GIS Software, Applications, and Management: Providing Public Access to MapInfo and ArcView

Barbara Levergood, University of North Carolina at Chapel Hill
Chapel Hill, NC

Good morning! I would like to do three things that I hope will give you a sense of what can be done with GIS in a depository on a relatively limited basis. First, I'll use a discussion of why we provide GIS services in our library as an excuse to show you two examples of common queries that can be answered and simple maps that can be made using GIS. Second, I'll outline the services that we provide for GIS in our library. Third, I'll take you through the major steps involved in making a thematic map using GIS.

I. Why We Provide GIS Services

There are two primary reasons that we wanted to get into GIS in our Documents Section. First, we wanted to provide access to census maps (e.g. block maps) that were not depository anymore in paper or fiche. Let me give you an example. With the GIS, and using the depository TIGER/Line data from the Census Bureau, we can create digital maps of streets (overhead 1), census blocks (overlay overhead 5), or census block groups (overlay overhead 4) and print them out. This level of geographic detail is not available in any other format from Census through the FDLP.

However, users commonly request this type of map from us. For example, a user might want to collect census data for a community (overhead 1). The problem is that they might have the names of the streets that bound the community but they would not normally know which blocks or block groups or census tracts define that community. Without that information, the user cannot collect census data. The GIS can help the user identify the census geographic areas.

We might start with a street map of the neighborhood in Durham County, North Carolina. This map was created using depository TIGER/Line data and MapInfo GIS software. Using the GIS, we would then overlay a block map of that area (overlay overhead 2). The user would identify the community and we could shade in the blocks that are a part of the community. We might overlay a map of the census tracts (overlay overhead 3); here we

discover that the community is entirely within census tract 9. But notice that the community is just a small part of the tract; collecting data for the entire tract might not correctly characterize the community. So, we overlay a map of the block groups (overlay overhead 4); here we see that all of block group 2 is within the community and part of block group 3 is also. The user might decide now to collect data for both block groups or for just block group 2 and ignore that part of the community within block group 3. These are the only options available if the user is interested in socio-economic information, since that data is available at the block group level but not at the block level. Alternatively, the user may be interested in block data also. So, we can number the blocks that define the community (overlay overhead 5). The user can print out this map or any of the other maps we have seen as is or considerably elaborated with titles, legends, scale, etc. The user will now have the needed information: the identification of the census geographic areas needed to define the community and to collect census data for.

Note that these maps include spatial (geographic) information: streets, blocks, block groups, and census tracts; along with street names, numbers, and map titles. But there was no statistical information represented on the maps.

This leads us to the second reason that we wanted to provide mapping capabilities for users, for the 1990 census statistical data especially (overhead 6). This is a simple thematic map showing the geographic distribution of population density and African Americans in Durham County census tracts. This map is more complex than the maps I showed you before in that it includes not only the spatial data (the census tract boundaries) but it also includes a representation of statistical data. The colors represent population density and the dots represent African Americans. This map could be used to answer the question: "Do African Americans in Durham County, North Carolina live in areas of high population density?" This type of map is also commonly requested.

We got into GIS, then, for two reasons. First, we wanted to provide access to small area census maps that were no longer depository. Second, we wanted to provide thematic mapping capabilities for 1990 census data.

II. Our Services

Next, I will outline the services that we provide for the GIS.

Our Documents Section is a Regional library. We provide public access to GIS software and spatial data in support of Government information. In Documents, we have three full-time librarians and three full-time paraprofessional staff. I am the only full-time staff member who works with the GIS, but I have a student or two whom I train to help me. Our Maps Library, a separate unit in a different building, holds most of our depository maps but it does not have GIS.

We provide public access to two pieces of GIS software, MapInfo Professional and ArcView, on a single workstation in Documents. Users can print out in black and white on a laser printer or in color on an inkjet printer.

We ask that people contact us ahead of time to discuss what software, spatial data, and statistical data are available and suitable, to set their expectations, and to make an appointment for instruction. Ahead of time, we prepare the spatial data files that the user will need. We support the depository TIGER/Line data from the Census Bureau as well as TIGER, which is commercial data based on TIGER/Line and produced by Wessex. During the appointment, we spend 30-60 minutes assisting the user in learning how to use the GIS software and how to access and manipulate the spatial and statistical data via the GIS in order to complete the project. We then leave the user alone to do so, remaining available to answer questions that inevitably arise. We provide documentation and custom handouts in support of the software and data.

The GIS is a very heavily used resource and it is growing in popularity, by folks from both on and off campus. Each of a user's sessions may be from 1 to 8 hours or more, with 2 hours being a typical minimum. We find that because of the complexity of the software and data, users need a high level of service.

We do not provide full service for the GIS. For instance, we do not create a map for a user. Rather, we familiarize the user with the software and data and they create the map. Recall that our mission is to provide access to the GIS in support of Government information. Unfortunately, we can provide a minimum level of service (or none at all) to those who are interested primarily in creating an attractive cartographic product not using Government information since that is not a part of our mission in Documents and because of limitations of staff and hardware resources.

