

## **Making Sense of the Census, Part I: Introduction to the U.S. Census of Population and Housing**

Thomas H. Grayson  
22 January 2002

### **Overview**

What is it and why do we care?  
How the data are collected?  
What data are available?  
Introduction to Census geography and summary levels  
A Quick Look at the Census documentation  
A Quick Look at some sample data

### **What Is It and Why Do We Care?**

Mandated by the [Constitution of the United States](#)  
The modern census of population and housing was established in 1940 with the incorporation of the housing component and the introduction of sampling techniques for the long form  
Conducted every ten years  
Attempts an actual count of population categorized by various criteria  
The only source for demographic data with a wide geographic scope  
The most reliable and detailed information for describing local areas: neighborhoods, cities, counties  
The most consistent source of time series demographic data available  
U.S. Congressional representatives are apportioned based on census counts. Federal dollars for schools, employment services, highway assistance, housing construction, hospital services, programs for the elderly, etc. are all distributed based on census figures.

### **A Quick Look**

View some Census data at [American FactFinder](#).

### **How the Data Are Collected**

Collected from households through a mail survey conducted every decade  
For the **2000 Census** more than 285,000 census takers and support personnel accounted for the 118 million households and 275 million persons in the United States.

- [2000 Census Home Page](#)

Two different census questionnaires were distributed:

- [short-form questionnaire](#) contains questions asked of everyone (summarized in Summary Tape File 1 (STF 1) for 1980 and 1990, Summary File (SF 1) for 2000)
- [long-form questionnaire](#) contains questions asked of a population sample (1/6 households) (summarized in Summary Tape File 3 (STF 3) for 1980 and 1990, Summary File 3 (SF 3) for 2000)

The long form is being replaced in the 2010 Census by the [American Community Survey](#). This program will survey homes every month and provide updated statistics every year instead of every ten years. The program begins in 2003.

### **What's Included: Information on Population, Employment and Housing Characteristics**

**Short Form: 100% Count (STF 1/SF 1)**

<b>Population Characteristics</b>	<b>Housing Characteristics</b>
<i>Age</i>	<i>Tenure</i>
<i>Gender</i>	<i>Value or Contract Rent</i>
<i>Race</i>	<i>Vacancy Status</i>
<i>Hispanic Origin</i>	<i>Number of Rooms</i>
<i>Marital Status</i>	<i>Units in Structure</i>
<i>Household Type</i>	<i>Congregate Housing</i>
<i>Household Relationship</i>	

[Sample Short Form from 2000 Census](#)

**Long Form: Sample Counts (STF 3/SF 3)**

<b>Population Characteristics</b>	<b>Housing Characteristics</b>
<i>Social Characteristics</i>	<i>Age of Housing</i>
<i>Education</i>	<i>Heating Fuel</i>
<i>Citizenship</i>	<i>Facilities</i>
<i>Ancestry</i>	<i>Vehicles</i>
<i>Language</i>	<i>Mortgage Status</i>
<i>Disability</i>	
<i>Children</i>	
<i>Place of Birth</i>	
<i>Economic Characteristics</i>	
<i>Income</i>	
<i>Labor Force Status</i>	
<i>Employment</i>	
<i>Place of Work</i>	
<i>Public Assistance</i>	
<i>Retirement Income</i>	

[Sample Long Form from 2000 Census](#)

**Why We Need to Know the Two Components**

Accuracy of the data varies and counts differ (Why?)

