MATH 161
Abby Lewis & Tenny Liu
Final Project

Spatial distribution of Starbucks in California and Wisconsin explained by gentrification-related demographic factors

Abstract

Gentrification, or the movement of affluent residents and businesses into diverse, low-income urban areas, is known to cause shortages of affordable housing, health consequences, and cultural displacement. In this study, we examine the geographical distribution of Starbucks, an upper-middle-class aesthetic landmark, to determine its association with gentrification-related demographic factors. We find that a linear model relating county-wide starbucks abundance to median age, white population, median income, and median household value captures most of the spatial variation in Starbucks abundance within the state of California and performs comparably to several types of spatially-informed models (SAR, CAR, GWR). We then used our linear model to predict another state with a drastically different population, Wisconsin, and found mixed results. There was no spatial autocorrelation in the residual error in Wisconsin, indicating that we adequately captured much of the spatial variability in Starbucks locations. However, the model almost universally overestimated the number of Starbucks in counties in Wisconsin and severely underestimated one major city. While gentrification-related demographic factors can be used to explain trends in the spatial distribution of Starbucks density, other regional factors likely modulate these trends on a national scale.

Introduction

The word "gentrification" was first coined to describe the movement of upper-middle-class residents into low-income urban areas of London (Glass, 1964). Since its conception, the term has brought increasing awareness to the racial/ethnic and class tension caused by the demographic and economic transformations that characterize the process of gentrification. Among the popular interpretations of this phenomenon are the "supply-side" interpretation, citing the economic pricing of acquiring housing and store-fronts, and the "demand-side" interpretation, identifying the attraction of cultural and social vibrancy and diversity (Zukin, 1987). In recent years, centers under the U.S. Department of Health & Human Services are recognizing the housing, health, and cultural issues, especially affecting the "vulnerable population" of urban centers, that accompany the shift in consumption and capital dynamics in these "previously run-down neighborhoods" due to the influx of affluent professionals/residents and businesses (Centers for Disease Control and Prevention, 2015). It is under this heightened acknowledgement of the negative consequences of capital-fueled urban redevelopment that we spotlight Starbucks, an upper-middle-class aesthetic landmark and a stereotypical proxy for gentrification (Kasperkevic, 2015).

In this paper, we attempt to analyze the spatial pattern of Starbucks stores locations open as of February 2017 and their correlation with various demographic factors for each census tract. In addition to displaying trends using geographical mapping, we incorporated both general linear models and spatially weighted models to capture the relationship between the number of Starbucks stores and the explanatory demographic factors. We also tested for spatial correlations before and after fitting the models. We narrowed the scope of our analysis to the state California and used the most reasonable model derived to predict another state with a drastically different population, Wisconsin, to determine the applicability of our results in context of the US. Based on media coverage of the population moving to city centers and frequenting stores like Starbucks (Bliss, 2018), we aim to reveal how total population, white population, age, income, and property value can help model the number of Starbucks stores. Ultimately, we hope our study can help inform urban planning and policy making in order to tackle the social consequences of gentrification.

Methods

We obtained our dataset of Starbucks locations worldwide owned and maintained by Starbucks on Kaggle.com. The dataset is said to contain “a record for every Starbucks or subsidiary store location currently in operation as of February 2017” (Starbucks, 2017). Each store entry includes its geographical information like the city, state/province, and country the store is in as well as its longitude and latitude. We then created one subset of all store locations in California and another one of those in Wisconsin for later analysis. The demographic information from all census tracts in California and Wisconsin is scraped from the American Community Survey 5-Year Data (2009-2017) via Census API (United States Census Bureau, 2018). The demographic information encapsulates:

- total population,
- population of white alone or in combination with one or more other races,
- the estimated Median household income in the past 12 months (in 2015 Inflation-adjusted dollars),
- the estimated median age of total population,
- the estimated median value of houses moved in from 2010 to 2014,
- and the estimated median value of houses moved in 2015 or later.

