HISTORICAL METHODS
2016, VOL. 49, NO. 4, 187–197
http://dx.doi.org/10.1080/01615440.2016.1151393

Routledge
Taylor & Francis Group

Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930

Allison Shertzer[a,b], Randall P. Walsh[a,b], and John R. Logan[c]

[a]Department of Economics, University of Pittsburgh; [b]National Bureau of Economic Research; [c]Department of Sociology, Brown University

ABSTRACT
Most quantitative research on segregation and neighborhood change in American cities prior to 1940 has utilized data published by the Census Bureau at the ward level. The transcription of census manuscripts has made it possible to aggregate individual records to a finer level, the enumeration district (ED). Advances in geographic information systems (GIS) have facilitated mapping these data, opening new possibilities for historical GIS research. This article reports the creation of a mapped public use dataset for EDs in ten northern cities for each decade from 1900 to 1930. The authors illustrate a range of research topics that can now be pursued: recruitment into ethnic neighborhoods, the effects of comprehensive zoning on neighborhood change, and white flight from black neighbors.

KEYWORDS
Cities; historical GIS; segregation; urban history

Most quantitative research on segregation and neighborhood change in American cities prior to 1940 has been based on census data tabulated for city wards. Basing his analysis on ward data, for example, Allen Spear (1967, 7) stated that in Chicago in 1890 "[m]ost Negroes, although concentrated in certain sections of the city, lived in mixed neighborhoods." Stanley Lieberson (1980) found that black isolation in Northern cities—the black population share in areas (i.e., wards) where the average black person lived—was very limited at that time (averaging only 0.067) and did not rise appreciably until 1930 (when the average was 0.299). Lieberson's (1963) earlier research focused on white ethnic groups, showing that there was little segregation between "native whites" and English or German foreign-born persons but much higher segregation of Russians and Italians.

A longstanding concern for historical researchers has been that the ward is too large a geographic unit to capture neighborhood differences, especially in the early years of migration when new groups were relatively small. This issue, referred to by geographers as the Modifiable Areal Unit Problem (MAUP), arises when results at one spatial scale differ from results at another scale. The strongest critique of relying on ward data was voiced by Thomas Lee Philpott (1978, 120–1),who complained that the 1900 ward map for Chicago "shows blacks scattered over all of the Southwest Side, most of the South Side, and much of the West Side as well." In fact, his view was that "the residential confinement of the blacks was nearly complete at the turn of the century." Chicago had only 35 wards in 1900, averaging nearly 50,000 persons per ward. Black neighborhoods, Philpott argued, were invisible within such large zones. One response to these critiques has been to try to adjust segregation scores for wards to be more consistent with those calculated at the finer level of census tracts. Nathan Kantrowitz (1979) compared ward and tract measures for six cities in 1930, concluding that there is approximately a 10-point difference between them (when the Index of Dissimilarity is scaled to a 0–100 range). More recently, David Cutler, Edward Glaeser, and Jacob Vigdor (1999) traced a century-long trend of black-white segregation, similarly relying on ward data for 1890–1930, with adjustments based on differences between ward and tract measures in 1940. Although on average this approach is reasonable, it must assume (contrary to what is known from 1940 when both ward and tract data are available) that the differential is constant across cities and over time. Further, it does not allow researchers to identify neighborhood variation within wards, which is often where meaningful changes were occurring.

It is now possible to do better by analyzing data at the level of enumeration districts (EDs) that are smaller even than census tracts (typically fewer than 2,000 residents). A major source of genealogical data, Ancestry.com, has transcribed portions of all the individual records from

CONTACT Allison Shertzer shertzer@pitt.edu Department of Economics, University of Pittsburgh, 4901 W. W. Posvar Hall, Pittsburgh, PA 15260, USA.
Color versions of one or more figures in this article are available online at www.tandfonline.com/vhim.

pre-1950 censuses and used them to construct a finder index for users of its web-based system. Coauthor Allison Shertzer obtained permission to assemble these data from Ancestry's webpage for the four census years 1900–30 for ten cities.[1] The cities are New York, Chicago, Boston, Baltimore, Philadelphia, Cincinnati, Cleveland, St. Louis, Pittsburgh, and Detroit. These cities included over 18 million residents in 1930 (about half the total in the largest 100 cities). All but Cincinnati were among the largest ten cities in 1940 (the other was Los Angeles, which was only the thirty-sixth largest city in 1900 when Cincinnati was tenth largest). We have cleaned the individual records and aggregated data to the ED level.

In addition, we have created historically accurate GIS maps of the EDs in each decade. Researchers now will be able to study areas that are closer to the scale of "neighborhoods" instead of the large urban districts represented by wards. Having comparable data for several decades also makes it possible to study how neighborhoods change. The boundaries of EDs shift over time, but their relatively small size and social homogeneity facilitates the use of interpolation methods to estimate data for areas with constant boundaries (Goodchild, Anselin, and Deichmann 1993). We present one approach to such interpolation, harmonizing the ED-level data to hexagon-shaped areas that are similar in size to EDs but have constant boundaries over time. We also describe an application that leverages this ability to make intertemporal comparisons across constant boundary neighborhoods.

In the following sections, we describe the data that have been made available in this way and the procedures used to create the ED maps. We then illustrate the potential uses of these files with initial findings from our own research: analysis of trends in segregation, recruitment into ethnic neighborhoods, native white flight from neighborhoods where new immigrant groups were settling, and the effects of the introduction of comprehensive zoning on neighborhood change.

