Routledge
Taylor & Francis Group

Mapping America in 1880
The Urban Transition Historical GIS Project

JOHN R. LOGAN
JASON JINDRICH
HYOUNGJIN SHIN
WEIWEI ZHANG
Department of Sociology
Brown University

Abstract. The Urban Transition Historical GIS [Geographic Information Systems] Project is a new data resource for U.S. counties and cities that takes advantage of the North Atlantic Population Project's 100-percent digital transcription of records from the 1880 U.S. census. The project has developed several additional resources to make possible analysis of social patterns at the level of individuals and households while taking into account information about their communities. One key contribution is the creation of historically accurate GIS maps showing the boundaries of enumeration districts in 39 major cities. These materials are now publicly available through a Web-based mapping system. Addresses of all households in these cities are also being geocoded, a step that will enable spatial analyses of residential patterns at any geographic scale. Preliminary analyses demonstrate the utility of multiple scales and the ability to combine information about individuals with data about their neighborhoods.

Keywords: cities, historical GIS, immigration, urban history

The Urban Transition Historical Geographical Information System Project (HGIS) is a new data resource on U.S. counties and cities at the end of the nineteenth century, when the nation was in transition from its rural origins toward a predominantly urban and industrial century. The project takes advantage of the 100-percent digital transcription of records from the 1880 U.S. Census that was organized by the Church of Latter Day Saints and prepared for scholarly use by the Minnesota Population Center (MPC). This file includes approximately 50 million Americans, organized by household and with information on residents' name, age, race, gender, relation to head of household, state or country of birth, each parent's state or country of birth, and occupation, and the enumeration district, ward, city, county, and state of residence. The Urban Transition HGIS has created several additional resources that facilitate analyses of urban structure and population geography. First, the data for

individuals and households have been aggregated into summary files for enumeration districts and counties throughout the country. Second, county data have been joined with accurate 1880 maps created by MPC's National HGIS (NGHIS) Project, and the Urban Transition HGIS has developed maps of enumeration districts in 39 major cities that can display these data at a neighborhood level. Third, these maps have been made publicly available on a fully featured Web-based mapping system (http://www.s4.brown.edu/utp) that requires no GIS training for users. For those users with more technical skills, all the data and the boundary files on which they are based can be downloaded. Fourth, street addresses have been added to the file for households in these cities (comprising about 5 million residents). Geocoding of these addresses is now in progress, making it possible to analyze urban spatial patterns at any scale, with no need to use administrative boundaries imposed by the census.

The intention is to promote quantitative historical research on social relations that goes beyond studies of samples of individuals to also take into account their embeddedness in surrounding communities. The year 1880 is a significant moment for such studies. The Western frontier would be declared closed only a decade later, while the transformation of the country from an agricultural economy to an industrial economy was gaining momentum, and African Americans were newly released from slavery. The year 1880 is generally considered the year that ends the dominance of farmers—or more precisely, the gainfully employed persons in agriculture—in the American labor force. It does not mean the predominance of the urban population over the rural population, which would take another 40 years (Nugent 1981, Anderson 1988). But this is the time when the United States was clearly poised to become a nation of cities. Midwestern cities like Chicago, already large, were almost new in 1880. Older cities like New York were still receiving heavy waves of German and Irish immigrants, and the nation was on the verge of the next great wave of population movements—especially

Address correspondence to John R. Logan, Department of Sociology, Brown University, Box 1916, Maxcy Hall, 112 George St., Providence, RI 02912, USA. E-mail: john_logan@brown.edu

that of blacks from South to North and immigrants from Southern and Eastern Europe. The successes and failures of cities in integrating newcomers into community and economic life in this era set the stage for basic processes of assimilation and segregation that have been more widely studied in the twentieth century. Unlimited access to the full 1880 census and especially the ability to link data on individuals with aggregate information on the places where people lived, at various levels of geography create new opportunities for research on this key period in American history. Our purpose here is to point to some of the substantive questions that these data are well suited to address and to provide enough documentation of how the files have been developed so that researchers will be prepared to use them knowledgeably.

Research Questions about Communities

A key urban question concerns how and when various white ethnic groups assimilated with native-born whites, while a color line limited the social mobility of African Americans. There has been considerable historical research on boundaries in the labor market. For example, Stanley Lieberson (1980, 328) noted "broad occupational similarities between [South, Central, and Eastern] Europeans and blacks in 1900." Stephan Thernstrom (1973) found that blacks' position was comparable to that of the Irish—at the bottom of the queue in Boston—in the mid and late 19th century, when less than 2 percent of Boston's population was black. But it is well known that white ethnics assimilated over generations in the occupational structure (Roediger 2005), whereas African Americans did not.

Social scientists have also paid attention to neighborhood patterns. Nathan Kantrowitz (1979) analyzed ward data for Boston extending back to the mid–nineteenth century. He calculated average levels of segregation among white ethnic groups, distinguishing older (United Kingdom, Ireland, Norway, Sweden, and Germany) and newer (Russia, Italy, Austria, Hungary, and Czechoslovakia) arrivals. Segregation measured by the Index of Dissimilarity (D) by wards between all pairs of white groups averaged 45 in 1880 and was virtually unchanged at 43 in 1920. These trends suggest that ethnicity was only a moderate force in organizing neighborhoods in 1880, and its importance had not changed 40 years later. The same appears to be true of Philadelphia. As Stephanie Greenberg (1981, 206) observes in Philadelphia, immigrant ethnic ghettoes such as she found by 1930 for Russians, Italians, Poles, and blacks had not developed in the 1800s for earlier immigrant groups. Stanley Nadel (1990) reports similar findings for New York City, though he noted greater residential separation of Germans than of Irish.

A limitation of most studies of residential segregation is that indices are based on frequency distributions of a single variable (e.g., race or occupation) within census tracts or other geographic units. But often, researchers on segregation are interested in the relationship between race or ethnicity and social class. One source of ethnic segregation is segrega-

tion in the workplace, and, as Greenberg (1981) points out, workers in a particular industry in 1880 tended to live within a mile around the main job locations for that industry. This may have been particularly relevant for Germans. Deborah Dash Moore (1994, 145) argued that Germans in New York settled near industries established by German entrepreneurs. They "constructed a fairly complete ethnic 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. In contrast, the Irish "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." Addressing these questions requires the analysis and comparison of multiple aspects of neighborhood composition, at levels of detail that are often not available to the researcher.

