Historical Population Estimates for 2010 U.S. States, Counties and Metro/Micro Areas, 1790-2010

DOCUMENTATION

Creator: Jonathan P. Schroeder
Minnesota Population Center
University of Minnesota, Twin Cities
jps@umn.edu
Creation dates: 2011-2016

Abstract

The core of this data set is a series of historical population estimates for each U.S. decennial census year, 1790-2010, for all U.S. counties and county equivalents (excluding Puerto Rico and other territories), using spatially fixed 2010 county definitions. The data set also includes additional statistics derived from the county data, including historical population estimates for 2010 states and for metropolitan and micropolitan statistical areas (i.e., core-based statistical areas, or CBSAs) following both 2009 and 2013 CBSA definitions. Errors may be substantial in many cases, so users should not give great weight to any single estimate in analyses. However, for broad studies covering large regions or timespans, errors may generally have a small impact, and using these spatially standardized data enables a broad range of visualization and analytical approaches that would not otherwise be applicable. The data were originally generated for use in the production of an atlas of U.S. population distribution and change, which has not been published. The author has also used the data in the production of conference posters.

Disclaimer

Data users should recognize that the estimates provided here are necessarily "rough" in many areas, with a potentially high degree of error, especially for earlier censuses and in older counties bordering "urban counties" that expanded over time—e.g., Baltimore County around Baltimore City, and several Virginia counties that border Virginia's independent cities. Although some additional steps were taken to reduce the severity of errors in such counties, mapping estimated population densities from early censuses reveals that the estimates near expanding urban counties are still markedly lower than other nearby counties', suggesting systematic *under*-estimation of the populations in such counties.

Use Examples

The author used the county and CBSA data to produce two conference posters, accessible through these links:

- U.S. County Population Trends, 1790-2010: New Perspectives Using Spatially Standardized Data. Association of American Geographers. New York, February 24-28, 2012.
- U.S. Metropolitan Populations, 1790-2010. North American Cartographic Information Society. Greenville, SC, October 9-12, 2013.

Data Dictionary

The data are provided in four spreadsheet files stored in Microsoft Excel 2013 format:

File/Worksheet/Field(s)	Description
state2010_hist_pops.xlsx	Historical populations of 2010 states, 1790-2010, with related derived statistics
s2010_hist_pops	State population estimates, 1790-2010
ST	State FIPS code
State	State name
Region	Census region
epop1790 – epop2000	Population estimates, 1790-2000
pop2010	2010 population
change_rates	Rates of population change in each consecutive decade, 1790-2010
ST	State FIPS code
State	State name
1790s – 2000s	Rate of population change in each identified decade
nation_totals	Total national population, 1790-2010
pop1790 – pop2010	U.S. national population in each census year, 1790-2010
national_shares	Each state's share of the nation's population in each census year, 1790-2010
…	Geographic identifiers, as above
share1790 – share2010	Share of nation's population (%)
max_scaled_shares	Each state's share of the nation's population in each census year, 1790-2010, *relative to* the state's maximum historical share
…	Geographic identifiers, as above
maxshare	State's maximum share of nation's population, 1790-2010 (%)
mss1790 – mss2010	Max-scaled share, 1790-2010 (range from 0 to 1). Values of 1 are highlighted in red to indicate year when maximum occurred. Values of exactly 0 are shown in gray to distinguish them from very small non-zero numbers that round to 0.
county2010_hist_pops.xlsx	Historical populations of 2010 counties, 1790-2010, with related derived statistics
c2010_hist_pops	County population estimates, 1790-2010
GISJOIN	NHGIS county ID
GEOID10	State + county FIPS code
STATE	State name
COUNTY	County name
epop1790 – epop2000	Population estimates, 1790-2000
pop2010	2010 population
densities	County population densities
…	Geographic identifiers, as above
land_sqmi	Land area in square miles
dens1790 – dens2000	Population density estimates, persons / sq. mile, land area, 1790-2010
20yr_pop_changes	Population changes, by county, for each consecutive 20-year period, 1790-2010
…	Geographic identifiers, as above
c[year1]_[year2]	Population change between year1 and year2

