CenSoc-Numident Codebook*

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	link_abe_exact_conservative	Flag for conservative ABE match
20	weight	CenSoc weight
21	weight_conservative	CenSoc weight (Conservative Sample)
22	Additional IPUMS variables	Additional variables include pernum, perwt, age, mbpl, fbpl, educd, empstatd, hispan, inconwg, marst, nativity, occ, occscore, ownership, race, rent, serial, statefip, urban

Summary: The CenSoc-Numident Demo dataset (N = 85,865) was constructed by (i) linking the CenSoc-Numident dataset to the IPUMS 1940 1% census sample and (ii) selecting a set of 20 mortality covariates from the 1940 census. The smaller size of the file — approximately 1% of the records in the full CenSoc-Numident dataset — makes it easier to work with. However, the demo file isn't conducive to high-resolution mortality research, and we recommend working with the complete CenSoc-Numident file for any final analysis.
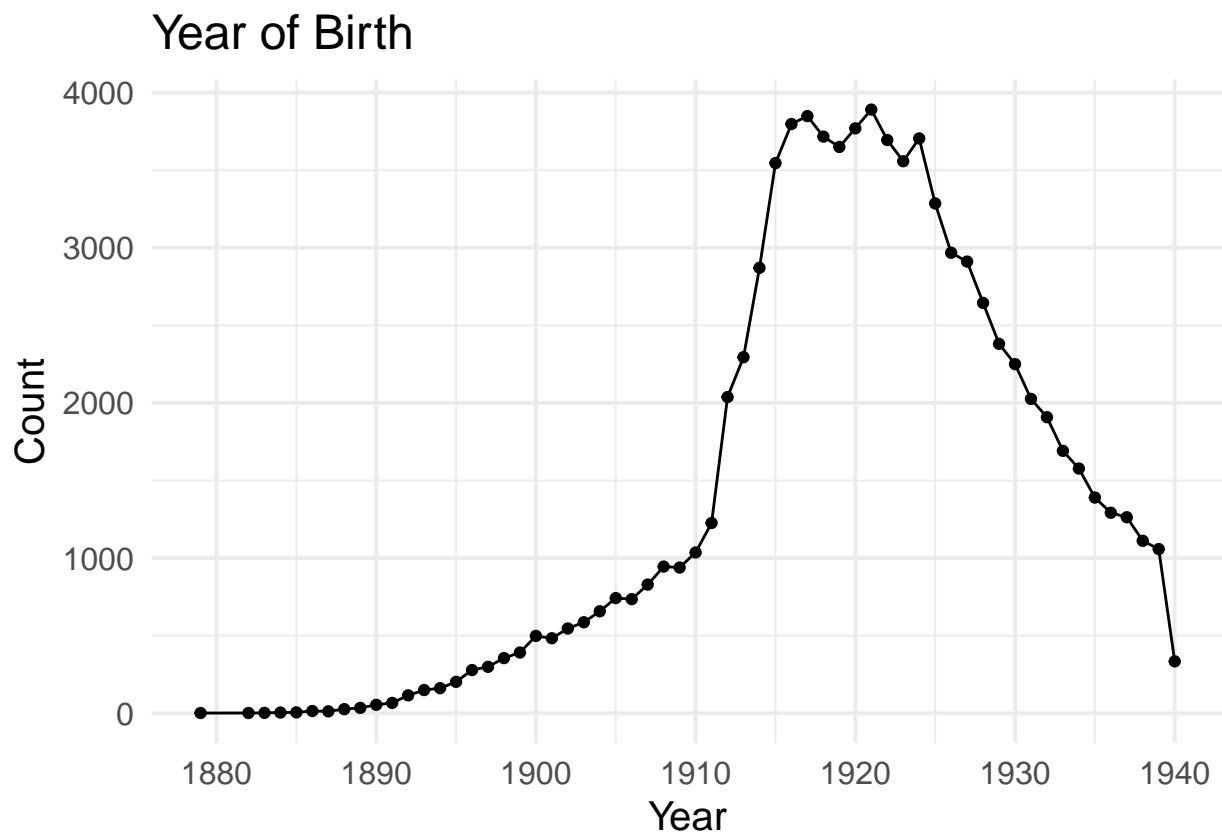
---