We estimated the median property value moved in since 2010, and used it instead of the last two variables on the list above, to hopefully not only match the opening times of the Starbucks in our dataset, but also have a sense of the most recent real estate values. This new estimate was calculated by simply taking the average of the estimated median value of houses moved in from 2010 to 2014 and that of 2015 or later. It should also be noted that there are two counties with missing median value of houses and, therefore, the two counties are omitted from any regression regarding house values. This is not particularly concerning since those two counties do not vary much from most of the others in terms of number of Starbucks or demographics. We then tallied the Starbucks stores in each census tract in California and Wisconsin respectively and combined them with their corresponding demographic data.

To understand the spatial pattern in Starbucks locations, we first computed the Moran's I statistics for the number of Starbucks in each census tract in California. Here, we made a deliberate choice to use Moran's I instead of Geary's C, because we are more concerned with how the number of Starbucks in each county compares to the state average rather than among themselves. For this reason and for consistency, only Moran's I (neighbor's determined by Queen's contiguity) is used to assess spatial autocorrelation throughout this study.

Detecting the non-normality in the number of Starbucks, we performed a log transformation to achieve better model fit. We then proceeded to model the logged number of Starbucks as a response variable to multiple demographic components. By including and excluding each demographic variable in a linear model and running Moran's I for the residuals, we found the linear model that would yield close to spatial randomness in the residuals. We next tested if spatial regression captures the distribution even better. We used our best fitting linear model for geographically weighted regression (GWR), simultaneous autoregression (SAR),and conditional autoregression (CAR). We evaluated the appropriateness of the models based on their AIC and Moran's I for their respective residuals.

To determine the applicability of our final model in other demographic contexts, we used the linear model that we created for California to predict the density of Starbucks in Wisconsin. We chose the standard linear regression model over the spatial regression models because of not only the drastic geographical differences between the two states but also the similar fitting results which we will discuss in the next section. Finally, we again used Moran's I to assess the spatial autocorrelation in the residuals.

Results

Our dataset of Starbucks locations included 25,599 restaurants worldwide. These restaurants are concentrated in North America, Europe, and Eastern Asia (Figure 1).

