

New Approaches to Human Mobility: Using Mobile Phones for Demographic Research

John R.B. Palmer · Thomas J. Espenshade · Frederic Bartumeus · Chang Y. Chung · Necati Ercan Ozgencil · Kathleen Li

Abstract This article explores new methods for gathering and analyzing spatially rich demographic data with mobile phones. It describes a pilot study in which mobile phone users around the world were successfully recruited to share GPS- and cellular-tower-based information on their movements and respond to dynamic, location-based surveys using an open-source Android application. The pilot study illustrates the great potential of mobile phone methodology for moving spatial measures beyond residential census units and investigating a range of important social phenomena, including the heterogeneity of activity spaces, the dynamic nature of spatial segregation and the contextual dependence of subjective well-being.

Keywords Spatial Demography · Activity Space · Segregation · Subjective Well-Being · Ecological Momentary Assessment · Human Mobility · Mobile Phones · GPS Data

J. R.B. Palmer (CORRESPONDING AUTHOR)

Woodrow Wilson School of Public and International Affairs and Office of Population Research,
Princeton University, 284 Wallace Hall, Princeton, NJ, 08544, USA
e-mail: jrpalmer@princeton.edu

T. J. Espenshade

Office of Population Research and Department of Sociology, Princeton University, Princeton,
NJ, USA

F. Bartumeus

Center for Advanced Studies of Blanes, CEAB-CSIC, Blanes, Spain

C. Y. Chung

Office of Population Research, Princeton University, Princeton, NJ, USA

N. E. Ozgencil

Syncsort, Woodcliff Lake, NJ, USA

K. Li

Google, New York, NY, USA

Introduction

Questions of human location and movement are of central importance in demography, but answers are limited by our frequent inability to observe and measure where people are and where they are going. Census data tell us about general patterns of human settlement and migration, but very little about where people spend their time or where and how they travel. Interviews and surveys give us more detail, but these are constrained by scale and self-reporting errors, and they often fail to provide a dynamic picture. We are thus confronted with an important obstacle as we try to better understand a wide range of human activity—from immigration and urbanization to social networks; from the environmental effects of population growth and the health effects of environmental exposure to the spread of diseases and the distribution of social or cultural capital.

Mobile phones offer a way around this obstacle. With their increasing worldwide popularity and sophistication, the research potential of these devices is quickly becoming apparent, and the possibilities for using them to record location and movement are especially promising. A given phone generally can be traced to the cellular tower with which it is in closest proximity, and a user's movement can be estimated by analyzing the phone's transitions from one tower to another. More accurate location data may be obtained if a phone is able to process signal information from multiple cellular towers, from nearby devices, or from distant satellites (Eagle et al. 2009). Phones with these capabilities are now firmly entrenched in the market and gaining widespread use around the world. That many of these phones have a variety of additional sensing capabilities, store huge amounts of data, and function as miniature computers adds to their usefulness (Raento et al. 2009).

In this article, we describe a pilot study that we conducted to explore the use of mobile phones in demographic research, and to test one particular technique for which they may be especially well-suited: a dynamic, location-based survey. The pilot study uses GPS and cellular tower data from mobile phones not only to infer subjects' locations of residence (to which we are limited when using census data) and observe how they respond to questions at a fixed time and place (to which we are limited when relying on traditional survey methods) but also to examine where they spend time when not at home, how they move around, and how they respond to questions at a variety of different times and places. The study provides a partial glimpse at the answers to such questions as how traditional concepts of neighborhood might be adjusted to better account for the varied spaces in which people carry out their day-to-day activities, how patterns of spatial segregation observed in residential housing permeate other aspects of life, and how subjective well-being depends on spatial context. It demonstrates the powerful ways in which demographers may use mobile phones to move beyond residential census units and "put people into place" (Entwisle 2007), more fully capturing the complex relationship between spatial behavior and local environment. And it identifies some of the limitations and risks associated with mobile phone studies, including the need for appropriate safeguards to protect subjects' privacy in light of the particular problems presented by spatial data (Gutmann et al. 2008; VanWey et al. 2005).

We start with an overview of the basic problem of studying human movement, the mobile phone technology that can be used for this purpose, and the research that has made use of this technology to date. We then describe our pilot study and we present our results. Finally, we evaluate the study methodology and discuss ways in which this research complements and extends the field of demography.

Mobile Phones as Demographic Instruments

The scientific value of observing human spatial variation has long been apparent (e.g., Du Bois 1899, Park and Burgess 1925; Anderson and Massey 2001), and spatial questions are increasingly at the core of research in a variety of disciplines. Specific research questions are diverse but may be grouped into three basic categories: (1) those that aim to describe and explain patterns of location and movement (e.g., Hägerstrand 1970; Golledge and Stimson 1997), (2) those concerned with how people influence their spatial context (e.g., Cassels et al. 2005), and (3) those concerned with how spatial context influences people (e.g., Inagami et al. 2007). In calling on demographers to take the lead in researching how neighborhoods influence health, Entwisle (2007) demonstrates the interconnectedness of the three categories and the importance of treating them together. People are not merely passive objects on which neighborhoods operate, she argues, but rather agents who make choices about where to spend their time and who modify those places through their presence and activities (see also Cummins et al. 2007). Entwisle's point extends beyond health research and highlights a need that mobile phone methods are well suited to address.

By making it possible to systematically follow people as they move, mobile phones help overcome the limitations of censuses and other methods that focus on static places.¹ They allow us to reconstruct detailed movement trajectories and analyze the space through which people move as they go about their daily activities (Golledge and Stimson 1997). Although this can be done using more traditional methods like time-space diaries or interviews (Wang et al. 2012; Wong and Shaw 2011; Matthews 2011; J. Y. Lee and Kwan 2011; Basta et al. 2010; Kwan 2008, 1999; Janelle et al. 1998; Goodchild and Janelle 1984), mobile phones avoid many of the accuracy and logistical problems inherent in self-reporting (Kwan and Lee 2004; Stopher et al. 2007; Murakami and Wagner 1999; Golob and Meurs 1986). Moreover, because people incorporate them into their daily lives naturally, mobile phones are particularly well-suited for measuring people's reactions and capturing information about their surroundings in real time, facilitating techniques like ecological momentary assessment or the construction of precise individual environmental exposure profiles (Borrell 2011; Ahas 2011; Kwan 2009; Raento et al. 2009; Shiffman et al. 2008; Nusser et al. 2006).

Mobile phone location tracking has been the subject of a body of research that has expanded rapidly over the past decade as an array of commercial and scientific applications have become technologically and socially feasible (Ahas 2011; Ahas and Mark 2005; Asakura and Hato 2004). Many phones are able to estimate their own locations using signals from cellular towers, satellites, or other beacons, but even phones that lack this capability can be localized externally based on the traces they leave on cellular networks and other receivers. Data collection techniques can be divided between active positioning, in which subjects volunteer to be tracked and data are collected from their phones or their phones' traces, and passive positioning, in which trace data are obtained directly from network operators without contacting the individuals being tracked (and also, consequently, without much information about these individuals apart from location) (Ahas 2011). Alongside these collection techniques has come the development of methods for organizing and evaluating spatio-temporal data, including techniques for analyzing

¹ The difference is between selecting people and observing the space through which they move and selecting places and observing the people who move through them. This is a basic choice that must be confronted in studying any collection of moving objects and it is often the case that the first approach (which mobile phones facilitate) is more illuminating but also harder to implement (Ökubo and Levin 2001).

geospatial lifelines—the paths formed in space-time by successive location estimates of each tracked individual (Laube et al. 2005; Laube et al. 2007).

Within the social sciences, mobile phones have been used to address such topics as social networks (Eagle et al. 2009; Pentland 2007), mobility and migration (Ahas 2011; Wesolowski and Eagle 2010; Gonzalez et al. 2008), crowds and urban dynamics (Calabrese et al. 2010; Calabrese et al. 2007; Reades et al. 2007), transportation behavior (Ahas 2011; Reddy et al. 2010; Asakura and Hato 2004; Asakura and Hato 2001), health behavior (Madan et al. 2010; Anokwa et al. 2009), and happiness (MacKerron 2011; Killingsworth and Gilbert 2010). All of this mobile phone research has come alongside the emergence within the social sciences of a “mobilities” paradigm, which emphasizes the role of movement in creating social and material realities (Büscher et al. 2011; Adey 2010; Büscher and Urry 2009; Urry 2007; Sheller and Urry 2006).

The results of all this research have been impressive, but much remains to be explored in terms of methods and their integration within theoretical frameworks (D’Andrea et al. 2011). Scholars have complained that existing methods deal poorly with the “fleeting,” transitory nature of movement (Büscher et al. 2011; Law and Urry 2004). D’Andrea et al. (2011: 151) note that methodological development has lagged behind empirical and conceptual advances and “the examination of basic research protocols has only emerged recently.” In particular, they emphasize the need for methods that link micro and macro level observation.

From the perspective of demography, mobile phone methods offer new sources of data to answer traditional questions, while also suggesting new ways of formulating questions and thinking about populations to begin with. Mobile phones can be used to measure traditional migration events in populations that are otherwise hard to reach. For example, Wesolowski and Eagle (2010) used cellular tower data to infer changes of residence within a Nairobi slum that lacked the administrative or census records that would have been used for this purpose in other settings. By taking this approach, they were able to measure a high rate of residential turnover, much of it due to movement within the slum—something that would have been very difficult to detect and quantify using traditional methods.

Mobile phones also make it possible to understand population in a much more dynamic way, untethered to residence, adding an empirical, demographic grounding to the emerging mobilities paradigm (D’Andrea et al. 2011). Isaacman et al. (2010) used mobile phone call records to analyze daily movement patterns in New York and Los Angeles, finding significant variation by day of the week, city, and even neighborhood. Calabrese et al. (2007), Reades et al. (2009), Carioua et al. (2010), Calabrese et al. (2010), Ratti et al. (2007), and Ratti et al. (2006) tracked the ebbs and flows of urban populations throughout the day as people go to and from work, attend concerts, visit their weekend homes and engage in all of the other activities that we tend to ignore when asking traditional demographic questions like, “What is the city’s population?”

Viewing population from this perspective can help us to think about what we even mean when we talk about a city and how we should draw urban boundaries. It can help us to better understand what Bertolini and Dijst (2003) describe as the increasing divergence between modern cities’ social and physical dimensions. Similarly, it can help us find more meaningful ways to define neighborhoods for research purposes given the inherent limitations of commonly used census units, which exclude a great deal of spatial information (Reardon et al. 2008; B. A. Lee et al. 2008) and fail to capture individuals’ actual experiences and exposures (Matthews 2011; Basta et al. 2010; Kwan 2009; Chaix et al. 2009; Entwisle 2007).

