

## 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, ownership, race, rent, serial, statefip, and urban

**Summary:** The CenSoc-Numident Version 3 Demo dataset (N = 70,320) 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).

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 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

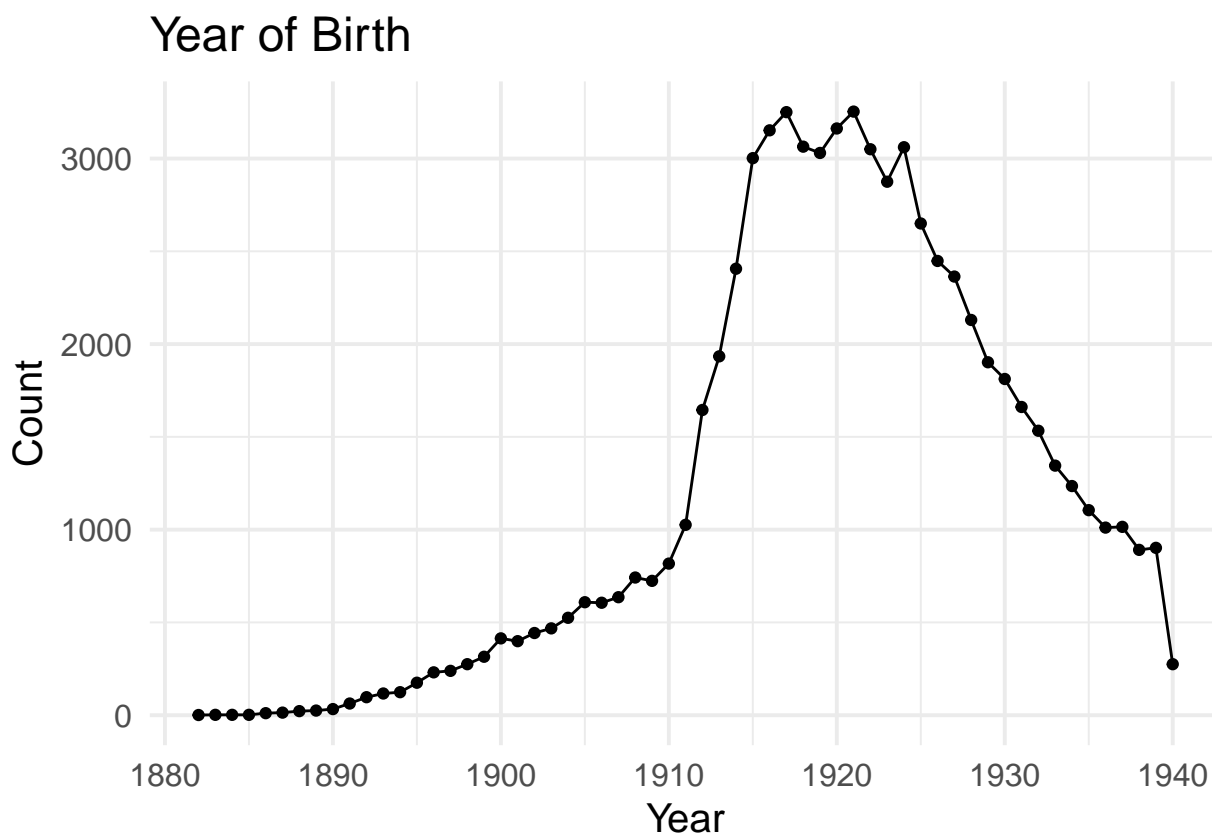
**Label:** Historical Unique Identifier

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

