

Annual Survey of Governments Finance Data

This archive contains data on the finances of U.S. municipal governments (counties and smaller units) from 1970 to 1999 and on the finances of state and federal governments in 1972 and from 1977 to 2000.

Genevieve Pham-Kanter and Sam Schulhofer-Wohl, research assistants to Michael Greenstone, compiled the archive in 2002 and 2003 from data provided by the U.S. Census Bureau. At the Census Bureau, Steve Poyta, stephen.m.poyta@census.gov, provided the data files.

For questions about the raw data, contact the Census Bureau. The Census Bureau's contact person for this data as of July 2003 is John Curry, john.l.curry@census.gov. The phone number for data inquiries at the Census Bureau is (800) 242-2184.

For questions about how the archive was compiled, contact Michael Greenstone, mgreenst@mit.edu, or Sam Schulhofer-Wohl, sschulh1@uchicago.edu.

1. Files

The files in this archive are:

ASG_Documentation.pdf, *ASG_Documentation.txt* – This file, in PDF and ASCII formats.

County_City/Data/local_asg.sas7bdat – County, city, village, town, school district and special function district finance data from the annual census of governments and annual survey of governments from 1970 to 1999. There are 1,181,319 observations on 540 variables. The file is in SAS 7 compressed format. Source: Compiled from U.S. Census Bureau data.

County_City/Data/local_asg_contents.txt – Output of SAS proc contents command for the local government data file, in text format.

County_City/Doc/Government_Finance_Manual.pdf – The Census Bureau's "Government Finance and Employment Classification Manual," which defines virtually all of the variables included in the data. (Definitions of variables not included in the manual appear below.) Source: <http://www.census.gov/govs/class/classfull.pdf>.

County_City/Doc/Sample_History.pdf – Documentation on the sample design for the municipal government data in survey years (those not ending in 2 or 7). Source: <http://www.census.gov/govs/sample/samplehistory.pdf>.

County_City/Doc/User_Guide.xls – A user guide to the municipal finance data. Source: U.S. Census Bureau.

County_City/FIPS/FIPS_Changes.xls, FIPS_Changes.pdf – A table of changes in FIPS county codes, in Excel and PDF formats. Updates may be listed at <http://www.census.gov/geo/www/tiger/ctychng.html>, which was the source for the table.

County_City/FIPS/nlookupall.sas7bdat – A lookup table of Census Bureau state and county codes and the corresponding FIPS state and county codes for each year from 1970 to 1999. The file is in SAS 7 format.

County_City/FIPS/Official_FIPS.txt – The Census Bureau’s official pairings of FIPS and Census state and county codes for 2002.

State_Fed/Data/State_Fed_Fin.dta – State and federal government finance data for 1972 and 1977 to 2000. There are 1,300 observations (one for each state, the District of Columbia and the federal government in each of 25 years) on 1,371 variables. The file is in Stata 7 format. Source: U.S. Census Bureau.

State_Fed/Data/State_Fed_Fin.cdb – Output of Stata codebook command for the state and federal government data file, in text format.

State_Fed/Doc/EstGuide.xls, EstGuide.pdf – A user guide to the state and federal finance data, in Excel and PDF formats. Source: U.S. Census Bureau.

2. Data changes

We produced the state and federal data by merging eight separate files provided by the Census Bureau. These files are extracts from the Annual Survey of State and Local Government Finances and Census of Governments. The combined file includes all observations and variables in the original; we did not generate additional variables or modify existing ones. The Census Bureau documentation (*State_Fed/Data/EstGuide.xls, State_Fed/Data/EstGuide.pdf*) contains a column (“Access Table”) indicating which of the eight files a variable came from.

We produced the municipal data file by merging annual files provided by the Census Bureau. These files also are extracts from the Annual Survey of State and Local Government Finances and Census of Governments. We generated several additional variables. These serve three purposes: identifying the FIPS code of the county in which the government is located, identifying government units for which data was collected continuously, and separating information on school enrollment and local population which the Census Bureau coded in the same variable. In addition, we recoded the *SURVEYYR* variable with four digits instead of two.

2-1. FIPS codes

The variable *FIPS_STC* represents the most accurate value we could identify for the FIPS state and county code for each observation, based on Census Bureau state and county codes and government unit names in the raw data.

FIPS_STC is a numeric variable that concatenates the FIPS state code (*FIPS_STn*) and the FIPS county code (*FIPS_Cn*) for each observation. *FIPS_STC* is either a four-digit number or a five-digit number. The first digit (in a four-digit number) or first two digits (in a five-digit number) are the FIPS state code, and the final three digits are the FIPS county code.

