

A Note on the Classification of Chinese in the IPUMS Full-Count Data Sets for 1880 and 1940

Susan Boslego Carter
Professor Emerita, Economics, University of California, Riverside
Susan.carter@ucr.edu

Prepared for the Census Full Count Sessions at the NBER DAE Meetings,
Cambridge, MA, July 8 and 9, 2015

July 1, 2015

Abstract

The IPUMS USA's Full Count data sets for the federal censuses through 1940 offer an unparalleled opportunity to document and analyze, for the first time, Chinese American economic and demographic history during the Exclusion Era. Before that work can begin, however, a substantial amount of additional data cleaning is required.

First, a Word of Thanks

Scholars rightly refer to the Chinese Exclusion Era as the “dark ages” of Chinese American history, ‘a deplorable lacuna in American historiography’.¹ Writing more than a quarter century ago, Roger Daniels called attention to the absence of any “dense corpus of scholarly books and articles [on Chinese American history] based on expertise in pertinent areas of history, economics, sociology, anthropology, and folklore.”² Daniels could have added the field of demography as well. Little has changed in the interim. The Chinese are mentioned only in passing in Michael Haines’s and Richard Steckel’s 736-page magnum opus, *A Population History of North America*, and in Richard A. Easterlin’s 43-page survey, “Twentieth-Century American Population Growth.”³

The problem stems from the limited reporting of basic demographic and economic information for the Chinese in the published censuses. Before 1940, little more than the number of Chinese persons, their gender, and place of birth were reported as state-wide totals. Even age was reported only sporadically, and only for the nation as a whole. There is no systematic reporting of year of immigration, industry, occupation, marital status, or living

¹ Sucheng Chan. *Entry Denied: Exclusion and the Chinese Community in America, 1882-1943*. Philadelphia: Temple University Press, 1991, quoted in Erica Lee, *At America’s Gates: Chinese Immigration During the Exclusion Era, 1882-1943*. Chapel Hill, NC: University of North Carolina Press, 2003, p. 8. In saying this I do not intend to minimize in any way the substantial and excellent scholarship on Chinese Americans during this period. My point is that this scholarship has, quite naturally, focused on topics for which research materials are available. A central topic has been the legal and political developments that defined Chinese Americans’ options. Others are Chinese social organizations, transnational connections, cultural contributions, and immigration strategies. See Lee (2003, pp. 8-11) for a relatively recent summary of the literature.

² Roger Daniels. *Asian America: Chinese and Japanese in the United States since 1850*. Seattle, WA: University of Washington Press, 1988: xiv.

³ Michael R. Haines and Richard H. Steckel. *A Population History of North America*. New York: Cambridge University Press, 2000 and Richard A. Easterlin, “Twentieth-Century American Population Growth.” In Stanley L. Engerman and Robert E. Gallman, eds., *The Cambridge Economic History of the United States*. Vol. III, *The Twentieth Century*. New York: Cambridge University Press: 2000: 505-548.

arrangements. Scholars' descriptions of fundamental topics such as the rate of population change, fertility, mortality, international migration, internal migration, living arrangements, literacy, English language skills, and industrial and occupational attachment are necessarily limited to averages over a wide geographic range or to case studies whose representativeness is difficult to assess.

Yet the census collected all the same data for the Chinese as it collected for the rest of the population. Except for the lost 1890 manuscripts, these are now available at the individual level through 1940. Some of these records have already been put into an electronic format as part of the IPUMS project.⁴ Nonetheless, because there were so few Chinese in America at the time and, because – prior to the development of the Full Count data sets -- the IPUMS samples themselves are small, it is difficult to draw precise inferences from them. Microdata on Chinese Americans in the IPUMS Full Count data sets will allow scholars to calculate, for the first time, basic demographic measures for the Chinese-American population. It's an exciting prospect. Thank you, Ancestry! Thank you, IPUMS!

