Education Demographic and Geographic Estimates (EDGE) Program

Geocodes: Public Schools and Local Education Agencies

ies NATIONAL CENTER FOR EDUCATION STATISTICS

Institute of Education Sciences

Education Demographic and Geographic Estimates (EDGE) Program

Geocodes: Public Schools and Local Education Agencies

MAY 2018

Douglas E. Geverdt
National Center for Education Statistics

U.S. DEPARTMENT OF EDUCATION

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

> NCES, IES, U.S. Department of Education
> 550 12th Street SW
> Washington, DC 20202

May 2018

The NCES Home Page address is http://nces.ed.gov and the Publications and Products address is http://nces.ed.gov/pubsearch. This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Acknowledgements

<p style="text-align:center">Contents</p>

1.0   Purpose
The National Center for Education Statistics (NCES) Education Demographic and Geographic Estimates (EDGE) program develops data resources and information to help data users investigate the social and spatial context of education. School point locations (latitude/longitude values) are a key component of the NCES data collection. These data are needed to address a variety of spatially-oriented tasks and research questions. They provide information needed to construct NCES school-based surveys; they provide indicators needed to help determine program eligibility; and they provide the foundation for determining geographic associations with other types of entities.

This document describes the content of geocode files developed for schools and LEAs reported in the Common Core of Data (CCD) school and agency universe. The CCD universe is a comprehensive collection of public elementary and secondary schools and the LEAs that operate and support them. The CCD provides administrative data about enrollment, staffing, and program participation for schools, LEAs, and states. State education agencies (SEAs) report these data to the U.S. Department of Education in a series of file submissions throughout the school year. The geocode files include the unique school and agency identifiers assigned by CCD, and this shared ID can be used to integrate the geocodes with the CCD directory files. Additional discussion and documentation for the CCD school and agency files is available at [NCES's CCD website](#).

2.0   Locating schools and agency administrative offices
The estimated location of schools and agency administrative offices is primarily derived from the physical address reported in the CCD directory files. The reported address is compared to the location of known address points and street locations using a process known as address geocoding. A geocoder is an algorithm that parses parts of an address (structure number, street name, city, and ZIP code) and attempts to match that information to addresses stored in geographic databases that contain verified latitude/longitude values. Commercial and non-commercial address point information is continuously updated and improved, so the estimated point location for an address may potentially change, even if the reported address has not. Estimated point locations may also change as part of regular manual review and curation by program staff to ensure that school points are positioned appropriately. All reported cases with a missing address are manually reviewed, and the results of all reviewed cases are maintained in subsequent years as long as the reported address remains stable. Schools that report a new address after being manually reviewed are reassessed so that the point location can be repositioned. Once all cases in the directory file have been assigned a final latitude and longitude value, these estimated locations are used to assign additional geographic indicators and identifiers.

3.0   Identifying geographic associations
One of the primary purposes for identifying school and agency point locations is to identify the spatial relationship between these locations and other types of geographic areas like counties, core based statistical areas (CBSAs), congressional districts, and other jurisdictions or statistical areas. These associations are determined using a geographic information system (GIS), an application that can manage and evaluate geographic data to identify spatial relationships. The associations are represented using unique geographic identifiers from the Federal Information Processing Series (FIPS).

NCES develops geographic indicators by associating school point locations with boundaries from the U.S. Census Bureau's TIGER (Topologically Integrated Geographic Encoding and Referencing) database, a continuously updated geographic database maintained by the Census Bureau's Geography Division that provides an authoritative collection of legal and statistical area boundaries for the United States. NCES also uses Census TIGER boundaries to develop NCES Locale boundaries that are used to assign Locale

codes to schools and local education agencies (LEAs). The native geographic coordinate system for the TIGER boundaries is North American Datum, 1983 (NAD83), and the data vintage corresponds to the starting year of the academic year. Additional documentation can be found at the [Census Bureau's TIGER website](#).

