GOOD DATA FOR GOOD DECISIONS ABOUT RESIDENT RETENTION

Analysis on Resident Retention in the City of Lafayette, Indiana

Abstract

The northern neighborhoods of Lafayette, Indiana, are plagued by high resident turnover, but the city strives to build a community where residents will want to live for an extended period. Through revitalization of these neighborhoods, the well-being of the community and the livability of the area for residents could be increased. The purpose of this study was to identify geographical and quantitative trends associated with these transient residents. In partnership with the City of Lafayette, the study analyzed geographical and numerical data about local households to understand these trends and aid in the city's effort to increase resident retention in the northern neighborhoods by identifying areas to focus retention efforts on. The study focused on analyzing household data to identify locations with patterns of households at high risk of low retention. The method used to identify these locations was plotting each household on a map of the studied neighborhoods and filtering households by different factors determined to significantly affect retention to derive geographical trends. Analyzing these trends showed that young low-income home renters were more likely to live in their home for shorter amounts of time. Geographical visualizations were produced focusing on these home renters and the northern neighborhood's likely transient residents, and geographical areas of these potentially transient home renters were identified. With these specific areas identified, it was recommended that the community partner, the City of Lafayette, focus resources on efforts of retention and revitalization in these specific areas of the northern neighborhoods of Lafayette.

Keywords

community well-being, resident retention, quality of life, livability, community indicators, neighborhood revitalization, data analysis

https://doi.org/10.7771/2158-4052.1531

Student Author

Eli Coltin is a rising junior majoring in economics honors, computer science, and data science at Purdue University. His research interest is exploring the applications of quantitative analysis on data involving quality of life and outcomes of human decisions. Coltin has been working with Dr. Jason Ware on researching community well-being since May 2020.

Mentor

Jason Ware earned a PhD in curriculum studies from Purdue University. He is a qualitative research methodologist with a focus on narrative inquiry and a recent turn to participatory action research vis-à-vis community-indicator and well-being projects. Ware and his research group are exploring the extent to which working with particular urban populations to establish and measure quality-of-life indicators at the neighborhood level can transform material realities and serve as an educative tool.

INTRODUCTION

Research on human well-being begins with a difficult question: How does one define well-being, and how can well-being be measured? While well-being may have no true definition, this study follows Veenhoven's specific definition of well-being as "quality of life-as-a-whole" and was used to "evaluate life-aspects such as dwelling conditions or employment chances" (2000, p. 1). This study focused on the life aspect of livability in relation to well-being and also understood quality of life as Veenhoven's definition of "quality of society" in some instances and "happiness of its citizens" in other contexts (2000, p. 1). For this study, well-being was applied at the community level with the purpose of understanding what can be done to improve community well-being. To measure community well-being, community indicators, which Phillips defines as "measurements that provide information about past and current trends," can be identified for a specific community and leveraged as a general overview of a community's trends on important issues (2003, p. 3).

This research applied these ideas of well-being and quality of life to the City of Lafayette, Indiana, as it looked at community well-being in the north-end neighborhoods of Lafayette with a focus on the specific community indicator of resident retention. Similar studies in Lafayette used a qualitative approach to this issue. Examples of these qualitative approaches are identifying resources and services in specific neighborhoods and visually assessing the livability of neighborhoods through evaluating infrastructure elements such as sidewalks and transportation routes (Lafayette City Council, 2016). While these studies may use quantitative data for background information and motivation for taking action, numerical data is not being used as a core piece of the analysis.

This research analyzed solely quantitative data on households in the north-end neighborhoods of Lafayette, such as household income, home value, and years lived in a home, to further understand well-being in Lafayette. Trends derived from this quantitative analysis were applied to the creation of household demographics maps. Qualitative analysis on these maps was performed to identify areas at high risk of low resident retention.

In collaboration with the City of Lafayette, this research was undertaken to help inform decisions regarding neighborhood revitalization and affordable housing in Lafayette. With less than 1% of relevant data on the topics analyzed by the City of Lafayette, the purpose of this research was to supply data-driven insights and results to enhance community well-being improvement efforts in Lafayette. This research focused on understanding the interaction between demographic variables and their impact on well-being and worked to answer the question of how the geographic distribution of demographic groups affected well-being.

