Data Shop

Chicago Multifamily Market Characterization: Developing a Comprehensive Picture of the Multifamily Housing Landscape

Rachel Scheu
Margaret Garascia
Elevate Energy

Abstract

A data-driven description of a community's housing stock can help identify community needs and inform decisionmaking regarding energy efficiency and other types of programs. This article presents the data and methods used in an analysis characterizing the multifamily building stock in Chicago, which segmented Chicago's multifamily buildings by age, size, construction type, and energy use. Conducting this analysis presented several thorny data challenges: building-level data are not collected in any central location; in Chicago and many other cities, the local property assessor has the most complete data of this kind, but the data are compiled for the purpose of tax assessment and not for the purpose of population-level building segmentation; and many disparate data sets must be combined with assessor data into a cohesive whole, presenting difficulty in matching, cleaning, and determining the appropriate level of granularity. This article describes a multifamily market characterization study in Chicago for which different data sources were merged for the analysis; presents a general methodology that could be used by other

Abstract (continued)

cities or program implementers; and discusses insights about the Chicago multifamily market. Identifying and locating geographic concentrations of certain building types enable more precise targeting for energy, housing, and other building programs.

Motivation

The analysis described in this article is part of a series of multifamily research studies, which were created in partnership with the Energy Efficiency for All (EEFA)[1] project, which has the goal of reducing energy consumption to support the larger goal of maintaining affordable housing. EEFA's aim is to encourage electric and gas utilities to spearhead programs that capture all cost-effective energy efficiency within the affordable multifamily housing sector, significantly benefiting low-income families, building owners, and utilities.

The Natural Resources Defense Council and other organizations have indicated a need for market characterization studies to document the building types, ownership structures, housing subsidy characteristics, and energy use characteristics of multifamily buildings at a local level. They recognize that in many cases municipalities are better positioned than states, public utility commissions, or utilities to design and implement locally tailored energy efficiency programs that meet city climate and sustainability goals. Local data could be used by municipal policymakers to complement utility energy efficiency potential studies, or the methodology can serve as a guide for municipalities to identify and characterize the broad and complex multifamily market. Market segmentation is needed because a one-size-fits-all approach to energy efficiency is ineffective in a diverse multifamily sector.

The approach presented here is focused on the affordable multifamily market in Chicago; similar characterizations are being done in New York City and Los Angeles (Yancy et al., 2015). The methodology is also broadly applicable to the other building sectors. For example, Elevate Energy used this methodology, augmented by extensive data transfer processes and Python scripting, to assist the City of Chicago in the implementation of its commercial building benchmarking and disclosure ordinance, as well as a countywide single-family market characterization.

Data Sources for the Chicago Multifamily Characterization

The authors faced expected issues of data availability, completeness, consistency, and granularity. These challenges stemmed from the number of data sources that were used to develop the database of Chicago multifamily buildings and the fact that each data source was originally designed and maintained for a purpose other than developing the multifamily characterization.

---

[1] EEFA is a joint effort of the Natural Resources Defense Council, the National Housing Trust, the Energy Foundation, and Elevate Energy. The mission of EEFA is to make multifamily homes healthier and more affordable through energy and water efficiency and access to clean energy, benefiting millions of Americans living on limited incomes.

Exhibit 1 includes the 13 data sources leveraged, shown in order of the number of observations. The primary data source used, which was from the Cook County Assessor, provided more than 173,000 observations. Although incomplete for some properties, the data source included variables for property age, units, stories, construction material, and assessed value. Critically, each observation was associated with a unique 10-digit number called a property identification number (pin10).

