

NEIGHBORHOODS BY NUMBERS:

An Introduction to Finding and Using Small Area Data

by Alan Mallach





**This volume is published and distributed by the:
Center for Community Progress**

111 E. Court Street, Suite 2C-1, Flint, MI 48502
(877) 542-4842

The publication is available in electronic format
via www.communityprogress.net

©2017 All rights reserved.



NEIGHBORHOODS BY NUMBERS:

An Introduction to Finding and Using Small Area Data

by Alan Mallach

CONTENTS

INTRODUCTION	6
---------------------------	----------

PART I: FINDING THE DATA	13
---------------------------------------	-----------

A. National Data Sources 14

1. CENSUS BUREAU DATA	15
The Decennial (10-Year) Census	15
The American Community Survey	15
OnTheMap	17
Other Census Products	18
2. OTHER IMPORTANT FEDERAL DATA	19
HMDA Data	19
US Postal Service Data	20
FBI Uniform Crime Reports	20
Commercial Data Providers	20
PolicyMap	20
Other Data Providers	20

B. Local Data Sources 22

1. DATA COLLECTED BY PUBLIC AGENCIES	23
Property Records	23
Foreclosure Records	25
Crime Records	27
Other Local Data Sources	28
2. LIMITATIONS OF PUBLIC AGENCY DATA	29
Getting the Data	29
What to Watch Out For	30
Converting Point Source into Small Area Data (Geocoding)	32
3. CREATING YOUR OWN DATA	32
Physical Condition Surveys	32
Sample Surveys (Interviews)	35
Informal Information-Gathering Methods	37

CONTENTS (CONTINUED)

PART II: USING THE DATA 39

A. Choosing the Geography 40

B. Identifying What You Want to Know 43

- 1. ANALYZING DATASETS 44
- 2. CREATING A MARKET INDEX 47
- 3. BUILDING A COMMUNITY INFORMATION DATABASE 50

C. Organizing the Data Project 51

- 1. ENSURING YOU HAVE THE RIGHT SKILLS FOR THE PROJECT 52
- 2. GETTING STARTED: THE IMPORTANCE OF COMMUNITY PARTNERS 53
- 3. FINDING DATA PARTNERS 54
- 4. MANAGING THE PROJECT 55

D. Telling Your Story With Data ... 58

- 1. THINKING IT THROUGH 59
- 2. PRESENTING DATA 60

PART III: APPENDICES 69

A. Information Resources 70

B. Glossary 72

ABOUT THE AUTHOR

ALAN MALLACH is a senior fellow at the Center for Community Progress in Washington DC. A city planner, advocate and writer, he is nationally known for his work on housing, economic development, and urban revitalization, and has worked with local governments and community organizations across the country to develop creative policies and strategies to rebuild their cities and neighborhoods. A former director of housing & economic development in Trenton, New Jersey, he currently teaches in the graduate city planning program at Pratt Institute in New York City. He has spoken on housing and urban issues in the United States, Europe, Israel, and Japan, and was a visiting scholar at the University of Nevada Las Vegas for the 2010-2011 academic year. His recent books include *A Decent Home: Planning, Building and Preserving Affordable Housing* and *Bringing Buildings Back: From Vacant Properties to Community Assets*, which has become a resource for thousands of planners, lawyers, public officials, and community leaders dealing with problem property and revitalization issues. He is a member of the College of Fellows of the American Institute of Certified Planners, and holds a B.A. degree from Yale University.



INTRODUCTION

We live in the age of data. This is nowhere more true than in America's cities and towns, where more and more data is becoming available every year that can be used to help local officials, neighborhood organizations and other stakeholders, even though many cities and towns have yet to take full advantage of this information. This data sheds light on many of the measurable indicators of neighborhood vitality, including property conditions, whether a property is owned by a homeowner or an investor, whether the property is in foreclosure, and more.

This data can thus inform plans for neighborhood revitalization or problem property strategies. Understanding what is going on with a building, a block, or a neighborhood is critically important when you have to make strategic decisions, identify priorities, and use limited resources in the most effective way possible. While data won't tell a planner or community developer everything, it can tell a lot, and often offer some surprises.

When we talk about data, we are talking about numbers, or quantitative information, that can be gathered, analyzed, and interpreted to help us understand what

is going on in some part of the world, from international trade or migration statistics to the code violations on a single building. The Youngstown Neighborhood Development Corporation in Youngstown, Ohio, uses data to design neighborhood plans, identifying the most appropriate strategy for each neighborhood. South Bend, Indiana, and Baltimore, Maryland, use data to tackle vacant properties, designing different strategies for different neighborhoods based on their market conditions. Minneapolis and other cities in Minnesota are using data to track problem rental properties, and design strategies to improve landlord stewardship.



Who are data users? Planners, community development practitioners, and others in communities around the United States are working to gather data about their communities and create databases and information systems to help guide their efforts. The data they are pulling together enable them to understand what is going on in a community, including crime patterns, tax foreclosures, vacant properties, and housing market trends. These individuals are all data users.

Every community has many potential data users. They include many people in local government, including city planners, as well as those working on community and economic development, housing inspections, public safety, neighborhood services, and more. They include people in the local nonprofit sector, including community development corporations (CDCs) as well as social service providers. They may include realtors, developers, lenders, and people in other parts of the local business community, as well as residents of the community's neighborhoods, who may be involved in civic associations, block clubs, and crime watch organizations. All of these people may be able to use data to help benefit their block, neighborhood, or community.

Often, however, people don't know which datasets can give them what information, how reliable they are, and where to find them. Many people are intimidated by what they see as the difficulties of obtaining and understanding data. Once they have the data, they may not know how to present it in the most effective way to get their point across.

The purpose of this guide is to make data more accessible and encourage people to use it to make better decisions about the future of their communities. It provides basic information about finding and using three types of information a local data user is likely to need:

- Information about individual properties
- Information about neighborhood housing market conditions
- Information about other neighborhood conditions relevant to neighborhood planning, strategies, and interventions.

As this suggests, the guide focuses on data that has to do with properties, physical conditions, and housing market conditions. All of this data is **spatially defined**; in other words, data that can be directly linked either to a small area, such as a block or census tract, within a larger city, county, or metro, or to a particular spatial location within that area, such as an address or building parcel. We generally refer to data that is available for individual parcels as "**point source data**," and that which is available for areas such as census tracts as "**small area data**." Many of the datasets we discuss are property-specific, such as vacancy or tax delinquency, while others, such as crime incidents, can be pinpointed to specific locations.

The guide is designed to help answer the following questions:

- Where to find the data
- How to extract the data
- What each dataset tells you, and what its limitations are
- How to use these datasets to better understand neighborhood conditions
- How data can be presented

Small area data is particularly important because, even in very small cities, housing and market conditions can vary significantly from area to area, sometimes even from block to block. In order to be most effective, neighborhood revitalization efforts will be tailored based on those small-area

DON'T FORGET THE BIG PICTURE

This guide focuses on small area data, but this doesn't mean that one should not look at the larger picture. Neighborhoods are not autonomous entities, but rise and fall as part of large towns, cities, counties, and regions. Thus, what's happening in the larger city or region, particularly with respect to job and population growth and decline, changing demographics, and changing economic sectors, directly affects every neighborhood in the community.

It is important to keep that bigger picture always in mind.



differences. To understand what's going on in a neighborhood or with a property, however, one must find the data that can allow one to zoom in on the particular spot of interest. That is the data that this guide is about.

When one uses data to answer questions about neighborhoods, it is important to look at both *conditions* and *trends*. Data on conditions provides a snapshot of the state of the neighborhood at present, or as close to the present as the data will allow. Trends measure the neighborhood's trajectory over time. Both are relevant. It is important to know what the homeownership rate is, for example, but it is equally important to know whether it is going up or going down.

While for many people data is interesting in itself, the principal purpose of gathering small area data is to help devise effective strategies and interventions that can reduce blight and problem properties and make neighborhoods better places for people to live or work. Good data can help answer questions that are critical for shaping either overall strategies for a block or neighborhood, or specific interventions for particular properties, such as:

- What would be the most suitable areas in which to carry out a homesteading program?
- Where are existing homeowners starting to give up on their neighborhood, even though it may still appear well-functioning?
- What other risk factors may be present in seemingly well-functioning neighborhoods that should be flagged before they potentially destabilize the area?
- What are likely to be the relative effects of demolishing vacant houses in one neighborhood or on one block compared to another?
- What is the likely market gap (the difference between the cost of rehab and the post-rehab market price) that will have to be filled in order to make rehab successful in a particular neighborhood?
- Which areas are facing the greatest challenges from an increased presence of absentee landlords?

USING STATISTICS

Contrary to what some people believe, *numbers never speak for themselves*. To use data to try to answer these questions, users must employ a variety of statistical methods to analyze and interpret the data. Statistical methods are the procedures one uses to take large piles of data, organize them in useful categories, understand what it means, and be able to make meaningful comparisons between one dataset and another. Without such systems, data is little more than noise.

Some statistical methods such as percentages or medians are straightforward and intuitive, while others are more complicated and may require formal training to understand, let alone apply. It is not realistic to expect users to become statisticians nor to understand sophisticated econometric methods. Users, however, need to understand why and when basic statistical methods are important, and have some idea of what they are about. In the course of this guide, we will try to flag key areas where the use of statistical methods is critical, and provide some idea of what is involved. We will not provide detailed descriptions of the methods or how to apply them. Some good introductory books on statistics are included in the reference section at the end of the guide.



What data one gathers, and how one chooses to use it, will vary depending on one's purpose. There are many different ways one can use small area data:

- Using parcel data to identify patterns and features, such as identifying vacant buildings on otherwise sound blocks or concentrations of absentee property ownership;
- Identifying individual variables or datasets of particular interest and exploring how they play out in different neighborhoods; for example, looking at the variation in house sales prices or the variation in vacant properties from one neighborhood to another;
- Combining different variables into a market index or market segmentation model, which provides a more comprehensive view of market conditions in neighborhoods (descriptions of a number of such models are given on page 49 of this guide);
- Creating a parcel-based neighborhood database, either for internal use by public sector and nonprofit organizations or as an online public-access resource for the community.

These categories are not mutually exclusive, and in fact can and should be combined. The same project can create a market index, pull out key variables to be looked at separately, flag problem properties meeting certain criteria, and create a database to make possible further use in the future. Whatever the precise way in which small area data is being analyzed, presented, and used, we refer to such activities in this guide as **small area data projects**.

The questions shown above are but a handful of those that good neighborhood or property data can help answer, whether for a municipal official, a CDC staffer, or a community activist. At different points in this guide we offer further examples, and try to suggest how different datasets can be used to address different community issues. While this guide can illuminate many ways small area data can be used and interpreted, it is not a comprehensive guide to the many potential uses of small area data, which could easily fill hundreds of pages and still leave many questions unanswered. At the end of the guide, we suggest a number of publications and other resources that can provide users with more information about how data can be used.

THE GUIDE CONSISTS OF TWO MAJOR SECTIONS

I. Finding the Data

This section looks at gathering data from both national and local sources. There are sources of public data that are available for every (or almost every) geographic area in the United States, such as Census data. Most of these data sources, including Census data, are free, but some are what are known as “proprietary” and must be purchased from the company that collects or compiles the data. Except for specific datasets that must be purchased, publicly gathered national data is generally available on the websites of the agencies that collect it. At the local level, state and local governments gather large amounts of data which can be useful for understanding neighborhood and property issues. This data needs to be obtained locally, although in some cases the agency that collects it makes it accessible on the internet. We discuss how organizations can gather their own data to supplement the information they can obtain from public sources.

II. Using the Data

This section describes some of the different ways in which small area data can be used, and looks at the practical considerations involved in doing data analysis or putting a small area data project together. It digs into some of the technical aspects of working with data. It also explores how to tell your story with data. There are many ways data can be presented, depending on the nature of the data and the purpose for which it is being used. These can range from printed reports to interactive maps. We discuss some of the different ways in which data can be presented for the greatest effect.

At the end we provide information on other useful resources, including publications about using data to address property and neighborhood issues, guides to data presentation, and links to good examples of community-based neighborhood and property analysis, and a glossary of terms used in the guide. Particularly valuable resources are highlighted as well in text boxes at different points in the guide.

The guide is designed to be useful for both those who have some background in data analysis and community



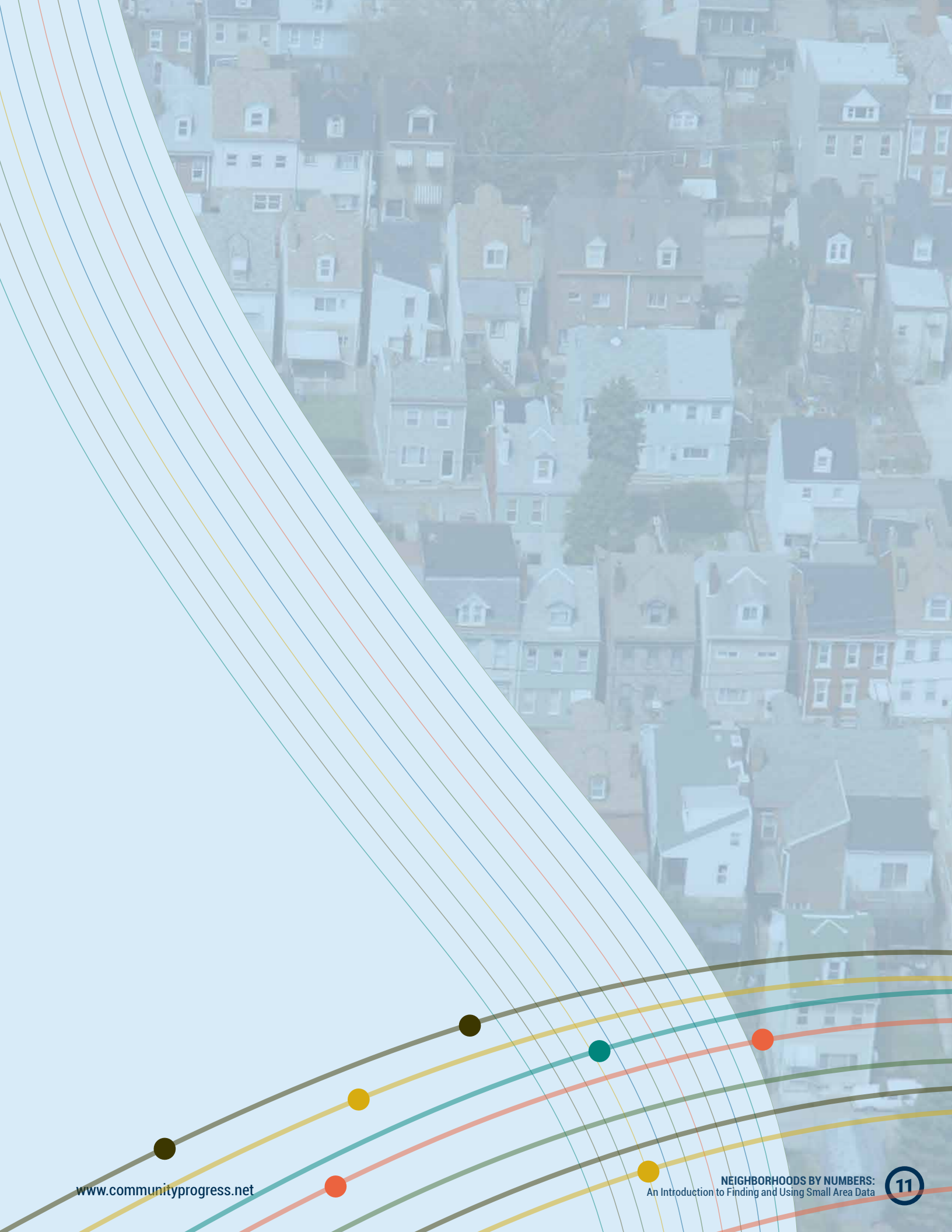
assessment, and those who are approaching this subject for the first time. For the latter group, we have tried to make sure that we define technical terms and concepts when we use them for the first time. At the same time, we have tried to provide enough technical detail that even fairly experienced data users may find new and valuable information here.

While there are a number of consulting firms around the country that will conduct analyses and prepare reports under contract to local governments or community-based organizations, we find the best approach to gathering, analyzing, and using small area data is often a homegrown one, where local officials, CDC staff, or community leaders work with a local college or university, or just use their own skills, to pull together, analyze, and present the information. This guide is designed to encourage local stakeholders to explore how they might do so in their communities.

NAVIGATING THIS GUIDE

The guide has many illustrations to accompany the text. Illustrations that present information in tabular form, whether numbers or text, are referred to as **tables**, while illustrations that present information in other forms, such as maps or graphs, are referred to as **figures**.

Italics are used where we want to emphasize a word, or a sentence or a particular point. **Bold** is used when a term that may be new to the reader is used for the first time. All terms shown in **bold** are defined where they appear in the text and in the glossary (Appendix B).





PART I: FINDING THE DATA

A. NATIONAL DATA SOURCES

This section discusses national data sources. It begins with a discussion of Census Bureau products, which are almost always the foundation of a neighborhood database. The rest of the section looks at other free national datasets along with data sources that can be obtained for a fee.





1. CENSUS BUREAU DATA

The United States Bureau of the Census is the principal data collection and publication vehicle of the federal government, and provides a wide range of data sources and products. Three Census datasets are particularly relevant for small areas – the decennial census, the American Community Survey (ACS), and the Longitudinal Employer-Household Dynamics (LEHD) program, or OnTheMap. Table 1-1 shows in summary form the information that is available from each of these three data sources, and the level of geography at which the information is offered. The principal small area unit used by the Bureau of the Census in all of the datasets is the **census tract**, which is further divided into **block groups** (see “What Is a Census Tract?” on the next page).

The Decennial (10-Year) Census

Every 10 years, the Census Bureau counts the population of the United States. In the past, the census included a “long form” which provided detailed social and economic information. The long form was abolished after the 2000 census. As a result, the only data available from the 2010 census is:

- Age, sex and race or ethnicity of the population
- Type and size of households
- Housing occupancy (occupied or vacant, and category of vacant) and tenure (owner or renter)

This information is limited, but still very useful. It can be used to track population shifts, including racial and ethnic changes; changes in household types, such as a decline in married-couple households or an increase in single person households, and changes in vacancy rates and homeownership rates, as well as current conditions in all of those areas. All of these are relevant to neighborhood planning and investment.



Decennial census data for 2000 and 2010 can be obtained from American FactFinder at <https://factfinder.census.gov/>.

The American Community Survey

Information about the demographic, social, and economic make-up of the population is now provided annually by the American Community Survey (ACS), which began in 2005. That information includes detailed information on household types, workforce activity, educational levels, incomes, housing costs and much more.

The most serious limitation of the ACS is that it is based on a much smaller sample of households than the old Census long form, so even though it is conducted annually, information from any one year is not reliable for small areas. In order to address this problem, each year the Census Bureau “pools together” data from five consecutive years to create five-year ACS estimates for small areas. The most recent data is

TABLE 1-1: Principal census small area data sources

DATA SOURCE	What Information Does It Contain?	What Periods Is It Available For?	What Small Areas Is It Available For?
Decennial (every ten years) Census	Basic population information including total count, household type, race and ethnicity, and housing occupancy and tenure (owner or renter)	Survey conducted every 10 years Most recent data is for 2010	Census tract Block group
Five-year American Community Survey	Detailed demographic, social and economic information about individuals, households, families and housing units	Survey conducted every year Most recent data is for the five-year period from 2011-2015	Census tract Block group
OnTheMap	Information about location of jobs and workers, and basic characteristics of industry sector and workforce	Data compiled every year since 2002 Most recent data is for 2014	Census tract User-defined areas



WHAT IS A CENSUS TRACT?

A census tract is a geographic unit first created in 1960 by the US Bureau of the Census to report and publish census and related data. It typically contains anything from 10 to 50 city blocks, and a population between 1,000 and 5,000 people. A vast array of demographic, social, economic, and housing data is available at the census tract level, which is used by researchers doing neighborhood studies. Census tracts are further subdivided into block groups, which are smaller clusters of city blocks, and which can also be used for small area analyses, but for which data availability is more limited and the data often less reliable.

for 2011-2015. Five-year ACS data must be used with care: even though 2013 is the midpoint of 2011-2015, one cannot call that data "2013" data. How to use this data correctly is discussed in more detail in the Census Bureau guidebook (see resources guide).

Another reason for being careful in using ACS data for small areas is its **margin of error**. Since it is based on a small sample, even when five years of data are pooled, there is still a large error factor. The Census Bureau acknowledges this limitation and prints the margin of error next to the estimate in their data. By way of example, Table 1-2 is a screenshot of a dataset from the 2009-2013 ACS for two census tracts in the city of Trenton, New Jersey, showing the distribution of households in each tract by income. The margin of

error is calculated at what is known as the 90% level, which means that there is a 90% probability that the actual number is within the range shown in the margin of error column. As one can see from Table 1-2, the margin of error for this data is very large relative to the number itself; in Census Tract 1, the number of households earning \$40,000 to \$44,999 is 49 +/-44. What that means is that the number of households in that category could be anywhere from 5 (49-44) to 93 (49+44), which is a very wide range. Generally speaking, the smaller the numbers in the ranges or cells, the larger the margin of error.

