Documentation for the *Nevada Census Microdata, 1860-1920*, version 1.0

Ronald M. James, Kenneth H. Fliess, and Eric Nystrom

November 24, 2014

Contents

1 Overview 1

2 Citation 2

3 Availability and License 2

4 Release History 2

5 History of the Project 2

6 Bibliography 3

7 Contact Information 4

8 Technical Documentation 4

9 Errata 28

1 Overview

The Nevada Census Microdata provides a complete-count transcription of the manuscript federal census pages from the state of Nevada, for the censuses of 1860 (Utah counties that became the state of Nevada), 1870, 1880, 1900, 1910, and 1920. It includes nearly all of the information that was recorded on the census forms, though the questions asked by census-takers changed over time. The data set contains 310,486 records.

This data can help researchers investigate historical questions related to immigration, migration, industrialization, work, and family. The multiple years covered by the data permit valuable longitudinal studies, at least within the state of Nevada. Future phases of the project will refine field names and values to make this data as compatible as possible with other sources used in historical demography, such as those published by the Integrated Public Use Microdata Series (IPUMS), and will strive to increase the accuracy of the data.

2 Citation

Please use one of the following citations:

- Ronald M. James, Kenneth H. Fliess, and Eric Nystrom. *Nevada Census Microdata, 1860-1920.* Version 1.0 (1991–2014). http://dx.doi.org/10.7910/DVN/27218
- James, Ronald M.; Fliess, Kenneth H.; Nystrom, Eric C., 2014, "Nevada Census Microdata, 1860-1920", http://dx.doi.org/10.7910/DVN/27218 Harvard Dataverse Network [Distributor] V1 [Version]

3 Availability and License

A web interface to the database which permitted users to find individual records was available as of Dec. 2013 from the Nevada State Historic Preservation Office: http://nvshpo.org/component/census/ though that link was not functional as of Nov. 2014.

The complete data set, available in a comma-separated file (as well as other formats in the future), is available at http://dx.doi.org/10.7910/DVN/27218 or http://thedata.harvard.edu/dvn/dv/nvcensus along with this documentation.

This data is licensed CC0, in the Public Domain. We request that any users include the following credit line:

Ronald M. James, Kenneth H. Fliess, and Eric Nystrom. *Nevada Census Microdata, 1860-1920.* Version 1.0 (1991–2014), http://dx.doi.org/10.7910/DVN/27218.

4 Release History

- 1.0 (Nov. 24, 2014): Initial release of complete dataset (in CSV format) with documentation. No efforts were made to refine this data, so that the 1.0 version can remain available for comparison with published work that used the dataset prior to 2014.

5 History of the Project

The project was initiated circa 1991 by Ronald M. James, the State Historic Preservation Officer of Nevada, to support a series of projects involving historical archaeology in the Comstock Mining Historic District. Initial efforts focused on the census of 1880, because that year was the first to feature street addresses in urban areas, making it useful to archeologists.

Shortly thereafter, Dr. Kenneth H. Fliess of the Anthropology Department of the University of Nevada–Reno began working on a transcription of the 1860 census records for the three westernmost counties of Utah Territory, which eventually became the Territory, then State of Nevada. Fliess coordinated his efforts with James to ensure record compatibility.

James presented research based on the 1880 and 1860 data at a conference for Nevada history in 1993, subsequently producing an article for the *Nevada Historical Society Quarterly*. These efforts attracted the attention of the Nevada Legislature, which provided funding during the 1993 and 1995 sessions to the State Historic Preservation Office (NVSHPO) to continue the work of transcribing the other available manuscript census years. NVSHPO contracted with Fliess, and students working for Fliess eventually transcribed the all of the Nevada 1860, 1870, 1880, 1900,

1910, and 1920 manuscript census records from microfilm. Data entry programs originally written for the *Texas Historical Demography Project*, with some modifications, were used to transcribe the data into a database in dBase/FoxBase format.[1] Fliess supervised the correctness of the data coding, and James focused on editing the names of people. Later, Fliess became too ill to continue, so James edited the 20th century entries.

In the mid to late 1990s, plans were made to release the data as a book combined with CD-ROMs, but due to difficulties with respect to publication the data was never released in this form. Some documentation of the database by Fliess, created as part of this effort, was later made available on the project website, but this documentation contains some differences with respect to the data that was eventually made available.[2]

Once it became clear that a book/CD-ROM combination was not possible, NVSHPO pushed forward an effort to distribute the data in a different way. By no later than 2003, the database was made available to the public on the Web through a search interface on the NVSHPO website.[3] The public could search for names of individuals and also perform some basic analysis of the data, though the complete data set was not available separately. The site was migrated to a different interface at least once since that time.

As of 2013, James retired as head of NVSHPO and it became reasonable to presume that new leadership may prioritize the database differently, as it is not a common activity for SHPOs of other states to host such databases. As of early 2014 there was no link to the Census Project web page from the main Nevada SHPO site, and it could not be found through searching the site, but it remained visible through Google and available at: http://nvshpo.org/index.php?option=com_census&view=demographics&Itemid=388 Due to its lack of visibility on the NVSHPO site, it likely became appropriate to presume that the database's future there could be short. By November 2014, the database had been removed from the SHPO site entirely.

In late 2012, historian Eric Nystrom began working with James to obtain a complete copy of the database, efforts that shifted in 2013-14 toward fully documenting the dataset and making it available to researchers as a complete set on a widely-available, stable data sharing platform. New NVSHPO leadership participated fully in arranging for permission for this effort. Nystrom subsequently worked to create documentation and archive the data set to ensure its continuing availability to the research community.[4] Future phases of work will include efforts to rework the data to be compatible with other existing microdata sets (such as those produced by NAPP/IPUMS) and to further improve the consistency and correctness of the data.

Also see the "Bibliography" section below.

