

LAB EXERCISE D: Working with Census Data

Querying and Analyzing Census Data

In this exercise, you will analyze a small portion of the 1990 Census data for the Greater Boston area. The Census data you need is in the file **C:\11_208\data\boston90.dbf**. The data dictionary for this file is at the end of this page.

In this exercise you may use both MS Access® and MS Excel®.

The database contains Census data for block groups in five counties in the Greater Boston area. The data dictionary at the end of this exercise lists the 3-digit census code for each county. Using Access, browse through the database file for a few minutes to get a general idea of its contents and size. It is overwhelmingly big, isn't it? It would be difficult to handle a database this large using only spreadsheet software. Using database management tools can make these projects easier.

Design a few queries and use the answers to fill in the following table:

Summary Statistics the Greater Boston Area, 1989

County Name	Number of Block Groups	Number of Census Tracts	Number of Households	Number of Households with Income under \$15,000	Percentage of Households with Income under \$15,000
Essex		150			
Middlesex	1231				
Norfolk			227868	31171	14%
Plymouth					
Suffolk					
Total		827		257716	

Some parts of the table have been filled out to help you check that you are on the right track. Think first about how to design queries that produce the needed information. Then, run the queries you need to complete the table. It is important to note that trying to design one single query to provide all the answers to this table is unlikely to work. In fact, a more straightforward approach is to design and run several queries. *One query at a time!*

Note: If at some point, you have a file that you'd like to save in order to perform spreadsheet-based calculations with in Excel, then you may export the file to Excel by choosing **Save As > Export...**, then selecting **To an External File or Database in the Save As** window. One approach is to *Save as type as (*.dbf)*, such as dBase 5. After it is saved, you may easily open the *.dbf file from Excel. Remember that if after performing calculations in Excel you want to bring the data from Excel into Access, you must either save it as a .dbf file or as an Excel file or as a .txt file. Then you must follow Import procedures from Access that you learned in Lab C (for importing *.dbf files) and in Lab B (for importing *.txt files) .

Note that before you save a spreadsheet as *.dbf format in MS Excel, you should make sure that 1) the entire spreadsheet is in 10 point or 12 point courier; 2) there are no formulas in columns, only values [you can **Copy**, then **Paste Special**, then **Paste Values**]; and 3) all column headings are re-sized.

MS Access Hint: Finding totals may require that you first set **Unique Values** in order to get a list without duplicates. At this point, however, counting this list may give you the same total as when you count the whole column because the computer is responding to the underlying table, not to the shortened column you've created. In order to get around this you should construct the query so that your window shows only the column(s) you want to count. (Note: At this point, you should have the **Unique Values** box checked and you shouldn't use any summary functions yet.)

To set **Unique Values** on: 1. Make sure you have opened a query in Design view. 2. Select the query by clicking anywhere in query Design view outside the design grid and the field lists. 3. Click

Properties  on the toolbar to display the query's property sheet. 4. Set the UniqueValues property to Yes. Note When the UniqueValues property is set to Yes, Microsoft Access

automatically sets the UniqueRecords property to No. If the query's results include more than one field, the combination of values from all fields must be unique for a given record to be included in the results. 5. To view the query's results, click **View** on the toolbar. 6. Optional: Use Query Make Table and save as new table. Or export data to Excel and put the data in a new worksheet. Save this data as a dbf file.

MS Access Hint: In some cases you may need to design separate queries for the statistics on the Total row and for those by county. The queries for the first three columns are relatively straightforward. The queries for the last two columns are somewhat more complicated: they require building expressions that combine several fields. You could do this by creating and saving several tables. Or more efficiently, by typing more complicated expressions the **Field** box in the Design grid. Examples of this include typing in expressions directly into the Field box that specify adding values: **field_name1 + field_name2** or dividing: **(field_name1 +field_name2)/(field_name3 + field_name4)** or even nesting expressions incorporating addition, division, and multiplication: **((field_name1+field_name2)/((field_name3+field_name4))*3**.

Extra credit: A more challenging question is to find the percentage of households who live *in block groups* with a median income of less than \$15,000. Hint: each record represents one block group.

Data Dictionary for boston90.dbf

BGRP	Blockgroup number (one digit)
COUNTY	County code
CTB_CODE	Composite county-tract-block group code (for mapping)
TRACT	Number of tracts (inconsistent record lengths)
TRACT2	Number of tracts (six-digit record lengths -- use this column if necessary)

HOUSEHOLD INCOME (Universe: Households)

HHTOTAL	Total number of households
HHY1	Households with less than to \$5,000
HHY2	Households with \$5,000 to \$9,999
HHY3	Households with \$10,000 to \$12,499
HHY4	Households with \$12,500 to \$14,999
HHY5	Households with \$15,000 to \$17,499
HHY6	Households with \$17,500 to \$19,999
HHY7	Households with \$20,000 to \$22,499
HHY8	Households with \$22,500 to \$24,999
HHY9	Households with \$25,000 to \$27,499
HHY10	Households with \$27,500 to \$29,999
HHY11	Households with \$30,000 to \$32,499
HHY12	Households with \$32,500 to \$34,999
HHY13	Households with \$35,000 to \$37,499
HHY14	Households with \$37,500 to \$39,999
HHY15	Households with \$40,000 to \$42,499
HHY16	Households with \$42,500 to \$44,999
HHY17	Households with \$45,000 to \$47,499
HHY18	Households with \$47,500 to \$49,999
HHY19	Households with \$50,000 to \$54,999
HHY20	Households with \$55,000 to \$59,999

HHY21	Households with \$60,000 to \$74,999
HHY22	Households with \$75,000 to \$99,999
HHY23	Households with \$100,000 to \$124,999
HHY24	Households with \$125,000 to \$149,999
HHY25	Households with \$150,000 and up

MEDIAN HOUSEHOLD INCOME (Universe: Households)

MEDY	Median household income for each blockgroup
-------------	---

CENSUS 3-DIGIT COUNTY CODES

County Name	County Code
Essex	009
Middlesex	017
Norfolk	021
Plymouth	023
Suffolk	025

EDUCATIONAL ATTAINMENT (Universe: Persons 25 years and over)

EDUTOTAL	Total Persons 25 years and over
EDU1	Less than 9th grade
EDU2	9th to 12th grade, no diploma
EDU3	High school grad (includes equivalency)
EDU4	Some college, no degree
EDU5	Associate's degree
EDU6	Bachelor's degree
EDU7	Graduate or professional degree