Exhibit 1

Data Sources Used in the Chicago Multifamily Market Characterization

Data Set Name	Description	Number of Buildings or Properties	Geography and Granularity	Public, Fee-Based, or Private
Cook County Property Assessor data	Residential properties that include the vintage, number of units, and other building characteristics	173,000	County (includes city of Chicago) parcel or building level	Public
Chicago Department of Buildings	Water meters in commercial, residential, and industrial buildings	71,000	City of Chicago building level	Private
Chicago Energy Use 2010	Aggregate electricity and gas use, at the census block level, for particular building types	67,000	City of Chicago census block	Public
Chicago Department of Buildings permit	Building permits for commercial, residential, and industrial buildings	54,000	City of Chicago building level	Public
CoStar	Commercial real estate database with multifamily module	9,000	National building level	Fee-based
ComEd 2013 Smart Meter data	Electricity usage data for ComEd customers with smart meters on the multifamily rate	3,000	Regional Illinois building and utility meter level	Private
Elevate Energy All-Electric database	Multifamily all-electric buildings in Chicago	1,600	Chicago building level	Private
National Housing Preservation Database	Aggregated database of federally subsidized properties	800	National census block	Public
Elevate Energy retrofit program	Energy use and characteristics for buildings that have applied for or completed retrofits	600	Regional Illinois building level	Private
City of Chicago benchmarking	Reported energy data for multifamily buildings > 250,000 square feet	300	Chicago building level	Private
ACS	Ongoing survey that provides demographic data	NA	National census block	Public
RECS	National survey of 12,000 households on energy consumption	NA	National census block	Public
Energy Score Cards/ Bright Power	Energy reporting data for owners of multifamily housing in Chicago	166		Private

ACS = American Community Survey. NA = not applicable. RECS = Residential Energy Consumption Survey.
Note: The ACS and RECS regional and national survey data are reported in aggregate.

To ensure that the database would be complete with respect to the number of multifamily buildings in Chicago, Elevate Energy complemented the assessor data with two sources from the Chicago Department of Buildings. These sources were complete for owner contact information but less complete for other building characteristics. CoStar, a commercial real estate database, contained about 9,000 buildings labeled as multifamily properties in the city of Chicago, many of which tended to be larger and newer properties. These buildings represent about 45 percent of the stock of multifamily buildings with at least five units.

To assess housing subsidy characteristics, Elevate Energy used the National Housing Preservation Database, which contains property-level subsidy information for the Low-Income Housing Tax Credit Program, public housing authority affiliation, and other subsidy information. Ownership structure was estimated using the most recent estimates from the American Community Survey (ACS).

Energy use information was unavailable at the population level. Instead, energy use was estimated using a variety of sources, including Elevate Energy's program data, which cover roughly 600 buildings in Chicago; the City of Chicago data portal, which published 2010 electricity and gas data for particular building types at the census block level; the Residential Energy Consumption Survey, which provides regional estimates on residential energy use; and Bright Power, which shared energy benchmarking data for approximately 150 multifamily buildings in Chicago.

Data Cleaning and Joining

The Cook County Assessor data for multifamily buildings required extensive data cleaning and manipulation. First, duplicates with the same unique pin10 were removed and any associated data fields were appended to the remaining observation. It is possible that these duplicates represented a property that had undergone significant improvements, like an added garage, added units, or renovation, which triggered a new instance in the assessor's data systems. Other causes for the extensive duplication in the pin10 were not immediately obvious.

The assessor data also presented a challenge because some single buildings were divided into multiple parcels. The City of Chicago publishes geospatial building footprint data on its data portal. These footprint data were analyzed in ArcGIS and overlaid with assessor data to identify potential duplication. An example is shown in exhibit 2, in which buildings, shown as shaded outlines, were represented as multiple observations (straight lines), according to the Cook County Assessor. Visual inspection revealed several possible explanations for the discrepancy: large buildings with multiple postal addresses or a mix of commercial and residential space could be represented as distinct assessed properties; newer development had been constructed on more than one parcel but had not been condensed to a single property; parking lots, vacant lots, and other nonbuilding objects that were subject to property tax assessment were included but were not relevant to the multifamily market study.

SQL scripts were employed to address each type of parcel or footprint discrepancy. For example, a large apartment building at a range of addresses was condensed into one observation for a given building footprint, and the associated data were aggregated for that observation. If this methodology resulted in different values for a variable—such as building age—the minimum value was always taken. Vacant or demolished buildings were removed from the data set.

