Learning from Place in the Era of Geolocation

Ryan T. Moore* Andrew Reeves†

forthcoming, *Analytics, Policy and Governance* (YUP, 2017), eds. Benjamin Ginsberg, Kathy Wagner Hill, and Jennifer Bachner

Abstract

In this chapter, we give an overview of the ways that scholars and policymakers are currently using individual geolocation data. We also describe some new analytic and service-provision possibilities that geolocation data enables. We discuss several challenges inherent in obtaining geolocation data and making that 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.

*Department of Government, American University, Ward Circle Building 226, 4400 Massachusetts Avenue NW, Washington DC 20016-8130. tel: 202.885.6470; fax: 801.922.7901; `rtm` (at) `american` (dot) `edu`; `http://ryantmoore.com`.

†Associate Professor, Department of Political Science Washington University in St. Louis, Campus Box 1063, One Brookings Drive, St. Louis MO 63130, http://www.andrewreeves.org, reeves@wustl.edu, (314) 935-8427.

1 Introduction

For centuries, policymakers and social scientists have utilized 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 Pew Research, by January 2014, 58 percent of American adults had a smartphone, up from 32 percent in May of 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 that scholars and policymakers are currently using individual geolocation data, as well as the new analytic and service-provision possibilities that geolocation data enables. 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 utilize 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.

2 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 questions 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.

2.1 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 policies[4] and generalized trust[5]. Key (1949), in particular, utilized county-level estimates of the percent black—a sensible choice in an era when geographic mobility was significantly more restricted than it is today. An individual's county of residence in the 1940's could circumscribe much of his or her life's experience. However, over the course of the twentieth 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[6].

Though scholars seek the characteristics of an individual's experience, they typically rely on the characteristics of a predetermined geographic container like the state, county, or census tract where 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 very 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 analysts choses 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 1, each county of 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 very 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%, virtually all census blocks are between 0 and 25% nonwhite.

