The Impact of Statistics on Government Policy: Statistics and the Census - A Brief Historical Journey

L. Billard*

Abstract

In this article, we trace a history of the United States census process since its U.S. Constitutional mandate from the first census in 1790 until the most recent census of 2010, as recorded primarily through the published records of the *Publications/Journal of the American Statistical Association* along with other relevant sources.

Key Words: Census, agriculture, industrial, vital statistics, Francis Walker, Carroll Davidson, Hollerith, data collection, undercounts, non-response

1. Introduction

The modern Census in the United States goes back to the 1787 US Constitution (drawn up in 1787, ratified in 1788, and entered into law in 1789). With reference to the Census, the Constitution includes the statement (see, e.g., Lunt, 1888) "The actual enumeration shall be made within three years after the first meeting of the Congress of the United States, and within every subsequent term of ten years, in such a manner as they shall by law direct."

Thus, it was that the first census took place in 1790, with an enumeration of 3,929,214 inhabitants occupying roughly 25% of today's area, a mostly rural population, and at a cost of $44,377.28.

When the plans for the 1840 census still depended on the antiquated methods of the first census of 1790, five men - William Cogswell, Richard Fletcher (first President), John Dix Fisher, Oliver Peabody, and Lemuel Shattuck (second President) - were so concerned they formed the American Statistical Society in December 1839 (name changed to the American Statistical Association, ASA, in 1840) to tackle this problem. The ASA and the census process were therefore intricately entwined over their respective histories. Indeed, as summarized in Billard (1998), "the journal writings of the 1800s were concentrated on the censuses in some way - how they were conducted, what information was to be sought, publication of the resulting statistical tables, what the data revealed." This thread - What, How, When - was to be collected and analysed permeates through all the censuses and the associated journal articles and census reports.

Today, everyone - citizens, legal residents, long-term visitors, illegal immigrants - everyone living in the US on census day should be counted (but, in reality, not everyone is ever actually counted). Census day became April 1 in 1930. Previously, it was mostly June or August, though curiously it was January 1 in 1920, after the shift to April, April 15, in 1910.

This article attempts to provide a profile of various aspects of the census history as reported primarily in the ASA journals along with other publications over the decades. In this brief journey, a broad brush is used, highlighting certain aspects along the census highway. Many historians have provided very detailed historical volumes of the census process; e.g., Anderson (2015), Anderson and Fienberg (1999),

*University of Georgia, Athens GA 30602

and Wright (1900). Lunt (1888) traces an early history highlighting the progression of statistics/data that were collected and for what reason.

We begin, in Section 2, with a brief outline of key components of the early censuses, primarily those up to 1910. One conclusion from that outline is the role played by agricultural, industrial and vital statistics questions beyond the simple mandate of population counts. Therefore, in Section 3, a deeper but still brief look is made on these three issues. Section 4 reviews the personna of Walker and Wright, giants who played key roles in the census processes of the late 1800's. The introduction of the Hollerith counting machines is discussed in Section 5. The role of ASA in establishing a permanent Census Bureau is the subject of Section 6. In Section 7, the methods employed to gather the census returns are outlined; while in Section 8, we are reminded that the accuracy of these returns cannot be guaranteed. Some concluding remarks are in Section 9.

2. Early Censuses

As the U.S. Secretary of State, Thomas Jefferson conducted the first Census in 1790 with an enumeration of the inhabitants. Going beyond the narrow strictures of the Constitution, census marshals classified

1. Names of heads of families,
2. Free white males sixteen and older,
3. Free white males under sixteen,
4. Free white females, including heads of families,
5. All other free persons,
6. Slaves.

For this and the next six censuses, the unit being counted was the family; in 1850, the unit became the individual.

The backdrop for these early censuses was that the nation was characterized by the so-called Jefferson agrarianism. While it is clear from the Constitutional mandate that the census is a population census, throughout the 1800's, the scope of the Census was enlarged, especially in the latter half of the century, as users of the data sought more and more data to expand their respective endeavors in addressing the needs of society. Indeed, it is possible to glean quite an interesting social history of the growth and expansion of the fledging nation by following the discussion as to what questions should be added to the census and why. At first, these questions were limited to reasonably simple notions such as "How many farms are there". Eventually, led first by Gen. Walker (see Section 4 below), there were demands that the census be a "scientific" census with questions added and designed to address substantive scientific questions of the day; most of these revolved around the ever-widening social constructs as researchers sought data to explain the composition and nature of the emerging new nation to itself.

In 1840, questions seeking statistics pertaining to universities and colleges were added. However, otherwise, this was to be the same "antiquated" census of the past; hence, the concerns leading to the formation of the ASA.

By 1850, there were over 100 inquiries with results assembled into six schedules, viz.,

1. *Free Inhabitants* - (many new features) - # dwellings, name/surname of every free white person, gender, age, color, education, marriage status, ...
2. *Slaves* - names of slave owners, age, gender, color, # fugitives, ...
3. *Mortality* - (new feature) - (details) of who died and why, ...
4. *Agriculture* - (improved coverage) - unimproved/improved acreage, stock, crops, value of farms and farm implements, ...
5. *Manufacturing Industry* - name/location of every establishment (to value of $500), # employed by gender, wage by gender, ...
6. *Social Statistics* - (new feature) - value of real and personal estate, taxes levied, details regarding schools, libraries, churches, newspapers, ...

Illustrating the still largely rural nature of the population, the final question revolved around the value of the average crop, what crops were short and to what extent. There was a marked change this time in the method of enumeration in that instead of heads of families being recorded, it was the individual person being recorded, a change that still exists today (Hunt, 1909).

Distracted by other events, the 1860 census essentially continued with the expanded set of questions and the modus operandi of the 1850 census, though there were publication delays in some areas.

Burdened with the inadequacies of the 1850 census laws, and under a newly appointed Superintendent General Walker, the 1870 census was not going to be easy. Walker wanted to modernize the process, to expand the census greatly, to enlighten the populace of the up-coming century-year-old nation with a vast exposition of its assets. While the House was supportive of his changes, ultimately the Bill that had to be approved with respect to all censuses was defeated in the Senate. Thus, with the exception of additional questions on machinery and machine power sources, census questions were restricted to those about population, vital statistics, wealth and industry, as in 1850.

However, after re-appointment as Superintendent for the 1880 census, though thwarted in 1870, Walker was determined to make changes for the 1880 census. His persistence paid off when many of his 1870 plans became practice. Primarily, he

(i) Wanted to establish a scientific census.

(ii) Replaced (politically appointed) marshals with Supervisors.

(iii) Appointed enumerators "without (political) affiliation", including women.

(iv) Introduced examinations for clerical staff.

(v) Added geographers to map the country systematically.

(vi) Greatly expanded the scope to produce an "encyclopedic census", where the number of questions asked across a wide swath of endeavors was nothing but mind-boggling most especially for the industrial sector; the results were reported in 22 volumes in 1880 - up from 4 volumes in one in 1870.

(vii) Questions covered "everything" – facts regarding population, agriculture, transportation, vital statistics, taxation, schools, public indebtedness, newspapers, shipbuilding. Plus forests, petroleum and building-stone industries, metals, cotton and other farm products, ...

