

GUIDE TO THE MILLENNIAL EDITION

Monty Hindman and Richard Sutch

Editions and Copyright

Previous editions. This is the fourth edition of *Historical Statistics of the United States*. The U.S. Bureau of the Census published the prior editions in 1949, 1960, and 1975, the last known as the Bicentennial Edition. Cambridge University Press publishes this, the Millennial Edition, with the permission of the Census Bureau. Some of the data and table documentation presented here are used without explicit quotation, but with permission, from the earlier editions. The Census Bureau takes no responsibility for the design of this edition or the accuracy of its content, which rests solely with the contributors, the editors, and Cambridge University Press.

Electronic edition. This edition of *Historical Statistics of the United States* is available in electronic form from Cambridge University Press. A compact disk containing the Bicentennial Edition of *Historical Statistics of the United States* is also available from the Press.

Copyright. Permission to quote or reprint copyright material should be obtained directly from the copyright owner. Much of the data reproduced in this work were originally published by agencies of the U.S. government and are in the public domain. Generally speaking, original data that have been published elsewhere under copyright protection may be freely used for educational, scholarly, or journalistic purposes (but not commercial purposes) with proper citation to the original source under the fair use provision of U.S. copyright law. Cambridge University Press has made every effort to secure, where necessary, permission to reproduce protected material. In almost every case the permission requested was freely granted. In a few instances, however, the copyright owner requested a specific citation. These citations may be found in the listing of Copyright Citations at the end of Volume 5.

Data Revisions and Updates

Reproduction and revision of data from prior editions. Although this volume provides many data series from prior editions of *Historical Statistics of the United States*, users should be aware that some data from these editions have subsequently been revised. Our contributors sought to present the most recently available data, and thus users probably will wish to use the data presented here rather than that in previous editions. In some cases, data from the earlier editions were judged to be unreliable or obsolete and were not reproduced.

Data updates. The data series in *Historical Statistics of the United States* do not have a uniform end date; instead, each table reports the data available at the time the contributor compiled the data. Many series in these volumes are continued on a regular basis with periodic updates and revisions by the agency, group, or individual responsible for the original data. Figures for many of the current series are presented in the *Statistical Abstract of the United States*, published annually by the U.S. Bureau of the Census. The updating of industrial statistics will be complicated by the switch in 1997 from the Standard Industrial Classification (SIC) system to the North American Industrial Classification System (NAICS); see the Introduction to Part D.

Additional data. In many cases, additional data can be found in the source documents, in references mentioned in the table documentation or chapter essays, and through the Internet sites of the groups or agencies noted in the sources for the data presented here.

Errors. In a work as large as this, errors of both commission and omission are likely to have occurred. Users who discover errors are urged to communicate them to Cambridge University Press, 40 West 20th Street, New York, New York 10011-4211, USA.

Data Selection

General principles. The criteria for the selection of data to be included in this edition varied broadly, depending on the particular subject matter. Generally, summary measures or aggregates at gross levels and immediately below were given highest priority for inclusion. Below such levels, selection was governed by the interplay of the following: the amount of space already devoted to a particular subject; the attempt to achieve a relatively balanced presentation among subject fields; whether other data already covered a particular topic; the quantity and quality of the data available; and the extent to which the data might enhance the value of other material in the book. During the early phases of the project these selection criteria were conveyed to our contributors, upon whose judgment we ultimately relied.

Data reliability. Our contributors have attempted to select data that they consider to be generally reliable and to reproduce faithfully the data reported in their sources. They have also provided citations and technical descriptions to assist users in making independent assessments of both the data's reliability and their suitability for a project at hand.

Original versus derived data. Primary emphasis was placed on the presentation of original, unmodified figures rather than derived data because they offer greater flexibility to users. Derived data – for example, averages, percentages, ratios, and index numbers – were provided if they were the accepted standard for presentation (for example, unemployment rates), if the table contributor judged that the derived data would be particularly helpful, or if the use of derived data saved a significant amount of space.

