Lee Kuan Yew School of Public Policy

Asia Competitiveness Institute

ACI Research Paper #18-2022

Neighborhood Asymmetries and Visits

Jing Zhi LIM

Lucas SHEN

October 2022

Please cite this article as:

Lim, Jing Zhi and Lucas Shen "Neighborhood Asymmetries and Visits", Research Paper #18-2022, *Asia Competitiveness Institute Research Paper Series (October 2022)*

Neighborhood Asymmetries and Visits*

Jing Zhi Lim† Lucas Shen‡

October 2022

Abstract

Using GPS records for Singapore, we study how mismatches in wealth and ethnic mix relate to daily neighborhood visits. We pay particular attention to asymmetries. Mismatch in wealth in the direction of poor to wealthy neighborhoods is related to visits that are 5 percent higher (standard error of 1 percent). In contrast, the estimates for wealthy to poor neighborhoods are statistically zero. Our results on mismatch by ethnic mix reveal similar insights, suggesting that asymmetries—where people come from and go to—matter for our understanding of segregation. We build on our context of study to revisit an extant policy that integrates residents by ethnicity. Our counterfactual analysis highlights the importance of asymmetries. The efficacy of the policy would have been three times lower had we not accounted for asymmetries.

*We are especially grateful to CITYDATA.ai (Apurva "Apu" Kumar and Lee Yew Leong) for an early data grant and advice. The data granters otherwise had no role in the study design, data preparation, analysis, and manuscript preparation. An earlier version of this working paper has been briefly discussed at the 2021 UEA Summer School and the Singapore Geospatial Festival 2021. We thank Paul Cheung, Huang Yuting, and colleagues for their feedback. We thank Li Changtai for helpful advice on the housing price data and Lee Shu En for research assistance on policies and background. All interpretations in this manuscript are those of the authors and not necessarily those of affiliated institutions. An earlier version of this working paper was circulated under the title, "Segregation Across Neighborhoods in a Small City." JZL Email: jlim@pardeerand.edu. LS Email: lucas@lucasshen.com.

†Pardee RAND Graduate School.

‡Asia Competitiveness Institute, Lee Kuan Yew School of Public Policy, National University of Singapore.

Several well-established and still-growing studies suggest that who people know and their networks influence their social and economic outcomes. Traditionally, social exposure is measured via census records depending on residency (Hutchens 2001; Jones and Pebley 2014; Krivo et al. 2013; Le Roux et al. 2017; Palmer et al. 2013; Rodriguez-Moral and Vorsatz 2016; Sin 2002a). The immediate problem with this measure of residential exposure is that social networks extend far beyond where people reside. Social interactions occur elsewhere, such as where people travel, eat, and spend their leisure time. Another limitation is that interactions need not be symmetric (e.g., Dong et al. 2020; Rao 2019).

To quantify experienced segregation in a small, densely urban, and highly connected city, we spatially join a novel set of daily neighborhood flow data, which we aggregate from anonymized GPS records in Singapore, to census and housing microtransaction records. Using this novel dataset, we examine how flows of individuals across neighborhoods depend on mismatches in wealth and ethnicity. In particular, we focus on how asymmetries in the mismatches affect neighborhood visits.

We use a simple stylized framework to sharpen our approach. In this framework, the representative individual's utility from visiting a given neighborhood depends on neighborhood-day-specific utility, utility from mismatches in neighborhood characteristics, costs of the within-city travel, and other factors. This framework makes explicit that the aggregated flows of individuals across neighborhood depend on the level of neighborhood mismatch, conditional on area-by-day fixed effects as well as neighborhood-specific demographics, businesses, rainfall, and places of interest.

We first examine mismatch by wealth. We define poor and wealthy neighborhoods based on whether they fall below or above the 25th percentile in housing price from the microtransaction records (we fully test sensitivity to this threshold later). We

find that mismatch by wealth increases neighborhood visits by 2 percent ($\approx 100 \times .021$, as log points, with a standard error of 0.9) relative to neighborhood pairs without mismatch by wealth. This positive estimate implies that individuals derive utility when visiting neighborhoods that locate in a different part of the wealth spectrum.

To reveal the first form of asymmetry in how mismatch by wealth affects neighborhood visits, we decompose the mismatch measure into the directions in which mismatch occurs. We find that the increased neighborhood flow from mismatch by wealth is driven by individuals from the poorer neighborhoods visiting the wealthier neighborhoods, rather than vice versa. Our estimate of 0.048 (standard error of 0.011) suggests that when the mismatch occurs in the direction of a poor to a wealthy neighborhood, neighborhood visit is 5 percent higher. The estimate of -0.008 (standard error of 0.011), on the other hand, suggests that mismatch in the direction of a wealthy to a poor neighborhood is inconsequential on neighborhood visits. This finding suggests that while individuals derive utility from visiting neighborhoods that are different in wealth, this effect is driven by one direction of social interaction—poor to wealthy—which implies that the poor are exposed more to the wealthy than vice versa.

To examine the second form of asymmetry in the mismatch by wealth effects, we allow the effect of mismatch on wealth to have different effects on neighborhood visits depending on whether or not the destination neighborhood lays in the central region (which houses the central business district). Our estimates suggest that mismatch by wealth heavily drives inflow into neighborhood in the central region, with the estimate of 0.04 (standard error of 0.01) implying that inflow into central region neighborhoods is 4 percent higher when there is a mismatch by wealth. However, this reverses drastically when the destination neighborhood is outside the central region,

with an estimate of -0.035 (standard error of 0.015). This form of asymmetry implies that having neighborhoods in a favorable central location matters for segregated interaction, and ignoring such asymmetries understates segregation.

The last set of evidence we adduce is on mismatch by ethnicity. By focusing only on two broad groups—ethnic minorities and ethnic majorities—we find that mismatch by ethnicity decreases neighborhood visits. The estimate of -0.21 (standard error of 0.10) implies that a one standard deviation in mismatch by ethnicity decreases neighborhood visits by 1.5 percent ($\approx 100 \times 0.21 \times 0.07$, p-value < 0.05). The key asymmetry we focus on for mismatch by ethnicity is the direction of flow arising from minority-ethnic to majority-ethnic neighborhoods and majority-ethnic to minority-ethnic neighborhoods. Our estimates suggest that only one direction—visits from majority-ethnic to minority-ethnic neighborhoods—drives all the implied segregated interaction arising from ethnicity differences, with further evidence refuting the alternative explanation that our estimates are capturing geographical idiosyncrasies in ethnic neighborhoods.

We put our results through a series of robustness tests. First, while all the main results use the broader administrative boundary of census planning area-by-day fixed effects for consistency, because it is the specification that allows identification of the asymmetries, we show that using neighborhood-by-day fixed effects only serves to strengthen our conclusion in terms of magnitude and statistical significance. For the mismatch by ethnicity results where the measure is continuous (as a measure of probability), the alternative specification in constant elasticities using log of mismatch by ethnicity does not substantially change our conclusion. In addition, we put all the mismatch by wealth results through comprehensive sensitivity tests. For the main analyses, we define a poor neighborhood as one that falls under the

25th percentile in house price. We iteratively go from the 1st to 99th percentile as thresholds in defining poor neighborhoods, with a focus on the 20-point band around the 25th percentile to show that the estimates are not sensitive around our chosen threshold. In addition, the estimates from the sensitivity tests ordered by magnitude mitigate concerns that we choose our threshold to inflate effect sizes.

We build on our estimates and context of study for a simple evaluation of a key policy that integrates residents by their ethnic group. Our counterfactual case study implies that the residential ethnic integration policy in Singapore increases social interactions via daily visits by 2.9 percent. For context, this is lower than the effect from eliminating all ethnic mismatch in the 1st percentile neighborhood-pair. However, focusing on the asymmetry in ethnic mismatch paints a very different story. The reduction in majority-to-minority ethnic mismatch implied by the policy increases visits by 8.7 percent. Moreover, we would need to completely eliminate majority-to-minority ethnic mismatch from the top 26 percent of neighborhood-pairs to be able to match the increase in social interactions implied by the ethnic integration policy. However one decides to judge the value of the integration policy, it is clear that ignoring asymmetries in segregated interaction would have severely understated the value of the policy.

The key substantive contribution of this study is to quantify asymmetries in experienced segregation, using novel data and methods, to the extent that the effect of neighborhood mismatches in wealth and ethnicity on daily neighborhood visits is mediated by where people come from. We add to a limited set of studies providing insights into asymmetries in segregated interactions (Dong et al. 2020; Hilman et al. 2021). Using computational methods to large-scale credit card transaction data, Dong et al. (2020) find that interactions from poorer to wealthier neighbor-

hoods disproportionally contribute to segregation. Hilman et al. (2021) also applies computational methods to social location data and find an upward bias in visits to more affluent places. Compared to Dong et al. (2020) and Hilman et al. (2021), who have richer data across regions, our focus is more direct and targeted in revealing asymmetries. Our study also evaluates a policy that concerns policymakers in cities that are ethnically diverse.

Our results speak to the literature on experienced social segregation to the degree that daily neighborhood visits are linked to neighborhood socioeconomic characteristics. Past research has found that social connectedness improves startup success (Banerji and Reimer 2019) and labor market opportunities (Banerjee and Ingram 2018; Hensvik and Skans 2016). Segregation is correlated with greater violence (Cook et al. 2018) and worsens socioeconomic markers such as schooling, employment, and marriage (Chay et al. 2014; Chetty et al. 2016). Social mixing and face-to-face interactions, on the other hand, reduce discrimination (Rao 2019) and lead to knowledge spillovers (Atkin et al. 2022).

More broadly, we build on a growing set of literature, which typically uses novel data to quantify real-time experienced segregation of individuals from different socioeconomic backgrounds (e.g., Athey et al. 2020; Davis et al. 2019; Dong et al. 2020; Moro et al. 2021), including segregation in the online space (Bastos et al. 2018; Cinelli et al. 2021; Dong et al. 2020; Eytan et al. 2015; Gentzkow and Shapiro 2011). Even with increasing residential integration (Glaeser and Vigdor 2012), this literature speaks to the significance of social segregation beyond geographical spaces (Wang et al. 2018), which is crucial because of polarization and echo chambers (Levy and Razin 2019).[1]

[1] Our study thus contributes to studies highlighting social segregation and inequality in Singapore (e.g. Choe 2016; Leong et al. 2020; Loo et al. 2003; Sin 2002a; Wong 2013; Teo 2018). Other studies

In general, the findings also fit in the literature that uses non-conventional approaches to quantify social exposure (e.g., Athey et al. 2020; Davis et al. 2019; Dong et al. 2020; Moro et al. 2021), as well as an emerging set of studies where the availability of GPS data provides further understanding of human behavior.[2]

The rest of the paper discusses in turn: (i) the geography of Singapore, (ii) our GPS and house microtransactions data, (iii) measuring neighborhood wealth levels using house prices, (iv) how utility from neighborhood visits depends on neighborhood mismatch, (v) results for mismatch by wealth, (vi) results for mismatch by ethnicity, (vii) sensitivity to wealth cutoff levels, and (viii) application to a residential ethnic quota. The final section closes with a discussion.

Geography of Singapore

Singapore is a city-state in Southeast Asia, widely considered as a small but developed economy. The size of the city is approximately 725km^2 or 280 square miles, which is about five times as large as San Francisco (121km^2), 1.2 times as large as Madrid city (667km^2), and 0.45 times as large as London city (1,570km^2). Unlike other regions, there are few concentrated suburban areas. These are mostly private housing which houses the minority of residences ($<$ 20 percent). Another factor linked to segregated interactions is that public transit usage is high. Train and bus trips is on average close to 6 million per day, for a population size of close to 6 million

focusing on segregation in cities include Jones and Pebley 2014; Krivo et al. 2013; Le Roux et al. 2017; Luo et al. 2016 and Wang and Li 2016.

[2] Examples include transport (Chua et al. 2020), within-city commuting and non-commuting patterns (Miyauchi et al. 2021), how partisanship affects family ties (Chen and Rohla 2018), how policing spatially affects criminal behavior (Blattman et al. 2021), how GPS can be used to predict poverty and wealth when census records are sparse (Blumenstock et al. 2015, Kreindler and Miyauchi [Forthcoming]), and in contemporaneous studies using GPS data to examine safe-distancing in the Covid-19 pandemic (Allcott et al. 2020; Gollwitzer et al. 2020).

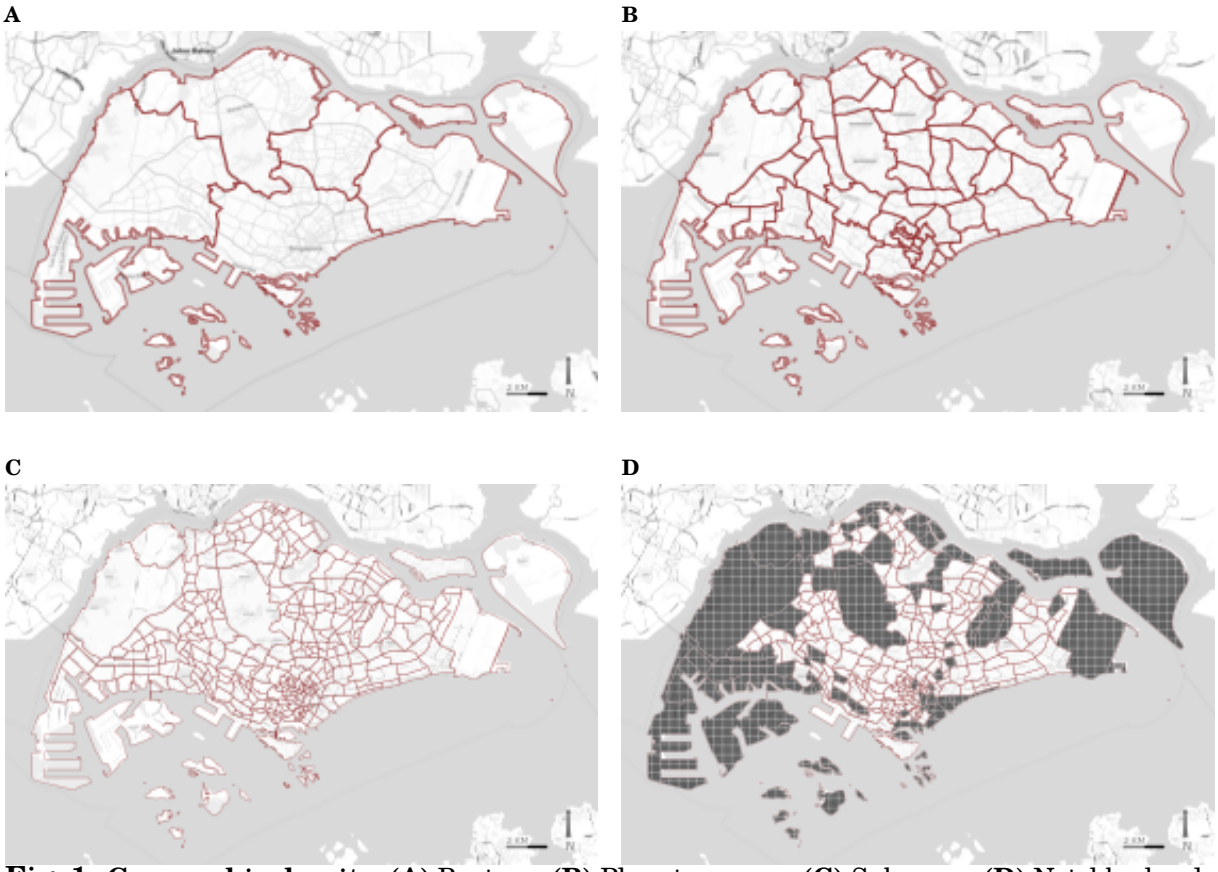

Fig. 1. Geographical units. **(A)** Regions. **(B)** Planning areas. **(C)** Subzones. **(D)** Neighborhoods (subzones that are not shaded; subzones with non-trivial density of residency).

(we account for transit-related amenities in our analyses).

For urban planning and census taking, Singapore is delineated, in increasing geographical subdivisions, into 5 regions, 55 planning areas, and more than 300 subzones (Fig. 1). For the purpose of this study, we term the subzones with census residential records as *neighborhoods* (Table 1) for ease of expositional. These neighborhoods are relatively small on average, with the representative neighborhood being about 1.35km^2 (0.52sq mi), which is comparable to the area of two geohash-6 grids (1.2km $\times$ 0.6km) stacked vertically, and with some neighborhoods as small as 0.05km^2 (0.02sq mi), which is comparable to two geohash-8 grids (38m $\times$ 19m).

Table 1. Descriptive statistics.

	Count/Mean ± s.d.	Min.	Max.
Device hashes	$> 17m$	.	.
Device hashes in sample (> 30 appearances)	$> 125k$	.	.
Census areas (coarser)	55	.	.
Subzones/neighborhoods (finer)	323	.	.
Census areas, with residential records	52	.	.
Subzones/neighborhoods, with residential records	219	.	.
Census areas, with GPS records	52	.	.
Subzones/neighborhoods, with GPS records	301	.	.
Subzone area (km^2)	2.23 ± 5.67	0.04	69.75
Neighborhood area (km^2) (subzones with residential records)	1.35 ± 1.24	0.05	8.45
Mismatch by wealth	0.3422 ± 0.4744	0	1
P $\longrightarrow$ W	0.1724 ± 0.3777	0	1
W $\longrightarrow$ P	0.1698 ± 0.3755	0	1
Mismatch by ethnicity	0.3685 ± 0.0686	0.0494	0.6955
Mnr $\longrightarrow$ Maj (ethnicity)	0.1878 ± 0.0685	0.0074	0.6880
Maj $\longrightarrow$ Mnr (ethnicity)	0.1807 ± 0.0744	0.0074	0.6880
$\lvert \text{wealth}_o - \text{wealth}_d \rvert$	0.0058 ± 0.0061	0	0.0319

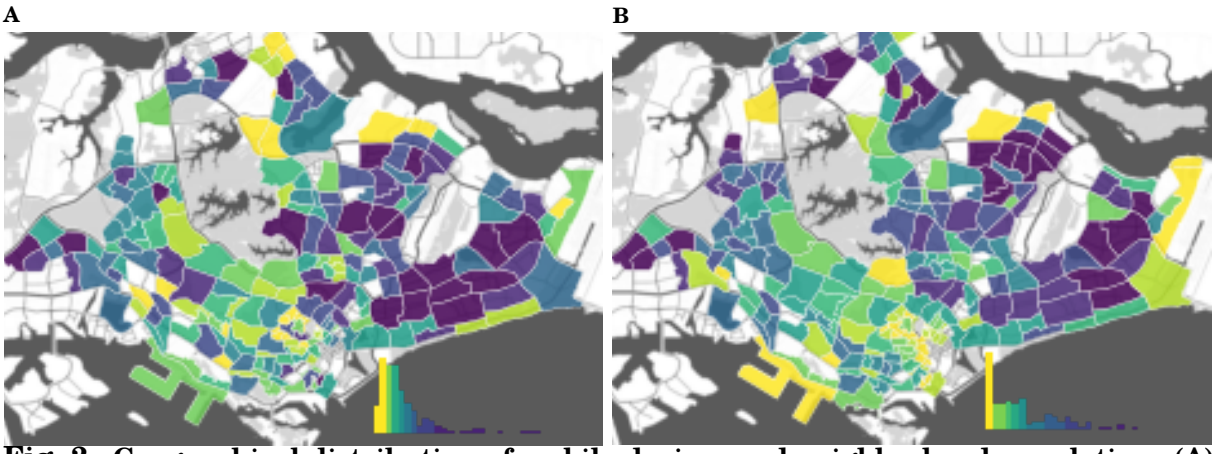

Fig. 2. Geographical distribution of mobile devices and neighborhood population. **(A)** Mobile devices. **(B)** Neighborhood resident population.

For a different sense of scale, our neighborhoods are comparable to US census tracts in terms of population size, but more comparable to the smaller and urban US census block groups in terms of land area because of the higher population density in Singapore with the majority of residents in high-rise flats (more below). These neighborhoods are the geographical units of analyses. In the discussion, we move up to the planning area level when using our neighborhood-derived estimates to

evaluate the ethnic integration policy because of historical data availability.[3]

Data

The backbone of our data is the GPS ping records from CITYDATA.ai, where we obtain the hashes of mobile devices present in a neighborhood in a given day. From this, we construct daily neighborhood flows by imputing the most frequented neighborhood as the origin neighborhood of a device. The sample period is for the 91 days in Jan–Mar 2020, which is before the city-wide COVID-19 lockdown on April 7.

We perform a set of tests to check for how representative the GPS pings are for the neighborhoods. GPS pings are higher in neighborhoods with larger residential population and neighborhood flows are highest when the destination neighborhood is geographically close to the origin neighborhood and highest with moderate levels of precipitation. The GPS pings otherwise do not vary much by various measures of neighborhood characteristics, as should be the case. Overall, neighborhood with denser residential population have more mobile devices captured (Fig. 2). Our supplementary appendix reports the full set of tests.

A second key data source is the microtransaction records for both public and private house prices from the URA REALIS database. We geocode the street addresses in the records to the neighborhood and then compute neighborhood wealth levels using the price per square meters, weighted by the proportion of residents in the neighborhood who live in private housing vs. public housings. This is the data we use to compute neighborhood wealth mismatch.

