



An Overview of IPUMS NHGIS

Geographic U.S. Census Data from 1790 to 2020

May 19, 2022

Questions & Answers

The following are questions received during the webinar and their answers. For additional questions or clarifications, contact IPUMS User Support at ipums@umn.edu.

Technical

- **Do you have other IPUMS NHGIS webinars or resources?**
Our [user guide page](#) includes training resources as well as a previous webinar that offers a similar overview, but includes a bit more detail on the time series/standardized data.
- **Can you use IPUMS NHGIS data in Stata?**
Yes, if you download our data files in the fixed width format, you will get a .do file to load the data directly into Stata.
- **Is there an option to obtain percentages in the tables?**
We do not offer that option in NHGIS.

Data Availability

- **Do you offer maps/data/shape files of redlined areas throughout each state?**
We do not. The [Mapping Inequality project](#) at the University of Richmond does provide shapefiles of redlined areas. Additionally, you may want to look at [Wenfei Xu's website](#) where she integrated the redlining shapefiles with NHGIS demographic data.
- **Do you anticipate that you will have more recent years of precipitation and temperature data and at tract and block group levels?**
We would like to extend the land use data to include the 2016 NLCD data, but do not currently have funding for this work. We are considering updating the precipitation and temperature data for census tracts as part of a forthcoming grant proposal.
- **What public health data are available from IPUMS NHGIS? For example, life expectancy data at the county-level.**

We have annual county and state-level data on births, death, stillbirths, infant deaths, and fetal deaths by place of occurrence and place of residence from 1922-1967 and for 1918. We also have county data on births and deaths from 1850, and state data on deaths for 1860, 1870, and 1880.

- **Do you have historical data on housing starts and demolitions/losses?**
NHGIS does not. I believe the Census data does publish information on this topic, but we do not include that in NHGIS.

- **What data are included in the environmental summaries?**
We currently summarize land cover from the National Land Cover Database, and we have precipitation and temperature from the PRISM and aggregate these items over various geographic units (e.g., counties, tracts).

- **Within years, do the data tables for lower-level geographies (blocks, tracts, for example) include identifiers for the higher-level geographies (counties, states, etc.)?**

Tables from recent decennial censuses include identifiers for most encompassing areas, but tables from the American Community Survey and older censuses include only a limited set of identifiers. This is due to the design of the source files, not something NHGIS has altered. If you download data tables at the census tract level (for example) from any source, the downloaded file will include columns with the state FIPS code, state name, county FIPS code, county name. If you download tract data from a recent decennial census, the file will also include codes for other encompassing areas (metropolitan areas, regions, divisions, etc.)

Standardization

- **When selecting multiple years of data, what data vintages are used for the geographic identifiers? Are they pre-harmonized?**

If you select source tables (e.g., tables from the 2010 decennial census, tables from the 1990 census), you will get the geographic identifiers that are specific to that census. For the time series tables, we use 2010 geographic identifiers for the “geographically standardized tables”. In other words, we standardize the data to 2010 census units. For the nominally standardized tables, we include the geographic identifiers for each year (e.g., 1970, 1980, 1990, 2000, 2010) in cases where they vary across time, and we include a single integrated code that usually corresponds to the unit’s most recent identifier.

- **Are time-series tables adjusted to a certain vintage of geography (e.g., all in 2000 geography) or do you have to adjust for that on your own?**

There are [two types of time series tables](#): nominally integrated (which do not standardized spatial extents) and geographically standardized (which do standardize the data to consistent spatial extents). At this time, the standardized tables are all standardized to 2010 geography. For nominally integrated TSTs, you would need to adjust on you own if you wanted to ensure you were measuring a consistent geographic footprint. There may, however, be cases where you would like to compare a conceptual

boundary rather than a consistent geographic footprint, such as for comparing city characteristics according to its changing legal definition across time.

- **Do you need to use weights with the crosswalks?**

Yes, [weights in crosswalks](#) indicate the proportion of each source zone's characteristics that should be allocated to each target zone. Note that these are very different from the sample weights that appear in microdata. There are no sample weights in NHGIS data tables or crosswalks. That weighting has already been applied.

- **Is it possible to customize the reference year of the geography when requesting to download time series data? For example, instead of using the 2010 census geography, can I request the data to be standardized according to the 2000 census geography?**

We don't support the ability to customize the reference year of the geography. We definitely understand the desire to do this, but preparing crosswalks and time series for each vintage of geography takes substantial effort. We first prioritized 2010 geography when it was the most recent vintage. We plan to add time series standardized to 2020 geography some time in the future, but we don't yet have time budgeted to standardize to earlier geographies..

- **Does IPUMS NHGIS plan to extend geographically standardized data further back in time than 1990?**

Yes, but there are no definitive plans for the timing of this. We are considering first wrapping up our work to standardize the 1980 block data to facilitate higher quality estimates of these earlier years.

- **Do you have any suggestions for harmonizing counties over time?**

For a separate project, Jonathan Schroeder has created a set of historical population estimates for 2010 counties going back to 1790. The [public repository](#) for those data includes documentation that explains how I standardized to 2010 counties. Similar methods could be applied for other county data. Another resource called [HISDAC-US](#) could help with modeling historical population distributions within counties.