This limitation *does not* mean one should not use ACS data. For much of the information directly relevant to neighborhood conditions, it is the only primary source of data that is available

TABLE 1-2: Five-year 2009-2013 ACS census tract table for median household income

	Census Tract 1, Mercer County, New Jersey		Census Tract 2, Mercer County, New Jersey	
	Estimate	Margin of Error	Estimate	Margin of Error
Total:	1,160	+/-87	1,111	+/-111
Less than \$10,000	181	+/-75	62	+/-48
\$10,000 to \$14,999	172	+/-71	57	+/-42
\$15,000 to \$19,999	101	+/-60	112	+/-71
\$20,000 to \$24,999	51	+/-37	40	+/-41
\$25,000 to \$29,999	47	+/-42	132	+/-85
\$30,000 to \$34,999	93	+/-56	52	+/-63
\$35,000 to \$39,999	53	+/-51	83	+/-77
\$40,000 to \$44,999	49	+/-44	47	+/-49
\$45,000 to \$49,999	21	+/-31	33	+/-35
\$50,000 to \$59,999	59	+/-38	65	+/-51
\$60,000 to \$74,999	67	+/-54	218	+/-97
\$75,000 to \$99,999	150	+/-71	114	+/-85

SOURCE: American Community Survey (2009-2013), U.S. Census Bureau, available at <https://factfinder.census.gov>

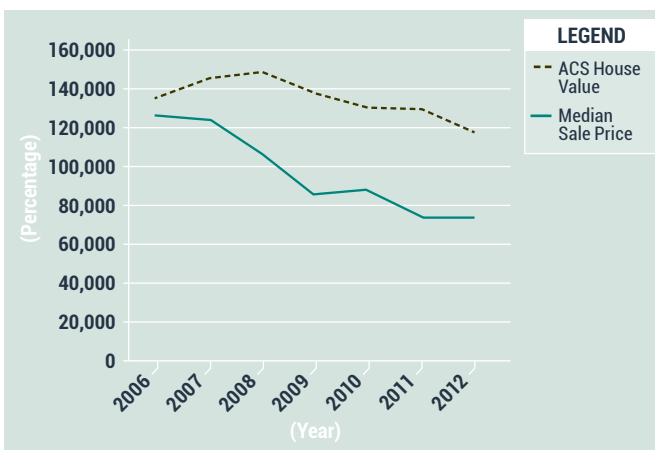


for small areas across the United States. It does mean that ACS data has to be used carefully. In the case of median incomes, a user should not try to present the data at the level of detail shown in the table, but instead to combine ranges, say 0–\$24,999, \$25,000–\$49,999, \$50,000–\$74,999 and so forth, in order to reduce the margin of error by increasing the number of cases in each cell. Great caution must be exercised if using block group data from the ACS, since in many cases data at that level may be simply too unreliable to use responsibly.

One particular ACS dataset we would strongly advise users to be very careful with is “median house value,” which is self-reported. The problem with this data is that the margin of error problem is compounded by the fact that owners (1) don’t necessarily know what their house is worth; and (2) are very slow to adjust their assessment of their home’s value to reflect reality as the market changes, particularly if it is moving downward. This is illustrated in Figure 1-1, which compares the house value figures in the ACS with median sales price by year for Milwaukee between 2006 and 2012.

As the figure shows, while the two sets of numbers were only about 10% apart in 2006, by 2012 the median self-reported value was 60% higher than the value based on sales transactions.

FIGURE 1-1: Comparing ACS median house value data based on owner self-reporting with sales transactions



Data for Milwaukee, Wisconsin. American Community Survey data can be obtained from American FactFinder at <https://FactFinder.census.gov/>

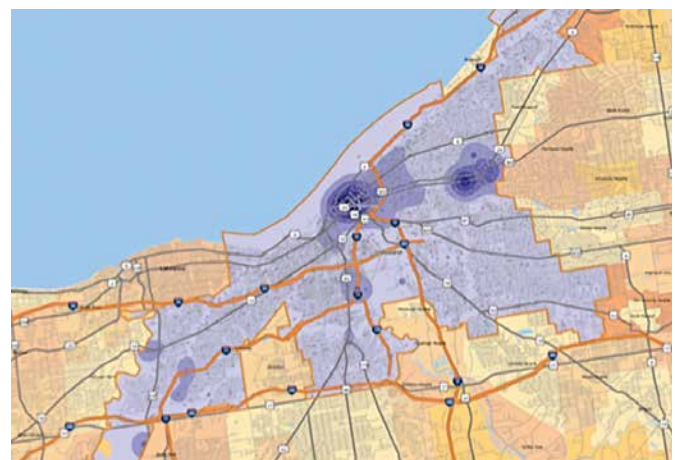
OnTheMap

The OnTheMap site is a product of the Longitudinal Employer-Household Dynamics (LEHD) program of the Census Bureau, which links federal and state data sources on jobs and workers to small areas. OnTheMap allows a user to pinpoint the number of jobs and the number of workers by census tract, as well as to find out the industry sector (manufacturing, retail, etc.) of the jobs in an area or of the workers who live in an area, basic information about the workers, including wage range, race, and educational level, and information about job flows in/out of the area. The site provides data as well as mapping capabilities; for example, Figure 1-2 shows a “heat map” of Cleveland from OnTheMap, showing that jobs in that city are tightly clustered in two areas, downtown and the University Circle area.

A special feature of OnTheMap, which is particularly useful for people doing neighborhood planning, is that one can use the site to get data for a **user-defined area**, by using a tool on the website to draw the boundaries of an area on the map. By way of example, Figure 1-3 shows a map of job flows for a neighborhood of Detroit in 2011. The map illustrates that almost no residents of this area work in the same neighborhood where they live.

This information may suggest opportunities for potential action by a creative local government or CDC. The data from

FIGURE 1-2: Job location “heat map” from OnTheMap website




Map depicts Cleveland, Ohio.



this website can also be used to reveal important trends in the neighborhood's job and workforce profile, as shown in Table 1-3.


Most data from the OnTheMap website is available for each year from 2002 to 2014, although certain data related to characteristics of the workforce is only available for more recent years.

 OnTheMap data can be obtained from <http://onthemap.ces.census.gov/>.

Other Census Products

In addition to these small area products, the Census Bureau publishes many other data reports.

The economic census, which is conducted in years ending in 2 and 7, provides detailed information on business activity, including manufacturing, retail, and services at the citywide and countywide levels, and may be useful in order to get background on economic activity and trends in the community as a whole.

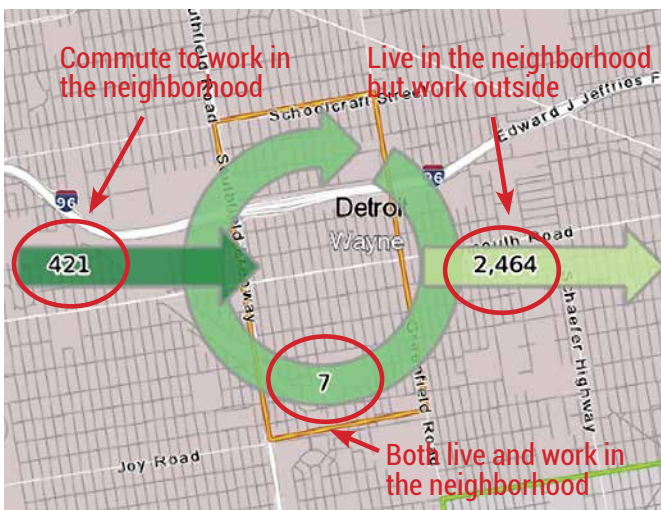
 Economic census data is also available at the American FactFinder website at <https://FactFinder.census.gov/>.

For more sophisticated users, the census provides a mapmaking tool that can be used to create maps from their data known as TIGER, which stands for Topologically Integrated Geographic Encoding and Referencing.

Data for the 2000 and 2010 decennial census, and for all American Community Survey years can be downloaded in a variety of formats from the American FactFinder website including as CSV (comma-separate value files) for statistical use, or as a PDF for presentation purposes.

 OnTheMap data can be downloaded by using the LEHD Extraction Tool on the census website at <http://ledextract.ces.census.gov/>.

FIGURE 1-3: Mapping job flows at the neighborhood level through OnTheMap



Map depicts 2011 data for Grandale neighborhood in Detroit, Michigan.

SOURCE: OnTheMap, available at <http://onthemap.ces.census.gov>

TABLE 1-3: Tracking neighborhood-level job/worker shifts through OnTheMap

	2002	2011	Change
Number of people who live in the neighborhood and have jobs	5119	2471	-52%
Number of jobs located in the neighborhood	835	428	-49%

2002-2011 data for Grandale neighborhood in Detroit, Michigan.



2. OTHER IMPORTANT FEDERAL DATA

While most federal data other than the census data described is only available for larger areas, such as counties or metropolitan areas, there are a couple of valuable federal data sources where the data is available by census tract. The two most important are data gathered about mortgage lending under the Home Mortgage Disclosure Act (HMDA), and data on vacant properties from the United States Postal Service (USPS).

HMDA Data

All mortgages and other loans secured by a home that are made by regulated financial institutions, including banks, savings associations, and credit unions, must be reported to the Federal Financial Institutions Examination Council (FFIEC), which publishes the data on its website (www.ffiec.gov). Data is organized for each metropolitan area in the United States, and can be accessed at <http://www.ffiec.gov/hmdaadwebreport/aggwelcome.aspx>. Data can also be accessed for individual lenders by metropolitan area.

The published data provide basic information by census tract, including the number of loans originated and denied and the average loan amount, by type of loan and by whether the borrower was an owner-occupant or non-occupant, such as an absentee or investor borrower. Sophisticated data users may want to download data files, so they can do more detailed analysis. The detailed data includes information on the income range, race, and ethnicity of the borrowers and the census tract, and on the nature of the loan, including how many loans were high-cost loans.

Users should never, however, use mortgages as a proxy for real estate transactions. Many houses, particularly in lower-value neighborhoods, are purchased with cash or with financing from informal non-HMDA reporting sources. To illustrate this point, Table 1-4 compares the relationship between total sales and purchase mortgages with household income for 10 lower-income and 10 upper-income census tracts in Baltimore. As the table shows, only 5% or 1 out of 20 sales in the lower-income tracts had mortgages, while 52% or more than half of the sales in the upper-income tracts had mortgages. This table illustrates how one can combine data from the American Community Survey, federal HMDA, and commercial data available on PolicyMap, a data resource described on page 20.

CENSUS TRACT	MEDIAN HOUSEHOLD INCOME (from ACS)	NUMBER OF HMDA PURCHASE MORTGAGES 2013 (from HMDA data)	NUMBER OF SALES 2013 (from PolicyMap)	MORTGAGES AS PERCENTAGE OF SALES
1402	\$14858	1	16	
1403	\$19338	3	48	
1501	\$21083	1	35	
1502	\$26739	1	33	
1503	\$32286	0	61	
1601	\$17268	1	43	
1602	\$35551	1	17	
1702	\$ 9877	2	24	
1701	\$29527	4	16	
1703	\$23843	1	14	
10 lower-income tracts		15	307	5%
101	\$90604	87	155	
104	\$85417	64	112	
1307	\$61318	56	112	
2201	\$57574	66	253	
2403	\$106641	56	85	
2712	\$111435	95	228	
2713	\$128688	30	54	
2714	\$108295	40	82	
2715.01	\$99375	63	108	
10 upper-income tracts		557	1189	47%

TABLE 1-4: Comparing sales and mortgages in lower-income and upper-income census tracts

2013 data for Baltimore, Maryland.

SOURCES:

American Community Survey, available at <https://factfinder.census.gov>;

Federal Financial Institutions Examination Council, available at <http://www.ffiec.gov>;

PolicyMap, available at <https://www.policymap.com>



US Postal Service Data

The United States Postal Service (USPS) routinely gathers information for mail delivery purposes about which addresses have been vacant for more than 90 days, and which addresses are what they call “no-stat,” meaning “addresses for businesses or homes under construction and not yet occupied, or addresses in urban areas identified by a carrier as not likely to be active for some time.” Thus, a building that is a vacant shell should be classified as no-stat. It includes information on business and residential addresses separately, and how long they have been vacant; (e.g. 90-180 days, 180-360 days, over 1 year).



Under an agreement between the USPS and HUD, this data is made available on a quarterly basis without charge to approved users, who must be affiliated with government agencies or nonprofit organizations. To submit your organization for approval as a user, go to <http://www.huduser.org/portal/usps/index.html>.

While the USPS data can be valuable in tracking neighborhood change, and is one of the very few datasets available as often as quarterly, it must be used carefully. While the tabulation of “no-stat” addresses would appear on its face to be a good proxy for abandoned properties, users have found many inconsistencies in that data. These tend to be principally of two kinds: differences between practices from one post office to the next that lead to inconsistency in labeling properties “no stat” or vacant, and changes in the data gathering procedures, particularly in 2010, that make it difficult to use the data to track longer-term trends. Users of USPS data should be on the lookout for inconsistencies, and check the data against field observation before relying on it. In addition to providing free data for small areas through the HUD User website, the Postal Service also licenses its data to commercial firms that sell address-level or point source data about vacant addresses.

FBI Uniform Crime Reporting

Crime data is available for individual towns and cities, but not for small areas, through the FBI Uniform Crime Reporting (UCR) Program. This data, however, can be useful for looking at overall trends, and for benchmarking individual neighborhoods where small area crime data can be obtained from one’s local police department.



UCR data can be found on the FBI site at <http://www.fbi.gov/about-us/cjis/ucr/ucr-publications>.

3. COMMERCIAL DATA PROVIDERS

In addition to data that the federal government makes available without charge, there are a number of sources of small area data available through commercial data providers that charge a fee for the data, although some make certain datasets available on their websites without charge. The cost of data from commercial providers varies widely, but is often within the means of many organizations.

PolicyMap

PolicyMap is a subscription-access website created by The Reinvestment Fund, a Philadelphia-based nonprofit organization. It aggregates data from many different sources, public and private, and makes it available to users in the form of maps or tables. An example of a map downloaded from the PolicyMap website, in this case showing the concentration of low-income families in the Newark, New Jersey, area, is shown as Figure 1-4. While the raw data used to make this map comes from the ACS and could be downloaded directly from the American FactFinder website, PolicyMap makes it available directly to users in map and other formats. Users can also download the underlying data used to create the map in the form of a table, which can then be exported to Excel or other database programs.

PolicyMap aggregates data from all of the federal sources mentioned above as well as from other datasets. A few of these additional datasets are available at the census tract or block group level, most notably data on sales transactions (number and price) collected by the market research firm Boxwood Means, as well as health-related data. Most other datasets are available by zip code, or by city or county. It is also possible to download most PolicyMap datasets at the census tract or block group level.



For further information about PolicyMap, go to <http://www.policymap.com/>.

Other Data Providers

With the development of automated tools for **data scraping** and the ability of firms to assemble vast databases, many national firms have emerged that provide real estate data to users on a fee basis. Many of these firms have data for all or most parts of the United States, while some have data



for only selected geographic areas. In addition, there are a number of locally oriented firms that provide more in-depth data resources for a particular city, state, or metro.

Typically, these firms can provide data on underlying property records, real estate transactions, mortgage foreclosures, and the like. Among the better-known firms in this area are Zillow, CoreLogic, RealtyTrac, and DataQuick, while a firm that specializes in providing rental data is RentJungle. This is not intended to be a comprehensive list, nor an endorsement of any specific firm's products or services.



Zillow provides a variety of datasets on its website without charge, which can be accessed at <http://www.zillow.com/research/data/>, including data for selected neighborhoods (not clearly defined on the website) in many major cities.

The value of buying data from a commercial source is that it can be tailored to the user's needs. For example, one can buy data on individual foreclosures, which can then be clustered by block, block group, census tract, or other geography. While most of the data that is provided by these sources comes from public records, it is often more cost-effective — particularly in communities where public agencies do not make data readily available on their websites — to buy the data than to spend

the many hours that may be necessary to pull it together from the source. The downside of much commercial data is that the providers accumulate vast quantities of data without necessarily checking for accuracy or duplicate entries. As a result, buyers of data may have to spend time **data cleaning** or **scrubbing** before it can be used.

In addition to raw data, these firms will also provide users with analytic or statistical reports at additional cost; in practice, however, it is better in most cases to buy raw data so users can do their own analyses, as well as reuse the data in different ways, based on particular local needs and interests.

Ground rules for buying data vary from firm to firm. Some firms will allow the data to be reused, made available to others, and published with few limitations; others, however, will offer one price for a limited license to use their data for a particular purpose, with strict restrictions on data-sharing, and a much higher price for the right to share the data and use it more freely. In many cases, however, these terms can be negotiated. In all cases, before buying any data from any commercial provider, prospective users should have a clear idea of what data they want and how they plan to use it, and go over their needs carefully with a representative of the firm to make sure that the purchase will be worth the money.

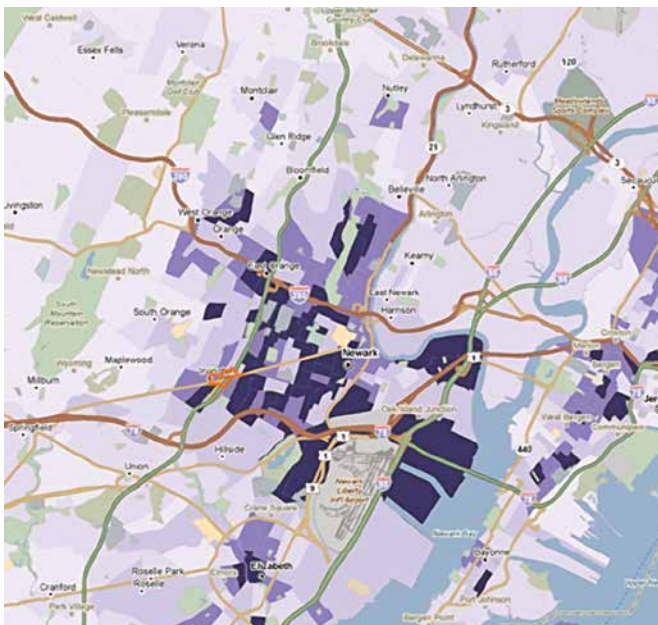
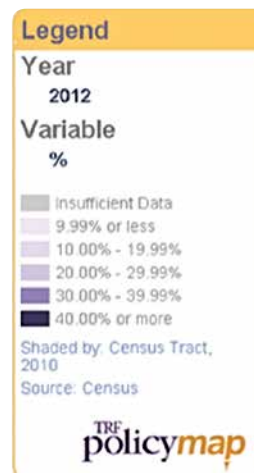


FIGURE 1-4: Map from PolicyMap showing concentrations of households with incomes below \$25,000



2012 data for Newark, New Jersey.
SOURCE: PolicyMap, available at <https://www.policymap.com>

B. LOCAL DATA SOURCES

While national data sources, particularly the census, are invaluable, the most valuable data for understanding property and neighborhood conditions is usually close at hand, in the many forms of information gathered routinely by local public agencies.





While it may sometimes be difficult or cumbersome to obtain and extract the data from local sources, it is usually worth the effort. A particular benefit of gathering data from local sources is that one can obtain **point source data**; that is, data for individual parcels, properties, or locations, as distinct from **aggregated data**; that is, data that has already been aggregated, or combined, for an area by someone else, like the Census Bureau. Point source data can be aggregated to any geographic unit for which a stakeholder needs information, or used as the material to create a community information database. Some of the potential uses of such a database are discussed on page 50.

1. DATA COLLECTED BY PUBLIC AGENCIES

Table 2-1 shows the property-related datasets that are usually collected by public agencies in most parts of the United States. Depending on the state, the data may be collected by a municipal or a county agency, and in some cases, may be assembled and made available by the state as well. The agency responsible for a particular dataset varies from state to state — in one state the agency responsible for property transactions may be called the county clerk, in another the county recorder, and in yet another, the register of deeds. Similarly, the terminology used in different states for the same action or procedure may vary. The sale of a property that has gone through foreclosure may be called a sheriff's sale in one state, a foreclosure sale in another and a trustee sale in yet another state. It is important to become familiar with local terminology to be sure that the data one is gathering is indeed the data one wants and needs.

Property Records

Just about every jurisdiction in the United States levies property taxes. As a result, they all keep records on properties in

TABLE 2-1: Data regularly gathered by municipal or county public agencies

DATASET	Information Usually Available	Source*
Property information	Name and address of owner, type of property, property features (square footage of lot and building, etc.), assessment and taxes levied.	Municipal or county tax assessor or other land records agency
Property transactions	Date of sale, name and address of buyer and seller, type of property, consideration (sales price), and (sometimes) whether a mortgage has been taken out on the property.	County recorder, register of deeds or clerk
Tax sale or tax foreclosure	Properties on which tax liens have been sold and name and address of lien buyer, or which have been sold in tax foreclosure auction with name and address of buyer (see foreclosure text box).	Municipal or county tax collector or treasurer
Mortgage foreclosure	Properties on which foreclosures have been filed, name and address of owner and name of foreclosing entity.	State or county court with foreclosure jurisdiction or county agency (see foreclosure text box)
Foreclosure sales	Properties which have been sold as a result of foreclosure, with name and address of buyer.	County sheriff
Crime incidents	Crime incidents reported by type and location.	Police or public safety department

*Names shown are typical names of agencies responsible for the subject area shown. Actual name may vary by state and locality.



order to track how much the property is worth, what the tax bill should be each year, and to whom the tax bill should be sent. In many cases, this information is posted online on the agency's website in the form of an Excel or similar spreadsheet, and can be downloaded by the user. Table 2-2 illustrates the format of such a list, in this case that used by tax assessors and tax collectors in the state of New Jersey, and known as the Mod IV file. The list provides the property address, including block and lot, the property class (Class 2 referring to 1 to 4 family residential properties), and the name of the owner.

