Subject	Geology
Paper No and Title	**Remote Sensing and GIS**
Module No and Title	GIS Data Structure: Raster vs. Vector
Module Tag	RS & GIS XXIII

Principal Investigator	Co-Principal Investigator	Co-Principal Investigator
Prof. Talat Ahmad ***Vice-Chancellor*** Jamia Millia Islamia Delhi	**Prof. Devesh K Sinha** Department of Geology University of Delhi Delhi	**Prof. P. P. Chakraborty** Department of Geology University of Delhi Delhi
Paper Coordinator	**Content Writer**	**Reviewer**
Dr. Atiqur Rahman Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia Delhi	**Dr. Atiqur Rahman** Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia Delhi	**Dr. Iqbal Imam** Aligarh Muslim University Aligarh

In GIS many times we noted about the term Data and Information that has been refer to the same thing. Here, it is necessary to understand the differences between the two. The concern of data is to describe the facts, characteristics and measurement of an object. For example, status of air pollution in Delhi, impact of global warming on mountains glacier, distance from Delhi to Mumbai of different transportation mode etc. Whereas, information refers to the knowledge and insight which is acquired through the collection and analysis of data. A GIS data structure is a mathematical construct for depicted geographic feature as data. The term GIS that can refer to many methods, processes and technologies. It is an associated with several applications related to environment, planning, transportation and many location-enabled services that can be analyse and visualize. Location as the key index variable is used in GIS that can be relate unrelated information. Location and position of features on earth surface has been recorded as x, y and z coordinates. The spatial-temporal location and extent references should be associated to one another to exact location or extent. Many datasets are being used in Geographic Information System (GIS) environment related to research, planning and utility projects etc. GIS data are generated from maps, aerial photographs, satellite imagery and field survey by scanning and digitising using GIS software. GIS data represented in the form of spatial data (location and geometry) and tabular data (characteristics of feature).

Spatial data is described in two ways i) Raster data systematically arranged rows and column in the form of cells ii) vector data represented in the form of point, line and area (polygon). Geospatial data consists of spatial component which is described the location of geographic feature on earth surface whereas non-spatial or tabular or attribute data component used to explain its characteristics. GIS data is divided into two broad categories:

1) Spatial data

The sources of spatial data are surveying and remote sensing. The surveyors used most advanced equipment such as Global Positioning System (GPS) and total station (an electronic theodolite) for more precise ground survey instead of

conventional tool. Aerial photographs and satellite images are the sources of real time geographic information. The storage of spatial data in GIS environment has several ways. Since 1990's GIS application has been used for spatial analysis and spatial databases (also known as geodatabases) for data storage. Spatial database use other techniques that is different from table to store spatial features. The usage of spatial data involves various discipline and the discipline that deals all aspect of spatial data handling is called Geoinformatics. Spatial data is also called geographic data that identified by geometry, geographic location and attribute that is described its characteristic, such as forest, ocean, town and others. The location and geometry of geographic features are stored in the form of coordinates (Latitude and Longitude) and topology. Spatial data manipulation or analyzed done with the help of attribute data in GIS environment that can be mapped.

GIS data models has a set of rules that is being used to described the aspects of real world in GIS domain. Two types of data models are being used to complete this task, Raster data models and Vector data models. The fundamental approaches to present the Spatial data in two ways:

a) **Raster data:** Raster data presented as a matrix or array of pixels (Picture Element). Pixel is the smallest unit of picture. It is stored spatial information in grid cells organized and accessed as rows and columns **(Fig. 1)**. The information of geographic features stored within the cells contain a number called Digital Number (DN) value. DN value of geographic features are dependent on feature's characteristics and reflectance value over the surface of the earth. Each geographic feature has own DN value arranged in regular grid. The spatial resolution of raster data is dependent on cell or pixel size. Spatial resolution of raster data increased with decreasing pixel size. Spatial details of raster data controlled by resolution. There are many sources of raster data such as satellite imagery, aerial photographs, scanned maps and other sources.

