1.4 Geographical Information System (GIS)

In the highly dynamic and complex world 'information' has become a critical resource for effective and efficient management of organisation. Information Technology in its various forms is enabling organizations to churn raw data into meaningful information for effective decision making. One such form of Information Technology (IT) is Geographic Information System (GIS). It is described as: "An organized collection of computer hardware, software, geographic data and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information".

According to this definition and as shown in Figure.1.3, GIS includes not only computing capability and data, but also manages the users, and organizations within which they function and institutional relationships that govern their management and use of information.

GIS system design and implementation planning are not a separate process. They must occur in conjunctions with one another.

1.4.1 Elements of GIS

The major elements of GIS are geography-the actual location, information- the description of the location, and the system to integrate and perform the required GIS functions.

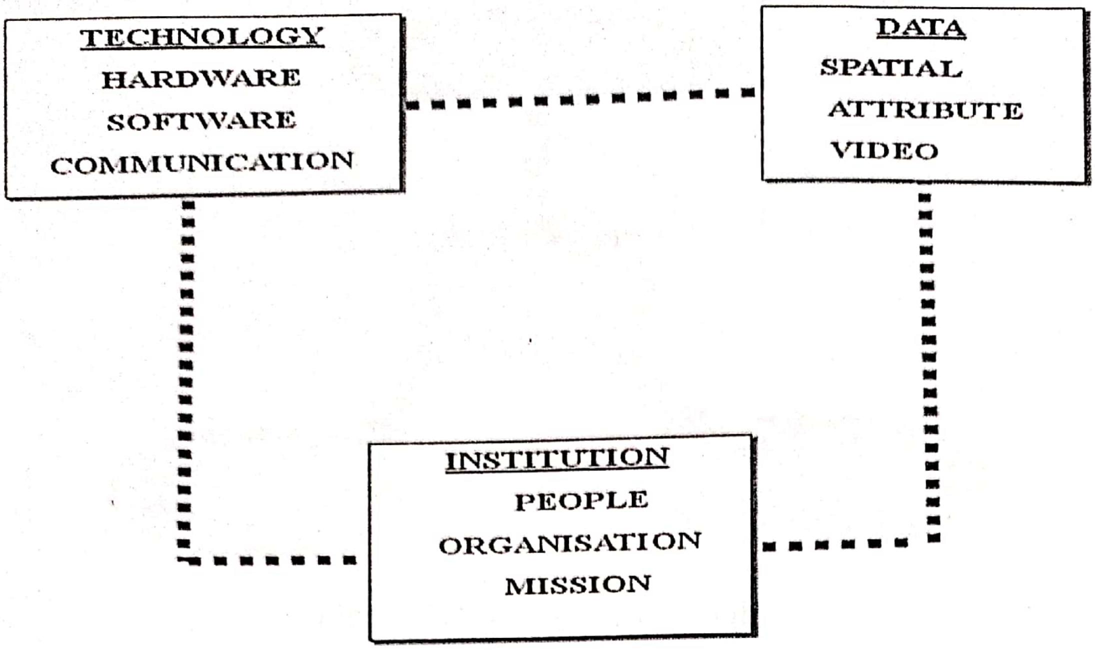


Figure 1.3: Domains of GIS

i) **Geographic:** The system is concerned with data relating to geography and geographic scales of measurement. This is referenced by some coordinate system to locations on the surface of the earth.

ii) **Information:** The system allows for the storage and extraction of specific and meaningful attributes information. These data are connected to some geography and are organized around a model of the real world. Spatial and non-spatial queries are made possible.

iii) **System:** An automated system should include an integrated set of procedures for the input, storage, manipulations and output of geographic information.

GIS is an integrated single platform of three areas viz. the relational database management system to store spatial and non-spatial data; cartographic capabilities to depict, graph and plot geographic information; and spatial analytical capabilities to facilitate manipulation and spatial analysis.

1.4.2 Components of GIS

There are four components of GIS: (1) data, (2) hardware, (3) software, and (4) users. As shown in Figure 1.4, the components must be integrated to support the management and analysis of spatial or mapped data. Data tends to be at the center of any GIS system, while the computer components of the system support the data management and analysis.

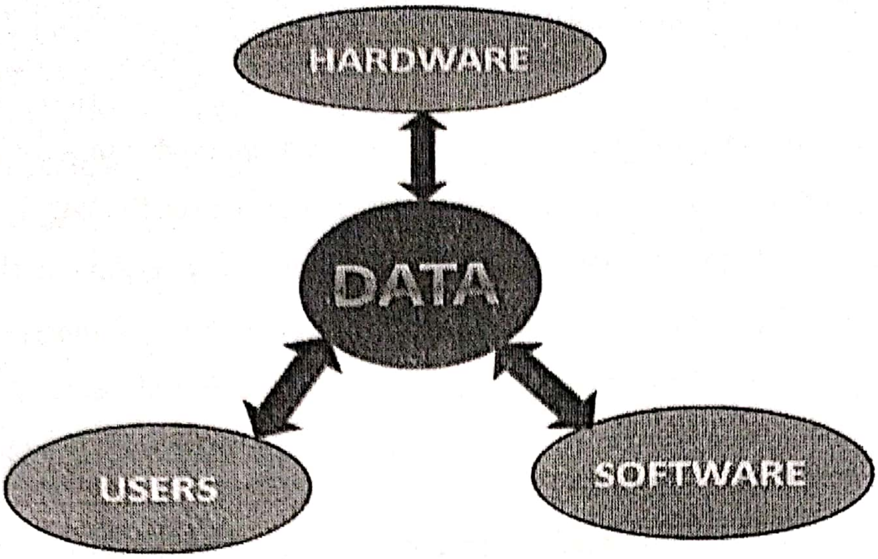


Figure 1.4: Components of GIS

1.4.2.1 Data

GIS components are dynamic; there has been rapid change in the computing industry as well as turnover of personnel involved in GIS projects. For this reason, GIS developers are often encouraged to adopt a data-centered approach. Simply stated, a data-centered approach views data as the central resource in the GIS. Though data may be shared among multiple users and multiple hardware/software environments, the data is collected and compiled by a person or organization to support the goal of that user. The other components provide the support needed to process that data.