At the time, it was considered the greatest, most elaborate, most valuable, census ever! When there were questions about apparent disparities in results for 1870, Walker went back and re-analysed relevant data; and where necessary, he made corrections to his own reports from 1870, such was his integrity, and such was his determination that the nation have a trustworthy census process.

Unarguably, the most characteristic feature of the 1890 census was the introduction of the Hollerith machine to record and tabulate census results; see Section 5. In most other respects, this census, now under Superintendent Carroll Wright, followed the lines established by Walker for the 1880 census. The number of questions continued to grow however, so much so, that by 1890, the number of inquiries had grown from four in 1790 to 13161, but this number was reduced to 7476 for 1900 (Wright, 1908).

With the act authorizing the 1910 census, the accepted practice to maintain confidentiality of census returns now became law. Later, 1978 legislation allowed for the National Archives and Records Administration to release census records after 72 years. This 72-year-rule was based on the 1952 estimate that a woman's life expectancy was 71.6 years. Presumably, this rule needs to be revised given today's life expectancy of 81+ years for American women.

Table 1: Population Counts and Costs by Census

Census No.	Census Year	Actual Cost	Cost in 2017 $'s	Population Count	Actual Cost per Person	2017 $'s Cost per Person
1	1790	44377.28	1144567.34	3929214	0.011	0.291
2	1800	66109.04	1235556	5308483	0.012	0.233
3	1810	178444.67	3384116.79	7239881	0.025	0.467
4	1820	208525.99	4137114.2	9633822	0.022	0.429
5	1830	378545.13	9571900.2	12866020	0.029	0.744
6	1840	833370.95	22389675.26	17069458	0.049	1.312
7	1850	1423351	42686861.96	23191876	0.061	1.841
8	1860	1969377	55210503.06	31443321	0.063	1.756
9	1870	3421198	61276843.74	38558371	0.089	1.589
10	1880	5790678	131010237.74	50155783	0.115	2.612
11	1890	11547127	297820516.53	62979766	0.183	4.729
12	1900	11854000	325250325.17	76303387	0.155	4.263
13	1910	15968000	396002433.52	91972266	0.174	4.306
14	1920	25117000	305928827.55	105710620	0.238	2.894
15	1930	40156000	585755812.46	122775046	0.327	4.771
16	1940	67527000	1174984270.1	131669275	0.513	8.924
17	1950	91462000	924498655.02	151325798	0.604	6.109
18	1960	127934000	1052875209.5	179323175	0.713	5.871
19	1970	247653000	1554871488.6	203302031	1.218	7.648
20	1980	1078488000	3188384857.6	226542199	4.761	14.074
21	1990	2492830000	4646219330.5	248718301	10.023	18.681
22	2000	4500000000	6365932055.8	281421906	15.990	22.621
23	2010	13000000000	14523053711	308745538	42.106	47.039

Census day moved to April 15, in 1910. Previously, this had varied but it was mostly June 1 or August 1, with stipulations that the census be completed within a set period of time, usually around nine months even though this time-frame was frequently hard to achieve. A statutory limit for completion was introduced for the 1900 census (Hill, 1920). An important precursor to the future mail-out mail-back step (of the 1970 census, see Section 7) was the sending out of the agricultural census to farmers in northern states ahead of census day so as to give them time to complete the relevant questionnaires properly; these were then collected by the numerators at the time of the census itself, see Durand (1910).

From the outset, led initially by Jefferon (see Anderson, 2015), there has been a tension between those who sought more and more information gleaned through expanded numbers of questions and those who wanted to reduce the burden of those responding to the census questionaires. Regardless of this dilemma, essentially however the numbers of questions asked continued to expand and continued to revolve around basic questions as to the economic and social growth and social fabric of the day, in addition to the constitutionally required mandate to gather counts of the population itself. This set-up changed in 1940, when for the first time what came to be called a "long-form" set of questions was asked of a sample of 5% of the population instead of everyone. While the percent sampled varied, this long-form sampling continued until the 2000 census (when 17% received the longer version of questionnaire). By 2010, only a short set of questions was used in the census itself; the long-form had been replaced by the American Community Survey, which was first implemented in 2005, and which annually surveyed roughly 250-300,000 households monthly, designed to give accurate up-to-date snapshots of the nation's communities. perhaps Jefferson's detractors would appalud this development.

As did the population count grow, so naturally, given the complexity of the nation's peoples and settlements, did the costs of executing the census process grow. Table 1 displays the actual cost by census decade, the cost in 2017 $s, the population counts, the actual costs per person, and the costs per person in 2017 $s. Actual costs and costs in 2017 $s along with population counts are plotted in Figure 1. The attainments of 100 million (in 1920), 200 million (in 1970) and 300 million (in 2010) population milestones are highlighted. The trends are obvious.

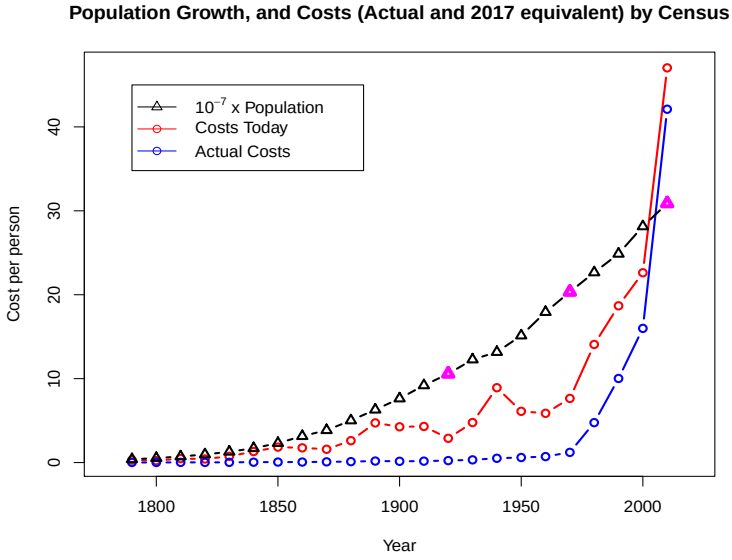


Figure 1: Population Growth, and Costs (Actual and 2017 equivalent) by Census

3. Agriculture, Industry, and Vital Statistics

Agricultural Components

Consistent with the Jefferson agrarianism that defined the 1800's, agriculture questions proliferated. Their first main entrance came with the Sixth Census of

1840. From this time, many agricultural questions graced the census documents.

According to Hill (1921), by 1910, there were over five hundred questions on agriculture alone, reduced after protests to 300+ in 1920. There were questions on the farmer/operator, tenure, acreage, farm value, encumbrances, expenditures, mortgage debt and labor costs, irrigation, with drainage being added for the first time in 1920. Farm products - crops of many types, fruit, dairy, poultry, the list goes on - and their anticipated income were all there. Ultimately, budgetary responsibility was transferred to the Department of Agriculture in 1997 though the census itself was still under the purview of the Census Bureau and the questions were still determined by the Census Bureau.

The agricultural census continued to run concurrently with the population census until 1950, after which it became a quinquennial census occurring in years ending in "4" and "9" for the first 20 years and subsequently in years ending in "2" and "7" after 1978.