The variables *STATEcx* and *COUNTYcx* contain the most accurate values we could identify for the Census Bureau state and county codes for each observation.

The variable *IDcx* is a unique identifying number for each government unit. It is constant throughout the sample even if the name of the government unit changed, so long as the basic function, area served and other characteristics of the unit did not change. The Census Bureau assigned the identifying numbers and determined when they should change.

The following variables come from the raw data, may be incorrect and should generally be ignored. *FIPS_ST* (FIPS state code), *STATE* (Census Bureau state code), *COUNTY* (Census Bureau county code), and *ID* (unique identifying code).

The variable *FIPS_err* has a value of 1 if any of *FIPS_ST*, *STATE* or *COUNTY* is incorrect, or if the given combination of *STATE* and *COUNTY* is logically impossible because the given county did not exist in the year of the observation. Otherwise, *FIPS_err* has a value of 0.

The variable *FIPS_predate* has an empty value unless the county identified by *STATE* and *COUNTY* came into existence after the year of the observation. In this case, *FIPS_predate* is the two-digit year in which the county came into existence, and *FIPS_STC* is the FIPS code that county received when it came into existence. Outside Alaska, this affects only La Paz County, Arizona, and all of the observations involve government units within the current boundaries of that county.

The variable *FIPS_postdate* has an empty value unless the county identified by *STATE* and *COUNTY* ceased to exist before the year of the observation. In this case, *FIPS_postdate* is the two-digit year in which the county ceased to exist, and *FIPS_STC* is the FIPS code that county had when it ceased to exist. This affects only Nansemond County, Virginia, which was combined into the independent city of Suffolk in 1972 but has observations for the county government in the dataset through 1976.

The FIPS coding system has changed over time, as have county boundaries. *FIPS_STC* reflects the FIPS state and county code for the county where the government unit was located in the year of the observation, unless:

- *FIPS_predate* or *FIPS_postdate* does not have an empty value.
- The observation is in Alaska. The frequency of changes in the FIPS coding system in Alaska prevented us from accurately identifying correct FIPS codes for each government unit.
- The observation is in Charlotte County or Charles City County, Virginia. Until 1979, the FIPS standard gave code 51039 to Charlotte and code 51037 to Charles City. Since that year, Charlotte County has had code 51037 and Charles City has code 51036. For all years, we coded these two counties according to the latter standard.
- The observation involves South Boston, Va. This city had its own FIPS code until 1995, when it was merged into another FIPS code; the Census Bureau has recoded the past data.
- The observation involves the cities of Manassas, Manassas Park or Poquoson, Va. These cities were part of larger counties until 1975, when they became independent and received their own FIPS codes. The Census Bureau has recoded the past data.

See the appendix for details on how we identified the FIPS codes.

2-2. Continuous data

All government units are observed in years ending in the digits 2 or 7. In other years, a non-random sample that emphasizes large governments is observed. The variable *consist_obs* has a value of 1 for observations from government units that were observed in all years from 1970 to 1999. It is zero otherwise. *County_City/Doc/Sample_History.pdf* describes the sampling method.

2-3. Population and enrollment

The variable *POP* in the data from the Census Bureau represents student enrollment for school districts and population for all other government units. For convenience, we generated *POPGEOG*, which equals *POP* for non-school units and is zero otherwise, and *ENROL*, which equals *POP* for school units and is zero otherwise.

3. Known data issues

3-1. Coding of missing data

When a government unit leaves an item blank on the survey form, the Census Bureau assumes that item does not apply to the government unit and records a value of zero. Therefore, the data do not distinguish between actual responses of zero and non-responses.

3-2. Aggregate variables

In many cases, aggregate variables in the data (such as “Total Expenditure”) were not reported directly by the government units but instead

were computed by the Census Bureau as the sum of several other reported variables. We do not know which variables were actually reported and which were computed by the Census Bureau. To determine which variables were actually reported, users may wish to review the Census Bureau's survey forms or inquire with the Census Bureau. As of July 2003, survey forms for the 2001 and 2002 survey and census of governments were available on the Web at <http://www.census.gov/govs/www/surveyforms.html>. These forms may differ from those used to collect the data in this archive, which covers 1970 to 1999.

3-3. Rounding

Rounding errors appear in some aggregated variables from 1970 to 1976. This is because, before 1977, values of variables obtained directly from surveys were recorded in whole dollars, but aggregates of these variables were rounded to thousands of dollars. Beginning in 1977, all variables were recorded in thousands of dollars rather than whole dollars.

