

Digitizing the US Social Statistics

by Vasily Rusanov, New York University, February 2020

This is a guide to the coding of 1850, 1860, and 1870 Social Statistics schedules.

- In this guide, I denote the columns that need to be filled by `column_name`. Example excel sheets are provided.
- All fields should be left empty if they are empty in the original. Do not code "0", "None" or "Nothing".
- Most numbers are in \$. Cents should be coded as decimals (eg, the wage of 70 cents should be coded as 0.7). Do not enter the dollar sign or the units, just enter the number (enter 2700.5 instead of \$2700.50). Ignore half cents (for example, \$5555.28 1/2 for State Tax in Daviess County, Missouri in 1860 should be entered as 5555.28).
- Sometimes, a comma is used as decimal separator, eg "\$1.50". This should be entered as "1.5" (dot, not comma).
- Some fields are open-ended, such as `valuedby`. In this case, you can take several values that I tried to describe by **sheriff**, **assessor**, etc, written in bold. When coding, use standard categories (that is, instead of *As assessed by the County Sheriff and Sheriff's assession* use **sheriff**, because **sheriff** is standard). See the descriptions for each variable below.
- Sometimes marshals reported only the total numbers, without breaking the total down by category. If this happens, report the number in the field that corresponds to the first entry in the form. Leave all other fields blank.
- The Census Marshals used the forms as their *notebooks/scratch paper*. Ignore everything that is not a part of the forms. In particular, every Marshal had to write an oath (*"I hereby certify that..."*). Ignore this text.
- Sometimes a range is given for some variables. In that case, enter the average of the range. For example, average crops can be given as *10-20 bushels per acre* instead of one number. In this case, you should code *15* as the average crop.

Reading the source

- Do or do* means "Ditto", "the same". This means the number/type is the same as on the line above.
- A weird symbol that looks like *f* means "ss". For example, "Assessor" is sometimes "AJeFor".
- The symbol "-" is often written to show the end of a number. For example, \$200,000.00-
- Every sheet is written by one person. If you are not sure about one letter or number, try to look at the other letters or numbers on the same sheet.
- There can't be fewer teachers than schools. Common schools often had 1 teacher in each (so-called one-house schools). If there are 48 schools, there can't be fewer than 48 teachers in them.
- Total value of estate is usually equal to the value of real estate + personal estate. This can help in deciphering digits that are hard to read.
- All taxes at that time were property taxes. Taxes were usually a small percentage of the value of total estate. This means that taxes cannot be higher than the value of estate.

1850-1860

ID variables

- `year`: Code **1850** or **1860**
- `filename`: Copy-paste the filename, for example **Cedar_co_IA_1850.jpg** or **Illinois_Adair.pdf**
- `statenam`: Code **Illinois, Wisconsin, Minnesota, Iowa, Missouri, or Arkansas**
- `countynam`: Name of the county, as it appears on top. Cannot be empty.
- `countycode`: Take the code GISJOIN from the standard codes, including the letter G. Codes can be found in `county_codes/List_counties_18XX.csv`. Cannot be empty.
- `sheet`: Sometimes several townships in one county are described individually. Thus, a county is described on multiple images, or "sheets". Use this field if there are multiple sheets (2 or more).
- `sheetname`: Sometimes one county is described on several sheets of paper or only a part of the county is described. the marshals wrote what part of the county is described on top of the page (eg, Township number X). Do not enter "whole" or "whole county".

Valuation of estate

- `realestat`: Real estate
- `perestate`: Personal estate
- `totestate`: Total estate
- `valuedby`: How Valued. Mostly, you should use one of the following:
 - sheriff** ("Sheriff's books", "From Sheriff", etc. should all be coded as **sheriff**)
 - assessor** ("County Assessor", "From Assessor", "Tax Assessor" etc.)
 - owners** ("Owner", "Self" etc.)
 - Other (specify!)
- `trueval`: True valuation. Do not fill if empty.

