

The Census Module

All Examples

Exercise 1. Exploring the Census Web Site

**Exercise 2. Accessing Census 2000 SF3 and SF4 Data and
Data Bases at ICPSR**

Exercise 3. Introduction to Excel

Exercise 4. Analyzing Census Data in Excel

Exercise 5. Sex Ratio

Exercise 6. Location Quotient

Exercise 7. Diversity Index

Exercise 8. Association between Variables

Exercise 9. Dependency Ratio

Exercise 10. Population Pyramid

Exercise 11. Population Growth

Exercise 12. Population Demographic Equation

Exercise 13. Accessing Census 2000 PUMS Data

Exercise 14. Analyzing Census 2000 PUMS Data

Exercise 15. Mapping Census 2000 Data

Exercise 16. Examining the Demographics of a Service Area

Exercise 17. Downloading Raw Census Data

Exercise 1 – Exploring the Census Web Site

The purpose of this exercise is to get a sense of where different resources can be located on the Bureau of the Census web site.

1. Log on to the Census web page at: www.census.gov
2. In the right column locate the *Find an Area Profile with Quickfacts* option. Under *Select a State* select *California*.

QuickFacts and QuickLinks

1. Now under the *California Counties* window (see above) select *Los Angeles County* from the list and select the *Go* button.

	California	USA
Population, 2003 estimate	35,484,453	290,809,777
Population, percent change, April 1, 2000 to July 1, 2003	4.8%	3.3%
Population, 2000	33,871,648	281,421,906
Population, percent change, 1990 to 2000	13.6%	13.1%
Persons under 5 years old, percent, 2000	7.3%	6.8%
Persons under 18 years old, percent, 2000	27.3%	25.7%
Persons 65 years old and over, percent, 2000	10.6%	12.4%

2. Compare some of the percents between the State and the County. It is often helpful to compare a place of interest to a larger total such as the state or the entire nation to get a sense of how your area varies from a reference value.

How does Los Angeles County compare in Percent Foreign-born, Homeownership, and Median Household Income to the State?

3. At the top of the page to the right of the *California counties* window, click on the *selection map* link to see the location of Los Angeles County within the State. When done, click the *Back* button on the Browser.

4. Select one of the *More California data sets* buttons at the top right of the page.



5. When the *California QuickLinks* page of other files appears, select from near the bottom, the *Historical population counts* link.

California QuickLinks

[PDF] denotes a file in Adobe's [Portable Document Format](#). To view a file, you will need a PDF viewer.

People QuickLinks

- Census 2000 population, demographic, and housing information:**
 Complete counts from questions collected on both the short form and the long form.
 - Population by Race and Hispanic or Latino origin: [California counties, 2000](#)
 - Housing unit counts: [California counties, Places in California](#)
 Estimates based on the one-in-six sample of housing units that received the full questionnaire.
 - [Social characteristics](#)
 - [Economic characteristics](#)
 - [Housing characteristics](#)
- 1990 Census population, demographic, and housing information:**
 Complete counts from questions collected on both the short form and the long form.
 - [General characteristics](#)
 - [Population Projections to 2025](#)
 - [Historical population counts, 1900 to 1990](#), for selected counties
 - [Income and Poverty](#) model-based estimates for selected counties

Look over this text file data set, then close it and browse over the *Quicklinks* web page to see what other data is available.

6. Finally, at the bottom of the *California QuickLinks* page under

U.S. Census Bureau

The 2007 Statistical Abstract
The National Data Book

[Main](#) | [Overview](#) | [Print Version](#) | [Earlier Editions](#) | [Order](#)

BROWSE SECTIONS: (SPREADSHEETS)
Accommodation, Food, & Other Services
Agriculture
Arts, Entertainment, & Recreation
Banking, Finance, & Insurance
Business Enterprise
Comparative International Statistics
Construction & Housing
Education
Elections
Energy & Utilities

What is the Statistical Abstract?

The *Statistical Abstract of the United States*, published since 1878, is the authoritative and comprehensive summary of statistics on the social, political, and economic organization of the United States.

Use the Abstract as a convenient volume for statistical reference, and as a guide to sources of more information both in print and on the Web

Sources of data include the Census Bureau, Bureau of Labor Statistics, Bureau of Economic Analysis, and many other Federal agencies and private organizations

Other QuickLinks select the **Statistical Abstract of the United States** link. This document contains pdf files of the annual data compiled for the *Abstract*. Note on the left side panel below are links to various parts of the *Abstract*.

When done, use the Browser's *Back* button to return to Home page and then click the *American Factfinder* page link from the left panel.

7. From the list of menu items on the left select the **Fact Sheet** button. The window below with statistics for the entire U.S. will open. Note that by default the latest *American Community Survey* data will appear in the table.

The screenshot shows the American Factfinder website interface. The left navigation menu has 'FACT SHEET' highlighted, with a black arrow pointing to it. The main content area displays 'United States' and '2005 American Community Survey Data Profile Highlights'. A search box on the right contains 'Los Angeles City' and a 'GO' button. The '2005' year tab is selected.

8. In the window below notice the *Margin of Error* for the ACS data. That means that 90% of the time the estimated value will fall within the *Estimate* plus or minus the *Margin of Error* value.

The screenshot shows the American Factfinder website interface with a data table. A black arrow points to the 'Margin of Error' column. Below the table is a note about the survey universe.

General Characteristics - show more >>	Estimate	Percent	Margin of Error	
Total population	288,378,137		*****	
Male	141,274,964	49.0	+/-20,305	rank
Female	147,103,173	51.0	+/-20,305	rank
Median age (years)	36.4	(X)	+/-0.2	rank
Under 5 years	20,267,176	7.0	+/-12,409	
18 years and over	215,246,449	74.6	+/-16,617	
65 years and over	34,760,527	12.1	+/-15,554	rank

Note: The 2005 American Community Survey universe is limited to the household population and excludes the population living in institutions, college dormitories, and other group quarters.

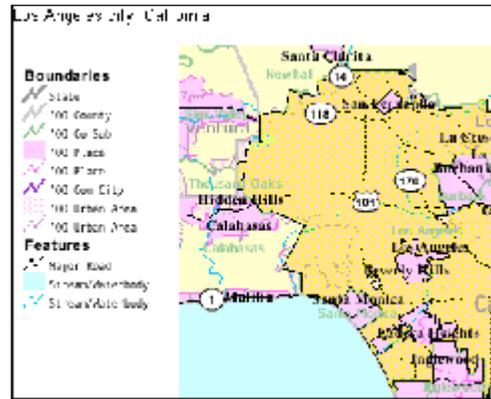
9. On the right side of the *Fact Sheet* page enter *Los Angeles City* in the city/town window. Then click *Go*. The U.S. fact sheet will be replaced by one for the City of Los Angeles in California.

10. On the right side of the data values select the *Reference Map* link.

Number	Percent	U.S.		
3,694,820			map	brief
1,841,805	49.8	49.1%	map	brief
1,852,015	50.2	50.9%	map	brief

Look over the map of LA and then click the *Close* button.

11. Use the *Back* button to return to the *Fact Sheet* page.



12. Enter your ZIP code in the upper right box and click *GO*.

Main ▶ Fact Sheet

FACT SHEET

 [United States](#) | 91325
Zip Code Tabulation Area 91325

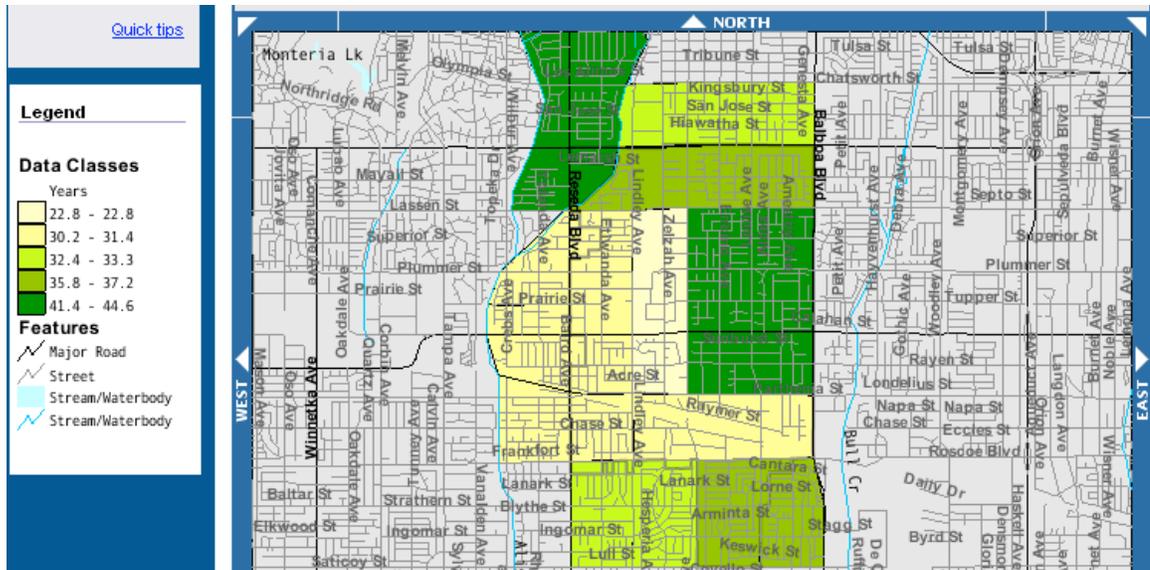
2000 2004 data [not available](#) for this geography

Census 2000 Demographic Profile Highlights: | [Reference Map](#)

General Characteristics - show more >>	Number	Percent	U.S.		
Total population	32,300			map	brief
Male	15,629	48.4	49.1%	map	brief
Female	16,671	51.6	50.9%	map	brief
Median age (years)	31.3	(X)	35.3	map	brief
Under 5 years	1,738	5.4	6.8%	map	
18 years and over	25,683	79.5	74.3%		

Because there is no ACS data for ZIP codes the data will change to Census 2000.

13. Locate the *Median age* item from the list and click on the map link on the right. A map of median age by census tract will appear.



When done with the map, click the Back button on the Browser.

14. Now click on the *Brief* link to the right of *Median age*. A pdf file will open that contains an analysis of the variable for the United States. Read over some of the file and then close it.

15. Click on the small *Home* link at the bottom of the page to return to the main Census page.

16. From the left panel of the main window select the *New on Site* link.



17. When the list of new items appears, scroll to the bottom and select the *Subscribe to the Census Product Update* link. From here you can put your email on a list server to receive notices of new census products.

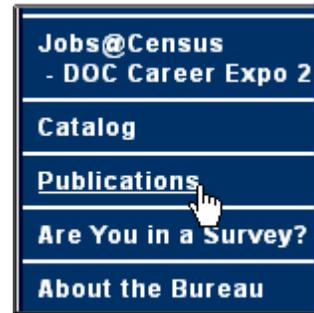


If you want to be on either the press release list or the new product list follow the instructions to submit your email. Then use the Back button to return to the main census page.

Publications

1. From the left panel of links select the *Publications* link. A list of subjects will appear from which you should select the *Population* link.

2. Scroll down the list of population topics and select any pdf file of interest to you. Check it out and then return to the main page.



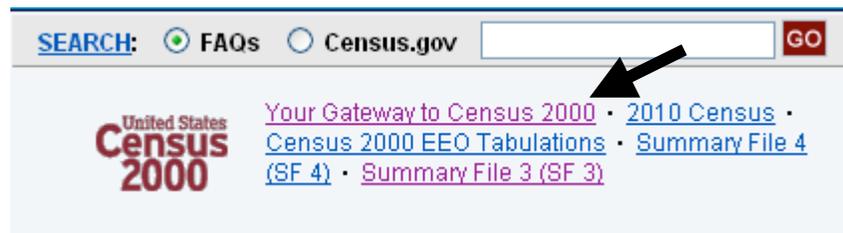
Income and Poverty	Estimates and Projections	Household Economic Studies	Population Characteristics	Special Studies	Technical Papers	American Community Surveys
------------------------------------	---	--	--	---------------------------------	----------------------------------	--

POPULATION		
INTERNATIONAL POPULATION REPORTS		
WP/02-1	Global Population at a Glance: 2002 and Beyond	Doc
WP/02	Global Population Profile: 2002	Doc
WP/02-2	The AIDS Pandemic in the 21st Century	Doc
P95 Series	An Aging World: 2001	Doc
INCOME AND POVERTY		
P60 Series	Consumer Income	Doc
ESTIMATES AND PROJECTIONS		
P25-1127	National and State Population Estimates: 1990 to 1994 (1 Mb)	Doc
P25-1129	Projections of the Number of Households and Families in the United States: 1995 to 2010	Doc
P25-1130	Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050 (3.9 Mb)	Doc
P25-1131	Population Projections, States 1995 to 2025	Doc
P25-1132	Projections of the Voting-Age Population for States: November 1998	Doc

Gateway to Census 2000 Link

1. From the main page click on the *Your Gateway to Census 2000* link.

The *Gateway* contains many useful links including those to the summary files.



U.S. Census Bureau

United States Census 2000

Your Gateway to Census 2000

Thank you, America, for your participation in Census 2000. The population of the U.S. on April 1, 2000 was **281,421,906** [PDF 2M].

Information Links

- [News Media](#) - News releases, web casts, tip sheets, videos, photos, and embargo access
- [Introduction to Census 2000 Data](#) - Overviews of Census 2000 and PowerPoint slides to download
- Release Schedules by:**
 - [Product](#)
 - [Geography](#)
 - [Subject](#)
- [Census Store](#) - Order products on-line
- [Evaluations](#) - Results from the Census 2000 Testing, Experimentation, and Evaluation Program
- [Special Tabulations](#) - Program for obtaining custom data tabulations
- [Count Question Resolution](#) - Program for challenging Census 2000 counts
- [Notes and Errata](#) [PDF 1.4M]

Access Data by Geography

American FactFinder

Tables and maps of Census 2000 data for all geographies to the block level

State & County QuickFacts

Summaries of the most requested data for states and counties

Data Highlights

Data highlights, documentation, and FTP access for the U.S., states, counties, places (cities & towns) and more, including Puerto Rico and [Island Areas](#)

Select a state:

Census Data

Rankings and Comparisons (PHC-T)

Tables showing population change, comparisons with 1990, Race and Hispanic or Latino origin, and other topics for states, counties, and places

Census 2000 Data Releases

- [Demographic Profiles](#)
- [Summary File 1](#)
- [Summary File 2](#)
- [Summary File 3](#)

2. For now, select the *Enter a Street Address* link.

3. When the window below opens, enter the address of this campus (1) and select *Go*. (2)



Search

You are here: [Home](#) > [Search](#) > [Advanced Geography Search](#)

keyword:

Choose a geography selection method

Select a year and program

Enter a street address, city, and state or a street address and zip code. Click [here](#) for instructions

City: State: Zip Code:

4. In the next window select Census tract 1152.02 and click Go.

Enter a street address, city and state, or a street address and zip code. Click 'Go'

Street Address [Quick tips](#)

City State Zip Code

Geographies Containing **18111 Nordhoff St, Northridge, California, 91330:**
 Select a geographic area and click 'Go'

State: California
 ... County: Los Angeles County
 ... County Subdivision: Los Angeles CCD
 ... Census Tract: **Census Tract 1152.02**
 ... Block Group: Block Group 1
 ... Block: Block 1001
 ... Place: Los Angeles city
 ... Congressional District - 106th: Congressional District 25 (106th Congress)

[Explain Census Geography](#)

You will get a lengthy list of resources for the tract.

5. Select a few links shown below.

Search results for Census Tract 1152.02

Reference Maps

2000
 ■ [Census Tract 1152.02, Los Angeles County, California](#)

Thematic Maps

Census 2000 Redistricting Data (Public Law 94 171) Summary File

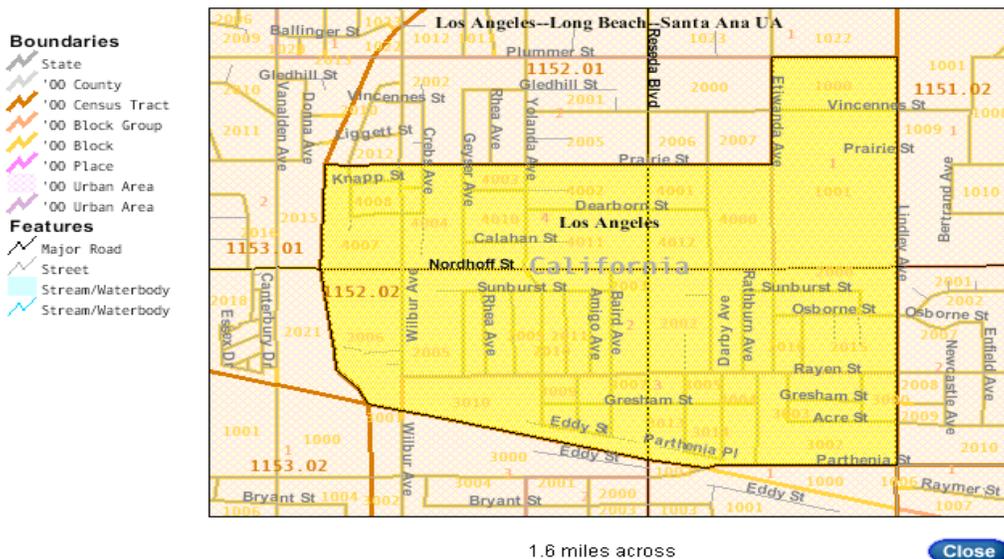
- TM-PL001: [Total Persons, 2000](#)
- TM-PL002: [Persons per Square Mile, 2000](#)
- TM-PL003A: [Percent of Persons Who Are White Alone, 2000](#)
- TM-PL003B: [Percent of Persons Who Are Black or African American Alone, 2000](#)
- TM-PL003C: [Percent of Persons Who Are American Indian and Alaska Native Alone, 2000](#)
- TM-PL003D: [Percent of Persons Who Are Asian Alone, 2000](#)
- TM-PL003F: [Percent of Persons Who Are Native Hawaiian and Other Pacific Islander Alone, 2000](#)
- TM-PL003F: [Percent of Persons Who Are Some Other Race Alone, 2000](#)
- TM-PL003F: [Percent of Persons of Two or More Races, 2000](#)
- TM-PL003H: [Percent of Persons Who Are Hispanic or Latin \(of any race\), 2000](#)

[11 more Thematic Maps](#)

Census 2000 Summary File 1 (SF 1) 100-Percent Data

Below is a map of the tract that shows streets, tract IDs, and block numbers.

Census Tract 1152.02, Los Angeles County, California



6. Return to the *Gateway* page and locate the *Geographic Products and Information* link from the center left column.

Geographic Products and Information
 Maps and digital geographic products for use in GIS and mapping software

7. From the *Map Products and Information* page select the *Census 2000 Map Series* link.

Census 2000 Maps

- [Census 2000 Map Series](#)
View the Census 2000 map Service and American Fact

8. A list of products will appear from which you should select the *Map Products* link.

- [WHAT'S NEW](#) - Lists the m
- [MAP PRODUCTS](#) - Links to Document Format [PDF]. O also provided.

9. From the list of map products scroll down to the *Reference Maps* set and select the *Census Block Maps: 2000* link.

REFERENCE MAPS ([description](#))

Census Geography

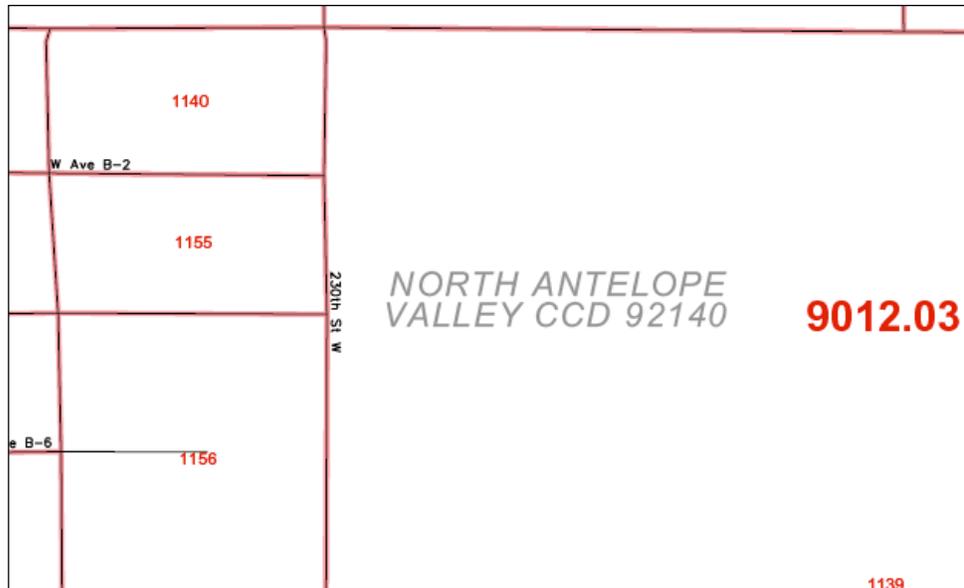
- Congressional District Map Products
[108th Congressional District Maps and 109th Congressional District Maps and Description 2000](#)
- American Indian Tribal Census Tract O
[Description 2000](#)
- American Indian/Alaska Native/Hawaiian
[Description 2000](#)
- Census Block Maps: ([Also available on Description 2000 | 1990](#))
- Census Tract Outline Maps: ([Also avai Description 2000 | 1990](#))



10. Select the *State* directory and then work your way down from to *County* to *Los Angeles* and select the following link:

 [CBC06037_006.pdf](#) 06-Feb-2002 07:11 98K

A color map 33 x 36" in size in pdf format will open. If you need a paper copy of a map of census geography you can obtain one if



you have a large-format printer.

11. Use the *Back* button to return to the *Census Bureau Map Products* page.

Map Products

The Bureau of the Census provides a number of other map products in both static and interactive form. The former often are suitable for printing on a color plotter.

1. Scroll down to the *Map Gallery* link at the bottom of the page.

2. On the *Map Gallery* page scroll down and select the *Mapping Census 2000: The Geography of U. S. Diversity* link.

Additional Maps, Special Reports and Publications

- [Mapping Census 2000: The Geography of U.S. Diversity](#) (Atlas)
This Census Bureau special report presents a synthesis of the basic data for the last decade. Each page features county-level detail for the 50 states and includes a small state-level map for a simplified view of the population in the U.S. Census Bureau Redistricting (PL 94-171) Summary File.
- [Map Products Web page](#)

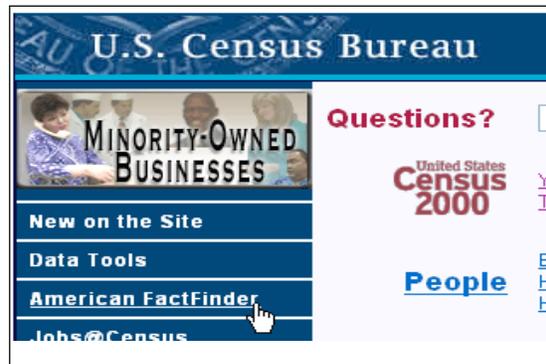
From the list of maps shown right examine a few of the maps in pdf format.

Census 2000 Special Reports (CENSR/01-1)

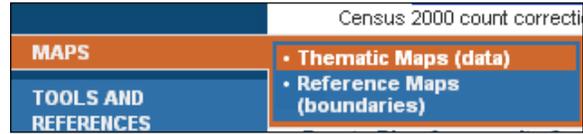
[Cover and Front Matter](#) (557k) [PDF]

1. [Location Maps](#) (1.3M) [PDF]
2. [Total Population](#) (2.8M) [PDF]
3. [Diversity](#) (1.4M) [PDF]
4. [White](#) (3.1M) [PDF]
5. [Black or African American](#) (3.4M) [PDF]
6. [American Indian and Alaska Native](#) (3.5M) [PDF]
7. [Asian](#) (3.7M) [PDF]
8. [Native Hawaiian and Other Pacific Islander](#) (4.1M) [PDF]
9. [Two or More Races](#) (1.2M) [PDF]
10. [Hispanic or Latino Origin](#) (1.5M) [PDF]
11. [White, Not Hispanic or Latino Origin](#) (3.2M) [PDF]

3. Return to the main Census Bureau web page and from the left panel select the *American Factfinder* link.



**4. From the left panel of the *American Factfinder* page select the *Maps* – *Thematic Maps* link.
Review the explanation and try a few maps.**



Working with Thematic Maps

You can change the view of the Thematic Map by

- selecting a different theme, geographic area, or data set by using the links at the top of the page
- changing the data values associated with each color (classes) for a different view of the same data
- clicking on the identify button, then clicking the map to find the data value for a shaded area
- clicking on the zoom button, then clicking the map to zoom in

Thematic Maps

You are here: [Main](#) > [All Data Sets](#) > [Data Sets with Thematic Maps](#) > [Geography](#) > [Themes](#) > [Data Sets with Thematic Maps](#)

Change...

- [Data Classes](#)
- [Boundaries](#)
- [Features](#)
- [Title](#)

Reposition on...

- [A street address or ZIP code](#)
- [A latitude and longitude](#)
- [The selected geography](#)

View...

- as a [ranking table](#)

[Quick tips](#)

United States by State
M0504. Percent of People Born in Latin America: Foreign-born population
 Universe: Foreign-born population
 Data Set: 2004 American Community Survey

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

Click map to: [Identify - geo name and data value](#)

Click map to: [Identify](#) [Zoom](#)

Legend

Data Classes

For help with using Thematic Maps, see the [Creating and Using Maps Tutorial](#).

Under *Change..* you can pick a different classing method, display different boundaries or features, and set a title. Under *Reposition* you may change the center point of the map. You may also click on a state to view the numbers associated with it. At the top of the map under the *Display map by:* window you may change the geographical units to counties or other areas. Also at the top of the map are a series of bars that will enable you to zoom in or out of the display by clicking on them.

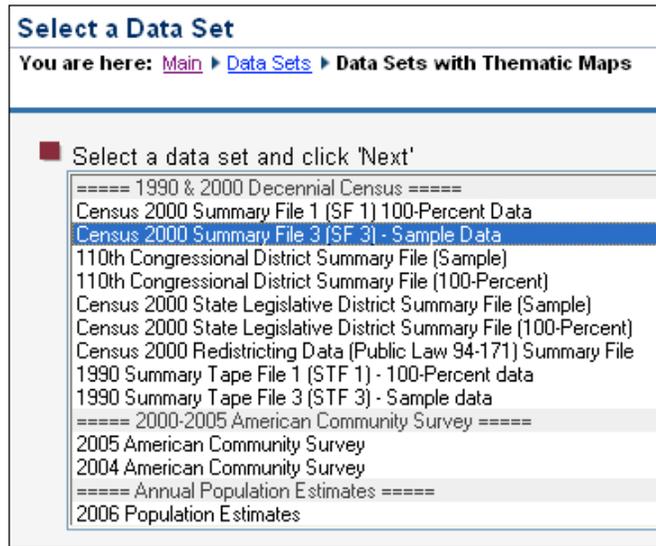
5. Experiment with a few of these map features.

6. When satisfied with the operation of the interactive map, select the link at the top of the map frame named *Data Sets with Thematic Maps*.

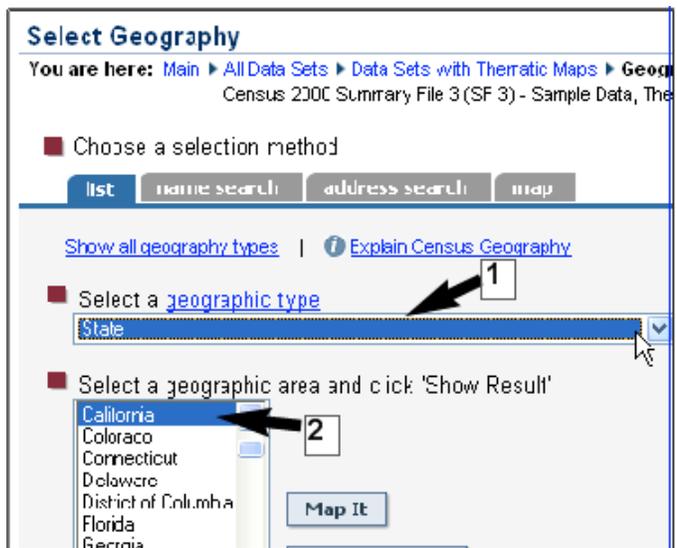


A window with a list of mappable data sets will open. The default file is *ACS* data from for the previous year.

7. Select the second item, *SF3*, and click the *Next* button.



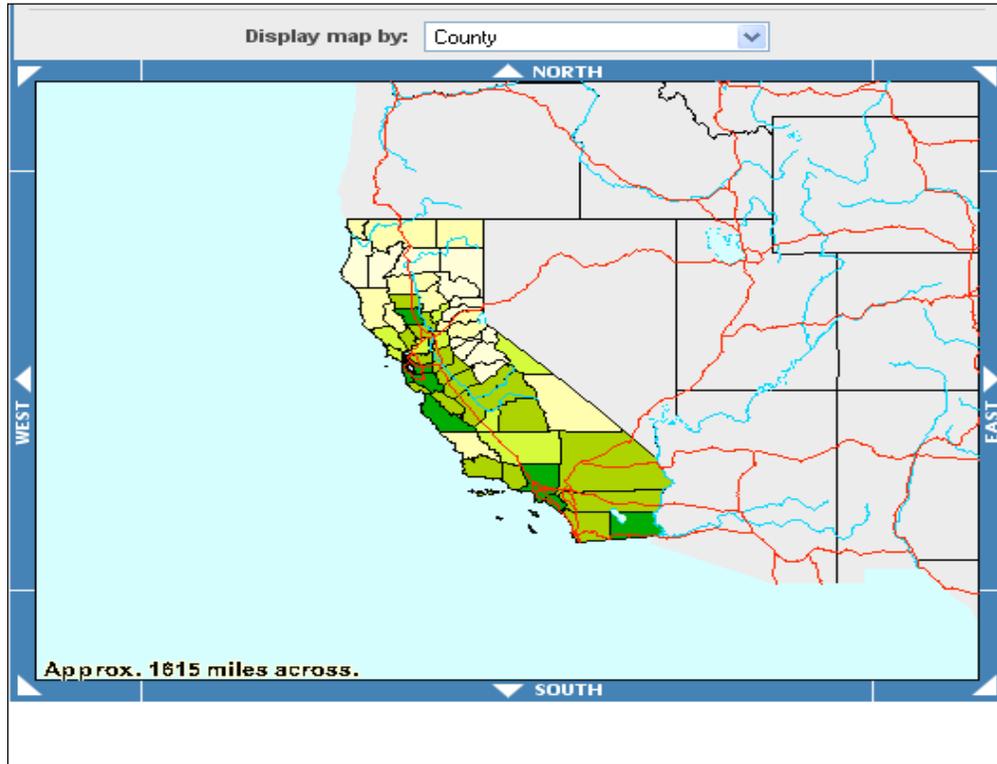
8. From the *Select Geography* window click on the arrow to the right of *Nation* (1) and scroll down to the *State* item in the popdown menu. Under *Geographic Area* select *California*. (2) Then click the *Next* button.



9. From the *Select Theme* window pick a variable of interest. Then click the *Show Result* button.



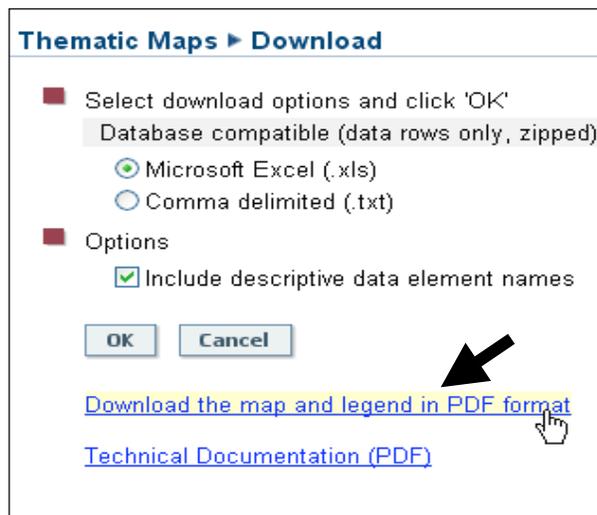
A map will appear (below is *Percent Foreign-born*). Note you may change the geographic detail from the *Display map by* window above the map. Adjust your map as desired and when done you may print it or save a copy in pdf format.



