

## CenSoc Enlistment-Numident V2 Codebook\*

Page	Variable	Label
2	id	Unique Identifier
3	sex	Sex
4	bpl	Place of Birth
5	byear_numident	Year of Birth (Numident)
6	bmonth_numident	Month of Birth
7	dyear_numident	Year of Birth
8	dmonth_numident	Month of Birth
9	death_age_numident	Age at Death (Years)
10	race_first_numident	Race on First Application
11	race_first_cyear_numident	First Race: Application Year
12	race_first_cmonth_numident	First Race: Application Month
13	race_last_numident	Race on Last Application
14	race_last_cyear_numident	Last Race: Application Year
15	race_last_cmonth_numident	Last Race: Application Month
16	zip_residence_numident	ZIP Code of Residence at Time of Death
17	socstate_numident	State where Social Security Number was Issued
18	age_first_application_numident	Age at First Social Security Application
19	byear_enlistment	Year of Birth (Enlistment)
20	date_of_enlistment_enlistment	Date of Enlistment
21	residence_state_enlistment	State of Residence at Enlistment
22	residence_county_enlistment	County of Residence at Enlistment
23	place_of_enlistment_enlistment	Place of Enlistment
24	education_enlistment	Education
25	grade_code_enlistment	Army Grade (Rank)
26	branch_code_enlistment	Army Branch
28	term_of_enlistment_enlistment	Term of Enlistment
29	race_enlistment	Race
30	citizenship_enlistment	Citizenship
31	civilian_occupation_enlistment	Civilian Occupation
32	marital_status_enlistment	Marital Status at Enlistment
33	height_enlistment	Height at Enlistment (Inches)
34	weight_before_march_1943_enlistment	Weight (Pounds)
35	weight_or_AGCT_enlistment	Weight (Pounds) or AGCT Score
36	component_enlistment	Army Component
37	source_enlistment	Source of Army Personnel
38	HISTID	HISTID (1940 Census)

**Summary:** The CenSoc Enlistment-Numident file links 1,692,279 men from World War II era army enlistment records (circa 1938-1947) to mortality records in the Berkeley Unified Numident Mortality Database (BUNMD). The suffix on each variable indicates whether the variable comes from Numident (BUNMD) records or from army enlistment records. For enlistment records that have been independently and additionally linked to the 1940 Census, we include the HISTID identifier variable that researchers may use to merge this dataset with IPUMS 1940 Census data. Matches between datasets were established using a conservative variant of the ABE method developed by Abramitzky, Boustan, and Eriksson (2012, 2014, 2017).

\*Last updated: 07 August, 2023

# id

**Label:** CenSoc Numident-Enlistment Unique Identifier

**Description:** id is a generated 11-digit string that uniquely identifies each record in the dataset.

# Sex

**Label:** Sex

**Description:** sex reports a person's sex.

sex	label	n	freq %
1	Men	1692279	100

# bpl

**Label:** Place of Birth

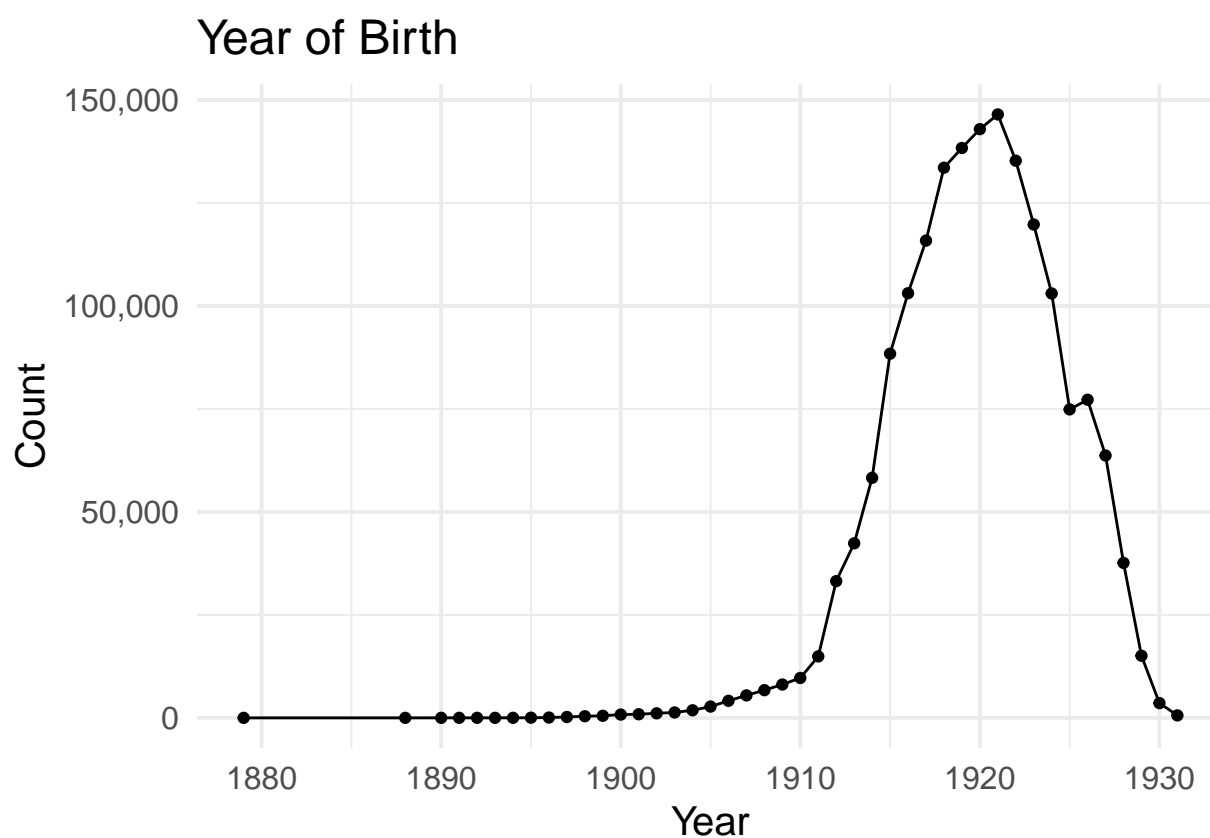
**Description:** bpl is a numeric variable that reports a person's place of birth, as reported in the World War II Army Enlistment Records and Numident records. The coding schema matches the **detailed** IPUMS-USA Birthplace coding schema. For a complete list of IPUMS codes, please see: [https://usa.ipums.org/usa-action/variables/BPL#codes\\_section](https://usa.ipums.org/usa-action/variables/BPL#codes_section).

