KCA UNIVERSITY

LECTURE NOTES

BIT 2307: GEOGRAPHICAL INFORMATION SYSTEMS

FOR:
BSC INFORMATION TECHNOLOGY & BSC BUSINESS INFORMATION TECHNOLOGY

DEVELOPED BY:

Edwin Omol-Lecturer, KCA UNIVERSITY
+254717197587/+254735458075
omoledwin@gmail.com

BIT 2307: GEOGRAPHICAL INFORMATION SYSTEMS

Table of Contents

1.0 INTRODUCTION TO GIS

1.1 Preamble

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

GIS is a set of tools (or technology) that allow for the processing of ***spatial data*** into ***information***. This set of tools is open ended, but will include data input, data storage, data manipulation, and a reporting system.

The 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**.

GIS lets us visualize, question, analyze, and interpret data to understand relationships, patterns, and trends. GIS benefits organizations of all sizes and in almost every industry. There is a growing interest in and awareness of the economic and strategic value of GIS

In a general sense, the term describes any **information system** that integrates, stores, edits, analyzes, shares, and displays **geographic** information. Information Systems allows the transformation of data into information via: structuring, formatting, conversion and modeling

GIS: transforms data with a spatial component. A spatial components here refer to location- based data, with coordinates (E,N,H or Latitude and longitude), on, below or above earths surface. The data with the "where".

GIS technology has been called by different names like: Geospatial Information Systems, Computer Aided Mapping, Automated Cartography, Land Information Systems, Environmental Information Systems, etc

1.2 Definition of GIS

Geographical Information System has been defined differently by different authors

§ "…a system of hardware, software, and procedures designed to support the capture, management, manipulation, analysis, modeling, and display of spatially referenced data for solving complex planning & management problems" (Rhind, 1989)

§ "…a computer system capable of assembling, storing, manipulating, and displaying geographically referenced information…" (USGS, 1997)

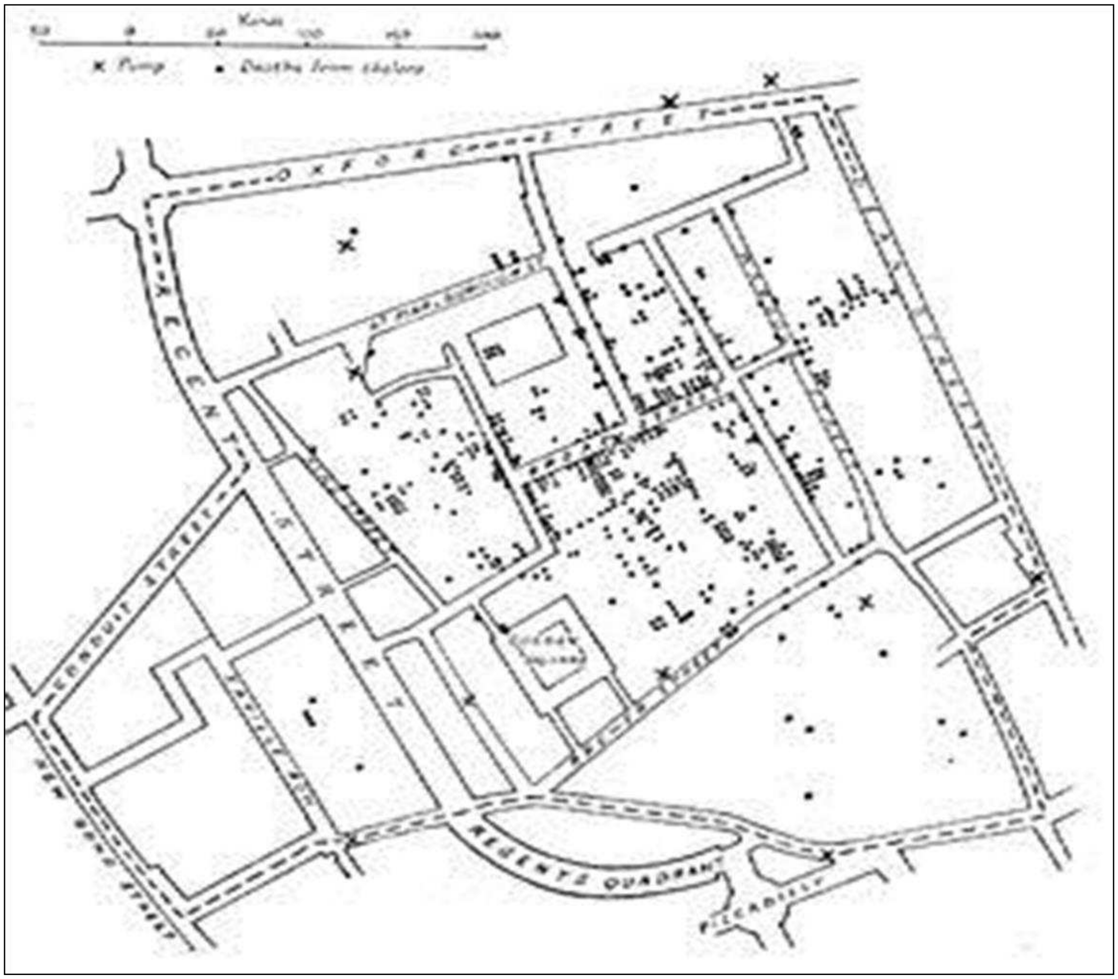
§ "…a set of computer-based systems for managing geographic data and using those data to solve spatial problems" (Lo & Yeung, 2002)

§ a computer system that allows the analysis and display of data with a spatial component (Phillips, 2002)

In general it can be defined as a ***"computer-based system that support capture, management, manipulation, analysis, modeling, and display of spatially referenced data that provides planning, management, development and humanitarian solutions".***

1.3 History of Geographic Information System

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". Tomlinson is also acknowledged as the "**father of GIS**".



***E. W. Gilbert's** version (1958) of **John Snow**'s 1855 map of the Soho cholera outbreak showing the clusters of cholera cases in the **London** epidemic of 1854*

Previously, one of the first applications of spatial analysis in **epidemiology** is the 1832. The **French geographer Charles Picquet** represented the 48 districts of the city of Paris by halftone color gradient according to the percentage of deaths by cholera per 1,000 inhabitants.

In 1854 **John Snow** depicted a **cholera** outbreak in **London** using points to represent the locations of some individual cases, an early successful use of a geographic methodology in epidemiology. While the basic elements of **topography** and theme existed previously in **cartography**, the John Snow map was unique, using cartographic methods not only to depict but also to analyze clusters of geographically dependent phenomena.

The early 20th century saw the development of **photozincography**, which allowed maps to be split into layers, for example one layer for vegetation and another for water. This was particularly used for printing contours – drawing these was a labor-intensive task but having them on a separate layer meant they could be worked on without the other layers to confuse the **draughtsman**. This work was originally drawn on glass plates but later **plastic film** was introduced, with the advantages of being lighter, using less storage space and being less brittle, among others. When all the layers were finished, they were combined into one image using a large process camera. Once color printing came in, the layers idea was also used for creating separate printing plates for each color. While the use of layers much later became one of the main typical features of a contemporary GIS, the photographic process just described is not considered to be a GIS in itself – as the maps were just images with no database to link them to.

Computer hardware development spurred by **nuclear weapon** research led to general-purpose computer "mapping" applications by the early 1960s.The year 1960 saw the development of the world's first true operational GIS in **Ottawa, Ontario**, Canada by the federal Department of Forestry and Rural Development. Developed by **Dr. Roger Tomlinson**, it was called the **Canada Geographic Information System** (CGIS) and was used to store, analyze, and manipulate data collected for the Canada Land Inventory – an effort to determine the land capability for rural Canada by mapping information about **soils, agriculture, recreation, wildlife, waterfowl, forestry and land use** at a scale of **1:50,000**. A rating classification factor was also added to permit analysis.

CGIS was an improvement over "computer mapping" applications as it provided capabilities for **overlay, measurement, and digitizing/scanning**. It supported a national coordinate system that spanned the continent, coded lines as **arcs** having a true embedded **topology** and it stored the attribute and locational information in separate files. As a result of this, Tomlinson has become known as the "father of GIS", particularly for his use of overlays in promoting the spatial analysis of convergent geographic data.

CGIS lasted into the 1990s and built a large digital land resource database in Canada. It was developed as a **mainframe**-based system in support of federal and provincial resource planning and management. Its strength was continent-wide analysis of complex **datasets**. The CGIS was never available commercially.

In 1964 Howard T. Fisher formed the Laboratory for Computer Graphics and Spatial Analysis at the **Harvard Graduate School of Design** (LCGSA 1965–1991), where a number of important theoretical concepts in spatial data handling were developed, and which by the 1970s had distributed seminal software code and systems, such as SYMAP, GRID, and ODYSSEY – that served as sources for subsequent commercial development—to universities, research centers and corporations worldwide.

By the early 1980s, **M&S** Computing (later **Intergraph**) along with Bentley Systems Incorporated for the **CAD** platform, Environmental Systems Research Institute (**ESRI**), **CARIS** (Computer Aided Resource Information System), **MapInfo Corporation** and **ERDAS** (Earth Resource Data Analysis System) emerged as **commercial vendors** of GIS software, successfully incorporating many of the CGIS features, combining the **first generation approach to separation of spatial and attribute information with a second generation approach to organizing attribute data into database structures**. In parallel, the development of two public domain systems (**MOSS** and **GRASS GIS**) began in the late 1970s and early 1980s.

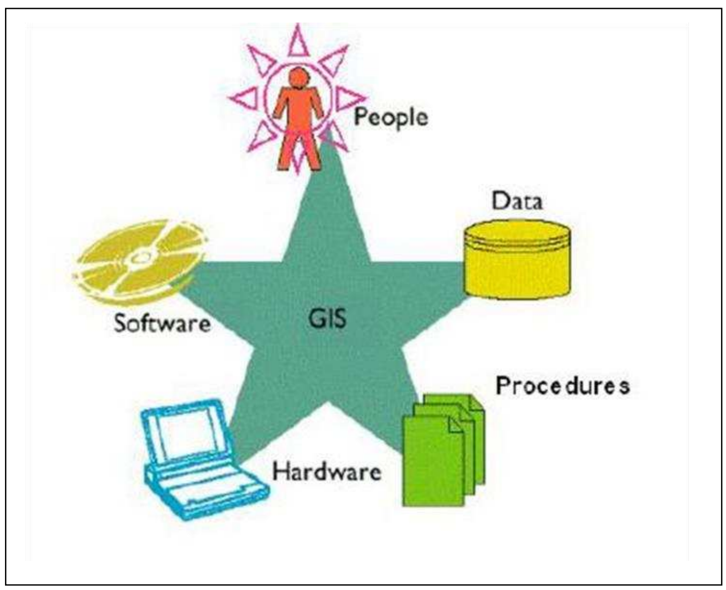
In 1986, Mapping Display and Analysis System (**MIDAS**), the first desktop GIS product emerged for the **DOS** operating system. This was renamed in 1990 to MapInfo for Windows when it was ported to the **Microsoft Windows** platform. This began the process of moving GIS from the research department into the business environment.

By the end of the 20th century, the rapid growth in various systems had been consolidated and standardized on relatively few platforms and users were beginning to explore viewing GIS data over the **Internet**, requiring data format and transfer standards. More recently, a growing number of **free, open-source GIS packages** run on a range of operating systems and can be customized to perform specific tasks. Increasingly **geospatial data** and **mapping applications** are being made available via the **world wide web.**

1.4 Components of GIS

An operational GIS also has a series of components that combine to make the system work. These components are critical to a successful GIS. There include: Hardware, software, data, people, methods/procedures

- **Hardware:** Hardware is the computer system on which a GIS operates. Today, GIS software 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. A review of the key GIS software subsystems is provided above.



- **Data:** Perhaps the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house, compiled to custom specifications and requirements, or occasionally purchased from a commercial data provider. A GIS can integrate spatial data with other existing data resources, often stored in a corporate DBMS. The integration of spatial data (often proprietary to the GIS software), and tabular data stored in a DBMS is a key functionality afforded by GIS.
- **People:** GIS technology is of limited value without the people who manage the system and develop plans for applying it to real world problems. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work. The identification of GIS specialists versus end users is often critical to the proper implementation of GIS technology.

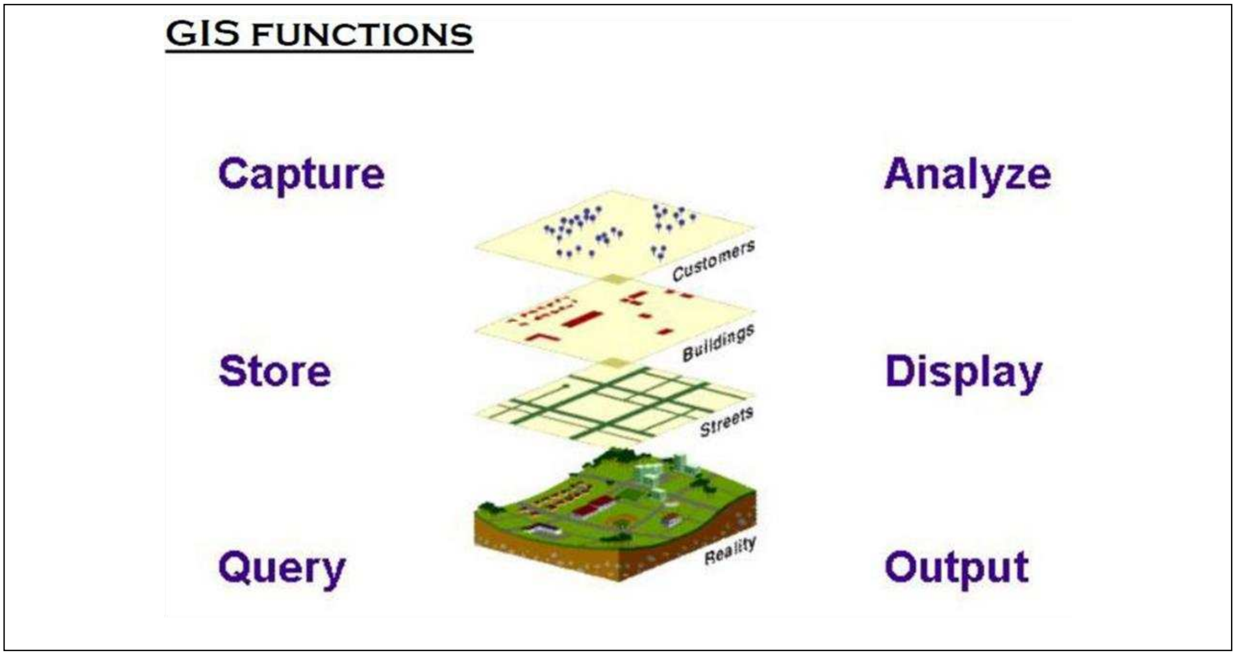
- **Methods:** A successful GIS operates according to a well-designed implementation plan and business rules, which are the models and operating practices unique to each organization.

1.5 GIS Subsystems and Functions

A GIS software has **four** main functional subsystems. These are:

→ a data input subsystem;
→ a data storage and retrieval subsystem;
→ a data manipulation and analysis subsystem; and
→ a data output and display subsystem.

- **Data Input:** A data input subsystem allows the user to capture, collect, and transform spatial and thematic data into digital form. The data inputs are usually derived from a combination of hard copy maps, aerial photographs, remotely sensed images, reports, survey documents, etc.



- Data Storage and Retrieval: The data storage and retrieval subsystem organizes the data, spatial and attribute, in a form which permits it to be quickly retrieved by the user for analysis, and permits rapid and accurate updates to be made to the database. This component usually involves use of a database management system (DBMS) for maintaining attribute data. Spatial data is usually encoded and maintained in a proprietary file format.
- **Data Manipulation and Analysis:** The data manipulation and analysis subsystem allows the user to define and execute spatial and attribute procedures to generate derived information. This subsystem is commonly thought of as the *heart of a GIS*, and usually distinguishes it from other database information systems and computer-aided drafting (CAD) systems.
- **Data Output:** The data output subsystem allows the user to generate graphic displays, normally maps, and tabular reports representing derived information products.

