Chapter 1
The census: An introduction

At the beginning of the last century, most U.S. households contained seven or more people. By 1950, most households contained two people.

In 1950, people living alone accounted for less than one out of ten households (9.5 percent). By 2000, more than one out of four households, about 26 percent, consisted of a person living alone.

Women headed one out of five households in 1970, but thirty years later they headed one out of every three households.

In 1950, about 78 percent of households were made up of married couples; in 2000, married couples represented only about half of all households (nearly 52 percent). During that same period, other types of family households (men and women living with family members but with no spouse) grew from 11.3 percent of all households to 16.4 percent.

Nearly half of all family households headed by a man with no spouse, and nearly three out of ten of family households headed by a woman with no spouse, had children under 18 in 2000 (compared to one in five for male-headed and about one in three female-headed households in 1950) (Hobbs and Stoops 2002, chapter 5).

Underlying this brief synopsis is an array of dramatic social and economic changes that have occurred over the past century—different mores regarding divorce and children born outside marriage, new economic opportunities for women, changed responsibility for elderly parents, and a housing market more accommodating to individuals living alone. Talk-show hosts and politicians may see in these numbers yet another indicator of the decline of U.S. standards of morality; others may applaud the decline of the "married majority" and the expansion of individual choice that these changing demographics indicate.

Regardless of one's personal attitudes, understanding changes in household structure (along with changes in the population, housing stock, commuting patterns, and economy) is the starting place for responding to the new challenges and opportunities they present. A recent cover story in *BusinessWeek* proclaims the latest round of social revolution: "Unmarried America: say goodbye to the traditional family. Here's how the new demographics will change business and society" (Conlin 2003, 106). The tax and pension penalties that unmarried people bear are the focus of the article; similar points could be made about job structure, health care, and housing. Governments, firms, and social organizations ignore changing social and economic structure at their peril.

As the source of the data quoted above, the U.S. census plays an essential role in charting and analyzing these changes. Exponential improvements in computer hardware and software over the past decades have enabled a wider range of analysts (even those with quite limited technical ability) to answer the questions that should shape intelligent public policy. While many types of researchers may make use of the census (from market researchers to investment analysts to presidential candidates), we focus here on researchers with a primarily urban and regional focus. For this group of users, the spatial analysis capabilities we have at the beginning of the twenty-first century offer an exciting way to enrich our understanding of the implications of the bald numbers in the census files.

This book focuses on the research needs of a broadly defined group of "urban analysts"—city planners, community development organizations, real estate development firms, economic development specialists, transportation planners and engineers, property appraisers, and social service providers. We assume readers are well acquainted with the substantive challenges entailed in analyzing cities and regions and have a basic level of familiarity with widely used computer programs such as spreadsheets and general-purpose databases. We assume readers have little or no familiarity with census data or GIS software, although we do not provide a manual for any particular GIS software package. Rather, we aim to provide the beginning or intermediate census user with the principles, skills, and techniques needed to locate, download, and analyze available data using any one of a number of GIS packages on the market. We expect the reader will consult the manual for the particular package used to find out how to create a thematic map, perform a travel-demand analysis, and so on. Our book, we hope, will answer all the essential questions a researcher may have about how to link census and other urban or regional data to any particular GIS package, how to design analyses to answer specific questions, and how to interpret the results.

This chapter introduces the key principles and concepts underlying the U.S. census. The remainder of this section defines the census and explains how it has evolved in the United States since it was first conducted in 1790. The role of GIS in census-based analyses is outlined. The second section describes the decennial census (the Census of Population and Housing) and explains why and how it has evolved to the present. In this section we discuss the major way the decennial census will change in the next decade as the American Community Survey (ACS) is

phased in to replace the so-called "long form"—the detailed demographic, economic, and housing questions asked of a sample of U.S. households. We address the issues raised by sampling, rather than enumerating, all these detailed characteristics. Finally, we discuss the accuracy of the data.

The third section of the chapter covers the basic concepts users should grasp to use the data appropriately. Census geography and geographic summary levels are explained. These concepts are the basis for the description of the geographic hierarchy and the FIPS (Federal Information Processing Standards) numbering system with which all spatial analysts work. The organization of summary files, tables, and variables is explained next. Finally, the major census data products are described. This section aims to demystify the initially confusing (but quite logically structured) array of data types confronting a new user. Where appropriate, we refer the reader to the detailed technical documentation provided by the Census Bureau.

What is "the census"?

"The census" is actually many different censuses and various related surveys. So what is a census? The word "census" comes to us from the Latin *censere* (or in English to *assess*) and originally referred to the enumeration and registration of people and property, often for the purpose of taxation. In its more modern sense (which entered the English language only in the middle eighteenth century) a "census" is usually understood to be an enumeration—that is, a count of everybody or everything. The Census of Population and Housing, the particular census that many professionals mean when they use the word "census," does include a 100 percent count of a population and housing. Surveys, which provide most of the information in the Census of Population and Housing, are usually understood to rely on a sample of people or households or businesses or whatever. New users may wonder whether the census would be better if none of it was based on a sample; the intuition here is that counting everything is surely better than counting only some things.

In fact, most demographers argue that a properly defined survey may give *more* accurate information than a full enumeration would. The Census Bureau, the agency responsible for conducting the Census of Population and Housing and various related surveys, uses surveys both because they are often more accurate than full enumerations, and also because they are much cheaper. However, the Census Bureau also continues to enumerate the population because, in January 1999, the Supreme Court ruled that Section 195 of Title 13 of the U.S. Code precludes the use of statistical sampling to produce congressional apportionment counts.[1] Sampling is however legal for other (nonapportionment) purposes, including redistricting.

What is a 100 percent count (a full enumeration) of the population? Does the Census Bureau ever miss people in its decennial Census of Population and Housing? In other words, is the 100 percent count really 100 percent? Are all homeless people counted? Are illegal immigrants ever missed? And what about those people who, for whatever reason, prefer to have no contact with government? In fact, it is likely that the census fails to count millions of people. Moreover, not all those missed are homeless, illegal immigrants, or radical libertarians. So a 100 percent census

[1]Title 13 of the U.S. Code specifies the powers and duties of the Census Bureau, the basic procedures of the census, and the sort of data that may be collected.

does not necessarily cover 100 percent of people—this is why surveys may be more accurate than censuses. Using statistical techniques, samples can be designed to deal explicitly with the problem of hard-to-count populations. Fortunately we have a *fairly* good idea of how many people have been missed, what they look like, and where they live. But *fairly* good is far from complete information. We have better data on how many African-Americans are missed than, say, how many illegal immigrants are missed. And counts of the homeless population are, in most cases, little better than broad estimates. So the Census of Population and Housing is not quite as complete as many would like to believe. But in many cases it provides us with the best data there is on people in the United States. Moreover, the data has status in law. The process of apportioning the 435 seats in the House of Representatives to the fifty states, mandated by Article 1 Section 2 of the U.S. Constitution and undertaken every ten years, is based on the portion of the nation's population in each state as measured by the decennial census.[2]

The census population numbers are also used to

- determine many other state and local political districts;
- distribute federal funds to states and local areas, state funds to local areas, and federal, state, and local funds to individual neighborhoods (around $283 billion is distributed annually this way from the federal government alone); and
- evaluate many federal, state, and local programs.

Regardless of its flaws, the census has enormous importance in the daily lives of state and local planners, government officials, and most land-development professionals.

The Economic Census is the other major census program in the United States (we discuss the Economic Census in more detail in chapter 4). It is conducted every half-decade, in years ending in two and seven (e.g., 1992, 1997, 2002). Again, much of this "census" information actually relies on surveys. Both federal and state governments also conduct a number of independent surveys and censuses (County Business Patterns, Current Population Survey, ES202, and so on). In the subject-area chapters that follow, we point readers to a variety of related data sources. The discussion in this chapter focuses on the broad organization of the Census of Population and Housing. A large amount of other official data follows the general organizational structure of the Census of Population and Housing.

GIS and the census

At its simplest, a geographic information system (GIS) is a database manager connected to another software program that is able to draw maps digitally. Queries of the database can be shown using maps. So if we wanted to know which counties in New Jersey had a population of more than a million people, we would query an appropriate state or national database (usually derived from the Census of Population and Housing), and the GIS would then be instructed to draw a map indicating which counties met this criterion and which did not.

GIS software has been around since the 1970s, but it is only over the past decade that GIS has become broadly institutionalized in local government, urban planning, and other land

[2]Each state is assigned one representative, and the remaining 385 seats are distributed according to population.

professions (for simplicity, we will refer to this group of users as urban researchers). Technical users with interests in the census have used GIS since the technology surfaced in order to help make sense of the huge amount of information provided by the census. Over the past half decade, the number of people using GIS has dramatically increased due to changes in the technology. GIS has become both more powerful and easier to use, and the connection between GIS software and census data is now straightforward.

In fact, since the 1990 Census, a GIS has been an all but necessary tool for any significant census analysis. Census data can now be incorporated into a GIS, with minimal preprocessing. And a GIS will allow the connection of this data to other data sources in a way that traditional databases and statistical software do not. For instance, if you were to analyze how school children's travel time would change with rural school district consolidation, it would be necessary to combine demographic information on school districts (available from the Census of Population and Housing) with the location of schools (possibly from a text-based database of school street addresses that would then be geocoded—in others words, turned into a map). Travel times for individual children to proposed consolidated schools could then be computed and compared to current commuting times.

Admittedly, census data can be used by researchers without computers or any knowledge of GIS. It is possible to go to the census books in so-called "depository" libraries[3] and, provided you understand the rudiments of census geography and the variables, look up data the old-fashioned way. However, computers allow much faster, easier access to data. Thousands, if not millions, of pieces of data may be downloaded at one time into one of a variety of user-friendly file formats that make the data relatively easy to use and analyze. Doing this by hand would take months, if not years. GIS makes the acquisition and analysis of data much faster. Since the census is organized and published around defined spatial units—what we call "census geography," which includes states, counties, cities, tracts, and blocks—the built-in spatial analysis capabilities of a GIS make analyzing census data that much more intuitive and powerful. Many of the ancillary census products, such as TIGER/Line geographic data and the Census Transportation Planning Package (CTPP) journey-to-work data, are most useful when used within a GIS framework. So although you need not know GIS to use the census, a GIS will greatly simplify (and in some instances, make practically possible) more advanced analyses of census data.

The Census of Population and Housing

The U.S. Constitution mandates that the Census of Population and Housing be undertaken every ten years to apportion seats in the House of Representatives. Over the years the census has grown in size and function, collecting an increasingly wide range of data on the U.S. public.[4] Costs have also increased. In 1790 the census cost $44,377 and counted 3,929,214 people. In 2000 the census cost $4.5 billion and counted 281,421,906 people. Staff (enumerators) have also increased, from 650 in 1790 to over a half million in 2000. Although the U.S. census may

[3]Depository libraries (often at larger state universities) are designated by the Census Bureau as storekeepers (the Federal Depository Library Program). There are about fourteen hundred depository libraries in the nation, at least one in each congressional district. The libraries keep copies of most federal publications.

[4]For a complete description of the changes to the questions asked, see U.S. Census Bureau (2002a). The chapters that follow discuss recent changes.

Table 1.1 Summary of questions asked	
100 percent Short Form	
Name	Race
Household relationships	Hispanic origin
Sex	Tenure (home owned or rented)
Age	
Sample long form	
Population	**Housing**
Social characteristics	Physical characteristics
Marital status	Units in structure
Place of birth, citizenship, and year of entry	Year structure built
School enrollment and educational attainment	Number of rooms
Ancestry	Year moved into residence
Residence five years ago	Plumbing and kitchen facilities
Language spoken at home and ability to speak English	Telephone
Veteran status, disability	Vehicles available
Grandparents as caregivers	Heating fuel
Economic characteristics	Farm residence
Labor force status	Financial characteristics
Place of work and journey to work	Value of house or monthly rent paid
Occupation, industry, and class of worker	Utilities, mortgage, taxes, insurance, etc.
Work status in 1999	
Income in 1999	

Source: Adapted from U.S. Census Bureau (2002b) at www.census.gov/mso/www/c2000basics/00Basics.pdf.

not be the modern world's oldest (the eleventh century Domesday book is arguably the oldest; England's Board of Trade conducted local censuses in the American colonies beginning in the early 1600s, and Sweden conducted a full national census in 1749), it is the world's oldest continuous national census.

