Modeling Susceptibility to Gentrification

Submitted as partial fulfilment of a Master of Science in Geospatial Technologies

Kevin G. Bogue

University of Washington Tacoma, 1900 Commerce Street Tacoma, WA 98402
Email: kbogue13@uw.edu

Abstract

Over the past three decades awareness around gentrification within America's urban centers has been increasing. As consciousness has risen amongst the general population, gentrification has become a much more widely debated subject which has spurred new focus into a long-established subject of study. To date, much of the gentrification research has examined the impacts on individual communities or has retroactively probed an area that has gentrified in order to determine latent origins. By comparison there has been relatively little research examining whether the process of gentrification can be preemptively predicted. Moreover, the predictive studies that have been undertaken tend to focus on a specific locality as opposed to an approach that seeks to encompass larger areas and trends. This research will examine whether the process of gentrification can be identified by quantitative indicators and explores the possibility of creating a technological framework to assess susceptibility to gentrification based on existing demographic factors within the United States.

1. Introduction

Thirty years ago, the term "gentrification" was largely unfamiliar to the American populace, but today its increased usage in the urban lexicon has fostered a growing awareness (Lees & Ferreri 2016). Despite this increased consciousness there is no singular all-encompassing definition of gentrification and this creates challenges not only for those seeking to better understand its processes, but for those attempting to mobilize a response to them (Hackworth, 2007; Lees et al., 2008). Many definitions focus on the effects, but there is no consensus about which effects are the defining ones as they can often vary from place to place (Atkinson 2003; Ellen & O'Regan 2011). Some definitions emphasize the displacement of existing residents (Atkinson 2000; Newman & Wyly 2006; Peck, Siemiatycki & Wyly 2014). Some definitions point to the change in a neighborhood's character ("Cupcake Gentrification" 2009; Kennedy and Leonard 2001; Zukin *et al.* 2009). Still others see changes in property values as an essential element of gentrification (Smith 1979; 1987; Zukin 1987).

The impacts of gentrification are arguably even more complicated than its causes and further obscuring its study is that there is no consensus as to gentrification's overall positive or negative effects. It is undeniably bad for some, but it is also supremely beneficial to others and sorting out winners and losers often depends on questions that are problematic to ask and difficult to answer such as who moves out, where they go, and why (Ellen & O'Regan 2011; Freeman 2009)? Proponents of gentrification usually focus on the aggregate effects which seem to be positive, while opponents typically focus on the distributional impact and point out how the negative effects are likely to fall disproportionately on the poor (Holland 2016). It is possible that both are correct, and should any resolution ever be reached it will undoubtedly have a profound impact on urban centers, those who live there, and the researchers that study them.

This research was conducted, not in an attempt to declare one side the victor, but to determine whether susceptibility to the gentrification process can be quantitatively identified. While the quantitative factors used can and should be fervently debated, this research is not an attempt to end the debate as a susceptibility analysis will typically identify the most underprivileged areas be as the most susceptible to gentrification which does not necessarily qualify them as the most likely to gentrify. The aim of this research is to provide a technological framework for assessing gentrification susceptibility that is easily repeatable and upon which further more localized analysis can be undertaken. If susceptibility to gentrification can be quantitatively identified this research could serve as a stepping stone to the quantitative prediction of the likelihood of future gentrification.

2. Approaches to Identifying Gentrification

As previously mentioned there is no consensus definition of gentrification but for decades scholars have been examining the innumerable ways in which revitalization of urban cores and central business districts impact the lives of established residents and entrenched communities. The term "gentrification" itself originated in the 1960's and has widely been attributed to sociologist Ruth Glass (*inter-alia* Butler 2003; Carmona 2003; Lees& Ferreri 2016) who noted that,

> "One by one, many of the working class quarters of London have been invaded by the middle classes – upper and lower…. Once this process of 'gentrification' starts in a district it goes on rapidly until all or most of the original working class occupiers are displaced and the social character of the district is changed" (Glass 1964: 18).

Despite the lack of an all-encompassing definition, most research undertaken on the topic has been conducted under a broad understanding similar to that put forth by Glass – that gentrification is a process by which higher-income households displace the lower-income households of a neighborhood, thus changing the neighborhood's essential character. In this way, this broad understanding of gentrification without specific definition bears resemblance to attempts to define pornography, about which United States Supreme Court Justice Potter Stewart infamously wrote, we "shall not today attempt further to define the kinds of material (we) understand to be embraced within that shorthand description, and perhaps (we) could never succeed in intelligibly doing so. But (we) know it when (we) see it..." (Jacobellis v. Ohio 1964). Similarly, scholars may never be able to intelligibly define all the facets necessary to produce an all-encompassing shorthand definition of gentrification, but research surrounding it has progressed based on the broad understanding presented by Glass because, like pornography, the process of gentrification is apparent when it is observed whether or not it can be specifically defined. This has led to a variety of approaches to better understand where and why the gentrification process takes place. While it would be impossible to cover every publication, what follows is a survey of some of the significant work in both the qualitative and quantitative methodological approaches to identifying gentrification.

2.1 Qualitative Approaches

When identifying gentrified neighborhoods, qualitative studies typically seek to identify a single or a small group of neighborhoods that have recently gentrified and attempt to deconstruct the change to identify the latent characteristics that preempted neighborhood change. A common way gentrification is examined in this way is through displacement; examples of which can be found in Newman & Wyly (2006) who collected qualitative survey data from New York residents in an attempt to estimate involuntary displacement or in Peck, Siemiatycki & Wyly's (2014) investigation into "Vancouverism" and its perpetuation of increasingly systemic forms of gentrification. While both examples undeniably have a quantitative aspect to their approach in that total displaced people were tabulated, the object of the research itself was to examine the underlying qualitative reasons for residents' displacement.

