

# Quantitative History

Margo Anderson

## WHAT IS QUANTITATIVE HISTORY?

Quantitative history is the term for an array of skills and techniques used to apply the methods of statistical data analysis to the study of history. Sometimes also called cliometrics by economic historians, the term was popularized in the 1950s and 1960s as social, political and economic historians called for the development of a 'social science history', adopted methods from the social sciences, and applied them to historical problems. These historians also called for social scientists to historicize their research and consciously examine the temporal nature of the social phenomena they explored. For both types of questions, historians found that they needed to develop new technical skills and data sources. That effort led to an array of activities to promote quantitative history.

Classical historical research methodology relies upon textual records, archival research and the narrative as a form of historical writing. The historian describes and explains particular phenomena and events, be they large epic analyses of the rise and fall of empires and nations, or the intimate biographical detail of an individual life. Quantitative history is animated by similar goals but takes as its subject the aggregate historical patterns

of multiple events or phenomena. Such a standpoint creates a different set of issues for analysis. A classic historical analysis, for example, may treat a presidential election as a single event. Quantitative historians consider a particular presidential election as one element in the universe of all presidential elections and are interested in patterns which characterize the universe or several units within it. The life-course patterns of one household or family may be conceived as one element in the aggregate patterns of family history for a nation, region, social class or ethnic group. Repeated phenomena from the past that leave written records, which read one at a time would be insignificant, are particularly useful if they can be aggregated, organized, converted to a electronic database and analyzed for statistical patterns. Thus records such as census schedules, vote tallies, vital (e.g., birth, death and marriage) records; or the ledgers of business sales, ship crossings, or slave sales; or crime reports permit the historian to retrieve the pattern of social, political, and economic activity in the past and reveal the aggregate context and structures of history.

The standpoint of quantitative history also required a new set of skills and techniques for historians. Most importantly, they had to

incorporate the concept of the data set and data matrix into their practice. Floud (1972: 17) defined the data set as 'a coherent selection of data from the whole range of historical data available to the historian, and it is selected because it relates closely to the questions that the historian wishes to consider.' The myriad instances of a phenomenon—for example, all United States presidential elections—form the cases of the data set. The pieces of information collected about the cases—for example, the candidates running, the year of the election or the vote totals—become the variable characteristics of the data set, that is, the varying characteristics of any particular case. The historian arranges the data in tabular form, that is, in a matrix of rows and columns, 'consisting of a number of rows, which will normally represent cases, and a number of columns, which will normally represent variables' (Floud, 1972: 18). The creation of quantitative data sets thus required the historian to carefully compile consistent information about the phenomenon to be investigated, and prepare the data in tabular form. Historians then were prepared to apply the techniques of statistical data analysis to the data set to answer the research question posed.

In short, to make effective use of quantitative evidence and statistical techniques for historical analysis, practitioners had to integrate the rapidly developing skills of the social sciences, including sampling, statistical data analysis and data archiving into their historical work. That task led to the development of new training programs in quantitative methods for historians, to the creation of new academic journals and textbooks, and to the creation of data archives to support the research.

## EARLY EFFORTS

Historians had made use of quantitative evidence prior to the 1950s, particularly in the fields of economic and social history. The *Annales* school in France pointed the way in the pre-World War II period. The rapid

growth and expansion of the United States had long required American historians to consider quantitative issues in their study of the growth of the American economy, population and mass democracy. Thus, for example, Frederick Jackson Turner's classic 1893 essay on 'The Significance of the Frontier in American History' was largely based on a reading and interpretation of the results of the 1890 population census.

But true 'data analysis' in the current sense had to await the growth of the social and statistical sciences in the first half of the twentieth century, and the diffusion to universities in the 1950s of the capacity for machine tabulation of numerical records, and then of mainframe computing in the 1960s. One can see the emerging field exemplified in seminal studies in the late 1950s and early 1960s. In 1959, for example, Merle Curti and his colleagues at the University of Wisconsin published *The Making of an American Community: A Case Study of Democracy in a Frontier County*. Curti et al. (1959) explored Turner's thesis with an in depth look at the mid-nineteenth century history of Trempeleau County, Wisconsin, including its records of newspapers, diaries, private papers and county histories. But they also added data analysis of the employment patterns derived from the individual-level federal census manuscripts for the censuses from 1850 through 1880.

Similarly, the 'new' economic historians of the 1950s challenged the conventional wisdom of the day on several key issues in economic history. One debate centered on the 'necessity' of the US Civil War. Historians at the time argued that the war had been 'unnecessary' since the institution of race-based slavery would collapse under the weight of its unprofitability. In contrast, economic historians employed economic theory and data on output of southern agriculture to argue that the southern agricultural economy could have survived profitably into the twentieth century using slave labor (Conrad and Meyer, 1958). Robert Fogel challenged the conventional wisdom on the centrality of railroads for the industrial development of

the United States. Making use of economic theory, carefully compiled data series, and the logic of the counterfactual, Fogel argued that canals would have also succeeded as a transportation system underpinning nineteenth-century American industrial development (1964).

'New political historians' such as Lee Benson, Allan Bogue, Richard P. McCormick, and political scientists with historical interests, such as Warren Miller and Walter Dean Burnham, translated the emerging techniques of political scientists analyzing contemporary election results and voter surveys to historical questions, and opened up dramatic new insights into American political history.<sup>1</sup> The new political historians identified the parameters of party systems, developed the theory of the critical election, and argued that underlying structures of electoral politics were accessible through historical analysis of voter turnout and election results. In 1964 in England, demographers and historians founded the Cambridge Group for the History of Population and Social Structure and began a forty-year project to retrieve, assemble and reconstruct 400 years of the family history of Britain.<sup>2</sup>

The new possibilities of quantitative history fit well with other trends within the discipline of history, particularly with the growth of social history and calls for what Jesse Lemisch (1967) called 'history from the bottom up'—that is, for historians to treat the lives of ordinary people, to complement the study of elites. By the mid-1960s, the interest in the new techniques led the American Historical Association to recognize that 'quantification in history' would require new skills and institutions within the historical profession. The AHA created a Quantitative Data Committee to consider the issues. Summer institutes and classes in quantitative methods for historians were held in 1965, 1967 and 1973 at the University of Michigan, Cornell University and Harvard University respectively. In 1968, the Inter-university Consortium for Political Research at the University of Michigan began offering a four-week course in quantitative historical

analysis as part of its summer program in quantitative methods. The course continues to be offered each summer. At the Newberry Library in Chicago, from 1971 to 1982 Richard Jensen spearheaded a summer program in quantitative methods for historians. By the early 1980s, about 40 percent of history graduate programs offered training in quantitative history as part of the graduate curriculum (Bogue, 1983: 220ff.).<sup>3</sup>

Additional institutional infrastructure of quantitative history can also be dated to the 1960s. New journals, textbooks, and edited collections also promoted the growth of quantitative history. The *Historical Methods Newsletter*, for example, began publishing in 1967, and was renamed *Historical Methods* in 1978. The *Journal of Interdisciplinary History* began publication in 1970. The Social Science History Association (SSHA) was founded in 1974 and the first issue of its journal, *Social Science History*, appeared in 1976. SSHA became the professional venue for bringing together historians who consciously adopted theory and methodology from the social sciences and social scientists doing historical work. The cross-fertilization has continued, and, as noted below, many of the innovations in quantitative history have been developed by scholars with formal training in the social sciences and appointments in departments of economics, demography, sociology, anthropology, geography and political science.