METHODOLOGY

Overall Design

Past community well-being research in Lafayette utilized interactions with residents as a method of gathering data. As this research was done during the COVID-19 pandemic, the option of resident interaction was limited. A pivot was therefore necessary to continue research on community well-being, and the new format of research that emerged out of this change concentrated on using data to understand well-being. While other studies use data as a motivation for action, this study used data analysis on secondary data as the core investigation technique. This method allowed for the analysis of many different variables at one time to determine variable interactions and effects.

This project analyzed household data from the U.S. Census Bureau and an online information directory, the Polk City Directory. Secondary data from these databases on the topics of aggregate statistics at the city and census tract levels was used to understand the community. Individual data about each household, such as estimated home value, was used to understand the livability and dwelling conditions for residents. These resources as well as communication with the City of Lafayette produced information necessary to determine ideal metrics to measure well-being in Lafayette. Information at different levels allowed for a more thorough understanding of issues related to livability by providing numerous data sources addressing different factors of well-being. Data on home value provided insight for research on affordable housing, and data about neighborhood median resident ages as well as each household's number of years living in their home provided information for analysis on transient residents. The target neighborhoods in this study were six north-end neighborhoods in Lafayette: Hanna, Historic Jefferson, Lincoln, Monon, St. Lawrence McAllister, and Vinton. This study selected these neighborhoods based on descriptions from City of Lafayette officials about disparity of high resident turnover relative to other neighborhoods in the north end of Lafayette.

Data from 3,093 households located in the six specified neighborhoods available in these databases was first analyzed for significant relationships to this project's chosen community indicator, resident retention. Resident retention was selected as the key indicator of livability and well-being due to these neighborhoods' disparity in turnover relative to other neighborhoods. Different household variables, such as age of residents and estimated household income, were analyzed at a high level for their correlation and connection to resident retention. This research also studied past research on the north-end neighborhoods to understand factors that were found to influence livability and retention in these areas. In addition, analysis on numerous variables with relation to resident retention, such as home ownership status, was performed to understand significant relations to resident retention. This analysis was done through comparing retention when different factors were controlled and filtered for. For example, the amount of years lived in a home for households was compared between income groups, with age held constant to analyze the effects of income on resident retention without the factor of age. The next step pursued further research to better understand the trends and their impact on retention.

The study mapped households using the significant variables as demographics after derivation of trends about the variables. For example, to look at the variable of age, the study plotted households on a map, with each age group given a different color. Then, visual analysis could be used to determine trends related to age groups by looking at the distribution of ages. By identifying the variables that had significant relationships to resident retention, the study isolated different areas exemplifying trends potentially indicating low retention to create a visual representation of locations of households at high risk of low retention.

This research created data-related facts and figures as well as maps of the north-end neighborhoods of Lafayette, highlighting areas at high risk of low retention, and presented them as the results of these methods. These formats were chosen for their ease of understanding for an audience, as the purpose of this research was to provide results that could be used to inform the City of Lafayette's decisions on neighborhood revitalization and affordable housing, so formatting results as easily digestible information was a key component of the project.

Variable and Indicator Choices

When using well-being indicators, there is no perfect indicator to represent a whole community. The indicators must be chosen to be representative of the community and provide insights that could benefit the community (Cuthill, 2002). In the north-end neighborhoods of Lafayette, the median household age is approximately 40 years old. In the broader scope of the City of Lafayette, the median age drops to 33 years old. This higher median age, which appears to correlate and influence the transience of residents, displays the issue of younger residents choosing not to remain in the area, which shifts the median age higher. Transient residents is a common issue plaguing these neighborhoods (Lafayette City Council, 2016). This study researched specific indicators to isolate certain trends used to better understand these indicators' effects on resident retention.

A lack of younger residents represents multiple issues, such as signaling that the area could have a limited quality of life for younger residents. In the Greater Lafayette area that includes the city of West Lafayette's Purdue University, which hosts over 45,000 enrolled students, a previous research study showed that over half of surveyed residents under the age of 40 plan to leave the region within the next four years (Next Generation Consulting, 2012). This trend, which is consistent with previous research on the area, could indicate that the Lafayette area is not an area where younger residents have as high a quality of life as their elder counterparts. The wants and needs of younger residents may not be readily attainable, leading to those residents viewing the environment of the north-end neighborhoods as less livable (Veenhoven, 2002).

