

Experienced Segregation

Susan Athey, *Stanford University and NBER**

Billy Ferguson, *Stanford University*

Matthew Gentzkow, *Stanford University and NBER*

Tobias Schmidt

July 2020

Abstract

We introduce a novel measure of segregation, *experienced isolation*, that captures individuals' exposure to diverse others in the places they visit over the course of their days. Using Global Positioning System (GPS) data collected from smartphones, we measure experienced isolation by race. We find that the isolation individuals experience is substantially lower than standard residential isolation measures would suggest, but that experienced and residential isolation are highly correlated across cities. Experienced isolation is lower relative to residential isolation in denser, wealthier, more educated cities with high levels of public transit use, and is also negatively correlated with income mobility.

*E-mail: athey@stanford.edu, billyf@stanford.edu, gentzkow@stanford.edu, tobyschmidt@web.de. We thank Jonathan Dingel, Jessie Handbury, and numerous seminar participants for helpful inputs and suggestions. We also thank our many dedicated research assistants for their contributions to this project. We acknowledge funding from the Stanford Institute for Economic Policy Research (SIEPR).

1 Introduction

Social outcomes are profoundly shaped by the extent to which groups are segregated from one another (Cutler and Glaeser 1997; Chetty and Hendren 2018a, 2018b; Chetty et al. 2016). As a result, large literatures have developed in economics, sociology, and related fields seeking to measure the extent of segregation across space and time.

Most of this empirical work focuses on segregation in where people live. A leading measure is the isolation index, which captures the share of individuals' neighbors who come from their own group.¹ If we view the object of interest as the exposure of one group to another (Massey and Denton 1988; Cutler et al. 1999; Echenique and Fryer 2007), residential measures have obvious limitations. Individuals living in highly segregated neighborhoods may be exposed to diverse others where they work, shop, and socialize, while those living in apparently mixed neighborhoods may have little contact with their neighbors and commute to highly segregated places. A corollary is that standard residential segregation measures are highly sensitive to the way in which neighborhood boundaries are defined (Cowgill and Cowgill 1951; Massey and Denton 1988).

In this paper we introduce a novel measure of segregation which addresses these limitations, and estimate it using Global Positioning System (GPS) data. This *experienced isolation* has the same form as the isolation index, but rather than assuming individuals are exposed uniformly to those in their neighborhood of residence, it averages exposure over the places individuals actually visit over the course of their days. This measure does not depend on arbitrary neighborhood boundaries, and it takes explicit account of the diversity experienced away from home. It can capture individual-level heterogeneity within neighborhoods (Echenique and Fryer 2007), and it can be disaggregated across times of day, locations, and activities, thus giving a richer picture of the forces that increase or decrease segregation.

Our main data are GPS signals from a sample of US smartphone users covering approximately 5% of the US population in the first four months of 2017. The data are obtained from a company that aggregates anonymous pings from a range of smartphone apps. We observe each device's home location as well as the location of every ping by the device recorded in the data. We map these locations to a grid of geographic units approximately 500 feet square, known as geohash7s. The sample of individuals is not random but is reasonably close to representative

¹See, for example, Cutler and Glaeser (1997), Cutler et al. (1999), Gentzkow and Shapiro (2011), and Davis et al. (2019).

along a number of dimensions, and has sufficient coverage that we can correct for deviations from representativeness using sample weights. We use the movement patterns we observe to compute experienced racial isolation.

Because we do not observe an individual's race directly, we define the two types whose segregation we study to be individuals with homes in majority white geohash7s and individuals with homes in majority non-white geohash7s. We refer to these two groups as WDs (White home geohash7 Devices) and NWDs (Non-White home geohash7 Devices) for simplicity. The median share white of majority white and non-white home geohash7s are 0.89 and 0.22, respectively. We discuss below the implications of using these geographic definitions in place of individual race, and we show robustness to an alternative strategy that imputes race at the individual level.

We present four main results: First, peoples' actual experiences as captured by our measure are substantially less segregated than traditional residential isolation would suggest. The average experienced isolation across all Metropolitan Statistical Areas (MSAs) is 0.46, compared to average *residential* isolation of 0.61.² This implies that the share of WD's exposures to other WDs is 46 percentage points greater than the share of NWD's exposures to WDs.

Second, experienced and residential isolation across MSAs are highly correlated. The overall correlation of the two measures among the 366 MSAs in our sample is 0.86. Among the 50 most populous MSAs, Milwaukee, WI; Detroit, MI; and Cleveland, OH rank in the top 5 in both residential and experienced isolation. Portland, OR; Seattle, WA; and Raleigh, NC rank in the bottom 5 for both measures.

Third, the variation in experienced relative to residential isolation is systematic. Experienced isolation is relatively lower in MSAs with higher population density and public transit use, consistent with the view that urban areas facilitate diverse interactions (Jacobs 1961). Experiences are also less isolated in MSAs with higher income and education and lower unemployment, possibly reflecting a role for social capital in reducing segregation (Putnam 2000). Finally, relative experienced isolation is negatively correlated with Chetty et al.'s (2014) measure of income mobility, consistent with both diverse interactions increasing mobility and with areas that facilitate opportunity also promoting diverse interactions.

Fourth, decompositions across time and space reveal the extent to which different activities increase or decrease segregation. Experienced isolation is lowest during the day and highest in

²Residential isolation based on our geographic definition of WD and NWD is larger than the standard measure of residential isolation based on individual race. We discuss the reasons for this difference below.

the morning and evening. Experienced isolation in home neighborhoods is *higher* than residential measures would suggest, whereas experienced isolation outside of home neighborhoods is much lower. Isolation is lowest at entertainment, retail, and eating establishments, while time at locations like churches and schools is somewhat more isolated.

These findings have several broader implications. They suggest that standard measures overstate the overall extent of segregation in the United States, and they highlight important forces such as commercial activity that reduce it. They suggest that residential measures may nevertheless be a good proxy when the main goal is to assess relative levels of segregation across cities. Finally, they suggest a more nuanced view of where the negative effects of segregation are likely to be largest. For example, local public goods such as schools or police services that are explicitly tied to residential boundaries are more likely to be provided in segregated environments. Any negative effects of segregation are likely higher for children and those who do not work, and others whose exposure is more tied to their local neighborhoods. Policies which affect the spatial distribution of commercial or leisure activities, or the transportation cost of accessing these activities, may be as or more effective than policies explicitly targeting housing.

We emphasize three main limitations of our analysis. First, we have no direct information about the individuals whose devices we see in our data, and so we define individual types based on the demographic composition of home geohash7s rather than based on individual race. This means we are targeting a slightly different concept than much of the prior literature on segregation. We discuss alternative approaches including imputing race at the individual level in the Online Appendix. Second, our sample is not fully representative, and the geolocation information we get about any given device is sparse. Third, while we can observe when devices occupy the same geographic space, we can not directly observe actual interaction between individuals. Under our construction, a restaurant-goer is just as exposed to the waiter or the cook in the kitchen as she is to the person sitting across the table. White (1983) highlights this subtlety by distinguishing geographic segregation (the concept we measure) and sociological segregation (based on actual interactions). Sunstein (2002), among others, argues that geographic segregation is of interest on its own.³

This paper builds on a large literature on measuring urban segregation. Important early work

³Sunstein (2002) writes that integrated physical spaces increase “the set of chance encounters with diverse others” and foster environments where “exposure is shared.” He argues that overhearing conversations while at a restaurant, a bus stop, or just walking down the street all contribute to individuals’ understanding of diverse others and open up opportunities for interaction.

on both the definition and measurement of segregation includes Duncan and Duncan (1955), Taeuber and Taeuber (1965), White (1983), Massey and Denton (1988), and Massey and Denton (1993). Cutler et al. (1999) provide a comprehensive analysis of segregation in US cities over the century from 1890 to 1990. Card et al. (2008) study the dynamics of neighborhood tipping.⁴ Our work is also related to a growing literature using GPS or similar location data to study social interactions.⁵

2 Data

2.1 Geography

We follow the literature in characterizing segregation at the level of MSAs and in using census tracts to approximate neighborhoods within MSAs.⁶ The finest geographic unit in our analysis is the geohash⁷, which as mentioned above is a unit of a grid roughly 500 feet square.⁷ We use census blocks to impute geohash⁷ demographics. Appendix Figures A1, A2, and A3 illustrate the relative sizes of geohash⁷s, census blocks, and census tracts, focusing on an urban census tract and a rural census tract respectively in Birmingham, AL.

We obtain information about the location of establishments and features of interest from two sources: InfoUSA and OpenStreetMaps (OSM). The 2015 InfoUSA US Businesses mailing list contains the names, addresses, industries, and latitude / longitude for 15.6 million businesses in the United States. We take from the full list all establishments that belong to the broad categories of “restaurants and bars,” “civil, social and religious organizations,” “accommodation,” “sports and recreation,” “entertainment,” and “retail,”⁸ 2,368,216 places all in all. We match each establishment to the geohash⁷s that contain its latitude / longitude. From OSM, we extract

⁴Park and Kwan (2018) define a notion of “multi-contextual segregation” that is closely related to our work in considering segregation over the varying geographic and temporal contexts of peoples’ daily lives.

⁵Glaeser et al. (2018) anticipate the value of such data. Blattman et al. (2018) track police patrols in Bogotá, Colombia using GPS to estimate how increased state presence affects violent and property crime. Chen and Rohla (2018) and Chen et al. (2019) use GPS data to measure the effects of political polarization on the length of Thanksgiving dinners and to measure racial differences in waiting times at polling places respectively. Davis et al. (2019) use data from Yelp to measure the segregation of restaurants in New York City, finding that restaurants are less segregated than residential neighborhoods. Caetano and Maheshri (2019) use data provided by the app Foursquare to quantify segregation by gender and by age in public places, and Phillips et al. (2019) use geotagged tweets to build an index capturing the extent to which residents in each neighborhood of a city travel to all other neighborhoods in equal proportions.

⁶We omit Micropolitan Statistical Areas.

⁷The geohash geocoding scheme divides the globe into grids of increasing fineness. Geohash1s divide the globe into 32 cells of equal size. Geohash2s divide each of these cells into 32 smaller cells, and so on.

⁸See Appendix Section A1 for our manual classification of NAICS code into these categories.

polygon data for outdoor spaces like parks, playgrounds, sports fields and gardens, and educational institutions like schools, kindergartens, universities and colleges (See Appendix Section 2.1 for details). We associate each OSM feature with all geohash7s that intersect the feature’s polygon. Appendix Figure A4 depicts geohash7s associated with civil, social, and religious organizations, education, outdoor spaces and restaurants and bars in downtown Birmingham, AL.

Many geohash7s are labelled with multiple features. We assume pings in a device’s home geohash7 (defined below) are at home regardless of what other features are present. We assign all pings in non-home geohash7s that contain transportation features to transportation. All other pings are allocated uniformly across features present in the geohash7.

2.2 GPS Device Movements

Our GPS data are provided by a company that collects anonymous location data from mobile applications on users’ smartphones. The sample is an unbalanced panel of GPS “pings” from more than 17 million devices spanning January to April 2017.⁹ Pings are logged whenever an application on a device requests location information. In some cases this will be the result of a device actively using an application, such as for navigation or weather information, while in other cases applications may request the information even while running in the background. Pings thus occur at irregular intervals. For each ping, we observe a timestamp, a device identifier, and the geohash7 in which the ping occurs. The data also contain the geohash7 of each device’s home, inferred probabilistically from the device’s nighttime and early-morning pings.

2.3 Demographics

We impute geohash7 demographics from the 2010 census. We match each home geohash7 to the census tract that contains its centroid. This yields a matching tract for 99.53 percent of devices in our sample. We match each home geohash7 to all census blocks that overlap its area. This yields a match to at least one census block with non-zero population for 98.12 percent of devices. We assign demographics to each home geohash7 by taking an area-weighted average of the demographics of the overlapping blocks.¹⁰ We define “white” population based on the

⁹We use “GPS” as a shorthand for a variety of means used by smartphones to determine their physical location. These include cell phone towers, the identity of nearby WiFi networks as well as the US GPS and the Russian GLONASS systems of satellites.

¹⁰We show robustness to alternative methods of demographic imputation in Appendix A2.2.

census designation “White Alone (Non-Hispanic),” and we group all other census race groups in the category “non-white.”

We use data on MSA characteristics from the 2010 American Community Survey (ACS) and the 2010 decennial census. These variables include the MSA’s median age, education level, unemployment rate, median income, population density, and share of residents using public transit to get to work.¹¹ We also use economic mobility measures from Chetty et al. (2020) indicating the share of individuals born to parents at the 25th percentile of the income distribution who make it to the top quintile for white and black populations. We compute MSA-level mobility measures by averaging across counties weighting by white and black county populations respectively.

2.4 Summary Statistics

We observe 17,730,615 devices with home locations identified in 7,292,623 distinct geohash7s. We match these home geohash7s to 72,785 census tracts and 6,186,564 census blocks. This matching procedure succeeds for 17,397,580 devices, the final sample used throughout the rest of the paper.

To assess the representativeness of the sample, we compare the average census tract demographics of devices in our sample to averages in the US population. We find that our sample is representative in terms of gender, age, and unemployment rate. We find that it slightly oversamples more educated and wealthy areas, with average median income across census tracts in our sample about a thousand dollars more than the U.S. mean, and census tract poverty rate about a percentage point lower. We address this imbalance by weighting as shown in equation 4. Details of this comparison and summaries of the average activity levels of devices in our sample are shown in Appendix Tables A2 and A3 respectively.

While our WD and NWD designations are not equivalent to individual race, they are highly correlated with it. The median share white in a device’s home geohash7 is 0.22 for NWDs and 0.89 for WDs. We plot the histogram of this share for both groups in Appendix Figure A7.

¹¹See Appendix Table A4 for a complete description and sources for census, ACS, and mobility variables.

3 Measure

3.1 Definition

Consider a population of individuals indexed by i and a set of MSAs or other geographic areas of interest indexed by a . We collect each individual who is a member of one of two groups which we denote W and NW . In our analysis below, W will be individuals from majority white geohash7s (WDs) and NW will be individuals from majority non-white geohash7s (NWDs). Each individual has a set of *exposures* to other individuals in area a . We let $e_i \in [0, 1]$ denote the share of individual i 's exposures that are to members of group W .¹²

A general form of the *isolation index* for area a captures the difference between the average value of e_i among individuals in the two groups (cf. Gentzkow and Shapiro 2011):

$$I_a = \frac{1}{|W_a|} \sum_{i \in W_a} e_i - \frac{1}{|NW_a|} \sum_{i \in NW_a} e_i. \quad (1)$$

Here W_a and NW_a are the sets of individuals making up the two groups in area a and $|\cdot|$ denotes the size of these sets. This measure ranges from zero—no isolation, with average e_i equal for the two groups—to one—perfect isolation, with $e_i = 0$ for all $i \in NW$ and $e_i = 1$ for all $i \in W$.

