

**Q.1 Attempt any TWO question of the following :** [10]

**Q.1(a) Explain the applications of GIS related to Crime Management.** [5]

(A) In more recent years GIS has been used for crime analysis, emergency planning, land records management, market analysis, and transportation applications. Here are some examples on the Internet:

- The Department of Homeland Security's National Incident Management System (NIMS) identifies GIS as a supporting technology for managing domestic incidents (<http://www.dhs.gov/>).
- The National Institute of Justice uses GIS to map crime records and to analyze their spatial patterns by location and time (<http://www.ojp.usdoj.gov/nij/maps/>).
- The Federal Emergency Management Agency links a flood insurance rate map database to physical features in a GIS database ([http://www.fema.gov/plan/prevent/fhm/mm\\_main.shtm](http://www.fema.gov/plan/prevent/fhm/mm_main.shtm)).
- Larimer County, Colorado allows public access to the county's land records in a GIS database (<http://www.larimer.org/>).

**Q.1(b) Explain State Plane Co-ordinate system with neat diagram.** [5]

(A) The SPC system was developed in the 1930s to permanently record original land survey monument locations in the United States. To maintain the required accuracy of one part in 10,000 or less, a state may have two or more SPC zones. As examples, Oregon has the North and South SPC zones and Idaho has the West, Central and East SPC zones (Figure 1). Each SPC zone is mapped onto a map projection. Zones that are elongated in the north-south direction (e.g., Idaho's SPC zones) use the transverse Mercator and zones that are elongated in the east-west direction (e.g., Oregon's SPC zones) use the Lambert conformal conic. (The only exception is zone 1 of Alaska, which uses the oblique Mercator to cover the panhandle of Alaska,) Point locations within each SPC zone are measured from a false origin located to the southwest of the zone.

Because of the switch from NAD27 to NAD83, there are SPC27 and SPC83. Besides the change of the datum, SPC83 has a few other changes. SPC83 coordinates are published in meters instead of feet. The states of Montana, Nebraska, and South Carolina have each replaced multiple zones with a single SPC zone. California has reduced SPC zones from seven to six. And Michigan has changed from transverse Mercator to Lambert conformal conic projections.

Some states in the United States have developed their own statewide coordinate system. Montana, Nebraska, and South Carolina all have a single SPG zone, which can serve as the statewide coordinate system. Idaho is another example. Idaho is divided into two UTM zones (11 and 12) and three SPC zones (West, Central, and East). These zones work well as long as the study area is within a single zone. When a study area covers two or more zones, the data sets must be converted to a single zone for spatial registration. But the conversion to a single zone also means that the data sets can no longer maintain the accuracy level designed for the UTM or the SPC coordinate system. The Idaho statewide coordinate system, adopted in 1994 and modified in 2003, is still based on a transverse Mercator projection but its central meridian passes through the center of the state (114° W). (A complete list of parameters of the Idaho statewide coordinate system is included in Task 1 of the applications section.) Changing the location of the central meridian means one zone for the entire state.



**Fig. 1** : SPC83 zones in the conterminous United States. The thinner lines are county boundaries, and the bold lines are state boundaries. This map corresponds to the SPC83 table on the inside of this book's back cover.

**Q.1(c) Explain the various map projection types with suitable diagram. [5]**

**(A)** The process of projection transforms the spherical Earth's surface to a plane. The outcome of this transformation process is a map projection: a systematic arrangement of parallels and meridians on a plane surface representing the geographic coordinate system.

**Types of Map Projections :**

- Cartographers group map projections by the preserved property into the following four classes :
  - (1) Conformal projection preserves local angles and shapes.
  - (2) An equivalent projection represents areas in correct relative size.
  - (3) An equidistant projection maintains consistency of scale along certain lines.
  - (4) An azimuthal projection retains certain accurate directions.
- The conformal and equivalent properties are mutually exclusive. Otherwise a map projection can have more than one preserved property.
- The conformal and equivalent properties are global properties, meaning that they apply to the entire map projection.
- The equidistant and azimuthal properties are local and may be true only from or to the center of the map projection.

Map projections can be grouped by their projection surfaces into cylindrical, conic and azimuthal.

- (1) Cylindrical projection - if a map can be constructed using a cylinder.
- (2) Conic projection - if a map can be constructed using a cone.
- (3) Azimuthal projection - if a map can be constructed using a plane.

**Q.1(d) Convert the following into degrees : [5]**

**(A)** (i)  $45^{\circ} 15' 45''$  (ii) 1745 rad

$$= 45 + \frac{15}{60} + \frac{45}{60 + 60} = 45 + 0.25 + 0.0125 = 45.2625 \text{ Degree}$$

(ii) 1745 radians  
1 radians = 57.2958  
1745 radians =  $1745 \times 57.2958$   
99981.171°

**Q.2 Attempt any TWO question of the following :** [10]

**Q.2(a) What is metadata? What constitutes metadata in GIS?** [5]

**(A)** Metadata provide information about geospatial data (Guptill 1999). They are therefore an integral part of GIS data and are usually prepared and entered during the data production process. Metadata are important to anyone who plans to use public data for a GIS project (Comber et al. 2005). First, metadata let us know if the data meet our specific needs for area coverage, data quality, and data currency. Second, metadata show us how to transfer, process, and interpret geospatial data. Third, metadata include the contact for additional information.

The FGDC has developed the content standards for metadata and provides detailed information about the standards at its website (<http://www.fgdc.gov/>). These standards have been adopted by federal agencies in developing their public data. FGDC metadata standards describe a data set based on the following categories:

- Identification information – basic information about the data set, including title, geographic data covered, and currency.
- Data quality information – information about the quality of the data set, including positional and attribute accuracy, completeness, consistency, sources of information, and methods used to produce the data.
- Spatial data organization information – information about the data representation in the data set, such as method for data representation (e.g., raster or vector) and number of spatial objects.
- Spatial reference information – description of the reference frame for and means of encoding coordinates in the data set, such as the parameters for map projections or coordinate systems, horizontal and vertical datums, and the coordinate system resolution.
- Entity and attribute information – information about the content of the data set, such as the entity types and their attributes and the domains from which attribute values may be assigned.
- Distribution information – information about obtaining the data set.
- Metadata reference information – information on the currency of the metadata information and the responsible party.

**Q.2(b) List the various sources of obtaining existing GIS Data. Explain any two.** [5]

**(A) Creating geospatial Data:**

Address geocoding, also called address matching, can create point features from street addresses. Street addresses are therefore an important data source for creating new data.

**Various data sources are :**

- |                                       |                                   |
|---------------------------------------|-----------------------------------|
| • Remotely sensed data                | • Field data                      |
| • Global positioning system           | • Text Files with x-y-Coordinates |
| • Digitizing Using a Digitizing Table | • Scanning                        |
| • On Screen digitizing                |                                   |

**(explain any one from the following)**

• **Remotely Sensed Data :**

- Satellite images can be digitally processed to produce a wide variety of thematic data for a GIS project. Land use/land cover data such as USGS National Land Cover Data are typically derived from satellite images. Other types of data include vegetation types, crop health, eroded soils, geologic features, the composition and depth of water bodies, and even snowpack. Satellite images provide timely data and, if collected at regular intervals, they can also provide temporal data that are valuable for recording and monitoring changes in the terrestrial and aquatic environments.

- Some GIS users fell in the past that satellite images did not have sufficient resolution, or were not accurate enough, for their projects. This is no longer the case with high-resolution satellite images. Ikonos and QuickBird images can now be used to extract detailed features such as roads, trails, buildings, trees, riparian zones, and impervious surfaces.
- **Field Data :**
  - Two important types of field data are survey data and global positioning system (GPS) data.
    - **Survey data -**
    - Survey data consist primarily of distances, directions, and elevations.
    - Distances can be measured in feet or meters using a tape or an electronic distance measurement instrument.
    - The direction of a line can be measured in azimuth or bearing using a transit, theodolite, or total station.
    - An azimuth is an angle measured clockwise from the north end of a meridian to the line. Azimuths range in magnitude from 0° to 360°. A bearing is an acute angle between the line and a meridian. The bearing angle always has the accompanied letters that locate the quadrant (i.e., NE, SE, SW, or NW) in which the line falls.
    - In the United States, most legal plans use bearing directions. An elevation difference between two points can be measured in feet or meters using levels and rods.
    - In GIS, field survey typically provides data for determining parcel boundaries. An angle and a distance can define a parcel boundary between two stations (points). For example, the description of N45°30'W 500 feet means that the course (line) connecting the two stations has a bearing angle of 45 degrees 30 minutes in the NW quadrant and a distance of 500 feet.
- **Global positioning System :**
  - Using GPS satellite in space as reference points, a GPS receiver can determine its precise position on the Earth's surface.
  - GPS data include the horizontal location based on geographic or projected coordinate system and the elevation of the point location.
  - A collection of GPS positions along a line can determine a line feature, and a series of lines measured by GPS can determine an area feature.
  - This is why GPS has become a useful tool for collecting geospatial data such as road networks, and for tracking point objects such as vehicles and people.
  - The GPS receiver measures its distance (range) from a satellite using travel time and speed of signals it receives from the satellite. With three satellites simultaneously available, the receiver can determine its position in space (x, y, z) relative to the center mass of the Earth. But to correct timing errors, a fourth satellite is required to get precise positioning . the receiver's position in space can then be converted to latitude, longitude and height based on a reference spheroid.
- **Text Files with x-,y-Coordinates :**

Geospatial data can be generated from a text file that contains x-, y-coordinates. The x-, y-coordinates can be geographic (in decimal degrees) or projected. Each pair of x-, y-coordinates creates a point. Therefore, we can create spatial data from a file that records the locations of weather stations, epicenters, or a hurricane track.
- **Digitizing Using a Digitizing Table :**

Digitizing is the process of converting data from analog to digital format. Manual digitizing uses a digitizing table (Figure 1). A digitizing table has a built-in electronic mesh, which can sense the position of the cursor. To transmit the x-, y-coordinates of a

point to the connected computer, the operator simply clicks on a button on the cursor after lining up the cursor's cross hair with the point. Large-size digitizing tables typically have an absolute accuracy of 0.001 inch (0.003 centimeter).

