GIS BASICS

EG 2105 GE

Unit 1: Database Management System

Public Namuna Secondary School

Lecture by

Er. Keshav Raj Bhusal

Table of contents

1.1 Introduction to Database Management System

- Data, information & Knowledge
- Databases and databases management system (DBMS)
- Component of database management system
- Define: tables, form, Query, relationship, reports
- Various DBMS softwares

1.2 Logical Data concept and Relationships

- Logical data concept :entities, data value, field/ attribute, records and relationships
- Types of relationships (one to one, one to many, many to many)
- Tables and field data types
- Primary key, candidate key and foreign key

Table of contents

1.3 Data models and DBMS applications

- Relational Data Model & types
- Importance and use of Database Management System (DBMS)
- Benefits of DBMS compared to file system

Data

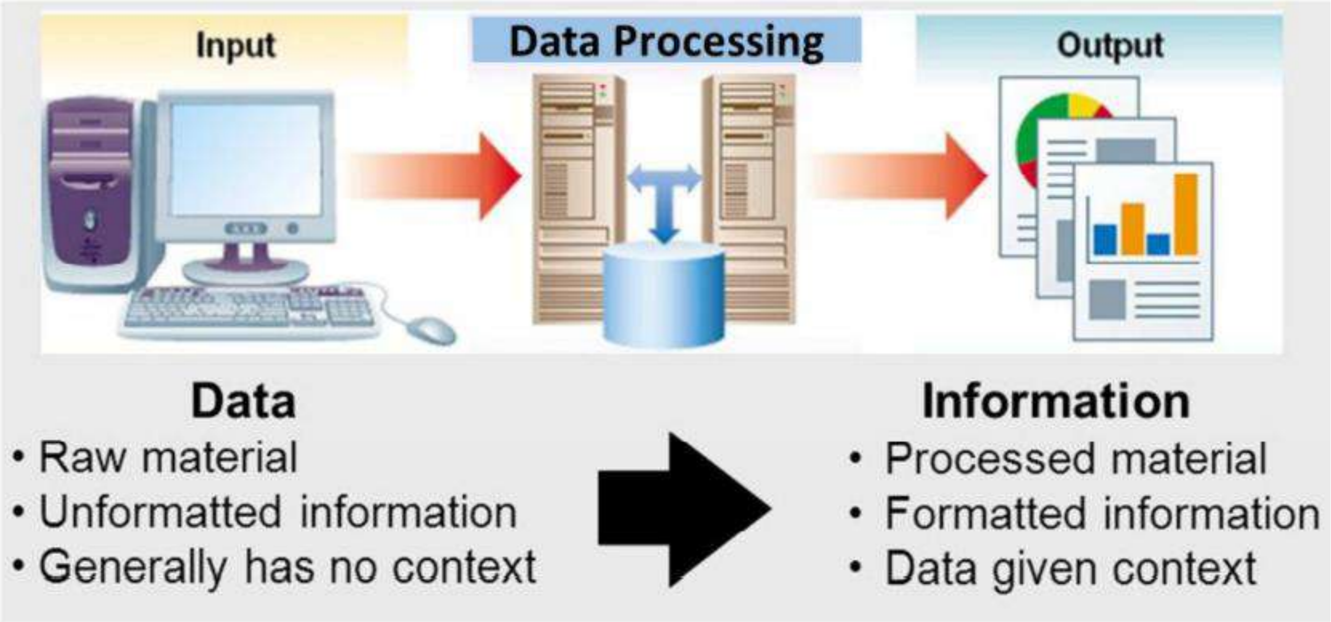
- Data is raw, unorganized facts that need to be interpreted to have meaning.
- Data are values from which the information or meaning is extracted by the users and used in decision making purposes.
- Observation and recording are done to produce data.
- Data by itself is not significant
- Data doesn’t depends on information.

- For example, a collection of numbers (e.g., 25, 30, 27, 22) without any context or labeling is data.

Information

- Information is the processed data. When data is processed, organized, structured or presented in a given context so as to make it useful, it is called information.
- Processing of data could be in the form of addition, subtracting, comparison, sorting, rearrangement etc.
- Information adds meaning to the data and makes the data more reliable and applicable.
- Analysis of data are done to obtain information.
- Information is significant.
- Information depends on data.
- For example, temperature readings for four days: 25°C, 30°C, 27°C, 22°C"), is information that provides a clearer understanding of the data.

Data vs Information



Source: https://portal.abuad.edu.ng/lecturer/documents/1554208765DATA_AND_INFORMATION.pdf

Knowledge

- It is the understanding, insights, and awareness gained from the processed information.
- Knowledge represents the ability to connect information to real-world experiences, concepts, and principles.
- For example, with the information from the table below, knowledge might involve recognizing that Sarah performed the best on the test, while John scored the lowest, and that there might be a need to provide additional support to John.

Student	Score
John	5
Alice	10
Mike	15
Sarah	20

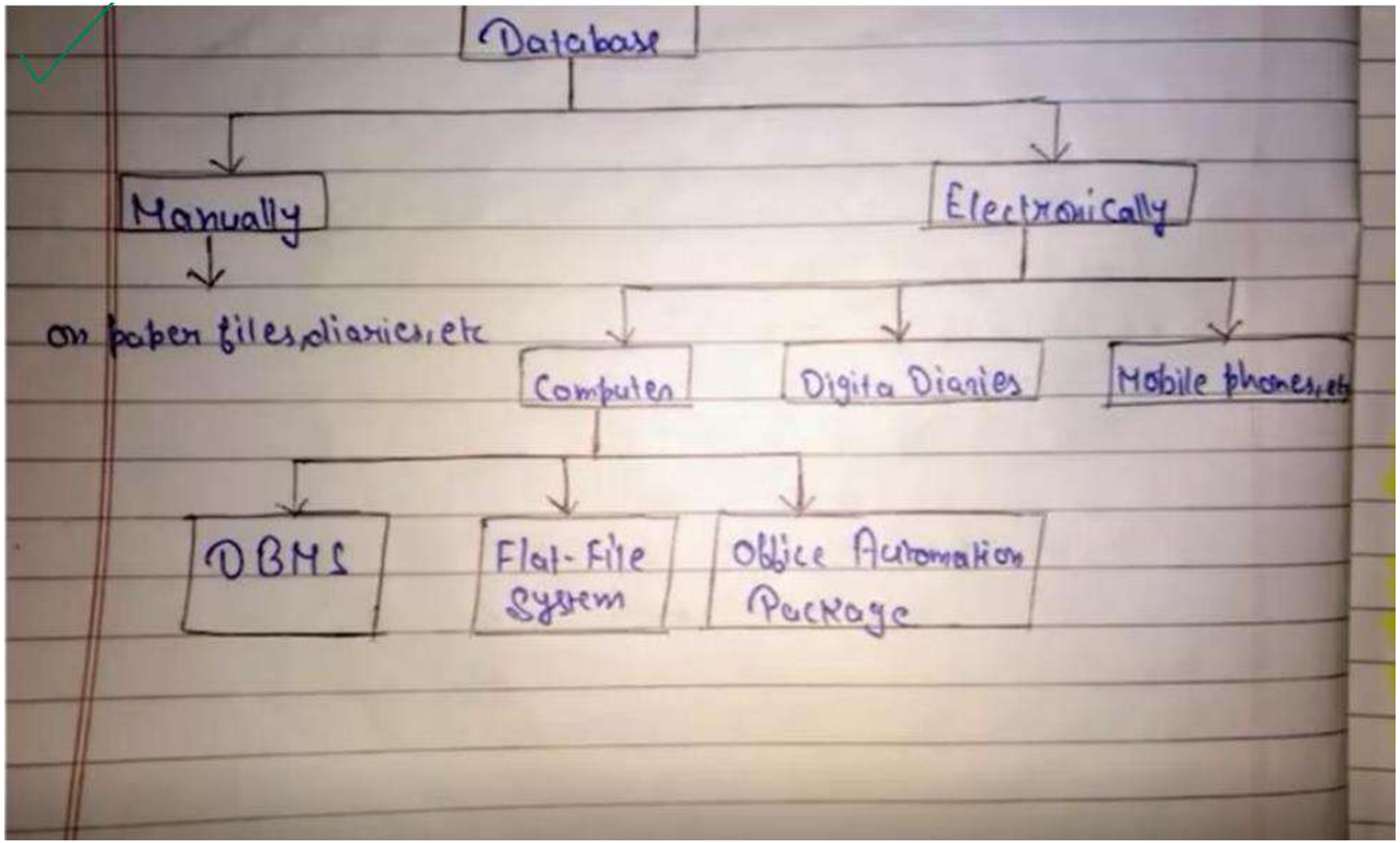
Source: https://chat.openai.com/

1.1 Introduction to Database Management System

Database

- A collection of stored operational data used by the application systems of some particular enterprise.
- Database is organized shared collection of logically related data in a systematic manner , that is stored to meet the requirement of different users of an organization, institution , government bodies that can be easily accessed, managed and updated.
- Database is actually a place where related piece of information is stored and various operation can be performed on it.
- Database can me maintained manually or through electronic devices such as : digital diaries, Mobile phones, Computers etc.
- The database are generally stored in the form of database tables.
- The table is set of elements or values that is organized in the form of rows and columns.
- The column is called as fields (attributes) and the row is called as records (tuples).
- A table has specific number of columns but may have multiple number of rows.

Database



Source: Er. Niraj KC; WRC

Database

attributes

column

SID	SName	SAge	SClass	SSection
1101	Alex	14	9	A
1102	Maria	15	9	A
1103	Maya	14	10	B
1104	Bob	14	9	A
1105	Newton	15	10	B

tuple

table (relation)

Source:https://www.tutorialspoint.com/dbms/pdf/dbms_quick_guide.pdf

Characteristics of Database

The data in a database should have the following features:

1) Organized/Related: It should be well organized and related.
2) Shared: Data in a database are shared among different users and applications.
3) Validity/integrity/Correctness: Data should be correct with respect to the real world entity that they represent.
4) Consistency: Whenever more than one data element in a database represents related real world values, the values should be consistent with respect to the relationship.
5) Security: Data should be protected from unauthorized access.
6) Non-redundancy: No two data items in a database should represent the same real world entity.
7) Easily Accessible: It should be available when and where it is needed i.e. it should be easily accessible.
8) Recoverable: It should be recoverable in case of damage.
9) Flexible to change: It should be flexible to change.

Database Management System (DBMS)

- A DBMS is a software (or a collection of programs) that enables users to create, store, modify, and extract information from database as per the requirements.
- A DBMS is a tool, that is used to perform any kind of operation on data in database.
- It has multi-user access and is designed to fulfill the need of small and large enterprises.
- Some DBMS examples: MY SQL, Oracle, System 2000, MS Excess, MY SQL Server, etc

Users of DBMS

a. Application Programmers

b. Database Administrators

c. End-Users

Database Management System (DBMS)

Users of DBMS

a. Application Programmers

- The Application programmers write programs in various programming languages to interact with databases.

b. Database Administrators

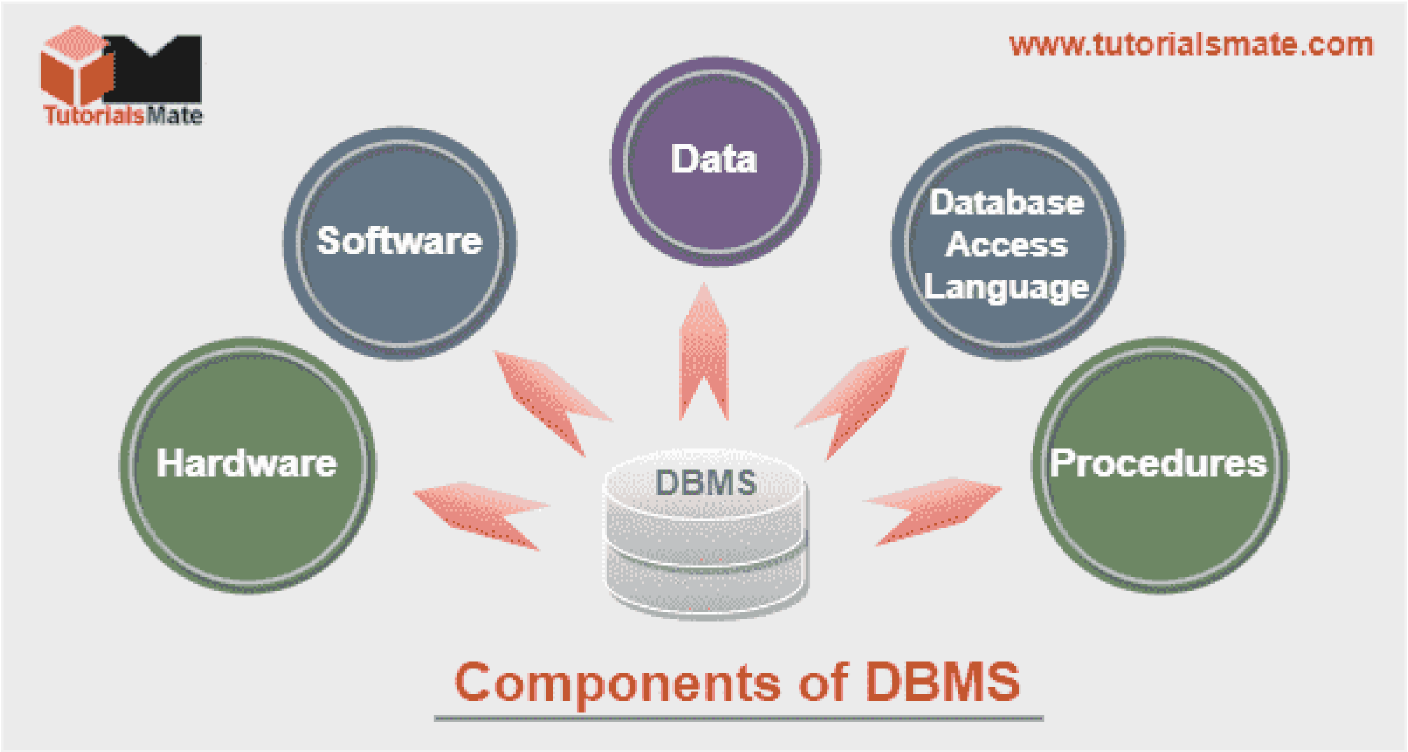
- Database Admin is responsible for managing the entire DBMS system. He/She is called Database admin or DBA.

c. End-Users

- The end users are the people who interact with the database management system.
- They are actually the one who reap the benefits of having a DBMS.
- They conduct various operations on databases like retrieving, updating, deleting, etc.

Components of DBMS

1. Software
2. Hardware
3. Data
4. Procedures
5. Database Access Language



Components of DBMS

Software

- It is the main component of a DBMS.
- It is the set of programs used to handle the database and to control and manage the overall computerized database.
- Various softwares of Database Management System are Oracle, Microsoft access, SQL server, Postgresql etc.

Components of DBMS

Hardware

- Hardware consists of a set of physical electronic devices such as computers (together with associated I/O devices like disk drives), storage devices, I/O channels, electromechanical devices that make interface between computers and the real world systems etc, and so on.
- It is impossible to implement the DBMS without the hardware devices.

Data

- Data is the most important component of the DBMS.
- The main purpose of DBMS is to process the data.
- In DBMS, databases are defined, constructed and then data is stored, updated and retrieved to and from the databases.
- The database contains both the actual (or operational) data and the metadata (data about data or description about data).

Components of DBMS

Procedures

- Procedures refer to the instructions and rules that help to design the database and to use the DBMS.
- The users that operate and manage the DBMS require documented procedures on hot use or run the database management system. These may include.

1. Procedure to install the new DBMS.
2. To log on to the DBMS.
3. To use the DBMS or application program.
4. To make backup copies of database.
5. To change the structure of database.
6. To generate the reports of data retrieved from database.

Components of DBMS

Database Access Language

- The database access language is used to access the data to and from the database.
- The users use the database access language to enter new data, change the existing data in database and to retrieve required data from databases.
- The user write a set of appropriate commands in a database access language and submits these to the DBMS. The DBMS translates the user commands and sends it to a specific part of the DBMS called the Database Jet Engine. The database engine generates a set of results according to the commands submitted by user, converts these into a user readable form called an Inquiry Report and then displays them on the screen.
- The most popular database access language is SQL (Structured Query Language).
- Relational databases are required to have a database query language.

Table

- In database, table is a collection of data which is organized in terms of rows and columns.
- Rows also known as tuple describes information about a single item, eg. A specific employee.
- Columns also known as attributes, describes a single characteristic (attributes) of its item, eg. Its EMP_ID, EMP_NAME etc
- In DBMS, the table is known as relation.
- A database can have one or more tables which may hold data about different, but related subject.

Some of the basic operation on table

- Create table
- Drop table
- Delete table
- Rename table

Table

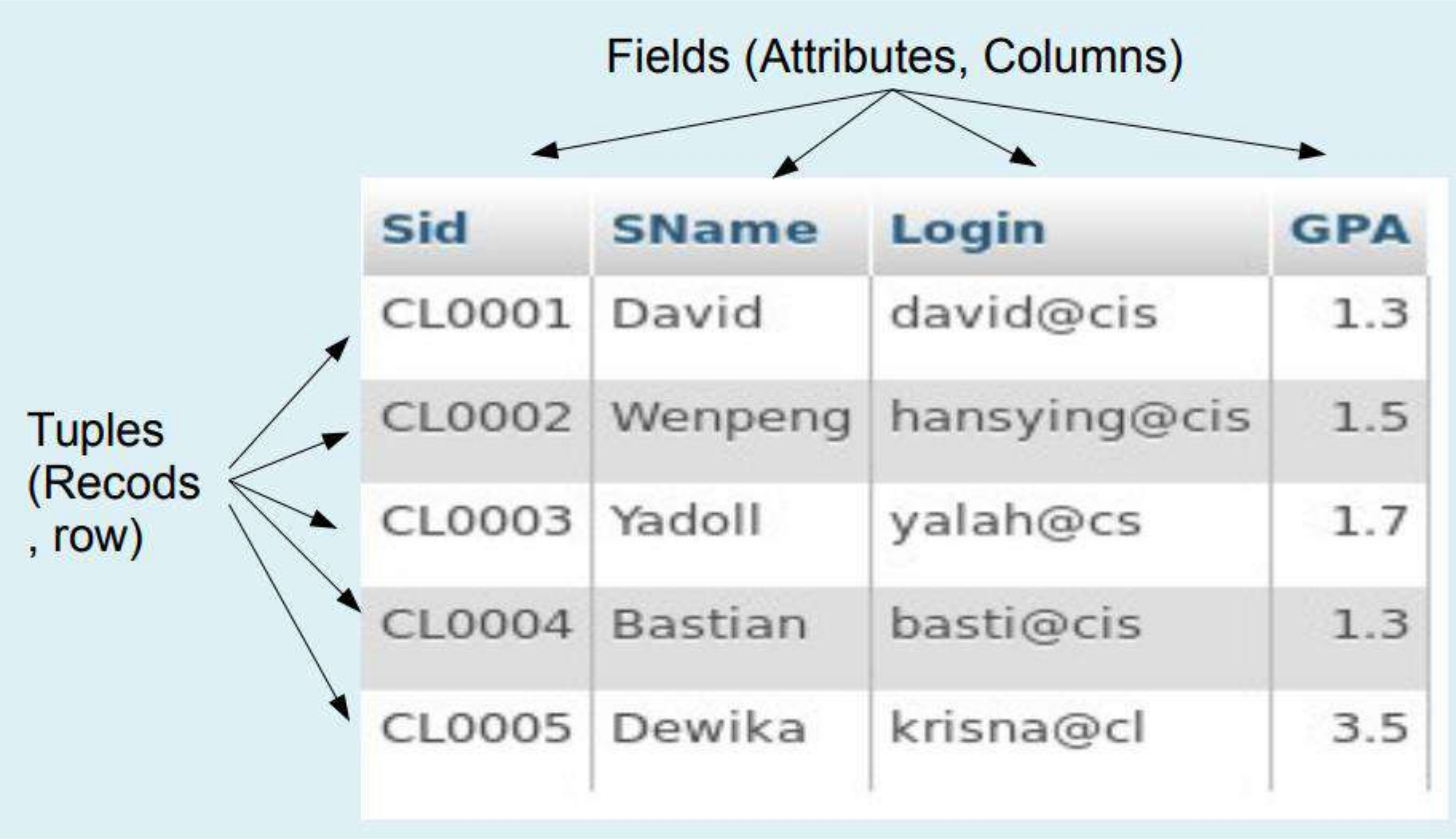
- Let's see an example of the EMPLOYEE table.

EMP_ID	EMP_NAME	CITY	PHONE_NO
1	Kristen	Washington	7289201223
2	Anna	Franklin	9378282882
3	Jackson	Bristol	9264783838
4	Kellan	California	7254728346
5	Ashley	Hawaii	9638482678

Source: https://www.javatpoint.com/dbms-sql-table

- In the above table, "EMPLOYEE" is the table name, "EMP_ID", "EMP_NAME", "CITY", "PHONE_NO" are the column names. The combination of data of multiple columns forms a row, e.g., 1, "Kristen", "Washington" and 7289201223 are the data of one row.

Table



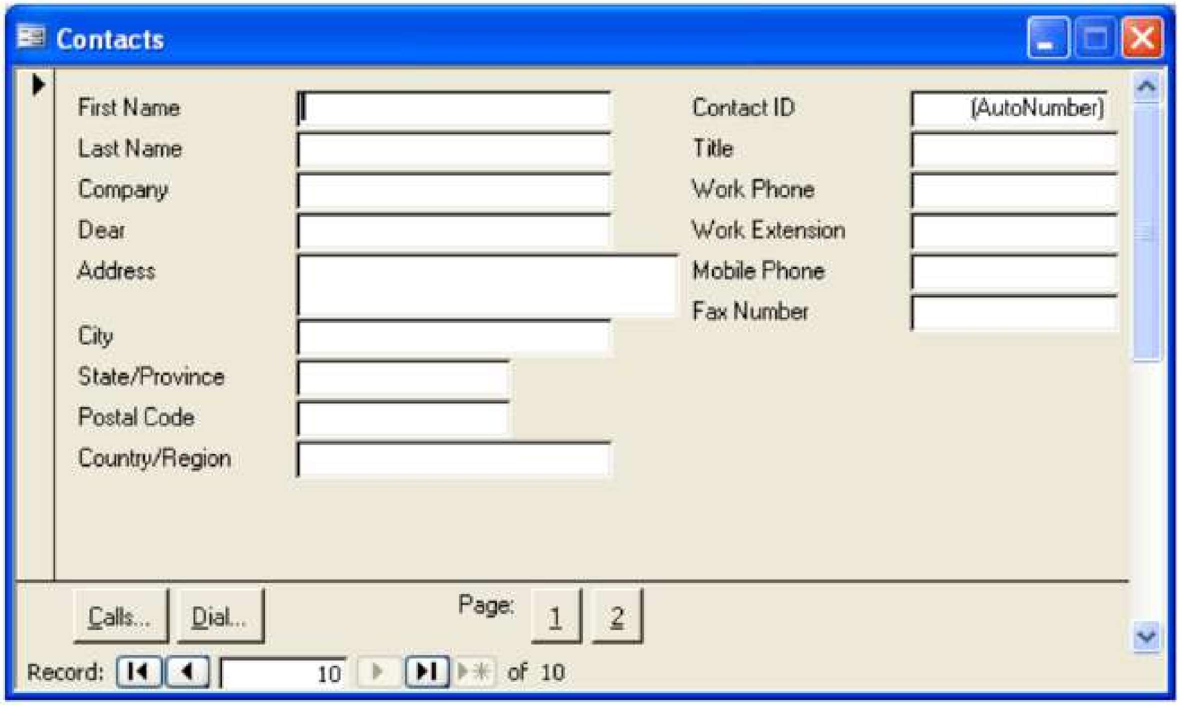
Sid	SName	Login	GPA
CL0001	David	david@cis	1.3
CL0002	Wenpeng	hansying@cis	1.5
CL0003	Yadoll	yalah@cs	1.7
CL0004	Bastian	basti@cis	1.3
CL0005	Dewika	krisna@cl	3.5

Source: https://www.cis.uni-muenchen.de/~hs/teach/14s/ir/rdbms.pdf

Form

- In the context of a Database Management System (DBMS), form refers to a graphical user interface (GUI) element used to input, view, and modify data in a database.
- Forms provide a user-friendly way to interact with the data stored in the database without the need to directly interact with the underlying database tables using SQL queries.

First Name	Last Name	Address	City	Age
Mickey	Mouse	123 Fantasy Way	Anaheim	73
Bat	Man	321 Cavern Ave	Gotham	54
Wonder	Woman	987 Truth Way	Paradise	39
Donald	Duck	555 Quack Street	Mallard	65
Bugs	Bunny	567 Carrot Street	Rascal	58
Wiley	Coyote	999 Acme Way	Canyon	61
Cat	Woman	234 Purrfect Street	Hairball	32
Tweety	Bird	543	Itotltaw	28



Source: *https://www.javatpoint.com/dbms-sql-table*

Query

- A query is simply a request for information from the database.
- Queries are typically written in a structured query language (SQL) for relational databases, such as MySQL, PostgreSQL, or Microsoft SQL Server. Non-relational databases, like MongoDB or Cassandra, may use different query languages or APIs according to their data models.

Query

Data Output | Explain | Messages | Notifications

Before Query

	id integer	dname character varying (20)	iname character varying (20)	salary integer
1	1	ComputerScience	ram	5000
2	2	Cartography	shyam	5200
3	3	English	lal	5100
4	4	GIS	rancho	2300
5	5	English	dal	5300
6	6	ProjectEngineering	chaula	5500
7	7	Community	dhami	5600
8	8	ComputerScience	kandel	6000
9	9	ComputerScience	hari	5000
10	10	Surveying	kandel	5100

Query Editor | Query History

```
select dname
from department
where salary> 5000;
```

SQL code

Data Output | Explain | Messages | Notifications

	dname character varying (20)
1	Cartography
2	English
3	English
4	ProjectEngineering
5	Community
6	ComputerScience
7	Surveying

Output

After Query

Above query gives dname whose salary is greater than 5000.

Source: https://mapadda.com/

Key (Key Fields)

- A key in DBMS is an attribute or a set of attributes that help to uniquely identify a tuple (or row) in a relation (or table).
- It is also used to establish relationship between the tables in a relational database.
- The individual values present in a key are commonly referred to as key values.

- There are mainly three types of key field in Database Management System. They are:

1. Primary Key
2. Candidate Key
3. Foreign Key

Primary key

- A primary key is a field in a table which uniquely identifies each row/record in a database table.
- Primary keys must contain unique values.
- The primary key field cannot be null.
- It is choosen from among the candidate keys and the remaining attributes in the list of candidate keys are called the alternate keys.
- In a table, there is only one primary key.

Primary key

Primary Key

STUD_NO	STUD_NAME	STUD_PHONE	STUD_STATE	STUD_COUNT	STUD_ Age
1	RAM	9865278251	Haryana	India	20
2	RAM	9655470231	Punjab	India	19
3	SUJIT	7514290359	Rajasthan	India	18
4	SURESH	8564103258	Punjab	India	21

Source: https://www.geeksforgeeks.org/difference-between-primary-key-and-foreign-key/

Here, STUD_NO, as well as STUD_PHONE both, are candidate keys for relation STUDENT but STUD_NO can be chosen as the primary key (only one out of many candidate keys).

Candidate key

- In a relational database, a candidate key is a set of one or more attributes (columns) that can uniquely identify each row (tuple) in a table.
- A table can have multiple candidate keys, but one of them is typically chosen as the primary key and the remaining attributes in the list of candidate keys are called the alternate keys.

GUID	GName	GPhone	GAddress
444444444444	Amit Ahuja	5711492685	G-35, Ashok Vihar, Delhi
111111111111	Baichung Bhutia	3612967082	Flat no. 5, Darjeeling Appt., Shimla
101010101010	Himanshu Shah	4726309212	26/77, West Patel Nagar, Ahmedabad
333333333333	Danny Dsouza		S -13, Ashok Village, Daman
466444444666	Sujata P.	3801923168	HNO-13, B- block, Preet Vihar, Madurai

Source: https://ncert.nic.in/textbook/pdf/lecs108.pdf

In above figure, table GUARDIAN has four attributes out of which GUID and GPhone always take unique values. No two guardians will have same phone number or same GUID. Hence, these two attributes are the candidate keys.

Foreign key

- Fields in a table that refer to the primary key in another table.
- This is sometimes called a referencing key.
- The data in this field must exactly match data contained in the primary key field.

STUD_NO	COURSE_NO	COURSE_NAME
1	C1	DBMS
2	C2	Computer Networks
1	C2	Computer Networks

Source: https://www.geeksforgeeks.org/difference-between-primary-key-and-foreign-key/

In above figure, table STUDENT_COURSE; STUD_NO is a foreign key and is referencing to STUD_NO in STUDENT table.

Foreign key

Primary Key STUD_NO	STUD_NAME	STUD_PHONE	STUD_STATE	STUD_COUNT	STUD_ *Age*
1	RAM	9865278251	Haryana	India	20
2	RAM	9655470231	Punjab	India	19
3	SUJIT	7514290359	Rajasthan	India	18
4	SURESH	8564103258	Punjab	India	21

Source: https://www.geeksforgeeks.org/difference-between-primary-key-and-foreign-key/

Fig: showing STUDENT table

Report

- A report refers to a formatted presentation of data retrieved from one or more database tables or queries.
- Reports are a way of presenting data in a structured and meaningful manner, making it easier for users to understand, analyze, and make informed decisions based on the information presented.

Relationship

- In the context of databases, a relationship refers to the association between tables in a relational database management system (RDBMS).
- The purpose of establishing relationships between tables is to organize and manage data efficiently, ensuring data integrity, reducing data duplication, and enabling data retrieval in a structured manner.
- Relationships are created between tables using the primary key field and a foreign key field.
- There are mainly three kind of relationship between tables. They are:

a. One to one relationship

b. One to many relationship

c. Many to one relationship

Relationship

One to one relationship

- In this relationship one record in a table relates to one record in another table.

One to many relationship

- In this relationship one record in a table can relate to many records in another table.

Many to Many relationship

- In this relationship many records in one table can relate to many records in another table.

1.2 Logical Data concept and Relationships

Relationship

Category

assignmentno integer	topic character varying (30)
1	Map Design
2	Digitization
3	Spatial Database
4	DBMS
5	Feature Selection

Identity

assignmentno integer	rollno integer
1	1
1	2
2	3
2	1
3	4
3	3
4	1
5	1

Student

	rollno integer	firstname character varying (30)	lastname character varying (30)
1	1	Sontosh	Yadav
2	2	Suraj	Thakur
3	3	Arun	Yadav
4	4	Yamlal	Lodh

Source: https://mapadda.com/

Relationship *(One-One relationship)*

Relationship between table 'Category' and 'Identity'

SQL code

```
select category.topic, identity.rollno
from category, identity
where
category.assignmentno = identity.assignmentno;
```

	topic character varying (30)	rollno integer
1	Map Design	1
2	Map Design	2
3	Digitization	3
4	Digitization	1
5	Spatial Database	4
6	Spatial Database	3
7	DBMS	1
8	Feature Selection	1

Source: https://mapadda.com/

Relationship *(One-One relationship)*

Relationship between table 'Identity' and 'Student'

	assignmentno integer	firstname character varying (30)
1	1	Sontosh
2	1	Suraj
3	2	Arun
4	2	Sontosh
5	3	Yamlal
6	3	Arun
7	4	Sontosh
8	5	Sontosh

SQL code

```
select identity.assignmentno, student.firstname
from identity, student
where
identity.rollno = student.rollno;
```

Source: https://mapadda.com/

Relationship

Category

assignmentno integer	topic character varying (30)
1	Map Design
2	Digitization
3	Spatial Database
4	DBMS
5	Feature Selection

Identity

assignmentno integer	rollno integer
1	1
1	2
2	3
2	1
3	4
3	3
4	1
5	1

Student

	rollno integer	firstname character varying (30)	lastname character varying (30)
1	1	Sontosh	Yadav
2	2	Suraj	Thakur
3	3	Arun	Yadav
4	4	Yamlal	Lodh

One-One relationship (solid line)

One-Many relationship (dashed line)

1.2 Logical data concept and relationship

Logical data concept :entities, data value, field/ attribute, records and relationships

Entities

- A real world thing or an interaction between two or more real world things.
- Entities are typically organized into tables in relational databases, and each entity has attributes that store specific pieces of information related to that entity.

For example, In a database named ‘library’ entities could be Book (having attributes: title, author), Author (name, date of birth), Borrower (contact name, address) etc.

Attribute

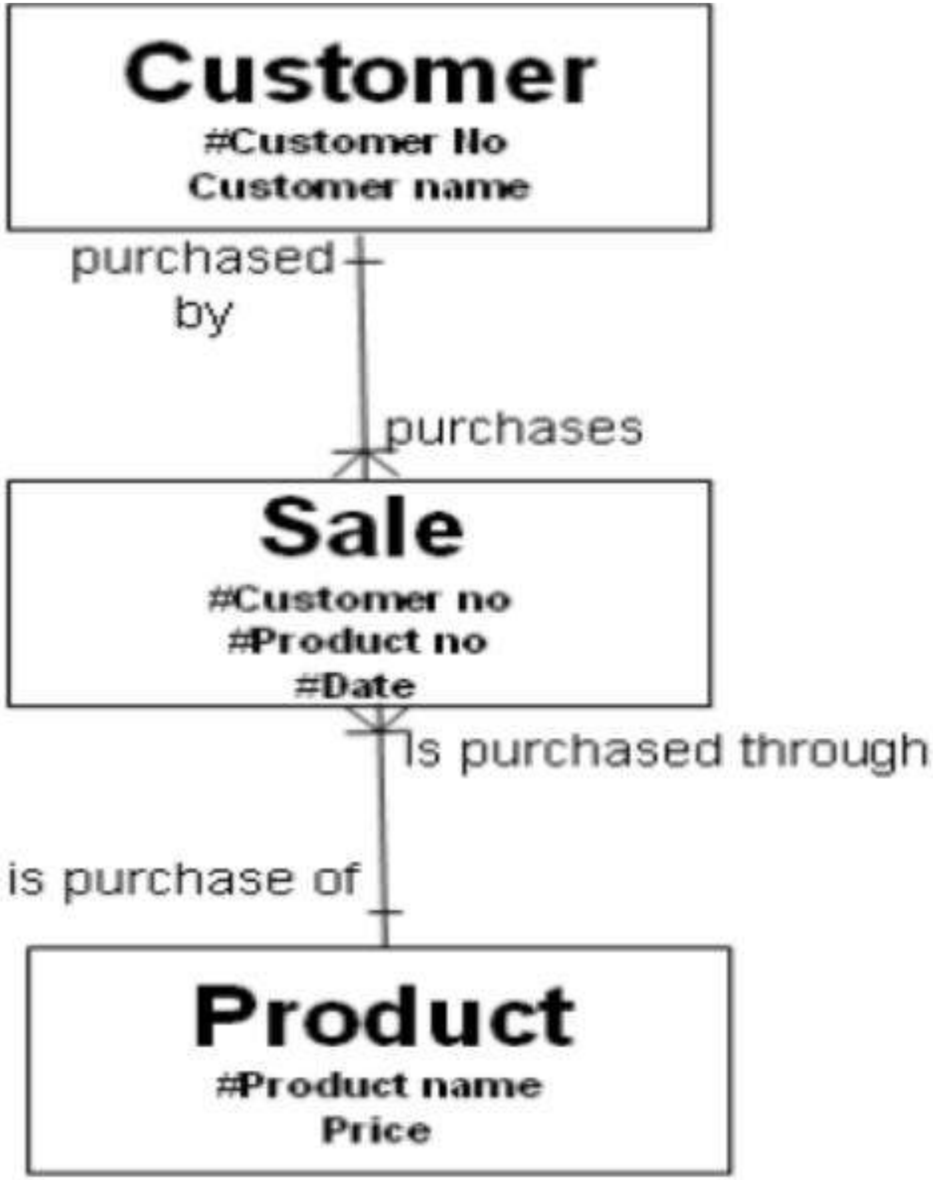
- Attribute are the the pieces of information about entities.

1.2 Logical data concept and relationship

Logical data concept :entities, data value, field/ attribute, records and relationships

For example,

- "Customer" is an entity.
- "Product" is an entity.
- For a "**Customer**" we need to know their "customer number" attribute and "name" attribute.
- For a "**Product**" we need to know the "product name" attribute and "price" attribute.
- "**Sale**" is an entity that is used to record the interaction of "Customer" and "Product".



Field data types

- Data type defines the type of data that can be stored in a particular column or field.
- It must be defined while creating a table in a relational database.

For example;

```
Create table Category(
AssignmentNo integer,
Topic varchar(40)
);
```

Above SQL code; creates a new table named Category having attributes AssignmentNo and Topic with data type integer and varchar(40).

1.2 Logical data concept and relationship

Field data types

Data type

	topic character varying (30)	rollno integer
1	Map Design	1
2	Map Design	2
3	Digitization	3
4	Digitization	1
5	Spatial Database	4
6	Spatial Database	3
7	DBMS	1
8	Feature Selection	1

Source: https://mapadda.com/

In above figure, fields topic and rollno have varchar(30) and integer data type respectively.

Field data types

Some of the most common field data types are:

1. Numeric data types:

- These are used to store numeric values.
- It includes INT, TINYINT, BIGINT, FLOAT, REAL, etc.

➢ Integer (INT): It is used to store whole numbers (e.g., 1, 10, -5).

2. Character and String data types:

- These are used to store character strings.
- It includes CHAR, VARCHAR, TEXT, etc.

Field data types

Some of the most common field data types are:

2.1 Character(CHAR):

- It refers to any number, letter, space or symbol that can be entered in a computer.
- It stores a fixed-length string of characters.
- When you define CHAR column in a database table, you must specify the maximum number of characters it can hold. For example, CHAR(10) will always occupy 10 characters of storage, regardless of the actual length of the data being stored.
- It is used for data entries with fixed length, like phone number.

2.2. Varchar() :

- It only uses storage as required by the actual content and doesn't add additional space to fulfil the size specified.
- It is used for particularly for data entries with variable length, like address.

Field data types

Some of the most common field data types are:

2.3 text:

- It is used to store large amount of textual data.
- It is designed to handle strings of characters that can be quite lengthy, and it is typically used for storing paragraphs, articles, or other forms of unstructured text.

3. Boolean:

- This data type is used for storing logical values TRUE/ FALSE or binary values (usually represented as 0 or 1).

4. Date and Time

- It includes DATE, TIME, DATETIME, etc.

Data Models

- A data model describes the structure of the database, including how data are defined and represented, relationships among data, and the constraints.
- It is the concept of tools that are developed to summarize the description of the database.
- A data model provides a way to describe the design of a database at the physical, logical and view levels.
- There are three different types of stages of data models produced while progressing from requirements to the actual database to be used for the information system.

a. Conceptual data model:

- It describes about what the system contains.
- It is this model, that is used in the requirement-gathering process i.e. before the Database Designers start making a particular database.
- It's a way to capture the essential concepts, entities, and their relationships in a format that can be easily understood by both technical and non-technical stakeholders.

Data Models

b. Logical

- It describes HOW the system will be implemented, regardless of the DBMS.
- It defines how data elements are organized, structured, and related to each other without concern for the specific physical storage details or implementation constraints.
- It's a higher-level representation that focuses on what the data means, rather than how it's stored.
- Key components of a logical data model includes entities, attributes, relationships, keys, normalization etc.

c. Physical

- It describes HOW the system will be implemented using a specific DBMS.
- It defines how the logical data model will be implemented in a specific database management system (DBMS) and the underlying hardware infrastructure.
- Key aspects of a physical data model include storage details, data types and constraints, indexing strategies, hardware configuration, data security etc.

Data Models

Types of Data Models

i. Relational Data Model
ii. Network Data Model
iii. Hierarchical Data Model
iv. Entity-Relationship (E-R) Models
v. The Object Oriented Model

Relational Data Model

- The relational model expresses the data and relationship among the data in the form of tables.
- Each table has columns and rows which are formally called attributes and tuples respectively.

Some of the key components of relational data model are:

a. Tables (Relations):

- In the relational model, data is organized into tables, also known as relations.
- Each table consists of rows (tuples) and columns (attributes).
- Each row represents a record or instance of data, while each column represents a specific attribute or characteristic of the data

b. Attributes (Columns):

- Attributes define the specific types of data that can be stored in each column.
- Each column has a name and a defined data type, such as text, numbers, dates, etc.

Relational Data Model

c. Tuples (Rows):

- Each tuple corresponds to a unique record in the table and contains values for each attribute.

d. Keys:

- Keys are attributes or combinations of attributes that uniquely identify individual rows in a table.

e. Relationships:

- Relationships describe how data in different tables are related to each other.
- Common relationship types include one-to-one, one-to-many, and many-to-many relationships

f. Normalization

g. SQL

Importance of DBMS

Some key reasons why a DBMS is important are:

a. Data Organization and Storage:

- A DBMS provides a structured way to organize and store data.
- It uses various data structures and algorithms to optimize storage and retrieval processes, ensuring that data is stored efficiently and can be accessed quickly.

b. Efficient Data Access

- In a database system, the data is managed by the DBMS and all access to the data is through the DBMS providing effective data processing.

c. Data Sharing :

- DBMS allows a user to share the data in any number of application programs.

Importance of DBMS

Some key reasons why a DBMS is important are:

d. Data Security:

- DBMS systems offer security mechanisms to control access to the database.
- Users can be granted specific levels of access based on their roles and responsibilities. This helps protect sensitive information and prevent unauthorized access.

e. Improved productivity:

- DBMS reduces the time and effort required to manage data, which increases productivity and efficiency.
- It also provides a user-friendly interface for data entry and retrieval, which reduces the learning curve for new users.

f. Data Consistency:

- DBMS ensures data consistency by enforcing data validation rules and constraints. This ensures that data is accurate and consistent across different applications and users.

