

DATA DRIVEN PLANNING



Data Driven Housing Planning

Data Literacy and Secondary Data for Housing Planning



The Florida Housing Coalition has produced this guidebook with funding from the Florida Housing Finance Corporation's Catalyst Program. The views expressed in this book do not necessarily reflect the views of the Florida Housing Finance Corporation.

TABLE OF CONTENTS

Introduction	4
Why Do Data Analysis?	5
Individuals, Organizations, and Data.....	7
Primary or Secondary Data	9
Questions to Ask	11
Generating Questions.....	11
Questions for Area Profiles and Background Data	13
Quick and Easy Data Points and Analysis	16
Quick Questions	17
More Advanced Analysis	29
Descriptive Statistics.....	29
Exploratory Data Analysis	32
Forecasting	37
Pivot Table	38
Visualization and Presentation	41
Already Visualized Data	43
Big Numbers	43
Simple Tables	44
Simple Excel Charts	47
Mapping.....	51
Interactive, Online Visualizations	54
Key Sources for Housing Planning Data in Florida	56
Local Consolidated Plans and Other Federal Plans	57
Federal Government Data Sources	57
Other Key Data Sources	66
List of Data Resources	68

Introduction

Governments today have access to far more data than ever before. Publicly available information from the US Census and State sources, data local governments automatically collect in offices like the tax collector or public works, and information generated through broadly available and accessible public outreach systems, are all available to help governments make the best decisions possible. Tools like Excel, ArcGIS, Tableau, RStudio, Survey Monkey, Bang the Table, and dozens of others have changed the way information is collected and processed. Governments have the power to collect and analyze information that would have been unimaginable outside of a super computing laboratory just two decades ago.

This rapid increase in the availability and accessibility of data has created a new challenge., How are local government employees to sift through the huge amounts of data available to find the best sources for answering their specific questions? How much can families in our community afford to spend on housing? How many houses do we have and, more importantly, how many housing units do we need? Is our population growing or shrinking, and how will these trends bear out over the next 10 years? Even more, how do local governments know what questions to ask? Many local governments know that there is a wide range of analyses available, but this array of choices may actually make it more difficult to select and apply the right skills and tools at the right time. With the right skills and tools, this data can save people time and lead to better decisions. Without the right skills and tools, the search for information can hold back decision making, overwhelm local governments with years of time-consuming planning processes, and frustrate elected officials and residents who feel that they are not getting the information or the solutions they need on time or in a way they can access.

The Florida Housing Coalition has spent the last decade working through these questions. Over that time, we have developed systems and methods for quickly getting the information local governments need to make decisions about housing policy. This publication's intent is to summarize this experience so that planners and housing professionals can do the analysis for themselves. This resource is geared towards different people at different levels of expertise, especially for people without a background in visualization, statistics, or mapping. This

publication will begin with a section on “Questions to Ask,” which can help you begin to understand what you do not know. It is followed by a section on how those questions can be answered in an area profile, including a look at some simple data visualizations to help answer questions quickly and efficiently. The publication then turns to still accessible but slightly more advanced analysis and visualization techniques and concludes with a list of key data sources for housing planning.

This publication also serves as a resource on the housing, disaster, demographic, transportation, and health data sources available to local planners, housing professionals, and elected officials. The second half of this document provides a list of these resources, a description of what they are, how to access them, and how to use them.

“Data Driven Planning” is intended to be practical: while many of the concepts in this publication are the domain of statistics, this is not a manual on complex statistical concepts. Instead, this publication provides an overview of secondary data sources most relevant to local planners, particularly to housing planners, and then provides easy to use analysis and visualization techniques so local governments and nonprofits can quickly and effectively gather, analyze, understand, and display data for their community.

Why Do Data Analysis?

Many local government staff feel like they need to do data analysis. However, the actual purpose of this analysis can be hazy. It is something their leadership, commission, or the public wants, something that looks professional and adds credibility to speaking points commonly repeated in the local government context. There are three major reasons for data analysis:

- to improve and evaluate our policies,
- to educate decision-makers and the public
- to maintain legitimacy.

We Do Data Analysis to Improve and Evaluate Policies

Accessing and understanding data is key to being an effective problem solver. Without the information we need at every stage of the problem-solving process, it is impossible to make accurate decisions that have the most impact. Knowing where affordable housing is needed, and for whom, what are the barriers to affordable housing, and what is already being done, is a critical first step in designing effective solutions. Data is also needed throughout the implementation and monitoring stages to ensure that policies are having their desired effect.

Six Sigma's DMAIC Roadmap

One regularly used system for understanding how to solve problems is **Six Sigma's DMAIC Roadmap**. While originally created to ensure manufacturing quality, this process is a helpful way to think about any problem we need to address, including public policy issues.

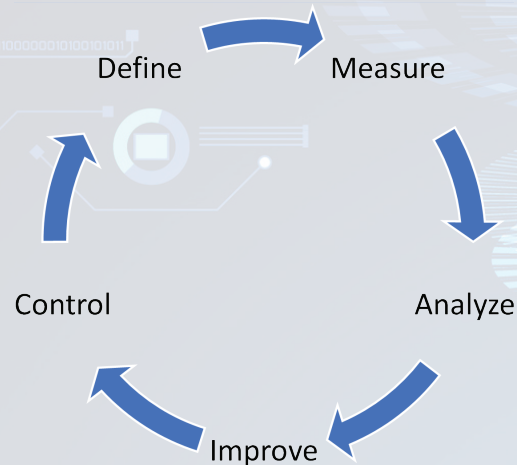


Figure 1: Six Sigma DMAIC Roadmap

Using this system, we first define the problem:

- What questions do we have?
- What are people in our community concerned about?
- What issues do we, as a government, keep running into over and over again?
- What improvement do we want to see?
- How will we know when we are done?

The next step is measurement:

- What is the system able to handle right now?
- Where are the problem areas?
- Where are we going on our current trajectory?
- Where is there missing capacity that we will need to improve if we are going to implement solutions?

Then we analyze:

- What are the underlying causes of our problem?
- Have I considered all factors?
- How are other local governments responding to this problem?

Once these initial three steps are completed, we can move forward to improve:

- Design experiments or “test of concepts” to quickly implement solutions and see how they work.
- Address inefficiencies and poor policies that are inadvertently making the problem worse.
- Make connections and improve knowledge across the board to improve the system.

Finally, once solutions are decided upon, it is important to control the process moving forward to make sure they are implemented correctly, and that implementation is having a meaningful impact. Collecting data on the implementation and effectiveness of programs allows local governments to invest in policies and programs that are working, improve flawed programs and policies, and end initiatives that do not work. Without evaluating changes using data, governments are working in the dark.

As local government planners and administrators, it is important to make sure our work solves the problems of the communities we serve. Data allows us to define, measure, and analyze problems in our community. Data provides a basis to formulate solutions to make improvements. Data helps us to ensure that those solutions are carried out. At every step of this problem-solving process, it is crucial that we collect data. We collect data before we begin planning to define our problems and measure how our system currently functions; during our analysis phase as we choose the most effective solutions; and as we improve and control our policies going forward to ensure our plans are actually implemented and that implementation is making a difference.

We Do Data Analysis to Educate Decision-Makers and the Public

As local government staff, we are often called upon to educate elected officials and the public so that they can make the best decisions possible. In these situations, our job is not so much to solve the community's problems but to provide as much information as possible so that our communities can solve problems for themselves. When data is deployed for educational purposes, it is extremely important that we present data in a way that is approachable and understandable. The section of this publication on "Visualization and Presentation" provides an overview of the ways we can take boring or inaccessible numbers and turn them into something people can grasp quickly and intuitively.

We Do Data Analysis to Maintain Legitimacy

A government is legitimate when the citizens it represents believe that it has a right to govern. Data analysis allows local governments to identify problems, share those problems with their constituents, and make decisions in a way that shows constituents that the local government is acting in their interests.

Individuals, Organizations, and Data

Not everyone in an organization needs to be an expert at collecting and analyzing data, but some level of organization-wide data literacy is key to the effective utilization and application of data. If data is just for a few experts at an organization, people will not be able to use it effectively. Data-literate organizations are organizations where decisions are grounded in data-informed knowledge; where even if specific staff are not the ones managing and organizing information, people know what they have access to; and where employees and leaders are intentionally working to improve their data systems and individual literacy.

For more information on creating data-literate public organizations, see "Data Literacy for the Public Sector: Lessons from Early Pioneers in the U.S" by the Data Foundation, The Data Lodge, and Deloitte.

Data Management

Public organizations collect a lot of data, from the tax appraiser to the public works department, the budget department, and the housing and community development department. Generally, each of these departments store data independently and have different privacy and storage requirements. Governments should work to make data more accessible to people who need it (without compromising privacy), particularly in other government offices. They can do this by asking and answering the following questions and making those answers available to their coworkers and the public.

Where Is Our Data?

Where is data stored, how is it accessed, and who gatekeeps that access? It should be clear to your co-workers what data is and is not available along with who they need to talk to in order to gain access. This can be as simple as an Excel or Word Document that lists the types of data available along with a contact person for that data. Many local governments create open data portals for information they are comfortable sharing widely.

What Does It Look Like?

A major barrier to using data can be formatting. Making sure that in-depth, complicated, or data files have data dictionaries to help people interact with them is key for generating data that is actually used.

Who Can Access It?

There are few things more frustrating than working to access data you know exists only to find that you do not have the correct permissions for it. Alternatively, it is extremely important to not share data with someone who is not supposed to have access. Having a clear set of criteria and protocols for sharing data can ensure that people are able to access the data they need and are entitled to without having to go through an elaborate discernment process every time someone new wants information. Making the steps needed to access clear to the administrators governing the information can also help ensure that data is not accidentally shared with someone who does not have permission to access it.

Primary or Secondary Data

Primary data is data that someone has collected for their own use. You are almost certainly already using primary data—community surveys, focus groups, interviews, and public meetings are all sources of primary data. If you work for a planning department, recent permit information is primary data, just like utility capacity information is primary data for public works.

Secondary data is data that has already been collected by someone else. Unlike primary data, which a planner or researcher needs to collect themselves, secondary data allows planners to quickly and easily answer questions about their community without going through the time-consuming process of collecting information themselves. Federal, state, and local governments generate huge amounts of secondary data, such as Census data, tax collector data, occupational and commuting data, and much more, that everyday planners can use to drive decision making in their communities. While this publication will discuss both types of data, it focuses on using publicly available secondary data. However, local planners should consider how past surveys, internal reports, and data the local government has chosen not to release can be used to increase understanding and help make better decisions.



Who Needs to Know What?

Data Lead

Organizations should have a go-to Data Lead or Team. This person or group should have a wide knowledge of the available data, as well as the technical skills to analyze and present that data.


Everyday Administrators and Staff

Everyone in an organization interacts with and generates data whether they know it or not, and almost all offices in a local government could make better decisions with better access to data. Everyone, from the County Administrator to the Office Assistant, generates data through the creation of timesheets and budgets, aside from staff that generate specific data products such as the Property Appraiser or Permits and Zoning. Staff not directly responsible for data analysis should still know what data they have access to and should participate in improving the systems they contribute to.

Even more fundamentally, staff are community members and organization members that can provide key quantitative information. Your coworkers are a resource for qualitative data and for helping to design and improve surveys, interview questions, etc.

Managers and Leaders

Managers and leaders, like their staff, should be aware of the data that is available to them. Many organizations are shifting to include regular key performance indicators or yearly data updates to help leaders make better data-driven decisions. It is not necessary for leaders to have all the skills to process and understand data, but they should understand what could be available to them and be in contact with staff who have these skills.



How Do These Two Types of Data Complement Each Other?

Secondary data does not negate the need for community input and other types of primary data collection, nor does community engagement mean that local governments can ignore secondary data. Instead, they complement each other. Without secondary data, plans that rely exclusively on primary data miss larger trends, only capture the needs and inputs of a select group of people who attend public meetings and fill out public surveys, and are generally less reliable when it comes to economic and demographic information. On the other hand, a plan that relies exclusively on secondary data without any unique, primary data collection can easily miss key questions, lack the feedback necessary to formulate the best and most community-specific solutions, and fail to receive wider community buy-in. Primary and secondary data interact in two main ways.

First, secondary data gives planners an objective look at mostly high-level statistics, while community input allows planners to understand the nuanced views, values, and preferences of a select group of respondents. Neither is “worse,” instead both types of data fill in different parts of the picture. Secondary data might provide answers on questions like “how many renters are cost burdened?” or “how many heirs’ properties have the tax accessor identified?” while primary data can provide qualitative context: “how have renters in our community made rent after an emergency?” or “do heirs’ property owners know about services available for clearing their title?” In this way, the two data sources fill in different parts of the picture for each other.

Second, primary and secondary analysis often have a “conversational” relationship. Questions generated through one type of data can inform future data collection. Public engagement can turn up issues the local government was not investigating, or a chart may raise questions to ask in your next focus group. For example, the local government may send out a survey and find that people are consistently writing in to say that they cannot find 1- or 2-bedroom units, something the local government was not previously focused on. The local government can then find secondary data to further explore this claim, and then present that data at a public meeting, eliciting further discussion. In this way, primary and secondary data build off each other, creating a fuller picture of the community than could be achieved with only one type of data.

This report primarily focuses on quantitative (numerical rather than narrative) secondary data. However, quantitative data analysis should be collected in coordination with narrative (qualitative) data collection through methods such as community engagement.

Questions to Ask

The first step for any analysis is to know what you are looking for. Ask yourself, what questions do I need to answer? Communities will have their own specific questions, but many questions are consistent across all communities. Once you know what you are asking, you can focus on only providing the analysis that you need, saving time and effort.

Analysts should strive for objectivity in the questions they ask and the analyses they provide. Data analysis should provide a straightforward overview of the facts, including data that both supports and contradicts our preferred understanding and solutions. However, both the questions asked, and the data provided (or not provided), are ultimately political. The choices of who and where to focus, what to prioritize, and how this data affects proposed solutions are all ultimately political choices. While it is impossible to avoid political pressure and bias, it is important to be aware of how data is being manipulated to serve a particular narrative or what questions are not being asked to avoid uncomfortable findings.

Generating Questions

Before we go to a list of specific questions, it is best to start with a brief overview to help us understand how we search for knowledge. An excellent way to think about the questions local governments need to ask comes from former Secretary of Defense Donald Rumsfeld:

Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tends to be the difficult ones.

Here we have two variables, known/unknown, and knowns/unknowns. They can be made into the following chart:

	Known	Unknown
Knowns	Known Knowns	Unknown Knowns
Unknowns	Known unknowns	Unknown Unknowns

Figure 2: Types of Questions

The first two boxes, known knowns and known unknowns, are for knowledge that is much easier to engage with and understand. A lot of the work of questioning and problem solving has to do with getting questions that fall into the second two categories (unknown knowns and unknown unknowns) into these first two categories.

Known Knowns

In this context, known knowns include the core competencies, well-documented programs, and the information the staffer tracks and interacts with every day. It can be helpful to document known knowns so that other people can understand what you know and so you can better see where information might be missing. This type of knowledge does not generate questions per se (as you already have the information), but should still be noted and, where appropriate, visualized and presented to make the process transparent for those who do not share your knowledge.

Known Unknowns

Known unknowns are only slightly more difficult to find than known knowns. This is information that is not “at your fingertips” but already has a well formatted question. You might ask, “What is our area’s median income?” “How many homes does our city have?” “What’s the population of this county?” Most people will not immediately know the answer, but it is something that can be easily found out through a secondary data source. Once the question is formulated, it is much easier to get to an answer than if (as in the two “unknown” categories) we do not know what questions to ask. One of the goals of this report is to move as many questions as possible into this known unknown category by giving you the questions and tools that you need to arrive at the answers as quickly and easily as possible.

Unknown Knowns

Unknown knowns are tricky because they involve information you possess but do not consciously recognize. What are these underlying assumptions and data points that are informing our decision-making but are not recognized consciously? Everyone has underlying assumptions

that color the decisions they make. Some of these assumptions may be true while some may be false. To the extent possible, we should make these unknown knowns clear, at the very least to ourselves. For example, a government staffer might make assumptions on the type of policy changes that would be politically feasible, the character of their community, or which neighborhoods are “bad” or “good.” Again, it is possible for these assumptions to be correct or incorrect, but when left unrecognized they can get in your way or cause you to make poor or irrational decisions. Because we do not know what assumptions we are making unconsciously, it can be hard to grapple with them. Just reading this paragraph though should help you realize you are making assumptions and challenge you to formulate them consciously, moving them to the “known known” category. The best ways to challenge these unconscious assumptions are speaking them out loud to other people who can then confirm or challenge your assumptions and, where possible, testing them against quantitative secondary data.

Unknown Unknowns

As Mr. Rumsfeld pointed out, unknown unknowns are the trickiest of the four. These are questions you know almost nothing about. The only way to identify unknown unknowns is to chip away at the other four categories: identify what you know, answer any questions you have, and identify and test your underlying assumptions. Once you have done this, you can look at your report, plan, or presentation and ask yourself “what is still missing? Where are there still holes in our information? What still does not make sense, even though I have answered all our questions?” This can help you generate new questions, moving your unknown unknowns into the known unknowns category and allowing you to formulate new questions.

Often, the knowledge exists and is accessible to key stakeholders, but may not be accessible to you as the data analyst/problem solver. Again, because you do not know what questions to ask, it can be difficult to dredge up this key information. This type of information is best accessed through qualitative secondary data that provides wide latitude to the subject, such as surveys with “fill in the blank” sections, public meetings, focus groups, and interviews. Also, you will learn over time new questions that you did not originally know to ask.

Questions for Area Profiles and Background Data

The next few sections provide a more in-depth look at how to ask the right questions and gather and present the right data in order to understand your community’s housing needs. Using information like this, it is possible to tailor new and existing programs and policies to get results.

Why So Many Questions?

The following section provides some “starter questions” to help kick off housing research. This list is not exhaustive, but can help you, as the problem solver, begin to move your analysis from “unknown unknowns” to “known unknowns,” and then finally to “known knowns.”

While we think of a united housing market, housing actually serves various functions in a community. Thus, understanding what policies could be implemented to positively improve housing involves understanding a broad range of perspectives. While in reality the housing market is far more complicated, the following four perspectives can help you think about housing in different ways and the questions you might need to ask to understand housing through that perspective.

Housing Is Infrastructure

People need housing to live, and economies need housing that is relatively affordable to attract and maintain workers. When housing is expensive, difficult to find, or located far from economic centers, housing is not functioning well as infrastructure, and the local economy will suffer from high staff turnover, long, expensive commutes, missed work, and a higher cost of living. Housing is most explicitly infrastructure where the government pays for and or maintains housing, but all housing serves this function. Understanding housing as infrastructure raises questions like:

- What income groups are most in need in our community?
- How much does housing cost, particularly compared to how much people can afford?
- How are wages tracking with housing costs?
- Are we seeing job growth or decline?
- Where are new homes being built in our community?
- Is new housing development located near job centers?
- Does the housing available meet the size needs of people in the community?
- How many subsidized housing units exist in our community?
- How many subsidized units are being built?
- How many subsidized units are being returned to the market due to expiring affordability periods?

Housing Is Central to the Economy and Tax Base

Property taxes on housing are a central component of a local government’s revenue, while the roads, utilities, and services provided to homes and the people that live inside of them are the local government’s primary expenditures. Development that generates a positive return on investment for local government – that is, that produces more taxable value than the government

must pay for in services and utilities – improves the economic fortunes of the community, while homes that require more government investment than taxable value are a net drain. In general, denser, older developments generate more taxable revenue, even when they are more in need of repair and occupied by lower-income households, while more sprawling neighborhoods characterized by fewer, larger homes on wide lots tend to be a net drain on city finances even when they appear to be in better repair and are occupied by wealthier households.