4.0   LEA geographic associations

Most LEAs are independent local governments that have authority to determine their geographic boundaries. These boundaries may or may not be consistent with boundaries for other types of legal and statistical areas. As a result, LEAs may have multiple spatial associations with other types of geographic areas. For example, an LEA boundary may include territory in two different counties or intersect three different congressional districts. The LEA geocode file assigns a single geographic association for counties, CBSAs, combined statistical areas (CSAs), New England city and town areas (NECTAs), and congressional districts to each agency based on the reported location of its administrative office. These associations are useful, but not necessarily complete. A complete set of associations is available from the EDGE School District Geographic Relationship Files (GRF). The GRFs are based on the Census Bureau's TIGER database, and they provide a separate record for each part of a school district that is uniquely associated with another type of geographic area. The files are designed to help answer spatially-oriented questions like: How many congressional districts are represented in a school district? How many school districts in the U.S. serve more than one county? Which ZIP codes or census tracts are included in a school district? And what portion of a school district is contained within a metropolitan area? The geographic associations on the LEA location file are most useful for identifying spatial relationships based on the location of the agency administrative office, or in cases where a single association or a primary association is required. The EDGE GRFs are most useful when a full set of geographic associations is needed. More information about the EDGE GRF tables is available on the [NCES's EDGE Geographic Relationship Files website](#).

5.0   Record layout and item descriptions

5.1   Record layout

Public School Geocode File (EDGE_GEOCODE_PUBLICSCH_*YYYY*) Record Layout

Field	Length	Type	Description
NCESSCH	12	String	School identification number
NAME	60	String	Name of institution
OPSTFIPS	2	String	FIPS state code for operating state
STREET	50	String	Reported location street address
CITY	25	String	Reported location city
STATE	2	String	Reported location state
ZIP	5	String	Reported location ZIP code
STFIP	2	String	State FIPS
CNTY	5	String	County FIPS
NMCNTY	100	String	County name
LOCALE	2	String	Locale code
LAT	10.6	Double	Latitude of school location
LON	11.6	Double	Longitude of school location
CBSA	5	String	Core Based Statistical Area
NMCBSA	100	String	Core Based Statistical Area name
CBSATYPE	1	String	Metropolitan or Micropolitan Statistical Area indicator
CSA	3	String	Combined Statistical Area
NMCSA	100	String	Combined Statistical Area name
NECTA	5	String	New England City and Town Area
NMNECTA	100	String	New England City and Town Area name
CD	4	String	Congressional District
SLDL	5	String	State Legislative District - Lower
SLDU	5	String	State Legislative District - Upper
SCHOOLYEAR	9	String	School year

Public Local Education Agency Geocode File (EDGE_GEOCODE_PUBLICLEA_*YYYY*) Record Layout

Field	Length	Type	Description
LEAID	7	String	Agency identification number
NAME	60	String	Name of agency
OPSTFIPS	2	String	FIPS state code for operating state
STREET	50	String	Reported location street address
CITY	25	String	Reported location city
STATE	2	String	Reported location state
ZIP	5	String	Reported location ZIP code
STFIP	2	String	State FIPS
CNTY	5	String	County FIPS
NMCNTY	100	String	County name
LAT	10.6	Double	Latitude of agency office location
LON	11.6	Double	Longitude of agency office location
CBSA	5	String	Core Based Statistical Area
NMCBSA	100	String	Core Based Statistical Area name
CBSATYPE	1	String	Metropolitan or Micropolitan Statistical Area indicator
CSA	3	String	Combined Statistical Area
NMCSA	100	String	Combined Statistical Area name
NECTA	5	String	New England City and Town Area
NMNECTA	100	String	New England City and Town Area name
CD	4	String	Congressional District
SLDL	5	String	State Legislative District - Lower
SLDU	5	String	State Legislative District - Upper
SCHOOLYEAR	9	String	School year
LOCALE	2	String	Locale assigned to district
LOC_OTHER	8	Number	Percentage of enrolled students attending schools in a locale other than the locale assigned to district
LOC_COUNT	8	Number	Number of locales assigned to schools operated by district
LOC_SCH	50	String	List of locales assigned to schools operated by district