1.3 Data models and DBMS applications

File system

- A file system is a method an operating system uses to store, organize, and manage files and directories on a storage device (hard disk, pen drive, DVD etc.).
- File system arranges the files and helps in retrieving the files, when required.
- It mostly consists of different types of files like mp3, mp4, txt, doc, etc. that are grouped into directories.
- A file system enables you to handle the way of reading and writing data to the storage medium.
- It is directly installed into the computer with the operating systems such as Windows and Linux.

File system

- Some common types of file systems are:

a. FAT (File Allocation Table)

- It is an older file system used by older versions of Windows and other operating systems.

b. NTFS (New Technology File System):

- It is a modern file system used by Windows.
- It supports features such as file and folder permissions, compression, and encryption.

c. HFS (Hierarchical File System):

- It is a file system used by macOS.

Comparison of DBMS over file system

a. Data structure:

- In a file system, data is typically organized into files and folders. Files can contain any type of data, such as text, images, videos, etc.
- In a DBMS, data is structured into tables, rows, and columns. This structured format helps with efficient organization and retrieval of data.

b. Data Relationships:

- File systems don't inherently support establishing relationships between different pieces of data.
- DBMS systems allow you to create relationships between different tables using keys. This enables complex data models and advanced querying.

Comparison of DBMS over file system

c. Data Retrieval and Querying:

- File systems can be inefficient for retrieving specific data, especially from large datasets.
- A DBMS provides query languages like SQL, making it easier to retrieve and manipulate data based on specific criteria. This enhances data searching and filtering.

d. Security:

- Security features in file systems are often limited and rely on the operating system's capabilities.
- DBMS systems offer advanced security features, including user authentication, access control, and encryption at the database level.

e. Data concurrency (many users can access data at the same time)

- File system does not offer concurrency whereas DBMS provides a concurrency facility.

Assignment

1. Differences between a database and a relational database.
2. What is a relational database management system (RDBMS) ?
3. Explain the importance and use of Database Management System (DBMS).
4. What is meant by normalization in database.
5. What are the applications/uses of DBMS ?
6. What are the different types of database ? Explain about relational database.
7. What are the advantages and disadvantages of database ?
8. Difference between primary key and foreign key.

References

- Data and Information; https://portal.abuad.edu.ng/lecturer/documents/1554208765DATA_AND_INFORMATION.pdf
- *Keys in database*
- Source: https://ncert.nic.in/textbook/pdf/lecs108.pdf
- https://www.geeksforgeeks.org/difference-between-primary-key-and-foreign-key/
- https://www.cis.uni-muenchen.de/~hs/teach/14s/ir/rdbms.pdf
- http://www.rjspm.com/PDF/BCA-428%20Oracle.pdf
- https://www.digitalocean.com/community/tutorials/sql-data-types
- Datatype
- https://www.geeksforgeeks.org/sql-data-types/
- File system vs DBMS
- https://ncert.nic.in/textbook/pdf/keip107.pdf
- Advantage Disadvantage of DBMS
- https://www.pvpsiddhartha.ac.in/dep_it/lecture%20notes/DBMS_19/PVP19%20DBMS%20UNIT-1%20Material.pdf
- https://www.geeksforgeeks.org/advantages-of-database-management-system/
- File system
- https://www.geeksforgeeks.org/file-systems-in-operating-system/
- https://www.lifewire.com/what-is-a-file-system-2625880

References

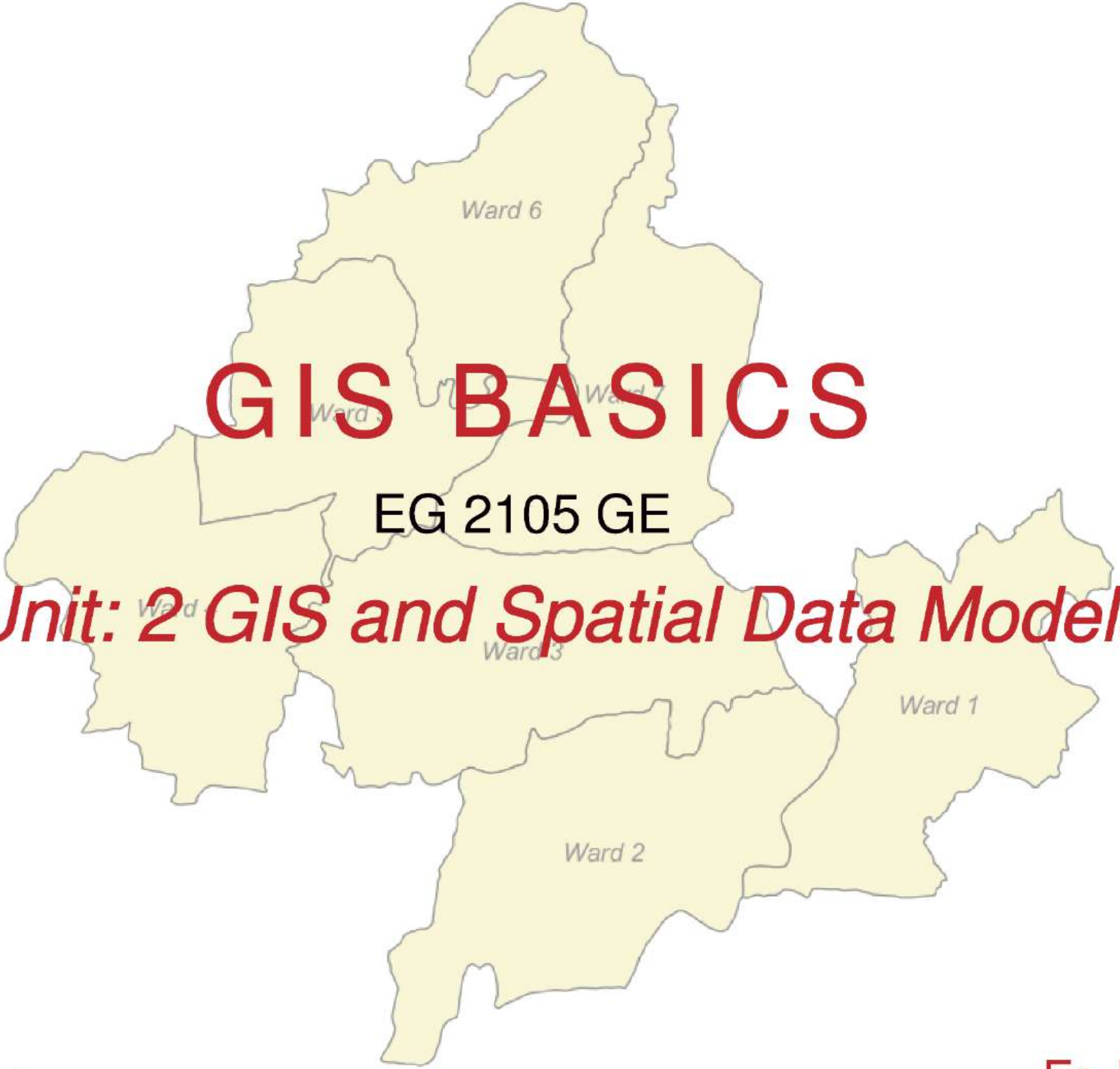
Difference between file system and dbms

- *https://www.geeksforgeeks.org/difference-between-file-system-and-dbms/*

GIS BASICS

EG 2105 GE

Unit: 2 GIS and Spatial Data Models



Public Namuna Secondary School

Lecture by

Er. Keshav Raj Bhusal

Table of contents

2.1. Introduction to GIS

- Geographic phenomena
- Definition of GIS,
- Component of GIS (Hardware, Software, People, Data, Method)
- Stages of GIS workflow (Data Preparation/ Acquisition, Data storage & Management, Data Analysis, and Visualization)
- Spatial and non-spatial data, Relation, tuple & attribute.
- Application area of GIS
- Various types of GIS users

2.1 Introduction to GIS

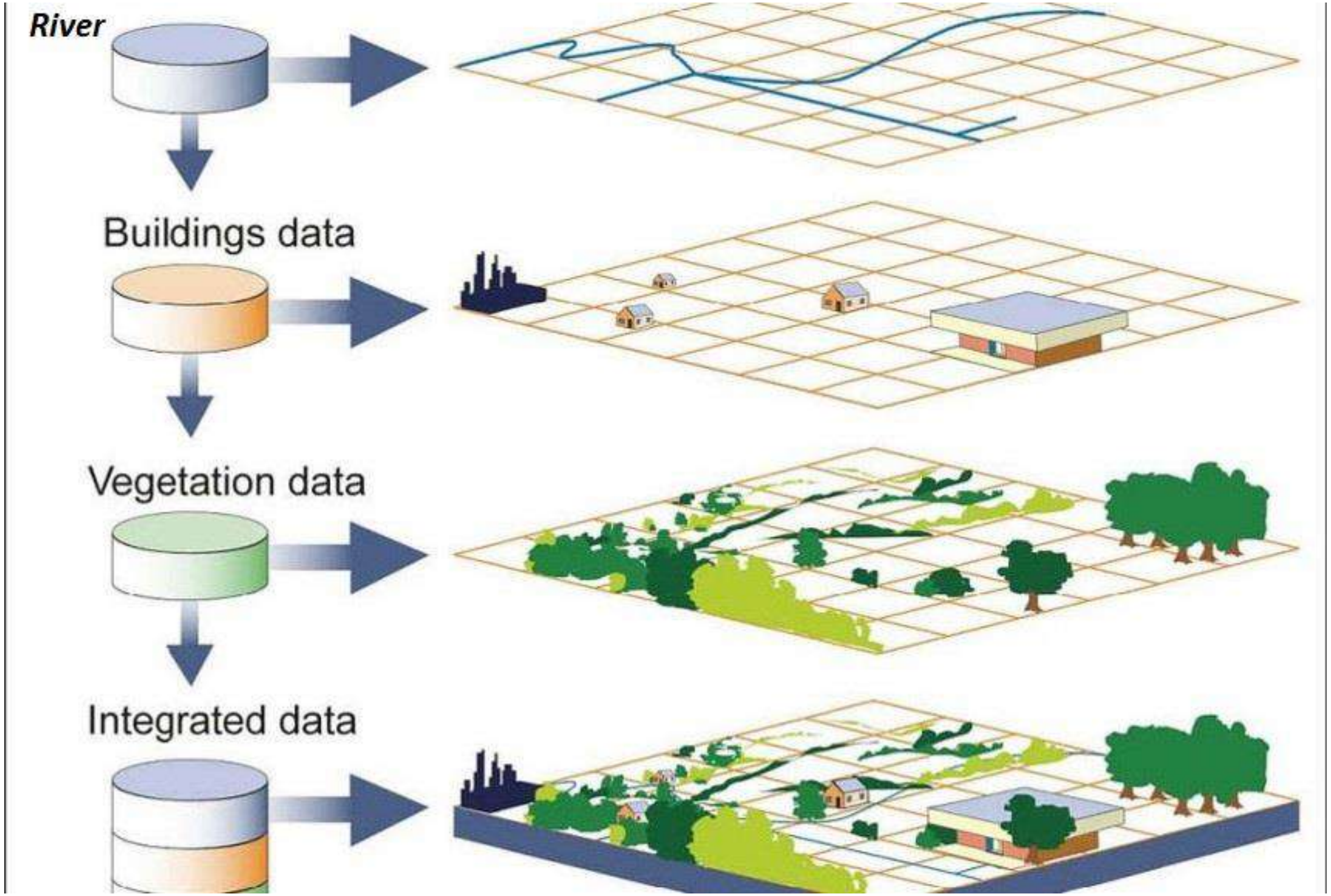
GIS

- A geographic information system (GIS) is a computer based system for capturing, storing, querying, analyzing and displaying data related to positions on Earth's surface.
- In practical terms, a Geographic Information System is a database to store data, a calculator which can manipulate and analyse data, a visualisation window to show results.
- GIS provides Geographic/geospatial information about the places on earth`s surface, knowledge about "what, where, when, how, how far etc.
- With GIS technology, people can compare the locations of different things in order to discover how they relate to each other.

For example, using GIS, a single map could include sites that produce pollution, such as factories, and sites that are sensitive to pollution, such as wetlands and rivers. Such a map would help people determine where water supplies are most at risk.

- GIS is an integrated technology related to a number of fields including remote sensing, Global Positioning System (GPS), cartography, surveying, geostatistics, Web mapping, programming, database management, and graphics design.

GIS



Source: https://education.nationalgeographic.org/resource/geographic-information-system-gis/

Fig1: showing different data layers in GIS

Components of GIS

Source: https://www.rst2.org/ties/GENTOOLS/comp_gis.html

Components of GIS

A working GIS integrates mainly four key components: hardware, software, data and people.

1. Hardware

- Hardware is the computer system on which a GIS operates.
- Today, GIS software runs on a wide range of hardware types, from personal computers to web servers.
- Some of the hardware components are: Motherboard, Hard driver, processor, graphics card, printer and so on. These all component function together to run a GIS software smoothly.

2. Software

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

➢ Tools for the input and manipulation of geographic information.

➢ A database management system (DBMS).

➢ Tools that support geographic query, analysis, and visualization.

➢ A graphical user interface (GUI) for easy access to tools.

- Commonly used GIS softwares are: ArcGIS, QGIS, ArcGISPro etc.

Components of GIS

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

3. Data

- Geospatial data is the core of any GIS.
- There are two primary types of data that are used in GIS: vector and raster data.
- Vector data is spatial data represented as points, lines and polygons. Raster data is cell-based data such as aerial imagery and digital elevation models.

4. 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.

Components of GIS

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

5. Method or Procedure

- A successful GIS operates according to a well-designed plan, business rules, procedures and methods unique to each organization.
- The procedures and software requirements will help to define a list of skills required to operate on the GIS functions, procedures, and make judgment.

Stages of GIS workflow

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

i. Data Preparation/ Acquisition
ii. Data storage & Management,
iii. Data Analysis
iv. Data Visualization

Stages of GIS workflow

i. Data Preparation/ Acquisition

- It is the first step. Without data GIS is meaningless.
- There are various techniques for spatial data acquisition in GIS. Survey data can be directly entered into GIS from digital data collection system from survey instruments like Total Station, GPS etc.
- Satellite remote sensing can be an important data source, the satellite imageries collected by the sensor can be an important data source.
- Aerial Photographs collected from airborne digital camera can be implemented into GIS.
- Existing data in paper form can be digitized or scanned to produce a digital data which can be then implemented into GIS.

Stages of GIS workflow

i. Data Preparation/ Acquisition

- It is the first step. Without data GIS is meaningless.
- There are various techniques for spatial data acquisition in GIS. Survey data can be directly entered into GIS from digital data collection system from survey instruments like Total Station, GPS etc.
- Satellite remote sensing can be an important data source, the satellite imageries collected by the sensor can be an important data source.
- Aerial Photographs collected from airborne digital camera can be implemented into GIS.
- Existing data in paper form can be digitized or scanned to produce a digital data which can be then implemented into GIS.

ii. Data Storage and Management

- After the data has been collected the data has to be used as input in GIS system.
- The input data has to be stored and effectivelly. The GIS provides the facility to contain and maintain the data.

Stages of GIS workflow

ii. Data Storage and Management

- The effective data management should contain the following aspects:

a. Data security

- Data security refers to the process of protecting data from unauthorized access.

b. Data integrity

- Data integrity means maintainance of assurance of accuracy.

c. Data storage and retrival

- Storing and collecting the require data from the collection of data.

d. Data maintainance ability

- Continual storage and constant check of the data.

Stages of GIS workflow

iii. Data Analysis

- GIS analysis allows you to solve complex location-oriented problems and better understand where and what is occurring in your world.
- GIS analysis is a process in which you model problems geographically, derive results by computer processing, and then explore and examine those results.

Analysis can be done for following real world problems:

i. Understanding where things are or where events occur.

ii. Measuring sizes, shapes, and distributions of things or measurements.

iii. Analyzing relationships and interactions between places.

iv. Optimizing locations for facilities, or routes for transportation.

v. Making predictions based on existing or theoretical patterns and relationships.

2.1 Introduction to GIS

Stages of GIS workflow

iv. Data Visualization

- In GIS, cartographic visualization is the translation or conversion of spatial data from a database into graphics(mainly maplike products).
- You can display your data as locations on an interactive map using the map visualization. For example, you can create a visualization that displays how customer households are clustered in different parts of the country using a density map, or display retail locations as a series of map markers that users can click to view additional information about stores in their area.
- Visualization of large amount of spatial data can be done with the help of various gis procedures.

Spatial Data & Non-spatial data

Spatial data

- Any data which are directly or indirectly referenced to a location on the surface of the earth are Spatial data.
- Spatial data are also known as geospatial data.
- Spatial data is information about a physical object that can be represented by numerical values in a geographic coordinates system (latitudes and longitudes).
- Spatial data represents the location, size and shape of an object on earth surface such as mountain, plain, township, people etc.
- It also provides all the attributes of an entity that is being represented.
- Spatial data represents features using point, line, polygon or pixel.

Spatial Data & Non-spatial data

Non-Spatial data (Attribute or Characteristics data)

- Non-spatial data, simply, is data that contains ‘what’ instead of ‘where’ and describe geographic regions or define characteristics of spatial features within geographic regions.
- Non-spatial data can help put spatial data into more context. It is a characteristic data.
- They are usually alphanumeric and provide information such as colour, texture, quantity, quality, and value of features.

Suppose you have got a data named ‘tree’ with attributes name type and height. Here ‘type’ and ‘height’ are non-spatial data and can provide meaning to location of tree.

- Non-spatial data are often derived from sources such as documents, files, and tables.
- In general, spatial data will have related non-spatial attributes, and thus some form of linkage must be established between these two different types of information.
- Such linkage is achieved with a common identifier that is stored with both the graphic and the nongraphic data.

Features & Attribute table

Features

- It is a representation of a real-world object on a map.
- A feature is a digital representation of a geographic object (e.g., a house, a road segment, a county) or event (e.g., a traffic accident) located in space.
- In GIS, features are shown as vector or raster data model.

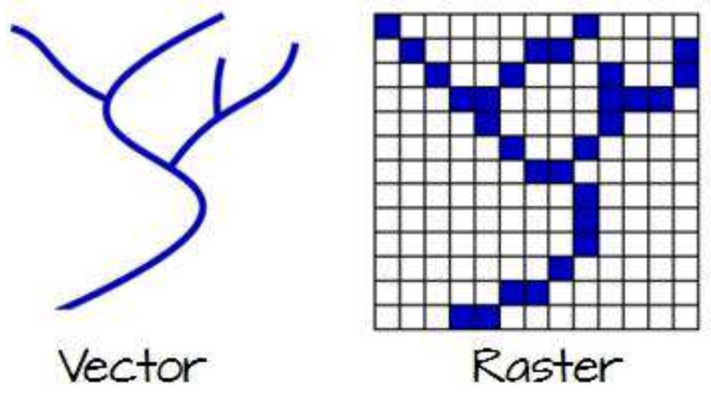


Fig: river being represented as vector and raster.

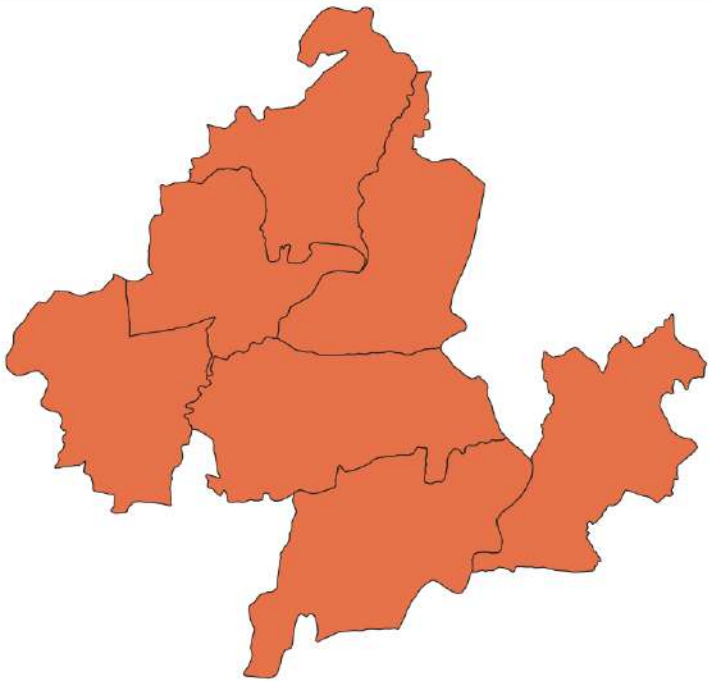
Source: https://mgimond.github.io/Spatial/chp02_0.html

Features & Attribute table

Attribute table

- Attribute table is the tabular file consisting of the information about the set of geographic features.
- They are usually arranged so that each row represent one feature and each column represent the feature attribute.
- Attributes are the pieces of information that determine the property of the features.
- Each row is called as record or tuple.
- Each column is called as field or attribute.
- They are usually arranged so that each row represent one feature and each column represent the feature attribute.

Features & Attribute table



Source: http://mapadda.com/

SiyariWard_44N — Features Total: 7, Filtered: 7, Selected: 0

	DISTRICT	PALIKA	TYPE	WARD	PondsNo	WardNo
1	RUPANDEHI	Siyari	Gaunpalika	6	39	Ward 6
2	RUPANDEHI	Siyari	Gaunpalika	7	12	Ward 7
3	RUPANDEHI	Siyari	Gaunpalika	1	2	Ward 1
4	RUPANDEHI	Siyari	Gaunpalika	5	263	Ward 5
5	RUPANDEHI	Siyari	Gaunpalika	4	282	Ward 4
6	RUPANDEHI	Siyari	Gaunpalika	3	77	Ward 3
7	RUPANDEHI	Siyari	Gaunpalika	2	14	Ward 2

Show All Features

Fig: showing attribute table having 7 features of Wardwise Siyari Gaupalika shapefile.
Here DISTRICT, PALIKA, TYPE, WARD, PondNo & WardNo are fieldnames or attributes.

Application areas of GIS

1. Business:

- Many retail businesses use GIS to help them determine where to locate a new store and to find the best routes to deliver their goods and services.

2. Emergency Preparedness and disaster management:

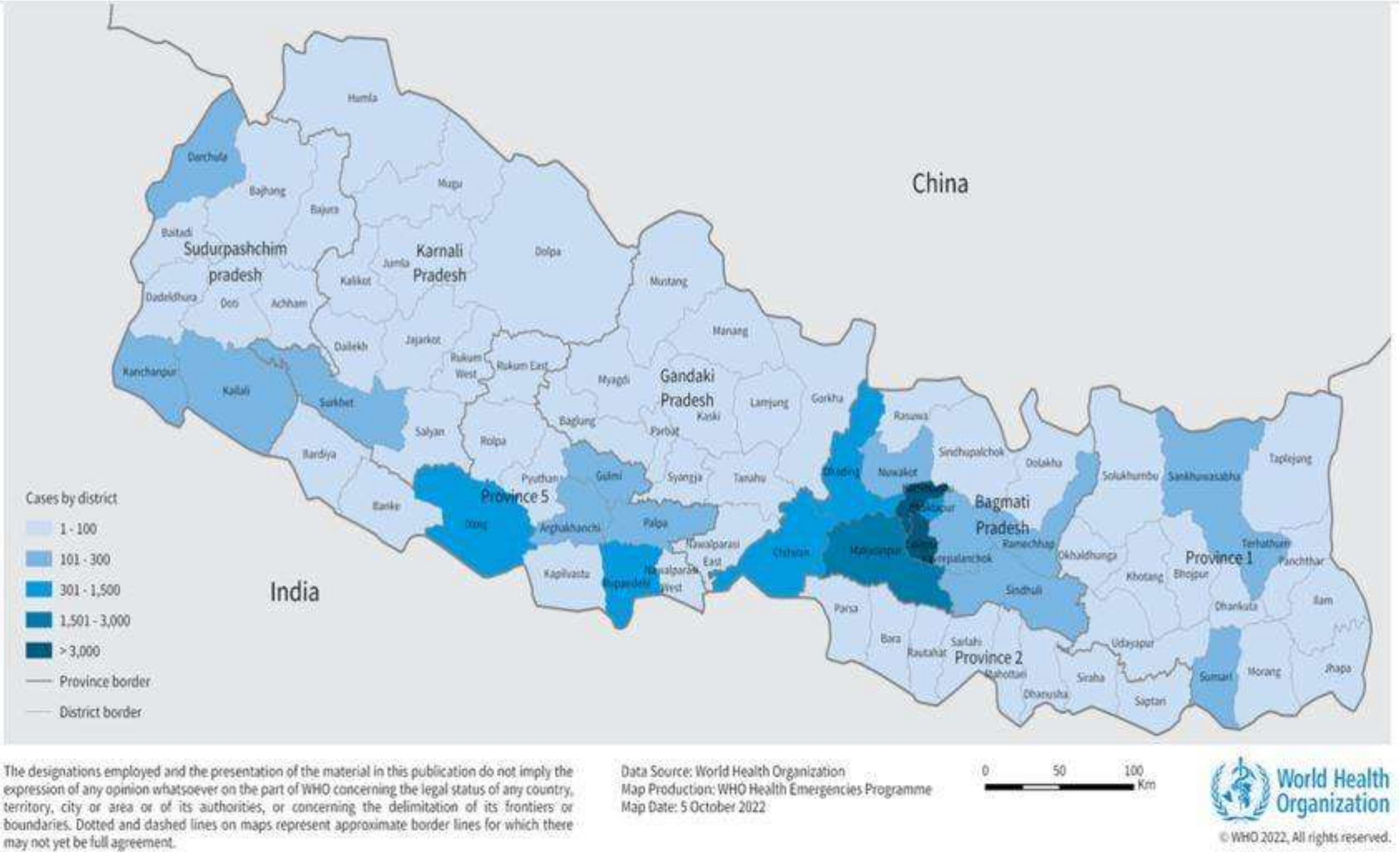
- City, state, or federal officials use GIS to help plan their response in the case of a natural disaster such as an earthquake or hurricane.
- GIS maps can show these officials what neighborhoods are most in danger, where to locate emergency shelters, and what routes people should take to reach safety.

Applications of GIS

3. Health sector:

- With GIS and spatial analysis, you can prioritize spending, site service locations, and identify vulnerable populations.
- The result is better outcomes for patients, stakeholders, and the public.

Applications of GIS



Source:https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON412

Applications of GIS

4. Military:

- Through GIS commanders and other military officials can get the information of terrain type, battlefield through relief maps. It helps in strategic planning.

5. Archaeology:

- GIS based applications help archaeologists to collect and map physical locations of their excavation findings from a wealth of sites.
- Preliminary Site Investigation: Searching ancient maps for buildings, cemeteries, roads, and fences as these sites present important clues to archaeological sites.
- Archaeological Site Prediction: GIS helps to predict archaeological sites using the Multiple Criteria Evaluation (MCE) by using criteria such as favorable slope, aspect, geology, hydrology, and distance to water.

Applications of GIS

6. Civil Engineering:

- Cut & Fill: Analysis of cut and fills estimation for major civil engineering projects such as major road constructions, excavation can be done through GIS.
- CAD Interoperability: Integrating CAD data (DWG, DXF) into GIS and vice versa.
- Building Permits – Helping the user determine whether or not a requested permit is in a historic district, an aquifer protection district, wetland, or floodplain.

7. Environment:

- Environmental Impact Assessment: Measuring anticipated effects on the environment of a proposed development project.
- Wetland Inventory: Delineating wetlands by types and function.

Applications of GIS

8. Real estate:

- Real Estate Metrics: Choosing the right house to buy or build based on the distance to schools, parks, transit stops, and other geospatial metrics.
- Construction: Considering slopes for construction and livability as 15 degrees or less is considered buildable.
- Parcel Dimensioning: Analyzing a property parcel shape to determine the area and dimensions of each identified parcel.

Types of GIS Users

a. Creator:

- A creator is a GIS specialist, asset manager, or data journalist who creates map and apps, perform spatial analyses, and share the results with colleagues and clients mainly via ready-to-use apps.

b. Viewer:

- A viewer is a GIS user who view the maps and apps created by others but don’t need to build them by oneself.
- In an organization a viewer may be an executive officer, manager, or sales manager.

c. Field worker:

- A field worker is one who collects real-time data, manage field assignments, inspect assets, and/or do surveying.
- Different type of mobile applications like SW maps, Kobo collect, Survey 123 or other surveying techniques can be used for data collection.

Types of GIS Users

d. Editor:

- They modify and add spatial data ensuring accuracy of the content.
- They have the ability review and change the existing data and create a new data to meet the needs of business.

e. GIS Professionals:

- They are the trusted source of information within the organization.
- They perform advanced spatial analysis, visualization and mapping and develop detailed layered maps to support decision making process.

QUESTIONS to think ?

1. Have you used GIS before ?

References

- GIS definition: https://education.nationalgeographic.org/resource/geographic-information-system-gis/
- ICIMOD: https://www.icimod.org/capacity-building/2021/WoGIT%20Afg/D1_basics_of_gis_pt.pdf
- highpointnc.gov: https://www.highpointnc.gov/DocumentCenter/View/1900/What-is-GIS-PDF?bidId=
- Indian institute: https://www.dspmuranchi.ac.in/pdf/Blog/Fundamental%20Concept%20of%20GIS.pdf
- GIS lounge: https://www.gislounge.com/what-is-gis/
- https://itsleeds.github.io/QGIS-intro/introduction-gis-ideas-and-concepts.html
- GIS concept: https://va.water.usgs.gov/GreatValley/ForumDec142005/ForumCarr1205.pdf
- GIS geography: https://gisgeography.com/gis-applications-uses/
- Spatial & Non-spatial data: https://old.amu.ac.in/emp/studym/100018292.pdf

Table of contents

Spatial Data Model

- Data model is a way of defining and representing real world surfaces.
- Spatial Data Models are the set of guidelines that convert the real world (geographic phenomenon) to digitally and logically represented spatial objects consisting of geometry and attributes.

There are three types of spatial data model.

2.2.1 Vector Data Model (or Vector Data Structure)

2.2.2 Raster Data Model (or Raster Data Structure)

2.2.3 TIN Data Model

2.2. Spatial Data Models

Vector Data Model

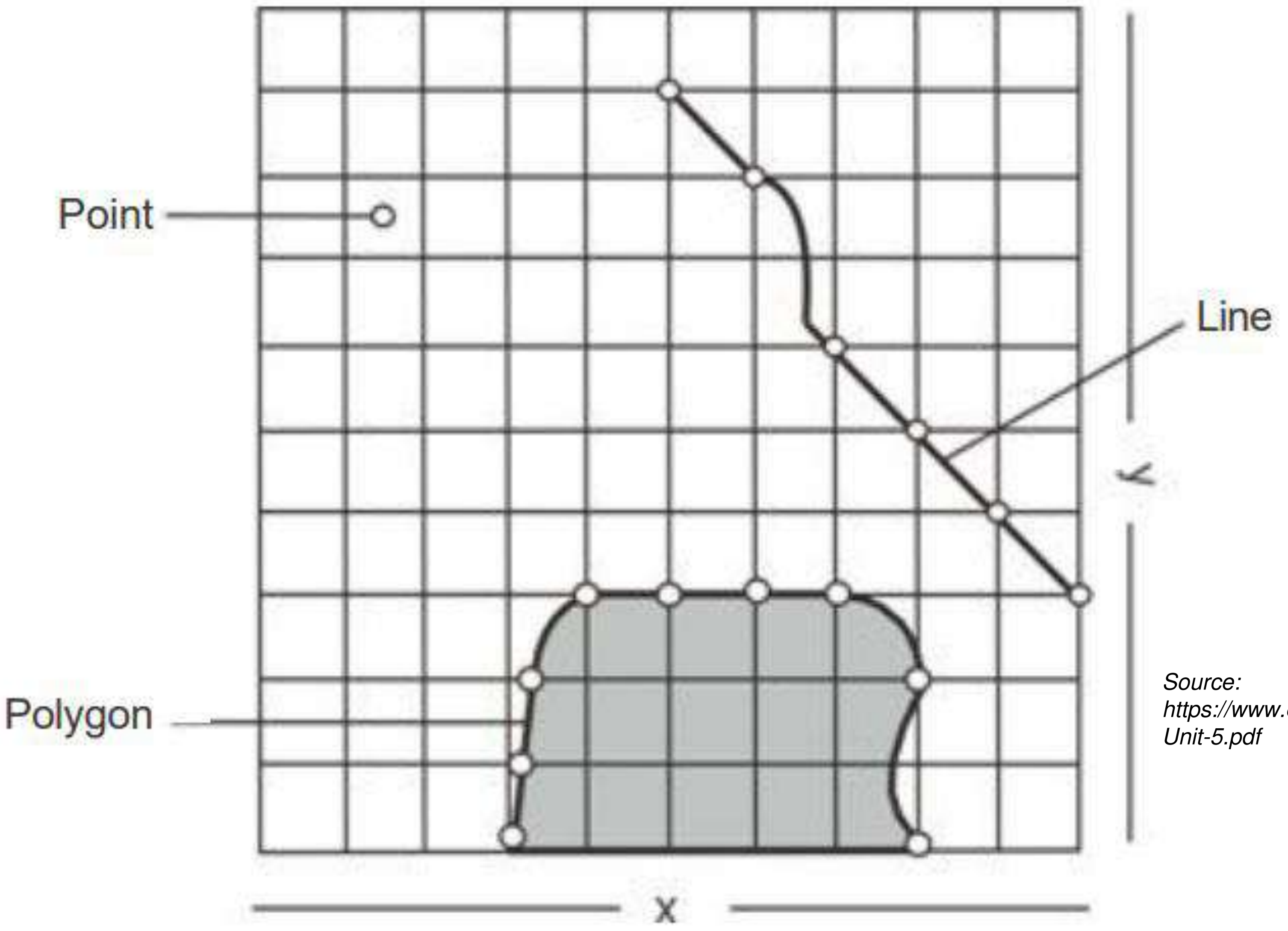
- The vector data model, also called the discrete object model, uses geometric objects of point, line, and polygon to represent spatial features on the Earth's surface.
- It is close to the traditional mapping approach.
- In a vector model, the positions of points, lines, and areas are precisely specified. The position of each object is defined by a (series of) coordinate pairs.
- Point features are defined by one coordinate pair, a vertex. Linear feature is defined by combination of coordinate pairs. Polygonal features are defined by a set of closed coordinate pairs.
- Vector models are useful for storing data that has discrete boundaries, such as country borders, land parcels, and streets.

Vector data models can be structured in many different ways. Two of them are:

i. Spaghetti Model

ii. Topological Model.

Vector Data Model



Source: https://www.egyankosh.ac.in/bitstream/123456789/39550/1/Unit-5.pdf

Geometry types of Vector Data Model

There are mainly three kinds of geometry for vector data model. They are:

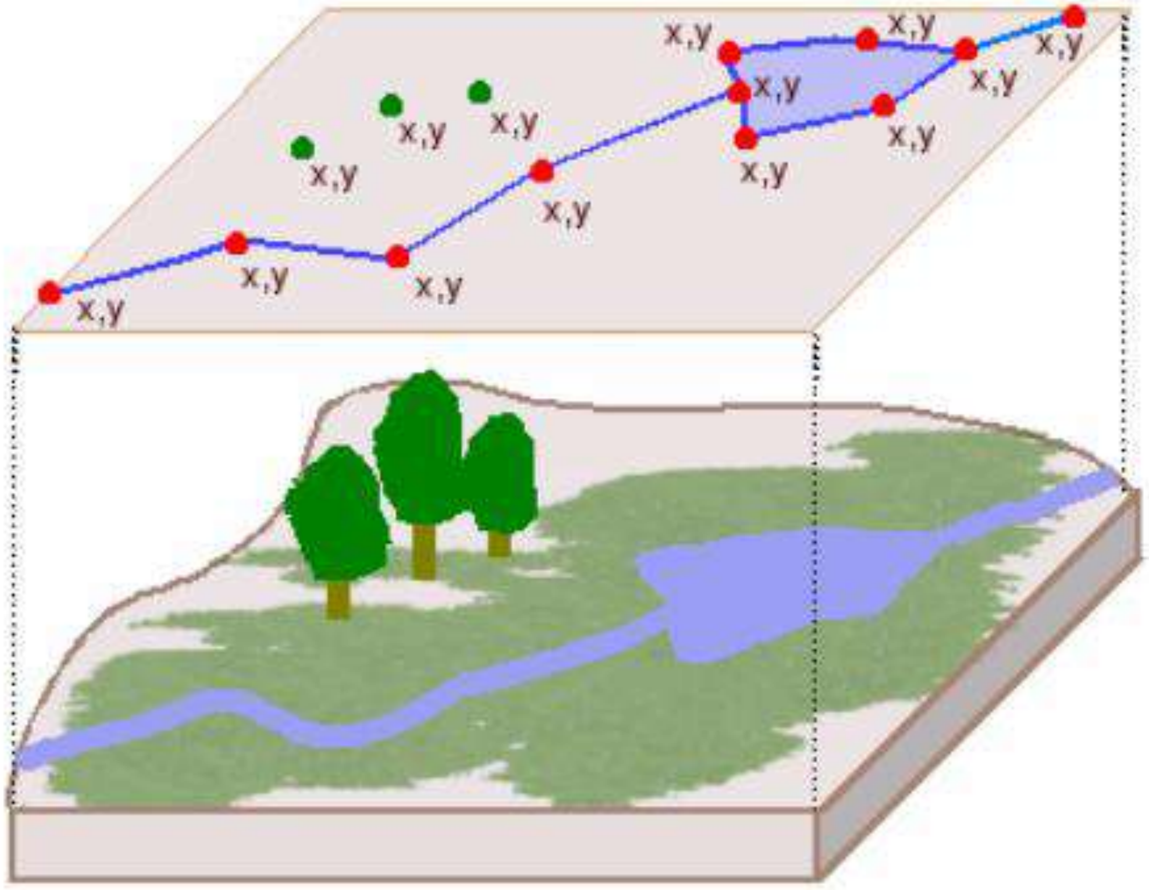
1. Point

- Points are zero-dimensional objects that contain only a single coordinate pair.
- Points are typically used to model singular, discrete features such as buildings, wells, power poles, sample locations, and so on.
- Points have only the property of location.
- Points can be spatially linked to form more complex features.

2. Line

- Lines are one-dimensional features composed of multiple connected points.
- Lines are used to represent linear features such as roads, streams, faults, administrative boundaries, and so on.
- Lines have the property of length.

Geometry types of Vector Data Model



Source: http://www.geography.hunter.cuny.edu/~jochen/gtech361/lectures/lecture05/concepts/03%20-%20Geographic%20data%20models.html#:~:text=The%20vector%20data%20model%20represents,by%20x%2Cy%20coordinate%20pairs.&text=In%20the%20raster%20data%20model,meter%20or%20a%20square%20mile.

Geometry types of Vector Data Model

3. Polygon

- Polygons are two-dimensional features defined by lines that close to form the polygon boundaries.
- In the case of polygons, the first coordinate pair (point) on the first line segment is the same as the last coordinate pair on the last line segment.
- Polygons are used to represent features such as city boundaries, geologic formations, lakes, soil associations, vegetation communities, and so forth.
- Polygons have the properties of area and perimeter.

Vector Data file formats

Some of the commonly used vector file formats in GIS are:

a. Shapefile
b. KML/KMZ
c. CSV
d. GPX
e. GeoJSON/TopoJSON
f. Geodatabase etc.

2.2. Spatial Data Models

Vector Data file formats

Some of the commonly used vector file formats in GIS are:

a. **Shapefile**

- A shapefile is a type of file used to store geographical data.
- It is a common data format for storing spatial data in geographic information systems (GIS) software.
- Shapefiles can store data about the shapes of geographical features, such as points, lines, and polygons, as well as their associated attributes, such as metadata, descriptions, and other information.
- It is one of the most common vector data formats; an open source format designed originally to be used with ArcGIS software.
- A shapefile typically contains a collection of files with the same name but different file extensions. The most common files included in a shapefile are:

i .shp

- This is the main file that contains the spatial data for the shapefile.
- It stores the geometry for the shapefile, such as the coordinates for the points, lines, and polygons.

Vector Data file formats

a. Shapefile

ii .dbf

- This file stores the attribute data for the shapefile.
- Attribute data is information about the features in the shapefile, such as their names, descriptions, or other characteristics.

iii .prj

- This file stores the coordinate system and projection information for the shapefile.
- This is used to ensure that the spatial data is displayed correctly on a map.

iv .shx

- This file stores the index for the spatial data in the .shp file.
- It is used to access spatial data more efficiently.

In addition to these files, a shapefile may also include others, such as .sbn and .sbx files that store the spatial index for the shapefile.

Vector Data file formats

b. KML

- KML stands for Keyhole Markup Language.
- This GIS format is XML-based and is primarily used for Google Earth.
- File extension => .kml

c. GPX

- GPX stands for GPS Exchange Format (GPX).
- It is an XML (Extensible Markup Language) schema that describes waypoints, tracks, and routes captured from a GPS receiver.
- File extension => .gpx

Advantage of vector data model

- Vector data model is efficient for topological relationship allowing spatial analysis such as proximity and network analysis;
- Easy coordinate transformation and projection;
- Vector data models tend to be better representations of reality due to the accuracy and precision of points, lines, and polygons over the regularly spaced grid cells of the raster model.
- Easy retrieval, updating and generalization of graphics;
- As each coordinate pair associated with a point, line, and polygon represents an infinitesimally exact location, zooming deep into a vector image does not change the view of a vector graphic in the way that it does a raster graphic.
- Graphic output is usually aesthetically more pleasing.
- Vector data tend to be more compact in the data structure, so file sizes are typically much smaller than their raster.

Disadvantage

- Complex data structure;
- Inefficient to describe/represent continuous data or data having high spatial variability.
- Not compatible with remote sensing imageries;

Vector Data Models

Applications of vector data model

Vector data model is widely used in GIS. Some of the applications while using vector data are:

1. Mapping:

- The vector data model is used to create and update digital maps for various purposes, such as urban planning, environmental management, transportation, and tourism.

2. Spatial analysis:

i. Overlay and Neighbourhood function

- Overlay includes dissolve, merge, clip, intersection operations.
- Neighbourhood function involves buffer zone delineation.

ii. Network analysis function: Routing, Shortest distance calculation etc.

3. Remote sensing:

- The vector data model is used to represent and analyze satellite and aerial imagery, which can be used for various applications, such as land use classification, crop monitoring, and disaster response.

