CenSoc Enlistment-DMF V2 Codebook*

Page	Variable	Label
2	id	Unique Identifier
3	sex	Sex
4	byear_DMF	Year of Birth (DMF)
5	bmonth_DMF	Month of Birth
6	dyear_DMF	Year of Death
7	dmonth_DMF	Month of Death
8	death_age_DMF	Age at Death (Years)
9	byear_enlistment	Year of Birth (Enlistment)
10	date_of_enlistment_enlistment	Date of Enlistment
11	bpl_enlistment	Birthplace
12	residence_state_enlistment	State of Residence at Enlistment
13	residence_county_enlistment	County of Residence at Enlistment
14	place_of_enlistment_enlistment	Place of Enlistment
15	education_enlistment	Education
16	grade_code_enlistment	Army Grade (Rank)
17	branch_code_enlistment	Army Branch
19	term_of_enlistment_enlistment	Term of Enlistment
20	race_enlistment	Race
21	citizenship_enlistment	Citizenship
22	civilian_occupation_enlistment	Civilian Occupation
23	marital_status_enlistment	Marital Status at Enlistment
24	height_enlistment	Height at Enlistment (Inches)
25	weight_before_march_1943_enlistment	Weight (Pounds)
26	weight_or_AGCT_enlistment	Weight (Pounds) or AGCT Score
27	component_enlistment_enlistment	Army Component
28	source_enlistment	Source of Army Personnel
29	HISTID	HISTID (1940 Census)

Summary: The CenSoc Enlistment-DMF files links 1,854,783 men from World War II era (circa 1938-1947) army enlistment records to the Social Security Death Master File (DMF). The suffix on each variable indicates whether the variable comes from DMF mortality 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 the conservative variant of the ABE method developed by Abramitzky, Boustan, and Eriksson (2012, 2014, 2017).

---

id

Label: CenSoc DMF-Enlistment Unique Identifier

Description: id is a generated 9-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	1854783	100

byear_DMF

Label: Year of Birth (DMF)

Description: byear_DMF is a numeric variable reporting a person's year of birth, as recorded in the DMF.