*Last updated: 11 January, 2023

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.
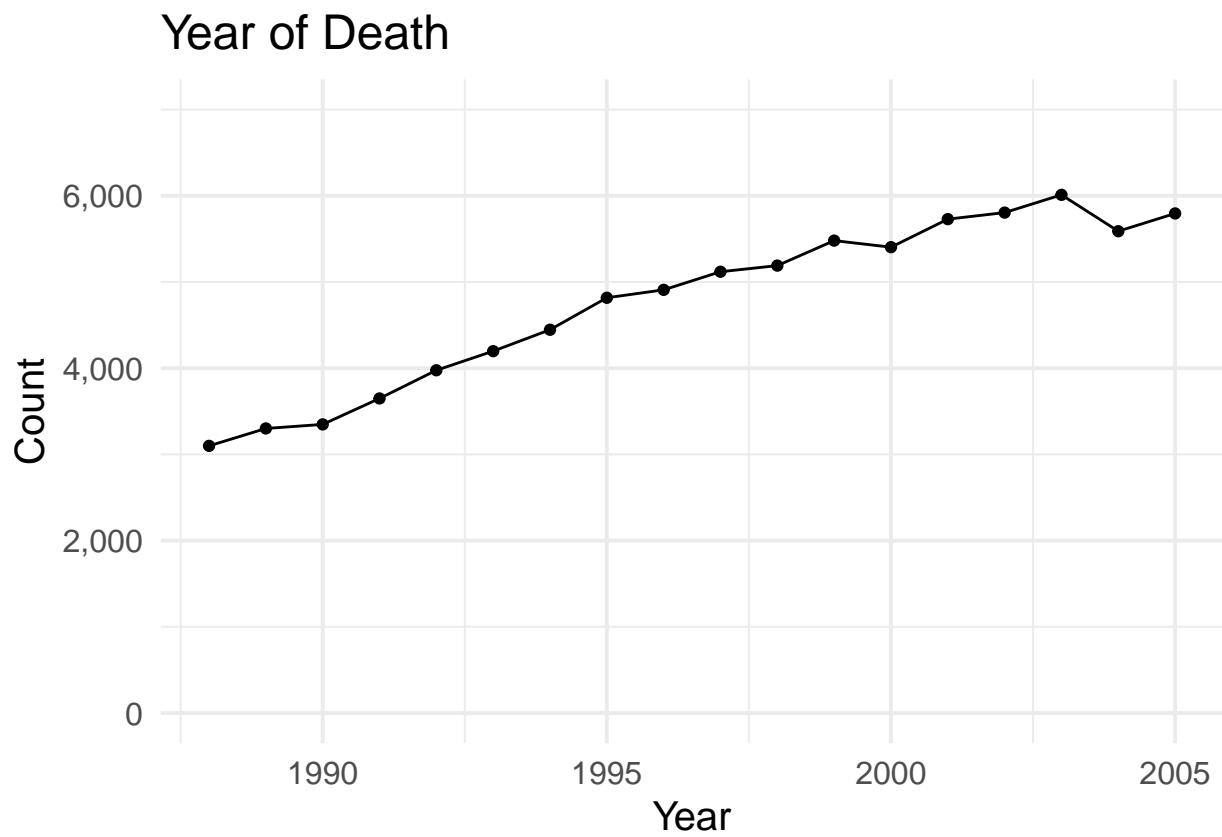
Year of Birth

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	7233	8.4
2	February	6877	8.0
3	March	7590	8.8
4	April	6909	8.0
5	May	6866	8.0
6	June	6967	8.1
7	July	7319	8.5
8	August	7762	9.0
9	September	7494	8.7
10	October	7205	8.4
11	November	6740	7.8
12	December	6903	8.0

dyear

Label: Death Year

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

Year of Death

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	7970	9.3
2	February	7155	8.3
3	March	7839	9.1
4	April	7009	8.2
5	May	7147	8.3
6	June	6722	7.8
7	July	6670	7.8
8	August	6683	7.8
9	September	6576	7.7
10	October	7145	8.3
11	November	7095	8.3
12	December	7854	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.