# byear\_\_numident

**Label:** Year of Birth (Numident)

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

**Note:** The matching algorithm used allows for up to two years of flexibility on birth year, so birth year in the Numident may disagree with birth year in enlistment records. Birth years are discordant by one year in 7.2% of records, and discordant by two years in 4.1% of records. Both birth year variables are published in this dataset.



# bmonth\_numident

**Label:** Birth Month

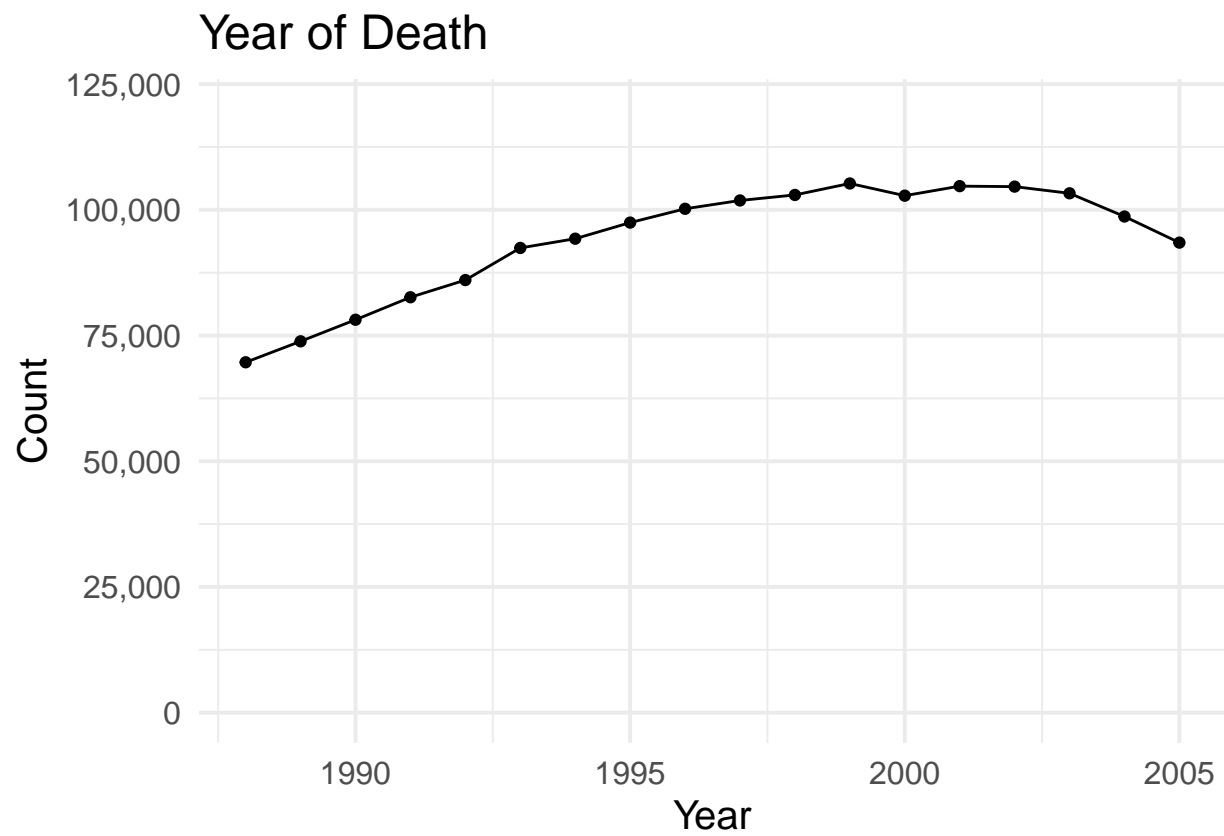
**Description:** bmonth\_numident reports a person's month of birth, as recorded in the Numident death records.

bmonth_numident	label	n	freq %
1	January	143035	8.45
2	February	134602	7.95
3	March	147238	8.70
4	April	137407	8.12
5	May	139691	8.25
6	June	135317	8.00
7	July	142848	8.44
8	August	148529	8.78
9	September	146588	8.66
10	October	144203	8.52
11	November	134368	7.94
12	December	138453	8.18

# dyear\_numident

**Label:** Death Year

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



# dmonth\_\_numident

**Label:** Death Month

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

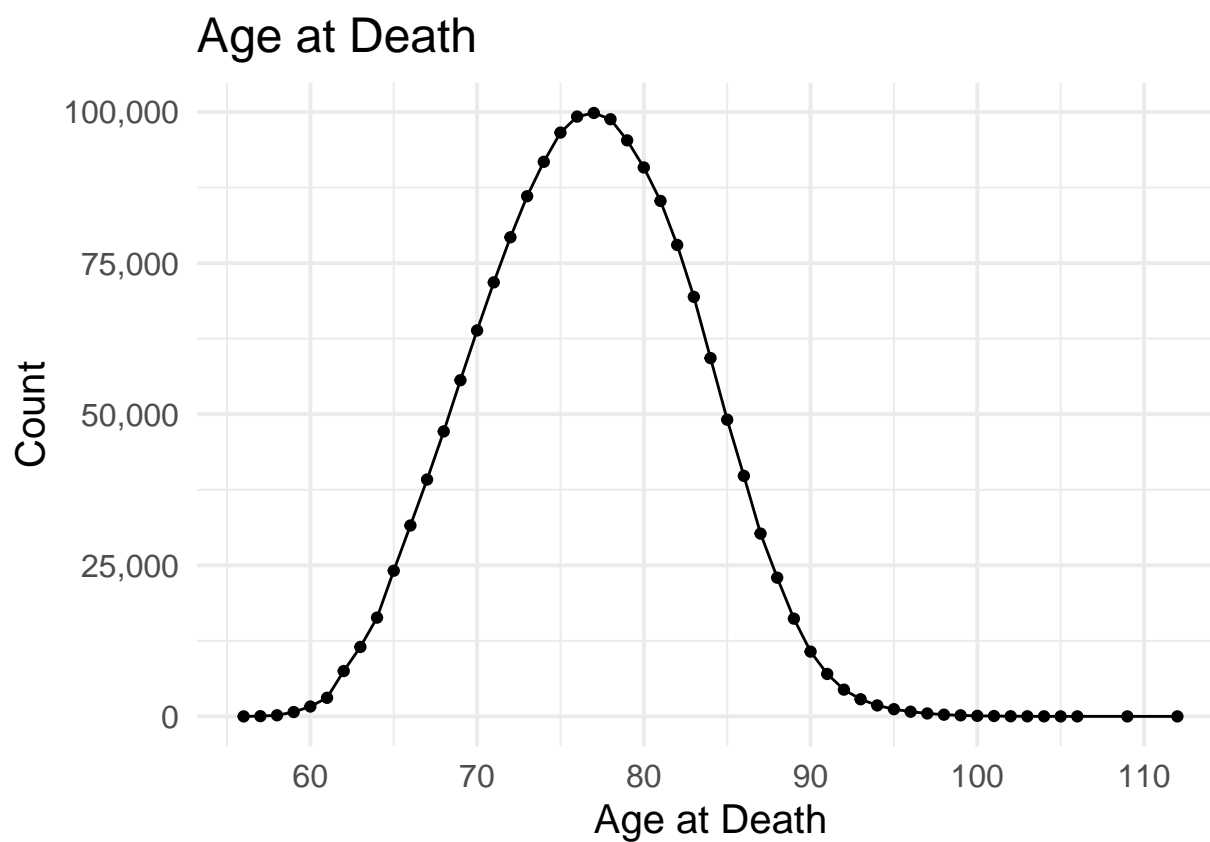
dmonth__numident	label	n	freq %
1	January	160178	9.47
2	February	141933	8.39
3	March	152020	8.98
4	April	139903	8.27
5	May	138976	8.21
6	June	131092	7.75
7	July	132675	7.84
8	August	131790	7.79
9	September	129298	7.64
10	October	140025	8.27
11	November	139584	8.25
12	December	154805	9.15



