

CenSoc Enlistment-Census-1940 V1 Codebook *

Page	Variable	Label
2	HISTID	1940 Census Unique Identifier
3	byear	Year of Birth
4	sex	Sex
5	date_of_enlistment	Date of Enlistment
6	bpl	Place of Birth
7	residence_state	State of Residence at Enlistment
8	residence_county	County of Residence at Enlistment
9	place_of_enlistment	Place of Enlistment
10	education	Education
11	grade_code	Army Grade (Rank)
12	branch_code	Army Branch
14	term_of_enlistment	Term of Enlistment
15	race	Race
16	citizenship	Citizenship
17	civilian_occupation	Civilian Occupation
18	marital_status	Marital Status at Enlistment
19	height	Height at Enlistment (Inches)
20	weight_before_march_1943	Weight (Pounds)
21	weight_or_AGCT	Weight (Pounds) or AGCT score
22	component	Army Component
23	source	Source of Army Personnel

Summary: The CenSoc Enlistment-Census-1940 file links 2,573,678 men from World War II era army enlistment records (circa 1938-1947) to the 1940 Census. We include the HISTID identifier to allow researchers to merge this file with IPUMS 1940 Census data. All variables included here, such as marital status and education, were measured at time of enlistment and are not taken from 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

HISTID

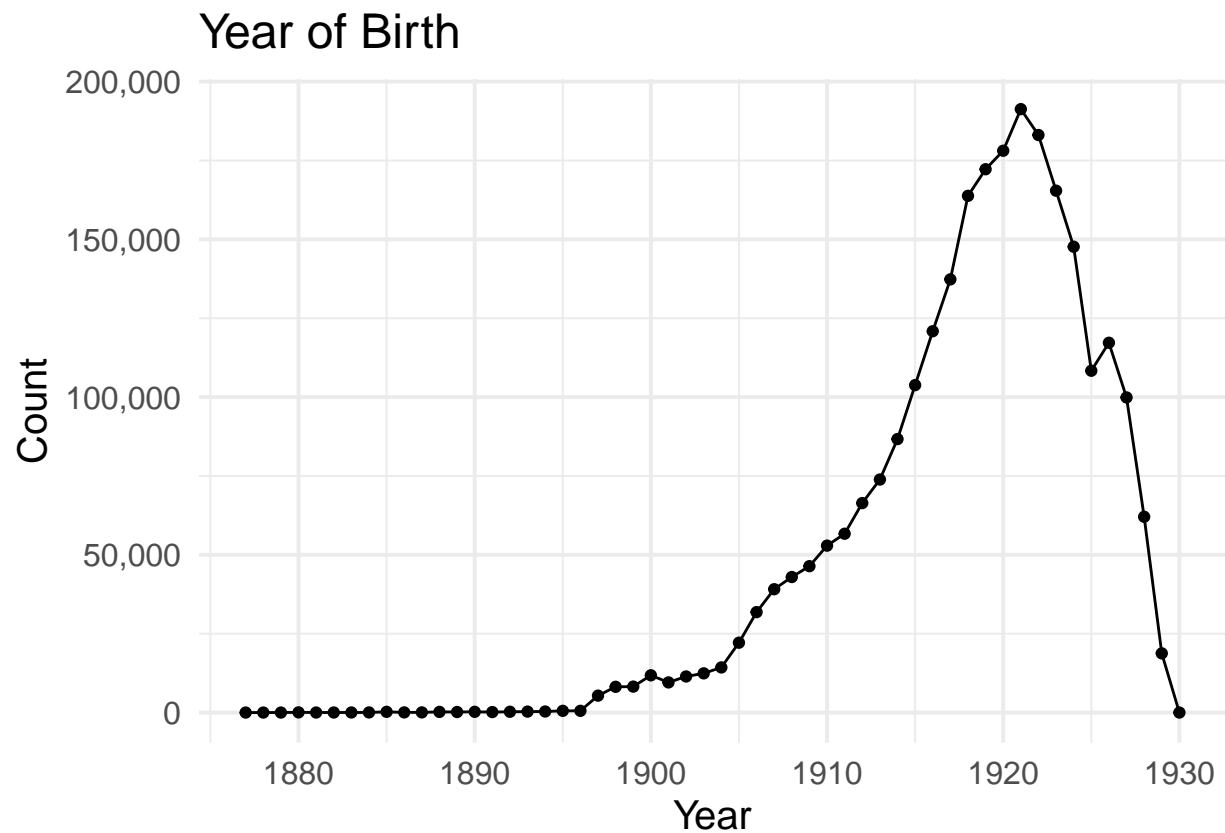
Label: 1940 Census Historical Unique Identifier

