Practical work in Geography

Part II

Textbook for Class XII

Contents

1 Data – Its Source and Compilation

You must have seen and used various forms of data. For example, at the end of almost every news bulletin on Television, the temperatures recorded on that day in major cities are displayed. Similarly, the books on the Geography of India show data relating to the growth and distribution of population, and the production, distribution and trade of various crops, minerals and industrial products in tabular form. Have you ever thought what they mean? From where these data are obtained? How are they tabulated and processed to extract meaningful information from them ? In this chapter, we will deliberate on these aspects of the data and try to answer these many questions.

What is Data?

The data are defined as numbers that represent measurements from the real world. **Datum** is a single measurement. We often read the news like 20 centimetres of continuous rain in Barmer or 35 centimetres of rain at a stretch in Banswara in 24 hours or information such as New Delhi – Mumbai distance via Kota – Vadodara is 1385 kilometres and via Itarsi - Manmad is 1542 kilometres by train. This numerical information is called data. It may be easily realised that there are large volume of data available around the world today. However, at times, it becomes difficult to derive logical conclusions from these data if they are in raw form. Hence, it is important to ensure that the measured information is algorithmically derived and/or logically deduced and/or statistically calculated from multiple data. **Information** is defined as either a meaningful answer to a query or a meaningful stimulus that can cascade into further queries.

Need of Data

Maps are important tools in studying geography. Besides, the distribution and growth of phenomena are also explained through the data in tabular form. We know that an interelationship exists between many phenomena over the surface of the earth. These interactions are influenced by many variables which can be

explained best in quantitative terms. Statistical analysis of those variables has become a necessity today. For example, to study cropping pattern of an area, it is necessary to have statistical information about the cropped area, crop yield and production, irrigated area, amount of rainfall and inputs like use of fertiliser, insecticides, pesticides, etc. Similarly, data related to the total population, density, number of migrants, occupation of people, their salaries, industries, means of transportation and communication is needed to study the growth of a city. Thus, data plays an important role in geographical analysis.

Presentation of the Data

You might have heard the story of a person who was travelling with his wife and a five-year old child. On his way, he had to cross a river. Firstly, he fathomed the depth of the river at four points as 0.6, 0.8, 0.9 and 1.5 metres. He calculated the average depth as 0.95 metres. His child's height was 1 metre. So, he led them to cross the river and his child drowned in the river. On the other bank, he sat pondering: "*Lekha Jokha Thahe, to Bachha Dooba Kahe* ?" (Why did the child drown when average depth was within the reach of each one ?). This is called statistical fallacy, which may deviate you from the real situation. So, it is very important to collect the data to know the facts and figures, but equally important is the presentation of data. Today, the use of statistical methods in the analysis, presentation and in drawing conclusions plays a significant role in almost all disciplines, including geography, which use the data. It may, therefore, be inferred that the concentration of a phenomena, e.g. population, forest or network of transportation or communication not only vary over space and time but may also be conveniently explained using the data. In other words, you may say that there is a shift from qualitative description to quantitative analysis in explaining the relationship among variables. Hence, analytical tools and techniques have become more important these days to make the study more logical and derive precise conclusion. Precise quantitative techniques are used right from the beginning of collecting and compiling data to its tabulation, organisation, ordering and analysis till the derivation of conclusions.

Sources of Data

The data are collected through the following ways. These are : 1. Primary Sources, and 2. Secondary Sources.

The data which are collected for the first time by an individual or the group of individuals, institution/organisations are called **Primary sources of the data**. On the other hand, data collected from any published or unpublished sources are called **Secondary sources**. *Fig. 1.1* shows the different methods of data collection.

Sources of Primary Data

1. Personal Observations

It refers to the collection of information by an individual or group of individuals through direct observations in the field. Through a field survey, information about the relief features, drainage patterns, types of soil and natural vegetation, as well as population structure, sex ratio, literacy, means of transport and communication, urban and rural settlements, etc. is collected. However, in carrying

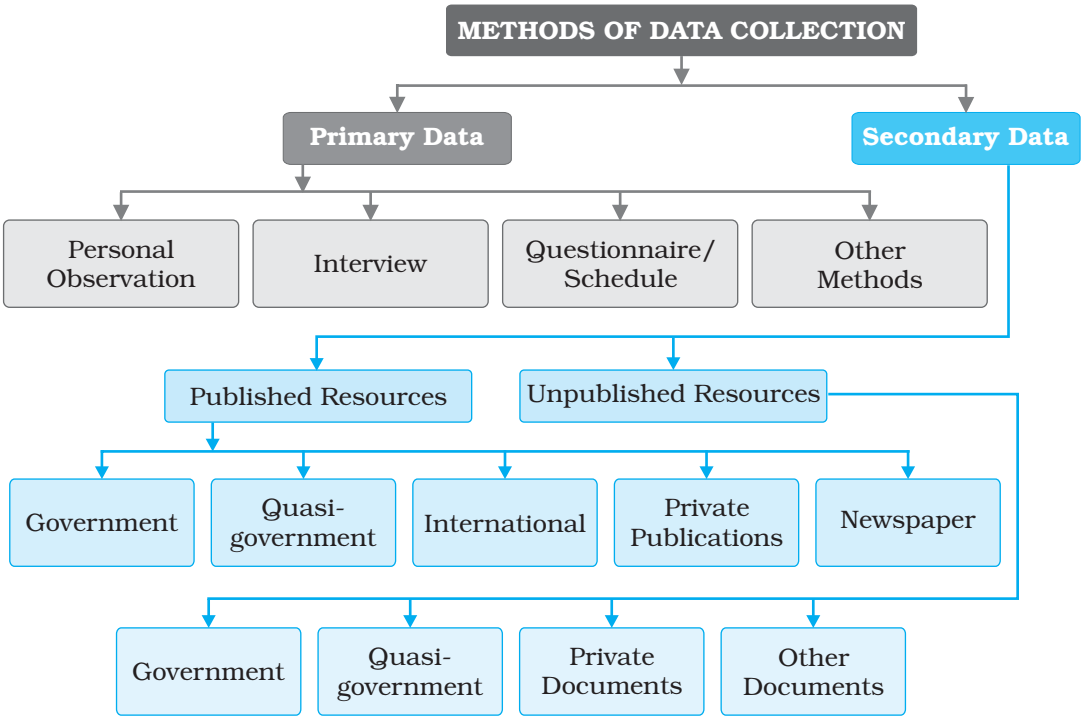


Fig. 1.1 : *Methods of Data Collection*

out personal observations, the person(s) involved must have theoretical knowledge of the subject and scientific attitude for unbiased evaluation.

2. Interview

In this method, the researcher gets direct information from the respondent through dialogues and conversations. However, the interviewer must take the following precautions while conducting an interview with people of the area:

(i) A precise list of items about which information is to be gathered from the persons interviewed be prepared.

(ii) The person(s) involved in conducting the interview should be clear about the objective of the survey.

(iii) The respondents should be taken into confidence before asking any sensitive question and he/she be assured that the secrecy will be maintained.

(iv) A congenial atmosphere should be created so that the respondent may explain the facts without any hesitation.

(v) The language of the questions should be simple and polite so that the respondents feel motivated and readily agree to give information asked for.

(vi) Avoid asking any such question that may hurt the self – respect or the religious feelings of the respondent.

(vii) At the end of interview, ask the respondent what additional information he/she may provide, other than what has already been provided by him/her.

(viii) Pay your thanks and gratefulness for sparing his/her valuable time for you.