Exhibit 2

Geospatial Building Footprint Data, City of Chicago

Note: Outlines of assessor parcels are overlaid onto footprints of existing buildings (shaded).
Sources: Cook County Property Assessor; city of Chicago geospatial building footprint data

Condominium units presented another duplication issue. To the property assessor, they are unique data entries with their own assessed value and other characteristics. The authors were concerned with the building level, not the property level, however, and therefore condominiums were collapsed to their respective buildings. They were identified as condos using their class code and were aggregated to the building level using the final four digits of their pin10, which were identical for units in the same building. Thus, condo units were aggregated to the building level and the number of units was included as a field.
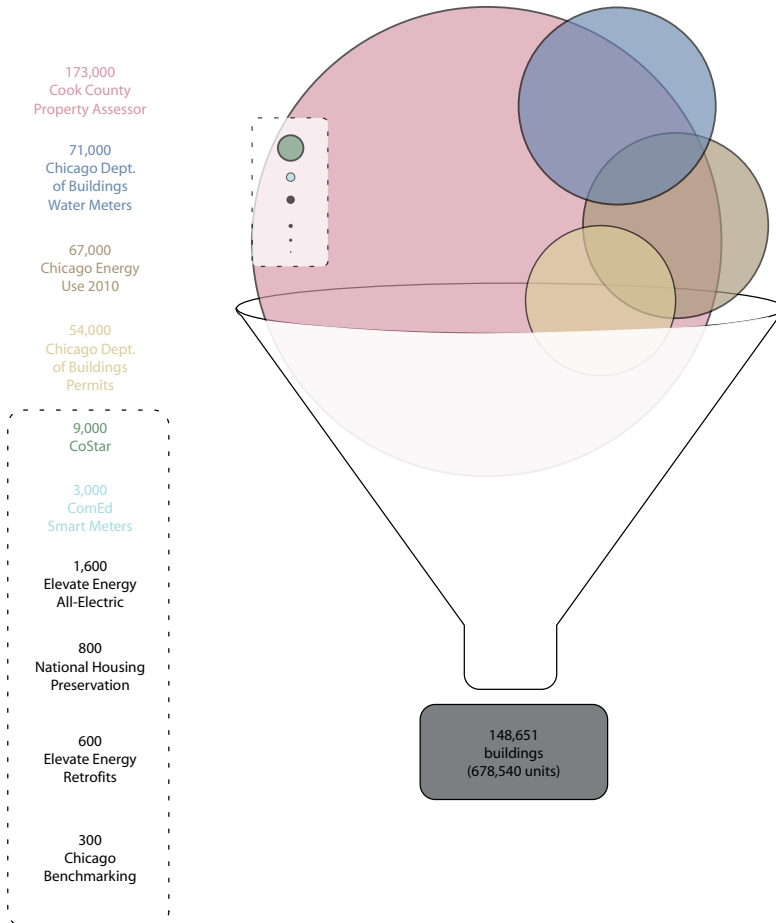
Whenever possible, the assessor data, which served as the primary data source, were joined to other sources via the pin10. For example, CoStar data were joined in this way, and the roughly 1,200 observations in the CoStar multifamily database that did not have matching pin10s in the assessor file were appended to the master data file. The slight mismatch between the assessor and CoStar files is attributed to a lack of standard methodology for classifying large apartment buildings as either residential multifamily or commercial, or as both.

Finally, each observation was geocoded with ArcGIS, using the address to assign a unique geospatial location to each building. Using this information, each building was linked to a census tract, which could be used to join the data from the National Housing Preservation Database, ACS, and other census-based survey data.

Exhibit 3 is a representation of the database.

Exhibit 3

Multiple Data Sources Combined To Build Characterization Database

173,000
Cook County
Property Assessor

71,000
Chicago Dept.
of Buildings
Water Meters

67,000
Chicago Energy
Use 2010

54,000
Chicago Dept.
of Buildings
Permits

9,000
CoStar

3,000
ComEd
Smart Meters

1,600
Elevate Energy
All-Electric

800
National Housing
Preservation

600
Elevate Energy
Retrofits

300
Chicago
Benchmarking

148,651
buildings
(678,540 units)

How To Conduct a Multifamily Market Characterization

The authors wrote a how-to guide for cities, program implementers, and others wishing to conduct their own segmentation study: *Understanding Your Multifamily Building Stock: A Framework for Cities and Municipalities* (Corso et al., 2016a). Following is a summary of the approach, and its intended audiences are cities only starting to identify uses and data sources. This approach could be applied to answer questions about the potential for energy efficiency savings in the building stock or to create a list of covered buildings under a benchmarking ordinance, and it is generalized to facilitate many other applications.