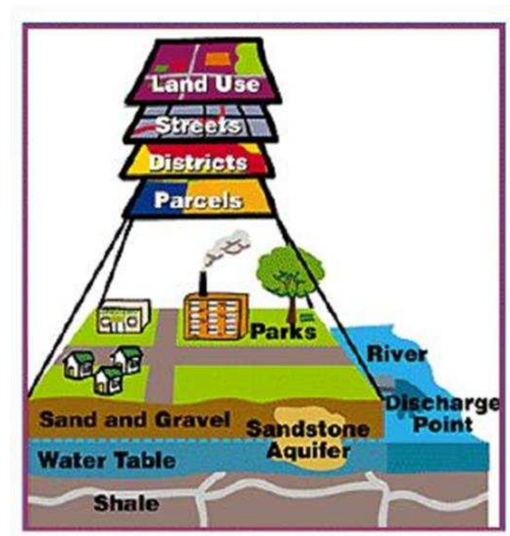
The critical function for a GIS is, by design, the analysis of spatial data.

Basic GIS functions are contained in the definition and include among others: data capture, storage, query, analysis, display and output/presentation or visualization.

1.6 GIS Data Models (Organization)

A GIS stores information about the world as a collection of thematic layers that can be linked together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to recording details of planning applications, to modeling global atmospheric circulation. ***The thematic layer approach allows us to organize the complexity of the real world into a simple representation to help facilitate our understanding of natural relationships***. Layers represents specific feature class like: roads/streets, parcels of land, District boundary and land use

1.7 Some Advantages of GIS



1) **Cost Savings from Greater Efficiency:** Typical implementations can result in a savings of 10 to 30 percent in operational expenses through reduction in fuel use and staff time, improved customer service, and more efficient scheduling. ***GIS helped the City of Woodland refine its fleet scheduling, saving fuel and labour.***
2) **Better Decision Making:** GIS is the go-to technology for making better decisions about location. Common examples include real estate site selection, route/corridor selection, evacuation planning, conservation, natural resource extraction, etc. Making correct decisions about location is critical to the success of an organization. For instance, ***GIS-based disaster decision support system helps Taiwan plan for and respond to typhoons***.

3) **Improved Communication:** GIS-based maps and visualizations greatly assist in understanding situations and in storytelling. They are a type of language that improves communication between different teams, departments, disciplines, professional fields, organizations, and the public. ***Michels Corporation improved collaboration and communication with GIS***

4) **Managing Geographically:** GIS is becoming essential to understanding what is happening and what will happen in geographic space. Once we understand, we can prescribe action. This new approach to management—managing geographically—is transforming the way organizations operate. ***Kuwait University used GIS to design and build a multibillion- dollar expansion***.

1.8 Application Areas

GIS are now used extensively in government, military, medical, industry, business, and academic/research for a wide range of applications including environmental resource analysis, land use planning, locational analysis, tax appraisal, utility and infrastructure planning, real estate analysis, marketing and demographic analysis, habitat studies, and archaeological analysis.

One of the first major areas of application was in **natural resources management**, including management of

- wildlife habitat,
- wild and scenic rivers,
- recreation resources,
- floodplains,
- wetlands,
- agricultural lands,
- aquifers,
- forests.

One of the largest areas of application has been in **facilities management**. Uses for GIS in this area have included

- locating underground pipes and cables,
- balancing loads in electrical networks,
- planning facility maintenance,
- tracking energy use.

Local, state, and federal governments have found GIS particularly useful in **land management**. GIS has been commonly applied in areas like

- zoning and subdivision planning,
- land acquisition,
- environmental impact policy,
- water quality management,
- Maintenance of ownership.

More recent and innovative uses of GIS have used information based on **street-networks**. GIS has been found to be particularly useful in:

- address matching,
- location analysis or site selection,
- development of evacuation plans.

1.9 GIS Users

§ Specialist: includes programmers, designers, developers
§ General Users: planners, medics, scientists, administrators, teachers (us)
§ Viewers: everyone (our “clients”)

Examples of Applied GIS

- **Urban Planning, Management & Policy**
  - Zoning, subdivision planning
  - Land acquisition
  - Economic development
  - Code enforcement
  - Housing renovation programs
  - Emergency response
  - Crime analysis
  - Tax assessment
- **Environmental Sciences**
  - Monitoring environmental risk
  - Modeling stormwater runoff
  - Management of watersheds, floodplains, wetlands, forests, aquifers
  - Environmental Impact Analysis
  - Hazardous or toxic facility siting
  - Groundwater modeling and contamination tracking
- **Political Science**
  - Redistricting
  - Analysis of election results
  - Predictive modeling
- **Civil Engineering/Utility**
  - Locating underground facilities
  - Designing alignment for freeways, transit
  - Coordination of infrastructure maintenance
- **Business**
  - Demographic Analysis
  - Market Penetration/ Share Analysis
  - Site Selection
- **Education Administration**
  - Attendance Area Maintenance
  - Enrollment Projections
  - School Bus Routing
- **Real Estate**
  - Neighborhood land prices
  - Traffic Impact Analysis
  - Determination of Highest and Best Use
- **Health Care**
  - Epidemiology
  - Needs Analysis
  - Service Inventory

2.0 GIS DATA TYPES AND MODELS

Spatial data are computer representations of spatial features of the real world-Representation of Real World. A modeling language for a GIS database is a spatial data model. A spatial database holds a digital representation of the real world. Among spatial data models, we can distinguish two major types, field and object-based models.

- Field-based models consider spatial phenomena to be of a continuous nature where in every point in space a value of the field can be determined. Examples of such phenomena are temperature, barometric pressure, or elevation.
- Object-based models consider space to be populated by well distinguishable, discrete, bounded objects with the space between objects potentially being empty. Examples include parcels, buildings, trees, among others

It is therefore common to see geographical phenomenon of spatial features in terms of field, objects and boundary- boundary simply defining extent of interest. i.e
Geographic phenomenon= Field, Object, Boundary

The primitives that are used to represent features in the object view comprise point, line, area/polygon, and volume items. It is often called the vector-based approach

In the field view, the part of space relevant to us is divided up completely into regular or irregular tessellations—usually squares, triangles, or cubes—that define a two- or three dimensional raster or grid. The more common name for the field view approach is the raster-based approach

2.1 GIS Data Types

A GIS deals with spatial data (e.g., parcels, rivers, wells,), their attributes and characteristics (e.g., location, area, length, name, depth …) and the relationships between the objects (e.g., a parcel boundary follows a river, a well is located in a certain parcel …). The objects are stored in the database with geometric primitives (volumes, areas, lines, and points) and the relationships between them (topology). Spatial data have the following characteristics:

Table 2.1: Characteristics of spatial data

Spatial reference (geographic location, coordinates)	where?
Attributes (non-spatial)	what?
Spatial relationships (topology, metric, order)	in what relationship?
Temporal component (different concepts of time)	when?

Source: Kainz, Wolfgang (2004)

It is common to equate GIS Data with three "triplex" characteristics- description, reference and time). Description is the attribute, reference is the spatial reference/data, and time as the reference and description may change over time.

The basic data type in a GIS reflects traditional data found on a map, and currently digitally in GIS or spatial databases. Accordingly, GIS technology utilizes two basic types of data. These are: spatial data (location and geometry), attribute data (descriptions) and spatial relationships

Spatial data: describes the absolute and relative location of geographic features.

Attribute data: describes characteristics of the spatial features. These characteristics can be quantitative and/or qualitative in nature. Attribute data is often referred to as tabular data.

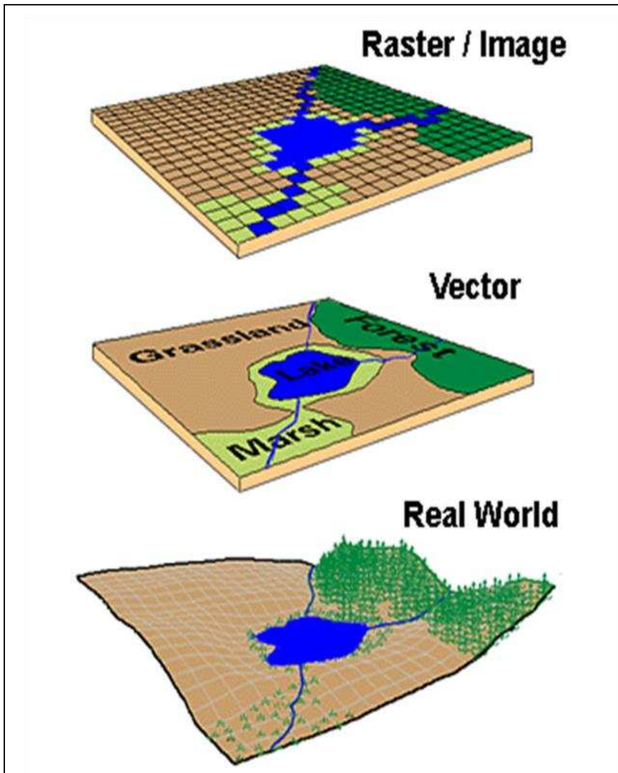
The coordinate location of a forestry stand would be spatial data, while the characteristics of that forestry stand, e.g. cover group, dominant species, crown closure, height, etc., would be attribute data.

Relationships (rules or behavior): topological relationship such as coincidence, adjacency and connectivity

2.2 Spatial Data Models (Formats)

Traditionally spatial data has been stored and presented in the form of a map. Three basic types of spatial data models have evolved for storing geographic data digitally. These are referred to as:

- Vector
- Raster
- Image



The following diagram reflects the two primary spatial data encoding techniques. These are vector and raster.

a. Vector Model (Format)

Vector model uses discrete points, lines and/or areas corresponding to discrete objects with name or code number of attributes.

b. Raster Model (Format)

Raster model uses regularly spaced grid cells in specific sequence. An element of the grid cell is called a pixel which contains a single value of attributes.

c. Image data utilizes techniques very similar to raster data, however typically lacks the internal formats required for analysis and modeling of the data. Images reflect *pictures* or *photographs* of the landscape.

The data model represents a set of guidelines to convert the real world (called entity) to the digitally and logically represented spatial objects consisting of the attributes and geometry.

2.3 GIS Data representations

Spatial data in GIS represents features that have a known location on the earth.

Points: X & Y Locations
Polygon: Connected X & Y Locations making aclose figure.
Line: Connected X & Y Locations without closing
Raster: Row and column matrix represent geographic space.

2.4 Vector Data Structure

The method of representing geographic features by the basic graphical elements of **points, lines and polygon** is said to be the **vector method**, or **vector data model.** Vector data represent geographic space that is intuitive and reminiscent of analog maps.

Vector data mostly incorporate the use to **topological structure**, especially the ***Arc/Node structure*** explained below.

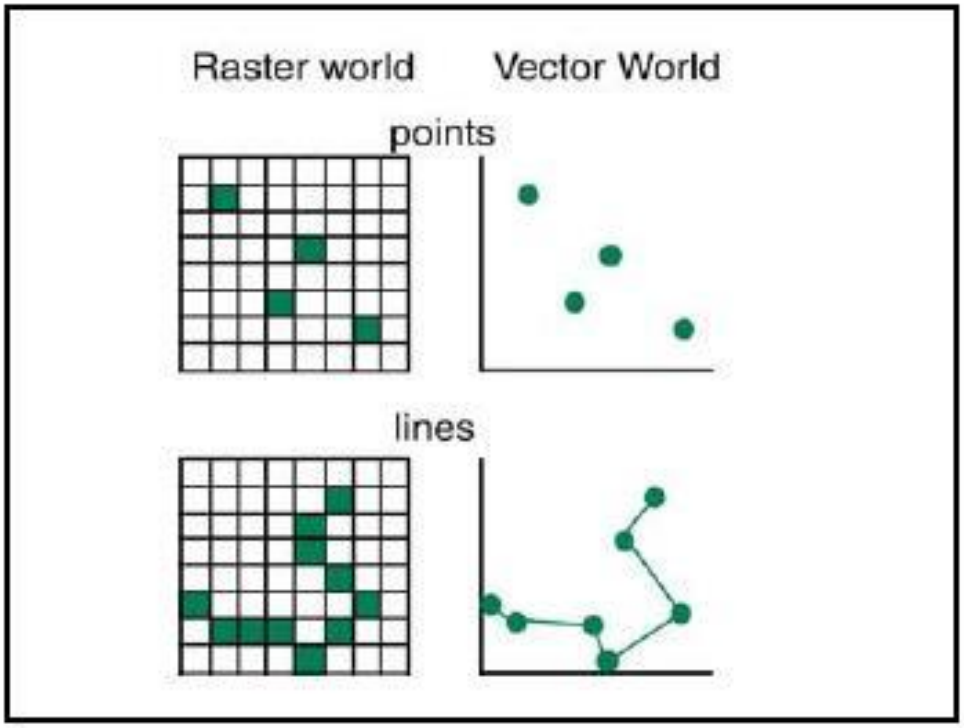
2.5 Rasters Data structure

Raster data models incorporate the use of a *grid-cell* data structure where the geographic area is divided into cells identified by row and column. A raster is a *tesselation* of a surface.• (A *tesselation* is defined as the process to cover a surface through the repeated use of a **single shape**.)

What is a raster data? A raster based system stores data by using a grid of cells. A unique reference coordinate represents each pixel either at a corner or in the middle of the cell. Each cell or pixel has discrete attributes assigned to it
Raster data resolution is dependent on the pixel or grid size and may vary from sub-meter to many kilometers. raster data stores different information in layers; elevation, soil type, geology, forest type, rainfall rate, etc.
Generally, raster data requires less processing than vector data, but it consumes more computer storage space. Examples include:

a. Remote sensors on satellites store data in raster format
b. Digital terrain models (DTM) and digital elevation models (DEM)
c. Continuous data (FIELD) suit a raster structure

A raster can use any ***reasonable*** geometric shape, as long as it can be connected in such a way as to create a ***continuous*** surface.



Comparison of Raster and Vector Data Models

Raster Model	Vector Model
Advantage:	Advantage:
1. It is a simple data structure.	1. It provides a more compact data structure than the raster model.
2. Overlay operations are easily and efficiently implemented.	2. It provides efficiently encoding of topology and as result more efficiently implementation of operations that require topological information, such as network analysis.
3.High spatial variability is efficiently represented in raster format.	3.The vector model is better suited to supporting graphics that closely approximate Hand-drawn maps.
4.The raster format is more or less required for efficient manipulation and enhancement of digital images.	

2.6 Attribute Data and Attribute Data Model

Attribute data are the information linked to the geographic features (spatial data) that describe them. That is, attribute data are the "[n]on-graphic information associated with a point, line, or area elements in a GIS."

Attribute data model is a separate data model is used to store and maintain attribute data for GIS software. These data models may exist internally within the GIS software, or may be reflected in external commercial Database Management Software (DBMS). A variety of different data models exist for the storage and management of attribute data. The most common: tabular, hierarchical, network, relational and object oriented models. Only relational models is commonly used in GIS currently

–

2.7 Relational Model

The relational database organizes data in *tables*. Each table, is identified by a unique table name, and is organized by *rows* and *columns*. Each column within a table also has a unique name. Columns store the values for a specific attribute, e.g. cover group, tree height. Rows represent one record in the table. In a GIS each row is usually linked to a separate spatial feature, e.g. a forestry stand. Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature. The following figure presents a sample table for forest inventory features. This table has 4 rows and 5 columns. The forest stand number would be the *label* for the spatial feature as well as the *primary key* for the database table. This serves as the linkage between the spatial definition of the feature and the attribute data for the feature.

UNIQUE STAND NUMBER	DOMINANT COVER GROUP	AVG. TREE HEIGHT	STAND SITE INDEX	STAND AGE
001	DEC	3	G	**100**
002	DEC-CON	4	M	**80**
003	DEC-CON	4	M	**60**
004	**CON**	**4**	**G**	**120**

Data is often stored in several tables. Tables can be joined or referenced to each other by common columns (relational fields). Usually the common column is an identification number for a selected geographic feature, e.g. a forestry stand polygon number. This identification number acts as the *primary key* for the table. The ability to join tables through use of a common column is the essence of the relational model. Such relational joins are usually ad hoc in nature and form the basis of for querying in a relational GIS product.

The relational database model is the most widely accepted for managing the attributes of geographic data.

–

The relational DBMS is attractive because of its:

a) Simplicity in organization and data modelling
b) Flexibility - data can be manipulated in an ad hoc manner by joining tables.

c) Efficiency of storage - by the proper design of data tables redundant data can be minimized; and
d) The non-procedural nature - queries on a relational database do not need to take into account the internal organization of the data.

