NBER WORKING PAPER SERIES

MEASURING GENTRIFICATION:
USING YELP DATA TO QUANTIFY NEIGHBORHOOD CHANGE

Edward L. Glaeser
Hyunjin Kim
Michael Luca

Working Paper 24952
http://www.nber.org/papers/w24952

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2018

We thank Byron Perpetua for excellent research assistance. We thank Susan Athey, Shane Greenstein, and Luther Lowe for valuable feedback, and Yelp for providing data for this analysis. Kim and Luca have done consulting for tech companies including Yelp, but their compensation and ability to publish are not tied to the results of this paper. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24952.ack

Measuring Gentrification: Using Yelp Data to Quantify Neighborhood Change
Edward L. Glaeser, Hyunjin Kim, and Michael Luca
NBER Working Paper No. 24952
August 2018
JEL No. D22,O18,O30,R11

ABSTRACT

We demonstrate that data from digital platforms such as Yelp have the potential to improve our understanding of gentrification, both by providing data in close to real time (i.e. nowcasting and forecasting) and by providing additional context about how the local economy is changing. Combining Yelp and Census data, we find that gentrification, as measured by changes in the educational, age, and racial composition within a ZIP code, is strongly associated with increases in the numbers of grocery stores, cafes, restaurants, and bars, with little evidence of crowd-out of other categories of businesses. We also find that changes in the local business landscape is a leading indicator of housing price changes, and that the entry of Starbucks (and coffee shops more generally) into a neighborhood predicts gentrification. Each additional Starbucks that enters a zip code is associated with a 0.5% increase in housing prices.

Edward L. Glaeser
Department of Economics
315A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

Hyunjin Kim
Harvard Business School
Soldiers Field Road
Boston, MA 02163
hkim@hbs.edu

Michael Luca
Harvard Business School
Soldiers Field Road
Boston, MA 02163
mluca@hbs.edu

I. Introduction

"Gentrification: New Yorkers can sense it immediately. It plumes out of Darling Coffee, on Broadway and 207th Street, and mingles with the live jazz coming from the Garden Café next door" – New York Magazine (2014)

Gentrification has emerged as an important policy issue in cities from New York to Edinburgh to Seoul. Urban leaders often applaud the in-migration of the prosperous and associated increases in property values, and gentrification enthusiasts point not only to housing markets, but also to the potential for a more vibrant commercial sector supporting local businesses and services. Yet, others fear that gentrification will both price out poorer renters and destroy the historic character of the neighborhood.

In practice, any policy response to gentrification requires both a reasonably up-to-date understanding of where gentrification is occurring, and a sense of how neighborhoods actually change with gentrification. Where are housing prices spiking? Which neighborhoods are experiencing changing demographics and an altering local economy? What do these changes look like?

Standard public data sources cannot fully answer these questions, both because they often come with multiple year lags and because there is limited public data about the types of businesses operating within an area. For example, the American Community Survey provides data at the local level, but small sample sizes mean that these are based on five years of data. For example, the current estimates cover 2012 to 2016 and are really most representative of 2014, almost four years ago. The Federal Housing Finance Agency has begun to provide repeat sales indices at the tract level, and these are available up to 2016. Yet, the indices are still in development, and prices may not be particularly representative for urban neighborhoods that are dominated by rental units. Moreover, data from government statistical agencies contain only coarse categorizations of business type.

In this paper, we explore the potential for Yelp data to provide city leaders and economists with real-time information on neighborhood change. To do so, we examine whether changes in the type of Yelp establishments in a neighborhood are associated with contemporaneous changes in five different measures of neighborhood change. The use of Yelp data as a complement to standard data sources has two advantages in this context. First, since data about the local economy is available virtually in real time on Yelp, it can help to detect which parts of their city are experiencing gentrification without waiting for public statistics to become available. Second, it provides a more granular understanding of how neighborhoods change during periods of gentrification.

As an illustrative example, suppose that Starbucks were to move into a struggling neighborhood. As the New York Magazine quote suggests, this may be an indicator of gentrification within a neighborhood. Combining Census data with Yelp data for the United States, we test this hypothesis and find that the entry of new Starbucks stores into a ZIP code is in fact a predictor of housing price growth. The entry of each additional Starbucks into an area is associated with a 0.5% increase in local housing prices. We find similar results when looking at coffee shops as a whole. We then expand this analysis to explore the broader changes to the local economy that accompany gentrification. We find that housing price growth is associated with entry of grocery stores, restaurants, barbers, and convenience stores – as well as increased activity at these businesses, as measured by Yelp data.

We then turn to three demographic measures of neighborhood change based on American Community Survey data (ACS) at the Census ZCTA level. We focus on New York City, and then compare results in New York with results for four other large cities. Our three measures of gentrification are the percent of population with a college degree, the percent aged between 25 and 34, and percent white – drawn from Census data. Because of the nature of the ACS, we are forced to consider only a long difference between the 2007-2011 period and the 2012-2016 period. To get comparable Yelp data, we also average over those two five-year periods. This work follows Waldfogel (2008) who examines the relationship between neighborhood composition and the location of different types of consumer goods.

Within New York City, we find that neighborhoods with rapid expansion of local businesses (ranging from grocery stores to bars to hair salons) also attract an influx of more educated residents. Growth in local businesses is also associated with an influx of younger residents in New York. There is a weaker relationship between changes in business activity and the racial composition of a neighborhood. We then repeat these analyses for four additional cities: Boston, Chicago, Los Angeles and San Francisco. Some of the patterns are consistent, such as the link between growth in education and growth in cafes and bars. Others vary across cities. For example, the expansion of laundromats as only a strong indicator of gentrification in New York City.

Our final outcome variable is the change in the value of Streetscore at the block level. Streetscore is a computer-generated measure of the way humans perceive the safety of an image of a streetscape. Naik et al. (2017) use changes in Streetscore as a measure of the change in the physical quality of the neighborhood, and find that it is particularly associated with education and density levels. We look at the relationship between Streetscore change and change in the number of establishments across a wide range of industries. We find that Streetscore change is most positively associated with increases in the number of vegetarian restaurants, Starbucks and cafes, and wine bars and bars more generally.

While these patterns provide further insight into the patterns of gentrification, the direction of causality is a priori unclear. In the final section of the paper, we explore the timing of changes using the notion of Granger causality. Do business openings typically precede increases in neighborhood education, or do they follow demographic shifts? With housing price changes, we find that openings of Starbucks precede price growth, but that price growth does not predict growth in Starbucks openings. Our preferred interpretation is that exogenous neighborhood changes spur store openings, which then encourage prices to rise.

We also look at whether demographic change precedes or follows our five strongest yelp variables: laundromats, groceries, cafes, bars and restaurants. We find that higher initial numbers of these establishments predict growth in the share of the population that is college educated, but that higher initial levels of college education do not predict growth in these variables. Once again, Yelp establishment growth seems to be a leading indicator, and perhaps even helps the process of neighborhood change.

Overall, our results suggest that changes in economic activity, as measured by Yelp, can provide leading indicators of gentrification within a neighborhood. Yelp data provides relatively up-to-date information that is useful for understanding where gentrification is occurring, before official statistics become available. Yelp data also provides additional insight about how gentrifying neighborhoods change. More generally, this type of data speaks to an additional benefit of digitization and the Internet age. While the intended purpose of Yelp data is to help customers identify local businesses, the same data is a valuable resources for policymakers and researchers looking to better understand economic activity.

II. Data Description

We begin by discussing our outcome variables. We then turn to the Yelp data, which we will use to predict changes in those variables.

Measures of Neighborhood Change

Our most straightforward measure of neighborhood change is the ZIP-code level housing price data provided by the Federal Housing Finance Agency (FHFA). This data is an annual repeat sales index for over 18,0000 five digit ZIP-codes in the U.S., described in Bogin, Doerner and Larson (2016). We use data from 2012 to 2016, and the average real growth of this index over this period is 3.1 percent (see Table 1).

Price indices are a tricky measure of neighborhood change, because economic theory predicts that they should be forward-looking. A neighborhood that is expected to gentrify should experience price increases before any changes in the neighborhood demographics occur, or before rents start to increase. Consequently, price indices may move before gentrification, or possibly if there is price stickiness, after gentrification. Moreover, price indices are likely to represent a better reflection of demand for a neighborhood when supply is relatively fixed, which can occur either because prices remain below construction costs or because strict land use controls limit the production of new housing.

