Exercise 16. Downloading Raw Census Data

Purpose: The goal of this exercise is to extract some 2000 census data using the Access data base program.

Accessing Raw Data on the Census Web Site

When using the Census web site one does not have to worry about the coded variables used within the various tables. However, when seeking raw data one will have to pay careful attention to the table and variable identifiers. Both are numbered and the summary file documentation will be required.

- 1. Start *Microsoft Explorer* (NOT *Netscape*) and go to the census web site: www.census.gov
- 2. Select the Summary File 3 link.



3. Under FTP Download select the All Files link.

Data: Access to all tables and maps in American FactFinder

Purchase Products: Summary Files for sale through Customer Services Center

Product Support: Summary File 3 Disc Product Support

FTP Download: All Files FTP Read me (MSWord | WordPerfect | Text)

Documentation: Technical Documentation [PDF] (6M)

- Comparing SF 3 Estimates with Corresponding Values in SF 1 and SF 2

A directory containing subdirectories of all states will appear.

At the top of the list are several useful files that describe how to read *SF3* files, in what files tables are located, and where to find the *Access* templates.

Further down the list are subdirectories containing the 77 files for each state. Later we will work with Nevada because of its small size.

	<u>Name</u>	Last modifie	<u>ed</u>	<u>Size</u>	Ī
	Parent Directory			-	
2	OREADME_SF3.doc	31-May-2002	12:26	142K	
	OREADME_SF3.txt	31-May-2002	12:26	12K	
?	OREADME_SF3.wpd	31-May-2002	12:26	43 K	
?	OSF3_File_Structure.doc	21-Jun-2002	13:05	23K	
	OSF3_File_Structure.txt	21-Jun-2002	13:05	83	
?	OSF3 File Structure.wpd	21-Jun-2002	13:05	7.3K	
?	OSF3 geo header.doc	27-Jun-2002	16:54	155K	
?	OSF3 table matrix.doc	12-Jul-2002	16:53	2.7M	

11-Jun-2003 12:03

19-Aug-2002 07:58

25-Sep-2002 00:01

n<u>v00076_uf3.zip</u>

U.S. Census Bureau

4. On the following page of file segmentation, study the distribution of the various SF3 tables among the 76 files.

0 National/

This is where you must use the *SF3 Documentation* to examine the tables and variables of interest. From *SF3* we will select the *geo* file, and data files *1,2,19*, and *21* because we are interested in the tables contained in those files. You must always get the *geo* file since all other files must use its codes to subset records.

Note that each file has a *uf3* suffix. This will have to be changed to *txt* so that *Access* will recognize it. Otherwise the file will not appear in the file input list of the program. Only the *geo* file has fixed field sizes, the other files are text files with comma delimited values.

- **5.** Return to your operating system. Create a directory where you can store information for this exercise.
- 6. Now return to *Explorer* and the *SF3* directory for *Nevada*. Scroll to the bottom of the list and double-click on *nvgeo uf3.zip*
- 7. When the archive opens in *WinZip* select the file and then click the *Extract* icon. Place the file in your newly created working directory.
- 8. Return to the list of files for Nevada and download file *nv00001.uf3*. Repeat the steps for *nv0002.uf3*, *nv00019.uf3*, and *nv00021.uf3* We will use these for this exercise.



SF1 File/Table Segmentation

Name	Number	Starting	Ending			
	Of Data	Matrix	Matrix			
	Items	Number	Number			
Geograph						
01	222	P1	P5			
02	238	P6	P18			
03	236	P19	P33			
04	149	P34	P45			
05	245	P12A	P12E			
06	241	P12F	P16I			
07	234	P17A	P27C			
08	247	P27D	P28E			
09	244	P28F	P30H			
10	229	P30I	P34I			
11	180	P35A	P35I			
12	235	PCT1	PCT9			
13	45	PCT10	PCT11			
14	209	PCT12	PCT12			
15	196	PCT13	PCT17			
16	209	PCT12A	PCT12A			
17	209	PCT12B	PCT12B			
18	209	PCT12C	PCT12C			
19	209	PCT12D	PCT12D			
20	209	PCT12E	PCT12E			
21	209	PCT12F	PCT12F			
22	209	PCT12G	PCT12G			
23	209	PCT12H	PCT12H			
24	209	PCT12I	PCT12I			
25	209	PCT12J	PCT12J			
26	209	PCT12K	PCT12K			
27	209	PCT12L	PCT12L			
28	209	PCT12M	PCT12M			
29	209	PCT12N	PCT12			
30	209	PCT12O	PCT12O			
31	245	PCT13A	PCT13E			
32	235	PCT13F	PCT15C			
33	225	PCT15D	PCT17B			
34	225	PCT17C	PCT17E			
35	225	PCT17F	PCT17H			
36	75 217	PCT17I	PCT17I			
37	217	H1	H20			
38	207 171	H11A	H15I			
39	1 / 1	H16A	H16I			

