

IPUMS Full-Count Microdatasets of Slave Inhabitants and Slaveholders in the United States in 1850 and 1860: Construction and Analysis

J. David Hacker, Lap Huynh, Matt A. Nelson, Brayden Rothe, and Matthew Sobek

This article describes the development of IPUMS full-count microdatasets of the censuses of slave inhabitants of the United States in 1850 and 1860. These datasets are a result of two collaborations. The 1850 slave dataset resulted from a collaboration between the Church of Jesus Christ of Latter-day Saints, whose volunteers transcribed the original manuscript forms, and IPUMS, which enhanced the raw data with editing, standardized coding procedures, constructed variables, and documentation. Constructed variables include linking slave holders named in the slave censuses to their records in the IPUMS full-count dataset of the free population in 1850. The 1860 dataset was the result of a similar collaboration between the genealogical company Ancestry.com and IPUMS. We discuss the features of these datasets, their limitations, and suggest possible research uses.

* This research was supported in part by funding from the Minnesota Population Center (P2C HD041023) and by grants from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (R01-HD060676-01 and R01-HD082120-01). Many individuals at the Minnesota Population Center contributed to the project, but we wish to especially acknowledge the work of Ronald Goeken, who contributed a significant amount of effort to all aspects of the project over a five-year period. Bigyan Khanal provided some initial programming. Some of the information in this paper was originally presented as Ronald Goeken, J. David Hacker, Lap Huynh, and Bigyan Khanal, "An Inquiry into Some Points of Record Linkage: Linking 19th Century U.S. Decennial Population Records to the Mortality and Slave Schedules," presented at "Putting the Pieces Together: Promise, Programs and Pitfalls in Linking Historical and Contemporary Records," at the Center for Economic History, Northwestern University, May 17-19, 2019.

IPUMS Full-Count Microdatasets of Slave Inhabitants and Slaveholders in the United States in 1850 and 1860: Dataset Construction and Analysis

This article describes new public-use IPUMS full-count microdatasets of the 1850 and 1860 slave censuses of the United States. The 1850 dataset was the result of a collaboration between the Church of Jesus Christ of Latter-Day Saints (LDS), whose volunteers performed the data entry, and the IPUMS Project at the University of Minnesota, which organized and cleaned the raw data, coded geographic and other string variables, linked slaveholders listed on the slave schedules to their records in the free population schedules, developed comprehensive on-line documentation, and distributes the data. The dataset contains individual information on each of the nation's 3.2 million slaves. The 1860 dataset was the result of a similar collaboration between IPUMS and the private genealogy company Ancestry.com and contains individual-level data on the nation's 3.9 million slaves in 1860. Funding for both projects was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

In addition to describing the construction, features, and limitations of the two datasets, this article suggests possible research projects that can be pursued using the data. It illustrates one possible use of the data in a preliminary analysis of slaveholding by age in 1860. Public-use versions of both datasets are now available for downloading at the IPUMS website (<http://ipums.org>). Restricted-use datasets, which include slaveholders' names, can be obtained by researchers with compelling research questions requiring slaveholder names and willing to agree to the terms of a confidentiality statement.

Background

At the time of its ratification in 1789, the United States Constitution required the federal government to conduct a census of the population every ten years to determine each state's representation in Congress. "Free Persons" and "all other Persons"—the latter a veiled reference to the

nation's enslaved inhabitants—needed to be counted separately, a result of the notorious “three-fifths” compromise that considered slaves as three-fifths of a person for the purposes of apportionment.¹ From the outset, the census collected more information than required by the Constitution. In 1790, the first census recorded the name of each household head and counted the number of free White males under 16 years of age, free White males aged 16 years and older, free White females, other free persons (free Blacks), and slaves in the household. The amount of information collected for both the free and slave populations increased over time but was always more detailed for the free population. In 1840, for example, the census counted the number of free White males and females in each household in 13 different age categories, while the number of enslaved Black males and females was counted in six age categories. The 1840 census also recorded the number of free White persons who could not read and write in the household but did not record the number of illiterate free Blacks or slaves.

The 1850 census introduced several significant improvements, the most important of which was to shift the census from a household survey with all information for household members summarized on one row of the enumeration form to an individual-level survey with a row dedicated to each person. Rather than counting the number of individuals in the household in different age categories, enumerators recorded the exact age of each person. Each free person's name, sex, race, marital status, place of birth, school attendance and literacy were also collected. The enslaved population was enumerated on a separate schedule with much less information. Although age, sex, color, and disability were recorded for each slave, names were not recorded, and no information was collected on occupation, marital status, birthplace, literacy or school attendance.

¹ It was an anti-slavery northerner, James Wilson of Pennsylvania, who proposed the compromise. Most southern delegates wanted slaves to count fully (while denying slaves the right to vote), while northern delegates with anti-slavery sentiments thought slaves should not be counted. Waldstreicher, D. 2010. *Slavery's constitution: from revolution to ratification*: Hill and Wang. Section 2 of the Fourteenth Amendment to the Constitution, adopted in 1868, repealed the three-fifths compromise.

