REPORT

Mapping Displacement Pressure in Chicago

December, 2017

TABLE OF CONTENTS

Project Overview

This report describes the methodology, data, and results of an Institute for Housing Studies (IHS) analysis to identify Chicago neighborhoods with different levels of risk associated with rising housing costs, lost affordability, and potential displacement. The goal of this project is to give stakeholders 1) a vital lens to understand neighborhood-level displacement and housing affordability pressures and 2) a framework for the development of proactive strategies to preserve affordability near planned public investments that are both sensitive and responsive to neighborhood needs.

Recent IHS research,"Measuring the Impact of the 606," found that the City of Chicago's new 606 linear park system had a dramatic effect on house prices in the lower-income and high-renter neighborhoods around the western half of the trail. Substantial house price increases in this area were driven by a premium that buyers were willing to pay to purchase properties within a half-mile of The 606. These findings further bolstered established research that public investments can increase demand for housing and lead to rising house prices. They also confirmed the concerns of local advocacy groups that long-time neighborhood residents may be at elevated risk of displacement due to rising costs and declining affordability.

Building on the findings of this report, IHS expanded its analysis citywide to understand which Chicago neighborhoods had a mix of demographic and housing market characteristics similar to areas along the western half of The 606. The following sections highlight the process for calculating the different layers used in this analysis and their results. The first section discusses a market segmentation analysis used to identify neighborhood types with similar demographic, economic, and housing stock characteristics. The second section describes a housing market conditions analysis used to classify areas based on the current market value of housing and recent price trends. The final section combines these two layers and focuses on neighborhoods with a combination of 1) vulnerable populations and 2) above-average price increases that indicate current or potential housing affordability concerns.

The report concludes with an appendix that describes project methodology and data in greater detail.

Market Segmentation Analysis

About the method

In order to assess the underlying vulnerability of populations to displacement in a rising-cost environment, IHS, with faculty and students from DePaul University's Department of Predictive Analytics, developed a market segmentation for City of Chicago neighborhoods. As described in previous IHS reports that use this technique,[1] a segmentation study can analyze the broad range of factors necessary to fully evaluate neighborhood conditions and needs by building an integrated model that identifies geographic units with similar characteristics.[2]

Clustering algorithms incorporate multiple variables to group items by their overall similarity. When applied to community-level data, clustering provides a way to compare different geographic units that share similar traits, regardless of their physical proximity. The goal of this analysis was to use clustering techniques to perform a market segmentation analysis of the City of Chicago. The model incorporated data on housing stock and affordability, resident demographics, crime, and socioeconomic indicators to identify communities with similar characteristics. The segmentation algorithm and associated techniques are discussed in greater detail in the technical appendix.

Overview of market segmentation analysis clustering results

The clustering application identified six distinct neighborhood types in the City of Chicago. Consistent with established and well researched geographic patterns of development and income segregation in the City of Chicago, the algorithm clustered wealthy, economically distressed, and high-density census tracts consistently and distinctly. Clusters were most strongly differentiated by household income, tenure, housing stock, levels of violent crime, and concentration of subsidized housing. The algorithm identified two higher-income clusters (Clusters 2 and 4), with high levels of educational attainment and low-levels of housing cost burden that were primarily differentiated by the age of the housing stock. It also identified two low- to moderate-income clusters (Clusters 5 and 6) with high-cost burden and lower housing values that were primarily differentiated by levels of violent crime and the relative concentration of subsidized housing. It also identified a primarily home ownership cluster (Cluster 3) and a diverse, high density renter cluster (Cluster 1). Figure 1 details each cluster and its key characteristics and Figure 2 maps clusters in the City of Chicago.

Race and ethnicity were not included in the clustering process due to the ways in which geographic patterns of racial segregation in the City of Chicago overwhelm the modeling algorithm and overpower other factors strongly associated with displacement risk, such as renter share, cost-burden, income, age, and etc. Post analysis shows that many clusters have high concentrations of a single race or ethnicity, however. For example, according to data from the 2011-2015 American Community Survey, Cluster 5 is 62.1 percent Latino, Cluster 4 is 70.6 percent white, and Cluster 6 is 89.1 percent African-American.