10. Under the *Print/Download* menu at the top of the page select the *Download* item.



11. From the *Download* window select the option *Download the map and legend in pdf format*. Then click *OK*.



12. When done, return to the main Census web page.

This concludes the introduction to the Bureau of the Census web site. Now would be a good time to review some of these materials, but choose your location or one of interest to you when

you explore the resources.

In the next exercise you will learn how to download raw data.

Exercise 2

Accessing Census 2000 SF3 and SF4 Data and Data Bases at ICPSR

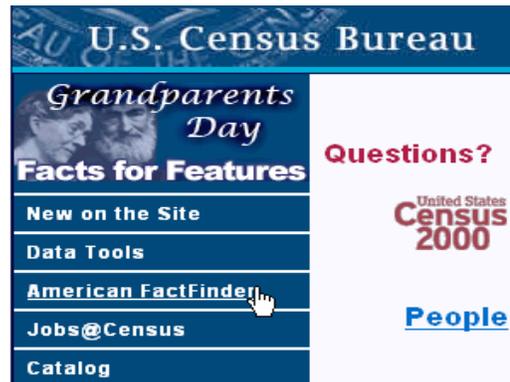
In this exercise you will learn how to do a basic download of census data from summary files. At the time of the actual download your results may vary because of the type of unzipping program you are using, the browser you are using, and what the security settings are on the browser. The Bureau of the Census uses popup features that *Internet Explorer* may not allow in its default setting.

Before proceeding with the exercise it is helpful to review the summary files provided by the Bureau of the Census. Complete-count data is found on SF1 and SF2. Sample data is found on SF3 and SF4. These latter files include data on income, education, occupation, ancestry, and housing that are commonly sought. Both SF2 and SF4 contain tables that are created for numerous race, Hispanic, Native American, and ancestry groups, but there must be at least 50 persons sampled in an area for it to be included.

Summary files contain data for multiple types of geography within a state. The user should know in advance what type of geography is needed and what tables are desired. There is a limit of 7000 areas per download. If the number of total items exceeds the 255 column limit of Excel, the tables will be split into multiple files.

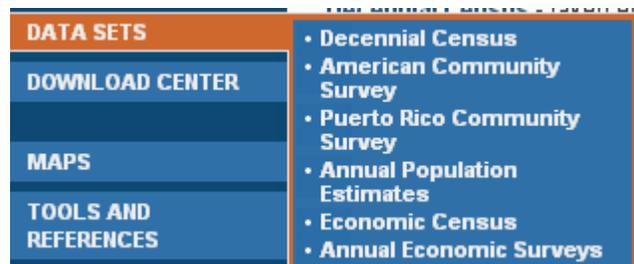
A. Accessing Summary File 3 on the Web

1. If you are on the main Census web page, locate on the left panel the link to *American Factfinder*.



Then from the left panel click on the *Data Sets* button.

From the list of options select the *Decennial Census*.



When the *Decennial Census* page opens you will see a tab for the last two censuses on the top of the page. See right.

Each of these will open a list of summary files for that decade. In the window right the first tab will contain the *Census 2000*.

2. For now, click on the button to select *Summary File 3*. Then click the *Detailed Tables* link to the right.

3. When the data extraction program opens the first thing you should do is select the level of geography you wish. Keep in mind you are limited to 7000 records so if you want all block groups in Los Angeles, you will have to download the raw data files.

Under *Geographic Type*: choose *County* (1).

4. After the window refreshes, choose *California* under *State*:(2).

5. Under the *Select one or more geographic areas...* click on *All Counties* (3) and then click the *Add* button (4) below the window. Wait for the program to display the counties in the window.

6. When the counties have been displayed, click the *Next* button (5). A list of tables will appear.

7. To check out the contents of a table, select one and click on the *What's This?* button to the right. If you want to add a table, select it and then click the *Add* button at the bottom of the window.

Select Tables

You are here: [Main](#) ▶ [Data Sets](#) ▶ [Data Sets with Detailed Tables](#) ▶ [Geography](#) ▶ [Tables](#) ▶ Results
 Census 2000 Summary File 3 (SF 3) - Sample Data, Detailed Tables

Choose a table selection method

[by subject](#) [by keyword](#) [show all tables](#)

Select one or more tables and click 'Add'

- P1. Total Population
- P2. Unweighted Sample Count of the Population
- P3. 100-Percent Count of the Population
- P4. Percent of the Population in Sample
- P5. Urban and Rural
- P6. Race
- P7. Hispanic or Latino by Race
- P8. Sex by Age
- P9. Household Type by Relationship
- P10. Household Size by Household Type by Presence of Own Children <18 Years

Abbreviations:
 Black - Black or Afri
 AIAN - American Ind
 NHPI - Native Hawaii
 SOR - Some Other F

[What's this?](#)

[Add](#) ▼

Current table selections:

[Remove](#)

[Show Result](#) ▶

8. Select the *PCT16* table (*First Ancestry Reported*) and click the *Add* button. Eventually the table will be listed in the window. Sometimes the operation of this program will take a little while and you can cause it to crash if you start clicking on various buttons after a step has been started.

9. To the right of the screen select a button labeled *What's this?*



You will find this window very helpful for examining the contents of any table in which you are interested. **Check out the table of ancestries and then close it.**

10. Now click *Show Result*.

Program: [Decennial Programs](#)
 Census: [Census 2000](#)
 Census Instance: [Census 2000 United States](#)
 [Census 2000 Puerto Rico](#)
 Data Set: [Census 2000 Summary File 3 \(SF 3\) - Sample Data](#)
 ▶ Table: **PCT16 ANCESTRY (FIRST ANCESTRY REPORTED) [111]**
 Universe : Total population

Subject Characteristics: [Ancestry](#)

Example:	Geography
Total:	
First ancestry reported:	
Acadian/Cajun	
Afghan	
Albanian	
Alsatian	
Arab:	
Egyptian	
Iraqi	
Jordanian	
Lebanese	

The program will then list part of the results for you to browse. (See below)

U.S. Census Bureau
American FactFinder

Main Search Feedback FAQs Glossary Site Map Help

Detailed Tables

You are here: [Main](#) > [All Data Sets](#) > [Data Sets with Detailed Tables](#) > [Geography](#) > [Tables](#) > [Results](#)

Use the links above to change your results | [Options](#) | [Print / Download](#) | [Related Items](#)

Note: use download to retrieve all selected tables and geographies

[PCT16_ANCESTRY \(FIRST ANCESTRY REPORTED\) \[111\] - Universe: Total population](#)
Data Set: [Census 2000 Summary File 3 \(SF 3\) - Sample Data](#)

geographies 1-10 of 58 [Next](#)

NOTE: Data based on a sample except in P3, P4, H3, and H4. For information on confidentiality protection, sampling error, nonsampling error, and definitions see <http://factfinder.census.gov/home/en/datatools/expst3.htm>.

	Alameda County, California	Alpine County, California	Amador County, California	Butte County, California	Calaveras County, California	Colusa County, California	Contra Costa County, California	Del Norte County, California	El Dorado County, California	Fresno County, California
Total:	1,443,741	1,208	35,100	203,171	40,554	18,804	948,816	27,507	156,299	799,401
First ancestry reported:	1,261,632	1,058	25,683	159,098	31,570	15,442	824,440	19,303	128,770	663,921
Acadian/Cajun	79	2	5	48	0	0	34	0	15	5
Afghan	7,917	0	0	13	0	0	2,900	0	33	8
Albanian	131	0	0	0	0	0	59	0	0	13
Alsatian	43	0	0	0	0	0	23	0	7	1
Arab:	6,199	0	72	608	70	0	4,285	21	284	2,248
Egyptian	666	0	0	32	0	0	640	8	11	184
Iraqi	147	0	0	0	0	0	98	0	0	2
Jordanian	197	0	0	31	0	0	170	0	0	5
Lebanese	1,322	0	27	178	15	0	1,011	0	99	53
Moroccan	175	0	1	16	0	0	86	0	0	3

In this case the first ten California counties have been listed. You can click the *Next* link to see the next ten. Note that the counties are listed as columns and each of the ancestries in a row.

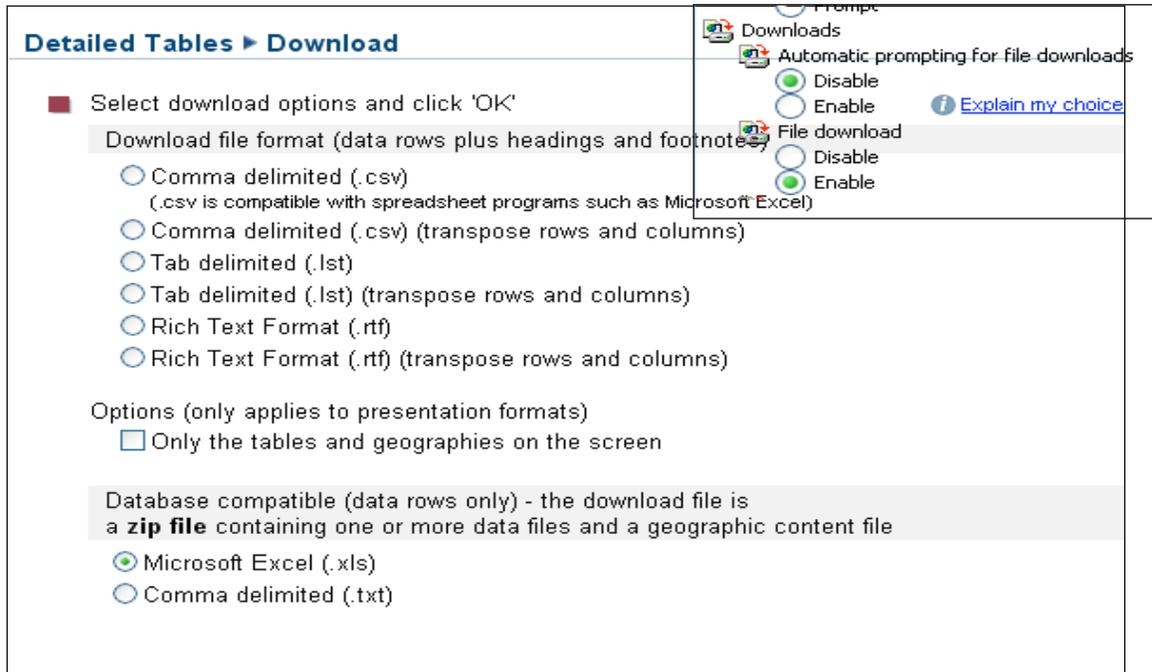
If you are only looking for some specific data you can locate a desired county and statistic and write it down or print out the contents of the page. However, in most cases you will want the contents of the entire table.

At this point you will notice that all counties are identified by their names and that no FIPS codes have been included. When you download this table a unique identifier (a combination of state and county FIPS codes labeled GEOID2) will be added to each row.

The provided FIPS codes may on occasion be converted in a spreadsheet from a character variable to a numeric variable and lose their leading zeros in subsequent processing. Thus, county code 005 would become 5 and code 037 would be 37. This is a nuisance, but simple to fix.

11. Select the *Print/Download* menu at the top of the screen and choose the *Download* option.

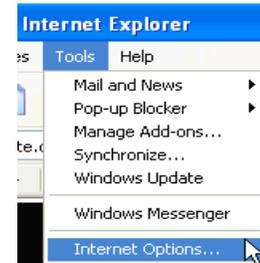
When the *Download* window opens, you may choose the *Database compatible* button for *Microsoft Excel* format (preferred) or you may choose the *CSV* version with rows and columns transposed. Then click *OK*.



Note that you will likely have download problems because of security settings in Explorer. You should look for a message at the top of the web page if the downloading does not seem to be taking place. Also, be sure to place the downloaded file in the *TEMP* directory, your personal directory, or a flash drive.

To enable downloads in Internet Explorer go to the *Tools* menu and the *Internet Options* option.

When the *Internet Options* window opens, click on the *Security* tab and then choose the *Custom Level* button.

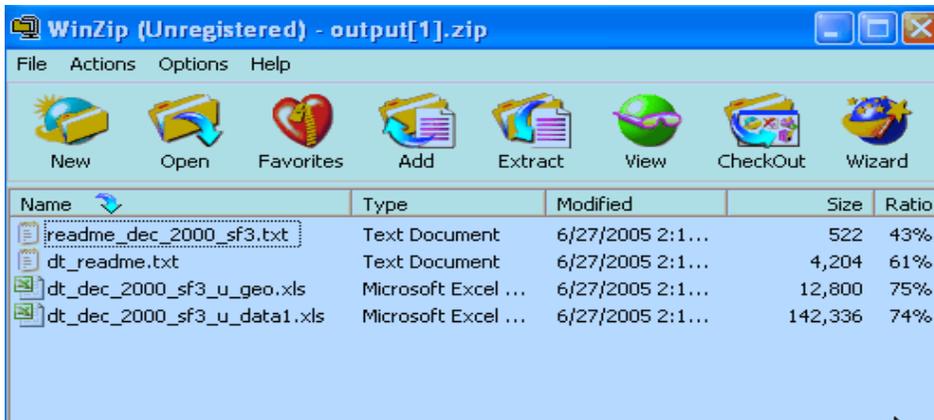


Scroll down through the list of options and make sure the file download is set to *Enable*.

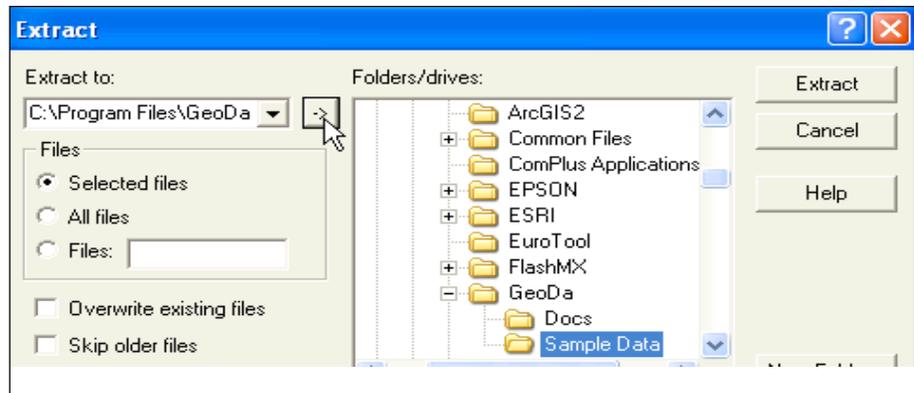
If you get a download warning you will need to again download the data after resetting the permissions.

B. WinZip

Once the download settings are correct the *WinZip* program will open with the contents of the Census download. Notice that you will get two *Excel* files. One (*geo*) contains the geographic descriptions for each county and the other (*data*) contains both the geography and the data values if you opted for the *Show Geographic Identifiers* under the *Options* menu. Read the text files if you want to.



1. Select all files. Click the *Extract* icon. When the *Extract* window opens (see below), click on the arrow shown below to navigate to a location open to you such as the *TEMP* directory or your personal directory.



2. ☒ After you have saved the files go to your directory and change the names of the geo and data files so that any future downloads from the Census Bureau will not over write on them. Unfortunately the names given by the download program are always the same.

3. Open the data file in *Excel* and look it over. You will use it in a future exercise.

This completes this exercise. **Close Excel**

C. The Geo within Geo Tab

An additional useful tool in selecting data for downloading is the *geo within geo* tab. This allows you to download a set of census units that are contained by a larger census unit. For example, you could download all tracts that fall entirely within a city or all block groups that fall within a county. Neither of these options is available under the default List method you just used.

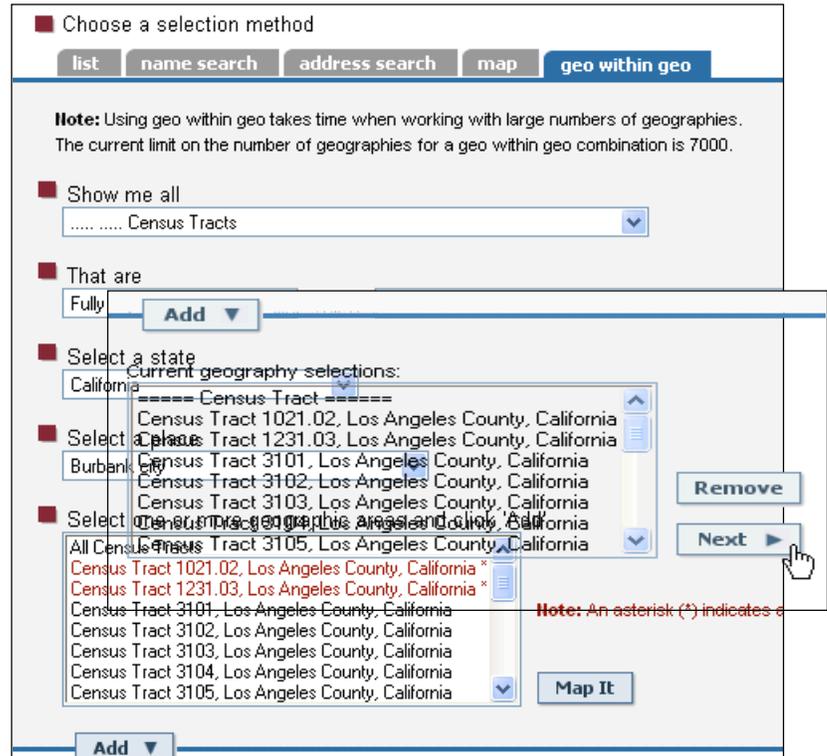
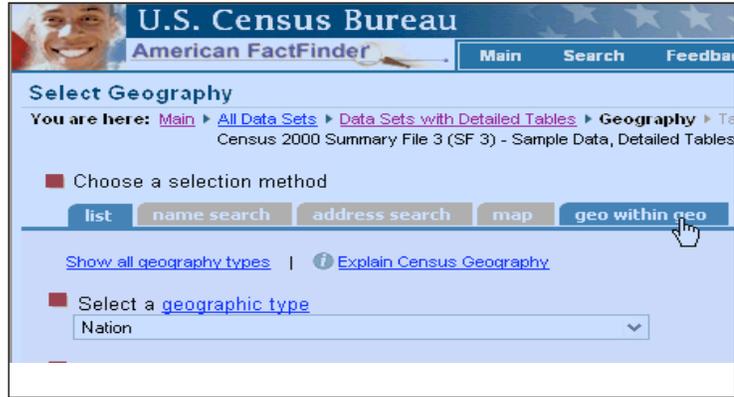
Note you do not need to do the following. Just read the steps to get a sense of how the program works.

If you had used the *Geo within Geo* you would have done the following steps.

a. Select the *geo within geo* tab.

b. From the geography selection window below under ‘*Show me all*’, *Census Tracts* were chosen. Under ‘*That are*’, the default *Fully or partially contained* was chosen. Under ‘*within*’, *Place* was chosen. Under ‘*Select a state*’, *California* was chosen. Under ‘*Select a place*’, *Burbank city* was chosen.

After the final selection all tracts that fall within the city are listed under the ‘*Select one or more geographic areas*’ window. Select the *All Census Tracts* option. Some tracts, such as the first two with brown letters, partially fall outside the city boundary. Thus, only those portions of tracts that lie within the city will be extracted. Note the content of the above window will change with your choices.



c. Click the *Add* button. All the tracts are listed

d. Click the *Next* button to select a table and download the data as before.

One thing to keep in mind here is that if you intend to map the city-only tracts that you must find a boundary file that contains only those tracts that fall within the city. These are generally not provided in free data sets. In most case only whole tract boundary files are provided and these often cross place boundaries, but not county boundaries.

D. Summary File 4

Summary File 4 contains more detail than other files, but most importantly, it is available for many individual ethnic groups. Because of this, there are some differences in the *American Factfinder* menus.

1. On the *American Factfinder* census web page select *Data Sets > Decennial Censuses*.

2. Make sure the *Census 2000* tab is selected and then scroll down and click the button to *Census 2000 Summary File 4*. Select the *Detailed Tables* link.

Census 2000 Summary File 4 (SF 4) - Sample Data
Summary File 4 contains tabulations of population and housing data collected from a sample of the population. The data are shown down to the census tract level for 336 race, Hispanic or Latino, American Indian and Alaska Native, and ancestry categories.

[SF 4 Thresholds](#)

Select from the following:

- [Detailed Tables](#)
- [Geographic Comparison Tables](#)
- [Quick Tables](#)
- [Reference Maps](#)
- [Custom Table](#)
- [Enter a table number](#)
- [List all tables](#)
- [About this data set](#)
- [Technical Documentation \(PDF\)](#)

This will open the data selection program.

3. For the *geographic type* select *County*

For *Select a state* choose *California*

For the *Select one or more geographic areas* choose *All Counties* and click the *Add* button.

Then click *Next*.

Choose a selection method

list name search address search map geo within geo

Show all geography types | Explain Census Geography

Select a geographic type
..... County

Select a state
California

Select one or more geographic areas and click 'Add'

All Counties
Alameda County
Alpine County
Amador County
Butte County
Calaveras County
Colusa County
Contra Costa County

Map It

Add

Current geography selections:

==== County =====
Alameda County, California
Alpine County, California
Amador County, California
Butte County, California
Calaveras County, California
Colusa County, California
Contra Costa County, California

Download data for more than 7,000 geographic areas using the [Download Center](#).

Remove

Next

4. From the list of tables add *PCT1*, *PCT89*, and *HCT2*.

Then click *Next*.

Choose a table selection method

by subject by keyword **show all tables**

Select one or more tables and click 'Add'

HCT1. Urban and Rural
HCT2. Tenure
HCT3. Tenure by Age of Householder
HCT4. Total Population in Occupied Housing Units by Tenure
HCT5. Household Size
HCT6. Tenure by Household Size
HCT7. Average Household Size of Occupied Housing Units by Tenure
HCT8. Tenure by Household Type and Presence and Age of Own Children
HCT9. Tenure by Household Type (Including Living Alone) by Age of Householder
HCT10. Tenure by Household Size by Age of Householder

Add ▾

Current table selections:

PCT1. Total Population
PCT89. Median Household Income in 1999 (Dollars)
HCT2. Tenure

5. The *Select Population Groups* window is unique to SF2 and SF4. It is here you will choose the ethnic groups for which you would like to obtain the tables. Note there is a tab for races and a tab for ancestries.

The tables for the total population have already been added to your list. To choose a particular group you first must click in the top window on the Table with which you want to work. In this example it is *PCT1*.

Race or Ethnic Groups Ancestry Groups

Select one or more tables

PCT1. TOTAL POPULATION [1] - Universe: Total population
PCT89. MEDIAN HOUSEHOLD INCOME IN 1999 (DOLLARS) [1] - Universe: Households
HCT2. TENURE [3] - Universe: Occupied housing units

Select one or more race or ethnic groups and click 'Add'

Total population
White alone
White alone or in combination with one or more other races
Black alone
Black alone or in combination with one or more other races
AIAN alone
.. American Indian alone
.. Alaska Native alone
AIAN alone or in combination with one or more other races
.. American Indian alone or in any combination

[Explain Race or Ethnic Groups](#)
[Print-ready Race or Ethnic Groups list](#)

Note: the following abbreviations are used to conserve space:
Black - Black or African American
AIAN - American Indian and Alaska Native
NHPI - Native Hawaiian and Other Pacific Islander

Add ▾

Current population group selections:

==== PCT1. TOTAL POPULATION [1] - Universe: Total population
Total population
==== PCT89. MEDIAN HOUSEHOLD INCOME IN 1999 (DOLLARS) [1] - Universe: Households
Total population
==== HCT2. TENURE [3] - Universe: Occupied housing units
Total population

Remove

Show Result ►

6. Scroll down the list of race groups and select *Chinese alone* and click the *Add* button. Note where it appears below.

Also select *Japanese* and *Korean alone*.

7. Click on the *PCT89* table in the top window. Then select and add *Chinese*, *Japanese*, and *Korean alone*.

8. Select the *HCT2* table in the top window and again select the three groups. Then click *Show Result*.

9. Scan down the list of tables to Median Household Income (*PCT89*). You will notice that values are missing for a number of counties because there was not at least 50 ethnic persons in the sample. Look over the table below and compare the median incomes of the three Asian groups. What group generally does better? Which does worse?

	Alameda County, CA	Alpine County, CA	Amador County, CA	Butte County, CA	Calaveras County, CA	Colusa County, CA	Contra Costa County, CA	Del Norte County, CA	El Dorado County, CA	Fresno County, CA
Median household income in 1999: Total	55,946	41,875	42,280	31,924	41,022	35,062	63,675	29,642	51,484	34,725
Chinese	58,707			36,125			74,424		82,802	47,083
Japanese	56,333			18,981			65,806		65,806	45,000
Korean	46,356						58,897			49,211

10. You can download the data if you would like to examine the incomes for more counties of CA.

E. Census Data at ICPSR

The Interuniversity Consortium for Political and Social Research (*ICPSR*) contains a wealth of census data and archives that could be valuable to anyone needing pre 1990 census information. Some of these data were originally compiled in digital form while other data represent special tabulations or collections that were converted from text to digital.

Unfortunately, use of this data will take much more effort since it is in raw form typically designed for computer tapes. In most cases there are not programs written to read the raw files and so a user will need to develop them using provided documentation.

1. To browse the ICPSR census data holdings enter the following in your browser:

<http://www.icpsr.org/>

The screenshot shows the ICPSR website homepage. At the top, the logo reads "ICPSR INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH". Below the logo is a navigation bar with five tabs: "Data", "Courses & Learning Tools", "Our Research", "Membership", and "About ICPSR". A banner image shows several people in a meeting. Below the banner is a "SEARCH" section with a search box, a dropdown menu set to "all fields", and checkboxes for "Web Site", "Data Holdings", and "Publications based on our data". There is also a "Search" button and links for "Advanced Search" and "Searching Tips". To the right is an "ANNOUNCEMENTS" section with an RSS icon and three entries: "2007.05.21 Subscribe to ICPSR announcements through RSS feeds...", "2007.05.21 New data releases...", and "2007.05.14".

2. Under the *Search* window select the *Advanced Search* link.

3. In the *Advanced Search* menu enter *census* under the first condition.

Search ICPSR's data holdings for documents that
 must contain in any field the words

4. Under the second condition enter *jail* and click on the option *must not contain*.

and
 must not contain in any field the words

5. Under the third condition enter *juvenile* and click on the option *must not contain*.

There are hundreds of data sets that involve jail censuses.

5. Click the *Search* button.

In the *Search Results* page there were still over 1500 data sets found. Note under each that there is a *description*, a *download* option, and a link to *related literature*.

6. Click on the *description* link to any data sets that seem interesting to you. You do not need to download any data sets unless you are prepared to create a program to read them. *SPSS* could be used to do this.

Search Results

If your search results are too broad, or if you're having trouble understanding the display of results below, you may wish to read our FAQ on [Searching the Web Site](#).

Results for: +census (1874) -jail (327) -juvenile (467)	Document count: +census (1874) -jail (327) -juvenile (467)
1543 results found, top 500 sorted by relevance score by date or sort by title 1-25 	
13569 Census of Population and Housing, 2000 [United States]: County-to-County Migration Flow Files	2007-03-07
United States Department of Commerce. Bureau of the Census description download related literature	
7529 Demographic, Social, Educational and Economic Data for France, 1833-1925	2007-02-16
Inter-university Consortium for Political and Social Research (ICPSR) description download related literature	
166 TIGER/Line Files at ICPSR: 1990 Census	2006-12-13
University of Michigan. Inter-university Consortium for Political and Social Research description	

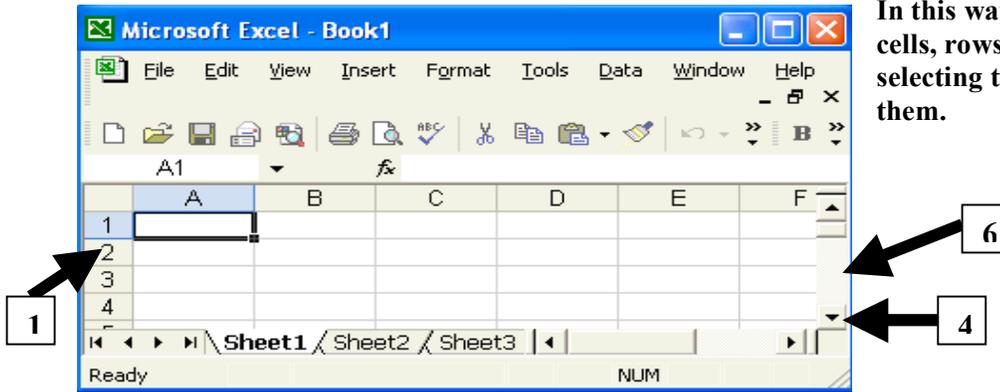
Exercise 3. Introduction to Excel

Purpose: This exercise will familiarize you with the basic operation of *Excel* and the procedures of data entry.

1. Select and open the *Excel* application icon.

5. With cell *A1* selected, hold the *shift* key and click on cell *D4*. In this way we can select an area of the spreadsheet for processing.

6. Select cell *A1* and then hold the *control* key and click on several other cells. In this way we can select specific cells, rows, or columns while not selecting those that lie between them.



Worksheet Features

Columns are identified with letters and rows with numbers. Note the characters *A1* in the small window to the left of the worksheet. (1) This is the currently active cell that is also indicated on the spreadsheet with a heavier outline. Also, the row and column of the selected cell are highlighted in blue on the spreadsheet margins.

At (2) are several tabs. You can have multiple spreadsheets within a file. Also, graphs you may generate can be on a sheet separate from the data. The chart will appear as a separate tab. To the left of the tabs are several arrows you can use to navigate through the different spreadsheets.

2. Select some different cells on the spreadsheet and note how these values change.

7. At (3) you can click on the corner and drag the edges of the spreadsheet to make it larger or smaller.

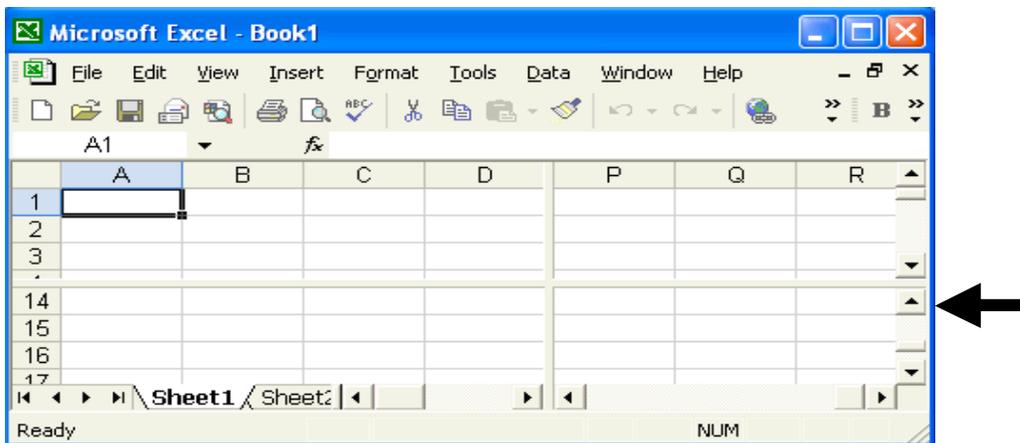
3. Now click on the margin letter *A* and note that all cells in the column are selected.