The dynamic perspective can also help us to investigate a variety of questions that blend demography with other disciplines. For example, the individual health consequences of

environmental exposure and the environmental and epidemiological impacts of population growth depend on the activities in which the population's members are engaged and the spaces through which they move (Cummins et al. 2007; Inagami et al. 2007; de Castro et al. 2006; Cassels et al. 2005; Martens and Hall 2000). Tracking human movement and environmental exposure can thus provide more insight than simply counting the number of residents within an administratively-defined boundary, and mobile phones offer a way to do this at high resolutions (Wiehe et al. 2008). Similarly, many of the social consequences of residential segregation depend on where people spend time when not at home (Ellis et al. 2004). While place of residence may sometimes be a good proxy for this, there is much to be gained by measuring movement outside the home more precisely. Studies based on self-reported places of work and travel diaries show notable differences in urban space use depending on individuals' socio-demographic characteristics (Ellis et al. 2004; Zandvliet and Dijst 2006).

Objectives

In our pilot study—the Human Mobility Project (HMP)—we tested the feasibility of using mobile phone positioning to obtain spatially rich demographic data and administer a dynamic, location-based survey using subjects around the world who would volunteer by simply downloading an application and installing it on their own phones. The linking of phone position with survey administration is at the forefront of methodological developments in this area (Ahas 2011), providing a way to merge localized response information with larger-scale movement patterns (D'Andrea et al. 2011). Moreover, the active positioning of volunteers with their own phones anywhere in the world is an approach that has been less well explored than the collection of cellular tower data (Blumenstock et al. 2010; Wesolowski and Eagle 2010; Isaacman et al. 2010; Calabrese et al. 2007; Reades et al. 2009; Carioua et al. 2010; Calabrese et al. 2010; Ratti et al. 2007; Ratti et al. 2006) or the tracking of volunteers who are given phones and who often travel through fixed areas containing pre-positioned beacons (Eagle and Pentland 2006; Wiehe et al. 2008; Eagle and Pentland 2009; Eagle et al. 2009; Raento et al. 2009; Asakura and Hato 2004). These latter two approaches can be quite powerful—cellular towers yield massive quantities of data while phone distribution improves sample representativeness and pre-positioned beacons boost accuracy—but the first lacks demographic detail and the second has limited scalability.

Because our approach makes it possible to ascertain respondents' characteristics, it has the potential to provide richer data than that obtained from cellular towers, while also allowing larger scales and broader ranges of geographic coverage than phone distribution studies. By harnessing the efforts of large numbers of volunteers to participate remotely in the creation of knowledge, our approach is in the spirit of what Goodchild (2007) describes as the volunteered geographic information movement. It is similar to approaches currently being implemented by researchers at the London School of Economics (MacKerron 2011)² and Harvard (Killingsworth and Gilbert 2010).³ Whereas those studies focus on, respectively, environmental and psychological determinants of happiness, we focused on methodology and explored its application to demographic and sociological questions.

² See www.mappiness.org.uk.

³ See www.trackyourhappiness.org.

We centered the study on three substantive topics: activity space, spatial segregation, and subjective well-being.

Activity space is the term behavioral geographers use to describe the “locations within which an individual has direct contact as a result of his or her day-to-day activities” (Golledge and Stimson 1997: 279). It is the component of the individual’s overall environmental interaction that involves movement and direct contact, as opposed to communication. Activity space is potentially a much better measure of environmental exposure than measures based on census units (Matthews 2011; Basta et al. 2010; Kwan 2009), but the latter continue to dominate research in many fields. It has long been recognized that people spend significant time far from home and that the very notion of neighborhood is immensely difficult to define (Matthews 2008; Foley 1950). Gathering data on activity space and refining our understanding of place have been flagged as important priorities, particularly for research on health and environment (Matthews 2011). In our study, we explored subjects’ activity spaces, analyzing the sizes of these spaces using the minimum convex polygon method, which takes the area of the smallest convex polygon that may be drawn around an individual’s locations (Buliung and Kanaroglou 2006).

Spatial segregation along lines of race, ethnicity, and class is an important topic of study because it is usually associated with deep and pervasive inequality (Anderson and Massey 2001; Massey and Denton 1993). What social scientists generally measure when examining this topic is residential segregation (e.g., Massey and Denton 1988), but segregation occurs throughout the day as people engage in a variety of activities (Wang et al. 2012; Wong and Shaw 2011; Ellis et al. 2004). This may be partly driven by the interplay between home locations and the locations of jobs, schools and other socially significant places. Indeed, the mismatch between suburban job locations and segregated central city neighborhoods provides one explanation of how segregation concentrates unemployment and poverty (Wilson 1996, 1987; Kain 1968). Segregation outside the home may be driven also by a variety of other factors ranging from transportation network design to policing tactics to individual preferences. As a starting point for investigating the question using mobile phones, we explored the amount of time subjects spent in census blocks with different racial characteristics.

Finally, *subjective well-being* has been a topic of social inquiry since ancient times, and it constitutes an important measure by which to judge the impact of policies and events on people’s lives (Helliwell 2003; MacKerron 2011). Subjective well-being is generally conceived to encompass both affective components, such as feelings of happiness or pain, and cognitive components, such as evaluations of life satisfaction (Diener et al. 1999). People’s self-reports about both types of components may be influenced by a variety of time-varying personal and environmental factors, and their reports about affective components, in particular, may depend on whether they are communicated as the event is being experienced or recalled afterwards (Kahneman and Krueger 2006). Consequently, there are important reasons to employ methods that are able to record people’s subjective feelings along with contemporaneous contextual data. This is what we attempted to do in the pilot study, employing a location-based survey drawn from the group of methods known as Experience Sampling (Csikszentmihalyi and Hunter 2003). While Experience Sampling methods generally ask a broader array of affect questions as well as questions about activities and surroundings, we kept our survey simple and relied for context on basic location information—distance from home—in order to test what may be gained by combining this type of survey with a positioning device.

Methods

The pilot study had three main components: (1) an open source application that subjects could install on any device running the Android platform, (2) an SQL database that communicated with the phone application to trigger location-based surveys and collect data, and (3) a project website where subjects could securely register and activate the application, and provide us with basic demographic information about themselves (age, sex, and race⁴) and view some of the data being collected. All user interfaces were in English.

We chose to write the application for Android because it is a free, open source platform used by a growing number and variety of mobile devices around the world, and because its ability to run applications and processes in the background makes it well-suited to location monitoring (Chang et al. 2010; Arnold 2010). Android is used in hundreds of phone models available on every continent, and by late 2011 it was installed on over half of all new smartphones sold worldwide (Yarow 2011).

We placed the application on the Android market for three weeks in early 2010. Those who downloaded it were directed to the HMP webpage, where they could activate the application and consent to participate in the study.

The application ran in the background of subjects' phones and recorded location estimates at intervals of 2, 5, 10, 30, or 60 minutes.⁵ The estimates were based on GPS satellite signals when available, and otherwise on cellular tower signals. The application transmitted estimates to the server and the server triggered surveys on phones when they entered predetermined locations. This feature would make it possible to focus, for instance, on subjective well-being near major landmarks. For the pilot study, our interest was in the ordinary places people visit throughout the day, so we continuously updated the trigger points to included subjects' actual locations, and we added a stochastic component to the trigger itself to avoid overwhelming subjects with too many surveys.

While it would be possible to use multiple survey questions, each one triggered at a different location or time, we used just one question for the pilot: "How happy are you?" Subjects could respond by touching the phone's screen to select between zero and five stars (with half-star increments) or "Ask Me Later."

Results

During the three weeks that the application was on the Android market a total of 869 people downloaded it, and 270 registered and consented to be research subjects. Their self-reported races and sexes are summarized in Table 1, and self-reported ages are shown in Figure 1.

[TABLE 1 ABOUT HERE]

[FIGURE 1 ABOUT HERE]

⁴ We use "race" throughout the article to refer to both racial and ethnic self-identification. Subjects could identify as "White," "Black, African, African American," "Spanish/Hispanic/Latino," "Asian," "American Indian or Alaska Native," "Pacific Islander," or "Other," and we treat these as non-overlapping categories for simplicity.

⁵ We gave subjects the ability to select the interval.

We collected location estimates and survey responses from subjects during a 5-week period. Subjects' participation times varied but 128 participated for at least 24 hours, 62 participated for over a week and 13 participated for more than 2 weeks. The mean and median participation times were, respectively, 6.3 days and 2.2 days. We received a total of 68,198 location estimates, of which 29% were based on GPS signals and the rest on cellular tower signals. We retained and analyzed 67,294, while discarding 904 (1.3%) due to accuracy concerns.⁶ Within this analyzed subset, the overall mean and median Android-reported accuracies were, respectively, 1,210 meters and 1,000 meters for all estimates combined, 261 meters and 48 meters for the GPS-based estimates, and 1,586 meters and 1,318 meters for the cellular-tower-based estimates. Figure 2 maps the approximate centroid of each subject's set of location estimates.⁷

[FIGURE 2 ABOUT HERE]

Activity Space

The individual movement patterns revealed by each subject's location estimates exhibit varying degrees of regularity. As an example, Figures 3 through 6 plot the movement of two subjects with notably different behavioral patterns. We have selected these subjects because their differences are so clear and because their participation times are among the highest, each lasting just over 30 days. Figures 3 and 4 show the subjects' full sets of location estimates (points), with sequential estimates connected by straight lines, and scale indicated by axis tick marks (1 km intervals). To protect privacy, we have transformed the coordinates such that the location estimates maintain their relative distances but it is impossible to identify subjects' actual locations on earth.⁸

[FIGURE 3 ABOUT HERE]

[FIGURE 4 ABOUT HERE]

Most of the estimates in these plots lie on top of one another and cannot be distinguished. Figures 5 and 6, therefore, plot the same subjects' patterns with separate panels for weekday (A) and weekend (B) locations and a small amount of random noise added to each estimate to produce artificial spreading and thus aid in visualization. Location estimates in these plots are

⁶ We discarded estimates with Android-reported accuracy measurements of over 5 kilometers and those that, based on our own calculations, would have required subjects to have been traveling faster than 54 meters per second.

⁷ The centroid is calculated as the mean longitude and mean latitude for each subject. A small amount of random noise is added to each centroid to aid in visualization and protect subjects' privacy, using the random perturbation masking technique described in Armstrong et al. (1999).