The census questionnaire

There are actually two questionnaires: the 100 percent short form questionnaire (which satisfies the legal requirements of apportionment) and the long form questionnaire completed by a sample of people and households. Nationwide, about one in six households receives the long form, but the rate varies considerably depending on the size of the area. In smaller areas the rate may be as high as one in two, while in densely populated areas it may be as low as one in eight.

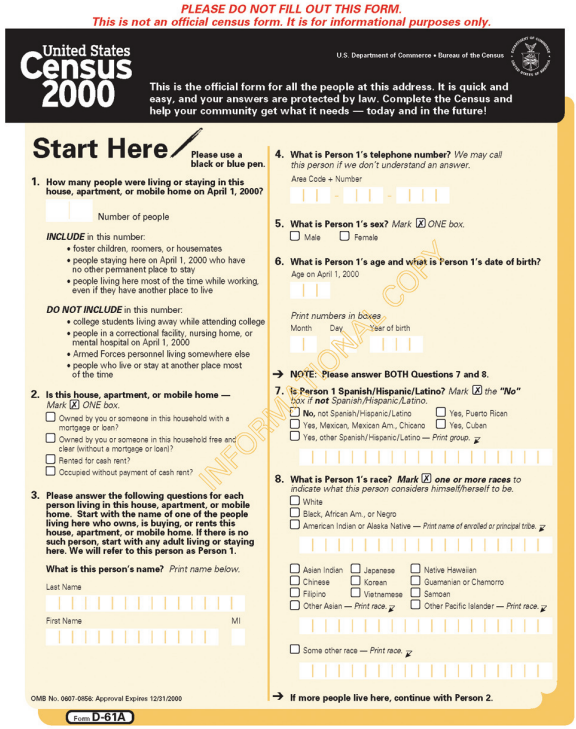
Table 1.1 summarizes the questions asked in the short and long forms. The actual short form questionnaire for Census 2000 is shown in box 1.1, and the long form questionnaire is shown

[4]For a complete description of the changing questions asked, see U.S. Census Bureau (2002a). The chapters that follow discuss recent changes.

in the appendix. Before actually using census variables, it is always a good idea to look at the questions asked to generate those variables. Both the short and the long forms were printed in six languages in 2000 (English, Chinese, Korean, Spanish, Tagalog, and Vietnamese). Non-English forms could be requested in advance.

The questionnaire has changed over time. The Census Bureau has documented most of these changes and differences in instructions given to census enumerators in a wonderful historical publication, *Measuring America: The decennial census from 1790 to 2000* (U.S. Census Bureau 2002a).

Box 1.1 The short form questionnaire



PLEASE DO NOT FILL OUT THIS FORM.
This is not an official census form. It is for informational purposes only.

United States Census 2000

U.S. Department of Commerce • Bureau of the Census

This is the official form for all the people at this address. It is quick and easy, and your answers are protected by law. Complete the Census and help your community get what it needs — today and in the future!

Start Here Please use a black or blue pen.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2000?
   Number of people
   INCLUDE in this number:
   - foster children, roomers, or housemates
   - people staying here on April 1, 2000 who have no other permanent place to stay
   - people living here most of the time while working, even if they have another place to live

   DO NOT INCLUDE in this number:
   - college students living away while attending college
   - people in a correctional facility, nursing home, or mental hospital on April 1, 2000
   - Armed Forces personnel living somewhere else
   - people who live or stay at another place most of the time
2. Is this house, apartment, or mobile home — Mark ☒ ONE box.
   ☐ Owned by you or someone in this household with a mortgage or loan?
   ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
   ☐ Rented for cash rent?
   ☐ Occupied without payment of cash rent?
3. Please answer the following questions for each person living in this house, apartment, or mobile home. Start with the name of one of the people living here who owns, is buying, or rents this house, apartment, or mobile home. If there is no such person, start with any adult living or staying here. We will refer to this person as Person 1.
   What is this person's name? *Print name below.*
   Last Name
   First Name MI
4. What is Person 1's telephone number? *We may call this person if we don't understand an answer.*
   Area Code + Number
5. What is Person 1's sex? *Mark ☒ ONE box.*
   ☐ Male ☐ Female
6. What is Person 1's age and what is Person 1's date of birth?
   Age on April 1, 2000
   Print numbers in boxes.
   Month Day Year of birth

→ NOTE: Please answer BOTH Questions 7 and 8.

7. Is Person 1 Spanish/Hispanic/Latino? *Mark ☒ the "No" box if not Spanish/Hispanic/Latino.*
   ☐ No, not Spanish/Hispanic/Latino
   ☐ Yes, Puerto Rican
   ☐ Yes, Mexican, Mexican Am., Chicano
   ☐ Yes, Cuban
   ☐ Yes, other Spanish/Hispanic/Latino — *Print group.*
8. What is Person 1's race? *Mark ☒ one or more races to indicate what this person considers himself/herself to be.*
   ☐ White
   ☐ Black, African Am., or Negro
   ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
   ☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
   ☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
   ☐ Filipino ☐ Vietnamese ☐ Samoan
   ☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
   ☐ Some other race — *Print race.*

→ If more people live here, continue with Person 2.

OMB No. 0607-0856: Approval Expires 12/31/2000

Form D-61A

Person 2

Your answers are important! Every person in the Census counts.

1. What is Person 2's name? *Print name below.*
   Last Name
   First Name MI
2. How is this person related to Person 1? *Mark ☒ ONE box.*
   ☐ Husband/wife
   ☐ Natural-born son/daughter
   ☐ Adopted son/daughter
   ☐ Stepson/stepdaughter
   ☐ Brother/sister
   ☐ Father/mother
   ☐ Grandchild
   ☐ Parent-in-law
   ☐ Son-in-law/daughter-in-law
   ☐ Other relative — *Print exact relationship.*

   If NOT RELATED to Person 1:
   ☐ Roomer, boarder
   ☐ Housemate, roommate
   ☐ Unmarried partner
   ☐ Foster child
   ☐ Other nonrelative
3. What is this person's sex? *Mark ☒ ONE box.*
   ☐ Male ☐ Female
4. What is this person's age and what is this person's date of birth?
   Age on April 1, 2000
   Print numbers in boxes.
   Month Day Year of birth

→ NOTE: Please answer BOTH Questions 5 and 6.

5. Is this person Spanish/Hispanic/Latino? *Mark ☒ the "No" box if not Spanish/Hispanic/Latino.*
   ☐ No, not Spanish/Hispanic/Latino
   ☐ Yes, Puerto Rican
   ☐ Yes, Mexican, Mexican Am., Chicano
   ☐ Yes, Cuban
   ☐ Yes, other Spanish/Hispanic/Latino — *Print group.*
6. What is this person's race? *Mark ☒ one or more races to indicate what this person considers himself/herself to be.*
   ☐ White
   ☐ Black, African Am., or Negro
   ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
   ☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
   ☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
   ☐ Filipino ☐ Vietnamese ☐ Samoan
   ☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
   ☐ Some other race — *Print race.*

→ If more people live here, continue with Person 3.

Person 3

Census information helps your community get financial assistance for roads, hospitals, schools, and more.

1. What is Person 3's name? *Print name below.*
   Last Name
   First Name MI
2. How is this person related to Person 1? *Mark ☒ ONE box.*
   ☐ Husband/wife
   ☐ Natural-born son/daughter
   ☐ Adopted son/daughter
   ☐ Stepson/stepdaughter
   ☐ Brother/sister
   ☐ Father/mother
   ☐ Grandchild
   ☐ Parent-in-law
   ☐ Son-in-law/daughter-in-law
   ☐ Other relative — *Print exact relationship.*

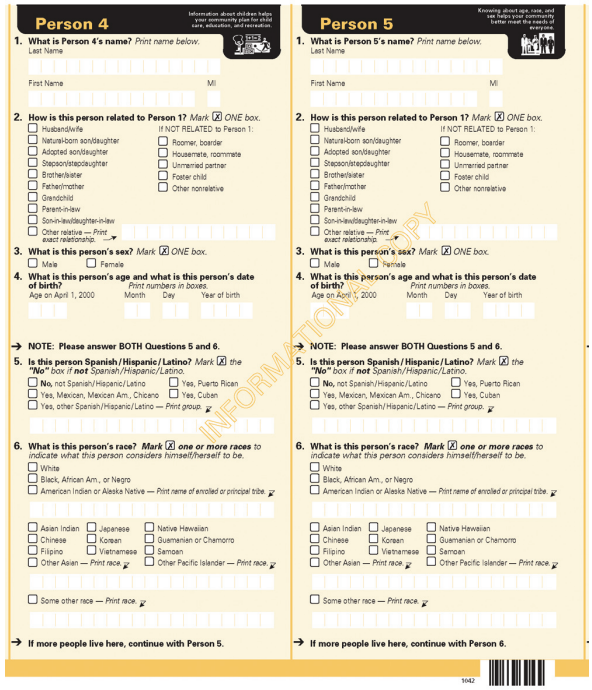
   If NOT RELATED to Person 1:
   ☐ Roomer, boarder
   ☐ Housemate, roommate
   ☐ Unmarried partner
   ☐ Foster child
   ☐ Other nonrelative
3. What is this person's sex? *Mark ☒ ONE box.*
   ☐ Male ☐ Female
4. What is this person's age and what is this person's date of birth?
   Age on April 1, 2000
   Print numbers in boxes.
   Month Day Year of birth

→ NOTE: Please answer BOTH Questions 5 and 6.

5. Is this person Spanish/Hispanic/Latino? *Mark ☒ the "No" box if not Spanish/Hispanic/Latino.*
   ☐ No, not Spanish/Hispanic/Latino
   ☐ Yes, Puerto Rican
   ☐ Yes, Mexican, Mexican Am., Chicano
   ☐ Yes, Cuban
   ☐ Yes, other Spanish/Hispanic/Latino — *Print group.*
6. What is this person's race? *Mark ☒ one or more races to indicate what this person considers himself/herself to be.*
   ☐ White
   ☐ Black, African Am., or Negro
   ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
   ☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
   ☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
   ☐ Filipino ☐ Vietnamese ☐ Samoan
   ☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
   ☐ Some other race — *Print race.*

→ If more people live here, continue with Person 4.

Box 1.1 The short form questionnaire (continued)



Person 4

Information about children helps your community plan for child care, education, and recreation.

1. **What is Person 4's name?** *Print name below.*
Last Name
First Name MI

2. **How is this person related to Person 1?** *Mark ☒ ONE box.*
☐ Husband/wife
☐ Natural-born son/daughter
☐ Adopted son/daughter
☐ Stepson/stepdaughter
☐ Brother/sister
☐ Father/mother
☐ Grandchild
☐ Parent-in-law
☐ Son-in-law/daughter-in-law
☐ Other relative — *Print exact relationship.*
If NOT RELATED to Person 1:
☐ Roomer, boarder
☐ Housemate, roommate
☐ Unmarried partner
☐ Foster child
☐ Other nonrelative

3. **What is this person's sex?** *Mark ☒ ONE box.*
☐ Male ☐ Female

4. **What is this person's age and what is this person's date of birth?**
Age on April 1, 2000
Print numbers in boxes.
Month Day Year of birth

→ **NOTE: Please answer BOTH Questions 5 and 6.**

5. **Is this person Spanish/Hispanic/Latino?** *Mark ☒ the "No" box if* ***not*** *Spanish/Hispanic/Latino.*
☐ **No,** not Spanish/Hispanic/Latino
☐ Yes, Puerto Rican
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Cuban
☐ Yes, other Spanish/Hispanic/Latino — *Print group.*

6. **What is this person's race?** ***Mark ☒ one or more races*** *to indicate what this person considers himself/herself to be.*
☐ White
☐ Black, African Am., or Negro
☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
☐ Filipino ☐ Vietnamese ☐ Samoan
☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
☐ Some other race — *Print race.*

→ **If more people live here, continue with Person 5.**

Person 5

Knowing about age, race, and sex helps your community better meet the needs of everyone.

