

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, sex, 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 ($N = 42,421$) 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).

We 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 full-count 1940 Census from IPUMS-USA, and merge data using on the individual-level, unique identifier HISTID variable. Please adhere to CenSoc and IPUMS citation guidelines when using this file.

*Last updated: 25 January, 2024

HISTID

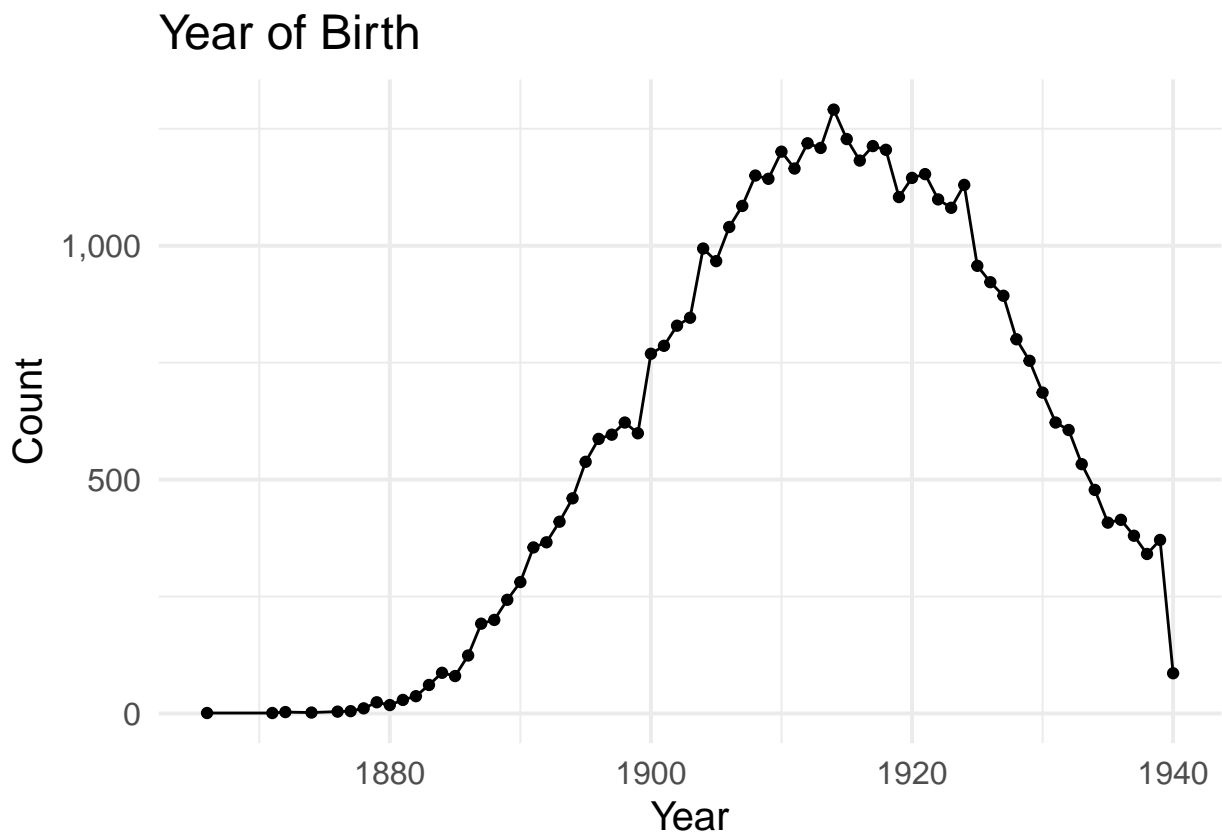
Label: Historical Unique Identifier

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

