

## CenSoc WWII Army Enlistment Dataset Codebook\*

Page	Variable	Label
2	serial_number	Army Serial Number
3	byear	Year of Birth
4	fname	First Name
5	mname	Middle Name
6	lname	Last Name
7	sex	Sex
8	date_of_enlistment	Date of Enlistment
9	bpl	Place of Birth
10	residence_state	State of Residence at Enlistment
11	residence_county	County of Residence at Enlistment
12	place_of_enlistment	Place of Enlistment
13	education	Education
14	grade_code	Army Grade (rank)
15	branch_code	Army Branch
17	term_of_enlistment	Term of Enlistment
18	race	Race
19	citizenship	Citizenship
20	civilian_occupation	Civilian Occupation
21	marital_status	Marital Status at Enlistment
22	height	Height at Enlistment (inches)
23	weight_before_march_1943	Weight (pounds)
24	weight_or_AGCT	Weight (pounds) or AGCT score
25	component_enlistment	Army component
26	source	Source of Army Personnel

**Summary:** The CenSoc WWII Army Enlistment Dataset captures 9,039,840 American men and women enlisted in the United States Army and the Women's Army Auxiliary Corps circa 1938-1946. This dataset is a cleaned and harmonized version of the National Archives and Records Administration's Electronic Army Serial Number Merged File, ca. 1938 - 1946 (2002). It contains information on enlistee sociodemographics, military service, and anthropometry.

