

## CenSoc-DMF Demo Codebook (V3.0)\*

Page	Variable	Label
2	HISTID	Historical Unique Identifier
3	byear	Year of Birth
4	bmonth	Month of Birth
5	dyear	Year of Death
6	dmonth	Month of Death
7	death_age	Age at Death (Years)
8	weight	CenSoc Sample Weight
9	Additional IPUMS variables	Additional 1940 Census variables, including: pernum, perwt, age, bpld, mbpl, fbpl, educd, educ_yrs, empstatd, hispan, incwage, incnonwg, marst, nativity, occ, occscore, ownershp, race, rent, serial, statefip, and urban

**Summary:** The CenSoc-DMF Version 3.0 Demo dataset ( $N = 45,907$ ) links the IPUMS 1940 1% census sample to the Death Master File (DMF) dataset, a collection of death records reported to the Social Security Administration. Records were linked using a conservative variant of the ABE method developed by Abramitzky, Boustan, and Eriksson (2012, 2014, 2017).

Note that this demo dataset is not conducive to high-resolution mortality research. We recommend using this file for exploratory and demonstrative purposes. To best conduct research with CenSoc data, researchers may download the full CenSoc-DMF from the CenSoc website, obtain an extract of the 1940 Census from IPUMS-USA, and merge data using on the individual-level, unique identifier HISTID variable. Please adhere to IPUMS citation guidelines when using IPUMS census data.

---

\*Last updated: 17 January, 2024

# HISTID

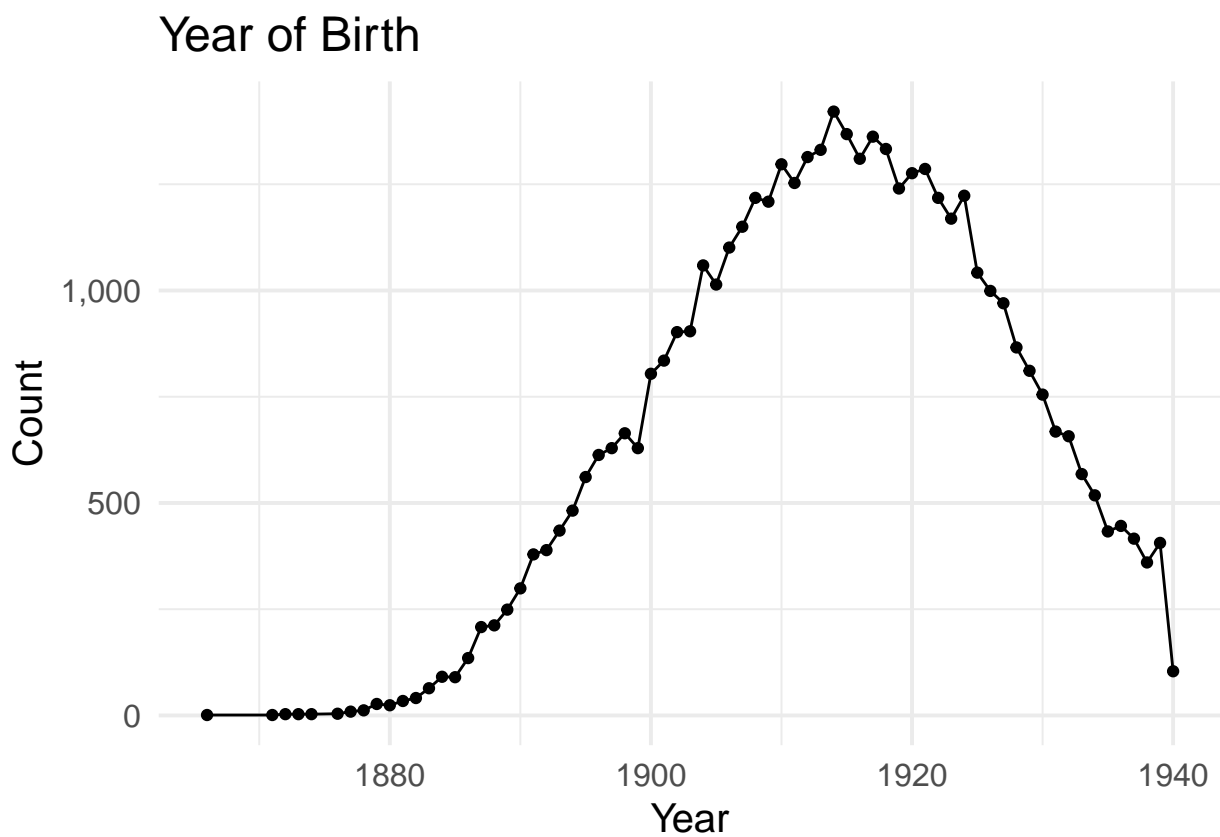
**Label:** Historical Unique Identifier

**Description:** HISTID is a unique individual-level identifier. It can be used to merge the CenSoc-DMF file with the 1940 Full-Count Census from IPUMS.

