Affordable Housing in Florida:
Systematic Literature Review and
Exploratory County-Level Data Analysis

Namratha Kulkarni
n01380801@UofNorthFlorida.onmicrosoft.com
Florida Blue
Jacksonville, FL, USA

Bharani Kothareddy
bharani0223@gmail.com
CSX
Jacksonville, FL, USA

Karthikeyan Umapathy
k.umapathy@unf.edu
School of Computing, University of North Florida
Jacksonville, FL, USA

Abstract

Despite the severe shortage of affordable housing in the United States, the social issue of homeownership is not well-understood and under-researched. We discuss two different studies in this paper. The first study presents the trends identified from articles relevant to affordable housing and data science approaches. The articles shortlisted for the first study were selected through a systematic literature review. The second study examines patterns and trends in housing characteristics, ownership, and occupancy at the Florida county level. More specifically, this research aims to understand the home affordability for renters vs owners using publicly available data. This paper analyzes housing data gathered from the U.S. Census Bureau. Using Tableau, we created interactive maps and visualizations to conduct exploratory data analysis and present the findings. Our analysis revealed significant disparities in housing conditions, ownership, and occupancy between different regions and counties in Florida. We also found that demographic variables such as age and income are key factors in relevance to housing patterns and trends.

Keywords: affordable housing, systematic review, exploratory data analysis, visual analytics, census datasets, demographic disparities.

1. INTRODUCTION

Housing is one of the most important factors in the socioeconomic development of a thriving society. Effective housing supply is one of the government's policies to ensure effective affordability for people from all income levels, most importantly for the middle- and low-income levels (Adabre & Chan, 2019). Housing affordability refers to getting a particular standard of housing at a price or rent that does not impose an undue strain on household incomes in the eyes of a third party (typically the government) (McCrone & Stephens, 2017). Affordable housing as a social issue has been undergoing considerable changes in many countries for a number of years, with a steady reduction in the availability of housing units and associated support mechanisms (Czischke & van Bortel, 2023).

A number of terms have been used to explain housing forms that are affordable to middle- and low-income earners or income-constraint families. Depending on the country, several of these phrases could refer to various housing tenures. Affordable housing is a commonly used term in the United States and can refer to both rental and ownership housing. Public and social housing terms are more strongly associated with rental housing in the context of Europe. Cooperative housing is used in some European countries and refers to housing where the entire building is owned in common by a homeowners' association (Czischke & van Bortel, 2023). Despite the benefits of maintaining housing affordability and accessibility for socioeconomic growth, the global housing affordability dilemma continues to be unsolvable (Galster & Lee, 2021).

In both developed and developing countries, the housing unaffordability crisis continues to be a major challenge. In both rich and developing countries, the housing crisis is demonstrated by a lack of housing facilities, which has the consequence of promoting the construction of slums (overcrowded and dilapidated urban residential facilities with inadequate infrastructure) (Golubchikov & Badyina, 2012). Affordable housing has become a key issue impacting the lives of many residents, particularly the population living in households that spend 40% or more of their disposable income on housing (Haffner & Hulse, 2021). Apart from the housing shortages, several of the existing affordable housing facilities are insufficient. If left unchecked, the housing scarcity and its inevitable outcome could get considerably worse, given that the world's population is expected to expand from 3.6 billion to 6.3 billion by 2050 (Golubchikov & Badyina, 2012). In response to the global housing problem, international organizations such as the World Bank and the United Nations (UN) have begun to develop policies to ensure enough housing. As a result, in addition to assuring housing affordability, achieving social and environmental sustainability goals for comprehensive sustainable development continues to be a hot topic in both developed and developing countries. The studies on affordable housing can broaden the knowledge base on which policymakers can base their decisions. More crucially, systematic and data-driven approaches to understanding affordable housing issues enhance the effect of existing knowledge on policy formation.

Research on affordable housing has mainly focused on descriptive explanations and policy causality analysis (Czischke & van Bortel, 2023). However, limited attention has been paid to understanding how affordable housing varies for population groups. To shed light on this gap, we conduct exploratory data analysis. As an exploratory data analysis would be effective when focusing on specific regions and for reasons for future work of the research team, we focus our research on Florida. Thus, this research paper aims to analyze the affordable housing situation in various counties in Florida using data from the U.S. Census Bureau datasets. The paper provides a detailed analysis of owner and renter-occupied housing units and the count of rooms in each unit for each Florida county. The population and demographic data of each county were also taken into consideration.

