

## README File for 2000-2004 Archives

### Data Available at this FTP Site

The American Community Survey (ACS) keeps a historical record of Data Products available through the Census FTP survey at [http://www2.census.gov/acs/downloads/Core\\_Tables/](http://www2.census.gov/acs/downloads/Core_Tables/). The Data Products are stored in directories by year from 2000 through 2004.

Six Data Products are provided on this site – Base Tables, Data Profiles, Subject Tables, Multiyear Profiles, and Selected Population Profiles.

**Base Tables** provide basic housing and population characteristics. These tables are the foundation upon which higher-level profiles are built. The Base Tables come complete with the lower bound and upper bound of a confidence interval. The lower bound and upper bound are based on the 90 percent level of confidence.

**Data Profiles** provide estimates of selected demographic, social, economic, and housing characteristics. The profiles come complete with the lower bound and upper bound of a confidence interval. The lower bound and upper bound are based on the 90 percent level of confidence.

**Multiyear Profiles** for 2004 show data from 2000 to 2004. They use estimates from the single-year profiles to show year-to-year changes, the related percentage distributions, and whether the differences are statistically significant at the 90 percent confidence level. This product shows detailed demographic, social, economic, and housing data about selected ACS geographic areas over several years in tabular form. The 2001-2002 multiyear profiles use 2 years of estimates from the single-year tabular profiles to show year-to-year changes, and the related percentage distributions, differences over the time period, margins of error for the differences, and whether the differences are statistically significant.

**Subject Tables** highlight a particular subject of interest. Each subject table is accompanied by its margin of error and the confidence intervals are based on the 90 percent level of confidence. They are similar to the Census 2000 Quick Tables, but they contain much more detail than Quick Tables. Subject tables display percent distributions rather than the estimates. Universe lines are displayed as numeric estimates to show the base of each distribution. Subject tables allow for other measures such as medians and means where appropriate, and include the imputation rates for relevant measures.

**Selected Population Profiles** are similar in format to the subject tables and highlight subjects of interest. Each subject table is accompanied by its margin of error and the confidence intervals are based on the 90 percent level of confidence. Groups covered by these profiles include teenagers, people in poverty, the older population, workers, children, grandchildren, and the foreign born.

The **Iterated Selected Population Profile** is a single table that can be viewed, or “iterated” for many race, ethnic, ancestry groups, and country of birth. The population group choices are similar to those found in the Census 2000 Summary Files 2 and 4. Selected Population Profile's are published for the nation, states, and other geographic areas that have a total population of 1,000,000 and a group or iterated population of 250,000 or greater.

The chart below provides a reference for which data products are available by year.

Table 1: Data Products Available by Year

Data Products Availability	Base Tables	Data Profiles	Subject Tables	Selected Population Profiles	Multiyear Profile	Iterated Selected Population Profiles
2000	X					
2001	X					
2002	X	X			X	
2003	X	X			X	
2004	X	X	X	X	X	X

## Geographic Overview

Geographic areas published for the 2000-2004 American Community Survey (ACS) can be divided into ACS Test Sites and the remaining ACS areas. Thresholds for publishing a geographic area in 2000-2004 were as follows:

	Single Year Estimates	Change Estimates
<b>Test Sites</b>	<b>65,000</b>	<b>250,000</b>
<b>Remaining Areas</b>	<b>250,000</b>	<b>1,000,000</b>

1. Data collected in the American Community Survey Test Sites were sampled at a higher rate. Single year estimates are produced for geographic areas with a household population of 65,000 or more. Change estimates are produced for geographic areas with a household population of 250,000 or more.

