CenSoc Enlistment-Numident V2.1 Codebook*

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

Summary: The CenSoc Enlistment-Numident file links 1,693,544 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

^{*}Last updated: 30 July, 2024

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).

id

Label: CenSoc Numident-Enlistment Unique Identifier

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

Note: new ids are generated for each version of this dataset, and are not compatible with previous releases.

\mathbf{Sex}

Label: Sex

Description: sex reports a person's sex.

sex	label	n	freq %
1	Men	1693544	100

bpl

Label: Place of Birth

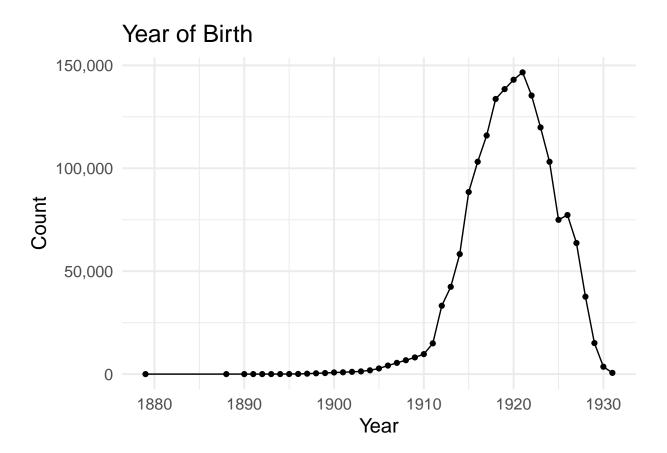
Description: bpl is a numeric variable that reports a person's place of birth, as reported in the World War II Army Enlistment Records and Numident records. The coding schema matches the **detailed** IPUMS-USA Birthplace coding schema. For a complete list of IPUMS codes, please see: https://usa.ipums.org/usa-action/variables/BPL#codes_section.

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

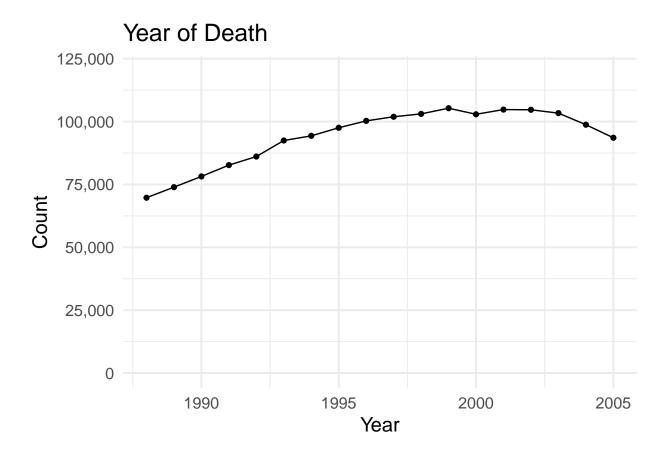
Description: bmonth_numident reports a person's month of birth, as recorded in the Numident death records.

bmonth_numident	label	n	freq %
1	January	143138	8.45
2	February	134716	7.95
3	March	147333	8.70
4	April	137528	8.12
5	May	139799	8.25
6	June	135436	8.00
7	July	142978	8.44
8	August	148594	8.77
9	September	146694	8.66
10	October	144296	8.52
11	November	134445	7.94
12	December	138587	8.18

dyear_numident

Label: Death Year

 $\textbf{Description:} \ \ \text{dyear_numident reports a person's year of death, as recorded in the Numident death records.}$



$dmonth_numident$

Label: Death Month

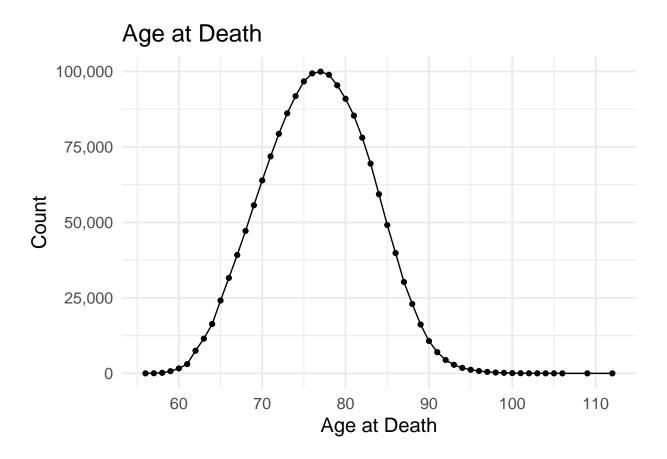
Description: dmonth_numident reports a person's month of death, as recorded in the Numident death records.