Description: HISTID is a unique individual-level identifier. It can be used to merge this file with full count 1940 Census data from IPUMS-USA.

byear

Label: Year of Birth

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



SEX

Label: Sex

Description: sex is a numeric variable that reports a person's sex, as inferred by army branch in World War II enlistment records.

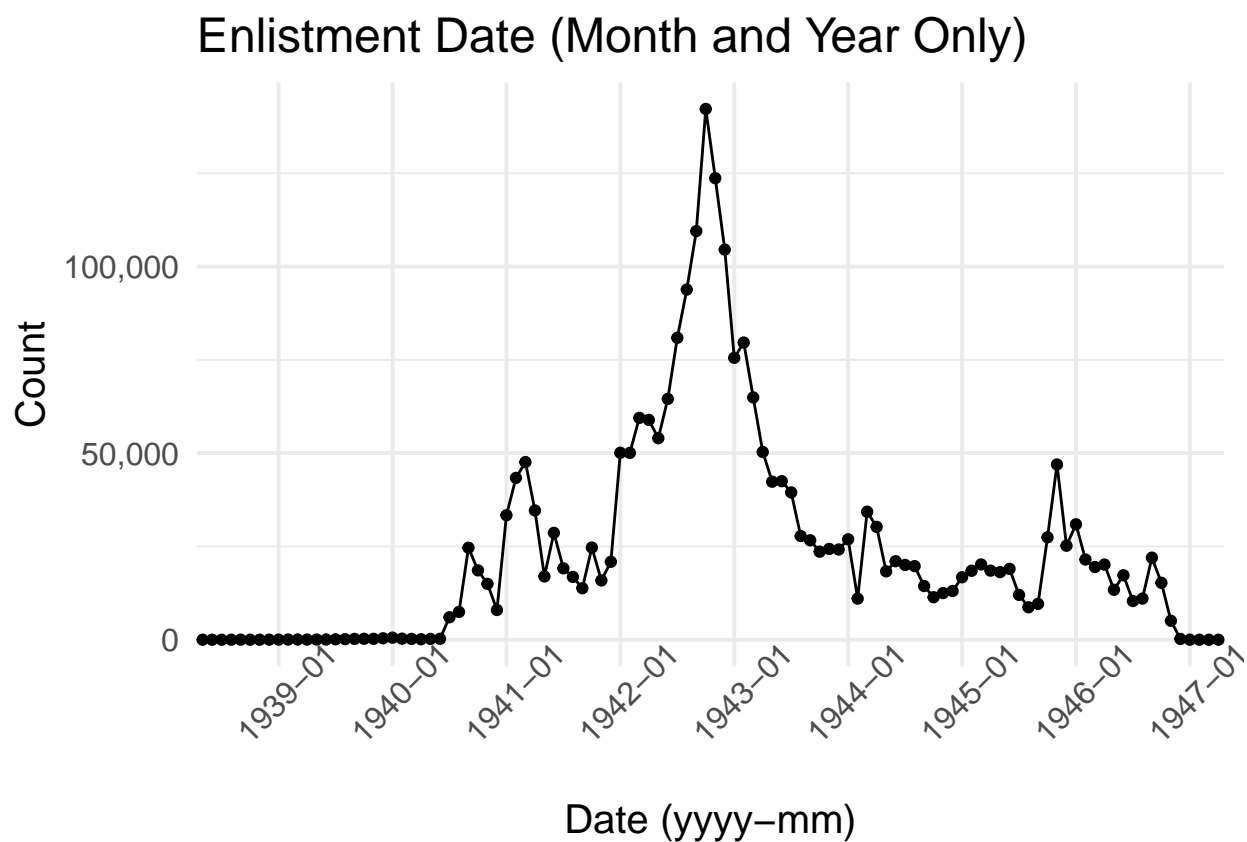
Note: Sex was not listed on enlistee punch cards and was therefore inferred from army branch in enlistment records.

sex	label	n	freq %
1	Men	2573678	100

date_of_enlistment

Label: Date of Enlistment

Description: date_of_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.