All data in a GIS are either **spatial** data or **attribute** data. As discussed above, spatial data tells us **where** something occurs. Attribute data tells **what** occurs; it tells us the nature or characteristics of the spatial data. Every GIS provides the ability to store and manipulate both the spatial data and the associated attribute data. Spatial data includes information pertaining to location of objects of interest, their distribution and extent, adjacency, proximity and connectivity, verses, attribute data, or observations about features.

1.4.2.2 Hardware

Computer hardware used to support GIS is a highly variable part of the overall system. Users will customize their hardware environment to best meet their own individual needs. In all cases, however, a fully functional GIS must contain hardware to support data input, output, storage, retrieval, display, and analysis.

1.4.2.3 Software

Software is also a highly dynamic part of the system with large number of GIS software packages existing now. These systems are available for almost all types of hardware platforms and come with a wide variety of functional capabilities. The range in software options goes from generic turnkey systems that are ready for use "right out of the box" to customized installations designed to support specific user needs, like TransCAD (a GIS software tailor made for transport applications).

Given that sometimes-bewildering array of choices for hardware and software, selection and use of a GIS should be approached strategically. In all cases, anyone considering a GIS software package should consider needs carefully and consult various references, including other users, vendors, and technical publications.

1.4.2.4 Users

The final component required for a true GIS is users. The term "user" may refer to any individual who will use GIS to support project or program goals, or to an entire organization that will employ GIS in support of its overall mission.

GIS users are often envisioned as hands-on computer processing people. While this is in part true, we choose to define a broader spectrum of GIS users. One classification scheme classifies users into two groups: **system users and end users**.

System users are those persons who have actual hands-on use of the GIS hardware and software. These persons have advanced technical skills in the application of GIS to problem solving. System users tend to be responsible not only for the day-to-day use of the system, but also for system maintenance and upkeep.

End users are those persons who do not have actual hands-on use of the system but who do make use of the information products generated via the GIS. End users do not necessarily have to possess hands-on technical skills. However, they must be able to communicate effectively and interact with system users in order to make requests for information products, and must also understand the limitations and requirements of GIS-based processing.

1.4.3 Conceptualization of GIS

Conceptually, a GIS can be envisioned as a stacked set of map layers, where each layer is aligned or registered to all other layers. Typically, each layer will contain a unique geographic theme or data type. **The GIS database stores both the spatial data (where something occurs) and the attribute data (characteristics of the spatial data) for all of the features shown on each layer.** These themes may include, for example, topography, soils, land-use, cadastral (land ownership) information, or infrastructure such as roads, Traffic Analysis Zones (TAZ), pipelines, power lines, or sewer networks. Figure 1.5 gives a schematic view of geographic layer system in GIS. By sharing mutual geography, all layers in the GIS can be combined or overlaid in any user-specified combination.