While ZIP-code level pricing data are available annually for a large number of ZIP-codes, our demographic data for ZIP codes is available only for five-year windows. This data comes from the American Community Survey, which replaced the Census long form after 2000. Since ZCTA, tracts and block groups are only surveyed sporadically, the Census aggregates years for these smaller geographic units. Moreover, there is surely error generated by the fact that some areas may be surveyed earlier in the five-year window while other areas are surveyed later in the window, but it is difficult to actually quantify this problem.

We use three different measures of demographic neighborhood change: percent college educated, percent aged between 25 and 34, and share of the population that is white. We see percent college educated as the most natural measure of gentrification, because education is so reliably correlated with both income and housing costs. In our sample, the average ZIP code saw the share of adults with college degrees increase by 2.6 percent in New York and 3.4 percent for our other four cities (Boston, Los Angeles, Chicago, and San Francisco).

Gentrification is also often associated with changes in the age distribution, with younger professionals replacing older longer-term residents. Consequently, we also use the share of the population that is between 25 and 34 as an outcome. The change in this variable is significantly correlated with the change in the share of college graduates in New York City; the correlation coefficient between the two variables is .4. The average ZIP code in our sample in New York saw the share of the population aged 25 to 34 grow by 0.3 percent. In the other cities, the average share of the population aged 25 to 34 rose by 1 percent.

Our final demographic variable is the share of the ZIP code's population that is white. Naturally, there can be gentrification without racial change, but many of the most explosive instances of gentrification occur when richer white tenants displace poorer African-Americans. The change in this variable is also correlated with both percent of adults with a college education (correlation coefficient of .32 in New York). In New York, the percent white decreased by 1.1 percent in our sample.

Our final measure of neighborhood change is Streetscore, borrowed from Naik et al. (2015) and Naik et al. (2017). This measure has its results in a crowd-source data set generated by Cesar Hidalgo, in which respondents were asked to rate which image from Google Streetview appears safer. The ratings generated by this crowd-sourcing were then used as training data for computer vision techniques, which generated a Streetscore for a much wider range of neighborhoods. Changes in the measure do correlate well with physical upgrading found in the records of the Boston Planning and Development Agency.

We interpret this measure as a proxy for the overall physical quality of the neighborhood, not of its safety per se. In fact, many areas with low Streetscores also experience low crime, presumably because people do not wander around such unappealing areas. Naik et al. (2017) have shown that changes in these measures correlate with density and education, and with neighboring Streetscore, density and education. Consequently, there appears to be a spread in the perceived physical quality of the neighborhood that resembles the usual process of spreading gentrification.

Streetscore data is generated from Google Streetview images, and consequently, we have this data for a limited number of years. We have a large number of images from around 2007, when Google Streetview was first put on line, and from the period after 2014. In this case, we will be using this long difference in Streetscore as our measure of gentrification. The mean increase in Streetscore in our New York sample is 1.6.

Yelp Data

For measures of changes in the types of establishments, we use data from Yelp, an online platform that publishes crowdsourced reviews about local businesses. Business listings on Yelp are sourced through Yelp's internal team, user submissions, business owner reports, and partner acquisitions, and checked by an internal data quality team. The data begin in 2004 when Yelp was founded, which enables U.S. business listings to be aggregated at the ZIP code, city, county, state, and country level for any given time period post-2004.

Despite its granularity and availability, Yelp data comes with limitations. Yelp's establishment classification is assigned through user and business owner reports and Yelp's internal quality check. As a result, businesses are not always categorized systematically, or equivalently to government data sets. Furthermore, the extent of Yelp coverage also depends on the degree of Yelp adoption, which has grown over time as the company has become more popular. Lastly, businesses with no reviews may receive less attention from users – and therefore may be less likely to be flagged as open or marked as closed even after they close. We discuss these issues in further detail in Glaeser, Kim and Luca (2017).

To account for these limitations, we only count businesses as open if they have received at least one recommended Yelp review. We also limit our sample to years post-2007 to match the ACS (and to post-2012 when possible), and to ZIP codes with at least one business recorded in this period in at least one of the categories we examine. The average ZIP code in New York City had 80 restaurants and 814 restaurant reviews on Yelp between 2007-2011, and experienced a mean increase of 36 restaurants and 1525 restaurant reviews between 2007-2011 and 2012-2016. Additional summary statistics for the number of establishments and reviews by categories are reported in Table 1.

III. Nowcasting Local Price Growth

In this section, we focus on the ability of Yelp data to predict contemporaneous changes in housing price growth at the ZIP-code level, as measured by the FHFA repeat sales index. This is an exercise in nowcasting, not prediction of the future. We will return to the whether Yelp data can predict future price growth in Section VI. Our exercise here is meant to ask whether Yelp data can be a useful indicated of price growth for policy-makers who have not yet received that data from FHFA about price increases. We look at the period from 2012 to 2016.

The Starbucks Effect

We begin by following Rascoff and Humphries (2015) who used Zillow price data and identified a "Starbucks effect," linking proximity to Starbucks stores and price growth. In our variant of their exercise, we examine whether price growth is correlated with contemporaneous growth in the number of Starbucks cafes, which allows us to understand whether Starbucks is a leading indicator of gentrification. We restrict the sample to include ZIP-codes that had at least one Starbucks during the initial year.

Our second specification in Table 2 regresses the percent growth in the home price index on the absolute increase in the number of Starbucks in the ZIP-code during that same year. We include year dummies to control for macroeconomic changes that impact housing price growth. We also cluster our standard errors by ZIP-code. We find that a one unit increase in the number of Starbucks during a given year is associated with a 0.5 percent change increase in housing prices.

This is a large effect, both economically and statistically, but the added explanatory power created by the Starbucks control is modest. The t-statistic is about six. The r-squared added by controlling for Starbucks, over and above the year dummies, is only .002. As found by Rascoff and Humphries (2015), the presence of Starbucks is associated with price growth, but Starbucks presence is hardly a great predictor of which areas will grow.

Naturally, we do not suggest, and do not think, that this is a causal estimate. There are two more plausible interpretations of this correlation. First, it is quite possible that Starbucks targets its cafes in places that are on the upswing, so the correlation reflects the endogeneity of Starbucks locations. Second, Starbucks may be correlated with other changes in the neighborhood, such as gentrification, and collectively these changes drive prices upwards.

To partially distinguish between these hypotheses, the third regression includes a ZIP-code fixed effect. Our time period is short, and if Starbucks is targeting growing areas, then a ZIP-code level fixed effect should eliminate much of the correlation. Including these fixed effects causes the estimated coefficient to fall to 0.17 and the r-squared to rise to 0.37 percent. A significant portion of the relationship between Starbucks and growth reflects the tendency of Starbucks to locate in communities that have a longer term positive trend.

We believe that our time period is too short to simultaneously estimate ZIP-code level fixed effects and anything using about the timing of the relationship between Starbucks and price growth. Hence we drop the ZIP code fixed effects for the remainder of the table.

In the fourth regression, we include both the current and lagged Starbucks growth. Both terms end up being about equally significant and have roughly equal magnitudes. In the fifth regression, we include two lags of Starbucks growth. All three terms are significant, but the second lag has a smaller coefficient.

The sixth regression includes two other Yelp variables: the increase in the number of Starbucks that are closed and the growth in the number of Starbucks reviews. Notably, the reviews are specific to the café, not the home location of the reviewer. An increase in the number of closed Starbucks does not predict price increases or declines. The growth in the number of Starbucks reviews provides our most powerful Yelp variable.

A 10-unit increase in the number of reviews is associated with a 1.4 percent increase in housing prices in the ZIP-code. Including this variable leads all the other Starbucks variables to become far less significant, and it increases the r-squared of the regression from .24 to .26. We take this variable as a proxy for the amount of business at the Starbucks, and it suggests that prices go up when a community has more members that use Starbucks.

Notably, this finding pushes against any interpretation that suggests that people are paying for proximity to Starbucks. The presence of a Starbucks is far less important than whether the community has people who consume Starbucks. Consequently, we think that this variable is likely to be a proxy for gentrification itself, reflecting an increase in the number of people who like to consume expensive coffee and then write about their experience online.

The Cafés Effect?

While Starbucks may be a particularly prominent upscale coffee house, there is little reason to think that it is the only possible retail establishment that can capture gentrification at the local level. In Table 3, we expand our analysis to include all of the cafes listed in Yelp over the same time period. This change enables us to considerably increase the number of ZIP-codes in our analysis. The number of ZIP-codes with a café in our sample is about twice as large as the number of Starbucks in our sample.