This poses a second issue of limiting the building of community. Community is important for improving well-being, as community can lead to a shared identity, civic participation, and lower crime (Mahmoudi Farahani, 2016). Past research by Cuthill (2003) indicates that citizen participation and building relationships in a community are important for not only well-being but also keeping residents involved in a community. So, measuring resident retention as a community indicator reveals information about the well-being of a community at the base layer of building a livable community, starting with keeping its human capital.

Quantitative Analysis

The Polk City Directory hosted aggregated data from 3,093 households in the north-end neighborhoods of Lafayette. The database contained over 50 variables about each household, such as household income and home price. The specific indicator being investigated was resident retention, so the variable of years in home was used as a proxy for resident retention. This study analyzed each variable in the database based on its relationship or correlation to resident retention to determine which variables would be further analyzed for their connection to well-being. This wide-scale quantitative analysis allowed for consideration of many different possibilities of relationships between variables and resident retention. Data analysis and processing for this work was executed using the programming languages Python and R with CSV files.

Choosing years in home as the proxy variable required keeping in mind multiple issues. First, when analyzing retention, years in home gave more weight to retention of older residents, as they had more of a lifetime to live in their home. Second, this metric only counted the years a resident had lived in a certain home. If there is a transient renter, such as a home or apartment renter who leaves her or his home to rent a home in another part of the north-end neighborhoods, the counter for years in home will restart at zero. So, with the data available, it was difficult to determine if transient renters' previous locations were in Lafayette or in a different location.

Another important gap to note is that this metric provided no information on why persons left their homes. In this study's approach of data-driven insights, causes of

transiency were not able to be understood other than the information available through data. Other studies with a focus on resident interaction could continue this study's line of inquiry about causes for transiency through methods involving anecdotal evidence by following the ideas presented in this study's results.

The next step was to determine the major trends associated with the proxy variable of years in home. This was done with correlations and visualizations to determine any trends related to years in home. After determining multiple major variables that strongly correlated with years in home, further investigation into the data about the relations was done by holding multiple variables constant using a multiple linear regression, with years in home as the dependent variable. This deeper research allowed for the determination of multiple key variables that appeared to vary with years in home.

Preprocessing the data required adjustment to some variables that adjusted the data's precision. For certain variables that would be expected to be strictly quantitative data, such as household income and home value, the Polk City Directory provided a categorical variable of a bucket of values, such as $50,000 to $59,999. In scenarios such as this, the average value was taken. So, for this situation the bucket would be coded to the quantitative value of $55,000 for ease of calculation with the least loss of precision.

Qualitative Analysis

This research also utilized qualitative analysis through mapping to derive visual trends and relationships to years in home in the north-end neighborhoods of Lafayette. The qualitative analysis involved the creation of maps of the 3,093 households studied. Each household was placed on a map to visually analyze the distribution of demographics on a household level. For example, Figure 1 shows the distribution of home renters and owners in Lafayette. Blue marks indicating households owned by the residents are overlaid over households

