

BUNMD Supplementary Geography File Codebook*

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

Summary and Methodology

The BUNMD Supplementary Geography File ($N = 42,922,321$) 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. Individuals records may contain only place of birth variables, only place of death variables, or both place of birth and place of death variables.

Place of death variables: To construct the place of death variables, we use the ZIP code of last residence available in Social Security Numident records. 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. Approximately 35 million records have a ZIP code mapped to state, county, and city information, though some fields may be missing for certain records. Place of death variables are primarily available for individuals who died in the 50 U.S. states and the District of Columbia, but some are also present for deaths in Puerto Rico and a small number of ZIP codes indicating military bases abroad.

Place of birth variables: To construct the place of birth variables, we use 12-character city/county of birth string from the Social Security Numident 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 the 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

*Last updated: 29 August, 2023

birthplace geography variables in the file. Birthplace variables are available for nearly 30 million records in this dataset. Place of birth variables are only available for locations within the 50 United States and the District of Columbia. Note that state/country of birth is already included in the BUNMD (see the variables **bpl** and **bpl_string**) and thus not present in this dataset.

SSn

Label: Social Security Number

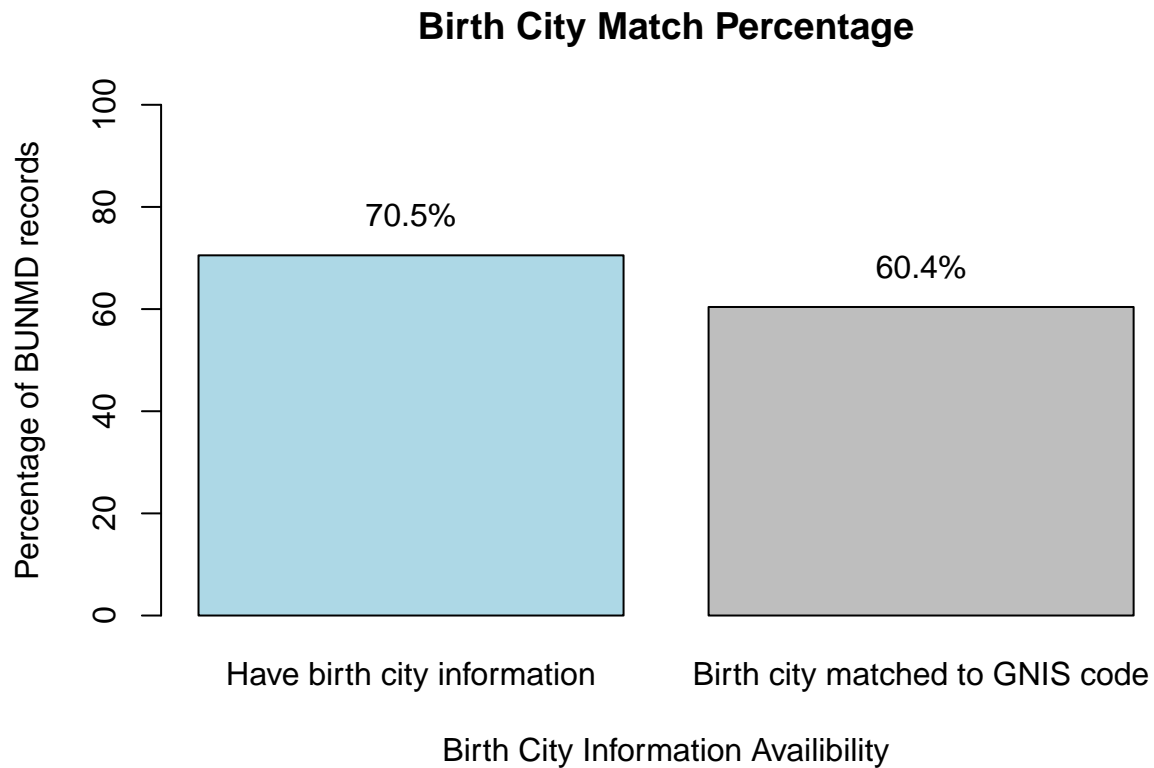
Description: ssn is a numeric variable reporting a person's Social Security number. This variable uniquely identifies all records in the dataset and can be used to merge onto the BUNMD dataset.

birth_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>.

Missing values occur if birth city is absent from Numident records, or if the original birth city string was not matched to a GNIS code.



birth_city

Label: Place of birth, uncleaned city/county string

Description: birth_city reports the 12-character city/county of birth string from the Numident Application records. This variable is generally uncleaned and unprocessed, and contains spelling errors and naming inconsistencies.

Missing values occur when birth city is not available in Numident records.

birth__county

Label: Place of birth, county

Description: birth__county is a character variable reporting a person’s county of birth, as determined by GNIS place of birth.

