Chapter II

An Introduction to GIS
(All Things Data)

One of the purposes of this book is to introduce community health groups to the potential of GIS, a technology that can help in understanding the spatial landscape of prenatal risk. Therefore, one of the first steps is to provide a brief overview of GIS. This introduction will be split over the next two sections, with this chapter focused on data issues associated with using a GIS, while the next presents an introduction to the functions of a GIS that make it so powerful: the ability to analyze and visualize spatial data. These two chapters are not meant for experienced GIS users (though even for these a few points and references from the non-geographic literature may prove to be useful). It is also not meant to be a comprehensive introduction to the science; there are several other excellent texts serving that need. These next two chapters are meant to give a basic understanding, and inform enough to encourage the adoption of a GIS approach.

Most people reading this book will probably be using a vector GIS. There are, however, two basic GIS formats, raster being the other type. Raster GIS is best suited for surface or complete coverage data (for example, vegetation cover) because the spatial surface is transformed into a grid, with each cell or pixel containing a relevant geographic attribute (such as 1 = water, 2 = forest, etc.). These pixels are fixed in space, allowing multiple layers at a single location to be compared and analyzed. Vector GIS, which is a more useful GIS format for the type of investigation likely to be performed by a health unit, contains points, lines, and areas. Unlike in a raster GIS, each spatial object has its own geography. The mother’s residence is a point on the city map, her bus route is a line, and she
resides in a zip code, which in vector GIS-speak is a polygon. Most of the following descriptions will refer to vector GIS topology.

Many texts have been written defining and describing the constituent components of a GIS. There is no point in replicating these here, though the reader is recommended to (Longley, Goodchild, Maguire, & Rhind, 2001) if he or she wishes to gain a better understanding of the subject. For this book, I will use the definition I always give to my class, or at presentations to non-geographic audiences: A GIS is a tool to work with, analyze, and visualize spatial data. As its input data are spatial, the map is the central part of the system. On this map a variety of spatial layers can be turned on or off according to the needs of the investigation. It is rather like a multilayered cake. One can take a huge mouthful, which may result in the individual tastes of the layers being missed, and even some crumbs spilling out and being lost. Or, the cake layers can be separated and eaten one at a time, or in smaller combinations to satisfy the taste buds.

If this seems like a rather flippant analogy, let’s use real examples. Figure 1 shows an online GIS available through the South Carolina Community Assessment Network (Department of Health and Environmental Control), otherwise known as SCAN. This GIS allows web visitors to select from a variety of databases to make charts, tables, and maps. Mappable data include the results of the South Carolina PRAMS (Pregnancy Risk Assessment Monitoring System) study (PRAMS will be described later in this chapter), pregnancy (live births, abortions, fetal deaths), mothers health and lifestyle, infant and child health, and infant mortality. If one of these data layers is selected, for example infant mortality, multiple GIS options become available as easy to navigate drop-down menus. Defaults are already checked on these menus for the easiest of map generations. The GIS user can select different subcategories of the outcome to be mapped — for example, race, ethnicity, age, birthweight, gestation, and ICD (international classification disease) codes. Figure 1 displays a snapshot of one of these maps. Any interested party could make maps for the years 1989-2002, save the images, and compare the outputs both spatially and temporally. Similarly, data can be outputted to tables (in Excel) for further analysis. An option is also available for more sophisticated analysis (such as chi-square, linear regression, and other forms of GIS analysis) through personal contact with the Division of Biostatistics and Health GIS.

A series of GIS operations are available on the bottom right of both maps. Notice that more options are available for the zoomed-in map. This is because the larger the scale of the map (the zoomed-in map is a larger scale than the state-wide map) the more spatial layers can be displayed. The user has the option to turn these spatial layers on or off as he or she desires. This is because more space is available on the larger scale map. If we tried to show all roads on a state level map the surface would be covered in a blackness of lines. We therefore make