Textbooks in quantitative history began to appear in the early 1970s, and many have been published since.<sup>4</sup> Numerous edited volumes introduced the new field and techniques to professional and student audiences.<sup>5</sup> Finally, researchers created data archives. In the United States, the Inter-university Consortium for Political Research (ICPR) was founded in 1962 primarily by political scientists. Renamed the Inter-university Consortium for Political and Social Research (ICPSR) in 1975, the Consortium has also pioneered in the creation and preservation of historical data collections. The United States National Archives and Records Administration (NARA)

created an electronic records preservation program in the early 1970s for federal government data that was 'born digital' (Ambacher 2003; Adams, 1995, forthcoming; Fishbein, 1973). Similar work began in Britain with the founding of the UK Data Archive in 1967.<sup>6</sup>

Thus by 1980, historians had taken major steps to establish the institutional structures necessary to integrate quantitative history into larger historical practice. That infrastructure has, if you will, both matured and faced challenges in the generation of work since, and in many ways quantitative history is still a work in progress. Nevertheless, it is possible to identify the types of questions quantitative history was intended to and has been able to address; the major types of data sets that have developed and the key characteristics of historical data sets; and the most commonly used techniques within the field. That background in turn provides the framework for a review of a number of methodological issues historians uniquely face, for a review of the achievements of quantitative history, and for a discussion of emerging issues.

### **QUESTIONS, DATA AND ISSUES IN CREATING HISTORICAL DATA SETS**

Quantitative history has been most successful in addressing big questions about long-term historical patterns of change. Practitioners have achieved important results by assembling substantial amounts of numeric or countable information, and organizing it into tabular data matrices for statistical analysis. The first generation of studies focused especially on the history of the family and social structure, trends in economic growth and change, patterns of electoral behavior and voter participation, or the record of inter-generational social mobility and living standards. More recently, the examples have proliferated. Historians of crime and the criminal justice system, for example, have retrieved court and newspaper records to examine the long-term patterns of crime and

violence in the past. Historians of the family have examined patterns of inheritance and the inter-generational transfer of wealth. The emerging work of 'anthropometric' history—the study of living standards and well-being in the past using measures of height, weight, stature and disease in the past—has cast an even wider net, aiming to evaluate comparative living standards over centuries and ultimately millennia.<sup>7</sup>

Making such studies possible was an explosive growth in the data sets informing quantitative history. Quantitative history, like other branches of the social sciences, requires what was once called 'machine-readable' (and are now known as 'electronic') data for analysis. Though there are some examples of large-scale data analysis undertaken by manual systems of tabulation and statistical analysis, most notably the nineteenth-century tabulations of census or vital registration records, social science data in the modern sense required the development of machine tabulation devices, counter sorters, and other mechanized calculators. The first system was the Hollerith system of punch-card tabulation used for the 1890 American population census; the social and statistical sciences grew with the new machinery. By the 1940s, social scientists had developed rules and procedures for collecting quantitative data to make best use of machine tabulation and analysis. These conventions included the fixed format data matrix, the classification of variables into nominal, ordinal, interval and ratio variables, the organization of questionnaires and survey forms to facilitate conversion to punch-cards for analysis, and coding systems such as the Likert scale. Quantitative historians inherited these practices and adapted this existing technology and set of conventions to their historical project. They soon recognized that they had to solve major new methodological and logistical problems before the potential for quantitative history could be achieved.

The first problem derives from the larger evidentiary issue faced by all historians, namely, that historical analysis must rely on the extant record of the past. Historians are at

the mercy of their subjects' penchant and capacity for preservation. And before 1890, that is, for most of the historical record of human history, no preserved data were 'machine-readable'. Thus all potential historical data had to be created from surviving, usually text-based, records and converted to machine-readable or electronic format. Even records collected in the twentieth century and informed by the conventions of the emerging social sciences frequently no longer exist in machine-readable format. Thus, the United States Census Bureau, for example, preserved the original paper census questionnaires from the eighteenth century forward. But census officials did not retain the punchcards they used to tabulate the censuses from 1890 to 1960. These cards were destroyed once the results of the census appeared in published form. Thus historians interested in reanalyzing the microdata from past censuses faced creating, or recreating, the machine-readable records.

Quantitative historians faced additional major methodological problems resulting from the recalcitrance of the existing archival historical records. All historians face the problems of missing data, and the difficulties of interpreting illegible, damaged, incomplete or destroyed records. For quantitative historians, though, aiming to translate the archival record to a data matrix for statistical analysis, these questions of data quality are particularly difficult. Cases and variables for a data matrix require precise conceptual and operational definitions, as do the allowable entries for particular cell values within the matrix, since the goal of statistical analysis is to assess extent, central tendency and dispersion of any particular characteristic. What does one do if the records for a year or period of years are missing? How does one handle illegible entries in the records of a company's finances? How does one know if the probate records found in a county archive are complete? Historians have had to confront the requirements for case and variable definition, classification and coding in building a data set. The solutions to these problems emerged with the overall field. The journal, *Historical*

*Methods*, in particular, became the venue for identifying, debating and proposing methodological solutions to these issues.

A related issue is the set of rules for extracting the information from a text-based evidentiary source to create a data set. Historical archives frequently contain text-based records that lend themselves to data set construction, but require considerable conceptual work before they can be manipulated statistically. Historians have made use of sales invoices, wills, parish registers and case files of charity or social welfare agencies, for example, and have had to create the cases and variables from the extant texts.

Historians have had to solve these methodological questions as they select the evidence to be analyzed and create the code-book for the data set. Whether one is analyzing existing tabular data from the past—for example, the records of imports and exports of a nation over a period of years, or the published results of a census—or whether one is creating a data set from text-based sources, the historian needs to define the case or unit of analysis, define the characteristics or variables to be selected to characterize the cases within the data set, and define the coding system used to organize the source information for the data set. Several examples of the issues involved best illustrate the work of quantitative historians.

## COMPILATION AND ANALYSIS OF PREVIOUSLY PUBLISHED DATA

The most accessible sources for quantitative historians were data that were already published in tabular format. The first generation of quantitative historians in particular compiled data sets from existing, usually aggregated, published data sources—for example, tabulated census results, election results, government reports of tax collections, imports and exports, and data from trade publications. Assembled into time series, such data permitted researchers to undertake basic analyses of historical trends and use regression models to correlate the

determinants of change. For example, Walter Dean Burnham's 950+-page study of count-level presidential election results, published in 1955, included a compilation of results from state archives and newspaper sources, and a discussion of the methodological issues he faced in compiling the data. Combined with denominator data from census results that allowed the researchers to measure turnout, the new data set permitted Burnham and his colleagues to begin the analysis of historical election analysis (Burnham, 1955). In similar ways, economic historians made particularly good use of the data compiled in statistical abstracts, such as the *Statistical Abstract of the United States*, published annually since 1878.

### CONVERTING TABULAR DATA IN MANUSCRIPT FORM TO ELECTRONIC FORMAT

A second source of quantitative data were archived tabular records in text-based format, probably best illustrated by individual-level census manuscript schedules. See Figure 14.1,<sup>8</sup> a facsimile of the 1950 US Census population schedule.

For the United States, such original census responses are available for all the federal censuses except 1890, and are available for public use through 1930.<sup>9</sup> The schedule is already in a matrix format, with rows of cases and columns of variables. The original difficulty with using these records is their volume. With one record per person for the censuses of 1850 and later, data set creation for a large portion of the population was beyond the capacities of an individual researcher. The first generation of quantitative historians resolved this problem by sampling, and usually by organizing a research project of a particular locale. The historical social mobility studies were designed as community studies to solve the problem of the volume of data.