Note: The matching algorithm used allows for up to two years of flexibility on birth year, so birth year in the DMF may disagree with birth year in enlistment records. Birth years are discordant by one year in 7.8% of records, and discordant by two years in 4.3% of records. Both birth year variables are published in this dataset.
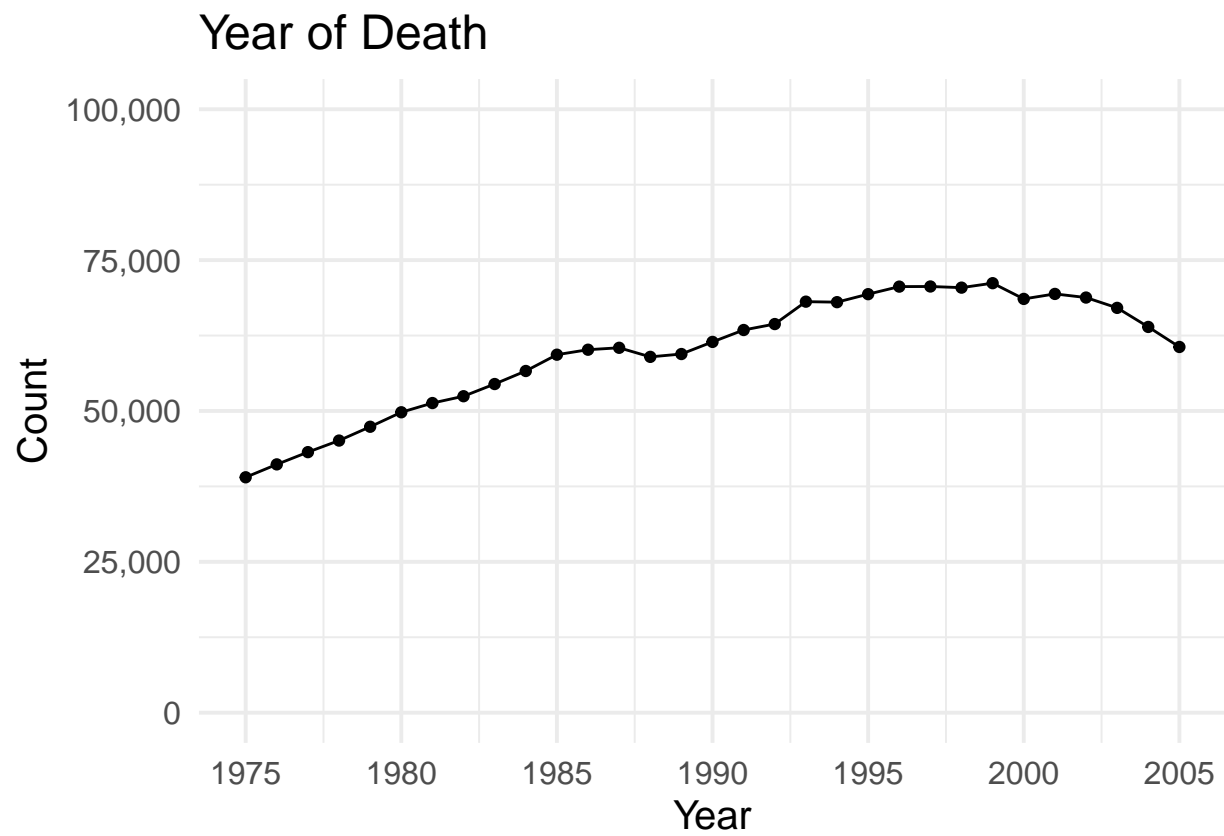
Year of Birth
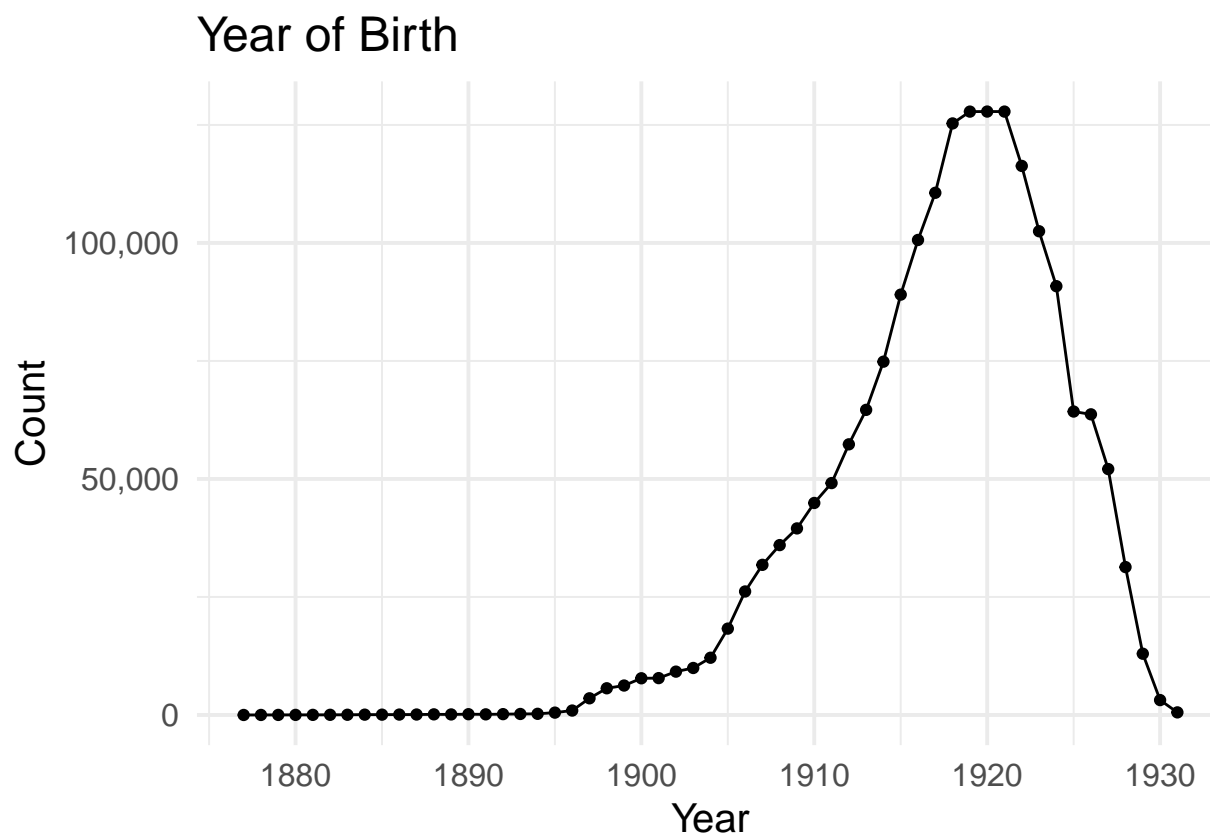
bmonth_DMF

