Using High-Resolution Population Data to Identify Neighborhoods and Establish Their Boundaries

Seth E. Spielman [a] [b] & John R. Logan [c] [d]

[a] Department of Geography, University of Colorado at Boulder

[b] Spatial Structures in the Social Sciences, Brown University

[c] Department of Sociology, Brown University

[d] Spatial Structures in the Social Sciences

PLEASE SCROLL DOWN FOR ARTICLE

Using High-Resolution Population Data to Identify Neighborhoods and Establish Their Boundaries

Seth E. Spielman* and John R. Logan[†]

*Department of Geography, University of Colorado at Boulder, and Spatial Structures in the Social Sciences, Brown University
[†]Department of Sociology, Brown University, and Spatial Structures in the Social Sciences

Neighborhoods are about local territory, but what territory? This article offers one approach to this question through a novel application of "local" spatial statistics. We conceptualize a neighborhood in terms of both space and social composition; it is a contiguous territory defined by a bundle of social attributes that distinguish it from surrounding areas. Our method does not impose either a specific social characteristic or a predetermined spatial scale to define a neighborhood. Rather, we infer neighborhoods from detailed information about individual residents and their locations. The analysis is based on geocoded complete-count census data from the late nineteenth century in four cities: Albany, New York; Buffalo, New York; Cincinnati, Ohio; and Newark, New Jersey. We find striking regularities (and some anomalies) in the spatial structure of the cities studied. Our approach illustrates the "spatialization" of an important social scientific concept. *Key Words: census, geocode, GIS, neighborhood boundaries, urban.*

设区是当地的领土，但它是什么领土呢？本文通过使用一个"局地"的空间统计的新应用，提供了针对这个问题一种方法。我们构思了一个在双方步伐和社会组成意义上的社区，它是由区别于周边地区的一堆社会属性所定义的一个连续的领土。我们的方法并没有强加一个特定的社会特征或预定的空间规模来定义一个街区。相反，我们从有关个别的居民和他们的位置的详细信息推断街区。该分析基于十九世纪末的四个城市的地理编码的普查数据的完整计数，即纽约州的奥尔巴尼，纽约州的布法罗市，俄亥俄州的辛辛那提市，和新泽西州的纽瓦克市。我们发现在这些所研究的城市的空间结构上的惊人规律（和一些异常）。我们的方法说明了一项重要的社会科学概念的"空间化"。*关键词：人口普查，地理编码，地理信息系统，邻里边界，城市。*

Los barrios o vecindarios tienen que ver con el territorio local, pero ¿de qué territorio se trata? En este artículo se presenta un enfoque sobre esta cuestión por medio de una novedosa aplicación de estadísticas espaciales "locales". Nosotros conceptualizamos un vecindario en términos tanto de espacio como de composición social; se trata de un territorio contiguo definido por un paquete de atributos sociales que lo distinguen de las áreas circundantes. Para definir un vecindario, nuestro método no impone una característica social específica ni una escala espacial predeterminada. Mejor que eso, a los vecindarios los inferimos a partir de la información detallada sobre residentes individuales y sus localizaciones. El análisis se basa en cuentas completas de datos censales georreferenciados de finales del siglo XIX en cuatro ciudades: Albany, en Nueva York; Buffalo, Nueva York; Cincinnati, Ohio; y Newark, Nueva Jersey. Encontramos notables regularidades (y algunas anomalías) en la estructura espacial de las ciudades estudiadas. Nuestro enfoque ilustra la "espacialización" de un importante concepto científico social. *Palabras clave: censo, geocódigo, SIG, límites barriales, urbano.*

Social scientists studying urban issues often lament that the lack of better data requires them to treat arbitrary administrative units as neighborhoods (Dietz 2002). Suppose that higher resolution data were available—a complete enumeration of residents of a city including their geocoded addresses and an array of social and economic characteristics for each person. In that case, how would one define neighborhoods? How would one establish their geographic scale and boundaries?

Since entering the social science lexicon in the early twentieth century, neighborhood has been a contentious unit of analysis. McKenzie (1923, quoted by Matthews 2011, 41) noted that, "probably no other term is used so loosely or with such changing content as the term neighborhood, and very few are more dif-ficult to define." Alihan (1938) wondered how social scientists could divide a continuously varying cityscape into discrete regions. In response, Quinn (1940) noted that many natural phenomena, like the electromagnetic spectrum, vary continuously, yet we identify distinct zones within them, naming colors in the visible portion of the electromagnetic spectrum even though color is a continuous rather than a discrete phenomenon. The challenge of defining the boundaries of neighborhoods is similar but with a major difference: People have reached consensus on what colors to look for, but there is no similar consensus on what kinds of neighborhoods exist, what characteristics should be used to identify them, or at what scale they occur.

In this article we present one approach to these questions using microdata from the 1880 decennial census. These microdata allow us to map the residential location of the entire population of Buffalo, New York; Cincinnati, Ohio; Albany, New York; and Newark, New Jersey. We develop a flexible method for identifying neighborhoods using these high-resolution spatial population data. We employ this method to gain insight into the social structure of industrializing cities in the late nineteenth century. This article is organized as follows: In the remainder of the introductory section we discuss the conceptualization of neighborhoods and relate it to the situation of urban life in the late nineteenth century. In the Methods section we operationalize our own understanding of neighborhoods and outline an approach to implementing it. In the Results section we apply the method to data from four cities and discuss what this approach to neighborhood research tells us about urban life in the late nineteenth century. Finally, in the Conclusion we relate our approach to current problems in urban geography and GIScience.

What Is a Neighborhood?

The use of neighborhoods as a unit of scientific analysis was advanced by the Chicago School of Sociology in the early twentieth century. The Chicago sociologists' interest in neighborhoods was motivated by an interest in "the effect of position in both time and space upon human institutions and human behavior" (Park, Burgess, and McKenzie 1925, 64). They observed that "geographical setting" was a fundamental aspect of neighborhoods (147) and proposed that, as "spatial relationships change the physical basis of social relations is altered" (64). In principle, neighborhoods could be defined along many different dimensions (political, cultural, and ecological) with boundaries that only partially overlap across dimensions. Yet Burgess's famous concentric zone model of the city, based mainly on social class and residential density, also identified specific salient combinations of class, housing type, race and ethnicity, and indicators of social disorganization that tended to be found together to constitute neighborhoods: "This differentiation into natural economic and cultural grouping gives form and character to the city" (Burgess 1925, 56). The ghetto, Little Sicily, and Deutschland were among the more colorful neighborhoods in a city that was being reorganized "into a centralized decentralized system of local communities" (52). Later in the development of the Chicago School tradition there emerged a stronger statement about the

continuity of natural areas of the city, as Shaw and McKay (1942) argued that patterns of social organization and disorganization tended to reproduce themselves over time in specific zones of the city. To identify neighborhoods, then, was to discover the fundamental ecological organization of the city.

More recent discussions of neighborhoods converge on the notion that they have a spatial dimension, including a location and (possibly) boundaries. There is agreement that the spatial organization of cities is multidimensional. A neighborhood is a "bundle of spatially based attributes" (Galster 2001, 2012). Galster identified ten general categories of attributes, ranging from the built environment and demographic composition to political behavior, social networks, and residents' place identities. He also suggested that these attributes might each vary in different spatial scales—at the extreme, one could imagine that the attributes all vary independently, so that the identification of neighborhoods along one attribute might look very different than the identification along another attribute. Dealing with such complexity, seeking to whittle down multiple indicators to a few main dimensions, is a long-standing interest of urbanists. Shevky and Bell (1961) argued that local areas within the city could be delineated on the basis of indicators of just three dimensions: social rank, urbanization, and segregation. Following in the tradition of social area analysis and factorial ecology, Sampson, Raudenbush, and Earls (1997) used factor analysis to aggregate census tracts in Chicago into neighborhood clusters that are geographically contiguous and socially homogeneous with respect to race, ethnicity, social class, housing density, and family structure. Health researchers Lebel, Pampalon, and Villeneuve (2007) offered a multiperspective approach to accomplish the same objective, including historical analysis (which streets appeared as boundaries between neighborhoods in historical documents), demographic analysis (mapping socioeconomic data), and expert perceptions (based on a committee of local experts equipped with historical and demographic information).

Such efforts reflect a place-based conception of neighborhood: a belief that neighborhoods exist and can be identified within more or less discrete boundaries despite ambiguity about what attributes to consider, how to combine them, and what scale to look for. Like other urban theorists, urban planner Kevin Lynch (1960) positsed that neighborhoods ("districts" in his terminology) can be defined by any number of architectural, social, or ecological attributes. He simplified the notion by arguing that what constitutes a neighborhood

is that these attributes come together in a distinctive way to create a "thematic unit." Some neighborhoods have hard edges, points of transition that clearly define the boundaries between thematic units. Others have soft edges, with a core that closely adheres to a thematic unit but with declining coherence at greater distances from the core. Regardless of the terminology that they use, many urban researchers concur that there is something essential about a geographically defined area, something about its composition, history, politics, or social relations among residents—or even its effects on residents—that makes it useful to think of it as a neighborhood. This is also our approach.

