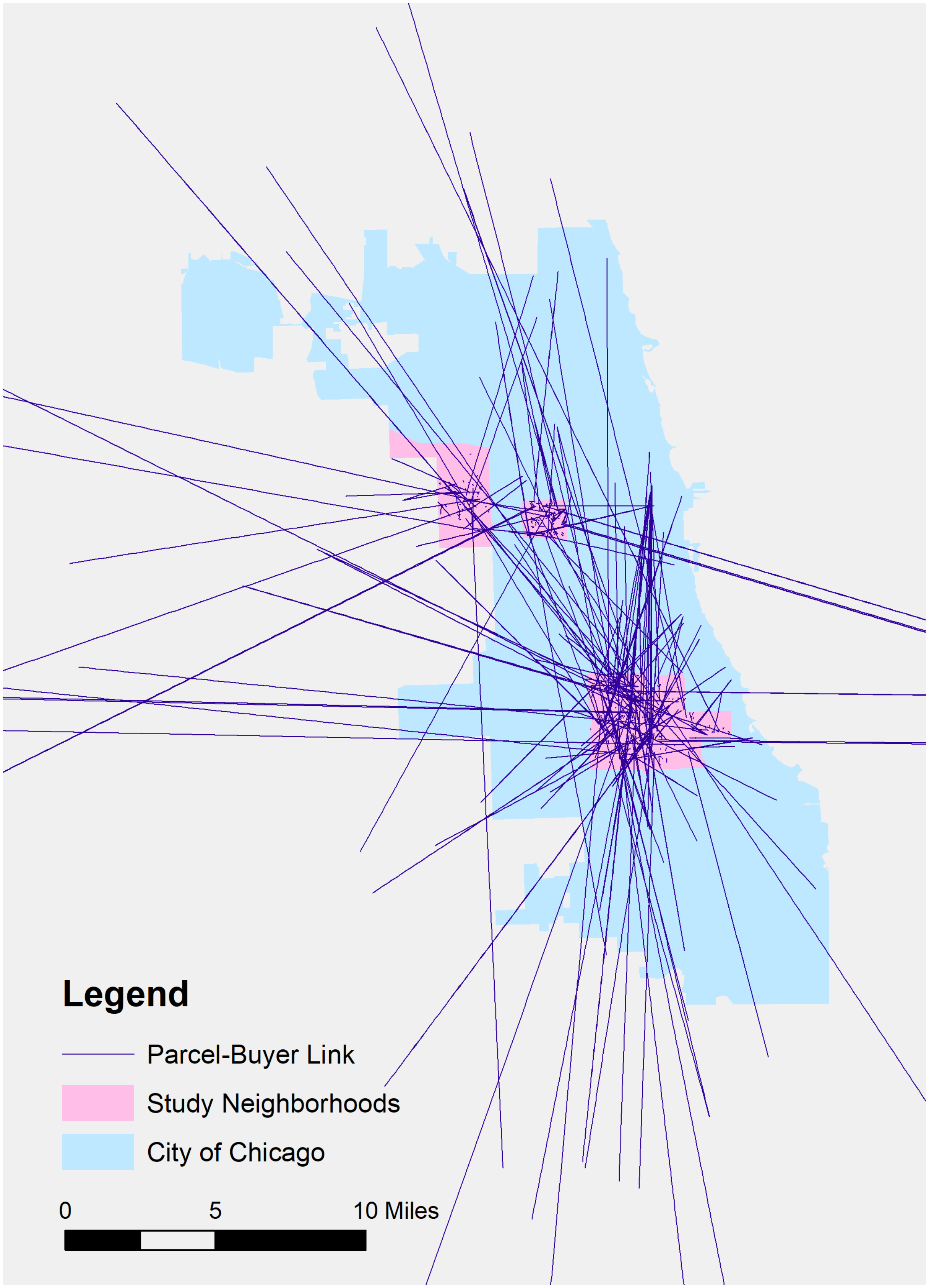
Despite the *de jure* 2 parcels per round maximum, 98 parcels (13.5%) appear to have been purchased by 22 buyers in violation of this rule. Culprits appear to be primarily developers and individual buyers operating in Englewood. No non-compliant purchases were made in Austin and non-compliant purchases dropped to near zero in East Garfield Park across program phases. In Englewood, non-compliance doubled across phases including the apparent purchase of 19 parcels by a single developer in a single phase. Table 9 describes these findings.

Table 9: Buyers by number of lots purchased

Buyers by quant. \ Phase	Englewood		East Garfield Park		Austin		All early (LL1-3)	All Late (LL6)	All buyers
	LL1	LL6	LL2	LL6	LL3	LL6			
Individuals who bought…									
1 parcel	103	92	55	13	29	25	187	130	317
2 parcels	34	60	22	3	2	1	58	64	122
3 parcels	2	4	1	1			3	5	8
4+ parcels	2	4	2				4	4	8
Community Orgs who bought…									
1 parcel	2	1	2				4	1	5
2 parcels	4	2					4	2	6
Developers who bought…									
1 parcel	3	2	3		1		7	2	9
2 parcels	1	5	1				2	5	7
3 parcels	1	1					1	1	2
5 parcels			1				1		1
19 parcels		1						1	1
Trusts, estates, etc who bought…									
1 parcel		4	1				1	4	5
2 parcels		1						1	1
Other organizations who bought…									
1 parcel	3	1	2	1	1	3	6	5	11
2 parcels		4						4	4
3 parcels	1		1				2		2
Total Purchasers of >2 parcels…									
# of purchasers	6	10	5	1	0	0	11	11	22
# of parcels	20	52	23	3	0	0	43	55	98

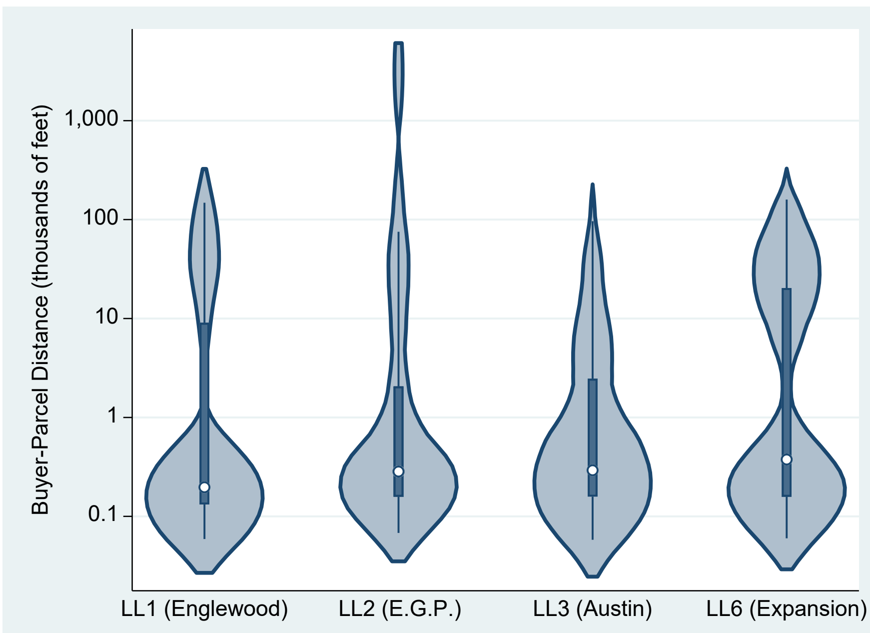
<u>Where are buyers located?</u> I start by plotting the distances between buyer mailing addresses and purchased parcels using lines, to visualize the extent to which outside purchasers may be accessing discounted parcels in the Large Lots program. A brief evaluation of Figure 10 reveals that buyers of Large Lots parcels reside not only in and around the program neighborhoods, but also in the predominantly wealthy northern neighborhoods and northern suburbs, in the southern suburbs, and occasionally outside the metropolitan area altogether.

Figure 10: Visualizing the distance between buyers and their parcels



I then I evaluate the distance between buyers and parcels over Large Lots program phase. This data is visualized in the violin plots of Figure 11 (Winter and Nichols 2008). These plots reveal that during the early, neighborhood-specific pilot phases, buyers were predominantly within 1000' of their parcels (a substitute measure for living on the same block) but that during the expansion phase the distribution becomes bimodal—while the largest body of buyers still had deeds sent to addresses less than 1000' from the purchased parcels a more substantial proportion of buyers used non-local addresses. Additional exploration of this apparent increase in outsider participation in LL6 is an important topic for further research, as it may indicate potential buyers with less legitimate claims of community membership find ways to participate in the program over time. If this is the case, program effectiveness can be expected to decline over multiple rounds.

Figure 11: Buyer-Parcel distance by Large Lots phase



I also evaluate buyer-parcel distance by buyer type. I find that individual buyers are predominantly supplying same-block addresses (<1000'), while non-individual entities are more widely distributed. Community organizations are bimodal but trend more local (although not as local as individuals), while developers are tightly distributed between 10 and 100 thousand feet away (in the metro area, but not in the same neighborhood). These findings are visualized in Figure 12.

Finally, I find that when dividing all sales by either count or total area into distance categories, 64% of sales were purchased by same-block buyers, and another 5-6% were purchased by buyers in the same neighborhood but on different blocks. The remaining 30% of buyers supplied mailing addresses outside the neighborhood (See Figure 13 and Table 10)

Figure 12: Buyer-Parcel distance by buyer type

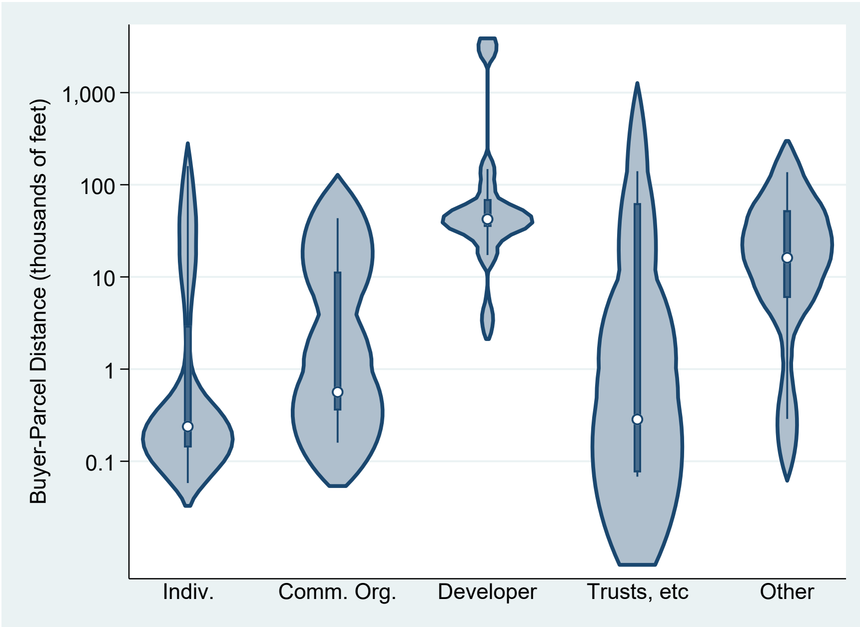


Figure 13: Distribution of sales by buyer mailing address

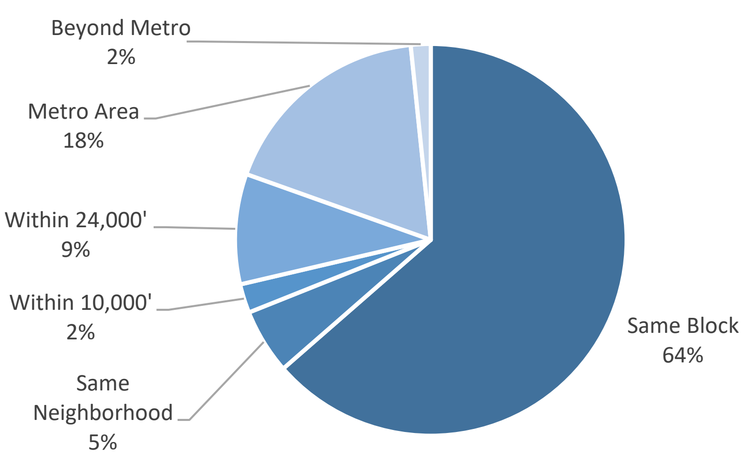


Table 10: Distribution of sales by buyer mailing address

	Parcel Count		Total Area	
	#	%	#	%
Same Block	426	63.6%	1,693,234	64.1%
Same Neighborhood	36	5.4%	149,326	5.6%
Within 10,000'	16	2.4%	75,799	2.9%
Within 24,000'	61	9.1%	223,477	8.5%
Metro Area	120	17.9%	459,272	17.4%
Beyond Metro	11	1.6%	42,092	1.6%

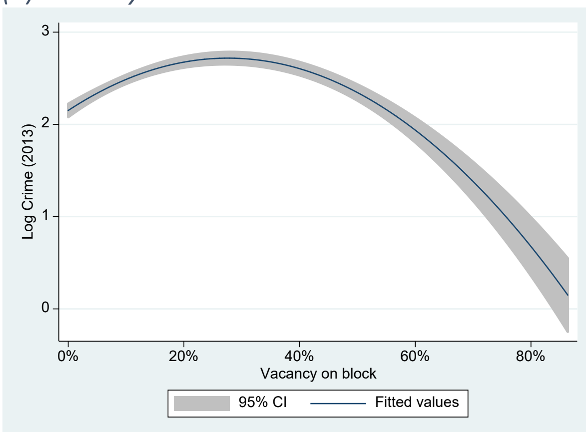
RQ2: Do Large Lots sales reduce crime?