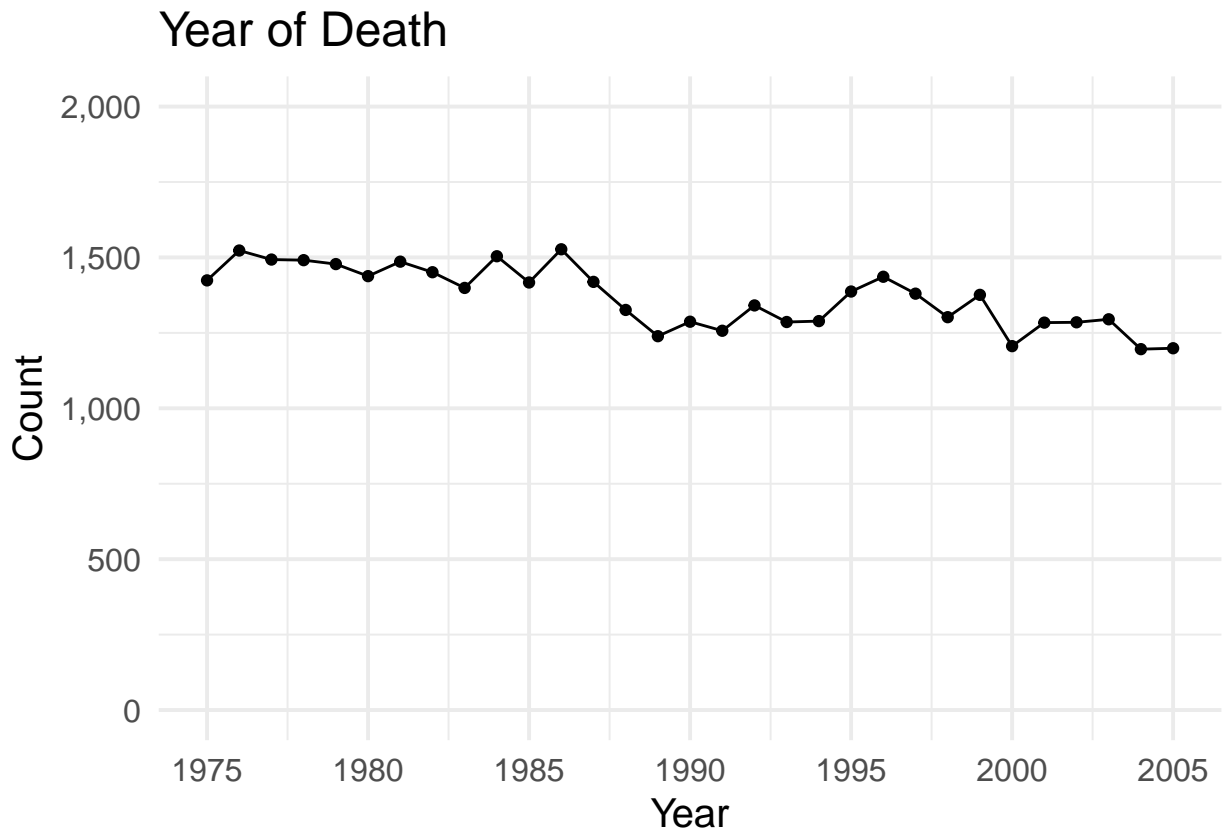
Description: bmonth reports a person's month of birth, as recorded in the Social Security Death Master File.

bmonth	label	n	freq %
1	January	3676	8.7
2	February	3420	8.1
3	March	3822	9.0
4	April	3410	8.0
5	May	3402	8.0
6	June	3273	7.7
7	July	3532	8.3
8	August	3797	9.0
9	September	3552	8.4
10	October	3629	8.6
11	November	3466	8.2
12	December	3442	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

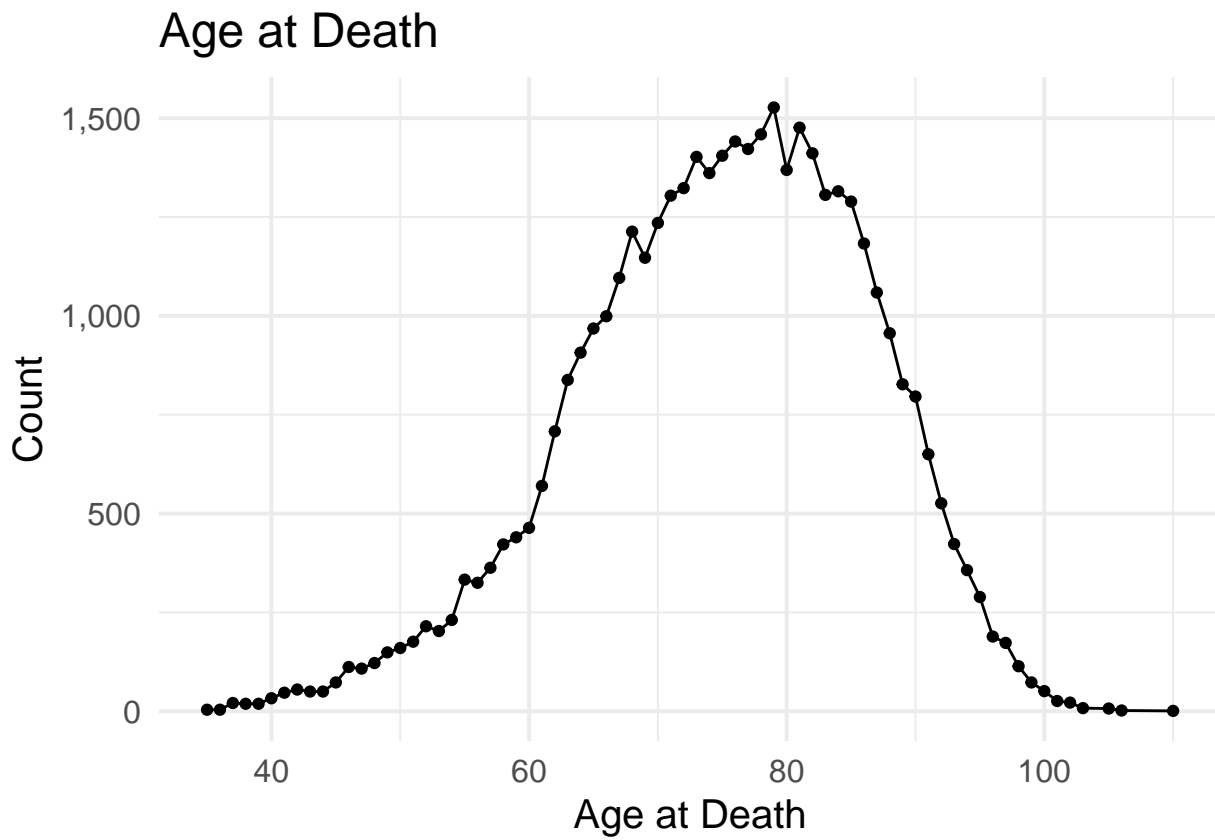
Description: dmonth reports a person's month of death, as recorded in the Social Security Death Master File.

