UNIT – IV GEOGRAPHIC INFORMATION SYSTEM

Introduction

A **geographic information system** (**GIS**) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data.

The acronym *GIS* is sometimes used for **geographical information science** or **geospatial information studies** to refer to the academic discipline or career of working with geographic information systems and is a large domain within the broader academic discipline of Geoinformatics.[1]

A GIS can be thought of as a system that provides spatial data entry, management, retrieval, analysis, and visualization functions.

The implementation of a GIS is often driven by jurisdictional (such as a city), purpose, or application requirements. Generally, a GIS implementation may be custom-designed for an organization.

Hence, a GIS deployment developed for an application, jurisdiction, enterprise, or purpose may not be necessarily interoperable or compatible with a GIS that has been developed for some other application, jurisdiction, enterprise, or purpose. What goes beyond a GIS is a spatial data infrastructure, a concept that has no such restrictive boundaries.

In a general sense, the term describes any information system that integrates, stores, edits, analyzes, shares, and displays geographic information for informing decision making. GIS applications are tools that allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations.[2][3]

Geographic information science is the science underlying geographic concepts, applications, and systems.[4]

The first known use of the term "Geographic Information System" was by Roger Tomlinson in the year 1968 in his paper "A Geographic Information System for Regional Planning".[5] Tomlinson is also acknowledged as the "father of GIS".

Introduction

Definition of GIS

Like the field of geography, the term Geographic Information System (GIS) is hard to define.

It represents the integration of many subject areas. Accordingly there use no absolutely agreed upon definition of a GIS (deMers, 1997). A broadly accepted definition of GIS is the one provided by the National Centre of Geographic Information and Analysis: a GIS is a system of hardware, software and procedures to facilitate the management, manipulation, analysis, modeling, representation and display of georeferenced data to solve complex problems regarding planning and management of resources (NCGIA, 1990)

Geographic information systems have emerged in the last decade as an essential tool for urban and resource planning and management. Their capacity to store, retrieve, analyse, model and map large areas with huge volumes of spatial data has led to an extraordinary proliferation of applications. Geographic information systems are now used for land use planning, utilities management, ecosystems modeling, landscape assessment and planning, transportation and infrastructure planning, market analysis, visual impact analysis, facilities management, tax assessment, real estate analysis and many other applications.

Functions of GIS include: data entry, data display, data management, information retrieval and analysis. A more comprehensive and easy way to define GIS is the one that looks at the disposition, in layers (Figure 1), of its data sets. "Group of maps of the same portion of the territory, where a given location has the same coordinates in all the maps included in the system". This way, it is possible to analyse its thematic and spatial characteristics to obtain a better knowledge of this zone.

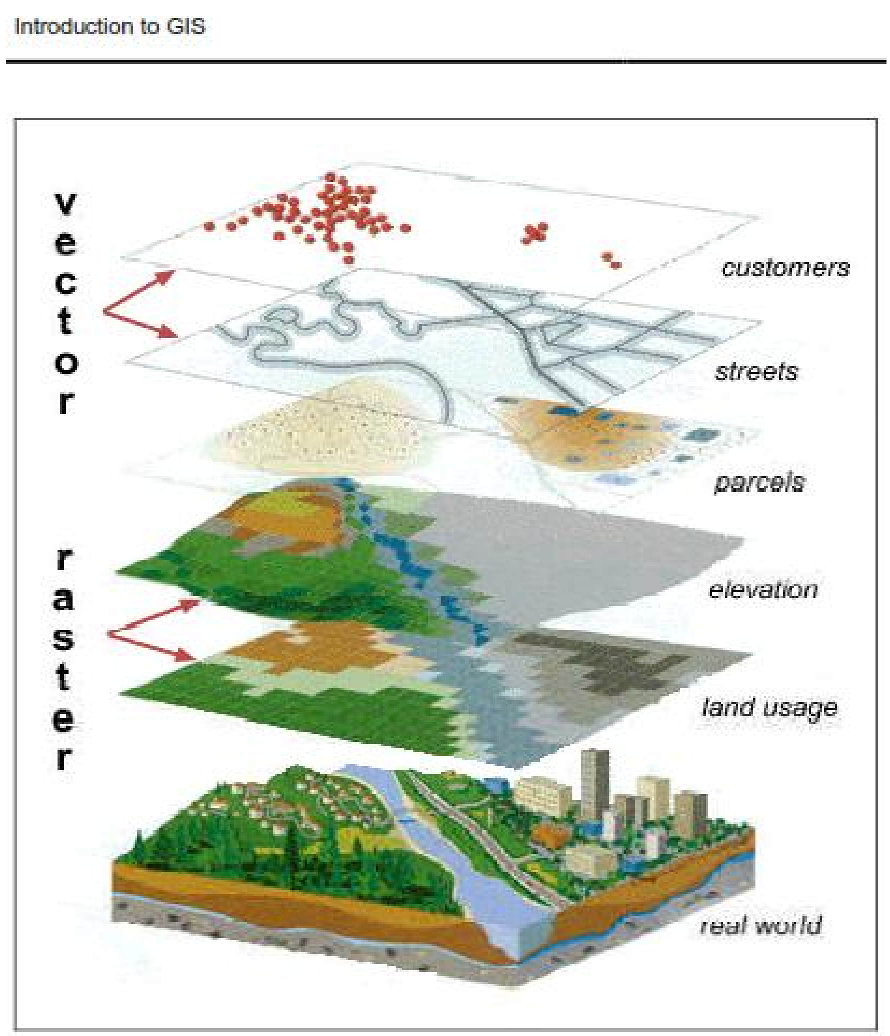


Figure. 1. The concept of layers (ESRI)

GIS applications

Mapping locations:

GIS can be used to map locations. GIS allows the creation of maps through automated mapping, data capture, and surveying analysis tools. **mapping quantities:** People map quantities, like where the most and least are, to find places that meet their criteria and take action, or to see the relationships between places. This gives an additional level of information beyond simply mapping the locations of features.

Mapping densities:

While you can see concentrations by simply mapping the locations of features, in areas with many features it may be difficult to see which areas have a higher concentration than others. A density map lets you measure the number of features using a uniform areal unit, such as acres or square miles, so you can clearly see the distribution.

Finding distances:

GIS can be used to find out what's occurring within a set distance of a feature.

Mapping and monitoring change:

GIS can be used to map the change in an area to anticipate future conditions, decide on a course of action, or to evaluate the results of an action or policy.

MAPS

Traditionally, the role of maps in the census has been to support enumeration and to present aggregate census results in cartographic form.

Cartographic automation has greatly expanded this role.

In addition to enabling more efficient production of enumerator maps and thematic maps of census results, GIS now plays a key role in census data dissemination and in the analysis of population and household data.

A map is a graphic representation (picture) of all or part of the Earth's surface (or another spatial object).

A map is usually a two-dimensional drawing of part of the three-dimensional world in which we live.

As a result, any map will always need to be changed (distorted) in some way to squeeze the three dimensions into two (more about this in the section on Map Projections).

A map maker (cartographer) uses many graphic communications methods to create a map including symbols, text, and color.

Maps are used to portray information in a geographic format so that individual features (streets, streams, houses, etc.) can be understood by their position in relation to other features.

DEFINITION:

A **map** is a **representation** of all or part of the Earth drawn on a **flat surface** at a specific scale.
Maps were devised because they are much easier to use, store, and transport than globes, and they facilitated the development of much larger scaled representations than was the case with a globe.

1.3. Mapping has been an integral part of census taking for a long time. Very few enumerations during the last several census rounds were executed without the help of detailed maps. In general terms, digital mapping serves several purposes in the census process:

Maps ensure consistency and facilitate census operations (pre-enumeration).
The census office needs to ensure that every household and person in the country is counted, and that no households or individuals are counted twice. For this purpose, census geographers partition the national territory into small reporting units. Maps thus provide an essential control device that guarantees consistency and accuracy of the census.

Maps support data collection and can help monitor census activities (during enumeration).
During the census, maps ensure that enumerators can easily identify their assigned set of households.
Maps are also issued to census supervisors to support planning and control tasks. Maps can thus also play a role in monitoring the progress of census operations. This allows supervisors to identify problem areas and implement remedial action quickly.

Maps make it easier to present, analyse and disseminate census results (post-enumeration).
Cartographic presentation of census results provides a powerful means for visualizing the results of a census. This supports the identification of local patterns of important demographic and social indicators. Maps are thus an integral part of policy analysis in the public and private sectors.
.

