GIS Spatial Analysis Final Project: Gentrification in the NYC Metro Area and its Social Effects

Manuel Rueda
UNI: MR3523

Introduction

Gentrification is a subject that has received significant media and academic attention in the last few decades, mainly because it's a process powered by the possibility of mobility which has been limited in earlier years. Given the recent rise of this phenomena, researchers have not been able to reach an agreement on what are its causes, its social and economical effects, or even on what constitutes a formal *process of gentrification.* Regardless of this, the subject is one that continues to gain importance, mainly because it has been accelerating on recent years, and it has significant implications in public policy; for example, in several studies it has been negatively associated with the displacement of population and loss of capital investment.

The aim of this study is to help narrow that gap, by analyzing the case of the New York City Metro Area and investigating which possible social covariates could be related to this process. To do this, first we begin by reviewing the literature available on the topic, from which we will be able to gather a number of variables that will be included in the analysis. Then we will proceed to define a specific methodology to test for association or dependence between census tracts that have gentrified in NYC, and 3 specific covariates: the percentage of population with a higher education degree, the percentage of population dedicated to the service industry, and the mean age. Special techniques that account for possible clustering in space will be utilized. Once the results have been attained, comments will be made on them, and suggestions for future research will be provided.

Literature Review

Before proceeding with an analysis of the effects or associated social variables of gentrification, it is convenient to revise what other academics have found on the subject. This phenomenon is relatively new, so most of the investigations that have been done with regards to it are also of recent publication. Now we proceed to review a number of them, as it might provide some insights into our own research.

London and Palen (1984) were two of the first sociologists to formally dive the issue, as they found that during the decade gentrification was becoming a subject of debate, and very little literature had been consolidated around it. Their main task was to review the scattered studies that had been done around the subject, gathering research on the topic performed by sociologists, city planners, geographers and urban studies faculty. Their published work does not seek to prove a particular hypothesis, but to serve as a starting point for the discussion. Some of the social variables they find relate with the gentrification process are social mobility, a change in the working class patterns, and residential displacement of the elderly.

Another seminal study, performed only 2 years after, is that of Ley (196). In his work he examines the Canadian urban system between 1971 and 1981, and tests a number of social, geographical and economical variables that could possibly explain the gentrification effect. With this, the author reviews four major explanations for it via correlation and regression analysis. The hypothesis reviewed are: demographic change, housing market dynamics, urban amenities, and changes in the economic base to a white collar industry. Based on his analysis economic and urban amenities factors perform most strongly.

Following the proposed thesis by Ley on 1986, Smith (1987) suggests that economic forces play the most important role on the process of gentrification. He exposes the basis of the rent-gap theory, an economic explanation for it. The theory is based on the disparity of the current rental income of a property and its potentially achievable rental income. If this exists, the author argues, then there is a strong incentive to invest in the property market of the neighborhood experiencing it. This is a purely economical approach that dismisses any sociological or

demographical explanations, and nevertheless the results obtained in this study are statistical significant.

Hays, Kreager & Lyons (1986) also conducted similar studies to the ones mentioned during the mid-80s. Their focus, however, is on the relationship between crime and gentrification. From a sociological perspective, the authors claim, security related variables are related with the inn-migration of wealthier population. To prove this, on their study they try to examine the relationship between crime and gentrification using data from the city of Seattle. The dataset consists on longitudinal tract-level records, from which the authors conclude that the gentrified areas in this region: (1) saw reductions in crime relative to those that did not gentrify, and (2) were also areas with higher than average crime rates at the beginning of the decade. Their final conclusion is that a curvilinear relationship exists between these two phenomena, where initially gentrification is associated with small increases in crime, but in latter stages the pattern reverses.

Another study from this decade which particularly relates to the case of New York City is that performed by Schaffer & Smith (1986). Asides from identifying possible driving factors for gentrification, they claim, it is also important to asses if there are others that could slow down or limit its development. For this, authors of this paper examine the impressionistic reports suggesting that Harlem might have undergone gentrification. A number of indicators from the 1980's census and mortgage data are examined, making this a social and economical approach to the subject. Their results suggesting that indeed the phenomenon started on Harlem, but several limitations made the process slow down. The concluding remarks point that if gentrification proceeds, this would result in white in-migration and the displacement of blacks.

Looking at more recent studies, we have that of Freeman (2014). On his paper the author retakes the displacement and mobility theory that have long been associated with gentrification, and proceeds to test if it is still relevant. To examine if there is a relationship between these variables, both gentrifying and non-gentrifying neighborhoods are investigated. His results suggest that no significant relationship takes place; demographic changes in gentrifying neighborhoods appear instead to be a consequence of low rates of intra-neighborhood mobility and relative affluence of in-movers.

Kennedy & Leonard (2001) also investigated the subject by taking a novel approach. They claim that, if one wants to identify the common drivers of gentrification, it might be useful to gather information from the people experiencing it. On their paper they examine the neighborhood's resident's view on the issue, providing insights from 'insiders' on the process. For this, Atlanta, Cleveland, Washington D.C. and the San Francisco Bay area are analyzed. The study suggest that long-term neighbors take very different positions and explanations for the phenomena, making it hard to form a decisive statement on whether they view the process as a threat or not.

