

CenSoc-Numident Supplementary Geography File Codebook*

Page	Variable	Label
3	HISTID	Historical unique identifier
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_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 code
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 Code

Summary and Methodology

The CenSoc-Numident Supplementary Geography File (N = 6,971,468) provides a set of supplementary geographic variables reporting place of birth and/or death for individuals in the CenSoc-Numident mortality file. This file can be linked onto the CenSoc-Numident at the individual-level using HISTID. Individual 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 ZIP code of residence as reported in Social Security Numident death records. We map these ZIP codes onto city, county, state, census region, and country information 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 6.3 million records in the CenSoc-Numident dataset have a ZIP code that can be 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 ZIP codes in Puerto Rico and military bases abroad.

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

*Last updated: 05 December, 2023

To construct other place of birth variable, we mapped the GNIS codes onto city, county, state, and census region information 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. Birthplace variables are available for nearly 6.5 million records in this dataset. These 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 CenSoc-Numident (see the variables **bpl** and **bpl_string**) and thus not present in this dataset.

HISTID

Label: Historical unique identifier

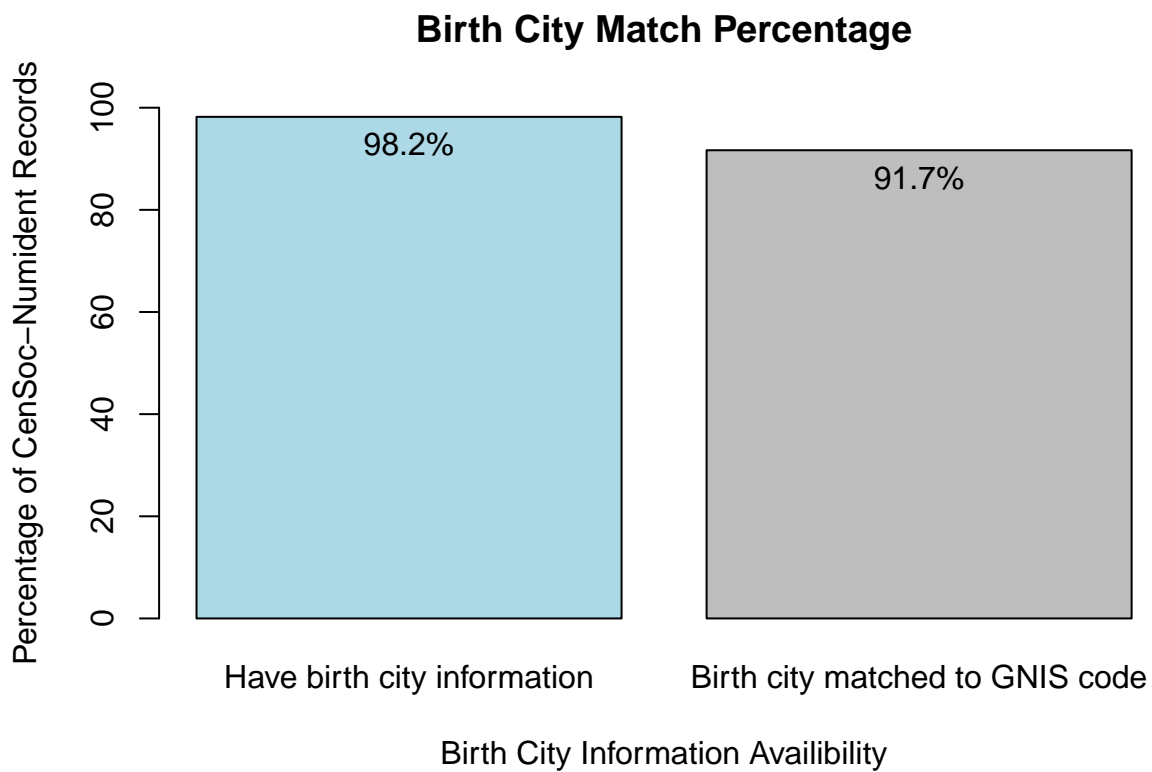
Description: HISTID is a unique individual-level identifier from IPUMS-USA 1940 Census data. This variable uniquely identifies all records in the dataset and can be used to merge onto the CenSoc-Numident mortality dataset.

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, city

Description: birth_city is a character variable that reports 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 the name of 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 county.

Notes: FIPS codes appear numeric but can contain leading zeros. County-level FIPS codes in this dataset 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>.