2. LITERATURE REVIEW OF DATA SCIENCE APPROACHES TO INVESTIGATE AFFORDABLE HOUSING

There has been a limited systematic review of the literature concerning data science and data analytics approaches to investigate the affordable housing context. Before we conducted our exploratory data analysis, we wanted to identify other literature that has created data visualizations and mapping of affordable housing trends.

Several systematic literature reviews on affordable housing have been conducted. Adabre and Chan (2019) conducted a review of 34 articles to determine the critical success criteria to attain sustainability in affordable housing. McCabe et al. (McCabe, Pojani, & van Groenou, 2018) focused their review on 67 studies in relevance to identifying the success factors,

barriers, and motivations for the application of energy technologies in social housing. Anderson et al. (2003) reviewed 23 articles and reports that studied affordable housing and the reduction of segregation of low-income families. Although these reviews contribute greatly to the literature on affordable housing and other sustainability goals, they were not focused on articles that utilize data science approaches to address the affordable housing issue. Thus, we decided to conduct a systematic literature review to identify articles that investigate affordable housing issues using data science approaches. We aimed to gain knowledge on data science tools, methods, models, and techniques employed by researchers to study the affordable housing context. Thus, we learn from and make meaningful contributions to important social issues.

Systematic Literature Review: Data Collection

Google Scholar was chosen as the key scientific database for data collection. Keywords including "affordable housing" and "low income housing" were used in combination with "data analytics", "machine learning", and "data science" were used to ensure thorough data retrieval.

Publish or Perish software (Harzing, 2007) was used to get the details such as title, score, Article URL, abstract, cites per year, cites per author, author count, volume, issue, ECC, DOI, ISSN, Citation URL, Volume, Issue, Start page, End page, Cites per author, age, and Related URL.

Some limitation criteria were put in place to ensure the retrieval of adequate and manageable articles. The 'document type' was only limited to journal articles because they form the broadest and collective knowledge base in any given academic field, given that we are seeking articles published by any researchers not just information systems researchers.

The search results were also refined to include articles published in the English language in the seven years ranging from 2015 to 2022, inclusive. We utilize the PRISMA flow diagram to depict the summarized view of the data collection and review article selection process. Figure 1 shows the PRISMA flow diagram. Aggregation of search results for different keyword combinations resulted in 911 articles.

After the aggregation of search results, the duplicates based on the Title, Author, Abstract, and Article URL were removed, which resulted in 888 articles for further consideration. We then

applied inclusion and exclusion criteria to filter the search result set further. The criteria we applied are 'If the article is English,' 'The article abstract is relevant to affordable housing,' 'Article is accessible through the university library,' and 'Article is related to US affordable housing.' After excluding articles that did not meet the criteria, we had 79 articles. We then excluded articles that were not peer-reviewed and not focused on the United States. We further assessed the remainder of the article by reading the title and abstract to ensure they investigated affordable housing context using data science or data analytics approaches. After applying all of the exclusion criteria, the resulting articles are 20.
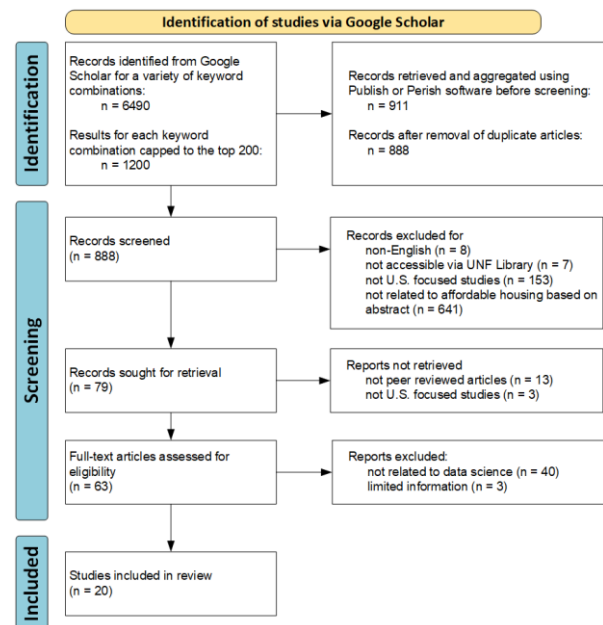

Figure 1 The PRISMA Flow Diagram of Search Result Screening

Systematic Literature Review: Data Collection

From the resulting 20 articles, we extracted the following details: data sources, tools used, analytical techniques applied, and research focus. After extracting relevant information, we created a categorical grouping of the information and calculated the frequency of the categories. Figure 2 depicts the frequency chart for data sources. Articles used multiple data sources in their research investigations. Federal and local government data sources were predominantly used by affordable housing researchers. In the figure 2, we show a breakdown of the sources for federal and local sources. It can be noted that Census, tax records, and Housing and Urban