Industrial and Economic Components

Although agriculture dominated in the early decades, by 1870, industrialism and mineral explorations had largely intruded and superseded this agrarianism. This began quietly in 1810 with questions on the industrialization of the country, when marshals were asked to gather information which tended to focus solely on whether or not there were manufacturing establishments in their districts. This was essentially the same in 1820 though there were some additional questions seeking details on employment, machinery and expenditures. Such economic components were omitted in 1830, but returned in 1840 with an expanded set of questions to include fisheries, commerce and mining, along with more detailed questions on manufacturing establishments. While there were penalties for failing to provide answers to population related questions, no such penalties existed when it came to industrial questions. Indeed for the 1840 census, this generated quite a bit of resistance especially in the South; one southern publication wondered "Is this federal prying into the domestic economy of the people a precursor to direct taxes?" (Lunt, 1888).

Reflecting the manufacturing/industrialzation social history of the country, questions in each successive census - often increasingly with considerably more detailed components - were added to measure progress or lack thereof of various enterprises. It becomes possible to trace the rise and influence of specific industrial sectors and their subsequent merging into the industrial fabric, their importance fading as new endeavors took center stage. Mineral exploration and railroad developments come immediately to mind here. For example, questions on railroads were added - their construction, materials, employees, costs, everything associated with this new mode of transportation particularly as it hastened the expansion of the West; sewing machines were added in 1860 - seen as an important factor in tracking the emergence of women to the salaried labor market; mineral products and fisheries in 1880; the rise of organized labor brought its share of new inquiries in 1880 - there were 167 questions just on wages and prices; and so on. Indeed, the total number of questions in 1880 on this industrialization aspect alone is mind-boggling (see Micarelli, 1998, who provides exceeding detailed details for all economic censuses over the 1800's and 1900's). Hunt's (1909) article nicely focuses on occupational questions. Prior to 1860, the country was rural; by 1860, one-third of the populace worked in manufacturing; by 1900, manufacturing goods were twice the value of agricultural goods.

The 1902 Act establishing the Census office also charged that there be quinquennial economic censuses. Thus, the industrial statistics component of the census was effectively siphoned off into its own domain. Over the years, these were sometimes run concurrently but independently of the decennial population census; on those occasions, population enumerators doubled up as follow-up enumerators for the economic census in small cites. Through the 1920's and 1930's, these had become biennial censuses, were suspended during the war years, to resume again for 1947, still as biennial compilations - but not for long, as from 1948, these returned to their quinquennial status taking place in years ending in "3" and "8" until it was changed to "2" and "7" starting in 1967.

Vital Statistics

The quest for vital statistics in their many forms dominated the census process and the questionnaires as well as being largely the primary focus of the ASA members for most of its first 100 years. By the 1850 census, there were murmurings and calls for the need of data which assisted in understanding the social development of the nation. These calls increased dramatically over the latter half of the century, but tended to fall under the broad rubric of vital statistics. Social scientists, historians, economists, educators, medical scientists, indeed researchers of all persuasions were clamoring for information to assist in their work to answer critical questions about the people – not just where and how they were employed, not just the value of their work and crops (those questions were gathered under the agricultural and industrial rubrics), but also about what can variously be described as societal, health, and living conditions.

Thus, e.g., since age was considered important for understanding social and economic problems (Chaddock, 1931), details on age and place of birth information, plus marital status, were added instead of simply the names of individuals. This of course became important in planning for schools. More broadly though, these data helped measure migration movements. Further, these migrations described the westward movement of the population in the 1800s. A different kind of migration occurred in the 1900s when there were population shifts from rural to urban areas especially in the 1920s when the urban economies boomed; or shifts from southern to northern regions in the 1920s, and then from the North to the South in the late 1900s, again driven by changing economic conditions. Migrant movements brought on by the Depression in the 1930s, overseas immigrants in the early part of 1900s also evolved. How these played out could be tracked from the various census returns.

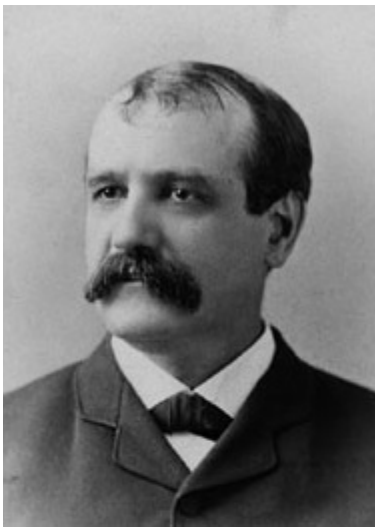
Date and causes of death became important issues in determining the health status of its people. Getting a real handle on death and its causes was proving to be most problematical, so much so, that the ASA in conjunction with the newly formed American Medical Association (an offshoot of the ASA) called for standardized death certificates in 1847. Tied in with deaths were the rates of diseases overall – tuberculosis, typhoid fever, dysentry, diarrhea, whooping cough, infant mortality, consumption, measles, yellow fever, you name it, all were pervasive and persistent. Medical researchers - primarily, but also other social entities - demanded data on these occurrences including geographical locations, not just for the medical purposes but also to help identify and isolate causes (such as inadequate housing conditions, poor sanitation in the workplace, and so on). Information on deaths by accidents was used to help improve railways, roads, factories, etc. Life tables and insurance emerged as related disciplines.

By and large, each successive census contained more and more questions, all

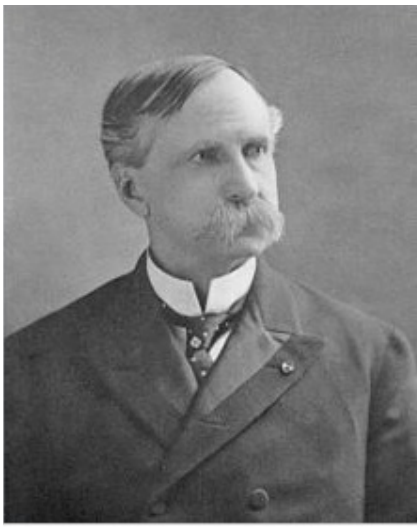
designed to gather information on these issues. Vial statistics were seen as vital. As reported in Willcox (1926), Wilbur (1907) wrote: "Vital statistics is the Cinderella of modern public health. ... (For) public health work vital statistics are our sole dependence."

4. Francis A. Walker and Carroll D. Wright

Francis Amasa Walker (Figure 2, left) was a giant on the American landscape in the late 1880's, and not just for his role with the census process, albeit a product of his times. He was lauded as a brilliant military man, exceptional organizer and administrator, a great writer and visionary, a scholar, an editor, a publisher, among many other accolades. At only twenty-nine years of age, he was appointed as Superintendent of the 1870 census and again for the 1880 census. Walker was President of the Massachusetts Institute of Technology from 1881-1897 during which time he made substantial changes all designed to put the Institute's unsteady prospects onto a solid financial footing along with the addition of new academic programs to ensure the future success of the institution. Furthermore, Walker was the ASA President from 1883-96 and the founding President of the American Economic Association (1886-92). In short, Walker was called upon to lead a breathtakingly busy life in several high profile positions, many simultaneously.



Francis Amasa Walker
July 2, 1840 - January 5, 1897,
ASA President 1883-1896



Carroll Davidson Wright
July 25, 1840 - February 20, 1909
ASA President 1897-1909

Figure 2: Giants of the 1800s

The census process was revolutionized under Walker's reign. He engaged geographers to help map out the land so as to be able to identify better where people lived and in what circumstances. In later years, Walker (1897) urged the nation to train and retain statisticians to work on the census.

