

CenSoc-Numident 3.0 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	weight	CenSoc weight

Summary: The CenSoc-Numident 3.0 dataset (N = 7,044,519) links the full-count 1940 Census to the National Archives' public release of the Social Security Numident file. Records were linked using the conservative variant of the ABE method developed by Abramitzky, Boustan, and Eriksson (2012, 2014, 2017). To merge 1940 Census variables with this dataset, researchers can obtain a copy of the 1940 Census from IPUMS-USA and link on the individual-level, unique identifier HISTID variable.

*Last updated: 20 October, 2023

HISTID

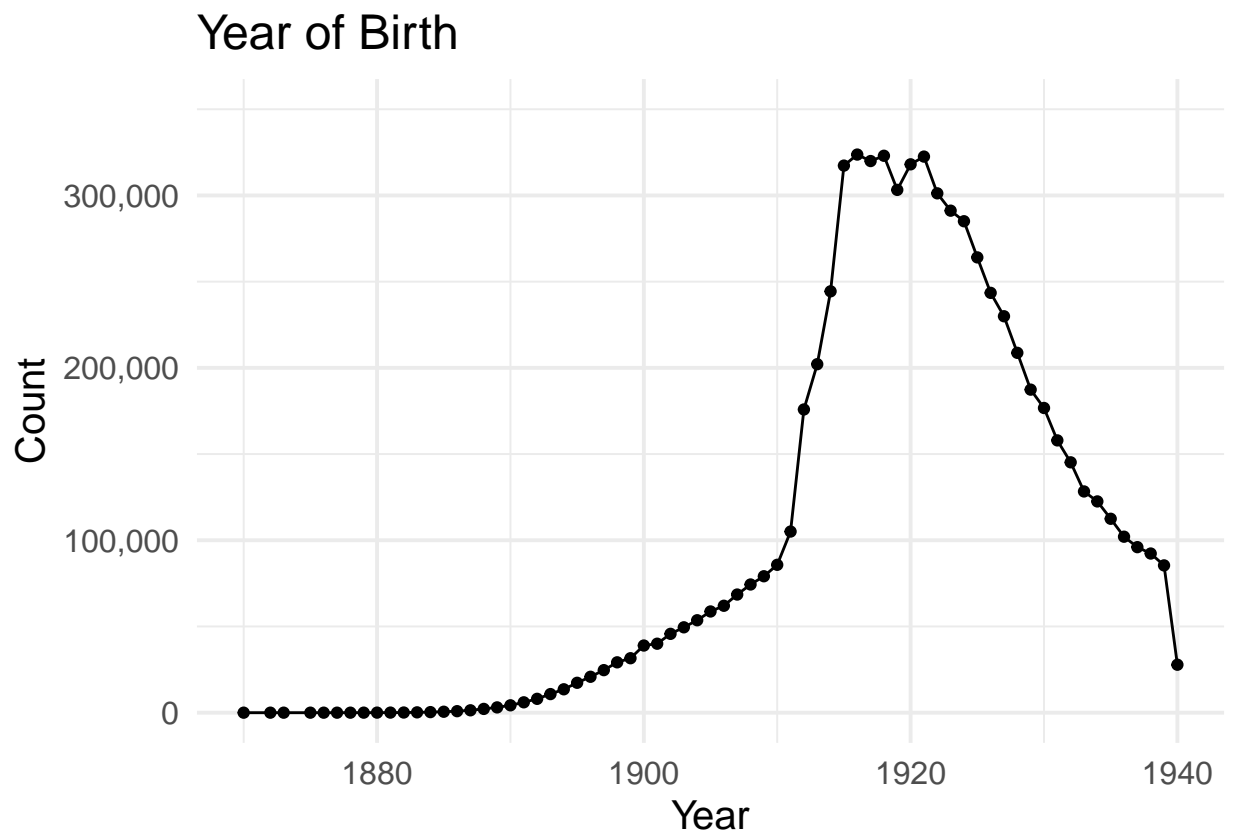
Label: Historical Unique Identifier

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.