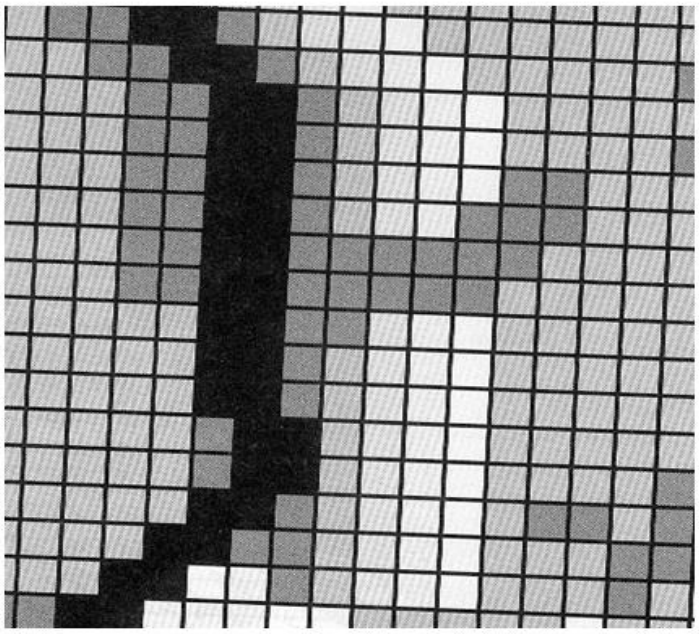


Fig. 1 Raster data

Properties of Raster data

- ✓ Raster data is a set of cells located by coordinate is used;
- ✓ Each cell is independently addressed with the value of an attribute.
- ✓ Each cell contains a single value and every location corresponds to a cell.
- ✓ One set of cell and associated value is a LAYER
- ✓ The linear dimension of each cell defines the spatial resolution of the data,
- ✓ The minimum mapping unit on 1:50000 scale is 3mmx3mm or 2.5 hectares
- ✓ Raster data models require a huge volume of data to be stored, fitness of data be limited by cell size

Raster data is divided into following two categories.

- **Discrete data**:- It is also called categorical, thematic, or discontinuous data. It is used to present discrete features in both raster and vector data models. The boundary of discrete features has known. For example, a forest is a discrete features within the surrounding landscape. A discrete feature is represented using same value neighbouring cells in raster data model. Roads, water bodies and built-up area are the examples of discrete features.

- **Continuous data:-** It is also called surface data or non-discrete data. Continuous data is divided into two types based on features they represent. first, to represent features using continuous data the value at each cell location measured by fixed registration point. For example height measured from ground surface as the fixed point. The another type of continuous data shows feature characterized by the way they move and represent features that gradually vary as they move across the surface from a source such as liquid and gas movement. It is impossible to measure every cell location of all continuous data because they are derived from discontinuous data. Interpolation method is used to obtained the continuous surface, which is based on the features characteristics.

There are various methods for encoding raster data from scratch. Few models are as follows:

1. Cell-by-cell raster encoding: This method encodes a raster by creating records for each cell value arranged in row and column. According to this method a big spreadsheet consists of cells and each cell represents a pixel in the raster data. This method is also known as exhaustive enumeration.

2. Run-length raster encoding: In this method cell values encodes in runs of similar valued pixels and can result in highly compressed image data. This method is useful when large groups of neighboring pixels have similar values and unuseful or less useful when neighboring pixel values vary widely.

