

Chapter 1

Accessing the Digital Census

A. About the Census

Over the decades the actual census questionnaire has undergone considerable modification. Changes have been made to its content, phrasing of questions, geographical units, and collection procedures. Though the censuses of 1980 and 1990 were very similar in the questions, the geographical units, and the tabulation of results, Census 2000 made a radical departure in the race category. See the discussion later in this section.

The last three censuses made extensive use of sampling that resulted in two questionnaires. On one, a basic short list of questions about gender, age, marital status, and housing was asked of everyone. The tabulated results are sometimes referred to as the *100 percent count* or *complete count* data. On the other, additional details were asked of only about a one in six sample of households. Tabulations are often referred to as the *sample count* or *sample* data.

B. Digital Census Data

The Bureau of the Census reports the population and housing census information in two major digital formats. The first is now called a *Summary File* and it contains population aggregations for selected variables. In 1990 the term was *Summary Tape File*. The second is the *Public-Use Microdata Sample (PUMS)*. This contains separate records for each household and individual. This file is very useful because it enables researchers to measure interrelationships between variables by person or housing unit rather than by geographical area. The researcher also has the ability to create custom tabulations.

In addition to population and housing, the Bureau of the Census provides a number of other tabulations such as government, business, foreign trade, manufacturing, and agriculture (moved to Dept. of Agriculture in 1996). There are also some historical population counts and special tabulations such as the county-to-county migration file. While important, these are beyond the scope of this module. Readers can browse the *Subjects Index* to look for numerous reports, studies, and data sets. (<http://www.census.gov/acsd/www/subjects.html>)

1. Summary Files

The SF files are tabulations and cross-tabulations that correspond to much of the census information in published volumes. Data include items such as counts of persons and households, persons by race by sex by age, housing type by tenure, and so on. Summary Files come as four major types: 1, 2, 3, and 4. In addition, there is the *Redistricting Data PL 94-171*

Summary File which is the first release of census information after a census. It includes only basic race tabulations for persons over and under age 18.

SF1 and SF2 contain information from the complete-count questionnaire on gender, ethnicity, marital status, and a few housing variables. SF3 and SF4 contain information from the sample-count questionnaires on education, occupation, income, migration, etc. Because the sample-count contains more questions, these files are much larger than SF1 or SF2

Summary Files 2 and 4 have tables repeated for up to 250 or 1000 ethnic groups respectively. The only condition for suppression is that there must be at least 50 ethnic persons sampled in a geographic unit for the data to appear. Thus, in SF2 and SF4 there are numerous missing locations for small groups within smaller geographic units. Both may be very useful when census tabulations are desired for a specific ethnic group such as Japanese, Cubans, Germans, or Cherokee Indians.

In the figure shows the number of tables provided in each of the four summary file types. A *P* variable is a population tabulation and an *H* variable is a housing tabulation. If a variable is preceded by a *PCT* or *HCT* then it will not be reported for units finer than census tracts. Some tabulations are broken out by individual ethnic groups and these special tabulations have a suffix of A through I appended to the variable name. See below for a list.

Table Type and Number	SF1	SF2	SF3	SF4
P	171		160	
H	56		121	
PCT	59	36	76	213
HCT		11	48	110
Race Crosstabs	14		51	
Race Categories		250		1000

Ethnic Group Suffixes for Tables

- A - White alone
- B - Black alone
- C - American Indian or Alaska Native alone
- D - Asian alone
- E - Hawaiian or Pacific Islander alone
- F - Some other race alone
- G - Two or more race alone
- H - Hispanic
- I - Non-Hispanic White alone.

2. Table Details

Before going too much further, it would be helpful to see something of the structure of a typical table. While it is easy to extract such data from the Census web site, you should be familiar with table structure in order to better use the resulting output or to understand how to extract data from raw census files should that ever become necessary.

Below is part of *Table P6* on race from Summary File 3. Several important pieces of information are included in the label. The *P6* indicates it is the sixth tabulation of population, the table title is *Race*, the *[8]* indicates there are eight items in the table, and the *Universe* indicates that the counts are based on the entire population. Many tables use subsets of the total population for the *Universe*. This table was generated for state totals at my request and the web page only displays the first ten states. I would have to click the *Next* button to see the next ten states.

This table was created for viewing on the screen. Data tables for downloading contain similar information, but the user must keep track of the labels and Universe population.