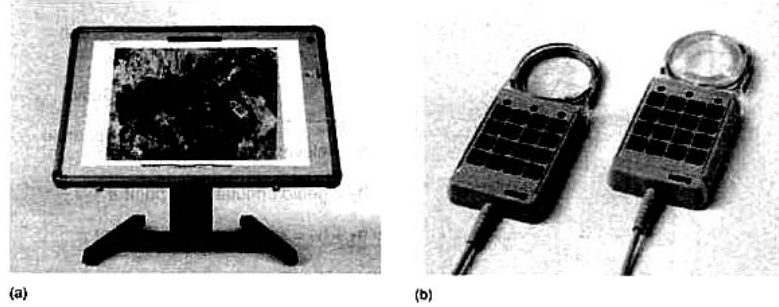


Fig. 1 : a. A large digitizing table    b. a cursor with a 16-button keypad

### Scanning

- Scanning is a digitizing method that converts an analog map into a scanned file, which is then converted back to vector format through tracing.
- A scanner converts an analog map into a scanned image file in raster format. The simplest type of map to be scanned is a black-and-white map: black lines represent map features, and white areas represent the background. The map may be a paper or Mylar map and it may be inked or penciled.
- Scanning converts the map into a binary scanned file in raster format; each pixel has a value of either 1 (map feature) or 0 (background). Map features are shown as raster lines, a series of connected pixels on the scanned file (Figure 2). The pixel size depends on the scanning resolution, which is often set at 300 dots per inch (dpi) or 400 dpi for digitizing. A raster line representing a thin inked line on the source map may have a width of 5 to 7 pixels (Figure 2).

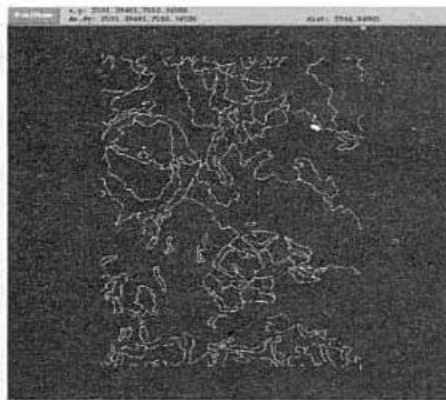


Fig. 2 : A binary scanned file: The lines are soil lines, and the black areas are the background.

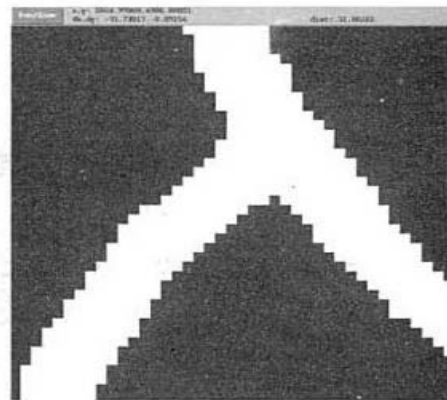


Fig. 3 : A raster line in a scanned file has a width of several pixels.

### On-Screen Digitizing :

On-screen digitizing, also called heads-up digitizing, is manual digitizing on the computer monitor using a data source such as a DOQ as the background. DOQs combine the image characteristics of a photograph with the geometric qualities of a map. Easily integrated in a GIS, DOQs are the ideal background for digitizing. On-screen digitizing is an efficient method for editing or updating an existing layer such as adding new trails or roads that are not on an existing layer but are on a new DOQ. Likewise, we can use the method to update new clear-cuts or burned areas in a vegetation layer.

Q.2(c) Explain the concept of 'Resampling of Pixel Values'.

[5]

(A) Resampling of Pixel Values

- The result of geometric transformation of a satellite image is a new image based on a projected coordinate system. But the new image has no pixel values. The pixel values must be filled through resampling.
- Resampling in the process of filling each pixel of the new image with a value or a derived value from the original image.

**Resampling Methods**

Three common resampling methods are - nearest neighbor, bilinear interpolation, and cubic convolution.

- **The nearest neighbor** - this resampling method fills each pixel of the new image with the nearest pixel value from the original image. For example, Figure 1 shows that pixel A in the new image will take the value of pixel a in the original image because it is the closest neighbor. The nearest neighbor method is computationally efficient. The method has the additional property of preserving the original pixel values, which is important for categorical data such as land cover types and desirable for some image processing such as edge detection.
- **Bilinear interpolation** - a resampling method that uses the distance-weighted average of the four nearest pixel values to estimate a new pixel value. Both bilinear interpolation and cubic convolution fill the new image with distance-weighted averages of die pixel values from the original image.

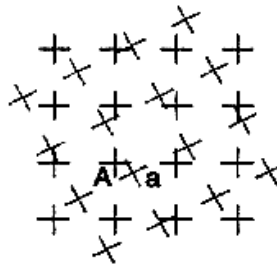


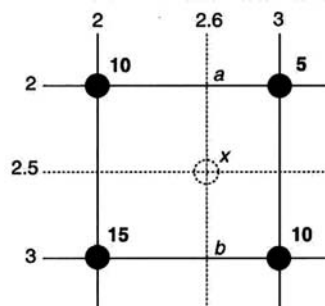
Fig. : Because a in the original image is closest to pixel A in the new image, the pixel value at a is assigned to be the pixel value at A using the nearest neighbor technique.

The bilinear interpolation method uses the average of the four nearest pixel values from three linear interpolations, whereas the cubic convolution method uses the average of the 16 nearest pixel values from five cubic polynomial interpolations.

- **Cubic convolution** tends to produce a smoother output man bilinear interpolation but requires a longer processing time.

Q.2(d) Perform bi-linear interpolation to find the value of x in the diagram below:

[5]



(A)  $a = 0.6(5) + 0.4(10) = 7$   
 $b = 0.6(10) + 0.4(15) = 12$

Then we perform the third linear interpolation between a and b to derive the interpolated value at x :

$x = 0.5(7) + 0.5(12) = 9.5$

**Q.3 Attempt any TWO question of the following :** [10]

**Q.3(a) Explain the concept of 'Layout' and 'Visual Hierarchy'.** [5]

**(A) Layout**

Layout, or planar organization, deals with the arrangement and composition of various map elements. Major concerns with layout are focus, order, and balance. A thematic map should have a clear focus, which is usually the map body or a part of the map body. To draw the map reader's attention, the focal element should be placed near the optical center, just above the map's geometric center. The focal element should be differentiated from other map element by contrast in line width, texture, value detail, and color.

After viewing the focal element, the reader should be directed to the rest of the map in an ordered fashion. For example, the legend and the title are probably the next elements that the viewer needs to look at after the map body. To smooth the transition, the mapmaker should clearly place the legend and the title on the map, with perhaps a box around the legend and a larger type size for the title to draw attention to them.

A finished map should look balanced. It should not give the map reader an impression that the map "looks" heavier on the top, bottom, or side. But balance does not suggest the breaking down of the map elements and placing them, almost mechanically, in every part of the map. Although in that case the elements would be in balance, the map would be disorganized and confusing. Mapmakers therefore should deal with balance within the context of organization and map communication.

Cartographers used to thumbnail sketches to experiment with balance on a map. Now they use computers to manipulate map elements on a layout page. ArcGIS, for example, offers two basic methods for layout design. The first method is to use a layout template. These templates are grouped as general, industry, USA, and world. Each group has a list of choices. For example, the layout templates for the United States include USA, conterminous USA, and five different regions of the country. The idea of using a layout template is to use a built-in design option to quickly compose a map.