---

\*Last updated: 23 July, 2023

# serial\_number

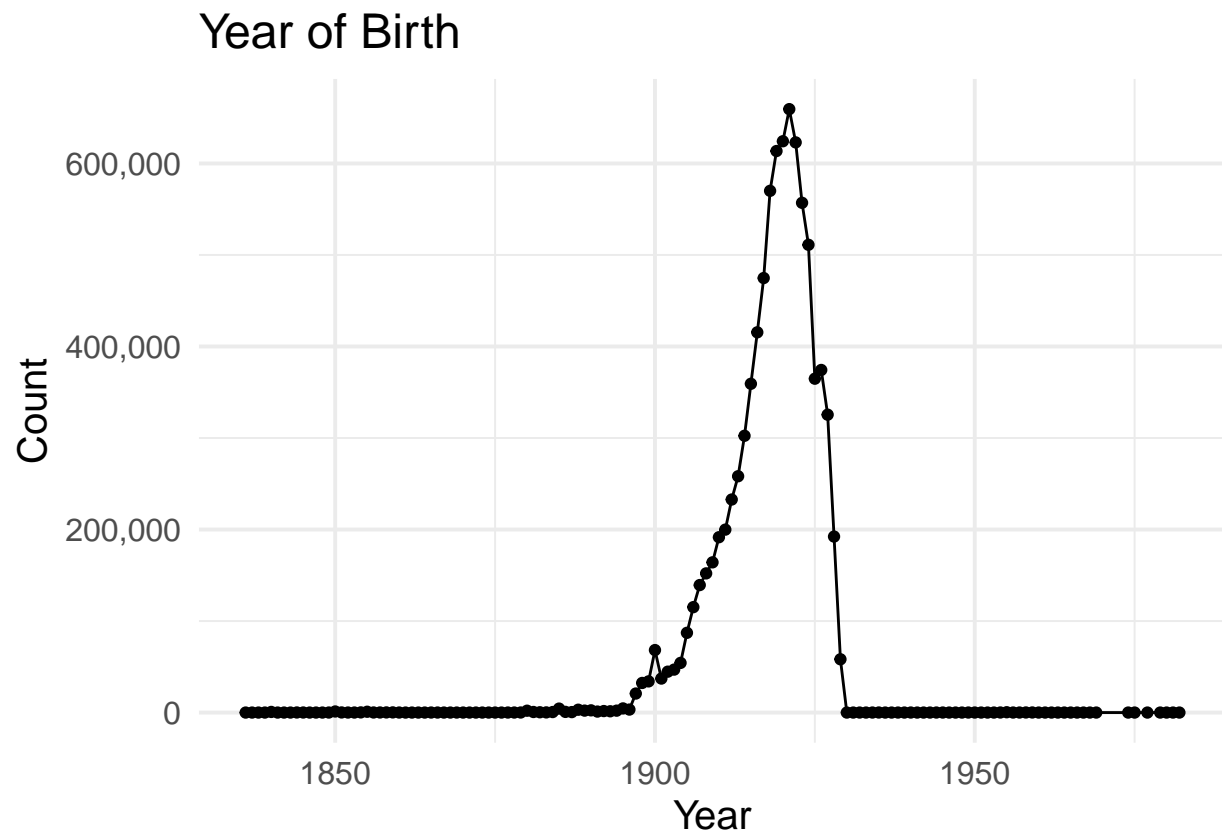
**Label:** Army Serial Number

**Description:** serial\_number is a character variable reporting a person's army serial number, as recorded in the World War II Army Enlistment Records. These numbers are intended to uniquely identify each enlistee, however, due to scanning or transcription errors some serial numbers in this dataset are non-unique.

# 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.



# fname

**Label:** First Name

**Description:** fname is a character variable reporting a person's first name, as recorded in the World War II Army Enlistment Records. This variable was cleaned from raw first names by removing non-alphabetical characters, whitespace, and titles (e.g., Dr.).

# mname

**Label:** Middle Name

**Description:** mname is a character variable reporting a person's middle initial or name, as recorded in the World War II Army Enlistment Records. This variable was cleaned from raw names by removing non-alphabetic characters. Most middle names consist of a single character.

# lname

**Label:** Last Name

**Description:** lname is a character variable reporting a person's last name, as recorded in the World War II Army Enlistment Records. This variable was cleaned and standardized to exclude non-alphabetic characters and whitespace (e.g., "O'Connor" is coded as "oconnor", and "Villa-Lobos" is coded as "villalobos").

# sex

**Label:** Sex

**Description:** sex is a numeric variable that reports a person's sex, as inferred by army branch in World War II Army Enlistment Records.

