Political Science Research and Methods (2020), 1–15
doi:10.1017/psrm.2020.10

RESEARCH NOTE

Defining racial and ethnic context with geolocation data

Ryan T. Moore[1]* and Andrew Reeves[2]

[1]Department of Government, American University, Kerwin Hall 226, 4400 Massachusetts Avenue NW, Washington, DC 20016-8130, USA and [2]Department of Political Science, Washington University in St. Louis, Campus Box 1063, One Brookings Drive, St. Louis, MO 63130, USA
*Corresponding author. Email: rtm@american.edu

(Received 23 April 2015; revised 9 September 2016; accepted 18 October 2016)

Abstract

Across disciplines, scholars strive to better understand individuals' milieus—the people, places, and institutions individuals encounter in their daily lives. In particular, political scientists argue that racial and ethnic context shapes attitudes about candidates, policies, and fellow citizens. Yet, the current standard of measuring milieus is to place survey respondents in a geographic container and then to ascribe all that container's characteristics to the individual's milieu. Using a new dataset of over 2.6 million GPS records from over 400 individuals, we compare conventional static measures of racial and ethnic context to dynamic, precise measures of milieus. We demonstrate how low-level static measures tend to overstate how extreme individuals' racial and ethnic contexts are and offer suggestions for future researchers.

Keywords: context; data collection; measurement; spatial methods

1 Introduction

In daily life, individuals are shaped by the people, places, and institutions they encounter—their *milieus*. Many studies argue that factors such as neighborhood racial and ethnic composition influence a variety of attitudes and behaviors (e.g., Key 1949; Huckfeldt 1986; Oliver and Mendelberg 2000; Hopkins 2010). Yet, others contend that "contextual influence is more the exception than the rule" (Hopkins 2013: 2). King (1996: 160) argues that "after we control for what we have learned about voters, there isn't much left for contextual effects." Other studies suggest that geographic context is simply a proxy for social networks, which, they argue, are the actual cause of political behavior (Eulau and Rothenberg 1986).

We argue that much scholarship in this vein uses measures of racial and ethnic context that fail to capture the theoretical concept of a milieu—the experience that an individual has in her environment while moving through space. The inability to measure the foundational concept of context compromises conclusions that these studies might draw. One study found that defining context as, for example, census tracts of residence instead of city can lead scholars to draw opposite conclusions about the effect of ethnic context (Cho and Baer 2011).

We document this problem and propose a potential solution through improved data collection. To begin, we demonstrate that quantifications of racial and ethnic context suffer from two measurement problems. The first problem stems from the researcher's choice of *level*, and the other—a less well-studied one—from measures' *static* nature. Though we focus on racial and ethnic context, our analysis is just as applicable to other studies of contextual effects across disciplines.[1]

[1]Social scientists consider an array of contextual influences including exposure to terrorism on tolerance (Peffley et al. 2015), polling locations on vote choice (Berger et al. 2008), unemployment on voting (Park and Reeves forthcoming), the

Early in a study of contextual effects, researchers choose the geographic containers that will define their subjects' contexts. Though researchers seek the characteristics of an individual's milieu, they are limited to characteristics of predetermined units like the ZIP code or census tract in which a respondent resides. Once the container is chosen, all of its characteristics are ascribed to the respondent as her contextual experience. Sometimes these containers are small, like a single census block group, and other times they are expansive, like an entire state. Small geographies may arbitrarily constrain an individual's measures to a few city blocks, while large geographies unrealistically assume an individual experiences an entire American state. Even the smallest state is hundreds of thousands of times larger than a city block, yet studies pick either container to measure the same context. Often, decisions about which container to use may be "choices driven by convenience and data availability" (Cho and Baer 2011: 418).

Recent research in measuring racial and ethnic context deepens our understanding of perceptions versus objective measures (Wong et al. 2012, 2013, 2015, 2020; Velez and Wong 2017; Hjorth et al. 2016) and overlapping residential communities (Baybeck 2006). However, we argue that even those researchers interested in objective rather than perceptual measures should employ individualized, precise, dynamic measures when possible. Below, we review approaches to measuring different contexts, arguing that traditional measures can be systematically misleading. Section 3 presents original individual-level geolocation data from over 400 individuals that we began recruiting in July 2013. Section 4 shows how the natural variation in individual experiences in our data belies static, aggregate measures. We evaluate particular levels of geographic aggregation and provide advice for scholars without access to dynamic measures. Finally, we detail the large individual variation in geographic context, even within particular American metropolitan areas.

2 Measuring racial and ethnic context

The existing census measures of racial and ethnic context have limited decades of studies using static geographic containers. As a result, measures of context may significantly misrepresent the local dynamic experiences that are posited to influence individuals' political behaviors and attitudes. Many applications measure context at the county level, for example.[2] However, the demographics and political environment of a respondent's county may differ greatly from those of the part of the county she actually lives in, works in, drives through, or experiences during her day. American counties are as large as 20,000 square miles—San Bernardino County, California, is larger than Costa Rica, Bosnia and Herzegovina, and Denmark—and encapsulate many different kinds of areas. As such, county-level averages likely misrepresent the experiences of many individuals. The average county includes about 1000 square miles; an average measure taken over a 50-mile by 20-mile area will significantly distort many individual experiences. We show several examples in Section 5.

Contextual factors are hypothesized to drive a number of attitudes related to prejudice. One seminal theoretic divide is whether in-group interactions with out-group members increase hostility or empathy toward the out-group. For example, Dinesen and Sønderskov (2015: 550) finds that "residential exposure to ethnic diversity reduces social trust." This work frequently uses static geographic containers including counties (Giles and Evans 1986; Fosset and Kiecolt 1989; Glaser 1994; Taylor 1998; Stein et al. 2000; Dixon 2006); cities or standard metropolitan statistical areas

socioeconomics of neighborhoods on divorce (South 2001), and even proximity to wolf packs on support for wolf hunting (Hopkins 2013). For an overview, see Cho and Gimpel (2012). Outside of political science, scholars have considered the effect of geographic context on a variety of health outcomes such as substance abuse (Mennis and Mason 2012), cardiovascular disease (Congdon 2009), and income inequality and racial context on overall individual health status (McLaughlin and Stokes 2002; Mellor and Milyo 2004).

[2] We count our own work among these studies (e.g., Gasper and Reeves 2011; Kriner and Reeves 2012, 2015; Moore and Ravishankar 2012).

(SMSAs) (Fosset and Kiecolt 1989; Bledsoe et al. 1995; Taylor 1998; Oliver and Mendelberg 2000; Oliver and Wong 2003; Dixon 2006); and states (Hero and Tolbert 2004). Lower levels of geography are also used, including housing projects (Ford 1973), ZIP codes (Oliver and Mendelberg 2000; Gilliam et al. 2002; Cho and Baer 2011; Weber et al. 2014), and neighborhoods (Bledsoe et al. 1995; Stolle et al. 2008; Cho and Baer 2011). Rarely, very low levels such as census blocks are used (Oliver and Wong 2003). As we show, each of these containers consistently mismeasures context when compared to individuals' actual dynamic experiences.[3]

Geographic contexts have also been connected to other political behaviors, such as state to turnout (Solt 2010), neighborhood to turnout (Cho et al. 2006), county to views on Ronald Reagan (MacKuen and Brown 1987), cities and census blocks to attitudes toward local government and one's neighborhood (Baybeck 2006), and neighborhood to various forms of participation (Huckfeldt 1979). Mutz and Mondak (2006) compares the workplace as a locus of political discussion with other contexts like the neighborhood and household. Baybeck and Huckfeldt (2002) and Berry and Baybeck (2005) deploy GIS data to demonstrate that geographic distance strongly predicts the density of political discussion networks and that interstate competition drives lottery adoptions (but not welfare benefit levels), respectively.