Label: Birth Month

Description: bmonth_DMF reports a person's month of birth, as recorded in the DMF.

bmonth_DMF	label	n	freq %
1	January	158654	8.55
2	February	149172	8.04
3	March	162121	8.74
4	April	150978	8.14
5	May	152849	8.24
6	June	147083	7.93
7	July	155962	8.41
8	August	162586	8.77
9	September	160580	8.66
10	October	157162	8.47
11	November	146632	7.91
12	December	151004	8.14

dyear_DMF

Label: Death Year

Description: dyear_DMF reports a person's year of death, as recorded in the DMF.

Year of Death

dmonth_DMF

Label: Death Month

Description: dmonth_DMF reports a person's month of death, as recorded in the DMF.

dmonth_DMF	label	n	freq %
1	January	175705	9.47
2	February	155374	8.38
3	March	164676	8.88
4	April	153248	8.26
5	May	153066	8.25
6	June	144447	7.79
7	July	146541	7.90
8	August	145519	7.85
9	September	142155	7.66
10	October	153282	8.26
11	November	152184	8.20
12	December	168586	9.09

death_age_DMF

Label: Age at Death (Years)

Description: death_age_DMF reports a person's age at death in years, calculated using the birth and death information recorded in the DMF.

Age at Death
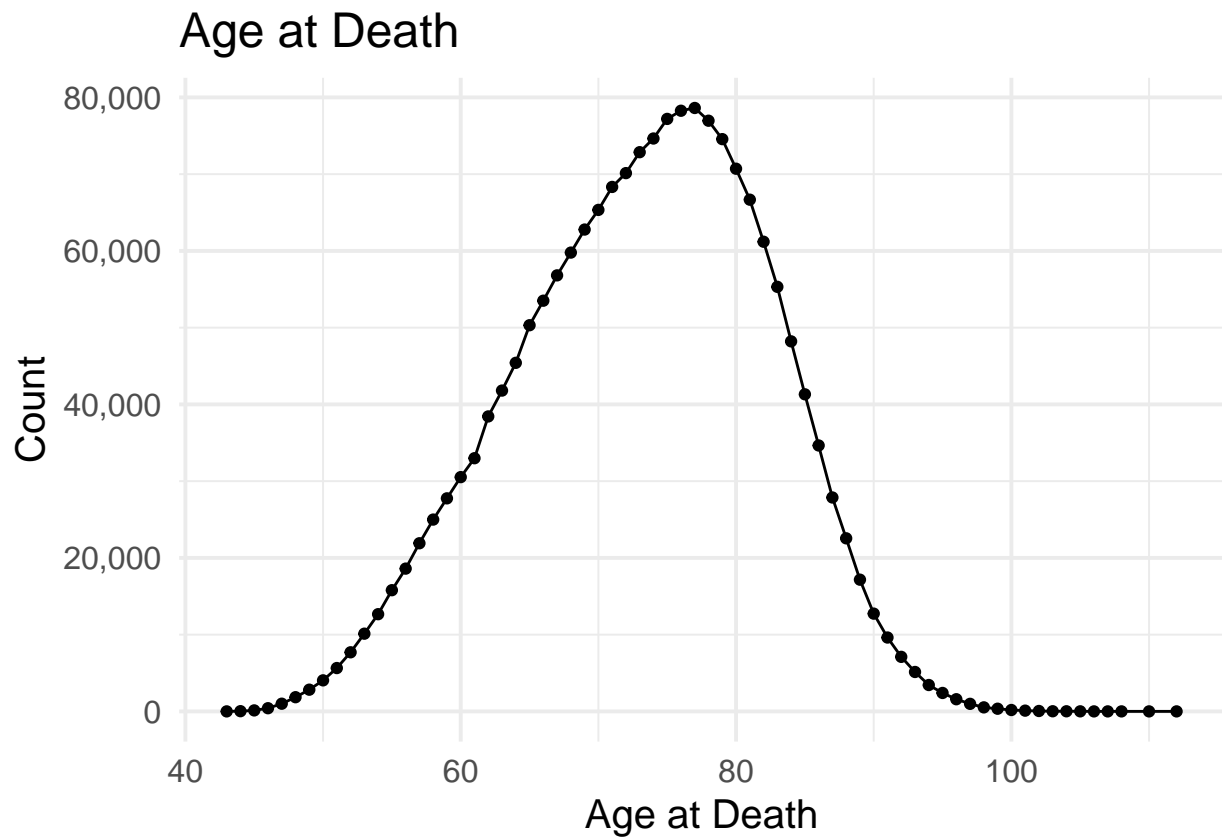
byear_enlistment

Label: Year of Birth (Enlistment)

Description: byear is a numeric variable reporting a person's year of birth, as recorded in the World War II Army Enlistment Records.
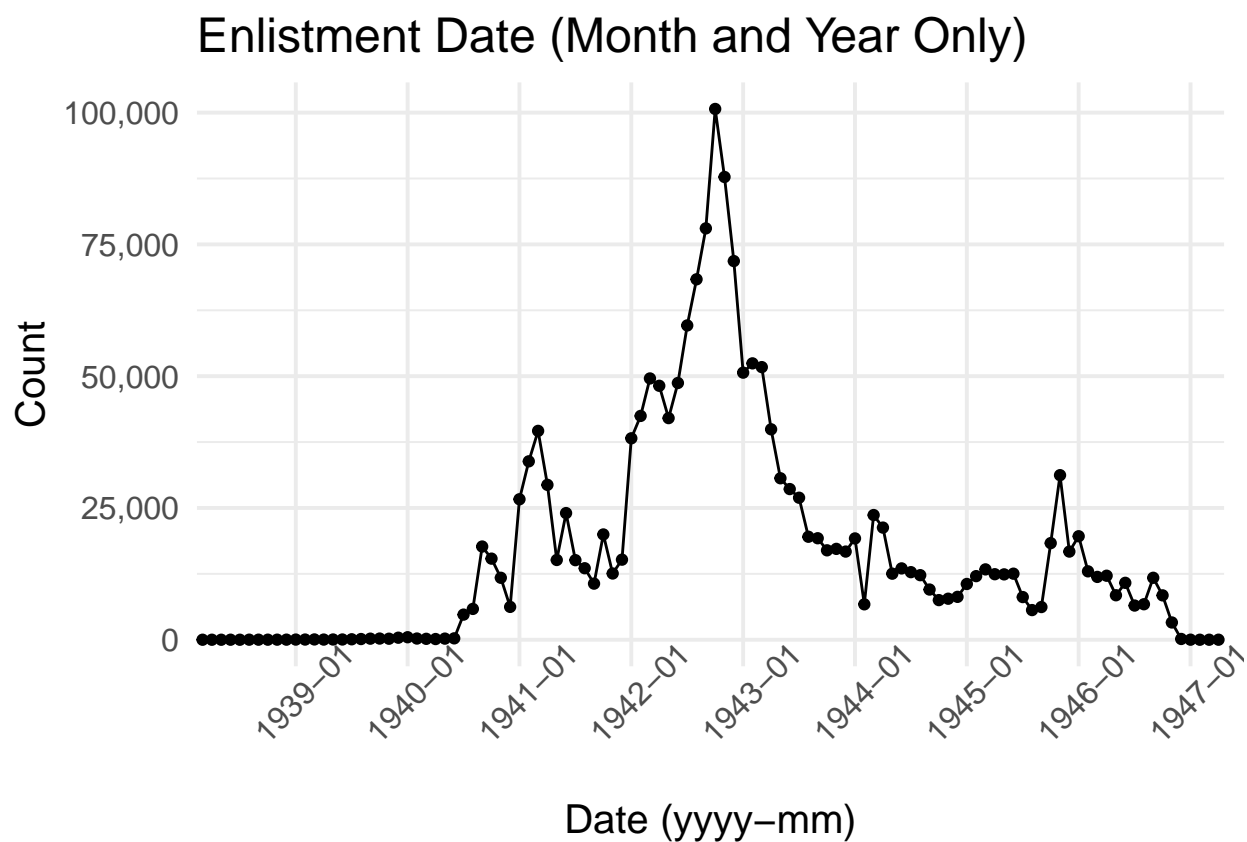
Note: The matching algorithm used allows for up to two years of flexibility on birth year, so birth year in the DMF may disagree with birth year in enlistment records. Birth years are discordant by one year in 7.8% of records, and discordant by two years in 4.3% of records. Both birth year variables are published in the dataset.

