Berkeley Unified Numident Mortality Database Codebook*

Page	Variable	Label
2	ssn	Social Security Number
3	fname	First Name
4	mname	Middle Name
5	lname	Last Name
6	fname_raw	First Name Raw
7	mname_raw	Middle Name Raw
8	lname_raw	Last Name Raw
9	sex	Sex
10	race_first	Race on First Application
11	race_first_cyear	First Race: Application Year
12	race_first_cmonth	First Race: Application Month
13	race_last	Race on Last Application
14	race_last_cyear	Last Race: Application Year
15	race_last_cmonth	Last Race: Application Month
16	race_change	Change in Race Response Flag
17	bpl	Place of Birth
18	byear	Year of Birth
19	bmonth	Month of Birth
20	bday	Day of Birth
21	dyear	Year of Death
22	dmonth	Month of Death
23	dday	Day of Death
24	death_age	Age at Death (Years)
25	zip_residence	ZIP Code of Residence at Time of Death
26	socstate	State where Social Security Number Issued
27	father_fname	Father's First Name
28	father_mname	Father's Middle Name
29	father_lname	Father's Last Name
30	father_fname_raw	Father's First Name Raw
31	father_mname_raw	Father's Middle Name Raw
32	father_lname_raw	Father's Last Name Raw
33	mother_fname	Mother's First Name
34	mother_mname	Mother's Middle Name
35	mother_lname	Mother's Last Name
36	mother_fname_raw	Mother's First Name Raw
37	mother_mname_raw	Mother's Middle Name Raw
38	mother_lname_raw	Mother's Last Name Raw
39	age_first_application	Age at First Social Security Application
40	number_apps	Number of Applications
41	number_claims	Number of Claims
42	weight	High-Coverage Sub-Sample Weight
43	ccweight	High-Coverage "Complete Case" Sub-Sample Weight

ssn

Label: Social Security Number

Description: ssn reports a person's Social Security number, as recorded in the Numident death records. Uniquely identifies all records in the dataset.

fname

Label: First Name

Description: fname reports the first 16 letters of the person's cleaned first name, as recorded in the Numident death records.

mname

Label: Middle Name

Description: mname reports the first 16 letters of the person's cleaned middle name, as recorded in the Numident death records.

lname

Label: Last Name

Description: lname reports the first 21 letters of the person's cleaned last name, as recorded in the Numident death records.

fname__raw

Label: First Name Raw

Description: fname_raw reports the first 16 letters of the person's raw first name, as recorded in the Numident death records.

mname_raw

Label: Middle Name Raw

Description: mname_raw reports the first 16 letters of the person's raw middle name, as recorded in the Numident death records.

lname_raw

Label: Last Name Raw

Description: lname_raw reports the first 21 letters of the person's raw last name, as recorded in the Numident death records.

sex

Label: Sex

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

sex	label	n	freq %
1	Men	25769568	52.20
2	Women	22686047	46.00
NA	Missing	882212	1.79

race_first

Label: race first

Description: race_first 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_first	label	n	freq %
1	White	28850917	58.476
2	Black	4066081	8.241
3	Other	689360	1.397
4	Asian, Asian American, or Pacific Islander	207656	0.421
5	Hispanic	286517	0.581
6	North American Indian or Alaskan Native	17485	0.035
NA	Missing	15219811	30.848

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	28417859	57.599
2	Black	4040665	8.190
3	Other	469636	0.952
4	Asian, Asian American, or Pacific Islander	305470	0.619
5	Hispanic	794762	1.611
6	North American Indian or Alaskan Native	89624	0.182
NA	Missing	15219811	30.848

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.

race_change

Label: Change in Race Response Flag

Description: race_change is a numeric dichotomous variable reporting whether an individual changed their race response across their Social Security applications.

race_change	label	n	freq %
0	No Change	33263321	67.40
1	Change	855047	1.73
NA	No Race Information	15219459	30.80

bpl

Label: Birthplace

Description: bpl is a numeric variable reporting the 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 numeric codes match the detailed IPUMS-USA Birthplace codes.