Age at Death
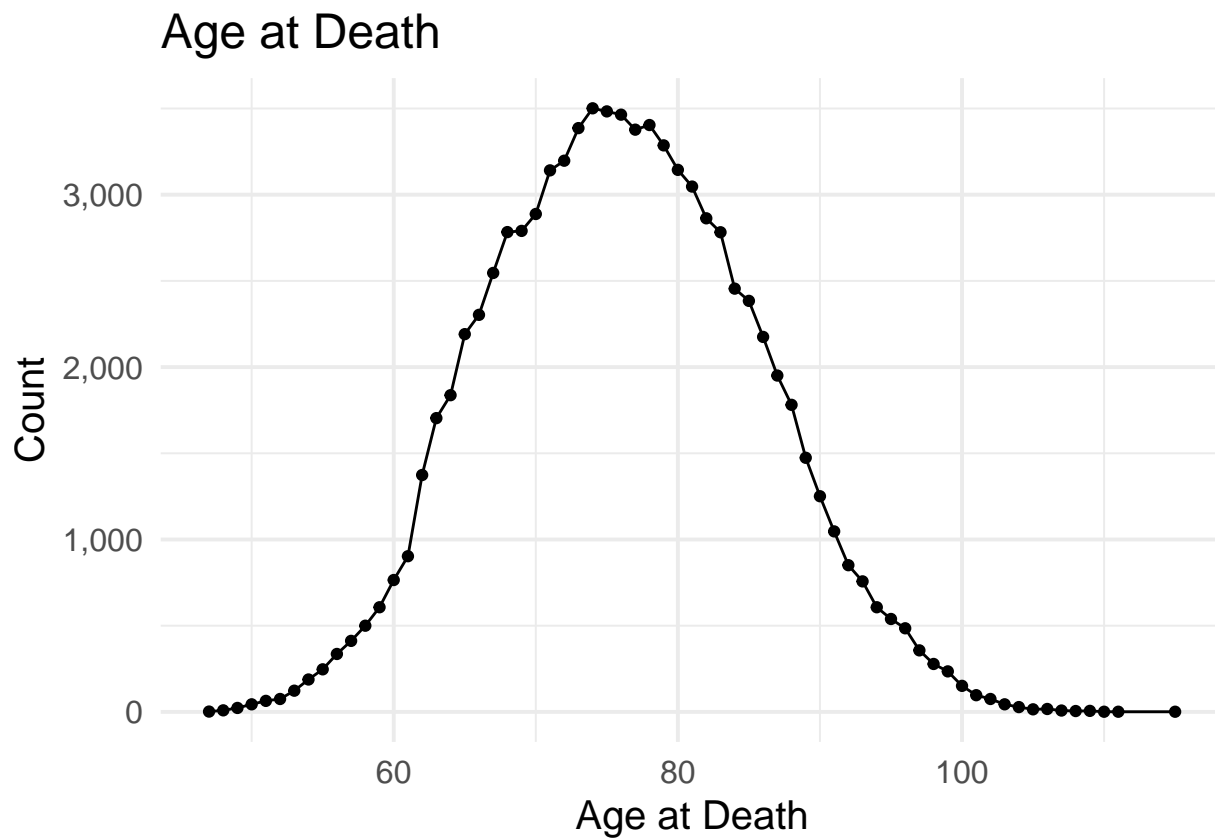
sex

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	41761	48.6
2	Women	44104	51.4

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	75633	88.1000
2	Black	6165	7.1800
3	Other	350	0.4080
4	Asian	99	0.1150
5	Hispanic	84	0.0978
6	North American Native	25	0.0291
NA	Missing	3509	4.0900

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	75355	87.800
2	Black	6144	7.160
3	Other	258	0.300
4	Asian	158	0.184
5	Hispanic	334	0.389
6	North American Native	107	0.125
NA	Missing	3509	4.090

