Documentation for the Longitudinal Imputed School Dataset (LISD) and Geographic Crosswalk, 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.*

The Common Core of Data (CCD), published by the National Center for Education Statistics (NCES), is a comprehensive dataset that includes annual fiscal and non-fiscal data on every public school and school district in the U.S. since the 1986-87 school year.[1] The CCD underlies many public policy and academic analyses, and CCD data on enrollment counts by race/ethnicity and free or reduced price lunch eligibility (FRLE) are frequently used in research on school composition and segregation.

However, school-level race/ethnicity and FRLE enrollment counts in the CCD are subject to data quality issues. Some states, districts, or individual schools do not report counts in some years, with missing data more prevalent in earlier years. FRLE counts are missing more frequently than race/ethnicity counts. FRLE data are also more likely to be reported but incorrect due to several issues. First, schools appear to sometimes report FRLE counts as zero instead of missing, despite NCES instructions to report unknown or unavailable counts as missing. Second, seemingly valid (non-zero, non-missing) FRLE counts are sometimes implausible, perhaps due to data entry error. Inspection identifies fewer such quality issues for race/ethnicity data, limited to a few states and years as described below. Finally, the Community Eligibility Provision (CEP), implemented in 2011-12, allows high-poverty schools and districts to count all students as FRLE to reduce administrative burden and ensure needy children receive meals. This renders FRLE counts unreliable where CEP is applicable.

We undertook an in-depth examination of CCD data to produce the **Longitudinal Imputed School Dataset (LISD)** from the 1991-92 through 2022-23 academic years (N>3 million school-year observations). We start with 1991-92 data because fewer than half of schools reported FRLE data until then. The LISD replaces missing and erroneous FRLE and race/ethnicity data with imputed values. We also provide a **Geographic Crosswalk** linking schools to geographic areas like counties or metropolitan areas to facilitate analyses. We describe our procedures below.

Missing FRLE and Race/Ethnicity Counts

We impute race/ethnicity and FRLE data when they are missing—not reported to NCES by state educational agencies. We also consider race/ethnicity and FRLE data to be missing if a school is not present in a given CCD annual school file but is present in the years before and after the missing year (0.13% of all school-year observations). Table 1 presents the percent of schools missing data on FLE or race/ethnicity enrollment counts in a given year. From 1991 to 2022 (years refer to fall of the academic year), race/ethnicity data were missing for up to 7% of schools in each year (with more missing data in the earlier years) and entirely missing in 30 of 1,664 state-years (1,664 = 32 years x 52 states or administrative units (states, Washington, DC, and the Bureau of Indian Education (BIE))).

FLE counts were missing for between 1 and 36% of schools in each year and entirely missing in 191 state-years. If FLE counts were missing and RLE counts were not, we set RLE counts to missing

and imputed (this is rare, N=355 school-year observations). If RLE counts were missing and FLE counts were not, we imputed RLE but did not set FLE counts to missing and impute—RLE counts had high missing rates when they were first reported starting in 1998. FRLE missing data were especially common prior to 2000 and after 2015, when states could report FRLE and/or direct certification counts for each school.[2] FRLE eligibility is based on student applications that include family income information. To reduce participation barriers, students can also be directly certified for school meals if their household participates in Supplemental Nutrition Food Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF), Food Distribution Program on Indian Reservations (FDPIR), or Medicaid (in some states) or if the child is migrant, homeless, in foster care, or in Head Start.[3] If states only report direct certification after 2015, we consider their FRLE data missing and impute all values (and we include direct certification counts as a predictor in our imputation model, see description below). Appendix A lists the state-years that only report direct certification data.

Implausible FRLE or Race/Ethnicity Data

We also impute FRLE or race/ethnicity data in cases where reported data appeared implausible. Our investigation of the CCD led us to identify a range of data quality issues: (1) schools erroneously reporting zero FRLE counts; (2) state-years where FLE data deviate from a state's general pattern; (3) state FRLE reporting changes; (4) district-years where FLE data deviate from a district's general pattern; (5) state-years where FRLE or race/ethnicity data are reported at the district rather than the school level; and (6) years where FLE is topcoded. We describe each case below.

(1) Schools Erroneously Reporting Zero FRLE Counts

Despite NCES reporting instructions, state educational agencies appear to sometimes report FRLE counts as zero rather than missing when true FRLE counts are unknown or unavailable, especially in early years of CCD data collection. We diagnosed this issue through visual plots of the number of schools reporting missing, zero, and non-zero FLE counts in each state over time, which revealed that state agencies appeared to oscillate between reporting unknown values as zero or missing in a given year, casting doubt that zero counts are true. For example, Figure 1 presents the plot for Alaska from 1991 to 2020. In Alaska, there are between 300-400 schools with non-zero FLE counts (blue portion of the bar) and an additional 100 to 200 schools with either missing FLE data (red) or FLE enrollment counts of zero (green) in most years. In some years, like 2003 and 2004, about 3-5% of zero or missing counts were zero and the rest were missing (note small green area). But in other years, such as 1994-1997, all zero or missing counts were recorded as

zeros, whereas all zero or missing counts were reported as missing from 1991-1993. This pattern of oscillation is evident in many other states, especially in earlier years. (Figure 1 also shows that FLE data are completely missing in AK in 1998-2001 and 2020). Therefore, we impute FLE data for schools originally reporting zero counts. If we impute observations where FLE counts were originally zero, we also impute RLE counts in those cases. After reporting begins in 1998, if RLE counts are zero, we impute RLE unless the reported FLE rate is either between 0 and 5% or greater than 95%. In these cases, it is plausible that there are zero RLE students.

It is possible that some schools truly had zero FLE students in some years. We used two criteria to identify state-years with plausibly valid zero values from 1991 to 2018 (the number of schools reporting zero FLE students declined starkly over time, so we ended our valid zero diagnosis and impute all zero values after 2018).