The relational DBMS has emerged as the dominant commercial data management tool in GIS implementation and application.

–

2.8 Topology (Spatial Relationships)

The topologic model is often confusing to initial users of GIS. ***Topology is a mathematical approach that allows us to structure data based on the principles of feature adjacency and feature connectivity. It is in fact the mathematical method used to define spatial relationships.*** Without a topologic data structure in a vector based GIS most data manipulation and analysis functions would not be practical or feasible.

The most common topological data structure is the *arc/node* data model. This model contains two basic entities, the *arc* and the *node*.

- The arc is a series of points, joined by straight line segments, which start and end at a node. Arc presents **linear features**
- The node is an intersection point where two or more arcs meet. Nodes also occur at the end of a *dangling* arc, e.g. an arc that does not connect to another arc such as a dead end street.
- Isolated nodes, not connected to arcs represent **point features**.
- **A polygon feature** is comprised of a closed chain of arcs.

Topology is the “way in which geographical elements are linked together”. Topology is how geographic features are related to one another and where they are in relation to one another.

Topology is the critical element that distinguishes a GIS from a graphics or automated cartography system. It is essential to the ability of a GIS to employ spatial relationships. Topology is what enables a GIS to emulate our human ability to discern and manipulate geographic relationships.

2.9 Labels

Affixed to data points, lines, or polygons.
• Used to describe the **feature** that you want to map.
• Can include text or numeric descriptors: i.e. nominal, ordinal, or interval/ratio data types.
• Must be careful in how the different data types are integrated and used – dangerous to mix and match.

2.10 GIS Database Creation/maintenance

Spatial data capture and edition goes hand in hand with GIS database creation. Sometime, or many a times, the database is designed long before data capture

Database is a large, computerized collection of structured data. In non-spatial domain, databases have been used since 1960s in bank account administration, stock monitoring, salary administration, stores management, flight reservations etc.

Designing a database is not an easy task. One has to consider the purpose, available data sources and formats before data can be entered into the database. The format means data structure. Database should be properly maintained and up-to-date.

Database Management Systems (DBMS) are a software package that allows the user to set up, use and maintain a database.

GIS have inbuilt Ms Access for small databases, but has ArcSDE for large databases. GIS stores data mainly in relational database discussed above

Reasons for database

- DBMS supports storage and maintenance of very large volume of datasets
- DBMS can be instruevcted to guard over data correctness or valididity
- DBMS supports concurrent use of the same dataset for many users
- Provides high level declarative query language. A query is is a computer programme that exctracts data from the database based on specified condition
- A DBMS supports use of data model. A data model is language which can apply to define data structure and in data manipulation and storage
- Include data backup and recovery functions
- Allows control over data redundancy

Relational data model

A data model is a language that allows: The data structures to be stored; the integrity constraints, and the computer programme to manipulate the data

For relational data model, the structure used to define the database is attributes, tuples and relations. Computer program can exctract information applying quelry and change content (update)

In relational data model, a database is vieed as a collection of relations, commonly known as tables. Atable or a relation is a collection of tuples or records. An attribute is a named field of tuple. Same attribute comes from single domain (text/string, integer, real, date, etc)

When creating a relation is created, we need to indicate the type of tuple to strore:

- Provide name of relation
- Indicate name of attributes
- Set domain for each attribute

Undelaying the attribute and their domain indicate primary key of the relations

The set of tuples in a relation at a given point is called a relation stance

Linking GIS and database

GIS software provide support for spatial data and attribute data. GIS stores spatial and attribute data separately, thus, it links them theorugh a primary key.

3.0 GIS DATA CAPTURE AND EDITING

3.1 Definition & Source of Data

GIS data capture and editing means the identification, collection, digitization and correction of errors for the data necessary in the building of a GIS database. This is the most expensive and critical phase in GIS setup. Building a GIS database could take up to 5-10 times the cost of hardware and software or 70-80% of the total cost.

The main sources of data for GIS are:

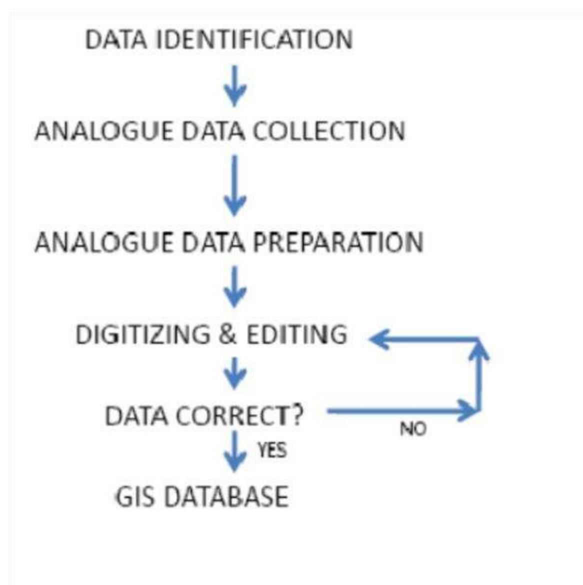
a) Analogue maps and plans;
b) Digital RS images
c) Surveying field notes
d) Aerial photographs
e) Tabular data e.g. census, rainfall, soils etc
f) GPS receivers data
g) Direct import from other GIS systems
h) Existing digital data

When collecting data, collect only the data needed to provide for the information needs of users as determined from user needs assessment. For cost effectiveness, go for the minimum quality that will get the job done.

3.2 The GIS Data Capture Process

a) Data Identification: a user needs assessment is carried out in order to identify user information needs. This helps identify the data required for their applications. Subsequently, the appropriate data sources and their location are identified.

b) Analogue data collection: in many cases, a lot of the available data are in the form of analogue maps. These maps are collected and evaluated as to their quality, completeness and complexity. If acceptable, they may then be prepared for digitization. In large projects, where maps will be digitized over a long time, maps should be stored in optimum temperature and humidity conditions while awaiting digitization.



GIS Data Capture Process

c) Analogue Data Preparation: at this stage, the features to be digitized are chosen and feature codes assigned to them. If necessary, they are highlighted for easy recognition. The manuscript may also require reformatting to conform to the digitization method to be used, e.g reduction in size, opaquing, cleaning etc.

d) Digitization and Editing: digitization is the conversion of analogue data to digital by a variety of techniques. During editing, the digital data are displayed, checked and corrected for errors. If no attribute data were entered during digitization, they are entered at this stage.

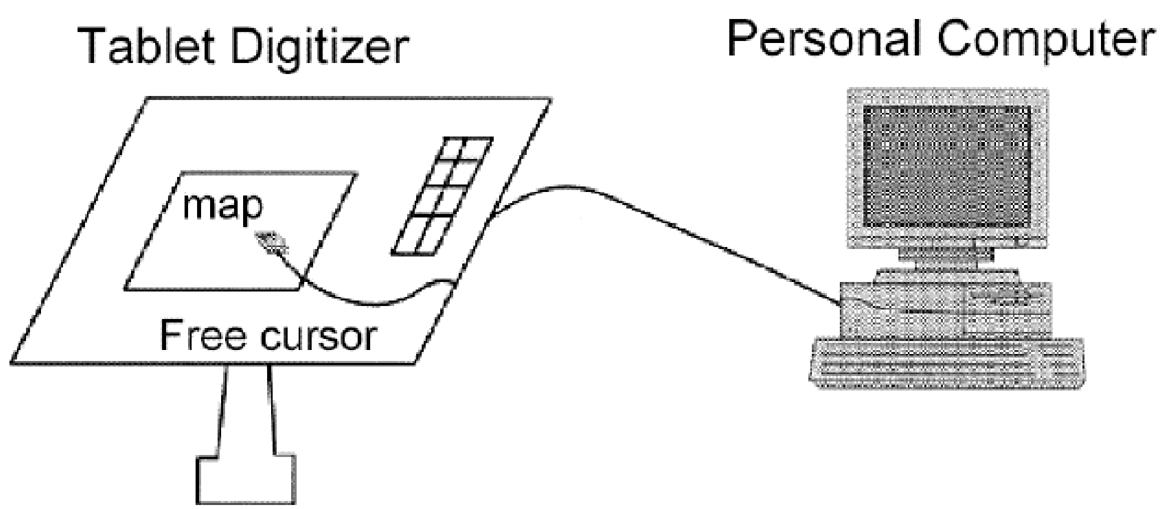
3.3 Methods of Digitization

Vector format: manual digitizing; onscreen digitization

Raster format: manual gridding; scanning; video digitizing

3.3.1 Manual Digitizing

This method uses a manual digitizer which consists of digitizing table and a host computer. The digitizing table consists of a two layer magnetized wire mesh that is sandwiched between a very flat top made of stable material and a metallic or plastic bottom. The electrical connections are such that when the cursor button is pressed at any point on the table surface, electrical impulses are sent to the control unit where they are interpreted versus digital table coordinates of the cursor position relative to a defined origin on the table. The cursor must have at least one button for coordinate registration but usually has more-4, 24 or 16 to server various command and coding functions. A menu box is usually provided on the table. This consists of a number of designated areas on the table where the activation of the cursor sends a particular command e.g.‘map mounting’ or ‘digitizer configuration’. The control unit provides interface to the host computer, while the user terminal enables keyboard data and command entry plus receipt of messages from the system.



Digitizing Software must be provided to enable activation/deactivation of the system, definition of the area to be digitized, map mounting and general control of the digitizing process. Map mounting involves the computation of transformation parameters between the table and map coordinate systems via an ***affine*** transformation.

A problem with manual digitizing is slow progress for dense data e.g. contours Operator stress and fatigue which deteriorates the quality of output. Thus is no longer is use currently

3.3.2 On-Screen Digitizing

Another method of manual digitizing that is common is the ***on-screen digitization***. It entails using a scanned map or image and displaying this on the screen and using a normal mouse, the features of interest are picked and their locations recorded.

3.3.3 Digitization in the Raster Format

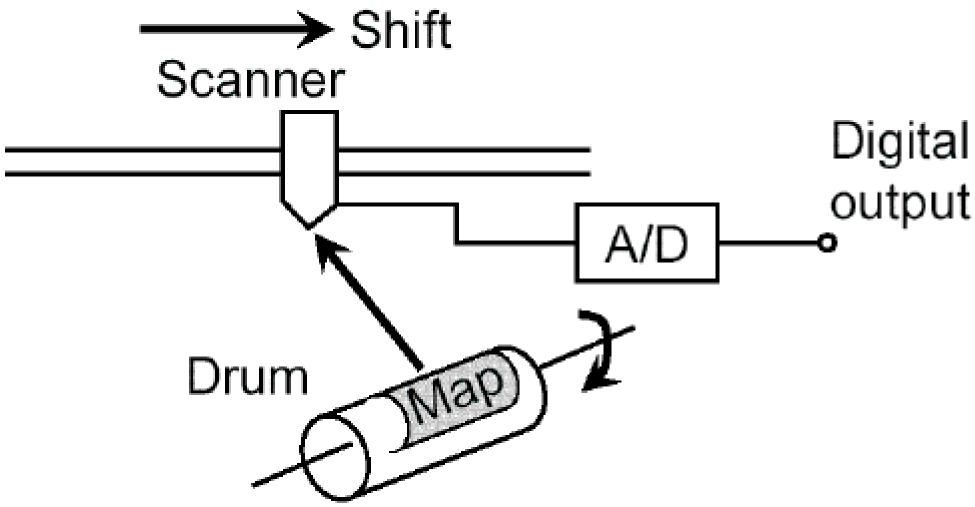
_ Manual gridding;
_ Scanning;
_ Video digitizing.

Only scanning has been covered here

Scanning

It is an automatic method of digitization carried out using a scanner. The scanner senses the binary grey tone or colour values of the analogue data and outputs them as a series of pixels in parallel scan lines. Scanners may be flat bed (document mounted on flat surface) or drum (document mounted on drum surface). In both types, the actual scanning is done by a scanning head which is able to sense reflected light (for opaque documents) or transmitted light (for transparent documents) and to turn the light intensity into a pixel value.

Drum scanners occupy less space and are faster but are more expensive. Before scanning, the document must be well prepared to ensure that line widths are resolvable, line separations exceed pixel sizes and unwanted data are opaque out



Parameters for evaluating a Scanner

a) Scanner resolution: smallest image size sensible by scanner expressed in dpi (dots per inch).
b) The larger dpi the fine the resolution and the slower the scan and greater the data volume and vice versa. Most scanners have a range of resolution of 100-200 dpi.
c) Maximum document size: modern large format scanners are drum.
d) Binary/grey tone/ colour capability: best scanners provide all the three capabilities.
e) Radiometric range: number of grey tone levels-the standard one is 256
f) Geometric accuracy: how much data is distorted by the scanning process?
g) Weight
h) Maximum document thickness: typically 3-5mm
i) Price.

Advantages of Scanning

1. Fast means of digitizing large or dense data formats
2. Process is largely automatic and puts minimum strain on operator
3. Output data can be easily integrated with satellite remotely sensed data.

Disadvantages of Scanning

1. High cost of hardware/software
2. Very intensive manuscript preparation
3. Selective digitizing impossible.

3.4 GIS Data Editing

This is the process of detecting and correcting the errors introduced into data during the data capture process. Editing may be carried out in two modes:

a. Batch mode;
b. Interactive mode.

Batch Mode

In this method, software is used to recognize and correct specific error condition in the whole dataset at ago e.g. find and correct all undershoots and overshoots. The main advantages of this mode are that errors of the sane type are corrected in bulk at speed and repeatedly. Batch programmes can also be left to run during slack period e.g at night.

Interactive Mode

This is a mode in which small portions of the digitized graphic are edited at a time by the issuance of appropriate edit commands.

The advantages of this mode are:

a. It is the only way to correct error conditions that batch mode cannot handle
b. Immediate verification of corrections
c. Enables simultaneous digitizing and editing

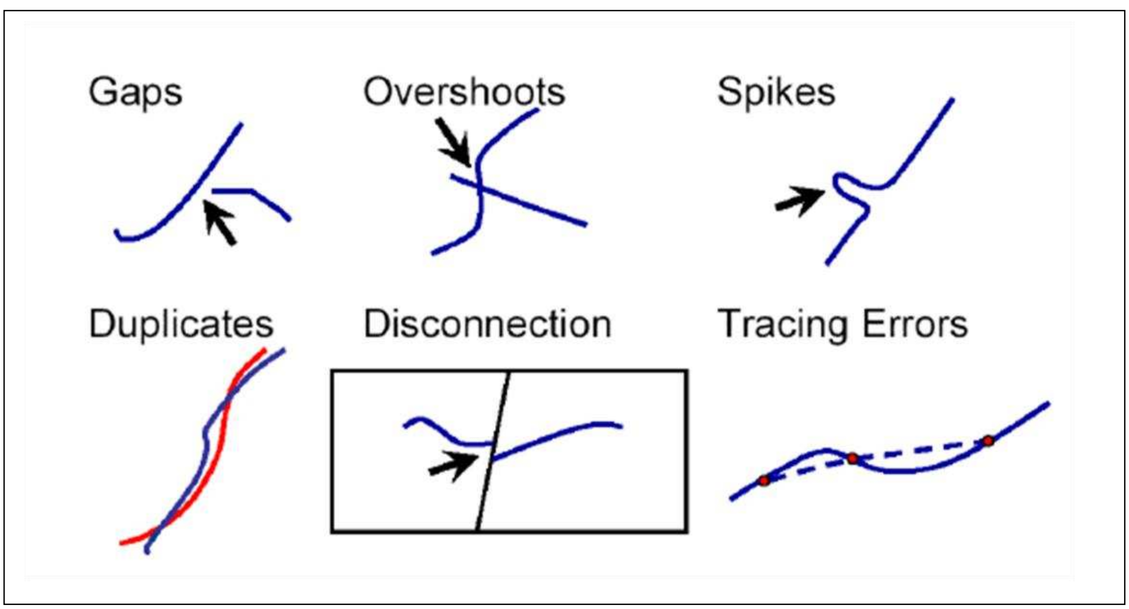
The disadvantage is that it is very slow and repetitive.

Most modern GIS packages enable editing in both batch and interactive modes. A part from the graphics, attributes and relationships may also need editing.