# byear

**Label:** Birth Year

**Description:** byear reports a person's year of birth, as recorded in the Social Security Death Master File.



# bmonth

**Label:** Birth Month

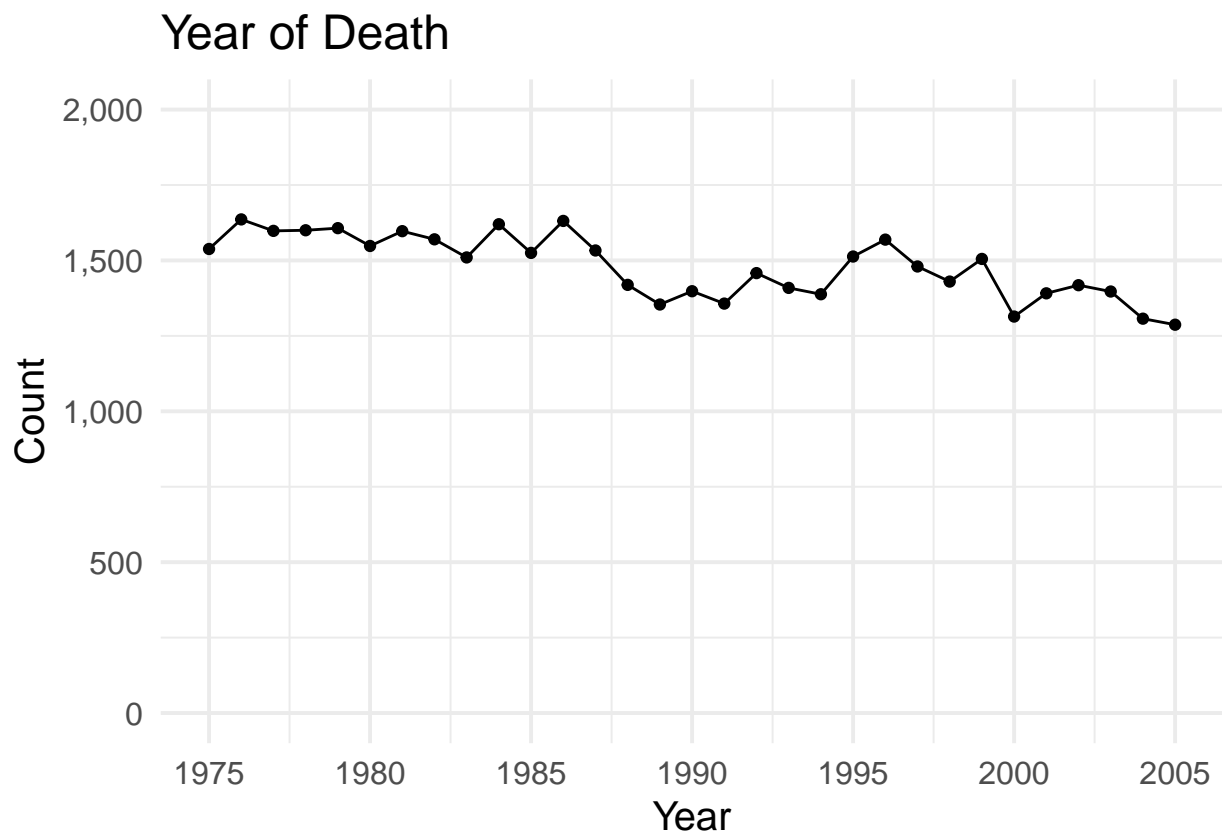
**Description:** bmonth reports a person's month of birth, as recorded in the Social Security Death Master File.

bmonth	label	n	freq %
1	January	3979	8.7
2	February	3700	8.1
3	March	4138	9.0
4	April	3699	8.1
5	May	3692	8.0
6	June	3556	7.7
7	July	3798	8.3
8	August	4104	8.9
9	September	3861	8.4
10	October	3936	8.6
11	November	3736	8.1
12	December	3708	8.1