# death\_age\_numident

**Label:** Age at Death (Years)

**Description:** death\_age\_numident reports a person's age at death in years, calculated using the birth and death information recorded in the Numident death records.



# race\_\_first\_\_numident

**Label:** Race First

**Description:** race\_\_first\_\_numident 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__numident	label	n	freq %
1	White	1519903	89.81
2	Black	118138	6.98
3	Other	15072	0.89
4	Asian	2305	0.14
5	Hispanic	1750	0.10
6	North American Native	395	0.02
NA	Missing	34716	2.05

# race\_\_first\_\_cyear\_\_numident

**Label:** First Race: Application Year

**Description:** race\_\_first\_\_cyear\_\_numident is a numeric variable reporting the year of the application on which a person reported their first race.

# race\_\_first\_\_cmonth\_\_numident

**Label:** First Race: Application Month

**Description:** race\_\_first\_\_cmonth\_\_numident is a numeric variable reporting the month of the application on which a person reported their first race.

# race\_\_last\_\_numident

**Label:** Race Last

**Description:** race\_\_last\_\_numident 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__numident	label	n	freq %
1	White	1514876	89.52
2	Black	117950	6.97
3	Other	12002	0.71
4	Asian	4436	0.26
5	Hispanic	6602	0.39
6	North American Native	1697	0.10
NA	Missing	34716	2.05

# race\_\_last\_\_cyear\_\_numident

**Label:** Last Race: Application Year

**Description:** race\_\_last\_\_cyear\_\_numident reports the year of the application on which a person reported their last race.

# race\_\_last\_\_cmonth\_\_numident

**Label:** Last Race: Application Month

**Description:** race\_\_last\_\_cmonth\_\_numident is a numeric variable reporting the month of the application on which a person reported their last race.

# zip\_residence\_numident

**Label:** ZIP Code of Residence at Time of Death

**Description:** zip\_residence\_numident 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\_\_numident

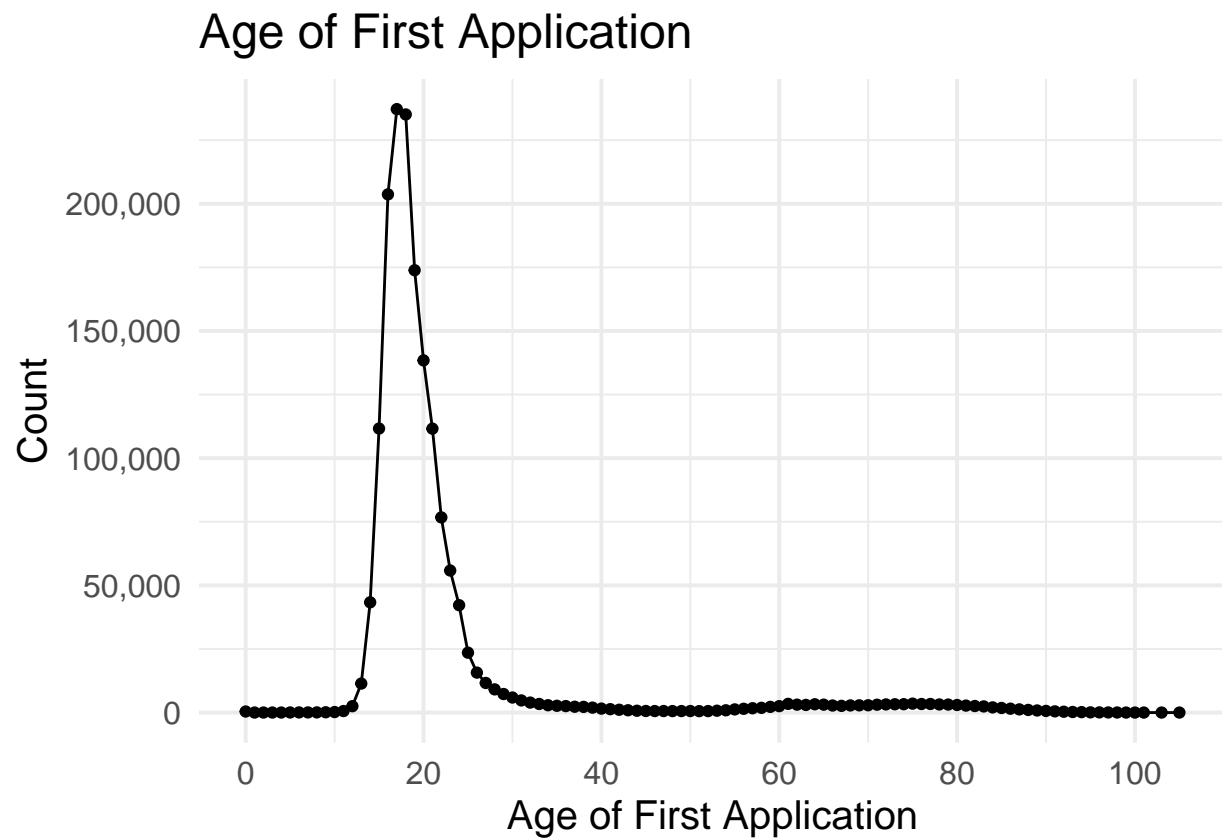
**Label:** State where Social Security Number Issued

**Description:** socstate\_\_numident is a numeric variable reporting the state in which a person's social security card was issued. It is determined by the first three (3) digits of the Social Security number, as recorded in Numident death records. The coding schema matches the **detailed** IPUMS-USA Birthplace coding schema. For a complete list of IPUMS codes, please see: [https://usa.ipums.org/usa-action/variables/BPL#codes\\_section](https://usa.ipums.org/usa-action/variables/BPL#codes_section).