The second major service that we offer is to provide spatial data files for users who want to use them on their own computers using their own GIS software. This is unproblematic in the legal sense for the Federal spatial and statistical data, which is public domain. But spatial data files are huge, so we have developed a number of ways to get them to users, including an FTP site that we use to store large files for users to download.

III. A Sample MapInfo Session

Please turn to the back page of the handout. I would like to guide you through this sample MapInfo session just to give you a feeling for the steps involved in making a thematic map. Please be aware that there are *many* other variations on how to work with the data and the software that I will briefly describe later.

Let's assume a very simple query: I would like a map showing the distribution of population in Durham County by census tract.

1. (overhead 10) We might begin by collecting the statistical data. We'll use the 1990 Census of Population and Housing, STF 3A, to create and download a dBASE III file giving the population of each census tract in Durham County, NC. That file must have the census tract numbers in one column, indicated by the arrow.
2. We will make the map using MapInfo, although we do not open the software up yet.
3. (overhead 11) In our library, we use TIGER/Line 1992 spatial data along with MapInfo.

4. (overlay overheads 11 and 12) GIS software in general, and MapInfo in particular, cannot read the raw TIGER/Line files directly off the CD-ROM. First, the spatial data that we want, census tract boundaries for Durham County, must be extracted from the TIGER/Line files and translated into a format that MapInfo can import. We use TigerBdry software to accomplish this. You are looking at the batch file that we use to create our desired map.
5. (overhead 13) Once the spatial data has been extracted and translated, we fire up MapInfo and import the file into the native format.
6. (overhead 14) We then open the resulting file within MapInfo and we display the map of Durham County census tracts (overlay overheads 14 and 15). Associated with this map is a database containing census tract numbers. The arrow is pointing to that column.
7. (overheads 10 and 15) Now, if we compare this spatial data database containing the census tract numbers that came from TIGER/Line with the statistical data database containing the numbers that we downloaded from the STF 3A file, we see that the format of the numbers is different. However, we want MapInfo to be able to join these two databases based on these census tract numbers. To do this, the numbers must be identical in every detail. Thus, we must edit the file to make the numbers identical.
8. We will also have to do the necessary calculations if the user wants to map density or a percentage or any other number that is not already in the statistical datafile.
9. (overhead 16) We are now ready to create the thematic map. This can be a fairly confusing process involving joining the statistical and spatial data files, choosing the thematic values for the map, and customizing the map.
10. We can label the map.
11. We can work with the layout of the map and the legend.
12. (overhead 17) Finally, we can print out the thematic map showing the population density within Durham County census tracts.

The estimated total staff and user time involved in making this map is 1-1/4 to 2-1/2 hours or more.

We just made a map using MapInfo along with depository TIGER/Line and census data. There are many other alternatives. Instead of MapInfo, you could use ArcView along with the depository data, as long as you have the appropriate extraction and translation software to use in step 4. The other steps would be similar.

Another alternative is to purchase spatial and/or statistical data that has been customized for use with either MapInfo or ArcView. Let me quickly take you through that same sample session, this time using ArcView along with TIGER and census data, both purchased from Wessex.

In step 1, instead of using the statistical data from the Census Bureau, you would create a file using the commercial package. In step 2, use ArcView instead of MapInfo. In step 3, instead of using TIGER/Line from Census, you would use commercial TIGER. In step 4, it would be necessary to extract the data from commercial TIGER from Wessex, but translation is not required. This can save considerable staff time. In step 5, the import step is not required, again saving time. Step 6 is similar. Step 7 may not be required, since the

files may already be customized. This is a very time consuming step and we are glad to save that time.

You may be wondering why we would ever make maps using the TIGER/Line spatial data when I indicated several times that we could save time and aggravation by using the Wessex TIGER data. If you look on page 4 of the handout, I outline some of the advantages and disadvantages of these two options. There are two major advantages of using depository TIGER/Line. First, block boundaries are included in TIGER/Line, they are not included in Wessex's TIGER. Second, TIGER/Line is public domain. Consequently, we extract and translate those files and give them away to our users fairly often.

Let me emphasize there are many variations on the procedures I just outlined that depend on the GIS software, the spatial and statistical data resources that you have through the FDLP or that you purchase, and any translation software that you purchase.

This then is a greatly simplified overview of the GIS services and spatial data that we provide to the public. I hope I have conveyed my belief in the importance of GIS as a tool working with statistical data, as well as an appreciation for the enormous resources that can be devoted to the creation and maintenance of these services, software, and data.

In the handout, I have included a lot of other information that you may find useful if you are thinking of getting into GIS. On page 2 is a step-by-step outline of how we have developed our GIS services and resources and some issues and problems with our services. On pages 2 and 3 is an estimate of the costs in terms of software, data, hardware, and staff. On page 4 is an outline of the spatial data that we support and how we make it accessible. This is a critical page and worth looking at in detail if you are interested when you get home. On page 5 is an overview of the statistical data that we frequently use with the GIS. Finally, on page 6 is the sample MapInfo session.