[P6. RACE \[8\] - Universe: Total population](#)

Data Set: [Census 2000 Summary File 3 \(SF 3\) - Sample Data](#)

NOTE: [Corrected counts](#) are available for one or more geographies displayed in this table.

geographies 1-10 of 52 [Next](#) ▶

NOTE: Data based on a sample except in P3, P4, H3, and H4. For information on confidentiality protection, sampling error, nonsampling error, and count corrections see <http://factfinder.census.gov/home/en/datanotes/expsf3.htm>.

	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delaware
Total:	4,447,100	626,932	5,130,632	2,673,400	33,871,648	4,301,261	3,405,565	783,60
White alone	3,161,671	434,225	3,871,715	2,137,166	20,122,959	3,558,579	2,777,794	584,68
Black or African American alone	1,153,044	21,968	154,316	417,881	2,219,190	159,279	305,902	148,82
American Indian and Alaska Native alone	22,897	97,012	253,542	18,492	312,215	43,101	9,419	3,11
Asian alone	29,908	25,496	91,223	19,081	3,682,975	93,306	82,277	16,05
Native Hawaiian and Other Pacific Islander alone	1,187	3,122	6,166	1,534	113,858	4,298	1,357	33
Some other race alone	29,155	9,869	597,173	41,164	5,725,844	310,552	148,809	16,24
Two or more races	49,238	35,240	156,497	38,082	1,694,607	132,146	80,007	14,35

U.S. Census Bureau
Census 2000

Below is a spreadsheet of data for two downloaded tables in *Excel* format from Summary File 3. The variables this time are reported for three different selected geographic units, the United States, California, and Los Angeles County. Note that each geographic unit has a *SUMLEVEL* code that identifies the type of geographic unit. Each also has a unique FIPS code (*GEOID2*) and a name that identifies the specific place. The *GEOID2* code is critical if you plan on linking this data to geographic units in a mapping program.

The first table, *Table 6 – Race*, is the same as that shown above. *P006001* is the first item in Table 6 and it is the value for the total population. Note these item values. The *P006* indicates *Population Table 6* and the *001* indicates it is the first item which in this case is the total population. These identifiers are important for data software that can not handle the lengthy column and row descriptions. The identifier definitions can be found in the summary file documentation.

The second table, *Table PCT74B – Median Earnings in 1999 for Black Alone population 16 years and over with earnings in 1999*, has 6 items that provide additional detail about the working Black population. Note the *B* suffix. This second table (and all other tables, for that matter) has a *Universe* that includes only Black or African American alone population 16 years and over with earnings in 1999. You need to be careful to use the proper *Universe* population in making subsequent calculations such as percents.

Each table has two identifiers, a brief variable name such as *P006001* and a description such as *Total population: Total*. In programs like *Excel* the description is helpful in precisely defining the variable, but if the table is to be converted to a *dbf* format table care must be taken to drop the long identifier since the *dbf* column type is capable of handling only one line of labels of no more than eight characters each. Thus one might want to generate descriptive labels. *P006001* might become *Totpop* and *P006002* might become *Totwhalo*. The *Universe* could be cleverly worked into the table name such as *SF3p6race_tot* or *SF3pct47b_16wearn*.

One also must use care when summing rows of a table. Some of the variables are subtotals of the *Universe* that would cause a column sum to be inflated. For example, *P006001* below amounts to the sum of all following rows within each of the three geographic units.

GEO_ID	Geography Identifier	01000US	04000US06	05000US06037
GEO_ID2	Geography Identifier		06	06037
SUMLEVEL	Geographic Summary Level	010	040	050
GEO_NAME	Geography	United States	California	Los Angeles Co., California
P006001	Total population: Total	281,421,906	33,871,648	9,519,338
P006002	Total population: White alone	211,353,725	20,122,959	4,622,759
P006003	Total population: Black or African American alone	34,361,740	2,219,190	916,907
P006004	Total population: American Indian and Alaska Native alone	2,447,989	312,215	68,471
P006005	Total population: Asian alone	10,171,820	3,682,975	1,134,263
P006006	Total population: Native Hawaiian and Other Pacific Islander alone	378,782	113,858	27,221
P006007	Total population: Some other race alone	15,436,924	5,725,844	2,262,925
P006008	Total population: Two or more races	7,270,926	1,694,607	486,792
PCT074B001	Black or African American alone population 16 years and over with earnings in 1999: Median earnings in 1999 ; Worked full-time; year-round in 1999 ; Total	27,264	33,982	34,175
PCT074B002	Black or African American alone population 16 years and over with earnings in 1999: Median earnings in 1999 ; Worked full-time; year-round in 1999 ; Male	30,000	36,391	36,313
PCT074B003	Black or African American alone population 16 years and over with earnings in 1999: Median earnings in 1999 ; Worked full-time; year-round in 1999 ; Female	25,589	31,728	32,180
PCT074B004	Black or African American alone population 16 years and over with earnings in 1999: Median earnings in 1999 ; Other ; Total	9,930	11,601	12,229
PCT074B005	Black or African American alone population 16 years and over with earnings in 1999: Median earnings in 1999 ; Other ; Male	10,402	11,766	12,319