SF3 File/Table Segmentation

File	Number	Starting	Ending	
(Cifsn)	Of Data	Matrix	Matrix	
	Items	Number	Number	
stgeo.uf3	31			
st00001.u	f3	248	P1	P14
st00002.u	f3	218	P15	P24
st00003.u	f3	241	P25	P37
st00004.u	f3	227	P38	P46
st00005.u	f3	220	P47	P50
st00006.u	f3	250	P51	P67
st00007.u	f3	213	P68	P91
st00008.u	f3	245	P92	P138
st00009.u	f3	203	P139	P145C
st00010.u	f3	245	P145D	P145H
st00011.u	f3	235	P145I	P146F
st00012.u	f3	246	P146G	P147I
st00013.u	f3	241	P148A	P149D
st00014.u	f3	245	P149E	P150I

st00015.uf3	239	P151A	P154D
st00016.uf3	240	P154E	P159G
st00017.uf3	239	P159H	P160E
st00017.uf3	164	P160F	P160I
st00019.uf3	247	PCT1	PCT8
	204	PCT9	PCT15
st00020.uf3			
st00021.uf3	222	PCT16	PCT17
st00022.uf3	235	PCT18	PCT19
st00023.uf3	233	PCT20	PCT24
st00024.uf3	233	PCT25	PCT27
st00025.uf3	221	PCT28	PCT32
st00026.uf3	106	PCT33	PCT34
st00027.uf3	221	PCT35	PCT37
st00028.uf3	162	PCT38	PCT43
st00029.uf3	205	PCT44	PCT48
st00030.uf3	224	PCT49	PCT51
st00031.uf3	205	PCT52	PCT56
st00032.uf3	243	PCT57	PCT61
st00032.uf3	243	PCT62A	PCT63C
st00033.uf3	234	PCT63D	PCT64H
st00034.ul3	234	PCT64I	PCT66C
st00036.uf3	233	PCT66D	PCT67E
st00037.uf3	223	PCT67F	PCT68C
st00038.uf3	245	PCT68D	PCT68H
st00039.uf3	247	PCT68I	PCT69I
st00040.uf3	243	PCT70A	PCT70I
st00041.uf3	245	PCT71A	PCT71E
st00042.uf3	196	PCT71F	PCT71I
st00043.uf3	240	PCT72A	PCT72B
st00044.uf3	240	PCT72C	PCT72D
st00045.uf3	240	PCT72E	PCT72F
st00046.uf3	240	PCT72G	PCT72H
st00047.uf3	215	PCT72I	PCT73A
st00048.uf3	190	PCT73B	PCT73C
st00049.uf3	190	PCT73D	PCT73E
st00050.uf3	190	PCT73F	PCT73G
st00050.uf3	190	PCT73H	PCT73I
st00051.ul3	231	PCT74A	PCT75C
st00052.uf3	236	PCT75D	PCT75G
st00053.ui3	234	PCT75H	PCT76D
	145	PCT76E	PCT76I
st00055.uf3			
st00056.uf3	127	H1	H18
st00057.uf3	249	H19	H26
st00058.uf3	216	H27	H44
st00059.uf3	250	H45	H68
st00060.uf3	248	H69	H86
st00061.uf3	250	H87	H104
st00062.uf3	59	H105	H121
st00063.uf3	171	HCT1	HCT3
st00064.uf3	115	HCT4	HCT4
st00065.uf3	143	HCT5	HCT5
st00066.uf3	248	НСТ6	HCT7
st00067.uf3	219	HCT8	HCT14
st00068.uf3	214	HCT15	HCT17
st00069.uf3	220	HCT18	HCT23
		HCT24	HCT31C
st00070.uf3	248		
st00071.uf3	246	HCT31D	HCT36E
st00072.uf3	246	HCT36E	HCT40I
st00073.uf3	243	HCT41A	HCT43I
st00074.uf3	224	HCT44A	HCT44C
st00075.uf3	247	НСТ44Н	HCT47F
st00076.uf3			
3100070.013	96	HCT47G	HCT48I

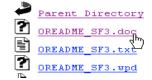
9. Now go to your working directory and change the suffix of the extracted data files from *uf3* to *txt*Access does not recognize the uf3 suffix.