Another qualitative analysis of gentrification is Cupcake Gentrification which postulates that culinary markers of neighborhood transformation change over time and that, "...cupcake shops can provide a more accurate and timelier guide to the frontiers of urban gentrification than traditional demographic and real estate data sets" ("Cupcake Gentrification" 2009). The point is that more so than any particular culinary fad, once a trendy shop opens in a traditionally necessitous neighborhood it is indicative of the beginnings of wholesale neighborhood change.

Qualitative measures of gentrification have traditionally been used to retroactively understand how or why gentrification took place. To do so they require researchers to actively engage within the community of study. Because the goal of my study is to generally assess the susceptibility to gentrification anywhere in the United States, the highly localized form of research undertaken to form qualitative study is not feasible. As such my research will be focused on quantitative assessments of gentrification because, although it will lack the highly localized viewpoint that can be gleaned through rigorous qualitative study, the large-scale datasets that are necessary to leverage to conduct analysis nationwide simply cannot be obtained through qualitative measures.

2.2 Quantitative Approaches

"(Q)uantitative studies (have) typically used a threshold strategy where neighbourhoods were identified as gentrifiable if they featured a particular characteristic or characteristics at the beginning of a decade and gentrified if they experienced a change in the characteristic or characteristics at a later time" (Barton 2014: 93). However, "(g)entrification is notoriously difficult to measure, and results are sensitive to the indicators chosen, the time periods under investigation, and thresholds used to distinguish among neighborhoods" (Wyly & Hammel 1999: 726). Additionally, even within the overarching umbrella of quantitative analysis, gentrification studies are extremely varied. There are numerous quantitative avenues, but some common lenses are to examine market factors or demographic factors as they pertain to gentrification.

A long-established market factor quantitative study of gentrification is the Rent Gap Theory – the disparity between the current rental income of a property and the potentially achievable rental income – to identify an area's propensity to gentrify (Smith 1979). "The whole point of the rent gap theory is not that gentrification occurs in some deterministic fashion where housing costs are lowest...but that it is most likely to occur in areas experiencing a sufficiently large gap between actual and potential land values" (Smith 1987: 464). "Gentrification occurs when the gap is wide enough that developers can purchase shells cheaply, can pay the builders' costs and profit for rehabilitation, can pay interest on mortgage and construction loans, and can then sell the end product for a sale price that leaves a satisfactory return to the developer" (Smith 1979: 545). By focusing on the flow or potential flow of capital, the Rent Gap Theory seeks to tie gentrification directly to a city or neighborhood's ability to attract productive capital.

Demographic factors are a compliment to the market factor theories and often focus on the demographic changes that contribute to gentrification. Often these types of quantitative analyses are performed using data collected from the United States Census Bureau to examine the change in census tract data and identify areas that have gentrified. In this type of analysis, researchers set parameters for what it means for a neighborhood to be either gentrifiable or non-gentrifiable and data is examined over a specified time period. Using methods similar to these, "…a (census) tract is designated as gentrifying if its classification switches from gentrifiable to non-gentrifiable during the time-period being analysed" (Bostic & Martin 2003: 2431).

A subset of quantitative gentrification analysis is the model-based approach. Three common types of modeling are the index, spatial statistics, and agent-based approaches. Using the index method to quantify gentrification typically involves the use of demographics to construct a weighted index based on each entry's subsequent positive or negative role in the likelihood of an area to gentrify (Chapple 2009). The spatial statistics method incorporates the index method and expands on it through the measure of spatial autocorrelation — whether characteristics tend to vary in contiguous areas and whether that variation is statistically significant (Eckerd 2011). Finally, some of the most complex quantitative analyses of gentrification are attributable to the agent-based modeling (ABM) approach. ABMs have traditionally been used to model from the bottom up. That is to say that this type of approach to studying gentrification relies on a model that employs simple decision-making rules to simulate interactions on a small scale between numerous actors that, combined, have a profound impact on the larger scale system (Jackson, Forest & Sengupta 2008; O'Sullivan 2002 2009).

The aforementioned approaches into the research of gentrification are by no means exhaustive but serve to illustrate that there is no single way to define or measure such a complex process. Though both the qualitative and quantitative research on gentrification is robust and varied, no method is above reproach. Qualitative approaches, because they generally require a high level of community engagement, have typically not been easily repeatable as the cost of acquiring qualitative data can be prohibitive. Similarly, quantitative approaches, because they have generally become increasingly technologically complex, require a level of technical skill on the part of the researcher that may be restrictive. Additionally, the methods deployed by both qualitative and quantitative research may not translate from the location of study to a different or larger area. By comparison there have been few attempts to produce an easily repeatable model to assess susceptibility to gentrification on a large scale that also has a relatively low technical bar for access. I propose that writing an open source script utilizing the Python programming language that autonomously retrieves and combines demographic indicators into an index that is repeatable anywhere in the United States may provide a more accessible avenue through which susceptibility to the gentrification process can be quantifiably identified by individuals and agencies that may not have the technical prowess or resources to conduct or commission an analysis of their own.

3. Identifying Areas Most Susceptible to Gentrification

To determine susceptibility to gentrification, this study gathered data from the United States Census Bureau's 2016 American Community Survey 5-year estimates at the block group level of enumeration. Block groups are statistical divisions of census tracts and generally contain between 600 and 3,000 people and never cross state, county, or census tract boundaries. The block group is the smallest geographical unit for which the US Census Bureau publishes sample data ("2010

Geographic Terms and Concepts - Block Groups" 2012). The block group level of enumeration was chosen due to its small size both in terms of geography and population to minimize data generalization to the greatest feasible extent.

As previously discussed, there are innumerable ways in which gentrification can be identified and assessed. However, the data that the US Census Bureau collects at the block group level is limited. Therefore, when assessing susceptibility, it was important to choose not only indicators broad enough to be widely available, but specific enough to provide insight and available at the block group level of enumeration. Given these constraints, the following indicators were chosen to assess gentrification susceptibility: household median income; rent as a percentage of household income; housing value; percentage of non-Hispanic white population; and educational attainment. The justification for each indicator is discussed below.