#### **Visual Hierarchy**

- (i) Visual Hierarchy is the process of developing a visual plan to introduce the 3D effect or depth to maps.
- (ii) Map Maker create the visual hierarchy by placing map elements at different visual levels according to their importance to the maps purpose.
- (iii) The most important element should be at the top of the hierarchy and should appear closest to the map reader.
- (iv) The least important element should be at the bottom
- (v) The concept of visual hierarchy is an extension of figure-ground relationship in visual perception
- (vi) The most important and effective principle in creating a visual hierarchy is interposition or superimposition.
- (vii) Interposition uses the incomplete outline of an object to make it appear as through it is behind another.  
Eg. Magazine, Newspaper

**Q.3(b) Explain various types of relationships in database tables.** [5]

**(A)** A relational database may contain four types of relationships (also called cardinalities) between tables, or more precisely, between records in tables : one-to-one, one-to-many, many-to-one, and many-to-many (Figure 1). The one-to-one relationship means that one and only one record in a table is related to one and only record in another table. The one-to-many relationship means that one record in a table may be related to many records in another table. For example, the street address of an apartment complex may include several households. The many-to-one relationship means that many records in a table may

be related to one record in another table. For example, several households may share the same street address. The many-to-many relationship means that many records in a table may be related to many records in another table. For example, a timber stand can grow more than one species and a species can grow in more than one stand.

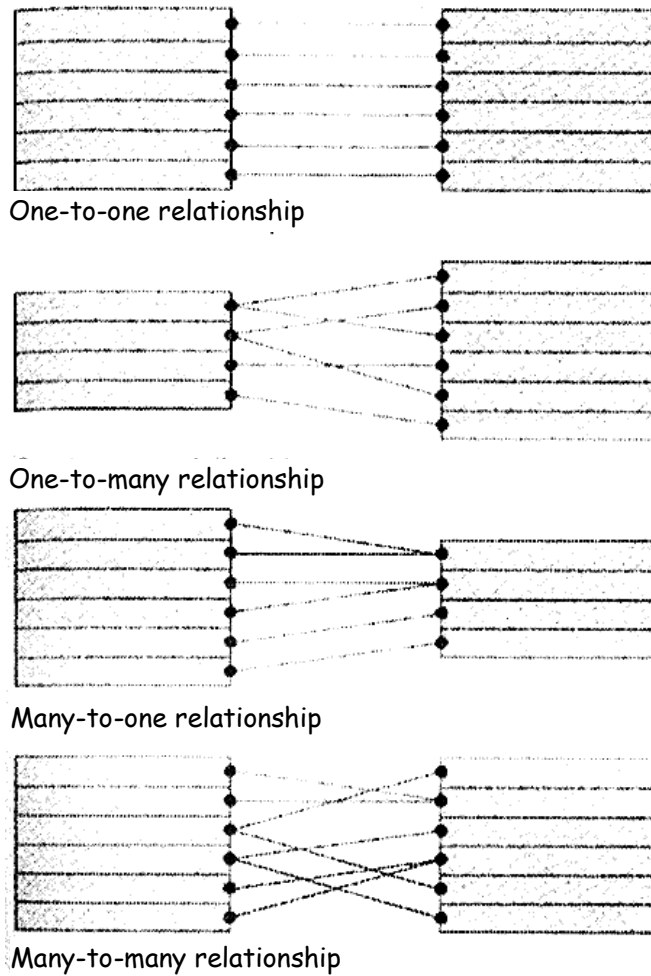


Fig. 1 : Four types of data relationships between tables; one-to-one, one-to-many, many-to-one, and many-to-many.

Q.3(c) Explain any two types of Database Designs with suitable diagrams. [5]

(A) Different types of database design are

- 1) Flat files
- 2) Hierarchical
- 3) Network
- 4) Relational

1) Flat files

It contains all data in large file. A feature attribute table is like a flat file. Another example is spreadsheet with large data.

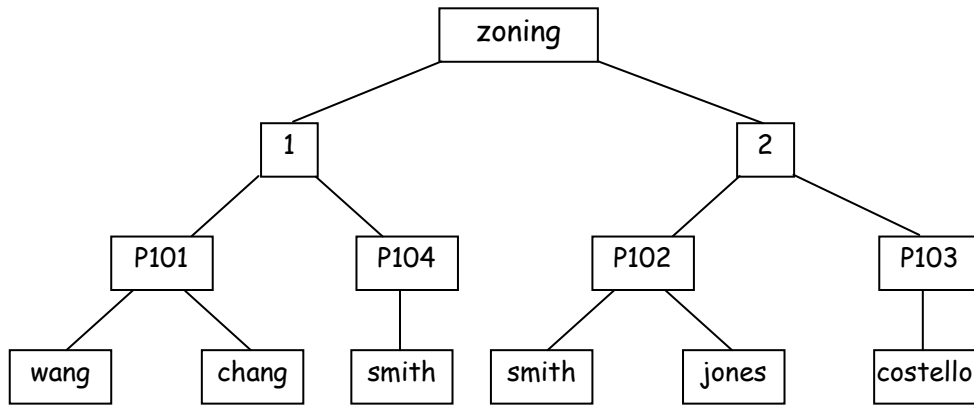
Example :

PIN	Owner	Zoning
P101	Wang	Residential (1)
P101	Chang	Residential (1)
P102	Smith	Commercial (2)
P102	Jones	Commercial (2)
P103	Costello	Commercial (2)
P104	Smith	Residential (1)

2) Hierarchical database :

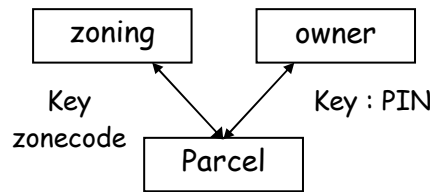
A hierarchical database organizes its data at different levels and uses only the one to many association between levels. The below example shows the hierarchical levels of zoning, parcel and owner based on the one-to-many associate each level is divided into different branches.





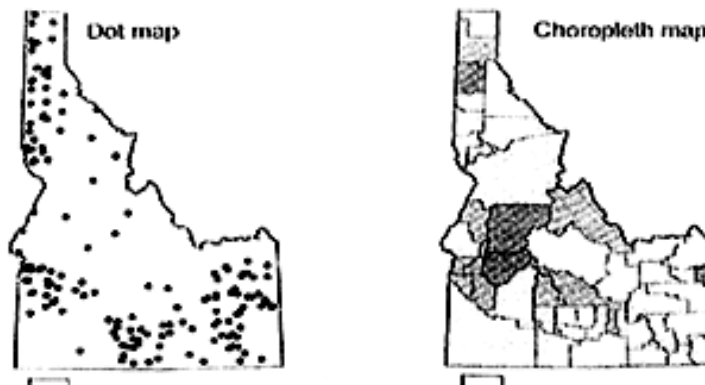
**3) Relational Database**

It is a collection of tables also called relations, which can be connected to each other by keys.



**Q.3(d) Explain : (i) Dot map, (ii) Choropleth map [5]**

- (A)
- The dot map uses uniform point symbols to show spatial data, with each symbol representing a unit value. One-to-one dot mapping uses the unit value of one, such as one dot representing one crime location. But in most cases, it is one-to-many dot mapping and the unit value is greater than one. The placement of dots becomes a major consideration in one-to-many dot mapping.
  - The choropleth map symbolizes, with shading, derived data based on administrative units. An example is a map showing average household income by county. The derived data are usually classified prior to mapping and are symbolized using a color scheme for quantitative data.



**Q.4 Attempt any TWO question of the following : [10]**

**Q.4(a) Explain the various 'Descriptive Statistics' used in GIS. [5]**

- (A)
- Similar to statistics, data exploration in GIS lets the user view the general patterns in the data set, query data subsets, and hypothesize about possible relationships between data sets (Anselin 1999; Andrienko et al. 2001; Haining 2003). But there are two important differences. First, data exploration in a GIS involves both spatial and attribute data. Second, the media for data exploration in GIS include maps and map features. For example, in studying soil conditions, we want to know not only how much of the study area is rated poor but also where those poor soils are distributed on a map. Therefore, besides descriptive statistics and graphics, data exploration in GIS must also cover attribute data query, spatial data query, and map manipulation.

**Descriptive Statistics :** Descriptive statistics summarize the values of a data set. They include the following :

- Range: the difference between the minimum and maximum values.
- Median: the midpoint value, or the 50<sup>th</sup> percentile.
- First quartile: the 25<sup>th</sup> percentile.
- Third quartile: the 75<sup>th</sup> percentile.
- Mean: the average of data values. The mean can be calculated by  $\sum_{i=1}^n x_i / n$ , where  $x_i$  is the  $i^{\text{th}}$  value and  $n$  is the number of values.
- Variance: the average of the squared deviations of each data value about the mean. The variance can be calculated by  $\sum_{i=1}^n (x_i - \text{mean})^2 / n$
- Standard deviation: the square root of the variance.
- Z score: a standardized score that can be computed by  $(x - \text{mean})/s$ , where  $s$  is the standard deviation.