3.5 Common Editing Problems

Vector Data

1) Polygon misclosures;
2) Overshoots/undershoots;
3) Polygon labels;
4) Knots, backtracks, wild lines
5) Slivers and gaps
6) Map sheet combination
7) Line generalization



Raster Data

1) Noise removal: uncharacteristic pixel values due to malfunctioning of system used to collect data or some items that were in the dataset if not required. Removed by digital filtering.
2) Line thinning: making lines to be one pixel wide
3) Gap removal: no pixel values due to malfunctioning of system-average corresponding scan lines
4) Stray pixel removal: clearing pixel falling in wrong locations
5) Re-assigning: giving pixels different values, happens when resampling size of the image

–

- Attribute Data Errors

The identification of attribute data errors is usually not as simple as spatial errors. This is especially true if these errors are attributed to the quality or reliability of the data. Errors as such usually do not surface until later on in the GIS processing. Solutions to these type of problems are much more complex and often do not exist entirely. It is much more difficult to spot errors in attribute data when the values are syntactically good, but incorrect.

Simple errors of linkage, e.g. missing or duplicate records, become evident during the linking operation between spatial and attribute data. Again, most GIS software contains functions that check for and clearly identify problems of linkage during attempted operations. This is also an area of consideration when evaluating GIS software

4.0 GIS SPATIAL DATA ANALYSIS

4.1 Introduction

Geographical analysis allows the study of real-world processes by developing and applying models to illuminate underlying trends in the geographical data and thus make new information available.

A GIS enhances this process by providing tools, which can be combined in meaningful sequences to develop new models. These models may reveal new or previously unidentified relationships within and between data sets, thus increasing our understanding of the real world.

Results of geographical data analysis can be communicated with maps, reports, or both. A map is best used to display geographical relationships whereas a report is most appropriate for summarizing the tabular data and documenting any calculated values. Charts can also be used

4.2 Objectives of GIS spatial analysis

Geographic Information Systems is a spatial analysis tool, which is used in the Spatial Analysis to include a spatial perspective.

- Whereby GIS data description answers the question "where?" GIS data analysis answers the question "why is it there?"
- *GIS Analysis* is the process of ***deriving information from one or more layers*** of spatial data. It can involve multiple steps and processes and it is perhaps the most important capability of a GIS.
- It ***allows discovering relationships between various spatial data*** that might not have been apparent otherwise.
- ***Spatial analysis*** is where all the hard work of digitizing, building a database, checking for errors, and dealing with the details of projections and coordinate systems finally ***pays off in results and better decisions.***

4.3 Scope of Analysis in GIS

Analysis sub-component of GIS software does four important functions:

1) **Selection** is a rather simple operation, but it is important because all subsequent work is based on the results of the selection process.
2) **Manipulation** has to do with aggregation, buffering, overlaying and interpolation.
3) **Exploration** is the first step in discovering any new kind of pattern or cluster in a data set. Explorative spatial data analysis (ESDA) uses the data in an inductive way to get new insight about spatial patterns and relations - "we let the data speak for themselves". {Spatial statistics as Moran's I and the G statistics are important tools in explorative spatial data analysis}.
4) **Confirmation** can be seen as tools for estimation of process models, simulation and forecasting.

Analysis in GIS is different from other statistical analysis in *that the attribute data have established links to maps for visual analysis.* Any statistic we can think of to describe the data then automatically has geographic properties and as a result can be placed on map for visual processing.

4.4 Analysis Procedure

Before starting any analysis, one needs to assess the problem and establish an objective. It is important to think through the process before making any judgments about the data or reaching any decisions; ask questions about the data and model; and generate a step-by-step procedure to monitor the development and outline the overall objective. The following steps outline the basic procedure for geographical analysis:

- Establish the objectives and criteria for the analysis. Define the problem and then identify a sequence of operations to produce meaningful results.
- Prepare the data for spatial operations. Prepare all map coverages for the proposed data analysis. Add one or more attributes to coverages in the database if necessary.
- Perform the spatial operations. Perform the spatial operations and combine the coverages, e.g. creating buffering zones around features, manipulating spatial features and performing polygon overlay.
- Prepare the derived data for tabular analysis. Make sure the feature attribute table contains all the items needed to hold the new values to be created.
- Perform the tabular analysis. Calculation and query the relational database using the model defined in step 1.
- Evaluate and interpret the results. Examine the results and determine whether the answers are valid. Simple map displays and reports can help in this evaluation.
- Refine the analysis if needed and repeat the analysis.

4.5 Types of GIS Analysis

4.5.1 Spatial measurements

GIS makes spatial measurements easy to perform. Spatial measurements can be the distance between two points, the area of a polygon or the length of a line or boundary. Calculations can be of a simple nature, such as measuring areas on one map, or more complex, such as measuring overlapping areas on two or more maps.

4.5.2 Information Retrieval {or selection}

With a GIS we can point at a location, object, or area on the screen and retrieve recorded information about it from the Database Management System (DBMS) which holds the information about the map's features.

In order for a GIS to answer the question "what is where?" we need to carry out retrieval or selection. "Geographic search or selection is the secret to GIS data retrieval" so GIS systems have embedded DBMSs, or link to a commercial DBMS.

a) **Searches by Attribute {or selection by Attribute}**

Most GIS systems include as part of the package a fairly basic relational database manager, or simply built on the exiting capabilities of a database system. All DBMS include functions for basic data display. Searches by attribute are then controlled by the capabilities of database manager.

Find is the basic attribute search. Find is intended to get a single record. Find can be browse or by searches. Examples include show attributes, show records, generate a report, find, recode, select, renumber, sort, compute allows the creation of new attributes based on calculated values, restrict, join, replace; all are examples of data reorganization.

Attribute queries are not very useful for geographic search as they don't or difficult to indicate location; so they just work as humble assistants in our geographical searching needs.

b) **Searches by geography {or spatial selection}**

Spatial selection is to extract specific features based on their location. The form of select used most is buffer operation. **Buffering** is a spatial retrieval around points, lines, or areas based on distance

c) **The query interface {or Query}**

The user must interact with the data in appropriate way, to do that, we need the query interface. Most GIS packages are fully integrated with the WIMP (windows, icons, menus, and pointers) and use the GUI (graphical user interface) of the computer's operating system, such as windows to support both a menu-type query interface and a macro or programming language. And fairly recent trend is that most GISs also contain a language or macro tool for automating repetitive tasks; e.g. ArcView's Avenue, MapInfo's MapBasic, and Arc/Info's AML

SQL (standard query language) has been developed to be a standard interface to relational databases and is supported by many GISs. These use interfaces have specific characteristics.

4.5.3 Spatial overlay

One basic way to create or identify spatial relationships is through the process of spatial overlay. Spatial overlay is accomplished by joining and viewing together separate data sets that share all or part of the same area. The result of this combination is a new data set that identifies the spatial relationships. Spatial overlay allow combining two or more (different) layers and applying the set theoretic operations of intersection, union, difference, and complement

4.5.3 Boundary analysis

Boundary analysis, which is often referred to as districting, helps define regions according to certain criteria. This procedure is used to define area of specific demographic characteristic for example. Since districting is normally an iterative process involving the development of numerous scenarios based on various combinations of desired criteria, the computing power of the GIS proves to be a real timesaver. Rather than struggling with paper maps and adding machines, it is able to interactively define proposed boundaries

4.5.4 Buffer analysis

Buffer analysis is used for identifying areas surrounding geographic features. The process involves generating a buffer around existing geographic features and then identifying or selecting features based on whether they fall inside or outside the boundary of the buffer.

This process is used to identify neighborhood.

4.5.5 Neighborhood Operations

Neighborhood operations can evaluate the characteristics of the area surrounding a specific location: Neighborhood functions operate on the neighbouring features of a given feature or set of features (location (s)).

a) **Search** functions allow the retrieval of features that fall within a given search window (rectangle, circle, or polygon).

b) **Line-in-polygon and point-in-polygon** functions determine whether a given linear or point feature is located within a given polygon, or they report the polygon(s) that a given point or line are contained in.

c) **Topographic functions** compute the slope or aspect from a given digital representation of the terrain (digital terrain model or DTM).

d) **Interpolation functions** predict unknown values using the known values at neighboring locations.

e) **Contour generation** functions calculate contours as a set of lines that connect points with the same attribute value. Examples are points with the same elevation (contours), depth (bathymetric contours), barometric pressure (isobars), or temperature (isothermal lines).

4.5.6. Connectivity Operations

Connectivity functions involve traversing an area and accumulating values: Contiguity measures, Proximity, Network functions, and Visibility functions

a) **Contiguity measures:** evaluate characteristics of spatial units that are contiguous (are connected with unbroken adjacency). An example would be the search for a contiguous piece of forest of a certain area and shape.

b) **Proximity functions:** The best known example of a proximity function is the buffer zone generation (or buffering).

c) **Visibility functions:** are used to compute the points that are visible from a given location (viewshed modeling or viewshed mapping) from a digital terrain model.

d) **Network Analysis:** Network analysis is used for ***identifying the most efficient routes or paths for allocation of services***, and ***for evaluation of it***. Identifying an efficient route or path is finding the shortest or least-cost manner in which to visit a location or a set of locations in a network. GIS can handle complex network problems, such as road network analysis. A GIS can work out travel times and the shortest path from A to B. This facility can be built into more complicated models that might require estimates of travel time, accessibility or impedance along a route system. Network analysis can also be used to ***optimize the allocation of resources***. Such allocation is performed by identifying and creating areas of influence or service zones based on certain criteria. It is accomplished by assigning portions of a network to a location based on impedance.

4.5.8 Digital Terrain Analysis

GIS can build three dimensional models, where the topography of a geographical location can be represented with an x, y, z data model known as Digital Terrain (or Elevation) Model (DTM/DEM). The x and y dimensions of a DTM represent the horizontal plane, and z represent spot heights for the respective x, y coordinates.

Network (TIN):- The data sets derived from a Digital Terrain Model can be used to analyze environmental phenomena or engineering projects that are influenced by elevation, aspect or slope. The visualization (display) power of the computer allows the terrain data to be visualized in three-dimensional form, often from any angle of view (this is known as point-of-view analysis).

5.0 DATA VISUALISATION AND MAP PRODUCTION

5.1 Data Visualization

Data visualization refers to the techniques used to communicate data or information graphically, by encoding it as visual objects (e.g., points, lines or bars) contained in graphics. The goal is to communicate information clearly and efficiently to users. It is one of the steps in spatial data analysis.
The "main goal of data visualization is to communicate information clearly and effectively through graphical means, usually as maps, graphic impressions, charts, etc. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often fail to achieve a balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose — to communicate information"

Data visualization is viewed by many disciplines as a modern equivalent of visual communication. It is not owned by any one field, but rather finds interpretation across many (e.g. it is viewed as a modern branch of descriptive statistics by some, but also as a grounded theory development tool by others). It involves the creation and study of the visual representation of data, meaning "information that has been abstracted in some schematic form, including attributes or variables for the units of information".

A primary goal of data visualization is to communicate information clearly and efficiently to users via the graphics selected, such as tables and charts. Effective visualization helps users in analyzing and reasoning about data and evidence. It makes complex data more accessible, understandable and usable.

Users may have particular analytical tasks, such as making comparisons or understanding causality, and the design principle of the graphic (i.e., showing comparisons or showing causality) follows the task. Tables are generally used where users will look-up a specific measure of a variable, while charts of various types are used to show patterns or relationships in the data for one or more variables.

Data visualization is both an art and a science. The rate at which data is generated has increased, driven by an increasingly information-based economy. Data created by internet activity and an expanding number of sensors in the environment, such as satellites and traffic cameras, are referred to as “Big Data”. Processing, analyzing and communicating this data present a variety of ethical and analytical challenges for data visualization.

5.2 Map Production

At the very basic, student ought to gain skills in data visualization to enable them accomplish the last of the functions of GIS, output. Map layout, cartographic design and symbology is at the core. Understanding of map production for spatial data is very crucial. Map production is the process of arranging map elements on a sheet of paper in a way that, even without many words, the average person can understand what it is all about. Maps are usually produced for presentations and reports. a map has to be effective in communicating spatial information. Maps can be hardcopy or digital (display, pdf, images, web maps)

Design of Maps in GIS

Maps are graphic representations of the real world or part of it, where all some of the features may be exaggerated to show them. It is a GIS visualization technique, in addition to reports, charts and tables. Map elements, which are components of a map, should be well arranged/designed for effective communication. Common elements of a map are the title, map body, legend, north arrow, scale bar, acknowledgement, and map border (see figure 5.1). Other elements that might be added are e.g. a graticule, or name of the map projection (CRS); datum, coordinate system, dates, producers/custodian.

Map body: The map body is, of course, the most important part of the map because it contains the map information. The other elements support the communication process and help the map reader to orientate himself and understand the map topic. For example, the title describes the subject matter and the legend relates map symbols to the mapped data.

Map Title: The map title is very important because it is usually the first thing a reader will look at on a map. It can be compared with a title in a newspaper. It should be short but give the reader a first idea of what the map is about.

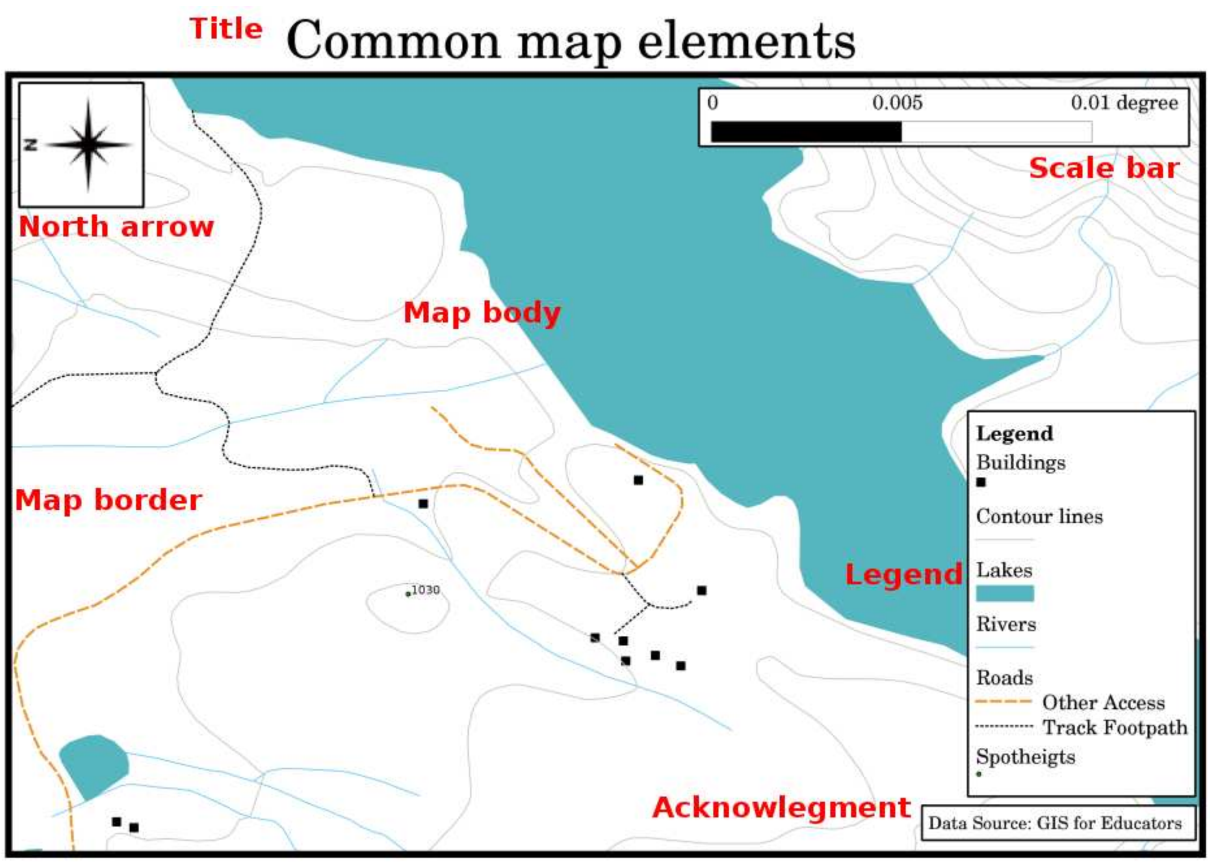


Figure 5.1: Common Map Elements

Map Border: The map border is a line that defines exactly the edges of the area shown on the map. When printing a map with a graticule (which we describe further down), you often find the coordinate information of the graticule lines along the border lines, as you can see in the figure 5.1

***Map Legend*:** A map is a simplified representation of the real world and **map symbols** are used to represent real objects. Without symbols, we wouldn't understand maps. To ensure that a person can correctly read a map, a map legend is used to provide a key to all the symbols used on the map. It contains icons, each of which will represent a type of feature represented in the map.