5.2 School and agency identification number (NCESSCH, LEAID)

Each record of the public school geocode file contains a unique NCES school identifier comprised of three components. The first two digits identify the state FIPS code. The next five digits identify the local education agency (LEA) code. The last five digits contain the school code. The combined twelve-digit ID provides a unique identifier for all schools on the file. Each record of the LEA location file includes a unique seven-digit LEA identifier comprised of the state FIPS code and the LEA code.

5.3 Name of institution or agency (NAME)

Name of school or LEA provided by the reporting agency.

5.4 FIPS state code for operating state (OPSTFIPS)

Schools and agencies reported in the CCD directory are identified by the state responsible for their administration, which is not necessarily the state where an agency or school may be located. Some SEAs

or charter school administrators occasionally operate schools in a neighboring state to accommodate unique program needs. This may also occur with schools that are centrally administered by the Bureau of Indian Education (BIE) and the Department of Defense Education Activity (DoDEA). BIE-operated schools are assigned an operation state code of 59 and DoDEA-operated schools are assigned an operation state code of 63. In cases of inter-state administration, the state abbreviations used in the physical location address will not necessarily reflect the SEA responsible for administering the school or agency.

5.5 Reported location street address (STREET)
The reported location address for the school or agency administrative office. If no location address is available on the directory file, the address fields are set to M.

5.6 Reported location city (CITY)
The reported location city for the school or agency administrative office.

5.7 Reported location state (STATE)
The reported location state for the school or agency administrative office.

5.8 Reported location ZIP code (ZIP)
The reported location ZIP code for the school or agency administrative office.

5.9 State FIPS (STFIP)
STFIP represents the two-digit FIPS code of the state where the school or agency administrative office is located, based on the latitude and longitude values of the school or agency. For a list of FIPS codes, see the Census Bureau's FIPS code website.

5.10 County FIPS (CNTY)
The county code is a five-digit code that uniquely identifies all counties in the United States. It includes a two-digit state FIPS prefix, followed by a three-digit county identifier. The county code is assigned to a school or agency administrative office using the latitude and longitude values of the school or agency. A county code is available for all counties and other geographic entities that function as county equivalents.

5.11 County name (NMCNTY)
The county name includes the legal area description reported by the U.S. Census Bureau for the county where a school or agency administrative office is located. In states with county equivalents, this legal descriptor may not be identified as "county." This includes independent cities in Virginia, parishes in Louisiana, and census areas and boroughs in Alaska.

5.12 Locale (LOCALE)
The locale code is a general geographic indicator that classifies the type of area where a school is located. Locale codes are based on a twelve-category framework that includes four primary classifications (city, suburban, town, and rural) that each have three sub-types. NCES uses locale codes for general description, analysis, sampling, and other statistical purposes. Locale classifications are primarily derived from urban and rural definitions determined by the U.S. Census Bureau. More discussion of the locale criteria is available in the EDGE Locale Boundaries File Documentation.

Locale assignments for schools are based on the estimated location of a school building. If a school does not have a locale code (i.e., is outside of the United States), this value will be 'N' for Not Applicable.

Unlike NCES school locale assignments that are based on the physical location of the school, NCES school district locale assignments are based on enrollment-weighted locale assignments of the schools operated by the district. If a single locale accounts for the majority of students in schools, that locale is also assigned to the district. If the district lacks a majority locale, the assignment is determined by first identifying whether a majority of students in schools are attributable to a basic type (city, suburban, town, rural), and the district is assigned to the subtype that accounts for a plurality of students in its schools. If the district lacks a majority basic type, the district is assigned to the locale that accounts for a plurality of enrollment-weighted schools. Although most district assignments are based on a majority locale, many school systems—particularly large, county-based districts in the mid-Atlantic and Southern states—contain substantial geographic variation that is not reflected by the single locale assigned to the district in CCD. LOCALE has been set to 'N' for all DoD agencies, as many operate schools not located in the United States and enrollment numbers are not reported. Note that unlike the locale assignment, school district geocodes for county, CBSA, and other types of geographic areas are based on the reported location of the school district administrative office.

