**NCES 2017–010 U.S. DEPARTMENT OF EDUCATION**



Education Demographic and Geographic Estimates (EDGE) Program,

Geographic Indicators and Identifiers

2014-2015



Education Demographic and Geographic Estimates (EDGE) Program,

Geographic Indicators and Identifiers

2014-2015

**SEPTEMBER 2016**

**Doug Geverdt**

National Center for Education Statistics

**Laura Nixon**

**Diane Snediker**

U.S. Census Bureau



****



**U.S. Department of Education**

**NCES 2017-010**

**U.S. DEPARTMENT OF EDUCATION**

John B. King, Jr.  
*Secretary of Education*

###### **Institute of Education Sciences**

###### Ruth Neild

###### Deputy Director for Policy and Research

###### Delegated Duties of the Director

###### **National Center for Education Statistics**

Peggy G. Carr

*Acting Commissioner*

###### **Administrative Data Division**

Ross Santy

*Associate Commissioner*

###### 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

###### September 2016

The NCES Home Page address is <http://nces.ed.gov>.

The NCES 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.

###### **Suggested Citation**

Geverdt, D., Nixon, L., Snediker, D. (2016). *Education Demographic and Geographic Estimates (EDGE) Geographic Identifiers* (NCES 2017-010). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from http://nces.ed.gov/pubsearch/.

**Content Contact**

Doug Geverdt  
(202) 245-8230  
Douglas.Geverdt@ed.gov

**Contents**

I. Purpose and Introduction 1

II. Locating Schools and Agency Administrative Offices 1

III. Geographic Indicators and Identifiers 2

Geographic Associations Based on Census TIGER 2

IV. LEA Geographic Associations 2

V. Data Content 3

Survey Collection Year (SURVYEAR) 3

NCES School ID and Local Education Agency ID (NCESSCH, LEAID) 3

State Codes (FIPST) 3

Location Street (LSTREE) 3

Location City (LCITY) 4

Location State (LSTATE) 4

Location ZIP Code (LZIP) 4

Latitude (LATCODE) 4

Longitude (LONGCODE) 4

County Code (CONUM) 4

County Name (CONAME) 5

Congressional District Code (CD) 5

Locale Code (LOCALE) 5

Core Based Statistical Area (CBSA) 6

Combined Statistical Area (CSA) 7

New England City and Town Areas (NECTA) 7

Metropolitan/Micropolitan Indicator (METMIC) 7

VI. Data Adjustments and Data Value Exceptions 7

Missing and non-applicable values 7

Addresses 8

Companion files 8

# I. Purpose and Introduction

This document describes geographic indicators and identifiers (geoIDs) for schools and local education agencies (LEAs) that supplement the 2014-2015 Common Core of Data (CCD) school and agency universe collections. The National Center for Education Statistics (NCES) Education Demographic and Geographic Estimates (EDGE) program develops this information to help data users investigate the social and spatial context of education. School points (latitude/longitude values that identify a school’s physical location) are a core component of the NCES data collection. These data are needed to answer a variety of spatially-oriented research questions, and they are a necessary input for assigning geographic locale indicators that are needed for NCES school-based surveys and for assessing program eligibility.

The CCD universe collection is a comprehensive listing of public elementary and secondary schools and the LEAs that administer and support them. The CCD includes directory, enrollment, and staffing data at state, LEA, and school levels. State education agencies (SEAs) report these data in a series of file submissions throughout the school year. The CCD school and agency geoIDs are derived from reported information about the physical location of schools and agency administrative offices. The school and agency geocode files include supplemental geographic information for the universe of schools and agencies reported in the 2014-2015 CCD school and agency directory files, and they can be integrated with the directory files through use of institutional identifiers included in both sources. NCES will publish annually updated versions of these files with each school year collection. Additional discussion and documentation for the CCD school and agency files is available at: <https://nces.ed.gov/ccd/ccddata.asp>

# II. Locating Schools and Agency Administrative Offices

The estimated location of schools and agency administrative offices is derived from the physical address reported in the CCD preliminary directory files (sch14pre\_txt, agn14pre\_txt). The reported address is compared to the location of known address points and street locations using a process known as 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 and longitude values. The NCES EDGE program uses data from two commercial sources (NavTech and HERE) to geocode the CCD directory addresses, and then reviews and synthesizes the results to retain the best estimated match. The addresses are then compared with the address reported in the prior CCD collection (if the school or LEA was present in the previous cycle) to determine whether the address changed. If not, the source and match quality of the current cycle and previous cycle are compared, and the latitude and longitude values of the best available match are assigned. 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. If the address changed, or if a school or LEA is new to the CCD collection, the latitude and longitude are based on the geocoded value. Addresses that cannot be successfully geocoded, or cases that can only be matched to a reported ZIP code or city are reviewed to identify a more proximate location. Once all cases in the directory file have been assigned a final latitude and longitude value, these estimated locations are used to develop additional geographic indicators and identifiers.