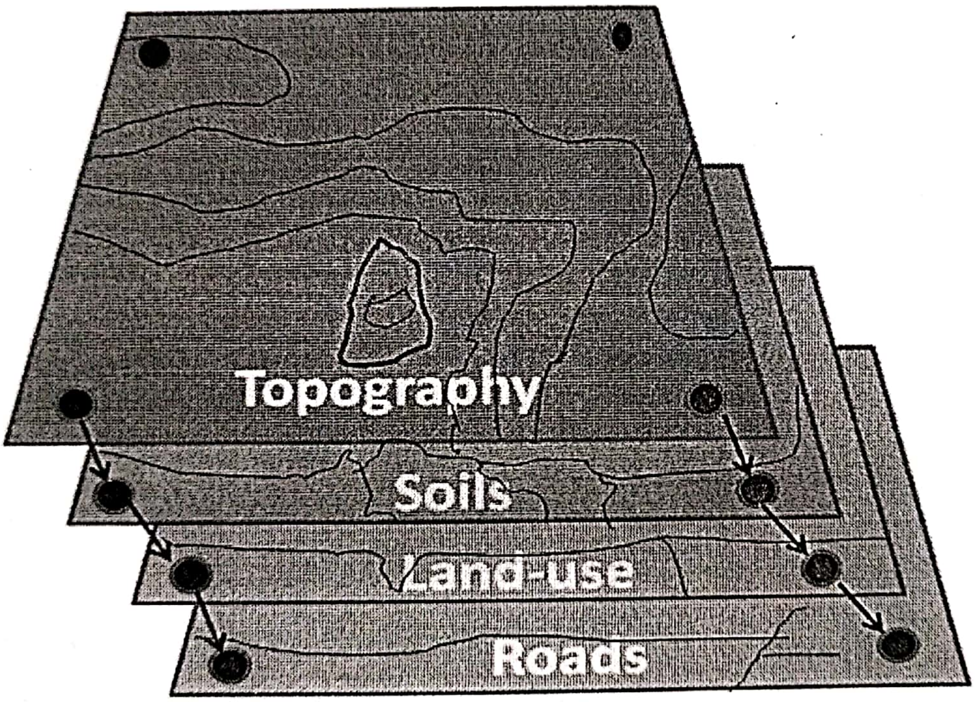


Figure 1.5: Mapping layers of GIS

1.4.4 Maps and Map Data Handling

Before examining the individual components and functions of GIS, it will be helpful to consider the nature of maps and mapped data. A GIS works with observations or measurements that can be tied to a specific geographic location on the ground. Another common term for mapped data is spatial data. Spatial data vary with location, so the nature of the data that we collect, measure, and interpret will change as we consider various locations on the earth's surface. Observations of the earth's surface are recorded on maps to portray the spatial data in a format that is easy for humans to comprehend. Maps, as devices used to communicate the nature of spatial data, are the focus of the next section.

All maps are simplifications of the real world. The true earth is infinitely complex and it is not possible to depict on a map all of the real earth features that we might be interested in. As a simplified image of the earth, maps can be called models of reality. Though they are simplifications, these models are quite sophisticated, and the science of map making, called cartography, is a formal geographic science.

A true map must accurately show not only what the nature of the mapped variable is (in correct proportion), but also must correctly place all mapped data in their true geographic locations. The geographic location of spatial data can be classified into two types: "**absolute**" and "**relative**". Absolute location refers to a unique and standardized place or position, while relative location defines position based on the location of other variables or phenomena. Absolute geographic position is specified using a universal coordinate system such as Latitude/Longitude or Universal Transverse Mercator (UTM) coordinates. These universal coordinate systems allow both the mapmaker and the map user to specify a unique and definite position for every location on both the earth and maps of the earth. This unique location, which is "tagged" to all spatial data, is critical to being able to store and analyze data in a GIS. It is this geographic tag or characteristic that distinguishes GIS as a technology that focuses on mapped data.

Generally there are three classes or families of map projections. Each geometric shape is used to transform the globe (a curved surface) to a plane (the map surface). These three families are called "developable" surfaces because planes, cones, and cylinders can be "flattened" without distortion (Figure 1.6). Azimuthal or planar projections use a flat two-dimensional surface to develop the map, conical projections are transformed onto a cone wrapped around the globe, which is then flattened, and the cylindrical family of maps is projected onto a cylinder wrapped around the globe. Different projections are selected to minimize specific types of distortions in distance, direction, shape and area found on all flat maps.

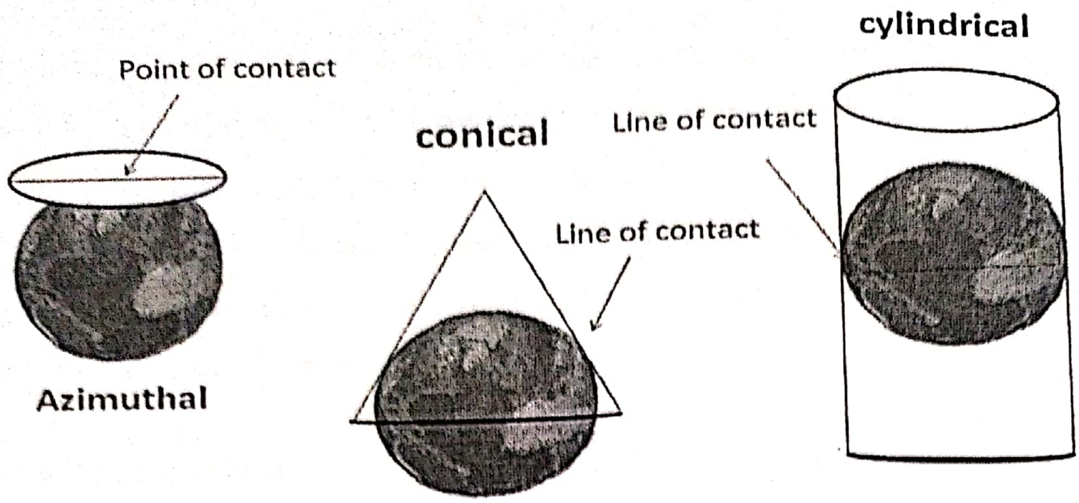


Figure 1.6: Map Projection

GIS developers and users must also concern themselves with map scale. Scale is the mathematical relationship of real earth distance (ground distance) to that same distance as it is shown on a map. This relationship is often stated as a ratio of the two distances. As mentioned above, all maps are models or simplifications of reality. Maps also are reductions of reality. That is, the ratio of map to ground distance is normally much less than one. Maps to be stored in a GIS must be similar in scale if they are to be manipulated together. Maps with widely varying scales cannot be accurately combined. Thus, the user of either traditional paper maps or computerized GIS maps is fundamentally restricted by the degree of scale difference between map manuscripts. Maps with large differences in scale (e.g., 1: 2, 50,000 vs. 1: 9,600) cannot be registered and overlaid without serious distortion and probable error.

1.4.5 Functions of GIS

As stated earlier in the chapter: a GIS is a computerized, integrated system used to compile, store, manipulate, and output mapped data. This section will examine each of these functions.

1.4.5.1 Data Compilation

Data compilation involves assembling all the spatial and attributes data that are to be stored in a computerized format within the GIS. Map data with common projections, scales, and coordinate systems must be pulled together in order to establish the centralized GIS database. Data must also be examined for compatibility in terms of content and time of data collection. Ultimately, the data will be stored in a GIS according to the specific format requirements set by both the user and the chosen GIS software/hardware environment.