⁸ We did this by applying what Armstrong et al. (1999) refer to as a scale transformation mask: We multiplied each subject's matrix of location estimates by a randomly generated number. While one might assume that it is sufficient to simply remove coordinates from the axes, the transformation is necessary because, depending on how the graphs are generated, their files may contain underlying coordinate data that can be extracted. We used a multiplicative transformation, rather than an additive one to make it impossible to reconstruct the original coordinates using scale information. Although this transformation does not preserve actual distances between points, it is possible to retain approximate map scale by measuring and transforming the actual distances between each subject's maximum and minimum latitude and longitude coordinates.

marked with points colored by the hour of the day when transmitted (in the subject's local time zone). As in the previous plots, sequential location estimates are connected by straight lines and scale is indicated by axis tick marks (1 km intervals).

Each of the displayed subjects' movement patterns is characterized by a high degree of clustering, and this is the case for many of our other subjects as well. In addition, each subject has at least two clusters, suggesting daily movements are not random but rather concentrated in specific locations that have some functional saliency to the subject. Moreover, we see from the clustering of colors in Figures 5 and 6 that when subjects return to locations on multiple occasions it is often at the same time of day. Finally, we see different patterns on weekends than on weekdays. Some areas are visited only on weekdays and some only on weekends. Even in areas visited on both weekdays and weekends, the timing of the visits may differ, as can be seen by the predominance of different sets of colors in each panel of Figures 5 and 6.

There are also notable differences between the activity spaces of these two subjects. Subject 1 (Figures 3 and 5) traveled through a much larger area than did Subject 2 (Figures 4 and 6). Likewise, the main clusters of location estimates are spread further apart in the case of Subject 1 than in that of Subject 2. The area of the smallest convex polygon within which each movement pattern fits (without the addition of any random noise) is 827 km² for Subject 1 but only 8 km² for Subject 2.

[FIGURE 5 ABOUT HERE]

[FIGURE 6 ABOUT HERE]

These types of differences can be seen in the movement patterns of our other subjects as well. Among the 96 who participated for over 24 hours and provided over 50 location estimates, the smallest convex polygons within which their movement patterns fit range from 0.01 km² to 56,000 km², with a median of 150 km² and an interquartile range of 60 to 394 km².

For subjects in the United States, we compared activity-space areas to the areas of the census blocks and tracts within which each subject resided, using census unit boundaries from 2000. We estimated the home census block for each subject as the one in which he or she spent at least four consecutive hours bracketing 3 am.⁹ We estimated the amount of time subjects spent in any given census block by summing the times elapsed between successive location estimates lying within that block. This is a modified version of the approach taken by Bartumeus et al. (2010) and Barraquand and Benhamou (2008), and it relies on the assumption that when successive points lie within the same block, the subject remained in that block during the intervening time. The approach ignores time spent between successive points that fall on either side of a block boundary, and time spent in a census block prior to the subject's first location estimate or subsequent to his or her last.

Using this approach, we were able to estimate home census blocks for 65 of the subjects for whom we calculated activity-space areas.¹⁰ We also identified the census tracts in which these blocks were located. The median home block area was 0.2 km², with an interquartile range of 0.03 to 0.7 km², and the median home tract area was 4.9 km², with an interquartile range of 2.7 to 10.9 km². In contrast, the median activity-space area for these subjects was 156.8 km², with an

⁹ When there was more than one such block, we used the one that occurred most frequently in the data.

¹⁰ In Figures 5 and 6, location estimates within the home block are marked with squares.

interquartile range of 61.0 to 377.2 km². If we take the ratio of each subject's residential tract area to activity-space area, the median is 0.037 and the interquartile range is 0.009 to 0.155. In terms of the percentage of each subject's total participation time spent within his or her home tract, the median was 61%, with an interquartile range of 23 to 84%. For home census blocks, the median was 38% and the interquartile range 15 to 65%. If we recalculate the percentages for home tracts excluding time spent in the home block itself, the median was 4% and the interquartile range 0 to 36%.

Spatial Segregation

The racial compositions of residential census units are the ingredients of traditional spatial segregation measures (Massey and Denton 1988), and we used data from the 2000 census to calculate the proportion of white residents (PWR) in subjects' home census block populations. For white subjects whose home blocks were estimated ($n=58$), the median PWR was 0.89, with an interquartile range of 0.79 to 0.98. For black and Latino subjects combined ($n=9$) it was 0.74, with an interquartile range of 0.70 to 0.77.

To move beyond the traditional focus on residence, we calculated, for each subject, the mean PWR in the census blocks in which that subject spent time, weighted by the amount of time spent in each block. Median results were similar to those based on residence, but the interquartile ranges move closer together, overlapping at 0.76 to 0.96 for whites versus 0.72 to 0.82 for blacks and Latinos. The ranges overlap further if we exclude time spent in the home block, although this requires a loss of data because some of these subjects spent insufficient time outside their home blocks to make the calculation possible. Subtracting home-block PWR from mean non-home-block PWR for each of these subjects yields a median difference of -0.02 for whites ($n=51$), with an interquartile range of -0.10 to 0.06, and 0.01 for blacks and Latinos ($n=8$), with an interquartile range of -0.10 to 0.03. In other words, at least half of the subjects had greater exposure to members of their own race in their home blocks than outside them. However, the reverse was also true: When they left their home blocks, at least a quarter of the whites spent time in higher PWR areas, and a quarter of blacks and Latinos spent time in lower ones. Thus, time spent outside the home block can both attenuate and intensify segregation.

Another way to view segregation outside the home block is in terms of the proportion of each subject's time spent in census blocks of given racial compositions. Figure 7 summarizes this information for blocks grouped into quartiles by PWR, with time in the home block excluded. To increase sample size for these calculations, we consider all subjects who we observed for at least four hours in non-home census blocks (including subjects for whom we could not estimate a home census block). The biggest differences can be seen in the percentage of time spent in blocks with PWR greater than 0.75: For white subjects ($n=81$), the median is 83% of the subject's time, with an interquartile range of 38 to 99%, while for blacks ($n=5$) it is 43%, with an interquartile range of 0 to 95%, and for Latinos ($n=9$) it is 7%, with an interquartile range of 0 to 25%.

[FIGURE 7 ABOUT HERE]

We do not attempt to analyze the extent to which these results are driven by spatial autocorrelation (cf. Zenk et al. 2005), and we recognize that our samples are small and likely biased. Nonetheless, it is apparent from these data how a dynamic, time-and-space approach to human movement can help us to advance beyond traditional measures of spatial segregation.

Subjective Well-Being

The approach also allows us to examine the link between subjective well-being and geographic context. During the study, the application sent a total of 485 surveys to subjects' phones, with each survey triggered by the location of the phone and asking subjects to rate their happiness as between 0 (least happy) and 5 (most happy) in increments of 0.5. We received 232 responses with a pooled mean and median of, respectively, 3.8 and 4. Given the high share of rejections, it is possible that these results are biased upward by a correlation between subjects' happiness and their willingness to answer the survey at a given time. In particular, it may be that subjects were more likely to respond when they were in a better mood. Leaving that concern aside, however, we explored the relationship between subjects' self-reported happiness and their distance from home.

Having already estimated census blocks of residence, we inferred that each subject's home was at the point where the bivariate normal kernel density estimate of the subject's locations within the home census block reached its maximum value (Venables and Ripley 2002).¹¹ We then calculated how far each subject was from home when sending each survey response. In total, we analyzed 168 survey responses from 36 subjects (Figure 8).¹²

[FIGURE 8 ABOUT HERE]

We fit a series of multilevel linear models to the data, treating each survey response i from subject j as a continuous random variable drawn from a normal distribution with mean μ_{ij} and standard deviation σ . The multilevel approach provided a way to account for subjects having different characteristic levels of subjective well-being and to focus on context-dependent changes relative to these levels. We modeled the mean of the survey response variable as

$$\mu_{ij} = \alpha_{j[i]} + X_i\beta \quad (1)$$

where $\alpha_{j[i]}$ is a subject-specific constant, X_i is a matrix of observed covariates,¹³ and β is a vector of estimated coefficients. For these models we rely on Bayesian parameter estimates that reflect compromises between pooled and unpooled (subject-level) means, depending on the number of responses given by each subject (Gelman and Hill 2007). We obtained the estimates using the Gibbs sampler implemented through the OpenBUGS software package (Lunn et al. 2009).

We find that male subjects tended to describe themselves as less happy the farther they were from their homes, whereas there was no statistically discernible relationship between female subjects and distance.¹⁴ This pattern holds even when we control for subjects' race. Each 10 km

¹¹ In Figures 3 through 6, this point is identified by the large red circle drawn around it.

¹² In addition to excluding responses from subjects whose home locations were not estimated (meaning all outside of the United States), we also excluded one extreme outlier in terms of distance from home.

¹³ These included variables for sex, race (black, white, and Latino were the only responses given in the data analyzed), distance from home, and the interaction between distance from home and sex. Additional models were tested using variables for time of day, weekend, working hours (9am–5pm), and census block racial characteristics, but these variables did not have clear relationships to the survey responses or sufficiently improve model fit to warrant inclusion in the final analysis.

¹⁴ These conclusions are based on our estimated coefficients on the variable interacting sex with distance from home.

increase in distance from home is associated with a decrease in reported male happiness of between 0.6 and 1.1 points, estimated with 95% confidence (Figure 9). This finding is also robust to alternate model specifications, including pooled-subject linear models and ordered logit models.

Figure 9 displays the relationship between distance from home and observed and predicted happiness responses in the multilevel model for white, Latino, and black men and women. The lines in panels A-C are drawn using coefficients estimated in each of 999 final iterations of the Gibbs sampler, saved after the sequences reached approximate convergence, in order to indicate uncertainty (see Gelman and Hill 2007). The lines in panel D of Figure 9 are drawn using the mean estimates.

[FIGURE 9 ABOUT HERE]

Discussion

Our study demonstrates that it is possible to obtain detailed information on the locations, demographic characteristics, and context-dependent responses of people all over the world simply by having them download and run an application on their mobile phones. It illustrates the great potential in adopting an approach that lies somewhere between the large-scale collection of coarse mobile network operator data and the small-scale collection of detailed data from devices that have been distributed to subjects or pre-positioned in bounded research areas. By capturing more subject-specific information than the former while offering greater scalability than the latter, this middle approach is well-suited for pursuing a wide range of demographic inquiries. The capacity of this approach to incorporate dynamic, location-based surveys is particularly valuable for a range of research questions that depend on subjects' descriptions of their surroundings, activities, or mental states as they are experiencing them.

