

# Exploring the U.S. Census

**Eugene Turner  
Department of Geography  
California State University, Northridge  
Northridge, CA 91330**

**(818) 677-3527  
eturner@csun.edu**

# Exploring the U.S. Census

## Table of Contents

Introduction  
    Organization  
    Data Sets

### Chapter 1. Accessing the Digital Census

- A. About the Census
- B. Digital Census Data
  - Summary Files
  - Table Details
  - The American Community Survey
- C. Public-Use Microdata Sample Files
- D. Geography in Summary Files
  - SF3 Summary Level Code Hierarchy for Selected Geographic Units
  - Other SF3 Summary Level Codes
  - Codng Geographic Units - FIPS Codes
- E. Census Comments
  - The Race Question
  - Hispanic Subgroups
  - Small-area Tabulations
  - Census Geography
- F. Exploring the Census Web Site

### Chapter 2. Place Description, Ranking, and Mapping

- A. Some Basic Population Data Describing a City
- B. Examining a Characteristic in All Cities - Ranking Places
  - Densely Populated Places
  - Ethnic Composition
- C. Describing a Distribution with Statistics
- D. Graphing a Distribution
- E. Mapping a Distribution
  - Census Geography
  - Mapping Counts and Percents
  - Choropleth Maps
  - Graduated Symbol Maps
  - Dot Maps
  - Mapping with ArcGIS

### **Chapter 3. Analyzing Other Population Characteristics**

- A. The Sex Ratio
- B. The Location Quotient
- C. The Entropy Index

### **Chapter 4. Association between Two or More Variables**

- A. Cross Tabulations
- B. Scattergrams
- C. Correlation
- D. Regression

### **Chapter 5. Describing the Age of Populations**

- A. Median Age
- B. Dependency Ratios
- C. Population Pyramids

### **Chapter 6. Population Growth**

- A. Describing Population Change
- B. The Effect of Migration and Residential Mobility on Population Change
  - The Demographic Equation
  - Births and Deaths
  - Migration and Local Residential Mobility
- C. The Challenge of Analyzing Change

### **Chapter 7. The Public-use Microdata Sample (PUMS)**

- A. Income Distribution Differences Among Ethnic Groups
- B. Occupational Differences Among Ethnic Groups

### **Chapter 8. Describing the Relative Location of Populations**

- A. Service Areas
- B. Centers of Population
  - The Mean Center
  - Population Potential

## Appendices

- A. State and County FIPS Codes
- B. Summary File Segmentation
- C. Summary Level Sequence Chart
- D. Race Codes
- E. Hispanic Codes
- F. Ancestry Codes
- G. State and Foreign Country Codes
- H. Occupation Codes
- I. Industry Codes
- J. SF3 Table Description

## Exercises

- Ex 1. Exploring the Census Web Site
- Ex 2. Accessing Census 2000 SF3 and SF4 Data and Data Bases at ICPSR
- Ex 3. Introduction to Excel
- Ex 4. Analyzing Census Data in Excel
- Ex 5. Sex Ratio
- Ex 6. Location Quotient
- Ex 7. Diversity Index
- Ex 8. Association between Variables
- Ex 9. Dependency Ratio
- Ex 10. Population Pyramid
- Ex 11. Population Growth
- Ex 12. Population Demographic Equation
- Ex 13. Accessing Census 2000 PUMS Data
- Ex 14. Analyzing Census 2000 PUMS Data
- Ex 15. Mapping Census 2000 Data
- Ex 16. Examining the Demographics of a Service Area
- Ex 17. Downloading Raw Census Data

# Tables

## Chapter 2

Table 1.	Ethnic Populations in Glendale, Los Angeles, and California, 2000
Table 2.	Ranking of States Based on Census Variable Counts

## Chapter 3

Table 3.	Sex Ratios, 2000
Table 4.	Sex Ratios by Age in California, 2000
Table 5.	Location Quotient by Occupation and Class of Worker Los Angeles County, 1990
Table 6.	Diversity Scores

## Chapter 4

## Chapter 5

Table 7.	Dependency Ratios, 2000
----------	-------------------------

## Chapter 6

Table 8.	Greatest Mexican Population Changes in California Counties, 1980 – 1990 - 2000
Table 9.	Population Change in California, 1990 – 2000
Table 10.	Citizenship in California, 2000
Table 11.	Country of Birth in California, 2000
Table 12.	Residence in 1995 for California Residents in 2000