PCT074B006	Black or African American alone population 16 years and over with earnings in 1999: Median earnings in 1999 ; Other ; Female	9,554	11,459	12,161
------------	--	-------	--------	--------

In their raw form, all the tables are organized sequentially into a series of files for each state. Each file contains part or several of the tables depending on how many items are involved, but the intent is to break up the volume of data into manageable chunks. Thus, you do not download an entire summary file, but only the portion (file) that contains the table of interest to you for your selected state. Summary File 1 in raw form contains 39 files for the various tables and Summary File 3 contains 76. You would need to consult a figure that lists which population and housing tables are contained within which files. For example, *Table 74B* above for California is contained in the 52nd file, *ca00052_uf3.zip*. The file contains *Tables 74A* through *75C* and its size is about 7 Mb.

The 1990 census was much like that of 2000 except that there were only P or H tables. There was for each summary tape file an A, B, or C tabulation that differed by the levels of geography that were included. The C tabulation, for example, covered the entire United States, but did not provide geographic detail below counties or places over 10,000 persons. For summary tape files 1 and 3 there also was a D tabulation for congressional districts. One structural difference within STF2 and STF4 is that ethnic tabulations were embedded as *b* records and totals as *a* records within the files. In 2000, the ethnic tabulations were represented as individual files.

3. The American Community Survey

In the mid-2000s the Bureau of the Census initiated a new file that will eventually replace SF3 and SF4. Called the *American Community Survey*, the file is based on an annual survey of 3 million households and will provide estimated counts for the previous year. For geographical units greater than 65,000 persons, the data will be reported annually. For units between 20,000 and 65,000 persons, the data will be based on a three-year average, and for units smaller than 20,000, data will be based on a five-year average. The results will be based on an accumulation of data that will be surveyed from household each month of the previous year rather than a single time period. For averaged data, the earliest year will be dropped from the average with each subsequent data collection. Group quarters will be handled separately and not included in the totals as in previous censuses. Recently, data has been published for the larger units, but the smaller units will not be published until 2010.

Although sampling has been a part of census statistics for some time, the American Community Survey makes this issue more evident than ever before. For each table, the Bureau of the Census publishes data containing the estimated values, the *Margin of Error* (MOE), and the *standard error*. These can be used to determine the statistical significance of a difference between two geographic areas.


For counts of the total population and for the population by age, sex, race, and Hispanic Origin, the Bureau of the Census recommends using the controlled population estimates that it

generates in its Population Estimates Program. When these values appear in tables (see below) they contain a series of asterisks under the *MOE* column.

In the partial data profile for Los Angeles County shown below the estimated count of sex and age appear in the second column. The *Margin of Error* is based on a confidence interval of 90% which is a value the Bureau of the Census prefers. This means that if the survey was conducted 100 times, the estimated value would fall within the range surrounding the estimate 90 times. Thus for females aged 5 to 9 years the confidence interval extends from 721,324 to 741,026. Note that for larger samples the margin of error becomes proportionately smaller.

One could calculate the standard error of the estimate by dividing the *MOE* by 1.65. The standard error is that due to sampling and from it one could calculate a higher confidence interval of 95 or 99% by multiplying the standard error by 1.96 or 2.58 respectively.



Los Angeles County, California
General Demographic Characteristics: 2005 
 Data Set: 2005 American Community Survey
 Survey: 2005 American Community Survey
Demographic - Sex and Age, Race, Relationship, Household by Type
[Social](#) - Education, Marital Status, Fertility, Grandparents...
[Economic](#) - Income, Employment, Occupation, Commuting to Work...
[Housing](#) - Occupancy and Structure, Housing Value and Costs, Utilities...
[Narrative](#) - Text profile with graphs for easy analysis...

