Jeff Hudson (jdh2182)
QMSS 4063 – Data Visualization
Final Project
May 11, 2015

Measuring Gentrification

1. Introduction

Gentrification has recently become a hot topic in the national conversation. Artists and celebrities such as Spike Lee publicly decry it, activist local residents stage protests, while television comedies such as "Girls" and "Broad City" skewer it for entertainment. Opponents charge that it displaces working-class families that often cannot afford to move, while proponents argue that it benefits neighborhoods by injecting an economic stimulus, improving public spaces, and lowering crime rates. While much is being said about gentrification, there have been few attempts to understand and quantify the process objectively.

An objective understanding is critically important as the issue continues to become more widespread. State actors and local governments are often in the position of enacting policies that favor or prohibit the kinds of neighborhood change that characterize gentrification. They need a balanced, fact-based perspective of how these policies will affect their constituents in order to make an informed decision. In particular, before any analysis of the effects can be done, it is necessary to establish objective criteria for quantifying a neighborhood's gentrification. I use New York City as the location to study due to its size, the existence of well-known gentrified neighborhoods, as well as my personal familiarity.

2. Literature Review

The term "gentrification" was coined by British sociologist Ruth Glass in 1964 as she sought to describe the influx of middle-class families displacing the poor from historically

working-class neighborhoods in London. At just over fifty years old, it is still a relatively recent

phenomenon and somewhat under-theorized, though most scholars agree that it is characterized

by two central components: rising housing prices and the in-migration of the middle-class.

Although it seems to be a simple formulation, empirical studies of gentrification differ greatly in

their operationalization of the concepts to identify gentrifying neighborhoods to study.

Some scholars have chosen to simply specify a priori the identity of neighborhoods

undergoing gentrification. Although they go on to perform a rigorous analysis of gentrification-

related displacement of low-income households, Freeman and Braconi do not establish an

empirical definition for gentrification. Instead, they rely on their own "familiarity with recent

trends in neighborhood change" to identify the New York City neighborhoods that they consider

gentrified (Freeman & Braconi, 2004). This strategy simplifies the analysis but it limits the

validity and generalizability of their findings considerably. Instead, I propose to use local

knowledge as a way to validate and verify that results from objective model predictions of

gentrification do indeed correspond to gentrified areas.

Most scholars choose to focus either on housing stock or on demographic characteristics.

Occupational status is particularly popular in studying gentrification of London (Atkinson, 2000)

and Paris (Clerval, 2011). On the other end of the spectrum, Smith and Defilippis use property

tax arrears as a proxy for (dis)investment in housing stock (1999). It is expected that these two

different kinds of measures would still capture the same phenomenon, and indeed, Hammel and

Wyly (1998) found that census data was highly correlated with changes in housing stock in four

American cities.

The present study builds off of these previous efforts to use census data to map

gentrification in cities across the United States. In particular, Gina Clemmer (2000) examined

patterns of gentrification in the Portland, Oregon metropolitan area using several key indicators from the American Community Survey conducted annually by the U.S. Census Bureau. Elvin Wyly and Daniel Hammel (1996) also used U.S. Census data to perform linear discriminant analysis to identify key features of gentrified neighborhoods. Their results agree that the following eight demographic and housing characteristics are of primary interest:

- Age
- Race
- Household Income
- Residential Mobility

- Occupation
- Educational Attainment
- Property Value
- Rent

I rely on this earlier work and take these variables as my primary indicators of gentrification.

3. Description of Data

The data for the study are taken from 1% Public Use Microdata Samples of the United States Census Bureau's American Community Survey (ACS). These data are maintained through the Integrated Public Use Microdata Series (IPUMS) project of the Minnesota Population Center at the University of Minnesota. IPUMS catalogues changes in the codebooks for each survey and recodes some variables to maintain consistency across survey years, making it much easier to analyze changes over time. The microdata samples contain individual and household-level data with weights for each which allows us to construct neighborhood-level statistics directly from the underlying population.

