

Introduction to IPUMS

Complete Count and Linked US Census Data

Matt Nelson

Research Scientist, Institute for Social Research
and Data Innovation

nels5091@umn.edu

IPUMS.ORG

- Introduction to IPUMS
- Data Availability
- Complete Count Data 1790-1950
- Linked census data 1850-1940
- Questions

INTRODUCTION TO IPUMS

What is IPUMS?

- US and international census and survey data
- NIH & NSF funding to improve data accessibility
- Integrated documentation and standardized data to streamline analysis
- Available free of charge

What is Harmonization?

- Creating a single, consistent data series from datasets collected at different times
- Applying a coding scheme to group broadly comparable categories but retain additional detail when available
- Documenting these decisions and flagging potential comparability issues

Changes to detail in legal marital status in NHIS

Pre-2004 Categories

B = Not in Universe

1 = Married – spouse present

2 = Married – spouse not in household

3 = Married – spouse in household unknown

4 = Widowed

5 = Divorced

6 = Separated

7 = Never Married

2004-forward Categories

B = Not in Universe

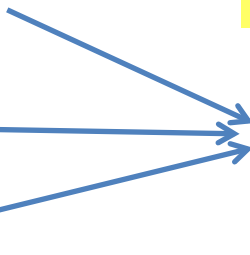
1 = Separated

2 = Divorced

3 = Married

4 = Single/Never Married

5 = Widowed



- The **first digit** of the code is consistent across all years
- The **second digit** of the code contains detail specific to pre-2004 data

Harmonized codes	
00	= Not in Universe
10	= Married
11	= Married – spouse present
12	= Married – spouse not in household
13	= Married – spouse in household unknown
20	= Widowed
30	= Divorced
40	= Separated
50	= Never Married

Benefits of Using IPUMS

- Harmonization
- Customized data extracts
- Integrated and supplemental documentation
- Online data analysis tools
- Additional data enhancements

Data Enhancements

- Family interrelationship variables
- Harmonized occupation and industry coding schemes
- Derived geographic identifiers

Benefits of Using IPUMS

- Harmonization
- Customized data extracts
- Integrated and supplemental documentation
- Online data analysis tools
- Additional data enhancements
- User support

User Support

- Video tutorials
- Sample exercises
- Data user forum <https://forum.ipums.org/>
- Email support ipums@umn.edu

Microdata

Summary data



U.S. Census and American Community Survey microdata from 1850 to the present.



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#).



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Tabular and GIS data from population, housing, and agricultural censuses around the world.



Historical and contemporary time use data from 1965 to the present.



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present).



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

U.S. data

International data



U.S. Census and American Community Survey microdata from 1850 to the present.



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#).



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Tabular and GIS data from population, housing, and agricultural censuses around the world.



Historical and contemporary time use data from 1965 to the present.



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present).



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

Census data

Survey data



U.S. Census and American Community Survey microdata from 1850 to the present.



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#).



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Tabular and GIS data from population, housing, and agricultural censuses around the world.



Historical and contemporary time use data from 1965 to the present.



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present).



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

Sub-state geographic detail



U.S. Census and American Community Survey microdata from 1850 to the present.



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#).



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Tabular and GIS data from population, housing, and agricultural censuses around the world.



Historical and contemporary time use data from 1965 to the present.



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present).



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

GIS data



U.S. Census and American Community Survey microdata from 1850 to the present.



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.



Health survey data for Africa and Asia, including harmonized data collections for [DHS](#) and [PMA](#).



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Tabular and GIS data from population, housing, and agricultural censuses around the world.



Historical and contemporary time use data from 1965 to the present.



Historical and contemporary U.S. health survey data from [NHIS](#) (1963-present) and [MEPS](#) (1996-present).



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

My expertise...



U.S. Census ~~and American Community Survey~~ microdata from 1850 to the ~~present~~.
1950



Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.



World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.



Health survey data for Africa and Asia, including harmonized data collections for DHS and PMA.



Tabular U.S. Census data and GIS boundary files from 1790 to the present.



Tabular and GIS data from population, housing, and agricultural censuses around the world.



Historical and contemporary time use data from 1965 to the present.



Historical and contemporary U.S. health survey data from NHIS (1963-present) and MEPS (1996-present).



Survey data on the science and engineering workforce in the U.S. from 1993 to the present.