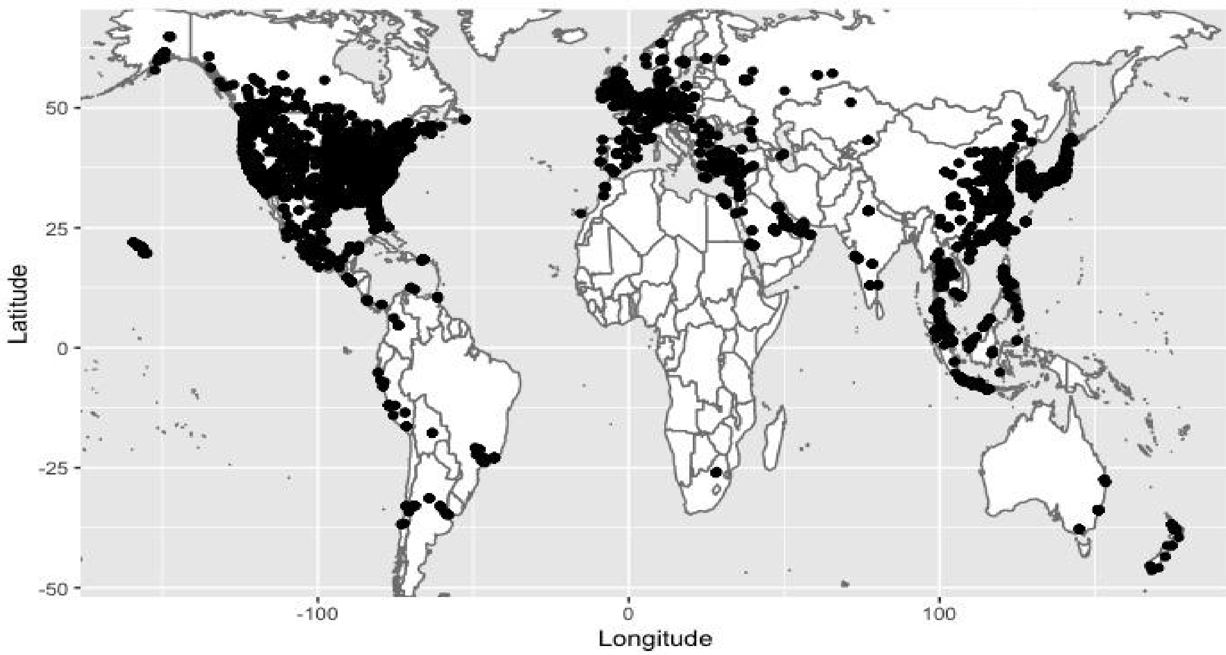

Figure 1: Starbucks locations worldwide

Within California, there are 2,821 Starbucks restaurants, and restaurants are primarily concentrated in larger metropolitan areas of San Diego and Los Angeles (Figure 2). The county-level abundance of Starbucks restaurants is significantly spatially autocorrelated within the state of California (Moran's I Statistic = 0.325, $p < 0.001$).

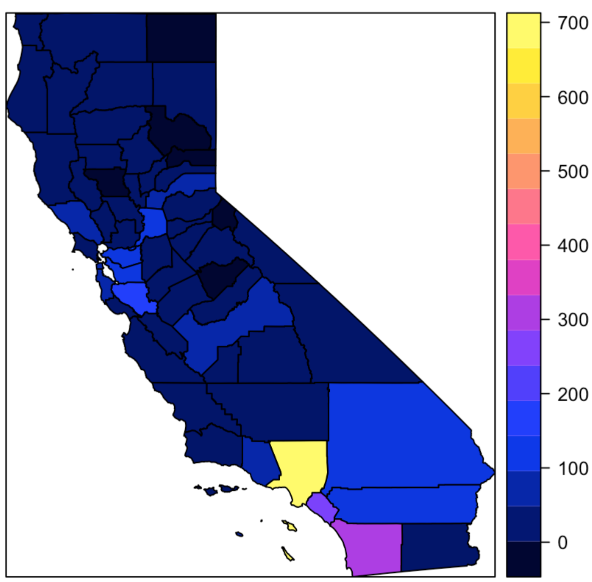

Figure 2: Number of Starbucks by county in California

When using standard regression to model the distribution of these Starbucks in California, we found that the linear model including median income, white population, total population, median age, and median house value best capture the variation in Starbucks distribution (Table 1). The residuals of this model are almost completely spatially random (Moran's I Statistic = 0.001309389, p = 0.409).

(Intercept)	Median_Income	White_Population
3.082e+00	3.738e-05	9.954e-07

Median_Age	Median_House_Value
-8.558e-02	5.141e-07

Table 1: Coefficients of final linear model

For spatial regression, the AIC value dropped from 132.52 of global regression to 124.52 of GWR and R-squared stayed around 0.82, indicating that the model was only improved slightly and that both were able to capture most of the variation. Although there appeared to be more pattern in the residuals of the GWR model(Moran's I Statistic = 0.032691207, p = 0.2745) than that of the linear model, it seemed to better capture the trend in Los Angeles and its neighboring counties (Figure 3). The SAR(AIC: 134.35) and CAR(AIC: 134.47) models both yield residuals with Moran's I around 0.03, which indicate that they were no improvement to the standard linear model.