3.1 Median Income

Areas with lower-income are vulnerable to displacement resulting from redevelopment projects or rising rents (Freeman & Braconi 2004). Median income is arguably the most significant indicator of susceptibility to gentrification and is therefore a reasonable starting point for a demographics-based analysis. Per the general understanding of gentrification – a process by which higher income households displace the lower-income households of a neighborhood – areas with high median income are more likely to have previously gentrified whereas areas with lower median income are likely to be more susceptible to gentrification.

3.2 Rent as a Percentage of Household Income

"As reinvestment occurs and property values rise, the potential for different forms of indirect displacement rises. Able to command higher rents on the market, landlords will raise rents to the extent permitted by law, increasing tenant turnover. While these increases may impact any tenant not residing in permanently affordable housing, they are most likely to displace those already paying a disproportionate share of their income for rent…" (Chapple 2009: 8). Typically, renters have a higher mobility than homeowners, thus assessing rent as a percentage of median income can provide insight into the likelihood of future neighborhood change with areas where the rent percentage of household income is higher being more likely to have renters become displaced than areas where the rent percentage is lower.

3.3 Housing Value

Housing value indicates susceptibility as there is a tendency for higher income residents to migrate into the less affluent neighborhoods that border more affluent neighborhoods. The in-migration of the more affluent residents into these border neighborhoods will bid up prices in those neighborhoods causing the original residents to migrate out (Guerrieri, Hurst, & Hartley 2013). By assessing housing value, it may be possible to identify low housing value areas that research indicates may be attractive for migration of more affluent populations.

3.4 Percentage of non-Hispanic White Population

The presence of a large population of non-white individuals indicates a greater susceptibility to gentrification. Simply put, "(t)he more non-Hispanic whites are in the area, the less likely it is to gentrify: the most susceptible areas are those where the majority is minorities" (Chapple 2009: 7). Assessing the percentage of non-Hispanic white population allows for assessment of gentrification susceptibility as areas with a higher percentage of non-Hispanic whites are more likely to have previously gentrified than areas with a lower percentage.

3.5 Educational Attainment

Educational opportunity and attainment have been recognized to be class-related with children from lower-class backgrounds performing less well on average than those from higher backgrounds (Butler & Hamnett 2007). There has long been a correlation between a college degrees and earnings, therefore measuring educational attainment as the percentage of the population with a Bachelor's degree or higher, gives another indication of an area's susceptibility to gentrification with areas with a higher percentage of degree holders being less susceptible to gentrification in the future than areas with a lower percentage.

4. Automating the Gentrification Susceptibility Analysis

A script was developed using the Python programming language to automate the demographic analysis model. The script is hosted online[1] and free to download. Once the user runs the script in their Python integrated development environment (IDE), the script autonomously opens the Google Chrome web browser and navigates to the United States Census Bureau's American FactFinder website[2], asks the user for State and County input and downloads the most recent American Community Survey 5-year estimate demographic tables for median income, rent as a percentage of household income, housing value, percentage of non-Hispanic white population, and educational attainment pertaining to the user's chosen State and County. After initiating the download of the demographic data, the script navigates to the US Census Bureau's TIGER/Line Shapefiles website[3] to download the geographic block group data that corresponds to the user's previously defined State.

Using the tabular demographic data, the script divides the data range for each of the indicators into five classes based on percentile. Depending on the class in which a data value resides, a score of 1-5 is assigned for each indicator based on that indicator's impact on susceptibility to gentrification for each block group. A score of '1' indicates a low susceptibility for a particular indicator and '5' indicates high susceptibility with scores of '2', '3', or '4' falling between. A score of '0' indicates that there was no data for that indicator in a certain block group. The script totals the scores for the indicators in each block group and calculates a final average score using only indicators for which there are non-zero values.

Finally, the script joins the processed tabular demographic data to the geographic data based on block group number, retains only the block groups within the county defined by the user and outputs a shapefile for use in the user's preferred visualization or Geographic Information Systems (GIS) software.

5. Methods

The methods deployed in this research were quantitative in nature and include the formulation and computation of a scored index. All analysis to construct the model was conducted using

[1] Script hosted at: https://github.com/kbogue13/gentrificationsusceptibility/releases

[2] American FactFinder website: https://factfinder.census.gov/faces/nav/jsf/pages/download_center.xhtml

[3] TIGER/Line Shapefiles website: https://www.census.gov/geo/maps-data/data/tiger.html

demographic data obtained from the United States Census Bureau's 2016 American Community Survey 5-year estimates.

5.1 Calculation of Median Income

Median income per block group was calculated using Census table B19013: Median Income in the Past 12 Months (in 2016 Inflation-Adjusted Dollars). The table was processed to replace all non-numeric values with numeric values. For instance, a block group with median income below $10,000 was initially represented in the table as "10,000-" but after processing is represented as "10000".

The data was then divided into five classes where each class received 20-percent of the data range. The classes were assigned a 'score' of 1-5 based on median income's impact on gentrification susceptibility. A score of '1' indicates median income at or above the 80^{th} percentile of the data range and a low susceptibility to gentrification for a particular block group, while a score of '5' indicates median income below the 20^{th} percentile of the data range and a high susceptibility to gentrification. Scores of '2', '3', or '4' represent median income at or above the 60th, 40th, or 20th percentile respectively. A score of '0' indicates that there was no data for median income for a block group. Sample data can be viewed in Table 1 under the columns 'MI' and 'MI_Score'.

5.2 Calculation of Rent as a Percentage of Household Income

Rent as a percentage of household income was calculated using Census table B25071: Median Gross Rent as a Percentage of Household Income in the Past 12 Months (Dollars). The table was first processed to replace all non-numeric values with numeric values.

As with median income, the data was divided into five classes where each class received 20-percent of the data range. The classes were assigned a 'score' of 1-5 based on the rent percentage of household income. A score of '1' indicates a rent percentage below the 20^{th} percentile of the data range and a low susceptibility to gentrification, while a score of '5' indicates a rent percentage at or above the 80^{th} percentile and a high susceptibility to gentrification. Scores of '2', '3', or '4' represent rent percentages at or above the 20th, 40th, or 60th percentile respectively. A score of '0' indicates that there was no data for rent as a percentage of median income in a block group. Sample data can be viewed in Table 1 under the columns 'RentPct' and 'Rent_Score'.