The data

Census data provide a summary of the composition of populations over time, and they are an essential component of quantitative historical research (Baker 2003; Holdsworth 2003; Sies 2001). Much research is now based on microdata that have been transcribed from individuals' records in the census manuscripts, providing information about households and persons within households. Although two decades ago historians must pore through these records manually, the Minnesota Population Center's IPUMS project has become the standard source for such samples, and they are available as early as 1790. Although data on individuals and households are valuable in themselves, Catherine Fitch and Steven Ruggles (2003) argued that the lack of usable historical census geography across multiple cities has limited the scope of research that can be done with these files.

Studies at the neighborhood, community, or larger geographic levels make use of summary data, tables showing the frequency distribution or a cross-tabulation of individual-level information, aggregated to a given unit of geography, such as a census tract or a city. The Census Bureau has a long tradition of providing such tabulations in published volumes, and much data are now available in digital format. 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; Logan and Zhang 2012). Major historical GIS projects outside the United States 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 Moore and Wiedemann 2001; http://www.hisgis.be/start_en.htm), and the China Historical GIS (Bol 2007; http://www.fas.harvard.edu/~chgis).

The project reported here provides new mapped data for American cities for the years 1900 to 1930, complementing sources that are already available for an earlier year (1880) and subsequent years (1940 onward). For the period from 1940 to 2000, MPC's National Historical GIS (NHGIS) project provides census tract data based on tabulations originally prepared by the Bureau of the Census. 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. The Urban Transition HGIS (Logan et al. 2011) pushes the time horizon for historical spatial analysis back to 1880. Built on the 1880 census, it provides additional layers of spatial data based on 100% microdata, including (1) contextual variables at the level of counties and EDs that can be added to individual person and household records for all persons across the nation, (2) accurate GIS maps of EDs 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. With such fine geographic detail, the researcher is free to study local areas at any scale (next-door neighbors, the street segment, block, or any larger area), or to define neighborhoods in terms of their ethnic composition, class composition, or any other measured social characteristic.

Table 1. Summary statistics for ED dataset.

Item	1900	1910	1920	1930
Population	1,551	1,460	1,260	1,414
	(620)	(661)	(705)	(832)
Maximum population	7,054	6,239	14,065	9,329
Area (square meters)	292,750	234,475	204,565	219,942
	(931,391)	(669,940)	(629,409)	(466,923)
Number of EDs per city	374	511	703	689
	(295)	(400)	(558)	(575)
Immigrants	482	491	363	366
	(380)	(422)	(330)	(316)
Blacks	53	53	65	117
	(128)	(150)	(211)	(344)
White natives*	455	415	403	477
	(276)	(284)	(343)	(441)
Total EDs	5,824	7,894	10,973	11,129

Note: The mean is provided with the standard deviation below in parantheses except for the maximum population.
*White natives are defined as white individuals who were both born in the United States and whose fathers were also born in the United States.

Because the 1900–30 data in the current project were drawn from Ancestry's indexed records for 1900–30, analysis at the ED level is limited to variables included in that index. The person's and parents' race and place of birth are the basis of categories of race and ethnicity. The concept of "generation" is central to immigration studies. Among whites, we have created ethnic categories based on country of birth and parents' country of birth. These categories include several groups in the earlier wave of immigration (England, Scotland, Canada, Ireland, Norway, Sweden, and Germany) and several newer groups (Russia, Italy, Austria, Hungary, Poland, and Czechoslovakia). First-generation immigrants are those born in the country of origin; second-generation persons are classified by the father's country of birth. Whites in the third or later generation are often referred to as "native whites" or native whites of native parentage. Blacks and "other race" persons (mostly Asian) are also identified.

U.S.-born persons can also be classified by their state of residence, particularly to compare those who were born in their state of residence to those who moved from another state. Migration within the United States is especially significant for blacks, and we have classified blacks born in states where slavery was legal prior to 1860 versus those born in other states (approximately a South vs. non-South distinction), with a separate category for those born abroad or whose parents were born abroad. Year of immigration is also available for every decade. This makes it possible to separately identify recent immigrants from foreign-born individuals who have long resided in the United States. Other data for every decade are sex, age, and geographic identifiers of ED, ward or assembly district, city, and state.[2]

Table 1 provides summary statistics for the ED data by year from 1900 to 1930. The average population of an ED fluctuates relatively little over the period, from a low of 1,260 in 1920 to a high of 1,551 in 1900. This finding is not unexpected given the manner in which the census was conducted during this period: One enumerator was tasked with surveying every individual in an ED during a fixed window of time, limiting the average maximum size of these districts. Nonetheless, there were a few EDs with much larger populations in each year (for instance as large as 9,000 people in 1930). These larger districts generally corresponded to apartment buildings occupying a single city block.

While the average population of an ED remained constant, the average area of an ED shrank over the period from 0.29 square kilometers in 1900 to 0.22 in 1930, consistent with increasing urban density occurring over the period. At the same time, the average number of EDs per city nearly doubled, from 374 in 1900 to 689 in 1930, reflecting the substantial urbanization-driven population growth of the early twentieth century. Table 1 also provides a simple demographic breakdown of the ED-level population in each census year. The average number of black residents per ED doubled over the period as the migration from the South got underway, while the number of immigrants per ED declined from 482 in 1900 to 366 in 1930, consistent with the slowdown in European migration associated with World War I and the anti-immigration National Origins Quotas Acts of 1921 and 1924.

Some researchers will be interested in combining ED data from this project with sample data for individuals that are available from the Integrated Public Use Microdata project at the University of Minnesota (www.ipums.org). These sample data include an ED identifier. In every decade, they also include individual-level indicators of race, ethnicity, and immigrant generation; marital status; and age. Occupation is the only available social class measure. It is typically included in analyses as an interval scale socioeconomic index (SEI) based on rankings of occupations' income, education, and prestige in 1950. Another important occupational characteristic is the category of domestic servant, important because many blacks and first-generation immigrants worked as servants and lived in their employers' home, especially in the earlier decades. We include below an example of how linked microdata and ED data can be used to study who lived in what kinds of neighborhoods.