3. Quad-tree raster encoding: In this method raster data is divided into hierarchy of quadrants and sub-divided based on similar valued pixels. When a quadrant is made completely from the similar value cells the division of the raster data is stops. A quadrant that cannot be sub-divided is known as leaf node.

b) Vector data: Vector data is used to present spatial information in point, line and polygon (area). Point data recorded in the form of (X,Y) coordinates. Line and polygon data is based on Arc-node having non-intersecting lines segment called arcs

and connecting set of arc forms area objects. The main merits of vector data has require minimum memory space, area and perimeter of polygon feature is estimated accurately, data handling/manipulation is fast and produce accurate results. Vector data used to captured the geographic location of discrete features such as roads, buildings, river, hospital and boundaries of other geographic features etc. The precise geo-location of features on earth surface are recorded in the form of (X, Y) in vector data.

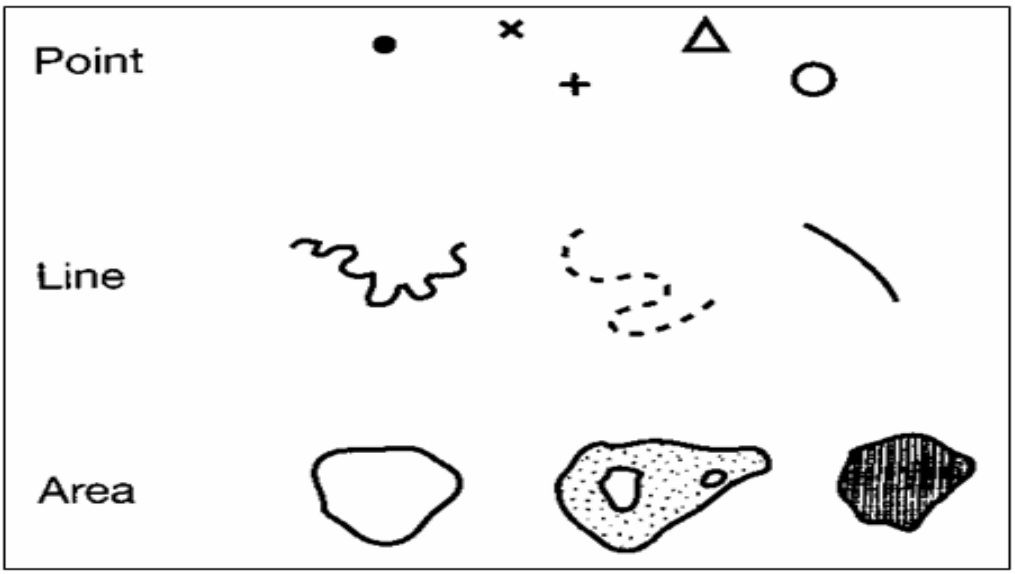


Fig. 2 Vector data

In GIS software, a point, line or a polygon feature shows as points or connect the points with line and displays them as line or the line enclosed area with a fill of colour and displays as polygon. There are many different ways of vector data models can be structured. Here, we examine two more common data structure.

1. **Spaghetti Data Model:** Each feature (Point, line and polygon) is presented as a string of X,Y coordinate group with no inherent structure in the spaghetti data model. In this model all lines to be a single strands of spaghetti that is formed into complicated shapes by the addition of many spaghetti strands. The polygons that lie adjacent to each other must be made up of their own strands of spaghetti. We can say in other words, all polygons

defined by its own set of X,Y coordinate pairs uniquely, even if the adjacent polygons share the same boundary. Due to this some redundancies within the data model is created and as a result efficiency is reduces.