All maps have the following characteristics:

1. The location of features. The accurate location of features on a map involves:

- Accuracy (See "Map Accuracy")
- Resolution
- Scale (See "Map Scale")

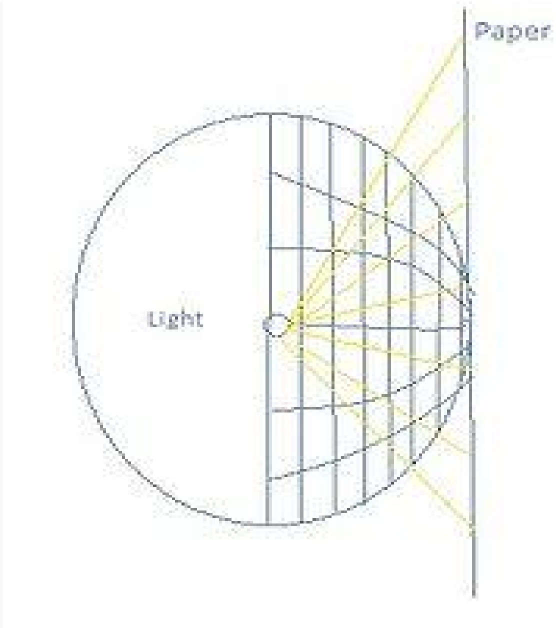
2. The information about (attributes of) each feature.

3. The transformation of a three-dimensional (globe) surface onto a two-dimensional plane (the map).

4. The portrayal of map information through symbols.

5. The abstraction of reality.

Map projection



Simple Sketch to illustrate the projection concept.

Map projection

A map **projection** is a method used to transfer the features of a globe, such as the lines oflatitude and longitude and the outlines of continents, onto the flat surface of a map. This wasoriginally done with the use of a light to project the shadow of a wire-skeleton globe onto a flat surface—hence the term "projection." The three major types of projections developed from thismethod are the **cylindrical**, **planar**, and **conic**.

In modern **cartography** (map making), the projection method has been largely replaced bymathematical constructions and computer-assisted programs; however, many of these projectionsstill have one of the three forms (cylindrical, planar, and conic) as their basis. Other projectionsare derived using non-perspective shapes and from mathematical calculations

A **map projection** is any orderly system of parallels and meridians on which a map can be drawn.[1]

These parallels and meridians help take real world locations and spatial references and put them in a flat plane.

Any mathematical function transforming coordinates from a curved surface to a plane is considered a projection.

Since the Earth is roughly the shape of an oblate spheroid, map projections are necessary for creating maps of the Earth or parts of the Earth that are represented on a plane such as a piece of paper or a computer screen.

To get a simple idea of how projections are created, imagine a light bulb placed at the center of a transparent globe, then placing a piece of paper around the globe turn on the light and trace what is "projected" onto the paper. The end result is a simple map projection.

In their attempt at planar representation of actual map features such as an island or continent on the curved surface of the earth, all map projections necessarily distort some aspects of these features.

Depending on the purpose of the map, some distortions are acceptable and others are not; therefore different map projections exist in order to preserve some properties of the spheroid-based features at the expense of other properties.

A map projection is an essential component of any modern map, and there are an infinite number of possible map projections. Since Gerardus Mercator presented his Mercator global map projection in 1569, numerous map projections have been developed and scores of projections are currently used by cartographers today.

Map projections are often named after the cartographer who developed them, after the method used in the projection, or a combination of both. All maps, such as wall maps and those in atlases, specify the type of projection they are based on. Well-known projections for world maps are the **Mercator**, **Peters,** and **Robinson** projections.

Mercator projection:

This projection, developed by Gerard us Mercator in 1569, is a cylindrical projection on which both the lines of latitude and lines of longitude appear as straight lines running parallel and perpendicular to each other.

Although the shapes of features it represents are accurate, the distances and areas are greatly distorted, particularly in higher latitudes.

This results in reasonably accurate area representation in equatorial regions, but greatly exaggerated areas in higher latitudes.

The common use of this projection in classrooms has been criticized by leaders of tropical countries, as they feel it unfairly represents their countries as very small in comparison to mid- and high-latitude countries.

This problem of perception was particularly troubling as many of the mid- and high-latitude countries (Britain and France, for example) were also world powers and colonizers of poorer tropical countries, and their exaggerated size on the map further emphasized their dominance in this relationship.

Peters projection:

This projection was developed by Dr. Arno Peters, a German historian and journalist, in 1973. His major aim in developing this projection was to create a map on which the sizes of countries were accurately represented unlike on the Mercator projection.

A projection showing accurate areas is called an equal-area map.

The Earth grid is similar to that on the Mercator projection, however, the spaces between the lines of latitude do not increase pole ward, thus reducing area distortion.

This characteristic appealed to organizations such as the United Nations Development Programmer and the National Council of Churches that are involved in development education and aid to the developing world.

These organizations believed that the Peters projection provides a more accurate perception of, and thus perhaps greater support for, the less-developed countries of the world.

Although this projection might satisfy those who dislike the Mercator projection, a major criticism of the Peters projection is that the shapes of continents are highly distorted, many of them appearing longer and narrower than their real shapes.

Furthermore, Peters was not a trained cartographer and, as a result, his projection did not get serious consideration by scientific-minded cartographers of his time.

The Peters projection, also known as the GallPeters projection,is seldom used in modern cartography.

Robinson projection:

This projection, developed by Professor Arthur E. Robinson in 1963, attempts to create a visually appealing (or right-appearing) view of the entire world.

Rather than trying to eliminate any single type of distortion, this is a compromise projection that attempts to keep all types of distortion to a minimum throughout the map area.

In this projection, the world appears somewhat like an oval; however, the poles appear as lines rather than as points.

The lines of latitude are straight and parallel and the lines of longitude are curved; however, they do not converge to a point.

The development of the Robinson projection is unique in that it was a response to a request by a map and atlas production company (Rand McNally) to develop a better visual representation of the world for use in its publications.

Not only is this projection still used extensively by Rand McNally, it has also been adopted by the National Geographic Society for use in many of the maps featured in its magazine and map products.

Types of Map Projections

There are four basic types of map projections:

1. **Conformal**

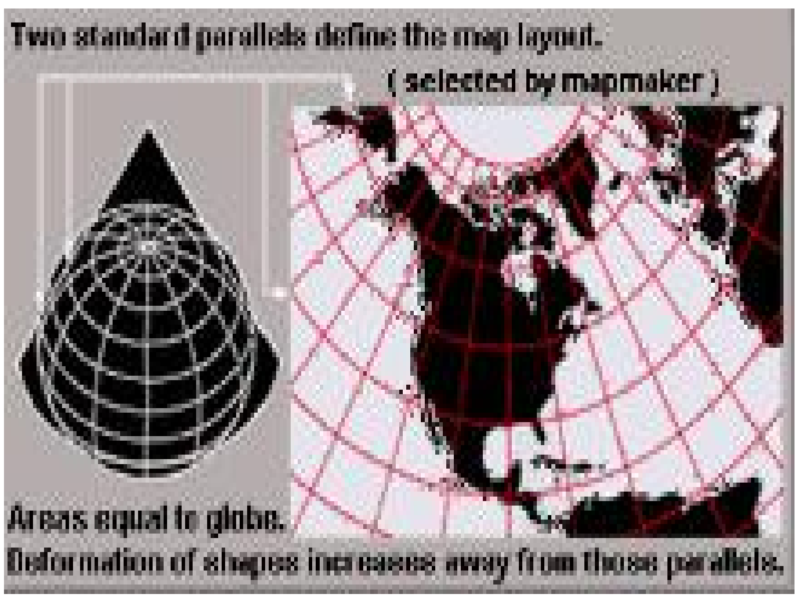
Preserves: Shape
Distorts: Area

2. **Equal Area**
   Preserves: Area
   Distorts: Shape, Scale or Angle (bearing)

