Spatial Demography ***2014*** 2(1): 1-29
http://spatialdemography.org

ISSN 2164-7070 (online)

RESEARCH

A Spatially Explicit Approach to the Study of Socio-Demographic Inequality in the Spatial Distribution of Trees across Boston Neighborhoods

Dustin T. Duncan[1], Ichiro Kawachi[1], Susan Kum[2], Jared Aldstadt[2], Gianfranco Piras[3], Stephen A. Matthews[4], Giuseppe Arbia[5], Marcia C. Castro[6], Kellee White[7], David R. Williams[1, 8]

[1] Department of Society, Human Development, and Health, Harvard School of Public Health, Boston, MA USA; [2] Department of Geography, University at Buffalo, State University of New York, Buffalo, NY USA ; [3] Regional Research Institute, West Virginia University, Morgantown, WV USA; [4] Department of Sociology, Department of Anthropology, and Population Research Institute, The Pennsylvania State University, University Park, PA USA; [5] Department of Statistical Sciences and Institute of Hygiene and Public Health, Faculty of Economics, Catholic University of the Sacred Heart, Rome, Italy; [6] Department of Global Health and Population, Harvard School of Public Health, Boston, MA USA; [7] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC USA; [8] Departments of African and African American Studies, and Sociology, Harvard University, Cambridge, MA USA

ABSTRACT

The racial/ethnic and income composition of neighborhoods often influences local amenities, including the potential spatial distribution of trees, which are important for population health and community wellbeing, particularly in urban areas. This ecological study used spatial analytical methods to assess the relationship between neighborhood socio-demographic characteristics (i.e. minority racial/ethnic composition and poverty) and tree density at the census tract level in Boston, Massachusetts (US). We examined spatial autocorrelation with the Global Moran's *I* for all study variables and in the ordinary least squares (OLS) regression residuals as well as computed Spearman correlations non-adjusted and adjusted for spatial autocorrelation between socio-demographic characteristics and tree density. Next, we fit traditional regressions (i.e. OLS regression models) and spatial regressions (i.e. spatial simultaneous autoregressive models), as appropriate. We found significant positive spatial autocorrelation for all neighborhood socio-demographic characteristics (Global Moran's *I* range from 0.24 to 0.86, all *P*=0.001), for tree density (Global Moran's *I*=0.452, *P*=0.001), and in the OLS regression residuals (Global Moran's *I* range from 0.32 to 0.38, all *P*<0.001). Therefore, we fit the spatial simultaneous autoregressive models. There was a negative correlation between neighborhood percent non-Hispanic Black and tree density (r_S=-0.19; conventional *P*-value=0.016; spatially adjusted *P*-value=0.299) as well as a negative correlation between predominantly non-Hispanic Black (over 60% Black) neighborhoods and tree density (r_S=-0.18; conventional *P*-value=0.019; spatially adjusted *P*-value=0.180). While the conventional OLS regression model found a marginally significant inverse relationship between Black neighborhoods and tree density, we found no statistically significant relationship between neighborhood socio-demographic composition and tree density in the spatial regression models. Methodologically, our study suggests the need to take into account spatial autocorrelation as findings/conclusions can change when the spatial autocorrelation is ignored. Substantively, our findings suggest no need for policy intervention vis-à-vis trees in Boston, though we hasten to add that replication studies, and more nuanced data on tree quality, age and diversity are needed.

KEYWORDS: neighborhood racial/ethnic composition, neighborhood poverty, racial/socioeconomic segregation, trees, spatial demography, spatial econometrics, Boston, US

Trees, Population Health and Community Wellbeing: Benefits of Trees

A multitude of physical, ecological, social, aesthetic and economic benefits of trees have been widely recognized and documented (Dwyer et al., 1992; Tyrväinen et al. 2005; Sarajevs,

Corresponding Author: Dustin T. Duncan, ScD, Harvard School of Public Health, Department of Society, Human Development, and Health, 677 Huntington Avenue, Kresge Building 7th Floor, Boston, MA 02115. Tel: 617-384-8732. Fax: 617-384-8730. Email: dduncan@hsph.harvard.edu

2011). Trees can improve and promote both *population health* and *community wellbeing*. Trees contribute to population health by: improving in air quality by exchanging and absorbing various gases and airborne pollutants (e.g. particulate matter), which can have implications for respiratory health, including lower rates of asthma (Lovasi et al., 2008; Nowak et al., 2006); providing shade and cooling ambient air temperature by diverting solar radiation and evapotranspiration, which in turn, helps minimize exposure to harmful ultraviolet radiation from the sun and reduce heat-related stress and deaths (Georgi and Zafiriadis, 2006; Basu and Samet, 2002; Heisler and Grant, 2000; Brown University, 2010; Oka, 2011); reducing stress and promoting more relaxed physiological states (Kaplan and Kaplan, 2003; Velarde et al., 2007); making open spaces more pleasant, encouraging physical activity (Bedimo-Rung, 2005; Sullivan et al., 2004; Corti et al., 1996); and mitigating the effects of noise (Fang and Ling, 2005; Gidlöf-Gunnarsson, 2007). Additionally, trees are associated with the reduced risk of poor pregnancy outcomes, i.e. a reduction in the number of small for gestational age births (Donovan et al., 2011) and improved child development (Taylor et al. 1998). With regards to community wellbeing, trees have been found to be associated with stronger social connections between neighbors (Sullivan et al., 2004) and greater perceived neighborhood safety (Kuo et al. 1998, Kuo, 2003), higher property values (Donovan and Butry, 2010; Payton et al., 2008), more favorable responses to shopping at local retailers (Wolf 2003; Wolf, 2005) and energy savings (Simpson, 2002). Other community-level benefits of trees include decreased incidents of property crime and violent crime (Kuo and Sullivan 2001) and increased attachment to place (Dwyer et al. 1992; Kuo et al. 1998).

Although these benefits of trees are applicable across diverse environments, arguably, they are more pronounced in urban geographies. The expansion of urban areas has been attributed to significant changes in climate and air quality. Impervious surfaces and buildings in urban areas, as well as, reduced vegetation cover, have created urban "heat islands" where evaporative cooling is inhibited and heat is stored, leading to temperature increases (Foley et al., 2005). Changes in emission patterns and atmospheric conditions enhanced by urbanization and the dependency on transportation modes requiring fossil fuels in urban areas, have compromised air quality (Foley et al. 2005; Srinivasan et al., 2003). Moreover, the increasingly isolated and sedentary lifestyles adopted by urbanites have negative social, economic and health consequences (Srinivasan et al., 2003). In summary, trees are an important aspect of neighborhoods related in multiple ways to population health and community wellbeing, and are particularly important in urban areas.

Residential Segregation: Implications for the Spatial Distribution of Trees

Racial and socio-economic residential segregation, i.e. the physical separation of certain population subgroups in space, are normative across communities in the United States (US), New Zealand, Brazil and other countries (Massey and Denton, 1993). While there have been substantial reductions in the magnitude of residential segregation in the US, the degree of segregation, specifically Black-White segregation, in several large metropolitan areas such as New York City, Chicago, Detroit, and Boston remain high (Logan and Stults 2011). Explanations for the persistence of high residential segregation among US Blacks may be in part attributable to explicit and implicit discrimination in housing and mortgage lending (Fix and Struyk 1993; Galster, 1987; Reibel, 2000). Additionally, local land and density zoning regulations are increasingly being recognized as an important contributor to perpetuating racial segregation (Rothwell, 2011; Lichter et al., 2007). Residential segregation (and the persistence of) can create neighborhoods with no access or inferior access to public services, amenities and resources (Marsh et al., 2010; Massey and Denton, 1993), including the distribution of trees across urban areas.

Neighborhood Socio-Demographic

Composition and the Spatial Distribution of Trees: Literature Review

Variations in the composition and configuration of urban trees within and between urban areas can be attributed to the confluence of current and historical conditions of biophysical and socio-demographic factors. While biophysical factors, such as temperature, soil and precipitation are important in explaining the differences in the abundance and diversity of particular species of trees among different urban areas, social contextual factors may be more influential in explaining the spatial distribution of trees within specific urban areas (Conway et al., 2011; Luck et al., 2009), especially areas with residential segregation and/or high concentrations of certain racial/ethnic groups. That is, in urban areas the relationship may be more pronounced. Social processes influencing neighborhood composition may have implications for the allocation of resources and management of urban trees. It is possible, for example, that the reputation of a neighborhood (which may be based on factors such as racial/ethnic concentrations) may influence the decisions of service planners and investors to locate certain amenities, such as trees, in particular neighborhoods. Therefore, given the importance of trees and potential inequitable spatial distribution of trees by neighborhood socio-demographic composition, we reviewed the literature on the relationship between the spatial distribution of trees and neighborhood socio-demographic composition.

Much of the empirical studies that indicate relationships between the inequitable spatial distribution of tree cover and socio-demographic factors have been of North American urban areas with recognized socio-demographic inequality (Kirkpatrick et al., 2011). Indeed, urban areas in the US, particularly those where large-scale gentrification has not occurred, the post-industrial core areas reflect residual residential and occupational segregation whereby racial/ethnic minority and socioeconomically disadvantaged residents are clustered together in amenity poor neighborhoods (Pickett et al. 2008; Wolch et al. 2005). The '*inequality hypothesis*' posits that environmental amenities, such as trees, are unevenly distributed across space among different socio-demographic groups. That is, a lower proportion of tree cover is expected in neighborhoods comprised of a higher proportion of racial/ethnic minorities and socioeconomically disadvantaged residents. Consequently, the benefits from urban trees may vary by geographic areas and population groups (Heynen et al. 2006; Landry and Chakraborty 2009). As such, residents of neighborhoods comprised of a higher proportion of racial/ethnic minorities and socioeconomically disadvantaged individuals may receive fewer benefits from urban trees as compared to residents in neighborhoods comprised of predominantly White and more affluent residents (Heynen et al. 2006; Landry and Chakraborty 2009).

Household income measures and other measures of neighborhood socioeconomic standing have consistently (for the most part) been found to be positively associated with urban trees. For example, neighborhood poverty is associated with fewer urban trees (Crawford et al. 2008; Heynen et al., 2006; Iverson and Cook, 2000; Jensen et al., 2004; Landry and Chakraborty, 2009; Loukaitou-Sideris and Steiglitz, 2002; Kirkpatrick et al., 2011; Lovasi et al., 2008; Neckerman et al., 2009; Pedlowski et al., 2002; Szantoi et al., 2012; Tooke et al., 2010; Wolch et al., 2005). "Poor" neighborhoods, based on percentage of resident living in poverty, have fewer street trees as compared to "non-poor" neighborhoods in New York City (Lovasi et al. 2008; Neckerman et al., 2009). Measures of urban tree cover were found to be positively correlated with household income levels with increases in urban land and lowest tree cover in lower income neighborhoods within the six-county Chicago, Illinois metropolitan region (Iverson and Cook 2000). Highest urban forest per unit area (urban forest is a general term to refer to a large amount of trees found in urban areas) was measured in areas with higher average annual household income in Miami-Dade County, Florida (Szantoi et al., 2012). A positive correlation between urban trees and median household income was found in Tampa, Florida (Landry and Chakraborty, 2009), New Orleans, Louisiana (Talarchek, 1990), Milwaukee, Wisconsin (Heynen et al. 2006; Perkins et al., 2004), Haute, Indiana (Jensen et al. 2004) and Phoenix, Arizona (Jenerette et al., 2007). Fewer trees in inner city areas, typically lower income and minority neighborhoods, have been noted in Los Angeles, California (Loukaitou-Sideris and Steiglitz, 2002; Wolch et al. 2005).

The relationship between urban tree cover and neighborhood income has also been established in several urban areas outside the US, although less established than US-based research. Public open spaces with more amenities, such as trees, were found to be in more affluent neighborhoods in Melbourne, Australia (Crawford et al. 2008). Household prosperity, as measured by house price in 1961 and median household income in 2006, was strongly associated with urban trees in both time periods in six east Australian cities (Kirkpatrick et al., 2011). In another Australian city (Ballarat), however, Kendal et al. (2012) found that education level (not household income) was associated with tree cover, i.e. areas of higher tree cover were associated with a higher proportion of residents with graduate education. Interestingly, the study also found that individuals with higher incomes and lower education levels settled in areas with newer housing developments with lower tree cover (Kendal et al., 2012). An uneven distribution of the abundance and diversity of trees among neighborhoods by income was detected in Rio de Janeiro, Brazil (Pedlowski et al., 2002). Median income, which provided the strong relationship with overall vegetation (not specific to but including trees), was positively correlated with vegetation in the three largest urban areas of Canada – Montreal, Toronto and Vancouver; areas of high vegetation and high income, as well as, areas of low vegetation and low income were identified in all three cities (Tooke et al., 2010). A socio-economic measure based on various factors such as average annual income, educational attainment, and vehicle ownership found that less affluent neighborhoods had lower tree cover, but the highest percentage of public trees in Santiago, Chile (Escobedo et al., 2006). It is important to note that some studies, however, did not find significant effects of measures of socioeconomic standing and trees. For example, a study that was conducted in three US cities (i.e. Birmingham AL, Houston TX and Los Angeles CA) found no significant association between neighborhood poverty and the spatial distribution of trees (Franzini et al., 2010).