bmonth

Label: Birth Month

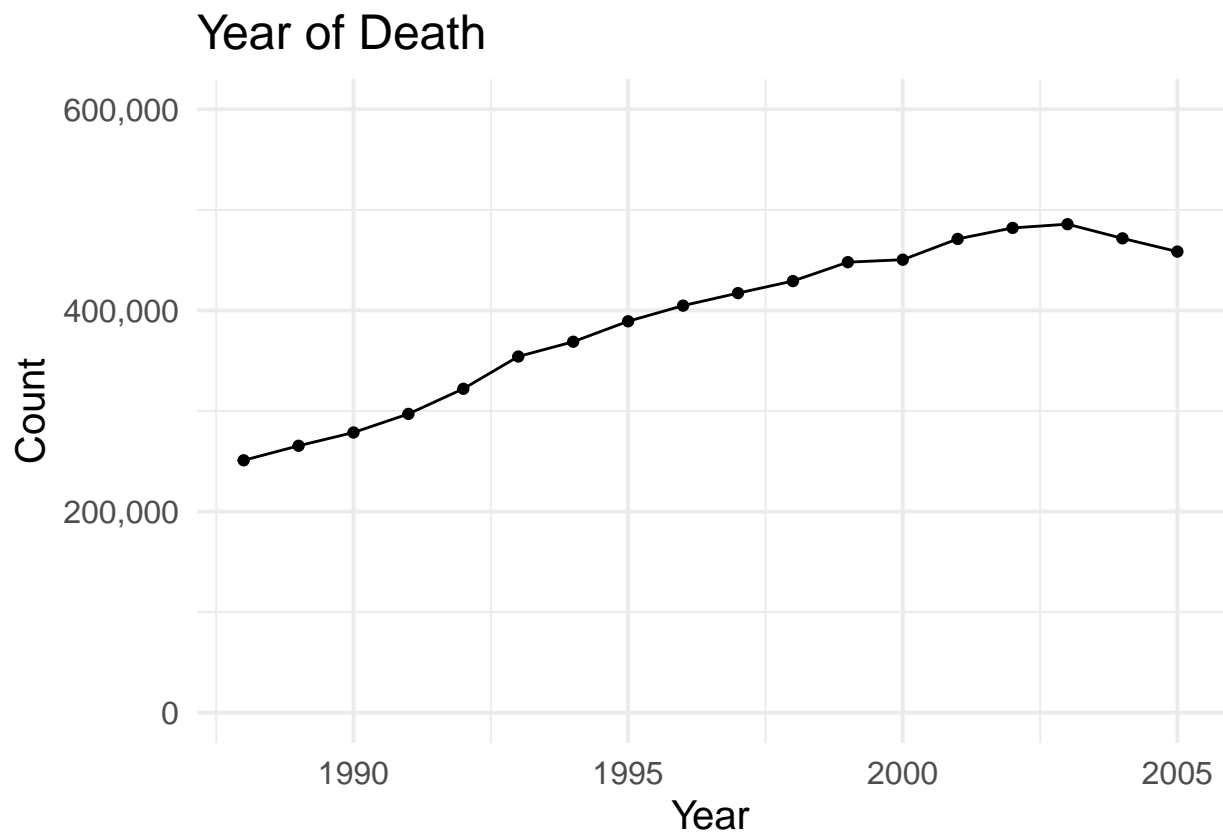
Description: bmonth reports a person's month of birth, as recorded in the Numident death records.

bmonth	label	n	freq %
1	January	594575	8.4
2	February	565389	8.0
3	March	619697	8.8
4	April	564763	8.0
5	May	577784	8.2
6	June	564560	8.0
7	July	603007	8.6
8	August	621631	8.8
9	September	609386	8.7
10	October	594495	8.4
11	November	555750	7.9
12	December	573482	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	665743	9.5
2	February	590986	8.4
3	March	627892	8.9
4	April	579140	8.2
5	May	576189	8.2
6	June	544328	7.7
7	July	552803	7.8
8	August	550451	7.8
9	September	540396	7.7
10	October	584498	8.3
11	November	580383	8.2
12	December	651710	9.3