Topical coverage. Because the last thirty years have witnessed the expansion of data collection into areas that were only inadequately covered, if at all, in the 1970s, this edition has a broader topical scope than its predecessors. A tentative list of topics emerged after extensive discussions between the project's editors in chief and Cambridge University Press. The outline was widely circulated to scholars, reference librarians, and government statistical bureaus. After a revision of that outline, the project recruited contributors, who offered additional suggestions. What emerged from this process was an outline for the project that was both designed by the profession and feasible to accomplish.

Temporal coverage. Contributors were asked to take the data series under their charge as far backward and forward in time as possible. They were also encouraged to include important lapsed series – those that begin and terminate in the past – because such series are sometimes available only in out-of-print documents. Most data series in *Historical Statistics of the United States* provide annual or decennial data spanning at least twenty years, with the main exceptions being for special topics (the colonial period and the Confederate States of America), for newly developed series providing the only data available to represent an important subject field, and for short series that served as important extensions of longer series.

Data frequency. Annual data were given preference for inclusion, but certain series are presented only for years in which a national census was conducted and, in some instances, only for scattered dates, as dictated by data availability. When both annual figures and benchmark data exist, both series are sometimes shown. A major exception was made for Chapter Cb, which presents many of its series on a monthly or quarterly basis. Although this volume mainly provides annual data, underlying data are sometimes available more frequently from the original sources.

Geographical coverage. The data in *Historical Statistics of the United States* generally cover the nation as a whole, defined by the recognized borders of the country for the year in question. As new states were admitted to the Union, the coverage of the typical statistical series in this volume expands to include the new additions, without any special notation in the table documentation. The documentation should be consulted to determine if such changes in the boundaries of the United States are likely to have affected the series. When the year of a state's inclusion in a series differs significantly from its year of statehood, this fact was noted in the documentation whenever possible. Refer to Appendix 2 for the dates of statehood.

Subnational data. Because of limitations of space, data are generally not shown for regions, states, or localities. The underlying sources sometimes provide data in finer geographical detail than shown here. Some tables provide data for U.S. census regions or divisions; see Appendix 2 for more information on such regional classifications.

Outlying areas. In almost all cases, outlying areas are not included in the national totals reported here. Refer to Chapter Ef for additional information on such areas.

Organization of the Volume

Arrangement of the data. In this edition of *Historical Statistics of the United States*, data are arranged by broad subjects in five parts, each published in a separate volume and each volume containing several chapters. The tables in most chapters are further organized into various subsections (see the Detailed Table of Contents in each volume).

Essays. Each chapter is introduced by one or more essays that provide a general guide to the data, the sources, and the historical trends that have been emphasized in the scholarly literature. They contain a list of references that may be consulted by those interested in more detail.

Series identifiers. Each data series is assigned a unique alphanumeric identifier. The two letters in the identifier indicate the chapter in which the series resides. Within a chapter, series are numbered sequentially. Sets of contiguous series are identified by means of a series range (for example, series Da42–47). Source citations and table documentation are linked to the data series by means of such identifiers, which may be preferred over page numbers for use in reference citations.

Table identifiers. An entire table is identified by the range of series that it contains. For example, the first two tables in the chapter on vital statistics contain ten and twenty series, respectively; thus, they are identified as Table Ab1–10 and Table Ab11–30. Similarly, a group of contiguous tables is identified by a series range. Using the same example, these two tables could be referred to jointly as Tables Ab1–30.

Table Documentation

Table contributors. Each table provides the names of the contributors who selected, collected, and described the data. The editorial staff also reviewed the data and table documentation for accuracy, completeness, and clarity of presentation.

Sources. In most cases, full citations are given for data sources; however, when numerous issues of a publication were used, the source citations are usually limited to “annual issues” or similar notations. When data are reproduced from the Bicentennial Edition, the source citation lists the original source rather than the Bicentennial Edition, except under special circumstances.

Unpublished data. Nearly all the data reported here have been previously published or accepted for publication. Rare exceptions for previously unpublished data were allowed if a contributor felt that the data were particularly important and if peer review accepted the data for inclusion.