Studies that have been conducted examining potential relationships between neighborhood racial/ethnic composition and the spatial distribution of trees have produced less consistent findings and have been limited to urban areas in the US (Flocks et al., 2011; Franzini et al., 2010; Heynen et al., 2006; Landry and Chakraborty, 2009; Lovasi et al., 2008). Lower proportions of public street trees were found in Black neighborhoods in Tampa, Florida (Landry and Chakraborty, 2009). Similarly, Black neighborhoods were found to have the lowest tree density, but the greatest percentage of street trees and potential for planting space for new trees among predominantly White, Black, and Hispanic neighborhoods in Miami Dade County, Florida (Flocks et al., 2011). While citywide urban tree canopy cover was negatively correlated with Hispanic neighborhoods, there was no signification correlation with Black neighborhoods in Milwaukee, Wisconsin (Heynen et al., 2006). Likewise, although there was no significant correlation between street tree density and neighborhood-level percent Black, neighborhood-level percent Latino was negatively correlated, but this association did not persists when accounting for other neighborhood characteristics in New York City (Lovasi et al., 2008). Franzini et al. (2010) found no association between minority racial/ethnic composition (i.e. predominantly Black and Hispanic neighborhoods) and trees in Birmingham AL, Houston TX and Los Angeles CA, but found that racially and ethnically mixed neighborhoods (defined as "no majority racial or ethnic group"), interestingly, had a higher density of trees.

The findings from the existing literature underscore the complexity of interacting forces associated with the distribution of trees in urban areas. While the aforementioned studies have been important in elucidating the existence of a relationship between neighborhood socio-demographic composition and spatial distribution of trees, some of the studies are investigating trees while others are investigating vegetation, which may be a reason that there are mixed results. It is also important to note that many of these studies provide evidence of associations without accounting for potential confounding covariates, so residual confounding and spurious associations are probable. Furthermore, most studies that perform multivariate analyses do not consider possible "spatial effects", specifically, spatial autocorrelation. The presence of spatial autocorrelation, i.e. interdependencies among observations in variables that exhibit a systematic pattern in attribute values due to spatial proximity, violates the classical statistics assumption of independence of observations, leading to biased parameter estimates, which influences subsequent interpretations of statistical significance (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b). Hence, there is a need to examine, and if necessary, account for spatial autocorrelation. We are aware of only two previously published studies examining the relationship between spatial distribution of trees and neighborhood socio-demographic composition that evaluated spatial autocorrelation (Landry and Chakraborty, 2009; Kendal et al., 2012). Additionally, we note that the vast majority of studies, especially in the same study, have not considered there may be threshold effects of neighborhood socio-demographic composition. Studies usually calculate and examine neighborhood racial/ethnic composition as a continuous variable despite the potential that the level of a dominant racial/ethnic group in a neighborhood may matter in where and whether amenities exist in a neighborhood. Indeed, a certain level of minority neighborhood racial/ethnic composition (e.g. 60% minority) or a certain level of poor residents in a neighborhood (e.g. 20% poor) may be necessary to influence where trees are located. We also recognize that spatial relationships between neighborhood socio-demographic characteristics and trees can vary across spatial contexts, which is another potential reason for the mixed results. Finally, we recognize that inconsistent results in the relationship between neighborhood socio-demographics and trees could be due in part to different measures of trees. Methods used for estimating trees include such measuring trees via census-takers counting street trees (e.g. Lovasi et al., 2008) and field collection of various tree measures of randomly selected samples (e.g. Pedlowski et al. 2002) as well as via remotely sensed methods including Landsat™ (e.g. Iverson and Cook, 2002), high-resolution satellite imagery (e.g. IKONOS) (e.g. Landry and Chakraborty, 2009), and aerial photography (e.g. Escobedo et al., 2006; Heynen et al., 2006; Kendal et al., 2012; Kirkpatrick et al., 2011).

Conceptual Framework and Study Purpose

This research is motivated by the 'inequality hypothesis' positing that environmental amenities, such as trees, are unevenly distributed across space among different socio-demographic groups, whereby a lower proportion of trees may exist in neighborhoods comprised of a higher proportion of racial/ethnic minorities and socioeconomically disadvantaged residents (Heynen et al. 2006; Landry and Chakraborty 2009). The 'inequality hypothesis' lends well to the examination of threshold effects, i.e. an association between an exposure and an outcome will be detected above a specified threshold value but none below it, among socio-demographic characteristics. Using a spatial analytical approach, the objective of this study was to evaluate the spatial distribution of trees across neighborhoods in Boston, Massachusetts, with a particular focus on whether there are socio-demographic disparities in tree locations, including the consideration of threshold effects. There are several reasons why Boston is an illustrative case-study. First, Boston is a racially, ethnically and socio-economically diverse city. Additionally, over the past three decades, Boston has consistently ranked as one of the top metropolitan areas in the US with a high degree of residential segregation (Logan and Stults 2011;

Iceland et al., 2002). This is based on several measures of segregation including the dissimilarity index (Duncan and Duncan, 1955), the most commonly used measure of segregation that measures the relative distributions across neighborhoods within a city (or metropolitan area) between two racial/ethnic groups (Logan and Stults 2011; Iceland et al., 2002). Specifically, in terms of Black-White segregation in the 50 metropolitan areas with the largest Black populations in 2010, the Boston-Quincy metropolitan area ranked eleventh and it ranked fourth in terms of Hispanic-White segregation in the 50 metropolitan areas with the largest Hispanic populations in 2010 (Logan and Stults, 2011). Importantly, Boston is also an understudied city in research on socio-demographic disparities and we are not aware of any study conducted in Boston examining socio-demographic disparities in trees, which highlights the need for such research. Boston was also chosen because we had excellent tree data.

METHODS

Spatial Area and Spatial Unit of Analysis

The spatial area for this study is Boston in Massachusetts, which is located in the northeastern corner of the United States (US). It is one of the oldest cities in the US, the largest city in Massachusetts and the largest city in New England, including six states: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island and Connecticut. Boston is also the cultural and economic hub of New England. The city of Boston includes a total area of 89.63 square miles (232.14 km^2); 54.03% of which is land. According to the 2010 US census, Boston has a total population of 617,594 and the Greater Boston area has an estimated population of over 4.5 million people across its neighborhoods. Consistent with much social science research in the US, the census tract was our spatial unit (i.e. definition of neighborhoods). As an example, census tracts have been used in Boston neighborhood research (Subramanian et al. 2006, Subramanian et al. 2005; Krieger et al. 2003, Krieger et al. 2002; Duncan et al., 2012). Census tracts in the US have an average of approximately 4,000 people and are designated to be homogenous spatial units "with respect to population characteristics, economic status, and living conditions" (US Census Bureau, 2012). The 2010 census tract boundaries were used in this study. There are 181 census tracts in Boston. However, consistent with past neighborhood research in Boston (Cradock et al. 2005; Chen et al. 2006; Duncan et al., 2012), we excluded the sparsely populated Harbor Islands (census tract # 980101) and a census tract that includes only the Massachusetts Bay (census tract # 990101). It is important to highlight that the small population (i.e. 535 individuals) on Harbor Islands in Boston are not typical of the rest of the city (there is a detoxification center on the Harbor Islands with permanent residents) and water does not include any individuals. These restrictions were important, because contiguity matters in spatial analysis of areal data, making the analytic sample the contiguous 2010 census tracts in Boston (n=179). We noticed that some of the remaining census tracts have very few residents (as low as 0 people). Therefore, also consistent with previous socio-demographic disparities research (Block et al. 2004; Duncan et al., 2012), our analytical sample was further restricted to those census tracts with >500 people (n=167). The removal of census tracts with extremely small populations removes missing/withheld American Community Survey data (which was used in this study) and ensures that census tracts with extremely small populations would not bias the results. Our final analytic sample of census tracts in Boston represents over 90% of its census tracts.

Spatial Socio-Demographic Data and Data for Control Variables

The socio-demographic data used in this study are percent non-Hispanic Black, percent Hispanic, percent of families in poverty and population density (i.e. total population per square kilometer) in census tracts. All of the demographic variables were extracted from the 2010 US Census. The socioeconomic disadvantage (i.e. poverty) data came from the

2006-2010 American Community Survey. These data were downloaded from Social Explorer. Percent of non-Hispanic Black residents and percent of Hispanic residents were measured both as continuous variables and categorical variables. We defined a neighborhood as over 60% Black and Hispanic as predominantly Black and Hispanic neighborhoods, respectively, which is consistent with prior research (Franco et al. 2008; Moore et al. 2008; Moore and Diez Roux 2006). Percent of families below poverty was also measured both as a continuous variable and a categorical variable. Consistent with prior research, high poverty neighborhoods were defined as at least 20% of families in poverty (Subramanian et al. 2005; Kelly et al., 2007; Franzini et al., 2010). Besides socio-demographic data, we collected data on privately or publicly owned protected and recreational open spaces (e.g. parks, playing fields) from the Office of Geographic Information (MassGIS), Commonwealth of Massachusetts, Information Technology Division (http://www.mass.gov/mgis), which was current as of January 2012. We then calculated open space density (i.e. open space per square kilometer for each census tract) in the Massachusetts state plane projection North American Datum (NAD) 1983.

Tree Data

Data on tree locations across Boston neighborhoods comes from the City of Boston, Department of Innovation and Technology. The GIS data collected included trees located on streets and in parks. These tree locations were mapped from high resolution (3" pixel) aerial photography flown in April of 2011. Some of the trees fell outside the Boston border because a 50 feet buffer around the city border was used (to make sure the data was collected in full). The aerial photos showed that there were 200,465 trees included in the shapefile (though, as expected, a few trees fell just outside the Boston boundary). Within Boston's boundaries, there were 199,490 trees. As stated previously, various methods can be and have been used to estimate trees including via remotely sensed methods including Landsat™. Using aerial photos, a form of remote sensing, for estimating the spatial distribution of trees has several important advantages including it can be more accurate when satellite data has a low resolution and estimating trees via aerial photos has been applied previously in research on socio-demographic inequality in the spatial distribution of trees (e.g. Escobedo et al., 2006; Heynen et al., 2006; Kendal et al., 2012; Kirkpatrick et al., 2011) as well as other research (e.g. Uuttera et al., 1998). Using the tree data, we calculated census tract-level tree density (i.e. trees per square kilometer for each census tract), using the Massachusetts state plane projection NAD 1983.

Spatial Statistical Analytic Plan

Taking an explicit spatial perspective to our research question ("What is the relationship between neighborhood socio-demographic composition and the spatial distribution of trees?"), we designed a comprehensive spatial modeling plan (outlined below). The methods used in this research are motivated by Tobler's First Law of Geography, which states, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970), which provides a clear reference to spatial effects, particularly spatial dependence. Thus, the motivation for the analytical strategy is accounting for spatial autocorrelation, if necessary and appropriate.

Geovisualization

After computing descriptive statistics for the study variables (e.g. means, standard deviations, ranges), we conducted geovisualization (GIS mapping), a method of exploratory spatial data analysis. Geovisualization should be the first step in any spatial analysis, as it can help researchers better understand potential spatial patterns in their data. All geovisualization was done in ArcGIS version 10 (ESRI, Redlands, CA). We created a choropleth map of the spatial patterns in the distribution of trees across Boston neighborhoods (census tracts) clipped to Boston's natural boundaries. Map colors were based on Color Brewer 2.0, a web-based tool for selecting map color schemes (Brewer and Harrower, 2012). We used the Jenks natural breaks classification method for this map. This classification method determines the best grouping of values in the data, by reducing the variance within classes, while maximizing the variance between classes (Jenks, 1967). We also present maps of census tracts that are predominantly Black, Hispanic and poor (colors for these maps were also based on Color Brewer 2.0) (Brewer and Harrower, 2012).

Global Spatial Autocorrelation

The next step in exploratory spatial data analysis is to test global spatial autocorrelation. We computed the Global Moran's *I* (Moran, 1950), which is the most commonly used test for global spatial autocorrelation (Cliff and Ord, 1981; Bailey and Gatrell 1995; Waller and Gotway 2004).