# III. Geographic Indicators and Identifiers

One of the primary purposes for identifying the point location of schools or agency administrative offices is to identify the spatial relationship between these locations and other types of geographic areas like counties, core based statistical areas (CBSA), congressional districts, and other jurisdictions or statistical areas. These associations are identified using a geographic information system (GIS), an application that can manage and evaluate geographic data to identify spatial relationships. In the same way that traditional databases rely on unique record identifiers or shared keys to join data tables together, a GIS can use shared geographic location as a means of joining data from different sources.

## Geographic Associations Based on Census TIGER

NCES develops geoIDs and indicators by associating school point locations with boundaries from the U.S. Census Bureau’s MAF/TIGER 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 U.S. NCES also uses Census TIGER boundaries to develop NCES Locale boundaries that are used to assign Locale codes to schools and 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. For example, the 2014 TIGER boundaries were used to develop the geoIDs and indicators for schools reported in the 2014-2015 CCD collection. Additional documentation about Census TIGER boundaries is available at: <http://www2.census.gov/geo/pdfs/maps-data/data/tiger/tgrshp2014/TGRSHP2014_TechDoc.pdf>.

# IV. LEA Geographic Associations

Most LEAs in the U.S. 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 like counties, congressional districts, or census tracts. 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 CCD agency geoID file assigns a single geographic association for counties, CBSAs, combined statistical areas (CSA), New England city and town areas (NECTA), and congressional districts to each agency based on the reported location of its administrative office. These associations are useful, but not necessarily complete.

The EDGE School District Geographic Relationship Files (GRF) were designed to provide a complete set of geographic associations between geographically defined school districts and other types of geographic areas including counties, CBSAs, CSAs, NECTAs, ZIP code tabulation areas (ZCTA), urban areas, congressional districts (CD), places, census tracts, and census block groups. The GRFs are based on the Census Bureau’s TIGER/Line database, and the tables 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 CCD assignments are a subset of the EDGE GRF assignments and are most useful for identifying geographic associations based on the location of the agency office, or in cases where a single association or a primary association is required. The EDGE GRF files are most useful when a full set of geographic associations is needed. More information about the EDGE GRF tables is available at <https://nces.ed.gov/programs/edge/geographicRelationshipFiles.aspx>.

# V. Data Content

## Survey Collection Year (SURVYEAR)

The survey year identifies the starting year of the annual CCD collection (e.g., 2014 indicates the 2014-2015 academic year).

## NCES School ID and Local Education Agency ID (NCESSCH, LEAID)

Each record of the CCD school file contains a unique NCES school identification number comprised of three components. The first two digits identify the state Federal Information Processing Series (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 file includes a unique seven-digit LEA identifier comprised of the state FIPS code and the LEA code.

## State Codes (FIPST)

Each record contains a state code based on the Federal Information Processing Series (FIPS). In the CCD, schools and agencies are included under the jurisdiction responsible for their administration, which is not necessarily the state where an agency or school is located. Some SEAs or charter school operators/administrators may 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). 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.

## Location Street (LSTREE)

The address for the physical location of a school or agency administrative office is based on the location street reported in the CCD Directory file. The location street address reported in the EDGE CCD school and agency geocode files is standardized to provide more uniform reporting of directionals and street type (e.g., Rd. and Road converted to RD), and some extraneous suffixes are removed (e.g., Suite numbers). If no identifiable address is available, LSTREE is set to ‘M’.

## Location City (LCITY)

The location city of a school or agency is based on the location city reported in the CCD Directory file. LCITY is not provided for some schools operated by the Department of Defense that are located in other countries.

## Location State (LSTATE)

The location state of a school or agency is based on the location state reported in the CCD Directory file, and is reported as the U.S. Postal Service two-digit state abbreviation. LSTATE is not provided for some schools operated by the Department of Defense that are located in other countries.

## Location ZIP Code (LZIP)

The location ZIP Code of a school or agency is based on the location ZIP reported in the CCD Directory file. The Directory file includes additional ZIP+4 information for some records.