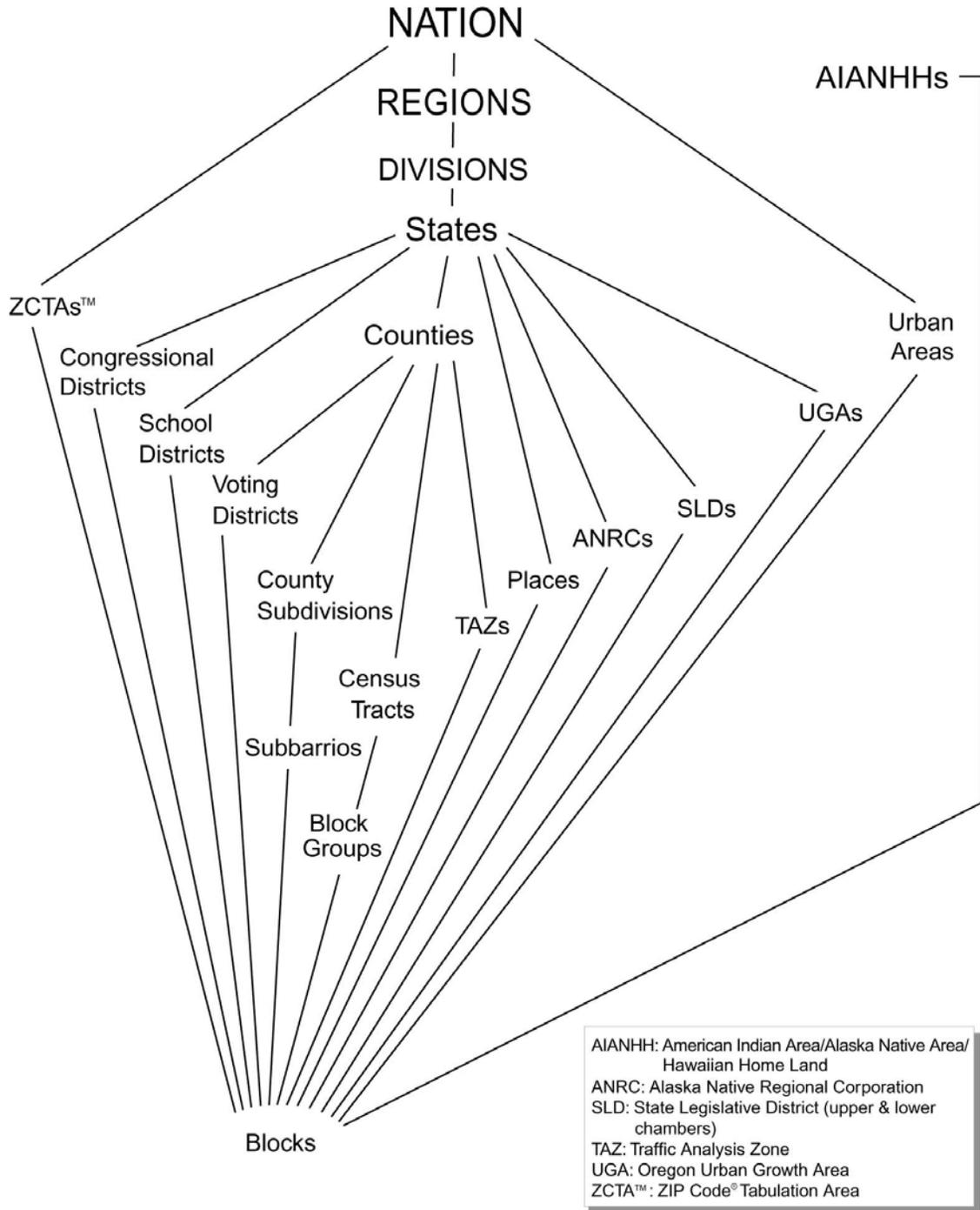
It helps us to understand how the data are organized in Summary Tape Files (STFs)

**Census Geography and Summary Levels**

**The Census organizes and aggregates data into a series of geographic hierarchies**

**Overview**

**Standard Hierarchy of Census Geographic Entities (from *Census 2000 Summary File 1 Technical*)**



Summary Level	Geographic Unit
010	United States
020	Region: Northeast (NE), Midwest (MW), South (S) and West (W) Regions