Housing development and maintenance is also a substantial portion of many local economies. In a midsized city, hundreds or even thousands of businesses exist to build, repair, assess, and maintain housing stock. Understanding housing as the basic component of a government's income and expenditures leads us to questions such as:

- What is the economic impact of new housing development on our economy?
- Does our housing tax base cover the cost of providing utilities and services to those houses?
- What parts of your jurisdiction are the most productive from a tax perspective?
- What areas are the least productive?
- What are the barriers to new housing development?
- Are there laws or policies that are driving up housing costs?
- What organizations are working on planning in the community?
- Where has money been spent in the past?
- How are our rules or laws making housing development easier or more difficult?
- Do we have the staff to manage more programs?
- Can we afford more staff if we need them?

Most Households' Primary Investment

In addition to being key infrastructure for the economy and a primary source of revenue for local governments, homeownership serves as the primary source of equity for most American households. Homeownership ties people to their community and homeowners are more likely to stay in and engage with their community than renters. Homeownership serves as a source of equity when families start businesses or send their children to college. Owning a home outright (after paying off a mortgage) is by far the most affordable way to own a home and a major source of housing that is affordable for the elderly. On the other side, not in my back yard-ism (NIMBYism), where homeowners oppose all new construction in their area because it might lower property values, is a driving force in American politics. While the research on the impact of subsidized housing on property values is mixed, local government officials should understand the way homeownership functions as a source of equity and work to provide access

to homeownership to as broad a base as possible. Understanding housing as an investment prompts the following questions:

- How much money do we have to spend and how do we want to spend it?
- Is housing development keeping pace with demand?
- Are housing prices rising, falling, or holding steady?
- Is there a geographic dimension to housing prices?
- How expensive is new development?
- What programs do we have to help renters move into homeownership?
- Are there homes at a variety of sizes and price points to help people enter the housing market?
- Are there parts of your community that have experienced redlining, segregation, or other forms of historic racism?
- Are there formerly low-income areas that are seeing growing cost pressures?
- Are there specific demographics (youth, people with disabilities, racial minorities, etc.) that have less access to homeownership?

Home

Ultimately, while most people experience housing as an investment, a utility, and a source of wealth, people experience housing most viscerally as “Home.” People have a deep, emotional connection to the places they live, and local governments should understand that housing is more than just an economic resource to the people they serve. These questions are generally answered through narrative, also known as qualitative, data.

- Why did people choose their neighborhood?
- What drew them to this community?
- What do community members want to see change?
- What do they want to stay the same?
- Do they see a future for them, their families, their children, and their friends in this community, or are there issues that make that future doubtful?

Quick and Easy Data Points and Analysis

This section provides an overview of the most easily accessible housing and demographic data that the Florida Housing Coalition regularly uses in its analyses. While this data lacks the depth of a more intensive analysis, in many cases these top line numbers are enough to identify housing problems and begin the process of formulating relevant solutions. These data sources require little to no transformation/visualization to make them accessible and are free to use.

“Avoiding excess work benefits you and the taxpayers you serve. Creating charts no one will look at, generating hundred-page long plans, duplicating effort, and answering irrelevant questions is unhelpful and can actually make your analysis less useful. Instead, local planners should follow the old adage, “work smarter, not harder.” Ask yourself: “How can I get to the answers I need as quickly and efficiently as possible with the least amount of work?” and then read the following section, thinking about how you can access and apply this data in your community.

Quick Questions

Before getting more specific, it is good to start with some fundamental questions to start defining housing needs in your community. The following questions are fundamental to housing planning and generally the Coalition’s starting place:

- How many housing units are there in our community?
- What income levels are most in need of housing?
- How many people in our community cannot afford housing?
- What are the neighborhoods or areas of highest need in our community?
- Where is housing affordable? Where is it expensive?
- Where are housing and transportation costs highest?
- How quickly are housing prices rising?



Data Sources for Quick Answers

Consolidated Plans and Analysis of Impediments to Fair Housing Choice

Website: Available on your local government's website.

What data is included?

- Economic and housing background
- Focus areas
- Affordable housing units and units expected to be lost
- Activities to be carried out
- Barriers to affordable housing
- Populations of Concern

The Consolidated Plan is a plan/application submitted by all HUD grantees in order to receive federal housing funding. The Consolidated Plan requires a process of public engagement and analysis to assess affordable housing and community development needs and market conditions. A consolidated plan is intended to push local communities to do the data-driven planning this publication recommends and serves as a great first resource for more in-depth analysis.

Pros:

- Extremely thorough
- Has pre-written narrative
- Is already a comprehensive look at your community
- Written by a member of local government staff or a consultant who is an accessible source to discuss the document

Cons:

- Most data are only available at the community-wide level
- Only updated every 5 years, so later in the cycle may be less relevant
- Data is visualized in tables rather than in a more accessible format and may require further processing or visualization for your purposes
- The plans are generally over 200 pages, extremely dense, and difficult to read

Example: Questions answered by the Tallahassee Consolidated Plan

Question: What housing problems do residents experience in our community?

Answer: The Consolidated Plan contains numerous charts, tables, and narrative sessions examining housing problems in your community.

The following narrative from the Tallahassee Consolidated Plan can be used to answer your question and included in any shorter, more accessible report you create with proper citation.

"The most common housing problem residents of Tallahassee face is housing cost burden. Table 7 shows the number of households with severe housing cost burden and housing cost burden far exceeds households living in substandard housing conditions or overcrowded living situations. Approximately 11,578 households experience housing cost burden and 17,994 experience severe cost burden. Renters have a greater rate of cost burden than owners. Table 7 also shows that 599 households live in substandard housing, 1,336 households are overcrowded, and 2,170 households with zero or negative income, that cannot actually have a cost burden, still require housing assistance.

The most common housing problem residents of Tallahassee face is housing cost burden. Table 7 shows that the number of households with severe housing cost burden and housing cost burden far exceed households living in substandard housing conditions or overcrowded living situations. Approximately 11,578 households experience housing cost burden and 17,994 experience severe cost burden. Renters have a greater rate of cost burden than owners. Table 7 also shows that 599 households live in substandard housing, 1,336 households are overcrowded, and 2,170 households with zero or negative income, that cannot actually have a cost burden, still require housing assistance.

Figure 3: Narrative from the Tallahassee Consolidated Plan

Example: Questions answered by the Gainesville Consolidated Plan

Question: How much money does our community receive in HOME dollars?

Answer: The Consolidated Plan and the attached Annual Action Plans contain the most up-to-date information on federal housing amounts. To answer the above question, attach the following chart and include narrative like: "The City of Tallahassee receives \$613,074 annually along with \$5,000 in program income, for \$2,372,296 over the life of the Consolidated Plan."

Table 1: Example of Home Fund Chart from Gainesville Consolidated Plan

Program	Source of Funds	Uses of Funds	Expected Amount Available Year 1				Expected Amount Available Remainder of ConPlan \$	Narrative Description
			Annual Allocation: \$	Program Income: \$	Prior Year Resources: \$	Total: \$		
HOME	public - federal	Acquisition Homebuyer assistance Homeowner rehab Multifamily rental new construction Multifamily rental rehab New construction for ownership TBRA	613,074	5,000	0	618,074	2,472,296	HOME is the largest Federal block grant to State and local governments designed exclusively to create affordable housing for low-income households. HOME funds are awarded annually as formula grants to participating jurisdictions.

Table 50 - Anticipated Resources

The Shimberg Center for Housing Studies Florida Housing Data Clearinghouse

Website: <http://flhousingdata.shimberg.ufl.edu/>

What Data Is Included?

- Housing affordability
 - Affordable and Available Housing
 - Housing Cost Burden
- Assisted Housing Inventory
- Comprehensive Plan Data
- Condos & Manufactured Housing
- Income & Rent Limits
- Data from the Home Mortgage Disclosure Act
- Parcels & Sales for All Counties in Florida
- Population & Household Projections
- Information on Special Needs Populations
- Maps & Visualizations
- COVID-19: Workforce & Housing Indicators
 - Recent Evictions
- Resilience and Energy Assessment of Communities and Housing (REACH) (Tampa Bay Area)
- Disaster Response from OPEN FEMA

The Shimberg Center for Housing Studies is a research center at the University of Florida established by the Florida Legislature. They provide data and applied research through the Florida Housing Data Clearinghouse, perhaps the most comprehensive source of affordable housing data in Florida. The Clearinghouse allows you to choose a community and select between dozens of different data sources provided in prebuilt tables. These tables can be copied directly into Word or Excel or downloaded from the Clearinghouse as an Excel document.

Pros:

- Extremely comprehensive look at affordable housing
- Updated regularly
- Provides information from a variety of sources
- Supported by the excellent staff at the Shimberg Center

Cons:

- Somewhat more difficult to navigate than some of the other listed sources
- Requires you to write your own narrative
- Data mostly available at the community-wide level rather than the neighborhood level

Examples: Questions Answered by The Shimberg Center for Housing Studies

Question: What income levels are most in need of housing in our area?

Answer: The Shimberg Center's Data Clearinghouse contains a huge amount of information on housing need broken out by income level. The following chart looking at the surplus/deficit of affordable and available units succinctly conveys this need in a way that is easy to understand. Add narrative like: "Housing need in the Northeast Region, which includes Bradford, Columbia, Dixie, Gilchrist, Hamilton, Lafayette, Levy, Madison, Suwannee, Taylor, and Union Counties, is extremely high for people making less than 50% AMI, with a deficit of 5,135 units for those making 40% AMI and 3,559 for those making 50% AMI, but there is a sufficient amount of housing at appropriate costs for those making 80% or 120% AMI."

Surplus/Deficit of Affordable/Available Units by Income, Florida Regions, 2019							
Geography	County	0-30% AMI	0-40% AMI	0-50% AMI	0-60% AMI	0-80% AMI	0-120% AMI
Northeast Nonmetropolitan Area (plus Gilchrist)	Bradford, Columbia, Dixie, Gilchrist, Hamilton, Lafayette, Levy, Madison, Suwannee, Taylor, Union	-4,460	-5,135	-3,559	-2,171	392	1,169
Notes: Affordable units are those for which a household at the given income limit (% AMI) would pay no more than 30% of income for gross rent. Affordable/available units are affordable at that income level and either vacant or occupied by a household below the income threshold. Surplus/Deficit of Affordable/Available Units for a particular income range shows the number of affordable/available units in the range minus the number of renter households in that range. For additional explanation of affordable/available methods, see the 2019 Rental Market Study , pp. 32-48							
Sources: Shimberg Center for Housing Studies analysis of 2019 American Community Survey PUMS							

Table 2: Surplus/Deficit of Affordable Rental Units in the Northeast Non-Metropolitan Area from the Shimberg Center

Examples: Questions Answered by The Shimberg Center for Housing Studies

Question: How many foreclosures did we see in 2020, and how did that compare to previous years?

Answer: The Shimberg Center Data Clearinghouse contains data on recent evictions for every county in the state, though some areas are missing 2019 data. Add narrative like: "Foreclosures fell dramatically in 2020 due to the foreclosure moratorium and forbearance."

Geography	Characteristics	Total
Florida	2019 Foreclosure filings	37126
Florida	2020 Foreclosure filings	13090
Sources: Florida Court Clerks & Comptrollers; county Clerk of the Court offices; U.S. Census Bureau, 2019 and 2015-2019 American Community		

Table 3: Foreclosures by year, 2019 and 2020 Shimberg Example

The Florida Home Matters Report

Website: <https://flhousing.org/home-matters-for-florida-report/>

What data is included?

- Count and type of homelessness
- Cost burden by tenure
- Affordable and available rental
- Households below income needed to afford median rent
- Affordable units end of affordability period next ten years
- Many others

The Florida Home Matters report is a yearly report on all the latest housing data for the State of Florida produced by the Florida Housing Coalition. In 2022, the Coalition plans to update the Home Matters Format to include a much larger variety of county- and MSA-level interactive charts and tables.

Pros:

- Data is previsualized alongside prewritten narrative
- Collects data from a huge variety of sources
- The Florida Housing Coalition staff can be contacted to assist with data needs
- Starting in 2020, interactive county level dashboards

Cons:

- Not as comprehensive as other sources
- In years prior to 2022, data is only at the statewide level or for a limited number of MSAs

Examples: Questions answered by the Home Matters Report

Question: How can I find information on how much jobs pay in our community and how that compares with wages in the community?

Answer: In 2022, the Florida Housing Coalition is creating a new interactive Home Matters Report web-based platform that aims to better provide local level data to communities across Florida. Dashboards like the one below provide local-level data by allowing you to click on your community in order to see how it compares to the state or nearby areas. Add narrative like: "In 2022, None of the top 10 occupations in the State of Florida could afford the mortgage on a median home, while in our community _____."

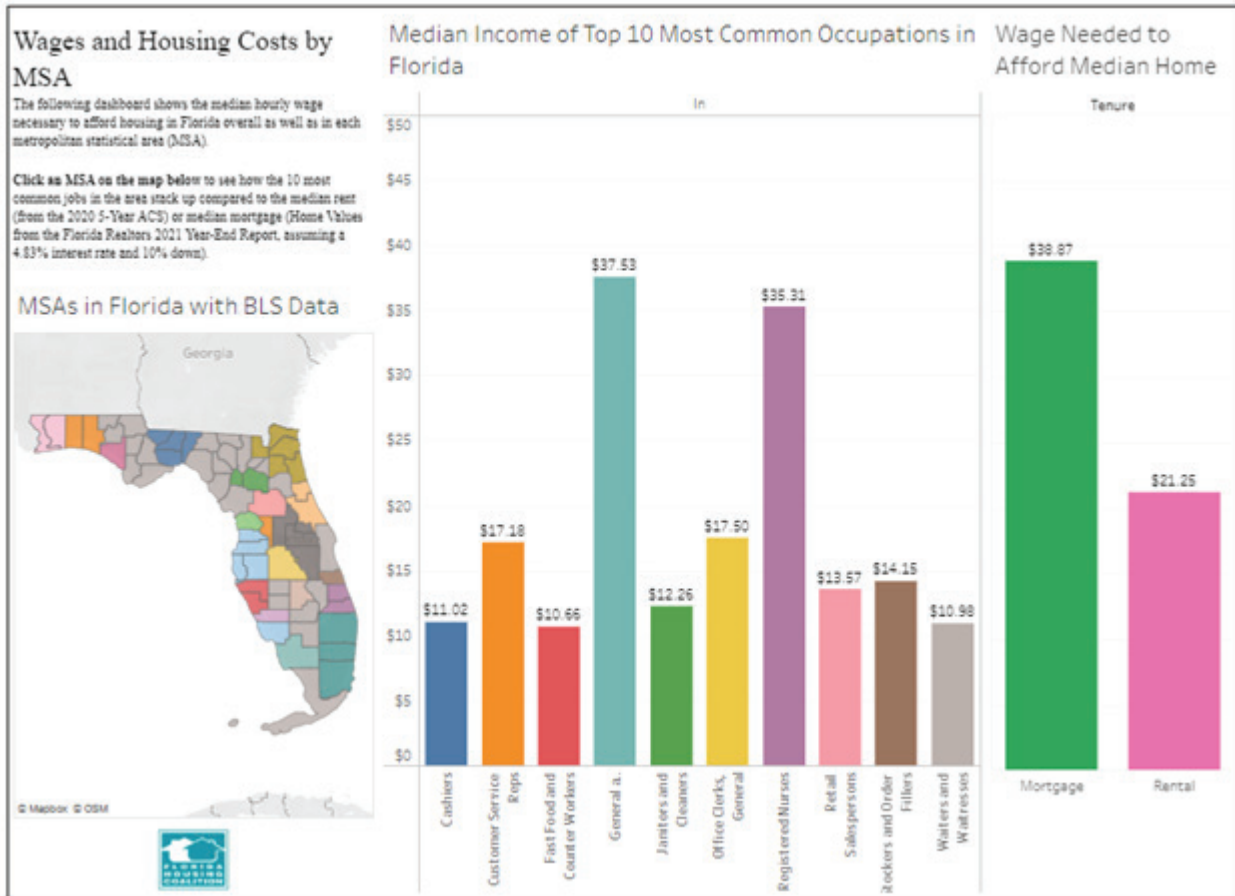


Figure 4: Still from Wages and Housing Cost by MSA Dashboard, Example (2022 Home Matters, Subject to Change)

Florida Realtor Data

Website: <https://www.floridarealtors.org/newsroom/market-data>

What data is included?

Complete overview of recent sales at the state level

Less in depth look at MSA level markets

The Florida Realtors release monthly, quarterly, and yearly data on sales at the statewide and MSA level. This data is the most comprehensive look at how many homes are selling at what price point. While far less comprehensive than the other data sources mentioned in this section, a quick look at the Florida Realtors data can be key to understanding your local housing market.

Pros:

- Most up to date sales data
- Breaks sales out by single and multifamily sales
- Provides an extremely in depth look at the market in Florida

Cons:

- Only available at the statewide and MSA levels
- Narrowly focused on housing sales

Example: Questions answered by the Florida Realtors MSA data

Question: How much does it cost to purchase the median single-family home in our area, and how has it changed over the last year?

Answer: In the Cape Coral-Fort Myers area in February 2022, home prices rose to \$381,481, a 21.1% increase over the previous February. Meanwhile fewer homes were being sold, suggesting a growing supply crunch that might push home prices even higher.”

State / Metropolitan Statistical Area	February 2022				Year-to-Date			
	Closed Sales	Y/Y % Chg.	Median Sale Price	Y/Y % Chg.	Closed Sales	Y/Y % Chg.	Median Sale Price	Y/Y % Chg.
State of Florida	23,661	-1.2%	\$381,481	21.1%	45,546	0.0%	\$379,500	22.4%
Cape Coral-Fort Myers MSA (Lee County)	1,239	-4.1%	\$410,000	24.3%	2,338	-7.1%	\$412,250	24.9%
Crestview-Fort Walton Beach-Destin MSA (Okaloosa and Walton counties)	730	3.0%	\$400,000	14.3%	1,404	6.8%	\$405,541	14.2%
Deltona-Daytona Beach-Ormond Beach MSA (Flagler and Volusia counties)	1,075	2.9%	\$330,000	24.7%	2,060	3.3%	\$325,000	22.8%
Gainesville MSA (Alachua and Gilchrist counties)	247	18.8%	\$310,000	5.1%	450	7.9%	\$300,000	3.5%
Homosassa Springs MSA (Citrus County)	311	4.0%	\$260,000	33.3%	613	6.2%	\$260,000	33.4%
Jacksonville MSA (Baker, Clay, Duval, Nassau, and St. Johns counties)	1,961	0.3%	\$358,961	26.0%	3,771	0.1%	\$355,000	24.6%

Figure 5: Example of Realtors Data at the MSA Level

Pre-Visualized Data, Such as Zillow’s Typical Home Sales Chart, The Opportunity Atlas, and the H+T Index

Zillow’s Home Value Index: <https://www.zillow.com/home-values/>

The Opportunity Atlas: <https://www.opportunityatlas.org/>

The H+T Index: <https://htaindex.cnt.org/>

What data is included?

Site dependent

There are a variety of data sources generated by government, nonprofits, for profits, and academic organizations that previsualize data relevant to affordable housing. The table in “List of Data Resources” includes a list of data sources that provide pre-visualized data. Becoming familiar with these data sources can allow you to quickly answer questions without the need for more work.

Pros:

- Beautifully pre-visualized data
- In-depth analysis of a specific issue

Cons:

- Most of these sources are focused on a specific issue
- Not as comprehensive

Examples: Questions answered by free, pre-visualized data from third party sources

Question: How have home prices changed over the last few years in our community and how does that compare to other places in the state?

Answer: The Zillow Home Value Index tracks “typical” home prices to chart how home prices are changing. This tool allows you to easily type in other areas in the “compare” box in the bottom right to better understand your community’s prices. Add narrative like: “Home prices have risen dramatically in Fort Lauderdale since the beginning of the COVID-19 Pandemic to a current typical home price of \$460k, higher than other communities across South Florida though mirroring similar trends.”