The standard version of this measure is *residential isolation*, which is equivalent to equation (1) under the assumption that each individual is exposed uniformly to others in her neighborhood of residence (Massey and Denton 1988; Cowgill and Cowgill 1951; Jahn 1950). In practice neighborhoods are typically defined to be census tracts. Letting $c(i)$ denote i 's census tract of residence, and letting r_c denote the share of the residents of tract c who are in group W , residential isolation is given by:¹³

$$RI_a = \frac{1}{|W_a|} \sum_{i \in W_a} r_{c(i)} - \frac{1}{|NW_a|} \sum_{i \in NW_a} r_{c(i)}. \quad (2)$$

¹²In our empirical analysis, we focus on the case where the groups W and NW partition the population, so that $1 - e_i$ is individual i 's exposure to members of group NW . Our measure is also well-defined in the case where some individuals in the population are neither in W nor NW . In this case, isolation where e_i is the share exposed to W may be different from isolation had we defined e_i as the share exposed to NW .

¹³This form of the isolation index is equivalent to Gentzkow and Shapiro (2011). Much of the literature using the isolation index studies simply the exposures of a group, without taking their difference. (White 1986, Iceland et al. 2002, Echenique and Fryer 2007). Massey and Denton (1988) provides a survey of other measures meant to encapsulate various qualitative aspects of segregation, and motivates our decision to capture segregation by measuring exposure.

Because this measure does not rely on any information other than the racial composition of each neighborhood, it can easily be computed using aggregate census data.

The new measure we introduce, *experienced isolation*, instead assumes that e_i is given by the composition of the individuals actually present in the locations that i visits over time. We index time by $t \in [0, 1]$ and consider a finite set of locations within area a indexed by l . We think of a location l as a specific place such as a restaurant, workplace, or park that is much smaller than a neighborhood. In our application, locations will be geohash7s. Letting $l(i, t)$ denote i 's location at time t , and letting $s(l, t)$ denote the share of individuals in location l at time t who are from group W , experienced isolation is defined to be:

$$EI_a = \frac{1}{|W_a|} \sum_{i \in W_a} \int_{t=0}^1 s(l(i, t), t) dt - \frac{1}{|NW_a|} \sum_{i \in NW_a} \int_{t=0}^1 s(l(i, t), t) dt. \quad (3)$$

3.2 Estimation

Estimating experienced isolation EI_a would be straightforward if we observed continuous location data for all individuals. While our GPS dataset is rich, it still falls well short of this ideal. There are two key limitations: (1) we observe locations only when a device pings rather than continuously; (2) we only observe a sample of individuals not the full population. We make several simplifying assumptions in order to address these limitations.

To address (1), we first assume that the times when an individual i visits a location l are not systematically selected to be times when $s(l, t)$ is unusually high or low. That is, letting \bar{s}_l denote the overall expectation of $s(l, t)$ over $t \in [0, 1]$, we have $E[s(l, t) | l(i, t) = l] = \bar{s}_l$ for all i . Provided this assumption holds, the expectation of the term $\int_{t=0}^1 s(l(i, t), t) dt$ is equal to $\bar{S}_i = \sum_l q_{il} \bar{s}_l$ where q_{il} is the expected share of i 's time that is spent in location l . We further assume that the times at which we observe pings are a random sample from $[0, 1]$ so we can estimate q_{il} and \bar{s}_l by the shares of i 's pings that occur in location l and the share of all pings in location l that come from W 's respectively.

Both of these are strong assumptions. The first would be violated, for example, if type W individuals tend to visit a particular park or restaurant in the morning while type NW individuals tend to visit it in the evening. The second would be violated if our data oversample periods in which the relative share of type W individuals is unusually high or low. In Appendix Section A2.3 we present robustness to an alternative specification allowing non-random weighting of pings across time.

To address (2), we re-weight home locations in our sample to match the distribution of population in the 2010 census. Because our data are relatively sparse at the geohash7 level, we re-weight by census tract. We define the weight for individual i to be

$$\lambda_i = \frac{N_{c(i)}}{\tilde{N}_{c(i)}} \quad (4)$$

where N_c is the census population of tract c and \tilde{N}_c is the number of devices in our sample with home locations in tract c .

Combining these assumptions, we form an estimator of S_i as follows. First, we form a leave-out estimate of \bar{s}_l :

$$\hat{s}_l^{-i} = \frac{\sum_{j \in \mathcal{P}_l^{-i} \cap W} \lambda_j}{\sum_{j \in \mathcal{P}_l^{-i}} \lambda_j}, \quad (5)$$

where \mathcal{P}_l^{-i} is the set of pings associated with individuals other than i who visit location l and we abuse notation by letting λ_j denote the weight of the individual associated with ping j . We omit visits by i from this measure to avoid a severe small-sample bias that can arise when some locations have a small number of observed visits (Cortese et al. 1976; Carrington and Troske 1997; Gentzkow et al. 2019). Second, we estimate \bar{S}_i by

$$\hat{S}_i = \frac{1}{|\mathcal{P}_i|} \sum_{j \in \mathcal{P}_i} \hat{s}_{l(j)}^{-i},$$

where \mathcal{P}_i is the set of pings associated with i and $l(j)$ is the location of ping j .

Finally, we estimate experienced isolation by

$$\hat{EI}_a = \frac{1}{|W_a|} \sum_{i \in W} \lambda_i \hat{S}_i - \frac{1}{|NW_a|} \sum_{i \in NW} \lambda_i \hat{S}_i.$$

We estimate residential isolation as

$$\hat{RI}_a = \frac{1}{|W_a|} \sum_{i \in W_a} \lambda_i \hat{r}_{c(i)} - \frac{1}{|NW_a|} \sum_{i \in NW_a} \lambda_i \hat{r}_{c(i)} \quad (6)$$

where \hat{r}_c is the share of devices in our sample with home census tract c that are WDs. This differs from the residential isolation measure typically reported in the literature because the types we consider are WDs and NWDs rather than white and black individuals and because we infer \hat{r}_c from our device data rather than census data.

3.3 Discussion

Our measure of experienced isolation considers an individual to be exposed to another if they are in the same location at the same time. This is what allows us to write equation 3 replacing the e_i of equation 1 with the average of $s(l, t)$ across space and time. The set of people that contribute to an individual’s exposure is, as discussed in the introduction, quite different from the set of people with whom an individual actually interacts. To the extent that we view actual interactions as the true object of interest, our measure can be seen as an approximation which significantly improves on residential measures but may still over- or understate isolation to the extent that interactions within different geohash7s are relatively more or less segregated.

In our empirical analysis, we define the types W and NW to be WDs and NWDs—devices from majority white and non-white home geohash7s—rather than white and non-white individuals. This is a departure from prior literature on residential segregation, where the assumption of uniform exposure within neighborhoods makes it possible to compute segregation based on individual race (using aggregate race shares measured in census data).

Therefore, the target of our estimation is subtly different than the standard target. To gain some intuition for the difference, note that individual geohash7s are perfectly segregated between WDs and NWDs by construction, whereas they are less than perfectly segregated by individual race. As noted in Section 2.4, the median WD lives in a home geohash7 which is 89 percent rather than 100 percent white, and the median NWD lives in a home geohash7 which is 78 percent rather than 100 percent non-white. We show below that this leads residential isolation between WDs and NWDs to be higher than between individual whites and non-whites. While the true level of segregation under our definition may be different, we expect the qualitative patterns we emphasize—e.g., the comparison of residential to experienced segregation—to be robust across alternative definitions.

As support for this, we report in Appendix Section A3 results using an alternative strategy where we impute race stochastically at the individual device level based on the composition of a home geohash7. This has the advantage of bringing our target concept closer to that in the prior literature. It has the disadvantage of introducing measurement error in the measure of a device’s type that could create a downward bias in experienced segregation estimates.¹⁴ While this alternative does change the level of segregation as expected, we confirm that our

¹⁴The random imputation strategy is equivalent to assuming that movement patterns are independent of individual race conditional on home geohash7. In simulations, we find that this tends to lead to a downward bias in estimates of experienced segregation.

main qualitative conclusions are indeed robust.

4 Main Results

Figure 1 shows estimated experienced and residential isolation for all MSAs in our sample.¹⁵ Two key facts are immediately apparent from these maps. First, experienced isolation is lower than residential isolation in large sections of the country. Second, the two measures are correlated across space, with both tending to be higher in the South, the Rust Belt, and in major cities, and tending to be lower in the upper Midwest and Northwest.

Figure 2 compares the two measures more directly, plotting experienced isolation against residential isolation. Experienced isolation is lower than residential isolation where residential isolation is high and higher than residential isolation where residential isolation is low. MSAs in the former category, however, account for the vast majority of the country's population, including all 15 of the most populous MSAs, with 87.9 percent of people living in MSAs where experienced isolation is less than residential isolation. The population-weighted average experienced isolation across all MSAs is 0.46, compared to average residential isolation of 0.61. The 10th and 90th percentiles of experienced isolation are 0.37 and 0.53, compared to 0.34 and 0.78 for residential isolation. This figure also confirms that experienced and residential isolation are highly correlated across MSAs, with a Pearson correlation coefficient of 0.864 and a Spearman rank correlation coefficient of 0.84. Among the 20 most populous MSAs, the ratio of experienced isolation to residential isolation is lowest (~ 0.6) in San Francisco-Oakland-Fremont, CA and Los Angeles, CA and highest (~ 0.8) in Atlanta, GA, and Riverside, CA.

To describe the factors that correlate with lower experienced segregation, we regress experienced isolation on observed MSA characteristics controlling for fifteen equal-sized bins of residential isolation. We focus on population-weighted univariate relationships, including a single observed characteristic in each case.¹⁶ We emphasize that these are purely descriptive correlations and need not imply anything about the causes or effects of segregation.

Figure 3 shows the results. Each panel plots residuals of experienced isolation against residuals of a given MSA characteristic after partialing out the residential isolation controls. Experienced isolation is relatively lower in MSAs with higher population density and more public

¹⁵ Appendix Figure A8 presents a map with the difference between experienced and residential isolation for each MSA. Appendix Tables A5-A7 report both experienced and residential isolation for each MSA.

¹⁶ Appendix Table A8 shows similar results in regressions that are unweighted but subset to the top 50, 100, and 200 most populous MSAs.

transit use. This is consistent with the fact that in dense areas residents from different neighborhoods are less separated by physical space, and may reflect the role of urban amenities such as parks and public facilities in facilitating diverse interactions (Jacobs 1961). Experiences are also relatively less isolated in MSAs with higher income, more education, and lower unemployment. This could reflect a number of forces including the role of social capital in reducing segregation (Putnam 2000). Experienced isolation is relatively lower where populations are younger, possibly reflecting the importance of schools and workplaces in reducing segregation. Finally, relative experienced isolation is negatively correlated with Chetty et al.'s (2014) measures of income mobility for both blacks and whites, consistent with both diverse interactions increasing mobility and with areas that facilitate opportunity also promoting diverse interactions.

5 Decomposing Experienced Isolation

5.1 By Time

We first ask how experienced isolation varies over hours of the day. To do this, we restrict both exposures and the set of devices to all those that occur in a specific hour according to the MSA's local time zone. Exposures are only estimated in geohash7s that are visited by devices that ping within that hour. For example, experienced isolation for 10 a.m. restricts our sample to pings that occur between 10 a.m. and 11 a.m. local time. After restricting the set of pings and devices, the estimation of experienced isolation is identical to our baseline measure.

Figure 4 plots experienced isolation over the course of the day, scaled relative to the level of residential isolation. The figure highlights the 10 most populous MSAs. The results are intuitive: Experienced isolation is lowest in the middle of the day as people move around and highest late at night as people withdraw into their homes. The ratio mostly differs in level between MSAs and almost all MSAs share the same time profile.

5.2 By Location

We next decompose experienced isolation by location. Much like restricting to pings within an hour, we restrict to pings that occur within a set of geohash7s of a particular type.¹⁷ These results are shown in Figure 5. The leftmost point in the plot shows the average of our baseline

¹⁷If an individual never visits a geohash7 of the type in question they are dropped from the sample.

measure of experienced isolation across MSAs, which includes all locations in our sample. The error bars in the plot indicate ± 1 standard deviation of the measure across MSAs.

The next two points in the figure show experienced isolation for locations within vs. outside of home census tracts. The results show that experienced isolation within home tracts (0.63 on average across MSAs) is higher than overall experienced isolation (0.46 on average), and actually higher than residential isolation (0.61 on average).¹⁸ As discussed above, this result is not mechanical: experienced isolation within the home tract could differ from residential isolation in either direction, both because within-tract exposure is not uniform and because it includes visitors who live outside the home tract. In contrast, experienced isolation outside of home tracts is much lower, with an average of 0.21 across MSAs. Thus, time spent away from home is the key force reducing segregation relative to what the standard residential measure would suggest.

Figure 5 summarizes the differences in experienced isolation for specific categories of features.¹⁹ The baseline category contains all features, as well as time spent at home. Average experienced isolation in outdoor spaces like parks, gardens, sports fields and playgrounds is only 50.3 percent of mean baseline isolation, and commercial establishments like restaurants and bars and retail stores have experienced isolation that is only 43.5 and 47.8 percent of baseline isolation respectively. Isolation is among its lowest in places of entertainment like theaters (24.3 percent of baseline) and accommodations like hotels (24.6 percent of baseline). Appendix Table A9 shows summary statistics for experienced isolation across a wider set of feature types.

5.3 By Race

Finally, we can decompose the differences in exposure that underlie the isolation index between WDs and NWDs. Experienced isolation is the difference between these groups in average exposure $E[s(l, t)]$. We ask how the experienced exposure relative to residential exposure differs by group. The results, which we present in Appendix Figures A10 and A11, show that the difference between experienced and residential exposure is relatively small for WDs and much larger for NWDs. It also shows that NWDs' experienced exposure varies much more across MSAs and across different feature types. This suggests that factors which reduce segregation away from home may have a particularly large impact on the experiences of non-whites.

¹⁸ Appendix Figure A9 depicts experienced isolation within and outside home tracts.

¹⁹ Appendix Figure A12 depicts ping activity across features by WD/NWD designation.

6 Robustness

The Online Appendix reports a number of additional specifications probing the robustness of our main result. We provide detail on these specifications in Online Appendix A2 and show the results in Online Appendix Table A10. They show that our main qualitative conclusions are robust to: (i) excluding pings that are likely to occur while devices are commuting or traveling; (ii) using alternative sources of demographic data; (iii) excluding devices with home locations outside the MSA; (iv) dropping the top 5 percent of devices in terms of number of pings per day; (v) excluding pings occurring between midnight and 6 a.m.; (vi) using only the first ping emitted by a device in a given hour (so as to avoid over-weighting hours with frequent pings). The final result in this table shows that we would over-estimate experienced segregation if we used a naive estimator rather than the leave-out correction in equation (5).