A small body of research takes a more circumspect approach to conceptualizing their measures of geography (Cho and Baer 2011; Reeves and Gimpel 2012; Wong et al. 2013; Christiansen et al. 2016; Gimpel et al. forthcoming). Analyzing economic attitudes, Reeves and Gimpel (2012) considers states, media markets, and counties, and then constructs multiple geographic regions based on social, political, and economic similarities of adjacent counties. Christiansen et al. (2016) argues that national economic perceptions are driven by extremely local conditions—unemployment rates within 80 m of home, but not unemployment within larger radii, such as 2500 m. Cho and Baer (2011) analyzes racial attitudes of blacks as a function of, among other things, multiple measures of racial context including census tract, ZIP code, neighborhood, and city. The study finds that the proportion of Latinos in a respondent's census tract is negatively associated with attitudes toward Latinos; however, these same attitudes are positively associated with the proportion of Latinos measured at the city level. Using ZIP code or neighborhood measures yields statistically insignificant relationships between the same variables. Cho and Baer (2011) then shows that incorporating *nearby* census tracts (as well as a measure of socioeconomic status) yields consistent results across geographic levels. Cho et al. (2012) similarly employs nearby geographic units in their spatial analyses.

The important finding of Cho and Baer (2011) is consistent with our work here, but there are differences. First, our sample here is geographically broader, touching every American state. Second, we dynamically measure the racial and ethnic milieus through which individuals actually pass, rather than statically assign them to a single set of nested geographic containers. Third, the lowest level of geography we employ is the census block, of which there are about 200 times as many as ZIP codes.

Even work that considers multiple levels of geography assumes that static measures—whatever their level—capture racial or economic experiences. As we demonstrate below, static measures of any level may not capture dynamic contexts, such as those of two people who live in the same block, but travel through different parts of a city during their daily lives.

We build on work in geography by Kwan (e.g., 2009, 2012) which asserts that contextual measures of exposure should be "people-based," rather than "place-based." Kwan argues that environmental influences on health and behavior unfold over the course of the day and depend on the

[3]Studies also address how characteristics of a place change over time (e.g., Green et al. 1998; Hopkins 2010). While we do not address the temporal dimension of an individual's milieu here, it merits further study. Others also relate context to political behaviors such as voting for California's immigration Proposition 187 (Morris 2000), for George Wallace for president (Wright 1976), and for a black mayoral candidate (Carsey 1995).

places passed through, rather than being defined by the single place we live. This holds whether we define where we live broadly, as in a US state, or narrowly, as in a single census block.

2.1 How aggregate census measures obscure static locations

Large geographies *could* represent their smaller constituent geographies quite well. For example, if a county was 60 percent white, each of the census blocks within that county could also be 60 percent white. However, larger geographies can also significantly misrepresent the demographics individuals encounter within them. A county that was 60 percent white could be a mixture of census blocks that were 100 percent white and 0 percent white. Further, where the county average misrepresents the nature of many blocks, we would not expect very different blocks to lie next to each other. American residential segregation (Massey and Denton 1993) implies that demographically similar blocks are likely to cluster. Thus, where block demographics vary within a county, we expect county measures to be a poor representation of individuals' much more localized experiences.[4]

For example, consider the city of Saint Louis, Missouri. The city is 48.5 percent black, but 35 percent of the city's census blocks are over 90 percent white or black.[5] If an individual was to experience all census blocks in the city, then 48.5 percent could be an accurate measure of her racial milieu. However, if she remained primarily in homogeneous black or white census blocks, then this summary measure would be a misleading characterization of her interaction with other racial groups. To address this, we define context dynamically, based on the characteristics of the areas through which an individual travels.

2.2 Toward low-level, dynamic measures of geography

Consider datasets commonly employed in social scientific research and the static geographic identifiers they sometimes make available. The American National Election Studies (ANES) makes available state and Congressional district indicators; it restricts access to county and ZIP code. The Cooperative Congressional Election Study (CCES) includes respondent region, state, county, Congressional district, and ZIP code, but ZIP code is missing for 60 percent of the sample in 2011. The General Social Survey (GSS) provides respondent region, but restricts access to census tract, as well as to higher levels of aggregation, including state, SMSA, and county.

By contrast, consider the geographic experience of the two individuals whose travels are plotted in Figure 1. These are two real people both of whom were observed in Los Angeles County, California, from data we describe in Section 3. The two individuals occupy strikingly different parts of the county—information that would be completely obscured by static measures of county. Despite both living in Los Angeles County, their milieus are distinct. Person 1, whose path is plotted with gray triangles (△), primarily occupies the West Los Angeles area. Person 2, whose path is plotted with black circles (◯), is centered in the La Crescenta-Montrose community. These two individuals' paths differ spatially, but also represent distinct racial and ethnic experiences.

Table 1 displays our measures of the racial and ethnic contexts these two individuals experience in the first and third columns. We then compare these contexts to several different geographic containers into which scholars might typically assign them—a census tract, Los Angeles County, and the state of California. We see that these Angelenos' different spatial patterns imply different contexts, as well. Person 1's movement through Los Angeles County leads her to experience places where, overall, about 26 percent of inhabitants are white, 19 percent percent are Hispanic, 14 percent percent are black, and 12 percent are Asian. Person 2's

[4]For related critiques of standard measures of contextual effects, see Matthews (2008) and Kwan (2009).
[5]Of St. Louis's 9577 census blocks, 1315 are over 90 percent white while 2023 are over 90 percent black.

Fig. 1 - B/W online, B/W in print

Figure 1. The paths of two individuals in Los Angeles County. Person 1 (△) is concentrated in West Los Angeles and Person 2 (◯) in La Crescenta-Montrose.

Table 1. Geographic experiences of two people in Los Angeles County

	Person 1's milieu (△)	West LA Tract	Person 2's milieu (◯)	Montrose Tract	LA County	California
White (not Hispanic)	26	32	64	51	28	40
Black	14	13	6	1	9	6
Hispanic/Latino	19	32	5	12	22	17
Asian	12	2	19	28	14	13

Cells contain percentage of geography of given race/ethnicity. Los Angeles County represents the milieu of Person 1 (△) fairly well, but not that of Person 2 (◯). Population estimates from the West LA Census Tract (2696.02) and Montrose Census Tract (3005.02) are reported from the American Community Survey, 2008 to 2012 and estimates for Los Angeles County and California are based on the 2010 decennial census.

experience is very different. Person 2 experiences places where most of the inhabitants are white (64 percent) with the next largest group being Asians (19 percent). If a survey researcher only had these respondents' county of residence, then the researcher would assume they have *identical* contextual experiences. While the county-level measures might reasonably approximate the experience of Person 1, they would mischaracterize the racial and ethnic context of Person 2. Whether each person's static local tract represents their experience better than the county is unclear; for some race/ethnicities tract is better, for others the county is better.

Comparing the patterns of these two individuals highlights three challenges in measuring contextual experiences. First, these individuals spend much of their time in different parts of the same county, a geographic context frequently used by studies measuring racial or ethnic composition. Second, these individuals do not exist in a single ZIP code, census block, or census tract. If we assume that their racial and ethnic experience matches even these lower levels of geography, we ignore the extent to which individuals may experience heterogeneity of conditions, though they might live in homogeneous neighborhoods. Third, neither of these individuals travels through the entire 5000 square miles of Los Angeles County, much less the entire state of California. Even these two individuals traveling through the southwestern part of Los Angeles County have markedly different experiences.

3 A new approach: racial and ethnic context via mobile geolocation

In order to create a dynamic measure of racial and ethnic context based on individual milieus, we rely on location data collected for over 400 users of a smartphone application, which automatically records users' latitude and longitude based on Global Positioning System (GPS) hardware embedded in their mobile phones.[6] Specifically, we obtain a sample of users of the OpenPaths application, developed and maintained by the Research and Development Lab at the New York Times Company. The Lab describes OpenPaths as a "secure data locker for personal location information" that allows the user to maintain geographic records of their daily travels. To use OpenPaths, the user downloads the application onto an iOS- or Android-based smartphone and signs up for an OpenPaths account either through the application or on the Internet. The Lab requires an email address to sign up. To successfully record data, then, a participant must have a smartphone, download the application, and sign in. The data are then encrypted and stored, so that the user can access it and control whether others can access it.[7]

Once the user begins using OpenPaths, the application runs in the background of their mobile phone and updates when there is "significant change."[8] The GPS in smartphones relies on several services, including GPS satellites, nearby Wi-Fi hotspots, and cellular towers, to pinpoint the user's coordinates.[9] Geolocations are still estimated with some uncertainty, and application developers must trade off between location accuracy and battery life. Zandbergen (2009), for example, finds that the median measurement error for the iPhone 3G (introduced in 2008) was between 74 and 600 m depending on the source of the information.

