

BUNMD Supplementary Geography Variables File Codebook

| Page | Variable | Label |
|------|----------------------|---------------------------------|
| 2 | ssn | Social Security Number |
| 3 | birth_gnis_code | GNIS Code of Birth Place |
| 4 | birth_city | City of Birth |
| 5 | birth_city_uncleaned | City of Birth Uncleaned String |
| 6 | birth_county | County of Birth |
| 7 | birth_region | Census Region of Birth |
| 8 | death_zip | ZIP Code of Death |
| 9 | death_state | FIPS code of Death County |
| 10 | death_state | State of Death |
| 11 | death_city | City of Death |
| 12 | death_county | County of Death |
| 13 | death_country | Country of Death |
| 14 | death_region | Census Region of Death |
| 15 | death_ruc1993 | USDA Rural-Urban Continuum Code |

Summary: The Berkeley Unified Numident Mortality Detailed Supplementary Geography Variables File (N = 42,878,885) provides information on a person’s birth and death locations. This dataset provides details on the death location based on the deceased’s ZIP code. It consists of 29,676,585 entries specifying birthplaces at city, county, and state levels, but it excludes data on individuals born in Washington D.C. or U.S. territories. To combine the Supplementary Geography Variables File with the BUNMD, use the individual’s social security number as the linking variable.

Methodology: “Place of death” variables were obtained by merging the death ZIP variable with a comprehensive dataset of all US zip codes obtained from the United States Postal Service (USPS) https://postalpro.usps.com/ZIP_Locale_Detail. For ZIP codes that have been decommissioned, we pulled from a dataset produced by UnitedStatesZipCodes.org. Once state and county of death were obtained, we were able to match this information to US Census regions and county Federal Information Processing System (FIPS) Codes.

The BUNMD dataset provides uncleaned birthplace strings, which often contain misspellings and unrecognizable values. To create more accurate and readable “place of birth” variables, we cleaned these strings and matched them against a Geographic Names Information System (GNIS) crosswalk. The crosswalk used was initially developed for the following research 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.

SSn

Label: Social Security Number

Description: ssn reports a person's Social Security number as recorded in the Numident death records. This variable uniquely identifies all records in the dataset and can be used to merge onto the BUNMD dataset.

birth_gnis_code

Label: GNIS Code of Birth Place

Description: birth_gnis_code is a numeric variable that denotes an individuals' Geographic Names Information System (GNIS) code for their place of birth. The GNIS code maps onto physical locations located in the U.S., including longitude and latitude coordinates and physical feature names.

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

birth__city

Label: City of birth

Description: birth__city reports the city of birth recorded in the BUNMD database. This variable is the result of the linkage between the BUNMD and GNIS crosswalk dataset. Approximately 94% of individuals in the BUNMD with city of birth information were matched with the crosswalk (29,676,585).

To construct the “place of birth” variables, we used a crosswalk originally 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.

birth__city__uncleaned

Label: City of birth Uncleaned String

Description: birth__city reports the city of birth recorded in the BUNMD database. This variable is uncleaned and unprocessed. As a result, it contains string values that could not be matched to actual city names (34,794,580) using the crosswalk mentioned above.

birth__county

Label: County of Birth

Description: birth__county reports the county of birth for an individual as a character string.

birth__region

Label: Census Region of Birth

Description: birth__region is a categorical variable that categorizes an individual's place of birth into one of the four Census Regions – Northeast, Midwest, South, and West – based on their state 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: Death ZIP Code

Description: death_zip is the ZIP code that is recorded at the time of death in the BUNMD. 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_fips

Label: Death County FIPS Code

Description: death_fips, a numeric variable, is the FIPS code for an individual's county of death as sourced from their ZIP code of death.

For a more detailed description of 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__state

Label: State of Death

Description: death__state reports the state the death ZIP code falls into.

For a more detailed description of ZIP codes, please see: <https://faq.usps.com/s/article/ZIP-Code-The-Basics>.

death__city

Label: City of Death

Description: death__city reports a ZIP code's primary city according to the USPS. In the case that a ZIP code crosses a city line, the city listed will be where the primary post office for that ZIP code is located.

For a more detailed description of ZIP codes, please see: <https://faq.usps.com/s/article/ZIP-Code-The-Basics>.

death__county

Label: County of Death

Description: death__county reports the county that the individual's ZIP code falls into, according to the USPS.

death_country

Label: Country of Death

Description: death_country indicates the country where the reported ZIP code is located. While the majority of these ZIP codes are within the United States, countries that host U.S. military bases on their territories may also have ZIP codes associated with them.

death__region

Label: Census Region of Death

Description: death__region is a categorical variable that categorizes an individual's place of death, based on their ZIP code and state of death, into one of the four Census Regions – Northeast, the Midwest, the South, and the West.

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_ruc1993

Label: USDA Rural-Urban Continuum Code

Description: death_ruc1993 is sourced from the United States Department of Agriculture (USDA) Urban-Rural Continuum Codes, which 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 |