**Q.4(b) Explain Raster Data Query with example.**

**[5]**

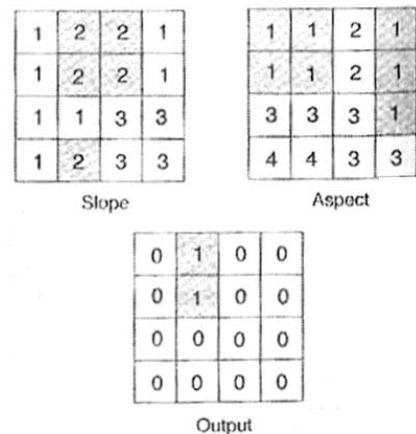
**(A) Raster Data Query**

Although the concept and even some methods for data query are basically the same for both raster data and vector data, there are enough practical differences to warrant a separate section on raster data query.

**Query by Cell Value**

The cell value in a raster typically represents a specific attribute value (e.g., land-use type, elevation value, etc.) at the cell location. Therefore, the operand in raster data query is the raster itself rather than a field as in the case of vector data query. This is illustrated in the following examples.

A raster data query uses a Boolean statement to separate cells that satisfy the query statement from cells that do not. The expression, [road] = 1, queries a road raster that has the cell value of 1. The Operand [road] refers to the raster and the operand 1 refers to a cell value, which may represent the interstate category. This next expression, [elevation] > 1243.26, queries a floating-point elevation raster that has the cell value greater than 1243.26. Again the operand [elevation] refers to the raster itself. Because a floating-point elevation raster contains continuous values, querying a specific value is not likely to find any cell in the raster.



**Fig. 1:** Raster data query involving two rasters: slope = 2 and aspect = 1. Selected cells are coded 1 and others 0 in the output raster.

Raster data query can also use the Boolean Connectors of AND, OR, and NOT to string together separate expressions. A compound statement with separate expressions usually applies to multiple rasters, which may be integer, or floating point, or a mix of both types. For example, the statement, ([slope] = 2) AND ([aspect] = 1), selects cells that have the value of 2 (e.g., 10-20% slope) in the slope raster, and 1 (e.g., north aspect) in the aspect raster (Figure 1). Those cells that satisfy the statement have cell value of 1 on the output, while other cells value of 0.

Querying multiple rasters directly is unique to raster data. For vector data, all attributes to be used in a compound expression must be from a table or a joint attribute table. Another difference is that a GIS package such as ArcGIS has dialogs specifically designed for vector data query but does not have them for raster data query. The user interface for raster data query is often mixed with raster data analysis.

**Q.4(c) What are the different types of graphs used for data exploration? [5]**

(A) Different types of graphs are used for data exploration.

- A line graph displays data as a line. The line graph example in figure 1 shows the rate of population change in the United States along the y-axis and the state along the x-axis. Notice a couple of "peaks" in the line graph.

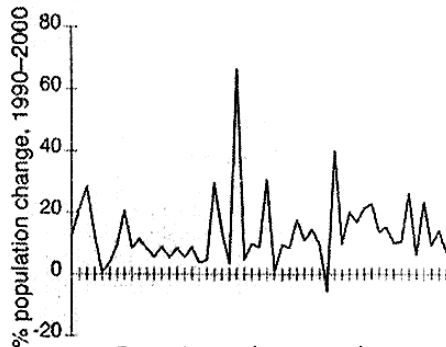


Fig. 1 : A line graph

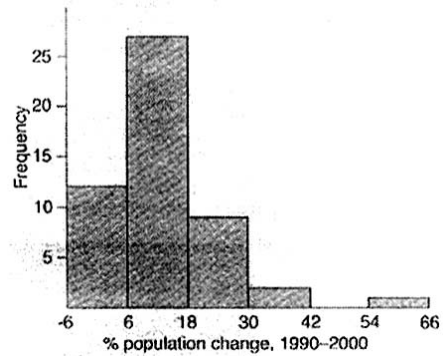


Fig. 2 : A histogram (bar chart)

- A bar chart, also called a histogram, groups data into equal intervals and uses bars to show the number of frequency of values falling within each class. A bar chart may have vertical bars or horizontal bars. Figure 2 uses a vertical bar chart to group rates of population change in the United States into six classes. Notice one bar at the high end of the histogram.

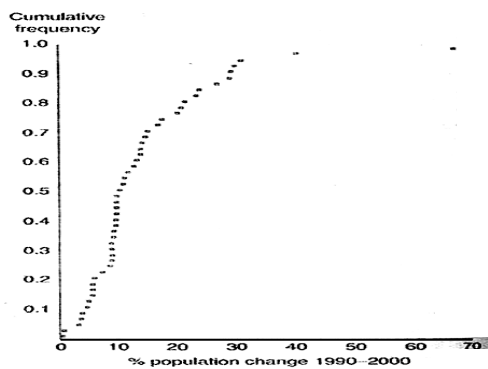


Fig. 3 : A cumulative distribution graph

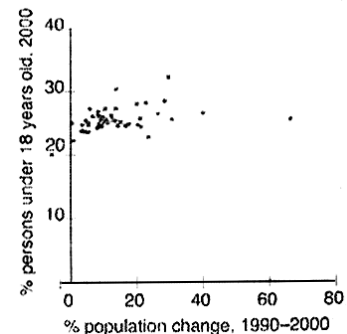
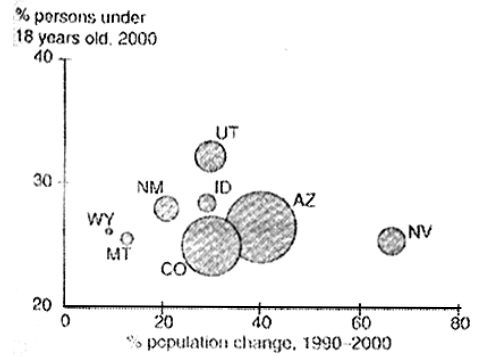


Fig. 4 : A scatterplot plotting percent persons 18 years old in 2000 against percent population change, 1990-2000. A weak-positive relationship is present between the two variables.

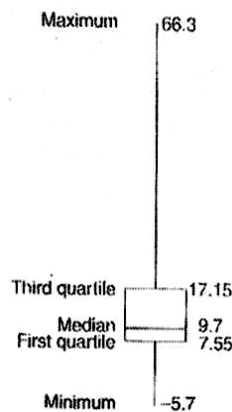
- A cumulative distribution graph is one type of line graph that plots the ordered data values against the cumulative distribution values. The cumulative distribution value of the  $i$ th ordered value is typically calculated as  $(i - 0.5)/n$ , where  $n$  is the number of values. This computational formula converts the values of a data set to within the range of 0.0 to 1.0. Figure 3 show a cumulative distribution graph.
- A scatterplot uses markings to plot the values of two variables along the x- and y-axes. Figure 4 plots percent population change 1990–2000 against percent persons under 18 years old in 2000 by state in the United States. The scatterplot suggests a weak positive relationship between the two variables.

- **Bubble plots** are a variation of scatterplots. Instead of using constant symbols as in a scatterplot, a bubble plot has varying sized bubbles that are made proportional to the value of a third variable. Figure 5 is a variation of Figure 4, the additional variable shows by the bubble size is the state population ins 2000. As an illustration, Figure 5 only shown states in the Mountain region, one of the nine regions defined by the U.S. Census Bureau.
- **Boxplots**, also called the "box and whisker" plots, summarized the distribution of five statistics from a data set the minimum, first quartile, median, third quartile, and maximum. By examining the position of the statistics in a boxplot, we can tell if the distribution of data values is symmetric or skewed and if there are unusually data points (i.e. outliers).
- **Quantile-quantile plots**, also called QQ plots, compare the cumulative distribution of a data set with that of some theoretical distribution such as the normal distribution, a bell-shaped frequency distribution.

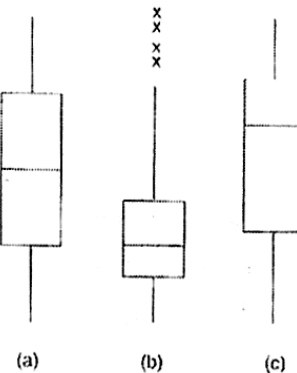


**Fig. 5 :** A bubble plot showing percent population change 1990-2000, percent persons under 18 years old in 2000, and state population in 2000.