1. We estimated a ratio of the observed number of schools with FLE rates of 0% in a state-year to the average number of schools in each of the three percentile bins from FLE rates greater than 0% to less than 3%. If there is a large spike of schools with FLE rates of 0% compared to the lower-tail distribution of FLE rates, the zero values are likely not valid. Therefore, we require the ratio of schools with zero FLE rates to the average number of schools with FLE rates in the lowest three percentile bins to be less than 3.

2. We geocoded schools to their Census tract and linearly interpolated tract poverty rates annually between the 1990 and 2000 decennial Censuses and the 2006-10 American Community Survey (ACS), which we assigned to 2008 CCD data. We assign subsequent ACS 5-year estimates to successive CCD years, assigning 2014-18 ACS to NCES years 2016 through 2018. We estimate the mean tract poverty rate in a state-year for schools originally reporting zero FLE students. If the mean tract poverty rate is 15% or less, we assume schools reporting zero in that state-year are valid.

One hundred twenty-four state-years met both criteria, and we then examined histograms of school FLE rates and tract poverty rates to confirm our diagnosis of valid zeros. The FLE rate histograms showed much lower spikes at zero in years with potentially valid zero counts (comparing within state), and tract poverty rate histograms were skewed right (towards lower values) in years with potentially valid zero counts. We were satisfied that our criteria identified state-years with valid zero counts, and we did *not* impute schools with original FLE counts of zero in these 124 state-years (representing 6% of all original zero observations). Appendix B provides a list of the state-years where we did *not* impute observations where FLE was originally zero. If a school's FLE zero count was determined to be valid, we also did not impute original RLE counts of zero for that observation.

Overall, we impute FRLE data for up to 5% of schools in each year that appear to have erroneously reported counts as zero instead of missing.

 (2) State-Years where FLE Data Deviate from a State's General Pattern

Our inspection of FLE data revealed some year-to-year anomalies within states. For example, Figure 2a shows that in Illinois, the distribution of FLE rates across schools looks very different in 2011 than in the surrounding years. Figure 2b presents the proportion of FRLE students at the state level and suggests a possible explanation—some FLE students were coded as reduced-price lunch students in 2011, perhaps due to data entry error. Further analyses indicate that this pattern was evident across districts in IL.

We used year-to-year variation in CCD data and triangulation with U.S. Department of Agriculture (USDA) data to identify state-years with implausible FLE counts. Via email from USDA staff, we received state-level data on participation in the National School Lunch Program (NSLP) from 1999-2000 to 2019-2020. The USDA data included the count of free lunches, total lunches, and number of students served monthly in each state, which we aggregated from September through May to get academic year estimates. We divided the number of free lunches served by total lunches served, and then multiplied this rate by total number of students served to estimate the number of FLE students in each state-year. The data capture free lunch participation rather than eligibility as measured in the CCD, but we assume that changes in free lunch participation should reflect underlying changes in eligibility. We examined scatterplots of USDA v. CCD counts within states over time. Figure 3a, for example, presents a scatterplot of CCD counts of FLE students and USDA counts of students that participated in the free lunch program in Kentucky. 2002 stands out as an outlier, with CCD counts much higher than expected given the relationship between USDA and CCD counts in other years. Figure 3b shows the trend in FRLE rates reported in the CCD for KY, and 2002 clearly stands out here as well.

We regressed CCD state-level FLE counts on USDA counts within states to estimate state-year studentized residuals. We also estimated residuals three ways within the CCD data:

a. Within state, we regressed state-level FLE rate on year and generated state-year studentized residuals to identify deviations from linear time trends.
b. To account for non-linear trends, we estimated state-year studentized residuals from a cubic model within states.
c. To account for both within-state and national trends, we estimated state-year studentized residuals from a model pooling states, in which we regressed state-level FLE rate on state and year dummy variables plus an interaction between state dummies and a continuous year variable.

We then identified state-year observations where the absolute values of studentized residuals was ≥ 2, following a common rule-of-thumb for distinguishing outliers. We identified 11 state-years that were outliers on all four diagnostic indicators (USDA and the three CCD models). We examined trends in states with any outlier years and identified an additional three state-years that appeared implausible but did not meet the residual threshold on all indicators. We only had USDA data since 1999-2000, so we relied only on CCD data for earlier years and identified six state-years that are outliers on the three CCD models. Preliminary data analyses uncovered additional state-years that had implausible FRLE data but didn't meet our outlier threshold. For

example, Utah's FRLE distribution in 2006 deviated substantially from Utah's data in 2005 or 2007. Visual inspection confirmed these values as implausible. Through this method, we identified 21 state-years (listed in Appendix C) that appear to have reporting issues for which we impute all FRLE data.

(3) State Reporting Changes

In Texas, FLE counts rose drastically from 2011-12 to 2012-13. CCD documentation reports that TX changed their reporting standard in 2012—the state began reporting both FRLE and "other economic disadvantage" counts in the FRLE count in 2012-13.[4] Therefore, we impute FRLE data for all schools in TX beginning in 2012-13.

(4) District-Years where FLE Data Deviate from District Patterns

Preliminary analyses also led to the discovery of data anomalies in FLE reporting in large districts in Texas prior to 2012-13. Figure 4 shows that, among the 20 largest districts in TX, several have implausible trends in FLE rates. For example, the FLE rate in Brownsville ISD (a district serving a high-poverty city) steadily declines to near 0 before returning to near 80% when the reporting changes in TX in 2012. Similarly, Houston between 2006 and 2011 and San Antonio between 2007 and 2012 show odd patterns. We used these plots to visually identify implausible years in these districts, and we imputed FRLE data for all schools in those district-years (N=54 district-years; 5,063 school-year observations).

(5) District-Level FRLE or Race/Ethnicity Data Reporting

In some states in some years, FRLE rates are reported at the district level rather than at the school level. In these instances, we imputed FRLE data. The affected state-years are Mississippi from 1991-2002, Nebraska from 1991-1993, Vermont in 1995, and Wisconsin from 1991-2000. Similarly, in Tennessee in 1997, schools have identical race/ethnicity rates within district, so we impute race/ethnicity data. After imputing in these cases, we adjust the school-level imputed values so that their weighted average sums to the original district-level reported values.