3.1 Our sample

The creators of OpenPaths allow researchers to request users' permission to access their geolocated data. We submitted a description of our project to the developers of OpenPaths, and they approved our project. We were then able to solicit users for permission to access their data. The users can approve, deny, or ignore the request, and they may revoke access at any time. We confined our requests to individuals who recorded coordinates within the United States. Every OpenPaths user who met this criterion was sent an email from OpenPaths requesting access for our project to their GPS coordinates. Approximately 8000 requests were made with 446 participants allowing us access to their data, 162 rejecting our request, and the remainder not responding. Included with the GPS coordinates are unique and randomly generated identification strings for each user, a timestamp, and information about the user's device, operating system, and version of the application. Thus, our sample is limited to individuals with smartphones who have self-selected into being geotracked for personal means and who are willing to share their personal data with researchers. No additional information is available and we have no way to contact these users for additional information.

---

[6]The application automatically records location data, so we avoid some potential pitfalls and shorter time horizons of time diaries (e.g., Kwan 1998, 1999).

[7]On the uses of smart phones to collect GPS data outside of political science, see Asakura and Hato (2004); Eagle et al. (2009) and Palmer et al. (2013).

[8]For technical documentation on Apple's iOS geotracking protocols, see http://j.mp/1dB3C5w.

[9]See http://support.apple.com/kb/ht4995 and https://support.apple.com/en-us/HT203033 for details on Apple's iOS.

Figure 2. The OpenPaths sample. The roughly 2.6 million GPS coordinates for all individuals in our OpenPaths sample. Every state is represented; coordinates are concentrated in the Northeast and West Coast. Darker shading represents areas with more observations.

Our sample of OpenPaths users includes 2.6 million data points from 446 individuals. Figure 2 displays all of these points. The number of GPS points for each individual range from 1 to over 111,000 with the median number of coordinate pairs being about 3200. On average, we have about a year's worth of geolocation for the individuals (364.4 days) with a maximum of over 4 years' worth of data. For just over 1 percent of our respondents, we have less than a day's worth of geolocations. We place each geographic coordinate into a 2010 US Census block indicated by a 15-digit Federal Information Processing Standards (FIPS) code, which uniquely identifies the point's census block, census block group, census tract, county, and state. We record coordinates in all fifty states and the District of Columbia. Even sparsely-represented states include a few hundred observed points (North Dakota: 203; Nebraska: 899; and South Dakota: 974), and the more commonly-observed states include hundreds of thousands of points (Pennsylvania: 6 percent of the total, or 174,693; New York: 12 percent or 312,129; and California: 22 percent or 569,175). Online appendix Figure 1 displays the distribution of coordinates by state. Despite the scope of geographic coverage, we again note that we have no way to access covariate data to determine if the sample is demographically nationally representative.

3.2 Summarizing dynamic geographies

For our OpenPaths sample, we define the racial and ethnic context users experience by using their geolocated data to identify the characteristics of the census blocks where the respondent actually spends her time. Specifically, we calculate a population-weighted average of the racial and ethnic profile of all census blocks she travels through. This summary reduces the distribution of characteristics in the visited blocks to a single measure of context. This measure is a useful summary; as we show below, however, we can leverage a great deal more information about the distribution of blocks each individual experiences.

4 How static measures fail to capture dynamic contexts

We compare our dynamic measure of context to a wide variety of static measures, representing a range of approaches scholars have adopted in describing racial and ethnic context. In doing so, we arrive at two core conclusions. First, static measures are poor proxies for individuals' dynamic experiences. This mismeasurement persists even at the census block level, the lowest geography for which data are available. Second, low levels of static geography can be used to estimate dynamic experiences, but not uniformly across racial and ethnic groups.

4.1 Describing the mismeasurement

For each OpenPaths participant, we designate modal geographies. These geographies, in which the participant appears most, represent our estimates of the geographic containers (block, block group, tract, county, and state) into which a survey researcher measuring context would place the respondent. These geographies serve as static measures that we compare to the participant's dynamic context. To summarize the dynamic context, we take the mean percentage of residents of a given race in the census blocks through which the participant travels, weighted by block population.

Figure 3 plots our dynamic measure of white racial context on the y-axis against a static measure on the x-axis.[10] Each figure compares measures for one racial or ethnic group, and each panel compares the dynamic measure to a different level of geography. If static measures of geography were perfect proxies for respondents' experiences, then every point would lie on the black $y = x$ line in each panel. Clearly, this is not the case, and there are important systematic differences between the two measures. Where points fall above the $y = x$ line, the static measure understates the individual's experience with members of that racial or ethnic group; where they fall below, the static measure overstates their experience.

First, individual experiences of racial and ethnic context are typically more moderate than static measures suggest. When a respondent's modal census block has a small percentage of a given racial or ethnic group, static representation understates how much the individual experiences members of that race or ethnicity. At the extreme, for example, for participants whose modal block is 0 percent white (those at the very left-hand side of the first panel in Figure 3), the static measure underestimates their average experience of whites, who actually comprise between 25 and about 80 percent of the census blocks these participants experience.[11] The y-intercept of the bivariate linear regression line gives a sense of the average "bias" in the static measure at this point—about 34 percentage points (Tufte 1973). At the other extreme, static blocks that are majority white tend to overrepresent participants' average experience of whites; most points at the right side of the panel fall below the $y = x$ line.

Second, it does not appear that the dynamic measure is simply the static measure "regressed to the mean." Comparing our dynamic block-level milieu to the typical static county measure, for example, over two-thirds of the sample has its dynamic diversity understated by the static measure.

Third, the comparison of dynamic to static measures at the same low level of geography (the census block) demonstrates that the mismeasurement of geographic context is not merely a problem caused by high levels of aggregation, but also a failure to account for how individuals travel through their environments. We observe the same pattern across racial and ethnic groups and geographies, though with a few exceptions. As the last panel of online appendix Figure 4 shows, for example, static state measures nearly always underestimate the percentage of Asians in the blocks participants travel through.

The intercept of each regression line indicates the bias in static measures for those whose modal location for a given geography lacks members of a particular race or ethnicity. The slopes provide guidance for how to translate from static measures to dynamic ones. For example, the slope in the first panel of Figure 3 is 0.4, indicating that on average, an individual from a

[10]See the online appendix for measures of other racial and ethnic contexts.

[11]We follow the convention of using "white" to mean white, non-Hispanic or -Latino. "Non-whites" include Hispanics, Latinos, and those giving non-white racial identities.