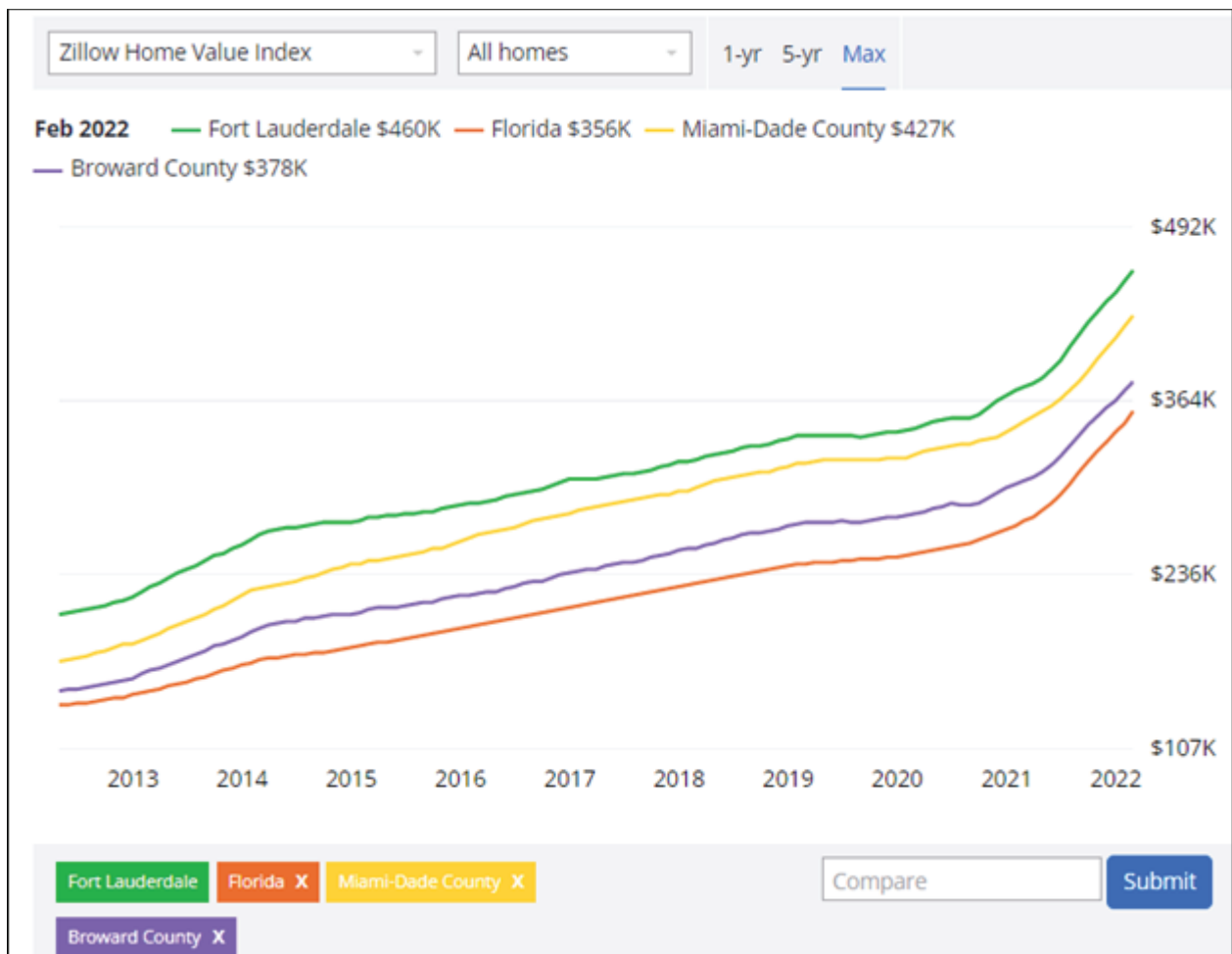


Figure 6: South Florida Home Price Chart from Zillow Home Value Index, Example

Examples: Questions answered by free, pre-visualized data from third party sources

Question: What parts of our community offer the best long-term outcomes for low-income children?

Answer: The Opportunity Atlas, an ingenious data set created by Harvard, Brown, and the US Census, used tax records to track children at age five to adulthood at age 35, to see which areas gave low-income children the best opportunity to enter the middle class. They have now provided this information free on their website with an easy-to-use interactive web application. For this map of Seminole County's Opportunity Index, add narrative like: "Children that were raised in low-income families in the western and south-central parts of the community thrived as adults, while children that grew up in low-income families around Sanford or in the southwest saw far worse outcomes as adults."

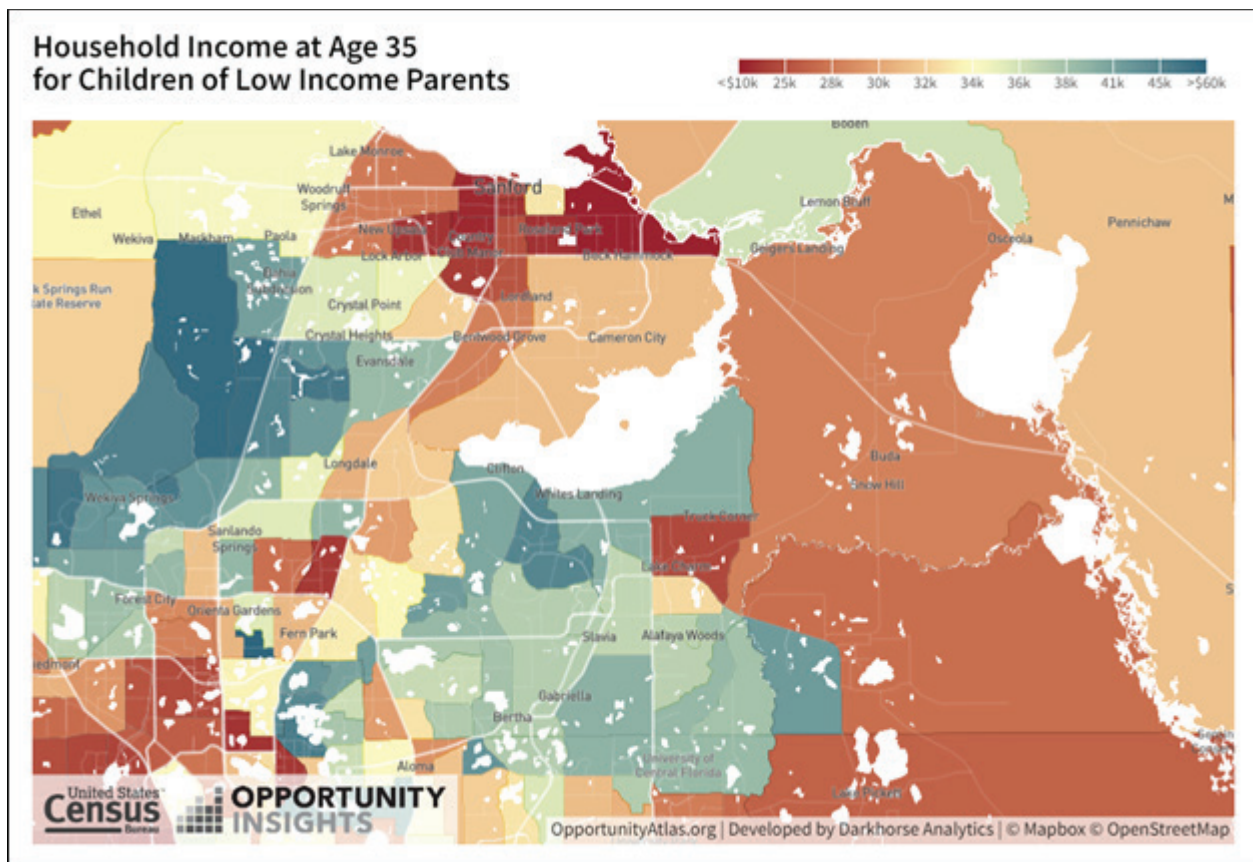


Figure 7: Opportunity Atlas for Seminole County, Example

Examples: Questions answered by free, pre-visualized data from third party sources

Question: What areas in our community have the lowest combined housing and transportation costs?

Answer: In most households, the two largest costs are housing and transportation. Combined, the two should not equal more than 45% of a household's budget, but in many areas housing and transportation can cost much more, often with an inverse relationship (households move to the suburbs to find lower cost housing only to find their transportation costs go up). The Center for Neighborhood Technology (CNT) created their H+T Index to track these costs at the block group, Census tract, or community wide level. Their easy-to-use maps and area profiles can be quickly synthesized in your housing reports. For a map like the one below, add narrative like: "Most of downtown Jacksonville has very low housing and transportation costs, with most of the area below the 45% of median household income that is considered healthy. However, costs rise as you move away from the downtown area."

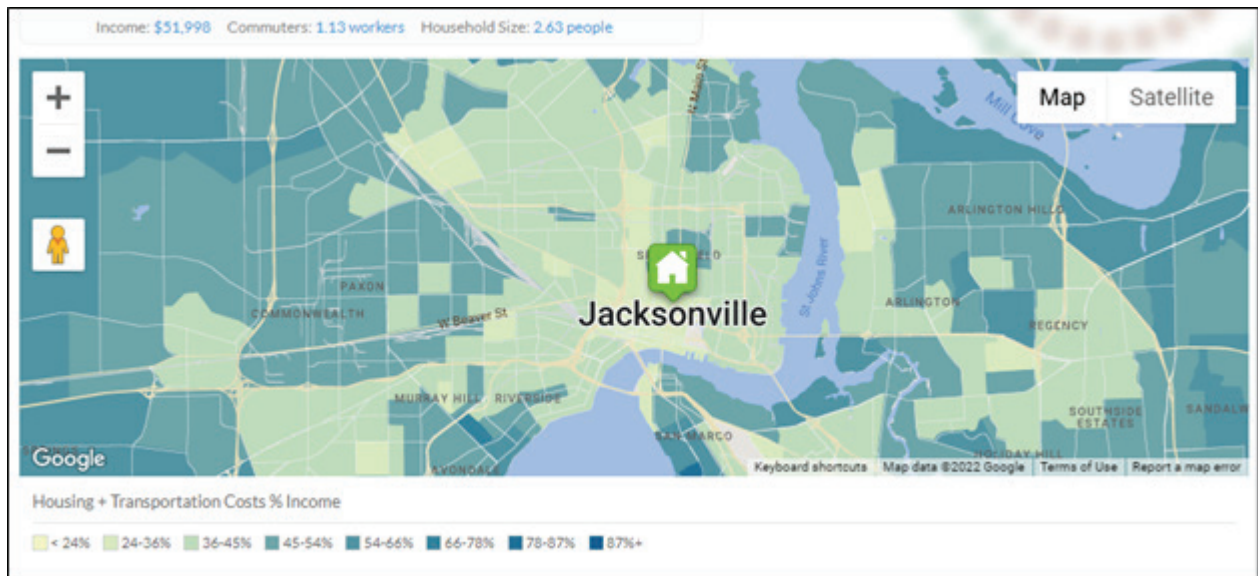


Figure 8: Downtown Jacksonville H+T, Example

More Advanced Analysis

Again, this publication is not providing a complicated look at statistical concepts. The following sections provide some basic ways of analyzing data sets without multivariate analysis or other complex statistical methods. Instead, these methods should help you quickly understand, interpret, and present large data sets to create actionable information.

Descriptive Statistics

Descriptive statistics are statistics that describe a data set. You almost certainly use some of these numbers already, like average, sum, and count. These numbers can be used in more complicated univariate and multivariate analysis, but even without these more advanced statistics, basic descriptive statistics can be extremely helpful for understanding your data set.

Descriptive statistics fall into the following four categories.

Measures of Central Tendency

Median and average are commonly used statistics that help us grasp “the middle” of a set. These numbers can help us to understand what is normal in our community and allow us to compare different values, answering questions like “What is the average wage in our community, and how does that compare to the wage needed to afford an averagely priced home?”

Mean/Average

Mean, also known as average, is perhaps the most commonly used statistic in the world. The mean is calculated using a simple equation with two other descriptive statistics: the sum of the sample divided by the count of that sample. For example, there were 102,783 homes in Marion County that received at least a partial homestead exemption. They were valued at a total of \$17,737,364,528. If we calculate \$17,737,364,528 divided by 102,783, we get the average value for a homesteaded property in Marion County in 2021, \$172,571.

Means are often used to forecast what an “average” or “normal” household can afford, or how much an “average” home costs. However, because homes have a price floor (\$0) but no price ceiling (some homes in your community are almost certainly valued at millions of dollars), the mean tends to show a number far higher than what is really “typical.” To calculate what is typical, median is the preferred method.

Median

The median is another measure of center that often better captures the reality of housing prices compared to the mean/average. The median is the most “in the middle” number. To calculate median, you line up all the numbers you have from smallest to largest and you pick the one in the middle. So, in the example above, there are 102,783 homesteaded properties in Marion County.

If we line them up from smallest to largest and pick the middle value (the 51,392nd row), we see that the community had a median just value for a homesteaded property of \$150,845 in 2021.

Mode

A third, less commonly used measure of central tendency is mode. Mode is the “most common” value in a set. It looks at the most repeated value. That is, if you have this set: 1,1,1, 3,2,5,6, then 1 would be the mode, as it occurs three times while all other numbers occur only once.

The mode for our example data set is not terribly useful; \$5,100 is the most common value in the Marion County just value of homestead properties data set. However, mode can be extremely useful when looking at things like “age first moved to the city” or “most common response on a survey” because it can show you where there is the most agreement or the sharpest trend.

Measures of Dispersion

Measurements of dispersion help us to track the “spread” within our data set. Is the “median” or “average” actually typical, or is there a substantial amount of variation within your records? Understanding this can help you to answer questions like, “Are all housing prices rising or are new homes just being built?” or “Is prosperity spread equally across our community or is there significant variation between neighborhoods?”

Range

The range is the measure of the total spread of values in a set. Put simply, you take the highest value in the set and subtract the lowest value and you have the range. For our data set, the lowest just value for a property receiving homestead in 2021 in Marion County was \$1,827, while the highest was \$7,341,672. That means the whole set has a range of \$7,339,845.

Standard Deviation

Like range, standard deviation measures the amount of dispersion within a set. Usually represented by a lower-case sigma (σ), standard deviation has one of the more complicated formulas for the statistics shown here:

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

Where σ means standard deviation, N is the count, x_i means each value in the population, and μ (μ) is the mean. However, you will never need to calculate this yourself. Instead, use a program like Excel to calculate the standard deviation. The higher the standard deviation is to the mean, the more dispersed your sample is, with a standard deviation near or above the mean generally considered “high.” For Marion County’s just values for homesteaded properties in 2021, the standard deviation is \$136,247, nearly as large as the area mean (\$172,521) suggesting a relatively high amount of variation within the home values of the county.

Measures of Rank

Measures of rank help us to track how something compares to the community as a whole, or how our community compares to other communities. While there are a number of measures of rank, in this section we will only discuss percent rank.

Percent Rank

Percent rank describes where a specific record sits in the population as a whole. You may remember this from standardized testing in high school. A standardized test score of 90% would mean that your score was higher than 90% of all students, while a score of 50% would mean your test scores were higher than 50% of all students. In Marion County in 2021, \$150,845 was ranked in the 50th percentile (the 50th percentile is the same as the median), while \$103,629 was ranked at the 25th percentile and \$208,860 was the 75th percentile. These three numbers (25th, 50th, and 75th percent rank) are also a measure of dispersion, showing the range of the middle 50% of cases in your population.

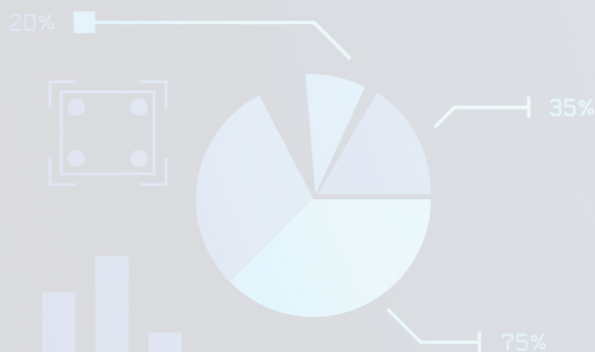
Measures of Frequency

Count

Count is perhaps the simplest descriptive statistic. It refers to the number of records you have in your population. In an Excel spreadsheet, you can determine this by looking at the number of rows and subtracting one for the first line of headings, or by using the =COUNT() function. In our Marion County data, there are 102,783 records or rows. Count can also be used when records have been divided up into categories showing the number of records that fit into each category. Figure 19: Race/Ethnicity in Florida 2020 5-Year ACS, Table DP05, Example 3 shows the count of households by race and ethnicity in Florida.

Percent

Percent frequency is a measure of the share of your population that falls within various groups. Table 6: Race/Ethnicity in Florida 2020 5-Year ACS, Table DP05, Example 2 shows the percentage of households by race and ethnicity in Florida, allowing the viewer to intuitively grasp how common people of each racial/ethnic group are in the state.



Exploratory Data Analysis

Visualization (as laid out in the section below) is not just for presentation. Instead, visualization can be key to helping you understand the information that you have. A good data analyst uses exploratory statistics and visual analysis to check their first assumptions, quickly identify patterns, and find holes or anomalies in their data. Charts such as histograms, bar charts, whisker plots, scatter plots, and other types of analysis can help you better understand the data set you are working with. These visualizations are necessarily rough; the purpose of them is not for presentation to an outside party, but to help you better understand what you are looking at. Before creating formal visualizations, try using Excel, Tableau, Power BI or another software to map, graph, and visualize your data to get a clearer picture of what is actually going on. Do not think of visualization just as an endpoint, but as a tool used to process information in a variety of ways.

Without pre-conceived notions, try taking your data set (like your community's parcel data, Census tract or block group data from the US Census, recent sales data from your property appraiser or Zillow, etc.) and create a few graphs and charts. As you begin exploring, note insights and then create additional graphs and charts to better explore those insights.



Breakout Example

Marion County Property Appraiser Data, Exploratory Charts in Tableau

Link to Charts: https://public.tableau.com/views/MarionExploratoryVisualizationExample/MarionExploratoryVisualizationsExample?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Avg. JV	130,110
Count of JV	270,596
JV	35,207,180,381
Max. JV	89,632,875
Median JV	74,185
Min. JV	0
Percentile (25) of JV	8,000
Percentile (50) of JV	74,185
Percentile (75) of JV	158,505
Std. dev. of JV	585,985
Variance of JV	343,378,618,784

Table 4: Descriptive Statistics for Marion County, Exploratory Statistics

Before jumping into analysis, it can often be helpful to create a series of charts and tables for “exploratory data analysis.” This helps you get a grasp on the large dataset you are looking at before making more complex decisions about how to process and analyze it. These charts are basic and rough: the purpose is not to create professional looking charts and numbers for display but visualizations that help you better understand what is going on. The link above provides an interactive Tableau dashboard with an analysis of just value in Marion County.

First, the dashboard examines some descriptive statistics for just value. There are 270,596 parcels in the county with an average just value of \$130,110 and a median just value of \$74,185. The standard deviation is huge (\$585,985), with a variance of \$343,378,618,784. However, the 25th, 50th, and 75th percentiles are relatively close together, suggesting variation may be a few extreme outliers. Tableau generates

these statistics automatically, or you can use simple excel formulas or the Excel Analysis tab (<https://www.excel-easy.com/examples/descriptive-statistics.html>) to generate these statistics.



Exploring that hunch, we might create a “box and whisker” plot. The box and whisker plot shows the median value, the 25th percentile, and the 75th percentile as a box with values three standard deviations above and below this box as a pair of lines or “whiskers.” Outliers, or values outside of these whiskers, are visualized as dots. Again, box and whisker plots can be automatically generated in Tableau after a dimension and metric are inserted, or can be generated in Excel as a “Static Chart” (<https://support.microsoft.com/en-us/office/create-a-box-and-whisker-chart-62f4219f-db4b-4754-aca8-4743f6190f0d>). Sure enough, the box for parcels is almost not visible, while there are a string of outliers at the high end. Again, this means a few parcels in Marion County are extremely valuable, while most are closer to that \$8,000-\$158,505 range. Again: this chart would not be a good visualization for display. The only purpose of these visualizations is to help you, as the data analyst, to better understand your data before you begin presenting it.