(6) FRLE Data Topcoding

Prior to 1999, FLE rates were topcoded at 0.95. We set FLE counts to missing and impute if the free lunch rate = 0.95 in these years. (If the imputed value is less than 0.95, we replace it to equal 0.95.) If we impute FLE counts due to topcoding in 1998, we also impute RLE counts and adjust imputed values to be less than .05 if needed.

Community Eligibility Provision

---

[4] Keaton, P. (2016). *Documentation to the NCES Common Core of Data Public Elementary/Secondary School Universe Survey: School Year 2012-13 Provisional Version 2a (NCES 2015-009rev).* U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved 9/16/2020 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015009rev.

In 2010, Congress authorized the Community Eligibility Provision (CEP) as part of the Healthy, Hunger-Free Kids Act. CEP was phased in over three years beginning in 2011-12 with a three-state pilot (IL, KY, MI) and was implemented nationwide by the 2014-15 school year. CEP allows schools and school districts in which 40% or more students were direct certified for free meals in the prior school year to classify all students as FLE without collecting household applications. Eligible schools therefore may report 100% FLE eligibility to the CCD even when not all individual students are eligible. The USDA and state departments of education provide lists of schools and school districts eligible for CEP since 2011 and the CCD reports participation since 2014.[5] Through a combination of these data sources, we flag schools in years they are CEP-eligible and impute their FRLE counts. Seven percent of all school-year observations and 23% of observations since 2014 are CEP eligible.

In sum, we impute FRLE and race/ethnicity data when it is missing, anomalous, or when a school is eligible for CEP. The data includes flag variables indicating when data are imputed and the reason for imputation (see codebook at http://edopportunity.org/segregation/get-the-data). Table 1 shows the share of observations for which we impute FRLE and race/ethnicity data by year. By 2020, we impute FRLE data for the majority of schools, pointing to the need for improved economic data collection efforts in the future.

Preparing the Data for Imputation

Longitudinal School IDs

Before imputing missing and implausible race/ethnicity and FRLE data, we harmonized NCES school ID numbers (variable **ncessch**) that changed over time. First, we relied on longitudinal CCD files created by Don McLaughlin, which correct for ID numbers that changed from 1986 to 1998.[6] We observe ID changes in 1.4% of schools (0.6% of school-year observations) from 1991-1998, mainly due to changes in district ID (e.g., a district merges or splits from another district).

We built on this crosswalk to create consistent longitudinal school IDs after 1998. In 2016-17, NCES adopted a unique 7-digit ID (2 digit FIPS state code + 5 digit school ID) to consistently identify schools over time (a full NCES ID is 2 digit FIPS state code + 5 digit school district + 5 digit school ID; this 7-digit ID omits the district so a school can be tracked if its district changes).[7] NCES assigns new school IDs sparingly—if the grade span changed by more than 3 grades (not including PK or K), if the school's physical location and attendance area changed significantly, or if two schools of about the same size or with different grade spans merge. NCES staff reviews and

---

[6] McLaughlin, Don. (2003). *CCD Data File: Thirteen Year Longitudinal Common Core of Data Non-Fiscal Survey Database: School Years 1986-87—1998-99*. Retrieved 1/19/2021 from https://nces.ed.gov/ccd/CCD13yr.asp
[7] Keaton, P. (2021). "Accessing the Common Core of Data (CCD)" NCES Blog, 1/4/21. Retrieved 3/31/21 from https://nces.ed.gov/blogs/nces/post/accessing-the-common-core-of-data-ccd

verifies state or district requests to assign a new NCES ID.[8] From 2016-17 onwards, the 7-digit ID is unique. Prior to 2016-17, there were duplicate school IDs within states if more than one district in a state assigned a school the same 5-digit school ID. We apply crosswalks provided by NCES to correct for this, which start in 2014-15.[9] From 1991-2014, 0.22% of school 5-digit IDs were not unique within state, so we apply the same method as the NCES crosswalks, working backwards and correcting for duplicate school IDs within state by making a minor change, e.g., changing a leading 0 to 1, in school IDs. In total, there are 5,498 schools (4.34% of school-year observations) with original NCES IDs that vary over time, which the 7-digit ID (variable **nces7_final**), corrects. This method identifies schools over time, which is important for our imputation method, but scholars interested in district changes will also want to examine district IDs alongside nces7_final to identify district secession or merges.

Longitudinal Race Categories

CCD race/ethnicity categories change beginning in 2008-09 from five (American Indian or Alaska Native, Asian or Pacific Islander, Black, Hispanic, or White) to seven mutually exclusive categories (American Indian or Alaskan Native, Asian, Black or African American, Hispanic/Latino, Native Hawaiian or Other Pacific Islander, White, or Two or More Races). Schools began transitioning to the new reporting categories in 2008-09 and were required to use the seven categories by 2010-11. We created six longitudinal race categories: American Indian or Alaska Native, Asian or Pacific Islander (combines Asian and Native Hawaiian or Other Pacific Islander after the adoption of seven categories), Black, Hispanic, White, and multiracial. Prior to 2008-09, zero students are coded as multiracial.

New York City Schools

Starting in 2005-06, CCD assigned New York City Public Schools to separate sub-districts. See Appendix D for a list of NCES district IDs. For consistency over time, we assign these schools to the original New York City Public Schools NCES ID (3620580). We also changed the full school ID (ncessch) to reflect this district in the first 7 digits (nces7_final is unaffected because it excludes district). **ncessch_orig** reflects the original ncessch ID. (ncessch_orig is missing for schools that were entirely imputed, i.e., schools missing entirely from the CCD file but that were present in the adjacent years).

Detroit Public Schools

In 2016, the city of Detroit's public schools were reorganized and renamed from the Detroit City School District (Local Educational Agency (LEA) ID number (**leaid**) 2612000) to the Detroit Public

---

[8] FS029 – Directory File Specifications (SY 2021-22), U.S. Department of Education, Washington, DC: ED*Facts*. Retrieved 4/1/22 from https://www2.ed.gov/about/inits/ed/edfacts/index.html

[9] "Document 4: NCES School ID Crosswalk SY 2015-16 to 2016-17"; "Document 3: NCES School ID Crosswalk SY 2014-15 & SY 2015-16." Retrieved 1/20/21 from https://nces.ed.gov/ccd/reference_library.asp

Schools Community District (leaid 2601103). To use a consistent ID over time, we replaced leaid and ncessch to reflect the new district ID 01103 in all years (ncessch_orig is unchanged).