1. Identify a Primary Data Source

To start, cities and municipalities should identify and document all the potentially useful data sets from various data sources and secure authorization to use them. After data sets are identified, it

is critical to identify the set that appears to be the most complete, comprehensive, and accurate and to consider this set as the primary data source. When the primary data set's strengths and weaknesses are well understood, identify secondary data sources to fill in any gaps. Ideally, the secondary data sets can be joined via a unique identifier, rather than an address, because addresses often require extensive cleaning.

The potential data sources include but are not limited to—

- Property tax assessors.

- Municipal open data portals.

- Housing agencies or university research centers.

- Utility energy efficiency studies.

- Energy use surveys.

- Municipal water agencies.

- Private real estate databases.

- Public surveys and data sources.

When choosing a primary data source, it is important to consider the original intended purpose of the data. A city's open data portal with information on building permits, for example, might not be intended to capture every building that exists in the city, because it tracks only those with pending or expired permits. By contrast, the purpose of a county assessor's database is to determine the tax base, and therefore it includes every property within its jurisdiction. Similarly, take note of the strengths and weaknesses of sources relative to one another; the quality and reliability of the building owners' contact information may be stronger in some data sources, because the data were used to create a contact list for a municipal agency. Another data set may lack phone numbers and addresses but include high-quality information about building construction materials and fuel mix.

Consider and document variable definitions as well and be aware that variables that appear to be identical in disparate data sets may have different data definitions, depending on the source of the information. For example, the size of a building measured in square footage might be reported as a distinct value across different data sets. A property assessor calculates square footage based on taxable area of a building, an energy program such as the U.S. Environmental Protection Agency's Portfolio Manager defines square footage as the conditioned heated and cooled areas of a building, and CoStar tracks gross leasable floor area. These different criteria for calculating square footage result in three distinct values for the same concept.

2. Consider the Granularity of Secondary Data Sources

Secondary data sources can include any data sources that are less complete, valid, or granular than the primary data source. For example, in an effort to protect the privacy of individual residents, a city might share energy or water data at the neighborhood level but not at the building level.

Because the data are at the neighborhood level, they cannot be linked to an individual building, although they are valuable to link by neighborhood. Given these constraints, it would not make sense to start a market characterization with such a secondary data source.

Many secondary data sources are less complete or available only in aggregate. Despite these constraints, these data can still complement the primary data source. A less complete source might cover only a particular neighborhood or type of housing. A conclusion can be drawn from this less complete source and extrapolated to a larger population, as long as the methods and assumptions are clearly explained.

3. Understand Privacy and Data Sharing

After primary and secondary data sources are selected, analysts must consider the privacy requirements associated with each. If the ultimate goal of a market characterization is to share a data file with a broad audience, it is paramount to ensure that subsidiary data sources can be shared more widely. Some data-sharing agreements prevent such wholesale sharing of data but enable sharing of aggregated, manipulated, or transformed data.

Building-level energy data in particular pose privacy challenges, because the industry is highly regulated. For example, a city that requires certain types of buildings to benchmark its energy data should take care to understand the data access limitations that may exist.

4. Consult Local Experts

Experts in housing, real estate, and energy policy can be invaluable partners in a market characterization study. Many of them have deep knowledge of the data sets available and the challenges associated with each. They can also assist in framing and communicating findings for various audiences. Advocates for affordable housing and experts in local housing markets are often especially helpful in providing context for working with the subset of affordable multifamily housing in a given market.