By clicking on the "More Info" tab for an individual property, we get the data shown in Table 2-3. We see what type of property it is, when it was built, and whether it is owner-occupied or owned by an absentee owner, depending on whether

the owner's address matches the actual property address. We also can see where the landlord lives, and how long he or she has owned the property. All of this information for all of the properties in a city or county can be directly downloaded as an Excel spreadsheet, so there is no need to code entries or transfer the information from the website to the user's spreadsheet property-by-property.

In addition to maintaining property records, local governments track real estate transactions, in order to make sure that their information on the owner is up to date, and to monitor price changes so that they can adjust property tax assessments accordingly. These records typically include the property address, the buyer's and seller's names and addresses, sale date, and the consideration or sales price. In some states it may include whether there is a mortgage on the property, and by whom.

This kind of data enables a user to get a reasonably accurate count of homeowners and absentee owners, or people buying as owner-occupants or as investors, by comparing the address of the property with the address that the owner has provided to which property tax bills are sent. Where the two are different, one can be almost certain that the owner is an absentee owner, rather than an owner-occupant. One should also look at the properties where the two addresses are the same, in order to screen out owner names that are clearly not individuals, such as corporate names or LLCs. A user can create a spreadsheet of properties for a small area such

TABLE 2-2: Excerpt from New Jersey Property Record (Mod IV)

	BL	LOT	CL	ADDRESS	OWNER
<u>More Info</u>					
<u>More Info</u>					
<u>More Info</u>					
<u>More Info</u>					
<u>More Info</u>					

New Jersey Property Record (ModIV) file (2016), available at <http://opr.s.co.monmouth.nj.us/opr.s/External.aspx?ild=12>

TABLE 2-3: Detailed property record from New Jersey Property Record (Mod IV) file

Block:	Prop Loc:		Owner:	Square Ft:
Lot:	District:		Street:	Year Built:
Qual:	Class:		City State:	Style:
Prior Block:	Acct Num:		Add. Lots:	EPL Code:
Prior Lot:	Mtg Acct:		Land Desc:	Statute:
Prior Qual:	Bank Code:		Bldg Desc:	Initial:
Updated:	Tax Codes:		Class4Cd:	Desc:
Zone:	Map Page:		Acreage:	Taxes:
Sale Date:	Book:	Page:	Price:	
Sr1a	Date	Book	Page	Price
				NU#
				Ratio
				Grantee
TAX-LIST-HISTORY				
Year	Owner Information	Land/Imp/Tot	Exemption	Assessed
				Property Class

New Jersey Property Record (ModIV) file (2016), available at <http://opr.s.co.monmouth.nj.us/opr.s/External.aspx?ild=13>

as a block, census tract or neighborhood, and then use it to track changing property conditions in the area, or changes to individual property conditions, such as resales, over time. By combining property record data with data on vacant properties or property conditions from a parcel survey (see page 33-34), one can get still more useful information. The data can also be mapped, to give a user a sense of where concentrations of homeowners, for example, may be located, or where large numbers of properties have turned over recently.

By **geocoding** (see page 32) this data to the census tract or other small area, a user may be able to answer some important questions, including:

- How many sales are taking place in the area? Is the sales volume unreasonably high or low in light of the number of properties? The actual level of sales can be compared to norms or rules of thumb used in the real estate industry to make this judgment.
- Are the sellers mainly homeowners or absentee owners?
- Are the buyers mainly homeowners or absentee owners? Is the share of absentee *buyers* more, less, or about the same as the share of *absentee* owners in the area?
- What are the typical prices? How do they compare to other parts of the city or region? Are they trending up or down?
- Are the absentee buyers local or out-of-town? Are properties being bought by large numbers of small (mom and pop) buyers, or a few large investors?

The last point may not always be easy to pin down, because some large investors use multiple shell entities or nominees to hold properties. In some cases, however, where the owner name is different, the address may be the same, indicating common ownership. All of these questions are directly relevant to key decisions about neighborhood strategy and investment.

Foreclosure Records

Both mortgage foreclosure and tax foreclosure are important in planning neighborhood strategies. Both are affected by the underlying conditions in the neighborhood, but abundant research evidence has shown that both, and particularly mortgage foreclosure, also have powerful effects on neighborhood conditions. Some basic information about how the foreclosure process works is given on the following page.

Mortgage Foreclosure

While mortgage foreclosures in the United States have declined sharply since peaking between 2008 and 2011, they are still a serious problem in many areas, particularly in lower-income communities of color. In **judicial foreclosure** states, it is important to track **foreclosure filings** (when the lender submits the initial motion to foreclose to the court) in addition to the actual foreclosure sale, since there is a long and unpredictable lag between the two. In addition, depending on market conditions in the community, the lender may or may not decide to pursue the expensive and time-consuming foreclosure process all the way to sale. In **non-judicial foreclosure** states, the only official public action is the foreclosure sale. The office usually responsible for foreclosure sales is the county sheriff's department.

Getting mortgage foreclosure data from its sources can be complicated. Courts and sheriffs vary widely in terms of how accessible and usable their foreclosure data is likely to be. Some put their data online, some maintain electronic data that is not accessible online, and some still use paper to store data. Moreover, when looking at data from these sources, it is often difficult to be sure exactly what the data means. The technical requirements for foreclosure in many states mean that each foreclosure may generate multiple data entries. Some sheriff's offices publish lists of properties that are scheduled for foreclosure sale, but do not always make clear which were actually sold, which were cancelled, and which were postponed. In many cases, getting good data may require many hours in the courthouse or the county office building. For these reasons, many users of foreclosure data, when they can afford to do so, buy the data from a commercial data provider rather than going through the hassle of getting it themselves.

Tax Foreclosure

Although in some cases it is a function of hardship, an owner's decision to not pay taxes and let her property go into foreclosure also often reflects lack of confidence in the neighborhood or community. Thus, a rising rate of tax delinquency, reflected in the number of properties that go up for tax sale or foreclosure auction, is an important indicator of increasing neighborhood instability.

The extent to which investors buy tax liens or tax foreclosed properties is an important indicator of how *the outside investor world* looks at the same neighborhood or community, and



FORECLOSURE BASICS

Foreclosure refers to “the process of taking possession of a property as a result of the owner’s failure to keep up either mortgage or property tax payments.” It is an important element in understanding what is going on in a neighborhood, but is easily misunderstood. There are two separate forms of foreclosure – tax foreclosure and mortgage foreclosure – and the rules for each vary from state to state. Here is some basic information about the two.

TAX FORECLOSURE

If a property owner fails to make property tax payments, his or her property may be subject to foreclosure. In some states, after the owner misses a certain number of payments the property is sold by the county or municipality in a foreclosure auction, and in others the property is put up for **tax sale**. In the latter states, the property itself is not sold – what is sold is a lien on the property for the amount due. The person who buys the lien then gets the right to foreclose on the property by filing a motion with the court (but is not obligated to foreclose) if the owner fails to pay back the taxes (redeem the property) within a specified time period.

MORTGAGE FORECLOSURE

If a property owner fails to make mortgage payments, the lender may foreclose on the property. In some states, the lender must file a foreclosure motion with the court, and the foreclosure sale only takes place after the lender receives a court order authorizing the sale. These states are called **judicial foreclosure** states. In other states, after the lender notifies the owner of the default, and the owner fails to cure, the lender goes directly to the sheriff who schedules the sale of the property. These states, are called **non-judicial foreclosure** states. A table showing foreclosure laws for each state can be found at <http://www.realtytrac.com/real-estate-guides/foreclosure-laws/>

whether they expect to make their money back from either the owner redeeming the property, or by taking and selling or using the property themselves.

In most states, tax foreclosure is handled by a county official, usually the county treasurer or tax collector. In some, it is handled by the municipal treasurer or tax collector. As indicated above, some states sell *liens* on tax delinquent properties, which can then be foreclosed by the lien buyer after a waiting period, while in other states, the property is put up for sale. This distinction is particularly important for practitioners trying to get a handle on the level of tax delinquency in the community.

In a state which sells tax liens (also called tax sale certificates), foreclosure is an option for the holder of the tax lien which may or may not be exercised, and in practice is difficult to track. Thus, in these states the most useful data is the *list of properties that are put up for tax lien sale*. In states that sell the property, the most useful data is the *list of properties in the foreclosure auction*. In both cases this data can usually be obtained from the county or municipality.

It is useful to track not only which properties or liens were put up for sale, but whether they found buyers or not. Figure 2-2 shows a map of neighborhoods in the city of Trenton, New

Jersey, ranked by the percentage of properties put up for tax sale that were not bought by investors (in New Jersey terminology, "struck off" to the city). In the city's stronger neighborhoods, investors bought all or almost all of the tax liens offered; in some weaker neighborhoods, half or more of the tax liens offered found no buyer.

Crime records

While crime data is not, strictly speaking, property data, we include it here because crime is one of the most significant factors — if not the most significant — that affects property values and neighborhood stability. Crime has a powerful impact on whether people move into a neighborhood, and whether they stay in their neighborhood when they have the means to move out. Research suggests that violent crime has a more pronounced effect on neighborhood stability than property crime. The FBI violent crime index includes murder and non-negligent manslaughter, rape, robbery, and aggravated assault.

Almost every police department in urban and larger suburban municipalities in the United States has a data system into which all reported crime incidents are entered, and which can pinpoint incidents by type and by location, usually with great precision. This data can then be geocoded to blocks,

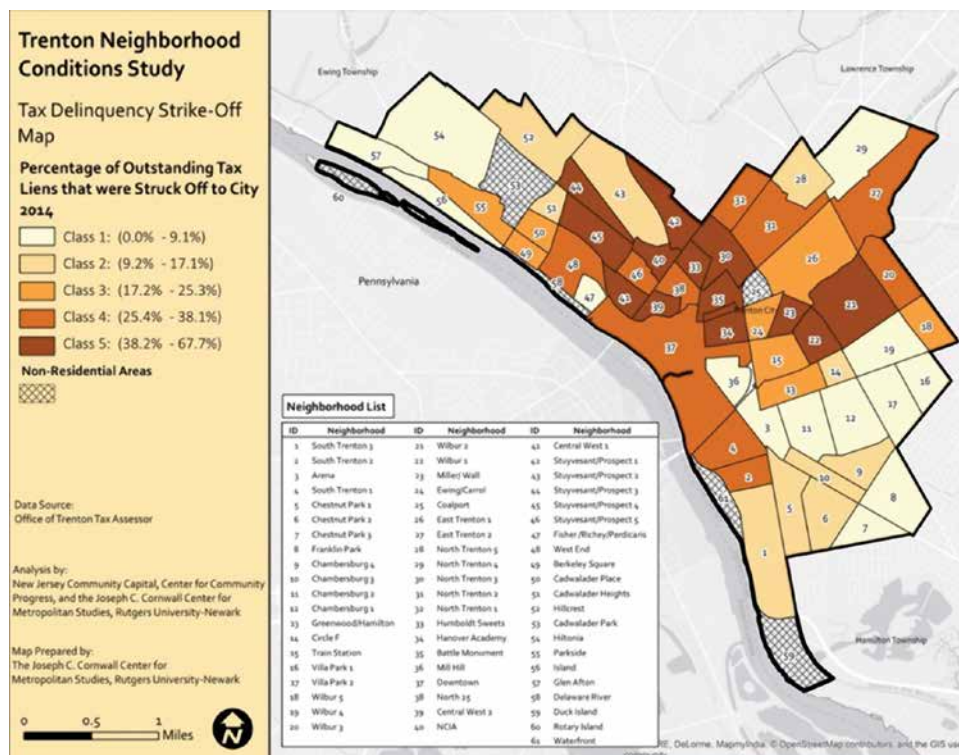


FIGURE 2-2:
Neighborhoods by percentage of tax liens offered for sale that did not find buyers

2014 Data for Trenton, New Jersey.



census tracts or community-defined neighborhoods. While this data is nominally public record data, in practice police departments vary widely in their ability and willingness to make this data available to others in usable form. This is discussed further on page 29, on getting the data from its owners.

Other local data sources

The three data sources shown above are generally available from public agencies in almost every community. Other property-related data is also likely to be available,

although the availability and quality will vary from community to community (See Table 2-4). Much of this data, however, may not add value for neighborhood planning or strategy purposes. For example, data on code violations in a city where code enforcement is handled by inspectors responding to complaints may not reflect actual neighborhood conditions. However, in a city that proactively inspects neighborhoods on the basis of explicit criteria, this information may be more useful. Similarly, while lists of licensed rental properties would appear to be useful, in practice in many cities the majority of landlords are not licensed, and the list tends to be disproportionately weighted to larger properties with professional management.

In other cases, however, even though a data source may not say much in and of itself about neighborhood conditions, it may be useful to add to the database. The inventory of publicly owned properties, for example, which may be held by the municipal government, a land bank, or a redevelopment agency, should be included, so that users can compare the location of those properties to where privately owned vacant properties or tax delinquent properties are located, as a guide to potential reuse or redevelopment opportunities. Despite the limitations on many of these datasets, a user should make an effort to find out what is available from different public agencies, and evaluate how comprehensive and useful it might be.

In addition to data from public agencies, some data may be available from local private sources. One potentially useful source is the local Board of Realtors or Realtors Association, which maintains records on real estate transactions from the Multiple Listing Service. As a measure of sales price, data from city or county records is preferable, since the MLS data is likely to be skewed upward relative to the total universe of real estate transactions, as very low value properties are more likely to be sold without a realtor's services. One dataset that is available from this source, however, which is generally not available from other sources is data on **days on market**, or the number of days a property is on the market before it is sold. This in itself can be a useful market indicator.

While it is worth canvassing other private sources, such as Chambers of Commerce, Builders Associations, and other similar organizations to see if they have useful data, almost all of these organizations rely on public data and few gather their own data.

TABLE 2-4: Other data that may be available from local public sources

DATASET	Likely source
Code enforcement citations	Municipal department of buildings, inspections, or neighborhood services
Building permits (including permits for rehab and home improvements)	Municipal department of buildings, inspections, or neighborhood services
CDBG and other public subsidy expenditures	Municipal department of housing or community development
Rental registration or licensing	Municipal department of buildings, inspections, or neighborhood services
Housing choice vouchers	Municipal housing authority or state/county agency
Vacant properties	Municipal department of buildings, inspections, or neighborhood services
Inventory of property held by public agencies	Municipal department of buildings, real estate or public property; redevelopment agency or authority; municipal or county land bank entity.
Condemned buildings and demolitions conducted	Municipal department of buildings, inspections, or neighborhood services
Days on market (real estate sales)	Local board of Realtors (Multiple Listing Service)

2. LIMITATIONS OF PUBLIC AGENCY DATA

While many of the questions that arise when using public agency data are noted in the previous section, this section goes into some of those challenges in further detail. In addition, this section discusses how to turn point source data, such as lists of properties, crime incidents, code violations, etc., into usable data for small areas, the process known as geocoding.

Getting data from public agencies

Although all of the datasets described above exist *some-where* within a city or county system, obtaining them in usable form for analysis is often not as straightforward as one might expect. Three distinct problems often arise when accessing and using public agency data: data format, data quality, and data access.

Data format

The form in which public agencies maintain their data varies widely. At one end, one finds agencies that place their information online in readily accessible format, downloadable into formats such as Excel, which users can then code, sort, and analyze at will. At the other, one still finds, particularly in smaller jurisdictions, some agencies that maintain their information on paper, in ledgers, index cards, or similar formats, and use their computers for little more than word processing and email. Many are in a middle ground, where the data exists in some electronic format, but is not available online. The electronic format they use may or may not be adaptable for searching and analysis.

The first step for users is to find out what form the agency's data is in. If it is in anything other than a publicly accessible, searchable, *and downloadable* database, the next step is to determine what level of effort is needed to obtain the data and turn it into a usable database. If the data needs to be obtained by searching and then either transmitting or transcribing individual property or incident records, which must then be turned into a database, the user should estimate how many records, and how much time is required for each, to determine whether it is feasible to gather that data.

If the time or cost of obtaining a particular dataset is prohibitive, it is worth exploring whether alternative datasets can get at the same subject. For example, while data from the county

recorder on real estate transactions is the best way to track sales prices, as well as whether the buyers are investors or future owner-occupants, and is usually available without difficulty, that is not always the case. Where it is not, data on sales prices may be potentially obtained either from the local Board of Realtors (from the Multiple Listing Service), or from PolicyMap, which buys data on real estate transactions from a commercial provider and posts them online by census tract and block group. Unfortunately, PolicyMap does not provide point source data, so it cannot be used for all of the purposes that can be pursued with data on individual properties.

Data quality

Even when data is available, it may not always be in a form that can be used without careful checking. All data must first be provided by someone, such as a buyer or lender, and must then be entered into the database by someone else. As a result, both incomplete entries and erroneous ones are likely to be present. As a result, any database that comes from any source must be scrubbed or carefully checked to correct erroneous entries (when they can be corrected) and eliminate duplicate or unusable ones. This can be easy and quick, or difficult and time-consuming, depending on the quality of the initial data.

In addition, datasets may be accurate, but may contain excessive unrelated or confusing information. Foreclosure datasets are particularly problematic. In many states, the foreclosure process involves multiple filings of different documents over time for a single foreclosure on a single property. Unless the data is carefully cleaned to eliminate such multiple filings, the resulting figures could be vastly out of proportion to the reality. In most cases, the first filing, which may be the filing of a document with the court or the filing of a **lis pendens** on the property, is the most important dataset to track.

Data access

The above issues are basically technical; they require technical solutions, for which a technical solution can often be found, or if not, are fairly quickly determined to be infeasible or unrealistic. Issues of gaining access to data are a different matter. While all of the datasets described above are matters of public record, some agencies are reluctant or unwilling to share information with people outside the city or county government, and sometimes even with their co-workers in other city or county departments. This is occasionally the case with property record or tax offices, but most often the case with police departments.



This is changing to some extent, as many cities — particularly large cities — are becoming part of what has been called the “Open Data Movement,” and making increasing amounts of data available through open data portals on their websites. One such city is Dallas, Texas, which provides access to police data, code violations, certificates of occupancy, building permits, and demolition permits on their open data portal (<https://www.dallasopendata.com/>). This data, of course, is subject to the same challenges of formatting and quality as any other public data.

There is no simple answer to the question of how to get data from an uncooperative public agency. It is a matter of relationships, and connecting with the right people in the right way. The data discussed here is public record information, and most states have an open public records law which requires agencies to make the data available. In practice, however, using such laws can be counterproductive. Not only can it establish an adversarial relationship with the agency, but they can effectively neutralize the intent of the law by providing the data in a format that is difficult or impossible to use, or demanding a fee to produce the data.

It is much better to try to get the agency’s support for the project. This is most likely to happen where some part of the city or county government is already involved in the project, and is willing to intercede with the agency for the data; the local planning or community development agency might well go to bat with the police department or tax office to help get the data needed for the analysis. A meeting with senior staff in the agency from which one is seeking data, focusing on how the analysis may be able to help their work, may be helpful. Once the basic ground rules for accessing and using the data have been agreed to, it may be useful to create a formal data-sharing agreement that outlines those ground rules, to ease any concerns the agency may have about use of and access to the data. Such an agreement may be first entered into with respect to a particular dataset or particular use, and ultimately be the basis for a longer-term agreement.

What to watch out for

Public record data can be a valuable resource for practitioners trying to understand conditions and trends in their neighborhoods. At the same time, it is not without pitfalls. Many datasets may be technically accurate, but may not actually measure what they seem to, or what the user is looking for. A couple of examples might be useful:

- Code enforcement citations indicate that the property is in violation of codes, but that does not mean that one can aggregate citations by area in order to get a measure of how property conditions vary from one neighborhood to the next. The reason is that in most cities citations are only issued as a result of complaints, so that it is impossible to know whether the complaints reflect the volume of problems, or a better organized civic association, or some other factor.
- While crime levels are a valuable indicator of neighborhood conditions, crimes reported to police do not always reflect actual crime levels. In one city that we studied, we learned by talking to neighborhood leaders that — as a result of sharp police staffing reductions — many crimes were no longer reported to police because victims considered it pointless to do so. As a result, even though the police department was willing to provide us with their data, we chose not to use it.

Table 2-5 flags potential problems with each of the datasets discussed earlier.

When the problems are data problems, they can be fixed. Foreclosure records can be cleaned to remove duplicate entries, and sales transactions can be screened to exclude any transaction for under \$100 (or some similar number). When the problems are with the underlying system, however, they are not likely to be solvable, except in some cases through more long-term system reform, such as by convincing the municipality to move from a complaint-driven to a more strategic code enforcement system. Finally, some data problems are inherent in the nature of the underlying process.