Integrated Public Use Microdata Series. A number of series reported in this edition are extracted from the Integrated Public Use Microdata Series (IPUMS). The IPUMS is composed of representative samples drawn from the returns of the decennial censuses of the population. All censuses from 1850 to 1990 are included, with the exception of 1930, which is under development, and 1890, the manuscripts for which were destroyed by fire. The IPUMS data and documentation are available over the Internet.*

Internet sources. Some data series in *Historical Statistics of the United States* are based on electronic sources; however, owing to the fleeting nature of specific Internet addresses or Web-based

* Steven Ruggles, Matthew Sobek, et al., *Integrated Public Use Microdata Series: Version 2.0* (Historical Census Projects, University of Minnesota, 1997).

file names, we do not use them when identifying sources. Instead, we use more general phrasing to direct users to the Internet source.

Table documentation. Most tables are accompanied by documentation defining relevant terms and concepts, providing methodological and historical background, noting unusual values or comparability issues, explaining methods used to calculate derived data, and providing references to sources containing more detailed data or more extensive discussion. Unlike prior editions, which consolidated table documentation at the beginning of chapters, this edition locates the documentation with the tables, the intent being to increase its visibility, convenience, and thus use. Many tables are fully self-documenting, without cross references to other parts of this work; however, when cross references to other tables or essays are provided, the user is encouraged to follow those references.

Footnotes. There is no sharp demarcation between the type of information conveyed in the ordinary table documentation and that conveyed in the footnotes. Roughly speaking, footnotes are used for two purposes: to draw attention to issues of particular importance (footnotes as warnings) or to comment on matters related to specific columns, rows, or cells in a table.

Footnote order. Within a table, footnotes are numbered sequentially as follows: first the general footnotes that apply to the entire table; then left-to-right across the table header (the footnotes governing specific series); and finally footnotes attached to the table stub and the data area, proceeding in top-to-bottom, then left-to-right fashion (as used here, the directional terms apply to tables with standard page orientation). A footnote's first appearance within a table determines its position within the sequential numbering.

Total and subtotals. In most cases, a table's header structure will clearly indicate the total–subtotal relationships among the series. The typical practice in this volume is to provide the total series first, followed by its components. Often the sum of the components will equal the total, perhaps with small deviations attributable to rounding or other causes; however, sometimes the breakdowns provided in a table are not exhaustive, and the components will add to an amount less than the total. Users should consult the table documentation and exercise caution in this regard.

Race and ethnicity. Many tables provide disaggregations by race or ethnicity. This volume typically uses the terms “white,” “black,” “Asian” (or “Asian American”), “Indian” (or “Amerindian” or “Native American”), and “Hispanic.” Note that a person identified as Hispanic may be of any race. See the essay on definitions and measurement of race and ethnicity in the Introduction to Part A for a discussion of racial classification and identification as it applies to the collection of historical statistics in the United States.

Dates

Date ranges. Throughout the table documentation and the chapter essays, date ranges are inclusive: for example, 1964–1987 includes both 1964 and 1987.

Year of record. The identification of the year of record – in other words, the precise meaning of the years shown in a table stub – was complicated by the failure of some sources to state whether the data were prepared on a calendar year, fiscal year, or some other basis; by changes in the year of record over time; and, in some instances, by imprecision or silence in the source concerning the beginning or ending date for the year of record. Table contributors

attempted to clarify such matters, but ambiguity remains in some tables.

Transition quarters. Sometimes the year of record changes in the middle of a table, and values are provided for the “transition quarter” – the gap between the end of the old year of record and beginning of the new. In such cases, users will see a (TQ) designation in the table stub. Nearly all transition quarters in this volume are associated with the year 1976, when the federal government changed the end of its fiscal year from June 30 to September 30. In rare cases, the (TQ) designation will be for a transition period that is not actually a quarter, but some other fraction of a year.