Later generations of quantitative historians have by and large solved the problem of

volume through collaboration and by building historical public-use microdata samples, or PUMS files. Starting with the 1900 census, historians proposed to create historical PUMS files that would be similar to the contemporary PUMS files that the Census Bureau has created since 1970. In the late 1980s, researchers at the University of Minnesota, initially led by Russell Menard, Steven Ruggles and Robert McCaa, began systematic retrieval of the historical census data from the United States, and more recently from other nations. The Integrated Public Use Microdata Sample (IPUMS) Project and the International IPUMS project have created microdata samples for the United States from all the censuses from 1850 to 2000, and are now collecting such data for many nations of the world. The data are easily downloadable from the web. The researchers have also built the code-books, technical support materials, and research bibliography necessary for the user to understand the context of the questions and responses to the census.<sup>10</sup>

### *Creating Tabular Data from Text Based Records*

The most time-consuming type of data set creation is the conversion of text-based records to matrix format. For existing tabular data, whether in manuscript or published form, the basic framework of the matrix is given in the original source. For text-based records with no tabular structure, it is up to the researcher to create the code-book, and thus all the variable definitions and coding rules. Figure 14.3,<sup>11</sup> an illustration of a record of a slave sale in antebellum America, illustrates the issues.<sup>12</sup>

There are thousands of such records in newspapers, private collections and archives, and, if marshaled for analysis, provide detailed, if somewhat gruesome, evidence of this chapter in American economic history. Robert Fogel and Stanley Engerman compiled such records for their study, *Time on the Cross* (1974) from the New Orleans Slave

QUANTIFICATION AND EXPERIMENT

(S1)

1. STATE *Iowa*

2. COUNTY *Stanton*

3. INCORPORATED PLACE OR TOWNSHIP *Orange City*

4. E. D. NUMBER *53-1*

5. HOTEL, LABOR ROOMING HOUSE, INSTITUTION, MILITARY INSTALLATION, ETC.

6. CONFIDENTIAL

FORM **P1** U. S. DEPT. OF COMMERCE 1950 CENSUS OF POPULATION AND HOUSING

HOUSEHOLD NUMBER	FOR HEAD OF HOUSEHOLD					FOR ALL PERSONS								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2022	9	No	No			Vacant								
2024	10	No	No			Stewart, James L. Head	W	M	32	Mar	Tennessee			
						Wife	W	F	30	Mar	North Carolina			
						Son	W	M	10	Mar	North Carolina			
						Daughter	W	F	6	Mar	North Carolina			
2029	11	No	No			Johnson, Edward Head	W	M	49	D	Virginia			
	12	No	No			Miller, George R. Head	W	M	66	Mar	Vermont			
						Son	W	M	27	Mar	Pennsylvania			
						Son-in-law	W	M	31	Mar	Pennsylvania			
						Daughter	W	F	29	Mar	Pennsylvania			
2064	18	No	No			Elmendorf, Eugene Head	W	M	61	Mar	Sweden			
						Wife	W	F	37	Mar	Colorado			
						Son	W	M	17	Mar	Michigan			
						Daughter	W	F	68	Mar	Sweden		No	
2067	19	No	No			Raney, Frances Head	W	F	53	D	Florida			
						Son	W	M	26	Mar	Texas			

HOUSEHOLD CONTINUED ON NEXT SHEET

THE QUESTIONS BELOW ARE FOR PERSONS LISTED

SAMPLE LINE	FOR ALL AGES											
	21	22	23	24a	24b	D	25	26	27	28		
1	Yes	Yes	Yes	County: <i>Stanton</i>	State or foreign country: <i>Iowa</i>	LEAVE BLANK	Father: <i>U.S.</i>	LEAVE BLANK	26	Yes	1	Yes
	No	No	No	County: <i>Stanton</i>	State or foreign country: <i>Iowa</i>	LEAVE BLANK	Mother: <i>U.S.</i>	LEAVE BLANK	27	No	2	No
6	Yes	Yes	Yes	County: <i>Stanton</i>	State or foreign country: <i>Iowa</i>	LEAVE BLANK	Father: <i>U.S.</i>	LEAVE BLANK	28	Yes	1	Yes
	No	No	No	County: <i>Stanton</i>	State or foreign country: <i>Iowa</i>	LEAVE BLANK	Mother: <i>U.S.</i>	LEAVE BLANK	29	No	2	No
20	Yes	Yes	Yes	County: <i>Cass</i>	State or foreign country: <i>New York</i>	LEAVE BLANK	Father: <i>Sweden</i>	LEAVE BLANK	30	Yes	1	Yes
	No	No	No	County: <i>Cass</i>	State or foreign country: <i>New York</i>	LEAVE BLANK	Mother: <i>Sweden</i>	LEAVE BLANK	31	No	2	No

30

Figure 14.1 Facsimile of 1950 Census Schedule for Orange City, Iowa

Market. ICPSR Study 7423 contains the data and code-book for the New Orleans Slave Sale Sample.<sup>13</sup>

For their sample, Fogel and Engerman converted the text-based records into cases and variables and codes, making decisions on unit

of analysis (the slave), sampling (2.5 percent or 5 percent, depending on the year of sale), number of variables (46), and codes. Each decision extracted a piece of information from the original text-based records, and had implications for ultimate analysis. The final

QUANTITATIVE HISTORY

The Census Bureau is authorized by Act of Congress (48 Stat. 11; 12 U. S. C. 261-263) which requires that a report be made. The information furnished is considered confidential treatment. The Census report cannot be used for purposes of taxation, investigation, or regulation.

U. S. DEPARTMENT OF COMMERCE  
BUREAU OF THE CENSUS

1950 CENSUS OF POPULATION AND HOUSING

C. DATE SHEET STARTED *April 1*, 1950  
SHEET NUMBER *1*

D. ENUMERATOR'S SIGNATURE  
*Harold W. Thompson*  
S. CHECKED BY  
*Julia Adams on April 5, 1950*  
(Over leader)

FOR PERSONS 14 YEARS OF AGE AND OVER

What State (or foreign country) was he born in?	What was this person doing most of last week, not counting work-keeping house, or something else?	1. If employed (WP in item 14, or Yes in item 16 or item 18), describe job or business held last week. If looking for work (Yes in item 17), describe last job or business.	2. For all other persons, leave blank.	What kind of work was he doing?	What kind of business or industry was he working in?	Class of worker
13	14	15	16	17	18	19
<i>Tennessee</i>	<i>Wk</i>			<i>48</i>	<i>Bus. driver</i>	<i>City Transit Co. G</i>
<i>North Carolina</i>	<i>H</i>	<i>No</i>	<i>No</i>	<i>No</i>		
<i>North Carolina</i>						
<i>North Carolina</i>						
<i>Virginia</i>	<i>Wk</i>			<i>40</i>	<i>Bricklayer</i>	<i>Building Contractors P</i>
<i>Vermont</i>	<i>Wk</i>	<i>No</i>	<i>No</i>	<i>No</i>		
<i>Pennsylvania</i>	<i>Wk</i>			<i>52</i>	<i>Lawyer</i>	<i>Law office O</i>
<i>Pennsylvania</i>	<i>Wk</i>			<i>40</i>	<i>Register operator</i>	<i>Paper mill P</i>
<i>Pennsylvania</i>	<i>H</i>	<i>No</i>	<i>Yes</i>		<i>Music teacher</i>	<i>City high school G</i>
<i>New York</i>	<i>No</i>			<i>40</i>	<i>Insurance salesman</i>	<i>Insurance Co. P</i>
<i>Sweden</i>	<i>Wk</i>			<i>35</i>	<i>Furniture finisher</i>	<i>Repair shop P</i>
<i>Colorado</i>	<i>H</i>	<i>No</i>	<i>No</i>	<i>No</i>		
<i>Michigan</i>	<i>Wk</i>	<i>Yes</i>		<i>8</i>	<i>Delivery man</i>	<i>grocery store P</i>
<i>Sweden</i>	<i>No</i>					
<i>Florida</i>	<i>H</i>	<i>Yes</i>		<i>40</i>	<i>Receptionist</i>	<i>Beauty shop P</i>
<i>Texas</i>	<i>Wk</i>			<i>48</i>	<i>quandy routeman</i>	<i>Laundry P</i>