The smallest geographical unit available in the microdata is the Public Use Microdata Area (PUMA), each of which contains at least 100,000 individuals. There are 55 PUMAs within the New York City limits. While it might be more instructive to analyze the data at a smaller geographical unit, such as the zip-code or tract level, due to privacy concerns, the ACS data is

not available at this level of specificity. Regardless, these PUMAs still correspond meaningfully to coherent neighborhoods and give us small enough units to isolate the occurrence of gentrification. The years of the survey I use are 2005 through 2011 as these are the only years with enough geographic specificity (at the PUMA level) to be meaningful in this analysis.

Figure 1 shows a map of the PUMAs in New York City (colored arbitrarily). The PUMAs are conceptualized as encompassing intervening bodies of water, thus the PUMAs representing Staten Island (for instance) are adjacent to PUMAs in Brooklyn, despite being separated by a river.

Figure 1. Map of Public Use Microdata Areas (PUMAs) in New York City
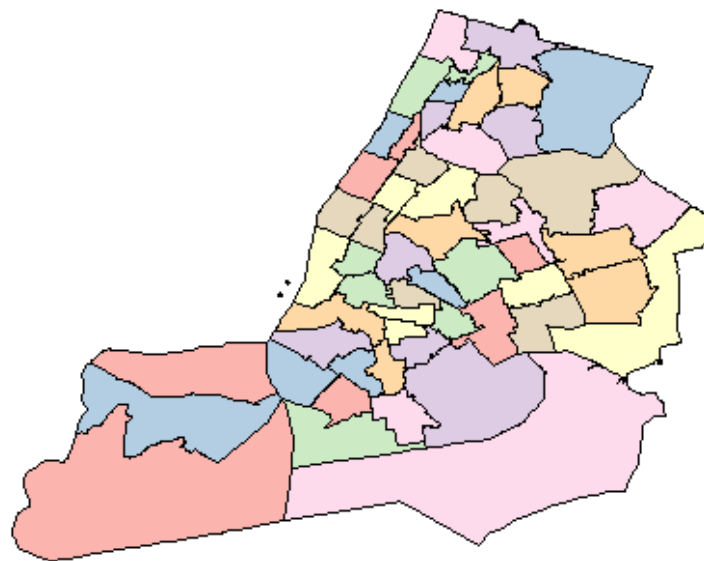
The data contain weights at both the household and individual level. For variables such as household income, the household level weight is clearly the most appropriate, while for age or race, individual weights must be used. For some variables, such as residential mobility, a case could be made for either. A couple with an infant might have been living in the same unit for six years, while the child would only have been there for one. In this case, I prefer to use the

household's residential mobility (six years) than the average over each individual in the household (4.33 years).

4. Construction of Variables

Because these neighborhood statistics are being calculated from individual and household observations, I have substantial latitude in how to construct them to maximize their usefulness for the identification of gentrification. As an example, the age variable could be constructed as the mean age, the median age, or even the percent of residents over/under a certain age. Each would alter the end results – some marginally, some dramatically. Additionally, some formulations will correspond better to our preconceived notions of the dynamics of gentrification. For instance, it is easy to understand how the percent of adult residents under 35 is associated with gentrification, while the connection between gentrification and the mean age in a neighborhood is less clear. I will discuss the factors motivating the specific construction of each variable in the analysis, including which weights are used for each.

- **Age**: As noted above, there are many ways this variable could be constructed, however, I use the percent of the population aged 18 to 35 as it has direct correspondence to gentrification, given that gentrifying residents tend to be young adults. Mean and median age can be pulled in either direction by a large number of children or the elderly, which would not be a good indication of the number of residents that are the age of the average gentrifier. In addition, this construction allows the variable to be positively correlated with increasing gentrification, which simplifies constructing a composite index. This variable uses the individual-level weights and is called **UTF** for percent under thirty-five.