The study also helps to clarify many of the methodological issues that must be confronted in this type of research. Representativeness and generalizability are clearly concerns when subjects self-select into a study that requires them to possess a smartphone and register on the Internet because we know there are systematic differences in mobile phone and Internet use across demographic groups (Blumenstock et al. 2010; Goodchild 2007). However, traditional survey methods present their own problems of representativeness and generalizability. Although the problems may be more serious when using mobile phone methods for certain purposes, the difference is one of degree and it is likely to shrink with time as the technology becomes more widely accessible. Moreover, mobile phone methods alleviate many of the problems presented by traditional methods, inducing participation by people who would not be willing to respond to paper surveys or interviews and minimizing reporting errors and logistical complications. Indeed, there may be much to be gained from using mobile phone methods in combination with traditional ones such that their strengths complement each other.

Another issue that must be confronted is privacy. One of the major complications of mobile phone localization studies is that the spatially explicit data they produce can be difficult to anonymize properly, particularly if the data are linked to information on social characteristics and particularly if the goal is to make data available publicly (Gutmann et al. 2008; VanWey et al. 2005). In our pilot study, access to the raw data we collected is limited to our research group, with one exception: We gave subjects on-line access to their own location data (as well as the option for them to request that we delete their data). In terms of processed data, we allowed

subjects to view on-line static maps showing aggregate survey responses from all subjects, and we have presented aggregate data in the present article. The only individual-level data we have presented here has been transformed to make it impossible to identify any subject or the actual locations they visited during the study.¹⁵

From a substantive perspective, the study provides strong evidence of the importance of moving beyond static, census-unit measurements when investigating the relationship between people and place (Entwisle 2007). We found a great deal of variation in the sizes of subjects' home census tracts, with the top quartile living in tracts over 4 times larger than those of the bottom quartile. This alone should make one cautious about relying too heavily on tract-based measures, as Matthews (2011) argues. The problem is made even clearer by our finding that the residential census tract accounts for less than 15.5% of the total area in which three quarters of the subjects moved and less than 3.7% for half of them. This is consistent with findings by Matthews (2011) and Basta et al. (2010), as well as earlier work by Foley (1950) and others (see Matthews 2008). If we care about the environment to which these people were exposed, their spatial choices, or the places they may have influenced through their presence or activities (Entwisle 2007), their residential census tracts alone tell only a small part of the story.

Incorporating time into the analysis gives more weight to home tracts but leads to the same conclusion. Half of the subjects we measured spent less than 61% of the study time within their census tracts of residence and a quarter spent less than 23% of the time in these tracts. Moreover, while in their census tracts of residence, subjects spent most of this time in their census *blocks* of residence—and much of it presumably in their homes (Basta et al. 2010). These subjects' census blocks of residence accounted for a small fraction of the areas of their census tracts of residence—less than 9% for three-fourths of them—and if we exclude time spent in the census blocks of residence, half of the subjects spent less than 4% of the time in their census tracts of residence and a quarter spent none of the time in these tracts. In other words, the residential census tracts are simultaneously over-inclusive and under-inclusive.

Mobile phone location tracking helps us overcome both problems—tightening measures of local environment to each individual's immediate surroundings while also letting these measures follow the individual as he or she moves through space. This gives us a dynamic “personal exposure” (Chaix et al. 2009) that includes much more information than residence-based measures. The racial composition of the populations our subjects encountered outside their home blocks differed notably from the racial composition of their home blocks themselves. Incorporating the non-home blocks into our analysis expanded our view of segregation and inter-racial exposure.

This is only part of what a moving local environment allows us to accomplish. We used census information on residential racial compositions as our environmental variable, but with a larger and more concentrated sample we could estimate the changing racial composition (or other characteristics) of each location based on the sampled individuals themselves. In this way, we would not only temporally refine our measure of exposure, but also illuminate the ways in which individuals choose the places they visit and modify the characteristics of these places through their presence and activities (Entwisle 2007).

Finally, just as we can tighten measures of local context around individuals as they move, so too can we tighten our measures of individuals' dynamic responses to this context. Our location-based survey made it possible to identify significant differences between male and female

¹⁵ See *supra* notes 7 and 8 and accompanying text. For an overview of masking techniques, see Armstrong et al. (1999), Gutmann et al. (2008).

subjects in terms of the relationship between subjective well-being and distance from home. That male subjects tended to respond differently (reporting lower happiness) the farther they were from home is additional evidence of the limits of residential measures. Had we surveyed these people only in their homes, our picture would have been much less complete.

Conclusions

Our work opens the door to new possibilities for incorporating mobile phones into demographic research. We explored the potential that these new technologies have for better understanding activity space, segregation outside the home, the geographic context of subjective-wellbeing and perhaps many other aspects of human behavior. That we were able to recruit 270 volunteers in 14 countries, examine their demographic characteristics and movement patterns, link these to census block data, and then have subjects respond in real time to surveys triggered by their locations, all in a low-cost and time-efficient manner, suggests the power of these new methodologies in a more broad-based inquiry.

Acknowledgements The authors thank Hazer Inaltekin, Spencer Lucian, and David Potere for their contributions to this project during its initial phases, and Matthew Salganik for his invaluable guidance and contributions throughout. The work was funded by a grant from the Center for Information Technology Policy at Princeton University. Institutional support was provided by National Institutes of Health Training Grant T32HD07163 and Infrastructure Grant R24HD047879.

References

- Adey, P. (2010). *Mobility*. London, UK: Routledge.
- Ahas, R. (2011). "Mobile Positioning." In M. Büscher, J. Urry, & K. Witchger (Eds.), *Mobile Methods* (pp. 183–199). London, UK: Routledge.
- Ahas, R. & Mark, Ü. (2005). "Location based services–new challenges for planning and public administration?" *Futures* 37, 547–561.
- Anderson, E. & Massey, D. S. (2001). *Problem of the century: Racial stratification in the United States*. New York: Russell Sage Foundation.
- Anokwa, Y., Hartung, C., Brunette, W., Borriello, G., & Lerer, A. (2009). "Open Source Data Collection in the Developing World." *Computer* 42(10), 97–99.
- Armstrong, M. P., Rushton, G., & Zimmerman, D. L. (1999). "Geographically masking health data to preserve confidentiality." *Statistics in medicine* 18, 497–525.
- Arnold, S. E. (2010). "Google and its strategy of 'meh'." *KM World* 19, 6–24.
- Asakura, Y. & Hato, E. (2001). "Behavioral monitoring of public transport users through a mobile communication system." *Journal of Advanced Transportation* 35, 289–304.
- Asakura, Y. & Hato, E. (2004). "Tracking survey for individual travel behaviour using mobile communication instruments." *Transportation Research Part C: Emerging Technologies* 12, 273–291.
- Barraquand, F. & Benhamou, S. (2008). "Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts." *Ecology* 89, 3336–3348.
- Bartumeus, F., Giuggioli, L., Louzao, M., Bretagnolle, V., Oro, D., & Levin, S. A. (2010). "Fishery Discards Impact on Seabird Movement Patterns at Regional Scales." *Current Biology* 20, 215–222.
- Basta, L. A., Richmond, T. S., & Wiebe, D. J. (2010). "Neighborhoods, daily activities, and measuring health risks experienced in urban environments." *Social Science and Medicine* 71, 1943–1950.
- Bertolini, L. & Dijst, M. (2003). "Mobility Environments and Network Cities." *Journal of Urban Design* 8, 27–43.
- Blumenstock, J., Gillick, D., & Eagle, N. (2010). "Who's Calling? Demographics of Mobile Phone Use in Rwanda." *AAAI Symposium on Artificial Intelligence and Development* 18, 116–117.
- Borrell, B. (2011). "Every bite you take." *Nature* 470, 320–322.
- Buliung, R. N. & Kanaroglou, P. S. (2006). "Urban Form and Household Activity-Travel Behavior." *Growth and Change* 37, 172–199.
- Büscher, Monika and John Urry. 2009. "Mobile Methods and the Empirical." *European Journal of Social Theory* 12:99–116.
- Büscher, M., Urry, J., & Witchger, K. (2011). "Introduction." In M. Büscher, J. Urry, & K. Witchger (Eds.), *Mobile Methods* (pp. 1–19). London, UK: Routledge.
- Calabrese, F., Colonna, M., Lovisolo, P.P.D., & Ratti, C. (2007). "Real-Time Urban Monitoring Using Cellphones: a Case-Study in Rome." *IEEE Transactions on Intelligent Transportation Systems* 12(1), 141–151.
- Calabrese, F., Pereira, F., Di Lorenzo, G., Liu, L., & Ratti, C. (2010). "The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events." In P. Floréen, A. Krüger, & M. Spasojevic (Eds.), *Pervasive Computing* (pp. 22–37). Berlin: Springer.
- Carioua, C., Ziemlickib, C., & Smoredab, Z. (2010). "Paris by Night." *NetMob 2010*, 62–66.

- Cassels, S., Curran, S. R., & Kramer, R. (2005). "Do Migrants Degrade Coastal Environments? Migration, Natural Resource Extraction and Poverty in North Sulawesi, Indonesia." *Human Ecology* 33, 329–363.
- Chaix, B., Merlo, J., Evans, D., Leal, C., & Havard, S. (2009). "Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur." *Social Science and Medicine* 69, 1306–1310.
- Chang, G., Tan, C., Li, G., & Zhu, C. (2010). "Developing mobile applications on the Android platform." In X. Jiang, M. Ma, & C. Chen (Eds.), *Mobile Multimedia Processing* (pp. 264–286). Berlin: Springer.
- Csikszentmihalyi, M. & Hunter, J. (2003). "Happiness in Everyday Life: The Uses of Experience Sampling." *Journal of Happiness Studies* 4, 185–199.
- Cummins, S., Curtis, S. Diez-Roux, A. V. & Macintyre, S. (2007). "Understanding and representing 'place' in health research: A relational approach." *Social Science and Medicine* 65, 1825–1838.
- D'Andrea, A., Ciolfi, L., & Gray, B. (2011). "Methodological Challenges and Innovations in Mobilities Research." *Mobilities* 6, 149–160.
- de Castro, M. C., Monte-Mór, R. L., Sawyer, D. O., & Singer, B. H. (2006). "Malaria risk on the Amazon frontier." *Proceedings of the National Academy of Sciences of the United States of America* 103, 2452–2457.
- Diener, E., Suh, E. M., Lucas, R. E., Smith, H. L. (1999). "Subjective well-being: Three decades of progress." *Psychological Bulletin* 125, 276–302.
- Du Bois, W.E.B. (1899). *The Philadelphia Negro: A social study*. Philadelphia: University of Pennsylvania.
- Eagle, N. & Pentland, A. (2006). "Reality mining: Sensing complex social systems." *Personal and Ubiquitous Computing* 10, 255–268.
- Eagle, N. & Pentland, A. (2009). "Eigenbehaviors: Identifying structure in routine." *Behavioral Ecology and Sociobiology* 63, 1057–1066.
- Eagle, N., Pentland, A., & Lazer, D. (2009a). "Inferring friendship network structure by using mobile phone data." *Proceedings of the National Academy of Sciences* 106, 15274–15278.
- Eagle, N., Quinn, J., & Clauset, A. (2009b). "Methodologies for Continuous Cellular Tower Data Analysis." In H. Tokuda, M. Beigl, A. Friday, A. Brush, Y. Tobe (Eds.), *Pervasive Computing* (pp. 342–353). Berlin: Springer.
- Ellis, M., Wright, R., & Parks, V. (2004). "Work together, live apart? Geographies of racial and ethnic segregation at home and at work." *Annals of the Association of American Geographers* 94, 620–637.
- Entwisle, B. (2007). "Putting people into place." *Demography* 44(4), 687–703.
- Foley, D. L. (1950). "The use of local facilities in a metropolis." *American Journal of Sociology* 56(3), 238–246.
- Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Golledge, R. G., & Stimson, R.J. (1997). *Spatial behavior: A geographic perspective*. New York: The Guilford Press.
- Golob, T. T. & Meurs, H. (1986). "Biases in response over time in a seven-day travel diary." *Transportation* 13, 163–181.
- Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A.-L. (2008). "Understanding individual human mobility patterns." *Nature* 453, 779–782.