Findings

The methods presented previously, when applied to the Chicago multifamily building stock, yielded insights about the size, geographic distribution, and segmentation of residential buildings in Chicago. Previous analyses of Chicago housing, most notably by the Institute for Housing Studies at DePaul University, used assessor data to understand the building stock. This analysis built upon those studies by adding layers of other municipal data, fee-based commercial databases, and energy data.
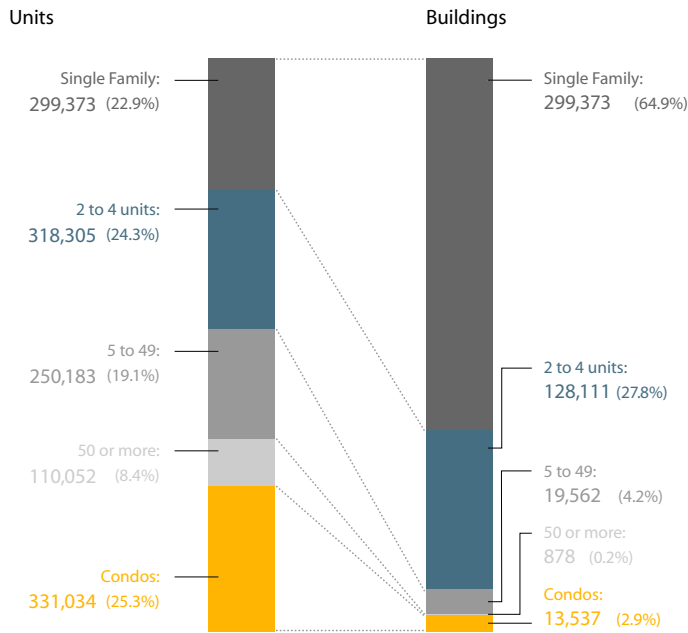
More Chicagoans live in multifamily housing than in any other type of residence. The city of Chicago is home to an estimated 1.3 million housing units, of which roughly 23 percent are single-family homes and the remaining 77 percent are broadly defined as multifamily housing. In other words, more than three-fourths of Chicago's housing units are in some kind of multifamily structure; these structures number nearly 150,000 and occur in every neighborhood in the city. See exhibit 4 for the segmentation of Chicago housing in buildings and units.

Exhibit 4

City of Chicago Residential Housing Stock in Buildings and Units

City of Chicago Residential Stock

Units

Single Family:
299,373 (22.9%)

2 to 4 units:
318,305 (24.3%)

5 to 49:
250,183 (19.1%)

50 or more:
110,052 (8.4%)

Condos:
331,034 (25.3%)

Buildings

Single Family:
299,373 (64.9%)

2 to 4 units:
128,111 (27.8%)

5 to 49:
19,562 (4.2%)

50 or more:
878 (0.2%)

Condos:
13,537 (2.9%)

Source: Elevate Energy analysis of the characterization database, as illustrated in Exhibits 1 and 3

Multifamily housing forms the backbone of affordable housing in Chicago. Nearly 90 percent of rental housing in Chicago is in a multifamily building. The two- to four-flat building is a dominant building type and accounts for 38 percent of rental housing in Chicago. The multifamily building with five or more units accounts for 52 percent of all rental housing in Chicago; rental units are divided into market-rate, subsidized housing, and that which is not subsidized but affordable. As seen in exhibit 5, the number of so-called "naturally occurring affordable housing" units (184,000) is double the number of subsidized housing units (91,000). Between 60 and 70 percent of multi-family housing in Chicago exists in neighborhoods where at least one-half of households earn well below the median income for the Chicago rental market (exhibit 6).