### 2000-2004 American Community Survey Test Sites

Jefferson County, AR  
 Pima County, AZ  
 San Francisco County, CA  
 Tulare County, CA  
 Upson County, GA\*  
 Broward County, FL  
 Black Hawk County, LA  
 Lake County, IL  
 Miami County, IN\*  
 DeSoto Parish, LA\*  
 Hampden County, MA  
 Calvert County, MD  
 Iron, Reynolds, and Washington  
 Counties, MO\*  
 Madison County, MS  
 Flathead and Lake Counties, MT

Douglas County, NE  
Otero County, NM\*  
Bronx County, NY  
Rockland County, NY  
Franklin County, OH  
Multnomah County, OR  
Fulton County, PA\*  
Schuylkill County, PA  
Sevier County, TN  
Fort Bend and Harris Counties, TX  
Starr and Zapata Counties, TX  
Petersburg, VA\*  
Yakima County, WA  
Oneida and Vilas Counties, WI\*  
Ohio County, WV

\*\* Test Sites with household populations of less than 65,000

2. Data for the remaining American Community Survey areas are available for geographies that are *wholly contained within self-representing Primary Sampling Units*. PSU is short for primary sampling unit. It is a county or a group of counties that are treated as a single sample unit. Each PSU falls within the boundary of a state.

The American Community Survey uses a two-stage sample of housing units for those areas. The first stage of sampling divides the state into PSUs and selects a sample of these PSUs. Then, the final (second stage) sample of housing units is taken from the chosen PSUs.

Certain PSUs (roughly, those with an estimated population of 250,000 or more) are designated as "self-representing" (SR). The remaining PSUs are referred to as "non-self representing" (NSR). The NSR PSUs are grouped into clusters and only two are chosen to represent all of the PSUs in the cluster. In the final sample, housing units are taken from all the SR PSUs and only the two selected NSR PSUs.

For more information on PSU definition and selection for sample, see the archived "Accuracy of the Data" section on the ACS website and review the Accuracy of the Data documents prior to 2005. These archived Accuracy of the Data documents are located on the following URL:

[http://www.census.gov/acs/www/UseData/Accuracy/Accuracy\\_Archived.htm](http://www.census.gov/acs/www/UseData/Accuracy/Accuracy_Archived.htm)

3. Metropolitan areas are defined as of June 30, 1999.

4. 2000 and 2001 data were published for the 106<sup>th</sup> Congressional Districts; 2002 and 2003 data are published for the 108<sup>th</sup> Congressional Districts.

5. Thresholds have been set to ensure that the estimates are statistically reliable. Listed below are the threshold criterion which will determine which estimates will be released for each of the geographic areas with the population exceeding 65,000 for the ACS Sites and 250,000 for the remaining ACS areas. Both of the criterion must be met in order to be released to the public:

- The weighted table universe count must be greater than or equal to 500 for the ACS Sites or 2500 for the remaining ACS geographic areas.
- The un-weighted count of the records which satisfy the table universe must be at least double the number of cells in the table, excluding total and sub-total cells, i.e. the average

cell size must be at least twice.

## Overlapping Geographic Summary Levels

Some geographies may appear under multiple summary levels. For example, Bronx, New York is shown under the following summary levels:

050 – County - Bronx County

060 - County subdivision (MCD) - Bronx borough

155 - Place-County - New York city, Bronx County pt.

## Data Products

### Base Tables

The Base Table files contain the Base Tables released for that geography. All base tables for a given year are contained in a comma-separated value (“.csv”) file. The name of each file begins with “BaseTables” and ends with the year of the tables. Between “BaseTables” and the year of the tables is a geographic code, which specifies whether the data refer to the national level, state level, and county level, or other geographic division. Table 2 describes these codes.

Table 2: Geographic Summary Level Codes

Geography summary level	Description
010	Nation Level
040	State Level
050	County Level
060	County Subdivision Level
155	Place County Part Level
160	Place Level
380	Primary Metropolitan Statistical Area
385	Metropolitan Statistical Area - Consolidated Metropolitan Statistical Area
500	Congressional District Level (108 <sup>th</sup> Congress)
A01	ACS Test Site
A40	ACS Test Site County

Example: the file “BaseTables0102003.csv” contains all base tables for the nation (that is, summary level 010) for the year 2003.

### *Layout of the Base Tables*

Table 3 shows the layout for the base tables files.

Table 3: File layout for Base Tables

Column Number	Column Header/Description*
1	Geographic Code
2	Table Number
3	Table Line Number
4	Stub

5	Estimate
6	Lower-bound Estimate
7	Upper-bound Estimate

\*This applies to the first row only.