Year of Birth
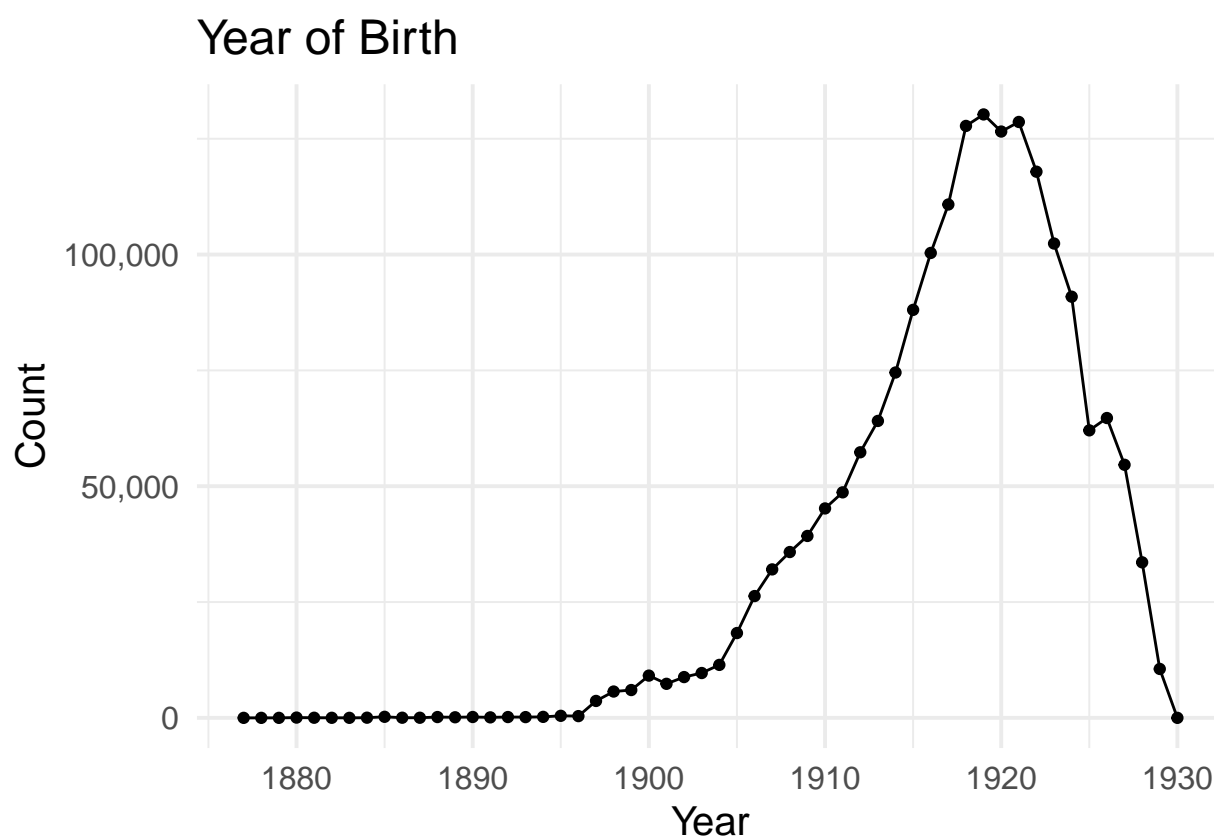
date_of_enlistment_enlistment

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.