# age\_first\_application\_numident

**Label:** Age at First Social Security Application

**Description:** age\_first\_application\_numident reports the age at which a person submitted their first Social Security Application.

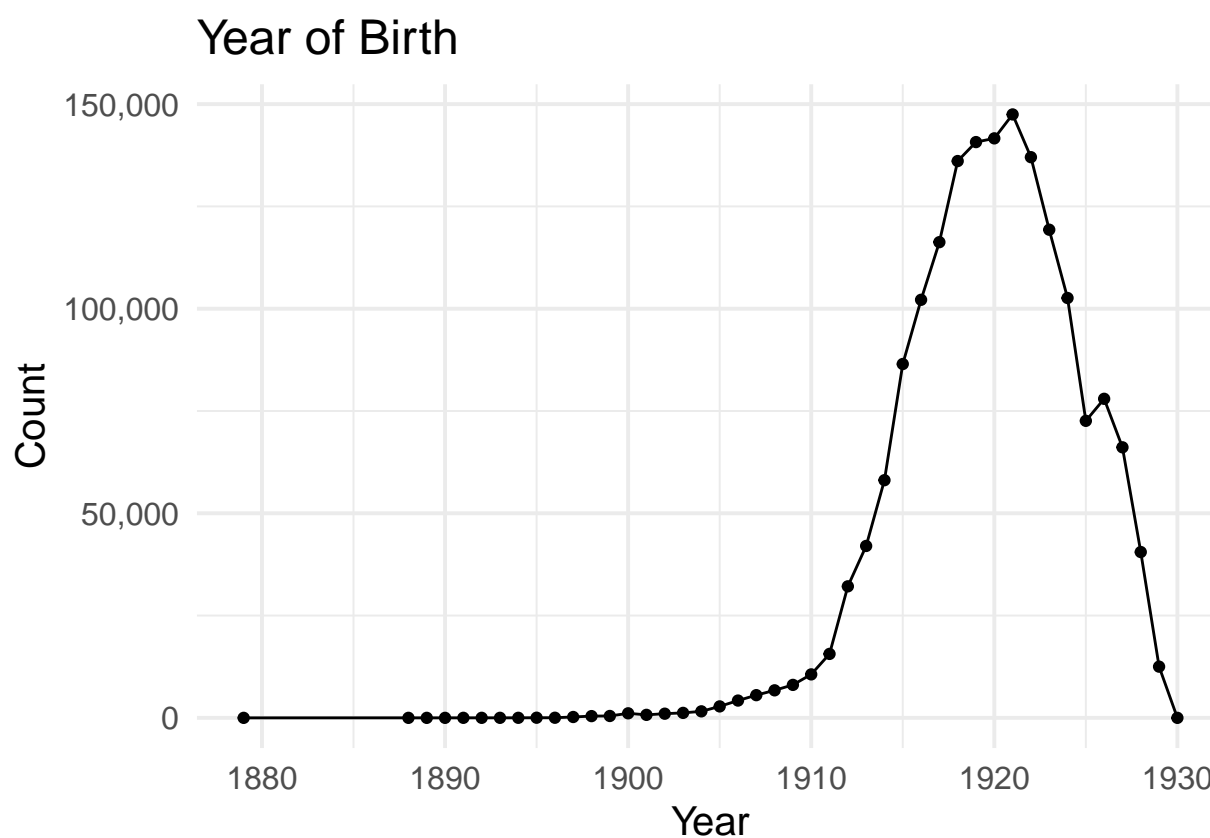


# byear\_\_enlistment

**Label:** Year of Birth (Enlistment)

**Description:** byear\_\_enlistment is a numeric variable reporting a person's year of birth, as recorded in the World War II Army Enlistment Records.

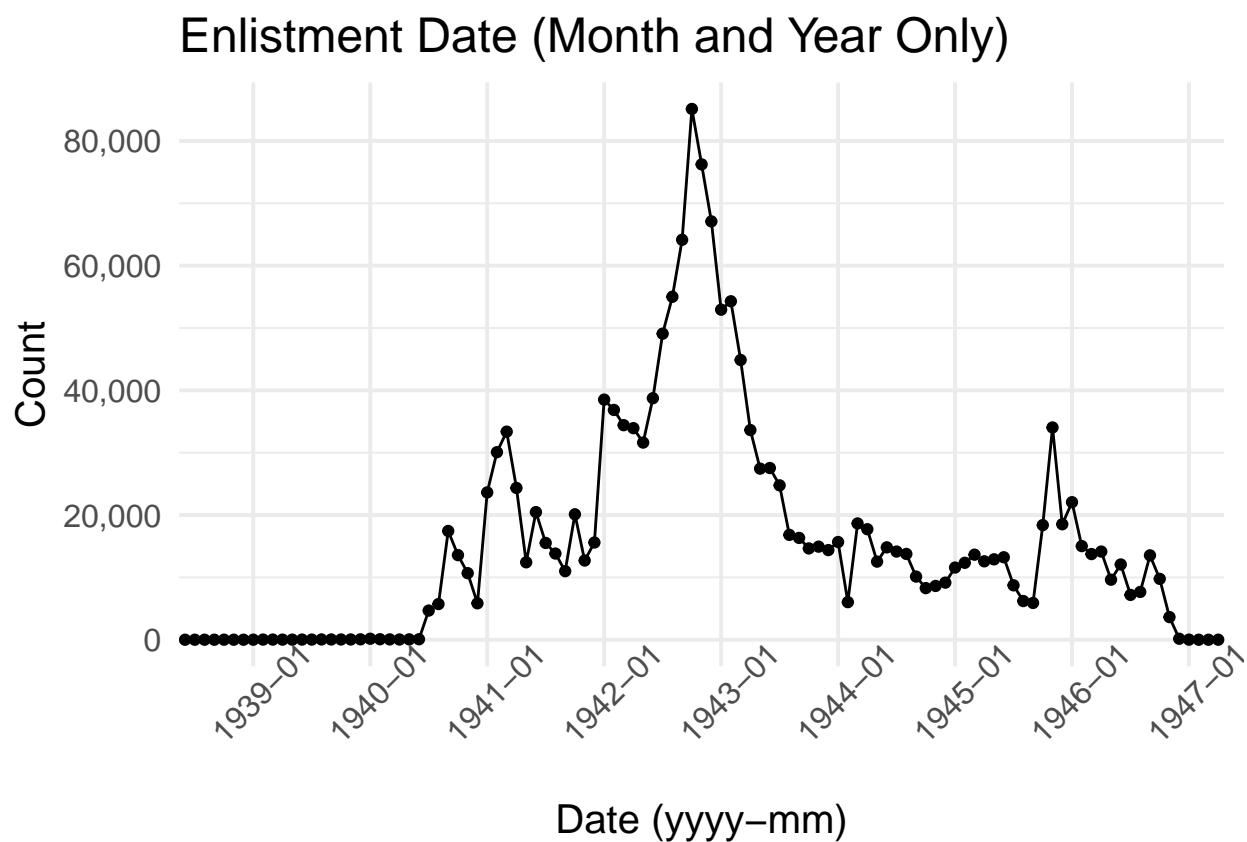
**Note:** The matching algorithm used allows for up to two years of flexibility on birth year, so birth year in the BUNMD may disagree with birth year in enlistment records. Birth years are discordant by one year in 7.2% of records, and discordant by two years in 4.1% of records. Both birth year variables are published in this dataset.