race_last_cyear

Label: First 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	1786	2.13	2900	Missouri	2669	3.19
200	Alaska	7	0.01	3000	Montana	388	0.46
400	Arizona	241	0.29	3100	Nebraska	1185	1.42
500	Arkansas	1453	1.74	3200	Nevada	38	0.05
600	California	2106	2.52	3300	New Hampshire	338	0.40
800	Colorado	774	0.92	3400	New Jersey	2281	2.72
900	Connecticut	1105	1.32	3500	New Mexico	336	0.40
1000	Delaware	151	0.18	3600	New York	7218	8.62
1100	District of Columbia	255	0.30	3700	North Carolina	2502	2.99
1200	Florida	799	0.95	3800	North Dakota	703	0.84
1300	Georgia	2132	2.55	3900	Ohio	4441	5.30
1500	Hawaii	19	0.02	4000	Oklahoma	1745	2.08
1600	Idaho	365	0.44	4100	Oregon	544	0.65
1700	Illinois	4808	5.74	4200	Pennsylvania	6837	8.17
1800	Indiana	2344	2.80	4400	Rhode Island	510	0.61
1900	Iowa	2014	2.41	4500	South Carolina	1325	1.58
2000	Kansas	1506	1.80	4600	South Dakota	569	0.68
2100	Kentucky	2314	2.76	4700	Tennessee	2023	2.42
2200	Louisiana	1610	1.92	4800	Texas	3841	4.59
2300	Maine	690	0.82	4900	Utah	416	0.50
2400	Maryland	1125	1.34	5000	Vermont	253	0.30
2500	Massachusetts	2861	3.42	5100	Virginia	1939	2.32
2600	Michigan	2775	3.31	5300	Washington	854	1.02
2700	Minnesota	1975	2.36	5400	West Virginia	1638	1.96
2800	Mississippi	1334	1.59	5500	Wisconsin	2429	2.90
				5600	Wyoming	153	0.18

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 (Native born only)

socstate	socstate_string	n	freq %	socstate	socstate_string	n	freq %
100	Alabama	1474	1.72	3000	Montana	326	0.38
200	Alaska	14	0.02	3100	Nebraska	917	1.07
400	Arizona	361	0.42	3200	Nevada	62	0.07
500	Arkansas	989	1.15	3300	New Hampshire	355	0.41
600	California	4472	5.21	3400	New Jersey	2585	3.01
800	Colorado	824	0.96	3500	New Mexico	303	0.35
900	Connecticut	1244	1.45	3600	New York	8028	9.35
1000	Delaware	174	0.20	3700	North Carolina	2301	2.68
1100	District of Columbia	482	0.56	3800	North Dakota	440	0.51
1200	Florida	1333	1.55	3900	Ohio	4962	5.78
1300	Georgia	1860	2.17	4000	Oklahoma	1356	1.58
1500	Hawaii	16	0.02	4100	Oregon	791	0.92
1600	Idaho	354	0.41	4200	Pennsylvania	6329	7.37
1700	Illinois	5050	5.88	4400	Rhode Island	526	0.61
1800	Indiana	2541	2.96	4500	South Carolina	1068	1.24
1900	Iowa	1644	1.91	4600	South Dakota	402	0.47
2000	Kansas	1274	1.48	4700	Tennessee	1877	2.19
2100	Kentucky	1756	2.05	4800	Texas	3991	4.65
2200	Louisiana	1504	1.75	4900	Utah	376	0.44
2300	Maine	672	0.78	5000	Vermont	210	0.24
2400	Maryland	1272	1.48	5100	Virginia	1858	2.16
2500	Massachusetts	2952	3.44	5300	Washington	1255	1.46
2600	Michigan	3345	3.90	5400	West Virginia	1419	1.65
2700	Minnesota	1908	2.22	5500	Wisconsin	2292	2.67
2800	Mississippi	1033	1.20	5600	Wyoming	196	0.23
2900	Missouri	2557	2.98	NA	NA	524	0.61

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.