State-Level Totals for the Full Count IPUMS and the Published Censuses Compared

Table 1 displays the Chinese population by state as shown in the IPUMS Full Count data and in the published census reports for 1880 and 1940. For the 1880 census the Full Count and published population totals are, with a few exceptions, in close agreement. For the 1940 census, on the other hand, the Full Count and published census totals are quite different.

For 1880 in the country as a whole the IPUMS Full-Count and the published census totals are within one percent of one another. In the Western states where the majority of the Chinese lived, the IPUMS Full Count and the published census totals are virtually identical. Discrepancies are limited to just a handful of states -- Delaware especially but also Missouri, Mississippi, Ohio, and Minnesota.

For 1940 the IPUMS Full-Count total exceeds that of the published census by 47 percent, but even this large figure understates the discrepancies because in some states – Delaware, Ohio, Kansas, Missouri, Alabama, North Carolina, Oklahoma, and Tennessee – the IPUMS Full Count total is far *larger* than in the published census while in others – New Hampshire, Indiana, North Dakota, and South Carolina – it is considerably *smaller*.

After comparing manuscript census records at the county level with the totals in the IPUMS Full Count and the published censuses I am convinced that the *published* census counts are highly accurate.

⁴ Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota. Available online at: <http://usa.ipums.org/usa/>.

In what follows I assume that all of the discrepancies are the result of errors in the IPUMS Full-Count data. These errors fall under three major headings.

Major Sources of Discrepancy

Enumerators Who Entered “C” for “Colored” instead of “C” for “Chinese”

In 1880 census enumerators were told that the race column is “always to be filled.” In 1940 they were told, “Write ‘W’ for white; ‘Neg’ for Negro; ‘In’ for Indian; ‘Chi’ for Chinese; ‘Jp’ for Japanese; ‘Fil’ for Filipino; ‘Hi’ for Hindu; and ‘Kor’ for Korean. For a person of any other race, write the race in full.” Nevertheless, in some cases both in 1880 and in 1940 some enumerators entered the letter “C” to indicate “Colored” or “Black.” Ancestry coded these “C”s as “Chinese.” This error clearly explains the excessive number of Chinese in Delaware in the Full Count data for 1880 and is probably the reason for the large Chinese overages in other Southern states in both 1880 and in 1940.

One Type of Ditto Error -- “Chinese” or “C” Transcribed as “White”

My analysis of county-level Chinese American population totals during the Exclusion Era shows that despite their small numbers the Chinese achieved wide geographic distribution.⁵ Many were the sole Chinese American in their community. Perhaps not expecting to see a Chinese person, data entry staff appear to have coded some Chinese as “White.” An example of this error in the 1940 census is the case of Henry Yum Vog of Exeter, Rockingham County, New Hampshire, his wife Nom and their six children, ranging in age from nine years to nine months. Henry and Nom were born in China and were recorded as “Chinese” by the census enumerator but were coded as “White” by Ancestry. Correcting the racial designation of the Vogs accounts for eight of the 31 Chinese missing from the New Hampshire total for 1940.

Another Type of Ditto Error -- Chinese Entered into the Manuscripts as “White”

Another type of ditto error occurred when census takers themselves designated a person who was clearly Chinese as “White.” An example in the 1940 census is Ong Sing, 57, an alien, China-born laundryman living in Bellows Falls, Vermont. Sing’s name, birthplace, occupation, and legal status all suggest that he was Chinese. Sing was the only Chinese person in Bellows Falls, Vermont in 1940. Perhaps because the enumerator, not anticipating a Chinese person, simply entered “W”s in all the little boxes up and down the race columns.

⁵ Susan Boslego Carter, “Confined to Chinatowns? A New Look at Chinese American Geographic Redistribution during the Exclusion Era, 1882-1943.” Work in progress.

Correcting the Discrepancies

The first error – where the census taker used “C” for “Colored” instead of “Chinese” -- is fairly easy to catch. The black population was geographically segregated. Census takers who entered “C” to mean “Colored” used “C” for *all* of the black people in their enumeration district. Tell-tale signs of this error are disagreement between the IPUMS Full Count and published census totals at a fine geographic level along with a high proportion of women and children and Southern birthplace among the purported “Chinese” population.