- **Race**: Since race is a categorical variable, some categories must be grouped together to get a single figure for each neighborhood. The construction that has the best

correspondence to the task of identifying gentrified neighborhoods is again to align it with the archetypal gentrifier, who is white. Thus for this variable I use the percent of the neighborhood respondents who identify as white. Most gentrified neighborhoods were formerly home to African-American or Latino communities, so tracking the increase in the percentage of white residents should give us an indication of gentrification. The ACS records "race" and "Hispanic origin" as separate measures, therefore our "percent white" figure will combine these variables and count only the number of non-Hispanic whites. It is important to note that in Brooklyn's Greenpoint neighborhood, for instance, the incumbent working-class residents are mostly of Polish descent, therefore also white. Thus, this may not help identify gentrification in every neighborhood, but can be a good indicator for some. I use **WHT** to refer to the percent white variable and it is constructed using the person-level weights.

- **Household Income:** I use the median, weighted at the household-level, as the statistic to measure household income. There are always large outliers for income (especially in NYC) so the mean is not a robust estimator of the neighborhood's average income level. Median household income will be encoded with the variable name **MHI**.

- **Residential Mobility**: The raw data for residential mobility are a categorical measure of the number of years that the respondent has lived in her or his current home or unit. These categories are in uneven intervals, beginning with one year, then moving to two, five and ten year intervals; thus direct calculation of statistics such as the mean or median is not possible. Instead, I combine the first three categories to find the total proportion of residents who have moved in to their dwelling within the past five years.

This is a good indicator of what proportion of residents are highly mobile. Residential mobility (**RMB**) is calculated using household-level weights.

- **Occupation**: The ACS's measurement of occupation is extremely detailed, including hundreds of different occupational categories. Past studies of gentrification generally group white collar workers, and those in professional, managerial, or technical fields. The codes are grouped thematically, making it easy to separate out groups of professions. Some groups included are: "Architecture and Engineering Occupations," "Financial Specialists," and "Management, Business, Science, and Arts Occupations" while some professions excluded are: "Protective Service Occupations," "Office and Administrative Support Occupations," and "Construction and Extraction Occupations." The percent of residents with professional/managerial occupations is given the variable name **OCC** and is calculated using the person-level weights.

- **Educational Attainment**: Educational attainment is also provided as a categorical variable in the ACS. Both Clemmer (2000) and Wyly & Hammel (1996) use the percent of the population with a Bachelor's degree as their indicator of educational attainment in studying gentrification, since middle-class gentrifiers tend to be college-educated. I believe that those who have received *some* college education are more similar to those who have a four-year degree, than to those who did not seek more schooling after high school. Thus I have decided to group the upper categories together to separate those with any college education from those who have up to a high school diploma. The educational attainment variable uses the person-level weights and is called **EDU**.

- **Property Value**: There are several different variables available in the ACS that could measure property value: house value, property taxes, and owner-occupied unit monthly

costs. Both house value and owner costs are given in real dollars, while property taxes are given in intervals of unequal size. Many New Yorkers live in apartment buildings, so these figure will be skewed for neighborhoods with smaller numbers of owner-occupied homes. Indeed, for all respondents who reported paying any amount of rent (non-zero value), the house value is top-coded at $10,000,000. Property values for what appear to be apartment buildings should be excluded, while it must be noted that some neighborhoods in New York City will have actual homes valued at $10 million or more, therefore home values must be evaluated in combination with other measures to determine which correspond to single units, and which to apartment buildings. For respondents who report a non-zero value for their rent, I ignore their property value figure in the calculation of this statistic. I use **MHV** to refer to the median home value as constructed in this way, using the household-level weights.

- **Rent**: The ACS includes data on both monthly contract rent and gross monthly rent (rent plus utilities). I use gross rent, as it better corresponds to the total cost of living. Some rent contracts include utilities like gas and electricity, while others do not, thus gross rent should allow for more consistent comparisons of total living costs in different areas. This figure can be combined with property value figures to give a more complete picture of how rent changes concurrently with housing stock. Rent may increase faster or earlier than property values because rental units turn over more quickly and can be renovated more quickly than a house. However, this may also lead rent figures to be more responsive to minor fluctuations in the market, while property values should be more stable over time. Median gross rent (**MGR**) uses the household weights.