7 Conclusion

The extent to which members of different groups are able to see, meet, and interact with one another can profoundly shape economic and social outcomes. Standard isolation indices capture such patterns under the assumption that people are uniformly exposed to others in their neighborhoods of residence. Our measure of experienced isolation relaxes this assumption, making it possible to leverage novel location data to describe the exposures people actually experience as they move around over the course of their days.

We find that the isolation people actually experience is substantially lower than residential measures would suggest. People spend substantial time away from their home neighborhoods, and when they do they are much more likely to encounter diverse others than they would at home. Commercial places like restaurants and retail shops are a particularly strong force pulling against segregation, while local amenities such as churches and schools tend to remain more segregated. One implication is that public goods that are tied to residential boundaries should be a particular focus of efforts to combat segregation. They also suggest that the negative effects of segregation are likely higher for those like children and the elderly whose exposure is more tied to their local neighborhoods.

While experienced and residential segregation are highly correlated across cities, the gap between them varies systematically, with relatively less experienced isolation in cities that are denser, wealthier, and more educated, that have greater use of public transport, and where in-

come mobility is higher. These correlations do not allow us to draw any direct conclusions about either the causes or consequences of segregation, but they point toward factors that will be especially fruitful for subsequent research to investigate.

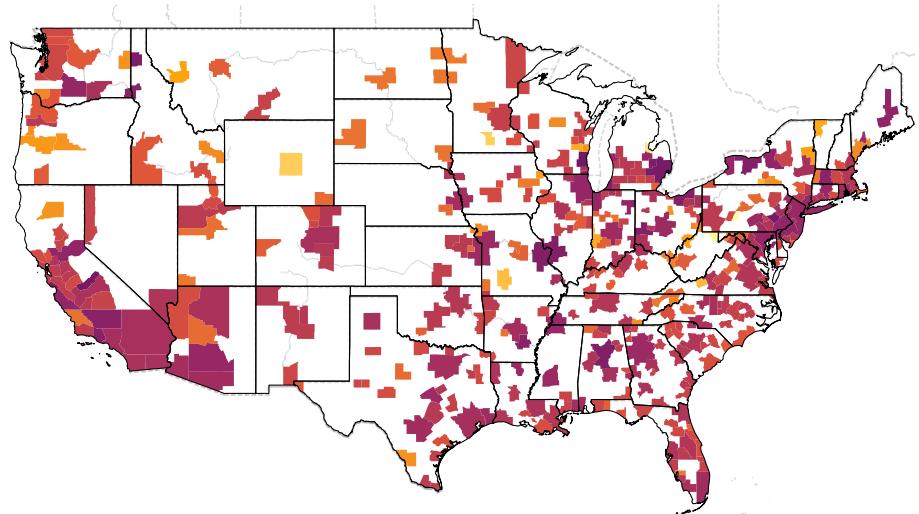
References

- Blattman, C., D. P. Green, D. Ortega and S. Tobon. 2019. "Place Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime." Working Paper.
- Caetano, G. and V. Maheshri. 2019. "Gender Segregation within Neighborhoods." *Regional Science and Urban Economics* 77:253-263.
- Card, D., A. Mas and J. Rothstein. 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics* 123(1):177-218.
- Carrington, William J., Kenneth R. Troske. 1997. "On Measuring Segregation in Samples with Small Units." *Journal of Business & Economic Statistics* 15(4):402-409.
- Chen, K.M., K. Haggag, D. Pope and R. Rohla. 2019. "Racial Disparities in Voting Wait Times: Evidence from Smartphone Data." NBER Working Paper No. 26487.
- Chen, K. M. and R. Rohla. 2018. "The Effect of Partisanship and Political Advertising on Close Family Ties." *Science* 360(6392):1020-1024.
- Chetty, R., N. Hendren, P. Kline, E. Saez. 2014. "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States." *The Quarterly Journal of Economics* 129(4):1553-1623.
- Chetty, R. and N. Hendren, 2018a. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects." *The Quarterly Journal of Economics* 133(3):1107-1162.
- Chetty, R. and N. Hendren, 2018b. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *The Quarterly Journal of Economics* 133(3):1163-1228.
- Chetty, R., N. Hendren and L. F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *The American Economic Review* 106 (4):855–902.
- Chetty, R., N. Hendren, M. R. Jones and S. R. Porter. 2020. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *The Quarterly Journal of Economics* 135(2):711-783.
- Cortese, C. F., R. F. Falk and J. K. Cohen. 1976. "Further Considerations on the Methodological Analysis of Segregation Indices." *American Sociological Review* 41(4):630–37.
- Cowgill, D. O. and M. S. Cowgill. 1951. "An Index of Segregation Based on Block Statistics." *American Sociological Review* 16:825-31.
- Cutler, D. M. and E. L. Glaeser. 1997. "Are Ghettos Good or Bad?" *The Quarterly Journal of Economics* 112(3):827–72.
- Cutler, D. M., E. L. Glaeser and J. L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *The Journal of Political Economy* 107(3):455–506.
- Davis, D. R., J. I. Dingel, J. Monras and E. Morales. 2019. "How Segregated Is Urban Con-

- sumption?” *Journal of Political Economy* 127(4).
- Diamond, R., T. McQuade and F. Qian. 2019. “The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco.” *American Economic Review* 109(9):3365-94.
- Duncan, O. D. and B. Duncan. 1955. “A Methodological Analysis of Segregation Indexes.” *American Sociological Review* 20(2):210–17.
- Echenique, F. and R. G. Fryer. 2007. “A Measure of Segregation Based on Social Interactions.” *The Quarterly Journal of Economics* 122(2):441–85.
- Gentzkow, M. and J. M. Shapiro. 2011.“Ideological Segregation Online and Offline.” *The Quarterly Journal of Economics* 126(4):1799–1839.
- Gentzkow, M., J. M. Shapiro, and M. Taddy. 2019.“Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech.” *Econometrica* 87(4):1307-1340.
- Glaeser, E. L., S.D. Kominers, M. Luca and N. Naik. 2018. “Big Data And Big Cities: The Promises and Limitations of Improved Measures of Urban Life.” *Economic Inquiry* 56:114-137.
- Humes, K. R., N. A. Jones and R. R. Ramirez. 2011.“Overview of Race and Hispanic Origin: 2010.” 2010 Census Briefs.
- Iceland, J., D. H. Weinberg and E. Steinmetz. 2002.“Racial and Ethnic Residential Segregation in the United States: 1980-2000.” *Census 2000 Special Reports*.
- Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York : Vintage Books.
- Jahn, J. A. 1950. “The Measurement of Ecological Segregation: Derivation of an Index Based on the Criterion of Reproducibility.” *American Sociological Review* 15:101-104.
- Massey, D. S. and N. A. Denton. 1988. “The Dimensions of Residential Segregation.” *Social Forces. A Scientific Medium of Social Study and Interpretation* 67(2):281–315.
- Massey, D. S. and N. A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.
- Park, Y.M. and M. Kwan. 2018. “Beyond Residential Segregation: A Spatiotemporal Approach to Examining Multi-Contextual Segregation.” *Computers, Environment and Urban Systems* 71:C.
- Phillips, N.E., B.L. Levy, R.J. Sampson, M.L. Small and R.Q. Wang. 2019. “The Social Integration of American Cities: Network Measures of Connectedness Based on Everyday Mobility across Neighborhoods.” *Social Methods & Research*.
- Putnam, R.D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
- Sunstein, C. R. 2002. *Republic.com*. Princeton University Press.

- US Census Bureau. 2017. “TIGER/Line Shapefiles, Technical Documentation.” <https://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2017>.
- Taeuber, K. E. and A.F. Taeuber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change*. Chicago. Aldine Pub. Co.
- White, M. J. 1983. “The Measurement of Spatial Segregation.” *American Journal of Sociology* 88(5):1008-1018.
- White, M. J. 1986. “Segregation and Diversity Measures in Population Distribution.” *Population Index* 52 (2):198–221.

Experienced isolation



Residential isolation

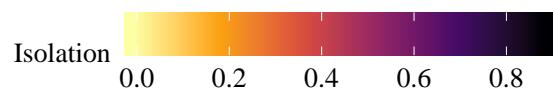
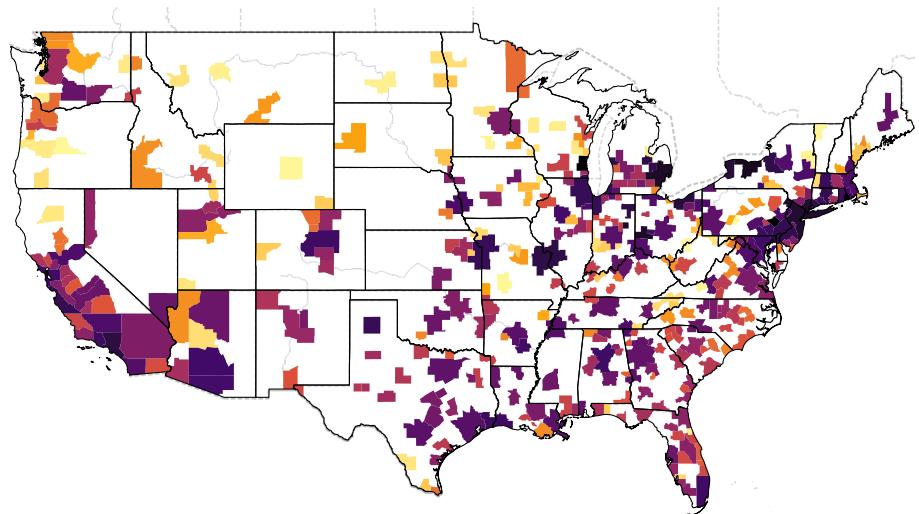


Figure 1: Experienced and Residential Isolation by MSA

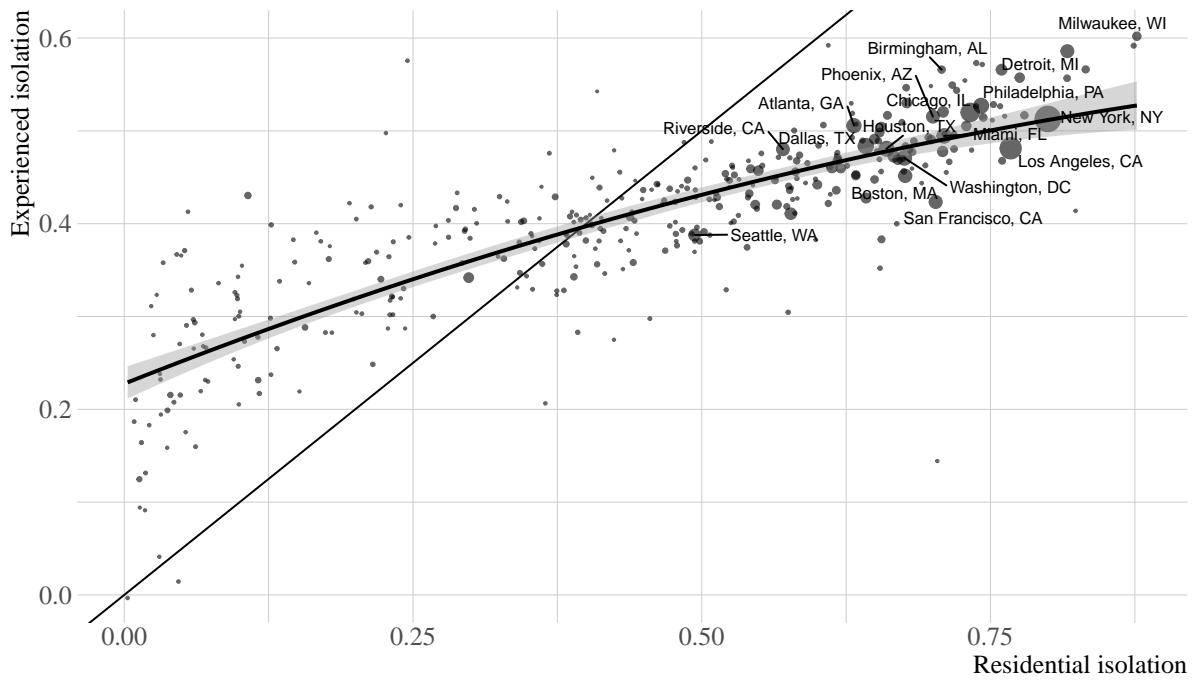


Figure 2: Experienced vs. Residential Isolation

Notes: Plot shows experienced and residential isolation for each MSA. The size of each point is proportional to the MSA's population. The labeled points designate the 15 most populous MSAs. We plot the 45 degree line and a local polynomial fit.