4. Click on the row number *1* and note all columns in that row are selected.

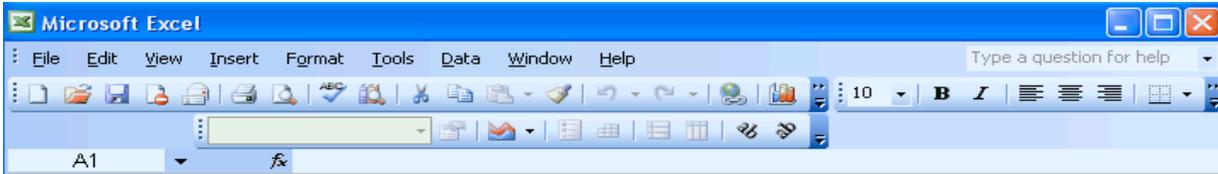
8. At (4) and (5) are two scroll bars. Click and drag each to move around the spreadsheet. If you click on the small arrows at the end of the scroll bars you can advance by single rows or columns. Remember you have a maximum of 255 columns and 15,756 rows to work with.

At (6) and (7) are two small bars. If you click and drag either you can divide the spreadsheet into two halves. This is a very useful feature for dealing with large census data files since you can click on the beginning row or column in the top or left sheet and then hold the shift key and click on the end of the row or column in the right or bottom spreadsheet. All cells in between your selection will be selected, though not visible, and you can fill out a formula without having to scroll over hundreds or rows or columns.

9. Drag the bar down to split the worksheet as shown in the illustration below. Then locate a similar bar on the bottom scroll bar and split the worksheet vertically. This feature is very important when performing calculations on large spreadsheets.



At this point you should be familiar with how to select a single cell, group of adjacent cells, or group of non-adjacent cells. You should be able to select an entire row and/or column and to scroll horizontally and vertically across the spreadsheet. You also should be able to divide the spreadsheet and to select a range of cells by clicking on a cell in each half of the divided sheet.

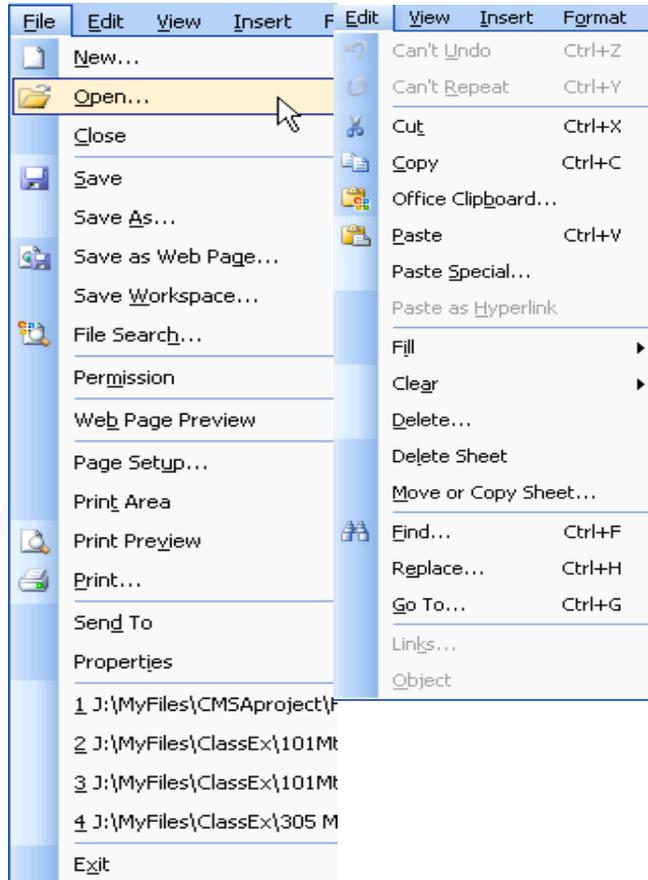


Look at the list of toolbars under the *View* menu and note which toolbars are now active (checked).

Menu Bar

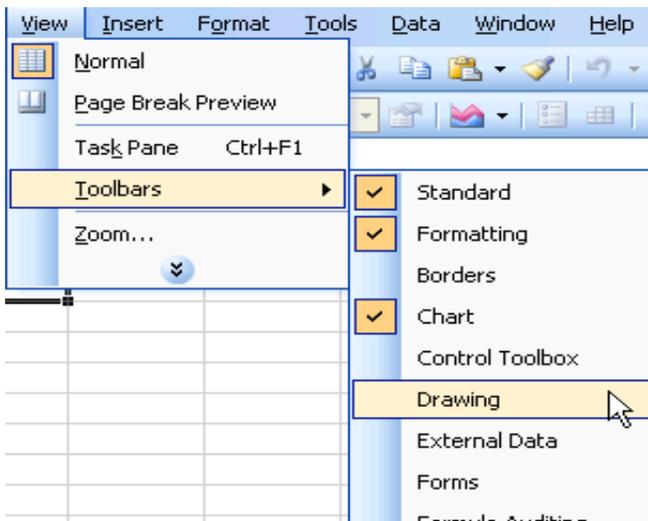
A menu bar extends across the top of the screen. It contains 10 pull down menus. These are indicated by the "Excel" symbol and the words: *File*, *Edit*, *View*, *Insert*, *Format*, *Tools*, *Data*, *Window*, and *Help*.

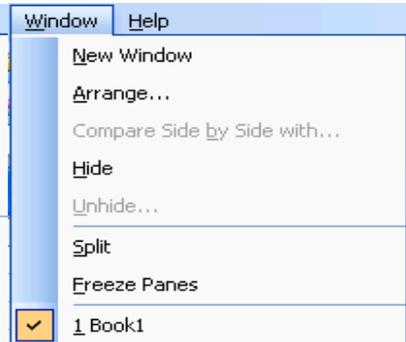
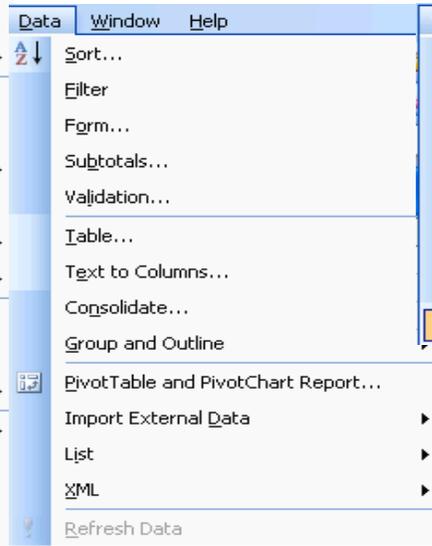
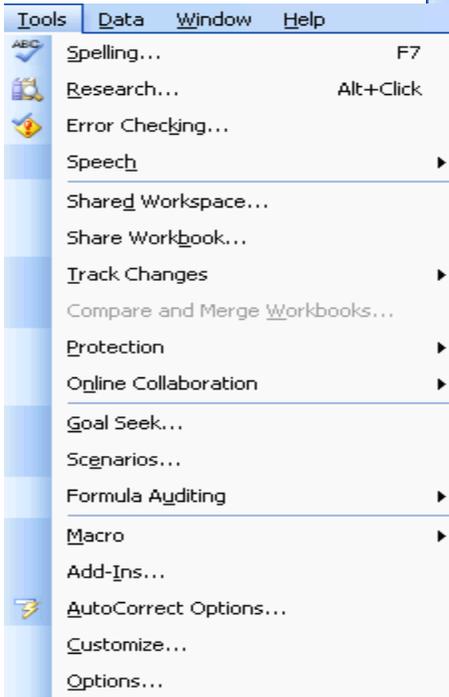
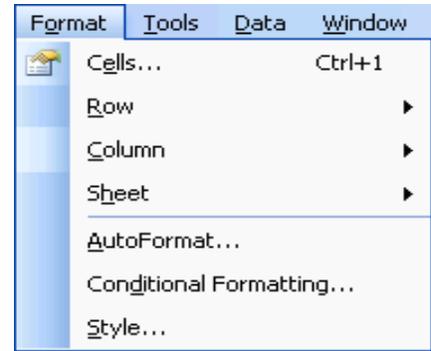
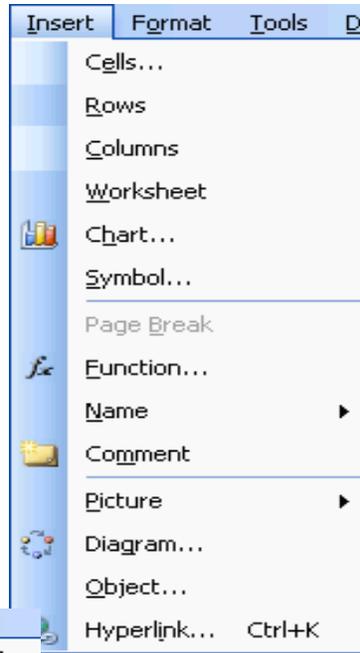
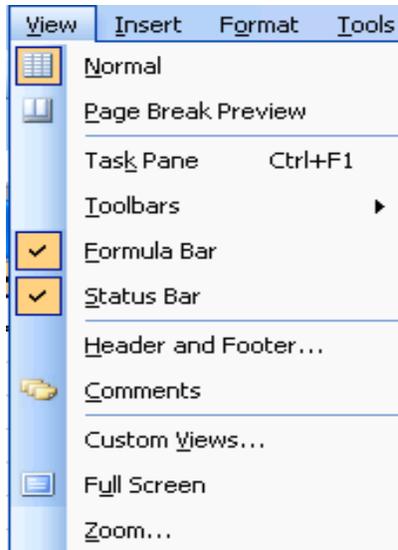
Select each *File* command in turn, except *Exit* with the selection arrow. Note that if two small arrows are visible at the bottom of the menu list that you may have to hold the cursor on them to see all menu options. For those options followed by a set of dots, an additional dialog box will appear with more options. Select the *Cancel* button for each of these additional dialog boxes.



Across the top of the spreadsheet and below the menu labels are a number of toolbar icons. More Toolbars may be invoked by selecting the *View* menu and the *Toolbars* option as shown right.

Select each of the other menus in turn from the menu bar and note the range of commands included in each.





	A	B	C	D	E	F	G	H
1	CNTY	ANPSADPI	AREALAND	P0010001	P0120001	P0120002	P0080003	P0080004
2	29	Kern County	21086815	543477	342300	28927	7239	16390
3	25	Imperial Coun	10813387	109303	32016	2573	1846	2171
4	37	Los Angeles C	10515327	8863164	3634722	946862	43689	955329
5	59	Orange Count	2045273	2410556	1557956	38825	12834	250136
6	65	Riverside Cou	18669111	1170413	757709	60063	11667	41556
7	71	San Bernardir	51961173	1418380	864830	110352	14271	58676
8	73	San Diego Co	10889601	2498016	1639845	150670	21509	198675
9	79	San Luis Obis	8558655	217162	177116	5066	2364	6319
10	83	Santa Barbara	7092635	369608	245074	9327	3889	16554
11	111	Ventura Count	4781006	669016	442015	14884	5041	34293
12								
13								
14								

B. Data Entry

1. Open a simple spreadsheet, *Ex3Excel.xls*, containing information related to several counties of Southern California.

2. The first thing is to replace the variable labels (columns) from the Census tables with more meaningful labels. Click on cell *A1*, and replace the variable name *CNTY* with *County*.

When you begin to enter information in a cell the contents appear in the cell and in an editing window shown in the menu bar. See *CNTY* above. Additionally, three icons will appear in the menu bar (see below); one is an "X" which if selected, cancels data entry in the cell. The other contains a "check," which if selected, loads the information into the cell. The last opens a window of functions that may be used to modify the contents of a cell.

	A	B
1	County	ANPSADPI
2	29	Kern County

3. In addition to *CNTY* replace each variable name with the more descriptive name shown in the list below.

Select cell	Contents
Replacement	
A1	CNTY County
B1	ANPSADPI Name
C1	AREALAND Area
D1	P0010001 TotalPop
E1	P0120001 NHWhite
F1	P0120002 NHBlack
G1	P0080003 Amer.Ind
H1	P0080004 AsianPI
I1	P0100001 Hispanic

In raw form, the census summary files are labeled with a table number and item number as shown in the figure above. These are being replaced with eight character names that will be preserved when this table is converted to a dbf format file.

Simple Calculations

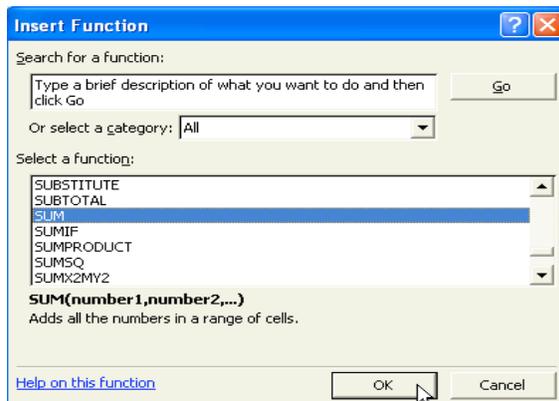
Excel can execute a broad range of arithmetic and statistical calculations which is why it is so popular for balancing check books and other numerical record keeping.. These functions are enabled through the creation of formulas in cells, columns, or rows.

1. Select cell *D13* and type in the "=" character. (see below)

	A	B	C	D
1	County	ANPSADPI	AREALAND	P0010001
2	29	Kern County	21086815	543477
3	25	Imperial Coun	10813387	109303
4	37	Los Angeles C	10515327	8863164
5	59	Orange Count	2045273	2410556
6	65	Riverside Cou	18669111	1170413
7	71	San Bernardir	51961173	1418380
8	73	San Diego Co	10889601	2498016
9	79	San Luis Obis	8558655	217162
10	83	Santa Barbara	7092635	369608
11	111	Ventura Count	4781006	669016
12				
13				=
14				

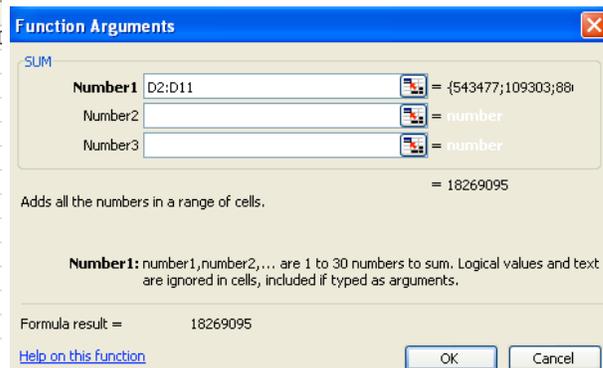
Anytime the "=" sign is entered, *Excel* expects a formula to follow. Note that as you type in cell *D13* the formula also appears in the formula bar at the top of the spreadsheet. If you wish to modify the formula later you will select the cell and then edit the equation in the formula bar, not the cell.

2. Open the *Insert* menu and select *Function...* option. A dialog box shown below will appear listing all the functions at your disposal. Scroll down to the function *SUM*. Check the explanation of the function below the Selection window. Then select *OK*



3. The *Function wizard* will appear to help you fill the necessary arguments for this function. In this case the function expects you to provide a range of cells to process. To do this you enter the beginning cell, a colon, and the last cell from a range of cells. The wizard will make a guess at what you want to do, but we do not wish to add cell *D12* since it is empty. Thus you will need to change the value of *D12* to *D11*.

Enter *D2:D11* in the *Number 1* box of the *Function wizard*.



Click *OK* and the function will appear in the formula bar and the value in Cell *D13*.

	A	B	C	D	E
1	County	Name	Area	Totpop	NHWhite
2	29	Kern County	21086815	543477	34
3	25	Imperial Coun	10813387	109303	3
4	37	Los Angeles C	10515327	8863164	363
5	59	Orange Count	2045273	2410556	155
6	65	Riverside Cou	18669111	1170413	75
7	71	San Bernardir	51961173	1418380	86
8	73	San Diego Co	10889601	2498016	163
9	79	San Luis Obis	8558655	217162	17
10	83	Santa Barbara	7092635	369608	24
11	111	Ventura Count	4781006	669016	44
12					
13				18269095	
14					

Please note that cell *D13* actually contains a formula, not the number shown. The cell is dynamically linked to the values above it and so if one of the cells above is changed, the sum will be updated. Also if you copy the contents of cell *D13* to another location you may get an error message because *Excel* can no longer find the values to process or you will get a wrong number because other columns are being accessed.

#REF!

To copy just the number to a new cell you must choose *Edit > Paste Special* and then click the *Value* option. This replaces the formula with the actual value. Keep this in mind since this operation is often used.

The equation also could have been typed directly in Cell *D13*. When complete, you would click on the "check box" to the left of the formula on the menu bar. The formula will disappear from the bar and the sum of the range *D2* to *D11* will appear in cell *D13*.

4. Select cell *D13* and drag to the right highlighting cells *E13* through *I13*.

5. Select *Edit > Fill > Right*. (See below) This will cause the original formula to be modified and inserted into each of the highlighted cells. In other words, we now know the sums of all the new columns without having to retype the equation. The sums of the ranges in columns *D* through *I* now appear. Click on some of the newly calculated cells and observe how the original formula values have been automatically modified.

	C	D	E	F	G	H	I
Area	Totpop	NHWhite	NHBlack	AmerInd	AsianPI	Hispanic	
21086815	543477	342300	28927	7239	16390	150558	
10813387	109303	32016	2573	1846	2171	71365	
10515327	8863164	3634722	946862	43689	955329	3306116	
2045273	2410556	1557956	38825	12834	250136	556957	
18669111	1170413	757709	60063	11667	41556	302286	
51961173	1418380	864830	110352	14271	58676	373632	
10889601	2498016	1639845	150670	21509	198675	498578	
8558655	217162	177116	5066	2364	6319	26869	
7092635	369608	245074	9327	3889	16554	97084	
4781006	669016	442015	14884	5041	34293	175414	
	18269095	9693583	1367549	124349	1580099	5558859	

6. Label cell *A13* as *TOTAL*.

7. Select cell *J2* and enter the following formula:

$$=E2 * 100 / \text{sum}(e2:i2)$$

(See below)

Select the check box next to the formula to accept it.

	C	D	E	F	G	H	I	J
Area	Totpop	NHWhite	NHBlack	AmerInd	AsianPI	Hispanic		
21086815	543477	342300	28927	7239	16390	150558	=E2*100/SUM	
10813387	109303	32016	2573	1846	2171	71365		
10515327	8863164	3634722	946862	43689	955329	3306116		

This is a very important and often-used formula for calculating percentages. Essentially it takes the contents of cell *E2* (Whites), multiplies it by *100* and then divides the result by the sum of all the population groups (to get a percentage).

It would also be nice if we could calculate the percentage of the total population for all the other groups. However, you must modify the equation before this will work.

6. Select *J2* and drag the cursor down highlighting cells *J2* through *J11*.

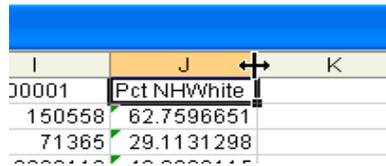
7. Select *Edit > Fill > Down*. This will cause the formula to be executed for each of the succeeding rows. Cells with formulas are linked dynamically to other cells.

	G	H	I	J
AmerInd	AsianPI	Hispanic		
7239	16390	150558		62.7596651
1846	2171	71365		29.1131298
43689	955329	3306116		40.9006115
12834	250136	556957		64.4660422
11667	41556	302286		64.580352
14271	58676	373632		60.8280857
21509	198675	498578		65.3512944
2364	6319	26869		81.3451275
3889	16554	97084		65.892861
5041	34293	175414		65.8106118
124349	1580099	5558859		

Note that each of the percentage values is different. This means that during the Fill Down procedure that each of the cell formula components was updated to access successive rows. That is, *J3* accesses columns *e3* to *i3*, *J4* accesses columns *e4* to *i4*, and so on. This feature makes calculations based on many rows very quick to execute.

8. Select *J1* and type *PNHWhite* Select the "check box" in the menu bar.

9. To enlarge column J, place the cursor on the vertical line separating column J from column K. (See right.)



A new cursor will appear as shown.

Press down the mouse button and drag to the right of the last letter in J1. This procedure allows columns to be enlarged, shrunk, or hidden.

10. To calculate the percent of the total population for the other groups you must make an important change to the formula. By adding a \$ sign before the cell identifiers for the total population the value can be fixed so that it will not increment upward whenever a Fill operation is done. Thus the number of persons in each group will be divided by the sum of the same set of cells. Enter the following in J2.

$$=E2 * 100 / \text{sum}(\$e2:\$i2)$$

11. Click on cell J2 and then hold the shift key and click in cell N2 so all between are highlighted. Select Edit > Fill > Right.

	J	K	L	M	N
1c	PctNHWhite				
50558	62.7596651	6.30367757	1.32724866	3.00505671	27.6043519
71365	29.1131298				
06116	40.9006115				
56957	64.4660422				
02286	64.580352				
73632	60.8280857				
98578	65.3512944				
26869	81.3451275				
97084	65.892861				
75414	65.8106118				
58859					

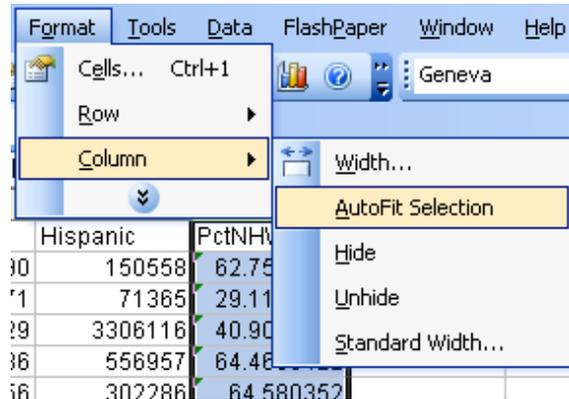
Click on some of the new columns and compare the numerator cells to the denominator cells. The denominator cells are fixed by the dollar sign while those in the numerator increment upward.

12. Now click on cell J2 and while holding the shift key, click on cell N11. Select Edit > Fill > Down. Percentage values will be calculated for all cells.

13. Place an appropriate label at the top of each percent column.

14. In the upper left corner of the spreadsheet click in the open box to the left of letter A and above the number 1 to highlight all cells in the sheet.

Select Format > Column > Autofit Selection. (See below.) This will ensure no values will be hidden - an essential property when working with dbf files later. Hidden numbers will be truncated when saved as dbf files.



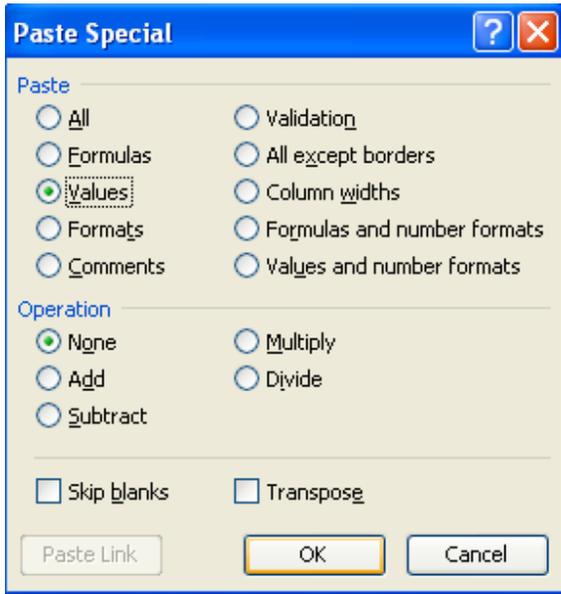
C. Removal of Formulas from Derived Data Ranges

Data manipulation often requires the elimination, insertion, and copying of cells, rows or columns. When such procedures are invoked, derived values may be altered because data contained in formula referenced cells has changed. In particular, when formula-based values are copied to new locations, their calculated contents will not be preserved.

In order to avoid this confusion the following procedure should be executed if automatic recalculation is no longer desired.

1. Select cell J2 and drag the cursor down to cell J11. Select Edit > Copy. This will cause the highlighted cells to be copied into the Clipboard (memory).

2. Now select *Edit > Paste Special....* A dialog box (below) will appear.

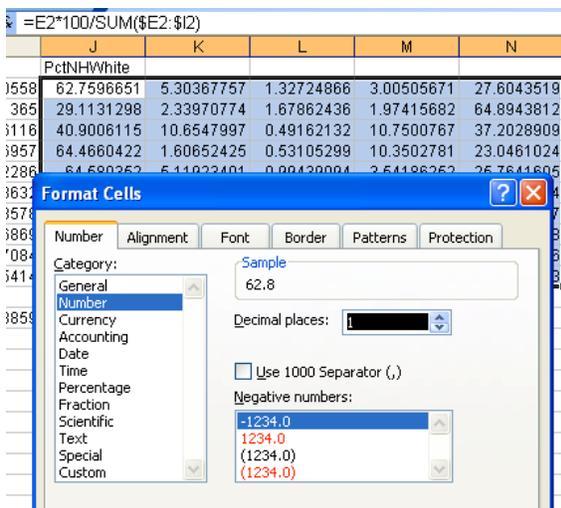


3. In the dialogue box select the *Values* button and click *OK*. This will cause only the values and not the formulas to be pasted into the cell range *J2* through *J11*.

D. Modifying Number Formats

Spreadsheet viewing frequently requires redefinition of number formats. For example cells *J2* through *J11* contain many more decimal places than necessary.

1. Select range *J2* through *N11*. Then select *Format > Cells*.

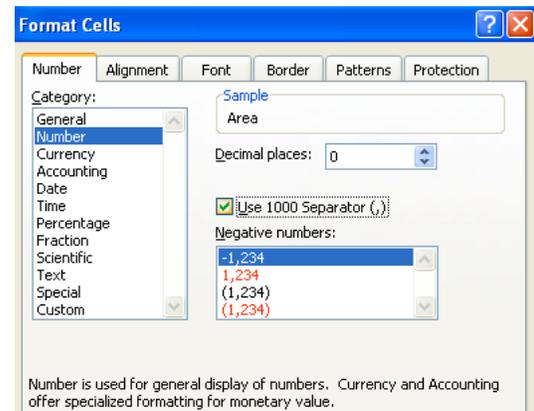


2. Select the *Number* tab and then under the *Category* window select *Number* and set the *Decimal places* to *1*. Click *OK*.

Note the percent columns have just one decimal. Make sure all your percent columns have just one decimal.

3. Now click on columns *C* through *I* so that all rows are highlighted. Again select *Format > Cells*.

In the *Format Cells* palette click the *Number* tab and the *Number Category*. Set the *Decimal places* to *0* and then click the *Use 1000 Separator* button. Then click *OK*.



4. Click on row *1* so that all columns in it are highlighted. Then select *Format > Cells > Alignment tab > Horizontal Alignment: > Center*.

Then select the *Font* tab and set the text to *Arial Bold*. Click *OK*. Note the changes.

To this point we have covered the basic steps for entering and modifying data in *Excel*. These include:

- Calculating a percent
- Entering a function in a cell
- Filling rows and columns
- Freezing cell values in an equation during a fill
- Replacing a formula with its numerical value
- Setting the number of decimals in a display
- Changing the cell font
- Aligning numbers in a cell

Don't hesitate to go back and review anything that seems unclear to this point.



Exercise 4. Analyzing Census Data in Excel

Purpose: Because Census data typically is provided in a spreadsheet form and usually in *Excel* format this exercise will cover some of the basic operations necessary to use *Excel* to analyze Census data..

Examining an Ancestry

1. Open the ancestry table (*STF3 PCT16*) that you created earlier in Ex. 2.

If you do not have the table, go to the Census web site and use *American Factfinder* to select *Summary File 3* and then table *PCT16* for all California counties.

	A	B	C	D	E	F	G
1	GEO_ID	GEO_ID2	SUMLEVEL	GEO_NAME	PCT016001	PCT016002	PCT016003
2	Geography Identifier	Geography Identifier	Geographic Summary Level	Geography	Total population: Total	Total population: First ancestry reported	Total population: First ancestry reported; Acadian/Cajun
3	05000US06001	06001	050	California	1443741	1261632	79
4	05000US06003	06003	050	California	1208	1058	2
5	05000US06005	06005	050	California	35100	25883	5

2. When you open the data table which will probably be labeled *dt_dec_2000_sf3_u_data1.xls* unless you changed it, note the various columns. Under *GEO_NAME* many of the county labels are hidden because of the narrow column width. Expand the column width to see all the names.

3. Scan across the columns and find an ancestry group you would like to examine in more detail. Avoid very small groups like *Acadian* since their low numbers are difficult to assess. Also, some categories like *Arab* are summaries of individual groups that follow.

When you have located a group, copy *GEO_ID2*, *GEO_NAME*, *PCT016001*, and your group to a new spreadsheet.

4. Compute the percent of the total population for each county represented by your group.

5. Change the column labels to eight character variable names and then delete the lengthy names in row 2. To delete the row, click on the number 2 to highlight the entire row. Then select *Edit > Delete*.

	A	B	C	D	E
1	GEO_ID2	County	TotalPop	Italian	Pctltal
2	06001	Alameda County, California	1443741	49120	3.4
3	06003	Alpine County, California	1208	60	5.0
4	06005	Amador County, California	35100	2262	6.4
5	06007	Butte County, California	203171	8941	4.4
6	06009	Calaveras County, California	40554	2338	5.8

In the figure above persons of Italian ancestry have been selected. Note in Column *B* the repetitious label of *County, California*. Since the file contains only counties we will delete the excessive wording.

6. Click on the top of Column *B* so that all of its cells are selected.

7. Select *Edit > Replace* and the window below will open. Enter *County, California* in the *Find what:* space. Click *Replace All*.

	A	B	C	D
1	GEO_ID2	County	TotalPop	Italian
2	06001	Alameda County, California	1443741	49120
3	06003	Alpine County, California	1208	60
4	06005	Amador County, California	35100	2262
5	06007	Butte County, California	203171	8941
6	06009	Calaveras County, California	40554	2338
7	06011	Colusa County, California	10000	1000
8	06013	Contra Costa County, California	100000	10000
9	06015	Del Norte County, California	10000	1000
10	06017	El Dorado County, California	156299	8500

Find and Replace

Find what: County, California

Replace with:

Replace All Replace Find All Find Next

Now your labels are much more practical.

	A	B	C	D	E
1	GEO_ID2	County	TotalPop	Italian	Pctltal
2	06001	Alameda	1443741	49120	3.4
3	06003	Alpine	1208	60	5.0
4	06005	Amador	35100	2262	6.4
5	06007	Butte	203171	8941	4.4
6	06009	Calaveras	40554	2338	5.8

These mass edits are frequently useful in creating useable labels when you have many values to process.

8. Select Data > Sort.

When the *Sort* window opens locate your group population in the *Sort by* window and then select the *Descending* button. Click **OK**.



This is a very useful function for isolating very high and low numbers in a distribution. Note that you can sort on more than one variable so, for example, you could sort by state and then by county within a state.

	A	B	C	D	E
1	GEO_ID2	County	TotalPop	Italian	Pctltal
2	06037	Los Angeles	9519338	214671	2.3
3	06059	Orange	2846289	107878	3.8
4	06073	San Diego	2813833	105667	3.8
5	06085	Santa Clara	1682585	71793	4.3
6	06013	Contra Costa	948816	53527	5.6

9. What are the leading counties with your ethnic population?

10. What are the counties with the fewest number of your ethnic group?

While it is interesting to note the numbers of ethnic persons it soon becomes apparent that in many cases the counties with the highest numbers ethnic also have the most people. Thus, we might want to control for population

differences to see where the proportion of ethnic persons is greatest and least.

11. Select Data > Sort.

This time sort by the percent ethnic column.

	A	B	C	D	E
1		County	TotalPop	Italian	Pctltal
2	06097	Sonoma	458614	34632	7.6
3	06041	Marin	247289	17654	7.1
4	06081	San Mateo	707161	46616	6.6
5	06005	Amador	35100	2262	6.4
6	06091	Sierra	3555	218	6.1

Sorting data this way is more interesting since it reveals something about the proportion of ethnic settlement. It seems that Italians are concentrated near San Francisco Bay and in nearby counties associated with wine grape growing.