5. **Results**

After calculating the demographic statistics for each neighborhood, I formed a composite index then examined the trend in the index over time to identify neighborhoods where the index is increasing. Since I do not have an objective measure of gentrification for these regions, I cannot perform a simple regression or classification analysis to identify gentrification.

Table 1 shows the primary variables of interest aggregated for all of New York City over the seven survey years. Immediately several trends are clear. Median Gross Rent (MGR) is steadily increasing. The effects of the market crash in 2008 are evident: Median Home Value (MHV) plummets after its high in 2007 at $625,000. The percent of adults with some college education (EDU) increases consistently, rising an entire 5% in just seven years, while the percent of adults under thirty-five (UTF) also rose steadily. Other variables do not display meaningful or consistent patterns.

All individual neighborhood trends must be compared to the city-wide trend to be meaningful. For instance, increases in rent are widely associated with gentrification, however, rent is increasing in aggregate across the city, thus a gentrifying neighborhood should experience particularly high rent increases compared to this average increase in order to distinguish it from other neighborhoods in the city.

Table 1. Aggregate Demographic Indicators in NYC between 2005 and 2011

YEAR	MHI	MGR	MHV	RMB	OCC	EDU	UTF	WHT
2005	$42,800	$910	$450,000	37%	22%	36%	23%	35%
2006	$49,645	$940	$625,000	35%	21%	36%	23%	35%
2007	$51,000	$990	$625,000	35%	21%	37%	23%	35%
2008	$54,000	$1,030	$600,000	36%	22%	38%	23%	35%
2009	$54,400	$1,080	$560,000	35%	23%	40%	25%	35%
2010	$50,800	$1,110	$550,000	35%	22%	40%	26%	33%
2011	$52,000	$1,170	$550,000	35%	22%	41%	26%	33%

Figure 2 shows these same variables over time for two neighborhoods, one expected to be experiencing gentrification (Williamsburg), and the other not (Upper West Side). Although the

Upper West Side has significantly higher values for most of the indicators (in particular, median household income and occupational prestige) there is no trend over time – the neighborhood is not changing noticeably. In Williamsburg, on the other hand, although indicators such as residential mobility and educational attainment are relatively low throughout, they are steadily increasing over the study period. This comparison highlights the importance of examining change over time across these variables, rather than their absolute values. Gentrifying neighborhoods are always in a process of *change*. Thus my analyses will focus primarily on the difference in these variables over time.

Figure 2. Gentrification Indicators in the Upper West Side and Williamsburg, Brooklyn
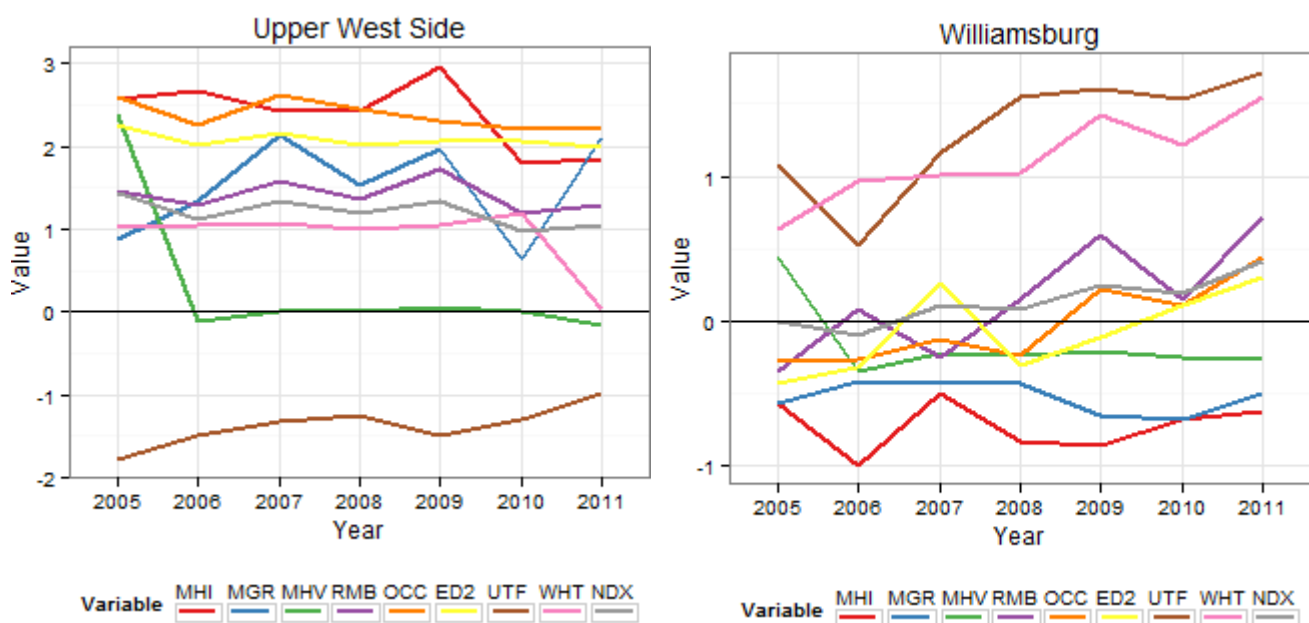