6 Bibliography

This section contains references to work that has used the dataset in any of its earlier forms. The list below is preliminary; please contact Eric Nystrom (see Contact information below) with other

[1] Kenneth H. Fliess, "Nevada's Federal Manuscript Censuses: Volume 1: The Existing 19th Century Censuses–1860, 1870, and 1880," 3-4. This was a booklet apparently intended to accompany the CD-ROM. The booklet was at one time available on the Web but has since apparently been taken down, though it is available through the Internet Archive's Wayback Machine. http://web.archive.org/web/20060925165603/http://dmla.clan.lib.nv.us/docs/shpo/NVCensus/docs/CensusIntro.pdf

[2] Fliess, "Nevada's Federal Manuscript Censuses..."

[3] This first web-based incarnation was housed at http://dmla.clan.lib.nv.us/docs/shpo/NVCENSUS and was active until at least June 2008.

[4] Email correspondence with Eric Nystrom, Ron James, and Karyn De Dufour, Dec. 21-26, 2012; Ron James, Rebecca Palmer, Karyn De Dufour, and Eric Nystrom, Dec. 16-19, 2013, copy in possession of author.

entries. Additionally, the dataset has been used by countless users through the SHPO website for more than 10 years. It is known to have been used particularly for historic archaeology projects in the Comstock Historic District and elsewhere, and it has perhaps naturally gained attention within the genealogical community.

- Ronald M. James, "Women of the Mining West: Virginia City Revisited," *Nevada Historical Society Quarterly* 36, no. 3 (1993): 153-177.
- Ronald M. James, Richard D. Atkins, and Rachel J. Hartigan, "Competition and Coexistence in the Laundry: A View of the Comstock" *Western Historical Quarterly* 25, no. 2 (1994): 164-184.
- Ronald M. James, *The Roar and the Silence: A History of Virginia City and the Comstock Lode* (Reno: University of Nevada Press, 1998).
- Ronald M. James and C. Elizabeth Raymond, eds., *Comstock Women: The Making of a Mining Community* (Reno: University of Nevada Press, 1998). Many of the individual essays in this collection make extensive use of the data set.
- Kenneth H. Fliess, "There's Gold in Them Thar–Documents? The Demographic Evolution of Nevada's Comstock, 1860 through 1910, and the Intersection of Census Demography and Historical Archaeology" *Historical Archaeology* 34, no. 2 (2000): 65-88.
- Ronald M. James, "Disasters on the Mining Frontier: A Look at Two Events on the Comstock" *Mining History Journal* 12 (2005): 25-39.
- Ronald M. James and Kenneth Fliess, "The Census and Demography of Nevada," *ONE: Online Nevada Encyclopedia* (19 July 2010), http://www.onlinenevada.org/articles/census-and-demography-

7 Contact Information

Eric Nystrom, Associate Professor of History at the Rochester Institute of Technology, currently maintains the dataset. Please send any additions and/or corrections to Eric:

- Email: eric.nystrom@rit.edu
- Twitter: @HistoryNystrom
- Phone: 585-475-6172
- Mail: 92 Lomb Memorial Drive, Rochester NY 14623

8 Technical Documentation

This section describes the dataset from a technical perspective, including documentation of the size, source, scope, and purpose of each field in the data.

In general, there are consistency problems with the data, including multiple ways of recording the same information, as well as a number of typos. This begins with the census sources themselves, which were not always accurate, and which are frequently difficult to read because of archaic handwriting. Additionally, transcribers introduced irregularities through mis-readings,

typos, and following the historical sources too closely, transcribing their irregularities with fidelity. Over time, we expect to refine many of these errors out of the data, but researchers should be aware of them if sorting or searching on particular fields. Such irregularities matter less if the data is used to find individuals (i.e. genealogy). An effort has been made in the documentation for each field to identify particularly troublesome or noteworthy irregularities.

The "years available" information comes from a listing of columns displayed on an old version of the SHPO website.[5] These indicate theoretical availability; the data should be checked to ensure that non-available fields provide a NULL value.

According to Fliess, the word "BLANK" was used any time an enumerator left a field blank on the census form. This was done to make data checking proceed easier, and to help users recognize that data coders did not accidentally skip fields.[6] This practice was often, but not universally followed.

Many fields contain NULL values in particular fields, most commonly when that field was not collected in a particular census year.[7] However, researchers will want to be careful with these. In numeric fields, NULL was often, though not always, represented by zero (0). Occasionally NULL values were not successfully converted, leaving their marker, a , instead. Sometimes fields were left blank instead of being properly converted to NULL. Further complicating matters is that while some databases are comfortable with NULL values, other programs – particularly some statistical packages, as well as the database system used by the Harvard Dataverse – do not support the NULL concept. This means NULLs will have to be recoded to some other value for maximum compatibility in the future. None of this should impact the genealogical researcher.

The data set contains 310,486 records.

8.1 RecordID

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Maximum length = 6
- Range: 1–310486

This is an integer unique ID for each record in the data set, consisting of up to six digits. It may not be NULL, must be unique, and is suitable for use as a primary key. Users should **NOT** count upon "recordid" being sequential in the order of people on the manuscript page, though it is frequently the case that they are. In a handful of cases where a line was inadvertently skipped in initial entry and a record added later, the "recordid" will be non-sequential.

8.2 Year

A four-digit integer indicating the census year of each record. Not NULL.

[5] http://web.archive.org/web/20070620074219/http://dmla.clan.lib.nv.us/docs/shpo/NVCensus/help/ColumnInfo.asp

[6] Fliess, "Nevada's Federal Manuscript Censuses...", p. 4.

[7] In database terms, NULL is neither true nor false, it means essentially that it just doesn't apply. For example, street addresses were not collected in the 1860 and 1870 census, but they were in 1880. If the field "StreetName" is blank in an 1880 record, that would clearly mean something different than if it is blank in an 1870 record, because no 1870 record would have a street name. So, done properly, the 1870 records would all have a NULL value for "StreetName," but the 1880 field in our example would actually be blank. Proper use of NULLs can help prevent problems when counting and summing, especially across years.

- Maximum length = 4

Table 1: Number of records for each census year

Year	Count
1860	6853
1870	41670
1880	62256
1900	42268
1910	81768
1920	75671

8.3 YrCoPgLine

- Integer, Maximum length = 5
- no NULL values found

It is not apparent what this field represents. Neither of the living co-authors can figure out what this is supposed to be from memory or surviving documentation.