030	Division: Northeast: New England, Mid Atlantic Midwest: East North Central, West North Central South: South Atlantic, East South Central, West South Central West: Mountain, Pacific
040	State (includes Washington, D.C. & Puerto Rico)
050	County
060	County Subdivision
070	Place
080	Census Tract / Block Numbering Area (average 4,000 persons)
090	Block Group (average 1,000 persons)
100	Block (average 85 persons)

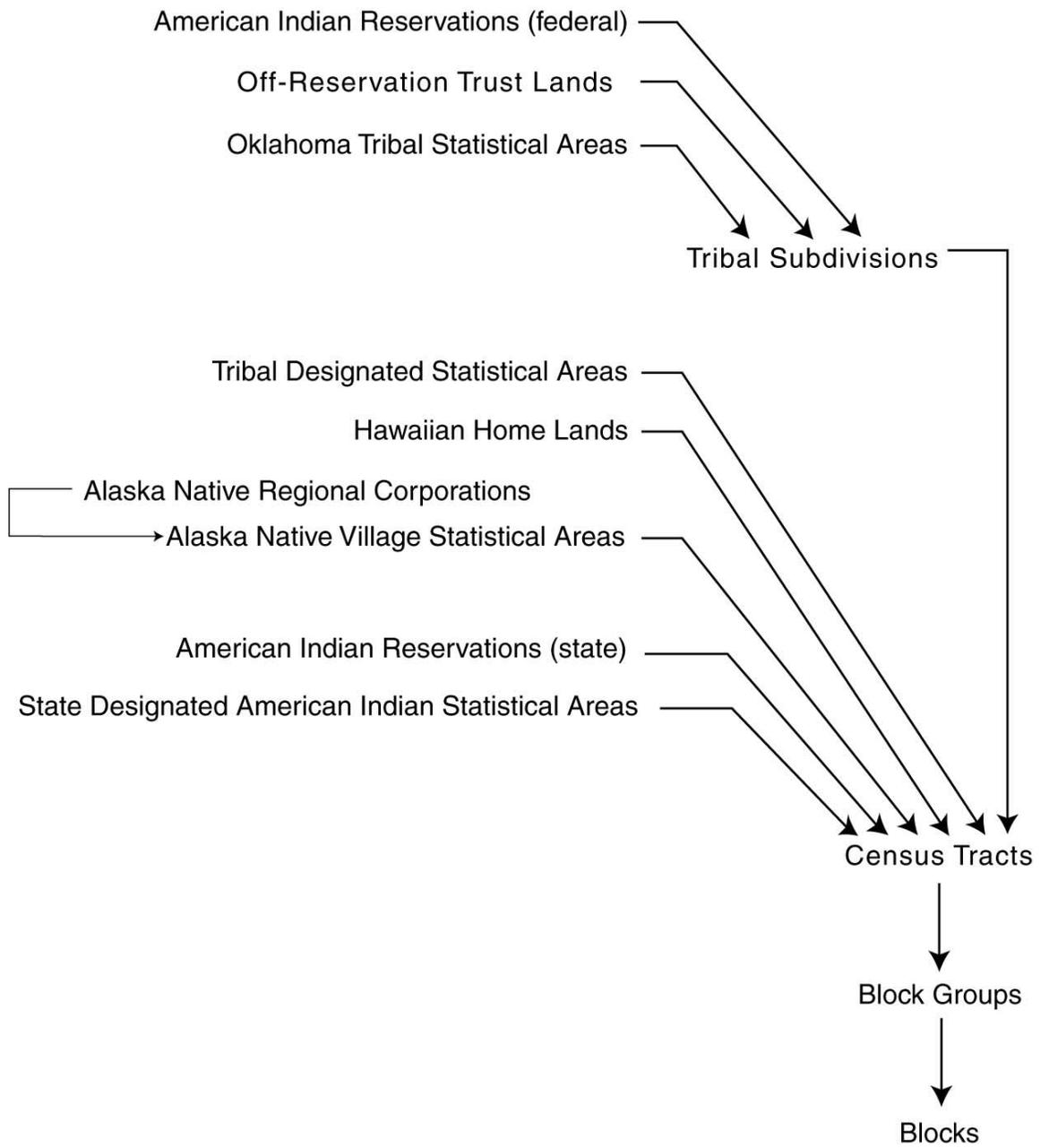
### State-County-Tract-Block Group Nesting

