Using Data Visualizations to Understand the Scope of Evictions and the Impact of ERA Programs

Good programs run on good data. For cities and their partners working to design and implement effective and equitable eviction prevention programs and services, it is critical to develop data systems that integrate data collection and analysis. Another key component of a data system is data visualization.

With the right visualizations, program administrators and policymakers will have a deeper understanding of the current eviction landscape, the effectiveness of programs and services, and who is accessing and not accessing emergency rental assistance (ERA) and other supports.

This brief provides guidance on the use of data visualizations to assess and improve the effectiveness of local eviction prevention programs such as emergency rental assistance. While not a comprehensive guide to creating visualizations, it offers recommendations for the early stages of compiling and cleaning data with a clear policy question or goal in mind. It is largely written with an aim of helping program administrators and city governments develop internally-facing visualizations, though the methods described can be adapted for public-facing deliverables if added consideration is given to data privacy.

This resource was developed by the National League of Cities and the Stanford Legal Design Lab based on partnerships with the five city participants of the Emergency Rental Assistance Technical Assistance Program. Additional resources for local governments implementing or refining an ERA program can be found at nlc.org/resource/emergency-rental-assistance-toolkit.

The Need for Data Visualization

Visualizing data can assist cities and program administrators in identifying structural problems and trends regarding evictions at the local level, as well as in assessing the impact of interventions such as emergency rental assistance. When datasets are visualized effectively, cities gain clarity about the local eviction crisis, as well as fresh insights on the best approach for eviction prevention programs.

At the most basic level, collecting eviction filing numbers and rates provides an indication of how many evictions are being pursued formally through the relevant court system. By tracking this data over time, stakeholders will be able to implement timely policies and programs focused on stemming rising eviction rates, while also evaluating whether current interventions are having an impact on housing stability.

Building the necessary capacity and data-driven culture to pursue robust data visualization within a local government, or any entity, takes time. It requires resources in terms of data system infrastructure, analysis tools and expertise, along with experience knowing what kinds of questions to pursue through data. But even basic data visualization presents an opportunity to both tell a story based on the underlying information, and to provide actionable policy insights.

This is achieved by first identifying the policy question that you seek to answer, based on the type, quality and quantity of data available.

Examples of Questions that Data Visualizations Can Address

◆ **Who is most at risk of eviction?**
By visualizing eviction rates or filings against potential risk factors (e.g., demographics, neighborhood, economic or housing market trends), policymakers can more clearly determine who in the community faces the greatest risk of being evicted, and therefore how to better target programs or outreach efforts.

◆ **How have local eviction trends shifted over time?**
Depending on the availability of data, local eviction rates can be visualized to help identify changes and trends historically, or in the more recent past (e.g., pre-and-post eviction moratorium, or the launch of an ERA program). These trends can also be explored in the context of a number of different individual factors, including race/ethnicity, neighborhood, family size, household income, or the average amount of rental arrears owed.

◆ **Who is currently accessing available assistance (e.g., ERA), and who is likely to be eligible but not participating in the program?**
As localized trends and risk factors are specified and target populations are identified, this information can be overlayed with ERA program data such as participation and disbursal rates. By visualizing this information, such as by graphing a target group's share of the population relative to their program participation rate, or by mapping it geographically, policymakers can more easily identify gaps in who is being served by a program.

◆ **What impact have policies, programs or outreach strategies had on local eviction rates?**
By visualizing data collected before and after a significant policy or program shift, program administrators and policymakers can identify correlations between those changes and shifts in data. Though often not demonstrating causal relationships or direct results, visualizations can provide insights that would be difficult to identify through raw data alone.

Data visualizations allow stakeholders to develop a more refined understanding of the problem and possible solutions in a way that goes beyond what can easily be garnered through data collection and spreadsheets alone.

By mapping eviction filing or judgement data and overlaying additional information such as race or ethnicity, or household income, cities and program administrators can gain insights into which populations or neighborhoods are disproportionately impacted by evictions and what characteristics may be associated. Armed with this information, city leaders can refine policies, programs and outreach strategies to more effectively reach those households likely to be affected.

Additionally, with a clearer understanding of the nature and scope of the local eviction problem through data visualization, program staff can better make the case for additional budget allocations or more staff to ensure that programs such as emergency rental assistance are able to target the problem more effectively.