The lack of information collected on the slave schedules is regrettable, especially given initial proposals. In 1849, an advisory board of some of nation's leading statisticians recommended that the slave schedule collect the name, age, sex, color, and place of birth of each slave; whether the slave was deaf, dumb, blind, insane, idiotic, or a fugitive; the number of children each enslaved woman had borne and the number of those children still surviving; and the name of each slave's owner. This information would have allowed the study of slave mortality, fertility, and family structure in much more detail than is currently possible. Although forms were printed, the proposed schedule ran into trouble in Congress, which was debating sensitive sectional issues regarding slavery, including a Fugitive Slave Act and the possible extension of slavery into California and new territories acquired in the Mexican American War. Southern Congressmen, fearing the possibility that information collected by the census could be used to attack their "peculiar institution," successfully reduced the number of questions on the slave schedule (Anderson 1988:37-41). The final slave schedule recorded only the age, sex, and color of each slave, whether the slave was currently a fugitive, deaf, blind, insane, or idiotic, and the slaveholder's name. (Slaveholders were typically owners, but sometimes individuals who rented, supervised, or held slaves in trust or as guardians for other owners were listed instead.) Despite the regrettable loss of potential data, the limited slave schedule approved by Congress resulted in the collection of better information on the slave population than obtained in prior censuses. These data, aggregated and analyzed by the Census Office in publications following both censuses (United States 1853, 1854, 1862, 1864), have proven useful for the study of slaveholding in the years leading up to the American Civil War (1861-1865) and the demographic growth, age structure, and distribution of the slave population (Hacker 2020).

Microdata on the slave population have much greater research potential than aggregated data. Researchers, for example, can use microdata to create custom tabulations and construct custom variables. Researchers can construct holding-level variables (e.g., the child to woman ratio or adult sex

ratio) and conduct multilevel analyses combining available and custom variables at the state, county, holding, owner, and individual slave levels.

The IPUMS full-count microdatasets described in this article are not the first IPUMS microdatasets based on the slave censuses. In 2004, IPUMS released microdata samples of the 1850 and 1860 slave censuses (Menard et al. 2004). In an earlier publication, Alexander et al. (2003) described the original IPUMS 1860 slave sample, including its design and method of construction, availability of variables, and potential research uses. The datasets were 5% random samples of slaves in most counties and 100% of slaves in a small non-random selection of counties. The latter group totaled 539,509 slaves in 13 southern states. Although enumerators did not identify slave dwellings, households, or combined holdings of each owner with unique serial numbers as they did for households and families on the free population schedules, the project identified slaves in the same “slaveholdings” using owners’ names, which were recorded on the same line as the first slave in the holding (and therefore assumed to be duplicated for all subsequent slaves in the holding). We use these same procedures in the construction of the IPUMS full-count slave datasets.

IPUMS Full-Count Cross Sectional Datasets of the Slave Population

The IPUMS full-count datasets should prove to be more useful for research than the low-density IPUMS samples. As the name implies, the full-count datasets include individual information on the complete universe of slaves in the United and therefore include more cases for analysis. Where the IPUMS sample of the 1860 slave census contained records for 195,027 slaves, for example, the full-count dataset includes individual-level information on 3,936,602 slaves.² Full-count datasets are especially

² There are small discrepancies between the full-count dataset and the official published numbers in both census years that we were unable to resolve but which should have negligible for most research. In 1860, the Census Office counted 3,952,838 slaves, indicating that the dataset is missing 16,236 slaves or about 0.41% of the total.

important for the study of small areas where low sample densities might represent a problem. There are many other benefits. As we describe below, full-count datasets are especially well-suited for linking slaveholders named in the slave censuses to their households in the free population census using machine-learning algorithms.

Our first step in constructing the full-count datasets was to assemble the data entered by LDS and Ancestry.com, sort the raw data by state, county, and place, and compare the returns to published totals made by the 1850 and 1860 Census Offices. Our initial audit uncovered inconsistent geographic information and several areas with missing records, which we corrected through in-house data entry at the University of Minnesota. We also discovered and corrected some errors made in data entry, such as the tendency to misinterpret the age of children given in months as their age in years (e.g., a child whose handwritten age was “3/12” was frequently entered as “3”, and frequent error we discovered in our auditing process).