Mapping procedures

Creating the ED map for every city in every decade is straightforward in principle but difficult in practice. Fortunately, in the 1900 to 1930 period, the Bureau of the Census created accurate maps showing the boundaries of EDs, and these are available in the National Archives.[3] These were photographed and used as the primary basis for determining ED boundaries. Figure 1 illustrates such

Figure 1. A section of the Census Bureau's 1910 ED map for Baltimore. *Note*: ED 323 in Ward 19, Tract S6, bounded by Saratoga, Mount, Lexington, Gilmor, Fayette, and Fulton is highlighted.

a map for a section of Baltimore in 1910. The quality of these maps is not uniform, however, and in some cases it was necessary to refer to other sources. Our secondary reference for ED boundaries was the written description of the EDs available on microfilm from the National Archives and transcribed by Stephen P. Morse.[4] Our third source of geographic data was a set of detailed maps, including all major street names, for the cities of Cleveland, Detroit, and St. Louis, made available by the University of Alabama Historical Map Collection.

Mapping EDs in a GIS system required a historically accurate street grid for each city. It would be possible to do this task in part with a contemporary street file, but in every city there are many new streets, older streets have been removed or renamed, and in some cases the entire alignment of streets has been altered. For historically accurate GIS street maps in six cities, we relied on 1930 maps in the Historical Urban Ecological Data Set created by the Center for Population Economics (CPE) at the University of Chicago (Villarreal et al. 2014). These include New York, Chicago, Philadelphia, Boston, Baltimore, and Cincinnati. To facilitate construction of the Pittsburgh ED maps, University of Pittsburgh team members georeferenced images of a comprehensive street survey conducted by the city of Pittsburgh beginning in 1923.[5] For Cleveland, St. Louis, and Detroit, we relied on 1940 street maps produced by the Urban Transition Historical GIS Project at Brown University that are based on paper maps created by the Census Bureau.

CPE street files were created from georeferenced historical maps, modern orthographic imagery, and Census TIGER Line files from 2007. The 1940 street files were drawn from 2010 shape files produced by the Census Bureau, edited to account for changes between 1940 and 2010. Finally, the Pittsburgh survey maps were georeferenced to match the geography of current day city streets taken from the 2010 U.S. Census TIGER/Line files. Hence, all of these maps can be linked to contemporary maps and data.

Figure 2 shows the street grid for Baltimore in 1930, corresponding to the area in Figure 1. The highlighted area shows the six city blocks of ED 323. ED boundaries were manually added to the street map to create a new layer in the shapefile. Many EDs are not polygons like this one, but rather are comprised of institutions or apartment buildings at a single address that fall within the boundaries of a larger ED. These institutional EDs were added to the map as smaller polygons at the approximate location of that address.

A major difficulty in making use of mapped data over time is that the boundaries of EDs shift from decade to decade. Forming consistent neighborhoods is challenging because unlike the case of modern-day census tracts, most changes were more complex than simple combinations or bifurcations. Our approach to this problem is to harmonize ED data to temporally invariant geographically defined areas that we treat as "synthetic neighborhoods" to study neighborhood change. We emphasize that scholars can use the ED shapefiles to create any size

Figure 2. The street grid for the same area of Baltimore in 1930. *Note*: The figure shows the street grid for Baltimore in 1930, and ED 323 is highlighted in red (see Villarreal et al. 2014 for details on how the street file was constructed).

or shape of synthetic neighborhoods desired. As an example, for this application, we cover each city with an evenly spaced grid of symmetric hexagons measuring 800 meters along their largest width. This strategy follows on Spencer Banzhaf and Randall Walsh (2008). We utilize hexagons because they are the most compact way to cover the plane with symmetric shapes.[6] The chosen size yields average populations that are comparable to modern-day census tracts (ranging from roughly 2,500 individuals in 1900 to just over 4,200 in 1930).[7]

The demographic composition of these synthetic neighborhoods is then imputed as the spatially weighted average of the underlying ED level data from each census. These areas are spatially invariant over time. This method is illustrated in Figure 3, using the same area of Baltimore (including ED 323) as in the previous maps. The figure shows the overlay of a hexagon drawn on top of the ED pattern. All population data for EDs that lie fully within the hexagon (as ED 323 does) are attributed to it. Data from other EDs are attributed to the hexagon in proportion to the share of land area of the ED that lies within it. This particular hexagon (which we will treat as a "neighborhood" for the purpose of studying change over time) encompasses all of three EDs and portions of 14 others. Areal interpolation is subject to some error (because populations are not uniformly spread within EDs). Given that the size, offset, and orientation of the hexagon grids were chosen based on factors that are independent of any neighborhood-level geography or characteristics, it is reasonable to presume that there is no systematic relationship between the grid geography and neighborhood characteristics.[8] Public availability of the ED level maps we have produced will facilitate future research into the impact of fixed neighborhood geography choice on empirical findings.

Uses of the 1900–30 mapped data

The public use files that will be made available through this project have three components for every decade. The first is a GIS shapefile that includes the street layer

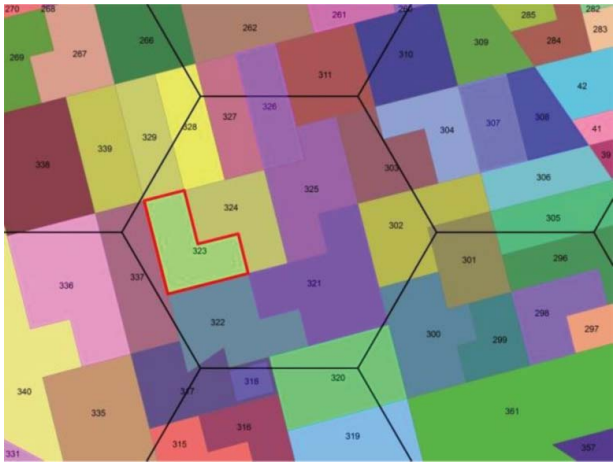

Figure 3. Illustration of a hexagon layer, showing the relationship between these hexagons and EDs. *Note*: The figure shows a sample of EDs for Baltimore in 1910. See Figure 1 for details of the location.

for 1930 or 1940, an ED layer specific to that decade, and a hexagon layer.[9] Second, for every ED, there are counts of several population variables from aggregating the 100% microdata. Third, we provide estimates of the same population variables from imputing from ED counts for hexagon neighborhoods. The included variables are total population, immigrants by country of birth, counts of the second generation based on father's country of birth, third or more-generation whites, blacks, and counts by age.

