2.2 ZIP Code-Level Data

ZIP Code Identifiers and Population Variables

Variable Name	Description
zip	5-digit ZIP code tabulation area code.
county	5-digit county FIPS code.
num_below_p50	Number of children with below-national-median parental household income. This variable is not constructed using Facebook data; it is obtained from publicly available data posted at the Opportunity Atlas website (Chetty et al. 2018).
pop2018	Population in 2018. This variable is not constructed using Facebook data; it is obtained from publicly available data posted at the Census website (American Community Survey).

ZIP Code Connectedness Statistics

Variable Name	Description
ec_zip	Baseline definition of economic connectedness: two times the share of high-SES friends among low-SES individuals, averaged over all low-SES individuals in the ZIP code. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document. This variable is mapped for the Los Angeles area in Figure 2b of Chetty et al. (2022a).
ec_se_zip	The standard error of economic connectedness, incorporating both sampling error and error from the addition of noise to protect privacy. The variance due to sampling error is calculated using a bootstrap approach described in Chetty et al. (2022a). We then add the noise variance that we apply to protect privacy to generate a combined standard error.
nbhd_ec_zip	Economic connectedness calculated using only within-neighborhood friends. We add noise to protect privacy, as described in Section 3 of this document. This variable is used to construct the green neighborhood bar in Figure 2A of Chetty et al. (2022b).
ec_grp_mem_zip	Two times the share of high-SES friends among low-SES individuals averaged over all low-SES individuals in the ZIP code, restricting attention to friendships that we can allocate to the group in which they were formed as described in Supplementary Information B.1 and B.2 of Chetty et al. (2022b). This variable is used in the first row of Table 2 of Chetty et al. (2022b).

ec_high_zip	Economic connectedness for high-SES individuals: two times the share of high-SES friends among high-SES individuals, averaged over all high-SES individuals in the ZIP code.
ec_high_se_zip	The standard error of economic connectedness for high-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy.
nbhd_ec_high_zip	High-type economic connectedness calculated using only an individual's neighborhood friends. We add noise to protect privacy, as described in Section 3 of this document. This variable is used to construct the orange neighborhood bar in Figure 2A of Chetty et al. (2022b).
ec_grp_mem_high_zip	Two times the share of high-SES friends among high-SES individuals averaged over all high-SES individuals in the ZIP code, restricting attention to friendships that we can allocate to the group in which they were formed.
exposure_grp_mem_zip	Mean exposure to high-SES individuals by ZIP code for low-SES individuals: two times the average share of high-SES individuals in individuals' groups, averaged over low-SES users. We assign Facebook users to groups within settings as described in Supplementary Information B.1 of Chetty et al. (2022b). We calculate SES as in Supplementary Information B.1 of Chetty et al. (2022a). This variable is mapped for the Los Angeles area in Figure 4B of Chetty et al. (2022b).
exposure_grp_mem_high_zip	Mean exposure to high-SES individuals by ZIP code for high-SES individuals: two times the average share of high-SES individuals in individuals' groups, averaged over high-SES users.
nbhd_exposure_zip	Exposure calculated using only users living in the relevant ZIP code. We add noise to protect privacy, as described in Section 3 of this document. This variable is used to construct the green and orange neighborhood bars in Figure 2B of Chetty et al. (2022b). Note that this is the same for high- and low-SES individuals who live in the same ZIP code.
bias_grp_mem_zip	ec_grp_mem_zip divided by exposure_grp_mem_zip, all subtracted from one. This variable is mapped for the Los Angeles area in Figure 4D of Chetty et al. (2022b).
bias_grp_mem_high_zip	ec_grp_mem_high_zip divided by exposure_grp_mem_high_zip, all subtracted from one.
nbhd_bias_zip	nbhd_ec_zip divided by nbhd_exposure_zip, all subtracted from one. This variable is used to construct the green neighborhood bar in Figure 2C of Chetty et al. (2022b).

nbhd_bias_high_zip	nbhd_ec_high_zip divided by nbhd_exposure_zip, all subtracted from one. (Note again that, within the same neighborhood, exposure is the same for low-SES and high-SES individuals) This variable is used to construct the orange neighborhood bar in Figure 2C of Chetty et al. (2022b).

ZIP Code Cohesiveness Statistics

Variable Name	Description
clustering_zip	The average fraction of an individual's friend pairs who are also friends with each other. See equations (4) and (5) of Chetty et al. (2022a). We include links to people outside the ZIP code when calculating individual clustering (equation 4), but only average individual clustering over users in the relevant ZIP code to compute clustering at the ZIP code level (equation 5). We add noise to protect privacy, as described in Section 3 of this document.
support_ratio_zip	The proportion of within-ZIP code friendships where the pair of friends share a third mutual friend within the same ZIP code. See equation (6) of Chetty et al. (2022a). We add noise to protect privacy, as described in Section 3 of this document.

ZIP Code Civic Engagement Statistics

Variable Name	Description
volunteering_rate_zip	The percentage of Facebook users who are members of a group which is predicted to be about 'volunteering' or 'activism' based on group title and other group characteristics. We do not include groups that have the privacy setting 'secret' enabled. We additionally manually review the 50 largest such groups in the United States and the largest group in each state, and remove the very small number of groups that are clearly misclassified. We add noise to protect privacy, as described in Section 3 of this document.
civic_organizations_zip	The number of Facebook Pages predicted to be "Public Good" pages based on page title, category, and other page characteristics, per 1,000 users in the ZIP code. We remove pages that do not have a website linked, do not have a description on their Facebook page or do not have an address listed. We then assign the page to a ZIP code on the basis of its listed address. We add noise to protect privacy, as described in Section 3 of this document.