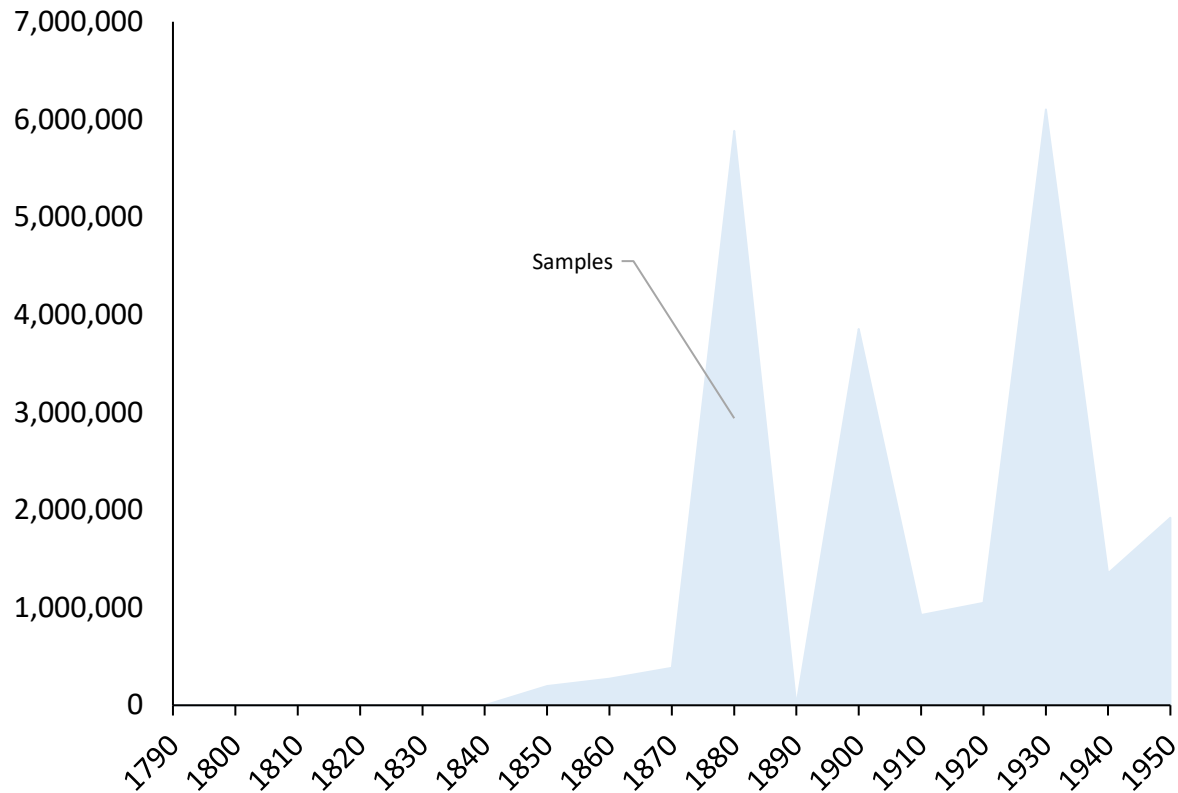
Data Availability

- Ancestry.com & FamilySearch
- 1790-1840 Households
- 1850-1950* Microdata
- Restricted Data 1790-1950*
 - ipumsres@umn.edu