---

[1] Institute for Housing Studies at DePaul University. "Regional Housing Solutions Data Tool." Chicago, IL: Institute for Housing Studies at DePaul University, 2017.
[2] An extensive history of housing segmentation and its applications can be found in Housing Market Segmentation: A review by Islam and Asami, and The Definition and Identification of Housing Submarkets written by C.A. Watkins.
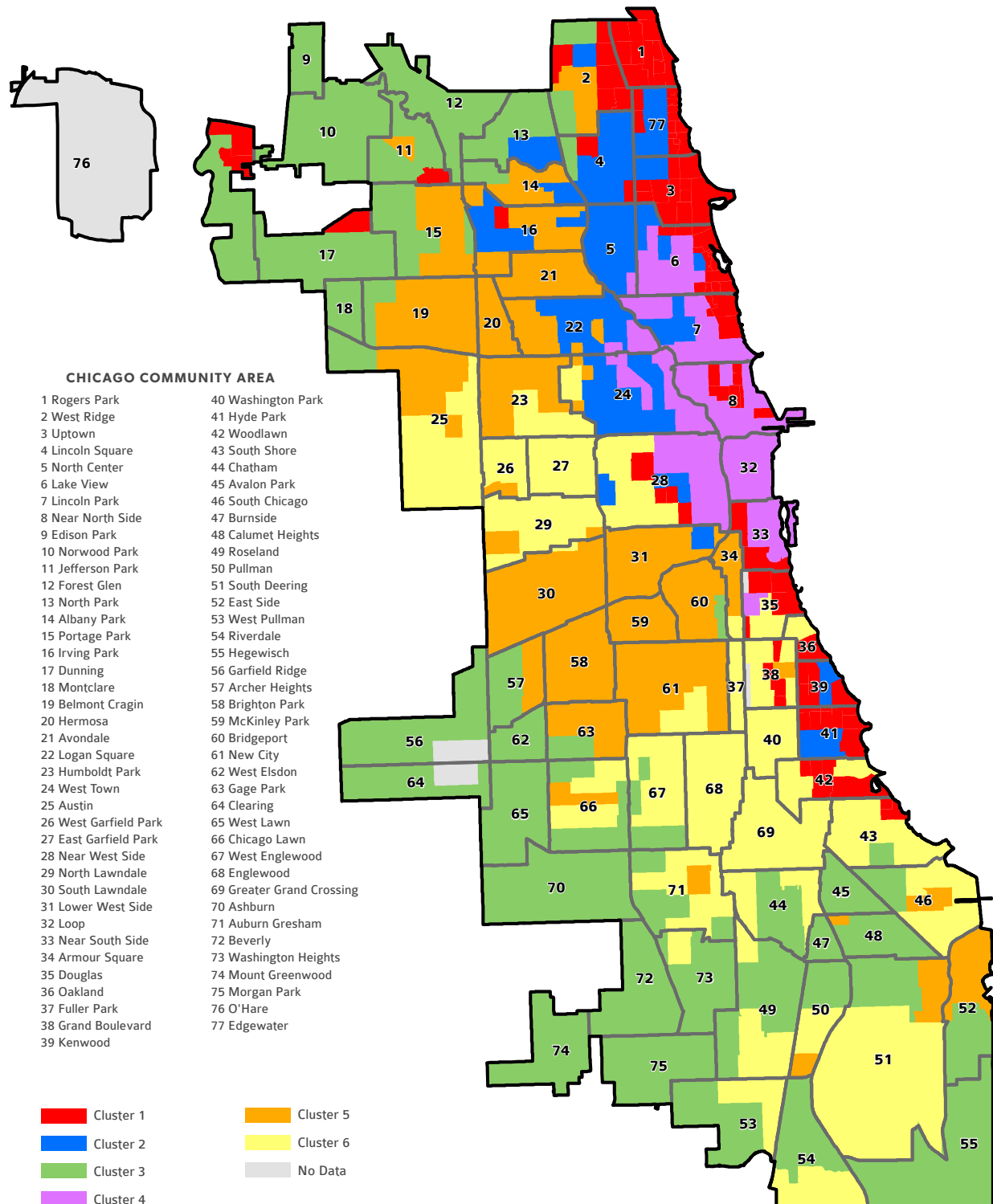
FIGURE 1

Results of the Market Segmentation Analysis for City of Chicago Submarkets

Cluster	Housing Affordability Factors	Housing Stock and Environmental Factors	Demographic and Socioeconomic Factors
1	•Generally lower cost housing but with large recent increases in rents •Recent increases in cost burden	•High density •Relatively high share of subsidized housing units •High share of condominium units and units in large buildings •High levels of renters	•Economically diverse cluster •Declining incomes •High share of seniors •High share of single-person households
2	•High cost and increasingly expensive housing •Low levels of cost burden	•Older, pre-1940s housing stock •Higher share of two-to-four unit buildings	•High and increasing share of higher income households •High levels of educational attainment •High and growing share of two to four person households •Declining share of households in poverty •High share of recent residents •Low share of seniors, low-income households, children, and large families
3	•Low rents and home values •Lower levels of cost burden	•High share of single-family homes •Low share of renter households •High share of homes built post-war	•High but declining share of seniors •High share of long-term residents •High share of larger households •Low levels of educational attainment
4	•Highest and increasing rents and home values •Low levels of cost burden	•High density cluster •Highest share of condo units •Highest levels of new housing developments	•Low share of seniors, children, low-income households, and large families •High income •Highest educational attainment •Lowest levels of households in poverty •Highest share of recent residents
5	•Low but increasing rents and home values •Highest and increasing levels of cost burden	•Highest share of units in two-to-four unit buildings •High share of older buildings •High share of renters •Low levels of subsidized housing	•High levels of low-income households •Declining incomes •Low levels of educational attainment •High share of larger households •High share of children
6	•Low but increasing rents •Low but declining home values •High and increasing cost burden	•High share of units in two-to-four unit buildings •High share of older buildings •High share of renters •High levels of subsidized housing •Highest levels of violent crime	•High levels of low-income households •Highest declines in income levels •Low levels of educational attainment •High share of larger households •High share of children

FIGURE 2

Map of Results of the Market Segmentation Analysis for City of Chicago Submarkets