ARE FOR PERSONS LISTED ON SAMPLE LINES

FOR PERSONS 14 YEARS OF AGE AND OVER

What is the highest grade of school that he has attended?	Did he attend school at any time since February last?	Has he attended school at any time since February last?	If looking for work (Yes in item 17), how many weeks has he been looking for work?	Income received by this person in 1949	Income received by this person in 1950	If this person is a family head (see definition below)—Income received by his relatives in this household	If Male—Did he ever serve in the U. S. Armed Forces during—						
26	27	28	29	30	31a	31b	31c	32a	32b	32c	33a	33b	33c
<i>10</i>	<i>Yes</i>	<i>Yes</i>	<i>1</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
	<i>No</i>	<i>No</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>No</i>	<i>No</i>	<i>No</i>
	<i>Yes</i>	<i>Yes</i>	<i>1</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
	<i>No</i>	<i>No</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>No</i>	<i>No</i>	<i>No</i>
	<i>Yes</i>	<i>Yes</i>	<i>1</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
	<i>No</i>	<i>No</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>No</i>	<i>No</i>	<i>No</i>
	<i>Yes</i>	<i>Yes</i>	<i>1</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
	<i>No</i>	<i>No</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>None</i>	<i>No</i>	<i>No</i>	<i>No</i>

34. To noncooperator: If worked last year (1 or more weeks in 1949) or in 1950, was he married in 1949 or 1950?  
 Yes—Skip to item 36  
 No—This subject is item 26, 28, and 35

35. What kind of work did this person do in his last job?  
*Cashier*

36. If ever married (Mar, Wid, D, or Sep in item 12): Has this person been married more than once?  
 Yes  
 No

37. If Mar.—How many years since this person was (last) married?  
 If Wid.—How many years since this person was widowed?  
 If D.—How many years since this person was divorced?  
 If Sep.—How many years since this person was separated?  
 2 years, or  Less than 1 year

38. If female and ever married (Mar, Wid, D, or Sep in item 20c): How many children has she ever born, not counting stillbirths?  
 1 children, or  None

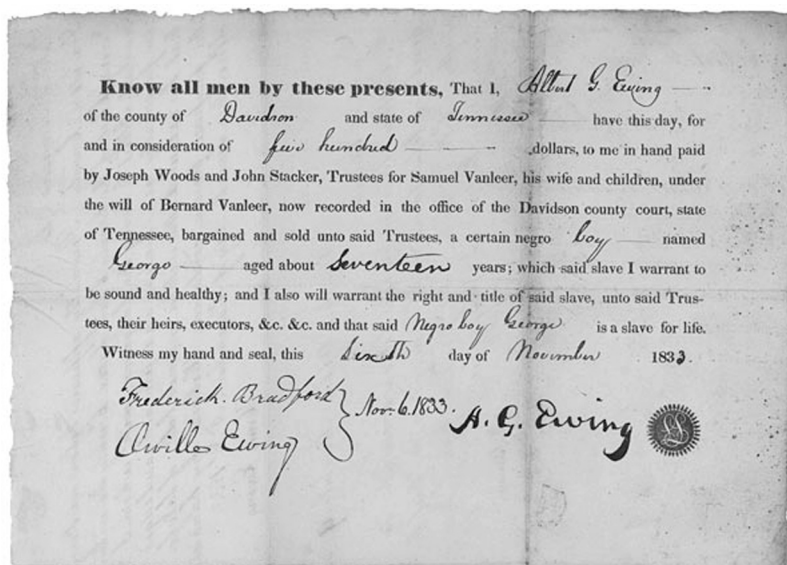
Figure 14.2 Facsimile of 1950 Census Schedule for Orange City, Iowa

data set contained 5009 records, and included information on the characteristics of the slave (e.g., age, sex, occupation, color), the terms of the sale (e.g., the date, price, whether paid in cash, the number of slaves sold together), and information on the buyer and seller.

Fogel and Engerman used the data to analyze the inter-state slave trade, and to address questions about the economic viability of the slave economy (Fogel and Engerman, 1974).

The work of building the corpus of machine-readable databases began in the





**Figure 14.3** Slave Bill of Sale, Davidson County, Tennessee, 1833

1960s, and continues both with small compilations and large collaborative data projects. In addition to the IPUMS project mentioned above, one can find large-scale historical data compilations of cost-of-living studies, election results, crime data, and the records of the heights and weights of people in the past. The creation and retrieval of historical data has also led to revision and improvement of data series compiled in earlier years and to the analysis of the history of data development. Most recently, for example, economic historians have produced a new 'millennial' edition of the *Historical Statistics of the United States* (Carter et al., 2006), which promises to provide opportunities for even more quantitative historical analysis.

## **ANALYZING HISTORICAL DATA SETS**

### ***Sampling and the Universe of Cases***

As have data analysts in the other social sciences, historians have made use of the theory of probability sampling to reduce the volume of information for a particular study

to a manageable level. Just as one does not need to survey the entire electorate to develop quite precise estimates of the ultimate election results, so historians studying family structure or economic activity or consumer behavior have not had to record all such behavior for study. As noted above, the process of creating historical data sets is sufficiently time-consuming to strongly recommend sampling strategies designed to reduce the volume of coding and data entry to the minimum necessary for robust analysis. Thus the original users and secondary users of the archived historical data sets need to attend to sampling strategy and introduce appropriate sample weights and measures of error into the analysis.

A more difficult issue is the one facing the historian who cannot be sure that she knows what the universe of cases actually is. Do the extant newspaper reports of lynchings, for example, encompass all lynchings (Griffin et al., 1997; Tolnay and Beck, 1995)? Are the records of wills filed with a particular county complete, or might some have been destroyed or lost over the centuries? These dilemmas have their analogues in non-quantitative

research. But as with code-book creation, the research must provide best estimates of answers to such questions before analysis, and a substantial methodological literature has emerged to address the issues, often with specific reference to the kind of data set being compiled.

### ***Techniques of Analysis***

Statistical analysis of historical data has ranged from elementary data analysis of the patterns of central tendency and dispersion of the phenomena under study to elaborate explanatory models of events and behavior. Much historical quantitative analysis has been descriptive, simply excavating and documenting patterns of change and activity in quantitative form that cannot be revealed by traditional historical analysis. Thus much work—important work—is simple counting of a phenomenon, and describing trends over time.

Somewhat more elaborate analysis involves determining the correlates of the phenomenon under study, or building a model to explicate more complex patterns in the data. Here the standard bivariate and multivariate techniques of statistics provide the tools necessary for the analysis. Quantitative historians have borrowed heavily from sociology, political science, demography and economics, and made use of the classic linear regression model and its variants as the workhorse technique for more complex analysis. Statistical packages, such as SPSS, SAS, STATA and the like underpin the analysis of quantitative historical work, as they do for the social sciences.

There is some evidence that quantitative history has begun to have an impact on the larger methodological practice of the social sciences, as quantitative historians have brought their methodological expertise to the social sciences. Two brief examples should illustrate that impact.

The first is development of the field of ecological regression, particularly for analysis of electoral patterns. Political scientists can supplement analysis of election results

with surveys of individual voters. Indeed the National Election Survey, conducted since 1948, has itself become an historical source of changing electoral behavior. But historians cannot go back and survey voters from the election of 1860, and thus must make use of the aggregate election results and the ecological characteristics of the voting units—e.g., precincts, districts or counties—that provided the vote. Ecological inference suffers from the threat of the ecological fallacy, that is, the danger of wrongly inferring individual level behavior from the patterns of aggregates. Practitioners of quantitative history have taken up new methods developed by political scientists and have devoted good effort to minimizing, if not completely solving, this dilemma. With historically minded political scientists, they have produced a methodological literature and new techniques that have produced rigorous results.<sup>14</sup>

The second contribution is serious attention to the development of statistical techniques to conceptualize and model time and temporal explanations. The methodological bread and butter for all historians is 'thinking in time' (Neustadt and May, 1986), and that standpoint has prompted historians and historically attuned social scientists to think about how to develop techniques of statistical analysis suitable for the goals of historical analysis.