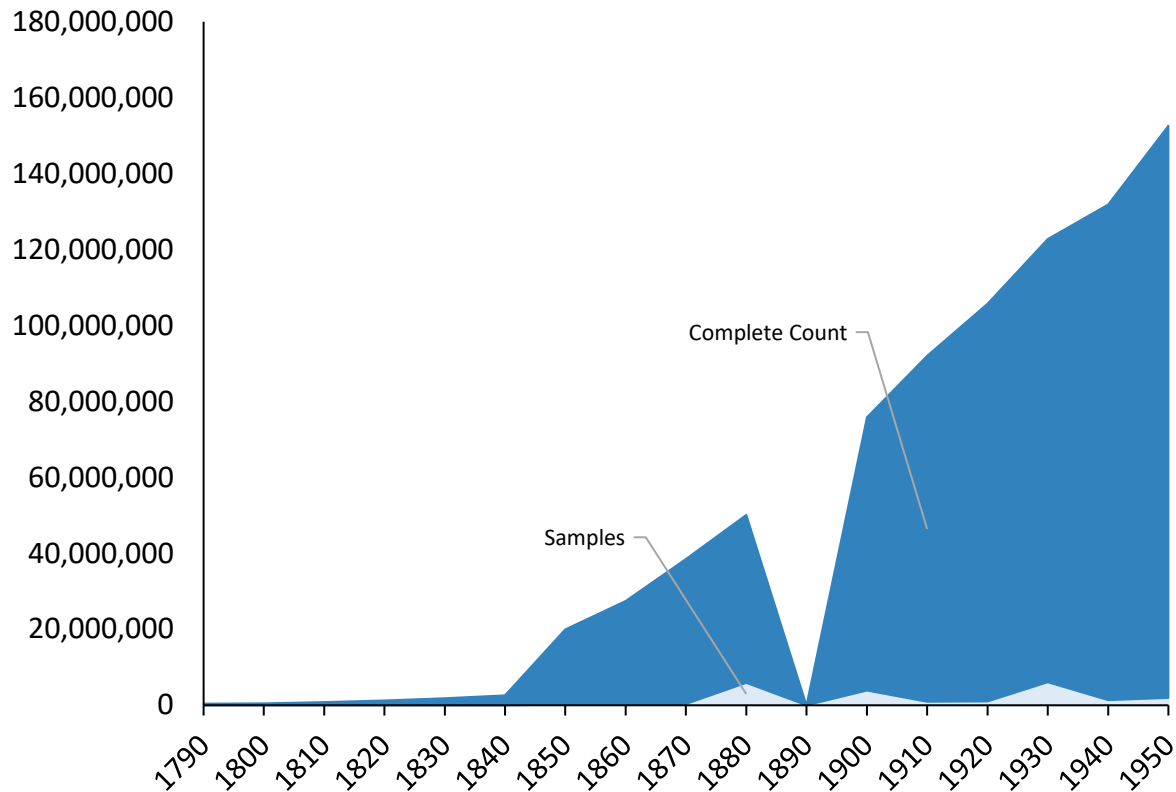
Important Takeaways

- HISTID
- Census Data Quality
- Examples of Research

Number of Records in IPUMS USA Data 1790-1950



Number of Records in IPUMS USA Data 1790-1950



"X" "V" and "F" indicates entries are to be used.
"O" indicates entries are to be omitted.

DEPARTMENT OF COMMERCE AND LABOR
THIRTEENTH CENSUS OF THE UNITED STATES

Smith River (part of) NAME OF INCORPORATED PLACE.

RELATION.	
Relationship of this sex	to the head of the family
4	
Head	(C)
Wife	
Son	
Head	
Wife	
Son	
Son	
Daughter	
Head	(C)
Wife	
Daughter	
Head	
Wife	
Son	
Daughter	
Head	
Wife	
Son	
Daughter	
Head	
Wife	
Son	
Head	
Daughter	
Daughter	
Head	
Wife	
Daughter	
Head	
Wife	
Son	
Daughter	
Head	
Sister	
Wife	
Sister	
Head	
Wife	
Son	
Son	
Son	
Head	
Wife	
Son	
Daughter	
Head	
Wife	
Head	(C)
Wife	

1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

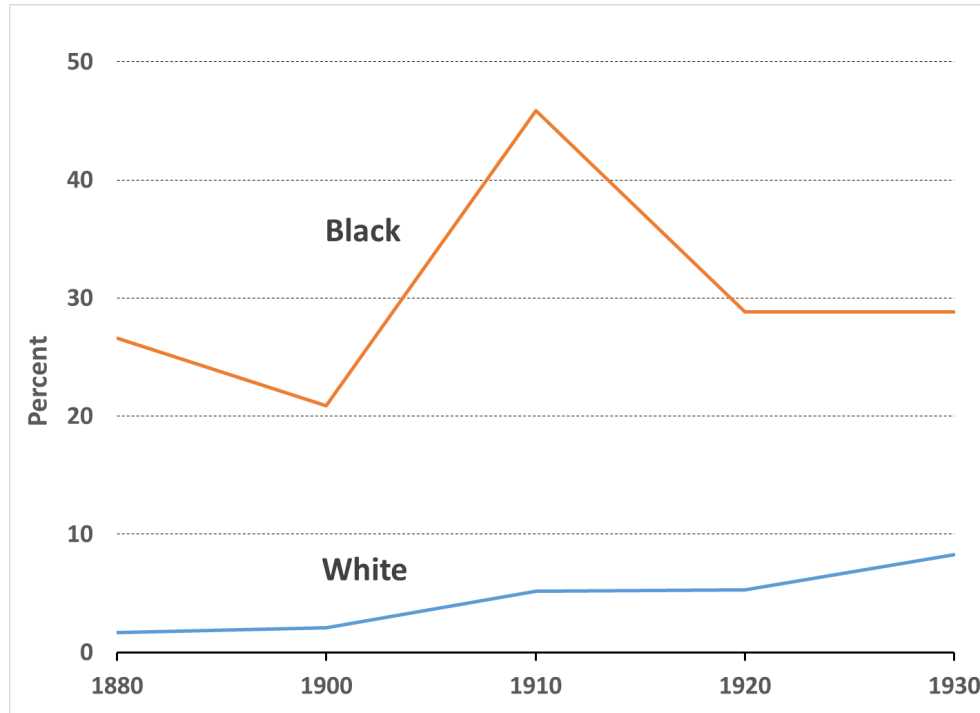
1910 Enumerator Instructions

154. Women doing farm work.—A woman working regularly at outdoor farm work, even though she works on the home farm for her husband, son, or other relative and does not receive money wages, should be returned in column 18 as a *farm laborer*. Distinguish, how-