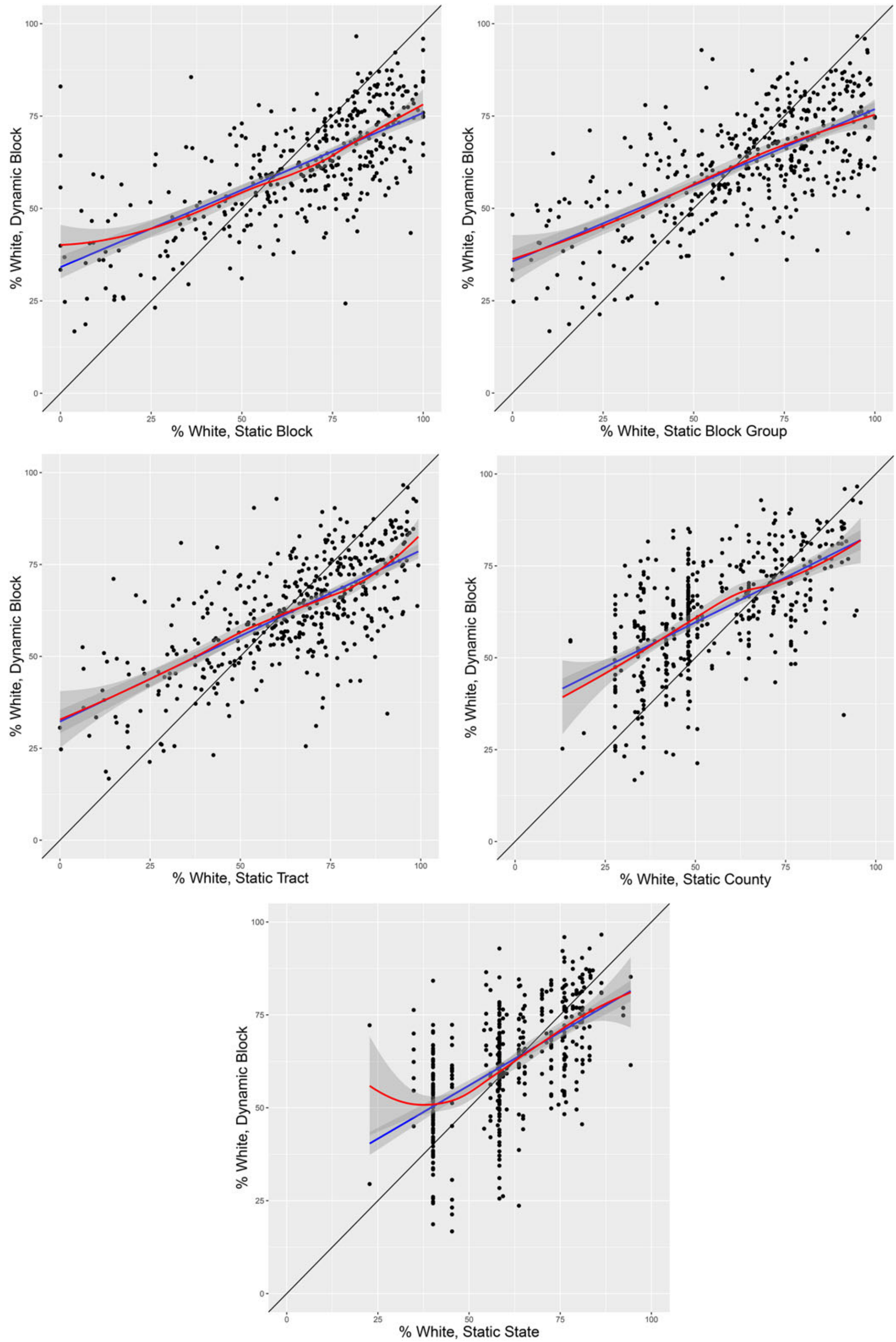

Fig. 3 - Colour online, B/W in print

Figure 3. Dynamic versus static measures of percent white. Panels show block, block group, tract, county, and state contexts. Modal location for a geography on *x*-axis; mean of dynamic contexts on *y*-axis. Respondents' dynamic experiences are more diverse than their static county where points lie above $y = x$. Linear regression (black dashed line) and loess smooths (white dashed line) are displayed.

block 10 percentage points more white than another actually only experiences about 4 percentage points more whites. We provide coefficient estimates for the full set of linear models in online appendix Table 1.

5 Individual variation in racial and ethnic context

We can further explore the experiences of an individual by examining the distribution of their contextual experiences. Figures 4 and 5 present the distribution of the racial and ethnic composition of the census blocks individuals travel through, grouped by modal county. The x-axis is the percent non-white, where densities farther to the right indicate more diverse (less white) racial and ethnic contexts. For each plot, the percent of the county population that is non-white is indicated by a vertical line.

Consider the participants from Buncombe County, North Carolina and Jefferson County, Colorado shown respectively in the left and center panels of Figure 4. These individuals spend most of their time in relatively homogeneous census blocks that match the characteristics of their counties as a whole. These distributions differ starkly from that of the individual from Wake County, North Carolina, who travels through a bimodal set of census blocks, represented in the right panel of Figure 4. This participant spends time in places that are either majority white or majority non-white, with less time spent in racially or ethnically mixed places. A static measure, even at the census block level, would clearly miss the variation in his or her experience. For example, the distribution of percent white in the census blocks our median respondent experiences has a standard deviation of about 21 percentage points, indicating that respondents tend to travel through heterogeneous groups of areas.

Bimodal variation in experiences is not confined to Wake County, North Carolina. Figure 5 displays the distributions of racial and ethnic experiences for several metropolitan counties for which we have multiple respondents. This figure illustrates the variation of experiences within a single geographic locality (in this case, county). If these individuals were part of a survey that measured context at the county level, their distributions would all collapse to the point on the x-axis at the vertical line. But as we see in Figure 5, the modal experience of one individual in Salt Lake County, Utah, whose mode is at about 50 percent non-white or Hispanic, is more diverse than most of the modal racial and ethnic experiences of our respondents from Brooklyn (Kings County), New York.

As with Figure 4, these plots also often reveal multimodal distributions of racial and ethnic experiences. We provide an example of a multimodal census block distribution within a county elsewhere, showing that even if an individual experienced every census block in her county, the summary measure could still mask extreme experiences (Moore and Reeves 2017). How might the attitudes of someone who spends their time in one type of census blocks vary from those of an individual who travels from one contextual extreme to another, even if their average experiences are the same?

6 Discussion

We demonstrate that static measures of lived experience, usually defined by residence at a single level of geography, fail to capture the variety of experiences that individuals are often posited to have. In some cases, these systematic misrepresentations represent a form of measurement error that plagues estimates of the effects of racial and ethnic context (e.g., Cho and Baer 2011).

Where scholars only have access to traditional measures of context, future modeling efforts may enable researchers to map these measures onto the likely dynamic contexts that typical residents of a geography experience. If one could collect dynamic geolocations and residential locations for a large set of Americans, for example, one could model dynamic context as a function of static contexts. If the predictions from such a model were sufficiently precise, then perhaps static

Fig. 4 - Colour online, B/W in print

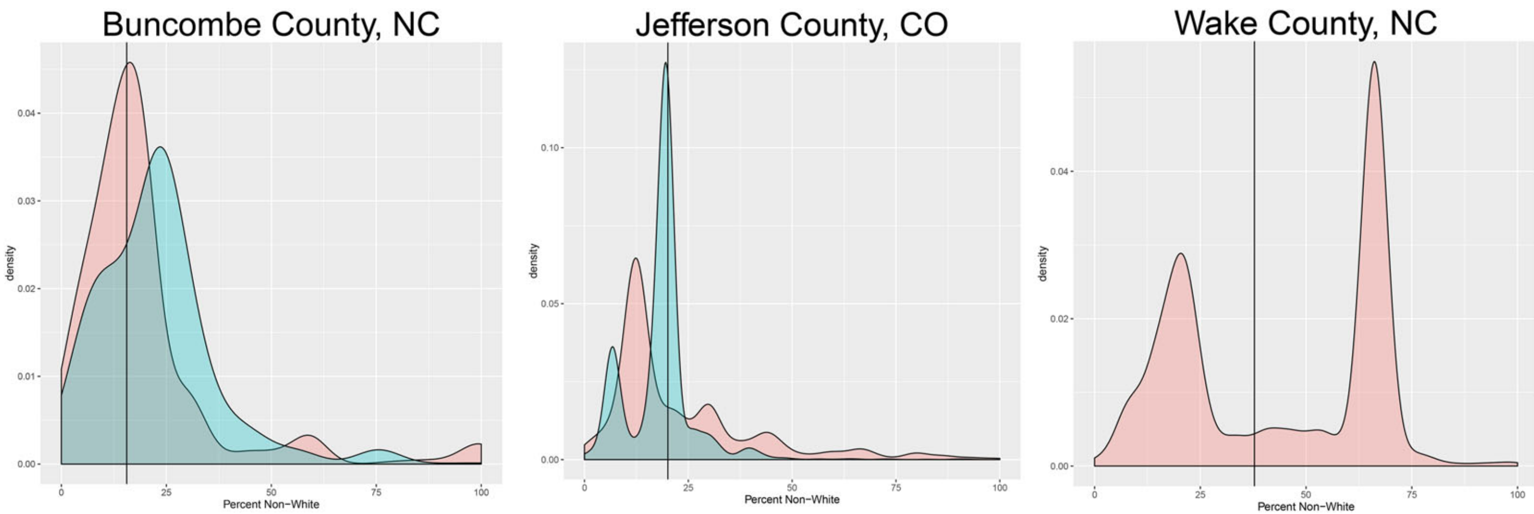

Figure 4. Racial experiences that are unimodal and bimodal. For some individuals, county captures both the mean and modal experience of the individual (left and center panels). For others (right panel), the county obscures the distribution of racial experiences. Panels show the distributions of racial and ethnic composition of the census blocks that individuals travel through, grouped by modal county. The *x*-axis is the percent non-white; densities farther to the right indicate more diverse (less white) racial and ethnic contexts. Vertical lines represent the percent of the county population that is non-white.