Summary Level	Geographic Unit
040	State (includes Washington, D.C. & Puerto Rico)
050	County
140	Census Tract
150	Block Group

### Supplemental Geographic Areas

Summary Level	Geographic Unit
400	Urbanized Areas
300	Metropolitan Areas (MSAs, CMSAs)
200	American Indian and Alaska Native areas
800	ZIP codes

**Hierarchy of American Indian, Alaska Native, and Native Hawaiian Entities (from *Census 2000 Summary File 1 Technical Documentation*, prepared by the U.S. Census Bureau, 2001, p. A-26)**



**A Visual Look at Census Geography**

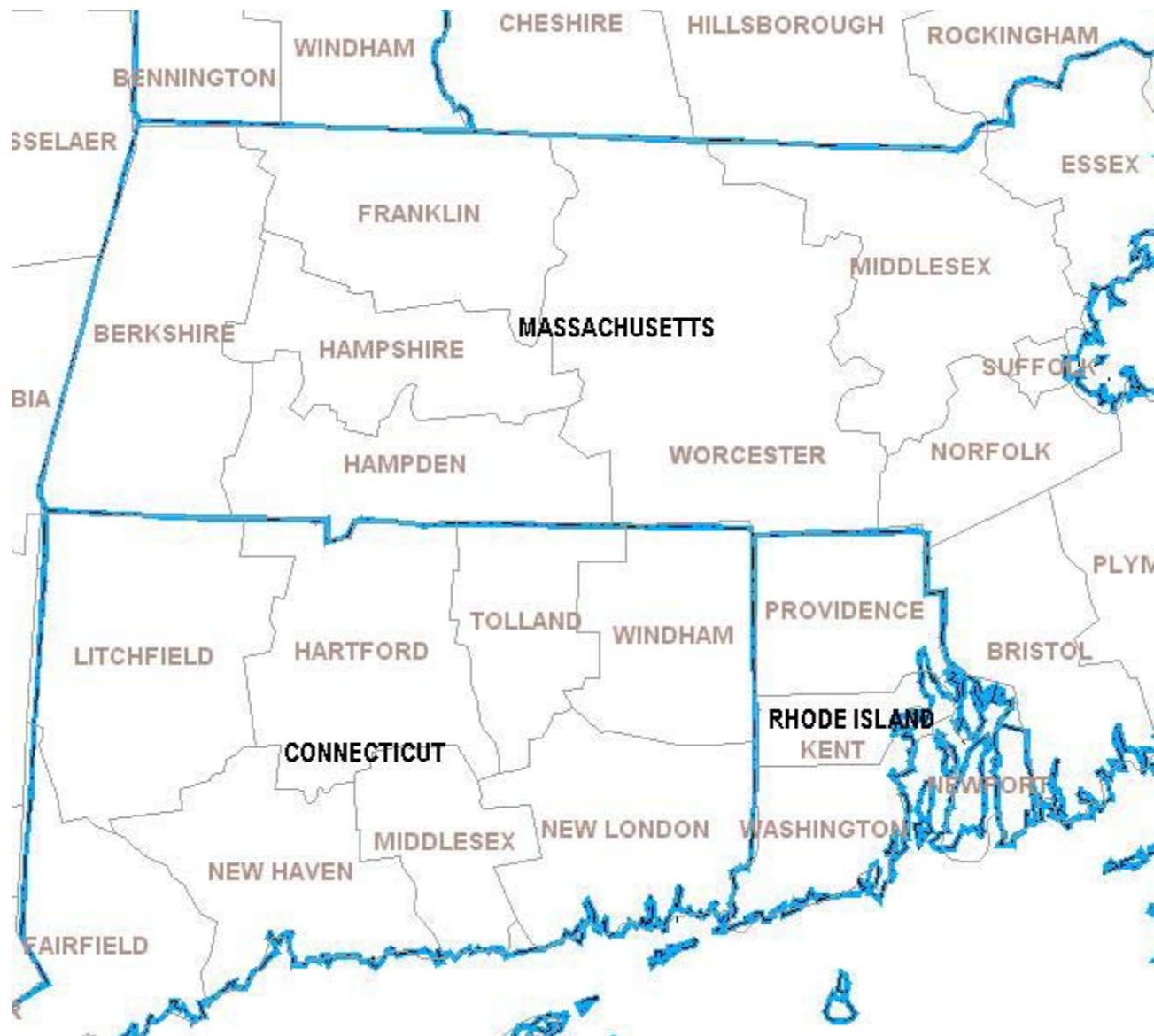
- **Continental United States (Regions in blue; Divisions in green; States in brown)**



