

**11.520: A Workshop on Geographic Information Systems**

**11.188: Urban Planning and Social Science Laboratory**

**Lab Exercise 5 Notes:  
Working with 2000 Census Data & MIT Geodata  
Repository**

---

## **Overview**

In this exercise you will use the Census STF3 files to create a thematic map indicating the percentage of workers in block groups in Middlesex County who drive alone to their jobs. You may find the notes from the "Making Sense of the Census" lecture a useful reference.

Your goal is to create a table that contains the percentage of workers who drive alone to work (by car, truck, or van) for each block group in Boston Metropolitan Area. Note that unlike the simple example shown in class, to complete this task you must normalize the data - that is, in order to compute a meaningful percentage, you will want to divide the number of workers who drive alone to work by the total number of workers who work outside their home.

### **Normalizing data & Finding the Appropriate Universe**

To further illustrate the ideas about normalizing data, let's consider a different context. Suppose we want to determine the fraction of rents that are less than \$500 in each block group. To compute this fraction, we need to normalize (that is, divide) our count of the number of units renting for \$500 or less by the count of the total number of units in each block group that are rented. We can add together several columns from the "Gross Rent" table to get the appropriate numerator. But, how do we determine the appropriate denominator. The technical documentation for the 'Gross Rent' tables indicates that the associated 'universe' is "Specified renter occupied housing units." The total count of renter-occupied housing units in each block group is listed in another table - the "Tenure" table. The sum of the rental units in various rent categories would need to be divided by this value. So the numerator and denominator in our fraction would be drawn from separate census tables (H8 for gross rent, and H43 for tenure). We can use the field "Logrecno," the logical record number, to link together the data that we extract from the two tables.

Why normalize the data? Comparing the raw numbers of housing units per block group may be deceiving, as the total number of renter-occupied housing units will vary from one block group to the next. By dividing the number housing units with rent less than \$500 by the total number of housing units, we obtain a fraction of the total occupied housing units with rents under \$500. This fraction may be compared fairly among block groups.

## **Data Preparation**

The geographical layers containing the census boundaries for this lab are available from the MIT Geodata Repository, instead of using the data from M: drive as we usually do. The census data tables that you will pull into MS-Access are fairly large. The original copies in the class locker are read-only. After making your own copy and extracting the data you need for Eastern Mass, you will want to clean up your working files and toss the large tables that you no longer need.

- Use the MIT GeoData Repository to get boundary files from the MIT Libraries Spatial Database Engine (SDE)
  - Use MS-Access to import and merge the appropriate census data tables, extract the needed rows and columns, and build a column that can be joined with the geometry from the MIT GeoData Repository. In order to do this, you will need to know how to use the Query Building part of MS-Access.
  - The read-only M:\data\census2k directory in the class locker is almost 100 MB. You'll need to copy some of these files to your private workspace. However, once you compute your 'drive-to-work-alone' ratios and confirm that you can match them to the eastern Mass block group map, you will want to erase some of the large census tables that you have copied into your workspace.
-