

CHAPTER FIVE Learning from Place in the Era of Geolocation

Ryan T. Moore and Andrew Reeves

IN THIS CHAPTER

In this chapter, we give an overview of the ways in which scholars and policy makers are currently using individual geolocation data. We also describe some new analytic and service-provision possibilities that geolocation data enable. We discuss several challenges inherent in obtaining geolocation data and making those data useful for social, political, and policy research and practice. We highlight the unique concerns over privacy that arise in this rich data environment and note some promising approaches for addressing them.

INTRODUCTION

For centuries, policy makers and social scientists have used geographic location data to improve services, solve vexing social problems, and learn more about how we affect and are affected by our environments. John Snow's investigation of the 19th-century London cholera epidemic, for example, relied on detailed knowledge of the locations of water suppliers, contaminants, and illness cases.¹ However, only recently has it become possible and inexpensive to collect fine, frequent, and automatic individual-level geolocation data for social and political research and practice.

The growth of smartphone adoption among Americans has expedited this new research possibility. The number of Americans with smartphones has grown rapidly in just the last few years. According to the Pew Research Center, in January 2014, 58 percent of American adults had a smartphone, up from 32 percent in May 2011.² With the widespread adoption of smartphones that include global positioning system (GPS) receivers, the collection of individuals' geolocation data is revolutionizing analysts' ability to understand how people interact with their social and physical environments.

In this chapter, we give an overview of the ways in which scholars and policy makers are currently using individual geolocation data, as well as the new analytic and service-provision possibilities that geolocation data enable. Researchers' ability

to characterize and assess individuals' contexts of various sorts is vastly richer now than in the past, when "county of residence" was the state of the art in describing individuals' experiences. Similarly, governments' ability to diagnose and remedy problems and to evaluate how citizens use public spaces are fundamentally different now that feedback from constituents can be instant, accurate, and sometimes even automatic. We discuss several challenges inherent in obtaining and analyzing geolocation data, and we highlight concerns over privacy in this rich data environment.

APPLICATIONS OF GEOLOCATION DATA

Researchers, entrepreneurs, and governments have all taken advantage of rapid advancements in geolocation technology to achieve a number of ends. We highlight three broad areas to which analysts bring geolocation to bear. First, we discuss how scholars use these tools to understand the social milieus that individuals encounter in their daily lives. Geolocation data allow us to determine the social characteristics of the places that individuals actually experience, shedding new light on both social networks and contextual experiences. Second, we discuss how geolocation data and smartphone applications allow governments new insight and efficiencies in delivering public goods. By allowing individuals to more easily and quickly report public works needs, for example, local governments can better respond to breakdowns in public infrastructure. These tools often go hand in hand with improved reporting of public works project data to the public. Increased transparency, then, helps constituents hold government accountable. Third, geolocation tools assist in understanding how individuals interact with their natural and social environments. For urban planners and local governments, these data provide direction about maximizing investment in parks, bike trails, and running paths. For researchers in public health and health administration, these data provide insights that can be leveraged to promote preventive care, track disease contagion, and eliminate health insurance fraud.

Learning About Geographic Contexts

Scholars have long been interested in how geographically defined contexts shape social and political behaviors. Early work focused on racial concentrations and voter turnout.³ More recent work has continued in a similar vein, focusing on the individual and social conditions under which contexts affect attitudes about racial

and ethnic groups and racially targeted policies⁴ and generalized trust.⁵ Political scientist V. O. Key, in particular, used county-level estimates of the percent black—a sensible choice in an era when geographic mobility was significantly more restricted than it is today.⁶ An individual's county of residence in the 1940s could circumscribe much of his or her life's experience. However, over the course of the 20th century, the average American went from a limited, slowly changing set of environments to traveling more than 13,000 miles per year, by car alone.⁷

Though scholars seek the characteristics of an individual's experience, they typically rely on the characteristics of a predetermined geographic container such as the state, county, or census tract in which the subject resides. Once the geographic unit of analysis is selected, its characteristics are assumed to be the experiences of the individual. The container may be small, like a census block group, or expansive, like a state. Small units of analysis arbitrarily assume individuals' experiences may be limited to a few city blocks, while large containers assume an individual experiences every corner of a state, for example. Despite the vast differences in these measures, studies may pick either container to try to capture the same individual's contextual experience.