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. 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

Label: State of Residence

Description: residence__state 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__county

Label: County of Residence

Description: residence__county 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. 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

Label: Place of Enlistment

Description: place__of__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

education

Label: Education

Description: education 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	label	n	freq %
1	Nursery school to grade 4	686708	26.68
3	Grade 9	213721	8.30
4	Grade 10	271080	10.53
5	Grade 11	249967	9.71
6	Grade 12	758652	29.48
7	1 year of college	162463	6.31
8	2 years of college	90552	3.52
9	3 years of college	44965	1.75
10	4 years of college	84032	3.27
11	5+ years of college	11538	0.45

grade__code

Label: Army Grade

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

grade__code	label	n	freq %
1	Master Sergeant	7933	0.31
2	First Sergeant	3216	0.12
3	Technical Sergeant	7255	0.28
4	Staff Sergeant	13420	0.52
5	Sergeant	19435	0.76
6	Corporal	23321	0.91
7	Private First Class	68550	2.66
8	Private	2395450	93.07
9	Aviation Cadet	12364	0.48
NA	NA	22734	0.88

branch__code

Label: Army Branch

Description: branch__code 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	label	n	freq %
0	Branch Immaterial - Warrant Officers, USA	1164986	45.265
1	General Staff Corps	57	0.002
2	No branch assignment	819257	31.832
3	Philippine Scouts	6738	0.262
4	Professor US Military Academy	315	0.012
5	Military Intelligence	489	0.019
6	Corps of Military Police	4779	0.186
7	Specialist Reserve	9	0.000
8	Detached Enlisted Men's List	3624	0.141
9	Inactive Reserve	915	0.036
10	Infantry	59876	2.326
20	Air Corps	305442	11.868
30	Field Artillery	29851	1.160
38	Transportation Corps	5525	0.215
40	Coast Artillery Corps or Army Mine Planter Service	21467	0.834
50	Quartermaster Corps	19679	0.765
70	Medical Department or Army Nurse Corps	15027	0.584
71	Medical Corps - For Officers only	13	0.001
72	Hospital Dietitians	273	0.011
73	Physical Therapy Aides	29	0.001
74	Dental Corps - For Officers only	38	0.001
75	Veterinary Corps or Service	72	0.003
77	Medical Administrative Corps - For Officers only	4	0.000
78	Adjutant General's Department	330	0.013
79	Pharmacy Corps	29	0.001
80	Chemical Warfare Service	995	0.039
82	Corps of Engineers	12808	0.498
87	Ordnance Department	17544	0.682
90	Signal Corps	28946	1.125
93	Cavalry	8130	0.316
96	Chaplains	74	0.003
97	Finance Department	803	0.031
98	Inspector General's Department	81	0.003
99	Judge Advocate General's Department	2558	0.099
NA	NA	42915	1.667

term_of_enlistment

Label: Term of Enlistment

Description: term_of_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	label	n	freq %
1	One year enlistment	58267	2.264
2	Enlistment for Alaska	7050	0.274
3	Enlistment for Hawaiian Department	124647	4.843
4	Enlistment for assignment to another corps area	4821	0.187
5	Enlistment for the duration of the War or other emergency	1814910	70.518
6	Enlistment for the Panama Canal Department	97018	3.770
7	Enlistment for the Philippine Department	24261	0.943
8	Enlistment for the Puerto Rican Department	1763	0.069
9	Three year enlistment	21442	0.833
NA	NA	419499	16.300

race

Label: Race

Description: race 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	label	n	freq %
100	White	2390531	92.88
150	Puerto Rican	293	0.01
200	Black	150234	5.84
300	Native American	4630	0.18
400	Chinese	1506	0.06
500	Japanese	3219	0.13
600	Filipino	209	0.01
630	Native Hawaiian	47	0.00
991	Other	5716	0.22
NA	NA	17293	0.67

citizenship

Label: Citizenship

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

citizenship	label	n	freq %
0	Non-Citizen	15260	0.59
1	Citizen	2541125	98.74
NA	NA	17293	0.67

civilian_occupation

Label: Civilian Occupation

Description: civilian_occupation 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

marital_status

Label: Marital Status

Description: marital_status 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