The Global Moran's *I* equation is:

$$I=\left(\frac{N}{\Sigma_i \Sigma_j w_{ij}}\right)\frac{\Sigma_i \Sigma_j w_{ij}\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)}{\Sigma_i\left(Y_i-\bar{Y}\right)^2}$$

where I is the Global Moran's I value; N is the total number of neighborhoods; Y is the variable of interest (tree density for a given census tract); $\bar{Y}$ is the mean of that variable; w_{ij} is the spatial weights matrix which is a measure of the closeness of areas i and j (one if i and j are neighbors, and zero if i and j are not neighbors); and the subscripts i and j refer to different neighborhoods (so Y_i and Y_j are observations for neighborhood i and j, respectively) (Cliff and Ord, 1981; Bailey and Gatrell 1995; Waller and Gotway 2004).

The Moran's I values range approximately between -1 to 1. A Moran value near zero suggests no overall spatial pattern, no spatial autocorrelation (the null hypothesis of complete spatial randomness). A negative spatial autocorrelation coefficient reflects neighboring areas with large inverse values–e.g. large values and small values are neighbors (i.e. dissimilarity). To illustrate, a significant negative Moran statistic would indicate that dissimilar levels of the distribution of trees, i.e. neighborhoods with high amounts of trees would have neighboring areas with low amounts of trees or vice-versa. A positive spatial autocorrelation coefficient reflects neighboring areas with similarly high or low values (i.e. similarity). To illustrate, significant positive Moran statistic would indicate that trees cluster in space, i.e. the number of trees in adjacent tracts are similar such that neighborhoods with high amounts of trees would have neighboring areas that would also have high amounts of trees and areas with low amounts of trees would have neighboring areas that also have low amounts of trees. Little theoretical and empirical work has been done to provide guidance with respect to choosing the "correct" spatial weights matrix (Anselin and Bera, 1998); and it has been suggested that one should experiment with different weighting matrixes. However, LeSage and Pace (2010) find little theoretical basis for the diffused evidence that estimates and inferences from spatial regression models are sensitive to a particular specification of the weights matrix. In this study, we employed a row-standardized binary contiguity spatial weights matrix based on the first-order Queen criteria for the Global Moran's *I* calculations, which is frequently done with areal data and which was used in our previous Boston neighborhood research (Duncan et al, 2012). The Queen's contiguity spatial weights matrix defines neighbors as census tracts that share a common boundary or a corner. A pseudo *P*-value for the Moran's *I* was calculated via a Monte Carlo simulation consisting of 999 random replications.

Correlation between Socio-demographic Characteristics and Tree Distribution

As our study variables had a non-normal distribution, we computed non-parametric Spearman correlations between socio-demographic characteristics and tree distribution. Because the existence of spatial autocorrelation is a violation of the independence assumption, if the variables exhibited spatial autocorrelation we computed the Spearman correlation accounting for spatial autocorrelation. It is important to account for spatial autocorrelation in correlation analysis since spatial autocorrelation can result in incorrect degrees of freedom in the conventional correlation tests of the significance, which may lead to a bias in the estimation of significance of effects (Clifford and Richardson 1985; Student 1914; Haining 1991). This is related to the concept of effective sample size (see, e.g. Cressie, 1993; Schabenberger and Gotway, 2008). Specifically, in a bivariate correlation analysis when positive spatial autocorrelation is present for both variables, the probability of Type I error exceeds the specified level (Clifford and Richardson 1985; Haining 1991). On the other hand, the specified level of significance is too conservative in the traditional bivariate correlation analysis when there is positive spatial autocorrelation in one variable and negative spatial autocorrelation in another, increasing the likelihood of the Type II error. In this study, we used the Clifford and Richardson effective sample size adjustment method to account for spatial autocorrelation in the Spearman correlation coefficients, which employs spatial correlation matrices for each variable to jointly measure the dependence between observations (Clifford and Richardson 1985; Haining 1991). Therefore, this methodology employs a spatial correlation matrix for each variable to adjust the effective sample size of the bivariate correlation test. We specifically used first order through sixth order queen contiguity matrices for generating empirical spatial correlation matrices if spatial autocorrelation was found in study variables (Haining 1991). Due to the adjusted sample size, the corresponding t-statistics and p-values change—sometimes dramatically. We report the Spearman correlation coefficients (r_S) and significance values.

Regression Analyses: Traditional A-Spatial and Spatial Methodologies

Preliminary data analyses indicated that dependent variable, tree density, was skewed. We computed a natural logarithm transformation on tree density to reduce the skewness. We developed a forward step-wise spatial modeling strategy, starting with a non-spatial model (Florax et al., 2003). This modeling strategy consists of estimating a standard linear regression model and then performing various Lagrange Multiplier (LM) tests for spatial autocorrelation on the residuals of this linear model. Although we are aware of potential problems deriving from pre-testing issues, this is a standard approach in the literature applying spatial econometric methodologies. As a first step, therefore, we fit traditional ordinary least squares (OLS) regression models for these log-linear models.

It is known that neglecting spatial autocorrelation, when present, can produce biased parameter estimates and/or incorrect standard errors (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b). Thus, when the data generation process is a spatial lag model the parameter estimates and standard errors both can be incorrect (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b). On the other hand, under the spatial error model, the OLS coefficients, while still unbiased, become inconsistent (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b). The spatial lag and spatial error models are well-known and widely used in the field of spatial econometrics to control for spatial autocorrelation in the data (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b). Thus, in the circumstance of substantive spatial autocorrelation (which is due to a substantive spatial process), the spatial lag model is applied. In contrast, in the instance of nuisance spatial autocorrelation (which is caused by spatial autocorrelation that arises from spillovers or spatial mismatch, i.e. the spatial scope of the phenomenon of interest does not match the scope of the spatial units used in the analysis), the spatial error model is applied. Because the causes of potential spatial inequality are unknown, both spatial models have some justification. Therefore, we estimated both the spatial error model and the spatial lag model, as suggested by the results of the LM tests. Moreover, this is consistent with available literature in that prior spatial socio-demographic disparities research has implemented both spatial regression models (Smiley et al. 2010).

The Spatial Error Model equation can be represented by the following equation:

$$y = \beta X + \varepsilon\text{, } where\ \varepsilon = \lambda W\varepsilon + u$$

Where y is the dependent variable; β is a vector of regression parameters associated with the explanatory variables matrix X; λ (lamba) is the spatial autoregressive coefficient that indicates the extent to which the spatial component of the errors are correlated with each other; and u is an independent error term. Therefore, in the model the overall error is decomposed into two parts, a spatially uncorrelated error term and, the term indicating the spatial component of the error term (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b). The interpretation of the coefficients in a spatial error model is the same as in a linear regression model.

On the other hand, the Spatial Lag Model can be represented by the following equation:

$$y = \rho W y + \beta X + \varepsilon$$

where y is the dependent variable; ρ (rho) is the spatial autoregressive coefficient for the lagged dependent variable Wy (a weighted average of the dependent variable of the neighboring locations); β is a vector of regression parameters associated with the explanatory variables matrix X; and ε is an independent and identically distributed error term. In the spatial lag model, therefore, spatial autocorrelation is introduced in the form of the spatially dependent variable (LeSage and Pace 2009; Ward and Gleditsch 2008; Waller and Gotway 2004; Bailey and Gatrell 1995; Anselin and Bera 1998; Anselin, 1988b).

We used the row-standardized first-order Queen's spatial weights matrix when computing the Global Moran's *I* statistic and the LM test for both spatial regression models to evaluate the OLS regression residuals for evidence of spatial autocorrelation (LeSage and Pace 2009; Anselin et al. 1996; Anselin and Bera 1998; Anselin 1988a; Anselin, 1988b). The Moran's *I* statistic is applied to the error terms of the OLS model to detect spatial autocorrelation. Spatial autocorrelation is evident when the Moran coefficient significantly deviates from zero. When the Moran's *I* is statistic significant, the LM test for spatial lag and spatial error dependence is used. The LM test statistic with the highest value and lowest *P*-value suggests the proper specification for the data. However, it is possible for the LM spatial error and LM spatial lag to both be significant. In this case, the robust forms are used to guide spatial model specification. Therefore, the LM test also suggests which spatial model should be used—spatial lag or spatial model (Florax et al., 2003). The OLS and spatial model were compared using the Akaike Information Criterion (AIC) (Akaike 1974). The AIC examines overall model fit and model complexity. A smaller AIC value indicates a better goodness-of-fit. If spatial error models were fit, we lastly computed the spatial Hausman test to compare the magnitude of the OLS and spatial error model coefficients (LeSage and Pace 2009, Pace and LeSage 2008). It is important to note that the AIC, which is easily interpretable, can only be computed when spatial models are estimated via maximum likelihood, which is the reason why spatial regression models were first estimated via maximum likelihood and the Jacobian in these models was computed using the Ord (1975) approximation.

Different spatial model estimating techniques can influence study findings and the assumptions of maximum likelihood might not be correct (e.g. homoskedasticity). We tested the assumptions required for maximum likelihood in the spatial models. Heteroskedasticity was assessed via the

Breusch-Pagan test (Breusch and Pagan, 1979). In the spatial econometrics literature, recently more flexible specifications have appeared that also require fewer model assumptions. In the circumstance of heteroskedasticity specifically, a variety of spatial models can be computed. All of these models are estimated through a two-stage least squares method where the spatially lagged right-hand side variables are used as instruments for the (endogenous) spatial lag term (Kelejian and Prucha, 1998). As an example of these more flexible specifications, in the present paper we consider the lag model with an error term representation given by (Kelejian and Prucha, 2007):

$$e = Ru$$

where u is a vector of innovations and R is an nxn non-stochastic matrix whose elements are not known. The model does not impose any specific structure on the disturbance process and is consistent with different forms of Heteroskedasticity and Autocorrelation (HAC) as well as with general patterns of spatial correlation. The spatial HAC estimation, although rarely applied in spatial demography, is a cutting-edge innovation for accounting for heteroskedasticity in a spatial framework (Piras, 2010). Kelejian and Prucha (2007) suggest estimating the elements of the variance covariance matrix in terms of the two-stage least square residuals, the instruments and a kernel function. The kernel function is defined in terms of a distance matrix that, along with a bandwidth, limit the number of sample covariances. In our application we use a variable bandwidth based on the fourth nearest neighbor distance (Anselin and Lozano-Gracia, 2008).

The regression modeling strategy included bivariate models followed by multivariate models, including all neighborhood socio-demographic characteristics, population density and open space density. The spatial distribution of the population may be a confounder in the studied relationship and is often included as a confounding covariate in socio-demographic disparities in neighborhood amenities research. Trees may be more likely to be located in open spaces, and there may be systematic variation in neighborhood socio-demographics by open spaces. Multivariate analyses including all neighborhood socio-demographic characteristics were also conducted in light of potential suppressor effects and/or interactive effects. We used the R statistical software version 2.15 (R Core Team, 2012). The *spdep* and *sphet* packages were used for all spatial data analyses in this study (Bivand et al., 2008; Bivand 2011; Piras, 2010). The designated level statistical significance was P <0.05.