Development (HUD) data sources are key data sources for affordable housing investigations.
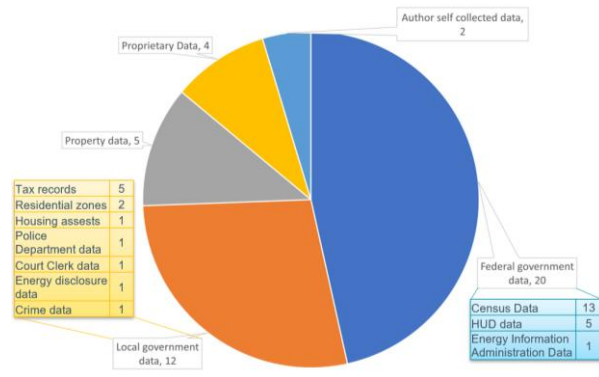

Figure 2 Data Sources

Figure 3 shows the frequency for data science tools specified in the articles. 50% of the researchers did not specify tools they utilized in their investigations. Authors, reviewers, and editors would have to ensure that tools used to perform analysis and solutions development are explicitly stated in the published articles; as these information are critical for replicability of the research. Of those specified, mapping tools which included GIS and other spatial tools, are mostly used by researchers.

Figure 4 provides counts of various data analytical approaches utilized in the identified articles. Some researchers did employ multiple techniques in their investigations. Our findings indicated that affordable housing researchers predominantly use regression analysis. Less usage of classification and clustering techniques indicates that researchers are using more statistical analysis over machine learning techniques.
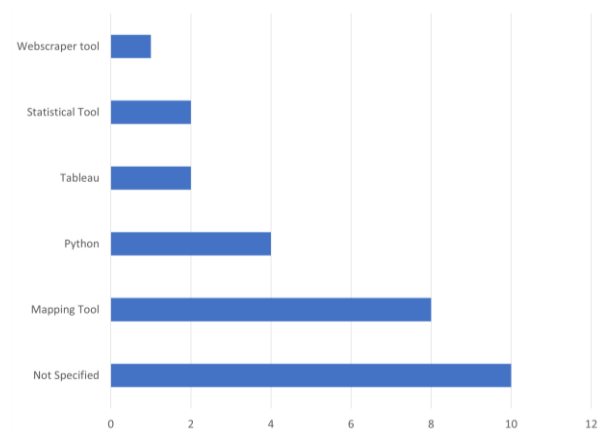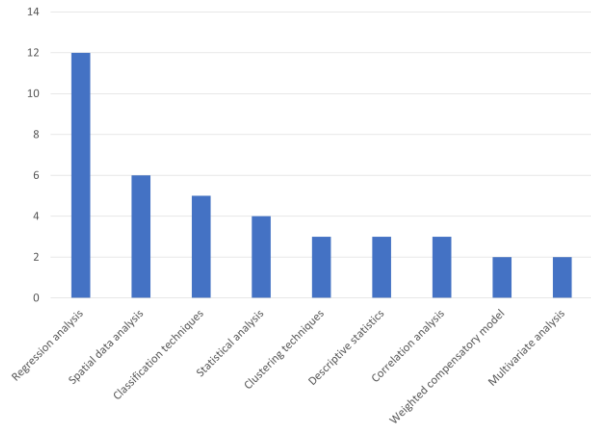

Figure 3 Data Science Tools

Figure 4 Data Analysis Techniques

With regard to key focus areas of the research, there were 12 articles where affordable housing was the primary focus, while for the remaining eight articles, affordable housing was the secondary focus area. In regards to data sizes, 11 articles did not specify the size of data used in their research, while five articles used big data and four used small datasets. Predominant of the articles (14 out of 20) did not partner with a community partner to study the affordable housing issue. The prevalent geographic focus for affordable housing research has been the city as the unit level of analysis, as shown in Figure 5.