Following this line of investigation, Snow & Turner (2001) investigates the phenomenon from the old-time neighbors point of view. Sometimes the process of gentrification is given a negative connotation, they claim, as it tends to occur inadvertently and might lead to the displacement of older population. For their study the authors examine the case of the District of Columbia, which on prior decades suffered from population loss and general disinvestment. Their work is focused on identifying the leading indicators of gentrification, in order to identify the areas where rapid reinvestment is likely to happen next. Their conclusion is that the negative consequences can be avoided if affordable housing is preserved in areas facing high demand for homeownership.

Also taking a similar stance, which relates to the public policy implications of gentrification, Bierbaum & Chapple (2015) present an up-to-date report with a very broad analysis of the literature available on the phenomena of gentrification, performed by academics from the University of California, Berkeley and UCLA. Topics such as historical perspectives, trends in mobility, the role of public investment, and simulation models are covered. They do not take a particular stance on the subject, but do identify a number of relevant variables, including level of education, house prices, and urban commodities.

On the year 2000, Bolitzer & Netusil retake some of the arguments brought up on earlier research, such as the one by Ley and others, and focus on the specific role of urban amenities. On their paper the authors create a proxy analysis of these amenities by looking at the impact of open-space proximity (i.e. parks) on home sale prices. They recur to pricing, geographical and home characteristics variables for the case of Portland, Oregon. Results show that, even though negative externalities might be induced due to this proximity, such as traffic congestion or noise,

in general it has a statistically significant effect on home prices, suggesting a positive correlation among the variables.

It is important to evaluate the extent to which neighborhoods are affected by gentrification, and if there are any common variables driving it across different territories. On this study performed by Governing (2015) (leading media platform on policy, politics and local government), the authors try to asses the extent to which gentrification has reshaped urban communities by analyzing the 50 largest cities in the U.S. The findings reveal that although the phenomenon has become more prevalent on recent years, it is still largely confined to selected areas. Some interesting correlations with gentrifying neighborhoods are: (1) general population increase; (2) population becoming whiter (particularly non-Hispanic); and (3) declining poverty rates.

Finally one of the most novel studies is that performed by Nara & Torrens (2007), taking a significantly different approach to the previous literature. Taking into consideration that in recent years' new computational methods have been made available, they try to implement them in the modeling social behaviors. On their paper the authors take an algorithmic approach to analyze gentrification, so their model makes use of cellular automata agents that allow for the representation of interactions among fixed and mobile agents in the urban environment. To backtest their approach, the case of Salt Lake City, Utah is examined, as it was undergoing a demographical transformation at the time. The results are positive, suggesting that the hybrid approach is useful in representing complex adaptive human systems. Although this approach will not be considered in this particular study, it leaves open a path where future researchers might want to focus their efforts.

Methodology

To evaluate the social variables associated with the phenomena of gentrification, we need to start by defining what we mean by it. To do this we will look at the work produced by *Governing* (2015), the agency that studied how this phenomenon evolved over time across the largest cities of the US. Broadly speaking, they define that a tract can be said to be gentrified if:

- The tract's median household income was in the bottom 40th percentile when compared to all tracts within its metro area at the beginning of the decade (2010).
- The tract's median home value was in the bottom 40th percentile when compared to all tracts within its metro area at the beginning of the decade.

And:

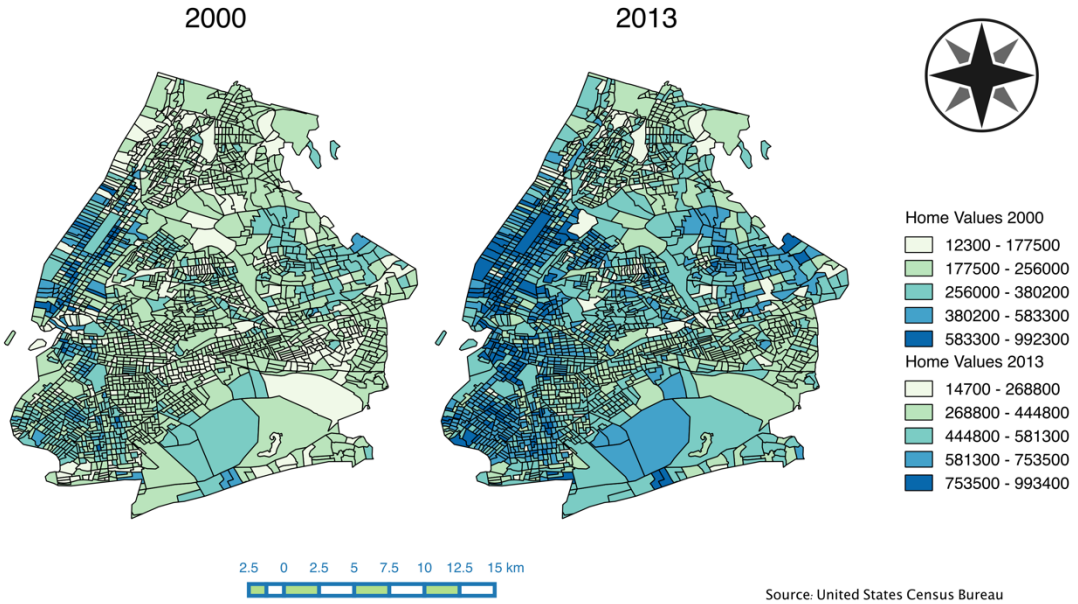
- A tract's median home value increased when adjusted for inflation.
- The percentage increase in a tract's inflation-adjusted median home value was in the top third percentile of all tracts within a metro area.