I start with baseline findings measuring correlations between vacancy and crime in 2013, the year before the Large Lots program was first announced. I find that crime is strongly quadratically correlated with vacancy ($p<.0001$): crime increases as vacancy increases from 0 to about 30% of the parcels on a blockface, and then strongly decreases with further rises in vacancy (Figure 14[a], Table 11[a]). This supports theoretical arguments about crime such as the routine activity theory which suggests that crime is a factor of potential offenders, targets, and an absence of capable guardians (Cohen and Felson 1979): as vacancy increases above a certain threshold the presence of potential offenders and worthwhile targets may drop. Furthermore, these are *reported* crime statistics: it may be that on highly vacant blocks crime is committed but unreported. When broken out by neighborhood these findings are replicable in both Englewood and Austin but not East Garfield Park (Table 11[b]).

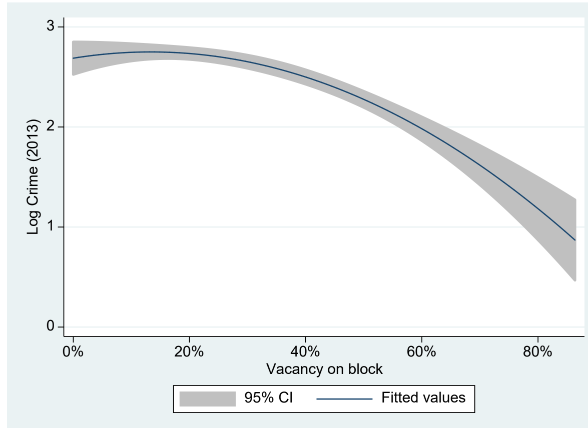
With this finding in mind, I also compare blocks with and without the city-owned vacant parcels that will become eligible for purchase in 2014 to 2013 crime. While blocks *without* available parcels retain a strong "up-then-down" correlation similar to the regression with all parcels blocks *with* available parcels primary see a *decrease* in crime over the possible vacancy range (Figure 14[b], Table 11[c]). Although this finding may appear counter-intuitive at first (why would crime decrease with vacancy on blocks where the city owned land?), this appears to be because parcels available for sale in the Large Lots program *are already on higher vacancy blocks*, which tend to have higher initial crime rates (Figure 14[c]). Indeed, a logistic regression finds that the presence of one or more Large Lots-eligible parcels is strongly correlated with the natural log of crime ($p<.01$, not pictured).

Figure 14: Pre-Large Lots vacancy and crime correlation

(a) Vacancy and crime: all blocks



(b) Vacancy and crime: blocks with available parcels



(c) Incidence of vacancy levels by presence of one or more parcels eligible for Large Lots sale

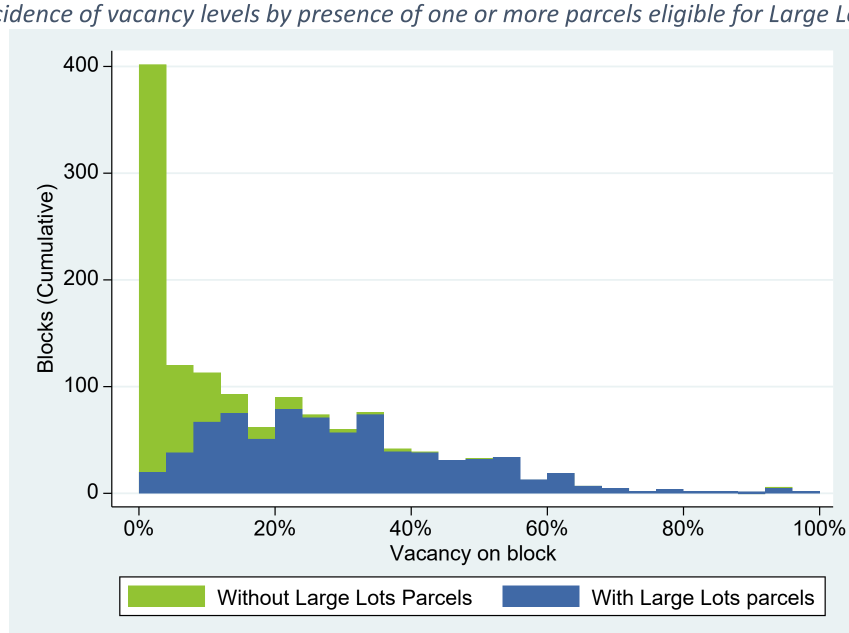


Table 11: Pre-Large Lots vacancy and crime correlation regression output

	Blocks	Vacancy Percentage			Crime Regression coefficients		
	n	min	mean	max	vacancy %	vacancy $\%^2$	R-squared
(a) All blocks	1332	0.0%	19.0%	100.0%	3.640***	-6.428***	0.142
(b) **Blocks, by neighborhood**							
Englewood	798	0.0%	26.7%	100.0%	1.968***	-4.513***	0.195
East Garfield Park	67	23.3%	27.6%	61.4%	-1.163	0.859	0.016
Austin	467	0.0%	4.5%	48.5%	9.306***	-18.14***	0.135
(c) **Blocks by presence of Large Lots eligible parcels**							
With eligible parcels	768	0.0%	29.8%	100.0%	0.997**	-3.619***	0.205
Without eligible parcels	564	0.0%	7.9%	94.1%	4.764***	-8.582***	0.065
*** p<0.01, ** p<0.05, * p<0.1							

With the pre-program linkage between vacancy and crime established, I turn to the panel regression. I analyze the full panel dataset with and without a 4 quarter lag to see whether changes in crime are best measured 1 year after new owners secure the deeds to their Large Lot parcels. I also perform the regression with increasingly conservative treatment variables that require buyers to be more and more local for a parcel to be considered treated. Because it is standard in the planning field I present OLS findings of the natural log of crime in Table 12. However, because of the characteristics of the outcome variable (namely, it is a count variable with a non-normal distribution), the negative binomial findings listed in Table 13 are preferred. These findings indicate an insignificant contemporaneous decrease in crime on blocks with Large Lots sales, and a statistically significant decline one year after the sale occurs (Table 13 Line A).

Table 12: Crime Panel – OLS Regression Output

OLS Outcomes	treated blocks	No lag†		4 quarter lag‡	
		Coefficient (β)	Interpretation§	Coefficient (β)	Interpretation§
a. All sales	346	-0.0235*	-2.32%	-0.0612***	-5.94%
b. Buyer in metro	295	-0.0297*	-2.93%	-0.0721***	-6.96%
c. Buyer < 24k'	249	-0.0474***	-4.63%	-0.1000***	-9.52%
d. Buyer < 10k'	233	-0.0465***	-4.54%	-0.0940***	-8.97%
e. Buyer in neighborhood	227	-0.0553***	-5.38%	-0.1030***	-9.79%
f. Buyer on block	213	-0.0420**	-4.11%	-0.0865***	-8.29%

† n =41,321, ‡ n = 37,038, *** p<0.01, ** p<0.05, * p<0.1
§ Interpretation of a logged outcome variable as % change = e^{β}-1

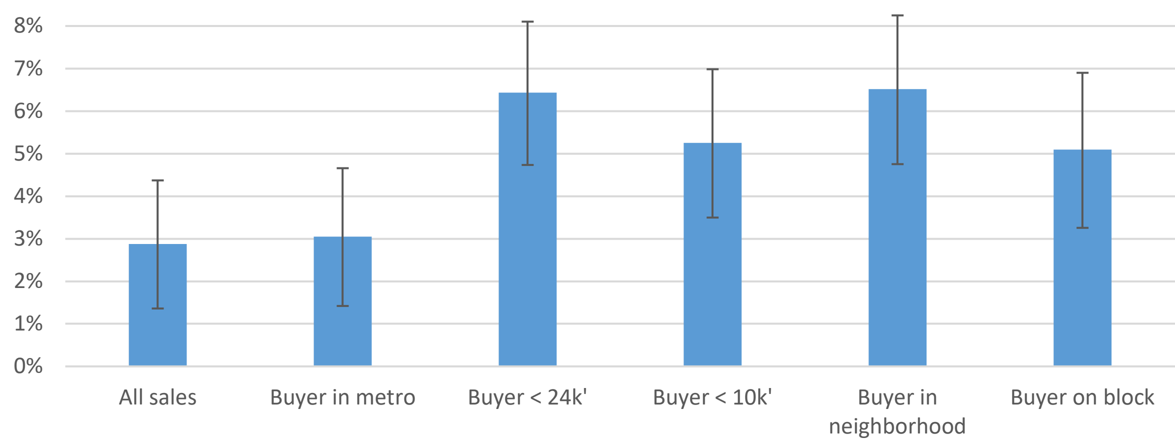
Table 13: Crime Panel – Negative Binomial Regression Output

Neg. Binomial outcomes	treated blocks	No lag†		4 quarter lag‡	
		Coefficient (β)	Interpretation§	Coefficient (β)	Interpretation§
a. All sales	346	-0.00382	-0.38%	-0.0292*	-2.88%
b. Buyer in metro	295	-0.00564	-0.56%	-0.0310*	-3.05%
c. Buyer < 24k'	249	-0.0326*	-3.21%	-0.0665***	-6.43%
d. Buyer < 10k'	233	-0.0209	-2.07%	-0.0540***	-5.26%
e. Buyer in neighborhood	227	-0.0381**	-3.74%	-0.0674***	-6.52%
f. Buyer on block	213	-0.0243	-2.40%	-0.0523***	-5.10%

† n =47,952 ‡ n = 42,624, *** p<0.01, ** p<0.05, * p<0.1
§ Interpretation of a logged outcome variable as % change = e^{β}-1

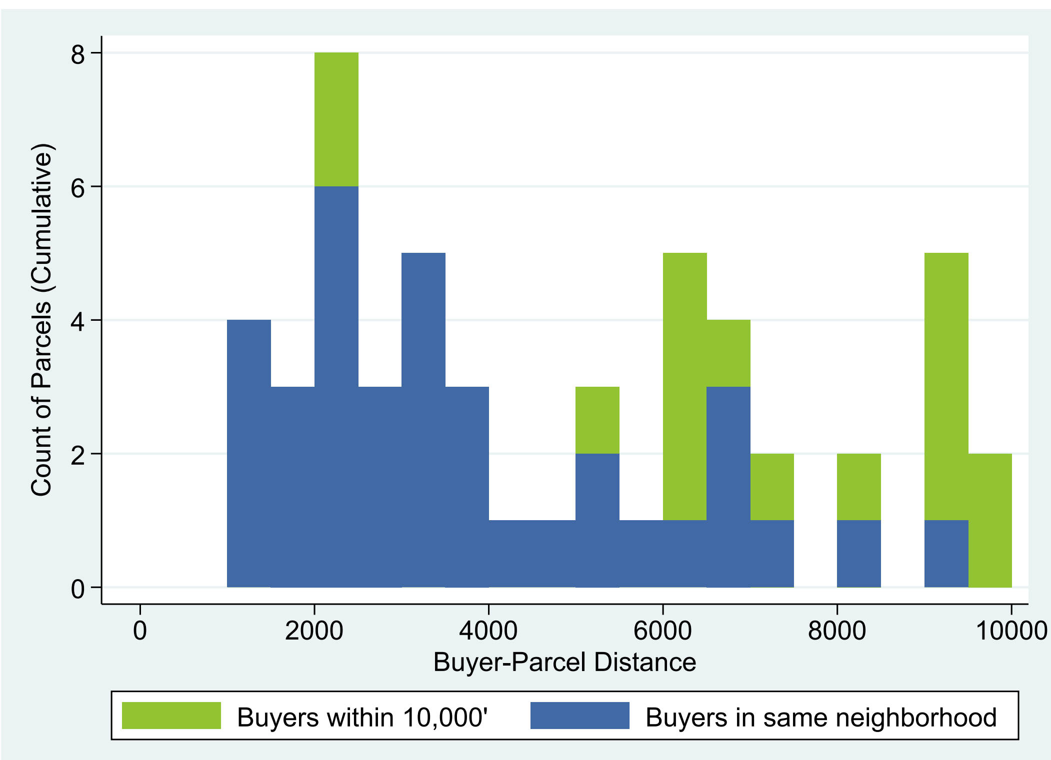
Furthermore, as described in Table 12, Table 13, and Figure 15, for both unlagged and lagged analyses reductions in crime resulting from Large Lots sales tend to increase as the treatment is measured more and more conservatively. This trend reverses before restricting buyers to the same block indicating that the largest crime reductions appear to be caused by sales to buyers whose mailing address is in the same neighborhood as the purchased parcel. The difference in crime reduction for buyers within 24,000 feet and buyers in the same neighborhood versus all large lots sales is statistically significant. It is also worth noting that buyers in the same neighborhood are more effective crime reducers than all buyers within 10,000’ of their parcels despite these categories having a relatively similar distribution of buyer-parcel distances (Figure 16).