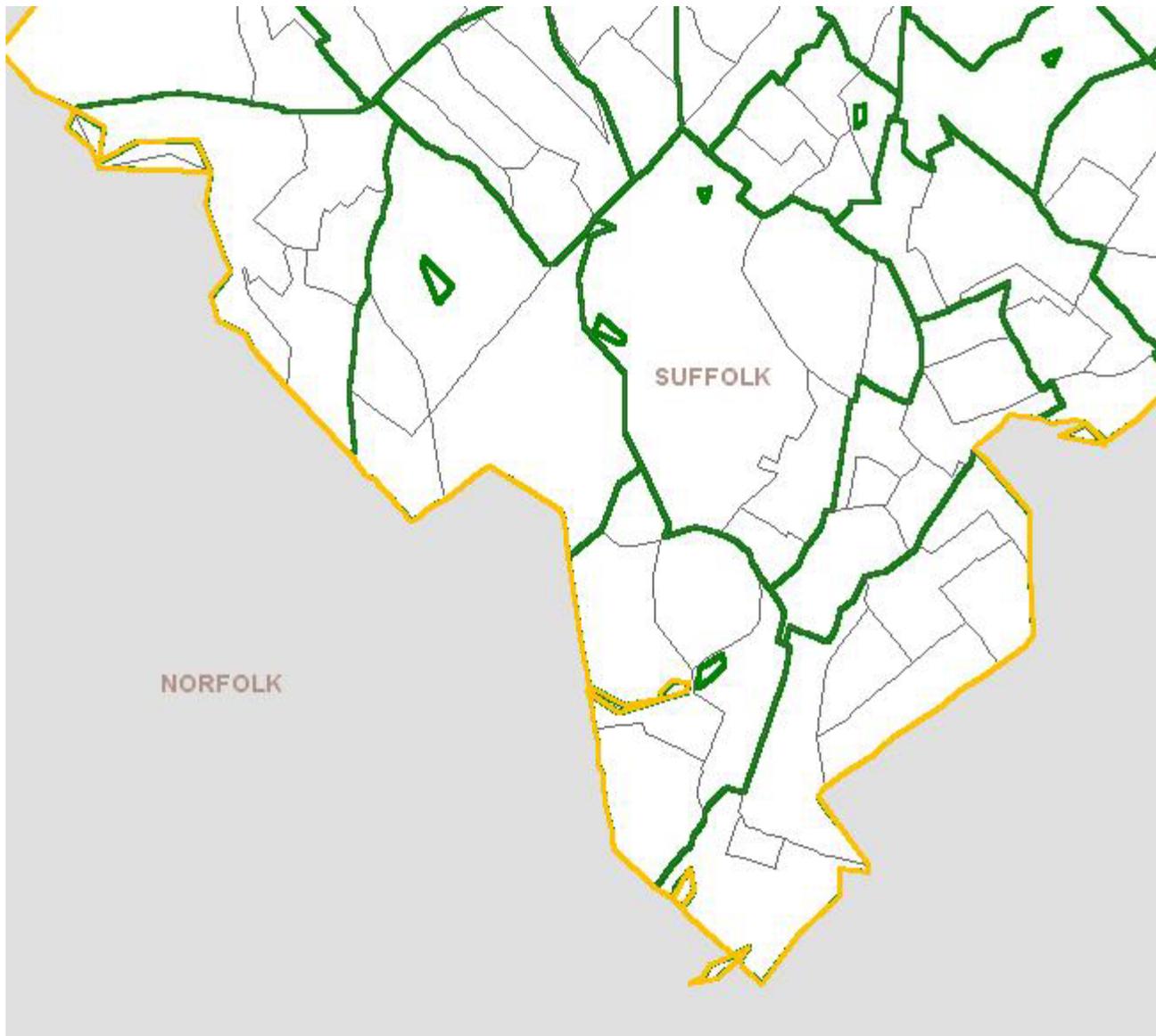
- Counties



- **A Closer Look at Southern New England Counties**



- Tracts (black lines) and Block Groups (green lines) in Suffolk County, MA



### **Census Geography Concepts**

- The census block is the basic level
- Confidentiality must be maintained, and data about individual persons and households are not revealed
- More detailed data are provided for higher levels of geography (Why?)
- Many, but not all, items are available at multiple summary levels

### **Potential Problems**

- The same geographic name is used for summary levels corresponding to different aggregations
- Geographic areas at lower levels may be subdivided by higher levels of geographic units
  - E.g., a census tract may be split by town boundaries
- The same variable names are used for different variables in the STF/SF 1 and [STF/SF 3](#)

- The way variable values are encoded makes identifying the meaning of variables difficult
- ZIP codes do not overlay other units cleanly
- Geographic boundaries change with time, making time-series analysis difficult (see "[An Analysis of Census Geography](#)," which looks at changes from 1980 to 2000)

### **Obtaining Census Geographic Boundary Files for Use in a GIS**

ArcView shapefiles and ArcInfo coverage formats are readily available for 1990 and 2000 Census geography boundaries

- [Boundary files from the U.S. Census Bureau](#)
- [Redistricting TIGER 2000 from ESRI's Geography Network](#)
- TIGER 1995 from ESRI's ArcData Online

