Rubbing Shoulders:
Class Segregation in Daily Activities*

Maxim Massenkoff† Nathan Wilmers‡

June 2023

Abstract

We use location data to study activity and encounters across class lines. Low-income and especially high-income individuals are socially isolated: more likely than other income groups to encounter people from their own social class. Using simple counterfactual exercises, we study the causes. While some industries cater mainly to low or high-income groups (for example, golf courses and wineries), industry alone explains only a small share of isolation. People are most isolated when they are close to home, and the tendency to go to nearby locations explains about one-third of isolation. Using our uniquely detailed data, we show that brands, combined with distance, explain about half the isolation of the rich. Casual restaurant chains, like Olive Garden and Applebee's, have the largest positive impact on cross-class encounters through both scale and their diversity of visitors. Dollar stores and local pharmacies like CVS deepen isolation. Among publicly-funded spaces, libraries and parks are more redistributive than museums and historical sites. And, despite prominent restrictions on chain stores in some large US cities, chains are more class diverse than independent stores. The mix of establishments in a neighborhood is strongly associated with cross-class Facebook friendships (Chetty et al., 2022). The results uncover how policies that support certain public and private spaces might impact the connections that form across class divides.

*For helpful comments we thank Kirill Borusyak, Keith Chen, Jacob Grumbach, Jeremy Pearce, Anna Stansbury, and seminar participants at the European Labor Symposium for Early Career Economists symposium and Copenhagen Business School.

†Naval Postgraduate School. Email: maxim.massenkoff@nps.edu

‡MIT Sloan. Email: wilmers@mit.edu

1 Introduction

Rising income inequality has renewed concerns about economic classes pulling apart. Recent research documents high levels of residential segregation by income (Reardon and Bischoff, 2011); educational polarization in political preferences (Gethin et al., 2022); income sorting across schools and universities (Chetty et al., 2020); and high levels of skill-based sorting across firms (Song et al., 2019). Many high-income Americans seem to live in their own enclaves, vote as a block, study at their own schools, and work at the same firms.

Yet, much of what defines identity and social interaction occurs not at home but in public spaces where we shop and eat and voluntarily spend time. This crucial interaction in "third places" (Oldenburg, 1999) could offset economic segregation in other spheres of society. If the poor are priced out of an expensive city center apartment, they may nonetheless commute in to downtown shopping districts and interact with a broad cross-section of other income groups. On the other hand, both the local (post offices and parks) and priced (expensive restaurants) nature of many third places could deliver exactly the economic segregation seen in other spheres.

In this article, we provide the first national estimates of economic segregation in location-based consumption and daily activity. We do so using geolocated data from SafeGraph, which allows for granular income proxies based on the neighborhoods where people live. A neighborhood's exposure to others is defined by the other people in the stores, restaurants, shops, parks, and libraries that they frequent. McDonald's locations, for example, tend to further isolate the poor because they are most likely to serve poorer customers.

We find that the most isolated Americans are not the poor, but the rich. Households from the top 20% of neighborhoods by income are twice as likely to encounter other high-income people as would be expected by chance. The bottom 20% of neighborhoods is also isolated, but at about half the rate. The isolation of the rich is largely an urban and suburban phenomenon. Rural areas provide less of an enclave for the rich. Instead, rural areas drive up isolation of the poorest residents compared to their urban and suburban counterparts. Isolation is higher for poor neighborhoods that are either majority Black or Hispanic.

Second, we ask how much class segregation reflects high- and low-income households frequenting different industries from others (high-end dining rather than fast food) or staying local to residentially segregated neighborhoods. In a counterefactual reweighting exercise, we find that there is only a small role for industry. High-income residents frequent different types of places—e.g., museums instead of libraries and full-service restaurants as opposed to fast food—but equating industry shares across classes would barely shift levels of isolation. On the other hand, distance matters, and can account for around one third of the isolation we observe. This suggests that activity segregation partially reflects residential segregation. But even adjusting for distance, the majority of activity segregation persists.

Finally, we zoom in on which specific chains and establishments contribute to socio-economic mixing and which exacerbate segregation. Some very poor-serving chains, like dollar stores, contribute to segregation. But, consistent with the importance of distance in the reweighting analysis, so do chains that have many local branches: while residents from all income quintiles shop at distinct CVS, they shop at CVS stores in their own neighborhoods. In contrast to these market segmented or highly local businesses, some chains contribute substantially to socio-economic mixing. Specifically, low-price full-service restaurants are frequented by a diverse range of residents: the rich and poor rub shoulders at Olive Garden and Applebee's. Indeed, the most socio-economically diverse places in America are not public institutions, like schools and parks, but affordable, chain restaurants.

1.1 Related literature

We extend a recent literature that goes beyond standard measures of residential segregation by using data from mobile phones, social networks, and financial transactions to study economic segregation. Athey et al. (2021) use mobile phone data to estimate racial segregation across metropolitan areas. We use a related measure of class segregation. Wang et al. (2018) employs similar data but focuses on mobility and racial segregation in the 50 largest US cities. Reme et al. (2022) study income segregation using income and mobile phone data in Oslo, Norway. Dong et al. (2020) use Twitter data and credit card transactions to show how online segregation recreates residential segregation. Chen and Pope (2020) use SafeGraph data to study mobility—how often and how far people travel—and how it varies across geography and class. Chetty et al. (2022) study connections between Facebook friends, capturing a similar inflection in homophily for the richest people in their data: the highest-income Facebook users are especially isolated.

Our study is the first to provide national estimates of experienced class segregation in daily activities. We also focus on an intuitive measure of isolation: the share of encounters between different groups. We provide the first decomposition quantifying sources of social isolation for the rich and poor. This approach distinguishes the roles played by distance, the types of places visited and specific chains. Finally, we study the types of locations that could promote economic integration, zooming in to even the brand level. The scale of our data provide a unique level of granularity for measuring the role of specific brands in promoting or curbing economic integration.

2 Data

We use SafeGraph Patterns data for the months December 2021 to July 2022. The SafeGraph data pools mobile location information from different sources to estimate the number of visitors to different places of interest. They also determine where each user lives based on where the device is for most of the night. To protect anonymity, data is not provided at the individual level.

Instead, the key measure we study is the count of visitors from a specific home census block group (CBG) to a specific point of interest in a certain month, restricted to cases where a minimum of two devices from the same CBG were present. This count, scaled by the inverse sampling probability of the CBG, provides an estimate of the number of people who visited that particular place from that neighborhood. CBGs are small: the average population is 1,483. The 10th percentile of population is 680 and the 90th percentile is 2,490.

In Table 1, we show the industry composition of visits, grouping industries following the categorizations in Appendix C. The largest category is essential retail, which accounts for 7.2 billion visits in the data and 22% of all visits. The top two categories, essential and non-essential retail, account for 37% of the visits in our data. The next two most common categories are full-service and limited-service dining, followed by entertainment. The data thus covers the places where people shop, eat, and socialize.

Note that some of these visits will be accounted for by the people working at those establishments. In this way, employment segregation between classes can appear in our data as consumption segregation. However, in the monthly data, visitors are counted as unique monthly visitors. As such, for most businesses, the small number of employees, although they make daily, frequent visits, will be swamped by the much larger number of unique customers who visit each month.

3 Methodology

Our core concept is exposure: the chance of encountering someone in a certain group, conditional on membership in a certain group. An encounter is when someone is at a place at the same time as someone else (Athey et al., 2021). This is based on the monthly aggregate visits, so it will misstate exposure if people from different income groups go to the same establishment at different times or on different days. We show in Figure A.2 using the American Time Use Survey (Flood et al., 2022) that, reassuringly, people tend to eat at similar times regardless of income.

3.1 Defining exposure

Let $\mathcal{I}$ represent the set of all CBGs, and $Visitors_{k(G)}$ be the number of visitors at establishment k from group G. Then in a given period, the exposure of some CBG $i \in \mathcal{I}$ to members of group G at establishment k is given by:

$$FirmExposure_i(G,k) = \frac{Visitors_{k(G_{-i})}}{\sum_{\{m \in I: m \neq i\}} Visitors_{k(m)}}. \tag{1}$$

In words, their exposure is the share of other visitors to establishment k who are from group G (not counting the focal CBG i).

These values are averaged together with weights given by i's the total visits to make our core exposure measures at the income quintile level. Income quintiles are defined by the median income in someone's home CBG. For example, let $Q1$ denote all CBGs in the bottom income quintile. Then the exposure of the top quintile, $Q5$, to members of the bottom quintile, $Q1$, is given by:

$$Exposure(Q1, Q5) = \frac{\sum_{i \in Q5} \sum_k FirmExposure_i(Q1,k) * Visitors_{k(i)}}{\sum_{i \in Q5} \sum_k Visitors_{k(i)}}. \tag{2}$$

In the numerator, the product gives the exposure of $Q5$ CBGs to residents of $Q1$ CBGs, $FirmExposure_i(Q1,k)$, multiplied by the total number of visitors to firm k from CBG i. This is summed over all firms k, and then over all CBGs in $Q5$, indexed by i. In the denominator, we sum total visits by residents of $Q5$ CBGs.

This is the average exposure that $Q5$ residents experience to $Q1$ residents, weighted by their number of visits. We mainly measure exposure across quintiles. For example, if Q5 to Q1 exposure, $Exposure(Q1, Q5)$ is 0.15, it means the average visit by a Q5 resident is to a place with 15% of visitors from Q1. With perfect mixing, each income quintile would have a 20% chance of encountering someone from another quintile.

The measure is most similar to the one employed by Athey et al. (2021), and related to standard dissimilarity indexes. A slight difference is that Athey et al. (2021) focus instead on $Exposure(Q1, Q5) - Exposure(Q5, Q5)$. One benefit of our measure, which is similar to the core metric of cross-class friendships in Chetty et al. (2022), is that it directly translates into the chance of an interaction with someone. For example, if people tend to speak with 2% of the visitors at the places they visit, then in general, if $Exposure(g, f) = 0.50$, we expect that people from group f have a 1% chance of interacting with a member from group g in their typical outing.

3.2 Decomposing isolation

To study the drivers of isolation, we perform a simple counterfactual exercise in the style of DiNardo et al. (1995). We reweight the data so that visitors are equally likely to travel to different kinds of places, whether that be industries, distances from their home, or brands. To see how, note that we can always partition the visits into groups of firms. Exposure at firms of type A is

$$Exposure_A(Q1, Q5) = \frac{\sum_{i \in Q5} \sum_{k \in A} FirmExposure_i(Q1, k) * Visitors_{k(i)}}{\sum_{i \in Q5} \sum_{k \in A} Visitors_{k(i)}} \tag{3}$$

where the only change to Equation 2 is that the firm sum is only over firms in A rather than all firms ($\sum_{k \in A}$), and the subscript A in $Exposure_A(Q1, Q5)$. If $w_A(g)$ is the share of visits that group g members have at firms of type A, then the overall exposure of $Q5$ to $Q1$ can be written as:

$$Exposure(Q1, Q5) = w_A(Q5) Exposure_A(Q1, Q5) + w_B(Q5) Exposure_B(Q1, Q5). \tag{4}$$

We can think of $\{A, B\}$ as capturing some important firm distinction, for example $A =$ chains and $B =$ non-chains. Then, in this framework, we can assess the importance of chain vs. non-chain visits by setting $\{w_A(g), w_B(g)\}$ to $\{\overline{w}_A, \overline{w}_B\}$ for all groups g.

4 Exposure to different income groups

How much do Americans of different income levels mix with one another? Who is exposed to a broad cross-section of income levels, and who is disproportionately exposed to others like themselves? In Figure 1 we provide the first ever national estimates of class segregation by activity patterns. On the x-axis, we array quintiles of neighborhoods by average income. Higher income neighborhoods are at the right and low income neighborhoods are at the left. We then study the establishments that residents of these different neighborhoods patronize. For example, residents of high-income neighborhoods go to Whole Foods more than Dollar General, while the reverse is true for residents of low income neighborhoods.

Specifically, for each establishment in the US, we calculate the share of visitors from each income quintile. A Whole Foods next to Golden Gate Park in San Francisco, CA receives 62% of its visitors from the top quintile of neighborhoods and only 3% from the bottom quintile. In contrast, at a Dollar General in Macon, GA, 40% of visitors are from the bottom income quintile. These establishment-level income distributions capture the extent of income segregation within the millions of places that people shop, work, relax, access government services, and socialize outside of the home.

We then average these establishment-level income distributions for each residential neighborhood, weighting by the number of visitors each establishment receives from the focal neighborhood. We remove the focal neighborhood from the establishment-level calculation, taking the leave-one-out average. This gives, for each neighborhood, a measure of the average exposure its residents have, via the places they frequent, to people from neighborhoods of different income levels.

In some neighborhoods, residents frequent establishments in which they see and meet others who are disproportionately from neighborhoods like their own. In other neighborhoods, residents frequent meccas like chain stores and parks (as we discuss below), that draw visitors from a widely representative swath of other neighborhoods. Our neighborhood-level exposure estimates quantify these differences in exposure.

In the columns of Figure 1, we average across neighborhood-level exposure estimates to show each neighborhood income quintile's exposure to visitors from different income quintiles. The antidiagonal of the matrix shows how exposed each quintile is to visitors like themselves. The bottom row shows how exposed each quintile is to visitors from low-income neighborhoods; the top row shows exposure to high-income Americans. The antidiagnoal shows that, in general, each quintile is more exposed to visitors from neighborhoods of a similar income level (about a quarter of exposed visitors, rather than a fifth). The values drop as you move away toward the top left and bottom right corners: the least likely encounters are those between the top and bottom income quintiles. The two cells at these corners are not equal because the denominator is the total number of visits by the x-axis income group: the rich visit more establishments, so they can represent a higher share of the encounters of the poor and the poor do for them.

High income neighborhoods are substantially more exposed to visitors like themselves: 41% of visitors they encounter are themselves residents of high income neighborhoods. This is around 60% more isolated than any other income quintile. While Americans in the middle three quintiles are more evenly exposed to other income groups, residents of the highest income neighborhoods are disproportionately exposed to others like themselves. This suggests that establishments frequented by the rich are mainly composed of the rich, whereas the places that residents of middle income neighborhoods patronize are a more representative mix of income levels.