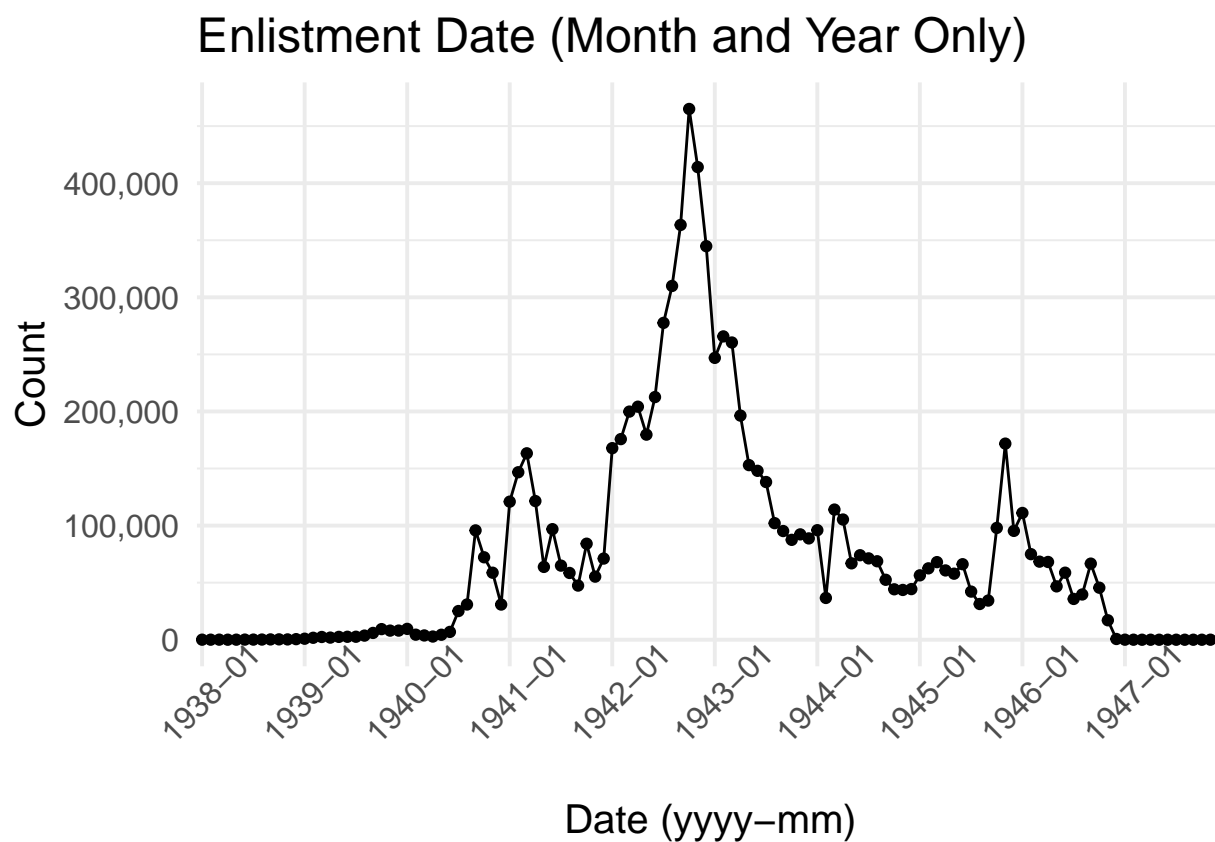
**Note:** Sex was not listed on enlistee punchcards. If an individual's numeric branch code or alphabetic branch code in the raw enlistment corresponds to the Women's Army Corps, we declare that person to be female. All other persons are assigned male. There is potential for error in this designation due to missing branch information or disagreements between army branch numeric codes, branch alphabetic codes (which we do not publish), and army component information. Thus, sex may not always correspond exactly to army branch and component information in the published data set.

sex	label	n	freq %
1	Men	8906784	98.50
2	Women	133056	1.47

# 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.

For a complete list of codes, please see: [https://usa.ipums.org/usa-action/variables/BPL#codes\\_section](https://usa.ipums.org/usa-action/variables/BPL#codes_section)

# residence\_\_state

**Label:** State of residence

**Description:** residence\_\_state is a character 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 US.

For a complete list of codes, please see: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=2046&dt=893&c\\_id=24996](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2046&dt=893&c_id=24996)

# 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. This variable is only available for enlistees residing in the US.

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)

# place\_\_of\_\_enlistment

**Label:** Place of enlistment

**Description:** place\_\_of\_\_enlistment is a character variable that reports the place a person was enlisted, as reported in the World War II Army Enlistment Records.

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

**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 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. Many enlistees with this code likely 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	2776711	30.7000000
3	Grade 9	773497	8.5600000
4	Grade 10	955260	10.6000000
5	Grade 11	843115	9.3300000
6	Grade 12	2376525	26.3000000
7	1 year of college	499514	5.5300000
8	2 years of college	315862	3.4900000
9	3 years of college	146253	1.6200000
10	4 years of college	279704	3.0900000
11	5+ years of college	73398	0.8120000
NA	NA	1	0.0000111

# grade\_\_code

**Label:** Army grade

**Description:** grade\_code is a numeric variable that records a person's grade (rank), as reported in the World War II Army Enlistment Records. Code list also is available at: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=2072&dt=893&c\\_id=24999](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2072&dt=893&c_id=24999)

grade__code	label	n	freq %
1	Master Sergeant	26283	0.291
2	First Sergeant	11421	0.126
3	Technical Sergeant	24471	0.271
4	Staff Sergeant	48140	0.533
5	Sergeant	71338	0.789
6	Corporal	89783	0.993
7	Private First Class	258778	2.860
8	Private	8215100	90.900
9	Aviation Cadet	77836	0.861
NA	NA	216690	2.400