Olivier Zunz (1982) did compare neighborhood composition for persons living in a sample of city blocks in Detroit. He found increasing levels of segregation from 1880 to 1920, measured in the following way: there was significant clustering (or overrepresentation of a particular group) in 30 percent of the sampled blocks in 1880, and this grew to 75 percent in 1920. By 1920, he found, Poles, Hungarians, Jews, and blacks "reached record levels of concentration in some blocks" (1982, 341). During this same period, however, Irish concentrations disappeared, and German clusters became more scattered. At the same time, Zunz found an increase in socioeconomic clustering, in the sense that a particular occupational group was overrepresented in the block. In 1880, it was common for ethnic blocks to include a wide range of occupations, from laborers to professionals and shopkeepers. By 1920, however, many whole city blocks comprised entirely people with the same ethnicity and occupation, like Polish factory workers, native white factory workers, or native white office workers. Zunz argues that the changing balance of residential patterns (from neighborhoods based mainly on ethnicity to neighborhoods segregated by both ethnicity and occupation) represented a relative "decline of ethnicity" and a "more uniform social contract ordering people primarily by social status" (1982, 401). What had declined, more precisely, were German and Irish settlements, and—except for those of Jews and blacks—mixed-class neighborhoods of a single group.

Other theoretically significant questions are whether the second-generation group members tended to move out from the original ethnic neighborhoods and what the implications of that movement for spatial assimilation are. Greenberg (1981) computed segregation indices separately for Philadelphia's first- and second-generation Germans and Irish. In both cases, segregation was lower in the second generation, but by a modest margin. In the most common version of assimilation theory, this difference should represent the concentration of immigrants in low-rent housing in the initial settlement areas. But Thomas Kessner (1977, 157), studying

New York with data from both the census and city directories, concluded that newer outlying areas included many new immigrants at the lowest occupational levels rather than a zone of "Americanization." In fact, Moore (1981, 12) reported that when second-generation Jews in New York in the 1920s left large immigrant neighborhoods behind, they became even more residentially segregated in dispersed smaller communities. They "chose to create and re-create new ethnic neighborhoods, constantly spreading clusters of Jewish settlements throughout the city while maintaining a high level of segregation."

An unusual advantage of the 1880 census data is that researchers now have complete control over the tabulations that they wish to analyze, and there is no need to limit research to a sample of blocks or to pretabulated data in printed sources. It is now possible to conduct comparative historical studies for multiple cities (including all 98 cities in the country that were separately identified in the 1880 census), not only a single city, and to test models that may account for differences between cities and between groups. It is also possible to link data for individuals with information about their neighborhoods (e.g., who lives in what kinds of places) and to invert these questions to ask how neighborhood context affects other outcomes (e.g., are people who live in an ethnic neighborhood more likely to work in the ethnic economy?). Finally, using HGIS methods, it is possible to introduce explicitly spatial analyses within the larger cities that make use of information about neighborhoods' location in relation to one another, and even to study individual households and their next-door neighbors.

Census Data at the Individual and Community Levels

Census data have become an invaluable and accessible resource for historical research on the population of the United States (Baker 2003; Holdsworth 2003; Sies 2001). These data generally are presented in two forms. *Microdata* are transcribed from individuals' records in the census manuscripts, and they provide information about households and persons within households. The Minnesota Population Center's IPUMS (Integrated Public Use Microdata Samples) project has become the standard source for such samples, and they are available as early as 1790. *Summary data* comprise tables showing the frequency distribution or a crosstabulation of individual-level information, aggregated to a given unit of geography, such as a census tract or a city. *GIS systems* offer additional information about where these geographic units are in space, and historical GIS efforts have become widespread in recent years (Knowles, 2000; Gregory and Healey 2007; Knowles and Hillier 2008).

Catherine Fitch and Steven Ruggles (2003) argue that the lack of usable historical census geography across multiple cities has limited the scope of research into basic issues in the social sciences. But now MPC's NHGIS project provides census tract data for the decades during the period 1940–2000 (and as early as 1910 for some cities),

based on tabulations originally prepared by the U.S. Bureau of the Census. Data for counties are provided for earlier years. NHGIS offers these data in conjunction with standardized boundary files unique to each decade that can be used in conjunction with GIS software. A companion project, Social Explorer, serves these maps through a browser, though some data are available only by subscription. Other comparable historical GIS projects include the Canadian Century Research Infrastructure (Gaffield 2007; http://www.canada.uottawa.ca/ccri), Great Britain HGIS (Gregory 2002; http://www.port.ac.uk/research/gbhgis), Belgian HGIS (De Moor and Wiedemann 2001; http://www.hisgis.be/start_en.htm), and the China HGIS (Bol 2007; http://www.fas.harvard.edu/~chgis).

In the United States, the Urban Transition HGIS pushes the time horizon for historical spatial analysis back to 1880. When completed, it will provide these additional layers of spatial data to augment the current NAPP data set and NHGIS boundary files: (1) contextual variables at the level of counties and enumeration districts that can be added to individual person and household records for all persons across the nation, (2) accurate GIS maps of enumeration districts in 39 cities, and (3) geocoded coordinates of residences in these 39 cities that allow the creation of local area units along any criteria that scholars may require for spatial analysis.

The prepared summary files for counties and enumeration districts include the numbers of persons in various categories, as well as computed variables: the number of households and total population, average occupational standing of residents (their socioeconomic index [SEI]), percentage of households comprising a married couple with a child, percentage of persons who are Yankees (native whites of native parents), and first or second generation Canadians, British, Irish, German, Swedish, Norwegian, Danish, French, Chinese, black, mulatto, or Native American individuals. It also separately counts second-generation ethnic group members. Age variables include the percentage who are 16 years or younger and 60 years or older. The sex ratio is calculated for persons aged 18–44 years. The percentage currently married among persons age 18 years or older is also calculated. Additional employment variables are the percentages of men and women (combined and separately) aged 15–64 years who have a recorded occupation.

Sources for Mapping in 1880

One advantage for researchers of the twentieth century is the relative stability of county boundaries. In the nineteenth century, boundaries were especially fluid along the frontier, but changes occurred everywhere. This is evident even at the national level as new counties were established, established places changed names, and existing boundaries adjusted to meet shifting populations. The definitive source for information about historical county maps is the Atlas Project of the Newberry Library (http://www.newberry.org/ahcbp). The Atlas Project documents and maps every legal change

to county boundaries from the seventeenth century to the present for every state. These maps can be viewed online, and they can also be downloaded as shape files.

The Urban Transition HGIS uses shape files made available through the NHGIS for county-level data. It combines the geographic information with a new aggregation of population data that are summed from the individual records of the 1880 census disseminated by NAPP, making possible calculation of a much wider array of variables than previously available. There is not, however, full coverage of the nation. In 1880 most of the area of the current lower 48 states was included in the census, the exceptions being the unorganized territories of the Dakotas and Indian Territory. Special enumerations by reservation or county equivalents were not tabulated or transcribed with the rest of the census, and these are therefore excluded from our data files. Some data are also missing for counties in established states. Because our population totals agree with those published for this census, we assume that the Census Bureau did not recognize those counties and that the residents have been included in the returns for neighboring counties. So while there is a loss of geographic specificity, the full population is accounted for.[1]

Creating an accurate, functional small-area urban HGIS for 1880 posed greater challenges, and it was not possible to complete this task for every city identified by the Census Bureau. A sample of 39 major cities was selected for detailed mapping. As shown in table 1, these include the 26 largest cities in the country, from New York (which at that time included only Manhattan and part of the Bronx) with 1.2 million residents to New Haven with fewer than 63,000. Additional cities were chosen mainly to extend the geographic range of the project to include smaller cities in the Midwest (e.g., Kansas City and Minneapolis), in the South (e.g., Charleston and Atlanta), and in the West (e.g., Denver and Oakland).