If any user knows what this field is for, please contact Eric Nystrom via email (see Contact).

This field may be slated for removal in future releases of the dataset.

8.4 Page

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 3
- no NULL values found
- Range: 0–268

This is intended to be the manuscript page number (with A/B disregarded), handwritten on the census form, **not** the later number stamped on it.

This can be used with fields "enumerator" and "line" to find an individual on the manuscript census forms. The field "county" will be needed to determine the correct roll of microfilm. The field "supervisor" may also be of assistance in locating the correct page.

Only 29 records with PAGE value of zero.

8.5 Subunit

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String, maximum length = 25
- 178 NULL values, 8 in 1910, 170 in 1920.

This is the location from the manuscript census form. Often this is a city, town, or settlement name. In larger settlements this might be a ward or other political subunit. These appear to be as-transcribed from the manuscript census. For 1900 and later, there are two fields on the manuscript form, one asking for the "township or other division of county" and the other for the name of the incorporated place. This field captures the township field. See the "city" field for the incorporated place names for those years.

8.6 County

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String, maximum length = 10
- No NULL values
- 142 "No_{Info}" records

This is the county name as recorded at the time of the census. (For example, Clark County, formed in 1908, is only listed in the 1910 and 1920 censuses; before that, people in present-day Clark County would have been found in Lincoln County.)

142 entries have the string "No_{info}" listed. These are all in the 1910 census.

1860 is a special case, because Nevada Territory was not organized until March 1861. The dataset contains records from Carson, Humboldt, and St. Mary's Counties in Utah Territory. This seems to exclude any possible residents living in parts of New Mexico Territory (later Arizona Territory) that were annexed to Nevada later; as well as Utah residents that became Nevada residents when the Nevada border was moved eastward in 1864 and 1866.

Table 2: Counties for 1860 records (Utah Terr., present-day Nevada)

County	1860 pop.
CARSON	6708
HUMBOLDT	40
ST. MARY'S	105

Possible values: CARSON, CHURCHILL, CLARK, DOUGLAS, ELKO, ESMERALDA, EUREKA, HUMBOLDT, LANDER, LINCOLN, LYON, MINERAL, NYE, No_{Info}, ORMSBY, PERSHING, ROOP, ST. MARY'S, STOREY, WASHOE, WHITE PINE

8.7 Day CONVERTNULL

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 2

This was the day of the month that the census was recorded in this particular place. Some records have a zero value. There are also some apparent errors, with values of 53, 81, 96, 98, and (supposed to be NULL).

8.8 Month CONVERTNULL

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 2

This the month that the census was conducted, converted into a numeric month. However, there are a number of irregularities in this column, as well as some values that may need to be converted to NULLs.

8.9 Line

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 3
- no NULL values

Line number on the manuscript census page of the entry. Generally ranges from 1-100 (with numbers from 51-100 on double-page census sheets).

There are a handful of unusual values. 21 records are "0" and one is "101"; most of these appear to be mistakes, but it is also possible these cases might be where an enumerator added a person at the bottom or top of a page.

8.10 HHN

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 4

Household Number, as recorded on the manuscript census forms. Enumerators numbered them in the order of their visitation. Should attempt to clarify and/or fix HHN = 0 and values.

Table 3: Records with HHN = 0 by year (total 5127)

Year	Count
1870	322
1880	2245
1900	21
1910	472
1920	2067

8.11 FN

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 4

Family Number, from the census form. Enumerators numbered them in the order of their visitation. 24640 records have FN = "0", and 8 others have unconverted NULL values ().

Table 4: Records with FN = 0 by year (total 24640)

Year	Count
1870	17295
1880	5740
1900	9
1910	183
1920	1413

8.12 FName

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String (uppercase), maximum length = 18
- 303 NULL values

First name, as listed on the census form. Includes unexpanded abbreviations such as "WM." for William, and also includes middle initials such as "JOHN H."

If first name not given, value is NULL. Many, though not all, records with NULL "Fname" values are for people who are Chinese or Native American. (See also records with a "FName" but no "Surname.")

Table 5: Records where FName value is NULL, by year (n = 303)

Year	Count
1870	50
1880	139
1900	14
1910	42
1920	58

8.13 Surname CONVERTNULL

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String (uppercase), maximum length = 12
- one blank value in 1880 and one in 1900
- 884 NULL values

Person's last name or surname, as listed on the census form.

884 records have a NULL "Surname," and two others have a blank entry, which was likely to have been intended to be NULL. These records are often minorities such as Chinese or Native Americans.

Table 6: Records where Surname value is NULL, by year (n = 884)

Year	Count
1870	154
1880	607
1900	99
1910	20
1920	4

8.14 $\text{Age}_{\text{Years}}$

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 3
- range: 0-100, exclusive of NULLs
- 148 () NULL values, need to convert
- 6965 records where $\text{Age}_{\text{Years}} = 0$

Fliess makes it clear that this field ($\text{AGE}_{\text{YEARS}}$) is the age information as recorded by the census, where the AGE field, below, was computed by the entry program and appears to have subsequently been damaged.

Fliess cautions that there appears to be evidence of "age-heaping" by the enumerators:

> There seems to be some degree of age-heaping, or ages ending frequently with a 5 or 0. As far as I can tell, the foreman or some other individual was reporting the names and ages of those [Chinese or other minorities] who worked under him. In all probability when the enumerator polled the foreman for information, in addition to giving his best translation of the name he also probably said something to the effect of 'he's about 20' or 'he's about 35'. There are statistical methods to test for age-heaping and to smooth out age distributions that are age-heaped (Shryock and Siegel, 1976), though these have not been done because the goal here was to preserve, as closely as possible the original manuscript data, not analyze them.[8]

8.15 $\text{Age}_{\text{Months}}$

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 2
- value of 0 when not applicable/null
- 3398 NULL ()
- possible error, value of "30"
- non-zero values in every census year

[8] Fliess, "Nevada's Federal Manuscript Censuses...", p. 6-7.

Need to convert to NULL. (though in the long run these should probably be converted to zeros to be consistent).