5.3 Calculation of Housing Value

Housing value was derived from Census table B25077: Median Value (Dollars). The table first was processed to replace all non-numeric values with numeric values.

As with median income and rent percentage, the housing value data was divided into five classes where each class received 20-percent of the data range. For housing value, a score of '1' indicates housing value at or above the 80^{th} percentile of the data range and a low susceptibility to gentrification, while a score of '5' indicates data at or below the 20^{th} percentile and a high susceptibility to gentrification. Scores of '2', '3', or '4' represent housing values at or above the 60th, 40th, or 20th percentile respectively. A score of '0' indicates that there was no data for housing value in a block group. Sample data can be viewed in Table 1 under the columns 'HouseVal' and 'HV_Score'.

5.4 Calculation of non-Hispanic White Population

Percentage of non-Hispanic white population was calculated using Census table B02001: Race. The pre-processed data does not contain a percentage but does contain a count for the estimated total population as well as counts for estimates of the populations of the Census-defined ethnic groups within each block group. The percentage of non-Hispanic white population was calculated by dividing the estimated count of non-Hispanic white population by the total estimated population for each block group.

As with previous indicators, the percentage of non-Hispanic white population was divided into five classes where each class received 20-percent of the data range. For non-Hispanic white population, a score of '1' represents the block groups with a percentage of non-Hispanic white population at or above the 80^{th} percentile of the data range and a low susceptibility to future gentrification, while a score of '5' represents block groups below the 20^{th} percentile of percentage of non-Hispanic white population and a high susceptibility to gentrification. Scores of '2', '3', or '4' represent a non-Hispanic white population percentage at or above the 60th, 40th, or 20th percentile respectively. A score of '0' represents that there was no data for non-Hispanic white population in a block group. Sample data can be viewed in Table 1 under the columns 'Total Pop', 'PctWhite' and 'Score'.

5.5 Calculation of Educational Attainment

This analysis is specifically concerned with the percentage of the population with a Bachelor's degree or higher and was derived from Census table B15003: Educational Attainment for the Population 25 Years and Over. The percentage of the population with a Bachelor's degree or higher was calculated by totaling the estimated counts for population with a Bachelor's degree, Master's degree, Professional degree, and Doctorate degree in each block group then dividing by the estimated total population for each block group that was included in Census table B02001: Race, which was previously used in the calculation of the percentage of non-Hispanic white population.

As with the previous indicators, the percentage of the population with a Bachelor's degree or higher was divided into five classes where each class received 20-percent of the data range. For educational attainment, a score of '1' indicates a block group with a percentage of the population with a Bachelor's degree or higher at or above the 80^{th} percentile of the data range and a low susceptibility to gentrification, while a score of '5' represents a block group below the 20^{th} percentile of the percentage of the population with a Bachelor's degree or higher and a high susceptibility to gentrification. Scores of '2', '3', or '4' represent median income at or above the 60th, 40th, or 20th percentile respectively. A score of '0' represents that there was no data for population with a Bachelor's degree or higher in a block group. Sample data can be viewed in Table 1 under the columns 'CollegeDegree', 'DegreePct' and 'Degree_Score'.

5.6 Index Calculation & Final Output

The scored data for median income, rent as a percentage of household income, housing value, percentage of non-Hispanic white population, and educational attainment for each block group were added together to produce a total score and divided by the number of indicators in the index. However, not all block groups have data for all five indicators. Therefore, merely totaling the scores and dividing by five distorts those block groups that do not have a full dataset. Block groups that do not contain a full dataset are also why the scores are unweighted, as there is no way to know which, if any, indicators will be missing data in a block group. Therefore, attributing a weight to any of the indicators could potentially distort the data. The final index calculation totals the scores then divides the total score by the number of indicators that contain non-zero values to produce an average score for each block group with a range of 1-5 where a score of ‘1’ represents block groups with the least susceptibility to gentrification and a score of ‘5’ represents block groups with the highest susceptibility with scores in the ‘2’, ‘3’, and ‘4’ range falling between. Sample data can be viewed in Table 1 under the columns ‘TotalScore’ and ‘AvgScore’.

After the index is calculated, the Python script joins the processed tabular demographic index to the previously downloaded user-defined State geographic data based on block group number, retains only the block groups pertaining to the user-defined County and outputs a shapefile for use in the user’s preferred visualization or Geographic Information Systems (GIS) software (Figures 1 & 2)

Table 1. Sample indexed data and calculated gentrification susceptibility average score