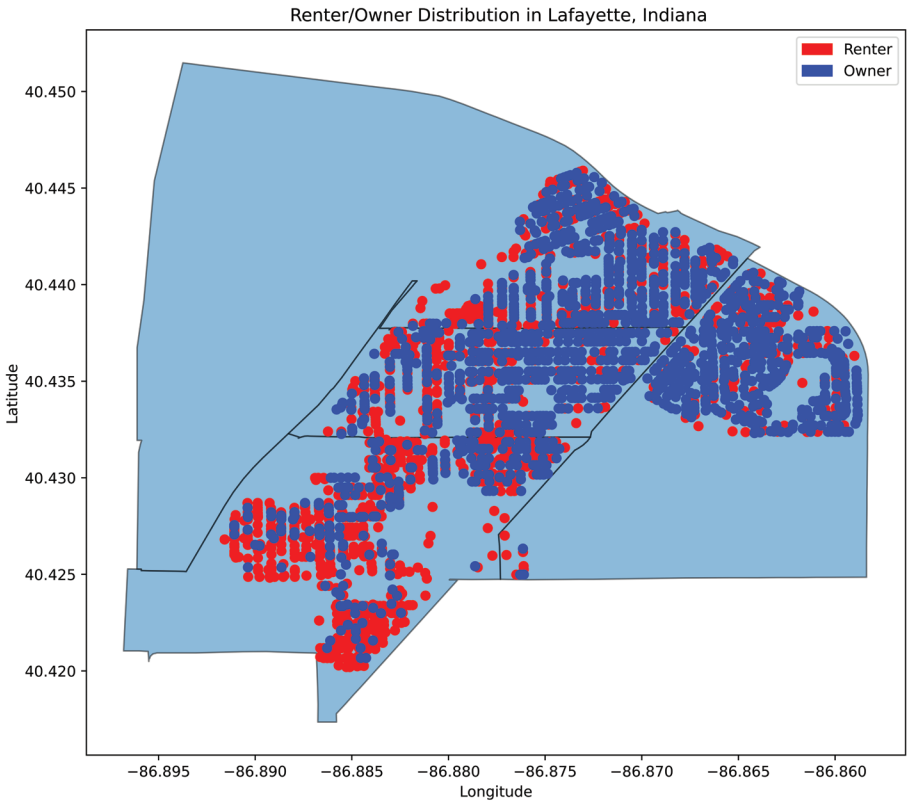

FIGURE 1. Map of distribution of household ownership demographics used for qualitative analysis.

rented by the residents. Visually, the trend in the southern neighborhoods, specifically the Lincoln and Historical Jefferson neighborhoods, of an elevated number of renters can be seen. This could indicate an issue with affordable housing and resident transiency in these regions. This qualitative analysis allowed for an intuitive way to determine geographic trends with relation to resident retention and livability.

RESULTS

Quantitative Results

The first step in the quantitative analysis of this research was to determine the high-level trends through correlations. Table 1 shows these correlations with years in home as the dependent variable. The independent variables, in order, are the average age of the residents, whether there are children present in the household, the estimated home value, the estimated income of the household, whether the household residents are married, and if a resident owns the home.

Unsurprisingly, age had the strongest correlation with years lived in a home. This was likely due to older residents having had more time to live in their home. The second factor, children in home, was a binary variable that appeared to have a decent correlation. Home value had minimal correlation with years in a home. Although this would most likely make home value not a significant variable and not necessary to include in the correlations, it would be expected for home value to have a stronger correlation in the positive direction, so its inclusion is important for showing its lack of importance. On the other hand, household income exhibited a stronger positive correlation with what was expected. The final two variables, married households and ownership, both were binary variable exhibiting positive correlations with years in home, meaning that in married households and in households where the resident is an owner residents to live in their home for a longer period of time.

With these basic correlations established, it was important to dig deeper into the data because other variables, such as age, could be an underlying influence on certain factors. To do this, a multiple linear regression model was used with years in home as the dependent variable and with age, children, income, marital status, and ownership as independent variables. The results are shown in Table 2. One difference from the rest of the study is that only 2,406 households were used for this model instead of 3,093, as some households did not have data for certain variables and so were omitted from the regression.

From the results of this regression, all variables in the model were significant at the .01 level except for the income variable. The income variable was an interesting case, as there was a large variability in low-income

TABLE 1. Correlation coefficient (r) of years in home against multiple independent variables.

Correlations (r)	Age	Children	Home Value	Income	Married	Ownership
Years in Home	0.6225	0.2104	–0.0857	0.2872	0.3842	0.5008

TABLE 2. Multiple linear regression model output with years in home as the dependent variable.

Variable	Estimate	Std. Error	t-value	Pr(>|t|)
(Intercept)	–11.5422	0.6260	–18.44	0.000 ***
Age	0.4026	0.0120	33.61	0.000 ***
Children	1.7007	0.5071	3.35	0.001 *
Income	4.46E-09	0.00002	0.00	1.000
Married	2.7941	0.5502	5.08	0.000 ***
Ownership	5.3198	0.5219	10.19	0.000 ***
Observations: 2,406		**Significance codes:** 0 '***' 0.01 '*'		
R-squared value: 0.4836		**Adjusted R-squared value:** 0.4825		

residents' years lived in home. The case of income and each analyzed variable will be covered later in more depth.

Age

As expected, age represented a large role in years lived in home. Figure 2 shows this disparity. While this may be indicative of the larger trend of younger residents leaving while older residents stay, it is difficult to definitively determine using the available data. That trend could be represented in the large increase in average years in home for those ages 60+ relative to the trend before, but further research would be necessary in this area.