While the nature of the problems may rule out use of some datasets, in other cases it may simply indicate that they be used only in certain ways rather than others. For example, knowing the geographic distribution of housing choice vouchers by area may not be useful as a measure of the area’s social or economic conditions, but may be useful in itself, to show how the vouchers may be affecting different neighborhoods’ rental markets and to recommend policies and strategies for using vouchers.

**TABLE 2-5:
Potential substantive
problems with
public sector data**

DATASET	POTENTIAL DATA PROBLEMS
Property information	Ownership data may not be current
Property transactions	Transaction lists may include many transactions that are not “arm’s length” sales, including conveyances between family members.
Tax sale or tax foreclosure	Tax foreclosures are not a meaningful measure in jurisdictions that do tax lien sales, because they are discretionary, and not all properties are foreclosed. It is important to make sure that the jurisdiction actually puts all eligible properties up for tax sale or foreclosure auction before relying on data.
Mortgage foreclosure	Foreclosure records often show multiple entries for the same property.
Foreclosure sales	Many properties on which foreclosures are filed never end up being foreclosed.
Crime incidents	Many crime incidents may not be reported to police, and police may not record all reports (although that is rarely a problem with violent crimes).
Vacant properties	Inconsistencies or uncertainties about how to classify properties as vacant, particularly where surveyors are not carefully trained.
Code enforcement citations	Most cities enforce codes based on complaints, and the geographic distribution of complaints is not likely to be representative of the distribution of problem properties.
Building permits (including permits for rehab and home improvements)	Much rehab and home improvement work takes place without permits, so permits may not be a good indicator of the level of activity.
CDBG and other public subsidy expenditures	Decisions about where and how to spend public subsidy money are often based on criteria other than need, such as varying degrees of capability on the part of recipients and political considerations.
Rental registration or licensing	In many cases, only a small percentage of rental units are actually registered or licensed. Small “mom and pop” properties are particularly likely to be under-represented.
Housing choice vouchers	Distribution reflects preferences or information, not actual social or economic conditions.
Municipal property inventory	Distribution reflects circumstances by which municipality acquired properties, not underlying property conditions.
Demolitions	Decisions to demolish properties are often based either on complaints or on redevelopment plans, rather than actual distribution of properties in need of demolition.



Converting Point Source into Small Area Data (Geocoding)

Most, if not all, public records data, whether involving property records, transactions, foreclosures, or vacant properties, will be what is known as point source data. This means that the data will be in the form of individual records for individual points, which can take the form of parcels, addresses or other location identifiers, such as “the corner of Smith and Elm Streets.” For purposes of doing any kind of neighborhood or other small area analysis, or to create maps of neighborhood conditions, that data has to be reorganized so that it corresponds to geographic units that can be mapped, whether they are census tracts, ZIP codes, or other units. The process of doing that is known as geocoding, which has been defined by ArcGIS as “the process of transforming a description of a location—such as a pair of coordinates, an address, or a name of a place—to a location on the earth’s surface.” The process of converting a large number of addresses — for example, all the addresses on a file of tax foreclosures — into spatial data for analysis or mapping is known as batch geocoding.

Converting point source data from local sources to data that can be analyzed at the small area level, and used to map neighborhood conditions or compare one neighborhood with another, requires that someone involved in the project be able to perform batch geocoding on the data.

There are a variety of programs through which this can be done. Most simply, users can contract with a service provider such as ESRI to geocode their data for a fee. PolicyMap has a feature that enables users to upload and geocode point source data where it has a clear street address. Alternatively, the data partner in the team can use one of a number of different GIS (geographic information system) software applications to geocode the data. A third option, Google Fusion Tables, Google’s cloud-based application for visualizing and sharing data tables and maps, is an attractive and free way to geocode lists of addresses without having to use GIS software. To use this option, users must have a free Google Drive account with the ability to create Google Sheets (the Google version of spreadsheets) and Google Fusion Tables.

3. CREATING YOUR OWN DATA

Existing data sources may not tell an organization everything it needs to know. Not only do important things happen in a neighborhood that may not show clearly in the data, but the data itself may be years out of date by the time it becomes available. As an example, the five-year American Community Survey for the period of 2010 to 2014 only became available in the fall of 2015. A resourceful organization does not have to limit itself to the data that it can get from existing public and private records. It can gather its own data, to fill the gaps between what it wants to know and what is available from existing sources.

Self-created data can include field surveys to gather information about concerns like vacant properties or housing conditions, survey interviews, and more informal approaches like focus groups to gather information which can help round out the picture of the community. This section will touch on three separate approaches: condition surveys, sample surveys, and informal ways to gather information. Since there is a great deal of information already available from other sources about the last two, most of this section will be devoted to physical condition surveys.

Physical condition surveys

Physical conditions, as reflected by vacancy or housing quality, are a key element in any neighborhood. While gathering this information requires a commitment of some time and effort, the process is well suited to organizations that have the capacity to mobilize volunteers on a short-term intensive basis, or to hire a small group of workers over a longer period. The period during which the survey is conducted, however, should not be too long — the shorter the better, and no more than three to four months if possible — because conditions can change rapidly in some neighborhoods, and too long a lag between survey activities in different parts of the city may make the information inconsistent and potentially less reliable. Surveys should be regularly updated. While annual updating is usually beyond the means of most communities, surveys should be updated approximately every three years if at all possible.

Property surveys are well suited for summer projects, because not only is the weather usually cooperative, but

students are often available at that time. *The time and cost involved, however, is often much less than many groups expect.* Such projects can also be turned into valuable learning experiences for students from high school to graduate school levels.

Vacant property surveys

Information on vacant properties can be useful for many different purposes. It can be used to assess neighborhood quality of life conditions, it is an important housing market indicator, and it can be used to assess opportunities and priorities for property acquisition, demolition or rehabilitation. It can be gathered by having people walk or drive slowly up and down streets, entering information on a paper form or into a hand-held computer. There are a number of apps available today which allow the information, along with a photograph of the property, to be entered into a hand-held device, and entered into a central database in real time.

Since any parcel survey requires that surveyors go street-by-street, if resources permit, vacant property surveys can be combined with overall housing condition surveys wherever feasible. The additional amount of time and money needed to do so is modest compared to the added value of the information obtained.

The key information to gather in a vacant property survey is as follows:

- Number of vacant structures
 - Condition of vacant structures
 - » Are they being maintained?
 - » Are the grounds maintained?
 - » Are they a visible blight or nuisance?
 - » Are they secured, or unsecured and open to entry?
 - » Are they being actively marketed, as reflected by a realtor's sign?
- Number of vacant lot
 - » Are they being used or maintained?
 - » Do they have evidence of trash or dumping?
 - » Are they fenced?

While in most cases the answer is fairly obvious, one cannot always be sure whether a property is or is not vacant. Even some dilapidated houses with broken windows, cracked foundations, sagging porches and the like are still occupied. Some surveys add a category "possibly vacant."

Surveyors should work with a city tax or property ID map to be able to make sure that each property is correctly identified in the field, so that each property can be matched with its official parcel ID, to make it possible to link the survey data to other information on that property, such as ownership or tax delinquency.

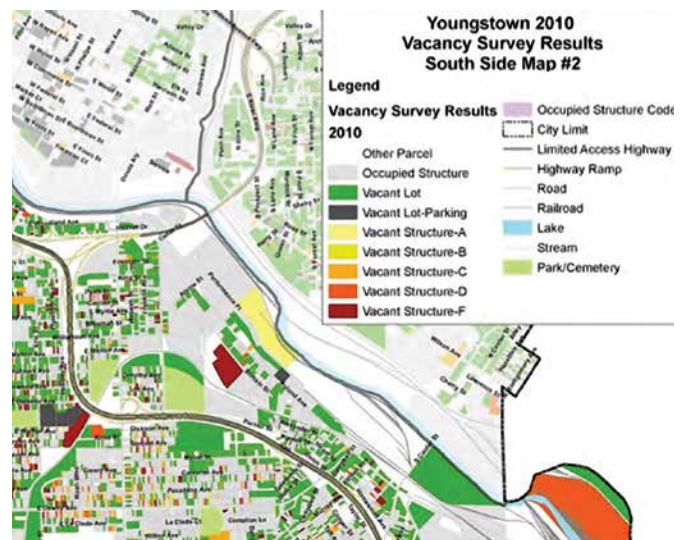
The Mahoning Valley Organizing Collaborative organized a survey of every parcel in the city of Youngstown, Ohio, in 2010 that identified every vacant structure and vacant lot, and classified every vacant structure from A to F, from structures that could readily be reoccupied to those that were an imminent health and safety hazard, based on exterior observation. In addition, the survey identified occupied properties that had a visible and significant exterior condition in need of repair. Figure 2-3 shows a segment of one of the maps generated from their study.



The Mahoning Valley Organizing Collaborative survey report is available at <http://mvorganizing.org/wp-content/uploads/2016/09/2010-youngstown-vp-report.pdf>

This survey, which involved checking over 60,000 separate parcels, was conducted by three full-time surveyors hired by MVOC over the course of a month and a half, or roughly 800

FIGURE 2-3: Mapping vacancy survey results



Mahoning Valley Organizing Collaborative map based on 2010 survey of Youngstown, Ohio. SOURCE: Mahoning Valley Organizing Collaborative—Youngstown Citywide Vacant Property Survey 2010 Results (2011), available at <http://mvorganizing.org/wp-content/uploads/2016/09/2010-youngstown-vp-report.pdf>

TABLE 2-6: Typical checklist for property condition survey

CATEGORY	CONDITION
Foundation	<ul style="list-style-type: none"> • In good condition • Minor repairs needed • Major repairs needed • Complete reconstruction needed
Walls/siding	<ul style="list-style-type: none"> • In good condition • Needs scattered painting, cleaning • Needs major painting, cleaning, or repair • In severe disrepair
Windows	<ul style="list-style-type: none"> • In good condition • Scattered broken window panes or boarded windows • Extensive broken window panes, damage to frames • In severe disrepair
Roof	<ul style="list-style-type: none"> • In good condition • Missing tiles or shingles • Extensive damage, evidence of leaks • In severe disrepair
Porch, deck, or steps	<ul style="list-style-type: none"> • In good condition • Minor cracks, need repainting • Major damage, porch sagging • In severe disrepair
Grounds	<ul style="list-style-type: none"> • Yard, driveway, fencing (if any) attractive and in good condition • Yard and driveway clean, but unattractive or poorly maintained • Yard and driveway in need of repair, cleaning • Visible trash and debris, loose dirt
Other visible conditions	<p>Visible evidence of any of the following:</p> <ul style="list-style-type: none"> • Electrical hazard • Sewerage or drainage problem • Visible other hazards

person-hours. This survey updated a similar survey that had been done in 2008 — comparing the change by area over those two years turned out to be a valuable way of assessing the different trends affecting various parts of the city.

Property condition surveys

Property condition surveys are similar to but more ambitious than vacant property surveys, because they also include giving occupied properties condition ratings based on what is visible from the street or sidewalk. An inventory of property conditions can be used in many ways, including understanding market conditions, assessing the extent of need for repairs and rehabilitation resources, identifying potential health and safety hazards, and more. The survey can also identify positive conditions such as houses undergoing rehabilitation or improvement.

The data can be matched with ownership data, and absentee-owned problem properties identified for targeted action. If a community is planning to do a vacant property survey, they should seriously examine whether they can do a survey of occupied properties as well. *The incremental cost, since one is already putting surveyors out into the neighborhoods, is quite modest.*

Property condition surveys are usually conducted using a checklist of conditions as shown in Table 2-6. Each community should evaluate how they want to use the checklist, and whether everything on the list should be included in their survey. For example, porches are likely to be much less important in a city in the Southwest than one in the Northeast.

There are a number of different ways to use a checklist. One way is to give each individual category/condition a score based on a somewhat subjective assessment of the importance of the category and the severity of the condition; thus, in the checklist, minor problems of painting or cleaning of the walls or siding might be given fewer points than scattered roof problems, on the basis of the assessment that scattered roof problems are more likely to be a sign of more serious underlying issues than comparable wall deficiencies. Another way to score is by ranking the conditions in each category 1 through 4, and giving different weight to different categories; for example, walls/siding might be worth 10 points, but grounds only 5. Finally, surveyors can be asked to do a scan of conditions, and give each property a single score from 1, for properties in all-around good condition, to 4 for severely deficient or dilapidated properties.



Careful training of surveyors is critical to an accurate property condition survey. Surveyors must be trained to understand and identify the different gradations in property condition, and enter the information accurately. This requires the organization doing the survey to prepare back-up materials for the checklist, including photographs of sample conditions and provide training in the field where surveyors can observe actual conditions, rather than simply being instructed in a classroom. Another reason for careful training is that different people, based on their backgrounds and the environments to which they are accustomed, may have very different ideas of what is “minor” or “severe.” Without careful training, survey data may be subject to inconsistencies that seriously diminish its usefulness.

Street-level or visual conditions surveys

A third form of field survey that is often useful, particularly in the context of designing a neighborhood market-building strategy, is to do a street-level or visual condition survey; that is, a survey that looks at the street-level environment of the neighborhood — including the streets, sidewalks, shade trees, front yards and fences that define the neighborhood’s visual image. This image is a major factor in affecting decisions by people to buy homes in an area, or remain in an area when they have the means to move elsewhere if they choose. Table 2-7 on the following page shows some of the elements that go into a neighborhood’s visual image.

In visual condition surveys, it is important to identify not only physical deficiencies, such as potholes or cracked sidewalks, but also indicators of disorder, such as trash in gutters or graffiti. Signs of disorder, *whether or not they reflect actual high crime incidence*, have been shown to have a significant impact on how people — both those who live in the area and those who might consider living there — feel about a neighborhood. Actively attractive features, such as a tree canopy of mature, healthy shade trees, or a series of well-maintained, attractive houses and front yards, should also be noted.

Sample surveys (interviews)

A lot of valuable information can be obtained simply by asking people questions. However, that is not really simple. If the city or neighborhood organization wants to be reasonably confident that the answers are meaningful, the process of asking the questions has to be carried out very carefully to make sure that the questions are correctly framed, the choice of people to interview or the **sample** is representative of the

larger community, and the information properly entered and evaluated. Table 2-8 on page 37 outlines the key questions to ask in designing and carrying out a survey. The document which contains the questions and answers is known as the **survey instrument**.

All surveys other than decennial censuses involve interviewing a sample of people. A sample is a smaller group of people that are representative of the larger group, often called the *universe*. There are well-established statistical procedures to determine how many people have to be interviewed in order to generalize from them, with a defined margin of error. The best-known examples are election polls, which are usually published with a sentence that reads “the data has a margin of error of +/- 4%.” That means that if 55% of the likely voters sampled expressed a preference for Candidate A, the actual preference of all voters is likely to be between 51% and 59%.

The greatest difficulty with doing surveys (other than perhaps raising the money in the first place) is reaching a large enough and representative enough group of people. In recent years that has become much more difficult, as many people do not have telephone land lines, and more and more people appear to be unwilling to be interviewed. While interviewing people in public places, such as on the street or at meetings, can be simpler, it is much harder in that case to make sure that the people interviewed are representative of the larger community. There are techniques available to address that problem, which should be explored with an individual expert in statistical sampling methods. In all settings, though, people tend to be impatient, so the survey should be designed so that the interview can be as short as possible.

There are many resources on conducting surveys, some of which are described in the section on key references at the back of this guide. Many nonprofit community development organizations are members of the Success Measures Project run by NeighborWorks America to help track their progress with neighborhood revitalization efforts; NeighborWorks provides support for member survey activities, including model questionnaires, a data analysis platform, and technical assistance.



For more information about NeighborWorks America’s Success Measures Project, see www.successmeasures.org.



TABLE 2-7: What do you look at in a visual conditions survey?

ELEMENT	KEY FEATURES	KEY POSITIVES	KEY NEGATIVES
Occupied houses	<ul style="list-style-type: none"> • Paint • Trim • Window treatment • Porches and stoops 	Clean, freshly painted and well-maintained houses are positive signals, as is evidence of ongoing property improvement taking place.	Signs of lack of maintenance such as peeling paint are negative signals, as is evidence of poor housekeeping, such as sheets in windows instead of shades. Poorly maintained or deteriorating porches and stoops are particularly visible.
Storefronts	<ul style="list-style-type: none"> • Condition • Window treatment • Signage 	Clean commercial areas without trash or graffiti, professionally made signs and professionally -dressed windows make a positive impression.	Signage or window treatments that suggest “fly-by-night” or shaky businesses, as well as trash or graffiti make a strongly negative impression.
Vacant houses	<ul style="list-style-type: none"> • Presence • Condition • Treatment 	Avoidance of boarding, or use of artistic or sensitive boarding treatments along with ongoing grounds maintenance, can help mitigate negative impacts.	Vacant houses are an inherent negative, particularly when they appear neglected or abandoned. Boarding of doors and windows, while often important for security, signals long-term abandonment, and should be avoided if possible.
Yard or lot	<ul style="list-style-type: none"> • Grass or ground cover • Plantings • Objects • Trash 	Yards with carefully -tended lawns, shrubbery and gardens, and without trash, are positive signals. Vacant lots can be considered positive if they are clearly in use, clean without trash, and attractively planted and fenced.	Yards with dirt instead of grass or other ground cover are a negative signal, as are untended lawns and vegetation, as well as trash or other objects — particularly if in poor condition — in front yards. Vacant lots, whether or not in use, have to be actively rendered attractive in order not to be seen as a problem. .
Fencing	<ul style="list-style-type: none"> • Material • Condition 	Well-maintained picket fences and wrought iron (or imitation wrought iron) fences can be positive.	Both the type of fence and its condition can be an issue; a chain-link or chicken wire fence tends to be seen as a negative signal, whatever its condition, while picket or wrought iron fences will be seen as negative if not well-maintained.
Public realm	<ul style="list-style-type: none"> • Trees • Sidewalk condition • Condition of verge (between sidewalk and street) • Street condition • Trash and debris • Visual signs of disorder 	A canopy of well-maintained, healthy mature trees along the sidewalk and in yards is a positive signal, as is an attractive planted verge between the sidewalk and the curb. Well-maintained sidewalks without breaks or pedestrian hazards, streets without potholes, and an absence of trash in the public realm, are positive signs.	Treeless streets in residential neighborhoods tend to leave a negative impression, while the condition of the sidewalk and of the strip of dirt or grass that is often bound between the sidewalk and the curb is highly visible. The presence of trash on the sidewalk, in the verge and in curbs and gutters, or graffiti on buildings or other signs of disorderly behavior, all make an immediate negative impression.



TABLE 2-8: Key questions in conducting a survey

ISSUE	KEY QUESTIONS
<p>Designing the survey instrument</p>	<ul style="list-style-type: none"> • What do you want to find out? • If the question is answered, will it tell you what you want to know? • Is the question clear enough that the people being interviewed will understand it? • Is the question phrased in a way that it might push people to answer in one direction or another, or is it neutrally phrased? • Is the number of questions on the survey the minimum number needed to get the information you need?
<p>Conducting the survey</p>	<ul style="list-style-type: none"> • Will you interview enough people for your information to be statistically meaningful? • Do you have a plan to make sure that the people you interview are representative of the larger community? • Have you made the survey as short as possible while still getting the information you need? • Are the interviewers well-trained, including providing information on why it will benefit people, and being able to clarify matters for people?
<p>Evaluating the information</p>	<ul style="list-style-type: none"> • Is the survey instrument designed so the answers are clear and easily coded? • Do you have a plan for entering and analyzing the data? • What can the data tell you, and how can that help in planning or decision-making?

At the local level, most colleges and universities have faculty members and/or graduate students trained in this area who can help community organizations or local governments design and carry out sample surveys. Many communities have **data intermediaries**, which are centers, sometimes but not always affiliated with a university, that gather local data. They may also be able to help in this area.

Informal information-gathering methods

Finally, it is worth remembering that not all useful information needs to be quantified. While measurable data is indispensable to understanding a neighborhood's story, it does not tell the whole story. Moreover, data, particularly national data such as the ACS, lags behind the present by a substantial period. Although most neighborhood change is gradual, market and other factors can often trigger a rapid acceleration in the pace of change, which may not be visible from data that can be a year or two old.

A good way to supplement the data is to gather information informally, by talking to selected individuals who are knowledgeable about what is going on in a particular neighborhood, and are willing to speak candidly about it. These individuals can be property owners, realtors, leaders of locally based CDCs and neighborhood associations, police officers, postal service workers, and others who are based in or actively engaged with the neighborhood.

While no one individual is likely to have a complete picture, or to be without personal biases, when it comes to a neighborhood, by creating a framework for listening to many different stories and comparing them, a sometimes remarkably accurate assessment of the changes taking place can emerge. One approach is to identify a panel of 5 to 10 knowledgeable individuals, like a focus group, and organize meetings on a regular basis, either quarterly or twice a year, to discuss what is going on in the neighborhood. Such a meeting should be semi-structured rather than free-form; the organizers should come with a list of specific questions, but allow the discussion to flow freely around each question.



PART II: USING THE DATA

Data can tell a story, but only if you ask it the right questions.

This section will explore some of the different ways in which data can be interrogated in order to extract the stories that it has to tell; in other words, what kinds of useful information can be obtained from data to help people understand what is going on in their neighborhood, and how they can use that information to help make better decisions and come up with better strategies.

