McKay Rytting

Applied Data Science Capstone

Final Report

# Introduction

Global economic activity is increasingly concentrated in densely-populated urban areas ([citation](https://ourworldindata.org/urbanization)). These high-paying jobs attract white-collar professionals to large cities, which house more and more of the population every year ([citation](https://ourworldindata.org/urbanization)). 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](https://www.theatlantic.com/business/archive/2018/03/chicago-segregation-poverty/556649/)). 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](https://www.forbes.com/2007/08/30/most-expensive-blocks-forbeslife-cx_mw_0831blocks.html#1204408238e6)), 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.

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. Additionally, the income data comes from the IRS only counts those who file a tax return. Many lower-income individuals will not file tax returns, so the actual income of the zip code is likely lower than the IRS data suggest.

Once I have time-dependent business similarity data, I will compare that data with median income and median home value across the same time frames to see how closely those indicators are correlated with the business similarity score. Median income data will come from the IRS, and median home value data will come from Zillow’s median sale price data which is available for many zip codes back to 2008. For the neighborhoods with the biggest change in business similarity to Lincoln Park, I will visualize the changes in median income, and median home value over time against the backdrop of the business similarity score.