Figure 15: Reductions in crime by buyer distance from parcel, negative binomial regressions



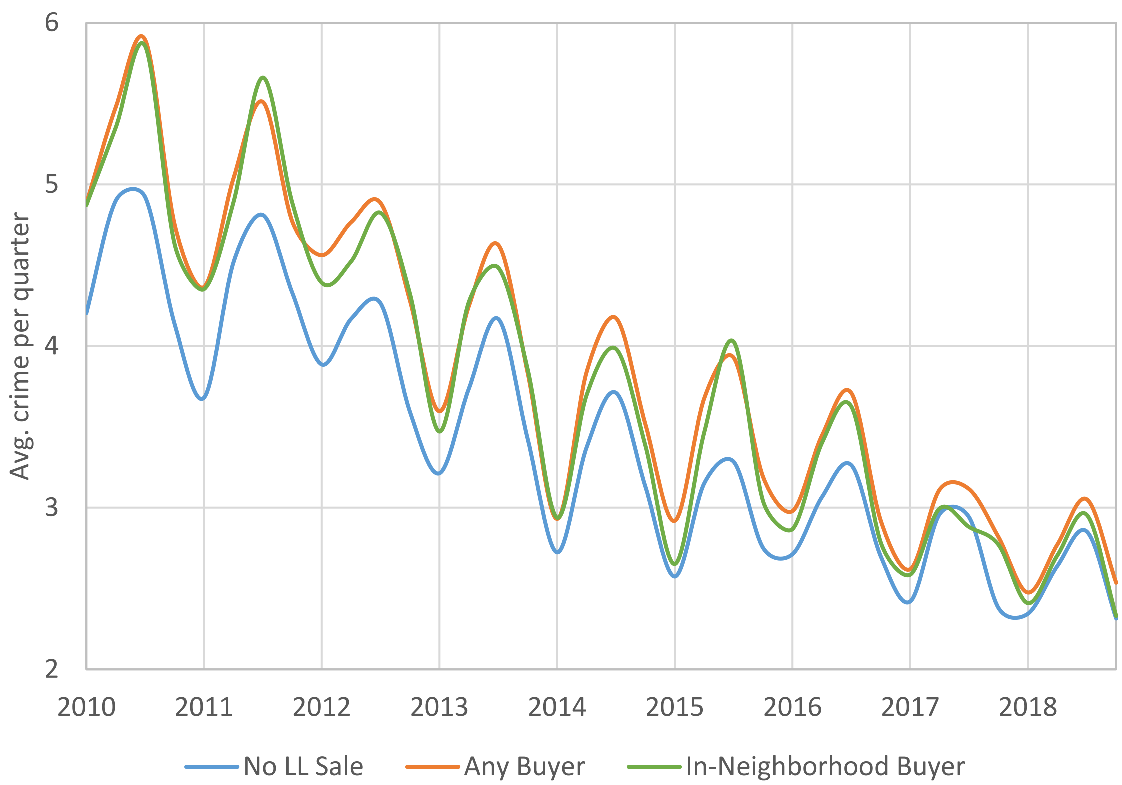
Columns represent the expected reduction in crime on blockfaces with Large Lots sales as a percentage of pre-sale crime levels on that blockface (e^{β}-1). Error bars represent the 95% confidence interval of each finding ($e^{\beta \pm SE}$-1). The inverse of each term is displayed, so positive findings are reductions in crime.

Figure 16: Count of parcels in “same neighborhood” and “less than 10,000 feet” distance classifications, by buyer-parcel distance



The findings of this regression are captured in a different way in Figure 17. Here, we see both the common seasonal cyclicality and overall downward trend in crime that Chicago is experiencing city-wide. We also see, in early 2010, that blocks on which Large Lots will eventually be purchased tend to have higher crime rates than blocks in the panel where sales will not occur. Over time, as lot sales occur, blocks with sales (red line) experience a greater decline in crime than blocks without sales (blue line), shrinking the gap by 2018. Furthermore, when looking only at average clime on blocks where at least one Large Lots buyer lives in the same neighborhood (green line), this decrease in crime is more extreme, and closing the gap entirely by Q4 2018.

Figure 17: Crime over time: No sales, Any Sales, and In-Neighborhood Sales



RQ3: How much wealth is generated?

I start by performing present dollar tabulations of total wealth generated through the program.[7] Using 2016 tax assessment data for all vacant parcels and calculating total value of all sold parcels using valuation per square foot (Model 1) I find that the total wealth generated in the 844 study parcels to be $27.3 million dollars. Using tax assessments for properties sold through the program only, valuations increase: using dollars per square foot (Model 2a) to $31.3 million, and using an average value per property to $34.8 million. This indicates that the parcels purchased through Large Lots are valued above average for vacant land. In Model 3, I use a weighted price per square foot average of private party vacant lot sales between 2014 and 2018. This model reports a total wealth generation of $83.1 million, between 2 and 3 times larger than the other models. These figures are described in Table 14.

[7] All values in 2016 dollars

Next, I dissect wealth generation by the two buyer typologies from Research Question 1: buyer type and buyer-parcel distance category. I find that 85% of wealth generated through the Large Lots program is accumulated by individual buyers while non-individual entities that appear to be property developers collected 7.5% of new wealth. I also find that 70% of wealth remained in the neighborhoods of the purchased parcels while the remaining 30% went to buyers whose deed mailing addresses are outside the neighborhood. More details can be found in Table 15 and Figure 18.

Table 14: Four models of Large Lots wealth generation

		Englewood	EGP	Austin	Total
	Number of LL parcels sold	583	179	82	844
	Cumulative area of LL parcels sold (sq. ft.)	2,442,797	630,018	368,330	3,441,145
Model 1 (vacant land assessment)					
	n (vacant parcels with assessment data)	5,395	800	638	6,833
	Mean value per sq. ft.	$ 6.95	$ 11.59	$ 8.18	$ 7.61
(1)	Wealth generated	$ 16,970,775	$ 7,300,336	$ 3,013,479	$ 27,284,590
Model 2 (LL parcels sold assessment)					
	n (LL sales with assessment data)	160	152	40	352
	Mean value per sq. ft.	$ 8.66	$ 11.57	$ 7.96	$ 9.84
	Sum value of sales with data	$ 6,998,640	$ 6,028,490	$ 1,487,860	$ 14,514,990
	Mean value per property	$ 43,742	$ 39,661	$ 37,197	$ 41,236
(2a)	Wealth generated (by sq. ft.)	$ 21,154,026	$ 7,287,623	$ 2,930,211	$ 31,371,860
(2b)	Wealth generated (by property)	$ 25,501,295	$ 7,099,340	$ 3,050,113	$ 34,802,987
Model 3 (Vacant land sales)					
	n (vacant land sales, 2014-2018)	822	146	143	1,111
	2016 sale price per sq. ft. (weighted average)	$ 25.54	$ 22.58	$ 17.69	$ 23.99
(3)	Wealth generated	$ 62,381,228	$ 14,223,116	$ 6,514,599	$ 83,118,943

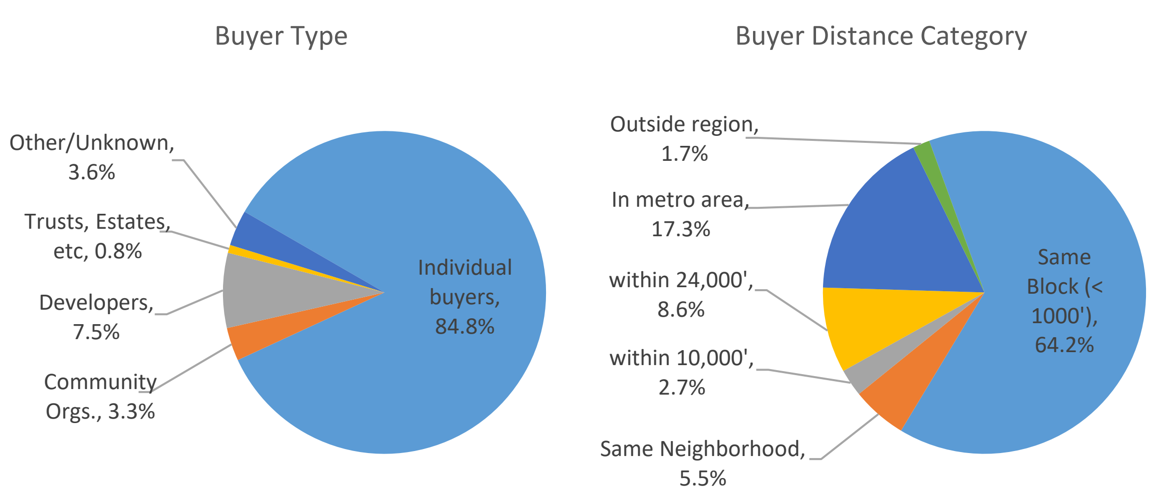
**all values in 2016 dollars*

Table 15: Wealth generation by model and buyer typologies

	Avg. %	Model 1	Model 2a	Model 2b	Model 3
Buyer Type					
Individual buyers	84.8%	$ 23,139,659	$ 26,522,686	$ 30,506,073	$ 70,002,845
Non-Indiv. Buyers	15.2%	$ 4,144,931	$ 4,849,175	$ 5,144,675	$ 13,116,098
- Community Org.	3.3%	$ 973,759	$ 1,145,658	$ 849,069	$ 3,105,217
- Developer	7.5%	$ 1,985,625	$ 2,359,340	$ 2,625,847	$ 6,514,759
- Trusts, Estates, etc.	0.8%	$ 195,005	$ 232,663	$ 348,788	$ 644,354
- Other/Unknown	3.6%	$ 990,542	$ 1,111,515	$ 1,320,972	$ 2,851,768
Buyer Distance Category					
Same Block (< 1000')	64.2%	$ 17,658,023	$ 20,223,758	$ 22,661,786	$ 53,220,338
Same Neighborhood	5.5%	$ 1,605,260	$ 1,759,435	$ 1,891,710	$ 4,422,418
Within 10,000'	2.7%	$ 796,878	$ 901,494	$ 849,569	$ 2,343,451
Within 24,000'	8.6%	$ 2,245,847	$ 2,623,507	$ 3,251,908	$ 7,101,004
In metro area	17.3%	$ 4,485,737	$ 5,316,554	$ 6,410,036	$ 14,687,907
Outside region	1.7%	$ 492,844	$ 547,112	$ 585,740	$ 1,343,825

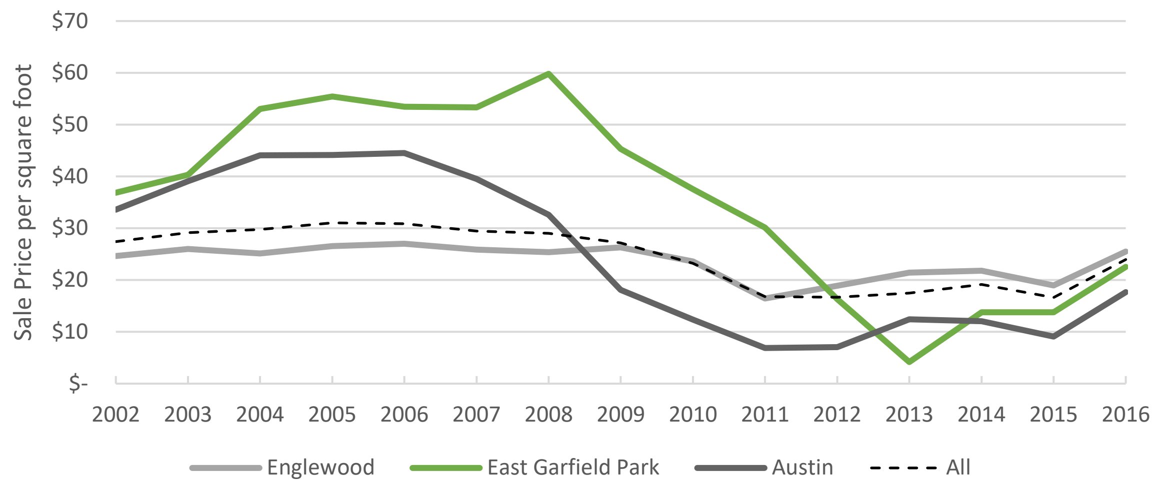
**all values in 2016 dollars*

Figure 18: Wealth generation by model and buyer typologies



Then, I look at trends of valuation over time to roughly project what the long term wealth generation impacts of the large lots program might be. These price trends, which reveal the impact of the recent economic cycle, are found in Figure 19. It is clear that although all three study regions have been increasing in land value since the recession, East Garfield Park and Austin's experience of the 2008 economic crash was much more severe than Englewood's.

Figure 19: Average vacant land sale prices over time

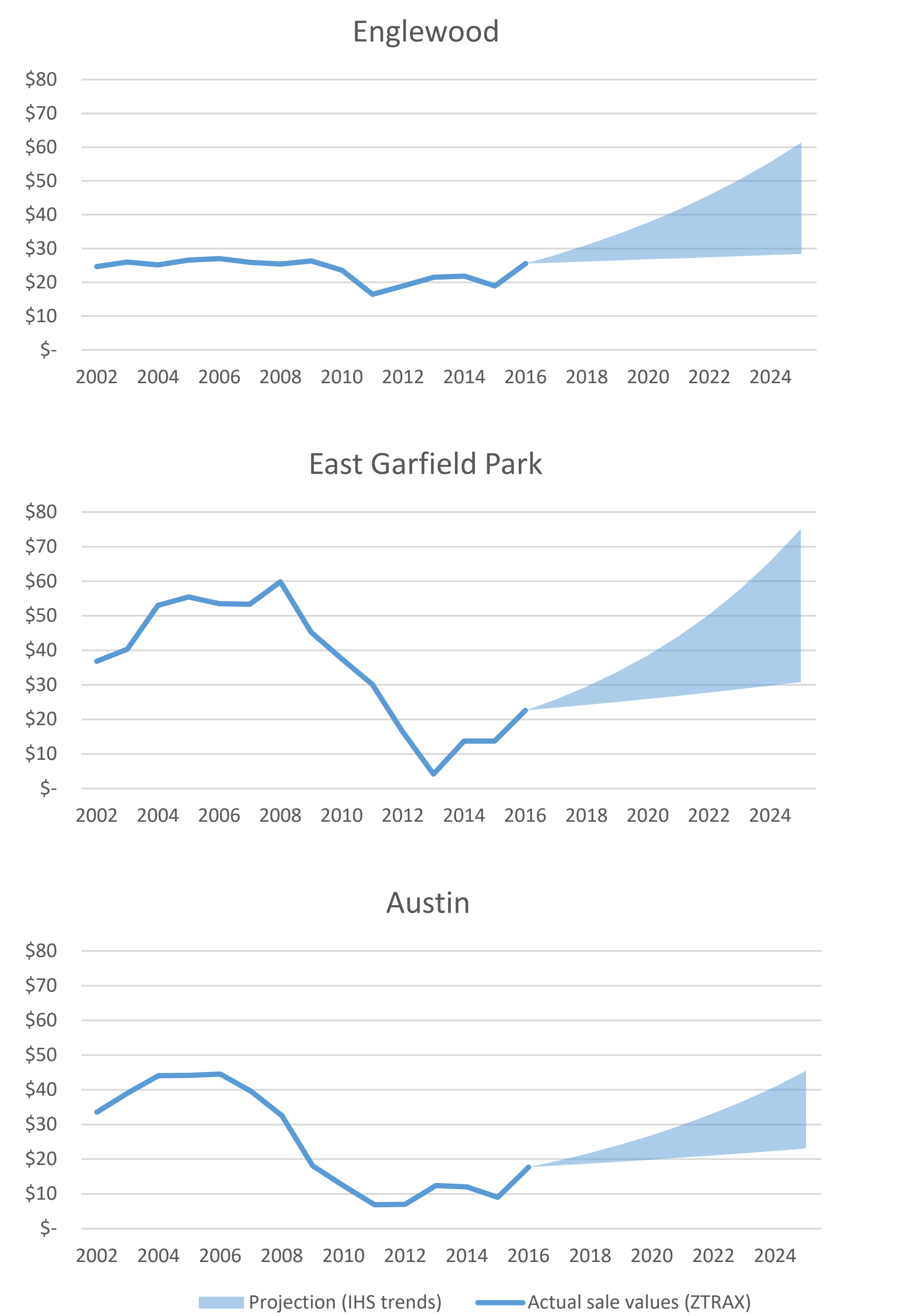


Finally, I use single family housing price trend data from DePaul University's Institute of Housing Studies to forecast total wealth generation in the years 2020 and 2025 based on actual sales of vacant lots in each neighborhood (Model 3). I annualize the growth rates experienced in each neighborhood since 2000, 2012, and 2017 and use the lowest and highest rates for each area as my upper and lower wealth projection bounds. Details are located in Table 16 while trends for each neighborhood are visualized in in Figure 20.