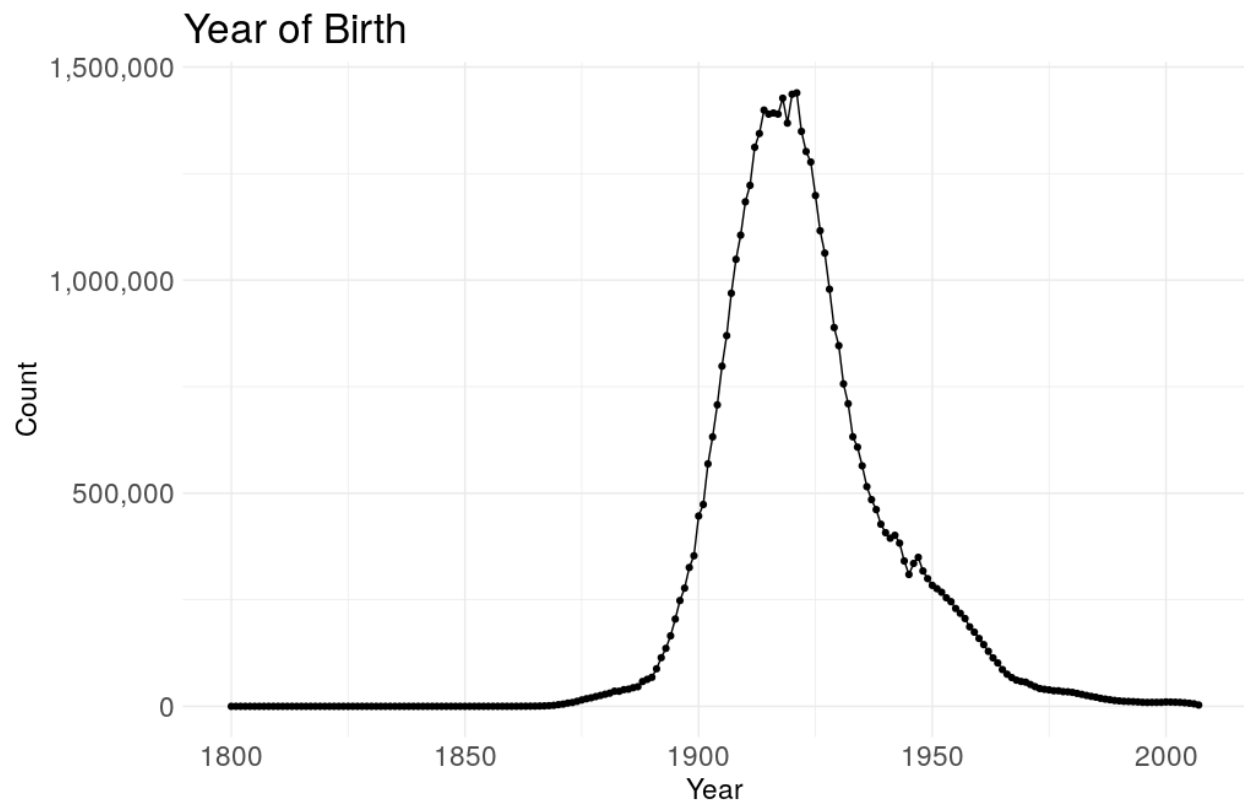For a complete list of IPUMS Birthplace codes, please see: https://usa.ipums.org/usa-action/variables/BPL

bpl	bpl_string	n	freq %	bpl	bpl_string	n	freq %
100	Alabama	916460	2.84	2900	Missouri	953475	2.96
200	Alaska	25406	0.08	3000	Montana	129791	0.40
400	Arizona	144300	0.45	3100	Nebraska	334691	1.04
500	Arkansas	624232	1.93	3200	Nevada	24990	0.08
600	California	1115176	3.46	3300	New Hampshire	102281	0.32
800	Colorado	277268	0.86	3400	New Jersey	845462	2.62
900	Connecticut	374288	1.16	3500	New Mexico	151357	0.47
1000	Delaware	64978	0.20	3600	New York	2646066	8.20
1100	District of Columbia	129166	0.40	3700	North Carolina	1100194	3.41
1200	Florida	462174	1.43	3800	North Dakota	190312	0.59
1300	Georgia	1011492	3.14	3900	Ohio	1547059	4.80
1500	Hawaii	91736	0.28	4000	Oklahoma	691459	2.14
1600	Idaho	125177	0.39	4100	Oregon	197939	0.61
1700	Illinois	1741873	5.40	4200	Pennsylvania	2480466	7.69
1800	Indiana	828095	2.57	4400	Rhode Island	160511	0.50
1900	Iowa	607860	1.88	4500	South Carolina	592424	1.84
2000	Kansas	476683	1.48	4600	South Dakota	168314	0.52
2100	Kentucky	866052	2.68	4700	Tennessee	868901	2.69
2200	Louisiana	718460	2.23	4800	Texas	1726094	5.35
2300	Maine	205832	0.64	4900	Utah	158295	0.49
2400	Maryland	436983	1.35	5000	Vermont	87739	0.27
2500	Massachusetts	970896	3.01	5100	Virginia	790692	2.45
2600	Michigan	1182488	3.67	5300	Washington	324126	1.00
2700	Minnesota	568850	1.76	5400	West Virginia	560213	1.74
2800	Mississippi	691061	2.14	5500	Wisconsin	713743	2.21
				5600	Wyoming	57795	0.18