Applications of vector data model

Vector data model is widely used in GIS. Some of the applications while using vector data are:

4. Emergency management:

- Vector data is used in emergency management to identify the location and extent of hazards, such as floods, fires, and earthquakes. This helps in developing evacuation plans and response strategies.

5. Urban Planning:

- Vector data models are used to represent land use, zoning, and building footprints. This data can be used to plan new developments, manage growth, and improve urban design.

Continous field vs Discrete field

Continous field

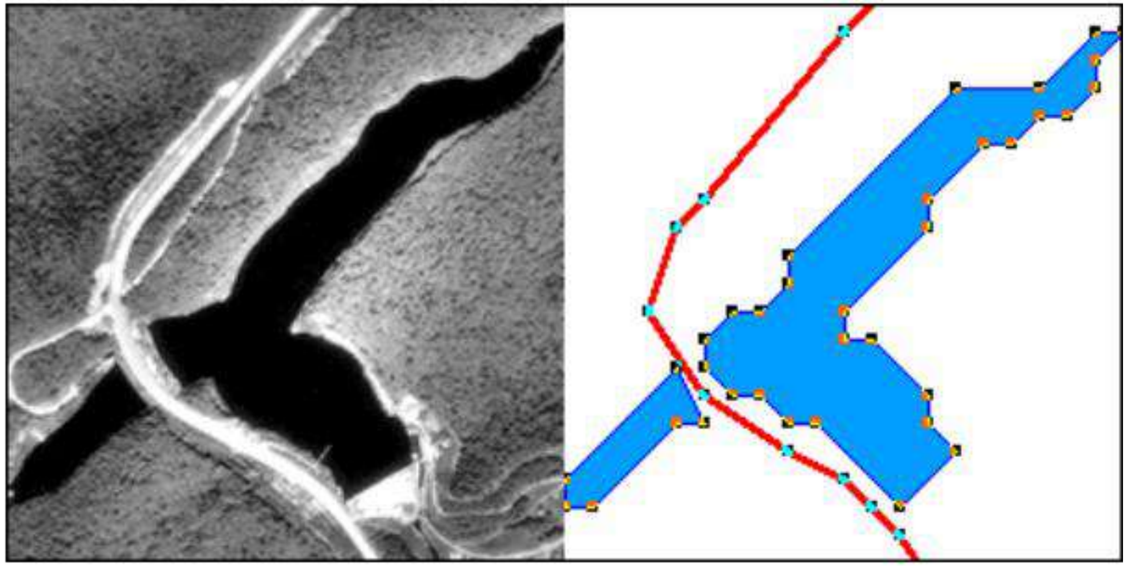
- A continuous field in GIS refers to a spatial data layer where the attribute values change gradually across space, forming a continuous surface. This type of data is typically represented by continuous variables such as elevation, temperature, precipitation, or population density.
- Continuous fields are often represented using raster data format, where the surface is divided into a regular grid of cells, and each cell contains a value representing the attribute of interest.
- The values can vary smoothly across the cells, and interpolation techniques can be used to estimate values at locations where data is not available.

Discrete field

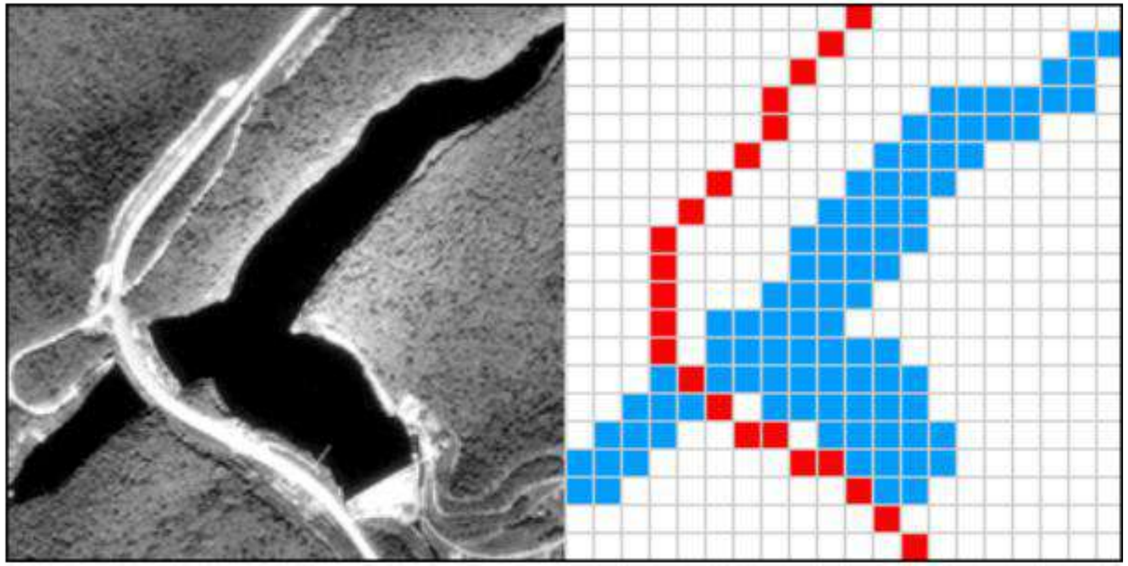
- A discrete field in GIS, on the other hand, refers to a spatial data layer where the attribute values are distinct and discrete, with no gradual change across space. This type of data is typically represented by categorical variables such as land use/land cover, land ownership, or types of vegetation.
- Discrete fields are often represented using vector data format, where discrete objects or points are used to represent the features of interest, and each object or point is associated with a specific attribute value.

Vector vs Raster approach of data representation

A



B



Source: https://www.tamiu.edu/cees/courses/fall2018/geol4460_labs/lecture6.pdf

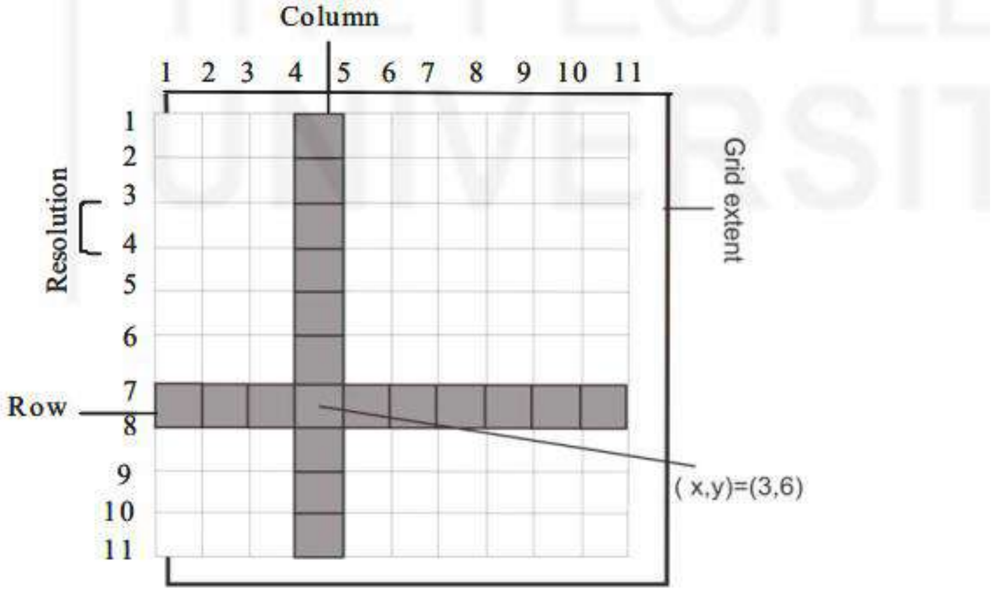
Fig: Representation of road and reservoir, (A) Vector approach and (B) raster approach

2.2. Spatial Data Models

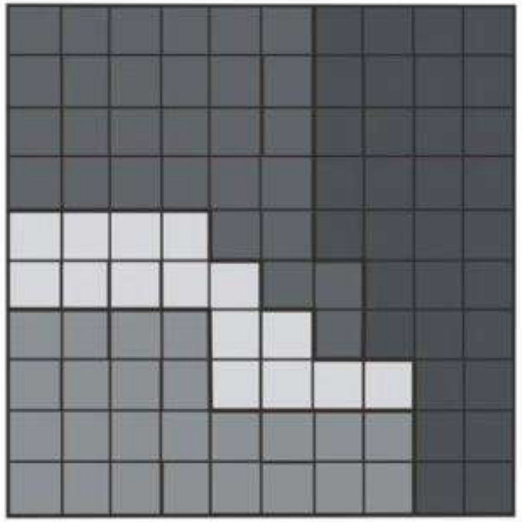
Raster data model (Field based model)

- The raster data model is composed of a regular grid of cells in specific sequence and each cell within a grid holds data.
- Each cell (also called pixel) contains a single value and is independently addressed with the value of an attribute. One set of cells and associated value is a layer.
- Cells are arranged in layers. A data set can be composed of many layers covering the same geographical areas e.g., water, paddy, forest, cashew.
- Raster data represent points with single cells, lines with sequences of neighboring cells, and polygons with collections of contiguous cells.
- Rasters are especially well suited for storing continuous data such as temperature and elevation values, but can hold discrete and categorical data such as land use as well.
- Raster can be Aerial photographs, satellite imagery, scanned maps etc.

Raster data model (Field based model)



(a)



(b)

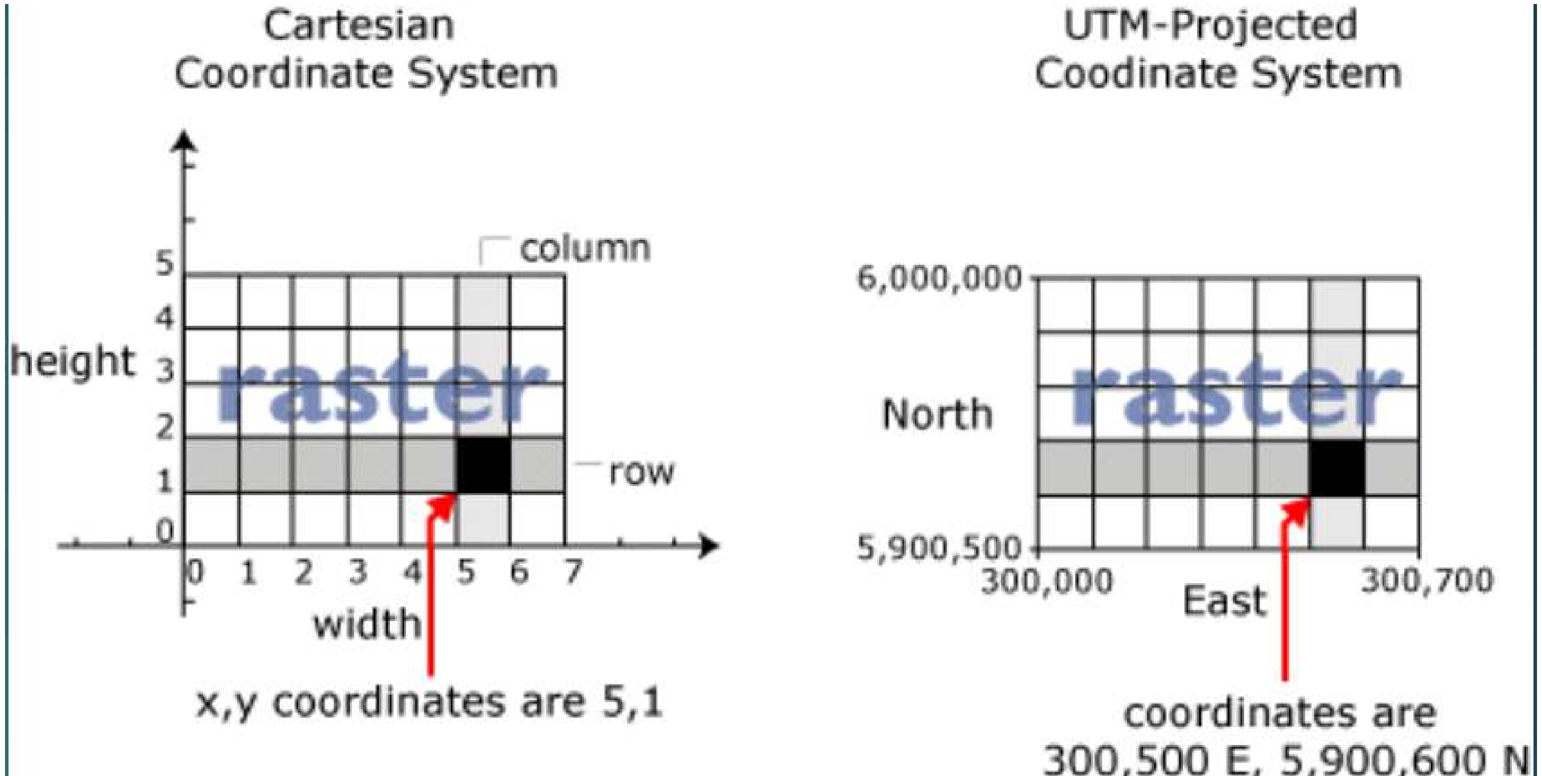
	Value	Count	Type
	19	15	Water
	38	29	Paddy
	62	32	Forest
	56	24	Cashew

Fig: (a) raster grid matrix with their cell location and coordinates, and (b) raster grid and its attribute table

Source: https://www.egyankosh.ac.in/bitstream/123456789/39550/1/Unit-5.pdf

Elements in Raster data model

- The rows and columns origin is the lower left Corner of the grid.
- The Rows Function as y – coordinates
- The Column functions as X – coordinates
- In line with rows and column in the raster, a cell is defined by its location.
- Resolution depends on grid cell size.



Raster data model (Field based model)

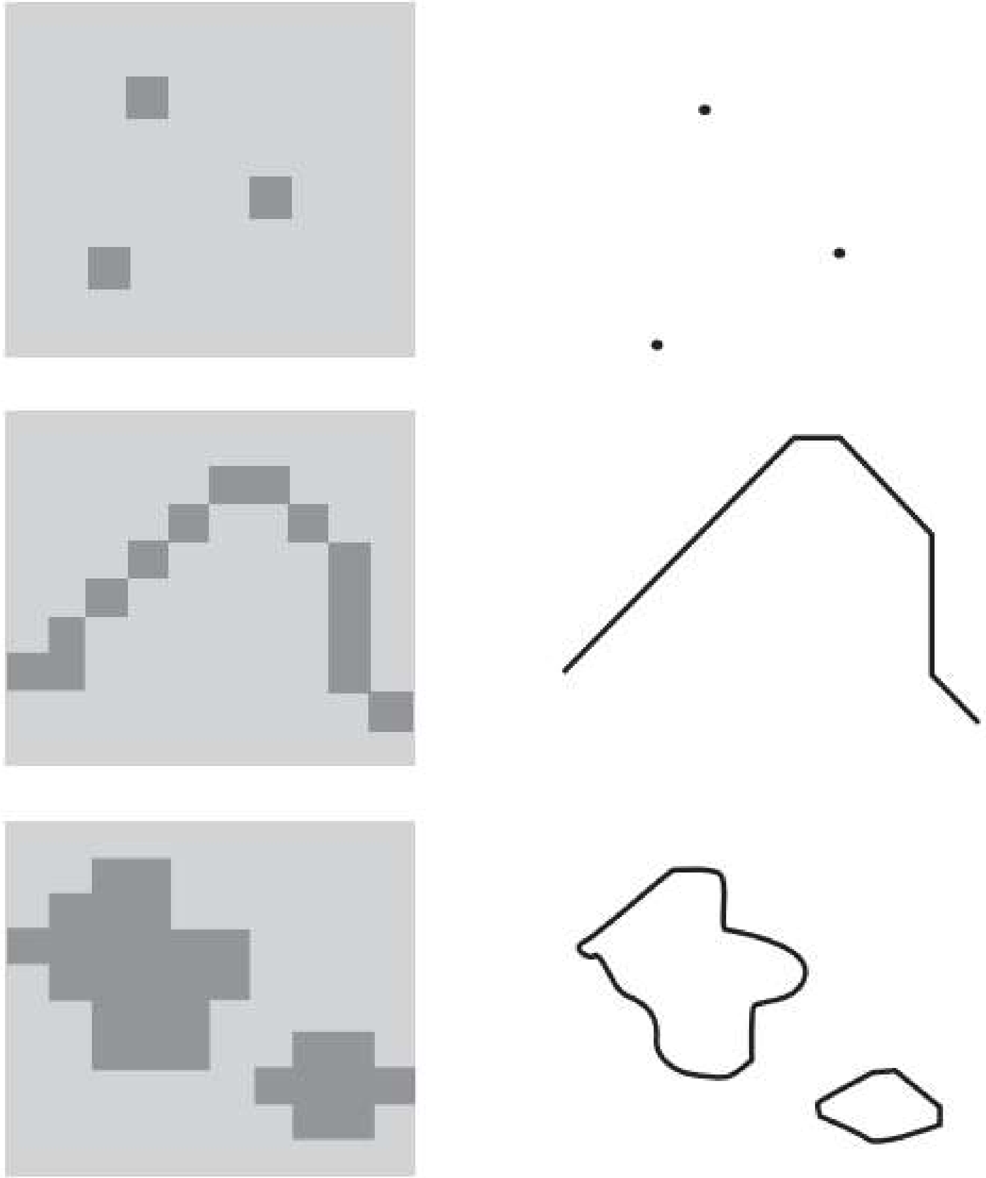


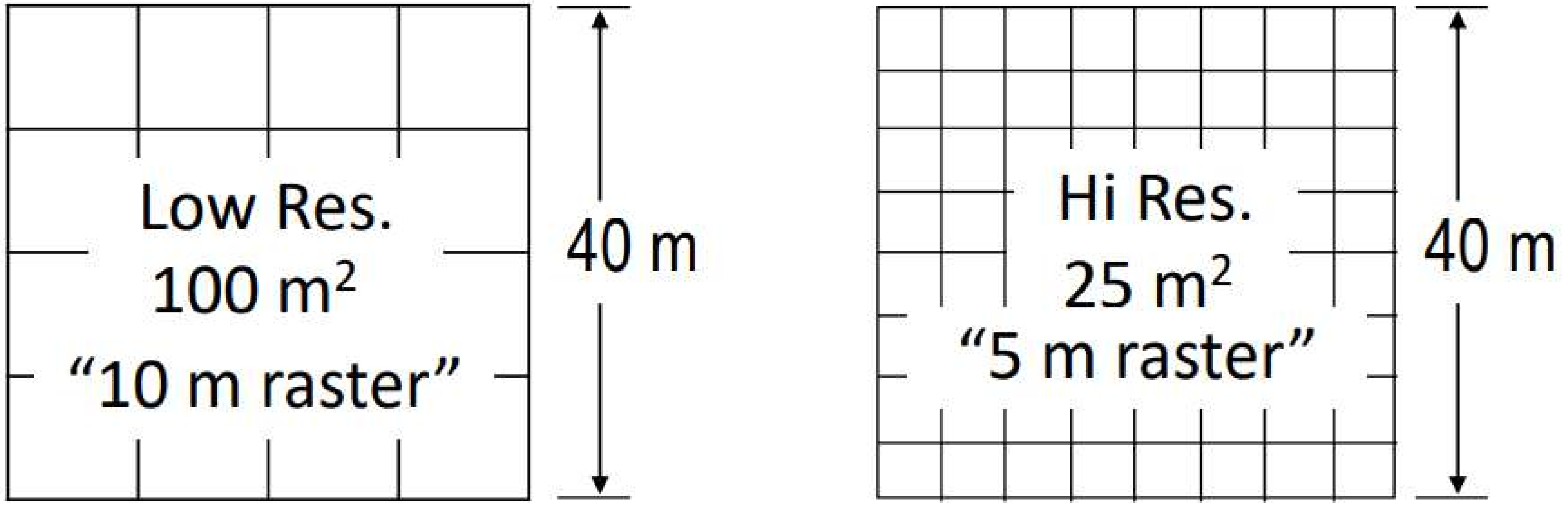
Fig: Representation of point, line, and polygon features:
raster format on the left and vector format on the right.

Source: chang, kang-tsung (2006). introduction to geographic information systems.

Raster data model (Field based model)

Resolution of raster dataset

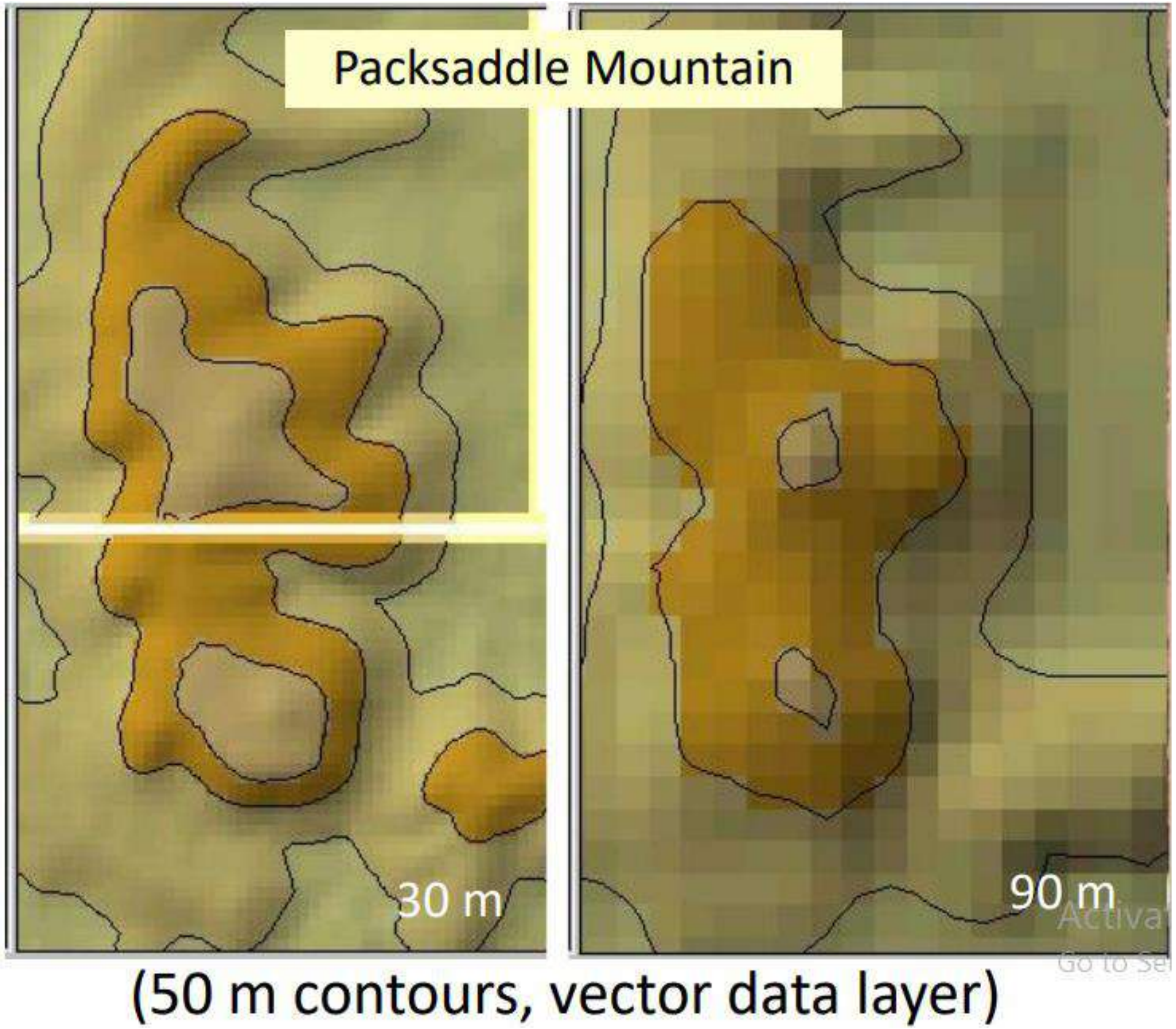
- In GIS, resolution of a raster dataset simply means the spatial resolution.
- It is defined by area or dimension of each cell.
- Spatial Resolution = (cell height) X (cell width)
- High resolution: cell represent small area.
- Low resolution: cell represent larger area.
- For example, a DEM of 5m resolution means its one pixel represents 5 by 5 metres on the ground.



Source: https://www.geo.utexas.edu/courses/371c/lectures/Spring22/Raster_datamodel_Spring22.pdf

Raster data model (Field based model)

Resolution of raster dataset



Source: https://www.geo.utexas.edu/courses/371c/lectures/Spring22/Raster_datamodel_Spring22.pdf

Fig: Resolution of 30 m data is 9 times better (9X as many pixels) than 90 m data.

Assignment

1. Write about various raster data file formats.

Raster data model (Field based model)

Raster data structure types

On basis of grid cells there are mainly two types of raster data structure. They are:

i. Regular Tessellation

ii. Irregular Tessellation

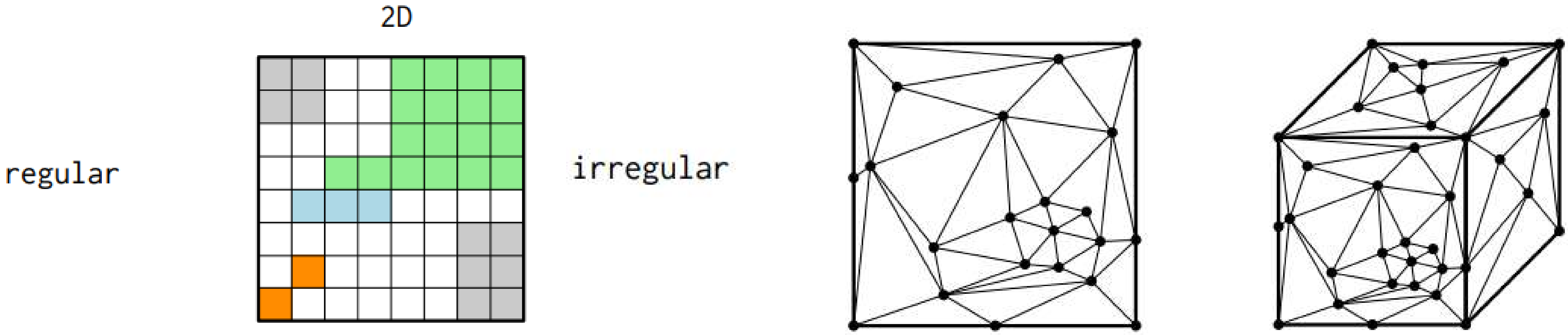
Regular Tessellation

- In a regular tessellation, the grid cells are uniform in size and shape, and the grid lines are parallel to each other.
- Examples of regular tessellation include the grid-based data structure used in digital elevation models (DEM) or satellite imagery.

Raster data model (Field based model)

Irregular Tessellation

- In an irregular tessellation, the grid cells are of varying shapes and sizes, and the grid lines are not parallel to each other.
- This type of tessellation is commonly used in applications where the spatial distribution of data is irregular, such as in hydrology, geology, and soil science.
- Examples of irregular tessellation include triangulated irregular networks (TIN) or Voronoi diagrams.



Source: https://3d.bk.tudelft.nl/hledoux/pdfs/17_aag_fields.pdf

Raster data model (Field based model)

Advantage

- It has got simple data structure.
- Overlay operations are easily and efficiently implemented.

Disadvantage

- It lacks the vector model's precision in representing the location and boundary of spatial features.
- File size: Raster files are typically very large.
- It is not suitable for some types of spatial analysis when multiple rasters of different scale and resolution are used.
- Changing map projections will alter the size and shape of the original input layer and frequently result in the loss or addition of pixels in output layer.
- The output of graphics is less aesthetically pleasing because boundaries tend to have a blocky appearance rather than the smooth lines of hand-drawn maps.

Raster data model (Field based model)

Applications:

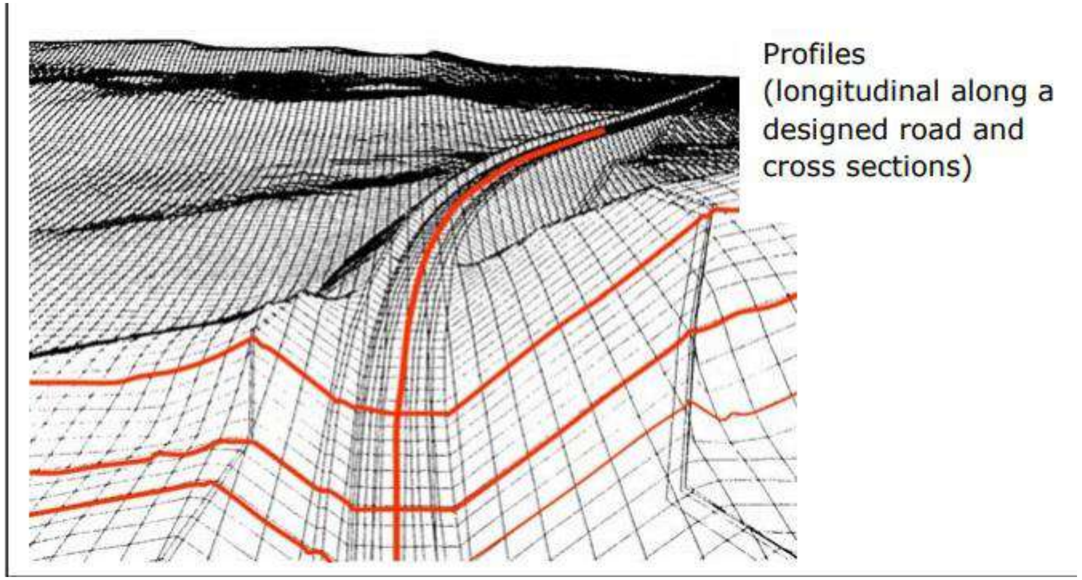
1. Map Making:

- Rasters like satellite imageries are digitalized for making topographic maps.

2. Civil Engineering

- It is used in the design, (construction and maintenance) of roads, railways, canals etc.
- Cut-Fill estimation: Digital Elevation Model (DEM) can also be used in analysis of cut and fill for construction projects like road, irrigation etc.

.



Raster data model (Field based model)

Applications:

3. 3D analysis:

- Rasters like DEM is used for generating slope, aspect, hillshade and other surface analysis.

4. Natural resource management:

- The raster data model is used in natural resource management applications such as forestry and agriculture. For example, satellite imagery can be used to map the extent of forest cover and monitor changes over time.

5. Land use/land cover mapping:

- The raster data model is used to classify and map different types of land use and land cover features such as forests, croplands, urban areas, water bodies, and wetlands.

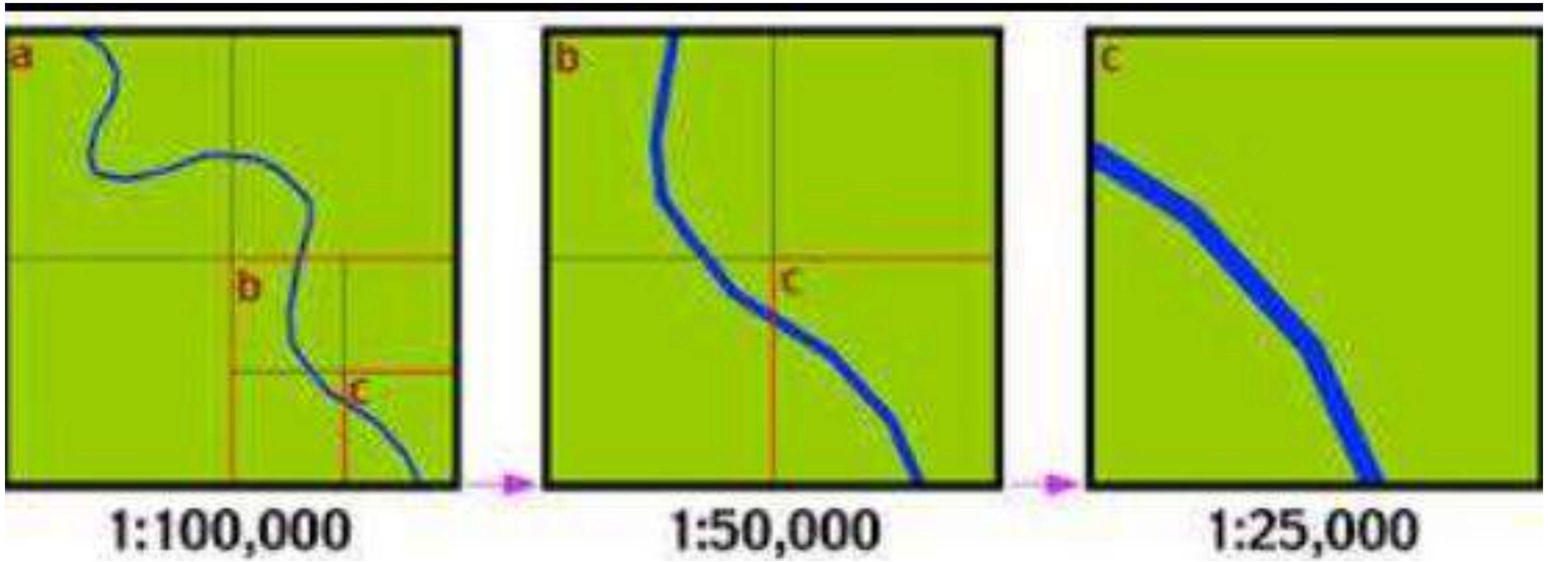
.

Scale

- It is the ratio of map distance to the distance on the ground.

Scale = map distance/ ground distance

- For example, a map of 1: 1200 means 1 cm on map will have 1200cm (i.e 12m) on ground.
- Small scale map means having ratio value small and hence there is less detail in the map.
- Large scale map means having ratio value large and hence there is more detail in the map.



TIN Data Models

Triangular Irregular Network (TIN)

- TIN data models (Triangulated Irregular Network) is a type of GIS data model that represents terrain or other surface features as a set of irregularly spaced points.
- These points are then connected by lines, which form a adjacent non-overlapping network of triangles.
- It is a vector based representation of land or water surface.
- It is based on Delaunay principle.
- TIN data models are useful for applications involving terrain analysis, such as hydrological modeling, land-use planning, and environmental impact assessment.
- Some examples of TIN data models are Esri TIN and GRASS TIN.

Triangular Irregular Network (TIN)

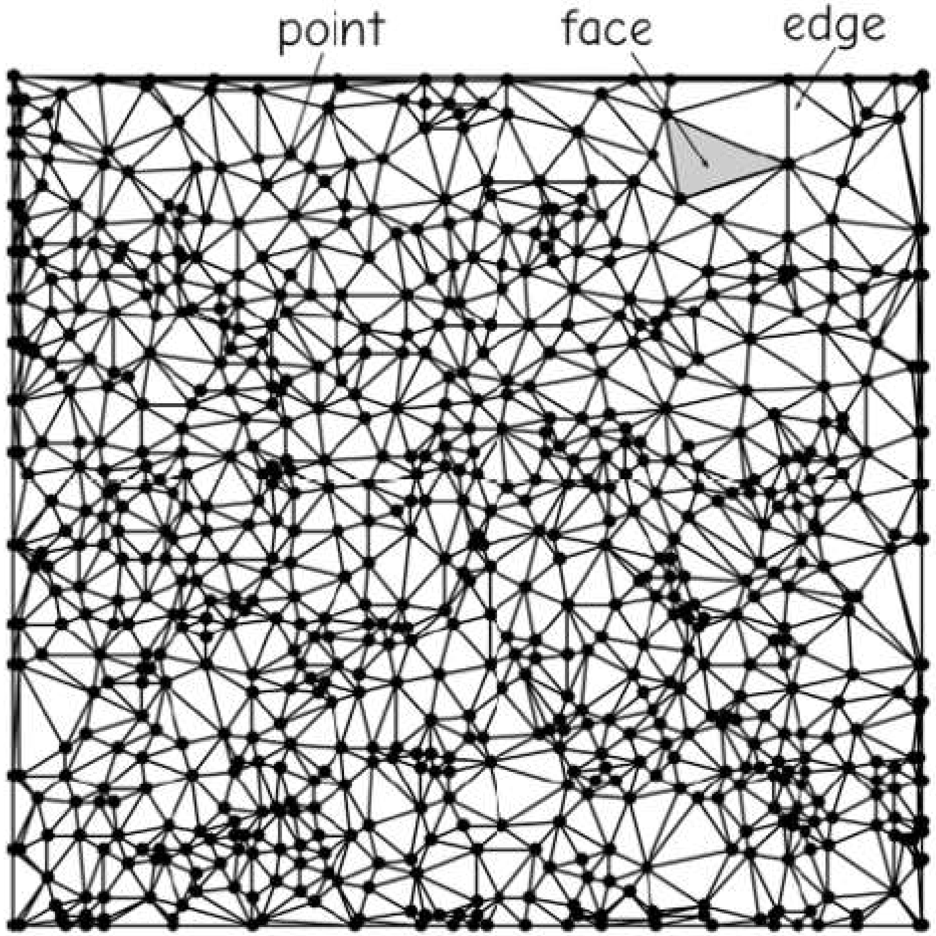
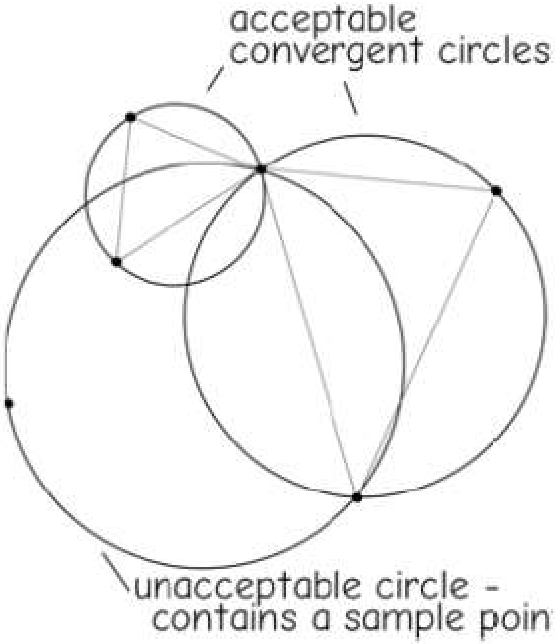


Figure A TIN data model defines a set of adjacent triangles over a sample space (left). Sample points, facets, and edges are components of TIN data models. Triangles are placed by convergent circles. These intersect the vertices of a triangle and contain no other possible vertices (below).



Source: https://geolearn.in/gis-data-models/#TIN_Data_Model

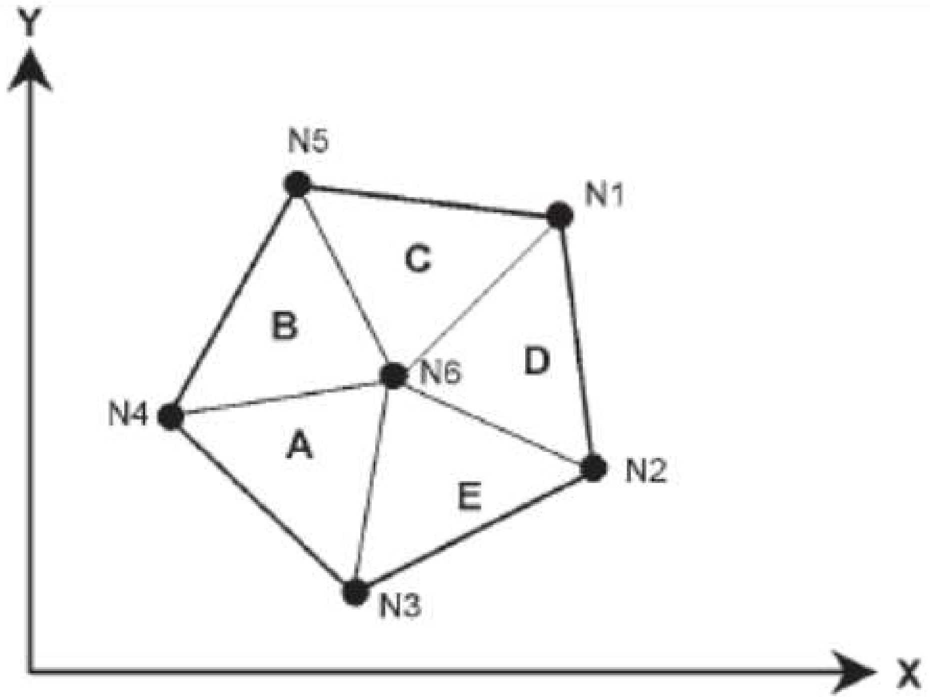
Watch: https://www.youtube.com/watch?v=GctAunEuHt4&t=68s&ab_channel=SCIco

TIN Data Models

TIN data structure

- A triangulated irregular network (TIN) is a data model commonly used to represent terrain heights.
- Typically the x, y, and z locations for measured points are entered into the TIN data model. These points are distributed in space, and the points may be connected in such a manner that the smallest triangle formed from any three points may be constructed.
- The TIN forms a connected network of triangles.
- Triangles are created such that the lines from one triangle do not cross the lines of another.
- Line crossings are avoided by identifying the convergent circle for a set of three points. The convergent circle is defined as the circle passing through all three points. A triangle is drawn only if the corresponding convergent circle contains no other sampling points.
- Each triangle defines a terrain surface, or facet, assumed to be of uniform slope and aspect over the triangle.
- The TIN model typically uses some form of indexing to connect neighbouring points. Each edge of a triangle connects to two points, which in turn each connect to other edges. These connections continue recursively until the entire network is spanned. Thus, the TIN is a rather more complicated data model than the simple raster grid when the objective is terrain representation.

TIN data structure



Triangle Table and Node

Triangle	Adjacent	Nodes
A	B, E	N3, N4, N6
B	A, C	N4, N5, N6
C	B, D	N1, N5, N6
D	C, E	N1, N2, N6
E	D, A	N2, N3, N6

Coordinate Table

Node	Coordinates
N1	X1, Y1, Z1
N2	X2, Y2, Z2
N3	X3, Y3, Z3
N4	X4, Y4, Z4
N5	X5, Y5, Z5
N6	X6, Y6, Z6

Source: Lecture notes by Bikash Sherchan

TIN Data Models

Advantages of TIN

- It can represent terrain roughness or rugged terrain more clearly than DEM by changing shape, size according to varying terrain.
- Can represent ridges, valley more accurately than DEM.
- Contains slope and aspect info of each triangles as attribute value.
- File sizes are typically much smaller since small number of data points are needed to represent a surface.