# 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	3976527	44.000000
1	General Staff Corps	1699	0.018800
2	No branch assignment	2829637	31.300000
3	Philippine Scouts	27402	0.303000
4	Professor US Military Academy	2537	0.028100
5	Military Intelligence	7835	0.086700
6	Corps of Military Police	19502	0.216000
7	Specialist Reserve	101	0.001120
8	Detached Enlisted Men's List	13125	0.145000
9	Inactive Reserve	56127	0.621000
10	Infantry	231058	2.560000
18	Women's Army Corps	71592	0.792000
20	Air Corps	955300	10.600000
30	Field Artillery	109267	1.210000
38	Transportation Corps	20736	0.229000
40	Coast Artillery Corps or Army Mine Planter Service	83355	0.922000
50	Quartermaster Corps	90281	0.999000
70	Medical Department - For Enlisted Men of the Medical Department (includes Enlisted Men of the Medical Department on Duty with the Dental Corps) or Army Nurse Corps	55386	0.613000
71	Medical Corps - For Officers only	74	0.000819
72	Hospital Dietitians	829	0.009170
73	Physical Therapy Aides	112	0.001240
74	Dental Corps - For Officers only	198	0.002190
75	Veterinary Corps or Service - For Officers of the Veterinary Corps and for Enlisted Men of the Veterinary Service	873	0.009660
77	Medical Administrative Corps - For Officers only	669	0.007400
78	Adjutant General's Department	1312	0.014500
79	Pharmacy Corps	267	0.002950
80	Chemical Warfare Service	6094	0.067400
82	Corps of Engineers	46725	0.517000
87	Ordnance Department	57965	0.641000
90	Signal Corps	95326	1.050000
93	Cavalry	28088	0.311000
96	Chaplains	1612	0.017800
97	Finance Department	2439	0.027000
98	Inspector General's Department	613	0.006780
99	Judge Advocate General's Department	11907	0.132000
NA	NA	233270	2.580000



# 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. The codes and full descriptions are 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	label (abbreviated)	n	freq %
1	One year enlistment	217859	2.41
2	Enlistment for Alaska	28447	0.31
3	Enlistment for Hawaiian Department	473506	5.24
4	Enlistment for assignment to another corps area	23166	0.26
5	Enlistment for the duration of the War or other emergency	6281919	69.49
6	Enlistment for the Panama Canal Department	339363	3.75
7	Enlistment for the Philippine Department	120956	1.34
8	Enlistment for the Puerto Rican Department	9950	0.11
9	Three year enlistment	100269	1.11
NA	NA	1444405	15.98

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

race	label	n	freq %
100	White	7797528	86.3000
150	Puerto Rican	63911	0.7070
200	Black	834688	9.2300
300	Native American	19943	0.2210
400	Chinese	15759	0.1740
500	Japanese	28153	0.3110
600	Filipino	29263	0.3240
630	Native Hawaiian	4275	0.0473
991	Other	36416	0.4030
NA	NA	209904	2.3200