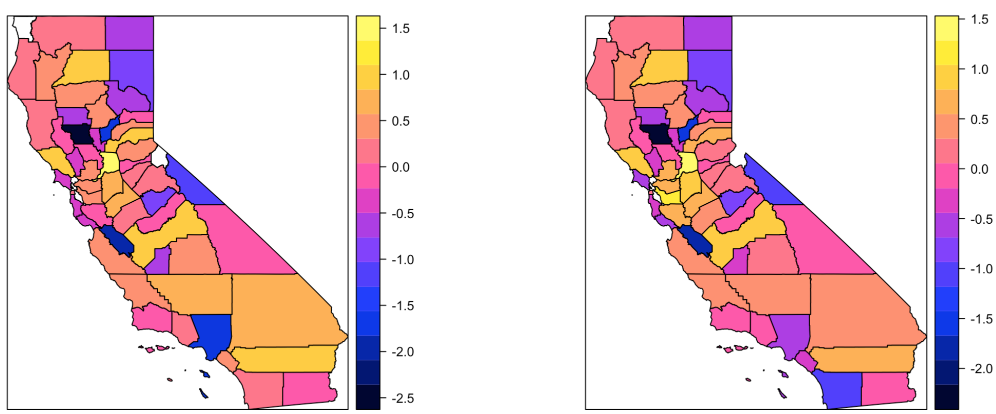

Figure 3: Modeling the number of Starbucks per county in California. Left: Residuals from final linear model. Right: Residuals from GWR model

Predicting Wisconsin

When we applied our final linear model from California to Wisconsin, it overestimated the number of starbucks throughout the state, except in Milwaukee County (Figure 4). Residuals were not significantly autocorrelated (Moran's I Statistic = 0.0388, p = 0.1731).

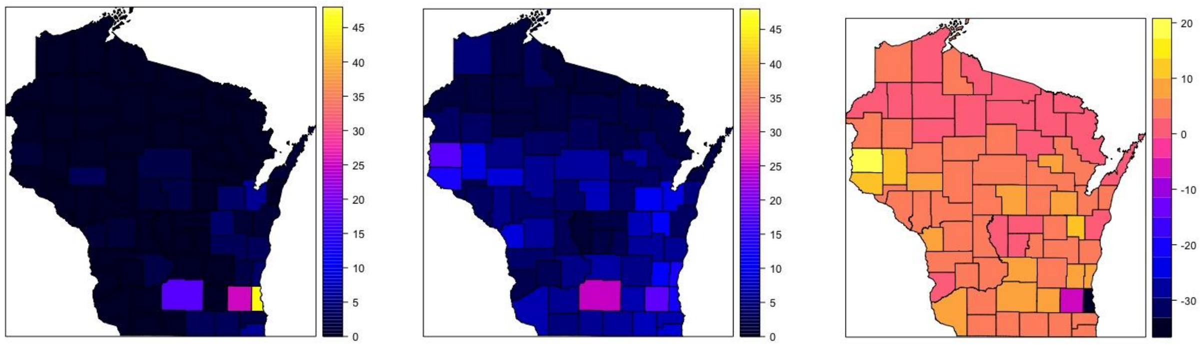

Figure 4: Predicting the number of Starbucks per county in Wisconsin. Left: Observed number of Starbucks per county. Center: Predicted number of Starbucks per county. Right: residual difference between observed and predicted.

Discussion

In this study, we demonstrate that demographic factors can be used to predict the locations of Starbucks in two demographically-distinct states in the USA. We built a model that captures much of the spatial variability in Starbucks locations in California based on purely demographic factors, including median age, white population, median income, and median household value (2010-present). We then extended the model to Wisconsin, with mixed results.

Many of the factors used in our final model are indicative of gentrification. For example, increases in median income and median house value are clear indicators of economic gentrification, and they were both included in our best model. According to our model, Starbucks are more likely to be found in areas where there are wealthier people and more expensive homes. This likely also reflects the level of urbanization, as cities in California typically have very high median income. Median age has been cited as an indicator of gentrification, because young single adults or childless couples tend to displace older people or families in the area and then move to the suburbs when having kids (Spain, 1992), and median age had a strong positive influence on the predicted number of Starbucks in a county. In addition to median income, this may play a role in capturing the level of urbanization because cities tend to have more young people. Gentrification also has strong racial ties, as it is often associated with cultural displacement and neighborhood "whitening" (Gibbons et al. 2016). Therefore, white population (included in our model) may also be an indicator of gentrification. Interestingly, our model faired better without a measure of total population or the percentage of the population that is white, indicating that the number of starbucks is most dependent on white population alone or that white population effectively integrates these other important factors. One reason that total population was not needed is likely that other demographic variables already captured the level of urbanization.