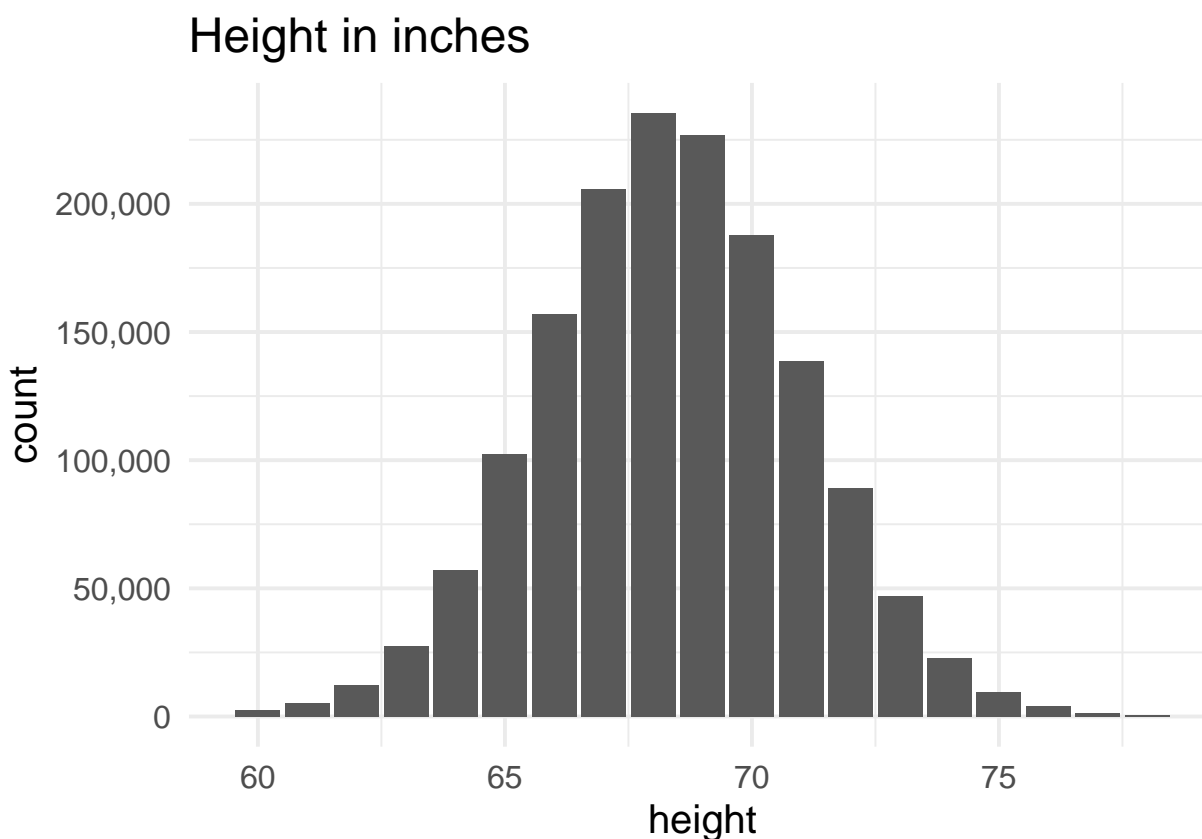
marital_status	label	n	freq %
1	Married	589592	22.91
3	Separated	34251	1.33
4	Divorced	48168	1.87
5	Widowed	9135	0.35
6	Single	1887849	73.35
NA	NA	4683	0.18

height

Label: Height

Description: height 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](#)).