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	3423848	48.6
2	Women	3620671	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	6305289	89.5000
2	Black	427034	6.0600
3	Other	30372	0.4310
4	Asian	7247	0.1030
5	Hispanic	6823	0.0969
6	North American Native	1821	0.0258
NA	Missing	265933	3.7800

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	6285489	89.200
2	Black	426055	6.050
3	Other	22482	0.319
4	Asian	12325	0.175
5	Hispanic	23764	0.337
6	North American Native	8471	0.120
NA	Missing	265933	3.780

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	140420	2.04	2900	Missouri	226926	3.30
200	Alaska	699	0.01	3000	Montana	38778	0.56
400	Arizona	23805	0.35	3100	Nebraska	109439	1.59
500	Arkansas	114870	1.67	3200	Nevada	5103	0.07
600	California	197375	2.87	3300	New Hampshire	31233	0.45
800	Colorado	71351	1.04	3400	New Jersey	199061	2.89
900	Connecticut	100541	1.46	3500	New Mexico	25260	0.37
1000	Delaware	16299	0.24	3600	New York	562570	8.18
1100	District of Columbia	23732	0.34	3700	North Carolina	166808	2.42
1200	Florida	66930	0.97	3800	North Dakota	61599	0.90
1300	Georgia	142994	2.08	3900	Ohio	368860	5.36
1500	Hawaii	1617	0.02	4000	Oklahoma	148309	2.16
1600	Idaho	36340	0.53	4100	Oregon	54065	0.79
1700	Illinois	407486	5.92	4200	Pennsylvania	521806	7.58
1800	Indiana	201473	2.93	4400	Rhode Island	45796	0.67
1900	Iowa	177557	2.58	4500	South Carolina	87144	1.27
2000	Kansas	131834	1.92	4600	South Dakota	57740	0.84
2100	Kentucky	165500	2.41	4700	Tennessee	145762	2.12
2200	Louisiana	128401	1.87	4800	Texas	299819	4.36
2300	Maine	61582	0.90	4900	Utah	44316	0.64
2400	Maryland	95020	1.38	5000	Vermont	28465	0.41
2500	Massachusetts	222992	3.24	5100	Virginia	138425	2.01
2600	Michigan	272250	3.96	5300	Washington	81638	1.19
2700	Minnesota	177659	2.58	5400	West Virginia	122476	1.78
2800	Mississippi	99552	1.45	5500	Wisconsin	213082	3.10
				5600	Wyoming	17144	0.25

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.