We also identified and corrected data entry errors in the identification of different slaveholdings. This error typically occurred in 1860 and was related to holdings with multiple slaveholders. As noted earlier, enumerators did not identify slaveholdings with unique serial numbers as they did for households and families on the free population schedules. We identified slaves in the same holdings using slaveholders’ names, which were typically recorded on the record of the first slave in the holding. In 1850, it is rare to find more than one slaveholder name listed and there was little ambiguity in determining the end of the previous holding and the beginning of a new one. The enumerator instructions in that year specified that in cases where “there are several owners to a slave, the name of one only need be entered“...”the principal object being to get the number of slaves, and not that of masters or owners.” The 1860 instructions, however, emphasized getting an accurate count of the number of owners and the names of all slaveholders. The instructions stated that “The person in whose charge, or on whose plantation the slave is to be found to be employed may return all slaves in his

charge (although they may be owned by other persons) provided they are not returned by their proper owner. The name of the bona fide owner should be returned as proprietor, and the name of the person having them in charge as employer.”³

A potential problem arises when there are multiple slaveholders associated with a single slaveholding and their names were recorded on multiple lines (e.g., when an overseer or employer is named on the first line on the holding and the owner is named on the second line. Most of these cases are clear in context. See Figure 1 for a typical example. The enumerator in this example wrote “Samuel Shadwell” on the first line of the holding (line 32 of the form) and “for William Hall” on the second line (line 22 of the form). The explanatory text before William Hall’s name—the preposition “for”—indicates that Hall was the owner of slaves in the holding while Shadwell on the line above was an employer or overseer. In this case, Samuel Shadwell and William Hall should have both been entered on the line associated with the first slave in the holding by the data entry operator, with Shadwell as the first slaveholder and Hall as the second slaveholder and with the “for” entered as explanatory text. We found that many data entry operators, however, missed the information associating the two slaveholders. In this case, Samuel Shadwell was listed as the first and only slaveholder for the female slave aged 65 on line 32 of the form, while William Hall was listed as the first and only slaveholder for the male slave aged 30 on line 33. From the raw data alone, this single slave holding of 9 slaves appears to be two holdings, with the first female slave associated only with Samuel Shadwell (a slave holding of just one slave), while the second and subsequent slaves were associated only with William Hall in a holding of 8 slaves. Given

³ Thus each holding should have an owner’s name on the slave schedule (who will usually also be the employer) while other holdings will have both an owner and employer name. In the latter case, it was up to the enumerator whether the slaves would be enumerated at the time of the owner’s or employer’s visitation (i.e., when the owner or employer was enumerated on the population schedule). For instructions to enumerators, see See <https://usa.ipums.org/usa/voliii/inst1860.shtml>. See also https://www.afrigenas.com/library/slave_schedule2.html.

the importance of the number of holdings, slaveholders, and slaves per holding to the dataset, we spent a considerable amount of time attempting to find and fix this type of error.

As the above examples indicate, in contrast to the prior IPUMS samples of the slave population, which were constructed entirely at the University of Minnesota, the full-count slave datasets were constructed without IPUMS oversight. We found that many cases and variables were haphazardly entered, most notably geographic place, disability information for individual slaves, slaveholder names, and the number of slave houses, manumissions, and fugitives for each holding. Stray marks and numbers written by what appears to be different individuals—such as the running totals made by Census Office staff when tabulating the manuscript returns—were sometimes entered into fields for individual slaves. Some variables were not collected, while others included significant percentages of missing information. In addition, Ancestry.com recorded a maximum of three slaveholders for each slave holding in 1860 (the earlier IPUMS samples for 1860 recorded information for up to eight slave holders). Although only a very small percentage of slave holdings have information on four or more slaveholders, this information was not collected. Moreover, information that helps identify the type of holder in 1860 (e.g., whether the person named in the census was an owner, renter, overseer, trustee, guardian, business partnership, etc.) was occasionally not recorded. Given the large number of records in the full-count datasets, the IPUMS project lacked the resources to re-enter all these data.

Unsurprisingly, given these problems, we found that many of the variables were unreliable and overall totals in the datasets, such as the number of slave fugitives and manumissions, differed markedly from published totals. So, despite their larger size and comprehensiveness, we anticipate that IPUMS full-count slave datasets will not completely replace the earlier IPUMS samples constructed from the slave censuses two decades ago. Although these earlier datasets were random samples of only 5 percent of the nation's slaves—combined with a small non-random selection of counties with full information—they were constructed entirely by the IPUMS project. Care was taken with all information

recorded on the manuscript returns, with concurrent data entry, checking, verification, and error correction, resulting in the highest possible data quality.⁴ Users needing individual-level data are therefore encouraged to rely on the earlier sample datasets where possible. Nonetheless, the complete-count datasets should prove to be a better choice for research tasks where low sample densities will not suffice and where information on houses, manumissions, and fugitives are not required. Because of the higher-quality data in the original 1860 slave dataset (Menard et al. 2004), we decided to replace the data entered by Ancestry.com with the data from the selection of counties in the earlier dataset with full-count data. In total, these counties included information on 539,509 slaves, representing about 13.6% of slaves in 1860. A variable in the IPUMS 1860 full-count slave dataset (“Menard”) indicates cases in which the data were obtained from the original dataset. These counties will contain higher quality data on manumissions, fugitives, slave houses, and multiple owners and most other variables. We caution, however, that these counties were a non-random selection of counties in 1860, and users wishing to confine their analyses to these cases should evaluate their representativeness and apply appropriate weights if possible.