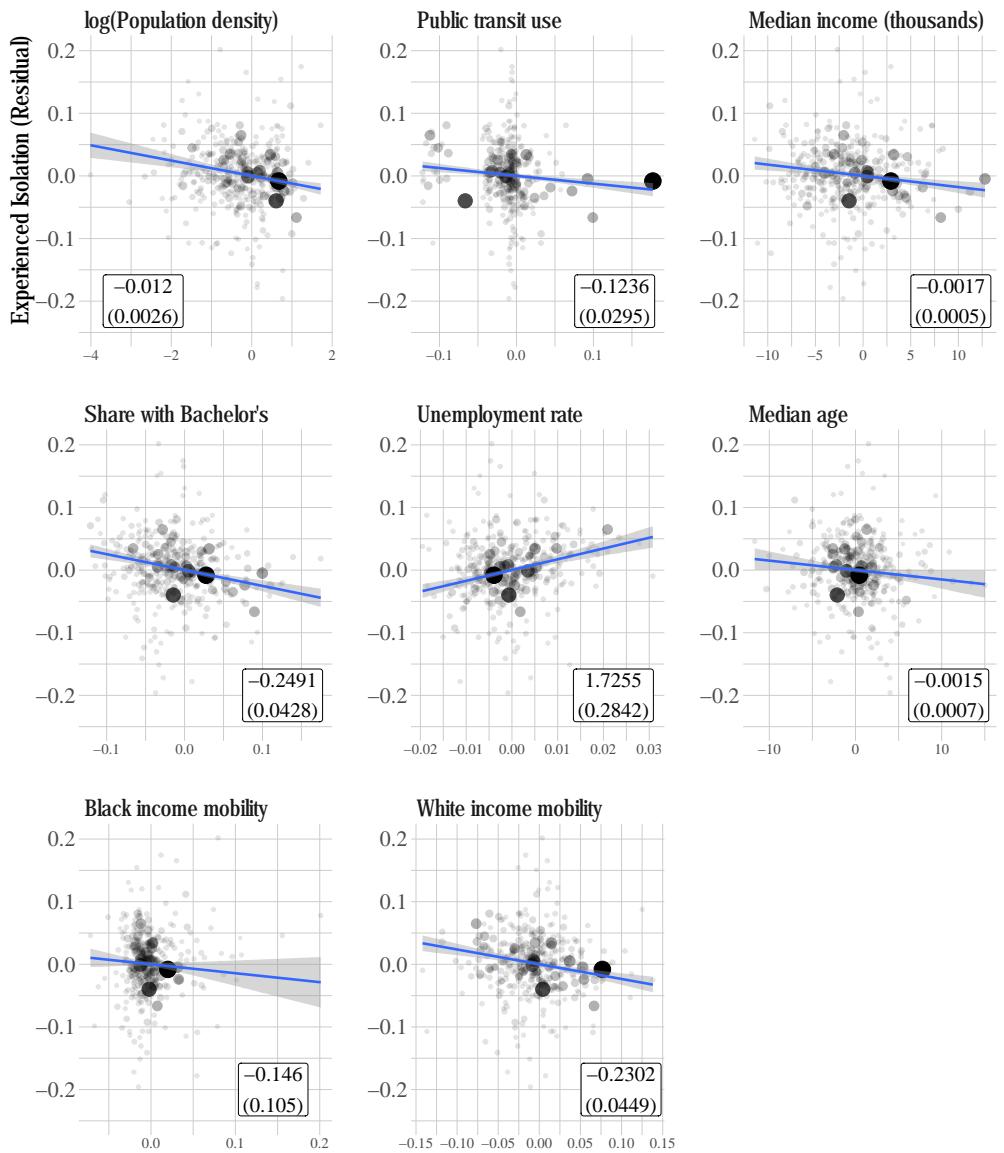


Figure 3: Residual Experienced Isolation and MSA Characteristics

Notes: On the y-axis, we plot the residual from a population weighted regression of experienced isolation on fifteen equal sized bins of residential isolation at the MSA level. The x-axis in each plot refers to the specified MSA characteristic. Each point refers to an MSA and is shaded and sized relative to total population. In the white box in the lower left corner, we show the coefficient and standard error from the population weighted regression of experienced isolation on the residential isolation bin fixed effects and the specified covariate. The blue line shows the population weighted linear fit. The share with bachelor's variable includes the percent of people in an MSA that have at least a bachelor's degree. The black and white income measures average Chetty et al.'s (2020) county estimates (pooled by race) of the share of individuals born in the 25th percentile of the income distribution who make it to the top quintile. Public transit use is the share of the working population that uses public transport to get to work.

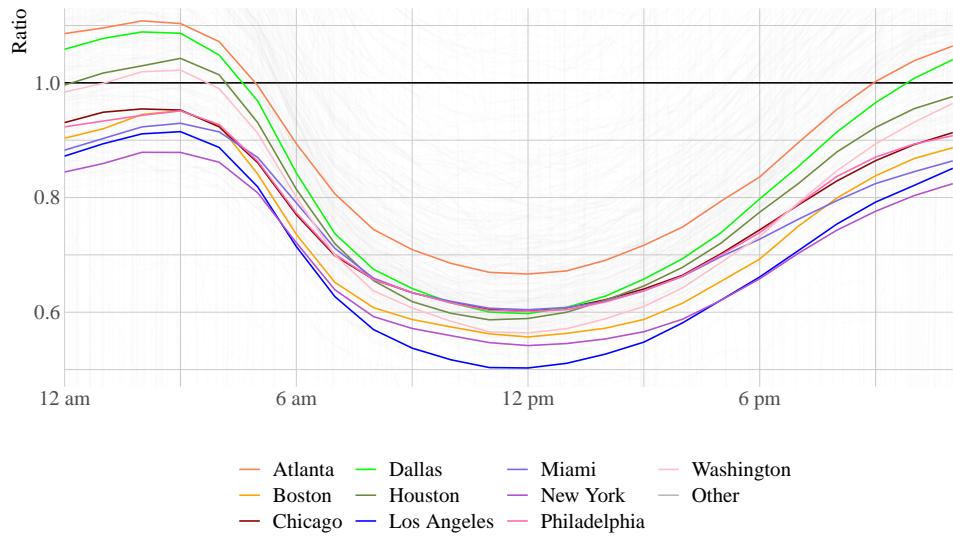


Figure 4: Experienced Isolation Relative to Baseline by Time of Day

Notes: We plot the ratio of experienced to residential isolation in each hour of the day, highlighting the 10 most populous MSAs. Note that isolation can only be calculated for the devices active in a given hour, so the sample does change for each hour specification.

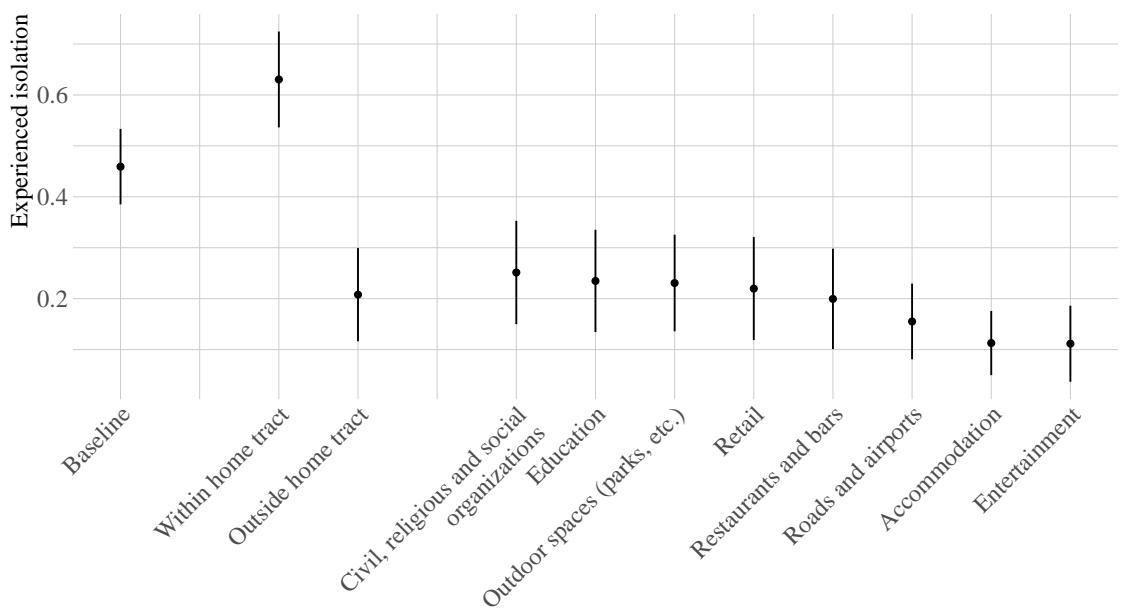


Figure 5: Experienced Isolation Relative to Baseline by Location

Notes: We plot the population weighted mean experienced isolation in a particular feature and compare with our baseline measure. Error bars show \pm the population weighted standard deviation of experienced isolation across MSAs.

Online Appendix: Experienced Segregation

Susan Athey, *Stanford University and NBER*

Billy Ferguson, *Stanford University*

Matthew Gentzkow, *Stanford University and NBER*

Tobias Schmidt

A1 Defining Geographic Features

A1.1 InfoUSA

We use data from InfoUSA to define the following features: (i) civil, religious, and social organizations; (ii) retail; (iii) restaurants and bars; (iv) accommodation; (v) entertainment; (vi) sports and recreation. We combine InfoUSA data with Open Street Maps data to define education features, as detailed in Appendix A1.2 below.

For each feature, InfoUSA provides latitude, longitude, and an 8-digit NAICS industry code. We focus on the top 334 NAICS codes in the data which together cover 95 percent of all establishments and assign them manually to our aggregated categories. The mapping between NAICS8 and categories is given in Table A1. We assign each feature to the geohash7 that contains its latitude-longitude pair.

A1.2 OpenStreetMaps

We use data from Open Street Maps (OSM) to define outdoor spaces and to supplement InfoUSA data in defining education features. OSM is an open source mapping project that defines geographic features and associates them with metadata tags. Whereas InfoUSA provides point locations for each feature, OSM defines two-dimensional polygons. These provide a more accurate representation for features like parks that occupy a large amount of space. As noted in the main text, we associate each OSM feature with all geohash7s that intersect its polygon.

We define features to be outdoor spaces if they are associated with the tags “leisure=park,” “leisure=playground,” “leisure=pitch,” or “leisure=garden.” We define features to be education-related if they are associated with the tags “amenity=school,” “amenity=kindergarten,”

“amenity=university,” or “amenity=college.” We define geohash7s to contain education features if they are associated with such a feature in either InfoUSA or OSM.

A1.3 Transportation Infrastructure

We define transportation features by combining polygon data on primary and secondary roads from the US Census’ TIGER database (U.S. Census Bureau 2017) with airport polygon data from OSM (polygons with the tag aeroway=aerodrome). We associate each such feature with all geohash7s that intersect its polygon.

A2 Robustness

Table A10 probes the robustness of our experienced isolation estimates. The first row repeats our baseline estimate. Each subsequent row reports a separate robustness check. For each, we report the mean, median, 5th, and 95th percentiles of estimated experienced isolation by MSA, as well as the correlation between that row’s estimates and the baseline.

A2.1 Excluding Transportation Infrastructure

The first two robustness checks show how the results change if we exclude pings that are likely to come while people are in transit. People sharing the same space while commuting or traveling (e.g., driving next to another car on a highway) may be relatively unlikely to have meaningful interactions, and so it is interesting to know whether these observations play a large role in our conclusions. Figure A13 shows the frequency of pings across geohash7s in Birmingham, AL, confirming visually that a substantial amount of activity indeed occurs on roads.²⁰

In row (2) of Table A10, we exclude all pings in geohashes we identify as containing roads or airports. In row (3) we exclude all pings that are part of a sequence suggesting the device is moving at more than 12 miles per hour.²¹ In both cases experienced isolation rises, consistent with time in transit having lower than average isolation, but the difference is modest and the correlation with the baseline estimates is high.

²⁰Online Appendix Table A11 reports summary statistics for pings that maybe in transit.

²¹We take the sequence of timestamped latitudes and longitudes, compute the Haversine distance between successive pings in the sequence, and divide by the time difference to estimate device speed.

A2.2 Imputing Geohash7 Demographics from Individual Data

Our baseline estimates impute geohash7 demographics from 2010 census data as described in Section 2.3. A weakness of this approach is that census blocks (the smallest units at which demographics are reported in the census) do not align exactly with geohash7s. An alternative approach is to impute geohash7 demographics from individual-level data sources where we observe individuals' exact addresses.

Row (4) of Table A10 reports estimates based on demographics imputed from an individual-level voter file provided by the company L2. These data provide home addresses of registered voters. The data include self-reported race in eight states that collect this as part of the registration process (Alabama, Florida, Georgia, Louisiana, North Carolina, South Carolina, Tennessee and Texas), and imputed race based on a proprietary algorithm elsewhere. We use the L2 category "European" as our measure of "white." We successfully match at least one L2 individual in 5,767,098 of our 7,288,958 geohash7s. For the remaining geohash7s, we use our baseline imputation.

Row (5) reports estimates based on demographics imputed from individual-level data from the company Infutor. These data provide names, address, immigration status and demographics of around 80% of the adult population in the US. Race is defined in accordance with census categories and is imputed as described in Diamond et al. (2019). We exclude all individuals over the age of 80 and determine the home-geohash7 of each individual based on the last address they were registered at. We successfully match at least one Infutor individual in 5,079,532 of our 7,288,958 geohash7s. For the remaining geohash7s, we use our baseline imputation.

The results show that these alternative approaches lead to even lower estimates of experienced isolation, strengthening our main conclusion. The correlation with our baseline estimates remains high in both cases.

A2.3 Alternative Temporal Weighting

As discussed in Section 3.2, our baseline estimates rely on an assumption of uniform sampling. One way this may be violated is if some devices emit large numbers of pings at specific times when relevant apps are used heavily. In row (6) of Table A10, we partially address this possibility by only using the first ping emitted by a device in a given hour in a particular geohash7. This effectively gives equal weight to each device-geohash7-hour tuple in the data. We find experienced isolation is even lower in this specification, suggesting that non-random weighting

may if anything lead us to underestimate the gap between residential and experienced isolation.

A2.4 Other Robustness Checks

The final rows of Table A10 consider other variations in the baseline sample. Row (7) drops pings from the sample devices whose home location is not in the same MSA as the ping, to give a sense of how out-of-town visitors influence the estimates. Row (8) drops the top 5 percent of devices in terms of number of pings per day, to address the possibility that such heavy users might have undue influence on the results. Row (9) excludes pings during late night hours from midnight to 6 a.m., to assess how much our results are influenced by the way we treat sleep time. None of these make a large difference to the estimates.

Finally, row (10) shows how the results are impacted by the leave-out correction in equation (5). If we instead use the naive estimator that includes a device's own observations in the estimation of \bar{s}_l , we would over-estimate segregation. This is consistent with prior literature showing that this small-sample bias leads segregation to be overstated.

A3 Alternative Definitions of Types

Our baseline results measure segregation between WDs (devices from majority-white home geohash7s) and NWDs (devices from majority-non-white home geohash7s). In Table A12, we present results for racial segregation under alternative definitions of types.

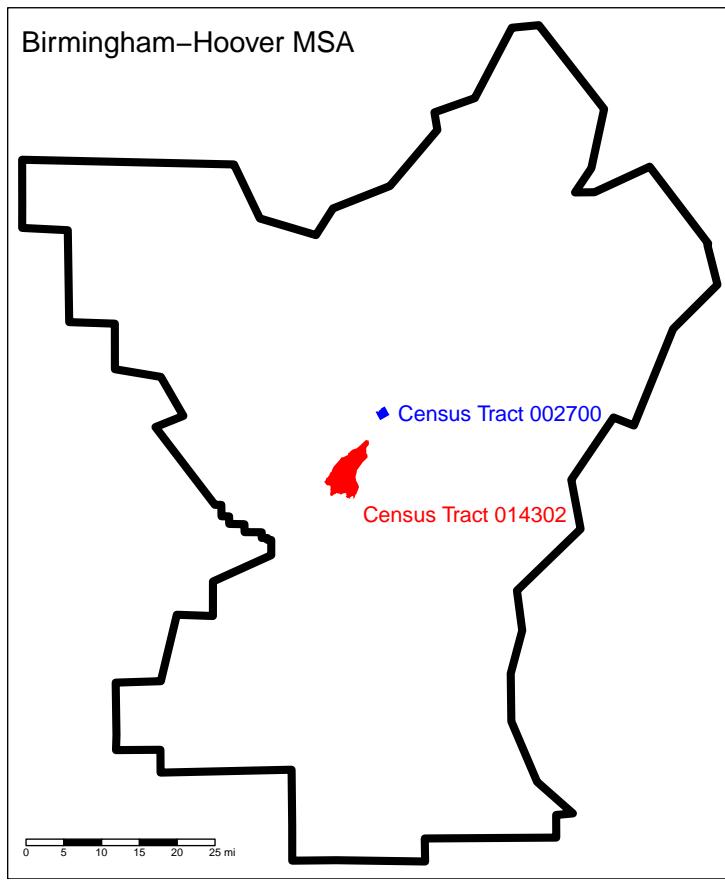
In row (2), we change the definition of WD and NWD to depend on whether a home geohash7 is above or below the overall US share white (63 percent). Results are similar to our baseline estimates.