RESULTS

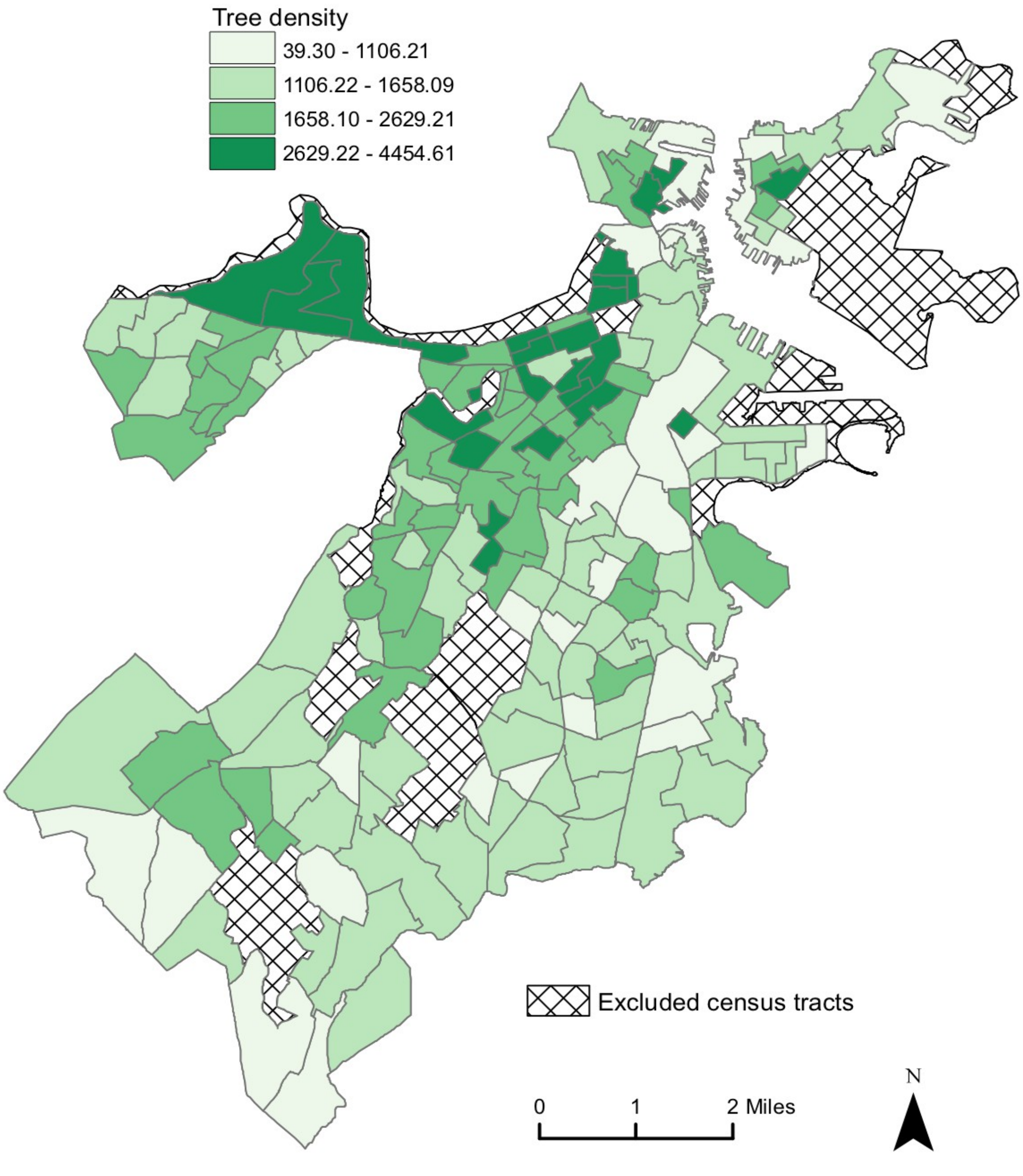
Descriptive Statistics

The mean tree count per census tract was 1025.50 (SD=941.23; range= 3 to 6731) and the mean tree density per census tract was 1792.44 (SD=813.32), with a range of 39.30 to 4455.00 for the analytic sample of 167 census tracts (Table 1). A large range exists for the neighborhood socio-demographic characteristics, especially for census tract percent non-Hispanic Black. The mean for census tracts that were predominantly non-Hispanic Black was 0.12 (n=20 census tracts), for predominantly Hispanic was 0.03 (n=5 census tracts), and predominantly poor was 0.31 (n=51 census tracts).

Geovisualization and Global Spatial Autocorrelation

Figure 1 shows a map of the spatial distribution of tree density across census tracts in Boston, which suggests spatial patterning—with trees being more abundant in Allston/Brighton, Fenway and Back Bay neighborhoods of Boston. The Global Moran's I for tree density was 0.45 (P=0.001) (Table 1). There was also positive significant global spatial autocorrelation in all of the neighborhood socio-demographic characteristics (Global Moran's I range from 0.24 to 0.86, all P=0.001). Figure 2, Figure 3 and Figure 4 shows the spatial distribution of predominantly Black, Hispanic, and poor census tracts in Boston,

Figure 1. Spatial Distribution of Trees Across Boston Neighborhoods (Census Tracts)



Note: Tree density was categorized in ArcGIS using jenks natural breaks classification methodology (n=4). Map colors from http:// www.colorbrewer2.org, by Cynthia A. Brewer, Penn State Geography

Table 1. Descriptive Statistics and Global Spatial Autocorrelation (n=167)

	Mean (SD)	Range	Moran's *I*	*P*-value
Tree Density	1792.44 (813.32)	39.30-4454.61	0.453	0.001
Log of Tree Density	7.38 (0.52)	3.67-8.40	0.373	0.001
Percent non-Hispanic Black	21.63 (24.08)	0.16-83.99	0.859	0.001
Percent Hispanic	17.70 (14.78)	0.68-66.60	0.706	0.001
Percent Families in Poverty	15.54 (14.81)	0.00-64.91	0.331	0.001
Predominantly non-Hispanic Black	0.120 (0.326)	0-1	0.633	0.001
Predominantly Hispanic	0.030 (0.171)	0-1	0.713	0.001
Predominantly Families in Poverty	0.305 (0.462)	0-1	0.244	0.001
Population Density	25110.91 (17271.54)	3444.00-110100.00	0.456	0.001
Open Space Density	0.07 (0.08)	0.00-0.59	0.115	0.012

respectively, as examples to show the stark spatial patterns in the neighborhood socio-demographic characteristics. To illustrate, predominantly Black neighborhoods are spatially concentrated in center of Boston, i.e. in the Roxbury, Dorchester and Mattapan neighborhoods of Boston. In contrast, predominantly Hispanic neighborhoods are spatially clustered in East Boston neighborhood. Visually comparing the spatial distribution of trees and the spatial patterns of predominantly Black census tracts suggests there may be a negative association.

Correlation Between Neighborhood Socio-Demographic Composition and Trees

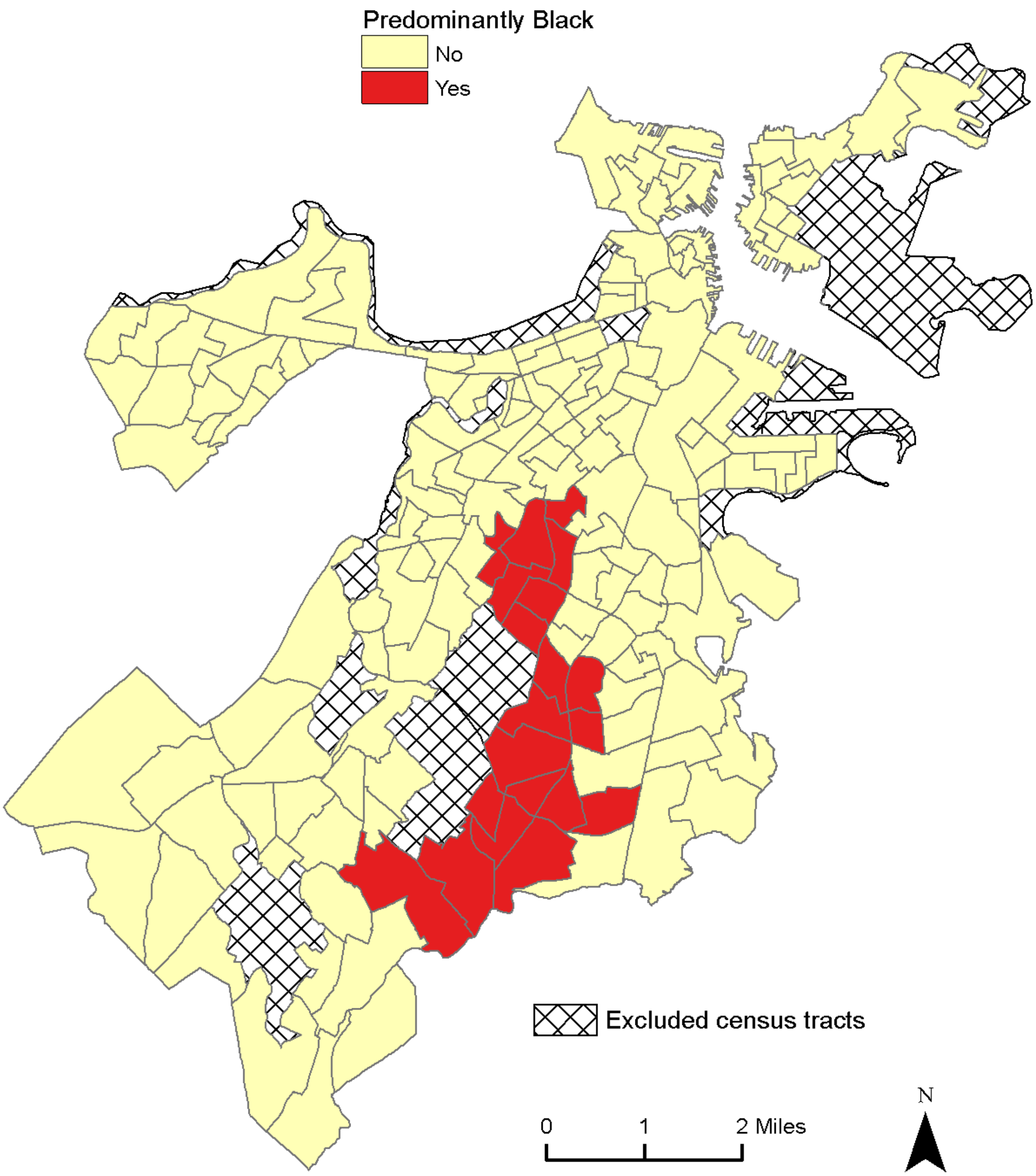
Table 2 shows all Spearman correlations. The Spearman correlations and the Spearman correlations accounting for spatial autocorrelation produced different findings; the *P*-values were more conservative when spatial autocorrelation was accounted for. Specifically, as an example, the *P*-values were only significant in the conventional significance correlation analyses between census tract percent non-Hispanic Black and tree density (r_S=-0.19; conventional *P*-value=0.016; spatially adjusted *P*-value=0.299) as well as between predominantly non-Hispanic Black census tracts and tree density (r_S=-0.18; conventional *P*-value=0.019; spatially adjusted *P*-value=0.180).

Spatial Regression Analyses of the Relationship Between Neighborhood Socio-Demographic Composition and Trees

The Moran's *I* for the OLS regression residuals were positive and significant across all models, including the bivariate and multivariate models (Global Moran's *I* range from 0.32 to 0.38, all *P*<0.001), which suggests significant spatial structure in the OLS residual residuals (Table 3). For example, the Global Moran's *I* for the multivariate model of percent neighborhood socio-demographic characteristics and log of tree density was 0.32 (*P*<0.001). The Lagrange Multiplier (LM) tests for the spatial models were significant, also confirming the need for spatial models (Table 3). Specifically, the LM lag and LM error were both significant across all models (all *P*<0.001). Across some models, the robust LM lag remained statistically significant while the robust LM error often did not. Generally, findings suggest that the spatial lag model was most appropriate.

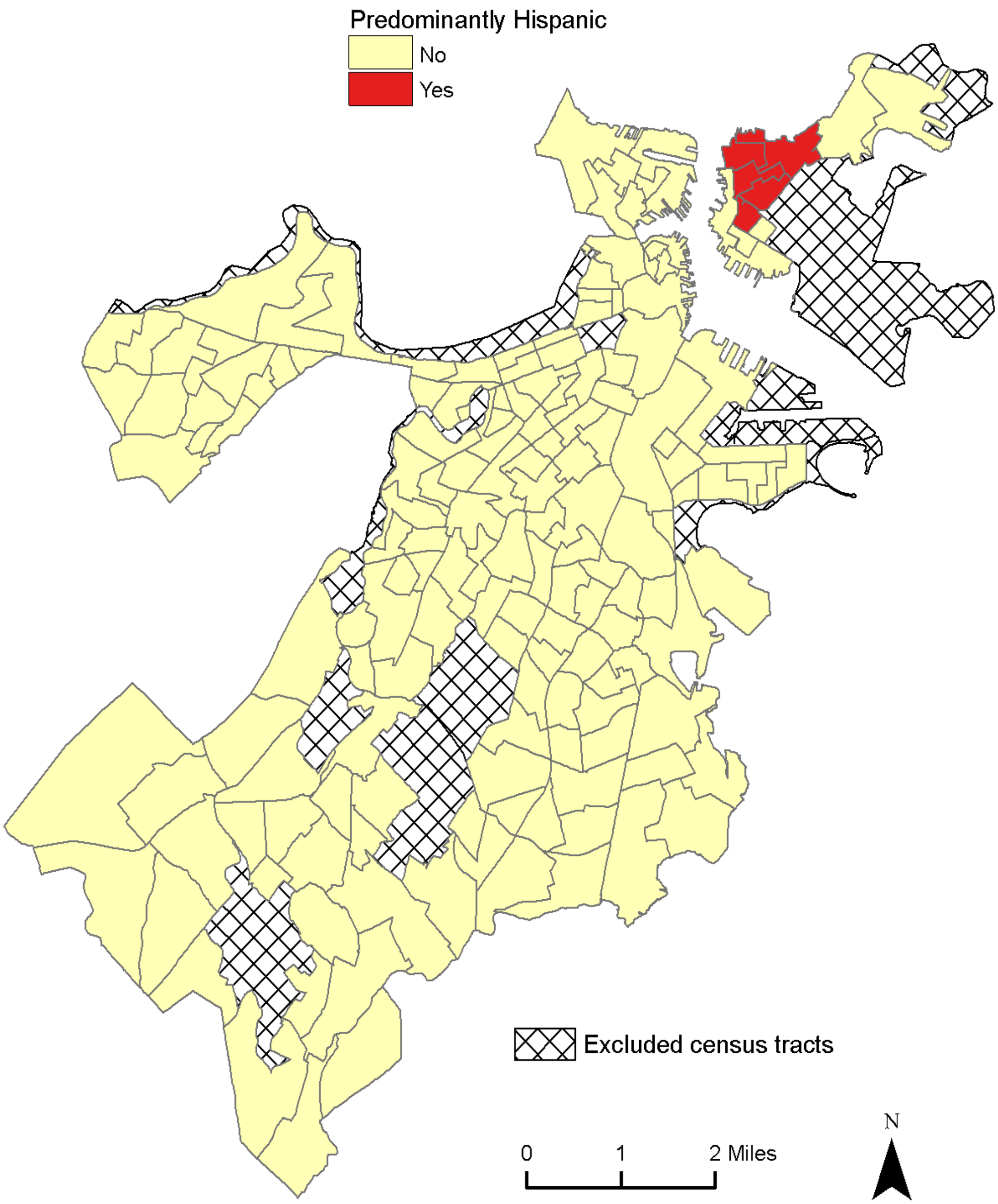
Compared to the OLS models (Table 4), the AIC values were lower in the maximum likelihood estimated spatial lag models (Table 5). For example, in the OLS multivariate model of percentages of the socio-demographic