Table 1 shows a partial list of the variables available in both datasets. Variables marked an “X” indicates a variable with information taken directly from the census, although the string data has been replaced with IPUMS standard coding where possible (e.g., the dataset includes stateicp codes instead of state names). Variables identified with a “C” indicates variables we constructed using the available information and logical rules (e.g., the number of slaves in each holding). Finally, variables that were incompletely collected and contain significant errors are designated with “inc.” We considered dropping these variables from the published datasets but thought that users might find them helpful for some areas, such as those in the full-count counties collected in the earlier Menard et al. dataset (2004). We

⁴ These datasets are available for free public download at <https://usa.ipums.org/usa/slavepums/data/data.html> .

caution, however, that users should conduct descriptive analyses for local areas and compare the results to published totals before using these variables.

Because of the importance of the names for linking slaveholders in the slave schedules to their records on the free population schedules, we spent a fair amount of effort correcting the names transcribed by LDS and Ancestry.com. In addition to improving name transcription accuracy, we also restored non-name information, when missing, that described the relationship between the slaveholder and the slaves. Names are available only in the restricted versions of the datasets.

We anticipate that one use of the slave datasets will be to investigate slave ownership (e.g., Oakes 1982: 245-250; Olsen 1972). Although census instructions make no mention of various types of owners in 1850 and are vague in 1860, we were able to classify most slaveholders in 1860 using one of six major holder types (owner, business/institution, employer, overseer/manager, trustee/guardian, and estate). Where possible, we also classified slaveholders into several minor types within these larger categories (e.g., we have detail codes identifying overseers, managers, agents, administrators, superintendents, executors, guardians, trustees, and many other types). In addition to the problems in the data entry and identification, we caution that differences in census instructions between the two censuses noted above makes the comparison of slave ownership rates between 1850 and 1860 problematic.

Linking Slaveholders to the Free Population Censuses

Research possibilities for the 1850 and 1860 slave censuses are increased significantly by linking slaveholders to their records in the free population censuses. The linked datasets make it possible for researchers to attach slaveholders' characteristics to slaves, enhancing analyses of the slave population. The linking process also results in the identification of most slaveholders and non-slaveholders in the

free population census, allowing analyses of slaveholding and slave ownership. The identification of slaveholding individuals can also be treated as an independent variable in studies of demographic behavior. Carter, Ransom, and Sutch (2004), for example, hypothesized that fertility rates among slaveholding couples should be lower than fertility rates among non-slaveholding couples because the former did not need to depend on children for farm labor and support in old age (see also, Hacker, Haines and Jaremski 2021). The linked dataset can be used to test this and other hypotheses.

Linking slaveholders to the free population census was challenging because names are the only explicit linkage information. The IPUMS 1850 and 1860 full-count slave datasets and the 1850 and 1860 full-count free population datasets were transcribed at different dates by different data entry operators. Nineteenth-century handwriting can be difficult for data entry operators to read, and it is no surprise to find many examples of what appears visually to be the same names on different census schedules to be spelled differently in the two databases, often dramatically different. Moreover, some enumerators used initials instead of first names, making it more difficult to determine if two names in different databases represent the same person, especially in areas where patrilineal kin resided in nearby households and the frequency of shared surnames is high (Nelson 2020).

Linking was possible, however, especially when we confined our searches to small areas. Both census schedules contained reliable information for state and county, which we used for “blocking” during the potential links generation process (i.e., searching only for potential matches in the same state and county). In addition, we relied on the order in which slave holdings were recorded on the manuscript pages. Typically, enumerators completed the free population, slave population, and mortality schedules during the same visitation to a household. Although enumerators numbered the sequential order of their visits to households and families only on the free population schedules, the order of each enumerator’s visit is preserved in the slave and mortality datasets and should follow the same general order, albeit skipping households without slaves or the death of a household member in

the previous year.⁵ The sequential order of enumeration, therefore, is a valuable clue in discerning the correct slave holder in the free population schedule, especially when there are significant differences in the spelling of the first and last names in the two datasets.

Our strategy of linking slaveholders to the free population schedules varied by census year. For 1850, which began at an earlier date as part of an earlier project, we reduced the slave census data to unique combinations of holdings and owner names, blocked potential matches by state and county, and calculated Jaro-Winkler similarity scores—a measure of the similarity of two strings based on the number of matching characters and transpositions—for first and last names (Winkler 1990). We assigned a Jaro-Winkler score of 0.8 to given names when one side contained a single letter and matches the other schedule's first letter in the name (e.g., 'J' and 'John'). We considered all potential links with first and last name similarity scores of 0.8 or above for potential slaveholders aged 18 years and above and accepted the link depending on its score, proper sequential ordering, and distance from the next best potential match. Most cases were machine linked, but approximately 12% of the slave holders in the 1850 dataset were hand linked. Hand links were sometimes necessary because of errors in transcription and because slave owners did not always live in the county where the slaves they owned were enumerated.