CHICAGO COMMUNITY AREA

1 Rogers Park	40 Washington Park
2 West Ridge	41 Hyde Park
3 Uptown	42 Woodlawn
4 Lincoln Square	43 South Shore
5 North Center	44 Chatham
6 Lake View	45 Avalon Park
7 Lincoln Park	46 South Chicago
8 Near North Side	47 Burnside
9 Edison Park	48 Calumet Heights
10 Norwood Park	49 Roseland
11 Jefferson Park	50 Pullman
12 Forest Glen	51 South Deering
13 North Park	52 East Side
14 Albany Park	53 West Pullman
15 Portage Park	54 Riverdale
16 Irving Park	55 Hegewisch
17 Dunning	56 Garfield Ridge
18 Montclare	57 Archer Heights
19 Belmont Cragin	58 Brighton Park
20 Hermosa	59 McKinley Park
21 Avondale	60 Bridgeport
22 Logan Square	61 New City
23 Humboldt Park	62 West Elsdon
24 West Town	63 Gage Park
25 Austin	64 Clearing
26 West Garfield Park	65 West Lawn
27 East Garfield Park	66 Chicago Lawn
28 Near West Side	67 West Englewood
29 North Lawndale	68 Englewood
30 South Lawndale	69 Greater Grand Crossing
31 Lower West Side	70 Ashburn
32 Loop	71 Auburn Gresham
33 Near South Side	72 Beverly
34 Armour Square	73 Washington Heights
35 Douglas	74 Mount Greenwood
36 Oakland	75 Morgan Park
37 Fuller Park	76 O'Hare
38 Grand Boulevard	77 Edgewater
39 Kenwood	

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
No Data

Housing market analysis

About the method

In order to better understand areas where current market conditions and recent price increases may indicate increased risk for lost housing affordability, IHS developed a housing market typology based on house prices for City of Chicago neighborhoods. Current house prices can tell us something about neighborhood affordability, while changing house prices illustrate shifting demand for housing and potential lost affordability. Both are valuable indicators for assessing the overall health of neighborhood housing markets and provide a helpful context for the development of housing market interventions.[3] For example, in areas experiencing increased demand for housing and rising prices, research indicates declining housing affordability as well as a growing risk of displacement for lower-income residents.[4]

For this analysis, IHS used parcel-level data on 1 to 4 unit property sales activity and geospatial techniques to derive a granular assessment of neighborhood-level prices in 2012 and 2016 relative to surrounding areas and to the City of Chicago as a whole. Using these data, IHS was able to classify neighborhoods based on current market conditions (high-cost, middle-cost, and low-cost) and how those conditions have changed over time (significantly rising, rising, stable, and declining). Changes in home values were classified based on standard deviations from the average level of change experienced in the City of Chicago. The geospatial techniques and the methodology to develop typologies are discussed in greater detail in the technical appendix.

Overview of housing market analysis results

The housing market analysis revealed a series of distinct market types based on current conditions and recent change in house prices in the City of Chicago. Markets that had the greatest price changes between 2012 and 2016 tended to be lower- or moderate-cost markets, while markets that had the highest cost had largely stable prices during this period, experiencing only modest price increases. While most census tracts saw some price gains during this period, three percent of census tracts saw price declines. The majority of these declining census tracts were in markets where house prices in 2016 were lower-cost. Conversely, low-cost markets also comprise the largest share of census tracts that saw rapid price increases since 2012: nearly 56 percent of all census tracts that saw a greater than 21.6 percent increase in house prices were in low-cost markets. Figure 3 illustrates the results of the housing market analysis, examining the current conditions observed across markets in 2016 with the trends in house prices in those same areas between 2012 and 2016. Figure 4 maps these patterns.

---

[3] See IHS Cook County House Price Index: Description of IHS Hedonic Data Set and Model Developed for PUMA Area Price Index, May 2015
[4] Joint Center for Housing Studies, "America's rental housing: Expanding options for diverse and growing demand" (December 9, 2015); Newman, K. and Wyly, E. K. "The Right to Stay Put, Revisited: Gentrification and Resistance to Displacement in New York City," Urban Studies 43 (1) (2006):51.

FIGURE 3

Results of the Housing Market Analysis for City of Chicago, 2016

Current Market Conditions in 2016	Change in Sales Prices, 2012 to 2016				Total Census Tracts
	Declining (negative percent change)	Stable (less than 9.5 percent change)	Rising (9.5 to 21.6 percent change)	Significantly Rising (greater than 21.6 percent change)	
High-cost	3	219	84	5	311
Moderate-cost	4	99	133	37	273
Low-cost	17	53	83	53	206
Total Census Tracts	24	371	300	95	790

FIGURE 4

Map of Results of the Housing Market Analysis for City of Chicago, 2016

CHICAGO COMMUNITY AREA

1 Rogers Park	40 Washington Park
2 West Ridge	41 Hyde Park
3 Uptown	42 Woodlawn
4 Lincoln Square	43 South Shore
5 North Center	44 Chatham
6 Lake View	45 Avalon Park
7 Lincoln Park	46 South Chicago
8 Near North Side	47 Burnside
9 Edison Park	48 Calumet Heights
10 Norwood Park	49 Roseland
11 Jefferson Park	50 Pullman
12 Forest Glen	51 South Deering
13 North Park	52 East Side
14 Albany Park	53 West Pullman
15 Portage Park	54 Riverdale
16 Irving Park	55 Hegewisch
17 Dunning	56 Garfield Ridge
18 Montclare	57 Archer Heights
19 Belmont Cragin	58 Brighton Park
20 Hermosa	59 McKinley Park
21 Avondale	60 Bridgeport
22 Logan Square	61 New City
23 Humboldt Park	62 West Elsdon
24 West Town	63 Gage Park
25 Austin	64 Clearing
26 West Garfield Park	65 West Lawn
27 East Garfield Park	66 Chicago Lawn
28 Near West Side	67 West Englewood
29 North Lawndale	68 Englewood
30 South Lawndale	69 Greater Grand Crossing
31 Lower West Side	70 Ashburn
32 Loop	71 Auburn Gresham
33 Near South Side	72 Beverly
34 Armour Square	73 Washington Heights
35 Douglas	74 Mount Greenwood
36 Oakland	75 Morgan Park
37 Fuller Park	76 O'Hare
38 Grand Boulevard	77 Edgewater
39 Kenwood	