Age of First Application
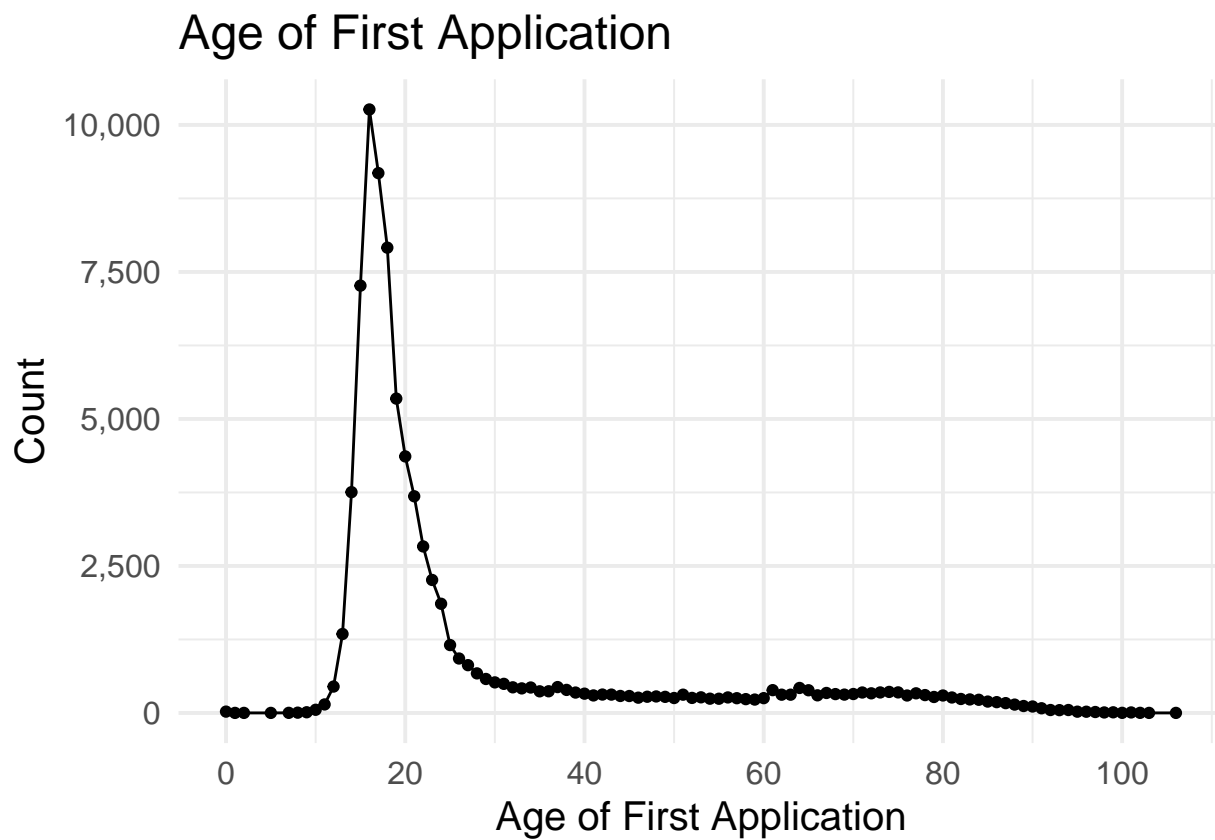
link_abe_exact_conservative

Label: Flag for conservative ABE match

Description: A flag variable reporting whether a match was established with the ABE conservative match with exact names.

Note: All matches were established with the standard ABE match. A subset of these records were also matched with the conservative ABE match.

link_abe_exact_conservative	label	n	freq %
1	Match Established with Conservative ABE Algorithm	64488	75.1
0	Not Established	21377	24.9

weight

Label: Sample Weight [1]

Description: A post-stratification person-weight to Human Mortality Database (HMD) totals for persons (1) born between 1895-1939 (2) dying between 1988-2005 (3) dying between ages 65-100. Please see the CenSoc Methods Protocol for more details on weighting procedure.

| Value	Label
2.33	Min Weight
24.47	Max Weight
NA	No Weight Assigned

Average weight by age at death and year of death

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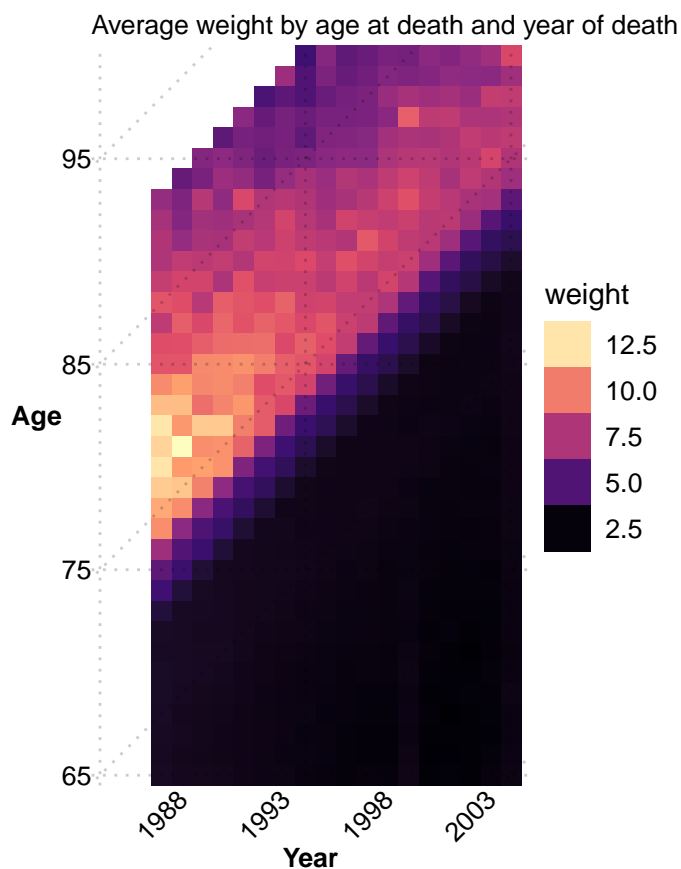
[1]The IPUMS-USA 1940 1% sample also includes a weight (`perweight`) to account for the 1940 sampling procedure (thus no weights for the 100% complete count 1940 census). For analysis, we recommend using both sets of weights. A final weight can be constructed by multiplying the two weights together.