Historians think about questions of what is an event, how is it bounded and measured; what is a turning point; what is a transition; what is a conjuncture or a rupture; and how is a period of time organized and bounded. Economists and other social and biological scientists have developed techniques to measure time series and temporal and cyclical events, for example, life cycles. The entry of quantitative historians into these discourses has been a useful clarification of the methodological issues involved. For example, the phrase, 'longitudinal analysis' that social scientists use does not necessarily privilege time as a central concern for analysis. Historians and social scientists who make temporal analysis such a central concern have thus

argued for the need to add methods that will address 'thinking in time' to the standard repertoire of statistical techniques. Such techniques as sequence analysis, event history analysis and the methodological discussions surrounding autocorrelation in time series analysis have usefully been enriched by the growth of the field of quantitative history.<sup>15</sup>

### THE COSTS OF DOING QUANTITATIVE HISTORY

The cost of scholarly work in quantitative history, like the cost of all scholarly work, can be measured in terms of both time and money required for the scholarship to flourish. The largest change in the working environment since the 1960s is that computing costs, which were quite expensive in the early years of the field, have dropped as the larger information revolution has developed. To my knowledge, there is no extant scholarly analysis of the costs of quantitative history versus traditional history, though I suspect that the underlying funding situation for quantitative historians has had an effect on the progress of the field.

In the early years of the development of quantitative history, in the United States the Social Science Research Council, the American Historical Association, the National Endowment for the Humanities, and the National Science Foundation, as well as research universities around the country, all provided sponsorship of the field by funding grants for data development, conference sponsorship and the institutional work required to promote the field. This early institutional support was aimed at jump-starting the field, not at providing sustained long-term support. Related to this, the National Endowment for the Humanities, the main federally sponsored grant agency for historians, has a much lower funding level than the federal funding agencies that support related social science research—for example, the National Science Foundation or the National Institutes of Health. The United States, unlike European nations, does not

include history in the main governmental foundation for funding academic research. Accordingly, quantitative history projects in the United States have had major difficulty in competing with both large-scale humanities grant projects, such as compilations of archival papers, and with large-scale long-term research projects in the social sciences such as the National Election Study or the Panel Study of Income Dynamics. Allan Bogue (1983) identified the chronic problems of funding faced by quantitative historians in the late 1970s. They remain unsolved as the concrete example which follows illustrates.

Robert Fogel, by any measure, represents one of the most successful and innovative quantitative history scholars in the field, yet even he has faced major funding obstacles. Fogel was awarded the Nobel Prize in Economics in 1993, and in his autobiographical statement prepared for the award, he described his career and acknowledged the problems of funding he faced, particularly, as he put it, for the 'current research projects on which I reported in the Prize Lecture'. The Center for Population Economics at the University of Chicago and the Walgreen Chair provided funding when federal grants would not. 'The data on health conditions', he wrote:

comes from a project called 'Early Indicators of Later Work Levels, Disease, & Death' which is tracing nearly 40,000 Union Army men from the cradle to the grave. It takes over 15,000 variables to describe the life-cycle history of one of these men. These life-cycle histories are created by linking about a score of data sets. It took more than half a decade of work to investigate the potential of these data sets, work out procedures for data retrieval and file management, and to establish the feasibility of the enterprise in our own minds.

The site committee of the National Institutes of Health which reviewed the original project proposal in 1986 agreed that such a project could in principle make a significant contribution to an understanding of the process of aging, but they were skeptical about the quality of some of the data, about whether the software and programming procedures we had developed by that time were adequate for the management of such a large data set, and about whether the project could be completed within the proposed budget.

To resolve these doubts it was necessary to draw a six percent subsample which linked together all of the separate sources and which demonstrated the effectiveness of the software by analyzing the information in the subsample. It took an additional four years to complete the second phase of the justification of the project. Thus nearly a decade of preliminary research, much of it funded by Walgreen and the CPE, was required before the project was accepted by the peer reviewers of NIH and NSF.<sup>16</sup> (Fogel, 1993)

Despite such barriers, quantitative historians have been able to take advantage of the technological developments in computing and data management to make major advances in the ease of analysis, in terms of both time and money. For example, historians of the 1960s through the 1980s who wished to have access to the archived data sets at ICPSR had to order tapes and paper code-books which were delivered by mail. The tape was then mounted on a mainframe computer, to be accessed in a statistical package run in a mainframe environment (with computer usage often charged by the university in the same way that phones or paper were charged). By the early 1990s, users could access files using FTP (file transfer protocol), and micro-computers on university desktops were providing direct access to statistical packages, even if those programs were sometimes still lodged on a mainframe. By the mid-1990s, desktop computing had replaced mainframe computing for most applications, and by the early 2000s, ICPSR initiated ICPSR Direct, the application that permitted an authorized user to download data files and PDF code-books directly to a desktop.

## CRITIQUES OF QUANTITATIVE HISTORY

From the outset of the development of the field of quantitative history, powerful critics have challenged practitioners on their work, and even challenged the usefulness of the field itself. In the early 1960s, Carl Bridenbaugh devoted a portion of his 1962 American Historical Association Presidential Address to a condemnation of quantitative history

(Bridenbaugh, 1962), memorably labeling it a 'bitch goddess' (Bogue, 1983). Even during the period of the rapid growth of quantitative history in the 1960s and 1970s, 'traditional' historians expressed doubts about the new methods, challenging them as reductionist, brittle and not pertinent to the main goal of the historical narrative. Critics were extremely dubious of the 'scientific' claims of quantitative historians, and resisted the challenge of the quantifiers that traditional historical writing was not theoretically rigorous or conceptually consistent.

In the 1980s, some of the original proponents of the field also renounced their earlier enthusiasm and suggested that quantitative methods had not fulfilled their promise. Most notable among these critics were Lee Benson and Lawrence Stone, early enthusiasts who had changed their minds (Benson, 1984; Stone, 1977, 1979). Such recantations gave support to the anti-quantifiers at a time when major new methodological challenges were facing historians, most notably from the postmodernists and what came to be called 'the cultural turn'. Through this welter of debate, quantitative practitioners continued their efforts, somewhat chastened by their fall from the heights of fashion of earlier years, but grounded sufficiently institutionally and intellectually to continue to work.<sup>17</sup>

Through some twenty years of debate, neither side of the traditional/quantitative divide 'won' their arguments. Rather, by the 1990s, the debate cooled into something of an uneasy truce, with practitioners acknowledging some of the points of their opponents, but agreeing to disagree on the larger validity of their enterprise.<sup>18</sup> In practical terms, quantitative techniques did not become a routine part of history graduate student training as they did in the social sciences, but have remained a specialty of some historians in some graduate training programs, considered more akin to language requirements for reading historical literature and texts of a non-English-speaking society than to a methodological necessity for all practicing historians. This compartmentalization of the skills of quantification for historians has in turn affected

the practice of quantitative historians within the larger history profession.

History as a field has maintained its roots as a 'humanities' discipline and quantitative historians' connections to the social sciences seem to many to be a betrayal of the historical project. The methodological 'training gap' has meant that when quantitative historians research and write for other historians, as opposed to other social scientists, they cannot expect their readers to appreciate or even understand the technical issues involved in their work. The history profession has maintained its commitment to accessible writing as well, and thus when writing for the broader audience of historians, quantitative historians have had to avoid technical jargon—for example, by avoiding the use of variable names in the explication of a model—and be mindful to explicate their arguments clearly.

The critiques have also encouraged quantitative historians to attend to the limitations in statistical methodology for analyzing historical processes, as discussed above. Much of this new work on statistical techniques for analyzing temporal processes is still in development and has yet to provide enough empirical work to demonstrate the robust nature of the new techniques, and hence convince non-quantitative historians, as well as the larger social science community, of the need to integrate explicitly temporal analysis into basic methods. But the promise is there, and as noted below, there are encouraging signs on the horizon.