In order to capture the net effect of changes in these eight different variables, I constructed a composite index measure. First I normalized the values of the eight variables using a standard z-score transformation, $z = \frac{x - \bar{x}}{\sigma}$, then I took the mean of the values for each PUMA and Year combination to get a standardized index (NDX) of gentrification. Note that all variables are expected to increase with gentrification, therefore the mean should correspond to an average

indication of gentrification. The literature does not suggest any particular scheme to give more

weight to some variables over others, so each is weighted equally.

To assess the relevance and meaningfulness of this index, I constructed a histogram of

the difference in index values between the first and last years in the dataset. These values

correspond to the total increase or decrease in the gentrification indicators in each neighborhood

over the course of the study. The histogram in Figure 3 shows that most neighborhoods are

clustered around 0, indicating that, regardless of their initial positions, most neighborhoods did

not see substantial changes in gentrification-associated indicators. There are, however, several

outliers that experienced substantial increases in the gentrification index.

Figure 3. Histogram of Differences in Gentrification Index
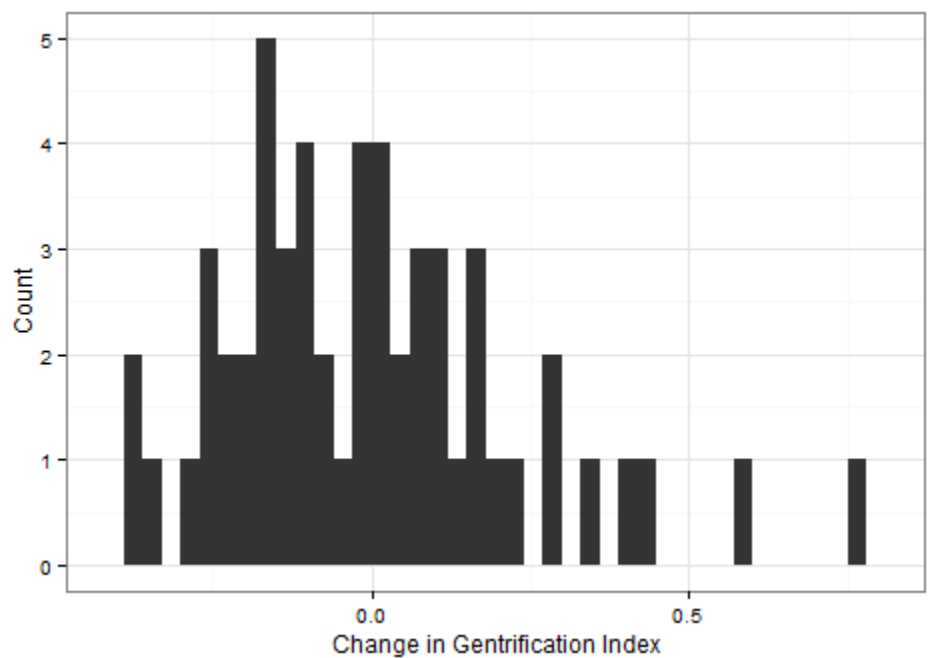
The project website (http://jeffhudson.github.io/nycgent/) displays a map of the changes