dmonth	label	n	freq %
1	January	4052	9.6
2	February	3550	8.4
3	March	3814	9.0
4	April	3595	8.5
5	May	3520	8.3
6	June	3238	7.6
7	July	3365	7.9
8	August	3387	8.0
9	September	3206	7.6
10	October	3508	8.3
11	November	3440	8.1
12	December	3746	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.

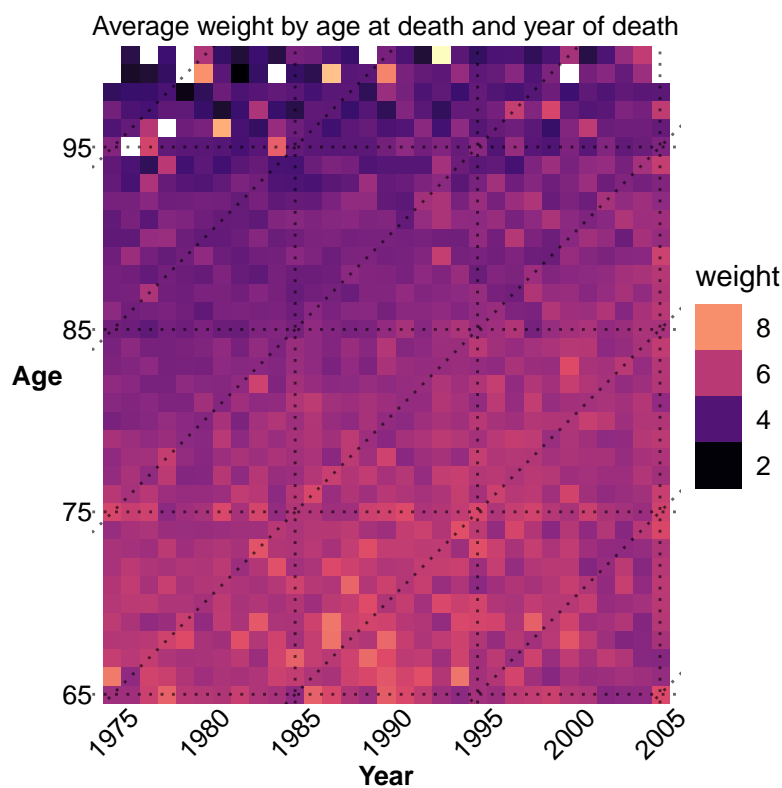


weight

Label: CenSoc Sample Weight¹

Description: weight is 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



¹The 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 ²
age	Age on April 1st, 1940
sex	Sex
bpld	Place of birth (detailed)
mbpl	Mother's place of birth ³
fbpl	Father's place of birth ⁴
educd	Educational attainment (detailed IPUMS codes)
educ_yrs	Educational attainment in years (constructed) ⁵
empstatd	Employment status (detailed)
hispan	Hispanic/Spanish/Latino origin (imputed) ⁶
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	Montly contract rent
serial	Household serial number
statefip	State of residence 1940 (FIPS code)
urban	Urban/rural status

²The 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

³This 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.

⁴This 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.

⁵`educ_yrs` is constructed from the IPUMS `educd` variable but not directly available from IPUMS.

⁶The 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.