From here, we spatially merge other data sets to get neighborhood demographics

[3] We emphasize that what we refer to as neighborhoods is related to but distinct from the government's definition of neighborhoods (such as when implementing residential ethnic quotas, see below).

and characteristics, mainly using official data sources. We compute the ethnic mismatch measure using the official census records, where available. Table 1 provides some basic descriptions of the data, and the data appendix in the supplementary appendix provides full details of how we construct the data.

Neighborhood Wealth Levels

The GPS data, being anonymized, is agnostic of individual records, including their location of residence. To impute the origin neighborhood of individuals, we use the most frequented neighborhood inside the sample time frame as the location of origin. Since the anonymized data is also agnostic of socio-economic characteristics, we impute wealth levels of neighborhoods using the housing price per square meters (weighted by residence-type share, additional details available in the Supplementary Materials).[4]

We begin by corroborating the use of housing prices from micro-transaction records as a proxy for income at the coarser geographical level of the census planning area, where income census records are available (Fig. 3). As proposed, areas with higher housing prices have higher income, with most points falling close to the linearly fitted line, with income explaining 79% of the variation in our wealth proxy (or, a correlation coefficient of 0.89). This relationship is similar to Xu et al. (2018) who find a correlation coefficient of 0.89 when correlating housing price with self-reported income from the Household Interview Travel Survey at the same census planning area level.[5]

Fig. 3 also shows how the geographical distribution of individual housing records

[4] We stress that aggregated income records at the neighborhood (subzone) level are not available to us, hence the use of house price as a proxy for wealth.

[5] We emphasize that the travel survey data is not available to us.

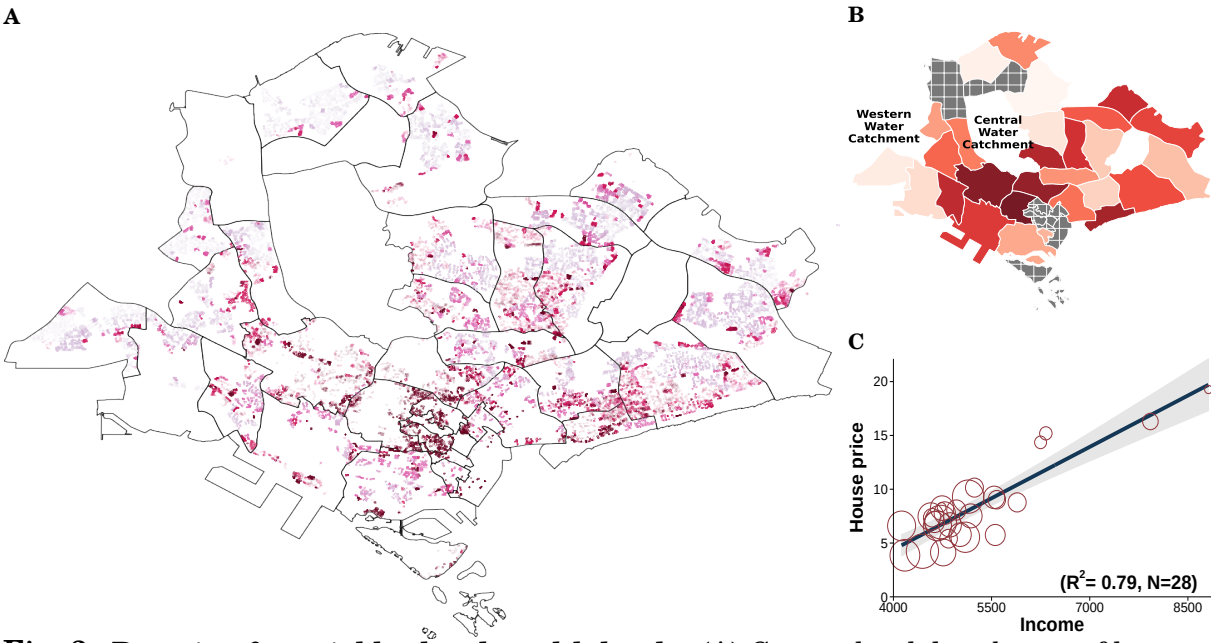

Fig. 3. Proxying for neighborhood wealth levels. **(A)** Geographical distribution of housing prices per square meters (past five years). Each point is a recorded transaction and darker shades indicate higher prices. **(B)** Geographical distribution of census income, by 28 (of 55) census planning areas. **(C)** House price and census income for the 28 areas. Marker size encodes resident population size.

compares to the distribution of census income at the 28 available census planning areas. Since the R^2 is itself a sample statistic, we resample the pairwise values from the 28 planning areas for 100,000 bootstrap samples and recompute the $\hat{R}^2$ values. This yields a left-skewed distribution with a 95% confidence bound of [0.43, 0.91], or equivalently, correlation coefficients of 0.66 to 0.95 (Fig. S23).[6]

[6] Another possible proxy for neighborhood wealth levels is nighttime lights. Certain measurements like the VIIRS (Visible Infrared Imaging Radiometer Suite) have high resolution. Still, with issues such as overglow, among others, we find house prices (especially after weighting by share of residents in public vs. private housing) to be a more direct proxy for neighborhood wealth.

Mismatch and Utility From Neighborhood Visits

Here we sketch a simple model where individuals derive utility via inter-neighborhood visits (as in Kreindler and Miyauchi (Forthcoming)]). An individual i going from neighborhood o to neighborhood d on day t derives total utility

$$U_{i,odt} = \frac{u_{dt}^{e_1} m_{od}^{e_2}}{D_{od}^{e_3}} \epsilon_{i,odt}$$

where u_{dt} is the utility individual i derives from visiting neighborhood d on day t, which subsumes wage-differentials across neighborhoods in a labor-specific context [as in Kreindler and Miyauchi (Forthcoming)] as well as utility from amenities (Miyauchi et al. 2021). For example, certain neighborhoods may coincide or be in close proximity to business districts and so would yield higher utility u_{dt}. Moreover, u_{dt} allows the utility of neighborhood d to differ across time, and in our analyses, it varies by day of the week. This nests the assumption that certain neighborhood would yield different utility depending on whether it is a weekday or a weekend (e.g. central business districts).

The key factor of our study is m_{od}. This is the utility (disutility if $e_2 < 0$) from visiting neighborhood d because of inherent mismatches in social dimensions with individual i's neighborhood of origin o. D_{od} is the disutility from travel costs between neighborhoods o and d. $\epsilon_{i,odt}$ is the individual-specific idiosyncratic utility shock.

Given the origin neighborhood o and day t, the probability that the representative individual i from neighborhood o visits d on day t is

$$f_{odt|ot} = \frac{u_{dt}^{e_1} m_{od}^{e_2} / D_{od}^{e_3}}{\sum_{\ell \neq o} \left(u_{\ell t}^{e_1} m_{o\ell}^{e_2} / D_{o\ell}^{e_3} \right)} \in [0, 1]$$

and taking logs gives

$$\log f_{odt|ot} = e_1 \log u_{dt} + e_2 \log m_{od} - e_3 \log D_{od} - \log \left\{ \sum_{\ell \neq o} exp\Big(e_1 \log u_{\ell t} + e_2 \log m_{o\ell} - e_3 \log D_{o\ell} \Big) \right\}$$

which can be estimated using

$$\log(\text{inflow})_{odt} = \theta_{dt} + \beta \log m_{od} + \delta \log D_{od} + \varepsilon_{odt}. \quad (1)$$

β in Eq. (1) is the key parameter of interest and captures the effect of mismatch in social dimensions across neighborhoods (defined in later in Eq. (2)) on cross-neighborhood movement. If the estimate of β is negative then this suggests that mismatch induces disutility from inter-neighborhood travels, and vice versa.[7]

θ_{dt} is the area-by-day fixed effects. Below, we use the planning census area-by-day fixed effects, where a census area is a geographical unit larger than a neighborhood, so that the regression analyses can identify the two components of the mismatch measures.[8]

δ captures the effect of distance between neighborhoods on neighborhood visits. In our analyses this includes centroid-based distances and contiguity indicators for each origin-to-destination neighborhood pair. More broadly, δ captures the effect of spatial frictions (as in Davis et al. 2019; Miyauchi et al. 2021) while β captures the effect of social frictions on flows between neighborhoods. To mitigate concerns that β approximates the true effect of neighborhood mismatches on neighborhood flows, we extend Eq. (1) to include the date fixed effects, which nests the day-of-week

[7] In the Results: Mismatch by Wealth and Results: Mismatch by Ethnicity analyses below, we switch from constant to semi-elasticities to accommodate the mismatch measure as a binary variable.

[8] We discuss this in more detail later. Using neighborhood-by-day fixed effects, however, does not change the conclusions as reported in our supplementary appendix (Table S8).

fixed effects, neighborhood-specific demographics, business, rainfall, and places of interest, where the effects are allowed to vary by time. The error term is ε_{odt}.

Letting z be a neighborhood-specific socio-economic characteristic, we define mismatch between neighborhoods as

$$\text{mismatch}_{od} = z_o(1 - z_d) + z_d(1 - z_o) \tag{2}$$

where z can be either an indicator variable or a proportion measure $\in$ [0,1]. For the mismatch by wealth measure, we let $z = 1$ when the neighborhood wealth proxy (house price per square meter) falls below the 25th percentile, and zero for neighborhoods above the 25th percentile.[9] For ease of exposition, $z = 1$ indicates a poor neighborhood while $z = 0$ indicates a wealthy neighborhood. This implies that mismatch defined in Eq. (2) is equal to one only when the origin and destination neighborhoods o and d fall into different categories—one poor and one wealthy, hence the mismatch, and zero otherwise.

The mismatch measure in Eq. (2) can be extended to continuous characteristics, in which case the mismatch measure is interpreted as the probability of mismatch between neighborhoods (Fig. S25). We do this for mismatch by ethnicity, where z is the proportion of minority-ethnic individuals in a neighborhood, so that mismatch by ethnicity defined using Eq. (2) is the probability of mismatch in ethnicity between the origin and destination neighborhoods.

To formalize how movement patterns across neighborhoods depends on mismatch, we use our novel dataset which we construct by merging the GPS-derived neighborhood flows with census and house microtransaction records, to estimate

$$\log(\text{inflow})_{odt} = \alpha + \beta\ \text{mismatch}_{od} + \Gamma_t X_{odt} + \varepsilon_{odt} \tag{3}$$

[9] We test all mismatch by wealth results for sensitivity to the 25th percentile cutoff below.

where inflow_{odt} is the probability that the representative resident from origin neighborhood o visits neighborhood d during day t. β is our key coefficient of interest, which tells us the effect of neighborhood mismatch in socioeconomic markers on neighborhood visits. As established in Eq. (1), the model includes spatial features that affect inter-neighborhood visits. These are all included as full interactions with the day fixed effects, allowing their effect to differ across days. Specifically, to account for spatial frictions, Eq. (3) includes the origin-to-destination neighborhood distances, a contiguity dummy, and neighborhood size.

We also include the fully interacted census area-by-day fixed effects, which subsumes the day-of-week fixed effects. This approach embeds the assumption that individuals derive utility from neighborhood visits that are area- and day-specific. To account for neighborhood density, we use both the census residence population and the transient urban population using the number of devices captured from the GPS records. This specification is our baseline model.[10]

More demanding specifications of Eq. (3) include neighborhood demographics such as the proportion of females (Dong et al. 2017) and age groups (below 20, 20 to 39, 40 to 64, and above 65);[11] the prevalence of services, manufacturing, and construction-related businesses (Miyauchi et al. 2021) with addresses registered in the neighborhoods; neighborhood- and day-specific rainfall (Fig. S24); and the POIs (places of interest and amenities, e.g., schools, transit stations, tourist attractions, etc.). These are included as full interactions with the day fixed effects. All estimations include date fixed effects, so that day-specific movement patterns (e.g. more leisure

[10] The area-by-day fixed effects subsumes the area-by-day utility derived from the neighborhood visits (Eq. (1)). This removes some of neighborhood visits relating to commute to workplaces, but at the same time, potentially removes certain non-commuting neighborhood visits. One study that disentangles commute from non-commute travels within the data pipeline is the one by Miyauchi et al. 2021. This is not possible with the resolution of the data we have in this study.

[11] These are the ones available to the public at the neighborhood (subzone) level.

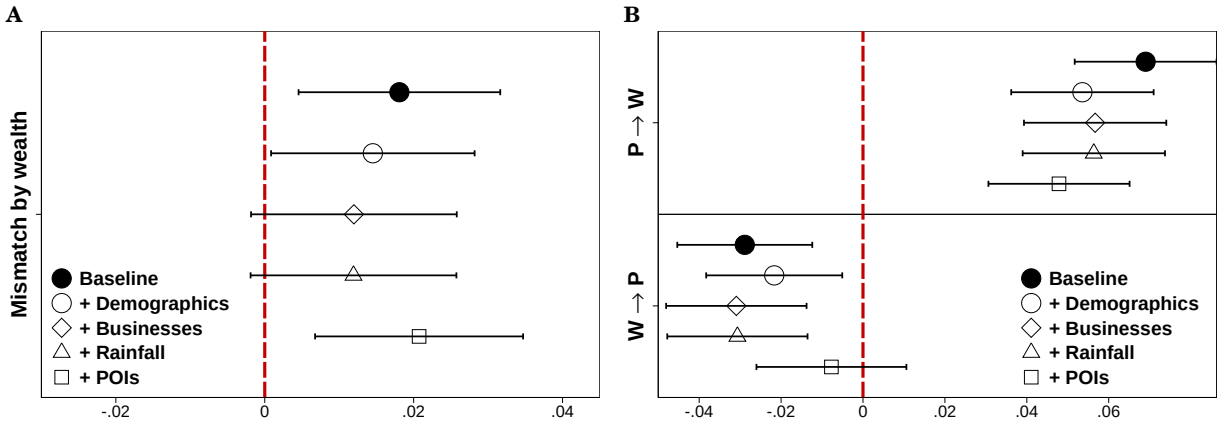

Fig. 4. Mismatch in wealth and movement. **(A)** Estimated β coefficients of the mismatch measure from estimating Eq. (3). **(B)** Asymmetry by direction. P $\rightarrow$ W indicates movement from poor to wealthy neighborhoods (or, $z_o(1 - z_d)$ in Eq. (2)). W $\rightarrow$ P indicates movement from wealthy to poor neighborhoods (or, $z_d(1 - z_o)$ in Eq. (2)). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Dependent variable is log inflow from origin o to destination d. Figure corresponds to Table S2 and Table S3. Capped horizontal lines are 90% confidence intervals from standard errors clustered at the origin-by-destination area level.

travels on weekends) are removed. Standard errors are clustered at the origin-by-destination census area level.

Results: Mismatch by Wealth

Fig. 4A reports the first set of results, from estimating Eq. (3), where mismatch in wealth between neighborhoods is associated with increased neighborhood visits on a daily basis. The solid black circle denotes the baseline specification, and the hollow markers show how the estimates change as additional controls are added. Using the final point estimate of 0.021 (standard error of 0.009) from the most demanding and our preferred specification (hollow square) suggests that mismatch in wealth increases neighborhood visits by 2.1 percent ($\approx 100 \times .021$, as log points; p-value $<$

0.05).[12]

This observed behavior where the mismatch by wealth increases neighborhood visits, however, could be driven asymmetrically by trip hops from the poorer to the wealthier neighborhoods than vice versa. Fig. 5 shows two contiguous neighborhoods with different wealth levels. We observe very different movement patterns for these two neighborhoods that are next to each other, such as differences in outflow frequency and choice of neighborhood visits. In particular, the farthest neighborhood that the representative individual visit from the wealthier neighborhood is similar in wealth levels (left panel), but the farthest neighborhood that the representative individual visit from the poorer neighborhood has a much higher wealth level (right panel).

To formalize this, Fig. 4B splits the mismatch by wealth measure into the directional components: $z_o(1 - z_d)$ and $z_d(1 - z_o)$. Specifically, instead of regressing inflow on the mismatch measure, we regress inflow on (i) $z_o(1 - z_d)$ which indicates movement from poor to wealthy ($P \rightarrow W$) and (ii) $z_d(1-z_o)$ which indicates movement from wealthy to poor ($W \rightarrow P$). This exercise reveals that the result in Fig. 4A is driven by the individuals from poorer neighborhoods visiting wealthier neighborhoods. The estimate of 0.048 for $P \rightarrow W$ implies that on average, the probability of an individual from a poor neighborhood visiting a wealthy neighborhood is 5

[12] In the Table S7 of our supplementary appendix, we directly use the euclidean distance of price per square meter to measure mismatch by wealth. The findings are similar: the point estimate of 6.45 implies that a standard deviation increase in euclidean distance of wealth increases inflow by approximately 3.9 percent ($\approx 100 \times 6.45 \times 0.006$, Table S7), with a corresponding standard error of 0.57.

The finding does not change if we swap out the census planning area-by-day fixed effects for the neighborhood-by-day fixed effects (Table S8 in the supplementary appendix). The estimate become slightly larger and more precise—an estimate of 0.027 (as opposed to 0.021) with a standard error of 0.0098 and p-value < 0.01. we retain the census planning area-by-day fixed effects in our main analyses to be consistent since the two individual components of the mismatch in Eq. (2) can only be identified with the broader planning area-by-day fixed effects.

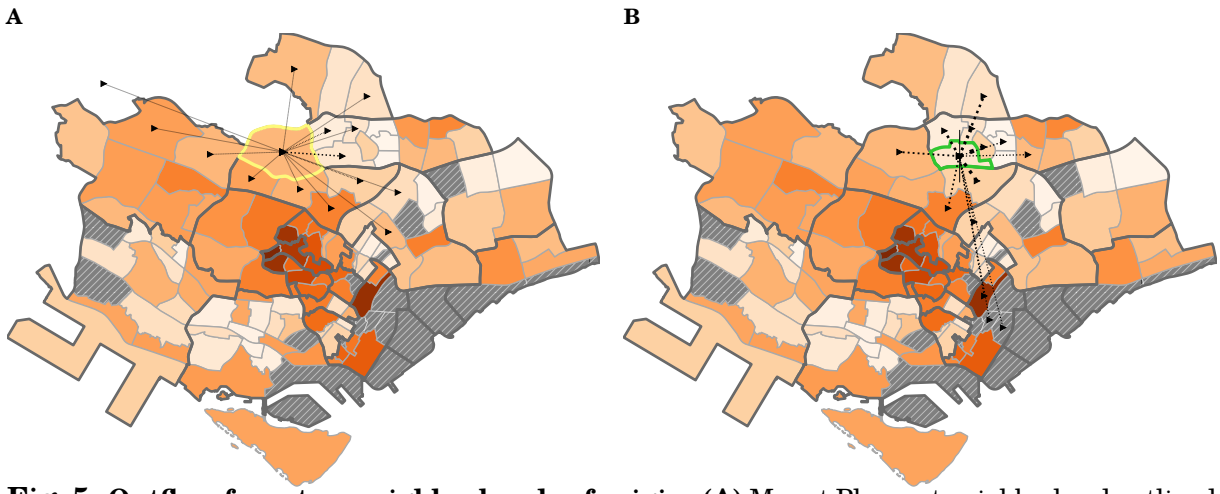

Fig. 5. Outflow from two neighborhoods of origin. **(A)** Mount Pleasant neighborhood outlined in yellow (wealthier). **(B)** Toa Payoh Central neighborhood outlined in green (poorer). Black lines show flow of movement from two neighborhoods of origin. Map is a cutout of the central region. Darker shades indicate wealthier areas. Shaded areas indicate non-residential areas. Thin gray lines indicate subzones while thicker gray lines indicate the coarser census area borders.

percent higher (standard error of 0.011, p-value < 0.01). The estimate of -0.008 for $W \rightarrow P$, on the other hand, is an order of magnitude lower and is not significant at conventional levels (with a standard error of 0.011).

Overall, what the estimates suggest is that the effect of mismatch by wealth depends on who travels out, with poor to wealthy and wealthy to poor having not just different magnitudes but different signs. Moreover, one direction of social interaction—poor to wealthy—drives all of our observed on mismatch by wealth. The results here suggest that the utility gain in neighborhood visits from a mismatch in wealth is asymmetric. Individuals from poorer neighborhoods have more to gain in visiting wealthier neighborhoods than vice versa. In terms of real-time (as opposed to residential) exposure of individuals from different socio-economic characteristics, the above suggests that the poor are exposed more to the wealthy than vice versa, a finding similar to Dong et al. (2020) who use shopping transaction records.

A second form of asymmetry we examine is in whether the destination neigh-

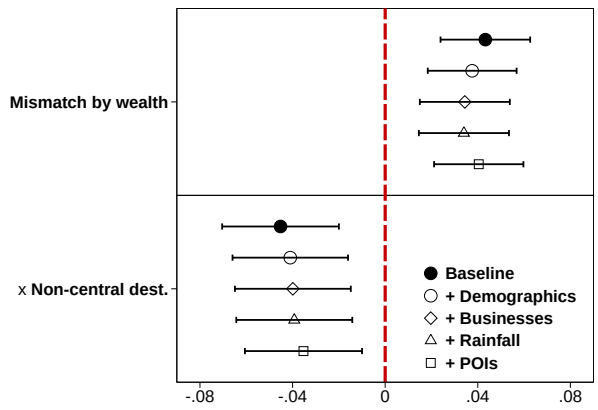

Fig. 6. Asymmetry by geography with mismatch in wealth. Top panel is mismatch, and bottom panel is the mismatch measure interacted with an indicator for destination neighborhoods outside the central region. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Dependent variable is log inflow from origin o to destination d. Figure corresponds to Table S4. Capped horizontal lines are 90% confidence intervals from standard errors clustered at the origin-by-destination area level.

borhood is in the central region, where the central business district lays.[13] Fig. 6 reports the results from estimating Eq. (3) with an additional interaction term for whether the destination neighborhood is outside of the central region. The top panel reports the coefficients for the mismatch by wealth measure, which is interpreted as the effect of mismatch by wealth on visits to central region neighborhoods relative to other neighborhoods without mismatch. The coefficient of 0.04 implies that mismatch by wealth increases neighborhood visits by 4 percent when the destination is in the central region (standard error of 0.012 and p-value < 0.01).