One county, Sierra, has a relatively high percent Italian, but the numbers involved are quite low. Thus, in addition to examining the percent ethnic, it is important to also consider the magnitude of the numbers behind the percents.

12. Look over the percentages for your ethnic group. Do you know of any reasons why the percentages are high in some counties? You could check the Internet or your library for the history of settlement for your group.

For even more detail you can prepare a graph of your distribution. Look over the next section to see about making a graph.

13. Save your table for use later.

Graphing an Ancestry

While sorting data does give us access to specific information, graphing and mapping the data helps us get a sense of the general trends in value and location of the counties.

1. Click on **Column B (County)** and while holding the **control** key, click on **Column E (Pct Ethnic)** in your ethnic table.

2. Select **Insert > Chart**. This will begin a “wizard” with four screens to help you design the chart. The first screen is shown below. From the first window select the **Bar** type and click **Next**.

	A	B	C	D	E	F
1		County	TotalPop	Italian	PctItal	
2	06097	Sonoma	458614	34632	7.6	
3	06041	Marin	247289	17654	7.1	
4	06081	San Mateo	707161	46616	6.6	

3. From the second screen select **Next**.

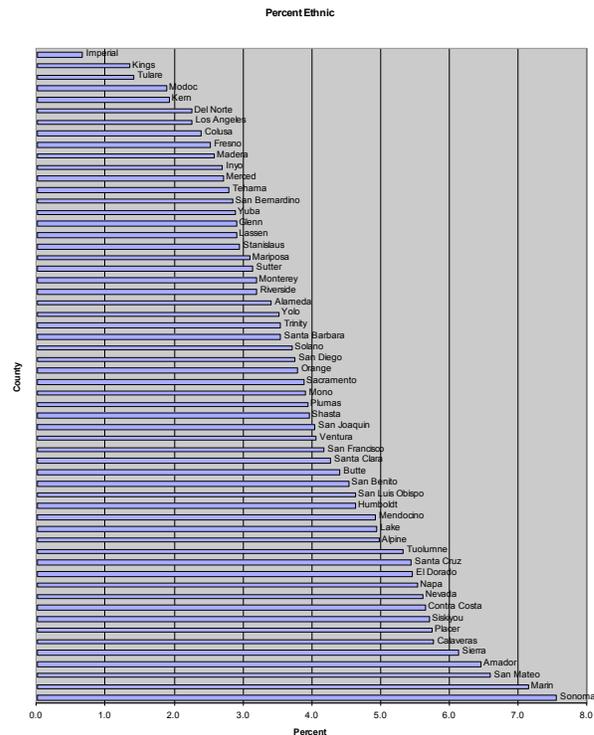
4. In the third step enter **Percent** in the **Y-axis** category and enter **County** in the **X-axis** category. (This seems the reverse of the display, but it may be a carry over from the **Column** chart design.)

Select the **Legend** tab and deselect the **Show Legend** button. It is not needed for just one variable.

Select the **Data Labels** tab and click the **Category Name** button.

Select the **Axes** tab and deselect the **Category (X) axis** button. Then click **Next**.

5. In the final step click on the **As New Sheet** button and then **Finish**. Your chart will appear as a new screen.



If your page layout is set to **Landscape**, select **File > Page Layout** and click the **Portrait** button to better visualize the distribution.

You can make various cosmetic changes to the look of your graph by double-clicking on the axes, lines, bars, or titles.

Note that a new tab, **Chart 1**, has been added to the bottom of your **Excel** window.

Look over your chart. Are there any extreme or unusual values? Any sudden changes in trends?

A Frequency Graph of Ancestry

Unfortunately *Excel* does not provide a tool to create a frequency graph. This graph shows the number of observations (counties) that fall within various categories that extend from the lowest values to the highest. The size of the categories (called *bins* in *Excel*) depends on the range of the data and the size of the paper the graph will appear on. The number of bins does have some impact on the look of the frequency graph, and so you might want to generate more than one version of the frequency graph.

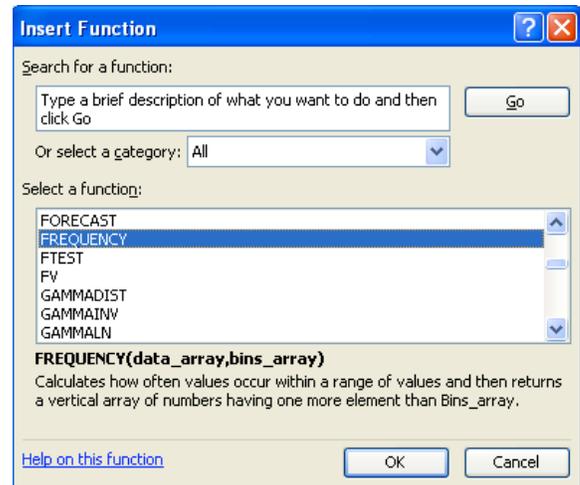
1. Open your percent ethnic table if it is not still open. Make a copy of your percent ethnic column and paste it to the right of the last column.

2. Sort the new column in ascending order to determine the range of the percentages. Fortunately, the orders of magnitude are much less than when dealing with raw numbers and this helps in setting the category interval. For the Italian data shown here the range of values extends from 0.7 to 7.6 percent. Thus we can set a category (bin) size of 0.5 to get 14 categories.

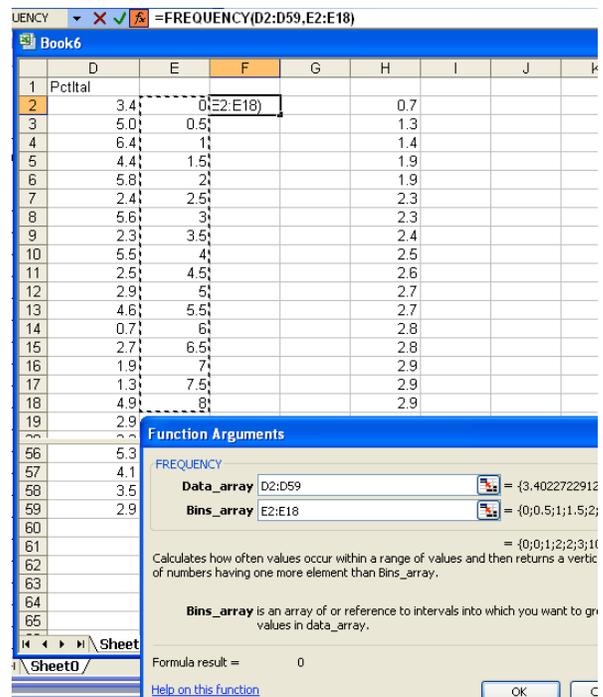
3. To carry out a *Frequency* function we need to set a series of category break values. In the column to the right of the percent column and next to the first data value enter a 0. Below that enter the equation $=d2 + 0.5$

Then fill down 16 rows so that the last value will be 8.0. Note the function is just a quick way to create a list of numbers. You could type them all in if you wish or if you choose to create irregular bin values.

4. Click in the column to the right of the bin values. Select *Insert > Function* and the *Insert Function* window will open. Under the *Or select a category:* window select *All*. Then under *Select a function* window scroll down to *FREQUENCY*. Note the explanation and example at the bottom of the window. For this function you will need your data array and the array of bins you just created.



5. Click OK.



6. In the *Function Arguments* window click on the *Data_array* window. Then in your spreadsheet click the first cell containing ethnic percent values. Then hold the *shift* key and click the last cell in that column. Here there are 58 cells, one for each California county.

7. Click on the *Bins_array* window and in your spreadsheet click on the first cell containing a bin value which is 0 here. Then hold the *shift* key and click the last bin value.

Then click *OK*. A single value will appear in the spreadsheet.

8. Select the cell holding the new frequency value and then hold the *shift* key and click to the right of the last bin value so all cells next to the bin values in the frequency column are highlighted.

9. Hit the *F2* key at the top of the keyboard.

10. Hold the *control* and *shift* keys and click the *enter* key. The frequencies for the remaining bins will be calculated.

	D	E	F
1	Pctltal	Bins	Frequency
2	3.4	0	0
3	5.0	0.5	0
4	6.4	1	1
5	4.4	1.5	2
6	5.8	2	2
7	2.4	2.5	3
8	5.6	3	10
9	2.3	3.5	5
10	5.5	4	10
11	2.5	4.5	5
12	2.9	5	6
13	4.6	5.5	3
14	0.7	6	6
15	2.7	6.5	2
16	1.9	7	1
17	1.3	7.5	1
18	4.9	8	1
19	2.9		

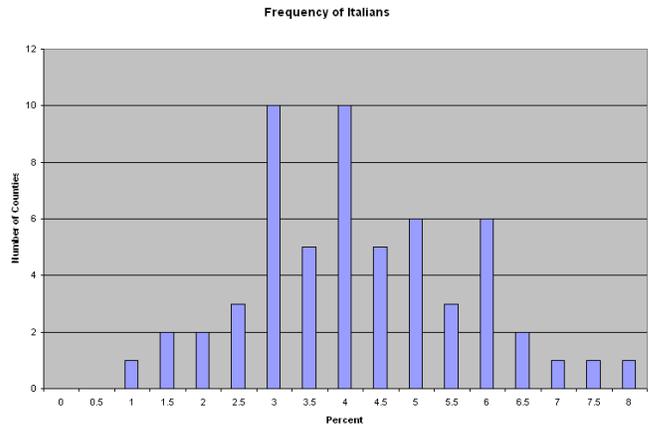
11. To be able to use the Bin values as labels you will need to convert the numbers to text. Click on the bin and frequency columns so all cells are highlighted in them. Select *Edit > Copy*. Then select *Edit > Paste Special* and be sure to hit the *Values* button. Click *OK*. This will remove the equations from the cells.

12. Now highlight only the bins column. Select *Data > Text to columns*.

In the *Convert Text to Columns Wizard* hit the *Next* button to get to *Step 3 of 3*. In that window click the *Text* button under the *Column data format* section. Then hit *Finish*. The numbers will shift leftward and now be considered as text.

13. Highlight the data values in both the *Bins* and *Frequency* columns. Then select *Insert > Chart*. At *Step 1* accept the vertical bar style and click *Next*. At *Step 2* click *Next*.

14. At *Step 3* under *Titles* enter *Frequency of [your group]*, enter *Percent* in the *X-axis* window, and *Number of Counties* under the *Y-axis* window. Under the *Legend* tab deselect the *Show legend* button. Then click *Next*. At *Step 4* click the *As a new sheet* button and click *Finish*. The frequency graph will appear.



Look over the frequency of counties across the categories. For the percent Italian shown above, the percents are quite normal looking with some potential breakpoints at 3.5 and 5.5

Exercises

1. Download from SF3 tables P159B, P159C, P159D, P159H, and P159I. These are Poverty Status in 1999 by Age for selected ethnic groups. Create a composite table for persons only below the poverty level from each of the ethnic groups. The table should be Race as the column variable by Age as the row variable. Each cell shows the number of persons below the poverty level.

Compute the percent of the total population that is in poverty for each group.

Compute the percent of the ethnic total in poverty for each age category.

Compare the race groups as to their poverty levels both for the totals and for the age categories.

2. Download from SF3 tables P145B, P145C, P145D, P145H, and P145I. These are Sex by Age for different ethnic groups.

Create a composite table of the total population for each of the ethnic groups by age. Collapse the age categories into larger categories of 0 – 17, 18 – 44, 45 – 64, and 65 and higher.

Compare the age structures for each of the ethnic groups. What groups have large numbers of young persons? Discuss some implications of this.



Exercise 5. Sex Ratio

Purpose: Because Census data typically is provided in a spreadsheet form and usually in *Excel* format this exercise will cover some of the basic operations necessary to use *Excel* to analyze Census data..

Calculating the Sex Ratio

1. Open the California cities file, *CalifCities*.
2. Create a new column showing the ratio of the number of males divided by the number of females.
3. Copy the City Name, Total Population, and Male/Female ratio to a new spreadsheet. Sort the cities by M/F ratio.
4. What cities have a greater proportion of males?
5. What cities have a greater proportion of females?
6. Are there any large cities with an extreme value?
7. Can you give any reasons for the rather extreme values?

Exercises

-Compute the sex ratio for persons under age 18 and over age 64. Normally the number of males exceeds females in the early years while the number of females is greater for older persons.

-Compute the ratio of female income to male income. For example, you could download median earnings for males and females (P85) to compare income by gender.

Compute the ratio of ethnic income to non-Hispanic White income. Download median household income by race and create ethnic to white income ratios for different groups. (p155A – P155I)

Exercise 6. Location Quotient

Purpose: The *Location Quotient* indicates whether a place has a greater or less than expected proportion of a subpopulation. Ideally, a place would have the same proportion as a larger reference area such as the entire state or the nation. In this case the proportion of the total population that is in an ethnic group will be examined. However, the measure is commonly used to gauge employment differences.

Calculating the Location Quotient

1. From the table below, calculate the proportion of the total population that is represented by each of the ethnic groups in the State of California.

Totalpop	NHWalo	Blackalo	Asianalo	Latino
33,871,648	15,771,163	2,219,190	3,682,975	10,969,132

2. Open the *CalifCities* file and then copy the City Names, Total Population, *NHWalo*, *Blackalo*, *Asianalo*, and *Latino* variables to a new spreadsheet.

3. Create four new columns for the ethnic groups and in each divide the city proportion ethnic by the proportion for the State. In other words for Non-Hispanic Whites the proportion white for each city would be divided by 0.4656, the value for the entire State.

4. When done, sort the cities based on the location quotient for each of the groups. Those places with scores above 1 have a greater than expected share of whites. A score of 2 would indicate twice as many whites as expected and a score of 0.5 would indicate half as many.

5. Can you offer any explanation of why some cities are very high or very low? For the latter, it usually means an especially high concentration of another ethnic group such as Black or Latino.

Exercises

1. Download occupational or industry data (*P50* or *P51*) and compute the proportion of males and females employed in different industries for California. Note you may want to use only the summary variables for the many values listed.

2. Use *SF4* to download the same data for selected ethnic groups (*PCT86* or *PCT87*) and compare ethnic ratios to those for all males and females in the entire State.

Exercise 7. Diversity Index

Purpose: There are several approaches for measuring diversity which is essentially the evenness in the proportion of several groups that comprise a total. In the case of ethnic data this would be an equivalent proportion of non-Hispanic Whites, Blacks, Asians, and Hispanics. Many more groups could be included such as American Indians or various Hispanic and Asian sub groups. However, the groups listed constitute the majority of ethnic categories.

Measuring Diversity

The Entropy Index

One measure of diversity is the *Entropy Index*. (H) This is essentially the sum of the terms that are comprised of the logs of the proportion of the total population for each group multiplied by the proportion of the population for each group. The higher the index, the more evenly balanced among the groups the population is.

$$H = -\sum \left(\frac{\text{White}}{\text{Totpop}} * \text{LN}\left(\frac{\text{White}}{\text{Totpop}}\right) + \frac{\text{Black}}{\text{Totpop}} * \text{LN}\left(\frac{\text{Black}}{\text{Totpop}}\right) + \frac{\text{AmInd}}{\text{Totpop}} * \text{LN}\left(\frac{\text{AmInd}}{\text{Totpop}}\right) + \frac{\text{Asian}}{\text{Totpop}} * \text{LN}\left(\frac{\text{Asian}}{\text{Totpop}}\right) + \frac{\text{Hispanic}}{\text{Totpop}} * \text{LN}\left(\frac{\text{Hispanic}}{\text{Totpop}}\right) \right) / 1.609$$

In this example there are a maximum of 5 groups which together generate a maximum score of 1.609 that is can be used to standardize the values of H .

This index increases with the number of groups that comprise the total, but may be reduced to a maximum of 1.0 by dividing all index values by the maximum possible index. This can be calculated by calculating the index for all proportions being equal.

1. Load the *CalifCities* file into *Excel*.

2. Copy the city names and the columns for non-Hispanic Whites, Blacks, Asians, and Hispanics to a new spreadsheet.

3. Enter the *Entropy Index* formula to the right of the last value. Don't forget the negative sign preceding the *SUM* function. Note only four terms are needed and the maximum value is now 1.386

4. Compute the *Entropy Index* for all California cities for the four major ethnic groups: non-Hispanic Whites, Blacks, Asians, and Hispanics.

5. Sort the cities in descending values of H .

Which places are most diverse? Which places are least diverse? Can you offer any reasons that explain what is happening in any of the cities? For example, does one ethnic group predominate? It might be interesting to look at other characteristics of the extreme places such as income, employment, and age to better understand their nature.

Exercises

1. Download all counties of the United States and determine which are the most and least diverse. Consider mapping this information to see if there are spatial patterns of diversity.

2. Select a county or city of interest and download all its census tracts. Compute the diversity and try to explain what might be causing variation. Consider mapping this information to see if there are spatial patterns of diversity

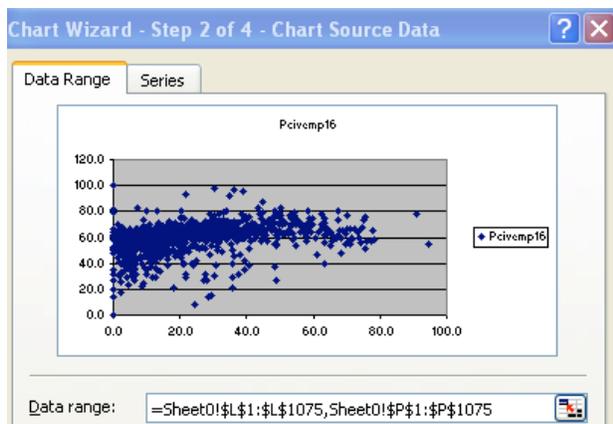
3. Calculate the diversity for 1990 and then calculate the change in diversity between the two censuses. Note increases can occur when a dominant group such as White or Hispanic is replaced by another group in an area that was previously low in diversity.

Exercise 8. Association Between Variables

Purpose: One important use of graphs is to look at the association of two or three variables. In most cases a scattergram of just two variables is created so that one can look to see if a change in one variable for an observation results in a systematic change in the second variable. Sometimes many variables are systematically paired in an array of scattergrams. One can then scan over the scattergrams to note any systematic patterns between various pairings of the variables.

Associations

1. Open *CalifCities* data file in *Excel*.
2. Select the columns labeled *Place*, *P25BAdeg*, and *Pcivemp16* so that all cells in each column are highlighted.
3. Select *Insert > Chart* and scroll down to *XY (Scatter)* in the first *Chart Wizard* window. Then click *Next*.
4. The second window will give you a quick look at your scattergram.



5. Click *Next* to move to *Step 3* of the *Chart Wizard*.
6. Under the *Titles* tab enter the title *Civilian Employed vs Education*. For the *X* axis enter *Percent Persons Age 25 and Older with a BA Degree or Higher*. For the *Y* axis

enter *Percent Persons Age 16 and Older Employed in Civilian Occupations*.

7. Under the *Legend* tab deselect the *Show legend* button. Then select *Next*.
8. At *Step 4* select the *As new sheet:* button and click *Finish*.

These two variables exhibit a fairly strong relationship since the points tend to form a positively oriented linear cluster. Thus, in cities with a higher percent of persons with at least a bachelor's degree there also tends to be higher percent of persons with civilian employment. However, for a given change in education there is not a lot of change in civilian employment.

Since there is a distinct trend in this data one might decide to apply correlation and regression methods to the variables.

While you might normally do correlation and regression in a program such as *SPSS* you can calculate the regression line in *Excel*.

9. Select *Chart > Add Trendline*.
10. In the *Add Trendline* palette under the *Type* tab click the *Linear* box if not already checked.
11. Select the *Options* tab and click the *Display equation on chart* button and then the *Display R-squared value on chart* button.

For these two variables the R^2 is 0.11 which is not very powerful for estimating employment from education. The trend line is $Y = 0.2055 + 55.041$

Exercises

1. Look through the *CalifCities* table to see if there are other pairs of variables that could be graphed in a scattergram. For example, education and income are strongly related.

Exercise 9. Dependency Ratio

Purpose: This exercise will give you the opportunity to calculate dependency ratios for California cities. These can be sorted and compared to look for similar locations or other qualities among the cities. In California one might expect cities with larger Mexican-origin populations to have higher Youth Dependency Ratios due to higher birth rates and cities that appeal to retirees to have higher Elder Dependency Ratios.

The Youth Dependency Ratio

1. Load *CalifCitiesAgeSex.xls* into Excel. This spreadsheet contains a set of age categories for males and then a repeat of the categories for females. You may find it helpful to split the spreadsheet for calculations.

2. In a new column to the right calculate the total population age 15 to 64 by adding all contained age groups for males with those for females.

3. In another new column to the right, calculate the number of persons 0 to 14 years of age by adding the three age columns for males with the three age columns for females.

4. Compute the Youth Dependency ratio for all the cities by dividing the age 0 – 14 population count by the age 18 – 64 population count.

5. Copy the YDR category and the city names to a new space and sort the cities by decreasing value of the ratio.

6. Which cities have the highest YDR and which have the lowest? Do any of these cities seem to have anything in common?

7. Compare the city values to those for the State of California and the U.S.

	United States	California	Los Angeles City	Glendale City
	Persons	Persons	Persons	Persons
Less than 15 years	60,253,375	7,783,683	839,417	36,030
15 - 64 years	186,176,778	22,492,307	2,498,274	131,829
Greater than 64 years	34,991,753	3,595,658	357,129	27,114
YDR	0.324	0.346	0.336	0.273
EDR	0.188	0.160	0.143	0.206
DR	0.512	0.506	0.479	0.479

The Elder Dependency Ratio

1. In a new column add all the age categories for males age 65 and older with all those for females age 65 and older.

2. Compute the Elder Dependency Ratio by dividing the age 65 and older population by the age 18 – 64 population.

3. Copy the EDR values and the city names to new columns and sort them by decreasing value of the EDR.

4. Which cities have the highest EDR and which have the lowest? Do any of these cities seem to have anything in common?

5. Compare the EDR and YDR cities.

6. Compare the high and low EDR values to those for the State of California and for the U.S.

The Dependency Ratio

1. Calculate an overall Dependency Ratio by adding the age 0 – 14 category with that of the age 65 and over category and then dividing the result by the age 18 – 64 category.

2. Again copy and sort the resulting calculation along with the city names in a new set of columns.

3. Compare the values at the ends of the ranking. Also compare the values with those for California and the U.S.

4. What types of needs can you foresee for areas that have either high EDR or YDR populations? In addition to the ratios you should look at the actual numbers of people behind the calculations.

Exercises

1. Use SF4 and download Table P8 for one or more specific ethnic groups by state to obtain the age/sex breakdown for those groups. Do various groups have distinctly different dependency ratios?

2. Calculate the dependency ratios for all states. What differences do you observe in different parts of the United States?

Exercise 10. Population Pyramids

Purpose: Population pyramids are a useful way of visualizing the age and sex structure of large populations. Within most populations the number of males exceeds that of females in the early years and then this reverses in the later years as males tend to die off more quickly. The pyramids also contrast developing nations with high birth rates and death rates with developed nations with lower birth rates and longer lived populations. Furthermore, abrupt changes in numbers between adjacent categories or cohorts indicate the effects of wars, migration, and general attitudes toward having more or fewer children.

Creating a Population Pyramid

It is possible to create a population pyramid in Excel with a little adjusting of the data values.

1. Open the file *CalifCitiesAgeSex* in *Excel*. This file has been modified so that all age categories are 5 year intervals. Furthermore all male values have been changed to negative numbers. This must be done so that the bars for males increase toward the left.

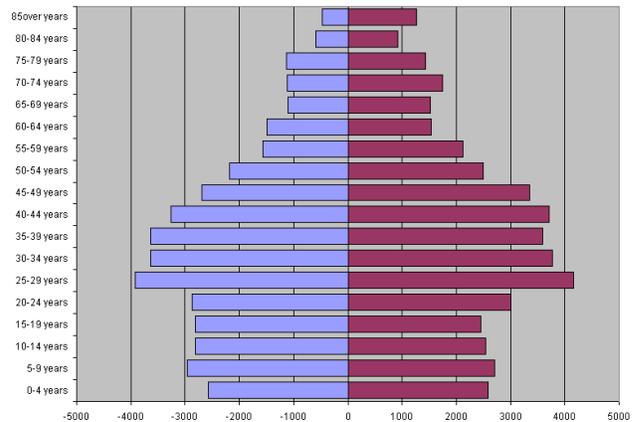
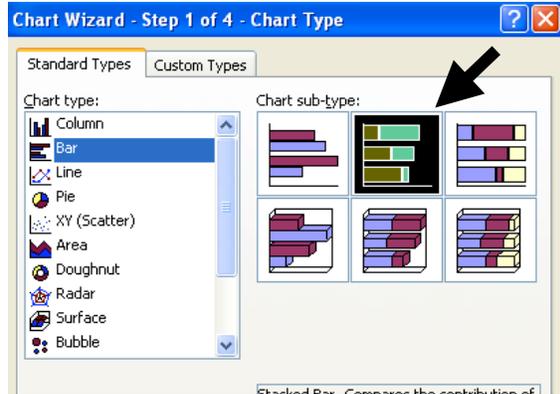
2. Copy a row of ages for a California city of your choice along with the column labels to a new file.

3. Copy all the female age groups in the new spreadsheet and paste them below the corresponding categories for males. See below.

	A	B	C	D	E	F
1	GEO_ID2	GEO_NAME	m0-4 yrs	m5-9 yrs	m10-14 yrs	m15-19 yrs
2	0600884	Alhambra city, California	-2573	-2954	-2809	-2818
3			2594	2715	2556	2470
4						

4. Highlight all cells from *C1* to *T3* and select *Insert > Chart*.

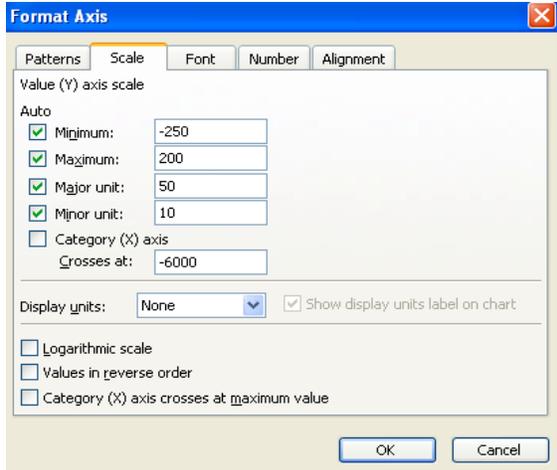
5. In the first window of the *Chart* wizard select the second type of *Bar* chart.



6. Proceed through the remaining windows to make the population pyramid.

Alhambra, shown above, has a sudden increase in the number of persons at age 25. Perhaps this is due to immigrant settlement or perhaps to people with children choosing to live somewhere else.

You can move the labels to the left margin by double-clicking on the bottom axis. When the *Format Axis* window appears, set the *Category X axis* value to cross at the lowest negative number shown. (-5000)



7. Repeat this for another city and compare the results for the two cities.

Exercises

1. Prepare a population pyramid for the entire state and compare it to your city.

2. Prepare a population pyramid for the entire U.S. and compare it to your state.

3. Download some age data for a developing country from the United Nations or the Population Reference Bureau and compare a population pyramid to that of the United States.

<http://www.prb.org/DataFind/datafinder7.htm>

<http://unstats.un.org/unsd/databases.htm>

Exercise 11. Population Growth

Purpose: This exercise will demonstrate ways of measuring change in growth of a population. Essentially two views are important. First is the absolute increase in numbers since it indicates where the overall impact of more people will be greatest. Second, is the percent increase or decrease in population. This indicates where the relative impact of growth will seem most pronounced. It is these latter areas that indicate shifts in people's behaviors and may indicate places that could become major population concentrations in the future as they catch up to slower growing places.

Absolute Population Change

1. Open the spreadsheet *UScoPop80-00.xls* This contains the total population and the Hispanic population for the last three decades. Note Alaska was not included due to large boundary changes.

2. In new columns compute the change in number of persons from 1990 to 2000 and from 1980 to 1990.

3. Select all rows in the spreadsheet and sort the counties in descending order of population change for 1990 - 2000.

4. List the top ten population gaining counties.

5. List the bottom population losing counties.

Where does population change seem largest? Use a map to find any counties that you don't recognize.

6. Copy the county names and the column based on the 1980 to 1990 population change to the right of the

previous values so that you can view both decades of change at once.

7. Again list the top ten population gaining and bottom ten population losing counties.

8. Are any counties included in both decades? Describe the changes that took place over the two decades.

Percent Population Change

1. In two additional columns compute the percent change from 1990 to 2000 and the percent change from 1980 to 1990. To calculate the percent change for 1990 to 2000 you would use the following formula $(\text{Pop } 2000 - \text{Pop } 1990) * 100 / \text{Pop } 1990$

2. Again sort the counties in descending value by percent change in population from 1990 to 2000.

3. Examine the top ten percent population gaining counties and the bottom ten population losing counties. Generally where are the gainers and losers?

4. Copy the county names and the percent change in population from 1980 to 1990 into two new columns on the right of the spreadsheet. Sort these values by declining percent change in population.

5. Are the same counties in the same parts of the United States appearing at the top and bottom of the sorted list?

6. Can you describe what generally seems to be happening?

Change in Share of the Total

1. Delete your calculations from parts A and B above if the spreadsheet is open.

2. Calculate the percent Hispanic population for 2000 and the percent Hispanic population for 1990.

3. Subtract the 1990 percent value from the 2000 percent value to get the difference in percentage points.

This difference indicates the increase or decrease in the Hispanic share of the total population.

4. Note the top ten gaining counties.
These increased their share of Hispanic persons.

5. Note the bottom losing counties.
These decreased their share of Hispanic population.

Exercises

1. Download the total population by state from SF1 for both 2000 and 1990. Compare the changes in total population and the percent change among the states.

2. Download the population by race and Hispanic by state for 2000 and for 1990. Compare the change in number and percent for one or more race groups.

3. Consider applying the above methods to education, poverty, foreign-born, or homeownership.

Exercise 12. The Demographic Equation

Purpose: This exercise will provide a basic introduction to the estimation of population change. Such procedures are used by the Bureau of the Census and the various states to estimate the composition and count of the population between censuses and for estimating what the likely values will be decades in the future. The models developed to do this can be quite complex as demographers try to account for detailed differences in the age, racial composition, birth and death measures, and migration within the overall population. Sometimes additional data sources like drivers license records, phone records, life expectancy tables, and membership in the armed forces are incorporated into the estimation process.

The basic three components of population change are births, deaths, and migration. The *Demographic Equation* expresses the interrelationship of these components.

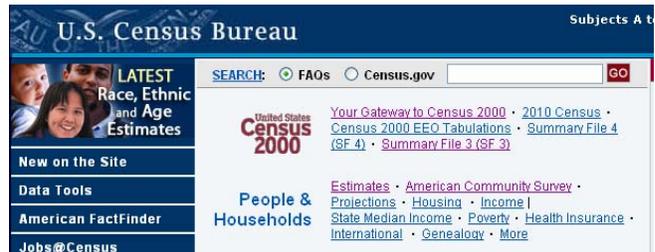
$$\text{Pop Change} = \text{Births} - \text{Deaths} + \text{In-migrants} - \text{Out-migrants}$$

The excess of births over deaths is called *natural increase* and the difference between in-migration and out-migration is called *net migration*.

For more discussion on this topic see: Current Population Reports: Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050. P25-1130
<http://www.census.gov/prod/1/pop/p25-1130.pdf>

U. S. Census Population Estimates

1. Go to the Census web site at www.census.gov and select the *Estimates* link from the *People & Households* category in the center of the page.

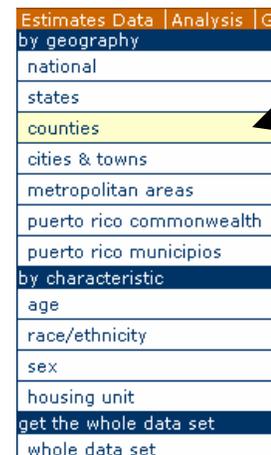


The population estimates web page contains a variety of resources related to the estimation of population in the U.S.