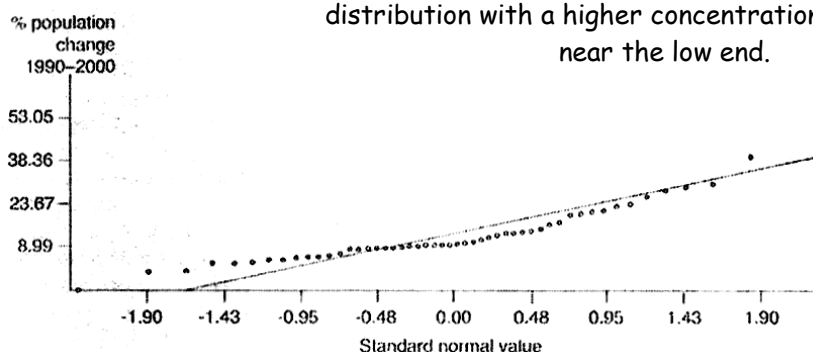
Figure 6 shows a boxplot based on the rate of population change in the United States. This data set is clearly skewed toward the higher end. Figure 7 summarizes three basic types of data sets in terms of distribution of data values. Boxplots are therefore useful for comparisons between different data sets.



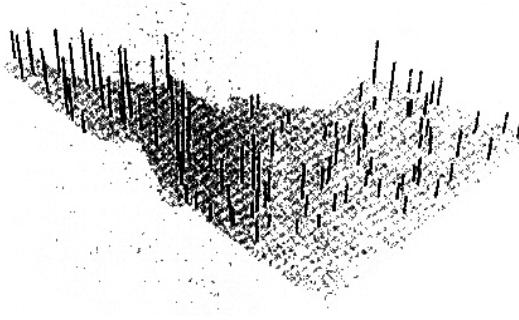
**Fig. 6 :** A boxplot based on the percent population change 1990-2000 data



**Fig. 7 :** Boxplot (a) suggests that the data values follow a normal distribution. Boxplot (b) shows a positively skewed distribution with a higher concentration of data values near the high end. The x's in (b) may represent outliers, which are more than 1.5 box lengths from the end of the box. Boxplot (c) shows a negatively skewed distribution with a higher concentration of data values near the low end.



**Fig. 8 :** A QQ plot plotting percent population change, 1990-2000 against the standardized value from a normal distribution.



**Fig.9** : A 3-D plot showing annual precipitation at 105 weather stations in Idaho. A north-to-south decreasing trend is apparent in the plot.

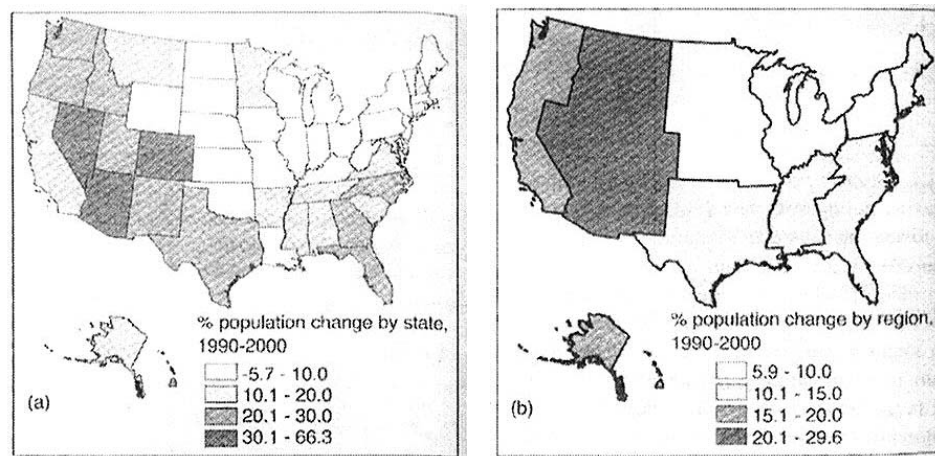
Some graphs are designed for spatial data. Figure 9, for example, shows a plot of spatial data values by raising a bar at each point location so that the height of the bar is proportionate to its value. This kind of plot allows the user to see the general trends among the data values in both the x-dimension (east-west) and y-dimension (north-south).

**Q.4(d) Explain the terms 'Spatial Aggregation' and 'Map Comparison'.**

**[5]**

**(A) Spatial Aggregation**

Spatial aggregation is functionally similar to data classification except that it groups data spatially. Figure 2, shows percent population change in the United States by state and by region. Used by the U.S. Census Bureau for data collection, regions are spatial aggregates of states. As shown in 16(b), a map by region gives more general view of population growth in the country than a map by state does. Other geographic levels used by the U.S. Census Bureau are county, census tract, block group, and block. Because these levels of geographic units form a hierarchical order, we can explore the effect of spatial scaling by examining data at different spatial scales.



**Fig. 2** : Two levels of spatial aggregation : by state (a), and by region

If distance is the primary factor in a study, we can aggregate spatial data by distance measures from points, lines, or areas. An urban area, for example, may be aggregated into distance zones away from the city center or from its streets (Batty and Xie 1994). Unlike the geography of census, these distance zones require additional data processing such as buffering and areal interpolation.

Spatial aggregation for raster data means aggregating cells of the input raster to produce a coarser-resolution raster. For example, we can aggregate cells of a raster by a factor of 3. Each cell in the output raster corresponds to a 3-by-3 matrix in the input raster, and the cell value is a computed statistic such as mean, median, minimum, maximum, or sum from the nine input cell values.

**Map Comparison**

Map comparison can help a GIS user sort out the relationship between different maps. For example, the display of wildlife locations on a vegetation layer may reveal the association between the wildlife species and the distribution of vegetation covers.

If the maps to be compared consist of only point or line features, they can be coded in different colors and superimposed on one another in a single view. But this process becomes difficult if they include polygon features or raster data. One option is to use transparency as a visual variable. A semitransparent layer allows another layer to show through. For example, to compare two raster layers, we can display one layer in a color scheme and the other in semitransparent shades of gray. The gray shades simply darken the color symbols and do not produce confusing color mixtures. Another example is to use transparency for displaying temporal changes such as land cover change between 1990 and 2000. Because one layer is semitransparent, we can follow the areal extent of a land cover type from both years. But it is difficult to apply transparency to more than two layers.

There are three other options for comparing polygon or raster layers. The first option is to place all polygon and raster layers, along with other point and line layers, onto the screen but to turn on and off polygon and raster layers so that only one of them is viewed at a time. Used by many websites for interactive mapping, this option is designed for casual users.

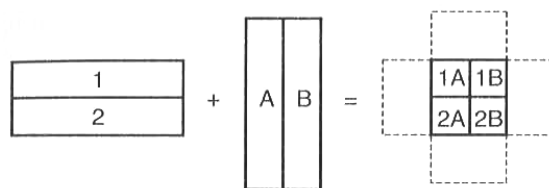
**Q.5 Attempt any TWO question of the following : [10]**

**Q.5(a) Explain the concept of Overlay with suitable diagram. [5]**

**(A) Overlay**

An overlay operation combines the geometries and attributes of two feature layers to create the output. The geometry of the output represents the geometric intersection of features from the input layers. Figure 1 illustrates an overlay operation with two polygon layers. Each feature on the output contains a combination of attributes from the input layers, and this combination differs from its neighbors.

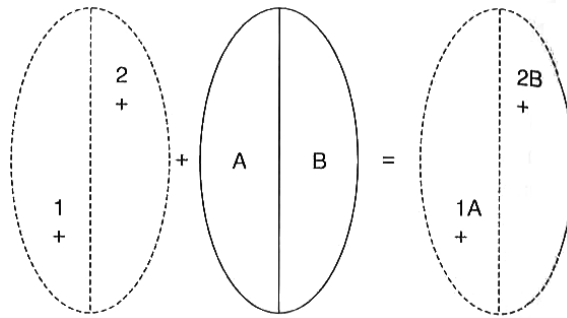
Feature layers to be overlaid must be spatially registered and based on the same coordinate system. In the case of the UTM (Universal Transverse Mercator) coordinate system or the SPC (State Plane Coordinate) system, the layers must also be in the same zone and have the same datum (e.g., NAD27 or NAD83).



**Fig. 1 :** Overlay combines the geometries and attributes from two layers into a single layer. The dashed lines are for illustration only and are not included in the output.

**Feature Type and Overlay**

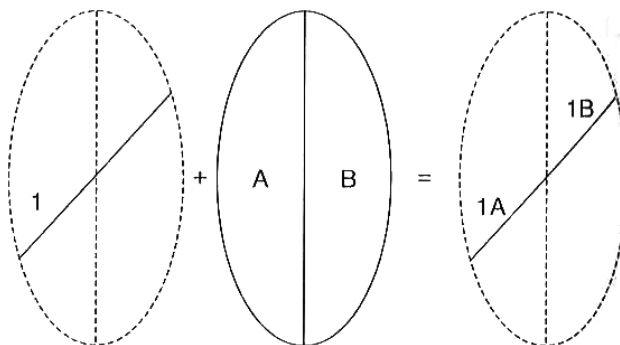
In practice, the first consideration for overlay is feature type. There are two groups of overlay operations. The first group uses two polygon layers as inputs. The second group uses one polygon layer and another layer, which may contain points or lines. Overlay operations can therefore be classified by feature type into point-in-polygon, line-in-polygon, and polygon-on-polygon. To distinguish the layers in the following discussion, the layer that may be a point, line, or polygon layer is called the input layer, and the layer that is a polygon layer is called the overlay layer.