Enlistment Date (Month and Year Only)

Count

100,000

75,000

50,000

25,000

0

1939–01   1940–01   1941–01   1942–01   1943–01   1944–01   1945–01   1946–01   1947–01

Date (yyyy–mm)

bpl_enlistment

Label: Place of Birth

Description: bpl_enlistment 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 codes, please see: https://usa.ipums.org/usa-action/variables/BPL#codes_section

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.

residence_county_enlistment

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 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 Documention.

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. County codes are numbers assigned to counties within states by alphabetical order, and are generally equivalent to FIPS county codes. However, they may not align exactly with modern FIPS codes due to county splits, merges, name changes, etc. since 1947.

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

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. 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_enlistment	label	n	freq %
1	Nursery school to grade 4	569153	30.69
3	Grade 9	165493	8.92
4	Grade 10	203082	10.95
5	Grade 11	172915	9.32
6	Grade 12	489908	26.41
7	1 year of college	95950	5.17
8	2 years of college	57602	3.11
9	3 years of college	29791	1.61
10	4 years of college	61326	3.31
11	5+ years of college	9563	0.52

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. Code list also is available at: https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2072&dt=893&c_id=24999

grade_code_enlistment	label	n	freq %
1	Master Sergeant	5522	0.30
2	First Sergeant	2352	0.13
3	Technical Sergeant	5178	0.28
4	Staff Sergeant	9874	0.53
5	Sergeant	14238	0.77
6	Corporal	17093	0.92
7	Private First Class	46374	2.50
8	Private	1729425	93.24
9	Aviation Cadet	7002	0.38
NA	NA	17725	0.96

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	910868	49.109
1	General Staff Corps	41	0.002
2	No branch assignment	562131	30.307
3	Philippine Scouts	4337	0.234
4	Professor US Military Academy	193	0.010
5	Military Intelligence	365	0.020
6	Corps of Military Police	3526	0.190
7	Specialist Reserve	14	0.001
8	Detached Enlisted Men's List	2596	0.140
9	Inactive Reserve	776	0.042
10	Infantry	44318	2.389
20	Air Corps	182691	9.850
30	Field Artillery	21663	1.168
38	Transportation Corps	3766	0.203
40	Coast Artillery Corps or Army Mine Planter Service	15919	0.858
50	Quartermaster Corps	14617	0.788
70	Medical Department or Army Nurse Corps	11074	0.597
71	Medical Corps - For Officers only	5	0.000
72	Hospital Dietitians	153	0.008
73	Physical Therapy Aides	13	0.001
74	Dental Corps - For Officers only	24	0.001
75	Veterinary Corps or Service	39	0.002
77	Medical Administrative Corps - For Officers only	1	0.000
78	Adjutant General's Department	255	0.014
79	Pharmacy Corps	15	0.001
80	Chemical Warfare Service	789	0.043
82	Corps of Engineers	9046	0.488
87	Ordnance Department	12032	0.649
90	Signal Corps	18654	1.006
93	Cavalry	5256	0.283
96	Chaplains	53	0.003
97	Finance Department	489	0.026
98	Inspector General's Department	45	0.002
99	Judge Advocate General's Department	1301	0.070
NA	NA	27718	1.494

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	n	freq %
1	One year enlistment	38657	2.08
2	Enlistment for Alaska	4284	0.23
3	Enlistment for Hawaiian Department	86679	4.67
4	Enlistment for assignment to another corps area	3933	0.21
5	Enlistment for the duration of the War or other emergency	1319875	71.16
6	Enlistment for the Panama Canal Department	57991	3.13
7	Enlistment for the Philippine Department	18976	1.02
8	Enlistment for the Puerto Rican Department	1403	0.08
9	Three year enlistment	17574	0.95
NA	NA	305411	16.47

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	1690268	91.13
150	Puerto Rican	5008	0.27
200	Black	114760	6.19
300	Native American	3859	0.21
400	Chinese	2708	0.15
500	Japanese	8698	0.47
600	Filipino	6512	0.35
630	Native Hawaiian	1145	0.06
991	Other	7062	0.38
NA	NA	14763	0.80

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	27111	1.46
1	Citizen	1812909	97.74
NA	NA	14763	0.80

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

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	443417	23.91
3	Separated	27810	1.50
4	Divorced	35823	1.93
5	Widowed	7200	0.39
6	Single	1336412	72.05
NA	NA	4121	0.22

height_enlistment

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 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 erroneous data remains in this field. See the National Archives' technical documentation on the Amy Serial Number Electronic File for further information.