# dyear

**Label:** Death Year

**Description:** dyear reports a person's year of death, as recorded in the Social Security Death Master File.



# dmonth

**Label:** Death Month

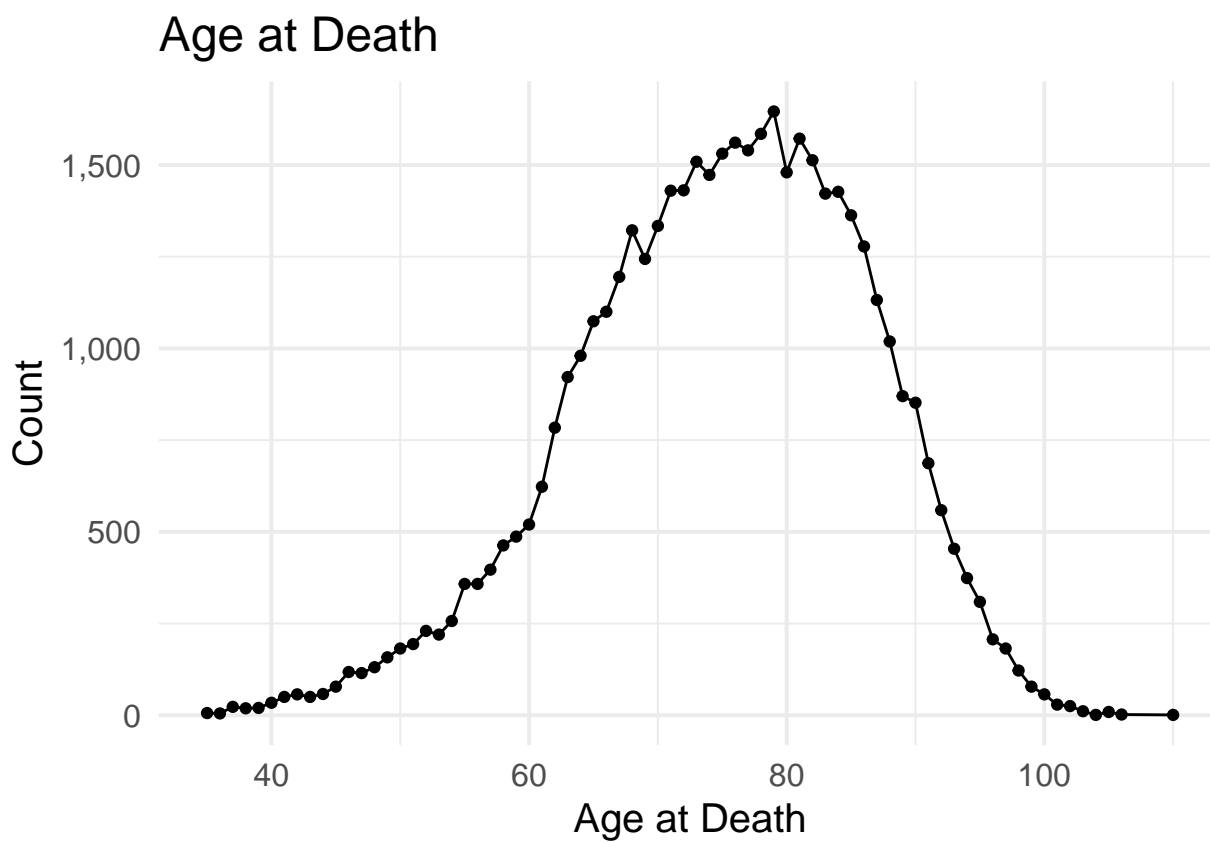
**Description:** dmonth reports a person's month of death, as recorded in the Social Security Death Master File.

dmonth	label	n	freq %
1	January	4398	9.6
2	February	3837	8.4
3	March	4142	9.0
4	April	3891	8.5
5	May	3786	8.2
6	June	3499	7.6
7	July	3628	7.9
8	August	3662	8.0
9	September	3490	7.6
10	October	3794	8.3
11	November	3724	8.1
12	December	4056	8.8

# death\_age

**Label:** Age at Death (Years)

**Description:** death\_age reports a person's age at death in years, calculated using the birth and death information recorded in the Social Security Death Master File.