TIP: Be mindful of blind spots that may result from incomplete or imperfect data. For example, following eviction trends based on eviction filing rates alone does not account for informal and illegal evictions that may be taking place in the community (e.g., through intimidation or "lock-outs," without an eviction being filed with the court).

Determining Your City's Level of Data Granularity

The usefulness of data visualization depends on the "granularity" of the data collected — the more granular, or specific, the available data is, the more opportunities there are for garnering actionable insights. That being said, useful insights can often even be drawn from data at the lowest levels of granularity.

To determine what kinds of data visualizations are possible, it is important for program administrators to have a good understanding of what data is available and how detailed it is. For example, if eviction filing rates are collected at the state, county, city and zip code level, then the most granular level of data would be the zip code-level data. Granularity can be measured by other characteristics such as time interval (e.g., years, months, weeks, days) or the amount of information associated with a particular entry (e.g., whether there is demographic data associated).

Prior to pursuing any type of visualization, consider the level of data granularity for the city. These levels include low, moderate and high granularity.

Level 1: Low Granularity

At Level 1, only the most basic information is being collected, such as eviction fillings per month or per year.

The limitation of data with low granularity is that it does not provide detailed insight into the scope of a community's eviction landscape. While it may provide an overview of how many evictions are being filed in the city, it does not indicate which neighborhoods or residents are being affected the most.

Level 2: Moderate Granularity

At Level 2, more detailed information is being collected to help understand the city's eviction landscape. In the case of eviction filings, this might include basic demographic information such as race, age and gender, along with geographic data such as a census tract or zip code.

Data that is more granular in terms of geography and demographics allows local leaders to gain a better understanding of where evictions are happening and who is being disproportionally affected.

Level 3: High Granularity

At Level 3, the dataset includes eviction filings, detailed demographic and geographic data, and program data such as participation and disbursal rates. More advanced demographic data includes household income, education level, employment status and household size, in addition to age, race and gender. Geographic data includes residential addresses or zip codes.

As cities or program administrators begin to establish a strategy for data collection, compilation and visualization, there is often an inherent tradeoff between the availability of data and its granularity.

Level 3 data with high granularity is often preferred because it gives analysts the opportunity to be creative and experiment with different types of visualizations. However, in practice, high granularity data is usually the exception rather than the rule. Data with this level of specificity is time-consuming and expensive to collect, and it may raise concerns about privacy.

TIP: Program administrators and data analysts should work to build a clear understanding of what data is already being collected by different agencies, departments or partner organizations, along with collection methods The way data is collected impacts how it can be analyzed and visualized. Additionally, there may be opportunities to streamline or standardize data collection processes, which opens up the possibility of visualizing data from a range of sources.

Compiling Data Prior to Visualization

After determining the level of data granularity, program administrators should compile the available data. This includes identifying the relevant datasets, coordinating with "data owners," and later preparing (or "cleaning") and analyzing the data prior to visualizing it.

Identify the Relevant Datasets

Begin by identifying available datasets related to eviction. This is likely to include:

◆ Eviction filing numbers or rates with residential addresses or zip codes

◆ Eviction filing numbers or rates with demographics

◆ Emergency rental assistance program distribution rates with addresses or zip codes

◆ Emergency rental assistance program distribution rates with demographics

◆ Demographic data such as age, gender, race, household income, education level, employment status, and household size for people who have faced eviction

If staff and stakeholders are looking to explore and visualize data related to other eviction prevention efforts beyond ERA, such as the impact of mediation or diversion programs, consider expanding your search for data to include datasets relating to eviction judgment decisions, default rates, and program-specific data.

Coordinate with Data Owners to Retrieve the Data

Every city or organization has different protocols and guidelines when it comes to data-sharing and privacy. This makes it difficult for cities to create a centralized data system, as there are different protocols that need to be followed before a data scientist or analyst can gain access to the data.

The first step to securing data is to identify the "data owner" for each dataset. The data owner is a person within an organization who is either collecting the data or monitoring the database. Data owners often know what the sharing protocols are. Once the data owner for each dataset has been identified, it will be easier to coordinate the respective protocols and formats of the different datasets into one holistic data landscape.

Common data owners include courts and program staff of eviction prevention programs and services (including both city and community-based programs), such as Court Clerks, Housing Strategy Directors, Housing Coordinators, and Court Administrators.