File/Worksheet/Field(s)	Description
mean_scaled_pops	Historical county populations *relative to* each county's mean across all census years, 1790-2010
…	Geographic identifiers, as above
avgpop	County's mean population across all census years, 1790-2010
msp1790 – msp2010	Mean-scaled population, 1790-2010. Values of exactly 0 are shown in gray to distinguish them from very small non-zero numbers that round to 0.
year_threshold_passed	For each county, the years that the county's population density first surpassed specified thresholds
…	Geographic identifiers, as above
year_d#	The census year when the county's population density (land area only) first surpassed # persons / sq. mile. Values of 999999 indicate that the county density never surpassed the threshold in any census year.
national_shares	Each county's share of the nation's population in each census year, 1790-2010. Values of exactly 0 are shown in gray to distinguish them from very small non-zero numbers that round to 0.
…	Geographic identifiers, as above
share1790 – share2010	Share of nation's population (%)
maxes	Historical maxima in county populations and shares, and the years they were attained
…	Geographic identifiers, as above
maxpop	County's maximum population estimate, 1790-2010
yrmaxpop	Year of county's maximum population, 1790-2010
yrhalfmax	First year that county's population exceeded half of its historical maximum
maxshare	County's maximum share of nation's population, 1790-2010 (%)
yrmaxshare	Year of county's maximum share of nation's population, 1790-2010
coef_var	Coefficients of variation in historical county populations
…	Geographic identifiers, as above
mpop	Mean population, census years 1790-2010
sdpop	Standard deviation in populations, census years 1790-2010
cvpop	Coefficient of variation in populations, census years 1790-2010
share_coef_var	Coefficients of variation in historical county shares of national population
…	Geographic identifiers, as above
mshare	Mean share of nation's population, census years 1790-2010
sdshare	Standard deviation in share of nation's population, census years 1790-2010
cvshare	Coefficient of variation in share of nation's population, census years 1790-2010
max_scaled_shares	Each county's share of the nation's population in each census year, 1790-2010, *relative to* the county's maximum historical share
…	Geographic identifiers, as above
maxshare	County's maximum share of nation's population, 1790-2010 (%)

File/Worksheet/Field(s)	Description
mss1790 – mss2010	Max-scaled share, 1790-2010 (range from 0 to 1). Values of 1 are highlighted in red to indicate year when maximum occurred. Values of exactly 0 are shown in gray to distinguish them from very small non-zero numbers that round to 0.
nation_totals	Total national population, 1790-2010 (sums of all county populations)
pop1790 – pop2010	U.S. national population in each census year, 1790-2010
cbsa2009_hist_pops.xlsx	Historical populations of 2009 core-based statistical areas, 1790-2010, with related derived statistics
cbsa2009_hist_pops	CBSA population estimates, 1790-2010
CBSA	2009 CBSA code
CBSA_NAME	CBSA name
MET_MIC	Metropolitan/micropolitan status (1 = metropolitan, 2 = micropolitan)
epop1790 – epop2000	Population estimates, 1790-2000
pop2010	2010 population
national_shares	Each CBSA's share of the nation's population in each census year, 1790-2010
…	Geographic identifiers, as above
share1790 – share2010	Share of nation's population (%)
summaries	Summary statistics for each CBSA's census-year populations, 1790-2010
…	Geographic identifiers, as above
avgpop	CBSA's mean population estimate, 1790-2010
maxpop	CBSA's maximum population estimate, 1790-2010
yrmaxpop	Year of CBSA's maximum population, 1790-2010
avgshare	CBSA's mean share of the nation's population, 1790-2010 (%)
maxshare	CBSA's maximum share of nation's population, 1790-2010 (%)
yrmaxshare	Year of county's maximum share of nation's population, 1790-2010
nation_totals	Total national population, 1790-2010
pop1790 – pop2010	U.S. national population in each census year, 1790-2010
cbsa2013_hist_pops.xlsx	Historical populations of 2013 core-based statistical areas, 1790-2010
cbsa2013_hist_pops	CBSA population estimates, 1790-2010
CBSA	2013 CBSA code
CBSA_NAME	CBSA name
MET_MIC	Metropolitan/micropolitan status (1 = metropolitan, 2 = micropolitan)
epop1790 – epop2000	Population estimates, 1790-2000
pop2010	2010 population

Derivation

There were two main stages in processing the historical county population estimates:

1. Compute historical populations for 2000 counties (completed before 2010 data were available)
2. Re-allocate historical populations to 2010 counties

(All other estimates and statistics—e.g., CBSA estimates, state estimates, national population shares, etc.—were derived from these estimates by generally straightforward means not documented here.)

Part 1: Historical populations for 2000 counties

Source datasets:

- County shapefiles for 1790-2000 from the National Historical Geographic Information System (NHGIS—http://nhgis.org)
- Water polygons from U.S. Census 2000 TIGER/Line Files
- 1990 block group polygons from NHGIS
- 1990 block group populations from NHGIS
- Historical county population totals 1790-2000 from NHGIS
- Alternative source of historical county populations: Forstall, R. L. 1996. *Population of States and Counties of the United States: 1790-1990*. U.S. Government Printing Office, Washington, DC.