For those non-central destination neighborhoods, the effect reverses. The estimate of -0.035 implies that the additional effect of mismatch by wealth on neighborhood visits is approximately -3.5 percent (standard error of 0.015 and p-value $<$ 0.05). On net, mismatch by wealth still confers some positive effect on neighborhood visit (confirmed by an F-test reported in Table S4 that fails to reject the null that the joint additive effect is zero). Nonetheless, the above results suggest that whether the neighborhood is in a favorable central location matters, and that the real-time expo-

[13] The central region, and more specifically the central business district, includes some mixed-use urban neighborhoods as part of a continual goal in urban planning, which entails rental incentives. This means that neighborhoods in the central region, even if residential, are more likely than non-central regions to have commercial buildings, hotels, museums, and concert halls, among others.

sure of individuals from poorer neighborhoods to wealthier neighborhoods is driven by visits to neighborhoods in the central region. Moreover, since neighborhoods in the central region is disproportionally wealthier, this result is consistent with the previous finding that inflows from mismatch by wealth comes from individuals going from the poor to wealthy neighborhoods.[14]

Results: Mismatch by Ethnicity

Our final set of results focuses on how mismatch by ethnicity affects neighborhood visits. Here we let z from the mismatch measure defined in Eq. (2) be the proportion of the minority ethnic residents in a neighborhood. The relaxation of the z to a continuous measure of proportion implies that the mismatch measure can be interpreted as the unconditional probability that, for two randomly drawn individuals i and j from neighborhood o and d, either i is from a minority ethnic, or j is minority ethnic, but not both. The larger this measure is, the larger the probability of mismatch by ethnicity between the two neighborhoods (see Fig. S25).

Fig. 7A reports the coefficients, and here we observe a negative effect. The estimate of -0.21 implies that a one standard deviation increase in the mismatch by ethnic measure decreases neighborhood visits by 1.5 percent ($\approx 100 \times 0.21 \times 0.07$, standard error of 0.102 and p-value < 0.05). On a more extreme scale, taking into account ethnic clusters in neighborhoods (Leong et al. 2020) with neighborhood pairs that are very different in terms of the mismatch measure, going from the median neighborhood pair to the extreme neighborhood pair in the sample reduces neighborhood visits by 6.9 percent ($\approx 100 \times .21 \times [0.7 - 0.37]$). This result is consistent with studies that find homophilic preferences in residence location (Büchel et al. 2020;

[14] Using neighborhood-by-day fixed effects yields very similar conclusions (Table S9).

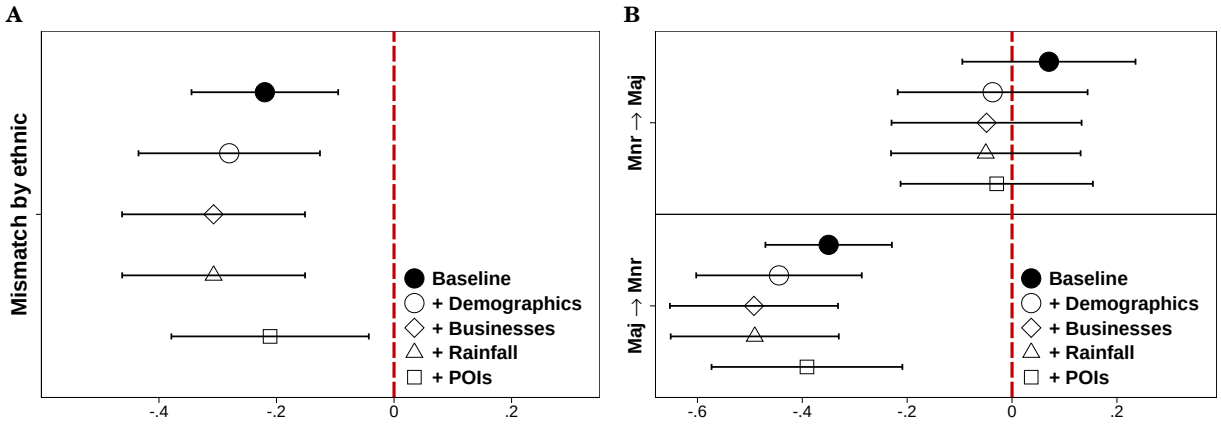

Fig. 7. Mismatch in ethnicity and movement. (**A**) Coefficients are for the mismatch measure from Eq. (2), with z_i as the proportion of minority-ethnic residents in neighborhood i. (**B**) Asymmetry by direction. Estimated coefficients of the two terms in Eq. (2): $z_o(1 - z_d)$ and $z_d(1 - z_o)$ indicating separately: movement from minority-ethnic neighborhood to majority-ethnic neighborhood and movement from majority-ethnic neighborhood to minority-ethnic neighborhood. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Dependent variable is log of inflow from origin o to destination d. Figure corresponds to Table S5 and Table S6. Capped horizontal lines are 90% confidence intervals from standard errors clustered at the origin-by-destination area level.

Wong 2013), and in this study, we detect ethnic preferences in daily neighborhood visits.[15]

Fig. 7B provides evidence of asymmetry in the mismatch in ethnicity effect. We once again decompose the mismatch measure into the two components: minority-ethnic to majority-ethnic neighborhoods and majority-ethnic to minority-ethnic neighborhoods.[16] The estimate of -0.39 implies neighborhood visits increases by 2.9 percent

[15] Again, like in the Results: Mismatch by Wealth section, our finding does not change if we swap out the census planning area-by-day fixed effects for the neighborhood-by-day fixed effects (Table S10 in the supplementary appendix). The estimate becomes larger and more significant—the estimate of -1.069 (as opposed to 0.21) with a standard error of 0.256 and p-value < 0.01. This estimate implies that a one standard deviation increase in the mismatch by ethnic measure decreases neighborhood visits by 7.5 percent ($\approx 100 \times -0.069 \times 0.07$).

In addition, using the constant elasticities specification implied in Eq. (1) leads to similar conclusion—the estimate of -0.067 (standard error of 0.033) for the log transformed mismatch by ethnic variable implies that a 10 percent increase in mismatch in ethnic decreases neighborhood visits by 0.7 percent (p-value < 0.05, Table S11).

[16] 'Minority' and 'majority' ethnicity is as defined at the national (city) level. A neighborhood can have a high proportion of minority-ethnic residents relative to the city-wide average.

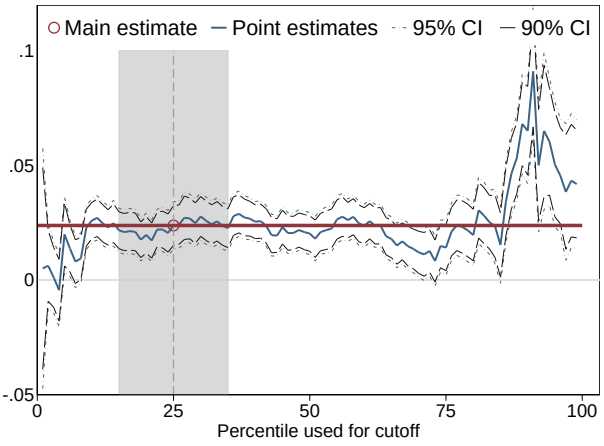

Fig. 8. Sensitivity for mismatch by wealth results. Sensitivity test for results in Fig. 4A and Table S2 where mismatch by wealth leads to more neighborhood visits to the percentile cutoff to define z in Eq. (2). Shaded gray areas indicates the 15th to the 35th percentiles for a $\pm$ 10 range around the 25th percentile. The specification is the one with the full controls. Fig. S3 reports the results after sorting by estimate size.

($\approx 100 \times -0.391 \times -0.074$, with a standard error of 0.11 and p-value < 0.01) when coming from a majority-ethnic to a minority-ethnic neighborhood. On the contrary, the estimate of -0.029 with a standard error of 0.12 implies that the direction of minority-ethnic to majority-ethnic neighborhood has no estimated effect on neighborhood visits. In the discussion section below, we discuss what our estimates imply about the ethnic quote in public housing residence in the city.

Sensitivity to Cutoffs

Mismatch by wealth above is based on an indicator variable defined by whether a neighborhood's wealth proxy falls below ($z = 1$) or above ($z = 0$) the 25th percentile. We fully test how sensitivity our mismatch by wealth results are to this threshold by repeating the analyses and reporting the coefficients of interest using the full range of possible cutoffs.

Fig. 8 reports the sensitivity of the estimated coefficients from Fig. 4A, where mismatch by wealth increases neighborhood visits. We report the full range of possible cutoffs (from the 1st to 99th percentile). First, the estimates are stable in the immediate range of the 25th percentile cutoff (indicated by the gray shaded

area and the dotted vertical line). This suggests that our results from Fig. 4 are not artifacts of the 25th percentile cutoff. Second, the estimates are precise, as indicated by both the 90% and 95% confidence intervals, with precision blowing up mostly in the tail ends of the distribution since sample imbalance in the poor vs. wealthy neighborhoods are exacerbated. Both the estimates and standard errors are erratic only at extreme values.

Fig. 9 reports the sensitivity of the coefficients from Fig. 4B, where the observed increased flow from mismatch by wealth is driven unidirectionally by individuals from poorer neighborhoods visiting the wealthier neighborhoods. Again, the estimates are insensitive to the choice of the cutoff, especially in the immediate range of the 25th percentile. In particular, the coefficients for the ($W \rightarrow P$) are positive and significant only erratically at certain cutoffs.

Moreover, we interpret Fig. 9 as suggesting that the asymmetry barely reverses (given the magnitude) and only reverses when we are overwhelmingly inclusive in the definition of 'poor' in the wealth distribution. Figure 9A also suggests that the preference for visiting higher-scale neighborhoods persists across the wealth spectrum and is in no way unique to those in the 25th percentile. This finding relates to those of Hilman et al. (2021) where people prefer to visit neighborhoods similar to their own, and if they visit other neighborhoods, it will likely be of a higher class.

Fig. 10 reports the sensitivity of the estimated coefficients from Fig. 6, which suggests a differential effect of mismatch by wealth on neighborhood visits when the destination neighborhood is outside the central region. Here, the finding that mismatch by wealth increasing neighborhood visits reverses for non-central destinations is still stable around the 25th percentile cutoff. However, as the cutoff approaches and increases from the median, the estimate reverses sign and becomes

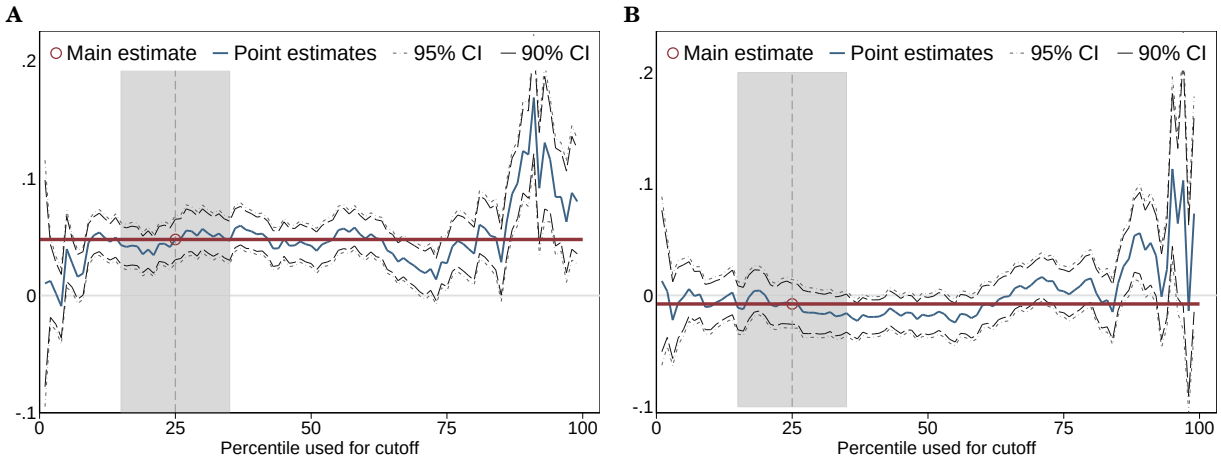

Fig. 9. Sensitivity tests for wealth mismatch by direction. **(A)** Sensitivity of the ($P \rightarrow W$) coefficient from the top panel of Fig. 4B. **(B)** Sensitivity of the ($W \rightarrow P$) coefficient from the bottom panel of Fig. 4B. Shaded gray areas indicate the 15th to the 35th percentiles for a $\pm$ 10 range around the 25th percentile. The specification is the one with the full controls. Figure corresponds also to Table S3. Fig. S4 reports the results after sorting by estimate size.

positive. Nonetheless, our results are still stable over a considerable range (shaded gray area) around the main 25th percentile cutoff.[17] [18]

Application to Ethnic Housing Quota

Context of Urban and Desegregation Policies. We build on our estimates and context of study to perform a simple counterfactual case centered on a key residential integration policy in Singapore. Our analysis serves two purposes: i) to quantify the

[17] In the supplementary appendix, we show that when ordering the estimates from the sensitivity tests by magnitude instead of threshold, the estimates from our chosen threshold always appear near the middle of the curve. This pattern should mitigate concerns that we have chosen our threshold to inflate the estimated effect sizes.

[18] We also do a form of placebo exercise where we allow the effect of mismatch by ethnic to have different effects for central and non-central region destination neighborhoods. This exercise is similar to the one in Fig. 6, except that ethnic levels should no longer be link to geographically location (as opposed to the pre-modern ethnic enclaves we discuss). Hence, we should not detect a differential effect for non-central neighborhoods and our results confirm this (Table S12 and Table S13 in the supplementary appendix). In fact, the differential effect of mismatch by ethnicity for non-central neighborhoods is a precisely estimated zero—with an estimate of 0.0008 and a standard error of 0.18.

In addition, we include both mismatch by wealth and mismatch by ethnic (Table S14 and Table S15) and show that the estimates do not change substantially—they have the same signs, magnitude, and level of statistical significance.

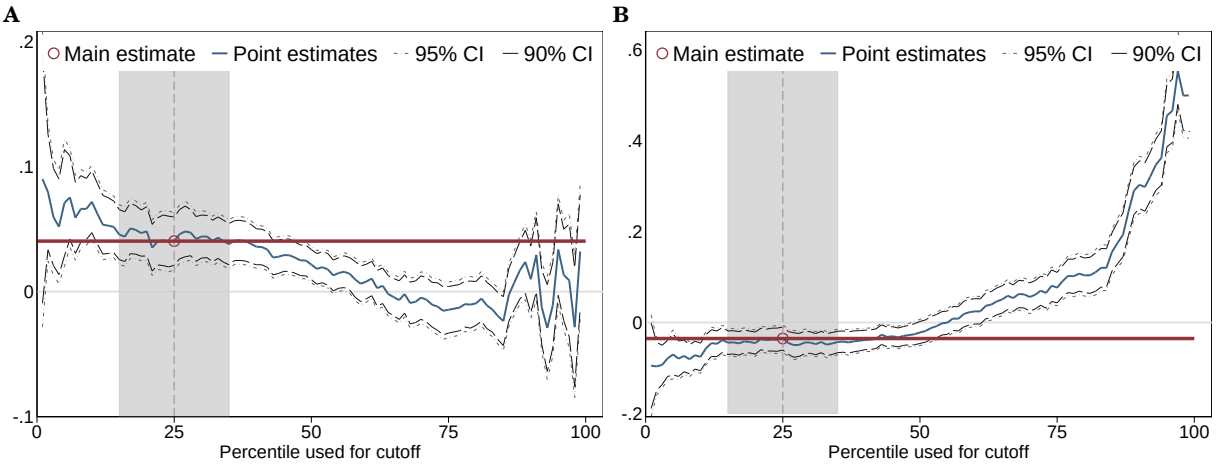

Fig. 10. Sensitivity tests for wealth mismatch by geography. **(A)** Sensitivity of the mismatch by wealth coefficient in the top panel of Fig. 6. **(B)** Sensitivity of the mismatch by wealth interacted with non-central destination coefficient in the bottom panel of Fig. 6. Shaded gray areas indicate the 15th to the 35th percentiles for a $\pm$ 10 range around the 25th percentile. The specification is the one with the full controls. Figure corresponds also to Table S4. Fig. S5 reports the results after sorting by estimate size.

value of residential ethnic quotas and ii) to demonstrate the value of quantifying asymmetric segregation.

Before the establishment of the government-run Housing and Development Board in 1960, communal enclaves were formed by immigrant ethnic groups settling and concentrating in different parts of the city under the previous colonial administration (Choe 2016). One historical administration was the Jackson Plan in the early 1800s, which segregated the city into ethnic subdivisions (Koh et al. 2006).

The Ethnic Integration Policy (EIP) was introduced in 1989 to prevent further formation of enclaves and to ensure a balanced mix of ethnic groups (Chinese 76%, Indian 7.5%, Malay 15%, Others 1.5%) in public housing which houses more than 80 percent of the population (Leong et al. 2020; gov.sg 2020). The EIP is implemented by hierarchical ethnic quotas which limits the percentage of residents of a certain ethnicity by block and by neighborhood so that the neighborhood ethnicity mix is close to the city-wide ethnicity mix. No further sale of flats to an ethnic group is

Table 2. Pre-EIP ethnic composition for selected towns.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1988 (Predicted)							
Town	Chinese	Indian/Others	Malays		Majority	Minority	Mismatch
Bukit Merah	88.4	6.1	5.5		88.4	11.6	0.35
Bedok	69.8	5.7	24.5		69.8	30.2	.
2000 (Predicted)							
Town	Chinese	Indian/Others	Malays		Majority	Minority	Mismatch
Bukit Merah	93.1	5.6	1.3		93.1	6.9	0.48
Bedok	52.0	5.0	43.0		52.0	48.0	.
2000 (Actual/Official statistics)							
Town	Chinese	Indian	Malays	Others	Majority	Minority	Mismatch
Bukit Merah	83.8	9.5	5.9	0.8	83.8	16.2	0.34
Bedok	73.1	7.2	17.4	2.4	73.1	26.9	.
2015 (Actual/Official statistics)							
Town	Chinese	Indian	Malays	Others	Majority	Minority	Mismatch
Bukit Merah	78.7	9.7	8.6	3.0	78.7	21.3	0.37
Bedok	72.1	8.7	15.2	4.1	72.1	27.9	.

EIP is the ethnic integration policy implemented since 1989. Predicted values of ethnic composition for 1988 and 2000, which comes directly from Sin 2002b who sourced it from Ooi 1993, are estimates of ethnic composition in the absence of the EIP. 2015 values from official statistics also used for the main analyses. Mismatch is ethnic mismatch as defined in Eq. (2) with z as the proportion of ethnic minorities. Towns are administrative boundaries that are larger than neighborhoods; they correspond more closely to the census areas. Systematic pre-EIP ethnic records at the neighborhood level is not available to us.

allowed once its ethnic limit is reached. The exception is when the buyer and seller are of that same ethnic group.[19] [20]

Evaluating the ethnic integration policy (EIP). To show what experienced segregation by ethnic mismatch would have been, we use the projected ethnic compositions from historical trends that predate the EIP (Ooi 1993; Sin 2002b). Table 2 shows four sets of values: predicted ethnic composition in 1988 (a year before the EIP) based on housing applications, predicted ethnic composition in 2000, and actual ethnic composition in 2000 and 2015. We focus on Bukit Merah and Bedok (the two extant

[19] Hence, in a block that is 'over-occupied' by ethnic X, while ethnic X owners can still sell their home to ethnic X buyers, non-ethnic X owners cannot sell to a ethnic X so as not to worsen the ethnic imbalance.

[20] Owing to the difference in the distribution of household income among different ethnic groups (Fig. S22), the EIP also indirectly integrates residents of varying income groups.

towns from Sin 2002b), two towns in the central and east region with legacies as ethnic enclaves. Comparing the 1988 to 2000 values makes it clear that Bukit Merah would have become a Chinese enclave while Bedok would have become a Malay enclave. The EIP, with a binding ethnic quota, prevents this as evident from the 2000 values (12 years after the start of EIP).