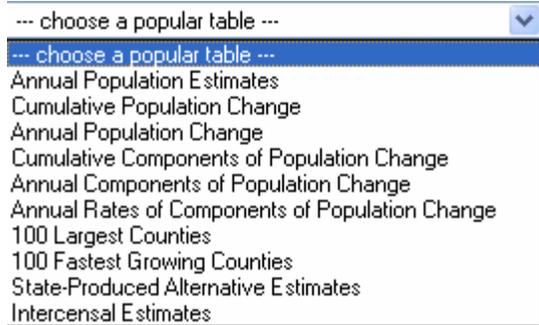


2. Check out the menus at the top of this page to see what is available.

3. Click on the *Estimates Data* menu. Select the *counties* link.



4. On the counties page select the *choose a popular table* window and select the *Annual Components of Population Change* link. Then click *Go*.



5. From the list of states choose *California* in *Excel* format. This will open in an *Excel* table.

Geographic Area	Total Population Change*	Natural Increase			Net Migration		
		Total	Births	Deaths	Total	Net International Migration	Net Internal Migration
California	303,402	325,333	561,364	236,031	-21,389	266,295	-287,684
Alameda County	6,361	11,358	20,985	9,627	-4,973	14,421	-19,394
Alpine County	42	1	9	8	41	-	41

6. Select the entire spreadsheet, copy it, open *Excel*, and then paste the sheet into a new spreadsheet within the *Excel* program.

Note the *Total Population Change* includes a residual value so if you add the components of change you will get a slightly different value for the total.

7. In a cell to the right of the last value for California calculate the total population change based on the demographic equation. Simply stated, it would be total natural increase + total net migration.

8. Select all data cells and sort the table by total population change. What five counties have the greatest change?

9. Resort the table by total natural increase. What five counties had the greatest natural increase?

10. Resort the table by net international migration. What five counties had the greatest net international migration?

11. Resort the table by net internal migration. What five counties had the greatest net internal migration?

12. Discuss what population changes seem to be taking place in California counties.

13. Compare change estimates in California to Arizona, Nevada, or Oregon.

14. Using the state totals compare changes in the estimated population for different ethnic groups in California.

Population Estimates from the California Department of Finance

The California Department of Finance provides a range of demographic and economic data about the state. The Department uses different procedures for estimating population and, in some cases, has arrived at different values than the Bureau of the Census. For example, the DOF has estimated a much lower loss of domestic population for California over the last few years. However, their intent is to focus on the state while the Bureau of the Census attempts to apply methods uniformly to all states.



1. Go to the California Department of Finance web page.

<http://www.dof.ca.gov/>

2. On the left panel select the *Demographic, Economic, and Financial Research* link. Browse some of the links on the resulting page to get a feel for what documents are available.



3. Under the *Demographic Research* heading click on the *Reports and Research Papers* link.

Estimates

- [E-1 City / County Population Estimates](#) with Annual Percent Change — January 1, 2006 and 2007 — **NEW**
- [E-2 California County Population Estimates](#) and Components of Change by Year, July 1, 2000–2006
- [E-3 California Race/Ethnic County Population Estimate](#) with Components of Change by Year, July 1, 2000–2004
- [E-3 California Race / Ethnic Population Estimates](#): Components of Change for California Counties, April 1990 to April 2000
- [E-4 Population Estimates](#) for California Cities and Counties, 1970–1980
- [E-4 Population Estimates](#) for California Cities and Counties, 1981–1990
- [E-4 Revised Historical](#) City, County and State Population Estimates, 1991–2000, with 1990 and 2000 Census Counts
- [E-4 Population Estimates](#) for Cities, Counties and the State, 2001–2007, with 2000 Benchmark — **NEW**
- [E-5 City / County Population](#) and Housing Estimates, 1991–2000, with 1990 Census Counts
- [E-5 City / County Population and Housing Estimates](#), 2007, Revised 2001–2006, with 2000 Benchmark — **NEW**
- [E-6 County Population Estimates](#) (.xls, 32k) — July 1, 1947–1969
- [E-6 County Population Estimates](#) and Components of Change (.xls, 224k) — July 1, 1970–1990
- [E-6 County Population Estimates](#) and Components of Change — July 1, 1990–2000
- [E-6 County Population Estimates](#) and Components of Change — July 1, 2000–2006
- [E-7 California Population Estimates](#), with Components of Change and Crude Rates, July 1,

4. Scroll down the list of tables to [E-6 County Population Estimates](#) and Components of Change — July 1, 2000–2006. Open the table.

5. Look over the components for the state of California.

	Population (July 1)	Percent Change*	Numeric Change	Births	Deaths	Natural Increase	Net Migration	Foreign Immigration	Net Domestic Migration
California	1999 33,418,578								
	Census 2000 33,873,086		225,654	129,778	54,032	75,746	149,908	59,110	90,798
	Apr-Jun 2000								
	2000 34,098,740	2.04	680,162	525,427	227,915	297,512	382,650	236,440	146,210
	2001 34,784,382	2.01	685,642	529,395	231,737	297,658	387,984	290,812	97,172
	2002 35,392,960	1.75	608,578	526,280	232,941	293,339	315,239	249,680	65,559
	2003 35,990,107	1.69	597,147	537,419	233,295	304,124	293,023	227,434	65,589
	2004 36,522,026	1.48	531,919	539,858	239,325	300,533	231,386	227,310	4,076
	2005 36,981,931	1.26	459,905	547,137	232,448	314,689	145,216	204,844	-59,628
	2006 37,444,385	1.25	462,454	551,855	234,977	316,878	145,576	212,921	-67,345
	Sum*		3,571,299	3,361,722	1,458,755	1,902,967	1,688,332	1,472,111	196,221
	Average*		571,408	537,876	233,401	304,475	266,933	235,538	31,395

The table below shows the figures of population change shown by the Bureau of the Census and the Dept. of Finance. Note the differences in some of the values.

	Population (July 1)	Percent Change*	Numeric Change	Births	Deaths	Natural Increase	Net Migration	Foreign Immigration	Domestic Migration
CA DOF 2006	37,444,385	1.25	462,454	551,855	234,977	316,878	145,576	212,921	-67,345
Census			303,402	561,364	236,031	325,333	-21,389	266,295	-287,684

6. Using the DOF data look up one of California's counties and note which components are adding or subtracting from the population.

7. In what year did your county gain most population? Is this part of a pattern?

8. Have births and deaths changed over the time period?

9. How has international and domestic migration changed?

10. Compare your data to that for the entire state. Are the patterns similar?

11. Use *Excel* to graph the percent change in population for your county.

12. Use *Excel* to graph the change in Net Domestic Migration. Can you offer any explanation for the trends?

13. Return to the DOF list of *Reports and Research Papers*. Scroll down and examine some of the other types of data tables available.

Exercise 13. Accessing Census 2000 PUMS Data

Purpose: The goal of this exercise is to extract some 2000 PUMS data for Asian Indians for PUMAs within California. You may either download the records for all states or filter the selection to a particular state or attribute. For states, you use the STATEFIP variable to limit your selection. What we would like to determine is if there are any notable changes in occupations between men and women in several selected states. You can determine this by calculating the percent employed in each occupation and noting the major occupational niches.

A. About PUMS

The *Public-Use Microdata Sample* is a collection of person and household records from the census of Population and Housing. This census file has become quite popular because it allows one to create custom tabulations.

The advantage of custom tabulation is somewhat offset by the limitations in geography. In 1990 and 2000 household records were tabulated by Public-use Microdata Area or PUMA. These units have a minimum of 100,000 persons. Another issue with PUMAs is that they often consist of disconnected areas. Thus, Glendale and San Fernando have been joined to form a PUMA. Apparently the designers tried to aggregate urban places into a PUMA before they would append adjoining rural or suburban space. The last 1990 PUMA in Los Angeles County is particularly poor - consisting of fragments from Signal Hill near Long Beach to Santa Clarita. For mapping purposes, there is a PUMA boundary file you may use to map tabulated variables.

PUMs data are available for a number of decades and these have been organized and integrated together by the Minnesota Population Center for easier access.

PUMs data may also be obtained in raw form from the Bureau of the Census. However, you will have to separate the housing and person records before making any tabulations.

More recent PUMs files consist of a sample of households and the persons in them. Usually these samples consist of a 1% national file and 5% state files, but a few other samples have been created such as a 0.1% national sample and a 3% sample for elderly persons. In 1980, estimates of the total population could be achieved by multiplying all records by a single factor, while in 1990 and 2000 each record must be multiplied by a weighting value.

A PUMs file consists of a housing record followed by person records for that household. The first person is the head of household, followed by the spouse, then children, and then others. The first column in each record identifies the record type with either an "H" or a "P" followed by the relevant data. The records are in text format and contain no delimiting characters. Thus one must be extra careful to properly specify field widths for each variable.

Because the housing and person records are nested together, one can not simply read a PUMs file into a spreadsheet and add the values in a column. The program must recognize each record type as it is input and possibly decide how to link household variables to the persons living there.

One common approach to simplifying the record processing is to link the housing data to each person in the household. This has been done by the *Minnesota Population Center*. The danger here is that you can not add housing variables to get a total since they have been repeated for every person. One can add housing data by selecting data only for persons who are heads of households. Another approach is to subset only household records from the raw PUMs data file.

1. Log in to the IPUMS web site:

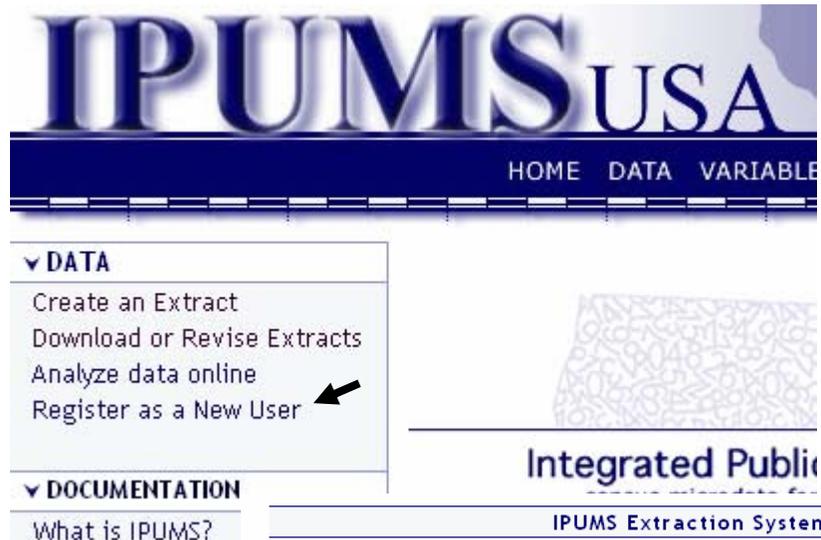
<http://www.ipums.umn.edu/>

2. Select the *IPUMS-USA* link.

In recent years *IPUMS* has also collected microdata for other countries and this could be a valuable resource for people who want to compare characteristics between the U.S. and other countries.

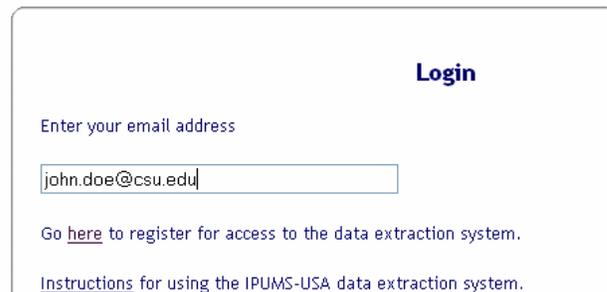


3. On the PUMS-USA page look for the *Data* links and register as a new user.



Then select the *Create an Extracts* link.

4. When the *Data Extraction System* program starts, you will need to enter your email address as the job name. Then click *Login*.



5. On the next page select the *Create New Extract* link. On the following page select the *Large* button and then click the *Continue to Sample Selection* button. We want the 5% sample.

IPUMS Extraction System
[logout](#)

Create a New Extract

Sample Density
[help](#)

Tiny Tiny samples contain approximately 5,000 adults and their households.

Small Small samples contain approximately 50,000 adults and their households.

Regular Most of the regular samples are 1% but the density of these samples varies. Refer to the [online documentation](#) for more information.

Large The large samples are 5% samples and are available for 1980-2000.

Regular & Large You can mix regular and large samples in the same extract

Sample Selection List
[help](#)

6. On the next *Sample Selection* page be sure to select the *2000 5% State sample* and then click the *Continue to Variable Selection* button.

Samples

1980 5% sample

1990 5% sample

2000 5% sample

Continue to Variable Selection

7. On the *Variable Selection* page are listed categories of household and person records. These are links to groups of variables that are listed below. Begin scrolling down this page.

You will note that some items are checked by default and the remainder must be selected as needed by you. Still others can be selected under *Case Selection* to limit the number of extracted records.

Technical Variables (Household) top					
Detailed Version	General Version	Variable	Label	Case Selection	2000
<input checked="" type="checkbox"/>		YEAR	Census year		x
<input type="checkbox"/>		DATANUM	Data set number		x
<input type="checkbox"/>		SERIAL	Household serial number		x
<input type="checkbox"/>		NUMPREC	Number of person records following	<input type="checkbox"/>	x
<input type="checkbox"/>		SUBSAMP	Subsample number	<input type="checkbox"/>	x
<input checked="" type="checkbox"/>		HHWT	Household weight		x
<input type="checkbox"/>		All Technical Variables			

8. Under *Geographic Variables (Household)* click the *STATEFIP* and *PUMA* items. Note that each household item will be appended to any person within that household later. To limit the records to a particular state now, you should select the *Case Selection* button to the right of the *STATEFIP* item.

Geographic Variables (Household) top					
Detailed Version	General Version	Variable	Label	Case Selection	2000
<input type="checkbox"/>		REGION	Census region and division	<input type="checkbox"/>	x
<input type="checkbox"/>		STATEICP	State (ICPSR code)	<input type="checkbox"/>	x
<input checked="" type="checkbox"/>		STATEFIP	State (FIPS code)	<input checked="" type="checkbox"/>	x
<input type="checkbox"/>		METRO	Metropolitan status	<input type="checkbox"/>	1, 5
<input type="checkbox"/>	<input type="checkbox"/>	METAREA	Metropolitan area	<input type="checkbox"/>	1, 5
<input type="checkbox"/>	<input type="checkbox"/>	CITY	City	<input type="checkbox"/>	1, 5
<input type="checkbox"/>		CITYPOP	City population		1, 5
<input checked="" type="checkbox"/>		PUMA	Public Use Microdata Area		5
<input type="checkbox"/>		PUMASUPP	Supp. Public Use Microdata Area		1, 5

9. Scroll down to *Demographic Variables (Person)* and select the *RELATE* and the *SEX Detailed* buttons.

Demographic Variables (Person) top					
Race, Ethnicity, and Nativity Variables (Person) top					
Detailed Version	General Version	Variable	Label	Case Selection	2000
<input checked="" type="checkbox"/>	<input type="checkbox"/>	RACE	Race	<input checked="" type="checkbox"/>	x
<input type="checkbox"/>	<input type="checkbox"/>	BPL	Birthplace	<input type="checkbox"/>	x
<input type="checkbox"/>	<input type="checkbox"/>	ANCESTR1	Ancestry, first response	<input type="checkbox"/>	x
<input type="checkbox"/>	<input type="checkbox"/>	ANCESTR2	Ancestry, second response	<input type="checkbox"/>	x
<input type="checkbox"/>		CITIZEN	Citizenship status	<input type="checkbox"/>	x

10. Under the *Race, Ethnicity, ...* category select the *Detailed Version* and the *Case Selection* buttons.

11. Under the *Work Variables (Person)* heading select the *Detailed Version* of the *Occupation* variable (*OCC*).

12. Under the *Income Variables (Person)* heading select the *Detailed Version* button of the *Total personal income (INCTOT)* variable. Later, if you wish, you can compare incomes for Asian Indian men and women and between occupations.

13. Go to the bottom of the *Variables Selection* page and click the *Continue* button.

14. The Case Selection page will open. This page allows you to limit the number of records according to one or more conditions. In this case we will limit the search to one state and to only persons of Asian Indian race.

15. For this exercise we will look at Asian Indians in California. Click on 06 California under the State (FIPS code) window and then scroll down to 610 Asian Indian under the Race (Person) window and select it.

Then select the Continue to Extract Request Summary button.

You will get a summary listing of the parameters set for this extract request.

State (FIPS code) (Household)

- 01 Alabama
- 02 Alaska
- 04 Arizona
- 05 Arkansas
- 06 California**
- 08 Colorado
- 09 Connecticut
- 10 Delaware
- 11 District of Columbia
- 12 Florida

Race (Person)

- 600 Filipino
- 610 Asian Indian (Hindu 1920-1940)**
- 620 Korean
- 630 Hawaiian
- 631 Hawaiian and Asian (1900,1920)
- 632 Hawaiian and European (1900,1920)
- 634 Hawaiian mixed

15. If all seems OK, enter a brief description of why you are extracting this information. In this case it is “to examine the differences in occupations among Asian Indian men and women.”

Then click the Submit Extract Request button.

16. You will receive confirmation of your extract request. You can monitor progress if you wish or wait about 15 minutes for an email notice to be sent to you.

17. Your email confirmation will appear similar to that at right.

Confirmation

Your extract request [eturner_csun_edu_032](#) has been submitted.

You will be notified by e-mail at eturner@csun.edu when it has been created.

When your extract is ready, [download](#) the data, codebook, and command files for your statistical software.

You may also return to the IPUMS-USA [homepage](#) or [create](#) a new extract.

Note: the extract will be available for 72 hours, after which time it will be deleted.

If you want to monitor progress, click the download link shown right. Hit the refresh button from time to time in your browser to update the file listing.

18. Be sure to download the Data, Codebook, desired Command files to a working directory. Note that IPUMS provides commands in SPSS, SAS, and STATA forms. We will use SPSS for this exercise.

IPUMS Extraction System [logout](#)

Extracts for Download: eturner@csun.edu

Click on the extract number in the leftmost column to revise that extract.
[Instructions](#) for opening an IPUMS extract on your PC.

Request	Created	Description	Data	Codebook	Command
031	May 14, 2007, 13:20:22	Looking at characteristics of people who left CA since 95	data	codebook	SPSS SAS STATA

The *data* file will appear similar to that below. Note that without the field descriptions it is useless.

```
File Edit Format View Help
00008212400820281700006001001210982013200006000011012100980332200012000002010524035000006000
30274000060000092123009204312000360000020147030290000060000062123006203911000270000010100620
00000101045200000000600000711240071024210000800000101173101300000600000981210098027110003600
40072100012000001001000999999006000008612100860052100006000001001000999999006000003712100620
00111121009804211000240000010159030670000060000050121006102911000120000010104000230000060000
999999006000017212101840201200006000002011476000090000600001721210184019120000600000201149400
0000101131300000006000007212400720222200006000002011936003200006000011112101100712100029000
04316000060000042141110470000060000086121006203111000270000010106230280000060000086121004904
00991710123027220001200000201052400300000600000711240071023110003600000101136501400000600000
```

19. Load the sps file into Word or a word processor
 The beginning of the SPSS program file appears right.
 The *Data list file* command provides a description of all the items and their field locations. You will eventually run this file in SPSS to input your data for analysis. However, you first need to make a couple of adjustments.

```
*IPUMS USER: note that you'll need to enter the path to the
*folder containing your data file in the program line below
*these comments.
*
*For instance, if your file is in the 'C:' drive in a
*folder called IPUMS, the "file =" statement below these
*comments would read:
*
*       file = 'C:\IPUMS\eturner_csun_edu_031'
*
*Please alter the following lines accordingly.

data list file = '<insert path to folder containing data file
here>eturner_csun_edu_031.dat' /
year          1-2
hhwt          3-6
statefip      7-8
```

20. Correctly set the path in the data list file command to the location of your data file and change the .dat to .txt. See below. To get the path, use Windows to locate the data file and then copy the path from the top of the window.

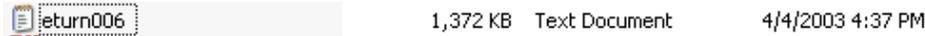
```
data list file = 'J:\MyFiles\CAAsianind\eturn006.txt' /
hhwt          1- 4
statefip      5- 6
```

Failure to get this correct will result in the following SPSS error message:

```
>Error # 31 in column 17. Text: J:\MyFiles\CAAsianind\eturn006.dat
>File not found.
>This command not executed.
```

21. You must open your data file (dat file) and resave it as a text file rather than an html file. In other words the file should have a .txt suffix.

Note that on some machines *Windows* has been set to suppress the file suffix in the listed name as is shown below.



22. If you would like to read this data into *SPSS*, proceed to the next exercise.

B. IPUMS Documentation

On the main *IPUMS* page is a link to documentation on the data. If you would like to know more about *IPUMS* and *PUMS* data select the *What is the IPUMS* link. At right is the *Users Guide* web page.

1. Under the *Contents of this page* links select the *Subject Content* link. The browser will jump to a discussion of the item.

2. Within the *Subject Content* discussion locate the “*Variable Availability*.” link and select it.

You will get a list of Person and Household variables.

3. Under the *Person Record* column select the *Race, Ethnicity, and Nativity Variables*.

Variable Availability	
A B C D E F G H I K L M N O P Q R S T U V W Y	
Household Record Technical Variables Geographic Variables Group Quarters Variables Economic Characteristic Variables Dwelling Characteristic Variables Appliances, Mechanical, Other Variables Constructed Household Variables Historical Oversample Variables Historical Technical Variables 1970 Neighborhood Variables	Person Record Technical Variables Family Interrelationship Variables Demographic Variables Race, Ethnicity, and Nativity Variables Education Variables Work Variables Income Variables Migration Variables Activity Five Years Ago Variables Disability Variables Veteran Status Variables Place of Work and Travel Time Variables Historical Oversample Variables Historical Technical Variables Other Variables

A matrix of variables and the decades for which they are available will appear. Note that because of the American Community Survey that values appear by year starting with 2000.

4. Locate the *BPL* variable (Birthplace) and select it. The *BPL* web page explains the nature of this variable across the various censuses.

5. At the top of the page is a link, *Codes and Frequencies*. Select it.

A code list of places starting with states will appear. These may be important should you not have labels for the code numbers in a program.

The default listing is *Category Availability View*. However, you can change this to *Case-Count View* to see how many records are available for each place as is shown below.

General Codes											
Case counts are for the most commonly requested regular sample for each year.											
<input type="radio"/> Category Availability View <input checked="" type="radio"/> Case-Count View											
	Code	1850	1860	1870	1880	1900	1910	1920	1930	1940	1950
UNITED STATES											
Alabama	001	3263	4698	9659	13436	19977	32537	25839	15337	34695	53331
Alaska	002	1	•	•	16	6188	5425	377	28	74	162
Arizona	004	•	•	18	87	5722	10140	1342	1102	3152	6993
Arkansas	005	682	1512	2874	5173	11270	21836	16514	9578	23506	35175
California	006	47	724	1731	3523	10375	21882	13836	10364	26371	56417
Colorado	008	14	•	69	355	2114	5139	4886	3346	8189	13365
Connecticut	009	4107	4789	4800	5387	6846	10803	9122	5750	13177	18772
Delaware	010	1050	1122	1305	1547	1705	2701	2005	1171	2404	3560

Exercise 14. Analyzing Census 2000 PUMS Data

Purpose: The goal of this exercise is to demonstrate how the *PUMS* data downloaded from the IPUMS web site can be brought into *SPSS* for processing. It is not the intent of this exercise to get too deeply into *SPSS* since that package can be a class in itself. However, you can read through the following pages to get an overview of how one might use *SPSS* on *PUMS* data.



Loading Data into SPSS

1. Locate and open the *SPSS* program.

2. When the *Data Editor* spreadsheet opens, cancel the option to open an existing file.

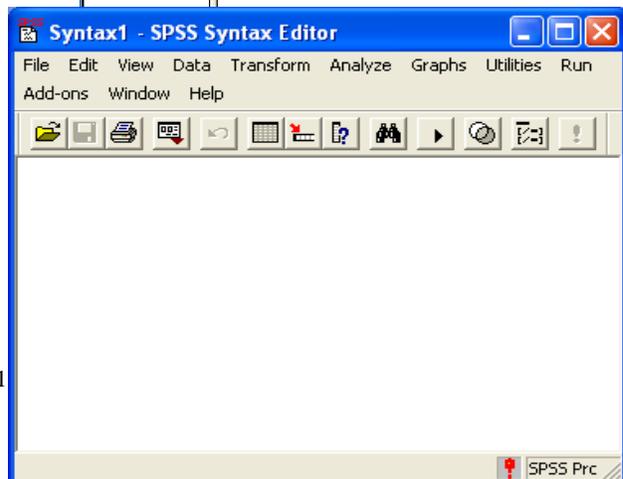
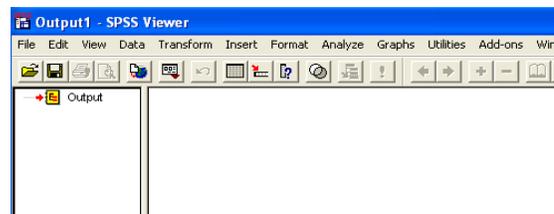
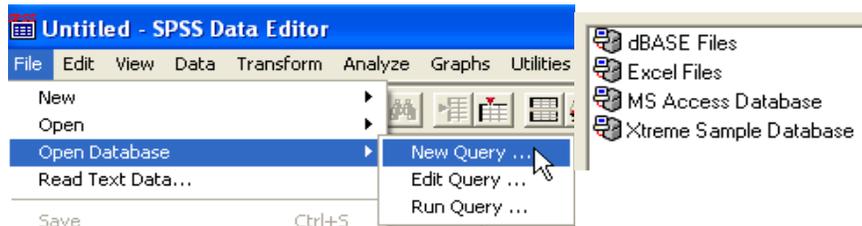
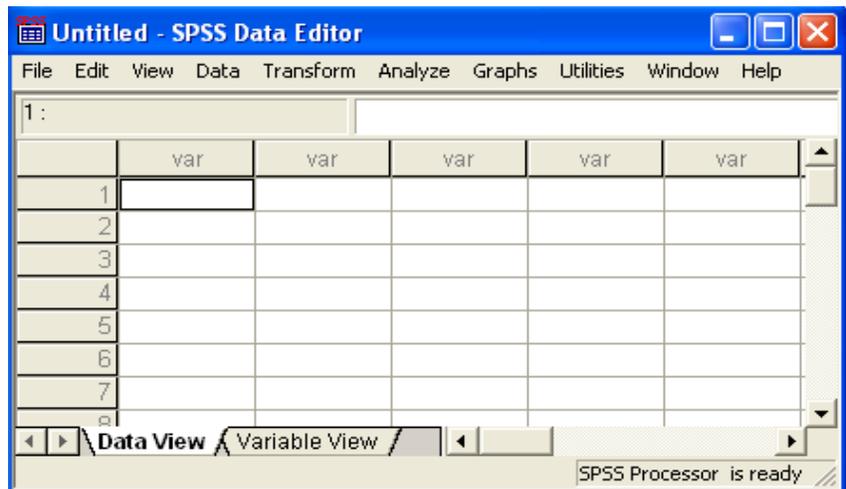
You will need to populate the database either manually or from existing sources.

SPSS can read several existing formats with the aid of a database wizard that is located in the *File* menu under *Open Database > New Query*.

However, your data is in text format and the field widths and column positions must be defined.

The *SPSS* program commonly uses several windows that depend on what tasks you need to perform. One of these is a *Data Editor* window shown at the top. Another is the *Output* window at right. Each window has accompanying tools.

In some cases one must revert to the old form of *SPSS* that is driven with various commands. For example, the *PUMS* data extracts from *IPUMS* are generated in a text



format that necessitates the use of *SPSS* syntax. Fortunately *IPUMS* also generates a syntax file to define and input the needed data. When read, the imported data can be saved as an *SPSS sav* file for future use. Data in this native format need only be opened.

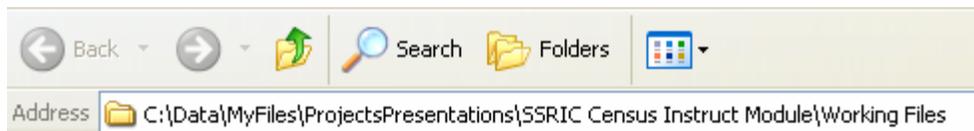
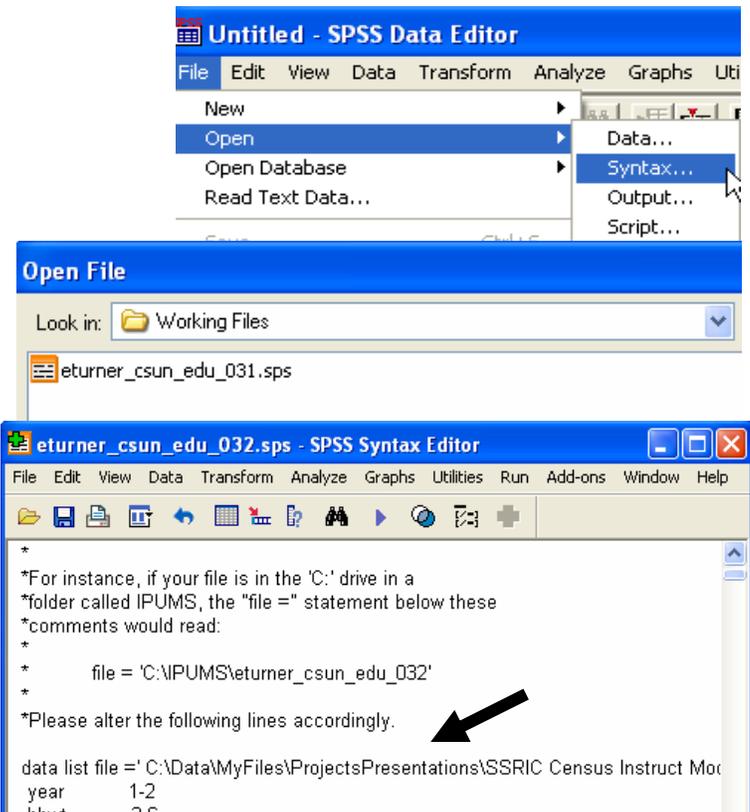
For program execution, *SPSS* provides a *Syntax* window (shown right) where necessary commands are specified in text form. Each ends with a period.

SPSS also provides a *Script* window for running customized programs.

2. Select *File > Open > Syntax*.

3. From the *Open File* window locate the *SPSS* syntax file (has *.sps* suffix) you downloaded from *IPUMS* earlier. Shown is *eturner_csun_edu_031.sps*. Click *Open*.

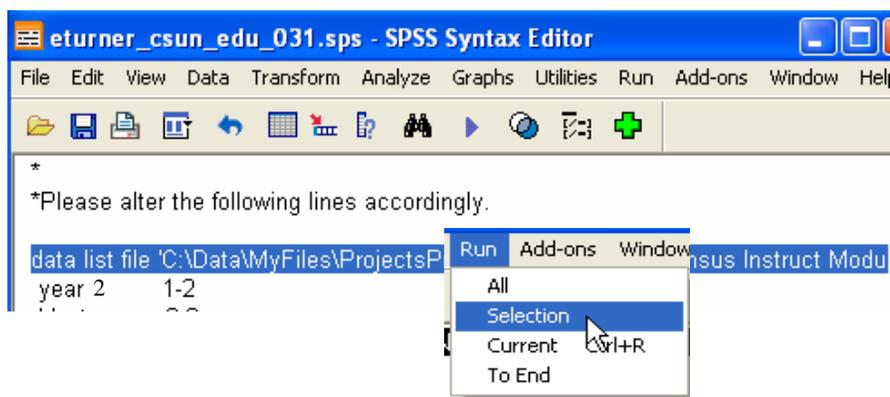
The *SPSS* command syntax will be listed in the *Syntax Editor* window shown right. From here you must change the path and file name so that the data can be loaded. Below, *Windows* was used to specify the path to the file location. The path can be copied and pasted into the first line in the *Syntax Editor*. Note the revised path below:



just the first line and then select *Run > Selection*.