1920 Enumerator Instructions

159. Women doing farm work.—For a woman who works *only occasionally*, or *only a short time each day* at outdoor farm or garden work, or in the dairy, or in caring for live stock or poultry, the return should be *none*; but for a woman who works *regularly*

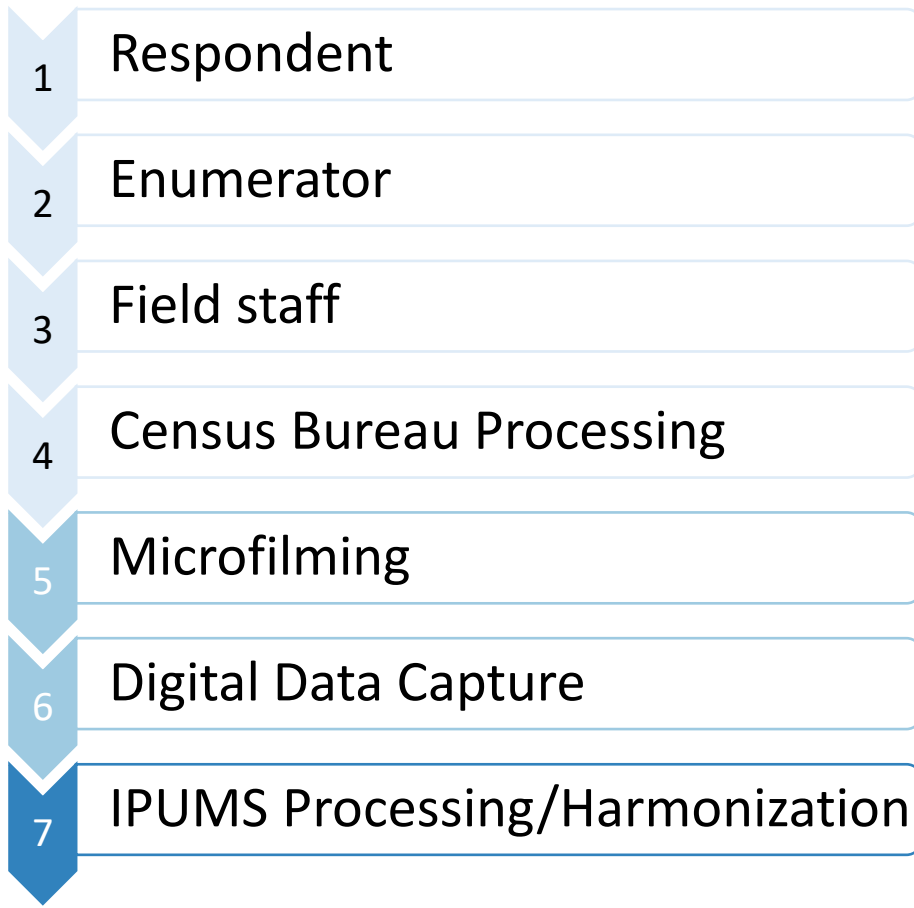
Occupational Responses for Married Women by Race 1880-1930



Occupational and Industrial Classification Summary Statistics 1900-1930

	Occupation 1950 Classification		Industrial 1950 Classification	
	Cases Uncoded	Strings Uncoded	Cases Uncoded	Strings Uncoded
1900 ¹	1.9%	56.8%	3.4%	63.4%
1910	2.3%	17.2%	2.2%	12.9%
1920	1.5%	16.3%	1.4%	12.4%
1930	1.6%	15.1%	1.4%	11.3%

¹ 1900 only collected occupation and no industry information.



Within Universe Agreement Rates between 1900-1930 Complete Count Data and Samples: Household

Variable Name	Variable Description	Universe	1920	1930
OWNERSHP	Home Ownership	Head	98.6%	98.7%
FARM	Farm Status	Head	98.1%	99.4%

Within Universe Agreement Rates between 1900-1930 Complete Count Data and Samples: Demographic

Variable Name	Variable Description	Universe	1920	1930
RELATE	Relationship to Head Status		98.1%	99.1%
SEX	Sex		99.3%	99.5%
AGE	Age		97.7%	97.0%
MARST	Marital Status	Age 12+	98.4%	99.2%

Within Universe Agreement Rates between 1900-1930 Complete Count Data and Samples: Race, Ethnicity, and Nativity