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

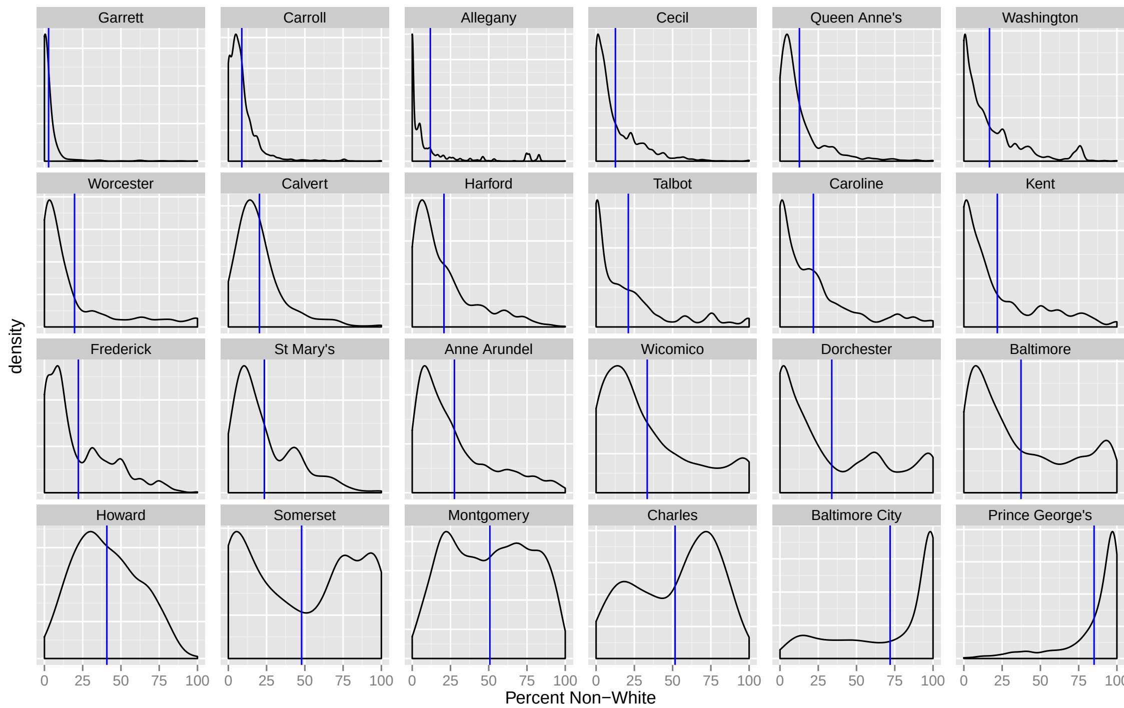

Figure 1: Distributions of Non-White Populations in Census Blocks by County in Maryland. Distributions weighted by block population. County means displayed as vertical lines. County means can describe block-level distributions well (e.g., Garrett County) or poorly (e.g., Somerset County).

of residence. The standard practice is to measure context as a single location, rather than the 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.

Beyond the MAUP, the context that influences a certain attitude or behavior may not correspond to a single, static container (regardless of level). Furthermore, its influence may depend on how long or exactly when an individual is in that context. These phenomena combine to create the *uncertain geographic context problem*[7].

In other research, we seek to overcome these problems via geolocation data by introducing individual-level, automatically recorded location data that allows 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 upon. 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.

2.2 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 where 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 fifty percent more potholes in 2014 than in 2013[10]. The data are also providing new insights into other factors that ail 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," and caused 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 their 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 travelled 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, Kriner and Reeves (2012) finds that county-level presidential incumbent vote share increases as a function of county-level federal spending. 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.

2.3 Understanding Interactions with Social and Physical Environments

Related to understanding the provision of public goods, geolocation data also helps 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 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 influences residents' patterns of exercise[20]. As a report from the National Park Service notes, "Parks and protected areas have long been recognized as important resources for public health."[21] This study utilizes freely available public data from MapMyRun.com, a smartphone application which records an individual's jogging speeds and location through GPS. Analyzing the factors associated with the utilization 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 their mission to "maintain the integrity of land and waterways for practical purposes, recreation and their natural beauty"[22].

Another study examines whether being in nature improves levels of happiness[23]. 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 participant was with, influenced their subjective wellbeing 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 above: 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[24].

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, Delaney et al. (2014) dynamically geolocate concentrations of men seeking sex with other men through a social network mobile phone application. 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 heatmaps of geolocated Twitter posts about thirty different conditions[25]. These maps can help identify where disease outbreaks may be occurring, and thus where treatment and research resources best allocated.

Static geographic container data can be used along the same lines when policymakers 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 policymakers. Musal (2010) demonstrates a method for detecting Medicare fraud by identifying providers who have unusually large fractions of their earnings from patients who have traveled unusually far.

Geolocation data can also be used to understand the scope and limits of programs to encourage vulnerable populations to utilize a health system. A study by ORB International used survey enumerators' GPS information to determine the distance from respondent households to the nearest health facility. Women near program facilities were more likely to give birth in a facility, when compared to those near non-program facilities, but the difference disappeared after a distance of about 10km. Beyond this, women were unlikely to give birth in any type of facility[26]. This household-level geolocation data can be incorporated into policymakers' plans for new clinic siting both inside and outside the program.

3 Challenges in Geolocated Data

Accompanying the wealth of opportunities that individual smartphone geolocations provide are significant challenges and decisions that must be faced before researchers can collect or utilize the data. Below, we focus on how to obtain geolocated data and how to make it useful by linking it to other sources.

3.1 Collecting Geolocated Data

Prior to the explosion of mobile smartphone usage, recording geolocations required technologies with more severe limitations. Time diaries, for example, require active, accurate participation of respondents, but respondent reports suffer from a variety of biases and inadequacies[27]. Now, even in "resource-poor" areas, a variety of relatively inexpensive commercially available GPS-logging devices can accurately record individual geolocations and weigh under 100 grams[28]. In more resource-abundant areas, geolocation data collection is even easier. With widespread adoption of mobile phones that include GPS antennas, researchers and policymakers can access data from inexpensive, lightweight devices that individuals naturally carry with them over the course of their daily lives.