Additional School Data

We retain several CCD school-level variables on the imputed file. More information about each can be found in CCD's documentation. **Type** identifies whether the school was regular, special education, vocational, other/alternative, or some other reportable program. **Grdlo** and **grdhi** are, respectively, the low and high grade served by a school; instructional **level** (e.g., elementary, secondary) is provided since 2008-09.[10] **Charter** and **magnet** indicators are available since 1998-99. Each of these variables can vary over time; we cannot determine whether these are true changes or data errors. An indicator for **virtual** schools is available starting in 2012. We construct an indicator (**bie**) for Bureau of Indian Affairs schools starting in 1998, when CCD identifies such schools using their own state FIPS codes. We construct an indicator for schools for the deaf or blind if the words blind, deaf, or hearing are in the school name (**deafblind**)**.**

Auxiliary Covariates in the Imputation Model

The imputation model includes school enrollment data from the CCD as well as auxiliary variables in select years: school economic disadvantage rate (ED), tract child poverty rate, percent of students direct-certified as FRLE, and school Title I status. ED rate is available from 2008 to 2018 from EdFacts data provided to the Stanford Educational Opportunity project. States report ED rates to EdFacts using state-specific definitions as used for accountability purposes; in some states, ED correlates highly with FLE rates, but other states use more or less expansive definitions of student disadvantage. To obtain tract child poverty rate, we spatially join schools to the Census tracts in which they are located using geodata provided in the CCD. We link schools to 1990 and 2000 Census data based on their location in the 1991 and 2000 CCD and to 2006-10 to 2018-22 five-year American Community Survey (ACS) estimates based on their location in 2008 to 2020, respectively. Direct certification rates are provided by the CCD from 2016 to 2022. School Title I status is reported in the CCD, and we use data from 2010 to 2019 (data are available since 1998 but have high missingness rates prior to 2010). We create two dummy variables indicating whether a school had schoolwide Title I or targeted assistance programs.

Imputation Procedures

We use Stata's -mi- command suite to impute missing racial/ethnic and FRLE rates and total enrollment using chained regression models. The imputation model includes these variables and auxiliary variables described above: school ED rate (2008-2018), tract child poverty rate (1991, 2000, 2006-2022 (five-year aggregations)), direct certification rate (2016-2022), and Title I status (2010-2019). Our data structure is "wide" format, with one observation per school and year-specific variables. We create ten imputed datasets.

---

[10] The level variable changed in SY 2017-18. See Document 11 "Changes to CCD-assigned schools and levels" https://nces.ed.gov/ccd/reference_library.asp

Because we impute racial/ethnic and FRLE *percentages*, we convert racial/ethnic and FRLE rates to logits and enrollment to log enrollment. (To avoid missing logit or log values, we replace (valid) racial/ethnic and FRLE rates of zero as 1/(8*total enrollment) and values of 1 as 1-(1/(8*total enrollment)). The imputed values we obtain are thus in logit or log scale, so we take the inverse logit of imputed values to recover race/ethnicity and FRLE rates and exponentiate to recover total enrollment. For all imputed values, we rescale racial/ethnic and FRLE rates by total number of students so the imputed composition variables sum to 100%. (We do not impute percent non-FRLE; we estimate it from the imputed values of FLE and, after 1998, RLE.)

Segregation Explorer Geographic Crosswalk

We created a geographic crosswalk file for use with the LISD that links schools to geographies of interest.

Geographic School District and State IDs

Some schools—chiefly charter schools and BIE schools—are located in a school district's geographic boundaries but are administered by their own LEA and thus have a different NCES administrative school district ID (leaid) than that of the school district in which it is located. For some analyses, it is desirable to identify all schools located in a school district's geographic boundaries. We assign schools to a geographic district starting in 1998-99 (the year that charter and magnet indicators became available) by spatially joining latitude and longitude coordinates of schools' physical locations to annual school district boundary shape files provided by NCES Education Demographic and Geographic Estimates (EDGE). Latitude and longitude data are not available prior to 2000-01; we assign schools in 1998-99 and 1999-2000 to the same geographic district to which they are geocoded starting in 2000. Some schools in the 1998 or 1999 data did not exist after 2000; we geocode these schools' street locations (provided by CCD) to obtain latitude and longitude and then spatially join them to school districts. Schools can be spatially joined to unified, elementary, and/or secondary school districts; for non-unified districts, we use high- and low-grade data to assign schools to either elementary or secondary districts. We created the variable **geolea** equal to a school's geographic district. For most schools, this is the same as their leaid; for charter and BIE schools, geolea is the geographic district in which a school was located, since their LEA is not geographically-based. We created a geographic state ID based on geolea (**geost**)—BIE schools (reported in the CCD since 1998-99) are assigned a FIPS code of 59, so the geographic state ID reflects the state in which they are actually located (compare to **admst,** which reflects ncessch and leaid). Prior to 1998, geolea equals administrative district (leaid).

County

The CCD provides data on schools' counties (**conum)** since 2002. Prior to 2002, and if a school is missing county data, we assign schools to their administrative district's county as defined by the CCD, which is the county in which most of the LEA's schools were located, weighted by student

enrollment. If both school and district county data are missing in a given year, we assign the school to its modal county over time.