We offer several examples of ways in which these data can be used. These applications reflect the authors' interests in racial segregation and racial or ethnic change. They include non-spatial analyses that depend only on the availability of data aggregated to EDs, spatial analyses that make use of EDs as they existed in each decade, and spatial analyses that make use of synthetic neighborhoods that have constant boundaries over time.

Non-spatial applications

John Logan and colleagues (2015) have used the ED data for these cities to assess the level of segregation between blacks and whites over time. Studies at the ward level have concluded that segregation was moderate in the early twentieth century, increasing to a high level only after the acceleration of the Great Migration of blacks from the rural South to selected northern cities after 1920. Figure 4 reproduces some of the results from their study. It shows the average values across these cities (weighted by the size of their black population) for two common measures of segregation. The first is the Index of Dissimilarity (D), which measures the evenness of the distribution of blacks and whites across areas. It ranges between 0 (where there is no difference in the two groups' distributions) to 1 (when there is no overlap at all between them). Both show the same trajectory from 1900 to 1930, but the levels of D at the scale of EDs are about 0.20 higher than at the ward scale, which is what scholars previously relied upon. At the ward scale, D barely reached the level that scholars consider to be "very high" segregation (0.60) by 1920; at the ED scale, the value was above 0.60 already in 1900. Although there is a high positive correlation between the values of D at the ED and ward scale, the ED data reveal substantial differences in the timing of extreme segregation that affect interpretation of the phenomenon.

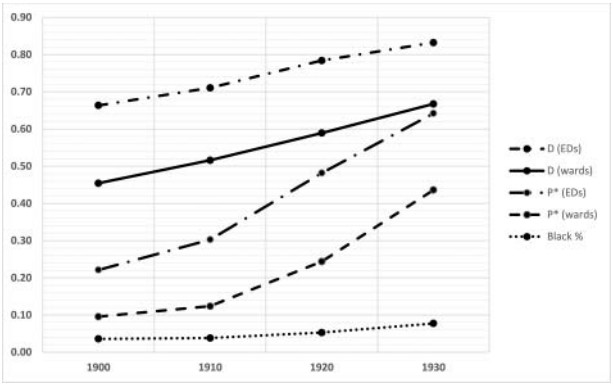

Figure 4. Trends in segregation for ten Northern cities at the ED and ward level: Index of Dissimilarity (D), Isolation Index (p*), and black population share. *Note*: Averages across cities weighted by the size of the black population in each year. Ward data were previously reported by Cutler and colleagues (1999), and are available from their Segregation Data Webpage (http://www.nber.org/data/segregation.html).

The other measure is the black Isolation Index, which is the percentage of black neighbors in the area where the average black person lived. As context, the figure also shows the actual black share of city residents, which was below 5% in 1900 and rose slowly toward close to 9% in 1930. At the ward scale, blacks were close to a majority in their area (about 43%) by 1940; at the ED scale, they were closer to a majority already in 1920. Of course, scholars are interested not only in the average value of segregation, but also how it varied across cities and what factors accounted for its increase over time. These are questions that can be tackled with the new ED data.

One approach to further analysis is to ask which black residents lived in areas with larger shares of black neighbors, and more specifically whether blacks with higher occupational status or those who were born in the same city experienced less racial isolation than migrants. Logan and colleagues (2015) studied this issue with a form of multilevel modeling, where information about where a person lived (the ED data from this project) was combined with information about individual adult black residents. Models of this type (i.e., locational attainment models) are increasingly used in the urban research literature. Other models using the same combination of individual and ED data could be applied to questions of neighborhood effects, where the ED characteristic is a predictor rather than an outcome. Such models are especially attractive when data are found at the individual level from non-census sources. A recent example (Xu, Short, and Logan 2014) combined ED-level census data from 1880 with information from children's death certificates in Newark, New Jersey. The analysis shows how the ethnic homogeneity of neighborhoods affects child health outcomes.

Table 2 reports results of a locational attainment analysis for black residents in 1900 and 1930 in our ten Northern cities (see Logan et al. 2015). The racial composition of people's neighborhoods depends very much on the overall size of the black population and its segregation in the city where they live (seen in the effects of

Table 2. Multilevel regression predicting percent black in the ED where a person lies, 1880–1940 (black persons aged 15 and older).