Residents of low income neighborhoods, often characterized as socially isolated, are also disproportionately self-exposed, albeit less than the rich. Only 11% of the visitors encountered by people from low income neighborhoods are from high income neighborhoods. In contrast, middle income neighborhoods are exposed to a fairly representative sample of other Americans. If high-income Americans estimated the size of the quintiles in the income distribution based on their face-to-face interactions outside of the home, they would overestimate the share of other high-income people. Middle income Americans would get their estimates roughly correct.

These results use the median income in an individual's home CBG as a proxy for their class. To what extent does this reflect the isolation of individuals? We explore this question in Appendix E. We show that our estimates of isolation are likely biased downward due to the ecological proxies we use, since places drawing rich people in general will also tend to draw the rich people from poorer neighborhoods. Next, we present two robustness checks exploiting detailed income bins from the American Community Survey. Results are identical assuming that each visitor is a random draw from their home CBG's income distribution. Further, we see similar patterns of isolation when we zoom in on highly homogeneous CBGs, with 75% of households belonging to the specified income group.

Given these findings, we next assess whether social isolation of the rich and poor hold for third places specifically; whether encounters measured in our colocation data are likely to contribute to real network ties; how urban, suburban and rural areas vary in isolation; and whether class segregation simply reflects racial segregation. We then turn to a counterfactual reweighting exercise to assess the role of distance, industry, and specific chains in the patterns we observe.

4.1 Connection with third places and friendship

Does our measure of encounters track real-world interactions? Two pieces of evidence bolster the main finding: first, we get the same results restricting to "third places" where people are more likely to interact. Second, our measure of exposure from daily activities is highly correlated with other granular data on cross-class friendships from Chetty et al. (2022).

Third places As shown in Table 1, the largest industry category is essential retail, which means much of our estimates could be driven by places like CVS, the largest pharmacy chain in the country, or 7-Eleven, a convenience store with around 10,000 locations as of 2023. These places are arguably impersonal and perhaps less likely to generate interactions between conversations compared to churches, for example.

As a first check on these results, we perform the same exposure calculations restricting to "third places" (Oldenburg, 1999). This subsample includes cafes, churches, gyms, civic organizations, beauty parlours, bars, libraries, bookstores, parks, and fast-food establishments. We show the results in Figure A.1. We find that exposure within potential third places matches the exposure documented in Figure 1, particularly for the top and bottom quintile. Even restricting to places that are more conducive to conversation and socialization, our data shows just as much isolation.

Friendships Another check on these results is to ask whether the encounters measured in our data are correlated with friendships. Chetty et al. (2022) provide ZIP code-level measures of cross-class friendships based on Facebook data. Specifically, their estimate is defined as "two times the share of high-SES [socioeconomic status] friends among low-SES individuals, averaged over all low-SES individuals in the ZIP code," where high-SES and low-SES are defined as being above or below median, respectively.

Our estimates of cross-class interactions are highly associated with this friendship measure. We show this in Figure 3(a). Since the Chetty et al. (2022) is based on the friendships that lower-income individuals have, we focus on interactions of the bottom quintile CBGs. The measure from our data tracks the exposure of low-income individuals to the rich. The plot shows that in ZIP codes where poor people encounter a higher share of rich people in their daily activities (x-axis), there's a higher degree of cross-class friendships as measured in the Facebook data (y-axis). This confirms that our activity-based measures are tied to another real-world outcome of social connectedness.

Residents of neighborhoods with more friendships across class lines might be expected to frequent more third places that foster interaction. We test this in Figure 3(b). The outcomes shown in the y-axes measure the share of visits to "third places" as defined in the previous section. We show results with and without fast food because it is such a large share of overall visits. The x-axis is deciles of the cross-class friendship measure from Chetty et al. (2022). The results show clearly that people in areas with more of these friendships tend to spend more time at the types of locations identified by Oldenburg (1999). ZIP codes in the top decile of the friendship measure have a 60-90% higher share of visits to third places compared to those in the bottom decile.

Some of this association could follow naturally from the income distribution in that particular area. For example, if a ZIP code consists mainly of high income people, the poorer people in the ZIP code will have more encounters with the rich and more friendships with the rich. Social connectedness would be higher, but due merely to the income of people who live there. As we show in Appendix D, however, these associations are robust to two more stringent tests. First, we add fine income controls at the zipcode level. Next, we swap out the actual firm-level exposure to what would be predicted only by the mix of chains operating in that zipcode, leaving out all the focal zipcode firms in the calculation. In both cases, cross-class friendships (Chetty et al., 2022) and cross-class interactions remain closely linked.

4.2 Urban, Suburban and Rural Isolation

The isolation of the rich is largely an urban and suburban phenomenon. In Figure 2(a), we show the anti-diagonal values from Figure 1: the exposure of each income quintile to people from the same income group. Except we split the CBGs into one of three county-level urban-rural designations.

For the most part, urban and suburban isolation move in lockstep at around 30% for the bottom quintile and 40% for the top quintile. The relationship is starkly different for rural income quintiles. Residents of poor CBGs in rural areas are the most likely to encounter people like themselves. In contrast, the richest rural areas see practically no isolation. The estimate of isolation for the top quintile group is just over 20%, about what would be expected with random mixing. Overall, urban and suburban areas deepen the isolation of the high-income groups and decrease isolation of low-income groups.

4.3 Segregation by race

To what extent do our main findings reflect racial segregation? In Figure 2(b), we perform the same exercise split out the series into two groups based on racial composition: CBGs that are majority white and those that are majority non-white (this means some falling in neither category are excluded).

The rightmost dots show that residents of rich white and non-white CBGs are similarly likely to encounter other rich people. However, majority white CBGs in income quintiles 2 and 3 are relatively more isolated. Finally, minority neighborhoods in the poorest quintile are distinctly more isolated from other income groups.

These results align with the main findings of Wang et al. (2018), who find using Twitter data that non-white neighborhoods are less likely to come into contact with people from richer areas. There are a few key distinctions: their outcome is the number of visits to white, non-poor neighborhoods, so exposure is measured using geography. In contrast, exposure in our context is defined by the people who also visited the same establishment. Second, they use Twitter data, which could be a more selected sample and presents more challenges for estimating home residence. Finally, while they study the 50 largest cities, our data are more comprehensive, covering all but a few counties in the United States.

5 Decomposing sources of segregation

5.1 Industry

Next, we use simple counterfactuals to uncover what aspect of activity drives the segregation of the rich and poor. One candidate is industry. Our data shows that income levels determines the kinds of establishments that people visit.

We illustrate this Figure 4(a), which shows the share of visits to broad income categories.[1] (The categories are explained in Appendix C.) For example, the table shows that visits by people from the top income quintile are composed more of full-service dining establishments (16%) compared to visits by the bottom quintile (13%). And the top quintile spends a smaller share of time at essential retail (e.g., drug stores and gas stations), which accounts for 17% or their visits compared to 25% in the bottom quintile.

These patterns are starker if we zoom in on specific industries. Figure 4(b) shows that some industries cater specifically to low- or high-income individuals. This plot shows the 5- or 6-digit NAICS codes with at least 5 million visitors that represent a much larger share of visits for the top quintile compared to the bottom quintile, or vice versa. The industry that accounts for the most disproportionate share of low-income visits is general merchandise stores, which is dominated by dollar store chains like Dollar Tree and Dollar General. Correctional institutions, wireless carriers, and credit unions are also a much larger share of bottom quintile visits. In contrast, visits to golf courses, country clubs, fitness centers, breweries, and wineries are up to a 3 times higher share of top quintile visits.

How much of the isolation of the rich is due to the fact that residents of high income neighborhoods patronize different industries? We assess this by reweighting so that all income groups go to the the same industries in similar proportions, as described in Section 3.2. Each industry is one of the 453 unique NAICS codes in the data. This measures the class mixing that would occur if all groups

[1] The exact shares are reported in Table A.1.

went to fast food restaurants, jewelry stores, and post offices in equal proportion.

The results in Figure 5 show that the industry mix has only a small influence on the self-exposure of different classes. The solid black line, showing the original result from Figure 1, and the dashed blue line, which incorporates the industry reweighting, are nearly identical. People continue to segregate within industry.

This suggests that industry does not contribute to the isolation of the bottom and top quintiles. A simple way to see why is that the sharp class divides are replicated in the two largest industry categories: essential and non-essential retail. These account for 37% of the visits in the US. In both industries, people segregate across stores in a way that reproduces class divisions. Top quintile consumers go to retail stores where 42% of visitors are from the same income group. In other words, when people go to drug stores, they go to locations that mirror their own attributes, recreating the patterns in Figure 1. Despite some stark patterns in Figure 4, industry matters little in comparison to the specific establishment.

5.2 Geography

One obvious sub-industry attribute is distance. To what extent does activity segregation simply reflect the fact that residences are geographically separated? For a high-income person, the closest drug store will probably be in a high-income neighborhood. And Chen and Pope (2020) show that richer individuals travel farther, which could take them to more exclusive establishments.

The two panels in Figure 6 explore the relationship between isolation and distance. The gray line in Figure 6(a) shows the relationship between distance and the share of total visits, averaged across all income groups. People are most likely to go to places that are five miles away, and 80% of visits are within ten miles of home. Next, the dashed lines in Figure 6(a) show the relationship between isolation and distance for the residents of high- and low-income neighborhoods. For both groups, isolation is highest close to home. This decreases sharply as distance increases, but stays above the random-mixing benchmark of 0.20. Even at 18-19 miles from home, where just a small fraction of visits occur, the high-income group is much more likely to encounter similar people. This implies that proximity can explain some, but not all, of the isolation that we observe in the data.

To quantify this descriptive finding, we perform a counterfactual exercise where were we assume that people are equally likely to visits places close and far from them, using five-mile bins and restricting to distances under 50 miles.[2] In our reweighting, we assume that people are equally likely to

[2]This accounts for 95% of visits in the data and replicates the main isolation results. The reweighting results are not

travel any of the five-mile distances. This essentially swaps out the observed gray line in Figure 6(a) for a flat one. The results, shown in Figure 6(b), suggest that distance does account for a portion of the observed segregation.

We quantify the change in isolation as the decrease in the excess self-exposure accomplished in the reweighting. The top quintile has 41% of their encounters with other top quintile residents (Figure 1). The reweighting decreases this to 35%, so this decreases the excess self-exposure by about 28%. The change in isolation for the bottom quintile is smaller, as foreshadowed by the shallower slope in Figure 6(a), achieving a decrease of about 18%.

5.3 Chains

We next turn to the role that particular companies play in increasing or decreasing isolation. 62 percent of visits in our data are to chain establishments with at least two locations, and 50 percent are to establishments with at least ten locations. Overall, independent (non-chain) locations are slightly more isolating for higher-income individuals: these people have 39.5% top quintile encounters at chains compared to 42.0% at non-chain establishments.

The analysis of the Chetty et al. (2022) data in Section 4.1 showed that the mix of chains at a location can predict its level of cross-class friendships. This means that something stable about brands might affect the kinds of interactions that happen in the locations where they operate. What would happen if people visited the country's chains in equal proportion, regardless of income?

We explore this counterfactual in Figure 7. We show the main isolation estimates from Figure 1 in the solid black line. Next, since this analysis necessarily restricts to multi-location places, we show in the dashed blue line what happens to isolation when considering chains only. This captures the fact that isolation for the rich is slightly lower at multi-location firms, but slightly higher for lower-income groups. (As a group, chains serve a higher share of low-income people, something we explore in the next section.)

We next reweight the income quintile-by-firm observations so that each income group visits every chain equally. For example, 2.3% of visits in our data are to McDonald's. In this reweighting, all income groups find themselves at McDonald's 2.3% of the time. Since high-income people are underrepresented at McDonald's, this upweights their visits there. Importantly, however, the mix of encounters experienced within McDonald's locations still matches the tendencies of each income group: the upweighting is applied to the typical exposure the particular group gets at McDonald's

sensitive to this cutoff.

(see Section 3.2).

For instance, McDonald's has a location in Danville, California, a wealthy suburb of Oakland, with predominantly high-income visitors. The typical McDonald's experience of a top-quintile resident will differ compared to other income groups because they are more likely to visit the McDonald's in Danville, and those in other wealthy areas. These effects allow for the high- and low-income visitors to remain isolated while visiting the same brand.

The red dotted line shows the results. We find that reweighting by chains alone explains a portion isolation, decreasing it by 2pp for the bottom quintile and 4pp for the top quintile of income relative to the chains-only isolation estimated in the blue dashed line. Compared to an equal mixing benchmark of 20%, this reduces the excess isolation of high-income visitors by 20%. This suggests a meaningful role for chains in explaining class isolation. Still, the vertical gap between the red dotted line and the 20% benchmark at the bottom of the plot suggests, perhaps surprisingly, that most of the isolation driven by chains comes not from brand but from the particular locations that people visit.

One way to carve out a role for these location effects is to categorize places based on proximity to home. This is a useful proxy because it will reflect broad patterns of residential segregation by class, as people tend to patronize the places closest to home. As mentioned above, some high-income residents of California are more likely to visit the McDonald's in the wealthy enclave of Danville.

How much isolation is explained by these combining these chain and distance effects? In a final step, we combine the reweighting exercises from the previous two sections, asking whether the combination of brand and distance explains the bulk of isolation. To do so, we capture the average exposure to other income groups at the (i) income quintile by (ii) firm by (iii) distance level. Instead of giving an equal weight at the firm level, we apply equal weights at the firm-by-distance level. Thus, to continue with the McDonald's example, each group is assumed to visit McDonald's equally; and to spend the same amount of time at nearby and faraway McDonald's locations. We use six distance bins with five-mile increments. This discretization is needed to have sufficient representation each income-by-firm-by-distance cell.

