INDICATORS OF GENTRIFICATION

Isha Goswami[1], Kalash Bhaiya[2], Nanxi Jiang[3], Raven (Ruiwen) Tang[4]

CONTENTS

ABSTRACT

Gentrification is a pervasive socioeconomic shift that interacts with various factors. This study analyzes gentrification trends across metropolitan areas in the United States from 2013 to 2023, focusing on income shifts, housing affordability, and demographic changes. Using the Ellen O'Regan model, we classify zip codes that have undergone gentrification based on relative income growth within their metropolitan areas. We integrate data from Zillow, the American Community Survey, and the U.S. Census Bureau to identify key indicators of gentrification and assess their impact on housing markets and racial demographics. Our findings highlight the accelerating influence of socioeconomic and environmental factors on urban transformation, particularly in cities like Miami, where climate change has played a role in displacement. By mapping economic disparities and housing trends, we contribute to a deeper understanding of gentrification's long-term effects on communities and urban policy.

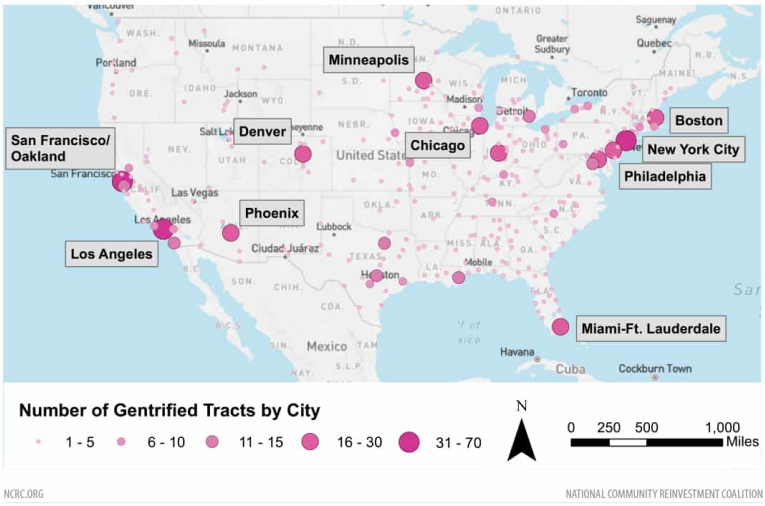
[1] *California Institute of Technology*
[2] *University of Toronto*
[3] *Stanford University*
[4] *University of Pennsylvania*

1 INTRODUCTION

The problem statement for the 2025 Citadel Securities Women's Datathon prompts an analysis of housing market trends and affordability, as homeownership in the United States, a classic milestone of the fabled "American Dream," has grown increasingly difficult to attain in recent years. Various factors, ranging from interest rates to employment trends to government policies, interact in complex ways to influence and be influenced by the housing market.

This report is thus motivated to analyze a prevailing, multifaceted force in the history of housing in America: gentrification. Driven by urbanization and industrialization, accelerated by climate change, and acting reciprocally with influences of immigration and job market shifts, gentrification has acted as a key player to alter the housing, social, and economic fabric of the country. Gentrification is defined by both the quick development and the displacement of lower-income residents of urban areas as wealthier people move into these areas.

These effects have already hit close to our datathon setting, right here in Miami. Rising sea levels from climate change have created a gentrification movement in neighborhoods in Miami, historically inhabited by communities of color, that are at higher sea levels. Wealthy homeowners, who have previously retained luxurious beachfront properties, are seeking refuge from the rising sea levels in places like Overtown, Allapattah, Liberty City, and Little Haiti, which sit on Miami's Rock Ridge and are 3 feet above the city's average sea level [1]. House prices, property taxes, and rent in these neighborhoods are skyrocketing, and current residents are being driven out of their homes.



A highlight of geographic areas most affected by gentrification, Miami among them, from https://ncrc.org/gentrification20/

Gentrification is a significant urban phenomenon that reflects a confluence of diverse social, political, and economic factors, yielding simultaneous positive and negative impacts. Our analysis utilizes the provided Zillow zip code-level dataset, American Community Survey datasets, integrated with data from the US Census Bureau for income, to identify indicators that flag both the influences and consequences of gentrification, particularly in the areas of housing and racial demographics. This is specifically relevant and critical to examine, because gentrification increasingly shapes long-term impact, altering the social and economic situation of millions more Americans as the wealth inequality gap widens [2], anthropogenic climate change events escalate [3], and the nation undergoes cultural, economic, and demographic transitions. Our research question asks: **In the most populated metropolitan areas of the United States, are income, housing, and racial demographic changes reflective of gentrification?**

2 TECHNICAL EXPOSITION

To analyze the interaction of house prices with the prevalence of gentrification, we narrowed our analysis to zip code-level in the United States, within the decade 2013 to 2023 (focusing on changes from January 2013 to January 2023 if multiple months' data was available). We selected zip code-level to best capture changes at a neighborhood scope, as counties, states, and national shifts would absorb too much inter-neighborhood movement. We also focused on the most recent, complete data available, which left us with the most recent decade up to 2023 (2024 being too close to the present day). Our approach can be outlined by several stages:

1. Selecting a classification model for gentrification and using the model to categorize zip codes that underwent gentrification between 2013 and 2023
2. Calculating the change in median house value (both direct numerical difference and percentage difference) at the zip code level, and conversely analyzing gentrified, non-gentrified, and overall neighborhoods
3. Analyzing outliers and special zip codes of interest, flagged by our plots

2.1 Gentrification Model, and Cleaning of & Application to Data

In housing research, there isn't a standard, quantified definition of gentrification. We considered 3 strong contenders for gentrification models. The Freeman model categorizes a neighborhood as undergoing gentrification if the median household income and share of housing built in the past 20 years are less than the average metropolitan area values, the proportion of residents with college educations is greater than the average value in the metropolitan area, and housing prices have increased. The Ellen & O'Regan model categorizes a neighborhood as gentrified if its household income is initially less than 70 percent of the metropolitan area, and this percentage ratio of neighborhood to metro area has increased by at least 10 percentage points 10 years later [4]. Thirdly, the McKinnish, et al model categorizes a neighborhood as gentrified if its average family income is in the bottom 20 percent of all metropolitan neighborhoods in the United States and the neighborhood's average family income has increased by at least $10,000 by the end of a decade.

We ultimately chose and applied the Ellen and O'Regan model to identify gentrified neighborhoods, because it captures a reasonable timeframe of 10 years (as opposed to the Freeman model selecting a broad range of 2 decades), because it's generally independent of inflation (unlike the McKinnish, et al model utilizing a strict threshold of $10,000 in income difference), and because it enables relative analysis (each zip code is compared only to others in the same metropolitan area, and it qualifies relative to its own percentage point increase).

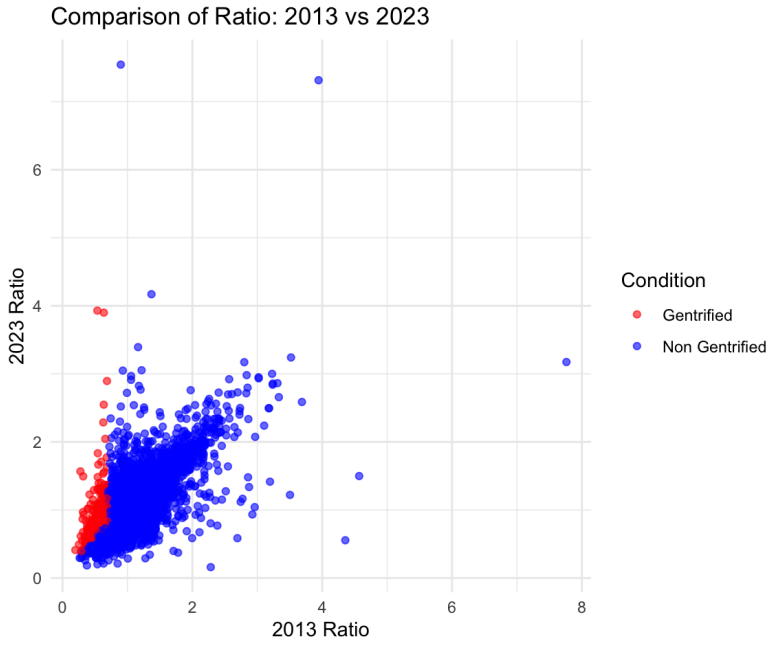
From the U.S. Census Bureau dataset, we used data on the median incomes in 2013 and 2023 per zip code [5]. Using the Zillow data set provided to us, we mapped each zip code to its membership metropolitan area. This required several steps of cleaning: dropping of null data rows, elimination of duplicate zip codes, discarding of rows with invalid metropolitan area values (in particular, several rows featured a city name but no state abbreviation, raising the possibility of two metropolitan areas of the same name in different states). Additionally, we specifically drew data from the month of January for both 2013 and 2023 to maintain consistency. We then took the weighted sum of all the zip codes within each metropolitan area to find the average household income of the metropolitan area, as shown in the equation below:

$$G_z = \frac{I_z}{\sum_{i \in \text{metro}} p_i \cdot I_i}$$

where G_z is the gentrification ratio for zip code z as per the Ellen and O'Regan model, I_z is the median income of the zip code z, and p_i is the percentage of

households in zip code i, where i belongs to the same Metro as z. Census Data for 2023 and 2013 (zip code population and mean income) was used to calculate G_z. In this way, the denominator represents the mean income of the entire metropolitan area, and the numerator is the individual income value of zip code z.

We then plotted each zip code on a 2-dimensional plot. Each zip code's x-axis value represents the zip code's income ratio to its metropolitan area in 2013, and each zip code's y-axis value represents the zip code's income ratio to its metropolitan area in 2023. Therefore, in order for a zip code to qualify as "gentrified" under the Ellen and O'Regan model, its x value must be less than 0.7, and its y value must be at least 0.1 greater than its x value. All points to the left of $x = 0.7$ and all points above the line $y = x + 0.1$ are thus categorized as "gentrified." In our plot, this corresponds to a red flag (also denoted by 1), while non-gentrified zip codes correspond to blue flags (also denoted by 0).