- Goodchild, M. F. (2007). "Citizens as sensors: The world of volunteered geography." *GeoJournal* 69, 211–221.
- Goodchild, M. F. & Janelle, D. G. (1984). "The city around the clock: Space-time patterns of urban ecological structure." *Environment and Planning A* 16, 807–820.
- Gutmann, M. P., Witkowski, K., Colyer, C., O'Rourke, J. M., & McNally, J. (2008). "Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality." *Population Research and Policy Review* 27, 639–665.
- Hägerstrand, T. (1970). "What about people in regional science?" *Papers in Regional Science* 24, 6–21.
- Helliwell, J. F. (2003). "How's life? Combining individual and national variables to explain subjective well-being." *Economic Modelling* 20, 331–360.
- Inagami, S., Cohen, D. A., & Finch, B. K. (2007). "Non-residential neighborhood exposures suppress neighborhood effects on self-rated health." *Social Science and Medicine* 65, 1779–1791.
- Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Rowland, J., & Varshavsky, A. (2010). "A tale of two cities." In *Proceedings of the Eleventh Workshop on Mobile Computing Systems and Applications* (pp. 19–24). Annapolis, MD: ACM.
- Janelle, D. G., Klinkenberg, B., & Goodchild, M. F. (1998). "The Temporal Ordering of Urban Space and Daily Activity Patterns for Population Role Groups." *Geographical Systems* 5, 117–137.
- Kahneman, D. & Krueger, A. B. (2006). "Developments in the measurement of subjective well-being." *Journal of Economic Perspectives* 20, 3–24.
- Kain, J. F. (1968). "Housing segregation, Negro employment, and metropolitan decentralization." *The Quarterly Journal of Economics* 82, 175–197.
- Killingsworth, M. A. & Gilbert, D. T. (2010). "A Wandering Mind Is an Unhappy Mind." *Science* 330(6006), 932.
- Kwan, M.-P. (1999). "Gender, the Home-Work Link, and Space-Time Patterns of Nonemployment Activities." *Economic Geography* 75, 370–394.
- Kwan, M.-P. (2008). "From oral histories to visual narratives: Re-presenting the post-September 11 experiences of the Muslim women in the USA." *Social and Cultural Geography* 9, 653–669.
- Kwan, M.-P. (2009). "From place-based to people-based exposure measures." *Social Science and Medicine* 69, 1311–1313.
- Kwan, M.-P. & Lee, J. (2004). "Geovisualization of human activity patterns using 3D GIS: a time-geographic approach." In M. F. Goodchild & D. G. Janelle (Eds.) *Spatially integrated social science: Examples in Best Practice* (pp. 48–66). Oxford University Press.
- Laube, P., Dennis, T., Forer, P., & Walker, M. (2007). "Movement beyond the snapshot: Dynamic analysis of geospatial lifelines." *Computers, Environment and Urban Systems* 31, 481–501.
- Laube, P., Imfeld, S., & Weibel, R. (2005). "Discovering relative motion patterns in groups of moving point objects." *International Journal of Geographical Information Science* 19, 639–668.
- Law, J. & Urry, J. (2004). "Enacting the social." *Economy and society* 33, 390–410.
- Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). "Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales." *American Sociological Review* 73, 766–791.

- Lee, J. Y. & Kwan, M.-P. (2011). "Visualisation of Socio-Spatial Isolation Based on Human Activity Patterns And Social Networks in Space-Time." *Tijdschrift voor economische en sociale geografie* [Journal of Economic and Social Geography] 102, 468–485.
- Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). "The BUGS project: Evolution, critique, and future directions." *Statistics in Medicine* 28(25), 3049–3067.
- MacKerron, G. (2011). "Happiness Economics From 35,000 Feet." *Journal of Economic Surveys*, doi: 10.1111/j.1467-6419.2010.00672.x.
- Madan, A., Cebrian, M., Lazer, D., & Pentland, A. (2010). "Social Sensing for Epidimiological Behavior Change." In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing* (pp. 291–300). New York: ACM.
- Martens, P. & Hall, L. (2000). "Malaria on the move: Human population movement and malaria transmission." *Emerging Infectious Diseases* 6, 103–109.
- Massey, D. S. & Denton, N. A. (1988). "The Dimensions of Residential Segregation." *Social Forces* 67, 281–315.
- Massey, D. S. & Denton, N. A. (1993). *American apartheid: Segregation and the making of the underclass*. Harvard University Press.
- Matthews, S. A. (2008). "The salience of neighborhood: Some lessons from sociology." *American Journal of Preventive Medicine* 34(3), 257–259.
- Matthews, S. A. (2011). "Spatial polygamy and the heterogeneity of place: Studying people and place via egocentric methods." In L. M. Burton, S. P. Kemp, M. Leung, S. A. Matthews, & D. T. Takeuchi (Eds.), *Communities, Neighborhoods, and Health: Expanding the Boundaries of Place* (pp. 35–55). New York: Springer.
- Murakami, E. & Wagner, D. P. (1999). "Can using Global Positioning System (GPS) improve trip reporting?" *Transportation Research Part C: Emerging Technologies* 7, 149–165.
- Nusser, S. M., Intille, S. S., & Maitra, R. (2006). "Emerging technologies and next-generation intensive longitudinal data collection." In T. A. Walls & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 254–277). New York: Oxford University Press.
- Ökubo, A. & Levin, S. A. (2001). *Diffusion and ecological problems: Modern perspectives*. New York: Springer.
- Park, R. E. & Burgess, E. W. (1925). *The city*. University of Chicago Press.
- Pentland, A. (2007). "Automatic mapping and modeling of human networks." *Physica A: Statistical Mechanics and its Applications* 378, 59–67.
- Raento, M., Oulasvirta, A., & Eagle, N. (2009). "Smartphones: An Emerging Tool for Social Scientists." *Sociological Methods and Research* 37, 426–454.
- Ratti, C., Sevtsuk, A., Huang, S., & Pailer, R. (2007). "Mobile landscapes: Graz in real time." In G. Gartner, W. Cartwright, & M. P. Peterson (Eds.), *Location Based Services and TeleCartography* (pp. 433–444). Berlin: Springer.
- Ratti, C., Williams, S., Frenchman, D., & Pulselli, R. M. (2006). "Mobile Landscapes: using location data from cell phones for urban analysis." *Environment and Planning B: Planning and Design* 33, 727–748.
- Reades, J., Calabrese, F., & Ratti, C. (2009). "Eigenplaces: Analysing cities using the space-time structure of the mobile phone network." *Environment and Planning B: Planning and Design* 36, 824–836.
- Reades, J., Calabrese, F., Sevtsuk, A., & Ratti, C. (2007). "Cellular census: Explorations in urban data collection." *IEEE Pervasive Computing* 6(3), 30–38.

- Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, S. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). "The Geographic Scale of Metropolitan Racial Segregation." *Demography* 45(3), 489–514.
- Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2010). "Using mobile phones to determine transportation modes." *ACM Transactions on Sensor Networks* 6(2), 13:1–13:27.
- Sheller, M. & Urry, J. (2006). "The new mobilities paradigm." *Environment and Planning A* 38, 207–226.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). "Ecological momentary assessment." *Annual Review of Clinical Psychology* 4, 1–32.
- Stopher, P., FitzGerald, C., & Xu, M. (2007). "Assessing the accuracy of the Sydney Household Travel Survey with GPS." *Transportation* 34, 723–741.
- Urry, J. (2007). *Mobilities*. Cambridge, UK: Polity.
- VanWey, L. K., Rindfuss, R. R., Gutmann, M. P., Entwisle, B., & Balk, D. L. (2005). "Confidentiality and spatially explicit data: Concerns and challenges." *PNAS* 102(43), 15337–15342.
- Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
- Wang, D., Li, F., & Chai, Y. (2012). "Activity spaces and sociospatial segregation in Beijing." *Urban Geography* 33, 256–277.
- Wesolowski, A. P. & Eagle, N. (2010). "Parameterizing the Dynamics of Slums." In *Proceedings of the AAAI Artificial Intelligence for Development Symposium* (pp. 103–108).
- Wiehe, S. E., Carroll, A. E., Liu, G. C., Haberkorn, K. L., Hoch, S. C., Wilson, J. S., & Fortenberry, J. D. (2008). "Using GPS-enabled cell phones to track the travel patterns of adolescents." *International Journal of Health Geographics*, doi:10.1186/1476-072X-7-22
- Wilson, W. J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.
- Wilson, W. J. (1996). *When work disappears: The world of the urban poor*. New York: Alfred Knopf.
- Wong, D. W. S. & Shaw, S.-L. (2011). "Measuring segregation: An activity space approach." *Journal of geographical systems* 13, 127–145.
- Yarow, J. (2011). "GARTNER: Android Market Share Doubles, iOS Drops In Q3." *Business Insider*. http://articles.businessinsider.com/2011-11-15/tech/30400455_1_ios-iphone-smartphone-market. Accessed 30 April 2012.
- Zandvliet, R. & Dijst, M. (2006). "Short-term Dynamics in the Use of Places: A Space-Time Typology of Visitor Populations in the Netherlands." *Urban Studies* 43, 1159–1176.
- Zenk, S. N., Schulz, A. J., Israel, B. A., James, S. A., Bao, S., & Wilson, M. L. (2005). "Neighborhood Racial Composition, Neighborhood Poverty, and the Spatial Accessibility of Supermarkets in Metropolitan Detroit." *American Journal of Public Health* 95, 660–67.