1. **What is Person 5's name?** *Print name below.*
Last Name
First Name MI

2. **How is this person related to Person 1?** *Mark ☒ ONE box.*
☐ Husband/wife
☐ Natural-born son/daughter
☐ Adopted son/daughter
☐ Stepson/stepdaughter
☐ Brother/sister
☐ Father/mother
☐ Grandchild
☐ Parent-in-law
☐ Son-in-law/daughter-in-law
☐ Other relative — *Print exact relationship.*
If NOT RELATED to Person 1:
☐ Roomer, boarder
☐ Housemate, roommate
☐ Unmarried partner
☐ Foster child
☐ Other nonrelative

3. **What is this person's sex?** *Mark ☒ ONE box.*
☐ Male ☐ Female

4. **What is this person's age and what is this person's date of birth?**
Age on April 1, 2000
Print numbers in boxes.
Month Day Year of birth

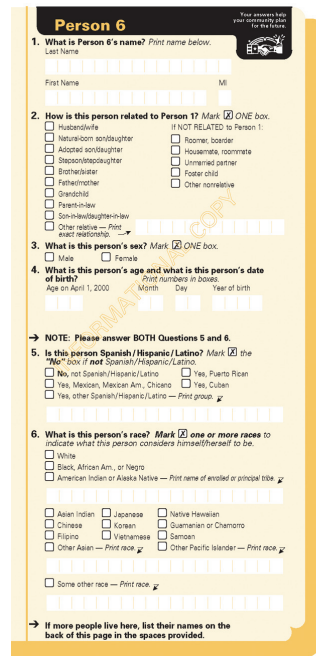
→ **NOTE: Please answer BOTH Questions 5 and 6.**

5. **Is this person Spanish/Hispanic/Latino?** *Mark ☒ the "No" box if* ***not*** *Spanish/Hispanic/Latino.*
☐ **No,** not Spanish/Hispanic/Latino
☐ Yes, Puerto Rican
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Cuban
☐ Yes, other Spanish/Hispanic/Latino — *Print group.*

6. **What is this person's race?** ***Mark ☒ one or more races*** *to indicate what this person considers himself/herself to be.*
☐ White
☐ Black, African Am., or Negro
☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
☐ Filipino ☐ Vietnamese ☐ Samoan
☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
☐ Some other race — *Print race.*

→ **If more people live here, continue with Person 6.**

1042



Person 6

Your answers help your community plan for the future.

1. **What is Person 6's name?** *Print name below.*
Last Name
First Name MI

2. **How is this person related to Person 1?** *Mark ☒ ONE box.*
☐ Husband/wife
☐ Natural-born son/daughter
☐ Adopted son/daughter
☐ Stepson/stepdaughter
☐ Brother/sister
☐ Father/mother
☐ Grandchild
☐ Parent-in-law
☐ Son-in-law/daughter-in-law
☐ Other relative — *Print exact relationship.*
If NOT RELATED to Person 1:
☐ Roomer, boarder
☐ Housemate, roommate
☐ Unmarried partner
☐ Foster child
☐ Other nonrelative

3. **What is this person's sex?** *Mark ☒ ONE box.*
☐ Male ☐ Female

4. **What is this person's age and what is this person's date of birth?**
Age on April 1, 2000
Print numbers in boxes.
Month Day Year of birth

→ **NOTE: Please answer BOTH Questions 5 and 6.**

5. **Is this person Spanish/Hispanic/Latino?** *Mark ☒ the "No" box if* ***not*** *Spanish/Hispanic/Latino.*
☐ **No,** not Spanish/Hispanic/Latino
☐ Yes, Puerto Rican
☐ Yes, Mexican, Mexican Am., Chicano
☐ Yes, Cuban
☐ Yes, other Spanish/Hispanic/Latino — *Print group.*

6. **What is this person's race?** ***Mark ☒ one or more races*** *to indicate what this person considers himself/herself to be.*
☐ White
☐ Black, African Am., or Negro
☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
☐ Asian Indian ☐ Japanese ☐ Native Hawaiian
☐ Chinese ☐ Korean ☐ Guamanian or Chamorro
☐ Filipino ☐ Vietnamese ☐ Samoan
☐ Other Asian — *Print race.* ☐ Other Pacific Islander — *Print race.*
☐ Some other race — *Print race.*

→ **If more people live here, list their names on the back of this page in the spaces provided.**

Form D-61A

Please turn to go to last page.

Persons 7 – 12

If you didn't have room to list everyone who lives in this house or apartment, please list the others below. *You may be contacted by the Census Bureau for the same information about these people.*

Person 7 — Last Name
First Name MI

Person 8 — Last Name
First Name MI

Person 9 — Last Name
First Name MI

Person 10 — Last Name
First Name MI

Person 11 — Last Name
First Name MI

Person 12 — Last Name
First Name MI

The Census Bureau estimates that, for the average household, this form will take about 10 minutes to complete, including the time for reviewing the instructions and answers. Comments about the estimate should be directed to the Associate Director for Finance and Administration, Attn: Paperwork Reduction Project 0607-0856, Room 3104, Federal Building 3, Bureau of the Census, Washington, DC 20233.

Respondents are not required to respond to any information collection unless it displays a valid approval number from the Office of Management and Budget.

Thank you for completing your official U.S. Census 2000 form.

The "Informational Copy" shows the content of the United States Census 2000 "short" form questionnaire. Each household will receive either a short form (100-percent questions) or a long form (100-percent and sample questions). The short form questionnaire contains 6 population questions and 1 housing question. On average, about 5 in every 6 households will receive the short form. The content of the forms resulted from reviewing the 1990 census data, consulting with federal and non-federal data users, and conducting tests.

For additional information about Census 2000, visit our website at **www.census.gov** or write to the Director, Bureau of the Census, Washington, DC 20233.

FOR OFFICE USE ONLY
A. JIC1 **B.** JIC2 **C.** JIC3 **D.** JIC4

If you need help completing this form, *call 1-800-XXX-XXXX between 8:00 a.m. and 9:00 p.m., 7 days a week. The telephone call is free.*

TDD — *Telephone display device for the hearing impaired. Call 1-800-XXX-XXXX between 8:00 a.m. and 9:00 p.m., 7 days a week. The telephone call is free.*

¿NECESITA AYUDA? *Si usted necesita ayuda para completar este cuestionario llame al 1-800-XXX-XXXX entre las 8:00 a.m. y las 9:00 p.m., 7 días a la semana. La llamada telefónica es gratis.*

There is a formal process to decide what questions are to be asked in any census; this naturally involves the Census Bureau, plus the Office of Management and Budget (OMB), the U.S. Congress, and census users around the country. Deciding what questions to ask involves balancing the federal government's need for information about its citizens with citizens' right to privacy. Generally speaking, for a question to be included, there must be some specific federal need for the data (data must be mandated, or required, by law or by the Constitution). Due concern must be given to the intrusiveness of questions. Continuity is also important; ideally, census questions should remain as constant as possible across censuses to enable historical comparisons. However, legal, social, or economic changes may require new sorts of data or new ways of measuring the same characteristics.

After the 1990 Census, the Census Bureau and the Office of Management and Budget (OMB) organized a review to determine what subjects to include in Census 2000. Federal agencies were asked to identify all legal mandates. The needs of a wide range of nonfederal users (state and local governments, community organizations, business, academics, religious groups, and the general public) were also surveyed. Various standing advisory committees, expert panels, users' groups, and so on were also consulted.

Two tests of proposed census content were then undertaken: the 1996 National Content Survey (also known as the U.S. Census 2000 Test) and the 1996 Race and Ethnic Targeted Test (also known as the 1996 Census Survey). Focus groups made comments on questionnaire design and content. Cognitive research was also undertaken, evaluating, among other things, behavioral responses to alternative visual designs of the proposed census. Some of this research is available at *www.census.gov/srd/www/byyear.html* (or search the U.S. Census Bureau home page by typing in "content determination"). One result was that both the short and the long forms had fewer subjects in 2000 than in 1990.

The Census Bureau submitted subjects planned for inclusion in Census 2000 to Congress on March 31, 1997, and the planned questions on March 30, 1998. Questions were then submitted to the OMB on June 30, 1998. The OMB has the statutory responsibility to ensure that questions meet essential data needs and that the time burden of filling out a questionnaire is kept to a minimum. For details on the full process, see U.S. Census Bureau (2002a).

Nevertheless, important changes were made to the 2000 questionnaire.[5] The change that has received the most public attention concerns race. The "race" question has undergone constant modification since the earliest days of the census. Respondents in 2000 were allowed to select one *or more* racial categories to describe their racial identity, or they could write in a

[5]For a detailed description of changes between the 1990 and 2000 questionnaire, see *Major differences in subject-matter content between the 1990 and 2000 Census questionnaire* at *www.census.gov/population/www/cen2000/90vs00.html.*

specific racial identification. In 1990 only one race category per respondent was allowed (this is explained in more detail in chapter 3). Other significant changes included a new question about grandparents as caregivers, a new definition of disabilities, and the elimination of a question about sewage disposal. Several questions were also moved from the short to the long form.

Many of the changes to the questionnaire reflect social shifts in U.S. society and law. Consider the case of the head-of-household in married-couple households. In 1970, the head-of-household was the person considered head by the household itself; but if a married woman was living with her husband, the husband was always defined as the head to simplify tabulations (Myers 1992, 51). By 1980 this was untenable. In the 1980 Census, the "head-of-household" terminology was replaced by "householder," and wives in married-couple households could be designated as the householder. Since the characteristics of the householder are, in some instances, assigned to the household, this complicated the identification of trends from previous censuses. Wives tend to be younger than their husbands, so this change may have resulted in an underestimation of the age of householders.

While there is a formal administrative process to determine the questions included in each census, external political voices also influence the process. In the buildup to the 1990 Census, many social researchers argued that it was important to ask questions about respondents' sexual behavior. Sexually transmitted diseases, particularly AIDS, had become a major public policy concern. There were, of course, good reasons not to ask such questions in the 1990 Census: for one the questions would be highly intrusive, which is why questions about sexual behavior had never been asked before. Because of privacy issues, a number of senators and congressional representatives were vehemently opposed to the questions, and they were never included in the 1990 Census questionnaire. However, the 2000 Census included a question about "unmarried partners" in nonfamily households, and same-sex unmarried partner households are now shown in the tables (see box 3.4, chapter 3).

Questions about religion have never been asked in the U.S. Census. However, they are common census questions in other countries; although religious classification in these countries is sometimes resisted in creative ways. In 2001, Statistics Canada reported that "[a]n astonishing 20,000 Canadians declared themselves to be followers of the religion of Jedi, the guardians of peace and justice in the Star Wars flicks" (Canada.com 2003*).* Canadians were outdone by the Australians and the British (at 70,000 and 400,000 followers of Jedi, respectively). A Canadian man circulated an e-mail encouraging people to identify themselves as followers of Jedi. He said the Jedi membership drive was his way of "thumbing his nose at the government for asking what he feels is an inappropriate question. 'My religion is my issue, not the government's'" (Canada.com 2003*).*

Changes to the census questionnaire, although undertaken in a formal administrative manner, nevertheless reflect political pressures on the Census Bureau, current federal and national policy concerns, and current social mores. The Census of Population and Housing is not a neutral and objective numerical description of the U.S. public. The census is also, in part, a political document reflecting current political, policy, social, and moral concerns. Nothing illustrates this better than the history of how questions about race and ethnicity have changed over time.

In the 1870 Census, the first census conducted after the abolition of slavery, the instructions to marshals (enumerators) read as follows:

"*Color*—It must not be assumed that, where nothing is written in this column, "White" is to be understood. The column is always to be filled. Be particularly careful in reporting the class *Mulatto.* The word is here generic and includes quadroons, octoroons, and all persons having any perceptible trace of African blood. Important scientific results depend upon the correct determination of this class in schedules 1 and 2" (U.S. Census Bureau 2002a, 14).

As in previous censuses, "Indians" (Native Americans) were not to be recorded on the schedule of population unless they were taxed (although marshals were, for the first time, encouraged to record nontaxed "Indians" living off reservations). In 1880, a separate schedule was developed to enumerate "Indians" living on reservations. By 1900, the instructions had been purged of detailed definitions of types of "mulattos" and instead read as follows, reflecting the immigration (and integration of Native Americans) that had occurred in the previous two decades:

"Column 5. Color or race. Write 'W' for white; 'B' for black (negro or negro descent); 'Ch' for Chinese; 'Jp' for Japanese, and 'In' for Indian, as the case may be" (U.S. Census Bureau 2002a, 36).

Another example is worth considering here. In the early 1990s, a few politicians argued that the traditional census had become a huge invasion of personal privacy and suggested the questionnaire be replaced by a new form, which was the size of a postcard. Obviously such a small form would mean that the vast majority of questions in the current census would have to be removed. This initiative got nowhere. Nevertheless, the Constitution mandates only the counting of people; it does not mandate the gathering of other census data.