Missing values occur if birth city is absent from Numident records, or if the original birth city string was not matched to a GNIS code.

Note: In 2022, the state of Connecticut began replacing its eight counties with nine county-equivalent planning regions. This dataset contains the eight “legacy” county regions of Connecticut.

birth__fips

Label: Place of birth, county FIPS code

Description: birth__fips is a character variable reporting a person’s county FIPS code of birth. FIPS codes are 5 digits, of which the first two are the FIPS code of the state to which the county belongs. For a list of modern FIPS codes, see: https://en.wikipedia.org/wiki/List_of_United_States_FIPS_codes_by_county.

Missing values occur if birth city is absent from Numident records, or if the original birth city string was not matched to a GNIS code or FIPS code.

Notes: FIPS codes appear numeric but can contain leading zeros. County-level FIPS codes are always 5 digits. In 2022, the state of Connecticut began replacing its eight counties with nine county-equivalent planning regions. This dataset contain FIPS codes for the the eight “legacy” county regions of Connecticut. FIPS codes occasionally change due to county boundary changes, name changes, etc., and thus codes contained in this dataset may not be completely consistent with external county-level data.

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>.

NA values are due to an inability to match uncleaned birth city values with the GNIS crosswalk, birth city being outside of a census region (in a different country for example), or a birth city not being entered in social security records.

birth_region	label	n	freq %
	NA	11324131	26.4
midwest	Midwest	9150959	21.3
northeast	Northeast	7747504	18.1
south	South	11938948	27.8
west	West	2760861	6.4

Births by Census Region

death_zip

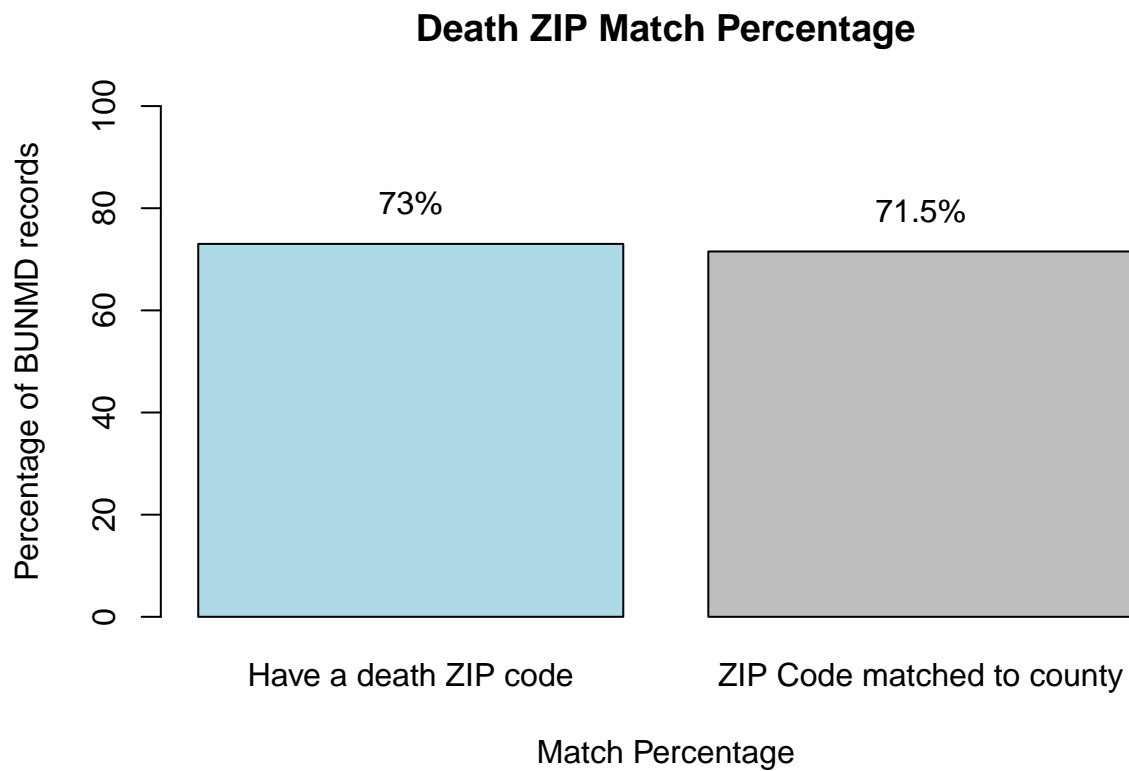
Label: Place of death, ZIP code

Description: death_zip is a character variable that represents the 5-digit ZIP code of the last residence, as obtained from Numident Death records

Missing values may occur if ZIP code was entered incorrectly in Numident records, if ZIP code is absent from Numident records, or if place of death occurred outside USPS identifiable ZIP codes (e.g., most locations abroad). However, Non-identifiable codes such as “XX768” and “00000” may remain in this field.