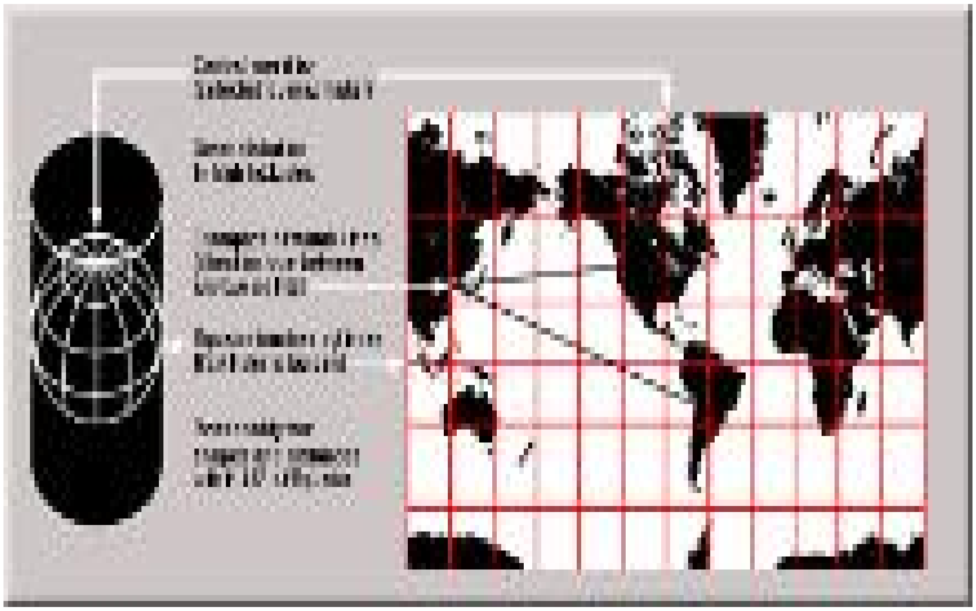
3. **Equidistant**
   Preserves: Distances between certain points (but not all points)
   Distorts: Other distances

4. **Azimuthal** (True Direction)
   Preserves: Angles (bearings)
   Distorts: Area and shape

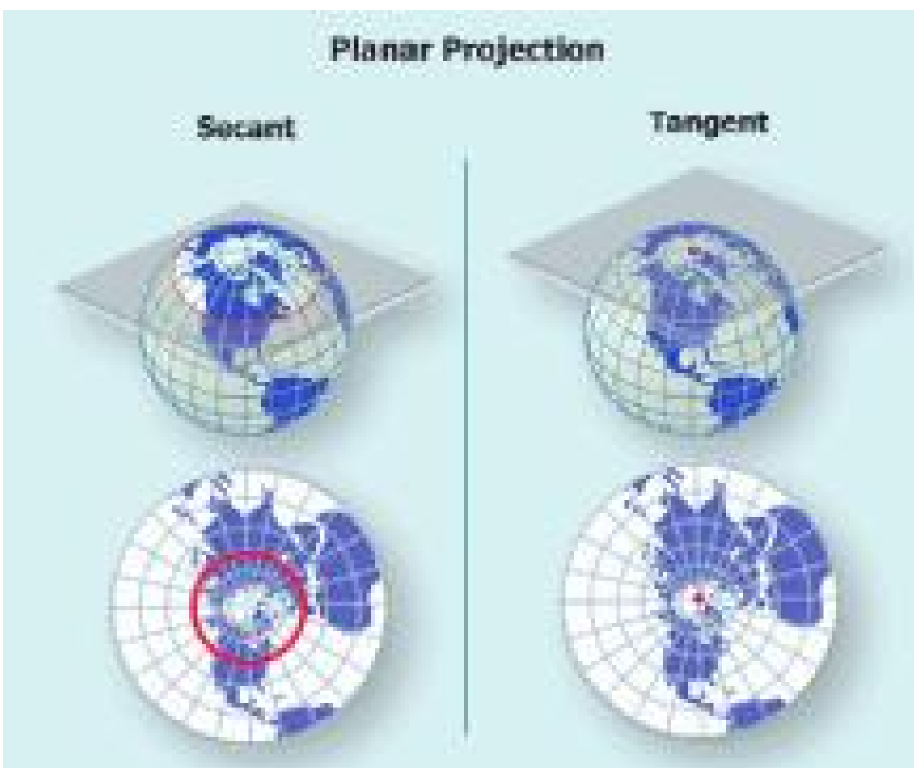
In addition, there are three basic classifications of **projection surfaces**:



A **conic** projection, as its name implies, is based on a cone which either touches the surface of the globe in a circle, or intersects the globe in two circles, passing through it.



A **cylindrical** projection is based on a flat surface, curved to make a cylinder into which the globe is placed so that the globe intersects the cylinder in either one circle (where the cylinder touches the outside of the globe, or two circles (where the cylinder intersects and passes through the globe).



A **planar** projection transforms all or part of the spherical surface onto a plane based either on one point of contact with the globe, or a circle of contact with the globe where the plane passes through the globe.

Many common map projections are classified according to the projection surface used: conic, cylindrical, and planar.

Conic projections

The simplest conic projection is tangent to the globe along a line of latitude. This line is called the ***standard parallel*.**

The meridians are projected onto the conical surface, meeting at the apex, or point, of the cone. Parallel lines of latitude are projected onto the cone as rings.

The cone is then 'cut' along any meridian to produce the final conic projection, whichhas straight converging lines for meridians and concentric circular arcs for parallels. The meridian opposite the cut line becomes the ***central meridian*.**

In general, distortion increases away from the standard parallel. Thus, cutting off the top of the cone produces a more accurate projection.

This is accomplished by not using the polar region of the projected data.

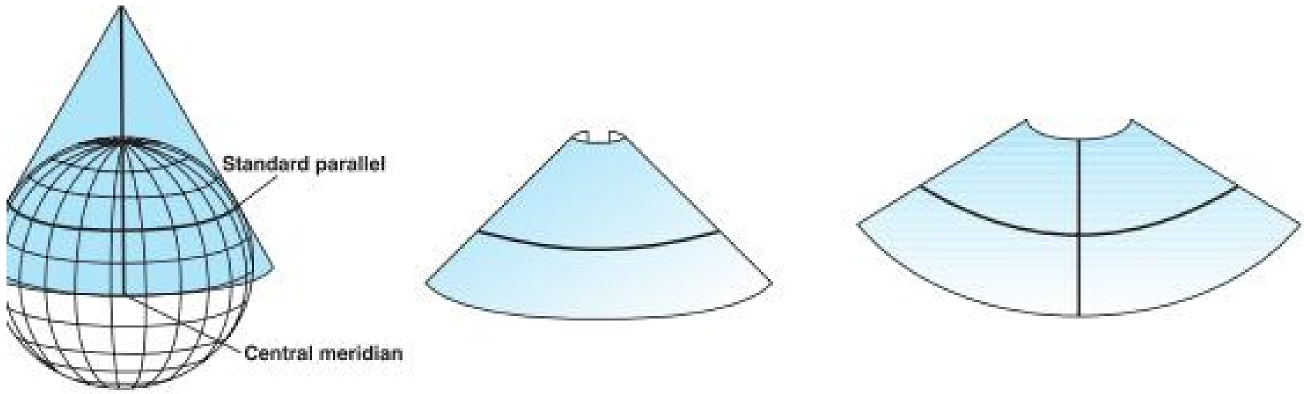
Conic projections are used for multitude zones that have an east-to-west orientation. Somewhat more complex conic projections contact the global surface at two locations.

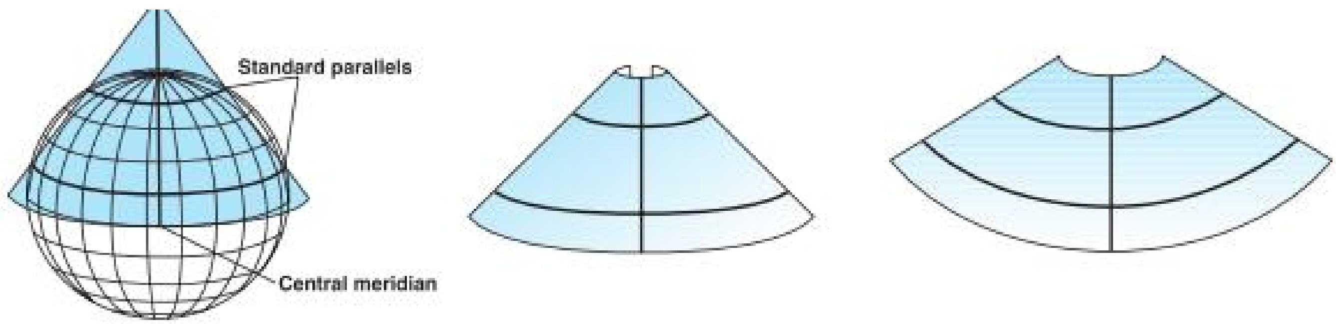
These projections are called *secant* conic projections and are defined by two standard parallels.

It is also possible to define a secant projection by one standard parallel and a scale factor. The distortion pattern for secant projections is different between the standard parallels than beyond them.