	1900		1930	
	b	SE	b	SE
Female	−2.41	(0.56)***	−0.88	(0.87)
Age	−0.01	(0.03)	0.00	(0.04)
Southern born	0.02	(0.56)	5.05	(0.95)***
Marital status (REF = single)				
Married	1.99	(0.80)*	−0.23	(1.28)
Divorced/widowed	4.43	(0.98)***	−0.84	(1.51)
Lives with non-relatives only	1.15	(0.74)	1.28	(1.16)
Literate	1.18	(0.83)	4.67	(2.32)*
Highest family member's SEI	0.12	(0.02)***	0.13	(0.03)***
Owner	−12.49	(0.97)***	−8.39	(1.33)***
City-level D_{bw}	0.91	(0.07)***	1.88	(0.05)***
City-level % black	4.09	(0.26)***	2.25	(0.16)***
Constant	−55.31	(5.89)***	−120.58	(5.86)***
R^2: overall	.08		.24	
R^2: within	.05		.02	
R^2: between	.57		.91	
N	5,801		4,923	

*$p < .05$; **$p < .01$; ***$p < .001$.
Note: Based on 100% microdata and ED racial composition in ten Northern cities as reported in Logan and colleagues (2015).

city-level variables and in the large between-city R^2). The family's occupational standing (measured by its SEI) also has a significant effect, but it is in the opposite direction of what might be predicted: Higher status blacks lived in EDs with higher black shares. This association possibly reflects the advantages for black professionals and business owners of living near their place of business and black customers. Black homeowners lived in more racially mixed neighborhoods, though this effect was limited by the very small share of blacks who owned their home.

Spatial analysis of EDs as defined in each decade

The most direct use of the mapped data is to visualize the spatial distribution of populations and how it changes over time. Figure 5 shows a map of EDs in St. Louis in 1900 and 1930. EDs that are shaded in black were more than 70% black; unshaded areas were less than 3% black. The maps show that as early as 1900, when St. Louis's black population was barely 6% of all residents, there were small pockets of concentrated black settlement in the downtown area just west of the Mississippi River. By 1930, these pockets had expanded and consolidated to form what many observers would describe as a black ghetto. The post-1940 era saw an expansion of this zone to include much of the city and (today) large sections of suburban St. Louis (Gordon 2009).

Another form of analysis is to relate population composition to other neighborhood characteristics. Here, we point out how data from other sources can be analyzed together with our population data when they are combined using GIS software. Allison Shertzer, Tate Twinam, and Randall Walsh (2016a) have digitized the universe of pre-zoning land use from a survey conducted by the city of Chicago in 1922 and compared it to the city's initial zoning code passed in 1923. Figure 6 illustrates a pattern that is consistent with Burgess's concentric zone model of the city at that time (Burgess 1925), in which the densest residential and commercial areas tend to be centrally located. The figure shows ED population density in 1920 plotted against the share of the district zoned for commercial uses in 1923. The upward-sloping relationship indicates that commercial uses tended to cluster around areas of highest population density.

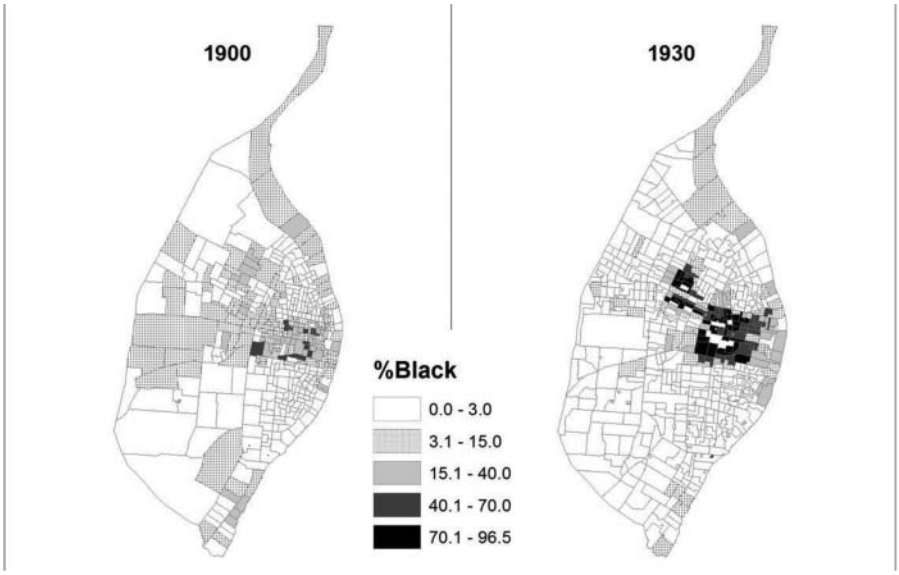

Figure 5. Distribution of EDs by Percent Black, St. Louis in 1900 and 1930. *Note*: The figure shows the digitized 1900 and 1930 ED maps for the city of St. Louis.

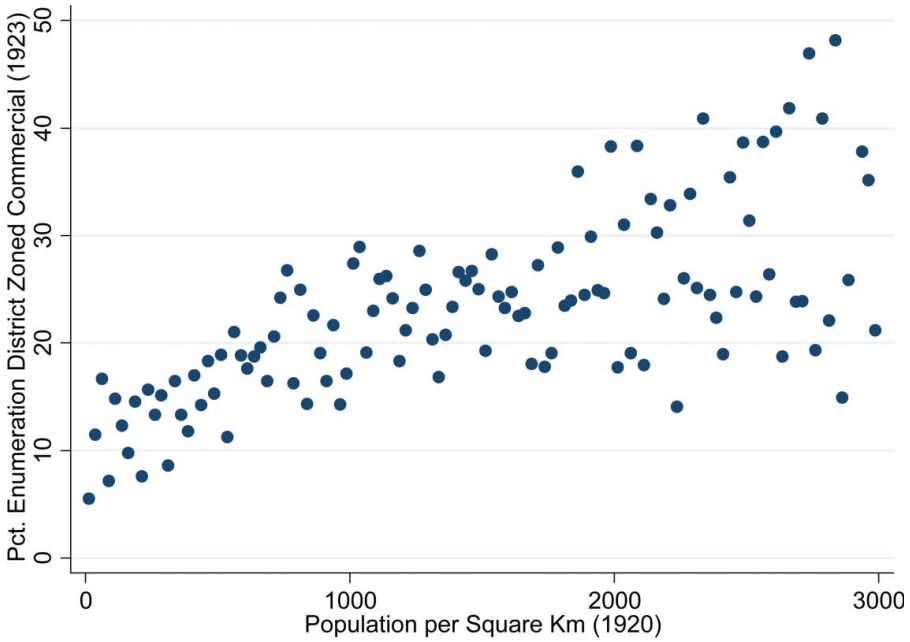

Figure 6. Zoning in Chicago 1920: population density by share of the ED that was zoned for commercial use. *Notes*: The population density is estimated using the population of 1920 census EDs. The commercial zoning share is taken from Chicago's initial zoning ordinance. See Shertzer and colleagues (2016a) for details of how the zoning ordinance was digitized.