**Fig. 2 :** Point-in-polygon overlay. The input is a point layer. The output is also a point layer but has attribute data from the polygon layer.

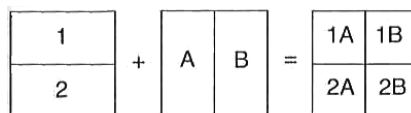
In a point-in-polygon overlay operation, the same point features in the input layer are included in the output but each point is assigned with attributes of the polygon within which it falls (Figure 2). For example, a point-in-polygon overlay can find the association between wildlife locations and vegetation types.

In a line-in-polygon overlay operation, the output contains the same line features as in the input layer but each line feature is dissected by the polygon boundaries on the overlay layer (Figure 3). Thus the output has more line segments than does the input layer. Each line segment on the output combines attributes from the input layer and the underlying polygon. For example, a line-in-polygon overlay can find soil data for a proposed road. The input layer includes the proposed road. The overlay layer contains soil polygons. And the output shows a dissected proposed road, each road segment having a different set of soil data from its adjacent segments.



**Fig. 3 :** Line-in-polygon overlay. The input is a line layer. The output is also a line layer. But the output differs from the input in two aspects: the line is broken into two segments, and the line segments have attribute data from the polygon layer.

The most common overlay operation is polygon-on-polygon, involving two polygon layers. The output combines the polygon boundaries from the input and overlay layers to create a new set of polygons (Figure 4). Each new polygon carries attributes from both layers, and these attributes differ from those of adjacent polygons.



**Fig. 4 :** Polygon-on-polygon overlay. In the illustration, the two layers for overlay have the same area extent. The output combines the geometries and attributes from the two layers into a single polygon layer.

For example, a polygon-on-polygon overlay can analyze the association between elevation zones and vegetation types.

**Q.5(b) Explain Silvers from overlay operation. [5]**

**(A) Silvers from overlay operation**

- A common error from overloading polygon layers is silvers, very small polygon along correlated or shared boundary lines of the input layer.
- The existence of silver often results from digitizing errors.
- Because of the high precision of manual digitizing or scanning, the shared boundaries on the input layers are rarely on the top of one another.
- When the layers are overlaid the digitized boundaries interest to form silvers.
- Other causes of silvers include errors in the source map or error in interpretation.
- Polygon boundaries on soil and vegetation maps are casually interpreted from field survey data, aerial photography and satellite images.

**Q.5(c) Explain the following map manipulation operations with example : [5]**

**(i) Dissolve (ii) Clip**

**(A) (i) Dissolve :** aggregates features that have the same attribute value or values (Figure 1). For example, we can aggregate roads by highway number or counties by state. An important application of Dissolve is to simplify a classified polygon layer. Classification groups values of a selected attribute into classes and makes obsolete boundaries of adjacent polygons, which have different values initially but are now grouped into the same class. Dissolve can remove these unnecessary boundaries and creates a new, simpler layer with the classification results as its attribute values. Another application is to aggregate both spatial and attribute data of the input layer. For instance, to dissolve a county layer, we can choose state name as the attribute to dissolve and county population to aggregate. The output is a state layer with an attribute showing the state population (i.e., the sum of county populations).

**(ii) Clip :** creates a new layer that includes only those features of the input layer that fall within the area extent of the clip layer (Figure 2). Clip is a useful tool, for example, for cutting a map acquired elsewhere to fit a study area. The input may be a point, line, or polygon layer, but the clip layer must be a polygon layer.

**Q.5(d) Explain Local Operations on Multiple raster inputs with suitable examples. [5]**

**(A) Local Operations with Multiple Rasters**

- Local operations with multiple rasters are also referred to as compositing, overlaying, or superimposing maps. Another common term for local operations with multiple input rasters is map algebra, a term that refers to algebraic operations with raster map layers. Because local operations can work with multiple rasters, they are the equivalent of vector-based overlay operations.
- A greater variety of local operations have multiple input rasters than have a single input raster. Besides mathematical functions that can be used on individual rasters, other measures that are based on the cell values or their frequencies in the input rasters can also be derived and stored on the output raster. Some of these measures are, however, limited to rasters with numeric data.

**Example :** Summary statistics, including maximum, minimum, range, sum, mean, median, and standard deviation, are measures that apply to rasters with numeric data. Figure 1, for example, shows a local operation that calculates the mean from three input rasters. If a cell contains no data in one of the input rasters, the cell also carries no data in the output raster by default.



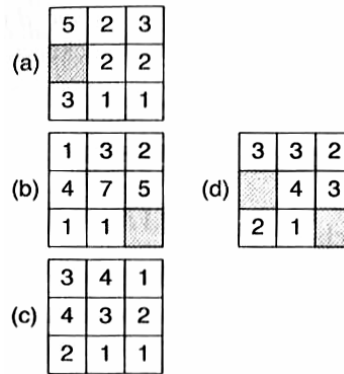


Fig. 1 : The cell value in (d) is the mean calculated from three input rasters (a, b and c) in a local operation. The shaded cells have no data.

Other measures that are suitable for rasters with numeric or categorical data are statistics such as majority, minority, and number of unique values. For each cell, a majority output raster tabulates the most frequent cell value among the input rasters, a minority raster tabulates the least frequent cell value, and a variety raster tabulates the number of different cell values. Figure 2, for example, shows the output with the majority statistics from three input rasters.

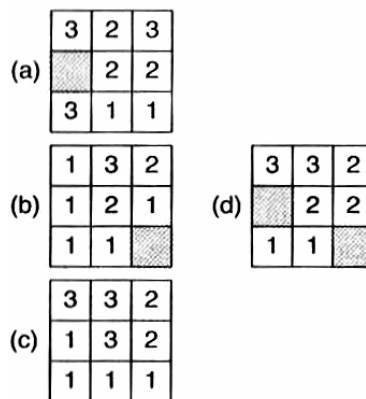


Fig. 2 : The cell value in (d) is the majority statistic derived from three input rasters (a, b and c) in a local operation. The shaded cells have no data.

**Q.6 Attempt any TWO question of the following :** [10]

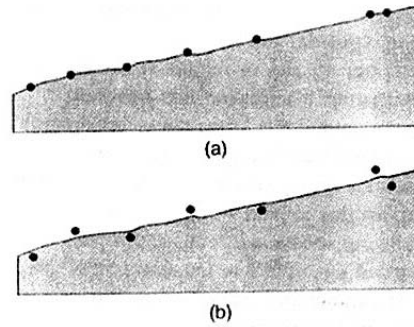
**Q.6(a) What is spatial interpolation? What are its types?** [5]

**(A)** Spatial interpolation is the process of using points with known values at other points. Through spatial interpolation, we can estimate the precipitation value at a location with no recorded data by using known precipitation readings at nearby weather stations. In GIS applications, spatial interpolation is typically applied to a raster with estimates made for all cells. Spatial interpolation is therefore a means of creating surface data from sample points so that the surface data can be used for analysis and modeling.

**Type of Spatial Interpolation**

Spatial interpolation methods can be categorized in several ways. First, they can be grouped into global and local methods. A global interpolation methods uses every known point available to estimate an unknown value. A local interpolation method, on the other hand, uses a sample of known points to estimate an unknown value. Because the difference between the two groups lies in the number of control points used in estimation, one may view the scale from global to local as a continuum.

Conceptually, a global interpolation method is designed to capture the general trend of the surface and a local interpolation method the local or short-range variation. For many phenomena, it is more efficient to estimate the unknown value at a point using a local method than a global method.



**Fig. 1 :** Exact interpolation (a) and inexact interpolation (b).

Faraway points have little influence on the estimated value; in some cases, they may even distort the estimated value. A local method is also preferred because it requires much less computation than a global method does.

Second, spatial interpolation methods can be grouped into exact and inexact interpolation (Figure 1). Exact interpolation predicts a value at the point location that is the same as its known value. In other words, exact interpolation generates a surface that passes through the control points. In contrast, inexact interpolation, or approximate interpolation, predicts a value at the point location that differs from its known value.

Third, spatial interpolation methods may be deterministic or stochastic. A deterministic interpolation methods provides no assessment of errors with predicted values. A stochastic interpolation method, on the other hand, offers assessment of prediction errors with estimated variances. The assumption of a random process is normally required for a stochastic method.

Table 1 shows a classification of spatial interpolation methods covered. Notice that the two global methods can also be used for local operations.