Table 3 reproduces Table 2 with growth in the number of cafes substituting for growth in the number of Starbucks. Our second regression shows that the coefficient on cafes is similar to our coefficient on Starbucks: .54. A ten unit increase in the number of cafes in a ZIP-code is associated with an approximately five percent increase in prices. The t-statistic for this coefficient is 16. The overall r-squared for this regression is lower than for regression 2 in Table 2, but this reflects the lower explanatory power of year effects in this larger sample. The r-squared that is added by the cafes variable is .008, which is four times larger than the r-squared added by the Starbucks variable, but still a modest contribution.

Just as in the case of Starbucks growth, this growth might reflect the strategic location of cafes in areas with rising incomes and housing prices. Regression 3 includes ZIP-code specific place effects. The r-squared rises to .21, and the coefficient on cafes falls to .02 and loses statistical significance. Once again, there is evidence for strategic location, but regressions with fixed effects can be problematic when the time series is this short.

In the fourth regression, we include a lag of café growth as well. Once again, both variables are quite significant, and the r-squared rises from .157 to .163, which is a modest increase but a larger one than that associated with adding lagged Starbucks growth. In regression 4, we include two lags and now all three variables are significant. The r-squared increases to .167.

In the final regression of this table, we include café closings and growth in the number of café reviews. The pattern is similar to that seen in Table 2, regression 6. The number of closings is negatively correlated with price. The growth in the number of café reviews is strongly associated with price growth, and it reduces the coefficient on contemporaneous and lagged café growth.

However, the magnitude of the power of general café reviews is somewhat weaker than Starbucks reviews. One interpretation is that growth in Starbucks reviews disproportionately measure growth in upscale patrons, and so is a better proxy for gentrification of the neighborhood.

The Correlation of Home Prices with Other Yelp Establishments

In this section, we expand our analysis to other Yelp industries. Table 4 provides the results when growth in housing prices is regressed on contemporaneous percent growth in establishments from a series of different retail clusters. All regressions are univariate with year fixed effects and standard errors are clustered by ZIP-code.

The first column shows the estimated coefficient and its standard error. The second column shows the r-squared from the regression. The third column shows the r-squared when home prices are regressed on the year dummies alone for the relevant sample. The fourth column shows the number of observations, which change across samples because in some cases, ZIP-codes have none of the relevant establishments at the beginning of the time period.

The results show that across almost all categories, more establishments are associated with more growth. The only real exception is the number of Chinese restaurants. The correlation with the number of high price restaurants is also statistically insignificant, and the correlation with the number of restaurants opened (which is essentially the change in the rate of change) is small.

While many of these effects are quite large in both statistical and economic terms, their predictive power is limited. The variables which add the most to r-squared, over the year fixed effects, are barbers (.013 added r-squared), cafes (.008 added r-squared), restaurants (.007) and wine bars (.007 added r-squared). The largest point estimate is for laundromats.

In Table 5, we look at the growth in the number of Yelp reviews instead of the number of Yelp establishments. In this case, all of the coefficients are significant and positive. The r-squared increases are generally larger. Including the number of restaurant reviews increases the r-squared from .128 to .159, which is a material increase in goodness of fit. Other measures of the number of reviews, including barber reviews and café reviews, added to the r-squared in a meaningful manner.

In this section, we have confirmed that home price index growth is correlated with growth in a variety of different types of Yelp establishments. In many cases, as in Starbucks, the number of Yelp reviews provided more predictive power than the number of Yelp establishments. This was not true when we tried to predict the number of establishments in County Business Patterns (Glaeser, Kim and Luca, 2017), suggesting that Yelp reviews may be a more important correlate of gentrification if a less important correlate of economic development.

IV. Nowcasting Demographic Change

We now turn to predicting demographic change with Yelp data. In this case, our outcomes are the change in demographic variables at the Census tract level. Our first period extends from 2007 to 2011. Our second period runs from 2012 to 2016. We look at the ability of Yelp data to predict changes between those two periods.

Results for New York City

In this case, we first focus on New York City, a place with significant gentrification. We then turn to correlations outside of New York City. We limit ourselves to the ZIP codes that had at least one example of an establishment type during the entire period. This restriction means that we lose a significant number of ZIP-codes for many establishment types. We list only those establishment types for which we have more than 100 ZIP-codes, except for Starbucks, which is included despite being represented in slightly less than 100 ZIP-codes.

Each row in Table 6 shows the pairwise correlation between the growth in the number of establishment of each type and the change in the demographic variable. We use the absolute change in the number of establishments, which eliminates the need to worry about cases where there are zero establishments in the pre-period. Beneath each correlation coefficient, we report the p-value, the estimated probability that the correlation is actually zero. We also show the number of observations, which are the same across columns but differ across rows, because the Yelp data is different for each ZIP code and category.

We order the results by the strength of the correlation with change in the share of the adult population in the ZIP-code that is college-education. Our first row shows the results for the change in the number of groceries, which has a correlation coefficient of .35 with the change in the share of adults with college degrees. This relationship is strongly significant.

The correlation of the change in the number of groceries with the shares of the population that are between 25 and 34 and the share of the population that is white are also significant at the five percent level, but they are smaller. These correlation coefficients are approximately .18, or about one-half the size of the correlation with change in percent college educated. These results seem compatible with the literature on “food deserts” that documents how poorer people live in areas with fewer options for healthy food.

The second row shows the .338 correlation between growth in laundromats and the growth in the share of the population with college degrees. We found this result somewhat surprising, as we do not typically think of laundromats as a natural indicator of gentrification. Yet, many older neighborhoods, especially in Brooklyn, lack indoor laundry machines. As the population in such areas becomes wealthier, perhaps they are willing to pay more for laundry facilities that are nearby.

The number of laundromats also correlates with the share of the population that is young, which is perhaps less surprising. The correlation coefficients between these two variables is .2, which is significant at the ten but not the five percent level. Laundromat growth is not significantly correlated with the change in the share of the population that is white.

The third row shows the .319 correlation between change in the share of the population with college degrees and the number of cafes. These results support the findings of the previous section that cafes are a strong indicator of gentrification. Again, the results are quite strong statistically for the share with college degrees.

The relationship between change in the number of cafes and our two other demographic variables are much smaller and statistically indistinct from zero. Apparently, the growth in the number of cafes are not strongly correlated with either growth in the younger population or growth in the white population. We found those results somewhat surprising, but perhaps café growth is also related to business location, while grocery growth is more strongly related to residential change.

The fourth row shows the growth in the number of bars, which has a .313 correlation coefficient with growth in the share of adults with college degrees. This correlation is almost exactly as strong as the correlation with cafes. In this case, the correlation with growth in the share of the population that is young is insignificant. We were somewhat surprising that the correlation between bars and youth was not stronger. The correlation with percent white was weaker still and insignificant.

The fifth row shows the .27 correlation between change in the share of the population with college degrees and the change in the number of restaurants. Better educated, better paid restaurants eat out more often and are willing to pay more for good food. Restaurant growth also correlates significantly with the share of the population that is young at the ten percent level. The correlation between racial change and this variable is insignificant.

The sixth row looks at the change in the number of barbers, which has a .237 correlation coefficient with change in the share of the population that is well educated. This variable again correlates significantly with the change in the share of the population that is young, but not the share of the population that is white.

The seventh row shows the .232 correlation between change in the number of wine bars and change in the share of the population that is college educated. The correlation between this variable and share that is young or share that is white are both .14 and insignificant.

The tables also show significant correlations between change in the share of the population that is college education and change in the number of convenience stores, change in the number of fast food restaurants, change in the number of florists, and change in the number of restaurants that are categorized by Yelp as being pricey. Florists also correlate with the number of people who are young, but otherwise these variables are unrelated to changes in the share of the young or share of the population that is white. The other variables are insignificantly correlated with change in the share of the population that is well-educated, young or white, except for the change in the number of restaurant openings, which is correlated with the change in the share of the white population. Of our sample, we found eleven variables that had significant correlations with educational improvements. The correlations with the other variables were almost uniformly weaker.

Table 7 repeats Table 6 but using the percentage increase in the number of Yelp reviews, rather than the change in the number of Yelp establishments as our dependent variable. The results show a similar pattern. For example, laundromats again correlate particularly highly with gentrification, but the correlations decline. This change reverses our finding with price increases, which were generally more correlated with the number of reviews than with the number of establishments. This change may reflect a difference between demographics and prices, or a difference between New York City and the rest of the country. We will return to this issue in the next subsection when we look at results outside of New York.

Table 8 looks at the total explanatory power using four of our strongest univariate predictors. Collectively, change in restaurants, cafes, bars and groceries can explain 16.2 percent of the variation in the education change variable across 170 New York ZIP-codes. Perhaps surprisingly, groceries have the strongest positive coefficient. The coefficient on restaurants actually flips sign, suggesting that the well-educated may be more interested in convenient places to shop than places to eat. In the second regression, we include initial controls for the three demographic measures. The r-squared rises to 31.1 percent. The coefficients remain relatively stable but generally decrease slightly, with the exception of the number of cafes, which rises by .05.