The two other errors are more difficult to detect. Birth in China offers one clue. In the 1880 census Ah Goon, a 40-year-old laundryman who lived in San Mateo, California was coded as “White” by Ancestry although the census enumerator recorded him as “Chinese.” In the 1940 census manuscripts, Mack Yee, 21, a restaurant owner in Burns, Oregon was recorded by the enumerator as “White” with a birthplace in China although Yee’s name and occupation suggest that he was Chinese. In both cases, Chinese birthplace allowed me to find these individuals and correct the errors.

Still, care must be taken. In the 1880 census the birthplace of Antonia Boziers, 32, a mill worker in Port Blakeley, Kilsap, Washington was coded by Ancestry as “China.” The same was true of the birthplace of her mother and father. The census enumerator had written “Chili” for those entries. In that same census the birthplace of A.A. Carr, 44, a hotel keeper in Wooster, Wayne County, Ohio, was coded by Ancestry as “China” although the census enumerator had entered “Ohio” as his birthplace. The same was true of Carr’s wife and daughter.

A surprisingly large number of whites in both 1880 and in 1940 actually were born in China. An example from the 1880 census is Emma Ames, 14, a schoolgirl living with her grandparents John and Martha Hayden in Bath, Maine. Emma’s parents were not living in the household. Perhaps they were missionaries who sent their daughter back home so she could obtain an American education. Another example is Susie Chase, a 26-year old single school teacher living as a boarder in Oneida, Madison County, New York. Miss Chase told the census enumerator that she and her parents were born in Shanghai, China. This example shows that even birth in China and mother’s and father’s birth in China do not guarantee that the person is Chinese.

In the 1880 census Ancestry reports only 1,037 “Whites” born in China so these can be checked individually. In the 1940 census, however, 11,643 persons coded as “white” reported being “Born in China.” Some errors in recording are relatively easy to catch. For example, white residents of Canton, Norfolk County, Massachusetts had their country of birth misclassified because the census enumerator entered the *town* under the heading “Place of Birth,” rather than the state or country. When Ancestry coded these entries “Glasgow” became “Scotland,” “Belfast” “Ireland,” and “Canton (MA)” -- “China.” In most cases, unfortunately, a trained coder will have to look at the entire entry in the census manuscripts and make a determination. Further compounding the problem for the 1940 census is the fact that “Father’s Place of Birth,” “Mother’s Place of Birth,” and “Native Tongue” are available only for sample line persons and for persons living with their parents.

Conclusion and Suggestion

In this note I document and examine discrepancies between state-level totals for the Chinese population in the IPUMS Full-Count data and in the published censuses for 1880 and 1940. I conclude that a substantial amount of additional data cleaning is required before scholars can put the IPUMS Full-Count data to effective use in studying the Chinese American population. Although I have not examined IPUMS Full-Count data sets for other census years I suspect that they, too, would require a considerable clean-up effort.

The erroneous coding of some blacks as Chinese should be relatively easy to identify and correct but other systematic errors will take quite a bit of time and effort to rectify. Comparing IPUMS and published census totals at the county level helps. Names help. Examination of the original manuscripts is essential.

Is this a project that can be crowd-sourced? Might Asian Studies classes or Chinese American historical or fraternal societies be induced to take on a city, county, or state? Suggestions welcome.

Thanks, again, to Ancestry and to IPUMS for their excellent efforts on this important project!

Appendix: Instructions to Enumerators Regarding the Race Question, 1880 and 1940.

1880: *Color.*-It must not be assumed that, where nothing is written in this column, "white" is to be understood. The column is always to be filled. Be particularly careful in reporting the class mulatto. The word is here generic, and includes quadroons, octofoons, and all persons having any perceptible trace of African blood. Important scientific results depend upon the correct determination of this class in schedules 1 and 5.