## Chapter 7

Table 13.	Income Distribution within Ethnic Groups
Table 14.	Ethnic Employment for Males and Females in PUMA 5200

## Chapter 8

Table 15.	Distances from Tract 3019.00
Table 16.	Mean Center of Glendale Tracts, 1980 and 1990
Table 17.	Accessibility Index to Two Glendale Tracts

# Exploring the U.S. Census

## Introduction

The United States Bureau of the Census collects and publishes a wide range of statistics about the population, housing, economy, productivity, and government in the United States. Data on these subjects are periodically tabulated and released to give a better understanding of American society.

Among the most sought-after data are the statistics on housing and population collected every decade. Demographers, planners, businessmen, and social scientists use this information to track differences between locations and over time. Government agencies use the information to help them decide what needs to address and to allocate funding to various social programs.

This module explores the tabulations of the *Census of Population and Housing* and some of the basic techniques used to describe and analyze the data contained within it. These techniques form the basis of more sophisticated techniques used by a great many researchers in universities, in businesses, and in government. However, advanced procedures are beyond the introductory scope of this module, and students should consult other sources in texts and journal articles for information on sampling and statistical analysis.

## Organization

The following chapter describes some of the resources that are easily accessible via the web site maintained by the Bureau of the Census. Most of these resources cover the censuses of 1990 and 2000, but data from earlier censuses are increasingly being converted to digital form and many of these can be accessed at the web site of the Interuniversity Consortium for Political and Social Research (ICPSR). See [www.icpsr.com](http://www.icpsr.com) In Chapter 2 of this module various methods are presented for describing populations through counts, percentages, density, and mapping. Chapter 3 presents methods for looking at gender and ethnic components of the population, and Chapter 4 explores the age components. Chapter 5 illustrates ways of looking at change in the population between two censuses. Through data provided from the California Department of Finance, population estimates and migration statistics are provided between 1990 and 1997 for California counties. Chapter 6 presents a few simple ways of measuring the spatial components of the population. Chapter 7 investigates some of the ways that the PUMS database can be used to create special tabulations to control for factors such as gender, education, and income when trying to understand differences between groups. Finally, Chapter 8 introduces a few of the ways that spatial analysis might be done with

census data. The methods introduced could be carried out with *Excel*. However, this opens a broad range of techniques now available with GIS software and suitable for separate modules.

Several important appendices have been attached which provide additional information on the content and structure of the digital census files. The codes for the selected variables will be especially useful after data have been downloaded for processing.

Seventeen exercises cover a variety of census related issues from locating and downloading data, to tabulating and analyzing tables, to creating charts and maps of the information. These exercises are intended to provide the user the opportunity to access digital information to answer basic kinds of demographic questions. Certainly not every demographic topic has been covered in this module and many could be added. For example, analysis of segregation, poverty, assimilation, and crowding to name but a few. Fortunately there are other modules that deal with some of these issues.

Over the last decade the U.S. Census has become easily accessible over the Internet. With the assistance of interactive software, not only technical documents, but data tables, graphs, and maps can be customized and downloaded. Data extracts can be downloaded in *Excel* or delimited formats and entire summary files can be accessed in raw form through *Access* and *SAS* software. For statistical analysis either *SAS* or *SPSS* may be used. One now has the ability to move the data files through various software packages as needs dictate.

For many of these exercises *Excel* will be used, and so the user should become familiar with it. For those not familiar with it, a basic introduction is included along with a few methods particular to handling census information.

Exercise 1 provides an overview of the resources available on the Census web site. Since this site seems to be constantly evolving, don't be surprised if things in the exercise don't exactly match the web pages in the future. As of May 2007, the exercise and the web pages are in agreement.

Exercise 2 demonstrates how to download data using the *American Factfinder* search engine. Essentially one selects a summary file and a level of geography and then a table is created. This can be downloaded in *Excel* format with rows and columns transposed from what is initially displayed. Remember that *Excel* only allows 255 columns and the number of geographic units typically exceeds that. Thus, the transposition is necessary.

Exercise 3 is an introduction to *Excel* and some of the capabilities that might be useful with census data. Experienced users may want to skip it.

Exercise 4 deals with analyzing data using Excel. Often simple descriptions and ranking of values are all that is needed. The creation of a bar graph and a frequency graph are introduced.

