

CenSoc-Numident Demo Codebook (V3)*

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	sex	Sex
9	race_first	Race on First Application
10	race_first_cyear	First Race: Application Year
11	race_first_cmonth	First Race: Application Month
12	race_last	Race on Last Application
13	race_last_cyear	Last Race: Application Year
14	race_last_cmonth	Last Race: Application Month
15	bpl	Place of Birth
16	zip_residence	ZIP Code of Residence at Time of Death
17	socstate	State where Social Security Number Issued
18	age_first_application	Age at First Social Security Application
19	weight	CenSoc Sample Weight
20	Additional IPUMS variables	Additional 1940 Census variables, including: pernum, perwt, age, mbpl, fbpl, educd, educ_yrs, empstatd, hispan, incwage, incnonwg, marst, nativity, occ, occscore, ownershp, race, rent, serial, statefip, and urban.

Summary: The CenSoc-Numident Version 3 Demo dataset (N = 64,686) links the IPUMS 1940 Census 1% sample to the National Archives' public release of the Social Security Numident file. 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-Numident 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

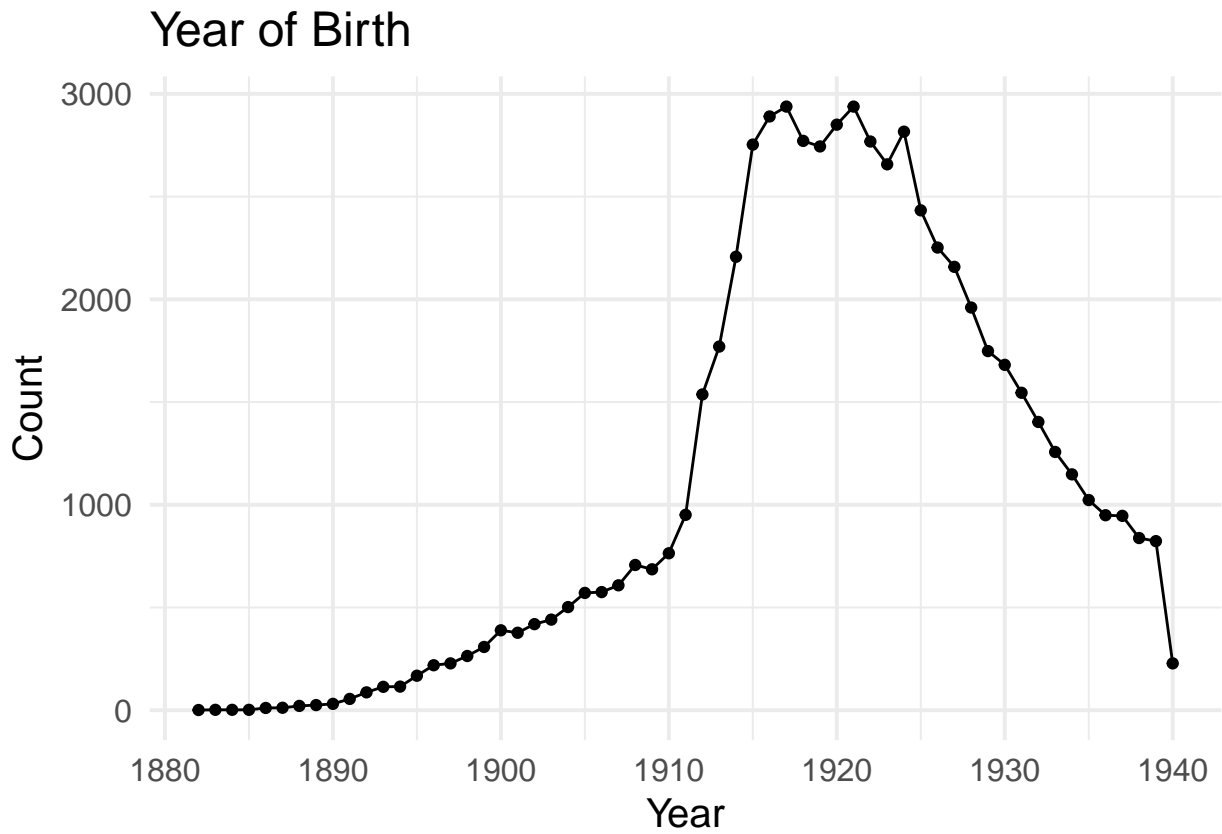
Label: Historical Unique Identifier

Description: HISTID is a unique individual-level identifier. It can be used to merge the CenSoc-Numident 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 Numident death records.