Figure 9: Box and Whisker Plot of Marion County Property Data, Descriptive Statistics Example

Creating a scatter plot with land square feet on one side and just value on the other can help us to see the relationship between parcel size and value (<https://support.microsoft.com/en-us/topic/present-your-data-in-a-scatter-chart-or-a-line-chart-4570a80f-599a-4d6b-a155-104a9018b86e>). There is a correlation, but it is not particularly strong ($R^2: .14$), suggesting that property values may be concentrated on a few sized parcels and may have a geographic component.

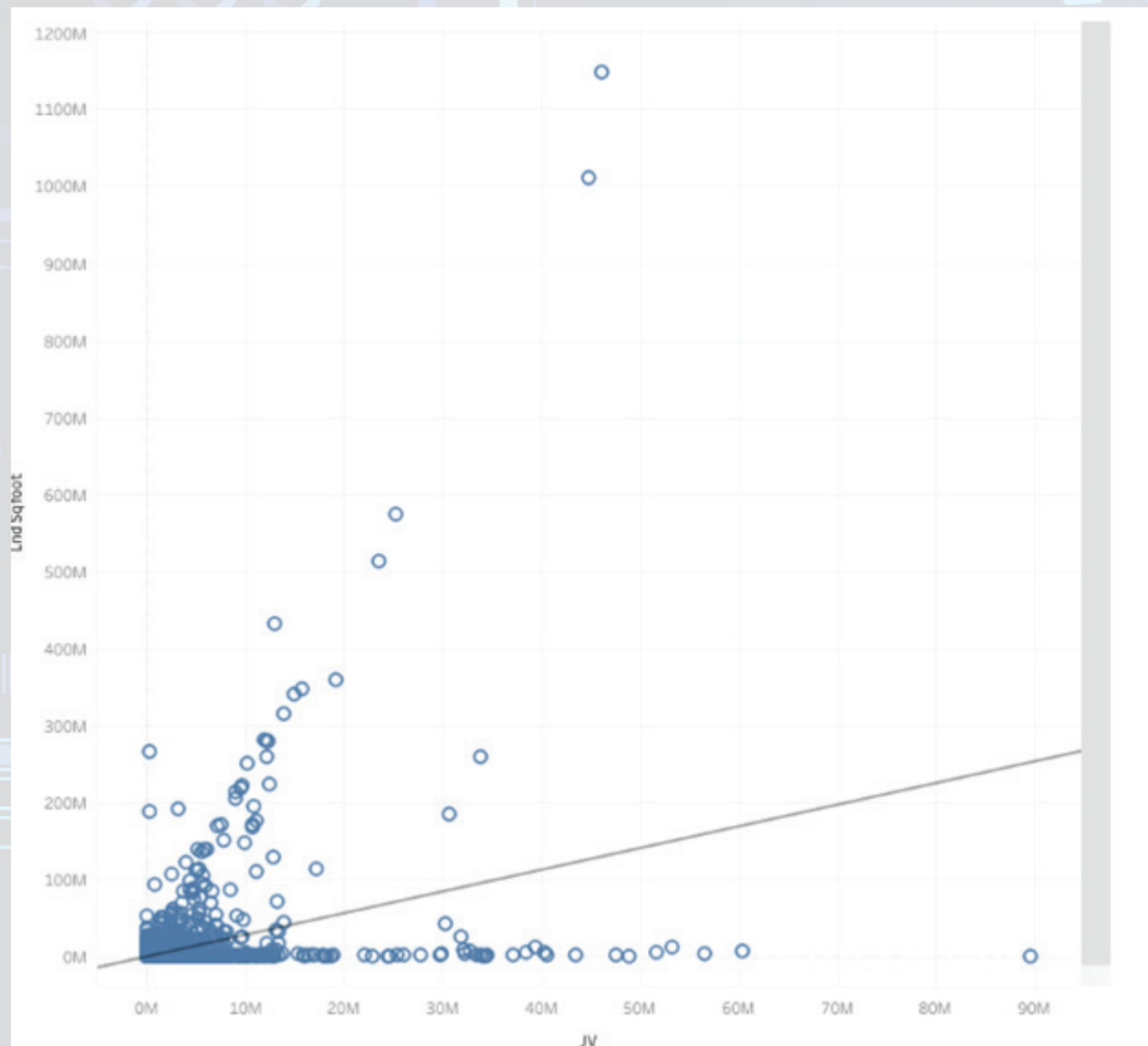


Figure 10: Scatter Plot of Acreage and Value in Marion County, Descriptive Statistics Example

This last insight points us to geography, so the next analysis should include some maps. Maps can be created in Tableau or ArcGIS with a Shapefile or other geographic file. The following maps look at value by acre and total value. They show that while some of the highest just value parcels are just extremely large parcels out in the more rural areas of the county, many are smaller/medium sized parcels in and around downtown.

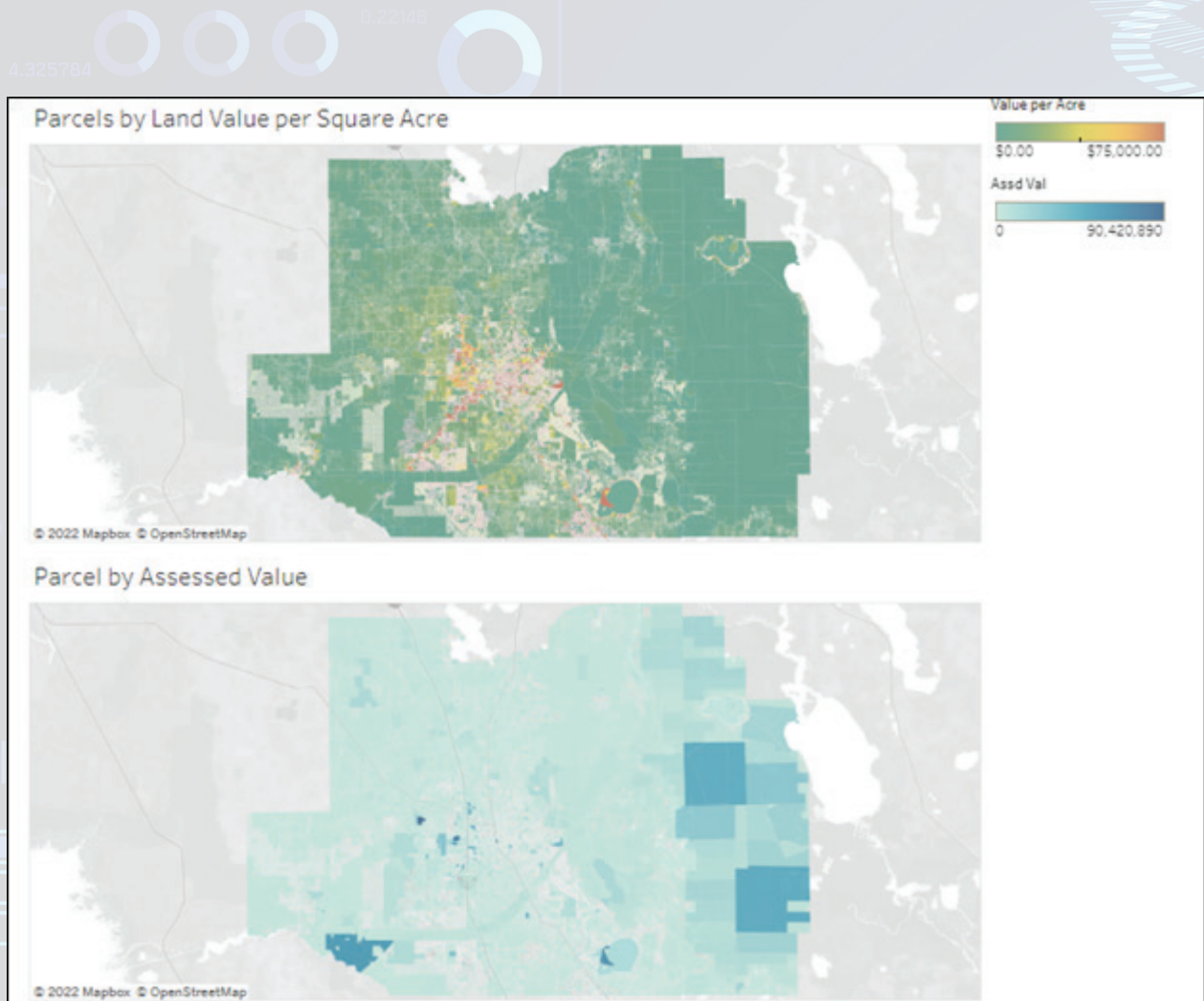


Figure 11: Map of Property Values, Marion County, Descriptive Statistics Example

https://public.tableau.com/views/MarionCountyParcelExample/Dashboard2?:language=en-US&:display_count=n&:origin=viz_share_link

After creating these visualizations, we have a better picture of the property values in Marion County and can create more visually appealing charts and graphics to present our findings to others.

Forecasting

Forecasting is the process of estimating and predicting future trends. A forecast can be extremely helpful for planning, whether you are predicting population growth, housing price change, etc. Of course, no one can see the future. However, with sufficient historical data, many forecasting systems can generate relatively reliable predictions that allow you to plan for your community. A good forecast incorporates data from a variety of sources to better predict which trends are likely to continue. When forecasting, consider using surveys, past trends, and trends in related variables to improve your model (e.g., when studying population growth look at recent jobs growth, when looking at housing price look at new housing construction and vacancy). However, even a simple forecast like the example below can be useful for understanding potential futures.

Example

The statistics behind quantitative forecasting are complicated and worthy of a university level statistics course. Thankfully, programs like Tableau, Excel, and Power BI now provide powerful forecasting tools that allow you to construct your own forecasts without doing the underlying statistics yourself. See examples below of walkthroughs on how to perform a forecast in your preferred program.

Excel: <https://support.microsoft.com/en-us/office/create-a-forecast-in-excel-for-windows-22c500da-6da7-45e5-bfdc-60a7062329fd>

Tableau: https://help.tableau.com/current/pro/desktop/en-us/forecast_how_it_works.htm

Power BI: <https://powerbi.microsoft.com/fr-be/blog/introducing-new-forecasting-capabilities-in-power-view-for-office-365/#:~:text=The%20Power%20BI%20Team&text=Forecasting%20in%20Power%20View%20utilizes,from%20a%20series%20of%20data.>

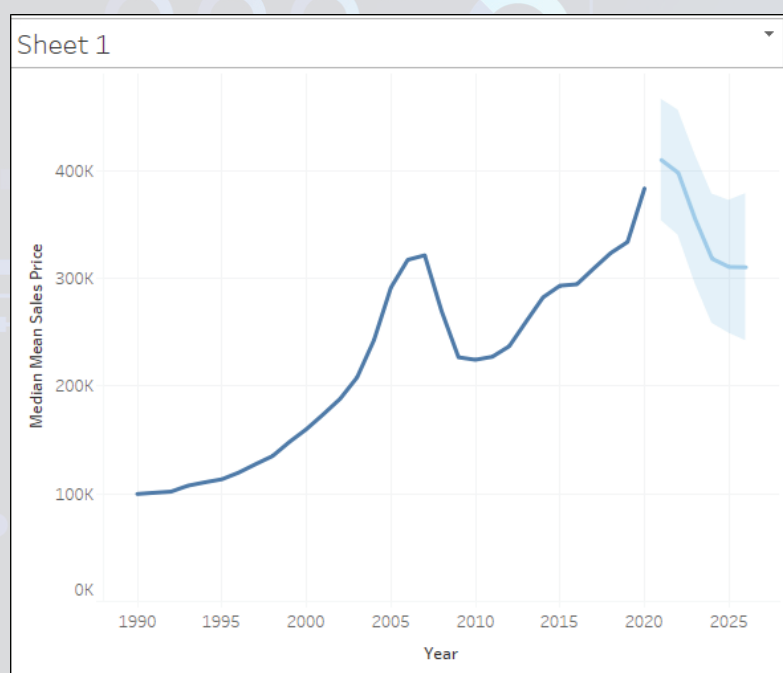


Figure 12: Tableau, Forecasting the Median Home Price in Florida through 2025 from Existing Sales Data

Pivot Table

Pivot tables are a simple but powerful tool for analyzing large data sets in Excel. If you develop no other data analysis skills, pivot tables are a great tool to analyze large data sets quickly and effectively.

Pivot tables require data sets to be structured in a list format with titled column headings. Many data sources, such as American Community Survey (ACS) data accessible at data.census.gov when “zip” is chosen as the download option, automatically include information in this format. The example below, which looks at median and mean sales price for the state, the MSA level, and for all counties along with other sales data, shows how data should be organized. Every column must have a heading for the pivot table to work. Before starting a pivot table, select the topmost leftward cell (A1), which will allow Excel to automatically select the data you want included in the pivot.

MSA Code	County Code	State/MSA/County	Year	Number of Sales	Mean Sales Price	Real Mean Sales Price (2021 \$)
		Florida	2021	144,008	\$459,619	\$459,619
		Florida	2020	386,515	\$383,141	\$401,795
		Florida	2019	364,812	\$333,545	\$354,025
		Florida	2018	348,000	\$323,241	\$349,373
		Florida	2017	326,982	\$308,991	\$342,147
		Florida	2016	310,725	\$294,285	\$332,787
		Florida	2015	277,482	\$292,991	\$335,518
		Florida	2014	241,769	\$282,188	\$323,556
		Florida	2013	233,368	\$259,685	\$302,483
		Florida	2012	194,400	\$236,789	\$279,898
		Florida	2011	166,732	\$227,040	\$273,983
		Florida	2010	164,254	\$224,276	\$279,085
		Florida	2009	158,879	\$226,557	\$286,655
		Florida	2008	171,722	\$269,391	\$339,744
		Florida	2007	202,704	\$321,172	\$420,483
		Florida	2006	341,084	\$317,031	\$426,797
		Florida	2005	450,209	\$291,199	\$404,667
		Florida	2004	409,568	\$242,475	\$348,373
		Florida	2003	359,649	\$207,876	\$306,618
		Florida	2002	319,430	\$188,091	\$283,757
		Florida	2001	206,313	\$173,633	\$266,335

Figure 13: Image of Data Organized for Pivot Table

The pivot table button is included in the far-left box under the Insert tab. Once you have selected cell A1, press the pivot table button and a command box will pop up. The command box should be auto filled with the correct settings. The Table/Range should include all data in your data set and the pivot table should be placed in a new worksheet. Click “OK.”

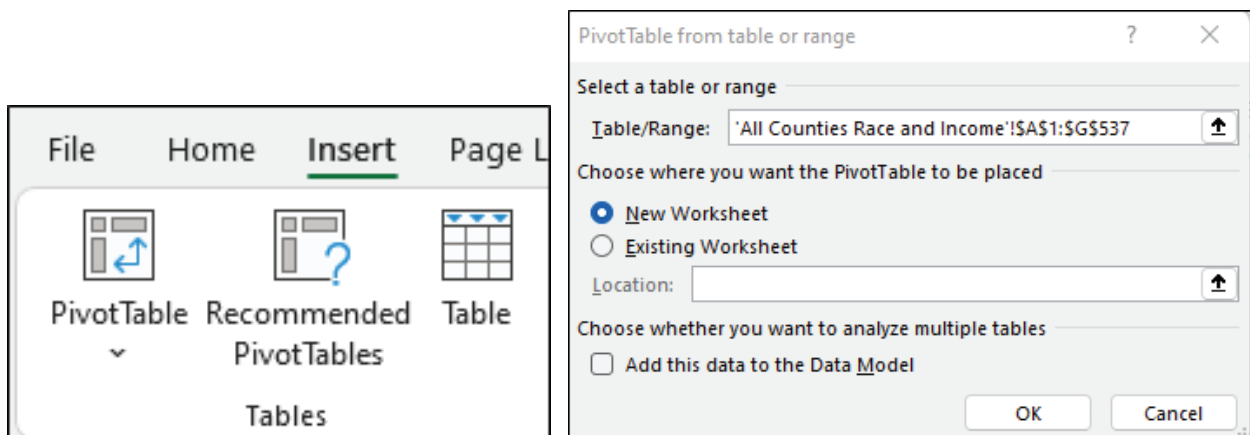


Figure 14: Images of How to Launch a Pivot Table

The pivot table is created on a separate worksheet, but at first the data is empty. On the right of the screen, you will see a box like the one below which includes the titles of the columns from your initial sheet. Drag the data you wish to look at into the “Rows,” “Values,” “Columns,” and “Filters” to create a table of all the information you are looking for.

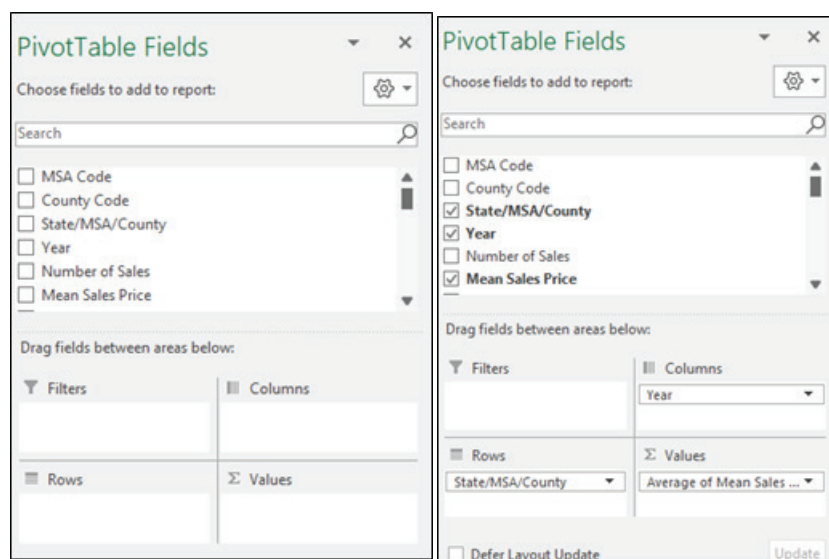


Figure 15: the Pivot Table Fields Box before and after selecting fields for viewing

In the example, this creates a table with average mean sales price per year by geography. It is ideally organized for the creation of a chart or table showing home prices over time.

Average of Mean Sales Price	Column Labels			
Row Labels	1990	1991	1992	1993
Alachua County	80003.99832	82479.6549	85786.94004	88082.5533
Baker County	0		66458.30556	66907.7888
Bay County	70408.58396	71026.0274	77162.26644	80997.3366
Bradford County	53739.68254	56050	54458.92857	62219.4214
Brevard County	87515.67489	87563.31452	87747.60873	90456.6500
Broward County				
Calhoun County	44630.74074	45935.27586	43785.75758	47811.8478
Cape Coral-Fort Myers, FL MSA	117346.7335	112574.5382	117251.6085	123154.15
Central Nonmetropolitan Area	58434.59468	57303.51837	59022.18836	59980.4597
Charlotte County	86697.83188	89468.89097	90873.63608	89198.3773
Citrus County	60876.37512	60769.02468	59995.75325	66659.3076
Clay County	87895.51053	88377.0883	91041.57595	92948.6555
Collier County	139084.7799	138600.9346	154022.4063	172206.064
Columbia County	57825.91235	61494.96	64190.33333	61972.6587
Crestview-Fort Walton Beach-Destin, FL MSA	77299.5183	81953.32352	85419.79211	90865.5052
Deltona-Daytona Beach-Ormond Beach, FL MSA	81291.75603	83168.83446	83119.04076	84283.7923
DeSoto County	60593.42105	62451.69492	62608.13008	61567.1755

Figure 16: Raw Pivot Table output

Filtering the data for the jurisdictions of interest, we can see Alachua and Leon Counties below. From here, it is easy to make a chart showing the way home prices have changed over time, as in the chart below.