The contemporary census builds geographic data from blocks, block groups, and census tracts. In the nineteenth century, the smallest unit for which data can be readily tabulated is the enumeration district (ED), which in urban areas is comparable in population size to modern census tracts. The practical function of the ED in 1880 was to define the area within which a given enumerator was contracted to gather data. A written description of the boundaries has survived for much of the country, in the form of a listing of enumerators, the payment owed to them, and the area for which they were responsible. Unfortunately they often do not have very clear boundaries. District supervisors were chosen for their familiarity with their region, and when defining enumeration districts, they provided only enough information to guide the census taker, who was also a resident familiar with the area. The result is descriptions that occasionally include prominent, but idiosyncratic features. Examples include fences between houses (Chicago), obsolete political boundaries (St. Louis's former city limit, and ward boundaries), minor water features that have since moved underground (many loca-

tions), shorelines that have been radically changed through infill or dredging, alleys, and hill crests. One recurring problem is the use of an outdated or informal name for features that are crucial for defining the limits of a boundary. Another common problem is a boundary described as following an extension of a street until it reaches the city limits, which is problematic if that street ceased to exist or was extended in a different direction after 1880.

A greater problem is that the Census Office's records of ED boundaries in 1880 are incomplete. No records remain for Alabama, Arizona Territory, Arkansas, California, Colorado, Connecticut, Montana Territory, Ohio, Oregon, Pennsylvania, or Wisconsin. In many other cases (such as New Jersey), EDs in major cities are described as portions of precincts (e.g., precinct 5 north of Main Street) that are undefined in the descriptions. Table 1 lists the 23 cities for which ED boundaries were available either as detailed street descriptions or by use of ward or precinct boundaries that could be identified from other sources. In other cases, as subsequently described, ED boundaries were inferred from address information in the original microdata, where each person's ED was also listed.

The Urban Transition HGIS provides maps of EDs in two forms. The first is referenced to a historical street map from the period, and this is the form employed in our Web-based HGIS. These ED boundaries have been drawn along the streets that are shown in the historical map as an annotation. This version is designed to facilitate visualization, because it allows the user to see the city as it was represented by a nineteenth-century cartographer, with street names and other features clearly labeled, as in the illustration of district 61 in St. Louis in figure 1.

Although we have modified the map through georeferencing, the historical map image is not an accurate projection. To be of use for systematic analysis, maps need to have accurate geographic coordinates. The second form of the boundary file is based on contemporary GIS maps of the cities, using the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing System (http://www.census.gov/geo/www/tiger/) and the Environmental Systems Research Institute–adapted version of those same files (http://www.esri.com/data/download/census200 0-tigerline/index.html). These files were edited by hand to correspond with the street layout of cities in 1880. Doing this involved deleting streets or highways constructed after 1880, adding those that were demolished, changing names of others, and in some cases correcting the alignment of streets. (Similarly edited TIGER shape files are also the basis for county boundaries made available by the NHGIS.) Household addresses that are geocoded to this map align as closely as possible with the current TIGER standard and locations. These maps are available for download and are intended for detailed spatial analyses. However, there is always something lost in translation. For example, we have not been able to adjust the land area of cities to take into account landfill or rechanneling of waterways over the last 130 years.

TABLE 1. Cities in Study Sample by Population Size and Rank in 1880

Rank	City	Population	Enumeration district descriptions	Address ranges
1	New York, NY	1,206,299	Street boundaries	Provided in full
2	Philadelphia, PA	847,170	Not available	Provided in full
3	Brooklyn, NY	566,663	Known ward/precinct boundaries	Provided in full
4	Chicago, IL	503,185	Street boundaries	Provided in full
5	Boston, MA	362,839	Street boundaries	Provided in full
6	St Louis, MO	350,518	Street boundaries	Provided in full
7	Baltimore, MD	332,313	Street boundaries	Provided in full
8	Cincinnati, OH	255,139	Not available	Provided for major streets
9	San Francisco, CA	233,959	Not available	Provided in full
10	New Orleans, LA	216,090	Street boundaries	Provided for major streets
11	Cleveland, OH	160,146	Not available	Provided in full
12	Pittsburgh, PA	156,389	Not available	Only street names listed
13	Buffalo, NY	155,134	Street boundaries	Provided in full
14	Washington, DC	147,293	Street boundaries	Provided in full
15	Newark, NJ	136,508	Not available	Provided in full
16	Louisville, KY	123,758	Known ward/precinct boundaries	Only street names listed
17	Jersey City, NJ	120,722	Not available	Provided in full
18	Detroit, MI	116,340	Street boundaries	Only street names listed
19	Milwaukee, WI	115,587	Not available	Provided for major streets
20	Providence, RI	104,857	Street boundaries	Provided in full
21	Albany, NY	90,758	Street boundaries	Provided in full
22	Rochester, NY	89,366	Street boundaries	Provided in full
23	Allegheny, PA	78,682	Not available	Only street names listed
24	Indianapolis, IN	75,056	Street boundaries	Provided in full
25	Richmond, VA	63,600	Not available	Provided in full
26	New Haven, CT	62,882	Not available	Provided in full
30	Kansas City, MO	55,785	Street boundaries	Only street names listed
33	Columbus, OH	51,647	Not available	Only street names listed
36	Charleston, SC	49,984	Known ward/precinct boundaries	Provided for major streets
38	Minneapolis, MN	46,887	Street boundaries	Only street names listed
40	Nashville, TN	43,350	Street boundaries	Only street names listed
43	Hartford, CT	42,015	Not available	Provided in full
45	St Paul, MN	41,473	Street boundaries	Provided for major streets
49	Atlanta, GA	37,409	Street boundaries	Provided for major streets
50	Denver, CO	35,629	Not available	Only street names listed
51	Oakland, CA	34,555	Not available	Only street names listed
54	Memphis, TN	33,592	Known ward/precinct boundaries	Provided for major streets
63	Omaha, NE	30,518	Street boundaries	Only street names listed
68	Mobile, AL	29,132	Not available	Only street names listed

The process of geocoding individual households has also begun, placing them in their approximate locations along the street grid. Grant support from the National Institutes of Health and the National Science Foundation was used to contract with the MPC to transcribe street addresses for residents of the 39 large U.S. cities. Address ranges for city blocks have changed from 1880, in many cases quite radically, but contemporary address ranges offer a first approximation of locations.

Many online sources guided our editing of the contemporary GIS maps to fit the historical city street grids. Because EDs typically are subdivisions of city wards, one key source is ward maps. The U.S. Library of Congress compiled a volume identifying *Ward Maps of United State: A Selective Checklist of Pre-1900 Maps in the Library of Congress* (Shelley 1975), and these maps are available on microfiche through the Cornell University Library Map Collection. The David Ramsey Map Collection (http://www.davidrumsey.com/) has an invaluable set of high quality digital historical map images of U.S. cities. The Sanborn Fire Insurance maps (http://sanborn.umi.com/) offer more detail than most other sources and proved particularly useful when additional information was needed to locate residential alleys or other features of densely developed downtown districts for some cities.