When we applied this model to Wisconsin, there was no spatial autocorrelation in the residuals, indicating that we adequately captured much of the spatial variability in Starbucks locations. However, the model almost universally overestimated the number of Starbucks in counties in Wisconsin. In California, many counties with low population and median income still have a few Starbucks restaurants due to major highways that pass through the county. The vast majority of Wisconsin does not have the same level of transit bringing people into the county. Furthermore, Wisconsin is a much less diverse state than California, so the fact that our model only includes the white population may artificially inflate predicted values in rural Wisconsin. The one county our model severely underestimated was Milwaukee county. Milwaukee is a large city but it has a very low median income, whereas cities in California typically have high median income. This may indicate that the presence of Starbucks is more dependent on urbanization than on income directly. Finally, our model significantly overestimated two counties in the northwestern part of the state: Peirce and St. Croix. These counties have all of the factors we would expect to predict a large number of Starbucks, but the observed number is surprisingly low. Upon doing more research, we found that these two counties are part of the Minneapolis metropolitan area, and they were the fastest growing counties in Wisconsin at the start of the 21st century. We predict that the number of Starbucks is likely to increase in the next few years in response to demographic changes, and the reason there are not many Starbucks yet is simply because of the time required to notice an opportunity, complete permitting, and start the store.

This study presents a first look at the distribution of Starbucks and the demographic factors that can be used to explain their distribution. Future studies could build off of what we have accomplished by studying the timing of when Starbucks restaurants are built and how that compares to changes in demographics. This would help identify whether gentrification is the cause or consequence of increased prevalence of Starbucks restaurants. Another useful next step would be to narrow down the specificity of the analysis to demographics of a given neighborhood or the few blocks surrounding a single restaurant. This would help confront potential biases associated with aggregating the number of starbucks in a given county. Such biases result from the modifiable areal unit problem (MAUP), as the patterns observed may be different on a regional level than on a local level, where the city planning decisions are ultimately made.

References

Bliss, Laura. “Yelp Reviews Can Track Gentrification. (And Might Encourage It.).” *CityLab*, 7 Sept. 2018, www.citylab.com/equity/2018/09/on-yelp-gentrification-is-in-the-stars/569419/. Accessed 29 Apr. 2019.

Centers for Disease Control and Prevention. "Health Effects of Gentrification". *Centers for Disease Control and Prevention*. 24 March 2015, https://www.cdc.gov/healthyplaces/healthtopics/gentrification.htm. Retrieved 8 May 2019.

Gibbons, J., & Barton, M. S. (2016). The association of minority self-rated health with black versus white gentrification. Journal of Urban Health, 93(6), 909-922.

Glass, R. (1964). *London : Aspects of change* (Centre for urban studies report, no. 3). London: MacGibbon & Kee.

Kasperkevic, Jana. “In Gentrified Cities Which Came First: Starbucks or Higher Real Estate Prices?” *The Guardian*, 3 Feb. 2015, www.theguardian.com/money/us-money-blog/2015/feb/03/starbucks-gentrification-real-estate-prices. Accessed 29 Apr. 2019.

Starbucks. “Starbucks Locations Worldwide.” *Kaggle*, 13 Feb. 2017, www.kaggle.com/starbucks/store-locations. Accessed 22 Apr. 2019.

Spain, D. (1992). A gentrification research agenda for the 1990s. *Journal of Urban Affairs*, 14(2), 125-134.

United States Census Bureau. “Detailed Table.” *2013-2017 American Community Survey 5-Year Estimates,* 6 Dec. 2018, https://www.census.gov/data/developers/data-sets/acs-5year.html. Accessed 22 Apr. 2019.

Zukin, Sharon. “Gentrification: Culture and Capital in the Urban Core.” *Annual Review of Sociology*, vol. 13, 1987, pp. 129–147. *JSTOR*, www.jstor.org/stable/2083243.