SPSS should be able to locate the data and happily list out in an *Output* window the variables from the

4. Highlight

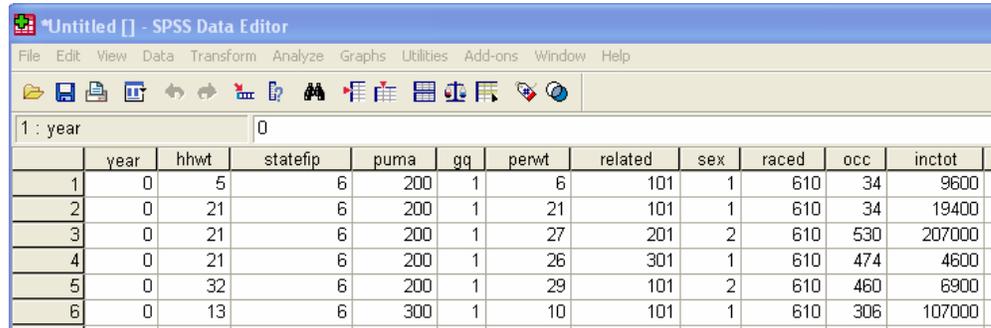


executed *Data List* command.

5. If all seems OK, select *Run > All* option.

SPSS should populate the spreadsheet. This may take awhile since PUMS files may be quite large, up to a million records in some cases. Note that occasionally you have to embed an *SPSS* command such as *Frequencies* at the end of the syntax file to make the command file finish executing properly.

At right is the populated *Data Editor* window.



	year	hhwt	statefip	puma	qq	perwt	related	sex	raced	occ	inctot
1	0	5	6	200	1	6	101	1	610	34	9600
2	0	21	6	200	1	21	101	1	610	34	19400
3	0	21	6	200	1	27	201	2	610	530	207000
4	0	21	6	200	1	26	301	1	610	474	4600
5	0	32	6	200	1	29	101	2	610	460	6900
6	0	13	6	300	1	10	101	1	610	306	107000

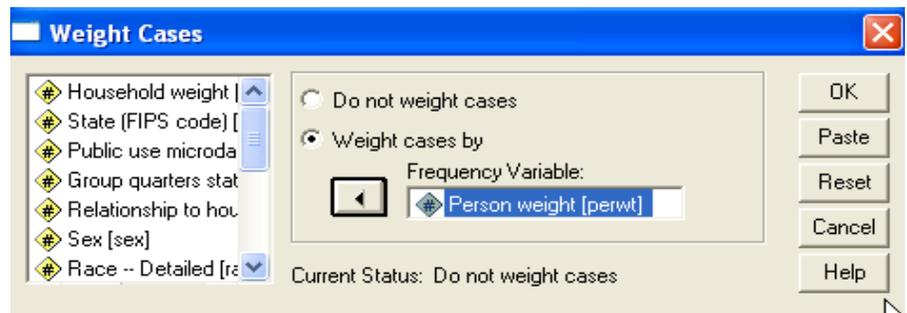
6. Now select *File > Save As*. When the window below

opens, name your file and save it in your work space. This creates an *SPSS.sav* file that you can use to directly open your data in the future. At this point you are ready to begin processing your data.

Analysis

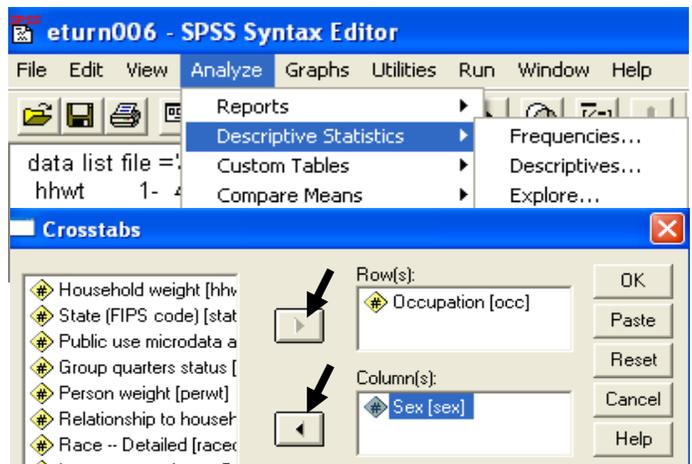
Our first goal is to determine how Asian Indian men and women have taken different jobs in different states. Select the FIPS codes for two states.

1. In the *Data Editor* window select *Data > Weight Cases*. When the window at right opens, click the *Weight Cases By* button and choose the *PERWT* variable. Click *OK*.



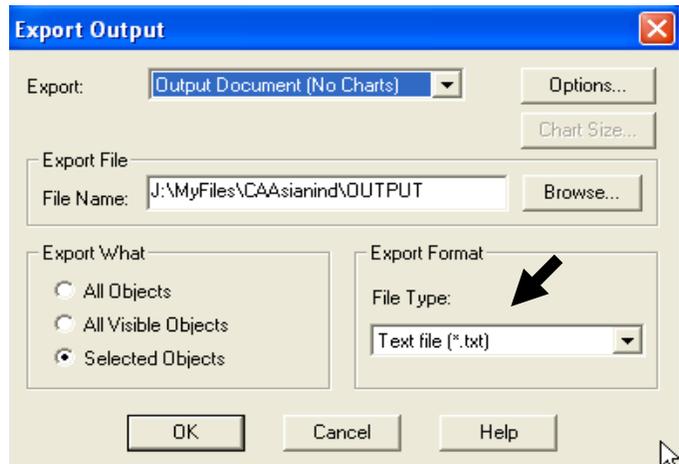
2. Select *Analyze > Descriptive Statistics > Crosstabs*.

3. When the *Crosstabs* window opens select *Occupation* from the list on the left and click on the *Rows* arrow to enter it in the top window. Do the same for *Sex* and click the *Columns* arrow in

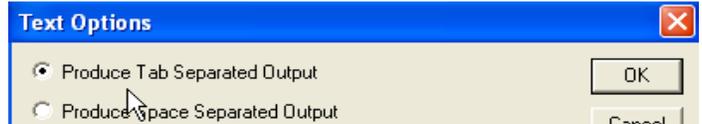


the lower window. Click *OK* and the table will be generated.

4. Look over your table in the *Output* window to make sure things are OK. Now select *File > Export*. When the *Export Output* window shown right opens, change the *Export Format* to *Text* file and click on the *Options* button.



5. Select the *Produce Tab Separated Output* and click *OK* two times.



6. You can now go to *Excel* and import your file. Compute the percentage of the employed population in each occupation and sort them in descending order. Note that you DO NOT want to include occupations coded with a 0 in your calculations since they are non-employed persons such as children.

What are the top ten categories for males and females?

**Major Asian Indian Occupations
By Number and Percent
for Males and Females in California, 1990**

Males			Females		
Code	No.	Pct.	Code	No.	Pct.
22	3619	6.5	276	1895	5.2
276	2497	4.5	95	1159	3.2
243	2290	4.1	313	1064	2.9

55	1693	3.0	447	1063	2.9
17	1571	2.8	274	933	2.6
64	1390	2.5	23	921	2.5
84	1322	2.4	379	903	2.5
53	1311	2.4	337	840	2.3
23	1257	2.3	785	754	2.1
804	1211	2.2	385	749	2.1

See the IPUMS documentation for Occupation Codes

Males

017	Managers, food serving and lodging establishments
022	Managers and administrators, n.e.c.
023	Accountants and auditors
053	Civil Engineers
055	Electrical and electronic Engineers
064	Computer systems analysts and scientists
084	Physicians
243	Supervisors and proprietors, sales occupations
276	Cashiers
804	Truck drivers

Females

023	Accountants and auditors
095	Registered nurses
274	Sales workers, other commodities
276	Cashiers
313	Secretaries
337	Bookkeepers, accounting, and auditing clerks
379	General office clerks
385	Data-entry keyers
447	Nursing aides, orderlies, and attendants
785	Assemblers

Personal Incomes for Men and Women

Another interesting issue that can be investigated in *PUMS* is the equity in income between men and women. Many believe that women are paid less than men for the same work and so you can check this by comparing the income from wages and salaries for similar groups of Asian Indian men and women. You can use the *Basic Tables* option to generally check on this issue.

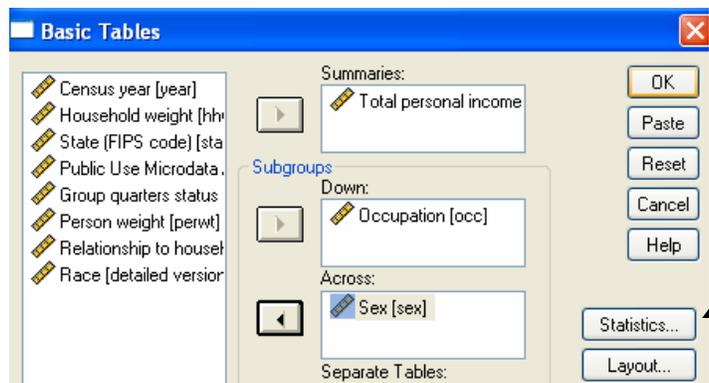
1. First go to the *SPSS Data Editor* and select the *Variable View* tab at the bottom of the window. Locate the *Missing* column and the cell for *inctot*. Set the *Missing* values to *Discrete* with values of *0* and *999999* so that those values will not be included in calculations. Also set the *Missing* value for *Occupation* to *0*.

	Name	Type	Width	Decimals	Label	Values	Missing	Co
1	year	Numeric	2	0	Census year	{0, 2000}...	None	6
2	hhwt	Numeric	4	0	Household wei	None	None	6
3	statefip	Numeric	2	0	State (FIPS co	{1, Alabama}...	None	10
4	puma	Numeric	5	0	Public Use Mi	None	None	7
5	gq	Numeric	1	0	Group quarters	{0, Vacant unit	None	4
6	perwt	Numeric	4	0	Person weight	None	None	7
7	related	Numeric	4	0	Relationship to	{101, Head/Ho	None	9
8	sex	Numeric	1	0	Sex	{1, Male}...	None	5
9	raced	Numeric	3	0	Race [detailed	{100, White}...	None	7
10	occ	Numeric	3	0	Occupation	None	0	5
11	inctot	Numeric	6	0	Total personal	None	0, 999999	8
12								
13								
14								
15								
16								
17								

Then click *OK*. Click the *Variable View* tab.

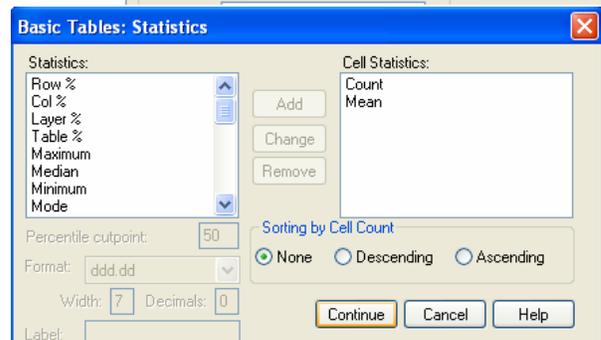
2. From the *Analyze* menu select *Tables > Basic Tables...*

3. Select the *Total income* item and click the arrow for *Summaries*. Select *Occupation* for the *Down Subgroup* and *Sex* for the *Across Subgroup*.



4. Select the *Statistics* button and the window right will open. Select the *Count* and *Mean* options.

Click the *Descending* button if you wish the *Occupations* sorted by counts. Then click *Continue*.



5. In the *Basic Tables* window click *OK* to start the processing.

In the *Output* window you will get a listing of the mean personal incomes for men and women in various occupations. Below, the *M/F Income Ratio* was created in *Excel* after exporting the table in the *Output* window.

**CA Asian Indian Wage and Salary Personal Income
by Selected Occupation, 1990**

Occup	Male		Female		M/F Ratio
	Count	Mean\$	Count	Mean\$	
0	27578	.	36149	.	
17	1571	32256	407	17655	1.8
22	3619	61210	727	25654	2.4
23	1257	30217	921	20871	1.4
53	1311	39181	79	35764	1.1
55	1693	45417	103	33861	1.3
64	1390	40599	313	31415	1.3
84	1322	100290	739	72692	1.4
95	74	12197	1159	30273	0.4
243	2290	24723	733	20166	1.2
274	854	13067	933	5093	2.6
276	2497	9151	1895	8050	1.1
313	68	10221	1064	18189	0.6
337	562	20708	840	14691	1.4
379	456	19526	903	11035	1.8
385	194	16746	749	12052	1.4
447	218	12186	1063	14766	0.8
785	556	16389	754	10116	1.6
804	1211	16381	13	12000	1.4

Using the percent employed in the top occupations compare the incomes of Asian Indian males and females. Do women and men earn similar incomes?

Exercises

- 1. Using the *Tables* analysis, compute the average Asian Indian income for males and females by PUMA. This could be mapped to see what areas in California pay higher wages to members of this group.**
- 2. Compute the male/female income ratio for the PUMAs to see where women are paid better.**
- 3. Go to the *IPUMS* web site and select the above data plus language spoken at home. See if different linguistic groups reflect different occupational niches among Asian Indians. Do some specialize in professional services, business, engineering, or health care?**
- 4. Go to the *IPUMS* web site and select the above variables for a different state or a different ethnic group. Make sure your choice has a sufficient number of the group.**

Exercise 15. Mapping Census 2000 Data

Purpose: In this exercise you will utilize *ArcMap* and existing digital data sources to produce a choropleth map of Los Angeles County. While other mapping packages are available, all CSU campuses have access to this software through a site license and so it will be used as an example.

GIS packages like *ArcGIS* are proving useful for general mapping purposes and they do offer the added advantage of being able to link an entire symbol set to a data attribute. Thus groups of symbols can be designed and modified quickly. *ArcGIS* also can convert all GIS layers into *Adobe Illustrator* layers. That graphic software is very useful for finishing the map and for creating posters from other graphic elements.

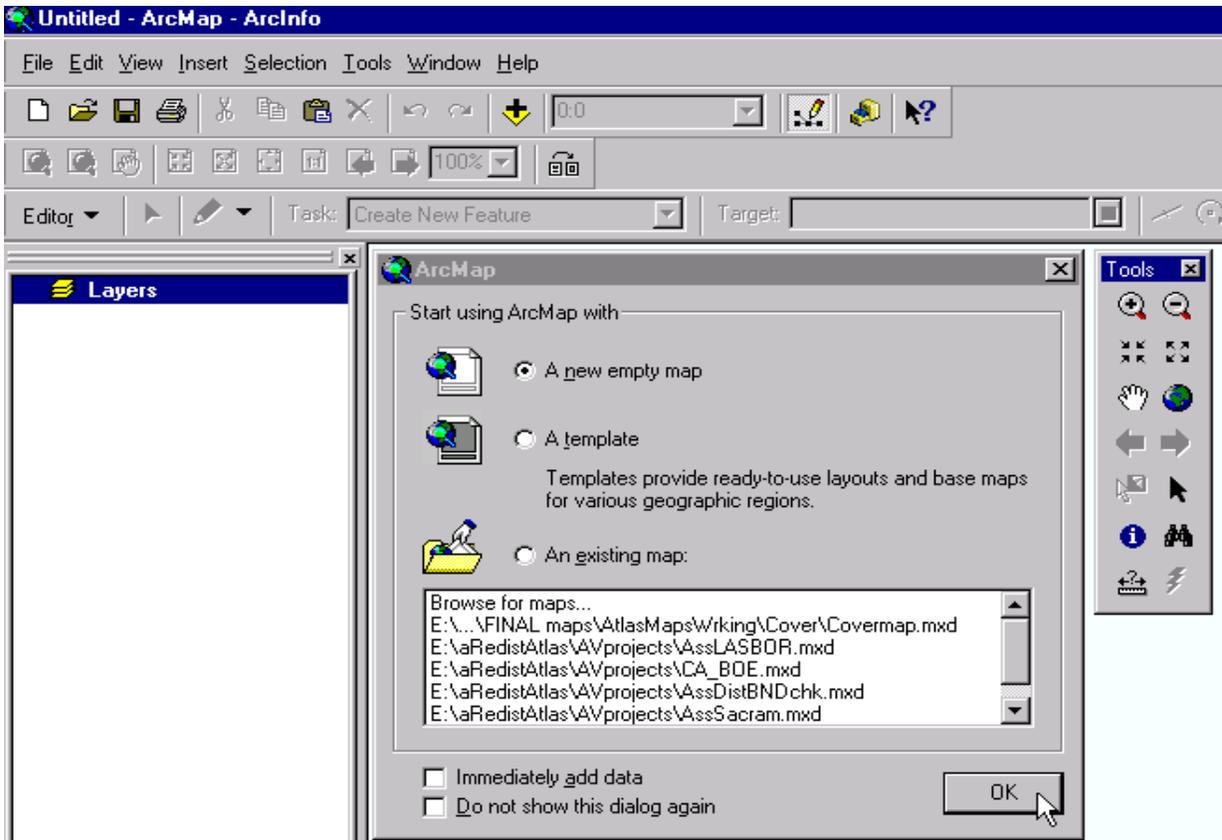
To produce the map, two files will be needed. One is the boundary file that contains the positions of all points that describe the outlines of counties or other geographic areas. The second is the data file that must be linked to the boundary file in order to map some variable of interest. Both files must contain a variable that uniquely identifies each geographic unit and these will serve to join the records of both files into one long record. Then, values from the second file can be displayed graphically within the boundaries.

Loading Files for Mapping

1. To begin, locate *ArcGIS* and *ArcMap* software on your machine. Then copy the *Mapping* folder to your machine.
2. Make sure the California county shapefile (*Caco*) is present.
3. From the *Start* button in *Windows* locate *ArcGIS*, the *ArcMap* option, and open it.

The window shown below will open.





The left part of the window in the illustration is the *Table of Contents*. A default *Layers* icon is shown below which any added data sets or layers will be listed.

The larger window in the center is the *Display* window (now mostly covered by the ArcMap window and it shows your map once the layers have been loaded.

On the far right is a *Tools* menu that contains mostly browsing features. Move the cursor over each to see what they do - which is pretty self explanatory.

In the middle of the screen is a gray window giving you the option of opening a blank template, using an existing template, or opening an existing project. Templates are basically predefined map layouts that are invaluable for doing a series of similar maps. They may be simply layout windows or may contain partially completed maps.

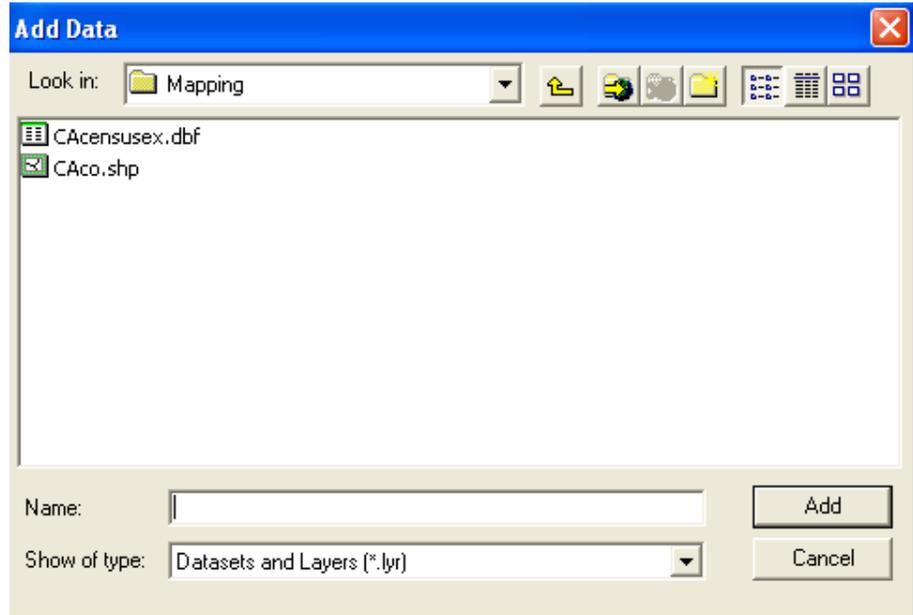
The most important icon for now is the diamond and cross at the top of the main menu. This allows you to add data to your project.



4. Click the *OK* button on the center window and then the *Add Data* icon at the top of the screen.

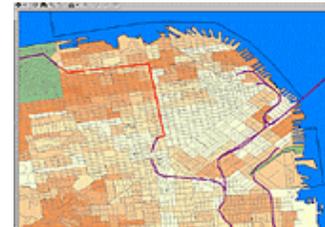
5. When the *Add Data* window opens (see below) select the California county outline boundary file named *Caco*. Then click the *Add* button.

Note that *ArcGIS* represents different data types with different icons. Point data are represented with three dots, line data with a line, and polygons with three joined polygons. *Dbf* files like *Ancestries* are shown with some column-like dashes.



Most mapping projects are composed of multiple layers though, in this exercise, we will keep things to an absolute minimum. The type of map you are going to create cartographers refer to as a *choropleth* map. In *Arcmap* it is called a *Graduated Color* map.

Basically statistical numbers are displayed within their sampled areas. Here that would be county units. Also, the data values are assigned to several classes so that the map takes on a sort of “quilt” effect from the colors assigned to the several classes. Cartographers have spent considerable effort on finding appropriate methods to determine the appropriate number of classes and where break points should occur within a distribution. The default method used in *Arcmap*, inappropriately called *natural breaks*, is an excellent starting point for setting break points in a distribution. As for the number of classes, the rule of thumb is to choose between three and eight. Fewer are needed when there are fewer areas such as with this map of California counties.



The choropleth map is a very common type of statistical or thematic map type, but it does have some important caveats.

1. Because of difference in area sizes, displaying numerical totals is usually not appropriate. The obvious result of this is that large areas will always appear in the highest classes and small areas will usually appear in the lower classes. For example, if you mapped number of children, retirees, singles, and executives the resulting patterns would all be virtually identical - reflecting the distribution of total population. Usually you want to map a percentage, percent change, density, or some other statistic adjusted for population or area.

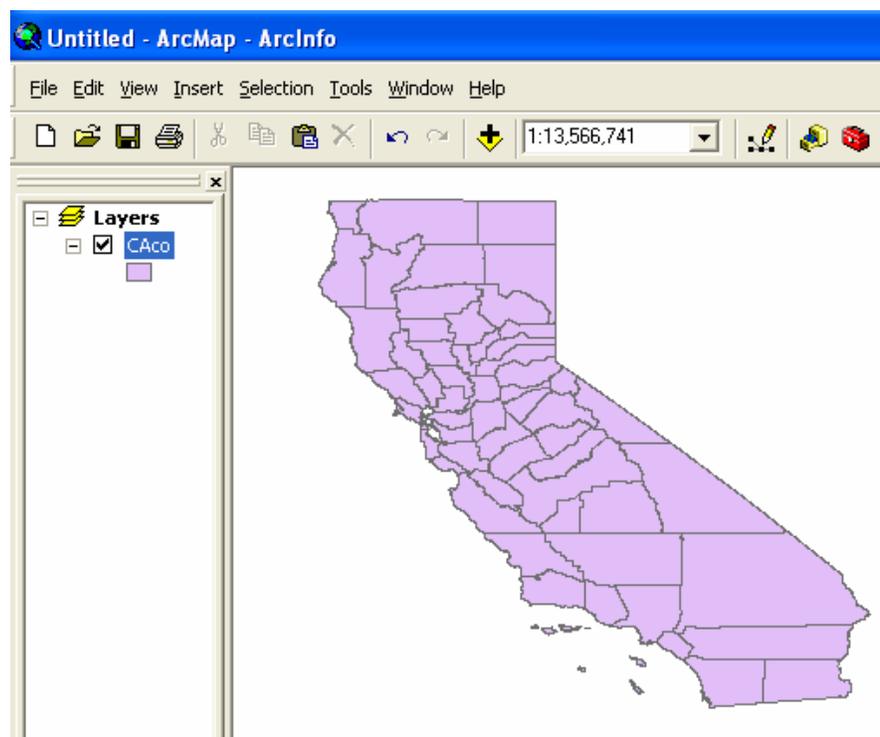
2. It is possible that important areas are missed because they are small and an inset map of an enlarged subarea might be necessary. A good example are the boroughs of New York that contain significant numbers of people but are virtually invisible on a page-size map of the entire United States. Because of the larger sizes of areas in the Western United States, one often gets a greater sense of importance for these areas than for those in the Eastern United States. In your map, San Francisco County is very small.

3. The sequence of colors selected for the map categories should increase in impact with the values in the classes. Thus the highest category should visually stand out most. The best way to accomplish this is to let the value of a color change with increasing magnitude. In ArcMap you can pick a strong, dark color for the highest category and a pale color for the lowest category and create a ramp. The software will calculate a series of transition colors between the extremes. All too often maps are created showing categories with different hues such as red, green, yellow, and blue and this makes estimating the values of intermediate categories difficult.

4. Be careful about areas with very small total populations since they can generate very high percentages. For example, in 1980 one census tract in Los Angeles County had over 12% American Indian. There were only 8 people in the tract and one was an American Indian. In some cases you might set a minimum population threshold to reduce this effect.

5. Always place a few locational references on these maps to help readers identify locations. These might include major cities, roads, rivers, or other significant regions. Other than countries or states, people are probably unfamiliar with exactly where polygons are located. In the case of our map, displaying San Francisco, Sacramento, Los Angeles, and San Diego would be helpful.

When the county layer is added, you will see a list of layers and a map as shown right. Any symbols shown are default and these can be changed. The map also has no projection. The lack of projection causes California to be 20-30% wider than it should be.



Another common problem in GIS is that data sets may have different coordinate systems or different datums. Usually if the coordinates are different, the second set will not appear on the map. If the datums are different, the layers will be offset at larger scales. ArcGIS will usually warn you when problems are encountered.

B. Setting up a Projection

Currently all the coordinates are in latitude and longitude and you can see the location values of your cursor in the bottom portion of the window.

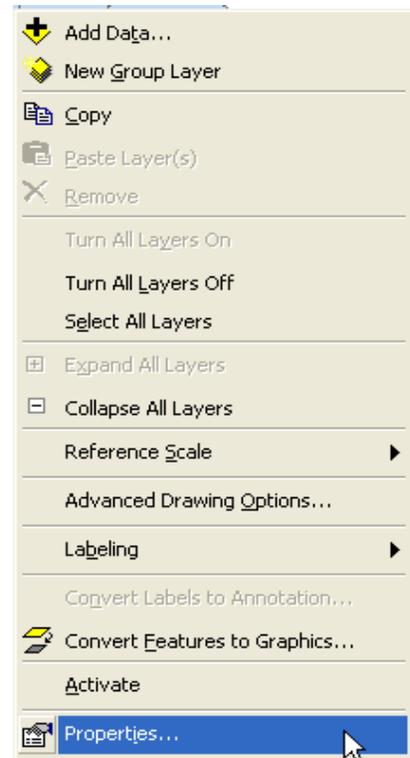
118°7'12.91"W 33°36'55.42"N

The default map is represented as if the coordinates were cartesian (x,y). This creates what is called a platte carree projection that has a great deal of distortion in higher latitudes.

Unprojected maps are increasingly appearing in the literature which indicates that many people are unaware or don't care about the distortion of area and directions being presented. The horizontal exaggeration exists even if a very small area is mapped at a very large scale. One could potentially make some serious measurement errors from such maps, and so it is worth a few moments to learn to assign a projection to a map. Fortunately we can create a projection "on-the-fly" in ArcGIS that does not change the actual coordinates. Let's pick a common projection for mid latitude areas, *Albers Equal Area Conic*.

The conic projection is typically used for mid-latitude places and especially for those with a prominent east-west dimension like the U.S., Canada, Europe, Russia, China, and Australia. For portraying geographic distributions a projection that preserves area also is desirable so that the sizes of places across the map are in proportion to their size on the earth. Your concern is to locate a central meridian in the middle of your desired area and two standard parallels that divide the vertical extent of your area of interest more or less into thirds.

For larger areas the Universal Transverse Mercator projection is often used. Each UTM zone covers six degrees of longitude and for the continental U.S. the first zone is Zone 10 on the west coast (includes northern California) followed by Zone 11 for Southern California. The zones proceed eastward in six degree steps. Furthermore, each zone is divided into a north and south half and all coordinates are laid out on a regular grid in meters. The origin for the grid is on the equator 500,000 meters to the west of the central meridian of the zone. All values in the zone are positive. One of the UTM projections is often useful for mapping one or more counties or even smaller areas.



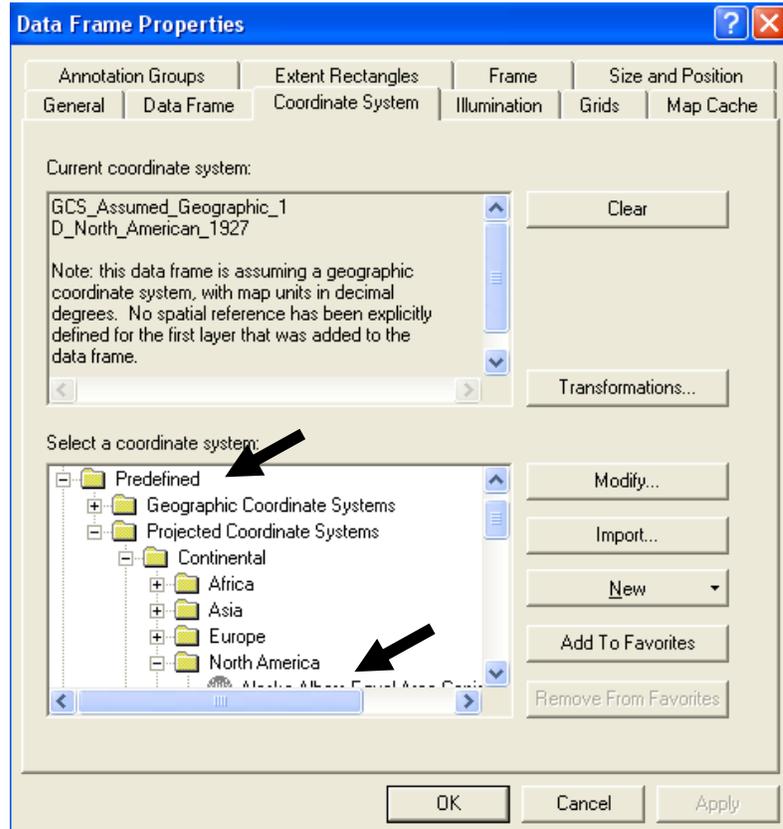
6. Right-click on the map or on the yellow *Data Frame* icon in the *Table of Contents*.

7. From the popup menu at right scroll down and select the *Properties* option.

Note well, that you will do a lot of "right-clicking" in *ArcGIS*. In particular, you will be accessing the *Properties* option of this menu.

When the *Properties* window opens make sure the *Coordinate System* tab is selected in the new window. This map has a global coordinate system (GCS) based on the *North American datum of 1927*.

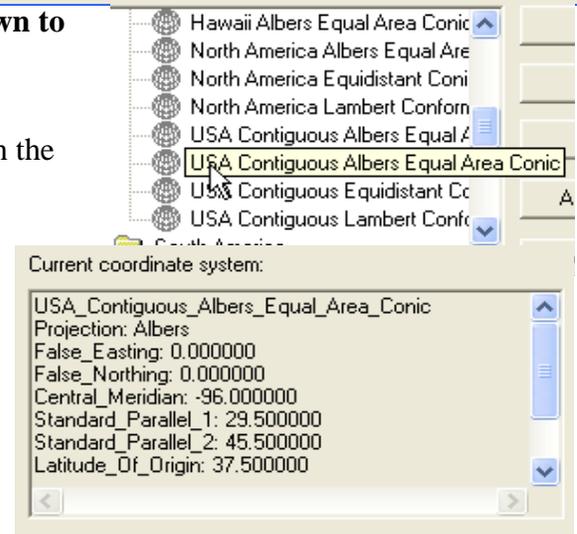
8. To select a projection from the *Select a coordinate system* window (see lower window), open the *Predefined* directory and click on the *Projected Coordinate Systems* directory, then the *Continental* directory, and finally the *North America* directory.