This variable is supposed to be the age, in months, of any person who is not yet one year old, I think. However, a quick comparison with "Age_{Years}" suggests it may have occasionally been recorded for people more than one year old.

8.16 Age

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 3
- range: 0-100
- 11 NULL () values, need to be converted to NULL
- 4936 records where "Age" = 0,

DO NOT USE THIS FIELD, IT IS INCORRECT.

Fliess notes that this was computed by the input program from the years and months recorded on the census. Unfortunately, this was computed as a decimal, and it appears likely that the field was not correctly converted at some point during a later data transfer, causing the decimals to be lost. As a result, this field is unneeded and also incorrect, and therefore should not be relied upon. (This field is a candidate for removal in future releases of the dataset.)

There are 4179 records where "Age" != "Age_{Years}".

8.17 FiveYear

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- Integer, maximum length = 2
- 8 NULL values () – need to convert

This is a constructed variable that permits easy grouping into five-year age blocks. So FIVEYEAR = 60 will include all records with ages of 60, 61, 62, 63, or 64.

There are at least a few errors in this. There are a handful of records that are NULL or have a FIVEYEAR value that is not actually a five-year division. There also appear to be some records that have ages outside the five-year span, though it is not clear whether that means that AGE or FIVEYEAR is the incorrect value. (This should be fixed in a future release.)

According to Fliess, a value of "99" indicates no age was reported for the individual in the census.

8.18 Sex

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String, maximum length = 1
- 157 blank
- values M, F

This is a single character, M for Male and F for female. There are 157 blank, and 135 records where the character is something other than M,F, or blank due to a typo. These should be fixed.

8.19 Color

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String, maximum length = 1
- 61 blank

This is a single-character code to indicate "race" or "color" as recorded by the census-takers. The possible choices ought to be constrained by the question asked each census year. The table below lists the known codes. The entire universe of codes is listed in another table. Some codes with low counts are likely to be errors; for example, the six from code "K" are Asian names. Though it is possible they are Korean, it is unlikely that the census recorded them this way. Similarly, the names of code "A" appear to be Asian names, mostly Chinese but with some apparently Japanese.

Fliess apparently attempted to adjust the racial codes by converting all "Mulatto" entries to "Black," as the former was "a truly abhorrent term, idea, and category."[9] Modern researchers are split on the value of this modification.

Note also that according to Fliess, Native American populations were for the most part ignored in 1860 and 1870, "unless they were living in a Euro-American context, though it is not clear that even in this type of context they were enumerated on a regular basis."[10]

1860 asked for Black/White/Mulatto. 1870, 1880 asked Black, White, Mulatto, Chinese, Indian.

Table 7: Expected COLOR code values

Code	Value
B	Black/Negro
C	Chinese
I	Indian/Native American
J	Japanese
M	Mulatto
W	White/Caucasian

8.20 Profession

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String (uppercase), maximum length = 30
- 6522 records are NULL
- 10 are blank

This is the occupation as recorded in the census, transcribed. Blank or no profession appear to have been handled in different ways, unfortunately. Many are NULL, a few are genuinely blank, but 67,194 have the string value NONE and another 32,664 have the string value BLANK.

[9] Fliess, "Nevada's Federal Manuscript Censuses...", p. 5-6.
[10] Fliess, "Nevada's Federal Manuscript Censuses...", p. 6.

Table 8: COLOR code values, possibly with errors

Color	Count
	61
A	39
B	1427
C	11506
I	18011
J	1736
K	6
M	74
O	3
P	4
U	5
W	277613
w	1

There are a variety of spellings, but it is not known whether an effort was made to preserve alternate spellings during transcription.

The enumerators' instructions directed that this field be filled out for those "over 15 years of age" (1860). In 1880, this question was not supposed to be asked of people under 10 years old.

8.21 POBirth

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String (uppercase), Maximum length = 20
- 151 NULL entries, 628 entries with string "BLANK", 218 with string "UNKNOWN"
- 32 entries with string "TORN PAGE"
- 75 entries with string "ILLEGIBLE"

This variable records an individual's place of birth, which in the manuscript census could be a state, if born within the United States, or a Country Name if born outside the US. These have (mostly) been coded by the transcribers into standard two-digit codes for those born in the United States, and non-US births have been transcribed in full. There are a few exceptions, however. There are 126 entries for "UT TERRITORY", 88 for "KA" (?), 946 for "USA"; 30 for "INDIAN TERRITORY"; 26 for "DAKOTA" though "SD" and "ND" are also both represented; 27 for "EUROPE"; 23 for "ID TERRITORY"; 19 for "SANDWICH ISLANDS" though others are "HI" for modern Hawaii; 15 for "AMERICAN CITY"; 14 for "WA TERRITORY"; 12 for "AMERICA"; and others along a similar vein, especially those with state/territory names. A handful are American city names, such as "CINNCINNATI". Many of the foreign places are richly evocative and smaller units than countries alone.

8.22 School

- Years available: 1860, 1870, 1880, 1900, 1910, 1920

- String, maximum length = 2
- 103503 are NULL

This variable contains potentially non-commensurate information as the question it represents was asked differently from one census year to another.

In 1860, answers were N and Y. ("Attended School within the year.")

In 1870 and 1880, answers were N, Y, many NULL, and perhaps a few errors. No age was specified in the question in 1870.

In 1900, answers were given in numeric form. The preponderance of answers were "0" (zero), with a small handful of NULL values.

In 1910, most answers were NULL; most of the non-NULL answers were Y or N, with some other answers including blanks and a variety of typos.

In 1920, most answers were NULL; the rest were U, Y, and N, with a few other blanks, typos, etc.

8.23 Read

- Years available: 1860, 1870, 1880, 1900, 1910, 1920
- String, maximum length = 1
- 39,256 are NULL

Most manuscript censuses asked if a person could read and/or could write. Some of these questions were negative ("cannot read") and were answered only if true; in other words, if the form was not marked, it was assumed that the person could read. These answers appear to have been coded into Y/N answers by the transcribers. There are a significant number of NULL values, which are distributed across all years. Since a non-answer on this question should result in a "Y" in most cases, this deserves further investigation. There are also a handful of typos.