Table 16: Forecast of value generated by the Large Lots Program in 2020 and 2025

	Englewood	EGP	Austin	Total
Number of LL parcels sold	583	179	82	844
Cumulative area of LL parcels sold	2,442,797	630,018	368,330	3,441,145
Model 3 2016 valuation	$ 62,381,228	$ 14,223,116	$ 6,514,599	$ 83,118,943
Annualized Growth rates (HIS)				
Since 2000	1.2%	3.5%	3.0%	2.6%
Since 2012	8.0%	14.3%	9.8%	10.7%
Since 2017	10.2%	12.1%	11.1%	11.1%
Model 3, Lower bound growth				
2020 - sq. ft.	$ 26.78	$ 25.92	$ 19.91	$ 24.20
2025 - sq. ft.	$ 28.42	$ 30.81	$ 23.07	$ 27.43
2020 - total value	$ 65,416,507	$ 16,331,769	$ 7,331,825	$ 89,080,101
2025 - total value	$ 69,419,125	$ 19,412,495	$ 8,498,993	$ 97,330,613
Model 3, Upper bound growth				
2020 - sq. ft.	$ 37.71	$ 38.52	$ 26.91	$ 34.38
2025 - sq. ft.	$ 61.39	$ 75.11	$ 45.46	$ 60.66
2020 - total value	$ 92,118,925	$ 24,267,866	$ 9,910,890	$ 126,297,681
2025 - total value	$ 149,957,274	$ 47,323,586	$ 16,745,336	$ 214,026,195

Figure 20: Average vacant land sale prices per square foot over time, with forecasts



VI. Discussion & Policy Implications

This study evaluates the impacts of Chicago's Large Lots program from multiple angles including studying who is buying parcels, whether the program is linked to reductions in crime, and how much wealth the program has generated. In this section, I summarize key findings and discuss pertinent policy implications.

Most, but not all, buyers are individuals. The Large Lots program allows any same-block landowner in good standing with the city (resident or otherwise) to buy parcels through the Large Lots program, with a maximum of two parcels per buyer per program round. The intention behind these regulations is to strike a balance between (a) restricting program participation to members of the communities where parcels are available, (b) creating a broad base of potential buyers, and (c) limiting participation to those prepared to manage the financial and logistical burden of landownership. I find that 86% of buyers were individuals. It is believed that over 7% of buyers were professional developers, while the remaining 7% were community organizations, family trusts & estates, and other non-individual entities. Furthermore, the presence of developers increased dramatically in Englewood across program rounds while dropping to zero in both East Garfield Park and Austin. Community organizations (including churches, block groups, and other non-profits) were more active in the early rounds, buying 12 parcels across two neighborhoods, than the latter round when they purchased 5 parcels in Englewood only.

Considering that locally-elected aldermen have veto power over any specific application to buy parcels through the Large Lots program, in the absence of actual application data, it can be assumed that the quantity of approved purchases for any buyer class is limited by total interest for those types commonly viewed positively and by aldermanic prerogative and/or interest for those types that may be viewed as more problematic—namely, professional developers.[8] In this light, the zeroing out of developer participation in East Garfield Park and Austin in the 2016 expansion round, as these neighborhoods began to experience growth in property values, suggests that local aldermen may have become more practiced in the use of their veto power to protect the public interest across program rounds. In Englewood, where land values remained low and flat, developer activity increased; this may be because of aldermanic oversight or because professional developer participation in the Large Lots program was viewed as a positive outcome. All things considered, it seems that close partnership with and careful participation of local elected officials in the program is an important program component.

Most, but not all, buyers are local: I found that 30% of buyers had deeds mailed to addresses not within the same neighborhood as the property being purchased. These findings align to a degree with the goals of the Large Lots program which include enabling those who maintain ties to these neighborhoods but no longer live in them to purchase land and accumulate wealth through additional landownership. However, many deeds were mailed to the wealthier north-side and northern suburbs as well as to

[8] This begs a larger and difficult question about what role professional developers should play in the early stages of distressed neighborhood reinvestment and whether cities should facilitate developers' access to cheap land through programs such as Large Lots. On one hand, developers tend to be profit-motivated and thus may have an extractive effect on communities as they change. Additionally, the simultaneous actions of many developers may eventually lead to gentrification and displacement. On the other hand, the revitalization of distressed neighborhoods will surely require substantial developer activity and private capital investment; perhaps discounting land for early capital risk-takers is a helpful way to jumpstart growth. See Kelly (2016), especially p. 1026.

addresses in downtown Chicago. This suggests that some buyers, even if they have close connections to these communities, may already have accumulated substantial wealth.

Additionally, individuals are much more likely to be highly proximal to the properties they are purchasing than organizations of any type but developers and other/uncategorized organizations supply addresses especially far from purchased parcels. With these findings in mind, the city may consider more tightly defining their desired buyer pool (e.g. limiting to those with true ties to the community and/or using income or wealth ceilings); this could be achieved by either tweaking program rules or giving aldermen more information with which to approve or veto certain sales.

Furthermore, considering that LLCs present distinct tax and liability advantages for property investors, the low frequency of uncategorized/other organizations (3.7% of total) and the higher distribution of parcel-buyer distance for this category indicate that few truly local buyers—one population the city hopes will build wealth through this program—are taking advantage of the LLC model to enhance financial outcomes. Incorporating training about the advantages of placing a side-yard lot in an LLC and providing discounted legal services could support the Large Lot's wealth-building goals.

<u>Some buyers are breaking the rules</u>: Using buyer names and mailing addresses to identify sets of parcels purchased by a single buyer, I found that over 13% of parcels were purchased by buyers who appear to have purchased more than two parcels in a single program round in violation of Large Lots policy. This includes one situation in which an apparent developer purchased 19 Englewood properties in the 2016 Expansion round. It is worth improving the way in which purchasers are identified and tracked to prevent this from occurring in the future and/or transparently clarifying the situations in which exceptions to the two parcels per buyer policy may be made.

<u>The large lots program reduces crime</u>: I find that the blockfaces on which Large Lots parcel sales reside experience an average 2.88% reduction in crime beginning one year after the sale occurs. Because of the methods employed in this study this decrease can be considered causal. This is an important and positive finding for the Large Lots program in that it confirms the existence of a positive benefit of the program that has been hitherto-unproven. Furthermore, when considering only those blocks with Large Lots purchased by buyers within 24,000 feet (roughly 4.5 miles) or within the same neighborhood as the purchased parcel, the average crime reduction is significantly greater (6.5%). This finding further motivates the suggestion that the city consider placing additional restrictions on the buyer pool as it appears that lower parcel-buyer-distances are correlated with the greatest reductions in crime.

The regressions that underpin this finding are agnostic to what owners have done with their new land and when any improvements were made. It seems likely that some owners have done more to activate their parcels than others and that this difference could have an impact on crime rates. Ideally, future research will link crime not only to Large Lots sales but data on parcel-specific activation—the sort of data collected by Gobster et al. (2018).

<u>The Large Lots program has generated substantial wealth</u>: I find that the 844 studied Large Lots transactions have generated between $27.3 and $83.1 million of wealth for program buyers ($32,000 to $98,000 on average per parcel). The dramatic range is due to the large difference between tax assessments (the lower bound) and the prices of actual non-zero-dollar purchases (the upper bound). If current land valuation trends in these neighborhoods continue, the upper bounds of the wealth generation forecast (using data from actual sales) suggest total generated wealth of $89M-$126M in

2020 and $97M-$214M in 2025. Applying buyer-parcel distance classifications to wealth generation suggest that 70% of wealth generation is remaining in the neighborhoods where sales are occurring, while the remaining 30% has been captured by those who live elsewhere. Additionally, 7.5% of generated wealth appears to have been captured by professional developers. These findings further support the consideration of additional limitations on the Large Lots buyer pool.

VII. Conclusion

Side-yard and other vacant land disposition programs that seek to widely distribute ownership of vacant land have many goals and multiple constituencies. Public-to-private property transfers lessen the municipal burden of managing thousands of discrete vacant parcels, decrease the city's liability, and directly increase property tax revenue, changes that serve the interest of many local government stakeholders. For neighborhoods, these transfers have the potential to increase local agency over land use decisions, increase vibrancy, foster community, decrease crime, and transfer wealth back into communities that been financially hard-hit by decades of anti-inner-city urban policy. Moreover, increased vibrancy, higher levels of private ownership, and lower crime rates may together improve market conditions and lead to further development which should in turn raise home values, generate additional wealth, and create a cycle of rebirth and growth. With so many parties standing to benefit it is no wonder why Chicago's Large Lots program is part of a rapidly growing national trend.

However, most of these linkages are not given. Do the new owners of vacant parcels pay their property taxes, and are those payments a financial burden? Does private, rather than public, management of vacant parcels decrease blight and produce more activated streets and neighborhoods? Is this activation actually tied to local reductions in crime or other neighborhood health benefits? Who is taking advantage of these wealth generation opportunities and are they the people who program implementers believe should benefit? How long do new owners hold on to program parcels and what do they do with them? And, if programs succeed at delivering market signals that generate additional development, does that growth eventually lead to gentrification and displacement that deprives the very population the program had hoped to benefit of their own reborn and vibrant communities?

Some scholars have begun to address these questions (Dewar 2006, 2015; Ganning and Tighe 2015; Mallach 2017, 2018; Gobster et al. 2018) but most remain unanswered. At the same time, others have raised legitimate concerns about the intentions and outcomes of programs that encourage private development in neighborhoods whose experience of hypervacancy is tied closely to redlining, urban renewal, subprime lending, and other policy choices with racist overtones (Safransky 2014; Rosenman and Walker 2016; Kelly 2016). This scholarship reinforces the need for evaluative research.

This study begins to address this gap in scholarship, supply evidence as to whether these programs are successful, and identify best practices. I find that the wealth generation impacts of the Large Lots program have been substantial and that most program participants have been local individuals. I also draw a direct causal link between the program and reductions in crime, and find that the highest crime reduction is measured when property transfers are limited to less-distant owners. However, I also find that the program has seen active participation by professional developers, that many (potentially up to 31%) of new owners may not actually live in the communities in which they are buying, and that over 13% of parcels were purchased by buyers who violated the 2-parcels-per-buyer-per-round rule. These findings hopefully provide evidence to the founders and leaders of Chicago's Large Lots program as they consider how to iterate and improve the program and encourage further study of vacant land disposition programs around the country.

Hypervacancy in urban neighborhoods is both a problem and an opportunity. Because urban land vacancy axiomatically occurs where property values are lowest, how these problems are solved and opportunities captured will have real impacts on the some of the United States' most economically and politically disenfranchised communities. The long term outcomes of programs designed to address hypervacancy are sure to be interpreted—with the memory of urban renewal and decades of disinvestment etched into the collective memory of poor urban communities and professional urban planners alike—as America's next attempt to "fix" urban poverty. With this in mind, it is wise to start studying the efficacy of programs like Chicago's Large Lots program now, and linking future policy iterations to evaluations of current programs.

Bibliography

596 Acres. 2018. “202 City-Owned Lots Sold to Housing Developers for $1 Each since 2014, Mapped at Onedollarlots.Org.” 596 Acres. Accessed April 1, 2019. http://596acres.org/one-dollar-lots/.

Arnio, Ashley N, Eric P Baumer, and Kevin T Wolff. 2012. “The Contemporary Foreclosure Crisis and US Crime Rates.” *Social Science Research* 41 (6): 1598–1614. https://doi.org/10.1016/j.ssresearch.2012.05.013.

Beauregard, Robert A. 2001. “Federal Policy and Postwar Urban Decline: A Case of Government Complicity?” *Housing Policy Debate* 12 (1): 129–51. https://doi.org/10.1080/10511482.2001.9521401.

Butler, Elizabeth. 2016. “Second Chances for the Second City’s Vacant Properties: An Analysis of Chicago’s Policy Approaches to Vacancy, Abandoment, & Blight.” *Kent Law Review* 91 (1): 233–65. https://scholarship.kentlaw.iit.edu/cklawreview/vol91/iss1/10.

Campbell, John, Stefano Giglio, and Parag Pathak. 2011. “Forced Sales and House Prices.” *American Economic Review* 101 (5): 2108–31.

Chicago Metropolitan Agency for Planning. 2017. “Land Use Inventory for Northeast Illinois, 2013.” CMAP Data Hub. Accessed April 3, 2019. https://datahub.cmap.illinois.gov/dataset/land-use-inventory-for-northeast-illinois-2013.