dmonth_numident	label	n	freq %
1	January	160296	9.47
2	February	142031	8.39
3	March	152140	8.98
4	April	139981	8.27
5	May	139098	8.21
6	June	131209	7.75
7	July	132772	7.84
8	August	131890	7.79
9	September	129383	7.64
10	October	140155	8.28
11	November	139687	8.25
12	December	154902	9.15

${\bf death_age_numident}$

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	1520787	89.80
2	Black	118274	6.98
3	Other	15189	0.90
4	Asian	2304	0.14
5	Hispanic	1842	0.11
6	North American Native	395	0.02
NA	Missing	34753	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	1515547	89.49
2	Black	118033	6.97
3	Other	12089	0.71
4	Asian	4436	0.26
5	Hispanic	6989	0.41
6	North American Native	1697	0.10
NA	Missing	34753	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.

$socstate_numident$

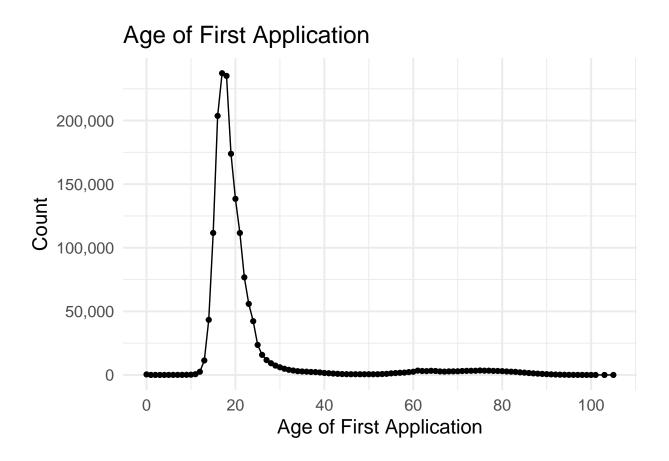
Label: State where Social Security Number Issued

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_sec tion.

${\bf age_first_application_numident}$

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.

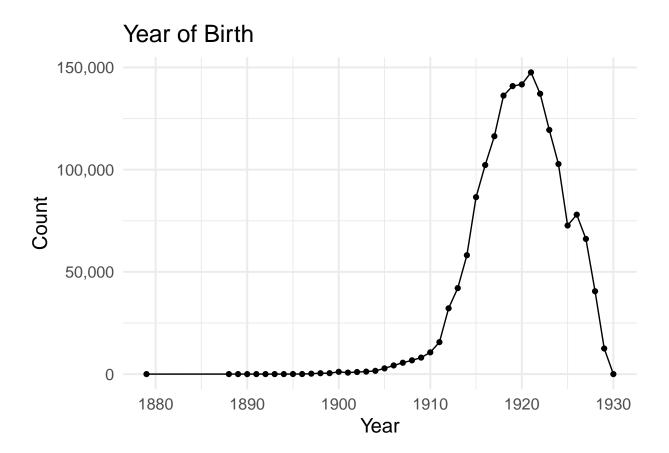


byear_enlistment

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.

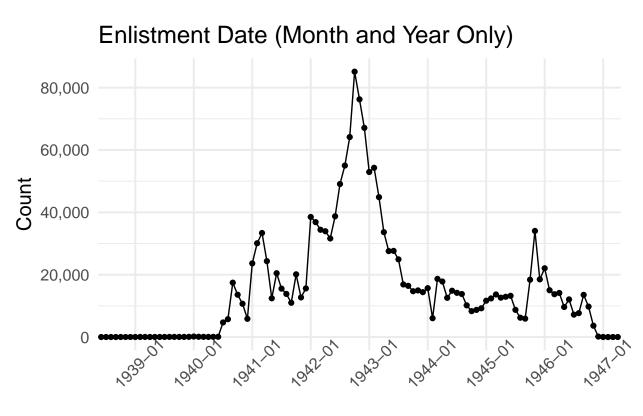
Note: The matching algorithm used allows for up to two years of flexibility on birth year, so birth year in the BUNMD may disagree with birth year in enlistment records. Birth years are discordant by one year in 7.2% of records, and discordant by two years in 4.1% of records. Both birth year variables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



Date (yyyy-mm)

$residence_state_enlistment$

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.

$residence_state_fips_enlistment$

Label: State FIPS Code of Residence

Description: residence_state_fips is a character variable (2 digits, including leading zeros) that reports the FIPS code of a person's state of residence at time of enlistment. This variable is only available for enlistees residing in the United States, Hawaii, Alaska, and Puerto Rico. For a complete list of FIPS codes, please see: https://transition.fcc.gov/oet/info/maps/census/fips/fips.txt

residence_county_fips_enlistment

Label: County FIPS Code of Residence

Description: residence_county_fips is a character variable (3 digits, including leading zeros) that reports a person's FIPS code of county 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. County is not available for residents of Hawaii and Alaska, which were not yet states.

