11.520: A Workshop on Geographic Information Systems

11.188: Urban Planning and Social Science Laboratory

Homework 1

Due Lecture 5, at start of class.

NOTE: Homework assignments will take longer and be much more important to your grade than any one lab exercise, so devote your energy accordingly.

In this exercise, you will explore the spatial patterns of the housing and socio-economic characteristics of communities in and around Boston. To assist in this task, we provide (1) demographic data at the census tract level from the 1990 US Census, and (2) boundary files for cities and towns, major roads, shopping center locations, and census tracts. We ask you to use these data to prepare three *exploratory* maps and one *explanatory* map.

Before starting the hands-on work, read through the entire assignment to get a sense of the datasets, analytic approach, and processing steps. Then, make sure you can access the datasets in the 11.520 Locker.

The census tract boundaries are saved in a 'shapefile' called `msa5_tr90.shp` and may be found in the **M:\data** directory. This shapefile contains only the bouandary geometry and it must be paired with a dbf-formatted file (called `msa5_tr90_data.dbf`) in order to relate the census data to specific census tracts. The attribute table for the census tract shapefile only contains a few geographic identifiers (like county, track number, etc.). Some additional socioeconomic data for these tracts have been pulled from the 1990 Census SF3A datasets and are stored in the same **M:\data** directory in a dbf-formatted table called `msa5_tr90_data.dbf`. This file must be 'linked' or 'joined' to the attrbiute table for the tract boundaries by a common field called "**STCNTYTR**" before you will be able to generate thematic maps using the census data. (STCNTYTR is the abbreviation for STate-CouNTY-TRact.) Use the ArcMap help files to see how to 'join' the data table to the attribute table if you want to get started with the homework before we show you how to do this in class.

The `msa5_tr90_data.dbf` table includes 60+ variables from the much longer list of all variables in the decennial Census. Take a look at the dictionary for the specific census data fields in msa5_tr90_data.dbf. (Note: this list is a subset of the full Census Bureau's

listing and technical documentation for the hundreds of population and housing variables from the 1990 census. This technical document is archived in the class locker. More details about the 1990 US Census are available in Tom Grayson's and Annie Thompson's notes on "Making Sense of the Census". Be aware, however, that some of the online references in these notes are no longer available and you will need ONLY the shorter list of 60+ variables mentioned above in order to do the homework.)

Besides the census data, which will be used primarily in Problem #1, you will need a map of major roads and shopping centers for Problem #2. The shopping center coverage (for the Boston metro area) is called ***shopcntrs*** and is stored in **M:\data**. The major roads layer is called ***majmhda1*** and can be also found in the directory **M:\data**. All of these coverages use the following coordinate system: Massachusetts State Plane, Mainland Zone, NAD 1983, meters. Be sure to set the map units and distance units in the ***Data Frame Properties*** window so you can measure distances in your Data View window and be sure that the distances you compute in Problem #2 are reasonable. The roads coverage comes from the Mass Highway Department via MassGIS and the shopping center coverage is proprietary data provided by SSR Research (circa 1995) for internal MIT educational use.

A map should always have a purpose. A good map should deliver the information that you want readers to understand. Therefore the map should be very intuitive without requiring reading the discussion of the map in your paper or report. Try to give the map to your friends who have no training in GIS to see if they can realize the message you were trying to deliver.

- Your goal is to gain some understanding of housing and socio-economic patterns in metro Boston.
- Try to have fun and explore ArcMap while doing this homework. Use the on-line help to experiment with ArcMap's capabilities. Here's another mapmaking hint: you can specify different colors for the foreground, background and outline of a symbol, and different widths and sizes for lines and symbols. Exploit this functionality when overlaying different data layers.
- To make your maps more readable, you should make use of map files showing the political boundaries of Massachusetts towns. Also consider using the water bodies in Eastern Massachusetts. The former coverage is located in **M:\data\matown00.shp**. A shapefile of Eastern Massachusetts water bodies is available in **M:\data\msa_water.shp**.