We should take note, however, of other conceptions. Some theorists (Wellman 2001; Miller 2007) believe that places are losing relevance to people's social connections and exposure to ideas and lifestyles, in part due to general improvements in transportation and communication technology and specifically due to social media and the Internet. More relevant to our thinking, others have retained geography in their conception but define neighborhoods egocentrically. The insight is that even people living adjacent to one another might experience very different local urban environments. Weber and Kwan (2003), for example, suggested that through the conduct of daily activities, individuals construct "personal cities." Matthews (2011) similarly argued that people are spatially polygamous and they have intimate attachments with multiple places and therefore measures of context should account for this polygamy. Hagerstrand (1982, 324) urged geographers to "rise up from the flat map, with its static patterns and think in terms of a world on the move, a world of incessant permutations." To this day, this kind of dynamic thinking remains a challenge and constitutes an important subdiscipline of GIScience. Kwan (1998); Miller (1991, 2005); Laube, Imfeld, and Weibel (2005); and others have made important progress in the visualization, computation, and analysis of dynamic time-geographic patterns. Kwan (2009) argued that these dynamic approaches are better able to capture the social and environmental exposures faced by people in their daily lives.

Another approach to egocentric neighborhoods is to consider every person's residential or work location as the center of his or her neighborhood (Chaix 2009; Chaix et al. 2009). The idea that neighborhoods exist "around" people is reinforced by studies of people's perceptions. Field experiments by Coulton et al. (2001) found that most residents placed their own home at the center of their neighborhood. Survey research by Hunter (1974) reported that neighborhoods had

"rolling" boundaries—people might agree on the name of their neighborhood, but those living near its edge tended to perceive it as extending further in that direction. More often, though, egocentric neighborhoods are constructed from documentary sources. For example, Frank, Engelke, and Schmid (2004) defined a 1-km circle around a person's home to study the effect of the built environment (street networks, land use mix, population density) on travel behavior. Static person-based neighborhoods can also be constructed at multiple scales. Lee et al. (2008) used a series of concentric rings to estimate spatial measures of residential segregation at different distances from a person's home.

The fact that there are multiple ways to define neighborhoods and that these different methods often disagree has significant implications for urban spatial analysis. Our approach is place based, by which we mean that we seek criteria by which we can define areas as different neighborhoods on the basis of their attributes. We introduce two features that are consistent with person-based approaches: (1) the starting point for our analysis is the construction of egocentric neighborhoods around individual persons, and (2) we make minimal assumptions about the relevant geographic scale. Our data are limited, however, to static information about people and their residential locations. It might be useful to think of neighborhoods as a fixed component of people's broader activity spaces, a component that they share with others who live near them. We leave to future studies the question of whether the neighborhood in this sense is more or less consequential for their lives than the actual locations that they routinely visit (their personal cities).

Neighborhoods in the Nineteenth Century

This is a study of neighborhoods in U.S. cities in the late nineteenth century. To draw out hypotheses about their composition and scale, we first describe the data source, pointing out both its limitations and its advantages. Then we offer a brief overview of the accounts provided by urban historians about this period.

The data used in these analyses were compiled by the Urban Transition Historical GIS Project (UTP) at Brown University (Logan et al. 2011; see also http://www.s4.brown.edu/utp). Historical geographic information systems (GIS) like the UTP are becoming increasingly common. Gregory and Healy (2007) noted that using GIS in historical inquiry allows one to ask questions about pattern and distribution in ways that archival sources outside of a GIS context do not. In

spite of the potential for historical GIS to provide novel insights into the geographic organization of society and its evolution, Gregory and Healy (2007) also observed that analytical studies of historical data in the context of GIS are rare. Although we generally agree with this observation, there are some notable exceptions (e.g., Hershberg 1981; Dorling et al. 2000; Gregory 2008).

The UTP takes advantage of the 100 percent digital transcription of records from the 1880 Census that was organized by the Church of Jesus Christ of Latter-Day Saints and prepared for scholarly use by the Minnesota Population Center (MPC). For thirty-nine major cities, UTP has added addresses for all residents and is geocoding those addresses based on historical sources. For this study we selected a subset of four cities: Albany, New York; Buffalo, New York; Cincinnati, Ohio; and Newark, New Jersey. These cities had a combined population of more than 600,000 in 1880. All were among the top twenty-five largest cities in the United States.

The key criteria for selection were that these cities had similar overall ethnic composition, which is important because ethnicity is one of two dimensions along which we seek to identify neighborhoods. The only population groups that were more than 10 percent of the population were Germans, Irish, and Yankees. Germans and Irish are defined here as persons who were born or had at least one parent who was born in Germany or Ireland. Yankees are whites born in the United States with U.S.-born parents. These groups made up about 80 percent of the population in each city.

The other variable that we draw on is socioeconomic status, determined using the occupation of each working member of a household. Occupations were ranked on a scale from 0 to 100 based on the average education and earnings of persons in each occupation as measured in 1950. The resulting scale, the Duncan Socioeconomic Index (SEI), is commonly used as a measure of socioeconomic status. Sobek (1996) compared the average income of men in each of 140 occupations in 1890 to the income of men in those occupations in 1950, finding a correlation between the two of 0.93, concluding that the scale is valid for the earlier time point. For each household we use the maximum SEI among the household members. Additional variables that could be used in a more thorough study include age, household composition, and U.S. versus foreign birth.

Mapping begins with a contemporary census TIGER street file of every county, which requires considerable editing to be useful for 1880 (deletion of new roads and other features, insertion of roads that had been eliminated, and correction of street names that had changed). Key resources for the reconstruction of historic street maps include descriptions of the boundaries of enumeration districts from the 1880 census (not available for Newark), city directories that sometimes include address ranges for most streets, and detailed street maps that often identify the political wards within which enumeration districts were formed. The microdata include each household's address and the enumeration district within which it was counted. Geocoding began with efforts to map enumeration districts and then to place households along the streets in each enumeration district. In each of the cities in this analysis, more than 95 percent of addresses have been successfully geocoded.

The late nineteenth century was a dynamic time for urban America. Cities were undergoing a series of rapid transitions driven by immigration and technological change; the nature of urban life was changing (Borchert 1967). New modes of production and transport technologies played an important role in transforming the economic and residential landscape of cities. American cities were in transition from a preindustrial form to a modern industrial configuration. But in 1880 cities were only beginning their transition. Although inter- and intracity rail transport existed, on a day-to-day basis the average city dweller still likely walked everywhere (Hershberg 1981).

The main geographic debates about residential patterns in cities at this time hinge on this question of transportation. Warner (1962) argued that transportation technology, namely, the streetcar, played a central role in neighborhood differentiation in Boston. He hypothesized that transportation technology supported the development of a "two-part" city where home and work were geographically separated. Bruegmann (2005) made a convincing historical argument that in almost every era of urban history those who have been able to afford it have moved, at least part-time, outside of the city center. The expansion of transportation technology lowered the economic barriers to suburbanization and let increasing numbers of Americans leave crowded city centers, thereby increasing geographic separation of rich and poor (Warner 1962; Hall 1996). At the extremes the ability of the middle and upper classes to separate home and work by commuting into the inner city supported the development of large slums in many cities (Hall 1996). In the United States, this basic pattern, the geographic separation of home and work, of rich and poor, continues to have profound social and political consequences.

Another dimension of segregation was ethnicity. Even in preindustrial cities, ethnic neighborhoods (or quarters) were often maintained through religious or cultural norms (Sjoberg 1965). These ethnic quarters were viable because modes of production favored small workshops; A person's home and workplace were often the same structure. In the industrial era production shifted from small workshops to factories. But in 1880 intraurban transport was far too expensive to be a part of the daily life of most Americans (Hershberg 1981), so most workers still walked to their jobs. Residential location was driven by employment (Burstein 1981), and ethnic employment niches therefore generated ethnic neighborhoods.

The consequences of this dense concentration of employment for residential patterns in the late nineteenth century are unclear. One likely effect is that the scale of neighborhoods was small, because most daily life was carried out on foot. Hershberg (1981), in the Philadelphia social history project, matched 3,711 individuals to potential employers (identifying the nearest firm to an individual's home where, based on his profession, he could have been employed). Individuals with generic professions like "laborer" were omitted from the analysis; 97 percent of doctors, 93 percent of confectioners, 68 percent of blacksmiths, 65 percent of cabinetmakers and carpenters, 24 percent of lawyers, and 11 percent of bookbinders worked within 0.5 km of their home (Hershberg 1981).

Both ethnicity and class probably influenced people's residential locations, but there is little evidence of which was more important. The inner city, because it afforded access to the city center's mix of high- and low-paying jobs, might have tended to be class-diverse in the late nineteenth century. There is some anecdotal support for this suggestion; Charles Booth's street-by-street maps of poverty in London in the 1890s are striking today because they show an intricate patterning of "class" that seems more structured by the hierarchy of streets than by broad residential zones (Dyos, Cannadine, and Reeder 1982). Zunz (1982) argued that segregation in Detroit in 1880 was primarily based on ethnicity. He reported significant ethnic clustering (overrepresentation of a particular group) in 30 percent of the blocks that he sampled in 1880, but at that time it was common for ethnic blocks to include a wide range of occupations, from laborers to professionals and shopkeepers. Others have suggested that occupation effects might be group specific. Moore (1994) argued that Germans in New York settled near industries established by German entrepreneurs. They "constructed a fairly complete eth-

nic economy that included workers as well as a range of mercantile establishments . . . thus German ethnicity permeated the urban class culture of the neighborhood" in places like Bushwick and Williamsburg (145). The Irish, in contrast, "rarely concentrated in such numbers throughout a neighborhood that they created a complete local ethnic economy. Instead they fashioned an ethnic network through politics and the church which did not require significant residential concentration." Following this reasoning, one would hypothesize that German ethnicity was more salient than Irish ethnicity, and that occupation might have mattered less for Germans, because they were more likely to live in mixed-class neighborhoods.