Walker's 1874 *Statistical Atlas of the United States* was revolutionary at the time, with its numerous charts and maps, in color, showing the movement of population as it expanded across the ever-changing landscape of "The West", and illustrating pictorially the state of the nation across many mostly social perspectives.

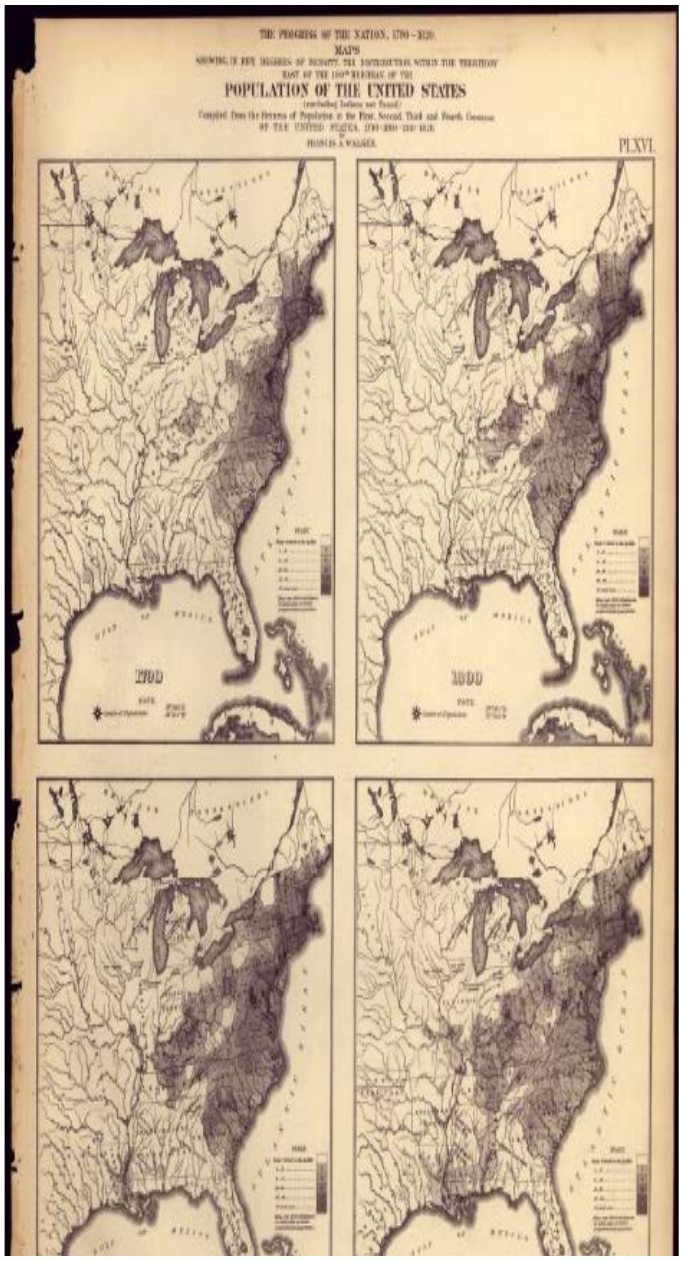


Figure 3: Walker's *Atlas* Population Map, 1790-1820 by Census Decade
Downloaded from the Library of Congress
<https://www.loc.gov/item/05019329/>

The historical setting is Post Civil War (or, Post War Between the States) with the uprooting of the Jefferson agrarianism mantra to one that saw industrial entities including mineral exploration becoming more prominent. The maps were constructed by exploiting new graphical methods allowing for complex data to be shown visually, all driven by the quantitative data arising from the censuses. Most were based on the 1870 census data. Some such as the population maps of Figure 3 showed trends over time as white settlement spread west. These results challenged traditional notions surrounding Jefferson's agrarianism. There was considerable focus on the movement west as mineral resources were revealing themselves (or so

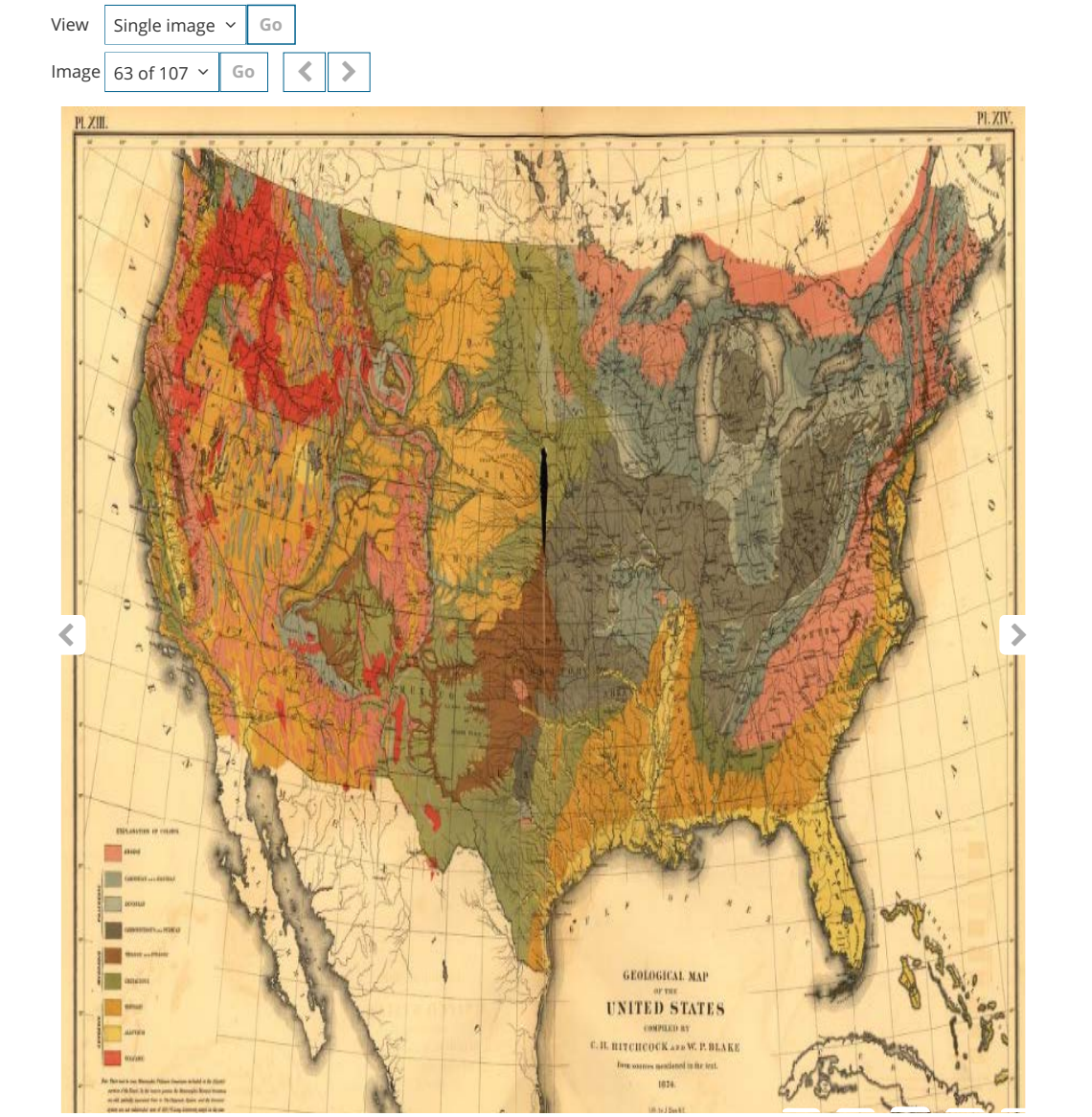


Figure 4: Walker's *Atlas* Geological Map, 1872
Downloaded from the Library of Congress
<https://www.loc.gov/item/05019329/>