When all of the common data requirements are set by the GIS user, a "base map" has been established. A base map is a set of standard requirements for data. It provides accurate standards for geographic control, and also defines a model or template that is used to shape all data into a compatible form. A base map is not necessarily a map; rather, it is a comprehensive set of standards established and enacted to ensure quality control for the spatial and attribute data contained in the GIS.

Once the data are assembled and base map parameters are set; the user must translate the map and attribute data into computer-compatible form. This conversion process referred to as "conversion" or "digitizing" converts paper maps into numerical digits that can be stored in the computer. Digitizing can be performed using various techniques. Scanning is one technique. Another technique of digitization uses a tablet and a tracing stylus (Figure 1.7). Digitizing simplifies map data into sets of points, lines, or cells that can be stored in the GIS computer. Each GIS software package will impose a specific form and design on the way that these sets of points, lines, and cells are stored as digital map files.

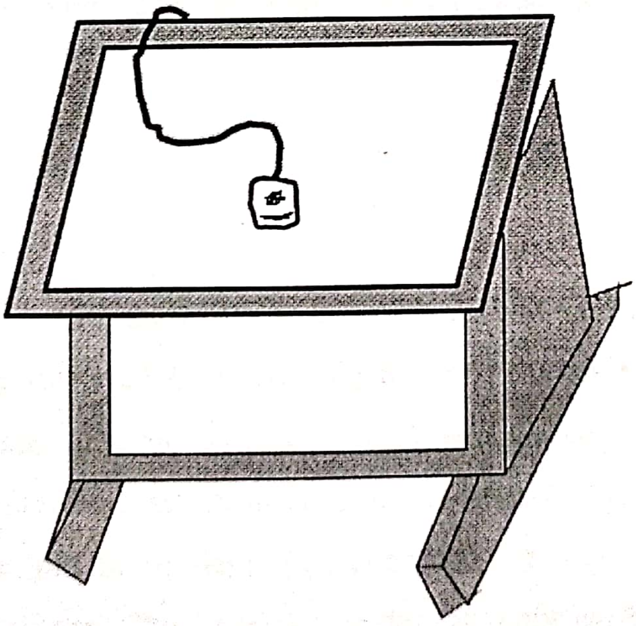


Figure 1.7: Digitizer

Digitization is a simplification process that converts all spatial data to a point (e.g., traffic intersection), a line (e.g., a transport link), a polygon formed by a closed, complex line [e.g., Traffic Analysis Zone (TAZ)], or a grid cell. Digitization reduces all spatial entities to these simple forms because they are easy to store in the computer. A GIS database cannot readily recognize features or entities as human map users do. For example, we cannot enter the entity

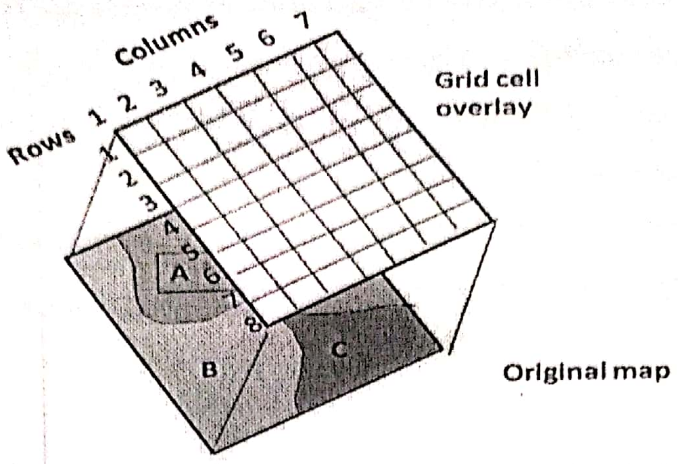
"TAZ" into a GIS. Rather, we enter the spatial data coordinates for the TAZ's boundary as a polygon. Later, the attributes of the TAZ will be entered into the GIS database and will be associated with the polygon. Following the digitization of map features, the user completes the compilation phase by relating all spatial features to their respective attributes, and by cleaning up and correcting errors introduced as a result of the data conversion process. The end result of compilation is a set of digital files each accurately representing all of the spatial and attributes data of interest contained on the original map manuscripts. These digital files contain geographic coordinates for spatial objects (points, lines, polygons, and cells) that represent mapped features.

1.4.5.2 Storage

Once the data have been digitally compiled, digital map files in the GIS are stored on magnetic or other (e.g., optical) digital media. Again, different GIS software packages will employ different storage formats. In most cases, however, data storage will be based on a generic data model that is used to convert map data into a digital form. The two most common types of data models are **raster** and **vector**. Both types are used to simplify the data shown on a map into a more basic form that can be easily and efficiently stored in the computer.

1.4.5.2.1 Raster Image

The raster approach for storing map data in a GIS is perhaps the most intuitive. Figure 1.8 shows the essential step involved in converting a map to a raster format. First, a gridded matrix is registered to and overlaid on the original map manuscript. The row and column coordinates of each cell define location in the grid. To encode the map data for each cell in the raster format, three pieces of data are recorded: the **row coordinate,** the **column coordinate**, and the **attribute**. Thus a triplet of data is recorded for each cell in the array, which is termed a raster. After map data are stored in a raster format, each cell in the raster corresponds to a location on the map and each cell's location in the raster is identified by row and column coordinates. By assigning a value to each cell, the corresponding attribute data for that location are also stored. The end result of this conversion process is a set of cells, each with a specified location and an attribute value. These data can then be entered into a computer-compatible file and stored in the GIS database.