These devices use a combination of sensors to locate users in space. Mobile phones can determine user locations with a mixture of direct location sensors (GPS antennas) and other functional sensors that can interact with bases that can be traced to specific locations. For example, cellular antennas connect to specific cell towers during a call, WiFi antennas connect to specific access points, and Bluetooth sensors can connect to beacons whose locations are known. Different applications use different mixtures of sensors and triggers for those sensors to collect and distribute geographic data. Montoliu, Blom, and Gatica-Perez (2013) describe a strategy for efficiently combining these data collection modes in an application.

At one extreme are applications that show the user on a map and provide turn-by-turn directions to the next destination. These applications rely on GPS antennas, and update the user location in real time to provide instructions that are useful in a moment. These applications tax the device's battery heavily, resulting in less usable time between recharges. On the other hand, applications specific to retail outlets (such as the Apple Store app, the Safeway grocery store app, or the American Eagle clothing store app) can use Bluetooth low energy signals to prompt only users who are in the store to visit certain aisles. Governments and public facilities can use these beacons to guide museum visitors to specific exhibits, to provide information when a visitor is in a certain part of a national park, or to make airports easier to navigate for the blind. Such applications preserve battery life, but only provide feedback or geolocated data for very specific places.

Between these two extremes are applications that are interested in locations beyond a single cultural site or retail outlet, but that want to preserve battery life to collect geolocations throughout the user's daily activities. Examples of such applications rely predominately on cell tower networks to determine user location, but then engage the GPS antenna when the device detects "significant movement".

Once researchers or governments decide to adopt mobile technology to provide information to or get feedback from individuals, they must decide whether to pursue a *native* application or a *mobile web* application. Native smartphone applications are written in the language that the operating system requires; third-party native applications are downloaded to the phone by the user. Mobile web applications work directly in a web browser, so programmers can deploy the application to any phone with a browser without having to rewrite it for iOS, Android, Windows Phone, or other platforms. In a political science example, a web application has been used to encourage engagement in and collect feedback from candidate debates[29].

Although native applications thus require more infrastructure in order to be available across operating systems, native applications "provide a richer, more compelling experience with a more responsive interface and superior interaction"[30]. Native applications can run at native speeds and can access the full functionality of the smartphone, including the camera, microphone, Bluetooth connectivity for a variety of external sensors, and the GPS antenna for geolocation[31]. Governments exploit these functionalities to find out, for example, exactly where a user-reported pothole is located and even what it looks like—users can send photographs of the nuisance directly to officials through native apps. Kuntsche and Labhart (2013) provide a typology and list of example strategies for the spectrum of automatic, manual, and hybrid data collection approaches in native applications.

Although mobile web applications are easier to distribute than native ones, centralized online software application markets like the Apple App Store and Google Play store make distribution of native apps relatively easy. At the federal level, native applications appear more common than web applications. The federal government's mobile application registry lists more than 150 registered native federal iOS apps, but fewer than half that many "mobile web" apps[32].

3.2 Linking Geolocations to Other Data

In social science and governance applications, researchers want to link individual geolocations to other data. Geolocations alone can provide a sense of how mobile individuals are, but many other interesting substantive questions will remain elusive. In particular, we see two significant linking challenges of interest to social scientists and policymakers: linking to aggregate geographic contextual data and linking to individual-level background data and survey responses.

First, researchers will often want to link geolocations to aggregate geographic contextual data, such as measures of neighborhood racial and ethnic composition from the Census. For each observation of a single individual (usually a latitude, longitude, and time triple), this process starts by taking the geographic coordinates and determining which census block that point is in. This can involve first reading in census block shapefiles, then locating the block code for each point. Next, the researcher retrieves the aggregate contextual measures for that block. This process can be computationally intensive, since the number of aggregate units and observations can both be large. In a moderately sized sample of about 450 individuals, for example, we observe about 2.6 million coordinate triples, each of which we then place within one of the roughly 8.2 million American census blocks. The process must be repeated for each geographic level of interest, since each geography will have its own shapefiles and, in some cases, files of aggregate measures.