North arrow: A north arrow (sometimes also called a compass rose) is a figure displaying the main directions, **North**, **South**, **East** and **West**. On a map it is used to indicate the direction of North.

Scale: The scale of a map is the value of a single unit of distance on the map, representing distance in the real world. The values are shown in map units (meters, feet or degrees). The scale can be expressed in several ways, for example, in words, as a ratio or as a graphical scale bar

Scale can be expressed as:

1) Words: e.g. one millimeter on maps represents a thousand meters on the ground.
2) Representative fraction (RF), where map distance and respective ground distance in the real world is represented as a ratio. E.g. 1:25,000 or 1/25,000. 25,000 in the ratio is called the **scale denominator**
3) **Bar scale:** where scale is represented as a graphic. It shows measured distances on the map

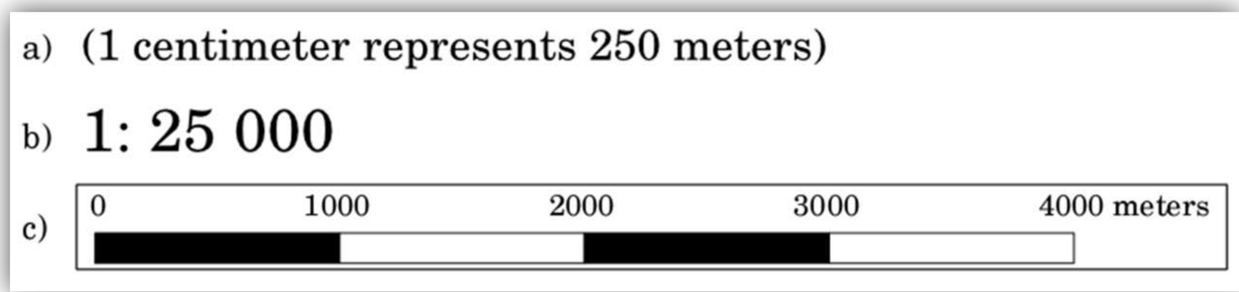


Figure 5.2: A map scale can be expressed in words (a), as a ratio (b), or as graphic or bar scale (c)

Acknowledgment: In the acknowledgment area of a map it is possible to add text with important information. For example information about the quality of the used data can be useful to give the reader an idea about details such as how, by whom and when a map was created.

Graticule and Grids: A graticule is a network of curved lines overlain on a map to make spatial orientation easier for the reader. The lines can be used as a reference. As an example, the lines of a graticule can represent the earth's parallels of latitude and meridians of longitude for geographical coordinate system. . Grids can also be applied for projected coordinate system instead of graticules. Grids are networks of parallel and orthogonal straight lines of Eastings and northings in projected cylindrical projections.

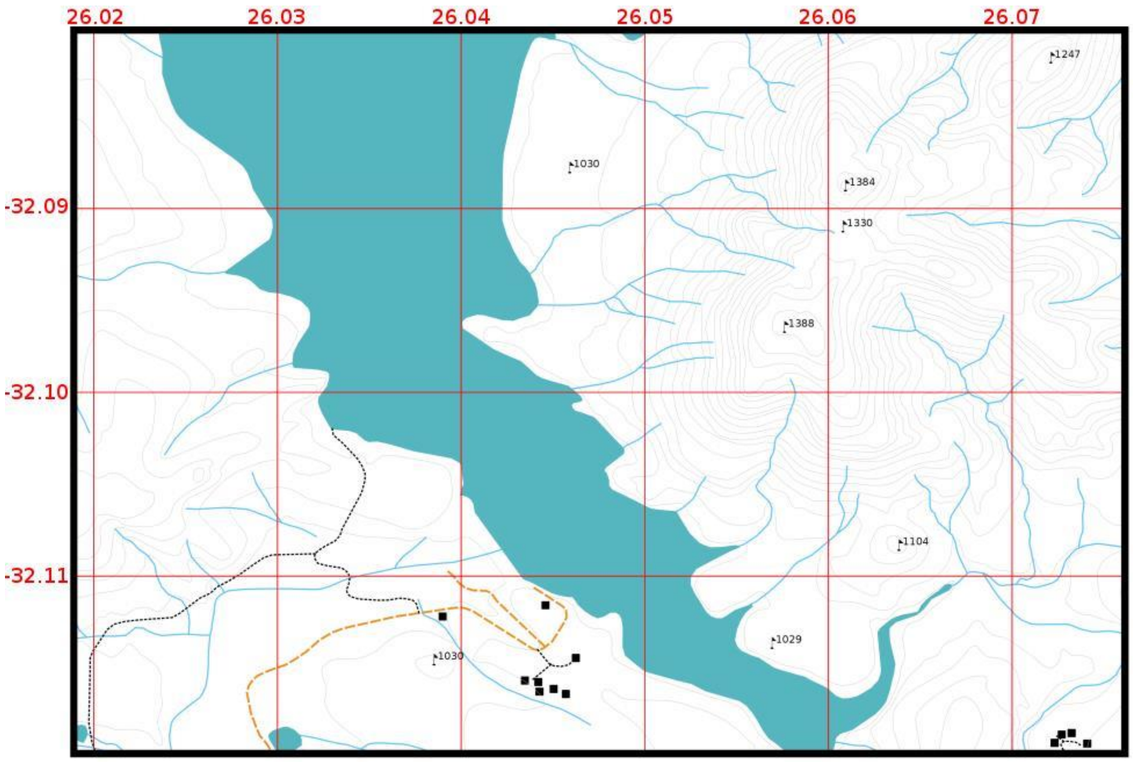


Figure 5.3: Graticule

6.0 SPATIAL REFERENCING

6.1 Overview

The spatial referencing deals with positioning and reference surface for mapping with include: Map Projection Systems, Coordinate Reference Systems and Coordinate Transformations. Satellite Based Positioning will be added because of its increasing importance in positional, mapping and navigation applications.

6.2 Reference Surface for Mapping

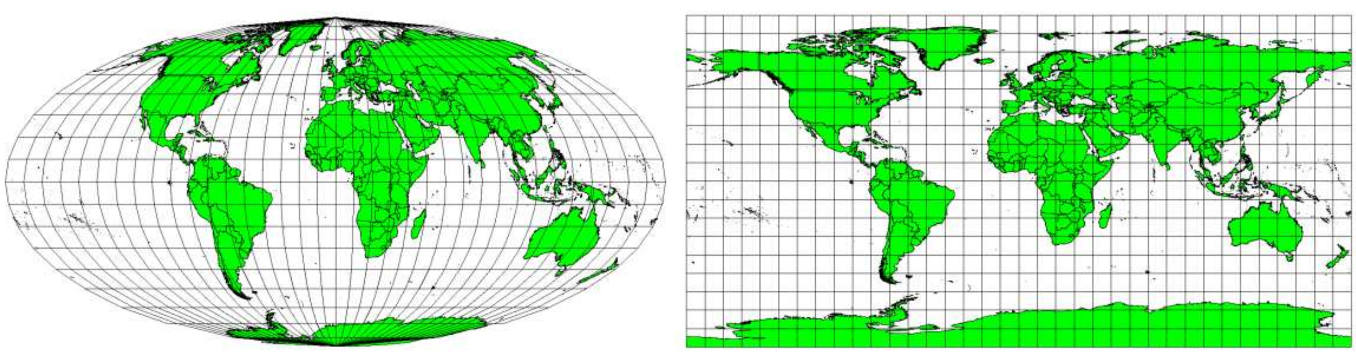
Ideally, reference surface for mapping is either curved for large area or plane for small area. Any area beyond length or breath of 1km is considered large, and may suffer from errors due to earth’s curvature. The earth is curved not flat, and transferring it into plane sheen of peper calls for appropriate map projection system, which is discussed below.

6.3 Map Projection Systems

A map projection tries to represent the curved (3-dimensional) Earth with all its features on a flat sheet of paper. This is very difficult as you can imagine, and even after hundreds of years there is no single projection that is able to represent the Earth perfectly for any area in the world. Every projection has advantages and disadvantages. To be able to create maps as precisely as possible, people have studied, modified, and produced many different kinds of projections. In the end almost every country has developed its own map projection with the goal of improving the map accuracy for their territorial area.

With this in mind, we can now understand why it makes sense to add the name of the projection on a map. It allows the reader to see quickly, if one map can be compared with another. For example, features on a map in a so-called Equal Area projection appear very different to features projected in a Cylindrical Equidistant projection. Map projection is a very complex topic and we cannot cover it completely here. You may want to take a look at our Coordinate Reference Systems if you want to know more about it.

Figure 6.1: Map Projection



The world in different projections. A Mollweide Equal Area projection left, a Plate Carree Equidistant Cylindrical projection on the right.

Map projections try to portray the surface of the earth or a portion of the earth on a flat piece of paper or computer screen. A **coordinate reference system** (CRS) then defines, with the help of coordinates, how the two-dimensional, projected map in your GIS is related to real places on the earth. The decision as to which map projection and coordinate reference system to use, depends on the regional extent of the area you want to work in, on the analysis you want to do and often on the availability of data.

A traditional method of representing the earth's shape is the use of globes. There is, however, a problem with this approach. Although globes preserve the majority of the earth's shape and illustrate the spatial configuration of continent-sized features, they are very difficult to carry in one's pocket. They are also only convenient to use at extremely small scales (e.g. 1:100 million).

Most of the thematic map data commonly used in GIS applications are of considerably larger scale. Typical GIS datasets have scales of 1:250 000 or greater, depending on the level of detail. A globe of this size would be difficult and expensive to produce and even more difficult to carry around. As a result, cartographers have developed a set of techniques called **map projections** designed to show, with reasonable accuracy, the spherical earth in two-dimensions.

When viewed at close range the earth appears to be relatively flat. However when viewed from space, we can see that the earth is relatively spherical. Maps, as we will see in the upcoming map production topic, are representations of reality. They are designed to not only represent features, but also their shape and spatial arrangement. Each map projection has **advantages** and **disadvantages**. The best projection for a map depends on the **scale** of the map, and on the purposes for which it will be used. For example, a projection may have unacceptable distortions if used to map the entire African continent, but may be an excellent choice for a **large-scale (detailed) map** of your country. The properties of a map projection may also influence some of the design features of the map. Some projections are good for small areas, some are good for mapping areas with a large East-West extent, and some are better for mapping areas with a large North-South extent.

The three families of map projections

The process of creating map projections can be visualised by positioning a light source inside a transparent globe on which opaque earth features are placed. Then project the feature outlines onto a two-dimensional flat piece of paper. Different ways of projecting can be produced by surrounding the globe in a **cylindrical** fashion, as a **cone**, or even as a **flat surface**. Each of these methods produces what is called a **map projection family**. Therefore, there is a family of **planar projections**, a family of **cylindrical projections**, and another called **conical projections.**

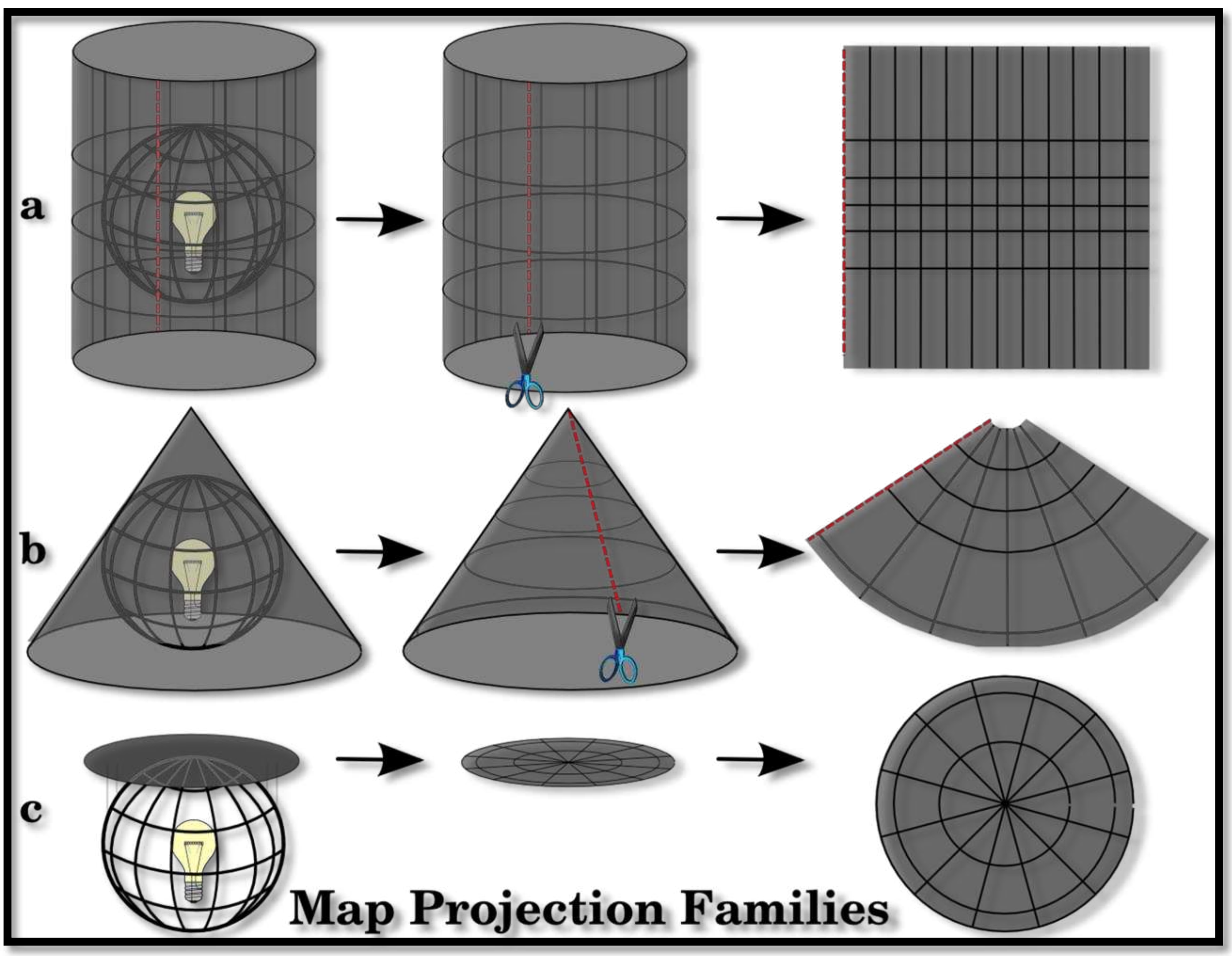
The three families of map projections. They can be represented by a) cylindrical projections, b) conical projections or c) planar projections.

Today, of course, the process of projecting the spherical earth onto a flat piece of paper is done using the mathematical principles of geometry and trigonometry. This recreates the physical projection of light through the globe.

