

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)

Summary: The CenSoc-Numident file ($N = 7,890,991$) links the full-count 1940 Census to the National Archives' public release of the Social Security Numident file. Records were linked using the standard and conservative variants 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: 31 October, 2020

HISTID

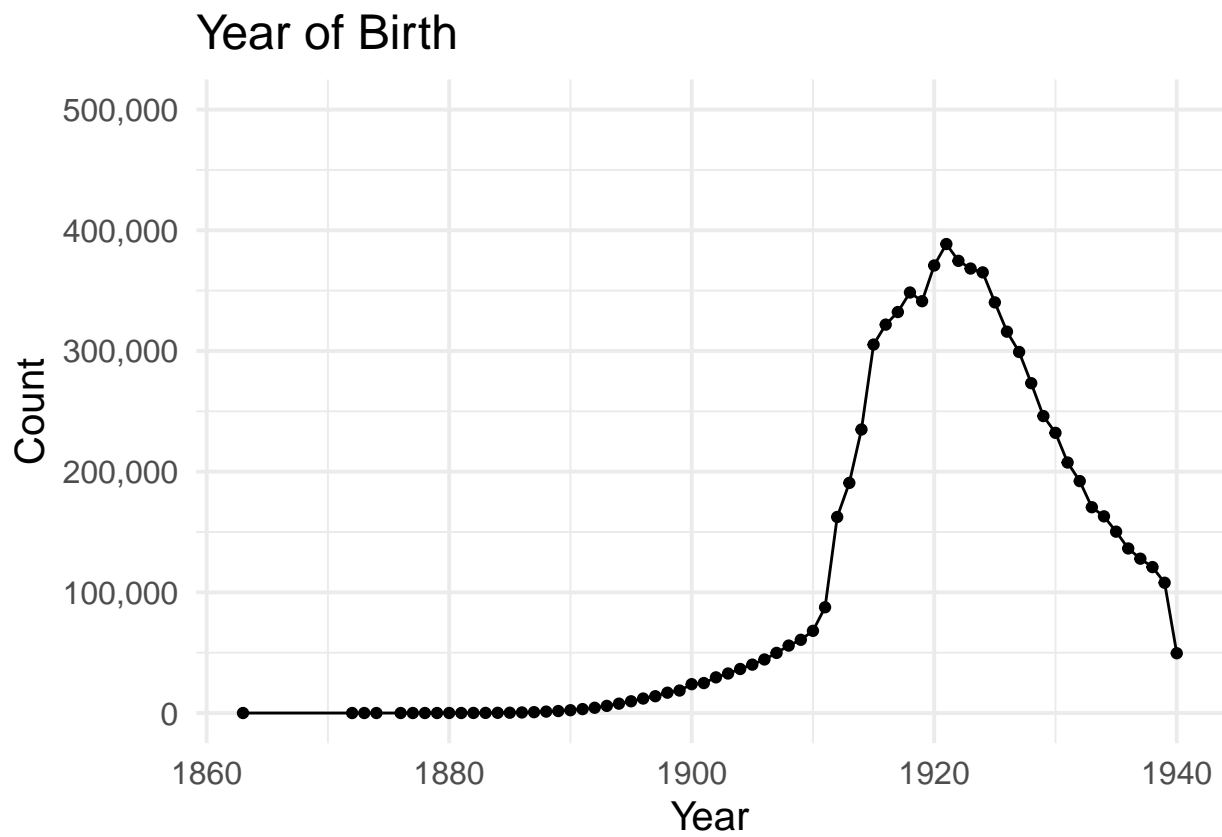
Label: Historical Unique Identifier

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

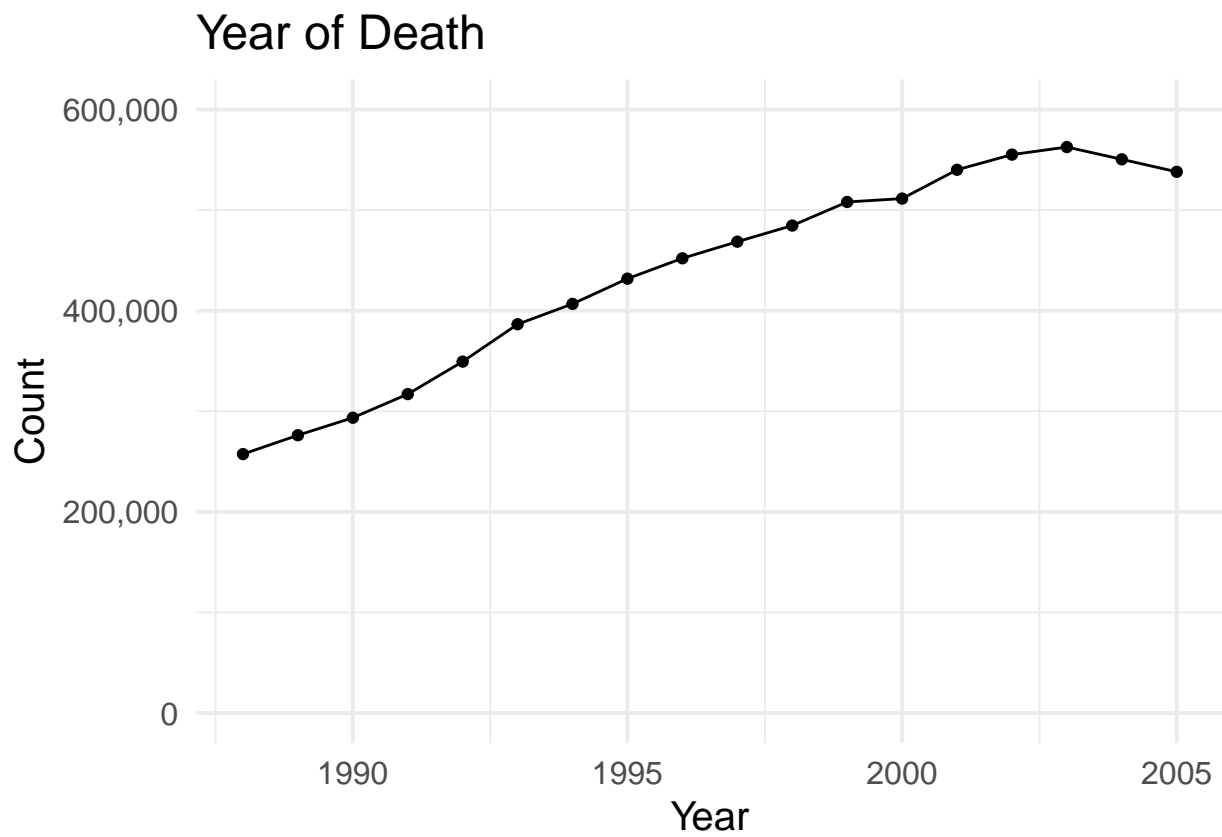
Description: bmonth reports a person's month of birth, as recorded in the Numident death records.

bmonth	label	n	freq %
1	January	677972	8.6
2	February	645279	8.2
3	March	707944	9.0
4	April	646387	8.2
5	May	648708	8.2
6	June	629288	8.0
7	July	669708	8.5
8	August	690057	8.7
9	September	672225	8.5
10	October	656419	8.3
11	November	613103	7.8
12	December	633901	8.0

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	739859	9.4
2	February	657462	8.3
3	March	699869	8.9
4	April	647872	8.2
5	May	646331	8.2
6	June	611979	7.8
7	July	622736	7.9
8	August	620045	7.9
9	September	608471	7.7
10	October	655634	8.3
11	November	651642	8.3
12	December	729091	9.2

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	4453736	56.4
2	Women	3437255	43.6

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	6984050	88.5000
2	Black	627441	7.9500
3	Other	35528	0.4500
4	Asian	7488	0.0949
5	Hispanic	9311	0.1180
6	North American Native	1902	0.0241
NA	Missing	225271	2.8500

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	6955193	88.100
2	Black	626025	7.930
3	Other	25745	0.326
4	Asian	13500	0.171
5	Hispanic	35200	0.446
6	North American Native	10057	0.127
NA	Missing	225271	2.850