Seasons and Crops

- To what extent crops are short (usually fraction, 1/4, 1/2, 3/4). Convert to decimals with 3 digits, e.g. 1/3 = 0.333, 7/8 = 0.875 etc. *"Entire"*, *"All"*, *"Almost All"*, *"Failed"* mean 1. "Good" or "Average" means 0 (no crops are lost or "short"). You should also report the average crops in bushels/acre (except when noted) using the `average` columns
 - `wheatshort` and `wheataverage`
 - `cornshort` and `cornaverage`: Include *Indian corn, Corn*.
 - `oatshort` and `oataverage`
 - `potatoesshort` and `potatoaverage`: Include *Potatoes, Irish potatoes*. *Sweet potatoes* should be in `other`!
 - `hayshort` and `hayaverage`: Include *Hay, Grass (Grass)*, *Hungarian Grass*, and *Timothy*. Average is in US tons/acre (1 ton = 2,000 pounds), not bushels. Code the number in tons, do not convert to pounds.
 - `ryeshort` and `ryeaverage`
 - `cottonshort` and `cottonaverage`: Average should be reported in pounds (lbs) per acre. 1 Bale = 480 pounds, convert when necessary.
 - `buckwheatshort` and `buckwheataverage`
 - `othershort` and `otheraverage`: Reserved for crops not listed. Put the name of this other crop in `othershort_name`.
- Sometimes the Census Marshals made a mistake and wrote the total crop (several thousand bushels) instead of the averages. In that case, the information is useless (we don't know the number of acres), so leave the `average` part empty.

Annual taxes

- Leave blank if a particular tax is not reported.
 - `statetax`
 - `countytax`: Include *Co Tax, County Tax*.
 - `schooltax`: Include *School, State School, Town school tax, Ccounty school tax*
 - `schhousetax`: Include *School house/building*.
 - `teachertax`: Include *teaching/teachers' fund*.
 - `districttax`
 - `roadtax`: Include *Highway*. Do not include *Railroad*, put it in `othertax1`.
 - `towntax`: Include *Township tax, Town tax, and also City tax*.
 - `polltax`: Include *poll/voting tax*.
 - `poortax`: Include *lacking or poverty*.
 - `asylumtax`: include *lunatic* here.
 - `othertax1`, with `othertax1_name` and `othertax1_how`: put any tax that is not listed above and write its name. *State Inst or State Interest Fund* (see Miller Co, MO), common in Missouri, means *State Interest*. - `othertax` variables run from 1 to 5.
- How paid. Most taxes were paid in cash. Leave the `_how` field empty if payment is made with cash. If one payment is split, right the majority (eg, "3/4 work, 1/4 cash" is coded as **work**). If in doubt, put cash (which is coded as empty).
 - cash* As described above, leave the `_how` field blank if cash. *Specie, Gold, Silver, Money* are all cash.
 - work or labor* (**work**)
 - county scrips, county orders, township scrips, or warrants* (code all of these as simply **scrips**)
- Sometimes the tax rate is reported instead of the amount collected (for example, in Linn County in Iowa in 1860 and in many counties in Arkansas). The rate is usually given in mills. One mill = 1/1000, or 0.1 percent. In that case, write the word **rateonly** in the `_how` field and calculate the amount manually. To calculate the amount, multiply the Total value of property by the tax rate.

For example, IA_1860_249.jpg, Linn county, Bertram Township. Total estate, valued by Tax Assessor, is \$172887.00. The State tax rate is 1.5 mill = 0.15%, or 0.0015 (1.5/1000). The amount of State tax paid should be = 0.0015 * \$172887.00 = 259.33. Put **rateonly** in `statetax_how` field. For `countytax`, you should put 518.66 (3 mills times \$172887.00), and you should put **rateonly, scrips** in `countytax_how` because county orders were used to pay the tax. See Conway Co, AR, 1860 for an example of a county where *both* the rate and the amount were recorded.

Colleges, Academies, Schools (by kind)

Sometimes the Census Marshals mistakenly listed every school separately, while they only needed to record the *total* number of schools and students in them for each kind of school. Only the totals should be recorded, do *not* record each school individually.