3. Questionnaire/Schedule

In this method, simple questions and their possible answers are written on a plain paper and the respondents have to tick-mark the possible answers from the given choices. At times, a set of structured questions are written and sufficient space is provided in the questionnaire where the respondent write their opinion. The objectives of the survey should be clearly mentioned in the questionnaire. This method is useful in carrying out the survey of a larger area. Even questionnaire can be mailed to far-flung places. The limitation of the method is that only the literate and educated people can be approached to provide the required information. Similar to the questionnaire that contains the questions pertaining to the matter of investigation is the **schedule**. The only difference between the **questionnaire** and the **schedule** is that the respondent himself/herself fills up the questionnaires, whereas a properly trained enumerator himself fills up schedules by asking question addressed to the respondents. The main advantage of schedule over the questionnaire is that the information from both literate and illiterate respondents can be collected.

4. Other Methods

The data about the properties of soil and water are collected directly in the field by measuring their characteristics using soil kit and water quality kit. Similarly, field scientist collect data about the health of the crops and vegetation using transducers (*Fig. 1.2*).

Fig. 1.2 : *Field Scientist taking Measures of Crop Health*

Secondary Source of Data

Secondary sources of data consist of published and unpublished records which include government publications, documents and reports.

Published Sources

1. Government Publications

The publications of the various ministries and the departments of the Government of India, state governments and the District Bulletins are one of the most important sources of secondary information. These include the Census of India published by the Office of the Registrar General of India, reports of the National Sample Survey, Weather Reports of Indian Meteorological Department, and Statistical Abstracts published by state governments, and the periodical reports published by different Commissions. Some of the government publications are shown in *Fig. 1.3*.

Fig. 1.3 : *Some of the Government Publications*

2. Semi/Quasi-government Publications

The publications and reports of Urban Development Authorities and Municipal Corporations of various cities and towns, Zila Parishads (District Councils), etc. fall under this category.

3. International Publications

The international publications comprise yearbooks, reports and monographs published by different agencies of the United Nations such as United Nations Educational, Scientific and Cultural Organisation (UNESCO), United Nations Development Programme (UNDP), World Health Organisation (WHO), Food and Agricultural Organisation (FAO), etc. Some of the important publications of the United Nations that are periodically published are Demographic Year Book, Statistical Year Book and the Human Development Report (*Fig. 1.4*).

Fig. 1.4 : *Some of the United Nations Publications*

4. Private Publications

The yearbooks, surveys, research reports and monographs published by newspapers and private organisations fall under this category.

5. Newspapers and Magazines

The daily newspapers and the weekly, fortnightly and monthly magazines serve as easily accessible source of secondary data.

6. Electronic Media

The electronic media specially internet has emerged as a major source of secondary data in recent times.

Unpublished Sources

1. Government Documents

The unpublished reports, monographs and documents are yet another source of secondary data. These documents are prepared and maintained as unpublished record at different levels of governance. For example, the village level revenue records maintained by the patwaris of respective villages serve as an important source of village level information.

2. Quasi-government Records

The periodical reports and the development plans prepared and maintained by different Municipal Corporations, District Councils, and Civil Services departments are included in Quasi – government records.

3. Private Documents

These include unpublished reports and records of companies, trade unions, different political and apolitical organisations and resident welfare associations.

Tabulation and Classification of Data

The data collected from primary or secondary sources initially appear as a big jumble of information with the least of comprehension. This is known as raw data. To draw meaningful inferences and to make them usable the raw data requires tabulation and classification.

One of the simplest devices to summarise and present the data is the **Statistical Table**. It is a systematic arrangement of data in columns and rows. The purpose of table is to simplify the presentation and to facilitate comparisons. This table enables the reader to locate the desired information quickly. Thus, the tables make it possible for the analyst to present a huge mass of data in an orderly manner within a minimum of space.

Data Compilation and Presentation

Data are collected, tabulated and presented in a tabular form either in absolute terms, percentages or indices.

Absolute Data

When data are presented in their original form as integers, they are called absolute data or **raw data**. For example, the total population of a country or a state, the total production of a crop or a manufacturing industry, etc. *Table 1.1* shows the absolute data of population of India and some of the selected states.

Table 1.1 : Population of India and Selected States/Union Territories, 2001

State/ UT Code	India/State/ Union territory	Total Population		
		Persons	Males	Females
1	2	3	4	5
	INDIA *	**1,027,015,247**	**531,277,078**	**495,738,169**
1.	Jammu & Kashmir #	10,069,917	5,300,574	4,769,343
2.	Himachal Pradesh	6,077,248	3,085,256	2,991,992
3.	Punjab	24,289,296	12,963,362	11,325,934
4.	Chandigarh ##	900,914	508,224	392,690
5.	Uttaranchal	8,479,562	4,316,401	4,163,161
6.	Haryana	21,082,989	11,327,658	9,755,331
7.	National Capital Territory of Delhi	13,782,976	7,570,890	6,212,086
8.	Rajasthan	56,473,122	29,381,657	27,091,465
9.	Uttar Pradesh	166,052,859	87,466,301	78,586,558
10.	Bihar	82,878,796	43,153,964	39,724,832

* inclusive of all territorial boundary of India
\# excluding PoK
\## Union Territory

Percentage/Ratio

Some time data are tabulated in a ratio or percentage form that are computed from a common parameter, such as literacy rate or growth rate of population, percentage of agricultural products or industrial products, etc. *Table 1.2* presents

literacy rates of India over the decades in a percentage form. Literacy Rate is calculated as :

$$\frac{\text{Total Literates}}{\text{Total Population}} \quad 100$$

Table 1.2 : Literacy Rate* : 1951 – 2001

Year	Person	Male	Female
1951	18.33	27.16	8.86
1961	28.3	40.4	15.35
1971	34.45	45.96	21.97
1981	43.57	56.38	29.76
1991	52.21	64.13	39.29
2001	64.84	75.85	54.16

* as percentage of total
Source: Census of India, 2001

Index Number

An index number is a statistical measure designed to show changes in variable or a group of related variables with respect to time, geographic location or other characteristics. It is to be noted that index numbers not only measure changes over a period of time but also compare economic conditions of different locations, industries, cities or countries. Index number is widely used in economics and business to see changes in price and quantity. There are various methods for the calculation of index number. However, the simple aggregate method is most commonly used. It is obtained using the following formula:

$$\frac{\sum q_1}{\sum q_0} \quad 100$$

$\sum q_1$ = Total of the current year production

$\sum q_0$ = Total of the base year production

Generally base year values are taken as 100 and index number is calculated thereupon. For example, *Table 1.3* shows the production of iron ore in India and the changes in index number from 1970 – 71 to 2000 – 01 taking 1970-71 as the base year.

Table 1.3 : Production of Iron ore in India

	Production (in million tonnes)	Calculation	Index Number
1970-71	32.5	$\frac{32.5}{32.5}$ 100	100
1980-81	42.2	$\frac{42.2}{32.5}$ 100	130
1990-91	53.7	$\frac{53.7}{32.5}$ 100	165
2000-01	67.4	$\frac{67.4}{32.5}$ 100	207

Source – India : Economic Year Book, 2005

Processing of Data

The processing of raw data requires their tabulation and classification in selected classes. For example, the data given in *Table 1.4* can be used to understand how they are processed.

We can see that the given data are ungrouped. Hence, the first step is to group data in order to reduce its volume and make it easy to understand.