The American Community Survey

One of the major problems faced by users of detailed demographic, economic, and housing data from the decennial census has been how to update estimates in the later years of each decade. There is usually a two- to three-year lag in the release of sample data, and by the end

of the period analysts are relying on data that is ten or twelve years old. In 1996, the Census Bureau began a new initiative that may resolve the problem by the end of this decade, assuming full congressional funding is provided. The American Community Survey (ACS) will be conducted annually for a sample of approximately 3,000,000 households (plus 36,000 households in Puerto Rico and 2.5 percent of the group quarters population). Once it is fully phased in, the ACS will provide the equivalent of long-form census data annually (U.S. Census Bureau 2003e, 3). A decennial census will still be conducted, but this will consist of the short form only. The sample will be drawn from every county, Native American Tribal Area, and Hawaiian Home Land, and although the sample size will not be equivalent to the approximately 17 percent sample on which the long form is based, preliminary evaluations suggest that sampling error will be only slightly increased for most places (Griffin and Obenski 2002, 27). Data from large jurisdictions (more than 65,000 people) will be available after the first year the ACS is fully implemented. For smaller jurisdictions (more than 20,000 people), a three-year average will provide an equivalent sample to that of the long form, and for places with less than 20,000 people, a five-year average will provide the equivalent. Assuming the ACS is fully implemented on schedule in July 2005, the equivalent of long-form data will be available for every location by 2010, eliminating the need for a long-form questionnaire for the decennial census. Updated data will be available annually from that point on. This will significantly enhance the census's usefulness for local planning efforts and overcome one of the major problems that data users have faced until this point.

While the ACS is intended to collect the same level of detail as the long-form census data, and at the level of statistical reliability, there are some important differences between the ACS and the current long form.

The ACS will be a rolling survey with four three-month survey cycles each year: A skilled core staff will be available to do follow-up phone and in-person interviews with nonrespondents during each cycle, rather than the much larger temporary staff of enumerators the decennial census relies on. So, it is possible that ACS data will be higher quality with less nonsampling error. Imputation rates (the amount of data for items with no response that is imputed; imputation rates are discussed in the following section of this chapter) were compared for the 2000 Supplemental Survey and the Census 2000 long form. Preliminary comparison suggests that improved follow-up for the Supplemental Survey resulted in significantly lower imputation rates for key population items (gender, race, age, household relationship, and Hispanic origin) (Griffin and Obenski 2002, 20). No similar comparisons are yet available for nonresponses by households (as opposed to nonresponses on particular questions), but these may be higher because the decennial census allows enumerators to collect information from neighbors if a household cannot be contacted, while the ACS requires data be obtained from a household member (Griffin and Obenski 2002, 17).

The ACS will use a different definition of "place of residence": The decennial census defines "place of residence" as the place people live the majority of the year. Instead, the ACS will define

"place of residence" as the place people live when surveyed, as long as they have lived there, or plan to live there, for two months or more. This will affect local population counts during each decade. (The decennial census will continue to enumerate people using the traditional definition of "usual place of residence.") "Snowbirds" who spend winter in a southern location will now be counted at their winter home if surveyed during winter and at their summer home if surveyed during summer; the other home would be counted as vacant. College students, traditionally counted at the place they attend college rather than their parents' home, may instead be counted at their parents' home if surveyed over the summer months (U.S. Census Bureau 2003e).

This change is likely to improve the accuracy of long-form census data because it will now include seasonal residents, which should provide a more accurate picture of the "typical" population in a place. However, it will affect the comparability of several long-form census variables between 2000 and 2010. For instance, we may see sharp increases in the number of occupied manufactured homes and apartments in so-called "snowbird" destinations, fewer cost-burdened young adult renters in college towns, and so on. Using two different definitions of the population base—one in the decennial enumeration and one in the ACS—will raise new questions about the methods used to weight sample respondents. However, the change will improve our ability to understand regular seasonal migration and the impact this has on local demographics, economies, and housing markets.

The ACS will provide annual data on population characteristics: This should improve the population estimates the Census Bureau produces in years between censuses (discussed in chapter 3). More precise data on international and internal migration, fertility differences, housing characteristics, seasonal residence, and racial characteristics, has the potential to significantly improve population estimates. By mid-2003, three Supplemental Surveys had been completed (from 2000 to 2002) and data had been released for several metropolitan areas and counties with populations over 65,000. While the coverage is not complete, the data provides a useful update for those areas, although it is not broken down at smaller geographic levels.

Annual Public Use Microdata Sample (PUMS) data, one of the few census resources that provides data by individuals rather than by geographic units (described in detail in chapters 3 and 6), will be available based on the ACS for geographic areas of 100,000 people or more. Other questions, such as how base tables, derived reports, and analytic reports will be produced and disseminated, are still being discussed. The ACS Web site *(www.census.gov/acs/www)* will provide more information as it becomes available.

Improved timeliness will involve tradeoffs: The small increase in sampling error overall will affect smaller places and some data items more than others. Sampling error is the inaccuracy that results from using a sample to represent the characteristics of a population—in general, the smaller the sample as a proportion of the population, the larger the sampling error may be (this is discussed in greater detail in the following section of this chapter). Sampling error is typically measured by the coefficient of variance (CV). For places with fewer than 20,000 people (most census tracts, small cities and towns, and many rural counties), ACS estimates are anticipated

to have CVs approximately one third higher than the CVs for the decennial long form (Griffin and Obenski 2002, 29). Timeliness may offset some of the losses in accuracy. Overall, data quality will improve more for larger places (more than 65,000 people) than for smaller places because of timeliness and lower nonsampling error. Differences in the time period over which data is collected will raise some dilemmas. When establishing fair market rents, for instance, should the Department of Housing and Urban Development (HUD) use the most recent data for larger places and five-year averages for small places, which may bias the estimate of rents in small places downward? Or, should it ignore the most recent available data and use five-year averages for every place to ensure consistency (U.S. Department of Housing and Urban Development 2002a, vi)?

A second issue is that sampling error may increase more for some population groups than for others, because rates of response to mailback surveys differ dramatically among racial and ethnic groups. The ACS will only sample nonrespondents for follow-up at a rate of one-in-three, instead of following up on every nonrespondent as in the decennial census. This problem is being addressed in the redesign of the sampling procedure, based on experience with the Supplemental Surveys and the ACS test sites, to attain a more uniform sampling error across demographic groups (Griffin and Obenski 2002, 29).

A less tractable problem is raised by differences in sampling error for different data items. Some items are correlated—they have responses that we would expect to cluster within the same household (such as reported race, ancestry, or language spoken at home). Consequently, a sample will include fewer independent observations and have a higher sampling error for those items. For instance, if most Laotian households are not included in the sample, the number of Laotian individuals in that location will be underestimated by a wider margin than, for instance, the number of employed individuals, an item less likely to be correlated within households.

Data on housing vacancy rates suffers from a related problem. Because vacant units will only be identified from the one-in-three sample selected for in-person follow-up from addresses with no response, vacant units will be undersampled compared to occupied housing units and those items based on vacant units (asking price if rented, type of unit, and so on) will have a higher CV (Griffin and Obenski 2002, 28).

Another tradeoff is that the ACS data will not be released for areas smaller than census tracts. Block group data on gross rents, linguistic isolation, or people with disabilities, for instance, may be important for many local planning applications. This is especially the case in rural areas, where counties may contain only one tract, and block groups may provide a more appropriate delineation of service areas. Block and block group data will still be released for items in the enumeration, but analysts will lose an important level of spatial detail.

Sampling error

The population characteristics shown in some census data products do not necessarily match those of other data products. For instance, the total population of people 65 years and older

shown in Summary File 1 for Dane County, Wisconsin, may be different than the total for the same location from the Public Use Microdata Sample (described in more detail in chapters 3 and 6). Why are they different if they come from the same census? Census data products based on the long form are generated from samples and are not exact representations of the characteristics of the total population. The sample-based data is obtained from one-in-six housing units on average. In other words, a single household was providing answers for five other households. The accuracy of the population estimate depends in part on how typical the people within the sampled housing units are to those within the geographic area. A complex weighting scheme is used by the bureau to improve estimates of the population characteristics, but small discrepancies are unavoidable.

Differences between the 100 percent count data and the sample data arise for two reasons: sampling and nonsampling errors. They have different causes and can be controlled to a different extent. Sampling errors occur simply because not every household is being asked to respond. Sampled households may not be representative of all households. For example, if sampling is based on the "luck of the draw," then all the households within the sample may have incomes lower than the average for all households (the "true" mean). We use the concepts of sampling error to estimate these effects for any sample by calculating confidence intervals. This is explained in box 1.2.

Unlike sampling error, nonsampling error is the result of the data collection, processing, and reporting stages. Nonsampling error has two components: nonrandom nonsampling error and random nonsampling errors.

- Nonrandom errors bias the results consistently in a positive or negative direction. For example, if householders constantly understate their age (or overstate the value of their home), then the results of the sample will have a negative (or positive) nonsampling error.
- Random nonsampling errors do not bias the sample in one direction or the other but (in theory) cancel each other out. For example, if people round their commuting time to the nearest five-minute interval, the errors of the higher estimates will cancel the errors of the lower estimates within the sample. Random nonsampling errors increase the variability of the sampled responses. However, the increase in variability caused by random nonsampling errors can be mathematically estimated as part of the sampling error calculations.

The accuracy of the data: The address list, the undercount, and adjustments to the census

Since 1970, households with city-style street addresses have received their census questionnaire by mail. For this system of delivery to work comprehensively, the Census Bureau must have an up-to-date accurate address list for the entire country. For 2000, the quality of the address list depended on the Local Update of Census Addresses (LUCA) program. The bureau worked with local governments to update the street address database—cities and counties should have the best information because in almost all places they control both the road development and

Box 1.2 Sampling error, standard error, and confidence intervals

Imagine drawing all possible samples from the same population of households: the average income of each is likely to differ slightly from the "true" average income of the population, but the average income of all samples will be the same as the "true" average. Each sample would have a sample error: the measure of how far it deviates (or varies) from this "true" average (mean). Of course, we could only calculate this precisely if we knew the "true" mean, but we can estimate sample error based on the proportion of the total population in the sample. For small populations, we would need a larger proportionate sample to ensure the same level of sample error as we would get with a much smaller proportionate sample of a large population. This is why households in very small places are oversampled (at a rate of about one-in-two) compared to households in large places (which are sampled at a rate of one-in-eight). Sample error is zero when the "sample" includes the total population.

Standard errors are estimated based on sample errors. The standard error measures the variation among estimates of the "true" mean from all possible samples. It provides an estimate of how much any particular sample's mean is likely to differ from the "true" mean; in practical terms, it allows us to say things like: "the average income of households in place *x* is $35,000, plus or minus $2,000." The Census Bureau calculates unadjusted standard errors for each variable in each data product; these are shown in the technical documentation for each product. Together with sampling rates (the percent of the population in the sample) and design factors (the ratio of the estimated standard error to the standard error of a simple random sample), also provided in the technical documentation, standard errors can be calculated for each data item. A detailed example of the calculation is shown in chapter 6.

The main purpose of calculating standard errors is to construct a confidence interval around estimates—the "plus or minus $2,000" mentioned above. A confidence interval is the range within which we can expect the average value of a characteristic (calculated over all possible samples) to fall with a specific level of probability. Thus, if the above estimate of "plus or minus $2,000" was based on a 95 percent confidence interval, we could say with a 95 percent probability of confidence that the income range $33,000 and $37,000 includes the average estimate from all possible samples (in practical terms, the "true" mean). As the reader may remember from elementary statistics classes, standard errors are related to the "normal curve" (assuming the variable is normally distributed). For an interval ranging one standard error above and below the estimated (sample) mean, we could say with a 68 percent probability of confidence that the interval includes the average estimate from all possible samples; for an interval from 1.645 times the standard error above and below the sample mean, we could say this with 90 percent confidence; and for an interval of two standard errors on either side of the sample mean, we could say this with 95 percent confidence.

address designation processes. The U.S. Postal Service validated addresses and identified missing addresses. New addresses were then added to the TIGER/Line system, the central street address database (TIGER/Line is described in greater detail in chapter 2). In some cases, census enumerators also went door-to-door looking for living quarters not in the address file. As a result, more than four-fifths of households received a questionnaire by mail. Households living in areas that did not have city-style street addresses (this includes much of the rural United States) had their questionnaires delivered directly (U.S. Census Bureau 2000a).

The questionnaires were delivered on March 13–15, 2000, and follow-up postcards were sent out to remind those who had not completed their questionnaires to do so and thank those that had. Finally census enumerators telephoned or visited those households that did not complete the form. The final response rate on the short form was 67 percent, better than the 1990 Census's 65 percent and considerably better than the expected rate of 62 percent. Things were not quite so good for the long form—the expected response rate was 60 percent, the actual rate 54 percent (U.S. General Accounting Office 2000).