BlockGroup	TotalPop	PctWhite	Score	MI	MI_Score	RentPct	Rent_Score	HouseVal	HV_Score	CollegeDegree	DegreePct	Degree_Score	TotalScore	AvgScore
530530612006	991	0.78506559	3	64464	3	21.3	1	196200	4	165	0.166498486	3	14	2.8
530530613001	980	0.813265306	3	20977	5	43.5	5	241200	3	142	0.144897959	3	19	3.8
530530613002	997	0.849548646	2	48000	4	28.8	3	175000	4	229	0.229689067	2	15	3
530530613003	795	0.558490566	5	45052	4	28.8	3	157600	5	113	0.142138365	3	20	4
530530613004	871	0.464982778	5	52344	4	37.2	4	142200	5	152	0.174512055	2	20	4
530530613005	707	0.47241867	5	35714	5	38.9	5	155800	5	83	0.117397454	4	24	4.8
530530613006	852	0.475352113	5	0	0	46.2	5	190000	4	127	0.149061033	3	17	4.25
530530614001	992	0.664314516	4	15833	5	44.9	5	0	0	60	0.060483871	5	19	4.75
530530614002	1448	0.625	4	18346	5	34.3	4	0	0	104	0.071823204	5	18	4.5
530530614003	1158	0.444732297	5	17264	5	34.1	4	0	0	79	0.068221071	5	19	4.75
530530615001	1787	0.902070509	2	48075	4	23.2	2	392200	1	682	0.381645215	1	10	2
530530615002	1244	0.745176849	3	40038	5	41.8	5	334200	1	468	0.376205788	1	15	3
530530615003	951	0.811777077	3	39226	5	27.6	2	0	0	207	0.217665615	2	12	3
530530615004	883	0.868629672	2	48021	4	28	2	0	0	187	0.211778029	2	10	2.5
530530616011	1950	0.603076923	4	22431	5	29.2	3	408800	1	582	0.298461538	1	14	2.8
530530616021	861	0.603948897	4	43869	4	28.9	3	221200	3	209	0.242740999	2	16	3.2
530530617001	1134	0.541446208	5	57634	3	25.9	2	212300	3	251	0.221340388	2	15	3
530530617002	1004	0.338645418	5	70000	2	32.6	4	198600	4	102	0.101593625	4	19	3.8
530530617003	740	0.618918919	4	52917	4	29.2	3	168900	4	140	0.189189189	2	17	3.4
530530617004	732	0.275956284	5	71429	2	28.1	3	161800	4	110	0.150273224	3	17	3.4
530530617005	1006	0.560636183	5	50547	4	24.2	2	158500	5	145	0.144135189	3	19	3.8
530530618001	1121	0.508474576	5	43825	4	34.1	4	154800	5	92	0.082069581	5	23	4.6
530530618002	991	0.415741675	5	0	0	50	5	137500	5	52	0.05247225	5	20	5
530530618003	790	0.481012658	5	48625	4	50	5	166000	4	94	0.118987342	4	22	4.4
530530619001	953	0.664218258	4	24453	5	31.2	3	189800	4	246	0.258132214	1	17	3.4
530530619002	953	0.661070304	4	37500	5	46.8	5	155100	5	161	0.168940189	3	22	4.4
530530620001	955	0.788481675	3	50357	4	24.9	2	181100	4	81	0.084816754	5	18	3.6
530530620002	1110	0.501801802	5	42031	5	33.2	4	155800	5	124	0.111711712	4	23	4.6
530530620003	1115	0.340807175	5	51000	4	31.7	3	147200	5	110	0.098654709	4	21	4.2
530530620004	1328	0.613704819	4	41023	5	27.7	2	155700	5	165	0.124246988	3	19	3.8
530530623001	1419	0.563072586	5	49766	4	47	5	157800	5	182	0.128259338	3	22	4.4
530530623002	1500	0.407333333	5	33393	5	38.8	5	140600	5	135	0.09	4	24	4.8
530530623003	1328	0.472891566	5	62015	3	18.7	1	169000	4	149	0.112198795	4	17	3.4

.