Metropolitan Areas and Commuting Zones

We assign schools to metropolitan areas and commuting zones based on their county. The Office of Management and Budget (OMB) periodically defines metropolitan and micropolitan areas, and we link counties to metropolitan areas using 2003, 2013, and 2023 delineation files.[11] We provide the school's core based statistical area (CBSA) code (**CBSACode03 CBSACode13 CBSACode23**) and name (**CBSATitle03 CBSATitle13 CBSATitle23**), its status as either a metropolitan (CBSA associated with an urban area with a population of at least 50,000) or micropolitan (CBSA associated with an urban area between 10,000 and 50,000 population) area (**metrostatus03 metrostatus13 metrostatus23**), and, for schools located in large CBSAs, its metropolitan division code (**MetroDivCode03 MetroDivCode13 MetroDivCode23**) and name (**MetroDivTitle03 MetroDivTitle13 MetroDivTitle23**) (divisions are delineated for metropolitan CBSAs containing a single Urban Area with a population of at least 2.5 million). Schools without values on these variables are located in counties outside of CBSAs.

We also assign schools to their 1980, 1990, 2000, and 2010 commuting zones (CZs) based on schools' county identifiers, drawing on CZ definitions from the USDA (**CommutingZoneID1980 CommutingZoneID1990 CommutingZoneID2000 CommutingZoneID2010**).[12]

---

[11] https://www.census.gov/programs-surveys/metro-micro.html
[12] 1980 through 2000: https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/; 2010: https://sites.psu.edu/psucz/data/

Table 1. Missing and Imputed School-Year Observation Rates, 1991-2022

Year (Fall)	Missing FLE	Missing Race/Ethnicity	Imputed FLE	Imputed RLE
1991	35.9%	6.7%	40.8%	--
1992	30.4%	3.9%	40.3%	--
1993	24.2%	1.2%	33.9%	--
1994	22.4%	1.0%	30.7%	--
1995	24.1%	1.1%	35.1%	--
1996	21.1%	3.9%	29.2%	--
1997	18.7%	1.1%	27.1%	--
1998	19.3%	1.1%	27.0%	44.0%
1999	16.7%	2.9%	22.8%	25.3%
2000	14.7%	2.1%	21.8%	23.8%
2001	13.3%	2.2%	17.7%	20.0%
2002	13.7%	2.3%	19.8%	22.0%
2003	15.6%	2.3%	18.5%	20.9%
2004	13.3%	3.3%	21.7%	24.3%
2005	7.5%	0.7%	12.3%	15.5%
2006	4.3%	1.0%	9.1%	13.1%
2007	8.2%	1.0%	11.2%	14.8%
2008	3.8%	1.2%	7.5%	11.0%
2009	4.8%	1.0%	10.9%	14.2%
2010	1.6%	1.0%	6.7%	10.5%
2011	10.3%	1.3%	21.0%	24.1%
2012	2.3%	1.3%	16.2%	20.8%
2013	2.6%	1.3%	17.1%	22.0%
2014	5.1%	1.0%	25.7%	30.1%
2015	7.6%	1.1%	29.7%	34.7%
2016	10.6%	0.8%	34.9%	39.1%
2017	9.3%	0.8%	35.5%	40.1%
2018	11.4%	0.7%	39.8%	44.3%
2019	11.3%	0.7%	40.6%	45.0%
2020	23.8%	0.6%	51.3%	55.6%
2021	23.3%	0.3%	50.7%	57.3%
2022	17.1%	0.0%	46.7%	51.8%

Note: "Missing" refers to data initially missing in raw CCD data. FLE = free lunch eligible; RLE = reduced lunch eligible (first reported in 1998). Imputed rates reflect school-year observations imputed for all reasons, not just missing (e.g., invalid zeros, CEP eligibility, or erroneous data).