Just how accurate are the resulting numbers? Inaccuracy derives from a number of sources. Respondents may lie or guess, census workers may make clerical or computational errors (these fall into the category of nonsampling error), and potential methodological problems with the way the Census Bureau deals with incomplete and contradictory questionnaire responses exist. With data from the long form, there is also sampling error (explained above).

From a policy point of view, the "undercount"—the failure to count some of the population—is the major problem. The undercount is divided into

- the "sheer undercount," the failure to count people who live in the nation; and
- the "differential undercount," when some groups are undercounted more than others.

For instance, young black males may be undercounted more than young white males, middle-aged people may be undercounted less than people in their twenties, and renters may be undercounted more than owners. The worrying aspect is that some places have a higher proportion of those groups likely to be undercounted (compare Detroit, Michigan, which has a very large African-American population, with Des Moines, Iowa, which has a small African-American population). Since political representation and a considerable amount of federal money is distributed according to census population counts, cities with high proportions of people likely to be undercounted usually lobby the Census Bureau and Department of Commerce to have their populations adjusted upward (and bring legal suits when lobbying fails) (Cantwell, Hogan, and Styles 2003). Unfortunately, it is also true that even if a city knew that its population had been counted perfectly, it would still be in that city's interest to claim that its population had been undercounted.

City planners are often responsible for spearheading local responses to a potential or suspected undercount. Estimating the size of the undercount is entirely feasible—the methods to do this are well-established though clearly are not 100 percent accurate. Two different methods are used in the United States:

- demographic analysis
- special post-census surveys (in 1990 and 2000)

Demographic analysis uses records of births, deaths, migration, Medicare enrollment, estimates of legal but unrecorded emigration, and estimates of illegal immigration to estimate how many persons (by age, sex, and race) should have been counted in the census. The method is entirely separate from the census count itself, and thus provides an independent evaluation of the completeness of census coverage. To be useful, the method relies on accurate and complete administrative records. Unfortunately, emigration is poorly recorded in the United States, and illegal immigration has increased over the past few decades. Procedures have to be developed to fill these data holes before a demographic analysis can provide valid numbers (U.S. Census Bureau 2001).[6] In 2000, these procedures were the source of some concern.

Demographic analysis indicates that the undercount has declined significantly over the past sixty years, down from 5.4 percent in 1940 to 1.2 percent in 1980, up slightly in 1990 to 1.8 percent, followed by a massive drop in 2000 to 0.1 percent (see table 1.2). Historically, the undercount of African-Americans is much greater than that of non-African-Americans, and the undercount of males (white and African-American) much greater than that of females. The age pattern of the undercount for African-American males is particularly interesting. In 2000, African-American males 10 to 17 years of age were slightly overcounted (-1.9 percent undercount). But the undercount shoots up for 18- to 29-year-olds (5.7 percent) and is even greater for 30- to 49-year-olds (9.9 percent). It is adult African-American males under 50 who are most likely to be missed, although African-American households are being counted adequately (U.S. Census Bureau 2001).

In 1990 the post-census survey was called the Post-Enumeration Survey (or PES). PES used a dual-system estimator to calculate the number of persons missed by the census. This method is akin to the tagging (or "capture-recapture") techniques often used to estimate animal populations in the wild. Essentially, the method tries to locate the same individual twice. Statistical methods allow the success rate of these attempts to be converted into an estimate of the total population.

Table 1.2 Estimated net census undercount by race and sex, 1940–2000

	1940	1950	1960	1970	1980	1990	2000
Total	5.4	4.1	3.1	2.7	1.2	1.8	0.1
African-American, male	10.9	9.7	8.8	9.1	7.5	8.5	8.1
African-American, female	6.0	5.4	4.4	4.0	1.7	3.0	3.1
Non-African-American, male	5.2	3.8	2.9	2.7	1.5	2.0	1.6
Non-African-American, female	4.9	3.7	2.4	1.7	0.1	0.3	0.5

Source: For 1940 through 1990, see Robinson et al. (1993). For 2000, see U.S. Census Bureau (2002c, table 1). Note that the latter document also gives figures for demographic analysis of the 1990 Census; these are slightly different from those reported here. See also U.S. General Accounting Office (1998, table II.2).

[6]There are a few nations that measure population not using a census or sampling but using national population registers of births, deaths, marriages, divorces, moves, and so on. This requires considerable oversight of a nation's citizenry (Lavin 1996). Demographic analysis parallels this method.

The estimated 1990 undercount according to the PES was 1.6 percent, compared to 1.8 percent from the demographic analysis. Like demographic analysis, PES showed a very high undercount for non-Hispanic African-Americans (4.6 percent using PES). It also shows high undercount for Hispanic (5 percent using PES) and Native Americans living on reservations (12.2 percent using PES) (U.S. Census Bureau 2003b, table 1). The estimated undercount for non-Hispanic whites was 0.7 percent.

For Census 2000 two separate series based on a post-enumeration survey were released: the Accuracy and Coverage Evaluation (ACE) estimates. ACE Revision II provides the best estimates (U.S. Census Bureau 2003b). The first version of ACE estimates indicated a largish undercount (1.2 percent), but demographic analysis indicated a tiny undercount (the final revised estimate was 0.12 percent). After some investigation, it emerged that there were substantial methodological problems with ACE I. ACE Revision II tries to solve as many of these problems as is feasible. The final net undercount Revision II estimate for the total U.S. population is -0.5 percent; in other words, there was a slight overcount. The upshot is that the original census count was 281,421,906, but the final revised demographic estimate was 281,759,858. The estimate from ACE Revision I was over 284.5 million, but from Revision II (and with adjustment for correlation bias) it was a mere 280,090,250 (U.S. Census Bureau 2002c). These are huge swings in estimation (a range of 4.5 million people). Ignoring the estimates with obvious methodological flaws, the revised demographic analysis suggests the census undercounted 337,952 people, while ACE Revision II suggests that the census found 1,331,656 more than there really were (a range of about 1.7 million people).

Which estimate is right? How could there be such huge swings in the Post-Enumeration Survey? And, more to the point, how could the population be overcounted? There are few simple answers to these questions, and the reader is encouraged to read the detailed material put out by the Census Bureau on these issues (U.S. Census Bureau 2001; 2002c; 2003b). As a general matter though it is unclear which estimate (of the actual census count, the revised demographic analysis, and ACE Revision II) is correct. Fortunately, all three are, by historical standards anyway, fairly close. Keep in mind that the ACE Revision II estimate of a -0.49 percent undercount has a standard error of 0.2; so if we chose a 95 percent confidence interval it would mean that the true undercount estimate is likely to fall in the range of -.09 percent (-.49 + .2 + .2) to -.89 percent (-.49 - .2 - .2).

The simple fact is that counting the U.S. population is a dauntingly difficult task—there were methodological problems with the census count, with the demographic analysis, and with ACE. There is little reason to prefer one set of the three revised sets of results over another. Thus the Census Bureau decided not to adjust the final census count for 2000 up or down based on the results of demographic analysis or ACE. In 1990 there also was no adjustment, although at that time there were better methodological grounds for adjustment.

There were a number of reasons for the swings in the ACE estimates; an important one is the existence of "duplicates." For instance, households with teenagers away at college may indicate

Table 1.3 Net undercount for various demographic groups, 2000 and 1990 (percent)				
	2000 ACE Revision II		1990 PES	
	Estimate	Standard error	Estimate	Standard error
Total	-0.49	0.20	1.61	0.20
Race/origin:				
Non-Hispanic white	-1.31	0.20	0.68	0.22
Non-Hispanic African-American	1.84	0.43	4.57	0.55
Hispanic	0.71	0.44	4.99	0.82
Hawaiian and Pacific Islander	2.12	2.73	2.36	1.39
Non-Hispanic Asian	-0.75	0.68	2.36	1.39
Native American on reservation	-0.88	1.53	12.22	5.29
Native American off reservation	0.62	1.35	0.68	0.22
Tenure				
Owner	-1.25	0.20	0.04	0.21
Nonowner	1.14	0.36	4.51	0.43
Age and sex				
0–9	-0.46	0.33	3.18	0.29
10–17	-1.32	0.41	3.18	0.29
18–29 male	1.12	0.63	3.30	0.54
18–29 female	-1.39	0.52	2.83	0.47
30–49 male	2.01	0.25	1.89	0.32
30–49 female	-0.60	0.25	0.88	0.25
50+ male	-0.80	0.27	-0.59	0.34
50+ female	-2.53	0.27	-1.24	0.29

Souce: This table has been adapted from U.S. Census Bureau (2003b, table 1). Note: Negative numbers indicate an overcount.

those children still living at home, but those same teenagers may also be counted in college dormitories or apartments. In fact, 19-year-olds were one of the most overcounted age groups in Census 2000. In other cases, a vacant housing unit may be misclassified as occupied, and households and persons may then be "imputed" (a process by which missing records for households, families, and persons are created by the Census Bureau) to that housing unit. There were also problems in the LUCA street address program that may have resulted in the duplication of some housing units. The swing in the ACE and ACE Revision II estimates is in large part the result of better methodologies to control for duplicates in Revision II (U.S. Census Bureau 2002d).

Before moving on it is worth having a more detailed look at the undercount in 1990 and 2000 for some important demographic groups (see table 1.3). Notice that the undercount (positive or negative) was often statistically insignificant (for instance, the ACE Revision II estimates for Native Americans, and Hawaiians and Pacific Islanders). The differential in undercount

between owners and renters has worsened since 1990, as has the undercount of all adult males younger than 50. Unsurprisingly, the net undercount varies by state. Fortunately, in most states with a positive undercount, the difference between the census count and the ACE Revision II estimate was smaller than the ACE Revision II standard error. For instance, California has a census count of 33,871,648 persons but ACE Revision II estimated the population at 33,915,728. However, the standard error of that estimate was 87,146, indicating that the census count was not statistically different from the ACE estimate (U.S. Census Bureau 2003c, table 1).

Using the Census of Population and Housing

In this section we cover three important issues necessary to understanding and using the census: census geography, the tables and variables, and the publication schedule. Of these, census geography is the most complex.

Census geography and summary levels

The census is organized around geographical units for which data is summarized. For the vast majority of census publications, it is impossible to get information on individuals. Title 13 guarantees the confidentiality of answers to the census questionnaire:

- Employees of the Census Bureau must take an oath of confidentiality.
- There is security for completed census questionnaires.
- There are detailed disclosure-avoidance programs implemented by the bureau to ensure that tabulations do not allow the identification of specific persons or households.

Thus the data must be *aggregated,* in other words, summed or averaged over a particular geographical area. So we usually talk of data for a particular *summary* level, meaning data at a specified level of geographical aggregation. Some of these geographical regions or units exist independently of the census such as cities (towns, villages, and so on), counties or their equivalents, states (and similar areas such as the District of Columbia and outlying territories like the U.S. Virgin Islands and Guam), or the United States as a whole. These are *governmental units.* There are also *statistical units,* things like blocks, block groups, tracts, urbanized areas, metropolitan areas, census regions, and census divisions. Statistical units are created by the Census Bureau and do not exist as independent governmental units. Most people intuitively understand what a governmental unit is—it's the statistical units that cause trouble. The focus of the following discussion is on the statistical units.

The main geographical hierarchy[7]

Table 1.4 summarizes the main geographical hierarchy. In the discussion that follows, government and statistical units are printed in bold type when they are introduced.

Although census questionnaires are assembled by actual street addresses, the lowest level of tabulated census geography is the block. A **census block** is usually defined by roads, though it may also be bounded by rivers, streams, railroad tracks, invisible boundaries such as city or

[7]This discussion of census geography units is based explicitly on U.S. Census Bureau (2003a, appendix A, and 2003d, chapter four).

Table 1.4 The main census geography hierarchy

United States
Region
Division
State (includes District of Columbia, Puerto Rico, and four Island Areas[a])
County
County subdivision
Place
Census tract
Block group
Census block

Note: [a]American Samoa, Guam, the Commonwealth of the Northern Marina Islands, and the Virgin Islands of the United States.

county limits, or even imaginary extensions of roads. In a typical city with a grid road structure, four intersecting streets delimit the boundaries of a block. In fact it is the centerline (an imaginary line in the center of all public roads) of those four intersecting streets that define the block. Blocks typically have a population of about 85 people. In rural areas, blocks may contain an area of many square miles. For Census 2000, blocks were completely renumbered using four-digit codes, for instance "4000" (in 1990, blocks were identified using a three- or four-digit code, the first three digits being numbers, the last being an alphabetic suffix). The treatment of water areas was changed between 1990 and 2000. In 1990 all water areas in a block group (see below) were given a single block number ending in "99." In 2000, a water area completely in a land block was given the same number as the land block, but if the water area touches two or more blocks then the water area gets its own code.