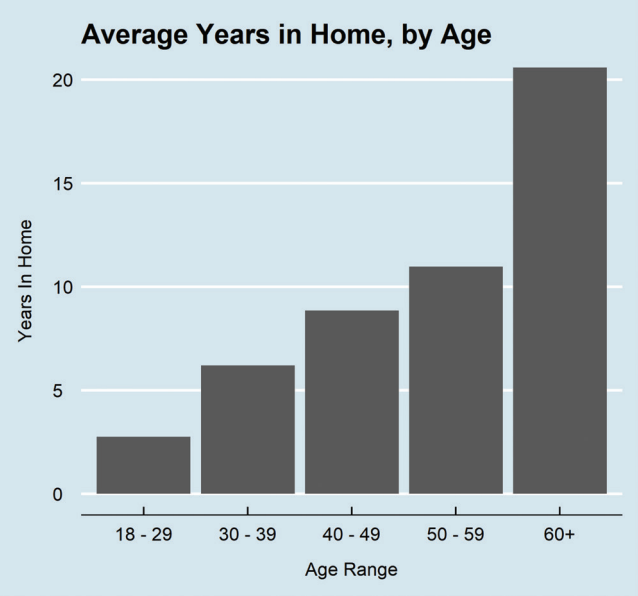

FIGURE 2. Trend of rising average years in home as age increases.

Children

With children, the main significance was the difference between having no children and having any children. Even with age held constant, the trend of households with children living longer in their residences on average than those without children held. Figure 3 shows this difference, and there was no significant difference in years in home based on the number of children in the household when the amount was greater than zero. On average, households with children lived in their homes for 6 more years on average. Households with children lived in their home for 15.96 years on average versus 9.94 years for those without children.

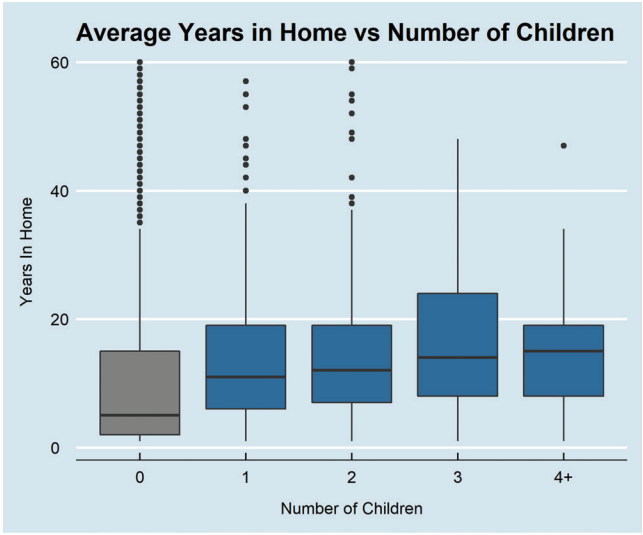

FIGURE 3. Disparity in years in home for households with children compared to households without children.

Income

Income was not seen to be significant in the linear regression, as shown in Figure 4. This was most likely due to the large amount of outliers in the lowest-income group that are due to older-aged households with lower income. The largest disparity was between the lowest and highest income groups. Households with an income of greater than $50,000 had lived in their homes for 22.79 years on average, which is about 13 more years than the 9.45 year average for households with an income under $20,000. This disparity is significant at the .001 significance level when using a student's t-test for difference in means with a t-statistic of 8.31 and a p value of 1.56e-16.

Married

Across all age groups, there was a large disparity between married households and single households, as shown in Figure 5. On average, married households had lived in their homes for 11 more years than single households, with a 19.95 year average versus the single household 8.56 year average.