Through the identification of neighborhoods we hope to address some basic questions about the residential landscape of late nineteenth-century cities:

- Were cities residentially differentiated by class, ethnicity, or some combination of both? Did language play an important role in the organization of cities? Were non-English-speaking immigrants (Germans) more separated from native whites than English-speaking immigrants (Irish)?
- Did class organize urban space? Transportation constrained the distance between residence and employment, and employment was concentrated in the city center. Had the suburban push begun or were central city areas still relatively well-off compared to peripheral areas?
- Finally, were there mixed neighborhoods, places in the city where ethnic or economic heterogeneity was the norm?

To explore such questions we must first identify neighborhoods.

Methods

Person-based representations of neighborhoods are situational in that they describe a person's position with regard to surroundings (Hagerstrand 1982). We identify neighborhoods by measuring how different types of people are situated with respect to each other through a novel application of local spatial statistics. We set conceptual boundaries through three simple assumptions about the nature of neighborhoods:

1. A neighborhood is a contiguous territory defined by a bundle of social attributes (Galster 2001). It is not a place where everyone is the same but

rather a place where the mix of people in terms of these attributes is distinct from surrounding areas; that is, where characteristics of individuals when aggregated form a thematic unit (in Lynch's terms). The thematic unit that defines neighborhoods is based on the characteristics of static person based neighborhoods constructed at multiple scales and captures the situation of people with respect to each other. Thematic units are place-based. Therefore, neighborhoods are regions in the sense that they contain groups of entities that are proximal and similar and that these groups of people are distinguished from other proximal groups (Montello 2003).

2. Neighborhoods have boundaries. Boundaries are defined by changes in the bundle of attributes between adjacent territories. These boundaries can be sharp, in which case they can be represented by linear features, or fuzzy (like a zone of transition), in which case they might be geographically extensive. Heikkila and Wang (2010) noted the utility of fuzzy set theory to urban research, arguing that the apparent dichotomy between urban and rural can be better addressed through fuzzy membership functions. Lynch (1960) described spatially extensive boundaries between urban districts. Neighborhoods, like colors of the spectrum, can bleed into each other.

3. Neighborhoods are not mutually exclusive. Every part of a city is in a neighborhood, but a location can be simultaneously in multiple neighborhoods. Neighborhoods are defined by a core and edges. Core areas unambiguously belong to a single thematic unit, but edge areas are often associated with multiple thematic units.

Egocentric Neighborhoods: An Example

Consider a simplified city that has two types of residents, light gray and dark gray (Figure 1). Light gray people are the minority group and dark gray people are the majority. In Figure 1 it is clear that the two groups generally live in separate parts of the city, and one might think of it as a city with two neighborhoods. In the center of the city, though, the groups seem to be more mixed. The question we investigate has two parts: (1) Is this area of overlap a diverse neighborhood in its own right, or is it the edge between two different monoethnic neighborhoods? (2) If there are two or three neighborhoods, what are the boundaries between them?


Figure 1. A simplified city showing two groups, light gray (the minority group) and dark gray (the majority group).

Our approach is based on the egocentric framework developed by Lee, Reardon, O'Sullivan, and others in a series of papers (Lee et al. 2008; Reardon et al. 2008; Reardon et al. 2009; Reardon and O'Sullivan 2004). In this approach each person is understood to be at the center of a series of concentric circles and the composition of each circle is used to describe the area around the person at a specific scale. If we superimposed concentric circles on the numbered people in Figure 1 and summarized the ethnic composition of each circle, person 1 would have a different profile than person 2. Person 1 is in a homogeneous dark gray context at short distances, but at greater distances the context becomes more mixed. Person 2 is in a mixed context at all scales. The profile of each person could be summarized by two curves, one describing the prevalence of light gray people as a function of distance and another describing the prevalence of dark gray people. Each individual has a distinct set of curves, the shape of which is determined by the spatial pattern of ethnic settlement and that person's location within it.

An individual person's profile can provide clues about the overall spatial pattern. Consider the profile in Figure 2 for a person in a light and dark gray city. At small distances the light gray group dominates, but as one moves outward, light gray people become an ever declining minority. This profile is suggestive of what we might call a minority neighborhood with a fairly abrupt edge, beyond which most people are dark gray. One might even consider the boundary of this person's neighborhood to be the inflection point of the light and dark curves.

Figure 2. The signature of a minority neighborhood in a city with two groups, light gray (minority) and dark gray (majority).
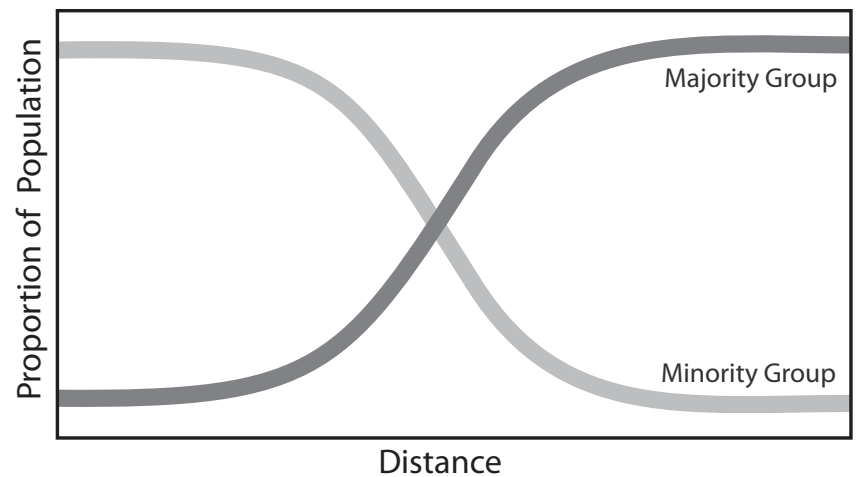
The bundle of attributes defining these profiles is at our discretion. For example, suppose we added information about people's socioeconomic status (SES) to each person's profile (the dashed in Figure 3). Figure 3 shows two possibilities. The left image depicts what we might call an affluent minority enclave because SES is highest at small distances; the right image suggests a low-income minority ghetto. Conceptually, there is no limit to the number of attributes that could be incorporated into a neighborhood signature (with the caveat that edge effects must be controllable).

Place-based neighborhoods cannot be based on a single person's egocentric scale profile. Yet when mapped onto geographic space the profiles of individuals will exhibit a high degree of spatial autocorrelation, allowing us to identify neighborhoods as clusters of people with similar profiles. Suppose we calculated profiles for all of the individuals in the red and green city depicted in Figure 1. The people next to person 1 would have similar curves (predominantly dark gray except at larger distances) and hence could be classified as members of the same type of neighborhood. People near person 2

would have curves that are more mixed at all distances. People in the light gray area would generally have curves with a high share of light gray, declining with distance. As the size of the ring increases, the spatial autocorrelation among individual signatures increases. At the extremes, if the outermost ring around each person encompassed the entire city, there would be no difference between individuals at that scale. This reminds us that at some point, increasing the size of the distance band adds little new information to the profile. Finding an appropriate scale that captures meaningful variation is critical to the design of neighborhood profiles.

Now that we have a profile for every person, we notice spatial autocorrelation in the profiles, and our problem is how to use these data to identify neighborhoods.

From Egocentric Signatures to Neighborhoods

Neighborhoods are regions that encompass multiple scales and multiple attributes. Egocentric scale profiles describe multiple attributes over multiple scales, but
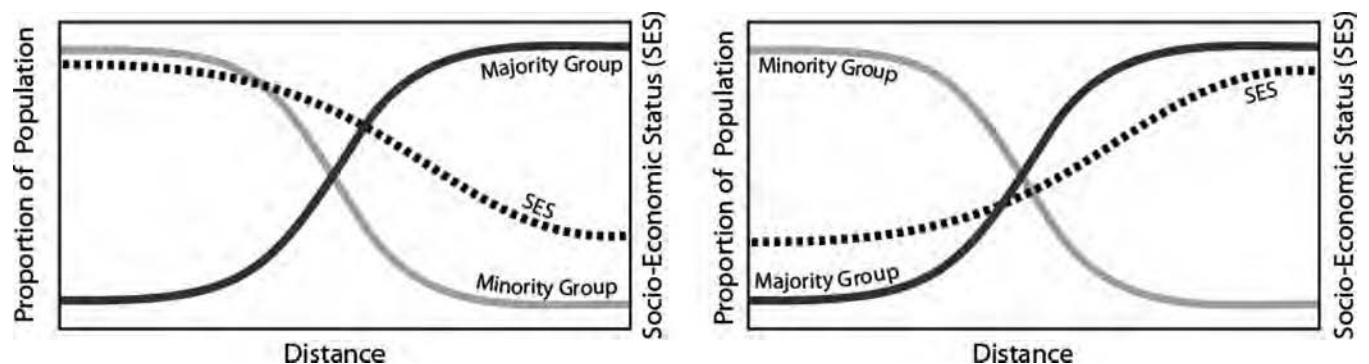
Figure 3. Extending a neighborhood signature to include multiple attributes. The addition of a third curve representing socioeconomic status (SES) alters the interpretation of Figure 2.

they are not regions. A region can be represented as a thematic unit using a generalized scale profile, as in Figures 2 and 3, where we sought to represent different kinds of minority neighborhoods. If one knew, or had a hypothesis, about the kinds of neighborhoods that existed in a city, one could sketch prototypical thematic units. Each person's observed egocentric profiles could be compared to the theoretical thematic units for fit. This deductive approach would be like image processing techniques commonly employed in remote sensing—the social signature of a neighborhood can be thought of like the spectral signature of vegetation or land cover (Campbell 2008).