New Approaches to Human Mobility: Using Mobile Phones for Demographic Research

TABLES AND FIGURES

Table 1 Race and sex of subjects.

Race	Female		Male		Missing		Total	
	N	%	N	%	N	%	N	%
Asian	1	0.37	15	5.56	0	0.00	16	5.93
Black	10	3.70	7	2.59	0	0.00	17	6.30
Latino	14	5.19	19	7.04	0	0.00	33	12.22
White	65	24.07	127	47.04	0	0.00	192	71.11
Missing/Other	7	2.59	4	1.48	1	0.37	12	4.44
Total	97	35.93	172	63.70	1	0.37	270	100.00

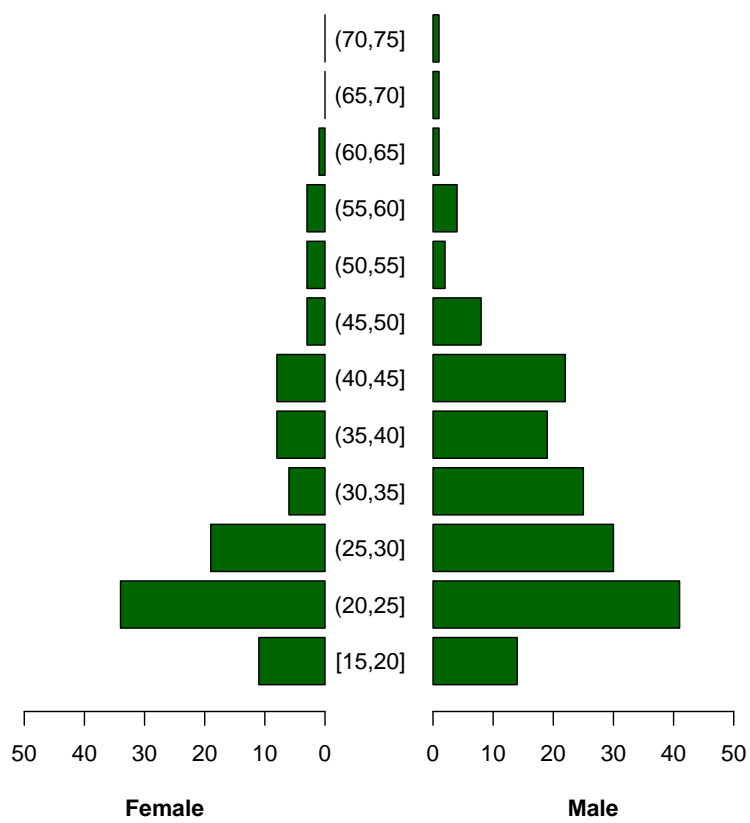


Fig. 1 Age and sex of subjects. Number of subjects is shown for each age group and sex.

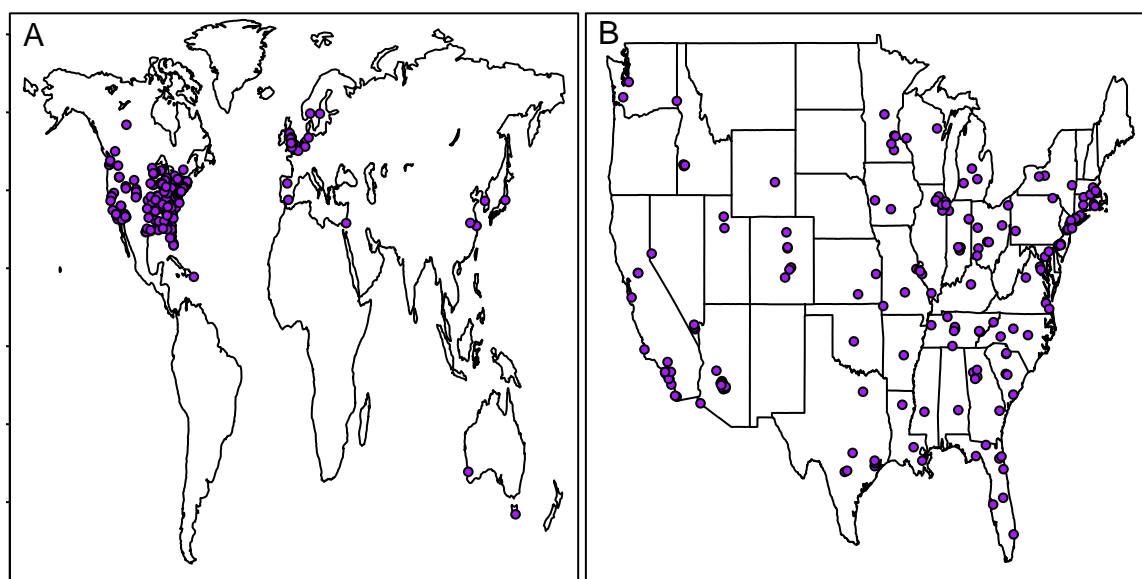


Fig. 2 Centroids of each subject's location estimates on world (A) and U.S. (B) maps. A small amount of random noise has been added to each location estimate to protect privacy and aid in visualization of high density areas.

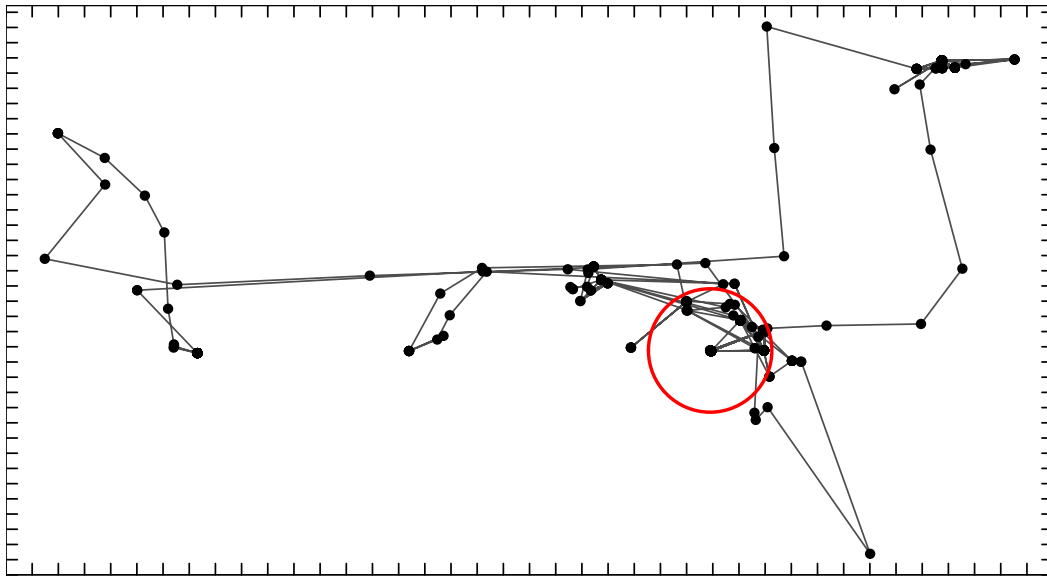


Fig. 3 Individual movement pattern of Subject 1. Straight lines connect sequential location estimates (points) and the red circle marks estimated place of residence (circle's center). Many location estimates lie on top of one another and so cannot be distinguished. Tick-marks are 1 km apart. This subject's total participation time was 30 days.

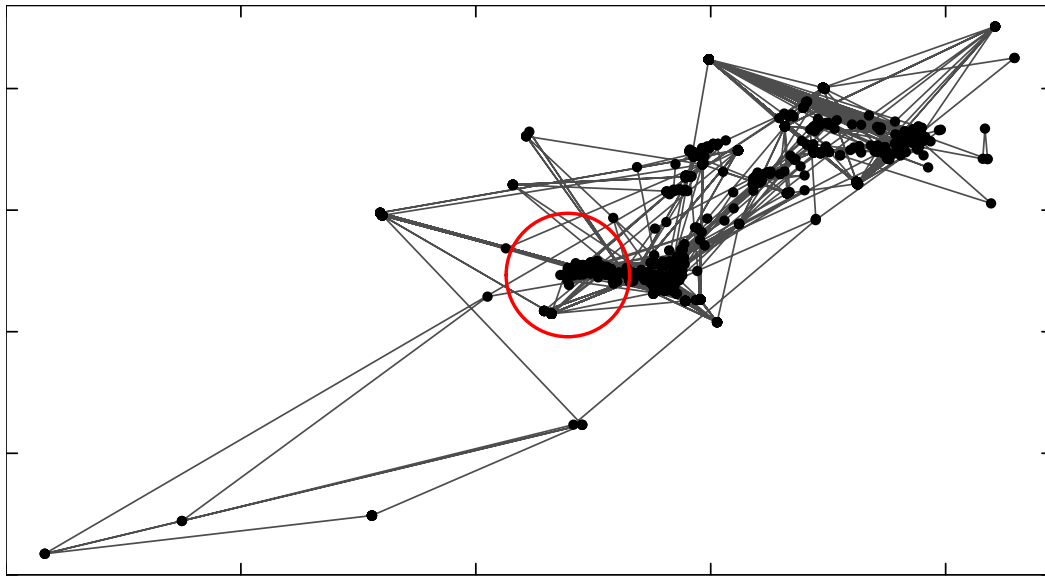


Fig. 4 Individual movement pattern of Subject 2. Straight lines connect sequential location estimates (points) and the red circle marks estimated place of residence (circle's center). Many location estimates lie on top of one another and so cannot be distinguished. Tick-marks are 1 km apart. This subject's total participation time was 31 days.