Figure 2. Spatial Distribution of Predominantly Black Boston Neighborhoods (Census Tracts)



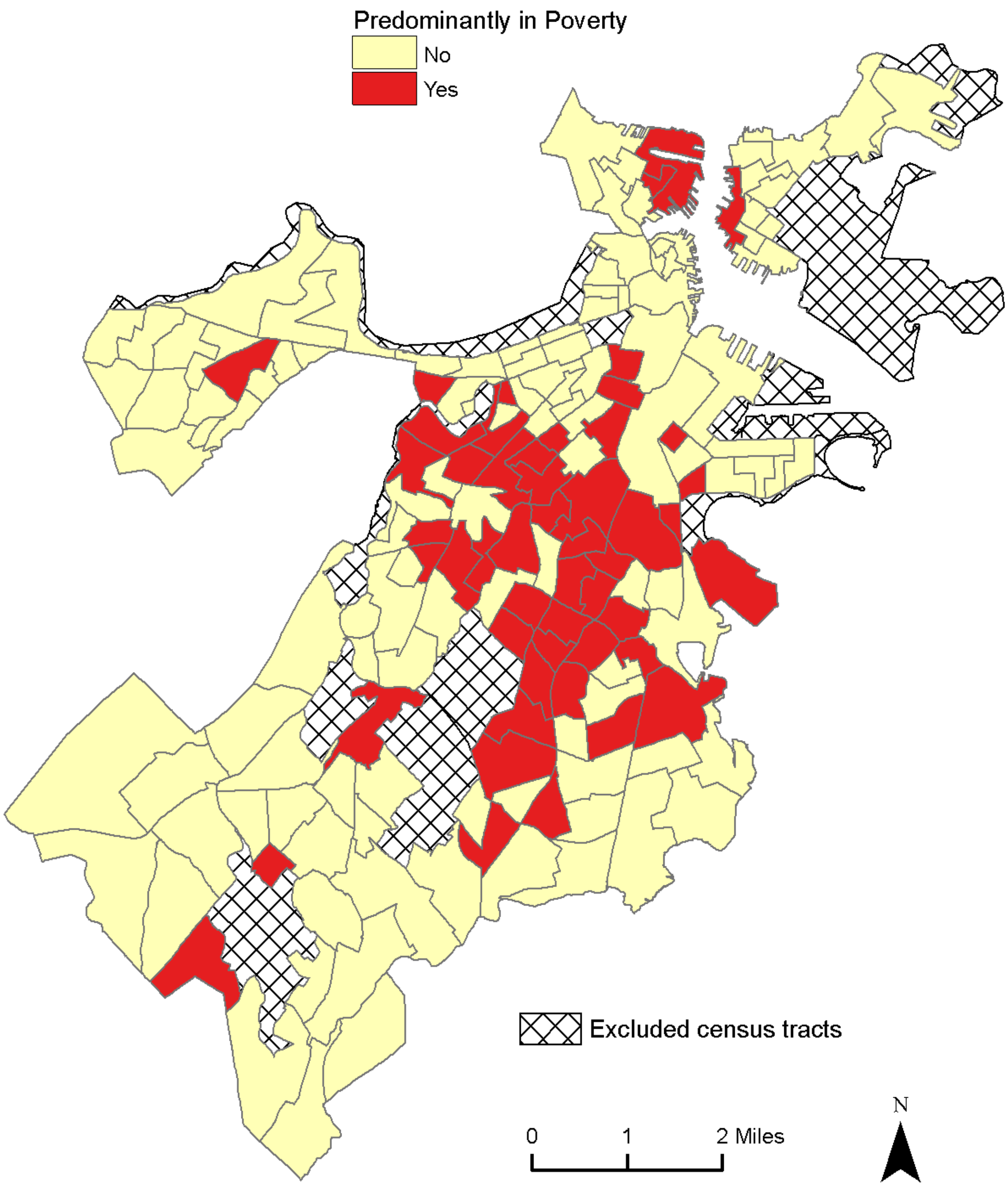
Note: Map colors from http:// www.colorbrewer2.org, by Cynthia A. Brewer, Penn State Geography

Figure 3. Spatial Distribution of Predominantly Hispanic Boston Neighborhoods (Census Tracts)



Note: Map colors from http:// www.colorbrewer2.org, by Cynthia A. Brewer, Penn State Geography

Figure 4. Spatial Distribution of Predominantly Poor Boston Neighborhoods (Census Tracts)



Note: Map colors from http:// www.colorbrewer2.org, by Cynthia A. Brewer, Penn State Geography

Table 2. Spearman Correlation Between Neighborhood-Level Socio-Demographic Characteristics and Tree Density

	r_S	Conventional *P*-value	Spatially adjusted *P*-value
Percent non-Hispanic Black	-0.186	0.016	0.299
Percent Hispanic	-0.084	0.281	0.594
Percent Families in Poverty	-0.031	0.691	0.810
Predominantly non-Hispanic Black	-0.181	0.019	0.180
Predominantly Hispanic	0.031	0.688	--
Predominantly Families in Poverty	0.022	0.782	0.851

Note: Due to the small sample size of predominantly Hispanic neighborhoods, the spatially adjusted correlation between neighborhood predominantly Hispanic and tree density was not able to be computed.

characteristics as predictors of tree density logged the AIC was 259.61 while the AIC of the spatial lag multivariate maximum likelihood estimation of this relationship between percentages of the socio-demographic characteristics as predictors of log of tree density was 225.52. In these models, the rho values (which incorporates the spatial effects) are significant across all spatial lag model maximum likelihood estimations (rho: all approximately 0.06, all *P*<0.001). However, there was heteroskedasticity across most spatial lag maximum likelihood estimated models, including all multivariate models. Since spatial lag models were determined to be most appropriate and because there was heteroskedasticity, we applied the instrumental variable (IV) two-stage least squares model for the spatial lag model with HAC standard errors.

In the two-stage least squares spatial lag bivariate and multivariate models estimated with HAC standard errors applied to the relationship between neighborhood socio-demographic composition and log of tree density, we found no statistically significant relationships (Table 6). Only for comparison purposes do we present the OLS models and the maximum likelihood spatial lag models for the studied relationships (Tables 4 and 5, respectively). In the OLS bivariate and multivariate models, census tract percent Black was associated with reduced tree density (both *P*<0.060). However, it is important to remember that OLS models do not account for spatial autocorrelation, demonstrating that not accounting for spatial autocorrelation can lead to incorrect inferences. The findings of the spatial HAC estimation were near identical to the parameter estimates obtained via maximum likelihood (e.g. magnitude of most coefficients), though the correct model for this set of data is applying HAC standard errors.

DISCUSSION AND CONCLUSION

In this study, we sought to examine the relationship between neighborhood socio-demographic characteristics and the spatial distribution of trees using a spatial perspective. Our motivation was due in part because past research has been equivocal but also because the prior work, with two exceptions (Landry and Chakraborty, 2009; Kendal et al. 2012), has been a-spatial, and there was good reason to suspect that a spatial approach might be necessary. Not surprisingly, we found significant positive spatial autocorrelation in the neighborhood socio-demographic characteristics, tree density and also in the OLS regression residuals. While there was a negative correlation between neighborhood percent non-Hispanic Black and tree density as well as a negative correlation between neighborhood predominantly non-Hispanic Black and tree density, these relationships were not significant when spatial autocorrelation was accounted for. Additionally, while the conventional OLS regression model found a

Table 3. Global Moran's *I* and Lagrange Multiplier (LM) Diagnostics for Spatial Autocorrelation in OLS Model Estimation of the Relationship Between Neighborhood Socio-Demographic Characteristics and Log of Tree Density

	Global Moran's *I*	*P*-value	LM Spatial Error Model Value (*P*-value)	LM Spatial Lag Model Value (*P*-value)	Robust LM SEM Value (*P*-value)	Robust LM SLM Value (*P*-value)
Bivariate Estimation						
Percent non-Hispanic Black	0.348	<0.001	45.184 (<0.001)	46.906 (<0.001)	2.618 (0.106)	4.340 (0.037)
Percent Hispanic	0.372	<0.001	51.690 (<0.001)	51.824 (<0.001)	6.603 (0.010)	6.737 (0.009)
Percent Families in Poverty	0.373	<0.001	51.883 (<0.001)	51.916 (<0.001)	0.033 (0.856)	0.066 (0.798)
Multivariate Estimation						
Multivariate model is controlled for population density, open space density and the other percent neighborhood socio-demographic characteristics	0.316	<0.001	37.213 (<0.001)	41.416 (<0.001)	3.283 (0.070)	7.486 (0.006)
Bivariate Estimation						
Predominantly non-Hispanic Black	0.359	<0.001	48.180 (<0.001)	49.444 (<0.001)	0.870 (0.351)	2.134 (0.144)
Predominantly Hispanic	0.375	<0.001	52.423 (<0.001)	52.122 (<0.001)	1.112 (0.292)	0.811 (0.368)
Predominantly Families in Poverty	0.375	<0.001	52.609 (<0.001)	52.339 (<0.001)	0.503 (0.478)	0.233 (0.629)
Multivariate Estimation						
Multivariate model is controlled for population density, open space density and the other predominantly neighborhood socio-demographic characteristics	0.339	<0.001	43.027 (<0.001)	45.391 (<0.001)	0.698 (0.404)	3.062 (0.080)

Table 4. OLS Model Estimation of the Relationship Between Neighborhood Socio-Demographic Characteristics and Log of Tree Density

Bivariate Estimation				*Bivariate Estimation*			
	Coefficient (SE)	*P*-value	AIC		Coefficient (SE)	*P*-value	AIC
Percent non-Hispanic Black	-0.003 (0.001)	0.055~	257.96	Predominantly non-Hispanic Black	-0.187 (0.124)	0.133	259.40
Percent Hispanic	-0.000 (0.003)	0.950	261.68	Predominantly Hispanic	0.098 (0.238)	0.681	261.52
Percent Families in Poverty	0.003 (0.003)	0.347	260.79	Predominantly Families in Poverty	0.042 (0.088)	0.633	261.46
Multivariate Estimation				*Multivariate Estimation*			
	Coefficient (SE)	*P*-value	AIC		Coefficient (SE)	*P*-value	AIC
Percent non-Hispanic Black	-0.004 (0.002)	0.054~		Predominantly non-Hispanic Black	-0.172 (0.128)	0.183	
Percent Hispanic	-0.001 (0.003)	0.763		Predominantly Hispanic	0.034 (0.239)	0.887	
Percent Families in Poverty	0.005 (0.003)	0.093~	259.61	Predominantly Families in Poverty	0.073 (0.091)	0.423	262.43
~ $p < 0.10$; * $p < 0.05$ (bold); ** $p < 0.01$ (bold)							

Multivariate models are controlled for population density, open space density and the other neighborhood socio-demographic characteristics.

Table 5. Spatial Lag Model Maximum Likelihood Estimation of the Relationship Between Neighborhood Socio-Demographic Characteristics and Log of Tree Density

Bivariate Estimation				*Bivariate Estimation*			
	Coefficient (SE)	*P*-value	AIC		Coefficient (SE)	*P*-value	AIC
Percent non-Hispanic Black	-0.001 (0.001)	0.438	220.68	Predominantly non-Hispanic Black	-0.084 (0.105)	0.421	220.63
Percent Hispanic	0.002 (0.002)	0.364	220.46	Predominantly Hispanic	0.111 (0.199)	0.577	220.96
Percent Families in Poverty	0.002 (0.002)	0.347	220.39	Predominantly Families in Poverty	0.059 (0.074)	0.420	220.62
Multivariate Estimation				*Multivariate Estimation*			
	Coefficient (SE)	*P*-value	AIC		Coefficient (SE)	*P*-value	AIC
Percent non-Hispanic Black	-0.002 (0.002)	0.311		Predominantly non-Hispanic Black	-0.088 (0.108)	0.416	
Percent Hispanic	0.002 (0.003)	0.536		Predominantly Hispanic	0.095 (0.201)	0.638	
Percent Families in Poverty	0.003 (0.003)	0.325	225.52	Predominantly Families in Poverty	0.079 (0.076)	0.302	226.15
~ $p < 0.10$; * $p < 0.05$ (bold); ** $p < 0.01$ (bold)							

Multivariate models are controlled for population density, open space density and the other neighborhood socio-demographic characteristics.