Name 🔺	Size	Туре	Date Modified
™nv00001.uf3	5,984 KB	UF3 File	7/29/2002 8:31 PM
🔟 nv00002.uf3	5,076 KB	UF3 File	7/29/2002 8:32 PM
🔟 nv00019.uf3	3,271 KB	UF3 File	7/29/2002 8:32 PM
🔟 nv00021.uf3	2,581 KB	UF3 File	7/29/2002 8:32 PM
🔟 nvgeo.uf3	3,192 KB	UF3 File	7/29/2002 8:35 PM

Name A	Size	Туре	Date Modified
🗐 nv00001.txt	5,984 KB	Text D	7/29/2002 8:31 PM
🗐 nv00002.txt	5,076 KB	Text D	7/29/2002 8:32 PM
🗐 nv00019.txt	3,271 KB	Text D	7/29/2002 8:32 PM
🗐 nv00021.txt	2,581 KB	Text D	7/29/2002 8:32 PM
nvgeo.txt	3,192 KB	Text D	7/29/2002 8:35 PM

Now that you have the geography file and four of the 76 data files, you are ready to extract some tables for analysis.

10. Go to the top of the list of Nevada files and select the *0README* document.



11. When the documentation opens, click on the second link: Structure files in Access 97 and other formats.

Contents

About the FTP Application
Other Sources of the Data
File Naming Convention
Technical Documentation
Data Records and Segmentation
File Record Layout
FTP File Transfer
UnZipping the Files
Spreadsheet and Data Base Aids
Maximum File Sizes

- For step-by-step instructions for moving the data and the structure into a data base format (including screen shots), please see www.census.gov/support/SF3ASCII.html.
- Structure files in Access97 and other formats are available at http://www.census.gov/support/2000/SF3/.
- We are unable to provide one-on-one support for applications of the data to specific spreadsheets or data base software.

12. Click on the *Acc2000.zip* file. (See right.)

Note there are instructions for four different data base programs, but you will be working with *Access*.



13. Extract the Access database file (.mdb) in WinZIP and then save it to your working directory.

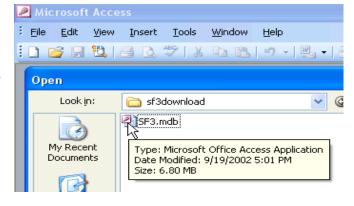


The *Access 2000* file consists of templates for the geography file and all 76 *SF3* files. Right now it is around 7 Mb, but it will grow very quickly as data files are appended to the file templates.

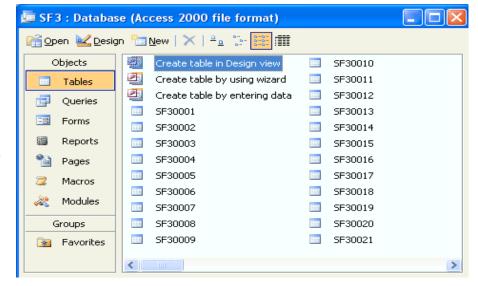
Loading Census Data into Access – Fixed Record Lengths

1. Start the Microsoft Access program and select File > Open. (See right.) Look for the SF3.mdb Access database under your working directory and open it.

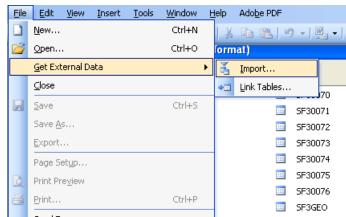
When the database opens as shown below, you will see three routines for creating tables and a list of all empty *SF3* templates to which data may be added.



At the end of the list is the SF3GEO file that contains the necessary geographic codes to extract data. This file has fixed record lengths whereas all the other data files use a commadelimited record format.