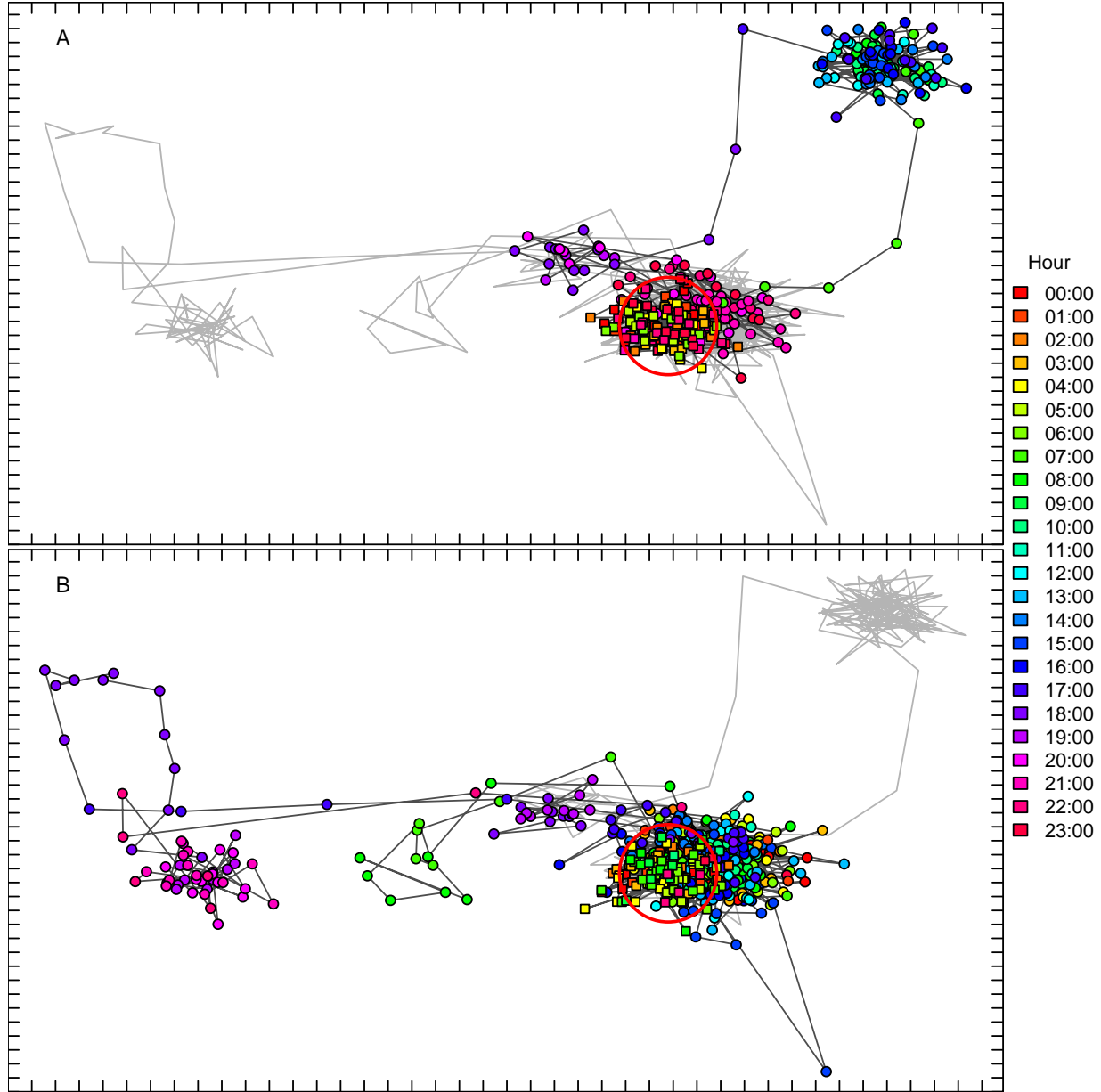


Fig. 5 Location estimates of Subject 1 on weekdays (A) and weekends (B) with random noise added to aid in visualization. Lines connecting sequential location estimates during both weekdays and weekends are displayed in both plots in light grey. In Panel A, weekday location estimates are plotted with colored markers and dark grey lines connect sequential estimates. In Panel B weekend location estimates are plotted with colored markers, and dark grey lines connect sequential estimates. In both cases, filled circles indicate locations outside subject's estimated census block of residence and filled squares indicate those within this block. The red circle marks estimated place of residence (circle's center). Axis tick-marks are 1 km apart. This subject's total participation time was 30 days

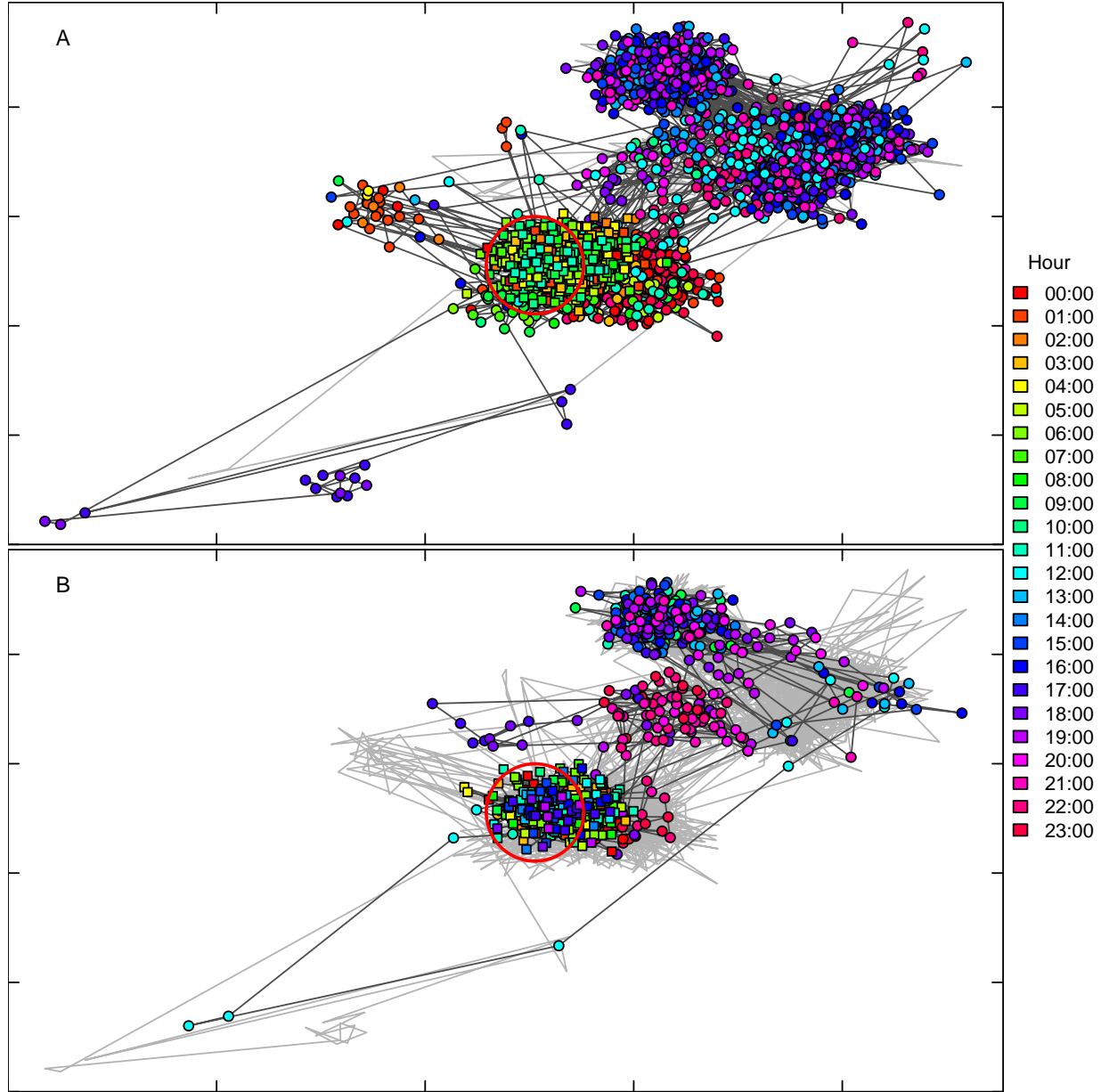


Fig. 6 Location estimates of Subject 2 on weekdays (A) and weekends (B) with random noise added to aid in visualization. Lines connecting sequential location estimates during both weekdays and weekends are displayed in both plots in light grey. In Panel A, weekday location estimates are plotted with colored markers and dark grey lines connect sequential estimates. In Panel B weekend location estimates are plotted with colored markers, and dark grey lines connect sequential estimates. In both cases, filled circles indicate locations outside subject's estimated census block of residence and filled squares indicate those within this block. The red circle marks estimated place of residence (circle's center). Axis tick-marks are 1 km apart. This subject's total participation time was 31 days