Second, augmenting individual geolocations with individual-level data allows researchers to answer substantive questions such as "do particular racial contexts affect individuals' attitudes about race?" and "do men and women have different degrees of access to urban opportunities?" To examine these individual-level questions, we developed an original application that records both geolocations and responses to survey questions[33]. The survey questions are served directly within the application itself, and responses can themselves be geolocated. To enable future researchers to take advantage of this new platform, the infrastructure is quite general, and survey content can be determined by other interested researchers. Surveys can be administered to selected samples, including geographically determined ones. Smartphone geolocation data are an example of passive, automatic measurement of social quantities of interest. Instead of relying on respondent self-reports, passive, automatic measurement can better capture social behaviors. In a typical older approach, Kwan (1999) tackles the question of gendered access to urban opportunities by pairing time diary reports with a demographic survey including gender, age, income, employment status, and automobile usage.

Individual survey responses can also help validate approaches for estimating people's activities at the locations they inhabit each day. When an individual regularly spends a lot of time at geolocations very near each other, clustering algorithms can identify his or her home and work locations quite accurately, for example. Montoliu, Blom, and Gatica-Perez (2013) validate their algorithmically-discovered locations by asking participants to label them by hand in a web application.

4 Privacy and Ethics

Geolocated data pose special problems for privacy and raise new ethical concerns for researchers and policymakers alike[34]. One source of these privacy concerns is the ability of phones to detect location with a high degree of accuracy. For example, iOS application developers can configure their software to record geolocations to within ten meters or so[35]. Of course, this accuracy is precisely what makes mobile phone geolocations valuable—if local government receives a report that there is a pothole within one kilometer of a site, this information is not likely to be useful. On the other hand, iOS guidelines recommend developers use "significant change" location services, which only respond when the user's position changes by 500 meters or so. Coupled with precise timestamps, extremely accurate geolocations could identify where individuals are in real time.

As developers and consumers of technologies that utilize geolocation data, it is important to consider best practices for preserving the privacy of individual users while providing the insights that the local governments, researchers, or businesses desire. While deidentified geolocation data can be purchased directly from wireless carriers without the direct consent of the user[36], many entities collect these data via mobile applications. In these cases it is important that the developer of the app provide transparency as to the type of data that will be collected, how it will be used, and how the users can terminate their participation in the study. Developers who must seek approval from institutional review boards, such as researchers at universities, may receive guidance on these questions. But all too often, these entities are ill-equipped to deal with the new challenges of emerging technologies and their oversight may do more harm then good[37].

In addition to specifying the type of data that will be collected, developers should collect only the data they need and no more than that. As discussed above, software protocols allow developers to define the precision of the GPS coordinates that are recorded, and developers should collect it at the coarsest level possible for their research purposes. Unless users consent, geolocation data should not be paired with identifying information like names, email addresses, or phone numbers. For example, our Milieu app, which combines survey data with geolocation data, provides users with a randomly generated user ID. This ID links a single (anonymous) user to geolocation data and to answers to survey questions delivered through the app. While this strategy means that users of the app get a new ID and appear as a new user if they reinstall the app, it further (and desirably) limits the developers' ability to identify the user. A model resting on user accounts identified by email addresses, for example, could ease direct identification of users.

Still, study after study shows that individuals are easily identified in ostensibly "deidentified" data with far less than information than geolocation data. For instance, one study using the 1990 U.S. Census found that 87% of the United States could be uniquely identified based on their ZIP code, gender, and date of birth[38]. In an "anonymized" dataset of credit card transactions in 10,000 stores over three months, researchers identified 90% the of 1.1 million people[39]. While we argue every effort should be made to not collect identifying information, identifying "deidentified" individuals is easy without geolocation data and somewhat easier with it. For this reason, we argue that transparency with the user is especially important. Additionally, other protections may be employed such as inducing random noise into the data at the time of collection or employing a third party to act as a firewall between researchers and the individual location records.

The regularities in human mobility can, however, further facilitate identifying where an individual is at a given moment. Roughly 90% of people in a Swiss study tended to visit two to four locations on an average day, and novel places tended to be visited in well-defined patterns – during meal times and on Saturdays, for example[40]. Further, researchers were able to label frequently-visited locations with greater than 80% true positive rates.