ChiHackNight. 2017. “Assessing Chicago’s Large Lot Program [Webcast].” Chicago: Scrappers Film Group. https://www.youtube.com/watch?v=6-v6odxrTXs.

City of Buffalo. n.d. “Urban Homestead Program.” City of Buffalo Website. Accessed April 1, 2019. https://www.buffalony.gov/306/Urban-Homestead-Program.

City of Chicago. n.d. “Adjacent Neighbors Land Acquisition Program.” City of Chicago Website. Accessed April 1, 2019. https://www.chicago.gov/city/en/depts/dcd/supp_info/adjacent_neighborslandacquisitionprogramanlap.html.

———. 2019. “Various Datasets.” Chicago Data Portal. Accessed April 3, 2019. https://data.cityofchicago.org/.

City of Columbus. 2016. “Application to Purchase a Vacant Land Bank Property as a Side Yard or Permanent Community Garden.” Accessed April 1, 2019. http://www.epa.gov/lead/pubs/leadpdfe.pdf.

Cohen, L.E., and M. Felson. 1979. “Social Change and Crime Rate Trends: A Routine Activity Approach.” *American Sociological Review* 44: 588–608.

Cook County Assessor. n.d. “Calculating an Estimated Residential Property Tax Bill.” Cook County

Assessor’s Office. Accessed April 3, 2019. http://www.cookcountyassessor.com/Resources/Residential-Tax-Bill.aspx.

Cook County Clerk. 2016. “Cook County Parcels 2014.” Cook County Open Data Catalog. Accessed April 3, 2019. https://datacatalog.cookcountyil.gov/GIS-Maps/ccgisdata-Parcel-2014/2m9h-cq6j.

Cui, Lin, and Randall Walsh. 2015. “Foreclosure, Vacancy and Crime.” *Journal of Urban Economics* 87: 72–84. https://doi.org/10.1016/j.jue.2015.01.001.

Dewar, Margaret. 2006. “Selling Tax-Reverted Land: Lessons from Cleveland and Detroit.” *Journal of the American Planning Association* 72 (2): 167–80. https://doi.org/10.1080/01944360608976737.

———. 2015. “Reuse of Abandoned Property in Detroit and Flint: Impacts of Different Types of Sales.” *Journal of Planning Education and Research* 35 (3): 347–68. https://doi.org/10.1177/0739456X15589815.

Dickhut, Kathy. 2018. “Personal Interview,” August 6, 2018.

Ellen, Ingrid Gould, Johanna Lacoe, and Claudia Ayanna Sharygin. 2013. “Do Foreclosures Cause Crime?” *Journal of Urban Economics* 74: 59–70. https://doi.org/10.1016/j.jue.2012.09.003.

Fahey, Mark. 2015. “Buffalo’s $1 Homes Aren’t as Cheap as They Seem.” *CNN Money*, March 3, 2015. Accessed April 1, 2019. https://money.cnn.com/2015/03/03/real_estate/buffalo-1-homes/index.html.

Frazier, Amy E., Sharmistha Bagchi-Sen, and Jason Knight. 2013. “The Spatio-Temporal Impacts of Demolition Land Use Policy and Crime in a Shrinking City.” *Applied Geography* 41: 55–64. https://doi.org/10.1016/j.apgeog.2013.02.014.

Gallagher, John. 2018. “Detroit Land Bank Notches Big Successes despite Some Missteps.” *Detroit Free Press*, September 23, 2018. Accessed April 1, 2019. https://www.freep.com/story/money/business/john-gallagher/2018/09/23/detroit-land-bank/1358663002/.

Ganning, Joanna P., and J. Rosie Tighe. 2015. “Assessing the Feasibility of Side Yard Programs as a Solution to Land Vacancy in U.S. Shrinking Cities.” *Urban Affairs Review* 51 (5): 708–25. https://doi.org/10.1177/1078087414555435.

Gobster, Paul H, William P. Stewart, Alessandro Rigolon, Carena J. van Riper, and Douglas A. Williams. 2018. “Visual Resource Stewardship at the Neighborhood Scale: Methods for Assessing a Vacant Land Reuse Program.” In *Visual Resource Stewardship Conference Proceedings*, edited by Paul H Gobster and Richard C. Smardon. https://doi.org/10.2737/NRS-GTR-P-183-VCS-4.

Goldstein, James, Michael Jensen, and Edward Reiskin. 2001. “Urban Vacant Land Redevelopment: Challenges and Progress.”

Großmann, Katrin, Marco Bontje, Annegret Haase, and Vlad Mykhnenko. 2013. “Shrinking Cities: Notes for the Further Research Agenda.” *Cities* 35: 221–25. https://doi.org/10.1016/j.cities.2013.07.007.

Hackworth, Jason. 2014. “The Limits to Market-Based Strategies for Addressing Land Abandonment in Shrinking American Cities.” *Progress in Planning* 90: 1–37. https://doi.org/10.1016/j.progress.2013.03.004.

Hummel, Daniel. 2015. “Right-Sizing Cities In The United States: Defining Its Strategies.” *Journal of Urban*

Affairs 37 (4). https://doi.org/10.1111/juaf.12150.

Hunter, Brandon. 2018. "Detroit Land Bank Buyback Program Gives Residents Second Chance at Home Ownership." *The Michigan Chronicle*, July 17, 2018. Accessed April 1, 2019. https://michiganchronicle.com/2018/07/17/detroit-land-bank-buyback-program-gives-residents-second-chance-at-home-ownership/.

Immergluck, Dan, and Geoff Smith. 2006. "The External Costs of Foreclosure: The Impact of Single-family Mortgage Foreclosures on Property Values." *Housing Policy Debate* 17 (1): 57–59. https://doi.org/10.1080/10511482.2006.9521561.

Institute for Housing Studies. 2018a. "2018-Q2 Housing Price Index Data." DePaul University. https://www.housingstudies.org/media/filer_public/2018/10/11/ihs_2018_q2_price_index_data.xlsx.

———. 2018b. "Cook County House Price Index: Second Quarter 2018." DePaul University. Accessed April 1, 2019. https://www.housingstudies.org/releases/cook-county-house-price-index-second-quarter-2018/.

Jackson, Kenneth. 1985. *Crabgrass Frontier*. Oxford: Oxford University Press.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York: Random House.

Kelly, James J. 2016. "Affirmatively Furthering Neighborhood Choice: Vacant Property Strategies and Fair Housing." *University of Memphis Law Review* 46 (4).

Koski, Andrew. 2015. "Lots to Love." South Side Weekly. Accessed March 25, 2018. https://southsideweekly.com/lots-to-love/.

LISC Chicago, and DataMade. n.d. "About Large Lots." Large Lots Website. Accessed April 3, 2019. https://largelots.org/about/.

Louis Land Reutilization Authority, St. n.d. "Mow To Own Program." City of St. Louis, MO Website. Accessed April 1, 2019. https://www.stlouis-mo.gov/government/departments/sldc/real-estate/mow-to-own-program.cfm.

Lutton, Linda. 2019. "The Middle Class Is Shrinking Everywhere — In Chicago It's Almost Gone." *WBEZ*, February 18, 2019. https://www.wbez.org/shows/wbez-news/the-middle-class-is-shrinking-everywhere-in-chicago-its-almost-gone/e63cb407-5d1e-41b1-9124-a717d4fb1b0b.

Mallach, Alan. 2011. "Demolition and Preservation in Shrinking US Industrial Cities." *Building Research and Information* 39 (4): 380–94. https://doi.org/10.1080/09613218.2011.573743.

———. 2017. "What We Talk about When We Talk about Shrinking Cities: The Ambiguity of Discourse and Policy Response in the United States." *Cities* 69: 109–15. https://doi.org/10.1016/j.cities.2017.01.008.

———. 2018. "The Empty House Next Door." https://www.lincolninst.edu/sites/default/files/pubfiles/empty-house-next-door-full.pdf.

Nassauer, Joan Iverson, and Julia Raskin. 2014. "Urban Vacancy and Land Use Legacies: A Frontier for Urban Ecological Research, Design, and Planning." *Landscape and Urban Planning* 125: 245–53. https://doi.org/10.1016/j.landurbplan.2013.10.008.

National Vacant Properties Campaign. 2005. "Vacant Properties: The True Costs to Communities." Washington, DC. https://www.hudexchange.info/resources/documents/VacantPropertiesTrueCosttoCommunities.pdf.

NORA. n.d. "Lot Next Door 3.0 Program." New Orleans Redevelopment Authority. Accessed April 1, 2019. http://www.noraworks.org/lotnextdoor.

Pagano, Michael, and Ann Bowman. 2000. "Vacant Land in Cities: An Urban Resource." https://www.brookings.edu/research/vacant-land-in-cities-an-urban-resource/.

Rosenman, Emily, and Samuel Walker. 2016. "Tearing down the City to Save It? 'Back-Door Regionalism' and the Demolition Coalition in Cleveland, Ohio." *Environment and Planning A* 48 (2). https://doi.org/10.1177/0308518X15609741.

Safransky, Sara. 2014. "Greening the Urban Frontier: Race, Property, and Resettlement in Detroit." *Geoforum* 56. https://doi.org/10.1016/j.geoforum.2014.06.003.

Schilling, Joseph. 2002. "The Revitalization of Vacant Properties: Where Broken Windows Meet Smart Growth." *International City/County Management Association*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.188.9674&rep=rep1&type=pdf.

Schilling, Joseph, and Jonathan Logan. 2008. "Greening the Rust Belt." *Journal of the American Planning Association* 74(4): 451–66. https://doi.org/10.1080/01944360802354956.

Spader, Jonathan, Jenny Schuetz, and Alvaro Cortes. 2016. "Fewer Vacants, Fewer Crimes? Impacts of Neighborhood Revitalization Policies on Crime." *Regional Science and Urban Economics* 60: 73–84. https://doi.org/10.1016/j.regsciurbeco.2016.07.005.

Spielman, Fran. 2016. "Emanuel Selling 4,000 Vacant Lots in 33 Neighborhoods for $1 Each." *Chicago Sun Times*, November 29, 2016. https://chicago.suntimes.com/politics/emanuel-selling-4000-vacant-lots-for-1/.

Tappendorf, J A, and B O Denzin. 2011. "Turning Vacant Properties into Community Assets through Land Banking." *Urban Lawyer* 43 (3): 801–12. http://www.scopus.com/inward/record.url?eid=2-s2.0-84859823103&partnerID=40&md5=72ce51a65eb0e994d91b2914f24c90b7.

Twinam, Tate. 2017. "Danger Zone: Land Use and the Geography of Neighborhood Crime." *Journal of Urban Economics* 100: 104–19. https://doi.org/10.1016/j.jue.2017.05.006.

US Census Bureau, and Erik Steiner. 2018. "United States Historical City Populations, 1790-2010." Center for Spatial and Textual Analysis, Stanford University. https://github.com/cestastanford/historical-us-city-populations.

Whitaker, Stephan, and Thomas J Fitzpatrick IV. 2013. "Deconstructing Distressed-Property Spillovers: The Effects of Vacant , Tax-Delinquent , and Foreclosed Properties in Housing Submarkets." *Journal of Housing Economics* 22 (2): 79–91. https://doi.org/10.1016/j.jhe.2013.04.001.

Wilson, J.Q., and G.L. Kelling. 1982. "The Police and Neighborhood Safety: Broken Windows." *The Atlantic*, 1982.

Winter, Nick, and Austin Nichols. 2008. "VIOPLOT: Stata Module to Produce Violin Plots with Curent Graphics." *Statistical Software Components S456902*. Boston College Department of Economics.

https://ideas.repec.org/c/boc/bocode/s456902.html.

Zillow. n.d. “ZTRAX FAQs.” Zillow Research. Accessed April 1, 2019. https://www.zillow.com/research/ztrax/ztrax-faqs/.

———. 2019. “ZTRAX: Zillow Transaction and Assessor Dataset, 2018-Q4.” *Zillow Group, Inc*. http://www.zillow.com/ztrax/.