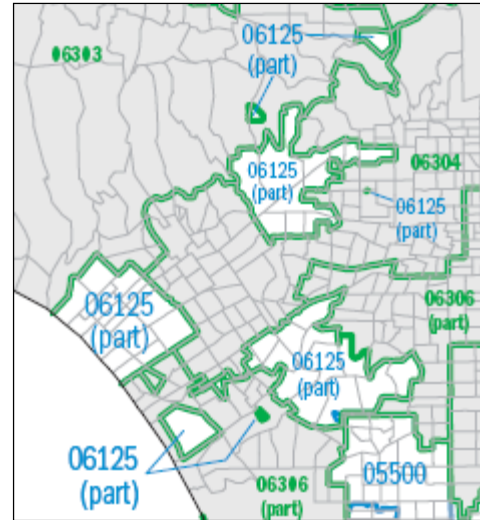
NOTE. Data are limited to the household population and exclude the population living in institutions, college dormitories, and other group quarters. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see [Survey Methodology](#).

General Demographic Characteristics: 2005	Estimate	Margin of Error
Total population	9,758,886	*****
SEX AND AGE		
Male	4,816,227	+/-676
Female	4,942,659	+/-676
Under 5 years	753,834	*****
5 to 9 years	731,175	+/-9,851
10 to 14 years	798,292	+/-9,850
15 to 19 years	701,441	+/-1,154
20 to 24 years	668,647	+/-1,079
25 to 34 years	1,430,627	+/-983

C. Public-Use Microdata Sample Files

There are two PUMS files, which contain data for either a 5% sample for all of the housing units in a state or a 1% sample of all the housing units in the United States. These data are particularly useful because they are for individual persons and housing units. In 1980 an estimate of the total number of persons in a state was obtained by multiplying the sample value by 20 or 100, but in 1990 and 2000 each person and housing unit received an individual weight that is used to estimate the total population. PUMS files provide considerable detail on a number of variables and the appendix lists the necessary codes to deal with these variables.

The 1990 and 2000 PUMS files contain a number of geographic areas called PUMAs (Public-Use Microdata Areas) or SuperPUMAs. See Appendix for a list of California PUMAs. PUMAs contain a minimum of 100,000 persons in the 5% sample and SuperPUMAs contain 400,000 persons in the 1% sample. In 1980 Los Angeles County had only 3 geographic units (Los Angeles City, Long Beach City, and the remainder of County). However, in 1990 and 2000 the county was divided into over 50 PUMAs that greatly expanded the geographic value of the PUMS data. In heavily populated places like the city of Los Angeles, PUMAs consist of aggregations of tracts while in other areas they may be aggregations of incorporated places. Unfortunately these places are often not contiguous. Note at right how PUMA 06125 in Los Angeles has been split among the cities of Santa Monica, Beverly Hills, Culver City, Marina Del Rey, and pieces of Los Angeles County.



The PUMS data set has a different structure than the Summary Files. It is arranged in a hierarchical structure in which both housing and person record types are found in the same file. Data for a housing unit appears first and then a person record follows for each person in the household. Each person record contains a household identifier and codes to indicate the position of that person in the household.

D. Geography in Summary Files

The boundaries used to aggregate census information have their origins in the *TIGER* files that the Bureau of the Census has been refining over the last 30 years. A *TIGER* file consists basically of descriptions of each street segment. A segment is usually the length of road between two intersections, but it may follow a city boundary, a stream, or a coastline. For each segment, variables describe the address ranges on both sides, the blocks, tracts, ZIP codes, Congressional districts, etc. on both sides, the street name, and the latitude and longitude coordinates of the end points. Using these files, the Bureau of the Census can determine which census unit a returned census form is in as well as the address coordinates. Also, from these files the boundaries of various geographic units can be created by looking for only those segments that have different area identifiers on each side. Those with the same value are eliminated. *TIGER* files are of little value to most people unless they have specialized software that can process the segments into other useful forms.