The fact that results can depend on which level of geography an analyst chooses is known as the "modifiable areal unit problem" (MAUP). To explicate the degree to which higher levels of geography may misrepresent lower ones in the United States, we focus on a particular state. In figure 5.1, each county in Maryland is represented by a panel, and the density shows the distribution of the proportion of non-white residents within that county's census blocks. The counties are sorted by their overall proportion non-white (represented by each panel's vertical line). Since an individual in a census block encounters more people of a given demography when the census block is more populous, we weight the census blocks by population.

In some cases, the aggregate county measure represents most of the census blocks that comprise it. For example, in the upper left-hand corner, Garrett County has a small percentage of non-whites, and the same is true of its constituent census blocks. Similarly for Carroll County, where the overall fraction of non-whites is 9 percent, virtually all census blocks are between 0 and 25 percent non-white.

However, where there is variation in the census block measures within a county, the county mean may poorly summarize the distribution. For example, Baltimore County is about 37 percent non-white, but the modal block is significantly less diverse, only around 10 percent non-white. Many blocks (about 27 percent) in Baltimore County are more than half non-white, as well. Even worse, in Somerset

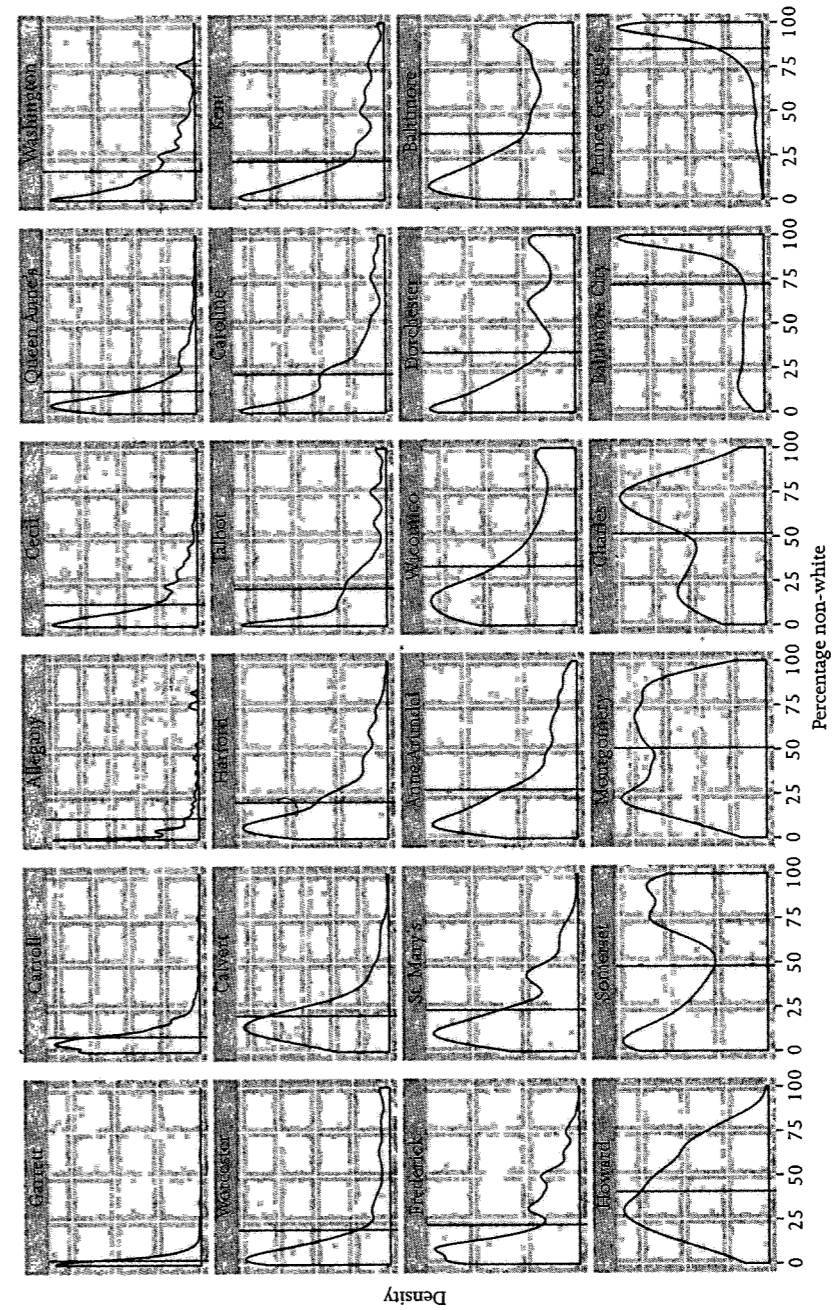


Figure 5.1. Distributions of non-white populations in census blocks by county in Maryland. Notes: Distributions are weighted by block population. County means are displayed as vertical lines. County means can describe block-level distributions well (e.g., Garrett County) or poorly (e.g., Somerset County).

County, blocks tend to be either extremely white, as represented by the mode to the left of the distribution, or highly non-white, as represented by the mode to the right. Though the average would summarize Somerset County as 50 percent non-white, this summary is not indicative of the concentration of relatively homogeneous blocks in the county.

Once the researchers select a geographic unit of analysis, they typically assume that individuals' experiences reflect only characteristics of that geographic container. Despite this assumption, an individual may spend a majority of her time in other geographies that are distinct from her place of residence. The standard practice is to measure context as a single location rather than as a dynamic experience across a number of locales over a period of time. However, we argue that this need not be the case in an era of rich geolocation data.

In other research, we seek to overcome these problems via geolocation data by introducing individual-level, automatically recorded location data that allow us to describe individuals' experiences more richly, more dynamically, and less constrained by single-geography measures.⁸ This allows us to compare individuals' dynamic experiences to traditional scholarly proxies for those experiences. We estimate milieus for hundreds of individuals and show that these precise, dynamic measures reveal different contextual experiences than those based on the coarse, static proxies that scholars have, up to this point, relied on. In particular, our research shows that previous static measures overstate how extreme racial experiences are for most individuals. Ultimately, our new measures of contextual experience will allow us to understand how place shapes both attitudes and behaviors.