However, among the most densely populated EDs, there is a wider range of commercial zoning, suggesting that there were two types of densely developed neighborhoods, some mainly residential and some mixed-use. Exploring these relationships will provide new insights into the organization of land uses in urban areas.

The urban geography literature has emphasized the competing objectives of these two land uses: Commercial activities demand locations as close as possible to residential areas, while homeowners desire neighborhoods that are purely residential (Song and Knapp 2004). Once zoning laws are enacted, they affect how areas develop in the future, so it is hard to know how they would have evolved in the absence of zoning (see Shertzer, Twinam, and Walsh 2016b). We can, however, explore whose interests were best protected by Chicago's new zoning ordinance. In research in progress, we find that (even after controlling for the extent of industrial land use prior to 1923) black neighborhoods were more likely than white neighborhoods to be zoned industrial. This is another example of how land use data can be combined with the ED population data using GIS to investigate questions in urban geography and economic history.

Spatial analysis of synthetic neighborhood change

The last example demonstrates the use of constant (synthetic) neighborhood areas over time. Allison Shertzer and Randall Walsh (2016) used the full sample of hexagon neighborhoods to revisit the question of why American cities are so segregated by race, focusing on the mechanism of "white flight" from central cities. This exercise shows how causal models of "push" factors that were developed using data from the postwar period (e.g., see Boustan 2010; Card, Mas, and Rothstein 2008) can now be applied to the early twentieth century, when segregation rose most rapidly in large cities. Many scholars have discussed the structural conditions that promoted segregation, including specific factors in the housing sector (e.g., deed restrictions and redlining) and broader inequalities in the other domains such as the labor market and political arena (Massey and Denton 1993; Sugrue 1996). Demographers have also emphasized pull factors that contributed to white flight from changing neighborhoods, such as new housing developments in suburbia (Frey 1979). This analysis shows how a particular type of synthetic neighborhoods constructed from the underlying ED dataset can be used to investigate how conditions at the neighborhood level influenced the specific patterns of racial change in a city.

Our conceptualization of white flight derives from Thomas Schelling's model of neighborhood "tipping" (1971). This model posits that below some threshold or tipping point of neighborhood black share, white population changes should be relatively stable or increasing. However, whenever the baseline black population share exceeds this tipping point, the model predicts that declines in white population will become self-reinforcing, and a rapid drop in white population will occur as the community moves to a new majority black equilibrium. In Figure 7, we present simple nonparametric estimations of the relationship between neighborhood black share in the baseline year and the

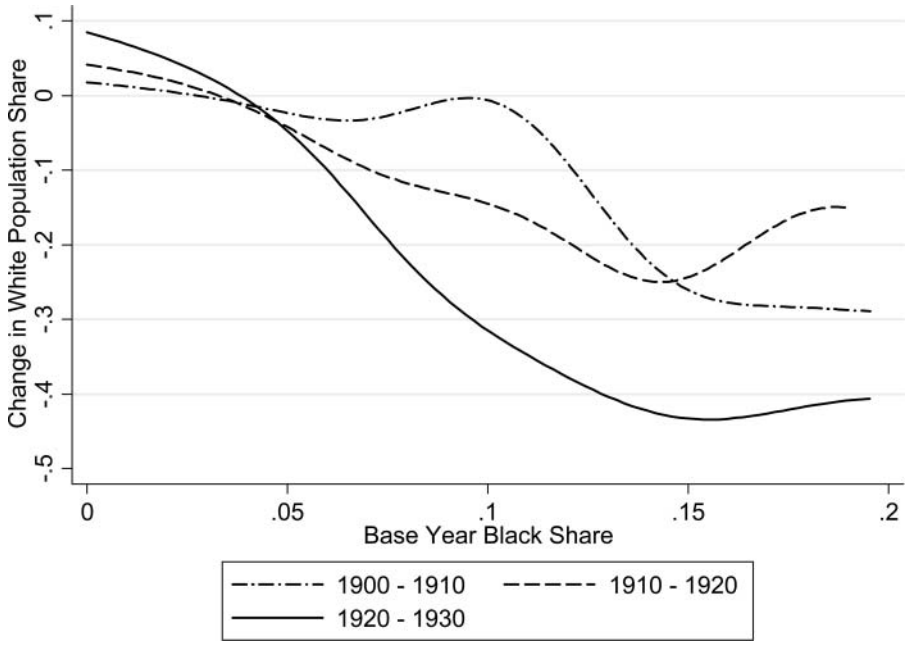

Figure 7. Analysis of racial tipping for synthetic neighborhood areas of Chicago, 1900–30. *Note*: The figure shows the white population share change as a function of the initial black share of the population at the hexagon level for Chicago. For estimation details, see Shertzer and Walsh (2016).

change in neighborhood white population share over the ensuing decade (see Shertzer and Walsh 2016 for a more detailed discussion of nonlinear estimation of demographic neighborhood change). In each of the first three decades of the 1900s, our analysis suggests a tipping point in the vicinity of 5% neighborhood black share. The overall magnitude of this estimated tipping effect roughly doubles over the 30-year period. By the third decade (1920–30), neighborhoods with baseline (1920) black shares above 10% on average experienced relative reductions in their white populations of greater than 40 percentage points, suggesting that white flight from black neighbors was a quantitatively important and accelerating phenomenon in prewar America.