Despite the review of existing affordable housing literature conducted using Google Scholar and multiple relevant keywords, only a small portion of the result sets had articles that focused on affordable housing using data science approaches. Findings from extracted information reveal that there is minimal commonality among the published research articles. This implies that greater investment in affordable housing research has a higher chance of making meaningful impacts for those affected by affordable housing issues.
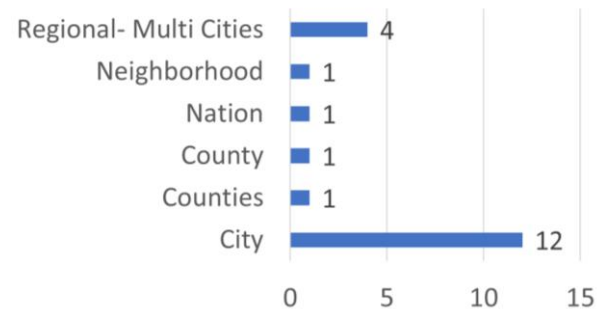

Figure 5 Geographic Focus

From the systematic literature, we noted there are limited studies focused on county as the unit level of analysis and none specifically focused on Florida. Thus, we are conducting an exploratory data analysis on affordable housing issues in Florida at the county level using publicly available datasets.

3. EXPLORATORY ANALYSIS OF AFFORDABLE HOUSING IN FLORIDA: DATA

Exploratory Data Analysis (EDA) is a crucial initial step in a data science study conducted by visualizing data to understand its characteristics, detect patterns & trends, and identify meaningful relationships. We use datasets from the Housing and Urban Development (HUD) and Census Bureau for the exploratory analysis.

As the literature has indicated, the housing cost burden and the ratio of owner-occupied to renter-occupied household units are closely associated with affordable housing issues. Thus, in this research, we aim to explore home affordability for renters vs. occupied household unit characteristics using publicly available data.

Datasets

In this section, we describe CHAS (Comprehensive Housing Affordability Strategy) and Census Bureau data profiles used in this research study.

Comprehensive Housing Affordability Strategy (CHAS)

The U.S. Department of Housing and Urban Development (HUD) obtains custom Census data from the Census Bureau that are not available through standard data releases. Using these data, HUD produces the CHAS (Comprehensive Housing Affordability Strategy) dataset to reveal the extent of housing problems and needs for low-income households (HUD, 2023). The dataset, along with household counts, contains HUD Area Median Family Income (HAMFI), which is a median family income calculated by HUD to determine Fair Market Rents and income limits for affordable housing programs. Thus, CHAS data can be used to estimate the number of households that have affordable housing problems and have incomes low enough to qualify for HUD housing programs.

Census Bureau

As CHAS data was calculated based on the 2016 - 2020 American Community Survey (ACS) 5-year estimation data, we obtained relevant 2016 - 2020 ACS 5-Year Data Census Profiles for this research. We gathered Census data that are relevant to demographics, occupancy characteristics, housing units, and financial characteristics. We have listed the Census data profiles used for this research in Table 1.

Data Profile	Description
S2501: Occupancy Characteristics	Provides renters and owners occupied housing units based on household size, occupants per room, and age of householder
S2502: Demographic Characteristics for Occupied Housing Units	Provides renters and owners occupied housing units based on race, ethnicity, age, and education attainment of householder
S2503: Financial Characteristics	Provides renters and owners occupied housing units based on household income and monthly housing costs
S2504: Physical Housing Characteristics for Occupied Housing Units	Provides renters and owners occupied housing units based on housing unit structures

Table 1: Census Data Profiles used in the research study

4. VISUAL ANALYSIS OF AFFORDABLE HOUSING IN FLORIDA

Data described in the above section was downloaded for Florida counties. Downloaded data columns were merged based on the Florida counties. In this section, we present mapping-based visuals for various data points gathered.

Population Distribution of Florida

We first created the population distribution of Florida as a baseline visual. Maimi-Dade is the most populous county, followed by Broward and Palm Beach counties. All three top populous counties in Florida are in the southeast region of Florida, stacked one over the other. The next three populous counties are Hillsborough, Orange, and Duval, which happen to be other metropolitan areas in Florida apart from Miami. The bottom five least populous counties are Liberty (being the least), Lafayette, Glades, Franklin, and Calhoun. The least populous

counties are rural counties, which are mostly in the northwest and southwest regions of Florida. Figure 6 provides a shading map of population distribution for Florida counties.
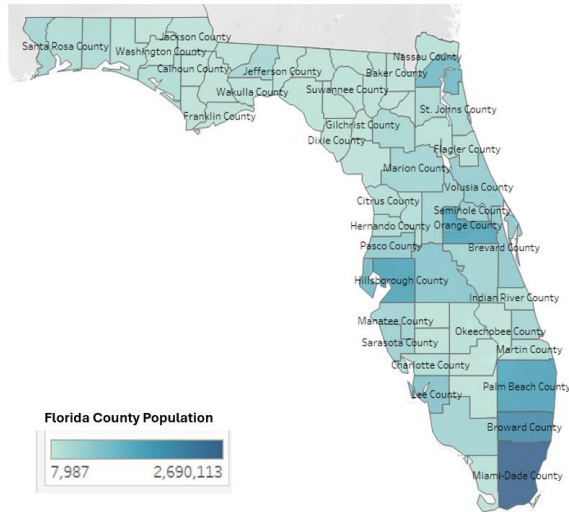