Figure 5. Examples of racial experiences in select metropolitan counties. Individuals living in the same county have disparate racial experiences. Panels show the distribution of the racial and ethnic composition of the census blocks that individuals travel through, grouped by modal county. The *x*-axis is the percent non-white; densities farther to the right indicate more diverse (less white) racial and ethnic contexts. Vertical lines represent the percent of the county population that is non-white.

Fig. 5 - Colour online, B/W in print

block, block group, tract, county, or state measures from traditional surveys could be used to impute dynamic experiences. We think such an approach has promise, but would require a large, dense network of Americans to be reliable.

For many contextual effects, research is still limited by what administrative data are available. We employ the lowest level of geography available in the census, the block, in our description of each individual's milieu. While we would prefer an even finer and more dynamic measure, we do not yet have, for example, a measure of the people within a participant's field of vision. Currently, we are limited to the characteristics of a milieu that are collected at that administrative level. Despite this limitation, we argue that the dynamic collection of relatively small administrative contexts approximates the actual environment one experiences better than current practices. In the future, we could develop additional dynamic measures by defining age contexts, for example, using block level data. Any time researchers have geographic data, dynamic contexts could be defined using the variables available for that container.

Our data collection strategy reflects larger trends toward exploiting geographic social scientific data. Significant geolocation efforts stand to revolutionize the study of disease patterns, for example, and the study of racial and ethnic contextual effects should follow. Scholars have long been interested in the effects of milieus on a variety of outcomes. However, only recently has the ability to precisely, dynamically measure where individuals travel in their daily lives become affordable and unintrusive enough to be feasible for large social scientific samples. With the ability to collect frequent individual geolocations, political science can better link empirical measures to theoretical concepts like racial and ethnic milieus.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/psrm.2020.10.

Data. All data and information necessary to replicate the results in this article are available in the Harvard Dataverse at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/G3YJXY.

Acknowledgments. We thank Brady Baybeck, Michael Kellermann, Hyun Woo Lim, Melissa Sands, Allen Settle, Stuart Soroka, Steven Webster, and participants of the Toronto Political Behaviour Workshop, the Women's and Minority Representation conference of McGill University's Centre for the Study of Democratic Citizenship, and the American University Government Research Seminar for comments and suggestions on earlier drafts.

References

Asakura Y and Hato E (2004) Tracking survey for individual travel behaviour using mobile communication instruments. *Transportation Research Part C* **12**, 273–291.

Baybeck B (2006) Sorting out the competing effects of racial context. *Journal of Politics* **68**, 386–396.

Baybeck B and Huckfeldt R (2002) Spatially dispersed ties among interdependent citizens: connecting individuals and aggregates. *Political Analysis* **10**, 261–275.

Berger J, Meredith M and Wheeler SC (2008) Contextual priming: where people vote affects how they vote. *Proceedings of the National Academy of Sciences* **105**, 8846–8849.

Berry WD and Baybeck B (2005) Using geographic information systems to study interstate competition. *American Political Science Review* **99**, 505–519.

Bledsoe T, Welch S, Sigelman L and Combs M (1995) Residential context and racial solidarity among African Americans. *American Journal of Political Science* **39**, 434–458.

Carsey TM (1995) The contextual effects of race on white voter behavior: the 1989 New York city mayoral election. *Journal of Politics* **57**, 221–228.

Cho WKT and Baer N (2011) Environmental determinants of racial attitudes redux: the critical decisions related to operationalizing context. *American Politics Research* **39**, 414–436.

Cho WKT and Gimpel JG (2012) Geographic information systems and the spatial dimensions of American politics. *Annual Review of Political Science* **15**, 443–460.

Cho WKT, Gimpel JG and Dyck JJ (2006) Residential concentration, political socialization, and voter turnout. *Journal of Politics* **68**, 156–167.

Cho WKT, Gimpel JG and Shaw DR (2012) The tea party movement and the geography of collective action. *Quarterly Journal of Political Science* **7**, 105–133.

Christiansen MB, Sønderskov KM and Dinesen PT (2016) Reconsidering the neighborhood effect: does exposure to residential unemployment influence voters' perceptions of the national economy?. *Journal of Politics* **78**, 719–732.

Congdon P (2009) A multilevel model for cardiovascular disease prevalence in the US and its application to micro area prevalence estimates. *International Journal of Health Geographics* **8**, 1–14.

Dinesen PT and Sønderskov KM (2015) Ethnic diversity and social trust: evidence from the micro-context. *American Sociological Review* **80**, 550–573.

Dixon J (2006) The ties that bind and those that don't: toward reconciling group threat and contact theories of prejudice. *Social Forces* **84**, 2179–2204.

Eagle N, Pentland A and Lazer D (2009) Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences* **106**, 15274–15278.

Eulau H and Rothenberg L (1986) Life space and social networks as political contexts. *Political Behavior* **8**, 130–157.

Ford SW (1973) Interracial public housing in a border city: another look at the contact hypothesis. *American Journal of Sociology* **78**, 1426–1447.

Fosset M and Kiecolt KJ (1989) The relative size of minority populations and white racial attitudes. *Social Science Quarterly* **70**, 820–835.

Gasper JT and Reeves A (2011) Make it rain? Retrospection and the attentive electorate in the context of natural disasters. *American Journal of Political Science* **55**, 340–355.

Giles M and Evans A (1986) The power approach to intergroup hostility. *The Journal of Conflict Resolution* **30**, 469–486.

Gilliam FD, Valentino N and Beckmann M (2002) Where you live and what you watch: the impact of racial proximity and local television news on attitudes about race and crime. *Political Research Quarterly* **55**, 755–780.

Gimpel JG, Lovin N, Moy B and Reeves A (forthcoming) The Urban-Rural gulf in American political behavior. *Political Behavior*. https://doi.org/10.1007/s11109-020-09601-w

Glaser JM (1994) Back to the black belt: racial environment and white racial attitudes in the south. *The Journal of Politics* **56**, 21–41.

Green DP, Strolovitch DZ and Wong JS (1998) Defended neighborhoods, integration, and racially motivated crime. *American Journal of Sociology* **104**, 372–403.

Hero RE and Tolbert CJ (2004) Minority voices and citizen attitudes about governmental responsiveness in the American states: do social and institutional contexts matter?. *British Journal of Political Science* **34**, 109–121.

Hjorth F, Dinesen PT and Sønderskov KM (2016) The content and correlates of subjective local contexts. In *Proceedings of the Annual Meeting of the American Political Science Association.*

Hopkins DJ (2010) Politicized places: explaining where and when immigrants provoke local opposition. *American Political Science Review* **104**, 40–60.

Hopkins DJ (2013) Misplaced: the limits of contextual influence on Americans' political attitudes. (Typescript).

Huckfeldt RR (1979) Political participation and the neighborhood social context. *American Journal of Political Science* **23**, 579–592.