TIP: In many cases, the primary data owner of eviction datasets is someone within the court system. Pursuing a formal data-sharing agreement with the court should often be among the top priorities for collaboration and partnership with court staff. This could mean arranging to receive regular reports (e.g., quarterly, monthly, daily or real-time) on eviction filings or judgment decisions taking place at the local level. Access to this data would enable program staff to target assistance to households actively facing eviction, while expanding the potential for impactful data visualizations.

Once a relationship with a data owner has been established, coordinate with them to gain access to the relevant data. While protocols will vary, this may include anonymizing the data, sharing encrypted files, and determining where data files will be stored (e.g., on a secured server or an encrypted file sharing platform).

Selecting a Type of Data Visualization

Selecting a type of data visualization will depend on several factors. These include:

◆ The availability and granularity of datasets

◆ The objective of the data visualization

◆ The resources available to produce the data visualization

The objective can often be framed as the policy question you are attempting to answer. For example, an ERA provider may want to gain an understanding of who is most at risk of eviction and where to target outreach efforts. Or a city may be interested in assessing whether a particular program or strategy has been effective in preventing evictions or improving access to a service. Alternatively, cities with access to long-term data may seek to understand how eviction trends have changed over time. Each of these objectives will be best served by a different type of data visualization.

Additionally, consider resource availability from the perspective of both capacity within the team, and the availability of data visualization tools. Common tools for data visualization include:

◆ ArcGIS

◆ Power BI

◆ Tableau

◆ Canva

◆ Excel

Once the team has assessed data availability/granularity, objectives and resources, program administrators can then determine what type of data visualization to produce.

Types of Data Visualizations

Data can be presented in a range of different chart and graph styles. The list below is not comprehensive, but includes the styles that are most likely to be relevant for visualizing data related to ERA programs. Additional guidance is available through the Suggested Resources section at the end of this guide.

BAR GRAPH

This data visualization can demonstrate a relationship between the two variables, typically with a categorical variable (e.g., race and ethnicity) on the x-axis and numerical data on the y-axis (e.g., number of eviction filings or filing rates).

Relevant data to visualize in this way include eviction filing rates by demographic variables such as race, gender or household income.

The limitation of bar graphs is the lack of detail. For example, bar graphs struggle to visualize large, multi-variate datasets and can quickly become difficult to visually interpret.

TIP: If capacity and resources are limited internally, work with a local university or community-based organization that has the necessary data capacity to collect and analyze data, both quantitative and qualitive, and produce data visualization such as geographic maps, bar graphs, scatterplots or timelines.

Example Bar Graph: Eviction Filings by Neighborhood Racial/Ethnic Majority

This representation of monthly eviction filings in Milwaukee, WI was developed by the Eviction Lab as part of the Eviction Tracking System, an open-source system for tracking eviction filings to help monitor and respond to eviction hotspots.

FILINGS BY NEIGHBORHOOD RACIAL/ETHNIC MAJORITY

FILING COUNTS | FILINGS RELATIVE TO AVERAGE

- WHITE
- LATINX
- BLACK
- OTHER/NONE
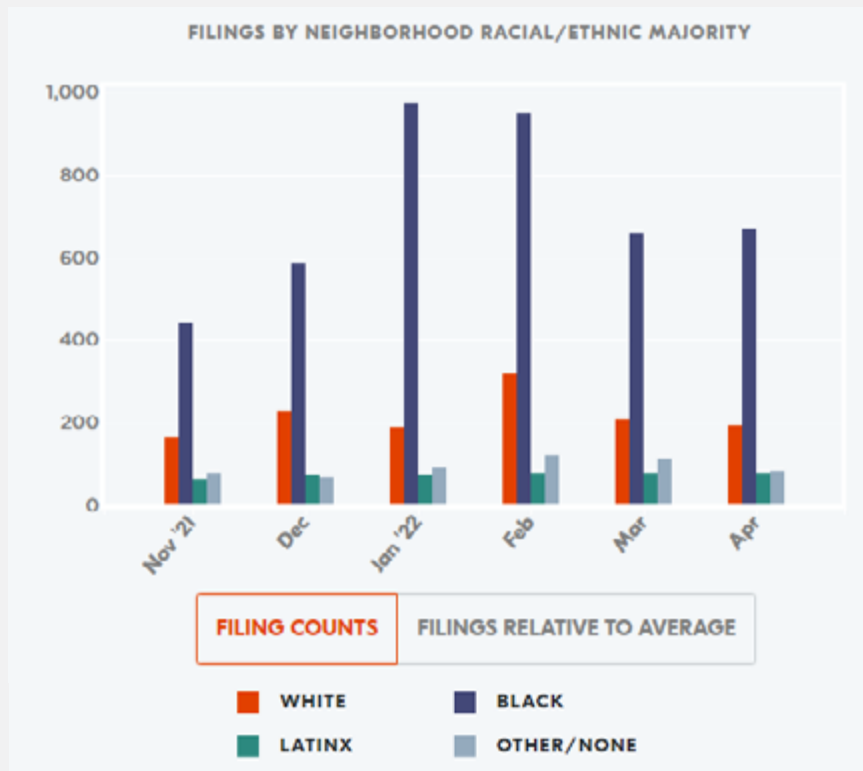
Source: The Eviction Lab at Princeton University