There was also considerable attention paid to the distribution of native American peoples, though admittedly there were also gaps in the knowledge for this segment of the population again reflecting the data collected, or not collected, on censuses. We remind ourselves that Walker was also Commissioner of Indian Affairs (1871-72) during the time these maps were being prepared; and he did not neglect that duty to the native American people. Kinnahan (2008) has a nice comprehensive summary of the impact of this volume. Kinnahan tells how the American identity at the time had become destabilized because of rapid changes in urbanization, immigration and industrialization, and in considerable detail explains how Walker's *Atlas* focussed the nation's lens on accurate and true data to shape and re-shape perceptions.

To those who knew him, Walker was distinguished by his "intellectual honesty [and] the integrity of his statements" (Wright, 1897). Indeed, there is frequent reference to Walker's letter to Wright upon the latter's appointment as Commissioner of Labor in 1893. Wright (1897) explains: "General Walker laid down the

enduring principles of official statistics. Whoever adheres to them will meet with success; whoever neglects them commits a crime." These words of wisdom surely echo down the halls of history to those of us today who profess to ply the trade of quantification/statistics.

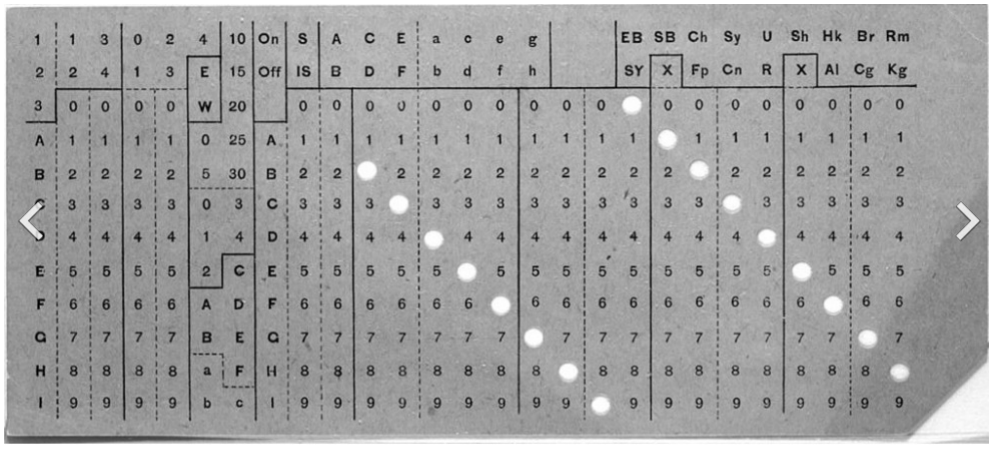
Carroll Davidson Wright (Figure 2, right) was a contemporary of Walker, both born in July 1840 mere months after the formation of ASA. Though both studied law, it was the emerging field of "quantification" (i.e., statistics) as it related to improving the social conditions through the census process that galvanised their energies. Wright himself was a key figure in the labor movement broadly described with his census innovations largely driven by his concerns for the industrial sector - both the workers and the industrial titans - he had the respect from all sides of the raging debates of the day. Though Walker was the acknowledged "giant", Wright was not far behind. They worked together well, collaborating and assisting each other, as they both worked assiduously for the improvement of society and the census process. It was Wright who became the first, and longtime, U.S. Commissioner of Labor (1885-1905). Wright, though reluctantly becoming census Superintendent for the period 1893-97, worked diligently with many of his proposals for the census being implemented. Both men were highly principled men; both though in some sense holding political appointments never let politics influence their decisions. Rather both were led by the compulsion to publish the facts, to ensure the data were right, and to leave any interpretation to the social scientists and political users of the data. As summarized by Wright (1904) "The official statistician['s] ... duty is to collect, classify, and publish facts relating to the conditions of the people"; they did this well. It was not surprising that the protege Wright became ASA President (1897-1909) when the mentor Walker (ASA President 1883-1896) died in January 1897.

5. Hollerith Machines

As indicated above in Section 2 and Section 3, changes in the censuses from about 1850 onwards were driven by a desire to convert the mere collection of population counts to one that provided answers to scientific questions governing the (very broadly defined) social structure of our society. From an operational point of view, a bigger issue emerged when trying to organize the results into tables that could be useful in those enterprises. Up to the 1880 census, counts were tallied by "paper and pencil" and with the ever expanding population by size and area covered, this was taking far too long to do (e.g., a reported 8 years for the 1880 census). According to Willcox (1926), John Billings, then in charge of the vital statistics program and later a supervisor of the 1880 and 1890 censuses, was an inspiration with his casual remark "There ought to be some mechanical way of doing this job, something on the principle of the Jacquard loom, perhaps, whereby holes on a card regulate the pattern to be woven." Therefore, in 1888 a competition was held to solicit a better way, i.e., could the tabulation stage be accelerated by mechanical means? Three contestants entered the fray. The challenge focused on capturing the data themselves and also on assembling said data into suitable tables. The machine developed by Herman Hollerith (1860-1929), a former census employee, far outshone his competitors in terms of its speed of operation, both in transcribing the returns and in tabulating the results; and so was given the contract for the 1890 census. Other countries soon adopted this machine also. Figure 5 (top) shows this machine along with its punch card of Figure 5 (bottom). [Incidentally, Hollerith's

machine was one of four "computer" companies that merged in 1911 into what later became the International Business Machine (IBM).] Hollerith (1889) gives a technical description of his mechanical machine and how the traditional means of tally-marks can be replaced by his punch cards.

Hollerith Machine, 1890



Hollerith Punched Card, 1890

Figure 5: The Hollerith Machine Revolutionized the Data Processing in 1890

These punch cards recorded the data from the census returns by series of punch holes, and were then feed into the Hollerith machine for subsequent tabulation and processing. One issue that came with this innovation was the concept of "recording errors"; the new capture-recapture sampling ideas (see Petersen, 1896) were

implemented to address these issues. Figure 6, from the 1920's, illustrates how completely and extensively the census bureau depended on their "computers" to process the collected counts. Versions of the Hollerith machine continued to be used until the then-new ENIAC (Electronic Numerical Integrator and Computer) machine replaced them for the 1950 census with its adaptation UNIVAC (Universal Automatic Computer) developed for population and housing tabulations. A second version of the UNIVAC was used for the 1960 census.