Understanding and Improving the Provision of Public Goods

One of the central roles of local government is to provide public goods in an efficient way that maximizes the benefit to residents. However, a local government may lack information about areas most in need of resources or the exact locations of breakdowns in public infrastructure. Smartphone applications that record geolocation data are transforming the accuracy, speed, and quantity of information that local governments have access to as they make decisions about the allocation of scarce resources to support public goods. Additionally, the delivery of these goods are often publicly posted under open government initiatives, which allow citizens, journalists, and researchers to analyze successes and failures of service delivery.

For example, many local governments have developed smartphone applications to supplement 3-1-1 services, non-emergency telephone numbers whereby citizens may report public works issues or request municipal services. Cities such as Philadelphia, Boston, New York, and San Francisco supplement their 3-1-1 service with smartphone applications. These apps allow residents to report noise complaints, rat sightings, lost items in taxis, illegal parking, broken parking meters, poor apartment conditions, graffiti, streetlight repair, and potholes, among other issues. These apps often allow users to take photographs and automatically record the location of nuisances and then relay that information directly to authorities.

In one innovative example, the city of Boston has made reporting road conditions even easier with its Street Bump application, produced by the Mayor's Office of New Urban Mechanics. This app "collects data about the smoothness of the ride" passively, eliminating the requirement for constituents themselves to report a nuisance.⁹ By at least some reports, this new technology is allowing the city to do a better job in filling potholes. The mayor of Boston has touted that the city filled 50 percent more potholes in 2014 than in 2013.¹⁰ The data are also providing new insights into other road conditions and safety hazards that inconvenience drivers. One of the insights gleaned from the app was that manhole and other grate covers accounted for "about eight times as many bumps and divots in the road as plain old potholes"; this led the city of Boston to pressure utility companies to correct the misaligned covers.¹¹ Prior to this sort of technology, residents could call 3-1-1 and report these nuisances, but authorities would have to rely on residents having a phone nearby, accurately reporting the location of the nuisance, and describing it faithfully. Local government 3-1-1 apps make the reporting process quick, detailed, and accurate.

Complementing the work of governments, firms have capitalized on the same sort of automated reporting to provide users with real-time transit information. The smartphone application Waze records users' routes, locations, and times to their destinations to calculate traffic speeds and immediately report conditions to its user base.¹² Automatically recording, processing, and delivering vast amounts of geolocation data led to Waze being acquired by Google in 2003 for over \$1 billion.