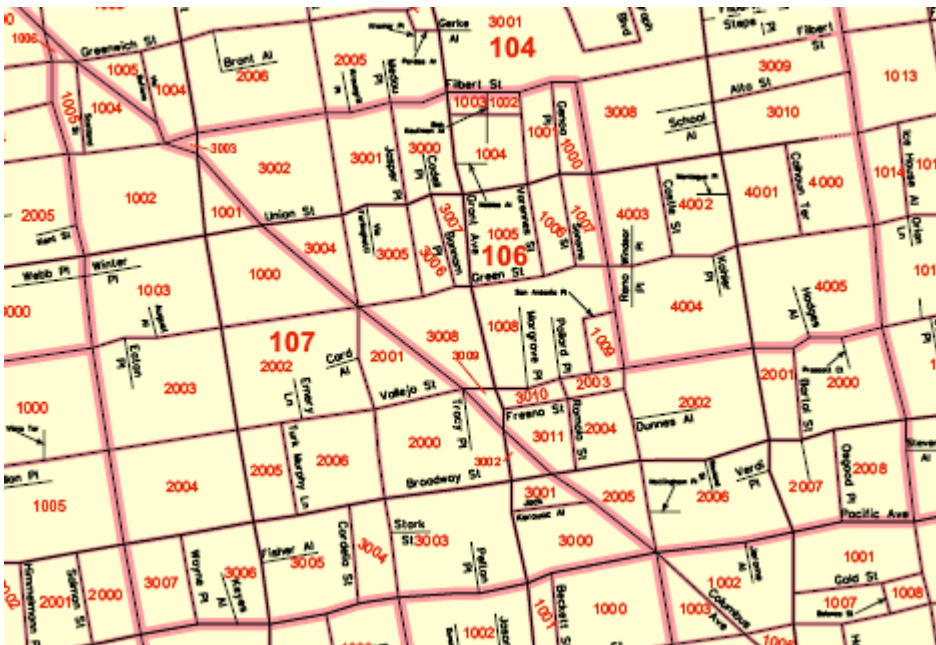
What makes the Summary Files large is that each of the tabulations is reported for multiple types of geographic units derived from *TIGER* files. These types are organized hierarchically from larger to smaller units and are defined by *Summary Level Codes*. When working with raw data one typically has to consult documentation to determine the appropriate code so that a desired set of geography can be extracted from all the geographic record types contained in a file. These codes are critical for extracting the proper records from the larger raw files and they

can be found on page 4-1 of the census documentation. They also are important in grouping data should you download different types of geographic units at the same time.

The diagram from the Bureau of the Census below illustrates the hierarchy of the various geographical units for which they report data.



The map below shows census blocks and tracts (heavier lines) in San Francisco.



Examine the following extract (ordered by size of unit) of census geography definitions to better understand some of the more significant smaller geographic types:

Consolidated metropolitan statistical area (CMSA)

A geographic entity defined by the federal Office of Management and Budget for use by federal statistical agencies. An area becomes a CMSA if it meets the requirements to qualify as a metropolitan statistical area, has a population of 1,000,000 or more, if component parts are recognized as primary metropolitan statistical areas, and local opinion favors the designation. Example: **Los Angeles--Riverside--Orange County, CA CMSA**

Primary metropolitan statistical area (PMSA)

A geographic entity defined by the federal Office of Management and Budget for use by federal statistical agencies. If an area meets the requirements to qualify as a metropolitan statistical area and has a population of one million or more, two or more PMSAs may be defined within it if statistical criteria are met and local opinion is in favor. A PMSA consists of one or more counties (county subdivisions in New England) that have substantial commuting interchange. When two or more PMSAs have been recognized, the larger area of which they are components then is designated a consolidated metropolitan statistical area. Example: **Los Angeles--Long Beach, CA PMSA**

Metropolitan statistical area (MSA)

A geographic entity defined by the federal Office of Management and Budget for use by federal statistical agencies, based on the concept of a core area with a large population nucleus, plus adjacent communities having a high degree of economic and social integration with that core. Qualification of an MSA requires the presence of a city with 50,000 or more inhabitants, or the presence of an Urbanized Area (UA) and a total population of at least 100,000 (75,000 in New England). The county or counties containing the largest city and surrounding densely settled territory are central counties of the MSA. Additional outlying counties qualify to be included in the MSA by meeting certain other criteria of metropolitan character, such as a specified minimum population density or percentage of the population that is urban. MSAs in New England are defined in terms of minor civil divisions, following rules concerning commuting and population density. Example: **Santa Barbara--Santa Maria--Lompoc, CA MSA**

County and equivalent entity

The primary legal subdivision of most states. In Louisiana, these subdivisions are known as parishes. In Alaska, which has no counties, the county equivalents are boroughs, a legal subdivision, and census areas, a statistical subdivision. In four states (Maryland, Missouri, Nevada and Virginia), there are one or more cities that are independent of any county and thus constitute primary subdivisions of their states. The District of Columbia has no primary divisions, and the entire area is considered equivalent to a county for statistical purposes. In Puerto Rico, municipios are treated as county equivalents.