1860 asked about "Persons over 20 years of age who cannot read or write."

In 1880, this question was not supposed to be asked of people under 10 years old.

8.24 PostOffice

- Years available: 1860, 1870
- All 1860 entries are "BLANK" as the enumerator apparently did not record them.
- Need to convert blank entries in other years to NULL.

This information comes from the header of each census page, where in 1860 and 1870 both the post office name and the locality were asked. However, the Carson County, Utah enumerator apparently did not record any value in this field, meaning essentially 1870 is the only year with this data.

This is most helpful for extremely small or remote localities whose names might not have survived to the present day.

8.25 RealEstate

- Years available: 1860, 1870
- Integer; value is zero for all other years

"Value of Real Estate," enumerated on a person-by-person basis. Also see "perestate" (Personal Estate). These are numeric values, presumably in dollars from the year of the census.

1860 has 3146 (unresolved NULL); 1870 has 14. 1870 also has 7 entries with value "99999" which suggests either an error or that 1870 was only 5 digits and a six-digit value was encountered.

8.26 PerEstate

- Years available: 1860, 1870
- Integer, value is zero for all other years.

"Value of Personal Estate." Enumerated on a person-by-person basis. Also see "realestate." 1860 has 3401 entries of zero, and another 3046 (unresolved NULL). 1870 has 33996 zeros, 8 "99999" values (see realestate), and 13 unresolved nulls ().

8.27 Deaf

- Years available: 1860, 1870, 1880, 1910
- Values: Y/N

Deafness was typically collected by making a slash mark on the manuscript census form. Non-responses were interpreted as "no" by the coders.

1870 has a few out-of-bounds answers. 1880 has a small handful of non- Y/N answers with a wider range including "S" and "U". 1900 appears to have some blanks where NULLS are needed, same for 1920 (the question was not asked on either year's form). 1910 was not supposed to be available, according to initial dataset documentation, but it clearly is; mostly N, but a small spectrum of U, M, H, E, Y, C, and blank.

8.28 MarStat

- Years available: 1870, 1880, 1900, 1910, 1920
- String, one character.

Marital status: Single, Married, Widowed, or Divorced. Should be S, M, W, D; most years also have some number of responses outside these bounds, some of which are clearly typos, but others may represent actual responses. I suspect U is supposed to be Unknown.

8.29 FatherFor

- Years available: 1870
- Responses: Y/N

Mark if Father is of foreign birth. Handful of non-null, but blank, responses in other years, and one error in 1870.

8.30 MotherFor

- Years available: 1870
- Responses: Y/N

Mark if Mother is of foreign birth. Handful of non-null, but blank, responses in other years, and one error in 1870.

8.31 MonthBorn

- Years available: 1870, 1880, 1900
- Integer, plus some unresolved NULLs

"If born within the year, state month (Jan., Feb., &c.)." This was converted to a numeric month by the enumerators and/or transcribers.

8.32 MonthMarr

- Years available: 1870, 1880
- Integer; 0 if N/A

"If married within the year, state month (Jan., Feb., &c.)." This was converted to a numeric month by the enumerators and/or transcribers.

8.33 Write

- Years available: 1870, 1880, 1900, 1910, 1920
- String, with some garbage in need of cleaning

This question asked if the person could write (in some years, English was specified). In 1880, this question was not supposed to be asked of people under 10 years old.

8.34 Citizen

- Years available: 1870
- String, two characters

The census form asked whether the person was a "Male citizen of the U.S. of 21 years of age and upwards." The dataset contains LOTS of potentially confusing answers. Further investigation is needed and caution is urged in use of this field.

8.35 Vote

- Years available: 1870
- String Y/N plus some blanks (and nulls of course)

The question on the form asks for an indication of whether the person CANNOT vote, but in transcribing apparently this was switched, so a "Y" answer indicates that voting is possible. (There are 19,523 "Y" and 4,475 "N".) As stated on the form: "Male citizen of the U.S. of 21 years of age and upwards whose right to vote is denied or abridged on other grounds than rebellion or other crime."

8.36 Supervisor

- Years available: 1880, 1900, 1910, 1920
- Integer, with some unresolved nulls; N/A seems to be zero
- Maximum length = 3

This is the number of the "Supervisor's District" from the manuscript census form. Few resources utilize this number but it can be helpful as a check against typographical errors in other fields.

8.37 Enumerator

- Years available: 1880, 1900, 1910, 1920
- Integer, maximum length = 3; N/A is probably zero

Enumeration district, as reported on manuscript census sheet. This is NOT the name of the enumerator.

8.38 StreetName

- Years available: 1880, 1900, 1910, 1920
- String

Street name, as provided by the enumerator in urban areas (written sideways on the far-left of the population schedule). Transcribers made few efforts to standardize street names, resulting in (to choose one example: "SIXTH STREET", "SIXTH STREET" (note extra spaces), "SIXTH ST.", "6TH ST.", "6", and "6TH STREET", not to mention variations with N/S/E/W etc. Thus caution and ingenuity is necessary to use this information.

8.39 StreetNumb

- Years available: 1880, 1900, 1910, 1920
- String, maximum length: 12

Street House Number. When an address was provided by the enumerator, this field was intended to capture the number portion. Typically though not always numeric, as there are some commas, some letters (307-A, etc), some of what are clearly "halves" in some cases made into .5, etc.

8.40 MainRel

- Years available: 1880, 1900, 1920
- String, maximum length = 2

"Main Relation." This used a system developed by Gutmann and Holmes, 1988, to code household relationships. It uses MAINREL, HEADREL, and REL2 together to provide different levels of detail.[11] (Also see explanations for HEADREL and REL2 below.)

According to this system, MAINREL has five possible choices.

Table 9: MAINREL code choices, after Gutmann and Holmes

Code	Relationship
H	Head of Family
W	Wife of Head
C	Child of Head
R	Related - other
U	Unrelated other

If MAINREL is H/W/C, then coding is done and HEADREL and REL2 are empty.