For 1860, slaveholder links to the free population schedules were made entirely by machine. We used a multiple round process where the most confident links were made in the first round. Subsequent rounds lowered the required thresholds needed to make a match but benefitted from more knowledge of the sequential ordering of links on both schedules. As was also the case in the 1850 linking project, we reduced the slave census data to unique combinations of holdings and slaveholder names, blocked

⁵ Because about three-in-ten southern households owned slaves, the typical pattern was to find slaveholders adjacent in the slave schedules to be separated by about three households on average in the population census. Individual cases varied, of course, by chance and the degree of slaveholding in each area.

potential matches by state and county, and calculated Jaro-Winkler similarity scores for first and last names. We again considered all potential links with first and last names similarity scores of 0.8 or above. In addition, however, we only considered potential links if the individual on the free population census had personal property of at least \$100. We also augmented the name similarity scores by determining whether potential links between individuals named as a slaveholder in the slave census and in the free population census were “neighbors” with other potential linked individuals, using the observed differences between their serial numbers in the dataset. We used the neighbor count and the potential link’s location in the free population file to reject some links we might have otherwise made. In general, this approach is conservative; we were willing to reject high name similarity potential links based on an implausible location of the potential slave holder in the free population file.

Subsequent link rounds included some cleaning of bad links and lowering the Jaro-Winkler threshold to 0.65 for both given and surnames, provided the sequential ordering of households and neighbors was viable. In the final round of linking, we blocked only by state and raised the Jaro-Winkler threshold to 0.9 for the last name, which allowed us to link some absentee owners who resided in another county of the same state.

The processes linked 291,934 slaveholders in 1850 (84.0% of those counted by the 1850 Census Office) and 317,257 slaveholders in 1860 (80.6% of those counted by the 1860 Census Office). Although it is impossible to evaluate the accuracy without external data, we are optimistic about the quality of the links. Our optimism is based on having links with reasonable name similarity that also meet minimum requirements for age and personal property wealth, along with the presence of linked neighbors and logical order in the population files. Unlinked holdings fell primarily into two general categories: absentee owners and holdings where there was insufficient name similarity to qualify as a potential link.

Research Possibilities Using the New Datasets

We anticipate many possible uses of the IPUMS full-count slave census microdatasets in several fields of study. These include the study of slave mortality and fertility, patterns of slaveholding, and the size, growth, and distribution of the mixed-race population, among other potential topics.

Demographic research on the slave population will benefit from access to the new datasets. Microdata allows researchers to use the full detail of the information collected by enumerators, create custom cross-tabulations, and conduct analyses at mixed levels ranging from the individual slave level to holding and owner levels, and at multiple levels of geography, including place, urban/rural residence, county, state, region, and nation. Prior studies of slave demography were forced to rely on the published age distributions of the slave population to estimate fertility, mortality, and population growth, which were limited above age 20 to ten-year age intervals (Hacker 2020; McClelland and Zeckhauser 1982; McDaniel and Grushka 1995). The full-count IPUMS dataset will allow researchers to tabulate the data by single years of age, which can be used to measure and correct for age-heaping errors—the tendency to round ages to preferred digits, typically those ending with “0” or “5.” As shown in Figure 2, age-heaping in the slave population was severe. There were, for example, 16.5 times more slaves aged 50 in the census than the average of the number of slaves aged 49 and 51. The microdata can also be used to examine how demographic patterns varied over time, space, and characteristics of the slaveholding or the slave owner. Researchers, for example, can examine how the adult sex ratio varied in regions specializing in different staple crops (e.g., cotton, tobacco, rice, sugar, and other), how child-to-woman ratios varied by size of slave holding and region (Steckel 1985), and how migration patterns by age and sex.

Both the 1850 and 1860 slave datasets recorded slaves’ “color.” In 1850, enumerators were instructed to record whether slaves were “Black” or “Mulatto,” the latter being a nineteenth-century for

mixed race individuals. In 1860, the instructions were more specific, noting that slaves “who are in any degree of mixed blood are to be termed mulatto.” The designation of mixed-race individuals in the datasets facilitates possible study of the prevalence of interracial relationships between slaveholders and their slaves, a topic that generated extensive comment and debate among both nineteenth-century abolitionists (e.g., Douglass 1845) and historians (e.g., Fogel and Engerman 1974; Genovese 1974; Malone 2000; Yarbrough 2005). Although census data cannot tell us how often interracial relationships took place or the character of those relationships, they can provide clues to its incidence, distribution, and correlates.

The slave microdatasets will also provide a better source of data to study the size and structure of slave holdings from the perspective of the slaves who were part of them and for the study of slaveholding from the perspective of owners. The 1850 Census Office published the number of slave holdings of various size categories in each state, while the 1860 Census Office published the number of slave holdings of various size categories in each county, facilitating some research on the distributions of different sized slave holdings. Although the data for larger-sized holding were binned into large categories (e.g., the number of slaveholdings in specified county with between 100 to 199, 200-300, 300-499, 500-999, and 1,000 and over slaves), researchers have used these data to estimate that the median slave lived on holding of 20.6 slaves in 1850 and 23.0 slaves in 1860 (Gray 1924: 530). Because slaves lacked the ability to travel without their owner’s permission, the size and distribution of slaveholdings have major implications for slaves’ work and social lives, including their ability to form families and maintain kin connections. Microdata will allow researchers to be more precise about the distributions of slaves across various size holdings, including precise estimates of the mean and median size of holdings, and how those distributions varied across time, space, and dominant crops.