# date\_of\_enlistment\_enlistment

**Label:** Date of Enlistment

**Description:** date\_of\_enlistment\_enlistment is a date variable that reports a person's date of enlistment in dd-mm-yyyy format, as reported in the World War II Army Enlistment Records.



# residence\_\_state\_\_enlistment

**Label:** State of Residence

**Description:** residence\_\_state\_\_enlistment is a numeric variable that reports a person's state of residence at time of enlistment, as reported in the World War II Army Enlistment Records. This variable is only available for enlistees residing in the United States, Hawaii, Alaska, and Puerto Rico. The coding schema matches the **detailed** IPUMS-USA birthplace coding schema. For a complete list of IPUMS codes, please see: [https://usa.ipums.org/usa-action/variables/BPL#codes\\_section](https://usa.ipums.org/usa-action/variables/BPL#codes_section).

# residence\_\_county\_\_enlistment

**Label:** County of Residence

**Description:** residence\_\_county\_\_enlistment is a character variable that reports a person's county of residence at time of enlistment, as reported in the World War II Army Enlistment Records. These codes are generally equivalent to FIPS county codes. This variable is only available for enlistees residing in the United States. County is not available for residents of Hawaii and Alaska, which were not yet states. For a complete list of codes, please see: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=2057&dt=893&c\\_id=24998](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2057&dt=893&c_id=24998). Or, for county codes organized by state, see page 1-65 of [Nara Compiled Technical Documentation](#). (Note that state codes in NARA Compiled Technical Documentation are *not* equivalent to either state FIPS codes or IPUMS state codes published in this dataset.)

**Note:** County codes are only unique within states. For example, code 111 refers to Ventura County for residents of California, but refers to Jefferson County for residents of Kentucky. These codes may not align exactly with modern FIPS codes due to county splits, merges, name changes, etc. since the 1940s. County codes within individual states have not been cleaned, and invalid codes may be present.

# place\_\_of\_\_enlistment\_\_enlistment

**Label:** Place of Enlistment

**Description:** place\_of\_enlistment\_enlistment is a character variable that reports where a person was enlisted, as reported in the World War II Army Enlistment Records. Place of enlistment may record a state, town, military post, or other ad hoc station of enlistment. For example, places of enlistment in the state of Florida include Florida, Tallahassee, Orlando Air Base, and Municipal Fair Grounds Tampa.

For a complete list of codes, please see: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=4123&dt=893&c\\_id=24997](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=4123&dt=893&c_id=24997)

# education\_\_enlistment

**Label:** Education

**Description:** education\_\_enlistment is a numeric variable that reports an individual's educational attainment at time of enlistment, as reported in the World War II Army Enlistment Records. Codes are also available at: [https://usa.ipums.org/usa-action/variables/EDUC#codes\\_section](https://usa.ipums.org/usa-action/variables/EDUC#codes_section)

**Note:** Exact number of years of schooling is unknown for enlistees who did not complete at least one year of high school. Such enlistees were given originally given a code meaning “grammar school”, though it is unknown if this corresponds to the modern equivalent of a 5th, 6th, or 8th grade education. Some enlistees with this code may have had little to no actual education. Before the attack on Pearl Harbor, literacy was required of army inductees, but this requirement relaxed as the war effort increased (Brandt, 2004). By 1942, the Army accepted illiterate men if they were deemed trainable, though academic standards for admission increased again in 1944 and 1945.

education__enlistment	label	n	freq %
1	Nursery school to grade 4	437626	25.86
3	Grade 9	144744	8.55
4	Grade 10	181809	10.74
5	Grade 11	166523	9.84
6	Grade 12	507866	30.01
7	1 year of college	102978	6.09
8	2 years of college	58056	3.43
9	3 years of college	29552	1.75
10	4 years of college	56274	3.33
11	5+ years of college	6851	0.40



# grade\_\_code\_\_enlistment

**Label:** Army Grade

**Description:** grade\_\_code\_\_enlistment is a numeric variable that records a person's army grade (rank), as reported in the World War II Army Enlistment Records.

grade__code__enlistment	label	n	freq %
1	Master Sergeant	4340	0.26
2	First Sergeant	1538	0.09
3	Technical Sergeant	4475	0.26
4	Staff Sergeant	8795	0.52
5	Sergeant	12186	0.72
6	Corporal	16598	0.98
7	Private First Class	50189	2.97
8	Private	1571868	92.88
9	Aviation Cadet	8525	0.50
NA	NA	13765	0.81

# branch\_\_code\_\_enlistment

**Label:** Army Branch

**Description:** branch\_\_code\_\_enlistment is a numeric variable that records the army branch in which the enlistee served, as reported in the World War II Army Enlistment Records.