Replications for the Other Large Cities

As a form of replication, and to test whether these results are particular to New York City, we now reproduce Table 6 and 7 using data from four other large cities: Boston, Chicago, Los Angeles and San Francisco. Again, we look at ZIP-code correlations between changes in the demographic variables. Table 9 looks at the correlations with Yelp establishment numbers. Table 10 examines correlations with the number of Yelp reviews.

Many of the patterns are broadly similar. The number of restaurants, cafes and bars are among the strongest correlates of gentrification across these four cities, although the correlation coefficients are somewhat smaller than in New York City. The number of laundromats is no longer a strong correlation of gentrification, suggesting that surprising finding was really describing an unusual New York City phenomenon. In these other cities, the number of florists was particularly strongly correlated with change in the share of the population with college degrees.

Just as in the case of New York, the correlations between the number of establishments and education are stronger than with either age or race. Also like New York, the correlation between reviews and education was weaker than the correlation between number of establishments and education, but with one significant difference. Several of the review counts correlated reasonably well with the number of younger people in the ZIP-code.

Consequently, one puzzle is why an increase the number of Yelp reviews in a ZIP-code correlates with an increase in the number of young people in these cities, but not in New York. One explanation might be that reviews for New York restaurants often come from people outside the ZIP-code. A second explanation is Yelp reviewers are more disproportionately young outside of New York City.

V. Nowcasting Physical Change

Our final outcome is the physical change in the neighborhood as measured by Streetscore. As before, we begin with our results on New York City and then turn to results for other large urban areas. As discussed in the data section, Streetscore represents a computer-generated measure of how safe the street looks to humans based on Google Streetview images in 2007 and 2014.

Predicting Streetscore Changes in New York City

To keep results comparable, we continue to look at ZIP-code level data, although there is no reason why we could not look at the block itself. Indeed, given that we have the precise location for every Yelp business and every Google Streetview image, it would be quite straightforward to estimate complex spatial models linking business location with change in the built environment. Again, we are not suggesting anything causal with these relationships, and we have ordered the table so that the strongest relationships are at the top. We have also excluded those restaurant classes for which we have fewer than 100 ZIP-codes, with the exception of Starbucks.

In Table 11, the strongest correlation is with the number of vegetarian restaurants, which had a much weaker correlation with the change in the share of the population that was well educated. The change in Streetscore has a .37 correlation with the change in the number of vegetarian restaurants. The second strongest correlation is with the change in the number of Starbucks restaurants, which is .355.

The stronger correlations of these variables with Streetscore than with neighborhood demographics might reflect a substantial difference between these two types of variables. Streetscore can change as readily for commercial areas as for residential areas, and indeed, business owners might be more likely to invest in obvious physical upgrades than residents. The many New Yorkers who live in historic districts are specifically prevented from doing much to change their physical appearance. Consequently, it may be that variables that are linked with commercial success correlate more strongly with Streetscore.

The third strongest correlation, a coefficient of .339, is with wine bars. This mirrors the results with demographic change. The significant correlations of Streetscore change with changes in cafes, florists, barbers, and restaurants overall, also all parallel the results of Table 6. There are also positive correlations with the number of fast food restaurants and convenience stores. Yelp changes the predict changes in the share of the population with a college degree also predict changes in the physical environment.

One notable difference is that growth in the number of laundromats does not predict upgrading in Streetscore, even though it is correlated with increases in the share of the population with a college degree. This difference may perhaps reflect the tendency of laundromats to locate in older residential areas that cannot be upgraded for regulatory reasons, which would suggest that laundromats appear specifically in areas that are getting richer but that cannot change their physical footprint.

VI. What Comes First?

We have documented correlations between gentrification and contemporaneous changes in Yelp openings and reviews. A second question is whether Yelp data predicts or "Granger causes" future gentrification. Does a recent opening of groceries or cafes predict subsequent gentrification? Granger causality does not imply true causality, as Yelp openings could precede gentrification solely because entrepreneurs expect gentrification that is to come. Still, understanding the timing of Yelp openings and gentrification has value from a purely predictive perspective, and might also enable us to reject some explanations for the observed correlations.

We begin with housing prices, which enable us to look at a relatively rich annual time series. With the demographic data and Streetscore, we can only look at whether early Yelp openings predict change in gentrification measures and whether early measures of gentrification predict changes in the Yelp variables.

The Timing of Housing Prices and Yelp Openings

While our time series is limited, we can still run regressions of annual percent changes in the ZIP-code level housing price indices on lagged changes in the growth of Starbucks or Cafes more generally. We begin this with Table 12, in which changes in housing price indices are first regressed on two lags in Starbucks growth. We do not include contemporaneous Starbucks growth.

The first regressions shows the both lags are economically and statistically significant. A ten-unit growth in Starbucks in the past year is associated with a 4.8 percent increase in housing prices. A ten-unit growth in Starbucks two years ago is associated with a 2.6 percent increase in housing prices during this year.

While these correlations are robust, they could just reflect the well-known correlation of housing prices changes with the one-year lag of housing prices changes (Case and Shiller, 1989, Glaeser and

Nathanson, 2017). To check whether the serial correlation of housing prices appears at the ZIP-code as well as the metropolitan area level, the second regression simply regresses price changes on two years lags of those price changes. Both coefficients are positive, although somewhat smaller than those typically reported at the metropolitan area level. The coefficient on the one year lagged value of price change is .324 and the coefficient on the two-year lag is .076.

Granger causality typically requires that lagged variables of one variable predict the change in a second variable, controlling for lagged values of the second variable. Consequently, in the third regression, we control for two lags of both Starbucks growth and past housing price growth. The coefficient on the one-year lag of Starbucks growth falls from .48 to .29, but it remains statistically significant. Starbucks growth does seem to predict future housing price growth, even controlling for the recent past of housing price growth.

In Table 13, we ask whether housing price growth predicts growth in Starbucks. The first regression shows that there is a small and statistically marginal effect of price growth in the previous period on Starbucks growth. A ten percent increase in price growth is associated with a .2 increase in Starbucks. The second regression shows that past Starbucks growth is not associated with current Starbucks growth during this period. The third regression shows that there is still a small and statistically modest impact of past price growth on future Starbucks growth.

Our overall finding here is that past Starbucks growth predicts future price growth much more strongly than past price growth predicts future Starbucks growth. Our preferred interpretation of this finding is that Starbucks locations are chosen by individuals with very good judgment about where prices are going to increase. A second possibility is that Starbucks enters early in the stage of gentrification because it caters to a crowd that is willing to move early into up-and-coming neighbors. We find the interpretation that Starbucks is actually causing price increases to be distinctly less plausible.

Which came first? Demographic Change or Yelp Establishments

We do not have rich enough data for our demographic variables to properly study the timing of change. Our closest approximation is to look at whether Yelp establishments during 2007-2011 predict demographic changes over the next five years, or whether demographics during 2007-2011 predict growth in Yelp establishments or both.

We will focus on five core types of Yelp establishments, and on the percent with a college education. We first look only at univariate relationships, and then also control for the initial value of the independent variable. Instead of just looking at New York, here we look at the entire country.

The odd-numbered regressions of Table 14 show the relationship between change in the percent college educated at the ZIP-code level and the average number of establishments between 2007 and 2011 in five Yelp categories: laundromats, restaurants, cafes, bars and groceries. The even numbered regressions include a control for the initial share of the population with a college degree. In most specifications, there is a negative correlation between the initial share of the population with a college degree and growth in the share of the population with a college degree, and controlling for this initial value strengthens the relationship between the Yelp variable and growth in the number of college educated.

The first two regressions examine the laundromat variable, and show that laundromats between 2007 and 2011 predict growth in the share of college educated over the next five years. Similarly, groceries, bars, cafes and restaurants all predict subsequent increases in the share of the population with a college degree. Over the past decade, neighborhoods that began with more entertainment establishments saw their levels of education increase more swiftly. As suggested by Glaeser, Kolko and Saiz (2001), rising skill levels are associated by increased demand for urban entertainment amenities.

In Table 15, we reverse the order and regress change in the number of Yelp establishments on the initial value of education. The odd-numbered regressions do not control for the initial number of Yelp establishments. The even-numbered regressions include that control.

In this case, initial share of the population with a college degree is an extremely strong predictor of growth in the number of cafes, bars and restaurants. There is little correlation between initial education and growth in the number of laundromats. The correlation with groceries is strong, but less strong than for growth in cafes, bars and restaurants.