Disadvantages of TIN

- TIN is a rather more complicated data model than the simple raster grid.
- Processing of tin takes more time than normal.
- Errors around the edges of TIN can be found.
- Any portion of dem can be extracted or clipped but can't be done in case of TIN.
- It can be difficult to integrate TINs with other data types in a GIS environment.

TIN Data Models

Applications of TIN

- TIN is used in case of rugged terrain; highly varying landscape like in himalayas regular raster grid can't contain complete elevation information.

- TIN data models are useful for applications involving terrain analysis, such as hydrological modeling, land-use planning, and environmental impact assessment.

- Used in animation, games etc. in constructing a 3D object/character with addition of different colors, textures.

- Map Making:

  - TINs are used to display topographic maps by showing the shape of the terrain.
  - Large the triangles higher the terrain variation.

- Landslides:

  - TINs are the most popular and most applied visualization for all types of landslides since there is high variation in elevation and raster can't represent it properly.

References

- Data model concept:
- https://www.egyankosh.ac.in/bitstream/123456789/39550/1/Unit-5.pdf
- http://www.geo.umass.edu/courses/geo494a/Chapter2_GIS_Fundamentals.pdf
- Vector Data Model:
- https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s08-02-vector-data-models.html
- https://gistbok.ucgis.org/bok-topics/vector-formats-and-sources
- Shapefile: https://mapscaping.com/what-is-a-shp-file/
- Vector/Raster data structure: https://2012books.lardbucket.org/books/geographic-information-system-basics/s08-data-models-for-gis.html

https://open.maricopa.edu/gist/part/chapter-6-introduction/

https://chat.openai.com/

Raster data model: https://www.geo.utexas.edu/courses/371c/lectures/Spring22/Raster_datamodel_Spring22.pdf

Regular/Irregular tessellation: https://3d.bk.tudelft.nl/hledoux/pdfs/17_aag_fields.pdf

TIN: https://gistbok.ucgis.org/bok-topics/triangular-irregular-network-tin-models

https://civilstuff.com/what-is-triangulated-irregular-network-tin/

GIS BASICS

EG 2105 GE

Unit 3: Spatial Data Acquisition and Preparation

Lecture by

Er. Keshav Raj Bhusal

Public Namuna Secondary School

Table of contents

3.1 Sources of Spatial Data

Sources of Spatial Data

Primary Data Sources / Primary Data Capture

- A primary data source is an original data source, that is, one in which the data are collected firsthand by the researcher for a specific research purpose or project.
- Primary data collection is quite expensive and time consuming compared to secondary data collection.
- The most common techniques of primary data sources are:

i. Field based technique: Plane Table, Total Station, GPS, DGPS

ii. Air-based Technique: Photogrammetry, UAV, LiDAR

iii. Space based Technique: Remote Sensing

3.1 Sources of Spatial Data

Primary Data Sources / Primary Data Capture

Field based technique:

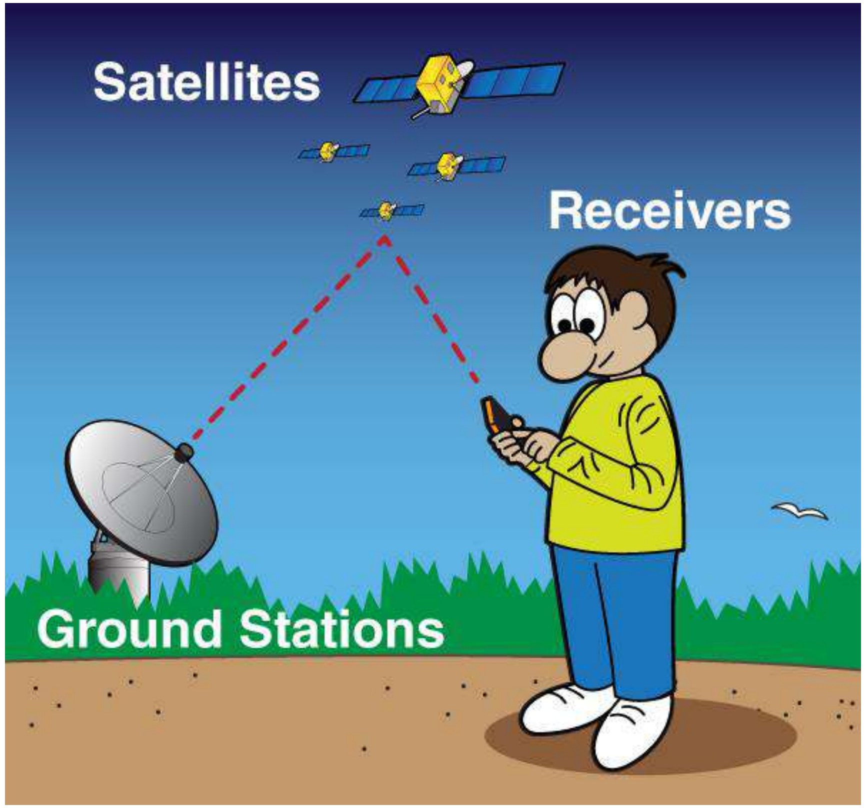
1. GPS

- A global positioning system (GPS) receiver can determine its precise position on the Earth's surface using satellites in space as reference points.
- It is a technique to capture vector data.
- GPS units can collect coordinates and then create GIS layers with them. They can store points (called waypoints), or capture the path followed by the user (called a track).
- GPS are used for verifying, maintaining, and updating the GIS data. It is also an excellent tool for validating geographic features, updating their attributes, and collecting new geographic features.
- GPS can be used in many diverse GIS data capture applications like road centreline mapping, utility pole mapping, and wetland boundary mapping.
- Depending upon the accuracy required different kind of GPS receivers like code-based differential GPS receivers, phase differential GPS, handheld GPS receiver or even smartphones receivers can be used.
- Handheld GPS receiver is a common low cost portable GPS receiver used for GIS task which requires less precision.

Primary Data Sources / Primary Data Capture

Field based technique:

1. GPS



Source: https://spaceplace.nasa.gov/gps/en/

Fig: GPS is made up of three parts: satellites, ground stations, and receivers.

Source: https://www.forestry-suppliers.com/p/39470/77651/garmin-etrex-10-gps-receiver

Fig: showing handheld GPS receiver

Primary Data Sources / Primary Data Capture

Field based technique:

2. DGPS

- DGPS stands for Differential Global Positioning System, is method of increasing the accuracy of positions derived from GPS receivers.
- Normally, GPS system based on the satellite technology can have the nominal accuracy of 15 meters whereas using DGPS can bring this accuracy nearly around 10cm.

How DGPS works?

- In DGPS two or more gnss receivers are used where one with known coordinate called base station and other rover. The base station sends the corrected GPS signals to rover and hence collected data is more precise.

Advantages:

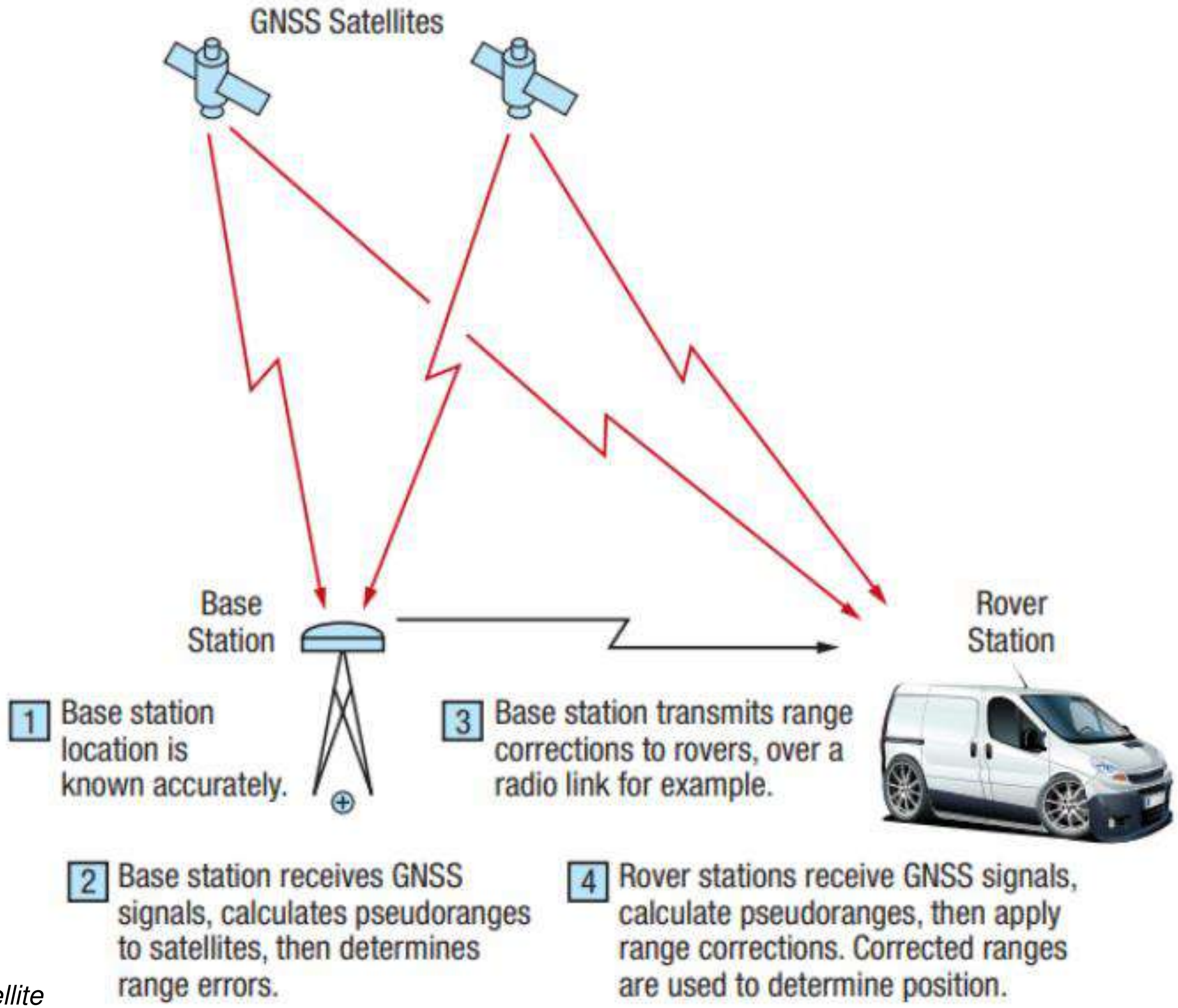
- Provides highly precise data.
- Data collection is faster.

Primary Data Sources/Primary Data Capture

Field based technique:

2. DGPS

Source: http://mapadda.com/



Source: Satellite Positioning; Dinesh Mani Bhandari

3.1 Sources of Spatial Data

Primary Data Sources/ Primary Data Capture

Field based technique:

3. Total Station

- A total station is an electronic/optical instrument based on the combination of electronic theodolite (for measuring horizontal & vertical angles), an electromagnetic distance measurement (EDM) device (for measuring distances) that records 3D coordinates.
- The digital data collected from TS is vector data.
- Once data is collected from TS it is transferred to computer using cable or other medium. It is then loaded into GIS and is used for making topographic maps, contour and other various GIS analysis.

3.1 Sources of Spatial Data

Primary Data Sources/ Primary Data Capture

Field based technique: Total Station

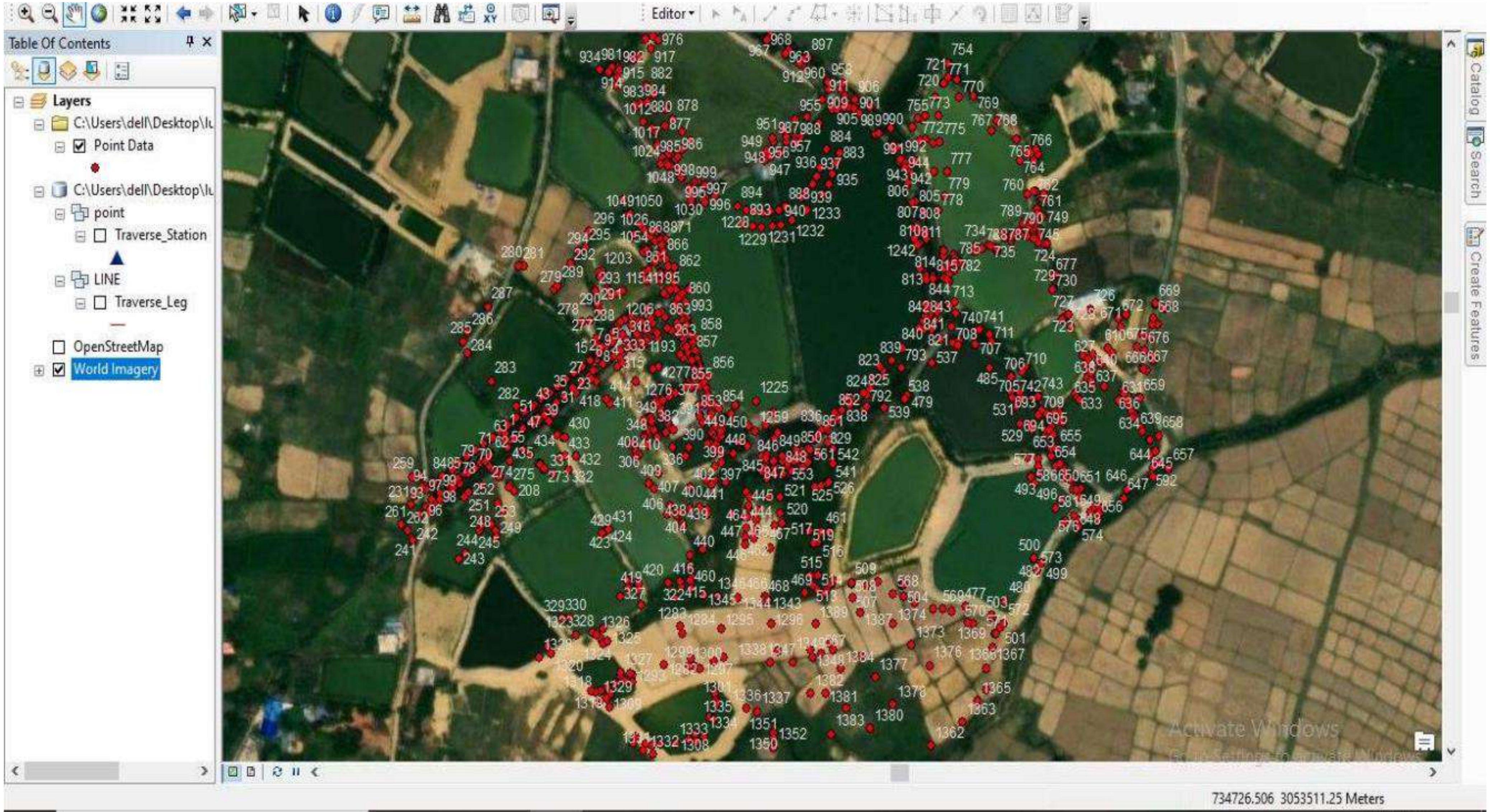


Fig: Data collected(red one) from TS loaded in ArcGIS

Source: http://mapadda.com/

3.1 Sources of Spatial Data

Primary Data Sources/ Primary Data Capture

Air-based Technique

1. Photogrammetry

- It is the art, science and technology of obtaining reliable information about the physical objects and environment through the process of recording , measuring; and interpreting photographic images and patterns of recorded radiant electromagnetic energy.
- The photographic images can be collected through Unmanned Aerial Vehicle (UAV), aircrafts with camera mounted etc.
- The end product of photogrammetry are orthomosaic, Digital Elevation Model (DEM), Digital Terrain Model(DTM) etc. which are later used in GIS for making topographic maps and different sorts of analysis.

Primary Data Sources/ Primary Data Capture

Air-based Technique: Photogrammetry

Fig: Orthomosaic of PNSS obtained after processing 222 images.

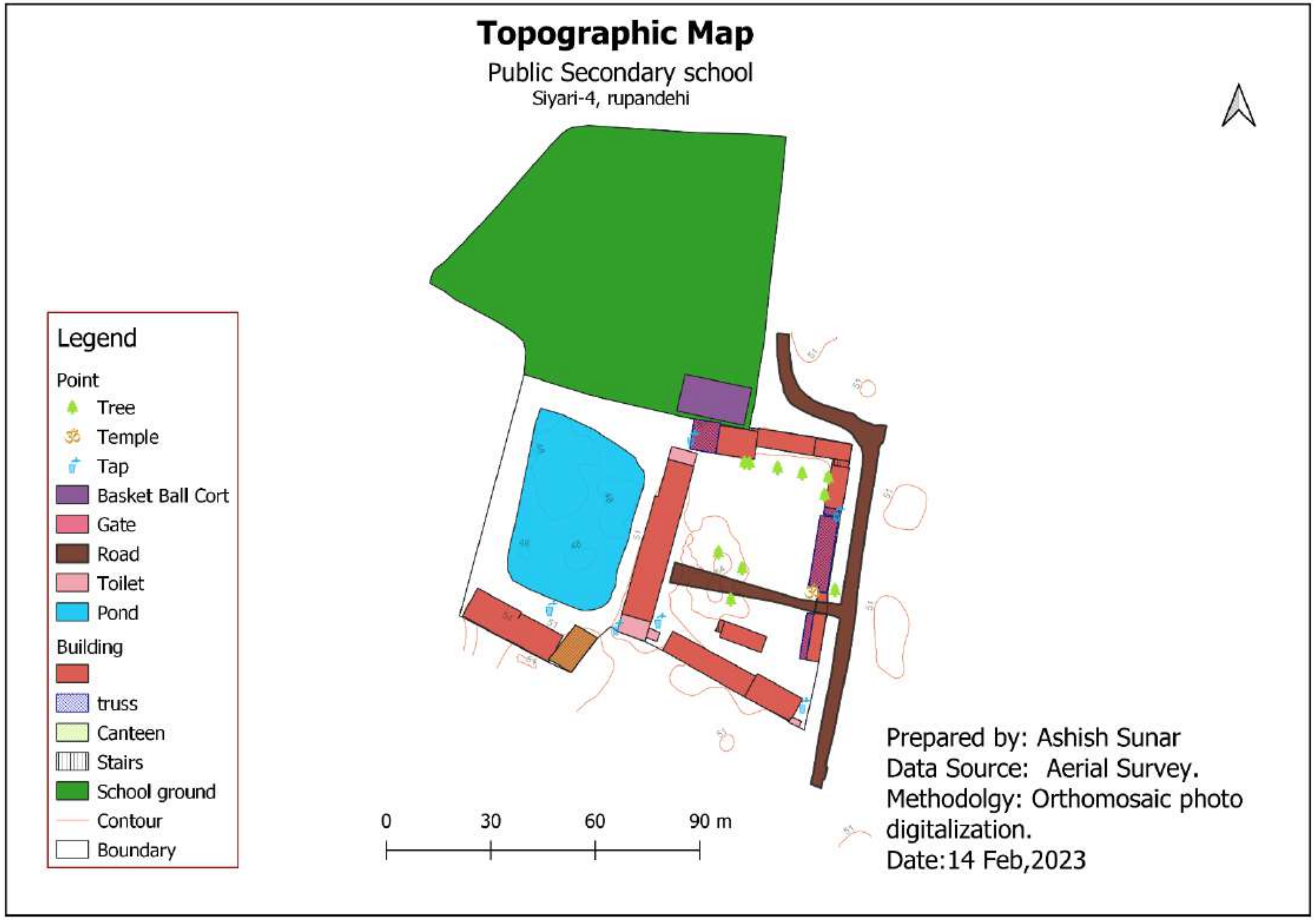


Fig: Topographic map obtained by using digitalizing Orthomosaic of PNSS.

Source: http://mapadda.com/

Primary Data Sources/ Primary Data Capture

Air-based Technique

LiDAR (Air-borne LiDAR)

- LiDAR stands for Light Detection and Ranging. It uses a laser scanner mounted on an aircraft or helicopter along with GPS technology to collect data about the Earth's surface.
- Through LiDAR DEM, contour, DTM and point clouds are obtained which are used as GIS data layer.

Lidar data can be used in GIS in several ways, including:

i. Terrain modeling: Lidar data can be used to create digital elevation models (DEMs) of the Earth's surface. These DEMs can be used to create 3D visualizations of the terrain, calculate slope and aspect, and generate contour lines.

ii. Flood mapping: Lidar data can be used to map flood zones by creating a digital surface model (DSM) and subtracting a digital terrain model (DTM) to create a digital flood model (DFM).

iii. Natural resource management: Lidar data can be used to map and monitor forests, wetlands, and other natural resources, which can help with conservation and management efforts.

Primary Data Sources/ Primary Data Capture

Air-based Technique

LiDAR (Air-borne LiDAR)

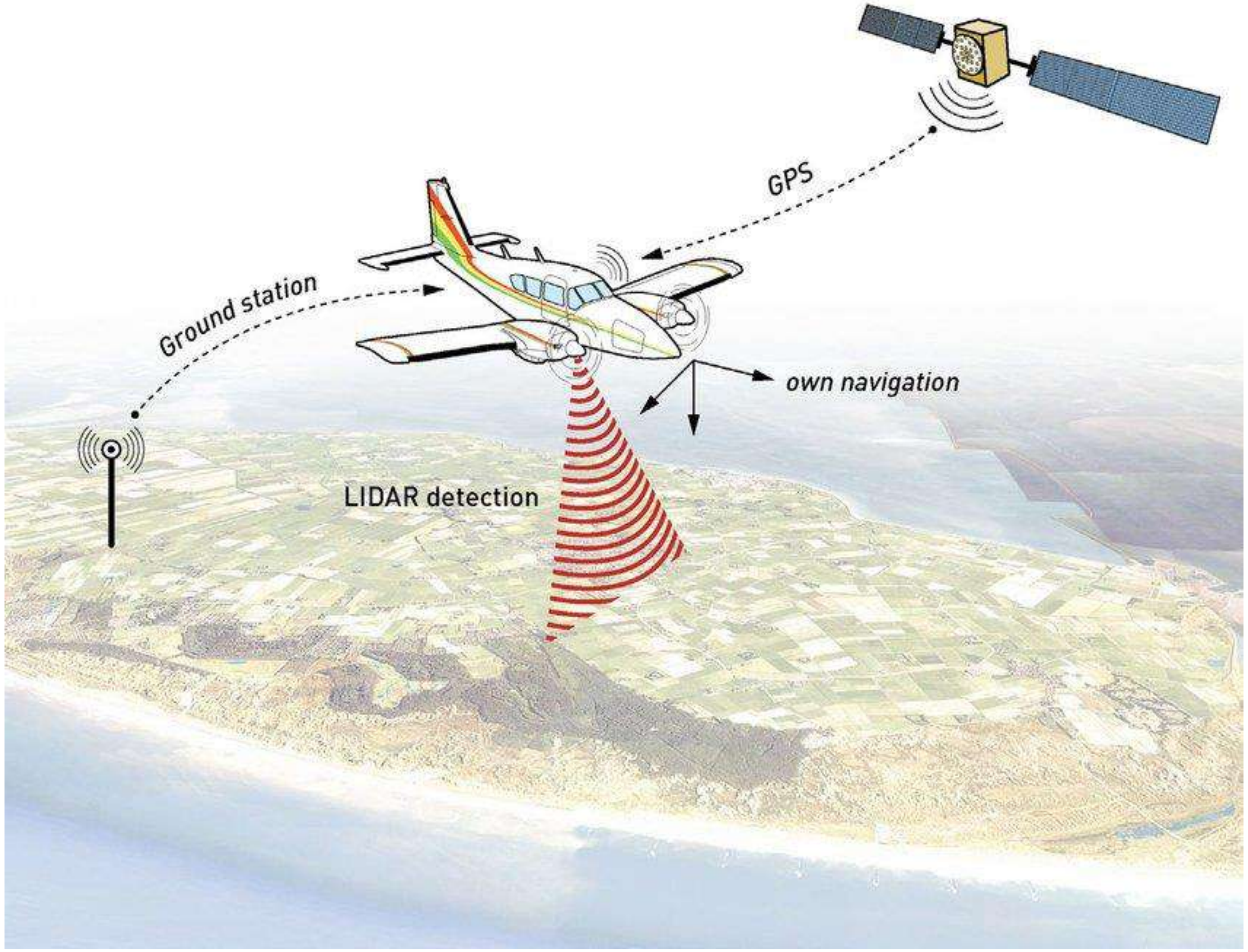


Fig: showing concept of Lidar data collection.

https://www.researchgate.net/figure/Collecting-liDar-data-from-an-aircraft-authors_fig1_323337401

Primary Data Sources/ Primary Data Capture

Space based Technique

Remote Sensing

- It is the art, science and technology of acquisition of information about an distant object or phenomenon without making physical contact with the object.
- It one of the most popular methods for gathering raster data.
- This type of GIS data collection is primarily carried out by satellites and aircraft sensors, which can assess the characteristics of a location by measuring the electromagnetic radiation objects on the surface emit.
- The data collected through remote sensing can be used to create various GIS layers, such as land cover maps, vegetation indices, elevation models, and surface temperature maps. These layers can be integrated with other GIS data, such as demographic data or infrastructure data, to create complex maps and spatial analysis.

Primary Data Sources/ Primary Data Capture

Space based Technique

Remote Sensing

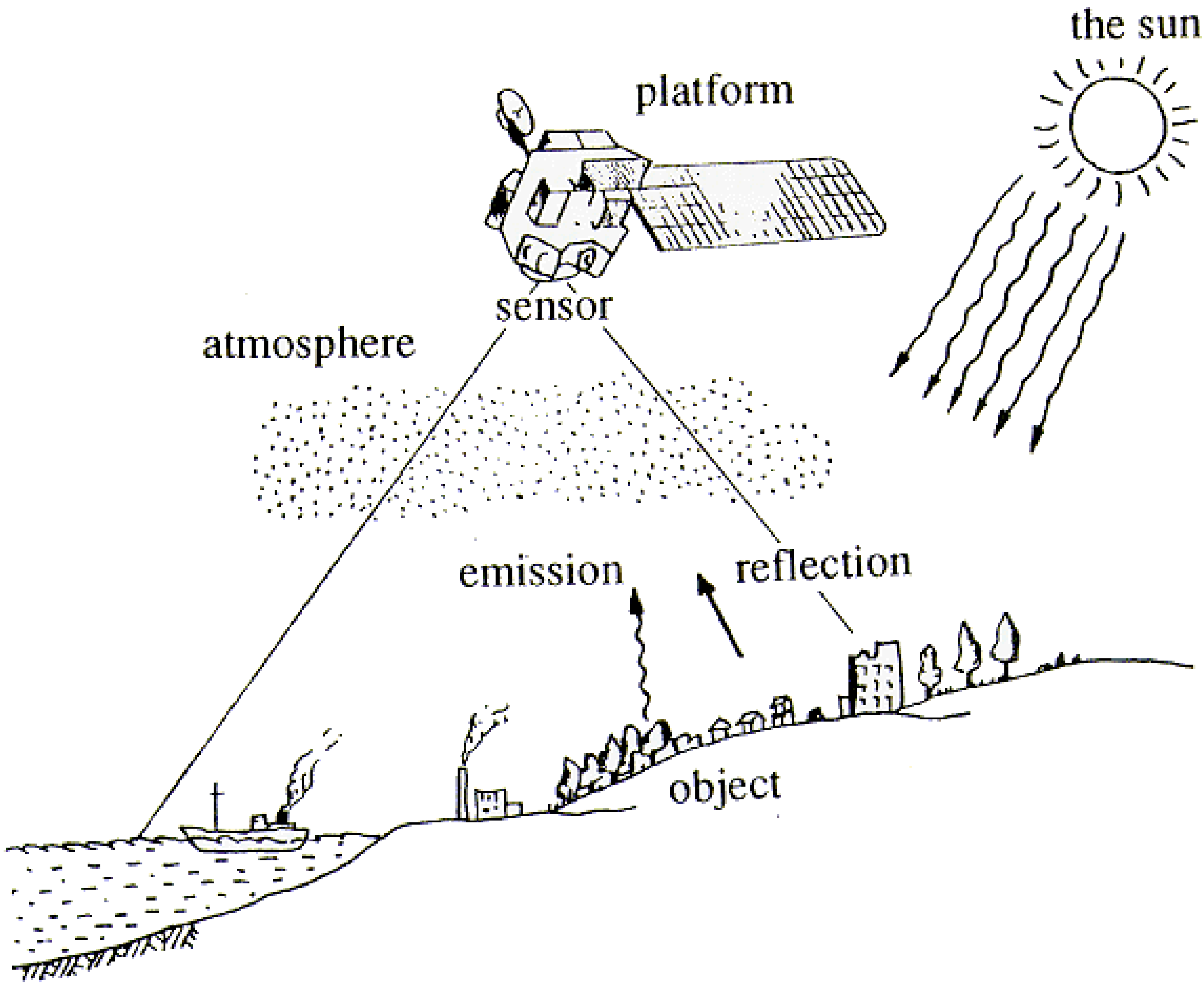


Fig: showing concept of remote sensing data collection.

Source: https://www.google.com/search?q=remote+sensing+collecting+data&tbm=isch&ved=2ahUKEwjVzr2apdn-AhXO0qACHRjFBgAQ2-cCegQIABAA&oq=remote+sensing+collecting+data&gs_lcp=CgNpbWcQAzoGCAAQBxAeOggIABAIEAcQHlC2CFjuJWC2K2gAcAB4AYABggSIAdgbkgELMC43LjYuMi4wLjGYAQCgAQGqAQtnd3Mtd2l6LWltZ8ABAQ&sclient=img&ei=a2ZSZJXtJM6lg8UPmlob&bih=722&biw=1536#imgrc=xi75oB4xRx1noM

Secondary Data Sources/ Secondary Data Capture

- Secondary data refers to data that is collected by someone other than the primary user.
- There are mainly two secondary data sources. They are:

i. Existing paper maps (Base and thematic maps)

ii. Data available in Web (Clearinghouse and online sources)

Existing paper maps (Base and thematic maps)

- Secondary data can be captured from existing paper maps by creating their raster and vector files.
- It involves:

i. Raster data capture

ii. Vector data capture

3.1 Sources of Spatial Data

Secondary Data Sources/ Secondary Data Capture: Existing paper maps

Raster data capture

- Raster data is simply captured by scanning the existing maps with the scanners.
- Once the digital file is created it is then georeferenced (setting a geographical context (coordinate system, control points, etc.), so the digitized elements produced are correctly referenced.

Vector data capture

- Vector data is simply captured by digitalizing the existing georeferenced map.
- It includes separating the different types of information from the map as a multiple independent layers, in GIS.
- Digitalization includes (i) Manual digitalization (ii) Automatic digitalization
- Manual digitalization: In this method an operator defines the features, trace its geometries and enter the associated attribute data.
- Automatic digitalization: In this method computer algorithm automatically analyzes the map and finds the elements that it contains, creates the corresponding vector layer elements from that.

o It is simply called vectorization.

Secondary Data Sources/ Secondary Data Capture:

Data available in Web

- These days many organization or agencies share the large amount of spatial data in web(internet) which can be downloaded by the user for a GIS project.
- To access these data you may need to pay or can get freely.
- Having data on the internet doesn't mean it would be useful for your project. You need to check the metadata before using it.
- In Nepal you can get governmental data in the web through national geoportal operated by Survey Department of Nepal.
- Some of the freely online sources of gis data are: openstreetmap, USGS Earth Explorer, humanitarian data exchange, open data Nepal etc.

Clearinghouse

- It can be defined as a service for searching, viewing, transferring, ordering, advertising and disseminating over the internet geo-data stored at many different locations in digital format.

3.2 Data Entry and Data Preparation

Data Entry

- Data entry is the procedure of encoding data into a computer-readable form and writing the data to the GIS database.
- It simply refers to the process of inputting or capturing spatial and attribute data into a GIS database.
- It typically involves converting data from physical or analog sources, such as paper documents, surveys, or forms, into a digital format.
- Data entry tasks may include typing or scanning information, verifying accuracy, and ensuring data consistency.
- The primary goal of data entry is to capture data accurately and efficiently, focusing on inputting the information as it is without much manipulation or transformation.
- Data entry operators perform these tasks, and their main responsibility is to transfer the data from its source to a digital form.

3.2 Data Entry and Data Preparation

Data Entry

- There are three types of data entry systems commonly used in a GIS. They are:

i. Manual (via typing on keyboard or importing digital files)

- Manual data entry on keyboard is done normally for attribute data and occasionally for spatial data.
- Digital files may be simple text files or binary files. Text files should have at least two columns with X and Y coordinates which allows association of it with specific geographic coordinate system.
- Binary files are usually a product of the software package associated with measuring device (like handheld GPS). They have X and Y data but in encoded format that could be read by special software.

ii. Digitizing

iii. Scanning

Data Preparation

- Data preparation (also known as data preprocessing) in GIS involves organizing, cleaning, and transforming data to make it suitable for analysis and visualization within a GIS environment.
- Spatial data preparation aims to make the acquired spatial data fit for use.
- For vector data it includes editing, such as the trimming of overshoots of lines at intersections, deleting duplicate lines, closing gaps in lines, and generating polygons.
- Data may require conversion to either vector format or raster format to match other data sets which will be used in the analysis.
- Additionally, the data preparation process includes associating attribute data with the spatial features through either manual input or reading digital attribute files into the GIS/DBMS.

Data Preparation

- Data preparation tasks include:

1. Data Cleaning: Identifying and correcting errors, inconsistencies, and missing values in the data.
2. Data Integration: Combining data from multiple sources into a unified format.
3. Data Transformation: Converting data into a consistent format, standardizing units of measurement, or normalizing variables.
4. Feature Selection: Identifying relevant features or variables for analysis and removing irrelevant or redundant ones.
5. Data Formatting: Ensuring the data remains to a specific structure.
6. Handling Missing Data: Dealing with missing values.
7. Data Aggregation: Summarizing or grouping data to a higher level of quality.

Map scanning process and Resolution

- Map scanning is one of the simplest way to convert physical map into digital formats with the help of scanner.
- It is used to convert an analog map into a raster data, which is again converted to vector format through tracing.
- Scanning automatically captures map features, text and symbols as individual cells, or pixels and produces an automated image.
- Pixel size depends upon the scanning resolution. Higher the resolution, lesser the size of pixels. The resolution in case of scanned map is represented in terms of dpi. Dpi is also called as dots per inch.
- Before scanning a document, resolution of scanning must be decided because the scanning resolution or pixel resolution or dpi is directly linked to information content, quality of information, project data quality standard and data volume or file size.
- High resolution scanning may result too noisy image and very low resolution may not separate the minor details. Therefore, scanning at different resolution should be tested to evaluate the result in order to get the optimal solution.

3.2 Data Entry and Data Preparation: THinK

Map scanning process and Resolution

Q.1: A hardcopy map is prepared at the scale of 1:750 was scanned at 500 dpi. Calculate the resultant cell size

Map scanning process and Resolution

Q.1: A hardcopy map is prepared at the scale of 1:750 was scanned at 500 dpi. Calculate the resultant cell size

Ans:

- Scale=1:750, dpi=500 dots per inch, cell size=?
- 500 dots are represented in a linear inch
- 500 pixels are represented in 1 inch
- 1 pixel is represented in (1/500) inch
- 1 pixel is represented in (2.54/500) cm
- 1 pixel is represented in 0.00508 cm on the image
- 1 pixel is represented in 0.00508*750 cm on the ground
- 1 pixel is represented in 3.81 cm on ground

*So, resultant cell size is 3.81 cm*3.81 cm*

Geo-referencing and map projection (Coordinate system)

Geo-referencing:

- Georeferencing is the process of adding geographic information to the digital image (scanned map or satellite image) such that the image gets placed in its appropriate real world location.
- It is the process of assigning real-world coordinates to each pixel of the raster.
- The process of georeferencing varies slightly based on the GIS software you are using and the characteristics of the raster data you are working with.
- If your scanned map has already got coordinates in grid markings then you can easily georeference by some simple clicks. But for map without coordinates, you need to determine the coordinates for visible features like building corners, street corners by connecting them to the same features in another map or dataset that is already georeferenced.
- It is a common GIS task where georeferenced image is later digitalized to generate vector data.

Geo-referencing and map projection (Coordinate system)

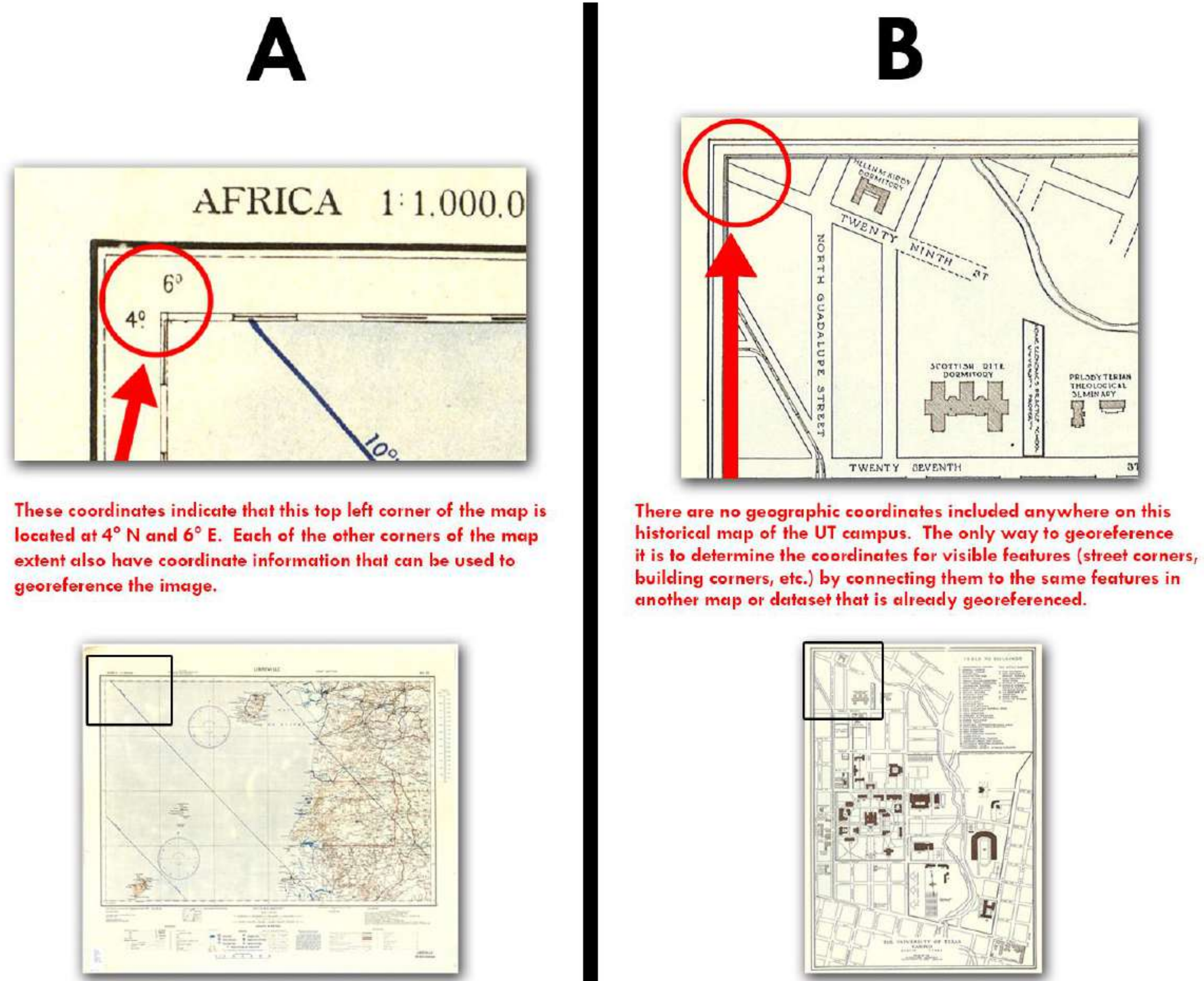
Geo-referencing:

Fig: old scanned map of Butwal being georeferenced with real world.

Source: http://mapadda.com/

Geo-referencing and map projection (Coordinate system)

Geo-referencing:



Source: https://libapps.s3.amazonaws.com/accounts/180230/images/georeferencing_method_comparison.jpg

Coordinate

- Coordinates define the location in two or three-dimensional space.
- They are used to define the shape and location of each spatial object or phenomenon.