- Kind (give the *total* number of those, sometimes you will have to add up). Do not report male and female schools separately (*Female common school* should be included together with *Male common schools* into *Common schools*)
 - `common`: *Common, Public, District, Town schools, Government schools*. If not specified, assume the school is "common".
 - `priv`: *Private*. *Select school* also means *private school*. In a few cases in Arkansas, some schools written as *Male* and *Female* should be classified as *priv*
 - `high`: *High schools*
 - `semin`: *Seminaries*. However, schools started by denominations (*Evangelical*) should go in `priv`.
 - `acad`: *Academies*. Include *Institutes* in Academies.
 - `colleg`: *Colleges and Universities*.
- # of teachers,
- # of pupils
- Amount realized from endowments
- Raised by taxes
- Received from public funds
- Received from other sources

Libraries

- `library_number_nonpriv`: Total number of libraries, *except* private libraries (*Private, Individual, Personal*). Do not break down by type.
- `library_volumes_nonpriv`: Total number of volumes in all libraries (you need to add up all volumes), *except* private libraries (as defined above). Do not break down by type.
- `library_number_private`: Total number of private libraries (*Private, Individual, Personal*). Do not break down by type.
- `library_volumes_private`: Total number of volumes in private libraries, (*Private, Individual, Personal*).
- Do not code: kind (instead, add up all the volumes together for non-private libraries)

Newspapers and periodicals

Only enter the *total* number and circulation of daily, weekly and monthly newspapers (you will need to add them up)

- `d_newsp`: total circulation of daily newspapers (*How often published* is described as *daily*)
- `d_newsp_circ`: total circulation of daily newspapers
- `w_newsp`: total number of weekly newspapers (*How often published* is described as *weekly*)
- `w_newsp_circ`: total circulation of weekly newspapers
- `m_newsp`: total number of monthly newspapers (*How often published* is described as *monthly*)
- `m_newsp_circ`: total circulation of monthly newspapers
- `oth_newsp`: total number of other newspapers (*How often published* is not any of the above)
- `oth_newsp_circ`: total number of other newspapers
- `oth_newsp_period`: periodicity of other newspapers, in the format [digit] per [period]
 - For example, enter "3 per month" or "2 per week". "3/month", "three per month" are wrong (do not use "/" or words instead of numbers)
- Do not code: Names of newspapers
- Do not code: Character of newspapers (political, misc)

Do not code: Religion

- Do not code: # of churches
- Do not code: Denomination
- Do not code: # each will accommodate
- Do not code: Value of church property

Pauperism

- `pauper_y_nat`: Whole No. of Paupers supported within the last year, Native
- `pauper_y_for`: Whole No. of Paupers supported within the last year, Foreign
- `pauper_jun1_nat`: Whole No. on 1st of June, Native
- `pauper_jun1_for`: Whole No. on 1st of June, Foreign
- `pauper_cost`: Cost of support

Crime

- `crim_y_convict_nat`: Whole No. of criminals convicted within the year, native
- `crim_y_convict_for`: Whole No. of criminals convicted within the year, foreign
- `prison_jun1_nat`: No. in prison on 1st of June, Native
- `prison_jun1_for`: No. in prison on 1st of June, Foreign

Wages

Enter all without \$ or c sign. Use dot as the separator. For example, "\$.50" is coded as **0.5**, and "75 c" is **0.75**.

- `m_wages_farm_board`: Av. monthly wages to a farm-hand with board
- `d_wages_laborer_board`: Av. to a day-laborer with board
- `d_wages_laborer_noboard`: Av. to a day-laborer without board
- `d_wages_carpent_noboard`: Av. day-wages to a carpenter without board
- `w_wages_female_board`: Weekly wages to a female domestic with board
- `w_board`: Price of board to labouring man per week

Quality

- `qualoverall`: rate the quality of the sheet you coded, from 0 to 3.
 - 0** Completely illegible, could not fill most of the fields even though the data is there
 - 1** Legible, but many fields (four or more) are likely to be wrong
 - 2** There was an issue with one, two or maybe three fields
 - 3** No issues while coding this sheet
- `problemfields`: Put all the `fields` that you are unsure about here, separated by a comma. If Quality was **0** or **1**, do not fill anything (so do not enter more than three fields).