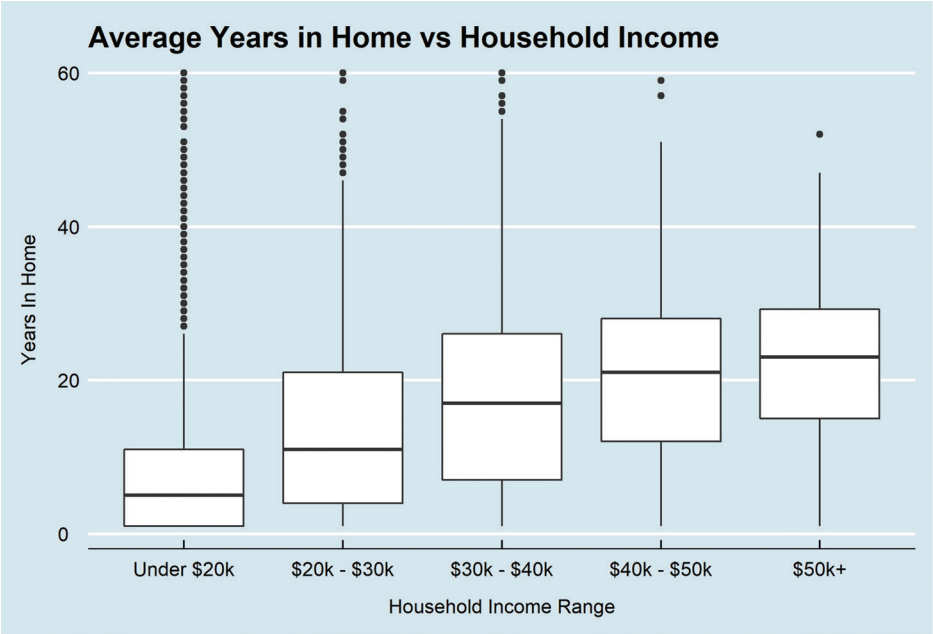

FIGURE 4. Upward trend in years in home as household income increases.

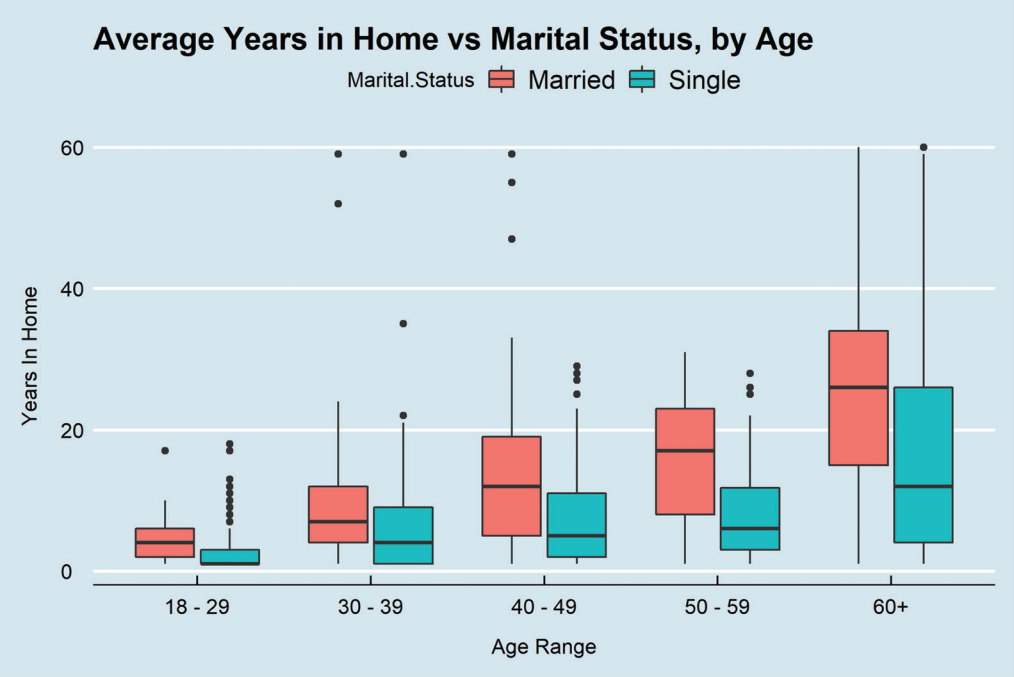

FIGURE 5. Trend of married households living in their homes for longer across all age groups.

