

Introduction

Global economic activity is increasingly concentrated in densely-populated urban areas ([citation](#)). These high-paying jobs attract white-collar professionals to large cities, which house more and more of the population every year ([citation](#)). Additional businesses in the service and construction sectors begin and flourish to serve these wealthier people food and entertainment and build or renovate their housing. This process of a wealthy population moving into a lower-income area and bringing with them a flurry of business activity is called gentrification. Gentrification improves areas on paper in several ways, including creating additional jobs, lowering crime rates, improving schools, and providing tax revenue. However, one of the potential problems of gentrification is that with all of these improvements comes an increase in the cost of living in the neighborhood, forcing previous residents to leave. This could create a feedback loop, where those who can't afford to stay leave and are replaced by more wealthy individuals, bringing additional economic activity but also continually increasing costs. Eventually, the old population would be replaced by the new, wealthy one.

Chicago has traditionally been one of the most segregated cities in America, with dramatic income differences between high- and low-income areas ([citation](#)). Those differences are likely accompanied by similarly significant differences in property value, crime rate, and business activity. One of the historically most wealthy areas is the Lincoln Park neighborhood ([citation](#)), the home of the Lincoln Park Zoo and DePaul University. As other Chicago neighborhoods gentrify, one might expect to see them begin to resemble Lincoln Park more closely. Therefore, one way to try to capture the effects of gentrification could be to compare the change in a neighborhood's median income and/or median home value to the similarity of that neighborhood's business venue profile to Lincoln Park. A strong correlation could indicate that a business venue profile similar to Lincoln Park could indicate high cost of living, and so a trend toward that profile could be correlated to loss of long-time residents as home prices increase. An understanding of the rate of gentrification and the accompanying changes in business activity can help cities formulate meaningful responses to help lower-income residents stay in place and receive the benefits of some gentrification without being completely erased from the neighborhood's population. This information could also help entrepreneurs find prospective areas for new businesses. If they can't afford to build their business in Lincoln Park, they may be able to afford it in a neighborhood that is becoming more similar to Lincoln Park but is still relatively affordable.

It should go without saying that any correlation would not necessarily prove that gentrification is pushing people out. It could be that many are staying put, even as property values rise. Those who own businesses in the neighborhood, would quite possibly see their own incomes rise as more affluent neighbors move in. Cost of remaining in the neighborhood also would not correlate exactly with cost of entering the neighborhood. Lower income individuals may not be able to afford to move to a gentrified neighborhood, but they may be able to afford to stay there if they owned their own property or had some form.

Data

The primary source of data I will use to investigate business activity is the Foursquare API. I plan to look at business venues, like the New York City and Toronto work we have previously done, but with monthly historical versions of the API going back as far as Foursquare has data. Next, I will see which neighborhoods become more like Lincoln Park as time goes on by using a similarity score between their venue profiles and looking at the change over time. This similarity score over time will be the indicator that I use to represent business similarity between communities.

Once I have time-dependent business similarity data, I will compare that data with median home value across the same time frames to see how closely those indicators are correlated with the business similarity score. For the neighborhoods with the biggest change in business similarity to Lincoln Park, I will visualize the changes in median home value over time against the backdrop of the business similarity score.

A caveat for any results from this is that the data is not perfect. The Zillow data includes home sales, but doesn't account for the value of homes that are not sold, so the home sales data could be much higher than the actual median home value in gentrifying neighborhoods, where those moving in are likely to be spending more money on housing than those already living there.

Methodology

First, a list of Illinois zip codes was pulled in and trimmed to keep only Chicago zip codes. These Chicago zip codes were then merged with a dataframe of home values from Zillow over the last ~12 years, where only zip codes that matched Chicago zip codes were kept. Location data was gathered for all Chicago zip codes and that location data was fed into the Foursquare API to determine the venue profiles of the zip codes within Chicago.

A function was written which would automatically pull data from Foursquare for a range of version numbers, encode and group the results, and then calculate a similarity score to Lincoln Park by calculating the dot product of the venue profile of each zip code with Lincoln Park. This series of similarity scores for each zip code was added to the original dataframe under a column named after the version number. Version numbers every month going back to 2008 were used.

I wanted to see which neighborhoods had significantly increasing home values and looked at average, standard deviation, max, and min home sale price values from Zillow for all of the zip codes in Chicago and then plotted both the neighborhoods with the largest proportional value increase as well as the largest raw value increase. I excluded data before 2011 to try to avoid the impact of the 2008 financial crisis.

Results

When these similarity scores were calculated, unfortunately I found that they did not change over time. That is, the dot product of each zip code's venue profile with Lincoln Park's venue profile did not change over the time period observed. It could be that the changes over this time were too small to impact the venue data pulled from Foursquare, or that the changes did impact the data but not enough to impact the dot product.

I did see some significant changes in average home sale price over the years. It may not be surprising that some of the neighborhoods with the largest proportional increases in home value were those where the value was quite low to begin with. Wealthier neighborhoods had larger raw increases in home value, but because these were starting from a higher starting point, the proportional increase was quite small.

Discussion

Someone looking to acquire property as an investment generally would be better served spending their money in areas with low prices, as those have the most room to grow. This may seem counterintuitive, because typically real estate investment is associated with high value properties, but for those who don't have the financial resources to acquire such high-value properties, low-value properties can provide good return on investment.

Because low-income areas are also the areas where home values are most likely to increase, it seems like business owners should not try to avoid such places as a matter of practice, as there can be economic growth there as well. Rather, a successful business is likely to be one that connects with the culture and economic state of the neighborhood where it is located. Location is important, but one must understand how to leverage their location to be successful.

Conclusion

Unfortunately, the Foursquare data do not provide the insights into changing venue environments that I was hoping to see with this exploration. The data from Foursquare were not dynamic enough to register differences in neighborhoods over time that could indicate the changing venue profile of a gentrifying neighborhood. Additionally, all of the neighborhoods observed had increases in home value over the time period studied, so any gentrification that may be happening is not obvious from home value data. Additional information such as income information may be necessary to probe these ideas more fully.