Annual city directories published during the period are another major source, often including detailed street map and boundaries of wards and electoral precincts. The most
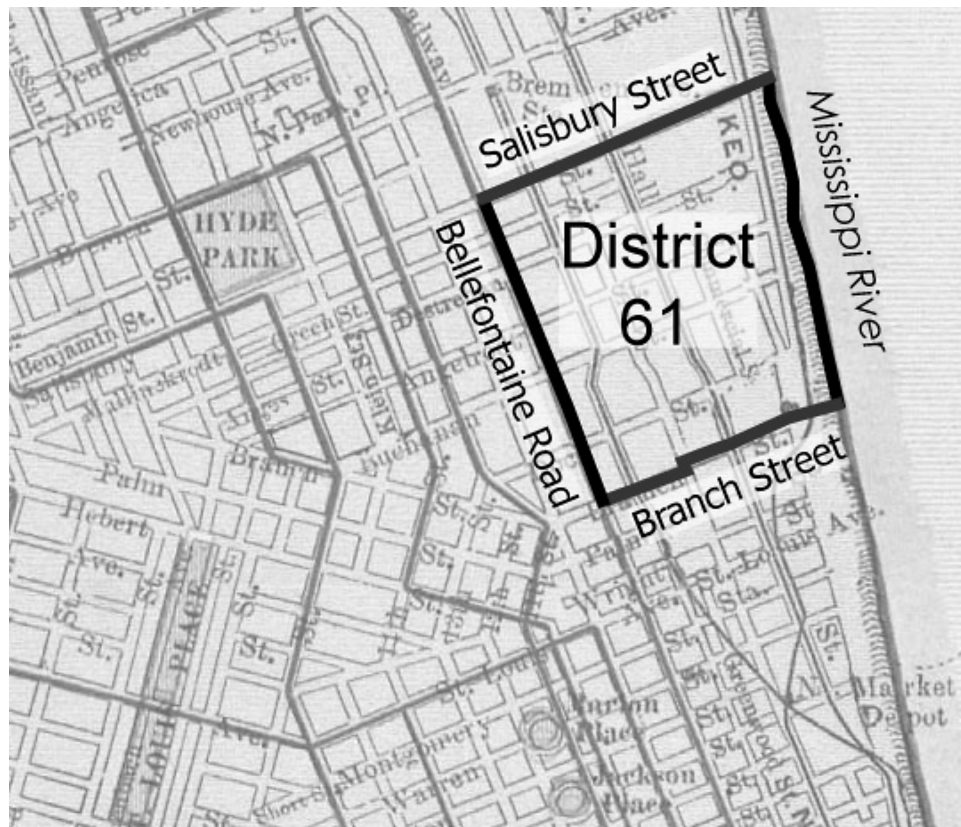
FIGURE 1. Outlines of enumeration district 61 in St. Louis drawn onto a historical street map.

useful information, however, is the comprehensive list of streets usually found in the back of these volumes, a list that often includes the address ranges between intersections. A guide to the volumes for each city used in this project is *City Directories of the United States, pre-1860 through 1901: A Guide to the Microfilm Collection* (Research Publications 1984). Many local sources, including libraries and local history associations, were used to trace street name changes over time in several cities, neighborhood maps, building names, and other historical information.
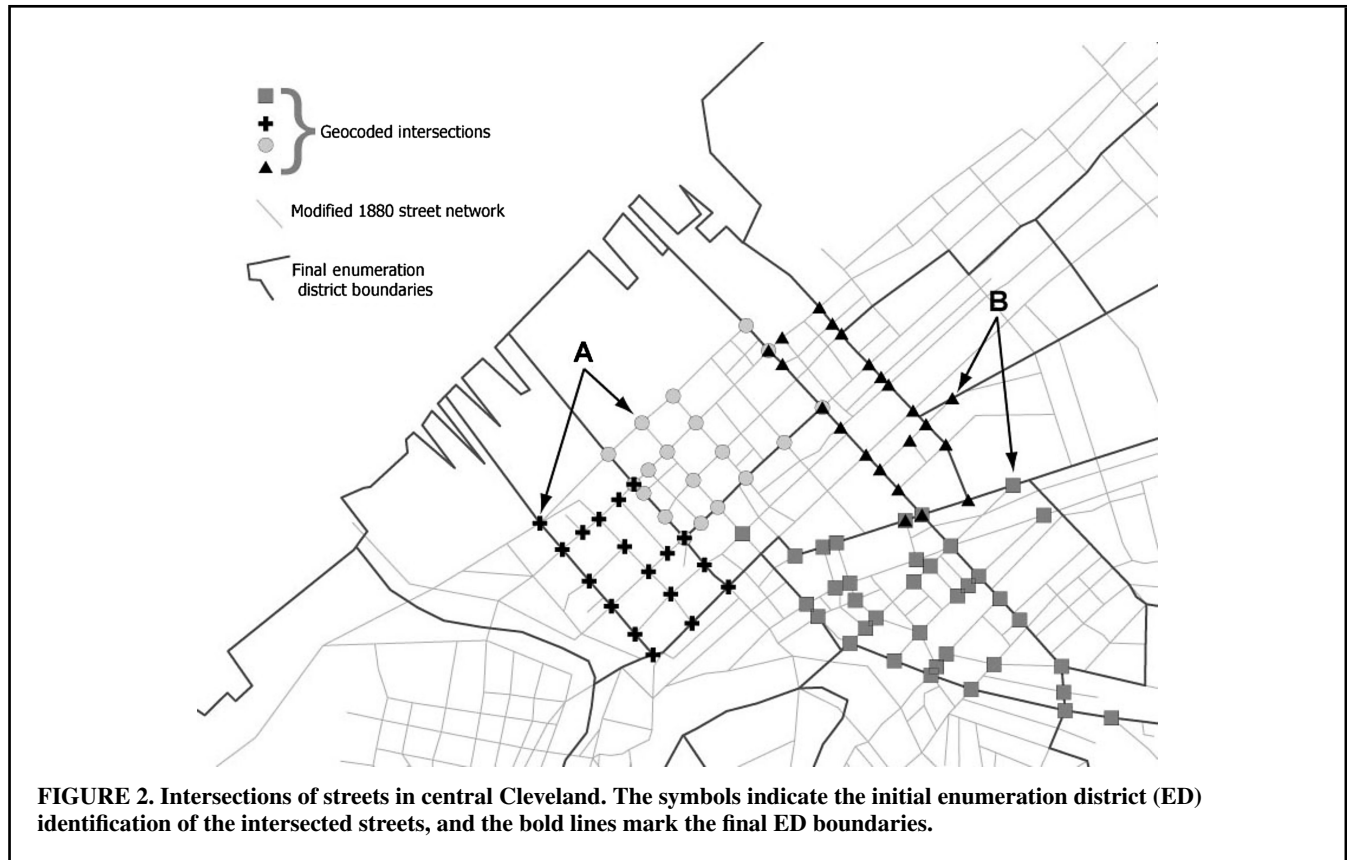
Mapping EDs

A first step in building maps of enumeration districts was to establish the 1880 limits of our 39 cities. This was easier in some locations than others. The simplest are those cities occupying their entire home county (e.g., Philadelphia, New Orleans, New York City, Baltimore, San Francisco, and St. Louis) and cities that have not annexed large areas since the 1880 census (Buffalo; Providence; Boston; Albany; Washington, DC; and Hartford). The majority of our cities occupy some portion of a larger county that is radically different from that of 130 years ago. Our task was to accurately

identify this area and reproduce it in the historical GIS. In these cases, the city limits were approximated by using a carefully georeferenced historical map, and when we were drafting the TIGER-referenced shape files, by taking guidance from modern details such as the location of streets and rivers.