Accuracy of map projections

Map projections are never absolutely accurate representations of the spherical earth. As a result of the map projection process, every map shows **distortions of angular conformity, distance and area**. A map projection may combine several of these characteristics, or may be a compromise that distorts all the properties of area, distance and angular conformity, within some acceptable limit

Figure 6.2: Map Projections



6.4 Coordinate Reference Systems

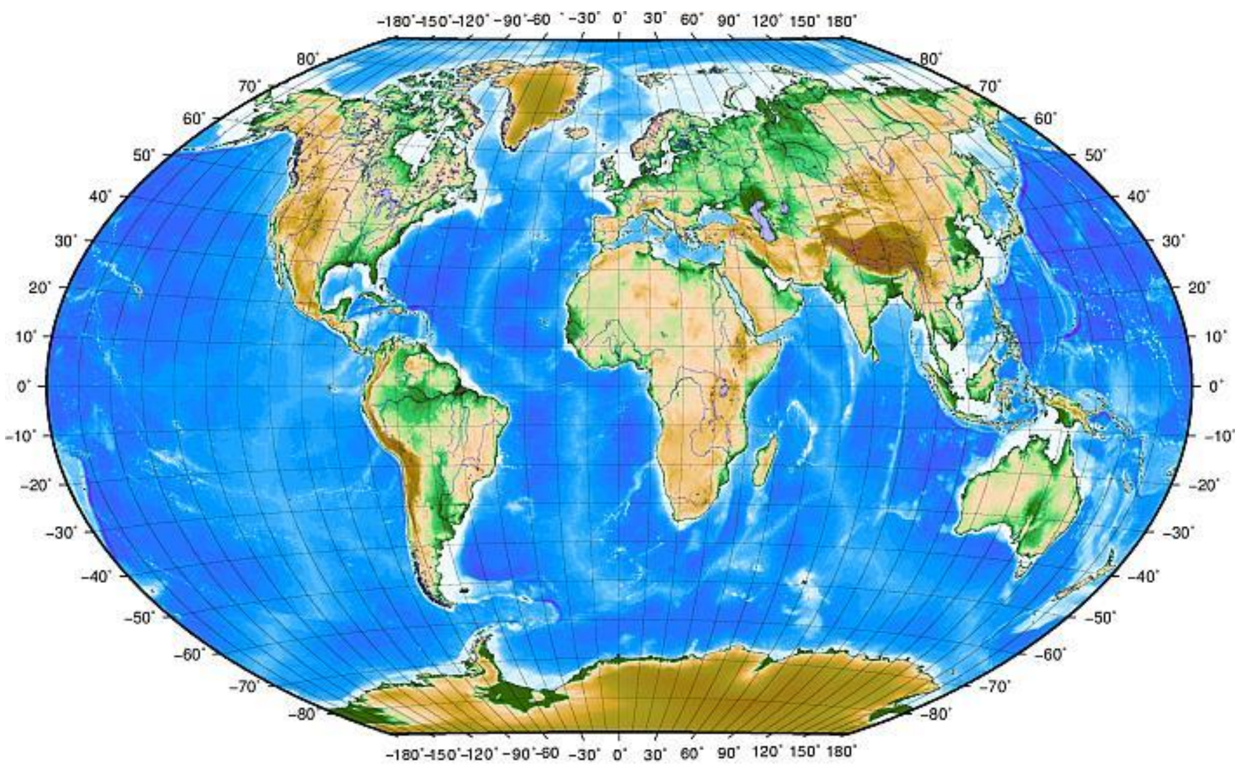
With the help of coordinate reference systems (CRS) every place on the earth can be specified by a set of three numbers, called coordinates. In general CRS can be divided into **projected coordinate reference systems** (also called Cartesian or rectangular coordinate reference systems) and **geographic coordinate reference systems**.

6.4.1 Geographic Coordinate Reference Systems

The use of Geographic Coordinate Reference Systems is very common. They use degrees of latitude and longitude and sometimes also a height value to describe a location on the earth's surface. The most popular is called **WGS 84**.

Lines of latitude run parallel to the equator and divide the earth into 180 equally spaced sections from North to South (or South to North). The reference line for latitude is the equator and each **hemisphere** is divided into ninety sections, each representing one degree of latitude. In the northern hemisphere, degrees of latitude are measured from zero at the equator to ninety at the north pole. In the southern hemisphere, degrees of latitude are measured from zero at the equator to ninety degrees at the south pole. To simplify the digitisation of maps, degrees of latitude in the southern hemisphere are often assigned negative values (0 to -90°). Wherever you are on the earth's surface, the distance between the lines of latitude is the same (60 nautical miles).

Figure 6.3: Geographical Coordinate Reference System



Geographic coordinate system with lines of latitude parallel to the equator and lines of longitude with the prime meridian through Greenwich

Lines of longitude, on the other hand, do not stand up so well to the standard of uniformity. Lines of longitude run perpendicular to the equator and converge at the poles. The reference line for longitude (the prime meridian) runs from the North pole to the South pole through Greenwich, England. Subsequent lines of longitude are measured from zero to 180 degrees East or West of the prime meridian. Note that values West of the prime meridian are assigned negative values for use in digital mapping applications.

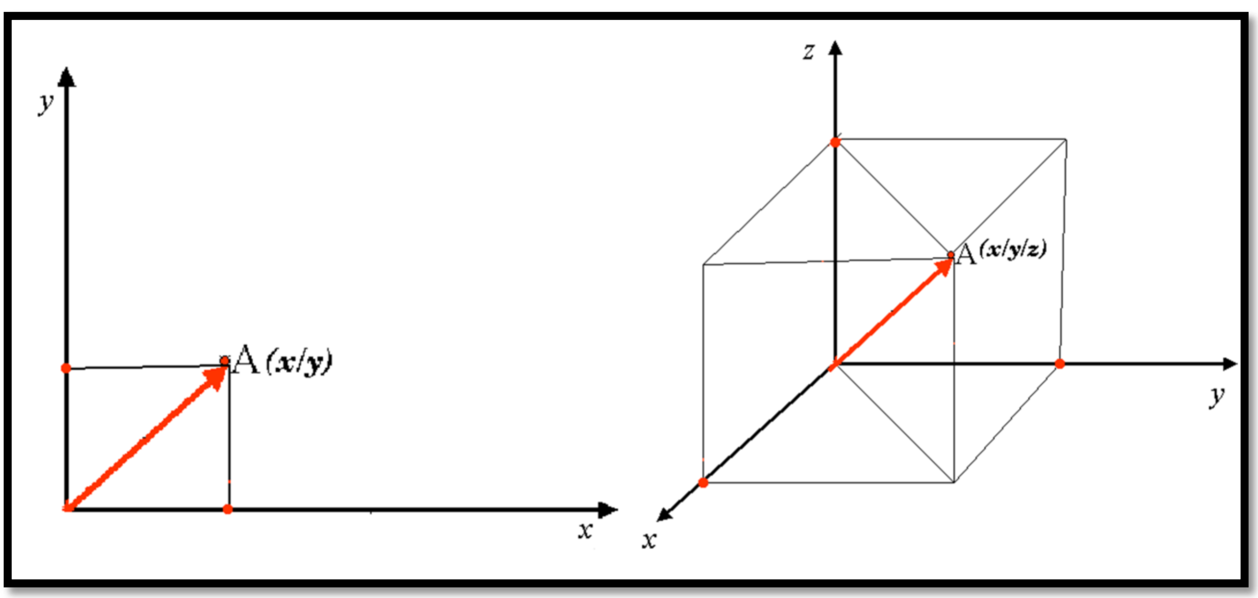
At the equator, and only at the equator, the distance represented by one line of longitude is equal to the distance represented by one degree of latitude. As you move towards the poles, the distance between lines of longitude becomes progressively less, until, at the exact location of the pole, all 360° of longitude are represented by a single point that you could put your finger on. Using the geographic coordinate system, we have a grid of lines dividing the earth into squares that cover approximately 12363.365 square kilometres at the equator — a good start, but not very useful for determining the location of anything within that square.

To be truly useful, a map grid must be divided into small enough sections so that they can be used to describe (with an acceptable level of accuracy) the location of a point on the map. To accomplish this, degrees are divided into **minutes** (') and **seconds** ("). There are sixty minutes in a degree, and sixty seconds in a minute (3600 seconds in a degree). So, at the equator, one second of latitude or longitude = 30.87624 meters

6.4.2 Projected Coordinate Reference Systems

A two-dimensional coordinate reference system is commonly defined by two axes. At right angles to each other, they form a so called **XY-plane** (see figure 6.4). The horizontal axis is normally labelled **X**, and the vertical axis is normally labelled **Y**. In a three-dimensional coordinate reference system, another axis, normally labelled **Z**, is added. It is also at right angles to the **X** and **Y** axes. The **Z** axis provides the third dimension of space. Every point that is expressed in spherical coordinates can be expressed as an **X Y Z** coordinate.

Figure 6.4: Projected CRS



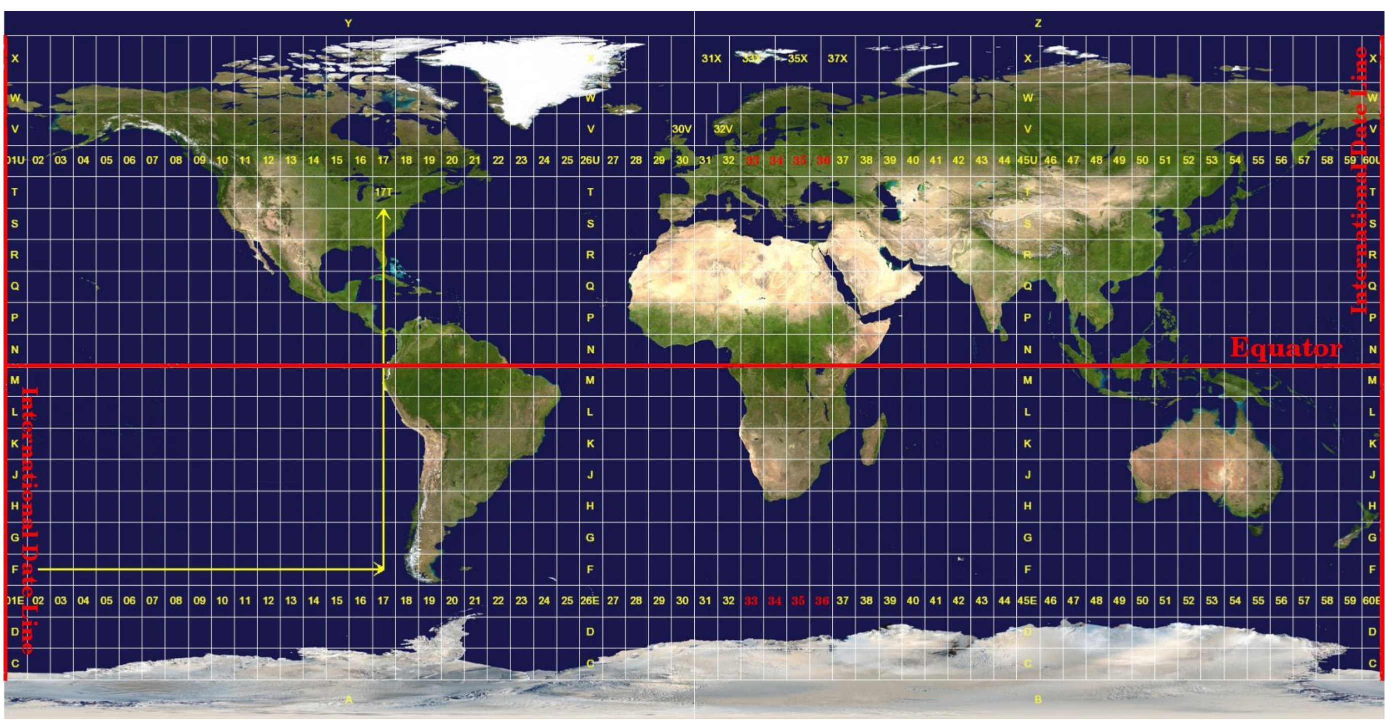
Two and three dimensional coordinate reference systems.

A projected coordinate reference system in the southern hemisphere (south of the equator) normally has its origin on the equator at a specific **Longitude**. This means that the Y-values increase southwards and the X-values increase to the West. In the northern hemisphere (north of the equator) the origin is also the equator at a specific **Longitude**. However, now the Y- values increase northwards and the X-values increase to the East. In the following section, we describe a projected coordinate reference system, called Universal Transverse Mercator (UTM) often used for South Africa.

6.4.3 Universal Transverse Mercator (UTM) CRS

The Universal Transverse Mercator (UTM) coordinate reference system has its origin on the **equator** at a specific **Longitude**. Now the Y-values increase southwards and the X-values increase to the West. The UTM CRS is a global map projection. This means, it is generally used all over the world. But as already described in the section 'accuracy of map projections' above, the larger the area (for example South Africa) the more distortion of angular conformity, distance and area occur. To avoid too much distortion, the world is divided into **60 equal zones** that are all **6 degrees** wide in longitude from East to West. The **UTM zones** are numbered **1 to 60**, starting at the **international date line, also called Greenwich meridian,** (**zone 1** at 180 degrees West longitude) and progressing East back to the **international date line** (**zone 60** at 180 degrees East longitude) as shown in Figure 6.5

Figure 6.5: Universal Transverse Mercator



The Universal Transverse Mercator zones

Normally for each of the 60 zones, comprising 6 degrees each, in practice three parameters are very important

- ⇒ Central Meridian: the longitude which possesses most centrally through the area of interest (AOI). In Kenya, it is normally 36 degrees
- ⇒ False Easting: usually +500,000 meters to make all coordinate figures positive. This happens for all areas left of central meridian, which ordinarily would be negative
- ⇒ False Northing: usually +10,000,000 for all areas in the southern hemisphere, which ordinarily would be negative.

6.4.4 On-The-Fly Projection

As you can probably imagine, there might be a situation where the data you want to use in a GIS are projected in different coordinate reference systems. For example, you might get a vector layer showing the boundaries of South Africa projected in UTM 35S and another vector layer with point information about rainfall provided in the geographic coordinate system WGS 84. In GIS these two vector layers are placed in totally different areas of the map window, because they have different projections.

To solve this problem, many GIS include a functionality called **on-the-fly** projection. It means, that you can **define** a certain projection when you start the GIS and all layers that you then load, no matter what coordinate reference system they have, will be automatically displayed in the projection you defined. This functionality allows you to overlay layers within the map window of your GIS, even though they may be in **different** reference systems.

6.5 Datum Transformations

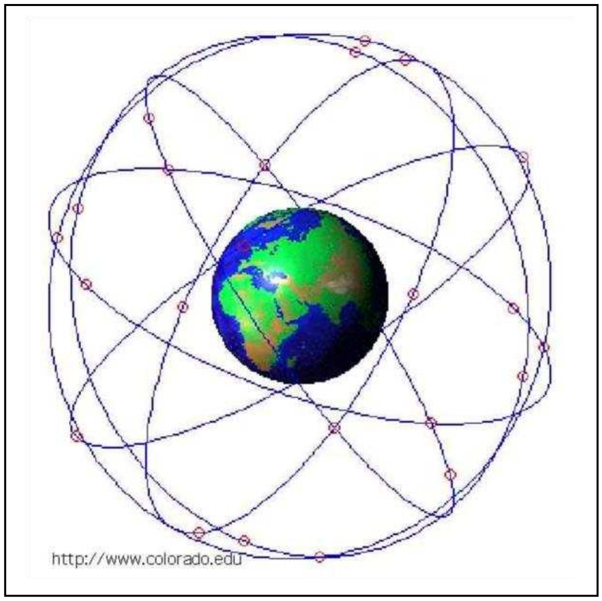
Coordinate Reference Systems, whether geographical or projected, are defined by datum and map projection systems. We have different datum, which are mathematical models representing the earth. Common world datum includes World Geographic Society (WGS) 1984. In Kenya, we have Arc Datum 1960, Clarke 1880, among others. We also have AFREF 2015 for Africa. Coordinates from one datum can be converted or transformed to another datum, whether geographic or projected. GIS software has made this process very easy at a click of a button.

6.6 Satellite Based Positioning

Use of satellite based technology in position, mapping and navigation has gained traction currently. It is a commonly used data capture tool for vector data by use of handheld or geodetic receivers. Common one is Global Positioning system developed by USA. Others are GloNASS developed by Russia and Galileo developed by EU. Their common name is Global Navigation Satellite System (GNSS).

6.6.1 Global Positioning System (GPS)

GPS (global Positioning System) is a system of 21-24 satellites in orbit around the earth.



Each satellite knows its position and uses a unique signal to continuously broadcast this information. Along with the position information is a time signal.

