Documentation for Segregation Estimates, Segregation Explorer Version 1.0

sean f. reardon, Ann Owens, Demetra Kalogrides, Heewon Jang, and Thalia Tom

May 2024

Please contact segxsupport@stanford.edu *with questions and errata.*

This work has been supported (in part) by Grant # 1911-19449 from the Russell Sage Foundation. Any opinions expressed are those of the principal investigator(s) alone and should not be construed as representing the opinions of the Foundation.

We estimate school segregation from 1991 to 2022 using a grade-level version of the Longitudinal Imputed School Dataset (LISD) 1.0. LISD combines, cleans, and imputes missing or erroneous data from the National Center for Education Statistics Common Core of Data (CCD) to provide school-year counts of students by race/ethnicity and free or reduced-price lunch (FRL) eligibility (for documentation and data download, see http://edopportunity.org/segregation/get-the-data).

Grade-specific race/ethnicity data is available since 1998 from the CCD. In prior years, and for school-year observations imputed in LISD, we assume that the racial composition of each grade is identical—i.e., if the school is 50% White, we assume each grade is 50% White—and generate grade-specific counts by race/ethnicity from LISD data and CCD grade enrollment data. The CCD does not provide grade-specific FRL eligibility data, so we assume the FRL eligibility composition of each grade is identical and generate grade-specific counts of FRL eligible students from LISD data and CCD grade enrollment data. We estimate segregation between students at each grade level and then aggregate to an estimate for each geography of interest by taking a weighted average, weighting the grade-specific segregation estimates by the number of students in each grade. This approach accounts for the different number of schools and students across grades. We estimate segregation only among "regular" (as opposed to special education, vocational, or alternative), non-virtual schools and we also exclude schools exclusively for the deaf and blind from segregation estimates.

The LISD comprises 10 imputed datasets. We estimate segregation separately in each dataset and then generate mean estimates over the 10 datasets; these mean estimates are available for download at http://edopportunity.org/segregation/get-the-data. Researchers interested in the imputation dataset-specific files or variance of estimates can contact segxsupport@stanford.edu.

We produce multiple measures of segregation, at multiple levels of geography, between schools and, for some geographic units, between school districts and sectors, and between various race/ethnicity- or FRL-based dyads, as described below.

Measures

We use the -seg- command in Stata (Reardon and Townsend 2018) to estimate segregation. Broadly, there are two classes of segregation measures: exposure and evenness (for an overview of segregation measures and more information on estimating each, see James and Taeuber (1985), Massey and Denton (1988), and Reardon and Firebaugh (2002)). We estimate two exposure measures, exposure (*X*) and isolation (*S*). Two-group exposure measures estimate the average enrollment share of one group in a second group's school (or subunit of interest; see below), taking only the presence of those two groups into account. For example, a Black-White exposure value of 0.3 indicates that the average share of White students in a Black student's school is 0.3 (considering only White and Black students as the total population). Two-group *isolation* measures estimate the average proportion of one's *own* group in a school (as a share of the two groups); a Black-White isolation value of 0.7 indicates that the average share of Black students in a Black student's school is 0.7 (considering only White and Black students' populations).

We estimate three two-group evenness measures: normalized exposure (*N*, also called relative diversity, variance ratio, or eta-squared), dissimilarity (*D*), and information theory (*H*). The interactive Segregation Explorer map displays normalized exposure, which measures the difference between two groups' exposure to one of the groups in a given geographic area (e.g., county, metropolitan area). For example,

a Hispanic-White normalized exposure value of 0.5 indicates that the proportion of Hispanic students in the average Hispanic student's school is 50 percentage points higher than in the average White student's school. (The normalized exposure measure ignores the presence of other groups aside from the racial dyad of interest.) The normalized exposure index ranges from 0 to 1. A value of 0 implies no segregation — the two groups have equal exposure to one group (all schools have identical proportions of the two groups). A value of 1 implies complete segregation— the two groups have no exposure to one another (no Hispanic student attends a school with any White students, and vice versa). For more on the normalized exposure index, see our brief here.

The dissimilarity index is interpreted as the proportion of one group that would need to change schools for the two groups to be evenly distributed across schools within a given geographic area. The information theory index assesses the degree to which schools' population diversity (in terms of the two groups of interest) deviates from the population diversity of the larger geographic area. These measures also range from 0 (no segregation—all schools have the same composition as one another (and as the larger unit)) to 1 (total segregation—no diversity in schools).