Figure 6: Census Tabulation 1920

Figure 7: Data Capture 2000

Up until this time, it was the census bureau that primarily drove the development of computer technology. After that however, it was largely the private sector that led the electronic-computer revolution we know today. Of course, census operations are still heavily dependent on the latest technology to effect the charge to deliver the numbers so-to-speak to the American people in a timely manner. Thus, Figure 7 shows the use of then-contemporary scanners in use to capture digitally the

data from the 2000 census. Enumerators, used today to follow-up non-respondents, depend on computer-based technology instead of paper and pencil. These and related technological innovations change rapidly today.

While the early development of computer technology was intricately tied to the census process, these uses were focussed on capturing the information on the returns and assembling the collected data into meaningful tables for other scientific and societal purposes. Given the current state of our technologically driven world, it has to be assumed that some time soon, the census forms themselves will be completed and/or filed electronically. That of course will bring other errors (and not simply the recording errors when computers were first used in 1890).

6. ASA and Permanent Census Office

When the plans for the 1840 census were the same "antiquated and ineffective machinery" (used since) 1790, five men gathered in Boston in 1839 and formed the American Statistical Association. They wanted changes; but were not really able to achieve much, at least for 1840. Clamors for a better organization continued to be raised however. Eventually, ASA President Shattock had some success in the 1849 establishment of a Census Board consisting of the Secretary of State (who till then had sole responsibility for the execution of a census), the Postmaster-General and the Attorney-General; see Lunt (1888). Then, again with ASA input, Congress in 1850 decided that the census should be moved to the Department of the Interior, and that a single person, a Superintendent, be responsible for a given census. However, the most significant advances to the census process - how it was conducted, what questions to be asked and how analysed, etc. - did not come until 1870 when Gen. Walker over a 20+ year period established essential and long-lasting principles for a proper census (Wright, 1897).

In January 1870, there were the beginnings of stirrings to establish a permanent census office, with a long Report arising from an extensive investigation led by Representative General James A. Garfield (later U.S. President) who was Chair of the House Committee on the 1870 Ninth Census, aided mightily by Francis Walker and then-ASA President Edward Jarvis. The Report examined perceived "principal defects in the existing method", criticized the time allowed to take the census (up to 9± months), criticized the time taken to publish the results (which time was running into years; see Section 5 on the introduction of mechanical means), and it criticized how the census was being run; in short, this was a scathing indictment on the whole census operation. As Lunt (1888) explained, even Col. Walker, esteemed as he was, had been unable to effect what he saw as key operational changes for the 1870 census, characterizing it as "clumsy, antiquated, and barbarous".

There were calls for a radical change in operational machinery. These included, in particular, the doing away of the politically tainted marshals; and a call to create a new organization that was completely independent of all government machinery. The resulting Bill was passed by the House "triumphantly" but defeated in the Senate "curiously" (see Lunt, 1888, and Wright, 1900). Importantly, Walker continued to fight for a permanent census office.

In 1896, both the ASA and the American Economic Association (which incidently had evolved out of the ASA) appointed committees to work together in presenting a proposal to Congress to establish a permanent and independent census bureau (see FitzPatrick, 1957). Not surprisingly, Walker was on both committees (the only overlap in membership). Wright (who was not on either committee) as the

current Superintendent of the Census presented the proposal to Congress in 1897. Congress finally established the office to become effective in 1902. Its first Director was William R. Merriam (1902-03) who was the then-current Superintendent. The list of all subsequent directors, along with those who served as Superintending Clerks or Superintendents of prior censuses is given in Table 2.

Table 2: Census "Directors" and Tenure

Secretaries of State		Directors	
1790	Thomas Jefferson	1915-21	Samuel L. Rogers
1800	John Marshall	1921-33	William M. Steuart
1810	Robert Smith	1933-41	William L. Austin
1820	John Quincy Adams	1941-49	James C. Capt
1830	Martin Van Buren	1950-53	Roy V. Peel
Superintending Clerks		1953-61	Robert W. Burgess
1840	William A. Weaver	1961-65	Richard M. Scammon
1850-53	Joseph C. G. Kennedy	1965-69	A. Ross Eckler
1853-55	James D. B. DeBow	1969-73	George H. Brown
1860-65	Joseph C. G. Kennedy	1973-76	Vincent Barabba
Superintendents		1977-79	Manuel D. Plotkin
1870	Francis A. Walker	1979-81	Vincent Barabba
1879-81	Francis A. Walker	1981-83	Bruce Chapman
1881-85	Charles W. Seaton	1984-87	John G. Keane
1889-93	Robert P. Porter	1989-93	Barbara Everitt Bryant
1893-97	Carroll D. Wright	1994-1998	Martha Farnsworth Riche
1899-1902	William R. Merriam	1998-2001	Kenneth Prewitt
Directors		2002-08	Charles L. Kincannon
1902-03	William R. Merriam	2008-09	Steve H. Murdock
1902-09	Simon N. D. North	2009-12	Robert M. Groves
1909-13	E. Dana Durand	2013-17	John H. Thompson
1913-15	William J. Harris		

It is perhaps interesting to note that Simon N. D. North was also a member of the ASA committee, and became a Census Director whose term 1903-1909 was followed by his ASA Presidency in 1910 (the first year that ASA presidencies were limited to one year). In a tribute to Carroll Wright, North (1908) opined that "No single thing, save only the requirement for a decennial census ... has done so much to promote the study and to perfect the methods of statistics as that legislation."

7. Collection of Census Returns

For early census takings, population counts were collected by U.S. marshals on horseback scouring the territory, still uncharted, still unmapped, still very much the wild unknown. Understandably, this took many months, and clearly could not vouch for being exactly and totally accurate. The census continued in this manner until the marshals were replaced by supervisors and enumerators in 1870, though horseback was still the mode of transportation. These emumerators/supervisors were supposed to be more independent than the politically aligned marshals they replaced. By 1880, after Walker had appointed geographers tasked to map out the

lay of the land, finding where families and people lived became easier (but not foolproof, of course). For the 1910 census, open competitive examinations were held to select the enumerators. In a move to speed up the completion of the census forms, in 1910, with over 500 agricultural questions to be answered, northern farmers were given advance copies so as to have their completed copies ready for the enumerator. It was not long before these armies of enumerators were housewives, going door-to-door to solicit the relevant census information.

With time, these armies of housewives who had supplanted the swarms of marshals, were themselves hard to find in sufficient numbers to complete the process adequately. Therefore, after testing as part of the 1960 census, in 1970, these armies were replaced by mailing out and mailing back the census forms. By now, geographers were no longer the innovation (of 1880) but an essential component working hand-in-hand with the U.S. Postal Service. As can be seen in Table 3, these mail-back rates progressively declined in subsequent census years, from 90% in 1970 to a projected 55% in 2000. Since by 2000 the cost of following up households who had not returned their forms was substantial, it became necessary to try something new. As it so happened the Census Director appointed in the 1990's, Martha Farnsworth Riche, had a background in journalism. For her, it was a natural avenue to pursue to develop advertising and partnership mechanisms to encourage the populace to mail back their returns. The partnerships involved citizens at all levels, local and nationally based; and the advertisements were strategically designed to appeal to different segments of our society and its cultures. Figure 8 is one such poster. This innovation was an unmitigated success, with the estimated 55% return rate increasing to an actual mail-back rate of 74%. These were matched in 2010 with a mail-back rate of approximately 74% again. A challenge for future census operations is to improve this record, in a climate of distrust of government enterprises. In 1790, many inhabitants were geographically isolated; in contrast, but equally difficult for census enumeration, in 2020, many will be figuratively isolated as today's generation confines itself to electronic isolation and so can be hard to reach.