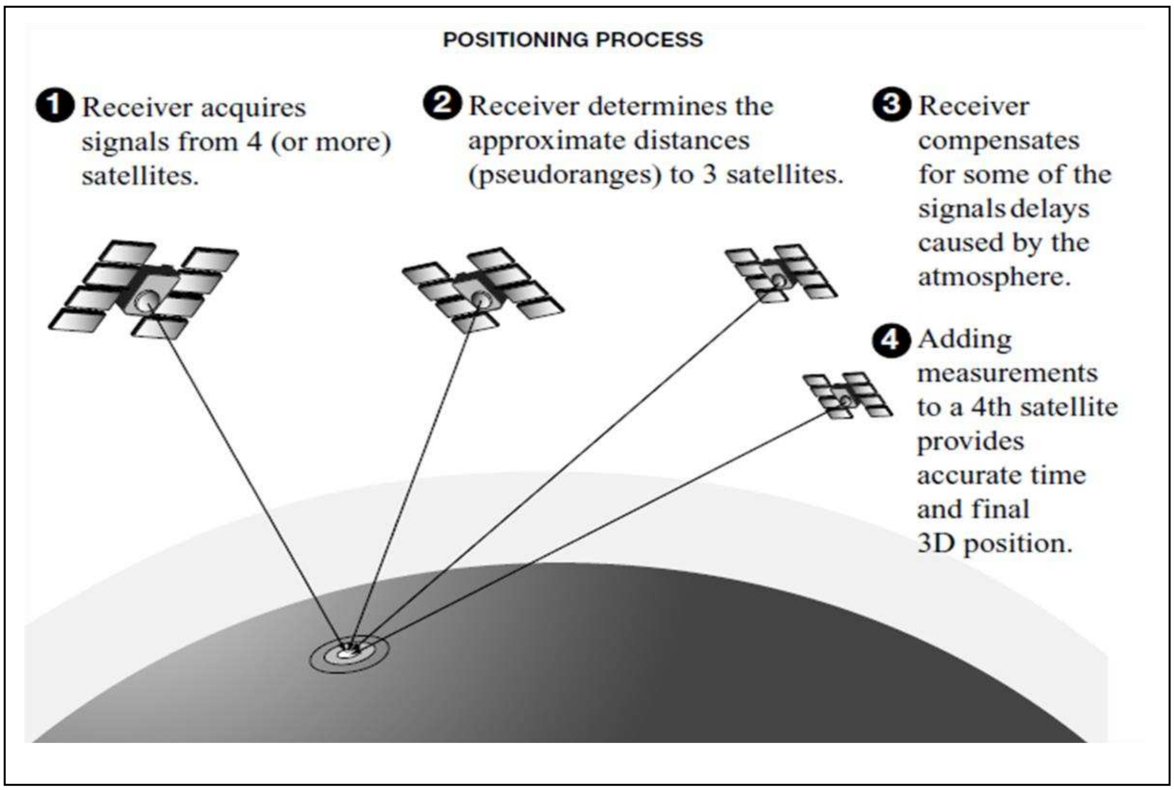
When a GPS receiver receives a signal from at least four (4) satellites it can compute its position by trilateration.

The receiver position can be expressed in degrees of latitude and longitude, or distance (meters) using Universal Transverse Mercator (UTM) coordinates.

GPS Determination of Positions

To compute your exact position, your GPS receiver determines the distance to each of several satellites by:

■ computing exactly where each satellite is in space,
■ measuring the travel time from there to here of radio signals broadcast by the satellites, and
■ accounting for delays the signals experience as they travel through the earth's atmosphere.



7.0 GIS SOFTWARE OVERVIEW

7.1 Main Players in GIS Softaware

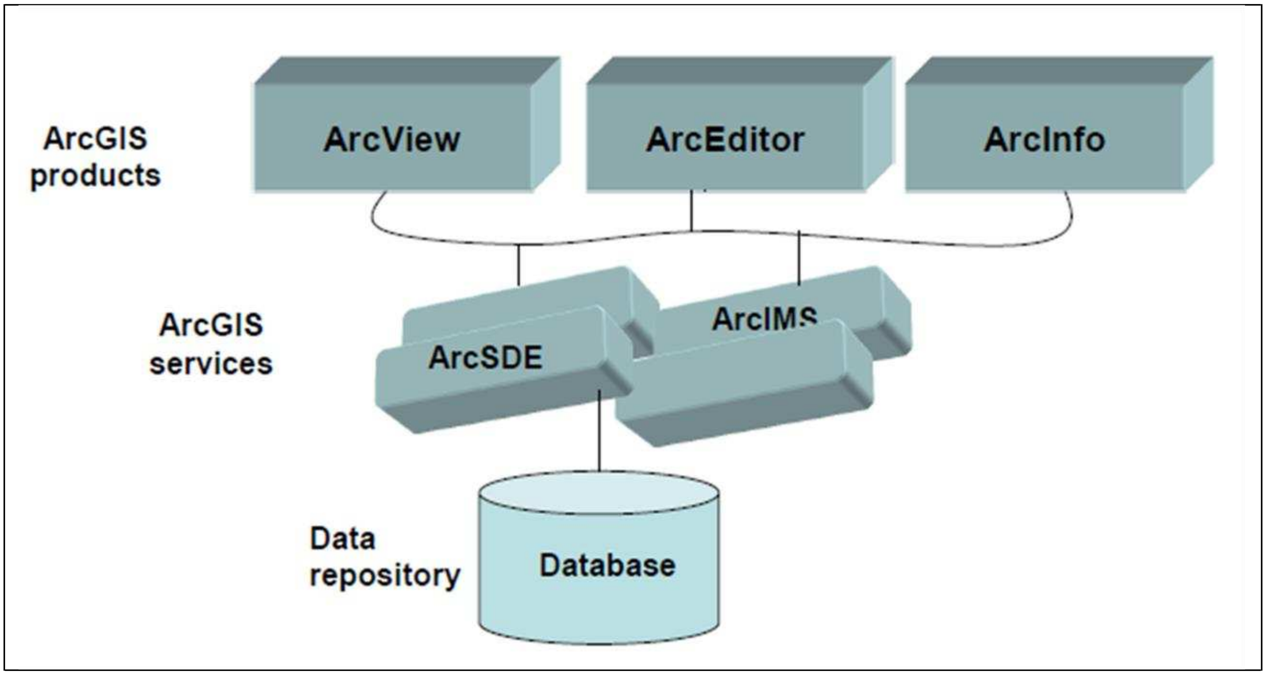
- **ESRI, Inc.,** Redlands, CA
  - clear market leader with about a third of the market
  - originated commercial GIS with their ArcInfo product in 1981
  - privately owned by Jack Dangermond, a legend in the field
  - Strong in gov., education, utilities and business logistics
- **MapInfo, Troy N.Y.**
  - Aggressive newcomer in early 1990s, but now well-established.
  - Strong presence in business, especially site selection & marketing, and telecom
- **Intergraph** (Huntsville, AL)
  - origins in proprietary CAD hardware/software
  - Older UNIX-based **MGE** (Modular GIS Environment) evolved from CAD
  - Current **GeoMedia** was the first true MS Windows-based GIS
  - strong in design, public works, and FM (facilities management), but weakening
- **Bentley Systems** (Exton, PA)
  - **MicroStation GeoGraphics,** originally developed with Intergraph, is now their exclusive and main product..
  - Strong in engineering; advertises itself as "geoengineering"
- **Autodesk** (San Rafael, CA)
  - Began as PC-based CAD, but now the dominant CAD supplier
  - First GIS product **AutoCAD *Map*** introduced in 1996
  - Primarily small business/small city customer base

QGIS Software is the latest open-source software that can be freely downloaded and used

7.2 ARCGIS Overview

ArcGIS is commercial GIS software by ESRI Inc, USA.

It comprised: 3 products, 3 application modules, ArcGIS Services, Database, extensions and ArcObjects



a) **Three products**: ArcView, ArcEditor and ArcInfo in the order of increasing functionality

Available capabilities within these modules are "tiered" in three levels

- **ArcView:** viewing, map production, spatial analysis, basic editing:
- **ArcEditor:** ArcView, plus specialized editing:
- **ArcInfo:** ArcView & ArcEditor plus special analyses and conversions:

b) It has **three applications modules**: ArcMAP, ArcCatalog and ArcToolbox

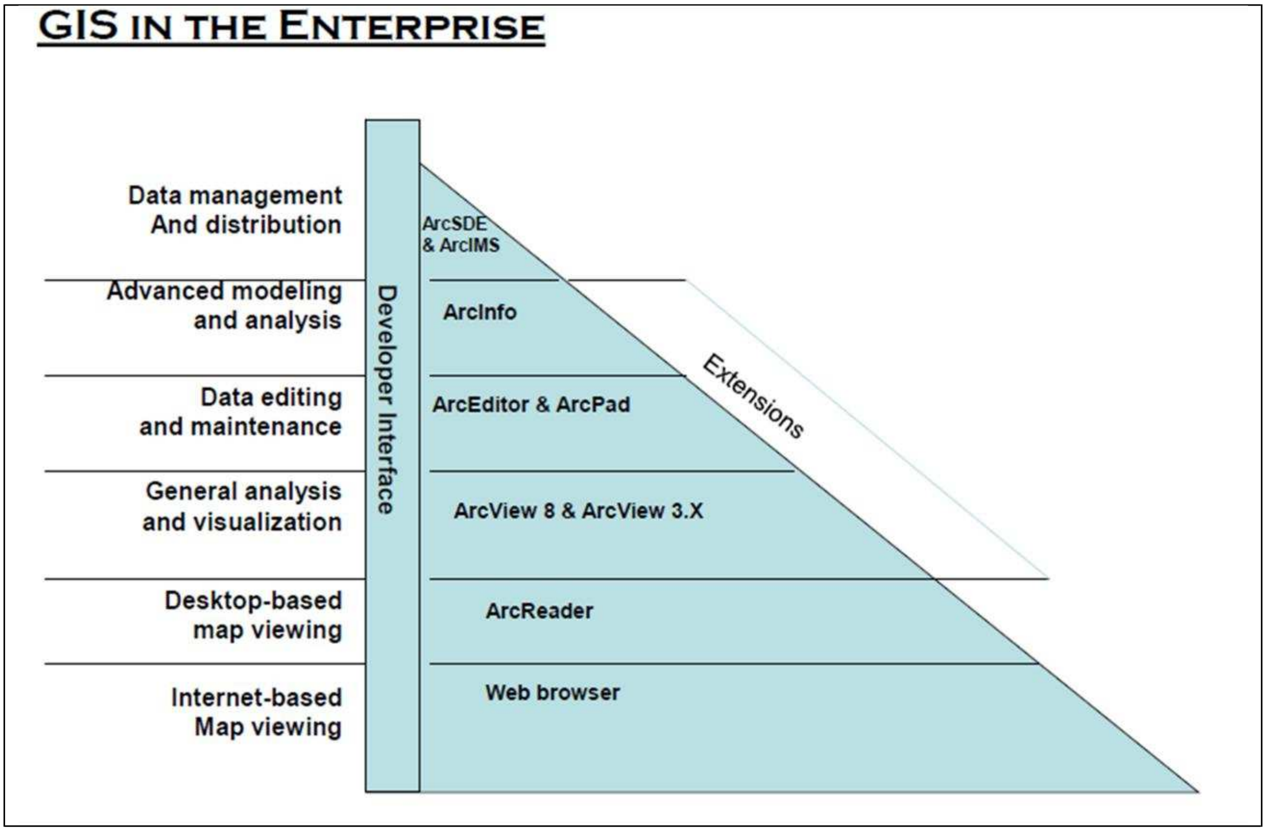
1. **ArcMap:** for data display, map production, spatial analysis, data editing
2. **ArcCatalog:** for data management and preview
3. **ArcToolbox**, for specialized data conversions and analyses, available as a window in both

c) **Extensions:** for special apps.: *Spatial Analyst, 3D Analyst, Geostatistics, Business Analyst, etc.*

d) **ArcObjects:** to build specialized capabilities within ArcMap or ArcCatalog using VB for Applications

e) **ArcGIS Services: ArcIMS** for internet mapping and **ArCSDE** for database management

f) **Others:** ArcScene for 3-D modelling; ArcGlobe for global mapping



ArcMap Interface

ArcMap does Primary display applications and •Perform map-based tasks like: Displaying, –Editing, Querying, Analyzing, Charting and –reporting. ArcMap interface is as seen below

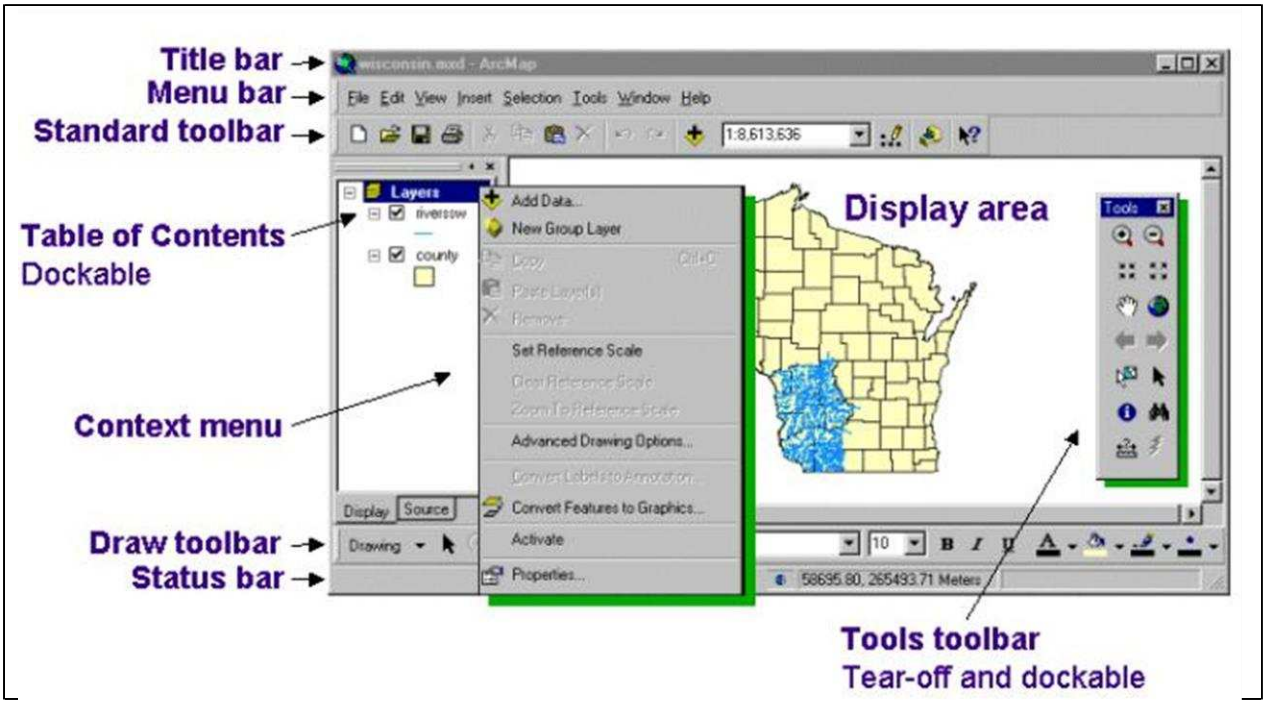
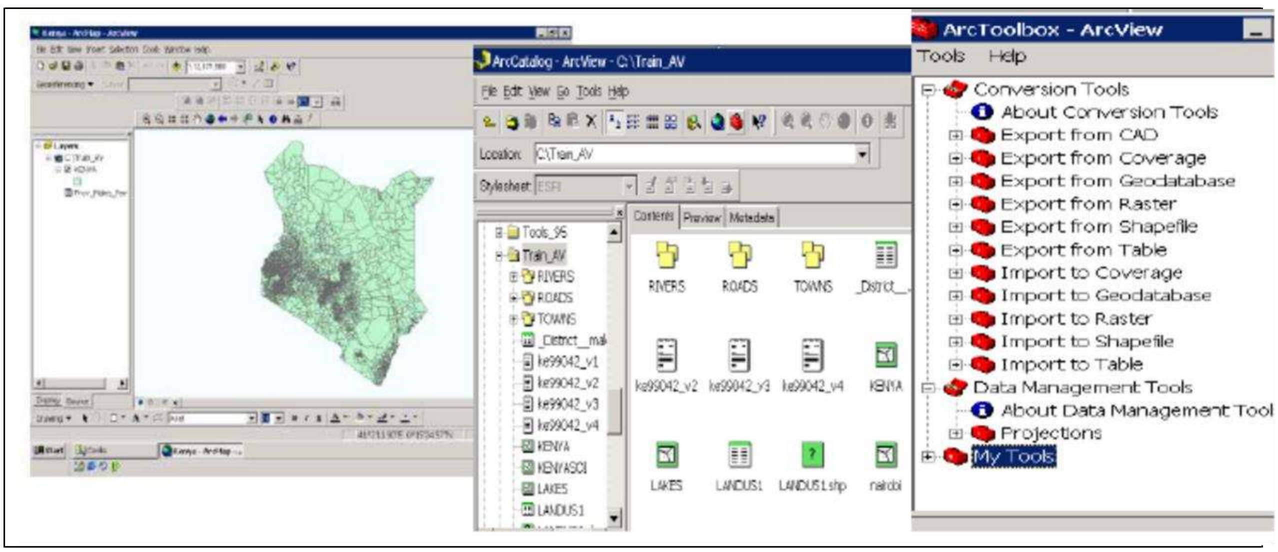