8.41 HeadRel

- Years available: 1880, 1900, 1920
- String, maximum length = 6

This used a system developed by Gutmann and Holmes, 1988, to code household relationships. It uses MAINREL, HEADREL, and REL2 together to provide different levels of detail.[12] (See detail for MAINREL, above, and REL2, below.)

HEADREL only has a value if MAINREL is "R", indicating a Related other relationship.

HEADREL indicates if the person is related to the Head of household or the Wife of the head-of-household, using H and W respectively. If HEADREL has a value, then REL2 explains that relationship in greater detail.

8.42 Rel2

- Years available: 1880, 1900, 1920
- String, maximum length = 6

This used a system developed by Gutmann and Holmes, 1988, to code household relationships. It uses MAINREL, HEADREL, and REL2 together to provide different levels of detail.[13] (Also see explanations for MAINREL and HEADREL above.)

REL2 contains additional data about Related or Unrelated Other relationships, as captured by MAINREL.

[11] Fliess, "Nevada's Federal Manuscript Censuses...", p. 43-45
[12] Fliess, "Nevada's Federal Manuscript Censuses...", p. 43-45
[13] Fliess, "Nevada's Federal Manuscript Censuses...", p. 43-45

Table 10: Possible Other Relationship values (REL2)

MAINREL	HEADREL	Relationship	REL2
R	H/W	Parent	PAR
R	H/W	Niece or Nephew	NIE
R	H/W	Cousin	COU
R	H/W	Grandchild	CHI
R	H/W	Brother or Sister	BRO
R	H/W	Grandparent	GRA
R	H/W	Brother/Sister-in-law	LAW
R	H/W	Uncle or Aunt	UNC
R	H/W	Son/Daughter-in-law	SON
R	H/W	Adopted Child	ADO
R	H/W	Child of a Previous Marriage	PRE
R	H/W	Illegitimate Child	ILL
R	H/W	Mother/Father-in-law	MOT
R	H/W	Other	OTH
R	H/W	Unknown	UNK
U	(blank)	Servant	SER
U	(blank)	Foundling	FOU
U	(blank)	In Jail	JAI
U	(blank)	Soldier	SOL
U	(blank)	Professional Helper	HEL
U	(blank)	Kids Without Parents	KID
U	(blank)	Religious Affiliation	REL
U	(blank)	Lodgers and Boarders	LOD
U	(blank)	Other	OTH
U	(blank)	Unknown	UNK

8.43 MosUnempl

- Years available: 1880, 1900, 1910
- Integer, with unresolved nulls, max of 2 digits

There's a note that hints at a desire by Fleiss to convert this to weeks unemployed so that 1910's answers are comparable to 1880/1900, but that does not appear to have been done.

In 1880, this question was not supposed to be asked of people under 10 years old. Also that year, this question was only to be asked if an occupation had been recorded as well.

8.44 Sickness

- Years available: 1880
- String, maximum length = 24

This field captures column 15 on the 1880 population schedules, which asks: "Is the person (on the day of the Enumerator's visit) sick or temporarily disabled, so as to be unable to attend to ordinary business or duties? If so, what is the sickness or disability?"

The enumerator was supposed to leave this field blank if not applicable, and those records were transcribed as "BLANK". All non-1880 records should be NULL but there are 670 blank (rather than properly null) records scattered through other years that should be fixed.

There are 502 non-BLANK entries in 1880.

8.45 Blind

- Years available: 1880, 1910
- String, single character with some debris

If person was blind, the enumerator was supposed to make a slash mark in column 16 of the 1880 manuscript census. The website claims that this was only collected for 1880, but it is clear it was captured for 1910 as well. This is supposed to have been transcribed as a Y/N answer, though there are some non-Y/N answers that should be further investigated; most are likely to be typos but it is possible a few capture non-standard answers on the population schedules. There are also some blank answers in other years to be converted to proper NULLs.

8.46 Idiotic

- Years available: 1880
- String, single character with some debris

If the person was "idiotic", the enumerator was supposed to make a slash mark in column 18 of the 1880 manuscript census. "Idiocy" in this context typically was understood as mentally disabled from birth or early childhood (in opposition to "insanity" where a person who has reason loses it). This is supposed to have been transcribed as a Y/N answer, though there are some non-Y/N answers that should be further investigated; most are likely to be typos but it is possible a few capture non-standard answers on the population schedules. There are also some blank answers in other years to be converted to proper NULLs.

8.47 Insane

- Years available: 1880
- String, single character with some debris

If the person was "insane", the enumerator was supposed to make a slash mark in column 19 of the 1880 manuscript census. "Insanity" in this context typically was understood as having at one time possessed reasoning but having lost that (as opposed to "idiocy" where a person was mentally disabled from birth or early childhood). This is supposed to have been transcribed as a Y/N answer, though there are a handful of non-Y/N answers that should be further investigated; most are likely to be typos but it is possible a few capture non-standard answers on the population schedules. There are also some blank answers in other years to be converted to proper NULLs.

8.48 Crippled

- Years available: 1880
- String, single character with some debris

The 1880 population schedule's 20th question asked enumerators to make a mark if the person was "maimed, crippled, bedridden, or otherwise disabled." This is supposed to have been transcribed as a Y/N answer, though there are a handful of non-Y/N answers that should be further investigated; most are likely to be typos but it is possible a few capture non-standard answers on the population schedules. There are also some blank answers in other years to be converted to proper NULLs.

8.49 FPOBirth

- Years available: 1880, 1900, 1910, 1920
- String

Father's Place of Birth. Apart from an apparent attempt to use two-letter state codes for state names within the US, little effort seems to have been made to regularize these. States and countries are intermixed here. Most state answers have been entered as the two-character state postal code (such as "NV") but there are a handful of exceptions which should be corrected.

8.50 MPOBirth

- Years available: 1880, 1900, 1910, 1920
- String

Mother's Place of Birth. Apart from an apparent attempt to use two-letter state codes for state names within the US, little effort seems to have been made to regularize these. States and countries are intermixed here. Most state answers have been entered as the two-character state postal code (such as "NV") but there are a handful of exceptions which should be corrected.