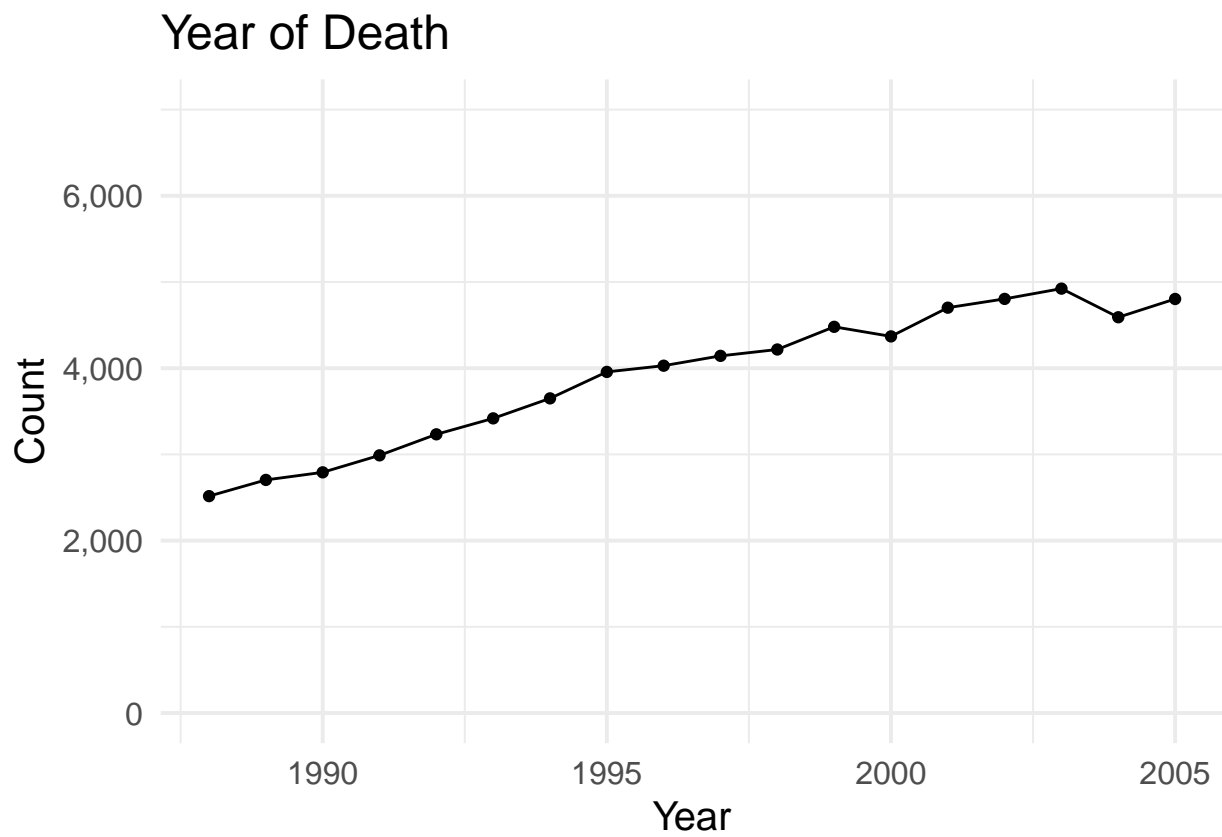
**Description:** bmonth reports a person's month of birth, as recorded in the Numident death records.

bmonth	label	n	freq %
1	January	5891	8.4
2	February	5593	8.0
3	March	6297	9.0
4	April	5564	7.9
5	May	5620	8.0
6	June	5709	8.1
7	July	5994	8.5
8	August	6383	9.1
9	September	6129	8.7
10	October	5935	8.4
11	November	5529	7.9
12	December	5676	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