Our understanding of cities in the late nineteenth century is not well enough developed to formulate a set of neighborhoods (thematic units) a priori. Instead, we use an inductive procedure in which we search the set of all egocentric signatures for people who have similar signatures. Our approach is a type of local spatial analysis, inspired by Getis and Franklin (1987), who developed a multiscale second-order neighborhood analysis, a local version of Ripley's K-function, to study the clustering of trees in a study area. One of the attractions of the K-function, like the egocentric signature, is that it allows one to assess patterns at multiple scales. We depart from the Getis–Franklin method in two ways. First, the K-function is based on a count of events and is typically evaluated by reference to somewhat arbitrary point generating processes (like a Poisson point process or Neyman–Scott process). Second, and related to the first, methods for the analysis of geolocated individuals (or other types of points) are often concerned with statistical significance. Getis and Franklin used the local K-function to make statistical inferences about geographic patterns, and this required that they sort out systematic variation in the location of trees (clustering) from background noise. Our use of egocentric signatures is more strictly descriptive, a means to characterize the environment around each person.

There are a number of different ways to identify groups of similar observations in a data set, generally known as cluster analysis. Cluster analysis is an unfortunate name because a cluster can be a geographic entity or a group of proximal events. In this case, a cluster is a group of people with similar egocentric scale profiles who might or might not live near each other. Among nongeographic cluster analysis methods, hierarchical and relocation techniques are the most common. Hierarchical cluster techniques are either agglomerative or divisive—that is, in each subsequent step of the clus-tering routine, they either combine observations into a group or divide existing groups. By contrast, relocation techniques begin with a user-specified number of clusters, defined by a centroid. This centroid is iteratively relocated until some stopping criterion is reached. Both hierarchical and relocation cluster analyses are awkward when it comes to determining the most appropriate number of groups in a data set. In addition, both techniques typically result in discrete class assignments in which a person (or location) can belong to only one group. Our conceptualization of neighborhoods is not discrete, a person can belong to multiple neighborhoods, for example, if they live in the transition zone between neighborhoods. Our suggested approach is probabilistic. We use both substantive and statistical criteria to determine the number and types of neighborhoods.

The simplified city depicted in Figure 1 has seventy residents, who we can denote as $y = (y_1, y_2, y_3, \ldots, y_{70})$. In the simplified city where the signature consists of only two curves y_i would be a vector of length $2|H|$ where H is the set of concentric rings used to construct the egocentric scale profile. Suppose that the simplified city consisted of a set of K types of neighborhoods (thematic units). We would like to estimate a vector $z = (z_1, z_2, z_3, \ldots, z_{70})$ where z_i is a vector of length $|K|$ indicating the probability that person i is in each type of neighborhood. Generally, when the bundle of attributes defining a neighborhood contains v variables y_i has a length of $v|H|$. To estimate z, the probability that each person belongs to each type of neighborhood, we must determine both the number of neighborhoods and the parameters defining each group; we do this using a model-based clustering procedure (Fraley and Raferty 1998; Witten and Frank 2005) based on the expectation maximization (EM) algorithm (Dempster et al. 1977).

Model-based cluster analysis assumes that the observed egocentric profiles (y) are generated by a set of K neighborhoods. Each of the K neighborhood types is described by a $v|H|$ dimensional Gaussian distribution with parameters μ and $\sum$. The model-based clustering procedure is used to estimate the parameters of the distribution defining each neighborhood. Model-based cluster analysis allows us to evaluate the set of neighborhoods in each solution $K = (k_1, k_2, k_3, \ldots, k_j)$, $j = 2, \ldots, j$ by calculating a formal measure of fit:

$$likelihood = \sum_i log \sum_j P(y_i|k_j)P(k_j)$$

This likelihood criterion simply assesses the conditional probability of the y_is given a particular set of

neighborhoods k_j. Specifically, we use the negative log likelihood, $-2log(likelihood)$ to search for the most appropriate number of neighborhoods. The solution that minimizes the negative log likelihood is the number of groups that provides the best fit to the data. Because the model-based clustering procedure is sensitive to initial conditions, each solution of j neighborhoods must be cross-validated using v-fold cross-validation. The cross-validation procedure involves fitting m models for each j by taking repeated random samples of 80 percent of the data. Plotting the average negative log likelihood for each of $m * j$ models provides some insight into the most appropriate number of neighborhoods. The negative log likelihood will always decrease as the number of neighborhoods j increases, so some judgment is required to balance parsimony and model fit. There is no guarantee that the "optimal" solution, the one that minimized the negative log likelihood, would be meaningful from an ecological or historical perspective. That is, the natural groups in the attribute space might not make sense when mapped onto geographic space.

Once we have estimated z we can map the results. Our expectation is that neighborhoods are all defined by a core and an edge. Residents of a core area (person 1 in Figure 1) belong to a single type of neighborhood. Depending on the nature of the edge, however, residents of the transition zones between neighborhoods (e.g., person 2 in Figure 1) might have a significant probability of belonging to multiple neighborhoods. The boundaries between neighborhoods are not always clearly defined. Individuals, like person 2 in Figure 1, can live in the interstices between more homogeneous regions. Estimating z (the probability that each person belongs to each type of neighborhood) should allow us to account for the indeterminacy or "fuzziness" of boundaries.

Computing Scale Profiles and Neighborhoods

In seeking balance between historical and methodological detail, we explore these questions by pooling data from four cities. Golledge (2002) noted that as geographic analyses increase in scale, geographic knowledge tends to become more categorical. In our case were there regular types of neighborhoods that occurred across each of the four cities? Did each city have unique types of neighborhoods because of its particular historical and geographical setting? Pooling these cities allows us to focus on general questions about neighborhood formation and sociospatial structure of cities in the late nineteenth century.

To reduce the computational load and redundancy we combined all of the residents living at a single residential address, resulting in 99,356 unique addresses. Each building was described by the number of people belonging to the Irish, German, Yankee, or "other" ethnic category. Every building was also assigned an SEI score based on the average SEI of the breadwinner (the adult with the highest SEI) in each household.

We consider a circle with a 0.5-km radius centered on a person's home to be a reasonable approximation of most people's "life environment"—that is, the environment experienced in the course of their daily activities. This definition might be conservative, but we find that beyond 0.5 km the egocentric profiles do not change much in response to marginal increases in size.

The signature for each address in the four cities ($N = 99,356$) contained four curves, one for proportion of the population that was German, one for Irish, one for Yankee, and one for SEI. Other ethnicities were not separately modeled because they represented a small portion of the population, but they were included in the denominator for density measurements. The curves were static person-based neighborhoods measured at eleven scales. The smallest scale was the building level and the subsequent ten measurements were concentric rings at 50-m radial increments, making the maximum diameter of the circles defining the egocentric profiles 1 km.

The 99,356 signatures describing the ethnic composition and income of each building in the four cities were analyzed using the model-based clustering procedure outlined in the previous section. We estimated z, the probability that each person belonged to each of K neighborhood types, where the number of types of neighborhoods ranged between two and twelve. A six-neighborhood solution was chosen because increasing the number of neighborhoods beyond six led to small changes in the negative log likelihood. On careful inspection, we determined that the six-neighborhood solution provides a set of neighborhoods that is both parsimonious and ecologically meaningful.

Results

Figure 4 shows three examples of signatures of specific buildings (with SEI omitted for simplicity). The ethnicity curves are scaled between 0 and 1, where 1 indicates that 100 percent of the population at a particular scale belonged to a single ethnicity. The first signature describes building 27706, where the population is not Irish, German, or Yankee. Moving outward

from the building scale the population is almost entirely Irish at 50 m, with a small but growing Yankee and German presence at greater distances. The second example (building 61412) is the signature of an Irish building in a Yankee neighborhood. At 50 m the Irish density is only around 0.4 and falling below 0.2, as the Yankee and German densities rise. Interestingly, individual buildings are sometimes quite different from their surroundings. The third profile describes building 75424, which is ethnically mixed but clearly in a German neighborhood; as one moves out from the building scale the German ethnic group is increasingly dominant.

In Figure 5 each building has been assigned to the neighborhood type for which it has the highest probability. Figure 5 shows the average prevalence of each ethnicity and the average SEI as a function of distance for each of the six neighborhood types. To facilitate visual comparisons we have rescaled SEI to a 0 to 1 range, placing it on the same scale as the measures of ethnic prevalence. For example, the line along the top of the left-most graph in Figure 5 represents class 1 in red. We

find that class 1 signatures on average are characterized by very high German share, slowly declining with distance, and with lower SEI than most other classes. Class 5 (orange) is also nearly 50 percent German at all distances, but it has more Yankees and Irish and a higher SEI than class 1. Table 1 summarizes each of the classes.