Table 3: Mail-back Rates by Census

Census Year	1970	1980	1990	2000	2010
Return %	90	78	65	(55) $\rightarrow$ 74	74

8. Accuracy, Undercounts and Sampling

Inaccuracy of the census return numbers has to be a fact of life. Even in the First Census of 1790, Jefferson famously wrote in the margins of his report that not everyone had been counted. Given that the country was expanding westward into uncharted territories at unknown rates, it would be hardly credible not to agree with this assessment. One hundred years later, the plans for the 1900 census were riddled with concerns for accuracy (Wilbur, 1907). For future censuses, Durand (1913), himself a census Director at the time, talks a lot about the need to produce more accuracy in the population counts, targeting in particular the competence or

Figure 8: An Advertising Poster Designed for Census 2000

otherwise of enumerators. Bullock (1899) provides a long detailed critical discussion of inaccuracies as they relate to wages across the censuses of the 1800s. Despite the pleas for greater accuracy, there is an unwritten assumption that inaccuracies cannot be avoided, only efforts to reduce them should be driving forces.

Discussion of undercounts and overcounts populated census reports for the 1800s. The reasons were many, but oft-times occurred because of changing definitions of who, how and why certain demographics should be counted. Stark reality hit after the 1940 census when the number of men of color enlisting in the armed services exceeded the number counted in the 1940 census; see Price (1947). Table 4 shows the percent undercount figures for the 1940-1990 censuses for the population overall and for black and non-black races by gender. Since recorded racial categories changed somewhat for the 2000 and 2010 censuses, a direct comparison by race cannot be made for these years. However, the overall population was estimated to be overcounted by 0.49% in 2000 and by 0.01% in 2010. Undercounts can occur for a variety of reasons; e.g., people may have more than one residence such as seasonal homes all listed as "occupied, or college students may be counted by their parents and also by their college community.

Table 4: Percent Under-Counts 1940-1990

Census	Total	Male	Female	Black	Male	Female	Non-Black	Male	Female
1940	5.4	5.8	5.0	8.4	10.9	6.0	5.0	5.2	4.9
1950	4.1	4.4	3.8	7.5	9.7	5.4	3.8	3.8	3.7
1960	3.1	3.5	2.7	6.6	8.8	4.4	2.7	2.9	2.4
1970	2.7	3.4	2.0	6.5	9.1	4.0	2.2	2.7	1.7
1980	1.4	2.2	0.3	4.5	7.5	1.7	0.8	1.5	0.1
1990	1.6	2.8	0.9	5.7	8.5	3.0	1.3	2.0	0.6

Undercounts are not limited to racial categories however. They can occur across regional lines, across rural-urban areas, across tenure categories (i.e., home ownership or rental). Hard-to-count population segments (e.g., those in shifting locations such as hospitals, prisons, homeless, etc.) are traditionally undercounted.

How these differential undercounts were to be estimated has received considerable statistical attention, starting with demographic analyses in 1940. Since then, we see the use of dual estimation methods, post-enumeration surveys, accuracy and coverage evaluations, census coverage measurement, all designed as an attempt to improve the accuracy of the overall count. In a broad sense, these methods are complicated sophisticated versions of the simpler capture-recapture methodologies, along with stratification principles. Detailed descriptions of the estimation methods can be found in, e.g., Wright (1998), Hogan (1992, 1993), Mulry and Spencer (1993), Robinson et al. (1993), Kostanich (2000), Wolter (1986), Bell and Cohen (2009), Mule (2012). Schirm (1991) looks at the undercount adjustment methods as they pertain to congressional apportionment.

Whatever the inherent difficulties associated with obtaining exact numbers, importantly however, this net undercount figure narrowed, as shown in Table 3. The differential undercount for men of color and other hard-to-count segments of the population remains larger than for the population as a whole, but these segments too seem to be more accurately estimated over time.

An undercount of a different kind occurred in the 1870 census counts of population in the Southern States. According to Wright (1897), the increase from 1870 to 1880 was so large that some, especially the Northern States were highly suspicious. There was considerable political pressure on Superintendent Walker to take "precipitate action". Instead, Walker quietly set about his own investigations, subsequently learning that the 1880 numbers were basically correct, and that the previous 1870 counts were very much undercounts which counts he corrected even though this meant contradicting his own report for 1870.

Sampling poses a parallel issue. In general, there are sampling and non-sampling issues, and measurement errors. Early efforts here included work when Wright developed sampling survey frames to obtain representative samples in his work on industrial statistics, and Holmes' (1890) search for representative sampling plans to deal with mortgage statistics. While versions of the capture-recapture sampling procedure have been used to estimate the undercount measures as outlined above, sampling issues by and large made their first, permanent, foray into the process for the 1940 census, after success in the 1937 test survey of unemployment with Neyman (1934, 1938) contributing here. Since then, sampling techniques have been essential to the testing and pre-planning stages of any census process. Morris Hansen, William Hurwitz, and Bill Madow, among many others, developed key guiding principles in these enterprises. The sampling book Hansen et al. (1953) was an important primer in setting the stage. In his Presidential address, Bowman (1964) endorses and urges the continued use of sampling efforts in the federal statistical domain. The afore-mentioned long form questionnaire and the current American Community Survey are totally based on appropriate sampling frames.

9. Conclusion

The constitutional mandate was primarily designed to obtain population counts for the purposes of apportionment of congressional seats. Up to the 1910 census, as the population grew, so did the number of seats expand. After that, the number of seats was fixed permanently at 435. Hence, apportionment by states and districts became a zero-sum game; for the politicians, the stakes were now higher. Fearful that certain congressional seats would be lost, this game became particularly loud in the lead-up to the 2000 census when, in an attempt to control costs (see Table 1

and Figure 1), the census procedures included additional sampling plans (over and beyond those traditionally used over the years for correcting counts, etc.) for the population non-response follow-up stage. Thus it was that, in the Disaster Relief Supplemental Appropriation of June 1997, Congress added a rider prohibiting the use of "sampling and any statistical procedure", or words to that effect, for the 2000 Census. If passed, this would be devastating to the census process as, e.g., this would prohibit even the collection of the data into tables for dissemination (which had been the backbone of census reports ever since 1790). The then-ASA President, along with all (non-government) past presidents of ASA, sent a letter (see *Amstat News*, 1997) to the US President and Vice-President and all members of Congress, pointing out that this would essentially nullify the constitutional mandate to have a census. Fortunately, the bill was signed into law without that rider. However, later, in 1999, the Supreme Court did rule that sampling could not be used for apportionment purposes. A related issue was the drawing of district boundaries though this too was not part of the constitutional mandate.

It was also constitutionally mandated that Congress was to oversee the conduct of the censuses. Therefore, ever since the first census in 1790, congress has approved, or disapproved (recall Superintendent Walker's failure to convince Congress in 1870 for an expanded census) the sets of questions to be asked on the census forms, appointed census overseers such as the Superintendents and after the establishment of a permanent office in 1902 approved the appointments of Census Directors.

