Education Demographic and Geographic Estimates (EDGE) Program

Geocodes: Private Schools

ies NATIONAL CENTER FOR
EDUCATION STATISTICS

Institute of Education Sciences

Education Demographic and Geographic Estimates (EDGE) Program

Geocodes: Private Schools

Doug Geverdt
National Center for Education Statistics

U.S. Department of Education

ies NATIONAL CENTER FOR EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
Miguel A. Cardona
Secretary

Institute of Education Sciences
Mark Schneider
Director

National Center for Education Statistics
Peggy G. Carr
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

This publication is only available online at the NCES Home Page. To download, view, and print the report as a PDF file, go to the NCES Publications and Products page.

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

Suggested Citation
Geverdt, D., (2018). *Education Demographic and Geographic Estimates (EDGE) Program, Geocodes: Private Schools* (NCES 2017-043). 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

Acknowledgements

Contents

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 reported in the Private School Survey (PSS) collection. The PSS is a biennial collection of private elementary and secondary schools that provides data about enrollment, staffing, type of program, and other basic administrative features. The geocode files include the unique school identifier assigned by PSS, and this shared ID can be used to integrate the geocodes with the PSS program data files. Primary discussion and documentation about the PSS program and data resources is available at [NCES's PSS homepage.](#)

2.0   Locating schools
The estimated location of a school is primarily derived from the address reported in the PSS collection. 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 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) in the Common Core of Data. 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 about Census TIGER boundaries is available at the Census Bureau's TIGER webpage.

4.0   Record layout and item descriptions

4.1   Record layout

Private School Geocode File (EDGE_GEOCODE_PRIVATESCH_*YYYY*) Record Layout

Field	Length	Type	Description
PPIN	8	String	School identification number
NAME	120	String	Name of institution
STREET	100	String	Reported street address
CITY	30	String	Reported city
STATE	2	String	Reported state
ZIP	10	String	Reported 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	250	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

4.2   School identification number (PPIN)
Each record of the private school geocode file contains a unique NCES school identifier.

4.3   Name of institution (NAME)
The name of school provided by the reporting agency.

4.4   Reported street address (STREET)
The reported address for the school. If no address is available on the directory file, the address fields are set to M.

4.5   Reported city (CITY)
The reported city for the school.

4.6   Reported state (STATE)
The reported state for the school.

4.7   Reported ZIP Code (ZIP)
The reported ZIP Code for the school.

4.8   State FIPS (STFIP)
STFIP represents the two-digit FIPS code of the state where the school is located, based on the latitude and longitude values of the school. The Census Bureau's Geography Division provides a list of FIPS codes on its website.

4.9   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 using the latitude and longitude values of the school. A county code is available for all counties and other geographic entities that function as county equivalents.

4.10   County name (NMCNTY)
The county name includes the legal area description reported by the U.S. Census Bureau for the county where a school 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.

4.11   Locale code (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 assignments for schools are based on the latitude and longitude values of the school. 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. 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. 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.

4.12   Latitude of school 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 campus. 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 is located. Coordinate degrees, minutes, and seconds have been converted to six-digit decimal degrees.

4.13   Longitude of school 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 campus. 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 located. Coordinate degrees, minutes, and seconds have been converted to six-digit decimal degrees.

4.14   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 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 [Locale Boundaries User's Manual](#).

4.15   Metropolitan/Micropolitan indicator (CBSATYPE)
This indicator identifies the location of a school 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 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.

4.16   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 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 on the [Metropolitan and Micropolitan program webpage](#).

4.17   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 using the latitude and longitude values. Additional information about NECTA identifiers is available at [Metropolitan and Micropolitan program webpage](#).

4.18   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 using the latitude and longitude values.

4.19   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 3-digit identifiers assigned to a school using the latitude and longitude values. The codes are unique within each state and must be combined with 2-digit state FIPS (STFIP) to create unique codes across the nation.

4.20   School year (SCHOOLYEAR)
This indicator identifies the academic year (e.g., 2019-2020).

5.0   Data adjustments and data value exceptions

5.1   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 N.

5.2   Supplemental location information
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 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. The missing addresses are not populated in the final file. Only the addresses reported by the PSS program are included.

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

6.1   Excel and SAS tables
The school Excel file has an .xlsx extension (EDGE_GEOCODE_PRIVATESCH_*YYYY*.xlsx), while the SAS data file has an .sas7bdat extension (EDGE_GEOCODE_PRIVATESCH_*YYYY*.sas7bdat).

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

6.2   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_PRIVATESCH_*YYYY*.<ext> where <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 for the .dbf file.

YYYY represents the four-digit year for the TIGER files used to assign the geographic areas to the private school locations.

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.