This example treats the racial composition of a neighborhood as the only factor influencing change over time. More elaborate studies could take other factors into account, such as the "pull" factors that motivated white residents to live elsewhere (see Frey 1979), or the deed restrictions or other constraints on where black residents could move. Any study of such topics requires data on how neighborhoods compared to others in the city and how they changed over time, which is the primary contribution of the new HGIS system described here.

Conclusion

The ED dataset described in this article will be of interest to a broad range of scholars in the social sciences. The capacity of researchers to investigate the early development of American cities was previously hampered by the lack of systematic fine-scale spatial data covering urban areas in the early twentieth century. Users who wish to study small neighborhoods in cities will find the ED summary files useful because they cover much smaller geographic units than the ward-level data published by the Census Bureau, yet provide information on spatial location. The ED dataset thus serves as a bridge between extant ward-level analyses and the more recent neighbor-to-neighbor measures of segregation that exploit the ordering of individuals on census manuscripts (Grigoryeva and Ruef 2015; Logan and Parman 2015).

For example, segregation indices for recently arrived Italian immigrants can be computed at the ED level, or the demographic characteristics of neighborhoods selected by black migrants from Alabama contrasted with those chosen by black migrants from Virginia. Scholars can investigate how segregation varies depending on how neighborhoods are defined and contrast traditional metrics with recently developed measures of how likely blacks or immigrants were to have white native neighbors, providing significant opportunities to broaden empirical inquiry into traditional segregation analyses. The early twentieth century saw the divergence of the socioeconomic fortunes of urban blacks from that of immigrants from southern and eastern Europe (Lieberson 1980). The trajectory of residential segregation experienced by these different racial and ethnic groups can be more fully studied with the ED data files, providing a new point of reference to a literature that has focused on timing of arrival and generational convergence (e.g., Glazer 1971).

The spatial data are intended to serve several types of users. For users who wish to visualize and inspect the data, the web-based mapping system now available for 1880 data (www.s4.brown.edu/utp) will be expanded to include these northern cities for each decade from 1900 to 1930. The GIS shapefiles and population data aggregated to the ED level will also be made available through the National Bureau of Economic Research. As public use files from the Minnesota Population Center become available, users will be able to create and map additional variables. Experienced users can add spatially referenced information from any other historical sources. For instance, scholars can create and merge data (as illustrated above) on city zoning or infrastructure characteristics, disease outbreaks from public health reports, the location of race riots reported in historical newspapers, or the location of particular industrial or retail establishments. The ability to combine the ED summary files with microdata from the Integrated Public Use Microdata project at the University of Minnesota provides even more avenues for potential scholarship.

Another application of these data is to combine them with existing tract-level maps for the years 1940 and beyond, linking questions from urban and economic history to contemporary phenomenon. The study of white flight reported above is an example of this type of scholarly inquiry. The synthetic (hexagon) neighborhood dataset was constructed using the most current methods in spatial analysis, allowing for the deployment of the same empirical methodologies used to study white flight from black arrivals in the 1940 to 1970 period. Estimates of the magnitude of white flight can thus be estimated consistently for the entire twentieth century. Especially in dense urban areas where contemporary census tracts cover small areas, the synthetic neighborhoods constructed here can be replicated for later decades. We expect this feature of the dataset to be particularly attractive to urban researchers.

A growing library of high-resolution spatial urban datasets has significantly enlarged the potential for scholarly inquiry into the history of American cities. The maps and population data assembled in this project build on the existing statistical and spatial infrastructure, and provide clear new advantages and opportunities for scholarly inquiry into the development of American cities in the early twentieth century.

Acknowledgments

Antonio Diaz-Guy, Phil Wetzel, Julia Burdick-Will, Weiwei Zhang, Jeremy Brown, Andrew O'Rourke, Aly Caito, Loleta Lee, and Zach Gozlan provided outstanding research assistance. Additional support was provided by the Central Research Development Fund and the Center on Race and Social Problems at the University of Pittsburgh. We thank David Ash and the California Center for Population Research for providing support for the microdata collection, Carlos Villarreal and the Center for Population Economics at the University of Chicago for the 1930 street files, Jean Roth for her assistance with the national Ancestry.com data, and Martin Brennan and Jean-Francois Richard for their support of the project. We are grateful to Ancestry.com for providing access to the digitized census manuscripts.

Funding

This research was supported by grants from National Science Foundation (SES-1355693, SES-1459847), National Institutes of Health (1R01HD075785-01A1), and by the staff of the research initiative on Spatial Structures in the Social Sciences at Brown University. The Population Studies and Training Center at Brown University (R24HD041020) provided general support.

Notes

1. In collaboration with Ancestry, the Minnesota Population Center is in the process of cleaning Ancestry's transcribed data files and coding variables in conformity with the procedures followed in the Integrated Public Use Microdata Samples (IPUMS) files for each of these census years. When the final versions of these files become publicly available, the research described here can be extended to other cities, for instance in the South.
2. Ward was not recorded on many census manuscript forms in 1930.
3. Census Enumeration District Maps; Enumeration District and Related Maps, Records of the Bureau of the Census, Record Group 29; National Archives Cartographic Branch, College Park, MD (National Archives Identifier 821491).
4. http://stevemorse.org/ed/ed.php
5. City of Pittsburgh Geodetic and Topographic Survey Maps, 1923–61 are available from Historic Pittsburgh at http://images.library.pitt.edu/g/geotopo/.
6. Squares or triangles would also cover the plane, but are less compact than hexagons. Circles are more compact than hexagons, but leave gaps and thus do not cover the plane.
7. Although hexagons that are comparable in size to modern census tracts were the most suitable for our application, we note that larger or smaller geographic units may be appropriate for other empirical questions. There is a trade-off between the size of the synthetic neighborhood and the extent of measurement error in the areal interpolation. This is an important area for future research as more high-resolution data (e.g., at the street or address level) become available.
8. Size was chosen to roughly match modern-day census tract population, orientation to match compass points, and specific offset to completely cover each city (including a 10 km buffer) with a rectangular hexagon grid.
9. The exception is Pittsburgh, for which we do not have a historically consistent street layer. To create the ED shapefile for this city, we worked directly from a combination of georeferenced contemporary maps and current TIGER files.