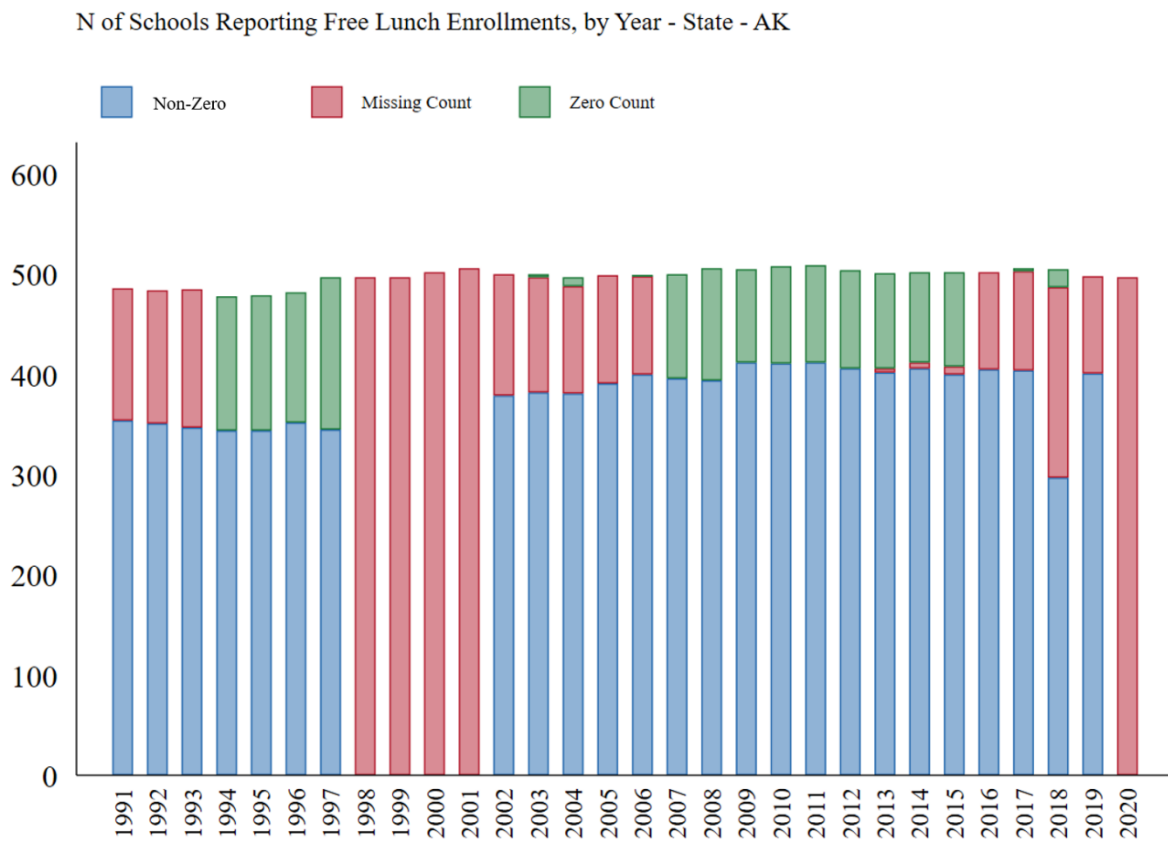**Figure 1. Number of Schools by FLE Count Type (Non-zero, Missing, or Zero), Alaska, 1991-2020**
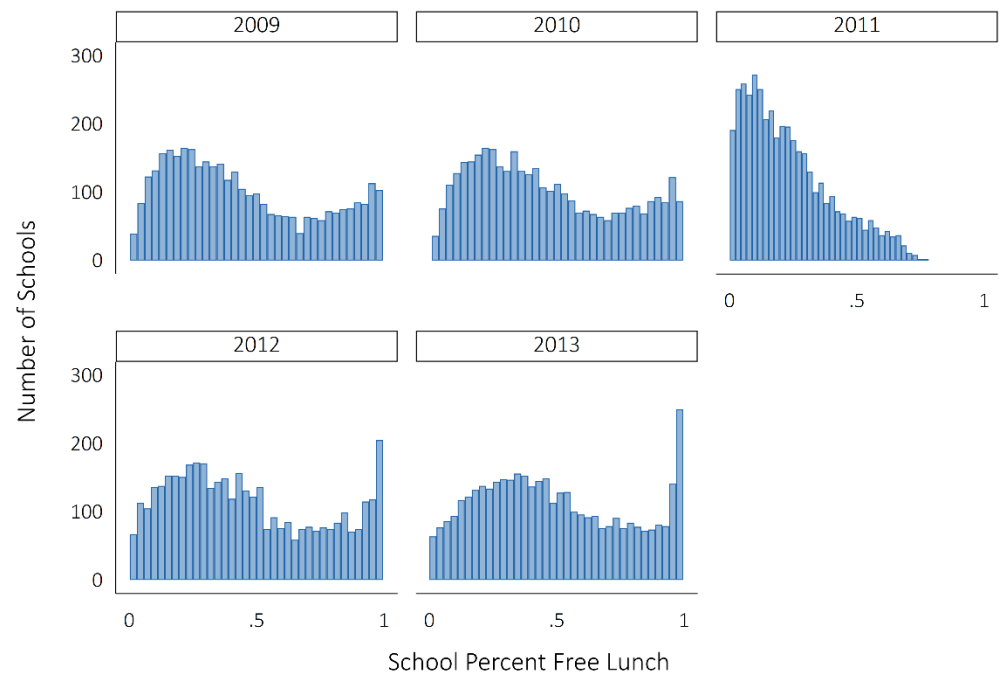
N of Schools Reporting Free Lunch Enrollments, by Year - State - AK

Figure 2a. Distribution of School-level FLE Rates, IL, 2009-2013

Graphs by year

Figure 2b. State-level FLE and Reduced-Price Lunch Rates, IL, 1991-2019
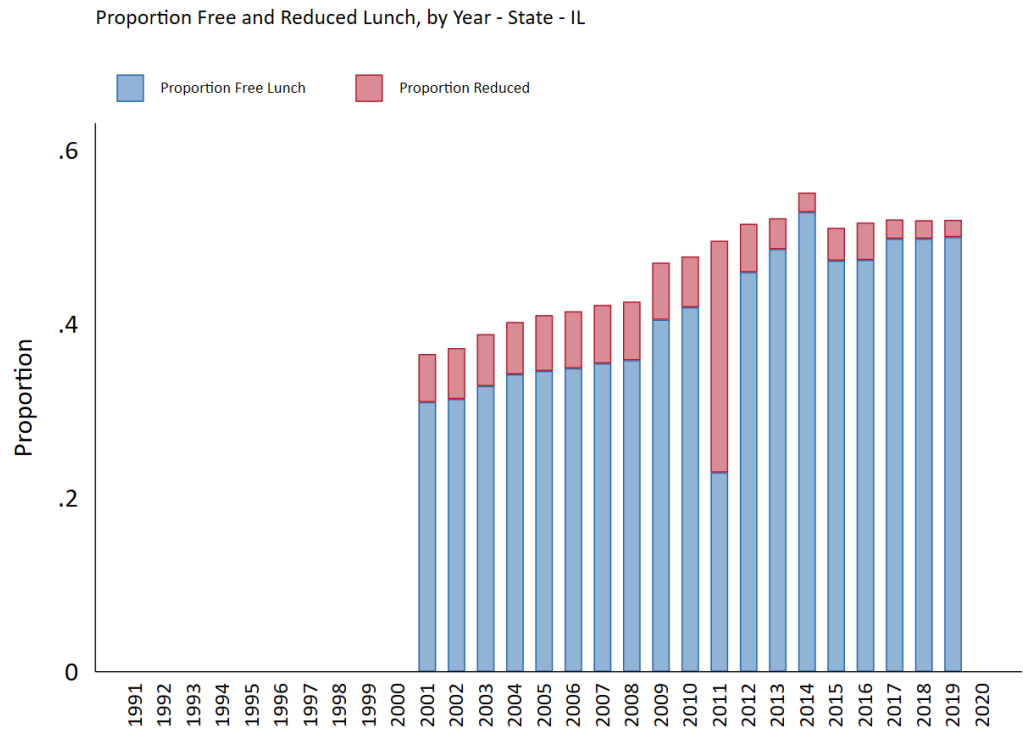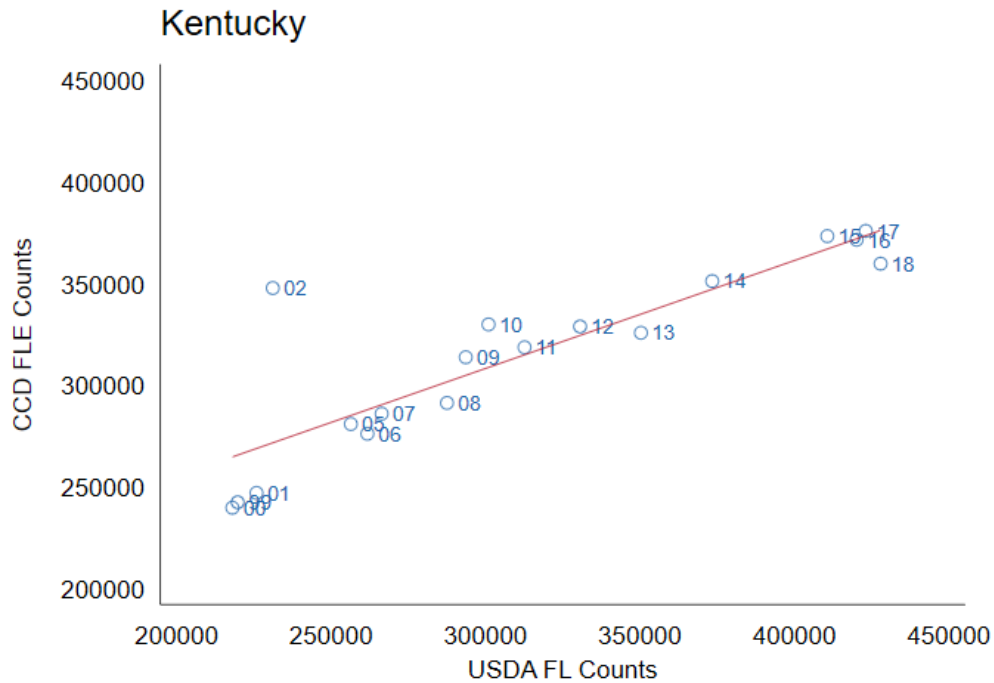
Figure 3a. CCD FLE counts v. USDA FL participation counts, KY, 1999-2018