Birthplace Frequencies (for Native Born)

byear

Label: Birth Year

Description: byear is a numeric variable reporting the person's year of birth, as recorded in the Numident death records.

Year of Birth

bmonth

Label: Birth Month

Description: bmonth is a numeric variable reporting the person's month of birth, as recorded in the Numident death records.

bmonth	label	n	freq %
1	January	4181375	8.47
2	February	3933555	7.97
3	March	4296940	8.71
4	April	3976307	8.06
5	May	4051432	8.21
6	June	3937147	7.98
7	July	4221224	8.56
8	August	4362872	8.84
9	September	4279353	8.67
10	October	4173868	8.46
11	November	3884566	7.87
12	December	4039185	8.19
NA	Missing	3	0.00

bday

Label: Birth Day

Description: bday is a numeric variable reporting the person's day of birth, as recorded in the Numident death records.
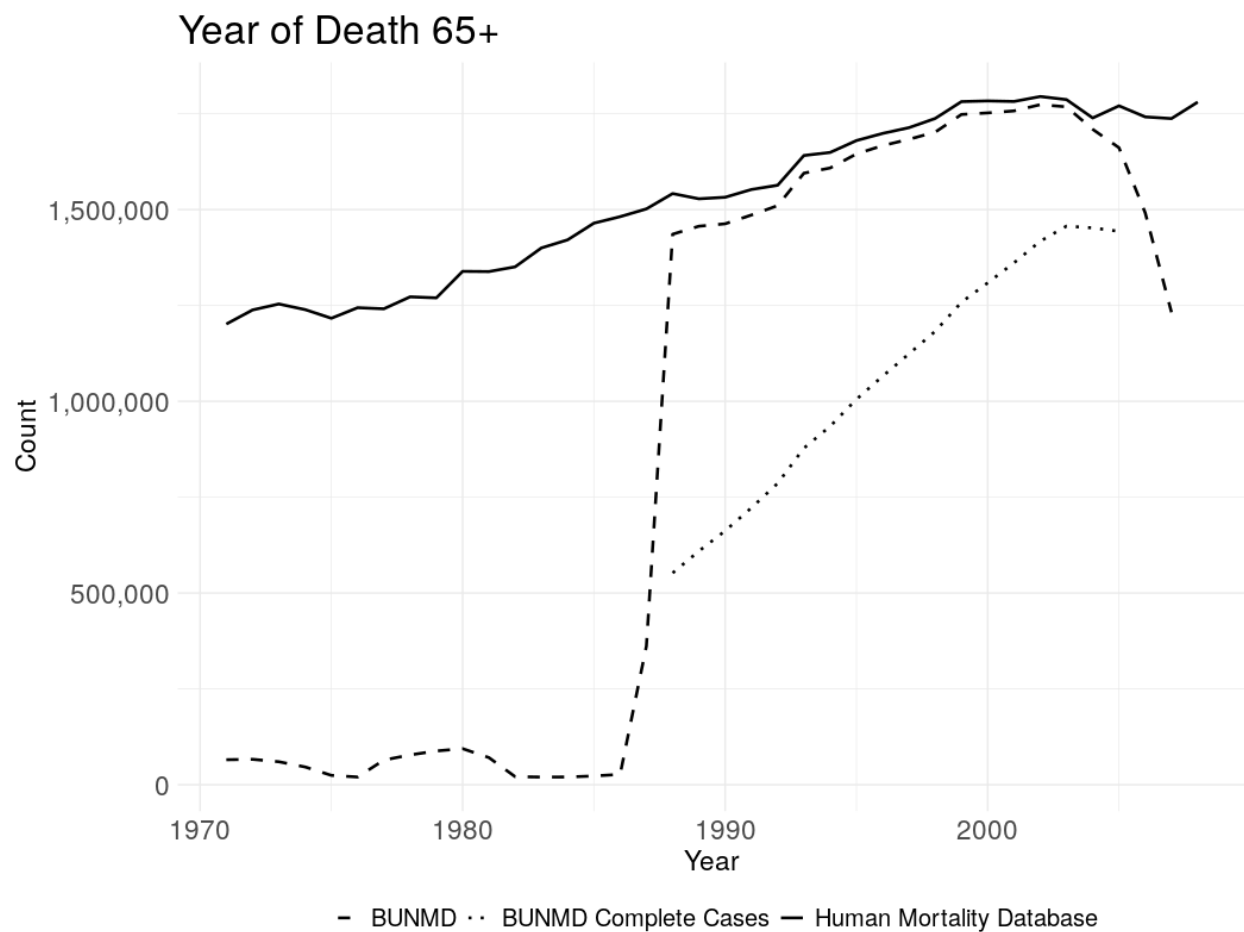
Note: It appears from the surplus of counts on the 1st and the 15th of the month that SSA might have done some imputation of missing values to these dates.