Results are in the gold dashed line in Figure 7. Flattening distance within chain has a large impact on the isolation of the top quintile, reducing it nearly to the random benchmark of 20%. In contrast, this interaction is less powerful for reducing the isolation of the bottom quintile. This reflects the fact, also captured in Figure 6, that distance tends to have a larger impact on high-income integration. And it shows that this difference is driven by multi-location establishments. (The adjustment matters less for non-chain establishments because the far-away places that high-income groups visit are more likely to be isolating such as golf courses and resorts.)

Together, these findings suggest that brand and distance go a long way toward explaining the isolation of the rich. Low-income residents, however, tend to travel along segregated corridors.

6 Which places increase socio-economic mixing?

Given the importance of chain firms to the isolation of high- and low-income neighborhood residents, we next identify the specific industries and chains that contribute to and offset isolation. Our establishment-level data allows us to track particular chains and provides the first assessment of particular companies' contribution to socio-economic mixing.

Different chains and types of establishments yield different levels of mixing. We show a handful of examples in Figure 8. For example, isolation tends to be higher in churches (panel a) than when calculated using all visits. The same is true of libraries (panel b). When people go to Starbucks locations (panel c), the level of isolation depends heavily on their own income quintile. The bottom three quintiles encounter almost a uniform distribution of income quintiles. Starbucks tends to attract people from richer CBGs, so it serves to integrate poeple from the lower income groups. On the other hand, top quintile visitors are slightly more isolated at Starbucks than overall. Dollar General serves plays the opposite role of Starbucks, tending to isolate the lower income groups and integrate the high income groups.

Removal exercise To quantify this contribution for each industry, we study how class isolation changes when a single chain is industry. Specifically, we again calculate exposure to either richer (for poor neighborhoods) or poorer (for rich neighborhoods) industry, except we remove all establishments in the referenced chain. The ensuing two measures reflect both the extent to which an industry mixes socio-economically diverse visitors and also the size of the chain in terms of visitors (and therefore how important it is to overall activity segregation levels). A small industry that has substantial class mixing will contribute little to the overall level of mixing in the data.

Note that our two measures of isolation—high-income exposure to lower income and low-income exposure to higher income—are subtly different. If a high-income resident visits a store that is otherwise composed entirely of poor visitors, that visit will increase the high-income resident's exposure to the poor. However, if a poor person visited that same store, it would increase her isolation, as, notwithstanding that single high-income visitor, all of the other visitors are poor. As such, we average the two isolation measures together, to index mixing simultaneously from the high- and low-income perspective.

6.1 Industries

How do different industries affect mixing? Figure 9, we show the impact of the 37 largest industries in the data. As described above, these estimates are based on calculating mixing and exposure if all visits to a particular industry were removed.

The two industries that contribute the most to mixing across class lines are limited-service and full-service restaurants. Part of this is due to their scale. But zooming in on the breakdown provided in Figure 9 shows that they impact class mixing through opposite channels. The existence of full-service restaurants makes integration of bottom quintile residents 0.71% higher (blue triangle in the bottom row). For rich residents, however, full-service restaurants have essentially no impact on mixing (red diamond in the bottom row). So full-service restaurants serve to increase mixing by integrating lower-income groups.

On the other hand, limited-service restaurants (e.g., McDonald's) tend to slightly decrease mixing for the lower-income groups (2nd row from the bottom, blue triangle). But these establishments have the highest impact on the integration of higher-income groups apart from gas stations.

In the third row from the top is supermarkets, which tend to increase isolation for both the rich and the poor. Schools, in the very top row, increase isolation for both groups, but far more so for the rich. These industries are all typically local, and their segregating role likely reflects in part residential segregation. Fitness and recreation and religious institutions are likewise highly class-segregated. Gyms substantially increase isolation for the rich. Churches, notwithstanding their oft-cited role in contributing to social capital, isolate both the poor and the rich.

These consistent differences across industry show that not only do high- and low-income people frequent different industries, but even within those industries some mix and some segregate across classes.

6.2 Chains

We next repeat the analysis for companies, rather than industries. This allows us to drill down into specific chains that contribute to class mixing.

Figure 10 shows the results. Several cheap, poor-serving chains reduce mixing, like Family Dollar and Dollar General. These reduce mixing specifically by isolating the poor. Chains with many local, neighborhood branches, like Walgreens and CVS, likewise reduce mixing, but do so because they isolate the rich in their own neighborhoods. In contrast Starbucks reduces average mixing entirely by reducing mixing among the rich; it actually contributes substantially to exposing the poor to non-poor visitors.

At the other end, some chains disproportionately contribute to socio-economic mixing. McDonald's and Wendy's add to mixing overall, because they do so much to expose top quintile visitors to others. However, they actually exacerbate the isolation of the poor. In contrast, several low-price, full-service restaurants, like Olive Garden, Applebee's, Chili's and IHOP contribute to mixing for both poor and rich visitors.

This exercise demonstrates that the chains that most contribute to mixing among the rich (often by exposing them to many poor visitors) are often different from those that expose the poor to non-poor (often rich) visitors. This suggests that decreasing isolation among the poor and isolation among the rich may require distinct planning and policy actions. However, we also identify a genre of chains that consistently mixes both rich and poor: full-service, low-price restaurants.

7 Policy-relevant locations

We next turn to the role of government in fostering places that serve specific groups. Several government spaces are primarily used by physically visiting them. A prominent example is parks. Los Angeles, Chicago, and Minneapolis spend 3-5 percent of their municipal budgets on public parks (Barron, 2023). And US states spend between $13 and $108 per capita on public libraries (Ebdon et al., 2019). Museums, historical preservation cites, and hospitals also account for line items in city and state budgets. Transaction costs and information frictions could mean that these expenditures are less redistributive than might be intended (e.g., Finkelstein and Notowidigdo, 2019).

Separate from government-run facilities, municipalities in the US also use zoning policies to promote economic development, diversity, and other community goals. Notably, some of this has involved limiting the operation of chains to promote independent businesses. For example, San Francisco and New York City have both enacted limits on chain stores to protect the local business landscape (SF Planning, 2023; Bobrowski, 2012; NYC Planning, 2012). Other laws make it harder for bars, liquor stores, and casinos to gain operating licences.

The way that a business mix affects a community has long been a focus of sociologists and city planners (e.g. Jacobs, 1961; Zukin et al., 2009). Within the US, the Department of Housing and Urban Development has for decades encouraged mixed-income communities as a way to decrease economic isolation and concentrated poverty (Smith, 2002). One arguably critical input into discussions around redistribution and gentrification might be information on who visits these places. This has been historically challenging. But the findings here can provide key data for studying the consequences of government policies and evolving commercial landscapes: information on the people who work, shop at, or otherwise visit each location.

In Figure 11, we provide an illustration. Instead of showing the absolute share of visitors from each location, we demean each location relative to its city average. We use within-city measures to focus on distribution across city residents. If the libraries in a largely wealthy municipality tend to serve the relative low-income groups, this should count as redistributive even if the city's bottom quartile has high socioeconomic status based on the national distribution.

Figure 11(a) shows the fraction of visitors from the bottom quintile, relative to the mean within the city. Positive values mean that a relatively higher share of low-income visitors appear at those locations. Churches (last row) draw the most-low income visitors than any other location. Their visitors are 1.4pp more likely to come from the bottom quintile relative to other places in the city. Libraries and parks, in this analysis, are redistributive in that the typical visitor comes from lower on the income distribution compared to the average visitor in the city. Interestingly, fast food and large chains (those with more than 500 locations) play a similar role in serving the lower-income residents of a city. The typical non-chain or independent establishment is slightly less likely to draw bottom-quintile visitors. And museums contrast with libraries in attracting richer visitors overall.

A related goal in discussions of gentrification and mixed-income neighborhoods is retaining business that service racially diverse groups. In Figure 11(b) we perform the same analysis except ranking places in terms of the share of their visitors from majority non-white census block groups (CBGs). These findings largely echo panel (a). For example, churches and large chains both attract more visitors from non-white neighborhoods compared to other locations in the city. However, libraries are now slightly less likely to garner visitors compared to panel (a).

8 Conclusion

Class segregation appears not just in where people live (Reardon and Bischoff, 2011) and work (Song et al., 2019), but also in the public and commercial places where people spend time and money. In this paper, we establish that residents of low- and especially high-income neighborhoods are exposed disproportionately to others like themselves.

In a counterfactual analysis, we show that this is not largely because of broad, industry-level differences in consumption patterns. Even within industry, the rich and poor end up in different places. This is partially because many frequently visited places are highly local, like drugstores, post offices, schools and parks. In context of residential segregation, these highly local establishments facilitate very little mixing.

Alongside these local and distance effects, some large chains tend to segregate and some tend to mix people. Specifically, we find that restaurants tend to mix people. Fast food places, like McDonald's and Wendy's, tend to mix people by substantially exposing rich residents to the non-rich. However, these places tend to do little to exposure the poor to the non-poor, as they predominately serve a relatively low-income clientele. In contrast, cheap, full service brands, like Olive Garden and Applebee's, increase mixing for both the rich and the poor.

Our results demonstrate that the places that contribute most to mixing by economic class are not civic spaces like churches or schools, but large, affordable chain restaurants and stores. Insofar as policy makers seek to increase exposure between different classes, they should pay attention to the role of firms in shaping class mixing.

References

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt, "Estimating experienced racial segregation in US cities using large-scale GPS data," *Proceedings of the National Academy of Sciences*, 2021, *118* (46), e2026160118.

Barron, James, "One Percent of the Budget for Parks? A Bargain, Says a Nonprofit," *New York Times*, March 2023, *22.* https://www.nytimes.com/2023/03/13/nyregion/one-percent-of-the-budget-for-parks-a-bargain-says-a-nonprofit.html.

Bobrowski, Mark, "The regulation of formula businesses and the dormant commerce clause doctrine," *Urb. Law.*, 2012, *44*, 227.

Chen, M Keith and Devin G Pope, "Geographic mobility in America: Evidence from cell phone data," Technical Report, National Bureau of Economic Research 2020.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan, "Income segregation and intergenerational mobility across colleges in the United States," *The Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

_ , Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob et al., "Social capital I: measurement and associations with economic mobility," *Nature*, 2022, *608* (7921), 108–121.

Cho, Wendy K Tam and Charles F Manski, "Cross-Level/Ecological Inference," in "The Oxford Handbook of Political Methodology," Oxford University Press, 2008.

DiNardo, John, Nicole Fortin, and Thomas Lemieux, "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach," 1995.

Dong, Xiaowen, Alfredo J Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland, "Segregated interactions in urban and online space," *EPJ Data Science*, 2020, *9* (1), 20.

Ebdon, Carol, Ji Hyung Park, Aimee L Franklin, and Jonathan Moore, "Institutional arrangements for public library funding and spending," *The American Review of Public Administration*, 2019, *49* (5), 532–544.

Finkelstein, Amy and Matthew J Notowidigdo, "Take-up and targeting: Experimental evidence from SNAP," *The Quarterly Journal of Economics*, 2019, *134* (3), 1505–1556.

Flood, Sarah M., Liana C. Sayer, and Daniel Backman, "American Time Use Survey Data Extract Builder: Version 3.1 [dataset]," *College Park, MD: University of Maryland and Minneapolis, MN: IPUMS*, 2022. https://doi.org/10.18128/D060.V3.1 (Accessed in 2023).

Gethin, Amory, Clara Martínez-Toledano, and Thomas Piketty, "Brahmin left versus merchant right: Changing political cleavages in 21 Western Democracies, 1948–2020," *The Quarterly Journal of Economics*, 2022, *137* (1), 1–48.

Jacobs, Jane, *The death and life of great American cities*, Vintage, 1961.

NYC Planning, "Projects & Proposals > Manhattan > Upper West Side: Special Enhanced Commercial District Upper West Side Neighborhood Retail Streets - Overview," 2012. https://web.archive.org/web/20230204063058/https://www.nyc.gov/assets/planning/download/pdf/plans/upper-west-side-neighborhood-retail-streets/uws.pdf.

Oldenburg, Ray, *The great good place: Cafes, coffee shops, bookstores, bars, hair salons, and other hangouts at the heart of a community*, Da Capo Press, 1999.

Piantadosi, Steven, David P Byar, and Sylvan B Green, "The ecological fallacy," *American journal of epidemiology*, 1988, *127* (5), 893–904.

Reardon, Sean F and Kendra Bischoff, "Income inequality and income segregation," *American journal of sociology*, 2011, *116* (4), 1092–1153.

Reme, Bjørn-Atle, Andreas Kotsadam, Johannes Bjelland, Pål Roe Sundsøy, and Jo Thori Lind, "Quantifying social segregation in large-scale networks," *Scientific reports*, 2022, *12* (1), 1–7.

SF Planning, "Policy Basis For Formula Retail (Chain Stores)," 2023. https://web.archive.org/web/20230401085025/https://sfplanning.org/project/policy-basis-formula-retail-chain-stores.

Smith, Alastair, *Mixed-income housing developments: Promise and reality*, Joint Center for Housing Studies, Graduate School of Design [and] John F . . . , 2002.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter, "Firming up inequality," *The Quarterly journal of economics*, 2019, *134* (1), 1–50.

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

Zukin, Sharon, Valerie Trujillo, Peter Frase, Danielle Jackson, Tim Recuber, and Abraham Walker, "New retail capital and neighborhood change: Boutiques and gentrification in New York City," *City & Community*, 2009, *8* (1), 47–64.