Huckfeldt RR (1986) *Politics in Context: Assimilation and Conflict in Urban Neighborhoods*. New York, NY: Agathon Press, Inc.

Key VO (1949) *Southern Politics in State and Nation*. Knoxville: University of Tennessee Press.

King G (1996) Why context should not count. *Political Geography* **15**, 159–164.

Kriner DL and Reeves A (2012) The influence of federal spending on presidential elections. *American Political Science Review* **106**, 348–366.

Kriner DL and Reeves A (2015) Presidential particularism and divide-the-dollar politics. *American Political Science Review* **109**, 155–171.

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

Kwan MP (1999) Gender and individual access to urban opportunities: a study using space-time measures. *The Professional Geographer* **51**, 210–227.

Kwan MP (2009) From place-based to people-based exposure measures. *Social Science & Medicine* **69**, 1311–1313.

Kwan MP (2012) The uncertain geographic context problem. *Annals of the Association of American Geographers* **102**, 958–968.

MacKuen M and Brown C (1987) Political context and attitude change. *American Political Science Review* **81**, 471–490.

Massey DS and Denton NA (1993) *American Apartheid: Segregation and the Making of the Underclass.*Cambridge, MA: Harvard University Press.

Matthews SA (2008) The salience of neighborhood: some lessons from sociology. *American Journal of Preventive Medicine* **34**, 257–259.

McLaughlin DK and Stokes CS (2002) Income inequality and mortality in US counties: does minority racial concentration matter?. *American Journal of Public Health* **92**, 99–104.

Mellor JM and Milyo JD (2004) Individual health status and racial minority concentration in US states and counties. *American Journal of Public Health* **94**, 1043–1048.

Mennis J and Mason MJ (2012) Social and geographic contexts of adolescent substance use: the moderating effects of age and gender. *Social Networks* **34**, 150–157.

Moore RT and Ravishankar N (2012) Who loses in direct democracy?. *Social Science Research* **41**, 646–656.

Moore RT and Reeves A (2017) Learning from Place in the Era of Geolocation. In Bachner J, Wagner Hill K and Ginsberg B (eds). *Analytics, Policy and Governance*. New Haven, CT: Yale University Press, pp. 118–136.

Moore RT and Reeves A (2020) *Replication Data for: Defining Racial and Ethnic Context with Geolocation Data*. Harvard Dataverse.

Morris I (2000) African American voting on proposition 187: rethinking the prevalence of interminority conflict. *Political Research Quarterly* **53**, 77–98.

Mutz DC and Mondak JJ (2006) The workplace as a context for cross-cutting political discourse. *Journal of Politics* **68**, 140–155.

Oliver EJ and Mendelberg T (2000) Reconsidering the environmental determinants of white racial attitudes. *American Journal of Political Science* **44**, 574–589.

Oliver EJ and Wong J (2003) Intergroup prejudice in multiethnic settings. *American Journal of Political Science* **47**, 567–582.

Palmer JRB, Espenshade TJ, Bartumeus F, Chung CY., Ozgencil NE and Li K (2013) New approaches to human mobility: using mobile phones for demographic research. *Demography* **50**, 1105–1128.

Park T and Reeves A (forthcoming) Local unemployment and voting for president: uncovering causal mechanisms. *Political Behavior*. https://doi.org/10.1007/s11109-018-9502-4

Peffley M, Hutchison ML and Shamir M (2015) The impact of persistent terrorism on political tolerance: Israel, 1980 to 2011. *American Political Science Review* **109**, 817–832.

Reeves A and Gimpel JG (2012) Ecologies of unease: geographic context and national economic evaluations. *Political Behavior* **34**, 507–534.

Solt F (2010) Does economic inequality depress electoral participation? Testing the Schattschneider hypothesis. *Political Behavior* **32**, 285–301.

South SJ (2001) The geographic context of divorce: do neighborhoods matter?. *Journal of Marriage and Family* **63**, 755–766.

Stein RM, Post SS and Rinden AL (2000) Reconciling contact and context effects on racial attitudes. *Political Research Quarterly* **53**, 285–303.

Stolle D, Soroka S and Johnston R (2008) When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions. *Political Studies* **56**, 57–75.

Taylor MC (1998) How white attitudes vary with the racial compositions of local populations: numbers count. *American Sociological Review* **63**, 512–535.

Tufte ER (1973) The relationship between seats and votes in two-party systems. *The American Political Science Review* **67**, 540–554.

Velez YR and Wong G (2017) Assessing contextual measurement strategies. *Journal of Politics* **79**, 1084–1089.

Weber CR, Lavine H, Huddy L and Federico CM (2014) Placing racial stereotypes in context: social desirability and the politics of racial hostility. *American Journal of Political Science* **58**, 63–78.

Wong C, Bowers J, Williams T and Simmons KD (2012) Bringing the person back in: boundaries, perceptions, and the measurement of racial context. *Journal of Politics* **74**, 1153–1170.

Wong C, Bowers J, Rubenson D, Fredrickson M and Rundlett A (2013) Does the context fit the outcome? When (or where) racial/ethnic context should affect politics. *American Political Science Association Annual Meeting*.

Wong C, Bowers J, Rubenson D, Fredrickson M and Rundlett A (2015) Do you see what I see?: Psuedoenvironments, ethnic diversity, and social capital. Manuscript (August).

Wong C, Bowers J, Rubenson D, Fredrickson M and Rundlett A (2020) Maps in people's heads: assessing a new measure of context. *Assessing Contextual Measurement Strategies* **79**, 160–168.

Wright GC (1976) Community structure and voting in the south. *Public Opinion Quarterly* **40**, 201–215.

Zandbergen PA (2009) Accuracy of iPhone locations: a comparison of assisted GPS, WiFi and cellular positioning. *Transactions in GIS* **13**, 5–25.

Cite this article: Moore RT, Reeves A (2020). Defining racial and ethnic context with geolocation data. *Political Science Research and Methods* 1–15. https://doi.org/10.1017/psrm.2020.10

Online Appendix for "Defining Racial and Ethnic Context with Geolocation Data"[*]

Ryan T. Moore[†] Andrew Reeves[‡]

April 14, 2020

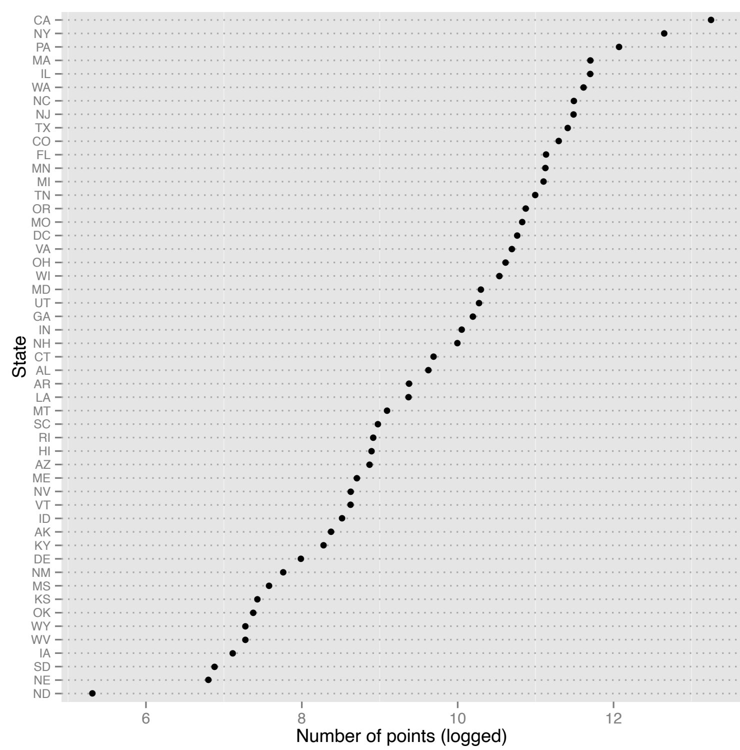

Figure 1: OpenPaths Observations by State. Each point presents the logged number of observed geolocations for all respondents for each state.