The Census Office also reported the number of slaveholders in each state in 1850 and the number in each county in 1860. In 1850, the total number of slaveholders in the United States was

347,725, while in 1860 it was 395,216. These figures indicate that the average slaveholder held 9.2 slaves in 1850 and 10.0 slaves in 1860 (Carter 2-380). Again, the microdatasets will allow more detailed investigations. Because we were unable to link every slaveholder to their record in the free population, the two datasets identify only 291,934 and 317,257 slaveholders respectively. That percentage is high enough, however, to support analyses of slaveholding.

As an example of the kind of analysis of that can be conducted, we examined how slaveholding varied across the life course in 1850 and 1860, using the slaveholders who were linked to the IPUMS 1850 and 1860 full-count datasets for the free population census. Researchers have observed that a large percentage of individuals in the United States did not own slaves, including a majority of southern white men who fought on behalf of the Confederate States of America during the American Civil War (Glatthaar 2011). But percentages are higher among southern families and among southern white men too old to enlist in the Confederate armed forces. Although few Confederate soldiers owned slaves themselves, many more of their fathers and family members did so, and many could anticipate owning slaves in their lifetimes.

Before examining age pattern of slaveholding, we imputed the 77,959 slave holders in 1860 we were unable to link to the free population. We began by constructing a logistic regression model predicting slave ownership, with age, sex, occupation group, the natural log of real estate wealth, the natural log of personal estate wealth, and a dummy variable for household heads as independent variables. The results were as expected, with the likelihood of holding slaves positively associated with age and wealth. We then assigned a propensity score to each person who was not identified as a slaveholder in the linking process to hold slaves. We used that score, the number of unidentified slave holders in each county, and a random function to impute the missing slaveholders.

In Figure 3, we plotted the percentage of white males in both census years who held slaves by age. We plotted the linked and the linked plus imputed slaveholders separately to show the effects of the imputation process. The figures indicate that slaveholding was strongly correlated with age, rising from less than 1% of white males in the South in South aged 15-19 in both censuses to 35.0% among white males aged 55-59 in 1850 (linked and imputed slaveholders) and 33.3% among white males aged 65-69 in 1860. Slaveholding declines from these peaks among men in older age groups, possible because of endowments of slaves to children. Interestingly, the results indicate that slaveholding declined for all age groups between 1850 and 1860, and differentials were especially wide in middle years. At age 40-44, for example, 25.6% of southern white men held slaves in 1850 compared to 20.2% of white men in the same age group in 1860. If we follow the same cohort of men between 1850 when aged 40-44 and 1860 when aged 50-54, the percentage holding slaves increased a modest 1.5% in the decade, from 25.6 to 27.1%.

Much more can be done, of course, including comparing slaveholding rates and trajectories across states, among areas with different dominant crops, and among different population subgroups. We look forward to seeing researchers use the new datasets to expand our understanding of the slave population, slaveholding, and the southern economy in the period immediately before the most destructive war in United States history.

Conclusion

New IPUMS full-count datasets of the slave censuses of United States in 1850 and 1860 represent powerful new sources of data to study the slave population and slaveholding in a critical period in U.S. history. Because the data are full-count, researchers can now study small subpopulations and small areas. Researchers can also construct custom variables, calculate various aggregated summary

statistics, and conduct multilevel analyses ranging from level of the individual slave to slaveholdings to various geographic levels.

Numerous other areas of research are possible, including the potential for new investigations of slaveholding, slave fertility and mortality, the distribution of slaves in different sized holdings, and slave and slave owner migration. Because of a few data transcription issues noted above, however, researchers should consider whether the earlier sample datasets (Menard et al. 2004) remain more appropriate for their analysis.

Public use datasets are now available for free downloading on the IPUMS website (<http://ipums.org>). Restricted versions of the datasets, which include slaveholder names, are also available from IPUMS. Accessing these data requires a compelling research question, and user agreements to protect data confidentiality. Interested users should contact ipums@umn.edu or ipumsres@umn.edu to request access to these data.