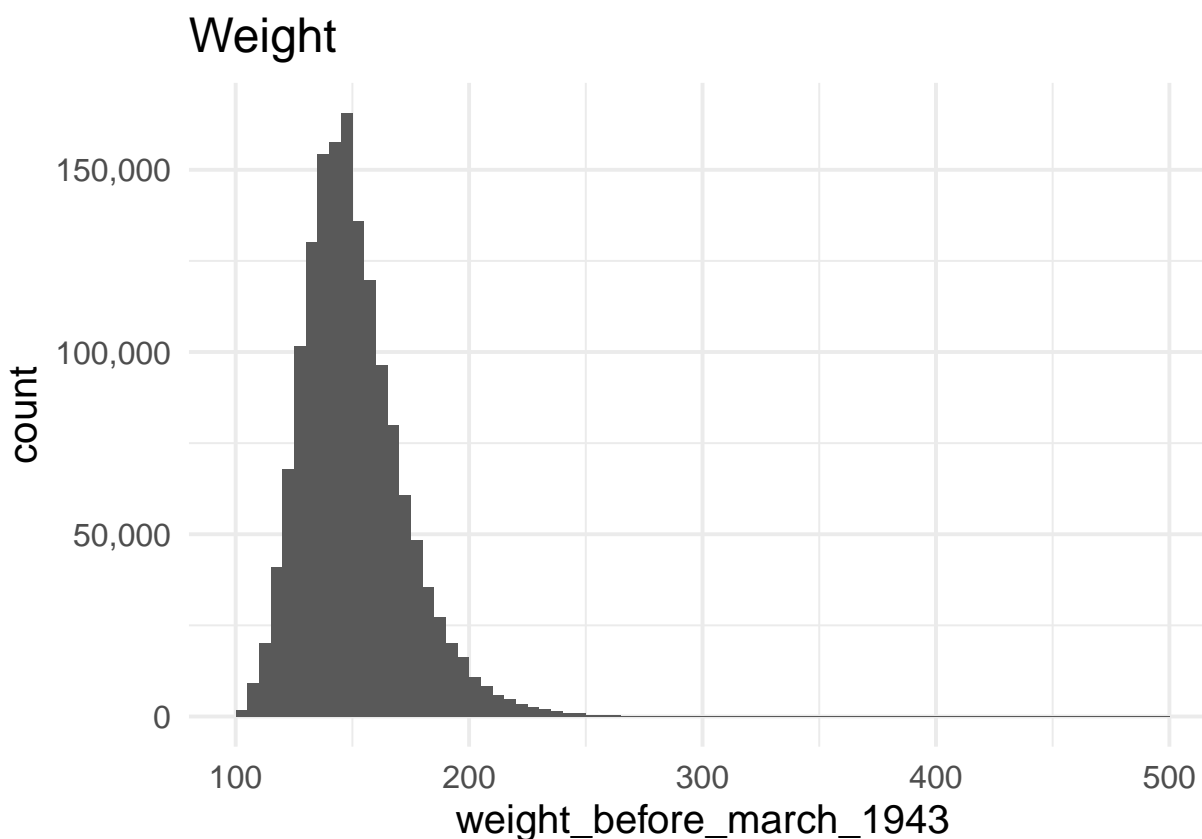
Note: Please be aware that instructions on how to use this field on enlistment punch cards varied over time. This field may encode military occupation or other information depending on time and site of enlistment. We have removed height data for persons who enlisted in March of 1943 or later, as the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



weight_before_march_1943

Description: weight_before_march_1943 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**.



