

2.3 High School Data

High School Identifiers and Population Variables

Variable Name	Description
high_school	12-digit NCES school ID.
high_school_name	Name of the high school.
zip	5-digit ZIP code tabulation area code.
county	5-digit county FIPS code.
students_9_to_12	Number of students from grades 9 to 12. This variable is not constructed using Facebook data; it is obtained from publicly available data posted at the National Center for Education Statistics website

High School Connectedness Statistics

Variable Name	Description
ec_own_ses_hs	Baseline definition of economic connectedness: two times the share of high-SES friends within three birth cohorts among low-SES individuals, averaged over all low-SES individuals in the school. See equations (1), (2), and (3) of Chetty et al. (2022a) for a formal definition. We estimate 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 used in Supplementary Information Figure 3A of Chetty et al. (2022b) .
ec_own_ses_se_hs	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.
ec_parent_ses_hs	Economic connectedness with parental SES: two times the share of high-parental-SES friends (who attended the same school within three birth cohorts of the individual) among low-parental-SES individuals, averaged over all low-parental-SES individuals at the school. See equations (1), (2), and (3) of Chetty et al. (2022a) for more details on the calculation. We link individuals to parents as described in Supplementary Information A1 of Chetty et al. (2022a) , and estimate parental 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 used in Figure 5A of Chetty et al. (2022b) .

<code>ec_parent_ses_se_hs</code>	The standard error of economic connectedness with parental SES, incorporating both sampling error and error from the addition of noise to protect privacy.
<code>ec_high_own_ses_hs</code>	Economic connectedness for high-SES individuals: two times the share of high-SES friends within three birth cohorts among high-SES individuals, averaged over all high-SES individuals in the school.
<code>ec_high_own_ses_se_hs</code>	The standard error of economic connectedness for high-SES individuals, incorporating both sampling error and error from the addition of noise to protect privacy.
<code>ec_high_parent_ses_hs</code>	Economic connectedness for high-parental-SES individuals using parental SES: two times the share of high-parental-SES friends (who attended the same school within three birth cohorts of the individual) among high-parental-SES individuals, averaged over all high-parental-SES individuals in the school.
<code>ec_high_parent_ses_se_hs</code>	The standard error of economic connectedness for high-parental-SES individuals using parental SES, incorporating both sampling error and error from the addition of noise to protect privacy.
<code>exposure_own_ses_hs</code>	Mean exposure to high-SES individuals by high school for low-SES individuals: two times the average share of high-SES individuals within three birth cohorts, averaged over low-SES users. This variable is used in Supplementary Information Figure 3A of Chetty et al. (2022b) .
<code>exposure_parent_ses_hs</code>	Mean exposure to high-parental-SES individuals by high school for low-parental-SES individuals: two times the average share of high-parental-SES individuals within three birth cohorts, averaged over low-parental-SES users. This variable is used in Figure 5A of Chetty et al. (2022b) .
<code>bias_own_ses_hs</code>	<code>ec_own_ses_hs</code> divided by <code>exposure_own_ses_hs</code> , all subtracted from one. This variable is used in Supplementary Information Figure 3A of Chetty et al. (2022b) .
<code>bias_parent_ses_hs</code>	<code>ec_parent_ses_hs</code> divided by <code>exposure_parent_ses_hs</code> , all subtracted from one. This variable is used in Figure 5A of Chetty et al. (2022b) .
<code>bias_high_own_ses_hs</code>	<code>ec_high_own_ses_hs</code> divided by <code>exposure_own_ses_hs</code> , all subtracted from one.
<code>bias_high_parent_ses_hs</code>	<code>ec_high_parent_ses_hs</code> divided by <code>exposure_parent_ses_hs</code> , all subtracted from one.

High School Cohesiveness Statistics

Variable Name	Description
clustering_hs	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 only links to friends within the school when calculating individual clustering (equation 4). We add noise to protect privacy, as described in Section 3 of this document.

High School Civic Engagement Statistics

Variable Name	Description
volunteering_rate_hs	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.