1870

ID variables

- `year`: Code **1870**
- `filename`: Copy-paste the filename, for example **Cedar_co_IA_1850.jpg** or **Illinois_Adair.pdf**
- `statenam`: Code **Illinois, Wisconsin, Minnesota, Iowa, Missouri, or Arkansas**
- `countynam`: Name of the county, as it appears on top. Cannot be empty.
- `countycode`: Take the code GISJOIN from the standard codes, including the letter G. Codes can be found in `county_codes/List_counties_18XX.csv`. Cannot be empty.
- `sheet`: Sometimes several townships in one county are described individually. Thus, a county is described on multiple images, or "sheets". Use this field if there are multiple sheets (2 or more).
- `sheetname`: Sometimes one county is described on several sheets of paper or only a part of the county is described. the marshals wrote what part of the county is described on top of the page (eg, Township number X). Do not enter "whole" or "whole county".

Valuation

- `realestat`: Real estate
- `perestate`: Personal estate
- `totestate`: Total estate
- `valuedby`: How Valued. Mostly, you should use one of the following:
 - sheriff** ("Sheriff's books", "From Sheriff", etc. should all be coded as **sheriff**)
 - assessor** ("County Assessor", "From Assessor", "Tax Assessor" etc.)
 - owners** ("Owner", "Self" etc.)
 - Other (specify!)
- `trueval`: True valuation. Do not fill if empty.

Public Debt

- `countydebt_bonds`
- `countydebt_nobonds`
- `towndebt_bonds`
- `towndebt_nobonds`

Taxation

Leave blank if a particular tax is not reported. Sometimes the tax rate is reported instead of the amount collected (in particular, in Arkansas); in that case, write the word **rateonly**.

- `statetax`
- `countytax`
- `towntax`
- `totaltax`
- `typetax`: Write, in words, the types of tax collected. Use comma to separate. Replace the word "*and*" with a comma. For example *School, School house, County expenses, Bridge*. Sometimes the taxes are described separately per level collected. For example, Carroll County in Missouri in 1870 is described as

State: Rev. & Int.
County: Rev. Int. Road, Courthouse & R.R.
Town: Rev. School

The above should be coded as

State: Rev, Int. County: Rev, Int, Road, Courthouse, RR. Town: Rev, School.

Do not use the symbol "&" (replace it with a comma), and only use dots to separate State, County, and Township (for example, *R.R.* should be written as *RR*)

Pauperism

- `pauper_y_nat`: Whole number of Paupers supported during the year, Native
- `pauper_jun1_for`: Whole number of Paupers supported during the year, Foreign
- `pauper_jun1_nat_w`: Whole number on 1st of June, Native white
- `pauper_jun1_nat_b`: Whole number on 1st of June, Native black
- `pauper_jun1_for`: Whole number on 1st of June, Foreign
- `pauper_cost`: Average cost of support

Crime

- `crim_y_convict_nat`: Whole No. of criminals convicted within the year, native
- `crim_y_convict_for`: Whole No. of criminals convicted within the year, foreign
- `prison_jun1_nat_w`: No. in prison on 1st of June, Native white
- `prison_jun1_nat_b`: No. in prison on 1st of June, Native black
- `prison_jun1_for`: No. in prison on 1st of June, Foreign

Libraries

- `library_number_nonpriv`: Total number of libraries, *except* private libraries (*Private*). Do not break down by type.
- `library_volumes_nonpriv`: Total number of volumes in all libraries (you need to add up all volumes), *except* private libraries (*Private*). Do not break down by type.
- `library_number_private`: Total number of private libraries, *except* private libraries (*Private*). Do not break down by type.
- `library_volumes_private`: Total number of volumes in private libraries, (*Private*).
- Do not code: kind (instead, add up all the volumes together for non-private libraries)