The classifications include:

11 = City, Large: Territory inside an urbanized area and inside a principal city with population of 250,000 or more.

12 = City, Midsize: Territory inside an urbanized area and inside a principal city with population less than 250,000 and greater than or equal to 100,000.

13 = City, Small: Territory inside an urbanized area and inside a principal city with population less than 100,000.

21 = Suburban, Large: Territory outside a principal city and inside an urbanized area with population of 250,000 or more.

22 = Suburban, Midsize: Territory outside a principal city and inside an urbanized area with population less than 250,000 and greater than or equal to 100,000.

23 = Suburban, Small: Territory outside a principal city and inside an urbanized area with population less than 100,000.

31 = Town, Fringe: Territory inside an urban cluster that is less than or equal to 10 miles from an urbanized area.

32 = Town, Distant: Territory inside an urban cluster that is more than 10 miles and less than or equal to 35 miles from an urbanized area.

33 = Town, Remote: Territory inside an urban cluster that is more than 35 miles from an urbanized area.

41 = Rural, Fringe: Census-defined rural territory that is less than or equal to 5 miles from an urbanized area, as well as rural territory that is less than or equal to 2.5 miles from an urban cluster.

42 = Rural, Distant: Census-defined rural territory that is more than 5 miles but less than or equal to 25 miles from an urbanized area, as well as rural territory that is more than 2.5 miles but less than or equal to 10 miles from an urban cluster.

43 = Rural, Remote: Census-defined rural territory that is more than 25 miles from an urbanized area and is also more than 10 miles from an urban cluster.

5.13 Latitude of school or agency office location (LAT)

Longitude and latitude values (often referred to as XY coordinates) are geographic coordinates that are used to identify the estimated location of a school building or an agency administrative office. Latitude is the north or south angular distance from the equator, with positive values going north and negative values going south. When combined with longitude, it reflects an estimation of where the school or office is located. Coordinate degrees, minutes, and seconds have been converted to six-digit decimal degrees.

5.14 Longitude of school or agency office location (LON)

Longitude and latitude values (often referred to as XY coordinates) are geographic coordinates that are used to identify the estimated location of a school building or an agency administrative office. Longitude is the east or west angular distance from the prime meridian, with positive values going east and negative values going west. When combined with latitude, it reflects an estimation of where the school or office is located. Coordinate degrees, minutes, and seconds have been converted to six-digit decimal degrees.

5.15 Core Based Statistical Area (CBSA) and name (NMCBSA)

A CBSA is a geographic entity associated with at least one population core of 10,000 or more, plus adjacent territory that has a high degree of social and economic integration with the core, as measured by commuting ties. CBSAs that contain a census urbanized area are designated as *metropolitan* statistical areas, while those that contain only an urban cluster are designated as *micropolitan* statistical areas. CBSAs consist of counties and equivalent entities throughout the United States and Puerto Rico. They are not delineated for other U.S. island areas. The largest city in each metropolitan or micropolitan statistical area is designated a "principal city." Additional cities qualify if specified requirements are met concerning population size and employment. The title of each metropolitan or micropolitan statistical area consists of the names of up to three of its principal cities and the name of each state into which the metropolitan or micropolitan statistical area extends. The CBSA code is a 5-digit identifier assigned to schools and agency administrative offices based on their latitude and longitude values. The CBSA classification is not an urban–rural classification; metropolitan and micropolitan statistical areas and counties outside CBSAs may contain both urban and rural territory. More discussion of urban and rural areas and their relationship to CBSAs is available in the [EDGE Locale Boundaries File Documentation](#).