### **Census Summary Files**

The most useful files distributed by the Census Bureau are the Summary Tape Files (now renamed simply Summary Files) that aggregate the individual census forms to various levels of census geography.

The Census Bureau distributed the 1990 Census files as DBF files on CD-ROMs. In what looks like a recent development, the Census Bureau has posted the contents of many 1990 CD-ROMs online. These are available via [HTTP](#) and [FTP](#). An online forms-based interface called [1990 Census Lookup](#) is available. Now, [American FactFinder](#) provides another forms-based online interface.

In fact, the **1980** STF 1 and STF 3 are now [online](#)! You can obtain the 1980 STF 1 via [HTTP](#) or [FTP](#) and the 1980 STF 3 via [HTTP](#) or [FTP](#). Documentation is available from the [Odum Institute for Research in Social Science](#).

The Census Bureau is distributing the 2000 Census files on CD-ROMs, DVD-ROMs in a proprietary format and online in flat ASCII format via [HTTP](#) and [FTP](#). [American FactFinder](#) provides a forms-based online interface.

#### **STF/SF 1: 100% count data from the short form**

For the 2000 Census, the SF 1 files encompass all summary levels.

For the 1990 Census, the STF 1 files came in four varieties:

- [A](#): States and subdivisions to the block group level
- [B](#): Block level
- [C](#): Entire U.S. and major subdivisions
- [D](#): Congressional Districts

#### **STF/SF 3: Sample data from the long form**

For the 2000 Census, the SF 3 files will encompass all summary levels.