These are historical codes that may not always align with modern FIPS codes due to county splits, merges, name changes, etc. since the 1940s. For a complete list of these codes, please see: 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 Documention. (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. Users may combine 2-digit residence_state_fips_enlistment codes with the 3-digit county FIPS codes to obtain unique 5-digit county codes. 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=89 $3\&c_id=24997$

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

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	Grammar school or less than grade 9	438361	25.88
3	Grade 9	144958	8.56
4	Grade 10	181868	10.74
5	Grade 11	166585	9.84
6	Grade 12	507940	29.99
7	1 year of college	103055	6.09
8	2 years of college	58069	3.43
9	3 years of college	29565	1.75
10	4 years of college	56281	3.32
11	5+ years of college	6862	0.41

${\bf grade_code_enlistment}$

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	1536	0.09
3	Technical Sergeant	4472	0.26
4	Staff Sergeant	8788	0.52
5	Sergeant	12190	0.72
6	Corporal	16593	0.98
7	Private First Class	50176	2.96
8	Private	1573157	92.89
9	Aviation Cadet	8525	0.50
NA	NA	13767	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_c	code_enlistment	label	n	freq $\%$
	0	Branch Immaterial - Warrant Officers, USA	752298	44.422
	1	General Staff Corps	50	0.003
	2	No branch assignment	540339	31.906
	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	3451	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	41954	2.477
	20	Air Corps	205410	12.129
	30	Field Artillery	21234	1.254
	38	Transportation Corps	4032	0.238
	40	Coast Artillery Corps or Army Mine Planter Service	15409	0.910
	50	Quartermaster Corps	14757	0.871
	70	Medical Department or Army Nurse Corps	10589	0.625
	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	8845	0.522
	87	Ordnance Department	10526	0.622
	90	Signal Corps	19328	1.141
	93	Cavalry	5646	0.333
	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	1.680

$term_of_enlistment_enlistment$

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

$term_of_enlistment_$			
enlistment	label (abbreviated)	n	freq $\%$
1	One year enlistment	42174	2.49
2	Enlistment for Alaska	4479	0.26
3	Enlistment for Hawaiian Department	89996	5.31
4	Enlistment for assignment to another corps area	3471	0.20
5	Enlistment for the duration of the War or other	1160170	68.51
	emergency		
6	Enlistment for the Panama Canal Department	65502	3.87
7	Enlistment for the Philippine Department	17593	1.04
8	Enlistment for the Puerto Rican Department	1248	0.07
9	Three year enlistment	15272	0.90
NA	NA	293639	17.34

$race_enlistment$

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.

 $Codes\ are\ also\ available\ at:\ https://usa.ipums.org/usa-action/variables/RACE\#codes_section$

race_enlistment	label	n	freq %
100	White	1540394	90.957
150	Puerto Rican	3216	0.190
200	Black	119660	7.066
300	Native American	2905	0.172
400	Chinese	1747	0.103
500	Japanese	5408	0.319
600	Filipino	2973	0.176
630	Native Hawaiian	670	0.040
991	Other	6040	0.357
NA	NA	10531	0.622

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	12198	0.72
1	Citizen	1670815	98.66
NA	NA	10531	0.62

civilian_occupation_enlistment

Label: Civilian Occupation

Description: civilian_occupation is a character variable (3 digits, including leading zeros) 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_i d=24986. Only meaningful occupational codes from this code list have been retained in this dataset, plus the addition of code "992", which corresponds to "Student".

Note: Original 3-digit occupation codes (including leading zeros) are organized by occupational group/classification. For more information, see pages 164-165 of Nara Compiled Technical Documention.. odes, please see: 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.

 $Codes\ are\ also\ available\ at:\ https://usa.ipums.org/usa-action/variables/MARST\#codes_section$

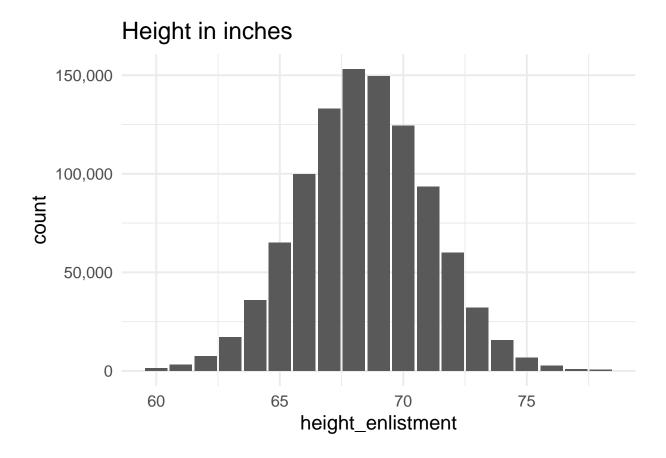
marital_status_enlistment	label	n	freq %
1	Married	357722	21.12
3	Separated	14502	0.86
4	Divorced	16710	0.99
5	Widowed	3017	0.18
6	Single	1298868	76.70
NA	NA	2725	0.16

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).