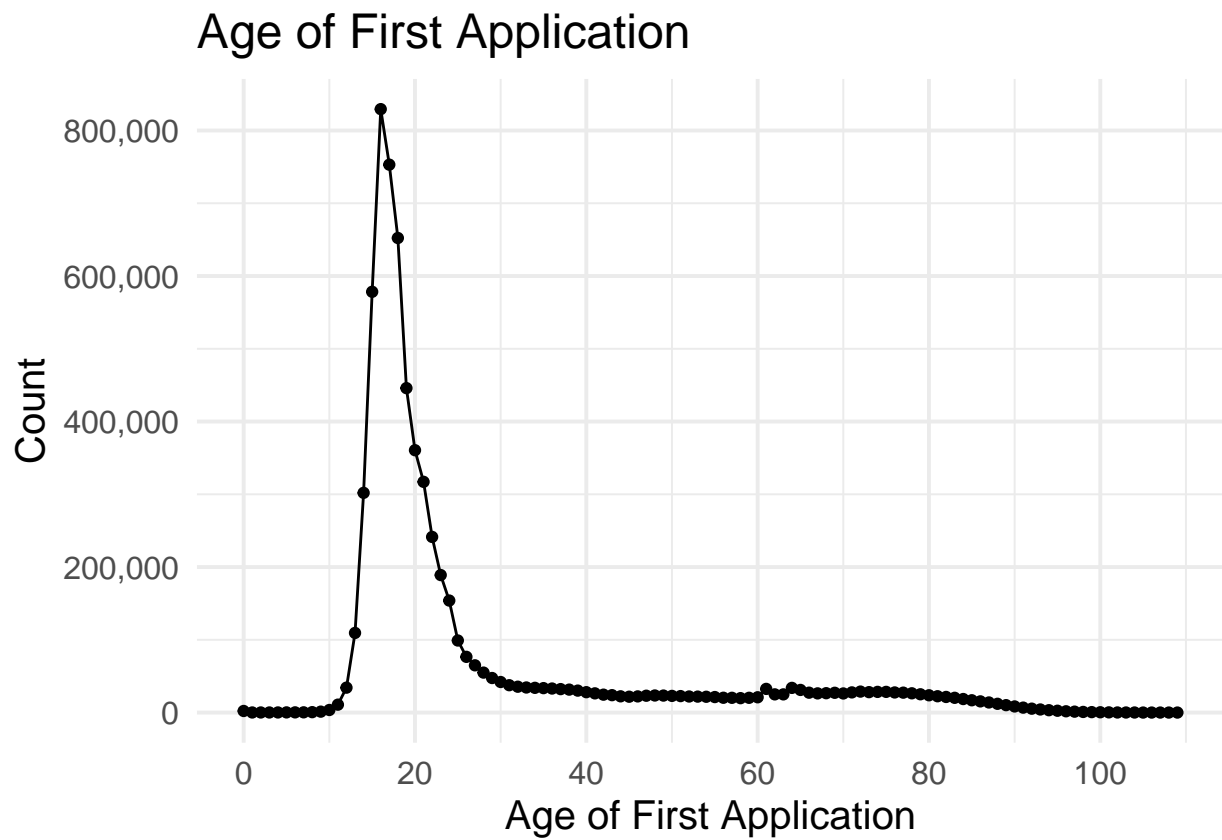
Tabulation of socstate (Native born only)

socstate	socstate_string	n	freq %	socstate	socstate_string	n	freq %
100	Alabama	115333	1.64	3000	Montana	30943	0.44
200	Alaska	1371	0.02	3100	Nebraska	83415	1.18
400	Arizona	32307	0.46	3200	Nevada	7521	0.11
500	Arkansas	79903	1.13	3300	New Hampshire	31294	0.44
600	California	409246	5.81	3400	New Jersey	220163	3.13
800	Colorado	75304	1.07	3500	New Mexico	24451	0.35
900	Connecticut	109931	1.56	3600	New York	631551	8.97
1000	Delaware	17558	0.25	3700	North Carolina	154754	2.20
1100	District of Columbia	39417	0.56	3800	North Dakota	37912	0.54
1200	Florida	108308	1.54	3900	Ohio	408497	5.80
1300	Georgia	124613	1.77	4000	Oklahoma	110731	1.57
1500	Hawaii	1198	0.02	4100	Oregon	77411	1.10
1600	Idaho	33832	0.48	4200	Pennsylvania	481833	6.84
1700	Illinois	427072	6.06	4400	Rhode Island	47675	0.68
1800	Indiana	212108	3.01	4500	South Carolina	71016	1.01
1900	Iowa	148846	2.11	4600	South Dakota	40178	0.57
2000	Kansas	106580	1.51	4700	Tennessee	137440	1.95
2100	Kentucky	125381	1.78	4800	Texas	317059	4.50
2200	Louisiana	120184	1.71	4900	Utah	39319	0.56
2300	Maine	58020	0.82	5000	Vermont	23914	0.34
2400	Maryland	106476	1.51	5100	Virginia	131299	1.86
2500	Massachusetts	227258	3.23	5300	Washington	113042	1.60
2600	Michigan	316753	4.50	5400	West Virginia	104225	1.48
2700	Minnesota	168296	2.39	5500	Wisconsin	202750	2.88
2800	Mississippi	73891	1.05	5600	Wyoming	16978	0.24
2900	Missouri	213664	3.03	NA		45370	0.64

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: Sample Weight

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
1.00	Min Weight
22.14	Max Weight
NA	No Weight Assigned