Table 1.4 : Score of 60 Students in Geography Paper

47	02	39	64	22	46	28	02	09	10
89	96	74	06	26	15	92	84	84	90
32	22	53	62	73	57	37	44	67	50
18	51	36	58	28	65	63	59	75	70
56	58	43	74	64	12	35	42	68	80
64	37	17	31	41	71	56	83	59	90

Grouping of Data

The grouping of the raw data requires determining of the number of classes in which the raw data are to be grouped and what will be the class intervals. The selection of the class interval and the number of classes, however, depends upon the range of raw data. The raw data given in *Table 1.4* ranges from 02 to 96. We can, therefore, conveniently choose to group the data into ten classes with an interval of ten units in each group, e.g. 0 – 10, 10 – 20, 20 – 30, etc. (*Table 1.5*).

Table 1.5 : Making Tally Marks to Obtain Frequency

Group	*Numerical of Raw Data*	*Tally Marks*	*Number of Individual*
0-10	02,02,09,06	////	4
10-20	10,15,18,12,17	𝍸	5
20-30	22,28,26,22,28	𝍸	5
30-40	39,32,37,36,35,37,31	𝍸//	7
40-50	47,46,44,43,42,41	𝍸/	6
50-60	53,57,50,51,58, 59,56,58,56,59	𝍸///	10
60-70	64,62,67,65, 63,64,68,64	𝍸 𝍸	8
70-80	74,73,75,70,74,71	𝍸 ///	6
80-90	89,84,84,80,83	𝍸 /	5
90-100	96,92,90,90	𝍸	4
			$\sum f = N = 60$

Process of Classification

Once the number of groups and the class interval of each group are determined, the raw data are classified as shown in *Table 1.5*. It is done by a method popularly known as **Four and Cross Method** or tally marks.

First of all, one tally mark is assigned to each individual in the group in which it is falling. For example, the first numerical in the raw data is 47. Since, it falls in the group of 40 – 50, one tally mark is recorded in the column 3 of *Table 1.5*.

Frequency Distribution

In *Table 1.5* we have classified the raw data of a quantitative variable and have grouped them class-wise. The numbers of individuals (places in the fourth column of *Table 1.5*) are known as frequency and the column represents the frequency

distribution. It illustrates how the different values of a variable are distributed in different classes. Frequencies are classified as **Simple** and **Cumulative frequencies**.

Simple Frequencies

It is expressed by ***'f'*** and represent the number of individuals falling in each group *(Table 1.6)*. The sum of all the frequencies, assigned to all classes, represents the total number of individual observations in the given series. In statistics, it is expressed by the symbol N that is equal to $\sum f$. It is expressed as $\sum f = N = 60$ (*Table 1.5* and *1.6*).

Table 1.6 : Frequency Distribution

Group	*f*	*Cf*
00-10	4	4
10-20	5	9
20-30	5	14
30-40	7	21
40-50	6	27
50-60	10	37
60-70	8	45
70-80	6	51
80-90	5	56
90-100	4	60
	$\sum f = N = 60$	

Cumulative Frequencies

It is expressed by ***'Cf'*** and can be obtained by adding successive simple frequencies in each group with the previous sum, as shown in the column 3 of *Table 1.6*. For example, the first simple frequency in *Table 1.6* is 4. Next frequency of 5 is added to 4 which gives a total of 9 as the next cumulative frequency. Likewise add every next number until the last cumulative frequency of 60 is obtained. Note that it is equal to N or $\sum f$.

Advantage of cumulative frequency is that one can easily make out that there are 27 individuals scoring less than 50 or that 45 out of 60 individuals lie below the score of 70.

Each simple frequency is associated with its group or class. The **exclusive** or **inclusive** methods are used for forming the groups or classes.

Exclusive Method

As shown in *Table 1.6*, two numbers are shown in its first column . Notice that the upper limit of one group is the same as the lower limit of the next group. For example, the upper limit of the one group (20 – 30) is 30, which is the lower limit of the next group (30 – 40), making 30 to appear in both groups. But any observation having the value of 30 is included in the group where it is at its lower limit and it is excluded from the group where it is the upper limit as (in 20-30 groups). That is why the method is known as exclusive method, i.e. a group is excluded of its upper limits. You may now make out where all the marginal values of *Table 1.4* will go.

The groups in *Table 1.6*, are interpreted in the following manner –

0 and under 10	10 and under 20
20 and under 30	30 and under 40
40 and under 50	50 and under 60
60 and under 70	70 and under 80
80 and under 90	90 and under100

Hence, in this type of grouping the class extends over ten units. For example, 20, 21, 22, 23, 24, 25, 26, 27, 28 and 29 are included in the third group.

Inclusive Method

In this method, a value equal to the upper limit of a group is included in the same group. Therefore, it is known as inclusive method. Classes are mentioned in a different form in this method, as is shown in the first column of *Table 1.7*. Normally, the upper limit of a group differs by 1 with the lower limits of the next group. It is important to note that each group spreads over ten units in this method also. For example, the group of 50 – 59 includes the ten values i.e. 50, 51, 52, 53, 54, 55, 56, 57, 58 and 59 (*Table 1.7*). In this method both upper and lower limit are included to find the frequency distribution.

Table 1.7 : Frequency Distribution

Group	f	Cf
0 – 9	4	4
10 – 19	5	9
20 – 29	5	14
30 – 39	7	21
40 – 49	6	27
50 – 59	10	37
60 – 69	8	45
70 – 79	6	51
80 – 89	5	56
90 – 99	4	60
	$\sum f = N = 60$	

Frequency Polygon

A graph of frequency distribution is known as the frequency polygon. It helps in comparing the two or more than two frequency distributions (*Fig. 1.5*). The two frequencies are shown using a bar diagram and a line graph respectively.

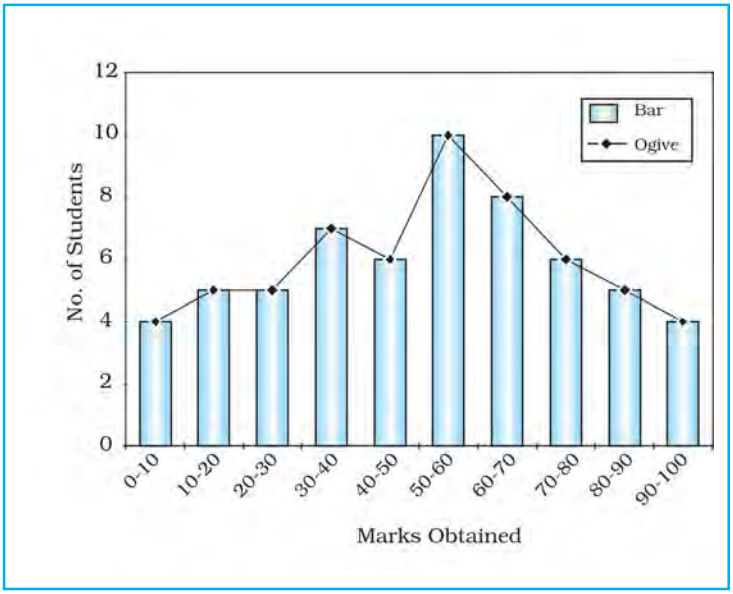


Fig. 1.5 : *Frequency Distribution Polygon*

Ogive

When the frequencies are added they are called cumulative frequencies and are listed in a table called cumulative frequency table. The curve obtained by plotting cumulative frequencies is called an **Ogive** (pronounced as ojive). It is constructed either by the **less than method** or the **more than method**.