# 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	106768	1.18
1	Citizen	8723168	96.50
NA	NA	209904	2.32

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

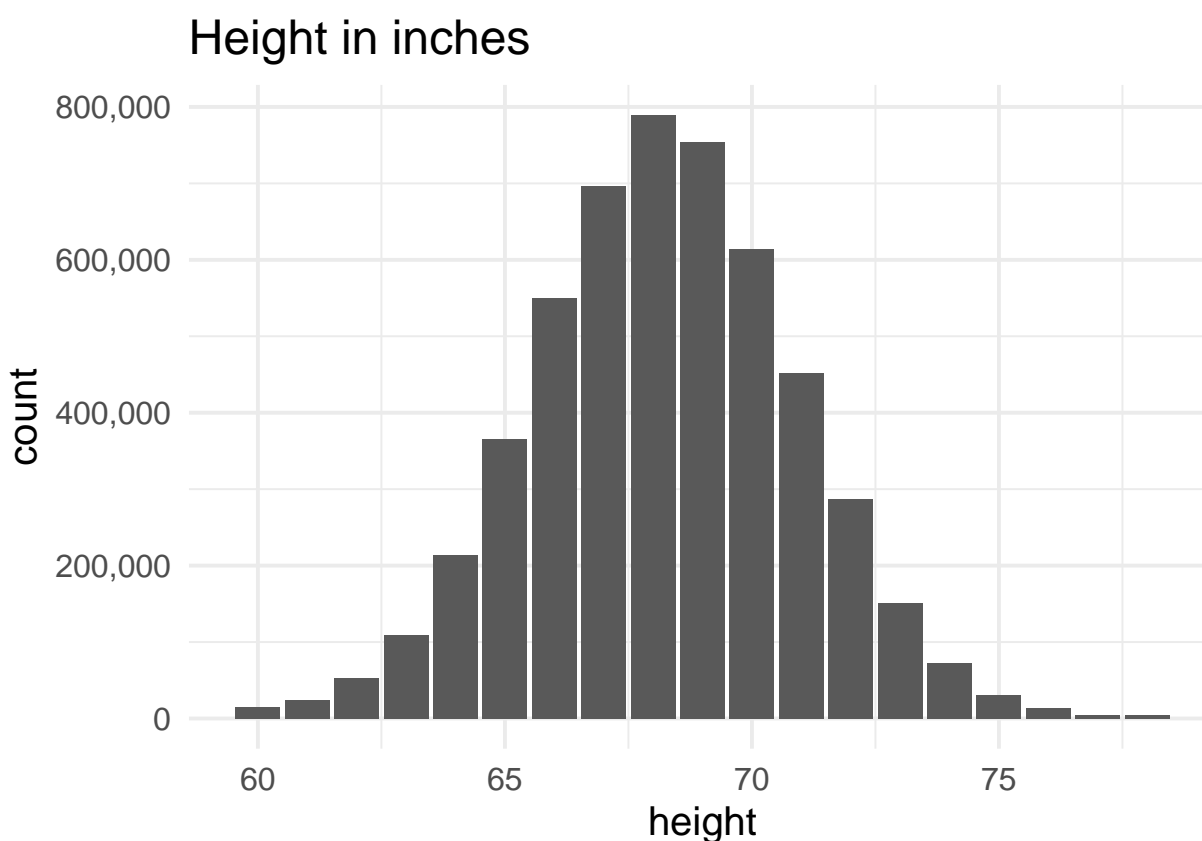
marital_status	label	n	freq %
1	Married	2048974	22.700
3	Separated	157700	1.740
4	Divorced	215371	2.380
5	Widowed	68436	0.757
6	Single	6471628	71.600
NA	NA	77731	0.860