The first part of this section looks specifically at that question — what can data tell a user? We show how data can not only provide useful information about specific issues — such as vacant properties or sales prices — but can both illuminate the market conditions in a neighborhood and identify the underlying factors that affect the neighborhood’s real estate market. In the second part, we focus on the practical considerations involved in putting together what we call a **small area data project**, which can be anything from pulling out a particular piece of information from the data to help solve a specific problem, to carrying out a multi-factor analysis to provide a broad **picture** of neighborhood conditions.

TWO KEY RESOURCES

Two valuable books were published in 2014: *What Counts: Harnessing Data for America’s Communities* published by the Federal Reserve Bank of San Francisco, and *Strengthening Communities with Neighborhood Data*, from the Urban Institute. Both contain valuable chapters discussing many of the different uses of small area data for planning and strategy, and both are available for free download or for hard copy purchase at a very modest price (information on downloads appears in the reference section at the end of this guidebook).

A. CHOOSING THE GEOGRAPHY

Before analyzing and presenting any small area data, one must define the geographic area for which the data will be analyzed and presented. As we described above, data from the census or American Community Survey, as well as much other federal data such as Home Mortgage Disclosure Act (HMDA) data, is organized and made available by census tract.



Most other data that one can obtain, particularly local public data such as real estate transactions or tax delinquency, is available as **point source** data, where each data entry represents a different property or an event such as a crime that can be defined in spatial terms. While census and HMDA data all start out as point source data, it is usually either not possible or not feasible for a local data user to obtain that data. Why does this matter?

If one is going to rely entirely or principally on federal data, then one must use census tracts or census block groups, which are subdivisions of census tracts, as the geographic basis for defining the areas for which the data is analyzed and presented to the public. Using census tracts is generally preferable to using census block groups, because many more datasets are available at the census tract level than at the block group level and the data is more reliable.

If, however, one is using local data sources, the option is available to create *community-defined* geographic areas, based on the community's unique characteristics and how the data will be used, and then geocode the data to those areas. This can be very valuable, because it means that the data can be organized in the ways that are most meaningful for the people most directly involved — neighborhood residents and neighborhood-based organizations.

The downside, however, of using community-defined areas is that by defining neighborhoods that are different from census tracts or block groups, one loses the opportunity to integrate federal and local data into a single picture. The one major exception to that rule is OnTheMap data, which allows the user to define the geographies for which the data is provided. Thus, it would not be possible to compare house prices or tax foreclosures with census-derived data such as the area's poverty rates or ethnic/racial mix.

Our experience with many data projects suggests that while there are exceptions, the bigger the city, the bigger its neighborhoods. Since census tracts are about the same size regardless of the city's size, that means that the larger the city, the more likely it will be that census tracts will roughly line up with neighborhood boundaries. In many small cities, however, tracts are a poor fit with neighborhoods. For that reason, for example, the project team for the Trenton data project from which we draw a number of examples in this guide decided to use community-defined neighborhood boundaries and geocode the data to those boundaries.

In order to create a neighborhood map, the project team brought together a focus group of people from the community to sit together over a big city map, and brainstorm the best places to draw the lines between neighborhoods. Large neighborhoods were further divided into subareas for the project.

It is clearly better to use census-defined areas *where it is appropriate to do so*. To determine whether it is appropriate, one should answer three questions:

- Is it important that the boundaries of small areas correspond to, or approximate neighborhood boundaries as understood by community residents?

If the data is going to be used to support a participatory neighborhood planning project, for example, it may be essential that the boundaries reflect residents' perception of their neighborhood. If it is going to be used to assess real estate market activity, it may be less important.

If the answer is no, clearly using census-defined areas is appropriate. If the answer is yes, continue:

- If yes, do individual census tracts or groups of census tracts reasonably correspond to the neighborhoods as understood by community residents?

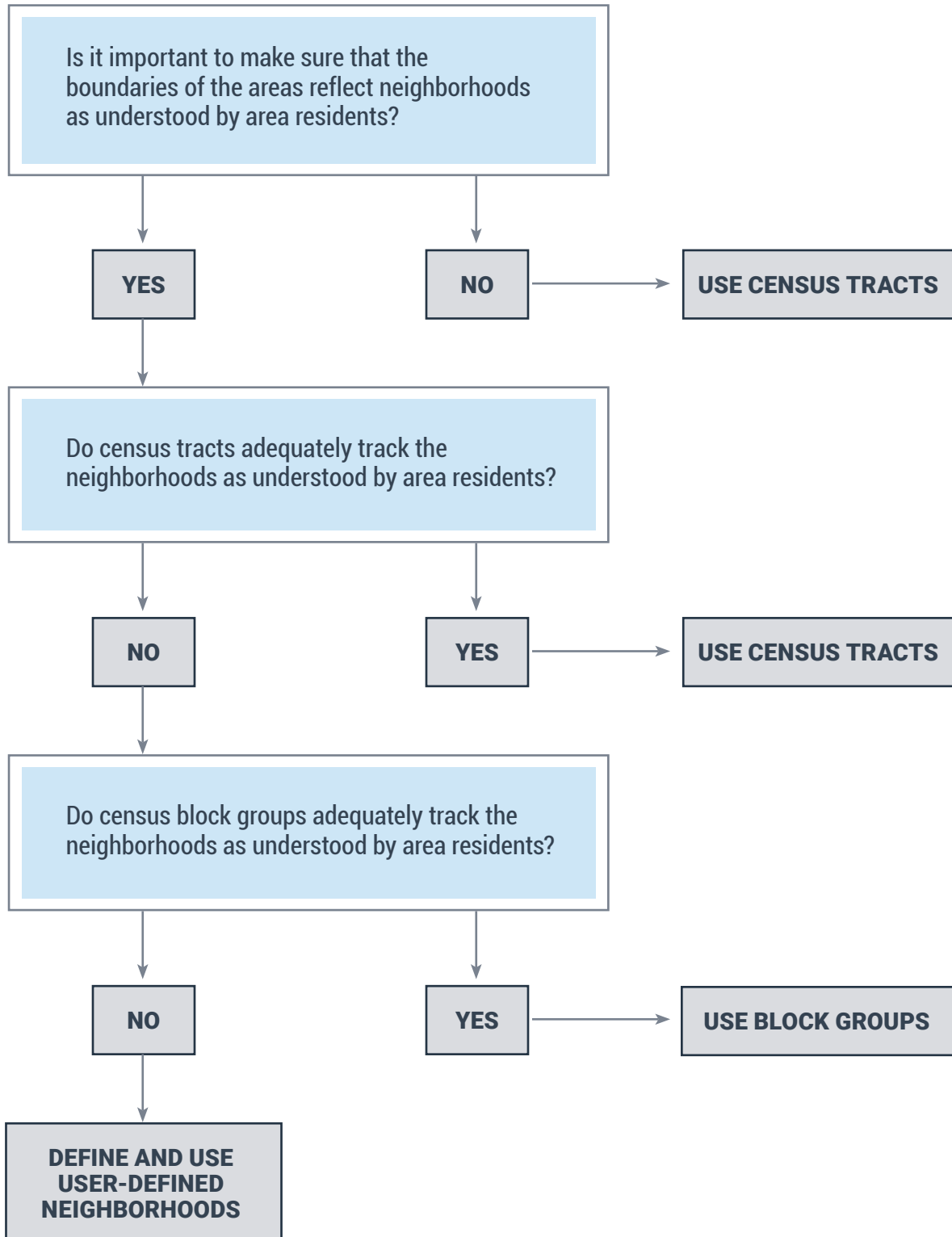
If the answer is yes, clearly using census tracts is appropriate. If the answer is no, continue:

- Do clusters of census block groups reasonably correspond to the neighborhoods as understood by community residents?

Again, if the answer is yes, using block groups is likely to be the best way to go. If clusters of block groups do not correspond closely enough to neighborhoods, then the best way is likely to be to define small areas that are based on how people in the community define their neighborhoods. This way the data will be most meaningful to the people who will be using the data, and affected by decisions based on the data. The process described above is shown in the flow chart on the following page (Figure 3-1).



FIGURE 3-1: Flow chart for determining small area geographies for analysis and presentation of data



IDENTIFYING WHAT YOU WANT TO KNOW

B.

The data that we focus on in this guide falls, generally speaking, into three broad categories:

1. Data that provides useful information about *properties*
2. Data that describes directly how the *neighborhood market* is working
3. Data that describes things going on that *affect* the neighborhood market





While data on the status or condition of a property is largely self-explanatory, it is important to understand how different datasets can help illuminate a particular dimension of a neighborhood’s market conditions. No one dimension can encapsulate the full reality of the market, but each one contributes to the total picture, while also conveying important information by itself.

Since each individual dataset only describes one piece of a complex reality, they are often combined to create a more comprehensive index of neighborhood market conditions. Many cities have created such indices and found them valuable tools for planning and strategy. First, one must define the “small area” for which the data is going to be used, as discussed above. Here we will discuss using individual datasets, and then show how they can be combined into a market index for a neighborhood or other small area.

1. ANALYZING DATASETS

Each of the datasets discussed earlier in this guide sheds light on a particular dimension of the neighborhood housing market. Table 3-1 shows the relevance of different indicators. These datasets can be analyzed in several different ways, including (a) analyzing each individual dataset for relevant information; (b) analyzing one dataset over a period of time to identify trends; and (c) comparing different datasets to see how indicators relate to one another.

Analyzing each individual dataset for relevant information

The datasets listed in Table 3-1, as well as data gathered from surveys of vacant properties or property conditions can often be useful in themselves. For example:

TABLE 3-1: What can one learn about the market from key public datasets?

CATEGORY	DATASET	RELEVANCE
Indicators that measure the market directly	Sales price	Higher prices are an indicator of market strength.
	Days on market or sales volume	Shorter periods between listing and sale are an indicator of market strength. Where days on market data is unavailable, a measure of sales volume can be created by comparing the number of 1 to 4-unit sales transactions to the total number of 1 to 4-unit properties in the area.
	Mix of investors and homebuyers	A higher percentage of homebuyers (buyers for owner-occupancy) is an indicator of market strength.
Indicators that measure factors that affect the market	Mortgage foreclosure	Higher foreclosure rates are an indicator of weakness because they are likely to lead to reduced maintenance, potential increases in vacancy, and neighborhood destabilization.
	Tax delinquency	Higher tax delinquency rates are an indicator of weakness because they indicate that property owners in the area are less committed to keeping their properties, and can lead to reduced maintenance and increased vacancies.
	Tax lien sales or tax auctions	The percentage of properties offered at tax sales or auctions and bought by investors is an indicator of how the outside investor world perceives the neighborhood. A high percentage of properties left unsold and retained by the city or county is an indicator of weakness.
	Crime	A higher crime rate is an indicator of weakness because crime risk (or the perception of risk) affects people’s desire to stay in, or move into, a neighborhood.
	Vacancy	High vacancy rates reflect low demand for the neighborhood’s housing supply and are an indicator of weakness.



ILLUSTRATION: LOOKING AT HOME PURCHASE TRENDS

The figure below shows annual trends for the percentage of single family homes bought by investors rather than homeowners in the city of Trenton and two neighborhoods of similar market strength. Investors were determined by comparing the address of the property to the address of the buyer of record as discussed in Section II of the guide. The figure demonstrates how home trend data can be used to infer information about neighborhood trajectories.

The figure shows, first, that the share of investor purchases has been rising steadily since 2006, although it appears to have leveled off where around 75% of all purchases citywide are by investors. The trajectories for the two neighborhoods, however, are very different. While the investor share of purchases in the St. Hedwig's area has remained about the same, it has risen in Villa Park at a rate even faster than citywide, although still low by citywide standards. Despite the apparent stability of the Villa Park neighborhood, this trend identifies a potential risk factor that could potentially lead to future destabilization.

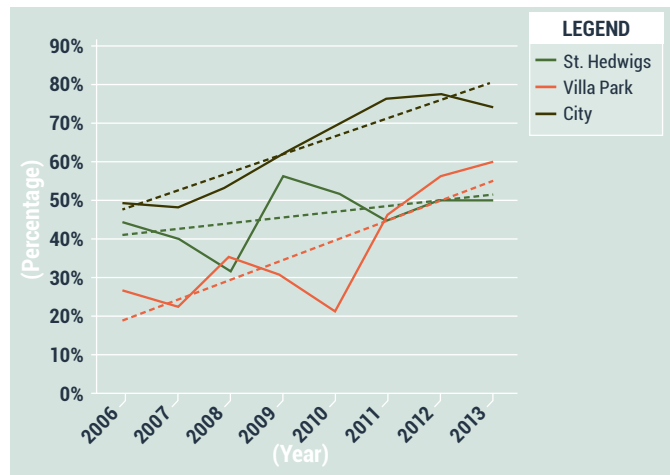
- Information on absentee owners or investor-buyers can be used to make a municipal rental licensing program more effective, by using the information for outreach to make sure that all absentee-owned properties are licensed;
- Information on vacant properties can be used to identify "tipping point" blocks, where scattered vacant properties are likely to be destabilizing entire otherwise stable blocks, and targeting those properties for action
- Information on property conditions can be used to target municipal housing inspection resources more effectively;

More examples are given in Table 3-3 on page 52.

Some of the datasets shown in Table 3-1 to some extent speak for themselves. Although there are always issues of interpretation, when using datasets such as median sales price or crime rates, that information can be analyze and presented directly. Pulling together other key datasets, however, may require varying degrees of data manipulation, or combining different data elements to create new datasets.

A good example is identifying the mix of investors and homebuyers in the market. This data is rarely if ever found as a specific dataset. Instead, it has to be created by the user.

FIGURE 3-2: Investor share of all single-family home purchases by year for city and two selected neighborhoods



Data is for Trenton, New Jersey (2016). NOTE: Dotted lines shown the trend line which averages out the year-by-year variations in the data for each area

ILLUSTRATION: COMPARING VACANCY RATE AND SALES PRICE

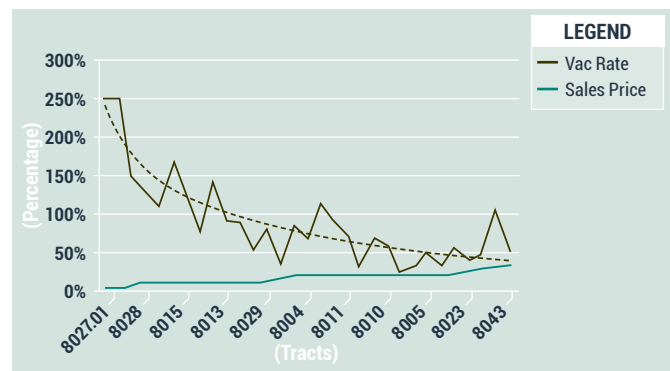
The figure below shows the relationship between vacancies and sales prices by census tract in Youngstown, Ohio, (sales price is shown as percentage of the citywide median; 250% on the graph means that the median sales price in that tract is 2.5 times the citywide median). The data shows, as one might expect, that as the vacancy rate goes up, the sales price goes down. However, it shows something more than that – it shows that when the vacancy rate rises by only a small amount, the sales price goes down sharply (what statisticians call a logarithmic relationship). It suggests that *even a small increase in vacancies can be either an indicator or a trigger of significant neighborhood destabilization.*

Depending on what data is available, there may be a couple of ways to go about it. The best way, by comparing addresses on property records, which is described on page 24, provides a highly accurate estimate. If that data is unavailable, a way to approximate the answer is by comparing the number of single family sales to the number of mortgages reported under HMDA. Since it is well-documented that homebuyers are much more likely to use HMDA-reported mortgages than investor-buyers, the ratio between sales and mortgage gives a rough estimate of the ratio of homebuyers and investor-buyers. Nationally, the ratio is approximately 2:1, or one mortgage for every two sales.

Analyzing a dataset over time to identify trends

The great value of getting time series data is that one can analyze both present conditions and trends. Trends do not predict the future, but can suggest where a neighborhood may be heading. For this reason, when asking for administrative data, it is important to try to get not only the most recent data, but annual data for as many years back as is reasonably available. In some cases, where the agency has been using good (or consistent) systems for managing data for many years, good data may go back 5 or 10 years; in other cases, it may only be available for the current year or a couple of years back. The text box shows an example of how a trend analysis can help illuminate neighborhood conditions, or in this case, a key difference between two seemingly similar neighborhoods.

FIGURE 3-3: Exploring the relationship between vacancy rate and sales price



Data is for Youngstown, Ohio.

Comparing datasets to establish how they relate to one another

Another way to learn what is going on in a neighborhood from different datasets is to compare them to each other in order to see how they relate to one another. The text box offers an example of how a comparative table illuminated an important relationship, in this case between housing vacancy and sales prices.



2. CREATING A MARKET INDEX

Any of the different datasets collected through national sources as described in Section 1, or from the local sources described in this Section, can be combined or *aggregated* to create an index, or overall measure, of conditions in small areas. When used in this fashion, they are generally referred to as **indicators**. Based on the index, the different neighborhoods in a community can be grouped using a market classification or **typology**, clustering neighborhoods with similar conditions into categories. In 2013 South Bend, Indiana, developed a typology which divided neighborhoods in that city into four classes (the map is shown on page 49 as Figure 3-4):

- **Conservation Area:** An area where the housing market is strong and little or no public intervention is needed. The private market should effectively address abandoned properties.
- **Stabilization Area:** An area where the housing market functions relatively well, but is beginning to show signs of decline. Without some attention, the area could decline further. This is a transitional area where private investment still occurs and minimal public intervention is needed in order to help stabilize the area.
- **Revitalization Area:** An area where the housing market does not function as well as it could, and assistance and/or incentives are needed. The area is somewhat distressed and some private investment is occurring; some public investment is needed to improve the area.
- **Reinvestment Area:** An area where the housing market has limited to no demand. This area has experienced significant decline, including widespread vacancy and abandonment. There is minimal private investment and public investment is needed.

As the descriptions indicate, South Bend did its analysis to serve as a tool to determine how best to address the challenge of restoring vacant properties. Typologies can be constructed that break a city's neighborhoods into as few as three, or as many as six, categories. Fewer than three is likely not to be meaningful, while more than six can be cumbersome and hard to work with. Within each category, however, users may want to create subcategories; for example, if the

data used includes trend data, one can distinguish between stabilization areas (to use the South Bend terminology) that are trending upward, downward, or remaining much the same.

The choice of variables can depend on what users consider meaningful, but can include a mix of direct market data such as median sales prices; market-influencing data such as crime or tax delinquency; and demographic data such as population change or poverty rate. Table 3-2 shows the datasets that were used in three neighborhood market analyses, in South Bend, Trenton, and Youngstown.

The overlap between the indicators used by the three cities is actually much greater than it might appear from the table, because some that seem different are actually different ways of getting at the same issue. Understanding what percentage of buyers are investors and what percentage will be owner-occupants, for example, is an important indicator of market strength. Trenton was able to get data that measured the actual percentages with great accuracy, but not the other cities. They used instead the ratio between sales transactions and mortgages as a *surrogate* for that measure, based on the knowledge that most owner-occupant buyers get mortgages (which are reported in HMDA data) but most investor-buyers do not. Similarly, sales volume and days on market are both different ways of estimating market strength in terms of how readily houses sell when they come on the market. The three cities used different geographic units for their analysis. Youngstown used census tracts, South Bend used census block groups, while Trenton used community-defined geographies, as discussed on page 41.

Figures 3-4 through 3-6 on page 49 show the maps that were created by the teams that did the neighborhood analyses in Trenton, South Bend, and Youngstown.

The different datasets that are being used to create the market index are aggregated by using some form of cluster analysis to create a single composite score. Cluster analysis is not a single statistical method, but a general approach based on the idea that by combining different datasets into "clusters," one can identify meaningful relationships between them. There are many different ways to conduct cluster analysis, which range from simple and straightforward to highly sophisticated. It is important that the results be credible from a methodological standpoint, but it is equally important to



make sure not only that the findings from the data, but the process of getting to those findings, can be both understood and used by people from many different backgrounds. Therefore, it is desirable to use a methodology that can be explained to people who do not have a technical background in statistics. The project data partner should be able to offer the team alternative approaches that meet both tests.

Neighborhood market assessments or indices can be used in many different ways. In Youngstown, the assessment is being used as a key tool in developing the city’s new comprehensive plan, and to assist in the development of strategies

for each of the city’s neighborhoods. In South Bend, it is being used as a guide to how to address vacant and abandoned properties, and what strategies should be used to either remove them or restore them to productive use. In Trenton, it is being used to help the city identify the target areas for a pilot vacant property homesteading program. In the future, it may be used to help target code enforcement efforts as well as problem landlord initiatives.