Coordinate system

- Coordinate system can be thought as a system used to identify locations on a graph or grid.
- There are various coordinate systems available to represent the location of any point. However, all of them fall into two broad categories

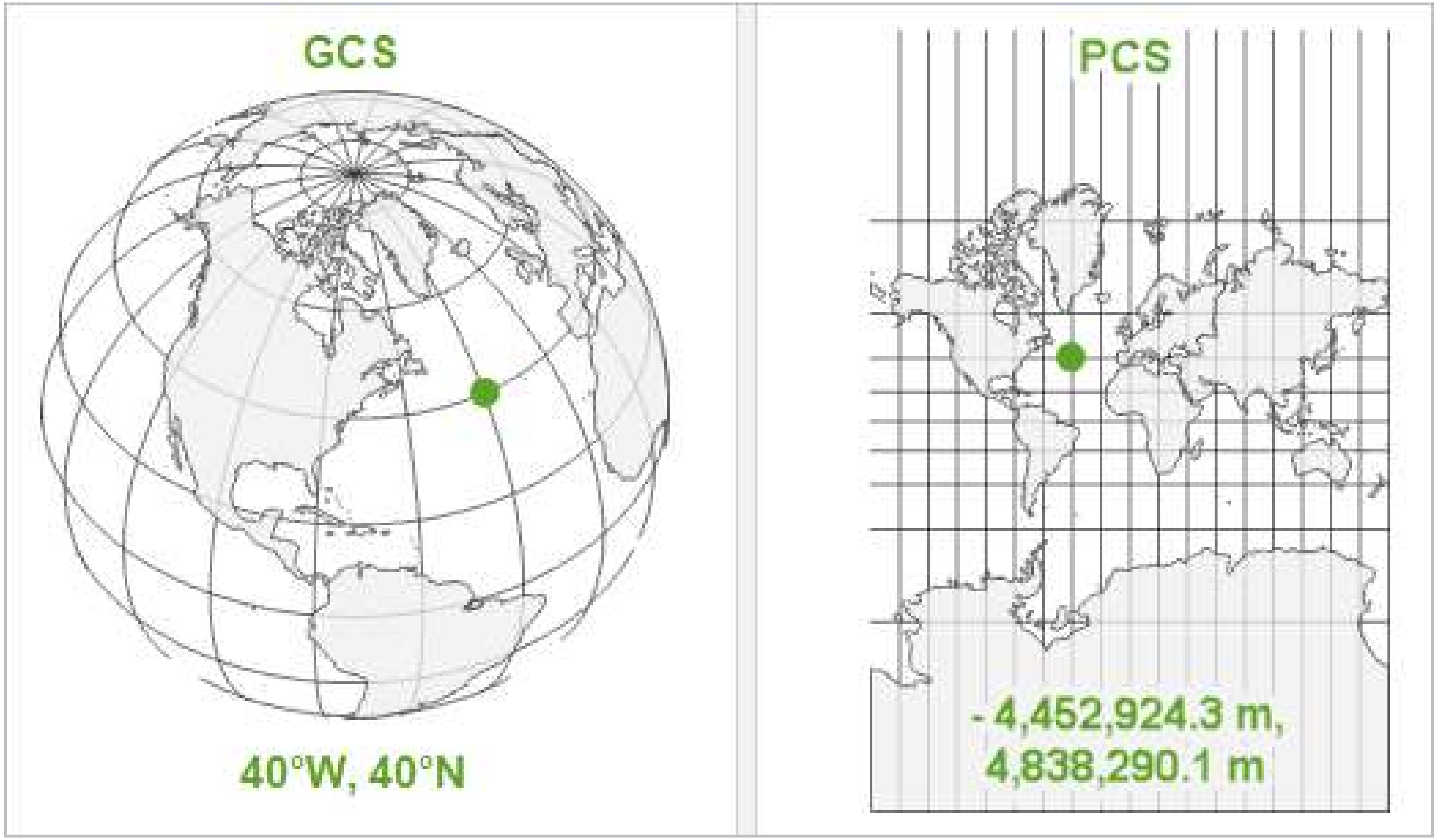
i. Curvilinear

- Curvilinear system uses angular measurements from the origin to describe one's position.
- For example, geographic coordinate system (GCS) that uses latitude/longitude measurements.

ii. Rectangular

- Rectangular coordinate system uses distance measurements from the origin.
- For example, cartesian coordinate system (CCS) that uses distance measurements.

Geographic Coordinate System (GCS) vs Projected Coordinate System



Source: https://www.google.com/search?q=coordinate+system+in+gis+file+type+pdf&hl=en&sxsrf=APwXEdcv6oBC-8ZOEJTReL5cFwyShYWYSg:1683621003702&source=lnms&tbm=isch&sa=X&ved=2ahUKEwietvL-6Of-AhVgXGwGHZr0AGIQ_AUoAXoECAEQAw&biw=1536&bih=722&dpr=1.25#imgrc=2JusKR0uzRmlhM

Geographic Coordinate System (GCS)

- The geographic coordinate system is the reference system for locating spatial features on the Earth's surface.
- The geographic coordinate system is defined by longitude and latitude.
- Latitude and longitude can be expressed in degrees-minutes-seconds (DMS) or in decimal degrees (DD).
- Depending upon the datum there are various geographic coordinate system used globally. Some of them are: Everest 1830, World Geodetic System (WGS) 84 etc.

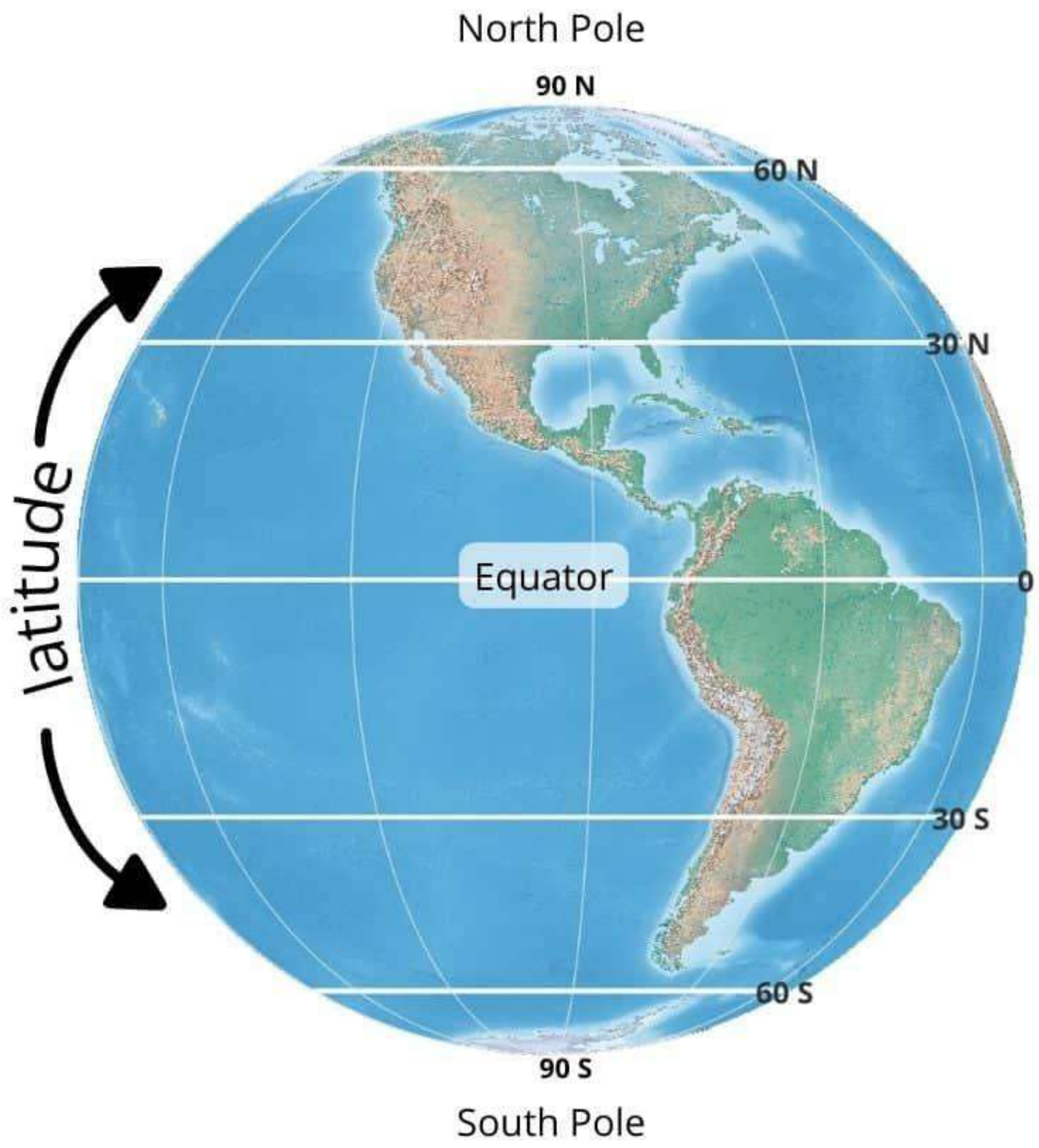
Latitude (λ)

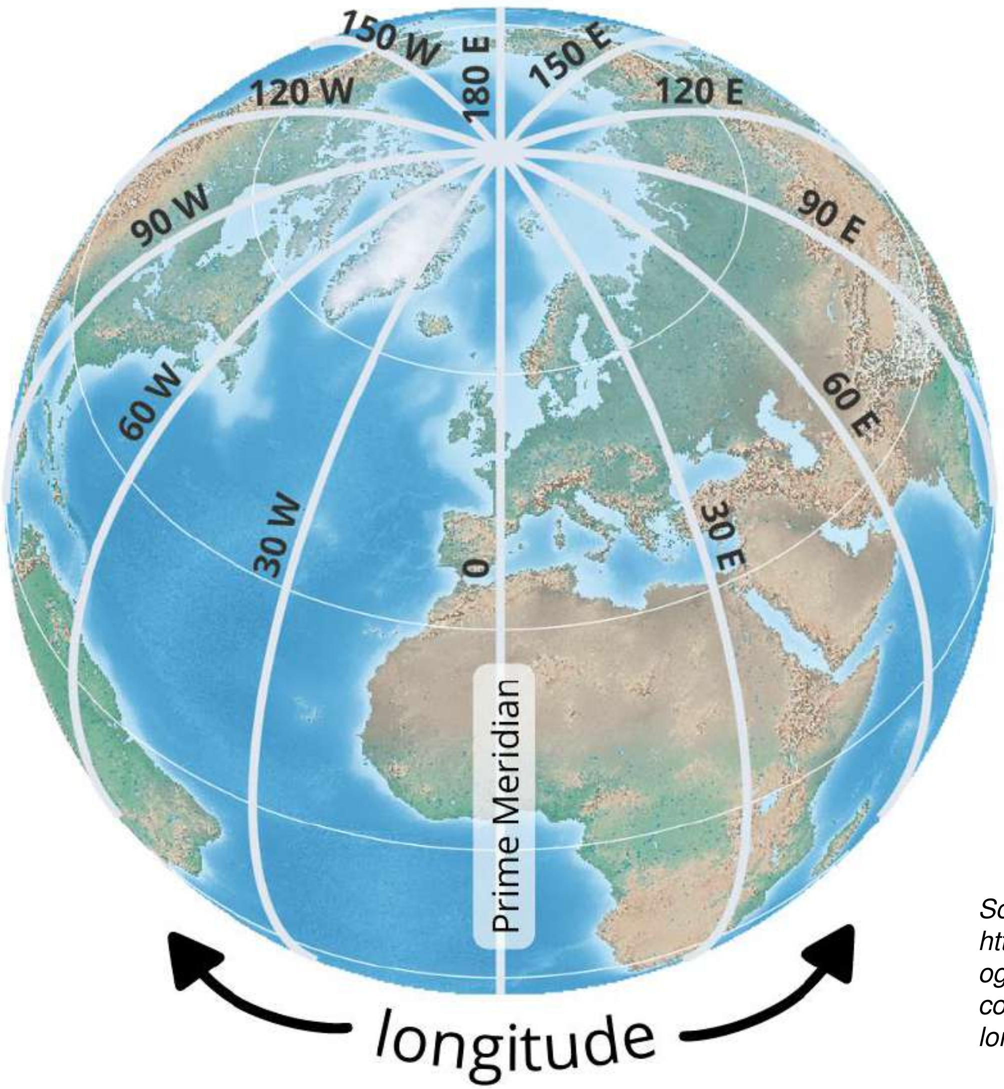
- Lines of latitude run across a map east-west, the latitude indicates the north-south position of a point on earth.
- Lines of latitude start at 0 degrees at the equator and end at 90 degrees at the North and South Poles (for a total to 180 degrees of latitude).
- Lines of latitude are also known as parallels.

Longitude (φ)

- Longitude are lines that run north-south and indicate the east-west position of a point on earth.
- Lines of longitude are called meridians.

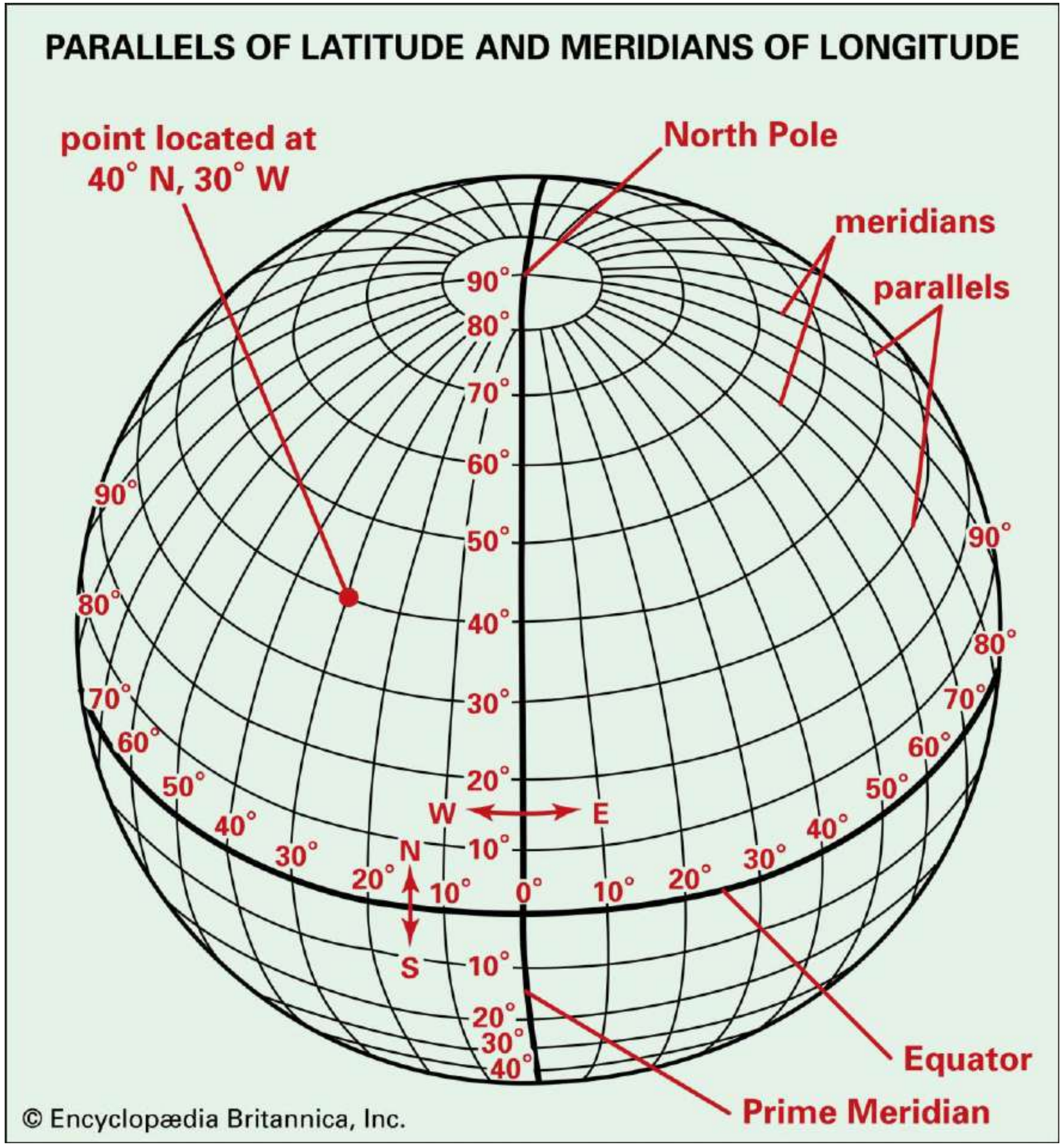
Latitude and longitude





Source: https://www.geographyrealm.com/latitude-longitude/

Latitude and longitude



Source: https://www.britannica.com/science/latitude

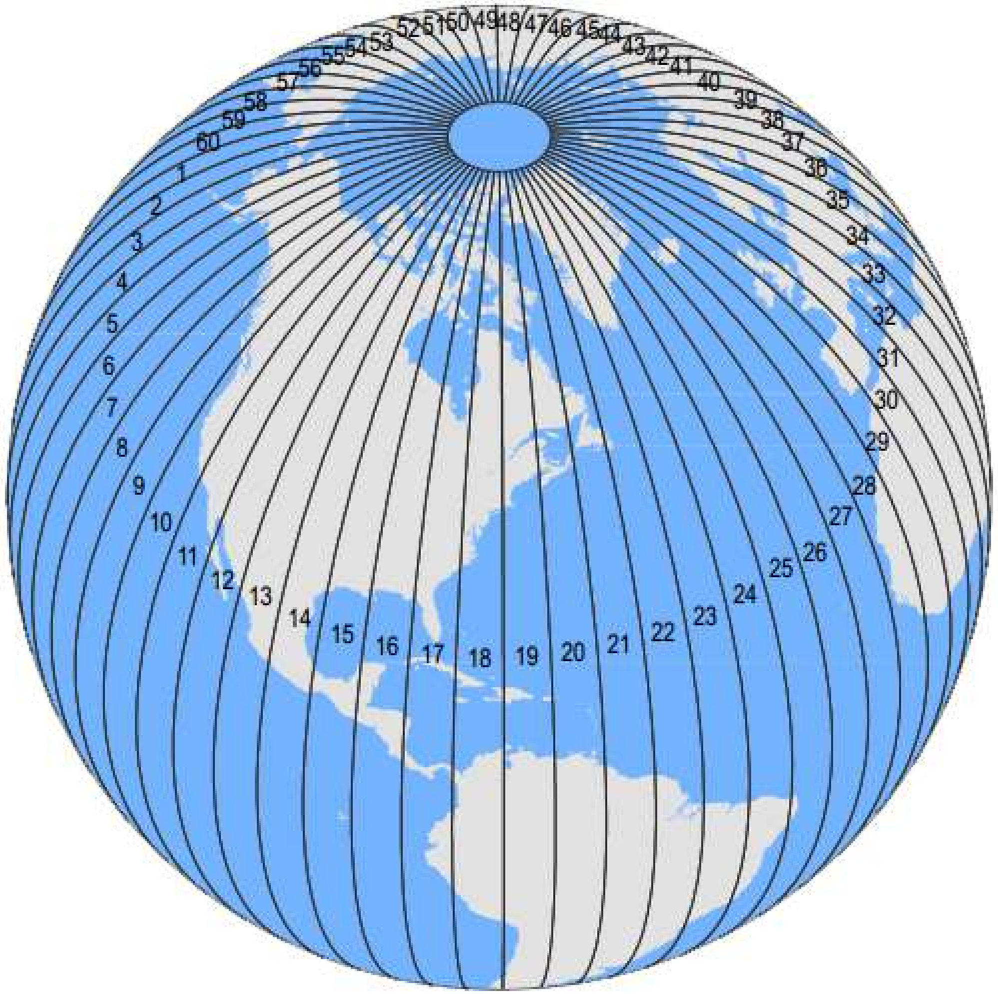
Projection (Map Projection)

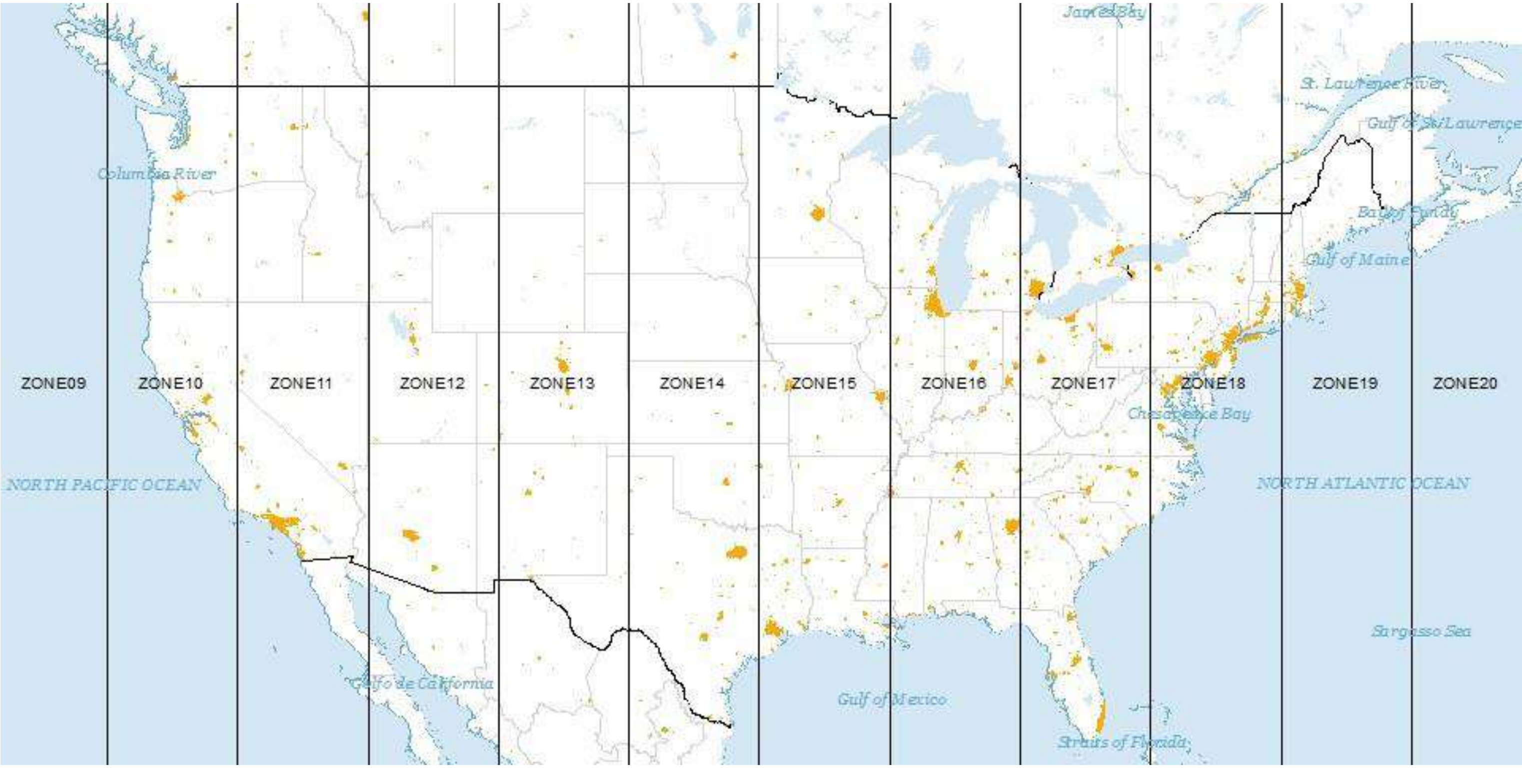
- It refer to the methods and procedures that are used to transform the spherical three-dimensional earth into two-dimensional planar surface.
- It is the mathematical formula that is used to translate location on the curved surface of the Earth with the geodetic coordinates (φ, λ) to planar map coordinates (x, y).
- The transformation from the surface of an ellipsoid to a flat surface always involves distortion, and no map projection is perfect.
- Every map projection preserves certain spatial properties while sacrificing other properties.
- Some of the common projection are: Mercator Projection, Projection, Lambert Conformal Projection etc.

Projected coordinate system

- A Projected Coordinate System is a 2D reference system that uses Cartesian coordinates (X: Easting,Y: Northing) to represent the location of a point on a two-dimensional surface.
- It consists of lines that intersect at right angles, forming a grid.
- To transform from geographic to projected coordinate system a projection is required.
- The advantage of using projected coordinate system is that the lengths, angles, and areas are constant across the two dimensions which is not true when working in a geographic coordinate system.
- Some of the common projected coordinate system are: WGS 84/44N, WGS 84/45N etc.

Projected coordinate system: WGS 84/44N



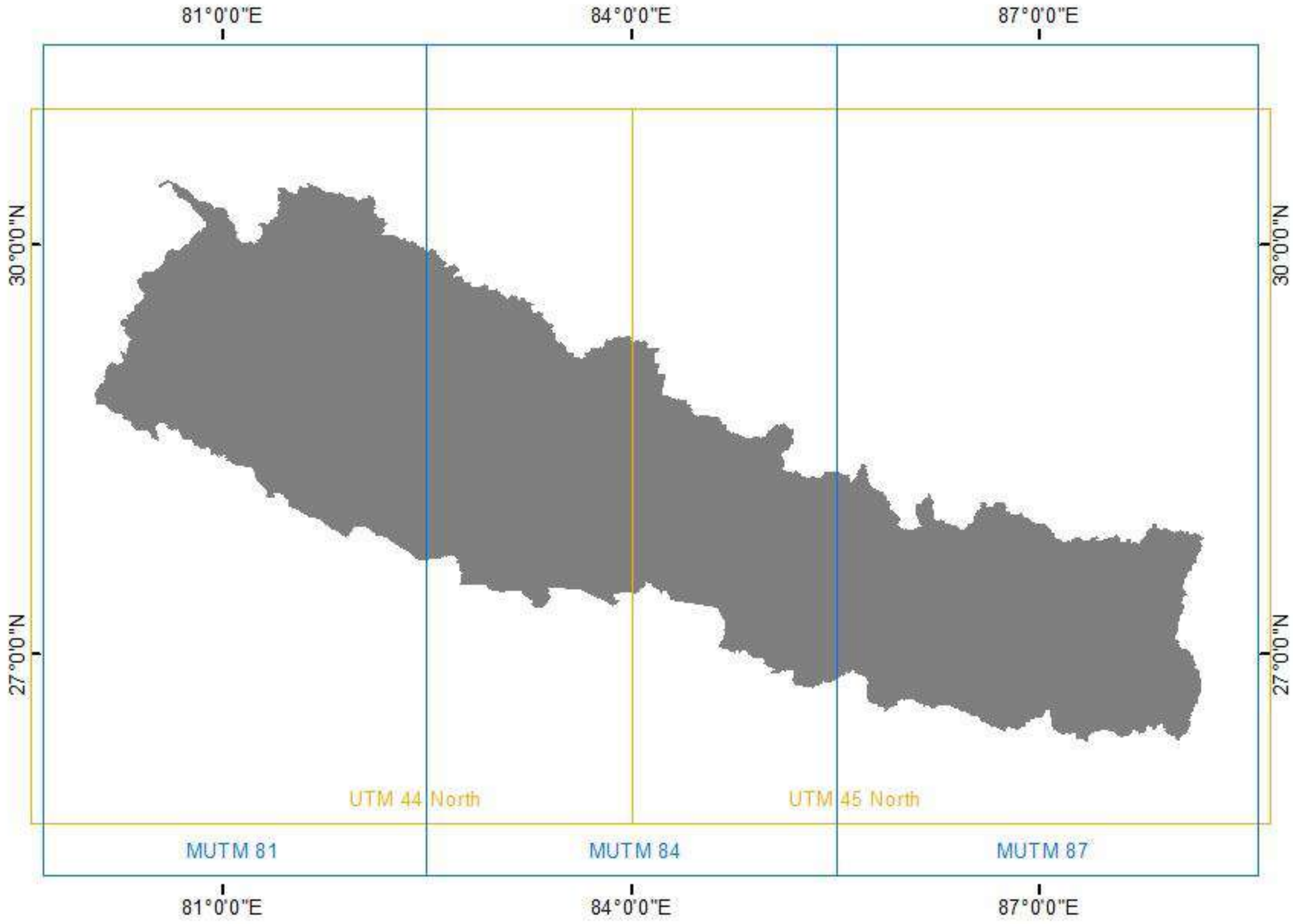


Source: https://gisgeography.com/utm-universal-transverse-mercator-projection/

Projected system used in Nepal

- In Nepal the projection system used is universal transverse Mercator (UTM) which may be based on either WGS84 or Everest 1830.
- Considering the WGS84 Nepal lies in two zones of UTM projection system which are 44 north and 45 north zones.
- When UTM is based on Everest 1830 projection system used is Modified Transverse Mercator (MUTM). It consists of three zones MUTM 81, MUTM 84 and MUTM 87 which covers west, central and eastern Nepal respectively.

Projected system used in Nepal



Source: https://spaceappnet.wordpress.com/2020/06/24/coordinate-systems-used-in-nepal/#:~:text=In%20Projected%20Coordinate%20System%20(PCS,north%20and%2045%20north%20zones.

Digitization

- Digitizing is the process of converting geographic features on a paper map into digital format.
- The raster image used must be georeferenced before digitization.
- It includes tracing points, polylines or polygons features from maps or images.
- It is a method to capture vector data.
- It is the most common and labor intensive method to create a spatial database.

There are mainly three methods of digitization. They are:

i. Manual
ii. Interactive Tracing (Semi-automatic)
iii. Automatic (Raster to Vector Conversion)

Digitization

OHT 18

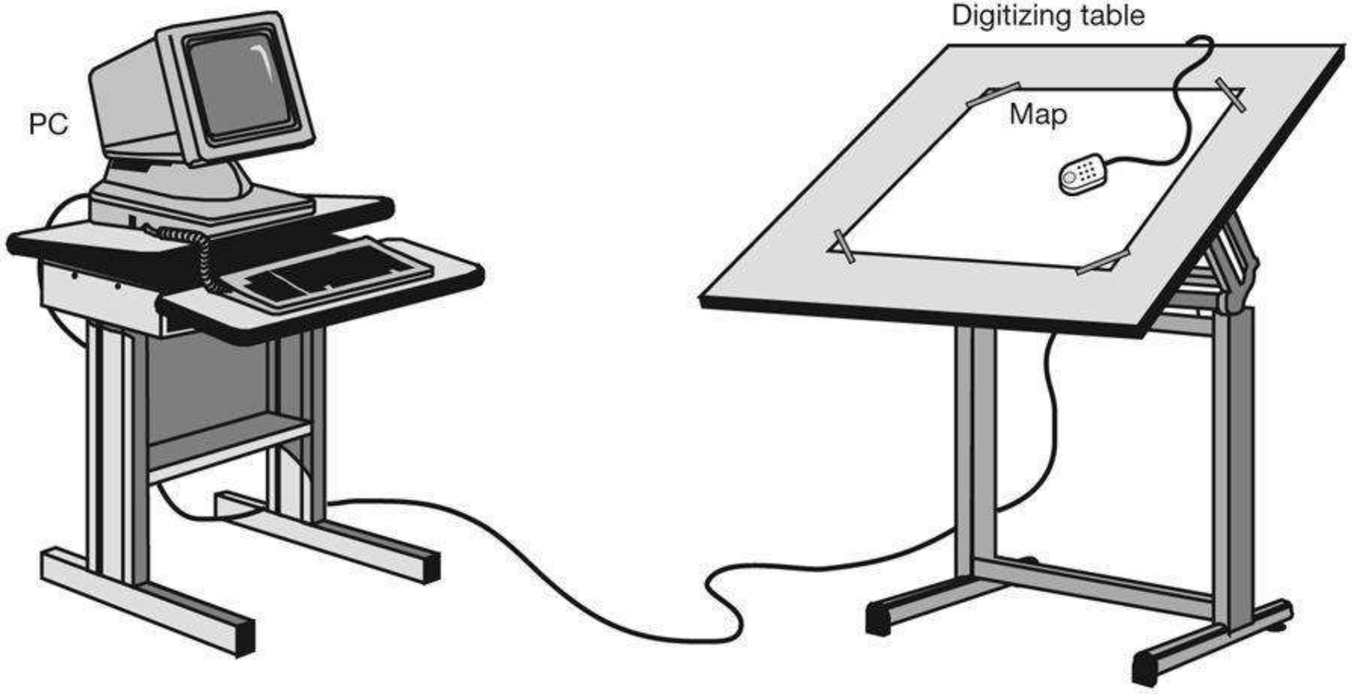


Figure 5.2 Digitizing table and PC workstation

Heywood *et al*., An Introduction to Geographical Information Systems, 2/e

Fig: showing manual digitization

Digitization

1. Manual digitization

- It involves digitization with the help of digitizing tablet (an electronic device consisting of a table upon which the map or drawing is placed).
- With this method, the operator manually traces all the lines from his hardcopy map using a pointer device (digitizer puck) and create an identical digital map on his computer.

2. Heads-Up Digitizing/Semi-automatic digitization/On-screen digitization

- It is the simplest way to create vectors from raster layers manually from a computer screen using a mouse or digitizing cursor.
- Main difference between the manual digitization and semi-automated digitization is that instead of the digitization table raster maps are displayed on the computer screen.
- In this process first of all scanning of paper maps is done to form a digital raster data and the raster data is eventually used for creation of vector lines by the method of tracing similar to the manual digitization.
- Advantage of this process are zooming facility gives higher accuracy, image processing techniques can be applied in the image to enhance the image than that of manual technique.

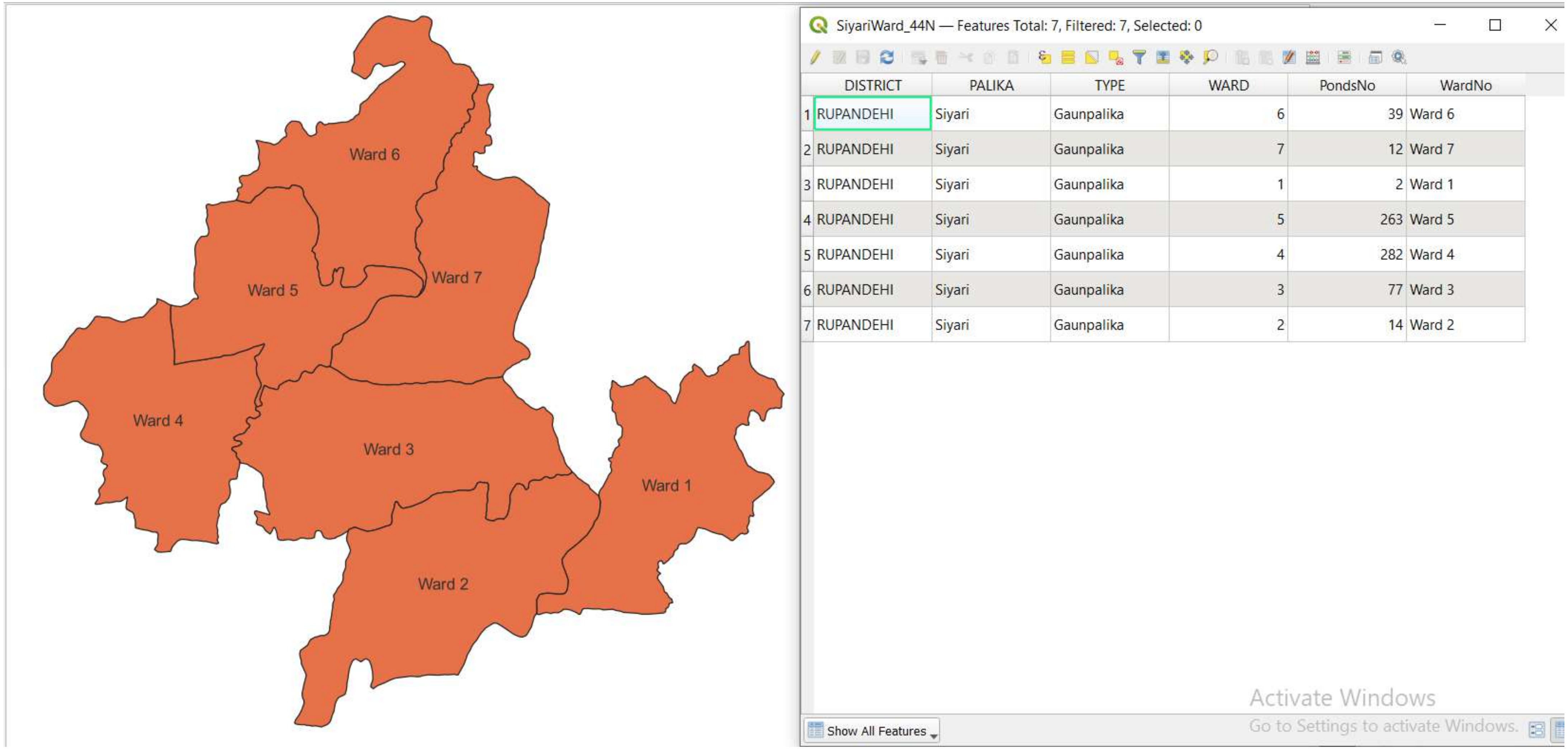
Digitization

3. Automatic digitization

- This method uses the computer directly to convert the digital raster image to vector format.
- Automatic digitizing or so called automated raster to vector conversion, traces lines automatically from the scanned raster image using image processing and pattern recognition techniques.

3.2 Data Entry and Data Preparation

Process of inserting attribute data in digitized data



	DISTRICT	PALIKA	TYPE	WARD	PondsNo	WardNo
1	RUPANDEHI	Siyari	Gaunpalika	6	39	Ward 6
2	RUPANDEHI	Siyari	Gaunpalika	7	12	Ward 7
3	RUPANDEHI	Siyari	Gaunpalika	1	2	Ward 1
4	RUPANDEHI	Siyari	Gaunpalika	5	263	Ward 5
5	RUPANDEHI	Siyari	Gaunpalika	4	282	Ward 4
6	RUPANDEHI	Siyari	Gaunpalika	3	77	Ward 3
7	RUPANDEHI	Siyari	Gaunpalika	2	14	Ward 2

Fig: Map showing polygon features with other attributes like ''DISTRICT', 'PALIKA', 'TYPE', 'PondNo' etc.'

Source: mapadda.com

Process of inserting attribute data in digitized data

- The process of inserting attribute data into digitized data involves associating specific attributes or characteristics with the corresponding digital data.
- Inserting attribute data in digitalized data may vary depending on the specific tools, software, or workflows employed in a digitization project. The general steps involved are as follows:

Process of inserting attribute data in digitized data

i. Import spatial data:

- The first step is to import the spatial data into your GIS software.
- This can include vector data such as points, lines, or polygons representing geographic features like roads, buildings, or land parcels..

ii. Prepare attribute data:

- It includes determining the specific attributes that need to be associated with the digitized data.
- Attribute data typically consists of non-spatial information, such as names, categories, numerical values, or descriptive text.

iii. Create attribute fields:

- Within your GIS software, create the necessary attribute fields that correspond to the types of data you want to store.
- For example, if you have a set of buildings and want to include attributes like building name, construction year, and number of floors, create separate fields for each attribute.

Process of inserting attribute data in digitized data

iv. Populate attribute values:

- Fill in the attribute values for each feature in the attribute table.
- You can either manually enter the data or use automated methods like field calculations or importing from external sources through join or relate.
- External sources required unique identifier or key field in spatial data.

v. Save and maintain data integrity:

- Save the updated data in a suitable format, such as a shapefile, geodatabase, or feature class.
- Ensure that you maintain data integrity by correctly managing the link between the spatial and attribute data.
- This can involve adding new features, deleting existing ones, or modifying attribute values. .

Create attribute data of digitized features

1. Identify the Features:

- Determine the features you want to digitize, such as points, lines, or polygons, based on your project requirements.

2. Create Attribute Fields:

- After identifying features we need to determine the necessary attribute fields for each feature. This may include names, descriptions, dates, measurements, or any other relevant information.
- This can be done through GIS software's attribute table.

3. Assign Data Types:

- Assign appropriate data types to the attribute fields. For example, text for names, numeric for measurements, date for dates, and so on.

Create attribute data of digitized features

4. Start Digitizing:

- Use the digitizing tools of your GIS software to create the features on the map.

5. Enter Attribute Data:

- Once you have digitized a feature, access its attribute table and enter the corresponding data for each field.
- This process should be repeated for all features you digitize.

6. Maintain Consistency:

- Ensure consistency in attribute data entry by following predefined standards or guidelines. This helps maintain data integrity and facilitates analysis.

7. Save and Backup:

- Regularly save your GIS project to preserve the attribute data and any changes you have made. Additionally, create backups to prevent data loss in case of system failures.

Topology

- Topology shows the spatial relationships between connecting or adjacent vector features (points, polylines and polygons) in a GIS.
- Topological or topology-based data are useful for detecting and correcting digitising errors.
- Topology is necessary for carrying out some types of spatial analysis, such as network analysis.

Topological errors:

- There are different types of topological errors and they can be grouped according to whether the vector feature types are polygons or polylines.
- Topological errors with polygon features can include unclosed polygons, gaps between polygon borders or overlapping polygon borders.
- A common topological error with polyline features is that they do not meet perfectly at a point (node). This type of error is called an undershoot if a gap exists between the lines, and an overshoot if a line ends beyond the line it should connect to.

Topological errors:

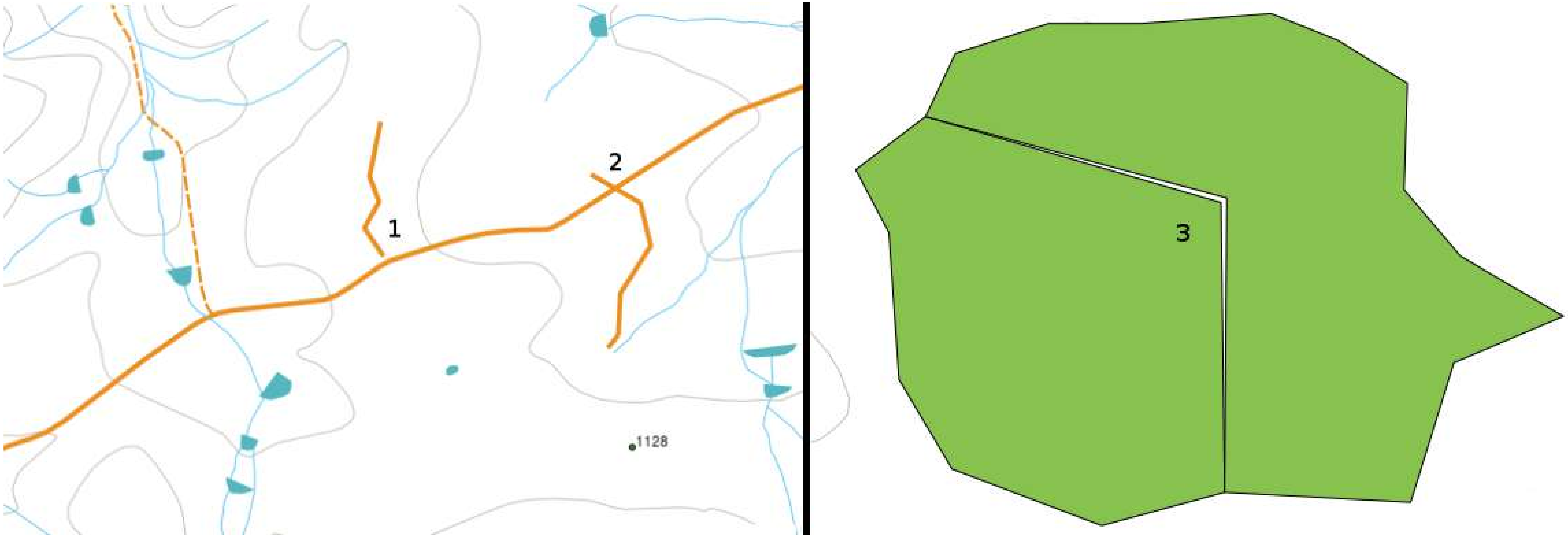


Fig: showing undershoot(1), overshoot(2) and two polygons whose vertices do not match on the border.