Raster data file

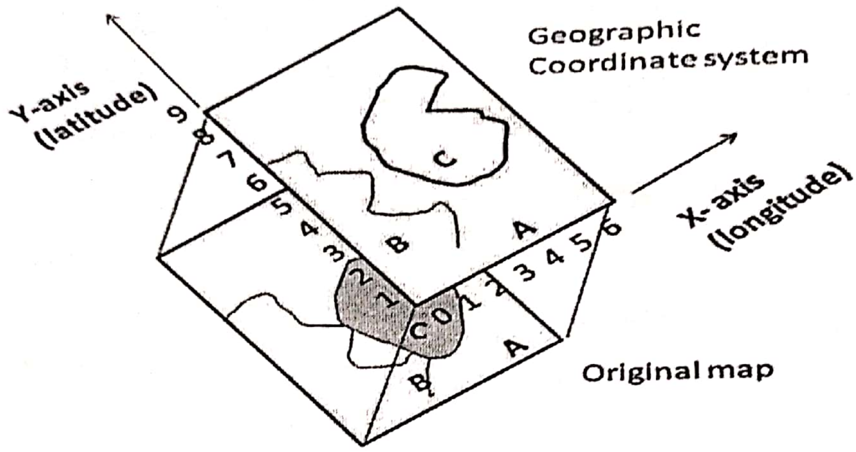
Row	column	attribute
1	1	B
1	2	B
1	3	A

Figure 1.8: Raster Image

Perhaps the most critical issue in using a raster GIS is the selection of an appropriate grid cell size. The user is forced to examine the trade-off between data resolution (how small grid cells are in the raster array) and storage requirements (increasing the number of rows and columns causes' exponential increase in storage requirements). The use of smaller cells records greater detail in the GIS, so the user would normally attempt to select the smallest practical cell size. The choice of cell size depends on many factors, including the resolution of the original map data, the degree of resolution needed in the GIS analysis, the time and money available for data compilation, available storage space on the GIS computer, and cell sizes already employed for previously existing raster data that the user may want to incorporate into the GIS database.

1.4.5.2.2 Vector Image

A vector data structure is very different from a raster data model. Whereas the raster data model uses sets of grid cells to record all data, a vector model stores all spatial data as a point, line, or polygon. These three types of spatial data are referred to as features, and a vector GIS can be termed a "feature-based technology". Figure 1.9 shows an example of a vector data model. When a vector model represents an entity as a point (e.g., a traffic intersection or a bus stop location), a single coordinate pair is used to specify its location. A feature represented as a line (e.g., a road or rail link) uses a linked set of coordinates, and a feature represented as a polygon, which is an alternative form of a line (e.g., a TAZ), must have the same beginning and ending point coordinates. In a raster model, a point is a single cell, a line is a linked set of cells, and a polygon is a group or neighborhood of similarly encoded cells. For all three types of features stored in a vector GIS, an attribute code is entered into the GIS files to identify the object. For example, a TAZ would be vectorized (added to the database as a vector map) as a polygon by storing the linked set of coordinates for the boundary line and the attribute code "TAZ" would be entered in the GIS files to identify that group of cells.