Figure 5 shows the average values for each class as a function of distance, but further information is found in the variance around the mean. Figure 6, which only reports SEI curves, shows this variance for a random sample of 10,000 buildings. The differences in means are visible (e.g., the green curves tend to be higher SEI than the red curves). There are also differences in the amount of variation. For example, the neighborhood type (green, class 3) with the highest average SEI has much less variability than the lowest type (class 4), meaning that wealthy neighborhoods have less variation in income than poor neighborhoods. This variation for all classes emphasizes that neighborhoods are defined by a thematic unit, a distribution with parameters μ and $\sum$. Some buildings with the most
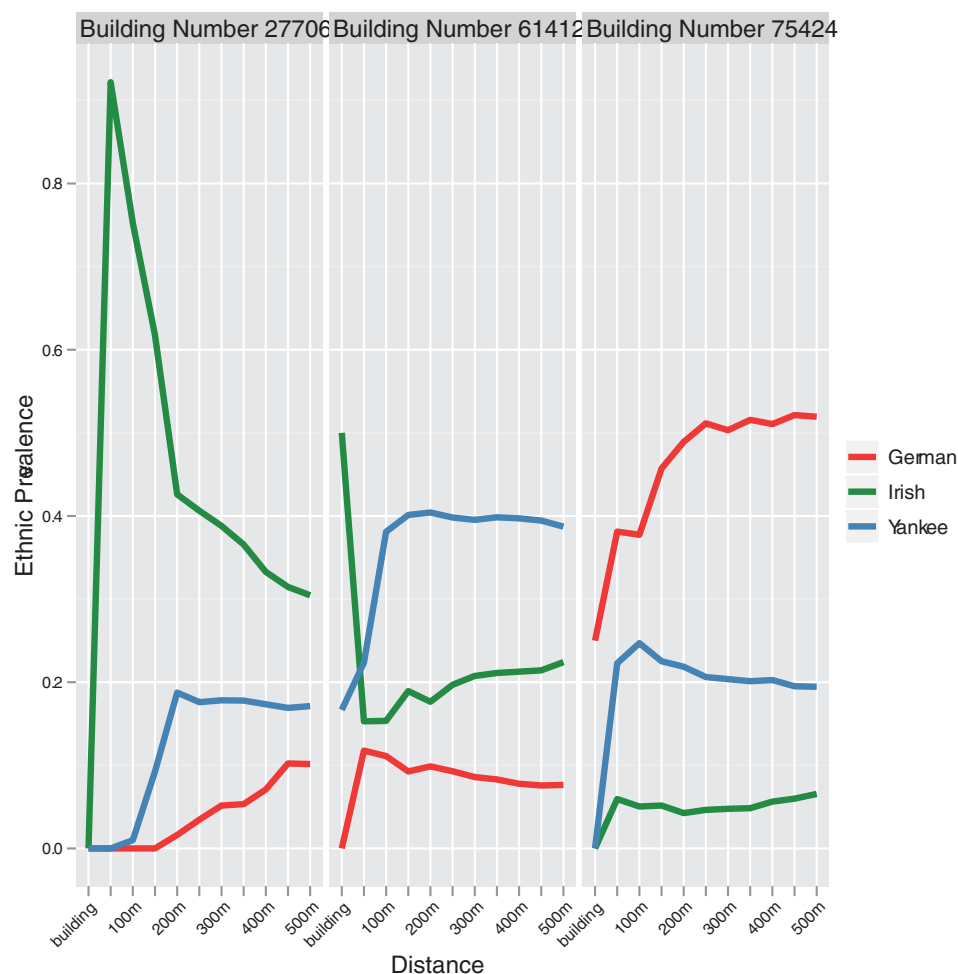
Figure 4. Profiles of three randomly selected buildings. Curves represent the prevalence of each ethnicity within a series of concentric rings. The horizontal axis (distance) refers to the radius of each ring. (Color figure available online.)

Figure 5. Neighborhood signatures: Average ethnic prevalence and socio-economic index (SEI) for each neighborhood type. *Note*: Colors correspond to regions in Figure 7. (Color figure available online.)
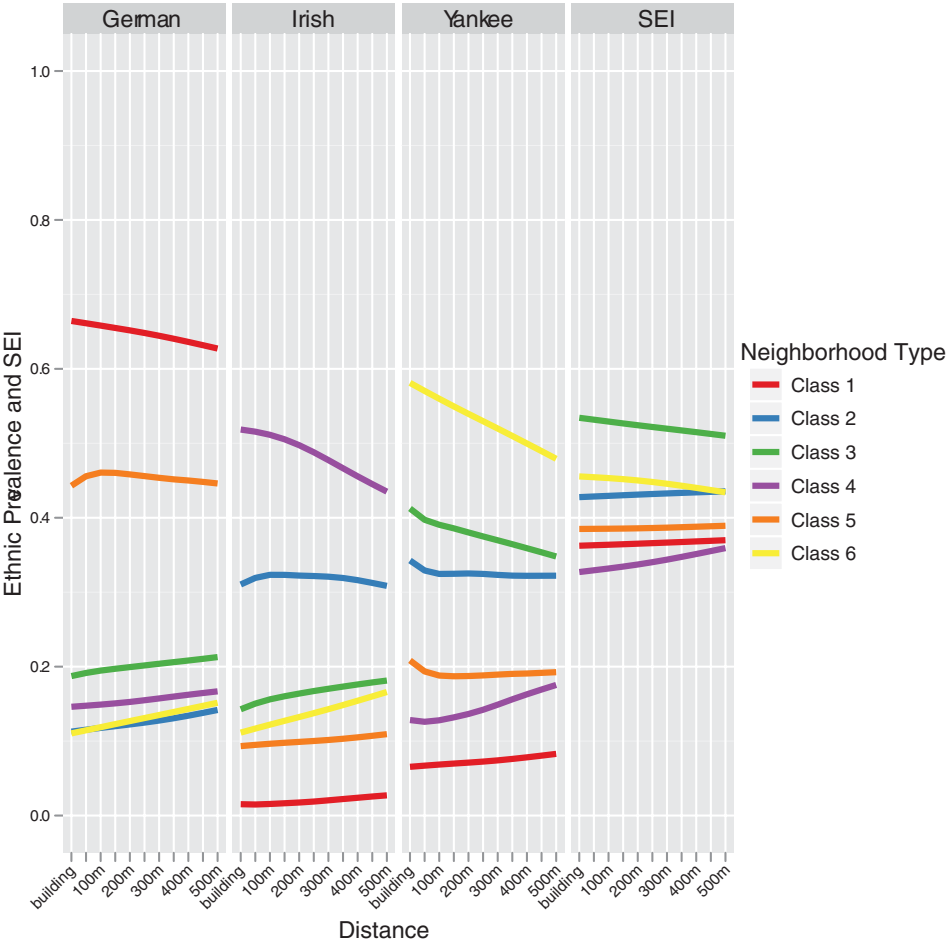
Table 1. Neighborhood types in Newark, Buffalo, Albany, and Cincinnati

Class	Color	Description	Location
1	Red	Low income, predominantly German, little ethnic mix	Near city center, often large zones
2	Blue	English speaking, Irish and Yankee, moderate income	Near city center, often borders classes 3 and 6
3	Green	High income, ethnically mixed	Near city center
4	Purple	Very low income, predominantly Irish	Edges of city, often waterfront
5	Orange	Low to moderate income, German with some ethnic mix	Surrounds lower income German areas in most cities
6	Yellow	Yankee, high income, moderate ethnic mix	City center and edges of city

probable type as class 3 (on average, very high SEI at any distance and high Yankee share especially at short distances) actually have profiles for which SEI is low at short distances (but then quickly increases with greater distance).

Each of these classes is defined solely by the model-based cluster analysis of the scale profiles of each building. The model-based clustering procedure is entirely aspatial, so there was no explicit spatial constraint in the clustering procedure. It was entirely possible that the classes identified in the model-based cluster analysis would not exhibit meaningful spatial patterns. For example, it could be the case that the classes that emerged would be city specific: Neighborhoods in Buffalo might be different from those in Cincinnati. It could also be the case that groups identified through the cluster analysis when mapped would form a dense irregular patchwork. If buildings in the same class were not spatially clustered, all geographic areas would be composed of a mixture of neighborhood types and we
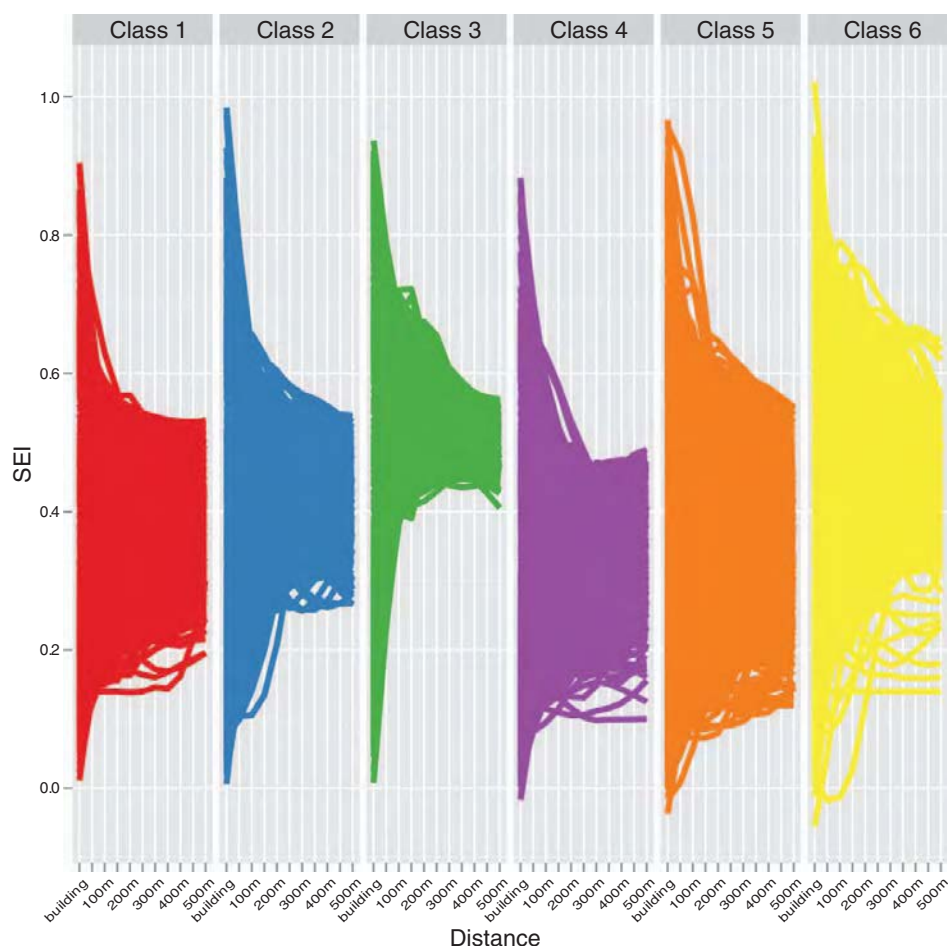
Figure 6. Variations in household income (SEI) within neighborhood types. Each subplot shows SEI profiles for a random sample of buildings. Neighborhood types are differentiated by both the mean income at each scale and the variability in income at each scale. (Color figure available online.)

would not be able to identify "neighborhoods" with the characteristics that we posited at the start.