Each point represents a single zip code z, with the x-value representing G_z in 2013 and the y-value representing G_z in 2023. Points highlighted in red satisfy the Ellen and O'Regan model of gentrification

2.2 Analysis with Median House Price Change

Upon determining which zip codes are classified as gentrified and not gentrified, we compared this classification with the median percentage change in house price to find that across all zip codes that were in both the U.S. Census Bureau data set and the Zillow data set (14,335 data points), the average percentage change in median house price is 111%. On the other hand, the average percentage change in median house price for gentrified zip codes (431 data points found by the requirements of the Ellen and O'Regan model) is 151%. Histograms for the change in median house price for each of these subsets is shown below:

The average change in median house price percentage for gentrified zip codes is higher than the average for the whole data set, indicating that zip codes where wealthier people moved in and lower-income people moved out did experience significant increases in house prices. This is further supported by the histogram for the gentrified zip codes, which shows a right-skewed distribution, suggesting that many areas experienced extreme growth in housing prices. The histogram for the gentrified zip codes is also shifted to the right, as a result of the increased mean. To determine if the difference in means is statistically significant, we first conducted

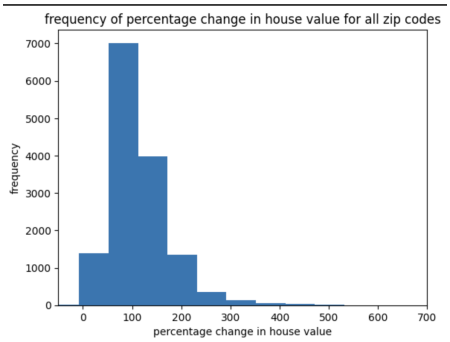


Figure 1: A histogram for the percentage change in median house price for all 14,335 zip codes we had access to.

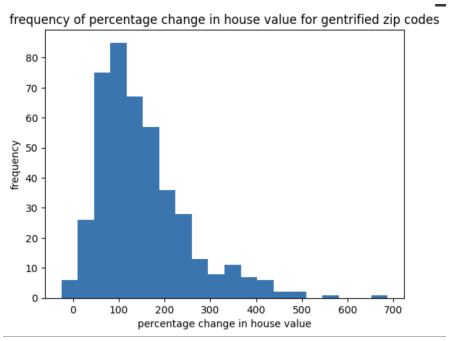


Figure 2: A histogram for the percentage change in median house price for only the gentrified zip codes we determined.

a Shapiro-Wilk test to see if the distribution of percentage change in median home value for all of the zip codes in the data set is normal, resulting in a p-value of 7.57×10^{-74}, indicating that the distribution is not normal. So, we conducted a Mann-Whitney U test, treating the distributions as two separate samples to obtain a p-value of 7.41×10^{-17}. This small p-value confirms the hypothesis that the difference in means of median house price percentage change from 2013-2023 is significant between the two distributions of all zip codes and gentrified zip codes.

2.3 Analysis of Outliers & Other Special Zip Codes of Interest

With the comparison of the G_z ratios between 2013 and 2023, certain postal codes stand out as outliers despite not being gentrified. These include Kylertown, PA (16847), Gibson Island, MD (21056), and Miami Beach, FL (33109), where we are now!). These outliers appear as in either 2013 or 2023, populations in the zip codes earned significantly higher income relative to the other year. For Kylertown and Miami Beach, the relative mean income increased significantly from 2013 to 2023, while it decreased significantly for Gibson Island. This is not too surprising or concerning, as it is simply the fluctuation of population income in places that were reasonably wealthy in both years.

3 FINAL CONCLUSION

3.1 Potential Extensions

The time frame of our analysis raises particularly interesting questions. For example, within 10 years, perhaps there are neighborhoods that have undergone and then undone gentrification (similar to Detroit during an earlier period of time), meaning that such reversal stories won't be captured by the checkpoints selected by our analysis, since we make an observation once at 2013 and one more time at 2023. This question is especially compelling within the context of the COVID-19 pandemic, which made a noticeable difference in migration to and from rural areas in the United States. Specifically, rural residents are less likely to leave their communities, and the numbers of incoming rural residents have also increased [6]. Though our analysis focuses on metropolitan rather than rural areas, it's incredibly likely that gentrification has combined with pandemic effects to influence community make-ups and migration rates in cities.

REFERENCES

[1] G. Iacurci. Climate change is gentrifying neighborhoods. 2024.

[2] R. Kochhar J. Horowitz, R. Igielnik. Trends in income and wealth inequality. 2020.

[3] United States Environmental Protection Agency. Climate change indicators: Weather and climate. 2025.

[4] Ingrid Gould Ellen and Katherine M. O'Regan. How low income neighborhoods change: Entry, exit, and enhancement. *Regional Science and Urban Economics*, 41(2):89–97, 2011.

[5] U.S. Census Bureau. Income by zip code tabulation area, 2022. Accessed: July 2024.

[6] J. Petersen R. Winkler. Fewer people are moving out of rural counties since covid-19. 2024.