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Alachua C	\$80,004.00	\$82,479.65	\$85,786.94	\$88,082.55	\$93,436.45	\$92,314.90	\$99,172.63	\$104,690.12	\$111,073.99	\$115,788.14	\$123,601.3
Leon Cour	\$83,087.42	\$85,279.89	\$90,613.68	\$93,891.20	\$101,945.42	\$114,262.24	\$120,444.67	\$118,324.39	\$128,893.95	\$128,440.12	\$127,390.0

Figure 17: Pivot Table filtered for the jurisdictions we are interested in

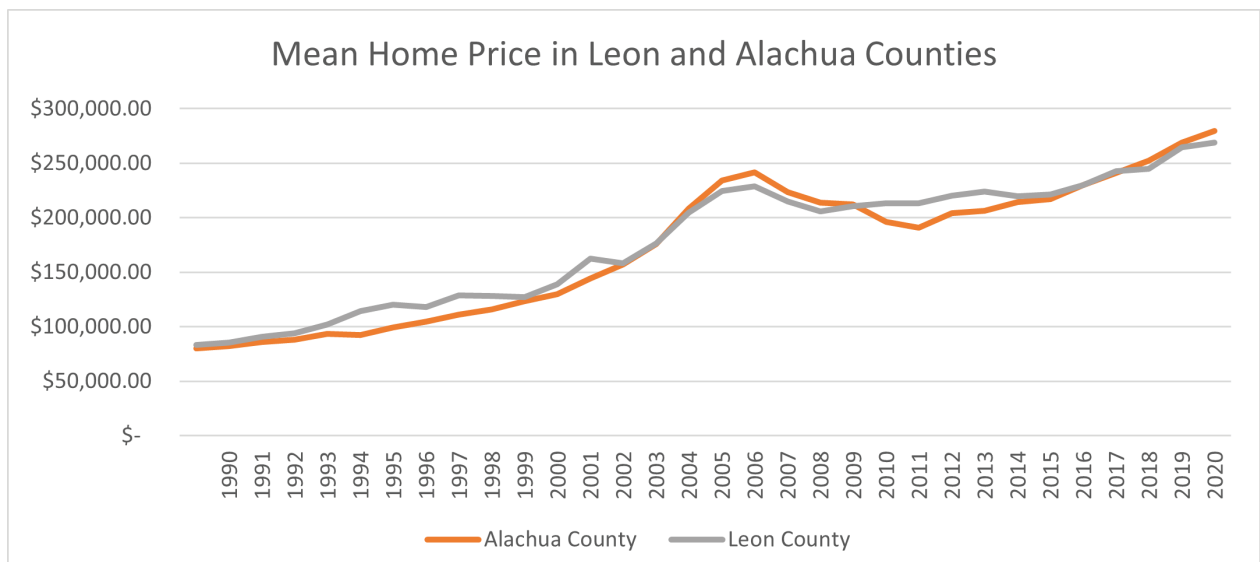


Figure 18: Mean Home Price in Leon and Alachua, Example of Pivot Table output used to create chart

Visualization and Presentation

Many reports are full of great, informative tables, yet they sit on a shelf somewhere gathering dust. It is very difficult for human beings to intuitively grasp large sets of numbers. Numbers are abstractions, and to convert them into something understandable takes mental effort. Instead, data should be presented visually whenever possible, at the very least with table headings and percentages, but generally with easily understandable charts and maps.

There are many different types of charts, graphs, and tables, but it is important to remember simplicity and focus. Fundamentally, data should be presented in the least cultured way possible, with extra information stripped out or included in separate charts. Good data visualizations also use various techniques, such as bolded lines, color changes, and size changes to draw the viewers' attention to the most important part of the table or graph.

Breakout, Example

The following example is intended to show how basic visualization techniques can turn unreadable numbers into something understandable. Here is a table directly copied from the American Community Survey (2020, a portion of Table DP05) without any edits or changes:

Hispanic or Latino (of any race) 5468826
Not Hispanic or Latino 15748098
White alone 11331222
Black or African American alone 3231108
American Indian and Alaska Native alone 39070
Asian alone 579476
Native Hawaiian and Other Pacific Islander alone 10889
Some other race alone 90892

Table 5: Race/Ethnicity in Florida 2020 5-Year ACS, Table DP05, Example 1

It is difficult to draw conclusions from this table, and to do so you must pause and really understand the information, making comparisons and working to understand what each column means. At the very least, such a table can be copied into Excel, cleaned to remove unnecessary rows, and a percent of total column can be added. Here is a slightly more visualized table of the same data from above:

Race/Ethnicity	Population	Percent
Hispanic or Latino (of any race)	5,468,826	25.8%
White alone	11,331,222	53.4%
Black or African American alone	3,231,108	15.2%
American Indian and Alaska Native alone	39,070	0.2%
Asian alone	579,476	2.7%
Native Hawaiian and Other Pacific Islander alone	10,889	0.1%
Some other race alone	90,892	0.4%
Two or more races	465,441	2.2%

Table 6: Race/Ethnicity in Florida 2020 5-Year ACS, Table DP05, Example 2

Here, after being copied into an Excel document, the data is better formatted. “Not Hispanic or Latino” has been removed since it encompasses seven different racial groups, and a percent column has been added (by typing `=B2/sum(B:B)` into cell C2 and then selecting the small box that appears in the bottom right corner of the cell and dragging down so that there is a percentage for each race/ethnic group). It is now much easier to see the relative population sizes of racial/ethnic groups in Florida. Still, this table is not as accessible as it could be.

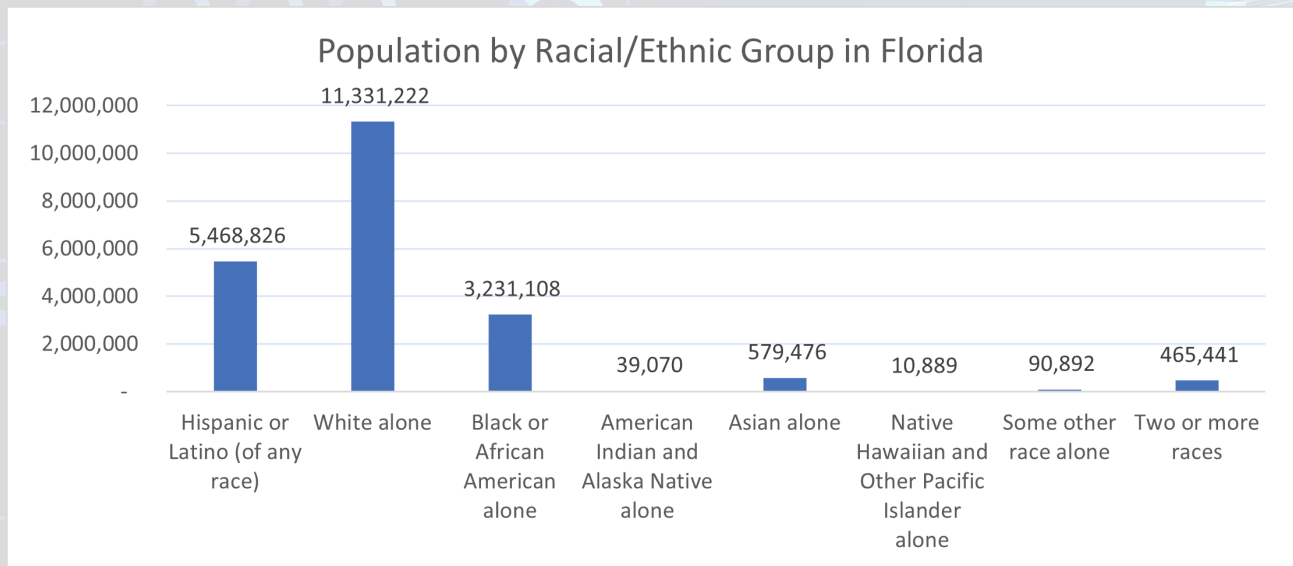


Figure 19: Race/Ethnicity in Florida 2020 5-Year ACS, Table DP05, Example 3

Instead, a simple bar chart allows the viewer to quickly understand the relative sizes of racial/ethnic groups in Florida. By selecting the information in Excel, then selecting Insert>Bar Chart in the ribbon at the top of Excel, we can quickly and easily create a visualization that makes the information dramatically more accessible.

Already Visualized Data

Definition: Information that someone else has formatted and presented in an easy-to-understand way.

Programs Needed: An internet browser

Discussion:

We have already discussed pre-visualized data in the Section “Quick and Easy Data Points and Analysis.” To reiterate, there is no reason to duplicate efforts. Many nonprofits, academic institutions, and other organizations provide data in previsualized formats. Many of the data sources listed in the “List of Data Resources” section at the end of this document have pre-visualized maps, dashboards, and tables. Make sure to cite the original source, but otherwise looking at and even presenting pre-visualized data is a great first step when creating a data profile.

Big Numbers

Definition: Large, attention-grabbing numbers to quickly highlight what you think is most important.

Programs Needed: Excel or another data analysis software can be useful to find your big numbers, but big numbers can be created in whichever program (PowerPoint, Word, etc.) you are using to display your information.

Discussion:

With visualization, simpler and more accessible is almost always better. Very few things are simpler and more accessible than large, easy-to-read numbers. Consider using 1-3 typographically large numbers at the top of reports or through infographics (like in the example below). The numbers below are all “sums.” Sum is one of the descriptive statistics described in the descriptive statistics section above and often a key number to highlight. Other descriptive statistics also make good “Big Numbers,” such as averages, medians, count, and rank. These can be clearly displayed at the top of a report or section.

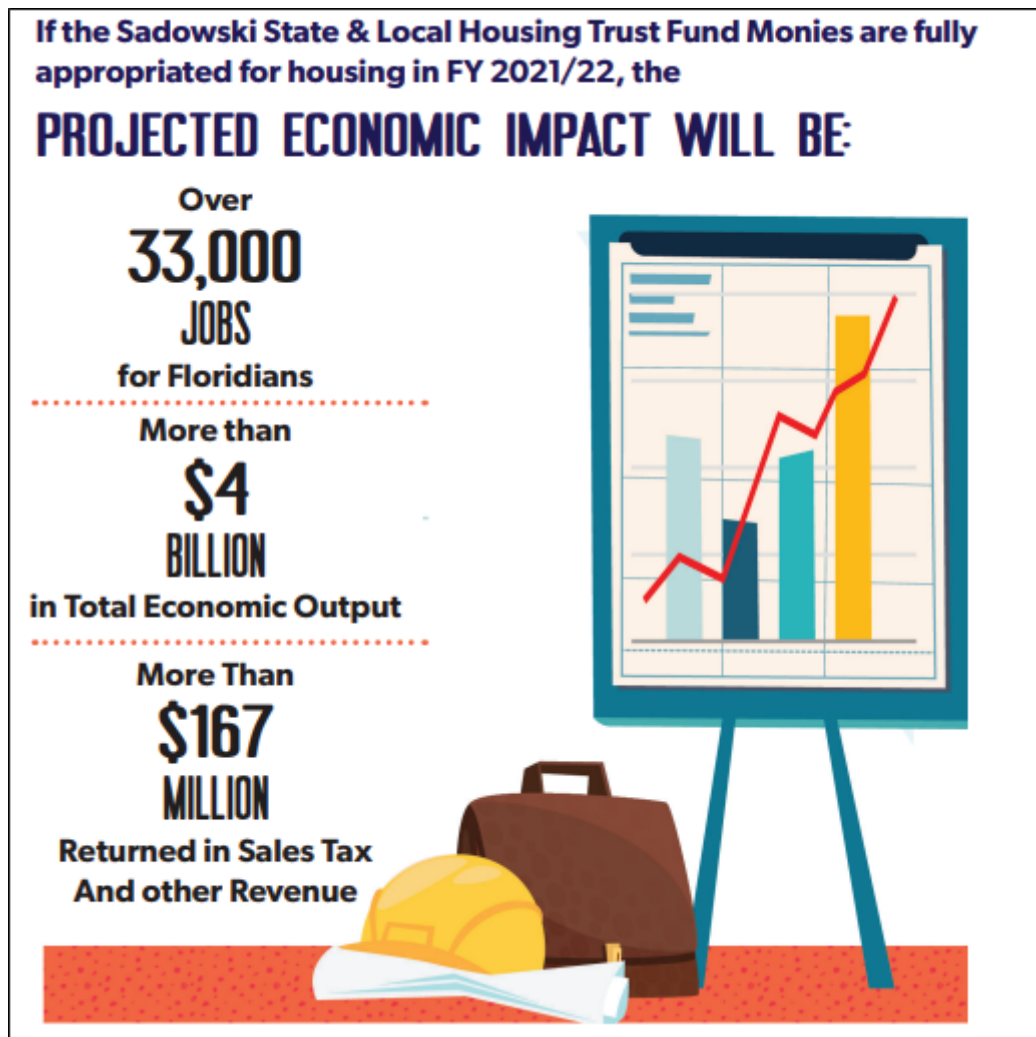


Figure 20: Big Numbers Infographic from 2021 Home Matters, Example

Simple Tables

Definition: A display of information in tabular form.

Programs Needed: Word, PowerPoint, and many websites allow the creation of tables, but tables can be generated with more functionality in Excel and other data analysis programs.

Discussion:

Even people with relatively little data analysis experience are often comfortable creating simple tables in Excel or Microsoft Word. Tables are extremely easy to create (many sources and programs even provide pre-built tables), only requiring you to type into a few Excel cells and then format in Excel, Word, or PowerPoint.

A good table is relatively simple. Tables should have columns and rows clearly labeled and should also include a title or heading, again for clarity. Tables can be an effective way to present information but lose effectiveness as they become more complicated. Tables with a single category are clearest:

Homeownership Rate (%), 1990 through 2015-2019				
	1990	2000	2010	2015-2019 5-Year Estimates
Alachua County	54	55	54	55
Sources: U.S. Census Bureau, 1990 Census, 2000 Census, 2010 Census, and 2015-2019 American Community Survey 5-Year Estimates				

Table 7: Homeownership in Alachua County from the Shimberg Center, Example

As you begin to add categories to the left of the table, it becomes more complicated, though tables with a few rows can still be helpful and accessible. In the chart below, Florida as a whole has been added to the Alachua chart above as a comparison:

Homeownership Rate (%), 1990 through 2015-2019				
	1990	2000	2010	2015-2019 5-Year Estimates
Florida	67	70	67	65
Alachua County	54	55	54	55
Sources: U.S. Census Bureau, 1990 Census, 2000 Census, 2010 Census, and 2015-2019 American Community Survey 5-Year Estimates				

Table 8: Homeownership Rate in Florida and Alachua County from the Shimberg Center, Example

Meanwhile, this chart from the Tallahassee Consolidated Plan contains far too much information to be read easily. It will not fit on a single page and contains two different subgroups (renters and owners). Tables like this are acceptable, if necessary, in an appendix or internal report, but when presenting information to the public, elected officials, or in PowerPoints, tables like this are more likely to cause people's eyes to glaze over than to meaningfully impart information. Consider separating tables like this into multiple charts or using a different visualization tactic (such as a chart, map, or interactive dashboard) to convey this information in a way that is more intuitive.

	Renter					Owner				
	0-30% AMI	>30-50% AMI	>50-80% AMI	>80-100% AMI	Total		>30-50% AMI	>50-80% AMI	>80-100% AMI	Total
NUMBER OF HOUSEHOLDS										
Substandard Housing - Lacking complete plumbing or kitchen facilities	265	200	25	65	555	0	0	40	4	44
Severely Overcrowded - With >1.51 people per room (and complete kitchen and plumbing)	345	150	120	55	670	0	0	0	4	4
Overcrowded - With 1.01-1.5 people per room (and none of the above problems)	135	124	264	30	553	19	0	75	15	109
Housing cost burden greater than 50% of income (and none of the above problems)	10,750	3,950	725	100	15,525	1,085	620	534	230	2,469

Table 9. Example Table from Consolidated Plan

Simple Excel Charts

Definition: Information organized into an accessible format, such as a table, graph, or diagram.

Programs Needed: Excel

Discussion:

Excel provides a variety of tools for visualizing data with minimal transformation. Because of Excel's ubiquity, many programs and data sources come pre-formatted into Excel chart-friendly tables, including ACS and Census data downloaded through data.census.gov using the "Excel" download option.

The image below shows the buttons for the various types of charts available in Excel. This section explores the various types of charts and how they are used.

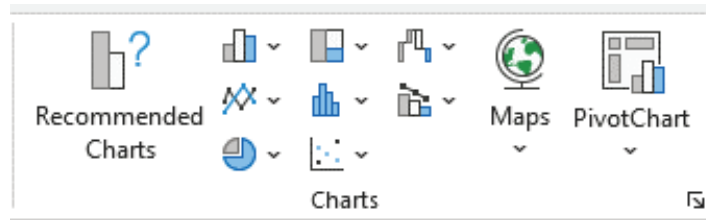


Figure 21. Buttons for Various Types of Charts in Excel

Column or Bar Graphs

Column or bar charts are perhaps the simplest and most effective way to present data. In a column or bar graph, the value of your variables is represented by the height of equally wide columns or bars. It is extremely easy for the viewer to immediately understand the relationship between variables. Column and bar graphs are perhaps the most commonly used chart today.

The previous example (*Table 6: Race/Ethnicity in Florida 2020 5-Year ACS, Table DP05, Example 2*) on page __ provides an overview of the way data should be organized for the creation of bar charts in Excel.

Line and Area Graphs

Line graphs are a type of graph that are generally (though not exclusively) used to show change in information over time. Area graphs are line graphs where the area below the line is shaded. These consist of a series of points that represent a specific date and value, connected by a line to show the trend over time.

Line and Area Graphs can be created by formatting data with horizontal rows that represent the record or area you want to track alongside columns that represent the various dates you are examining. By selecting this data and then clicking “line graph” under Insert>Charts, Excel will automatically generate a line graph like the one below.

Homeownership Rate (%), 1990 through 2015-2019 5-Year Estimates				
	1990	2000	2010	2015-2019 5-Year Estimates
Florida	67	70	67	65
Alachua County	54	55	54	55

Sources: U.S. Census Bureau, 1990 Census, 2000 Census, 2010 Census, and 2015-2019 American Community Survey 5-Year Estimates

Table 10: Homeownership Rate over Time, Example of Line Graph Formatting

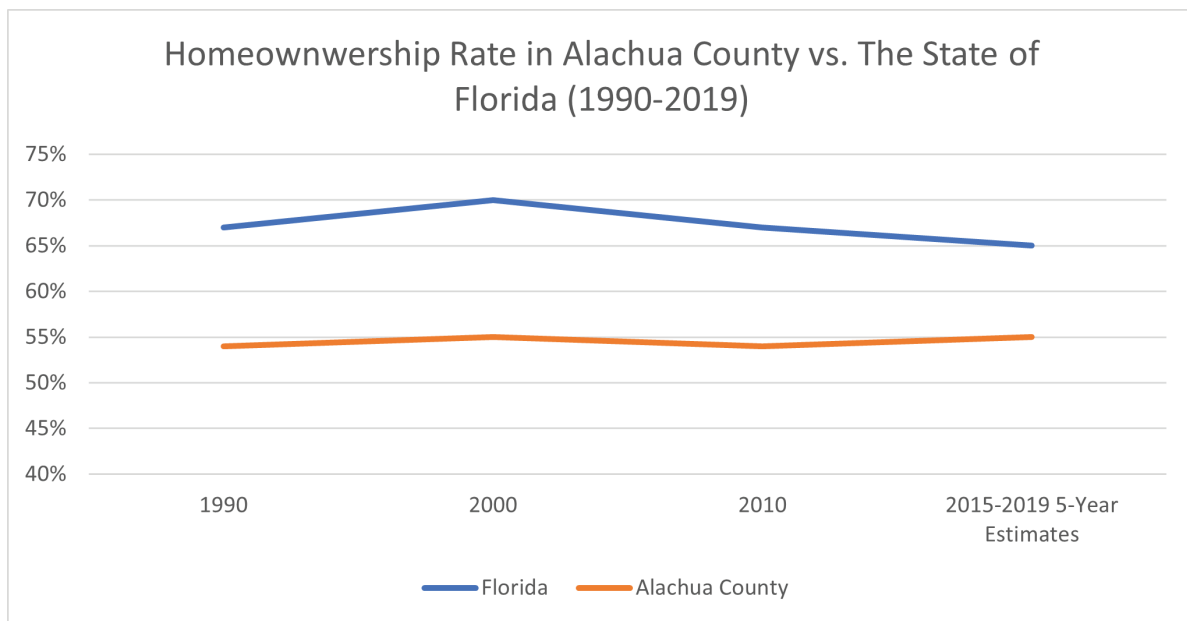


Figure 22: Homeownership Rate Line Chart Example

Pie Charts and Hierarchy Graphs

Pie charts and hierarchy graphs are created with two different buttons in Excel but show data in functionally the same way: a shape is divided into proportional sectors that represent the proportions of the underlying data. Both types of charts are widely used, but the pie chart has been falling out of fashion with data analysts in recent years because it can be more difficult to intuitively understand similar sized sectors compared to hierarchy charts.