Missing values occur if birth state/country is absent from Numident records or outside of the United States.

birth_region	label	n	freq %
	NA	142323	2.0
midwest	Midwest	2388590	34.3
northeast	Northeast	1763495	25.3
south	South	2083543	29.9
west	West	593517	8.5

Births by Census Region

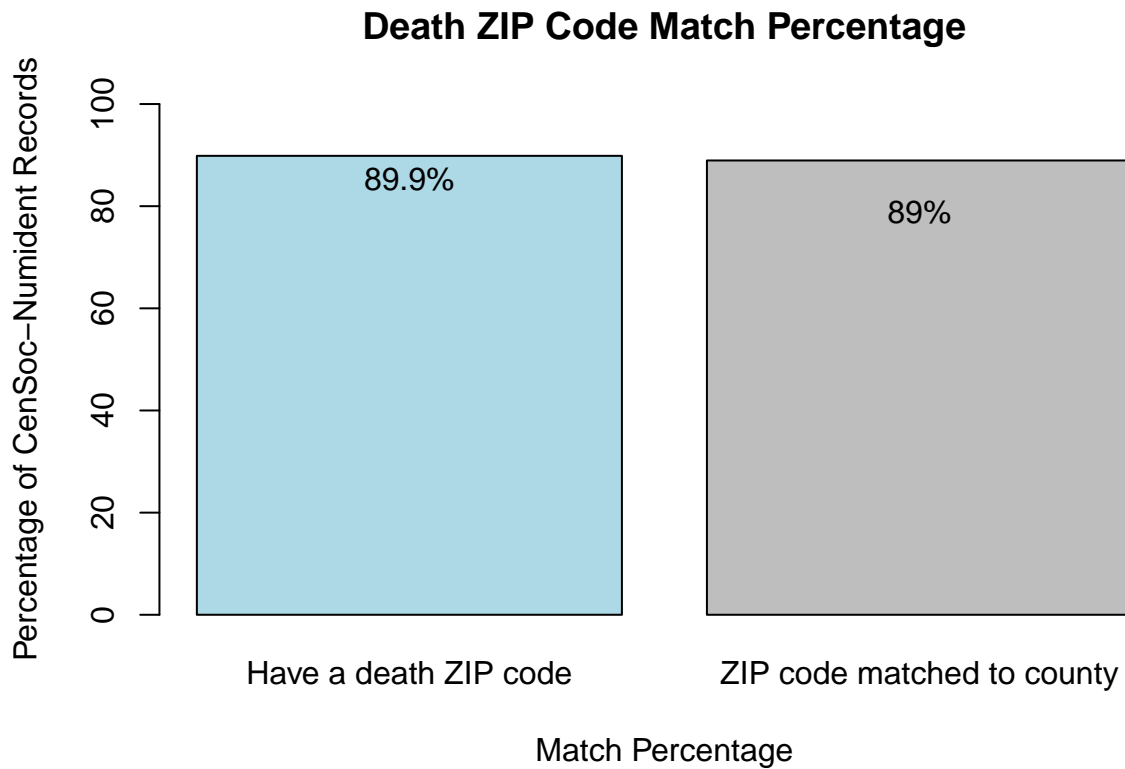
death_zip

Label: Place of death, ZIP code

Description: death_zip is a character variable that reports 5-digit ZIP code of 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 residence at time of death was outside USPS identifiable ZIP codes (e.g., most locations abroad).

Notes: ZIP codes appear numeric, but can contain leading zeros. They should always be 5 digits. Non-identifiable codes such as “XX768” and “00000” may remain in this field. 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 identifiable USPS ZIP code.