Using information from the US Census Bureau, specifically from their *Fact Finder* platform, we can use this criterion to identified which tracts within the NYC Metro area have undergone this process. The data we will utilize on this first step is:

- Median household income between 2000 & 2013 (tables DEC_00_SF3_HCT012 & ACS_13_5YR_B19013).
- Median house value between 2000 & 2013 (DEC_00_SF3_H085 & ACS_13_5YR_B25077).

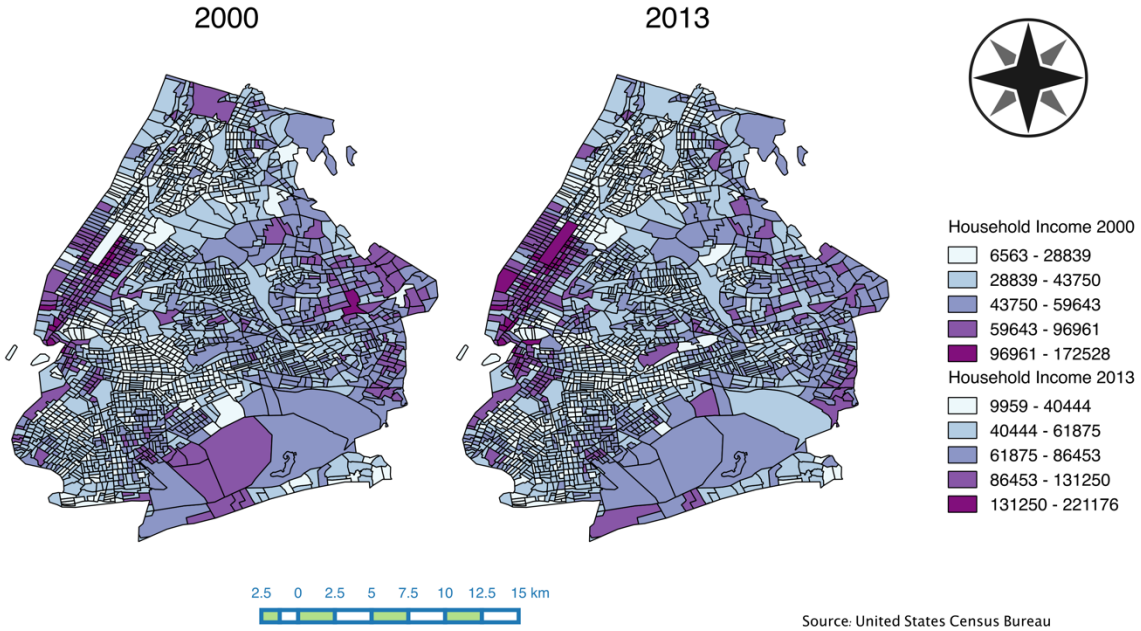
Using the data from 2000, we identify and label those tracts that were on the bottom 40% percentile of the sample. Then, with the 2013 data, we label the tracts that experienced the largest percentage growth in home value between this period. We take the tracts where these labels intersect, and we got the subset of them that we can classify as "gentrified". Utilizing QGIS are able to visually follow this process.