branch_code_enlistment	label	n	freq %
0	Branch Immaterial - Warrant Officers, USA	752288	44.454
1	General Staff Corps	50	0.003
2	No branch assignment	539070	31.855
3	Philippine Scouts	4945	0.292
4	Professor US Military Academy	183	0.011
5	Military Intelligence	310	0.018
6	Corps of Military Police	3453	0.204
7	Specialist Reserve	8	0.000
8	Detached Enlisted Men's List	2037	0.120
9	Inactive Reserve	602	0.036
10	Infantry	41949	2.479
20	Air Corps	205413	12.138
30	Field Artillery	21234	1.255
38	Transportation Corps	4032	0.238
40	Coast Artillery Corps or Army Mine Planter Service	15413	0.911
50	Quartermaster Corps	14762	0.872
70	Medical Department or Army Nurse Corps	10590	0.626
71	Medical Corps - For Officers only	9	0.001
72	Hospital Dietitians	197	0.012
73	Physical Therapy Aides	19	0.001
74	Dental Corps - For Officers only	33	0.002
75	Veterinary Corps or Service	37	0.002
77	Medical Administrative Corps - For Officers only	2	0.000
78	Adjutant General's Department	205	0.012
79	Pharmacy Corps	21	0.001
80	Chemical Warfare Service	579	0.034
82	Corps of Engineers	8847	0.523
87	Ordnance Department	10527	0.622
90	Signal Corps	19329	1.142
93	Cavalry	5646	0.334
96	Chaplains	35	0.002
97	Finance Department	495	0.029
98	Inspector General's Department	38	0.002
99	Judge Advocate General's Department	1476	0.087
NA	NA	28445	1.681

# term\_of\_enlistment\_enlistment

**Label:** Term of Enlistment

**Description:** term\_of\_enlistment\_enlistment is a numeric variable that reports the length of a person's enlistment or their department of enlistment, as reported in the World War II Army Enlistment Records. Codes are also available at: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=2076&dt=893&c\\_id=24979](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2076&dt=893&c_id=24979)

term_of_enlistment_enlistment	label (abbreviated)	n	freq %
1	One year enlistment	42182	2.49
2	Enlistment for Alaska	4481	0.26
3	Enlistment for Hawaiian Department	90019	5.32
4	Enlistment for assignment to another corps area	3471	0.21
5	Enlistment for the duration of the War or other emergency	1159505	68.52
6	Enlistment for the Panama Canal Department	64873	3.83
7	Enlistment for the Philippine Department	17595	1.04
8	Enlistment for the Puerto Rican Department	1248	0.07
9	Three year enlistment	15272	0.90
NA	NA	293633	17.35

# race\_\_enlistment

**Label:** Race

**Description:** race\_\_enlistment is a numeric variable that reports an individual's race, as reported in the World War II Army Enlistment Records.

Codes are also available at: [https://usa.ipums.org/usa-action/variables/RACE#codes\\_section](https://usa.ipums.org/usa-action/variables/RACE#codes_section)

race__enlistment	label	n	freq %
100	White	1540362	91.023
150	Puerto Rican	2003	0.118
200	Black	119655	7.071
300	Native American	2905	0.172
400	Chinese	1747	0.103
500	Japanese	5408	0.320
600	Filipino	2973	0.176
630	Native Hawaiian	670	0.040
991	Other	6041	0.357
NA	NA	10515	0.621

# citizenship\_\_enlistment

**Label:** Citizenship

**Description:** citizenship\_\_enlistment is a numeric variable that reports an individual's citizenship, as reported in the World War II Army Enlistment Records.

citizenship__enlistment		label	n	freq %
0	Non-Citizen	12132	0.72	
1	Citizen	1669632	98.66	
NA	NA	10515	0.62	

# civilian\_\_occupation\_\_enlistment

**Label:** Civilian Occupation

**Description:** civilian\_\_occupation\_\_enlistment is a numeric variable that reports a person's occupation at time of enlistment, as reported in the World War II Army Enlistment Records.

For complete codes, please see: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=3323&dt=893&c\\_id=24986](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=3323&dt=893&c_id=24986)

# marital\_\_status\_\_enlistment

**Label:** Marital Status

**Description:** marital\_status\_enlistment is a numeric variable that reports a person's marital status at time of enlistment, as reported in the World War II Army Enlistment Records.

Codes are also available at: [https://usa.ipums.org/usa-action/variables/MARST#codes\\_section](https://usa.ipums.org/usa-action/variables/MARST#codes_section)

marital_status_enlistment	label	n	freq %
1	Married	357552	21.13
3	Separated	14495	0.86
4	Divorced	16705	0.99
5	Widowed	3015	0.18
6	Single	1297789	76.69
NA	NA	2723	0.16