Generally, a secant projection has less overall distortion than a tangent case. On still more complex conic projections, the axis of the cone does not line up with the polar axis of the globe. These are called ***oblique*.**

The representation of geographic features depends on the spacing of the parallels. When equally spaced, the projection is equidistant in the north–south direction but neither conformal nor equal area such as the Equidistant Conic projection. For small areas, the overall distortion is minimal. On the Lambert Conic Conformal projection, the central parallels are spaced more closely than the parallels near the border, and small geographic shapes are maintained for both small-scale and large-scale maps. Finally, on the Albers Equal Area Conic projection, the parallels near the northern and southern edges are closer together than the central parallels, and the projection displays equivalent areas.





Cylindrical projections

Cylindrical projections can also have tangent or secant cases. The Mercator projection is one of the most common cylindrical projections, and the equator is usually its line of tangency.

Meridians are geometrically projected onto the cylindrical surface, and parallels are mathematically projected, producing graticular angles of 90 degrees.

The cylinder is 'cut' along any meridian to produce the final cylindrical projection.

The meridians are equally spaced, while the spacing between parallel lines of latitude increases toward the poles. This projection is conformal and displays true direction along straight lines.

Rhumb lines, lines of constant bearing, but not most great circles, are straight lines on a Mercator projection.

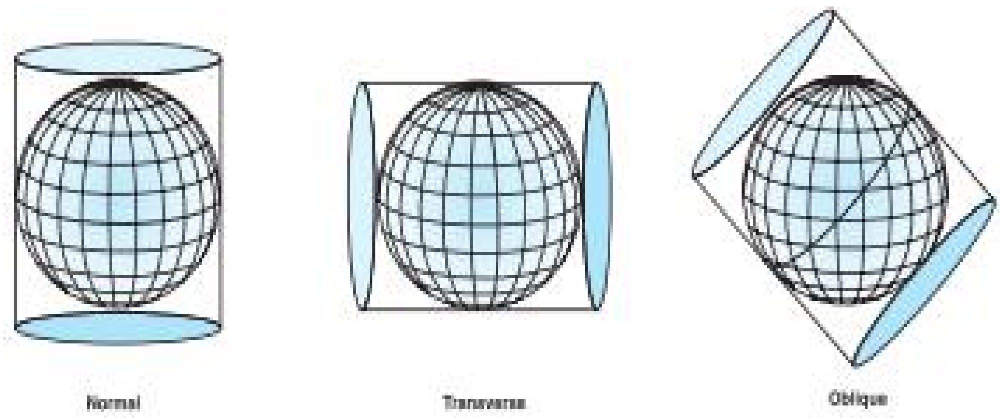
For more complex cylindrical projections the cylinder is rotated, thus changing the tangent or secant lines.

Transverse cylindrical projections such as the Transverse Mercator use a meridian as the tangential contact or lines parallel to meridians as lines of séance.

The standard lines then run north and south, along which the scale is true. Oblique cylinders are rotated around a great circle line located anywhere between the equator and the meridians.

In these more complex projections, most meridians and lines of latitude are no longer straight of equidistance.

Other geographical properties vary according to the specific projection.



In all cylindrical projections, the line of tangency or lines of secancy have no distortion and thus are lines

Planar projections

Planar projections project map data onto a flat surface touching the globe. A planar projection is also known as an azimuthal projection or a zenithal projection. This type of projection is usually tangent to the globe at one point but may be secant.

The point of contact may be the North Pole, the South Pole, a point on the equator, or any point in between. This point specifies the *aspect* and is the focus of the projection. The focus is identified by a central longitude and central latitude. Possible aspects are *polar*, *equatorial*, and *oblique*.

Projections. Perspective points may be the center of the earth, a surface point directly opposite from the focus, or a point external to the globe, as if seen from a satellite or another planet.

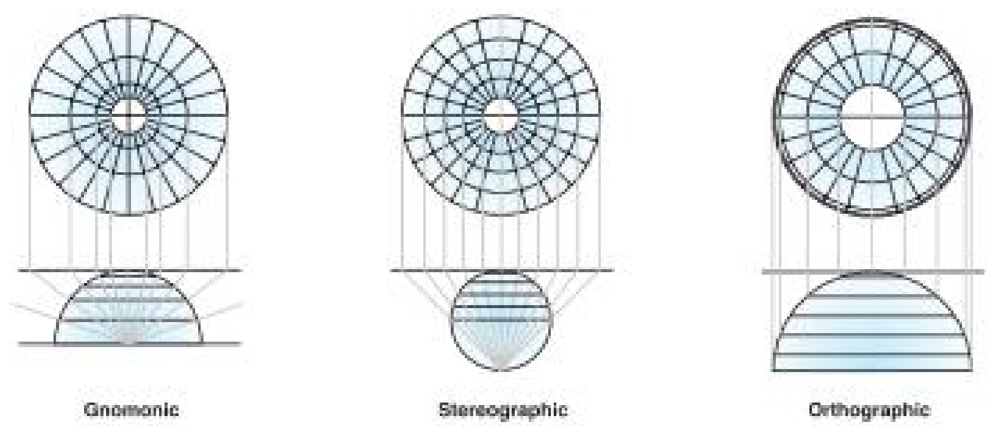
Polar aspects are the simplest form. Parallels of latitude are concentric circles centered on the pole.

Polar aspects are the simplest form. Parallels of latitude are concentric circles centered on the pole, and meridians are straight lines that intersect at the pole with their true angles of orientation. In other aspects, planar projections will have graticular angles of 90 degrees at the focus. Directions from the focus are accurate.

Great circles passing through the focus are represented by straight lines; thus the shortest distance from the center to any other point on the map is a straight line. Patterns of area and shape distortion are circular about the focus. For this reason, azimuthal projections accommodate circular regions better than rectangular regions. Planar projections are used most often to map polar regions.

Some planar projections view surface data from a specific point in space. The point of view determines how the spherical data is projected onto the flat surface. The perspective from which all locations are viewed varies between the different azimuthal

Azimuthal projections are classified in part by the focus and, if applicable, by the perspective point. The graphic below compares three planar projections with polar aspects but different perspectives. The Gnomonic projection views the surface data from the center of the earth, whereas the Stereographic projection views it from pole to pole. The Orthographic projection views the earth from an infinite point, as if viewed from deep space.
Note how the differences in perspective determine the amount of distortion toward the equator.



MAP ANALYSIS

●**ANALYSIS**

The heart of GIS is the analytical capabilities of the system. What distinguish the GIS system from other information system are its spatial analysis functions. Although the data input is, in general, the most time consuming part, it is for data analysis that GIS is used.

The analysis functions use the spatial and non-spatial attributes in the database to answer questions about the real world. Geographic analysis facilitates the study of real-world processes by developing and applying models.

Such models illuminate the underlying trends in geographic data and thus make new information available. Results of geographic analysis can be communicated with the help of maps, or both.

The organization of database into map layers is not simply for reasons of organizational clarity, rather it is to provide rapid access to data elements required for geographic analysis.

The objective of geographic analysis is to transform data into useful information to satisfy the requirements or objectives of decision-makers at all levels in terms of detail.

An important use of the analysis is the possibility of predicting events in another location or at another point in time.

●**ANALYSIS -How?**

Before commencing geographic analysis, one needs to assess the problem and establish an objective. The analysis requires step-by-step procedures to arrive at the conclusions.

The range of geographical analysis procedures can be subdivided into the following categories

Use of Spatial Analysis:

It helps us to:

- **Identify trends on the data**.
- **Create new relationships from the data**.
- **View complex relationships between data sets**.
- **Make better decisions**.

GEOGRAPHIC ANALYSIS

It is the analysis of problems with some Geographic Aspects.

- **Alternatives are geographic locations or areas**.
- **Decisions would affect locations or areas**.
- **Geographic relationships are important in decision-making or modeling**.

Relationship of Modeling to Analysis:

- Decision Models search through potential alternatives to arrive at a recommendation.
- Decision support models process raw data into forms that are directly relevant to decision making.
- Data characterization models are used to develop a better understanding of a system to help characterize a problem or potential solutions.

NETWORK ANALYSIS

Network models are based on interconnecting logical components, of which the most important are:

o "Nodes" define start, end, and intersections.
o "Chains" are line features joining nodes.
o "Links" join together points making up a chain.

This network can be analyzed using GIS.