Appendix A: Stata log of ZTRAX Interpretation program

```
-----------------------------------------------------------------------------------------------
      name:  ZTRAX interpretation, complete run
       log:  Z:\Chicago\DataExtracts\log_31Mar2019_203142.smcl
  log type:  smcl
 opened on:  31 Mar 2019, 20:31:42
. * display local variables already set, for the record
. display "`loc_subfn'"
Z:\Chicago\MATT_CODE\import_sub_function.do
. display "`loc_layout'"
Z:\Chicago\17_new\layout.xlsx
. display "`loc_data'"
Z:\Chicago\17_new
. display "`loc_output'"
Z:\Chicago\DataExtracts
. display "`using_fips'"
17031
. }

.
. **program definitions.
. local output_name = "ZTrans" + "`using_fips'"

. local output_name2 = "ZAsmt" + "`using_fips'"

. local output_name3 = "ZCombined" + "`using_fips'"

. set more off

.
. cd `loc_output'
Z:\Chicago\DataExtracts

.
. local run_this = 0

. ************************
. ***** TEST FUNCTION *****
.
. if `run_this' == 1 {
. *this tests the sub function on a very small table.
.
. local table = "BorrowerMailAddress"
. local keepvars ""TransId BrwrMailCareOfName BrwrMailHouseNumber BrwrMailHouse
> NumberExt""
. do `loc_subfn' `table' "ZTrans" `loc_layout' `loc_data' `keepvars'
. }

.
.
. local run_this = 1

. ********************************
. **** Import ZAsmt Main table ****
.
. if `run_this' == 1 {
. local table = "Main"
. local keepvars ""RowID ImportParcelID FIPS AssessorParcelNumber UnformattedAs
> sessorParcelNumber PropertyZoningDescription TaxAmount TaxYear LotSizeSquareF
> eet""
.
. di c(current_time)
20:31:42
```

```
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZAsmt" `loc_layout' `loc_data' `keepvars'
. di c(current_time)
20:52:10
.
. keep if FIPS == `using_fips'
(4,012,323 observations deleted)
.
. save `output_name2'.dta, replace
file ZAsmt17031.dta saved
. }

.
.
. local run_this = 1

. *********************************
. **** Import ZAsmt Value table ****
.
. if `run_this' == 1 {
. local table = "Value"
. local keepvars ""RowID LandAssessedValue ImprovementAssessedValue TotalAssess
> edValue AssessmentYear LandMarketValue ImprovementMarketValue TotalMarketValu
> e MarketValueYear LandAppraisalValue ImprovementAppraisalValue TotalAppraisal
> Value AppraisalValueYear""
.
. di c(current_time)
20:52:14
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZAsmt" `loc_layout' `loc_data' `keepvars'
. di c(current_time)
20:53:01
.
. **merge with existing dataset
. merge 1:1 RowID using `output_name2'

    Result                           # of obs.
    -----------------------------------------
    not matched                     4,012,323
        from master                 4,012,323  (_merge==1)
        from using                          0  (_merge==2)

    matched                         1,859,435  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : record in this file but not in master file
. *          (presumably because it is not in correct FIPS code)
. *    _merge==2 : record in master file but there are no details to add in thi
> s file
. *    _merge==3 : record in master file AND new details are in this file
.
.
. **remove records that don't belong
. drop if _merge==1
(4,012,323 observations deleted)
. drop _merge
.
. **drop fields with no values
. missings dropvars, force

Checking missings in RowID LandAssessedValue ImprovementAssessedValue
    TotalAssessedValue AssessmentYear LandMarketValue ImprovementMarketValue
    TotalMarketValue MarketValueYear LandAppraisalValue
    ImprovementAppraisalValue TotalAppraisalValue AppraisalValueYear
    ImportParcelID FIPS AssessorParcelNumber UnformattedAssessorParcelNumber
    PropertyZoningDescription TaxAmount TaxYear LotSizeSquareFeet:
```

```
1859435 observations with missing values

note: LandMarketValue ImprovementMarketValue LandAppraisalValue
ImprovementAppraisalValue TotalAppraisalValue AppraisalValueYear
PropertyZoningDescription dropped
.
. ** fix parcelID
. gen  pin_length = length(UnformattedAssessorParcelNumber)
. tab pin_length

 pin_length |      Freq.     Percent        Cum.
------------+-----------------------------------
         10 |  1,391,294       74.82       74.82
         14 |    468,141       25.18      100.00
------------+-----------------------------------
      Total |  1,859,435      100.00
. keep if pin_length == 10 | pin_length == 14
(0 observations deleted)
. gen pin14 = UnformattedAssessorParcelNumber if pin_length == 14
(1,391,294 missing values generated)
. replace pin14 = UnformattedAssessorParcelNumber + "0000" if pin_length == 10
(1,391,294 real changes made)
. gen  pin_length2 = length(pin14)
. assert pin_length2 == 14
. drop AssessorParcelNumber UnformattedAssessorParcelNumber pin_length pin_leng
> th2
.
. drop FIPS TaxAmount TaxYear
. order RowID ImportParcelID pin14 LotSizeSquareFeet AssessmentYear LandAssesse
> dValue ImprovementAssessedValue TotalAssessedValue MarketValueYear TotalMarke
> tValue
.
. **and export the master table...
. save `output_name2', replace
file ZAsmt17031.dta saved
. }

.
.
.
. local run_this = 1

. *******************************
. ** Import ZTrans PropertyInfo **
.
. if `run_this' == 1 {
. local table = "PropertyInfo"
. local keepvars ""TransId FIPS AssessorParcelNumber UnformattedAssessorParcelN
> umber ImportParcelID AssessmentRecordMatchFlag""
.
. di c(current_time)
20:53:51
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZTrans" `loc_layout' `loc_data' `keepvars'
. di c(current_time)
21:37:40
.
. **keeping only records in specified FIPS code...
. keep if FIPS == `using_fips'
(11,765,700 observations deleted)
.
. save `output_name'.dta, replace
file ZTrans17031.dta saved
.
. ** create a list of transactions in study area
. keep TransId
```

```
. duplicates drop

Duplicates in terms of all variables

(859,781 observations deleted)
. save  `output_name'_TransId_only.dta, replace
file ZTrans17031_TransId_only.dta saved
. }

.
.
. local run_this = 1

. ************************
. ** Import ZTrans  Main **
.
.
. if `run_this' == 1 {
. local table = "Main"
. local keepvars ""TransId DocumentTypeStndCode RecordingDate DocumentDate Sign
> atureDate EffectiveDate SalesPriceAmount""
.
. di c(current_time)
21:40:18
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZTrans" `loc_layout' `loc_data' `keepvars'
. di c(current_time)
22:40:45
.
. * could create an egen here to count number of parcels involved in each trans
> actions
. * in order to remove multi-parcel transactions later on when evaluating price
> s
.
. **merge with existing dataset
. merge 1:m TransId using `output_name'

    Result                           # of obs.
    -----------------------------------------
    not matched                    10,781,215
        from master                10,781,215  (_merge==1)
        from using                          0  (_merge==2)

    matched                         9,323,334  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : record in this file but not in master file
. *          (presumably because it is not in correct FIPS code)
. *    _merge==2 : record in master file but there are no details to add in thi
> s file
. *    _merge==3 : record in master file AND new details are in this file
.
.
. **remove records that don't belong
. drop if _merge==1
(10,781,215 observations deleted)
. drop _merge
.
. **and export the master table...
. save `output_name', replace
file ZTrans17031.dta saved
. }

.
.
. local run_this = 1
```

```
. ************************************
. ** Import ZTrans  BuyerMailAddress **
.
. if `run_this' == 1 {
. local table = "BuyerMailAddress"
. local keepvars ""TransId BuyerMailFullStreetAddr BuyerMailCity BuyerMailZip B
> uyerMailSequenceNumber""
.
. di c(current_time)
22:42:29
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZTrans" `loc_layout' `loc_data' `keepvars'
. di c(current_time)
22:54:33
.
. ** keep only first buyer mailing address per transaction.
. ** (hopefully, drop 0 or near 0 records)
. drop if BuyerMailSequenceNumber > 1
(0 observations deleted)
.
.
. merge 1:1 TransId using `output_name'_TransId_only.dta

    Result                           # of obs.
    -----------------------------------------
    not matched                     9,504,861
        from master                 4,963,417  (_merge==1)
        from using                  4,541,444  (_merge==2)

    matched                         3,922,109  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : address record exists but transaction is not in transactions
>  list
. *          (presumably because it is not in correct FIPS code)
. *    _merge==2 : record in transactions list but no addresses associated
. *    _merge==3 : record in transactions list AND address is found
.
. **remove records that don't belong
. drop if _merge != 3
(9,504,861 observations deleted)
. drop _merge
.
. **and export BuyerMailAddress table...
. save `output_name'_BuyerMailAddress.dta, replace
(note: file ZTrans17031_BuyerMailAddress.dta not found)
file ZTrans17031_BuyerMailAddress.dta saved
. }

.
.
. local run_this = 1

. *****************************
. ** Import ZTrans  BuyerName **
.
. if `run_this' == 1 {
. local table = "BuyerName"
. local keepvars ""TransId BuyerIndividualFullName BuyerNonIndividualName Buyer
> NameSequenceNumber BuyerMailSequenceNumber""
.
. di c(current_time)
22:55:19
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZTrans" `loc_layout' `loc_data' `keepvars'
```

```
. di c(current_time)
22:57:35
.
. ** keep only first buyer mailing address per transaction.
. ** (hopefully, drop 0 or near 0 records)
. drop if BuyerMailSequenceNumber > 1
(140 observations deleted)
.
. ** bring in buyer addresses
. merge m:1 TransId using `output_name'_BuyerMailAddress.dta

    Result                           # of obs.
    -----------------------------------------
    not matched                     7,704,772
        from master                 7,704,423  (_merge==1)
        from using                        349  (_merge==2)

    matched                         5,883,080  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : record in name file but not in address file
. *          (presumably because it is not in correct FIPS code)
. *    _merge==2 : record has buyer address but not buyer name
. *    _merge==3 : record has buyer name and buyer address
.
. drop if _merge == 1
(7,704,423 observations deleted)
. drop _merge BuyerMailSequenceNumber
.
. ** keep only first three buyer names
. drop if BuyerNameSequenceNumber > 3
(45,795 observations deleted)
.
. ** generate single name field
. replace BuyerNonIndividualName = BuyerNonIndividualName + " (Org)" if BuyerNo
> nIndividualName != ""
(1,272,597 real changes made)
. gen BuyerName = "_noname", after(TransId)
. replace BuyerName = BuyerIndividualFullName if BuyerIndividualFullName != ""
variable BuyerName was str7 now str81
(4,565,037 real changes made)
. replace BuyerName = BuyerNonIndividualName if BuyerNonIndividualName != ""
(1,272,597 real changes made)
. drop BuyerIndividualFullName BuyerNonIndividualName
.
. ** reshape so every transaction has only a single row
. reshape wide BuyerName, i(TransId BuyerMailFullStreetAddr BuyerMailCity Buyer
> MailZip) j(BuyerNameSequenceNumber)
(note: j = 1 2 3)

Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                  5.8e+06   -> 3.9e+06
Number of variables                   6   ->       7
j variable (3 values)BuyerNameSequenceNumber-> (dropped)
xij variables:
                              BuyerName   ->   BuyerName1 BuyerName2 BuyerName3
-----------------------------------------------------------------------------
.
.
. ** merge with existing dataset
. merge 1:m TransId using `output_name'

    Result                           # of obs.
    -----------------------------------------
    not matched                     4,874,276
```

```
        from master                         0  (_merge==1)
        from using                  4,874,276  (_merge==2)

    matched                         4,449,058  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : transaction has buyer details but is not in master file
. *          (should be near 0)
. *    _merge==2 : record in master file but there are no buyer details to add
. *    _merge==3 : record in master file AND buyer details have been added
.
. drop if _merge == 1
(0 observations deleted)
.
. generate BuyerDetails = 0
. replace BuyerDetails = 1 if _merge == 3
(4,449,058 real changes made)
. drop _merge
.
. **and export the master table...
. save `output_name', replace
file ZTrans17031.dta saved
. }

.
.
. local run_this = 1

. ******************************
. ** Import ZTrans SellerName **
.
. if `run_this' == 1 {
. local table = "SellerName"
. local keepvars ""TransId SellerIndividualFullName SellerNonIndividualName Sel
> lerMailSequenceNumber SellerNameSequenceNumber""
.
. di c(current_time)
23:02:15
. *Starting import subfunction. This could take a while.
. run `loc_subfn' `table' "ZTrans" `loc_layout' `loc_data' `keepvars'
. di c(current_time)
23:04:47
.
. ** keep only first buyer mailing address per transaction.
. ** (hopefully, drop 0 or near 0 records)
. drop if SellerMailSequenceNumber > 1
(9,960,786 observations deleted)
. drop SellerMailSequenceNumber
.
. ** keep only first three seller names
. drop if SellerNameSequenceNumber > 3
(78,310 observations deleted)
.
. ** generate single seller name
. replace SellerNonIndividualName = SellerNonIndividualName + " (Org)" if Selle
> rNonIndividualName != ""
(1,495,326 real changes made)
. gen SellerName = "_noname", after(TransId)
. replace SellerName = SellerIndividualFullName if SellerIndividualFullName !=
> ""
variable SellerName was str7 now str81
(2,957,614 real changes made)
. replace SellerName = SellerNonIndividualName if SellerNonIndividualName != ""
(1,495,326 real changes made)
. drop SellerIndividualFullName SellerNonIndividualName
.
```

```
. ** reshape so every transaction has only a single row
. reshape wide SellerName, i(TransId) j(SellerNameSequenceNumber)
(note: j = 1 2 3)

Data                               long   ->   wide
-----------------------------------------------------------------------------
Number of obs.                  4.5e+06   -> 2.7e+06
Number of variables                   3   ->       4
j variable (3 values)SellerNameSequenceNumber->(dropped)
xij variables:
                             SellerName   ->   SellerName1 SellerName2 SellerNa
> me3
-----------------------------------------------------------------------------
.
. ** merge with existing dataset
. merge 1:m TransId using `output_name'

    Result                           # of obs.
    -----------------------------------------
    not matched                     9,293,656
        from master                 1,442,745  (_merge==1)
        from using                  7,850,911  (_merge==2)

    matched                         1,472,423  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : transaction has seller details but is not in master file
. *          (presumably because it is not in correct FIPS code)
. *    _merge==2 : record in master file but there are no seller details to add
. *    _merge==3 : record in master file AND seller details have been added
.
. drop if _merge == 1
(1,442,745 observations deleted)
.
. generate SellerDetails = 0
. replace SellerDetails = 1 if _merge == 3
(1,472,423 real changes made)
. drop _merge
.
. **and export the master table...
. save `output_name', replace
file ZTrans17031.dta saved
.
. erase `output_name'_BuyerMailAddress.dta
. }

.
.
. local run_this = 1

. **********************
. *** Cleanup records ***
.
. if `run_this' == 1 {
.
. ** fix parcelID
. gen  pin_length = length(UnformattedAssessorParcelNumber)
. tab pin_length

 pin_length |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |     25,947        0.28        0.28
          1 |        442        0.00        0.28
          2 |          8        0.00        0.28
          3 |          4        0.00        0.28
          4 |         15        0.00        0.28
```

```