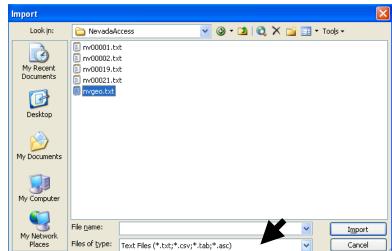


2. Scroll to the end of the list of tables and select *SF3GEO*. Then select *File > Get External Data > Import*.



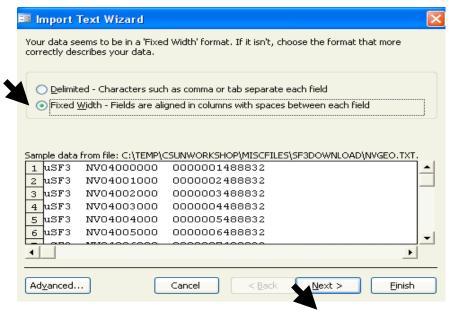
3. When the *Import* window opens set the *Files of type* option at the bottom to *Text*. (If you did not change the *uf3* suffixes to *txt* earlier you will not see the files.)

Then select the *NVGEO* file and click the *Import* button.



The *Import Wizard* will begin to lead you through a series of options.

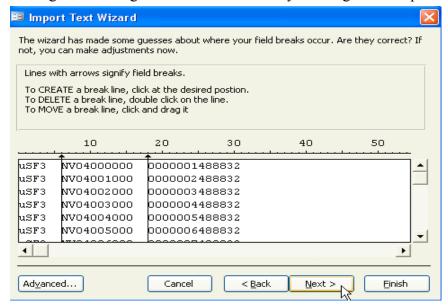
- 3. The first few records of the *NVGEO* file will be displayed. In this file the records are fixed in width and each variable is assigned a specific column range. The *Fixed Width* button should be checked.
- 4. Click the *Next* button.



The second window shows a default method for partitioning each record into fields of specific widths. You can drag a delimiting line or add new lines by clicking on the top

lines. It is a very helpful way of visually selecting fields and their sizes. For now, forget this option since we will use a template that already has defined this information

5. Click Next.



6. In the third window select the *Advanced* button at the lower left corner of the window.

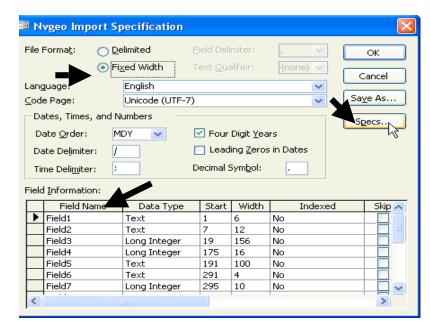
The SF3GEO file is the only fixed-field file in the set and so you must select it using the Advanced tools.



7. When the *Advanced* options window opens, note the *Fixed Width* button is selected.

Note also that the field names are generic (ie *Field1*, *Field2*, etc.).

Now select the *Specs*.. button.



8. From the *Import/Export*Specifications window select the
SF3GEO Import Specification file
from the list of files.

Then click Open.

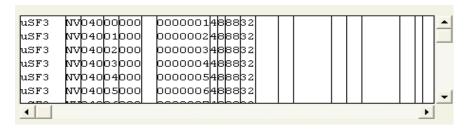


Back at the *Field Information* window you will now see the names of the variables and each has a *Start* column and *Width*.



9. Click *OK* and you will return to the *Import Text* Wizard window.

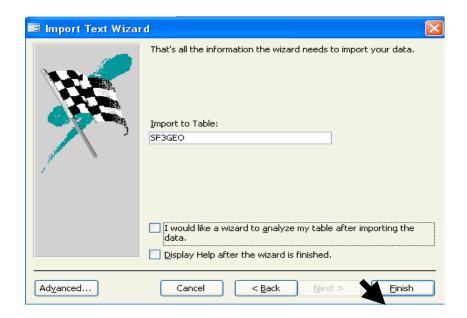
The records have now been partitioned.



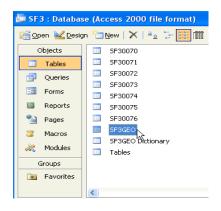
- 10. From the *Import Text* Wizard select the *In an Existing Table* button and then locate the *SF3GEO* file. We want to put the data in this file.
- 11. Click Next.