Median Home Values across New York City



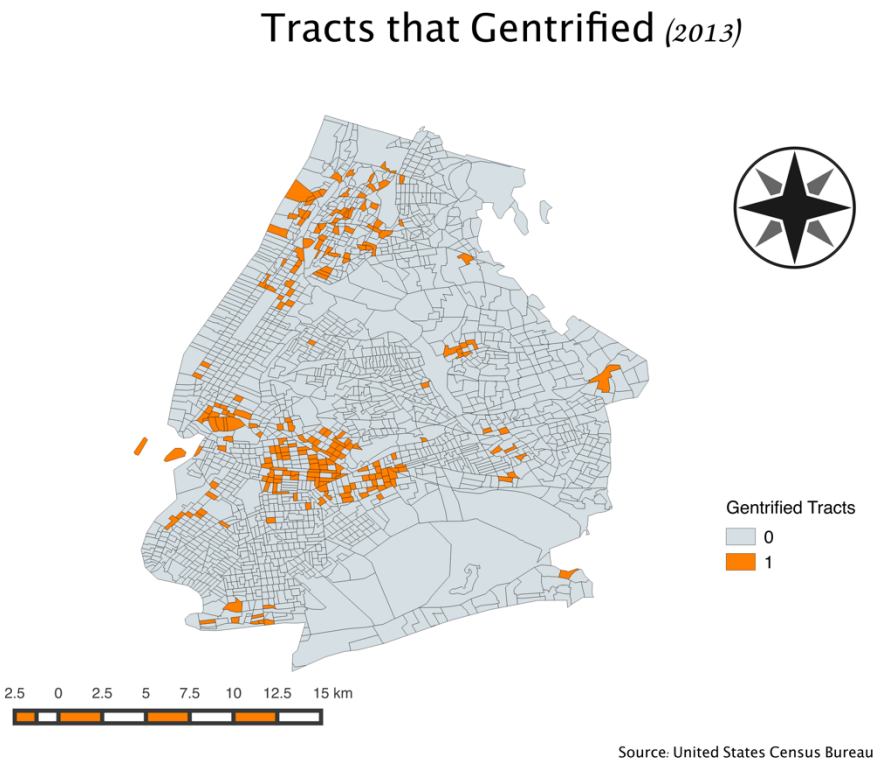
Source: United States Census Bureau

Median House Income across New York City



Source: United States Census Bureau

In this case, the lightest areas from the left panels are those eligible for gentrification as of the year 2000 (i.e. Midtown Manhattan was not). Joining this information with the growth rate up to 2013, we get the following subset of gentrified tracts.



We can see some interesting patterns, with the gentrification occurring mostly on the Bronx and upper Brooklyn areas.

For our analysis, first we will investigate the distribution of our independent variable, mainly to identify if there could be a spatial component to it (are there clusters where it is more likely to occur?). For this we will make use of the *Moran's I*, probably the most widely utilized measure of spatial autocorrelation, defined as:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2},$$

where *N* is the number of observations, w_{ij} an element of the spatial weight matrix, and *X* the variable of interest. A value of zero indicates no correlation, while those close to +1 or -1 indicate

a strong positive or negative spatial correlation, respectively. Making use of the information provided by this indicator we will also be able to build a map where significant clusters are identified.

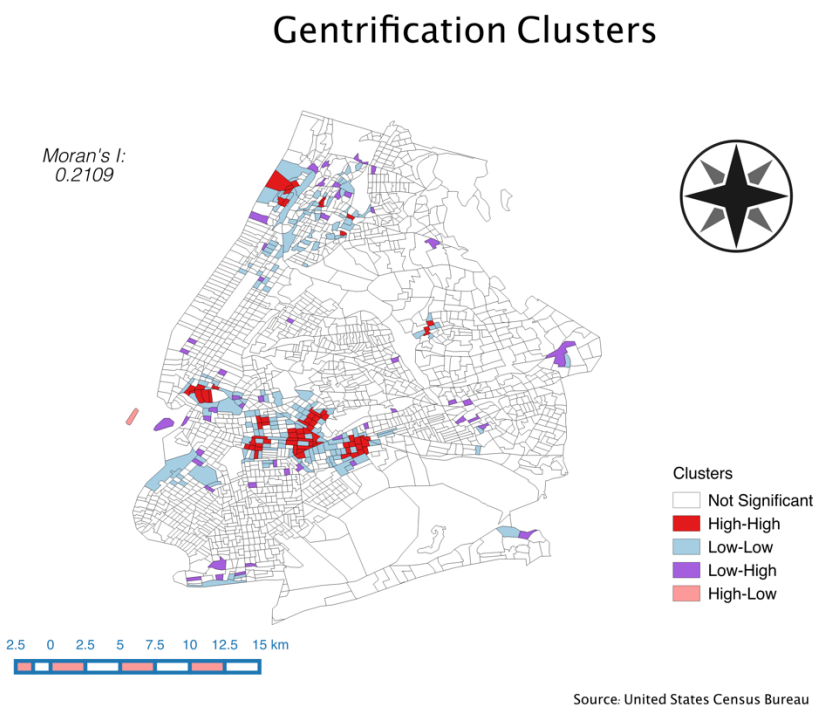
Next, we will then try to asses if there exists a relationship between this process and 3 other social / demographical variables of relevance (as of 2013):

- *% Higher Education.* Measured as percentage of population with a bachelor's degree or higher. The literature (Governing Data, 2014 and London, B. & Palen, J., 1984) suggest that this relationship will be positive.
- *% of Population Engaged in Service Works*. Relates to the work performed by Ley on 1984. When testing the Canadian case he found a positive relationship, so we wish to confirm if in this different scenario it's the same.
- *Mean Age*. This variable has not been explored too much on the literature, so we wish to test against it. Both Kennedy. & Leonard (2001) and London & Palen (1984) suggest that in-movers associated to the gentrification effect tend to be relatively young, displacing the older population, but this point is not fully explored. Here we will test this hypothesis.