Newspapers and periodicals

Only enter the *total* number and circulation of daily, weekly and monthly newspapers (you will need to add them up)

- `d_newsp`: Total circulation of daily newspapers (*How often published* is described as *daily*)
- `d_newsp_circ`: Total circulation of daily newspapers
- `w_newsp`: total Number of weekly newspapers (*How often published* is described as *weekly*)
- `w_newsp_circ`: Total circulation of weekly newspapers
- `m_newsp`: total Number of monthly newspapers (*How often published* is described as *monthly*)
- `m_newsp_circ`: Total circulation of weekly newspapers
- `oth_newsp`: total number of other newspapers (*How often published* is not any of the above)
- `oth_newsp_circ`: total number of other newspapers
- `oth_newsp_period`: periodicity of other newspapers, in the format [digit] per [period]
 - For example, enter "3 per month" or "2 per week". "3/month", "three per month" are wrong (do not use "/" or words instead of numbers)
- Do not code: Names of newspapers
- Do not code: Character of newspapers (political, misc)

Wages

Enter all without \$ or c sign. Use dot as the separator. For example, "\$.50" is coded as **0.5**, and "75 c" is **0.75**.

- `m_wages_farm_board`: Av. monthly wages to a farm-hand with board
- `d_wages_laborer_board`: Av. to a day-laborer without board
- `d_wages_laborer_noboard`: Av. to a day-laborer with board
- `d_wages_carpent_noboard`: Av. day-wages to a carpenter without board
- `w_wages_female_board`: Weekly wages to a female domestic with board
- `w_board`: Price of board to laboring man per week

Schools

- Kind (give the *total* number of those, sometimes you will have to add up because the Marshals listed them individually).
 - `normal`: number of *Normal Schools*
 - `high`: number of *High schools*
 - `graded`: number of *Graded common schools*
 - `ungraded`: number of *Ungraded common schools*
 - `privday`: number of *Private schools: Day*
 - `boarding`: number of *Private schools: Boarding*
 - `colleg`: number of *Colleges and Universities*. Include both colleges and universities here, adding everything for them together.
 - `acad`: number of academies
 - `other1` - `others`: Use those fields for all other types of schools. There are no other named columns because it is very rare to have *Technological schools* or *Schools of mining* etc. Code their type in `other1-5_type` exactly as it appears in the source, for example "Law" or "Schools of art and music".
 - For each type (above), also code:
 - `_teachers_male`, `_teachers_female`
 - `_pupils_male`, `_pupils_female`
 - `_endowment`, `_taxes`, `_pubfunds`, `_otherfunds`: This corresponds to columns *Income from endowment*, *Income Raised by Taxation*, *Income Received from Public Funds*, *Income From other source, including Tuition*
 - `pub_total`: Sometimes, only the total income is reported for all public schools, without breaking down by type of school (for example, in MO_1870_Buchanan.jpg). In this case, use the columns `pub_total_endowment`, `pub_total_tax`, `pub_total_pubfunds`, `pub_total_otherfunds`. Note that those columns are only intended to describe the totals for rows *Normal, High, Grammar, Graded common, Graded common*.
 - Sometimes, the schedules contain notes written that break down the funds proportionately between schools. Usually, these notes are in pencil and different handwriting (see Wisconsin and Missouri). Ignore them, since they do not add any information.

Do not code: Religion

- Do not code: # of churches
- Do not code: Denomination
- Do not code: # each will accommodate
- Do not code: Value of church property

Quality

- `qualoverall`: rate the quality of the sheet you coded, from 0 to 3.
 - 0** Completely illegible, could not fill most of the fields even though the data is there
 - 1** Legible, but many fields (four or more) are likely to be wrong
 - 2** There was an issue with one, two or maybe three fields
 - 3** No issues while coding this sheet
- `problemfields`: put all the `fields` that you are unsure about here, separated by comma. If Quality was **0** or **1**, do not fill anything (so do not enter more than three fields).