Comparatively little census information is made available for blocks—only the information on the short form. One reason for this is the confidentially requirement of all census data. No data may be made public that would allow users to identify individuals, individual families, or households—the long-form sample data would, in some cases, allow individual households or persons to be identified at the block level. Another important reason is that long-form-derived data would have very large sampling errors at the block level.

Blocks are then assembled into **block groups.** A block group with the identification number "4" will include all blocks with numbers between "4000" and "4999." Obviously, blocks cannot cross block group boundaries. Most block groups were designated locally as part of the Census Bureau's Participant Statistical Areas Program. The Census Bureau undertook the designation only where state, local, or tribal authorities declined to, or where potential local participants could not be found. Block groups range from 600 to 3,000 people, but the bureau considers a population of 1,500 the optimum (1,000 for Native American reservations). Block groups (the lowest level for which sample data is available) are then assembled into **census tracts.** Block groups cannot cross tract boundaries. Tracts are delineated locally wherever possible. Tracts are meant to be fairly homogenous and permanent areas with similar demographic and economic characteristics and living conditions. In other words, they are intended to resemble something

approximating small neighborhoods. The area covered by tracts depends on population density. Population size usually ranges between 1,500 and 8,000 with an optimum size of 4,000 (2,500 in Native American reservations and island areas). Counties with fewer people than this have a single census tract (tracts do not cross county lines). In 1990, some counties had tracts and others had block numbering areas (BNAs). For 2000, all BNAs were replaced with tracts. Blocks, block groups, and census tracts are assigned FIPS codes that are unique within each county. When appended to the state and county FIPS codes, each has a unique number. FIPS codes are explained in more detail later in this section.

As places grow, roads are built, new blocks are created, and new block groups and tracts defined (and the opposite may be true in declining communities). Thus block, block group, and tract boundaries may change from census to census, making some historical comparisons in fast-growing areas tricky, to say the very least. From an analytic point of view, this is most often a problem at the tract level where over a decade there may be a considerable redefinition of tracts (this issue is dealt with in some detail in chapter 3).

Tracts are part of **counties** or **county equivalents,** and counties compose **states.** Each county has a unique five-digit FIPS code made up of a three-digit county code and a two-digit state code. A full listing of state and county FIPS codes is given in appendix A of U.S. Census Bureau (2003d). Few counties change boundaries between censuses, but enough do that users doing county analyses should consult the Census Bureau publication *Significant changes to counties and county equivalent entities: 1970–present* (*www.census.gov/geo/www/tiger/ctychng.html).* States almost never change their boundaries, and if they do it is almost always in ways that are insignificant from the point of view of census analysis.

There are many different types of "county equivalents." In Louisiana, **parishes** have the functions that counties do in most other states. Alaska has no counties; the statistically equivalent areas are **organized boroughs** (these are governmental units) and **census areas** (these are statistical units designated cooperatively by the Census Bureau and the state of Alaska). In Maryland, Missouri, Nevada, and Virginia there are **incorporated places** (cities) that are independent of any counties. For the purpose of the census, they perform the same functions and are treated as county equivalents. They are known as **independent cities.** Confusingly, some data tabulations on independent cities are given at the county summary level and some at the place summary level. The District of Columbia is treated as both a state-equivalent and a county-equivalent unit.

County subdivisions fall into three main categories: **minor civil divisions** (MCDs), **census county divisions** (CCDs), and **unorganized territories.** Twenty-eight states have MCDs; the legal status of MCDs varies considerably across those states.[8] In some states, MCDs can include places (cities), but in others all places are their own MCDs. In twelve states,[9] MCDs have much the same legal functions as incorporated places (cities)—in these states the Census Bureau tabulations for places includes MCDs. In ten states with MCDs, portions of some counties are not covered by MCDs—these are called **unorganized territories.** States without MCDs have CCDs;

[8]They are variously known within states as American Indian Reservations, assessment districts, boroughs, charter townships, election districts, election precincts, gores, grants, locations, magisterial districts, parish governing authority districts, plantations, precincts, purchases, road districts, supervisor's districts, towns, or townships.

[9]Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Wisconsin.

unlike MCDs (which are governmental units of census geography), CCDs are statistical units determined by the Census Bureau in cooperation with state and local governments. Finally, in Alaska, **census subareas** are statistical divisions of boroughs and census areas.

States are then organized into **divisions,** and divisions into **regions.** These are defined in table 1.5.

Table 1.5 Regions and divisions

Region	Division	Constituent states
Northeast	New England	Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
	Middle Atlantic	New York, New Jersey, Pennsylvania
Midwest	East North Central	Ohio, Indiana, Illinois, Michigan, Wisconsin
	West North Central	Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas
South	South Atlantic	Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
	East South Central	Kentucky, Tennessee, Alabama, Mississippi
	West South Central	Arkansas, Louisiana, Oklahoma, Texas
West	Mountain	Montana, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Idaho
	Pacific	Washington, Oregon, California, Alaska, Hawaii

The main geographical unit hierarchy and the associated FIPS number system

Each state has a two-digit code number—technically its FIPS (Federal Information Processing Standards) code.[10] Iowa, for instance, has the FIPS code "19," Illinois the code "17." Each county in each state will have a three-digit county code. Our home county is Johnson County, Iowa. Its code is "103." Thus Johnson County, Iowa, is uniquely defined by the FIPS code "19103." In Illinois the county code "103" refers to Lee County. To uniquely identify Lee County, the five-digit state and county FIPS code must be used: "17103."

Each tract in the United States is uniquely identified by the state and county codes plus a further six-digit tract code, the last two digits working as implied decimal places. The first tract of Dane County, Wisconsin, has the code "000100" (sometimes the code will be given as "1.00") (see figure 1.1). Thus the full identification of this tract is "55025000100." The most northeasterly tract of Dane County is "011800" (or "118.00"). Its full code is "55025011800." A tract that has been divided up into multiple tracts over time will usually have numbers in the decimal place positions of the tract code. This is the case for tracts "120.01" and "120.02" in Dane County ("55025012001" and "55025012002").

Individual tracts are then divided into block groups and block groups into blocks. The most northeasterly tract in Dane County, "55025011800," has three block groups, "1," "2," and "3." This provides the full block group code: "550250118001," "550250118002," and

[10]FIPS codes are a standardized numeric or alphabetic code issued by the National Institute of Standards and Technology (NIST).

"550250118003." The most northeasterly of these three is "550250118002." This block group consists of sixty-four blocks, the lowest level of census geography. Each block is identified with an additional three numbers. Thus the first block is "550250118002000" and so on. As we noted before, the unit identification system is hierarchical—we start with a state and then move down from there.

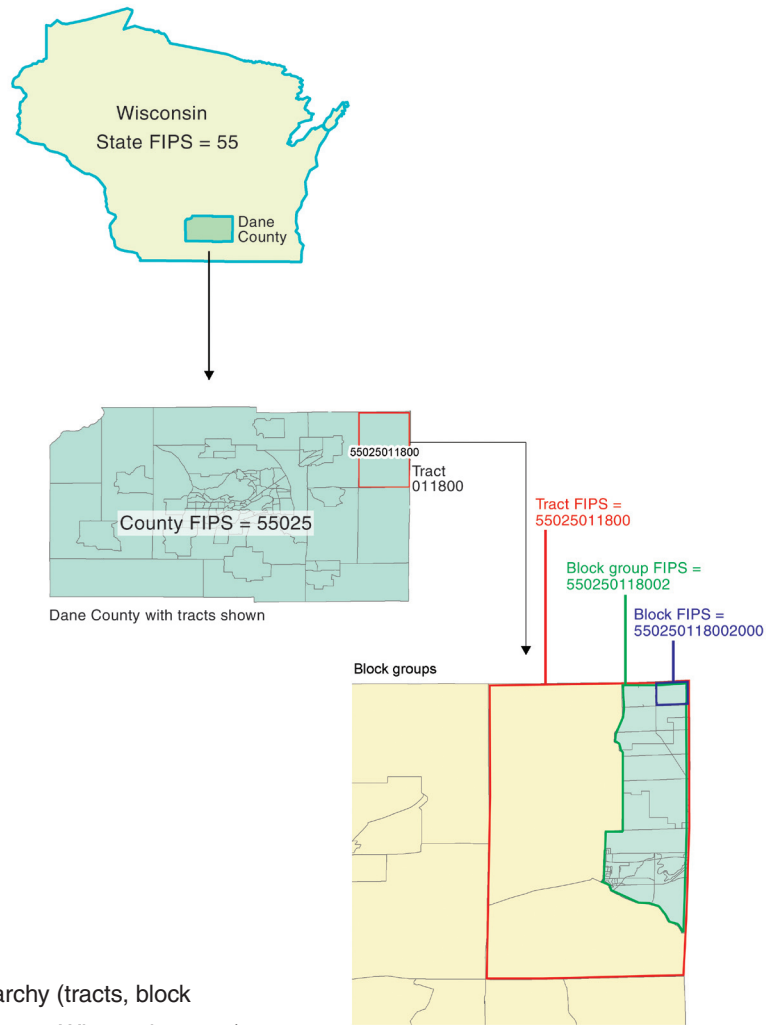


Figure 1.1 FIPS code hierarchy (tracts, block groups, and blocks, Dane County, Wisconsin, 2000).

The basic census hierarchy for Puerto Rico and U.S. indigenous people

Unfortunately, there are a vast number of other units of census geography beyond the basic hierarchy described above. Puerto Rico and the census units for U.S. indigenous people pose some special problems.

- **American Indian areas:** Here the hierarchy goes **tribal block group → tribal census tract → tribal subdivision → American Indian reservation (federal)** or **off-reservation trust land** or **Oklahoma tribal statistical area** (OTSA). In some instances there is an alternative hierarchy: tribal block group → tribal census tract → **tribal designated statistical area** (TDSA) or **American Indian reservation (state)** or **state designated American Indian statistical area** (SDAISA).
- American Indian Reservations are lands where the federal government has—by treaty, statute, or court order—recognized that Native American tribes have primary governmental authority. Common names for reservations are: colonies, communities, pueblos, rancherias, ranches, reservations, reserves, tribal towns, and tribal villages. Reservations may cross county and state lines.
- TDSAs are statistical units for recognized tribes that do not have a reservation or off-reservation trust land.
- OTSAs are statistical units for tribes in Oklahoma that formerly had a reservation in the state but no longer do.
- A state reservation is land held in trust by a state for a particular tribe.
- SDAISAs are statistical entities for state-recognized Native American tribes that do not have a state recognized reservation.
- **Alaska native areas:** Here the hierarchy goes block group → tract → **Alaska native village statistical area** (ANVSA) → **Alaska native regional corporation** (ANRC). Twelve ANRCs cover all of Alaska, except for the Annette Island Reserve, which is a Native American reservation. A thirteenth corporation covers Alaska natives not living in Alaska and not identifying with any of the other twelve ANRCs. ANVSAs are the settled portions of Alaska native villages—this unit aids in the presentation of data but has little meaning on the ground since native villages often have no determinate boundary.
- **Hawaiian Home Lands:** Here the hierarchy goes block group → tract → **Hawaiian Home Lands** (HHL). This is a new unit for Census 2000 and includes land held in trust for native Hawaiians by the state of Hawaii.
- In Puerto Rico, the **municipio** is treated as the equivalent of a county in the United States. The municipio is then divided into **barrios** or **barrio-pueblos,** and barrio-pueblos and some barrios are divided into **subbarios.** Puerto Rico has no incorporated places—instead the Census Bureau designates the **zona urbana,** essentially the governmental center of each municipio, and **comunidades,** or other urban areas.

The units described in this section are, like all units of the census hierarchy, composed of blocks.