8.51 City

- Years available: 1900, 1910, 1920
- String, maximum length = 24

E.g. 1910: "Name of Incorporated Place." Instructions were to "insert proper name and, also, name of class, as city, village, town, or borough." Note that this is different than the "township or other division of county" field on the forms, which appears to have been captured for these years in the "subunit" field of this dataset.

Data in this field tends to be inconsistent, sometimes including the name of the class and other times not doing so, apparently following the practice of each enumerator. Records for unincorporated locations were supposed to be left blank and were transcribed as "BLANK" in the dataset.

8.52 Institutio

- Years available: 1900, 1910, 1920
- String, maximum length = 24

"Institution" is a line on the header of the Population Schedule for 1900, 1910, 1920, to indicate when people being enumerated live in some sort of formal group quarters such as a jail, an Indian reservation, a mine boardinghouse, or a hospital. There are some inconsistencies in this field, including multiple spellings/abbreviations of the same institution over time. Some seem to have been badly abbreviated. Non-answers are supposed to be coded "BLANK" but there are some that are "NONE."

Note the missing "n" on the field name.

8.53 Ward

- Years available: 1900, 1910, 1920
- String, maximum length = 24

Field "Ward of City" from the population schedule, left blank if it was an unincorporated place. There are a handful of irregularities with the data, including misspellings.

8.54 Rank

- Years available: 1900, 1910, 1920
- Integer, two digits

Rank represents the order within the household of the records as they were recorded by the enumerator. Seems to be equivalent to the PERNUM field in the NAPP/IPUMS system. It is not clear why this information was only collected for 1900, 1910, and 1920, as it could conceivably been collected for earlier years. Rank of zero is NULL or N/A.

8.55 YearBorn

- Years available: 1900
- Integer, with unresolved NULLs

This field collects information from a wonderful question that was only asked by the 1900 manuscript census: "Date of Birth, Year" (column 7). The field is 4-digit integer, NULL value represented by zero, though there are a few unresolved NULLs in there ().

8.56 YrsMarried

- Years available: 1900, 1910
- Integer, 2 digit, with some unresolved nulls

Question "number of years married," column 10 in 1900. In 1910, column 9 asked "number of years of present marriage." NULL in other years should be zero; non-married persons in 1900 and 1910 should be zero also.

8.57 NumChildren

- Years available: 1900, 1910
- Integer, 2 digit, with some unresolved nulls

Column 11 in 1900 asked "Mother of how many children." Column 10 in 1910 asked "Mother of how many children: number born." Both years had follow-up questions to determine how many of those children were still alive, see field "NumAlive." Only females should have non-zero responses; NULL in other years should be zero. Some unresolved NULLs ().

8.58 NumAlive

- Years available: 1900, 1910
- Integer, 2 digit, with lots of zeros and some unresolved nulls

Number of living children, asked of mothers (see NumChildren)

Column 12 in 1900 asked "Number of those children living." Column 11 in 1910 asked "Mother of how many children: number now living." Both years had asked previous questions to determine how many children each woman mothered, see field "NumChildren." Only females should have non-zero responses; NULL in other years should be zero. Some unresolved NULLs ().

8.59 YearImgrt

- Years available: 1900, 1910, 1920
- Integer, 4 digit, plenty of typos, zeros, and unresolved nulls

Year the person immigrated to the United States if foreign-born. For example column 16 on 1900 population schedule asked for the "Year of immigration to the United States," and column 15 from 1910 had the same wording.

There are lots of unresolved NULLs, also some typos. The expectation would be that a zero would represent a non-answer.

8.60 YrsInUS

- Years available: 1900
- Integer, 2 digit, with lots of zeros and some unresolved nulls

Number of years the person has lived in the United States if foreign born, asked by column 17 on the 1900 population schedule. Should be left blank if the person was not an immigrant, and these should be converted to zeros. Also there are some unresolved NULLs.

As an additional check, this could be compared with "YearImgrt" to find possible transcription errors in those fields.

8.61 Naturalize

- Years available: 1900, 1910, 1920
- String, one character.

Question 18 on the 1900 form simply states "Naturalization." On the 1910 form, question 16 asks "Whether naturalized or alien." (A very spotty glance suggests enumerators put "NA" for naturalized, "AL" for alien, on the handful of sheets I examined.) Question 14 on the 1920 form asks whether naturalized or alien as well.

This dataset field has a range of answers and should be used with care. The majority are "A" and "N" but there are lots of others that need to be investigated. Chances are good most of these are typos but it is possible some are accurately-transcribed edge cases.

8.62 English

- Years available: 1900, 1920
- String, should be a single character, much bigger range than there should be, so some garbage here.

Question 24 on the 1900 form was "Can speak English." Question 25 asked "whether able to speak English." Also note that a different question on the 1920 form documented the person's mother tongue, captured in this dataset's "PLanguage" field.

This dataset field has a range of answers and should be used with care. The majority are "Y" and "N" but there are others that need to be investigated, including a number of "U" answers (presumably "unknown"). Chances are good many of the others are typos but it is possible there are some additional accurately-transcribed edge cases.

8.63 Own

- years available: 1900, 1910, 1920
- String, should be a single character, much bigger range than there should be, so some garbage here.

This field captures the answer to 1900's question 25 ("Ownership of Home: Owned or rented"), 1910's question 26 (same wording), and 1920's question 7 ("Home owned or rented"). Most responses are "O" (oh), or "R" for Owned or Rented, respectively. The data also has a number of other responses which are likely typos, though the "U" answers might be "unknown." Non-heads of household should have NULL even in featured years.

8.64 Free

- Years available: 1900, 1910, 1920
- String, should be a single character, much bigger range than there should be, so some garbage here.

This field captures the answer to 1900's question 26 ("Ownership of Home: Owned free or mortgaged"), 1910's question 27 (same wording), and 1920's question 8 ("If owned, free or mortgaged"). Most responses are "F" or "M" for Owned Free or Mortgaged, respectively. The data also has a number of other responses which are likely typos, though the "U" answers might be "unknown." Non-heads of household should have NULL even in featured years.

8.65 Farm

- Years available: 1900, 1910
- String, should be a single character, much bigger range than there should be, so some garbage here. Schedules asked whether it was a Farm or a House.