Figure 7 maps the most probable class for every building in our study cities. The spatial clustering is apparent. We see large areas of class 6 (yellow: Yankee and relatively high SEI) in Albany and Newark, with smaller such areas in Buffalo and Cincinnati. Class 3 (green: Yankee and highest SEI) areas are prominent in center-city portions of every city but covering a very small area in Albany. The large red, German, and relatively low SEI class 1 zones are highly concentrated in Buffalo, Newark, and Cincinnati but are not found in Albany. Class 5 (orange: plurality German but with substantial Irish and Yankee minorities and modest SEI) areas of Cincinnati cover a large area extending to the north and more peripheral areas of the other three cities. In every city except Albany class 5 tends to surround class 1 zones, suggesting areas of transition from a German core to a less German buffer area. Class 4 (purple: Irish and very low SEI) areas are found around the edges of the

city, which people familiar with these cities will recognize as waterfronts in Albany, Buffalo, and Cincinnati; there is a similar riverside class 4 area in Newark and another larger such zone in the northwest of the city. Class 2 (blue: mixed Irish and Yankee, medium SEI) is a major feature of Albany and Newark but less prominent in Buffalo and Cincinnati. Classes 2 and 4 (both with many Irish residents) and classes 2 and 6 (both with many Yankees) seem generally to be adjacent to one another.

It is interesting that in the highest SEI neighborhoods (Class 3) Irish and German immigrants together make up 30 to 40 percent of the population compared to 50 to 60 percent Yankee. These neighborhoods are almost always in the city center. The concentration of affluence in the city center suggests that in the transportation-constrained 1880s, the benefits of lower density living were more than offset by the lack of economic opportunity. This is significant when one considers just how dense these cities were: Parts of Cincinnati

Figure 7. Neighborhood types in four cities. *Note*: Colors correspond to neighborhood signatures in Figure 5. (Color figure available online.)
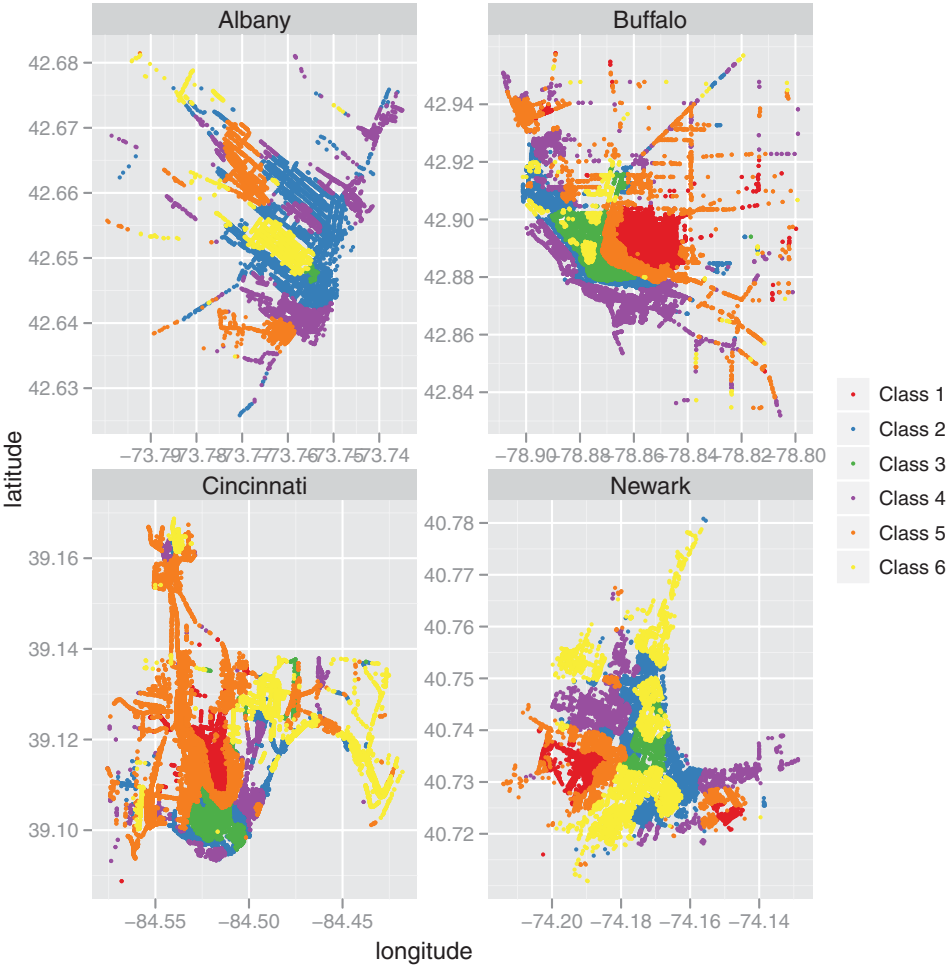
contained more than 30,000 people per square kilometer, Albany and Newark contained nearly 20,000 people per square kilometer in their densest sections, and Buffalo exceeded 14,000 people per square kilometer in its densest spots. This level of density was occurring in a time when the vast majority of buildings were two or three stories tall (Hershberg 1981).

As a counterpoint to the concentration of wealth in the city center there are also clear pockets of affluence on the edges of cities. In both Cincinnati and Newark there were well-off Yankee-dominated neighborhoods (class 6) on the edge of the city. In ongoing work we are examining these pockets of affluence in more detail; they seem to be associated with the annexation of nearby municipalities by the city and the presence of streetcar lines. For the city of Newark we have mapped the streetcar lines as they appeared in 1889, nine years after the 1880 census. For each building we computed the distance to the nearest streetcar line. Streetcars increase the accessibility to the city center. Following Alonso (1964), one would

expect this increased accessibility to translate into the price of housing. The correlation between streetcar access and the SEI measured at various geographic scales is quite high and in the expected direction (Table 2).

Table 2. Pearson correlation between distance to the nearest streetcar line and Duncan Socioeconomic Index (SEI) for Newark, New Jersey: 1880

SEI	r
Building SEI	−0.15
SEI 50 m	−0.30
SEI 100 m	−0.37
SEI 150 m	−0.39
SEI 200 m	−0.39
SEI 250 m	−0.39
SEI 300 m	−0.37
SEI 350 m	−0.37
SEI 400 m	−0.37
SEI 450 m	−0.35
SEI 500 m	−0.33
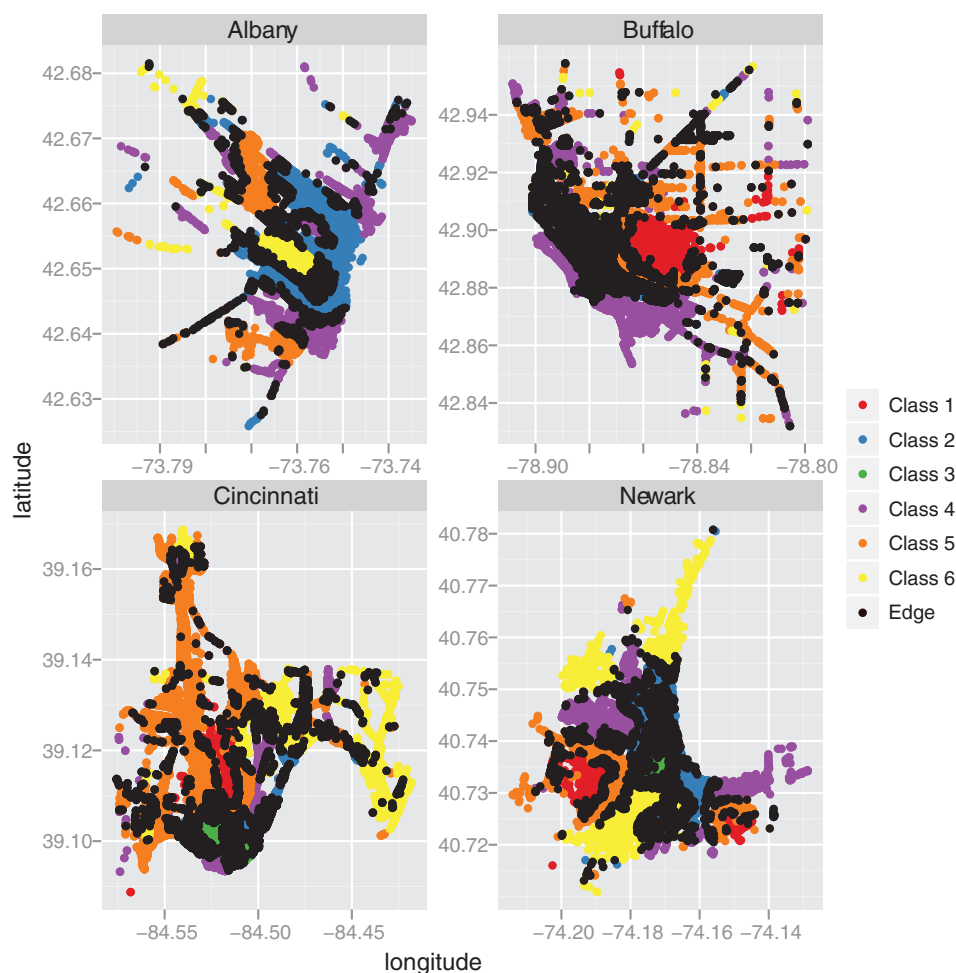
Figure 8. Black dots indicate edges of neighborhoods: Buildings that have a probability of belonging to multiple neighborhood types. (Color figure available online.)