SCATTER PLOTS

This data visualization shows a correlation between two different data indicators along an x-and y-axis. A scatter plot can visualize a certain degree of correlation (or relationship) between the two variables.

Relevant data to visualize in this way include eviction filing rates by demographic variables such as race, gender or household income.

A limitation of scatter plots is that they are predominantly used to examine and display the relationship between just two numeric data points.

Example Scatter Plot: Housing Loss Among Majority Black Census Tracts

This representation of housing loss due to eviction and foreclosure in Forsyth County, NC was developed as part of Displaced in America. This project from New America visualizes historic housing loss at the county level nationwide, and looks to predict where future housing loss will occur as a result of the COVID-19 crisis.

Forsyth County: Housing Loss Among Majority Black Census Tracts
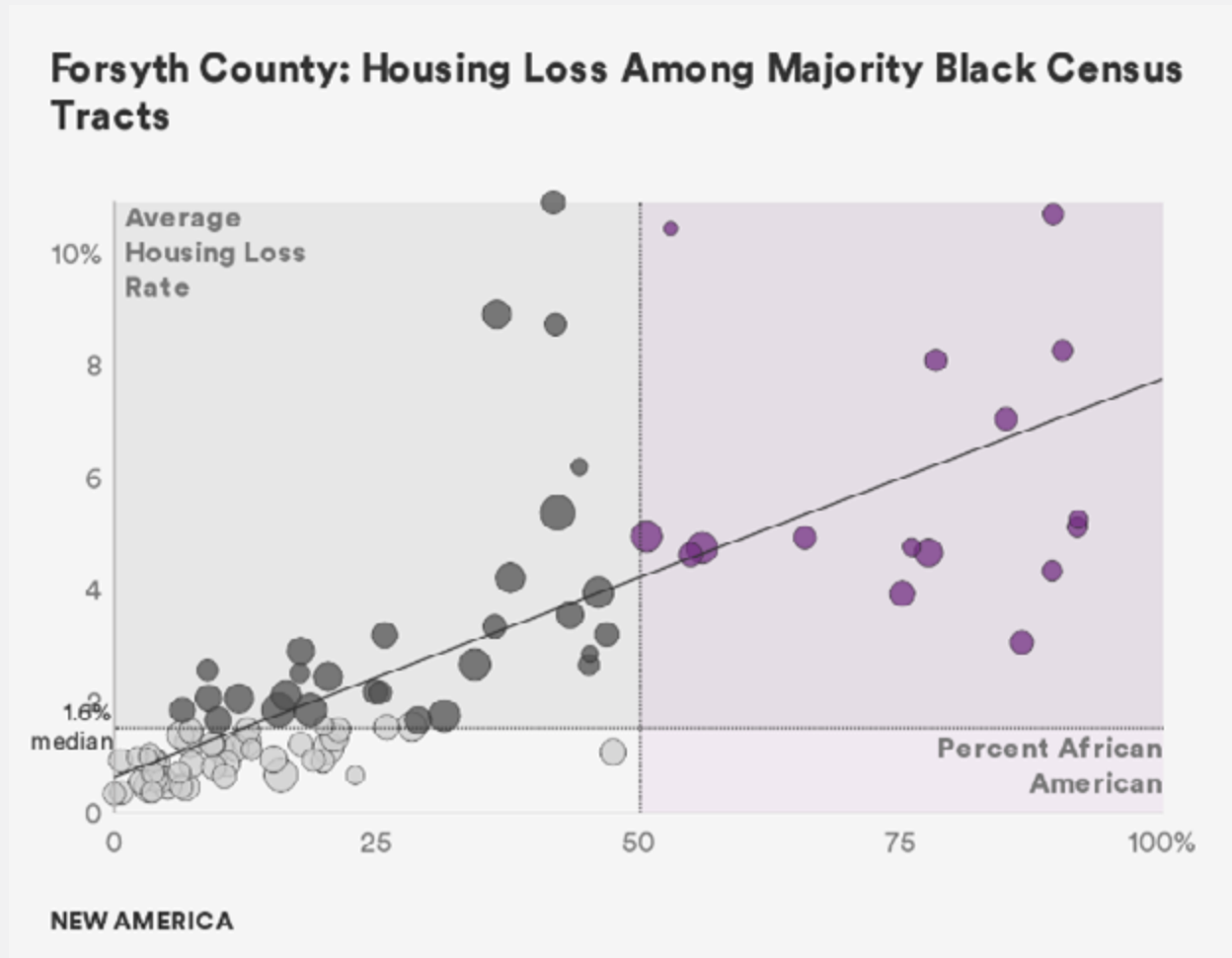
Source: New America-Future of Property Rights (via DataKind)