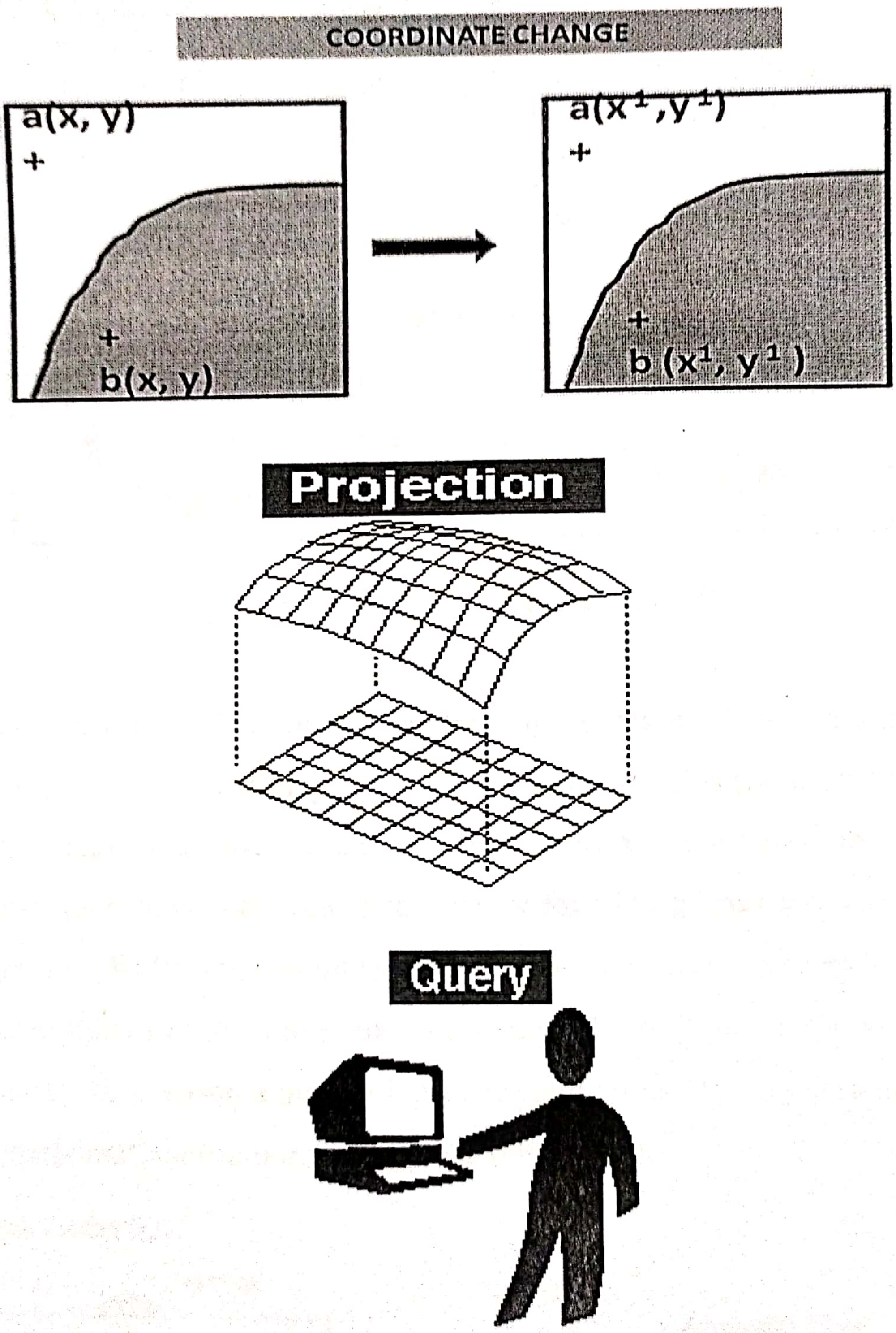


ENITIY	X,Y COORDINATES	NAME
A	5,1	Well
B	0,7;1,7;1,6;2,5;1,4;2,3;3,3;2,1	stream
C	4,3;5,3;6,4;6,5,6,6;7,7;6,8;4,7;3,5;4,3	lake

Figure 1.9: Vector Data

1.4.5.3 Manipulation

Once data are stored in a GIS, many retrieval, analysis, and output options are available to users. These functions are often available in the form of "toolkits." A toolkit is a set of generic functions that a GIS user can employ to manipulate and analyze geographic data. Toolkits provide processing functions such as data retrieval, measuring area and perimeter, overlaying maps, performing map algebra, and reclassifying map data. A GIS usually includes a basic set of computer programs or "tools." The functions provided by the toolkit vary with the software package. Figures 1.10 and Fig 1.11 provides an overview of various tool functions. Data manipulation tools include coordinate change, projection, and edge matching, which allow a GIS to reconcile irregularities between map layers or adjacent map sheets called "tiles." Query and windowing are spatial retrieval tools. Query provides a way to retrieve user-specified data from the database. Windowing allows the user to select a specified area from a map displayed on the monitor to examine it in greater detail as shown in Figure 1.10.