As a last step we consider the question of neighborhood edges. In the core of a neighborhood, buildings are assigned to a single neighborhood type. At the edges of a neighborhood there is some uncertainty. Let us define an edge as a building with a probability of belonging to the assigned type of neighborhood that is less than 1. Figure 8 shows these edges as black dots. These edges are sometimes lines surrounding the core of a neighborhood. For example, in parts of Newark the edges fall along some major streets (e.g., Broad and Market streets). Major urban thoroughfares are logical dividing lines between districts and the fact that our method identified these edges is reinforcement for our method. In other places, large clusters of buildings are classified as an edge. For example, by our definition, most of the high-SEI sections of Albany and Buffalo (class 3) are edges because they have a chance of belonging to class 2 or class 6. Perhaps there is a type of neighborhood that is unique to Albany and Buffalo (combining features of classes 2, 3, and 6); perhaps in Albany this area is the extended edge of class 2 (the

type of neighborhood that surrounds it on three sides). The edges that we have identified warrant further investigation and formal characterization, a project we hope to pursue in the future.

Although the six neighborhood types do not appear in all cities, each has at least five. Albany is most distinctive because it does not have a large pronounced German low-SEI enclave. There is a German neighborhood but it tends to be higher income than those in the other cities. The central city of Albany is dominated by a high-income, high-Yankee neighborhood, whereas most cities have a higher income but slightly more mixed population in the core.

Conclusion

In 1880 there were no municipal administrative units that could be justifiably called neighborhoods. Political wards were quite large in most cities; electoral precincts were quite small. Enumeration districts were used by the

Bureau of the Census, but they were laid out mainly for the convenience of supervising data collection with no intention of representing natural social areas. Walter Laidlaw, the originator of the census tract, would not conceive of his "convenient and scientific city map system" for another twenty-five years. What is the correct unit of local analysis?

We define neighborhoods as categorizations of urban space defined by both characteristics of individuals (or residential units) and their spatial context. We assess both individual (building) and contextual characteristics through the development of egocentric scale profiles. Our purpose is to identify neighborhoods based on the composition of their residents at multiple scales—in this case Irish, German, or Yankee neighborhoods, rich or poor neighborhoods, and homogeneous or mixed neighborhoods. Our method does not depend on prior knowledge of the types of neighborhoods or their number, and it allows for the possibility that the boundaries between neighborhoods might be sharp edges or larger zones of transition. We believe this method, or similar methods, can contribute significantly to social scientists' understanding of what we mean by a neighborhood. This method is increasingly relevant as rich sources of disaggregate spatial data become available. We are not suggesting that ours is the only solution to the definition of neighborhoods. Our method is not appropriate for assessing individual exposures but rather is a way to identify the social structure of a city. The empirical example presented here considers only two attributes, ethnicity and class. Some readers might prefer to think of the results as social areas or as spatial clusters of these specific attributes and to reserve the term *neighborhood* for community areas that combine a larger array of common features. Regardless of how much information is used to identify these areas, however, the key to our approach is not what we call these units but what they tell us about the social structure of cities.

The neighborhood types indicated on the maps are based on a local statistical analysis of data pooled across four cities. The "local turn" in spatial analysis is often associated with descriptive analyses of particular relationships in particular places. Local spatial techniques are not designed to build generalized or generalizable models (Fotheringham 1997). In this analysis, however, by applying the same technique to multiple cities, we have found that common patterns are evident. These cities were selected because they had similar ethnic compositions, but we could not be assured that there would be geographic regularities in their spatial structure be-

cause they have very different geographic settings and historical circumstances.

Our procedure bridges work on geodemographic classification and regionalization. *Regionalization* is the process of defining regions within a geographic data set. Graph-based algorithms by Assuncao et al. (2006) and Guo (2008) provide a means to flexibly define regions. These algorithms face computational constraints when dealing with a large number of entities and are ill suited to the analysis of discontiguous study sites like ours. Geodemographic systems are typologies in which the city is viewed as a collection of distinct types of places (Longley and Tobon 2004; Harris, Sleight, and Webber 2005). A key difference between regionalization and geodemographic cluster analysis is that the former includes an explicit spatial constraint. We use local statistics to create a geodemographic system. Because local statistics computed for overlapping regions exhibit spatial autocorrelation, aspatial classification techniques will tend to produce regions. Hence, our procedure could be considered a "soft" regionalization technique.

A concern with either of these methods is whether they produce meaningful results. Both are inductive means of summarizing information, and neither is assured of being "right." One potential problem with our approach is that spatial autocorrelation could potentially lead to spurious classifications. If the profile is overly large, meaningful small-scale variation might be overwhelmed by meaningless large-scale similarities. Additionally, cluster analyses tend to be disproportionately influenced by the variables with the most variance. Both of these problems can be addressed by weighting variables. Another concern is that the criteria for the appropriate number of clusters are difficult to quantify. Is the appropriate number of neighborhoods the one that minimizes or maximizes some criterion, or is the appropriate number of neighborhoods the solution that is the most ecologically meaningful?

In our case, however, we had no substantive foundation for weighting variables—the relative importance of SES and ethnicity in the differentiation of cities was part of our question. We also lack a theory of neighborhood types that would provide some insight into the appropriate number of neighborhoods. Although the neighborhoods we identify mesh with our understanding of nineteenth-century cities, it is possible that the patterns are simply an artifact of the method. Therefore the neighborhoods we present here would be more convincing if there were some form of external validation, which requires detailed archival research. We

have started the validation process by mapping the churches in Newark, New Jersey, and identifying the language used in church services. This allowed us to map eleven German churches. Of these, eight were in the core or edge of neighborhoods that we identified as German. Of the three German churches that were outside of German neighborhoods, one was located downtown and two were within a few hundred meters of German neighborhoods. The initial validation results are promising, but much more needs to be done.

Generally, uncertainty about the definition of neighborhoods raises concerns about the modifiable areal unit problem (MAUP). The MAUP concern is that as one changes the size and shape of areal units, the relations one observes (e.g., the size and direction of observed regression and correlation coefficients) can change, sometimes dramatically (Openshaw and Taylor 1979; Arbia 1989; Fotheringham and Wong 1991). In analyses that use fixed areal units like census tracts, the MAUP is an important but hidden concern, revealed only when researchers are able to repeat analyses using different units (Hipp 2007). With person-based measures of neighborhoods, however, the MAUP is exposed and vexing. Person-based egocentric neighborhoods, whether dynamic or static, give researchers enormous flexibility in the definition of neighborhoods, potentially exacerbating the MAUP by making neighborhoods highly modifiable. Kwan (2009) suggested that dynamic person-based neighborhoods "can go a long way" toward addressing the MAUP because they are "frame independent" (Tobler 1989).

Although there are many perspectives on the MAUP, it is our view that the solution to the MAUP is to minimize the modifiability of geographic units of analysis. The MAUP is frequently viewed as a technical concern, but we believe, following King (1997) and Openshaw (1996), that it is best conceived as a theoretical problem rooted in the conceptualization of the unit of analysis. Questions about the definition of neighborhoods create ambiguity about their geographic realization. Our research illustrates one approach to the MAUP, the explicit "spatialization" of a social-scientific unit of analysis. Through this spatialization we have done our best to reduce the modifiability of the neighborhoods we identify. We do this by carefully linking our method to a conceptualization of neighborhoods based on three simple assumptions and using our limited knowledge of individual spatial behavior in the late nineteenth century to set the maximum extent of the egocentric profiles. The advantage of a historical study is that we have a public source of population information at the smallest possible spatial unit, the household. But the questions we raise and the approach that we take to answering them are relevant to modern studies of urban social structure. As new technologies make high-resolution disaggregate spatial data increasingly available, the soft regionalization procedure we describe here could prove to be a useful tool to develop detailed maps and statistical descriptions of regions.

References

Alihan, M. 1938. *Social ecology: A critical analysis.* Lanham, MD: Cooper Square.

Alonso, W. 1964. *Location and land use: Toward a general theory of land rent.* Cambridge, MA: Harvard University Press.

Arbia, G. 1989. *Spatial data configuration in statistical analysis of regional economic and related problems.* Dordrecht, The Netherlands: Kluwer Academic.

Assuncao, R., M. Neves, G. Camara, and C. Da Costa Freitas. 2006. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science* 20 (7): 797–811.

Booth, C., and H. Pfautz. 1967. *Charles Booth on the city: Physical pattern and social structure: Selected writings.* Chicago: University of Chicago Press.