In the following rows, we shift focus to segregation between white and black home locations rather than white and non-white home locations. Row (3) defines the two types to be devices from home geohash7s with at least 50 percent white and at least 50 percent black respectively. Row (4) uses cutoffs of at least 70 percent white and at least 70 percent black respectively. Row (5) uses cutoffs of at least 90 percent white and at least 90 percent black respectively. Note that unlike in our baseline specification, some devices here fall into neither category and so are omitted from the analysis. Results are again similar to baseline, with somewhat lower estimates of both experienced and residential isolation when we use the more extreme cutoffs.

In rows (6)-(7), we present an approach to measuring experienced isolation by individual race. The differences between this and our baseline approach based on home geography are discussed in Section 3.3. We probabilistically impute individual race at the device level using the race shares in their home geohash7s (as measured by our baseline imputed demographics) as probabilities.²² This probabilistic imputation would correctly estimate segregation by individual race under the assumption that expected movement patterns are the same for devices of different races that come from the same geohash7. When this assumption is violated, we expect the systematic measurement error in imputed race to bias our experienced isolation measure downward.

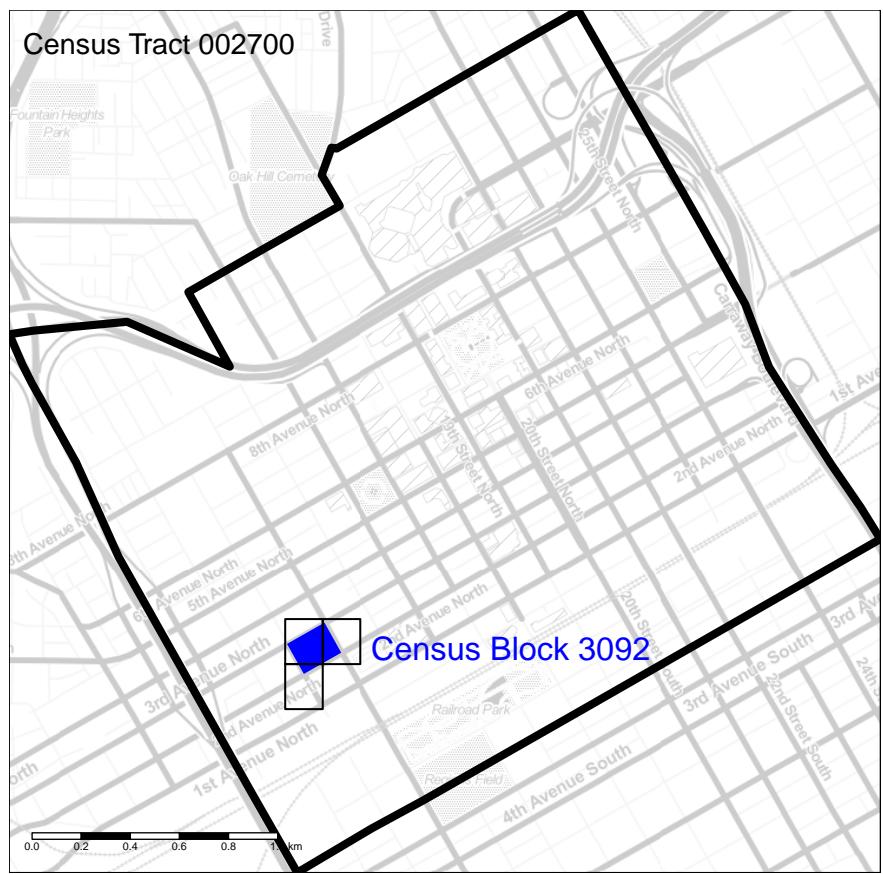
Row (6) shows results where we probabilistically impute white and non-white race. Row (7) shows results where we probabilistically impute white and black race (allowing for an omitted category of devices that are neither black nor white). Consistent with expectations, estimates of both experienced and residential isolation fall significantly in these specifications. However, we find that experienced isolation is substantially lower than residential isolation and that the two are highly correlated across MSAs continue to hold.

²²The direct imputation of race assigns to each device the home geohash7 share white and share nonwhite (or black) as constructed in Section 2.3. When we observe a device’s ping in a geohash7, that device contributes the imputed share white to the exposure of that geohash7 instead of contributing a wholly white or nonwhite visit. Furthermore, when we construct the average exposure amongst white and nonwhite (or black) devices, we cannot simply take the average over devices of each type since each device is probabilistically assigned to both groups. Therefore, the average exposure of each group is calculated as a weighted average with the imputed home geohash7 demographic shares as the weights.



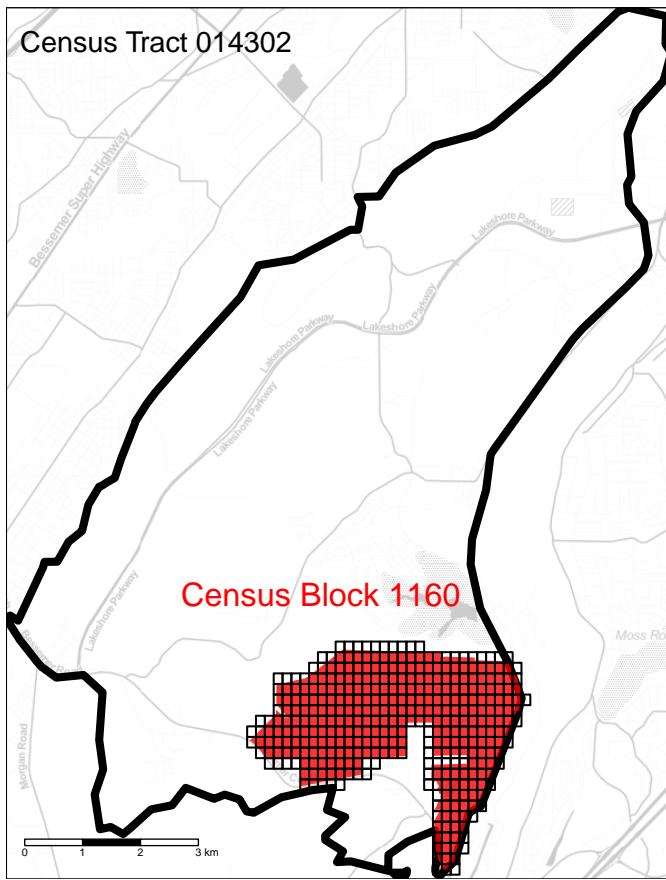
Online Appendix Figure A1: Birmingham-Hoover MSA and Census Tracts 002700 & 014302

Notes: We depict the relative size of urban and rural tracts within the Birmingham-Hoover MSA, which is the final aggregate unit of analysis.



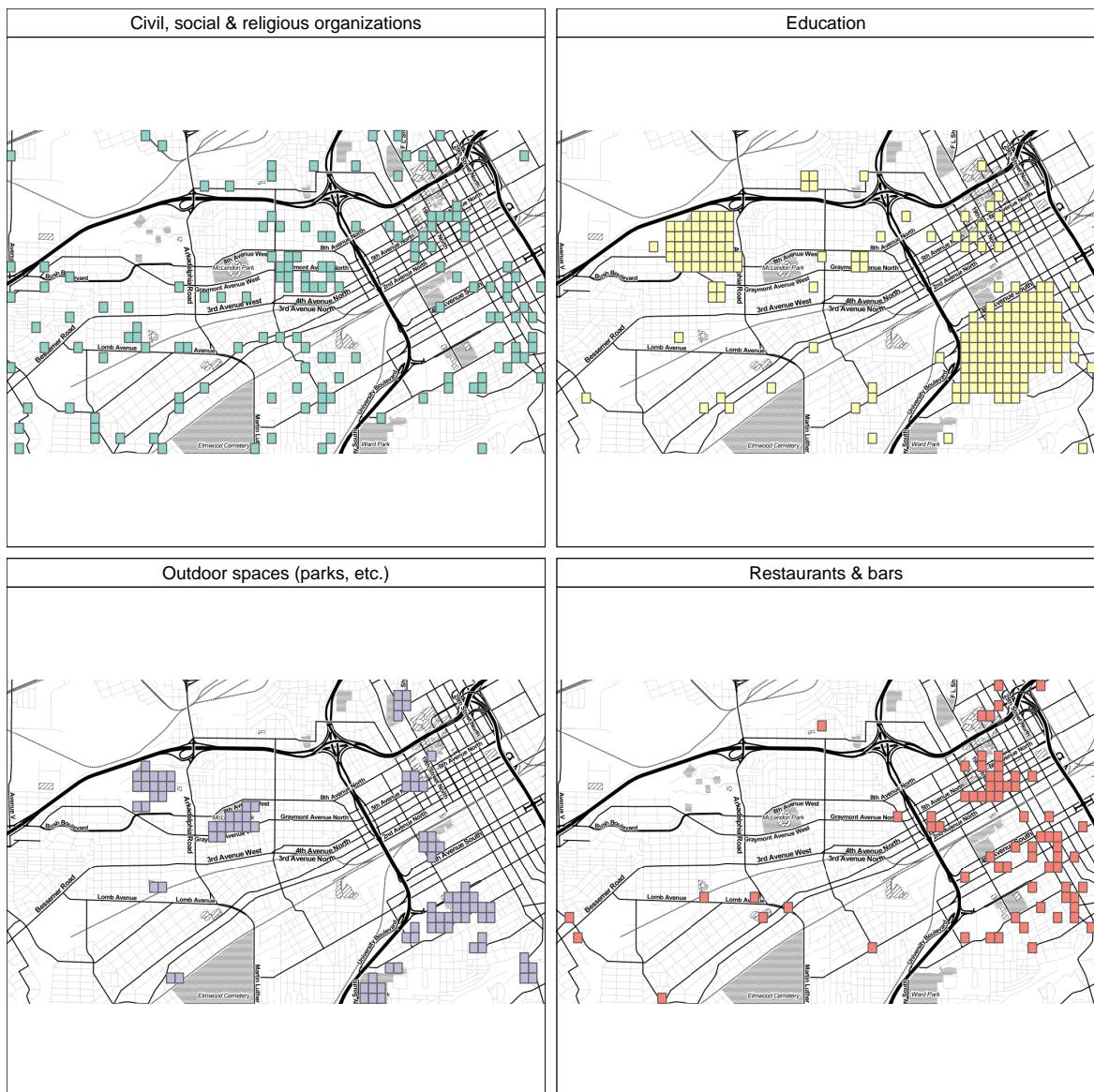
Online Appendix Figure A2: Relative Geographies in an Urban Center

Notes: This figure illustrates the relative size of census tracts / blocks and geohash7s in an urban area. The larger black outline depicts census tract 002700 in urban Birmingham, AL. The black grid consists of three geohash7s that overlap census block 3092.



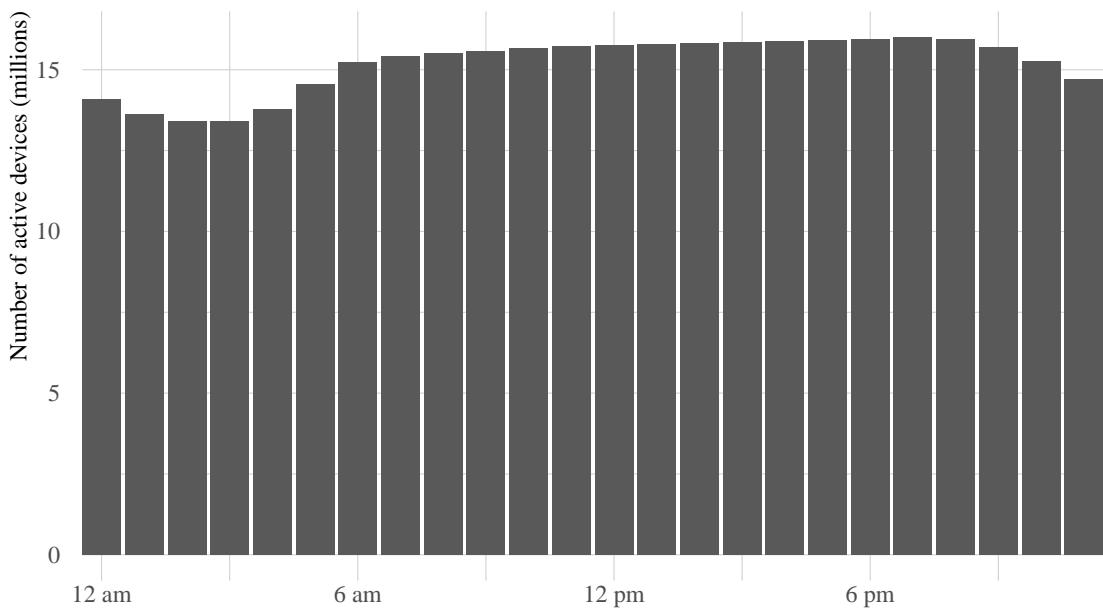
Online Appendix Figure A3: Relative Geographies in a Rural Setting

Notes: This figure illustrates the relative size of census tracts, blocks, and geohash7s in a rural area. The larger black outline depicts census tract 014302 in rural Birmingham, AL. The black grid consists of 367 geohash7s that overlap the census block 1160.



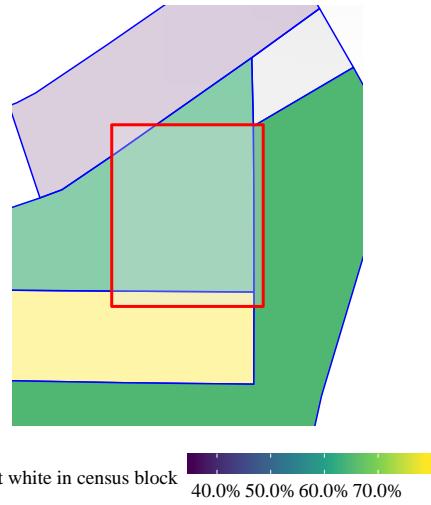
Online Appendix Figure A4: Features in Downtown Birmingham, AL

Notes: Figure highlights geohash7s that contain (i) civil, social, and religious institutions, (ii) educational institutions, (iii) outdoor spaces, and (iv) restaurants and bars respectively, in downtown Birmingham, AL. A single geohash7 can contain multiple features.