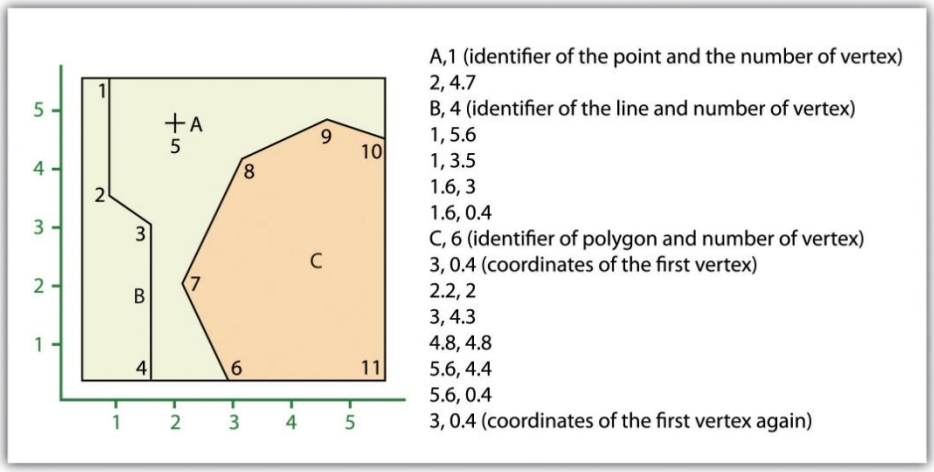


Fig. 3 Spaghetti Data Model

2. **Topological data model:** It is characterized by the inclusion of topological information within the dataset. Topology is a set of rules that model the relationships between neighboring points, lines, and polygons and determines how they share geometry **(Fig. 4).** For example, consider two adjacent polygons. In the spaghetti model, the shared boundary of two neighboring polygons is defined as two separate, identical lines. The inclusion of topology into the data model allows for a single line to represent this shared boundary with an explicit reference to denote which side of the line belongs with which polygon. Topology is also concerned with preserving spatial properties when the forms are bent, stretched, or placed under similar geometric transformations, which allows for more efficient projection and re-projection of map files.

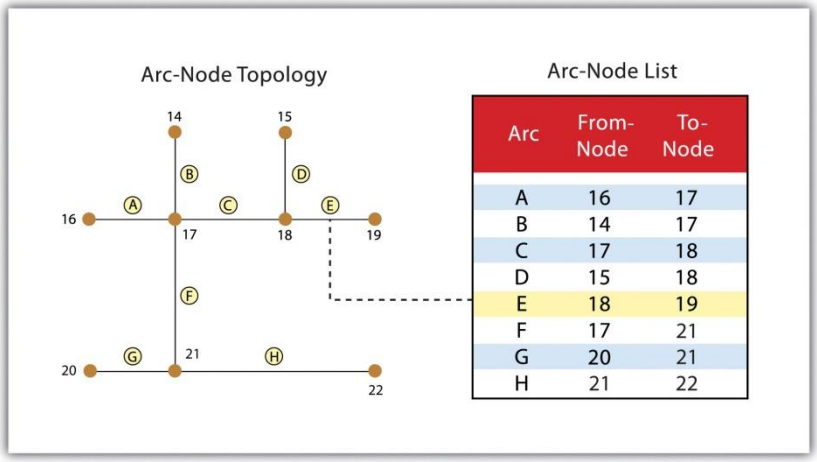


Fig. 4. Arc-Node Topology

Arc-polygon topology requires that all arcs in a polygon have a direction (a from-node and a to-node), which allows adjacency information to be determined **Fig. 5.** Polygons that share an arc are deemed adjacent, or contiguous, and therefore the "left" and "right" side of each arc can be defined. This left and right polygon information is stored explicitly within the attribute information of the topological data model.