Height in inches
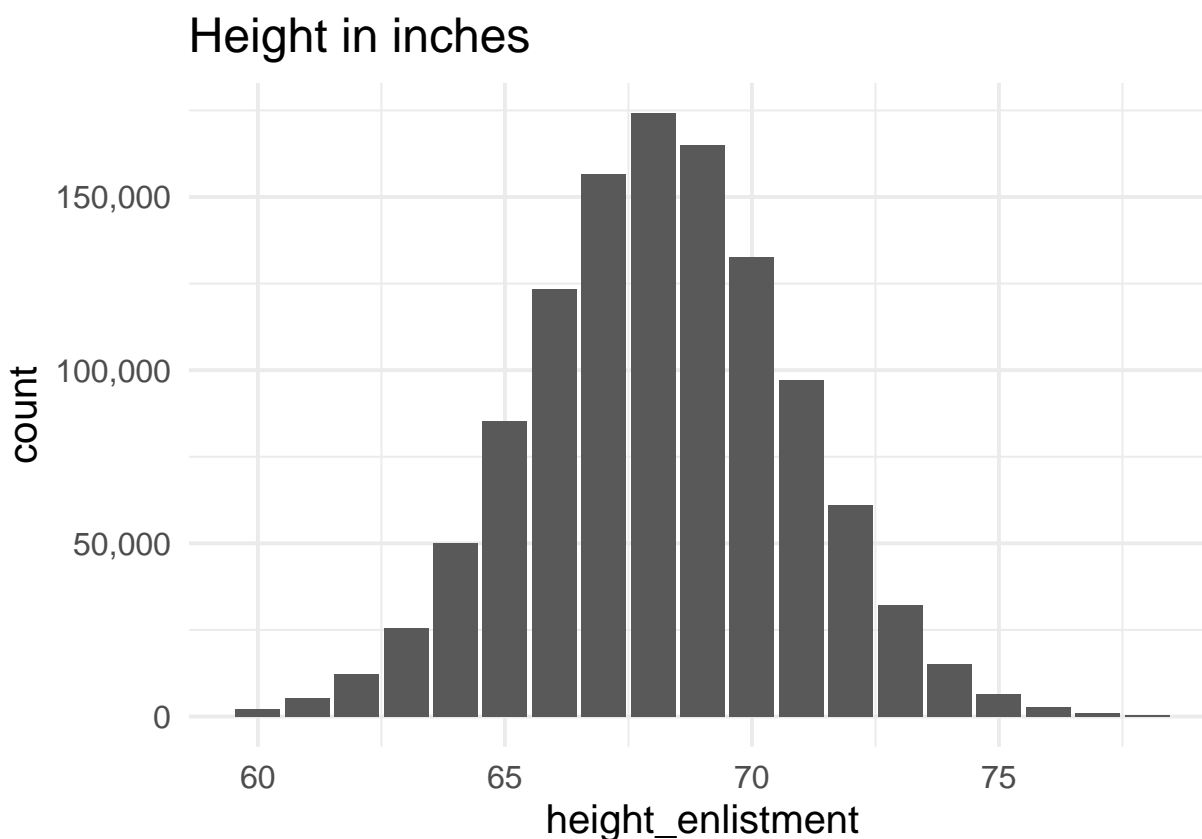
weight_before_march_1943_enlistment

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 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 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.