Perhaps just as relevant for user identification, though, extremely accurate timestamps and geolocations are not required to uniquely characterize individual travels through environments. Geolocated data is surprisingly unique, even when coarsened. Even when characterized by just a handful of latitude, longitude, and time coordinate triples, the paths users take through their environments tend not to be shared with others[41]. This finding, apparently inherent in human mobility, poses challenges to anonymity and deidentification of a different type than simple demographics, say. Indeed, some 30% of smartphone users recently reported that they have "[t]urned off the location tracking feature on your cell phone because you were worried about other people or companies being able to access that information"[42]. Several intellectual initiatives have sprung up in response to recent data privacy concerns, including the American Statistical Association's Committee on Privacy and Confidentiality, the Data Privacy Lab at Harvard University's Institute for Quantitative Social Science, and the resources on geospatial confidentiality provided by NASA's Socioeconomic Data and Applications Center.

This uniqueness of individual geolocation data runs counter to recent social science trends regarding data openness and replication. In particular, scholars highlight the value of registration to counteract fishing for statistically significant results[43], including in observational studies[44]. Others highlight the value in replicating others' findings[45] and in sharing and archiving data[46]. We support these trends and actively provide our own materials in public archives. However, since one can uniquely characterize mobility patterns with relatively few, relatively coarse observations, publicly-shared replication data with geolocations must take special precautions to avoid reidentification, especially if the data are adjoined to static demographic measures. Even fine geolocated data can be shared publicly, however, if proper precautions are taken[47]. For example, researchers can release several datasets in which the individuals' geographic positions are multiply imputed, rather than releasing the original values[48]. Such an approach can preserve correlations between all the measures, geographic and non-geographic, without requiring the release of any true geolocations[49]. In political science, one can find several applications of overwriting observed data with multiply-imputed values[50].

One model for the future of sensitive data such as geolocations seeks to eliminate redundancies and to put control in the hands of each user. In an "openPDS" (for "open personal data store") infrastructure model, each individual amasses his or her own data in a personal data storage site. The individual then controls who can access his or her data study-by-study. Separate researchers do not need to duplicate data collection; they simply need to ask the user's permission for access to a part of the data[51].

OpenPaths, a geolocation project of the Research and Development Lab of the New York Times, represents one such data store. OpenPaths bills itself as a "secure data locker for personal location information"[52]. This project allows individuals who store their geolocation data with OpenPaths to receive invitations to participate in artistic, academic, and commercial projects. The user decides on a case-by-case basis whether to allow the project access to his or her data. Projects do not need to recollect the individual's geolocation data, creating a significant research efficiency. Each project must, however, have each user's explicit consent. Elsewhere, we employ OpenPaths geolocation data to redefine racial and ethnic geographic contexts[53].

5 Discussion

Governments at all levels have begun to embrace mobile computing. From local councils to federal agencies, governments have released mobile applications to facilitate two-way communication with constituents and visitors. On one hand, governments use smartphone applications to provide users with information, ranging from precinct-level real-time election results to features of local tourist sites (in Jefferson Parish, LA and Sparks, NV, respectively). On the other, governments encourage users to provide immediate, geolocated, and often visual feedback on local infrastructure. Citizens can report rat sightings, potholes, and street sign repairs to local governments via the burgeoning category of 3-1-1 smartphone applications. These two primary functionalities are complemented by the suite of technical guides packaged as applications, such as the federal Terrestrial Mollusc Key app, "designed for federal, state and other agencies or organizations within the U.S. that are concerned with the detection and identification of molluscs of [quarantine] significance"[54].

Academic researchers have also begun to utilize these new tools to answer long-standing difficult questions in the social sciences. When researchers posit theoretical mechanisms that involve individuals' experiences rather than simply the location of their residences or workplaces, rich geolocation data can and should be brought to bear. We can better reflect the modern realities of human mobility by measuring social and political contexts in dynamic ways; we are no longer tied to "county of residence" to describe people's experiences.

Of course, geolocations collected by policymakers to improve service delivery can inform important social scientific questions about the distribution of resources and the impact of political conditions. Similarly, as social scientists pair geolocations with social survey data, we see significant opportunities to improve policymakers' responsiveness to constituents' needs and to further deepen campaigns' knowledge of whom to try to turnout or persuade in an election.

The realities of modern geolocated data suggest that deidentification and anonymization strategies that were appropriate for long, in-person interviews in people's homes may no longer suffice to ensure respondent privacy. New statistical and technical solutions can help, but preserving both scientific value and individual privacy will be an ongoing challenge.