weight_conservative

Label: Sample Weights (Conservative Sample)

Description: A post-stratification person-weight to Human Mortality Database (HMD) totals for persons (1) born between 1895-1939 (2) dying between 1988-2005 (3) dying between ages 65-100. Please see the CenSoc Methods Protocol for more details on weighting procedure.

Value	Label
3.18	Min Weight
36.18	Max Weight
NA	No Weight Assigned

Average weight by age at death and year of death
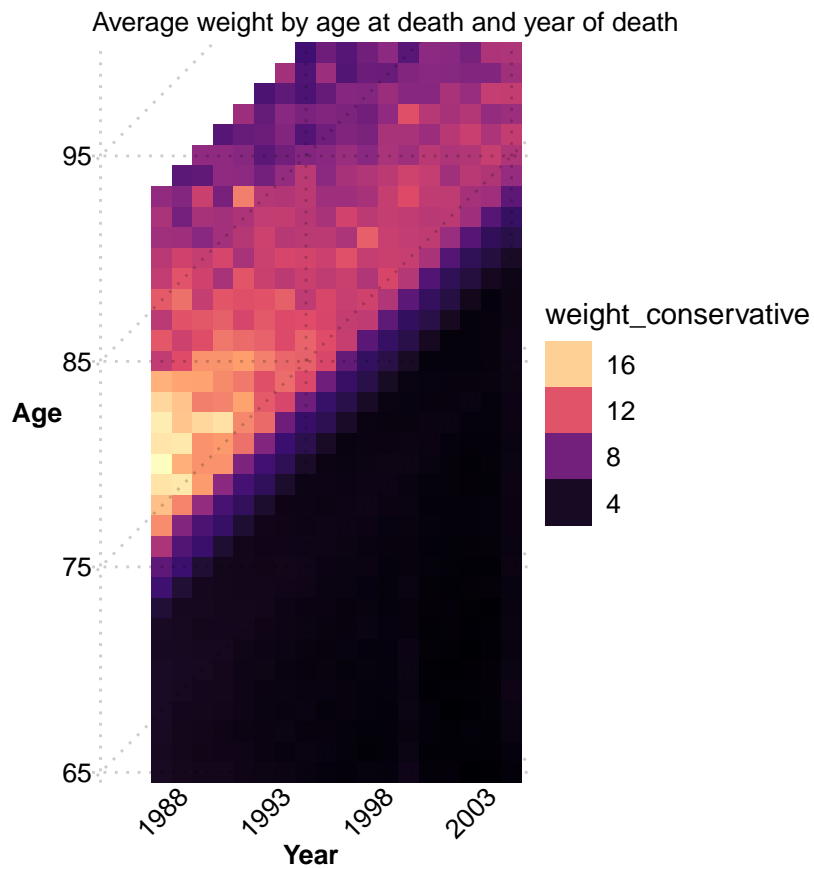
IPUMS 1940 Census Variable

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 sample weight
perwt	IPUMS person weight[2]
age	Age in 1940
mbpl	Mother's place of birth
fbpl	Father's place of birth
educd	Educational attainment (detailed)
empstatd	Employment status (detailed)
hispan	Hispanic/Spanish/Latino origin
inconwg	Had non-wage/salary income over $50
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
urban	Urban/rural status

---

[2]The `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