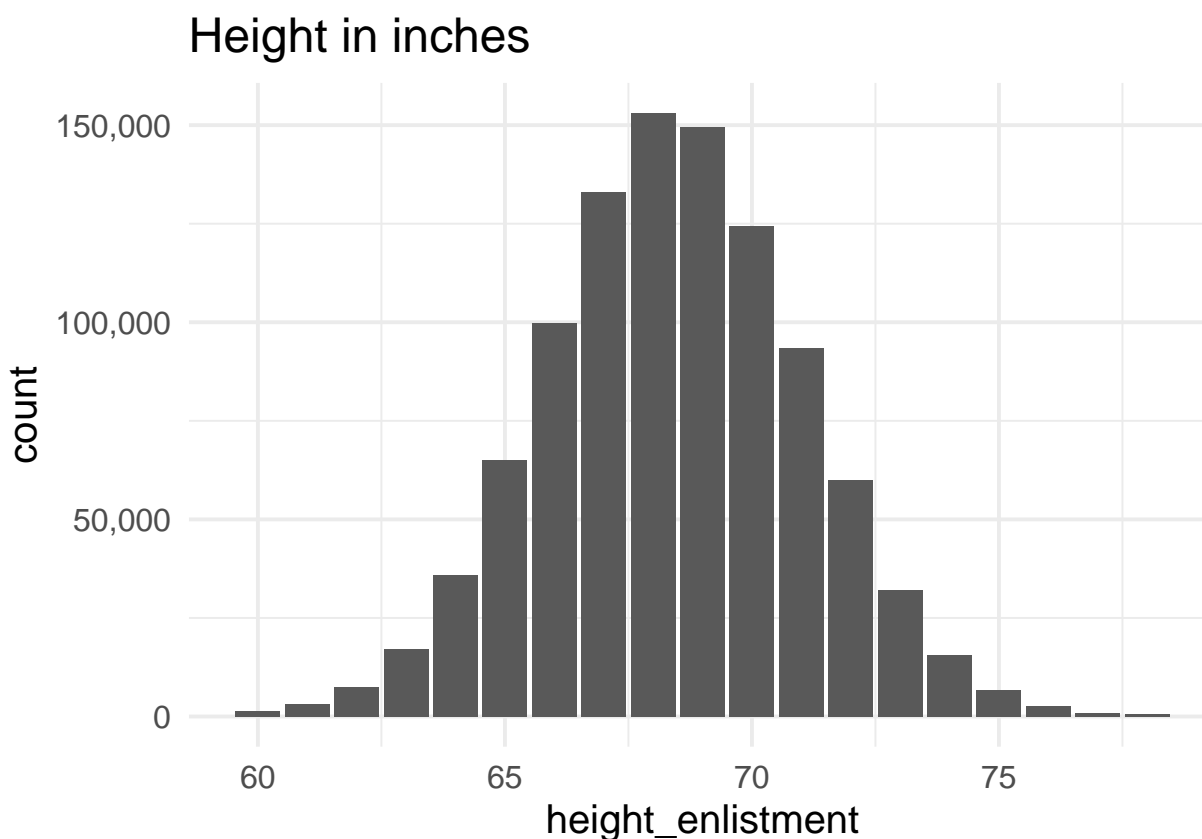


# height\_enlistment

**Label:** Height

**Description:** height\_enlistment is a numeric variable that reports a person's height in inches at time of enlistment, as reported in the World War II Army Enlistment Records. Height was filtered based on the range of permissible heights for enlistment in the military, 60 - 78 inches ([Foster et al, 1967](#)).

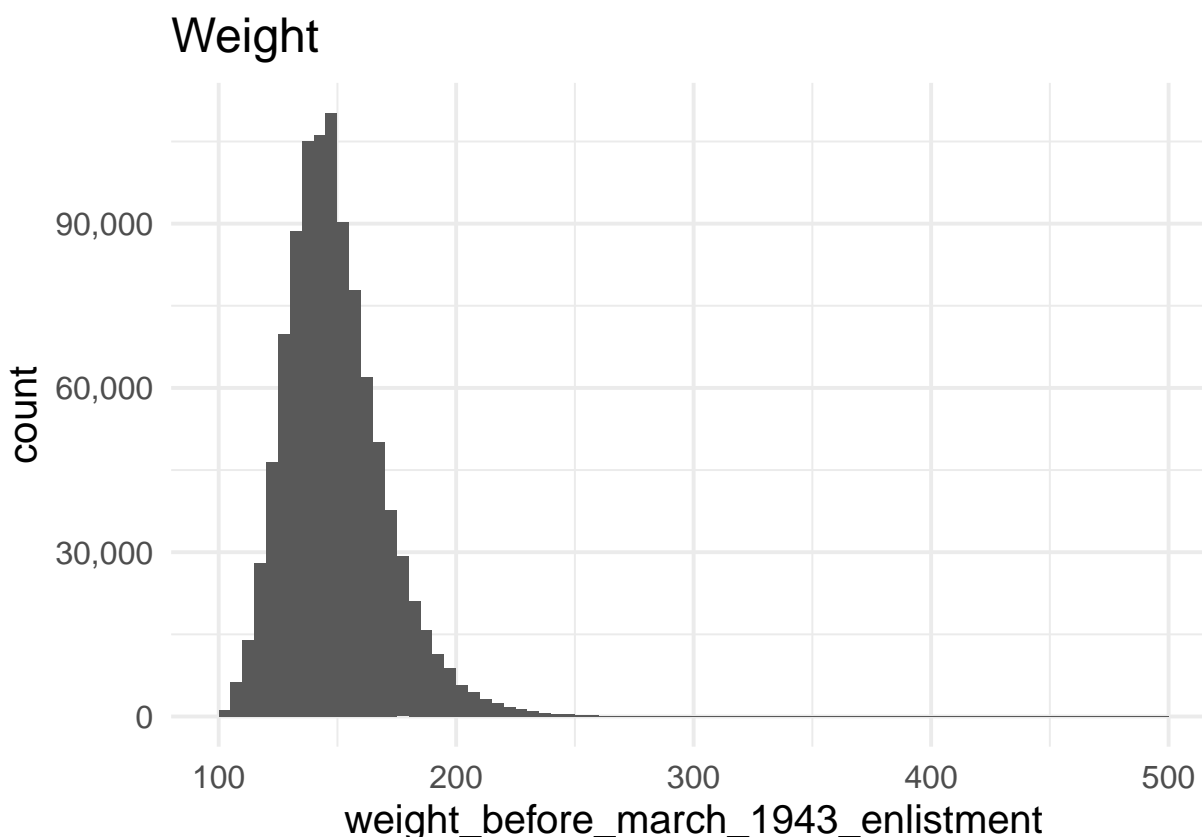
**Note:** Please be aware that instructions on how to use this field on enlistment punch cards varied over time. This field may encode military occupation or other information depending on time and site of enlistment. We have removed height data for persons who enlisted in March of 1943 or later, as the distribution of this variable changes significantly at that point. However, it is possible that erroneous data remain in this field. See the [National Archives' technical documentation on the Amy Serial Number Electronic File](#) for further information.



## weight\_before\_march\_1943\_enlistment

**Description:** weight\_before\_march\_1943\_enlistment is a numeric variable that reports a person's weight in pounds at time of enlistment as reported in the World War II Army Enlistment Records. This variable is only available for enlistees who enlisted before March of 1943. Weight is filtered to a minimum of 105 pounds, the minimum permissible weight for a 60-inch-tall enlistee ([Foster et al, 1967](#)). While there was no clear maximum allowable body weight for army enlistees, we have removed values of over 500.

**Note:** On enlistment punchcards, the field used to record weight was also used for Army General Classification Test (AGCT) scores, and may also encode military occupation or other information. Weight may have been recorded through 1943 at some enlistment sites, but the recording of AGCT score instead of weight at many enlistment sites likely began in March 1943. Because of this ambiguity, we only make this variable available for enlistees who enlisted before March 1943. Refer to the [National Archives' technical documentation on the Amy Serial Number Electronic File](#) for further information. For enlistees who enlisted in March 1943 and later, we have created an additional variable called **weight\_or\_AGCT\_enlistment**.