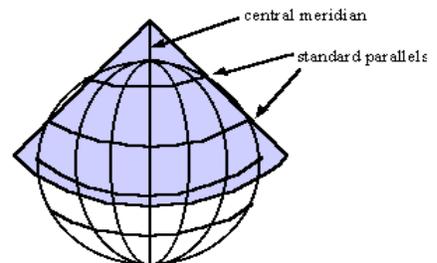
A list of projections will appear.

9. Under the *North America* directory scroll down to the *USA Contiguous Albers Equal-Area Conic*.

The projection properties will now be listed in the *Current Coordinate System* window.



The cone is usually centered over a pole and its surface contacts the earth along a parallel. The best representation of the earth occurs along that parallel (called the *Standard Parallel*) and as you can see, conic projections are ideal for mid-latitude regions. Usually the cone is made to cut into the earth at one parallel and re-emerge at another to further improve the area of coverage. Thus, two standard parallels are called for along with a central meridian to center the map on.



You will need to make a couple changes to this projection since it was set for a map centered on North America not California. You will have to move the central meridian westward and pick two lines of latitude that lie well within California. Note that longitude values in the Western Hemisphere are given negative values.

10. From the set of buttons in the right of the *Projection Coordinate System* window select the *Modify* button.

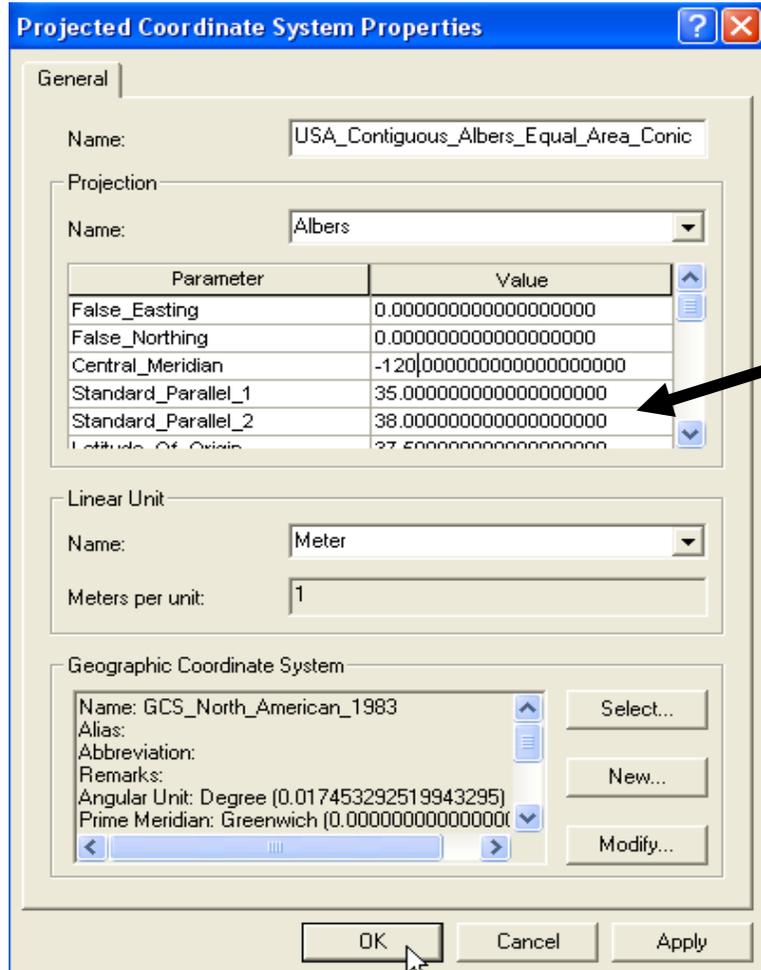


When the *Projection Properties* window opens set the *Central Meridian* to *-120*. (note the negative sign), the *First Standard Parallel* to *35*, and the *Second Standard Parallel* to *38*. Click *OK*.

Note the change in the shape of California.

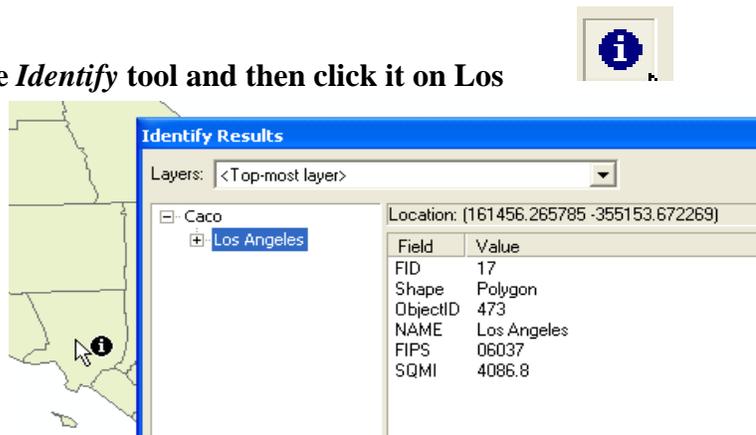
FRUSTRATION ALERT!!

This would be a really good time to save your *ArcMap* project. Select the *File* menu and save the *mxd* file in your working directory. Should *ArcMap* fail, you can open the program at this stage of completion. Note that once you have created this map file you should not move it or the map layers.



11. From the *Tools* menu select the *Identify* tool and then click it on Los Angeles County.

The *Identify Results* window right will open listing the attributes of the area.

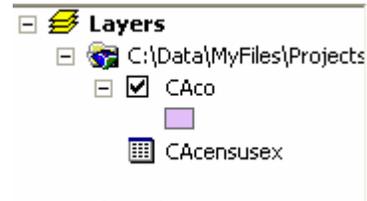


Note what variables come with the California County boundary file. Check the form of the geographic ID of the county and its name (*GEOID2*). Fortunately, the Census Bureau in its files does create unique FIPS code IDs for the counties as shown here.

C. Importing Data into the Map

1. Click on the *Add Data* icon and add the *CAcensusEx* file.

Note the icon in the *Table of Contents* for a data table.



You will need to join this data table to the attributes of the boundary file in order to map the information.

2. Right-click on the *Caco* layer in the *Table of Contents* and from the popup menu select *Joins and Relates > Join*.

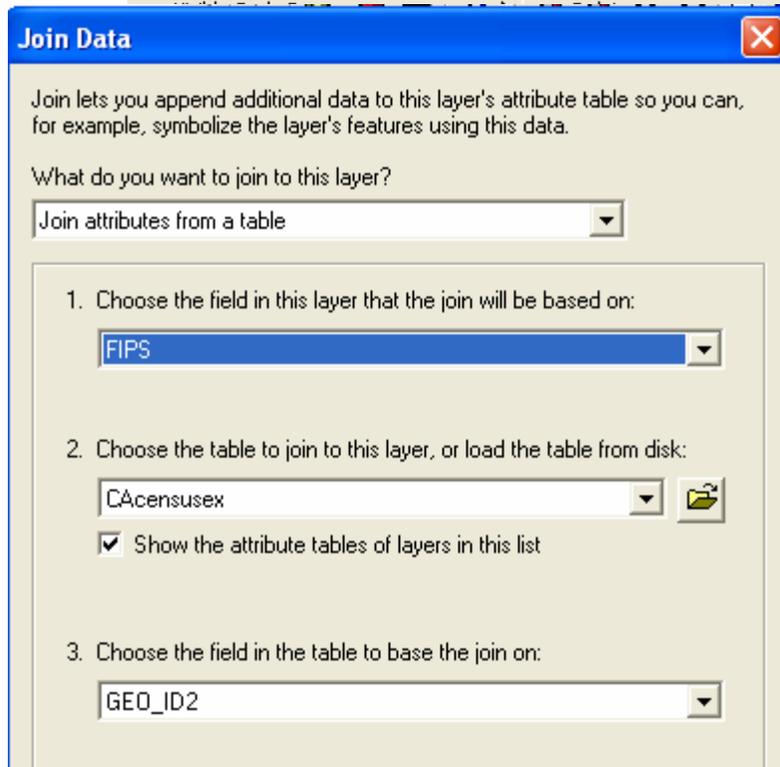


3. From the *Join Data* window make sure that *FIPS* (your ID field in the boundary file) is selected for the first join field at 1.

Next at 2, select the name of your data table, *CAcensusex*.

Then at 3 select the *GEO_ID2* variable from the data table. Then click *OK*. Say *Yes* to creating an index.

4. Again right-click on the *Caco* layer and select the *Open Attribute Table* option. Scroll to the right to see the new appended variables. Note that if you see the word *<null>* that there was a problem with the join.

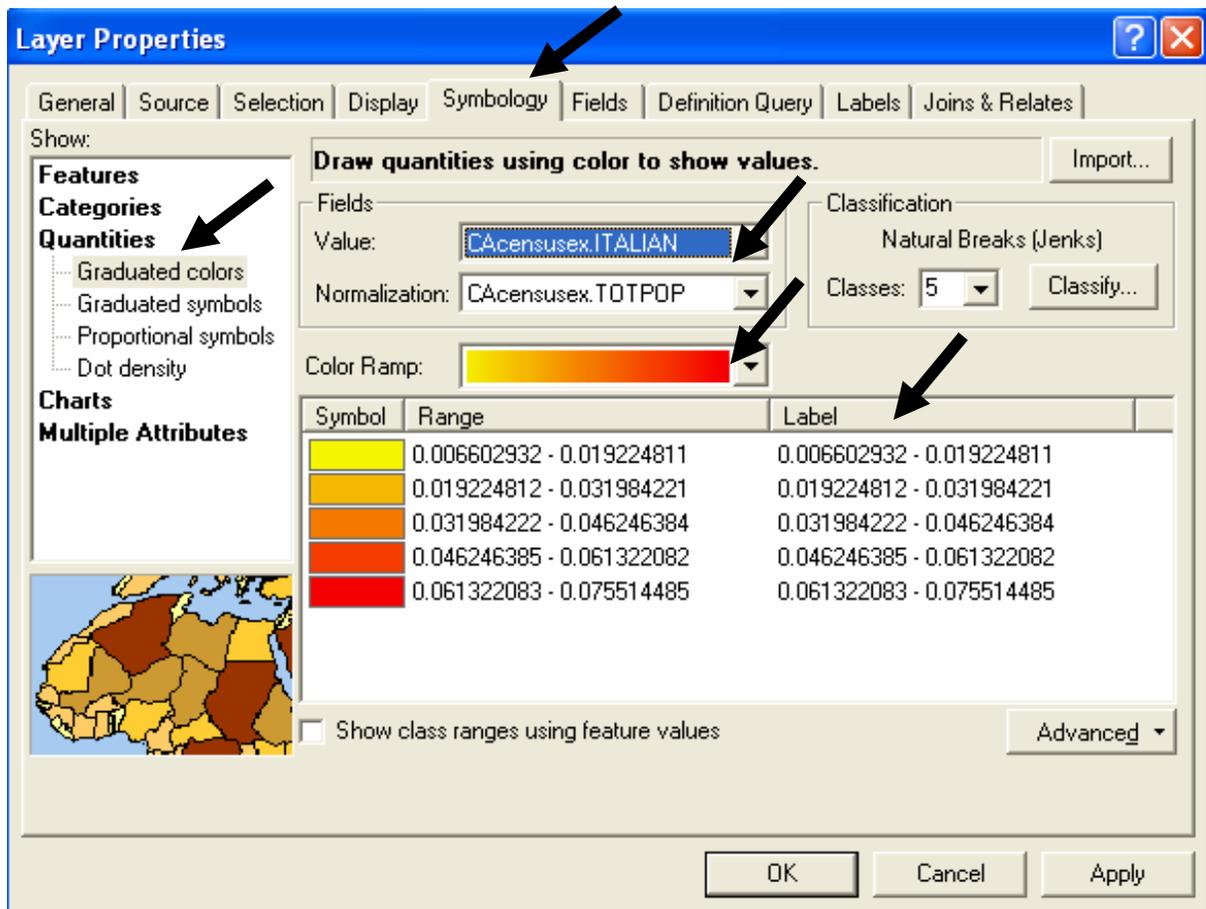


Attributes of CAco							
CAco.STATE_FIPS	CAco.CITY_FIPS	CAco.FIPS	CAcensusex.OID	CAcensusex.GEO_ID2	CAcensusex.GEO_NAME	CAcensusex.TOTPOP	
06	093	06093	46	06093	Siskiyou County, California	44301	
06	015	06015	7	06015	Del Norte County, California	27507	
06	105	06105	52	06105	Trinity County, California	13022	
06	089	06089	44	06089	Shasta County, California	163256	

Close the *Attribute Table* when satisfied.

D. Making the Map

1. Right-click on the *Caco* layer and choose the *Properties* option. When the window below opens, select the *Symbology* tab.
2. In the left-side window click on the *Quantities* label and note the four map types shown. Make sure *Graduated colors* is chosen.

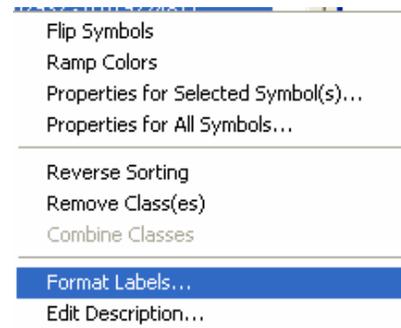


3. In the *Value:* window select the *Italian* variable and in the *Normalization* window select the *Totpop* variable. This will create a proportional value that can be converted to a percent.

A default set of 5 classes with a default color ramp will appear.

4. Click on the triangle to the right of the *Color Ramp* window and then select the yellow to red color ramp from the popup list.

5. Note that all the class labels are small decimals. To convert these to percents with fewer decimal values right-click on one of the label values. Select the *Format Labels...* option.

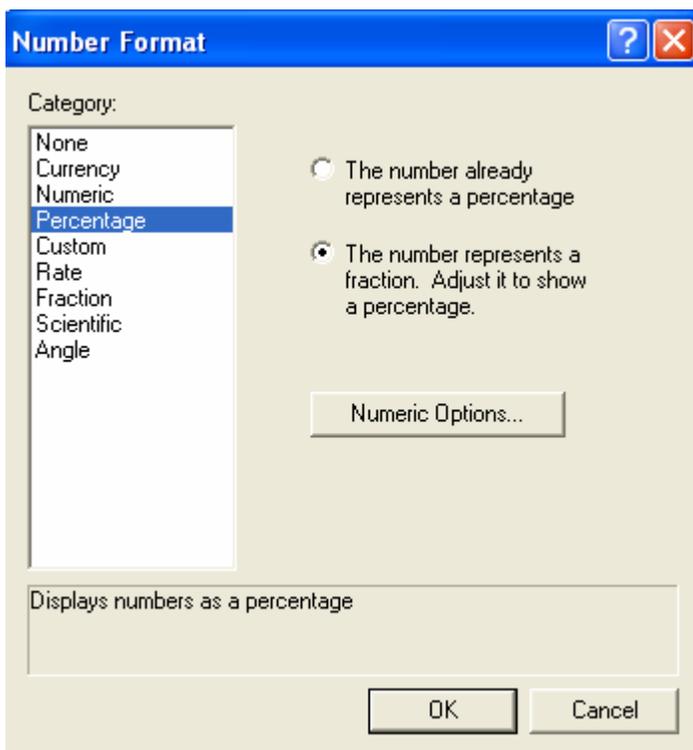


6. In the *Number Format* window select the *Percentage* category. Then click the button for *The number represents a fraction*.

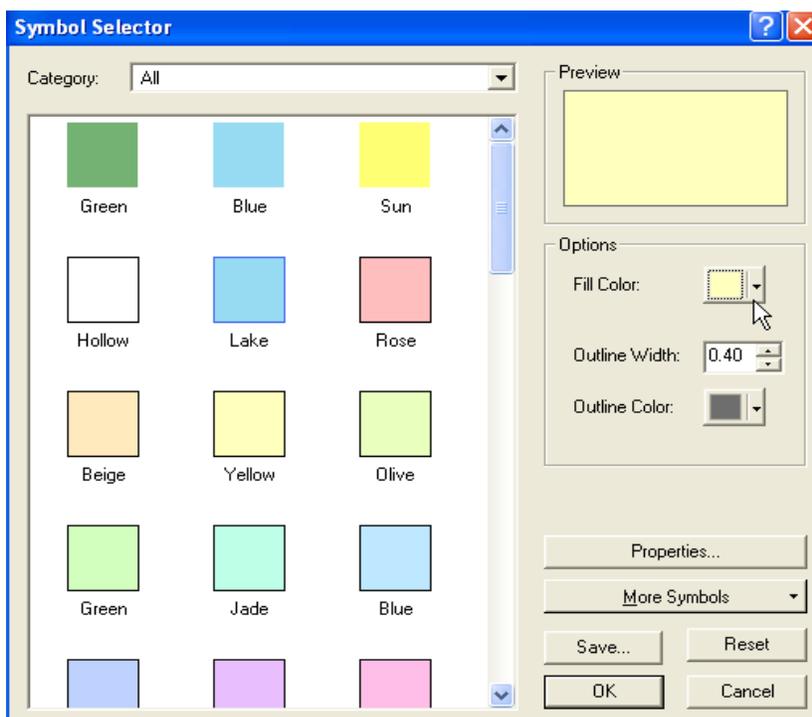
7. Click the *Numeric Options* button. Set the number of decimals to 1 and then click *OK*.

Click *OK* a second time to see the map.

Also in the *Properties* window above you can change the number of classes, the classing method, the color ramp, and you can manually re-label the classes under the *Label* column.

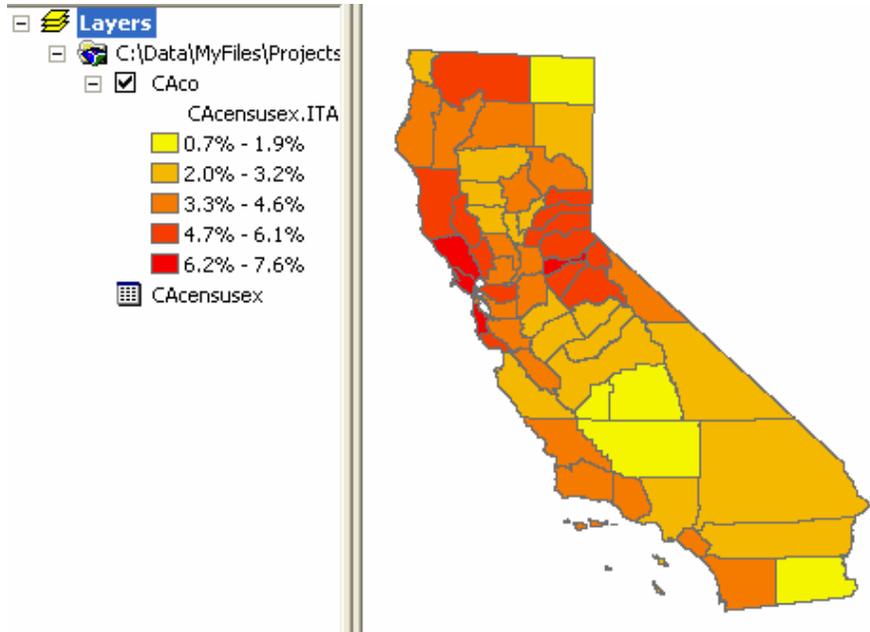


8. Double-click on the yellow color symbol in the *Table of Contents* to bring up the *Symbol Selector* window. Here you can change its fill shade to a different shade or the stroke color and width. Change the yellow to a lighter shade and then repeat the step for a few of the other lower classes. When done, click *OK* twice.



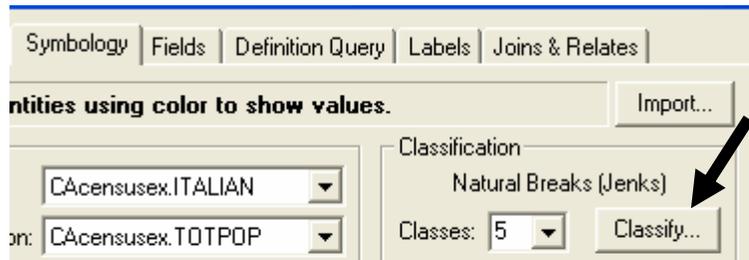
9. When the map at right of Percent Italians appears, look at it and the class values to see if it seems reasonable. In other words, check your work.

Any problems?



10. Again right-click on the *Caco* layer and select the *Properties* option. From the *Symbols* tab window select the *Classify* button.

In the *Classification* window you can modify the classes in various ways.

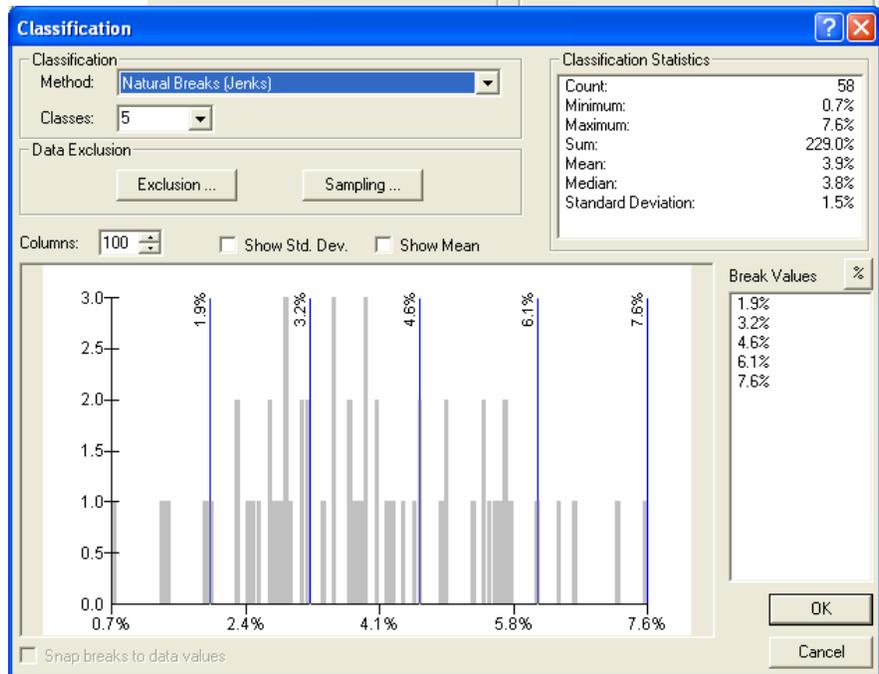


11. From the *Classification Statistics* window (right) what is the count of counties? _____

What is the minimum percentage? _____

What is the maximum percentage? _____

What are the number of classes? _____



Look at the current class break values and the shape of the distribution in the frequency diagram. You also may drag the blue lines if you want to manually shift the class breaks.

Using the *Exclusion* feature you could exclude values that fail to meet a chosen criteria. For example you could eliminate any counties with percentages less than 0.1 or with fewer than 1000 persons. On the map these could be assigned a unique color.

12. Close the *Classification* window.

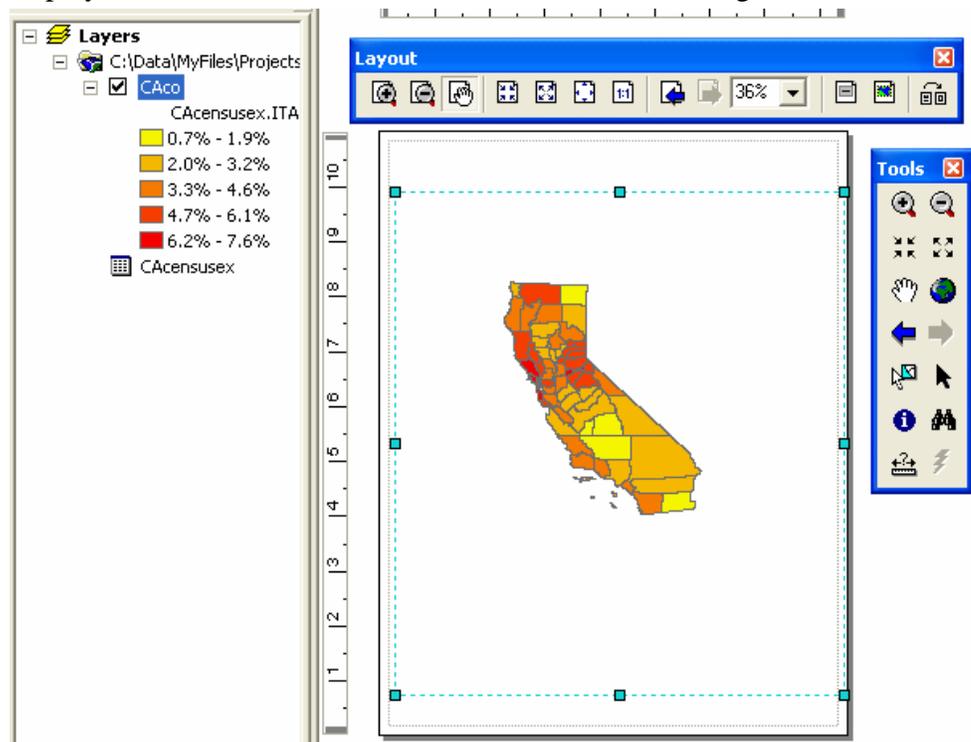
E. The Layout Window

ArcGIS provides two views of your map, a *Data View* and a *Layout View*. The former is used to compose a particular map and only one *Data View* can be displayed at a time. The latter is used to prepare a map for publication and might display several data views, plus a legend, scale, and other graphic elements.

1. From the *View* menu select *Layout View*. From here you can compose your final map for printing.

When the *Layout* window opens you will note the map in the *Data View* is displayed inside a dashed line. This sets the map display size and it can be changed by dragging its handles. Outside the dashed line are two additional lines. The first indicates the printing area of the output device and the second indicates the size of the paper the map will be printed on. On the far right of the display is the old *Tools* menu. These can be used to enlarge, reduce, or move the map within its data frame. You should NOT use these tools to zoom or move the layout page since they act on the map itself.

To move around the *Layout* window you will notice another set of tools designed for it at the top of the page above. Remember that because menus can be dragged that they may be in different locations on your machine.



2. Use the original *Tools* menu and *Zoom in* tool to size the map to fit nicely within the page.

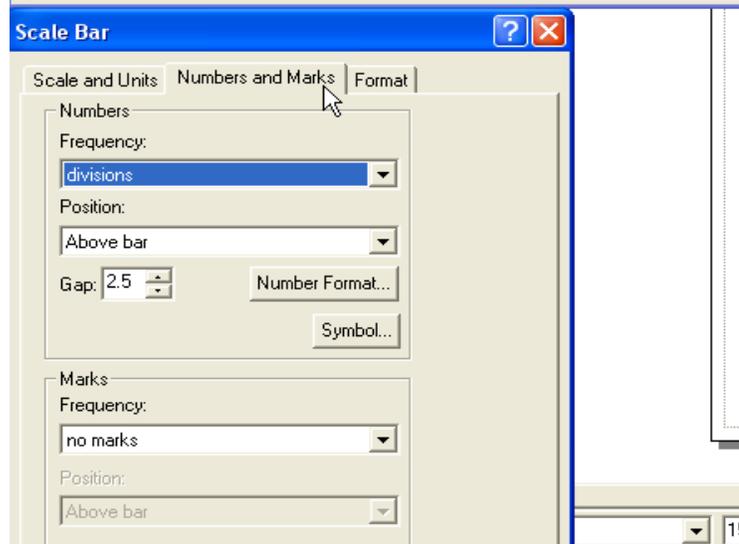
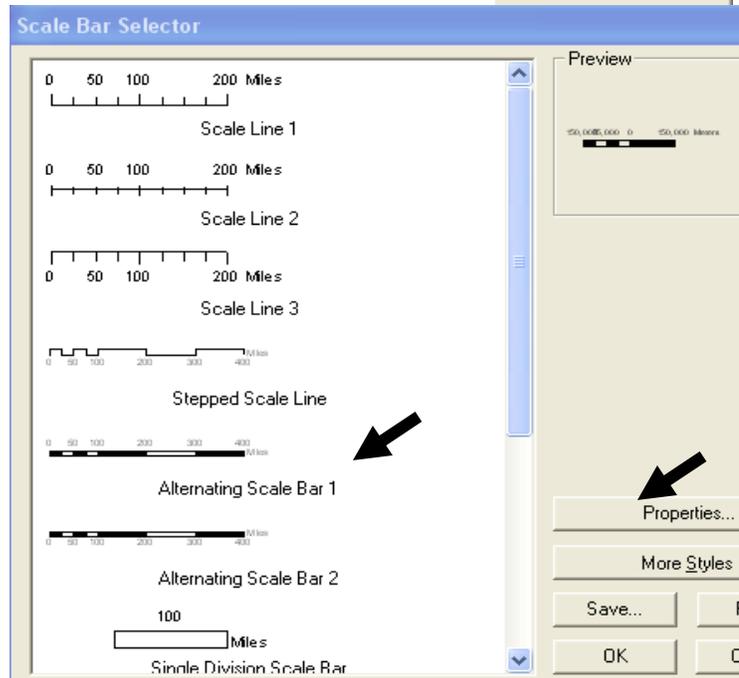
3. From the top menu select the *Insert* menu and and the *Legend* option. A *Legend Wizard* will present you with a number of design options, but for now accept all the defaults.

You can make some changes later by double-clicking on the legend itself. Basically you can add a box or background, change the nature of the symbol boxes, and determine what items are included. Not every map symbol needs to be in the legend and some would argue that the subtitle *Legend* is redundant.

4. Now drag a box over the area on the map page where you would like to place a legend.

From the *Insert* menu you may also add a north arrow and a bar scale. However, remember that on intermediate and small scale maps a north arrow is not appropriate since the direction varies over the map.

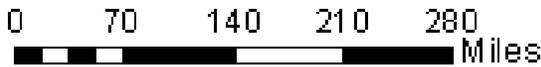
5. Select the *Insert* menu and the *Scale bar* option. Choose *Alternating Scale 1* and then select the *Properties* button. Click the *Numbers and Marks* tab and set the value to *Divisions*.



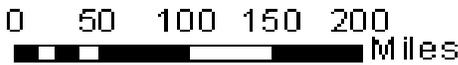
6. Now select the *Divisions and Units* tab. Set the *Division Units* to *Miles* and de-select the button to *Show one division below zero*.

Click *OK* to return to the main *Scale Bar Selector* window and then *OK* again to see the bar scale on the map.

It is quite possible your bar scale will come out similar to that shown. Unfortunately these odd break points are not as helpful as they should be. A bar scale should have evenly divisible units from which interpolation is not an effort. You will notice that the first break would be at 17.5 miles.