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	181784	2.35	2900	Missouri	244823	3.16
200	Alaska	600	0.01	3000	Montana	37162	0.48
400	Arizona	24169	0.31	3100	Nebraska	104282	1.35
500	Arkansas	132435	1.71	3200	Nevada	4511	0.06
600	California	220455	2.85	3300	New Hampshire	30388	0.39
800	Colorado	71230	0.92	3400	New Jersey	216436	2.79
900	Connecticut	102348	1.32	3500	New Mexico	27717	0.36
1000	Delaware	15986	0.21	3600	New York	665375	8.59
1100	District of Columbia	25124	0.32	3700	North Carolina	226844	2.93
1200	Florida	78681	1.02	3800	North Dakota	58290	0.75
1300	Georgia	194387	2.51	3900	Ohio	409592	5.29
1500	Hawaii	1434	0.02	4000	Oklahoma	167351	2.16
1600	Idaho	34457	0.44	4100	Oregon	52723	0.68
1700	Illinois	442254	5.71	4200	Pennsylvania	611169	7.89
1800	Indiana	217711	2.81	4400	Rhode Island	46218	0.60
1900	Iowa	178912	2.31	4500	South Carolina	118064	1.52
2000	Kansas	132351	1.71	4600	South Dakota	53520	0.69
2100	Kentucky	202361	2.61	4700	Tennessee	183573	2.37
2200	Louisiana	144416	1.86	4800	Texas	367266	4.74
2300	Maine	63299	0.82	4900	Utah	42531	0.55
2400	Maryland	103253	1.33	5000	Vermont	26588	0.34
2500	Massachusetts	263117	3.40	5100	Virginia	175809	2.27
2600	Michigan	288867	3.73	5300	Washington	82924	1.07
2700	Minnesota	175295	2.26	5400	West Virginia	140749	1.82
2800	Mississippi	128895	1.66	5500	Wisconsin	210131	2.71
				5600	Wyoming	15950	0.21

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.

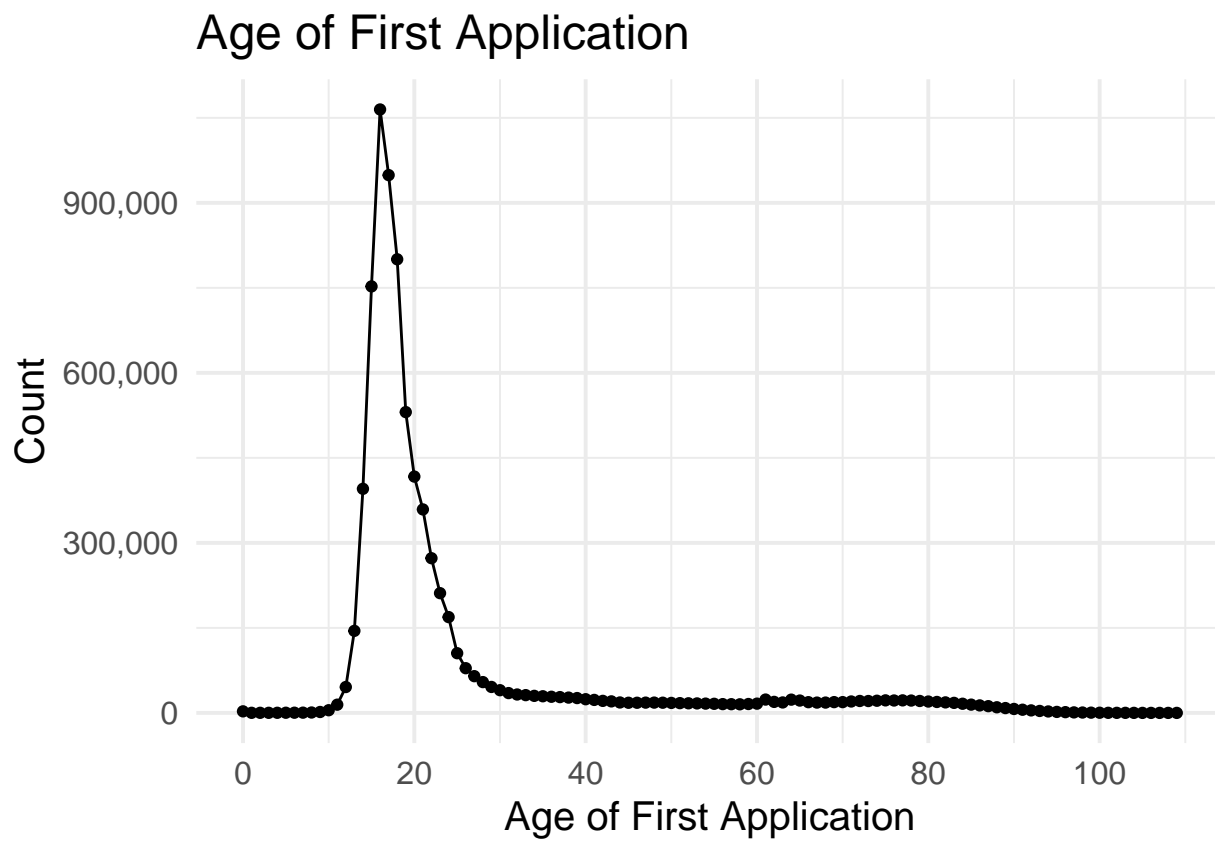
Tabulation of socstate (Native born only)