6 For Further Reading

- Do, Trinh Minh Tri, and Daniel Gatica-Perez. 2014. "The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data." IEEE Transactions on Mobile Computing 13(3): 638–648.
- Gutmann, Myron P., and Paul C. Stern, eds. 2007. Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data. Washington, DC: National Academies Press.

- Kuntsche, Emmanuel, and Florian Labhart. 2013. "Using Personal Cell Phones for Ecological Momentary Assessment: An Overview of Current Developments." European Psychologist 18(1): 3–11.
- Kwan, Mei-Po. 2012. "The Uncertain Geographic Context Problem." Annals of the Association of American Geographers 102(5): 958–968.
- Moore, Ryan T., and Andrew Reeves. 2015. "Defining Racial and Ethnic Context with Geolocation Data." Manuscript.

Notes

[1](Freedman, 1991, describes the case in detail.)

[2]See http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/ and http://www.pewinternet.org/data-trend/mobile/device-ownership/ for details, accessed 30 October 2014.

[3]Key (1949).

[4]Hopkins (2010); Oliver and Mendelberg (2000).

[5]Stolle, Soroka, and Johnston (2008).

[6]See http://www.fhwa.dot.gov/ohim/onh00/bar8.htm, accessed 9 June 2015.

[7]Kwan (2012).

[8]Moore and Reeves (2015).

[9]http://www.cityofboston.gov/DoIT/apps/streetbump.asp.

[10]Andrew Ryan. April 13, 2005. “City touts pothole numbers, but what exactly qualifies?” *The Boston Globe*, https://www.bostonglobe.com/metro/2015/04/12/potholes-count-does-counting-them/eP0jqKHac7wz0lGIU0sgYM/story.html.

[11]Eric Moskowitz, December 16, 2012, “App shows jarring role of cast-metal covers in Boston,” *The Boston Globe*, https://www.bostonglobe.com/metro/2012/12/16/pothole/2iNCJ05M15vmr4aGHACNgP/story.html.

[12]See https://support.google.com/waze/answer/6078702?hl=en for details.

[13]http://www.streetlightdata.com/case-studies/.

[14]Max Smith, May 21, 2015, “Tracking data aims to speed commutes in Virginia,” *WTOP.com*, http://wtop.com/virginia/2015/05/tracking-data-aims-to-speed-commutes-in-virginia/.

[15]Alex Salkever, Sept. 2, 2014,“How Big Data Reveals The Secret Life Of Cities,” *readwrite*, http://readwrite.com/2014/09/02/streetlight-data-laura-schewel-big-data-tracking-surveillance.

[16]Laswell (1936).

[17]Kriner and Reeves (2015); Gasper and Reeves (2011).

[18]Berry, Burden, and Howell (2010).

[19]Reeves (2011).

[20]Adlakha et al. (2014).

[21]http://www.nps.gov/public_health/hp/hphp/press/HPHP_Science%20Plan_accessible%20version.final.23.july.2013.pdf.

[22]http://www.openspacestl.org/missionabout-us.

[23]MacKerron and Mourato (2013).

[24]Mass et al. (2009).

[25]See http://nowtrending.hhs.gov/heatmap, accessed 16 June 2015.

[26]http://www.opinion.co.uk/article.php?s=using-gis-mapping-in-aid-evaluation, accessed 18 June 2015.

[27]Stopher, Bullock, and Horst (2002).

[28]Vazquez-Prokopec et al. (2009).

[29]Boydstun et al. (2014).

[30](Serrano, Hernantes, and Gallardo, 2013, page 24).

[31]Schobel et al. (2013).

[32]See http://www.usa.gov/mobileapps.shtml, accessed 27 October 2014.

[33]**?**

[34]For an overview of the challenges to civil liberties posted by digital technology see Richards (2015).

[35]See http://j.mp/1dB3C5w for technical specifications.

[36]Russell Brandom, April 9, 2013, “Can you find me now? How carriers sell your location and get away with it,” *The Verge*, http://www.theverge.com/2013/4/9/4187654/how-carriers-sell-your-location-and-get-away-with-it.