Ownership

Ownership held the second-highest correlation with years in home behind age during the initial look at correlations. Figure 6 shows the large disparity between owners and renters. While it was difficult to determine if renters had this status because they are in transient housing and may leave their residence for another home in the north-end neighborhoods, there was still a significant difference between the two groups. Homeowners have lived in their homes for 17.72 years on average, which is about 13 more years than the 4.88-year

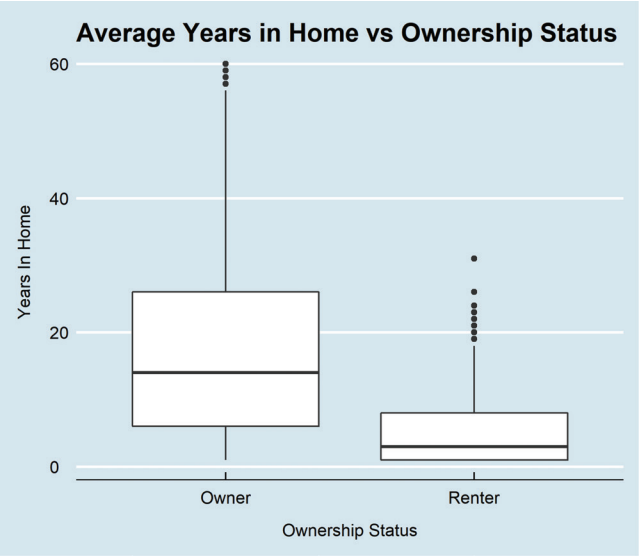

FIGURE 6. Large disparity between owners and renters for years lived in their home.

average for renters. A similar trend to this holds across all age groups.

Qualitative Results

With the understanding of age, children, married households, and ownership on resident retention, maps of the north-end neighborhoods of Lafayette were created to visualize areas at high risk of low resident retention. In the map in Figure 7, households that were unmarried, had no children, and were renters had average owner ages of less than 30 and an estimated household income of less than $20,000, highlighted in red. All other households are highlighted in gray. Clearly, the most southern neighborhoods, which are the Lincoln and Historic Jefferson neighborhoods, visually appear to be at high risk of low retention. These findings align with