Source: https://docs.qgis.org/3.28/en/docs/gentle_gis_introduction/topology.html

Topological errors:

- Topological errors break the relationship between features.
- These errors need to be fixed in order to be able to vector analysis like network analysis, distance measurement and other analysis.

Checking and repairing Geometry of spatial data

- After digitization the topological error that may be present can be identified automatically, after which manual editing methods can be applied to correct the errors. Alternatively, some software may identify and automatically correct certain types of errors.

Some basic steps for checking and repairing geometry of spatial data are:

1. Identify the data with geometry issues:

- Start by identifying the spatial datasets that need to be checked for geometry problems.
- This could include shapefiles, feature classes in a geodatabase, or any other spatial data format.

2. Validate geometry:

- It includes checking the geometry of the features in the dataset using a spatial data validation tool or function.
- This process involves examining the integrity of points, lines, and polygons to identify errors such as self-intersections, gaps, overlaps, or invalid geometries.

Checking and repairing Geometry of spatial data

Some basic steps for checking and repairing geometry of spatial data are:

3. Repair invalid geometries:

- Once the geometry errors are identified, they need to be fixed. Depending on the GIS software you are using, there are different methods to repair invalid geometries. Some common approaches include:

a. Deleting or removing the problematic features:

- If the invalid features are not crucial or can't be easily repaired, they can be deleted from the dataset.

b. Correcting geometry manually:

- In some cases, you may need to manually edit the geometry to fix specific issues. This can involve adjusting vertices, merging or splitting features, or digitizing new features.

c. Automated geometry repair tools:

- Many GIS software provide automated tools to repair invalid geometries. These tools can often fix simple errors like gaps or overlaps automatically.

Checking and repairing Geometry of spatial data

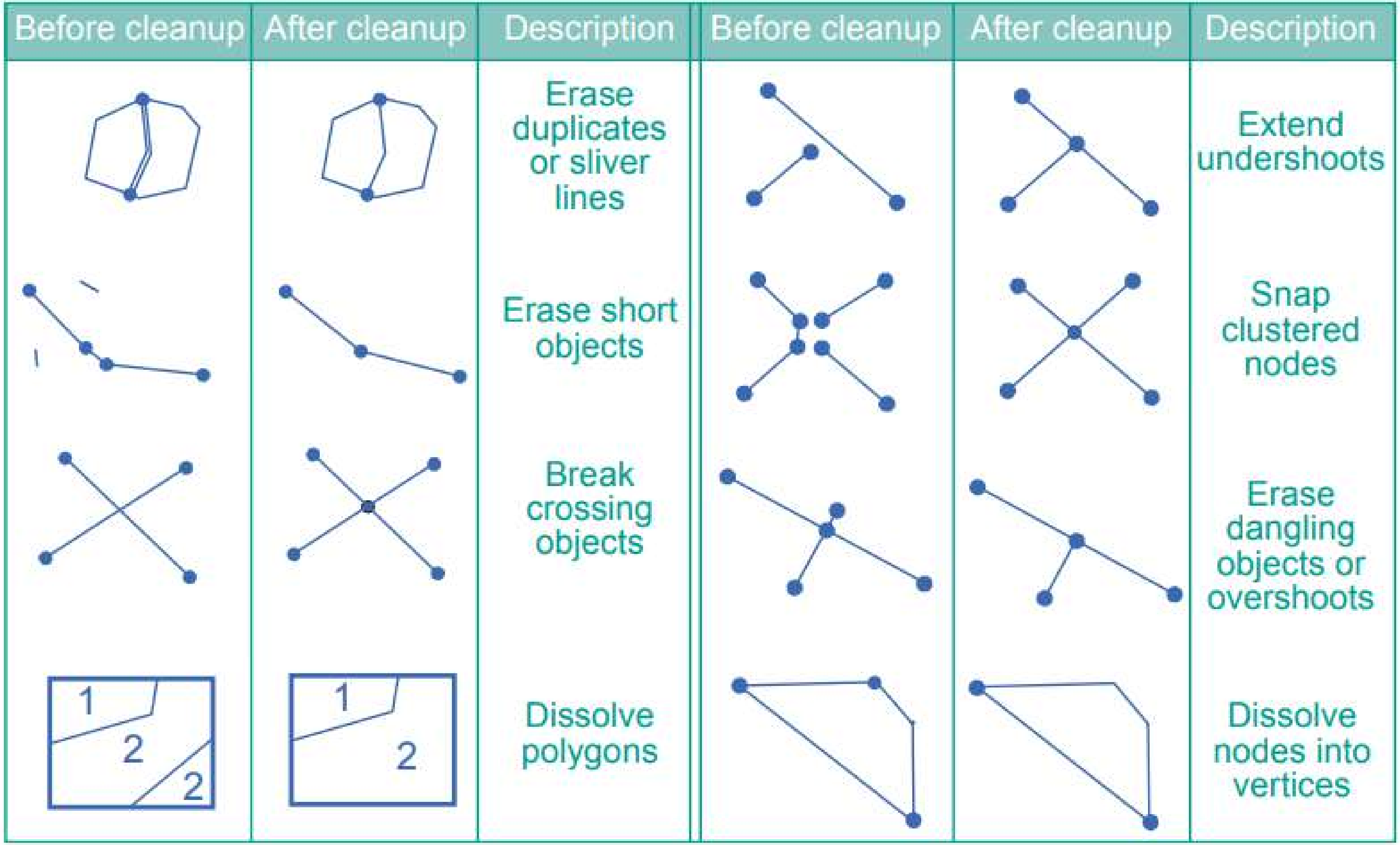
Some basic steps for checking and repairing geometry of spatial data are:

4. Revalidate and verify:

- After the geometry repairs, it is crucial to validate the data again to ensure that all errors have been resolved.
- This step helps confirm that the repaired dataset now complies with the desired geometry rules and standards.

3.2 Data Entry and Data Preparation

Checking and repairing Geometry of spatial data



Source: https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesgis.pdf

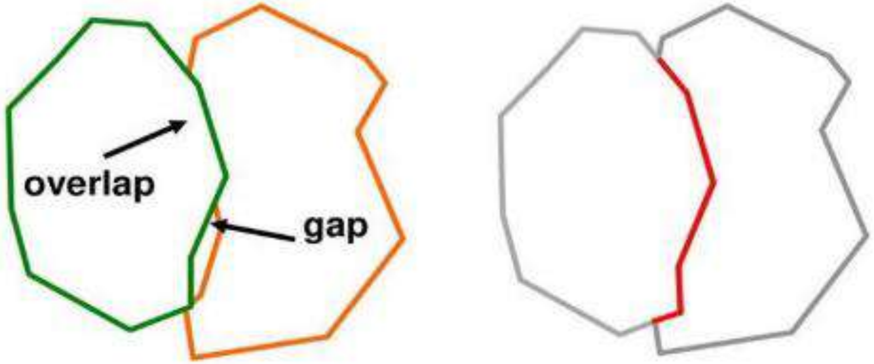
Fig: showing features being geometrically corrected after repairing them.

Data Topology GIS Operations and Map Composition and topological rules

Topological rules

Some of the common topological rules are:

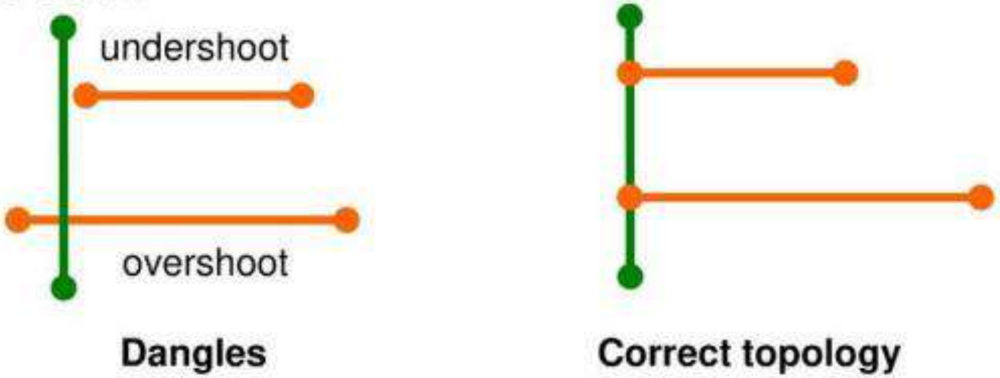
(1) Adjacent polygons must share a coincident boundary that is exactly the same for both. There are no gaps or overlaps between adjacent polygons.



(2) Lines should always end on other lines. Failure of two lines to meet is called a dangle.

– Undershoots occur when one line is not quite long enough to meet the other.

– Overshoots occur when one line crosses too far over the other.

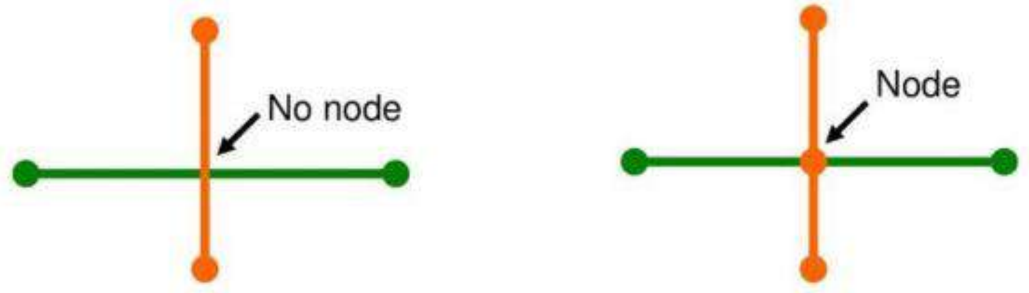


Data Topology GIS Operations and Map Composition and topological rules

Topological rules

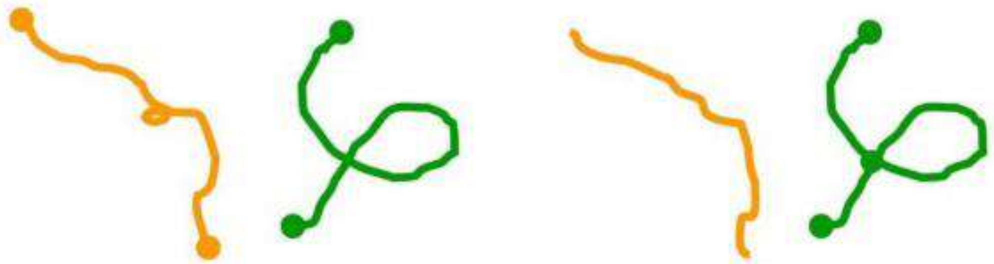
Some of the common topological rules are:

(3) Lines that intersect should always each have a node (endpoint) at the intersection. Lines crossing without nodes are termed improper intersections.



Improper intersection (2 lines)   **Proper intersection (4 lines)**

(5) Lines or polygon boundaries should not cross over themselves and form loops.



Loops   Correct topology

Data Topology GIS Operations and Map Composition and topological rules

Topological rules

Some of the common topological rules are:

6. Polygons Must Be Closed: Ensures that polygon features have closed boundaries, with no gaps or dangling segments.

Assignment

1. What do you mean by base map and thematic map ?.

References

- GPS/DGPS

https://volaya.github.io/gis-book/en/Data_sources.html

https://spaceplace.nasa.gov/gps/en/

https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/284522e.pdf

https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch33.pdf

https://sjce.ac.in/wp-content/uploads/2018/01/01_Total-Station_GPS-1.pdf

https://id.land/blog/gis-data-collection-where-it-comes-from-how-it-gets-there

GIS data capture: https://www.tamiu.edu/cees/courses/fall2018/geol4460_labs/lecture7.pdf

https://ekappiahadjei.files.wordpress.com/2018/04/unit-3-data-entry.pdf

https://gis-university.com/gis-data-input-techniques/

http://www.gitta.info/DerivedSourc/en/multimedia/scanning.pdf

Georeferencing: https://www.qgistutorials.com/en/docs/georeferencing_basics.html

Map projection along with true/false easting concept:
http://www.geo.hunter.cuny.edu/~jochen/gtech201/lectures/lec6concepts/06%20-%20Projected%20coordinate%20systems.html

https://docs.qgis.org/3.4/en/docs/gentle_gis_introduction/coordinate_reference_systems.html

Digitization:
http://ccnet.vidyasagar.ac.in:8450/pluginfile.php/1948/mod_resource/content/2/2ND%20SEM%20CBCS%20RSG%20204.2%20DIGITIZATION%20%281%29.pdf

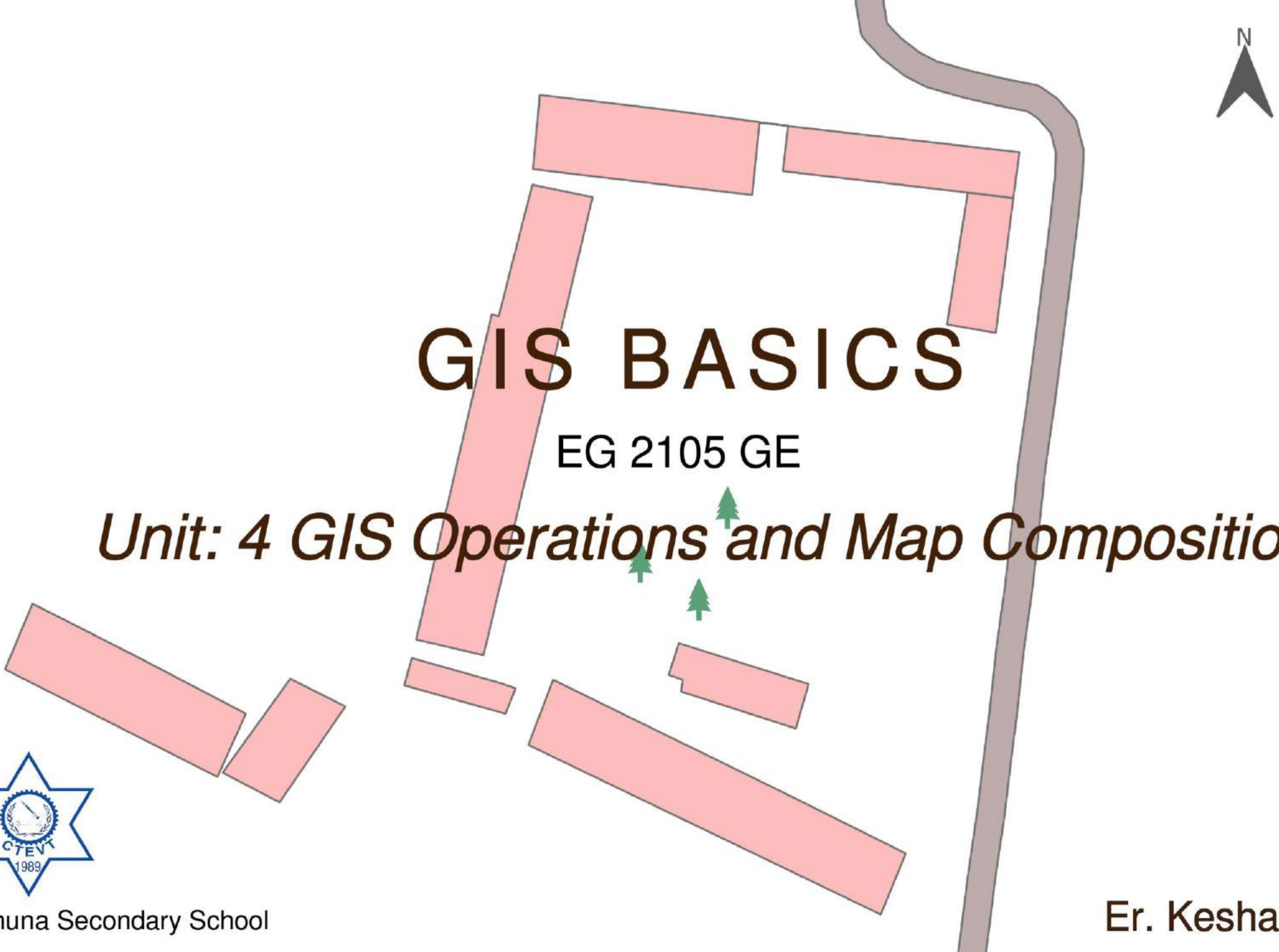
References

Topology

- *https://docs.qgis.org/3.28/en/docs/gentle_gis_introduction/topology.html*
- *https://gisgeography.com/topology-rules-arcgis/*

Data preparation:

https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesgis.pdf pg. 304



GIS BASICS

EG 2105 GE

Unit: 4 GIS Operations and Map Composition

Public Namuna Secondary School

Lecture by

Er. Keshav Raj Bhusal

Table of contents

4.1 Querying Databases

• Define querying database

• Understand structure of query language (SQL)

• Define and explain The terms: attribute query, Spatial query (location based query)

• Differentiate Spatial and database query

4.2 Overlay Operation and Geo-processing

• Define and explain following overlay operation with examples

- Clipping
- Intersection
- Union
- Merge
- Dissolve

4.3 Data Visualization

• Output map preparation

• Map symbolization

• Map design and map elements

4.1 Quering databases

Query

- A query is simply a request for information from the database.
- In GIS, it is the retrieval and display of data from spatial database about one or more features, their attributes and the relationship between them.
- Queries may be simple, only searching for one feature, or more complex, selecting a group of features.
- Queries in GIS are like ‘List the cities that can use water from a river which are 100km far from it’, ‘Find all districts that are neighbor of Rupandehi district ’ and so on.

- There are two basic methods for searching and querying in GIS:

1. Query by attribute
2. Query by geography/ Spatial Query/ Location based query

i. Selection by cursor

ii. Selection by graphical objects

iii. Selection by spatial relationship

4.1 Quering databases

Query

- There are two basic methods for searching and querying in GIS:

1. Query by attribute

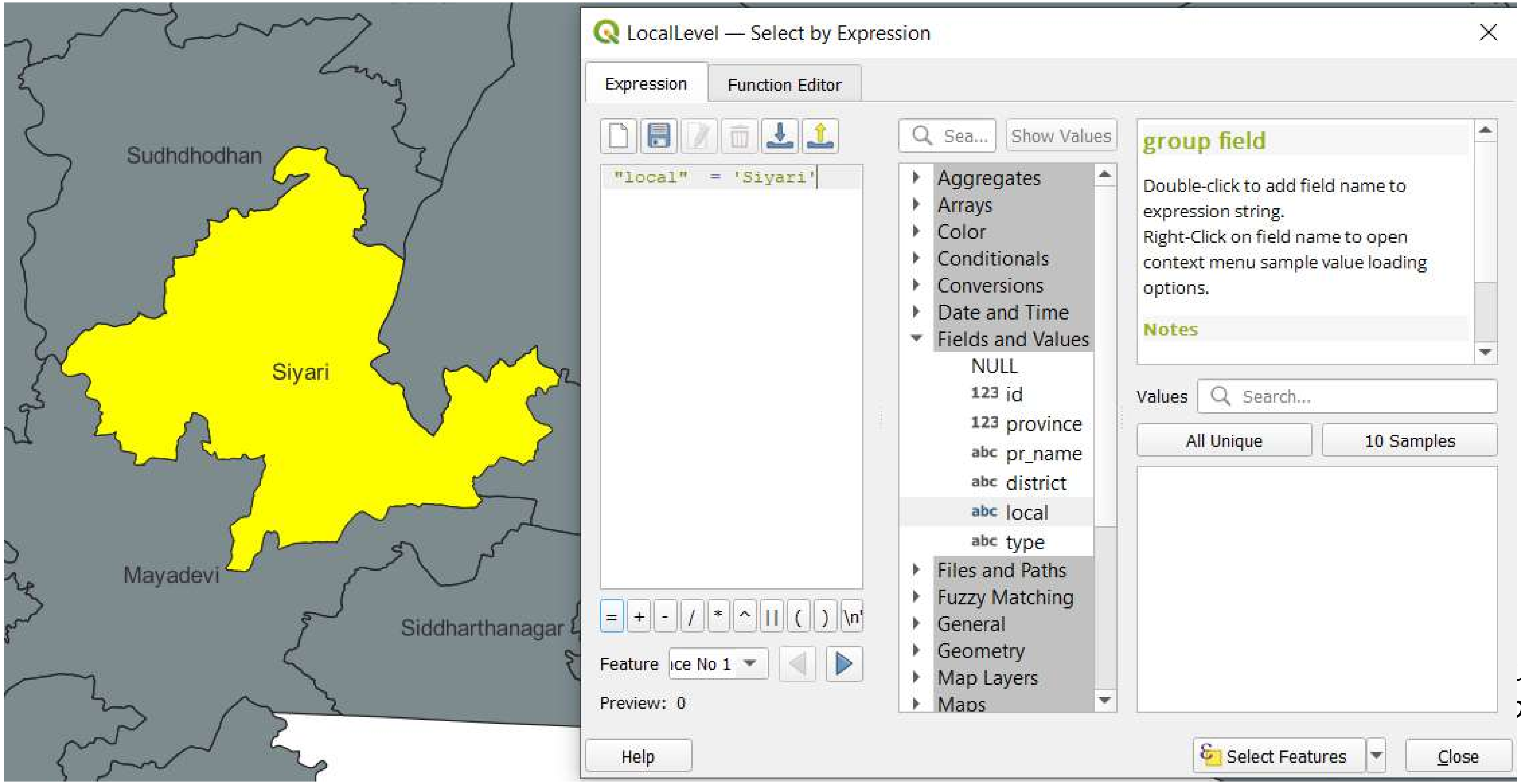
- An attribute query is the process of searching and retrieving records of features in a database based on attribute values.
- Such queries are a fundamental part of managing and analyzing GIS data. Typically, this is performed using a query language, most commonly SQL.
- It includes operators which can be relational (>, <, >=, <=), arithmetic (+, -, /), or Boolean (AND, OR) and will typically appear inside of conditional statements in the WHERE clause for performing operations.
- Many GIS softwares have user friendly interface for SQL formulation and with a few clicks the desired SQL commands are compiled. Behind the interface, the SQL syntax is used to query the database.

4.1 Quering databases

Query

- There are two basic methods for searching and querying in GIS:

1. Query by attribute



urce:
ps://mapadda.com/

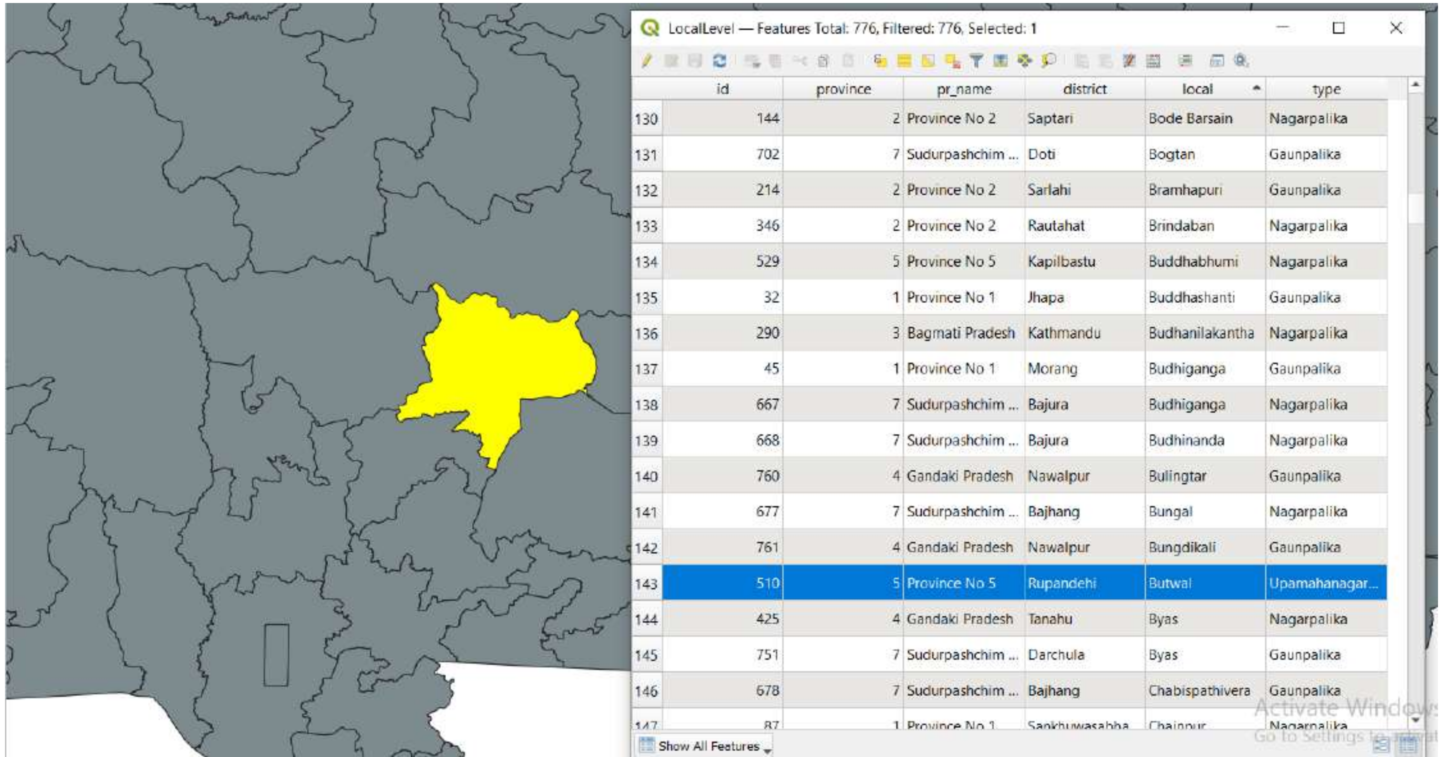
For example; below figure shows selection of 'Siyari' with query by attributes: 'local' = 'Siyari'

Query

- There are two basic methods for searching and querying in GIS:

2.1. Selection by cursor

- For this one selects to the feature of interest (points, lines, and polygons) by using the mouse cursor by clicking the feature.
- Selecting features highlight those features of interest, both on-screen and in the attribute table.
- It is the easiest way to search and query spatial data in a GIS.



Source: https://mapadda.com/

4.1 Quering databases

Query

- There are two basic methods for searching and querying in GIS:

2.2. Selection by graphical objects

- This kind of query uses graphical objects such as rectangle, circle, polygons to select the map features.
- For example, selecting all the districts that falls under the drawn rectangle.

Query

- There are two basic methods for searching and querying in GIS:

2.3. Selection by spatial relationships

- It allows one to select particular features by examining their position relative to other features.
- It includes “target layer” (feature dataset whose attributes are selected) and “source layer” (feature dataset on which the spatial query is applied).
- For example, finding the buildings, which are completely within the forest as shown in below figure.

Query

- There are two basic methods for searching and querying in GIS:

2.3. Selection by spatial relationships

Layer1

Layer 2

Input

Find the buildings, which are completely within the forest.

Query

TOPO

Shape	ID	AREA
Polygon	1	105
Polygon	2	50
Polygon	3	50
Polygon	4	110
Polygon	5	110
Polygon	6	115
Polygon	7	50
Polygon	8	50
Polygon	9	50
Polygon	10	50
Polygon	11	50
Polygon	12	50

Output

Source: http://www.gitta.info/

Query

- There are two basic methods for searching and querying in GIS:

2.3. Selection by spatial relationships

- Spatial relationship in query can be divided into following subtype:

1. Containment:

- Selects the features that are contained within the other specified feature. For example selecting all the Palikas that fall under the given district.

2. Adjacency:

- Selects the adjacent features used for the selection.For eg. Select all the district that are adjacent to the selected district.

3. Intersect:

- Selects the features that intersect to the given feature. For example select all the districts that the highway passes.

4. Proximity:

- Selects the features which is at the specified distance from the given object. For example select all the blood banks at a distance of 200 meters from the hospital.

Structured Query Language (SQL)

- SQL is a computer language for storing, manipulating and retrieving data stored in relational database.
- It provides a set of commands or statements that allow users to interact with a database, perform various operations, and retrieve or modify data stored within the database.
- SQL is the standard language for Relational Database Management System (RDBMS). All relational database management systems like MySQL, MS Access, Oracle, Postgres and SQL Server use SQL as standard database language.
- SQL commands includes:

i. Data Definition language

Command	Description
CREATE	Creates a new table, a view of a table, or other object in database
ALTER	Modifies an existing database object, such as a table.
DROP	Deletes an entire table, a view of a table or other object in the database.

4.1 Quering databases

Structured Query Language (SQL)

- The basic SQL syntax is

SELECT <attributes/ fieldname>

FROM < table>

Where < Conditional statements>

Structured Query Language (SQL)

- SQL commands includes:

ii. Data Manipulation language

Command	Description
INSERT	Creates a record
UPDATE	Modifies records
DELETE	Deletes records

iii. Data Control Language

Command	Description
GRANT	Gives a privilege to user
REVOKE	Takes back privileges granted from user

Structured Query Language (SQL)

- SQL commands includes:

iv. Data Query Language

Command	Description
SELECT	Retrieves certain records from one or more tables

Structured Query Language (SQL)

Data Output | Explain | Messages | Notifications

Before Query

	id integer	dname character varying (20)	iname character varying (20)	salary integer
1	1	ComputerScience	ram	5000
2	2	Cartography	shyam	5200
3	3	English	lal	5100
4	4	GIS	rancho	2300
5	5	English	dal	5300
6	6	ProjectEngineering	chaula	5500
7	7	Community	dhami	5600
8	8	ComputerScience	kandel	6000
9	9	ComputerScience	hari	5000
10	10	Surveying	kandel	5100

Query Editor | Query History

```sql
select dname
from department
where salary> 5000;
```

SQL code

Data Output | Explain | Messages | Notifications

	dname character varying (20)
1	Cartography
2	English
3	English
4	ProjectEngineering
5	Community
6	ComputerScience
7	Surveying

Output

After Query

Source: https://mapadda.com/

4.1 Quering databases

Differentiate Spatial and database query

- A database query and a spatial query are both used to retrieve information from a database, but they have different purposes and functionality.

Database query

- It is a general term that encompasses various types of queries performed on structured databases.
- Database queries are used to work with general data stored in a database, which can include text, numbers, dates, and other types of information by operating on attributes and relationships within the database tables.
- It uses set algebra (less than, greater than, equal to and not equal to) and Boolean algebra (AND, OR).
- Database queries enables operations like retrieving specific columns or rows, filtering data based on conditions, joining tables, aggregating data, and performing calculations.
- A basic SQL syntax for database query is

```
SELECT tree
FROM type
WHERE type = 'Peepal';
```

Here tree is table name, type is column name; this query selects only tree whose name is Peepal

4.1 Quering databases

Differentiate Spatial and database query

Spatial query

- A spatial query is a specialized type of query used in spatial databases or GIS to retrieve information about spatial data that has location or geometry attributes.
- Spatial queries enable the retrieval, analysis and display of spatial data about one or more features, their attributes and the relationship between them.
- It uses spatial logic and focus more on the physical relationships between datasets. For eg. finding all the rivers that are within a given state or all the states that are adjacent to a river.
- The answers to spatial queries are derived directly from the location of features on a map.
- This query enables operations like finding objects within a specified area, determining distances between objects, calculating intersections or overlaps, and conducting spatial analysis.

4.1 Quering databases

Spatial query

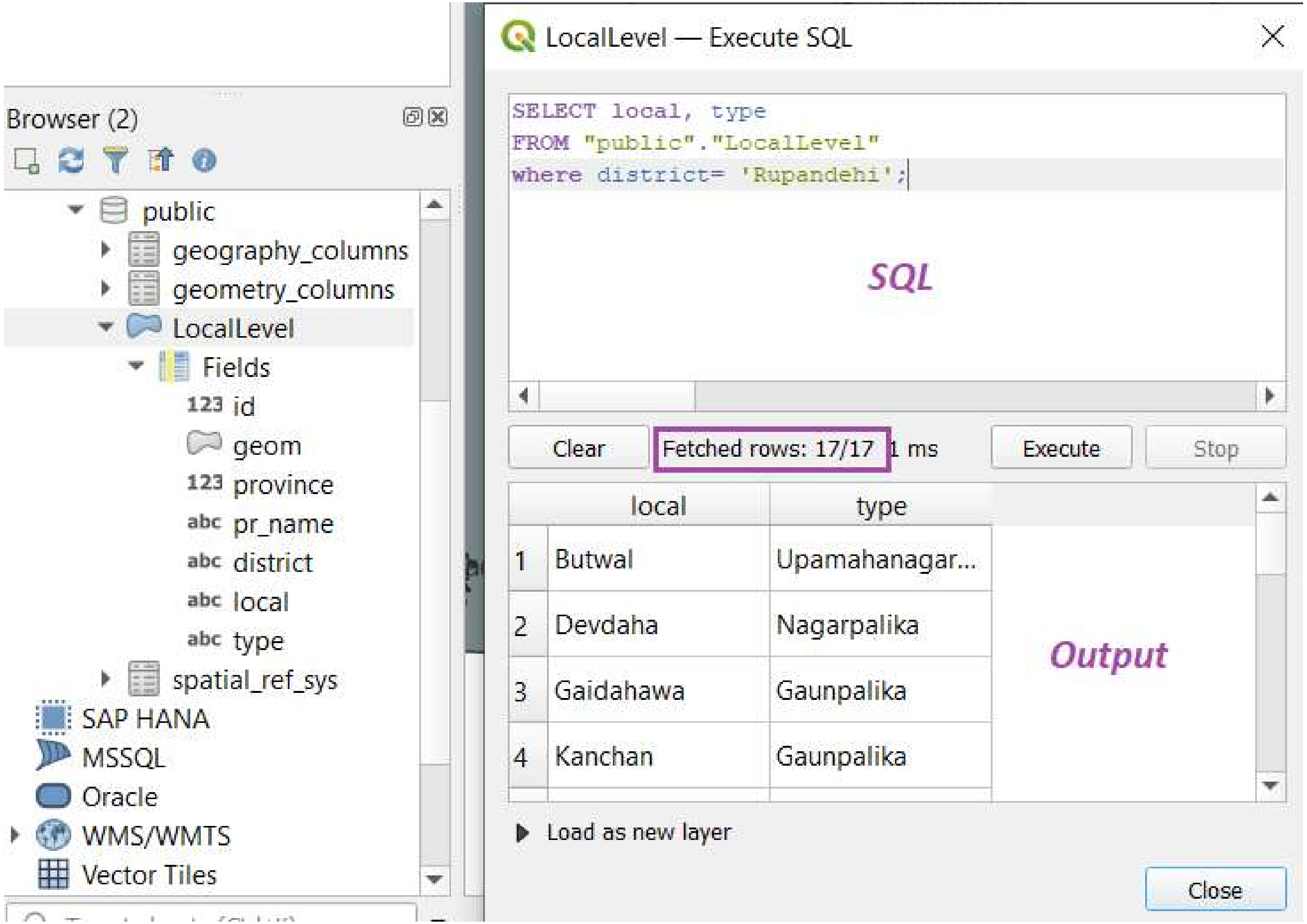


Fig: showing simple SQL query to select only local levels that lies in Rupandehi district.

Here public is schema while LocalLevel is table name.

Source: https://mapadda.com/

Overlay

- Overlay is the spatial operation in which two or more map layers are registered to a common coordinate system and are superimposed for the purpose of showing relationship between the map layers.
- It is a technique of combining two spatial data layers and producing a third from them using spatial overlay operators.
- The principle of spatial overlay is to compare the characteristics of the same location in both data layers, and to produce a new characteristic for each location in the output data layer.
- Both the vector and raster data can be overlayed.
- An overlay operation combines the geometries and attributes of two feature layers to create the output.
- Overlay methods are based on the Boolean connectors AND, OR, and XOR.
- Some of the overlay operation used in GIS are:

i. Clipping
ii. Intersection
iii. Union
iv. Merge
v. Dissolve

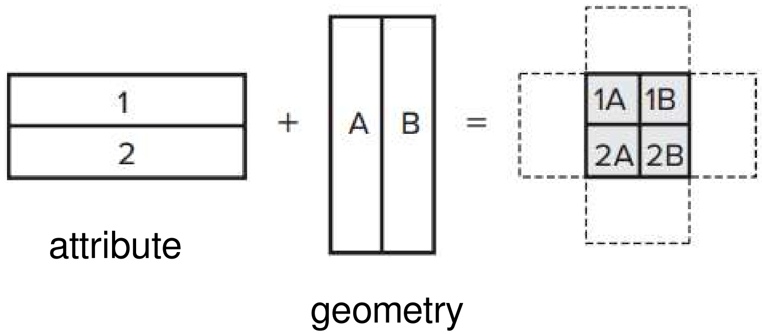


Fig: illustrates an overlay operation with two simple polygon layers. Each feature on the output contains a combination of attributes from the input layers.

4.2 Overlay Operation and Geo-processing

Geo-processing

- Geoprocessing in GIS (Geographic Information Systems) refers to the set of operations and techniques used to manipulate, analyze, and transform geographic data.
- It involves performing various spatial operations on geographic data to derive meaningful information, make informed decisions, and create maps or other spatial products.
- Some of the geoprocessing tools are: Clip, Merge, Union, Erase, Intersection etc.

Overlay

Clipping

- It is a kind of overlay operation where a input data layer spatial extent is restricted to the generalized outer boundary of a second input layer (mainly polygon).
- This is advantageous when an analyst only needs to work with a certain focus area; he/she can discard the unnecessary spatial information with no loss to his/her core data.
- It is commonly used GIS operation can be applied for both vector and raster data.
- For example, clipping a satellite image with a study area vector data.

Overlay

Clipping

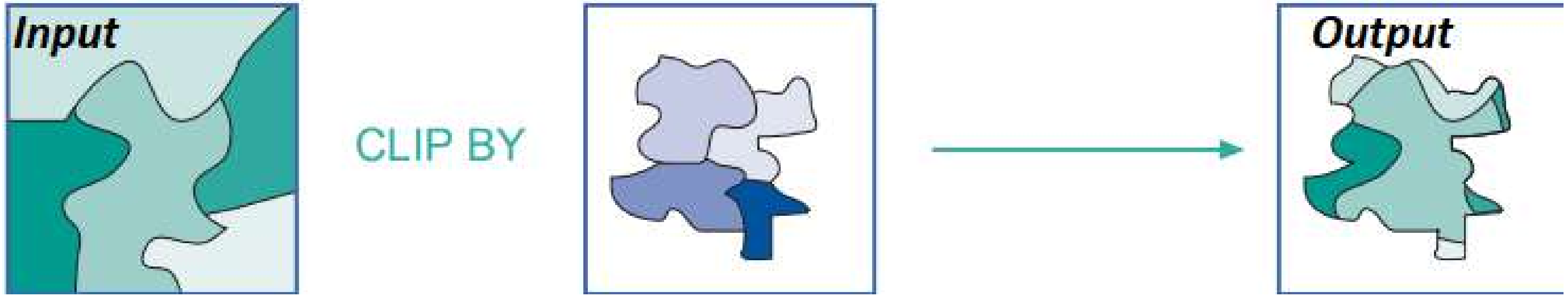


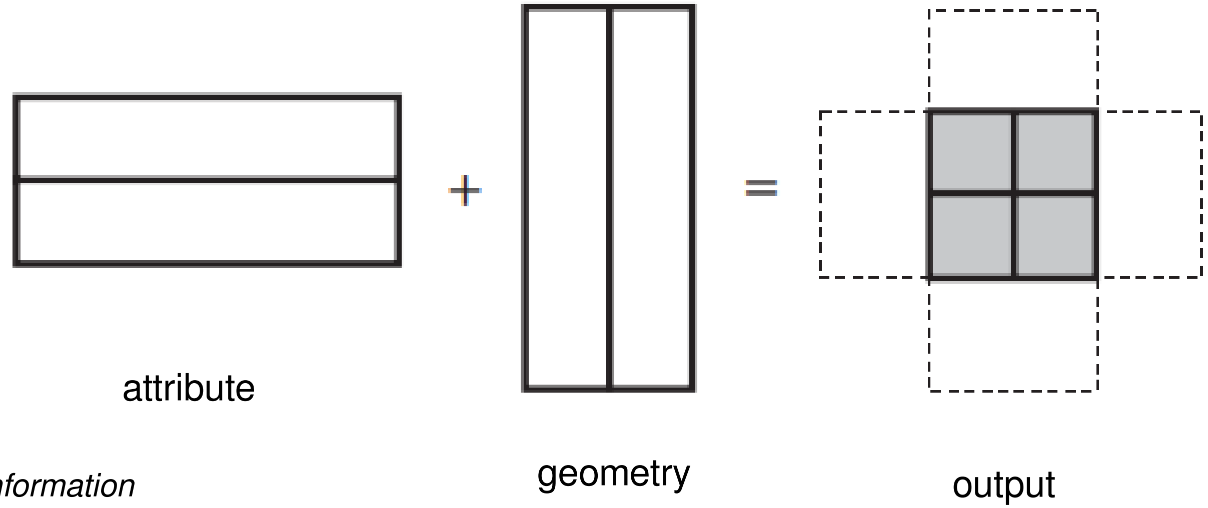
Fig: showing clipping operation

Source: Principles of Geographic Information System - Rolf A. de By (ed.) (ITC Education Text Book Series; 1)

Overlay

Intersection

- Intersect preserves only those features that fall within the area extent common to the inputs.
- Intersect uses the AND Boolean connector.
- This overlay operation yields the attribute table which is a join of the two input attribute tables. So, it is also called spatial join.
- For example, intersection can be useful to do analysis of LULC in local units if given two datasets one with LULC and other with ward data.