In the case of two cities—Philadelphia and Louisville—we relied on existing maps of enumeration districts created by local institutions. The next simplest case is when complete ED boundaries are available from the historical census record. When available, the descriptions of most enumeration districts are short, simple lists of the streets bounding the districts. (Also common were references to political units, like precincts or wards, in the descriptions of EDs. When the boundaries of such units were unavailable, such descriptions were of less use for our purpose.)

Additional efforts were required to infer ED boundaries for the 16 cities whose descriptions were not preserved by the Census Bureau. These are listed in table 1. Creating EDs in cities with missing descriptions involved an inductive process similar to a logic puzzle, where multiple sources of incomplete information are brought together to suggest a solution. We used three different kinds of information relating to ED boundaries for this purpose:

FIGURE 2. Intersections of streets in central Cleveland. The symbols indicate the initial enumeration district (ED) identification of the intersected streets, and the bold lines mark the final ED boundaries.

1. Minor civil divisions. The Census Office instructed district supervisors to make efforts to confine EDs within wards and other minor political units. As a result, these political boundaries feature prominently in ED descriptions. Where there was documentation for these wards or precincts, their creation was our first step. In practice we found examples of ED boundaries crossing one or more minor political boundaries within a city, but these were always relatively few.

2. Intersecting streets. We extracted street names from the transcription of addresses created by the MPC matched to the microdata file from NAPP. Using this information, we determined the names of streets within any given ED, and by pairing those names, we arrived at a set of all possible intersections. By geocoding and marking these intersections with their ED number we were able to further estimate and refine our boundaries. This procedure is illustrated by a result from Cleveland in figure 2. The symbols (square, plus, circle, and triangle) represent intersections located in different EDs as defined by this method, while the bold lines indicate our final ED boundaries; note where errors occur. Some of these intersections (such as those indicated as type A on the map) are in a zone where the regular gridiron of streets and their densities lends itself to an accurate bounding of the EDs. Intersections marked as

type B on the map are in a zone where the streets are not as regular and where more information was needed to establish ED boundaries (note that a triangle from the initial coding is misplaced into an area where all other points are square).

3. Geocoded household addresses. Minor civil boundaries, intersections, and in some cases the official ED descriptions provide approximations of ED boundaries. However the final and definitive guide to ED boundaries is geocoded addresses of residents whose ED location was coded by NAPP. These addresses create a much more detailed point pattern than does the map of street intersections. *Geocoding* in the sense used here refers to the automated procedure of determining a point location for an address using GIS software. This requires two data files. One is a standardized list of addresses. This we adapted from the transcribed address file provided by the MPC. The second file is the street shape file identifying the location and address range found between the endpoints of every street segment.

Location of street segments was resolved through our detailed editing of contemporary TIGER files, but address ranges for each segment were more difficult to establish.

By 1880, most major cities had adopted some form of uniform addressing, but that may or not be the same as in

use today (Rose-Redwood 2007). Addressing systems that were more or less uniform across the city were in place in the largest U.S. cities by the 1870s. Nevertheless, there were sometimes remnants of a previous, irrational, addressing scheme, or a system that was in transition from an older system to a newer one. Among the largest cities, aggressive annexations doubling or tripling the area of a city occurred throughout the nineteenth century. Consolidations of cities with their smaller neighbors often resulted in redundant street names in different parts of a city (e.g., more than one Main Street with similar house numbers). There are many other variations. The three largest Ohio cities changed from names to numbers for the north-south streets and then reused these names for east-west routes in the developing districts at the urban fringe. St. Louis cobbled together the street networks of many smaller communities into a discontinuous grid. Some cities reformed their uniform address model after 1880, many for the very good reason that the older model was rational, but needlessly confusing (for example, Chicago had two Cartesian grids originating from widely separated points, and Milwaukee had three).

Fortunately, city directories are a reasonably complete source for historical street names, directions, address ranges, and intersecting streets. Table 1 lists the 20 cities whose directories provided detailed address ranges. In seven other cities, address ranges were listed only for major streets. Unfortunately, the city directories for the remaining cities offered no information on addressing Allegheny, Columbus, Denver, Detroit, Kansas City, Louisville, Minneapolis, Mobile, Nashville, Oakland, Omaha, and Pittsburgh. The system dominant in the Midwest and West (most of those with missing address range information) has house numbers assigned through reference to a central point in the city. This Cartesian (or Philadelphia) system typically has a single point of origin for all addresses demarcated by the crossing of two baseline streets, with breaks at a regular distance, most often the distance of one city block—hence the 100 block, the 200 block, and so on. In such cities, geocoding can be done with a high degree of accuracy even where historical city directories do not provide address ranges.

In the worst case, when no information about addresses was available, we determined the range of addresses of all residents along a street within the boundaries of an ED and changed the HGIS accordingly. Their actual location along the street is based on linear interpolation. In some cities it is hard to improve on this approach, because the addressing system gives a relative location in respect to other addresses on that street, but it does not indicate an absolute location. This is most likely in places where blocks are of irregular length, and streets tend to follow the contours of the landscape, as is often true on the outskirts of a city.

In practice, there are multiple sources of error for all cities, from the ambiguities in ED descriptions, to mistakes by enumerators, to errors in transcription. Our approach in every city was to make multiple iterations: estimate ED boundaries, map residents using geocoded addresses, then correct obvious boundary errors and redo the geocoding of residents. The ED boundaries mapped as a result of this process have a high degree of accuracy.

Applications of the Urban Transition HGIS

We have begun to take advantage of these new resources in studies of urban residential segregation. Here we present some illustrative results from research on Newark, NJ, which is one of the first major cities for which mapping of enumeration districts and geocoding of individual household locations were completed. Figure 3 provides a thematic map of the percentage of the population in each ED that comprised first- or second-generation Germans. About one in three residents of Newark was German in 1880, and evidently they were unequally distributed around the city. Neighborhoods in the northern part of the city, especially along the Passaic River waterfront, mostly were less than 15-percent German. Higher concentrations were found in neighborhoods south of the Passaic and especially to the west of the central business district. Comparing this map with similar maps for the other major population groups—Yankees, Irish, and British—is a step toward understanding the ethnic character of neighborhoods at this time.