Weight or AGCT Score
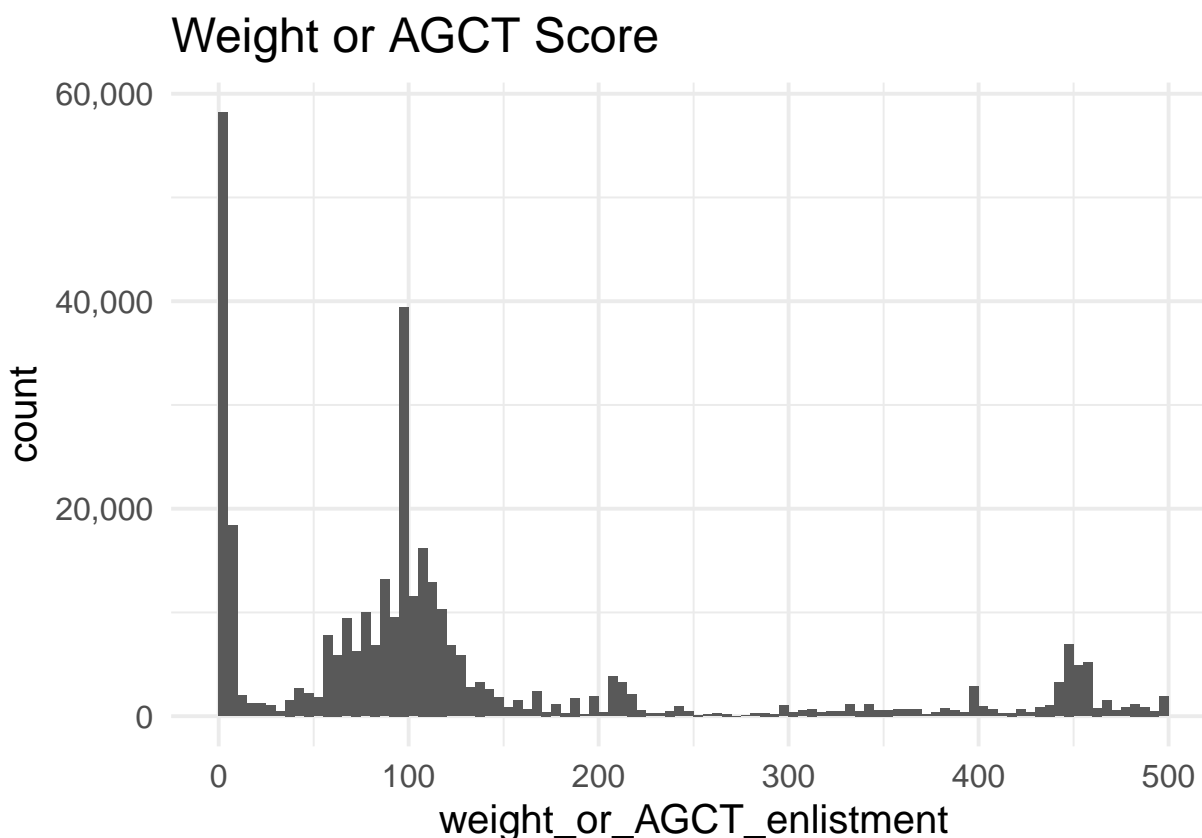
component_enlistment

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.

Codes are also available at: https://aad.archives.gov/aad/popup-codelist.jsp?cl__id=2078&dt=929&c__id=25014

component_enlistment	label	n	freq %
1	Regular Army	229988	12.40
2	Regular Army Reserve	87	0.00
3	Reserves	90610	4.89
4	National Guard	51211	2.76
5	Philippine Scouts	1551	0.08
6	Army of the United States	140685	7.58
7	Selectees	1336322	72.05
9	Women's Army Corps	617	0.03
NA	NA	3712	0.20

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. Codes are also available at: https://aad.archives.gov/aad/popup-codelist.jsp?cl_id=2077&dt=893&c_id=24981

source_enlistment	label	n	freq %
0	Civil Life	1566463	84.455
1	Enlisted Man, Regular Army, within 3 months of Discharge or former WAAC Auxiliary	4552	0.245
2	National Guard in Federal Service, within 3 months of Discharge	70727	3.813
3	Enlisted Man, Regular Army, after 3 months of Discharge	10628	0.573
4	National Guard in Federal Service, after 3 months of Discharge	8741	0.471
5	Enlisted Reserve or Medical Administrative Corps (MAC) Officer	29454	1.588
6	Enlisted Man, Philippine Scout or recall to AD of an enlisted man who had been transferred to the ERC	68866	3.713
7	National Guard	52944	2.854
8	Returned to Military Control, Surrendered	994	0.054
9	Returned to Military Control Apprehended	6	0.000
11	Selective Service within 3 months of Discharge	7214	0.389
12	Selective Service after 3 months of Discharge	28	0.002
NA	NA	34166	1.842

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	779389	42.02
NA	1075394	57.98