1940: Write "W" for white; "Neg" for Negro; "In" for Indian; "Chi" for Chinese; "Jp" for Japanese; "Fil" for Filipino; "Hi" for Hindu; and "Kor" for Korean. For a person of any other race, write the race in full.

Table 1

**Chinese Populations by State in the Published Census Volumes and in the IPUMS Full Count Data
Censuses of 1880 and 1940**

<u>State</u>	<u>1880</u>			<u>1940</u>		
	IPUMS	Published Census	IPUMS as % of Published Census	IPUMS	Published Census	IPUMS as % of Published Census
<u>Connecticut</u>	120	123	98	239	292	82
<u>Maine</u>	9	8	113	83	92	90
<u>Massachusetts</u>	247	229	108	3103	2513	123
<u>New Hampshire</u>	14	14	100	32	63	51
<u>Rhode Island</u>	27	27	100	221	257	86
<u>Vermont</u>	2	0	--	14	21	67
<u>Delaware</u>	366	1	36600	1105	39	2833
<u>New Jersey</u>	178	170	105	1482	1200	124
<u>New York</u>	1,055	900	117	19778	13731	144
<u>Pennsylvania</u>	262	148	177	1004	1477	68
<u>Illinois</u>	216	209	103	1888	2456	77
<u>Indiana</u>	33	20	165	120	208	58
<u>Michigan</u>	30	27	111	622	924	67
<u>Ohio</u>	224	100	224	6483	921	704
<u>Wisconsin</u>	15	16	94	230	290	79
<u>Iowa</u>	48	33	145	57	81	70
<u>Kansas</u>	32	19	168	1062	133	798
<u>Minnesota</u>	47	24	196	513	551	93
<u>Missouri</u>	385	91	423	4260	334	1275
<u>Nebraska</u>	31	18	172	371	102	364
<u>No. Dakota</u>	8	238	0	15	56	27
<u>So. Dakota</u>	230			25	36	69
<u>Virginia</u>	7			165	208	79
<u>Alabama</u>	4	4	100	11177	41	27261
<u>Arkansas</u>	145	138	105	375	482	78
<u>Florida</u>	20	18	111	1920	214	897
<u>Georgia</u>	28	17	165	421	326	129
<u>Louisiana</u>	475	489	97	303	112	271
<u>Mississippi</u>	158	51	310	710	743	96
<u>No. Carolina</u>	0	0	--	1056	83	1272

<u>So. Carolina</u>	8	9	89	15	27	56
<u>Texas</u>	160	136	118	1025	1081	95
<u>Kentucky</u>	2	10	20	119	100	119
<u>Maryland</u>	5	5	100	326	437	75
<u>Oklahoma</u>				1750	112	1563
<u>Tennessee</u>	17	25	68	5509	60	9182
<u>W. Virginia</u>	5	5	100	37	57	65
<u>DC</u>	15	13	115	492	656	75
<u>Arizona</u>	1,641	1,630	101	1245	1449	86
<u>Colorado</u>	611	612	100	269	216	125
<u>Idaho</u>	3,365	3,379	100	167	208	80
<u>Montana</u>	1,763	1,765	100	217	258	84
<u>Nevada</u>	5,428	5,416	100	254	286	89
<u>New Mexico</u>	57	57	100	85	106	80
<u>Utah</u>	529	501	106	214	228	94
<u>Wyoming</u>	914	914	100	81	102	79
<u>California</u>	75,067	75,132	100	38551	39556	97
<u>Oregon</u>	9,534	9,510	100	2046	2086	98
<u>Washington</u>	3,149	3,186	99	2364	2345	101
TOTAL	106,686	105,485	101	113,600	77,504	147

Sources: *IPUMS*: Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010 and published census volumes for 1880 and 1940. *Published censuses*: Michael R. Haines and Inter-university Consortium for Political and Social Research. *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002* [Computer file]. ICPSR02896-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-21. doi:10.3886/ICPSR02896.v3.