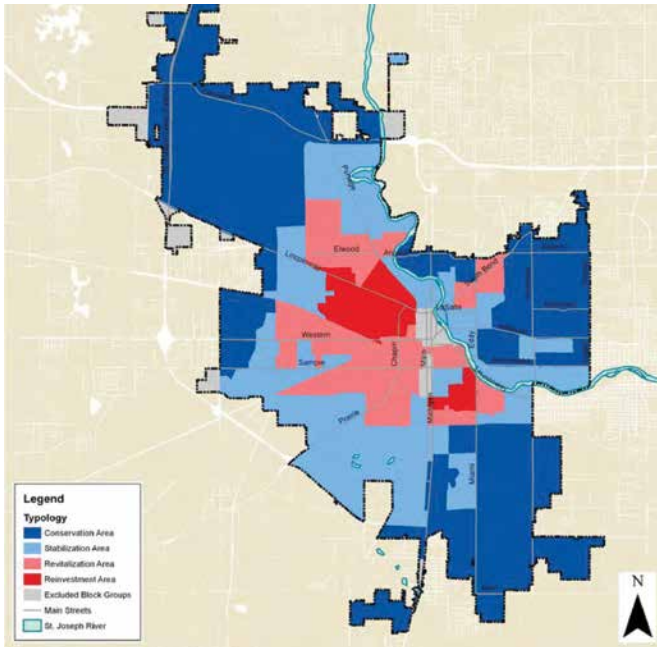


To see how Youngstown uses this data in its neighborhood planning process, go to <http://www.yndc.org/planning>

TABLE 3-2:
Indicators and data source used to create small area market index in three cities

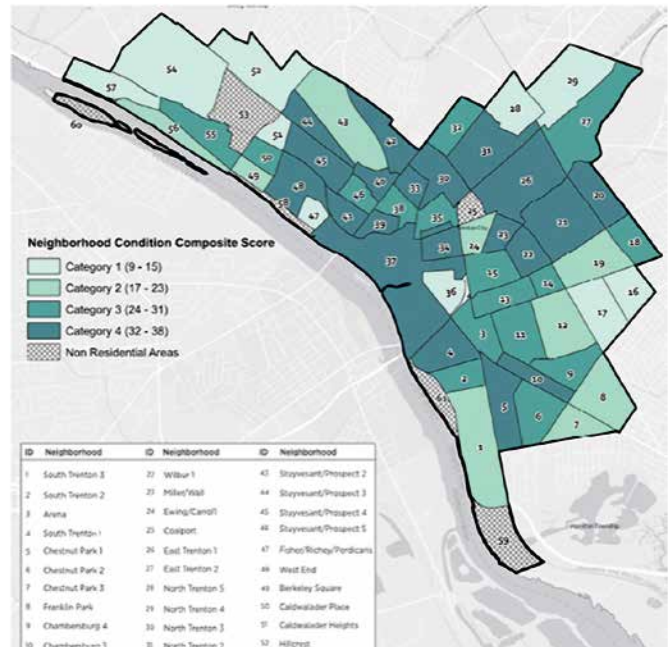
INDICATORS	SOUTH BEND	TRENTON	YOUNGSTOWN
Change in Assessed value	PUBLIC		
Vacant or abandoned properties	SURVEY	SURVEY	CENSUS
Property tax delinquency		PUBLIC	PUBLIC
Properties offered but not purchased at tax sale	PUBLIC	PUBLIC	
Mortgage foreclosures	PUBLIC	PROPRIETARY	
Home sales price	REALTORS	PUBLIC	PUBLIC
Home sales days on market	REALTORS		
Sales volume (sales to properties ratio)			PUBLIC & ACS
Mortgage to sales ratio	PROPRIETARY & HMDA		PUBLIC & HMDA
Share of investor purchases		PUBLIC	
Homeownership rate	CENSUS	PUBLIC	CENSUS
Change in homeownership rate			CENSUS
Crime level		PUBLIC	PUBLIC
Poverty			ACS
Population change			CENSUS
TOTAL NUMBER OF INDICATORS	8	8	10

FIGURE 3-4: Neighborhood market conditions map



Depiction of South Bend, Indiana; created in 2013.

FIGURE 3-5: Neighborhood market conditions map



Depiction of Trenton, New Jersey; created in 2015.

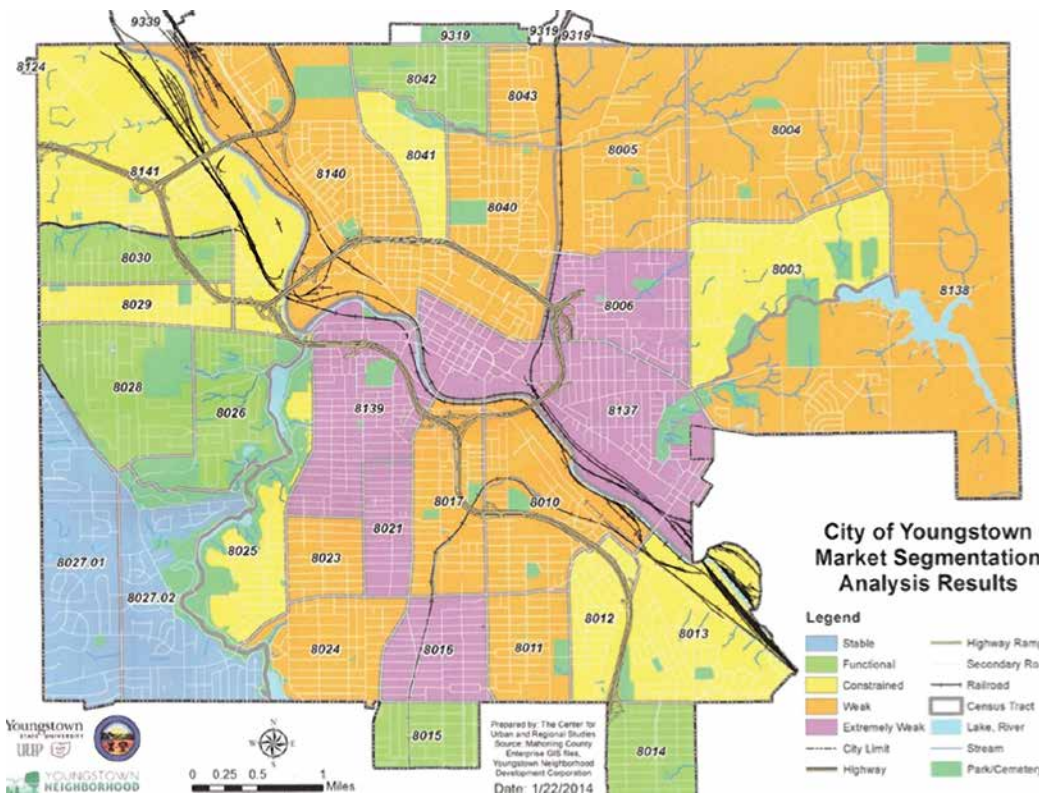


FIGURE 3-6: Market segmentation map

For Youngstown, Ohio, based on 2014 data. SOURCE: Youngstown Neighborhood Development Corporation - "Neighborhood Conditions Report," available at http://www.yndc.org/sites/default/files/NCR_EXEC_SUMMARY_140206_0.pdf



3. BUILDING A COMMUNITY INFORMATION DATABASE

The uses of data that have been described above all involve working with data that has been collected and then aggregated into statistics for a geographic area, whether a census tract, block group, or community-defined area. However, where point source data is the starting point, the value of the project can be greatly enhanced by creating a database where that data can be made available, and the different datasets connected with one another.

As Table 3-2 on page 48 shows, the Trenton project team now has access to the following datasets:

- Individual vacant buildings and vacant lots
- Tax delinquent properties
- Tax lien certificates held by the city of Trenton
- Properties with mortgage foreclosure filings since 2006
- Sale price of properties purchased since 2006 by property and date
- Purchasers of properties since 2006 by owner-occupant/absentee status
- Properties broken down by ownership status (owner-occupant or absentee owner)
- Crime incident reports by location since 2009

Each of the datasets shares common parcel identifiers in the form of individual tax block and lot numbers. Thus, once the necessary technical work is done to set up the database, users can connect any of the datasets to get a multidimensional picture of what is happening to or at a particular location. As additional datasets are identified, they can be added to the database.



At this time, the Trenton database is still a work in progress. Interested readers can see what information is available at <http://www.restoringtrenton.org/>

This database will be designed to zoom in on any and all of the above datasets for individual blocks where a CDC may be focusing its efforts, and to answer many specific program and strategy questions that are likely to come up in the work of a municipal government, CDC, or other stakeholder. A few examples are shown in Table 3-3 on page 52.

The database can be maintained as an internal resource shared by public sector and organizational users, or it can be put online so it is accessible to the general public. It can also be both external and internal, where certain datasets are openly available to any user, while in other cases, passwords can be used to protect data that for legal or other reasons needs to be limited to certain users.

To be most valuable, a database needs to be regularly updated. It is far better if the process by which this takes place, including arrangements with the data sources for periodic submission of data and an individual or organization that will take responsibility for updating, is organized in advance so that it is in place by the time the database goes live.

ORGANIZING THE DATA PROJECT

C.

Once public agencies or organizations in a community have decided that they want to explore the idea of carrying out some form of data-driven investigation of neighborhood or small area conditions, and have decided what type of analysis they want to do, there are a number of practical issues they should think through before proceeding.





A small area data project can be as simple a matter as pulling out particular property data for a particular use, or the analysis of one or two key neighborhood features such as vacant properties or tax delinquency. It can also be a more ambitious effort, such as the creation of a market index for the neighborhoods or census tracts in a city or county, or the development of a user-friendly and interactive neighborhood database. In all cases, however, it will involve gathering, analyzing, interpreting and presenting neighborhood-level information, with the goal of having it used by practitioners and stakeholders in the community. This section will address some of the practical challenges involved in such an effort, and give the reader a sense of how to go about putting the pieces together.

TABLE 3-3: Examples of potential use of property database to address information needs

INFORMATION NEED	HOW DATABASE CAN HELP
Identify individual vacant buildings on city blocks with high homeownership rate for a homesteading or vacant property rehab program	Link vacant building data with dataset of owner-occupied properties
Identify owner-occupants who live next to city-owned vacant lots to market a side lot sales program.	Link vacant lot data with dataset of owner-occupied properties
Identify principal investor buyers for targeted landlord strategy	Sort list of investor buyers by name and by address of record (even though many names may be different, they may share the same address)
Identify blocks with strong homeownership demand for marketing strategy	Identify owner-occupant share of home purchases by block
Identify areas with high potential for land assembly	Link city-owned properties, vacant properties and tax delinquent properties by block

1. ENSURING YOU HAVE THE RIGHT SKILLS FOR THE PROJECT

The process of assembling, interpreting and presenting neighborhood data is not rocket science, in the sense of being a daunting, monumental task, but neither is it entirely simple or straightforward. If it is going to be done well, a number of different skills must be brought to bear on the project. Those skills include the ability to:

- Frame clearly and as precisely as possible the questions that one hopes to answer through the neighborhood or small area analysis.
- Review and evaluate the different data sources available, and determine which are most suitable to be used in the analysis.
- Create data spreadsheets, combine the data into appropriate geographic categories (geocoding), and organize it in ways that enable it to be used to compare neighborhoods, create maps, etc.
- Interpret the data and reach reasonable conclusions about the neighborhood conditions one is trying to measure.
- Present the information in reports, visual displays and presentations to community stakeholders.

Although beyond the scope of this report, an equally relevant skill that needs to be in the mix is the ability to develop effective and productive strategies based on the information that comes out of the analysis.

It is extremely unlikely that any one person can, or should, be expected to have or apply all of those skills. In some cases, a city may have an organization that can apply them all and carry out the project themselves. Where there is no single organization that can take on all aspects of the small area data project, it is very likely, however, that people and organizations exist who collectively can contribute all of the skills and experience needed. Creating a partnership between those people and organizations is often the key to carrying out a successful project.

An alternative approach is to hire an outside firm or organization to do the entire project on behalf of the local stakeholders; that is, assemble the data, analyze it, identify indicators,



interpret them, create a report and/or maps with the information, and present it to the community.

Considerations related to working with an outside firm

One strong benefit of working with an outside firm is that it does not require local organizations to devote significant time and energy to the tasks involved. On the other hand, the direct cash outlay involved with hiring an outside vendor is likely to be significantly greater than when the work is done by a local organization or collaborative.

When relying on an outside vendor, it is critically important for the city or organization commissioning the work to contractually clarify in advance that they can not only use the report or work product, but that the underlying database will also be provided in its entirety in usable form. The city or other organization should not accept any limitations on its use of both the report and the underlying data. The vendor should also be willing to spend whatever time is reasonably needed to make sure that key local stakeholders fully understand the relevance of the datasets used, and the manner in which they are assembled to create the market or other analysis, so that they can both use the product as well as communicate it effectively to others in the community.

Considerations related to working with a local data partnership

A homegrown product offers some significant benefits:

- The approach and the datasets can be more closely tailored to local conditions and local information needs.
- It offers greater opportunity to adjust the scope of the analysis, or add new elements, during the course of the project and afterward, rather than being largely locked into a scope determined in advance.
- It offers the opportunity to build a local database that can be used for more detailed analyses after the initial data project has been completed, and can become a public-access resource for the entire community (this is particularly valuable if any or all of the data is point source data).
- It builds the skills and knowledge base of local stakeholders, by engaging them in an ongoing process of exploring what is going on in their neighborhoods.

2. GETTING STARTED: THE IMPORTANCE OF COMMUNITY PARTNERS

The goal of any neighborhood or small area data project is to ensure that the data is used to help guide decision-making, whether by local government, CDCs, or other stakeholders. For that reason, it is important to engage a broad cross-section of community partners in the process. While local officials and neighborhood-based organizations are likely to be most directly involved and affected, others including major institutions, realtors, lenders and business organizations also have a stake in the outcome. Engagement should take place from the very beginning for two extremely important reasons:

- Having a diverse group of people at the table is important to addressing the first aspect of the project, namely framing the questions one hopes to answer.

Different people have different perspectives on what is important, both in terms of what questions they want answered about a neighborhood, and what decisions should be guided by the data. A city planner may want help in where to allocate scarce CDBG money, while a CDC may want to identify the best target areas for a homeownership strategy. In contrast to tasks which require specific technical skills, this step benefits from broad public participation, making sure that the full range of issues of concern to local government, CDCs, civic associations, and others are on the table.

- Getting people involved from the beginning increases the level of buy-in, and reduces the “surprise” factor later on.

Some of the information that results from the project may be unexpected or unwelcome, particularly if stakeholders feel that they have not been involved, or that the information has, in some fashion, been sprung on them. While how the information is presented, which we discuss in the next section of this guide, is important, most important is to eliminate the surprise factor, by making sure that people are broadly aware of what the project is about and what the data is likely to say.

- Community partners can provide qualitative insights about the neighborhood that may supplement and help interpret the data being gathered in the data project.



Not all information is quantitative, or statistical, in nature. Data that may reflect observations, anecdotes, and insights grounded in personal experience, while not a substitute for statistical measures, is also important and can add valuable dimensions to any data project. Community partners can bring that additional information to bear.

One way to ensure input from key community partners is to create a task force or working group that is involved not only at the beginning, but throughout. In South Bend, the small area analysis was conducted under the auspices and with the active involvement of a Vacant and Abandoned Property Task Force. The Task Force was co-chaired by the mayor and a respected Notre Dame law professor, and included representatives from city and county government, nonprofit organizations, banks and realtors. In Trenton, the study team worked closely with the Trenton Neighborhood Restoration Campaign, a coalition of nonprofits and neighborhood associations, in which city government representatives also participated.

The task force must be small enough to be workable and therefore won't reach everyone. For that reason, it is important to reach out beyond the task force — as well as urge task force members to reach out to their constituencies — to inform and elicit ideas from people who may be affected by the outcomes of the study. Meetings with neighborhood associations and small business organizations, for example, are desirable, as are one-on-one meetings with key individuals, particularly influential or highly visible civic and community leaders such as members of city council, who are likely to weigh in on the study when it comes out and who may have a voice in deciding how it will be used.

3. FINDING DATA PARTNERS

A small area data project will only be successful with the involvement of an individual or organization with the skills to evaluate, analyze and explain the data. In some respects, this is really two separate skills: those involved in the data scrubbing or cleaning, taking the raw data and removing duplications and errors, and organizing it in different ways; and those involved in analysis, which is the ability to look at the data, interpret it, and understand its implications. The first requires technical skills, but those skills are generally within the ability of most individuals with professional training in planning and related areas, graduate students in planning, geography or related fields, and many undergraduates as well.

Analysis does not necessarily require more advanced technical skills, but it requires someone who has had enough experience working with data to understand which datasets are likely to be most useful for decision-making, how they can be organized so they are most meaningful and reliable, and how to interpret what they are trying to say. In addition, some level of practical experience actually working in urban neighborhoods, or working on housing and community development issues, is invaluable in any analyst.

Both of these are often found in two types of organization, one or both of which may be found in many communities:

- Data centers
- University or college research centers or institutes

Many communities have data centers, often members of the National Neighborhood Indicators Partnership. These organizations take many different forms, but all are involved in developing and maintaining, as the NNIP website states, “advanced information systems with integrated and recurrently updated information on neighborhood conditions in their cities.”

Thirty-four cities, counties, or regions currently have one or more NNIP partner organization. While some of the NNIP data centers are based in local universities, most are not, and they include foundations, nonprofit agencies (including one United Way), local governments, and independent organizations, usually supported by local foundations.



Many universities also have research centers, institutes or programs which focus either on local concerns and needs, or on urban and community research. Quite a number of these organizations have become involved with community small area data projects. The University of Miami created an Office of Civic and Community Engagement, with the mission of “engaging the university’s academic resources in the enrichment of civic and community life in South Florida.” That entity is currently partnering with a coalition of CDCs and nonprofit organizations in Miami-Dade County to develop a housing and neighborhoods database that can be used in planning and program decisions. When the Youngstown Neighborhood Development Corporation embarked on their data project, they enlisted the help of the Youngstown State University Center for Urban and Regional Studies, while the Trenton project brought in the Joseph C. Cornwall Center for Metropolitan Studies at Rutgers University in Newark to provide data crunching, and a local experienced analyst to provide the analysis and write the project report.

If there is a four-year college or community college in or near the community which does not contain a formal entity that might become a data partner, it might still have individual faculty members in fields such as economics, statistics, business or geography, who may be interested in the subject, and who can be enlisted in the project.

4. MANAGING THE PROJECT

While this guide will not go into detail about project management, it is appropriate to end this section with a few pointers and observations:

- **Plan ahead**

A small area data project has a lot of moving parts. Make sure that key elements, particularly outreach and participation, are in place early. It is a good idea to develop a project flow chart or critical path chart, identifying all of the elements that have to come together for the project to be a success. Allow time to ensure that all participants in the project team, and all key community stakeholders, are kept aware of the progress of the project, and have the opportunity to weigh in at key decision points.

- **Line up funding**

Community data projects are not particularly expensive, but do cost money. How much is likely to vary widely, depending on how the project is designed, whether it includes physical surveys or just uses public record data, how extensive the outreach and dissemination efforts are going to be, and how much of the work can be done by people who are already on the payroll of one of the organizations that are part of the project team. The last is very important. Having access to AmeriCorps volunteers, for example, can make doing a vacant property survey feasible. While most local data partners, such as university research centers, will require some funds to cover their costs, those costs are likely to be much lower than if the project contracted with a consulting firm to do the same work. Unless the organization leading the project has some flexible funds at its disposal, getting at least some outside funds is likely to be helpful, and in many cases necessary.

The most common sources of funds for data projects are local governments and community foundations. Both have a compelling interest in seeing good data and analysis being used in the community, the first because it can enable limited resources to be used more effectively, and the second because it helps them target their grant-making more productively. The Trenton project described earlier was supported by a grant from the Princeton Area Community Foundation, while data work in Youngstown has been



TABLE 3-4: Elements in a small area data project

PROJECT PHASE	ELEMENTS
ORGANIZING	Assemble project team
	Identify and contact key stakeholders
	Raise funds
	Define purpose of project: why will the data be valuable and how will it be used?
	Define key questions that the data will try to answer
	Create ongoing task force or working group
	Determine geographic areas for analysis (census- or user-defined areas)
DATA ASSEMBLY	Identify potential available datasets
	Contact data sources
	Obtain data for available datasets
	Clean data
	Geocode data and create spreadsheets for analysis based on geographic areas selected
DATA ANALYSIS	Analyze and interpret data
	Create tables, charts and maps for initial review
	Elicit feedback on initial analysis and interpretation from task force/working group
	Finalize analysis
REPORT PREPARATION	Draft and circulate report for review
	Develop rollout plan
	Hold small group/one-on-one meetings with key stakeholders to describe preliminary findings and implications of analysis
	Prepare executive summary and other materials for review
	Elicit feedback on initial analysis and interpretation from task force/working group and other stakeholders
	Finalize report and other materials for distribution/ publication
ROLLOUT	Hold rollout event to release report and other materials
	Meet with stakeholder groups to discuss report findings and implications
	Hold community/neighborhood meetings to discuss report findings and implications

supported both by the City of Youngstown and the locally based Raymond John Wean Foundation.

- **One person has to be in charge**

One individual needs to be the project manager. That person should be senior enough so that she has no difficulty getting access to key individuals, whether among the project partners, county, and local government, or actual/potential funders.

- **When it comes to data, be ready to let it go**

Try as one might, it may not be possible to obtain some datasets, either in a timely enough fashion to meet the project timetable or at all. Other datasets will be available, but are too unreliable or uninformative. In some cases, it might be possible to find another way of getting at the same issue, but in some cases that may not be possible. In either case, it is better to move on, and not spend unreasonable time and effort pursuing problematic datasets. It's far better to work with what one has, or can readily get.