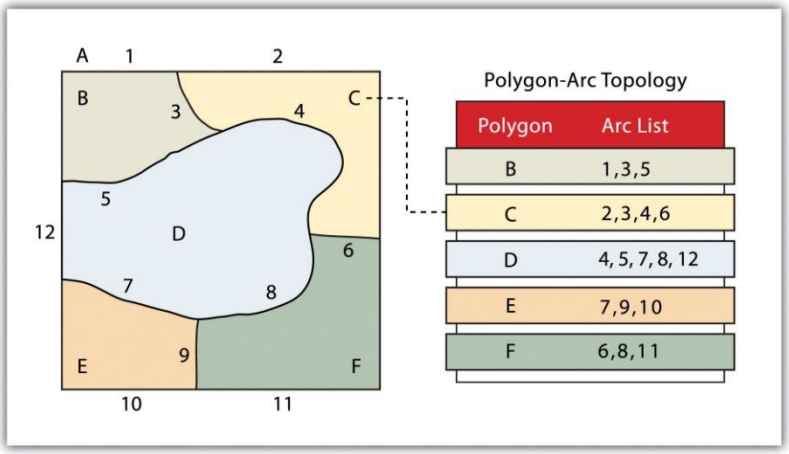


Fig. 5. Arc-Polygon Topology

Advantages/Disadvantages of the Vector Model

In comparison with the raster data model, 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. This results in vector data tending to be more aesthetically pleasing than raster data.

Vector data tend to be more compact in data structure, so file sizes are typically much smaller than their raster counterparts. Although the ability of modern computers has minimized the importance of maintaining small file sizes, vector data often require a fraction the computer storage space when compared to raster data. Vector data is that topology is inherent in the vector model. This topological information results in very simple spatial analysis e.g. network analysis, proximity analysis, and spatial transformation by using a vector model.

The disadvantage 0f vector data structure tends to be much more complex than the simple raster data model. As the location of each vertex must be stored explicitly in the model, there are no shortcuts for storing data as that of raster models (e.g., the run-length and quad-tree encoding methodologies).

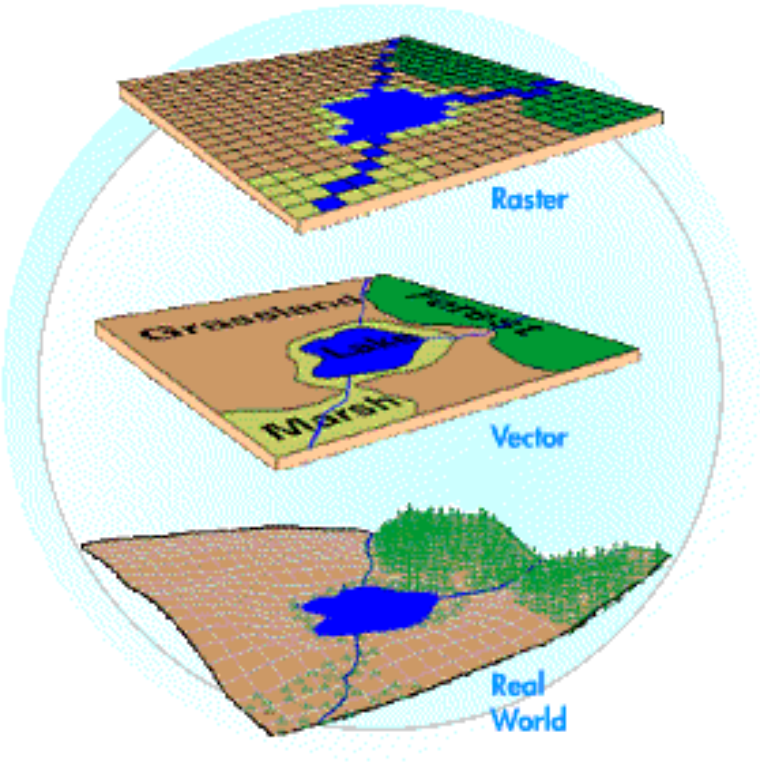
Another disadvantage is the implementation of spatial analysis can also be relatively complicated due to minor differences in accuracy and precision between the input datasets. Similarly, the algorithms for manipulating and analyzing vector data are complex and can lead to intensive processing requirements, particularly when dealing with large datasets.

Properties of Vector data

- Vector data utilizes points, lines, and polygons to represent earth surface feature features in a map.
- Topology is an informative geospatial property that describes the connectivity, area definition, and contiguity of interrelated points, lines, and polygon.