Finally, we estimate racial-economic segregation (*Z*), or the difference between two racial groups in their school's economic composition. This measure is the difference in the average free lunch eligibility (FLE) rate in each group's schools. For example, a Black-White racial-economic segregation value of 0.5 implies that the FL eligible rate in the average Black student's school is 50 percentage points higher, on a 0 to 1 scale, than in the average White student's school.

Levels of Geography

We use the Segregation Explorer Geographic Crosswalk to identify each school's geographic district, county, commuting zone, Core-Based Statistical Area (CBSA), metropolitan division, and geographic state. We estimates segregation within administrative districts (the local education agency (LEA) that administers the school), geographic districts (all schools located within the geographic boundaries of the LEA, including charter and Bureau of Indian Affairs schools not administered by that LEA), counties, commuting zones (separate estimates based on 1980, 1990, 2000, and 2010 definitions), CBSAs and, for the large CBSAs that have them, metropolitan divisions (separate estimates based on the 2003, 2013, and 2023 definitions), states, and the entire nation. See LISD documentation for more information on the Geographic Crosswalk (available at http://edopportunity.org/segregation/get-the-data).

Subunits

Within all units of geography described above, we estimate segregation between schools. Within the nation, states, CBSAs, metropolitan divisions, commuting zones, and counties, we also estimate segregation between administrative districts and geographic districts. For between-district measures, exposure and isolation measures capture exposure to a particular group in a student's district, rather than their school. Evenness measures assess the distribution of students by race/ethnicity or FRL eligibility between districts, rather than schools. Within geographic districts, we estimate segregation between charter schools and all other school to capture segregation between the charter and non-charter sectors. *H* and *N* are decomposable, meaning that one can divide the total value of segregation between schools by the value of segregation between districts (or sectors) to assess what proportion of total segregation in a place occurs between districts (or sectors).

Groups

We estimate segregation between racial/ethnic and FLE/free and reduced-priced lunch eligibility (FRLE) dyads: White-Black, White-Hispanic, White-Asian, White-Native American, White-non-White, White – Black + Hispanic + Native American students (referred to as "minority" in the data), White + Asian – Black + Hispanic + Native American, FLE-not FLE, and FRLE-not FRLE. Racial/ethnic categories are defined by the CCD. White, Black, Asian, Native American, and Multiracial refers to non-Hispanic students. Asian includes Asian, Hawaiian, and Pacific Islander students. When estimating White-non-White segregation, non-White refers to Black, Native American, Asian, and Hispanic students. FLE or RLE students are those whose families earn less than 130% and 185%, respectively, of the poverty threshold for their family size. We estimate racial-economic segregation as the White-Black, White-Hispanic, White – Black + Hispanic + Native American, and White + Asian – Black + Hispanic + Native American difference in average school FL rate.

Data Files

Each segregation file and its data dictionary are available at http://edopportunity.org/segregation/get-the-data. The data dictionary provides a description of each variable—our general variable naming convention is

[segregation measure][subunit]_[group1]_[group2]

Segregation measures: x (exposure), s (isolation), n (normalized exposure), d (dissimilarity), h (information theory), and z (racial-economic segregation)

Subunit: s (between schools), d (between administrative districts, g (between geographic districts), and c (between charter and non-charter schools)

Groups: wht (non-Hispanic White), blk (non-Hispanic Black), hsp (Hispanic), asn (non-Hispanic Asian, Pacific Islander, or Hawaiian Native), nam (Native American), min (blk+hsp+nam), was (wht+asn), nwh (asn+blk+hsp+nam), flu (FL eligible), nfl (not FL eligible), frl (FRL eligible), and nfr (not FRL eligible)

The variable ns_wht_blk, for example, is White-Black normalized exposure between schools.

Please contact segxsupport@stanford.edu with questions and errata.

References

James, David R. and Karl E. Taeuber. 1985. "Measures of Segregation." Sociological Methodology, 14:1-32.
Massey, Douglas S. and Nancy A. Denton. 1988. "The dimensions of racial segregation." Social Forces. 67:281-315.
Reardon, Sean F., and Glenn Firebaugh. 2002. "Measures of multigroup segregation." Sociological Methodology 32: 33-67.
Reardon, Sean F. & Joseph B. Townsend. 2018. "SEG: Stata module to compute multiple-group diversity and segregation indices," Statistical Software Components S375001, Boston College Department of Economics.