in gentrification index by neighborhood. Outlier neighborhoods can be identified (in purple).

These outliers (with total increases in the gentrification index greater than 0.25) are listed by

name in Table 2. They include well-known gentrifying neighborhoods such as Williamsburg in

Brooklyn, but also several neighborhoods in the Bronx, which is less commonly associated with gentrification. The most significant increase is in the West Harlem neighborhoods of Hamilton Heights and Manhattanville. These results confirm my expectations based on the recent literature identifying gentrified neighborhoods in New York City (Barton 2014). The project website can also be used to view the map for any variable and year combination, as well as the 2005-2011 trend over time.

Table 2. Neighborhoods with Significant Increase in Gentrification Index 2005-2011

Neighborhood Names (Borough)	Index Increase
Hamilton Heights, Manhattanville & West Harlem (Manhattan)	0.76
Hunts Point, Longwood & Melrose (Bronx)	0.59
Morris Heights, Fordham South & Mount Hope (Bronx)	0.44
Greenpoint & Williamsburg (Brooklyn)	0.42
Brooklyn Heights & Fort Greene (Brooklyn)	0.34
Bedford-Stuyvesant (Brooklyn)	0.29
Co-op City, Pelham Bay & Schuylerville (Bronx)	0.29

While these initial results conform to my expectations regarding gentrifying neighborhoods in New York City, the index measure is a crude approximation and may not capture important nuances between different component variables. To incorporate more variability into this characterization, I performed k-means clustering (k=5) using all eight differenced variables. Though the output of the k-means algorithm is not deterministic due to its random initialization, the results were surprisingly consistent over a number of repetitions.

West Harlem, Hunts Point, and Morris Heights consistently made up a single cluster. The centroid corresponding to this cluster was characterized by the highest index increase as well as a large difference in median home values and substantial increases in median household income and the adult population under thirty-five. The cluster with the next highest gentrification index increase was slightly more variable but most often included Williamsburg, Bedford-Stuyvesant, Brooklyn Heights, Crown Heights, and Astoria. These neighborhoods (mostly in Brooklyn and

west Queens) had noticeable increases across the board, but particularly in residential mobility, occupational prestige, percentage of adults under thirty-five and percentage of white residents. At the other end of the spectrum, most of Manhattan (south of Harlem) was a single cluster characterized by declining property values and decreasing percentage of white residents. The final two clusters encompassed the neighborhoods that saw very little change and were not consistently well-defined.

6. Discussion

Although it was very simple, the composite gentrification index was a good measure of gentrification. There is no objective "ground-truth" about socially constructed phenomena such as gentrification, but the literature confirmed that the identified neighborhoods are recognized by other scholars as undergoing gentrification. In particular it is a validation of the work pioneered by Hammel and Wyly, and continued by Clemmer, to organize and catalogue demographic indicators of gentrification. Though none of those authors had based their work on studies of New York City, the variables they reported were successful in identifying gentrification in NYC. This suggests that while gentrification may have different flavors in different regions, the general socio-demographic profile is similar everywhere.

The cluster analysis lends some support to theories of multi-stage gentrification processes and offers a characterization of two stages. The first cluster was primarily defined by sharply rising home values in West Harlem and the Bronx. The process in these neighborhoods could be classified as "emerging gentrification." A likely narrative is that savvy real estate developers begin buying plots in these neighborhoods, banking on their sharp increase in value once the neighborhood becomes gentrified. Thus the beginning of the process in the real estate market is reflected by sharp increases in home values and rent. After developers renovate or build new

units, middle-class gentrifiers arrive en masse, driving the population-based changes such as age, educational attainment and residential mobility that characterize the second cluster of later-stage gentrified neighborhoods in Brooklyn.