12. Click *Finish* to begin loading the data into the *SF3GEO* file.



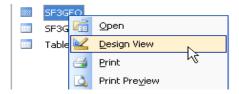
13. When the program finishes, double-click on the *SF3GEO* file to open it.



14. Look at the various columns and values. You now have one table completed, but it is the most important one since any desired data tables must be linked via the variable *LOGRECNO* to it in order to extract data.

	■ SF3GEO: Table								
	FILEID	STUSAB	SUMLEV	GEOCOMP	CHARITER	CIFSN	LOGRECNO	REGION	DIVISION
▶	uSF3	N∀	040	00	000		0000001	4	8
	uSF3	N∨	040	01	000		0000002	4	8
	uSF3	N∨	040	02	000		0000003	4	8
	uSF3	N∨	040	03	000		0000004	4	8
	uSF3	N∨	040	04	000		0000005	4	8
	uSF3	N∨	040	05	000		0000006	4	8
	uSF3	N∨	040	06	000		0000007	4	8
	uSF3	N∨	040	07	000		0000008	4	8
	uSF3	N∨	040	08	000		0000009	4	8
	uSF3	N∨	040	09	000		0000010	4	8
	uSF3	N∨	040	10	000		0000011	4	8
	uSF3	N∨	040	11	000		0000012	4	8
	uSF3	N∨	040	12	000		0000013	4	8
	uSF3	N∨	040	13	000		0000014	4	8
	u≤E3	MV	OAO.	1/	nnn		0000015	4	8

15. Close the *SF3GEO* table and then either rightclick over the *SF3GEO* file name or select it and then click on the *Design View* icon from the main data base menu.



16. When the *Design View* of the *GEO* table opens (right), locate the *LOGRECNO* variable and right-click on it.

When the new menu pops up, click on *Primary Key*.

The *Primary Key* will be used to link tables which all contain this particular variable. This declaration will only be done with the *GEO* file since it was a fixed-record type of file.



17. At right the *LOGRECNO* variable now has a small key next to its name. **Close the table and save your changes.**

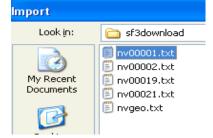
■ SF3GEO : Table					
	Field Name	Data Type			
	FILEID	Text			
	STUSAB	Text			
	SUMLEV	Text			
	GEOCOMP	Text			
	CHARITER	Text			
	CIFSN	Text			
৪▶	LOGRECNO	Text			
	REGION	Text			
	DIVISION	Text			
	STATECE	Text			
	STATE	Text			
	COUNTY	Text			

Loading Census Data into Access – Comma-Delimited Records

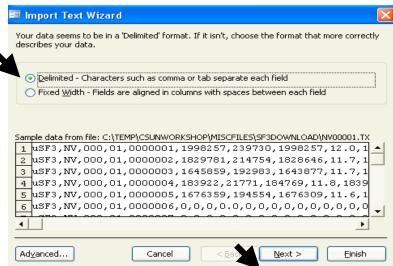
1. Select the *SF3001* file in the *Database* window.



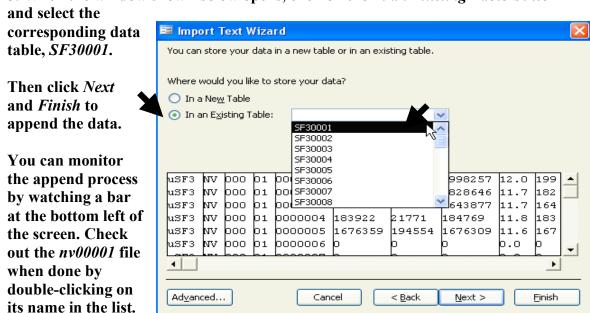
2. Select *File > Get External Data > Import*. When the list of files appears, select the *nv00001.txt* file. (See right.)



- 3. When the *Import* wizard begins, the file will be recognized as being *Delimited* and the first few records will have commas between them. Click *Next*.
- 4. Click *Next* on the following window.



5. When the window shown below opens, click on the In an Existing Table button



6. Now repeat steps 1 through 4 to append data into *nv0002*, *nv0019*, and *nv0021*. You should have five of the 77 total files with data in them when done with this part of the exercise.