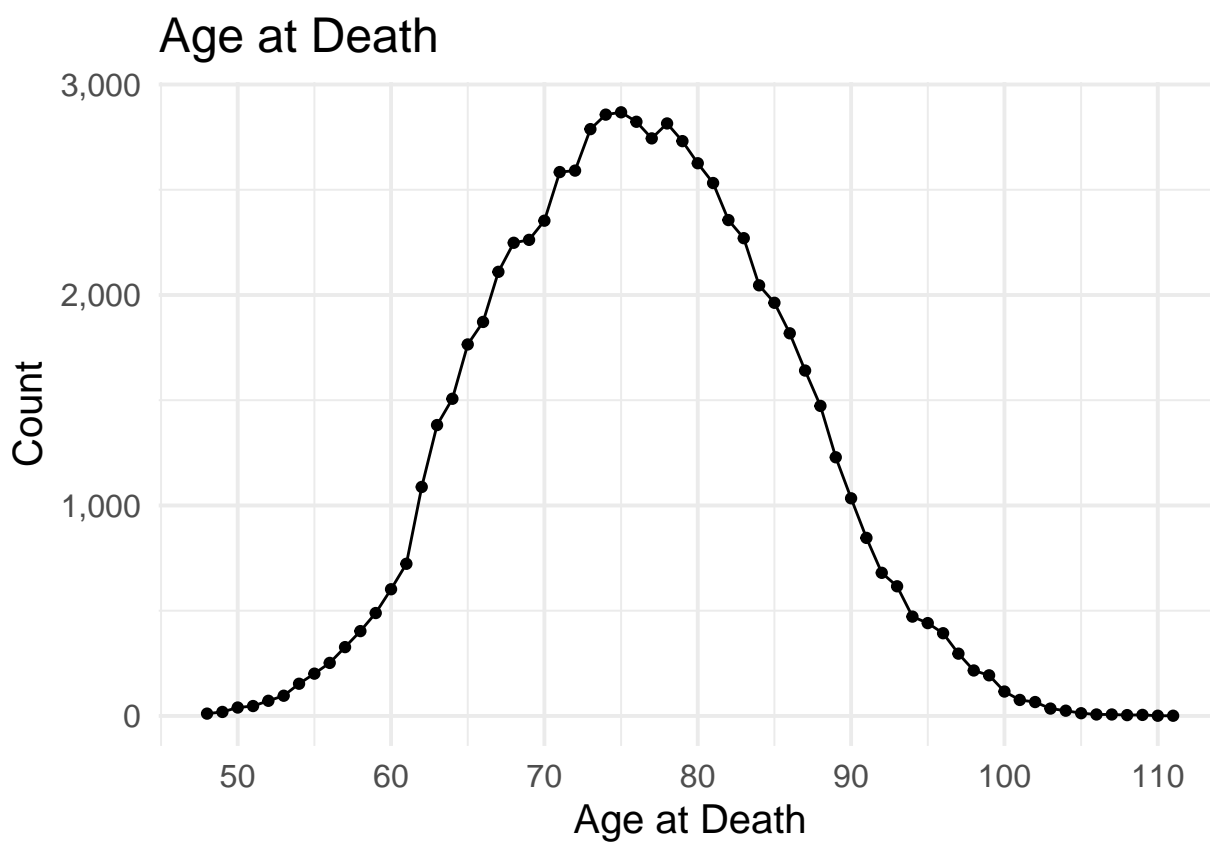
**Description:** dmonth reports a person's month of death, as recorded in the Numident death records.

dmonth	label	n	freq %
1	January	6530	9.3
2	February	5831	8.3
3	March	6442	9.2
4	April	5806	8.3
5	May	5832	8.3
6	June	5501	7.8
7	July	5448	7.7
8	August	5474	7.8
9	September	5394	7.7
10	October	5857	8.3
11	November	5810	8.3
12	December	6395	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	34119	48.5
2	Women	36201	51.5



# race\_\_first

**Label:** Race First

**Description:** race\_\_first reports a person's race, as recorded on their first 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	62789	89.3000
2	Black	4444	6.3200
3	Other	328	0.4660
4	Asian	83	0.1180
5	Hispanic	54	0.0768
6	North American Native	21	0.0299
NA	Missing	2601	3.7000

# 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 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	62598	89.000
2	Black	4435	6.310
3	Other	235	0.334
4	Asian	131	0.186
5	Hispanic	218	0.310
6	North American Native	102	0.145
NA	Missing	2601	3.700

# 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	1289	1.88	2900	Missouri	2181	3.17
200	Alaska	7	0.01	3000	Montana	385	0.56
400	Arizona	235	0.34	3100	Nebraska	1091	1.59
500	Arkansas	1166	1.70	3200	Nevada	49	0.07
600	California	1771	2.58	3300	New Hampshire	324	0.47
800	Colorado	699	1.02	3400	New Jersey	1978	2.88
900	Connecticut	1007	1.47	3500	New Mexico	280	0.41
1000	Delaware	143	0.21	3600	New York	5700	8.29
1100	District of Columbia	254	0.37	3700	North Carolina	1758	2.56
1200	Florida	655	0.95	3800	North Dakota	689	1.00
1300	Georgia	1491	2.17	3900	Ohio	3639	5.29
1500	Hawaii	17	0.02	4000	Oklahoma	1422	2.07
1600	Idaho	340	0.49	4100	Oregon	530	0.77
1700	Illinois	3988	5.80	4200	Pennsylvania	5395	7.85
1800	Indiana	2020	2.94	4400	Rhode Island	490	0.71
1900	Iowa	1802	2.62	4500	South Carolina	959	1.40
2000	Kansas	1311	1.91	4600	South Dakota	548	0.80
2100	Kentucky	1768	2.57	4700	Tennessee	1500	2.18
2200	Louisiana	1344	1.96	4800	Texas	2886	4.20
2300	Maine	631	0.92	4900	Utah	381	0.55
2400	Maryland	1002	1.46	5000	Vermont	262	0.38
2500	Massachusetts	2221	3.23	5100	Virginia	1509	2.20
2600	Michigan	2430	3.54	5300	Washington	777	1.13
2700	Minnesota	1787	2.60	5400	West Virginia	1338	1.95
2800	Mississippi	960	1.40	5500	Wisconsin	2176	3.17
				5600	Wyoming	152	0.22