### THE FUTURE OF THE FIELD

The intellectual achievements of quantitative history in conjunction with the larger information technology revolution makes the prognosis for the future of the field better today than it has been for many years.<sup>19</sup> Almost a half-century on, one can look back at steady development, though not always in a satisfyingly linear pattern.<sup>20</sup> Perhaps the most interesting recent development is the impact of the information technology revolution on the larger practice of historians.

When quantitative history as a field was in its most rapid initial development, most traditional historians labored much as their nineteenth-century predecessors had with pen, pencil, typewriter and note-card as technological support. Bibliographic work entailed using library card catalogs or reading large indexed tomes of articles, books, compilations, and the like. 'Data management' meant developing a file of index cards, not an electronic spread sheet or database. Secretaries typed manuscripts for publication, and though some large research institutions had introduced line editors for manuscript production by the 1970s, these were machines for staff, not faculty or students. By the 1980s, the situation changed. Desktop computers proliferated and for most historians, word processing opened up the possibility of the electronic future. By the 1990s, email replaced typed letters. After 1995, the content on the internet exploded, and first bibliographical work, and then much actual archival work, shifted to a computerized format. In short, non-quantitative historians had come to operate in a technological environment that was very similar to their quantifying peers. Most recently, cheap computing has made multimedia evidence—visual and oral, video and audio—accessible to the practicing historian. One can see these developments in particularly acute form in the developing field of historical geographic information systems, or historical GIS. GIS, was until quite recently, a very expensive technology, and thus adding historical maps to geographic databases has only just begun. As with the digitizing projects of the 1960s and 1970s, the payoff for the large initial costs of first translating maps to a new medium to become 'data', and then the development of new theory, software programs and methods to make the best use of these new data, are just beginning (Knowles, 2006).

More broadly, the effect of these technological changes has been to produce a convergence of work of what one might call 'technologically enabled' history. Traditional historians and humanists in general—for example, in the work of Franco Moretti

(2005)—also now work with electronic databases, learn new computer programs to analyze the rapidly proliferating data, and explore new forms of presentation of the results of their analysis. Quantitative historians had to learn the skills necessary to prepare and present statistical results in print. Historians more generally are using visual images, audio and video in their presentations, not as 'illustration' to enhance or supplement an analysis but as core evidence for analysis.<sup>21</sup>

Richard Steckel (2005) recently proposed an agenda for what he called 'Big Social Science History', which would extend the capacities of quantitative history and translate some of its methods of work to non-quantitative projects.<sup>22</sup> Andrew Abbott (2005) has also proposed such possibilities. As with the first generation of quantitative history, these large agendas will require collaborative efforts to manage the enormously expanding data infrastructure and the myriad computer technologies required to make best use of the expanding corpus of digitized historical evidence, and to develop appropriate theoretical approaches to such historical work.

## ACKNOWLEDGEMENTS

\* This essay has been improved considerably by comments from colleagues, particularly Peggy Adams, Erik Austin, Morgan Kousser, Jim Oberly, Lex Renda, Jack Reynolds, Steve Ruggles, Carole Shammass and Dan Scott Smith. The overall interpretation and remaining errors are all mine.

## REFERENCES AND SELECT BIBLIOGRAPHY

Abbott, Andrew (2001) *Time Matters: On History and Method*. Chicago: University of Chicago Press.

——— (2005) 'Looking backward and looking forward, social science history at 2000' in Harvey Graff, Leslie Page Moch and Philip McMichael with Julia Woesthoff (eds), *Looking Backward and Looking Forward: Perspectives on Social Science History*. Madison: University of Wisconsin Press, pp. 69–72.

- Abbott, Andrew and Tsay, A. (2000) 'Sequence analysis and optimal matching methods in sociology', *Sociological Methods and Research*, 29: 3–33.
- Adams, Margaret (1995) 'Punch card records: Precursors of electronic records', *American Archivist*, 58 (Spring): 182–201.
- (forthcoming) 'Analyzing archives and finding facts: Use and users of digital data records', *Archival Science*.
- Alter, George (1988) *Family and the Female Life Course*. Madison: University of Wisconsin Press.
- Alter, George and Gutmann, Myron (1999) 'Casting spells: Database concepts for event-history analysis', *Historical Methods*, 32 (4): 165–76.
- Ambacher, Bruce I. (ed.) (2003) *Thirty Years of Electronic Records*. Lanhan, MD: The Scarecrow Press.
- Annual Report of the American Historical Association* (1893).
- Aydelotte, William, Bogue, Allan and Fogel, Robert (eds) (1972) *The Dimensions of Quantitative Research in History*. Princeton: Princeton University Press.
- Benson, Lee (1957) 'Research problems in American political historiography', in Mirra Komarovsky (ed.), *Common Frontiers of the Social Sciences*. Glencoe, IL: Free Press, pp. 113–83, 418–21.
- (1961) *The Concept of Jacksonian Democracy: New York as a Test Case*. Princeton: Princeton University Press.
- (1984) 'The mistransference fallacy in explanations of human behavior', *Historical Methods*, 17 (3): 118–31.
- Bogue, Allan G. (1983) *Clio and the Bitch Goddess: Quantification in American Political History*. Beverly Hills: Sage Publications.
- (1986) 'Systematic revisionism and a generation of ferment in American history', *Journal of Contemporary History*, 21 (2): 135–62.
- (1990) 'The quest for numeracy: Data and methods in American political history', *Journal of Interdisciplinary History*, 21 (1): 89–116.
- Bourke, Paul, DeBats, Donald and Phelan, Thomas (2001) 'Comparing individual-level voting returns with aggregates: A historical appraisal of the King solution', *Historical Methods*, 34 (3): 127–34.
- Bridenbaugh, Carl (1962) 'The Great Mutation', AHA Presidential Address. Available at: [http://www.historians.org/info/AHA\\_History/cbridenbaugh.htm](http://www.historians.org/info/AHA_History/cbridenbaugh.htm). (Accessed on May 15 2006.)
- Burnham, Walter Dean (1955) *Presidential Ballots, 1836–1892*. Baltimore: Johns Hopkins University Press.
- (1970) *Critical Elections and the Mainsprings of American Politics*. New York: W.W. Norton & Co.