In the **less than method**, we start with the upper limit of the classes and go on adding the frequencies. When these frequencies are plotted, we get a rising curve as shown in *Table 1.8* and *Fig. 1.6*.

In the **more than method,** we start with the lower limits of the classes and from the cumulative frequency, we substract frequency of each class. When these frequencies are plotted, we get a declining curve as shown in *Table 1.9* and *Fig 1.7*.

Both the *Figs. 1.5* and *1.6* may be combined to get a comparative picture of less than and more than Ogive as shown in *Table 1.10* and *Fig. 1.7*.

Table 1.8 : Frequency Distribution less than Method

Less than Method	*Cf*
Less than 10	4
Less than 20	9
Less than 30	14
Less than 40	21
Less than 50	27
Less than 60	37
Less than 70	45
Less than 80	51
Less than 90	56
Less than 100	60

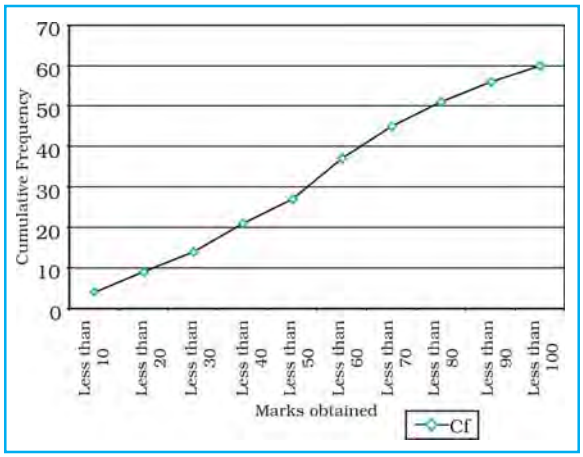


Fig. 1.6 : *Less than Ogive*

Table 1.9 : Frequency Distribution more than Method

More than Method	*Cf*
More than 0	60
More than 10	56
More than 20	51
More than 30	44
More than 40	38
More than 50	28
More than 60	20
More than 70	14
More than 80	9
More than 90	4

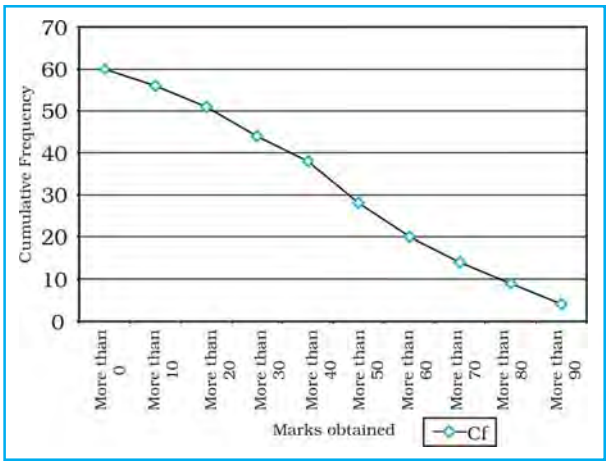


Fig. 1.7 : *More than Ogive*

Table 1.10 : Less than and more than Ogive

Marks obtained	*Less than*	*More than*
0 - 10	4	60
10 - 20	9	56
20 - 30	14	51
30 - 40	21	44
30 - 40	27	38
50 - 60	37	28
60 - 70	45	20
70 - 80	51	14
80 - 90	56	9
90 - 100	60	4

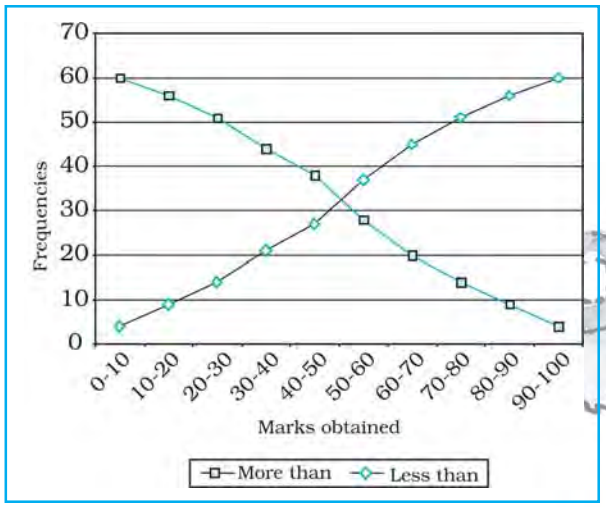


Fig. 1.8 : *Less than and more than Ogive*

Excercises

1. Choose the right answer from the four alternatives given below:

(i) A number or character which represents measurement is called
(a) Digit (b) Data (c) Number (d) Character

(ii) A single datum is a single measurement from the
(a) Table (b) Frequency (c) Real world (d) Information

(iii) In a tally mark grouping by four and crossing fifth is called
(a) Four and Cross Method (b) Tally Marking Method
(c) Frequency plotting Method (d) Inclusive Method

(iv) An Ogive is a method in which
(a) Simple frequency is measured
(b) Cumulative frequency is measured
(c) Simple frequency is plotted
(d) Cumulative frequency is plotted

(v) If both ends of a group are taken in frequency grouping, it is called
(a) Exclusive Method (b) Inclusive Method
(c) Marking Method (d) Statistical Method

2. Answer the following questions in about 30 words:

(i) Differentiate between data and information.

(ii) What do you mean by data processing?

(iii) What is the advantage of foot note in a table?

(iv) What do you mean by primary sources of data?

(v) Enumerate five sources of secondary data.

3. Answer the following questions in about 125 words:

(i) Discuss the national and international agencies where from secondary data may be collected.

(ii) What is the importance of an index number? Taking an example examine the process of calculating an index number and show the changes.

Activity

1. In a class of 35 students of Geography, following marks were obtained out of 10 marks in unit test – 1, 0, 2, 3, 4, 5, 6, 7, 2, 3, 4, 0, 2, 5, 8, 4, 5, 3, 6, 3, 2, 7, 6, 5, 4, 3, 7, 8, 9, 7, 9, 4, 5, 4, 3. Represent the data in the form of a group frequency distribution.

2. Collect the last test result of Geography of your class and represent the marks in the form of a group frequency distribution.

2 Data Processing

You have learnt in previous chapter that organising and presenting data makes them comprehensible. It facilitates data processing. A number of statistical techniques are used to analyse the data. In this chapter, you will learn the following statistical techniques:

1. Measures of Central Tendency
2. Measures of Dispersion
3. Measures of Relationship

While measures of central tendency provide the value that is an ideal representative of a set of observations, the measures of dispersion take into account the internal variations of the data, often around a measure of central tendency. The measures of relationship, on the other hand, provide the degree of association between any two or more related phenomena, like rainfall and incidence of flood or fertiliser consumption and yield of crops.

Measures of Central Tendency

The measurable characteristics such as rainfall, elevation, density of population, levels of educational attainment or age groups vary. If we want to understand them, how would we do ? We may, perhaps, require a single value or number that best represents all the observations. This single value usually lies near the centre of a distribution rather than at either extreme. The statistical techniques used to find out the centre of distributions are referred as **measures of central tendency**. The number denoting the central tendency is the representative figure for the entire data set because it is the point about which items have a tendency to cluster.

Measures of central tendency are also known as statistical averages. There are a number of the measures of central tendency, such as the **mean**, **median** and the **mode**.

Mean

The mean is the value which is derived by summing all the values and dividing it by the number of observations.