Note: Each circle represents a year identified by its last two digits

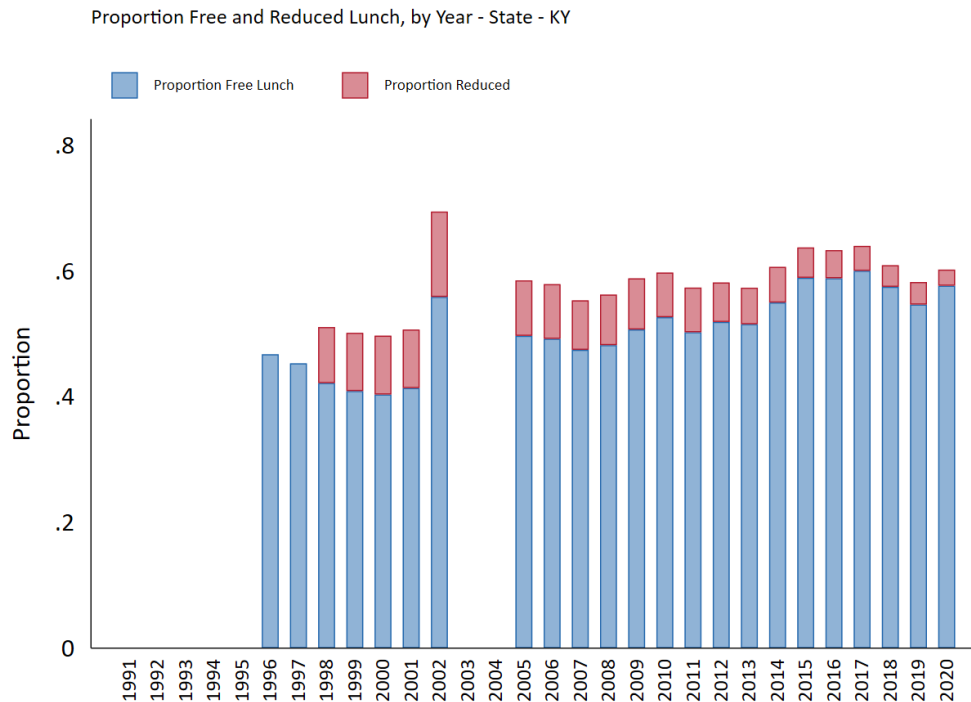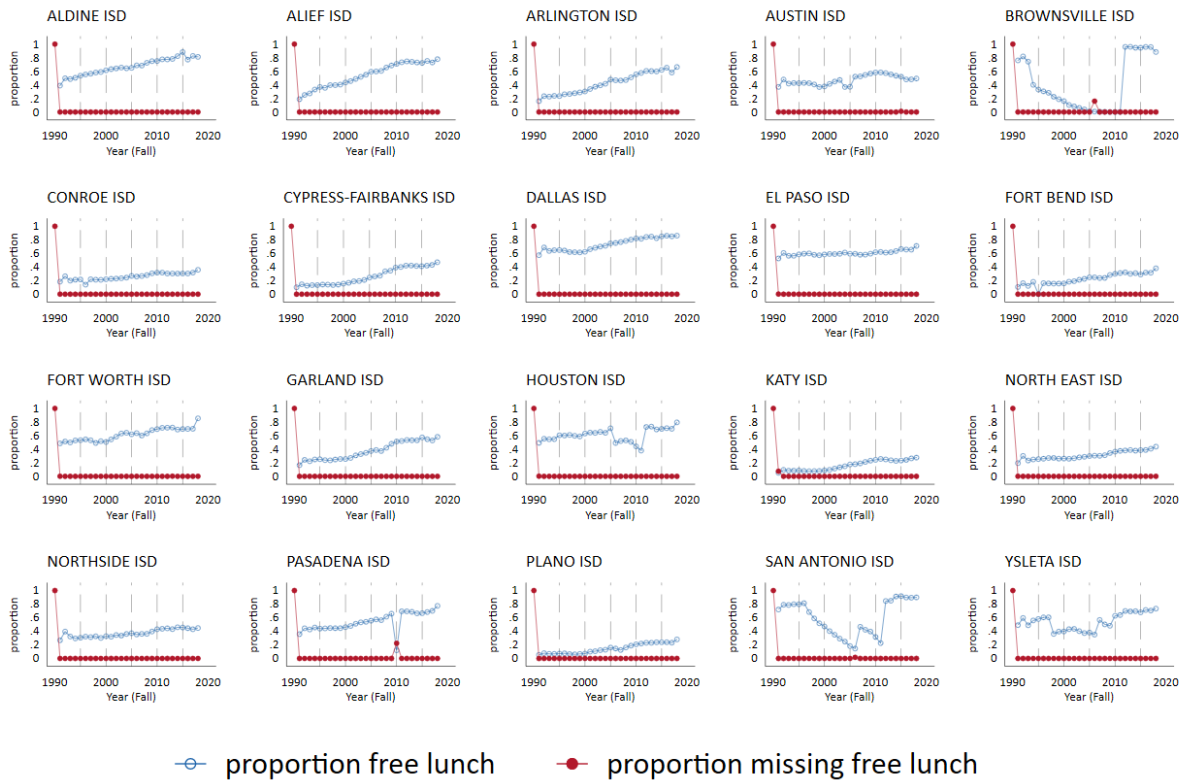**Figure 3b. Trends in CCD FLE rate, KY, 1991-2020**