Displacement risk typology

About the method

According to available research, a number of socioeconomic and demographic factors are indicators for displacement risk in a rising-cost environment. These factors include high concentrations of renters, low- and moderate-income households, cost-burdened households, households with large families, and seniors.[5] Additionally, a number of neighborhood- and location-specific factors may anticipate future, increased demand for housing and indicate increased risk of lost affordability. These include adjacency to stronger housing markets, strong access to public transit service, historical architecture, planned large public investments, and upward trends in sales prices.[6]

Recently, a growing number of studies have focused on ways to analyze data associated with lost affordability and displacement risk to help policymakers be more proactive in preventing displacement. While these analyses are not predictive, they represent a critical starting point for assessing conditions and needs at the neighborhood level. Many of these studies use scoring rubrics to classify neighborhoods at high risk for displacement.[7] For example, in its study on gentrification and displacement, the City of Portland used a number of variables associated with vulnerability to housing displacement, demographic changes, and housing market appreciation to classify census tracts at different stages of gentrification. Portland has since used these data to determine where displacement needs to be addressed immediately, what neighborhoods need to be observed for ongoing monitoring, and where and how to mobilize resources that could mitigate displacement.[8]

Building on IHS's work documenting changes to the housing market along The 606, IHS created a census tract-level typology of displacement pressures. For this analysis, IHS utilized the results of the demographic and socioeconomic segmentation and housing market studies to identify census tracts where 1) the underlying population consists of a high proportion of individuals or households at high risk for displacement and 2) rising house prices may indicate a shifting demand for housing that may lead to lost affordability.

Overview of displacement risk typology

According to available research, areas with rapid price appreciation likely increase pressures on the affordability of housing for vulnerable populations and may lead to displacement.[9] In order to identify neighborhoods with current or potential future risk for displacement, IHS identified three market types from the segmentation model that had high concentrations of vulnerable populations, high shares of renters, low- and moderate-income households, households with large families, and seniors. Within these vulnerable clusters, IHS identified census tracts with three types of housing market conditions that indicate current or potential lost affordability risk: high-cost markets where prices are rising; moderate-cost markets where prices are rising; and low-cost markets where prices are rising.

Each of these markets have different drivers of demand, different current conditions, and therefore different risks associated with displacement. For high-cost areas, displacement pressure is likely coming from higher-

[5]Bates, L. K. (2013). Gentrification and displacement study: Implementing an equitable inclusive development strategy in the context of gentrification (pp. 1-95, Tech.). Portland, OR: City of Portland Bureau of Planning and Sustainability.; U.S. Department of Housing and Urban Development. (2017). Worst case housing needs 2017 report to Congress (Rep.).
[6] Kennedy, M., and Leonard, P. (2001, April). Dealing with neighborhood change: A primer on gentrification and policy choices (Rep.).; Zuk, M., and Chapple, K. (2015). Case studies on gentrification and displacement in the San Francisco Bay Area (Rep.).
[7]Zuk, 2015; Bates, 2013; and Goetz, E. G., Damiano, A., & Hicks, J. (2015). On racially concentrated areas of wealth (Paper).
[8]Bates, 2013
[9]Kennedy, 2001; Bates, 2013

income households and investors targeting the higher-income market, while pressure in more affordable areas is likely due to investors seeking value in potentially rising markets. For lower-cost areas, rising prices may indicate a stabilizing market, and long-term disinvestment is likely a central concern and driver of neighborhood change rather than gentrification.

The results of the segmentation study show high levels of vulnerable populations found along the lakefront in the otherwise economically diverse Cluster 1 and in more concentrated low- and moderate-income neighborhoods on the west and southwest sides of Chicago in Clusters 5 and 6.[10] Despite sharing the demographic and socioeconomic characteristics of vulnerability to displacement due to lost affordability, these communities have a range of housing market conditions. Figure 5 shows that the combined results of the housing market and segmentation studies indicate that in 2016, nearly 70 percent of census tracts in Cluster 1 were high-cost compared to only 20.7 percent of census tracts in Cluster 5 and 2.2 percent of census tracts in Cluster 6. Conversely, 18.5 percent of census tracts in Cluster 5 and nearly 79 percent of census tracts in Cluster 6 were low-cost in 2016 compared to just 1.9 percent of census tracts in Cluster 1.

FIGURE 5

Results of the Housing Market Analysis for Vulnerable City of Chicago Submarkets, 2016

| | Current Market Conditions in 2016 | Change in Sales Prices, 2012 to 2016 | | | | Total Census Tracts |
| | | | | *Rising Prices* | | |
		Declining (negative percent change)	Stable (less than 9.5 percent change)	Rising (9.5 to 21.6 percent change)	Significantly Rising (greater than 21.6 percent change)	
Cluster 1	High-cost	1.0%	53.3%	15.2%	0.0%	69.5%
	Moderate-cost	0.0%	13.3%	10.5%	4.8%	28.6%
	Low-cost	0.0%	1.9%	0.0%	0.0%	1.9%
	Total Census Tracts	1.0%	68.6%	25.7%	4.8%	100.0%
Cluster 5	High-cost	0.0%	3.3%	15.2%	2.2%	20.7%
	Moderate-cost	1.1%	14.1%	35.3%	10.3%	60.9%
	Low-cost	0.0%	2.2%	13.6%	2.7%	18.5%
	Total Census Tracts	1.1%	19.6%	64.1%	15.2%	100.0%
Cluster 6	High-cost	0.0%	0.6%	1.1%	0.6%	2.2%
	Moderate-cost	0.0%	1.7%	10.1%	7.3%	19.0%
	Low-cost	7.3%	20.1%	25.7%	25.7%	78.8%
	Total Census Tracts	7.3%	22.3%	36.9%	33.5%	100.0%

[10] While Cluster 3 has many features associated with vulnerability to displacement such as high levels of moderate-income households and seniors, IHS did not classify this cluster as vulnerable due to the high share of homeownership in Cluster 3.