***Data Profiles (Single-year and Multi-year)***

Data Profiles provide estimates of selected demographic, social, economic, and housing characteristics for each geographic area. There are two types of profiles, single-year and multi-year. Single-year profiles contain estimates for the year indicated on the file name; multi-year estimates provide estimates from the year 2000 to the year indicated on the file name. The multi-year profiles include “Significant Difference” columns, which indicate whether an estimate from a previous year is statistically significantly different at the 90 percent level of confidence from the most recent year on the file.

The data profile files begin with “Profiles” for single-year profiles and “MultiYearProfiles” for multi-year profiles. Like the base table files, this is followed by a geographic summary level code and then by the year.

Examples:

- The file “Profile0402004.csv” contains all single-year profiles for the states (that is, summary level 040) for the year 2004.
- The file “MultiyearYearProfiles0502002.csv” contains all multi-year profiles for the counties (that is, summary level 050) for the year 2002.

***Layout of the Single-Year Profiles***

Table 4a shows the layout for the single-year profiles.

Table 4a: File layout for Profile Tables

Column Number	Column Header/Description*
1	Geographic ID Code
2	Geographic Name
3	Table Number
4	Order
5	Stub
6	Estimate
7	Lower Bound
8	Upper Bound

***Layout of the Multiple-Year Profiles***

Table 4b shows an example of the layout for the multi-year profile. In this case, the layout is for the years 2000-2004.

Table 4b: File layout for Profile Tables

Column Number	Column Header/Description*
1	Geographic ID Code
2	Geographic Name
3	Table Number
4	Table Order
5	Stub

6	Estimate 2004
7	Estimate 2003
8	Significant Difference Interval: 03-04
9	Estimate 2002
10	Significant Difference Interval: 02-04
11	Estimate 2001
12	Significant Difference Interval: 01-04
13	Estimate 2000
14	Significant Difference Interval: 00-04

\*This applies to the first row only.

\*\*This is a flag that indicates whether there were significant differences at the 90 percent confidence level between the two years listed.

### ***Layout of the 2001-2002 Multiyear Profiles***

Table 4c shows an example of the layout for the 2001-2002 multi-year profile tables.

Column Number	Column Header/Description*
1	Geographic Code
2	Geographic Name
3	Table Number
4	Table Order
5	Stub
6	Estimate 2002
7	Percent Estimate 2002
8	Estimate 2001
9	Percent Estimate 2001
10	Percent Change
11	Margin of Error of Difference **
12	Statistically Significant **

\*This applies to the first row only.

\*\*This is a flag that indicates whether there were significant differences at the 90 percent confidence level between the two years listed.

### **Subject Tables**

Subject tables highlight a particular subject of interest. Each subject table is accompanied by its margin of error. Confidence intervals are based on the 90% level of confidence. Subject Tables are available for 2004. Subject tables begin with either “ST” (for “subject table”) or “SPP” (for “selected population profile”). They are followed by a geographic summary level code and then by the year. Subject tables are available for 2004 only.

Examples:

- The file “SPP0402004.csv” contains the selected population profiles for the states (that is, summary level 040) for the year 2004.
- The file “ST0502004.csv” contains all subject tables for the counties (that is, summary level 050) for the year 2004.

### ***File Layout for Subject Tables (non-SPP)***

Table 5: File layout for Subject Tables

Column Number	Column Header/Description	Variable Name*
---------------	---------------------------	----------------

1	Geographic Code	GEOID
2	Name of Area	Name of Area
3	Table Number	Table Id
4	Table Order	order
5	Stub	STUB
6	Estimate Column 1	cest1
7	Margin of Error Column 1	moe1
8	Estimate Column 2	cest2
9	Margin of Error Column 2	moe2
10	Estimate Column 3	cest3
11	Margin of Error Column 3	moe3
12	Estimate Column 4	cest4
13	Margin of Error Column 4	moe4
14	Estimate Column 5	cest5
15	Margin of Error Column 5	moe5
16	Estimate Column 6	cest6
17	Margin of Error Column 6	moe6
18	Estimate Column 7	cest7
19	Margin of Error Column 7	moe7

For the “ST” files, forty-three tables with different layouts comprise the subject tables. Since the subject tables are bundled together into one file, table columns mean different things depending on the table being listed. At the top of each table, however, are headings that describe the columns’ meanings. Variable names were referenced in the column headers instead of labels for the subject tables because each subject table contains a different number of columns and a variable format.