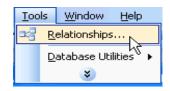
Linking Tables in Access

Now that the geography table and some data tables have been filled out, they will have to be related (linked) through a common variable. This is what the logical record number

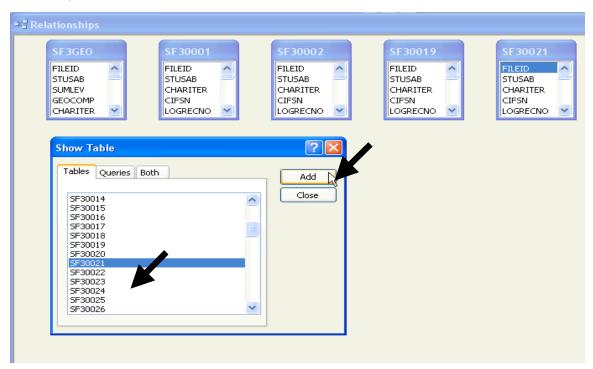
(LOGRECNO) is for. Each geographic record has a unique record number that can be used to link tables. At right is a part of the sf30001 table.

LOGRECNO	P001001	P002001	P003001	PI
0000001	1998257	239730	1998257	
0000002	1829781	214754	1828646	
0000003	1645859	192983	1643877	
0000004	183922	21771	184769	
0000005	1676359	194554	1676309	
0000006	0	0	0	
0000007	0	0	0	
0000008	1314296	151210	1314357	
0000009	0	0	0	
0000010	304013	36444	303689	
0000011	0	0	0	

1. From the *Tools* menu select the *Relationships* option.

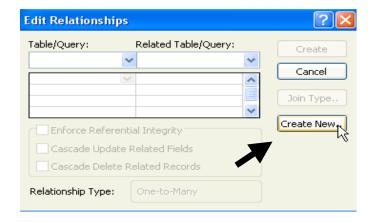


2. When the *Relationships* window opens, click on *SF3GEO* and select the *Add* button. A small template will appear in the window. Repeat this for *SF30001*, *SF30002*, *SF30019* and *SF30021* and then close the *Show Table* window.



3. Double-click on *LOGRECNO* in the *SF3GEO* table and the *Edit Relationships* window will open. (Right.)

Click the *Create New* button. You will use this to create the table linkage.



4. Under *Left Table Name:* select *SF3GEO* and below that in *Left Column Name:* select *LOGRECNO*.

Under *Right Table Name*: select *SF300001* and below that enter: *LOGRECNO*.

Then click OK.

5. When the *Edit Relationships* window opens, click *Create*.





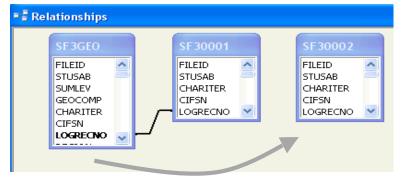
A dark angular line connecting *LOGRECNO* will join the two tables in the

Relationships window as shown below.

6. To more quickly establish links do the following: on the SF3GEO window at right click on the LOGRECNO variable and then drag over to the LOGRECNO variable on table

SF30002. (See gray arrow right.)

The *Edit Relationships* window will open. Then click on the *Create* button to join this table to *SF3GEO*. Repeat this quick method for the remaining two files so that all are linked to *SF3GEO*.





You can join multiple tables this way, but I have had *Access* crash when the total size of the database approached 2 gigabytes. Thus you can not process all census tables in one database.

Now that the needed tables are linked, you can begin to extract information by location or type of geography.

Querying Access to Extract Data

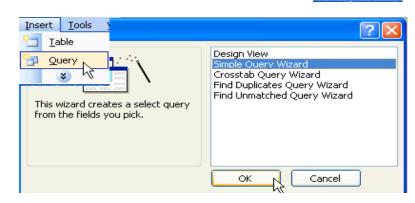
This is where knowledge of FIPS codes and Summary Level Codes becomes important. To begin, you will make a simple query to list the FIPS codes for the cities in Nevada.

1. Either select the *Query* option under the *Database* window or select *Insert* > *Query* as shown right.



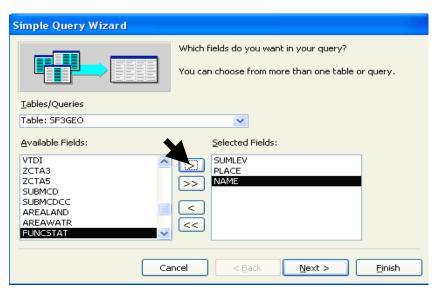
When the *New Query* window opens, select the *Simple Query Wizard* (far right) and then click *OK*.