[37]Schneider (2015)

[38]Sweeney (2000).

[39]de Montjoye et al. (2015).

[40]Do and Gatica-Perez (2014).

[41]de Montjoye et al. (2013).

[42]Boyles, Smith, and Madden (2012), pg. 16.

[43]Humphreys, de la Sierra, and van der Windt (2013).

[44]Monogan (2013).

[45]King (1995).

[46]King (2007).

[47] Gutmann and Stern (2007).
[48] Wang and Reiter (2012).
[49] Little and Rubin (2002); Schafer (1997).
[50] Blackwell, Honaker, and King (2015).
[51] Tarran (2014).
[52] See `https://openpaths.cc/about`, accessed 31 October 2014.
[53] Moore and Reeves (2015).
[54] Available at `http://j.mp/1wvgobw`, accessed 27 October 2014.

References

Adlakha, Deepti, Elizabeth L. Budd, Rebecca Gernes, Sonia Sequeira, and James A. Hipp. 2014. "Use of emerging technologies to assess differences in outdoor physical activity in St. Louis, Missouri." *Frontiers in Public Health* 2: 1–8.

Berry, Christopher R., Barry C. Burden, and William G. Howell. 2010. "The President and the Distribution of Federal Spending." *American Political Science Review* 104(4): 783–799.

Blackwell, Matthew, James Honaker, and Gary King. 2015. "A Unified Approach to Measurement Error and Missing Data: Overview." *Sociological Methods & Research* 44(1): XX–XX.

Boydstun, Amber E., Jessica T. Feezell, Rebecca A. Glazier, Timothy P. Jurka, and Matthew T. Pietryka. 2014. "Colleague Crowdsourcing: A Method for Fostering National Student Engagement and Large-N Data Collection." *PS: Political Science & Politics* 47(4): 829–834.

Boyles, Jan Lauren, Aaron Smith, and Mary Madden. 2012. "Privacy and Data Management on Mobile Devices." Report of the Pew Internet & American Life Project.

de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The Privacy Bounds of Human Mobility." *Scientific Reports* 3: 1–5.

de Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex "Sandy" Pentland. 2015. "Unique in the shopping mall: On the reidentifiability of credit card metadata." *Science* 347: 536–539.

Delaney, Kevin P., Michael R. Kramer, Lance A. Waller, W. Dana Flanders, and Patrick S. Sullivan. 2014. "Using a Geolocation Social Networking Application to Calculate the Population Density of Sex-Seeking Gay Men for Research and Prevention Services." *Journal of Medical Internet Research* 16(11): e249.

Do, Trinh Minh Tri, and Daniel Gatica-Perez. 2014. "The Places of Our Lives: Visiting Patterns and Automatic Labeling from Longitudinal Smartphone Data." *IEEE Transactions on Mobile Computing* 13(3): 638–648.

Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21: 291–313.

Gasper, John T., and Andrew Reeves. 2011. "Make it Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters." *American Journal of Political Science* 55(2): 340–355.

Gutmann, Myron P., and Paul C. Stern, eds. 2007. *Putting People on the Map: Protecting Confidentiality with Linked Social-Spatial Data.* Washington, DC: National Academies Press.

Hopkins, Daniel J. 2010. "Politicized Places: Explaining Where and When Immigrants Provoke Local Opposition." *American Political Science Review* 104(February): 40–60.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21(1): 1–20.

Key, Valdimer Orlando. 1949. *Southern Politics in State and Nation.* Knoxville: University of Tennessee Press.

King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics* 28(3): 444–452.

King, Gary. 2007. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." *Sociological Methods & Research* 36(2): 173–199.

Kriner, Douglas L., and Andrew Reeves. 2012. "The Influence of Federal Spending on Presidential Elections." *American Political Science Review* 106(2): 348–366.

Kriner, Douglas L., and Andrew Reeves. 2015. "Presidential Particularism and Divide-the-Dollar Politics." *American Political Science Review* 109(1): 155–171.

Kuntsche, Emmanuel, and Florian Labhart. 2013. "Using Personal Cell Phones for Ecological Momentary Assessment: An Overview of Current Developments." *European Psychologist* 18(1): 3–11.