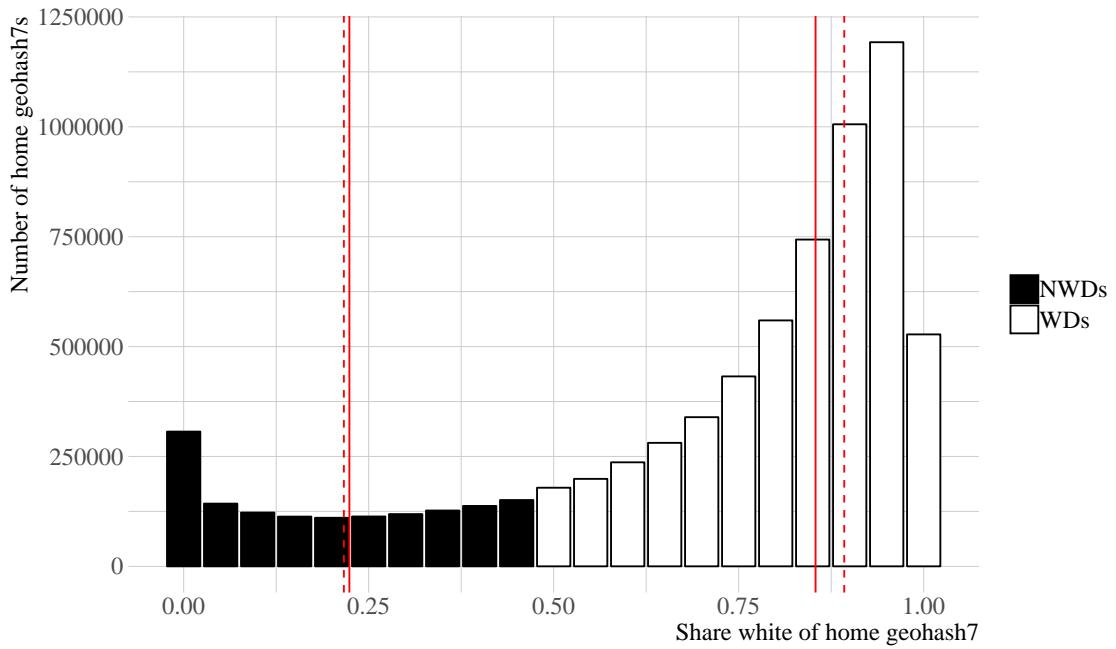
Online Appendix Figure A5: Number of Active Devices by Hour

Notes: We plot the number of active devices in millions by hour. A device is considered active if we ever observe at least one ping within the given hour.



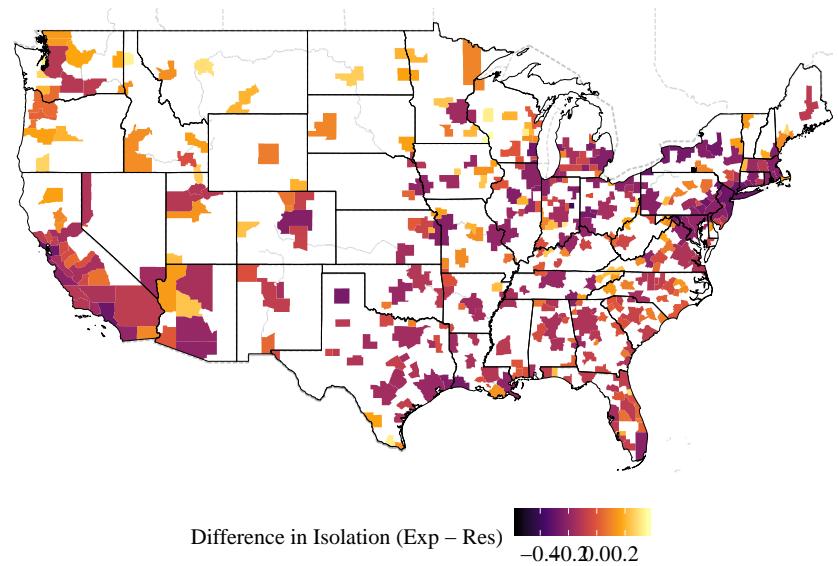
Online Appendix Figure A6: Matching Home Geohash7 to Blocks

Notes: Geohash7 djfq8cs in Jefferson county, AL is the rectangle outlined in red. Census blocks are the polygons outlined in blue. There are five census blocks overlapping the geohash7 which we color in relation to their share white. The grey census block is uninhabited.



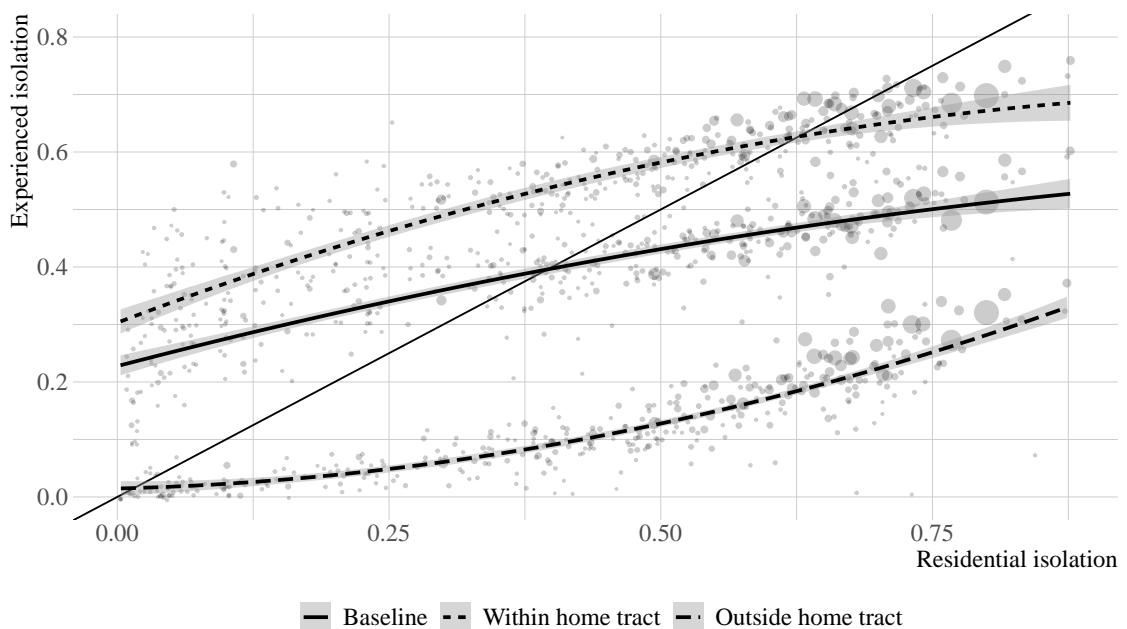
Online Appendix Figure A7: Home Geohash7 Percent White Histogram by Majority Race

Notes: Figure plots the number of WD and NWD home geohash7s by the share white of the home geohash7. The mean and median by majority race of the home geohash7 are represented by solid and dashed red lines respectively. The mean and median share white of NWD home geohash7s are both 0.22 and of WD home geohash7s are 0.85 and 0.89 respectively.



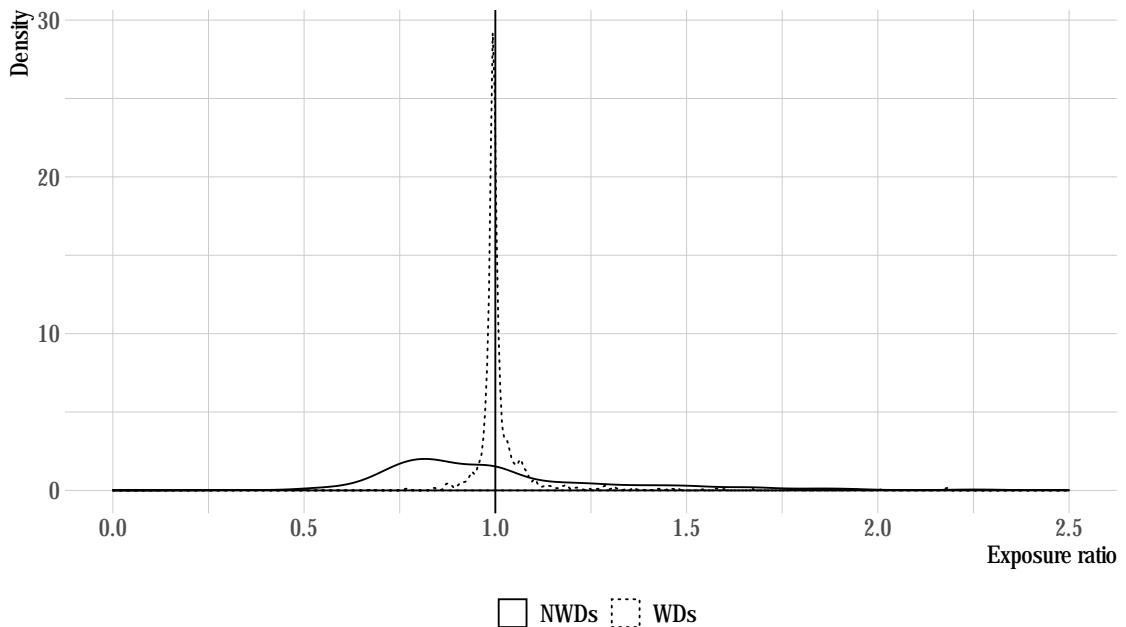
Online Appendix Figure A8: Difference Between Experienced and Residential isolation by MSA

Notes: We color each MSA relative to the difference of experienced minus residential isolation.



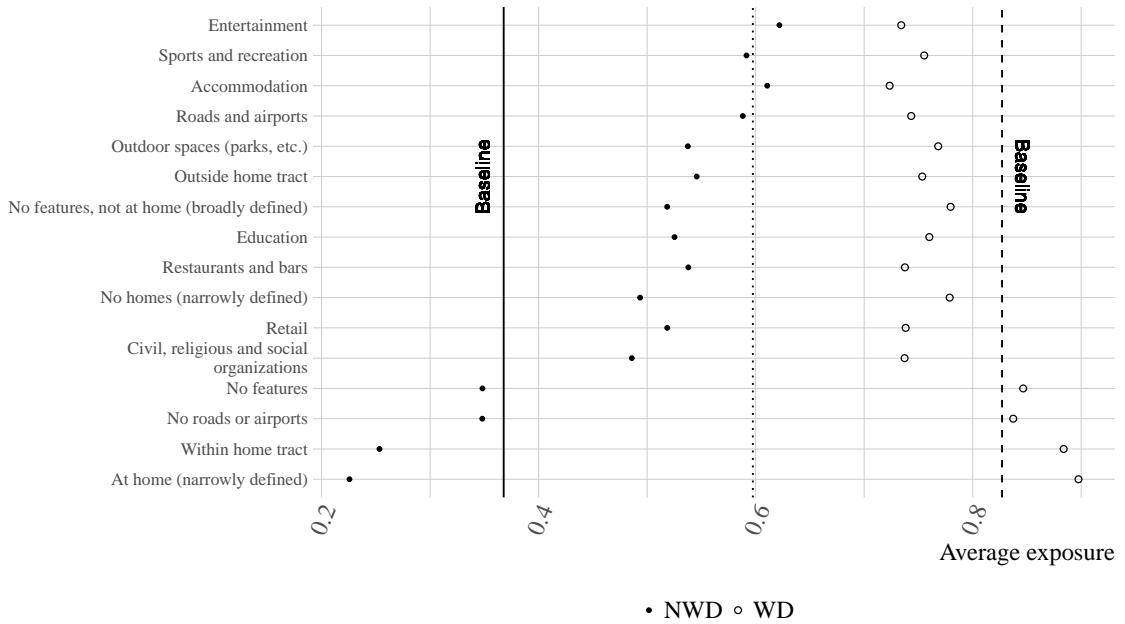
Online Appendix Figure A9: Experienced vs Residential Isolation Relative to Devices' Homes

Notes: We plot three specifications of experienced isolation against residential isolation with each point representing an MSA. The within and outside home tract specifications only include exposures in geo-hash7s within or outside individuals' home census tract. The size of each point is proportional to the MSA's population. We plot the 45 degree line and fit local polynomials to the data.



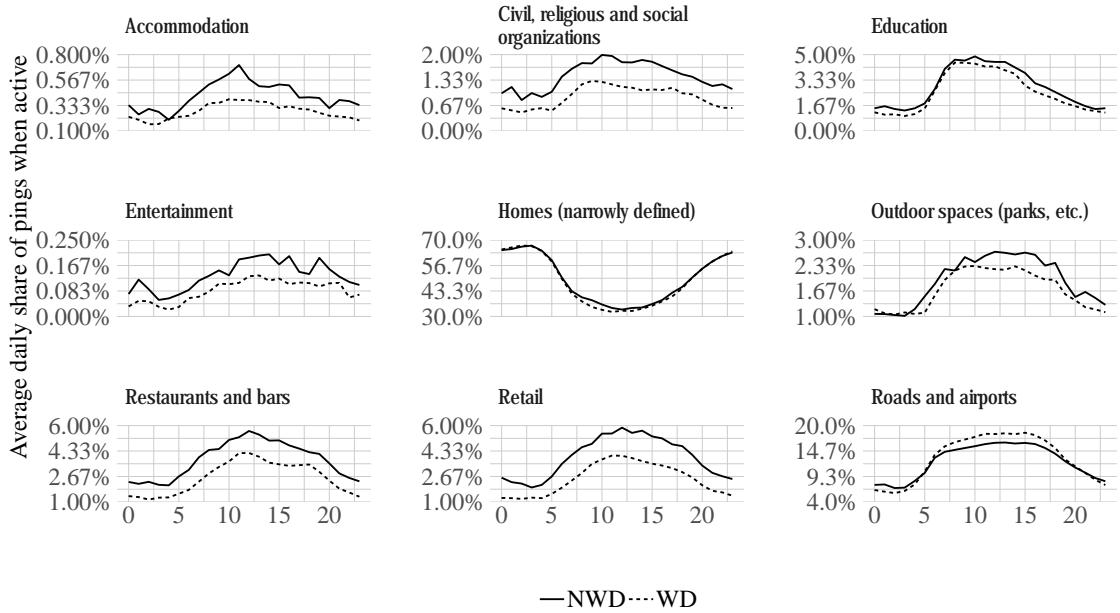
Online Appendix Figure A10: Experienced / Residential Exposure for WDs and NWDs

Notes: We plot the distribution of exposure ratios for all WDs and NWDs in our sample. The exposure ratio is $\hat{S}_i/\hat{r}_{c(i)}$, the exposure to WDs under the experienced measure over the exposure to WDs for the residential measure. There is much more variation in exposures for NWDs than WDs, suggesting that the primary mechanism for the greater integration we measure relative to residential isolation is driven by NWD exposures.