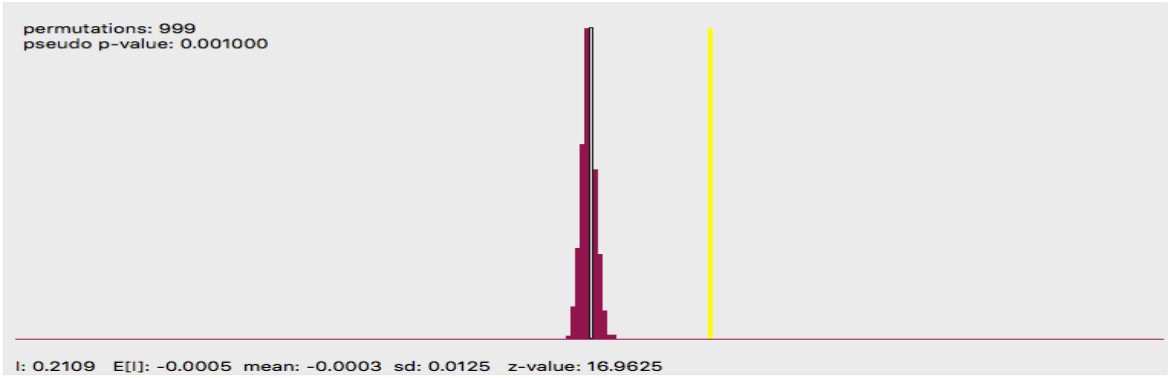
In order to test these relationships we will begin by making a visual inspection of the distribution of the covariates, and test for spatial concentration on them too (Moran's I). After that we will run a traditional OLS model, to test if there is a statistically significance interaction between the response (*Gentrified Tract)* and the predictors. Looking at the initial distribution of the median household income and median house value, we suspect there is some sort of spatial lag effect; if the diagnostic for spatial dependence confirms this, then we will proceed to run the respective correction model (spatial lag or error lag). After this some conclusions will be presented.

Results

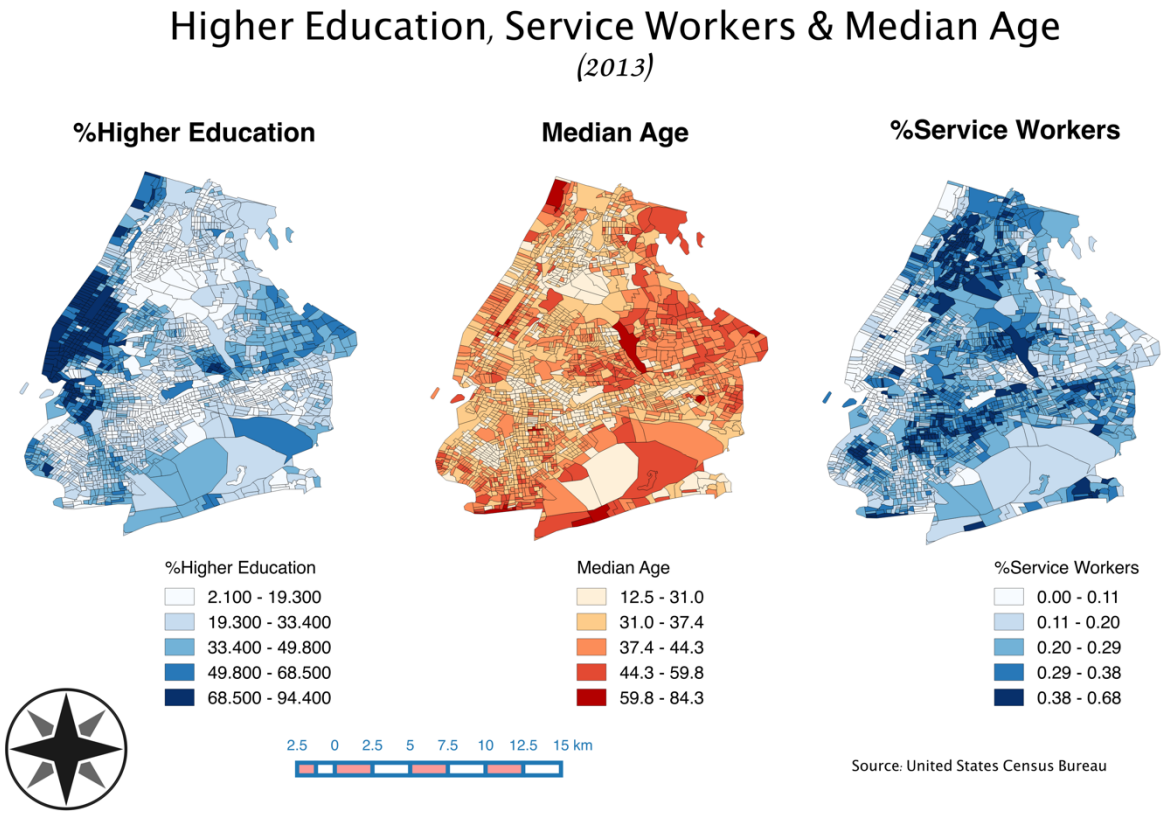
We begin by analyzing a possible clustering effect on our independent variable, the indicator of whether or not a tract gentrified. Using the Moran's I and LISA clustering analysis we get the following results.