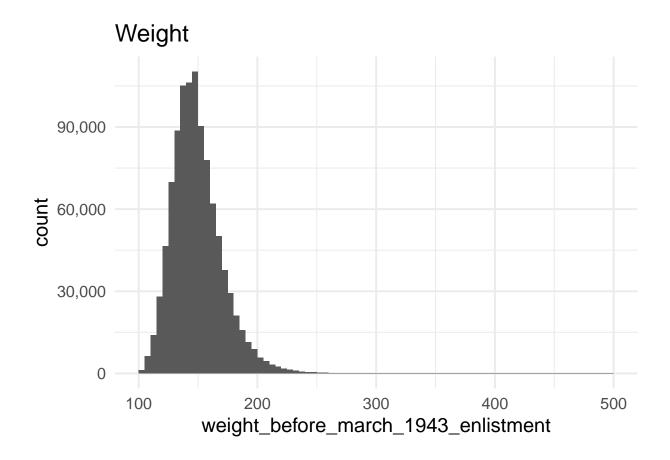
Note: Please be aware that instructions on how to use this field on enlistment punch cards varied over time. This field may encode military occupation or other information depending on time and site of enlistment. We have removed height data for persons who enlisted in March of 1943 or later, as the 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.



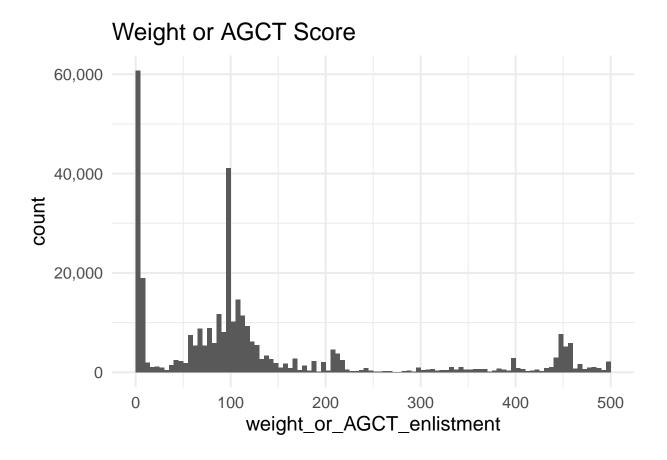
weight_or_AGCT_enlistment

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 Amy 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=92 $9\&c_id=25014$

component_enlistment	label (abbreviated)	n	freq %
1	Regular Army	246140	14.53
2	Regular Army Reserve	60	0.00
3	Reserves	105441	6.23
4	National Guard	49067	2.90
5	Philippine Scouts	810	0.05
6	Army of the United States	140732	8.31
7	Selectees	1148295	67.80
9	Women's Army Corps	121	0.01
NA	NA	2878	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_			freq
enlistment	label	n	%
0	Civil Life	1401230	82.740
1	Enlisted Man, Regular Army, within 3 months of Discharge or former	3555	0.210
	WAAC Auxiliary		
2	National Guard in Federal Service, within 3 months of Discharge	75398	4.452
3	Enlisted Man, Regular Army, after 3 months of Discharge	7228	0.427
4	National Guard in Federal Service, after 3 months of Discharge	5766	0.340
5	Enlisted Reserve or Medical Administrative Corps (MAC) Officer	34064	2.011
6	Enlisted Man, Philippine Scout or recall to AD of an enlisted man who had	77543	4.579
	been transferred to the ERC		
7	National Guard	50837	3.002
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	0.436
12	Selective Service after 3 months of Discharge	23	0.001
NA	NA	29339	1.732

$enlistment_status_enlistment$

Label: Status of Enlistment

Description: enlistment_status_enlistment is a numeric variable that reports whether the enlisted person was drafted by Selective Service or voluntarily enlisted. This variable was constructed using information from the **component** variable.

enlistment_status_enlistment	label	n	freq %
0	Voluntary Service	542371	32.026
1	Draftee	1148295	67.804
NA	NA	2878	0.170

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	761551	44.968
NA	931993	55.032