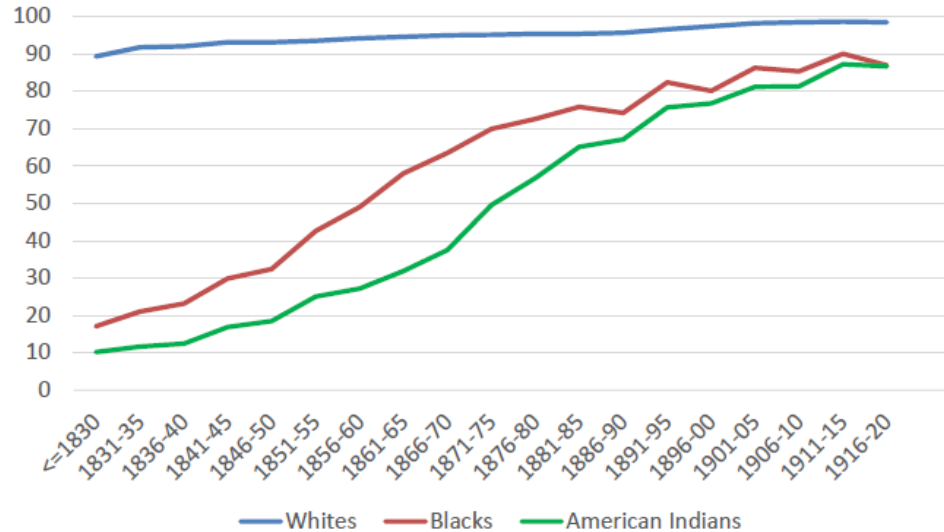
Variable Name	Variable Description	Universe	1920	1930
RACE	Race		99.4%	99.7%
BPL	Birthplace		95.5%	99.1%
FBPL	Father's Birthplace		96.4%	98.7%
MBPL	Mother's Birthplace		95.9%	98.6%
MTONGUE	Mother Tongue	Foreign-Born	93.6%	98.1%
YRIMMIG	Year of Immigration	Foreign-Born	91.7%	95.4%
CITIZEN	Citizenship Status	Foreign-Born ²	96.4%	97.7%
SPEAKENG	Speaks English	Age 10+	98.3%	99.3%

Within Universe Agreement Rates between 1900-1930 Complete Count Data and Samples: Education and Work

Variable Name	Variable Description	Universe	1920	1930
SCHOOL	School Attendance	³	98.8%	98.5%
LIT	Literacy	Age 10+	98.5%	98.9%
OCC1950	Occupation 1950 Classification		89.2%	92.5%
IND1950	Industry 1950 Classification		89.9%	91.1%
CLASSWKR	Class of Worker		96.6%	97.7%

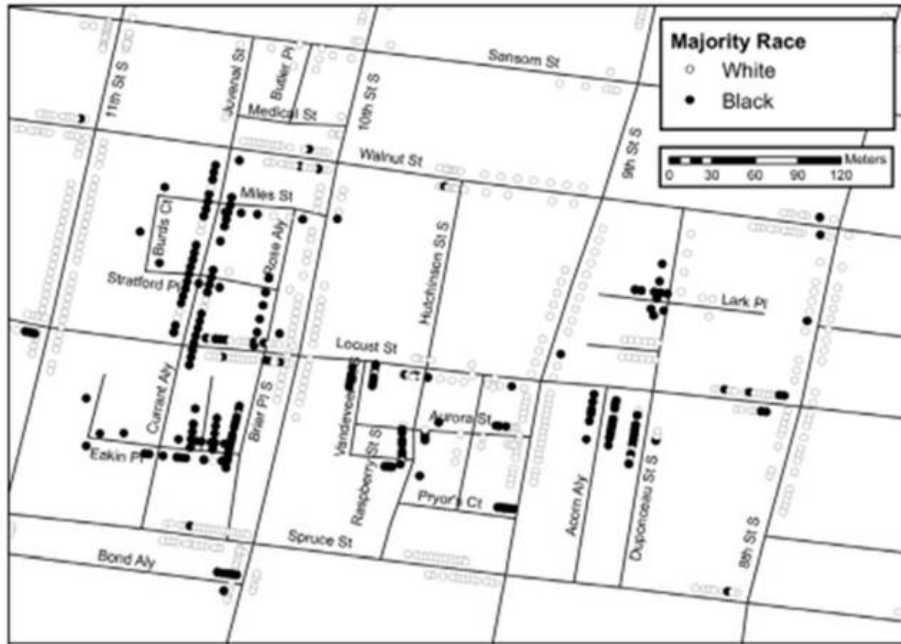
Small Subpopulations

Figure 1. Percent literate at the time of the 1900-1930 census
by race and birth cohort



Thornton & Young-DeMarco (2021). “Literacy Among American Indians: Levels and Trends from 1900 to 1930 and Across Birth Cohorts from 1830 to 1920,” *Sociological Studies of American Indians Working Paper 2021-0443*