In this case, we find striking differences between the odd regressions and the even regressions. If we don't control for initial number of establishments, then education strongly predicts growth in the number of cafes, restaurants, bars and groceries. Once we do control for the initial share, we find that the relationship with education either reverses or becomes small and marginally significant.

The overriding fact in the Yelp data is that the places with more establishments initially added more establishments over the next five years. This may reflect something real about the growth of consumer-oriented establishments in places that had natural strengths, or it may represent something more artificial about the roll-out of Yelp. It may be that Yelp coverage increased gradually and places that had more restaurants overall had more restaurants and in 2011, but also saw more of those restaurants added to Yelp by 2016.

Yelp establishments are solid predictors of growth in education. Education is not a strong predictor of growth in Yelp establishments, when we control for the initial level of establishments. We would not want to attribute too much meaning for these facts, but they are at least compatible with a view that consumption establishments are attracting the more skilled—not just that they follow the location of the more skilled.

VII. Conclusion

In this paper, we illustrate how the local business landscape, as measured by Yelp data, correlates with measures of neighborhood gentrification. Housing price increases at the ZIP-code level are correlated with increases in the number of Yelp establishments, and more strongly related to changes in the number of Yelp reviews. Demographic change is also correlated with the changes in the number of Yelp establishments, both in New York City and in other large cities. For predicting demographics, the number of establishments is a better predictor than the number of reviews.

Changes in the number of Yelp establishments and reviews appear to be reliably correlated to process of gentrification, measured either with housing prices or with increases in the share of the population that is well educated. The predictive fit is far from perfect, but Yelp can provide signs that a neighborhood is changing. Consequently, given the slow appearance of reliable neighborhood level Census data, Yelp may be helpful for measuring up-to-date development in a neighborhood.

We have little to say about causation. The most natural hypothesis to us is that restaurants respond to exogenous changes in neighborhood composition, not that restaurant availability is driving neighborhood change. Yet, it seems true that Yelp establishments from 2007-2011 predict changes in education levels over the next five years, but education from 2007 to 2011 does not predict increases in the number of Yelp establishments, once we control for the initial level of Yelp establishments. Consequently, it is still also possible that Yelp is measuring neighborhood amenities that also help drive neighborhood change.

References

Bogin, Alexander, William Doerner, and William D. Larson (2016). “Local house price growth acceleration” (2016). FHFA Staff Working Paper Series, No. 16-02, Federal Housing Finance Agency, Washington, DC.

Case, Karl E., and Robert J. Shiller (1989). “The efficiency of Home Buyers in Boom and Post-Boom Markets. American Economic Review 79(1): 125-137.

Davidson, Justin, “Is Gentrification all that bad?” New York Magazine. Feb 2, 2014. <http://nymag.com/news/features/gentrification-2014-2/>. Accessed Jan. 4, 2018.

Glaeser, Edward L. and Kim, Hyunjin and Luca, Michael (2017). “Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity.” NBER Working Paper No. w24010.

Glaeser, Edward L., Jed Kolko, and Albert Saiz (2001). "Consumer city," Journal of Economic Geography, Oxford University Press, vol. 1(1): 27-50.

Glaeser, Edward L., and Charles G. Nathanson (2017). "An extrapolative model of house price dynamics," Journal of Financial Economics, vol 126(1), pages 147-170.

Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser (2015). “Do people shape cities, or do cities shape people? The co-evolution of physical, social, and economic change in five major U.S. cities.” NBER Working Paper No. w21620.

Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser (2017). “Computer vision uncovers predictors of physical urban change.” Proceedings of the National Academy of Sciences 114(29): 7571-7576.

Rascoff, Spencer, and Stan Humphries (2015). “Confirmed: Stabucks knows the next hot neighborhood before everybody else does.” Quartz. Jan. 28, 2015. < https://qz.com/334269/what-starbucks-has-done-to-american-home-values/>. Accessed Jan. 4, 2018.

Waldfogel, Joel (2008). "The median voter and the median consumer: Local private goods and population composition." *Journal of Urban Economics* 63(2): 567-582.

Table 1 Summary Statistics

Summary Statistics: ZIP Codes across Cities

	New York City			Boston, Los Angeles, Chicago, San Francisco		
	mean	sd	count	mean	sd	count
Change in percent of college educated (from 2007-2011 to 2012-2016)	2.595	3.933	180	3.382	3.909	165
Change in percent of ages 25 to 34 (from 2007-2011 to 2012-2016)	0.304	3.071	180	1.011	4.074	166
Change in percent white (from 2007-2011 to 2012-2016)	-1.127	5.695	180	-0.259	8.428	166
Percent of college educated in 2007-2011	36.518	21.527	180	40.250	25.348	165
Percent of ages 25 to 34 in 2007-2011	17.882	6.881	180	20.014	8.444	166
Percent white in 2007-2011	48.116	27.013	180	50.532	24.974	166
Change in street score 2007-2014	1.638	0.703	150	-	-	-
Change in the number of winebars (from 2007-2011 to 2012-2016)	5.190	7.931	139	3.693	4.691	133
Change in the number of vegetarian restaurants (from 2007-2011 to 2012-2016)	1.725	2.094	110	1.224	1.525	136
Change in the number of Starbucks (from 2007-2011 to 2012-2016)	1.258	1.738	97	1.153	1.773	128
Change in the number of laundromats (from 2007-2011 to 2012-2016)	1.469	1.404	122	0.954	1.028	101
Change in the number of groceries (from 2007-2011 to 2012-2016)	3.731	3.523	174	1.881	1.858	166
Change in the number of florists (from 2007-2011 to 2012-2016)	1.197	1.478	143	0.997	2.635	148
Change in the number of fastfood (from 2007-2011 to 2012-2016)	3.295	2.772	176	2.852	2.433	165
Change in the number of convenience (from 2007-2011 to 2012-2016)	2.632	2.239	165	1.849	1.701	159
Change in the number of Chinese restaurants (from 2007-2011 to 2012-2016)	3.607	8.188	178	1.398	2.239	162
Change in the number of cafes (from 2007-2011 to 2012-2016)	10.133	12.169	184	8.210	7.341	168
Change in the number of bars (from 2007-2011 to 2012-2016)	9.989	13.913	180	7.299	9.146	162
Change in the number of barbers (from 2007-2011 to 2012-2016)	2.755	2.471	161	2.261	2.397	157
Change in the number of $$$$ restaurants (from 2007-2011 to 2012-2016)	0.360	0.986	185	0.313	0.777	176
Change in the number of restaurants opened (from 2007-2011 to 2012-2016)	0.958	3.269	185	0.109	2.727	176
Change in the number of restaurants (from 2007-2011 to 2012-2016)	36.101	34.715	185	25.718	20.879	176
Change in the number of reviews for winebars (from 2007-2011 to 2012-2016)	302.235	548.439	139	262.460	373.859	133
Change in the number of reviews for vegetarian restaurants (from 2007-2011 to 2012-2016)	87.598	150.347	110	71.235	111.674	136
Change in the number of reviews for Starbucks (from 2007-2011 to 2012-2016)	14.402	15.476	97	15.741	16.565	128
Change in the number of reviews for laundromats (from 2007-2011 to 2012-2016)	4.407	5.966	122	4.016	7.376	101