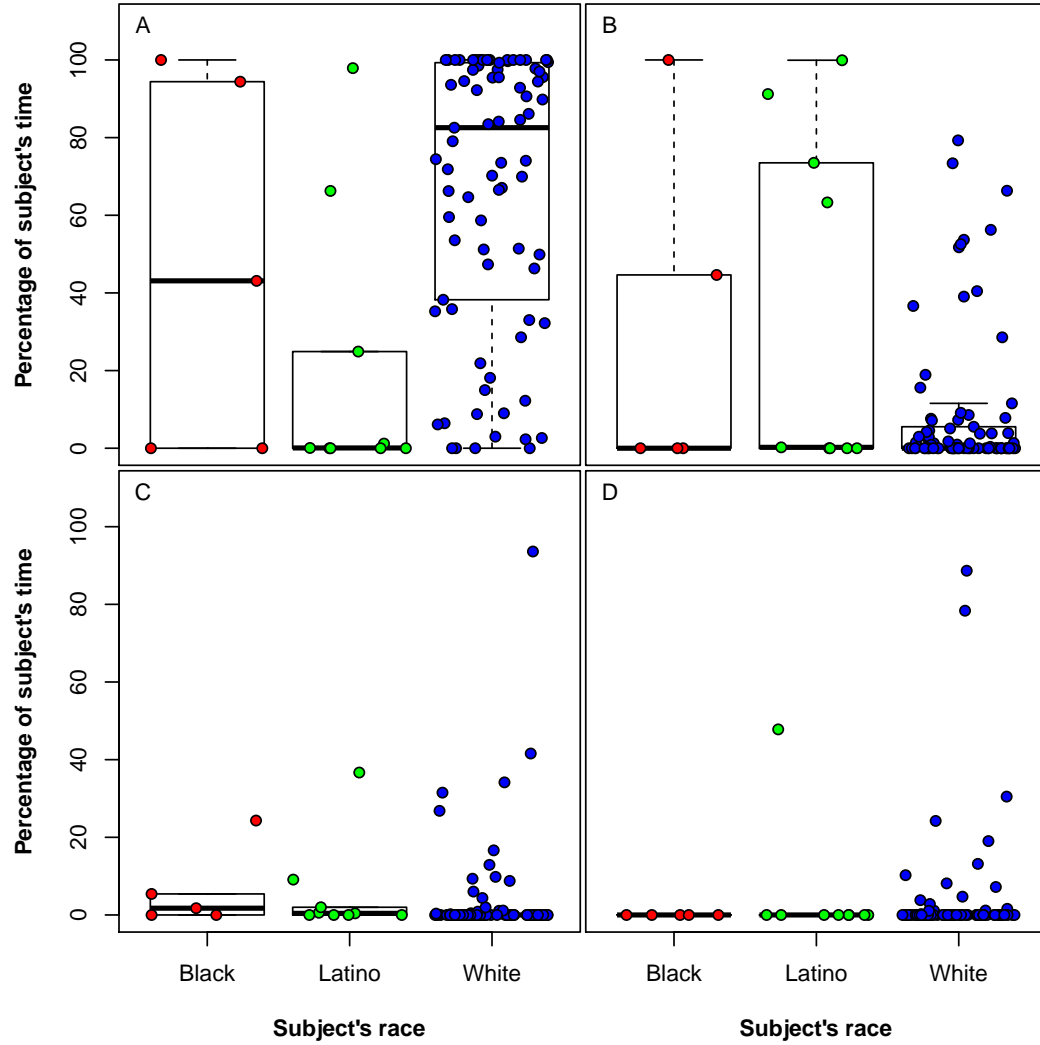


Fig. 7 Box-whisker plots of percentage of time that black, Latino, and white subjects spent in census blocks in which the proportion of white residents was $(0.75, 1]$ (A), $(0.50, 0.75]$ (B), $(0.25, 0.50]$ (C), or $[0, 0.25]$ (D), based on the 2000 census. Boxes show medians and interquartile ranges (calculated as Tukey's lower and upper hinges). Each whisker extends to the maximum and minimum values lying at a distance of up to 1.5 times the length of the interquartile range. All data points are displayed (with a small amount of noise added to their x-coordinates to avoid overlap), rather than simply the outliers, for better visualization of different sample sizes and distributions. Figures exclude subjects with less than 4 hours observed outside of home and they also exclude time spent in the census block of subject's residence. Not displayed here is time spent in census blocks with total census populations of zero.

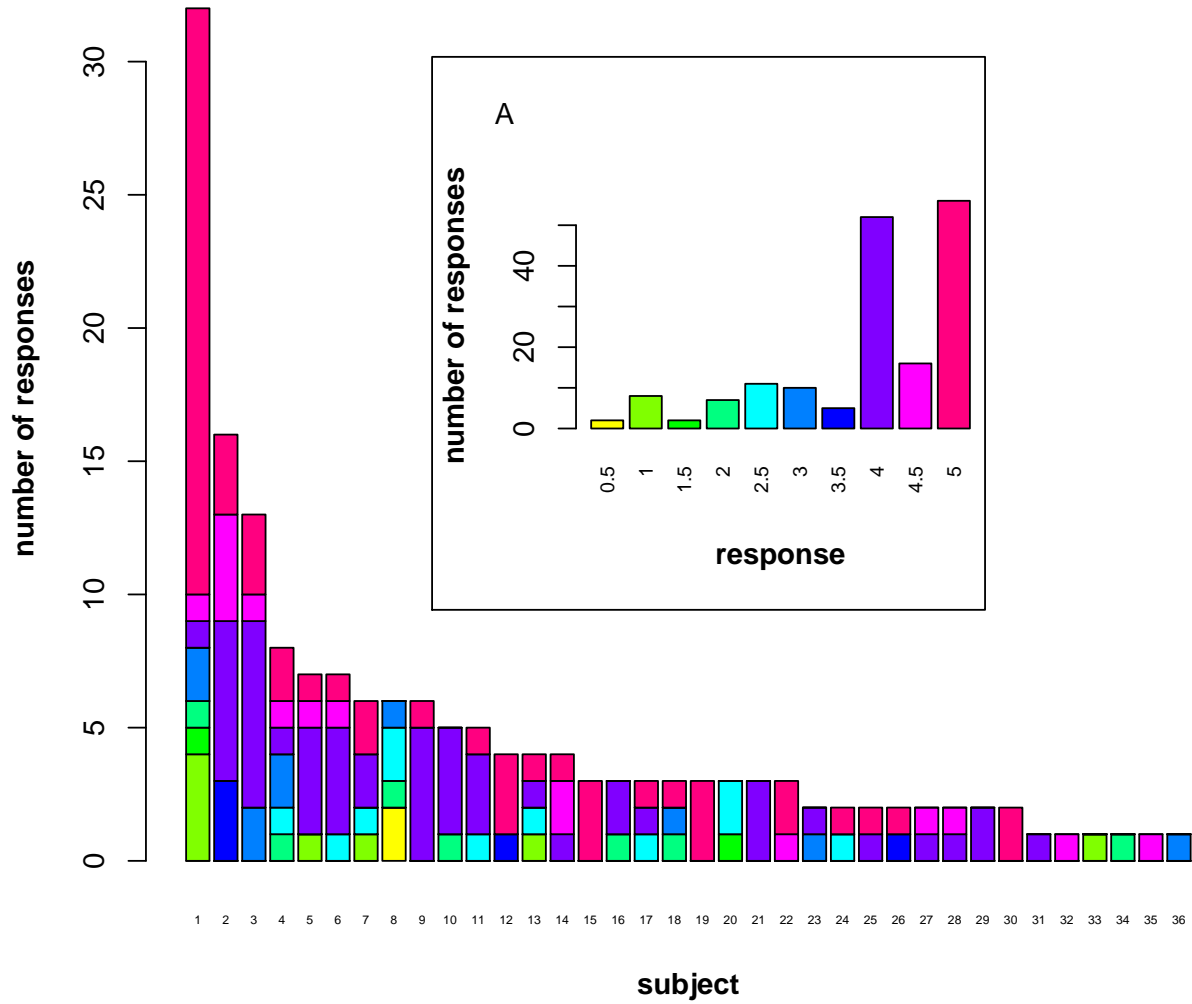


Fig. 8 Survey responses analyzed in models with distance from home as a covariate and without census block racial characteristics. Main plot shows responses grouped by research subject, with colors corresponding to the pooled responses shown in inset A.

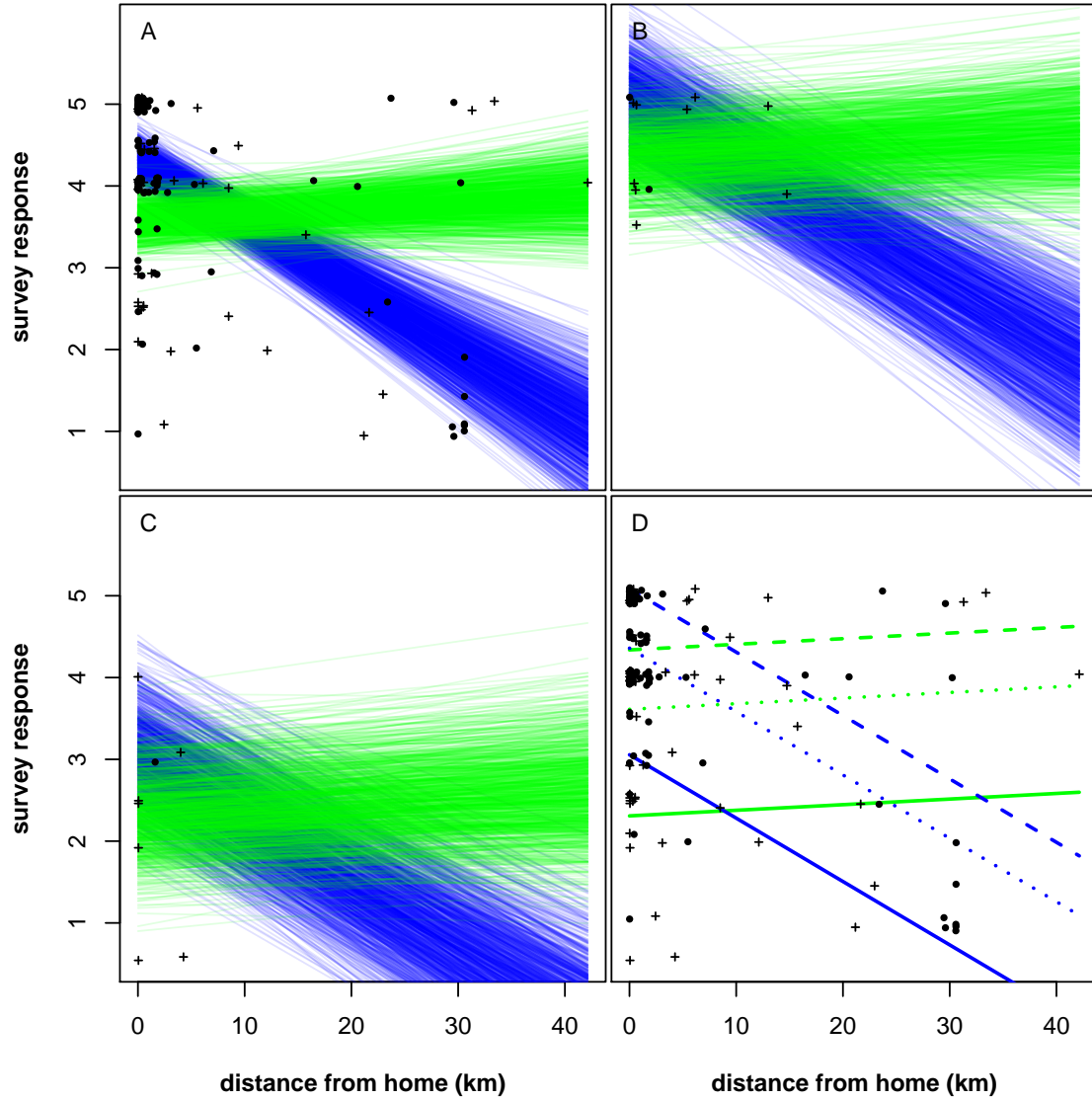


Fig. 9 Observed and predicted happiness by distance from home for white (A), Latino (B), and black (C) men (crosses and blue lines) and women (filled circles and green lines). Lines in Panels A-C represent uncertainty and are drawn using coefficients estimated in each of 999 final iterations of the Gibbs sampler. Lines in Panel D are drawn using mean estimates for each gender and race (dotted = white, dashed = Latino, solid = black).