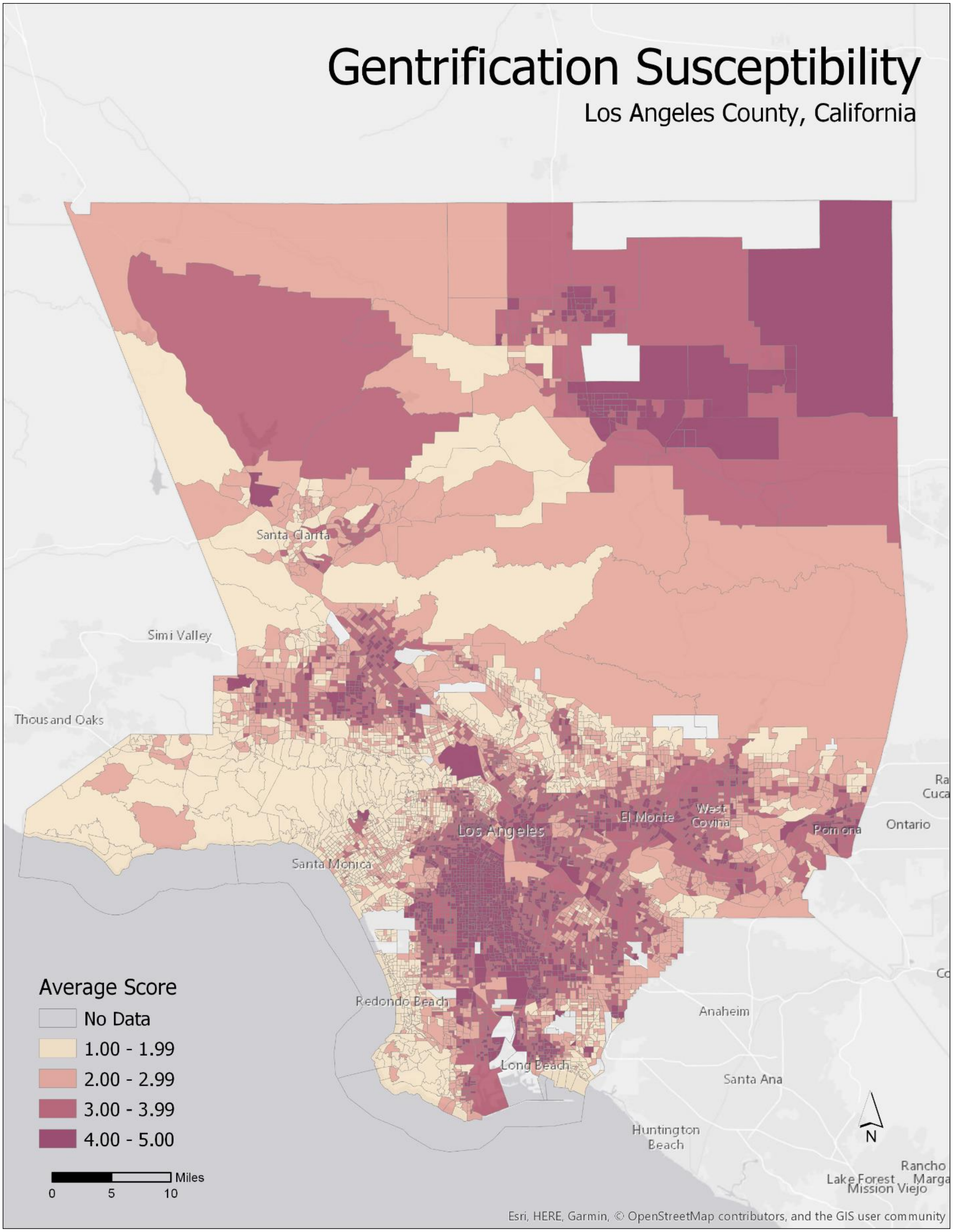

Figure 1. Gentrification susceptibility in Los Angeles County, California

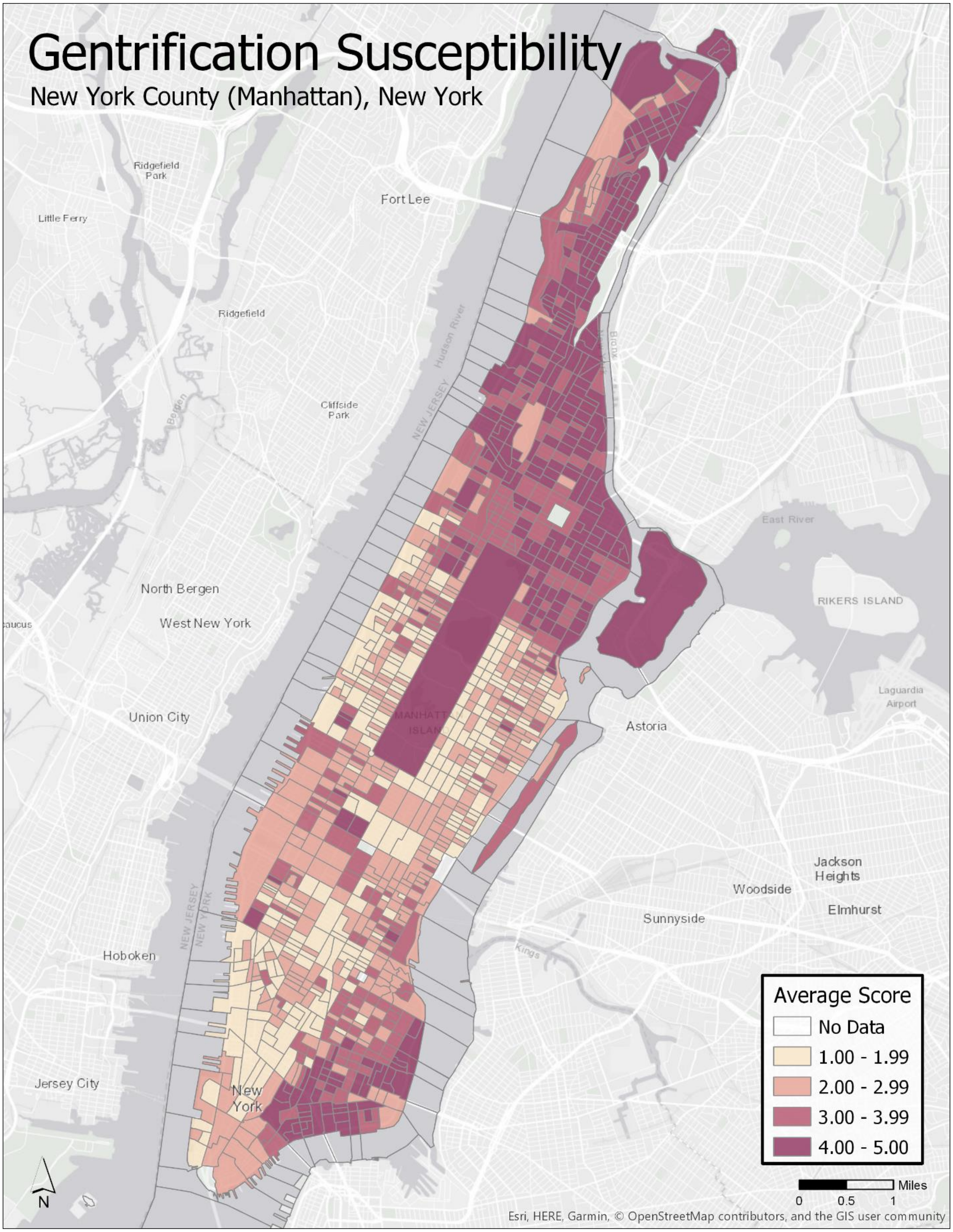

Figure 2. Gentrification susceptibility in New York County, New York

6. Discussion

The Python-scripted model developed to carry out the automation of demographic gentrification analysis illustrates that it is possible to develop a repeatable technological framework for quantitatively identifying susceptibility to gentrification anywhere in the United States. While the study of quantitatively identifying gentrification is not novel, to date the research has tended to focus on individual communities or has retroactively probed an area that has gentrified in order to determine latent origins. While these highly specific and localized analyses are certainly significant and should not be diminished, the Python-scripted model is not meant to be a replacement, but rather seeks to address the need to be able to be to quickly and easily locate susceptible areas anywhere in the United States.

As Figures 1 illustrates, the model finds nearly the entire coast of Los Angeles County, California – which encompasses Beverly Hills, Malibu, Santa Monica, Redondo Beach, etc. – to be the least susceptible to gentrification, while inner-city Los Angeles – Inglewood, Compton, etc. – is calculated to be the most susceptible. Similarly, in Figure 2, the island of Manhattan, which comprises nearly the entirety of New York County, New York, shows that the northern area of the island – encompassing Harlem, Sugar Hill, Washington Heights, etc. – is calculated to be the most susceptible to gentrification while the southern area of the island – encompassing Midtown, Wall Street, Greenwich Village, etc. – is calculated to be far less susceptible, with the exception of Chinatown and the Lower East Side.