While a large share of Cluster 1 census tracts are high-cost, nearly 70 percent are price-stable, meaning that prices between 2012 and 2016 increased below the average increase seen Citywide. Figure 5 shows that roughly 30 percent of census tracts in Cluster 1 are experiencing rising or rapidly rising prices, and these census tracts are split between areas where prices are currently moderate-cost and where prices are already high.

Conversely, a much larger share of census tracts in Clusters 5 and 6 experienced heightened or rapid price appreciation since 2012: 79.3 percent of census tracts in Cluster 5 and 70.4 percent of census tracts in Cluster 6 experienced rising or rapidly rising prices during this period. While both Clusters 5 and 6 saw the types of price gains that may indicate lost affordability risk, those gains occurred in very different types of markets. For example, rising-cost markets in Cluster 5 are typically moderate-cost while rising-cost markets in Cluster 6 are low-cost.

These differing market conditions across Clusters are likely connected to various attributes surfaced in the segmentation model that tend to limit or attract investment. For example, high levels of violent crime may deter investment, while high concentrations of properties that are less susceptible to turnover such as large rental buildings or subsidized units, are likely to slow market transitions.[11] Similarly, areas dominated by a more easily transacted housing stock with single family homes and 2-to-4 unit buildings, may be more susceptible to investment pressures and rapid turnover than other clusters. IHS will continue to examine conditions and factors that impact neighborhood change in future analysis.

For the development of its displacement risk typology, IHS focused on identifying vulnerable neighborhoods undergoing rapid price appreciation where increased prices may lead to lost affordability and potential displacement. Figure 6 maps census tracts with rising costs that are in vulnerable clusters, separately, by current housing market conditions in 2016. It shows an array of neighborhoods with rising costs and large shares of populations vulnerable to displacement in a rising-cost environment, in areas currently classified in 2016 as high-cost, moderate-cost, and low-cost.

High-cost, rising prices - These are high-value markets with recent, strong increases in demand for housing that are thought to be highest risk of displacement, as high costs and rapid price appreciation will likely put significant pressure on housing costs for vulnerable households. These communities likely require immediate interventions to preserve affordability and include a small handful of neighborhoods surrounding The 606 in Logan Square and Humboldt Park and neighborhoods surrounding the medical district on the Near West Side. High-cost areas where prices are rising but at a slightly lower rate include a larger number of communities, and include parts of west Logan Square, Avondale, Irving Park, Albany Park, and Edgewater. These areas likely require thoughtful preservation strategies that are effective in higher cost markets.
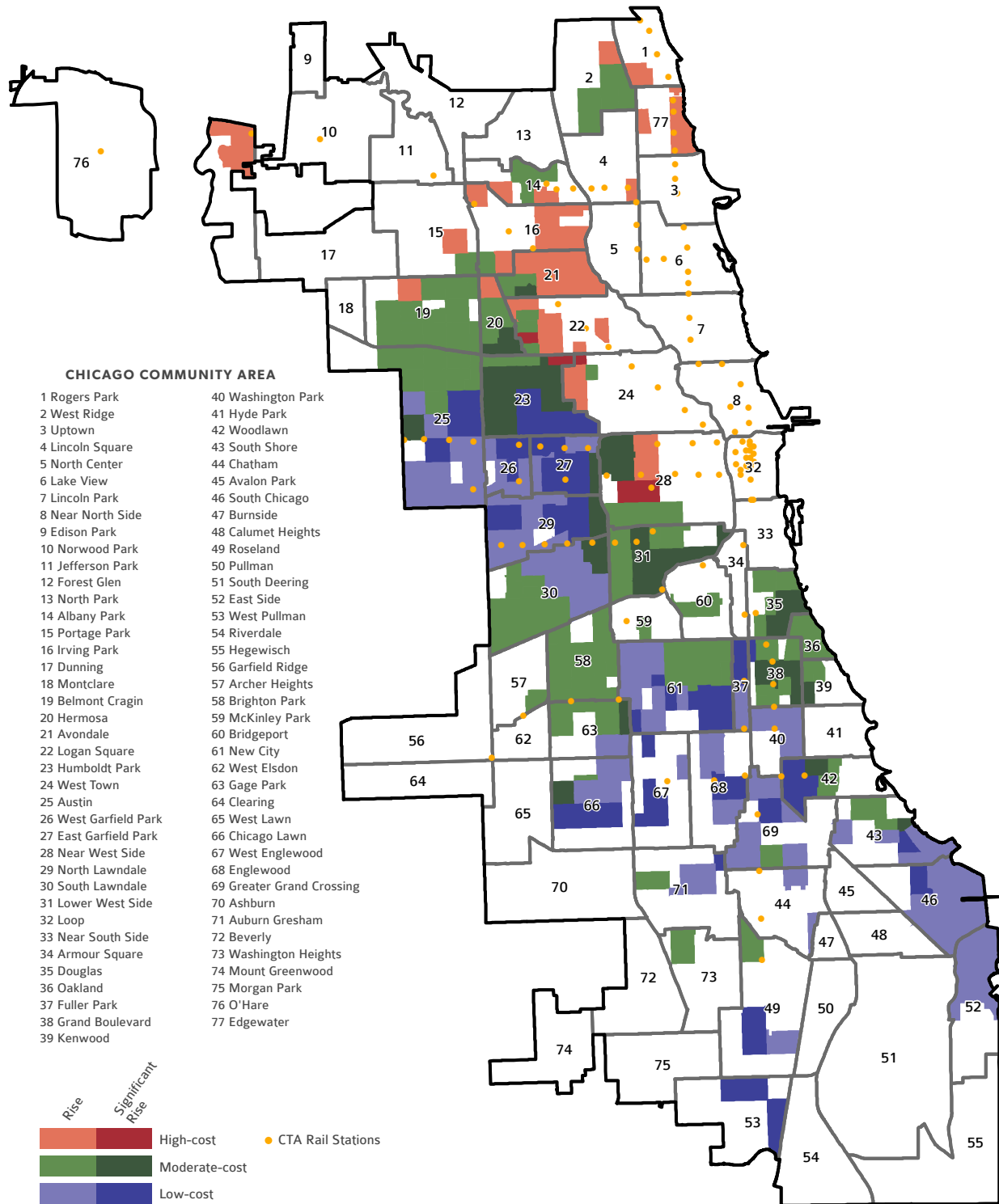
Moderate-cost, rising prices - A large number of vulnerable communities that are experiencing rising prices are located in affordable markets scattered across the west and near south sides of the City of Chicago. If these areas are also near stronger real estate markets or amenities, this could lead to more rapid shifts in future housing demand and increased pressure on housing costs for vulnerable populations. Depending on proximity pressures, the displacement risk may be less immediate and the options for preserving affordability more varied. Neighborhoods with this risk characteristic where prices are rapidly rising include much of Humboldt Park and the Lower West Side, and parts of Douglas, Oakland, South Lawndale, and Woodlawn that are well-served by rapid transit.