References

- Alexander, J.T., S. Condon, J.C. Digman, and J.D. Hacker. 2003. "A public use microdata sample of the 1860 census of slave inhabitants." *Historical Methods* 36(1):21-26.
- Anderson, M.J. 1988. *The American Census: A Social History*. New Haven, Conn.: Yale University Press.
- Carter, S.B., R.L. Ransom, R. Sutch, T. Guinnane, W. Sundstrom, and W. Whatley. 2004. "Family matters: The life-cycle transition and the antebellum American fertility decline." *History matters: Essays on economic growth, technology, and demographic change*:271-327.
- Douglass, F. 1845. *Narrative of the life of Frederick Douglass, an American slave, written by himself*. Boston: American Anti-Slavery Society.
- Fogel, R.W. and S.L. Engerman. 1974. *Time on the Cross: The Economics of American Negro Slavery*. Boston: Boston, Little, Brown.
- Genovese, E.D. 1974. *Roll, Jordan, roll; the world the slaves made*. New York,: Pantheon Books.
- Glatthaar, J.T. 2011. *Soldiering in the Army of Northern Virginia: A Statistical Portrait of the Troops Who Served under Robert E. Lee*: UNC Press Books.
- Hacker, J.D. 2020. "From '20. and odd'to 10 million: the growth of the slave population in the United States." *Slavery & abolition* 41(4):840-855.
- Hacker, J.D., M.R. Haines, and M. Jaremski. 2021. "Early fertility decline in the United States: Tests of alternative hypotheses using new complete-count census microdata and enhanced county-level data." Pp. 89-128 in *Research in Economic History*: Emerald Publishing Limited.
- Malone, A.P. 2000. *Sweet chariot: Slave family and household structure in nineteenth-century Louisiana*: Univ of North Carolina Press.
- McClelland, P.D. and R.J. Zeckhauser. 1982. *Demographic Dimensions of the New Republic : American Interregional Migration, Vital Statistics, and Manumissions, 1800-1860*. New York: Cambridge University Press.
- McDaniel, A. and C. Grushka. 1995. "Did Africans live longer in the antebellum United States? The sensitivity of mortality estimates of enslaved Africans." *Historical Methods* 28(2):97-105.
- Menard, R., T. Alexander, J. Digman, and J.D. Hacker. 2004. "Public Use Microdata Samples of the Slave Population of 1850-1860." edited by M.P. Center. Minneapolis, MN: IPUMS.
- Nelson, M.A. 2020. "The decline of patrilineal kin propinquity in the United States, 1790–1940." *Demographic research* 43:501.
- Oakes, J. 1982. *The ruling race*: Knopf.

- Olsen, O.H. 1972. "Historians and the extent of slave ownership in the Southern United States." *Civil War History* 18(2):101-116.
- Steckel, R.H. 1985. *The Economics of U.S. Slave and Southern White Fertility*. New York: Garland Publishing.
- United States, B.o.t.C. 1853. "Abstract of the seventh census." Washington, D.C.: Government Printing Office.
- . 1854. "Compendium of the seventh census, 1850." Washington, D.C.: Government Printing Office.
- . 1862. "Preliminary report of the eighth census." Washington, D.C.: Government Printing Office.
- . 1864. "Census of agriculture." Washington, D.C.: Government Printing Office.
- Waldstreicher, D. 2010. *Slavery's constitution: from revolution to ratification*: Hill and Wang.
- Winkler, W.E. 1990. "String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage."
- Yarbrough, F.A. 2005. "Power, perception, and interracial sex: Former slaves recall a multiracial south." *The journal of southern history* 71(3):559-588.

Figure 1. Selection of 1860 Slave Census Manuscript page for Talladega County, Alabama

<i>Samuel Shadock</i>	1	05	f	B						32
<i>For William Hall</i>	1	30	m	B						33
	1	30	m	B						34
	1	30	m	B						35
	1	25	m	B						36
	1	15	m	B						37
	1	13	m	B						38
	1	9	m	B						39
	1	8	m	B						40

Table 1. Partial list of Variables in the IPUMS Full-Count Slave Census Datasets, 1850 & 1860

Variable name	Description	1850	1860
year	Census year	X	X
serial	Slaveholding serial number	C	C
slave_owner_order	Slaveholder order number	C	C
slavenum	Slave number in holding	C	C
slave_count	Holding size	C	C
sex	Sex	X	X
race	Color/Race	X	X
age	Age of slave	X	X
agemonth	Age of slave in months	X	X
Blind	Blind		inc.
Deaf	Deaf		inc.
Idiotic	Idiotic		inc.
Insane	Insane		inc.
mpcid	Historical ID number for slaves	C	C
numholders	Number of slave holders	C	C
sh1typed	First slave holder, type of holder		C
sh2typed	Second slave holder, type of holder		C
sh3typed	Third slave holder, type of holder		C
link_status	One or more slaveholders linked to population	C	C
mpcid_pop1	Historical ID number for slave holder 1	C	C
mpcid_pop2	Historical ID number for slave holder 2		C
mpcid_pop3	Historical ID number for slave holder 3		C
stateicp	State ICPSR code	X	X
statefip	State FIPS code	X	X
countyicp	County ICPSR code	X	X
place	Place name	X	
stdtownship	Township name		X
urban	Urban/rural status	C	C
city	City name code	C	C
citypop	City population (incorporated places)	C	C
Number of slaves fugitives	Number of slaves fugitives		inc.
Number of slaves manumitted	Number of slaves manumitted		inc.
Number of slave houses	Number of slave houses		inc.
Menard	Data taken from Menard et al. (2004) dataset		X
Reel	Microfilm reel number	X	X
Page	Microfilm page number	X	X
Ancestry.com url	Link to ancestry.com stable url		C
Sh1first	First name of first slaveholder	Res.	Res.
Sh1last	Last name of first slaveholder	Res.	Res.
Sh2first	First name of second slaveholder		Res.
Sh2last	Last name of second slaveholder		Res.
Sh3first	First name of third slaveholder		Res.
Sh3last	Last name of third slaveholder		Res.