Data for pie and hierarchy charts should be organized the same as a bar chart, with categories included as one column followed by the values you wish to visualize. Then, the data is selected, and the pie or hierarchy button is pressed in the Charts box.

<i>Race/Ethnicity</i>	<i>Population</i>	<i>Percent</i>
<i>Hispanic or Latino (of any race)</i>	<i>5,468,826</i>	<i>25.8%</i>
<i>White alone</i>	<i>11,331,222</i>	<i>53.4%</i>
<i>Black or African American alone</i>	<i>3,231,108</i>	<i>15.2%</i>
<i>American Indian and Alaska Native alone</i>	<i>39,070</i>	<i>0.2%</i>
<i>Asian alone</i>	<i>579,476</i>	<i>2.7%</i>
<i>Native Hawaiian and Other Pacific Islander alone</i>	<i>10,889</i>	<i>0.1%</i>
<i>Some other race alone</i>	<i>90,892</i>	<i>0.4%</i>
<i>Two or more races</i>	<i>465,441</i>	<i>2.2%</i>

Table 11: Data formatted for Pie or Hierarchy Chart, 2020 ACS

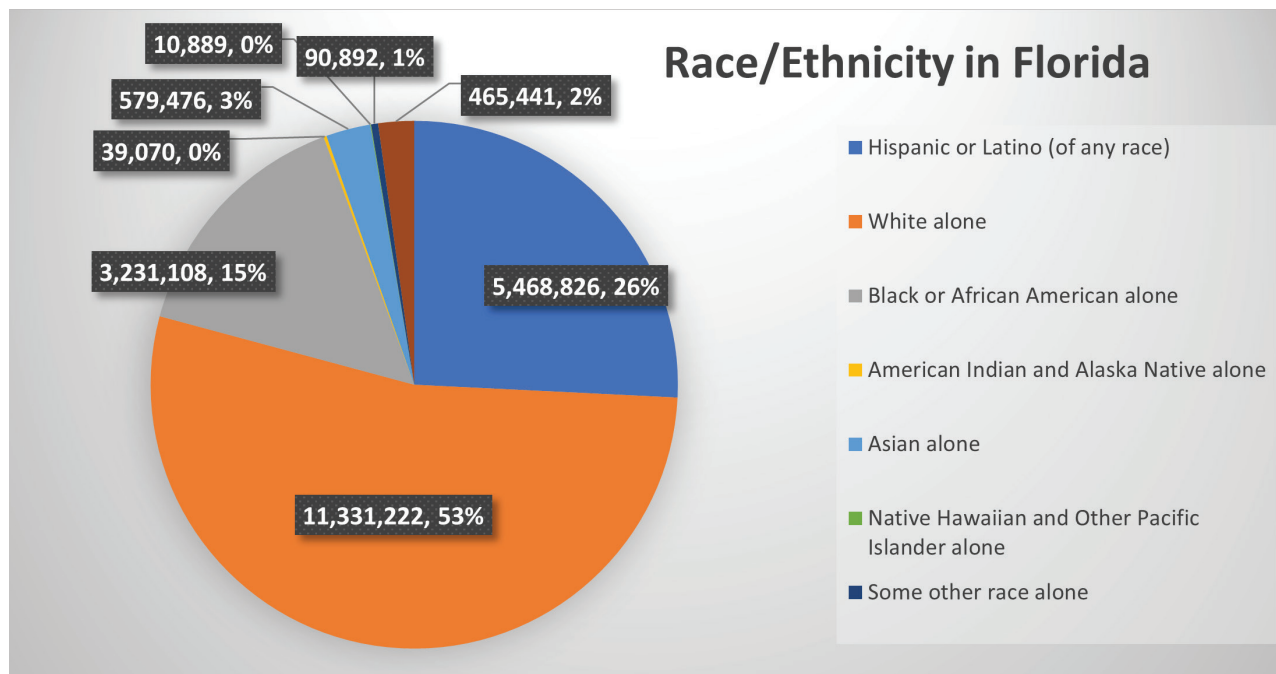


Figure 23: Pie Chart

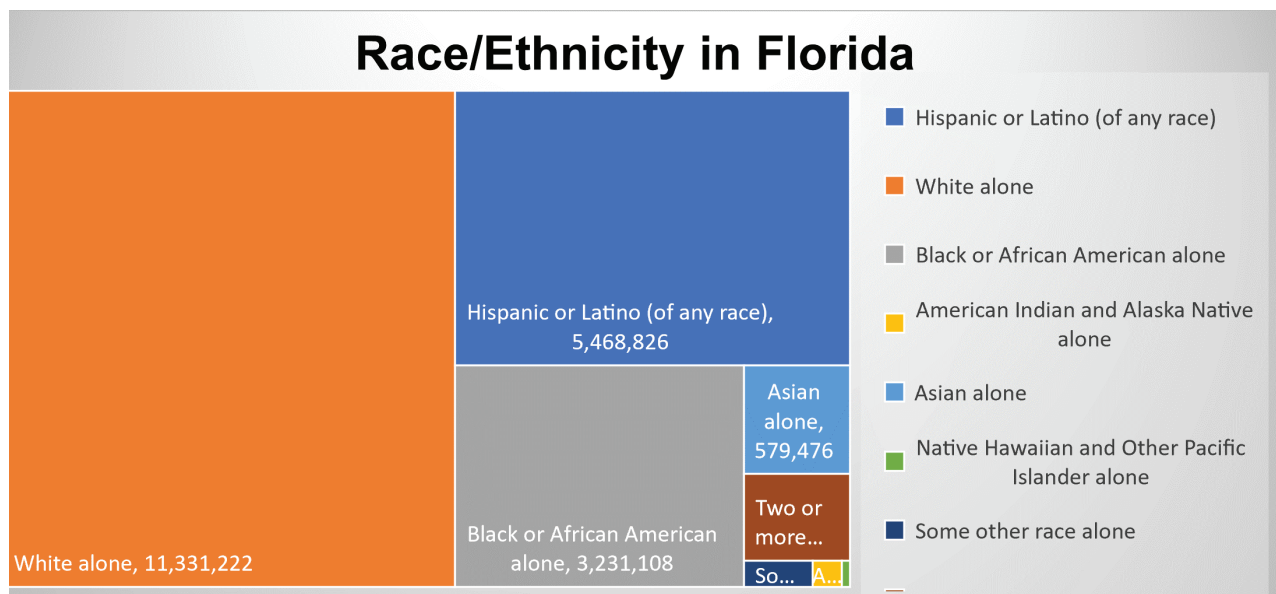


Figure 24: Hierarchy Graph Example

Scatter Charts/Plots

Scatter charts or scatter plots are a way of showing the relationship between two variables, similar to line/area charts except that, while line charts are primarily used to analyze a variable over time, scatter plots are used to detect trends in relationships more generally. Scatter plots have two axes, vertical and horizontal, and individual records are plotted on their value for each at the same time. This way, it is possible to see if the values relate and the slant and direction of their trend.

Scatter plots for Excel should be organized as two columns (vertical) of information that have a relationship with each other. The example below shows the relationship between overall homelessness and sheltered homelessness within continuums of care (CoC). It is clear there is a relationship between the overall homelessness population in a CoC and the number of people living in shelters, which makes sense, but there is significant variation from the overall trend with some areas having a far higher or lower ratio between these two numbers.

Overall Homeless, 2021	Sheltered ES Homeless, 2021
526	292
870	713

Table 12: Example of how data should be organized for a scatter plot, 2021 PIT Count

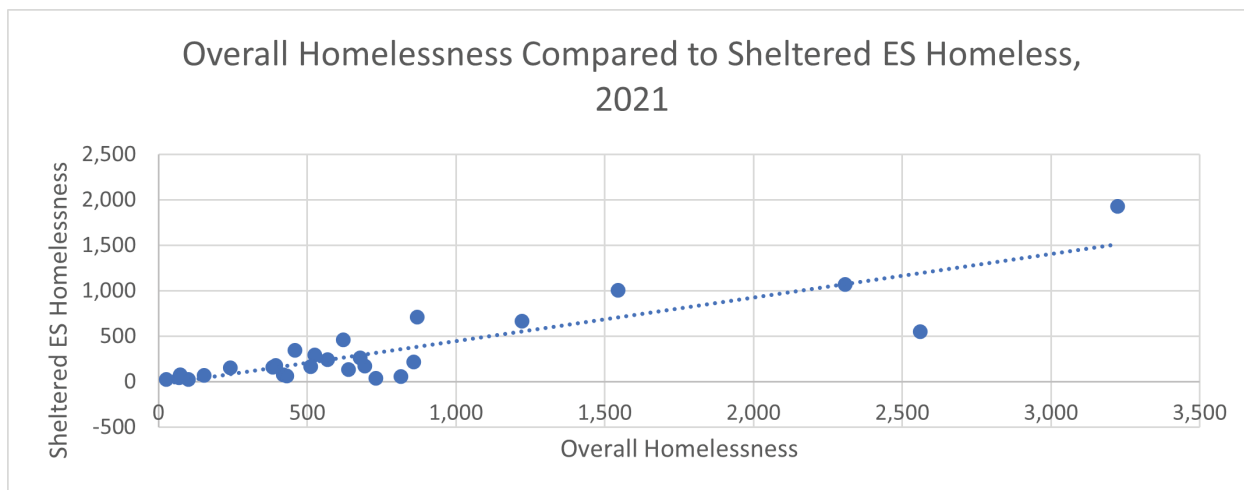


Figure 25: Overall Homelessness Compared to Sheltered Homelessness, Scatterplot Example, 2021 PIT

Mapping

Definition: A two-dimensional representation of a geography, often overlaid with other information.

Programs Needed: ArcGIS, Tableau, Python, and others

Discussion:

Mapping is an extremely powerful form of visualization. A good map can be grasped quickly and intuitively and shows variations across an area. While area-wide sums, averages, and medians can be helpful, they can accidentally hide variations between different areas of your community. The map below shows median age by Census tract in Marion County, and it is immediately obvious to the observer where people are younger and older in the area. This information is in many ways more useful than an area-wide average that would hide very young and very old pockets within the county.

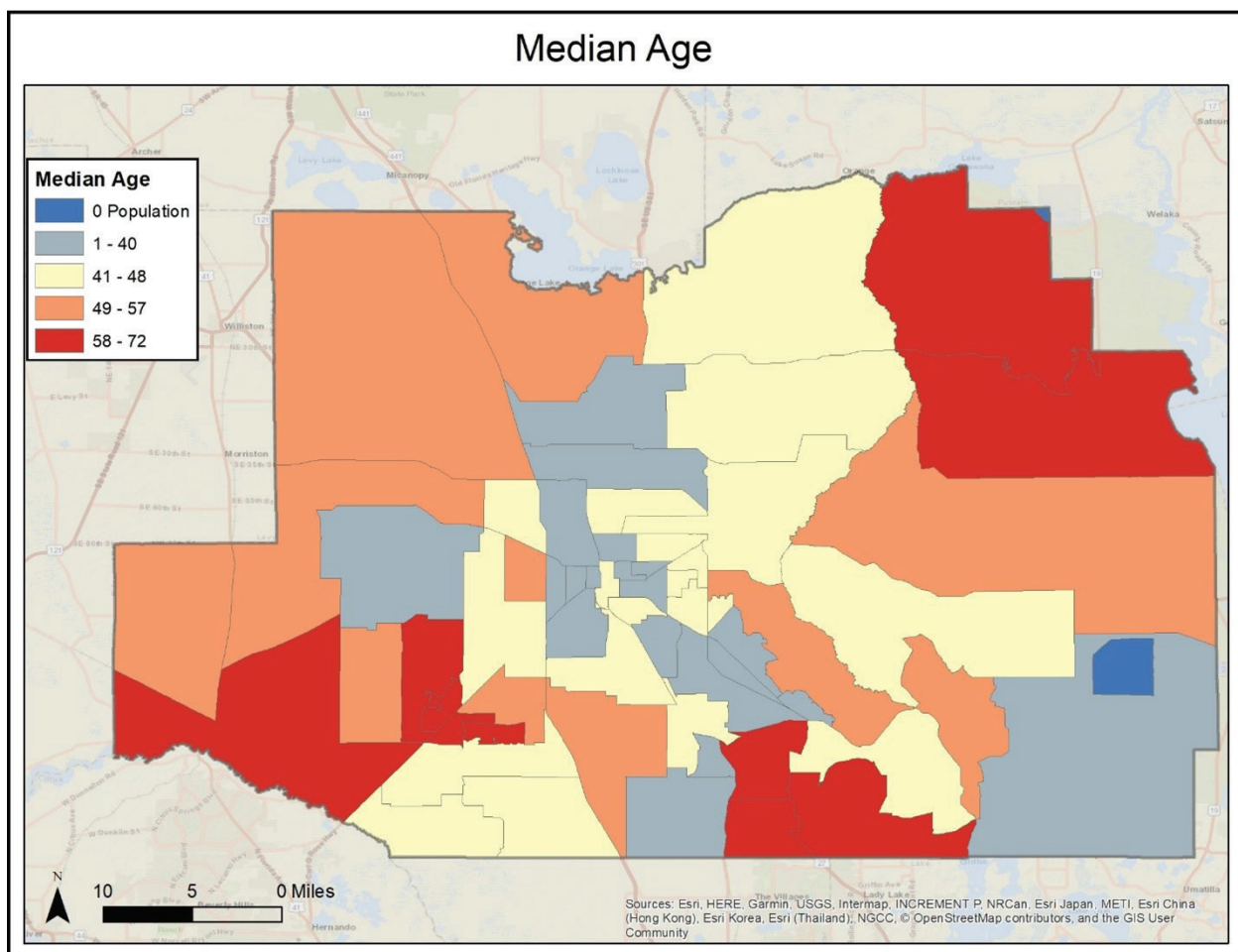


Figure 26: Median Age by Census Tract in Marion County, County from the Marion County Workforce Housing Study. 2019 5-Year ACS, Example

To create ArcGIS maps may require ArcGIS skills, but most administrators can work with their colleagues to produce such maps or find online interactive maps they can grab (with appropriate citation) using the Snipping Tool that comes with most computers. Incorporating maps where possible makes your analysis and presentation more engaging and, with the broad adoption of programs that allow for visualization, is much easier than even five or six years ago. Programs like Tableau and Power BI also include mapping capabilities that are much simpler than ArcGIS and other older and more powerful programs.

This section is not intended to be an introduction to any specific program (though Excel tables were addressed in more detail). However, as a broad rule, it is surprisingly easy to make a bad map, but most of the worst offenses can be easily avoided by keeping it simple. See the map below.

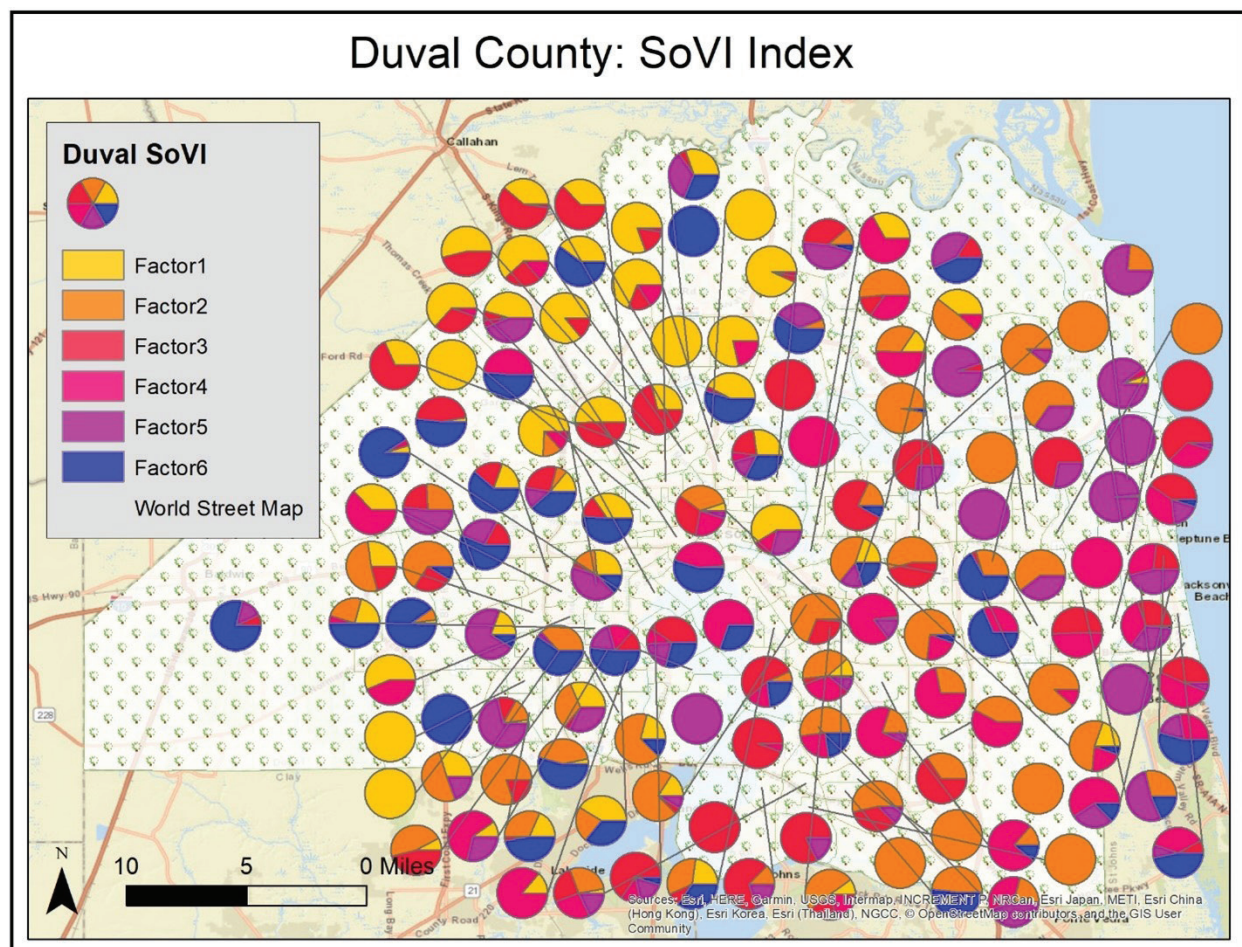


Figure 27: Example of Extremely Ugly Map

This map provides the level of a series of vulnerability factors, but it is completely unreadable. Instead, separate complex maps into multiple, simpler maps, using red-green, red-blue, or other simple gradients, and making sure that only key information is included in the legend. That way, the map becomes much clearer.