9 Figures

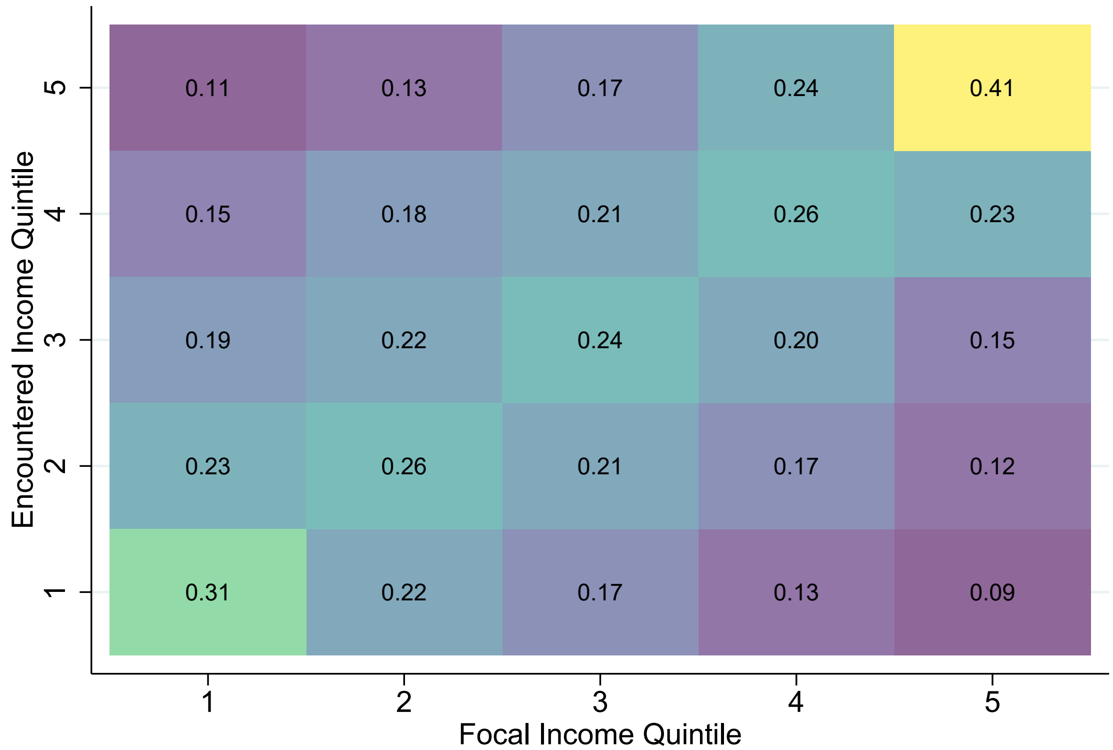
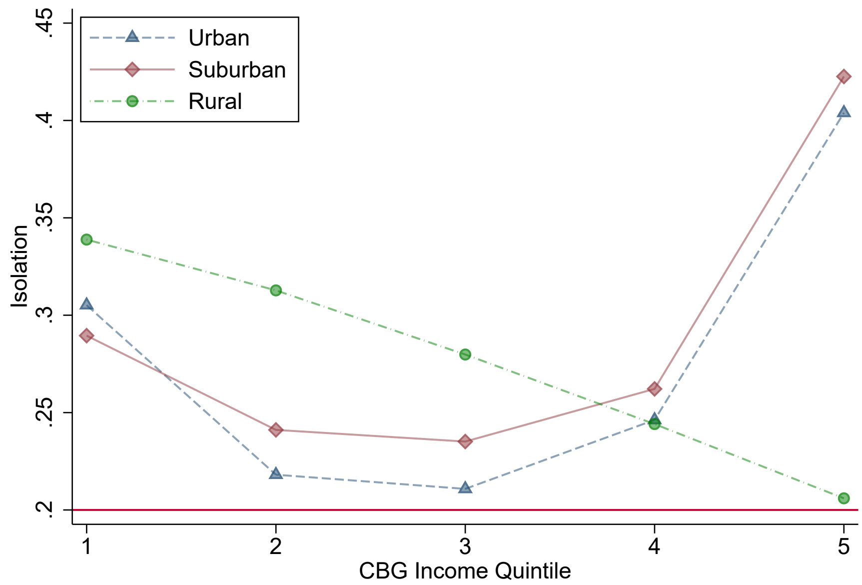
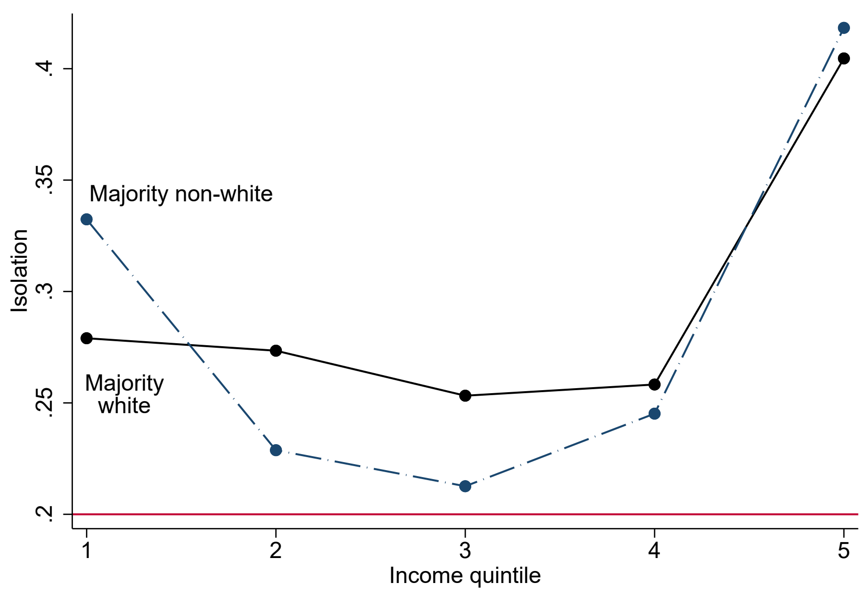


Figure 1: Exposure to different income groups, by quintile

Notes: Each cell gives the chance that a member of the x-axis quintile encounters someone from the group indicated in the y-axis in their average visit. So columns sum to 1, while rows do not. Example: 9% of top quintile encounters are with bottom quintile (bottom right cell). But 11% of bottom quintile exposures are with top quntile (top left). The level of isolation of each group is captured in the anti-diagonal: the exposure of an income quintile has with members of the same group. This is 31% for the bottom quintile (bottom left) and 41% for the top quintile (top right). (Back to section)



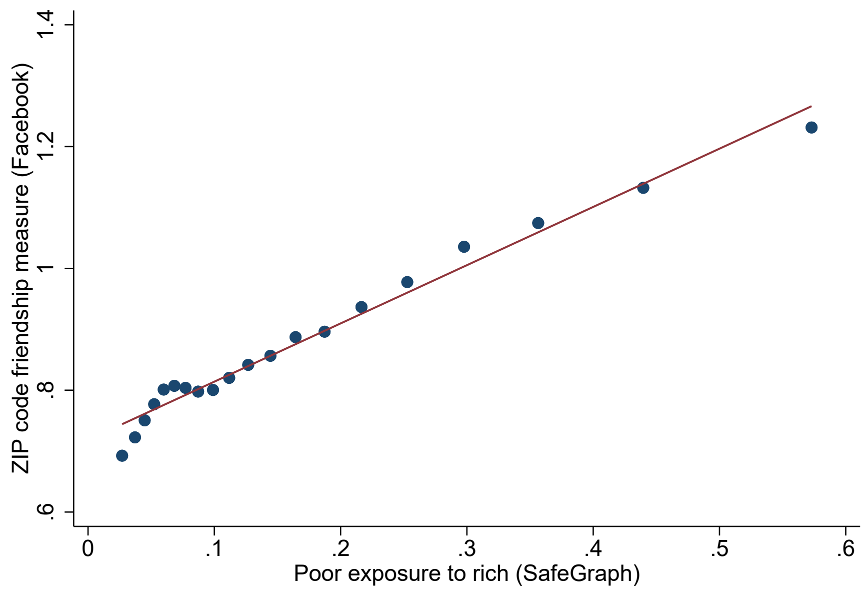
(a) Split by urban-rural designation



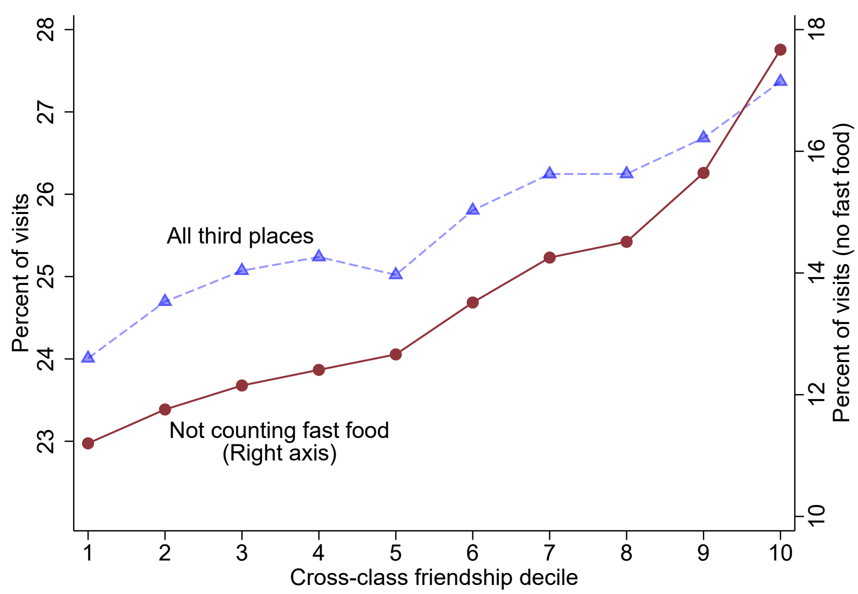
(b) Split by majority CBG race

Figure 2: Exposure to different income groups, split by urban-rural designation and race

Notes: Panel (a): This plot shows the level of isolation (the anti-diagonal from Figure 1) split out by three types of census block groups: those in urban, suburban, and rural counties. Rich isolation is driven by urban and suburban areas. The rural rich are not isolated at all, while the rural poor are the most isolated from other income groups. Panel (b): This plot shows the level of isolation (the anti-diagonal from Figure 1) split out by two types of census block groups: those with majority white residents (N=144,860 CBGs) and those with majority non-white residents (N=66,168). Rich isolation is similar between the two groups (rightmost dots), but non-white low-income areas are especially isolated (leftmost dots). (Back to section)



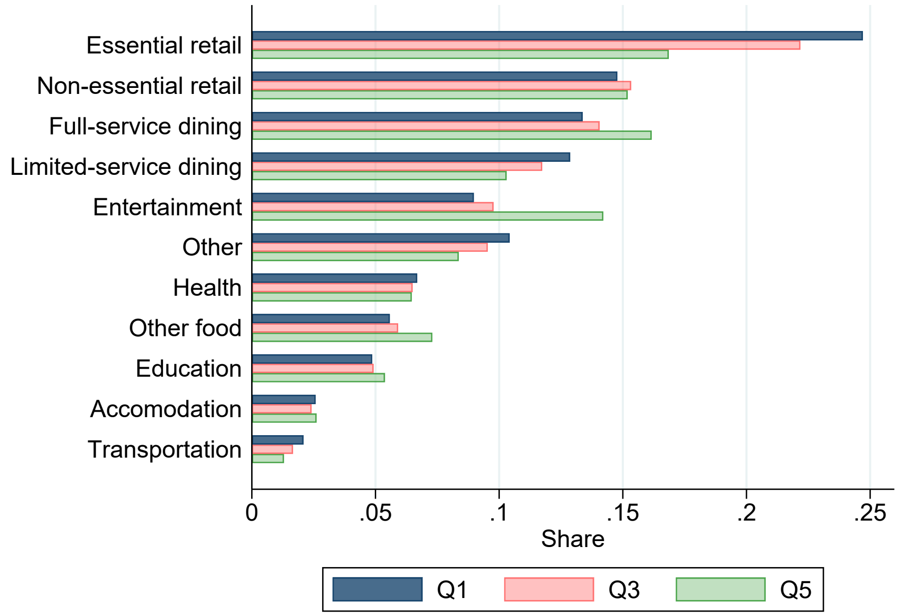
(a) Correlation between cross-class friendships and interactions



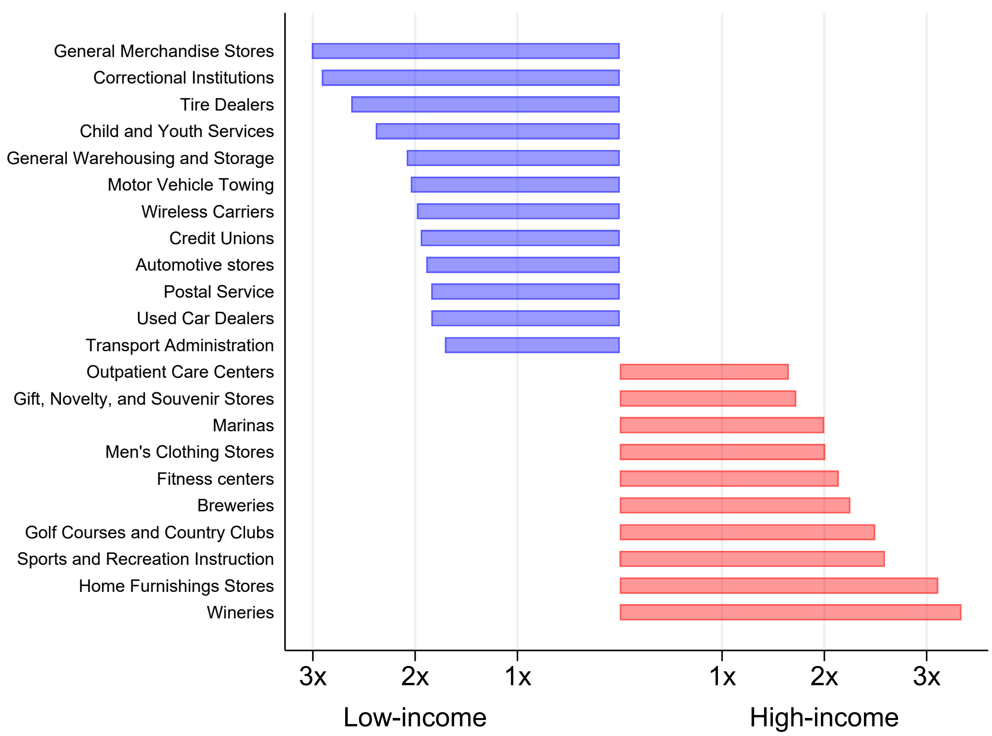
(b) Correlation visits to third places and cross-class friendships