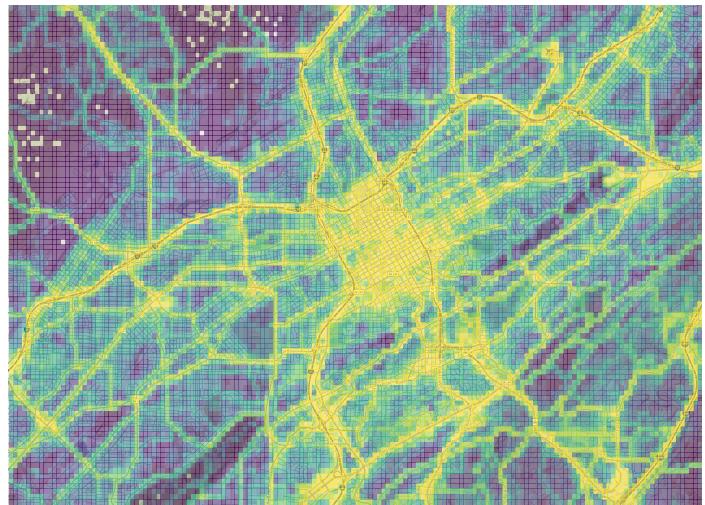
Online Appendix Figure A11: Exposure to WDs in Different Features, Decomposition by Race

Notes: The vertical lines show mean exposures in our baseline specification. The population weighted mean across MSAs of exposure for WDs and NWDs is represented by open and filled points respectively. The distance between any pair of points represents the isolation index in that feature. If the points overlap, isolation is zero. If the WD and NWD populations were contributing equally to their change in exposure, the points would meet at the dotted line splitting the difference between the baseline estimates.



Online Appendix Figure A12: Average Share of Pings in Features by Device Group

Notes: We plot the average share of pings in each feature for WDs and NWDs by hour. The solid and dashes lines depict the average for NWDs and WDs respectively. For several features there are interesting differences between groups: NWDs spend more time at civil, religious and social organizations, restaurants and bars, retail establishments and outdoor spaces but less time on roads and at airports. We note that time spent in one's home geohash7, however, is similar for both groups.



Online Appendix Figure A13: Activity in Birmingham, AL

Notes: We depict the level of activity in pings across geohash7s in Birmingham, AL. The number of pings increases as the color moves from blue to yellow. Activity seems to be concentrated on roads and in the central area of the city.

Online Appendix Table A1: InfoUSA NAICS8 Categories and Combined Feature Category

Combined category	NAICS8 Category	# of items	Share of all items in infoUSA dataset
Retail	Supermarkets/Other Grocery (Exc Convenience) Strs	90850	0.58 %
Retail	Pharmacies & Drug Stores	73215	0.47 %
Retail	Convenience Stores	72947	0.47 %
Retail	Used Merchandise Stores	62472	0.4 %
Retail	All Other General Merchandise Stores	59747	0.38 %
Retail	Gift, Novelty & Souvenir Stores	54629	0.35 %
Retail	Women's Clothing Stores	40032	0.26 %
Retail	Beer, Wine & Liquor Stores	39694	0.25 %
Retail	Other Clothing Stores	33709	0.22 %
Retail	Retail Bakeries	29162	0.19 %
Retail	Hobby, Toy & Game Stores	26008	0.17 %
Retail	Department Stores (Except Discount Dept Stores)	24993	0.16 %
Retail	Optical Goods Stores	22419	0.14 %
Retail	Hardware Stores	22188	0.14 %
Retail	All Other Specialty Food Stores	21094	0.13 %
Retail	Food (Health) Supplement Stores	20193	0.13 %
Retail	All Other Health & Personal Care Stores	17774	0.11 %
Retail	Book Stores	15910	0.1 %
Retail	Clothing Accessories Stores	15578	0.1 %
Retail	Office Supplies & Stationery Stores	12213	0.08 %
Retail	Paint & Wallpaper Stores	10619	0.07 %
Retail	Children's & Infants' Clothing Stores	9963	0.06 %
Retail	Electronic Shopping	9871	0.06 %
Retail	Men's Clothing Stores	9420	0.06 %
Retail	Meat Markets	9398	0.06 %
Restaurants_bars	Full-Service Restaurants	607719	3.89 %
Restaurants_bars	Snack & Nonalcoholic Beverage Bars	66575	0.43 %
Restaurants_bars	Limited-Service Restaurants	16778	0.11 %
Civil_social_religious_organizations	Religious Organizations	386741	2.47 %
Civil_social_religious_organizations	Civil & Social Organizations	64645	0.41 %
Education	Elementary & Secondary Schools	174380	1.12 %
Education	Colleges, Universities & Professional Schools	27442	0.18 %
Education	Libraries & Archives	26059	0.17 %
Education	Museums	19108	0.12 %
Accommodation	Hotels (Except Casino Hotels) & Motels	70789	0.45 %
Accommodation	All Other Traveler Accommodation	12420	0.08 %
Accommodation	Bed-&-Breakfast Inns	10338	0.07 %
Sports_recreation	Fitness & Recreational Sports Centers	65877	0.42 %
Entertainment	Motion Picture Theaters (Except Drive-Ins)	8763	0.06 %
Entertainment	Theater Companies & Dinner Theaters	6484	0.04 %

Online Appendix Table A2: Home Census Tract Summary Statistics for Devices in the Sample.

	US mean	Sample mean
Female	0.508	0.509
Bachelor's Degree	0.114	0.121
Median Age	37.385	37.431
Median Income (in 1000s of USD)	28.618	29.727
Population in Poverty	0.133	0.124
Unemployment Rate	0.039	0.038

Notes: We aggregate the demographics for the home census tracts in which we observe devices. Data are collected from the 2010 census. Columns show US averages as well as mean for the unweighted device sample. Including sample weights as described in Section 2.4 allows us to exactly recover US averages with the device sample.

Online Appendix Table A3: Summary Statistics for Measures of Activity of Devices in Sample.

	Median	Mean
Number of days active	51.00	56.92
Number of hours / active day	7.10	9.45
Number of geohash7s visited / active day	9.68	22.95
Number of pings / active day	33.88	86.84
Percent of pings at home (narrowly defined)	36.79	42.15
Number of geohash7s visited overall	195.00	720.85

Notes: All statistics are weighted using the sample weights described in Section 2.4. An active day is a day on which we see at least one ping for the device.

Online Appendix Table A4: Summary of Variables and Sources

Variable	Description	Source
Median Age	Median Age	2010 ACS variable B01002_001
Median Income	Median Income In The Past 12 Months (In 2010 Inflation-Adjusted Dollars)	2010 ACS variable B06011_001
Population in Poverty	Count Of Individuals With Income Below Poverty Level For The Past 12 Months	2010 ACS variable B17001_002
Unemployment Count	Unemployment Count	Sum of 2010 ACS variables B17005_006, B17005_011, B17005_017 and B17005_022
Black Alone	Single Race Non-Hispanic Black Population Count	2010 Decennial Census variable P009006
Black Alone or in Combination	Single Or Multiracial Non-Hispanic Black Population Count	Sum of 2010 Decennial Census variables P009013, P009018, P009019, P009020, P009021, P009029, P009030, P009031, P009032, P009039, P009040, P009041, P009042, P009043, P009044, P009050, P009051, P009052, P009053, P009054, P009055, P009060, P009061, P009062, P009063, P009066, P009067, P009068, P009069, P009071 and P009073
Total Population	Total Population	2010 Decennial Census variable P009001
White Alone	Single Race Non-Hispanic White Population Count	2010 Decennial Census variable P009005
Population Density	Population per square mile	2010 Decennial Census variables P009001 and SUBHD0303
Public Transit Use	Share of working population using public transprotation to get to work	2010 ACS variable B08101
Share with Bachelor's	Share of population with at least a Bachelor's degree	2010 ACS variables B06009_005 and B06009_006
Black Income Mobility	share of black individuals born in the 25th percentile of the income distribution who make it to the top quintile	Average Chetty et al.'s (2018) pooled by race county estimate kfr.top20.black.pooled.p25
White Income Mobility	share of white individuals born in the 25th percentile of the income distribution who make it to the top quintile	Average Chetty et al.'s (2018) pooled by race county estimate kfr.top20.white.pooled.p25

Online Appendix Table A5: Experienced and Residential Isolation by MSA

MSA	Exp	Res	MSA	Exp	Res
Abilene, TX	0.30	0.46	Lansing, MI	0.42	0.48
Akron, OH	0.46	0.61	Laredo, TX	0.22	0.05
Albany, GA	0.47	0.54	Las Cruces, NM	0.39	0.35
Albany, NY	0.48	0.68	Las Vegas, NV	0.46	0.62
Albuquerque, NM	0.43	0.51	Lawrence, KS	0.31	0.10
Alexandria, LA	0.44	0.58	Lawton, OK	0.32	0.24
Allentown, PA	0.48	0.68	Lebanon, PA	0.41	0.44
Altoona, PA	-0.00	0.00	Lewiston, ID	0.54	0.41
Amarillo, TX	0.48	0.73	Lewiston, ME	0.50	0.23
Ames, IA	0.35	0.10	Lexington, KY	0.36	0.41
Anchorage, AK	0.40	0.44	Lima, OH	0.44	0.46
Anderson, IN	0.41	0.41	Lincoln, NE	0.30	0.27
Anderson, SC	0.37	0.36	Little Rock, AR	0.49	0.65
Ann Arbor, MI	0.43	0.47	Logan, UT	0.34	0.08
Anniston, AL	0.43	0.45	Longview, TX	0.38	0.41
Appleton, WI	0.42	0.21	Longview, WA	0.24	0.03
Asheville, NC	0.36	0.21	Los Angeles, CA	0.48	0.77
Athens, GA	0.41	0.39	Louisville, KY	0.45	0.63
Atlanta, GA	0.51	0.63	Lubbock, TX	0.39	0.49
Atlantic City, NJ	0.45	0.58	Lynchburg, VA	0.39	0.43
Auburn, AL	0.36	0.32	Macon, GA	0.47	0.58
Augusta, GA	0.44	0.46	Madera, CA	0.59	0.61
Austin, TX	0.44	0.60	Madison, WI	0.40	0.40
Bakersfield, CA	0.55	0.68	Manchester, NH	0.36	0.18
Baltimore, MD	0.52	0.71	Manhattan, KS	0.42	0.42
Bangor, ME	0.53	0.63	Mankato, MN	0.09	0.02
Barnstable Town, MA	0.30	0.13	Mansfield, OH	0.21	0.36
Baton Rouge, LA	0.47	0.58	McAllen, TX	0.43	0.11
Battle Creek, MI	0.43	0.48	Medford, OR	0.33	0.06
Bay City, MI	0.04	0.03	Memphis, TN	0.52	0.66
Beaumont, TX	0.50	0.72	Merced, CA	0.38	0.34
Bellingham, WA	0.38	0.23	Miami, FL	0.49	0.71
Bend, OR	0.21	0.01	Michigan City, IN	0.43	0.51
Billings, MT	0.40	0.20	Midland, TX	0.38	0.48
Binghamton, NY	0.22	0.12	Milwaukee, WI	0.60	0.88
Birmingham, AL	0.57	0.71	Minneapolis, MN	0.41	0.58
Bismarck, ND	0.28	0.03	Missoula, MT	0.18	0.05
Blacksburg, VA	0.13	0.02	Mobile, AL	0.49	0.64
Bloomington, IL	0.34	0.13	Modesto, CA	0.44	0.47
Bloomington, IN	0.32	0.10	Monroe, LA	0.51	0.63
Boise City, ID	0.34	0.22	Monroe, MI	0.32	0.23
Boston, MA	0.45	0.68	Montgomery, AL	0.48	0.62
Boulder, CO	0.39	0.29	Morgantown, WV	0.01	0.05
Bowling Green, KY	0.32	0.37	Morristown, TN	0.27	0.07
Bremerton, WA	0.37	0.05	Mount Vernon, WA	0.36	0.21
Bridgeport, CT	0.47	0.76	Muncie, IN	0.41	0.82
Brownsville, TX	0.42	0.29	Muskegon, MI	0.51	0.75
Brunswick, GA	0.40	0.53	Myrtle Beach, SC	0.33	0.24
Buffalo, NY	0.52	0.78	Napa, CA	0.35	0.39
Burlington, NC	0.45	0.53	Naples, FL	0.48	0.55
Burlington, VT	0.16	0.01	Nashville, TN	0.45	0.63
Canton, OH	0.38	0.48	New Haven, CT	0.48	0.72
Cape Coral, FL	0.45	0.52	New Orleans, LA	0.45	0.65
Cape Girardeau, MO	0.48	0.42	New York, NY	0.51	0.80
Carson City, NV	0.45	0.44	Niles, MI	0.55	0.73
Casper, WY	0.09	0.01	North Port, FL	0.42	0.52
Cedar Rapids, IA	0.27	0.13	Norwich, CT	0.38	0.49
Champaign, IL	0.33	0.52	Ocala, FL	0.39	0.39
Charleston, SC	0.36	0.44	Ocean City, NJ	0.33	0.34
Charleston, WV	0.33	0.38	Odessa, TX	0.39	0.44
Charlotte, NC	0.47	0.62	Ogden, UT	0.38	0.50

Notes: We report baseline estimates of experienced and residential isolation for each Metropolitan Statistical Area in alphabetical order here and in Tables A6 and A7.