General steps:

- Check historical county populations for unusual zero counts and apply corrections.
- Create "county atoms"—the union of all counties 1790-2000.
- Estimate atoms' 1990 population through area weighting of 1990 block group data.
- Estimate atoms' pre-1990 populations using "cascading density weighting" (Schroeder 2009) with a few adjustments
- Sum the historical atom population estimates for each 2000 county

Detailed steps:

- Check historical county populations for unusual zero counts and apply corrections.
    - 1820: DC "county" population = 0
        - Assign correct population from state-level counts
    - 1850: Data were lost for 3 counties in CA (Contra Costa, Santa Clara, San Francisco)
        1. Acquire 1852 Census of California county population counts from www.learncalifornia.org/doc.asp?id=441 (which excludes "foreign residents" counted separately).
        2. To compute 1860 population of 1850 San Francisco County (which was split in two), sum 1860 populations of San Francisco and San Mateo Counties
        3. Use area weighting to allocate 1860 population of Alameda Co (a new county) to 1850 areas of Contra Costa and Santa Clara
        4. Assume no 1850 population lived in the small portion of 1850 Contra Costa that was added to 1860 San Joaquin Co.
        5. Using the 1852 and 1860 county population counts, apply linear extrapolation to estimate the 1850 counts:
            - pop50 = pop52 – 0.25 * (pop60 – pop52)
        6. Round extrapolated figures to 2 significant digits to produce final estimates
    - 1870: Baltimore City population = 0 (misallocated to Baltimore County)
        - Use proper NHGIS variable & verify against 1870 city population data
    - 1870: Numerous counties in MI have boundaries defined but no population. A 1900 report indicates that several counties' populations were all credited to Marquette County in 1870, but 1870 and 1880 reports indicate otherwise.
        - Where possible, obtain 1870 counts from Forstall (1996)…
            - Allocate 799 of Marquette Co's population to Schoolcraft Co.
        - Other counties with no data are allocated no population.
    - 1880-1940: Data compiled by Census staff omitted AK & HI territorial counties

        - I included counties for AK starting in 1880 and HI starting in 1900, corresponding with the first U.S. census counts of those areas
    - 1880: NHGIS shapefile has polygon for newly formed Dickenson County, VA, but that county was not recognized in census, meaning the population for its area was reported for Wise and Buchanan Counties (out of which Dickenson was carved).
        - Change 1880 shapefile to match 1870 boundaries in this area.
    - 1910 & 1920: Counts for Williamsburg City, VA, were 0 (misallocated to James City)
        - Use Forstall's counts for James and Williamsburg Cities
    - 1910: NHGIS shapefile has polygon for newly formed Greenlee County, AZ, but its population was apparently reported with Graham County, AZ, in 1910
        - Change 1910 shapefile to match 1900 boundaries in this area.
- Create "county atoms"—the spatial union of all counties 1790-2000.
- Estimate atoms' 1990 population through areal weighting of 1990 block group data.
    - Intersect the county atoms with 1990 block groups and erase water polygons so area weighting is based only on land area.
        - To simplify processing, erase only "substantial" water areas (> 10 acres)
    - Problem: Because the NHGIS 1990 block group boundaries do not consistently align with 1990 county boundaries, simple area weighting results in several allocations of 1990 block group populations to areas outside of the proper 1990 county.
        - Solution: Allocate 1990 county populations among atoms in proportion to the 1990 atom population estimates from the block groups
- Estimate atoms' pre-1990 populations using cascading density weight (CDW; Schroeder 2009) with some small adjustments
    - Use the 1990 atom population estimates to guide allocation of 1980 tract data among atoms; use the 1980 estimates to guide 1970 allocation; and so on back to 1790.
    - Problem: In a few cases, the later year's estimates for a county's population sum to 0, making it impossible for the CDW process to use those estimates as a guide for allocation (a divide-by-zero problem).
        - Solution: Apply area weighting in these cases
    - Problem: When "urban counties" (e.g., Virginia independent cities, Baltimore City, Denver County, etc.) expand their boundaries into surrounding counties, CDW tends to perform poorly. The population change rate in an area of expansion is typically much higher than in other parts of the original county, but CDW assumes a uniform change rate. Thus, it typically overestimates the population that originally lived in the area of expansion and underestimates the population in the rest of the county. These errors propagate backward and produce noticeable anomalies on historical density maps, appearing as rings of very low density around 2010's urban counties.
        - Solution: Adjust the estimates in "urban expansion atoms"…
        - First identify cases of urban county expansion, atoms where the county ID changes between censuses, the atom's estimated density in the later year is "urban" (> 500 persons / sq. mi.), and the estimated density of the county the atom had been in is less than half the atom's estimated density in the later year.
        - Instead of using CDW's assumption of a uniform change rate throughout the original county, assume that in the urban expansion atoms, the change rate is similar to the rate in the following decade.