A cluster effect does seem to take place, mainly on the Bronx and upper Brooklyn areas (where our visual inspection suggested some form of concentration). We see that these clusters take the form of "High-Highs" and "Low-Lows". The Moran´s I indicator (0.2109) confirms this direct association, with a highly significant *p-value*.



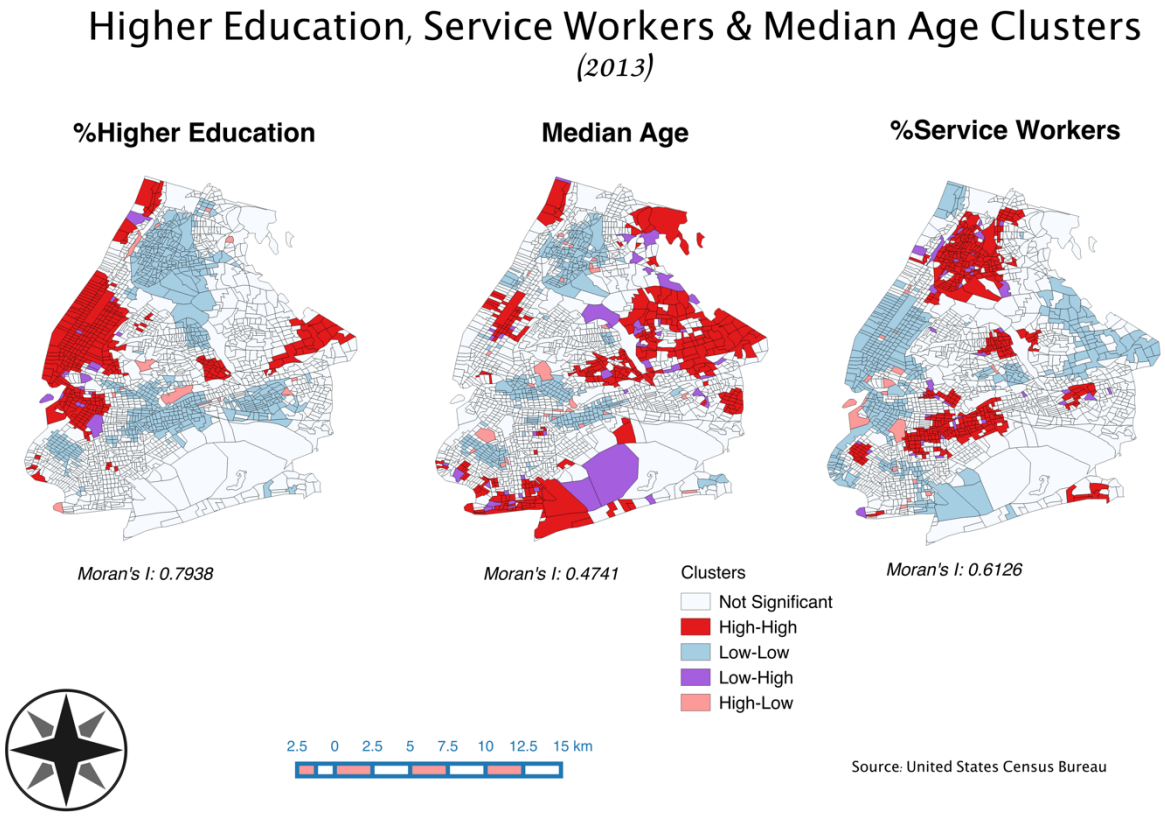
Now we perform a visual analysis of the covariates in space.



We can start seeing some patterns already:

- *%Higher Education* is heavily concentrated in Manhattan. Compared to the gentrified tracts, no discernible relationship is found.
- In general *Median Age* doesn't appear to be so spatially clustered. We do see some regions, particularly on the eastern Kings County region, where there is a larger concentration of high age population.
- High levels for *%Service Workers* seem to be clustered around the gentrified areas, mainly upper Brooklyn and on the Bronx.

Testing for spatial autocorrelation and generating the respective cluster maps will give us a better picture of their distribution, so we proceed to do that.



Compared to our initial cluster analysis on tracts that gentrified, we see that the variable that overlaps the most in this sense is *%Service Workers*, with "High-Highs" on the Bronx and upper Brooklyn.

Since visual results are not conclusive, we proceed with the implementation of the OLS regression, which will also help us test for a possible spatial effects. The results after performing it are shown below.