TIMELINE

This data visualization shows variation in a dataset over time. Time is noted as the x-axis and the y-axis is the data variable that is typically shown as a line or bar.
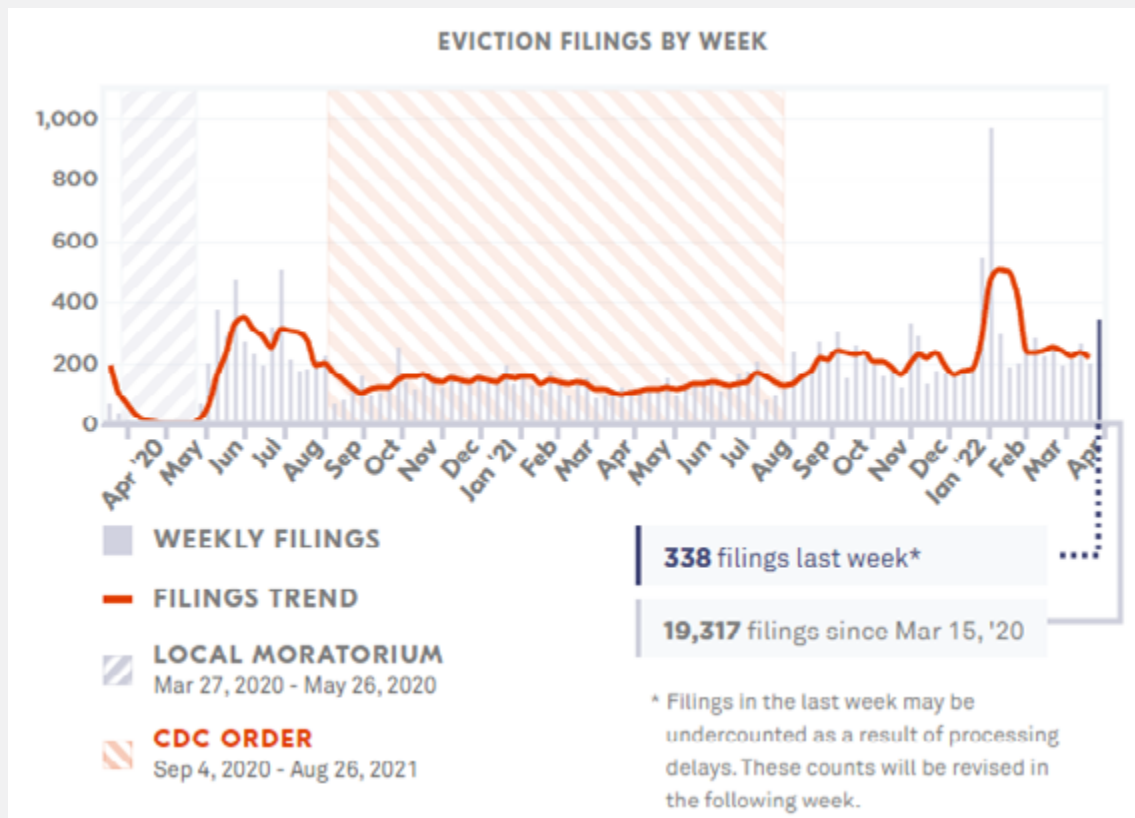
Relevant data to visualize in this way include eviction filing rates over a period of time, such as by months or years. This can be supplemented with indicators of the timing of significant events, such as the implementation or removal of an eviction moratorium, or the launch of an eviction prevention program.