Census county division (CCD)

A subdivision of a county that is a relatively permanent statistical area established cooperatively by the Census Bureau and state and local government authorities. Used for presenting decennial census statistics in those states that do not have well-defined and stable minor civil divisions that serve as local governments.

Place

A concentration of population either legally bounded as an incorporated place, or identified as a Census Designated Place (CDP, comprising a densely settled concentration of population that is not within an incorporated place, but is locally identified by a name) including comunidades and zonas urbanas in Puerto Rico. Incorporated places have legal descriptions of borough (except in Alaska and New York), city, town (except in New England, New York, and Wisconsin), or village.

Town

A type of minor civil division in the New England states, New York, and Wisconsin and a type of incorporated place in 30 states and the Virgin Islands of the United States.

Census tract

A small, relatively permanent statistical subdivision of a county delineated by a local committee of census data users for the purpose of presenting data. Census tract boundaries normally follow visible features, but may follow governmental unit boundaries and other non-visible features in some instances; they always nest within counties. Census tracts average about 4,000 inhabitants and are designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment. They may be split by any sub-county geographic entity.

Block group (BG)

A subdivision of a census tract (or, prior to 2000, a block numbering area), a block group is the smallest geographic unit for which the Census Bureau tabulates sample data. A block group consists of all the blocks within a census tract with the same beginning number.

Example: block group 3 consists of all blocks within a 2000 census tract numbering from 3000 to 3999. In 1990, block group 3 consisted of all blocks numbered from 301 to 399Z.

Census block

A subdivision of a census tract (or, prior to 2000, a block numbering area), a block is the smallest geographic unit for which the Census Bureau tabulates 100-percent data. Many blocks correspond to individual city blocks bounded by streets, but blocks -- especially in rural areas -- may include many square miles and may have some boundaries that are not streets. The Census Bureau established blocks covering the entire nation for the first time in 1990. Previous censuses back to 1940 had blocks established only for part of the nation. Over 8 million blocks are identified for Census 2000.

1. SF3 Summary Level Code Hierarchy for Selected Geographic Units

```

40 - State
    50 - County
        60 - County subdivision
            70 - Place or place part
                80 - Census tract
                    90 - Block group
  
```

The above geographic units may be split by a higher level unit. For example, many tracts are split by place boundaries and many places are split into separate, non-contiguous areas. However, if only data within a place is wanted, one would use a summary level of 80 to extract only tracts or parts of tracts that fell entirely within a specific city.

To accommodate the need for unsplit units such as would be found in most counties, additional records are available. It is these records (140 and 150) that geographers often seek for mapping contiguous geographic units within a county.

2. Other SF3 Summary Level Codes

```

160 - Place
140 - Census tract
150 - Block group
500 - Congressional district
170 - Consolidated city
390 - Metropolitan area
    871 - ZIP code (ZCTA)
  
```

3. Coding Geographic Units - FIPS Codes

All geographic units have a standardized number identifier referred to as a *FIPS* (Federal Information Processing Standards) *Code*. The appendix lists FIPS codes for all U.S. states and counties. For named places such as MSAs, states, counties, and cities, the FIPS codes follow an alphabetical organization of the names. For example, in California, Alameda County has a FIPS code of *001* and Yuba County has a code of *115*.

Often these FIPS codes are used to limit a search of the state records to a specific area of interest. Thus for all data within the State of California, one could limit the search area to Los Angeles County by specifying a county FIPS code of *037* and to only tracts by specifying a summary level of *140*.