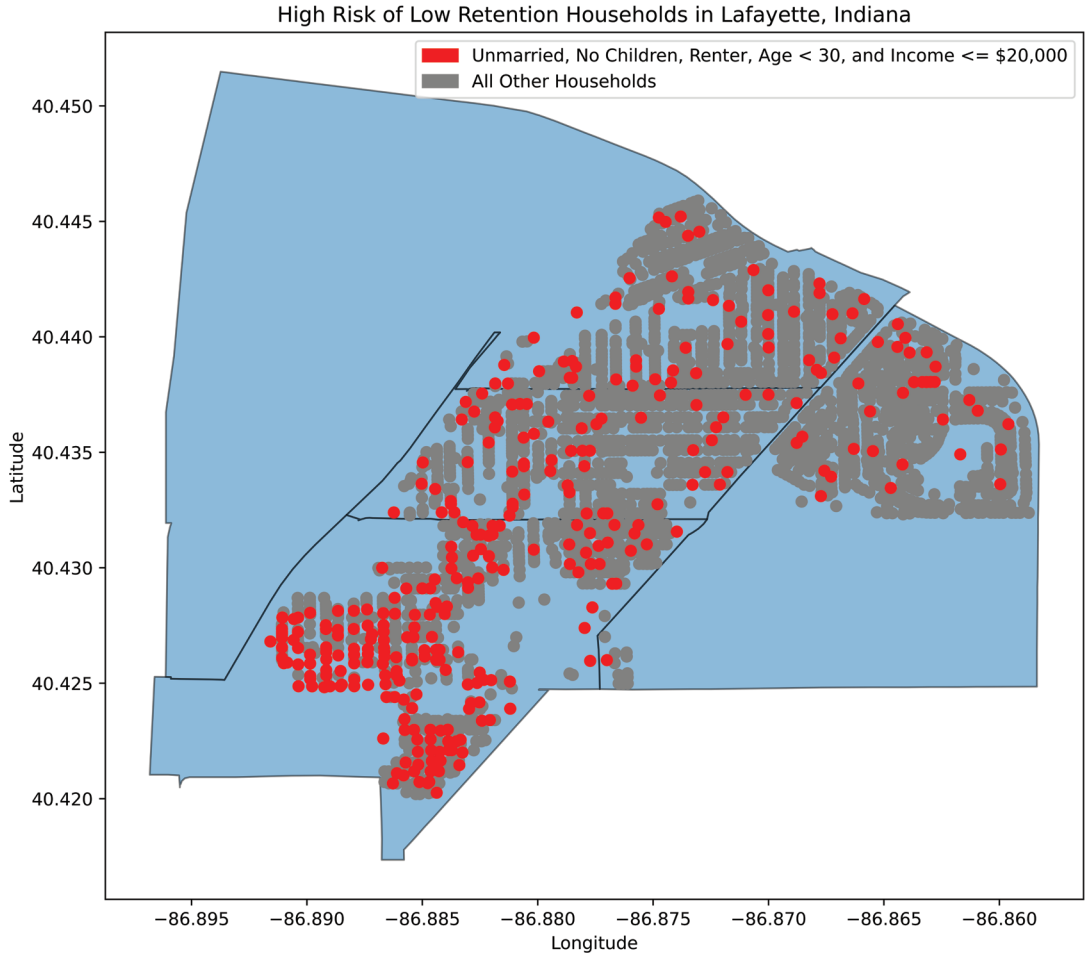

FIGURE 7. Map of households with a demographic at high risk of low retention.

past research in the area (Lafayette City Council, 2016). Similarly, patches in other neighborhoods also appear to exhibit this risk. Some streets have numerous red households in large patches of gray. Further investigation into these areas would be useful to determine why they fit the demographics of high risk of low retention while the surrounding households do not.

DISCUSSION

One of the first steps in the processes of neighborhood revitalization and increasing resident well-being and community livability is community building, and human capital is one of the most important parts of the fabric of a community (Cuthill, 2002). With the City of Lafayette's focus on the ideas of neighborhood revitalization, affordable housing, and livability in the north-end neighborhoods, an important place to begin was with resident retention due to the high ages of residents in those neighborhoods compared to other areas of the city. Retaining younger and transient residents would allow for growth in human capital.

By taking a data-driven approach while most similar studies take a qualitative approach, this research was able to validate the findings of qualitative studies through hard data. At the same time, the experiences of those in the community of Lafayette validated the results of the quantitative research. The conclusions about the significance of resident age, presence of children, income, married residents, and homeowners are all key takeaways that could be used to assist the City of Lafayette's efforts in revitalization and the elevation of well-being and livability. Similarly, the qualitative approach of mapping allows the city to see areas where steps could be taken to increase resident retention or have a more educated approach to the issues.

Future research in this area could be done to further investigate the causes of the significant indicators of low retention found through data analysis. Qualitative studies could be utilized as a form of cross-validation to better understand resident experiences and how they align with results from the quantitative analysis. While data cannot replace the reality of human experiences, its utilization in well-being studies can be a useful tool for supplementing and validating knowledge gained from residents. But with understandable results and visualizations, data provides a powerful evidence-based approach to understanding well-being, livability, community indicators, and issues in a community.

ACKNOWLEDGMENTS

I would like to thank Dr. Jason Ware for his support and guidance throughout my research process on well-being. I would also like to thank the Purdue Office of Undergraduate Research and the Purdue Honors College for their help in obtaining necessary resources as well as the City of Lafayette for providing the background information necessary to complete my research.

REFERENCES

Cuthill, M. (2002). Coolangatta: A portrait of community well-being. *Urban Policy and Research, 20*(2), 187–203. doi:10.1080/08111140220144489

Cuthill, M. (2003). The contribution of human and social capital to building community well-being: A research agenda relating to citizen participation in local governance in Australia. *Urban Policy and Research, 21*(4), 373–391. doi:10.1080/0811114032000147412

Lafayette City Council. (2016, October 4). Lincoln neighborhood livability study." Paper presented at a regular meeting of the Lafayette City Council, Lafayette, IN.

Mahmoudi Farahani, L. (2016). The value of the sense of community and neighbouring. *Housing, Theory and Society, 33*(3), 357–376. doi:10.1080/14036096.2016.1155480

Next Generation Consulting. (2012). From good to great. https://www.lafayette.in.gov/DocumentCenter/View/274

Phillips, R. (2003). *Community indicators.* Chicago: American Planning Association.

Veenhoven, R. (2000). The four qualities of life." *Journal of Happiness Studies, 1*(1), 1–39. doi:10.1023/A:1010072010360

Veenhoven, R. (2002). Why social policy needs subjective indicators. *Social Indicators Research, 58,* 33–46. doi:10.1023/A:1015723614574