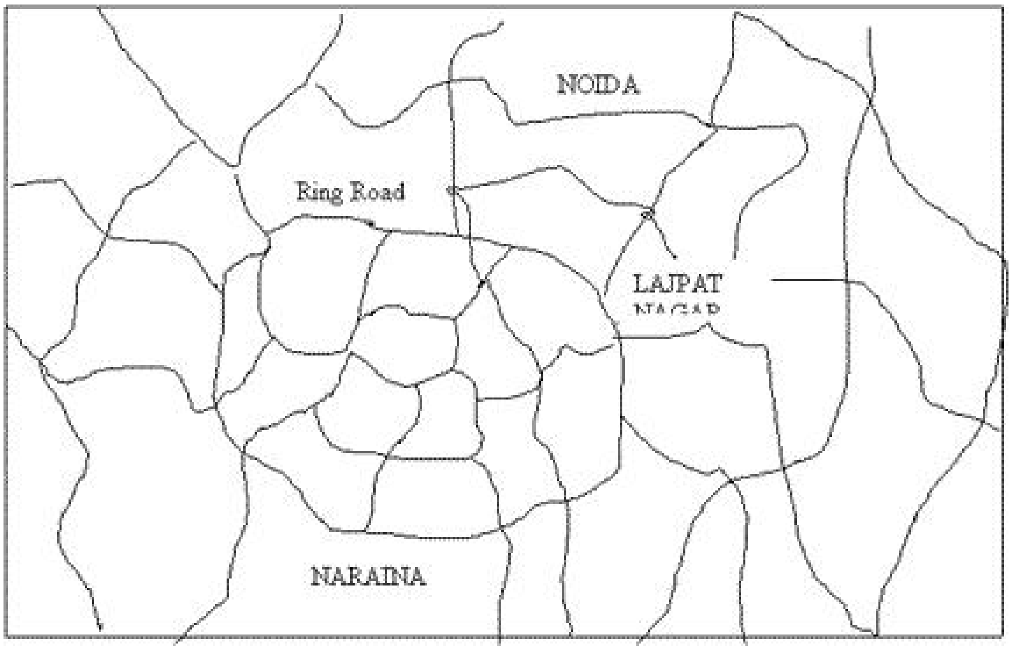
A simple and most apparent network analysis applications are:

o Street network analysis.
o Traffic flow modeling.
o Telephone cable networking.
o Pipelines etc.

Basic forms of network analysis simply extract information from a network. More complex analysis, process information in the network model to derive new information.

One example of this is the classic shortest-path between two points.

The vector model is more suited to network analysis than the raster model.



A ROAD NETWORK

GIS

GEOGRAPHICAL INFORMATION SYSTEM

Meaning

The expansion of GIS is Geographic Information System which consists of three words, viz. Geographic, Information and System.

Here the word 'Geographic' deals with spatial objects or features which can be referenced or related to a specific location on the earth surface.

The object may be physical / natural or may be cultural / man made. Likewise the word 'Information' deals with the large volume of data about a particular object on the earth surface. The data includes a set of qualitative and quantitative aspects which the real world objects acquire.

The term 'System' is used to represent systems approach where the complex environment (consists of a large number, of objects / features on the earth surface and their complex characteristics) is broken down into their component parts for easy understanding and handling, but is considered to form an integrated whole for managing and decision making.

Now-a-days this is possible in a very short span of time with the development of sophisticated computer hardware and software.

Therefore, GIS is a computer based information system which attaches a variety of qualities and characteristics to geographical location (Fig.5)and helps in planning and decision making.

A Geographic Information System (GIS) may be defined in different manners. International Training Centre (ITC), Holland defined Geographic Information System (GIS) as a computerized system that facilitates the phases of data entry, data analysis and data presentation especially in cases when we are dealing with geo referenced data.

A Geographic Information System (GIS) may be defined in different manners. International Training Centre (ITC), Holland defined Geographic Information System (GIS) as a computerized system that facilitates the phases of data entry, data analysis and data presentation especially in cases when we are dealing with geo referenced data.

Indian Society of Geomatics (ISG) and Indian Space Application Centre (ISRO) defined GIS as a system which provides a computerised mechanism for integrating various geoinformation data sets and analysing them in order to generate information relevant to planning needs in a context.

According to Centre for Spatial Database Management and Solutions (CSDMS), GIS is a computer based tool for mapping and analysing things that exist and events that happen one earth.

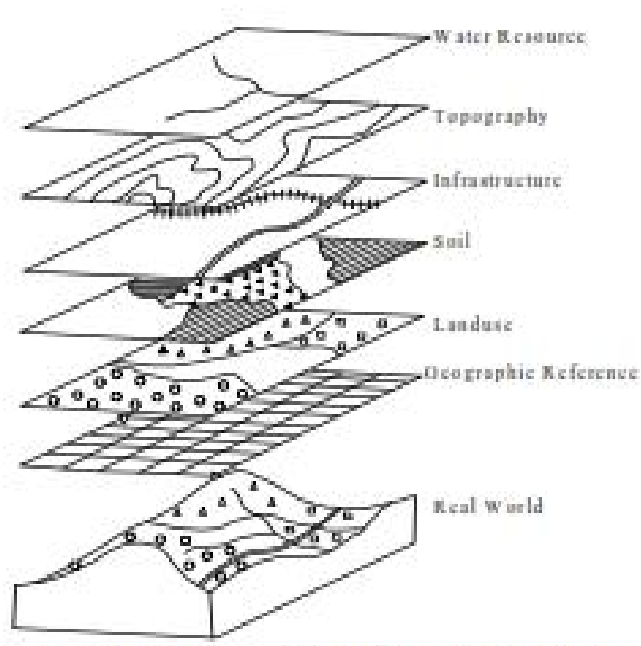
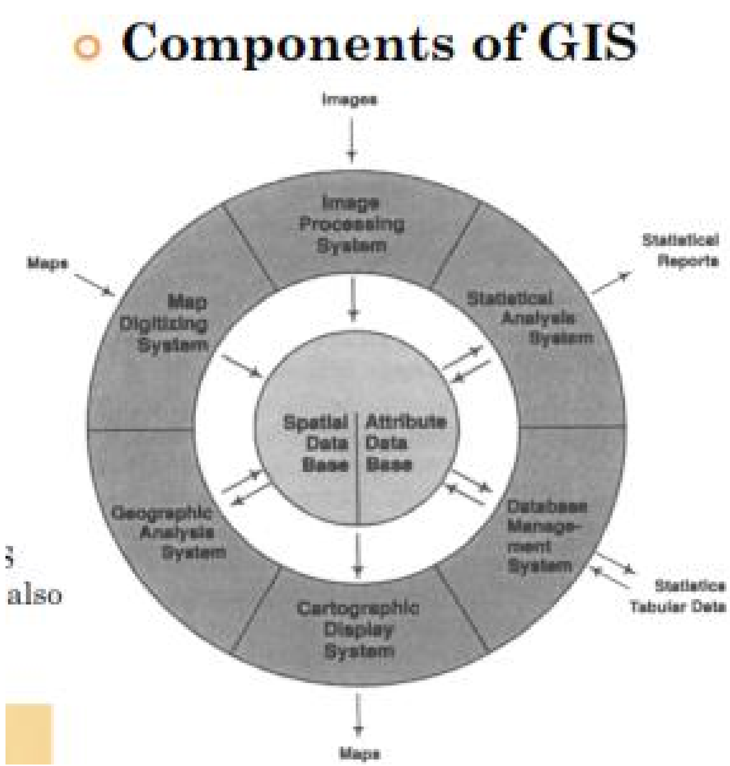


Fig.5 : Informaton sored as theme layers with each layer linke to a common spatial featurereferenced to a specific location

Components of a GIS

A working GIS integrates these five key components: hardware, software, data, people, and methods.



Hardware

Hardware is the computer on which a GIS operates. Today, GIS runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are

- a database management system (DBMS)
- tools for the input and manipulation of geographic information
- tools that support geographic query, analysis, and visualization
- a graphical user interface (GUI) for easy access to tools

Data

Maybe the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or bought from a commercial data provider. Most GISs employ a DBMS to create and maintain a database to help organize and manage data.

People

GIS technology is of limited value without the people who manage the system and to develop plans for applying it. GIS users range from technical specialists who design and maintain the system to those who use it to help them do their everyday work.

Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

GIS software

It encompasses not only to the GIS package, but all the software used for databases, drawings, statistics, and imaging.

The functionality of the software used to manage the GIS determines the type of problems that the GIS may be used to solve. The software used *must* match the *needs* and *skills* of the end user.

\ **Popular GIS Software**

Vector-based GIS
- ArcGIS (ESRI)
- ArcView
- MapInfo

Raster-based GIS
- Erdas Imagine (Leica
- ENVI (RSI)
- ILWIS (ITC)
- IDRISI (Clark Univ.)