E. Census Comments.

1. The Race Question

In Census 2000, persons could indicate more than one race on the questionnaire, and fortunately only 2.4% of the national population did so. (4.7% of Californians did so) Of the multiracial population 93% indicated only two races and 32% indicated *White plus Some Other Race*. The latter *Other* category proved mostly to be Latino. Thus, most researchers use the tabulations for single-race only categories. See *Overview of Race and Hispanic Origin: Census 2000 Brief* (<http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf>)

This issue is particularly of concern to people trying to estimate changes since 1990. In 1990, persons were forced to choose only one race response, but in 2000 they could check as many as they wished. For larger groups such as Whites, Blacks, and Asians the difference between comparing the 1990 race count with the 2000 single-race only count or single race plus any other race count is not great. However, for small groups such as Asian subgroups like Thai or Hmong, the differences in percent change between using the single race only population versus the single race plus any other race can be very large. This topic, called *Bridging*, can be examined in greater detail at (<http://www.census.gov/Press-Release/www/2001/raceqandas.html>)

The question on race in the U.S. Census is separate from the question on Hispanic origin. People can indicate a particular race such as White, Black, American Indian, any of several Asian groups, or Other. Then they also may indicate if they are or are not of Spanish/Hispanic origin such as Mexican, Cuban, or Puerto Rican. Hispanics often indicate their race as white, yet Whites are commonly seen as distinct from Hispanics. Thus, tabulations based on the total reported White race are complicated by two distinctly different groups. To compensate for this it is usually better to use the non-Hispanic White category when tabulating data for "whites." This removes those persons of white race who indicated that they also were of Hispanic origin. In Census 2000 a number of tables have been included for this special race category.

2. Hispanic Subgroups

A problem has been noted in counts of small Hispanic subgroups like Dominican and Salvadoran. Apparently because of the wording in of the Hispanic question in Census 2000 many Hispanics in smaller groups chose to respond *Other Hispanic* rather than write in their subgroup as they did in 1990. Thus smaller counts have been noted in 2000 that seem to run counter to increases noted in immigration data. Larger groups like Mexican, Cuban, and Puerto Rican were listed on the Census 2000 questionnaire and do not appear to have this problem.

Another problem observed in 1990 was that there were far too many Hispanic Black persons in census tracts. This later proved to be related to the allocation of these characteristics to non-responding individuals, but the data was not updated.

3. Small-area Tabulations

In small areal units such as census tracts unusual results may appear because most if not all the population is in an institution such as a jail or college campus. It is sometimes helpful to identify and exclude such tracts from analysis.

Another problem when calculating change from 1990 is that many persons in institutions may be shifted between adjacent tracts between the last two censuses. Perhaps due to a change in the address of a college dormitory or jail, this often appears on a map as a large gain and loss in adjacent tracts.

4. Census Geography

Some geographical units undergo change between censuses. These include boundaries for places, ZIP codes, Congressional districts, school districts, tracts, block groups, and blocks. Although census tracts theoretically are only to be split if the population grows significantly or to be joined if the population drops, it is common for boundaries to be moved into adjacent tracts. Thus, comparison of data at the block group and tract level between two census decades can be complicated. The Bureau of the Census publishes equivalency tables to draw attention to where changes have occurred and some private companies redistribute the population into tracts with common boundaries. Typically, the numbers are redistributed based on percent of common area or are based on streets, but the new numbers are estimates and not actually based on the addresses of respondents to the census.

F. Exploring the Census Web Site

The Bureau of the Census web site provides a rich source of information in published and raw forms. There are various research reports that cover major variables like race, income, education, and marriage in the United States, summary tables of statistics for various geographic units, and printed maps of census units and selected

variables. There are interactive programs that allow a user to generate a map or graph of desired data. And finally, there are the raw statistics and the digital boundary files for mapping that a user can download to produce custom tables, graphs, and maps.



The figure above shows the opening web page from the Bureau of the Census. Its look and content change as new products become available though the overall appearance has been relatively similar the last few years. On the left panel are links to various resources provided by the Bureau. In the center panel are links to specific tables, maps, and data sets and on the right are links to several search engines to help a user find data on a topic or place.

In Exercise 1 you will follow some of the links shown above. The first link, *A*, takes you to basic information on a place of interest. *B* is the link to the *American Factfinder* where data from the last two censuses can be found. *C* is the link to the new *American Community Survey* data. *D* is the link to various news releases and research documents on various population and housing subjects. *E* is the link to census reference maps in pdf format and to boundary file data for GIS programs.

G. Exercises

- Ex 1. Exploring the Census Web Site
- Ex 2. Accessing Census 2000 SF3 and SF4 Data and Data Bases at ICPSR
- Ex 17. Downloading Raw Census Data