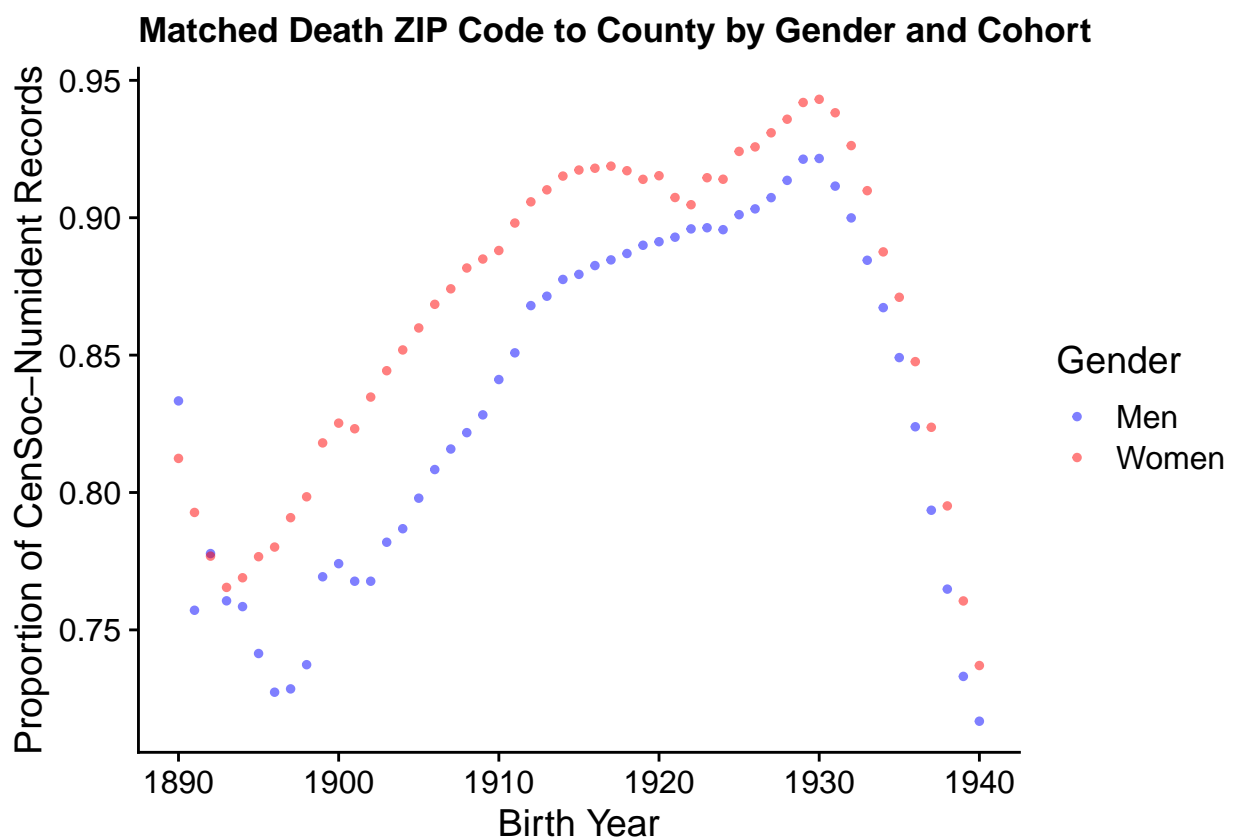
death_county

Label: Place of death, county

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

Missing values may occur if ZIP code was entered incorrectly in Numident records, if ZIP code is absent from Numident records, or if ZIP codes 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__county__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 in this dataset are always 5 digits. In 2022, the state of Connecticut began replacing its eight counties with nine county-equivalent planning regions. This dataset contains FIPS codes for 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. Values in this field 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.

death__state	label	n	freq %	death__state	label	n	freq %
	NA	705247	10.1	ms	Mississippi	59350	0.9
ak	Alaska	4833	0.1	mt	Montana	23905	0.3
al	Alabama	96717	1.4	nc	North Carolina	158663	2.3
ar	Arkansas	68530	1.0	nd	North Dakota	20996	0.3
az	Arizona	108487	1.6	ne	Nebraska	57204	0.8
ca	California	541103	7.8	nh	New Hampshire	31644	0.5
co	Colorado	81786	1.2	nj	New Jersey	197142	2.8
ct	Connecticut	93149	1.3	nm	New Mexico	30969	0.4
dc	Washington DC	10451	0.1	nv	Nevada	37500	0.5
de	Delaware	19606	0.3	ny	New York	377049	5.4
fl	Florida	377124	5.4	oh	Ohio	327561	4.7
ga	Georgia	118774	1.7	ok	Oklahoma	93291	1.3
hi	Hawaii	4667	0.1	or	Oregon	95311	1.4
ia	Iowa	96989	1.4	pa	Pennsylvania	343227	4.9
id	Idaho	29976	0.4	pr	Puerto Rico	1216	0.0
il	Illinois	291387	4.2	ri	Rhode Island	36341	0.5
in	Indiana	161364	2.3	sc	South Carolina	73069	1.0
ks	Kansas	79917	1.1	sd	South Dakota	26570	0.4
ky	Kentucky	100157	1.4	tn	Tennessee	126017	1.8
la	Louisiana	102850	1.5	tx	Texas	335119	4.8
ma	Massachusetts	158004	2.3	ut	Utah	35333	0.5
md	Maryland	114550	1.6	va	Virginia	141986	2.0
me	Maine	45951	0.7	vt	Vermont	18809	0.3
mi	Michigan	248418	3.6	wa	Washington	133667	1.9
mn	Minnesota	119164	1.7	wi	Wisconsin	165813	2.4
mo	Missouri	169809	2.4	wv	West Virginia	61322	0.9
				wy	Wyoming	13384	0.2

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, or if place of death could not be matched to an identifiable USPS ZIP code in the United States.

birth_region	label	n	freq %
	NA	706351	10.1
midwest	Midwest	1538179	22.1
northeast	Northeast	1301316	18.7
south	South	2057688	29.5
west	West	1367934	19.6

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, or 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 is a numeric variable that reports 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 Rural-Urban 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, see: <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>.

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

death_ruc1993		label	n	freq %
0		Central counties of metro areas of 1 million population or more	2448589	35.1
1		Fringe counties of metro areas of 1 million population or more	239080	3.4
2		Counties in metro areas of 250,000 to 1 million population	1410715	20.2
3		Counties in metro areas of fewer than 250,000 population	556801	8.0
4		Urban population of 20,000 or more, adjacent to a metro area	291959	4.2
5		Urban population of 20,000 or more, not adjacent to a metro area	172429	2.5
6		Urban population of 2,500 to 19,999, adjacent to a metro area	511661	7.3
7		Urban population of 2,500 to 19,999, not adjacent to a metro area	422880	6.1
8		Rural or fewer than 2,500 urban population, adjacent to a metro area	80740	1.2
9		Rural or fewer than 2,500 urban population, not adjacent to a metro area	128967	1.8
NA		NA	707647	10.2

Rural-Urban County of Death