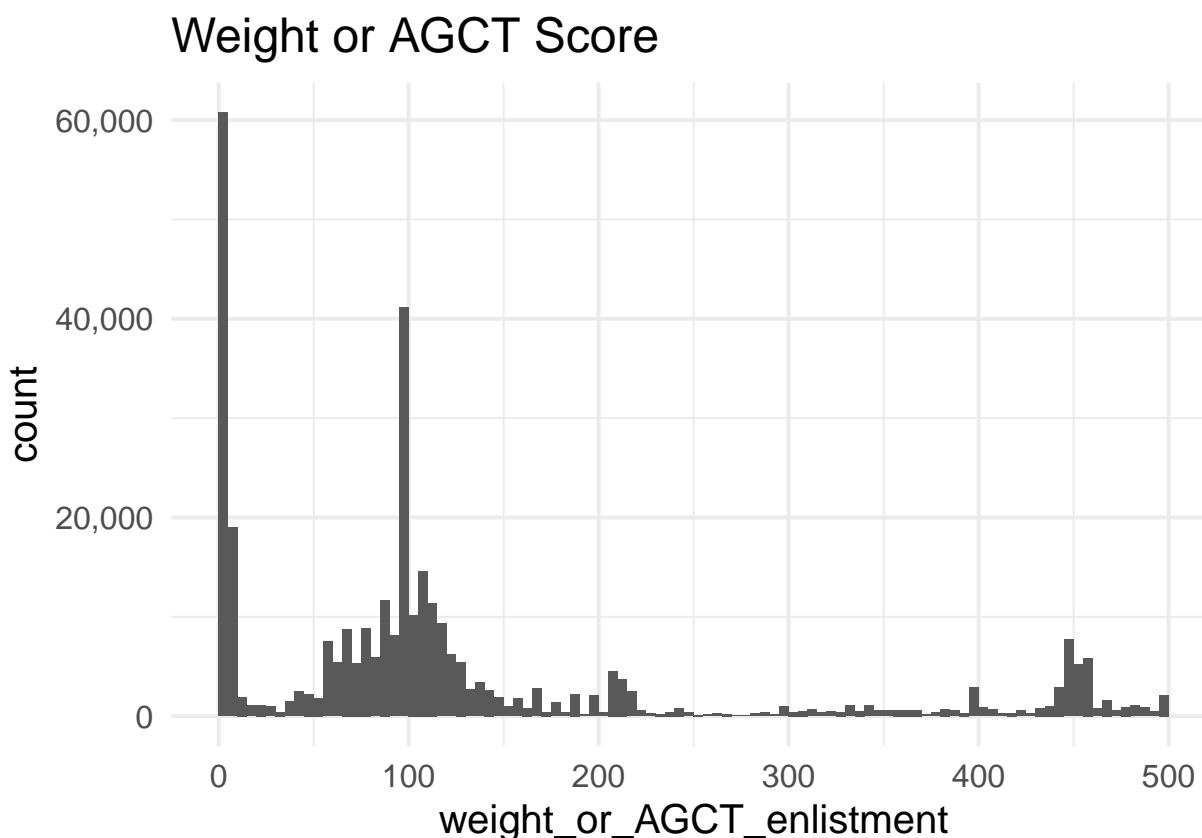
# weight\_or\_AGCT\_enlistment

**Label:** Weight (Pounds) at Enlistment or AGCT Score

**Description:** weight\_or\_AGCT\_enlistment is a numeric variable that reports a person's weight in pounds at time of enlistment or their Army General Classification Test (AGCT) score, as reported in the World War II Army Enlistment Records. This variable is only available for enlistees who enlisted in March 1943 or later. We have removed zeros and values above 500, which likely exceed the maximum realistic body weights and AGCT scores, but otherwise leave the data in this field unfiltered and unsorted.

**Note:** On enlistment punch cards, the field used to record weight was also used for Army General Classification Test (AGCT) scores, and may also encode military occupation or other information. Weight may have been recorded through 1943 at some enlistment sites, but the recording of AGCT score instead of weight at many enlistment sites likely began in March 1943. Because of this ambiguity, we created this variable for persons enlisted in March 1943 or later, which may contain weight or AGCT or other information.

Refer to the [National Archives' technical documentation on the Army Serial Number Electronic File](#) for further information. For enlistees prior to March 1943, we used the information recorded in this field to create the weight\_before\_march\_1943\_enlistment variable.



# component\_\_enlistment

**Label:** Army Component

**Description:** component\_\_enlistment is a numeric variable that records the component of the army in which an individual served, as recorded in the World War II Army Enlistment Records.

For more detailed variable labels, see: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=2078&dt=929&c\\_id=25014](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2078&dt=929&c_id=25014)

component__enlistment	label (abbreviated)	n	freq %
1	Regular Army	246185	14.55
2	Regular Army Reserve	60	0.00
3	Reserves	105437	6.23
4	National Guard	49068	2.90
5	Philippine Scouts	810	0.05
6	Army of the United States	140730	8.32
7	Selectees	1146992	67.78
9	Women's Army Corps	121	0.01
NA	NA	2876	0.17

# source\_\_enlistment

**Label:** Source of Army Personnel

**Description:** source\_\_enlistment is a numeric variable that reports the source of the enlisted person, as reported in the World War II Army Enlistment Records. This denotes if an enlistee was a civilian prior to enlistment, or if they were already enlisted in the National Guard or another faction of the military.

source__ enlistment	label	n	freq %
0	Civil Life	1399921	82.724
1	Enlisted Man, Regular Army, within 3 months of Discharge or former WAAC Auxiliary	3555	0.210
2	National Guard in Federal Service, within 3 months of Discharge	75401	4.456
3	Enlisted Man, Regular Army, after 3 months of Discharge	7230	0.427
4	National Guard in Federal Service, after 3 months of Discharge	5768	0.341
5	Enlisted Reserve or Medical Administrative Corps (MAC) Officer	34065	2.013
6	Enlisted Man, Philippine Scout or recall to AD of an enlisted man who had been transferred to the ERC	77575	4.584
7	National Guard	50838	3.004
8	Returned to Military Control, Surrendered	1164	0.069
9	Returned to Military Control Apprehended	8	0.000
11	Selective Service within 3 months of Discharge	7389	0.437
12	Selective Service after 3 months of Discharge	23	0.001
NA	NA	29342	1.734

# HISTID

**Label:** Historical Unique Identifier (1940 Census)

**Description:** HISTID is a unique individual-level identifier assigned to individuals in IPUMS-USA 1940 Census data. This variable is available for individuals where a link directly between army enlistment records and the 1940 Census has been established.

HISTID	n	freq	%
non-missing	761518		45
NA	930761		55