

# BUNMD Supplementary Geography File Codebook

Page	Variable	Label
2	ssn	Social Security Number
3	birth_string_uncleaned	Place of birth, uncleaned string
4	birth_gnis_code	Place of birth, GNIS code
5	birth_city	Place of birth, city
6	birth_county	Place of birth, county
7	birth_region	Place of birth, census region
8	death_zip	Place of death, 5-digit ZIP code
9	death_city	Place of death, city
10	death_county_fips	Place of death, FIPS county codes
11	death_county	Place of death, county
12	death_state	Place of death, state
13	death_region	Place of death, census region
14	death_country	Place of death, country
15	death_ruc1993	Place of death, rural-urban continuum

## Summary and Methodology

The BUNMD Supplementary Geography File (N = 42,878,885) provides a set of supplementary geography variables reporting place of birth and death for individuals in the BUNMD. This file can be linked onto the BUNMD at the individual-level using Social Security number.

**Place of death variables:** To construct the place of death variables, we use the ZIP code of last residence available in the Numident Death record. It is likely that this reflects the ZIP code where an individual last lived. We map these ZIP codes onto city, county, state, census region, country, and rural urban continuum codes using a database from the United States Postal Service (USPS) Link. For ZIP codes that have been decommissioned, we use a secondary database from UnitedStatesZipCodes.org.

**Place of birth variables:** To construct the place of birth variables, we use the uncleaned 12-character city/county of birth string from the Numident Application records. These strings are uncleaned and contain misspellings and other inconsistencies. We mapped these uncleaned strings onto Geographic Names Information System (GNIS) codes using the crosswalk developed for this paper:

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. 2015. "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review*. 105(2):477–503. doi: 10.1257/aer.20120642.

To construct other place of birth variable, we mapped the GNIS codes onto city, county, state, and census regions using a database from the U.S. Board on Geographic Names Link. Please cite Black et al. (2015) if you are using any of the birthplace geography variables in the file.

# SSn

**Label:** Social Security Number

**Description:** ssn is a numeric variable reporting a person's Social Security number. This variable uniquely identifies all records in the dataset and can be used to merge onto the BUNMD dataset.

# birth\_string\_uncleaned

**Label:** Place of birth, uncleaned city/county string

**Description:** birth\_string\_uncleaned reports the 12-character city/county of birth string from the Numident Application records. This variable is uncleaned and unprocessed, and any many contain spelling errors or inconsistencies.

# birth\_gnis\_code

**Label:** Place of birth, GNIS code

**Description:** birth\_gnis\_code is a numeric variable that reports a person's Geographic Names Information System (GNIS) code for their place of birth. Each GNIS code maps onto physical locations located in the U.S., including longitude and latitude coordinates, physical feature names, counties, and states.

For more information on GNIS codes, please see <https://www.usgs.gov/faqs/what-geographic-names-information-system-gnis>.

# birth\_\_city

**Label:** Place of birth, city

**Description:** birth\_\_city is a character variable reporting a person's city of birth.

# birth\_\_county

**Label:** Place of birth, county

**Description:** birth\_\_county is a character variable reporting a person's county of birth.

# birth\_\_region

**Label:** Place of birth, census region

**Description:** birth\_\_region is a character variable reporting a person's census region of birth. For more information on Census Regions, please see <https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html>.

Births by Census Region

Region	n	freq %
midwest	8722283	20.3
northeast	7459050	17.4
south	10870153	25.4
west	2625099	6.1
NA	13202300	30.8

# death\_\_zip

**Label:** Place of death, ZIP code

**Description:** death\_\_zip is a numeric variable reporting the 5-digit ZIP code of last residence, as recorded in the Numident Death record. We note that ZIP codes are primarily used for USPS mail delivery purposes, not as geographic units. For a more detailed description of ZIP codes, please see: <https://faq.usps.com/s/article/ZIP-Code-The-Basics>.



# death\_city

**Label:** Place of death, city

**Description:** death\_city is a character variable reporting a person's city of death, as sourced from the ZIP code of their residence at time of death. For ZIP codes that cross city lines, we report the city where the primary post office for that ZIP code is located.

# death\_\_county\_\_fips

**Label:** Place of death, county FIPS code

**Description:** death\_\_fips is a numeric variable reporting a person's county FIPS code of death.

For a more detailed description of county FIPS codes, please see: <https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt#:~:text=FIPS%20codes%20are%20numbers%20which,to%20which%20the%20county%20belongs>.

# death\_\_county

**Label:** Place of death, county

**Description:** death\_\_county is a character variable reporting a person's county of death, as sourced from the ZIP code of their residence at time of death.

# death\_\_state

**Label:** Place of death, state

**Description:** death\_\_state is a character variable reporting a person's state of death, as sourced from the ZIP code of their residence at time of death.

# death\_\_region

**Label:** Place of death, census region

**Description:** death\_\_region is a character variable reporting a person's census region of death, as sourced from the ZIP code of their residence at time of death.

Deaths by Census Region

region	n	freq %
midwest	7445766	17.4
northeast	7372376	17.2
south	12843523	30.0
west	7347540	17.1
NA	7869680	18.4

# death\_country

**Label:** Place of death, country

**Description:** death\_country reports the country in which a person died, as sourced from the ZIP code of their residence at time of death. While the majority of ZIP codes are within the U.S., countries that host U.S. military bases on their territories may also have ZIP codes associated with them.

# death\_ruc1993

**Label:** Place of death, rural-urban continuum code

**Description:** death\_ruc1993 report the Rural-Urban continuum code for a person's county of death, as sourced from the United States Department of Agriculture (USDA). The Urban-Rural Continuum Codes are a classification system developed to categorize U.S. counties based on their degree of urbanization and adjacency to metropolitan areas. The system consists of ten codes, ranging from 0 to 9, where lower codes represent more urbanized counties and higher codes represent more rural areas.

For more information: <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>

## Rural-Urban Deaths

code	description	n	freq %
0	Central counties of metro areas of 1 million population or more	14394896	33.6
1	Fringe counties of metro areas of 1 million population or more	1224950	2.9
2	Counties in metro areas of 250,000 to 1 million population	7848375	18.3
3	Counties in metro areas of fewer than 250,000 population	2901397	6.8
4	Urban population of 20,000 or more, adjacent to a metro area	1529092	3.6
5	Urban population of 20,000 or more, not adjacent to a metro area	933918	2.2
6	Urban population of 2,500 to 19,999, adjacent to a metro area	2802950	6.5
7	Urban population of 2,500 to 19,999, not adjacent to a metro area	2261344	5.3
8	Rural or fewer than 2,500 urban population, adjacent to a metro area	441397	1.0
9	Rural or fewer than 2,500 urban population, not adjacent to a metro area	663520	1.5
NA	NA	7877046	18.4