Data type for GIS

A GIS without data is like a car without fuel. Without fuel, a car cannot move, likewise without data a GIS will not produce anything.

Data for GIS can be obtained from different sources like aerial photographs, satellite imageries, digital data, conventional maps, Census, Meteorological department, field data (surveys/GPS) etc.

These data obtained from various sources can be classified into two types – spatial data which describes location and attribute data which specifies the characteristics at that location. Spatial data tells us, "where the object is?" Attribute data tells us "What the object is?" or "How much the object is?" In other words, it tells the characteristics at that location.

Spatial Data

The spatial data or real world features are very complex. So, spatial data is simplified before they are entered into the computer. The common way of doing this is to break down all geographic features into three basic entity types – points, lines and areas.

Points are 'one dimensional' objects, used to represent features that are very small, e.g. a post box, an electric pole, a well or tube well etc. Only latitudinal and longitudinal values or a coordinate reference can be given to these features to explain their location. Lines are two dimensional objects and are used to represent linear features, for example roads and rivers. Lines are also used to represent linear features that do not exist in reality, such as administrative boundaries and international boundaries.

Areas are three dimensional objects and are represented by closed set of lines and are used to define features such as agricultural fields, forest areas, administrative areas etc. Area entities are often referred to as polygons.

The representation of real world features using the point, line and area entity types appears relatively simple.

However, the appropriate entity to represent real world features is often difficult and it depends upon the scale of the map. On a world map, cities are represented by points. It only gives information about number of cities shown on the world map. At national or regional scale, the 'point' entity to represent cities is considered too simple, as it tells us nothing about the real size of the city.

In this case, cities are represented by 'area' entity. At the local scale, 'area' entity to represent cities would be considered too simple. In this case, cities are represented by mixture of 'point', 'lines' and 'areas' as entities.

Points may be used to represent features such as electric poles, post boxes etc. Likewise lines and areas may be used to represent road networks and residential blocks respectively. So, the decision makers decide the 'entities' through which different features of real world would be represented.

Attribute Data

As it is mentioned earlier, attribute data tells the characteristics of different objects / features on the earth surface. These are descriptions, measurements or classification of geographic features.

Attribute data can be both qualitative (like land use type, soil type, name of the city/river etc.) and quantitative (like elevation, temperature, pressure of a particular place, crop yield per acre etc.).

So, the attribute can be both numeric and textual. The examples of attribute data of different spatial features like point (well), line (river), area (village) are shown in box 1. The attribute data are generally in tabular form.

DATA TYPE:

Vector

In the vector data model, features on the earth are represented as:

Points

Lines

Polygons

Raster

In the raster data model, ageographic feature like land cover is represented as:

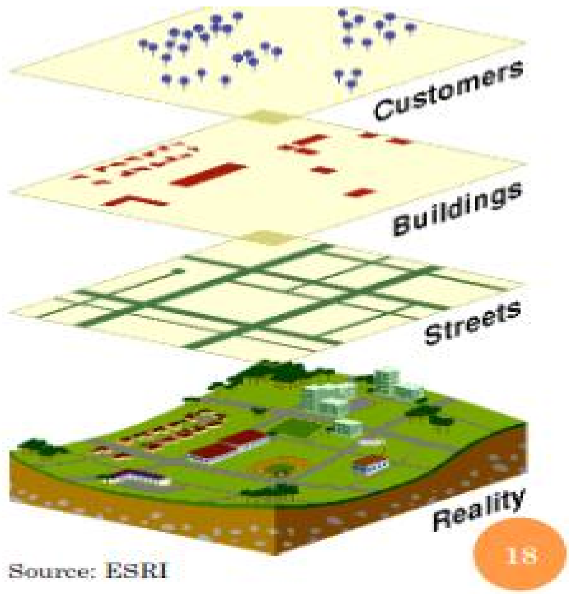
single square cells

Attribute

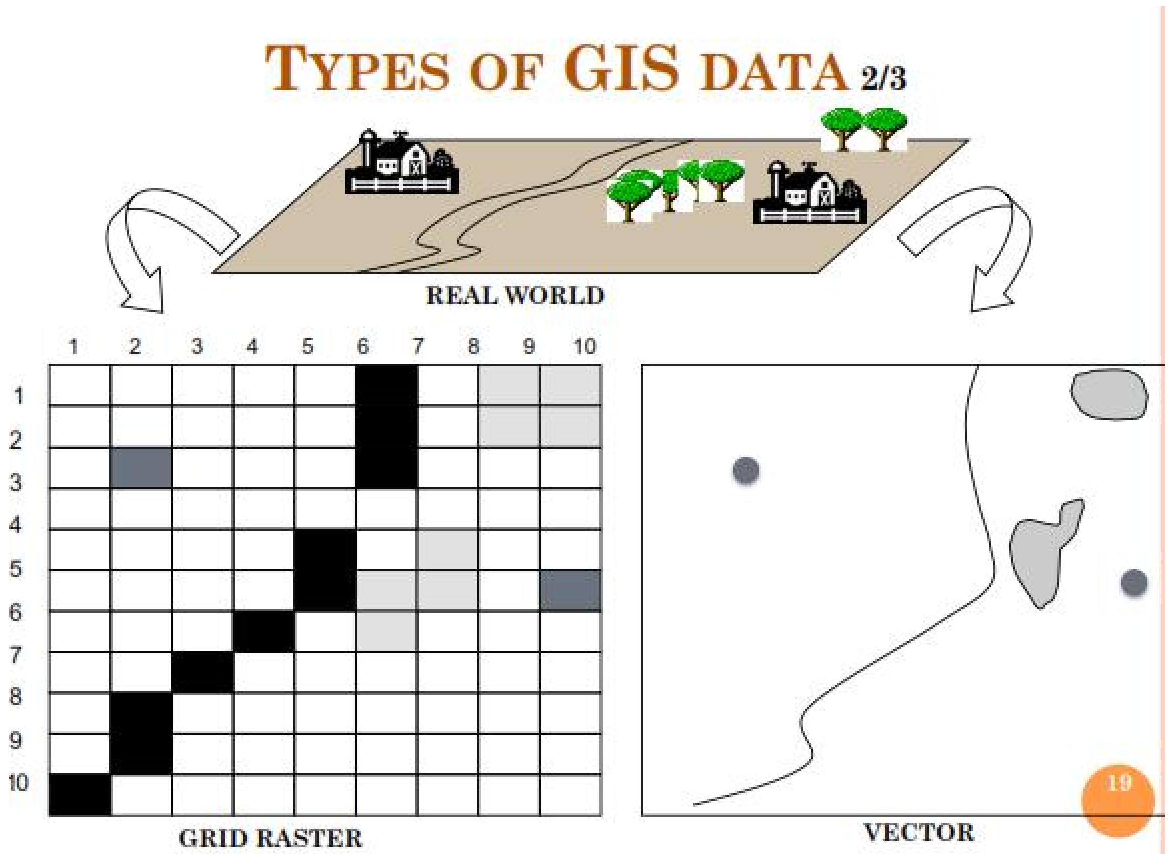
Attribite values in a GIS arestored as *relational database* tables.

Each feature (point, line, polygon,or raster) within each GIS layer will be represented as a record in a table.

A GIS stores information about the world as layers of spatial features (customers, buildings, streets, and so on).



Customers
Buildings
Streets
Reality
18
Source: ESRI



TYPES OF GIS DATA 2/3
REAL WORLD
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
GRID RASTER
VECTOR
19

Spatial Data

The spatial data or real world features are very complex. So, spatial data is simplified before they are entered into the computer. The common way of doing this is to break down all geographic features into three basic entity types – points, lines and areas.

Points are 'one dimensional' objects, used to represent features that are very small, e.g. a post box, an electric pole, a well or tube well etc. Only latitudinal and longitudinal values or a coordinate reference can be given to these features to explain their location.

Lines are two dimensional objects and are used to represent linear features, for example roads and rivers. Lines are also used to represent linear features that do not exist in reality, such as administrative boundaries and international boundaries. Areas are three dimensional objects and are represented by closed set of lines and are used to define features such as agricultural fields, forest areas, administrative areas etc. Area entities are often referred to as polygons.

The representation of real world features using the point, line and area entity types appears relatively simple. However, the appropriate entity to represent real world features is often difficult and it depends upon the scale of the map.

On a world map, cities are represented by points. It only gives information about number of cities shown on the world map.

At national or regional scale, the 'point' entity to represent cities is considered too simple, as it tells us nothing about the real size of the city.

In this case, cities are represented by 'area' entity.