This field captures the response to 1900's question 27 and 1910's question 28, asking whether the home (owned or rented; owned free or mortgaged; see fields "own" and "free") was a Farm or a House. Most answers are "F" or "H" for Farm and House, respectively. There are a smattering of other answers which are likely typos and should be resolved, but over all these answers are cleaner than many others. Presumably non-heads of household should have NULL even in featured years.

8.66 SchedNum

- Years available: 1900, 1910, 1920
- Integer, up to three digits.

This the response to the follow-up question asking if it was a farm, what the farm schedule number was. In 1920, the last column of the page (field 29) asked for the "number of farm schedule" which presumably was only filled in for farms.

Non-farms in both available years and non-available years given a zero. (No NULLs)

Some of these numbers seem quite high and might be worth spot-checking.

8.67 PLanguage

- Years available: 1910, 1920
- String, maximum length 24

The person's "mother tongue" or primary spoken language. In 1910, this represents field 17, "Whether able to speak English; or, if not, give language spoken." For 1920, this is question 20, part of the series of questions about the place of birth and mother tongue of the person and each parent.

This field should be used with care, as the transcribers seem to have faithfully reproduced the varied spellings of enumerators, and likely introduced additional typographic errors and mistakes.

8.68 MLanguage

- Years available: 1910, 1920
- String, maximum length 24

Mother's Language and/or birthplace. In 1910, this represents field 14, "place of birth of the mother of this person." For 1920, the person's mother's "mother tongue" or primary spoken language. This is question 24, part of the series of questions about the place of birth and mother tongue of the person and each parent.

This field should be used with care, as the transcribers seem to have faithfully reproduced the varied spellings of enumerators, and likely introduced additional typographic errors and mistakes. It also appears that for the 1910 census, transcribers coded this field NULL if the answer indicated that the mother was born inside the US.

8.69 FLanguage

- Years available: 1910, 1920
- String, maximum length 24

Father's Language and/or birthplace. In 1910, this represents field 13, "place of birth of the father of this person." For 1920, the person's father's "mother tongue" or primary spoken language. This is question 22, part of the series of questions about the place of birth and mother tongue of the person and each parent.

This field should be used with care, as the transcribers seem to have faithfully reproduced the varied spellings of enumerators, and likely introduced additional typographic errors and mistakes. It also appears that for the 1910 census, transcribers coded this field NULL if the answer indicated that the father was born inside the US.

8.70 WorkStatus

- Years available: 1910, 1920
- String, single character. Might be a bit of garbage.

Occupation – Work Status. 1910, question 20: "Employer, employee, or working on own account." 1920, question 28: "Employer, salary or wage worker, or working on own account."

A handful of spot-checks suggests that this field has been unreliably transcribed and must be used with substantial caution.

8.71 YearNat

- Years available: 1920
- Integer, lots of zeros and unresolved nulls

Question 15 on the 1920 form asked "If naturalized, year of naturalization."

This should only apply to naturalized immigrants. There are a number of obvious typos that should be corrected. Non-1920 records should have a zero value. There are a number of 1920 records that are unresolved NULLs which should be converted to zeros.

8.72 Industry

- Years available: 1910, 1920
- String, maximum length = 36

This represents a more general accounting of the area or industry in which a person was employed, as opposed to their more-specific job title, captured for these and other census years in the "profession" field. For 1910, this is question 19, which asked for the "General nature of industry, business, or establishment in which this person works, as in /cotton mill, dry goods store, farm,/ etc." Question 27 on the 1920 form uses the same wording.

The answers have not been systematized.

8.73 OCCode

- Years available: 1920
- Integer, three digits

This field captures a three-digit occupational code that was written in the right margin of the 1920 manuscript census schedules after the census was taken. Occasionally the writers would spill this code over into empty columns on the right side of the form.

The codes group specific jobs into occupational categories for compilation and comparison purposes. The complete list of codes can be found on the IPUMS website: https://usa.ipums.org/usa/volii/92occup.shtml

Note that these codes may be helpful for verifying that occupational information was transcribed correctly.

Entries for years other than 1920 are coded as zero. In 1920, there are a number of zeros, and a greater number of unresolved NULLs. It should be ascertained if the zeros in 1920 are as-transcribed. Additionally, there are a small number of "998" (Illegible) codes, and a greater number of "999" (No occupation reported) codes. These too should be determined if they were as-transcribed. Then the unresolved NULLs from 1920 should be converted, likely to 999 (no occupation reported) but possibly to zero.

8.74 Unemployed

- Years available: 1910
- String, but lots of garbage here because a lot of times extra numbers were written in the margins in or near this column when it was otherwise blank.

Surviving documentation suggests that this field should only capture 1910's question 21, "Whether out of work on April 15, 1910," which was a yes/no question; coded by transcribers to a code of U, N, or Y. For 1910, this was done, with some answers featuring other characters that are likely typos.

However an examination of the data shows that entries were also made for the 1900 census in this field. The 1900 entries are not U/N/Y answers (or corresponding typos), but integers representing the number of months the person has been unemployed. Luckily, however, this data for 1900 was also captured in the "MosUnempl" field and the counts of each answer appear

to be the same (with a few minor discrepancies with NULLs/zeros). Thus this information is a candidate for removal in future releases of the dataset.

Also of note is that 1910's question 22 asks for the number of weeks the person was out of work in **1909**. This information is related to that found in the "MosUnempl" field for earlier years, but was not recorded anywhere in the dataset, likely because it is not quite comparable to the earlier data.

See also "MosUnempl"

9 Errata

Corrections to the data and documentation are welcome. Please send corrections or notice of errors to Eric Nystrom (eric.nystrom@rit.edu).

- Most fields have typos and/or faithfully-transcribed errors from the original documents.
- The documentation for each field (above) attempts to provide notice about particularly pernicious issues with each field.
- All fields need to be checked for fields that ought to be NULL (because they are out of scope) but aren't (e.g. are blank).

Smaller stuff:

- Check for "enumerator" of zero in 1880 and after. Found some in Hall"ick" Precinct, Elko Co., 1910; which should be E.D. 15.