Figure 3: Connection with Facebook friendship measure

Notes: Panel (a) shows a binned scatterplot portraying the ZIP code level relationship between cross-class friendships (the variable `ec_zip` from Chetty et al. (2022) and our ZIP code level measure of poor (bottom quintile neighborhood residents) encounters with the rich (top quintile). Panel (b) shows the relationship between the percent of visits to "third places" in the y-axes vs. the ZIP code's `ec_zip` decile, from Chetty et al. (2022). In the dashed blue line, third places includes cafes, churches, gyms, civic organizations, beauty parlours, bars, libraries, bookstores, parks, and fast-food. The solid maroon line excludes fast food. (Back to section)



(a) Sector share of visits, by income quintile



(b) Difference in granular industries

Figure 4: Differences in industry by income

Notes: Panel (a) shows the share of visits to broad industry categories (defined in Appendix C) for people in the first, third, and fifth quintile of CBG median income. Panel (b) shows industries (defined by their 5 or 6 digit NAICS codes) with at least 5m visitors that are disproportionately frequented by residents of the bottom or top income quintiles. The blue bars at the top show places that are visited more by the poor. For example, General Merchandise Stores (mostly dollar stores) represent a 3x higher share of visits for residents of bottom quintile neighborhoods. The red bars below show places visited more by people from top income quintile neighborhoods. Golf Courses are a 2.5x higher share of visits for the rich, for example. (Back to section)

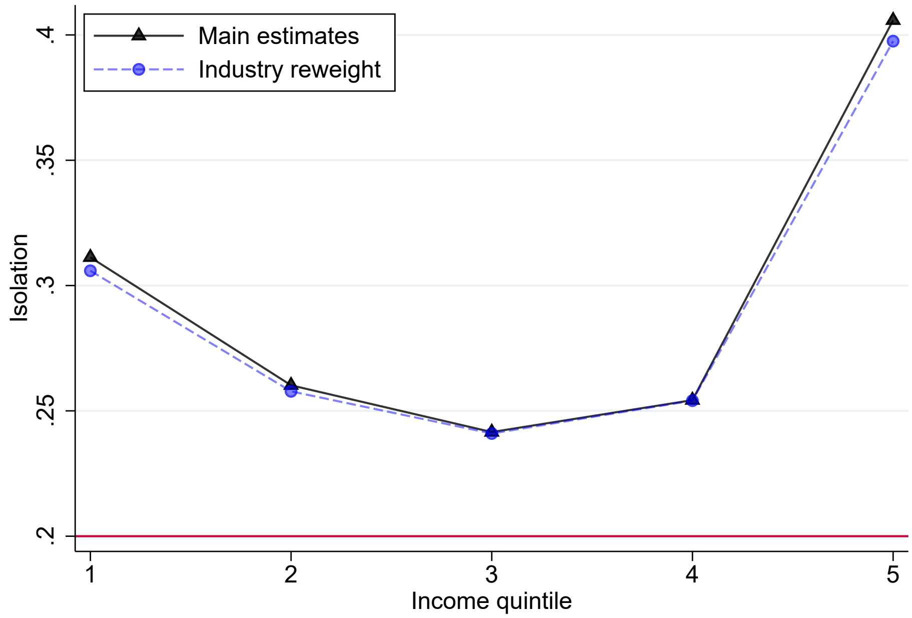
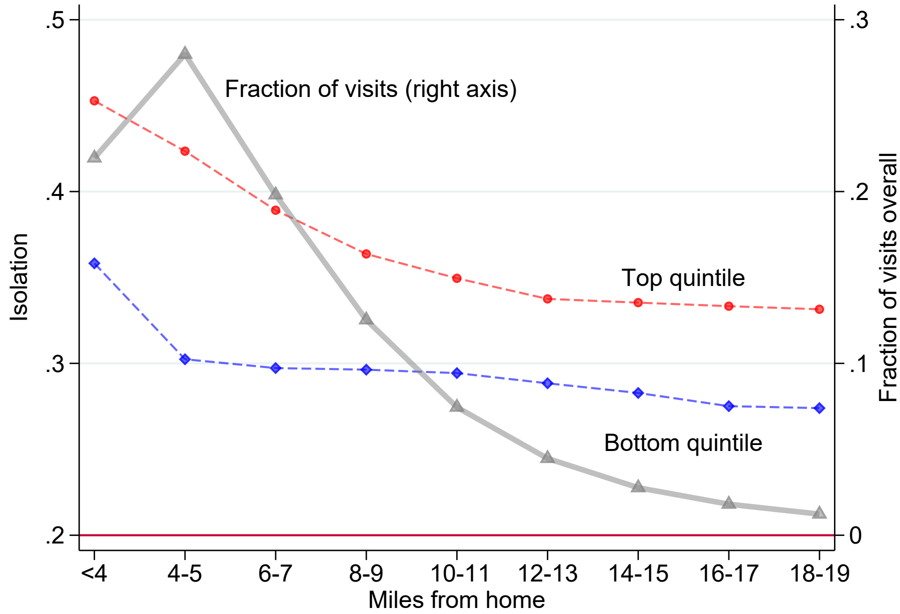
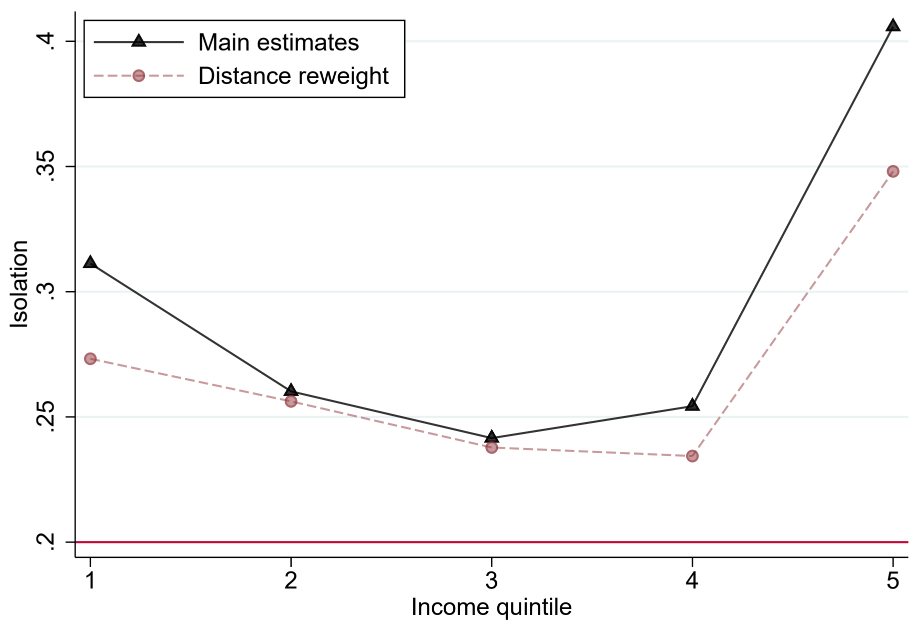


Figure 5: Isolation of different income quintiles, Industry reweighting

Notes: This figure shows what happens to class-based isolation if we assume that people from the five income quintiles visited all 453 industries in equal proportions. For example, this amounts to upweighting the visits to essentail retail establishments for top quintile residents and downweighting such visits for bottom quintile residents (see Figure 4). The black solid line shows the main estimates of isolation, the anti-diagnoal from Figure 1. The blue dashed line shows the counterfactual isolation with identical industry shares. (Back to section)



(a) Isolation vs. distance from home



(b) Isolation with and without reweighting for distance

Figure 6: The role of distance

Notes: This figure shows the role of distance in class-based isolation. In panel (a), the solid gray line (right axis) shows the fraction of overall visits to each distance bin, aggregating across all income groups. The dashed lines show the relationship between isolation and distance. Residents of high- and low-income neighborhoods are most likely to encounter people from a similar social class in their trips within five miles of their home. Panel (b) shows isolation results assuming that each income quintile has equal likelihood of visiting different distances. In other words, this assumes that the gray line in panel (a) is flat. (Back to section)

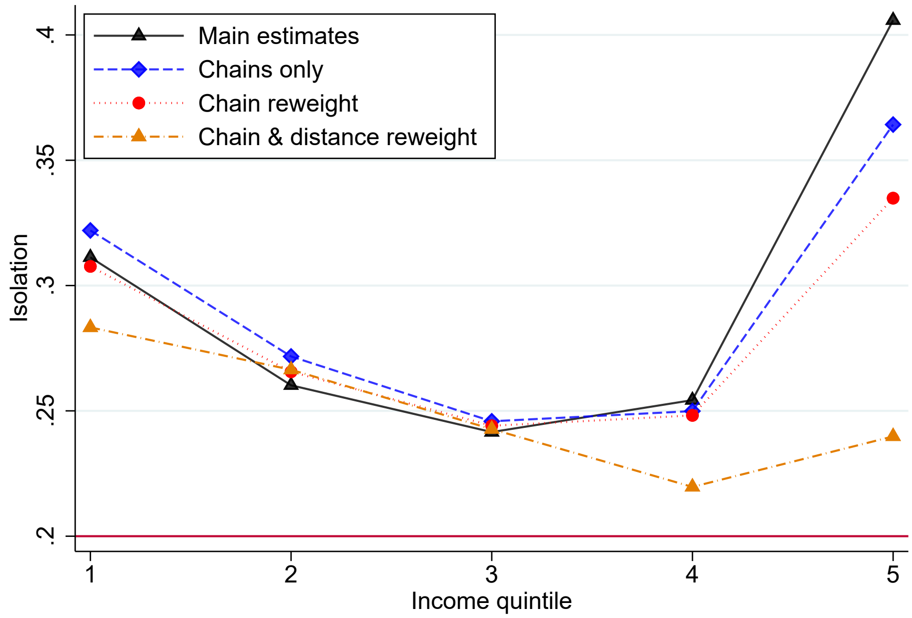
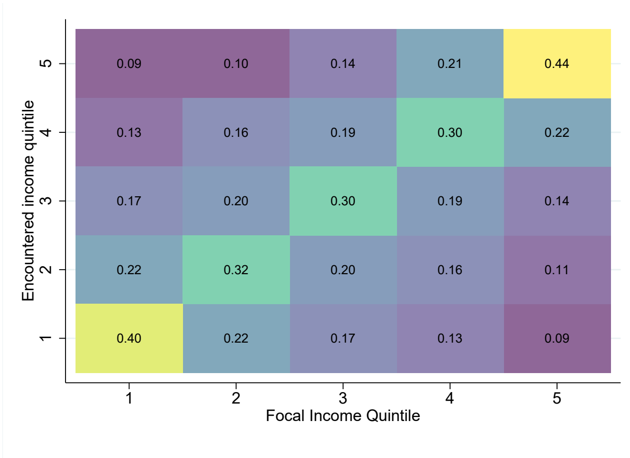
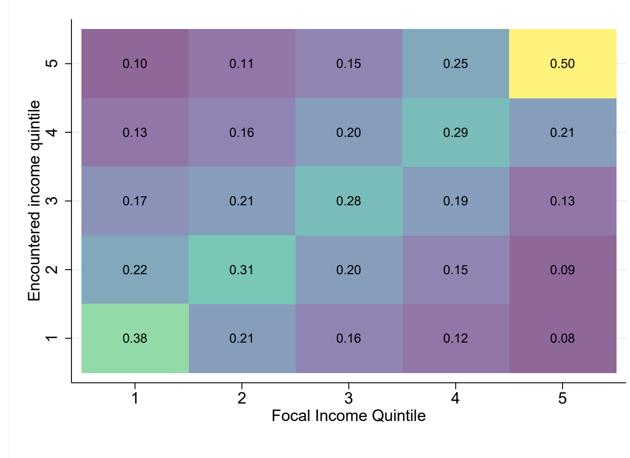


Figure 7: The role of chains

Notes: The solid black line shows the main isolation estimates from the anti-diagonal of Figure 1. The dashed blue line shows isolation restricting to chains only. The dotted red line shows isolation restricting to chains only and reweighting so that each income quintile visits each chain in equal proportion. The remaining variation is just the chain *locations* visited. Finally, the gold dotted and dashed line shows isolation when reweighting by both chain and distance. In other words, each income group is assumed to go to McDonald's in equal proportion, and to be just as likely to go to a near vs. far McDonald's location. (Back to section)

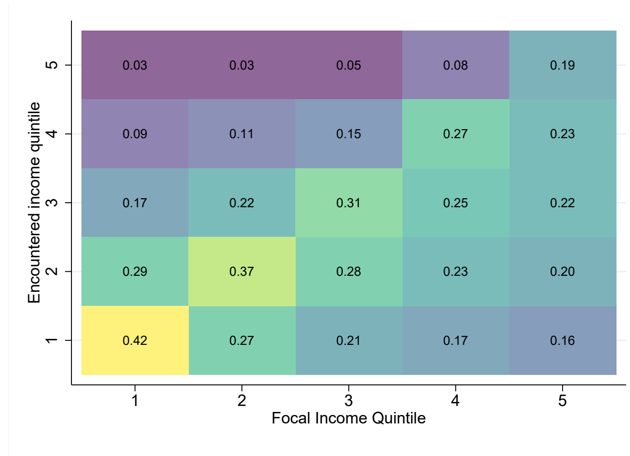


(a) Churches

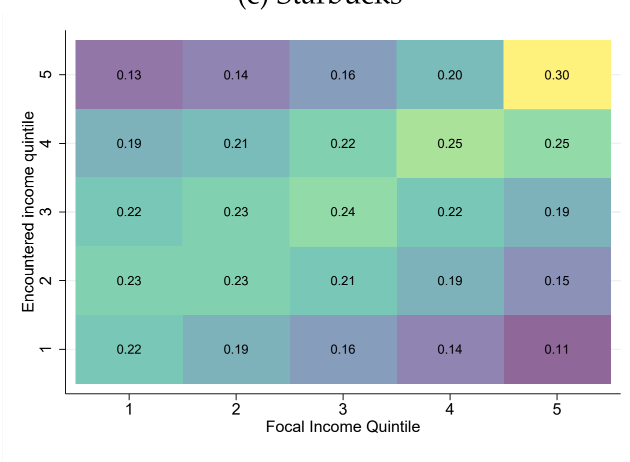


(b) Libraries

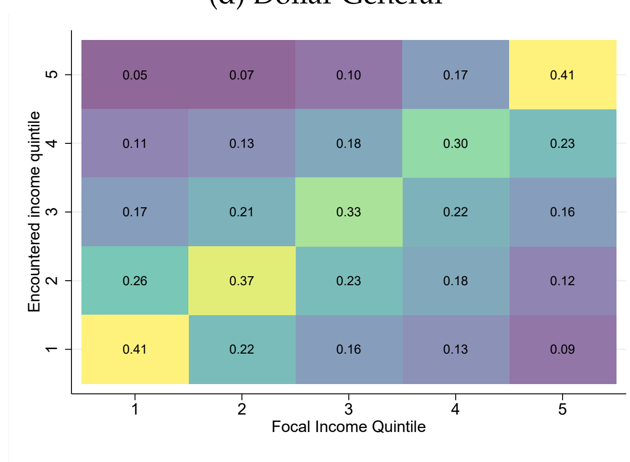
(c) Starbucks



(d) Dollar General



(e) Olive Garden



(f) Post Offices

Figure 8: Mixing at different places

Notes: This figure shows what mixing looks like in different kinds of places: churches (Number of locations=348,998), libraries (N=12,852), Starbucks (N=11,045), Dollar General (N=18,007), Olive Garden (N=789), and post offices (N=26,772). Each panel is identical to Figure 1, except that exposure is calculated using only the specified places. (Back to section)