Change in the number of reviews for groceries (from 2007-2011 to 2012-2016)	32.001	53.499	174	30.087	77.080	166
Change in the number of reviews for florists (from 2007-2011 to 2012-2016)	8.852	14.010	143	13.697	35.396	148
Change in the number of reviews for fastfood (from 2007-2011 to 2012-2016)	41.815	95.815	176	67.476	97.901	165
Change in the number of reviews for convenience (from 2007-2011 to 2012-2016)	7.812	8.579	165	6.681	13.300	159
Change in the number of reviews for Chinese restaurants (from 2007-2011 to 2012-2016)	108.589	311.913	178	88.049	177.104	162
Change in the number of reviews for cafes (from 2007-2011 to 2012-2016)	294.848	529.102	184	393.068	516.407	168
Change in the number of reviews for bars (from 2007-2011 to 2012-2016)	484.387	931.167	180	411.165	668.562	162
Change in the number of reviews for barbers (from 2007-2011 to 2012-2016)	32.588	63.946	161	27.330	40.262	157
Change in the number of reviews for $$$$ restaurants (from 2007-2011 to 2012-2016)	34.541	118.341	185	17.499	51.411	176
Change in the number of reviews for restaurants (from 2007-2011 to 2012-2016)	1524.451	2698.427	185	1444.518	1864.140	176
Number of winebars in 2007-2011	3.521	6.591	139	3.726	4.851	133
Number of vegetarian restaurants in 2007-2011	2.187	3.762	110	2.038	2.443	136
Number of Starbucks in 2007-2011	1.938	2.920	97	2.306	2.373	128
Number of laundromats in 2007-2011	1.767	1.655	122	1.354	1.630	101
Number of groceries in 2007-2011	7.692	6.679	174	10.461	10.297	166
Number of florists in 2007-2011	3.020	3.459	143	4.870	5.115	148
Number of fastfood in 2007-2011	4.667	4.423	176	8.659	6.151	165
Number of convenience in 2007-2011	3.874	3.305	165	4.990	4.558	159
Number of Chinese restaurants in 2007-2011	9.344	11.821	178	7.616	10.506	162
Number of cafes in 2007-2011	9.553	13.231	184	14.337	14.314	168
Number of bars in 2007-2011	14.896	28.031	180	18.968	24.225	162
Number of barbers in 2007-2011	3.626	3.740	161	4.680	3.737	157
Number of $$$$ restaurants in 2007-2011	1.534	4.402	185	1.015	1.991	176
Number of restaurants opened in 2007-2011	11.215	13.693	185	10.272	8.339	176
Number of restaurants in 2007-2011	79.693	97.853	185	88.627	72.714	176
Number of reviews for winebars in 2007-2011	90.445	214.000	139	221.904	388.014	133
Number of reviews for vegetarian restaurants in 2007-2011	48.027	130.732	110	82.837	134.027	136
Number of reviews for Starbucks in 2007-2011	3.734	5.670	97	7.364	9.408	128
Number of reviews for laundromats in 2007-2011	1.126	2.063	122	2.287	7.634	101
Number of reviews for groceries in 2007-2011	23.191	46.083	174	55.254	84.339	166
Number of reviews for florists in 2007-2011	3.698	14.279	143	14.176	21.640	148

Number of reviews for fastfood in 2007-2011	16.953	52.384	176	47.998	73.635	165
Number of reviews for convenience in 2007-2011	4.372	8.071	165	14.123	59.802	159
Number of reviews for Chinese restaurants in 2007-2011	54.898	219.712	178	119.407	267.364	162
Number of reviews for cafes in 2007-2011	95.189	232.859	184	255.510	435.677	168
Number of reviews for bars in 2007-2011	244.324	661.819	180	618.407	1019.271	162
Number of reviews for barbers in 2007-2011	5.768	15.837	161	14.183	23.073	157
Number of reviews for $$$$ restaurants in 2007-2011	30.942	108.883	185	43.145	117.373	176
Number of reviews for restaurants in 2007-2011	813.758	1973.270	185	1947.623	2922.978	176

Summary Statistics: All ZIP codes across US

	mean	sd	count
Change in percent of college educated between 2007-2011 and 2012-2016	1.772	6.871	19495
Percent of college educated in 2007-2011	26.033	16.357	19515
Number of laundromats in 2007-2011	0.751	0.838	2359
Change in the number of laundromats between 2007-2011 and 2012-2016	0.551	0.676	2359
Number of groceries in 2007-2011	2.869	3.485	13374
Change in the number of groceries between 2007-2011 and 2012-2016	1.033	1.388	13374
Number of cafes in 2007-2011	2.644	4.644	14380
Change in the number of cafes between 2007-2011 and 2012-2016	2.368	3.546	14380
Number of bars in 2007-2011	3.813	7.637	14877
Change in the number of bars between 2007-2011 and 2012-2016	2.585	4.622	14877
Number of bars in 2007-2011	3.813	7.637	14877
Change in the number of bars between 2007-2011 and 2012-2016	2.585	4.622	14877
Percent change in Housing Price Index (HPI) between 2012-2016	3.067	6.544	75594
Change in the number of Starbucks between 2012-2016	0.138	0.429	25100

Table 2 **Correlations between Annual Percent Change in HPI and Annual Absolute Change in the Number of Starbucks across ZIP codes (2012-2016)**

	(1)	(2)	(3)	(4)	(5)	(6)
	% Change in HPI	% Change in HPI	% Change in HPI	% Change in HPI	% Change in HPI	% Change in HPI
Yelp Starbucks Growth		0.536***	0.171*	0.539***	0.538***	0.206*
		(0.082)	(0.075)	(0.082)	(0.082)	(0.087)
Yelp Starbucks Growth (lag1)				0.486***	0.486***	0.261**
				(0.086)	(0.086)	(0.086)
Yelp Starbucks Growth (lag2)					0.259***	0.195**
					(0.070)	(0.070)
Yelp Growth in Closed Starbucks						-0.042
						(0.149)
Yelp Starbucks Reviews Growth						0.136***
						(0.007)
Constant	-0.790***	-0.858***	-0.826***	-0.888***	-0.957***	-0.952***
	(0.056)	(0.057)	(0.056)	(0.057)	(0.061)	(0.061)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
ZIP FE	No	No	Yes	No	No	No
Observations	24865	24865	24865	24865	24865	24865
Adjusted R^2	0.238	0.240	0.372	0.241	0.241	0.256

All regressions include a full set of calendar year dummies and cluster standard errors at the ZIP Code level. * p<0.10, ** p<0.05, *** p<0.01.

Table 3 Correlations between Annual Percent Change in HPI and Annual Absolute Change in the Number of Cafes across ZIP codes (2012-2016)

	(1)	(2)	(3)	(4)	(5)	(6)
	% Change in HPI	% Change in HPI	% Change in HPI	% Change in HPI	% Change in HPI	% Change in HPI
Yelp Cafes Growth		0.535***	0.020	0.431***	0.368***	0.250***
		(0.023)	(0.023)	(0.022)	(0.022)	(0.024)
Yelp Cafes Growth (lag1)				0.475***	0.396***	0.277***
				(0.025)	(0.025)	(0.024)
Yelp Cafes Growth (lag2)					0.363***	0.292***
					(0.023)	(0.024)
Yelp Growth in Closed Cafes						-0.077*
						(0.033)
Yelp Cafes Reviews Growth						0.009***
						(0.001)
Constant	-1.233***	-1.523***	-1.231***	-1.631***	-1.789***	-1.679***
	(0.043)	(0.044)	(0.043)	(0.045)	(0.048)	(0.048)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
ZIP FE	No	No	Yes	No	No	No
Observations	59180	59180	59180	59180	59180	59180
Adjusted R^2	0.149	0.157	0.211	0.163	0.167	0.172

All regressions include a full set of calendar year dummies and cluster standard errors at the ZIP Code level. * p<0.10, ** p<0.05, *** p<0.01.

Table 4 **Correlation between Annual Percent Change in HPI and Annual Change in the Number of Establishments by Category across ZIP Codes (2012-2016)**

	(1)	(2)	(3)	(4)
	% Change in HPI	Adjusted R^2	Adjusted R^2 (Year Effects Only)	Obs.
Yelp Growth in laundromats	1.157*** (0.159)	0.249	0.245	11178
Yelp Growth in barbers	1.041*** (0.045)	0.218	0.205	36831
Yelp Growth in convenience stores	0.916*** (0.057)	0.169	0.164	49920
Yelp Growth in florists	0.708*** (0.084)	0.198	0.196	35520
Yelp Growth in winebars	0.672*** (0.046)	0.197	0.190	30499
Yelp Growth in Starbucks	0.536*** (0.082)	0.240	0.238	24865
Yelp Growth in cafes	0.534*** (0.023)	0.157	0.149	59153
Yelp Growth in vegetarian restaurants	0.521*** (0.074)	0.217	0.215	19394
Yelp Growth in fastfood	0.517*** (0.033)	0.158	0.154	54639
Yelp Growth in bars	0.403*** (0.021)	0.155	0.149	60932
Yelp Growth in groceries	0.332*** (0.044)	0.154	0.154	57253
Yelp Growth in $$$$ restaurants	0.256 (0.133)	0.128	0.128	75594
Yelp Growth in restaurants	0.199*** (0.008)	0.135	0.128	75594
Yelp Growth in Chinese restaurants	0.064 (0.038)	0.170	0.170	50047
Yelp Growth in restaurants opened	0.019* (0.007)	0.128	0.128	75594

All regressions include a full set of calendar year dummies and cluster standard errors at the ZIP Code level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 5 **Correlation between Annual Percent Change in HPI and Annual Change in the Number of Yelp Reviews by Category across ZIP Codes (2012-2016)**