bmonth

Label: Birth Month

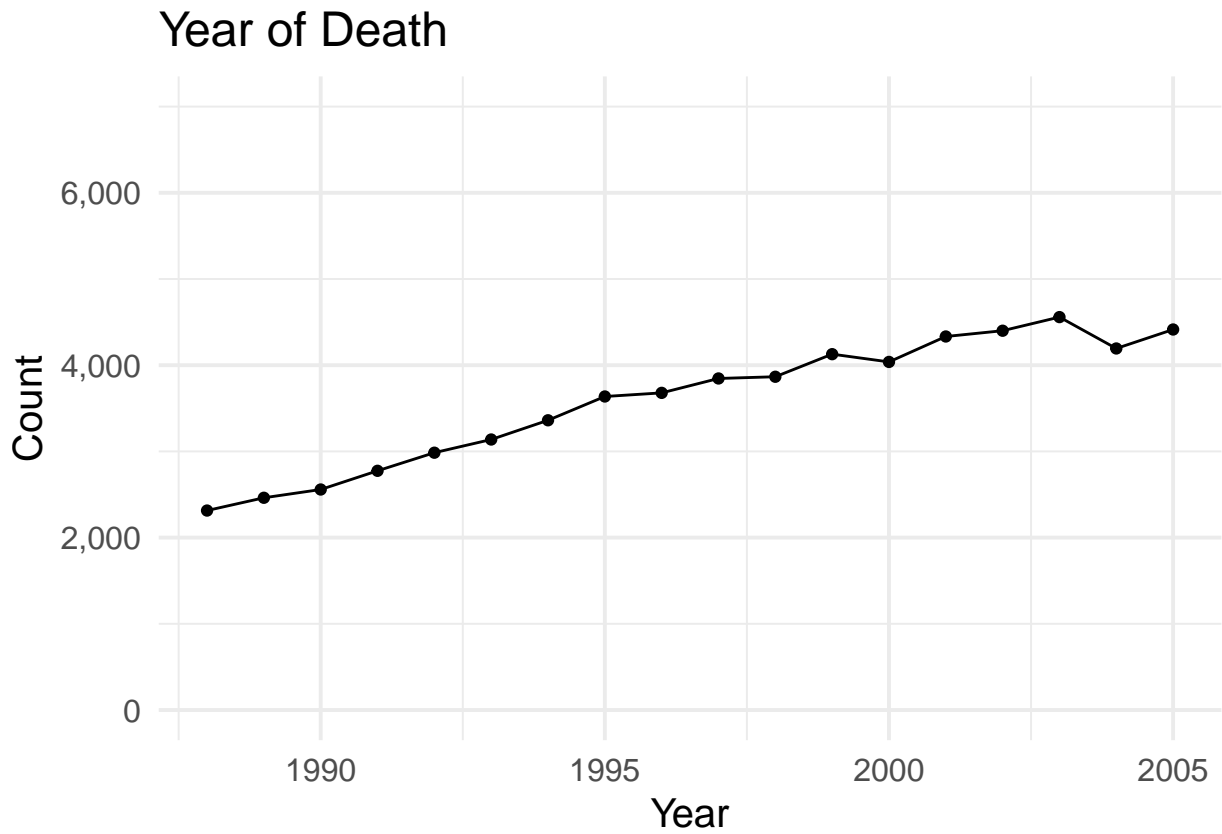
Description: bmonth reports a person's month of birth, as recorded in the Numident death records.

bmonth	label	n	freq %
1	January	5409	8.4
2	February	5140	7.9
3	March	5761	8.9
4	April	5149	8.0
5	May	5145	8.0
6	June	5247	8.1
7	July	5537	8.6
8	August	5889	9.1
9	September	5659	8.7
10	October	5456	8.4
11	November	5082	7.9
12	December	5212	8.1

dyear

Label: Death Year

Description: dyear reports a person's year of death, as recorded in the Numident death records.