# 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:** The state in which a person's social security card was issued. Determined by first three (3) digits of 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.

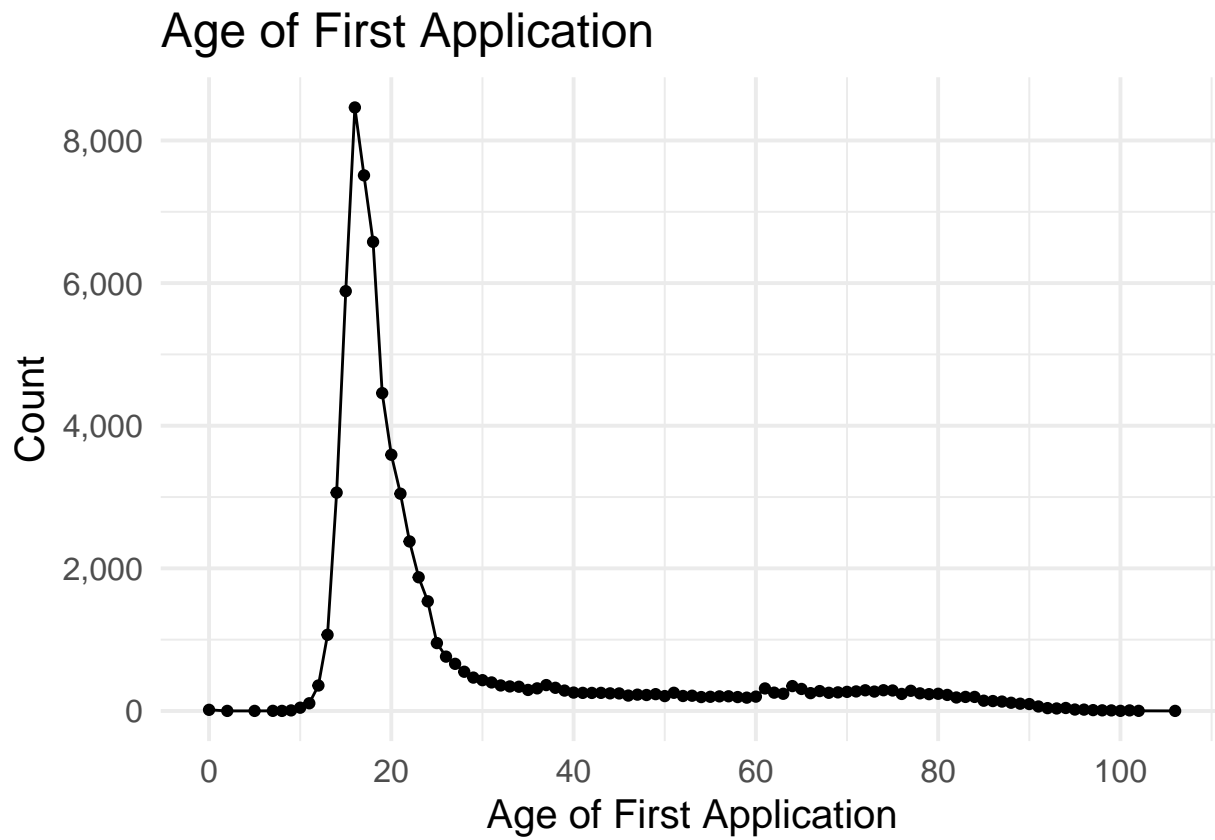
Tabulation of socstate

socstate	socstate_string	n	freq %	socstate	socstate_string	n	freq %
100	Alabama	1077	1.53	3100	Nebraska	847	1.20
200	Alaska	15	0.02	3200	Nevada	63	0.09
400	Arizona	319	0.45	3300	New Hampshire	325	0.46
500	Arkansas	769	1.09	3400	New Jersey	2208	3.14
600	California	3742	5.32	3500	New Mexico	251	0.36
800	Colorado	722	1.03	3600	New York	6362	9.05
900	Connecticut	1116	1.59	3700	North Carolina	1636	2.33
1000	Delaware	152	0.22	3800	North Dakota	417	0.59
1100	District of Columbia	422	0.60	3900	Ohio	4021	5.72
1200	Florida	1054	1.50	4000	Oklahoma	1120	1.59
1300	Georgia	1326	1.89	4100	Oregon	753	1.07
1500	Hawaii	12	0.02	4200	Pennsylvania	5024	7.14
1600	Idaho	319	0.45	4400	Rhode Island	483	0.69
1700	Illinois	4201	5.97	4500	South Carolina	784	1.11
1800	Indiana	2165	3.08	4600	South Dakota	386	0.55
1900	Iowa	1467	2.09	4700	Tennessee	1399	1.99
2000	Kansas	1085	1.54	4800	Texas	3050	4.34