# 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](#)).

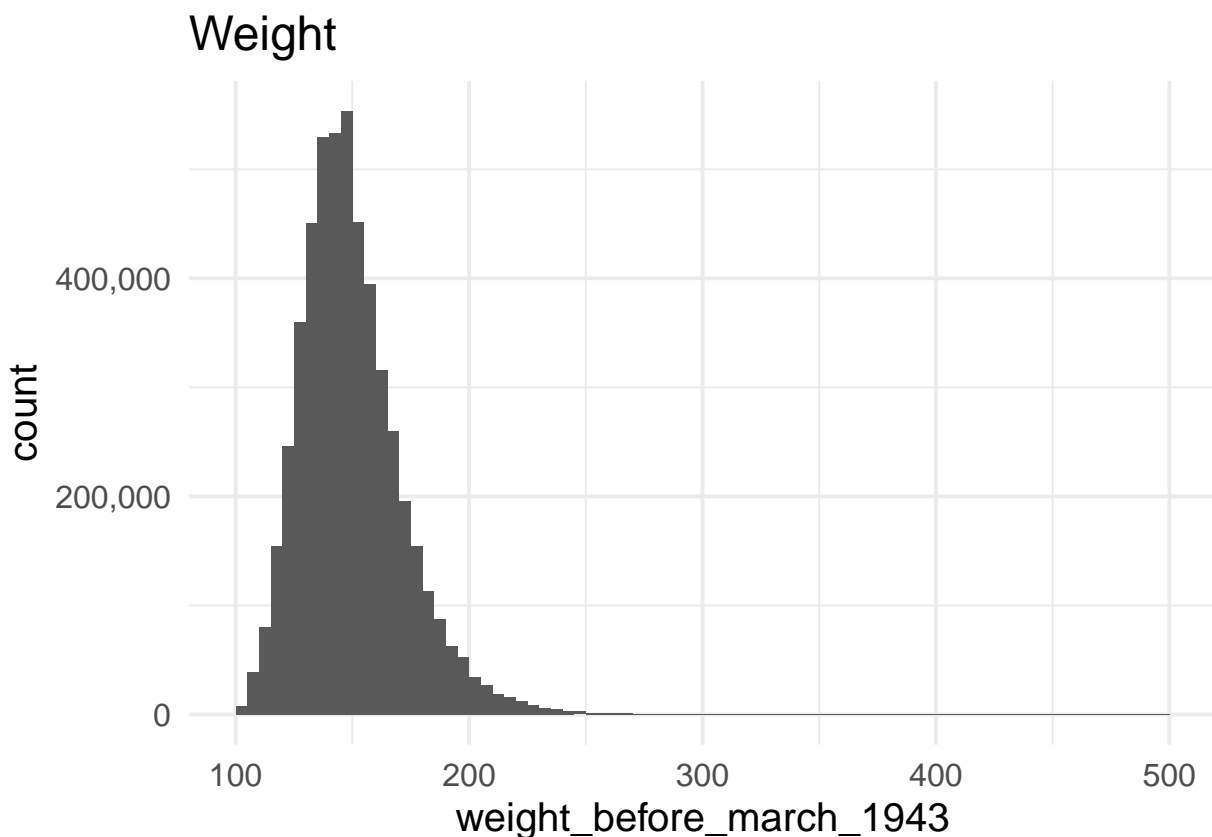
**Note:** Please be aware that instructions on how to use this field on enlistment punch cards varied over time. This field may encode army occupation or other information depending on time and site of enlistment. We have removed height data for men who enlisted in March of 1943 or later, as the distribution of this variable changes significantly at that point. However, it is possible that other erroneous data remains in this field. See the [National Archives' technical documentation on the Amy 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 the 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 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 1943. Refer to the [National Archives' technical documentation on the Army Serial Number Electronic File](#) for further information. For enlistees who enlisted starting in March 1943, we have created an additional variable called **weight\_or\_AGCT**.