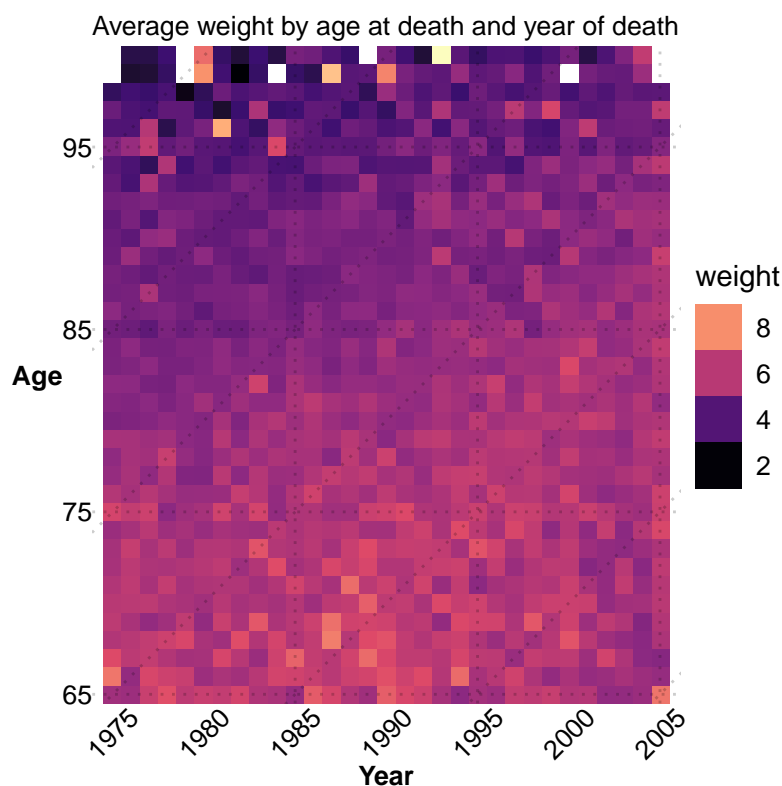


# weight

**Label:** CenSoc Sample Weight<sup>1</sup>

**Description:** A post-stratification person-weight to National Center for Health Statistics (NCHS) totals for persons (1) dying between 1975-2005 (2) dying between ages 65-100. Weights are based on age at death, year of death, sex, and race, and place of birth. Please see the [technical documentation on weights](#) for more information.

Value	Label
1.00	Min Weight
27.89	Max Weight
NA	No Weight Assigned



<sup>1</sup>The IPUMS-USA 1940 1% sample also includes a weight (**perweight**) to account for the 1940 sampling procedure (thus no weights for the 100% complete count 1940 census). For analysis, we recommend using both sets of weights. A final weight can be constructed by multiplying the two weights together.



# IPUMS 1940 Census Variables

The variables below are from the IPUMS-USA 1940 1% census sample. We recommend looking at the terrific documentation on the IPUMS-USA website: <https://usa.ipums.org/usa/index.shtml>

Variable	Label
pernum	Person number in household
perwt	IPUMS person weight <sup>2</sup>
age	Age on April 1st, 1940
sex	Sex
bp1d	Place of birth (detailed)
mbpl	Mother's place of birth <sup>3</sup>
fbpl	Father's place of birth <sup>4</sup>
educd	Educational attainment (detailed IPUMS codes)
educ_yrs	Educational attainment in years (constructed) <sup>5</sup>
empstatd	Employment status (detailed)
hispan	Hispanic/Spanish/Latino origin (imputed) <sup>6</sup>
incwage	Wage and salary income in 1939
incnonwg	Had non-wage/salary income over \$50 in 1939
marst	Marital status
nativity	Foreign birthplace or parentage
occ	Occupation
occscore	Occupational income score
ownership	Ownership of dwelling (tenure)
race	Race
rent	Monthly contract rent
serial	Household serial number
statefip	State of residence 1940 (FIPS code)
urban	Urban/rural status

<sup>2</sup>The IPUMS **perweight** accounts for the 1940 sampling procedure to construct the 1% sample, and thus is only available in the 1940 1% sample. For analysis, we recommend using both the IPUMS **perweight** and the CenSoc **weight**. A final weight can be constructed by multiplying the two weights together

<sup>3</sup>This variable is only available for sample-line persons (a one-in-twenty sample asked additional questions in the 1940 Census) or those living with their mother.

<sup>4</sup>This variable is only available for sample-line persons (a one-in-twenty sample asked additional questions in the 1940 Census) or those living with their father.

<sup>5</sup>**educ\_yrs** is constructed from the IPUMS **educd** variable but not directly available from IPUMS.

<sup>6</sup>The 1940 Census did not directly inquire about Hispanic ethnicity or origin. This variable is determined by IPUMS using information such as one's birthplace or a parent's birthplace.