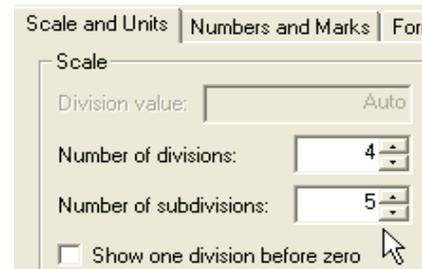
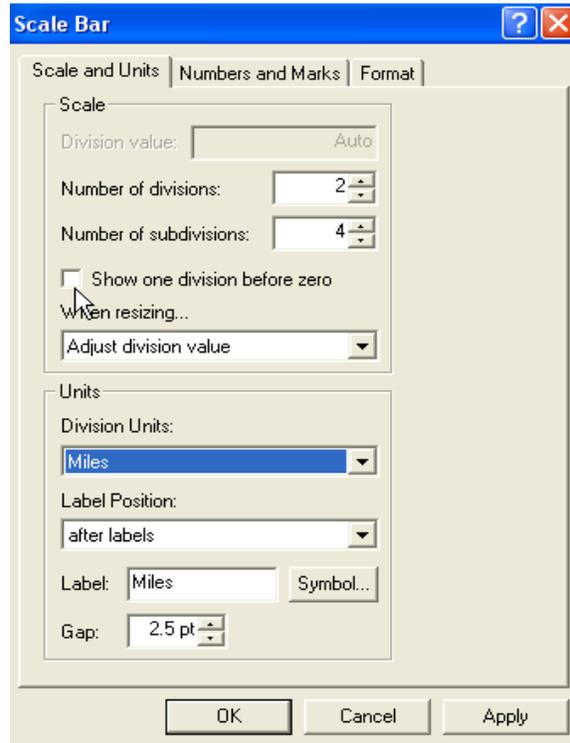


To fix this problem you can double-click on the bar to adjust it or, more simply, drag one of its handles to the left or right until the units are evenly divisible.



However, the problem with this bar scale is that there are four subdivisions at the beginning and so the breaks will occur at odd numbers.

By double-clicking on the bar scale and selecting the *Scale and Units* tab, the number of subdivisions can be set to 5 so that the subdivisions are even between 0 and 50.



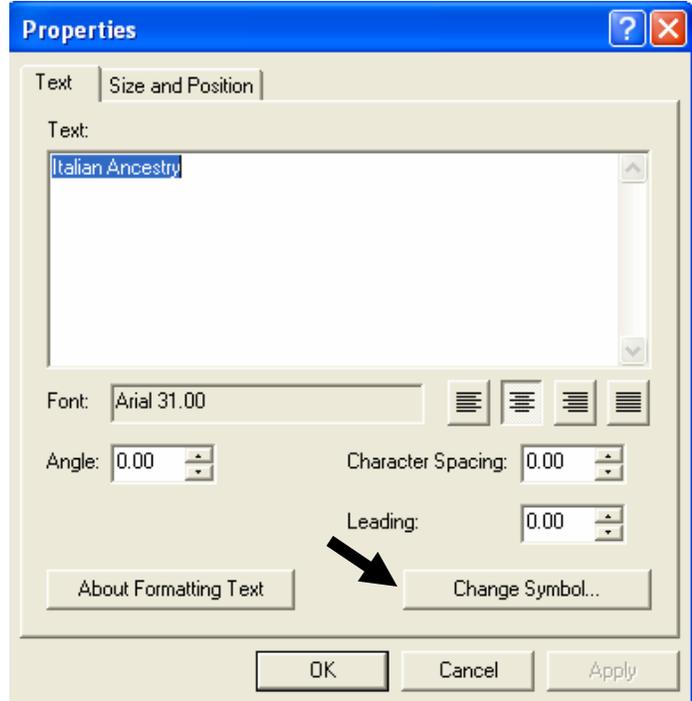
7. Select the *Insert* menu and the *Title* option. *ArcGIS* will open a box at the top of the map in which you can type your title. For this map it is *Italian Ancestry*.



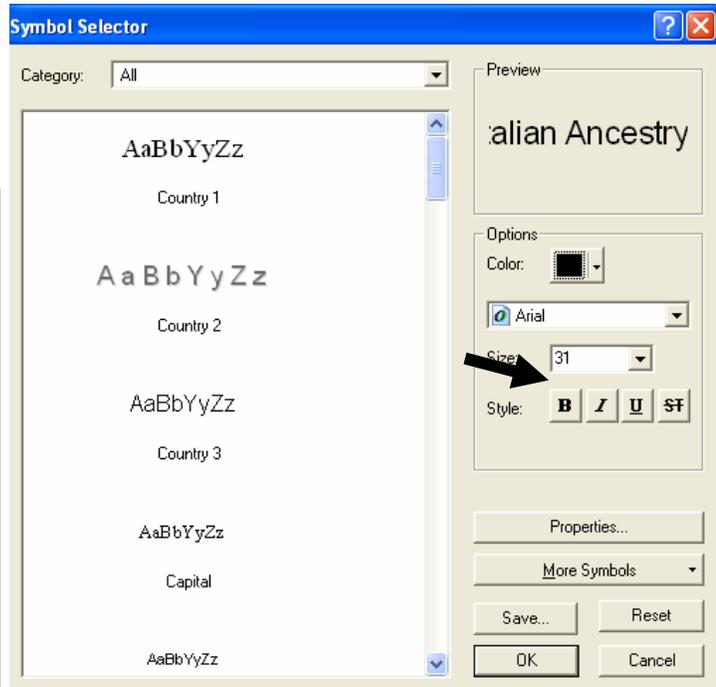
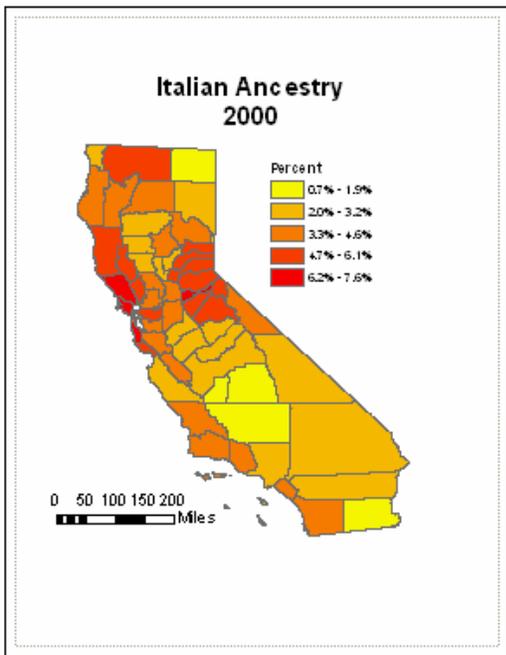
Italian Ancestry

To change the character of the title double-click on it to open the *Properties* window. Then select the *Change Symbol* button at the bottom. (See both right.)

8. Change your text (Font) to bold by clicking on the *B* button. Then click *OK* twice to see the result. Note you can make some other cosmetic changes to the legend labels by clicking once on their names in the *Table of Contents* and retyping a new value there.



9. Your map is now done. Print a copy for yourself and look at other census variables if you wish.



F. Querying the Map

One of the significant applications of software like ArcGIS is the ability to perform a variety of queries and analyses with one or more sets of data. This is generally beyond the scope of this module, but a few simple queries and tabulations of the data can be useful.

1. Select *View > Data View*. From the *Tools* menu select the *Identify* tool and click on one of the counties with a high percentage of Italians. A window of attributes will pop open. Look over the different variable values for the county.

Unfortunately, the list does not include the percent Italian which was computed in the *Properties* window. To work with this information we would need to create a new field in the *Attribute Table* and then populate it with the calculated percent Italian.

2. Close the *Identify Results* window. Then select *Selection > Select by attributes*..

From this window we can write a query to select a subset of counties from the map.

Double-click on “*CAcensusex.ITALIAN*” variable.

Click on the \geq button.

Enter *10000* at the end of the statement as shown right.

Click *Verify* to check your syntax and then click *OK*. All counties with 10,000 or more Italians will be highlighted. In this way we can see not only where the percentage of Italians is high, but where large numbers are involved. Note we could select counties based on other variables as well.

3. Close the *Select by Attributes* window. Note that in *ArcGIS* you can export selected geography into a new layer or the data into a new spreadsheet.

The image shows two overlapping windows from ArcGIS. The top window is titled "Identify Results" and shows a table of attributes for a selected feature named "Napa". The bottom window is titled "Select By Attributes" and shows a list of variables with "CAcensusex.ITALIAN" selected. The query text in the bottom window is "SELECT * FROM CAco_CAcensusex WHERE: 'CAcensusex.ITALIAN' >= 10000".

Field	Value
CAco.FID	12
CAco.Shape	Polygon
CAco.ObjectID	27
CAco.NAME	Napa
CAco.STATE_NAME	California
CAco.STATE_FIPS	06
CAco.CNTY_FIPS	055

Method: Create a new selection

Variables list: "CAcensusex.IRISH", "CAcensusex.ISRAELI", "CAcensusex.ITALIAN", "CAcensusex.LITHU", "CAcensusex.NORWEG", "CAcensusex.POLISH"

Query: SELECT * FROM CAco_CAcensusex WHERE: "CAcensusex.ITALIAN" >= 10000

4. Right-click on the *Caco* layer in the *Table of Contents*. Then select *Open Attribute Table* from the popup menu.

Note that in the *Attribute Table* that the selected counties are highlighted. At the bottom of the table you can see that 24 of the 58 counties have been selected.

5. On the bottom of the table click on the *Selected* button. Only the selected counties will be displayed. Then click the *All* button.

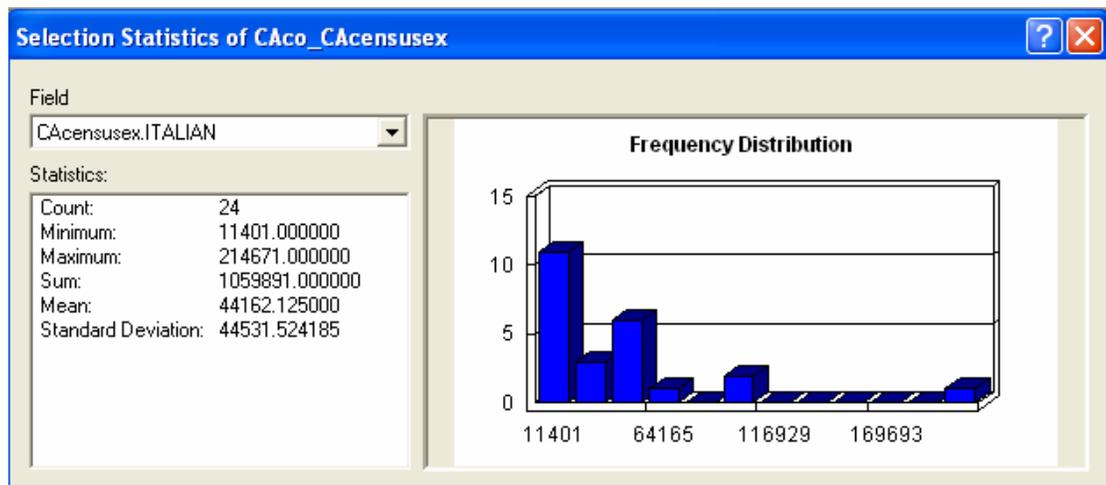
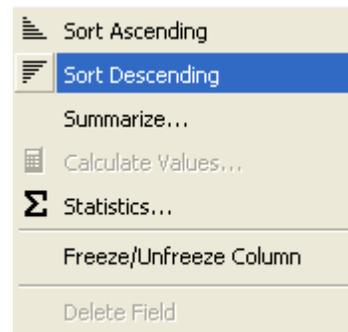
Attributes of			
	CAcensussex.ISRAELI	CAcensussex.ITALIAN	CAcensussex.LITHU
	0	2525	56
	0	619	16
	0	461	8
	0	6446	131
	24	5851	110
	0	1565	38
	4	4247	50
	0	766	5
	253	17654	584
	0	2879	74
	0	449	0
	28	5925	177
	0	6861	136
	124	34632	700
	27	14624	404
	511	53527	1107
	643	49120	1659
	553	32433	1162
	248	48846	707

Record: 1 Show: All Selected Records (24 out of 58 Selected.)

6. Right-click on the top of the *CAcensussex.ITALIAN* column. From the popup menu select the *Sort Descending* option. Look over the column. What counties have the most Italians?

7. Again right-click on the top of the *CAcensussex.ITALIAN* column. Then select the *Statistics...* option.

Note the resulting statistics are relevant to your selected subset of counties. When done, close the *Selection Statistics* window.

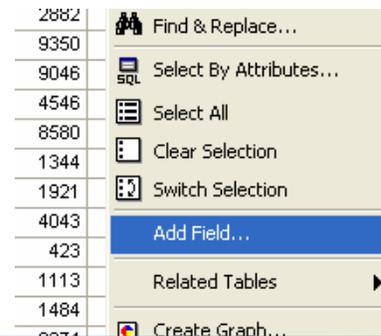


9. Another way to select features is to use the *Select Features* tool from the *Tools* menu. You may click and drag a rectangle over the map or hold the shift key and click on a series of counties to select them. Use *Selection > Clear Selected Features* to release the selected counties.

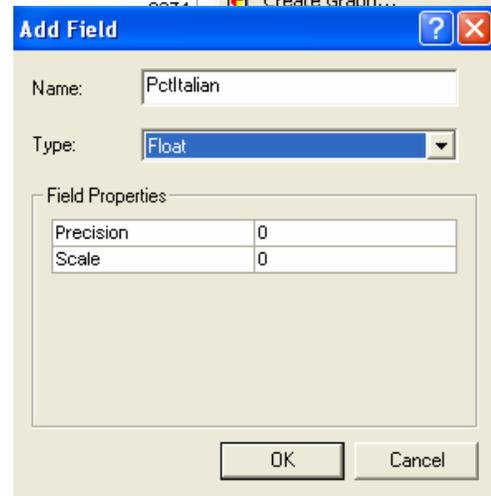


10. Right-click on the *Caco* layer and select the *Open Attribute Table* option to make the attribute table visible.

11. At the bottom of the *Attribute Table* click the *Options* button. From the popup menu select the *Add Field...* option.



12. In the *Add Field* window enter the Name of *PctItalian* and set the *Type* to *Float*. Leave the *Precision* and *Scale* set to 0. The former is the number of characters and the latter is the number of decimals. *ArcGIS* can determine these values.



Click *OK* and note the new field will be added at the end of the list of boundary file attributes. (i.e. ahead of the joined variables)

13. Look for the *Editor Toolbar* symbol at the top of the screen. If it is not visible click *Tools > Editor Toolbar*.



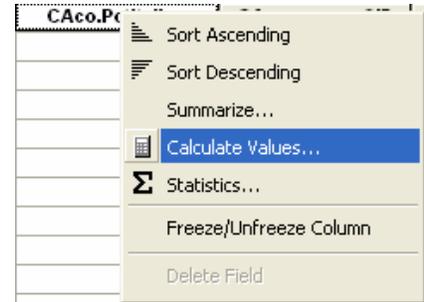
14. Locate the *Editor* menu and select the *Start Editing* option.



In the *Attribute Table* the fill behind the column labels will turn white.

Attributes of				
	CAco.FIPS	CAco.PctItalian	CAcensussex.OID	CAcensussex.GEO_ID2
	06047	0	23	06047
	06069	0	34	06069
	06051	0	25	06051
	06109	0	54	06109
	06007	0	3	06007
	06091	0	45	06091

15. Right-click on the top of the new column (which is currently filled with 0s) and select *Calculate Values* from the popup menu.



16. In the *Field Calculator* window double-click on the item *CAcensussex:ITALIAN*. Note spaces are required between items in the equation.

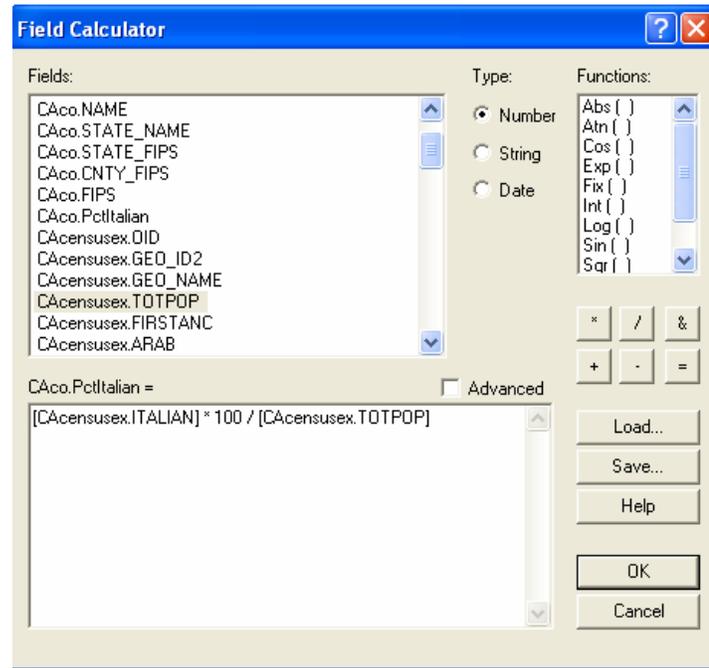
Select the * symbol.

Enter a space and type 100.

Enter a space and select the / symbol.

Enter a space and double-click on *CAcensussex:TOTPOP*

Click *OK* to run the calculation. The new percent Italian field will be populated.



17. Again locate the *Editor* button and select the *Stop Editing* option. Save your edits.

You now have a permanent value for percent Italian that can be sorted and processed just as the count of Italians was.

G. Exercises

1. Select one of the other ancestries in the *Attribute Table* and map it by percent of total population. Create a new field with the actual percentages. In what counties is the group especially concentrated? In what counties are there large numbers? Can you offer any explanation about why the group is located where it is? Consider time of initial settlement, and particular occupations.
2. Make choropleth maps for an ethnic group of both the number and the percent of total population. Compare the patterns on the two maps. Are they similar? Why do you think they might be different?
3. Compute the correlation of two census variables in a statistics program. Export both the correlations and the residuals making sure that you have useable geographic identifiers as well. Map the two variables, the correlation values, and the residuals. Describe the patterns.
4. Calculate the percent ethnic and make choropleth maps for counties and census county divisions. How do the two patterns differ? Is either one more helpful or descriptive?

Exercise 16. Examining the Demographics of a Service Area

You might want to review the first *ArcMap* exercise if you are not yet familiar with the product.

Purpose: This exercise will give you an idea of how spatial queries may be made with census data. In this example you will perform a common type of marketing query to determine what the demographic characteristics are within a service area. Normally a distance of a one to three mile radius from a specified location is used. Then, socio-economic data is aggregated from those census units that fall within the specified distance.

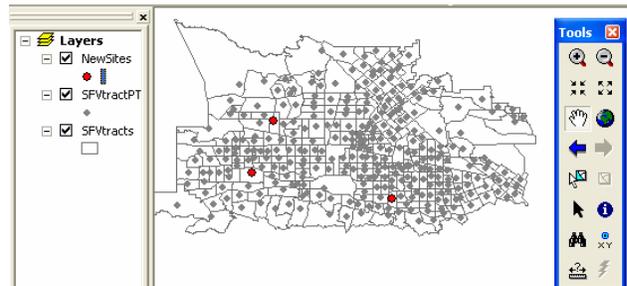
For this exercise there are three possible locations in the San Fernando Valley, California. One is in the northwest part of the valley in an area called Chatsworth. To the south is the second location in Woodland Hills. Farther to the east is a third site in an area called Encino. The sites are visible on the *NewSites* layer.

The second map layer, *SFVtractPT*, contains the centroids of census tracts along with a series of census variables deemed useful to help select a site. These include the populations of Non-Hispanic Whites, Blacks, Asians, and Hispanics; the number of males and females, the number of persons less than age 18, from 18 to 64, and over age 64; the number of households for Non-Hispanic Whites, Blacks, Asians, and Hispanics; and the aggregate household income in each tract for each of the same four groups. Depending on what group we want to target, we can selectively tabulate demographic statistics for an area (a buffer) that surrounds a site.

The third map layer of tract outlines is to provide a locational background for the points.

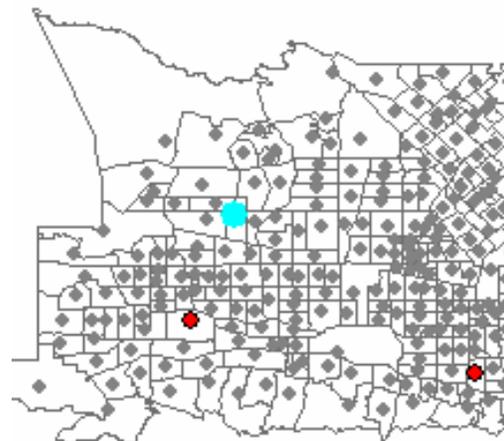
1. Start *ArcMap*, select the *Add Data* button, and load the following three files from the *Mapping* directory: *NewSites*, *SFVtractPT*, and *SFVtracts*.

A *dataframe* called *Layers* will appear that contains the three map layers. Right-click on the *Layers* dataframe and set a projection if you can. *UTM Zone 11* works.

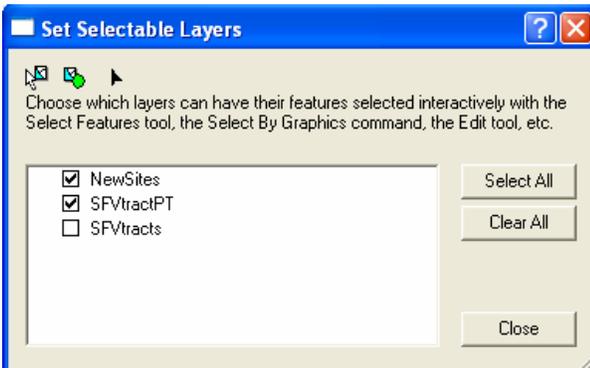


2. Right-click on the *NewSites* layer and select the *Open Attribute Table* option. Click to the left of the *Chatsworth* site to select its row. Then close the *Attribute Table*. The site should be highlighted in blue.

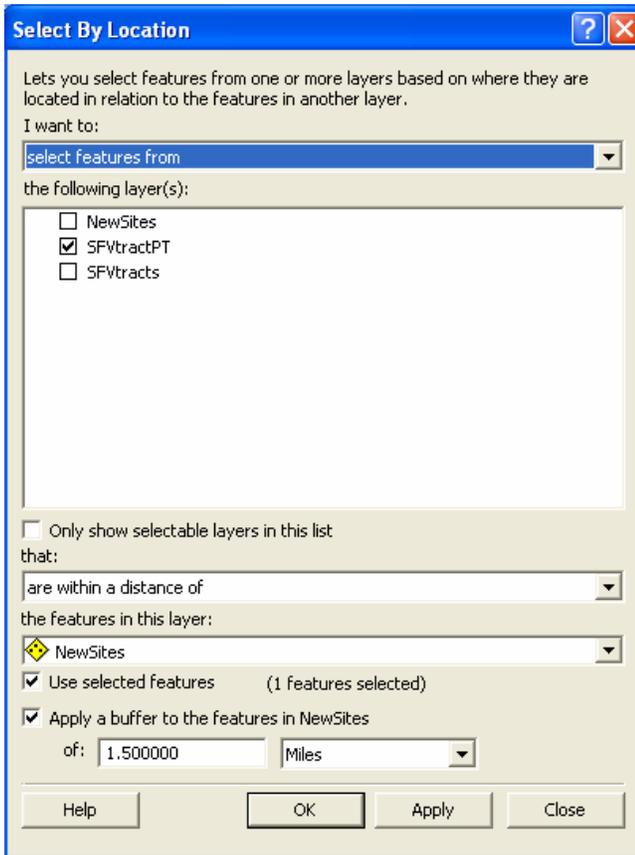
FID	Shape *	TRACT2KID	SQMILES	LAT	LONG	OID_	GEOID	
0	Point	06037113301	0.8757	34.247453	-118.559528	83	06037113301	Chatsworth
1	Point	06037124600	0.5954	34.160156	-118.426516	169	06037124600	Woodland Hills
2	Point	06037134902	1.5937	34.188771	-118.582780	242	06037134902	Encino



3. Select *Selection > Set Selectable Layers* and make sure the *SFVtracts* layer is not checked. Then click *Close*.



4. Click on the *NewSites* layer in the *Table of Contents*. Then select *Selection > Select by Location*.



5. In the *Select by Location* window...

Choose the *select features from* option at the top.

Under the *the following layer(s)* window check the *SFVtractPT* layer.

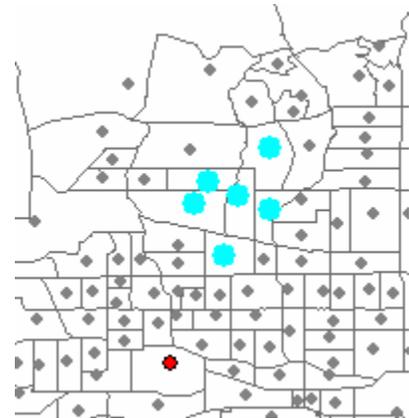
Under *that choose are within a distance of*.

Under *the features in this layer* window choose *NewSites*.

Make sure *Use selected features* is checked.

Make sure *Apply a buffer to the features in NewSites of 1.5 miles* is chosen.

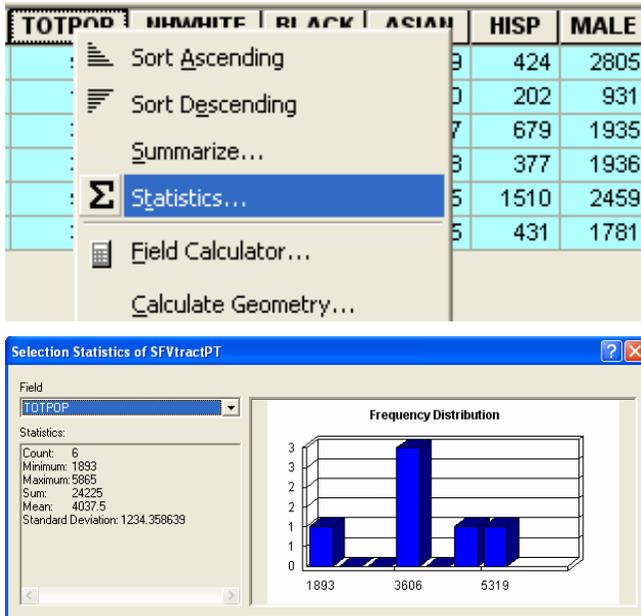
Then click *OK*. Six tracts within a radius of the Chatsworth site will be selected.



6. Right-click on the *SFVtractPT* layer in the *Table of Contents* and select the *Open Attribute Table* option. At the bottom of the spreadsheet click on the *Selected* button to see only the six selected tracts.

Selected Attributes of SFVtractPT							
FID	Shape *	TRACT2KID	SQMILES	LAT	LONG	OID	GEOID
87	Point	06037111204	1.3073	34.264402	-118.548373	70	06037111204
80	Point	06037113301	0.8757	34.247453	-118.559528	83	06037113301
81	Point	06037113303	1.5793	34.244903	-118.575044	84	06037113303
83	Point	06037113322	0.7058	34.252960	-118.569841	86	06037113322
84	Point	06037113401	0.9494	34.226439	-118.564130	87	06037113401
91	Point	06037115301	0.6366	34.242954	-118.547868	94	06037115301

7. Right-click on the top of the *Totpop* attribute in the table and from the popup menu select the *Statistics* option.



8. A table that summarizes various qualities of the *Totpop* variable will appear. Note the *Count* of tracts is 6 and the *Sum* of the *Totpop* is 24,225.

9. Add this number to the table above and then select the *Field* window and select the next variable in the *Attribute Table*, and the *Sum* of *NHWhite* which is 15,195. Add this to the table above and proceed through the other variables.

10. When the *Sum* for all attributes have been recorded for Chatsworth, close the *Attribute Table*. Choose *Selection > Clear Selected Features*.

11. Return to Step 2 above and select the second site, Woodland Hills. Then follow the steps to buffer it at 1.5 miles, select the tract centers within the buffer, and then sum the attributes of the selected tracts. Write the result for each variable in the table above.

	Chatsworth	Woodland Hills	Encino
Total Pop			
NH-White			
Black			
Asian			
Hispanic			
Males			
Females			
<18 yrs			
18-64 yrs			
>64 yrs			
Black HH			
Asian HH			
Hisp. HH			
NHWh HH			
Agg Inc Blk			
Agg Inc Asian			
Agg Inc Hisp			
Agg Inc NHW			
Blk Inc p HH			
Asian Inc p HH			
Hisp Inc p HH			
NHW Inc p HH			

12. Again repeat the selection process for Encino.

13. Calculate the *Aggregate Income per Household* for each ethnic group by dividing the *Aggregate Income* by the count of households.

14. Note the ethnic make up of each site's service area. Compute the percent if you wish.

15. Which site has the highest percent of females?

16. Which site has the greatest number of persons under 18 and which site has the greatest number over 64?

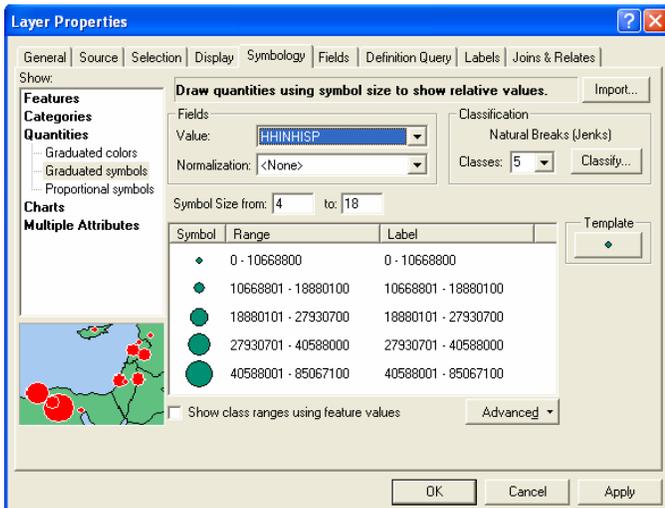
17. Which site seems to have the most people with the highest per household income?

Mapping Census Centroids

It is usually helpful to look at the patterns of data values over a larger area to get some sense of where higher values of interest occur. This might be done initially to focus on potential sites for more detailed analysis.

Since the data are associated with points we will use graduated symbols to examine the quantities in each tract. Graduated symbols are preferred for displaying raw counts.

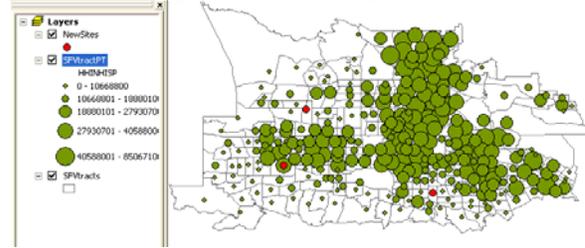
1. Right-click on the *SFVtractPT* layer and select the *Properties* option and *Symbology* tab.



2. From the *Show* window on the left select the *Quantities* option and the *Graduated symbols* option. A set of five circles will appear representing ranges of dollar categories of the sum of Hispanic income at each point.

3. In the *Value* window select the *HHINHISP* variable (Hispanic Aggregate Income). Click *OK*.

A map showing the amount of Hispanic income will appear. Note the pattern of tracts.

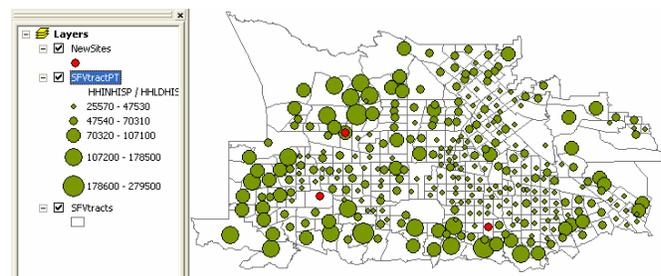


If you were interested in just dollars you might look at the areas of large numbers of large circles. However, many of these areas also have a greater number of Hispanic households.

For marketing purposes we are really interested in locations where the per household income is high since these people are more likely to purchase a more expensive product. In this case you can compute this “on the fly” by dividing the aggregate income by the number of Hispanic households. This is called *Normalization* in ArcGIS.

4. Again right-click on the *SFVtractPT* layer and from the *Properties* option select the *Symbology* tab. Click on the *Normalization* window and select the *HHldHisp* variable (Hispanic households). Then click *OK*.

On the new map note the locations of well-to-do Hispanic households. Compare this pattern to that of the previous map.



Map some of the other variables and consider where you might want to locate a marketing site based on greater potential for customers.