Notes: ZIP codes appear numeric, but can contain leading zeros. They should always be 5 digits.

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 determined 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.

Missing values may occur if ZIP code was entered incorrectly in Numident records, if ZIP code is absent from Numident records, or if place of death could not be matched to an USPS identifiable ZIP code.

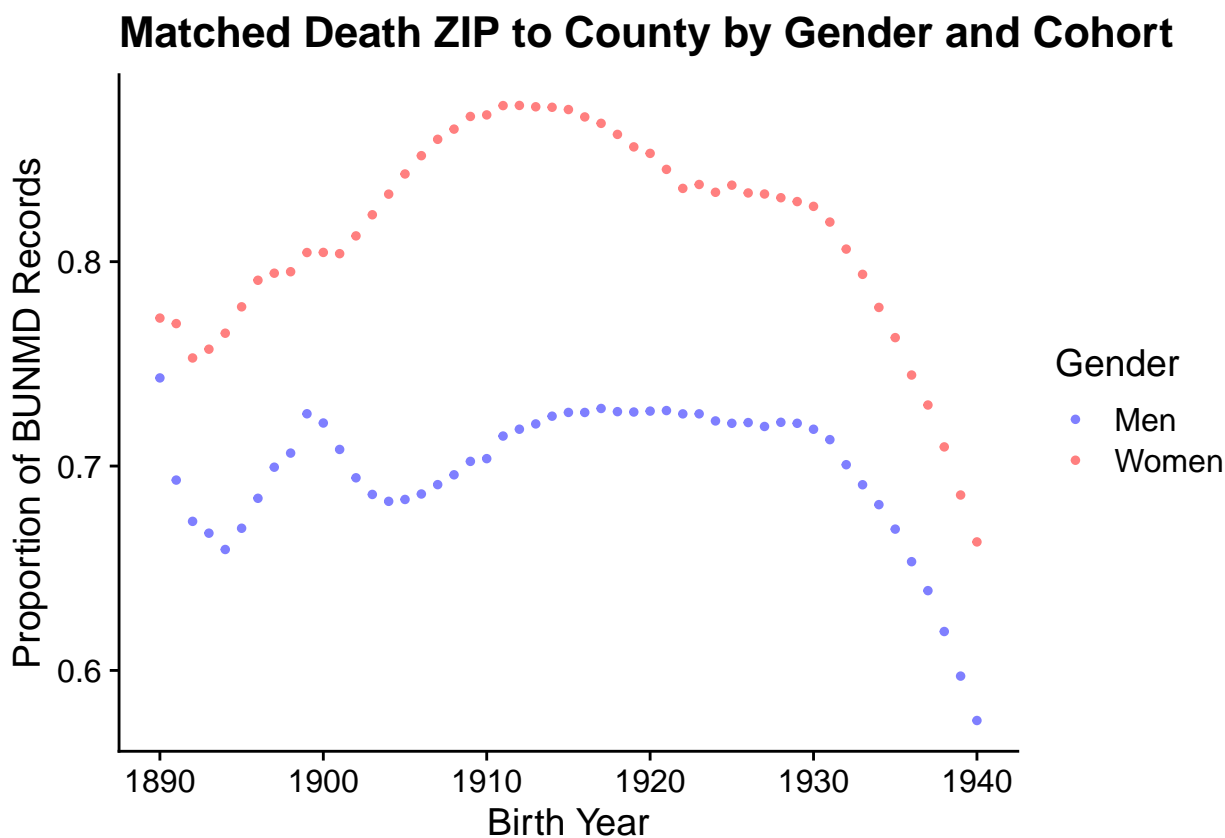
death_county

Label: Place of death, county

Description: death_county is a character variable reporting a person's county of death, as determined by the ZIP code of their residence at time of death.

NA values are due to ZIP code not being entered correctly in social security records or a ZIP code that could not be matched against USPS records.

Note: In 2022, the state of Connecticut began replacing its eight counties with nine county-equivalent planning regions. This dataset contains the eight “legacy” county regions for Connecticut.



death_fips

Label: Place of death, county FIPS code

Description: death_fips is a character variable reporting a person's county FIPS code of birth. FIPS codes are 5 digits, of which the first two are the FIPS code of the state to which the county belongs. For a list of modern FIPS codes, see: https://en.wikipedia.org/wiki/List_of_United_States_FIPS_codes_by_county.

Missing values may occur if ZIP code was entered incorrectly in Numident records, if ZIP code is absent from Numident records, if place of death could not be matched to an identifiable USPS identifiable ZIP code, or if county could not be matched to a FIPS code. County FIPS codes are not available for deaths occurring in Puerto Rico.