Figure 6 Florida County Population

Age Group Under 35

As the cost of home ownership and overall cost of living is increasing, younger population age group individuals would face several challenges to find affordable housing. Thus, we created a visual of the renter population for the age group under 35. We calculated a ratio of renters under 35 years old and the sum of renters of all age groups for each county. Figure 7 depicts the top 10 counties that have the highest renters ratio. Interestingly, we noted that Washington County has the highest ratio. Washington County is in the northwest region, and it is a rural county. The highest ratio is due to a high number of the younger population but less number of homeowners, leading to the majority of the younger population renting.

In a similar vein, we calculated the ratio of owners under 35 and the sum of owners for all age groups. Figure 8 shows the top 10 counties that have the highest owners ratio. We note that Collier County, where Naples is located, is the highest, followed by Duval, where Jacksonville is located, then Hillsborough, where Tampa is located, and then Orange County, where Orlando is located. Thus, it seems younger populations tend to become homeowners in metropolitan cities like Jacksonville, Tampa, and Orlando more than high the populous Miami metropolitan area.
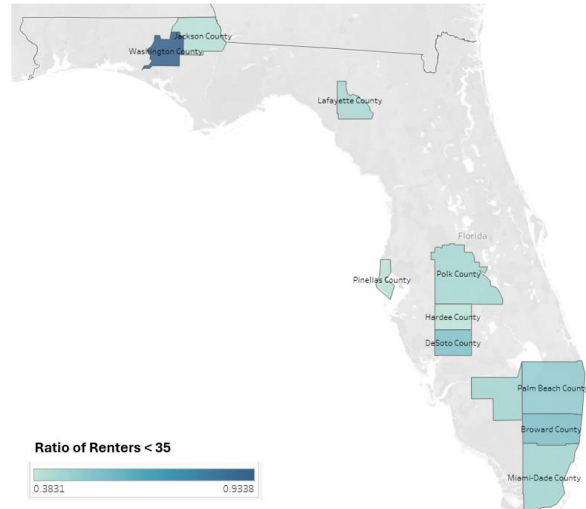

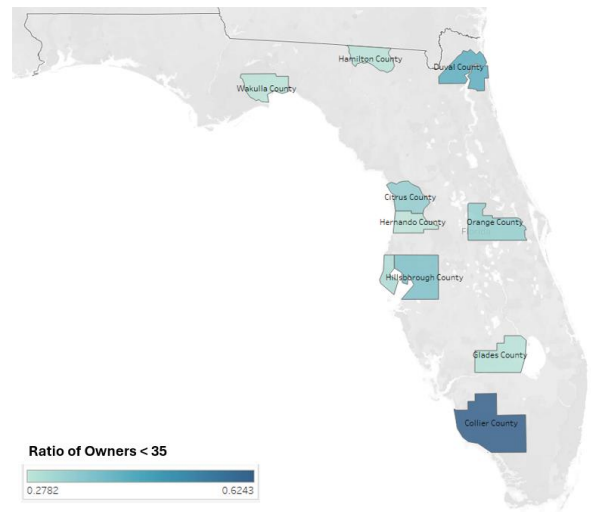**Figure 7 Top 10 Counties with Highest Renter Ratio for Age Group under 35**

Figure 8 Top 10 Counties with Highest Ratio of Owners for Age Group under 35

Renters vs. Owners Home Built Year

Next, we created visuals based on the year the home was built. We focus on newly built home as it adds to the supply of available homes and would be preferred to live in due to lesser maintenance costs. We created a ratio of renters living in homes built in 2019 and after and all renter homes, and similarly, we created a ratio of owners living in homes built in 2019 and after and all owner homes. Figures 9 and 10 show shading maps for Florida counties for the ratio of renters and owners, respectively, for homes built years after 2019. Regards to newly built homes with the highest renters ratio, Okaloosa, Leon, and Bradford counties in north Florida, along with