For the 1990 Census, the STF 3 files came in four varieties:

- [A](#): States and subdivisions to the block group level
- [B](#): 5-digit ZIP codes
- [C](#): Entire U.S. and major subdivisions
- [D](#): Congressional Districts

The 1980 STF 1 and STF 3 files had varieties similar to those of the 1990 Census.

### **A Quick Look at the Census Data and Documentation**

#### **1790-1960 Censuses**

[State- and County-Level Data](#)

#### **1970 Census**

[Fifth Count Special Tabulation at UC Berkeley](#)  
[1970 Census Data at the University of Wisconsin](#)  
[CensusCD 1970](#) from [GeoLytics](#)

## 1980 Census

[Overview](#) from SUNY Albany's Center for Social and Demographic Analysis  
Data sets available from IPCSR (For MIT affiliates: [Information on Accessing ICPSR Data](#)  
1980 STF 1 via [HTTP](#) or [FTP](#) from the U.S. Census  
1980 STF 3 via [HTTP](#) or [FTP](#) from the U.S. Census  
1980 STF 1 and STF 3 Documentation from the [Odum Institute for Research in Social Science](#).  
[CensusCD 1980](#) from [GeoLytics](#) (available at Rotch Library)

## 1990 Census

[STF 3A Documentation Table of Contents](#)  
[STF 3A Variable Locator](#)  
[Subject Locator](#)  
[Table Definitions \(Matrix\)](#)  
[Using the File \(List of Census Tables by DBF File\)](#)  
[State/County FIPS Codes](#)  
[Summary Level Sequence Charts](#)  
[1990 Census Lookup](#) (a forms-driven Census information access system)  
[Census Data at the Center for Disease Control and Prevention](#)  
[CensusCD + Maps 1990](#) from [GeoLytics](#) (available at Rotch Library)  
[CensusCD Blocks 1990](#) from [GeoLytics](#)  
[Online Census Data at the University of California at Berkeley](#)

## 2000 Census

[Census 2000 Home Page at the U.S. Census Bureau](#)

[American FactFinder](#): A user-friendly web interface for Census 2000  
[Response Rates for 2000 Census](#)  
[Data Release Dates](#)  
[Subjects Areas of Questions Asked](#)

Public Law 94-171 (PL 94-171): Redistricting Data

- [Home Page](#)
- [Documentation](#)
- [Help on Using Browser Software on the CD-ROM](#)
- [Data](#)

100% Count (Short Form) Data

- Summary File 1 (SF 1): Data Down to the Block/Block Group Level
  - [Home Page](#)
  - [Documentation](#)
  - [Help on Processing Data Files in ASCII Format](#)
  - [Data](#)
- Summary File 2 (SF 2): Data for 249 Different Population Groups to the Tract Level
  - [Home Page](#)
  - [Documentation](#)
  - [Help on Processing Data Files in ASCII Format](#)
  - [Data](#)

#### Sample (Long Form) Data

- Summary File 3 (SF 3): Data Down to Block/Block Group Level
  - [Documentation](#)
- Summary File 4 (SF 4): Data for 249 Population Groups and 86 Ancestry Categories Down to the Tract Level
  - [Documentation](#)

#### **Time Series**

[CensusCD 40 Year](#) (1970-2000 time-series Census data) from [GeoLytics](#)

[Census Tract Data, 1940-1970: Elizabeth Mullen Bogue File](#) (from the National Archives and Records Administration; *not available online*)

#### **Censuses in Other Countries**

[International Statistics Agencies](#)

*The section of these notes entitled "Introduction to the U.S. Census of Population and Housing" is adapted from a Microsoft PowerPoint presentation created by Qing Shen for a lecture he delivered to the MIT class 11.208 on January 21, 1997.*