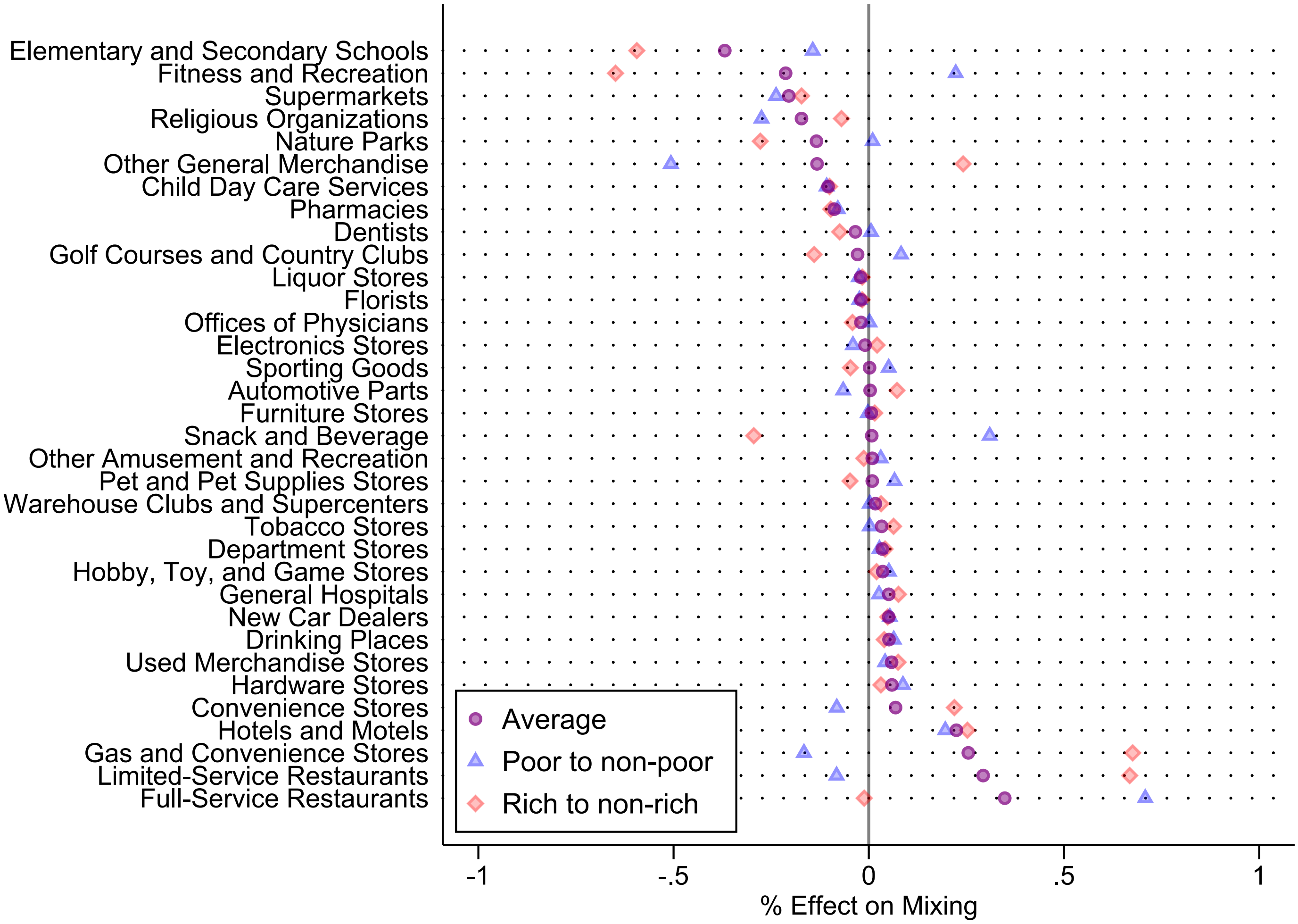


Figure 9: Mixing at different industries

Notes: This plot shows how the largest industries impact experienced isolation, based on the removal exercise described in Section 6. For example, the net effect of full-service restaurants is to increase mixing by 0.35% (last row, purple circle). This is based on the simple averages of two quantities: (i) the impact on exposure of quintile 1 to quintiles 2 to 5 (blue triangle) and (ii) the impact on exposure of the top quintile to quintiles 1 to 4 (red diamond). (Back to section)

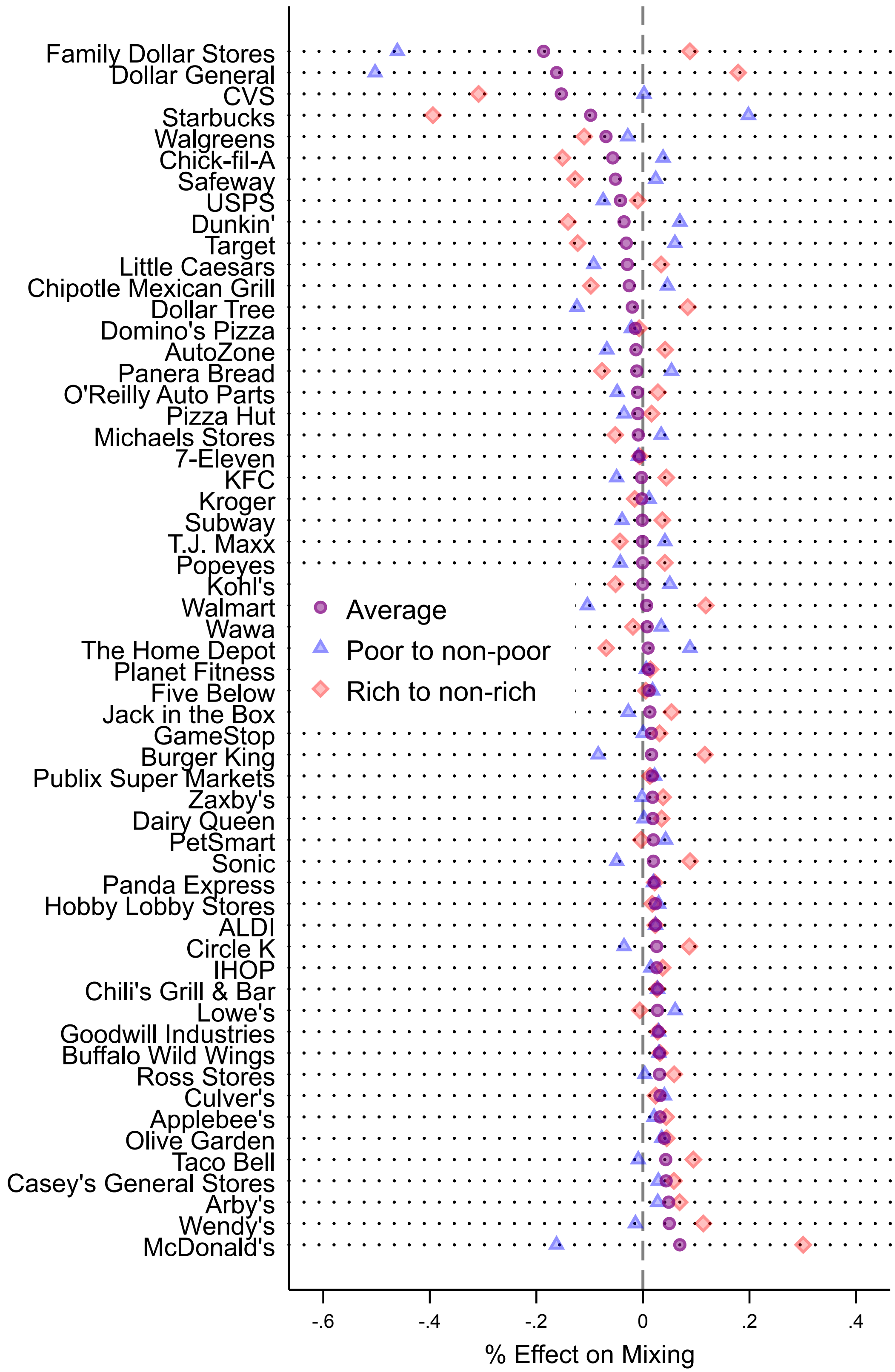
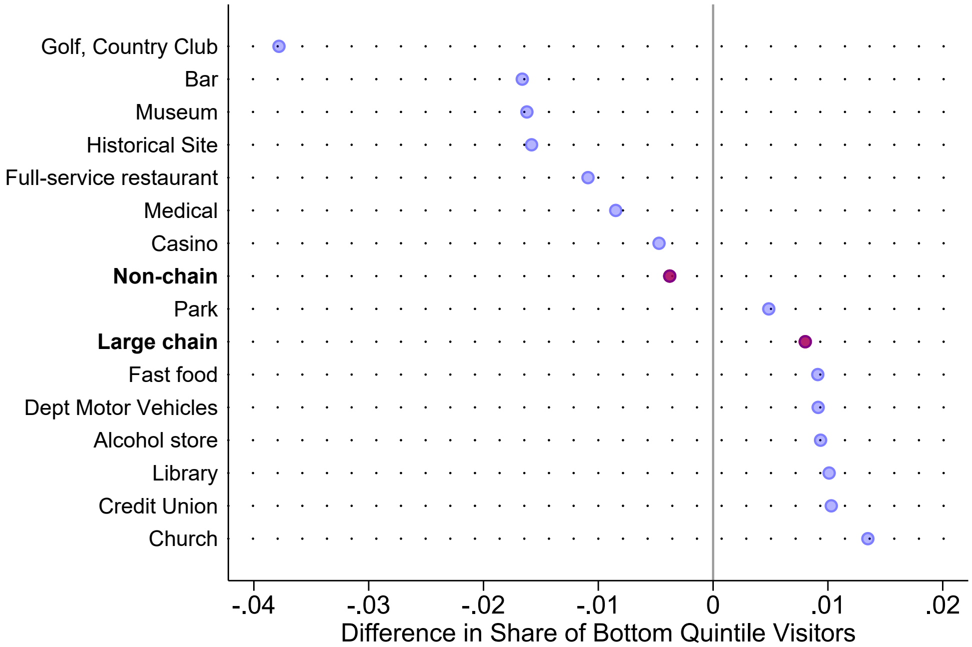
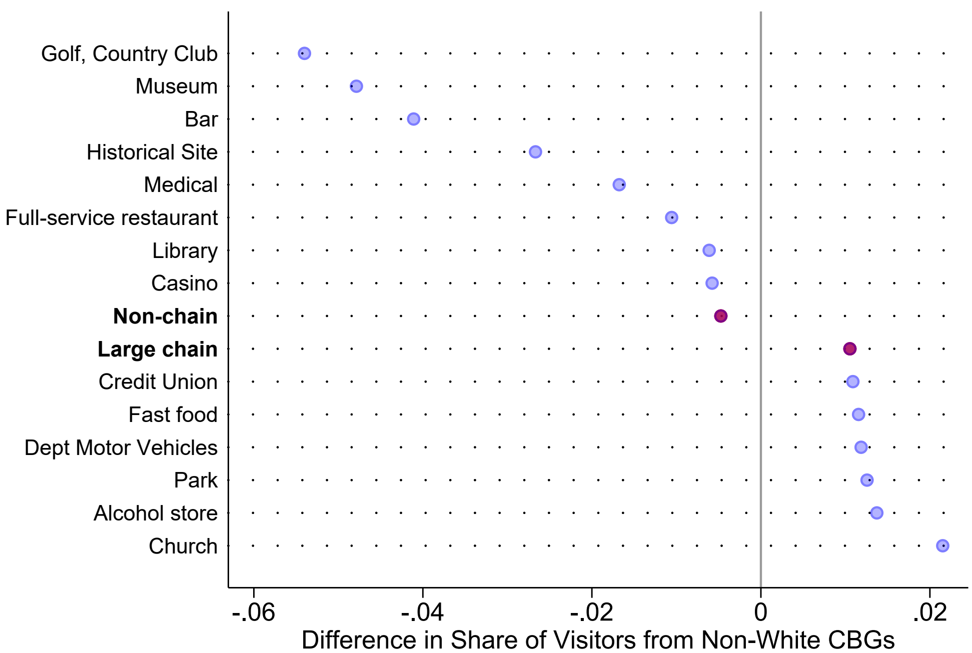


Figure 10: Contributions to Income Mixing by Large Chain

Notes: This plot shows how the largest multi-establishment firms impact experienced isolation, based on the removal exercise described in Section 6 and analogous to Figure 9. (Back to section)