Duval County: Social Vulnerability by Census Tract

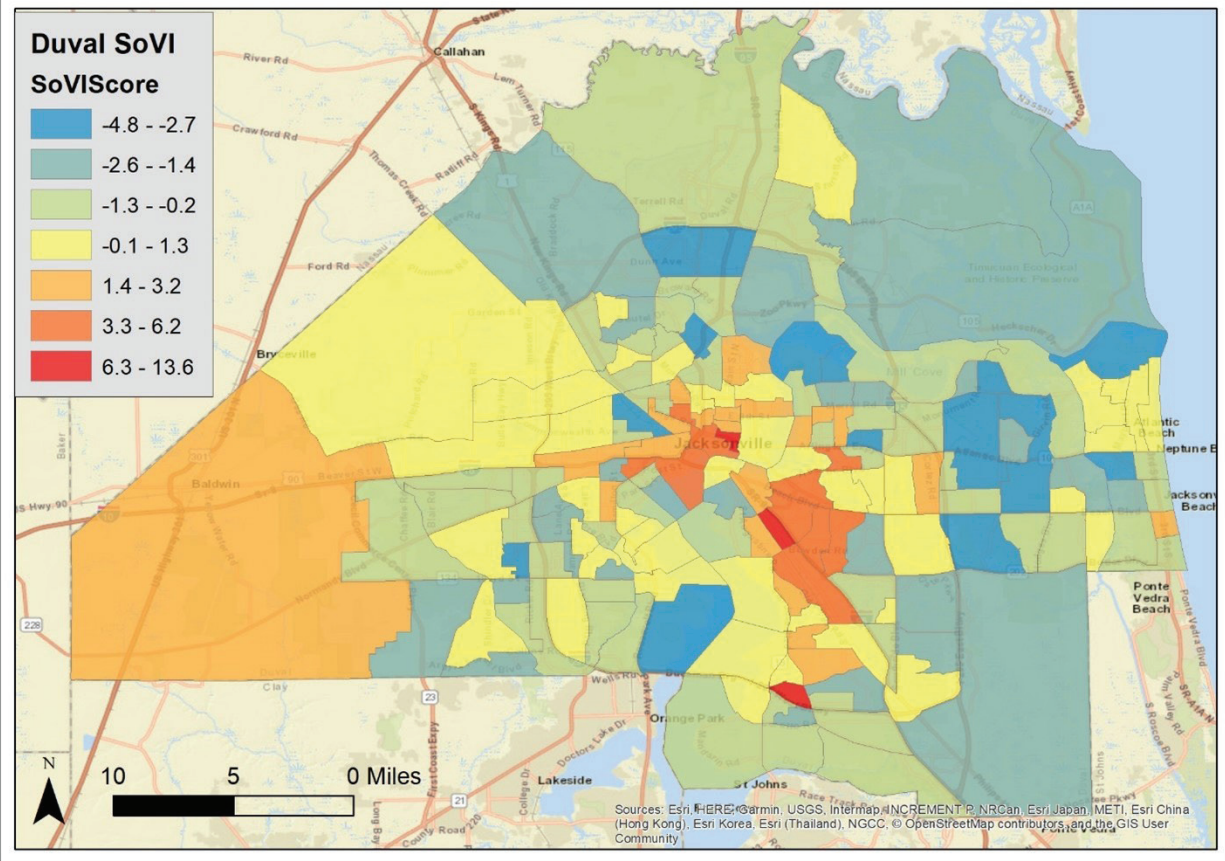


Figure 28: Ugly map fixed

Interactive, Online Visualizations

Definition:

Programs Needed: Tableau, ArcGIS, Power BI, Python, and others.

Discussion:

Over the last few years, programs like ArcGIS online, Tableau, Power BI, and others have made information easier to process and display. These programs also allow anyone to access online, interactive dashboards. Dashboards collect multiple charts, graphs, maps, etc. from multiple underlying data sources and present them together in a single display. Using this software, these dashboards can be automatically updated and/or interactive, allowing people to navigate your information themselves. Rather than just producing static maps and charts, it is now possible to create colorful, online displays that stakeholders can interact with.

Producing useful interactive dashboards requires producing useful charts, tables, maps, and large numbers. A good dashboard is quickly accessible, and simple charts and large numbers can help draw the audiences' eyes to your key points. Good dashboards also utilize the interactive capabilities of these programs to allow the audience to filter data so they can access the information that is most relevant to them. The dashboard below presents two large numbers (percent of people experiencing homelessness and total change in the homeless population) for the state of Florida as well as a bar chart of the number of individuals that fall within specific groups/types of homelessness. The map provides a look at where homelessness is concentrated and, by clicking on a specific CoC, allows the viewer to filter both the big numbers and chart to better understand a local area.

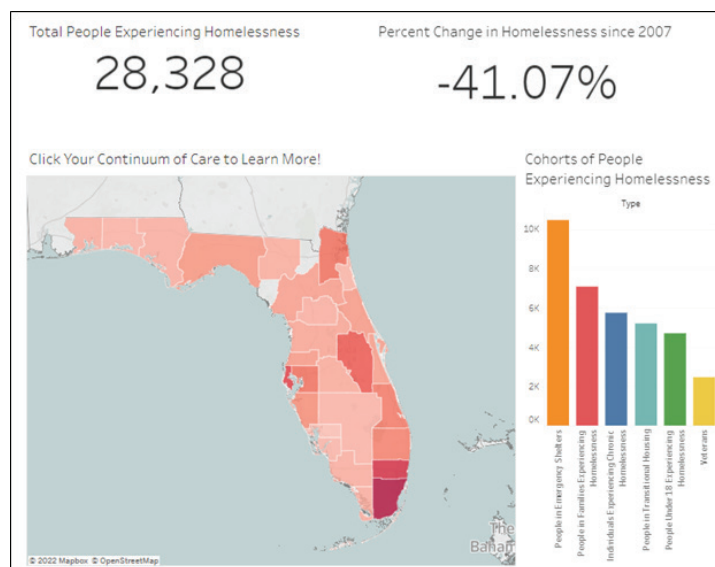


Figure 29: Interactive Dashboard of Continuum of Care Data, 2019

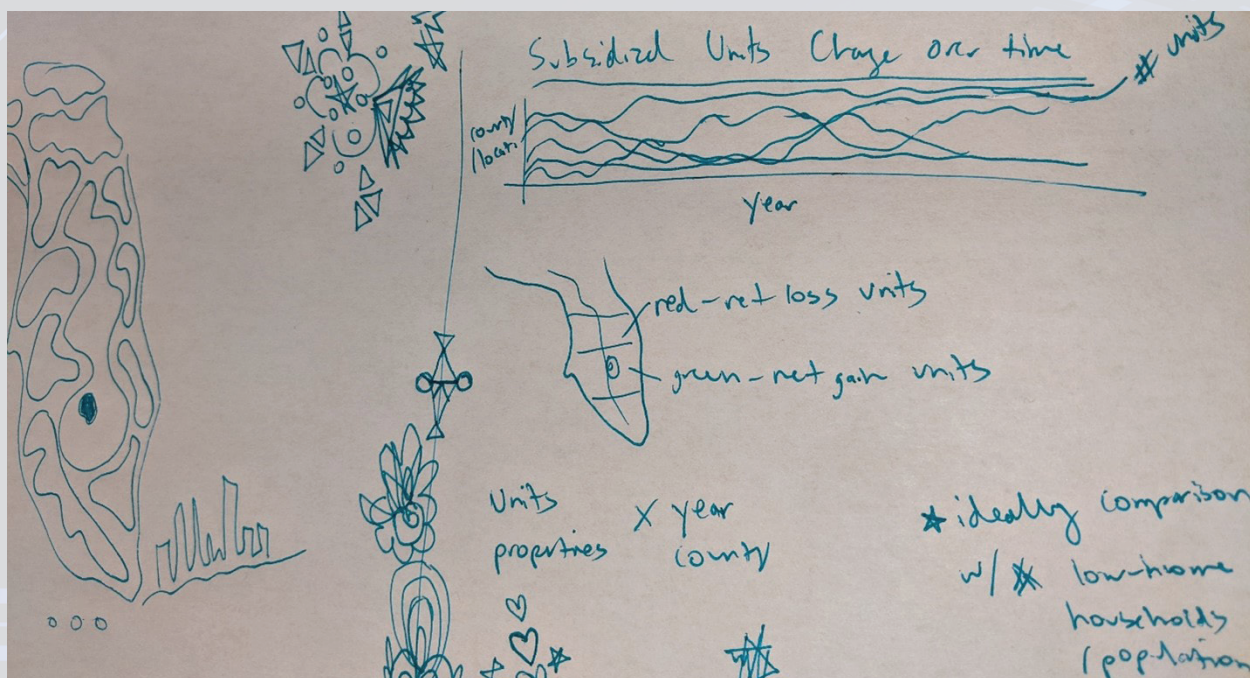


Figure 30: Wireframe of a Dashboard for the 2022 Home Matters, Example

Wireframing

Wireframing is the process of laying out elements and understanding how they will work together before creating a final product. Wireframing can save a lot of time and be very useful: rather than creating a final product sight unseen, it gives the creator a chance experiment with look and content before putting in all the work with the final product.

For app designers or other, more advanced visualizations, it makes sense to have a digital wireframing tool such as Azure. But for simpler dashboards, writing and drawing out ideas on paper is easier, faster, and allows input from a wider variety of people. When thinking about visualizing data like the dashboard above, consider what information you want included. Do you want bar charts, pie charts, hierarchy charts? Do you want a few big numbers? What filters will you have, how will they interact, and where will they sit on the dashboard?

Drawing out a simple design, but it is even more useful if it is possible to incorporate the ideas of a team of people. Have three or four people draw out their ideas for the information before talking to each other. You will be surprised at the different ways people choose to visualize the same information and, after discussing the final product as a group, will be able to create a better product more quickly than would have been possible without this simple planning exercise.

Notice the informality of the example on this page. The map of Florida is drawn simply. A first attempt to the left has been abandoned and then, as the creator was thinking about the second draft, used as a basis for a doodle. During the wireframing stage it is important to let your mind wander, abandon early attempts, give yourself the space to think through what you are doing without judgement.

Key Sources for Housing Planning Data in Florida

A list of over 50 data sources is included in Appendix 1. This section provides broad overviews and descriptions of the resources available and the most important sources for housing-focused data in Florida.

The Shimberg Center for Housing Studies Data Clearinghouse

Website: <http://flhousingdata.shimberg.ufl.edu/>

Data Provided:

- Housing affordability
 - Affordable and Available Housing
 - Housing Cost Burden
- Assisted Housing Inventory
- Comprehensive Plan Data
- Condos & Manufactured Housing
- Income & Rent Limits
- Data from the Home Mortgage Disclosure Act
- Parcels & Sales for All Counties in Florida
- Population & Household Projections
- Information on Special Needs Populations
- Maps & Visualizations
- COVID-19: Workforce & Housing Indicators
 - Recent Evictions
- REACH (Tampa Bay Area)
- Disaster Response from OPEN FEMA

As highlighted in the “Quick and Easy Data Points and Analysis” section, The Shimberg Center for Housing Studies at the University of Florida conducts research into housing policy and planning with a central focus on housing affordability in the state of Florida. More importantly for the purposes of this guide, the Center provides data and applied research to the State, local planners, for-profit and nonprofit developers, and anyone else interested in housing policy in Florida.

At the Coalition, the Shimberg Center is our first go-to source for housing data. Even more important than the huge amount of information provided for free on their website is their responsive staff who can provide data (such as parcel data) not available on their website and answer any questions you have about housing affordability in your area.

Local Consolidated Plans and Other Federal Plans

Website: Available on your local government's website.

Data Provided:

- Economic and Housing Background
- Focus Areas
- Affordable Housing Units and Units Expected to be Lost
- Activities to be Carried Out
- Barriers to Affordable Housing
- Populations of Concern
- Over 40 tables!

As highlighted in the “Quick and Easy Data Points and Analysis” section, federal plans such as the Consolidated Plan provide a plethora of data and analysis already conducted by staff or a consultant in your community. While the data in Consolidated Plans can be dense and somewhat inaccessible, they are a great place to start when looking for data to present and visualize. Before generating your own information from a variety of sources that have not been visualized or formatted, consider looking at what is already prepared in the Consolidated Plan.

Federal Government Data Sources

The federal government collects and disseminates an incredible amount of data, paralleled by few organizations on earth. This data is distributed through dozens of government offices. This section provides a list of the sources the Coalition has found most applicable to housing policy research.

US Census

The US Census serves as the nation's leading provider of data. The US Census conducts hundreds of surveys and Censuses, but the following four are among the most accessible and useful for housing planning. This section will highlight three of the data sources most important for housing, but all surveys conducted by the Census are included in the chart on page 61 and can be accessed on the Census' website.

Decennial Census

Website: <https://data.census.gov/cedsci/>

Data Provided:

The 10 Year or Decennial Census is the measurement of every person living in the United States. Required by the Constitution, the Census is used to apportion the number of seats in the House of Representatives. The first Census was conducted in 1790 when the population of the US was just under four million.

Data from the Census is extremely accurate because nearly everyone in the US responds. It provides detailed demographic data but lacks many of the questions around income and housing used by the American Community Survey. Also, because the Census data is only collected every 10 years, it is most useful early in each decade (e.g., 2010 or 2011) and less useful later in the decade (e.g. 2008, 2009), because by the end of the decade the information is almost 10 years old.

American Community Survey

Website: <https://data.census.gov/cedsci/>

Data Provided:

- Demographics
- Economics
- Housing
- Tenure

The American Community Survey (ACS) is a yearly, ongoing survey that asks a far broader set of questions than the US Census and is used by HUD and other agencies to pinpoint resources. The ACS includes questions on demographics, economics, housing, and tenure, providing relatively recent and accurate information for every part of the country. Tables DP02, DP03, DP04, and DP05 within the Census viewer provide broad overviews of the social, economic, and housing characteristics of an area and are an ideal place to start when beginning to look into the ACS data available for your community, though there are hundreds of tables that break down this data more granularly.

Sidebar on Accuracy, Reliability, and Validity

When thinking about accuracy, whether the information you are collecting adequately measures reality and “fit” or whether the data is useful when answering the questions you are posing, the concepts of reliability and validity become relevant.

Fundamentally, the analyst wants their number to be accurate – that is, as close to the truth as possible – and fit, or relevant to the questions. The ACS provides two main products, the 1-Year and 5-Year estimates. How do we choose between them, and in general, how do we choose between any different data sources?

The 5-Year ACS uses 5 years of surveys for greater reliability. With the 5-Year Estimates, the Census asked a sample (1,000 for this example) how much their house was worth in 2015, ‘16, ‘17, ‘18, and ‘19. They then used all 5,000 of those surveys and created a “5-Year Estimate.”

However, this raises an obvious problem: while the 5-Year Estimate is extremely reliable because it has a much larger sample size that makes the risk of random error (of getting one or two extremely high or low samples that throw off the final estimate) very low, is it actually measuring what we want it to measure (validity), since it’s incorporating data from 5 years ago? In a normal place at a normal time, yes, the greater reliability more than makes up for slightly older information. Some of the samples used are from 5 years ago, but in general, things don’t change dramatically over 5 years. This need for reliability is particularly true at smaller geographies, where a small city or a single Census tract might, if the 1-Year Estimate was used, be based on only a few survey responses and be extremely imprecise and unreliable. This is also true when trying to understand smaller cohorts (such as racial or age cohorts), because the number of surveys within each cohort may be very small, leading to high standard deviations and low reliability.

However, for large areas or for fast moving indicators, using the 1-Year Estimates may be more fit for this purpose because it is more accurately measuring what you are looking for (recent data). For population or housing prices in fast growing areas, 1-Year Estimates or other sources (such as data from the Florida Realtors) may be more accurate, even if it lacks some of the broad statistical reliability of a much larger and longer-term survey.

While there are statistical ways to evaluate data fit, for most of the sources discussed here it makes more sense to look at the standard deviations provided and think about what exactly you are looking for: the most reliable number provided via a larger, longer-term sample, or the most recent data even if that data is less reliable.

American Housing Survey

Website: <https://www.census.gov/programs-surveys/ahs.html>

Data Provided:

- Housing Cost
- Housing Quality
- Household Characteristics

The American Housing Survey (AHS) is the most comprehensive survey on housing conducted by the US Census. It provides a plethora of information on housing quality, cost, and other factors. Unfortunately, the AHS data is only available for certain communities, so it may not be available depending on where you are attempting to study.

US Census Permits and Construction Surveys

Website: <https://www.census.gov/construction/bps/>

Data Provided:

- Construction
- Permits
- Sales
- Housing Characteristics

The various construction surveys (looking at permits, sales, housing characteristics, and permits) are far narrower than the other two highlighted surveys. They are both extremely relevant to housing planning (providing statistics specifically around home construction) and, in the case of permits, updated monthly, allowing local governments to access the most up to date information.

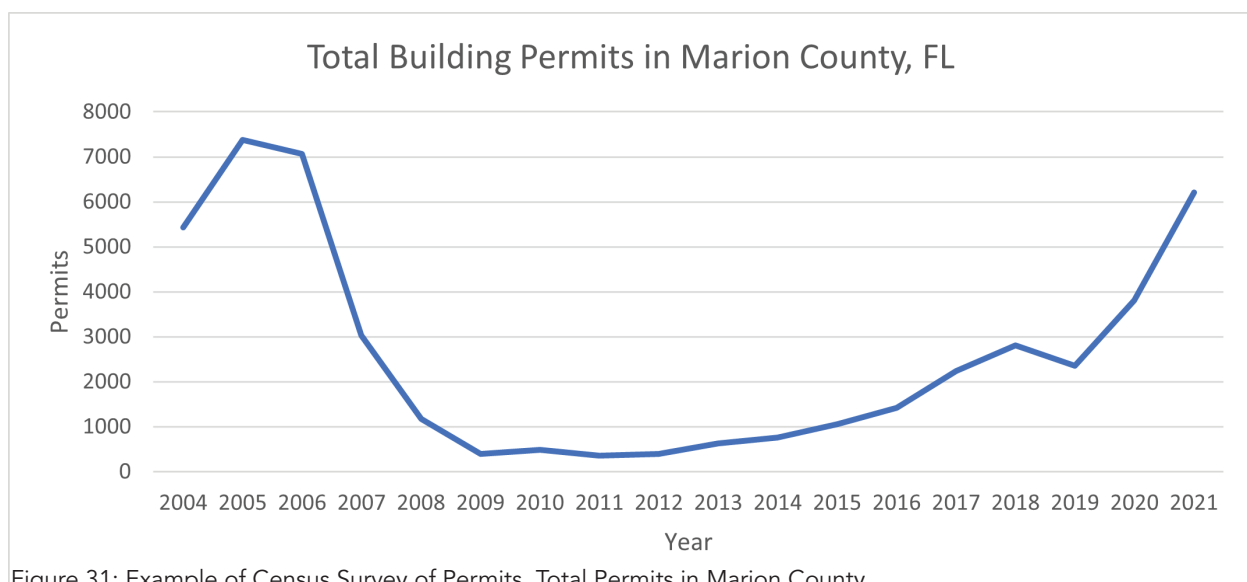


Figure 31: Example of Census Survey of Permits, Total Permits in Marion County

Surveys Conducted by the US Census

- Census
- Household Pulse Survey (COVID-19)
- Small Business Pulse Survey
- Academic Libraries Survey
- Advance Monthly Sales for Retail and Food Services (MARTS)
- American Community Survey (ACS)
- American Housing Survey (AHS)
- American Time Use Survey (ATUS)
- Annual Business Survey (ABS)
- Annual Capital Expenditures Survey (ACES)
- Annual Parole Survey & Annual Probation Survey
- Annual Retail Trade Survey (ARTS)
- Annual Services Report (Service Annual Survey)
- Annual Survey of Entrepreneurs
- Annual Survey of Jails
- Annual Survey of Manufactures (ASM)
- Annual Survey of Public Employment & Payroll (ASPEP)
- Annual Survey of School System Finances
- Annual Survey of State Government Tax Collections (STC)
- Annual Survey of State and Local Government Finances
- Annual Wholesale Trade Survey (AWTS)
- Ask U.S. Panel Pilot
- Beginning Teacher Longitudinal Study (BTLs)
- Building Permits Survey (BPS)
- Business Enterprise Research and Development Survey (BERD)
- Business and Professional Classification Survey
- Census of Governments
- Census of Jails
- Census of Juveniles in Resident Placement (CJRP)
- Census of State and Federal Adult Correctional Facilities
- Commodity Flow Survey (CFS)
- Common Core of Data (CCD)
- Construction Progress Reporting Survey (CPRS)
- Consumer Expenditure Survey (CE)