Exhibit 5

City of Chicago Affordable Multifamily (Five or More) Housing Units

	Two- to Four-Unit Buildings		Multifamily (Five or More) Buildings	
	Units	Percent	Units	Percent
Market rate	111,632	30	264,359	49
Unsubsidized affordable	261,502	70	183,860	34
Subsidized	NA	NA	90,747	17
All multifamily housing units	373,149		538,966	

NA = data not available.
Sources: National Housing Preservation Database; American Community Survey

Exhibit 6

Chicago Median Income, by Census Tract

Median Household Income by Chicago Census Tract
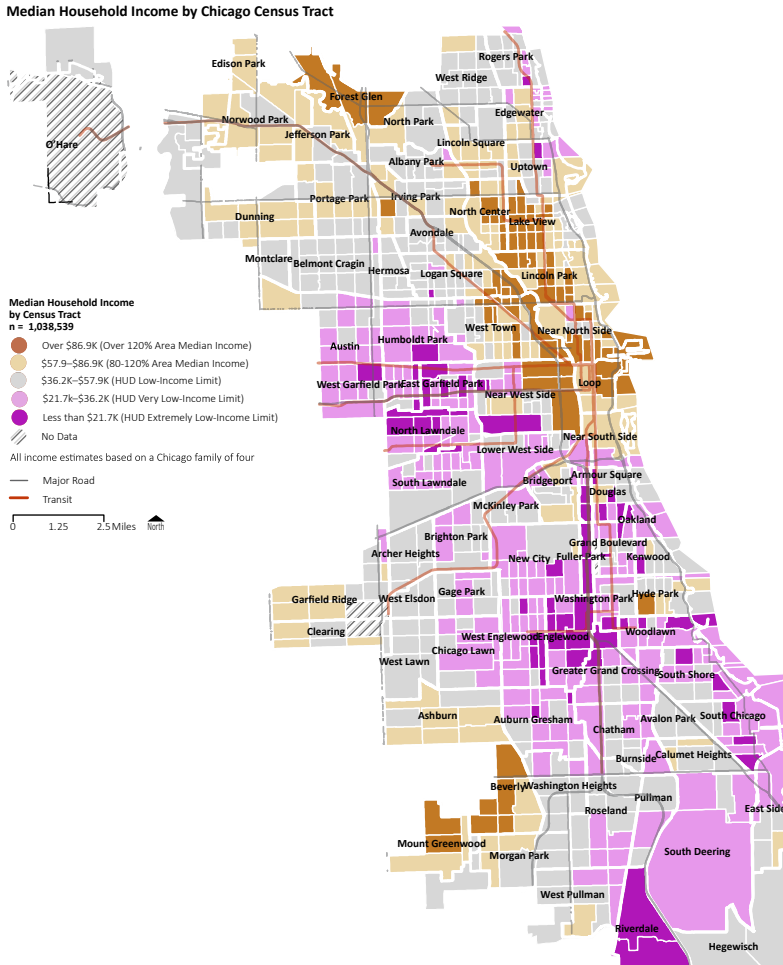
Last Updated: September 2015

HUD = U.S. Department of Housing and Urban Development. n = number of households.
*Source: 2013 American Community Survey 5-year estimates*

Targeting five building segments would capture 86 percent of Chicago multifamily buildings and 94 percent of multifamily units. These segments are—

1. Five-or-more-unit, lowrise, prewar buildings.

2. Five-or-more-unit, lowrise, mid-century buildings.

3. Five-or-more-unit, lowrise, post-1978 buildings.

4. Two- to four-unit, prewar, masonry buildings.

5. Two- to four-unit, prewar, frame buildings.

The largest segment consists of the lowrise prewar building with at least five units, which is also the most common type of multifamily building to undergo energy efficiency retrofit services from Elevate Energy. Exhibit 7 shows the number of multifamily retrofits completed in each Chicago neighborhood contrasted with the size of the total market.