## Latitude (LATCODE)

Longitude and latitude (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. This value is derived from the reported location address. Latitude is the north or south angular distance from the equator. When combined with longitude, it reflects an estimation of where the school is physically located. Coordinate degrees, minutes, and seconds have been converted to six-digit decimal degrees.

## Longitude (LONGCODE)

Longitude and latitude (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. This value is derived from the reported location address. 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 is physically located. Coordinate degrees, minutes, and seconds have been converted to six-digit decimal degrees.

## County Code (CONUM)

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, therefore the CONUM state prefix may differ from FIPST in cases where schools or agency administrative offices are located outside the reporting state. A county FIPS code is applied to all counties and other geographic entities that function as county equivalents.

## County Name (CONAME)

The county name is the legal area description reported by the U.S. Census Bureau for the county where a school or agency administrative office is located. States with county equivalents are named using their legal descriptors. This includes independent cities in Virginia, parishes in Louisiana, and census areas and boroughs in Alaska.

## 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 two-digit numeric code used to represent the congressional districts of each multi-district state of the United States. For example, the first congressional district is identified as “01,” the second congressional district as “02,” etc. A congressional district in a state with only a single representative elected at large is designated as “00.” For an entity with a nonvoting delegate—the District of Columbia, American Samoa, Guam, Puerto Rico (whose delegate is referred to as a “resident commissioner”), and the U.S. Virgin Islands—the representational area is designated as “98.” The congressional district codes are prefixed with the two-digit state FIPS code to ensure each entity is uniquely identified. The CD code is assigned to a school or agency administrative office using the latitude and longitude values.

## Locale Code (LOCALE)

The locale code is a general geographic indicator that classifies the type of area where a school or LEA administrative office 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 local codes for general description, analysis, sampling, and other statistical purposes. Locale assignments for schools are based on the estimated location of a school building, and may not necessarily characterize an entire attendance area or the residences of all enrolled students. 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 Locale Boundaries User’s Manual (<https://nces.ed.gov/programs/edge/docs/NCES_LOCALE_USERSMANUAL_2016012.pdf>). 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 of 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.

LEA Locale Assignments

Unlike school locale assignments that are based on the physical location of a school, LEA locale assignments are based on enrollment-weighted locale assignments of the schools operated by the district. If a single locale accounts for a majority of students, that locale is assigned to the agency. If the agency lacks a majority locale, the assignment is determined by first identifying whether a majority of students are attributable to a basic type (city, suburban, town, rural). If so, the agency is assigned to the subtype within that basic area that accounts for a plurality of students. If a basic type does not account for a majority of students, the agency is assigned to the single locale that accounts for a plurality of students. Although most agency 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 may be poorly characterized by a single locale assignment. Note that unlike the locale assignment, agency geocodes for county, CBSA, and other types of geographic areas are based on the reported location of the agency administrative office.

## Core Based Statistical Area (CBSA)

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 CBSA classification is not an urban–rural classification; metropolitan and micropolitan statistical areas and counties outside CBSAs may contain both urban and rural areas. More discussion of urban and rural areas and their relationship to CBSAs is available in the Locale Boundaries User’s Manual (<https://nces.ed.gov/programs/edge/docs/NCES_LOCALE_USERSMANUAL_2016012.pdf>). The CBSA code is a 5-digit identifier that is assigned to a school or agency administrative office using the latitude and longitude values.

## Combined Statistical Area (CSA)

A combined statistical area consists of two or more adjacent core based statistical areas (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. Additional information about CSA names and identifiers is available at <http://www.census.gov/population/metro/data/def.html>.

## New England City and Town Areas (NECTA)

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 categorized as metropolitan or micropolitan. 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 <http://www.census.gov/population/metro/data/def.html>.

## Metropolitan/Micropolitan Indicator (METMIC)

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 METMIC 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.

# VI. Data Adjustments and Data Value Exceptions

## 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 ‘N’. 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 geoIDs and indicators for these records are assigned as ‘N’. Missing values for latitude and longitude default to 0.

## Addresses

The physical addresses originally reported in the school or agency directory files are standardized to provide consistent formatting of directional prefixes and street types, but the content is otherwise unaltered. Missing address data are not imputed. Point locations are based on reported address information to the greatest extent possible, however points may also be assigned based on visual review of satellite imagery and other supplemental spatial data sources. In these cases, a reverse geocode of the estimated point location may not necessarily produce an address consistent with a reported address.

## Companion files

Additional information about variable names, category values, item frequencies, and record layouts is available in the Companion worksheets for the school and agency geography files.