Median

The median is the value of the rank, which divides the arranged series into two equal numbers. It is independent of the actual value. Arranging the data in ascending or descending order and then finding the value of the middle ranking number is the most significant in calculating the median. In case of the even numbers the average of the two middle ranking values will be the median.

Mode

Mode is the maximum occurrence or frequency at a particular point or value. You may notice that each one of these measures is a different method of determining a single representative number suited to different types of the data sets.

Mean

Mean is the simple arithmetic average of the different values of a variable. For ungrouped and grouped data, the methods for calculating mean are necessarily different. Mean can be calculated by direct or indirect methods, for both grouped and ungrouped data.

Computing Mean from Ungrouped Data

Direct Method

While calculating mean from ungrouped data using the direct method, the values for each observation are added and the total number of occurrences are divided by the sum of all observations. The mean is calculated using the following formula:

$$\bar{X} \quad \frac{x}{N}$$

Where,

$\bar{X}$ = Mean
= Sum of a series of measures
x = A raw score in a series of measures
x = The sum of all the measures
N = Number of measures

Example 2.1 : Calculate the mean rainfall for Malwa Plateau in Madhya Pradesh from the rainfall of the districts of the region given in *Table 2.1:*

Table 2.1 : Calculation of Mean Rainfall

Districts in Malwa Plateau	*Normal Rainfall in mms*	*Indirect Method*
	x Direct Method	***d= x-800****
Indore	979	179
Dewas	1083	283
Dhar	833	33
Ratlam	896	96
Ujjain	891	91
Mandsaur	825	25
Shajapur	977	177
x and d	**6484**	**884**
$\frac{x}{N}$ and $\frac{d}{N}$	**926.29**	**126.29**

* Where 800 is assumed mean.
d is deviation from the assumed mean.

The mean for the data given in *Table 2.1* is computed as under :

$$\bar{X} = \frac{x}{N}$$

$$= \frac{6,484}{7}$$

$$= 926.29$$

It could be noted from the computation of the mean that the raw rainfall data have been added directly and the sum is divided by the number of observations i. e. districts. Therefore, it is known as **direct method**.

Indirect Method

For large number of observations, the indirect method is normally used to compute the mean. It helps in reducing the values of the observations to smaller numbers by subtracting a constant value from them. For example, as shown in *Table 2.1*, the rainfall values lie between 800 and 1100 mm. We can reduce these values by selecting 'assumed mean' and subtracting the chosen number from each value. In the present case, we have taken 800 as assumed mean. Such an operation is known as **coding**. The mean is then worked out from these reduced numbers (Column 3 of *Table 2.1*).

The following formula is used in computing the mean using indirect method:

$$\bar{X} = A + \frac{d}{N}$$

Where,

A = Subtracted constant

d = Sum of the coded scores

N = Number of individual observations in a series

Mean for the data as shown in *Table 2.1* can be computed using the indirect method in the following manner :

$$\bar{X} = 800 + \frac{884}{7}$$

$$= 800 + \frac{884}{7}$$

$$\bar{X} = 926.29 \text{ mm}$$

Note that the mean value comes the same when computed either of the two methods.

Computing Mean from Grouped Data

The mean is also computed for the grouped data using either direct or indirect method.

Direct Method

When scores are grouped into a frequency distribution, the individual values lose their identity. These values are represented by the midpoints of the class intervals in which they are located. While computing the mean from grouped

data using direct method, the midpoint of each class interval is multiplied with its corresponding frequency (f); all values of fx (the X are the midpoints) are added to obtain fx that is finally divided by the number of observations i. e. N. Hence, mean is calculated using the following formula :

$$\bar{X} \quad \frac{fx}{N}$$

Where :

$\bar{X}$ = Mean
f = Frequencies
x = Midpoints of class intervals
N = Number of observations (it may also be defined as f)

Example 2.2 : Compute the average wage rate of factory workers using data given in *Table 2.2*:

Table 2.2 : Wage Rate of Factory Workers

Wage Rate (Rs./day)	***Number of workers (f)***
Classes	***f***
50 - 70	10
70 - 90	20
90 - 110	25
110 - 130	35
130 - 150	9

Table 2.3 : Computation of Mean

Classes	***Frequency (f)***	***Mid-points (x)***	***fx***	***d=x-100***	***fd***	***U = (x-100)/ 20***	***fu***
50-70	10	60	600	-40	-400	-2	-20
70-90	20	80	1,600	-20	-400	-1	-20
90-110	**25**	**100**	2,500	**0**	0	0	0
110-130	35	120	4,200	20	700	1	35
130-150	9	140	1,260	40	360	2	18
fx and fx	f =99		fx = 10,160		fd = 260		fu = 13

Where N = f = 99

Table 2.3 provides the procedure for calculating the mean for grouped data. In the given frequency distribution, ninety-nine workers have been grouped into five classes of wage rates. The midpoints of these groups are listed in the third column. To find the mean, each midpoint (X) has been multiplied by the frequency (f) and their sum (fx) divided by N.

The mean may be computed as under using the given formula :

$$\bar{X} \quad \frac{fx}{N}$$

$$\frac{10,160}{99}$$

$$= 102.6$$

Indirect Method

The following formula can be used for the indirect method for grouped data. The principles of this formula are similar to that of the indirect method given for ungrouped data. It is expressed as under

$$\bar{x} \quad A \quad \frac{fd}{N}$$

Where,

A = Midpoint of the assumed mean group
(The assumed mean group in *Table 2.3* is 90 – 110 with 100 as midpoint.)

f = Frequency

d = Deviation from the assumed mean group (A)

N = Sum of cases or f

i = Interval width (in this case, it is 20)

From *Table 2.3* the following steps involved in computing mean using the direct method can be deduced :

(i) Mean has been assumed in the group of 90 – 110. It is preferably assumed from the class as near to the middle of the series as possible. This procedure minimises the magnitude of computation. In *Table 2.3*, A (assumed mean) is 100, the midpoint of the class 90 – 110.

(ii) The fifth column (u) lists the deviations of midpoint of each class from the midpoint of the assumed mean group (90 – 110).

(iii) The sixth column shows the multiplied values of each f by its corresponding d to give fd. Then, positive and negative values of fd are added separately and their absolute difference is found (fd). Note that the sign attached to fd is replaced in the formula following A, where ± is given.

The mean using indirect method is computed as under :

$$\bar{x} \quad A \quad \frac{fd}{N}$$

$$= 100 + \frac{260}{99}$$

$$= 100 + 2.6$$

$$= 102.6$$

Note : The Indirect mean method will work for both equal and unequal class intervals.

Median

Median is a **positional average**. It may be defined "as the point in a distribution with an equal number of cases on each side of it". The **Median** is expressed using symbol M.

Computing Median for Ungrouped Data

When the scores are ungrouped, these are arranged in ascending or descending order. Median can be found by locating the central observation or value in the arranged series. The central value may be located from either end of the series arranged in ascending or descending order. The following equation is used to compute the median :

$$\text{Value of } \frac{N \quad 1}{2} \text{ th item}$$

Example 2.3: Calculate median height of mountain peaks in parts of the Himalayas using the following:

8,126 m, 8,611m, 7,817 m, 8,172 m, 8,076 m, 8,848 m, 8,598 m.

Computation : Median (M) may be calculated in the following steps :

(i) Arrange the given data in ascending or descending order.