bday	n	freq %
1	1752583	3.600
2	1658244	3.400
3	1609261	3.300
4	1637955	3.300
5	1622040	3.300
6	1626699	3.300
7	1605110	3.300
8	1631598	3.300
9	1590390	3.200
10	1710297	3.500
11	1586088	3.200
12	1670822	3.400
13	1568796	3.200
14	1621064	3.300
15	1782188	3.600
16	1613051	3.300
17	1604846	3.300
18	1616778	3.300
19	1576661	3.200
20	1643349	3.300
21	1546797	3.100
22	1631722	3.300
23	1592459	3.200
24	1591364	3.200
25	1657548	3.400
26	1568447	3.200
27	1577419	3.200
28	1624799	3.300
29	1471946	3.000
30	1414317	2.900
31	901308	1.800
NA	31881	0.065

dyear

Label: Death Year

Description: dyear is a numeric variable reporting the person's year of death, as recorded in the Numident death records.

Year of Death 65+

Count

1,500,000

1,000,000

500,000

0

1970   1980   1990   2000

Year

– BUNMD ·· BUNMD Complete Cases — Human Mortality Database

dmonth

Label: Death Month

Description: dmonth is a numeric variable reporting the person's month of death, as recorded in the Numident death records.

dmonth	label	n	freq %
1	January	4610022	9.34
2	February	4119282	8.35
3	March	4353899	8.82
4	April	4036794	8.18
5	May	4023999	8.16
6	June	3828836	7.76
7	July	3924851	7.96
8	August	3887262	7.88
9	September	3835198	7.77
10	October	4125914	8.36
11	November	4102011	8.31
12	December	4489759	9.10

dday

Label: Day of Death

Description: dday is a numeric variable reporting the person's day of death, as recorded in the Numident death records.

Note: It appears from the surplus of counts on the 1st, 4th, and the 15th of the month that SSA might have done some imputation of missing values to these dates. Also, different periods have differ allocation patterns.

dday	n	freq %
1	1432114	2.9
2	1240840	2.5
3	1258505	2.6
4	1281969	2.6
5	1241857	2.5
6	1238121	2.5
7	1234527	2.5
8	1237353	2.5
9	1234152	2.5
10	1236852	2.5
11	1234080	2.5
12	1234552	2.5
13	1230163	2.5
14	1231013	2.5
15	3392922	6.9
16	1227374	2.5
17	1228022	2.5
18	1223495	2.5
19	1223837	2.5
20	1230608	2.5
21	1218728	2.5
22	1217328	2.5
23	1219353	2.5
24	1216433	2.5
25	1214084	2.5
26	1216347	2.5
27	1215884	2.5
28	1217904	2.5
29	1130910	2.3
30	1104857	2.2
31	713927	1.4
NA	9559716	19.0

death_age

Label: Age at Death (Years)

Description: death_age is a numeric variable reporting the person's age at death in years, calculated using birth and death information recorded in the Numident death records.