- County Business Patterns (CBP)
- Current Population Survey (CPS)
- Decennial Census of Population and Housing
- Economic Census
- Economic Census Industry Classification Report
- Export Statistics
- Exports from Manufacturing Establishments
- Federal Assistance Awards Data System (FAADS)
- Federal Audit Clearinghouse (FAC)
- Household Pulse Survey (COVID-19)
- Housing Vacancy Survey (HVS)
- Identity Theft Supplement – part of the National Crime Victimization Survey (NCVS)
- Import Statistics
- Justice Assistance Data Survey
- Juvenile Residential Facility Census (JRFC)
- La Encuesta sobre la Comunidad Estadounidense
- La Encuesta sobre la Comunidad de Puerto Rico
- Management and Organizational Practices Survey (MOPS)
- Manufactured Housing Survey (MHS)
- Manufacturers' Shipments, Inventories, and Orders Survey (M3)
- Manufacturers' Unfilled Orders Survey (M3UFO)
- Manufacturing Energy Consumption Survey (MECS)
- Medical Expenditure Panel Survey (MEPS)
- Monthly Retail Trade Survey (MRTS)
- Monthly Wholesale Trade Survey (MWTS)
- Mortality in Correctional Institutions (MCI)
- National Ambulatory Medical Care Survey (NAMCS)
- National Crime Victimization Survey (NCVS)
- National Health Interview Survey (NHIS)
- National Hospital Ambulatory Medical Care Survey (NHAMCS)
- National Household Education Survey (NHES)
- National Juvenile Justice Directory Program
- National Prisoner Statistics (NPS)
- National Sample Survey of Registered Nurses (NSSRN)
- National Survey of Children's Health (NSCH)
- National Survey of College Graduates (NSCG)

- National Survey of Fishing, Hunting, & Wildlife-Associated Recreation
- National Teacher and Principal Survey (NTPS)
- National Training, Education, and Workforce Survey (NTEWS)
- New York City Housing and Vacancy Survey (NYCHVS)
- Nonemployer Statistics
- Police-Public Contact Survey (PPCS)
- Principal Follow-Up Survey (PFS)
- Private School Universe Survey (PSS)
- Public Libraries Survey (PLS)
- Puerto Rico Community Survey (PRCS)
- Quarterly Financial Report (QFR)
- Quarterly Services Survey (QSS)
- Quarterly Summary of State and Local Government Tax Revenue (QTAX)
- Quarterly Survey of Plant Capacity Utilization (QPC)
- Quarterly Survey of Public Pensions (QSPP)
- Rental Housing Finance Survey (RHFS)
- Report of Organization
- School Crime Supplement (SCS)
- School Pulse Panel
- School Survey on Crime and Safety (SSOCS)
- Small Business Pulse Survey
- Special Census Program
- State Library Administrative Agencies Survey (SLAA)
- Statistics of U.S. Businesses (SUSB)
- Survey of Business Owners and Self-Employed Persons (SBO)
- Survey of Construction (SOC)
- Survey of Income and Program Participation (SIPP)
- Survey of Market Absorption of New Multifamily Units (SOMA)
- Survey of Sexual Victimization (SSV)
- Teacher Follow-Up Survey (TFS)
- Telephone Point of Purchase Survey (TPOPS)
- Value of Construction Put in Place Survey (VIP)
- Vehicle Inventory and Use Survey (VIUS)

Consumer Financial Protection Bureau: Home Mortgage Disclosure Act Data

Website: <https://www.consumerfinance.gov/data-research/hmda/historic-data/>

Data Provided:

- Location of Loans
- Characteristics of Borrowers
- Denial Statistics and Characteristics

The Home Mortgage Disclosure Act (HMDA) requires the majority of financial institutions to collect and disclose loan-level information about mortgages. This data includes race, income, outcome of the loan, debt-to-income ratio, reason for denial, and other factors. The HMDA information is important for understanding where loans are going into your community and who is or is not accessing loans.

HUD Picture of Subsidized Housing

Website: <https://www.huduser.gov/portal/datasets/assthsg.html>

Data Provided:

- Public Housing Authority Data
- Other Multifamily Programs
- Populations and Numeric Codes

HUD provides an overview of data at the public housing authority (PHA) level, as well as other HUD programs at various levels. While it does not cover funding through CDBG, Rural Housing Services, Indian Housing, or HOME, this program provides data on the myriad of ways housing is provided.

HUD AFFH Data Sets

Website: <https://egis.hud.gov/affht/>

Data Provided:

The Fair Housing Act requires HUD and its recipients to intentionally work to overcome patterns of segregation and exclusion; that is most local governments are required to affirmatively further fair housing (AFFH). While different executive governments have interpreted the AFFH language differently, HUD provides a variety of maps and data sources to help local governments affirmatively further fair housing and address discrimination. These are all available through an easy-to-use data viewer which, whether planning for AFFH or not, can provide useful information on a variety of demographic and economic topics.

Federal Emergency Management Agency: Open FEMA

Website: <https://www.fema.gov/about/reports-and-data/openfema>

Data Provided:

- Disaster
- Hazard Mitigation
- Disaster Response

The Federal Emergency Management Agency collects a huge amount of information on a variety of aspects of emergency management, including individual and public assistance, the National Flood Insurance Program, and information on specific disasters and hazards. This information is generally available at the Census tract level and larger geographies and can be useful for understanding how communities have or are likely to react to disaster.

US Bureau of Labor Statistics

Website: <https://www.bls.gov/>

Data Provided:

- Labor Market Activity
- Working Conditions
- Price Changes
- Productivity

The Bureau of Labor Statistics measures all things related to labor in order to support data-informed decision making. The Coalition regularly uses Bureau wage and occupation statistics to better understand what the workforce can afford.

Other Key Data Sources

While less comprehensive than many of the sources above, many nonprofits and universities create data sources that answer specific questions extremely well. These data sources often are also pre-visualized, saving you time and energy. Here are a few that the Florida Housing Coalition has found most helpful.

Florida Realtors

Website: <https://www.floridarealtors.org/newsroom/market-data>

Data Provided:

- Recent sales and prices

While much of the information created by multiple listing services (MLS) and Realtors Property Resource (RPR) is proprietary or not free to access, the Florida Realtors do provide monthly, quarterly, and yearly reports at the statewide and MSA level. The information in these reports is the most up-to-date and all-inclusive data available for free on housing sales.

United Way, United for ALICE (Asset Limited, Income Constrained, Employed) Reports and Dashboards

Website: <https://unitedforalice.org/florida>

Data Provided:

- ALICE numbers, demographics, and breakdowns
- Overview of income needed to afford basic necessities

The Asset Limited, Income Constrained, and Employed (ALICE) reports by the United Way provide an extremely detailed account of the income necessary to afford basic necessities by different family types (single, married couple with two small children, elderly). Unlike the poverty rate, which uses an archaic formula based on the cost of food, the ALICE team at United Way carefully tracks how much low-income families are paying for things like food, housing, transportation, insurance, clothing, etc. This report is only available at certain geographies, but is a far more useful benchmark than the poverty rate as a measure of need and gives policy makers a fuller understanding of what it actually costs to live in their community. In the last few years, ALICE has also begun adding interactive dashboards that providing more detailed information by area or demographic group.

Zillow Data Resources

Website: <https://www.zillow.com/research/data/>

Data Provided:

- Home Values
- Forecasts
- Rental Index
- Inventory Statistics
- List and Sale Price
- Sales Count and Price Cut

Zillow, the online real estate marketplace, provides a huge amount of its data for free. As with the information from Florida Realtors, Zillow tracks and releases information on an ongoing basis. This data is among the most up-to-date available. Zillow is also the only place for up-to-date rental information, though this is focused on rentals available on Zillow and may not include lower cost, less formal rental housing.

Appendix 1: List of Data Resources

Data Source	Website	Comments	Type	Pre-Visualized Data
Shimberg Center for Housing Studies	http://www.shimberg.ufl.edu/	Broad, easily accessible information across Florida	Housing	Yes
ACS	https://data.census.gov/cedsci/	Most general and broadest source of data available, key source of demographics and economic data	General	Yes, but difficult to navigate
American Housing Survey	https://www.census.gov/programs-surveys/ahs.html	Ideal source for housing related data but only provides some information at the MSA level and no information at the county level	Housing	No
Building Permits	https://www.census.gov/construction/bps/	Provides up to date (monthly) information rather than yearly or bi-yearly like many other sources on housing permits at Region, State, MSA, and for some places	Housing	No
National Housing Preservation Database	https://preservationdatabase.org/	Free account allows you to download housing preservation data	Housing	No
HUD Location Affordability Index	https://www.hudexchange.info/programs/location-affordability-index/	Similar to H+T Index	Housing	Yes
H+T Index (Housing + Transportation)	https://htaindex.cnt.org/	Provides information and housing costs at an extremely granular level as well as at the community level		Yes
HMDA Data	https://www.consumerfinance.gov/data-research/hmda/historic-data/	Data on loan origination, also available through the Shimberg Center for Housing Studies upon request	Housing	No
Special Request Shimberg Center for Housing Studies	http://www.shimberg.ufl.edu/	Evictions and parcel level appraisal data are accessible, reach out to Anne Ray for data sharing agreement	Housing	No
LIHTC HUD Data	https://lihtc.huduser.gov/	Also available on Shimberg's site under "Assisted Housing Inventory"	Housing	No
AIDSVU HIV/AIDS Data	https://aidsvu.org/	Provides the best resource on HIV/AIDS related data, only available in some communities	Health	Yes
HUD Picture of Subsidized Housing	https://www.huduser.gov/portal/datasets/assthsg.html	Overview and HUD Subsidized Housing. Most of this information can be found on Shimberg's website	Housing	Yes

ASTDR Social Vulnerability Index	https://www.atsdr.cdc.gov/placeandhealth/svi/publications/publications_materials.html	One of several useful social vulnerability index	Emergency	Yes
Smart Location Database	https://www.epa.gov/smartgrowth/smart-location-mapping#SLD	Provides data to encourage smarter, more efficient growth as well as walkability	Housing	Yes
CHAS data	https://www.huduser.gov/portal/datasets/cp.html	Provides all of the data available in your consolidated plan	Housing	No
FEMA Open Data	https://www.fema.gov/about/openfema/data-sets	Broad data on disaster and recovery	Housing	No
Mapping Inequality Redlining Maps	https://dsl.richmond.edu/panorama/redlining/	Map of historic redlining maps, only available in select communities but an excellent resource for understanding racial inequity	Equity	Yes
Opportunity Index	https://opportunityindex.org/	Provides a map of areas of opportunity, but less useful and user friendly than the Opportunity Atlas	Economic	Yes
Opportunity Atlas	https://www.opportunityatlas.org/	Maps areas of opportunity and other demographic factors, extremely usable	Economic	Yes
Federal Reserve Economic Data	https://fred.stlouisfed.org/	Plots a huge variety of economic indicators without the need for further visualization	Economic	Yes
USDA Food Environment Atlas	https://www.ers.usda.gov/data-products/food-environment-atlas/	Research and data on food deserts	Health	Yes
CDC Statistics	https://www.cdc.gov/datastatistics/index.html	Most comprehensive data on disease and health	Health	Yes
Civil Rights Data Collection	https://ocrdata.ed.gov/	Information collected on civil rights violations at the local level	Equity	No
EPA Environmental Justice Mapping and Screening Tool	https://www.epa.gov/ejscreen	Tracks the unequal fallout of environmental pollution and disaster	Equity	Yes
HRSA Mapping	https://data.hrsa.gov/	Useful for seeing availability of health resources in a neighborhood	Health	Yes
Access Across America	https://access.umn.edu/research/america/	Information on accessibility for people with disabilities	Equity	No

SBA Data	https://www.sba.gov/about-sba/sba-performance/open-government/digital-sba/open-data/open-data-sources	Provides all Small Business Association data	Economic	No
HUD School Proficiency Index	https://hudgis-hud.opendata.arcgis.com/	Maps the quality of schools in an area, which ties into equity, housing costs, etc.	Education	Yes
Florida Realtors Data	https://www.floridarealtors.org/newsroom/market-data	Provides up to date information on recent home sales at the state and MSA level	Housing	Yes
Qualified Opportunity Zones	https://www.irs.gov/credits-deductions/businesses/opportunity-zones	Provides information on where Qualified Opportunity Zones are	Economic	Yes
EPA Geo-Spatial Download	https://www.epa.gov/frs/geospatial-data-download-service	All EPA geo-spatial (mapping) downloads	Health	No
R/ECAPs	https://hudgis-hud.opendata.arcgis.com/	Information on racially/ethnically concentrated areas of poverty	Equity	Yes
LEHD Data	https://lehd.ces.census.gov/	Census data on longitudinal employment and housing trends		Yes
FCC Broadband Data	https://broadbandmap.fcc.gov/#/	Information on broadband accessibility	Economic	Yes
BLS Data	https://www.bls.gov/data/	Best breakdown of income by occupation, only available at MSA level	Economic	Yes
Census Transportation Planning Products	https://ctpp.transportation.org/	Transportation planning data from the Census		No
County Business Patterns	https://www.census.gov/programs-surveys/cbp.html	Industry Data by County	Economic	Yes
Eviction Lab	https://evictionlab.org/	More detailed information available at Shimberg, currently housed under COVID tab	Housing	Yes
National Air Toxics Assessment	https://www.epa.gov/national-air-toxics-assessment	Environmental Justice	Health	Yes

Housing and Vulnerability Research Institute Emergency Vulnerability Index	http://artsandsciences.sc.edu/geog/hvri/hvri-resources	County social vulnerability	Emergency	Yes
Bureau of Economic Analysis	https://www.bea.gov/data	Economic analysis data from the Bureau of Economic Analysis	Economic	Yes
TIGER Line Files	https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html	Central source for government Shapefiles	General	No
Florida National Hydrography Data	https://floridadep.gov/dear/watershed-services-program/content/about-florida-national-hydrography-dataset	Watershed information, useful for emergencies and climate change mapping	Emergency	No
FHFA Housing Price Index	https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index.aspx#:~:text=The%20FHFA%20HPI%20is%20a%20weighted%2C%20repeat%2Dsales%20index%2C,refinancings%20on%20the%20same%20properties.&text=The%20FHFA%20HPI%20serves%20as,trends%20at%20various%20geographic%20levels.	Appreciation Index	Housing	Yes
Freddie Mac House Price Index	http://www.freddiemac.com/research/indices/house-price-index.page	Preferred source for charting appreciation	Housing	No
AARP Surveys	https://www.aarp.org/research/data-tools/datasets/	Surveys on health, housing, lifestyle for older adults	Health	No
FBI Hate Crimes Data	https://ucr.fbi.gov/hate-crime	Source for all information on hate crimes in your community	Equity	No
UpJohn Data	https://www.upjohn.org/data-tools	Incentives and Taxation, New Hires, and Promise Scholarships	Economic	No
Zillow Research	https://www.zillow.com/research/data/	Useful for tracking home sales data	Housing	Yes

All HUD Data sets	https://www.huduser.gov/portal/pdrdatas_landing.html	Broad information on housing	Housing	No
The Real Cost of Housing	https://www.policymap.com/issues/housing-quality/	City Level Data on Home Repair Need. Working on zip code and Census Tract level data	Housing	Yes
Survey of Household Economics and Decision Making	https://www.federalreserve.gov/consumerscommunities/shed_data.htm	Great source for housing economics data but very raw/ needs a lot of processing	Economic	No
Voter Study Group	https://www.voterstudygroup.org/publication/nationscape-data-set	Extremely granular demographics information needs lots of processing		No
Travel Time Map Demo	app.traveltime.com	Isochrone map of travel time by mode (car, public, bike, walk) at a certain time		Yes
Open Street Map	openstreetmap.org	Road map, shows road type, also some other attributes like parks, cemeteries, churches, bus stops, interstate on/off ramps		Yes
HUD Point in Time and Housing Inventory Count Data	https://www.hudexchange.info/programs/hdx/pit-hic/	Most in depth overview of homelessness counts and services	Housing	No

Appendix 2: List of Key Terms

Accuracy is the degree to which your data represents reality.

Analysis is the systematic exploration of data. Analysis can take the form of statistical, narrative, and/or logical techniques with the purpose of condensing and evaluating data. Simplified, analysis is the conversion of data points into understandable and actionable content that you and other people can use.

Area Median Income (AMI) refers to the income “in the middle” of all incomes in the geography being analyzed.

Cost Burden is when a household is paying 30% or more of its income towards housing. Households that are cost burdened, particularly when they are also low-income, have difficulty paying for other necessary expenses such as food or medical care, and may find it difficult or impossible to save for emergencies, putting them at risk of homelessness.

Data is any information that can be used for reasoning. More colloquially, when people talk about data, they often mean information that has been collected into a digital form and primed for easy analysis.

Data Collection is the gathering together of information. Data collection can be as simple as downloading a pre-collected secondary data set but can also include accessing and cleaning existing government information or sending out surveys or conducting public meetings.

Data Dictionaries are resources put together to describe what a data set contains and how it is formatted and structured. Large data sets are often extremely complicated and difficult to access for people other than the creator, so it is important to have good data dictionaries to make sure other people can use your data.

Data Literacy is the ability to understand, create, and use data. Different organizations, and different people within organizations, will have different appropriate levels of data literacy.

Data Management is the entire process of collecting, processing, and storing data. Your organization already has (probably several) data management systems.

Extremely Low-Income refers to households making 30% of the area median income or less.

Fit for Purpose means a data source is appropriate for its intended use. Data analysts should check to make sure that data is reliable and appropriate to answer the question posed.

Geography is an area for which data is collected. You may be interested in data at the state, county, city, neighborhood, or parcel level and there are different data sources available for these different geographies.

Low-Income refers to households making 80% of the area median income or less.

Metropolitan Statistical Areas (MSAs) are regions with a central city. The city and the surrounding counties that have significant economic and social interactions with the central city are part of the MSA.

Precision is the degree to which your data is repeatable under similar circumstances. Precision is similar to reliability, but with a focus on both the changing nature of your sample as well as the consistency of your measurement methods.

Primary Data is data that you have collected. Governments collect huge quantities of data automatically (from property appraiser information, code enforcement, zoning and permitting, utilities, etc.). For specific questions, local governments also conduct surveys and focus groups which are another type of primary data.

Reliability is when your data or data collection tools are collecting information in a consistent way.

Secondary Data is data that has already been collected and is published for analysis. Much of the data discussed in this publication is secondary data.

Shapefiles are a data storage format for storing geographic information. While originally created for ESRI's ArcGIS software, Shapefiles are now used in a variety of visualization software.

Validity is the extent to which your measurement or data is actually measuring what you want it to measure.

Very Low-Income refers to households making 50% of the area median income or less.



The Florida Housing Coalition, Inc., is a nonprofit, statewide membership organization which brings together housing advocates and resources so that all Floridians have a quality affordable home and suitable living environment.

FOR MORE INFORMATION,
CONTACT:

The Florida Housing Coalition
1311 N. Paul Russell Road, B-201
Tallahassee FL 32301
(850) 878-4219
info@flhousing.org
FLHousing.org