(ii) Apply the formula for locating the central value in the series. Thus :

$$\text{Value of } \left(\frac{N \quad 1}{2} \right) \text{ th item}$$

$$= \frac{7 \quad 1}{2} \text{ th item}$$

$$\frac{8}{2} \text{ th item}$$

4th item in the arranged series will be the Median.

Arrangement of data in ascending order –

7,817; 8,076; 8,126; 8,172; 8,598; 8,611; 8,848

↓

4th item

Hence,

M = 8,172 m

Computing Median for Grouped Data

When the scores are grouped, we have to find the value of the point where an individual or observation is centrally located in the group. It can be computed using the following formula :

$$M \quad l \quad \frac{i}{f} \quad \frac{N}{2} \quad c$$

Where,

M = Median for grouped data
l = Lower limit of the median class
i = Interval
f = Frequency of the median class
N = Total number of frequencies or number of observations
c = Cumulative frequency of the pre-median class.

Example 2.4 : Calculate the median for the following distribution :

class	50-60	60-70	70-80	80-90	90-100	100-110
f	3	7	11	16	8	5

Table 2.4 : Computation of Median

Class	*Frequency (f)*	*Cumulative Frequency (F)*	*Calculation of Median Class*
50-60	3	3	
60-70	7	10	
70-80	11	*21c*	
80-90 (median group)	***16 f***	**37**	$M=\frac{N}{2}$
90-100	8	45	
100-110	5	50	$=\frac{50}{2}$
	f **or** **N= 50**		= 25

The median is computed in the steps given below :

(i) The frequency table is set up as in *Table 2.4*.

(ii) Cumulative frequencies (**F**) are obtained by adding each normal frequency of the successive interval groups, as given in column 3 of *Table 2.4*.

(iii) Median number is obtained by $\frac{N}{2}$ i.e. $\frac{50}{2}$ = **25** in this case, as shown in column 4 of *Table 2.4*.

(iv) Count into the cumulative frequency distribution (**F**) from the top towards bottom until the value next greater than $\frac{N}{2}$ is reached. In this example, $\frac{N}{2}$ is 25, which falls in the Class interval of 40-44 with cumulative frequency of 37, thus the cumulative frequency of the pre-median class is 21 and actual frequency of the median class is 16.

(v) The median is then computed by substituting all the values determined in the step 4 in the following equation :

$$M \quad l \quad \frac{i}{f}(m \quad c)$$

$$80 \quad \frac{10}{16} (25 - 21)$$

$$80 \quad \frac{5}{8} \quad 4$$

$$80 \quad \frac{5}{2}$$

$$80 \quad 2.5$$

$$M \quad 82.5$$

Mode

The value that occurs most frequently in a distribution is referred to as **mode**. It is symbolised as **Z** or $\mathbf{M_0}$. Mode is a measure that is less widely used compared to mean and median. There can be more than one type mode in a given data set.

Computing Mode for Ungrouped Data

While computing mode from the given data sets all measures are first arranged in ascending or descending order. It helps in identifying the most frequently occurring measure easily.

Example 2.5 : Calculate mode for the following test scores in geography for ten students :

61, 10, 88, 37, 61, 72, 55, 61, 46, 22

Computation : To find the mode the measures are arranged in ascending order as given below:

10, 22, 37, 46, 55, **61, 61, 61**, 72, 88.

The measure 61 occurring three times in the series is the **mode** in the given dataset. As no other number is in the similar way in the dataset, it possesses the property of being **unimodal**.

Example 2.6 : Calculate the mode using a different sample of ten other students, who scored:

82, 11, 57, 82, 08, 11, 82, 95, 41, 11.

Computation : Arrange the given measures in an ascending order as shown below :

08, 11, 11, 11, 41, 57, 82, 82, 82, 95

It can easily be observed that measures of 11 and 82 both are occurring three times in the distribution. The dataset, therefore, is **bimodal** in appearance. If three values have equal and highest frequency, the series is **trimodal**. Similarly, a recurrence of many measures in a series makes it **multimodal**. However, when there is no measure being repeated in a series it is designated as **without mode**.

Comparison of Mean, Median and Mode

The three measures of the **central tendency** could easily be compared with the help of normal distribution curve. The normal curve refers to a frequency distribution in which the graph of scores often called a bell-shaped curve. Many

human traits such as intelligence, personality scores and student achievements have normal distributions. The bell-shaped curve looks the way it does, as it is symmetrical. In other words, most of the observations lie on and around the middle value. As one approaches the extreme values, the number of observations reduces in a symmetrical manner. A normal curve can have high or low data variability. An example of a normal distribution curve is given in *Fig. 2.3*.

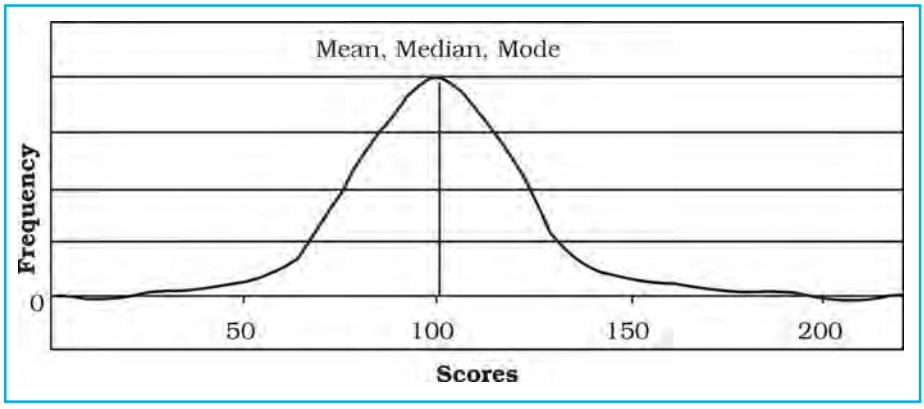


Fig. 2.3 : *Normal Distribution Curve*

The normal distribution has an important characteristic. **The mean, median and mode are the same score** (a score of 100 in *Fig. 2.3*) because a normal distribution is symmetrical. The score with the highest frequency occurs in the middle of the distribution and exactly half of the scores occur above the middle and half of the scores occur below. Most of the scores occur around the middle of the distribution or the mean. Very high and very low scores do not occur frequently and are, therefore, considered rare.

If the data are skewed or distorted in some way, the mean, median and mode will not coincide and the effect of the skewed data needs to be considered (*Fig. 2.4 and 2.5*).