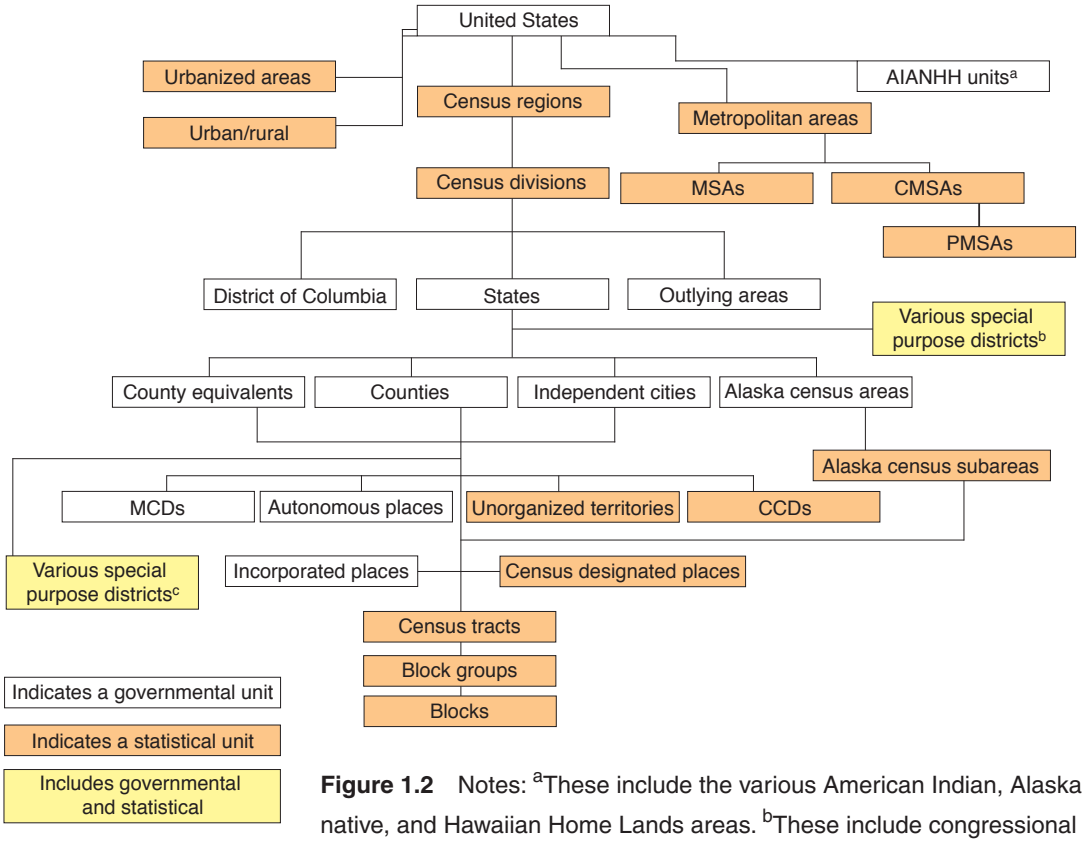


Figure 1.2 Notes: [a]These include the various American Indian, Alaska native, and Hawaiian Home Lands areas. [b]These include congressional districts, state legislative districts, school districts, urban growth areas, and Alaska native regional corporations. [c]These include voting districts and Traffic Analysis Zones.

The extended census geography hierarchy

The main hierarchy does not cover a wide range of geographical units crucial to the work of urban land professionals. Figure 1.2 summarizes the most important elements in the extended hierarchy. Note again that blocks are the "building blocks" of all higher levels of census geography.

Cities, towns, villages, boroughs, etc.: These are all **places.** Places have a special assigned five-digit FIPS code based on the alphabetical order of the place's name within a state. There are three main categories of place: **incorporated places, census designated places** (CDP), and **consolidated cities.** None of these crosses state lines, but they may cross county lines.

Incorporated places are cities, towns, and so on that have legal existence according to the laws of their states. This usually means they have defined governmental functions. However, there are various naming complications across states. Towns in New England, New York, and Wisconsin, and boroughs in New York are treated as MCDs; they are not treated as incorporated places. As we indicated earlier, boroughs in Alaska are treated as county equivalents. Maryland, Missouri,

Nevada, and Virginia have independent cities that are not part of any county; nevertheless these are treated by the Census Bureau as counties (and also as county subdivisions and places).

CDPs are statistical units (they have no legal status outside of the census) for built-up areas with dense population settlement that have not incorporated (in other words, are not recognized as a city, town, or village by state law). CDPs are designated by the Census Bureau in cooperation with local officials. Note that for 2000 there were no minimum size thresholds for CDPs (though there were in previous censuses). From 1950 through 1970, CDPs were called "unincorporated."

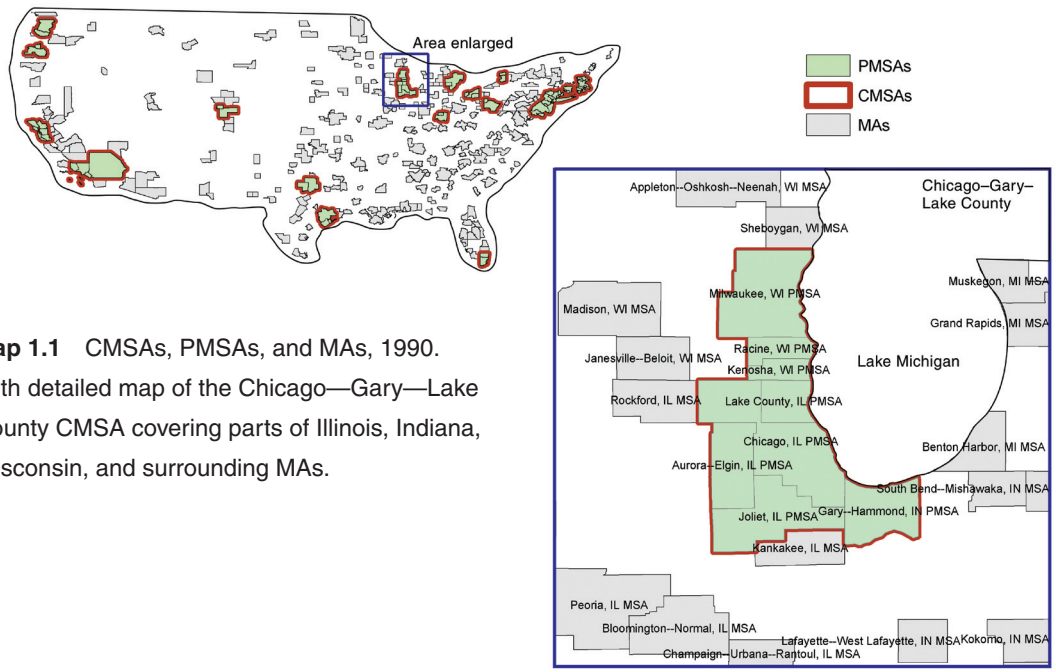
Consolidated cities are places where the legal functions of an incorporated place (city, town, and so on) have merged with its county or MCD but where the county or MCD continues to have a separate legal status. In the data hierarchy, data for a consolidated city will be shown at the county or MCD level (depending on the nature of the consolidation). In some consolidated cities there may be semi-independent places. In some tabulations, data for consolidated cities is not provided. Rather, each semi-independent place will have its own data record. Data for what is known as the **consolidated city (balance)** will also be provided; the balance will be the numbers for the entire consolidated city minus the numbers for its semi-independent places.

Metropolitan areas: A **metropolitan area** (MA) is a county or set of counties with a population over 100,000 (75,000 in New England), and a central city population of at least 50,000. Outlying counties are included in a particular MA if enough workers in the adjacent counties commute to the center county for work—in other words, the outlying counties should be functionally integrated with the central county. The outlying counties must also meet additional criteria of population density, urban population, and population growth. The population of the MA is then divided into those **living inside the central city** and those **outside the central city.**

MAs are then categorized into three further units: **metropolitan statistical areas** (MSAs), **primary metropolitan statistical areas** (PMSAs), or **consolidated metropolitan statistical areas** (CMSAs). Think of an MSA as an independent MA, in other words, an MA not functionally linked to any other MA (although two independent MSAs may still be contiguous). CMSAs are interconnected groups of MAs, akin to a conurbation. The individual MAs in a CMSA are then called PMSAs, not MSAs.

Note that in New England, MAs consist of sets of cities and county subdivisions rather than counties. However, New England also has an alternative county-based definition of MSAs called **New England county metropolitan areas** (NECMAs). The main frame of map 1.1 shows all MSAs, CMSAs, and PMSAs in the United States and the smaller detailed frame shows those (with names) for parts of Illinois, Indiana, and Wisconsin. Table 1.6 lists all CMSAs as of 1990. There is a four-digit FIPS code to identify each MA and in the case of a PMSA a further code to identify the CMSA to which it belongs.

Urban/rural and urbanized areas: People and housing units in **urbanized areas** (UA) and **urban clusters** (UC) are classified **urban.** Rural areas include people and territory outside of UAs and UCs. UAs and UCs are densely settled areas defined as sets of



Map 1.1 CMSAs, PMSAs, and MAs, 1990. With detailed map of the Chicago—Gary—Lake County CMSA covering parts of Illinois, Indiana, Wisconsin, and surrounding MAs.

- blocks or block groups with at least 1,000 persons per square mile
- surrounding blocks or block groups with 500 persons per square mile
- less densely settled blocks or block groups that nevertheless form connections across more densely populated blocks and block groups

A place, an MCD, a county and, most confusingly to new users of the census, a metropolitan area, may include both urban and rural areas. Technically a UC is a densely settled territory with a population between 2,500 and 50,000 while a UA has more than 50,000 people. Practically then, "urban" will include densely settled places with a population greater than 2,500 and rural will include (besides agricultural areas and open country) places—cities, towns, villages, and so on—with less than 2,500 people.

The definition of a UA changed markedly between 1990 and 2000. A UA or UC may contain more than one place. The dominant place in a UA is called the **urban area central place.** A place may also be partly within and partly outside an urban area, in which case it is referred to as an **extended place.**

Purpose-defined districts: Census data is tabulated for several other kinds of districts:

- **School districts:** Each school district has a five-digit code unique within a state. Three types of school district are recognized: elementary, secondary, and unified.
- **State legislative districts:** These are the districts represented in the upper and lower houses of state assemblies.
- **Voting districts:** This covers the election districts, precincts, wards, and so on used by state, local, and tribal government for elections.

Table 1.6 CMSAs in 1990	
Boston–Lawrence–Salem, MA–NH	Los Angeles–Anaheim–Riverside, CA
Buffalo–Niagara Falls, NY	Miami–Fort Lauderdale, FL
Chicago–Gary–Lake County, IL–IN–WI	Milwaukee–Racine, WI
Cincinnati–Hamilton, OH–KY–IN	New York–Northern New Jersey–Long Island, NY–NJ–CT
Cleveland–Akron–Lorain, OH	Philadelphia–Wilmington–Trenton, PA–NJ-DE–MD
Dallas–Fort Worth, TX	Pittsburgh–Beaver Valley, PA
Denver–Boulder, CO	Portland–Vancouver, OR–WA
Detroit–Ann Arbor, MI	Providence–Pawtucket–Fall River, RI–MA
Hartford–New Britain–Middletown, CT	San Francisco–Oakland–San Jose, CA
Houston–Galveston–Brazoria, TX	Seattle–Tacoma, WA

- **ZIP Code tabulation areas** (ZCTA™)**:** These are areas that approximate the areas covered by the U.S. Postal Service's five-digit or three-digit ZIP Code. ZCTA-based tabulations replace ZIP Code tabulation provided in the 1990 and earlier censuses.
- **Traffic Analysis Zone** (TAZ)**:** These are areas created by local transportation officials (those in metropolitan planning organizations, or MPOs) and are used for specialized tabulations of journey-to-work, place-of-work, and traffic-flow data. The data is published as part of the Census Transportation Planning Package (CTPP). Each TAZ has a six-character code that uniquely identifies it within a county. TAZs and the CTPP are described more fully in chapter 6.

Beyond these there are a number of other specialized areas that are either limited to particular areas of the country (such as **urban growth areas** in Oregon) or are associated with particular data sources (such as **Public Use Microdata Areas,** or PUMAs). Some of these will be discussed in the more detailed chapters that follow.

Summary level and tabulated data

Almost all census data is aggregated to a particular geographical scale, the "summary level," each with its own code. If we wanted data summed to the state level, we would need summary level "040" data. County level data is "050." Tract data is "140." Table 1.7 provides the most important summary level codes for urban land professionals.