dmonth

Label: Death Month

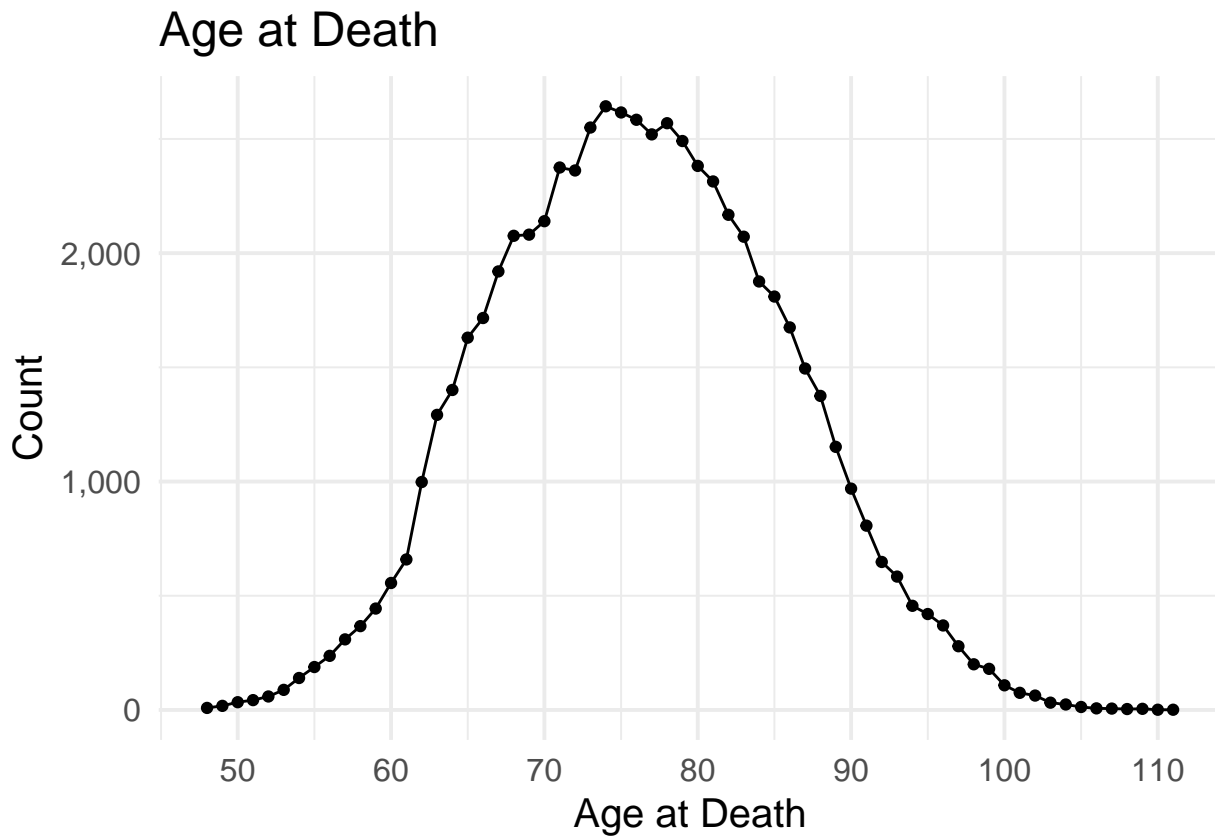
Description: dmonth reports a person's month of death, as recorded in the Numident death records.

dmonth	label	n	freq %
1	January	5995	9.3
2	February	5366	8.3
3	March	5936	9.2
4	April	5309	8.2
5	May	5394	8.3
6	June	5092	7.9
7	July	5005	7.7
8	August	5026	7.8
9	September	4965	7.7
10	October	5398	8.3
11	November	5329	8.2
12	December	5871	9.1

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 Numident death records.



SEX

Label: Sex

Description: sex reports a person's sex, as recorded in the Numident death, application, or claim records.

sex	label	n	freq %
1	Men	31237	48.3
2	Women	33449	51.7

race_first

Label: Race First

Description: race_first reports a person's race, as recorded on their first Social Security application entry.