The difference in ethnic mismatch from the 2000 predicted to actual ethnic composition implies that visits to and from these two towns on a daily basis is on average 2.9 percent ($\approx 100 \times -0.21 \times [0.34 - 0.48]$) higher than it would have been without the EIP. To contextualize this figure, a one standard deviation decrease in ethnic mismatch today increases visits by 1.5 percent ($\approx 100 \times -0.21 \times 0.069$). On the other hand, a mismatch in wealth on average increases neighborhood visits by 2.1 percent (as discussed above).[21]

On the other hand, the change in the majority-to-minority ethnic mismatch with and without the EIP implies that visits are 8.7 percent ($\approx 100 \times -0.391 \times [0.269(1 - 0.162) - 0.48(1 - 0.069)]$) higher than it would have been without the EIP. For comparison, a one standard deviation decrease in today's majority-to-minority ethnic mismatch implies only a 2.9 percent increase in visits ($\approx 100 \times -0.391 \times 0.074$).[22]

Fig. 11 systematically contextualizes how effective the EIP is, we compute all values of increase in visits implied by a percentile reduction in modern mismatch (our 2020 sample). Each point on the curve indicates percentage increase in neighborhood visits implied by the corresponding decrease in ethnic mismatch by the value of the percentile. For instance, the middle 50th percentile value for mismatch by ethnic is

[21] Using the alternative (and larger) estimate from the specification with neighborhood-by-day fixed effects (Table S10) implies that visits are 14.9 percent ($\approx 100 \times -1.069 \times [0.34 - 0.48]$) than it would have been.

[22] Our counterfactual analyses nominally require assumptions about similar distributions in amenities between the two towns for our sample period and in the past. However, since our focus is on the difference with and without accounting for asymmetry, such concerns loom less large.

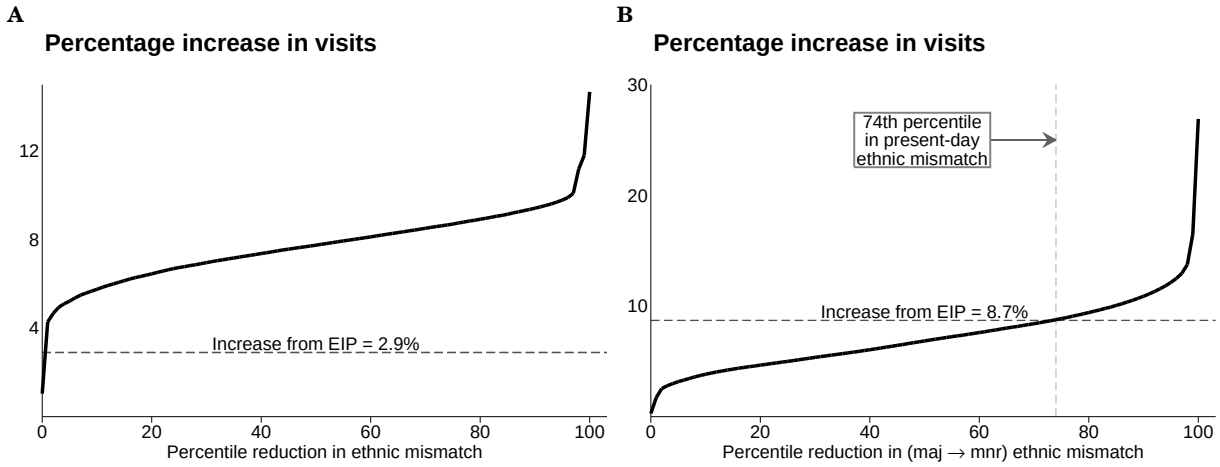

Fig. 11. Reduction in ethnic mismatch and increase in visits. **(A)** Without accounting for asymmetry: Percentile reduction in ethnic mismatch and the percentage increase in visits implied by the estimate of -0.21 from Fig. 7A. **(B)** Accounting for asymmetry: Percentile reduction in (maj $\rightarrow$ mnr) ethnic mismatch and the increase in visits implied by the estimate of 0.391 from Fig. 7B. Each point implies what the increase in visits would be if mismatch was reduced to zero. Each point is computed as follows: we take the value of ethnic mismatch implied by each centile derived from the 75,390 neighborhood-pairs in our sample and compute ($100 \times -.21 \times$ value) for panel A and ($100 \times -.391 \times$ value) for panel B. The dashed horizontal line indicates the percentage increase in visits implied from our counterfactual case study of the residential ethnic integration policy.

0.37 and going from a value of 0.37 to 0 in ethnic mismatch implies an increase in visits of 7.7 percent ($\approx 100 \times -0.21 \times 0.37$).

The increase in visits implied by the reduction in ethnic mismatch from the EIP is no larger than a 1st percentile reduction in today's ethnic mismatch. Cognizant of the importance in asymmetry, we further examine the value of EIP specifically in terms of majority-to-minority ethnic mismatch. The implied value of EIP is now more substantial. We would need a reduction from the 74th percentile value to zero in modern-day majority-to-minority mismatch to have the increase in social interaction implied by the EIP. Even without taking a position on the EIP, we have evidence that the value of the EIP is much higher ($\sim$factor of 3) than it would have otherwise been had we not accounted for asymmetries in segregated interaction.[23]

[23] Comparing the 2000 predicted values to the 2015 actual values shows an even larger difference in efficacy measure ($\sim$factor of 4).

Discussion and Conclusion

This study uses a new dataset constructed from location-based GPS data to quantify how neighborhood visits are affected by mismatch in neighborhood wealth and ethnicity. The context of our study is the city of Singapore, which is densely urban, has high public transit usage, and has a key binding ethnic quota to ensure neighborhoods have a similar ethnic mix to the city average. We find that mismatch by wealth increases neighborhood flows, but is driven asymmetrically from individuals from poor neighborhoods visiting the wealthier neighborhoods. Mismatch by ethnicity decreases neighborhood visits, and this is driven asymmetrically by lower neighborhood visits to minority-ethnic neighborhoods from individuals in majority-ethnic neighborhoods.

Our finding is of relevance since we know that social exposure can have profound consequences along both social and economic dimensions (Ananat 2011; Cook et al. 2018; Cutler and Glaeser 1997; Cutler et al. 2008) and create schisms in everyday social life (Low 2018; Teo 2018).

The asymmetries where the poor are more likely to visit the wealthy and where the majority ethnic are less likely to visit the minority ethnics is consistent with urban inequality. Certain neighborhoods may end up with disproportionally large amenities, including public transit accessibility, because these areas attract more individuals in a way that disadvantage the poorer and minority-ethnic neighborhoods. This implies urban planners may need to consider asymmetries in segregated interaction when planning the distribution of public goods.

We build on our context of study, Singapore, by applying our estimates on asymmetry to a counterfactual case study of a binding ethnic residential quota for a pair

of towns with legacies as ethnic enclaves. Our estimates on mismatch by ethnicity allow us to estimate the increase in neighborhood visits and, therefore, a reduction in daily segregated interactions resulting from the ethnic quota. Overall, we find that the ethnic quota's effect on increasing visits would have been severely understated had we not accounted for the asymmetry in ethnic mismatch.

While we leverage GPS data to overcome the limits of studying segregation only via residence, this approach is not without its limitations. One important limitation is that we are not able to disentangle individuals' purpose of visit. All we observe is the daily presence of individuals across neighborhoods. Nonetheless, even if our neighborhood visitation patterns include trip chains within longer commuting-related trips (e.g., Miyauchi et al. 2021), trip hops still constitute an important form of interaction where individuals share physical activity space as they go about their daily lives (Athey et al. 2020; Cagney et al. 2020).

In addition, what we use is a form of geolocation data (see Cagney et al. 2020), which means we observe the physical occupation of space but not the actual social interactions. One may share a space with another individual but not actually socially interact and share experiences. Nonetheless, the sharing of physical space has implications for segregation since integrated physical spaces foster interactions and expose one to individuals from different backgrounds (see Sunstein 2018; Athey et al. 2020). Using GPS data also implies that we do not have full coverage of the population, although our analyses in the supplementary appendix show that the GPS data correlates meaningfully with key neighborhood demographics and characteristics. Finally, the anonymized data necessitates our use of proxies in computing where people live and the characteristics of their neighborhood.

In spite of the limitations, we are optimistic that mobile phone data, and more

broadly geolocation data, have utility in helping us understand social and economic problems, as demonstrated in this study. In addition to studies on urban inequality and segregated interactions, we believe the granularity offered by GPS data can reveal insights into spatial-temporal behavior and the use of amenities in a way that is not possible with more traditional data sources.

References

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. 2020. "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic." *Journal of Public Economics* 191 104254. https://doi.org/10.1016/j.jpubeco.2020.104254. 7

Ananat, Elizabeth Oltmans. 2011. "The wrong side(s) Of the tracks: The causal effects of racial segregation on urban poverty and inequality." *American Economic Journal: Applied Economics* 3 (2): 34–66, http://www.jstor.org/stable/41288628. 30

Athey, Susan, Billy A Ferguson, Matthew Gentzkow, and Tobias Schmidt. 2020. "Experienced segregation." 10.3386/w27572. 6, 7, 31

Atkin, David, Keith Chen, and Anton Popov. 2022. "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley." 10.3386/w30147. 6

Banerjee, Mitali, and Paul Ingram. 2018. "Fame as an illusion of creativity: Evidence from the pioneers of abstract art." https://dx.doi.org/10.2139/ssrn.3258318. 6

Banerji, Devika, and Torsten Reimer. 2019. "Startup founders and their LinkedIn connections: Are well-connected entrepreneurs more successful?" *Computers in Human Behavior* 90 46–52. https://doi.org/10.1016/j.chb.2018.08.033. 6

Bastos, Marco, Dan Mercea, and Andrea Baronchelli. 2018. "The geographic embedding of online echo chambers: Evidence from the Brexit campaign." *PLoS ONE* 13 (11): 1–16. 10.1371/journal.pone.0206841. 6

Blattman, Christopher, Donald P Green, Daniel Ortega, and Santiago Tobón. 2021. "Place-based interventions at scale: The direct and spillover effects of policing and city services on crime." *Journal of the European Economic Association* 19. 10.1093/jeea/jvab002. 7

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting poverty and wealth from mobile phone metadata." *Science* 350 (6264): 1073–1076. https://doi.org/10.1126/science.aac4420. 7

Büchel, Konstantin, Maximilian V Ehrlich, Diego Puga, and Elisabet Viladecans-Marsal. 2020. "Calling from the outside: The role of networks in residential mobility." *Journal of Urban Economics* 119 103277. https://doi.org/10.1016/j.jue.2020.103277. 21

Cagney, Kathleen A., Erin York Cornwell, Alyssa W. Goldman, and Liang Cai. 2020. "Urban Mobility and Activity Space." *Annual Review of Sociology* 46 (1): 623–648. 10.1146/annurev-soc-121919-054848. 31

Chay, Kenneth Y, Jonathan Guryan, and Bhashkar Mazumder. 2014. "Early life environment and racial inequality in education and earnings in the United States." Working Paper 20539, National Bureau of Economic Research. 10.3386/w20539. 6

Chen, M Keith, and Ryne Rohla. 2018. "The effect of partisanship and political advertising on close family ties." *Science* 360 (6392): 1020–1024. https://doi.org/10.1126/science.aaq1433. 7

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment." *American Economic Review* 106 (4): 855–902. 10.1257/aer.20150572. 6

Choe, Alan F. C. 2016. *The early years of nation-building: Reflections on Singapore's urban history*. Chap. CHAPTER 1 3–21, World Scientific, . 10.1142/9789814656474_0001. 6, 26

Chua, Alvin, Serene Ow, Kevin Hsu, Wang Yazhe, Michael Chirico, and Huang Zhongwen. 2020. "Distilling actionable insights from big travel demand datasets for city planning." *Research in Transportation Economics* 83 (December 2019): 100850. 10.1016/j.retrec.2020.100850. 7

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The echo chamber effect on social media." *Proceedings of the National Academy of Sciences* 118 (9): e2023301118. 10.1073/pnas.2023301118. 6

Cook, Lisa D, Trevon D Logan, and John M Parman. 2018. "Racial segregation and southern lynching." *Social Science History* 42 (4): 635–675. https://doi.org/10.1017/ssh.2018.21. 6, 30

Cutler, David M, and Edward L Glaeser. 1997. "Are ghettos good or bad?" *The Quarterly Journal of Economics* 112 (3): 827–872. https://doi.org/10.1162/003355397555361. 30

Cutler, David M, Edward L Glaeser, and Jacob L Vigdor. 2008. "When are ghettos bad? Lessons from immigrant segregation in the United States." *Journal of Urban Economics* 63 (3): 759–774. https://doi.org/10.1016/j.jue.2007.08.003. 30

Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales. 2019. "How segregated is urban consumption?" *Journal of Political Economy* 127 (4): 1684–1738. https://doi.org/10.1086/701680. 6, 7, 14

Department of Statistics, Singapore. 2021. "Census of population 2020." https://www.singstat.gov.sg/find-data/search-by-theme/households/household-income/publications-and-methodology. 75

Dong, Xiaowen, Alfredo J Morales, Eaman Jahani et al. 2020. "Segregated interactions in urban and online space." *EPJ Data Science* 9 (1): 20. 10.1140/epjds/s13688-020-00238-7. 2, 5, 6, 7, 19

Dong, Xiaowen, Yoshihiko Suhara, Burçin Bozkaya, Vivek K. Singh, Bruno Lepri, and Alex Pentland. 2017. "Social bridges in urban purchase behavior." *ACM Transactions on Intelligent Systems and Technology* 9 (3): . 10.1145/3149409. 16

Eytan, Bakshy, Messing Solomon, and Adamic Lada A. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348 (6239): 1130–1132. 10.1126/science.aaa1160. 6

Gentzkow, Matthew, and Jesse M Shapiro. 2011. "Ideological segregation online and offline." *The Quarterly Journal of Economics* 126 (4): 1799–1839. 10.1093/qje/qjr044. 6

Glaeser, Edward, and Jacob Vigdor. 2012. "The end of the segregated century: Racial separation in America's neighborhoods, 1890-2012." *Civic Report* 66 1–36, https://media4.manhattan-institute.org/pdf/cr_66.pdf. 6

Gollwitzer, Anton, Cameron Martel, William J Brady, Philip Pärnamets, Isaac G Freedman, Eric D Knowles, and Jay J Van Bavel. 2020. "Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic." *Nature human behaviour* 4 (11): 1186–1197. https://doi.org/10.1038/s41562-020-00977-7. 7

gov.sg. 2020. "HDB's Ethnic Integration Policy: Why it still matters." https://www.gov.sg/article/hdbs-ethnic-integration-policy-why-it-still-matters. 26

Hensvik, Lena, and Oskar Nordström Skans. 2016. "Social networks, employee selection, and labor market outcomes." *Journal of Labor Economics* 34 (4): 825–867. 10.1086/686253. 6

Hilman, Rafiazka Millanida, Gerardo Iñiguez, and Márton Karsai. 2021. "Socioeconomic biases in urban mixing patterns of US metropolitan areas." https://doi.org/10.48550/arXiv.2110.04183. 5, 6, 24

Hutchens, Robert. 2001. "Numerical measures of segregation: desirable properties and their implications." *Mathematical social sciences* 42 (1): 13–29. https://doi.org/10.1016/S0165-4896(00)00070-6. 2

Jones, Malia, and Anne R Pebley. 2014. "Redefining neighborhoods using common destinations: social characteristics of activity spaces and home census tracts compared." *Demography* 51 (3): 727–752. 10.1007/s13524-014-0283-z. 2, 7

Koh, Tommy, Timothy Auger, Jimmy Yap, and Ng Wei Chian. 2006. *Singapore: The Encyclopedia*. Singapore: Editions Didier Millet. 26

Kreindler, Gabriel E, and Yuhei Miyauchi. Forthcoming. "Measuring commuting and economic activity inside cities with cell phone records." *Review of Economics and Statistics* to appear. https://doi.org/10.1162/rest_a_01085. 7, 13

Krivo, Lauren J, Heather M Washington, Ruth D Peterson, Christopher R Browning, Catherine A Calder, and Mei-Po Kwan. 2013. "Social isolation of disadvantage and advantage: The reproduction of inequality in urban space." *Social Forces* 92 (1): 141–164. 10.1093/sf/sot043. 2, 7

Le Roux, Guillaume, Julie Vallée, and Hadrien Commenges. 2017. "Social segregation around the clock in the Paris region (France)." *Journal of Transport Geography* 59 134–145. https://doi.org/10.1016/j.jtrangeo.2017.02.003. 2, 7

Lee, Shu En, Jing Zhi Lim, and Lucas Shen. 2021. "Segregation Across Neighborhoods in a Small City." *Asia Competitiveness Institute Research Paper Series* Research Paper #07-2021, https://lkyspp.nus.edu.sg/docs/default-source/aci/acirp202107.pdf?sfvrsn=a862240a_2. 79

Leong, Chan-Hoong, Eugene Teng, and William Weiliang Ko. 2020. "The state of ethnic congregation in Singapore today." In *Building Resilient Neighbourhoods in Singapore*, edited by Leong, Chan-Hoong, and Lai-Choo Malone-Lee 29–50, Springer Nature Singapore Pte Ltd, . 10.1007/978-981-13-7048-9. 6, 21, 26

Levy, Gilat, and Ronny Razin. 2019. "Echo chambers and their effects on economic and political outcomes." *Annual Review of Economics* 11 (1): 303–328. 10.1146/annurev-economics-080218-030343. 6

Loo, Lee Sim, Ming Yu Shi, and Sheng Han Sun. 2003. "Public housing and ethnic integration in Singapore." *Habitat International* 27 (2): 293–307. https://doi.org/10.1016/S0197-3975(02)00050-4. 6

Low, Minmin. 2018. "Class - not race nor religion - is potentially Singapore's most divisive fault line." https://www.channelnewsasia.com/news/cnainsider/

regardless-class-race-religion-survey-singapore-income-divide-10774682. 30

Luo, Feixiong, Guofeng Cao, Kevin Mulligan, and Xiang Li. 2016. "Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago." *Applied Geography* 70 11–25. https://doi.org/10.1016/j.apgeog.2016.03.001. 7

Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J Redding. 2021. "Consumption access and the spatial concentration of economic activity: Evidence from smartphone data." *National Bureau of Economic Research Working Paper Series* No. 28497. 10.3386/w28497. 7, 13, 14, 16, 31

Moro, Esteban, Dan Calacci, Xiaowen Dong, and Alex Pentland. 2021. "Mobility patterns are associated with experienced income segregation in large US cities." *Nature Communications* 12 (1): . 10.1038/s41467-021-24899-8. 6, 7

Ooi, G L. 1993. "The Housing and Development Board's ethnic integration policy." In *The Management of Ethnic Relations in Public Housing Estates*, edited by Ooi, G L, S Siddique, and K C Soh 4–24, Times Academic Press. 27

Palmer, John R B, Thomas J Espenshade, Frederic Bartumeus, Chang Y Chung, Necati Ercan Ozgencil, and Kathleen Li. 2013. "New approaches to human mobility: Using mobile phones for demographic research." *Demography* 50 (3): 1105–1128. 10.1007/s13524-012-0175-z. 2

Rao, Gautam. 2019. "Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools." *American Economic Review* 109 (3): 774–809. 10.1257/aer.20180044. 2, 6

Rodriguez-Moral, Antonio, and Marc Vorsatz. 2016. *An overview of the measurement of segregation: Classical approaches and social network analysis*. 93–119, Cham: Springer International Publishing, . 10.1007/978-3-319-40803-3_5. 2

Sin, Chih Hoong. 2002a. "Segregation and marginalisation within public housing: The disadvantaged in Bedok New Town, Singapore." *Housing Studies* 17 (2): 267–288. https://doi.org/10.1080/02673030220123225. 2, 6

Sin, Chih Hoong. 2002b. "The Quest for a Balanced Ethnic Mix: Singapore's Ethnic Quota Policy Examined." *Urban Studies* 39 1347–1374. 10.1080/00420980220142673. 27, 28

Sunstein, Cass R. 2018. *#Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, , ned - new edition edition, http://www.jstor.org/stable/j.ctv8xnhtd. 31

Teo, You Yenn. 2018. *This is what inequality looks like*. Singapore: Ethos Books. 6, 30

Wang, Donggen, and Fei Li. 2016. "Daily activity space and exposure: A comparative study of Hong Kong's public and private housing residents' segregation in daily life." *Cities* 59 148–155. https://doi.org/10.1016/j.cities.2015.09.010. 7

Wang, Qi, Nolan Edward Phillips, Mario L Small, and Robert J Sampson. 2018. "Urban mobility and neighborhood isolation in America's 50 largest cities." *Proceedings of the National Academy of Sciences* 115 (30): 7735 LP – 7740. 10.1073/pnas.1802537115. 6

Wong, Maisy. 2013. "Estimating ethnic preferences using ethnic housing quotas in Singapore." *The Review of Economic Studies* 80 (3): 1178–1214. 10.1093/restud/rdt002. 6, 22

Xu, Yang, Alexander Belyi, Iva Bojic, and Carlo Ratti. 2018. "Human mobility and socioeconomic status: Analysis of Singapore and Boston." *Computers, Environment and Urban Systems* 72 51–67. 10.1016/j.compenvurbsys.2018.04.001. 11

Supplementary Materials

Table of Contents

A. Data Appendix

The primary source of data is CITYDATA.ai, which aggregates anonymized GPS (global positioning system) ping records as a third party via SDKs (Software Development Kits) within applications installed on mobile phones. All ping records are anonymized and include a device hash ID and a day record, where the underlying location signals have horizontal accuracy of up to 25m. See Table S1 for sources of mobile phone GPS ping data. We receive these records as flat files of `area-day` for the presence of devices in an area on a given day for the 91 days in Jan–Mar 2020. This period is before the two-months long city-wide lockdown on 7th April (announced three days prior). The areas are assigned according to the URA (Urban Redevelopment Authority) 2014 Master Plan map and devices are recorded. We

combine this GPS ping records to the official census records (where available) and to both public and private housing transaction records as proxy for neighborhood wealth levels.