Age of Death

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 reported in the Numident death records.

socstate

Label: State where Social Security Number Issued

Description: The state in which the 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 numeric codes match the detailed IPUMS-USA Birthplace codes.

socstate	socstate_string	n	freq %	socstate	socstate_string	n	freq %
100	Alabama	1010756	2.07	3000	Montana	163411	0.33
200	Alaska	35207	0.07	3100	Nebraska	390427	0.80
400	Arizona	274363	0.56	3200	Nevada	59479	0.12
500	Arkansas	621147	1.27	3300	New Hampshire	154388	0.32
600	California	3381722	6.91	3400	New Jersey	1510334	3.09
800	Colorado	435935	0.89	3500	New Mexico	207319	0.42
900	Connecticut	644841	1.32	3600	New York	4980871	10.18
1000	Delaware	103907	0.21	3700	North Carolina	1349016	2.76
1100	District of Columbia	298017	0.61	3800	North Dakota	181420	0.37
1200	Florida	1119359	2.29	3900	Ohio	2460960	5.03
1300	Georgia	1171471	2.39	4000	Oklahoma	779894	1.59
1500	Hawaii	165054	0.34	4100	Oregon	423383	0.87
1600	Idaho	179125	0.37	4200	Pennsylvania	3266642	6.68
1700	Illinois	2855031	5.84	4400	Rhode Island	252536	0.52
1800	Indiana	1248048	2.55	4500	South Carolina	637802	1.30
1900	Iowa	764643	1.56	4600	South Dakota	174252	0.36
2000	Kansas	579408	1.18	4700	Tennessee	1115600	2.28
2100	Kentucky	917858	1.88	4800	Texas	2562712	5.24
2200	Louisiana	899017	1.84	4900	Utah	208380	0.43
2300	Maine	283261	0.58	5000	Vermont	110991	0.23
2400	Maryland	728497	1.49	5100	Virginia	1014527	2.07
2500	Massachusetts	1465566	3.00	5300	Washington	675565	1.38
2600	Michigan	2034399	4.16	5400	West Virginia	646790	1.32
2700	Minnesota	817858	1.67	5500	Wisconsin	1021002	2.09
2800	Mississippi	687994	1.41	5600	Wyoming	86773	0.18
2900	Missouri	1299368	2.66	NA		471923	0.96

Socstate Frequencies

father_fname

Label: Father's First Name

Description: father_fname is a character variable reporting the first 16 letters of the person's father's cleaned first name, as recorded in the Numident application records.

father_mname

Label: Father's Middle Name

Description: father_mname is a character variable reporting the first 16 letters of the person's father's cleaned middle name, as recorded in the Numident application records.

father_lname

Label: Father's Last Name

Description: father_lname is a character variable reporting the first 21 letters of the person's father's cleaned last name, as recorded in the Numident application records.

father_fname_raw

Label: Father's First Name Raw

Description: father_fname_raw is a character variable reporting the first 16 letters of the person's father's raw first name, as recorded in the Numident application records.

father_mname_raw

Label: Father's Middle Name Raw

Description: father_mname_raw is a character variable reporting the first 16 letters of the person's raw father's middle name, as recorded in the Numident application records.

father_lname_raw

Label: Father's Last Name Raw

Description: father_lname_raw is a character variable reporting the first 21 letters of the person's father's raw last name, as recorded in the Numident application records.

mother_fname

Label: Mother's First Name

Description: mother_fname is a character variable reporting the first 16 letters of the person's mother's cleaned first name, as recorded in the Numident application records.

mother_mname

Label: Mother's Middle Name

Description: mother_mname is a character variable reporting the first 16 letters of the person's mother's cleaned middle name, as recorded in the Numident application records.