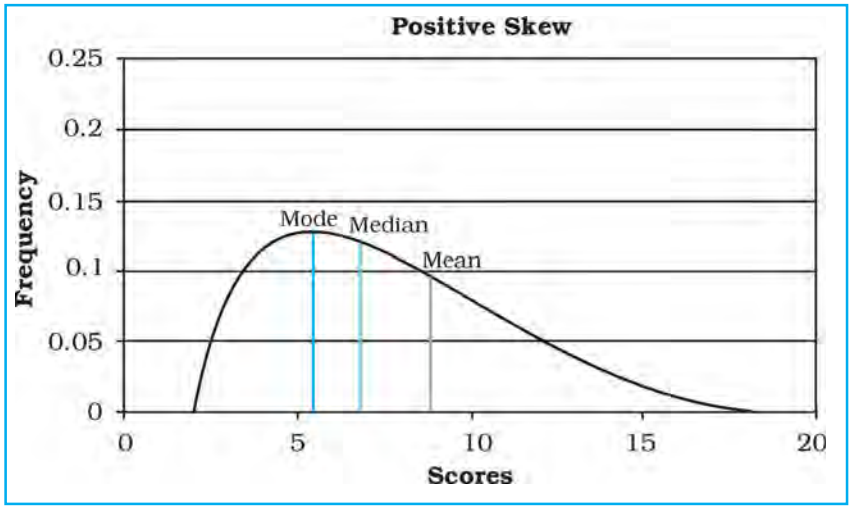


Fig. 2.4 : *Positive Skew*

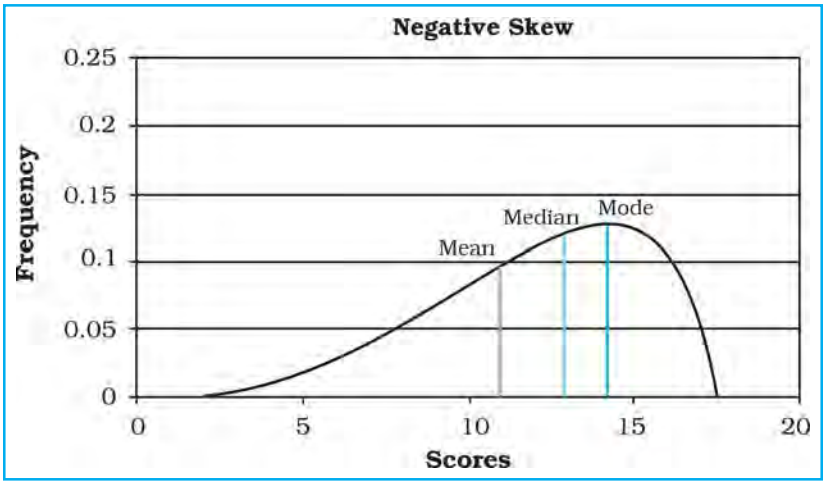


Fig. 2.5 : *Negative Skew*

Measures of Dispersion

The measures of Central tendency alone do not adequately describe a distribution as they simply locate the centre of a distribution and do not tell us anything about how the scores or measurements are scattered in relation to the centre. Let us use the data given in *Table 2.5* and *2.6* to understand the limitations of the measures of central tendency.

Table 2.5 : Scores of Individuals

Individual	*Score*
X1	52
X2	55
X3	50
X4	48
X5	45

Table 2.6 : Scores of Individuals

Individual	*Score*
X1	28
X2	00
X3	98
X4	55
X5	69

$\bar{X}$ = 50 for both the distributions

It can be observed that the mean derived from the two data sets (*Table 2.5* and *2.6*) is same i. e. 50. The highest and the lowest score shown in *Table 2.5* is 55 and 45 respectively. The distribution in *Table 2.6* has a high score of 98 and a low score of zero. The **range** of the first distribution is 10, whereas, it is 98 in the second distribution. Although, the mean for both the groups is the same, the first group is obviously **stable or homogeneous** as compared to the distribution of score of the second group, which is highly **unstable or heterogeneous**. This raises a question whether the mean is a sufficient indicator of the total character of distributions. The examples provide profound evidence that it is not so. Thus, to get a better picture of a distribution, we need to use a measure of **central tendency** and of **dispersion** or **variability.**

The term **dispersion** refers to the scattering of scores about the measure of central tendency. It is used to measure the extent to which individual items or numerical data tend to vary or spread about an average value. Thus, the

dispersion is the degree of spread or scatter or variation of measures about a central value.

The dispersion serves the following two basic purposes :

(i) It gives us the nature of composition of a series or distribution, and

(ii) It permits comparison of the given distributions in terms of stability or homogeneity.

Methods of Measuring Dispersion

The following methods are used as measures of dispersion :

1. Range
2. Quartile Deviation
3. Mean Deviation
4. Standard Deviation and Co-efficient of Variation (C.V.)
5. Lorenz Curve

Each of these methods has definite advantages as well as limitations. Hence, there is a need to use either of the methods with great precautions. The Standard Deviation (s) as an absolute measure of dispersion and Co-efficient of Variation (**cv**) as a relative measure of dispersion, besides the Range are most commonly used measures of dispersion. We will discuss how each one of these measures is computed.

Range

Range (R) is the difference between maximum and minimum values in a series of distribution. This way it simply represents the distance from the smallest to the largest score in a series. It can also be defined as the highest score minus the lowest score.

Range for Ungrouped Data

Example 2.7 : Calculate the **range** for the following distribution of daily wages:

Rs. 40, 42, 45, 48, 50, 52, 55, 58, 60, 100.

Computation of Range

The **R** can be calculated with the help of the following formula :

$$R \quad L \quad S$$

Where

'R' is Range,

'L' and 'S' is the largest and smallest values respectively in a series.

Hence,

$$\mathbf{R} = L - S$$

$$= 100 - 40 = \mathbf{60}$$

If we eliminate the 10th case, **R** becomes 20 (60 – 40). The elimination of one score has reduced the **R** to just one-third. It is obvious that the difficulty with **R** as a measure of variability is that its value is wholly dependent upon the two extreme scores. Thus, as a measure of dispersion **R** functions much the same way as **mode** does as a measure of central tendency. Both the measures are **highly unstable**.

Standard Deviation

Standard deviation (SD) is the most widely used measure of dispersion. It is defined as the square root of the average of squares of deviations. It is always calculated around the mean. The standard deviation is the most stable measure of variability and is used in so many other statistical operations. The Greek character denotes it.

To obtain SD, deviation of each score from the mean **(*x*)** is first squared (***x*²**). It is important to note that this step makes all negative signs of deviations positive. It saves SD from the major criticism of mean deviation which uses modulus ***x***. Then, all of the squared deviations are summed - ***x*²** (care should be taken that these are **not** summed first and then squared). This sum of the squared deviations **(*x*²)** is divided by the number of cases and then the square root is taken. Therefore, **Standard Deviation is defined as the root mean square deviation**. For a given data set, it is computed using the following formula :

$$\sqrt{\frac{x^2}{N}}$$

During these steps, we come across a term before taking its square root. It is assigned a special name, the **variance**. The variance is widely used in advanced statistical operations. Its square root is standard deviation. That way, the opposite is also true i.e. square of SD is variance.

Standard Deviation for Ungrouped Data

Example 2.8 : Calculate the standard deviation for the following scores :

01, 03, 05, 07, 09

$$\sqrt{\frac{x^2}{N}}$$

$$\sqrt{\frac{40}{5}}$$

$\sqrt{8}$ 2.828

2.83

Table 2.7 : Computation of Standard Deviation

X	$x(X \quad \overline{X})$	x^2
1	–4	16
3	–2/–6	4
5	0	0
7	2	4
9	6-Apr	16
X 25		
N =5		
∴ = 5		

Let us summarise the steps used in the above computation :

(i) All the scores have been placed in the column marked **X**.

(ii) Summing the raw scores and dividing by N have found mean.

(iii) Deviation of each raw score **(*x*)** has been obtained by subtracting the mean from them. A check on our work is that the sum of the ***x*** should be zero. We find that this is true for our exercise.

(iv) Each value of ***x*** has been squared and summed.

(v) Sum of the **x²s** has been divided by N. Recall that the resultant is the variance.

(vi) Its square root has been found to obtain Standard Deviation.