A limitation to using timeline data visualization is that they tend to denote a point in time (e.g., when an eviction filing has occurred, or when a tenant or landlord has received ERA funds).

Timelines often do not showcase detailed geographic data, which can have significant historic trends.

Example Timeline: Eviction Filings by Week

This representation of weekly eviction filings in Milwaukee, WI was developed by the Eviction Lab as part of the Eviction Tracking System, an open-source system for track- ing eviction filings to help monitor and respond to eviction hotspots.

EVICTION FILINGS BY WEEK

WEEKLY FILINGS

FILINGS TREND

LOCAL MORATORIUM
Mar 27, 2020 - May 26, 2020

CDC ORDER
Sep 4, 2020 - Aug 26, 2021

338 filings last week*

19,317 filings since Mar 15, '20

* Filings in the last week may be undercounted as a result of processing delays. These counts will be revised in the following week.

Source: The Eviction Lab at Princeton University

GEOGRAPHIC MAP

This type of visualization allows program staff, policymakers and others to gain a more detailed understanding of the impact of eviction prevention services at a neighborhood level. It is useful for assessing eviction trends and program impact at specific locations through the use of addresses or data at the zip code, census tract or census block level.

Geographic maps can visualize both quantitative (numeric) and qualitative (categorical) data, including eviction filings, demographics and program data.

Geographic maps can offer a more comprehensive understanding of:

◆ What neighborhoods, zip codes, or census tracts are/are not receiving assistance

◆ Where outreach and engagement should be targeted to ensure the equitable distribution of emergency rental assistance

◆ What neighborhoods have the highest eviction filling rates

Example Geographic Data Map: Housing Loss Rate by Census Tract

This representation of housing loss due to eviction and foreclosure in Forsyth County, NC was developed as part of Displaced in America. This project of New America visualizes historic housing loss at the county level nationwide, and looks to predict where future housing loss will occur as a result of the COVID-19 crisis.
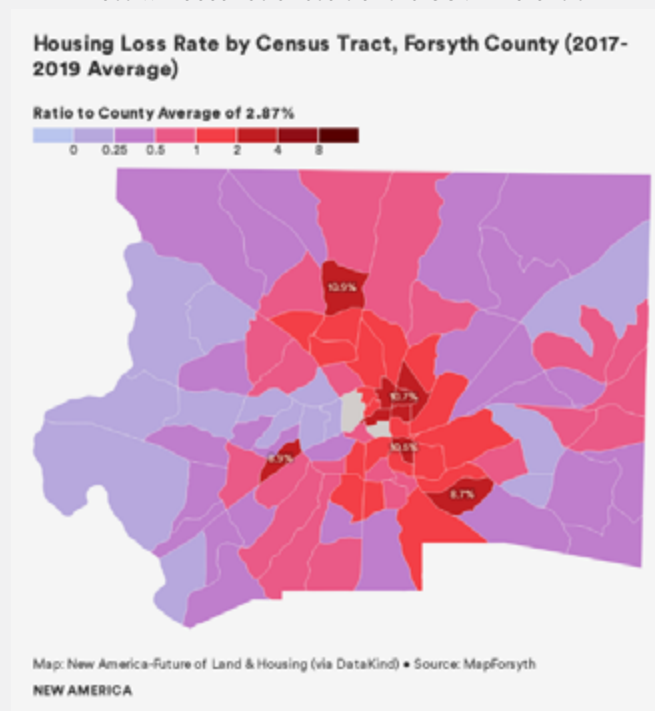
Source: New America-Future of Property Rights (via DataKind)

Preparing Data and Conducting the Initial Analysis

Once the data visualization type has been selected, program teams can begin the process of preparing the data by "cleaning" it, followed by completing any initial analysis, ahead of visualizing the datasets. This will involve acquiring data dictionaries, checking for data entry errors, standardizing variables, validating the data and documenting changes.