**Table 1 :** A Classification of Spatial Interpolation Methods

Global		Local	
Deterministic	Stochastic	Deterministic	Stochastic
Trend surface (inexact)*	Regression (inexact)	Thiessen (exact)	Kriging (exact)
		Density estimation (inexact)	
		Inverse distance weighted (exact)	
		Splines (exact)	

**Q.6(b) What are Thiessen Polygons? Give an example.**

**[5]**

**(A) Thiessens Polygons**

- Thiessens polygons assume that any point within a polygon is closer to the polygon's known point than any known points. Thiessen polygons were originally proposed to estimate areal averages of precipitation by making sure that any point within a polygon is closer to the polygon's weather station than any other station. Thiessen polygons, also called Voronoi polygons, are used in a variety of applications, especially for service area analysis of public facilities such as hospitals.
- Thiessen polygons do not use an interpolator but require initial triangulation for connecting known points. Because different ways of connecting points can form different sets of triangles, the Delaunay triangulation-the same method for constructing a TIN-is often used in preparing Thiessen polygons. The Delaunay

triangulation ensures that each known point is connected to its nearest neighbours, and that triangles are as equilateral as possible. After triangulation, Thiessen polygons can be easily constructed by connecting lines drawn perpendicular to the sides of each triangle at their midpoints (Figure).

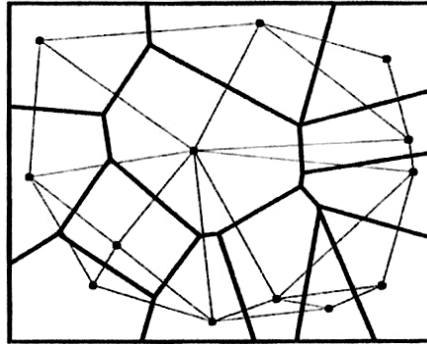


Fig. : Thiessen polygons (in thicker lines) are interpolated from the known points and the Delaunay triangulation (in thinner lines).

- Thiessen polygons are smaller in areas where points are closer together and larger in areas where points are farther apart. This size differentiation is the basis, for example, for determining the quality of public service. A large polygon means greater distances between home locations and public service providers. The size differentiation can also be used for other purposes such as predicting forest age classes, with larger polygons belonging to older trees.

Q.6(c) Explain trend surface model with suitable example. [5]

(A) Trend Surface Models.

- An inexact interpolation method, trend surface analysis approximates points with known values with a polynomial equation. The equation or the interpolator can then be used to estimate values at other points. A linear or first-order trend surface uses the equation :
 
$$z_{x,y} = b_0 + b_1x + b_2y \quad \dots (1)$$
 where the attribute value  $z$  is a function of  $x$  and  $y$  coordinates. The  $b$  coefficients are estimated from the known points. Because the trend surface model is computed by the least-squares method, the "goodness of fit" of the model can be measured and tested. Also, the deviation or the residual between the observed and the estimated values can be computed for each known point
  - The distribution of most natural phenomena is usually more complex than an inclined plane surface from a first model. Higher-order trend surface models are required to approximate more complex surfaces. A cubic or a third-order model, for example, includes hills and valleys. A cubic trend surface is based on the equation :
 
$$Z_{x,y} = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 + b_6x^3 + b_7x^2y + b_8xy^2 + b_9y^3 \quad \dots (2)$$
- A third-order trend surface requires estimation of 10 coefficients (i.e.,  $b_i$ ), compared to three coefficients for a first-order surface. A higher-order trend surface model therefore requires more computation than a lower-order model does. A GIS package may offer up to 12th-order trend surface models.
- There are variations of trend surface analysis. Logistics trend surface analysis uses known points with binary data (i.e. 0 and 1) and produces a probability surface. Local polynomial interpolation uses a sample of known points to estimate the unknown value of a cell. Conversion of a triangulated irregular network (TIN) to a DEM, for instance, uses local polynomial interpolation.



Fig. : An isohyet map in inches from a third-order trend surface model. The point symbols represent known points within Idaho.

Q.6(d) Define following :

[5]

(i) **Nugget**    (ii) **Range**    (iii) **Sill**    (iv) **Partial Sill**    (v) **Anisotropy**

- (A) (i) **Nugget** : A parameter of a covariance or semivariogram model that represents independent error, measurement error, or microscale variation at spatial scales that are too fine to detect. The nugget effect is seen as a discontinuity at the origin of either the covariance or semivariogram model.
- (ii) **Range** : A parameter of a variogram or semivariogram model that represents a distance beyond which there is little or no autocorrelation among variables.
- (iii) **Sill** : A parameter of a variogram or semivariogram model that represents a value that the variogram tends toward when distances become large. Under second-order stationarity, variables become uncorrelated at large distances, so the sill of the semivariogram is equal to the variance of the random variable.
- (iv) **Anisotropy** : Having non uniform spatial distribution of movement or properties, usually across a surface.

Q.7 Attempt any THREE question of the following :

[15]

Q.7(a) Explain modern applications of GIS related to GPS and wireless technology.

[4]

- (A) Integration of GIS with the global positioning system (GPS), wireless technology, and Internet has also introduced new and exciting applications (e.g., Tsou 2004). Here are some examples :
- Location-based services (LBS) technology allows mobile phone users to be located and to receive location information, such as nearby ATMs and restaurants.
  - Interactive-mapping websites let users select map layers for display and make their own maps.
  - In-car navigation systems find the shortest route between an origin and destination and provide turn-by-turn directions to drivers.
  - Mobile mapping allows field workers to collect and access geospatial data in the field.
  - Precision farming promotes site-specific farming activities such as herbicide or fertilizer application.

**Q.7(b) What is geometric transformation? Explain map-to-map and image-to-map [4] transformation methods.**

**(A) Geometric Transformation**

Geometric transformation is the process of using a set of control points and transformation equations to register a digitized map, a satellite image, or an aerial photograph onto a projected coordinate system. As its definition suggests, geometric transformation is a common operation in GIS, remote sensing, and photogrammetry. But the mathematical aspects of geometric transformation are from coordinate geometry (Moffitt and Mikhail 1980).

#### **Map-to-Map and Image-to-Map Transformation**

A newly digitized map, either manually digitized or traced from a scanned file, is based on digitizer units. Digitizer units can be in inches or dots per inch. Geometric transformation converts the newly digitized map into projected coordinates in a process often called map-to-map transformation.

Image-to-map transformation applies to remotely sensed data (Jensen 1996; Richards and Jia 1999). The term suggests that the transformation changes the rows and columns (i.e., the image coordinates) of a satellite image into projected coordinates. Another term describing this kind of transformation is georeferencing (Verbyla and Chang 1997; Lillesand et al. 2004). A georeferenced image can register spatially with other feature or raster layers in a GIS database, as long as the coordinate system is the same.

Whether map-to-map or image-to-map, a geometric transformation uses a set of control points to establish a mathematical model that relates the map coordinates of one system to another or image coordinates to map coordinates. The use of control points makes the process somewhat uncertain. This is particularly true with image-to-map transformation because control points are selected directly from the original image. Misplacement of the control points can make the transformation result unacceptable.

**Q.7(c) Explain Zonal Operations on Raster Data with suitable example. [4]**

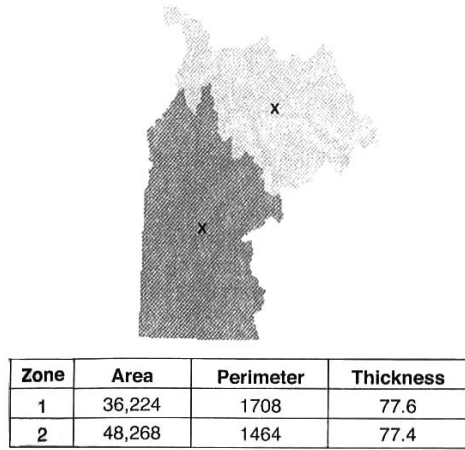
**(A) Zonal Operations**

A zonal operation works with groups of cells of same values or like features. These groups are called zones. Zones may be contiguous or noncontiguous. A contiguous zone includes cells that are spatially connected, whereas a noncontiguous zone includes separate regions of cells.

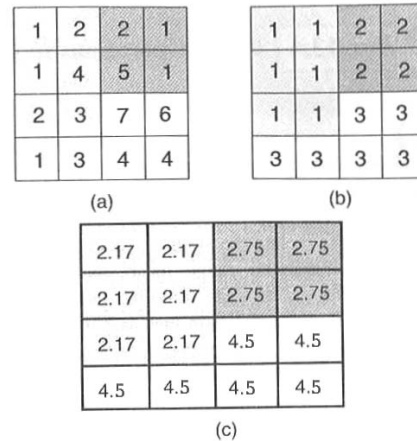
#### **Zonal Statistics**

A zonal operation may work with a single raster or two rasters. Given a single input raster, zonal operations measure the geometry of each zone in the raster, such as area, perimeter, thickness, and centroid (Figure 1). The area is the sum of the cells that fall within the zone times the cell size. The perimeter of a contiguous zone is the length of its boundary, and the perimeter of a noncontiguous zone is the sum of the length of each region. The thickness calculates the radius (in cells) of the largest circle that can be drawn within each zone. And the centroid is the geometric center of a zone located at the intersection of the major axis and the minor axis of an ellipse that best approximates the zone.