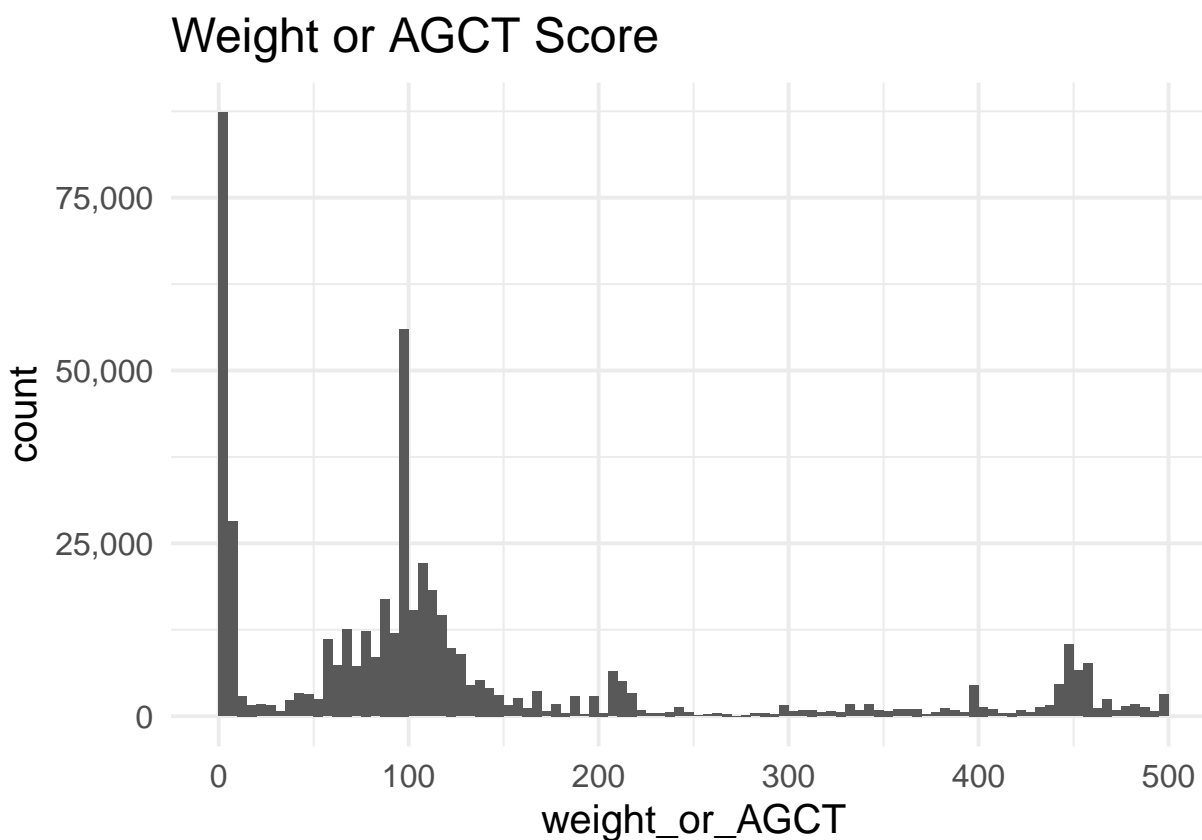
weight_or_AGCT

Label: Weight (Pounds) at Enlistment or AGCT Score

Description: weight_or_AGCT 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** variable.



component

Label: Army Component

Description: component 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

component	label (abbreviated)	n	freq %
1	Regular Army	348983	13.56
2	Regular Army Reserve	83	0.00
3	Reserves	173148	6.73
4	National Guard	70893	2.75
5	Philippine Scouts	33	0.00
6	Army of the United States	211352	8.21
7	Selectees	1764681	68.57
9	Women's Army Corps	217	0.01
NA	NA	4288	0.17

source

Label: Source of Army Personnel

Description: source 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	label	n	freq %
0	Civil Life	2130200	82.769
1	Enlisted Man, Regular Army, within 3 months of Discharge or former WAAC Auxiliary	5676	0.221
2	National Guard in Federal Service, within 3 months of Discharge	122331	4.753
3	Enlisted Man, Regular Army, after 3 months of Discharge	14935	0.580
4	National Guard in Federal Service, after 3 months of Discharge	12436	0.483
5	Enlisted Reserve or Medical Administrative Corps (MAC) Officer	56790	2.207
6	Enlisted Man, Philippine Scout or recall to AD of an enlisted man who had been transferred to the ERC	104052	4.043
7	National Guard	73822	2.868
8	Returned to Military Control, Surrendered	1381	0.054
9	Returned to Military Control Apprehended	6	0.000
11	Selective Service within 3 months of Discharge	9116	0.354
12	Selective Service after 3 months of Discharge	51	0.002
NA	NA	42882	1.666