Typically, the model identifies the most underprivileged areas as the most susceptible to gentrification which may not qualify them as the most likely to gentrify. It is far less likely that the gentrification process would originate in inner-city Los Angeles and migrate outward, than that the areas directly abutting previously gentrified areas would be the next to capitulate to the gentrification process as Guerrieri, Hurst, and Hartley (2013) noted. In general, a user of this model should visually identify areas with higher scores (3,4,5) that are in close proximity to areas with lower scores (1,2) and examine them further. It is likely that the areas with lower scores have previously gentrified, whereas the proximity of block groups with higher scores to those with lower scores is likely to influence real-world gentrification probability. The model's lack of ability to identify these block groups and calculate a score based on gentrification susceptibility of the nearby block groups could be considered a limitation. However, due to the intentionally flexible and broad nature of model it was beyond the scope of this research to consider all the potential phenomena that may impact gentrification susceptibility on a local level – particularly land use but also zoning, proximity to transit, travel times, etc. Because of the many potential localized phenomena which can impact susceptibility, in its current form it is unclear to what extent including an additional proximity-based measurement into this model would aid in assessment and perhaps this is a reason why gentrification research has tended to be highly localized. However, including a method to calculate a score based on both susceptibility and proximity in a future version of the model is being examined.

A further potential limitation of the model lies in the way the average index score is calculated. The model totals the scores for the five indicators and divides that total score by the number of indicators that have data for each block group. Ideally each block group would have data for each of the five indicators so the total score would be divided by five. However, many block groups do not contain complete datasets (Table 1). While all the indicators were carefully researched before selection, it is possible a score can become compromised if there are fewer indicators. Therefore, the fewer indicators for a block group which contain data, the more potentially distorted the average index score may become. This is illustrated in Figure 2 where Central Park has been identified as an area with the highest susceptibility to gentrification based on incomplete data which show a population of five individuals, of which two were non-Hispanic white, none had received a Bachelor's degree or higher, none reported income, none reported rent, and there were no homes for which to assess housing value. Central Park should contain zero values for each of the five indicators and it is unknown why data exists but it is likely that the data reflect caretakers, a small homeless contingent, respondents giving a false address, or a mix (Feuer 2011). Central Park is thus an extreme outlier in the model but does serve to illustrate the impact incomplete data can carry. A potential solution to this would be to conduct a type of statistical interpolation to calculate an estimate for missing data but doing so could potentially compromise findings by assigning values for indicators that may or may not be influenced by the values for those same indicators in the nearby block groups and may or may not accurately reflect the area.

When utilizing data from the United States Census Bureau's American Community Survey (ACS), the data itself serves as a limitation. While the ACS provides a repository for data for the entirety of the United States, the 5-year estimates are approximately two years behind real time, which, in this case, means that it is possible that the areas identified by the model as particularly susceptible to gentrification have started the gentrification process or already gentrified. Additionally, the ACS calculates estimates using far fewer responses than the decennial census, which can lead to data values with large margins of error. However, the extent to which this impacts the model's findings require much more widespread testing by local experts than was within the scope of this research. Furthermore, as noted by Wyly and Hammel, "(g)entrification is notoriously difficult to measure, and results are sensitive to the indicators chosen, the time periods under investigation, and thresholds used to distinguish among neighborhoods" (1999: 726) which creates a problem when trying to utilize the relatively inflexible and small amount of data collected by the US Census at the block group enumeration level. There may be additional or different indicators in certain areas that this model simply cannot consider.

Finally, gentrification is commonly considered to be solely an urban problem – and has largely been discussed as one in this study – but the Python model considers all block groups in a county whether or not they are deemed urban because whether the understanding of gentrification as solely an urban problem has merit is debatable (Ghose 2004; Phillips 1993; 2005) but creating an index that includes values county-wide may be viewed by some as a limitation though that critique is contrary to the stated purpose of providing a model that can assess susceptibility anywhere in the United States.

Despite the limitations of the model, it does succeed in its aim to provide a technological framework to quickly and easily locate areas susceptible to gentrification anywhere in the United States. While it lacks the specificity of a traditional localized analysis, the model typically outputs its product in approximately one minute for even the largest and most populous counties.

7. Conclusion

Gentrification is arguably one of the most controversial issues in American cities today, but it also remains one of the least understood. Few agree on how to define it or whether its impacts are overall positive or negative. There have been numerous avenues through which the gentrification process has been examined both qualitatively and quantitatively. Though both the qualitative and quantitative research on gentrification is robust and varied, there have been comparatively few attempts to produce an easily repeatable model to assess susceptibility to gentrification on a large scale. This research was conducted, not in an attempt to determine whether a particular approach is superior to another, but to determine whether susceptibility to the gentrification process can be quantitatively identified and to provide a technological framework upon which further and more localized analysis can be undertaken.

By creating a Python-scripted model to automate an index utilizing data obtained from the US Census Bureau's American Community Survey 5-year estimates pertaining to median income, rent as a percentage of household income, housing value, percentage of non-Hispanic white population, and educational attainment, the model can be used to quickly assess susceptibility to gentrification anywhere in the United States. Furthermore, by hosting the model online this research has the potential to provide access to this information to individuals and agencies that may not have the resources or technical prowess to commission or conduct a traditional localized gentrification analysis and has the potential to inform policy seeking to mitigate gentrification's impacts.

While the model demonstrates that it is possible to quantitatively identify areas based on their susceptibility to gentrification, future iterations of the model will look to develop a method to account for the proximity of susceptible areas to less susceptible areas as it is likely that proximity has a significant impact on which neighborhoods will be the next to undergo the gentrification process. Adding a method to calculate a score based on both susceptibility and proximity could provide an avenue through which it is possible to quantitatively predict the gentrification process.

References

2010 Geographic Terms and Concepts - Block Groups. (2012, September 01). Retrieved July 17, 2018, from https://www.census.gov/geo/reference/gtc/gtc_bg.html

Atkinson, R. (2000). Measuring Gentrification and Displacement in Greater London. Urban Studies, 37(1), 149-165.