3-4. Boundary changes

County boundaries change over time. To construct time series or panel datasets from this data, one must account for the boundary and coding changes described in *County_City/FIPS/FIPS Changes.xls*. The best way to connect observations will vary depending on the intended use of the data. For example, if one is aggregating government expenditures and revenues by geographic area, one may wish to combine pairs of counties whose common boundary has changed.

3-5. Multicounty governments

Some government units, such as school districts, can serve multiple counties. However, the Census Bureau records a single county for each unit's location. There is no flag in the data for government units that serve multiple counties, so we could not determine which units are affected by this or how many such units exist. The FIPS code in the data for each unit serving multiple counties corresponds to the location recorded by the Census Bureau.

4. Similar data and documentation elsewhere

ICPSR distributes "Annual Survey of Governments: Government Finance File" and "Annual Survey of Governments: Finance Data" files for various years. In principle, the ICPSR files contain the same data as this archive. However, we found that numerous Census Bureau and FIPS geocodes are incorrect in the ICPSR files. Some variables also appear to be missing or garbled for some years. Nonetheless, the codebooks for the ICPSR data may be useful in working with this archive.

The Census Bureau distributes Annual Survey of Governments and Annual Census of Governments data in various forms on its Web site.

Documentation there may be useful. In particular, recent versions of survey forms are posted at <http://www.census.gov/govs/www/surveyforms.html>.

Appendix: Generation of FIPS Codes

We started with a copy of the ASG dataset from ICPSR that included both Census and FIPS state and county codes for many observations in many years. This copy of the dataset could not be used as the final dataset because some variables were missing or miscoded, possibly because the data file was in ASCII format. We obtained a clean dataset in SAS format from the Census Bureau, but it did not have FIPS codes.

We compiled a table of all combinations of Census and FIPS codes that occurred in the ICPSR dataset, together with the number of observations in which each combination occurred. We identified each case in which a given Census code matched more than one FIPS code and vice versa. For each such combination, as well as each combination observed fewer than 10 times in the entire dataset, we either verified its accuracy or removed it from the table. We verified combinations using a Census Bureau table of equivalences between Census and FIPS codes as of 2002 (*County_City/FIPS/Official_FIPS.xls*, provided by the U.S. Census Bureau). We also used a Census Bureau list of every change in county FIPS codes since 1970 (summarized in *County_City/FIPS/FIPS_Changes.xls* and available on the Web at <http://www.census.gov/geo/www/tiger/ctychng.html>).

In some cases, a given combination of Census Bureau and FIPS codes is not valid in a particular year because the FIPS code did not exist in that year. This happens when counties are created or are merged into other counties. In other cases, a given Census Bureau state and county code must be paired with different FIPS codes in different years, because the FIPS coding system has changed.

We expanded our table of Census and FIPS codes by creating a separate row for each pair in each year in which it was valid. The expanded table is *County_City/FIPS/nlookupall.sas7bdat*. The variable *v1v3ln* contains the concatenated Census state and county codes, the variable *FIPS_STC* is the FIPS state and county code, and the variable *Year* is final two digits of the year for which the pair is valid.

We checked for errors in the lookup table by using the Census state and county codes to generate a matching FIPS code for each observation in the ICPSR dataset. For about 200 observations outside Alaska, the looked-up FIPS code did not match the FIPS code in the dataset; these are the pairs we rejected when we built the lookup table. We determined the correct FIPS code for each of these 200 observations by contacting the government unit to ask what county it is in. In all but eight cases, the values in the lookup table were correct. In the eight cases, the FIPS codes in the original dataset were correct but the Census codes were incorrect. We did not verify FIPS codes in Alaska.

Next, we used the lookup table of Census and FIPS codes to generate FIPS state and county codes for each observation in the clean Census Bureau dataset. Outside Alaska, we were unable to generate a FIPS code for only one observation, and the values of all variables in that observation are zero.

If the county identified by *STATE* and *COUNTY* did not exist in the year of the observation, we recorded in *FIPS_STC* the FIPS code that would have applied had the county existed. We used *FIPS_predate* or *FIPS_postdate* to note when the county was created or ceased to exist, respectively, and flagged the problem with a value of 1 in *FIPS_err*.

According to the Census Bureau, values in the second dataset of *ID* and, in some cases, *STATE* and *COUNTY* have been changed to reflect the location of the government unit in 2002, regardless of the year of the observation. This makes it possible to track all government units from year to year.

We searched the Census dataset for the eight cases that produced incorrect looked-up FIPS codes in the ICPSR dataset. In four cases, the data errors that led to incorrect FIPS values had been corrected, so there was no error in the new dataset. In the remaining four cases, the errors remained, and we modified the matching program to insert the correct FIPS codes and flag the affected observations.

July 8, 2003