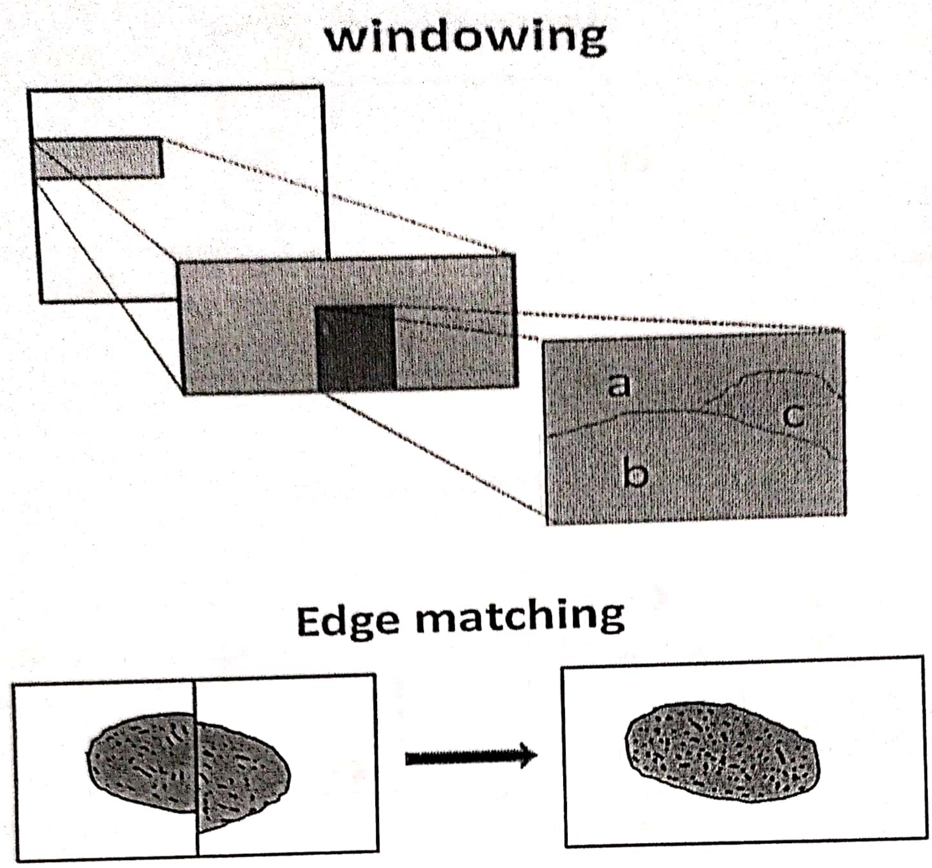
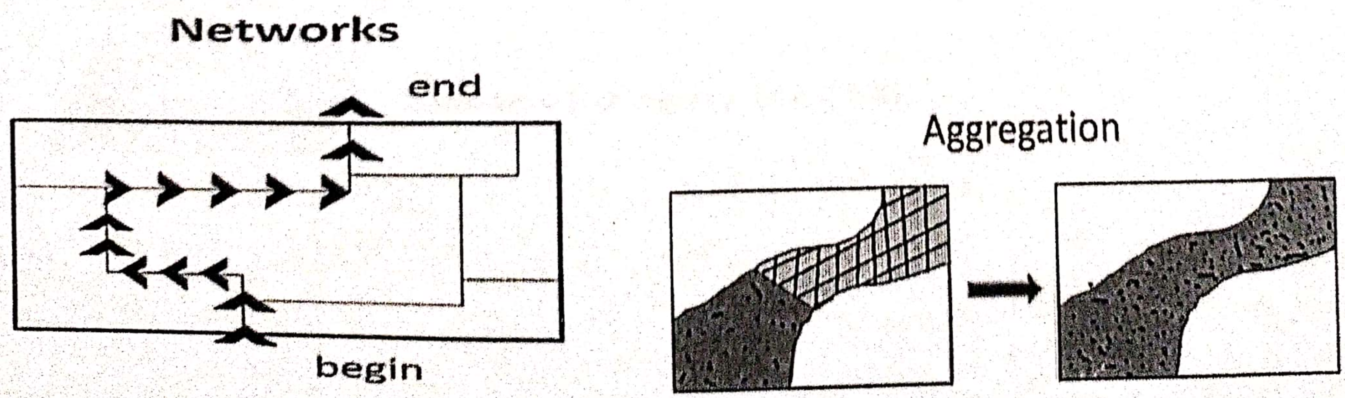
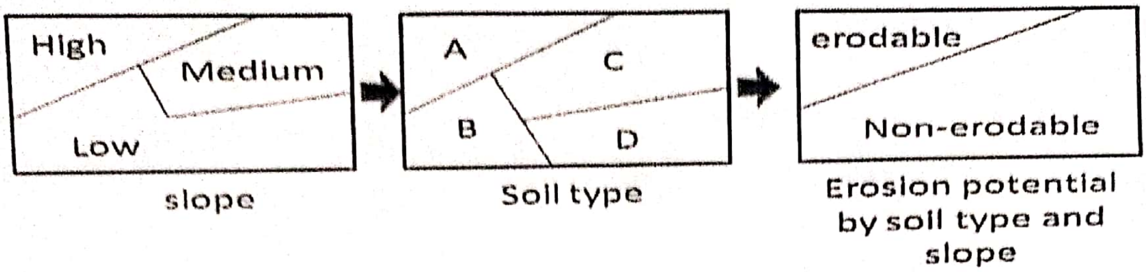


Figure 1.10: Manipulation tools of GIS

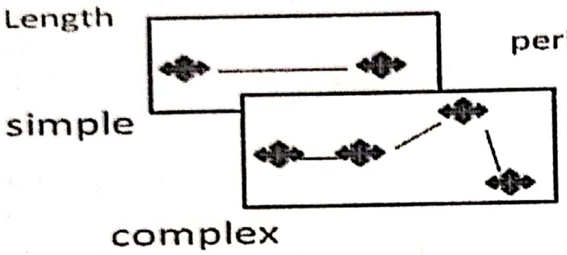
Data analysis tools shown in Figure 1.11 include aggregation, classification, measurement, overlay, buffering, networks, and map algebra. Aggregation helps the user in interpreting the data, classification allows the user to classify areas within a map, and measurement can be used to determine the size of any area. The overlay function allows the user to "stack" map layers on one another. Buffering examines an area that surrounds a feature of interest such as a point. Network functions examine the movement of objects along an interconnected pathway (e.g., traffic flow along a map of highway segments). Map algebra utilities allow the user to specify mathematical relationships between map layers.