Note: Before 1980, the race schema in the Social Security application form contained three categories: White, Black, and Other. In 1980, the SSA added three categories: (1) Asian, Asian American, or Pacific Islander, (2) Hispanic, and (3) North American Indian or Alaskan Native. The Other category was also removed.

race_first	label	n	freq %
1	White	57965	89.6000
2	Black	3857	5.9600
3	Other	284	0.4390
4	Asian	77	0.1190
5	Hispanic	46	0.0711
6	North American Native	21	0.0325
NA	Missing	2436	3.7700

race__first__cyear

Label: First Race: Application Year

Description: race__first__cyear is a numeric variable reporting the year of the application on which a person reported their first race.

race__first__cmonth

Label: First Race: Application Month

Description: race__first__cmonth is a numeric variable reporting the month of the application on which a person reported their first race.

race__last

Label: Race Last

Description: race__last reports a person's race, as recorded on their most recent Social Security application entry.

Note: Before 1980, the race schema in the Social Security application form contained three categories: White, Black, and Other. In 1980, the SSA added three categories: (1) Asian, Asian American, or Pacific Islander, (2) Hispanic, and (3) North American Indian or Alaskan Native. They also removed the Other category.

race__last	label	n	freq %
1	White	57790	89.300
2	Black	3845	5.940
3	Other	208	0.322
4	Asian	121	0.187
5	Hispanic	198	0.306
6	North American Native	88	0.136
NA	Missing	2436	3.770

race__last__cyear

Label: Last Race: Application Year

Description: race__last__cyear reports the year of the application on which a person reported their last race.

race__last__cmonth

Label: Last Race: Application Month

Description: race__last__cmonth is a numeric variable reporting the month of the application on which a person reported their last race.

bpl

Label: Birthplace

Description: bpl is a numeric variable reporting a person's place of birth, as recorded in the Numident application or claims records. The accompanying `bpl_string` variable reports the person's place of birth as a character string. The coding schema matches the **detailed** IPUMS-USA birthplace coding schema.

For a complete list of IPUMS Birthplace codes, please see: <https://usa.ipums.org/usa-action/variables/BPL>

BPL Tabulation (Native born only)

bpl	bpl_string	n	freq %	bpl	bpl_string	n	freq %
100	Alabama	1173	1.86	2900	Missouri	2039	3.23
200	Alaska	7	0.01	3000	Montana	356	0.56
400	Arizona	213	0.34	3100	Nebraska	1023	1.62
500	Arkansas	1075	1.70	3200	Nevada	40	0.06
600	California	1641	2.60	3300	New Hampshire	289	0.46
800	Colorado	657	1.04	3400	New Jersey	1809	2.86
900	Connecticut	914	1.45	3500	New Mexico	253	0.40
1000	Delaware	134	0.21	3600	New York	5242	8.29
1100	District of Columbia	213	0.34	3700	North Carolina	1579	2.50
1200	Florida	583	0.92	3800	North Dakota	634	1.00
1300	Georgia	1346	2.13	3900	Ohio	3372	5.33
1500	Hawaii	17	0.03	4000	Oklahoma	1286	2.03
1600	Idaho	329	0.52	4100	Oregon	491	0.78
1700	Illinois	3688	5.83	4200	Pennsylvania	4951	7.83
1800	Indiana	1875	2.97	4400	Rhode Island	457	0.72
1900	Iowa	1673	2.65	4500	South Carolina	872	1.38
2000	Kansas	1232	1.95	4600	South Dakota	514	0.81
2100	Kentucky	1639	2.59	4700	Tennessee	1362	2.15
2200	Louisiana	1215	1.92	4800	Texas	2667	4.22
2300	Maine	571	0.90	4900	Utah	349	0.55
2400	Maryland	889	1.41	5000	Vermont	231	0.37
2500	Massachusetts	2047	3.24	5100	Virginia	1353	2.14
2600	Michigan	2252	3.56	5300	Washington	712	1.13
2700	Minnesota	1674	2.65	5400	West Virginia	1231	1.95
2800	Mississippi	857	1.36	5500	Wisconsin	2039	3.23
				5600	Wyoming	144	0.23

zip_residence

Label: ZIP Code of Residence at Time of Death

Description: zip_residence is a string variable (9-characters) reporting a person's ZIP code of residence at time of death, as recorded in the Numident death records.

socstate

Label: State where Social Security Number Issued

Description: socstate is a numeric variable reporting the state in which a person's Social Security card was issued. It is determined by the first three (3) digits of a person's Social Security number, as recorded in Numident death records. The accompanying `socstate_string` variable reports the state in which a person's Social Security card was issued as a character string. The coding schema matches the **detailed** IPUMS-USA birthplace coding schema.