Given two rasters in a zonal operation, one input raster and one zonal raster, a zonal operation produces an output raster, which summarizes the cell values in the input raster for each zone in the zonal raster. The summary statistics and measures include area, minimum, maximum, sum, range, mean, standard deviation, median, majority, minority, and variety. (The last four measures are not available if the input raster is a floating-point raster.) Figure 2 shows a zonal operation of computing the mean by zone. Figure 2(b) is the zonal raster with three zones. Figure 2(a) is the input raster, and Figure 2(c) is the output raster.



**Fig. 1 :** Thickness and centroid for two large watersheds (zones). Area is measured in square kilometers, and perimeter and thickness are measured in kilometers. The centroid of each zone is marked with an x.



**Fig. 2 :** The cell values in (c) are the zonal means derived from an input raster (a) and a zonal raster (b). For example, 2.17 is the mean of {1, 1, 2, 2, 4, 3} for zone 1.

**Applications of Zonal Operations**

Measures of zonal geometry such as area, perimeter, thickness, and centroid are particularly useful for studies of landscape ecology (Forman and Godron 1986; McGarigal and Marks 1994). Many other geometric measures can be derived from area and perimeter. For example, a shape index called the areal roundness is defined as 354 times the square root of a zone's area divided by its perimeter (Tomlin 1990). Essentially, the areal roundness compares the shape of a contiguous zone to a circle. The area roundness value approaches 100 for a circular shape and 0 for a highly distorted shape. The expression for computing areal roundness is:  $354 \times \text{Sqrt}(\text{zonalarea}([\text{raster1}])) / \text{zonalperimeter}([\text{raster 1}])$ , where raster1 is the input raster and Sqrt is the square root function.

Zonal operations with two rasters can generate useful descriptive statistics for comparison purposes. For example, to compare topographic characteristics of different soil textures, we can use a soil raster that contains the categories of sand, loam, and clay as the zonal raster and slope, aspect, and elevation as the input rasters. By running a series of zonal operations, we can summarize the slope, aspect, and elevation characteristics associated with the three soil texture.

**Q.7(d) Explain with suitable example cell-by-cell encoding raster data structure. [4]**

**(A) Cell by Cell Encoding**

- The cell-by-cell encoding method provides the simplest raster data structure. A raster is stored as a matrix, and its cell values are written into a file by row and column (Figure 1). Functioning at the cell level, this method is an ideal choice if the cell values of a raster change continuously.
- DEMs use the cell-by-cell data structure because the neighboring elevation values are rarely the same. Satellite images also use the cell-by-cell encoding method for data storage. With multiple spectral bands, however, each pixel in a satellite image has more than one value.
- Multiband imagery is typically stored in the following three formats. The band sequential (.bsq) method stores the values of an image band as one file. Therefore, if an image has seven bands, the data set has seven consecutive files, one file per band. The band interleaved by line (.bil) method stores, row by row, the values of all the bands in one file. Therefore the file consists of row 1, band 1; row 1, band 2 ... row 2. band 1; row 2, band 2 ... and so on. The band interleaved by pixel (.bip) method stores the values of

all the bands by pixel in one file. The file is therefore comprised of pixel (1,1), band 1; pixel (1,1), band 2 ... pixel (2, 1), band 1; pixel (2, 1), band 2 ... and so on.

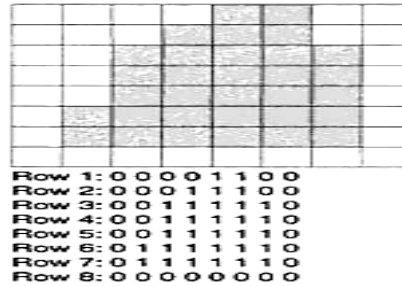


Fig. 1 : The cell-by-cell data structure records each cell value by row and column. The gray cells have the cell value of 1.

**Q.7(e) Write short note on Map production.**

[4]

- (A)
- GIS users design and make maps on the computer screen. These soft-copy maps can be used in a variety of ways. They can be printed, exported for use on the Internet, used in overhead computer projection systems, exported to other software packages, or further processed for publishing.
  - Many a times, color symbols from the color printers do not exactly match those on the computer screen. This discrepancy results from the use of different media and color models.
  - Data display on the computer screen uses either CRT (cathode ray tube) or LCD (liquid crystal display).
  - With either a CRT or an LCD, a color symbol we see on a screen is made of pixels, and the color of each pixel is a mixture of RGB (red, green, and blue). The intensity of each primary color in a mixture determines its color. The number of intensity levels each primary color can have depends on the number of bit-planes assigned to the electron gun in a CRT screen or the variation of the voltage applied in an LCD screen. Typically, the intensity of each primary color can range over 256 shades. Combining the three primary colors produces a possible palette of 16.8 million colors (256 x 256 X 256).
  - Many GIS packages offer the RGB color model for color specification. But color mixtures of RGB are not intuitive and do not correspond to the human dimensions of color perception. That is why other color models have been developed to specify colors that are based on the visual perception of hue, value and chroma.

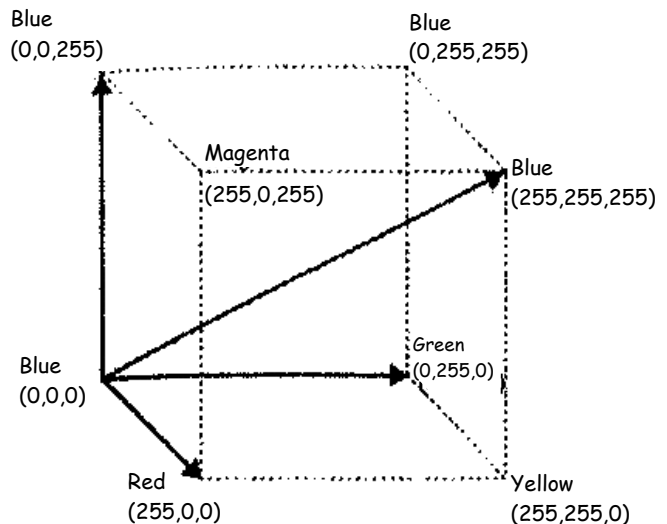


Fig. 1: The RGB (red, green and blue) color model

**Q.7(f) Explain universal kriging.**

[4]

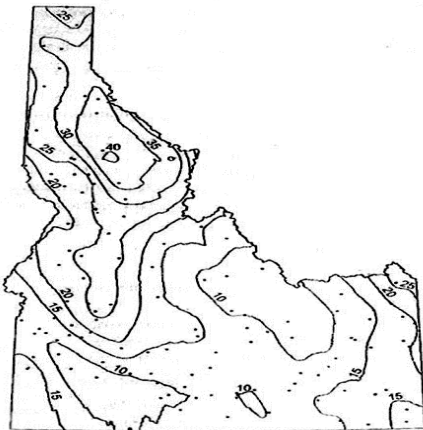
- (A)
- Universal kriging assumes that the spatial variation in z values has a drift or a trend in addition to the spatial correlation between the sample points. Typically, universal kriging incorporates a first-order (plane surface) or a second-order (quadratic surface) polynomial in the kriging process. A first-order polynomial is :  

$$M = b_1x_i + b_2y_i \quad \dots (1)$$

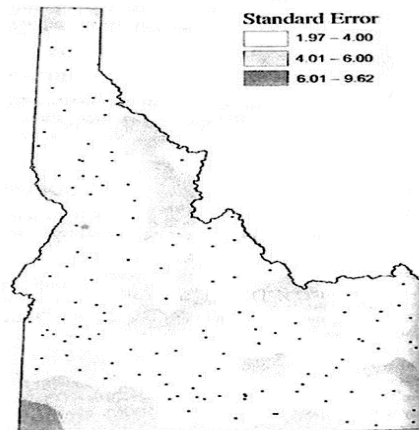
where  $M$  is the drift,  $x_i$  and  $y_i$  are the  $x, y$ -coordinates of sampled point  $i$ , and  $b_1$  and  $b_2$  are the drift coefficients. A second-order polynomial is :

$$M = b_1x_i + b_2y_i + b_3x_i^2 + b_4x_iy_i + b_5y_i^2 \quad \dots (2)$$

- Higher-order polynomials are usually not recommended for two reasons. First, kriging is performed on the residuals after the trend is removed. A higher-order polynomial will leave little variation in the residuals for assessing uncertainty. Second, a higher-order polynomial means a larger number of the  $b_i$  coefficients, which must be estimated along with the weights, and a larger set of simultaneous equations to be solved.
- Figure 1 shows an annual precipitation surface created by universal kriging with the linear (first-order) drift, and figure 2 shows the distribution of the standard error of the predicted surface. Universal kriging produces less reliable estimates than ordinary kriging in this case.



**Fig. 1 :**  
An isohyet map created by universal kriging with the linear drift and the spherical model.



**Fig. 2 :**  
Standard errors of the annual precipitation surface in figure 1.