In the first window you will select the tables and the variables from which you will extract data.



2. In the window shown below make sure *SF3GEO* is selected under the *Tables/Queries* window. Then click on the *SUMLEV* variable and the move right

button (>). Repeat this for the *PLACE*, and *NAME* variables.



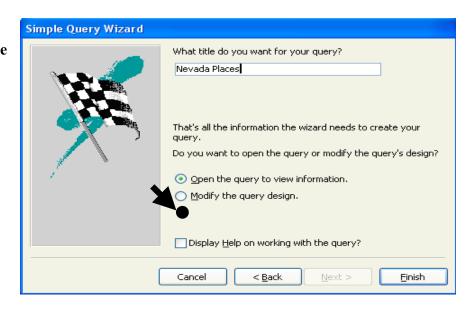
3. Return to the *Tables/Queries* window and select *SF30021*. From this table select the variable *PCT016001* and add it to the list. In this way any variable from any populated table can be added to your extract table.

Then click Next.

On the next window click Next.

3. On the final window change the name of the table from SF3GEO Query to Nevada Places and select the Modify the query design button.

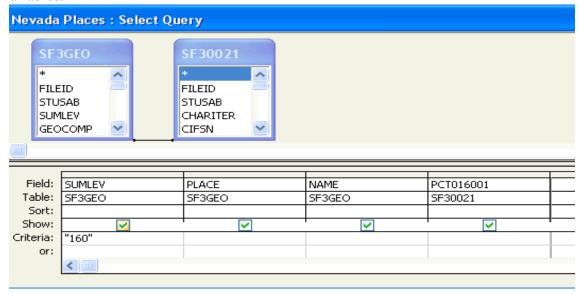
Then click Finish



Note well that many of the Census 2000 tables have too many variables (over 255 columns) for *Excel* to read. If you export the entire table, *Excel* will read a given number and then ignore the remainder. Therefore you must select the *Modify the query design* button to select which variables will be written and which will not. You may have to export a table twice to get all the variables should you include too many of them.

In the *Modify Design* window the desired variables are listed along with various attributes including whether each will be shown. Of importance here is the opportunity to restrict the query to certain types of geography. As mentioned earlier, the summary level codes (*SUMLEV*) and *FIPS* codes (e.g. *CNTY*, *TRACT*, *PLACE*, etc.) are used to do this.

4. Under *Criteria* enter "160" (quotes included). This will limit the query to places (ie cities). Note you also can limit the output by deselecting the green checks below the variables.



- 5. From the *Query* menu select the *Run* option and a table of Place *FIPS* codes and names will be generated.
- 6. When you have finished looking at the table, close and save it.

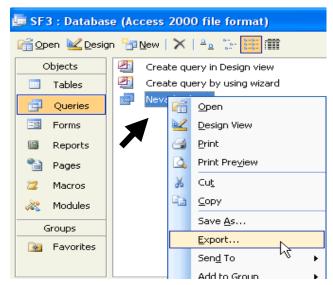
	SUMLEV	PLACE	NAME	PCT016001
•	160	04900	Battle Mountain	2967
	160	05100	Beatty CDP	1085
	160	05700	Blue Diamond C	307
	160	06500	Boulder City city	14966
	160	08100	Bunkerville CDF	1086
	160	08500	Caliente city	1132
	160	08600	Cal-Nev-Ari CDF	286
	160	08900	Carlin city	2187
	160	09700	Carson City	52457
	160	14090	Cold Springs C[3831
	160	17500	Dayton CDP	5902
	160	22500	Elko city	16599
	160	23500	Ely city	4027
	160	23770	Enterprise CDP	14437
	160	24100	Fallon city	7460
	160	24500	Fallon Station C	1391
	160	24900	Fernley CDP	8555
	l			

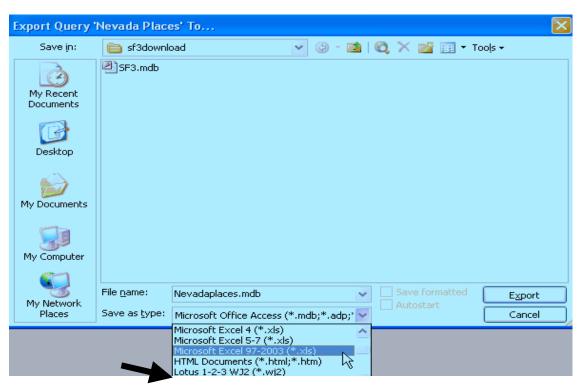
7. Back in the *Database* window, right-click on the new *Nevada Places* table.