OLS Regression Estimator

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.1241391	0.05128332	2.420652	0.01558
ED13	-0.0005143144	0.0004978641	-1.033042	0.30172
WK13	0.4137846	0.08825192	4.688675	0.00000
AG13	-0.002907017	0.001003688	-2.896336	0.00382

DIAGNOSTICS FOR SPATIAL DEPENDENCE

```
FOR WEIGHT MATRIX : Weights.gal
   (row-standardized weights)
TEST                           MI/DF        VALUE          PROB
Moran's I (error)              0.1753       14.0373        0.00000
Lagrange Multiplier (lag)        1         200.4901        0.00000
Robust LM (lag)                  1           7.9187        0.00489
Lagrange Multiplier (error)      1         192.8689        0.00000
Robust LM (error)                1           0.2976        0.58539
Lagrange Multiplier (SARMA)      2         200.7877        0.00000
```

Looking at the *p-values* of the estimation, we see that both *%Service Workers* and *Median* Age are significant (<1%). Education seems to hold no relationship, which is in line with what we observed on the map inspection. There seems to be a negative correlation with age (in line with Kennedy, M. & Leonard, P.) and a positive correlation with the percentage of service workers. Additionally, we can see that a spatial effect indeed exists, and we can control best for it by using the Spatial Lag model (Robust LM (lag) with *p-value* of 0.00489).

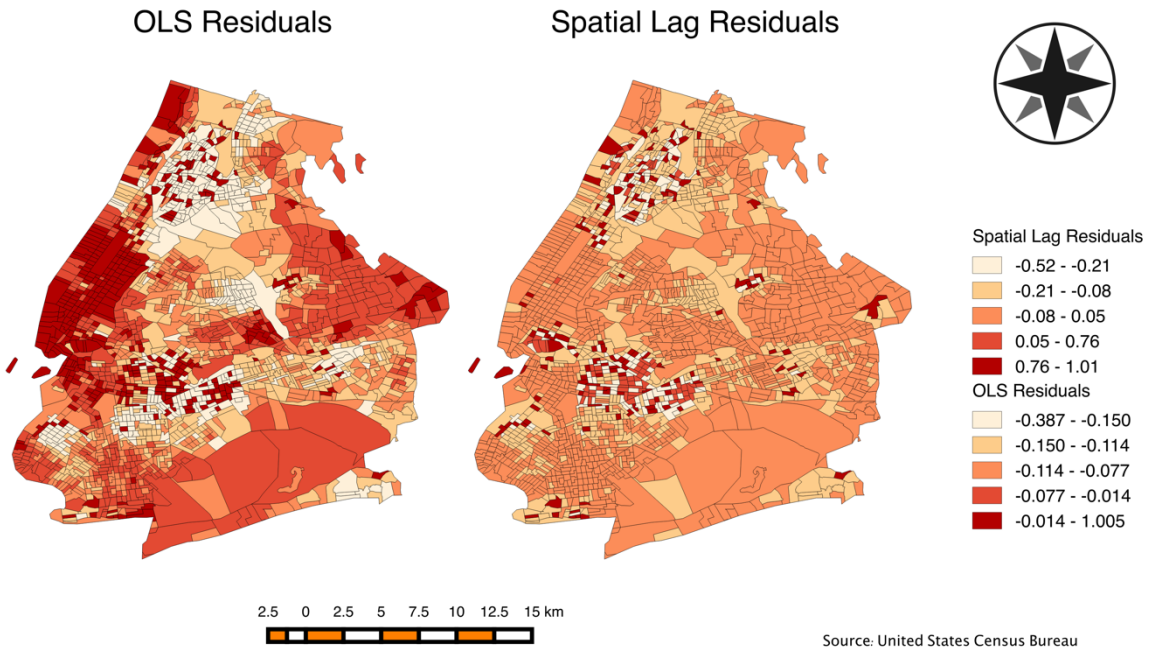
Performing this regression:

SPATIAL REGRESSION ESTIMATOR

```
------------------------------------------------------------------------
    Variable    Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------
  W_ELIGIBLE     0.3800665     0.03095216       12.27916     0.00000
    CONSTANT    0.05678975      0.0490469       1.157866     0.24692
        ED13  -0.0004434239   0.0004757224     -0.9321064     0.35128
        WK13     0.2911792     0.08408642       3.462857     0.00053
        AG13  -0.001321534   0.0009556557      -1.382856     0.16671
------------------------------------------------------------------------
```

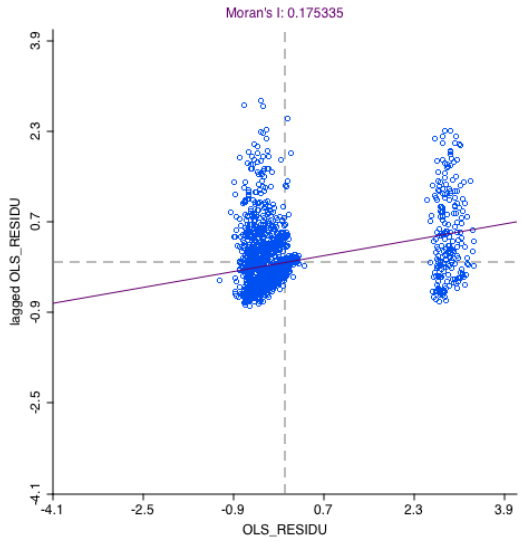
The coefficient for the control variable *W* is highly significant, and now age loses its relevance. We can confirm if indeed a spatial effect took place by looking at the distribution of the residuals across both models.