We have also calculated standard measures of segregation between these groups, drawing on the summary files at the ED level (Logan and Zhang 2010). We use the Index of Dissimilarity (D), which varies between 0 (no segregation) and 1 (complete segregation). These are presented in the first column of table 2. Germans were substantially more segregated from Yankees than were the Irish or British. German segregation from Irish and British was at almost the same level. The values of D for Germans (in the range of .50 to .54) are similar to those registered between Hispanics and non-Hispanic whites at the current time. Values of D between the British and either Yankees or Irish (in the range of .20–.30) are similar to those found today between non-Hispanic whites of different ancestry and are surprisingly low even considering that these groups share a common native language.

One question that is often raised about historical segregation measures is whether they are calculated at the right geographic scale. Cities in 1880 were, for most of their residents, walking cities where people tended to live close to their jobs. This fact has given rise to speculation that ethnic groups would tend to be intermingled except perhaps at the scale of buildings or street segments. The Urban Transition HGIS makes it possible to explore segregation at any scale above the household (even between households within the same building). The second column of table 2 provides an alternate calculation for street segments in Newark (aggregating residents of a single street from one intersection to the next). At this scale, segregation between Yankees and Germans, Irish and Germans, and British and Germans is in the range
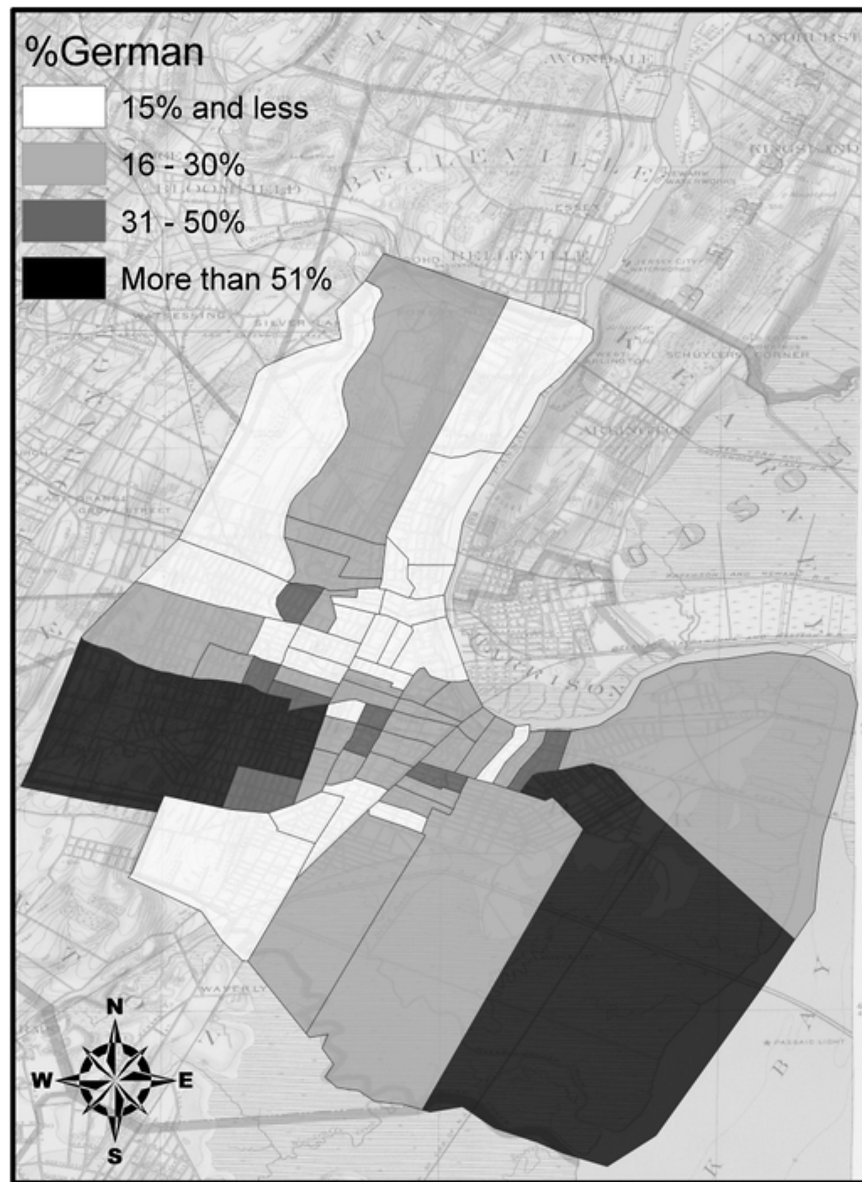
FIGURE 3. Thematic map of German population share in Newark, 1880, by enumeration district. *Source:* **John R. Logan and Weiwei Zhang, "White Ethnic Residential Segregation in Historical Perspective: U.S. Cities in 1880" (paper presented at the Annual Meeting of the Population Association of America, Dallas, TX, April 2010).**

of .63– .67. This is comparable with tract-based measures of segregation between African Americans and non-Hispanic whites in 2000, much higher than values previously reported in the literature on nineteenth-century cities. However, the relative rank ordering of segregation between pairs of ethnic groups is not much affected by the change in scale.

Having such detailed information about where people lived raises the possibility of combining information about individual residents with information about the people who lived around them to study the determinants of their location. An early version of such a study (White, Dymowski, and Wang

1994) used a public-use sample of data from 1910 and analyzed the ethnicity of people's next-door neighbors. As expected, people with higher status occupations and those who spoke English were more likely to have a "native white with native parents" neighbor; those in the immigrant generation were less likely, especially among Jews. Having enumeration district information allows researchers to study the characteristics of a larger neighborhood area. John R. Logan and Weiwei Zhang (2010) have analyzed the 1880 microdata for Germans, Irish, and British in conjunction with ED characteristics for 66 major cities. They show, for example, that

TABLE 2. Segregation (*D*) between Major White Ethnic Groups in Newark in 1880, by Enumeration District and Street Segment

Variable	Enumeration district	Street segment
Yankee and Irish	0.410	0.572
Yankee and German	0.537	0.666
Yankee and British	0.217	0.335
Irish and. German	0.529	0.650
Irish and British	0.307	0.454
German and British	0.503	0.637

Source: John R. Logan and Weiwei Zhang, "White Ethnic Residential Segregation in Historical Perspective: U.S. Cities in 1880" (paper presented at the Annual Meeting of the Population Association of America, Dallas, TX, April 2010).

being employed as a domestic servant and, alternatively, having an occupation of higher-than-average status were among the strongest predictors for an Irish person's living in an ED with a higher share of Yankees. Compared with immigrants, second-generation Irish lived in EDs with more Yankees. These findings are from multivariate analyses that included

controls for the relative size of Yankee and Irish populations in the city. One other city characteristic stands out for its importance: Irish in cities where they were more segregated from Yankees by occupation were likely to live in a significantly more Irish neighborhood. For example, two cities with very high occupational segregation between Irish and Yankees were Chicago and Boston, and these two cities also had among the highest levels of residential segregation. The ability to make comparisons across many cities, rather than limiting research to a single locale, draws attention to structural features of places that affect social boundaries between groups.