[*]All data and information necessary to replicate the results in the article Moore and Reeves (2020*a*) are available in the Harvard Dataverse at Moore and Reeves (2020*b*) at `https://doi.org/10.7910/DVN/G3YJXY`.

[†]Associate Professor, Department of Government, American University, Kerwin Hall 226, 4400 Massachusetts Avenue NW, Washington DC 20016. `http://www.ryantmoore.org`, `rtm@american.edu`, 202.885.6470.

[‡]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.

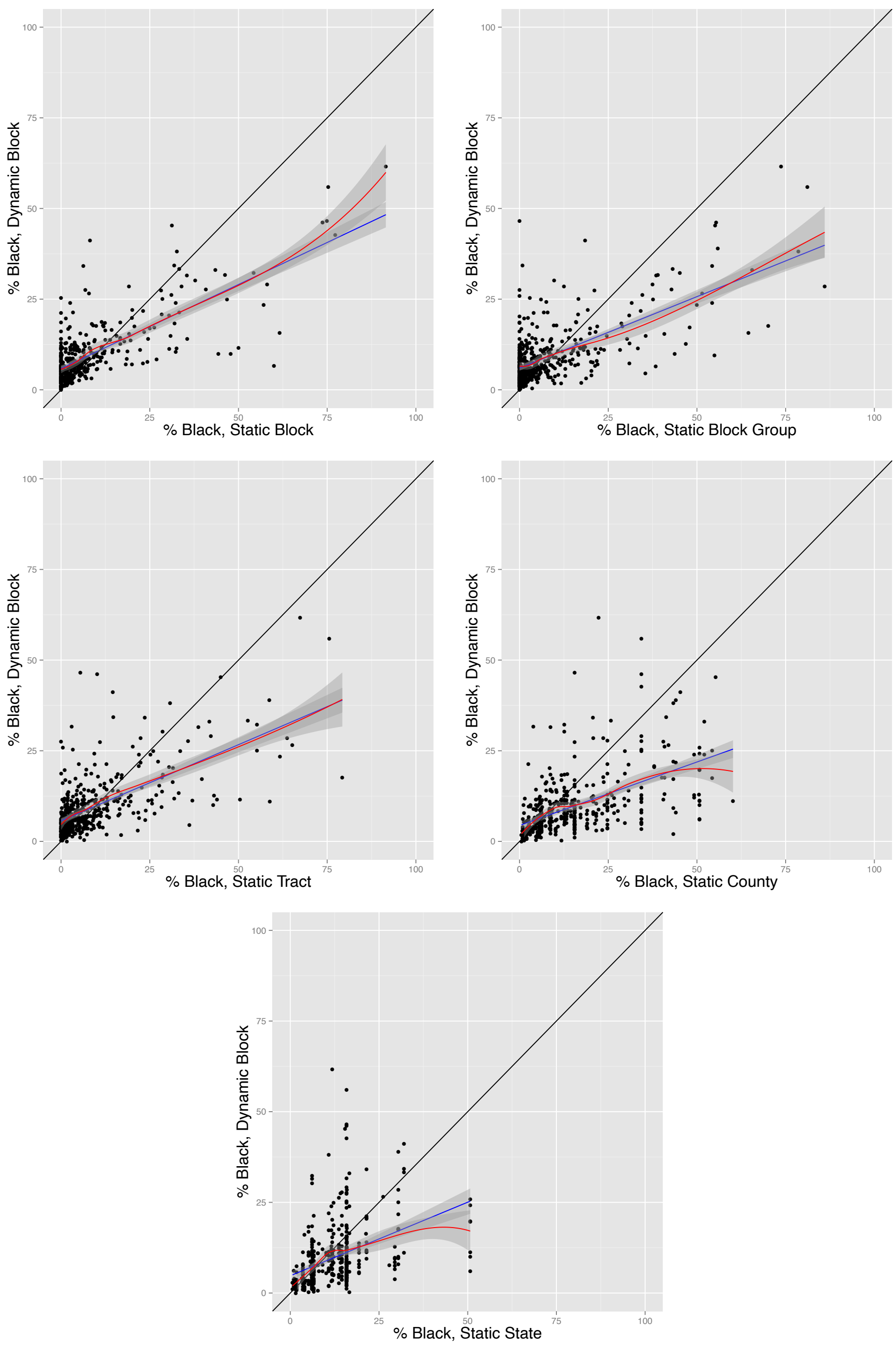

Figure 2: **Dynamic vs. Static Measures of Percent Black.** Panels show block, block group, tract, county, and state contexts. Modal location for a geography on x-axis; mean of dynamic contexts on y-axis. Linear regression (blue) and loess smooths (red) displayed.

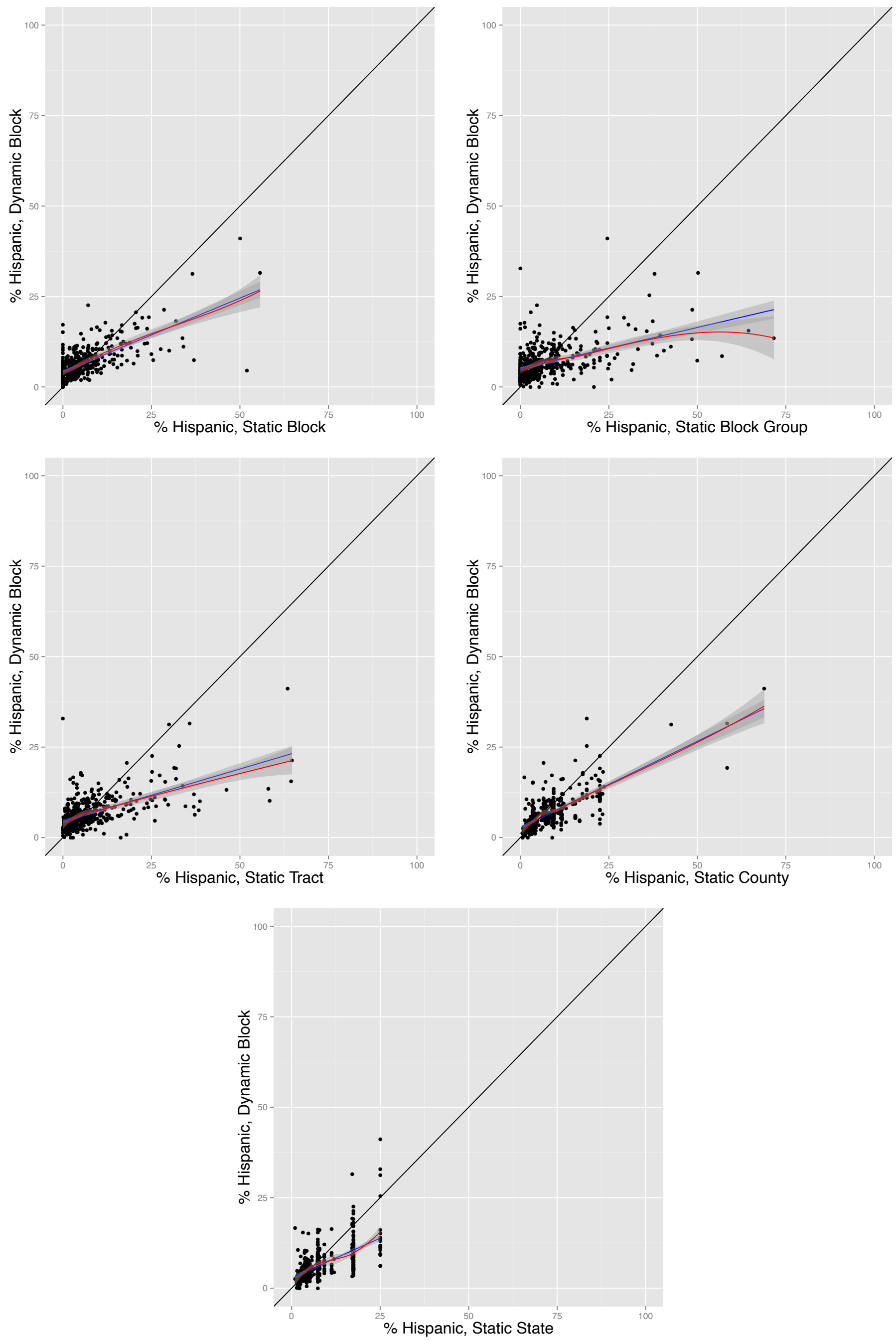

Figure 3: **Dynamic vs. Static Measures of Percent Hispanic.** Panels show block, block group, tract, county, and state contexts. Modal location for a geography on x-axis; mean of dynamic contexts on y-axis. Linear regression (blue) and loess smooths (red) displayed.