Low-cost, rising prices - Many vulnerable communities with rising costs are located in more challenged markets on the west and south sides of the City, where housing costs are low but where prices are increasing. While there is likely a low, short-term risk of displacement due to rising costs, a neighborhood near a high-value or rapidly rising market or another amenity may be at risk for land speculation. Neighborhoods with these risk features and rapidly rising prices include parts of East Garfield Park, North Lawndale, Humboldt Park, and west Woodlawn.

[11] Case, K. E., and Mayer, C. J. (1996). Housing price dynamics within a metropolitan area. *Regional Science and Urban Economics*, 26, 3, 387–407.; Lynch, A., and Rasmussen, D. (2001). Measuring the impact of crime on house prices. *Applied Economics*, 33, 15, 1981-1989.; Lacoe, J., and Bostic, R. W. (2017). Crime and private investment in urban neighborhoods (pp. 1-45, Working paper).; Tita, G., Petras, T., and Greenbaum, R. (2006). Crime and residential choice: A neighborhood level analysis of the impact of crime on housing prices. *Journal of Quantitative Criminology*, 22, 4, 299-317.; National Apartment Association. (2016). 2016 NAA Survey of Operating Income & Expenses in Rental Apartment Communities

FIGURE 6

Vulnerable City of Chicago Submarkets with Rising Sale Values, 2016

CHICAGO COMMUNITY AREA

1 Rogers Park	40 Washington Park
2 West Ridge	41 Hyde Park
3 Uptown	42 Woodlawn
4 Lincoln Square	43 South Shore
5 North Center	44 Chatham
6 Lake View	45 Avalon Park
7 Lincoln Park	46 South Chicago
8 Near North Side	47 Burnside
9 Edison Park	48 Calumet Heights
10 Norwood Park	49 Roseland
11 Jefferson Park	50 Pullman
12 Forest Glen	51 South Deering
13 North Park	52 East Side
14 Albany Park	53 West Pullman
15 Portage Park	54 Riverdale
16 Irving Park	55 Hegewisch
17 Dunning	56 Garfield Ridge
18 Montclare	57 Archer Heights
19 Belmont Cragin	58 Brighton Park
20 Hermosa	59 McKinley Park
21 Avondale	60 Bridgeport
22 Logan Square	61 New City
23 Humboldt Park	62 West Elsdon
24 West Town	63 Gage Park
25 Austin	64 Clearing
26 West Garfield Park	65 West Lawn
27 East Garfield Park	66 Chicago Lawn
28 Near West Side	67 West Englewood
29 North Lawndale	68 Englewood
30 South Lawndale	69 Greater Grand Crossing
31 Lower West Side	70 Ashburn
32 Loop	71 Auburn Gresham
33 Near South Side	72 Beverly
34 Armour Square	73 Washington Heights
35 Douglas	74 Mount Greenwood
36 Oakland	75 Morgan Park
37 Fuller Park	76 O'Hare
38 Grand Boulevard	77 Edgewater
39 Kenwood	

Rise / Significant Rise

- High-cost
- Moderate-cost
- Low-cost
- CTA Rail Stations

Technical appendix

About the data used in this analysis

For the segmentation analysis, data were collected from four sources: 2015 5-year sample estimates from the US Census Bureau's American Community Survey (ACS); data on violent crime from the City of Chicago data portal; data from the IHS Data Clearinghouse on the housing stock; and data from the 2009-2016 Department of Housing and Urban Development Picture of Subsidized Households. Additionally, data from the 2010 Decennial Census and data from the 2000 Decennial Census processed using the Brown University US 2010 project algorithm were used in post-analysis. The City of Chicago was studied at the census-tract level, with census tracts defined by the year 2010 Decennial Census; a total of 796 census tracts were analyzed.