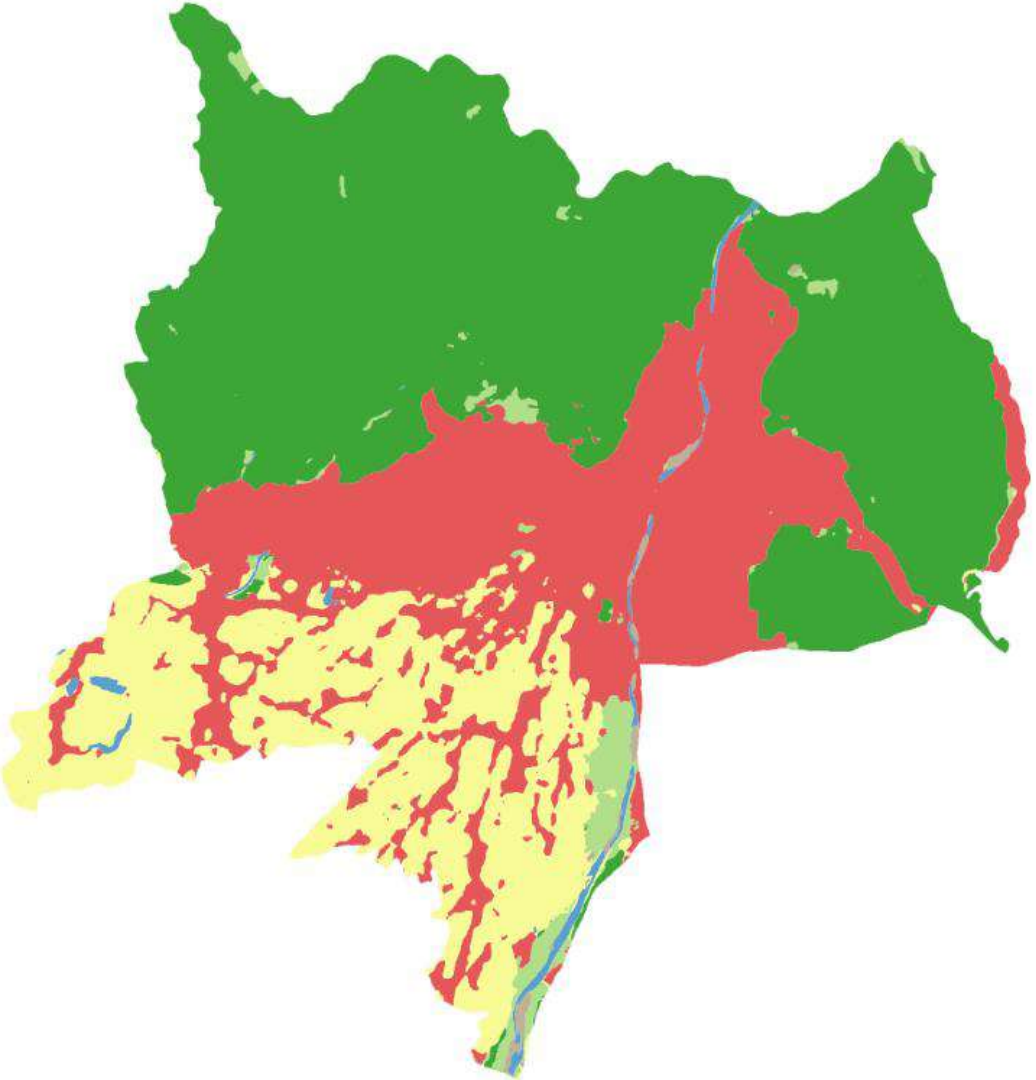
Source: Introduction to geographic information systems/Kang-tsung Chang,

Overlay: Intersection

LULC_vectorized — Features Total: 421,

	DN	Type
1	1	WaterBodies
2	1	WaterBodies
3	1	WaterBodies
4	1	WaterBodies
5	1	WaterBodies
6	1	WaterBodies
7	1	WaterBodies
8	1	WaterBodies
9	1	WaterBodies
10	1	WaterBodies
11	1	WaterBodies
12	1	WaterBodies
13	1	WaterBodies
14	1	WaterBodies
15	1	WaterBodies
16	1	WaterBodies
17	1	WaterBodies
18	1	WaterBodies
19	1	WaterBodies
20	1	WaterBodies
21	1	WaterBodies

Input 1: Attribute table



Source: http://keshavrajbhusal.com.np/uploads/materials/GISApplications_KRB.pdf

Overlay: Intersection

ButwalWards_44N — Features Total: 19, Filtered: 19, Selected: 0

	PROVINCE	PR_NAME	DISTRICT	PALIKA	TYPE	WARD
1	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	1
2	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	2
3	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	3
4	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	4
5	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	5
6	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	6
7	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	7
8	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	8
9	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	9
10	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	10
11	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
12	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
13	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	13
14	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	14
15	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	15
16	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	16
17	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	17
18	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	18
19	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	19

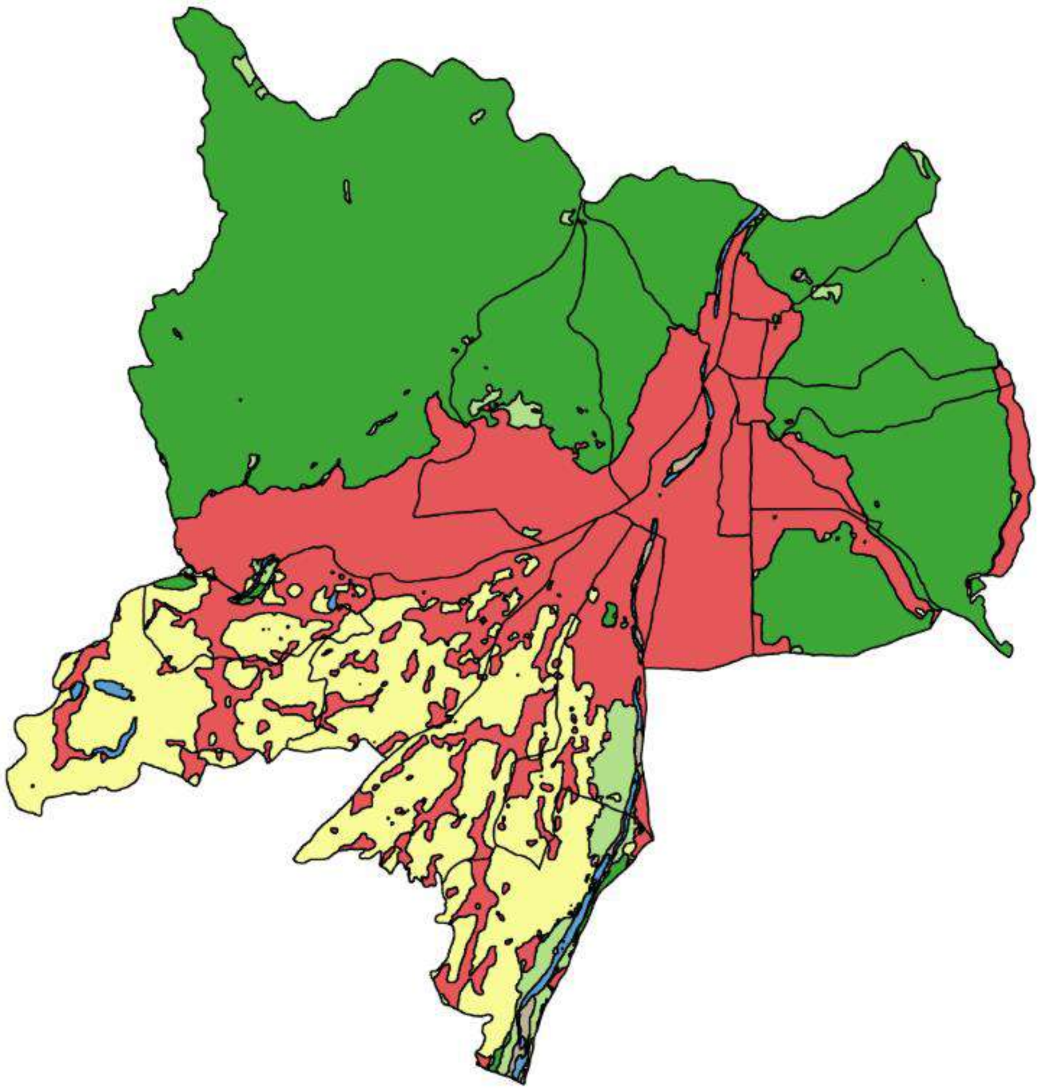
Input 2: Attribute table

Source: http://keshavrajbhusal.com.np/uploads/materials/GISApplications_KRB.pdf

Overlay: Intersection

LULCWards_Intersect — Features Total: 493, Filtered: 493, Selected: 0

	DN	Type	PROVINCE	PR_NAME	DISTRICT	PALIKA	TYPE_2	WARD
1	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	1
2	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	1
3	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	2
4	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
5	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
6	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
7	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
8	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
9	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
10	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
11	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
12	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
13	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
14	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
15	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
16	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
17	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
18	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	12
19	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
20	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	11
21	1	WaterBodies	5	Province No 5	RUPANDEHI	Butwal	Upamahanagar...	17

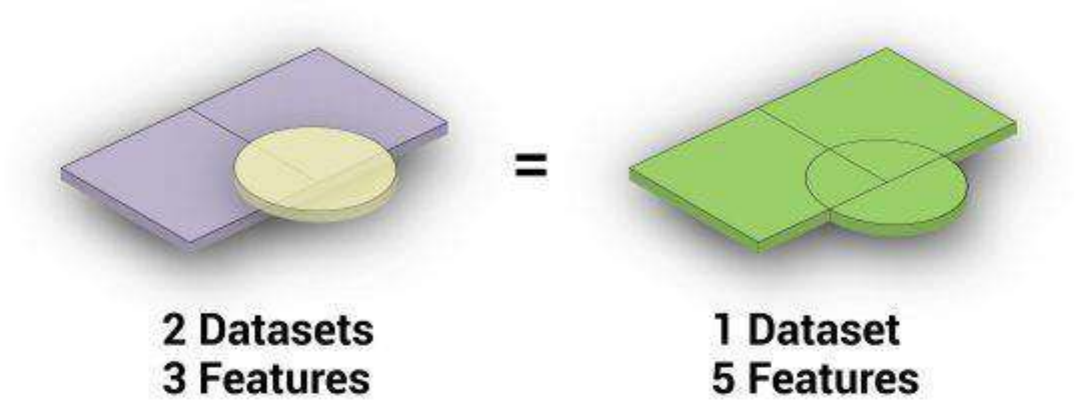


Output after intersection

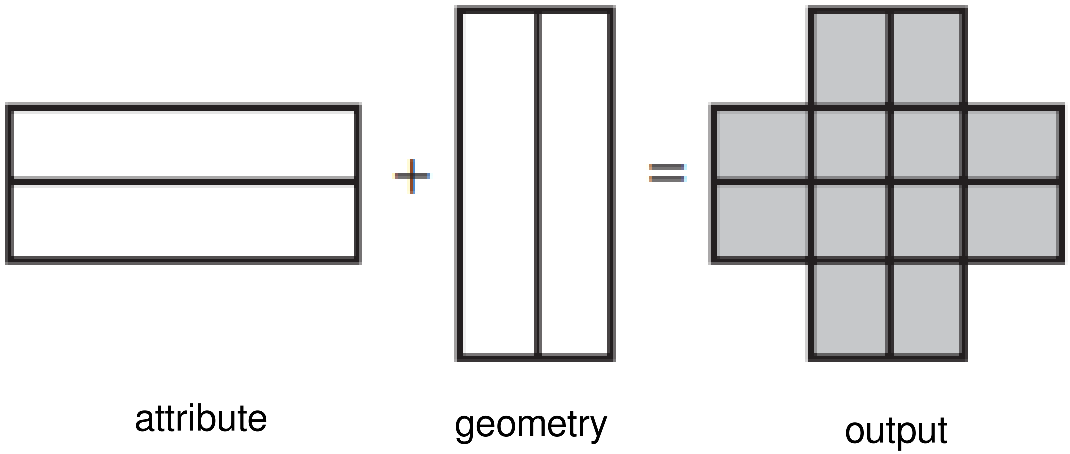
Source: http://keshavrajbhusal.com.np/uploads/materials/GISApplications_KRB.pdf

Overlay: ***Union***

- Union preserves all features from the inputs.
- The area extent of the output combines the area extents of both input layers.
- Union uses the OR Boolean connector.
- It creates lot of features which includes the input layer features along with the intersected region.



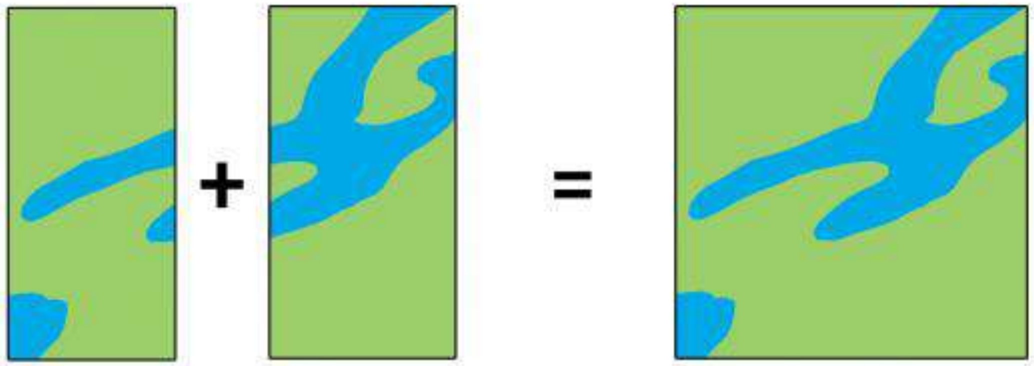
Source: https://gisgeography.com/geoprocessing-tools/



Source: Introduction to geographic information systems/Kang-tsung Chang,

Overlay: ***Merge***

- It is the overlay operation that combines data from multiple sources, then adds them into a new data set.
- It not only combines geometry, but it also merges attributes with the option to match fields from input datasets.
- For this, features in both datasets should have same geometry type (points, lines, or polygons).

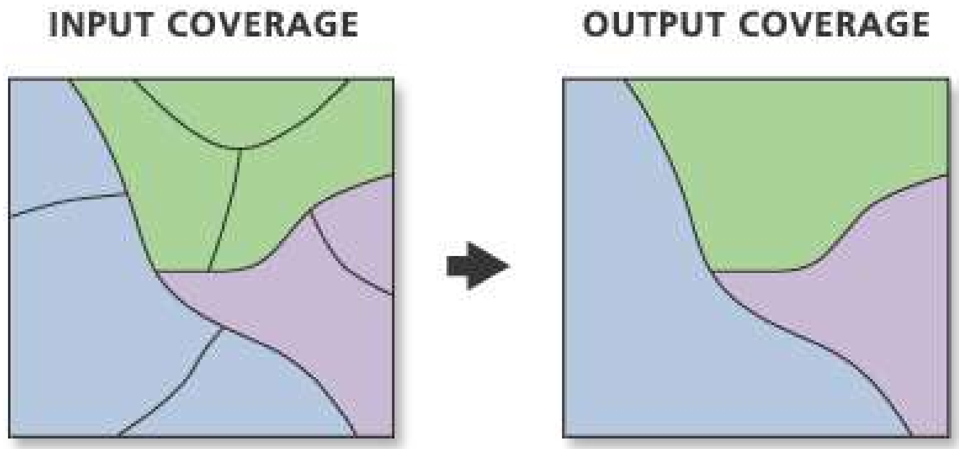


Layer 1 Layer 2 output

Source: https://gisgeography.com/merge-tool-gis/#:~:text=The%20Merge%20Tool%20combines%20data,%2C%20lines%2C%20or%20polygons).

Overlay: ***Dissolve***

- The Dissolve Tool unifies adjacent boundaries based on common attribute values.
- Dissolving features can help you to simplify the data and reduce unnecessarily large file sizes.
- This process is particularly useful when working with polygon layers, where adjacent polygons with the same attribute value can be dissolved into a single feature to reduce complexity and improve visualization.



Source: https://desktop.arcgis.com/en/arcmap/latest/tools/coverage-toolbox/dissolve.htm

Overlay: ***Dissolve***

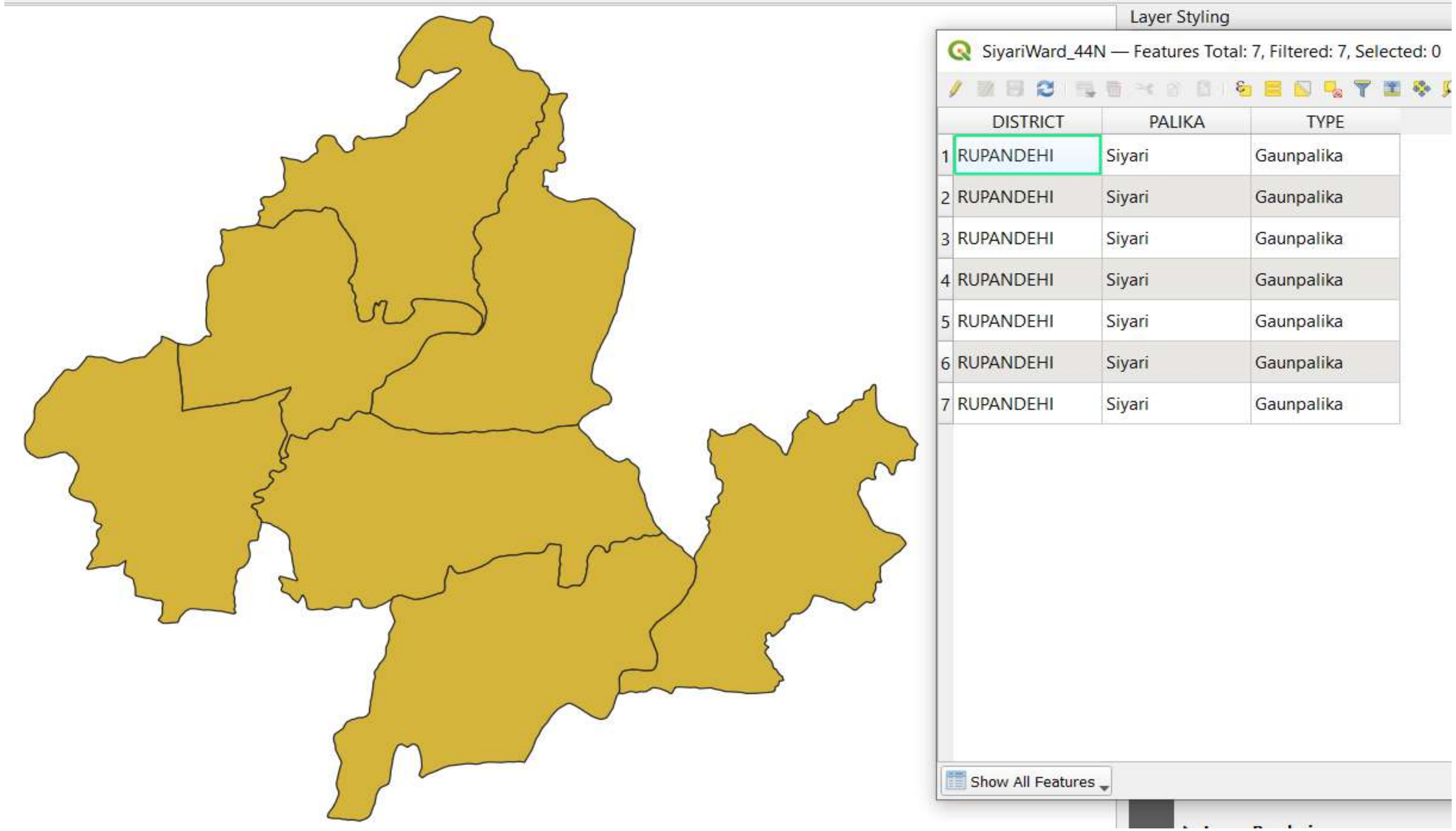


Fig: Before Dissolve

Source: https://mapadda.com/

Overlay: ***Dissolve***

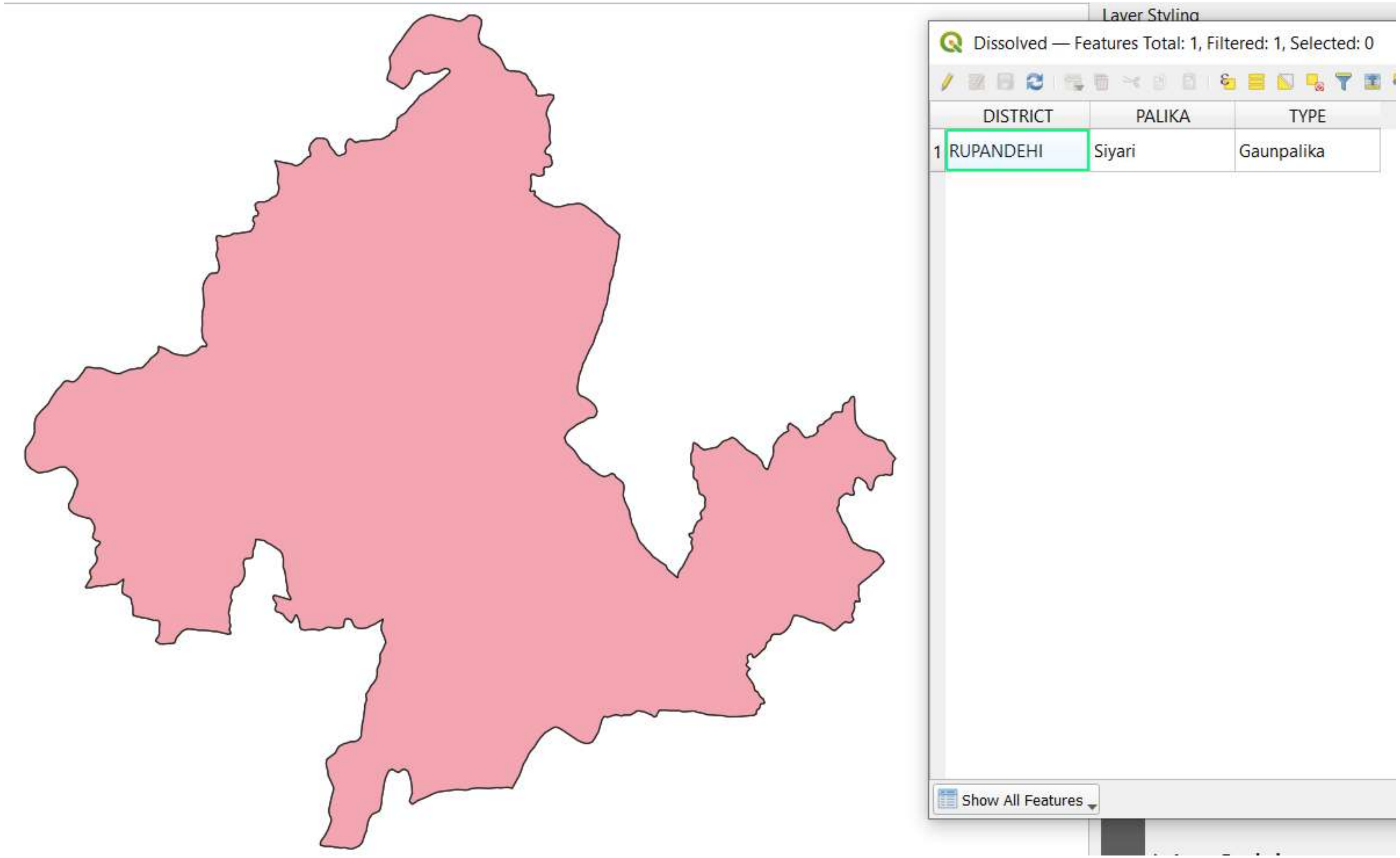


Fig: After Dissolve

Source: https://mapadda.com/

4.3 Data Visualization

Data visualization

- Visualization is simply the representation of an object or set of information as a chart or other image.
- In GIS, data visualization (or cartographic visualization) is the translation or conversion of spatial data from a database into graphics(mainly map like products).
- During the visualization process, cartographic methods and techniques are applied.
- Visualization process can be simple or complex, while the production time can be short or long depending on visualization purpose and environment (personal computer, WWW etc.).

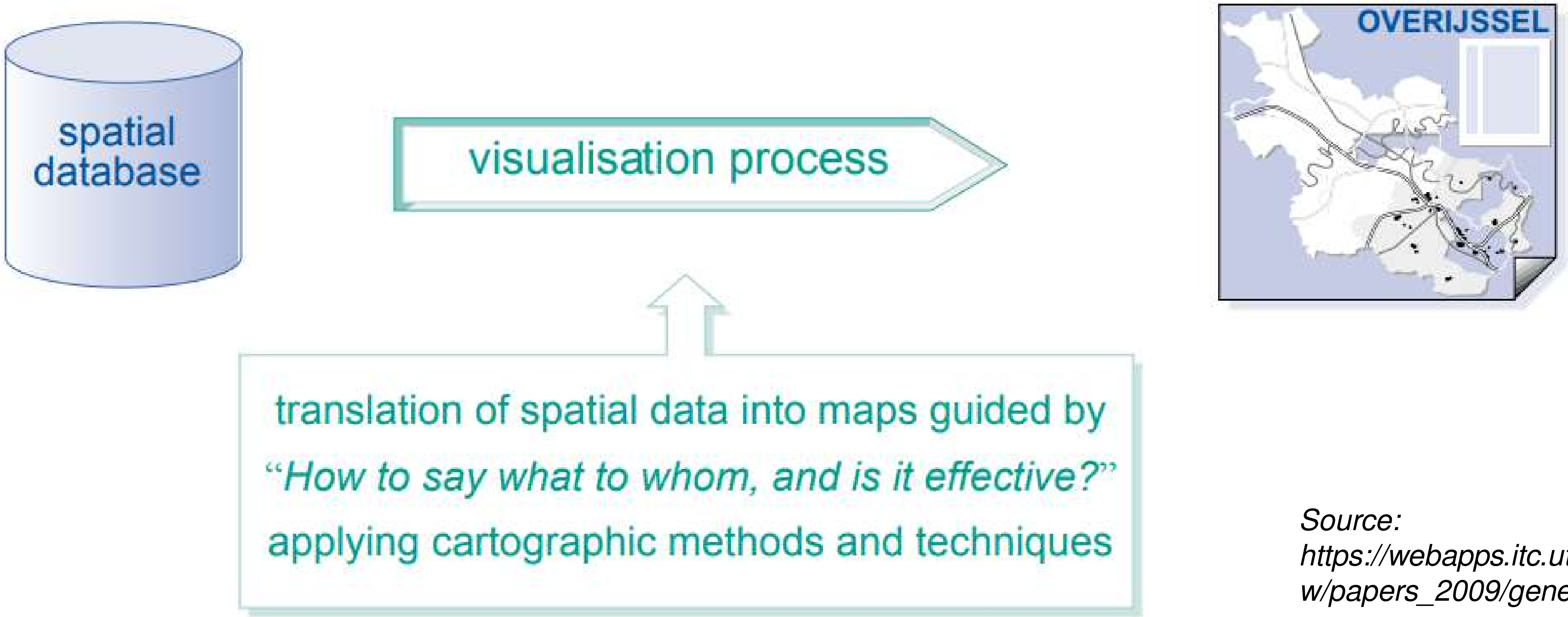


Fig: showing visualization process

Source: https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesgis.pdf

4.3 Data Visualization

Map elements / Parts of map

- The main elements of map are:

1. Map face / Data frame / Map frame – main body of map
2. Neat line – thick line to enclose map face
3. Border – space outside neat line, filled with grid or graticule value
4. Frame line – line surrounding border
5. Map title
6. Legend
7. Scale
8. North Arrow
9. Inset Map
10. Graticule/Grid
11. Map Citation / Marginal Information

- Marginal information includes: Author information, data source, date and other information

Map elements / Parts of map

- The main elements of map are:

1. Map face / Data frame / Map frame / Map Body

- It is the portion of the map that displays the data layers.
- This section is the most important and central focus of the map document.

2. Legend

- It is simply the description of color or layers used in map face.
- Without the legend, the color scheme on the map would make no sense to the viewer.

3. Map title

- It gives description of the subject matter of the map.
- It should be an answer to "What? Where? When?“
- The Map Title should reflect the purpose of the map clearly to the viewer.

Map elements / Parts of map

- The main elements of map are:

4. North Arrow

- It is used to show direction normally.
- We can orient the map according to maps north direction arrow and navigate or follow the map accordingly.
- It is normally placed at the top of map.

5. Scale

- The purpose of a map scale is to let the reader know the ratio of the distance on a map to the distance in the real world.
- Larger the scale features can be seen more distinctly and are zoomed in. Smaller the scale features are zoomed out and features are hard to recognize.
- Various type of scale is used in map. They may be verbal, graphical or numeric.
- Scale varies according to map size.

Map elements / Parts of map

- The main elements of map are:

6. Map Citation

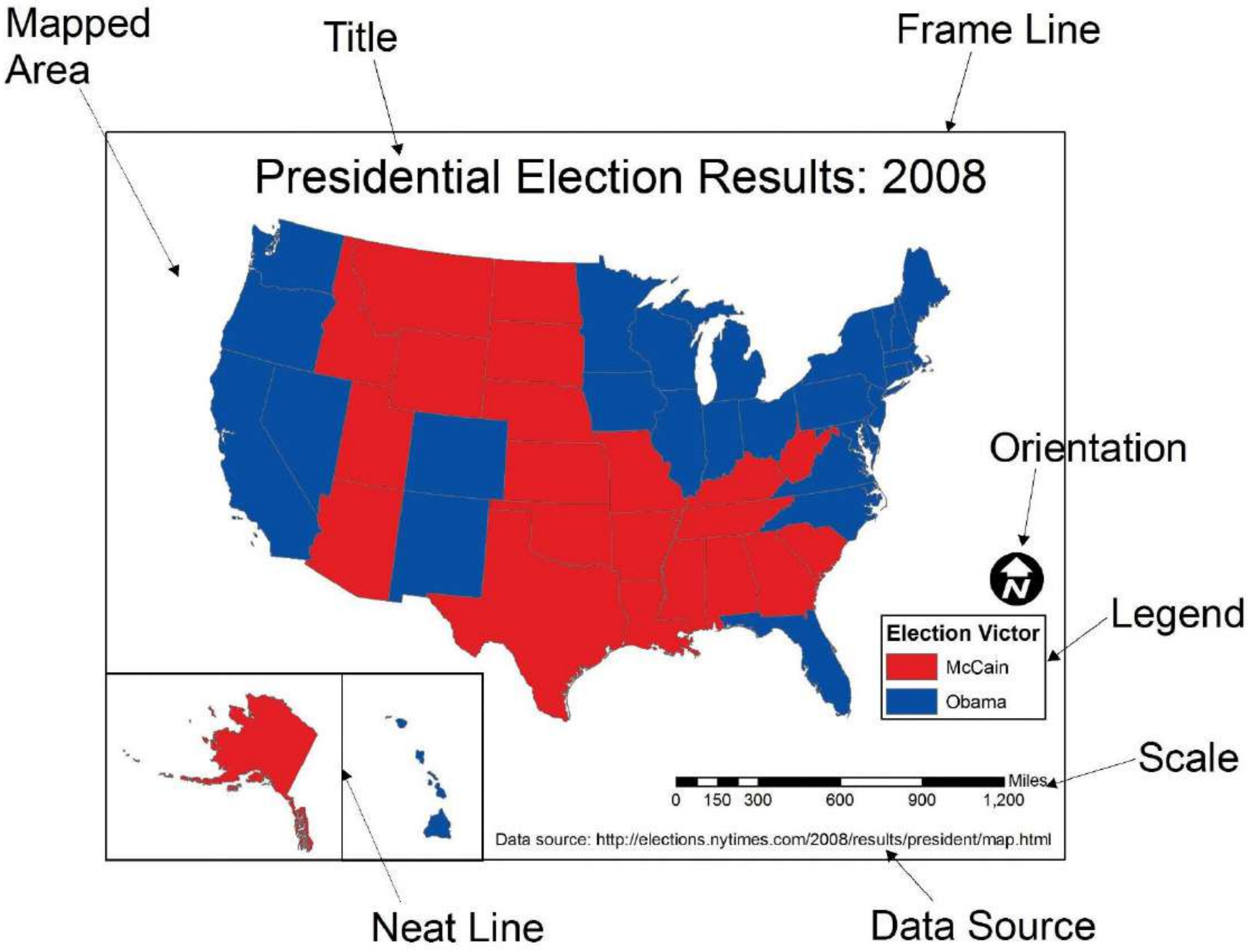
- The citation portion of a map constitutes the metadata of the map.
- It is the area where information related to data source, author's name, projection and other kind of things are written.
- Citations help the viewer determine the use of the map for their own purposes.

7. Neatline

- Neat lines are border boxes that are placed around individual map elements.
- A map layout may contain multiple neat lines surrounding other map elements helps to focus on a map elements.

Map elements / Parts of map

7. Neatline



Source: https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s13-03-cartographic-design.html

Map elements / Parts of map

8. Inset Map

- It is a second kind of map face.
- It is generally used to show the geographic extent of our main map face so that audience having no idea about that place can also get some location reference.
- If the audience are quite familiar with the place then it is not necessary to keep inset map.

9. Grid / Graticule

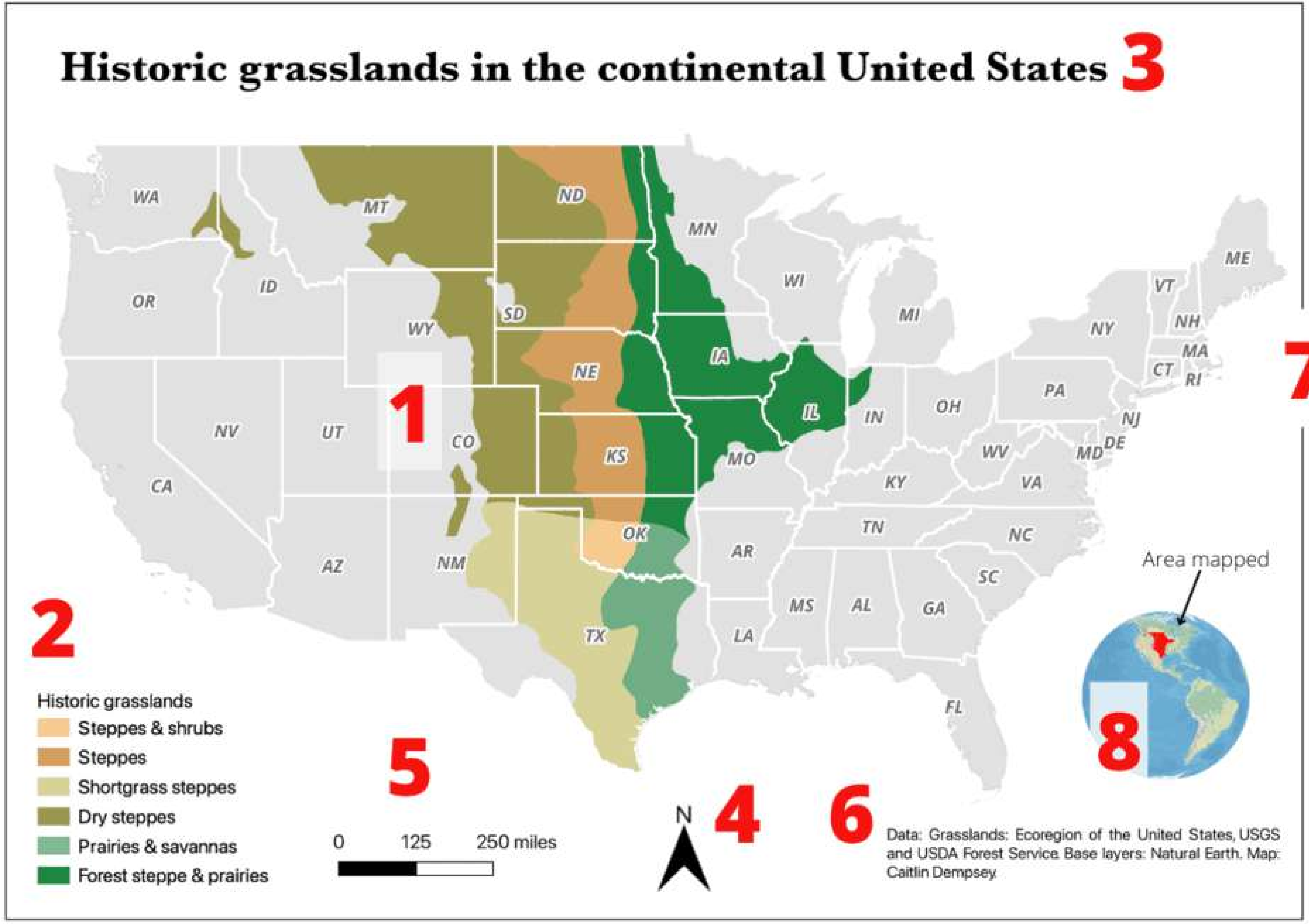
- They are the horizontal and vertical lines of latitude/longitude or easting/northing.
- The point of intersection of these two lines give us the coordinate (i.e location) of a map feature.

Map elements / Parts of map

10. Frame line

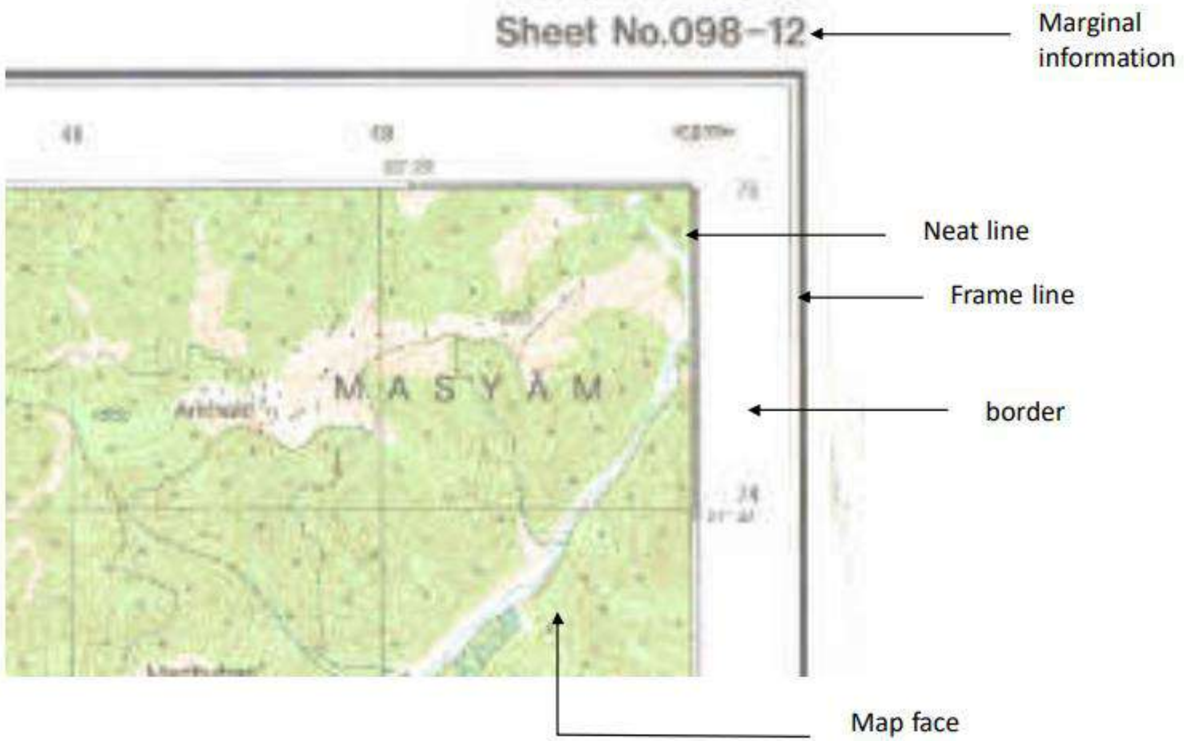
- A frame line encloses all elements on the map layout.

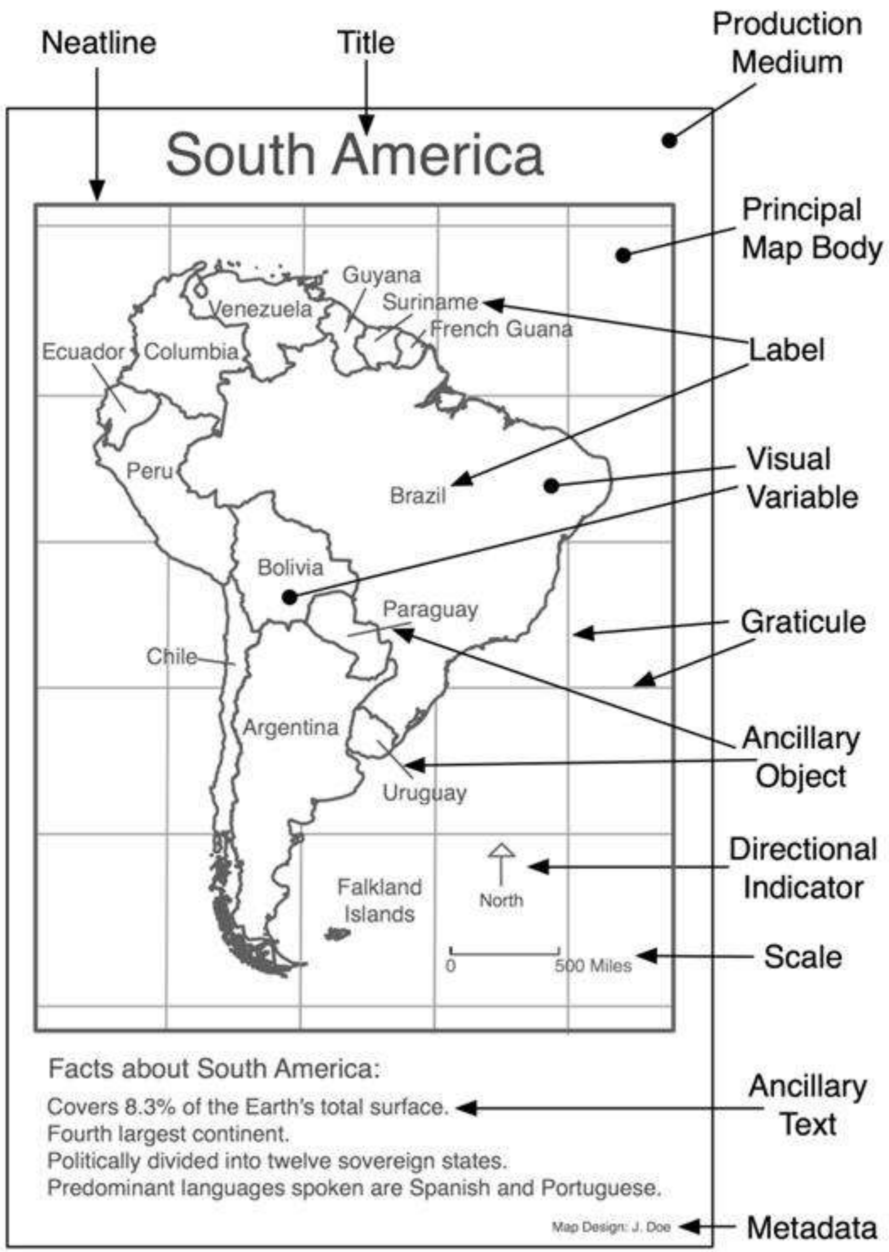
Map elements / Parts of map



Source: https://www.gislounge.com/whats-in-a-map/

Map elements / Parts of map





Source: https://open.maricopa.edu/gist/chapter/4-2-map-design-and-map-elements/

Map Design

- The purpose of map design is to enhance a map so that it is easy to understand and able to communicate the correct message or information.
- A well-designed map is balanced, coherent, ordered, and interesting to look at, whereas a poorly designed map is confusing and disoriented.
- Map design is both an art and a science.
- Cartographers usually study map design from the perspectives of layout and visual hierarchy.