It currently only exists in *Access* and if you want to use it in other programs you must export it from Access.

When the menu shown right pops up, select the *Export* option.

8. You can export your table in a number of formats. Select *Excel* from the list and export the table.

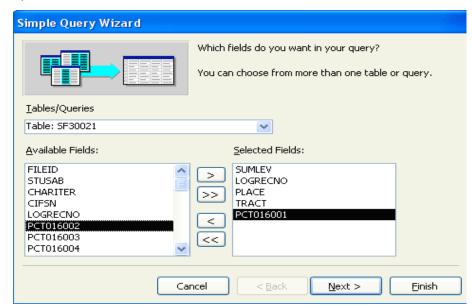




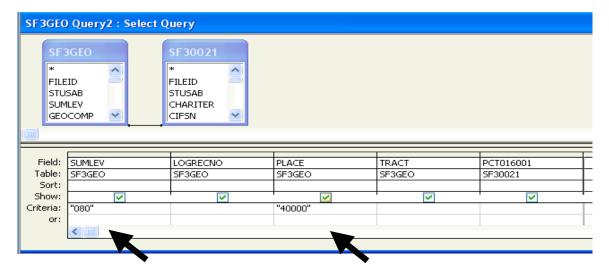
9. If you wish, import your table into Excel and print just the first page.

At this point you have covered the basic processes in processing and extracting census data from the raw files.

- 10. From your Nevada Places table look up the PLACE code for Las Vegas. The code is
- 11. Again select the *Insert* menu and the *Query* option. Choose the *Simple Query* Wizard and the *SF3GEO* table. Select the *SUMLEV*, *LOGRECNO*, *PLACE*, *TRACT* items from *SF3GEO*, the *PCT16001* item from *SF30021*. Click *Next*.
- 12. When you reach the last window make sure you have selected the *Modify the Query design* button. Then click *Finish*.



13. This time note the *Criteria* used to pull out census tracts within the city of Las Vegas. SUMLEV should be set to "080" to get tracts entirely in the city and PLACE should be set to "40000" to limit the tract search to just Las Vegas.



Below is the result of the query. Note all *PLACE* values are set to 40000 and Summary Levels are 080 reflecting only those tracts and tract parts that lie within

the city of Las Vegas. This is an example of a fairly typical query. Another typical query would be to pull out complete tracts within a county.

	SUMLEV	LOGRECNO	PLACE	TRACT	PCT016001
▶	080	0000282	40000	005810	0
	080	0000285	40000	005819	0
	080	0000766	40000	000101	6401
	080	0000772	40000	000102	7166
	080	0000778	40000	000103	5470
	080	0000784	40000	000104	7821
	080	0000790	40000	000105	3458
	080	0000795	40000	000201	3844
	080	0000800	40000	000203	4426
	080	0000806	40000	000204	1338
	080	0000810	40000	000301	3501
	080	0000817	40000	000302	5348
	080	0000823	40000	000400	9178
	080	0000831	40000	000503	7023
	080	0000837	40000	000504	5954
	080	0000842	40000	000510	5581
	080	0000847	40000	000511	6618
	080	0000852	40000	000512	9879
	080	0000858	40000	000513	3387
	080	0000863	40000	000514	6146
	000	0000000	40000	000545	4000

Exercises

- 1. Determine the *FIPS* code of Clark County, Nevada. Pull out the total population of all complete tracts (*Summary Level Code 140*) within that county. Print out the first page of the list of tracts.
- 2. For a state of interest determine what *SF3* files contain population and housing tables that are of interest to you. Download the *SF3GEO* file and those files that contain your tables. Link the tables in *Access* and then extract the variables from your collection of tables. Remember not to exceed 255 total items in one extraction if you plan to use *Excel*.