We offer one further illustration of how detailed geographic information can be exploited. Consider the wealth of information provided for a section around Bowery Street in Newark in figure 4. The historical street map is in the background. The map shows as circles the geocoded locations of every occupied building. Shading of the circles denotes the ethnicity of employed males in each building (Irish, Germans, or "others" in those cases where not all residents were either Irish or German). The numbers next to buildings identify the average occupational standing of employed male residents, based on the SEI that ranges from 0 to 100. Only the first digit of the value is provided here, so a 0 or
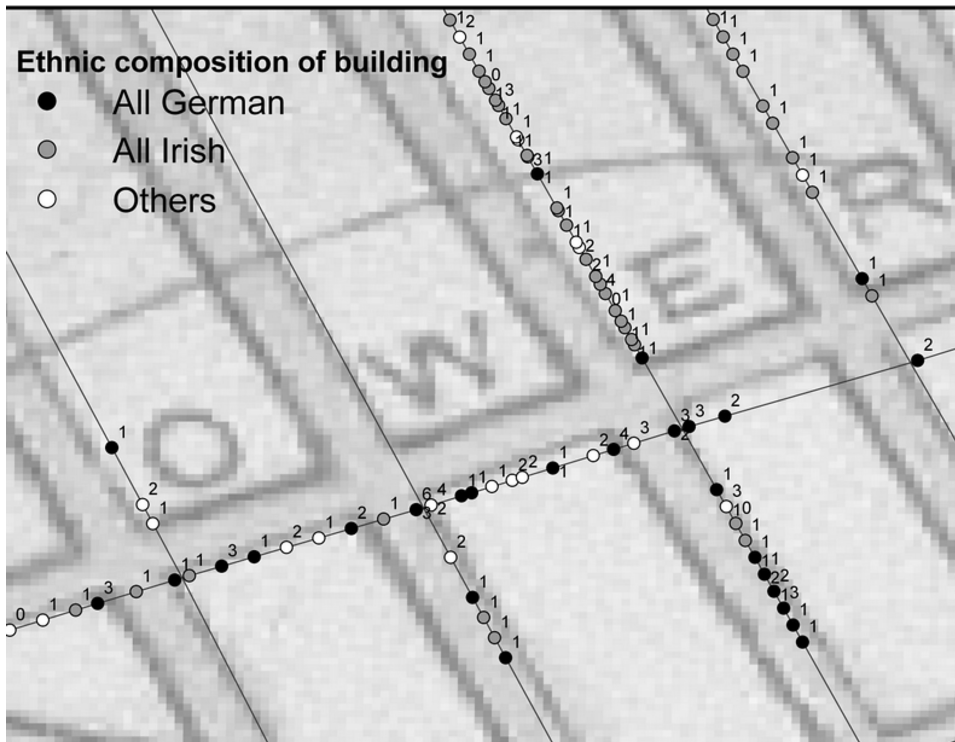
FIGURE 4. Section of Newark in 1880, showing locations of buildings, with race/ethnicity and average occupational socioeconomic index of adult male residents (one-digit code). *Source:* **John R. Logan and Weiwei Zhang, "White Ethnic Residential Segregation in Historical Perspective: U.S. Cities in 1880" (paper presented at the Annual Meeting of the Population Association of America, Dallas, TX, April 2010).**

1 represents an occupation of the lowest unskilled workers, and a value of 6–9 is mostly for professional occupations. Visual inspection suggests several patterns. This is clearly a working-class neighborhood, although there is some mixing especially along Bowery Street itself. The street leading north from Bowery is almost entirely Irish, with a larger share of Germans in the southern portion. Note that with this level of detailed information, it is possible to describe the residential setting of every individual in terms of nearest neighbors' ethnicity and occupation, or the whole street segment, or a larger area.

Logan and Hyoung-jin Shin (2010) have applied discrete choice models to tease out the contributions of race/ethnicity and class to this mapping of people. Their analysis of location refers to the approximately 1,500 residential street segments (the block along a single street bounded by intersections at each end) in Newark. They estimate separate models for more than 27,000 adult men—Irish, Germans, and British—in which the characteristics of their actual location are compared to characteristics of all other possible locations. Coefficients of the models for Irish men reveal, for example, that both the ethnic composition and the average class standing of neighbors have significant effects on the choice of location. On average, Irish were less likely to live in a street segment that had more residents of other groups or that had higher SEI. First-generation Irish were especially likely to live in street segments with more coethnics. But the main variation among the Irish came from their own occupational standing: the higher was their own SEI, the more likely they were to live in a location with more Yankees and with higher status neighbors of any ethnicity.

Conclusion

The Urban Transition HGIS is designed to serve many kinds of potential users. Scholars who are primarily interested in rural areas will find the nearly complete coverage of the county maps and the broad set of variables in the associated data especially useful. A great majority of the U.S. population in 1880 lived in relatively low-density rural areas, even if only about half the employed labor force was in agriculture. The county-level summary files include a standard set of population variables that are aggregated from the 1880 microdata. Large-scale geographic patterns can be readily visualized—where in the country first-generation immigrants were concentrated; what regions had more people of German, Irish, or other ethnic backgrounds; where African Americans (divided into subcategories of black and mulatto in 1880) were; where were the more elderly residents or children? An advantage of working with a Web-based map is that a scholar interested in any specific county or cluster of counties can easily find key population characteristics for the area of interest and download the data into a spreadsheet or other application. Representing data in a map also

makes it convenient to make comparisons with surrounding counties or with counties in other states. For scholars who are accustomed to working with large data files, the county-level data can be downloaded in a single file for the entire country.

An attraction of the 1880 NAPP files is that they allow research even on very small racial and ethnic groups (like Swedes or Chinese), whose representation is slender in the 1-percent or 5-percent sample files now available for other years. The Urban Transition HGIS provides limited information about many such groups, and working with these variables will help researchers decide which more specialized tabulations, if any, are required for their research. The original data are available from NAPP to be used for this purpose. For example, for any county or enumeration district in the country, it would be straightforward to count the number of immigrant Irish women under age 25 years who were working as domestics, or to evaluate the likelihood of intermarriage between people of British and Irish descent, or to calculate the volume of migration of people by state of birth to the newly expanding cities in the Midwest.

The Urban Transition HGIS adds a geographic dimension to such data. The additional spatial context can enrich research by making it more convenient to think about people in relation to those who are around them, and to study cities in terms of their diverse neighborhoods. For some urban scholars, the most obvious use of a GIS map is to place a neighborhood into a larger citywide setting. Urban historians typically work with a single city and become familiar with many specialized sources of information for that city. Our hope is that historians will supplement the data and GIS maps provided by this project with other place characteristics, like the location of churches, retail establishments, major employers, transit lines, election results, public health indicators, or significant historical events—all of which can be studied in relation to the demographic information available from the census. Because the boundary files can be downloaded, it is relatively simple to add new features to a city map.