Charlotte and Hernando counties, were in the top 5. All of the top 5 counties were rural counties. Regarding newly built homes with the highest owners ratio, Palm Beach County, which is north of Miami, has a large number of condos. Followed by Wakulla County, which is a rural county that has been seeing a growing population and growth of newly built homes.
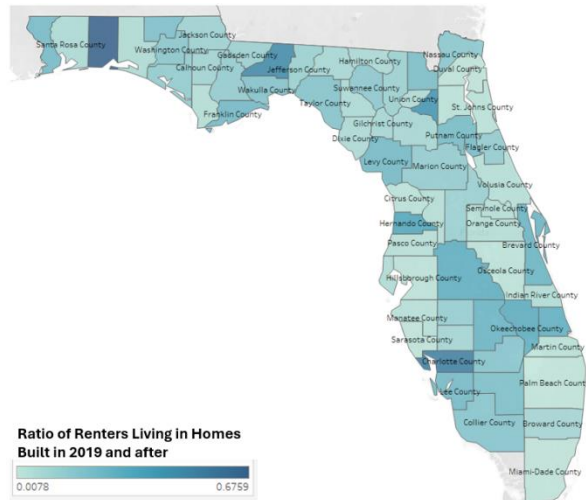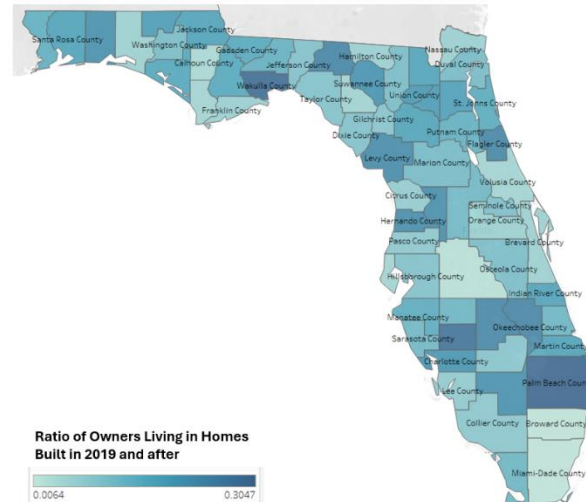

Figure 9 Renters Home Built in 2019 and after

Figure 10 Owners Home Built in 2019 and after

Income Constrained Household

Median household income for Florida as Census ACS has been around 55K for the past few years, including 2020 (FLHealthCharts, 2024). Thus, we focus our attention on household income in the range of 25K to 50K. Exploratory analysis of the number of renters with incomes in the lower income buckets may indicate a need for affordable housing programs or rent control policies. We created a calculated field of ratio of owners with household income 25K to 50K and renters with household income 25K to 50K. From Figure 11, the majority of Florida counties are in light teal, indicating that the majority of income-constrained families are renting. Calhoun and Pinellas counties are the top two counties that have higher rates of those who are income-constrained who are homeowners. In contrast, in other counties majority of income-constrained households are renters.
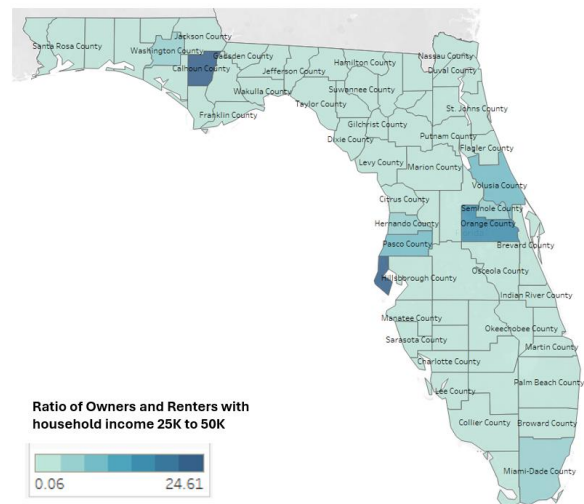

Figure 11 Ratio of Owners and Renters with Household Income 25 to 50K

Middle-Income Households

The closest ACS income grouping that can be considered as middle income, as per Pew Research Center (Kochhar, 2024), is households with income in the range of 50K to 100K. Exploratory analysis on the middle-income bracket can provide insights into the economic status of homeowners. Thus, we created a ratio of renters by owners with household income of 50K to 100K. We used owner household income as the denominator for this calculation as we anticipate more homeowners than renters at the county level for this income group. From Figure 12, it can be noted that Okeechobee and Hamilton counties have the highest rate of renter over to owners, followed by Dixie and Franklin counties. These top counties have larger rural areas while they have less number of households compared to metro areas, they have a higher relative percentage of renters compared to owners.
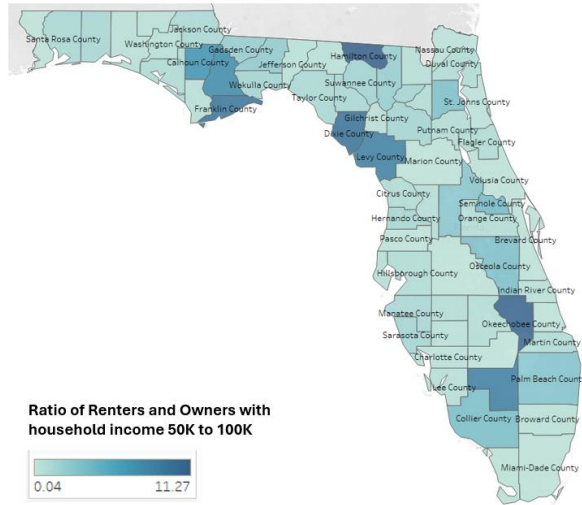
Figure 12 Ratio of Renters and Owners with Household Income 50 to 100K