Problem 1: Exploratory Mapping
Metropolitan Area Census Data (60 points)

1. [20 points] Create a thematic (or chloropleth) map showing the population density of the MSA. You should calculate density as persons (**P0010001**) per acre (**landacre**). Normalizing by the 'landacre' variable in the census data is more reliable than using the 'area' variable since the tracts extend into Boston harbor, the Charles River, etc. whereas the 'landacre' variable is the census estimate of land acreage within each Tract. Classify

the data into six quantiles, and be sure to include only those polygons (census tracts) where `landacre` > 0 and the relevant census data are not missing. (For example, census data are missing from a cluster of tracts north of Boston.) Use the **Select > Select by Attribute** menu option to select the tracts you want to include.

2. [20 points] Map the homeownership ratio--the ratio of owner occupied housing units to the total occupied housing units. Remember that the '**tenure**' variables count the number of owner-occupied and renter-occupied housing units. Be careful about what is the numerator and what is the denominator. Just as for the previous map, you will also need to exclude tracts that lack adequate data. Include a brief few sentences explaining your choice of classification scheme.

3. [20 points] Map another Census attribute of your choosing with interesting spatial patterns using the same process as described in number two.

- For all three maps, be sure to include a legend, title, source, logo, north arrow, and scalebar. The quality of your presentation **does** count!
- In separate text explain the "story" that your map tells. What spatial patterns, if any, are suggested. (Don't try to overexplain the map!) Explain the reasoning behind your classification choices, how you handled missing values, and any other relevant judgments and assumptions. Please limit this discussion to a total of one page. We do **not** want lengthy explanations. There is more than one reasonable choice for your classification and normalization choices. We simply want to be sure that you've thought about the issues and made reasonable choices.

Problem 2: Introductory Spatial Analysis
Relationships between Roads, Shopping Centers, and Residences

For this problem, you are asked to investigate the relationships among the locations of shopping centers, major roads, and residential clusters. Compared with the exploratory mapping you did in the first problem, you are asked to dig a bit further into the data, develop a few specific measures that carefully exclude incomplete or inapplicable data, and then develop maps that successfully visualize the results and reasoning behind your analysis.

The shopping center data are stored in a shapefile at `M:\data\shopcntrs.shp`. (These data are proprietary and not to be used or redistributed for non-MIT purposes.) Included in these data are characteristics such as square footage of retail space (**totalsf**) and type of center (**propertysu**). Explore these two variables to try to determine if a relationship exists between them. To do this you may want to calculate the average size of each type of shopping center. Note that not all observations include a value in the **totalsf** variable field. Note that these shopping 'centers' do *not* include places like Central Square (Cambridge) where commercial/retail activity is present among individually owned parcels and buildings along a city street. This dataset focuses on shopping center developments where a large tract of land or strip mall under common ownership is divided up into clusters of businesses.

1. [20 Points] Create one or two maps showing the relationship(s) between shopping center location and the location of major roads and population centers. Be sure to use different symbols and/or sizes on your map for the different types of shopping centers. Likewise differentiate major and not-so-major roads -- use the **class** field in the `majmhda1` attribute table. Use high population density as an indicator of where population centers are located and shade the tracts based on population density (as you computed it in problem 1. Be sure to turn the tract outlines off so they don't clutter the map. In fact, it will take some effort to develop maps with good symbolization and cartographic choices so that they are both readable and informative.

2. [20 Points] Select the Interstate Highways and Routes 2 and 3 (not 3A) from the major roads layer. Use the buffer tools to create a one mile buffer around these selected roads and then calculate the share of shopping centers by type (among those within the 5-county region) that lie within the buffer. Map your results. Include on your layout a map of the entire area, as well as a more detailed map zoomed into an area of the region with interesting spatial patterns. In addition to your map(s), create a table showing, for each type of shopping center (**propertysu**), the number of such shopping centers within the 5-county **msa5_tr90** region, and the number and percentage of each type of shopping center that fall within the buffer areas.

Explain in a couple of paragraphs (a) the steps you took to select the roads and shopping centers that you used to compute these statistics, and (b) your interpretation of any general pattern that you observe regarding the location of shopping centers, major roads, and population centers. In particular, be sure that your discussion covers:

- What conceptual relationships are these maps intended to portray?
- Which classification schemes did you choose to use, and why?
- Which shopping centers and tracts have you excluded from the analysis, and why?
- Do there appear to be any interesting spatial relationships shown on the map?

Howework Requirement

Don't just turn in the maps! Write up a short report that integrates the maps, tables and the explicit answers to both questions. Use the maps and tables in the paper to illustrate and amplify your verbal reasoning rather than simply to produce maps without a stated context and purpose.