As already mentioned, there is much potential for studies that encompass many cities and that seek to understand the differences and similarities among them—between North and South, older and newer regions, smaller and larger cities. Our sample of 39 cities, including all of the largest cities and some representation of every part of the country, offers a strong initial research design for such work. Because data are available for residents and for the EDs in which they lived, a natural approach is hierarchical linear modeling (HLM) in which contextual effects are estimated simultaneously with individual-level associations. A typical constraint on HLM models is whether there are enough cases in each higher-level grouping to represent the population, which makes the availability of the entire population a unique advantage of the 1880 data.

Much recent effort has been given to sophisticated methods of spatial analysis that use information about where

people live or events occur. For researchers who are familiar with these methods, accurate GIS maps of enumeration districts and eventually of individual household locations will be invaluable. For geographers, historical and otherwise, the geocoded data for individuals offer a unique opportunity to explore issues of scale. The novelty of a complete count census with accurate spatial attributes opens doors to questions of how neighborhoods are defined. At what distance is there no longer much relationship between people or any effect of surrounding neighborhoods? Working with contemporary census data in most countries means accepting the administrative units for which data are routinely reported and pretending that these are reasonable representations of socially relevant spaces. Some researchers will be attracted to studying late nineteenth-century U.S. cities especially because no boundaries within them are imposed.

Finally, another relatively recent development has been the creation of linked files across censuses. The MPC, for example, has used automated data mining techniques to link the 100-percent sample from 1880 with the smaller Integrated Public-Use Microdata Series samples that are already available for other years. If neighborhoods, cities, or counties have effects on people's subsequent lives (e.g., their subsequent rural-urban or interregional mobility, intergenerational or intragenerational occupational mobility, or marriage choices), these effects should be identifiable with a data set that includes their 1880 community characteristics in addition to individual attributes in 1880 and in a later decade.

There has been a significant spatial turn in the social sciences in the last decade. The aim of the Urban Transition HGIS is to facilitate the same direction of scholarly development for research in the late nineteenth century.

NOTE

1. NHGIS advises that a variable named GISJOIN can be used to link county shape files to population data downloaded from their Web site (see http://www.nhgis.org/mapping/using-data-in-a-gis/). This variable cannot be used to link shape files to population data aggregated from NAPP. Those users who wish to create their own specialized aggregate-variables data should use a different procedure, as follows. Both the NHGIS shape files and NAPP microdata include codes for states and counties. In NHGIS these are ICPSRST and ICPSRCTY; in NAPP they are STATICUS and COUNTYUS. These can be combined into a single unique identifier: in each file, multiply state code by 10,000 and add the county code. The 20 counties in the Idaho and Wyoming territories need to be handled separately, because the NHGIS shape files do not include ICPSRST and ICPSRCTY codes for these areas. But they can be identified by name and other codes.

REFERENCES

Anderson, M. 1988. *The American census: A social history*. New Haven, CT: Yale University Press.

Baker, A. R. H. 2003. Geography and history: Bridging the divide. New York: Cambridge University Press.

Bol, P. K. 2007. *China Historical Geographic Information System: Choices faced, lessons learned*. Harvard University. http://www.fas.harvard.edu/˜chgis.

De Moor, M., and T. Wiedemann. 2001. Reconstructing territorial units and hierarchies: A Belgian example. *History & Computing* 13:71–97.

Fitch, C., and S. Ruggles. 2003. Building the National Historical Geographic Information System. *Historical Methods* 36:41–60.

Gaffield, C. 2007. Conceptualizing and constructing the Canadian Century Research infrastructure. *Historical Methods* 40:54–64.

Greenberg, S. 1981. Industrial location and ethnic residential patterns in an industrializing city: Philadelphia, 1880. In *Philadelphia: Work, space, family, and group experience in the nineteenth century*, ed. Theodore Hershberg, 204–32. New York: Oxford University Press.

Gregory, I. N. 2002. The Great Britain historical GIS project: From maps to changing human geography. *Cartographic Journal* 39:37–49.

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

Holdsworth, D. W. 2003. Historical geography: New ways of imaging and seeing the past. *Progress in Human Geography* 27:486–93.

Kantrowitz, N. 1979. Racial and ethnic residential segregation in Boston, 1830–1970. *Annals of the American Academy of Political and Social Science* 441:41–54.

Kessner, T. 1977. *The golden door: Italian and Jewish immigrant mobility in New York City, 1880–1915*. New York: Oxford University Press.

Knowles, A. K. 2000. Historical GIS: The spatial turn in social science history. *Social Science History* 24:451–70.

Knowles, A. K., and A. Hillier. 2008. *Placing history: How maps, spatial data, and GIS are changing historical scholarship*. Redlands, CA: ESRI.

Lieberson, S. 1980. *A piece of the pie: Black and White immigrants since 1800*. Berkeley: University of California Press.

Logan, J. R., and H.-J. Shin. 2010. *Residential choice of immigrant groups in Newark City, 1880*. Paper presented at the Annual Meeting of the Eastern Sociological Society, Cambridge, MA, March.

Logan, J. R., and W. Zhang. 2010. *White ethnic residential segregation in historical perspective: U.S. cities in 1880*. Paper presented at the Annual Meeting of the Population Association of America, Dallas, TX, April.

Moore, D. D. 1981. *At home in America: Second generation New York Jews*. New York: Columbia University Press.

———. 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.

Nadel, S. 1990. *Little Germany: Ethnicity, religion, and class in New York City, 1845–80*. Urbana: University of Illinois Press.

Nugent, W. 1981. *Structures of American social history*. Bloomington: Indiana University Press.

Research Publications. 1984. *City directories of the United States, pre 1860 through 1901: A guide to the microfilm collection*. Woodbridge, CT: Author.

Roediger, D. 2005. *Working toward whiteness: How America's immigrants became white*. New York: Basic Books.

Rose-Redwood, R. S. 2007. Indexing the great ledger of the community: Urban house numbering, city directories, and the production of spatial legibility. *Journal of Historical Geography* 34:286–310.

Shelley, M. H. 1975. *Ward maps of United States cities: A selective checklist of pre-1900 maps in the Library of Congress*. Washington, DC: Not published.

Sies, M. C. 2001. North American suburbs, 1880–1950: Cultural and social reconsiderations. *Journal of Urban History* 27:313–46.

Thernstrom, S. 1973. *The other Bostonians: Poverty and progress in the American metropolis, 1880–1970*. Cambridge, MA: Harvard University Press.

White, M. J., R. F. Dymowski, and S. Wang. 1994. Ethnic neighbors and ethnic myths: An examination of residential segregation in 1910. In *After Ellis Island: Newcomers and natives in the 1910 census*, ed. S. C. Watkins, 175–208. New York: Russell Sage Foundation.

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