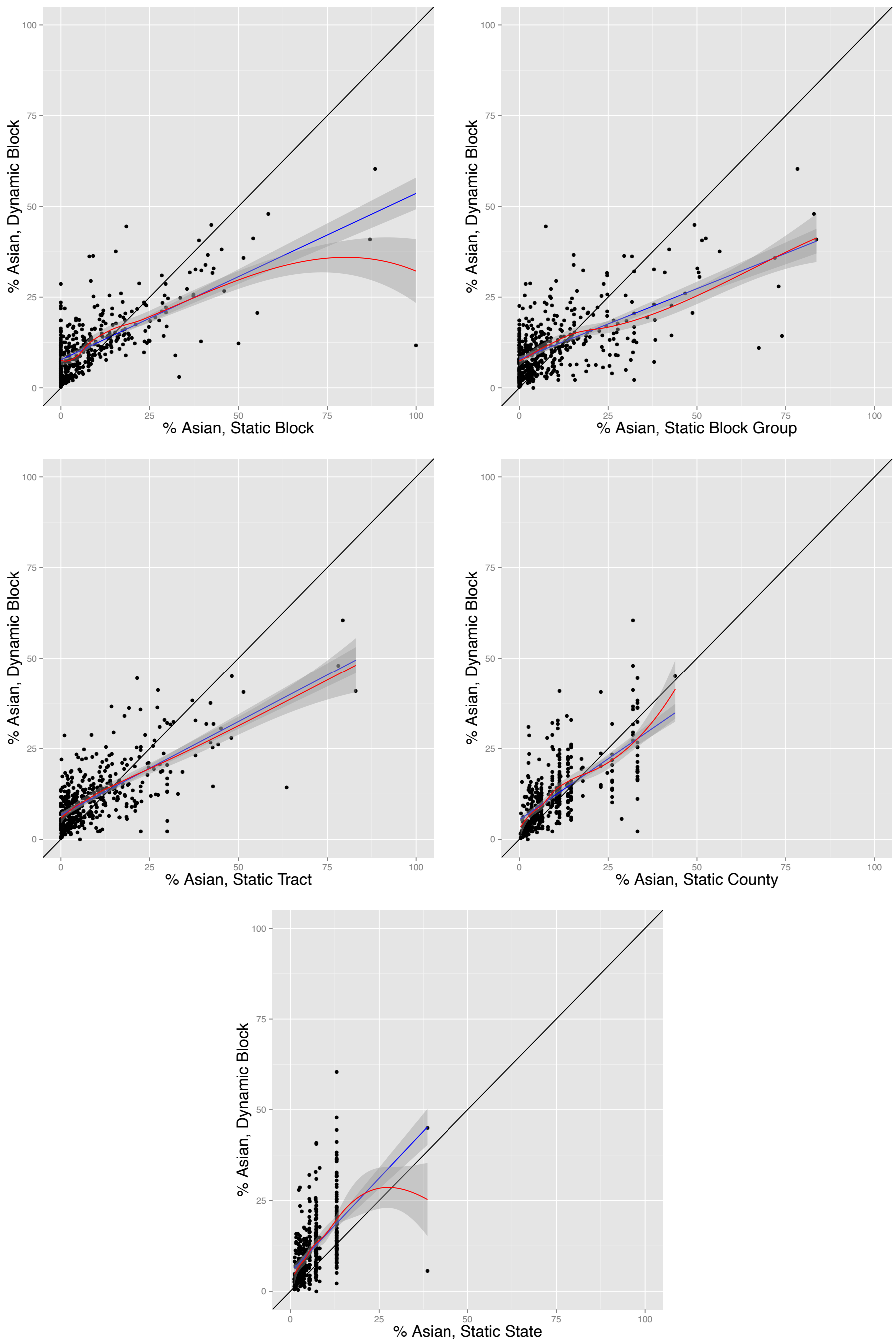

Figure 4: **Dynamic vs. Static Measures of Percent Asian.** Panels show block, block group, tract, county, and state contexts. Modal location for a geography on x-axis; mean of dynamic contexts on y-axis. Linear regression (blue) and loess smooths (red) displayed.

Race/Ethnic Group	Geography	Statistic	Value
White	Block	Intercept	34.1
White	Block	Slope	0.4
White	Block	RMSE	10.7
White	Block Group	Intercept	35.6
White	Block Group	Slope	0.4
White	Block Group	RMSE	11.3
White	Tract	Intercept	32.2
White	Tract	Slope	0.5
White	Tract	RMSE	10.9
White	County	Intercept	35.2
White	County	Slope	0.5
White	County	RMSE	12.0
White	State	Intercept	27.3
White	State	Slope	0.6
White	State	RMSE	12.3
Black	Block	Intercept	5.9
Black	Block	Slope	0.5
Black	Block	RMSE	5.8
Black	Block Group	Intercept	6.0
Black	Block Group	Slope	0.4
Black	Block Group	RMSE	6.3
Black	Tract	Intercept	5.7
Black	Tract	Slope	0.4
Black	Tract	RMSE	6.5
Black	County	Intercept	4.4
Black	County	Slope	0.4
Black	County	RMSE	7.2
Black	State	Intercept	4.8
Black	State	Slope	0.4
Black	State	RMSE	7.9
Hispanic	Block	Intercept	4.2
Hispanic	Block	Slope	0.4
Hispanic	Block	RMSE	3.4
Hispanic	Block Group	Intercept	5.0
Hispanic	Block Group	Slope	0.2
Hispanic	Block Group	RMSE	4.2
Hispanic	Tract	Intercept	4.5
Hispanic	Tract	Slope	0.3
Hispanic	Tract	RMSE	3.8
Hispanic	County	Intercept	2.8
Hispanic	County	Slope	0.5
Hispanic	County	RMSE	3.2
Hispanic	State	Intercept	2.9
Hispanic	State	Slope	0.4
Hispanic	State	RMSE	3.8
Asian	Block	Intercept	7.7
Asian	Block	Slope	0.5
Asian	Block	RMSE	6.4
Asian	Block Group	Intercept	7.9
Asian	Block Group	Slope	0.4
Asian	Block Group	RMSE	6.8
Asian	Tract	Intercept	6.7
Asian	Tract	Slope	0.5
Asian	Tract	RMSE	6.2
Asian	County	Intercept	5.4
Asian	County	Slope	0.7
Asian	County	RMSE	6.5
Asian	State	Intercept	5.2
Asian	State	Slope	1.0
Asian	State	RMSE	7.4

Table 1: Models of dynamic contexts as a function of static contexts. The dependent variables are the white, black, Hispanic, and Asian dynamic contexts. Each is modeled as a function of the static measure at the level of the census block, census block group, census tract, county, and state. For each of the 20 models (4 racial/ethnic groups $\times$ 5 geographic levels), we present the slope and y-intercept coefficients and root mean squared errors displayed in paper Figure 3 and appendix Figures 2 through 4.

References

Moore, Ryan T., and Andrew Reeves. 2020*a*. "Defining Racial and Ethnic Context with Geolocation Data." *Political Science Research and Methods* .

Moore, Ryan T., and Andrew Reeves. 2020*b*. "Replication Data for: Defining Racial and Ethnic Context with Geolocation Data." Harvard Dataverse, DOI: 10.7910/DVN/G3YJXY.