Renter vs. Owner Vacant Homes and Severe Housing Problems

The higher amount of vacant rental homes and rental properties with severe housing problems can exacerbate the affordable housing issues (Alvarez & Steffen, 2021). In the areas where these problems are higher, rental assistance programs in the form of vouchers could help in reducing worst-case scenarios of affordable housing issues. Rental assistance programs reduce the rent burden to fall within the program limits, and landlords participating in these programs are crucial for addressing affordable housing issues. We gain insights and visualize these factors; we created a scatter plot of substandard rental housing that is vacant against rental houses that have severe plumbing or kitchen facilities problems and cost burden of 50% of income. We create a similar scatter plot for owner based housing. The scatter plots for renters and owner data can be seen in figures 13 and 14.

In Figure 13, we can note that counties that have larger populations and metropolitan cities can be separated from other counties that are grouped in the lower quadrant. Among the larger counties, we can note the problems of vacant rental homes that may have severe problems are quite higher rate in comparison to other counties. A similar trend can be noted for owners data in Figure 14, with the exception of Lee and Polk counties, which are in the southeast and central regions of Florida.
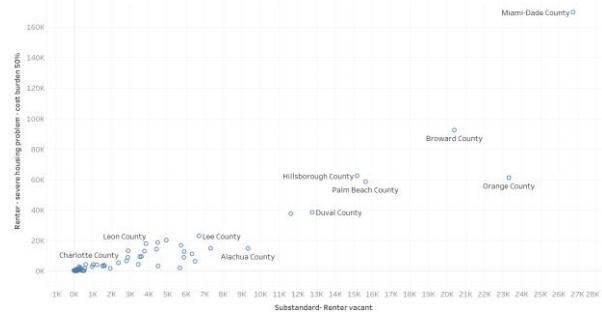

Figure 13 Scatter Plot of Substandard Vacant Rental Homes vs. Rental Homes with Severe Housing Problems
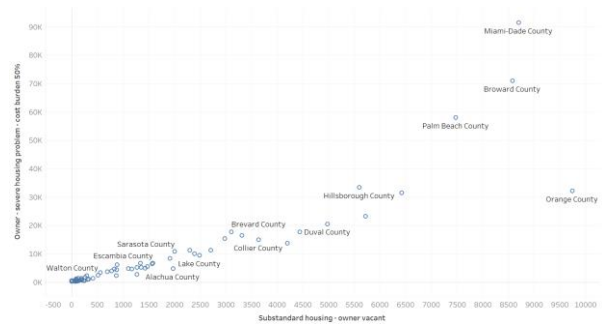

Figure 14 Scatter Plot of Substandard Vacant Owner Homes vs. Owner Living Homes with Severe Housing Problems

5. CONCLUSION

In this paper, we combine and present two related studies wherein the first study's findings encouraged us to proceed with the second study. The first study was a bibliographic review of existing affordable housing literature that investigated the issue using data science methods. We used Google Scholar as a search source. Despite research describing affordable housing increasing in the seven-year search period (2015-2022), articles identified matching our investigation criteria represent a small portion of research published in affordable housing journals. Attempted keyword analysis and comparison of data science-relevant items did not produce any major commonality. The lack of accumulation of coherent literature and scientific findings implies greater investment in data-driven affordable housing research is much warranted. Data science focused research studies could bring up meaningful interventions to effectively and efficiently address the societal challenge of affordable housing.

It is important to note that this study has limitations in the interpretation of the findings of

the research. All analyses on affordable housing literature were created mainly from journal articles. Future studies could consider including papers from reports, books, and conferences. The number of bibliometric records retrieved is limited due to the choice of keywords. Therefore, the keywords used for casting the wide net may not have produced complete coverage of all available relevant literature.