Table 6. Spatial Two Stages Least Square Lag Model Estimation with HAC Standard Errors of the Relationship Between Neighborhood Socio-Demographic Characteristics and Log of Tree Density

Bivariate Estimation			*Bivariate Estimation*		
	Coefficient (SE)	*P*-value		Coefficient (SE)	*P*-value
Percent non-Hispanic Black	-0.002 (0.002)	0.359	Predominantly non-Hispanic Black	0.040 (0.091)	0.660
Percent Hispanic	0.002 (0.003)	0.456	Predominantly Hispanic	0.138 (0.231)	0.550
Percent Families in Poverty	0.002 (0.003)	0.402	Predominantly Families in Poverty	0.006 (0.220)	0.979
Multivariate Estimation			*Multivariate Estimation*		
	Coefficient (SE)	*P*-value		Coefficient (SE)	*P*-value
Percent non-Hispanic Black	-0.000 (0.001)	0.949	Predominantly non-Hispanic Black	-0.016 (0.075)	0.831
Percent Hispanic	0.003 (0.003)	0.196	Predominantly Hispanic	0.147 (0.247)	0.551
Percent Families in Poverty	0.001 (0.003)	0.835	Predominantly Families in Poverty	0.084 (0.065)	0.197
~ $p < 0.10$; * $p < 0.05$ (bold); ** $p < 0.01$ (bold)					

Multivariate models are controlled for population density, open space density and the other neighborhood socio-demographic characteristics.

marginally significant inverse relationship between Black neighborhoods and tree density, we found no statistically significant relationship between neighborhood socio-demographic composition and tree density when estimating the spatial models, which were necessary in this research. Thus, our study provides knowledge about the spatial distribution trees, including that the relationship varies across geographic contexts.

As previously discussed, the existing research evaluating relationships between neighborhood socio-demographic characteristics (e.g. minority racial/ethnic composition and poverty) have been mixed in terms of direction of effect and significance. Interestingly, both previous studies examining the relationship between neighborhood socio-demographic composition and the spatial distribution of trees that evaluated and subsequently detected spatial autocorrelation implemented the spatial error model (Landry and Chakraborty, 2009; Kendal et al. 2012) in contrast to our study that determined the spatial lag model to be more appropriate. In the Landry and Chakraborty (2009) study, the Lagrange Multiplier test suggested that the spatial error model was most appropriate. Moreover, the association between the uneven distribution of trees and racial composition was only identified after accounting for spatial dependence in multivariate models (Landry and Chakraborty 2009). Kendal et al. (2012) state that spatial regression was carried out to identify physical (e.g. built age) and social factors (e.g. household income, percent graduates) associated with tree cover in different land uses – residential properties, residential road services, and pubic parks – because of significant spatial autocorrelation detected using the Moran's *I*, but they do not provide any results or explanations for why they used and report results from spatial error models. Household income was neither included in the best non-spatial OLS model nor the best spatial error model (the model with the lowest AIC) (Kendal et al., 2012). Model results suggest that tree cover across the land uses was explained by education level rather than income level; areas of higher tree cover were associated with higher proportion of residents with graduate education (Kendal et al., 2012). Moreover, individuals with higher incomes and lower education levels settled in areas with newer housing developments with lower tree cover (Kendal et al., 2012). Our findings suggest that the explicit application of spatial econometric methodologies is important substantively when studying spatial demographic phenomenon, such as neighborhood socio-demographic characteristics, which could influence the spatial distribution of trees. Our explicit utilization of a spatial perspective in our analysis also enabled us to carefully examine and account for heteroskedasticity. Previous studies evaluating socio-demographic disparities in trees using spatial regressions did not report on potential heteroskedasticity, but this can impact findings. In this study, we found heteroskedasticity in spatial models estimated via maximum likelihood. Therefore, we computed spatial heteroskedasticity and autocorrelation consistent (HAC) estimations.

There are several reasons why the relationship may be different depending on the setting and there are several *explanations* for our findings of no socio-demographic disparities in the spatial distribution of trees. First, areas of higher than expected tree cover in urban areas may be attributed to legacy effects of past socio-political processes and socio-cultural preferences, notably the preservation of forested areas as preserves and other public lands and programs promoting residential landscaping and street-tree plantings (Heynen et al. 2006; Iverson and Cook, 2000; Kirkpatrick et al., 2011). Past socio-demographic composition of neighborhoods may better explain vegetation patterns of neighborhoods because shifts in neighborhood composition and ecological changes occur at different rates (Pickett et al., 2008). Second, changes in the composition and life cycle of trees typically occur over a longer period time as compared to changes in neighborhood composition and public/private investments in trees (Grove et al. 2006; Kirkpatrick et al., 2011; Pickett et al., 2008; Troy et al., 2007). Furthermore, some neighborhoods often lack planting space, especially in highly built areas, and residents may even associate existing trees with liabilities rather than benefits

(Heynen et al., 2006; Iverson and Cook, 2000). Heynen et al. (2006) discovered that predominantly Black neighborhoods in Milwaukee have significant 'fence-line' forests, i.e. trees that grow along fences and foundations of buildings in poorly maintained properties. Although these fence-line trees may contribute to the urban tree canopy, they were largely perceived as 'weed' trees by many residents because of their potential to damage property. 'Fence-line' trees were also an issue in densely settled Hispanic neighborhoods, but residents took preemptive measures to remove 'weed' trees. Decisions to maintain and invest in quality trees are critical for the equitable distribution of trees (Perkins et al., 2004). The removal of poorly maintained trees may outpace new plantings and growth of properly maintained trees, which in turn, may contribute to the unequal distribution and benefits of trees over time.

Public investment and policy decisions promoting tree-planting programs are likely to have shaped the spatial distribution of trees in Boston. For instance, in 1992, a community tree program was initiated by the Boston Parks Department in three neighborhoods – East Boston, Mattapan, and the South End (City of Boston Parks & Recreation Department, 2002). It is interesting to note the present racial/ethnic composition of East Boston and Mattapan. East Boston has historically attracted immigrants, and more recently, there has been an influx of immigrants from Central and South America; the 2010 US Census estimates indicate about half of the residents in this neighborhood are Hispanic or Latino (City of Boston, 2012; Boston Redevelopment Authority Research Division, 2011). Mattapan is predominantly comprised of Black (African-American and Caribbean) immigrants; the 2010 U.S. Census estimates indicate about three quarters of the residents in this neighborhood are Black (City of Boston, 2012; Boston Redevelopment Authority Research Division, 2011). Between 1994 and 2001, more than 7,000 trees were planted throughout Boston. Furthermore, in 2007, the City of Boston and partners in the Boston Urban Forest Coalition started a campaign to plant 100,000 new trees in Boston by 2020 (Russell, 2007). Neighborhoods with low canopy cover were prioritized for the planting of new trees. According to a report by the Urban Ecology Institute (2008), 1,000 trees were planted in 2007 and 3,000 in 2008; as of 2008, about half of Boston's neighborhoods had at least 30% tree cover, but the neighborhoods of East Boston, South Boston, and the Central City (including Chinatown) had less than 10% tree cover.

Future Research Directions and Study Implications

While we found no socio-demographic disparities in trees, suggesting no need for policy intervention for trees in Boston, there remains a need for continued research. Replication studies are indeed needed. Future studies should consider monitoring the tree distribution in Boston. Furthermore, additional research should be conducted in other geographic areas and consider different definitions of a neighborhood, including those that may be more germane to the spatial process of the geography of trees, such as perhaps via egocentric neighborhood definitions (Matthews, 2011). George Galster (2001) noted that urban social scientists "have treated 'neighborhood' in much the same way as courts have treated pornography: a term that is hard to define precisely, but everyone knows it when they see it" (P. 2111), which is perhaps one reason neighborhoods have been defined variously. For the analytical reasons previously described, we felt that a Queen contiguity matrix was best in this instance. Future research could consider utilizing other spatial weights matrices that may relate better to the spatial distribution of trees. Additionally, another possible avenue for future research is to consider tree quality, age and type and including these variables (e.g. the spatial distribution of healthy trees or account for tree health) in the regression models (which we were not able to do given the structure of the data). Why are the health of trees, their age and tree species diversity important? Diversity of tree species and condition of trees for a given area can be used to evaluate the quality of the urban forest. Tree species monoculture may provide some positive effects in the short-term, specifically in terms of economic costs, but in the long-term, lack of tree species diversity leaves the urban forest more vulnerable to pests, diseases and natural disaster events (Flocks et al., 2011; Pedlowski et al., 2002; Zhao et al., 2010). Young growing trees in poor condition may indicate that there may be physical stresses (e.g. soil compaction, disease, water, pollution) or social stresses (e.g. damages made by human actions) that may make trees more susceptible to pests and diseases, as well as, inhibit proper functioning of trees, such as pollution removal and temperature modification (Nowak et al., 2002; Welch, 1994). Species diversity and condition of trees also reflects policies implemented and resources allocated in the management and care of trees (Flocks et al., 2011). As noted, the age of the trees might also be a target for future research. Younger trees, for example, may be more easily damaged (e.g. by deer) than older trees and thus their beneficial effects might not be as salient. It is also worth noting that while we implemented a spatially explicit approach (consisting of various spatial methods) to the study of socio-demographic disparities in the spatial distribution of trees, other spatial approaches exist for doing so that were not used in this study (such as local indicators of spatial association statistical methods [Anselin, 1995] and agent-based spatial modeling methods [O'Sullivan, 2009]), which could be used in future research. This future research might suggest the need for policy intervention to ensure equitable spatial distribution of diverse trees, as an example. When spatial inequalities are found in the distribution of trees, land use planning policies (e.g. tree ordinances) could be implemented but care should be taken to ensure that trees are *equitably* placed. It is also important to recognize that housing policies may influence where trees are located and as such could help ensure an equitable distribution of trees.

Research Limitations

This research has limitations that should be stated. First, this study was conducted in one city: Boston, Massachusetts. As such, our findings may not be generalizable to other urban areas. Second, in our study, the definition of neighborhood was the US census tract, which are a commonly used neighborhood definition (including in Boston-based neighborhood research) (Duncan et al., 2012; Subramanian et al. 2006, Subramanian et al. 2005; Krieger et al. 2003, Krieger et al. 2002). There are limitations of census tracts in neighborhood analyses (including census tracts are arbitrary units, there can considerable heterogeneity in the population size of census tracts and census tracts might be the most relevant neighborhood definition for understanding peoples spatial behavior/ exposure) [Matthews, 2011; Osypuk and Galea, 2007] and while other neighborhood definitions in Boston exists (including those based on the Boston Public Health Commission [Chen et al. 2006] and the Boston Redevelopment Authority [Li et al., 2009a; Li et al., 2009b]) these other neighborhood definitions are much larger than census tracts and a consequence of this is both a smaller sample size in our analysis and more relevant the likelihood of less variation among the units being measured due to spatial aggregation at coarser scales exists (Reynolds and Amrhein, 1997). We recognize that the modifiable areal unit problem is a concern in spatial research (Openshaw and Taylor, 1979; Arbia, 1989; Wong, 2009). Third, our analysis does not account for "edge effects" (i.e. census tracts adjacent to but outside of our Boston study area) and therefore we may have underestimated spatial associations. It is also worth noting that low density tracts could be due to built environment factors (e.g. airport) and natural factors (e.g. water), and that our analysis did not account for neighboring census tracts that were removed from the analysis. However, substantively, we do not believe this is necessary or appropriate in this instance. This study included both park and street trees, which may be a limitation. We recognize that there are various approaches to define predominantly minority racial/ethnic and high-poverty neighborhoods (e.g. predominantly minority neighborhoods could be defined as 50% and high-poverty neighborhoods could be defined as 30% or more of the population living in poverty), and that this may influence study findings. Additionally, another threshold approach is majority-rule (Farrell et al., 2011). As stated previously, our categorizations were based on prior published research. Additionally, a priori, we decided to evaluate neighborhood poverty because it is arguably the most frequently used measure of neighborhood disadvantage. Neighborhood poverty, which we believe to be an important neighborhood feature, has exhibited pronounced and consistent associations with various outcomes, ranging from social and health phenomenon. We are aware that other aspects of neighborhood disadvantage exist. Besides neighborhood poverty, another aspect of neighborhood disadvantage is percent of households receiving public assistance, for example. Different measures can be combined (e.g. using principal component analysis) to create a single measure of neighborhood disadvantage (e.g. Sampson et al., 2008; Morenoff et al., 2007; Sampson et al., 1997). However, composite measures suffer the limitation of not being able to distinguish which features are salient in any association. Despite our efforts to control for various potential confounding covariates, there is potential for residual confounding.