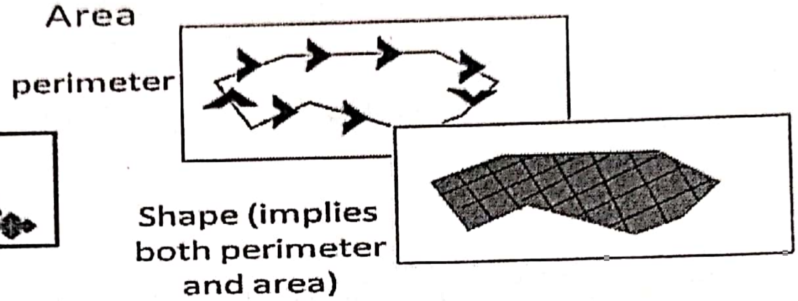


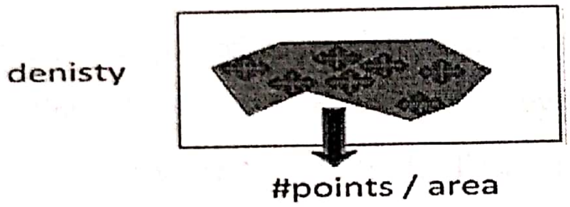
classification



Measurement



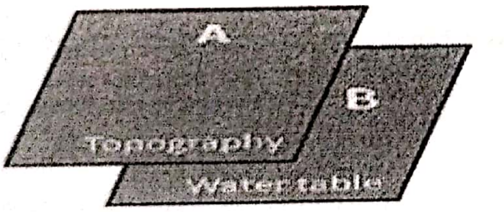




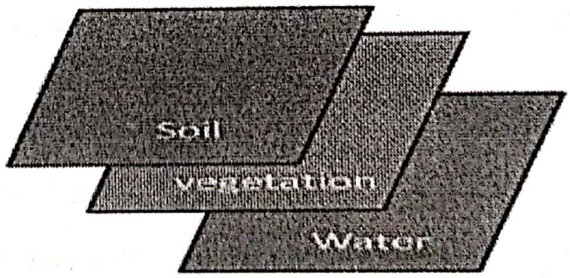
Unit conversion

Feet = meters / 0.3048

Map algebra



A-B = Depth of water



(Soil + vegetation + water) =Habitat

$Slope\Delta = (\Delta Z / \Delta Y) / (\Delta Z / \Delta X)$

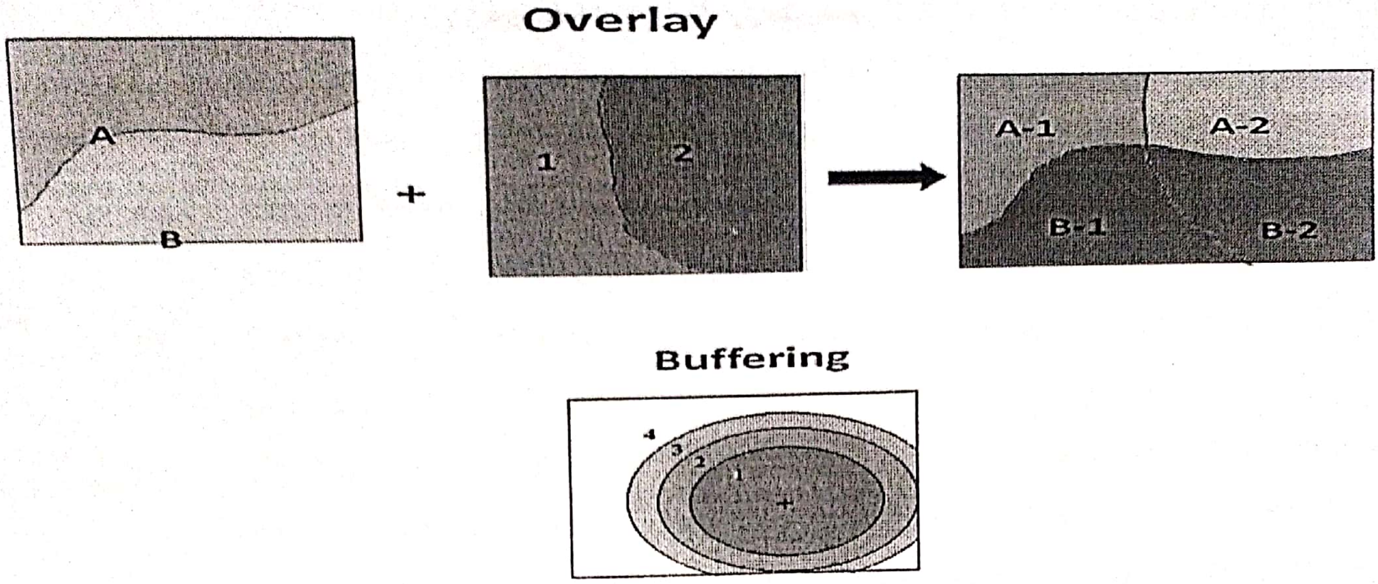


Figure 1.11: Data Analysis Tools of GIS

1.4.5.4 Output

The final functional task of a GIS is to generate output; usually a map. GIS-generated maps are compiled from the many data sets contained in the digital GIS and match exact user specifications. Map output may employ several color and symbology schemes, and will be sized and scaled to meet user needs. These output products resemble hand-drafted maps and fulfill essentially the same purposes. However, it is incorrect to refer to GIS simply as a mapping system. Although GIS is able to generate high-quality map output, its ability to perform analysis and management sets it apart from the more limited computer-mapping packages.

Another form of output from a GIS is tabular or reports information. Data summarized according to user-defined classes or within user-defined areas can readily be generated in a textual format. This output may also be routed to another computer application such as a statistical analysis package or a graphing package for subsequent analysis and display.

1.4.6 Common terms in GIS

The followings are the very frequently used terms in the context of GIS.

i) Map: Usually a two-dimensional representation of all or part of the earth's surface.

ii) Map Projection: The basic system of coordinates used to describe the spatial distribution of elements in a GIS.

iii) Layer: A logical separation of mapped information according to theme. Many geographical information systems and CAD/CAM systems allow user to choose and work on a single layer or any combination of layers at a time.

iv) Mosaic: Process of putting various parts of the study area together, or the result of putting various parts together.

v) Overlay: A record on a transparent medium to be superimposed on another record.

vi) Thematic map: A map displaying selected kinds of information relating to specific themes, such as land use, population, vehicle ownership, and so on.

vii) Topology: The way in which geographical elements are linked together and relate to one another in space. Topological structure essentially maintains the spatial relationships inherent in the terrain data and that is obvious to the human mind when visually examining a graphic representation of data.

viii) Network: In general, this term refers to a set of components connected by channels. In the context of transportation, it normally refers to a network of highways, arterial roads, public transport, rail etc.