	(1)	(2)	(3)	(4)
	% Change in HPI	Adjusted R^2	Adjusted R^2 (Year Effects Only)	Obs.
Yelp Growth in laundromats	0.104^{***} (0.021)	0.247	0.245	11178
Yelp Growth in barbers	0.073^{***} (0.010)	0.218	0.205	36831
Yelp Growth in convenience stores	0.064^{**} (0.021)	0.166	0.164	49920
Yelp Growth in florists	0.091^{***} (0.008)	0.202	0.196	35520
Yelp Growth in winebars	0.012^{***} (0.001)	0.202	0.190	30499
Yelp Growth in Starbucks	0.142^{***} (0.007)	0.255	0.238	24865
Yelp Growth in cafes	0.013^{***} (0.001)	0.165	0.149	59153
Yelp Growth in vegetarian restaurants	0.017^{***} (0.001)	0.224	0.215	19394
Yelp Growth in fastfood	0.037^{***} (0.002)	0.172	0.154	54639
Yelp Growth in bars	0.010^{***} (0.001)	0.164	0.149	60932
Yelp Growth in groceries	0.042^{***} (0.007)	0.159	0.154	57253
Yelp Growth in $$$$ restaurants	0.024^{***} (0.006)	0.128	0.128	75594
Yelp Growth in restaurants	0.006^{***} (0.0002)	0.159	0.128	75594
Yelp Growth in Chinese restaurants	0.027^{***} (0.002)	0.181	0.170	50047

All regressions include a full set of calendar year dummies and cluster standard errors at the ZIP Code level. * p<0.10, ** p<0.05, *** p<0.01.

Table 6 **Correlations between Changes in Demographics and Yelp Number of Establishments between 2007-2011 and 2012-2016 across New York City ZIP Codes**

	(1)	(2)	(3)	
	Change in percent of college educated	Change in percent of ages 25 to 34	Change in percent white	Obs.
Change in the number of groceries	0.352***	0.178*	0.189*	173
	(0.000002)	(0.019)	(0.013)	
Change in the number of laundromats	0.338***	0.200*	0.120	122
	(0.0001)	(0.027)	(0.187)	
Change in the number of cafes	0.319***	0.093	0.084	179
	(0.00001)	(0.216)	(0.264)	
Change in the number of bars	0.313***	0.140	0.114	176
	(0.00002)	(0.064)	(0.132)	
Change in the number of restaurants	0.270***	0.152*	0.098	180
	(0.0003)	(0.041)	(0.191)	
Change in the number of barbers	0.237**	0.197*	0.084	160
	(0.003)	(0.012)	(0.291)	
Change in the number of winebars	0.232**	0.143	0.144	136
	(0.007)	(0.097)	(0.094)	
Change in the number of convenience stores	0.222**	0.079	0.128	162
	(0.004)	(0.320)	(0.104)	
Change in the number of fastfood	0.200**	0.024	0.046	173
	(0.008)	(0.758)	(0.544)	
Change in the number of $$$$ restaurants	0.193**	0.125	0.066	180
	(0.009)	(0.094)	(0.378)	
Change in the number of vegetarian restaurants	0.175	0.067	0.054	108
	(0.069)	(0.490)	(0.580)	
Change in the number of florists	0.173*	0.185*	0.053	142
	(0.039)	(0.028)	(0.534)	
Change in the number of restaurants opened	0.089	0.091	0.168*	180
	(0.237)	(0.226)	(0.024)	
Change in the number of Starbucks	0.067	-0.099	-0.010	95
	(0.522)	(0.338)	(0.923)	
Change in the number of Chinese restaurants	-0.130	0.038	-0.117	176
	(0.084)	(0.613)	(0.123)	

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 7 **Correlations between Changes in Demographics and Yelp Number of Reviews between 2007-2011 and 2012-2016 across New York City ZIP Codes**

	(1)	(2)	(3)	
	Change in percent of college educated	Change in percent of ages 25 to 34	Change in percent white	Obs.
Change in the number of reviews for groceries	0.146	-0.052	0.066	173
	(0.055)	(0.497)	(0.385)	
Change in the number of reviews for laundromats	0.311^{***}	0.195^{*}	0.122	122
	(0.0005)	(0.031)	(0.179)	
Change in the number of reviews for cafes	0.216^{**}	0.117	0.080	179
	(0.004)	(0.119)	(0.289)	
Change in the number of reviews for bars	0.228^{**}	0.140	0.068	176
	(0.002)	(0.064)	(0.371)	
Change in the number of reviews for restaurants	0.203^{**}	0.113	0.055	180
	(0.006)	(0.132)	(0.463)	
Change in the number of reviews for barbers	0.134	0.152	0.020	160
	(0.092)	(0.055)	(0.802)	
Change in the number of reviews for winebars	0.189^{*}	0.137	0.086	136
	(0.027)	(0.112)	(0.318)	
Change in the number of reviews for convenience stores	0.154	0.095	0.102	162
	(0.050)	(0.231)	(0.198)	
Change in the number of reviews for fastfood	0.121	-0.016	0.013	173
	(0.114)	(0.835)	(0.862)	
Change in the number of reviews for $$$$ restaurants	0.142	0.105	0.016	180
	(0.057)	(0.160)	(0.829)	
Change in the number of reviews for vegetarian restaurants	0.085	0.049	0.062	108
	(0.384)	(0.614)	(0.523)	
Change in the number of reviews for florists	0.046	0.079	-0.013	142
	(0.585)	(0.350)	(0.882)	
Change in the number of reviews for Starbucks	0.032	0.078	0.034	95
	(0.760)	(0.454)	(0.744)	
Change in the number of reviews for Chinese restaurants	0.035	0.064	0.004	176
	(0.643)	(0.397)	(0.953)	

* $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 8 **Variation Explained in Change in Percent of College Educated between 2007-2011 and 2012-2016 by Yelp Establishment Categories across NYC ZIP codes**

	(1)	(2)
	Change in percent of college educated	Change in percent of college educated
Change in the number of groceries	0.402***	0.312**
	(0.110)	(0.106)
Change in the number of cafes	0.121*	0.172**
	(0.054)	(0.057)
Change in the number of bars	0.083	0.072
	(0.054)	(0.047)
Change in the number of restaurants	-0.066*	-0.083**
	(0.031)	(0.028)
Percent of college educated in 2007-2011		-0.109***
		(0.020)
Percent of ages 25 to 34 in 2007-2011		0.315***
		(0.072)
Percent white in 2007-2011		0.029**
		(0.010)
Constant	1.475**	-0.933
	(0.496)	(1.080)
Observations	170	170
Adjusted R^2	0.162	0.311

All regressions include robust standard errors. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Table 9 **Correlations between Changes in Demographics and Yelp Number of Establishments between 2007-2011 and 2012-2016 across Boston, Los Angeles, Chicago, and San Francisco ZIP Codes**

	(1)	(2)	(3)	
	Change in percent of college educated	Change in percent of ages 25 to 34	Change in percent white	Obs.
Change in the number of florists	0.322***	0.219**	0.124	146
	(0.00007)	(0.008)	(0.135)	
Change in the number of vegetarian restaurants	0.227**	0.052	-0.139	131
	(0.009)	(0.556)	(0.114)	
Change in the number of cafes	0.221**	0.143	-0.012	161
	(0.005)	(0.071)	(0.880)	
Change in the number of restaurants	0.217**	0.110	0.054	165
	(0.005)	(0.161)	(0.487)	
Change in the number of bars	0.202*	0.170*	0.041	159
	(0.011)	(0.032)	(0.612)	
Change in the number of winebars	0.196*	0.101	-0.015	130
	(0.025)	(0.255)	(0.867)	
Change in the number of barbers	0.156	0.109	0.029	155
	(0.052)	(0.178)	(0.723)	
Change in the number of convenience stores	0.102	0.003	0.069	154
	(0.206)	(0.967)	(0.392)	
Change in the number of restaurants opened	0.091	-0.011	-0.035	165
	(0.244)	(0.886)	(0.654)	
Change in the number of Starbucks	0.090	-0.041	-0.013	125
	(0.318)	(0.646)	(0.884)	
Change in the number of groceries	0.039	0.019	0.008	163
	(0.621)	(0.813)	(0.917)	
Change in the number of fastfood	0.039	-0.023	0.117	159
	(0.623)	(0.773)	(0.141)	
Change in the number of $$$$ restaurants	0.036	-0.052	0.103	165
	(0.645)	(0.508)	(0.387)	
Change in the number of Chinese restaurants	0.020	0.085	-0.025	157
	(0.800)	(0.289)	(0.760)	
Change in the number of laundromats	-0.033	0.067	-0.061	100
	(0.744)	(0.505)	(0.544)	

ZIP codes for each city were based on ZIP code designations in County Business Patterns. * p<0.10, ** p<0.05, *** p<0.01.