The list of codes is also available at: <https://usa.ipums.org/usa-action/variables/BPL>

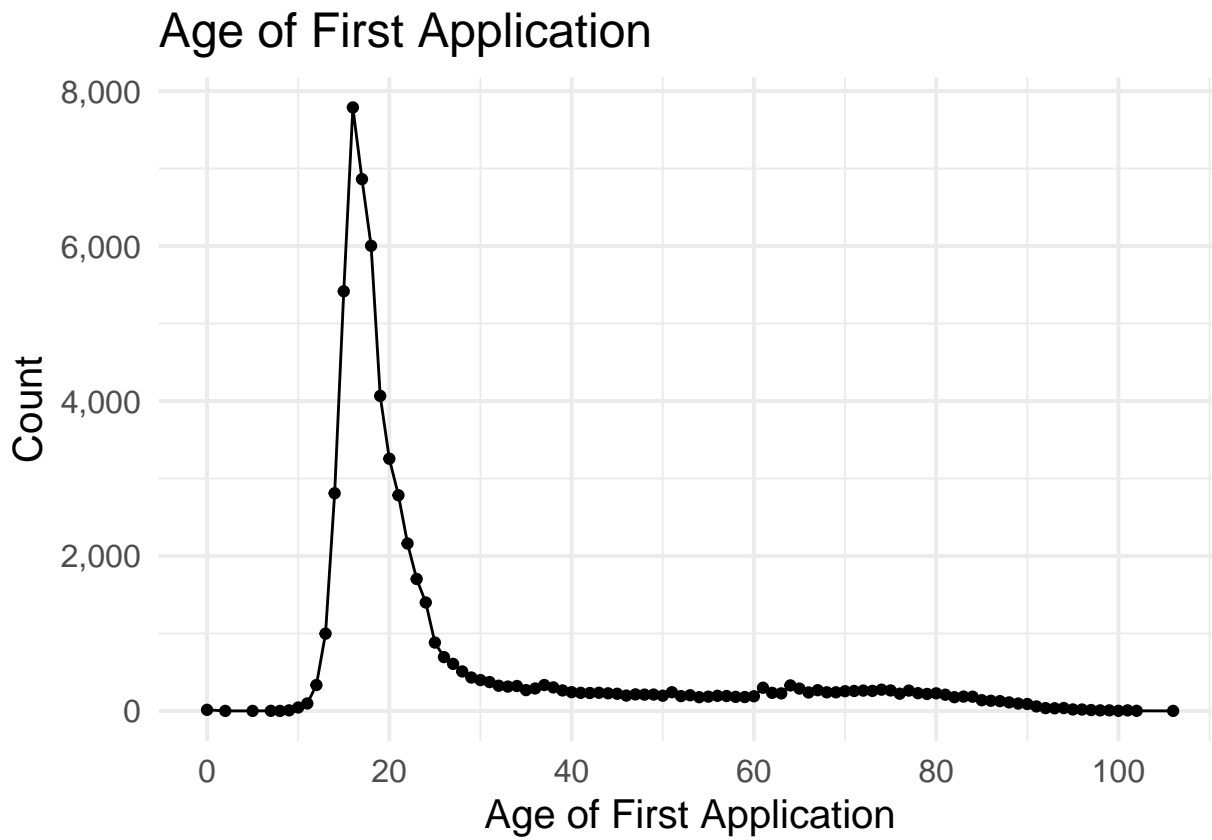
Tabulation of socstate

socstate	socstate_string	n	freq %	socstate	socstate_string	n	freq %
100	Alabama	987	1.53	3100	Nebraska	799	1.24
200	Alaska	14	0.02	3200	Nevada	59	0.09
400	Arizona	291	0.45	3300	New Hampshire	287	0.44
500	Arkansas	710	1.10	3400	New Jersey	2012	3.11
600	California	3462	5.35	3500	New Mexico	229	0.35
800	Colorado	679	1.05	3600	New York	5844	9.03
900	Connecticut	1017	1.57	3700	North Carolina	1472	2.28
1000	Delaware	139	0.21	3800	North Dakota	384	0.59
1100	District of Columbia	371	0.57	3900	Ohio	3721	5.75
1200	Florida	950	1.47	4000	Oklahoma	1013	1.57
1300	Georgia	1194	1.85	4100	Oregon	689	1.07
1500	Hawaii	12	0.02	4200	Pennsylvania	4603	7.12
1600	Idaho	306	0.47	4400	Rhode Island	450	0.70
1700	Illinois	3883	6.00	4500	South Carolina	711	1.10
1800	Indiana	2022	3.13	4600	South Dakota	360	0.56
1900	Iowa	1357	2.10	4700	Tennessee	1266	1.96
2000	Kansas	1012	1.56	4800	Texas	2815	4.35
2100	Kentucky	1244	1.92	4900	Utah	315	0.49
2200	Louisiana	1135	1.75	5000	Vermont	179	0.28
2300	Maine	541	0.84	5100	Virginia	1287	1.99
2400	Maryland	981	1.52	5300	Washington	1051	1.62
2500	Massachusetts	2135	3.30	5400	West Virginia	1053	1.63