socstate	socstate_string	n	freq %	socstate	socstate_string	n	freq %
100	Alabama	148241	1.88	3000	Montana	30076	0.38
200	Alaska	1542	0.02	3100	Nebraska	81110	1.03
400	Arizona	33657	0.43	3200	Nevada	7464	0.09
500	Arkansas	93878	1.19	3300	New Hampshire	31507	0.40
600	California	431397	5.47	3400	New Jersey	240486	3.05
800	Colorado	75795	0.96	3500	New Mexico	27880	0.35
900	Connecticut	112534	1.43	3600	New York	731653	9.27
1000	Delaware	18203	0.23	3700	North Carolina	208168	2.64
1100	District of Columbia	45537	0.58	3800	North Dakota	36300	0.46
1200	Florida	122831	1.56	3900	Ohio	453556	5.75
1300	Georgia	167803	2.13	4000	Oklahoma	124243	1.57
1500	Hawaii	1557	0.02	4100	Oregon	75427	0.96
1600	Idaho	32568	0.41	4200	Pennsylvania	566672	7.18
1700	Illinois	464772	5.89	4400	Rhode Island	48640	0.62
1800	Indiana	232466	2.95	4500	South Carolina	95701	1.21
1900	Iowa	152321	1.93	4600	South Dakota	37532	0.48
2000	Kansas	109715	1.39	4700	Tennessee	173503	2.20
2100	Kentucky	154323	1.96	4800	Texas	380334	4.82
2200	Louisiana	135517	1.72	4900	Utah	38664	0.49
2300	Maine	60105	0.76	5000	Vermont	23216	0.29
2400	Maryland	118066	1.50	5100	Virginia	166310	2.11
2500	Massachusetts	262405	3.33	5300	Washington	112381	1.42
2600	Michigan	335746	4.26	5400	West Virginia	121904	1.55
2700	Minnesota	165617	2.10	5500	Wisconsin	200381	2.54
2800	Mississippi	94928	1.20	5600	Wyoming	16408	0.21
2900	Missouri	231048	2.93	NA	NA	56800	0.72

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.



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.

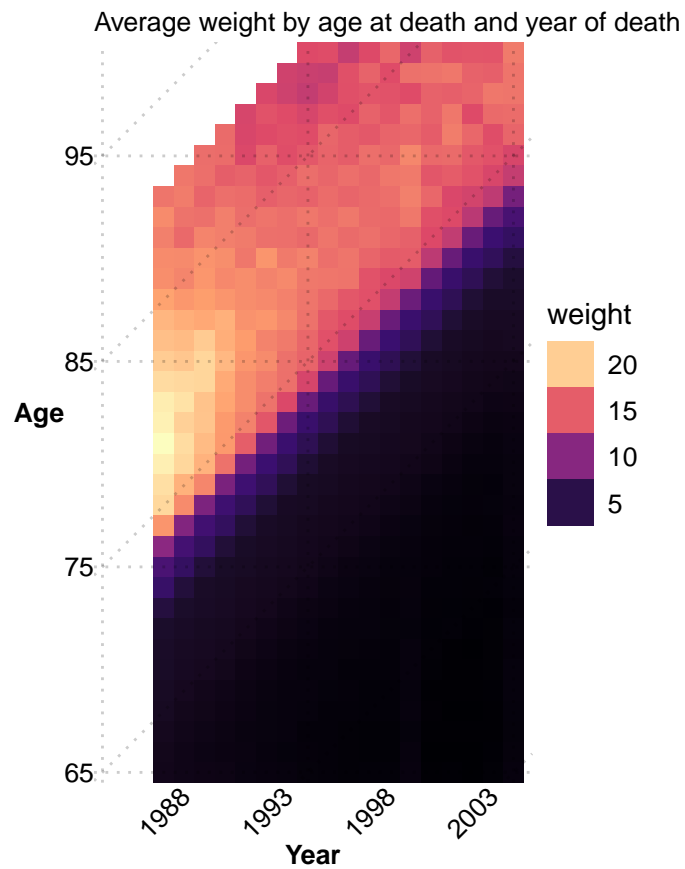
link__abe__exact__conservative	label	n	freq %
1	Match Established with Conservative ABE Algorithm	5797286	73.5
0	Not Established	2093705	26.5

weight

Label: Sample Weight

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
2.40	Min Weight
25.06	Max Weight
NA	No Weight Assigned



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
3.26	Min Weight
36.96	Max Weight
NA	No Weight Assigned