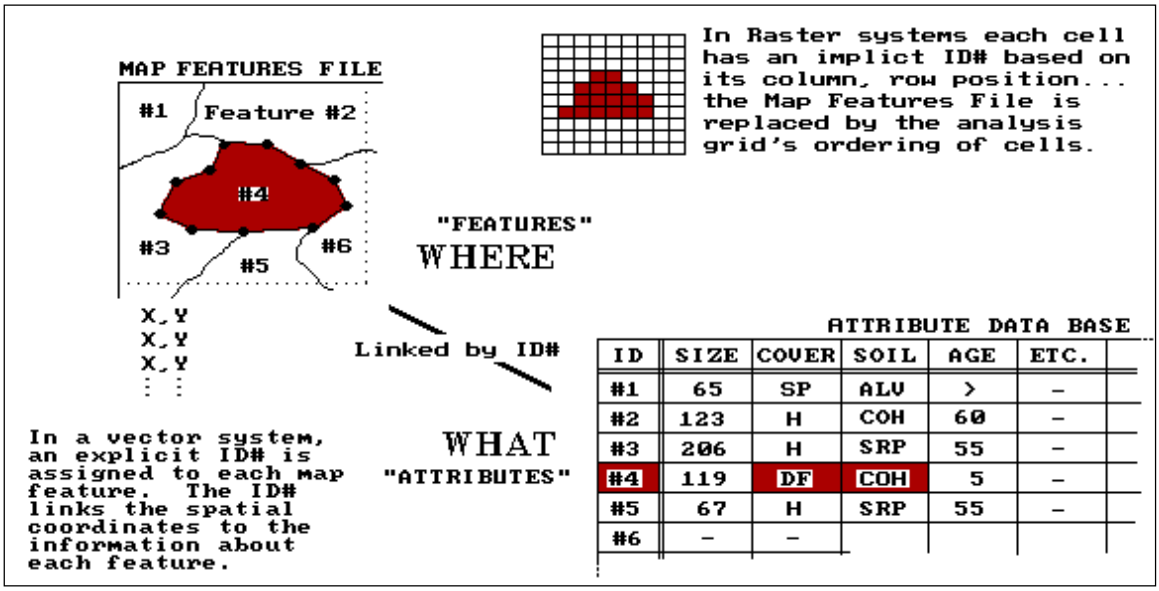
- Vector data may or may not be topologically explicit, depending on the file's data structure.
- Care should be taken to determine whether the raster or vector data model is best suited for your data and/or analytical needs.

The following **Fig. 6** let you understand in a most easy way about the Vector data and Raster data and how the real world looks like.



2) Non-spatial data:

Non-spatial data (also called attribute or tabular data) describe the characteristics of features associated with vector data **Fig. 7.** It is stored in database file (.dbf) and usually managed by Database management systems (DBMS) in GIS environment. Unique identification number used by database to link the non-spatial data with spatial data.



Basic linkages between a vector spatial data (topologic model) and attributes maintained in a relational database file (From Berry)

Frequently Asked Questions (FAQs):

1. **Define Geographic Information System (GIS) data structure?**

A Geographic Information System (GIS) data structure is a mathematical construct for depicted geographic feature as data. The term GIS that can refer to many methods, processes and technologies. It is an associated with several applications related to environment, planning, transportation and many location-enabled services that can be analyse and visualize. Location as the key index variable is used in GIS that can be relate unrelated information. Location and position of features on earth surface has been recorded as x, y and z coordinates. Many datasets are being used in GIS environment related to research, planning and utility projects etc. GIS data are generated from maps, aerial photographs, satellite imagery and field survey by scanning and digitizing using GIS software. GIS data represented in the form of spatial data (location and geometry) and tabular data (characteristics of feature).