- **Plan the rollout well in advance**

When the project is on a small scale, such as a limited analysis of data for a specific or immediate need, the issue of 'rolling out' the results in a public forum is not likely to arise. Where, however, the project is likely to lead to a major body of information being presented to the community — whether a report, public database, or other medium — thinking through how best to present the information so that it will have the greatest impact, as well as the greatest acceptance, is important.

It is important to carefully think through, once the project has been completed, how to present its findings and recommendations from the project to the community. Preparing for this should begin as soon as one has a reasonably firm completion date. In addition to meetings with key stakeholders to present the findings and recommendations, it is often a good idea to schedule one or more open community meetings about the project.

Table 3-4 summarizes the elements involved in carrying out a small area data project by project phase — beginning with the preparation and concluding with the rollout. With the exception of the roll-out, these phases are basically the same whether the project is on the smallest or largest scale. The difference lies in the scope and complexity of the tasks involved in each phase.

The most important phase, however, is implementation, or what happens after the project is completed and has been rolled out. While it is technically not part of the data project, how the data project is conducted, not just from a technical standpoint but also from the standpoint of how well key community stakeholders were engaged in the process, will largely determine how effective the subsequent integration of the data into the community's planning and decision-making will be.

D. TELLING YOUR STORY WITH DATA

Even the best analysis can be rendered ineffective if it is not presented in a way that people will understand and respond to it. As writer John Allen Poulos puts it “innumeracy, an inability to deal comfortably with the fundamental notions of number and chance, plagues far too many otherwise knowledgeable citizens” (*Innumeracy: Mathematical Illiteracy and its Consequences* (2001), pp3-4).





For this reason, if one is going to communicate the story that the numbers tell, one must use a variety of different methods to do so, recognizing that different people respond to different ways of presenting information — just as some people prefer written directions, others prefer to look at a map, and still others today just rely on a GPS system to tell them where to go.

1. THINKING IT THROUGH

The purpose of presenting information to stakeholders is to tell them a story about the neighborhood, a story that will be useful in terms of helping them not only understand what is going on in the area, but act on the basis of that understanding. When we refer to this as a story, we are not suggesting that it is fictional, or exaggerated. The information must be factual and reliable, but it needs to be brought to life and organized so that it takes a coherent form, not unlike the way a good story has what is called a “narrative arc.”

What needs to be presented is information, not data. Data and information are not the same thing. Data is the raw material, information is the product. There is a phrase often used in the information field, “data rich, information poor,” to describe the situation where too much data overloads the user and just adds to the confusion. Data needs to be presented in ways that turn it into information. That, in turn needs to be driven by what information is needed, who needs to understand it and how it will be used. Users should think through the following questions:

- What are the most important types of information that need to be presented, and to whom?

Small area data can be used in many different ways, and there are many different potential audiences for information about a city’s neighborhoods. These audiences include community residents, other community stakeholders such as social service providers or small business people, public officials, CDC staff and board members, and people in the private sector like realtors or developers who work or might work in the neighborhood in the future. They may also include state and federal officials as well as potential private or philanthropic funders.

Many of these different audiences may be looking for answers to very different questions. Some may want data to see how a particular neighborhood is doing today, or has been trending, while in others they will want to see how

different neighborhoods compare with one another, either on particular variables, or with respect to their markets generally. Some will want an overall sense of where the neighborhood is going, and others will want much more specific information, either about a particular location within the neighborhood, or a particular feature of the neighborhood, such as house prices.

- How familiar are the audiences with using numbers?

Among these many audiences, some are likely to be more familiar with using numbers, or with the subject of how housing markets affect neighborhood conditions than others. The city’s planning staff, for example, is likely to be quite familiar with numbers, but the members of city council are less likely to be. Other people may be familiar with many aspects of their neighborhoods, but may still benefit from help “connecting the dots” to understand the larger trends and conditions in the area. Still others, including some people in the local business community, may not even be familiar with the neighborhoods and their features. Generally speaking, most people, as Poulos suggests, do not “get” numbers automatically.

- What media will you use to communicate the information?

While there have always been many different ways to convey information, the growth of the Internet and social media over recent years have changed the ground rules in ways that are still emerging. While printed materials such as reports and flyers are still important, many organizations today use their websites — either an organizational website or a website dedicated to property or neighborhood information — as well as webinars and other electronic or social media as ways to convey information. Face-to-face communications, whether through formal presentations or informal meetings, are still an essential part of any effort.

- How do people plan to use the information?

As we saw above, there are many different ways information derived from small area data can be used. If it is being used for overall neighborhood planning, as is the case in Youngstown, the information that is most appropriate to be presented is likely to be very different than in South Bend, where it is being used to target vacant properties for different types of intervention.

One basic principle flows from all three questions — *focus the information being provided on what is relevant to the*



audience's needs and interests. Extraneous or irrelevant information, however interesting it may seem, can distract from the critical messages.

How best to organize the information also depends on the purpose of the particular meeting, event, or presentation. In some cases, information is presented — particularly early on in the process — as a way of fostering dialogue, and generating responses from participants, to help drive the project. In other cases, it may be for the purpose of informing audiences about the project's findings and conclusions. Each requires a different approach.

2. PRESENTING DATA

There are many ways to present data, turn data into information, and communicate that information to an audience. Many projects and presentations that have valuable information to offer are unsuccessful in getting that information across because of the methods they use. This is particularly likely to be the case when the underlying basis of the information is numerical or statistical, what is generically referred to as “quantitative” information. The subject of communicating quantitative information is complex far beyond what can be conveyed in this short guide; indeed, information design, defined as “the defining, planning and visualization of the contents of a message with the intention of achieving particular objectives in relation to the needs of the target users” by the International Institute for Information Design, is now recognized as a discipline in its own right.

This section will introduce the reader to the principal tools by which this information can be conveyed, and offer a few pointers on what approaches to these tools appear to be

most effective. We do this in two sections: the first looks at ways of visualizing information, such as maps or graphs; the second explores methods for delivering or sharing the information, such as reports or websites.

Visualizing information

This is a vast and complex subject, and this section is designed to do no more than whet the reader's appetite for further investigation. Readers are encouraged to explore the two books cited in the text box below as well as other materials in the reference section of this guide.

Charts, graphs and tables

Charts, graphs and tables are the bread and butter of presenting quantitative material whether in reports, presentations or web-based materials. We use the term **charts** to refer generically to all presentations involving the representation of numbers or statistics. **Graphs** refer to visual presentations in which **variables** are represented compared visually on a flat plane containing two **axes**, as illustrated in Figure 4-1 (which replicates Figure 9 presented earlier). It is customary to refer to the horizontal axis (along the bottom of the graph) as the X axis, and the vertical axis (along the side of the graph) as the Y axis. The terms chart and graph, however, are often used interchangeably. Figure 4-1 is referred to as a **line graph**, because the variables are represented as lines. The two variables in Figure 4-1 are vacancy rate and median sales price, the X axis represents census tracts, while the Y axis represents the values for both variables relative to city-wide averages. The table shows the relationship of vacancy rate to sales prices across census tracts in the city, and how sales prices go up as vacancy rates go down.

TWO KEY RESOURCES

Two outstanding books on visual presentation of quantitative information are *Designing Information: Human Factors and Common Sense in Information Design* by Joel Katz (John Wiley & Sons, 2012), and *The Visual Display of Quantitative Information* by Edward Tufte (Second Edition, Graphics Press, 2001). They are very different in their approach, and do not always agree on what are good or bad ways of presenting information, but are both well worth reading to gain a better understanding of this challenging subject.

Figure 4-2 is a bar graph, which represents the change in the total number of households, married couples, and married couples raising children under 18, in Baltimore between 1960 and 2012. In this case the X axis shows households by category, while the Y axis is the number of households in each category for 1960 and 2012. The graph presents vividly how, even though the total number of households in Baltimore has not changed much over that extended period, the number of married couples has dropped sharply, and the number of those raising children, even more sharply.

There are many other types of graph, which serve well for different purposes. The third most common form of graph is a pie graph (often called a pie chart), which is very useful for showing the quantity distribution of a single variable, such as the distribution of households or housing units by type in a community. Figure 4-3 shows a pie graph that illustrates the

distribution of rental properties by housing type in the United States, showing that over all of all rental units are in 1 to 4 family properties. There are many other forms of graph as well; an experienced information designer may be able to recommend creative ways of representing data to the project team.

Tables are simply collections of numbers that have been organized in a meaningful way to present information. The same information shown in Figure 4-2 can easily be presented in a table, which appears below as Table 4-1. In order to make the same point in numbers that is made visually in the graph, we have added a column showing the percentage change in each household category. Even so, the information contained in Table 4-1 is unlikely to come across effectively except to people very much accustomed to working with numbers.

FIGURE 4-1: Example of a line graph

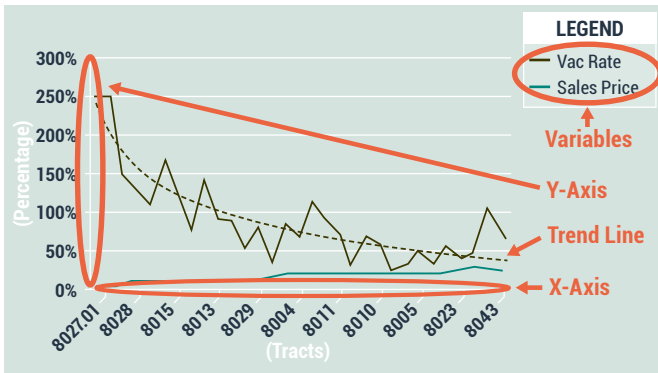


FIGURE 4-3: Example of a pie graph

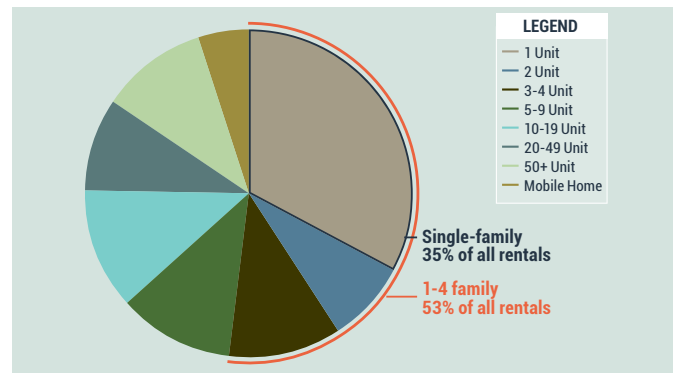


FIGURE 4-2: Example of a bar graph

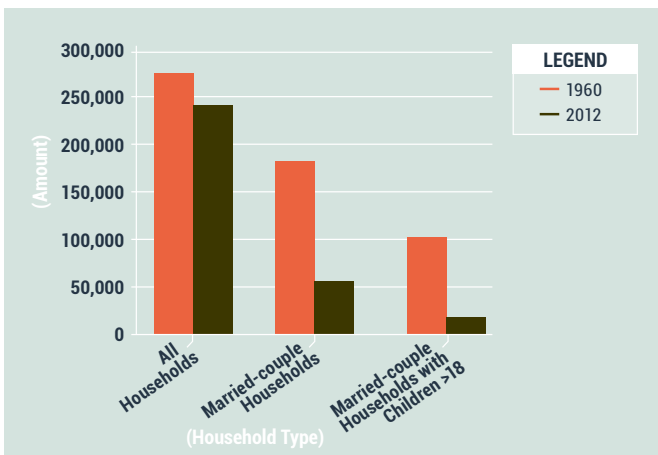
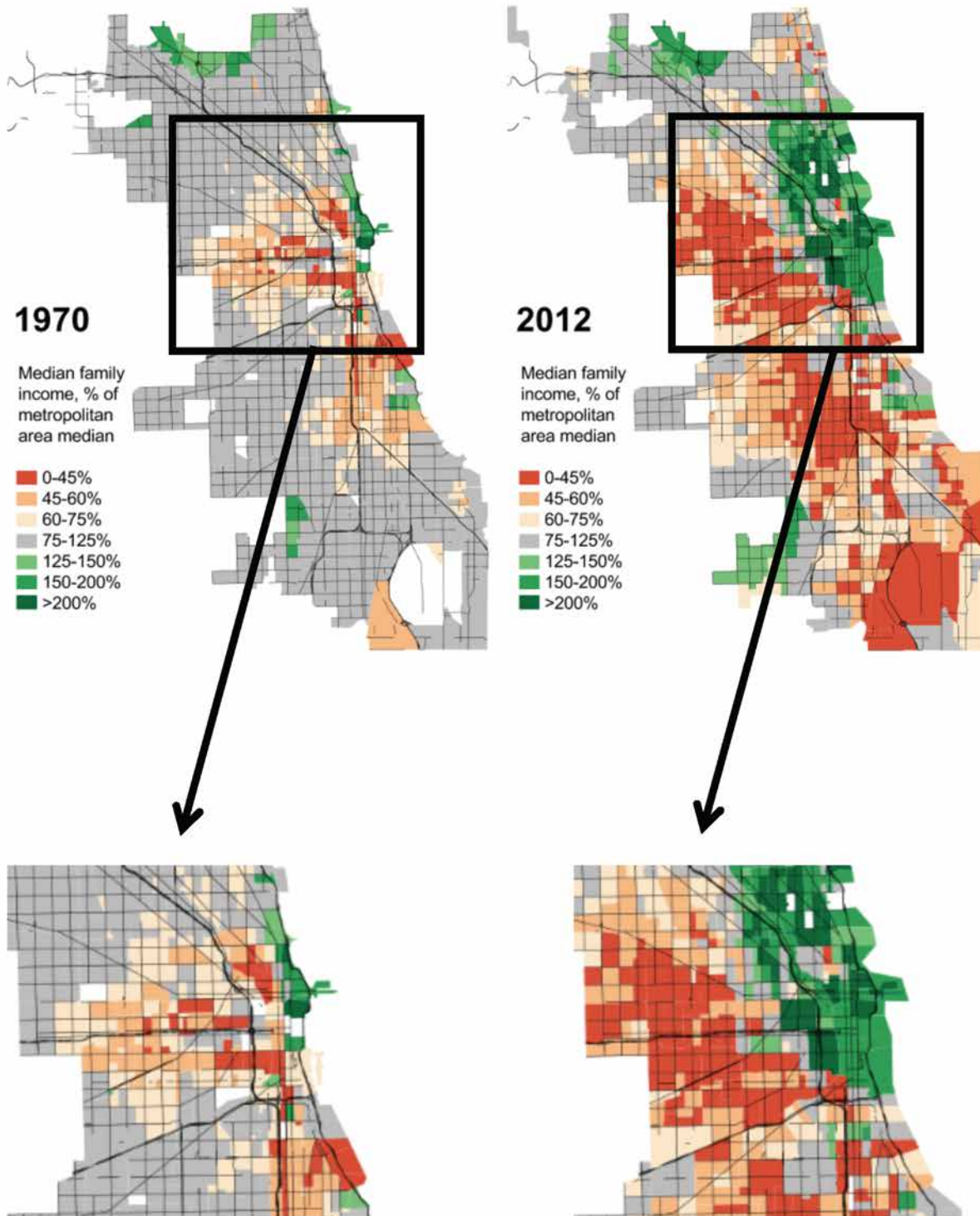


TABLE 4-1: Tabular representation of the information presented in Figure 4-2

	1960	2012	CHANGE 1960-2012
All households	277,243	244,397	- 12%
Married-couple households	184,545	57,086	- 69%
Married-couple households with children under 18	103,744	18,285	- 82%

FIGURE 4-4: Distribution of income by census tract in 1980 and 2009



Depiction of Chicago, Illinois. Source: Ed Zotti, Straight Dope Chicago (2012), available at <http://chicago.straightdope.com>



While a published project report can contain a mix of tables and graphs, it is usually a good idea to concentrate on graphs in presentations, except where the *actual numbers themselves* are the important information that one is trying to convey. When designing a graph, the most important thing is to be clear about what point or points one is trying to make, and to avoid trying to make too many separate points within the same graph. Figure 4-1 makes two distinct points, but they are very closely related. It shows not only that as the vacancy rate goes up, the sales price goes down, but also that when the vacancy rate rises by only a small amount, the sales price goes down sharply. Unless one can do so very clearly, it is usually preferable to make only one point in each graph.

While the above description describes the most commonly used ways that quantitative information is presented in graphic form, there are many other ways of presenting data graphically. Good examples of a variety of different forms of graph and chart can be found in the two books cited earlier, or in the Wikipedia entry for “data visualization.”

Maps

Maps are a powerful way of conveying spatial or geographic information, particularly when one wants to compare conditions in one area with those in another. Figure 4-4, which shows the change between 1980 and 2009 in the distribution of income in Chicago relative to the countywide median, is a good example. While the citywide map shows significant changes, when one zooms in on the area moving north and

northwest from the Loop, the transformation is remarkable — a large section of the city has gone from being largely lower income to predominantly upper income over a period of less than 30 years. One thing that makes the Chicago maps so powerful is the mapmaker’s use of the full color spectrum — from dark red for the lowest income group to dark blue for the highest. Maps that use only one part of the spectrum can still be highly effective, but tend to be less dramatic in their impact.

Color choices can affect how people will perceive a map, and must be carefully considered. A palette which represents the weakest neighborhoods (or the greatest distress with respect to a particular measure) in the darkest colors, and the strongest in the lightest colors, may not unreasonably be seen by some readers as trying to send a signal that the authors of the report may not intend. The palette used in the Chicago maps in Figure 4-4 is neutral in that respect, yet the association of low incomes with red may make some readers think of redlining. This does not mean that that color scheme should not be used; after all, it is unlikely that *any* palette is completely without potential associations, but each choice should be given careful thought.

Maps can also compare be used to multiple variables. Figure 4-5 compares the percentage of loan originations to African-Americans in Cleveland with the location of subprime foreclosures. That information could have been presented in a graph, but becomes more powerful by being shown as a map, particularly for people who are familiar with Cleveland’s racial geography.

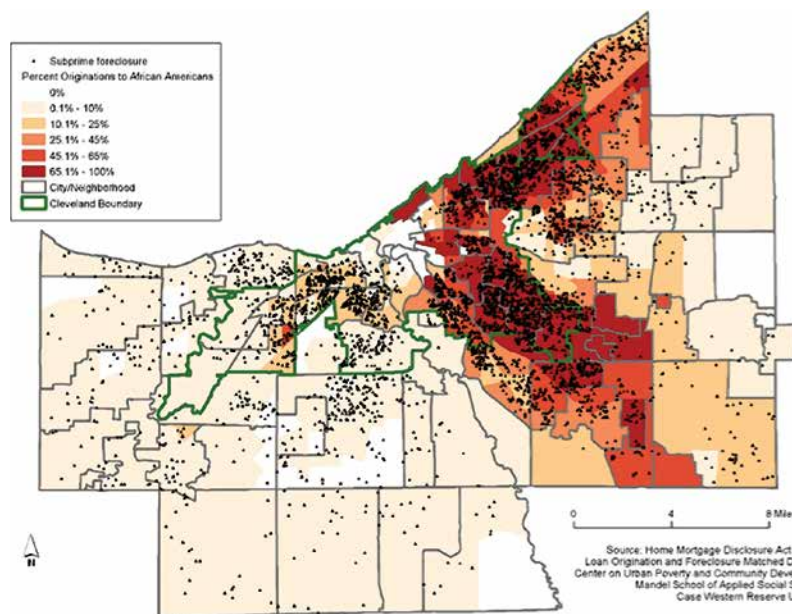


FIGURE 4-5: Percentage of loan originations to African-Americans and location of subprime foreclosures

Depiction of Cleveland, Ohio.

SOURCE: Coulton, Claudia and Tsui Chan, Michael Schramm, Kristen Mikelbank (2008), “Pathways to Foreclosure: A Longitudinal Study of Mortgage Loans, Cleveland and Cuyahoga County, 2005-2008,” Center on Urban Poverty and Community Development, Case Western Reserve University, available at <https://pdfs.semanticscholar.org/7d43/34f720271c-4ba8e441cf29eacb364c49031e.pdf>

Maps can be produced through a variety of means, including software such as ArcGIS and by manipulating Census Bureau TIGER files. A straightforward way of creating a wide variety of maps, that requires no special computer or other skills, is by using PolicyMap, a subscription service described on page 20. PolicyMap permits a user to create a map from any dataset in the site. Although the options for visual display of information are limited, they are adequate for many uses. PolicyMap also has a feature which allows subscribers to upload their datasets to the site, and then use the site's features to create maps.