Notes: Variables denoted by "X" are census questions with available data in a given year and coded, where applicable, using IPUMS codes. Variables denoted by "C" were constructed using logical rules. "Inc." indicates variable with significantly incomplete data or with significant errors in the dataset. These variables are complete when data was taken from the Menard et al. 1860 slave dataset (2004), designated with the Menard variable. "Res." are variables available only in the restricted-use versions of the datasets. Accessing these data does require specific stipulations in order to use, and interested users should contact ipums@umn.edu or ipumsres@umn.edu to request access.

Figure 2. Age Distribution of the Slave Population of the United States in 1860, both sexes

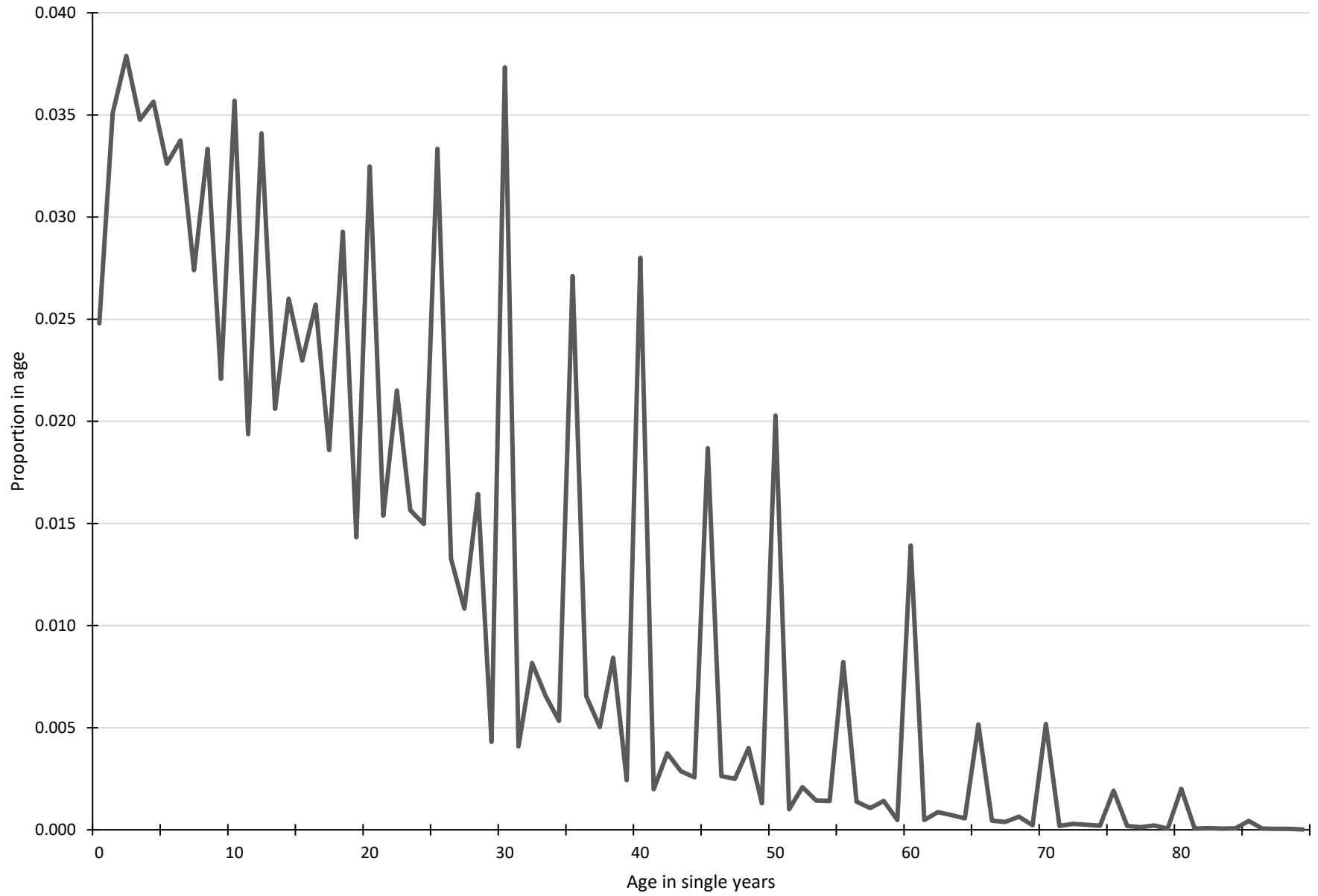


Figure 3. Percentage of all white males holding slaves in South by age, 1850 and 1860 IPUMS full-count datasets