1. Layout (Focus, Order, Balance)
2. Visual Hierarchy

4.3 Data Visualization

Map Design/Cartographic Principles

- There are mainly five principles of map design. They are:

1. Visual Contrast
2. Legibility
3. Figure-Ground Organization
4. Hierarchical Organization
5. Balance

Map Design Principles

1. Visual Contrast

- Visual contrast relates how map features and page elements contrast with each other and their background.
- A well-designed map with a high degree of visual contrast can result in a crisp, clean, sharp-looking map. The higher the contrast between features, the more some features will stand out (usually features that are darker or brighter).
- Contrast in color can separate the figure from the ground. Cartographers often use a warm color (e.g., orange to red) for the figure and a cool color (e.g., blue) for the ground.
- Features that have less contrast appear to belong together.

Map Design Principles

1. Visual Contrast

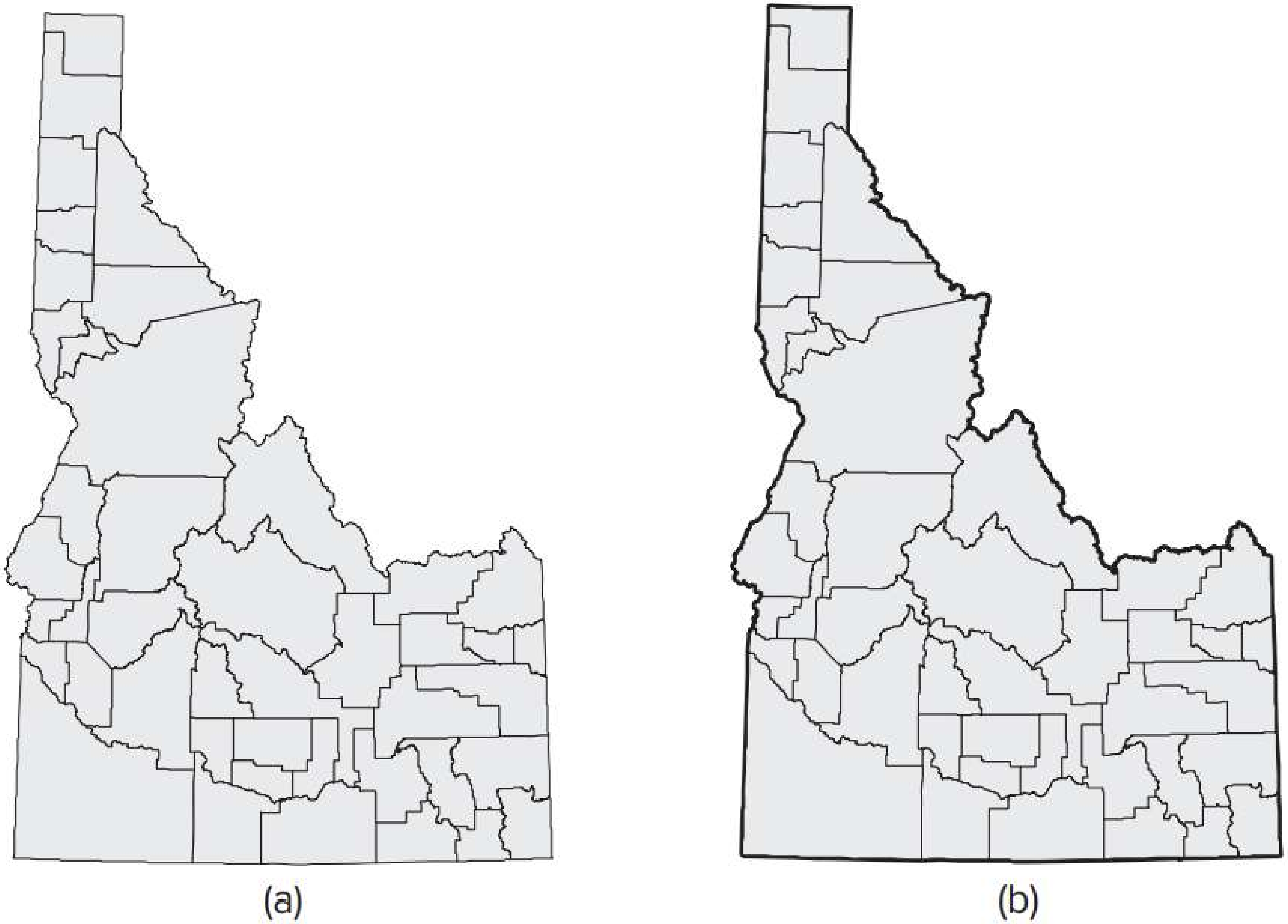
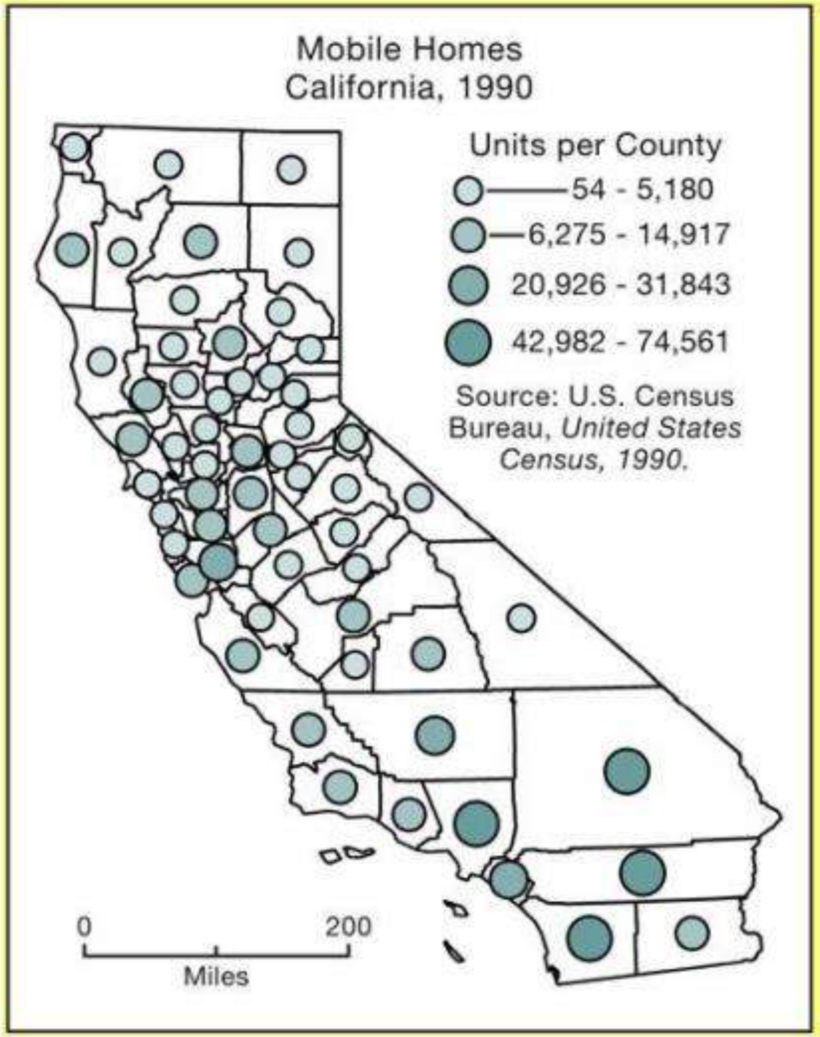


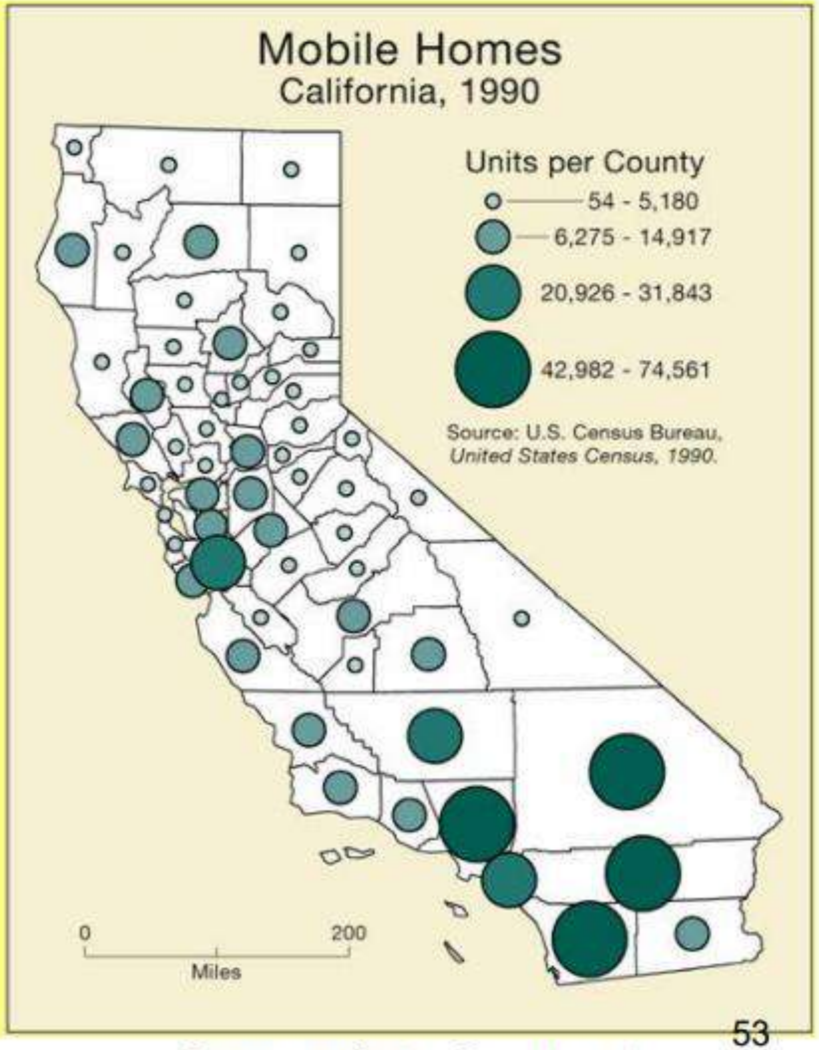
Fig: Contrast is missing in (a), whereas the line contrast makes the state outline look more important than the county boundaries in (b). Contrast in size or width can make a state outline look more important than county boundaries and larger cities look more important than smaller ones.

Map Design Principles

1. Visual Contrast



Inappropriate Contrast



Appropriate Contrast

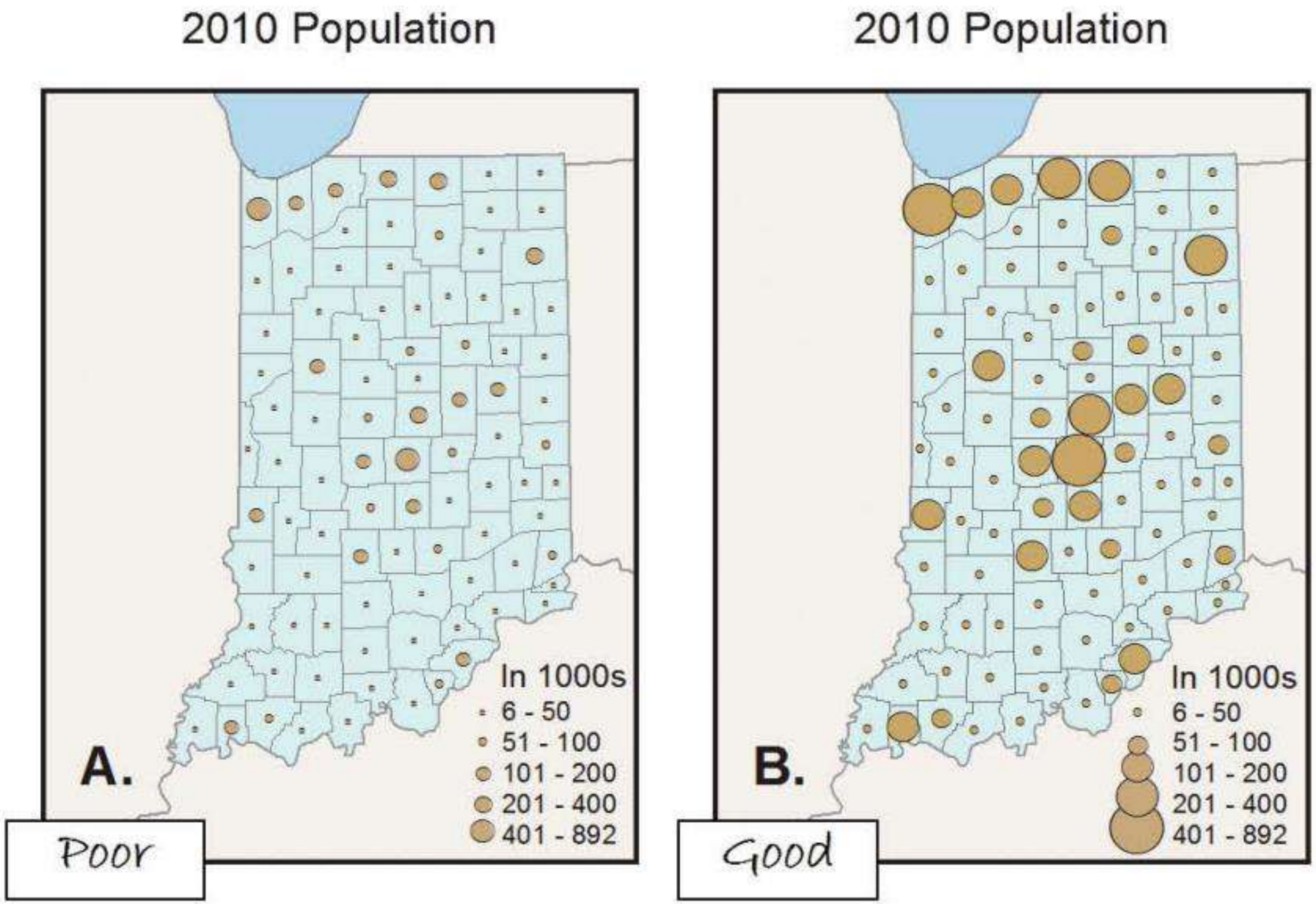
Map Design Principles

2. Legibility

- Legibility is the ability to be seen and understood.
- Legibility depends on good decision making when selecting symbols. Choosing symbols that are familiar and are appropriate sizes results in symbols that are effortlessly seen and easily understood.

Map Design Principles

2. Legibility

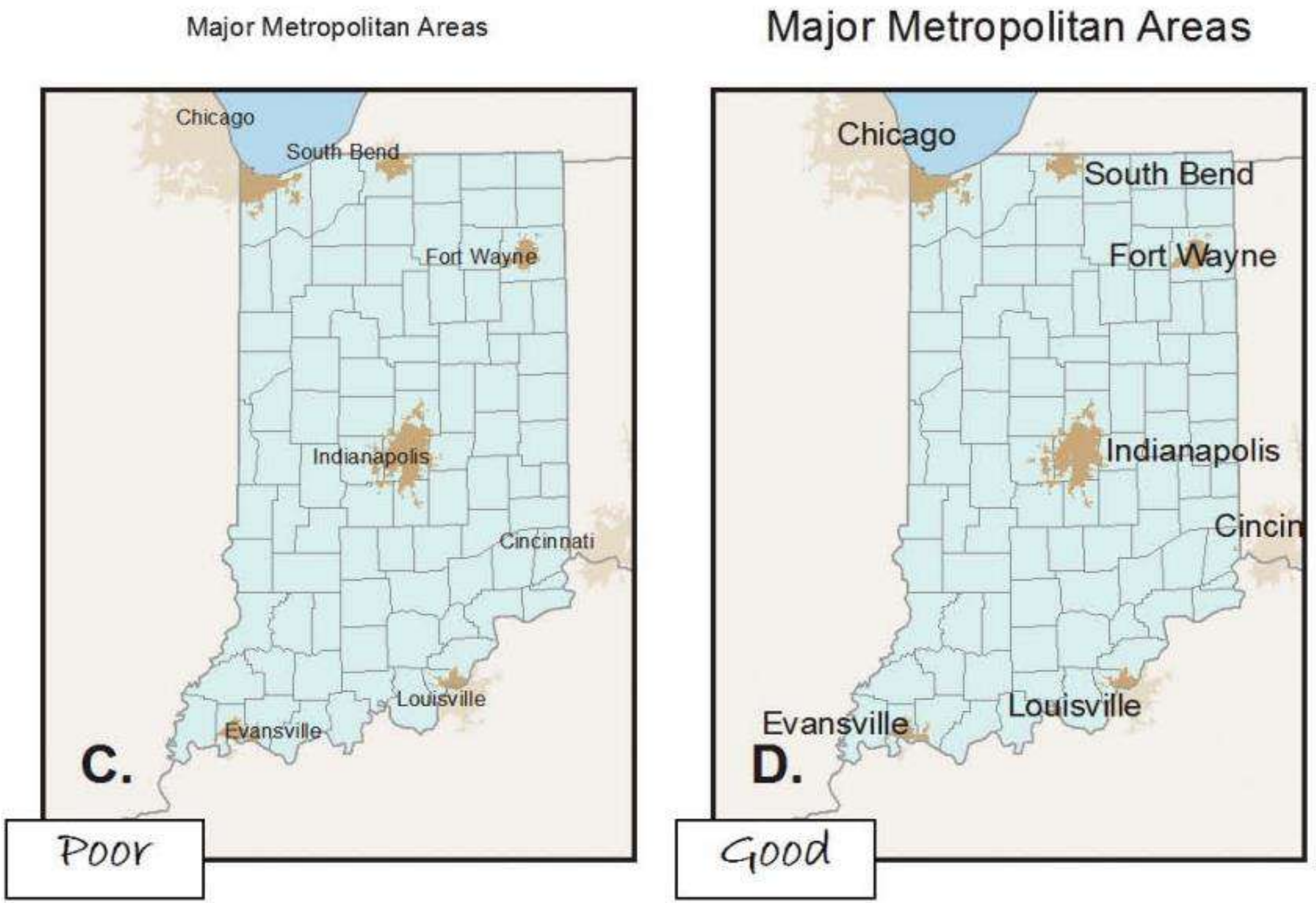


Source: https://www.esri.com/news/arcuser/0112/files/design-principles.pdf

Fig: Symbols (A) and text (C) that are too small are illegible. Appropriately sized symbols (B) and text (D) can be easily distinguished and read

Map Design Principles

2. Legibility



Source: https://www.esri.com/news/arcuser/0112/files/design-principles.pdf

Fig: Symbols (A) and text (C) that are too small are illegible. Appropriately sized symbols (B) and text (D) can be easily distinguished and read

Map Design Principles

3. Figure-Ground Organization

- It simply refers to the distinction between the map face(figure) and the background (ground) in a map.
- Cartographers use this design principle to help map readers focus on a specific area of the map.
- Color and contrast play an important role in figure-ground relationships. The less contrast in a map, the poorer the figure-ground effect, and likely the poorer the overall hierarchy.

Map Design Principles

3. Figure-Ground Organization

Fig: Color and brightness contrast let the land come forward from the ocean, and things like the river come forward from the land (Good figure-ground).

Map Design Principles

3. Figure-Ground Organization

Fig: Not so good figure-ground. Low contrasts make it harder and slower to pick out any salient elements.

Map Design Principles

4. Visual Hierarchy

- Visual hierarchy is used to emphasize certain important features on a map over less important features according to the map's purpose.
- An effective visual hierarchy attracts the map user's eyes to the most important aspects of the map first, and to less important aspects later.
- It results in maps that are easier to interpret, and are more attractive.
- It is the graphical representation of the intellectual hierarchy.

Map Design Principles

Intellectual hierarchy

- It is the ordering or ranking of all symbols and elements in the map according to their relative importance.
- The following is a general hierarchy for thematic maps.

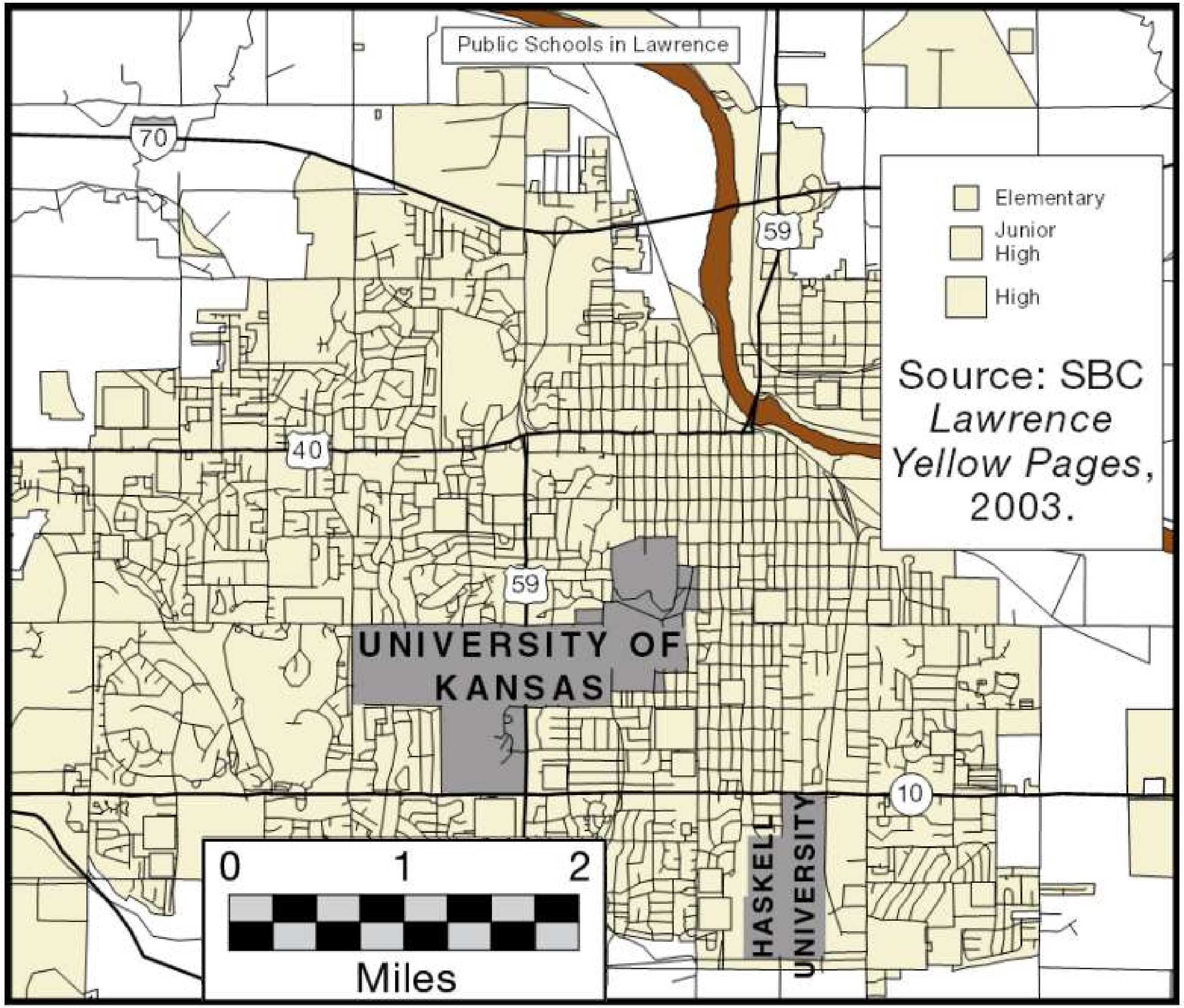
Most Important

- Thematic symbols and type labels
- Title, subtitle, and legend
- Base information (boundaries, roads, etc.)
- Scale and north arrow
- Data source and notes
- Frame and neat lines

Least Important

Map Design Principles

4. Visual Hierarchy



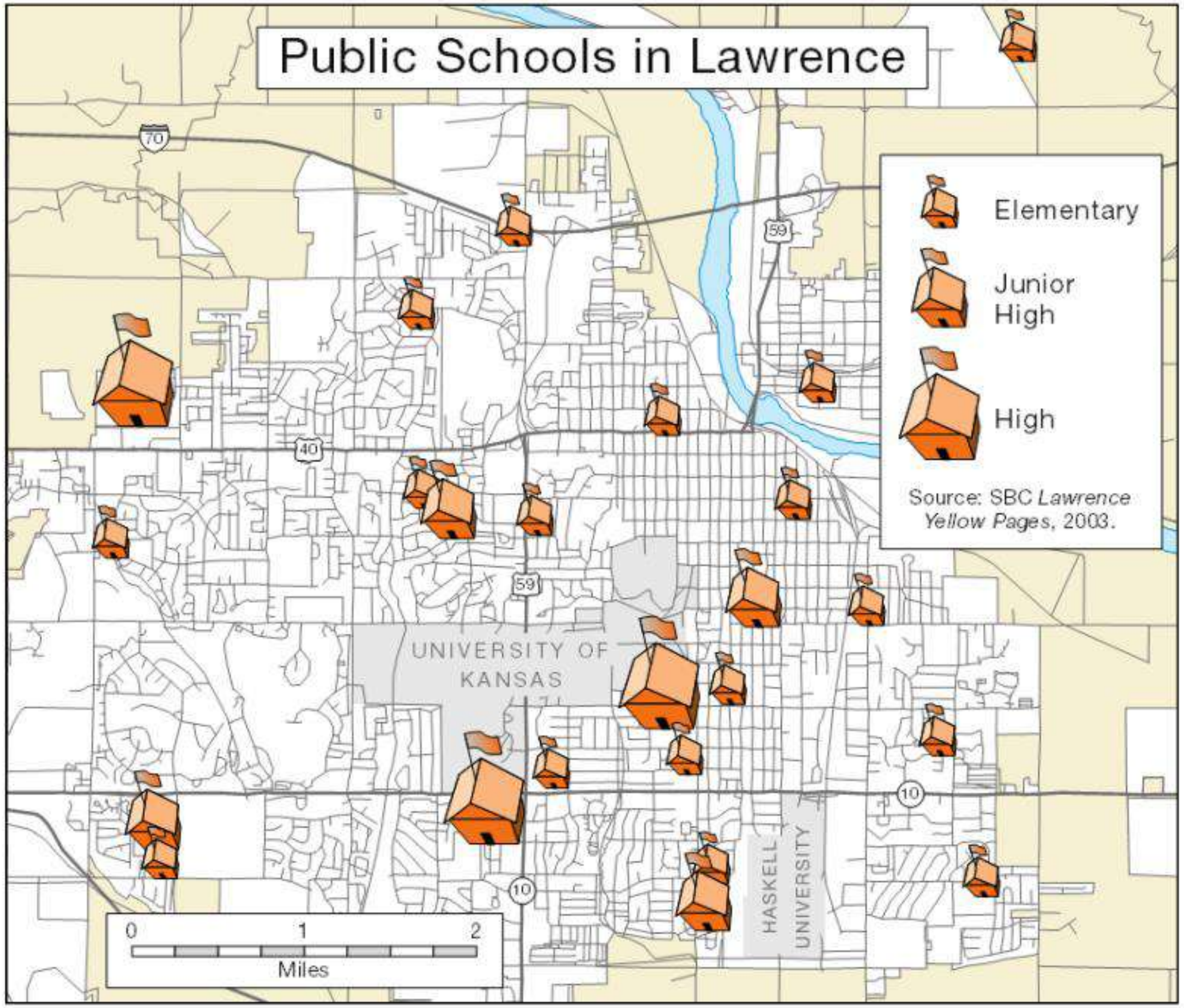
Inverted
(incorrect)
Visual Hierarchy

Based on an Inverted Intellectual Hierarchy

- Frame and neat lines
- Data source and notes
- Scale and north arrow
- Base information
- Title, subtitle, and legend
- Thematic symbols and type

Map Design Principles

4. Visual Hierarchy



Correct
Visual Hierarchy

Based on a Correct Intellectual Hierarchy

- Thematic symbols and type
- Title, subtitle, and legend
- Base information
- Scale and north arrow
- Data source and notes
- Frame and neat lines

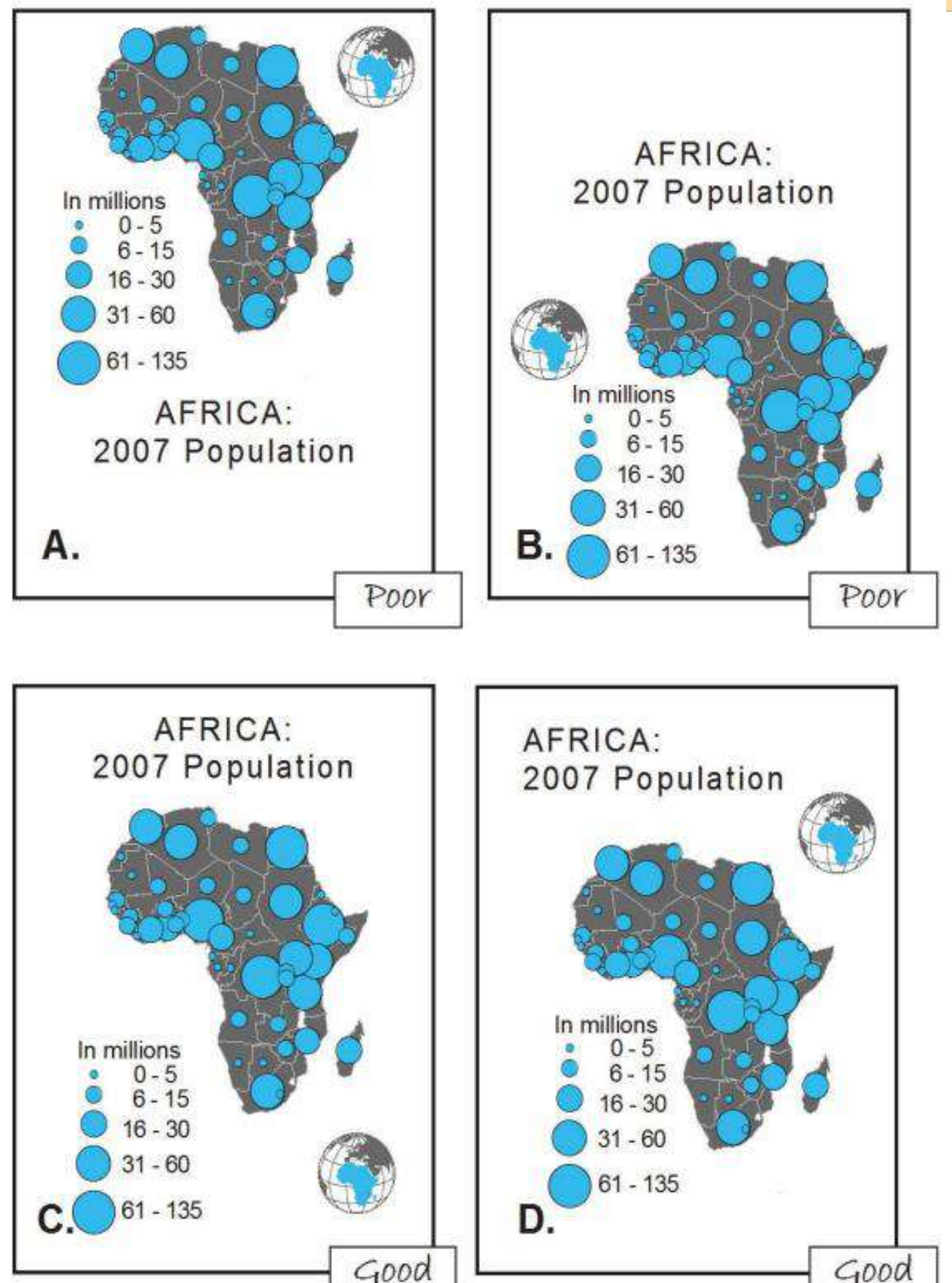
Map Design Principles

5. Balance

- Balance involves the organization of the map and other elements on the page.
- A well balanced map page results in an impression of equilibrium and harmony.
- It should not give the map reader an impression that the map “looks” heavier on the top, bottom, or side.

Map Design Principles

5. Balance



Source: https://www.esri.com/news/arcuser/0112/files/design-principles.pdf

4.3 Data Visualization

Output map preparation

The process of preparing a complete map within GIS involves the following main steps:

1. Data collection

- It includes collecting geographic data (vector or raster data) from primary or secondary source.

2. Importing data

- It involves loading spatial data into GIS softwares like QGIS, ArcGIS etc.

3. Data Styling and Symbolization:

- It includes customizing the appearance of map features through symbology by defining the colors (fill or stroke), line styles, point symbols, and transparency for each layers.

4. Layers Order and Labelling

- Layer order refers to arrangement of the layers in map according to visual hierarchy ensuring important layers are on top and less important ones are below.
- Label of features are turned on showing their attribute field information in case of necessity.

Output map preparation

The process of preparing a complete map within GIS involves the following main steps:

5. Map Layout and Composition:

- It includes designing the map layout by adding elements like a map title, legend, scale bar, north arrow etc.
- It also involves adjusting the size and orientation of map and setting the paper size that is intend to print.

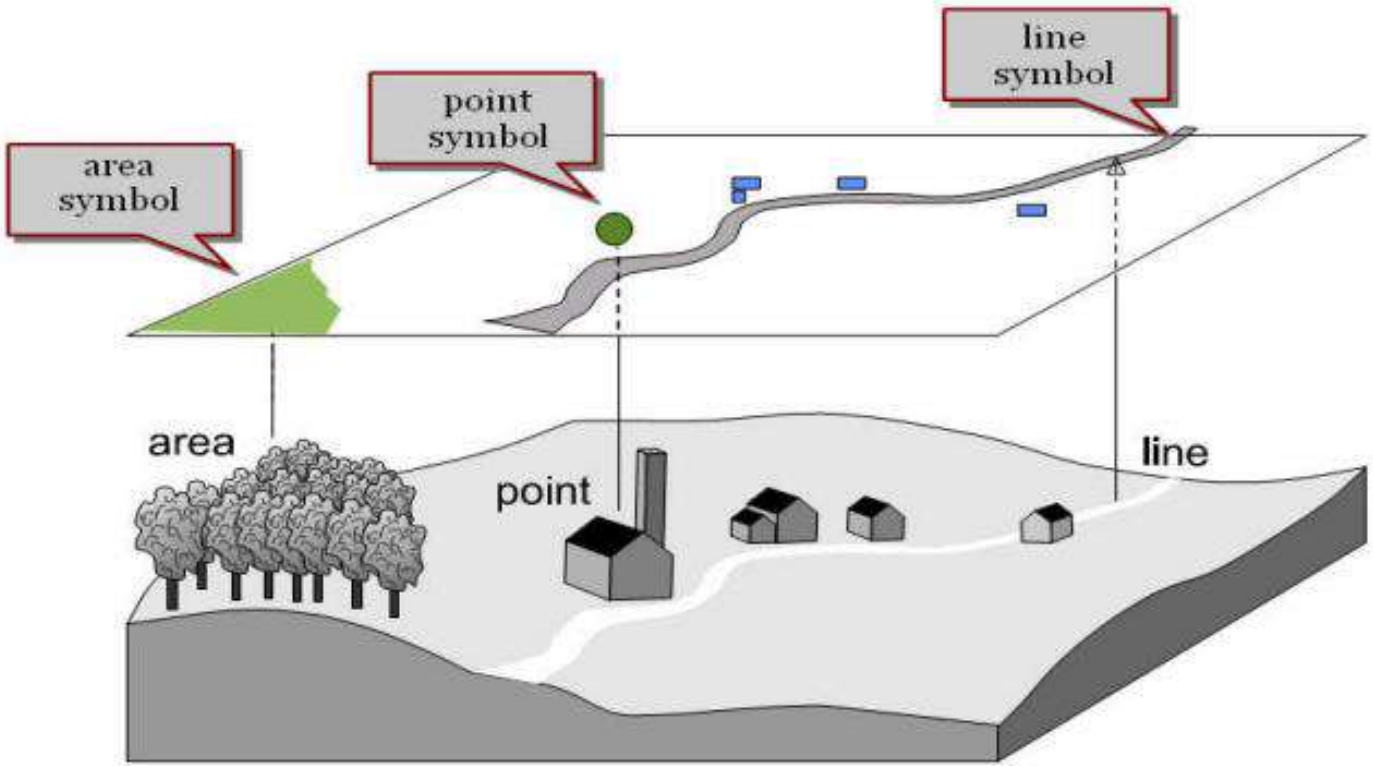
6. Map Export:

- It includes exporting map in different formats like PDF, JPEG, PNG, or even web formats for online use once you are satisfied with the map's appearance.

4.3 Data Visualization

Map symbolization

- Map symbolization or symbolization is the process of assigning symbols to represent features on a map.
- Symbols can be iconic (pictorial), geometric or abstract, textual.
- We can symbolize point features as dots, squares, triangles, flags, or other shapes and we can symbolize line features using solid, dashed, or other patterns.



- Point symbols
- Line symbols
- Area symbols

4.3 Data Visualization

Map symbolization

Cartographic symbols

a. Point symbols

- It represents location and characteristics of features of small territorial extent (in addition to map scale).
- There are 3 types of point symbols depending upon their construction. They are:

1. Geometric / abstract
2. Pictorial / descriptive
3. Textual

Map symbolization

Cartographic symbols

a. Point symbols

Geometric point symbol

• Simple geometrical shapes such as circles, squares, rectangles, triangles etc.

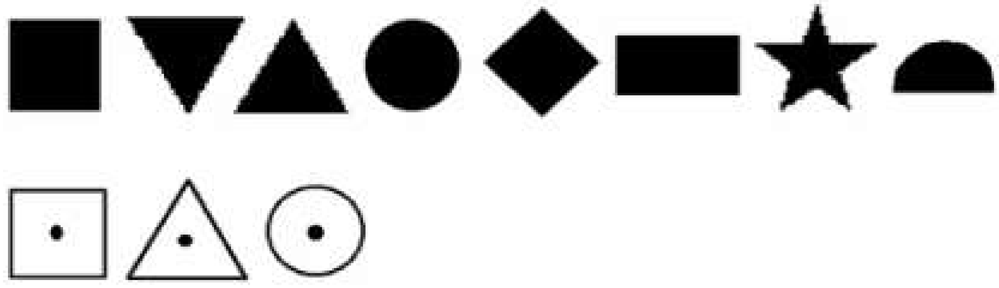
Advantage

• Easy to construct

• Cover relatively small area of map with high positional accuracy

Disadvantage

• No direct visual relationship to the object represented.



Map symbolization

Cartographic symbols

a. Point symbols

Pictorial point symbol

Advantage

- Easiest to interpret, giving as they do, a life-like image of the actual feature.

Disadvantage

- Difficult to find exact location of the object being represented.
- Difficult to construct.
- Occupy relatively more space.
- Difficult to adjust the optical / visual weight.

4.3 Data Visualization

Map symbolization

Cartographic symbols

a. Point symbols

Textual symbol

- Abbreviation or short forms of the feature category
- For example, P for phosphorous, Fe for iron.

Advantage

- Easy to construct.
- Easy to understand and remember.

Disadvantage

- Do not exactly show the location.
- Difficult to adjust the optical / visual weight.
- Mixing up with the map text.
- Graphics more efficient.
- Occupy much space.

Map symbolization: Cartographic symbols

b. Line symbols

- It is used to represent the linear, in geometric sense, ground objects.
- Width of such linear objects can not be represented in the map scale.
- For example, administrative boundary, telephone and electricity lines, road, river, canal.

c. Area symbol

- It is used to represent the ground objects having considerable extent in all direction which can be shown at the map scale.
- It may be abstract or pictorial.

References

Database query

- *https://surveyinggroup.com/what-is-a-spatial-database-and-why-do-we-need-it/#:~:text=What%20makes%20it%20different%20from,to%20speed%20up%20database%20performance*.
- *https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/s10-02-searches-and-queries.html*

Spatial SQL

- *https://forrest.nyc/what-is-spatial-sql/*

Map elements

- *https://open.maricopa.edu/gist/chapter/4-2-map-design-and-map-elements/*

Map Design concept:

https://www.igntu.ac.in/eContent/IGNTU-eContent-403493872964-BA-GeographyandRegionalDevelopment-4-Dr.RichaChaturvedi-CartographicTechniqueinGeography-4.pdf

https://www.esri.com/news/arcuser/0112/files/design-principles.pdf