1. To build the gravity panel by `origin-destination-date`, we start by using the device list records from CITYDATA.ai which are stored as flat files. For each census subzone and each day, we have a recorded list of captured devices in that subzone using the device hash. To get cross-area movement flows, we treat each device hash as an individual and aggregate the CITYDATA.ai records up to the `origin-destination-date` level. This gives for any `origin-destination-date` the count of individuals going from the `origin` area to the `destination` area on a given `date`. We then divide this inflow count by the number of devices for that `origin-date` to get the origin-to-destination flow measure.

2. To infer the ”origin” area of an individual, we aggregate the CITYDATA.ai records up to the `device-area`, and make the simple assumption that the area with the highest appearance count for a given device hash is the origin area. Devices that appear < 30 times in the 91-days sample period are dropped. Certain devices have ties in the rank and we treat them as different individuals. This yields records from approximately 125k devices.

   As a form of validation exercise, and to test the extent to how representative the GPS ping record and the constructed neighborhood flows data is of the population, we correlate the GPS data to neighborhood demographics in the supplementary appendix. We find that captured GPS pings are increasing with resident population size and do not vary too much by neighborhood-level characteristics such as age group, gender, ethnic composition, and house

type. We also find that geographical distance and rainfall affect neighborhood visits in an expected manner. More neighborhood visits is observed when the origin-destination pair in neighborhoods are adjacent and less when not. This observation extends to the centroid and edge-based distances between neighborhoods. More neighborhood visits is observed with moderate amount of precipatation, with the dryest and wettest periods having the least amount of neighborhood visits. For the 91-day sample period, we observe progressively fewer GPS pings in the weeks leading up to the lockdown. Captured GPS pings in the origin neighborhoods are decreasing with housing price, and this likely captures the fact that more expensive houses are private properties in areas that are less densely populated.

3. Census income data at the geographically smaller subzone level is unavailable to the public. To measure wealth/poverty at geographical units (census subzones) smaller than the available census income records (census/planning area), we use the public HDB housing micro-transaction records from the official and public repository (https://data.gov.sg/dataset/resale-flat-prices) and the proprietary private housing micro-transaction records are from REALIS maintained by the Urban Redevelopment Authority (URA).

   Transaction prices for public housing at the unit level are not available and only the aggregated storey range is provided to identify the unit. Further, the public housing transaction prices only have street addresses as geographical information. To crosswalk from the public transaction prices to subzones, we first query the OneMap database for postcodes using the street address. We then use the postcodes to obtain latitude and longitude coordinates from

which we can perform point-in-polygon analyses with the shape files to assign transaction prices to subzones. If a postal code maps to multiple coordinates, we will take the mean of the coordinates to get a representative coordinate for the postal code. Finally, we do a simple point-in-polygon query to see if the coordinate falls in a subzone to crosswalk from addresses to subzones.

Transaction prices for private housing is available at the unit level and postcodes are directly available in the REALIS data set. We crosswalk from the private transaction prices to subzone using postcodes, which is available in REALIS, in the same way above. As REALIS also contains data on land transactions, we removed it from our sample.

4. We retain all transaction records in the past three years (2017, 2018, 2019) before our sample year 2020, this yields about 133k public and private transactions, and then aggregate the housing price (in terms of price per square meters), separately for public and private transactions, up to the census subzone level.

   For each subzone, we compute neighborhood wealth as the average housing price (per square meters) weighted by the share of public residence and private residence:

$$\text{share}_i^{\text{public}}\text{PSM}_i^{\text{public}} + \text{share}_i^{\text{private}}\text{PSM}_i^{\text{private}}$$

   where share is the proportion (between 0 and 1) of residence in neighborhood i who live in the public HDB residences or private residences; PSM is the housing price per square meters. To validate the use of the weighted housing price per square meters measure, we aggregate the measure from the subzone level up to the census planning area level, and then correlate the aggregated wealth

proxy with the census income data available at 28 planning areas which yields a correlation coefficient of approximately 0.89. Fig. 3 shows the geographical distribution of the transactions and the census planning area level correlation with the census income records.

5. As an alternative measure of wealth/poverty at the smaller subzones using the census records, we use the the proportion of residents living in 1–2-room public housing flats—this is the count (rounded off to tens in the official records) of local residents living in 1–2-room HDB (Housing & Development Board) flats divided by the total number of local residents in the subzone. Records for "HUDC Flats (excluding those privatised)" and "Others" are excluded from both numerator and denominator. All residence type records are for the year 2019, available at https://storage.data.gov.sg/singapore-residents-by-subzone-and-type-of-dwelling-jun-2018/resources/planning-area-subzone-age-group-sex-and-type-of-dwelling-june-2011-2019-2020-03-06T03-39-39Z.csv. The results using residence type as a proxy for poverty and wealth is available in the supplementary materials.

6. Census variables for basic demographics at the subzone level are derived in the same manner. Demographic data available at the smaller subzone level are limited to age, gender, ethnicity, and population size. Age and gender records are from https://storage.data.gov.sg/resident-population-by-planning-area-subzone-age-group-and-sex-2015/resources/resident-population-by-planning-area-age-group-and-sex-2019-07-30T03-02-18Z.csv. Ethnicity records are from https://storage.data.gov.sg/resident-population-by-planning-area-subzone-ethnic-group-and-sex-2015/resources/resident-

population-by-planning-area-ethnic-group-and-sex-2019-08-01T03-23-57Z.csv. In our sample, we group age demographics at the smaller subzone level as: (i) below 20 years old, (ii) 20–39 years old, (iii) 40–64 years old, and (iv) 65 and above. For ethnicity demographics, four main ethnicities are recorded (Chinese, Indians, Malay, and Others with approximate composition of 76%, 7.5% 15%, 1.5%) where the first group is the majority ethnicity. In our sample and results, we group the neighborhood ethnicity into a single non-majority ethnic variable as the proportion of residence in a subzone that are from the non-majority ethnic to compute mismatch by ethnicity.

7. To measure poverty/wealth at the larger census planning area unit, we simply use the census records of both residential house types and (gross) income from work for resident working persons aged 15 years and above. This is available from https://storage.data.gov.sg/resident-working-persons-aged-15-years-over-by-planning-area-gross-monthly-income-from-work-2015/resources/resident-working-persons-aged-15-yrs-over-by-pa-gross-monthly-income-from-work-2019-08-08T04-36-54Z.csv. Income is a per-population weighted-average: for each income bin reported we take the midpoint and multiply it by the number of residents in that income bin, then we aggregate up to the area level and divide by the total number of working residents in that area.

8. To get distance measures, we use the 2014 masterplan data from https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-web. The geographic data is stored in a projected coordinate system - SVY21, which allows for a more accurate representation of the Singapore area. For each

subzone, we calculate the coordinates of its centroid. We then take the distance between all pairs of centroids to obtain the inter-subzone distance, `subzone-centroid-distance`. Additionally we obtain the edge-to-edge distance between all pairs of subzones. If the edge distance equals 0, the 2 subzones are considered to be contiguous.

9. For the area-specific POIs (places of interests), we mostly default to the official records found in https://data.gov.sg. For POIs records stored in a shapefiles or its equivalent, we simply do a point-in-polygon query to match POIs to areas. Number of POIs in a subzone includes: libraries, supermarkets, parks, preschools, schools (primary and secondary), silverzones, sport facilities, train stations, and tourist attractions.

   For businesses, the listing of corporate entities from the Accounting and Corporate Regulatory Authority come from https://storage.data.gov.sg/acra-information-on-corporate-entities/resources/acra-information-on-corporate-entities-a-2021-01-15T00-37-18Z.csv. From the entity status description of the official records we retain those that are "live", and then focus on three main industry divisions, based on the SSIC (Singapore Standard Industrial Classification), that account for a large portion of the employment force: construction (SSIC 41), manufacturing (SSIC 10), and services (SSIC 46, 47, 49). The corporate entities record have only street addresses as geographical information.

10. Daily rainfall (mm) records come from the MSS (Meteorological Service Singapore) at http://www.weather.gov.sg/climate-historical-daily/. To derive rainfall for each `subzone-date` we map each subzone to the nearest recorded

weather station using the centroid of the subzone. Our data includes 46 weather stations on record (Fig. S24). We apply linear interpolation for days where the rainfall record is missing. Temperature and wind records are sporadically available only for certain weather stations and dates and are thus not included in our data sample.

Table S1. Sources of mobile location data.

App Type	Share
AppGenre\Games	15%
AppGenre\Non-Games Apps\Tools	13%
AppGenre\Games\Puzzle	8%
AppGenre\Non-Games Apps\Media & Video	6%
AppGenre\Games\Casual	5%
AppGenre\Non-Games Apps\Communication	4%
AppGenre\Games\Arcade	4%
AppGenre\Non-Games Apps\Lifestyle	3%
AppGenre\Games\Action	3%
AppGenre\Non-Games Apps\Productivity	3%
AppGenre\Games\Simulation	3%
AppGenre\Non-Games Apps\Music & Audio	2%
AppGenre\Non-Games Apps\Photography	2%
AppGenre\Non-Games Apps\Sports	2%
AppGenre\Non-Games Apps\Books & Reference	2%
AppGenre\Non-Games Apps\Social	2%
AppGenre\Games\Word	2%
AppGenre\Non-Games Apps\Music	2%
AppGenre\Games\Board	2%
AppGenre\Non-Games Apps\Entertainment	2%
AppGenre\Non-Games Apps\Travel	2%

Table S1. Sources of mobile location data.

App Type	Share
AppGenre\Non-Games Apps\Social Networking	1%
AppGenre\Non-Games Apps\Photo & Video	1%
AppGenre\Games\Racing	1%
AppGenre\Games\Trivia	1%
AppGenre\Games\Music	1%
AppGenre\Non-Games Apps\News & Magazines	1%
AppGenre\Games\Card	1%
AppGenre\Non-Games Apps\Utilities	1%
AppGenre\Games\Role Playing	1%
AppGenre\Games\Strategy	1%
AppGenre\Games\Adventure	1%
AppGenre\Non-Games Apps\Personalization	1%
AppGenre\Non-Games Apps\Comics	<1%
AppGenre\Non-Games Apps\Navigation	<1%
AppGenre\Games\Sports Games	<1%
AppGenre\Non-Games Apps\Education	<1%
AppGenre\Games\Casino	<1%
AppGenre\Non-Games Apps\Weather	<1%
AppGenre\Non-Games Apps\Health & Fitness	<1%
AppGenre\Non-Games Apps\Travel & Local	<1%
AppGenre\Games\Sports	<1%
AppGenre\Non-Games Apps\Dating	<1%
AppGenre\Non-Games Apps\Business	<1%
AppGenre\Non-Games Apps\News	<1%
AppGenre\Non-Games Apps\House & Home	<1%
AppGenre\Non-Games Apps\Finance	<1%
AppGenre\Non-Games Apps\Art & Design	<1%
AppGenre\Non-Games Apps\Transportation	<1%

Table S1. Sources of mobile location data.

App Type	Share
AppGenre\Non-Games Apps\Medical	<1%
AppGenre\Non-Games Apps\Beauty	<1%
AppGenre\Non-Games Apps\Parenting	<1%
AppGenre\Non-Games Apps\Book	<1%
AppGenre\Non-Games Apps\Reference	<1%
AppGenre\Non-Games Apps\Shopping	<1%
AppGenre\Games\Educational	<1%
AppGenre\Non-Games Apps\Events	<1%
AppGenre\Non-Games Apps\Autos & Vehicles	<1%
AppGenre\Non-Games Apps\Food & Drink	<1%
AppGenre\Non-Games Apps\Libraries & Demo	<1%

B. Supplementary Tables to Main Figures

Table S2. Effect of wealth mismatch on movement. Coefficients estimated from Eq. (3). Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Corresponds to Fig. 4A. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by wealth	0.0181**	0.0145*	0.0120	0.0119	0.0207**
	(0.0082)	(0.0083)	(0.0084)	(0.0084)	(0.0085)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7332	0.7351	0.7353	0.7354	0.7367
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Table S3. Asymmetry in wealth mismatch. Coefficients estimated from Eq. (3). P $\rightarrow$ W indicates movement from poor to wealthy neighborhoods (or, $z_o(1-z_d)$ in Eq. (2)). W $\rightarrow$ P indicates movement from wealthy to poor neighborhoods (or, $z_d(1-z_o)$ in Eq. (2)). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Table also reports p-values for the null that (P$\rightarrow$W) + (W$\rightarrow$P) = 0. Corresponds to Fig. 4B. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
P $\rightarrow$ W	0.0690^{***}	0.0536^{***}	0.0567^{***}	0.0564^{***}	0.0479^{***}
	(0.0105)	(0.0106)	(0.0106)	(0.0106)	(0.0105)
W $\rightarrow$ P	-0.0289^{***}	-0.0217^{**}	-0.0309^{***}	-0.0307^{***}	-0.0077
	(0.0100)	(0.0101)	(0.0104)	(0.0104)	(0.0111)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: (P $\rightarrow$ W) $\neq$ (W $\rightarrow$ P)	.015	.054	.123	.125	.018
R^2	0.7336	0.7353	0.7357	0.7357	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Table S4. Asymmetry in wealth mismatch by geography. Coefficients estimated from fully interacting the wealth mismatch measure with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta \text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Table also reports p-values for the null that $\beta + \gamma = 0$. Corresponds to Fig. 6. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by wealth	0.0433^{***}	0.0376^{***}	0.0344^{***}	0.0340^{***}	0.0404^{***}
	(0.0118)	(0.0117)	(0.0118)	(0.0118)	(0.0117)
Mismatch by wealth × Non-central dest.	-0.0452^{***}	-0.0410^{***}	-0.0399^{***}	-0.0393^{***}	-0.0353^{**}
	(0.0153)	(0.0152)	(0.0152)	(0.0152)	(0.0154)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} \neq \hat{\gamma}$	.859	.755	.62	.634	.65
R^2	0.7333	0.7352	0.7354	0.7355	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Table S5. Effect of ethnic mismatch on movement. Coefficients estimated from Eq. (3). Mismatch is as defined in Eq. (2), computed using the proportion of minority ethnic residents (z in Eq. (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Corresponds to Fig. 7A. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by ethnicity	-0.2198^{***}	-0.2804^{***}	-0.3071^{***}	-0.3070^{***}	-0.2109^{**}
	(0.0758)	(0.0938)	(0.0946)	(0.0945)	(0.1020)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7338	0.7345	0.7347	0.7348	0.7360
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

Table S6. Asymmetry in ethnic mismatch. Coefficients estimated from Eq. (3). The two key independent variables [(Mnr → Maj) and (Maj → Mnr)] are the first and second term in the mismatch measure defined in Eq. (2). Mnr → Maj is the probability the origin neighborhood has high minority-ethnic population while the destination neighborhood does not. Mnr → Maj is the probability the origin neighborhood has high majority-ethnic population while the destination neighborhood does not. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Table also reports the p-value for the null that (Mnr → Maj) + (Maj → Mnr) = 0. Corresponds to Fig. 7B. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mnr → Maj	0.0703	−0.0370	−0.0486	−0.0500	−0.0292
	(0.1004)	(0.1100)	(0.1101)	(0.1098)	(0.1115)
Maj → Mnr	−0.3498***	−0.4445***	−0.4922***	−0.4907***	−0.3911***
	(0.0733)	(0.0959)	(0.0974)	(0.0974)	(0.1106)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: (Mnr → Maj) ≠ (Maj → Mnr)	.081	.011	.005	.005	.04
R^2	0.7341	0.7347	0.7349	0.7350	0.7362
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

C. Additional Robustness Tests

Table S7. Effect of distance in wealth on movement. Coefficients estimated from Eq. (3). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
$\lvert\text{wealth}_o - \text{wealth}_d\rvert$	8.6386^{***}	6.0795^{***}	6.0222^{***}	6.0103^{***}	6.4518^{***}
	(1.0152)	(0.9503)	(0.9528)	(0.9539)	(0.9471)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7344	0.7356	0.7358	0.7359	0.7372
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

Table S8. Effect of wealth mismatch on movement (neighborhood-by-day fixed effects). Coefficients estimated from Eq. (3). Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of neighborhood-by-day fixed effects. Corresponds to Table S2, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. Estimates persists after column (2) because any remaining variation in the neighborhood-specific covariates have been subsumed by the neighborhood-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by wealth	0.0345***	0.0268***	0.0268***	0.0268***	0.0268***
	(0.0100)	(0.0098)	(0.0098)	(0.0098)	(0.0098)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7485	0.7484	0.7484	0.7484	0.7484
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,408	1,228,357	1,228,357	1,228,357	1,228,357

Table S9. Asymmetry in wealth mismatch by geography (neighborhood-by-day fixed effects). Coefficients estimated from fully interacting the wealth mismatch measure with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta \text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of neighborhood-by-day fixed effects. Estimates persists after column (2) because any remaining variation in the neighborhood-specific covariates have been subsumed by the neighborhood-by-day fixed effects. Corresponds to Table S4, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by wealth	0.0595^{***}	0.0512^{***}	0.0512^{***}	0.0512^{***}	0.0512^{***}
	(0.0162)	(0.0155)	(0.0155)	(0.0155)	(0.0155)
Mismatch by wealth × Non-central dest.	-0.0407^{**}	-0.0396^{**}	-0.0396^{**}	-0.0396^{**}	-0.0396^{**}
	(0.0191)	(0.0185)	(0.0185)	(0.0185)	(0.0185)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.111	.32	.32	.32	.32
R^2	0.7486	0.7485	0.7485	0.7485	0.7485
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,408	1,228,357	1,228,357	1,228,357	1,228,357

Table S10. Effect of ethnic mismatch on movement (neighborhood-by-day fixed effects). Coefficients estimated from Eq. (3). Mismatch is as defined in Eq. (2), computed using the proportion of minority ethnic residents (z in Eq. (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of neighborhood-by-day fixed effects. Estimates persists after column (2) because any remaining variation in the neighborhood-specific covariates have been subsumed by the neighborhood-by-day fixed effects. Corresponds to Table S5, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by ethnicity	-0.7238^{***}	-1.0694^{***}	-1.0694^{***}	-1.0694^{***}	-1.0694^{***}
	(0.2678)	(0.2560)	(0.2560)	(0.2560)	(0.2561)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7484	0.7478	0.7478	0.7478	0.7478
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,088	1,255,555	1,255,555	1,255,555	1,255,555

Table S11. Effect of ethnic mismatch on movement (log-log specification). Coefficients estimated from Eq. (3). Mismatch is as defined in Eq. (2), computed using the proportion of minority ethnic residents (z in Eq. (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Corresponds to Fig. 7A. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
log(Mismatch by ethnicity)	-0.0819^{***}	-0.0868^{***}	-0.0944^{***}	-0.0943^{***}	-0.0667^{**}
	(0.0271)	(0.0308)	(0.0310)	(0.0310)	(0.0331)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7338	0.7345	0.7347	0.7348	0.7360
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

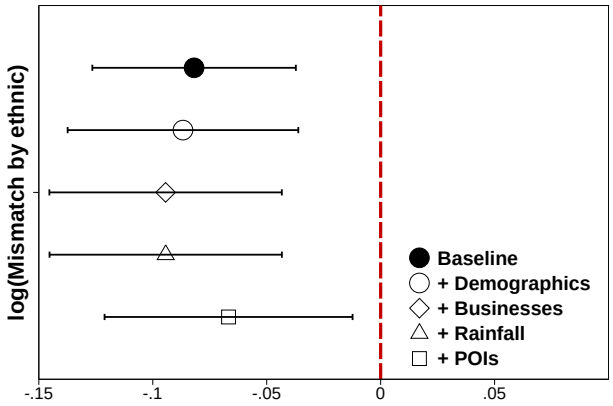

Fig. S1. Mismatch in ethnicity — log-log specification. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Dependent variable is log inflow from origin o to destination d. Figure corresponds to Fig. 7 and Table S11. Capped horizontal lines are 90% confidence intervals from standard errors clustered at the origin-by-destination area level.

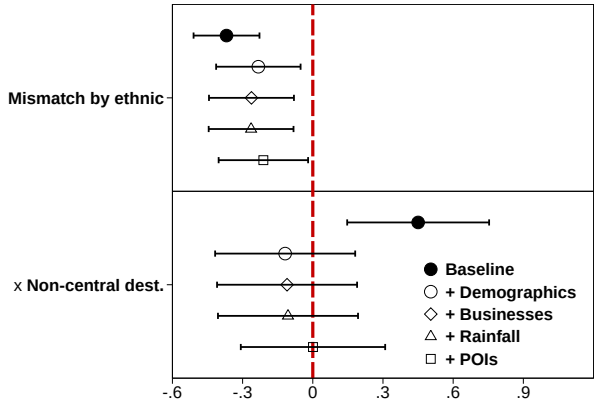

Fig. S2. Asymmetry in ethnic mismatch by geography. Baseline model includes neighborhood land area, population density (by both census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Dependent variable is log inflow from origin o to destination d. Figure corresponds to Table S12. Capped horizontal lines are 90% confidence intervals from standard errors clustered at the origin-by-destination area level.