Data for this study were collected to analyze a number of topics related to housing demand and supply and the vulnerability of underlying populations, including variables associated with current and changing demographic and socioeconomic conditions, housing affordability, and the housing stock. These data included information on population change, income level, household size, age, housing tenure, rents and home values, cost burden, vacancy, educational attainment, density, housing stock age and type, poverty status, and crime. IHS excluded data on race and ethnicity so that these features would not influence the clustering pattern and could be analyzed separately.

For the analysis of current market patterns and trends, IHS utilized parcel-level data on sales activity and property characteristics for one-to-four unit properties from its Data Clearinghouse for 2012 and 2016. These data are sourced from administrative records obtained from the Cook County Recorder of Deeds via Property Insight and the Cook County Assessor, and include variables related to sales price and unit count, property class, and building square footage. Sales price data were normalized by the square footage of the property, analyzed at the raster level using geospatial interpolation techniques, and aggregated at the census tract level. Census tracts were defined by the year 2010 Decennial Census. While data were collected for 796 census tracts, only 790 census tracts had sufficient data for analysis. Additionally, IHS used spatial data from the City of Chicago Data Portal on zoning districts to adjust the analysis for non-residential land uses.

Market segmentation analysis

Data preprocessing, Normalization of original data values

In clustering applications, a typical preprocessing step is to standardize variables so that all data are transformed to a comparable range of value. This is because variables measured at different scales will likely skew an analysis, where a variable with a larger range might outweigh variables with smaller ranges. To correct for this, the following transformations were applied:

- Each count variable was converted to percentages ranging in the [0,1] interval.

- Each continuous variable with a dollar amount such as median household income, home value or contract rent was converted to a new variable in the [0,1] range using a Min-Max Scaling. Year 2000 variables were also adjusted for inflation.

- Additionally, variables describing changes between year 2015 and year 2000 were computed by subtracting the percentage values in 2015 from the percentage values in 2000 and/or subtracting the inflation adjusted year 2000 amounts from year 2015 amounts.

Data preprocessing, Methods for standard errors

The Census ACS estimates are based on a sample and as a result, they may be affected by high levels of sampling variability. The reliability of each ACS estimate can be analyzed using the published margin of error that is based on a 90-percent confidence level.[12] The margin of error (MOEs) measures the variation in the random samples due to chance.

A commonly used technique to decide whether a certain ACS variable estimate is reliable employs the coefficient of variation (CV) of the sample estimate. The coefficient of variation is defined as the ratio between standard error and estimated value, and measures the relative amount of variability associated with the sample estimate. Low CV values indicate more reliable estimates. In line with this criterion, only ACS estimates with CV values below 30 percent were used in this analysis. In order to include certain ACS variables with CV values exceeding 30 percent, IHS followed Census Bureau protocols to create a new derived variable with a reduced and acceptable margin of error. Then the CV of the aggregated estimate was computed to assess its reliability and the new aggregated variable was used in the analysis if the newly computed CV was below 30 percent.

K-Medoids clustering technique, About the method

This analysis uses a K-Medoid technique as the method for defining clusters of census tracts with similar characteristics. K-Medoid is a distance-based partitioning method that divides the set of data points into non-overlapping subsets (or clusters) such that each data point is exactly in one subset. Objects within a subset are more similar to one another and different from the objects in other clusters.

The K-Medoid technique groups data points by calculating their pairwise distance from a central point in each cluster. The central-most point (medoid) of the cluster can be regarded as the representation of that cluster. Each data point is then assigned to the closest medoid, and the collection of points assigned to a medoid forms the associated cluster. Extensive discussion of this technique can be found in the textbook by L. Kaufman and P.J. Rousseeuw.[13] K-medoid clustering was chosen for this analysis as it can better handle the variation and outliers

---

[12]U.S. Census Bureau (2008). A compass for understanding and using American Community Survey data: What general data users need to know. Washington DC: U.S. Government Printing Office https://www.census.gov/content/dam/Census/library/publications/2008/acs/ACSGeneralHandbook.pdf
[13]Kaufman, L. and Rousseeuw, P. J. (2005). Finding groups in data: An introduction to cluster analysis. Wiley.

present in housing data utilized for the study.[14] The analysis was computed using the PAM implementation in the R "cluster" package. Eight clusters were created using the Euclidian distance measure.

K-Medoids clustering technique, Choosing the number of clusters "k"

One major challenge among clustering methodologies is the need to pre-select an appropriate number (k) of clusters. The intended use of the final clustering results can cause additional complexity. If there are too few clusters the segmentation is coarse, and results in broad, non-specific clusters. With too many clusters, clusters are differentiated by very small differences among variables, and it becomes difficult to characterize the clusters. One common quantitative approach to choosing the appropriate number of clusters is to cluster the data multiple times and choose a different number of clusters each time. An internal validity metric measuring the quality of the clustering results is recorded for each trial and the optimal k is selected according to some criterion specific to the chosen metric.

Silhouette width is a common internal validity measure for clustering and has been shown to be robust when applied to many clustering algorithms.[15] To choose the appropriate number of clusters for this study, silhouette width was recorded for values of k from 2 to 10. Six clusters were selected as they were associated with both a narrow silhouette width and an acceptable level of granularity for the intended use.

K-Medoids clustering technique, Qualitative testing

Clustering seeks to create useful, understandable, and insightful groupings. Considering these goals, qualitative evaluations of cluster quality are also relevant. For this study, mapping and evaluation of geographic patterns and trends verified that the algorithms produced clusters with merit by assessing whether clusters made sense intuitively and accurately reflected the observed characteristics of the areas.