Online Appendix Table A6: Experienced and Residential Isolation by MSA: Continued

MSA	Exp	Res	MSA	Exp	Res
Charlottesville, VA	0.30	0.23	Oklahoma City, OK	0.44	0.62
Chattanooga, TN	0.46	0.67	Olympia, WA	0.40	0.13
Cheyenne, WY	0.30	0.10	Omaha, NE	0.49	0.68
Chicago, IL	0.52	0.73	Orlando, FL	0.46	0.55
Chico, CA	0.39	0.28	Oshkosh, WI	0.37	0.05
Cincinnati, OH	0.47	0.67	Owensboro, KY	0.34	0.10
Clarksville, TN	0.39	0.38	Oxnard, CA	0.55	0.72
Cleveland, OH	0.56	0.78	Palm Bay, FL	0.36	0.33
Cleveland, TN	0.22	0.07	Palm Coast, FL	0.31	0.02
Coeur d'Alene, ID	0.58	0.25	Panama City, FL	0.35	0.42
College Station, TX	0.41	0.53	Parkersburg, WV	0.19	0.01
Colorado Springs, CO	0.45	0.53	Pascagoula, MS	0.49	0.51
Columbia, MO	0.30	0.10	Pensacola, FL	0.39	0.48
Columbia, SC	0.45	0.52	Peoria, IL	0.46	0.69
Columbus, GA	0.48	0.65	Philadelphia, PA	0.53	0.74
Columbus, IN	0.37	0.05	Phoenix, AZ	0.52	0.70
Columbus, OH	0.50	0.66	Pine Bluff, AR	0.55	0.70
Corpus Christi, TX	0.42	0.55	Pittsburgh, PA	0.43	0.64
Corvallis, OR	0.22	0.15	Pittsfield, MA	0.28	0.07
Crestview, FL	0.32	0.10	Pocatello, ID	0.37	0.39
Cumberland, MD	0.25	0.09	Port St. Lucie, FL	0.41	0.39
Dallas, TX	0.48	0.64	Portland, ME	0.23	0.12
Dalton, GA	0.40	0.39	Portland, OR	0.34	0.30
Danville, IL	0.41	0.40	Poughkeepsie, NY	0.44	0.58
Danville, VA	0.45	0.41	Prescott, AZ	0.30	0.06
Davenport, IA	0.36	0.36	Providence, RI	0.49	0.70
Dayton, OH	0.56	0.82	Provo, UT	0.29	0.16
Decatur, AL	0.40	0.48	Pueblo, CO	0.43	0.49
Decatur, IL	0.44	0.52	Punta Gorda, FL	0.27	0.07
Deltona, FL	0.38	0.40	Racine, WI	0.47	0.50
Denver, CO	0.48	0.71	Raleigh, NC	0.41	0.44
Des Moines, IA	0.37	0.54	Rapid City, SD	0.28	0.18
Detroit, MI	0.59	0.82	Reading, PA	0.59	0.87
Dothan, AL	0.38	0.36	Redding, CA	0.21	0.04
Dover, DE	0.39	0.25	Reno, NV	0.40	0.53
Dubuque, IA	0.19	0.03	Richmond, VA	0.46	0.58
Duluth, MN	0.39	0.30	Riverside, CA	0.48	0.57
Durham, NC	0.43	0.51	Roanoke, VA	0.46	0.66
Eau Claire, WI	0.36	0.03	Rochester, MN	0.36	0.15
El Centro, CA	0.48	0.37	Rochester, NY	0.57	0.83
El Paso, TX	0.34	0.39	Rockford, IL	0.46	0.53
Elizabethtown, KY	0.38	0.27	Rocky Mount, NC	0.43	0.29
Elkhart, IN	0.46	0.43	Rome, GA	0.39	0.35
Elmira, NY	0.14	0.70	Sacramento, CA	0.53	0.68
Erie, PA	0.35	0.65	Saginaw, MI	0.52	0.76
Eugene, OR	0.20	0.04	Salem, OR	0.44	0.41
Evansville, IN	0.39	0.41	Salinas, CA	0.53	0.74
Fairbanks, AK	0.29	0.23	Salisbury, MD	0.37	0.44
Fargo, ND	0.27	0.10	Salt Lake City, UT	0.43	0.54
Farmington, NM	0.44	0.46	San Angelo, TX	0.37	0.49
Fayetteville, AR	0.44	0.45	San Antonio, TX	0.49	0.65
Fayetteville, NC	0.40	0.40	San Diego, CA	0.47	0.67
Flagstaff, AZ	0.47	0.62	San Francisco, CA	0.42	0.70
Flint, MI	0.57	0.74	San Jose, CA	0.42	0.57
Florence, AL	0.30	0.23	San Luis Obispo, CA	0.28	0.39
Florence, SC	0.38	0.30	Sandusky, OH	0.38	0.18
Fond du Lac, WI	0.16	0.04	Santa Barbara, CA	0.43	0.54
Fort Collins, CO	0.36	0.29	Santa Cruz, CA	0.57	0.74
Fort Smith, AR	0.40	0.50	Santa Fe, NM	0.44	0.59
Fort Wayne, IN	0.53	0.76	Santa Rosa, CA	0.37	0.47
Fresno, CA	0.47	0.62	Savannah, GA	0.43	0.51

Notes: We report baseline estimates of experienced and residential isolation for each Metropolitan Statistical Area in alphabetical order here and in Tables A5 and A7.

Online Appendix Table A7: Experienced and Residential Isolation by MSA: Continued

MSA	Exp	Res	MSA	Exp	Res
Gadsden, AL	0.44	0.54	Scranton, PA	0.40	0.35
Gainesville, FL	0.39	0.51	Seattle, WA	0.39	0.49
Gainesville, GA	0.47	0.62	Sebastian, FL	0.40	0.33
Glens Falls, NY	0.18	0.02	Sheboygan, WI	0.21	0.10
Goldsboro, NC	0.41	0.39	Sherman, TX	0.42	0.33
Grand Forks, ND	0.27	0.06	Shreveport, LA	0.46	0.63
Grand Junction, CO	0.33	0.10	Sioux City, IA	0.45	0.56
Grand Rapids, MI	0.44	0.58	Sioux Falls, SD	0.28	0.12
Great Falls, MT	0.32	0.03	South Bend, IN	0.47	0.65
Greeley, CO	0.45	0.52	Spartanburg, SC	0.40	0.47
Green Bay, WI	0.39	0.38	Spokane, WA	0.22	0.04
Greensboro, NC	0.47	0.58	Springfield, IL	0.41	0.58
Greenville, NC	0.42	0.30	Springfield, MA	0.54	0.72
Greenville, SC	0.38	0.38	Springfield, MO	0.12	0.01
Gulfport, MS	0.41	0.40	Springfield, OH	0.44	0.69
Hagerstown, MD	0.30	0.57	St. Cloud, MN	0.29	0.05
Hanford, CA	0.44	0.48	St. George, UT	0.27	0.05
Harrisburg, PA	0.51	0.67	St. Joseph, MO	0.23	0.07
Harrisonburg, VA	0.42	0.24	St. Louis, MO	0.57	0.76
Hartford, CT	0.51	0.74	State College, PA	0.30	0.21
Hattiesburg, MS	0.40	0.40	Steubenville, OH	0.42	0.20
Hickory, NC	0.32	0.23	Stockton, CA	0.45	0.49
Hinesville, GA	0.44	0.36	Sumter, SC	0.41	0.36
Holland, MI	0.43	0.32	Syracuse, NY	0.53	0.75
Honolulu, HI	0.38	0.66	Tallahassee, FL	0.39	0.48
Hot Springs, AR	0.29	0.24	Tampa, FL	0.46	0.61
Houma, LA	0.36	0.23	Terre Haute, IN	0.16	0.06
Houston, TX	0.48	0.66	Texarkana, TX	0.37	0.41
Huntington, WV	0.25	0.22	Toledo, OH	0.47	0.67
Huntsville, AL	0.47	0.59	Topeka, KS	0.44	0.49
Idaho Falls, ID	0.23	0.07	Trenton, NJ	0.49	0.63
Indianapolis, IN	0.50	0.65	Tucson, AZ	0.49	0.70
Iowa City, IA	0.38	0.15	Tulsa, OK	0.45	0.56
Ithaca, NY	0.32	0.23	Tuscaloosa, AL	0.46	0.49
Jackson, MI	0.38	0.60	Tyler, TX	0.43	0.58
Jackson, MS	0.51	0.61	Utica, NY	0.40	0.67
Jackson, TN	0.43	0.58	Valdosta, GA	0.42	0.47
Jacksonville, FL	0.46	0.54	Vallejo, CA	0.46	0.55
Jacksonville, NC	0.34	0.35	Victoria, TX	0.35	0.43
Janesville, WI	0.43	0.43	Vineland, NJ	0.46	0.53
Jefferson City, MO	0.32	0.20	Virginia Beach, VA	0.42	0.55
Johnson City, TN	0.25	0.10	Visalia, CA	0.40	0.34
Johnstown, PA	0.38	0.17	Waco, TX	0.45	0.60
Jonesboro, AR	0.39	0.17	Warner Robins, GA	0.40	0.31
Joplin, MO	0.28	0.17	Washington, DC	0.47	0.68
Kalamazoo, MI	0.44	0.49	Waterloo, IA	0.43	0.61
Kankakee, IL	0.46	0.68	Wausau, WI	0.23	0.03
Kansas City, MO	0.51	0.73	Wenatchee, WA	0.34	0.16
Kennewick, WA	0.47	0.57	Wheeling, WV	0.24	0.13
Killeen, TX	0.50	0.58	Wichita Falls, TX	0.38	0.43
Kingsport, TN	0.29	0.06	Wichita, KS	0.42	0.57
Kingston, NY	0.33	0.35	Williamsport, PA	0.30	0.20
Knoxville, TN	0.42	0.61	Wilmington, NC	0.35	0.34
Kokomo, IN	0.27	0.42	Winchester, VA	0.40	0.27
La Crosse, WI	0.41	0.06	Winston, NC	0.48	0.60
Lafayette, IN	0.33	0.37	Worcester, MA	0.39	0.50
Lafayette, LA	0.39	0.41	Yakima, WA	0.52	0.63
Lake Charles, LA	0.46	0.71	York, PA	0.50	0.71
Lake Havasu City, AZ	0.37	0.22	Youngstown, OH	0.48	0.68
Lakeland, FL	0.43	0.37	Yuba City, CA	0.40	0.28
Lancaster, PA	0.47	0.71	Yuma, AZ	0.49	0.48

Notes: We report baseline estimates of experienced and residential isolation for each Metropolitan Statistical Area in alphabetical order here and in Tables A5 and A6.

Online Appendix Table A8: Regression Coefficients across Samples

	Baseline	Top 50	Top 100	Top 200
Share with Bachelor's	-0.2491 (0.0428)	-0.426 (0.1221)	-0.3269 (0.0812)	-0.1403 (0.0647)
Median income (thousands)	-0.0017 (0.0005)	-0.0032 (0.0014)	-0.0025 (0.0008)	-0.0007 (0.0008)
Unemployment rate	1.7255 (0.2842)	1.0394 (0.918)	1.8567 (0.4544)	1.5621 (0.4083)
White mobility measure	-0.2302 (0.0449)	-0.4903 (0.1105)	-0.2561 (0.0938)	-0.0848 (0.087)
Black mobility measure	-0.146 (0.105)	-1.3877 (0.3774)	-0.5263 (0.2158)	0.0567 (0.1525)
log(Population density)	-0.0015 (0.0007)	-0.0044 (0.0026)	-0.0003 (0.0011)	-0.0013 (0.0009)
Public transit use	-0.012 (0.0026)	-0.0273 (0.0073)	-0.0122 (0.0052)	0.0015 (0.0041)
Median age	-0.1236 (0.0295)	-0.3949 (0.0881)	-0.3376 (0.0888)	-0.245 (0.104)

Notes: We report the coefficient and standard error from our baseline population weighted regression of experienced isolation on fifteen residential isolation bin fixed effects and the specified covariate. We also consider the same regression unweighted and estimated on subsamples of the top 50, 100, and 200 most populous MSAs.

Online Appendix Table A9: Experienced isolation by Feature Type

	q5	Mean	Median	q95	Correl. with base- line	N
Baseline	0.323	0.459	0.477	0.557	1.000	366
Features						
Accommodation	0.007	0.113	0.115	0.205	0.674	366
Civil, Religious And Social Organizations	0.044	0.251	0.269	0.400	0.888	366
Education	0.037	0.235	0.253	0.352	0.859	366
Entertainment	0.000	0.112	0.117	0.199	0.635	362
Outdoor Spaces (Parks, Etc.)	0.050	0.231	0.259	0.356	0.858	366
Restaurants And Bars	0.021	0.200	0.210	0.334	0.831	366
Retail	0.025	0.220	0.233	0.356	0.863	366
Roads And Airports	0.023	0.155	0.157	0.264	0.813	366
Sports And Recreation	0.008	0.164	0.172	0.267	0.761	365
No Features	0.342	0.498	0.526	0.593	0.944	366
Homes						
No Features, Not At Home (Broadly Defined)	0.065	0.261	0.278	0.392	0.887	366
At Home (Narrowly Defined)	0.509	0.672	0.703	0.744	0.939	366
No Homes (Narrowly Defined)	0.117	0.285	0.297	0.393	0.947	366
Outside Home Tract	0.029	0.208	0.216	0.327	0.886	366
Within Home Tract	0.450	0.630	0.662	0.713	0.961	366

Notes: We report summary statistics for different specifications of our measure weighted by MSA population. We consider measure experienced isolation restricted to various features and degrees of home proximity.

Online Appendix Table A10: Robustness

Experienced Isolation							
		q5	Mean	Median	q95	Correl.	N
						with base- line	
1	Baseline	0.323	0.459	0.477	0.557	1.000	366
Robustness checks							
2	No Roads Or Airports	0.339	0.489	0.509	0.586	0.991	366
3	Only Pings < 12mph	0.363	0.507	0.524	0.599	0.996	366
4	All (L2 Imputation)	0.244	0.390	0.409	0.481	0.908	366
5	All (Infutor imputation)	0.251	0.421	0.437	0.530	0.910	366
6	All (Day-Hour Weighting)	0.260	0.409	0.427	0.512	0.984	366
7	Without Out-Of-Towners	0.341	0.476	0.491	0.577	0.990	366
8	Without Top 5% Active Users In Terms Of Pings Per Day	0.333	0.476	0.495	0.572	0.992	366
9	Exclude Night Hours	0.305	0.441	0.459	0.542	0.999	366
10	All (Without Leave-One-Out Exposures)	0.384	0.499	0.506	0.596	0.937	366

Notes: We report summary statistics for different specifications of our measure. We consider excluding transportation features like roads, airports, or devices moving fast enough to be considered in transit. We also consider different subsamples of users, weighting schemes, and demographic data sources.

Online Appendix Table A11: Sample Statistics Restricting Exposure During Transportation

	Devices	Geohash7s	Pings
Baseline	17,397,580	98,853,493	101,989,194,959
No roads or airports	17,328,912	91,277,728	76,324,902,186
Only pings < 12mph	17,381,896	79,837,642	68,019,968,506
Only pings < 8mph	17,381,803	75,775,799	65,123,085,955
Only pings < 4mph	17,381,553	69,193,133	60,991,437,300
Only pings < 4mph & no roads or airports	17,307,535	62,639,284	53,991,029,734

Notes: We remove pings emitted at speeds exceeding different thresholds or on transport infrastructure and report counts of devices, geohash7s, and pings on these subsamples.

Online Appendix Table A12: Summary Statistics for Alternative Measures of Isolation

		Exp. mean	Res. mean	Correl. with res.	Correl. with base- line
1	Baseline	0.459	0.605	0.864	1.000
Home geohash7 race					
2	White-nonwhite (63/37)	0.456	0.598	0.860	0.988
3	White-Black (50/50)	0.470	0.655	0.867	0.703
4	White-Black (70/70)	0.474	0.658	0.857	0.612
5	White-Black (90/90)	0.387	0.450	0.875	0.436
Direct race imputation					
6	White-nonwhite (direct)	0.185	0.311	0.930	0.838
7	White-Black (direct)	0.215	0.311	0.985	0.762

Notes: We report mean experienced and residential isolation along with their correlation under the alternative measures. We also report the correlation of the alternative measures with baseline. All estimates are weighted by MSA population. The specification White/Black (XX/YY) indicates that isolation is estimated between devices from white and black home geohash7s that are considered white if the share white is above XX and considered black if the share black is above YY. Direct means that individuals are assigned the indicated device groups probabilistically.