          5 |         25        0.00        0.28
          6 |         40        0.00        0.28
          7 |        210        0.00        0.29
          8 |        225        0.00        0.29
          9 |        900        0.01        0.30
         10 |  7,131,391       76.49       76.79
         11 |      4,680        0.05       76.84
         12 |        733        0.01       76.85
         13 |      1,107        0.01       76.86
         14 |  2,156,474       23.13       99.99
         15 |        749        0.01      100.00
         16 |        129        0.00      100.00
         17 |        142        0.00      100.00
         18 |         69        0.00      100.00
         19 |         12        0.00      100.00
         20 |          7        0.00      100.00
         21 |          3        0.00      100.00
         22 |          7        0.00      100.00
         23 |          4        0.00      100.00
         25 |          1        0.00      100.00
         26 |          2        0.00      100.00
         27 |          2        0.00      100.00
         28 |          3        0.00      100.00
         30 |          1        0.00      100.00
         35 |          1        0.00      100.00
         36 |          1        0.00      100.00
------------+-----------------------------------
      Total |  9,323,334      100.00
. keep if pin_length == 10 | pin_length == 14
(35,469 observations deleted)
. gen pin14 = UnformattedAssessorParcelNumber if pin_length == 14
(7,131,391 missing values generated)
. replace pin14 = UnformattedAssessorParcelNumber + "0000" if pin_length == 10
(7,131,391 real changes made)
. gen  pin_length2 = length(pin14)
. assert pin_length2 == 14
. drop AssessorParcelNumber UnformattedAssessorParcelNumber pin_length pin_leng
> th2
.
. ** date-ify all dates
. * per Zillow, DocumentDate > SignatureDate > RecordingDate
. foreach var of varlist RecordingDate DocumentDate SignatureDate {
  2.         display "working on `var'..."
  3.         generate str4 dxyr1 = substr(`var',1,4) if `var' != ""
  4.         generate str2 dxmo1 = substr(`var',6,7) if `var' != ""
  5.         generate str2 dxda1 = substr(`var',9,10) if `var' != ""
  6.         destring dx*, replace
  7.         gen `var'_2 = mdy(dxmo1, dxda1, dxyr1)
  8.         format `var'_2 %d
  9.         drop dxyr1 dxmo1 dxda1 `var'
 10.         rename `var'_2 `var'
 11. }
working on RecordingDate...
(1 missing value generated)
(1 missing value generated)
(1 missing value generated)
dxyr1: all characters numeric; replaced as int
(1 missing value generated)
dxmo1: all characters numeric; replaced as byte
(1 missing value generated)
dxda1: all characters numeric; replaced as byte
(1 missing value generated)
(1 missing value generated)
working on DocumentDate...
(1,599,637 missing values generated)
(1,599,637 missing values generated)
```

```
(1,599,637 missing values generated)
dxyr1: all characters numeric; replaced as int
(1599637 missing values generated)
dxmo1: all characters numeric; replaced as byte
(1599637 missing values generated)
dxda1: all characters numeric; replaced as byte
(1599637 missing values generated)
(1,599,637 missing values generated)
working on SignatureDate...
(9,287,865 missing values generated)
(9,287,865 missing values generated)
(9,287,865 missing values generated)
dxyr1: all characters numeric; replaced as byte
(9287865 missing values generated)
dxmo1: all characters numeric; replaced as byte
(9287865 missing values generated)
dxda1: all characters numeric; replaced as byte
(9287865 missing values generated)
(9,287,865 missing values generated)
.
. ** create best date
. * per Zillow, DocumentDate > SignatureDate > RecordingDate
. gen BestDate = DocumentDate
(1,599,637 missing values generated)
. replace BestDate = SignatureDate if BestDate == . & SignatureDate != .
(0 real changes made)
. replace BestDate = RecordingDate if BestDate == . & RecordingDate != .
(1,599,637 real changes made)
. format BestDate %d
. drop EffectiveDate SignatureDate FIPS
.
. ** create buyer type variable and remove "(Org)" from BuyerNames
. gen BuyerType = ""
(9,287,865 missing values generated)
. foreach var of varlist BuyerName* {
  2.         display "`var'..."
  3.
.         *create buyer type variable:
.         gen str1 buyertype_builder = "-"
  4.         replace buyertype_builder = "1" if substr(`var',-5,.)=="(Org)"
  5.         replace buyertype_builder = "0" if buyertype_builder != "1" & `var
> ' != ""
  6.         replace BuyerType = BuyerType + buyertype_builder
  7.         drop buyertype_builder
  8.
.         *remove "(Org)"
.         replace `var' = subinstr(`var', " (Org)", "", .)
  9. }
BuyerName1...
(1,028,586 real changes made)
(3,404,265 real changes made)
(9,287,865 real changes made)
(1,028,586 real changes made)
BuyerName2...
(428,519 real changes made)
(1,559,895 real changes made)
variable BuyerType was str1 now str2
(9,287,865 real changes made)
(428,519 real changes made)
BuyerName3...
(60,216 real changes made)
(68,297 real changes made)
variable BuyerType was str2 now str3
(9,287,865 real changes made)
(60,216 real changes made)
.
```

```
. ** create seller type variable and remove "(Org)" from SellerNames
. gen SellerType = ""
(9,287,865 missing values generated)
. foreach var of varlist SellerName* {
  2.         display "`var'..."
  3.
.         *create buyer type variable:
.         gen str1 sellertype_builder = "-"
  4.         replace sellertype_builder = "1" if substr(`var',-5,.)=="(Org)"
  5.         replace sellertype_builder = "0" if sellertype_builder != "1" & `v
> ar' != ""
  6.         replace SellerType = SellerType + sellertype_builder
  7.         drop sellertype_builder
  8.
.         *remove "(Org)"
.         replace `var' = subinstr(`var', " (Org)", "", .)
  9. }
SellerName1...
(548,126 real changes made)
(919,446 real changes made)
(9,287,865 real changes made)
(548,126 real changes made)
SellerName2...
(338,805 real changes made)
(447,243 real changes made)
variable SellerType was str1 now str2
(9,287,865 real changes made)
(338,805 real changes made)
SellerName3...
(65,248 real changes made)
(43,469 real changes made)
variable SellerType was str2 now str3
(9,287,865 real changes made)
(65,248 real changes made)
.
. **order variables
. order pin14 ImportParcelID TransId BestDate DocumentDate RecordingDate SalesP
> riceAmount DocumentTypeStndCode BuyerDetails SellerDetails BuyerMailFullStree
> tAddr BuyerMailCity BuyerMailZip BuyerType BuyerName1 BuyerName2 BuyerName3 S
> ellerType SellerName1 SellerName2 SellerName3
.
. **and export the master table...
. save `output_name', replace
file ZTrans17031.dta saved
. }

.
.
. local run_this = 1

. ************************************************
. ** Select most recent transaction per property **
.
. if `run_this' == 1 {
. ** keep only most recent transaction
. sort pin14 TransId
. gen seq=1, before(TransId)
. replace seq=seq[_n-1]+1 if pin14==pin14[_n-1]
(7,556,407 real changes made)
. gen nseq = -seq, before(TransId) //generates an inverse sequence rank
. sort pin14 nseq //within each PIN, places last transaction first
. by pin14 : keep if _n==1 //keeps first record (last transaction) per pin
(7,556,407 observations deleted)
. drop seq nseq
.
. **export this table...
```

```
. save `output_name'_mostrecenttrans, replace
file ZTrans17031_mostrecenttrans.dta saved
. }

.
.
. local run_this = 1

. **************************************
. ** Merge ZTrans_MostRecent with ZAsmt **
.
. if `run_this' == 1 {
. merge 1:1 pin14 using `output_name2'.dta

    Result                           # of obs.
    -----------------------------------------
    not matched                         335,565
        from master                     103,794  (_merge==1)
        from using                      231,771  (_merge==2)

    matched                           1,627,664  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : transaction in ZTrans but no record in ZAsmt
. *    _merge==2 : transaction in ZAsmt but no record in ZTrans
. *    _merge==3 : Record in ZAsmt and at least one trans in ZTrans
.
. drop ImportParcelID TransId RowID AssessmentRecordMatchFlag _merge
. save `output_name3'_mostrecenttrans, replace
file ZCombined17031_mostrecenttrans.dta saved
. }

.
.
.
. local run_this = 1

. *******************************************************
. ** Export various study area files using MostRecentTrans**
.
. if `run_this' == 1 {
. import delimited `parcellist', stringcols(2) clear
(13 vars, 71,533 obs)
. drop objectid shape_length
. rename shape_area parcel_size
. merge 1:1 pin14 using `output_name3'_mostrecenttrans

    Result                           # of obs.
    -----------------------------------------
    not matched                       1,892,980
        from master                         642  (_merge==1)
        from using                    1,892,338  (_merge==2)

    matched                              70,891  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : record in study area but no transaction in ZTRAX
. *    _merge==2 : record in ZTRAX but not in study area
. *    _merge==3 : record in study area and has a ZTRAX transaction
. drop if _merge == 2
(1,892,338 observations deleted)
. drop _merge
.
. ** first, keep all parcels in study area
. save `output_name3'_mostrecenttrans_studyarea, replace
file ZCombined17031_mostrecenttrans_studyarea.dta saved
```

```
.
. ** second, keep only residential vacant land per DPD or CMAP
. keep if status_improved > 0 | landuse == "VACANT_RES"
(59,339 observations deleted)
. save `output_name3'_mostrecenttrans_studyarea_vacant, replace
file ZCombined17031_mostrecenttrans_studyarea_vacant.dta saved
. export delimited using `output_name3'_mostrecenttrans_studyarea_vacant.csv, r
> eplace
file ZCombined17031_mostrecenttrans_studyarea_vacant.csv saved
.
. ** third, keep only LL transactions
. keep if status_improved > 0
(11,353 observations deleted)
.
. tab BuyerDetails

BuyerDetail |
          s |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         30        3.68        3.68
          1 |        786       96.32      100.00
------------+-----------------------------------
      Total |        816      100.00
. tab SellerDetails

SellerDetai |
         ls |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         62        7.60        7.60
          1 |        754       92.40      100.00
------------+-----------------------------------
      Total |        816      100.00
. tab DocumentTypeStndCode

DocumentTyp |
  eStndCode |      Freq.     Percent        Cum.
------------+-----------------------------------
       COCA |          4        0.49        0.49
       DEED |          4        0.49        0.98
       FASD |          7        0.86        1.84
       INTR |          2        0.25        2.08
       MTGE |         13        1.59        3.68
       NFCM |          2        0.25        3.92
       NTSL |          8        0.98        4.90
       OTHR |         10        1.23        6.13
       QCDE |        739       90.56       96.69
       RRDE |          3        0.37       97.06
       SHDE |          5        0.61       97.67
       SPWD |          2        0.25       97.92
       TXDE |         16        1.96       99.88
       WRDE |          1        0.12      100.00
------------+-----------------------------------
      Total |        816      100.00
. tab SellerType

 SellerType |      Freq.     Percent        Cum.
------------+-----------------------------------
        --- |         62        7.60        7.60
        0-- |         19        2.33        9.93
        00- |          1        0.12       10.05
        1-- |        734       89.95      100.00
------------+-----------------------------------
      Total |        816      100.00
. drop SellerName2 SellerName3
.
. keep if DocumentTypeStndCode == "QCDE"
```

```
(102 observations deleted)
.
. tab SellerName1

                         1 SellerName |      Freq.     Percent        Cum.
--------------------------------------+-----------------------------------
                      CITI OF CHICAGO |          1        0.14        0.14
                      CITY OF CHICAGO |        725       99.45       99.59
                     CITY OF MORTGAGE |          1        0.14       99.73
              HIGHLAND COMMUNITY BANK |          1        0.14       99.86
                        JEWEL SPENCER |          1        0.14      100.00
--------------------------------------+-----------------------------------
                                Total |        729      100.00
. keep if SellerName1 == "CITY OF CHICAGO" | SellerName1 == "CITI OF CHICAGO"
(13 observations deleted)
.
. tab BuyerType

  BuyerType |      Freq.     Percent        Cum.
------------+-----------------------------------
        0-- |        598       82.37       82.37
        00- |         25        3.44       85.81
        1-- |         98       13.50       99.31
        100 |          1        0.14       99.45
        11- |          4        0.55      100.00
------------+-----------------------------------
      Total |        726      100.00
. drop BuyerName3 SellerType
. drop DocumentTypeStndCode BuyerDetails SellerDetails SalesPriceAmount SellerN
> ame1
.
. drop if BestDate < date("20140101","YMD")
(1 observation deleted)
.
. save `output_name3'_studyarea_LLtrans, replace
file ZCombined17031_studyarea_LLtrans.dta saved
. export delimited using `output_name3'_studyarea_LLtrans.csv, replace
file ZCombined17031_studyarea_LLtrans.csv saved
. }

.
.
. local run_this = 1

. *************************************************************
. ** Export all transactions for vacant land in study area **
.
. if `run_this' == 1 {
. import delimited `parcellist', stringcols(2) clear
(13 vars, 71,533 obs)
. drop objectid shape_length
. rename shape_area parcel_size
.
. **merge in assessment data
. merge 1:1 pin14 using `output_name2'.dta

    Result                           # of obs.
    -----------------------------------------
    not matched                     1,789,668
        from master                       883  (_merge==1)
        from using                  1,788,785  (_merge==2)

    matched                            70,650  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : Record in study area but no ZAsmt entry
```

```
. *    _merge==2 : ZAsmt record, but not in study area
. *    _merge==3 : Study area parcel with ZAsmt record
.
. drop if _merge == 2
(1,788,785 observations deleted)
. drop _merge
.
. ** merge in transaction data
. merge 1:m pin14 using `output_name'.dta

    Result                           # of obs.
    -----------------------------------------
    not matched                     8,947,224
        from master                    11,950  (_merge==1)
        from using                  8,935,274  (_merge==2)

    matched                           352,591  (_merge==3)
    -----------------------------------------
. * Where...
. *    _merge==1 : record in study area but no transaction in ZTRAX
. *    _merge==2 : record in ZTRAX but not in study area
. *    _merge==3 : record in study area and has a ZTRAX transaction
. drop if _merge == 2
(8,935,274 observations deleted)
. drop _merge
.
. drop AssessmentRecordMatchFlag RowID
.
. ** keep sales since 2000 only
. gen year = year(BestDate)
(11,950 missing values generated)
. drop if year < 2000
(52,170 observations deleted)
.
. save `output_name3'_studyarea, replace
file ZCombined17031_studyarea.dta saved
.
. ** keep if residential vacant land per DPD or CMAP
. keep if status_improved > 0 | landuse == "VACANT_RES"
(284,226 observations deleted)
.
. save `output_name3'_studyarea_vacant, replace
file ZCombined17031_studyarea_vacant.dta saved
. }

.
.
.
.
. ********************
. ***** Close Log *****
. if `tolog' == 1 {
.         log close
      name:  <unnamed>
       log:  Z:\Chicago\DataExtracts\log_31Mar2019_203142.smcl
  log type:  smcl
 closed on:  31 Mar 2019, 23:24:27
--------------------------------------------------------------------------------
```