Atkinson, R. (2003). Introduction: Misunderstood Saviour or Vengeful Wrecker? The Many Meanings and Problems of Gentrification. Urban Studies, 40(12), 2343-2350.

Barton, M. (2016). An exploration of the importance of the strategy used to identify gentrification. *Urban Studies, 53*(1), 92-111.

Bostic, R., & Martin, R. (2003). Black Home-owners as a Gentrifying Force? Neighbourhood Dynamics in the Context of Minority Home-ownership. *Urban Studies, 40*(12), 2427-2449.

Butler, T. (2003). Living in the bubble: Gentrification and its 'others' in north London. Urban Studies, 40(12), 2469-2486.

Butler, T., & Hamnett, C. (2007). The Geography of Education: Introduction. Urban Studies, 44(7), 1161-1174.

Carmona, M. (2003). Public places, urban spaces : The dimensions of urban design. Oxford ; Boston: Architectural Press.

Chapple, K. (2009). Mapping Susceptibility to Gentrification: The Early Warning Toolkit. Retrieved February 3, 2018, from http://communityinnovation.berkeley.edu/publications

Cupcake Gentrification. (2009). Retrieved February 01, 2018, from http://www.ediblegeography.com/cupcakegentrification/

Eckerd, A. (2011). Cleaning Up Without Clearing Out? A Spatial Assessment of Environmental Gentrification. *Urban Affairs Review, 47*(1), 31-59.

Ellen, I., & O'Regan, K. (2011). Gentrification: Perspectives of Economists and Planners. The Oxford Handbook of Urban Economics and Planning, The Oxford Handbook of Urban Economics and Planning.

Ellen, I., & O'Regan, K. (2011). How low income neighborhoods change: Entry, exit, and enhancement. Regional Science and Urban Economics, 41(2), 89-97.

Freeman, L. (2009). Neighbourhood diversity, metropolitan segregation and gentrification what are the links in the US? Urban Studies, Vol. 46 (2009), Nr. 10, S. 2079-2102

Freeman, L., & Braconi, F. (2004). Gentrification and Displacement New York City in the 1990s. Journal of the American Planning Association, 70(1), 39-52.

Feuer, A. (2011). Census Apparently Did Check Behind Every Tree. Retrieved July 27, 2018, from https://www.nytimes.com/2011/03/26/nyregion/26census.html

Ghose, R. (2004). Big Sky or Big Sprawl? Rural Gentrification and the Changing Cultural Landscape of Missoula, Montana. Urban Geography, 25(6), 528-549.

Glass, R. "Introduction: Aspects of Change" in London: Aspects of Change, ed. Centre for Urban Studies (London: McKibbon and Kee, 1964), 18-30.

Guerrieri, V., Hurst, E., & Hartley, D. (2013). Endogenous gentrification and housing price dynamics. Journal of Public Economics, 100(C), 45-60.
Hackworth, J. (2007). *The neoliberal city: Governance, ideology, and development in American urbanism* (Cornell paperbacks). Ithaca: Cornell University Press.
Holland, S. (2016). Gentrification: Causes and Consequences. Journal of Lutheran Ethics, 16(1). Retrieved from http://elca.org/jle/articles/1135#_edn11
Jackson, J., Forest, B., & Sengupta, R. (2008). Agent-Based Simulation of Urban Residential Dynamics and Land Rent Change in a Gentrifying Area of Boston. *Transactions in GIS, 12*(4), 475-491.
Jacobellis v. Ohio, 378 U.S. 184 (1964)
Kennedy, M., Leonard, Paul, Brookings Institution. Center on Urban Metropolitan Policy & Policy. (2001). *Dealing with neighborhood change: A primer on gentrification and policy choices* (Discussion paper series (Brookings Institution. Center on Urban and Metropolitan Policy)). Washington, DC.: Brookings Institution Center on Urban and Metropolitan Policy.
Lees, & Ferreri. (2016). Resisting gentrification on its final frontiers: Learning from the Heygate Estate in London (1974–2013). Cities, 57(C), 14-24.
Lees, L., Slater, Tom, & Wyly, Elvin K. (2008). *Gentrification*. New York: Routledge/Taylor & Francis Group.
Newman, K., & Wyly, E. "The right to stay put, revisited: Gentrification and resistance to displacement in New York City," Urban Studies 43.1, (2006): 23-57.
O'Sullivan, D. (2002). Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems, 4*(3), 251-274.
O'Sullivan, D. (2009). Changing Neighborhoods-Neighborhoods Changing: A Framework for Spatially Explicit Agent-Based Models of Social Systems. *Sociological Methods and Research, 37*(4), 498-530.
Peck, J., Siemiatycki, E., & Wyly, E. (2014). Vancouver's suburban involution. *City, 18*(4-5), 386-415.
Phillips, M. (1993). Rural gentrification and the processes of class colonisation. Journal of Rural Studies, 9(2), 123-140.
Phillips, M. (2005). Differential productions of rural gentrification: Illustrations from North and South Norfolk. Geoforum, 36(4), 477-494.
Smith, N. (1979). Toward a Theory of Gentrification: A Back to the City Movement by Capital, not People. *Journal of the American Planning Association, 45*(4), 538-548.
Smith, N. (1987). Gentrification and the Rent Gap. Annals of the Association of American Geographers, 77(3), 462-465.
Wyly, E., & Hammel, D. (1999). Islands of decay in seas of renewal: Housing policy and the resurgence of gentrification. *Housing Policy Debate, 10*(4), 711-771.
Zukin, S. (1987). Gentrification: Culture and Capital in the Urban Core. Annual Review of Sociology, 13, 129-147.
Zukin, S., Trujillo, V., Frase, P., Jackson, D., Recuber, T., & Walker, A. (2009). New Retail Capital and Neighborhood Change: Boutiques and Gentrification in New York City. City & Community, 8(1), 47-64.