Exercise 5 demonstrates how to compute the Sex Ratio in Excel.

Exercise 6 introduces the calculation of the Location Quotient which is helpful in determining if a location has a more or less than expected share of a characteristic.

Exercise 7 demonstrates how to calculate a measure of diversity using the Entropy Index. It indicates how evenly numbered several groups are within an area.

Exercise 8 illustrates how a scattergram can be helpful for examining the association of two variables.

Exercise 9 focuses on calculating several forms of the Dependency Ratio. The ratio gives a sense of the relative population support a dependent group has in a location compared to the support in another location.

Exercise 10 shows how Excel can be used to prepare a population pyramid. This graphic device is very helpful in understanding the age and sex structure of a population.

Exercise 11 illustrates three methods for expressing the change in population over time. Each can result in a very different set of places having the most change.

Exercise 12 introduces the Demographic Equation. It is a useful tool in estimating the change in population in the years between the censuses. Resources at the California Department of Finance are explored.

Exercise 13 introduces the Public Use Microdata Sample data set and the IPUMS web site at the University of Minnesota for accessing this information. This web site provides PUMS data for a number of census decades as well as data for other countries.

Exercise 14 uses SPSS to aggregate PUMS data into useful tables for analysis. In this example tables of occupations for Asian Indian men and women are created to determine what occupational niches may exist. In a followup exercise income differences between Asian Indian men and women are explored.

Exercise 15 shows how to use ArcMap to create a map of census data.

Exercise 16 illustrates how to use ArcMap to select data that surround a site of interest. The characteristics of such a service area are important for marketing studies.

Exercise 17 demonstrates how to download raw census data and import it into Access for further extraction.

## Data Sets

Only six databases have been extracted and included with this module. Because it is so easy to download census data from that web site, users may wish to obtain data that represents their own area of interest. The accompanying databases are:

1. *Califcities.xls* Selected race and housing variables from SF3 for all 1074 California cities.
2. *CalifcitiesAgeSex.xls* Sex by age from SF3 for all 1074 California cities.
3. *Ex3\_Excel.xls* Selected ethnic variables for ten California counties
4. *UScoPop80-00.xls* Total and Hispanic populations for 1980, 1990, and 2000 for 3140 counties

The following files are located in the Mapping folder.

1. *CAcensussex.dbf* Data file to be joined to the California county boundary file
2. *CAcensusVarIDs.xls* More detailed labels of variable names in CAcensussex.dbf
3. *CAco* California county outlines
4. *SFVtractPT* Tract centroid file with associated data
5. *SFVtracts* Tract boundary file in the San Fernando Valley within Los Angeles
6. *NewSites* Three sites within the San Fernando Valley

## Other Resources

### **CensusScope**

A data extraction program supported by the Social Science Data Analysis Network (SSDAN)  
<http://www.censusscope.org/>

### **Kids Count in the Classroom Project**

A variety of tools, data sets, and modules for demographic analysis.  
<http://www.ssdan.net/kidscount/index.shtml>

### **DataCounts! Exploring Society by the Numbers**

Yet another SSDAN project with special data analysis software and useful extractions from various censuses. A large number of modules are available.  
<http://www.ssdan.net/datacounts/>

See also two publications:

*America By The Numbers: A Field Guide To The U.S. Population* by William H. Frey, Bill Abresch, and Jonathan Yeasting

*Investigating Change in American Society: Exploring Social Trends with U.S. Census Data* by William H. Frey

### **The Population Reference Bureau**

Basic population information for the U.S. and the world.  
<http://www.prb.org/>

### **The United Nations Population Information Network**

A variety of reports, data sets, and other resources on world population issues.  
<http://www.un.org/popin/data.html>

### **Geolytics**

A company that sells repackaged census data for business. A source for pre 1990 census information in common geographic units..  
<http://www.geolytics.com/>

### **TGR2SHP and TGR2MIF**

Free software written by Bruce Ralston that converts TIGER files to boundary files.  
<http://tnatlas.geog.utk.edu/downloadfree.htm>

### **Proximity**

A source for demographics, census mapping files, and various resources. Some are free.  
<http://proximityone.com/maps2000.htm>

### **UC San Diego Social Science Data Collection**

Links to many census and demographic sources.  
<http://ssdc.ucsd.edu/ssdc/cen2k.html>