Appendix B: Detailed Data Assembly and GIS Methodology

Bold names refer to feature classes stored in the analysis geodatabase. Geodata is available from the author upon request.

1. Prepped Large Lots 1 data
    a. Downloaded from website https://largelots.org/about/ – has status on (0, 3) scale. Only has eligible parcels.
        i. 0 = not applied for
        ii. 1 = applied for but not approved
        iii. 2 = applied for and approved
        iv. 3 = applied for and sold
    b. To get all parcels in area, downloaded cook county assessor parcel data from 2014: https://datacatalog.cookcountyil.gov/GIS-Maps/ccgisdata-Parcel-2014/2m9h-cq6j
    c. Intersected parcel data with Large Lots boundary shapefile from https://largelots.org/about/ and deleted all fields but PIN14.
    d. Joined eligible parcels to all parcel data on PIN14.
    e. Assigned a status of -1 to any parcel that existed in the assessor data but not the large lots data (as this was not eligible for sale)
2. Combined Large Lots 1 data with Large Lots 2 and 3 data from https://largelots.org/about/
    a. Large Lots 2 (EGP) data has status on (-1, 3) scale, which is ideal. Large Lots 3 (Austin) data doesn't use the numeric range, instead containing an 'eligible' field with a F or T and no data on actual sales. These two datasets have all parcels in region.
        i. -1 = ineligible lot
        ii. 0 = not applied for
        iii. 1 = applied for but not approved
        iv. 2 = applied for and approved
        v. 3 = applied for and sold
    b. Merged all three datasets
    c. Field calculated Austin Lots so that eligible = F → status = -1 and eligible = T → status = 0, then removed eligible field. Note this means Austin sales not yet recorded.
    d. This is **Parcels_1**
3. Combined with CMAP land use data
    a. Downloaded 2013 land use inventory from https://datahub.cmap.illinois.gov/dataset/land-use-inventory-for-northeast-illinois-2013
    b. Converted Parcels_1 to points, then spatial joined points to the land use inventory. Finally, rejoined points to polygons from Parcels_1.
    c. Removed roads and TCU land uses
    d. This is **Parcels_2** (field 'descr' is land use)
4. Joined Parcels_2 with data of confirmed large lots sales received from DPD staff in August 2018
    a. Manufactured clean PIN14 in excel from tabular data provided by DPD
    b. Imported as feature class into ArcGIS and joined in PIN14
    c. Deleted 9 polygons without PINs in Austin that were overlapping polygons with parcel PINs

    d. This is **Parcels_3**
5. Joined Parcels_3 with data with confirmed large lots sales received from DPD staff in September 2018 (hopefully better data)
    a. Manufactured clean PIN14 in excel from tabular data provided by DPD for selected three pilots and expansion. Imported as table **DPD1**
    b. First, Joined DPD1 with all cook county parcels from assessor data (See step 1b) to confirm that all join on PIN. 100% (n=1187) joined. Exported only lot sales from these four programs. This is **Parcels_4**
    c. Second, joined DPD1 with Parcels_3 to focus only on the parcels in the three study regions. 841 parcels joined. Now, status_source1 refer to LL.org data. sale_source2 refers first round of DPD data, and sale_source3 as well as address and buyer fields refer to the second round of DPD data. This is **Parcels_5**.
    d. Because 1187-841=346, 346 expansion sales occurred outside these three neighborhoods. Because 698 total expansion sales, 352 expansion sales occurred in these neighborhoods. Counted parcels in Parcels_4 that were not identical to parcels in Parcels_5 to confirm a difference of 346.
6. Calculated new fields
    a. a best guess status (status_improved) field, which uses the following codes (coded from highest number to lowest):
        i. -1 = ineligible (assigned to all status_improved = null rows after all the other statuses were assigned)
        ii. 0 = available in initial pilot but not purchased in pilot or expansion (availability from status_source1={0,1,2} but only if not indicated as sold in sales_source3)
        iii. 1 = purchased in LargelLots1 (Englewood) (per sale_source3)
        iv. 2 = purchased in LargeLots2 (East Garfield Park) (per sale_source3)
        v. 3 = purchased in LargeLots3 (Austin) (per sale_source3)
        vi. 4 = Purchased in Expansion (per sale_source3)
    b. A neighborhood field, which uses the 1,2,3, codes from above for Englewood, EGP, and Austin.
    c. This is **Parcels_6.** It contains all of these various sale fields. An analysis in Excel (**quality of improved status field.xlsx**) indicates that 99.9% of parcels are coded such that their status is corroborated by the largelots.org data.
    d. This data is then exported with the other status fields dropped as **Parcels_7**
7. Added other geographies
    a. Downloaded Chicago neighborhoods from https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9
    b. Spatial Joined Parcels_7 to the above to capture neighborhood of each parcel. Had to manually assign 5 parcels that fell across neighborhood boundaries. Used "pri_neighborhood" field as neighborhood
    c. Downloaded Chicago census tracts from https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik
    d. Used spatial join tool to perform one-to-one join using join as tracts shapefile, method of 'HAVE THEIR CENTER IN'.

    e. Removed unnecessary fields. Renamed certain fields from DPD for clarity. This is **Parcels_8**
8. Produced table with all parcels in three neighborhoods for ZTRAX pull, and pulled ZTRAX transaction data
    a. Turned off unnecessary fields from Parcels_8 and exported table into geodatabase as **ZTRAX_parcelrequest**. This table contains 71,533 records, and should be every polygon in the three study neighborhoods.
    b. Exported this table as a text file for ZTRAX pull.
    c. Ran ZTrans data pull process in STATA using the above text file (See Appendix A). This produced a file with all the records that are certainly LL sales. Created a pin14str field, because pin14 was imported as a double. This is a table in the geodatabase called **ZTRAX_LLdata**
9. Prepped roads data, Part 1
    a. Downloaded roads from https://data.cityofchicago.org/Transportation/Street-Center-Lines/6imu-meau (metadata https://data.cityofchicago.org/api/assets/06DEC62C-ACAB-42D3-A540-378F8464F83D)
    b. ADDR_START as min of LOGICRF, LOGICLF, unless one of those fields was 0, in which case max taken
    c. Generated BLOCKNAME as VBSCRIPT: [addr_start]& " " & [pre_dir]& " " & [street_nam]& " " & [street_typ]& " " & [suf_dir]
    d. Dropped unnecessary fields
    e. Eliminated highways, onramps, and other non-block roads with: class in ('2', '3', '4', '5', '7')
    f. Counted frequency of occurrences of block name (created freq_thisblockname]
    g. This is **Roads_1**
10. Prepped roads data, Part 2
    a. Selected roads that were within 100 ft of Parcels_3. This is **Roads_2**
    b. Manually selected blocks that would be good for analysis, which were marked as BlockType=1. These are predominantly residential on both sides of the street, with primarily single-family-sized parcels; relatively straight blocks; residential streets (avoided major through-ways and streets that had substantial commercial). Blocks not good for analysis marked BlockType=0
    c. To better visualize roads, used the “shorten polylines” toolbox downloaded from internet to shorten all roads by 50 ft from both sides (25’ each side). This is **Roads_3**.
11. Prepped roads for block-making and parcel attachment
    a. used “shorten polylines” to shorten roads 100 ft from both sides (50’ each side). This is **Roads_4**
    b. joined Roads_4 to Roads_2 to recollect important fields. Deleted unnecessary fields from join process. This is **Roads_5**
    c. Used definition query to obtain only “good blocks” (BlockType=1).
    d. Manually inspected all blocks under 200’ long and edited blocks to combine them with adjacent blocks or deleted them as logical. This is **Roads_6**
    e. Used buffer tool to buffer Roads_6 using inputs distance=75’; sidetype=full, endtype=flat. Made some custom edits to the resulting polygons to include the right

pacels—specifically joining adjacent polygons on same block, correcting for cut-offs in cul-de-sacs, etc.

f. Created a fixed BlockID for numerical block-tracking that is distinct from objectID.
g. Created approximate Block_length as (shapelength – 300)/2 + 100
h. Deleted unnecessary fields
i. Manually expanded blockID 1178 (4700 W West End Ave) due to street width—75' wasn't quite enough.
j. This is **Roads_7**

12. Joined parcel data with blocks
    a. Used spatial join tool to perform a one-to-many intersect using target=Parcels_8, join=Roads_7. The outcome of this join is that parcel polygons on a good-for-analysis block now have blockIDs attached. Any parcel near two good-for-analysis blocks is now duplicated, so that it appears in the analyses of both blocks.
    b. An excel analysis (**duplicate parcels checker.xlsx**) finds a total of 212 parcels (out of 45,213) apply to two blocks. Of these, 202 are ineligible parcels (no impact on final analysis). Only 10 were eligible or sold.
    c. Removed unnecessary fields from join. This is **Parcels_9**
13. Gathered buyer address data from ZTRAX database
    a. Geocoded buyer mailing address data from **ZTRAX_LLdata** using DM10 geocoder obtained from Phil McDaniel. Went through errors to correct typos in Buyer Address field (lowercase text = correction), and then geocoded again. 55 of 725 failed to match (7.6%). This is **Buyers_1**
    b. Changed coordinate system to NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet, and calculated x and y locations of buyer for 670 sales. This is **Buyers_2.**
    c. Joined parcels_8 (not 9 because we don't want to limit to crime study blocks) with buyers_2 on pin14/pin14str. Calculated x and y locations of parcels. This is **Buyers_3**.
    d. Used xy to line tool to draw lines between buyer and parcel (id = pin14). Calculated distance of line as new field. Classified parcel buyer into one of four categories (dist_type). This is **Buyers_4**.
        i. 1 = buyer within 1000 ft
        ii. 2 = buyer in neighborhood (determined by doing a spatial join where the line is entirely within Chicago neighborhood areas from step 7) (all are within 10,000 ft)
        iii. 3 = buyer within 10,000 ft, but not same neighborhood
        iv. 4 = buyer within 24,000 ft
        v. 5 = buyer within metro area
        vi. 6 = out of metro
    e. Joined parcels_9 with buyers_4 and kept only distance from buyer and distance classification variables.
    f. Created a vacancy type field to generally interpret vacancy levels in each study neighborhood:
        i. 0 = not believed to be vacant (not a higher class)
        ii. 1 = vacant, privately owned (CMAP land use is 'VACANT*', not a higher class)
        iii. 2 = vacant, city owned (status_improved = 0)

iv. 3 = vacant, sold through LL program (status_improved > 0)

g. This is **parcels_10**.

14. Retrieved crime data
    a. Downloaded csv from https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2 (updated download March 2019) (variable dictionary available at same link)
    b. Note: IUCR code data dictionary can be found at https://data.cityofchicago.org/Public-Safety/Chicago-Police-Department-Illinois-Uniform-Crime-R/c7ck-438e/data
    c. Kept only crimes between 2010 and 2018 (dropped 4M records from pre-2010, 36k records from 2019.)
    d. Kept only records with geolocations (xcoordinate and ycoordinate) (dropped 16.9k)
    e. Dropped unnecessary variables. Exported to **/data sources/crime/crime.csv**. This has 2.69M crimes.
    f. Imported this into ArcGIS using xcoordinate and ycoordinate and coordinate system NAD_1983_StatePlane_Illinois_East_FIPS_1201_Feet. This is **Crime_0**.
    g. Selected for crimes within 150 ft of Roads_2. This is **Crime_1**
15. Joined crime data with blocks
    a. Joined Crime_1 to Roads_7 (spatial join, falls inside) to give crimes the attributes of blocks. This is **Crime_2**
    b. Exported only crimes that fell in a Roads_7 polygon (in other words, on a block to be analyzed), and only desired fields. This is **Crime_3**
16. Exported **Parcels_10** and **Crime_3** to **0_parcels.csv** and **0_crimes.csv** for stata analysis.
17. For city-wide statistics on city-owned property
    a. Downloaded Chicago city-owned land inventory from https://data.cityofchicago.org/Community-Economic-Development/City-Owned-Land-Inventory/aksk-kvfp
    b. Matched on PIN14 with Cook county assessors parcel file. https://datacatalog.cookcountyil.gov/GIS-Maps/ccgisdata-Parcel-2014/2m9h-cq6j
    c. Kept only matched records and desired fields. This is **CityOwned_1**
    d. Converted CityOwned_1 to point. Joined point to CMAP land use inventory file (see previous step for link) to collect land use data. Exported desired fields to **CityOwned_2**.
    e. Rejoined CityOwned_1 polygons to CityOwned_2 on PIN14 to bring land use data to polygon file. This is **CityOwned_3**.
    f. Exported **CityOwned_3** to csv for excel analysis.
18. For city-wide statistics on vacancy
    a. Selected all vacant land from CMAP land use inventory file within the city of Chicago shapefile (from https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-City/ewy2-6yfk). This is **Vacant_1**. (note: these are regions, not parcels)
    b. Selected all parcels from the cook county assessors file with centroids inside Vacant_1. This is **Vacant_2**.
    c. Converted Vacant_2 to points. Spatial joined to Vacant_1 to collect specific vacancy class. Joined Vacant_2 to this points file to get data back to polygons. This is **Vacant_3**.