The US Constitution's concern rested solely with obtaining an accurate count of the population, along with the oversight of effecting this count. However, right from the very beginning, the census process expanded through its myriad of questions as it sought to obtain data, information, that would help explain, help describe the nation, to itself. In turn, these data were used to guide researchers, planners, and the like, as they sought to improve services to the peoples. This was often in the form of funds distribution to states and local communities, including Medicaid, the federal highway program (e.g., where should roads be built?), grants to local education agencies (such as the school lunch program), social services block grants, rehabilitation services, community development block grants, and so on. Data gathered assist in the planning of where, if at all, schools, or hospitals, care facilities, shopping centers, should be built. Identified population concentrations assist in the planning of emergency services availabilities. Researchers may wish to study the aging of the population, poverty patterns, residential movements, etc. The list goes on.

In short, census data are crucial to the vibrant functioning of our society. In his brief overview of censuses from 1790 to 1960, Hauser (1963) discussed why the census was important, apart from the constitutional mandate, and what data are collected and utilized to answer the needs of society. Echoing North (1908) on the importance of the 1902 legislation to the census process, Hauser concludes that "statistics was, not only a by-product of the social morphological revolution, but, also, an essential ingredient of it." Reading the reports and journal articles on the successive census operations, sheds light on the social history of our country; it makes for good reading!

REFERENCES

Amstat News (1997), "ASA Objects to Sampling Prohibition Bill," *Amstat News*, 244, 1,3.

Anderson, M. J. (2015), *The American Census: A Social History* (2nd ed.), New Haven: Yale University Press.

Anderson, M. J., and Fienberg, S. E. (1999), *Who Counts? The Politics of Census-Taking in Contemporary America*, USA: Russell Sage Foundation.

Bell, R. M., and Cohen, M. L. (eds.) (2009), *Coverage Measurement in the 2010 Census*, USA: National Research Council.

Billard, L. (1998), "The Role of Statistics and the Statistician," *The American Statistician*, 52, 319-324.

Bowman, R. T. (1964), "The American Statistical Association and Federal Statistics: Presidential Address," *Journal of the American Statistical Association,* 59, 1-17.

Bullock, C. J. (1899), "Wage Statistics and the Federal Census," *Publications of the American Economic Association*, 2, 343-368.

Durand, E. D. (1910), "Changes in Census Methods for the Census of 1910," *Publications of the American Statistical Association*, 12, 52-66.

Durand, E. D. (1913), "The Census Methods of the Future," *Publications of the American Statistical Association*, 13, 563-582.

Fienberg, S. E. (1992), "Bibliography on Capture-Recapture Modelling with Application to Census Undercount Adjustment," *Survey Methodology*, 18, 143-154.

FitzPatrick, P. J. (1957), "Statistical Societies in the United States in the Nineteenth Century," *The American Statistician*, 11, 13-21.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), *Sample Survey Methods and Theory*, New York: John Wiley.

Hauser, P. M. (1963), "Statistics and Society: Presidential Address," *Journal of the American Statistical Association*, 58, 1-12.

Hill, J. A. (1920), "The Essentials of a Good Census," *Quarterly Publications of the American Statistical Association*, 17, 1-12.

Hill, J. A. (1921), "Progress of the Census of Agriculture," *Quarterly Publications of the American Statistical Association*, 17, 779-780.

Hogan, H. (1992), "The 1990 Post-Enumeration Survey: An Overview," *The American Statistician*, 46, 261-269.

Hogan, H. (1993), "The 1990 Post-Enumeration Survey: Operations and Results," *Journal of the American Statistical Association*, 88, 1047-1060.

Hollerith, H. (1889), "Tabulation of Statistics," *Publications of the American Statistical Association*, 1, 279-282.

Holmes, G. K. (1890), "Mortgage Statistics," *Publications of the American Statistical Association*, 2, 1-21.

Hunt, W. C. (1909), "The Federal Census of Occupations," *Publications of the American Statistical Association*, 11, 467-485.

Kinnahan, T. P. (2008), "Charting Progress: Francis Amasa Walker's *Statistical Atlas of the United States* and Narratives of Western Expansion," *American Quarterly*, 60, 399-423.

Kostanich, D. (2000), "Accuracy and Coverage Evaluation Survey: Final Post-Stratification Plans for Dual System Estimation," *DSSD Census 2000 Procedures and Operations Memorandum Series #Q − 24*. Washington DC: Bureau of the Census.

Lunt, E. C. (1888), "History of the United States Census," *Publications of the American Statistical Association*, 1, 63-93.

Micarelli, W. F. (1998), "Evolution of the United States Economic Censuses: The Nineteenth and Twentieth Centuries," *Government Information Quarterly*, 15, 335-377.

Mule, T. (2012), "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States," *DSSD 2010 Census Coverage Measurement Memorandum Series* #2010-G-01.

Mulry, M. H., and Spenser, B. (1993), "Accuracy of the 1990 Census Undercount Adjustments," *Journal of the American Statistical Association*, 88, 1080-1091.

Neyman, J. (1934), "On the Two Different Aspects of the Representative Method," *Journal of the Royal Statistical Society*, 97, 558-625.

Neyman, J. (1938), "Contribution to the Theory of Sampling Human Populations," *Journal of the American Statistical Association*, 33, 101-116.

North, S. N. D. (1908), "The Outlook for Statistical Science in the United States," *Publications of the American Statistical Association*, 11, 17-26.

North, S. N. D. (1909), "The Life and Work of Carroll Davidson Wright, Fifth President of the American Statistical Association," *Publications of the American Statistical Association*, 11, 447-466.

Petersen, C. G. J. (1896), "The Yearly Immigration of Young Plaice into the Limfjord from the German Sea," *Report of the Danish Biological Station (1895)*, 6, 1-48.

Price, D. O. (1947), "A Check on Under-Enumeration in the 1940 Census," *American Sociological*

Review, 12, 44-49.

Robinson, J. G., Ahmed, B., Das Gupta, P., and Woodrow, K. A. (1993), "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Coverage," *Journal of the American Statistical Association*, 88, 1061-1071.

Schirm, A. L. (1991), "The Effects of Census Undercount Adjustment on Congressional Apportionment," *Journal of the American Statistical Association*, 86, 526-541.

Walker, F. A. (1897), "Remarks of President Walker at Washington," *Publications of the American Statistical Association*, 5, 179-187.

Wilbur, C. L. (1907), "The Outlook for a General System of Registration of Vital Statistics in the United States," *Publications of the American Statistical Association*, 10, 395-420.

Willcox, W. F. (1926), "The Past and Future Development of Vital Statistics in the United States. I John Shaw Billings and Federal Vital Statistics," *Journal of the American Statistical Association*, 21, 254-266.

Wolter, K. M. (1986), "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, 81, 338-346.

Wright, C. D. (1897), "Francis Amasa Walker," *Publications of the American Statistical Association*, 5, 245-275.

Wright, C. D. (1900), *The History and Growth of the United States Census*, U.S. Government Printing Office.

Wright, C. D. (1904), "Science and Economics," *Science*, 20, 897-909.

Wright, C. D. (1908), "Address of Carrol D. Wright, President of the American Statistical Association, at its Annual Meeting in Boston, Jan. 17, 1908," *Publications of the American Statistical Association*, 11, 1-16.

Wright, T. (1998), "Sampling and the Census 2000: Concepts," *The American Scientist*, 86, 245-353.