Notice the difference between level "080" and level "140." If we wanted tract data for Dane County, Wisconsin, and we decided to download using the "140" summary level, we would get data on all tracts in the county. If we selected multiple counties, or even multiple states, then we would get tract information for all counties and states selected, organized by county. But if we downloaded using "080," the data would be organized differently. Tracts would be organized into the places in which they exist and those places would be organized into the county subdivisions

Table 1.7 Selected summary level codes	
Summary level code	**Geographical element**
010	United states
020	Region
030	Division
040	State
050	State-county
060	State-county-county subdivision
070	State-county-county subdivision-place/remainder
080	State-county-county subdivision-place/remainder-census tract
090	State-county-county subdivision-place/remainder-census tract-urban/ rural-block group
140	State-county-census tract
150	State-county-census tract-block group
160	State-place
170	State-consolidated city
390	State-MSA/CMSA
391	State-MSA/CMSA-central city
850	Five-digit ZCTA
Note: These are only the most commonly used summary level codes. See U.S. Census Bureau (2003a, chapter 4) for a full set of codes.	

in which they exist. Tracts not in places would be categorized into the remainder (the nonplace part) of the county subdivision.

For those who plan to download and use the raw data tables provided by the Census Bureau (see chapter 2 for details), it is absolutely crucial to become familiar with the summary level code system. The codes will be necessary to extract information from the tables.

Data variables and data tables

The census is distributed as a series of data tables (sometimes called matrices) and associated data variables. Each summary file (in fact, each data product) has technical documentation in which tables and variables are described. For instance, chapter 5 of Summary File 3's technical documentation describes the available tables, and chapter 7 of the documentation lists the variables and their constituent data dictionary reference names (U.S. Census Bureau 2003a).

Counts for each of these subvariables are shown for a particular summary level. To find out the percentage of African-Americans in a particular geographic area, variable "P006003" would need to be divided by variable "P006001."

"Table P6: Race," part of the basic population tables based on the population (not a sample), is made up of eight subvariables, each with its own data dictionary reference name:

Total	P006001
White alone	P006002
Black or African-American alone	P006003
American Indian and Alaska Native alone	P006004
Asian alone	P006005
Native Hawaiian and other Pacific Islander alone	P006006
Some other race alone	P006007
Two races or more	P006008

"Table H79: Aggregate value (dollars) for all owner-occupied housing units by units in structure" shows not a count for each subvariable, but the total value of all homes in each category. It has eight constituent subvariables:

Aggregate value	H079001
1, detached	H079002
1, attached	H079003
2	H079004
3 or 4	H079005
5 or more	H079006
Mobile home	H079007
Boat, RV, van, etc.	H079008

To obtain the average value of mobile homes, for instance, the aggregate value for H079007 would be divided by the number of owner-occupied mobile homes (from "Table H32: Tenure by Units in Structure").

The system of tables and variables may seem very complex to new users of the census, but with a little practice the system is quite straightforward. Following two simple principles will simplify the process:

- First, choose a table making sure you understand the universe from which the table is drawn and the levels of geographic aggregation at which the table is available (for Summary File 3 this information will be provided in U.S. Census Bureau 2003a, chapter 5).
- Find out the constituent data dictionary reference names of the constituent variables (for Summary File 3 this will be provided in U.S. Census Bureau 2003a, chapter 7).

Census data products

The Census Bureau produces a number of standard data products commonly used by urban land professionals. These are listed in table 1.8. This data is available on computer tapes, CD–ROM, and on the Web. Acquiring census data for analysis in a GIS is discussed in detail in chapter 2.

Table 1.8 Major census data tabulations	
Name of product	**Variables, lowest level of geography**
100 percent data	
Redistricting Data Summary File	Population counts, blocks
Demographic Profile	Selected population and housing characteristics, tracts
Congressional District Demographic Profile	Selected population and housing characteristics, congressional districts
Summary File 1 (SF1)	Counts and cross tabulations on short-form questions, blocks
Summary File 2 (SF2)	Similar to SF1, but with detailed breakdowns by race, Hispanic origin, and American Indian, Alaska Native tribes, tracts
Sample data	
Demographic Profile	Selected population and housing characteristics, tracts
Congressional District Demographic Profile	Selected population and housing characteristics, congressional districts
Summary File 3 (SF3)	Social, economic, and housing characteristics, block group/tracts
Summary File 4 (SF4)	Similar to SF3, but with detailed breakdowns by race, Hispanic origin, and American Indian, Alaska Native tribes, tracts
Public Use Microdata Samples (PUMS)	Raw long-form data but with confidentiality screening. One percent sample for the nation, states, and some substate areas, 5 percent sample for state and substate areas. One percent Super Public Use Microdata Areas (SuperPUMAs), 5 percent PUMAs

There are also "special reports" put out by the Census Bureau (such as *Demographic trends in the 20th century* or *Racial and ethnic residential segregation in the United States: 1980–2000*) and various shorter "census briefs" (such as *The black population: 2000,* or *The 65 years and over population, 2000).* The advantage of these publications is that analysis has already been performed on the raw census numbers. You do not have to do the work yourself. However, these publications usually present data at a broad geographical scale, not the scale most relevant to the needs of urban analysts. Box 1.3 lists the full set of briefs and special reports available at the time of writing.

The Census Bureau also produces various specialized data publications that combine information from the Census of Population and Housing, the Economic Census, and various other sources. These include the following:

- *Census Transportation Planning Package (CTPP).* The Census 2000 version of the CTPP was not ready at the time of writing (the Census Bureau was experiencing difficulty compiling that data), but the 1990 version was available. The CTPP is used in an economic example

Box 1.3 Census 2000 briefs and special reports prepared by the Census Bureau

At the time of publication, the following briefs and special reports had been published:

Briefs:

1. *Overview of race and Hispanic origin* (C2KBR/01-1)
2. *Population change and distribution: 1990–2000* (C2KBR/01-2)
3. *The Hispanic population* (C2KBR/01-3)
4. *Race and Hispanic or Latino origin by age and sex for the United States: 2000* (PHC-T-8)
5. *The white population: 2000* (C2KBR/01-4)
6. *The black population: 2000* (C2KBR/01-5)
7. *The two or more races population: 2000* (C2KBR/01-6)
8. *Congressional apportionment* (C2KBR/01-7)
9. *Households and families: 2000* (C2KBR/01-8)
10. *Multigenerational households for the United States, states, and for Puerto Rico: 2000* (PHC-T-17)
11. *Gender: 2000* (C2KBR/01-9)
12. *Male-female ratio by race alone or in combination and Hispanic or Latino origin in the United States: 2000* (PHC-T-11)
13. *The 65 years and over population: 2000* (C2KBR/01-10)
14. *Population and ranking tables of the older population for the United States, states, Puerto Rico, places of 100,000 or more population, and counties* (PHC-T-13)
15. *The United States in international context: 2000* (C2KBR/01-11)
16. *Age: 2000* (C2KBR/01-12)
17. *Housing characteristics: 2000* (C2KBR/01-13)
18. *The Native Hawaiian and other Pacific Islander population: 2000* (C2KBR/01-14)
19. *The American Indian and Alaska Native population: 2000* (C2KBR/01-15)
20. *American Indian and Alaska Native tribes for the United States, regions, divisions, and states* (PHC-T-18)
21. *The Asian population: 2000* (C2KBR/01-16)
22. *Disability status: 2000* (C2KBR-17)
23. *Employment status: 2000* (C2KBR-18)
24. *Employment status of the population in households for the United States, states, counties, places, and for Puerto Rico: 2000* (PHC-T-28)
25. *Poverty: 1999* (C2KBR-19)
26. *Home values: 2000* (C2KBR-20)
27. *Housing costs of renters: 2000* (C2KBR-21)
28. *Veterans: 2000* (C2KBR-22)
29. *The Arab population: 2000* (C2KBR-23)
30. *Educational attainment: 2000* (C2KBR-24)
31. *Occupations: 2000* (C2KBR-25)

Box 1.3 Census 2000 briefs and special reports prepared by the Census Bureau (continued)

32. *School enrollment: 2000* (C2KBR-26)
33. *Housing costs of homeowners: 2000* (C2KBR-27)
34. *Geographical mobility: 1995 to 2000* (C2KBR-28)
35. *Language use and English-speaking ability: 2000* (C2KBR-29)
36. *Summary tables on language use and English ability: 2000* (PHC-T-20)
37. *Marital status: 2000* (C2KBR-30)
38. *Marital status for the population 15 years and over for the United States, regions, states, Puerto Rico and metropolitan areas: 2000* (PHC-T-27)
39. *Grandparents living with grandchildren: 2000* (C2KBR-31)
40. *Structural and occupancy characteristics of housing: 2000* (C2KBR-32)

Special reports:

1. *Mapping Census 2000: The geography of U.S. diversity* (CENSR/01-1)
2. *Emergency and transitional shelter population: 2000* (CENSR/01-2)
3. *Population in emergency and transitional shelters* (PHC-T-12)
4. *Racial and ethnic residential segregation in the United States: 1980–2000* (CENSR-3)
5. *Demographic trends in the 20th century* (CENSR-4)
6. *Married-couple and unmarried-partner households: 2000* (CENSR-5)
7. *Hispanic origin and race of coupled households* (PHC-T-19)
8. *Adopted children and stepchildren: 2000* (CENSR-6RV) and (PHC-T-21)
9. *Domestic migration across regions, divisions, and states: 1995 to 2000* (CENSR-7)
10. *State-to-state migration flows: 1995 to 2000* (CENSR-8)
11. *Migration for the population 5 years and over for the United States, regions, states, counties, New England minor civil divisions, metropolitan areas, and Puerto Rico: 2000* (PHC-T-22)
12. *Migration and geographic mobility in metropolitan and nonmetropolitan America: 1995 to 2000* (CENSR-9)
13 *Internal migration of the older population: 1995 to 2000* (CENSR-10)
14. *Migration by sex and age for the population 5 years and over for the United States, regions, states, and Puerto Rico: 2000* (PHC-T-23)
15. *Migration of natives and the foreign born: 1995 to 2000* (CENSR-11)
16. *Migration by nativity for the population 5 years and over for the United States and states: 2000* (PHC-T-24)
17. *Migration of the young, single, and college educated: 1995 to 2000* (CENSR-12)
18. *Migration by race and Hispanic origin: 1995 to 2000* (CENSR-13)
19. *Migration by race and Hispanic origin for the population 5 years and over for the United States, regions, states, and Puerto Rico: 2000* (PHC-T-25)

All of these are available at *landview.census.gov/population/www/cen2000/briefs.html*

in chapter 4 and is discussed in a transportation context in chapter 6. The CTPP data is distributed on CD–ROM and also on the Web.

- *Statistical abstract of the United States.* This publication is produced annually and includes data from the Census of Population and Housing, various updates to that census, the Economic Census, and various other official data sources. The data is best at the national and state levels. It is produced as a paper publication, as a CD–ROM, and is available on the Web.
- *State and metropolitan area data book.* This is also produced annually and includes data from the Census of Population and Housing, the Economic Census, and various surveys and updates. It is produced annually and is available on CD–ROM and the Web.
- *City and county data book.* This is produced annually and includes data from the Census of Population and Housing, the Economic Census, and various surveys and updates. It is produced annually and is available on CD–ROM and the Web. Data is provided for larger cities and counties.
- *TIGER/Line system.* This is the Census Bureau's main mapping database. It is described in detail in chapter 2.

Almost all of this data can be downloaded from the Web (there are complete instructions in chapter 2), or users may order CD–ROMs or DVDs directly from the Census Bureau. Depository libraries will usually have all the necessary information in electronic format.

Organization of the rest of the book

Chapter 2 is a hands-on discussion of downloading Census of Population and Housing data, and the ways that data can be brought into a GIS—it assumes almost no knowledge of the census beyond topics covered in this chapter. It also assumes very limited knowledge of GIS. The chapter is considerably more introductory than the ones that follow. Before more complex census analyses can be attempted, users should become familiar with the basic methods of data acquisition.

The four chapters following chapter 2 focus on particular sets of data. In each of these chapters we describe the relevant variables included in the decennial and other censuses, consider supplementary sources that can be used to extend or update analyses, and develop a few practical examples of analyses using census data and sometimes other sources. The discussion of particular examples is intended to clarify general principles of data access and use. Examples were chosen to illustrate typical methods of dealing with the major challenges census-based spatial analyses pose.

Chapter 3 looks at demographic and social data, perhaps the most commonly used census data. Chapter 4 examines the economic and occupation data in the Census of Population and Housing and also describes the Economic Census and various related economic data sources. Chapter 5 covers housing and community development data. Chapter 6 focuses on travel and commuting data, particularly the Public Use Microdata Sample (PUMS) and the Census Transportation Planning Package (CTPP).

Chapter 7 considers distribution strategies, particularly using the Web as a way of providing clients and the public easy access to information.