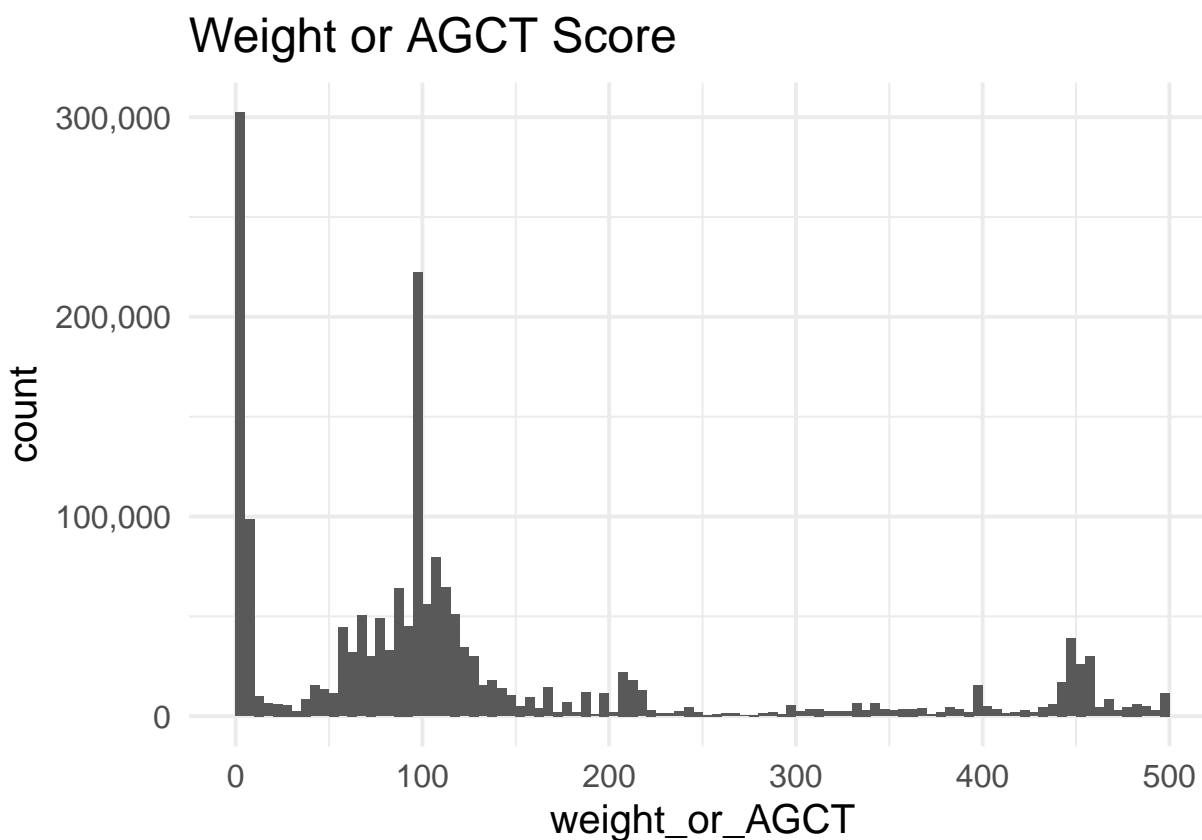
# weight\_or\_AGCT

**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 January 1943 or later. We have removed zeros and values above 500, which presumably exceed the maximum realistic body weights and AGCT scores, but otherwise leave the data in this field unfiltered and unsorted.

**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 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 January 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.

Codes also available at: [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	label	n	freq %
1	Regular Army	1294899	14.3000
2	Regular Army Reserve	2901	0.0321
3	Reserves	511487	5.6600
4	National Guard	251739	2.7800
5	Philippine Scouts	15377	0.1700
6	Army of the United States	679919	7.5200
7	Selectees	6063465	67.1000
9	Women's Army Corps	140580	1.5600
NA	NA	79473	0.8790

## 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. Codes also available at: [https://aad.archives.gov/aad/popup-codelist.jsp?cl\\_id=2077&dt=893&c\\_id=24981](https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2077&dt=893&c_id=24981)

source	label	n	freq %
0	Civil Life	7345292	81.30000
1	Enlisted Man, Regular Army, within 3 months of Discharge or former WAAC Auxiliary	26398	0.29200
2	National Guard in Federal Service, within 3 months of Discharge	397528	4.40000
3	Enlisted Man, Regular Army, after 3 months of Discharge	52222	0.57800
4	National Guard in Federal Service, after 3 months of Discharge	44337	0.49000
5	Enlisted Reserve or Medical Administrative Corps (MAC) Officer	181675	2.01000
6	Enlisted Man, Philippine Scout or recall to AD of an enlisted man who had been transferred to the ERC	389605	4.31000
7	National Guard	259526	2.87000
8	Returned to Military Control, Surrendered	6570	0.07270
9	Returned to Military Control Apprehended	1361	0.01510
11	Selective Service within 3 months of Discharge	32986	0.36500
12	Selective Service after 3 months of Discharge	312	0.00345
NA	NA	302028	3.34000