Data cleaning is the necessary precursor to any data analysis or visualization. As the adage goes: *garbage in, garbage out*. Without cleaning the data so it is standardized and rid of incorrect, corrupted, poorly formatted, duplicate, or incomplete entries, you run the risk of processing faulty data that will produce faulty results.

When it comes to eviction-related data, it is likely the case that there will be some data entry inconsistencies, as is often the case in courts records. Another common hiccup comes when combining multiple data sets (e.g., data from two ERA providers). If no standard data reporting is agreed upon in advance (e.g., consistent naming for particular variables), these data sets may be difficult to combine.

The process of cleaning, analyzing and visualizing data can be complex, and may require additional resources such as the involvement of a city's data team or outside support. However, while this is not intended as a comprehensive guide, data cleaning will generally involve a few main steps:

◆ **Acquiring data dictionaries**
  A data dictionary should have a record of each variable of interest, its possible values and how to interpret those values. If no formal data dictionary is available, request that information for the variables of interest.

◆ **Checking for data entry errors**
  Particularly when data is entered manually, there is ample possibility for data entry errors. Scan the data for possible errors (e.g., values that seem too high or low, variables that have no clear meaning, and missing values). Flag any potential errors and if possible, consult the person or organization who supplied the data. If this is not possible, exclude any entries that are cause for concern.

◆ **Standardizing variables**

Especially when combining datasets from multiple sources, variable names, categories and collection methods can often vary. For example, zip codes may be labeled as "ZIP" or "Zipcodes"; or one dataset might have demographic data from recipients who "opt-in" to report their race/ethnicity, while another might require recipients to report their demographic information. Know what and why certain decisions were made before determining how to move forward with standardizing data across multiple datasets.

◆ **Validating the data**

Before moving on to "officially" analyzing the data, validate it by running simple, preliminary tests. For example, create a basic bar graph of evictions by month in a given year — if there are any anomalies (e.g., an abnormally high number of evictions in a month that historically has a lower number of evictions), that is good indication that there may be an error in the data.

◆ **Documenting your decisions**

Write down any discrepancies, missing values, or other concerns as they appear in the data cleaning process, keeping in mind what analysis will eventually be conducted. Especially if other people and teams will be working with the data, knowing how the data transformed over time is key for quality assurance purposes.

Data is not purely a numbers game. Often, data roadblocks are due to unclear communication about a particular variable or why certain decisions were made in data entry. Data cleaning is also an iterative process — not only for the person who is manually in charge of manipulating the data itself, but also for stakeholders that have supplied the data. Anticipate back and forth communication across the entities supplying the data and the individuals who clean it, particularly about what, why and how certain decisions were made at each stage.

But ultimately, armed with clean data and a clear question or objective for analysis, program administrators and policymakers will be positioned to use data visualization to both answer and generate additional questions about local eviction prevention efforts and programs such as emergency rental assistance.

Suggested Resources

The resources below provide additional guidance on forms of data visualization, how to select the best type of visualization based on the data and objective, and how to use data visualizations to drive program design and policymaking.

Storytelling with Data: Chart Guide

Learn the basics about different styles of graphs, and how each style can be used to accurately and effectively present data to tell an action-oriented story. This guide offers strategies for styles such as line graphs, bar charts, area graphs, pie charts, scatterplots, slope graphs, bubble charts and more. It also identifies common pitfalls and offers tips for avoiding them.

Data Visualization Catalog

(PolicyViz)
This collection of more than 700 graphic visuals can offer inspiration for creative ways to present data. These real-world examples help to demonstrate a wide variety of graph styles, and ways to use typography, color, platform and design to communicate effectively through data visualizations.

Effective Data Visualizations Should Focus on Narrative, Not Numbers

(The Data Visualization Society)
This article offers four ways that people visualizing data can tell better stories using data, instead of prioritizing solely the visual impact of the design. This includes focusing on the narrative behind the data, and using data to persuade readers with facts instead of simply telling the reader what to believe.

Why is Eviction Data so Bad?: Recommendations for Improving the Local and National Landscape

(New America)
This report highlights current gaps in eviction data, details why these inadequacies matter, and lays out an 'ideal' eviction data landscape. It ends with concrete recommendations for fixing the county's eviction data gaps and offers a framework for creating local eviction databases that feed into a national database.