Fig of ArcMAP, ArcCatalog and ArcToolbox



ArcCatalog: is a window into GIS database used for Browse your data, Manage your data and Create and view data documents (metadata)

ArcToolBox: •Geographic processing functions like Data management, ,analysis, and conversion. –Tools vary between ArcGIS products

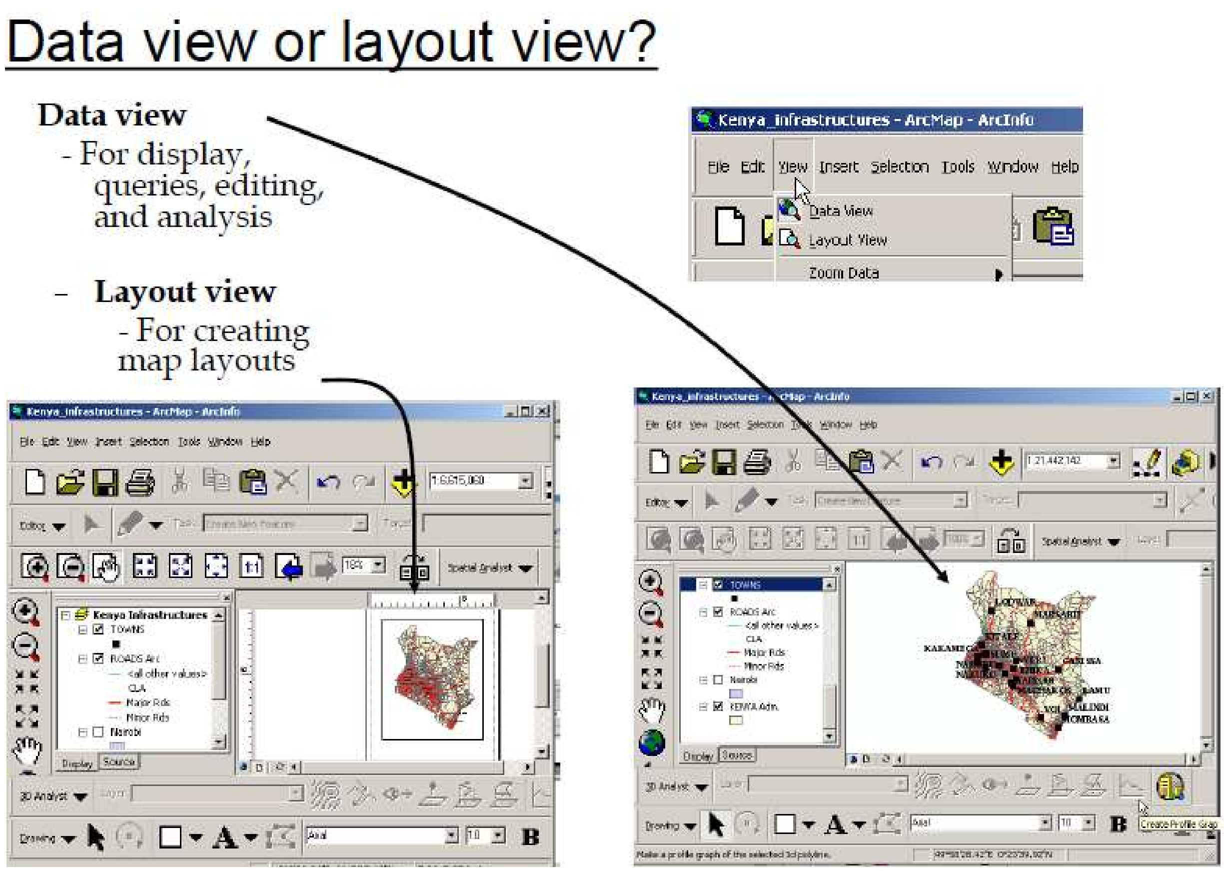
Practical exercises with ArcGIS Desktop

A) Demonstration

- •Observe ArcCatalog
  - –Directory structure and navigation
  - –Documentation
- •Observe ArcMap
  - –Add data
  - –Display and query data
  - –Zoom and pan tools
  - –Save a map document
- •Observe ArcToobox
  - –toolkits

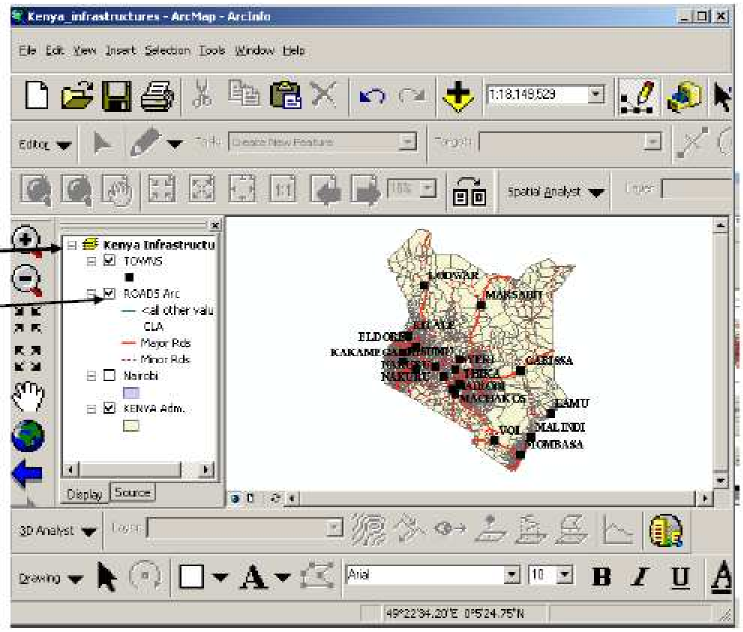
B) Observe and appreciate Data View and Layout View

C) View layers, data frame and maps



Layers, data frames, and maps

- Layer
  - Represent symbolized spatial data



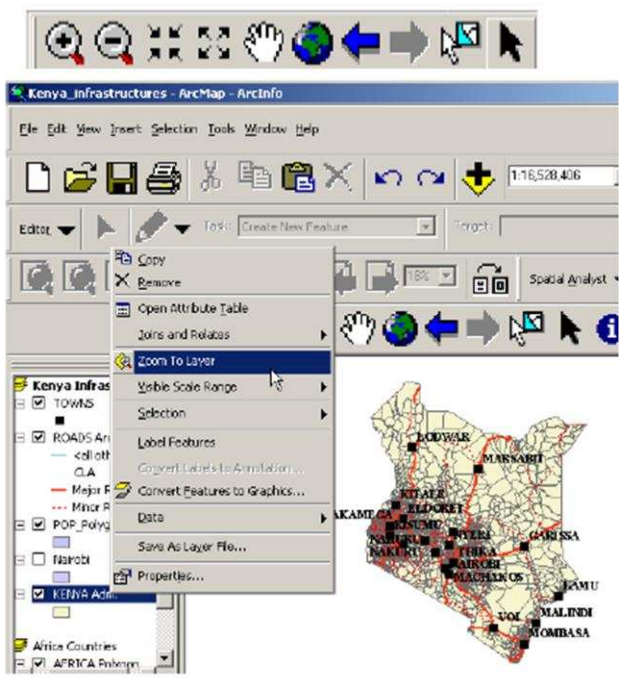
- Data frame
  - Organizes layers

- Map contains
  - Data frames
  - layers
  - Map elements

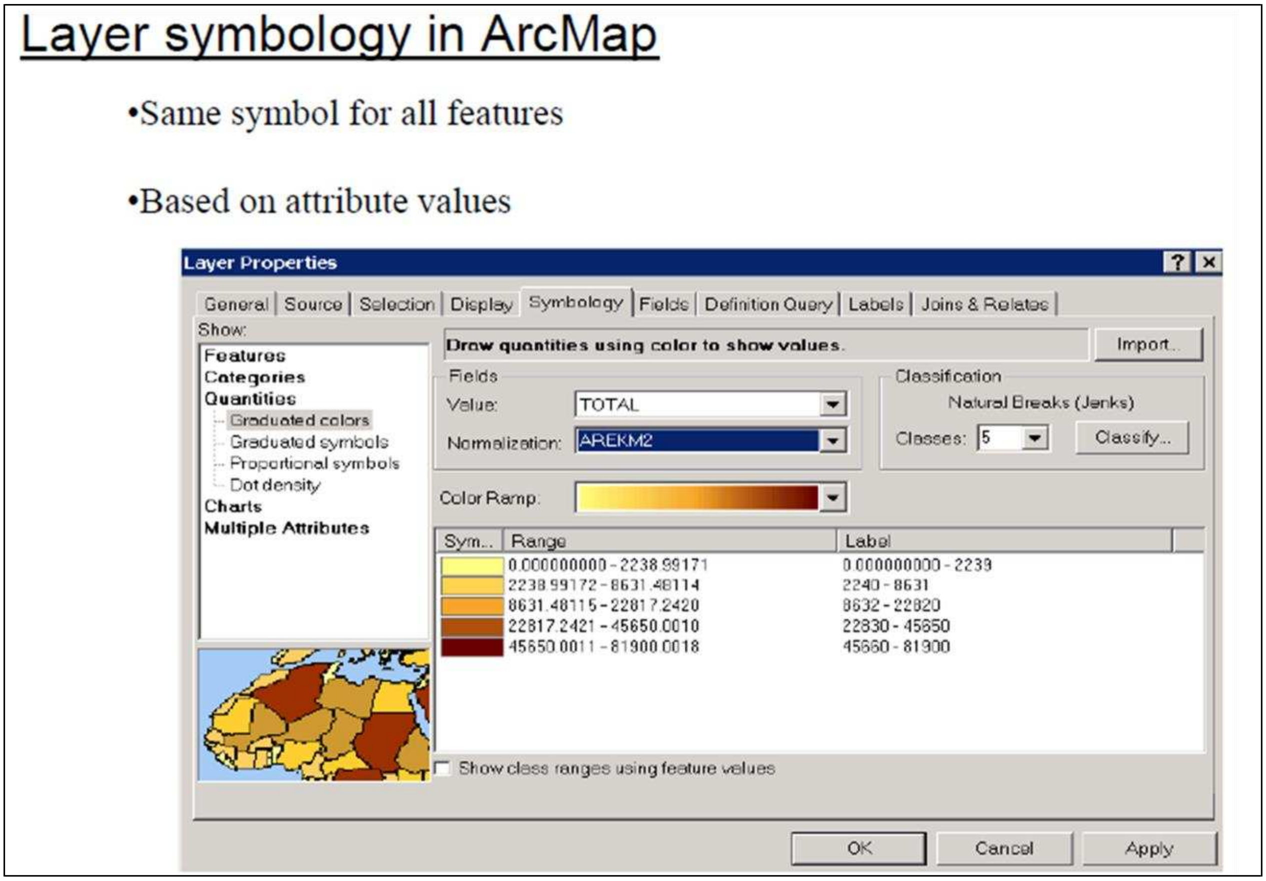
Moving around the map

- Zoom in or out
- Pan the display
- Full extents
- Back or forward one display

- Zoom to a layer



D) Move around the maps through zooming and panning



E) Symbolize GIS datasets

F) Conduct selection/searches by attribute, Query analysis, and Buffering and proximity analysis

G) Calculating summary statistics

7.3 QGIS Overview

QGIS in full is Quantum GIS. A Free and Open Source Geographic Information System for desktop GIS. It can be downloaded from **www.qgis.org**. It can Create, edit, visualise, analyse and publish geospatial information on Windows, Mac, Linux. The current version is 2.18. It does all that ArcGIS does. It doesn't have different products and applications though. It is all ensuite.

Why QGIS

Here are only some of the reasons:

- *It's free.* Installing and using the QGIS program costs you a grand total of zero money. No initial fee, no recurring fee, nothing.
- *It's free, as in liberty.* If you need extra functionality in QGIS, you can do more than just hope it will be included in the next release. You can sponsor the development of a feature, or add it yourself if you are familiar with programming.
- *It's constantly developing.* Because anyone can add new features and improve on existing ones, QGIS never stagnates. The development of a new tool can happen as quickly as you need it to.
- *Extensive help and documentation is available.* If you're stuck with anything, you can turn to the extensive documentation, your fellow QGIS users, or even the developers.
- *Cross-platform.* QGIS can be installed on MacOS, Windows and Linux.

Note:

Students are required to install QGIS in their laptops for GIS exercises. QGIS manual will be shared, but can also be downloaded

8.0 REFFERENCES

1) Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
2) David J. Buckley (1997). *The GIS Primer- An Introduction to Geographic Information Systems.* Fort Collins, Colorado, February 1997. http://bgis.sanbi.org/rootfiles/Introduction_to_GIS.pdf
3) DeMers, Michael N. (2005). Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195
4) Principles of Geographic Information Systems: An Introductory Textbook. Otto Huisman and Rolf A. de By (Eds). ITC, Enschede, The Netherlands, 2009. www.itc.nl/library/papers_2009/general/Principles**GIS**.pdf
5) Ormsby, et. al, *Getting to Know ArcGIS Desktop* 2nd Ed. (ESRI Press, 2004)
6) Wolfgang Kainz (2004). Geographic Information Science (GIS). Cartography And Geoinformation Department Of Geography And Regional Research, University Of Vienna
7) RCMRD Training Manuals for GIS Exercises and Presentations
8) Website Resources

COURSE OUTLINE

COURSE CODE**: BIT 2307** COURSE TITLE**: GEOGRAPHICAL INFORMATION SYSTEMS**

a) Course Objective:

The course generally introduces basics Geographical Information to learners. By the end of the course learners should be able to:

- Familiarize with definition, history, basics functions and applications of GIS
- Understand GIS data, data capture, store, manipulation, query, analysis, display and visualization (output)
- Carry out basic GIS analysis and visualization in a GIS software
- Gain basic hands-on experience in using at least one GIS software (ArcGIS 10.1 or QGIS).
- Improve communication skill, leadership and cooperative attitudes of work with others during group practical exercises in GIS lab.

b) Mode of Course Delivery

The module will be delivered through lectures, presentations, tutorials, demonstrations, Q & As and practical exercises in GIS laboratory using ArcGIS 10.1 in small managed groups. Students will be encouraged to actively undertake self-study and field exercises.

c) Mode of Course Evaluation

Components of Assessments	Methods	Weighting
Coursework	CAT and Practical Assignments	30%
Exam	Written Exams	70%

d) Course Content

1) **Introduction to GIS: Concept (**Definition, History & Evolution, GIS Components, GIS Functions, GIS data organization, Advantages, Applications and Users)

2) **GIS Data Types and Models:** Models and Representations of the Real World; GIS Data Models/Formats, Representations; Vector and raster data Structure; Attribute & Relational data Models; Topological data models

3) **Data Capture and Database creation:** Concept of data capture, Data sources; Entry (Direct, Indirect, Other Sources); data capture Process; and Database creation and management
4) **GIS Data analysis:** Retrieval (selection), classification and Measurements; Overlays Analysis; Neighborhood Functions (proximity analysis and Buffering, Diffusion, Flow and raster Surface Analysis); Network Analysis
5) **Data Visualization and Map Production:** GIS and Maps, Visualization Process, Visualization Strategies, Map Production (Map elements and design), Map dissemination (Hardcopy or Web mapping).
6) **Spatial Referencing:** Reference Surface for Mapping; Coordinate systems; Map projections; and Coordinate Transformations; Satellite Based Positioning
7) **Lab Practical:** Using ArcGIS Software (or QGIS) to demonstrate GIS data, data format/representation, data capture & editing, database creation, analysis, visualizations and map production; and integration of GPS and Remote Sensing data.

e) References

1) Principles of Geographic Information Systems: An Introductory Textbook. Otto Huisman and Rolf A. de By (Eds). ITC, Enschede, The Netherlands, 2009. www.itc.nl/library/papers_2009/general/Principles**GIS**.pdf
2) David J. Buckley (1997). The GIS Primer- An Introduction to Geographic Information Systems. Fort Collins, Colorado, February 1997.
3) Longley, Goodchild, Maguire, Rhind (2005). *Geographic Information Systems and Science* 2nd Ed. Wiley,
4) Wolfgang Kainz (2004). Geographic Information Science (GIS). Cartography And Geoinformation Department Of Geography And Regional Research, University Of Vienna
5) Michael Schmandt. *GIS Commons: An Introductory Textbook on Geographic Information Systems-E-text.* http://giscommons.org/
6) Lecture Notes, Presentations, & Demonstrations; and Website Resources