2600	Michigan	2649	4.10	5500	Wisconsin	1909	2.95
2700	Minnesota	1627	2.52	5600	Wyoming	165	0.26
2800	Mississippi	663	1.02	10500	Guam	1	0.00
2900	Missouri	1955	3.02	11000	Puerto Rico	2	0.00
3000	Montana	283	0.44	NA	NA	391	0.60

age_first_app

Label: Age at First Social Security Application

Description: age_first_application reports the age at which a person submitted their first Social Security application.

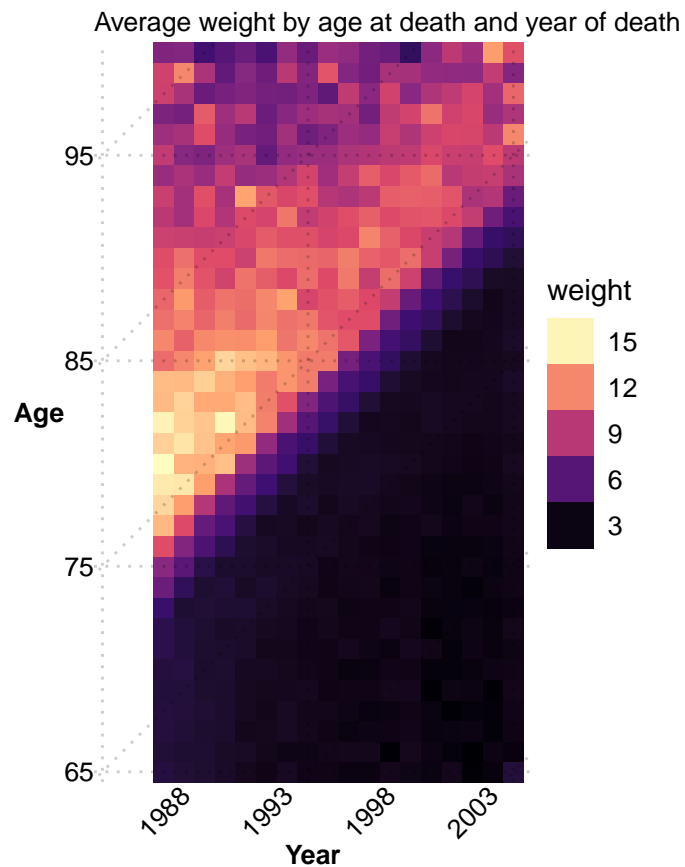


weight

Label: CenSoc Sample Weight ¹

Description: weight is a post-stratification person-weight to National Center for Health Statistics (NCHS) totals for persons (1) dying between 1988-2005 (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
1.01	Min Weight
22.14	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 Census 1% 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
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 ⁷
rent	Monthly contract rent
serial	Household serial number
statefip	State of residence 1940 (FIPS codes)
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.

⁷The IPUMS **race** variable reports race as recorded in the 1940 Census. In contrast, the **race_first** and **race_last** variables in this dataset contain race as self-reported on Social Security applications.