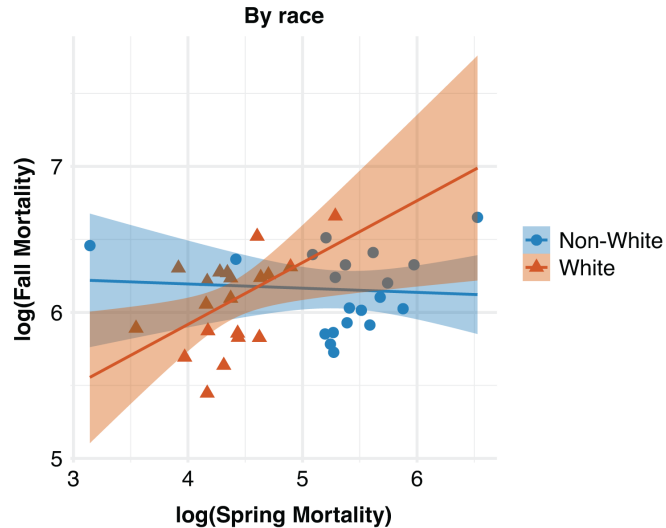
Geocoding



Logan & Bellman (2020).
“Before *The Philadelphia Negro*: Residential Segregation in a Nineteenth-Century Northern City,” *Social Science History* 40(4), 683-706.

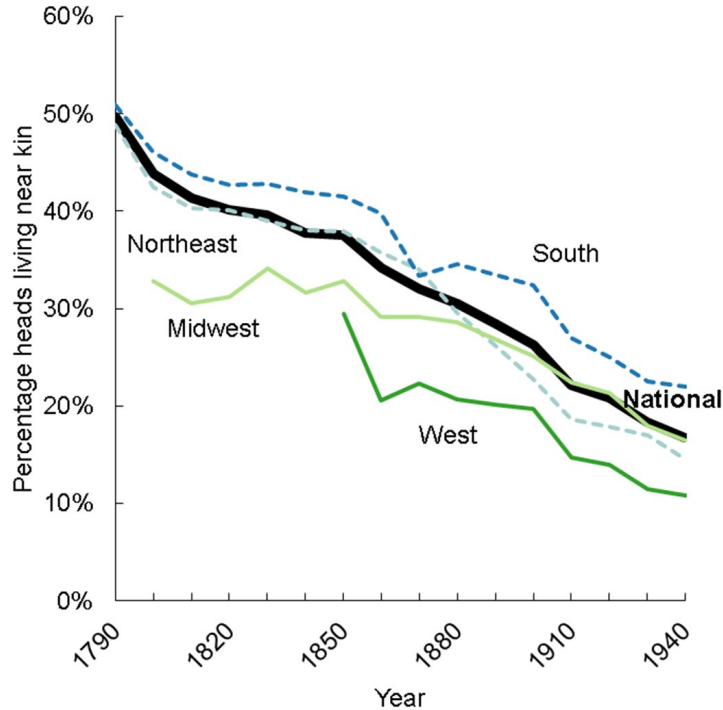
Population at Risk

Figure 4: Fall 1918 influenza and pneumonia mortality as a function of spring 1918 influenza and pneumonia mortality, logged, by race.



Eiermann, Wrigley-Field, Feigenbaum, Helgertz, Hernandez, & Boen (2022). “Racial Disparities in Mortality During the 1918 Influenza Pandemic in United States Cities,” *Demography* 59(5), 1953-1979.

Names



Nelson (2020). "The Decline of Patrilineal Kin Propinquity in the United States, 1790-1940," *Demographic Research* 43, 501-532.

Link Data to Census

Ordinary Least Squares Regression Models of
Baseline Z-Score of Cognitive Function

	Model 2 (adding HRS childhood SES measures to Model 1) b (95% CI)	Model 3 (adding 1940 Census childhood SES measures to Model 1) b (95% CI)
Father's education		
<8 years	Ref	Ref
8–11 years	0.04 (–0.05, 0.13)	0.07 (–0.02, 0.16)
12+ years	0.14** (0.03, 0.25)	0.12* (0.01, 0.23)
Mother's education		
<8 years	Ref	Ref
8–11 years	0.01 (–0.08, 0.10)	0.03 (–0.06, 0.13)
12+ years	–0.04 (–0.16, 0.09)	–0.02 (–0.15, 0.10)
Childhood finances		
Poor/bottom 35%	Ref	Ref
Average/middle 60%	–0.10 (–0.21, 0.00)	0.10* (0.02, 0.17)
Well-off/top 5%	–0.18* (–0.35, –0.01)	0.25** (0.07, 0.43)

Waren, Lee, & Osypuk (2022).
“The Validity and Reliability of
Retrospective Measures of
Childhood Socioeconomic
Status in the Health and
Retirement Study: Evidence
from the 1940 U.S. Census,”
The Journal of Gerontology
77(9), 1661-1673.