Residual Distribution across Models

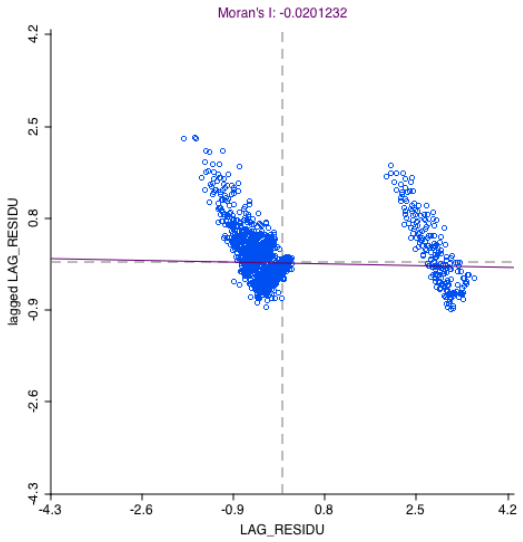


To complement this graph, we present also the Moran's I for each.

OLS Residuals



Spatial Lag Residuals



Conclusions

In this study we were interested in analyzing the phenomenon of *gentrification,* which has received significant media coverage and attracted the academics attention in recent years, but has not been clearly defined nor the social and demographic effects associated with it. This paper approached the subject by first defining the New York Metro area as the region of study, and then implementing a technique utilized by a well-known reporting agency (*Governing*) to identify tracts that have gentrified over the last 15 years. This measure is mostly based on two economic variables: median house price and median household income. We then try to asses if there is a relationship between this phenomenon and 3 other social/demographic variables, two of which have been already covered by the academic literature on previous years and on different regions. The variables are: the percentage of the population with a higher education degree, the percentage of population working on the service industry, and the median age. We try to see if in the case of NYC any significant relationship holds.

Or study suggested that of the variables examined, only the percentage of population working on the service industry is closely associated with the fact that a tract gentrified, while age and the education level do not seem to be related. There is also a spatial component to the process, as confirmed by the implementation of the spatial lag model. Our results imply that regions that have recently gentrified tend to focus heavily on the service industry, which is a hypothesis that Ley formulated back on 1986 by suggesting that changes in the economic base (towards a service industry) are closely related to the gentrification process. Gentrification also seems to have a propagation feature, where zones near others that are experiencing this phenomenon are more likely to follow this path as well.

For future studies, it is suggested that a spatial *logistic* model is tested (not currently available on Geoda), as it is a better approach for modeling probabilities (i.e. the probability that a tract gentrifies). Comparing if the results vary would be an interesting exercise. Additionally, future researchers are advised to test the regression analysis by incorporating more variables identified

on the literature, as our OLS output indicated that an error component was also prevalent in the estimation, suggesting that variables might be missing from the analysis.

Bibliography

Bierbaum, A.; Chapple, K.; et al. (2015) “Gentrification, Displacement and the Role of Public Investment: A Literature Review”. Federal Reserve Bank of San Francisco (http://www.frbsf.org/community-development/files/wp2015-05.pdf).

Bolitzer, B.; Netusil, N. (2000). “The Impact of Open Spaces on Property Values in Portland, Oregon”. Journal of Environmental Management, Vol. 59 (http://www.sciencedirect.com/science/article/pii/S0301479700903517).

Freeman, L. (2014). “Displacement or Succession? Residential Mobility in Gentrifying Neighborhoods”. (http://uar.sagepub.com/content/40/4/463.abstract).

Governing Data (2014). “Gentrification Report Methodology”. (http://www.governing.com/gov-data/gentrification-report-methodology.html).

Hays, Z.; Kreager, D.; Lyons, C. (1986). “Urban Revitalization and Seattle Crime, 1982–2000”. Social Problems, Vol. 58 (http://www.jstor.org/stable/10.1525/sp.2011.58.4.615).

Kennedy, M.; Leonard, P. (2001). “Dealing with Neighborhood Change: A Primer on Gentrification and Policy Choices”. The Brookings Institution Center on Urban and Metropolitan Policy: The Discussion Paper Series (http://www.brookings.edu/research/reports/2001/04/metropolitanpolicy).

Ley, D. (1986). “Alternative Explanations for Inner-City Gentrification: A Canadian Assessment”. Annals of the Association of American Geographers, Vol. 76 (http://www.jstor.org/stable/2562708).

London, B.; Palen, J. (1984). "Gentrification, Displacement, and Neighborhood Revitalization". Sunny Press.

Nara, A.; Torrens, P. (2007). “Modeling gentrification dynamics: A hybrid approach”. Computers, Environment and Urban Systems, Vol. 31 (http://www.sciencedirect.com/science/article/pii/S0198971506000718).

Smith, N. (1987). "Gentrification and the Rent Gap". Annals of the Association of American Geographers, Vol. 77 (http://www.jstor.org/stable/2563279).

Snow, C.; Turner, M. (2001). “Leading Indicators of Gentrification in D.C. Neighborhoods”. Urban Institute (http://webarchive.urban.org/publications/900461.html).