References

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

Banzhaf, H. S., and R. Walsh. 2008. Do people vote with their feet? An empirical test of Tiebout's mechanism. *American Economic Review* 89:843–63.

Bol, P. K. 2007. The China historical geographic information system: Choices faced, lessons learned. http://www.fas.harvard.edu/~chgis (accessed January 14, 2010).

Boustan, L. P. 2010. Was postwar suburbanization "white flight"? Evidence from the black migration. *The Quarterly Journal of Economics* 125 (1):417–43.

Burgess, E. W. 1925. (1967). The growth of the city: An introduction to a research project. In *The city*, ed. R. E. Park, E. W. Burgess, and R. D. McKenzie, 47–62. Chicago: University of Chicago Press.

Card, D., A. Mas, and J. Rothstein. 2008. Tipping and the dynamics of segregation. *Quarterly Journal of Economics* 123 (1):177–218.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor. 1999. The rise and decline of the American ghetto. *Journal of Political Economy* 107:455–506.

De Moor, M., and T. Wiedemann. 2001. Reconstructing territorial units and hierarchies: A Belgian example. *History & Computing* 13 (1):71–97.

Fitch, C., and S. Ruggles. 2003. Building the national historical geographic information system. *Historical Methods* 36 (1):41–60.

Frey, W. H. 1979. Central city white flight: Racial and nonracial causes. *American Sociological Review* 44 (3):425–48.

Gaffield, C. 2007. Conceptualizing and constructing the Canadian century research infrastructure. *Historical Methods* 40 (2):54–64.

Glazer, N. 1971. Blacks and ethnic groups: The difference, and the political difference it makes. *Social Problems* 18 (4):444–61.

Goodchild, M. F., L. Anselin, and U. Deichmann. 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25:383–97.

Gordon, C. 2009. *Mapping decline: St. Louis and the fate of the American city*. Philadelphia: University of Pennsylvania Press.

Gregory, I. N. 2002. The Great Britain historical GIS project: From maps to changing human geography. *Cartographic Journal* 39 (1):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 (5):638–53.

Grigoryeva, A., and M. Ruef. 2015. The historical demography of racial segregation. *American Sociological Review* 80 (4):814–42.

Holdsworth, D. W. 2003. Historical geography: New ways of imaging and seeing the past. *Progress in Human Geography* 27 (4):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.

Knowles, A. K. 2000. Historical GIS: The spatial turn in social science history. *Thematic Issue of Social Science History* 24 (3):451–70.

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

Lieberson, S. 1963. *Ethnic patterns in American cities*. New York: The Free Press.

Lieberson, S. 1980. *A piece of the pie: Blacks and white immigrants since 1880*. Berkeley: University of California Press.

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.

Logan, T., and J. Parman. 2015. The national rise in residential segregation. NBER Working Paper 20934. Cambridge, MA: National Bureau of Economic Research.

Logan, J. R., and W. Zhang. 2012. White ethnic residential segregation in historical perspective: U.S. cities in 1880. *Social Science Research* 41 (5):1292–1306.

Logan, J. R., W. Zhang, R. Turner, and A. Shertzer. 2015. Creating the black ghetto: Black residential patterns before and during the great migration. *The Annals of the American Academy of Political and Social Science* 660 (1):1–35.

Massey, D. S., and N. A. Denton. 1993. *American apartheid: Segregation and the making of the underclass*. Cambridge: Harvard University Press.

Philpott, T. L. 1978. *The slum and the ghetto: Neighborhood deterioration and middle-class reform, Chicago, 1880–1930*. New York: Oxford University Press.

Schelling, T. C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* 1:143–86.

Shertzer, A., T. Twinam, and R. P. Walsh. 2016a. Race, ethnicity, and discriminatory zoning. *American Economic Journal: Applied Economics* 8 (13):217–46.

Shertzer, A., T. Twinam, and R. P. Walsh. 2016b. Zoning and urban persistence. Manuscript.

Shertzer, A., and R. P. Walsh. 2016. Racial sorting and the emergence of segregation in American cities. NBER Working Paper 22077. Cambridge, MA: National Bureau of Economic Research.

Sies, M. C. 2001. North American suburbs, 1880–1950: Cultural and social reconsiderations. *Journal of Urban History* 27 (3):313–46.

Song, Y., and G. Y. Knaap. 2004. Measuring the effects of mixed land uses on housing values. *Regional Science and Urban Economics* 34:663–80.

Spear, A. 1967. *Black Chicago: The making of a negro ghetto 1890–1920*. Chicago: University of Chicago Press.

Sugrue, T. 1996. *The origins of the urban crisis*. Princeton, NJ: Princeton University Press.

Villarreal, C., B. Bettenhausen, E. Hanss, and J. Hersh. 2014. Historical health conditions in major U.S. cities: The HUE data set. *Historical Methods* 47 (2):67–80.

Xu, H., J. R. Logan, and S. Short. 2014. Integrating space with place in health research: A multilevel spatial investigation using child mortality in 1880 Newark, New Jersey. *Demography* 51 (3):811–34.