- Burton, Orville Vernon (2002) *Computing in the Social Sciences and Humanities*. Urbana: University of Illinois Press.
- Cameron, Sonja and Richardson, Sara (2005) *Using Computers in History*. New York: Palgrave Macmillan.
- Carter, Susan, Gartner, Scott, Haines, Michael R., Olmstead, Alan, Sutch, Richard and Wright, Gavin (2006) *Historical Statistics of the United States: Millennial Edition*. New York: Cambridge University Press.
- Chambers, William Nisbet and Burnham, Walter Dean (eds) (1967) *The American Party Systems: Stages of Political Development*. New York: Oxford University Press.
- Chopwhittle, Darcy and Taleglad, Lars Mooson (2001) 'Review of Philinda Blank's *When the Cows Come Home: Barn Architecture and Changes in Bovine Public Space*', *Social Science History*, 25 (5): 609–14. (See note 19.)
- Conrad, Alfred and Meyer, John (1958) 'The economics of slavery in the ante-bellum South', *Journal of Political Economy*, 66 (2): 95–130.
- Curti, Merle, Curti, Margaret Wooster, Daniel, Robert, Livermore Jr., Shaw, and Van Hise, Joseph (1959) *The Making of an American Community: A Case Study of Democracy in a Frontier County*. Stanford: Stanford University Press.
- Darcy, R. and Rohrs, Richard C. (1995) *A Guide to Quantitative History*. Westport, CT: Praeger.
- Dollar, Charles M. and Jensen, Richard J. (1971) *Historian's Guide to Statistics: Quantitative Analysis and Historical Research*. New York: Holt, Rinehart and Winston, Inc.
- Feinstein, Charles and Thomas, Mark (2002) *Making History Count: A Primer in Quantitative Methods for Historians*. New York: Cambridge University Press.
- Fishbein, Meyer H. (ed.) (1973) *The National Archives and Statistical Research*. Athens, OH: Ohio University Press.
- Fitch, Nancy (1986) 'Statistical fantasies and historical facts: History in crisis and its methodological implications', *Historical Methods*, 18 (4): 239–54.
- Floud, Roderick (1972) *An Introduction to Quantitative Methods for Historians*. Princeton: Princeton University Press.
- Floud, Roderick, Wachter, Kenneth and Gregory, Annabel (eds) (1990) *Height, Health and History: Nutritional Status in the United Kingdom, 1750–1980*. New York: Cambridge University Press.
- Fogel, Robert W. (1964) *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press.
- (1993) Nobel Prize in Economics Autobiographical Statement. Available at: <http://www.nobelprize.org/economics/laureates/1993/fogel-autobio.html>. (Accessed on May 15, 2006.)
- (2003) *The Slavery Debates: A Retrospective*. Baton Rouge: Louisiana State University Press.
- (2004) *The Escape from Hunger and Premature Death, 1700–2100: Europe, America, and the Third World*. New York: Cambridge University Press.
- Fogel, Robert and Costa, Dora L. (1997) 'A theory of technophysio-evolution, with some implications for forecasting population, health care costs, and pension costs', *Demography*, 34 (1): 49–66.
- Fogel, Robert William and Elton, G.R. (1983) *Which Road to the Past?* New Haven: Yale University Press.
- Fogel, Robert and Engerman, Stanley (1974) *Time on the Cross*, Vols. 1 & 2. Boston: Little Brown.
- Goldin, Claudia (1997) 'Exploring the "Present Through the Past": Career and family across the last century', *The American Economic Review*, 87 (2): 396–9.
- Graff, Harvey, Moch, Leslie Page and McMichael, Philip with Woesthoff, Julia (eds) (2005) *Looking Backward and Looking Forward: Perspectives on Social Science History*. Madison: University of Wisconsin Press.
- Greif, Avner (1997) 'Cliometrics after 40 Years', *The American Economic Review*, 87 (2): 400–3.
- Griffin, Larry (1993) 'Narrative, event-structure analysis, and causal interpretation in historical sociology', *American Journal of Sociology* 98: 1094–133.
- Griffin, Larry and Isaac, Larry W. (1992) 'Recursive regression and the historical use of "time" in time-series analysis of historical process', *Historical Methods*, 25: 166–79.
- Griffin, Larry J., Clark, P. and Sandberg, J. (1997) 'Narrative and event: Historical sociology and lynching', in W. Fitzhugh Brundage (ed.), *Under Sentence of Death: Lynching in the New South*. Chapel Hill: University of North Carolina Press, pp. 24–47.
- Gutmann, Myron and Alter, George (1993) 'Family reconstitution as event history analysis', in David Sven Reher and Roger S. Schofield (eds), *Old and New Methods in Historical Demography*. Oxford: Oxford University Press, pp. 159–77.
- Hacker, David and Fitch, Catherine (eds) (2003a) 'Building Historical Data Infrastructure: New Projects of the Minnesota Population Center', *Historical Methods* 36 (1).
- (eds) (2003b) 'Building Historical Data Infrastructure: New Projects of the Minnesota Population Center', *Historical Methods* 36 (2).
- Harvey, Charles and Press, Jon (1996) *Databases in Historical Research: Theory, Methods and Applications*. London: Macmillan Press Ltd.
- Haskins, Loren and Jeffrey, Kirk (1990) *Understanding Quantitative History*. New York: McGraw-Hill.
- Heckman, James J. (1997) 'The value of quantitative evidence on the effect of the past on the present', *The American Economic Review*, 87 (2): 404–8.

- Hudson, Pat (2000) *History by Numbers: An Introduction to Quantitative Approaches*. New York: Oxford University Press.
- Isaac, Larry W. and Griffin, Larry (1989) 'Ahistoricism in time-series analyses of historical process: Critique, redirection, and illustrations from US labor history', *American Sociological Review*, 54: 873–90.
- Jarausch, Konrad and Hardy, Kenneth (1991) *Quantitative Methods for Historians: A Guide to Research, Data, and Statistics*. Chapel Hill: University of North Carolina Press.
- King, Gary (1997) *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ.: Princeton University Press.
- Kousser, J. Morgan (1973) 'Ecological regression and the analysis of past politics', *The Journal of Interdisciplinary History*, 4 (2): 237–62.
- (1974) *The Shaping of Southern Politics: Suffrage Restriction and the Establishment of the One-party South, 1880–1910*. New Haven: Yale University Press.
- (1984) 'The revivalism of narrative: A response to recent criticisms of quantitative history', *Social Science History*, 8 (2): 133–49.
- (1986) 'Must historians regress?' *Historical Methods* 19 (2): 62–81.
- (1989) 'The state of social science history in the late 1980s', *Historical Methods*, 22: 13–20.
- (2001a) 'Evaluating ecological inference: An introduction', *Historical Methods*, 34 (3): 100.
- (2001b) 'Ecological inference from Goodman to King', *Historical Methods*, 34 (3): 101–26.
- Knowles, Anne Kelly (2006) 'GIS and history', Paper presented at the Annual Meeting of the American Historical Association, Philadelphia.
- Lemisch, Jesse (1967) 'The American Revolution seen from the bottom up', in Barton Bernstein (ed.), *Towards a New Past: Dissenting Essays in American History*. New York: Vintage Books, pp. 3–45.
- Lewis, Jeffrey B. (2001) 'Understanding King's ecological inference model: A method-of-moments approach', *Historical Methods* 34 (4): 170–88.
- Lorwin, Val and Price, Jacob (1972) *The Dimensions of the Past: Materials, Problems, and Opportunities for Quantitative Work in History*. New Haven: Yale University Press.
- McCormick, Richard P. (1966) *The Second American Party System; Party Formation in the Jacksonian Era*. Chapel Hill: University of North Carolina Press.
- McDonald, Terrence (1986) *The Parameters of Urban Fiscal Policy: Socioeconomic Change and Political Culture in San Francisco, 1860–1906*. Berkeley: University of California Press.
- Meyer, John R. (1997) 'Notes on cliometrics' fortieth', *The American Economic Review*, 87 (2): 409–11.
- Monkkonen, Eric (2001) *Murder in New York City*. Berkeley: University of California.
- Neustadt, Richard E. and May, Ernest R. (1986) *Thinking in Time: The Uses of History for Decision Makers*. New York: Free Press.
- Moretti, Franco (2005) *Graphs, Maps, Trees: Abstract Models for a Literary History*. New York: Verso.
- North, Douglass C. (1997) 'Cliometrics – 40 years later', *The American Economic Review*, 87 (2): 412–14.
- Palmquist, Bradley (2001) 'Unlocking the aggregate data past – Which keys fit?' *Historical Methods*, 34 (4): 159–69.
- Redding, Kent and James, David R. (2001) 'Estimating levels and modeling determinants of Black and White voter turnout in the South, 1880 to 1912', *Historical Methods* 34 (4): 141–58.
- Reiff, Janice (1991) *Structuring the Past: The Use of Computers in History*. Washington, DC: American Historical Association.
- Reynolds, John F. (1998) 'Do historians count anymore', *Historical Methods*, 31 (4): 141–8.
- Rowney, Don Karl and Graham, James Q. Jr. (eds) (1969) *Quantitative History: Selected Readings in the Quantitative Analysis of Historical Data*. Homewood, Ill.: Dorsey Press.
- Shreibman, Susan, Siemens, Ray and Unsworth, John (2004) *A Companion to Digital Humanities*. Malden, MA: Blackwell Publishing.
- Shammas, Carole (1990) *The Pre-industrial Consumer in England and America*. Oxford: Clarendon Press.
- Shammas, Carole, Salmon, Marylynn and Dahlin, Michel (1987) *Inheritance in America from Colonial Times to the Present*. New Brunswick, NJ: Rutgers University Press.
- Shorter, Edward (1971) *The Historian and the Computer: A Practical Guide*. Englewood Cliffs, NJ: Prentiss Hall.
- Silbey, Joel, Bogue, Allan and Flanigan, William (eds) *The History of American Electoral Behavior*. Princeton: Princeton University Press.
- Smith, Dan Scott (1984) 'A mean and random past: The implications of variance for history', *Historical Methods*, 17: 14– 48.
- (1992) 'Context, time, history', in Peter Karsten and John Modell (eds), *Theory, Method, and Practice in Social and Cultural History*. New York: New York University Press, pp. 13–32.
- Steckel, Richard (2005) 'Presidential Address', Annual Meeting of the Social Science History Association, Portland, Oregon.
- Steckel, Richard H. and Floud, Roderick (eds) (1997) *Health and Welfare During Industrialization*. Chicago, IL: University of Chicago Press.