At the local scale, 'area' entity to represent cities would be considered too simple. In this case, cities are represented by mixture of 'point', 'lines' and 'areas' as entities.

Points may be used to represent features such as electric poles, post boxes etc.

Likewise lines and areas may be used to represent road networks and residential blocks respectively.

So, the decision makers decide the 'entities' through which different features of real world would be represented.

Attribute Data

As it is mentioned earlier, attribute data tells the characteristics of different objects / features on the earth surface. These are descriptions, measurements or classification of geographic features.

Attribute data can be both qualitative (like land use type, soil type, name of the city/river etc.) and quantitative (like elevation, temperature, pressure of a particular place, crop yield per acre etc.).

So, the attribute can be both numeric and textual. The examples of attribute data of different spatial features like point (well), line (river), area (village) are shown in box 1. The attribute data are generally in tabular form.

A. INTRODUCTION

- the world is infinitely complex
- the contents of a spatial database represent a particular view of the world
- the user sees the real world through the medium of the database
    - the measurements and samples contained in the database must present as complete and accurate a view of the world as possible
    - the contents of the database must be relevant in terms of:
        - themes and characteristics captured
        - the time period covered
        - the study area

- this unit looks at techniques for sampling the world, and associated issues of accuracy, standards

B. REPRESENTING REALITY

- a database consists of digital representations of discrete objects
- the features shown on a map, e.g. lakes, benchmarks, contours can be thought of as discrete objects
    - thus the contents of a map can be captured in a database by turning map features into database objects

- many of the features shown on a map are fictitious and do not exist in the real world
    - contours do not really exist, but houses and lakes are real objects

- the contents of a spatial database include:
  - digital versions of real objects, e.g. houses
  - digital versions of artificial map features, e.g. contours
  - artificial objects created for the purposes of the database, e.g. pixels

Continuous variation

- some characteristics exist everywhere and vary continuously over the earth's surface
  - e.g. elevation, atmospheric temperature and pressure, natural vegetation or soil type

- we can represent such variation in several ways:
  - by taking measurements at sample points, e.g. weather stations
  - by taking transects
  - by dividing the area into patches or zones, and assuming the variable is constant within each zone, e.g. soil mapping
  - by drawing contours, e.g. topographic mapping

- each of these methods creates discrete objects
  - the objects in each case are points, lines or areas

- a raster can be thought of as:
  - a special case of a point sample where the points are regularly spaced
  - a special case of zones where the zones are all the same size

- each method is approximate, capturing only part of the real variation
  - a point sample misses variation between points
  - transects miss variation not on transects

  - zones pretend that variation is sudden at boundaries, and that there is no variation within zones
  - contours miss variation not located on contours

- several methods can be used to try to improve the success of each method
- e.g. for zones:
  - map the boundaries as fuzzy instead of sharp lines
  - describe the zones as mixtures instead of as single classes, e.g. 70% soil type A, 30% soil type B

C. SPATIAL DATA

- phenomena in the real world can be observed in three modes: spatial, temporal and thematic
  - the spatial mode deals with variation from place to place
  - the temporal mode deals with variation from time to time (one slice to another)
  - the thematic mode deals with variation from one characteristic to another (one layer to another)

- all measurable or describable properties of the world can be considered to fall into one of these modes - place, time and theme
- an exhaustive description of all three modes is not possible
- when observing real-world phenomena we usually hold one mode fixed, vary one in a controlled" manner, and measure"the third (Sinton, 1978)
  - e.g. using a census of population we could fix a time such as 1990, control for location using census tracts and measure a theme such as the percentage of persons owning automobiles

- holding geography fixed and varying time gives longitudinal data
- holding time fixed and varying geography gives cross- sectional data
- the modes of information stored in a database influence the types of problem solving that can be accomplished

Location

- the spatial mode of information is generally called location

Attributes

- attributes capture the thematic mode by defining different characteristics of objects
- a table showing the attributes of objects is called an attribute table
  - each object corresponds to a row of the table
  - each characteristic or theme corresponds to a column of the table
  - thus the table shows the thematic and some of the spatial modes

Time

- the temporal mode can be captured in several ways
  - by specifying the interval of time over which an object exists
  - by capturing information at certain points in time
  - by specifying the rates of movement of objects

- depending on how the temporal mode is captured, it may be included in a single attribute table, or be represented by series of attribute tables on the same objects through time

D. SAMPLING REALITY

Scales of measurement

- numerical values may be defined with respect to nominal, ordinal, interval, or ratio scales of measurement
- it is important to recognize the scales of measurement used in GIS data as this determines the kinds of mathematical operations that can be performed on the data
- the different scales can be demonstrated using an example of a marathon race:

1. Nominal

- on a nominal scale, numbers merely establish identity
    - e.g. a phone number signifies only the unique identity of the phone

- in the race, the numbers issued to racers which are used to identify individuals are on a nominal scale
    - these identity numbers do not indicate any order or relative value in terms of the race outcome

2. Ordinal

- on an ordinal scale, numbers establish order only
    - phone number 9618224 is not more of anything than 9618049, so phone numbers are not ordinal

- in the race, the finishing places of each racer, i.e. 1st place, 2nd place, 3rd place, are measured on an ordinal scale
    - however, we do not know how much time difference there is between each racer

3. Interval

- on interval scales, the difference (interval) between numbers is meaningful, but the numbering scale does not start at 0
    - subtraction makes sense but division does not
    - e.g. it makes sense to say that 200C is 10 degrees warmer than 100C, so Celsius temperature is an interval scale, but 200C is not twice as warm as 100C
    - e.g. it makes no sense to say that the phone number 9680244 is 62195 more than 9618049, so phone numbers are not measurements on an interval scale

- in the race, the time of the day that each racer finished is measured on an interval scale
    - if the racers finished at 9:10 GMT, 9:20 GMT and 9:25 GMT, then racer one finished 10 minutes before racer 2 and the difference between racers 1 and 2 is twice that of the difference between racers 2 and 3
    - however, the racer finishing at 9:10 GMT did not finish twice as fast as the racer finishing at 18:20 GMT

4. Ratio

- on a ratio scale, measurement has an absolute zero and the difference between numbers is significant
  - division makes sense
  - e.g. it makes sense to say that a 50 kg person weighs half as much as a 100 kg person, so weight in kg is on a ratio scale
  - the zero point of weight is absolute but the zero point of the Celsius scale is not

- in our race, the first place finisher finished in a time of 2:30, the second in 2:40 and the 450th place finisher took 5 hours
  - the 450th finisher took twice as long as the first place finisher (5/2.5 = 2)

- note these distinctions, though important, are not always clearly defined
  - is elevation interval or ratio? if the local base level is 750 feet, is a mountain at 2000 feet twice as high as one at 1000 feet when viewed from the valley?

- many types of geographical data used in GIS applications are nominal or ordinal
  - values establish the order of classes, or their distinct identity, but rarely intervals or ratios

- thus you cannot:
  - multiply soil type 2 by soil type 3 and get soil type 6
  - divide urban area by the rank of a city to get a meaningful number
  - subtract suitability class 1 from suitability class 4 to get 3 of anything

- however, you can:
  - divide population by area (both ratio scales) and get population density

- subtract elevation at point a from elevation at point b and get difference of elevation

Multiple representations

- a data model is essential to represent geographical data in a digital database
- there are many different data models
- the same phenomena may be represented in different ways, at different scales and with different levels of accuracy
- thus there may be multiple representations of the same geographical phenomena
- it is difficult to convert from one representation to another
  - e.g. from a small scale (1:250,000) to a large scale (1:10,000)

- thus it is common to find databases with multiple representations of the same phenomenon
  - this is wasteful, but techniques to avoid it are poorly developed

E. DATA SOURCES

Primary data collection

- some of the data in a spatial database may have been measured directly
  - e.g. by field sampling or remote sensing

- the density of sampling determines the resolution of the data

  - e.g. samples taken every hour will capture hour-to- hour variation, but miss shorter-term variation
  - e.g. samples taken every 1 km will miss any variation at resolutions less than 1 km

- a sample is designed to capture the variation present in a larger universe
  - e.g. a sample of places should capture the variation present at all possible places
  - e.g. a sample of times will be designed to capture variation at all possible times

- there are several standard approaches to sampling:
  - in a random sample, every place or time is equally likely to be chosen
  - systematic samples are chosen according to a rule, e.g. every 1 km, but the rule is expected to create

    no bias in the results of analysis, i.e. the results would have been similar if a truly random sample had been taken