**Selected Population Profiles**

Selected Population Profiles provide detailed information for a specific population group for selected geographic areas.

***File Layout for Selected Population Profile Tables (SPP)***

Table 6: File Layout for SPPs

Column Number	Column Header/Description*
1	Geographic ID Code
2	Geographic Name
3	Table Number
4	Order
5	Stub
6	Total Pop Estimate
7	Margin of Error
8	Selected Pop Estimate
9	Margin of Error

**Iterated Selected Population Profiles**

The Iterated Selected Population Profile is a single table that can be viewed for many specific population groups, including race, ethnic and ancestry groups. The population group for this product are similar to those found in the

**File Layout for the Iterated Selected Population Profile Tables (SPP)**

Table 7: File Layout for Iterated SPPs

Column Number	Column Header/Description*
1	Geographic Identifier
2	Geographic Name
3	Table Number
4	Order
5	Population Group
6	Subject (Stub)
7	Total Population Estimate
8	Margin of Error
9	Group Estimate
10	Margin of Error

**About the FTP Application**

This FTP (File Transfer Protocol) application is intended for users of census data, compressed files, and spreadsheet/database software. It provides quick access to data users, such as State Data Centers and news media, who need to begin their analysis immediately upon data release.

Each state directory provides all files available for the identified state. Once downloaded, the data are in Excel or PDF format. No software is provided. Users of the FTP application can then import it into the spreadsheet/database software of their choice for data analysis and table presentation.

**FTP File Transfer**

To facilitate transferring files, we suggest using features commonly found in most vendors' FTP utility. When testing the download in a PC environment, we used the ws\_ftp product. This product, and many other FTP products developed for the PC environment, allows individual or multiple file selection using the control key or block multiple file selection using the shift key.

In the UNIX environment, the "mget" command allows the transferring of single or multiple files. Using the ftp command mget, followed by a single filename, will allow a single file to be downloaded. Using the command mget followed by a wildcard character, such as \* or \*.\* , allows multiple files to be downloaded. For example, once you have navigated into the 2003 Profile directory for California, you can download all of the spreadsheets in the County Level subdirectory with the following ftp commands:

```
ftp>binary          (set file transfer mode to binary)
ftp>prompt off      (to avoid being asked for verification of each file, optional)
ftp>mget *.*         (to copy multiple files)
ftp>quit            (to exit ftp)
```

**Where to find more information:**

A fact sheet explaining the differences between the income and poverty estimates from the

American Community Survey and the Annual Social and Economic Supplement to the Current Population Survey can be found on the following link:

<http://www.census.gov/Press-Release/www/2004/factsheet.html>

In addition, data from previous years can either be located at the American Community Survey home page at < <http://www.census.gov/acs/www/>> or on the *American FactFinder* main page at < [http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?\\_pageId=sp1\\_acs](http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?_pageId=sp1_acs)>.

Any updates made to the data due to found error are displayed in Errata section of the ACS website located on: <http://www.census.gov/acs/www/UseData/Errata.htm>.

To better understand the results of the American Community Survey, 53 population variables, and 42 housing variables have identified and defined in the Subject Definitions document, located on <http://www.census.gov/acs/www/UseData/Def.htm>. This document includes information on how the data were collected and how they are presented in data products.

Between 2003 and 2004 the ACS detailed table numbering scheme was modified. Please reference the following ACS 2004 Base Table Mapping List for more details on the changes made to each detailed table and table ID: [ftp://www2.census.gov/acs/downloads/shells/2004\\_Tables.xls](ftp://www2.census.gov/acs/downloads/shells/2004_Tables.xls)

For additional information, contact Kai Wu or Nick Spanos in the Census Bureau American Community Survey Office. Kai can be contacted on (301) 763-2626 or at [kai.t.wu@census.gov](mailto:kai.t.wu@census.gov). Nick can be contacted on (301) 763-6841 or at [nicholas.m.spanos@census.gov](mailto:nicholas.m.spanos@census.gov).