5.16 Metropolitan/Micropolitan indicator (CBSATYPE)

This indicator identifies the location of a school or agency administrative office relative to a CBSA. The indicator distinguishes between schools located in metropolitan, micropolitan, and non-CBSA areas. The CBSATYPE code is a single-digit indicator assigned to a school or agency administrative office using the latitude and longitude values. The indicator is coded as '1' for Metropolitan, '2' for Micropolitan, and '0' if not included in a CBSA.

5.17   Combined Statistical Area (CSA) and name (NMCSA)

A Combined Statistical Area (CSA) consists of two or more adjacent CBSAs that share a high degree of interchange between workers who live in one area but commute to work in another area. The CSA code is a 3-digit identifier assigned to a school or agency administrative office using the latitude and longitude values. The title of a Combined Statistical Area will include the names of the two largest principal cities in the combination of the component CBSAs and the name of the third-largest principal city, if present. Additional information about CSA names and identifiers is available at the [Census Bureau's Metropolitan and Micropolitan website](#).

5.18   New England City and Town Area (NECTA) and name (NMNECTA)

NECTAs are a supplemental set of geographic entities, similar in concept to the county-based CBSAs, that the Office of Management and Budget delineates in New England based on county subdivisions— usually cities and towns. NECTAs are delineated using the same criteria as county-based CBSAs, and, similar to CBSAs, NECTAs are also categorized as metropolitan or micropolitan. The largest city or town in each NECTA is designated a "principal city." Additional cities qualify if specified requirements are met concerning population size and employment. The title of each NECTA consists of the names of up to three of its principal cities and the name of each state into which the NECTA extends. The NECTA code is a 5-digit identifier assigned to a school or agency administrative office using the latitude and longitude values. Additional information about NECTA identifiers is available at [Census Bureau's Metropolitan and Micropolitan website](#).

5.19   Congressional District code (CD)

Congressional districts are legislatively defined subdivisions of a state for the purpose of electing representatives or delegates to the House of Representatives of the United States Congress. A state or equivalent entity may comprise a single congressional district or similar representational area. The congressional district code is a four-digit numeric code used to represent the congressional districts of each multi-district state of the United States. The congressional district codes are prefixed with the two-digit state FIPS code to ensure each entity is uniquely identified. For example, the first congressional district of Alabama is identified as "0101," the second congressional district as "0102," etc. A congressional district in a state with only a single representative elected at large is designated as State FIPS + "00." For an entity with a non-voting delegate—the District of Columbia, American Samoa, Guam, the Commonwealth of the Northern Mariana Islands, Puerto Rico (whose delegate is referred to as a "resident commissioner"), and the U.S. Virgin Islands—the representational area is designated as State FIPS + "98." The CD code is assigned to a school or agency administrative office using the latitude and longitude values.

5.20   State Legislative District – Lower (SLDL) and Upper (SLDU)

State legislative districts are the areas from which members are elected to state or equivalent entity legislatures. State legislative districts embody the upper (senate—SLDU) and lower (house—SLDL) chambers of the state legislatures for all 50 states, the District of Columbia, and Puerto Rico. Nebraska has a unicameral legislature and the District of Columbia has a single council, both of which the Census Bureau treats as upper-chamber legislative areas for the purpose of data presentation. Therefore, the lower house of the state legislative districts for Nebraska and the District of Columbia are coded as N. The SLDL and SLDU codes are 5-digit identifiers assigned to a school or agency administrative office using the latitude and longitude values. The first two digits of the SLDL and SLDU identifiers are the state FIPS (STFIP) codes.

5.21   School year (SCHOOLYEAR)

This indicator identifies the academic year (e.g., 2017-2018).