Table S12. Asymmetry in ethnic mismatch by geography. Coefficients estimated from fully interacting the wealth mismatch measure with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta\text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Table also reports p-values for the null that $\beta + \gamma = 0$. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by ethnic	-0.3687^{***}	-0.2327^{**}	-0.2621^{**}	-0.2637^{**}	-0.2112^{*}
	(0.0856)	(0.1096)	(0.1103)	(0.1102)	(0.1162)
Mismatch by ethnic × Non-central dest.	0.4504^{**}	-0.1181	-0.1099	-0.1060	0.0008
	(0.1843)	(0.1819)	(0.1818)	(0.1819)	(0.1875)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.606	.023	.016	.017	.2
R^2	0.7340	0.7345	0.7347	0.7348	0.7360
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,127	1,255,583	1,255,583	1,255,583	1,255,583

Table S13. Asymmetry in ethnic mismatch by geography (neighborhood-by-day fixed effects). Coefficients estimated from fully interacting the wealth mismatch measure with an indicator variable for the destinations outside the central region:

$$\log(\text{inflow})_{odt} = \alpha + \beta\text{mismatch} + \gamma(\text{mismatch} \times \mathbb{1}^{\text{Non-central dest.}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{Non-central dest.}}$ is an indicator for destinations outside the census central region. Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of neighborhood-by-day fixed effects. Corresponds to Table S2, but with neighborhood-by-day fixed effects instead of census area-by-day fixed effects. Estimates persists after column (2) because any remaining variation in the neighborhood-specific covariates have been subsumed by the neighborhood-by-day fixed effects. Table also reports p-values for the null that $\beta + \gamma = 0$. Corresponds to Table S12. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by ethnic	-0.9852^{***}	-1.1058^{***}	-1.1058^{***}	-1.1058^{***}	-1.1058^{***}
	(0.2692)	(0.2626)	(0.2626)	(0.2626)	(0.2627)
Mismatch by ethnic × Non-central dest.	0.7485^{***}	0.1036	0.1036	0.1036	0.1036
	(0.2307)	(0.2155)	(0.2155)	(0.2155)	(0.2156)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
p-val: $\hat{\beta} + \hat{\gamma} = 0$	.46	.001	.001	.001	.001
R^2	0.7487	0.7478	0.7478	0.7478	0.7478
Days	91	91	91	91	91
Clusters	1,592	1,510	1,510	1,510	1,510
Observations	1,394,088	1,255,555	1,255,555	1,255,555	1,255,555

Table S14. Effect of wealth and ethnic mismatch on movement. Coefficients estimated from Eq. (3). Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. Mismatch by ethnic is computed using the proportion of minority ethnic residents (z in Eq. (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by wealth	0.0214^{**}	0.0182^{**}	0.0159^{*}	0.0158^{*}	0.0233^{***}
	(0.0085)	(0.0084)	(0.0084)	(0.0084)	(0.0085)
Mismatch by ethnic	-0.2873^{***}	-0.2713^{***}	-0.2945^{***}	-0.2934^{***}	-0.2223^{**}
	(0.0924)	(0.0959)	(0.0964)	(0.0962)	(0.1034)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7336	0.7352	0.7354	0.7355	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,308,012	1,228,385	1,228,385	1,228,385	1,228,385

Table S15. Effect of wealth and ethnic mismatch on movement (neighborhood-by-day fixed effects). Coefficients estimated from Eq. (3). Mismatch measure is as defined in Eq. (2). Wealth is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. Mismatch by ethnic is computed using the proportion of minority ethnic residents (z in Eq. (2)) in the neighborhoods. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of neighborhood-by-day fixed effects. Estimates persists after column (2) because any remaining variation in the neighborhood-specific covariates have been subsumed by the neighborhood-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Mismatch by wealth	0.0339^{***}	0.0323^{***}	0.0323^{***}	0.0323^{***}	0.0323^{***}
	(0.0103)	(0.0097)	(0.0097)	(0.0097)	(0.0097)
Mismatch by ethnic	-0.9170^{***}	-1.1885^{***}	-1.1885^{***}	-1.1885^{***}	-1.1885^{***}
	(0.3268)	(0.2593)	(0.2593)	(0.2593)	(0.2594)
Day fixed effects	✓	✓	✓	✓	✓
Neighborhood-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7486	0.7487	0.7487	0.7487	0.7487
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,307,973	1,228,357	1,228,357	1,228,357	1,228,357

D. Sensitivity Tests Ordered by Effect Size

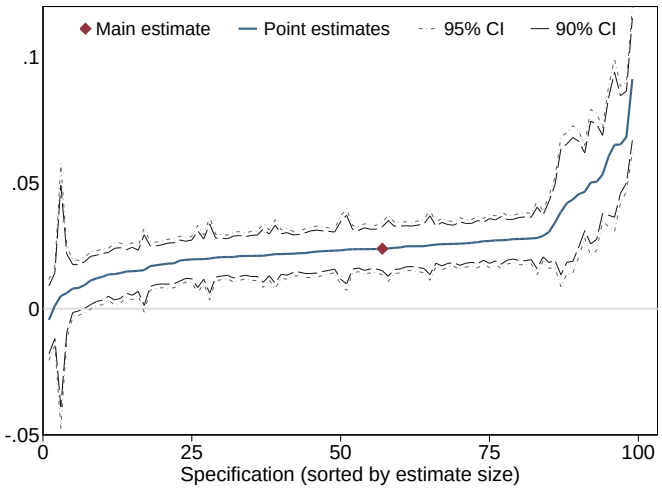

Fig. S3. Sensitivity for mismatch by wealth results (sorted). Corresponds to the unsorted plot in Fig. 8 and Table S2. Coefficients are sorted by size for a sense of where the main estimates (by 25th percentile cutoff) fall in the distribution.

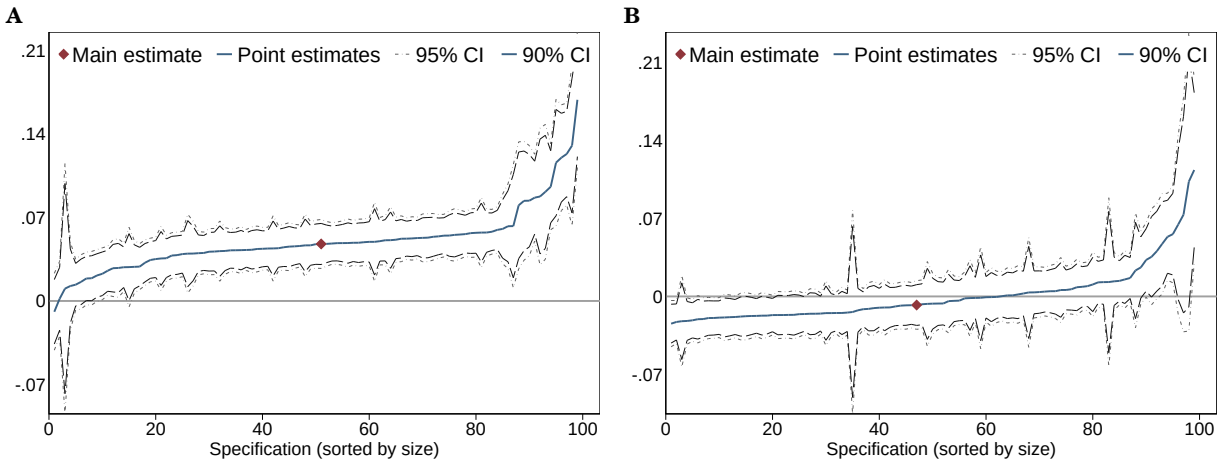

Fig. S4. Sensitivity tests for wealth mismatch by direction (sorted). **(A)** Sensitivity of the $(P \rightarrow W)$ coefficient from the top panel of Fig. 4B. **(B)** Sensitivity of the $(P \rightarrow W)$ coefficient from the bottom panel of Fig. 4B. Corresponds to the unsorted plot in Fig. 9 and Table S3. Coefficients are sorted by size for a sense of where the main estimates (by 25th percentile cutoff) fall in the distribution.

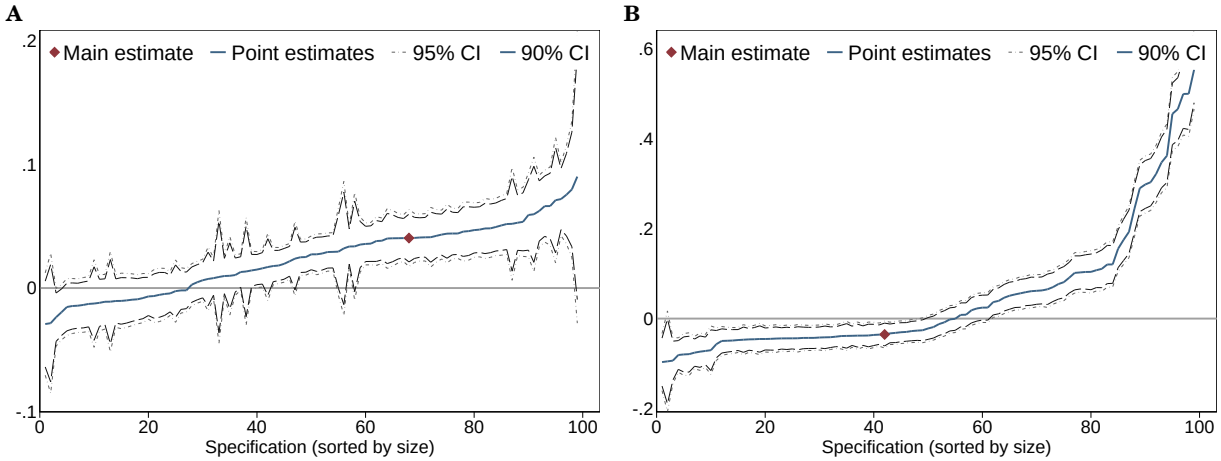

Fig. S5. Sensitivity tests for wealth mismatch by geography (sorted). **(A)** Sensitivity of the mismatch by wealth coefficient in the top panel of Fig. 6. **(B)** Sensitivity of the mismatch by wealth interacted with non-central destination coefficient in the bottom panel of Fig. 6. Corresponds to the unsorted plot in Fig. 10 and Table S4. Coefficients are sorted by size for a sense of where the main estimates (by 25th percentile cutoff) fall in the distribution.

E. Representativeness of GPS pings by neighborhood demographics

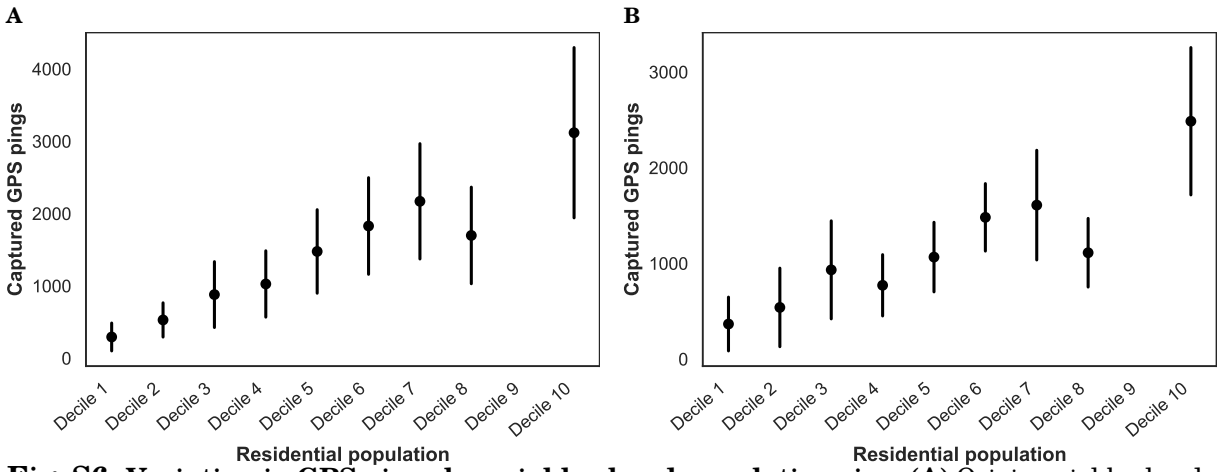

Fig. S6. Variation in GPS pings by neighborhood population size. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is bins for neighborhood populaton size. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

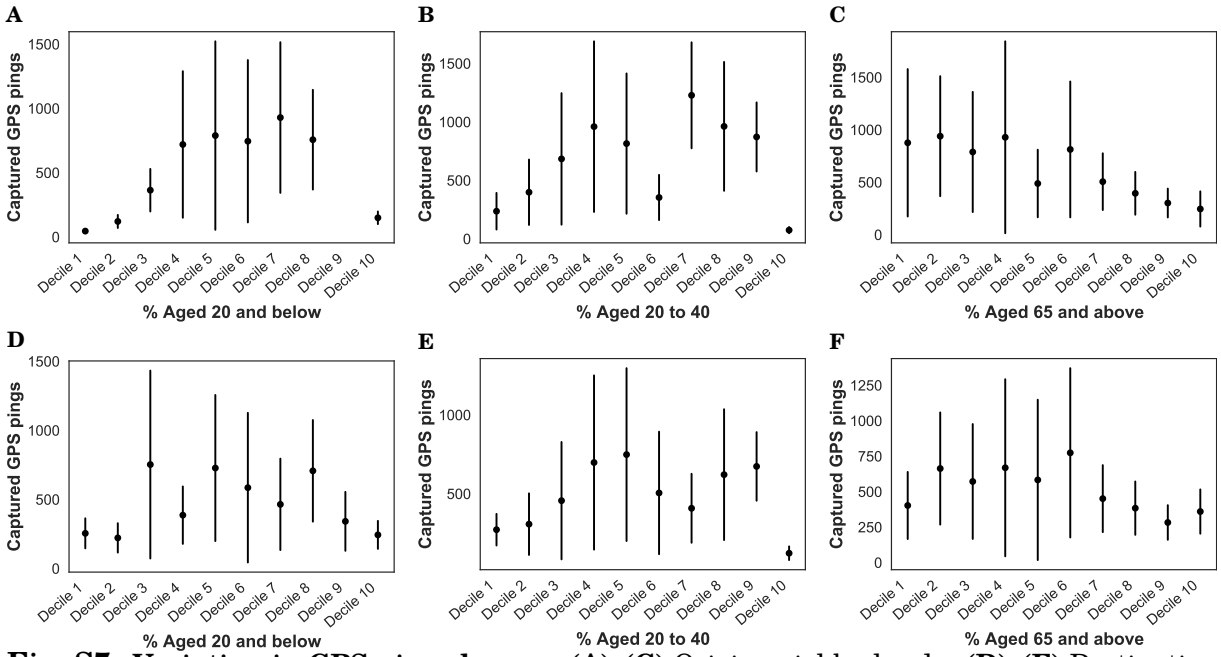

Fig. S7. Variation in GPS pings by age. **(A)–(C)** Origin neighborhoods. **(D)–(F)** Destination neighborhoods. Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

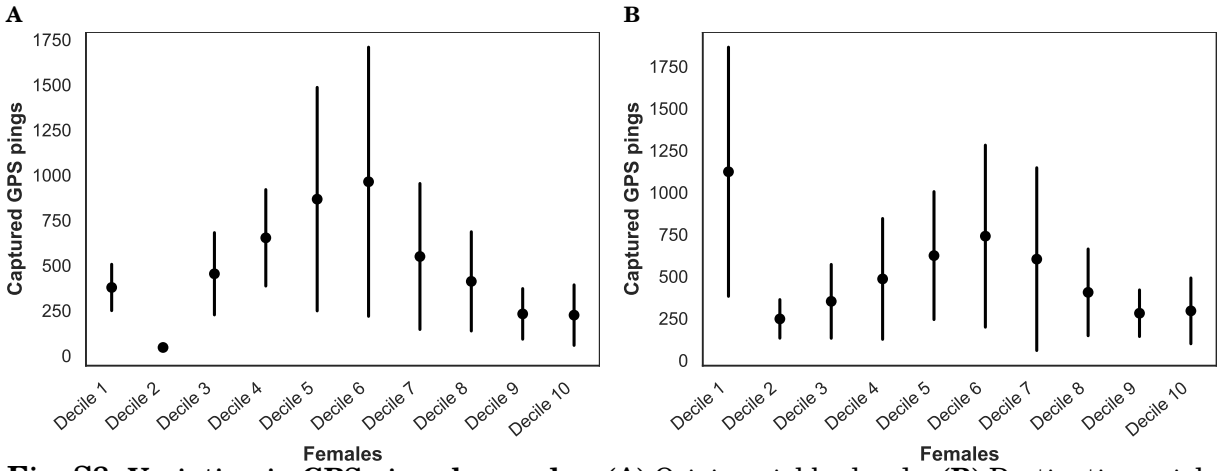

Fig. S8. Variation in GPS pings by gender. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is female proportion in a neighborhood. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

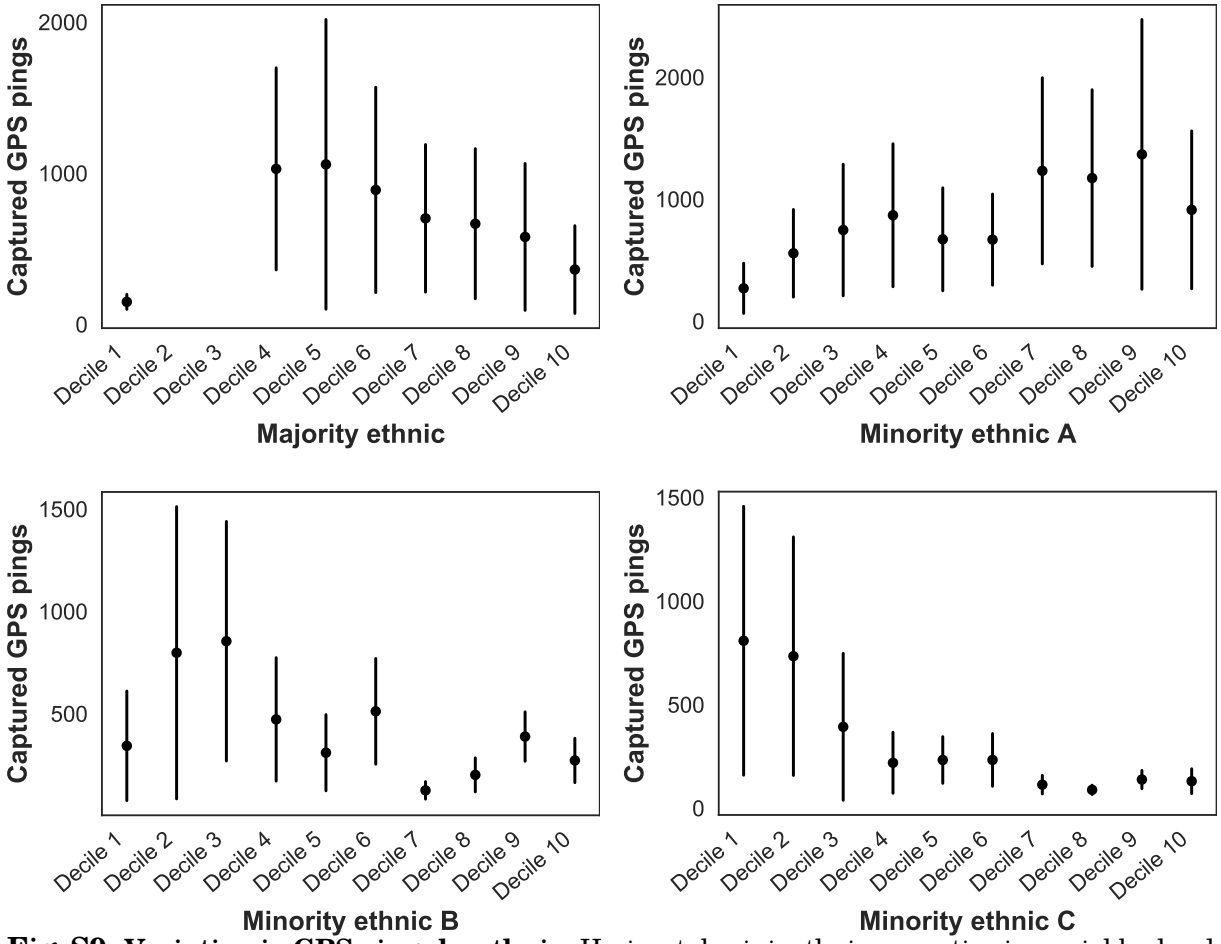

Fig. S9. Variation in GPS pings by ethnic. Horizontal axis is ethnic proportion in a neighborhood. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