Multigenerational Longitudinal Panel (IPUMS MLP)

- Census to Census Links 1850-1940*
- Uses as many features as possible

Features

Last Name	First Name	Birth year	Last Name	First Name	Birth Year
Thomas	Arthur	1893	Thomas	Arthur	1893
Thomas	Arthur	1893	Thomas	Arthur	1893
Thomas	Arthur	1893	Thomas	Arthur	1893
Thomas	Arthur	1893	Thomas	Arthur	1893
Thomas	Arthur	1893	Thomas	Arthur	1893
Thomas	Arthur	1893	Thomas	Arthur	1893

Features

Last Name	First Name	Birth year	Father	Mother	Last Name	First Name	Birth Year	Father	Mother
Thomas	Arthur	1893	William	Margaret	Thomas	Arthur	1893	John	Rebecca
Thomas	Arthur	1893	William	Margaret	Thomas	Arthur	1893	Will	Margaret
Thomas	Arthur	1893	William	Margaret	Thomas	Arthur	1893	George	
Thomas	Arthur	1893	William	Margaret	Thomas	Arthur	1893		
Thomas	Arthur	1893	William	Margaret	Thomas	Arthur	1893	Geo H	Martha
Thomas	Arthur	1893	William	Margaret	Thomas	Arthur	1893		

Description

- Census to Census Links 1850-1940*
- Uses as many features as possible
- Two Step Linking Process

Step 1

Last Name	First Name	Source A Birth Year	Last Name	First Name	Source B Birth Year
Beech	Robert	1869	Beech	Robt	1868
Beech	Robert	1869	Beers	Robert	1869
Beech	Robert	1869	Beers	Robert	1870
Beech	Robert	1869	Beers	Robt	1870

Step I

Last Name	First Name	Source A Birth Year	Spouse	Street	Last Name	First Name	Source B Birth Year	Spouse	Street	Shared Neighbors
Beech	Robert	1869	Emma	Walnut	Beech	Robt	1868		Fourth	
Beech	Robert	1869	Emma	Walnut	Beers	Robert	1869	Emma	Walnut	1
Beech	Robert	1869	Emma	Walnut	Beers	Robert	1870	Sarah		
Beech	Robert	1869	Emma	Walnut	Beers	Robt	1870		Brownell	

Step 1

First Name	Last Name	Relationship to Head	Birth year	State of Birth	Years Married
Isaac B	Cody	Head	1869	Alabama	1
Martha J	Cody	Wife	1877	Alabama	1
Albert O	Cody	Child	1899	Alabama	

First Name	Last Name	Relationship to Head	Birth year	State of Birth	Years Married
Albert O	Cody	Child	1900	Alabama	

Current Improvements

- Linking Women in Step I
- 20/30 Year Links
- NUMIDENT Social Security Death Records

How to Access MLP Links

- Extract System for Full Count Data and MLP Links
- Static Page with Crosswalks using *HISTID*, Extract System for Full Count Data



Historical Identification Key (*HIK*)

- If the MLP linking process identifies new or different links for a currently existing *HIK*, that *HIK* will be discarded and assigned a new *HIK*.
- For this reason, we strongly recommend you continue to download *HISTID*, which is a stable unique within census id that will never change.

Case Selection Limitations

- You can include other criteria for case selection, but...
- **Case selection only includes “AND” operator and not “OR” operator--this can exclude legitimate links**
- If you case select *STATEICP = 33*, *LINK1900 = 1*, *LINK1910 = 1*

Case Selection Limitations

The case in Minnesota in 1900 is selected but not the linked record in 1910 because that case is in South Dakota

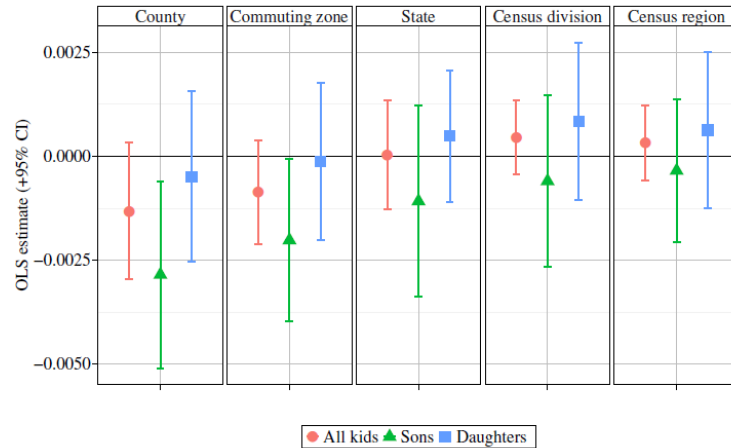
YEAR	SERIAL	PERNUM	AGE	HIK	STATEICP	Selected
1900	1632424	2	17	---32dIEO3V1J1Qgbpu1q	33	YES
1910	5434844	2	27	---32dIEO3V1J1Qgbpu1q	37	NO