Notes: FIPS codes appear numeric but can contain leading zeros. County-level FIPS codes are always 5 digits. In 2022, the state of Connecticut began replacing its eight counties with nine county-equivalent planning regions. This dataset contain FIPS codes for the the eight “legacy” county regions of Connecticut. FIPS codes occasionally change due to county boundary changes, name changes, etc., and thus codes contained in this dataset may not be completely consistent with external county-level data.

death_state

Label: Place of death, state

Description: death_state is a character variable reporting a person's state of death, as determined by the ZIP code of their residence at time of death. Includes the 50 U.S states, the District of Columbia, and Puerto Rico.

Missing values may occur if ZIP code was entered incorrectly in Numident records, if ZIP code is absent from Numident records, or if place of death could not be matched to an identifiable USPS identifiable ZIP code in the United States.

death_state	label	n	freq %	death_state	label	n	freq %
	NA	7630590	17.8	ms	Mississippi	431890	1.0
ak	Alaska	33259	0.1	mt	Montana	99675	0.2
al	Alabama	668450	1.6	nc	North Carolina	1062206	2.5
ar	Arkansas	434480	1.0	nd	North Dakota	85747	0.2
az	Arizona	571718	1.3	ne	Nebraska	224847	0.5
ca	California	3248159	7.6	nh	New Hampshire	128350	0.3
co	Colorado	398331	0.9	nj	New Jersey	1112089	2.6
ct	Connecticut	466399	1.1	nm	New Mexico	175151	0.4
dc	Washington DC	81134	0.2	nv	Nevada	175917	0.4
de	Delaware	96280	0.2	ny	New York	2454973	5.7
fl	Florida	2358276	5.5	oh	Ohio	1588469	3.7
ga	Georgia	858496	2.0	ok	Oklahoma	550056	1.3
hi	Hawaii	109597	0.3	or	Oregon	438795	1.0
ia	Iowa	430480	1.0	pa	Pennsylvania	1910092	4.5
id	Idaho	136070	0.3	pr	Puerto Rico	283217	0.7
il	Illinois	1592202	3.7	ri	Rhode Island	168624	0.4
in	Indiana	779401	1.8	sc	South Carolina	457600	1.1
ks	Kansas	378621	0.9	sd	South Dakota	95703	0.2
ky	Kentucky	605874	1.4	tn	Tennessee	831360	1.9
la	Louisiana	607693	1.4	tx	Texas	2011891	4.7
ma	Massachusetts	864433	2.0	ut	Utah	149011	0.3
md	Maryland	613376	1.4	va	Virginia	848772	2.0
me	Maine	188574	0.4	vt	Vermont	78840	0.2
mi	Michigan	1301649	3.0	wa	Washington	659701	1.5
mn	Minnesota	480117	1.1	wi	Wisconsin	713379	1.7
mo	Missouri	871309	2.0	wv	West Virginia	325070	0.8
				wy	Wyoming	56010	0.1

Deaths by State

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. This variable is available for persons dying within the 50 United States and the District of Columbia.

Missing values may occur if ZIP code was entered incorrectly in Numident records, if ZIP code is absent from Numident records, if place of death could not be matched to an identifiable USPS identifiable ZIP code in the United States.

death_region	label	n	freq %
	NA	7913076	18.4
midwest	Midwest	7445768	17.3
northeast	Northeast	7372380	17.2
south	South	12843615	29.9
west	West	7347564	17.1

Deaths by Census Region

death_country

Label: Place of death, country

Description: death_country is a character variable reporting a person's country of death, as determined by the ZIP code of their residence at time of death. While the vast 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. Codes follow [ISO 3166-1 alpha-2 country codes](#).

Missing values may occur if ZIP code being entered incorrectly in Numident records, if ZIP code is absent from Numident records, if place of death could not be matched to an identifiable USPS identifiable ZIP code.

death_ruc1993

Label: County of death, rural-urban continuum code

Description: death_ruc1993 report the 1993 Rural-Urban continuum code for a person's county (or county equivalent) 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.

Missing values may occur if ZIP code being entered incorrectly in Numident records, if ZIP code is absent from Numident records, or if ZIP code could not be matched to a county.

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

death_ruc1993		label	n	freq %
0	Central counties of metro areas of 1 million population or more		14394985	33.5
1	Fringe counties of metro areas of 1 million population or more		1224974	2.9
2	Counties in metro areas of 250,000 to 1 million population		7848382	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		2802951	6.5
7	Urban population of 2,500 to 19,999, not adjacent to a metro area		2261345	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	7920442	18.5

Rural-Urban County of Death