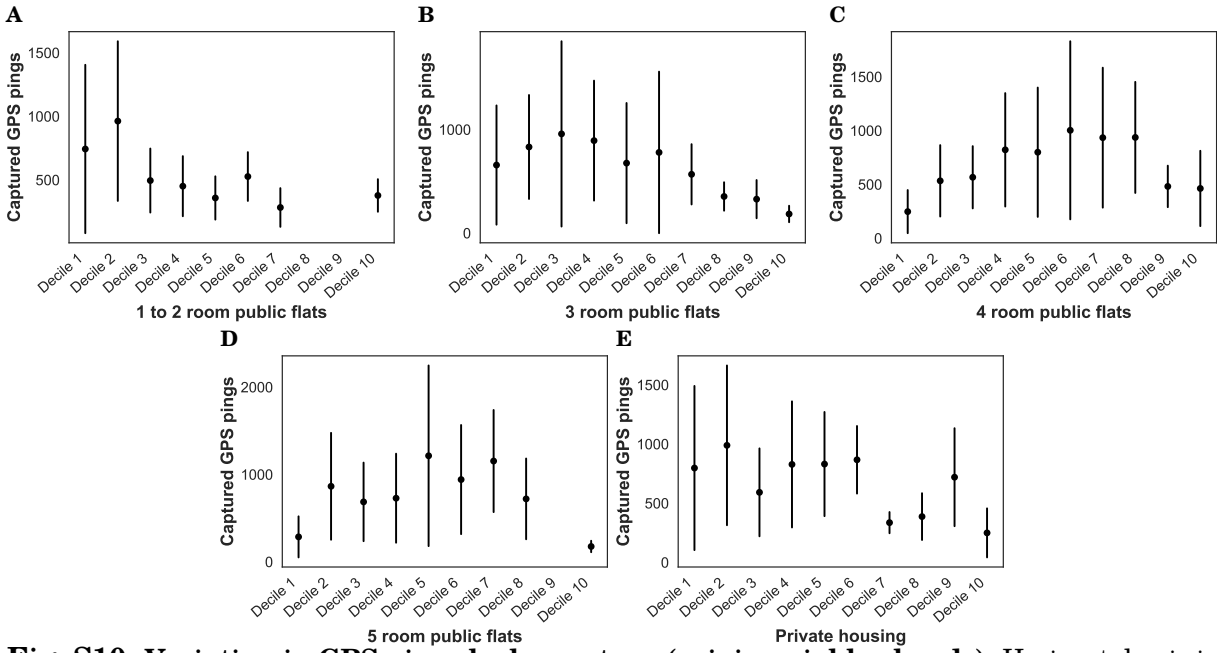

Fig. S10. Variation in GPS pings by house type (origin neighborhoods). Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

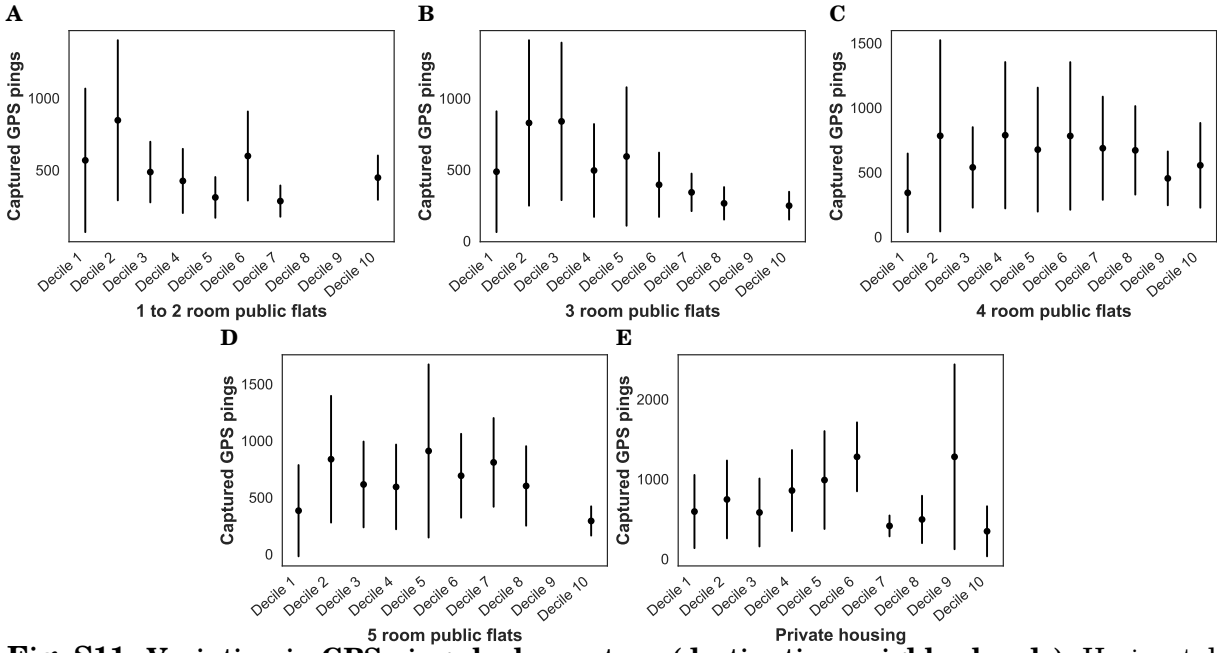

Fig. S11. Variation in GPS pings by house type (destination neighborhoods). Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date.

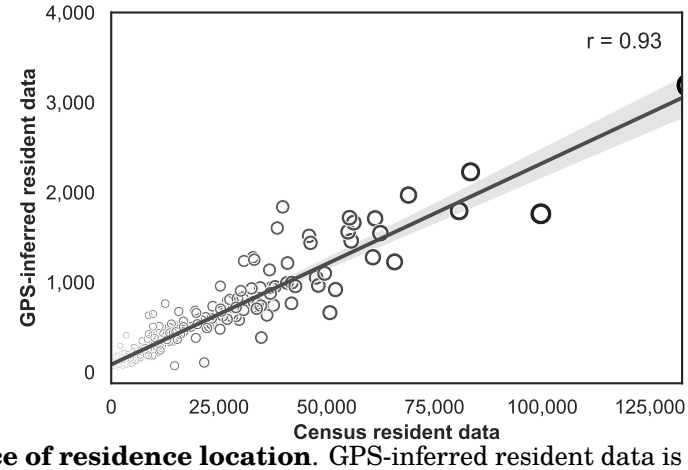

Fig. S12. Inference of residence location. GPS-inferred resident data is the average number of captured devices across neighborhood-days inferred as residing in that neighborhood. Supplementary materials section A. Data Appendix provides more details on inferring residence. Figure reports the correlation between the GPS-inferred resident size and the census resident size.

F. Representativeness of GPS pings by other characteristics

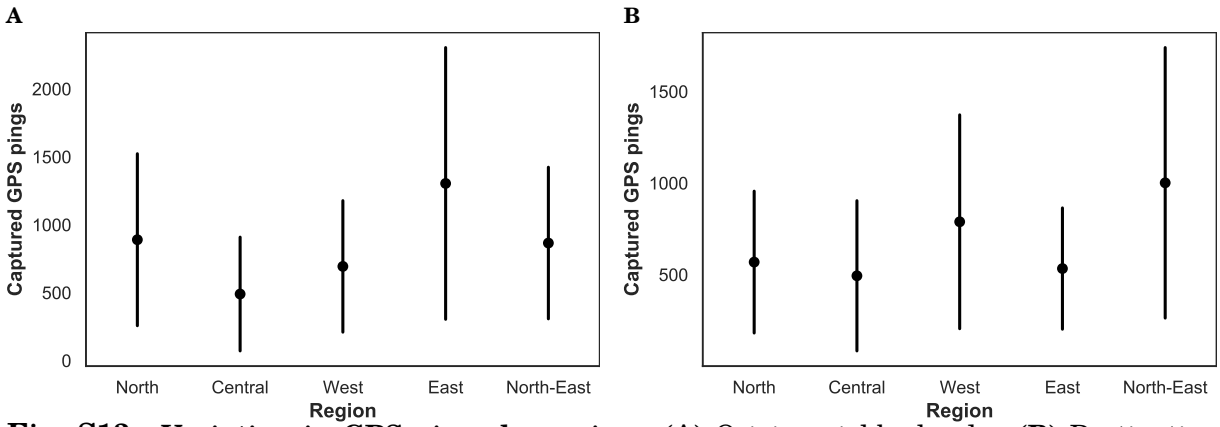

Fig. S13. Variation in GPS pings by region. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is for the five regions of administrative boundaries. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

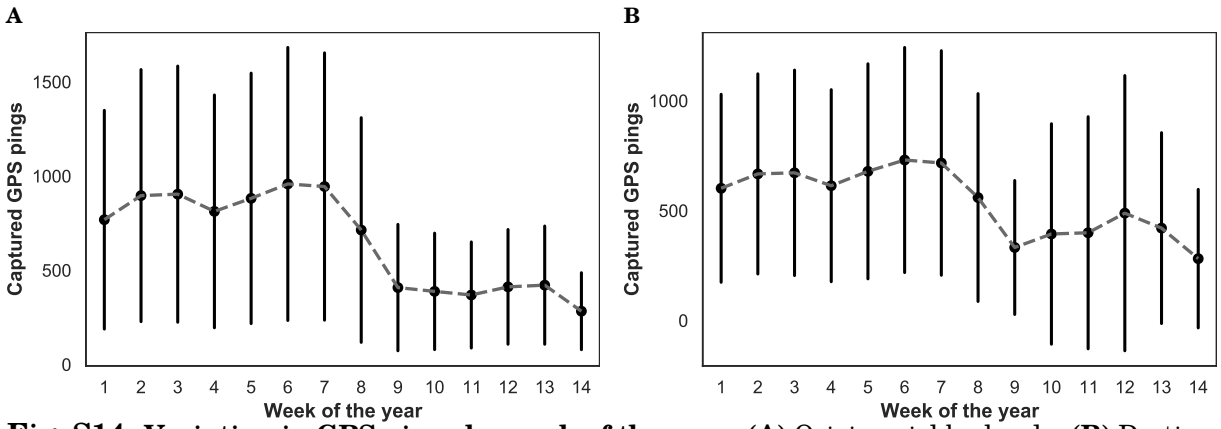

Fig. S14. Variation in GPS pings by week of the year. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is the week of the year. Vertical axis is the number of captured GPS pings by neighborhood-date. Vertical bars are standard deviations.

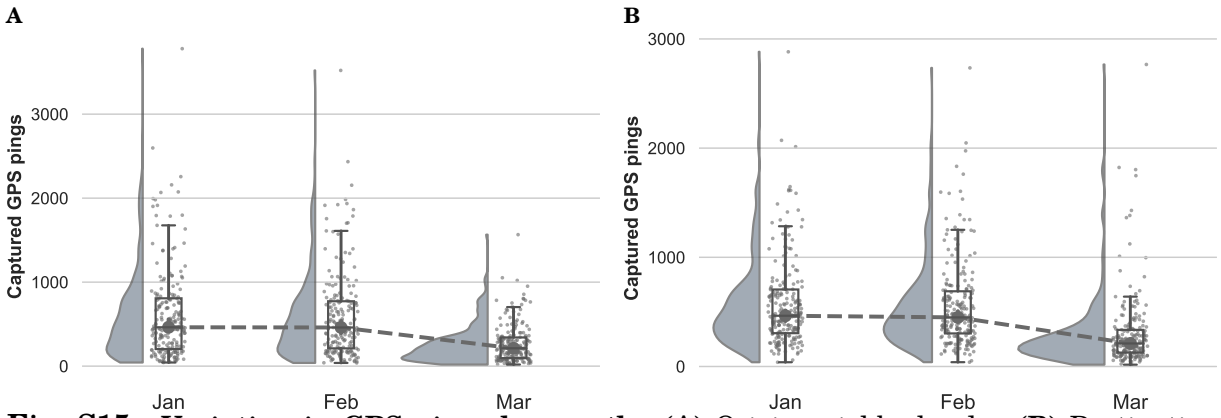

Fig. S15. Variation in GPS pings by month. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is month. Vertical axis is the number of captured GPS pings by neighborhood-date.

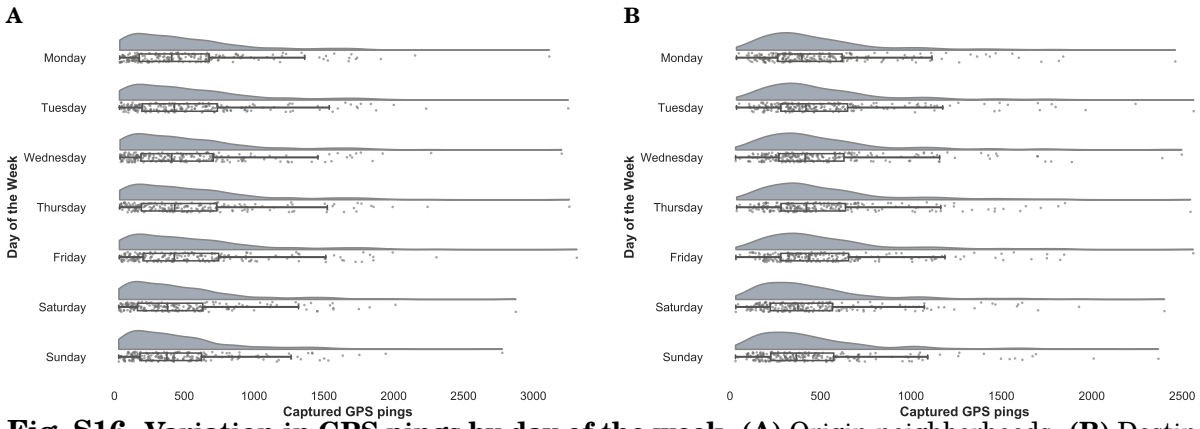

Fig. S16. Variation in GPS pings by day of the week. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. Horizontal axis is day of the week. Vertical axis is the number of captured GPS pings by neighborhood-date.

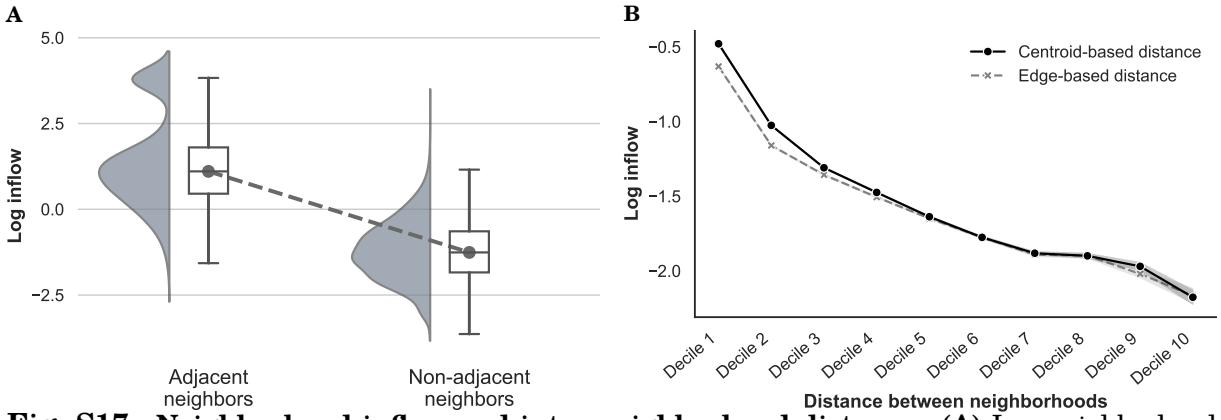

Fig. S17. Neighborhood inflow and inter-neighborhood distance. **(A)** Log neighborhood inflow for adjacent vs non-adjacent neighborhoods. **(B)** Log neighborhood inflow and distance for non-adjacent neighborhoods. Shaded areas are the 95% confidence intervals constructed using 10,000 bootstraps.

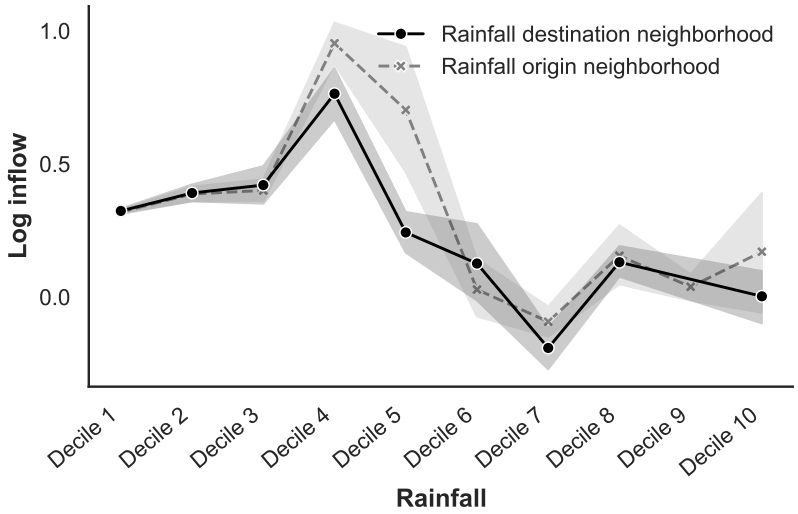

Fig. S18. Neighborhood inflow and precipitation. Solid black line is the level of precipitation in the destination neighborhood. Dashed gray line is the level of precipitation in the origin neighborhood. Corresponding shaded areas are the 95% confidence intervals constructed using 10,000 bootstraps.

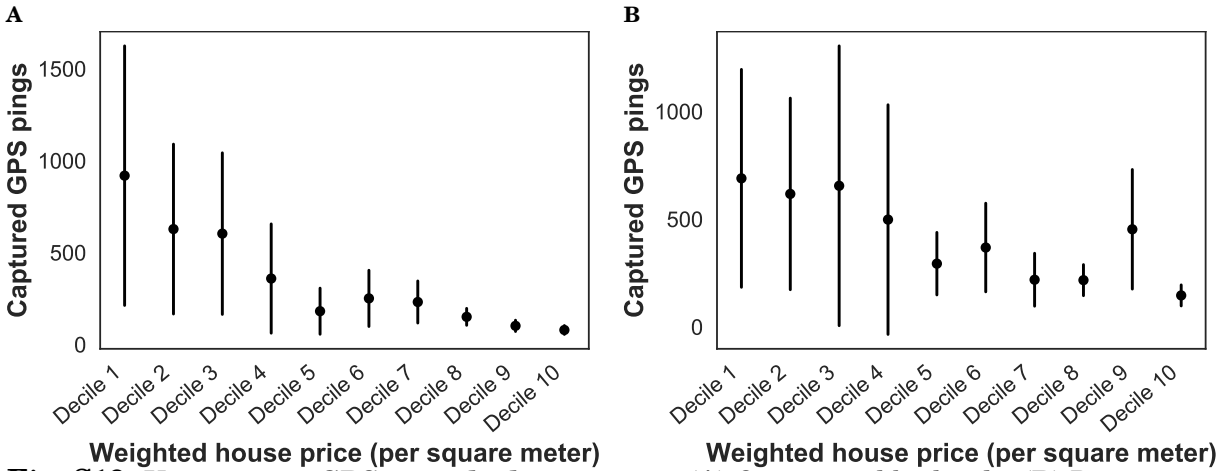

Fig. S19. Variation in GPS pings by house price. **(A)** Origin neighborhoods. **(B)** Destination neighborhoods. House price is the weighted price per square meters determined by the proportion of neighborhood residents who reside in public vs private housing. Vertical bars are standard deviations.

G. Additional Institutional Information

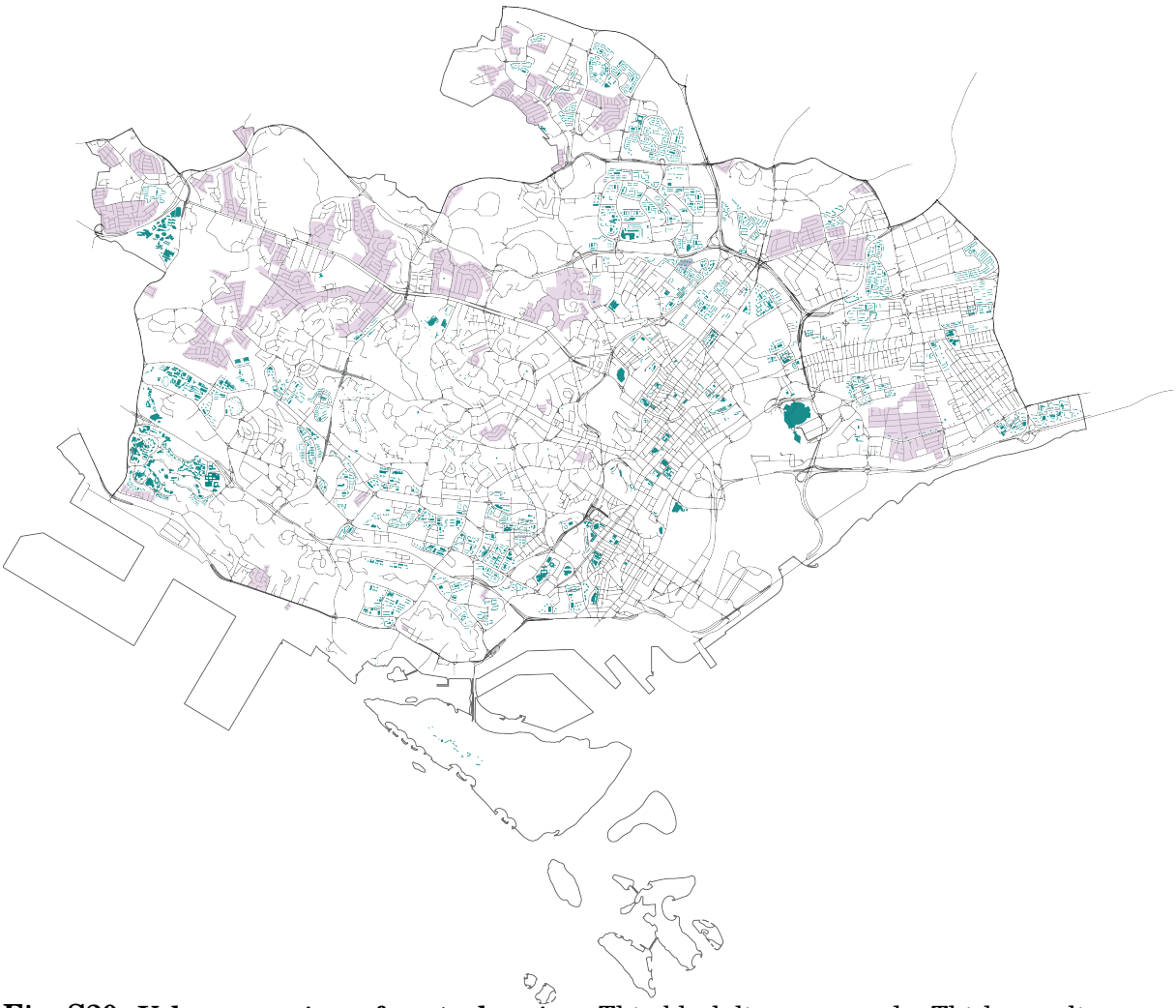

Fig. S20. Urban overview of central region. Thin black lines are roads. Thick gray lines are highways. Blue polygons are buildings, which include the public housing flats. Pink polygons are the single-family houses with dedicated land space. Corresponds to Fig. 5 and Fig. S21.