Units, Measures, and Monetary Values

Units of measure. Series are usually expressed in the units reported in the original source. In some cases, however, units were converted to make two or more data series comparable, or to create a single series when splicing data from multiple sources. The approach taken in these volumes was to restrict the units information to true *measures* and to rely on the table title and layered headers to convey other details about the things being counted or measured. Sometimes series are expressed in units too complex for pithy statement; in these rare cases, a generic unit of measure is given, with further elaboration left to the table documentation.

Billion and trillion. The American and Canadian definitions of billion (10^9) and trillion (10^{12}) are used throughout, not the definitions used in England, Germany, and many other countries.

Index numbers. Some series are expressed in terms of index numbers. In such cases, the base period of the index is provided where the unit of measure would normally be found. For a discussion of index numbers, see the essay on prices and price indexes in Chapter Cc and the essay on national income and product in Chapter Ca.

Weights and measures. Most data series are expressed in American units (the U.S. Customary System) rather than metric units (the International System). For a discussion of these two systems and for conversion information, see Appendix 1.

Monetary values. Unless otherwise noted, monetary values are expressed in current or nominal terms – in other words, the actual historical values (usually U.S. dollars), not adjusted for previous or subsequent changes in prices. This standard was adopted to avoid attaching the word “current” or “nominal” to every reference to a monetary unit. When monetary values have been adjusted in some fashion, this is stated explicitly and the relevant base period is given. For a discussion of monetary values, see Appendix 1 and the essay on prices and price indexes in Chapter Cc.

Data Values

Data precision and significant digits. In making decisions regarding the precision with which data values should be presented, fidelity to sources was our primary consideration. Thus, the underlying data files for *Historical Statistics of the United States* – available in the electronic edition – retain the full precision provided by table contributors, even though this level of detail might be deemed excessive by scientific standards for the reporting of significant digits. In most cases, the detail comes straight from the sources themselves; therefore, exact reproduction provides a valuable check for researchers wanting to trace the provenance of a number or hunt down an anomaly. In other cases, excessive

precision comes from spreadsheet calculations made by table contributors (for example, in the computation of derived data). Here, too, we did not impose our judgments concerning the appropriate precision and instead retained the full detail provided by contributors. Users should note that historical sources sometimes change the precision with which they report data over time. Also, some tables contain series reported in the sources at different levels of detail but that, for ease of comparison, are provided here in consistent units. The usual indication of varying precision – whether in a single series or across multiple series within a table – is a run of data values with trailing zeros, either before or after the decimal point. In such cases, users will need to exercise judgment concerning the precision of the data.

Decimal precision for display purposes. While the underlying data files retain all of the detail provided by table contributors, the data displayed in the print edition of *Historical Statistics of the United States* are shown in rounded fashion, typically with no more than three digits following the decimal point. Similarly, tables generated for display purposes by the electronic edition are formatted using the same rounding conventions; however, the underlying files available for downloading provide the values at full precision.

Zero values and (Z). A zero in a data series means exactly that: a reported value of zero. In some cases, an underlying data value may be so small that it rounds to zero when displayed at the

level of decimal precision chosen for the series. In such cases, a (Z) marker is used rather than a zero value. Stated more precisely, the (Z) notation indicates a *nonzero value that is not shown or possibly not known*. In the former case – a nonzero value not shown – (Z) means that the value falls below the threshold of our rounding convention: the number rounds to zero, as displayed in this volume (full precision for such values is available through the electronic edition). In the latter case – a nonzero value not known – (Z) means that the original source did not provide a specific value. Owing to these complexities, the meaning of the (Z) marker is specifically documented in every table that uses the device.

Dash as a data value. The “—” marker means that a value is not being reported. There are several possible reasons: the data are not available anywhere; the data were not provided in the source but conceivably could be found with sufficient research; the data were available in the source but the table contributor decided that they should not be reported (for example, unreliable data); or the data might conceivably be reported as a zero, but the table contributor decided for conceptual reasons to represent it as “no value reported” (for example, if a category or program covered by the series did not yet exist). Some sources do not carefully distinguish between zero values and missing data. Table contributors attempted to eliminate such confusion, but in some cases the “—” marker could mean that the value, if shown, would be zero.