(a) Share of visitors from bottom income quintile



(b) Share of visitors from majority non-white CBGs

Figure 11: Policy-relevant locations

Notes: Panel (a): This shows the composition of different types of places relative to the average composition in its city. For example, the visitors to golf courses and country clubs (top row) are about 4pp less likely to live in bottom quintile neighborhoods compared to visitors to other places in the same city. Panel (b): This performs the same exercise as panel (a), except using the share of visitors from majority non-white CBGs. For example, visitors to golf courses and country clubs are about 5pp less likely to reside in a majority non-white CBG compared to visitors to other places in the same city. (Back to section)

10 Tables

Industry	Visits (100m)	Percent	Cumulative Percent
Essential retail	72	22	22
Non-essential retail	49	15	37
Full-service dining	48	15	52
Limited-service dining	39	12	64
Entertainment	32	10	73
Other	29	9	82
Health	19	6	88
Other food	19	6	94
Education	13	4	98
Accomodation	6	2	99
Transportation	2	1	100
Total	328	100	100

Table 1: Industry composition of visits

Notes: This table shows all the visits in the SafeGraph data, aggregated by broad industries. The industry groupings are defined in Appendix C. (Back to section)

A Additional figures

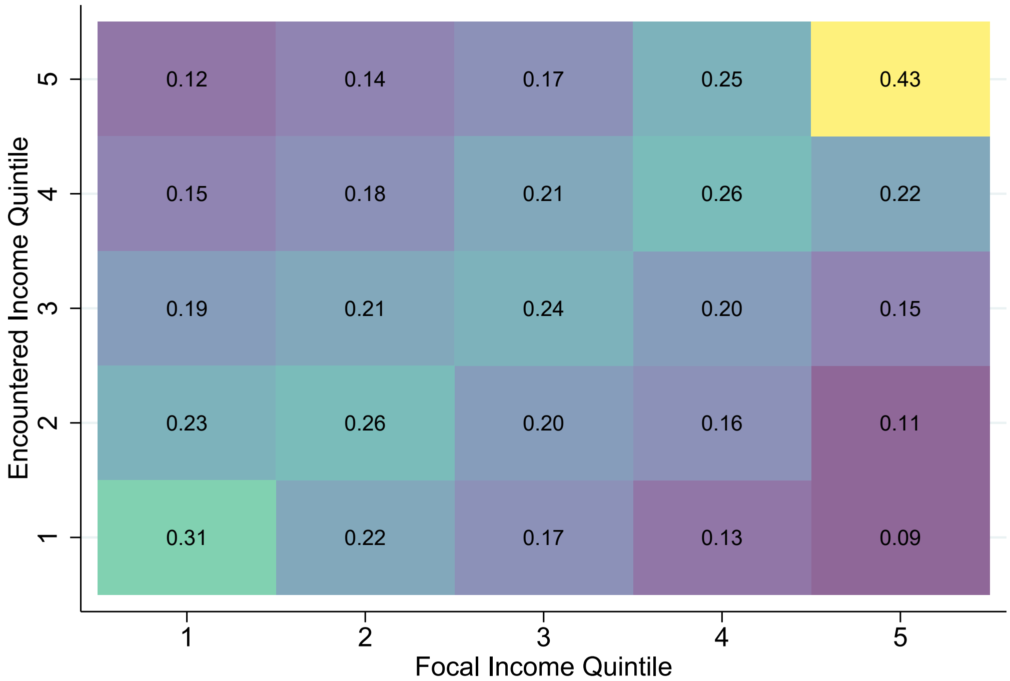


Figure A.1: Exposure to different income groups, third places only

Notes: This figure is identical to Figure 1, except it restricts to "third places:" cafes, churches, gyms, civic organizations, beauty parlours, bars, libraries, bookstores, parks, and fast-food establishments (Oldenburg, 1999). The results suggest that isolation is similar looking at the entire SafeGraph data (Figure 1) or third places alone. (Back to section)

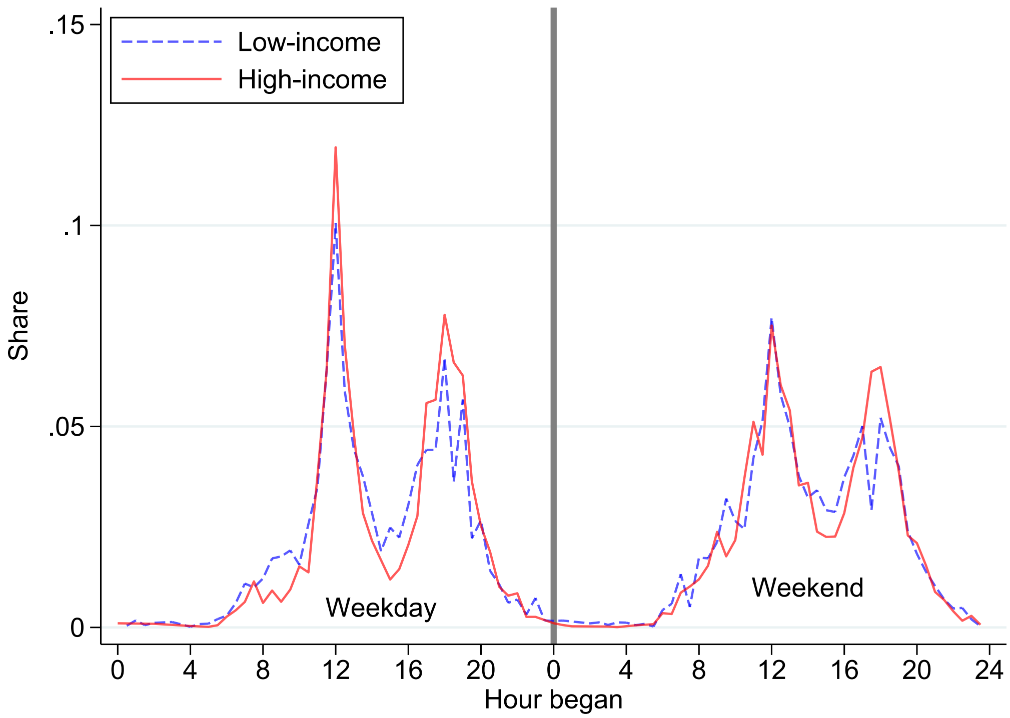


Figure A.2: Eating hours in the ATUS

Notes: This plot shows a histogram of eating times at restaurants for respondents to the American Time Use Survey (Flood et al., 2022). Low-income is defined as having family income of less than $49,999. High-income is defined as having family income of $75,000 and above. The start time is discretized into 30-minute intervals. Eating is defined as any activity with activity code 110000-119999 ("Eating and drinking") and with where code 104 ("Restaurant or bar"). We use 2006 weights.

B Additional tables

	Income quintile				
Industry	Q1	Q2	Q3	Q4	Q5
Essential retail	25.0	24.2	22.5	20.3	17.0
Non-essential retail	14.3	14.8	15.0	14.9	14.7
Full-service dining	12.9	13.4	14.0	14.8	16.1
Limited-service dining	12.4	12.0	11.4	10.9	9.8
Other	10.0	9.5	9.2	8.9	8.1
Entertainment	8.8	8.8	9.7	11.2	14.0
Health	6.0	6.0	6.0	6.0	6.1
Other food	4.6	4.9	5.5	6.2	7.2
Education	3.8	4.0	4.2	4.2	4.6
Accomodation	1.8	1.8	1.9	1.9	2.1
Transportation	0.9	0.9	0.8	0.7	0.6

Table A.1: Industry composition of visits, by income quintile

C Industry categories

Here we show the broad industry categories we use along with examples.

- Essential retail (stores mostly selling food, gas, or health items)

  NAICS Codes: 445000-447999.

  Examples: 7-Eleven, BP, CVS, Safeway, Walmart.

- Nonessential retail (stores mostly selling clothing or durable goods)

  NAICS Codes: 440000-459999.

  Examples: Ace Hardware, Staples, The Men's Wearhouse, Party City

- Full-service restaurants

  NAICS Code: 722511

  Examples: Applebee's, Red Lobster, Waffle House.

- Limited-service restaurants

  NAICS Code: 722513

  Examples: Arby's, KFC, Panda Express, Taco Bell.

- Accommodation.

  NAICS Codes: 721000-721999

  Examples: Comfort Inn, Motel 6, Days Inn.

- Entertainment

  NAICS Codes: 710000-719999

  Examples: Myrtle Beach State Park, Planet Fitness, Magnolia Greens Gold Course

- Health

  NAICS Codes: 620000-629999

  Examples: YMCA, Quest Diagnostics, Kaiser Permanente, MyEyeDr.

- Education

  NAICS Codes: 610000-619999

  Examples: Kumon, Kaplan Test Prep, Stuyvesant High School

- Transportation

  NAICS Codes: 480000-499999

  Examples: Greyhound, Fedex Ship Center, Millbrae BART Station

- Other services (mostly personal services).

  Primary NAICS Codes: 517312 (Wireless telecommunications), 522110 (Commercial banking), 524210 (Insurance agencies), 813110 (Religious orginizations).

- Other food establishments (bars, snack stands, etc.)

  NAICS Codes: 722000-722999

  Examples: Starbucks, Smoothie King, Cold Stone Creamery

D Friendship regressions

This section builds on the correlations with friendship discussed in Section 4.1. To separate out the role of income composition and other factors, we scrutinize these findings in regressions shown in Table A.2. In all cases, the unit of observation is ZIP code. The friendship measure is the outcome, our measure of poor-to-rich exposure is the primary covariate, and we incrementally add controls.

Column (1) includes no controls. The coefficient on poor exposure to the rich is 1.08, suggesting, for example, that if activity-measured poor exposure to the rich increases by 20pp, the Chetty et al. (2022) measure increases by 0.22, equivalent to a 11pp increase in the share of high-SES friends among low-SES individuals in that ZIP code. The coefficient drops to 0.70 when we include controls for demographics (column (2)), and drops further to 0.67 when we add fine income controls (column (3)). Finally, it increases slightly with the addition of city fixed effects (column (4)). Overall, however, the association remains statistically significant and meaningfully large in magnitude.

To what extent do these estimates track the impacts of establishments? In a final step, we use the establishment mix of a ZIP code to predict its level of cross-class friendships. In Table A.3, we swap out our ZIP code mix of exposure. Instead, we use the predicted exposure based on the ZIP code's mix of firms. We take the firm-level leave-one-out mean of exposure, omitting establishments in the focal ZIP code's commuting zone. This therefore relies on firms with at least two locations. The coefficient on this firm-based measure in Table A.3 is similar in magnitude and highly significant. Taken together, interactions as measured by SafeGraph data are highly predictive of friendships.

	(1)	(2)	(3)	(4)
Poor exposure to rich	1.079***	0.704***	0.673***	0.704***
	(0.011)	(0.012)	(0.014)	(0.027)
Demographic controls		×	×	×
Fine income controls			×	×
City fixed effects				×
R-squared	0.502	0.787	0.804	0.933
Observations	18,637	18,637	18,637	18,637

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Association between activities and friendship

Notes: This table shows the association between our measure of exposure across class lines and a ZIP code level measure of cross-class friendships from Chetty et al. (2022). Every observation is a ZIP code. The outcome in all models is ec_zip from Chetty et al. (2022): "two times the share of high-SES friends among low-SES individuals, averaged over all low-SES individuals in the ZIP code." The covariate of interest, "Poor exposure to rich," is the exposure of the bottom income quintile to the top income quintile at the ZIP code level, calculated using SafeGraph data. Demographic controls includes controls for ZIP code age, sex, race, employment, and education. Fine income controls includes a count of all 17 income bins enumerated by the American Community Survey. The coefficient in column (1) suggests, for example, that if poor exposure to the rich increases by 20pp, ec_zip increases by 0.22, equivalent to a 11pp increase in the share of high-SES friends among low-SES individuals in that ZIP code. (Back to section)

	(1)	(2)	(3)	(4)
Poor exposure to rich, firm level	2.960***	1.216***	0.969***	0.602***
	(0.044)	(0.045)	(0.041)	(0.066)
Demographic controls		×	×	×
Fine income controls			×	×
City fixed effects				×
R-squared	0.299	0.715	0.758	0.916
Observations	18,582	18,582	18,582	18,582

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.3: Association between activities and friendship using chains

Notes: This table is analogous to Table A.2 except the primary covariate is a measure of firm-level income mixing. Each ZIP code is assigned a value of poor to rich exposure based on the typical exposure in the mix of establishments it has. This leaves out establishments in the focal ZIP code, so relies on chains. (Back to section)

E Probing our exposure measure

Our main results are the estimates in Figure 1. These show how often each income quintile encounters people from other income quintiles. If we think of Census block groups as the units of our analysis, this is an accurate representation assuming the sample that SafeGraph provides is representative.

However, these estimates are also useful to the extent that they measure the exposure of *individuals* in these income groups, not just neighborhoods. Since we lack individual data, our analysis is vulnerable to the ecological fallacy (Cho and Manski, 2008; Piantadosi et al., 1988) in that neighborhood aggregates do not necessarily map directly to individual traits. What looks like isolation between neighborhoods may mask substantial mixing among groups.

In what follows, we present checks addressing potential sources of problematic sorting within neighborhoods. We argue that:

1. Our CBG-level measures of isolation are conservative.
2. Even using more accurate measures of class membership, our findings of increased isolation for the rich and poor quintiles persist.
3. Our ranking of firms is also robust to alternative measures of visitor class.

In the following, we first elaborate on the possibility for bias that arises from our neighborhood-level measures (Section E.1). We then conduct a series of empirical tests to determine the likely direction of bias. In Section E.2, we unpack CBGs in the simplest possible way, assuming that each visitor's income is a random draw from their home CBG. This produces identical results. But if the richest and poorest residents *within* the same CBG tend to go to different places, there is still a potential for bias. So in Section E.3 we zoom in on homogenous CBGs, where the majority of residents fall into that income category. Finally, Section E.4 shows that firm-level exposure looks very similar when we rank firms using only homogeneous CBGs.