- Steckel, Richard and Rose, Jerome (eds) (2002) *The Backbone of History: Health and Nutrition in the Western Hemisphere*. New York: Cambridge University Press.
- Stone, Lawrence (1977) 'History and the social sciences in the twentieth century', in C. Delzell (ed.), *The Future of History*. Nashville: Vanderbilt University Press. pp. 3–42.
- (1979) 'The revival of narrative: Reflections on a new old history', *Past and Present*, 89: 3–24.
- Swierenga, Robert (ed.) (1970) *Quantification in American History: Theory and Research*. New York: Atheneum.
- Tolnay, Stewart E. and Beck, E.M. (1995) *A Festival of Violence: An Analysis of the Lynching of African-Americans in the American South, 1882–1930*. Urbana: University of Illinois Press.
- Turner, Frederick Jackson (1893) 'The Significance of the Frontier in American History', *Annual Report of the American Historical Association*: 199–227.
- Whaples, Robert (1991) 'A quantitative history of the Journal of Economic History and the cliometric revolution', *Journal of Economic History*, 51 (2): 289–301.

## NOTES

1 See, for example, Benson (1957; 1961); Burnham (1970); Chambers and Burnham (1967); Richard P. McCormick (1966).

2 For information on the Cambridge Group, see their website, <http://www-hpss.geog.cam.ac.uk>.

3 See also Kousser (1989) and Reynolds (1998).

4 See for example, Darcy and Rohrs (1995); Dollar and Jensen (1971); Feinstein and Thomas (2002); Floud (1972); Haskins and Jeffrey (1990); Hudson (2000); Jarausch and Hardy (1991); Shorter (1971).

5 See, for example, Aydelotte et al. (1972); Lorwin and Price (1972); Rowney and Graham (1969); Silbey et al. (1978); Swierenga (1970).

6 ICPSR, founded in 1962 as ICPR, changed its name to the Inter-university Consortium for Political and Social Research (ICPSR) in 1975.

7 See, for example, Floud et al. (1990); Monkkonen (2001); Shamma et al. (1987); Shamma (1990); Steckel and Floud (1997); Steckel and Rose (2002).

8 The schedule in Table 14.1 is available on the IPUMS website at <http://www.ipums.umn.edu/usa/voliii/form1950.html>.

9 The United States maintains census schedules as confidential records for 72 years. The 1890 Census manuscript schedules were destroyed by fire in 1921.

10 See the special issues of *Historical Methods* (Hacker and Fitch 2003a; 2003b) on 'Building Historical Data Infrastructure: New Projects of the Minnesota Population Center' and the website of IPUMS at [www.ipums.org](http://www.ipums.org) for details.

11 National Archives and Records Administration, 'Inside the National Archives – Southeast Region, 1825-1863 Slave Sale Documents'. Available at: <http://www.archives.gov/southeast/exhibit/2.php>.

*Transcription of Slave Sale Document in Figure 14.3*

Know all men by these presents, That I, Albert G. Ewing, of the county of Davidson and state of Tennessee have this day for and in consideration of five hundred dollars, to me in hand paid by Joseph Woods and John Stacker, Trustees for Samuel Vanleer, his wife and children, under the will of Bernard Vanleer, now recorded in the office of the Davidson county court, state of Tennessee, bargained and sold unto said Trustees, a certain negro boy name George aged about seventeen years; which said slave I warrant to be sound and healthy; and I also will warrant the right and title of said slave, unto said Trustees, their heirs, executors, &c. &c. and that said negro boy George is a slave for life.

Witness my hand and seal, this Sixth day of November 1833.

A.G. Ewing

Frederick Bradford  
Orville Erving  
Nov. 6. 1833.

12 For other examples of slave sale documents, see the Slave Documents Collection from the Enoch Pratt Free Library, Baltimore, Maryland, available at <http://www.pratt.lib.md.us/exhibits/slavery/>

13 The data set and code-book are available at: <http://www.icpsr.umich.edu/cocoon/ICPSR-STUDY/07423.xml>.

14 See, for example, Kousser (1973, 1974). For recent methodological developments in the field and their impact in history, see King (1997), and the articles in the Summer and Fall 2001 (34 (3 & 4)) issues of *Historical Methods* on the time period by: Kousser (2001a, 2001b); Bourke et al. (2001); Redding and James (2001); Palmquist (2001); and Lewis (2001).

15 See Abbott (2001); Abbott and Tsay (2000); Alter and Gutmann (1999); Alter (1988), Gutmann and Alter (1993); Griffin (1993); Griffin and Isaac (1992); Isaac and Griffin (1989); Reher and Schofield (1993). On time series, see also McDonald (1986). On quantification and historical explanation, see Smith (1984; 1992).

16 For the results of this research, see Fogel (2004); Fogel and Costa (1997).

17 For discussion of Benson's change in position and critiques of the change, see Bogue (1986; 1990) and Kousser (1986). See also Fogel and Elton (1983); Kousser (1984); Fitch (1984); and Fogel (2003).

18 For a hilarious parody of the issues involved, see the Winter 2001 issue of *Social Science History*. Outgoing editors Paula Baker and Elizabeth Faue published reviews by Darcy Chopwhittle and Lars Mooson Taleglad of Philinda Blank's (2001) *When the Cows Come Home: Barn Architecture and Changes in Bovine Public Space* (2001).

(The reviewed book does not exist, though perhaps it might. Many people contributed to the review; Paula Baker and Elizabeth Faue take responsibility for it.)

19 On anthropometric history, for example, see the Summer 2004 Special Issue of *Social Science History*, Volume 28, no. 2, guest edited by John Komlos and Jorg Baten. For the impact of the IPUMS project, see the bibliography of work listed on the IPUMS website, <http://www.ipums.org>. For recent evaluations of 'social science history' as a field, see Graff et al. (2005).

20 For retrospectives on quantitative history, see Reynolds (1998). For retrospective analysis of 'cliometrics', see the special section of the 'Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association' in *The American Economic Review* (1997), 87 (2), on

'Cliometrics After 40 Years'. Papers in this section include Goldin (1997); Greif (1997); Heckman (1997); Meyer (1997); and North (1997). See also Whaples (1991).

21 See, for example, Burton (2002); Cameron and Richardson (2005); Harvey and Press (1996); Reiff (1991); Shreibman et al. (2004).

22 Steckel listed the large data projects social science historians have produced in the last generation and then added his own wish list: including an inventory all archeological sites; an inventory all artifacts at these sites; a database on natural disasters and human history; and an international catalogue of films and photos. He called for extending the digitization of all extant manuscript censuses in the past; a digitized and annotated collection of diaries; voting records at the precinct level; and probate records.