These limitations notwithstanding, our findings provide methodological and substantive insights. Methodologically, our study suggests the need to take into account spatial autocorrelation and we demonstrate a case where findings/conclusions can change when the spatial autocorrelation is ignored. Substantively, our findings suggest no need for policy intervention vis-à-vis trees in Boston, though we hasten to add that replication studies, and more nuanced data on tree quality, age and diversity are needed.

Acknowledgment

D.T. Duncan was supported by the Alonzo Smythe Yerby Postdoctoral Fellowship at Harvard School of Public Health. The Yerby Postdoctoral Fellowship and the National Institutes of Health Grant # 1P50CA148596 to the Lung Cancer Disparities Center at Harvard School of Public Health (David R. Williams, PhD, MPH, Principal Investigator) supported participation in the Spatial Econometrics Advanced Institute in Rome, Italy (May-June 2012) (Giuseppe Arbia, PhD, Director of SEAI), which this paper benefited from substantially. D.T. Duncan also acknowledges that this paper substantially benefited from participation in the Spatial Pattern Analysis Workshop (July 2010) and the Spatial Regression Modeling Workshop (June 2011) held at the University of California-Santa Barbara and The Pennsylvania State University, respectively, both which are part of the Advanced Spatial Analysis Training Program for Population Scientists supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant #R25-HD057002) (Stephen A. Matthews, PhD, Principal Investigator). We thank Jeff Blossom for providing technical assistance with building the geospatial dataset used in this research and for commenting on our description of the tree data. We also thank the City of Boston's Department of Innovation and Technology for in releasing the tree data.

References

Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716-723.

Anselin, L., Bera, A.K., Florax, R., and Yoon, M.J. (1996). Simple Diagnostic Tests for Spatial Dependence. *Regional Science and Urban Economics* 26(1):77-104.

Anselin, L. and Bera, A.K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A. and Giles, D.E.A. (eds). *Handbook of Applied Economic Statistics*. New York: Marcel Dekker: 237-289.

Anselin, L. (1988a). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis* 20(1): 1-17.

Anselin, L. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers, 1988b.

Anselin, L. (1995). Local Indicators of Spatial Association - LISA. *Geographical Analysis 27*(2), 93-115.

Anselin L and Lozano-Gracia N. (2008) "Errors in variables and spatial effects in hedonic house price models of ambient air quality" *Empirical Economics*, 34, 5-34.

Arbia, G. (1989) *Spatial data configuration in the* *statistical analysis of regional economics and related problems,* Kluwer, Dordrecht.

Bailey, T. C. and Gatrell, A. C. (1995). *Interactive Spatial Data Analysis*. Harlow Essex, England: Longman Scientific & Technical; J Wiley.

Basu, R., & Samet, J. M. (2002). Relation Between Elevated Ambient Temperature and Mortality: A Review of the Epidemiologic Evidence. *Epidemiologic Reviews*, *24*(2), 190–202.

Bedimo-Rung, A. L., Mowen, A. J., & Cohen, D. A. (2005). The significance of parks to physical activity and public health: A conceptual model. *American Journal of Preventive Medicine*, *28*(2, Supplement 2), 159–168.

Bivand, R.S., Pebesma, E.J., and Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York, NY: Springer.

Bivand, R. (2011). spdep: Spatial dependence: weighting schemes, statistics and models. [electronic resource]. http://cran.r-project.org/web/packages/spdep/index.html

Block, J.P., Scribner, R.A., and DeSalvo, K.B. (2004). Fast food, race/ethnicity, and income: a geographic analysis. *American Journal of Preventive Medicine* 27(3): 211-7.

Boston Redevelopment Authority Research Division (2011). 2010 U.S. Census, Summary File 1.

Brewer, C., and Harrower, M. (2012). Color Brewer 2.0. [electronic resource]. http://www.colorbrewer2.org

Breusch T.S. & A.R. Pagan (1979), A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47, 1287–1294.

Brown University, Center for Environmental Studies. (2010.) “Trees and the Urban Health Island Effect: A Case Study for Providence Rhode Island”.

Chen, J.T., Rehkopf, D.H., Waterman, P.D., Subramanian, S.V., Coull, B.A., Cohen, B., Ostrem, M., and Krieger, N. (2006). Mapping and measuring social disparities in premature mortality: the impact of census tract poverty within and across Boston neighborhoods, 1999-2001. *Journal of Urban Health* 83(6): 1063-84.

City of Boston (2012). Boston: A City of Neighborhoods. Retrieved from http://www.cityofboston.gov/neighborhoods/default.asp.

City of Boston. Parks and Recreation Department (2002). Open Space Plan 2002-2006.

Cliff, A.D., and J.K. Ord. (1981) *Spatial processes: models and applications*, Taylor & Francis.

Clifford, P. and Richardson, S. (1985). Testing the association between two spatial processes. *Statistics and Decisions* 2(Supp. issue): 155-160.

Conway, T. M., Shakeel, T., & Atallah, J. (2011). Community groups and urban forestry activity: Drivers of uneven canopy cover? *Landscape and Urban Planning*, *101*(4), 321–329.

Corti, B., Donovan, R. and Holman, C. (1996). “Factors influencing the Use of Physical Activity Facilities: Results From Qualitative Research.” *Health Promotion Journal Australia,* 6: 16–21.

Cradock, A.L., Kawachi, I., Colditz, G.A., Hannon, C., Melly, S.J., Wiecha, J.L., and Gortmaker, S.L. (2005). Playground safety and access in Boston neighborhoods. *American Journal of Preventive Medicine* 28(4): 357-63.

Cressie, N. (1993). *Statistics for Spatial Data*, Wiley, NY.

Crawford, D., Timperio, A., Giles-Corti, B., Ball, K., Hume, C., Roberts, R., Andrianopoulos, N., et al. (2008). Do features of public open spaces vary according to neighbourhood socio-economic status? *Health & Place*, *14*(4), 889–893.

Donovan, G. H., & Butry, D. T. (2010). Trees in the city: Valuing street trees in Portland, Oregon. *Landscape and Urban Planning*, *94*(2), 77–83.

Donovan, G.H., Michael, Y.L., Butry, D.T., Sullivan, A.D., Chase, J.M. (2011). Urban trees and the risk of poor birth outcomes. *Health & Place*. Jan;17(1):390-3.

Duncan, D.T., Aldstadt, J., Whalen, J., White, K., Castro, M.C., Williams, D.R. (2012). Space, Race and Poverty: Spatial Inequalities in Walkable Neighborhood Amenities? *Demographic Research*. 26(17): 409-448.

Duncan, O. D. and Duncan, B., (1955). A methodological analysis of segregation indexes.

American Sociological Review, Vol. 20, 210–217.

Dwyer, J. F., E. G. McPherson, H. W. Schroeder, and R. A. Rowntree. (1992). Assessing the benefits and costs of the urban forest. *Journal of Arboriculture* 18:227–227.

Escobedo, F. J., Nowak, D. J., Wagner, J. E., De la Maza, C. L., Rodríguez, M., Crane, D. E., & Hernández, J. (2006). The socioeconomics and management of Santiago de Chile's public urban forests. *Urban Forestry & Urban Greening*, *4*(3–4), 105–114.

Farrell, C.R. and Lee, B.A. (2011). "Racial Diversity and Change in Metropolitan Neighborhoods." *Social Science Research* 40 (July):1108-23.

Fang, C.-F., & Ling, D.-L. (2005). Guidance for noise reduction provided by tree belts. *Landscape and Urban Planning*, *71*(1), 29–34.

Fix, M. and Struyk, R.J. (1993). *Clear and convincing evidence: Measurement of discrimination in America*. Washington: Urban Institute Press.

Flocks, J., Escobedo, F., Wade, J., Varela, S., & Wald, C. (2011). Environmental Justice Implications of Urban Tree Cover in Miami-Dade County, Florida. *Environmental Justice*, *4*(2), 125–134.

Florax JGMR, Folmer H, Rey SJ. 2003. "Specification searches in spatial econometrics: the relevance of Hendry's methodology," *Regional Science and Urban Economics*, 33(5): 557-579.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., et al. (2005). Global consequences of land use. *Science*, *309*(5734), 570–574.

Franco, M., Diez Roux, A.V., Glass, T.A., Caballero, B., and Brancati, F.L. (2008). Neighborhood characteristics and availability of healthy foods in Baltimore. *American Journal of Preventive Medicine* 35(6): 561-7.

Franzini L, Taylor W, Elliott MN, Cuccaro P, Tortolero SR, Janice Gilliland M, Grunbaum J, Schuster MA. Neighborhood characteristics favorable to outdoor physical activity: disparities by socioeconomic and racial/ethnic composition. *Health & Place*. 2010 Mar;16(2):267-74.

Galster, GC. The Ecology of Racial Discrimination in Housing: An Exploratory Model. *Urban Affairs Quarterly*, v23 n1 p84-107 Sep 1987.

Galster G. On the nature of neighborhoods. *Urban Studies*. 2001;38(12):2111-2124.

Georgi, N. J., & Zafiriadis, K. (2006). The impact of park trees on microclimate in urban areas. *Urban Ecosystems*, *9*(3), 195–209.

Gidlöf-Gunnarsson, A., & Öhrström, E. (2007). Noise and well-being in urban residential environments: The potential role of perceived availability to nearby green areas. *Landscape and Urban Planning*, *83*(2–3), 115–126.

Grove, J. M., Troy, A. R., O'Neil-Dunne, J. P. M., Burch, W. R., Cadenasso, M. L., & Pickett, S. T. A. (2006). Characterization of households and its implications for the vegetation of urban ecosystems. *Ecosystems*, *9*(4), 578–597.

Haining, R. (1991). Bivariate correlation with spatial data. *Geographical Analysis* 23(3): 210-227.

Heisler, G. M., & Grant, R. H. (2000). Ultraviolet radiation in urban ecosystems with consideration of effects on human health. *Urban Ecosystems*, *4*(3), 193–229.

Heynen, N., Perkins, H. A., & Roy, P. (2006). The Political Ecology of Uneven Urban Green Space The Impact of Political Economy on Race and Ethnicity in Producing Environmental Inequality in Milwaukee. *Urban Affairs Review*, *42*(1), 3–25.

Iverson, L. R., & Cook, E. A. (2000). Urban forest cover of the Chicago region and its relation to household density and income. *Urban Ecosystems*, *4*(2), 105–124.

Iceland, J., Weinberg, D. H., and Steinmetz E.. (2002). Racial and Ethnic Residential Segregation in the United States: 1980-2000. Washington, DC: Government Printing Office. http://www.census.gov/hhes/www/housing/housing_patterns/pdf/censr-3.pdf

Jenerette, G. D., Harlan, S. L., Brazel, A., Jones, N., Larsen, L., & Stefanov, W. L. (2007). Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem. *Landscape Ecology*, *22*(3), 353–365.

Jenks GF: The Data Model Concept in Statistical

Mapping. *International Yearbook of Cartography* 1967, 7:186-190.

Jensen, R., Gatrell, J., Boulton, J., & Harper, B. (2004). Using remote sensing and geographic information systems to study urban quality of life and urban forest amenities. *Ecology and Society*, *9*(5), 5.

Kaplan, S., & Kaplan, R. (2003). Health, Supportive Environments, and the Reasonable Person Model. *American Journal of Public Health*, *93*(9), 1484–1489.

Kendal, D., Williams, N. S. G., & Williams, K. J. H. (2012). Drivers of diversity and tree cover in gardens, parks and streetscapes in an Australian city. *Urban Forestry & Urban Greening*, *11*(3), 257–265.

Kirkpatrick, J. B., Daniels, G. D., & Davison, A. (2011). Temporal and spatial variation in garden and street trees in six eastern Australian cities. *Landscape and Urban Planning*, *101*(3), 244–252.

Kelejian, H.H. and Prucha, I.R. (1998) A Generalized Spatial Two Stage Least Square Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17, pages 99–121.

Kelejian, H.H., Prucha, I.R. (2007). “HAC Estimation in a Spatial Framework.” *Journal of Econometrics*, 140, 131–154.

Kelly CM, Schootman M, Baker EA, Barnidge EK, Lemes A. The association of sidewalk walkability and physical disorder with area-level race and poverty. *Journal of Epidemiology and Community Health*. 2007 Nov;61(11):978-83.

Krieger, N., Chen, J.T., Waterman, P.D., Rehkopf, D.H., and Subramanian, S.V. (2003). Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures—the Public Health Disparities Geocoding Project. *American Journal of Public Health* 93(10): 1655-1671.