E.1 Sources of bias

First we discuss potential sources of bias in our main results. Using our main equation of exposure (Equation 2), exposure of group G to quintile $Q1$ is:

$$Exposure(Q1, G) = \frac{\sum_{i \in G} \sum_k FirmExposure_i(Q1, k) * Visitors_{k(i)}}{\sum_G \sum_k Visitors_{k(i)}}$$

To clarify how our aggregated measures could lead to bias, we make two abstracting changes: first, we remove the denominator to focus on the total number of visitors from group G. Then, unpacking the summation for the very first CBG i in group G, we would have the visit-weighted exposure as:

$$\sum_k FirmExposure_i(Q1, k) * Visitors_{k(i)}$$

When we estimate individual-level exposure, both firm exposure and the visitor count are measured with ecological proxies. Assume the true number of visitors to firm k from CBG i is $Visitors^*_{k(i)}$. Then

$Visitors_{k(i)} = Visitors^*_{k(i)} + \widetilde{Visitors}_{k(i)}$, where $\widetilde{Visitors}_{k(i)}$ is the unobserved error and $Visitors_{k(i)}$ is our estimate from the data. Defining firm exposure similarly, our equation becomes:

$$\sum_k (FirmExposure^*_i(Q1,k) + \widetilde{FirmExposure}_i(Q1,k)) * (Visitors^*_{k(i)} + \widetilde{Visitors}_{k(i)})$$

In words, the error arises for two reasons: error in calculating the income of the other firm visitors and error in calculating the income of the focal visitors (which amounts to a mistaken weighting). If the correlation between these errors is zero, then we should have little bias as we aggregate across all firms. However, when measuring isolation, we argue that these errors will tend to move together because these are both estimates of the number of people from group G visiting firm k.

For example, many top quintile neighborhoods have residents who go to dollar stores. If these are always the lowest-income residents of the top quintile neighborhoods, then both errors are positive: we are over-estimating exposure to the rich at these stores and overestimating the number of rich people who visit them. This leads to an overestimate of $Exposure(Q1, Q5)$ and an underestimate of $Exposure(Q1, Q1)$. Similarly, a golf course might attract the wealthiest residents of poor CBGs, leading to an overestimate of $Exposure(Q5, Q1)$ and an underestimate of $Exposure(Q5, Q5)$. In both cases, these tend to deflate values along the anti-diagonal and inflate values away from it. Following the logic of these examples, our estimates of isolation are likely conservative.

What about the comparison of isolation across quintiles: that the rich are more isolated than the poor, and the poor more isolated than the middle? We present two robustness checks of these findings below.

E.2 Income bins

Here, we test what would happen to our measures of exposure if each person is a random draw from their home CBG income distribution. At the CBG level, the American Community Survey provides estimates of the number of households in each of sixteen income bins:

1. Less than $10,000
2. $10,000 to $14,999
3. $15,000 to $19,999
4. $20,000 to $24,999
5. $25,000 to $29,999
6. $30,000 to $34,999
7. $35,000 to $39,999
8. $40,000 to $44,999
9. $45,000 to $49,999
10. $50,000 to $59,999
11. $60,000 to $74,999
12. $75,000 to $99,999
13. $100,000 to $124,999
14. $125,000 to $149,999
15. $150,000 to $199,999

16. \$200,000 or more

Let $P_i(a)$ give the share of respondents in CBG i who have a household income in bin a. Instead of treating all CBG residents as belonging to a certain quintile, we assume that the probability that both visitors and encountered people belong to bin a is simply $P_i(a)$. A similar weighted average can be taken to calculate the exposure of bin a to bin b, for all combinations of bins a and b.

Exposure to bin b in store k is:

$$Share_{b,k} = \frac{\sum_{j \in CBGs} Visitors_{k(j)} * P_j(b)}{\sum Visitors_{k(j)}}$$

Then the exposure of bin a to bin b is given by:

$$Exposure(a, b) = \frac{\sum_{i \in CBGs} \sum_k P_i(k) * Visitors_{i(k)} * Share_{b,k}}{\sum_{i \in CBGs} \sum_k P_i(k) * Visitors_{i(k)}}$$

This analysis "breaks up" the census block groups in both dimensions. Both the visiting and encountered people can belong to different income groups. The probability that they belong to any given income bin is the share of people from their CBG in that income bin. This addresses the simplest kind of measurement error. If each visitor is a random draw from their CBG, the probability that they belong to a certain income bin is the share of CBG residents in that bin. Figure A.3 shows the results.

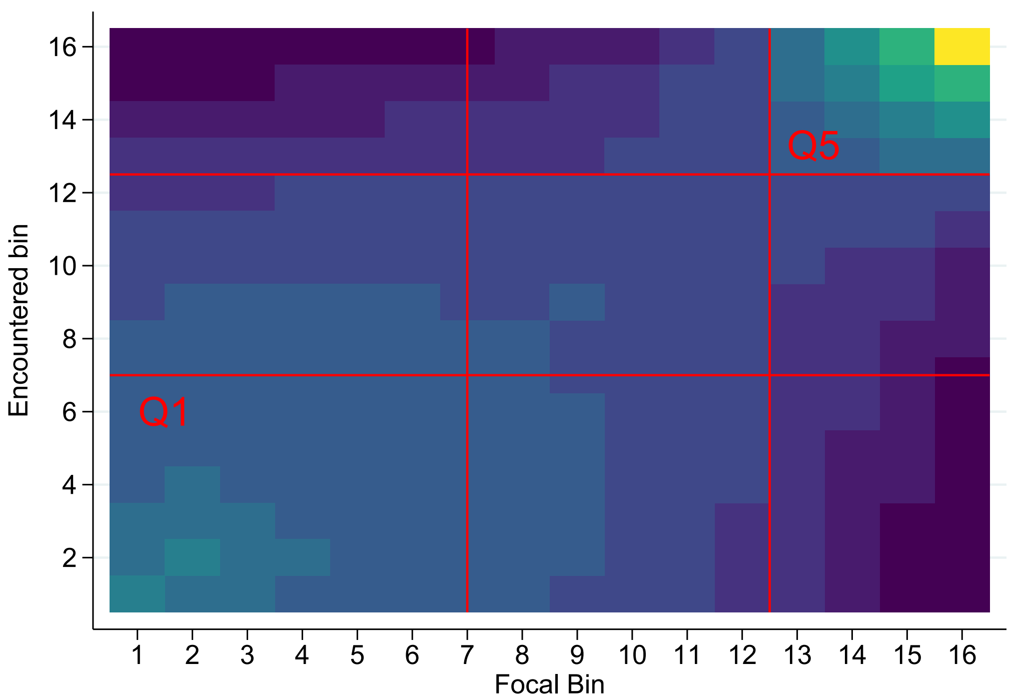


Figure A.3: Exposure results using income bins

The matrix is warped relative to our main results because the ACS bins do not exactly match the percentiles of the income distribution. However, the results mirror what we found in Figure 1.

Isolation is higher at the bottom and top, but highest at the top. Even incorporating the entire CBG income distribution as estimated by the ACS, we have similar findings. This addresses concerns that there is wide variation within CBGs belonging to a given quintile. Having a median household income in a certain quintile could be associated with a wide range of income distributions. Even adjusting for this within-CBG income variability, isolation estimates are similar.

E.3 Homogeneous CBGs

The previous analysis still leaves open the possibility that there is sorting within CBG residents. Indeed, it assumes that rich and non-rich residents of a rich neighborhood frequent similar locations. This is a strong assumption: given the substantial differences in visit patterns across neighborhoods, we expect that within-neighborhood, residents of different income levels will also visit different places.

Another way to probe ecological bias is to study what happens when we restrict to highly homogeneous Census block groups. This allows us to relax the strong assumption of within-neighborhood similarity across income groups. By focusing on the subset of neighborhoods that have a high level of income homogeneity, we isolate neighborhoods in which our aggregated measures introduce little distortion relative to individual-level measures.

As mentioned above, the ACS income bins do not align perfectly with the cutoffs for income quintiles. We study four groups of CBGs. There are two homogenous Q5 groups because the cutoff for the 5th quintile is $85,000, which is in the middle of two of the bins. We use the two closest bin borders. The homogenous middle group uses a lower cutoff because very few CBGs have 75% or more of households within that band (note that all other groups stretch to the very bottom or top of the income distribution).

- Homogenous Q1: CBGs with bottom quintile median income and 75%+ households in bins 1-6 (N=6,096 CBGs). The upper limit of bin 6 ($34,999) happens to line up closely with the upper limit of Q1 income, $37,216.
- Homogenous middle: CBG with median income in Q2 or Q3 and with 50+% of households in bins 7-11 (N=1,144 CBGs)
- Homogenous Q5 ($75K): CBGs in the top quintile of median income with 75% or more households over $75k (N=10,281 CBGs)
- Homogenous Q5 ($100K): same but with 75%+ households over $100k (N=3,134 CBGs)

We restrict our analysis entirely to these four groups. That is, we treat these as the four focal groups of CBGs, measuring exposure for each. And when we measure exposure, we ignore other visitors who do not come from a homogeneous CBG. In this way, both the visitors and the encountered people are much more likely to belong to their defined income group.

Unlike the quintiles, these groups do not represent an even share of the population. So isolation of group G is calculated as:

$$Isolation_G = \frac{Exposure(G,G)}{\frac{\sum_{i \in G} Visitors_i}{\sum_i Visitors_i}}$$

The denominator is simply the share of overall visits accounted for by group G, and the numerator is the actual exposure of group G with itself. So this is the ratio of realized exposure to the exposure we would expect if encounters were entirely random. A value of 2 would mean that people in group G are twice as likely to encounter each other than what we would expect with random mixing.

Table A.4 shows the results.

Group	Isolation	Exposure to Q1	Exposure to Q5	N CBGs
Homogenous Q1	4.46	0.35	0.12	6,096
Homogenous Middle	1.78	0.22	0.14	1,144
Homogenous Q5 ($75K)	4.08	0.08	0.48	10,281
Homogenous Q5 ($100K)	6.60	0.07	0.53	3,134

Table A.4: Isolation of homogenous Census block groups

Overall, the results echo the main findings. The rich are the most isolated, the poor slightly less so, and the middle are less isolated than both. For example, the homogenous Q5 ($100K) group (last row) is 6.60 times more likely to encounter a co-member compared to fully random encounters. The homogenous Q1 group (top row) is still highly isolated, 4.46 times more likely to encounter its own group than random. In the third and fourth columns, we show exposure to quintiles as measured in the main analysis of the paper. Understandably, these homogenous CBGs are in general more isolated than the quintiles they represent. For example, the homogenous Q5 ($100K) group's exposure to the top quintile is 0.53, as compared to 0.43 for the full sample of Q5 residents.

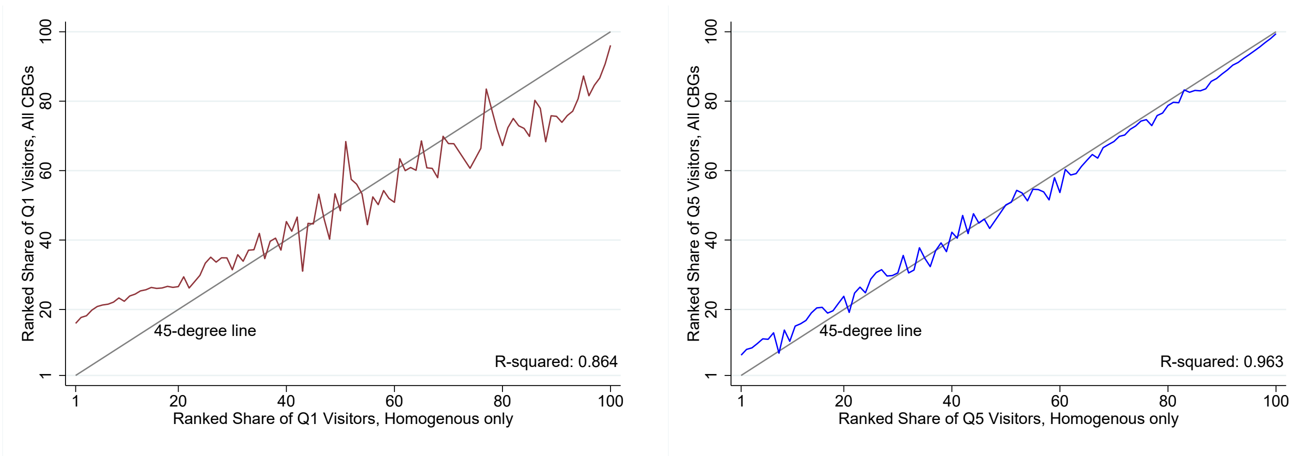
E.4 Firm ranking with homogeneous CBGs

In this section, we show how the composition of places looks when we only count people from homogeneous CBGs, using Homogenous Q1 and Homogenous Q5 ($75K) as described in Section E.3. We will show that using only these narrow subsets of homogenous neighborhoods, we get very similar rankings of firm-level exposure.

Procedure For every unique location name in the data, we calculate its ranking in terms of the share of visitors from the top and bottom quintile. These are the measures that go into **??**, our rankings of firms. Next, we calculate those same rankings but using only the homogenous CBGs. We study the rank-rank correlation between these two measures of firm-level exposure.

Figure A.4 shows the results. We find that the two measures are highly correlated. The x-axis is the firm exposure ranking using only homogenous CBGs. The y-axis is the average all-CBG ranking, with a unique average taken for each of the 100 values in the x-axis. In the bottom right corner of each panel, we show the r-squared from a rank-rank regression (without a constant). For example, Panel (b) shows that the firms ranked 2nd out of 100 in terms of visitors from the homogenous Q5 CBGs are ranked 8th in terms of visitors from the top quintile in general. Throughout panel (b), the ranks track each other closely with an r-squared of 0.963 in the regression.

Panel (a) shows that the correlation is less tight for bottom quintile visitors. For example, the top places for homogenous bottom quintile CBGs are only ranked 20th out of 100 when we use the whole bottom quintile. Apart from the top-ranked places in panel (a), the ranks track each other closely, with an r-squared of 0.864.



(a) Bottom quintile

(b) Top quintile

Figure A.4: Firm rankings with all vs. homogenous CBGs