Exhibit 7

Penetration of Elevate Energy's Multifamily Program in Chicago Neighborhoods
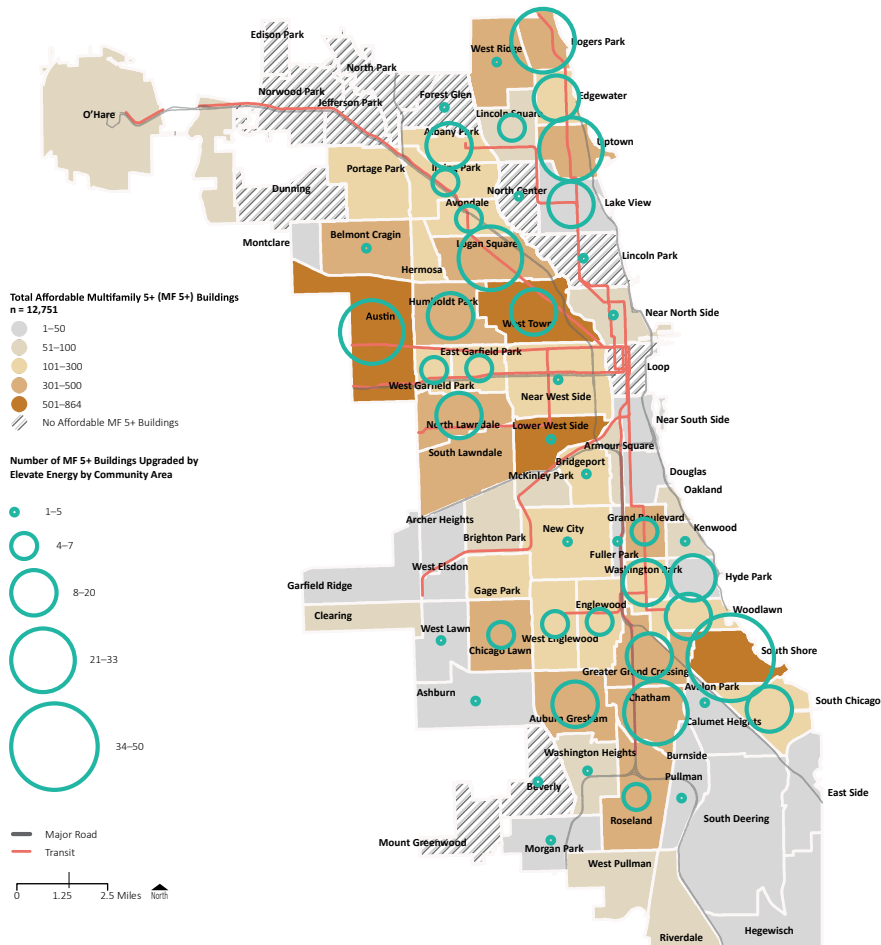
Conclusion

This article presented the data and methods used in an analysis characterizing the multifamily building stock in Chicago, which segmented Chicago's multifamily buildings by age, size, construction type, and energy use. The article described the 13 data sources that were merged to complete the analysis and discussed recommendations for other researchers conducting a similar market characterization. This article was adapted from the forthcoming reports *Segmenting Chicago*

Multifamily Housing to Improve Energy Efficiency Programs (Corso et al., 2016b) and *Understanding Your Multifamily Building Stock: A Framework for Cities and Municipalities* (Corso et al., 2016a), which are directed at municipal policymakers, energy-efficiency program implementers, and utility stakeholders.

These findings are a key step in designing and implementing energy efficiency programs that meet the needs of a diverse set of nearly 150,000 multifamily buildings in Chicago. As more cities and municipalities commit to energy reduction goals and pass legislation such as energy benchmarking and reporting, analysis of local data and subsets of the building stock can provide insights and identify areas of opportunity for energy efficiency programs. By using these local data sets and forming partnerships among policymakers, utilities, and program implementers, cities can continue to lead the way in making our urban areas more livable, sustainable, and economically viable.

Acknowledgments

Authors

Rachel Scheu is Director of Research and Innovation at Elevate Energy.

Margaret Garascia is a senior research analyst at Elevate Energy.

References

Corso, Abigail, Margaret Garascia, and Rachel Scheu. 2016a. *Understanding Your Multifamily Building Stock: A Framework for Cities and Municipalities*. Chicago: Elevate Energy.

———. 2016b. *Segmenting Chicago Multifamily Housing To Improve Energy Efficiency Programs*. Chicago: Elevate Energy.

Yancy, Richard, Ellen Abramowitz, Adam Hinge, Jeffrey Perlman, Conor Laver, and Yetsuh Frank. 2015. *Retrofitting Affordability: Evaluating New York City's Multifamily Building Energy Data for Savings Opportunities*. New York: Building Energy Exchange.