Clustering results were analyzed using a three-step process. First, silhouette distances were computed as a quantitative assessment of clusters quality. Census tracts and associated clusters were then mapped to determine whether the results were consistent with the observed characteristics in the region. Finally, the values for each variable included in the segmentation were compared among clusters to identify significant differences among clusters and to descriptively characterize each cluster. The results were further refined through meetings with project partners, resulting in the final housing market segmentation results presented in this document.

---

[14] Park, H. S. and Jun, C. H. (2009). A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl., vol. 36, no. 2 PART 2, pp. 3336–3341.
[15] Vendramin, L. Campello, R. and Hruschka, E. (2010). Relative clustering validity criteria: A comparative overview. Stat. Anal. Data Min., vol. 3, no. 5.

Housing market analysis

Data preprocessing, Normalization of sale values

Data for this analysis were sourced from administrative records from the Cook County Recorder of Deeds and the Cook County Assessor and originated at the parcel level. Data on sales prices were cleaned for irregularities and multiple transactions were deduplicated to retain the highest transaction within a year. Sales data were normalized using the square footage of the improvement based on property characteristics from the Cook County Assessor. Data were further standardized as statistical analyses work best when the histogram of the data is sorted into a symmetric, unimodal distribution and data on price per square foot create a skewed histogram, where there is a much larger number of extremely low value sales compared to extremely high value sales. To correct for this and create a histogram much closer to a symmetric distribution, the natural log of the price per square foot was calculated and used for this analysis.

Kriging interpolation technique, About the method

This analysis uses a kriging interpolation technique as the method for classifying the value range of sales activity to census tracts. Interpolation is a geospatial method of analysis that predicts values across a study area from a limited number of sample data points. Kriging is a specific type of interpolation that explores statistical relationships among measured data points and uses a fixed number of nearby data points to predict a value for every location.

The kriging interpolation technique predicts values by first running a statistical analysis to determine the distance at which the value of a data point no longer influences the value of another. After this step, a computation is run using a fixed number of data points closest to each location. Data points located closer to a location are weighted more highly than points farther away. The distance determined by the statistical analysis in the first step is accounted for in this computation. The results of the interpolation creates an image of raster cells, where each cell is assigned a value based on the analysis. Kriging interpolation was chosen for this analysis because of its ability to create a nuanced, localized analysis despite strong differences in home values across the City of Chicago. The analysis was computed using the method of Ordinary Kriging, located in the Spatial Analyst toolbox in ArcMap 10.5.

Further, the kriging interpolation technique was executed for both 2012 and 2016 in order to calculate the change of the average sales value of a location between the two points in time. The values all the raster cells in both years were calculated to create a map displaying percent change in sales value throughout the city. The analysis for this step was computed using the Raster Calculator tool, located in the Spatial Analyst toolbox in ArcMap 10.5.

Kriging interpolation technique, Aggregation of rasters to census tracts

A challenge with the kriging interpolation technique is that the resulting map is made up of raster cells, which are difficult to compare with the segmentation technique which is computed at the census tract level. A census tract, especially one with a large area, may contain both high and low values that would be difficult to classify for this analysis. To create a classification for each census tract, values were aggregated using the following transformations:

- Each raster cell was converted to points using Raster to Point, located in the Conversion Tools toolbox in ArcMap 10.5.
- A layer containing areas of non-residential land use larger than 500,000 square feet was created to eliminate value predictions in large non-residential areas; points that fell into these regions were eliminated from the analysis.

- Remaining points were aggregated by census tract by calculating the mean value of all points located within each census tract's boundaries.

Kriging interpolation technique, Qualitative testing

Kriging is a robust statistical prediction tool, however, qualitative evaluations were vital for verifying the localized nuances presented by the analysis. For this step, IHS tested the kriging results for alignment with staff knowledge, conversations with project partners, and ground truthing by comparing observed physical characteristics of housing with the averaged value presented in the analysis.

Another challenge for the analysis was to ensure that the typology and color gradation chosen to represent our results also characterized real values with meaning. This resulted in an iterative process of exploring various classification methods and assessing the interpretation of each method. The final typology choices, listed above, allowed for a clean description of the sales analysis, an optimized grouping of change between two points in time, and a method of analysis that could continually be updated and compared with past points in time.

Typology development, About the method

Creating a classification structure where sales values were both nuanced and representational when compared to the rest of the City of Chicago was a key factor for creating this analysis. After aggregating the kriging interpolation results from IHS sales data into census tracts, the aggregated values were distributed into seven separate classes by equal intervals, where the top category represented approximately the highest 14 percent of sales values for the entire range of values in the given year, and so on. These were further grouped into three for easier visualization. In order to determine how these classes would best be grouped, IHS normalized these ranges by taking the inverse log of the range of values and multiplying by 1500 to capture the dollar value of a 1500 square foot property. From this information, we consolidated our classes into the following typology:

High-cost: Greater than $265,000
Middle-cost: Between $80,000 and $265,000
Low-cost: Less than $80,000

The aggregated kriging results of percent change between 2012 and 2016 were distributed into four classes by Jenks natural breaks. This classification method divides the range of data so that the variation of values within the same class is minimized for the given number of classes and maximized between classes. These classes were then placed in the following typology:

Decline: Negative percent change
Stable: Between 0 and 9.5 percent change
Rise: Between 9.5 and 21.6 percent change
Significant Rise: Greater than 21.6 percent change

For the entire City of Chicago, the average change in sales value was 11.3 percent and the median change was 9.5 percent.

INSTITUTE FOR
HOUSING STUDIES AT
DE PAUL UNIVERSITY

DE PAUL UNIVERSITY

1 East Jackson, Suite 5400
Chicago IL 60604
312-362-7074

housingstudies.org
@housingsudies