        - Specifically, assume the proportional change rate is equal to the square root of the subsequent decade's ("half" the proportion): pop[year2] / pop[year1] <= sqrt(pop[year3] / pop[year2]).
    - Problem: In some cases, this assumption would estimate the density in an urban expansion atom to be about the same as or less than the outlying county's.
        - (We do expect areas of urban county expansion to have a higher density than outlying areas, just not as high as CDW predicts.)
        - Solution: Cap the assumed change rate so that the estimated population does not go beyond the midpoint between the CDW and area weighting estimates (i.e., the midpoint between assuming a uniform change rate and a uniform density in the original county)
    - Author's note: the effects of these adjustments are still not satisfactory.
        - The final estimates for counties surrounding urban counties generally remain too low, judged by comparing their densities to the densities of other nearby counties.
        - More work could be done to improve the interpolation in these areas!
- Sum the historical atom population estimates for each 2000 county

Part 2: Historical populations for 2010 counties

Source datasets:

- 2010 county populations from 2010 PL 94-171 Redistricting Data (via NHGIS)
- 2010 county boundaries from 2010 TIGER/Line (via NHGIS)
- 2000 populations for 2010 counties: Census's 2009 Population Estimates 2000 estimates base
- Pre-2000 population estimates for 2000 counties from Part 1
- 1790-1990 county intersections with 1990 block groups from Part 1 (with water erased)

Steps:

- Combine 2000 and 2010 populations for 2010 counties into a new file
- For all 2010 counties that are not "substantially" different from 2000 counties, transfer pre-2000 population estimates directly from Part 1
    - Problem: Part 1 processing used original NHGIS boundary data, which were based on 2000 TIGER/Line files. 2010 boundary data, based on much-improved 2010 TIGER/Line files, do not align well with the original NHGIS boundaries. Therefore, it would be problematic to complete estimates for 2010 counties by "extending CDW forward" universally because this would require overlay of mismatched boundary data and result in numerous sliver polygons, causing misallocations of data.
        - Solution: Because there are few significant boundary changes between 2000 and 2010, limit the "CDW extension" to areas where there was a significant boundary change and elsewhere just allocate data from each 2000 county to the corresponding 2010 county.
    - To determine qualifying "substantial" changes, use U.S. Census Bureau's list of "Substantial Changes to Counties and County Equivalent Entities" (www.census.gov/geo/www/tiger/ctychng.html), which lists "*changes affecting at least an estimated population of 200 plus with additional changes of at least one square mile where no estimated population was provided and research indicated that the affected population may have been 200 people or more or "large" annexations of unpopulated territory (10 square miles or more.)*"

- For 2010 counties that *are* substantially different from 2000 counties, interpolate historical population estimates from the Part 1 county atoms, weighting by 1990 block group populations
    - Using the county intersections with 1990 block groups (BG) from Part 1, estimate the historical populations of each county-BG atom by assuming that the proportion of each historical county's population in each atom is the same as the estimated 1990 proportion from Part 1.
    - Intersect Part 1 county-BG atoms with 2010 counties (mixing 2000 TIGER-based and 2010 TIGER-based data, resulting in some erroneous slivers)
    - Apply area weighting of 1990 block group populations onto county-BG atoms
    - Eliminate slivers between counties that were not involved in "substantial" changes
    - As in Part 1, because the 1990 block group boundaries do not align properly with 2010 county boundaries, allocate 1990 county populations among atoms in proportion to the 1990 atom population estimates from the block groups.
        - (Don't estimate 1990 populations using misaligned block group data alone. Just use the estimates from block group data as a guide.)
    - Allocate pre-1990 populations using the same proportions as with the 1990 data.
    - Sum historical populations by 2010 county and combine with the 2000 & 2010 counts.

Funding

The creation of these files occurred at intervals over several years while working on several projects. Initial work was begun under a contract with the U.S. Census Bureau. Most subsequent work was completed with support from grants supporting NHGIS (National Institutes of Health [NICHD RO1HD057929], National Science Foundation [SES-1324875]) and the Minnesota Population Center (National Institutes of Health [NICHD R24HD041023]).

References

Forstall, R. L. 1996. *Population of States and Counties of the United States: 1790-1990*. U.S. Government Printing Office, Washington, DC.

Schroeder, J. P. 2009. *Visualizing Patterns in U.S. Urban Population Trends.* Doctoral dissertation, University of Minnesota.