2100	Kentucky	1344	1.91	4900	Utah	350	0.50
2200	Louisiana	1247	1.77	5000	Vermont	198	0.28
2300	Maine	600	0.85	5100	Virginia	1433	2.04
2400	Maryland	1090	1.55	5300	Washington	1132	1.61
2500	Massachusetts	2313	3.29	5400	West Virginia	1163	1.65
2600	Michigan	2862	4.07	5500	Wisconsin	2034	2.89
2700	Minnesota	1741	2.48	5600	Wyoming	171	0.24
2800	Mississippi	734	1.04	10500	Guam	1	0.00
2900	Missouri	2080	2.96	11000	Puerto Rico	4	0.01
3000	Montana	308	0.44	NA	NA	423	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.

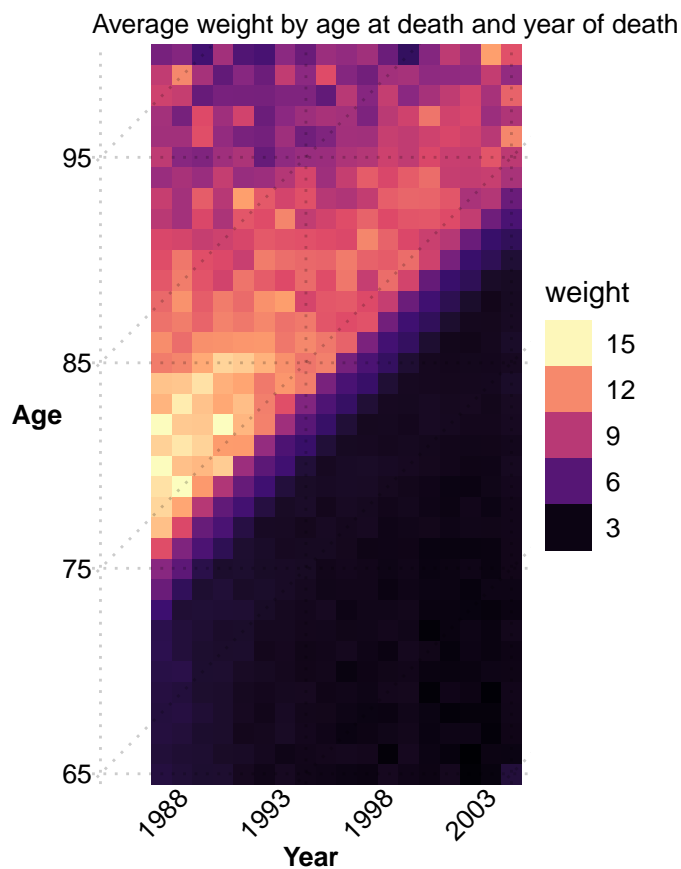


# weight

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

**Description:** 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



<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 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 <sup>2</sup>
age	Age on April 1st, 1940
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 <sup>7</sup>
rent	Monthly contract rent
serial	Household serial number
statefip	State of residence 1940 (FIPS codes)
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.

<sup>7</sup>The 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.