Figure 4. Trends in FLE Rates, 20 Largest Texas Districts

trends in free lunch rates, 20 largest TX districts

proportion free lunch   proportion missing free lunch

Appendix A: List of state-years with fully imputed free/reduced price lunch data due to direct certification reporting only

Year	State
2016	DE, DC, MA, TN
2017	DE, DC, MA, TN, WV
2018	DE, DC, MA, OK, TN
2019	DE, DC, MA, OK
2020	DE, DC, MA, MT, OH, OK, TN
2021	DE, DC, GA, MA, MT, OH, OK, OR, TN, WV
2022	DE, DC, MA, MT, OK, TN, WV

Notes: List only includes states or administrative units included in the imputed data (50 states, Washington, DC, and BIE).

Appendix B. List of state-years with valid zero FLE counts

Year	State
1992	LA, MD, NJ, NV, WI
1993	ME, NV, WI
1994	MD, NV, WI
1995	MD, ME, WI
1996	MD, ME, NV, WI
1997	CT, MD, WI
1998	CT, MD, ME, NV, TX, UT
1999	MA, MD, ME, PA, TX, WI
2000	DE, GA, MA, MD, ME, NC, NJ, NV, TX, WI
2001	GA, IA, MD, ME, NC, NV, TX, VA,
2002	AL, GA, MD, ME, NV, RI, TX, VA
2003	AZ, CO, GA, MA, MD, ME, NJ, TX
2004	MD, ME, TX
2005	CO, KY, ME, NY,
2006	LA, NC, NH, NY, TN, WI
2007	CA, KY, NH, RI, TN
2008	GA, IA, KY, ME, NJ, OK, PA, RI
2009	CA, CO, KY, ME, NC, NJ, NM
2010	AR, LA, OK, PA
2011	GA
2012	CA, NC, TN
2013	IL, NC
2014	CA
2015	CA, CO, IA, MD
2016	NC, WI
2017	CA, KS, LA, MI
2018	CA, MD, MI

Note: The number of schools reporting zero FLE students declined starkly over time so we ended our valid zero diagnosis and impute all zero values after 2018.

Appendix C. List of state-years with fully imputed free/reduced lunch data due to data quality issues

Year	State
1991	MS, NE, WI
1992	CT, MS, NE, WI
1993	HI, MS, NE, WI
1994	HI, MS, WI
1995	HI, MO, MS, VT, WI
1996	MS, WI
1997	MS, VT, WI
1998	MS, WI
1999	MS, WI
2000	MS, WI, WY
2001	MS
2002	KY, MS
2004	NY
2006	UT
2009	NY
2011	CA, IL, ME
2012	TX, UT
2013	DC, TX
2014	DC, TX
2015	DC, TX
2016	AR, MD, TX
2017	AR, TX
2018	TX
2019	TX
2020	TX
2021	TX
2022	TX

Note: These state-years include those that deviate from the state's general pattern, state reporting changes, and state-years where district-level, rather than school-level, rates are reported.

Appendix D. NCES LEAIDs Recoded to New York City District LEAID 3620580

Table D-4   Local education agency identification numbers (LEAID) used in the Common Core of Data (CCD) for the 33 geographic
districts associated with the New York City Public Schools district: School year 2010–11

Name of district	CCD LEAID
New York City Geographic District #1	3600076
New York City Geographic District #2	3600077
New York City Geographic District #3	3600078
New York City Geographic District #4	3600079
New York City Geographic District #5	3600081
New York City Geographic District #6	3600083
New York City Geographic District #7	3600084
New York City Geographic District #8	3600085
New York City Geographic District #9	3600086
New York City Geographic District #10	3600087
New York City Geographic District #11	3600088
New York City Geographic District #12	3600090
New York City Geographic District #13	3600091
New York City Geographic District #14	3600119
New York City Geographic District #15	3600092
New York City Geographic District #16	3600094
New York City Geographic District #17	3600095
New York City Geographic District #18	3600096
New York City Geographic District #19	3600120
New York City Geographic District #20	3600151
New York City Geographic District #21	3600152
New York City Geographic District #22	3600153
New York City Geographic District #23	3600121
New York City Geographic District #24	3600098
New York City Geographic District #25	3600122
New York City Geographic District #26	3600099
New York City Geographic District #27	3600123
New York City Geographic District #28	3600100
New York City Geographic District #29	3600101
New York City Geographic District #30	3600102
New York City Geographic District #31	3600103
New York City Geographic District #32	3600097
NYC Special Schools District 75	3600135

SOURCEL U.S. Department of Education, National Center of Education Statistics, Common Core of Data (CCD), "Local Education
Agency Universe Survey," 2010–11, Version Provisional 2a.