2. What is Spatial data?

The sources of spatial data are surveying and remote sensing. Aerial photographs and satellite images are the sources of real time geographic information. Spatial database use other techniques that is different from table to store spatial features. The usage of spatial data involves various discipline and the discipline that deals all aspect of spatial data handling is called Geoinformatics. Spatial data is also called geographic data that identified by geometry, geographic location and attribute that is described its characteristic, such as forest, ocean, town and others. The location and geometry of geographic features are stored in the form of coordinates (Latitude and Longitude) and topology. Spatial data manipulation or analyzed done with the help of attribute data in GIS environment that can be mapped.

3. What is non-spatial data?

Non-spatial data (also called attribute or tabular data) describe the characteristics of features associated with vector data. It is stored in database file (.dbf) and usually managed by Database management systems (DBMS) in GIS environment. Unique identification number used by database to link the non-spatial data with spatial data.

4. Define Raster data?

Raster data presented as a matrix or array of pixels (Picture Element). Pixel is the smallest unit of picture. It is stored spatial information in grid cells organized and accessed as rows and columns. The information of geographic features stored within the cells contain a number called Digital Number (DN) value. DN value of geographic features are dependent on feature's characteristics and reflectance value over the surface of the earth. Each geographic feature has own DN value arranged in regular grid. The spatial resolution of raster data is dependent on cell or pixel size. Spatial resolution of raster data increased with decreasing pixel size. Spatial details of raster data controlled by resolution. There are many sources of raster data such as

satellite imagery, aerial photographs, scanned maps and other sources. Raster data is divided into following two categories.

5. What is Vector data?

Vector data is used to present spatial information in point, line and polygon (area). Point data recorded in the form of (X,Y) coordinates. Line and polygon data is based on **Arc-node** having non-intersecting lines segment called arcs and connecting set of arc forms area objects. The main merits of vector data has require minimum memory space, area and perimeter of polygon feature is estimated accurately, data handling/manipulation is fast and produce accurate results. Vector data used to captured the geographic location of discrete features such as roads, buildings, river, hospital and boundaries of other geographic features etc. The precise geo-location of features on earth surface are recorded in the form of (X, Y) in vector data.

Multiple Choice Questions (Quiz)

1. GIS data represented in the form of

(i) Spatial data
(ii) Non-spatial data
(iii) Both (i) and (ii)
(iv) None of the above

2. Which of the following is not a spatial data model?

(i) Raster data model
(ii) Vector data model
(iii) Attribute data
(iv) None of the above

3. Which data describe the characteristics of features associated with vector data?

(i) Spatial data
(ii) Non- spatial data
(iii) Raster data
(iv) Vector data

4. The Vector data model is based on which of the following?

(i) Pixels or grid cells
(ii) Collections of points joined by straight lines
(iii) Cartesian coordinate system
(iv) None of the above

5. The Raster data model is based on which of the following?

(i) Grid cells or pixels grouped to form spatial entities
(ii) Discrete XY coordinate pairs
(iii) Grid cells
(iv) Tesselations

Suggested Readings:

1. Lo, Char P., & Yeung, Albert K. W. (2006). Concepts and Techniques of Geographic Information System, 2nd Edn. Pearson Education. ISBN: 013149502X, 978-0131495029.
2. Heywood, I., Cornelius, S. & Carver, S. (2011). An Introduction to Geographical Information Systems, 4th Edn. Prentice Hall. ISBN: 027372259X, 978-0273722595.
3. Burrough, Peter A., & McDonnell, Rachael, A. (1998). Principles of Geographical Information Systems, 2nd Edn. OUP Oxford. ISBN: 0198233655, 978-0198233657.
4. Sahu, Kali C. (2007). Textbook of Remote Sensing and Geographical Information. Atlantic Publications, New Delhi. ISBN: 8126909099, 978-8126909094.
5. Bernhardsen, T. (2002). Geographic Information System: An Introduction, 3rd Edn. John Wiley & Sons, New York. ISBN: 0471419680, 978-0471419686.
6. Chang, Kang-tsung (2017), Introduction to Geographical Information Systems, 4th Edn., McGraw Hill Education, India. ISBN: 0070658986, 978-0070658981.