Table 10 **Correlations between Changes in Demographics and Yelp Number of Reviews between 2007-2011 and 2012-2016 across Boston, Los Angeles, Chicago, and San Francisco ZIP Codes**

	(1)	(2)	(3)	
	Change in percent of college educated	Change in percent of ages 25 to 34	Change in percent white	Obs.
Change in the number of reviews for florists	0.235**	0.076	0.055	146
	(0.004)	(0.362)	(0.514)	
Change in the number of reviews for vegetarian restaurants	0.206*	0.130	-0.086	131
	(0.018)	(0.138)	(0.328)	
Change in the number of reviews for cafes	0.160*	0.233**	0.011	161
	(0.043)	(0.003)	(0.894)	
Change in the number of reviews for rests	0.199*	0.215**	0.035	165
	(0.010)	(0.006)	(0.660)	
Change in the number of reviews for bars	0.184*	0.227**	0.014	159
	(0.020)	(0.004)	(0.861)	
Change in the number of reviews for winebars	0.149	0.143	-0.020	130
	(0.091)	(0.104)	(0.819)	
Change in the number of reviews for barbers	0.218**	0.204*	0.002	155
	(0.007)	(0.011)	(0.980)	
Change in the number of reviews for convenience stores	0.035	0.002	0.089	154
	(0.669)	(0.980)	(0.273)	
Change in the number of reviews for Starbucks	0.115	-0.008	-0.048	125
	(0.203)	(0.933)	(0.595)	
Change in the number of reviews for groceries	0.178*	0.063	0.011	163
	(0.023)	(0.424)	(0.888)	
Change in the number of reviews for fastfood	0.078	0.045	0.019	159
	(0.330)	(0.577)	(0.813)	
Change in the number of reviews for $$$$ restaurants	-0.039	0.029	0.015	165
	(0.623)	(0.709)	(0.846)	
Change in the number of reviews for Chinese restaurants	0.076	0.105	0.010	157
	(0.346)	(0.189)	(0.902)	
Change in the number of reviews for laundromats	-0.064	0.071	-0.111	100
	(0.526)	(0.483)	(0.270)	

ZIP codes for each city were based on ZIP code designations in County Business Patterns. * p<0.10, ** p<0.05, *** p<0.01.

Table 11 Correlations between Changes in Yelp Number of Establishments (between 2007-2011 and 2012-2016) and Street Score (2007-2014) across NYC ZIP Codes

	(1)	
	Change in street score 2007-2014	Obs.
Change in the number of vegetarian restaurants	0.372***	100
	(0.0001)	
Change in the number of Starbucks	0.355***	88
	(0.001)	
Change in the number of wine bars	0.339***	119
	(0.0002)	
Change in the number of bars	0.327***	147
	(0.00005)	
Change in the number of cafes	0.318***	150
	(0.00007)	
Change in the number of barbers	0.316***	140
	(0.0001)	
Change in the number of florists	0.290***	127
	(0.001)	
Change in the number of restaurants	0.275***	150
	(0.001)	
Change in the number of fastfood	0.270***	148
	(0.001)	
Change in the number of convenience stores	0.208*	141
	(0.014)	
Change in the number of $$$$ restaurants	0.148	150
	(0.070)	
Change in the number of groceries	0.103	146
	(0.218)	
Change in the number of laundromats	0.034	109
	(0.729)	
Change in the number of Chinese restaurants	0.001	149
	(0.991)	
Change in the number of restaurants opened	-0.097	150
	(0.238)	

* p<0.10, ** p<0.05, *** p<0.01.

Table 12 Correlations between Annual Percent Change in HPI and Lagged Annual Absolute Change in the Number of Yelp Starbucks across ZIP Codes (2012-2016)

	(1)	(2)	(3)
	% Change in HPI	% Change in HPI	% Change in HPI
Yelp Starbucks Growth (lag1)	0.482***		0.291***
	(0.087)		(0.079)
Yelp Starbucks Growth (lag2)	0.260***		0.155*
	(0.070)		(0.066)
% Change in HPI (lag1)		0.324***	0.323***
		(0.013)	(0.013)
% Change in HPI (lag2)		0.076***	0.076***
		(0.011)	(0.011)
Constant	-0.890***	0.900***	0.835***
	(0.060)	(0.065)	(0.068)
Year FE	Yes	Yes	Yes
Observations	24865	24819	24819
Adjusted R^2	0.239	0.332	0.333

All regressions include a full set of calendar year dummies and cluster standard errors at the ZIP Code level. * p<0.10, ** p<0.05, *** p<0.01.

Table 13 Correlations between Annual Absolute Change in the Number of Yelp Starbucks and Lagged Annual Percent Change in HPI across ZIP Codes (2012-2016)

	(1)	(2)	(3)
	Yelp Starbucks Growth	Yelp Starbucks Growth	Yelp Starbucks Growth
% Change in HPI (lag1)	0.002^{**}		0.002^{**}
	(0.001)		(0.001)
% Change in HPI (lag2)	0.001		0.001
	(0.001)		(0.001)
Yelp Starbucks Growth (lag1)		-0.007	-0.008
		(0.011)	(0.011)
Yelp Starbucks Growth (lag2)		0.004	0.003
		(0.010)	(0.010)
Constant	0.135^{***}	0.126^{***}	0.135^{***}
	(0.006)	(0.006)	(0.006)
Year FE	Yes	Yes	Yes
Observations	24907	24907	24907
Adjusted R^2	0.026	0.026	0.026

All regressions include a full set of calendar year dummies and cluster standard errors at the ZIP Code level. * p<0.10, ** p<0.05, *** p<0.01.

Table 14 Correlations between Change in Percent of College Educated between 2007-2011 and 2012-2016 and Yelp Number of Establishments in 2007-2011 across ZIP Codes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated	Change in percent of college educated
Number of laundromats in 2007-2011	0.364***	0.325***								
	(0.085)	(0.086)								
Number of groceries in 2007-2011			0.063***	0.078***						
			(0.009)	(0.011)						
Number of cafes in 2007-2011					0.090***	0.152***				
					(0.008)	(0.013)				
Number of bars in 2007-2011							0.056***	0.090***		
							(0.005)	(0.008)		
Number of restaurants in 2007-2011									0.011***	0.026***
									(0.001)	(0.002)
Percent of college educated in 2007-2011		0.014***		-0.017***		-0.050***		-0.055***		-0.080***
		(0.004)		(0.004)		(0.007)		(0.007)		(0.007)
Constant	1.812***	1.384***	1.672***	2.117***	1.603***	2.891***	1.674***	3.091***	1.574***	3.363***
	(0.090)	(0.137)	(0.061)	(0.106)	(0.063)	(0.162)	(0.060)	(0.163)	(0.065)	(0.147)
Observations	2348	2348	13141	13141	13851	13851	14382	14382	19495	19495
Adjusted R^2	0.009	0.013	0.002	0.005	0.005	0.022	0.005	0.025	0.002	0.033

All regressions include robust standard errors. * p<0.10, ** p<0.05, *** p<0.01.

Table 15 **Correlations between Change in Yelp Number of Establishments between 2007-2011 and 2012-2016 and Change in Percent of College Educated in 2007-2011 across ZIP Codes**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Change in the number of laundromats	Change in the number of laundromats	Change in the number of groceries	Change in the number of groceries	Change in the number of cafes	Change in the number of cafes	Change in the number of bars	Change in the number of bars	Change in the number of restaurants	Change in the number of restaurants
Percent of college educated in 2007-2011	-0.001	-0.002	0.011***	0.003***	0.060***	0.002	0.066***	-0.002	0.219***	-0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)	(0.003)	(0.002)	(0.007)	(0.004)
Number of Yelp establishment type in 2007-2011		0.143***		0.221***		0.593***		0.499***		0.380***
		(0.035)		(0.008)		(0.019)		(0.019)		(0.008)
Constant	0.572***	0.497***	0.721***	0.329***	0.701***	0.788***	0.776***	0.786***	3.296***	2.358***
	(0.032)	(0.036)	(0.022)	(0.022)	(0.062)	(0.037)	(0.073)	(0.047)	(0.164)	(0.107)
Observations	2349	2349	13147	13147	13863	13863	14388	14388	19515	19515
Adjusted R^2	-0.0002	0.030	0.019	0.317	0.079	0.615	0.055	0.684	0.073	0.712

All regressions include robust standard errors. * p<0.10, ** p<0.05, *** p<0.01.