5.22   Supplemental information for districts (LOC_OTHER, LOC_COUNT, LOC_SCH)

These indicators were added to the LEA file beginning with the 2016-2017 school year and provide additional information about school locale assignments for agencies. The LOC_OTHER field gives the percentage of enrolled students attending schools in a locale other than the locale assigned to the district. It does not contain a value for DoD agencies, because they have not been assigned locale values. The LOC_COUNT field is the number of distinct locales assigned to schools operated by the school district and the LOC_SCH field lists those locales. This supplemental locale information is not applicable for districts that do not operate any schools listed in the CCD schools file (where LOCALE was assigned based on office location). For those districts, the value is set to N for the character variable LOC_SCH and -2 for numeric variables LOC_OTHER and LOC_COUNT.

6.0   Data adjustments and data value exceptions

6.1   Missing and not applicable values

In cases where an expected response was missing, the cell value was set to M. In cases where field values were not applicable, the cell value was set to either -2 (for numeric variables) or N (for character variables). For example, DoDEA schools located outside the U.S. or island territories may have missing address information ('M'), but these locations are outside the scope of the official legal and statistical areas used for the geographic associations and indicators. Therefore, geographic indicators for these records are set to N.

6.2   Supplemental location information

Point locations are based on reported location address information to the greatest extent possible, however points may also be assigned based on visual review of satellite imagery and other supplemental data sources. In these cases, a reverse geocode of the estimated point location may not necessarily produce an address consistent with a reported address. In instances of missing address information, internet searches are employed in an attempt to find a viable location address for the school or agency administrative office. The missing location addresses are not populated in the final CCD files. Only the location addresses reported by the CCD program are included in the data products.

7.0   File types

NCES provides the geocode files in multiple formats to facilitate different types of uses. Excel and SAS files provide the data in traditional tabular formats, while the shapefiles provide the data in a geographic format.

7.1   Excel and SAS tables

The school and agency Excel files have an .xlsx extension (EDGE_GEOCODE_PUBLICSCH_*YYYY*.xlsx and EDGE_GEOCODE_PUBLICLEA_*YYYY*.xlsx), while the SAS data files have a .sas7bdat extension (EDGE_GEOCODE_PUBLICSCH_*YYYY*.sas7bdat and EDGE_GEOCODE_PUBLICLEA_*YYYY*.sas7bdat).

YYYY represents the abbreviated academic year represented by the public school and agency locations (e.g., 1718).

7.2   Shapefiles

A shapefile is a geographic data format composed of multiple files that combine to define the geometry and characteristics of geographic features. All geographic files developed from Census TIGER are in Global Coordinate System North American Datum of 1983 (GCS NAD83).

The name of each file is:

        EDGE_GEOCODE_PUBLICSCH_*YYYY*.<ext>
        EDGE_GEOCODE_PUBLICLEA _*YYYY*.<ext>

Where:

        SCH = School
        LEA = Local Education Agency
        <ext> = file extension:

        • .shp – The .shp file contains information about feature geometry and encapsulates information for all of the vertices needed to construct the locale polygon.
        • .dbf – The .dbf file is a table that provides attributes for each feature. The table contains a unique record for each feature identified in the .shp file.
        • .shx – The .shx file provides an index that supports the link between feature geometry and table attributes.
        • .prj – The .prj file specifies the spatial coordinate system applied to the features. It identifies how the features are referenced and centered relative to an ellipsoidal representation of the earth.
        • .shp.xml – The .shp.xml file contains metadata about the shapefile.
        • .sbn – The .sbn and .sbx files are additional index files that divide features into regions to improve processing efficiency.
        • .cpg – The .cpg file defines the character encoding used fo the .dbf file.

The geocode shapefiles include metadata that describe various characteristics about data quality, purpose, spatial extent, publication date, attribute descriptions, valid field values, contact information, and various other features. The metadata file is compatible with a text editor, web browser, and common GIS applications, and is provided in Extensible Markup Language (XML) format, the Federal Geographic Data Committee's (FGDC) content standard for digital geospatial metadata.