Kwan, Mei-Po. 1999. "Gender and Individual Access to Urban Opportunities: A Study Using Space-Time Measures." *The Professional Geographer* 51(2): 210–227.

Kwan, Mei-Po. 2012. "The Uncertain Geographic Context Problem." *Annals of the Association of American Geographers* 102(5): 958–968.

Laswell, Harold D. 1936. *Politics: Who Gets What, When, How.* New York: McGraw Hill.

Little, Roderick JA, and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data.* 2nd ed. Hoboken, NJ: Wiley.

MacKerron, George, and Susana Mourato. 2013. "Happiness is greater in natural environments." *Global Environmental Change* 23(5): 992–1000.

Mass, J, R A Verheij, S de Vries, P Spreeuwenberg, F G Schellevis, and P P Goenewegen. 2009. "Morbidity is related to a green living environment." *Journal of Epidemiology and Community Health* 63: 967–973.

Monogan, James E. 2013. "A Case for Registering Studies of Political Outcomes: An Application in the 2010 House Elections." *Political Analysis* 21(1): 21–37.

Montoliu, Raul, Jan Blom, and Daniel Gatica-Perez. 2013. "Discovering Places of Interest in Everyday Life from Smartphone Data." *Multimedia Tools and Applications* 62(1): 179–207.

Moore, Ryan T., and Andrew Reeves. 2015. "Defining Racial and Ethnic Context with Geolocation Data." *Manuscript.* .

Musal, Rasim Muzaffer. 2010. "Two Models to Investigate Medicare Fraud within Unsupervised Databases." *Expert Systems with Applications* 37(December): 8628–8633.

Oliver, Eric J., and Tali Mendelberg. 2000. "Reconsidering the Environmental Determinants of White Racial Attitudes." *American Journal of Political Science* 44(July): 574–589.

Reeves, Andrew. 2011. "Political Disaster: Unilateral Powers, Electoral Incentives, and Presidential Disaster Declarations." *Journal of Politics* 73(4): 1142–1151.

Richards, Neil. 2015. *Intellectual Privacy: Rethinking Civil Liberties in the Digital Age.* New York: Oxford University Press.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data.* New York, NY: CRC press.

Schneider, Carl E. 2015. *The Censor's Hand.* Cambridge MA: MIT Press.

Schobel, Johannes, Marc Schickler, Rüdiger Pryss, Hans Nienhaus, and Manfred Reichert. 2013. Using Vital Sensors in Mobile Healthcare Business Applications: Challenges, Examples, Lessons Learned. In *Proceedings of the International Conference on Web Information Systems and Technologies.*

Serrano, Nicolás, Josune Hernantes, and Gorka Gallardo. 2013. "Mobile Web Apps." *IEEE Software* (September/October): 22–27.

Stolle, Dietlind, Stuart Soroka, and Richard Johnston. 2008. "When Does Diversity Erode Trust? Neighborhood Diversity, Interpersonal Trust and the Mediating Effect of Social Interactions." *Political Studies* 56(1): 57–75.

Stopher, Peter R., Philip Bullock, and Frederic Horst. 2002. Exploring the Use of Passive GPS Devices to Measure Travel. In *Applications of Advanced Technologies in Transportation.* American Society of Civil Engineers pp. 959–967.

Sweeney, Latanya. 2000. "Simple Demographics Often Identify People Uniquely." Carnegie Mellon University, Data Privacy Working Paper 3 Pittsburg, PA.

Tarran, Brian. 2014. "A New Model for Data Sharing." *Significance* 11(October): 15–16.

Vazquez-Prokopec, Gonzalo M., Steven T. Stoddard, Valerie Paz-Soldan, Amy C. Morrison, John P. Elder, Tadeusz J. Kochel, Thomas W. Scott, and Uriel Kitron. 2009. "Usefulness of Commercially Available GPS Data-loggers for Tracking Human Movement and Exposure to Dengue Virus." *International Journal of Health Geographics* 8(68): 1–11.

Wang, Hao, and Jerome P. Reiter. 2012. "Multiple Imputation for Sharing Precise Geographies in Public Use Data." *Annals of Applied Statistics* 6(1): 229–252.