Based on the findings of the first study, we decided to engage in exploratory data analysis at the Florida county level to gain a higher-order level of understanding of the affordable housing issue. During a conversation with one of the nonprofit stakeholders working on the affordable housing issue, they mentioned the disparity between individuals who are homeowners and renters while both groups might face affordable housing issues. The stakeholders informed this as an anecdotal fact, and they have not seen anyone investigate and compare data about those living in affordable housing conditions against those who own a home and are facing affordable housing issues.

We gathered data from the ACS Census and CHAS data sources for the exploratory study. From the analysis, it is clear that the majority of the housing units in Florida are owner-occupied. However, there are significant variations in the ratio of owner-occupied to renter-occupied units in different counties. Some counties have a much higher proportion of renter-occupied units than others. We created several Tableau visualizations of the issue and presented key visuals that shed light on the renters vs. owners scenarios. Visuals indicate very few Florida counties have homeowners for low-income population groups, while rural counties have more renters than homeowners for middle-income population groups. We hope the visualization of the distribution of housing units and the income of renters and owners helps researchers and policymakers create more equitable and sustainable housing policies that benefit everyone in a given community.

This study has several limitations that should be considered when interpreting the findings. First, the data used in this study were collected in 2019 and, therefore, did not reflect the impact of recent events, such as the COVID-19 pandemic, on housing patterns and trends. Second, the data used in this study are self-reported and may be subject to reporting bias. Third, the sample size for some counties may need to be bigger to draw statistically significant conclusions. Finally, the study is limited to the United States and cannot be generalized to other countries or regions.

There are several avenues for future research in this area. One potential area for future research is to examine the impact of housing policy on the patterns and trends observed in this study. Another potential area for future research is to investigate the role of urban planning and design in shaping housing conditions, ownership, and occupancy.

6. REFERENCES

Adabre, M. A., & Chan, A. P. C. (2019). Critical success factors (CSFs) for sustainable affordable housing. *Building and Environment, 156*, 203-214. doi:https://doi.org/10.1016/j.buildenv.2019.04.030

Alvarez, T., & Steffen, B. L. (2021). Worst Case Housing Needs 2021 Report to Congress. Retrieved from https://www.huduser.gov/portal/sites/default/files/pdf/Worst-Case-Housing-Needs-2021.pdf

Anderson, L. M., St. Charles, J., Fullilove, M. T., Scrimshaw, S. C., Fielding, J. E., & Normand, J. (2003). Providing affordable family housing and reducing residential segregation by income: A systematic review. *American Journal of Preventive Medicine, 24*(3, Supplement), 47-67. doi:https://doi.org/10.1016/S0749-3797(02)00656-6

Czischke, D., & van Bortel, G. (2023). An exploration of concepts and polices on 'affordable housing' in England, Italy, Poland and The Netherlands. *Journal of Housing and the Built Environment, 38*(1), 283-303. doi:10.1007/s10901-018-9598-1

FLHealthCharts. (2024). Florida Median Household Income (Census ACS). Retrieved from https://www.flhealthcharts.gov/ChartsDashboards/rdPage.aspx?rdReport=NonVitalIndGrp.Dataviewer&cid=0293

Galster, G., & Lee, K. O. (2021). Housing affordability: a framing, synthesis of research and policy, and future directions. *International Journal of Urban Sciences, 25*(sup1), 7-58. doi:10.1080/12265934.2020.1713864

Golubchikov, O., & Badyina, A. (2012). *Sustainable Housing for Sustainable Cities: A Policy Framework for Developing Countries*. Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2194204

Haffner, M. E. A., & Hulse, K. (2021). A fresh look at contemporary perspectives on urban housing affordability. *International Journal of Urban Sciences, 25*(sup1), 59-79. doi:10.1080/12265934.2019.1687320

Harzing, A.-W. (2007). Publish or Perish. Retrieved from https://harzing.com/resources/publish-or-perish

HUD. (2023). Consolidated Planning/CHAS Data. Retrieved from https://www.huduser.gov/portal/datasets/cp.html

Kochhar, R. (2024). The State of the American Middle Class: Who is in it and key trends from 1970 to 2023. Retrieved from https://www.pewresearch.org/race-and-ethnicity/2024/05/31/the-state-of-the-american-middle-class/

McCabe, A., Pojani, D., & van Groenou, A. B. (2018). The application of renewable energy to social housing: A systematic review. *Energy Policy, 114*, 549-557. doi:https://doi.org/10.1016/j.enpol.2017.12.031

McCrone, G., & Stephens, M. (2017). *Housing Policy in Britain and Europe*. London: Routledge.