Krieger, N., Chen, J.T., Waterman, P.D., Soobader, M.J., Subramanian, S.V., and Carson, R. (2002). Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. *American Journal of Epidemiology* 156(5): 471-482.

Kuo, F. E. (2003). The role of arboriculture in a healthy social ecology. *Journal of Arboriculture*, *29*(3), 148–155.

Kuo, F.E., W.C. Sullivan, R.L. Coley, and L. Brunson. “Fertile ground for community: inner-city neighborhood common spaces.” *American Journal of Community Psychology* 26 (1998):823-851.

Kuo, F.E., and W.C. Sullivan. “Environment and crime in the inner city: does vegetation reduce crime?” *Environment and Behavior* 333 (2001):343-367.

Landry, S. M., & Chakraborty, J. (2009). Street trees and equity: evaluating the spatial distribution of an urban amenity. *Environment and Planning A*, *41*(11), 2651–2670.

LeSage, J. and Pace, K.R. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.

LeSage, J.P. and Pace, K.R. (2010). The Biggest Myth in Spatial Econometrics. Available at SSRN: http://ssrn.com/abstract=1725503

Li, W., Land, T., Zhang, Z., Keithly, L., and Kelsey, J.L. (2009a). Small-area estimation and prioritizing communities for tobacco control efforts in Massachusetts. *American Journal of Public Health* 99(3): 470-9.

Li, W., Kelsey, J.L., Zhang, Z., Lemon, S.C., Mezgebu, S., Boddie-Willis, C., and Reed, G.W. (2009b). Small-area estimation and prioritizing communities for obesity control in Massachusetts. *American Journal of Public Health* 99(3): 511-9.

Lichter, D.T., Patisi, D., Grice, S.M., Taquino, M. Municipal underbounding: Annexation and racial exclusion in small southern towns. *Rural Sociology* 2007;72(1):47-68.

Logan, J.R. and Stults, B. (2011). The Persistence of Segregation in the Metropolis: New Findings from the 2010 Census. Census Brief prepared for Project US2010. http://www.s4.brown.edu/us2010.

Loukaitou-Sideris A, Stieglitz O: Children in Los Angeles parks: a study of equity, quality and children's satisfaction with neighbourhood parks. *The Town Planning Review* 2002, 73(4):467-488.

Lovasi, G.S., Quinn, J.W., Neckerman, K.M., Perzanowski, M.S., Rundle, A. Children living in areas with more street trees have lower prevalence of asthma. *Journal of Epidemiology and Community Health*. 2008 Jul;62(7):647-9.

Luck, G. W., Smallbone, L. T., & O'Brien, R. (2009). Socio-economics and vegetation change in urban ecosystems: patterns in space and time. *Ecosystems*, *12*(4), 604–620.

Marsh, B., Parnell, A.M., Joyner, A.M. Institutionalization of racial inequality in local political geographies. *Urban Geography* 2010;31(5):691-709.

Massey, D. S. and Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Matthews, S.A. Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods. In: Burton, L.M., Kemp, S.P., Leung, M., Matthews, S.A., Takeuchi, D.T., eds. *Communities, Neighborhoods and Health: Expanding the Boundaries of Place*. New York: Springer; 2011:35-55.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena, *Biometrika* 37, pp17-23.

Morenoff, J.D., House, J.S., Hansen, B.B., Williams, D.R., Kaplan, G.A., Hunte, H.E. (2007). Understanding social disparities in hypertension prevalence, awareness, treatment, and control: the role of neighborhood context. *Social Science & Medicine*. 65(9):1853-66.

Moore, L.V. and Diez Roux, A.V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health* 96(2): 325-331.

Moore, L.V., Diez Roux, A.V., Evenson, K.R., McGinn, A.P., and Brines, S.J. (2008). Availability of recreational resources in minority and low socioeconomic status areas. *American Journal of Preventive Medicine* 34(1): 16-22.

Neckerman, K. M., Lovasi, G. S., Davies, S., Purciel, M., Quinn, J., Feder, E., Raghunath, N., et al. (2009). Disparities in Urban Neighborhood Conditions: Evidence from GIS Measures and Field Observation in New York City. *Journal of Public Health Policy*, *30*, S264–S285.

Nowak, D. J., Crane, D. E. & Dwyer, J. F. (2002). Compensatory value of urban trees in the United States. *Journal of Arboriculture* 28, 194–199

Nowak, D. J., Crane, D. E. & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening* 4, 115–123.

Oka, M. (2011). The Influence of Urban Street Characteristics on Pedestrian Heat Comfort Levels in Philadelphia. *Transactions in GIS*, *15*(1), 109–123.

Openshaw, S. and Taylor, P. (1979). A million or so correlation coefficients: three experiments on the modifiable area unit problem. In: Wrigley, N. (ed.). *Statistical Applications in the Spatial Sciences*. London, UK: Pion Ltd: 127-144.

Ord, K. (1975). "Estimation Methods for Models of Spatial Interaction." *Journal of the Statistical Association*, 70:120–126.

O'Sullivan, D. (2009). Changing Neighborhoods–Neighborhoods Changing: A Framework for Spatially Explicit Agent-Based Models of Social Systems. *Sociological Methods & Research* Volume 37 Number 4, pp 498-530.

Osypuk, T., Galea, S. What Level Macro? Choosing Appropriate Levels to Assess the Relation Between Space and Population Health. In: *Macrosocial Determinants of Population Health* S. Galea (ed.). 2007. New York, NY: Springer Media: 399-436.

Pace, K. and LeSage, J.P. (2008). A Spatial Hausman Test. *Economic Letters* 101(3): 282-284.

Payton, S., Lindsey, G., Wilson, J., Ottensmann, J. R., & Man, J. (2008). Valuing the benefits of the urban forest: a spatial hedonic approach. *Journal of Environmental Planning & Management*, *51*(6), 717–736.

Pedlowski, M. A., Da Silva, V. A. C., Adell, J. J. C., & Heynen, N. C. (2002). Urban forest and environmental inequality in Campos dos Goytacazes, Rio de Janeiro, Brazil. *Urban Ecosystems*, *6*(1), 9–20.

Perkins, H. A., Heynen, N., & Wilson, J. (2004). Inequitable access to urban reforestation: the impact of urban political economy on housing tenure and urban forests. *Cities*, *21*(4), 291–299.

Pickett, S. T. A., Cadenasso, M. L., Grove, J. M., Groffman, P. M., Band, L. E., Boone, C. G., Burch Jr, W. R., et al. (2008). Beyond urban legends: an emerging framework of urban ecology, as illustrated by the Baltimore Ecosystem Study. *BioScience*, *58*(2), 139–150.

Piras, G. sphet: Spatial Models with Heteroskedastic Innovations in R. *Journal of Statistical Software*. 2010; 35(1): 1-21. URL: http://www.jstatsoft.org/v35/i01/.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Reibel, M. (2000). Geographic Variation in Mortgage Discrimination: Evidence from Los Angeles. *Urban Geography 21*(1), 45-60.

Reynolds, H. and Amrhein, C. (1997). Using a spatial data set generator in an empirical analysis of aggregation effects on univariate statistics. *Geographical and Environmental Modelling* 1: 199-220.

Rothwell, J. T. (2011). Racial Enclaves and Density Zoning: The Institutionalized Segregation of Racial Minorities in the United States. *American Law and Economics Review* 13(1): 290-358.

Russell, J. (2007). A plan takes root: City to plant more than 100,000 trees. *The Boston Globe*. Available at: http://www.boston.com/news/local/articles / 2007/04/28/a_plan_takes_root_city_to_plant _more_than_100000_trees/

Sampson, R.J., Raudenbush S., Earls, F. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277: 918-24.

Sampson, R. J., Sharkey, P. and Raudenbush, S. W. (2008). "Durable Effects of Concentrated Disadvantage on Verbal Ability Among African-American Children." *Proceedings of the National Academy of Sciences* 105: 845-852.

Sarajevs, V. (2011). "Health Benefits of Street Trees". The Research Agency of the Forestry Commission.

Schabenberger, O. and Gotway, C. 2008. *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC.

Simpson, J. R. (2002). Improved estimates of tree-shade effects on residential energy use. *Energy and Buildings*, *34*(10), 1067–1076.

Smiley, M.J., Diez Roux, A.V., Brines, S.J., Brown, D.G., Evenson, K.R., Rodriguez, D.A. (2010). A spatial analysis of health-related resources in three diverse metropolitan areas. *Health & Place*. Sep;16(5):885-92.

Srinivasan, S., O'Fallon, L. R., & Dearry, A. (2003). Creating Healthy Communities, Healthy Homes, Healthy People: Initiating a Research Agenda on the Built Environment and Public Health. *American Journal of Public Health*, *93*(9), 1446–1450.

Student. (1914). The elimination of spurious correlation due to position in time or space. *Biometrika* 10(1): 179-180.

Subramanian, S.V., Chen, J.T., Rehkopf, D.H., Waterman, P.D., and Krieger, N. (2005). Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts. *American Journal of Public Health* 95(2): 260-5.

Subramanian, S.V., Chen, J.T., Rehkopf, D.H., Waterman, P.D., and Krieger, N. (2006). Comparing individual- and area-based socioeconomic measures for the surveillance of health disparities: A multilevel analysis of Massachusetts births, 1989-1991. *American Journal of Epidemiology* 164(9): 823-34.

Sullivan, W.C., Kuo, F.E., and Depooter, S.F. (2004). The Fruit of Urban Nature: Vital Neighborhood Spaces. *Environment and Behavior*, 36(5): 678-700.

Szantoi, Z., Escobedo, F., Wagner, J., Rodriguez, J., & Smith, S. (2012). Socioeconomic Factors and Urban Tree Cover Policies in a Subtropical Urban Forest. *GIScience & Remote Sensing*, *49*(3), 428–449.

Talarchek, G.M. (1990). "The urban forest of New Orleans – an exploratory analysis of relationships." *Urban Geography* 11 65–86.

Taylor A. F. Wiley A, Kuo F. E. Sullivan W. C. (1998). "Growing up in the inner city: Green spaces as places to grow." *Environment and Behavior* 30, no. 1 3–27.

Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*. 46(2): 234-240.

Tooke, T. R., Klinkenber, B., & Coops, N. C. (2010). A geographical approach to identifying vegetation-related environmental equity in Canadian cities. *Environment and Planning. B, Planning & Design*, *37*(6), 1040.

Troy, A. R., Grove, J. M., O'Neil-Dunne, J. P. M., Pickett, S. T. A., & Cadenasso, M. L. (2007). Predicting opportunities for greening and patterns of vegetation on private urban lands. *Environmental Management*, *40*(3), 394–412.

Tyrväinen, L., Pauleit, S., Seeland, K., & de Vries, S. (2005). Benefits and uses of urban forests and trees. *Urban forests and trees. Springer-Verlag, Berlin Heidelberg*, 81–114.

Urban Ecology Institute (2008). State of the urban forest: a summary of the extent and condition of Boston's urban forest.

U.S. Census Bureau (2012). Decennial Management Division Glossary [electronic resource]. http://www.census.gov/dmd/www/glossary.html

Uuttera, J., Haara, A., Tokola, T., Maltamo, M. (1998). Determination of the spatial distribution of trees from digital aerial photographs. *Forest Ecology and Management*. Volume 110, Issues 1–3: 275–282

Velarde, M.D., Fry, G., Tveit, M.S. (2007). Health effects of viewing landscapes: landscape types in Environmental Psychology. *Urban Forestry & Urban Greening* 6:199-212.

Waller, L.A. and Gotway, C.A. (2004). *Applied Spatial Statistics for Public Health Data*. Hoboken NJ: Wiley-Interscience.

Ward, M.D. and Gleditsch, K.S. (2008). *Spatial Regression Models*. Thousand Oaks, CA: Sage Publications, Inc.

Welch, J. M. (1994). Street and park trees of Boston: a comparison of urban forest structure. *Landscape and Urban Planning*, *29*(2–3), 131–143.

Wolch, J., Wilson, J., & Fehrenbach, J. (2005). Parks and Park Funding in Los Angeles: An Equity-Mapping Analysis. *Urban Geography*, *26*(1), 4–35.

Wolf, K. L. (2003). Public Response to the Urban Forest in Inner-City Business Districts. *Journal of Arboriculture*, *29*(3), 117–126.

Wolf, K. L. (2005). Business district streetscapes, trees, and consumer response. *Journal of Forestry*, *103*(8), 396–400.

Wong, D. (2009). The Modifiable Areal Unit Problem (MAUP). In: Fotheringham, A.S. and Rogerson, P.A. (eds.). *The SAGE Handbook of Spatial Analysis*. London: SAGE Publications: 105-124.

Zhao, M., Escobedo, F. J., & Staudhammer, C. (2010). Spatial patterns of a subtropical, coastal urban forest: Implications for land tenure, hurricanes, and invasives. *Urban Forestry & Urban Greening*, *9*(3).