mother_lname

Label: Mother's Last Name

Description: mother_lname is a character variable reporting the first 21 letters of the person's mother's cleaned last name, as recorded in the Numident application records.
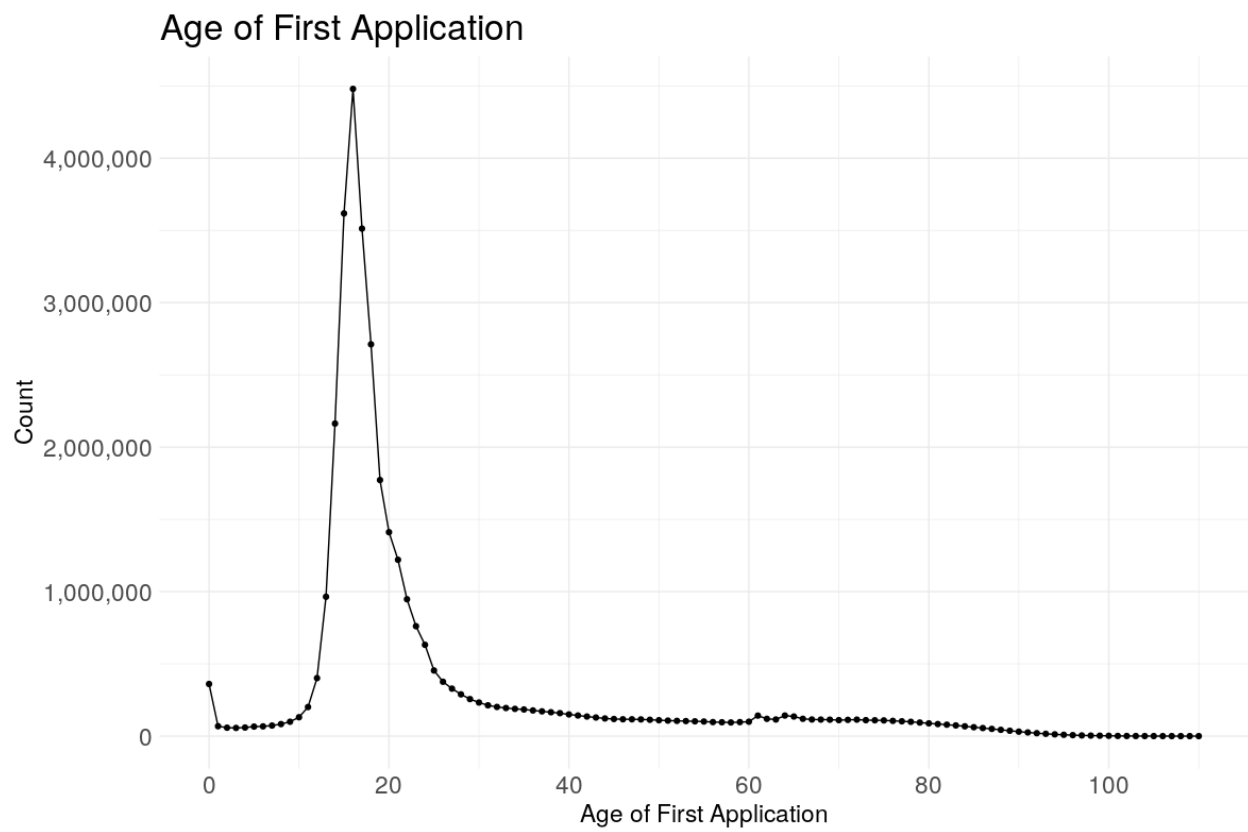
mother_fname_raw

Label: Mother's First Name Raw

Description: mother_fname is a character variable reporting the first 16 letters of the person's mother's rawfirst name, as recorded in the Numident application records.

mother_mname_raw

Label: Mother's Middle Name Raw

Description: mother_mname is a character variable reporting the first 16 letters of the person's mother's raw middle name, as recorded in the Numident application records.

mother_lname_raw

Label: Mother's Last Name Raw

Description: mother_lname is a character variable reporting the first 21 letters of the person's mother's raw last name, as recorded in the Numident application records.

age_first_application

Label: Age at First Social Security Application

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

Age of First Application

number_apps

Label: Total Number of Applications

Description: number_apps reports the person's total number of Social Security applications, as recorded in the Numident application records. This variable was top-coded at 10.

number_apps	n	freq %
0	14166703	29.000
1	17646617	36.000
2	10107569	20.000
3	4309930	8.700
4	1710582	3.500
5	719300	1.500
6	323112	0.650
7	155220	0.310
8	80363	0.160
9	43690	0.089
10+	74741	0.150