Data format

YEAR	SERIAL	PERNUM	AGE	HIK	HISTID
1900	1632424	2	17	---32dIEO3V1J1Qgbpu1q	5527E906-AD5F-4E34-9006-A928DF1BA69F
1910	5434844	2	27	---32dIEO3V1J1Qgbpu1q	231B4F7A-7957-45A5-A722-F3B3CB80E21C

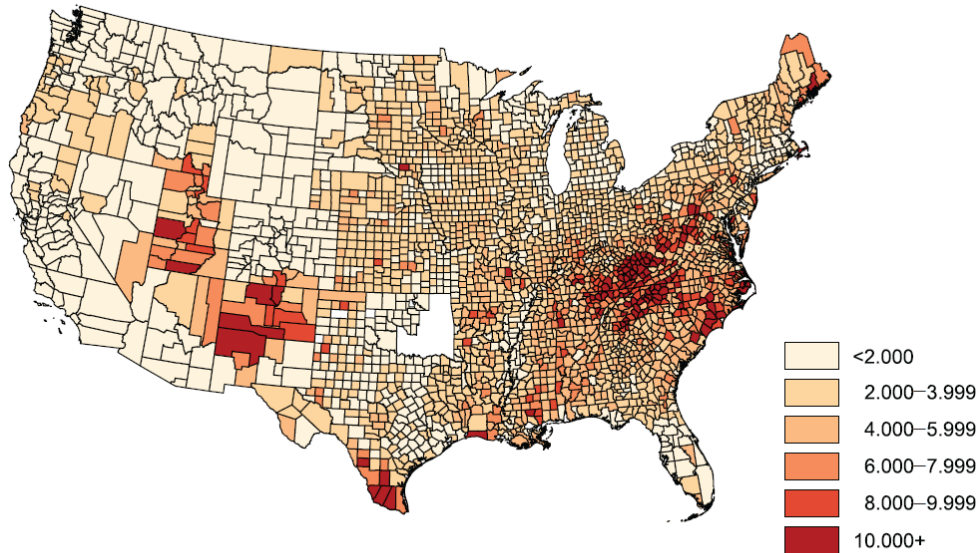
Intergenerational Effects

Figure 7: The impact of father's later Social Security coverage on children's propensity to die outside their father's 1930 place of residence



Fetter, Lockwood, & Mohnen (2022). “Long-Run Intergenerational Effects of Social Security”

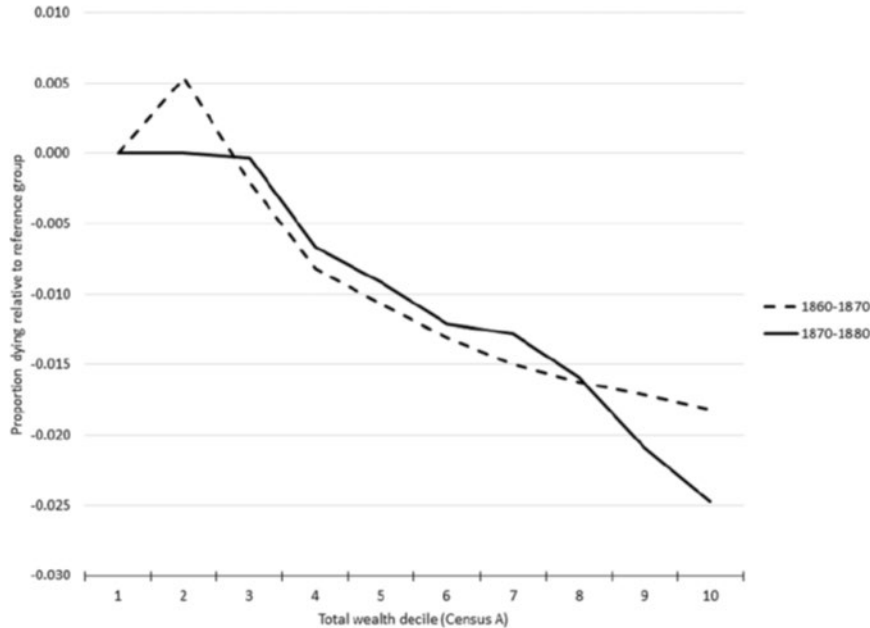
Fertility Outcomes



Hacker, Helgertz, Nelson, & Roberts (2021). “The Influence of Kin Proximity on the Reproductive Success of American Couples, 1900-1910,” *Demography* 58(6), 2337-2364.

Fig. 2 Average surname kin density per 1,000 adults by county, 1900 census

Mortality Outcomes



Hacker, Dribe, & Helgertz (2023). “Wealth and Child Mortality in the Nineteenth-Century United States: Evidence from Three Panels of American Couples 1850-1880,” *Social Science History* 47, 333-366.

Figure 3. Child mortality by parents' total wealth decile.
Source: See Table 2.