Borchert, J. 1967. American metropolitan evolution. *Geographical Review* 57 (3): 301–32.

Bruegmann, R. 2005. *Sprawl: A compact history.* Chicago: University of Chicago Press.

Burgess, E. W. 1925. The growth of the city. In *The city*, ed. R. E. Park, E. W. Burgess, and R. D. McKenzie, 47–62. Chicago: University of Chicago Press.

Burstein, A. 1981. Immigrants and residential mobility: The Irish and Germans in Philadelphia, 1850–1880. In *Philadelphia: Work, space, family, and group experience in the nineteenth century*, ed. T. Hershberg, 174–203. New York: Oxford University Press.

Campbell, J. B. 2008. *Introduction to remote sensing.* 4th ed. New York: Guilford.

Chaix, B. 2009. Geographic life environments and coronary heart disease: A literature review, theoretical contributions, methodological updates, and a research agenda. *Annual Review of Public Health* 30:81–105.

Chaix, B., J. Merlo, D. Evans, C. Leal, and S. Havard. 2009. Neighborhoods in eco-epidemiologic research: Delimiting personal exposure areas: A response to Riva, Gauvin, Apparicio and Brodeur. *Social Science & Medicine* 69:1306–10.

Coulton, C. J., J. Korbin, T. Chan, and M. Su. 2001. Mapping residents' perceptions of neighborhood boundaries: A methodological note. *American Journal of Community Psychology* 29:371–83.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39:1–38.

Dietz, R. D. 2002. The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research* 31:539–75.

Dorling, D., R. Mitchell, M. Shaw, S. Orford, and G. S. Smith. 2000. The ghost of Christmas past: Health effects of poverty in London in 1896 and 1991. *British Medical Journal* 321 (7276): 1547–51.

Dyos, H., D. Cannadine, and D. Reeder. 1982. *Exploring the urban past: Essays in urban history.* New York: Cambridge University Press.

Fotheringham, A. S. 1997. Trends in quantitative methods I: Stressing the local. *Progress in Human Geography* 21:88–96.

Fotheringham, A. S., and D. Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23:1025–44.

Fraley, C., and A. E. Raftery. 1998. How many clusters? Which clustering methods? Answers via model-based cluster analysis. *Computer Journal* 41:578–88.

Frank, L. D., P. Engelke, and T. L. Schmid. 2004. Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventive Medicine* 27 (2): 87–96.

Galster, G. 2001. On the nature of neighbourhood. *Urban Studies* 38 (12): 2111–24.

Getis, A., and J. Franklin. 1987. Second-order neighborhood analysis of mapped point patterns. *Ecology* 68:473–77.

Golledge, R. G. 2002. The nature of geographic knowledge. *Annals of the Association of American Geographers* 92 (1): 1–14.

Gregory, I. N. 2008. Different places, different stories: Infant mortality decline in England & Wales, 1851–1911. *Annals of the Association of American Geographers* 98 (4): 773–94.

Gregory, I. N., and R. G. Healey. 2007. Historical GIS: Structuring, mapping and analysing geographies of the past. *Progress in Human Geography* 31:638–53.

Guo, D. 2008. Regionalization with dynamically constrained agglomerative clustering and partitioning (REDCAP). *International Journal of Geographical Information Science* 22 (7): 801–23.

Hagerstrand, T. 1982. Diorama, path and project. *Tijdschrift voor Economische en Sociale Geografie* 73 (6): 323–39.

Hall, P. G. 1996. *Cities of tomorrow: An intellectual history of urban planning and design in the twentieth century.* Malden, MA: Wiley-Blackwell.

Harris, R., P. Sleight, and R. Webber. 2005. *Geodemographics, GIS, and neighborhood targeting: Mastering GIS: Technology, applications, and management.* Chichester, UK: Wiley.

Heikkila, E., and Y. Wang. 2010. Exploring the dual dichotomy within urban geography: An application of fuzzy urban sets. *Urban Geography* 31 (3): 406–21.

Hershberg, T. 1981. *Philadelphia: Work, space, family, and group experience in the nineteenth century: Essays toward an interdisciplinary history of the city.* New York: Oxford University Press.

Hipp, J. R. 2007. Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review* 72:659–80.

Hunter, A. 1974. *Symbolic communities.* Chicago: University of Chicago Press.

King, G. 1997. *A solution to the ecological inference problem.* Princeton, NJ: Princeton University Press.

Kwan, M.-P. 1998. Space–time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis* 30 (3): 191–216.

———. 2009. From place-based to people-based exposure measures. *Social Science & Medicine* 69:1311–13.

Laube, P., S. Imfeld, and R. Weibel. 2005. Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science* 19 (6): 639–68.

Lebel, A., R. Pampalon, and P. Villeneuve. 2007. A multiperspective approach for defining neighbourhood units in the context of a study on health inequalities in the Quebec city region. *International Journal of Health Geographics* 6 (1): 27.

Lee, B. A., S. F. Reardon, G. Firebaugh, C. R. Farrell, S. A. Matthews, and D. O'Sullivan. 2008. Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review* 73 (5): 766–91.

Logan, J. R., J. Jindrich, H. Shin, and W. Zhang. 2011. Mapping America in 1880: The urban transition historical GIS project. *Historical Methods* 44 (1): 49–60.

Longley, P. A., and C. Tobon. 2004. Spatial dependence and heterogeneity in patterns of hardship: An intra-urban analysis. *Annals of the Association of American Geographers* 94 (3): 503–19.

Lynch, K. 1960. *The image of the city.* Cambridge, MA: MIT Press.

Matthews, S. A. 2011. Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods. In *Communities, neighborhoods, and health: Expanding the boundaries of place,* ed. L. Burton, S. Kemp, M. Leung, S. A. Matthews, and D. Takeuchi, 35–55. New York: Springer.

Miller, H. J. 1991. Modeling accessibility using space–time prism concepts within geographical information systems. *International Journal of Geographical Information Systems* 5:287–301.

———. 2005. A measurement theory for time geography. *Geographical Analysis* 37:17–45.

———. 2007. Place-based versus people-based geographic information science. *Geography Compass* 1:503–35.

Montello, D. R. 2003. Regions in geography: Process and content. In *Foundations in geographic information science,* ed. M. Duckham, M. F. Goodchild, and M. F. Worboys, 173–89. London and New York: Taylor & Francis.

Moore, D. D. 1994. Class and ethnicity in the creation of New York City neighborhoods: 1900–1930. In *Budapest and New York: Studies in metropolitan transformation, 1870–1930,* ed. T. Bender and C. E. Schorske, 139–60. New York: Russell Sage Foundation.

Openshaw, S. 1996. Developing GIS-relevant zone-based spatial analysis methods. In *Spatial analysis: Modeling in a GIS environment,* ed. P. Longley and M. Batty, 55–74. Cambridge, UK: GeoInformation International.

Openshaw, S., and P. Taylor. 1979. A million or so correlated coefficients. In *Statistical applications in the spatial sciences,* ed. N. Wrigley, 127–44. London: Pion.

Park, R., E. W. Burgess, and R. D. McKenzie. 1925. *The city.* Chicago: University of Chicago Press.

Quinn, J. 1940. The Burgess zonal hypothesis and its critics. *American Sociological Review* 5 (2): 210–18.

Reardon, S., C. Farrell, S. Matthews, D. O'Sullivan, K. Bischoff, and G. Firebaugh. 2009. Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research* 38 (1): 55–70.

Reardon, S., S. Matthews, D. O'Sullivan, B. Lee, G. Firebaugh, C. Farrell, and K. Bischoff. 2008. The geographic scale of metropolitan racial segregation. *Demography* 45 (3): 489–514.

Reardon, S., and D. O'Sullivan. 2004. Measures of spatial segregation. *Sociological Methodology* 34 (1): 121–62.

Sampson, R., S. Raudenbush, and F. Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 277:918–24.

Shaw, C. R., and H. D. McKay. 1942. *Juvenile delinquency and urban areas.* Chicago: University of Chicago Press.

Shevky, E., and W. Bell. 1961. *Social area analysis: Studies in human ecology.* Berkeley, CA: University of California Press.

Sjoberg, G. 1965. *The preindustrial city, past and present.* New York: The Free Press.

Sobek, M. 1996. Work, status, and income: Men in the American occupational structure since the late nineteenth century. *Social Science History* 20 (2): 169–207.

Tobler, W. R. 1989. Frame independent spatial analysis. In *The accuracy of spatial databases*, ed. M. Goodchild and S. Gopal, 115–22. London: Taylor and Francis.

Warner, S. B. 1962. *Streetcar suburbs: The process of growth in Boston 1870–1900.* Cambridge, MA: Harvard University Press.

Weber, J., and M. Kwan. 2003. Evaluating the effects of geographic contexts on individual accessibility: A multilevel approach 1. *Urban Geography* 24 (8): 647–71.

Wellman, B. 2001. Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research* 25:227–52.

Witten, I., and E. Frank. 2005. *Data mining: Practical machine learning tools and techniques.* 2nd ed. San Francisco: Morgan Kaufmann.

Zunz, O. 1982. *The changing face of inequality: Urbanization, industrial development, and immigrants in Detroit, 1880–1920.* Chicago: University of Chicago Press.

Correspondence: Department of Geography, University of Colorado at Boulder, 260 UCB/110 Guggenheim Hall, Boulder, CO 80309, e-mail: seth.spielman@colorado.edu (Spielman); Department of Sociology/S4, Brown University, Maxcy Hall, Providence, RI 02912, e-mail: John_Logan@brown.edu (Logan).