number_claims

Label: Total Number of Claims

Description: number_claims reports the person's total number of Social Security claims, as recorded in the Numident claim records. This variable was top-coded at 3.
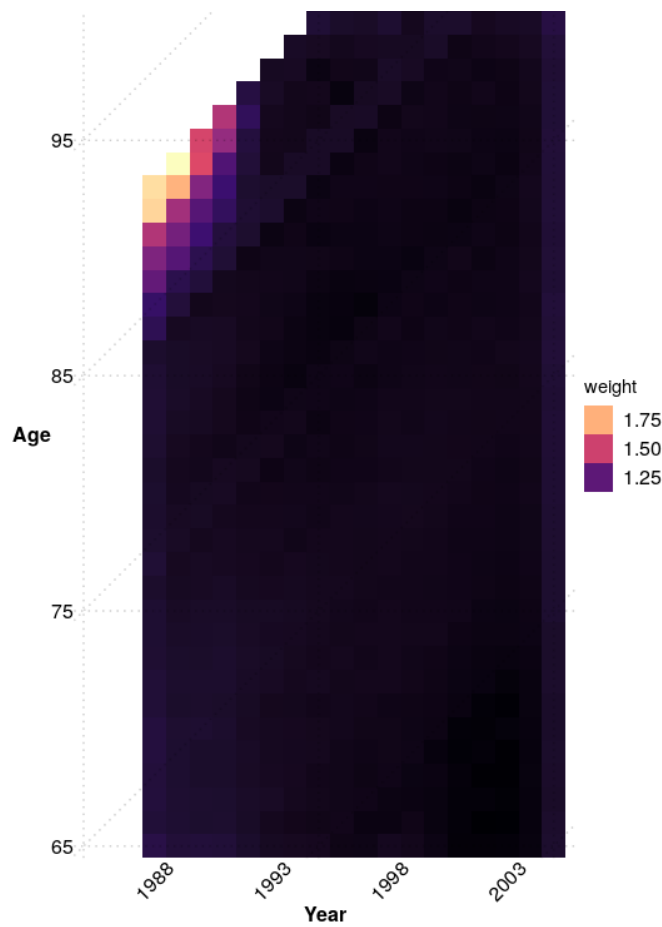
number_claims	n	freq %
0	32931311	66.75
1	16360228	33.16
2	45187	0.09
3+	1101	0.00

weight

Label: High Coverage Sample Weight

Description: A post-stratification person-weight to Human Mortality Database (HMD) totals for persons (1) born between 1895-1940 (2) dying between 1988-2005 and (3) dying between ages 65-100 (4) with a non-missing value for sex. Please see the CenSoc methods protocol for more details on weighting procedure.

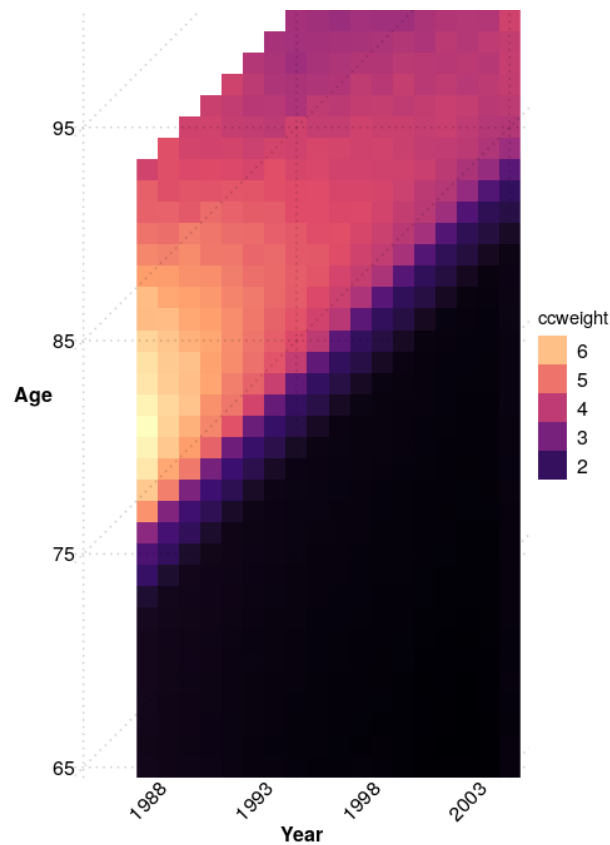| Value	Label
NA	No weight assigned
Min Weight	0.961
Max Weight	2.383

Average weight by age of death and year of death.

ccweight

Label: High Coverage "Complete Case" Sample Weight

Description: A post-stratification person-weight to Human Mortality Database (HMD) totals for persons (1) born between 1895-1940 (2) dying between 1988-2005 (3) dying between ages 65-100 (4) with a non-missing value for sex (5) with a non-missing value for birth place and (6) with a non-missing value for race. Please see the CenSoc methods protocol for more details on weighting procedure.

Value	Label
NA	Incomplete case
Min Weight	1.02
Max Weight	18.71

Average ccweight by age of death and year of death.