Problem Set Tips Sheet

Note: We expect your answers to the Problem Set to reflect the skills you have learned and to provide analysis of your results. We also expect a professional presentation based on your individual effort.

Clarifications

Some students have noticed inconsistencies in the problem set. We have updated the problem set a few times in an effort to clear these up. For grading purposes, we will expect you to use a reasonable interpretation of the version of the problem set that was online as of January 25, 2002 at 8:00 PM or any version we have posted since then, including the current online version. The later versions are intended as corrections to unintended errors or ambiguities, but we will not hold you accountable for the changes. Here is a summary of the changes:

> **Question 2B.** We intended to ask you about commuting **distance** rather than commute **time** in both questions 2A and 2B. However, in the original problem set, question 2B specified time instead of distance. The online version of the problem set has been updated to reflect this intent. However, if you answered this question using commute time instead of distance, you will be fine.
>
> **Question 3A.** The original problem set included this sentence:
>
> > Your answer should include tract number(s) and percentage for **each** of these counties.
>
> Actually, this should have mentioned **block group** numbers instead of **tract** numbers.
>
> Furthermore, what we meant by "tract number" or "block group number" is actually the full 11- or 12-digit identifier that you create to connect the Census data to a shapefile in ArcView® (e.g., 25017300100 to identify a tract or 250173001001 to identify a block group). The TRACTBNA field in the STF tables can contain repeats between counties (e.g., if county 017 has a tract 387202, then county 025 can also have a tract 387202). Similarly, the BLCKGR column in the STF tables will repeat block group numbers between tracts (e.g., if tract 387202 has a block group 6, then tract 387203 can also have a block group 6).
>
> The original problem set did not specify that we wanted the concatenated identifier, however, even though you need it to do the mapping required in Question 3B. For the purpose of the assignment, indicating the county, tract, and block group in separate columns is fine. The online problem set has been updated to reflect our intent.
>
> **Queston 3B.** The original problem set included this sentence:
>
> > Use the census block group map K:\11.208\arcviewfiles\stateplane\mablkgrp.shp to make a thematic map in ArcView® of Boston metropolitan area census tracts using the data on employment that you just collected for part (A).
>
> Instead of **Boston metropolitan area census tracts,** you should have been asked for block groups for the **three counties** that you worked with in part (A). Hopefully this was clear from the context, but we updated the problem set to make this explicit.

Tips

General Microsoft Access Tips:

Read the Tips sheet.

Add *only* those tables that you need for your analysis.

If you need to **join** tables, make sure that you join them on *unique* identifiers.

After adding a table, first display all the data in your results window to become familiar with the values. Then remove them from the results window and begin your query.

Be careful about designing complex queries to answer complicated questions. You are more likely to get correct results if you break down the question into parts. Some queries require multiple steps!

Use the Access query design grid to restrict your results to include only the records you want to analyze.

Limit your results columns in ways that allow you to group your answers appropriately.

Wrong answers are often due to overzealous or incorrect query design. Check to be sure you know what criteria you"ve included and that they fit the constraint of the problem.

Question 1

Note that SURVEY_NO in the employee.dbf does ***not*** refer to a unique ***record*** number. SURVEY_NO refers to the survey number for *each* DOT_BRANCH. However, each record is a unique survey response. Viewing the data displays this fact. Therefore, you may create queries and discuss the results in terms of "the number of people (i.e., survey respondents)" or "x percentage of total respondents responded y to question z." For example, you could say 761 of respondents drove alone as their primary mode of work that day.

Refer to Lab C.

Where the question refers to the four largest agencies, it is referring to the contents of the DOT_BRANCH column.

HINT: Read the survey questions carefully to help you understand the Data Dictionary. Pay careful attention to the wording of the assignment question to help you determine which survey variables to use. Note that some of the fields are very specific and actually refer to sub-parts of the survey questions. For example, the field PUB_DISCNT contains the codes 1 (Very likely), 2 (Somewhat likely), and 3 (Not at all likely) that refers to whether those respondents who presently drive alone, carpool, or vanpool are likely to take public transportation to DOT given the incentive of a discount bus or rail passes sold in their building.

Question 2

Try to understand what the variables actually mean.

Consider each question separately. Break the question into parts and query accordingly.

Question 3

Note that a county map (**K:\arcviewfiles\stateplane\macnt.shp**) is available--though not necessary--to enhance your map of tract-level analysis. You may add a second theme to display boundary outlines of tracts that otherwise might be blank space in your map. That is, add the theme again so that you can show all tract boundaries.

You will see several similar variable names in the Census data dictionary. Pay attention to field definitions and to how specific variables make up your universe. In this question, normalization refers to the process of identifying the fields you use for your universe and the process of creating percentages.

Use Lab E as a reference.

While there are several options for defining your universe for normalization, one good dependable approach is to add together the columns from the table category that you have selected fields from, since all together those fields comprise a universe. The STF files you need for any of these options are in **J:\stf3a1990\ma_nh**.

Before saving your final results as a .dbf file, make sure your calculated field is in number format (use at least 4 decimal places).

In this case, Boston metropolitan area includes the five counties in your analysis. However, if you like, you may add an inset map to better display the tracts in the Boston city area.

Question 4

The field **SqrMeter** in Bostown.shp refers to area in square meters. You may wish to add a field in ArcView® and calculate a square kilometer value that may be more meaningful in determining density.