Table S16. Average monthly household income from work per household member (including employer CPF contributions) among resident employed households by deciles. The average and median monthly household incomes per household member is $4,022 and $2,886 in 2020 (Department of Statistics, Singapore (2021), for resident employed households only and includes Employer CPF contributions).

Deciles	2015	2020
1st - 10th	541	560
11th - 20th	1,040	1,141
21st - 30th	1,446	1,609
31st - 40th	1,857	2,085
41st - 50th	2,274	2,603
51st - 60th	2,780	3,201
61st - 70th	3,409	3,940
71st - 80th	4,276	4,972
81st - 90th	5,804	6,712
91st - 100th	12,816	13,400

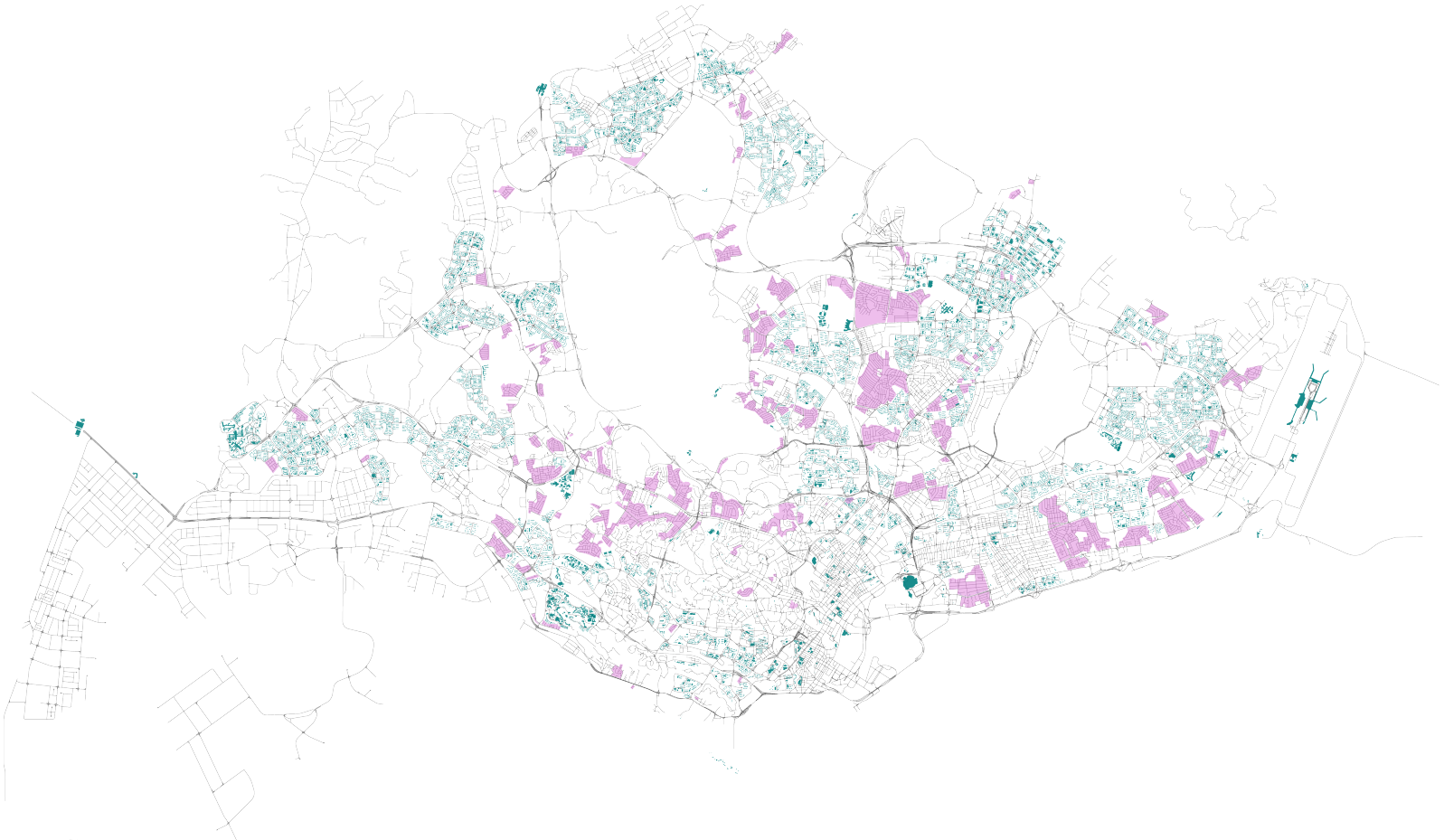

Fig. S21. Urban overview of Singapore. Thin black lines are roads. Thick gray lines are highways. Blue polygons are buildings, which include the public housing flats. Pink polygons are the single-family houses with dedicated land space.

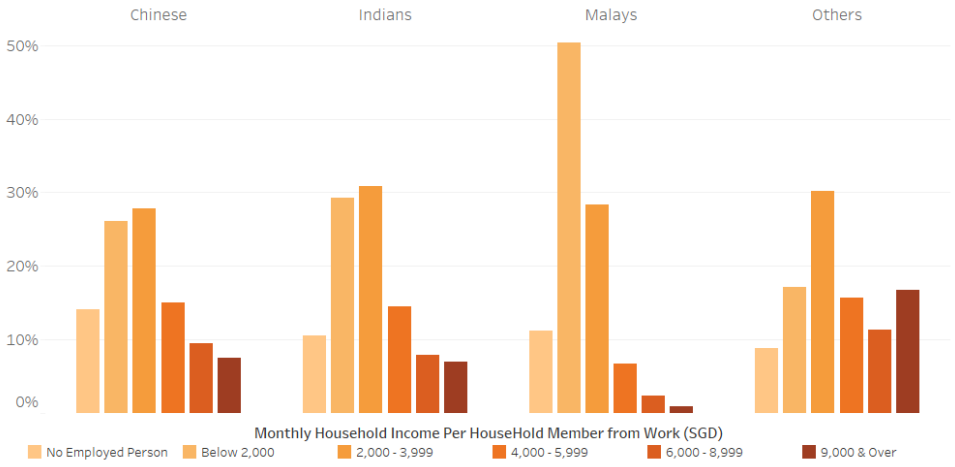

Fig. S22. Percentage of Resident Households by Monthly Household Income Per Household Member and Ethnic Group. Each bar is calculated by taking the number of residents in each income group and dividing by the total number of residents across income groups for each ethnic group. The distribution of monthly household income per household member of Malay residents skews largely towards the lower income groups with 50.4 per cent and 1.0 per cent earning below $2,000 (below median income) and earning above $9,000 (highest bracket), respectively. The distributions for Chinese, Indian, and 'Others' Residents are more spread out with 26.0 and 7.5 per cent, 29.3 and 6.9 per cent, and 17.2 and 16.7 per cent earning below $2,000 and earning above $9,000, respectively.

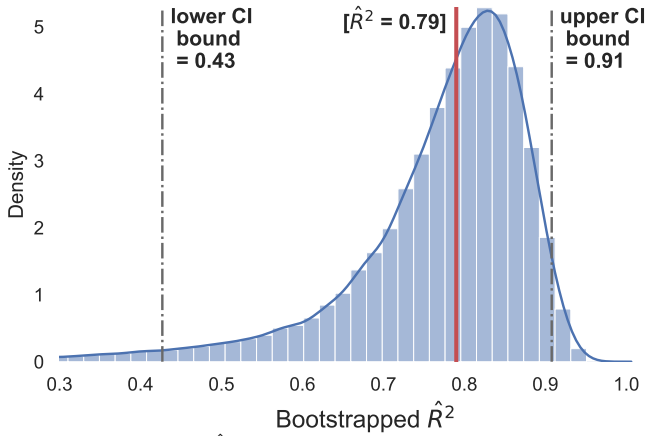

Fig. S23. Bootstrapped values of $\hat{R}^2$ from regressing house price on census income. At the 28 census planning areas where census income data is available. Number of bootstrap samples is 100,000. Red vertical line indicates the sample $\hat{R}^2$ value. The gray dotted lines indicate the lower and upper 95% confidence interval bounds. Corresponds to Fig. 3.

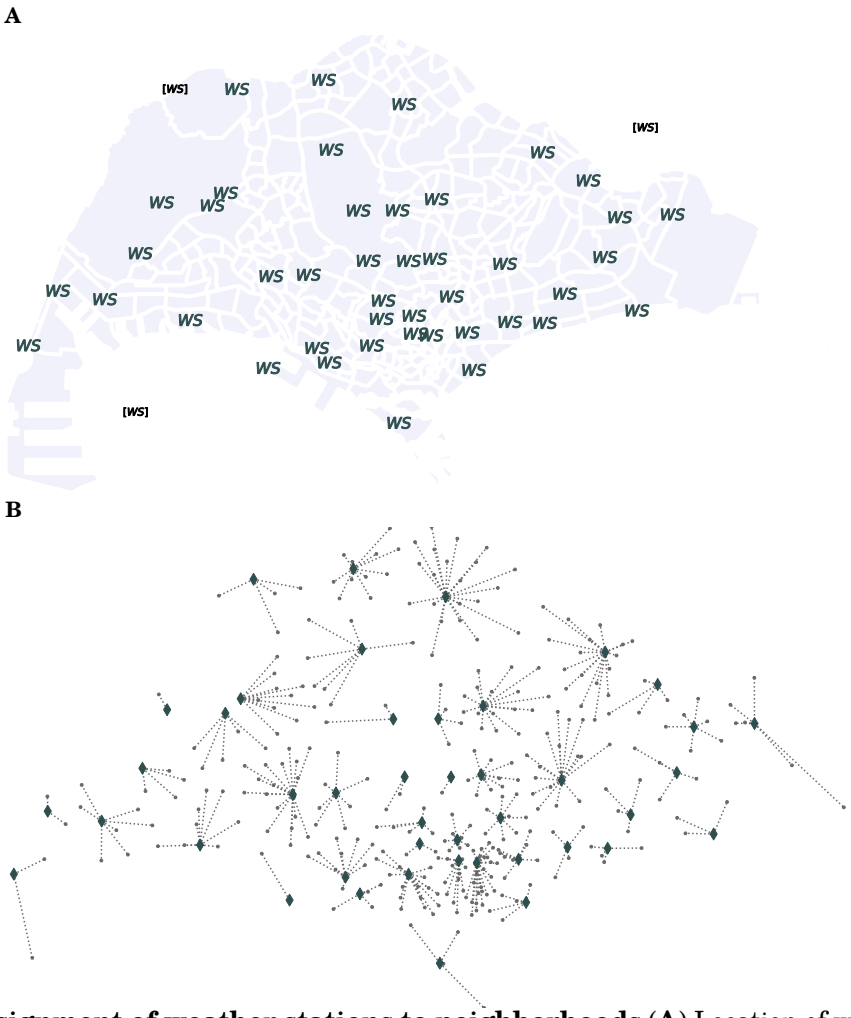

Fig. S24. Assignment of weather stations to neighborhoods (**A**) Location of weather stations in Singapore. The three weather stations that end up with no area mapping is shown in brackets. (**B**) Matches encoded by dotted lines, connecting weather stations to neighborhoods.

H. Neighborhood Flows by House Type

The earlier version of this study examines how poverty—proxied using residence type—predicts neighborhood visits. See Lee et al. (2021).

I. Neighborhood Flows During Chinese New Year

This analysis tests whether intergenerational mobility can be detected, to the extent that the neighborhood visiting patterns reverses during a key period where younger individuals and families disproportionately travel to visit their elders, including parents. Specifically, we pivot the analysis on the two days from 25–26 Jan for the Chinese New Year holidays where Chinese individuals visit their elders.

In Table S17, we find that neighborhood visits is highly correlated with wealth levels. In origin neighborhoods, higher wealth is correlated with lower outgoing flows, while in destination neighborhoods, higher wealth is‘ correlated with higher incoming flows.

If there is intergenerational mobility, then the estimated effects from Table S17 should reverse during the Chinese New Year period where the young visit their elder relatives. Specifically, during the Chinese New Year period, origin neighborhoods with higher wealth should be linked to higher outflows while destination neighborhoods with higher wealth should be linked to lower inflows. We model this reversal by interacting a Chinese New Year indicator variable with the origin and destination wealth levels, with the interactions allowing for the differential effects during Chinese New Year discussed above.

From Table S18, the coefficients of the interactions are all statistically indistinguishable from zero, suggesting that our sample is unable to capture the hypothesized

Table S17. Effect of wealth levels on movement. Coefficients estimated from Eq. (3). Wealth levels are measured by the price per square meters (in thousands) of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Wealth in origin	-0.0026^{*}	-0.0028^{*}	-0.0029^{*}	-0.0029^{*}	-0.0029^{*}
	(0.0014)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Wealth in destination	0.0046^{***}	0.0054^{***}	0.0056^{***}	0.0056^{***}	0.0058^{***}
	(0.0010)	(0.0012)	(0.0013)	(0.0013)	(0.0013)
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7334	0.7353	0.7355	0.7356	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

movement patterns during the Chinese New Year period. This could be because of a lack of statistical power or because of non-linearities.

Table S18. Movement patterns during CNY. Coefficients estimated from Eq. (3), with the poverty measures fully interacted with an indicator variable for the Chinese New Year period (24–27 Jan 2020):

$$\log(\text{inflow})_{odt} = \alpha + \beta_d P_d + \beta_o P_o + \gamma_d (P_d \times \mathbb{1}^{\text{CNY}}) + \gamma_o (P_o \times \mathbb{1}^{\text{CNY}}) + \Gamma_t X_{odt} + \varepsilon_{odt},$$

where $\mathbb{1}^{\text{CNY}}$ is an indicator variable for the Chinese New Year period (25–26 Jan 2020). Wealth levels is the price per square meters of houses weighted by the share of public and private residence in the neighborhood population. The model in column (1) includes neighborhood land area, population density (both by census and real-time records), neighborhood contiguity and distance, and the full interaction of census area-by-day fixed effects. Table also reports the one-sided tests for the hypothesized interaction effect. Standard errors are clustered at the origin-by-destination census planning area level. *** Significant at the 1 per cent level. ** Significant at the 5 per cent level. * Significant at the 10 per cent level.

	Dependent variable: Log neighborhood inflow				
Independent variable	**1**	**2**	**3**	**4**	**5**
Wealth in origin	-0.0027^{*}	-0.0028^{*}	-0.0029^{*}	-0.0029^{*}	-0.0029^{*}
	(0.0014)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
Wealth in destination	0.0047^{***}	0.0054^{***}	0.0056^{***}	0.0056^{***}	0.0057^{***}
	(0.0010)	(0.0012)	(0.0013)	(0.0013)	(0.0013)
Wealth in origin $\times \mathbb{1}^{\text{CNY}}$	0.0005	0.0001	0.0001	0.0003	0.0003
	(0.0012)	(0.0014)	(0.0014)	(0.0014)	(0.0014)
Wealth in destination $\times \mathbb{1}^{\text{CNY}}$	-0.0015	0.0006	0.0007	0.0007	0.0012
	(0.0009)	(0.0011)	(0.0011)	(0.0011)	(0.0012)
H_a : (Wealth in origin $\times \mathbb{1}^{\text{CNY}}) < 0$, p-val	.67	.519	.516	.581	.576
H_a : (Wealth in dest. $\times \mathbb{1}^{\text{CNY}}) > 0$, p-val	.946	.31	.268	.264	.165
Day fixed effects	✓	✓	✓	✓	✓
Census area-by-day fixed effects	✓	✓	✓	✓	✓
Demographics		✓	✓	✓	✓
Businesses			✓	✓	✓
Neighborhood-specific rainfall				✓	✓
Places of interest					✓
R^2	0.7334	0.7353	0.7355	0.7356	0.7368
Days	91	91	91	91	91
Clusters	1,439	1,434	1,434	1,434	1,434
Observations	1,321,467	1,228,385	1,228,385	1,228,385	1,228,385

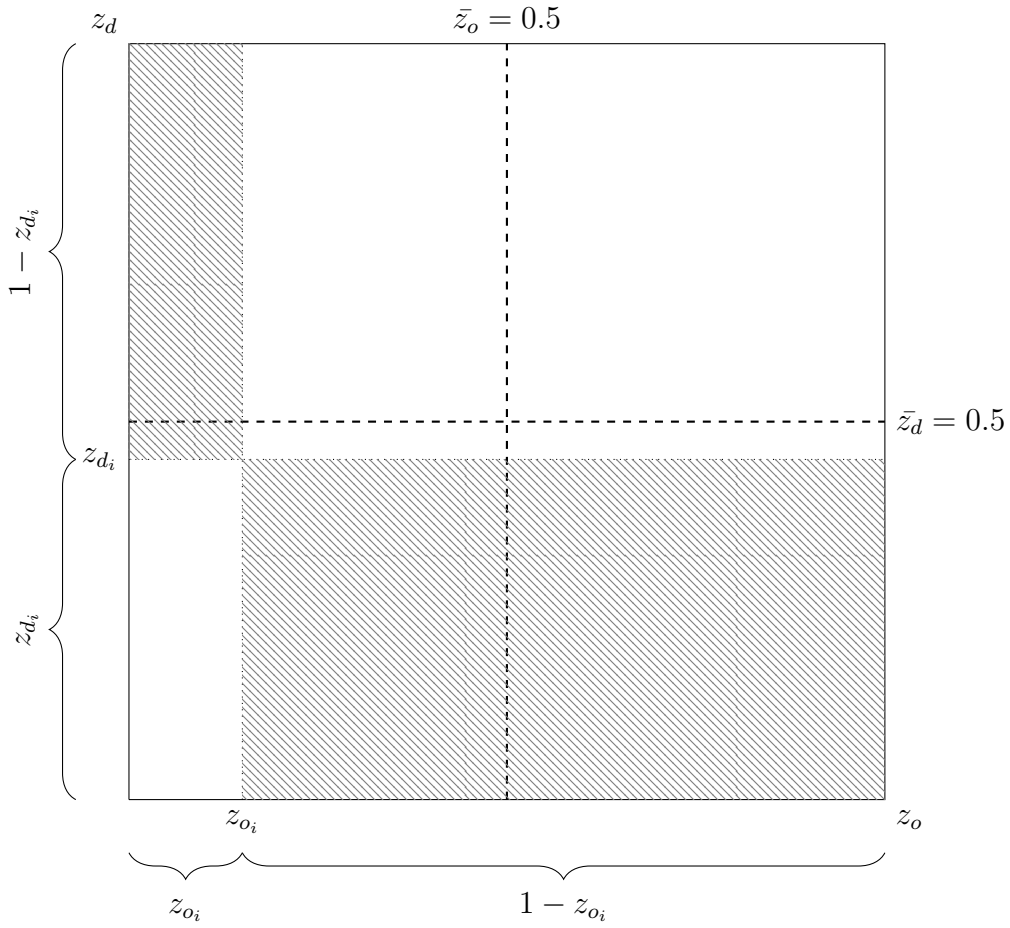

Fig. S25. Sample space of mismatch. Unit square with length one on all sides. Shaded area is $z_o(1-z_d) + z_d(1-z_o)$ which indicates the magnitude of the level of mismatch between neighborhoods o and d as defined in Eq. (2). $\bar{z_o}$ and $\bar{z_d}$ indicated by the dotted lines are thresholds. Without loss of generality, given z_o, an increase in z_d increases or decreases the shaded area depending on whether z_o is more or less than $\frac{1}{2}$:

$$\frac{\partial \{z_o(1-z_d) + z_d(1-z_o)\}}{\partial z_d} = 1 - 2z_o \lessgtr 0 \text{ if } z_o \lessgtr \frac{1}{2}$$

and z is bounded above by $\frac{1}{2}$ in the sample (ignoring extreme values on the right tail), so an increase in z_d increases the mismatch measure in the sample.