Other firms have innovated by developing new methods to analyze geolocation data to improve understanding of traffic, commerce, and other human interactions. For instance, StreetLight Data purchases anonymized cell phone data and, based on its analyses, advises companies where to locate stores,¹³ state departments of transportation where to build roads,¹⁴ and local communities how to woo businesses to build in underdeveloped but heavily traveled parts of their city.¹⁵ As in many other

contexts where big data are available, there are numerous challenges to analyzing the data to gain insights into the questions that local governments, businesses, and individuals are asking.

For governments, these services create opportunities to more efficiently deliver public resources to citizens. They also create vast new datasets that allow researchers to better understand the classic political questions of "who gets what, when, how."¹⁶ Do politicians reward particular subsets of voters with distributive resources? Are they more likely to do so near election time? These are questions that scholars have examined by considering, for example, the distribution of federal resources across counties,¹⁷ congressional districts,¹⁸ and states.¹⁹ In one such study, researchers found that county-level presidential incumbent vote share increases as a function of county-level federal spending.²⁰ As local governments collect more fine-grained data on the provision of public goods and especially the public demand for such goods, the findings of these studies stand to be subjected to new, rigorous tests.

Understanding Interactions with Social and Physical Environments

Related to understanding the provision of public goods, geolocation data also help us understand how individuals interact directly with their social and physical environments. Knowing how individuals interact with public green space, public transit, or bicycle paths can aid urban planners as they design cities that maximize public health and happiness while reducing congestion, air pollution, and other urban blights. In the health care sector, knowing how individuals tend to interact with others and with nearby providers can be used to detect disease outbreaks and insurance fraud and to encourage health system utilization at critical junctures such as childbirth.

For example, one study examines how economic characteristics within an urban area influence residents' patterns of exercise.²¹ As a report from the National Park Service notes, "Parks and protected areas have long been recognized as important resources for public health."²² This study uses freely available public data from MapMyRun.com, a smartphone application that records an individual's jogging speeds and location through GPS. By analyzing the factors associated with the use of public green space for running, jogging, and walking, this study aids public health researchers in understanding the relationship between socioeconomic context and the healthy use of urban space. At the same time, it provides information to such groups as the Open Space Council for the St. Louis Region, a non-profit conservation

group and one of the funders of the study, in its mission to "maintain the integrity of land and waterways for practical purposes, recreation and their natural beauty."²³

Another study examines whether being in nature improves levels of happiness.²⁴ In a sample of over 1.1 million responses from nearly 22,000 respondents, this study analyzed how the environment, including things like weather, land cover type, and whom the participants were with, influenced their subjective well-being in the moment. Using Mappiness, an iPhone app the researchers developed, they periodically surveyed respondents about their happiness and also recorded their geolocation and other observations about their environment. This study's data analysis lends credence to the assertion of the National Park Service's report described earlier: individuals reported being significantly happier when they were outdoors in nature. Previous studies had been limited to correlating green space around an individual's home with health factors.²⁵

Governments and non-governmental organizations also use geolocation data in efforts to improve health and health systems. For example, to target areas for possible follow-up intervention, researchers dynamically geolocated concentrations of men seeking sex with other men through a social network mobile phone application.²⁶ They demonstrate how the profile features of app users (such as their HIV status) can be used to identify neighborhoods for recruitment into research studies or provision of treatment clinics. Where there are more users from HIV-positive or vulnerable populations, researchers and clinicians can better focus their resources. Similarly, to describe where particular health conditions are becoming of local interest, the U.S. Department of Health and Human Services provides heat maps of geolocated Twitter posts about 30 different conditions.²⁷ These maps can help identify where disease outbreaks may be occurring and thus where treatment and research resources are best allocated.

Static geographic container data can be used along the same lines when policy makers and researchers exploit dynamic relationships between fixed points. For example, knowing where providers' offices are located and where their patients live can help detect suspicious billing behavior. Comparing the distances between individuals and their health providers can suggest instances of fraud to policy makers. Researchers have developed a method for detecting Medicare fraud by identifying providers who derive an unusually large proportion of their earnings from patients who have traveled unusually far.²⁸

Geolocation data can also be used to understand the scope and limits of programs to encourage vulnerable populations to use a health system. A study by ORB