7. Conclusion

Gentrification is a complex process that can be identified, influenced and impacted by a number of varied factors. The existing literature suggested some variables to use in identifying gentrifying neighborhoods, but there was no clear model or scheme on how to use these variables to predict the presence of gentrification. A simple composite index measure was sufficient to demonstrate that a few neighborhoods were changing rapidly, while many others were not changing at all. Furthermore, clustering on the overall trend in these variables revealed two distinct groups of gentrifying neighborhoods, which might be characterized as "early-stage" and "late-stage" gentrification. These results confirmed that gentrification can be readily identified by measuring the change in a handful of publicly available demographic indices.

As gentrification continues to define more and more neighborhoods in the urban United States, it is important to understand and be able to quantify the process. An objective evaluation of the impact of gentrification is vital to be able to hold state actors accountable when enacting policies that either promote or discourage it.

8. References

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

Atkinson, R. (2004). The Evidence on the Impact of Gentrification: New Lessons for the Urban Renaissance? *European Journal of Housing Policy*. *4*(1). 107-131.

Barton, M. (2014). An Exploration of the Importance of the Strategy Used to Identify Gentrification. *Urban Studies*. *20*(1).

Cameron, S. (2003). Gentrification, Housing Redifferentiation and Urban Regeneration: 'Going for Growth' in Newcastle upon Tyne. *Urban Studies*. *40*(12). 2367–2382.

Clerval, A. (2011). The Spatial Dynamics of Gentrification in Paris: a Synthesis Map. *Cybergeo: European Journal of Geography, Espace, Société, Territoire,* document 553.

Clemmer, G. (2000). Quantitative and Spatial Analysis Techniques for Analyzing Gentrification Patterns. Unpublished white paper.

Dutton, P. (2003). Leeds Calling: The Influence of London on the Gentrification of Regional Cities. *Urban Studies*. *40*(12). 2557-2572.

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

Hamilton, T. & Curran, W. (2013). From ''Five Angry Women'' to ''Kick-ass Community'': Gentrification and Environmental Activism in Brooklyn and Beyond. *Urban Studies*. *50(*8). 1557-1574.

Hammel, D. & Wyly, E. (1998). Modeling the Context and Contingency of Gentrification. *Journal of Urban Affairs*. *20*(3). 303-326.

Hammel, D., & Wyly, E. (1996). A Model for Identifying Gentrified Areas with Census Data. *Urban Geography*. *17*(3). 248-268.

Hamnett, C. (2003). Gentrification and the Middle-class Remaking of Inner London, 1961–2001. *Urban Studies*. *40*(12). 2401-2426.

Lees, L. (2000). A Reappraisal of Gentrification: Towards a 'Geography of Gentrification'. *Progress in Human Geography*. *24*(3). 389-408.

Lees, L. (2008). Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance? *Urban Studies*. *45*(12). 2449-2470.

Pratt, A. (2009). Urban Regeneration: From the Arts 'Feel Good' Factor to the Cultural Economy: A Case Study of Hoxton, London. *Urban Studies*. *46*(5&6). 1041-1061.

Smith, N. & Defilippis, J. (1999). The Reassertion of Economics: 1990s Gentrification in the Lower East Side. *International Journal of Urban and Regional Research*. *23*(4). 638-653.

Williams, P. & Smith, N. (1986). From 'Renaissance' to Restructuring: the Dynamics of Contemporary Urban Development, in: N. Smith and P. Williams (Eds) *Gentrification of the City*. 204–224. London: Allen and Unwin.

Wyly, E. & Hammel, D. (1999). Islands of Decay in Seas of Renewal: Housing Policy and the Resurgence of Gentrification. *Housing Policy Debate*. *10*. 711–771.