  - in a stratified sample, the researcher knows for some reason that the universe contains significantly different sub-populations, and samples within each sub-population in order to achieve adequate representation of each
    - e.g. we may know that the topography is more rugged in one part of the area, and sample more densely there to ensure adequate representation
    - if a representative sample of the entire universe is required, then the subsamples in each subpopulation will have to be weighted appropriately

Secondary data sources

- some data may have been obtained from existing maps, tables, or other databases
    - such sources are termed secondary

- to be useful, it is important to obtain information in addition to the data themselves:
    - information on the procedures used to collect and compile the data
    - information on coding schemes, accuracy of instruments

- unfortunately such information is often not available
    - a user of a spatial database may not know how the data were captured and processed prior to input
    - this often leads to misinterpretation, false expectations about accuracy

F. STANDARDS

- standards may be set to assure uniformity
    - within a single data set
    - across data sets
    - e.g. uniform information about timber types throughout the database allows better fire fighting methods to be used, or better control of insect infestations

- data capture should be undertaken in standardized ways that will assure the widest possible use of the information

Sharing data

- it is not uncommon for as many as three agencies to create databases with, ostensibly, the same information
  - e.g. a planning agency may map landuse, including a forested class
  - e.g. the state department of forestry also maps forests
  - e.g. the wildlife division of the department of conservation maps habitat, which includes fields and forest

- each may digitize their forest class onto different GIS systems, using different protocols, and with different definitions for the classes of forest cover
- this is a waste of time and money
- sharing information gives it added value
- sharing basic formats with other information providers, such as a department of transportation, might make marketing the database more profitable

Agency standards

- state and national agencies have set standards for certain environmental data
  - the Soil Conservation Service (SCS) has adopted the "seventh approximation"as the national taxonomy
  - the US Geological Survey has set standards for landuse, transportation, and hydrography that are used as guidelines in many states
  - forest inventories are not standardized; agencies may use different systems while managing a contiguous region of forest land

- Unit 69 covers standards for GIS in greater depth

G. ERRORS AND ACCURACY

- note: Units 45 and 46 discuss this topic in detail
- there is a nearly universal tendency to lose sight of errors once the data are in digital form
- errors:
  - are implanted in databases because of errors in the original sources (source errors)
  - are added during data capture and storage (processing errors)
  - occur when data are extracted from the computer
  - arise when the various layers of data are combined in an analytical exercise

Original Sin - errors in sources

- are extremely common in non-mapped source data, such as locations of wells, or lot descriptions
- can be caused by doing inventory work from aerial photography and misinterpreting images
- often occur because base maps are relied on too heavily
  - a recent attempt in Minnesota to overlay Department of Transportation bridge locations on USGS transportation data resulted in bridges lying neither beneath roads, nor over water, and roads lying apparently under rivers
  - until they were compared in this way, it was assumed that each data set was locationally acceptable
  - the ability of GIS to overlay may expose previously unsuspected errors

Boundaries

- boundaries of soil types are actually transition zones, but are mapped by lines less than 0.5 mm wide
- lakes fluctuate widely in area, yet have permanently recorded shorelines

Classification errors

- are common when tabular data are rendered in map form
- simple typing errors may be invisible until presented graphically
  - floodplain soils may appear on hilltops
  - pastureland may appear to be misinterpreted marsh

- more complex classification errors may be due to the sampling strategies that produced the original data
- timber appraisal is commonly done using a few, randomly selected points to describe large stands
  - information may exist that documents the error of the sampling technique
  - however, such information is seldom included in the GIS database

Data capture errors

- manual data input induces another set of errors
- eye-hand coordination varies from operator to operator and from time to time
  - data input is a tedious task - it is difficult to maintain quality over long periods of time

Accuracy standards

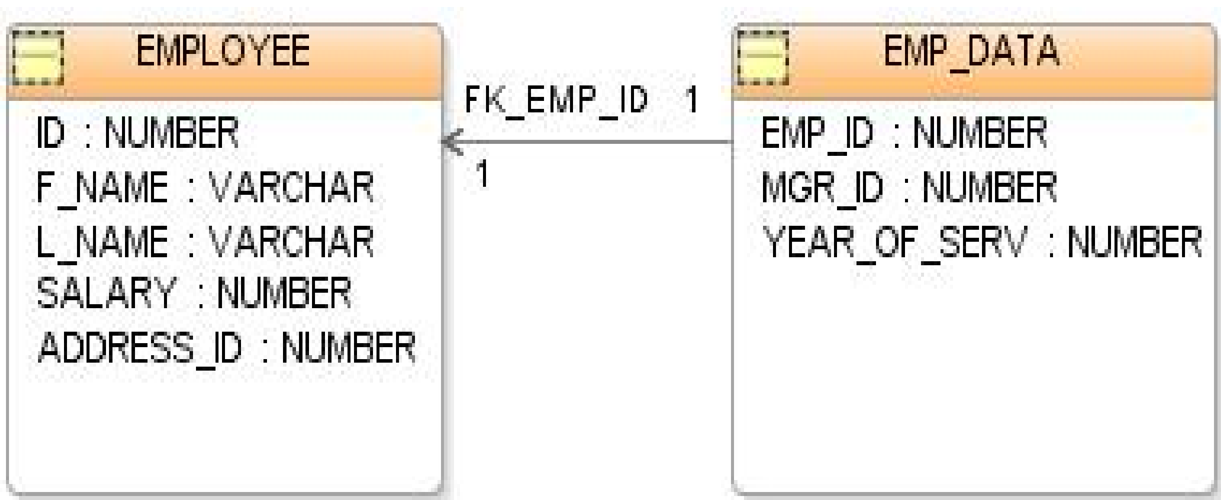
- many agencies have established accuracy standards for geographical data
  - these are more often concerned with accuracy of locations of objects than with accuracy of attributes

- location accuracy standards are commonly decided from the scale of source materials
  - for natural resource data 1:24,000 scale accuracy is a common target
  - at this scale, 0.5 mm line width = 12 m on the ground

- USGS topographic information is currently available in digital form at 1:100,000
  - 0.5 mm line width = 50 m on the ground

- higher accuracy requires better source materials
  - is the added cost justified by the objectives of the study?

- accuracy standards should be determined by considering both the value of information and the cost of collection



A **Database Management System** (**DBMS**) is a set of computer programs that controls the creation, maintenance, and the use of the database of an organization and its end users. It allows organizations to place control of organization-wide database development in the hands of database administrators (DBAs) and other specialists. DBMSes may use any of a variety of database models, such as the network model or relational model. In large systems, a DBMS allows users and other software to store and retrieve data in astructured way. It helps to specify the logical organization for a database and access and use the information within a database. It provides facilities for controlling data access, enforcing data integrity, managing concurrency controlled, restoring database.

Meta-data repository

Metadata is data describing data. For example, a listing that describes what attributes are allowed to be in data sets is called "meta-information". The meta-data is also known as data about data.

DBMS Current Trends

As of 1998 database management was in need of new style databases to solve current database management problems. Researchers realized that the old trends of database management were becoming too complex and there was a need for automated configuration and management [3]. Surajit Chaudhuri, Gerhard Weikum and Michael Stonebraker, were the pioneers that dramatically affected the thought of database management systems [3]. They believed that database management needed a more modular approach and that there are so many specifications needs for various users [3]. Since this new development process of database management we currently have endless possibilities. Database management is no longer limited to “monolithic entities” [3]. Many solutions have developed to satisfy individual needs of users. Development of numerous database options has created flexible solutions in database management. Today there are several ways database management has affected the technology world as we know it. Organizations demand for directory services has become an extreme necessity as organizations grow. Businesses are now able to use directory services that provided prompt searches for their company information [3]. Mobile devices are not only able to store contact information of users but have grown to bigger capabilities. Mobile technology is able to cache large information that is used for computers and is able to display it on smaller devices [3]. Web searches have even been affected with database management. Search engine queries are able to locate data within the World Wide Web [3]. Retailers have also benefited from the developments with data warehousing. These companies are able to record customer transactions made within their business [3]. Online transactions have become tremendously popular with the e-business world. Consumers and businesses are able to make payments securely on company websites. None of these current developments would have been possible without the evolution of database management. Even with all the progress and current trends of database management, there will always be a need for new development as specifications and needs grow.

Examples of Database Management Systems

Adabas Adaptive Server Enterprise Alpha Five Computhink's View Wise CSQL Daffodil DB DataEase	FileMaker Firebird Glom IBM DB2 IBM UniVerse Ingres	Informix InterSystems Caché Kexi WX2 Linter SQL RDBMS Lotus Approach Mark Logic Microsoft Access Microsoft SQL Server