Infographics

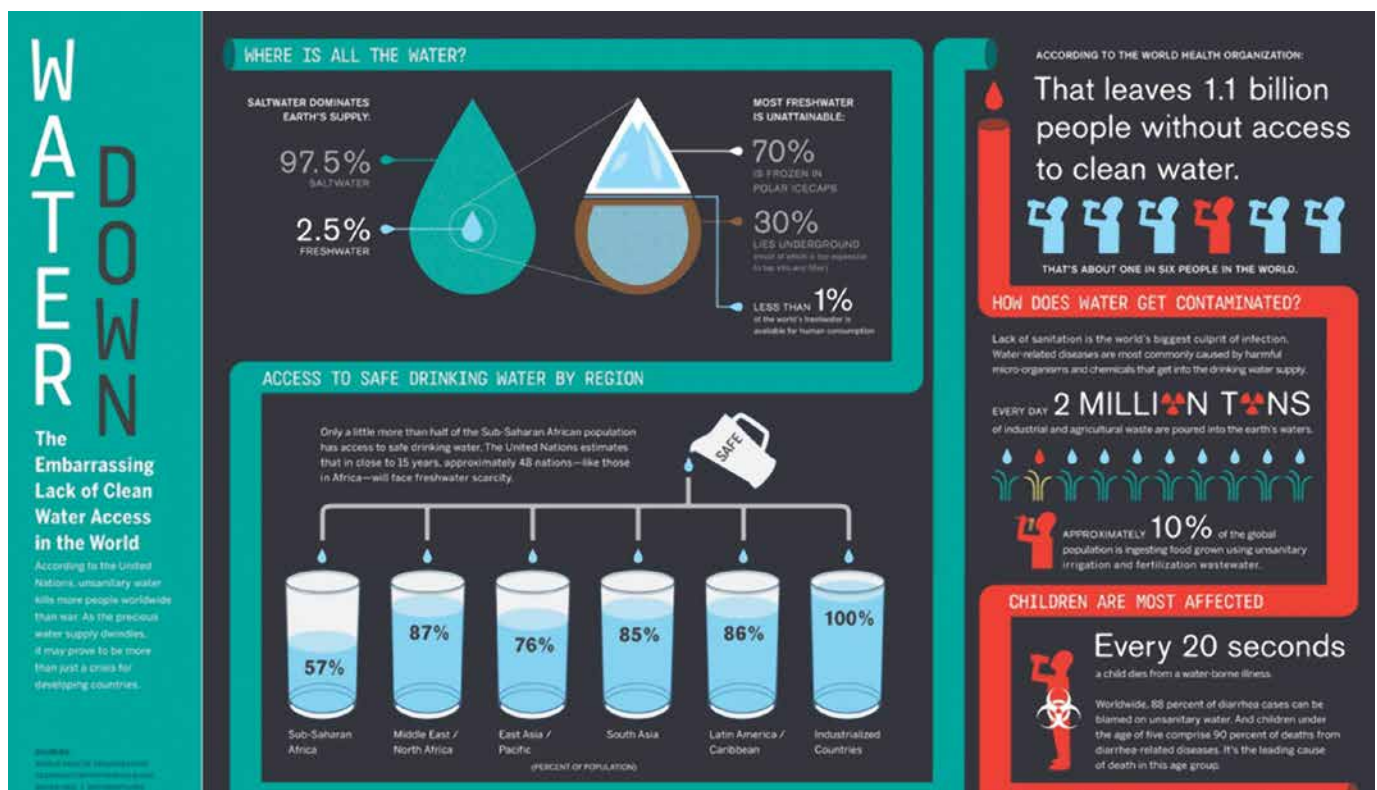
An infographic (short for "information graphic") has been defined as "a representation of information in a graphic format designed to make the data easily understandable at a glance" (from <http://whatis.techtarget.com/definition/infographics>). By using graphic presentation to make the information more visible, or to highlight key issues or points, it can convey information that does not necessarily lend itself to

charts, graphs or maps — although it can use charts, graphs and maps — more directly than simply including it in the basic text. The process of creating infographics is often called data visualization.

Many infographics use the basic chart and graph tools described earlier, but approach them in ways that are designed to highlight particular pieces of information or connections, often combining different types of data representation, or mixing them with images, color combinations and the like, to make them more readily visualizable. Figure 4-6 shows a representative infographic, which illustrates the shortage of clean water access in the world.

An example of a straight-forward infographic created to convey the results of the Detroit property survey conducted early in 2013 appears in Figure 4-7. Through the use of graphics, it highlights key information that the authors of the Blight Task Force report wanted to make sure caught the eye of the reader, and were not buried in the text.

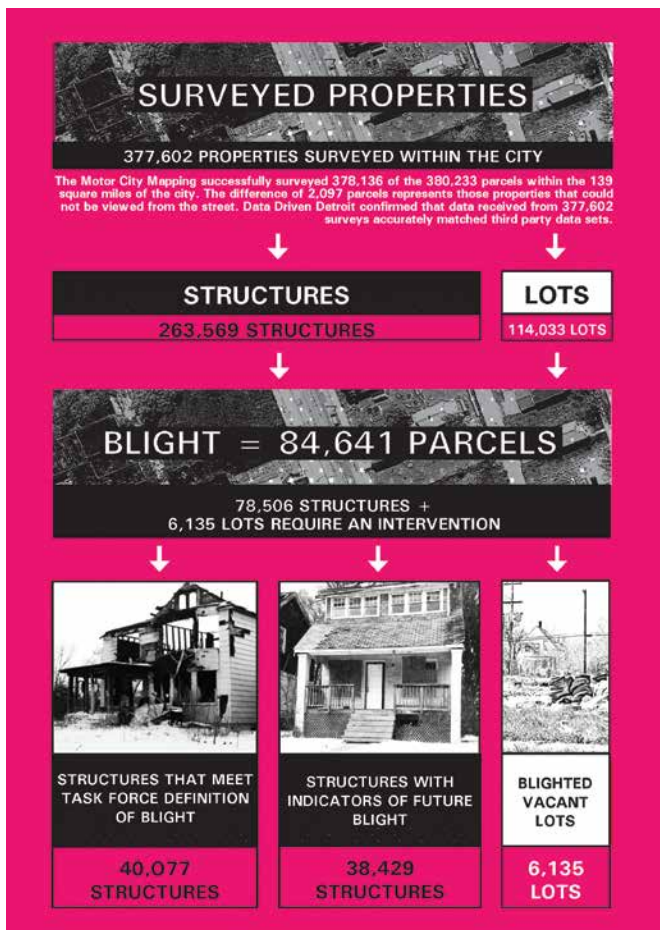
FIGURE 4-6: Infographic illustrating lack of access to clean water



SOURCE: GOOD and Column Five Media (2011), available at <https://www.good.is/infographics/infographic-lack-of-clean-water-access-worldwide>

Infographics can be effective, and can help make important points. They are not easy to do well, however. Many designers of infographics tend to try to put too much information for readers to readily grasp in a relatively small area (something that Figure 4-6 may be guilty of), or getting too carried away with how the infographic looks that they lose track of the fact that it's designed to convey information. It is usually risky for the leaders of a data project to try to use infographics unless they are working with a highly skilled graphic designer; furthermore, if they are, it is critical that they be clear about what information they are trying to convey, and look closely at the designer's product to make sure that it conveys the information clearly and effectively.

FIGURE 4-7: Infographic from Detroit Blight Task Force Plan



SOURCE: Detroit Blight Task Force Plan (2014), Chapter 1, available at <https://s3.amazonaws.com/detroit-blight-taskforce/CHAPTER+01.pdf>

Delivering information

Whatever the way in which the information is organized and visualized, it needs to be delivered to its audiences through some medium, or in all likelihood, multiple media. Although we would not go as far as Marshall McLuhan, who classically declared that “the medium is the message,” the media that a city, CDC, or other entity uses to disseminate its findings is a critical element not only in who receives it, but how they receive it, and how they are likely to make use of it.

Reports

A project report is still the principal end product of most substantial projects, particularly when they incorporate a volume and variety of information beyond that which can easily be conveyed in any one presentation or meeting. A project report provides the documentation of the activity, pulls together both the process and the findings in a single place, and becomes the basic reference guide to what was achieved for future users who may not be familiar with the project itself. It can also serve as a guide to alternative strategies and initiatives that might be suggested by the information that flows from the data.

Like any good reference document, the project report should be written and designed to be as visually attractive and readable as possible, and convey its information clearly and effectively. It should contain easily grasped graphics illustrating its principal points. *The report should be designed by someone with enough professional training and experience to make sure the layout is attractive and the information clearly presented, but not over-designed.* The report should be designed so that it can be easily read as a PDF online as well as in hard copy, and that a copy printed from the PDF on an inexpensive printer will be easily readable.

At the same time, most stakeholders will not get their key information about those findings that are most relevant to them from the report. A report describing a substantial small area data project, including a description of the indicators that were used, summary datasets for each neighborhood in the community, key findings and recommendations will by its nature be a fairly long document, which most people will not spend the time to read. From that standpoint, it makes little or no difference whether it is 30 pages long or 100. At most, they will browse through it, picking up a point or kernel of information here and there.



Although reports should be converted to a PDF file and posted on the websites of participating organizations as well as on the dedicated property information website, if any (see discussion of websites below), the project budget should provide funds if at all possible to have it printed in small numbers (100 to 250 copies). There are still many people who are more likely to read a document they hold in their hands than one they see on a monitor.

However long the report may be, an executive summary should be prepared, which should be kept short while capturing the most important findings and recommendations of the report. The executive summary should be printed in larger numbers, distributed widely, and made available at public meetings.

Other extracts may be appropriate for particular purposes. A report that compares neighborhood market conditions lends itself to the creation of “neighborhood dashboards,” which describe the conditions in an individual neighborhood, and which can be separately laid out and printed up for distribution at neighborhood-level meetings.

With smaller scale projects, where a full-scale report may not be necessary or desirable, it is worth considering more modest printed products that can be distributed at meetings and made available at public places such as libraries, schools or community centers. These can take the form of flyers, pamphlets, or newspaper inserts, which can be reproduced inexpensively.

Presentations

Although many people will look at written materials, either on the web or in hard copy, many people can be more readily reached to inform them about the project and elicit their ideas through face-to-face presentations, either in large-scale public meetings or smaller group meetings. While some of these sessions will be informal, more like conversations, many will involve more or less formal presentations.

Careful preparation is essential. “Winging it” is rarely a successful strategy. Be selective — almost any small area data project generates far more data, and far more interesting findings, than it is possible to share with the audience. Think through carefully beforehand what information is most important to convey to the audience and what information best sets the stage for the conversation that one hopes to have with that audience and focus on it. Much of the material

one is likely to present may be new and unfamiliar, and some of it — especially if it focuses on problems in a neighborhood — may be painful or unpleasant. Avoid the traps of talking down, or appearing to talk down, to neighborhood audiences, or becoming defensive if faced with criticism — however much the presenter may consider it unjustified — from audience members.

Many presentations are likely to involve the use of PowerPoint or similar software. PowerPoint has been controversial in the past. Edward Tufte, an author mentioned earlier, has expressed his objections to the use of PowerPoint in a paper “The Cognitive Style of PowerPoint” (see text box). He concludes that “the PP cognitive style routinely disrupts, dominates, and trivializes content. PP presentations too often resemble the school play: very loud, very slow, and very simple.” That is arguably overstated, but it is true that many PowerPoint presentations are dull and do a bad job conveying the information that is their justification.

One basic and widely observed mistake to avoid is to treat the PowerPoint slides as a “visible outline” of the presentation. *A good rule of thumb is if the slides of a PowerPoint presentation are mostly words, the slides are unnecessary.* In fact, researchers have found that when a slide contains only an image, a graph, or something similar without text, and the presenter provides the words, the audience retains more than when the slide contains text explaining the image. See Andy Goodman’s *Why Bad Presentations Happen to Good Causes* for more information (see Resources on page 67).

Fortunately, presentations of small area data are rich in opportunities to use PowerPoint to provide meaningful graphic images, whether maps, graphs, or photographs illustrating points made in the maps or graphs that tell the story in parallel with the presenter’s words. A few thoughts:

- Keep words to a minimum. We would suggest that as a rule the only text-heavy slide should be the last one, which sums up the key points or “takeaways” from the presentation.
- Where a slide is text-driven, it should still be mostly image. The image should be carefully chosen to support or reinforce the text.
- Graphs, images, and text should be large and clear enough to be easily followed by the audience. In most settings where community meetings will be held, the



size of the screen or wherever the slides are being projected will not be large. PowerPoint, in any event, is a low-resolution medium.

- As elsewhere, tables should be avoided unless the numbers themselves are the story.
- The sequence of slides should be carefully planned to tell the story.

In the end good preparation and careful planning make possible successful presentation and community buy-in, especially if the foundation for trust has been laid from the beginning of the project. Also, it is a good idea to serve food.

Websites and other web-based modes

Websites are an important way to present information. Not only do more and more people use websites as their principal means of obtaining information, but web formats make possible data and information presentations, such as interactive maps, that are difficult if not impossible to create through other means. The City of South Bend has created an interactive map which allows users to identify the locations of:

- Abandoned properties
- Repaired properties
- CDC partner properties
- Properties contracted for demolition
- Properties contracted for deconstruction
- Properties demolished.

RESOURCES

Andy Goodman's book can be downloaded without charge from <https://4good.org/andy-goodman--13/why-bad-presentations-happen-to-good-causes>. It contains many good pointers about presentations generally, not just about PowerPoint.

The paper by Tufte can be accessed online at <http://users.ha.uth.gr/tgd/pt0501/09/Tufte.pdf>. There are many guides to good presentations available online.



The map can be accessed at <https://www.southbendin.gov/government/content/vacant-abandoned-properties-taskforce-maps>

While the South Bend map and other information are on the city's website, many organizations create dedicated websites to present similar information. After organizations in Detroit completed a citywide parcel survey, they created <https://www.motorcitymapping.org> to provide public access to the data. The site contains an interactive map on which users can plot a variety of different types of information, as well as access photographs and other information on individual properties.



PART III: APPENDICES

APPENDIX A.

INFORMATION RESOURCES

This section pulls together a variety of both online and printed resources that may be useful. While many of these resources have already been mentioned in the text, they are pulled together here to be more accessible.

KEY WEB RESOURCES

Census and American Community
Survey data from 2000 on

<http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>

A guidebook published by the Census Bureau to using
and interpreting ACS data can be downloaded at:

[http://www.census.gov/acs/www/Downloads/
handbooks/ACSGeneralHandbook.pdf](http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf)

Data on jobs and workers for small areas

<http://onthemap.ces.census.gov/>

A guide to extracting OnTheMap data

<http://ledextract.ces.census.gov/>

Data on mortgages

<http://www.ffiec.gov/hmdaadwebreport/aggwelcome.aspx>

US Postal Service data on vacant addresses

<http://www.huduser.org/portal/usps/index.html>

Uniform crime reports

<http://www.fbi.gov/about-us/cjis/ucr/ucr-publications>

Accessing New Jersey property and
sales transactions records

[http://tax1.co.monmouth.nj.us/
cgi-bin/prc6.cgi?menu=index&ms_
user=monm&passwd=data&district=1301&mode=11](http://tax1.co.monmouth.nj.us/cgi-bin/prc6.cgi?menu=index&ms_user=monm&passwd=data&district=1301&mode=11)

Additional data resources discussed in text

<http://www.policymap.com>

<http://www.zillow.com/research/data/>

www.socialexplorer.com

<http://www.realtytrac.com/real-estate-guides/foreclosure-laws/>

South Bend, Indiana, small area conditions study

[https://www.southbendin.gov/government/
content/vacant-abandoned-task-force-report](https://www.southbendin.gov/government/content/vacant-abandoned-task-force-report)

Trenton, New Jersey, neighborhood conditions study

<http://www.restoringtrenton.org/#neighborhoods/zqams>

Youngstown, Ohio, small area market analysis

<http://www.yndc.org/planning/market-analysis>



The Reinvestment Fund, a Philadelphia-based nonprofit organization, conducts small area market value analysis studies for many cities

<https://www.reinvestment.com/initiatives/market-value-analysis/>

Youngstown, Ohio, property survey

<http://mvorganizing.org/wp-content/uploads/2016/09/2010-youngstown-vp-report.pdf>

Website with information on Detroit parcel survey

<https://www.motorcitymapping.org>

Two good municipal open data websites

<https://www.dallasopendata.com>
<https://www.opendataphilly.org>

This website, for the Success Measures project of NeighborWorks America, contains many valuable resources on community surveys

www.successmeasures.org

KEY BOOKS

Three general books about data, statistics, and surveys, which may not be directly related to community data projects, but are useful (and often entertaining) background are:

John Allen Poulos. *Innumeracy: Mathematical Illiteracy and its Consequences* (Hill and Wang, 2001)

Nate Silver. *The Signal and the Noise: Why So Many Predictions Fail – But Some Don't* (The Penguin Press, 2012)

Charles Wheelan. *Naked Statistics* (W.W.Norton & Co., 2013)

Two recently published and particularly valuable books about community data are:

What Counts: Harnessing Data for America's Communities (Federal Reserve Bank of San Francisco, 2014). To download or order, go to <http://www.whatcountsforamerica.org/>

Strengthening Communities with Neighborhood Data (Urban Institute, 2014). To download a PDF or ebook, go to <http://www.urban.org/strengtheningcommunities>

Both contain valuable chapters discussing many of the different uses of small area data for planning and strategy, and both are available for free download or for hard copy purchase at a very modest price.

For information about local data intermediaries that are part of the National Neighborhood Indicators Project

<http://www.neighborhoodindicators.org/partners/profiles>

Two outstanding books on visual presentation of quantitative information are:

Designing Information: Human Factors and Common Sense in Information Design by Joel Katz (John Wiley & Sons, 2012)

The Visual Display of Quantitative Information by Edward Tufte (Second Edition, Graphics Press, 2001).

They are very different in their approach, and do not always agree on what are good or bad ways of presenting information, but are both well worth reading to gain a better understanding of this challenging subject.

Two publications on presentations are:

Andy Goodman, *Why Bad Presentations Happen to Good Causes*, which can be downloaded without charge from <https://4good.org/andy-goodman--13/why-bad-presentations-happen-to-good-causes>.

Edward Tufte, "The Cognitive Style of PowerPoint" can be accessed online at <http://users.ha.uth.gr/tgd/pt0501/09/Tufte.pdf>.

There are many other guides to good presentations available online.

APPENDIX B.

GLOSSARY OF TERMS

Aggregated data

Data that has been assembled by a third party from individual point sources and is only available for an area, such as a block group, census tract, or zip code.

Block group

A smaller area within a census tract delineated by the US Bureau of the Census.

Census tract

A geographic unit by the US Bureau of the Census to report and publish census and related data, typically containing between 10 to 50 city blocks, and a population of between 1,000 and 5,000 people. It is the principal small area unit for reporting federal data.

Data cleaning

The manual process of going through a spreadsheet or other database in order to remove duplicate entries and correct inaccurate ones. It is also referred to as data scrubbing.

Data intermediary

A local organization, often based in a college or university, that gathers, analyzes and disseminates data on the city and region in which it is located. Many such organizations are members of the National Neighborhood Indicators Project.

Data scraping

Use of automated programs that allow computers to extract large amounts of data from websites.

Data scrubbing

See data cleaning.

Days on market

The number of days a property remains on the market from when it is first listed until sold. Sometimes used as an indicator of market conditions.

Foreclosure filing

The initial official step in a mortgage foreclosure proceeding.

Geocoding

Defined by ArcGIS as “the process of transforming a description of a location—such as a pair of coordinates, an address, or a name of a place—to a location on the earth’s surface.” The process of converting a large number of addresses into spatial data for analysis or mapping is known as batch geocoding.

Indicators

Data on an area, such as sales prices or foreclosures, that indicate the nature of market or other conditions in the area more generally.

Judicial foreclosure

Mortgage foreclosure where the lender must obtain a judgment from the court before proceeding to foreclosure on a property.



Lis pendens

Literally “lawsuit pending” in Latin. A notice filed with the property record to place parties on notice that a lawsuit or other action has been filed against the property.

Margin of error

A statistical term used to describe the error range for a particular datum based on a particular significance level; e.g., “the margin of error for this result is + or - 5% at a 90% significance level.” That means that there is a 90% likelihood that the actual result is within 5% of the figure.

Non-judicial foreclosure

Mortgage foreclosure where the lender can move directly for a sale of the property without court action after the owner defaults.

Point source data

Data that is available for individual point sources, which can be a specific property or a specific and precise geographic location.

Sample

In surveys, a smaller number of cases or respondents that is considered to be representative of the universe, or entire population.

Small area data

Data, either point source or aggregated, that is available for areas as small as a census tract or smaller area.

Small area data project

A research, planning or other project using small area data.

Survey instrument

In surveys, the questionnaire or interview guide being used to gather data.

Tax sale

A sale where the county or municipality sells liens on tax-delinquent properties to investors.

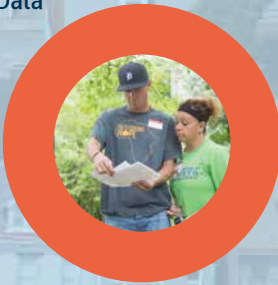
Typology

A classification of different areas on the basis of a common set of variables or characteristics

User-defined (or community-defined) area

A small area such as a neighborhood or sub-neighborhood, where the boundaries are defined by the users rather than on the basis of Census tracts or block groups.

NEIGHBORHOODS BY NUMBERS:
An Introduction to Finding and Using Small Area Data



Center for
**COMMUNITY
PROGRESS**
Vacant Spaces into Vibrant Places

HEADQUARTERS

111 E. Court St.
Suite 2C-1
Flint, MI 48502

NATIONAL OFFICE

1001 Connecticut Ave. NW
Suite 1235
Washington, DC 20036

(877) 542-4842

🌐 communityprogress.net

🐦 [@CProgressNews](https://twitter.com/CProgressNews)

📘 facebook.com/CenterForCommunityProgress

Design by: blankblank.com