Computation of Standard Deviation for Grouped Data

Example : Calculate the standard deviation for the following distribution:

Groups	120-130	130-140	140-150	150-160	160-170	170-180
f	2	4	6	12	10	6

The method of obtaining SD for grouped data has been explained in the table below. The initial steps upto column 4, are the same as those we followed in the computation of the mean for grouped data. We begin with assuming our mean to exist in the interval group of 150-160, hence a deviation value of zero has been assigned to the group. Likewise other deviations are determined. Values in column 4 (fx') are obtained by the multiplication of the values in the two previous columns. Values in column 5 (fx'^2) are obtained by multiplying the values given in column 3 and 4. Then various columns have been summed.

(1) ***Group***	(2) ***f***	(3) ***x'***	(4) ***fx'***	(5) ***fx'2***
120 - 130	2	–3	–6	18
130 - 140	4	–2	–8	16
140 - 150	6	–1	$\frac{6}{20}$	6
150 - 160	12	0	0	0
160 - 170	10	1	10	10
170 - 180	6	2	$\frac{12}{22}$	24
	N=40		fx' 2	fx'^2 =74

The following formula is used to calculate the Standard Deviation :

$$SD \quad i^2| \quad fx'^2 \quad \frac{fx'}{N}$$

Coefficient of Variation (CV)

When the observations for different places or periods are expressed in different units of measurement and are to be compared, the coefficient of variation (CV) proves very useful. **CV expresses the standard deviation as a percentage of the mean.** It is determined using the following formula :

$$\text{CV} \quad \frac{\text{Standard Deviation}}{\text{Mean}} \quad 100$$

$$CV = \frac{\sigma}{\bar{X}} 100$$

The CV for the dataset given in *Table 2.7* will, hence, be as under :

$$CV = \frac{\sigma}{\bar{X}} 100$$

$$CV = \frac{2.83}{5} 100$$

CV = **56%**

Coefficient of Variation for grouped data can also be calculated using the same formula.

Rank Correlation

The statistical methods discussed so far were concerned with the analysis of a single variable. We will now discuss the methods of exploring relationship between two variables and the way this relationship is expressed numerically. When dealing with two or more sets of data, curiosity arises for knowing whether or not changes in one variable produce changes in some other variable.

Often our interest lies in knowing the nature of relationship or interdependence between two or more sets of data. It has been found that the **correlation** serves useful purpose. It is basically a measure of relationship between two or more sets of data. Since, we study the way they vary, we call these events **variables**. Thus, the term **correlation refers to the nature and strength of correspondence or relationship between two variables**. The terms **nature** and **strength** in the definition refer to the **direction** and **degree** of the variables with which they co-vary.

Direction of Correlation

It is our common experience that an input is made to get some output. There could be three possibilities.

1. With the increase in input the output also increases.
2. With the increase in the input the output decreases.
3. Change in the input does not lead to change in the output.

In the first case, the direction of the relationship between the input and output is in the same direction. It is called that both are positively correlated.

In the second case the direction of change between the input and output is in the opposite direction and it is called that they are negatively correlated.

In the third case, change in the input has no relationship with the output, hence, it is said that these do not have a statistically significant relationship.

Let us now consider *Fig. 2.7* which looks just opposite of *Fig. 2.6*. The plotted values run from the upper left to the lower right of the graph. Notice that for every increase of one unit on the X-axis, there is a corresponding decrease of two units on the Y-axis. It is an example of a **negative correlation**. It means that the two variables have a tendency to move opposite to each other, i.e. if one variable increases, the other decreases and vice versa. We can find such relationships existing between various geographical pairs of variables. Correlations between

height above sea level and air pressure, temperature and air pressure are few examples. It implies that the obtained figure of correlation must precede with the arithmetical sign (plus or minus), more importantly in the negative correlation.

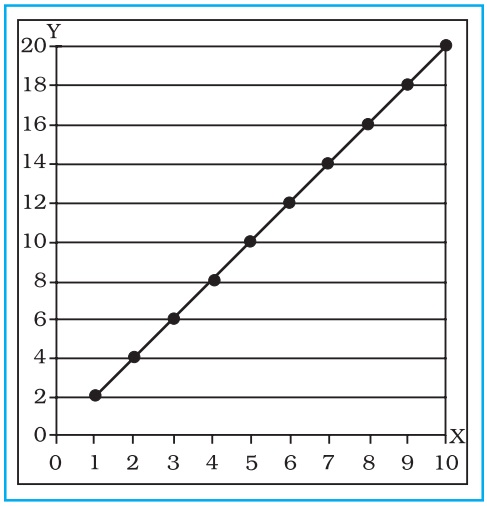


Fig. 2.6 : *Perfect Positive Correlation*

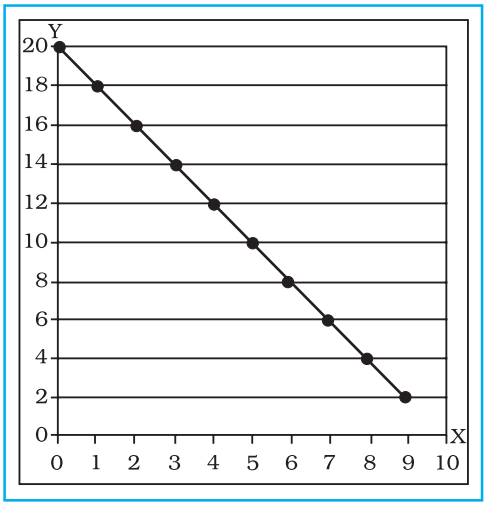


Fig. 2.7 : *Perfect Negative Correlation*

Degree of Correlation

When reference has been made about the direction of correlation, negative or positive, a natural curiosity arises to know the degree of correspondence or association of the two variables. The maximum degree of correspondence or relationship goes upto **1 (one)** in mathematical terms. On adding an element of the direction of correlation, it spreads to the **maximum extent of –1 to +1 through zero**. It can never be more than one. The spread can also be translated into linear shape, as shown in the *Fig. 2.8*. Correlation of **1** is known as **perfect correlation** (whether positive and negative). Between the two points of divergent, perfect correlations lies **0 (zero) correlation,** a point of **no correlation** or absence of any correlation between the variables.

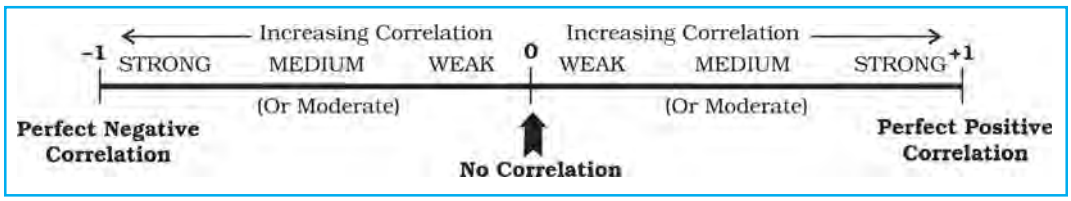


Fig. 2.8 : *Spread of Direction and Degree of Correlation*

Perfect Correlations

Figs. 2.6 and *2.7* have been constructed to show the typical relationship between two variables. Notice that these graphs show the scattering of X – Y values. Therefore, such graphs are referred to as **scatter gram** or **scatter plot**. It may be noted from *Fig. 2.6*, that the pairs of values like these, when plotted, fall along a straight line and when this straight line runs from the lower left of the scatter plot to the upper right, it is an example of a **perfect positive correlation (1.00)**.

Fig. 2.7 is just opposite of this. All the points again fall along a straight line which now runs from the upper left-hand part of the scatter gram to its lower right. It is an example of a **perfect negative correlation** (with a value of – 1.00). **No Correlation** (or Zero Correlation) is one when any of the variables in the pair does not respond to the changes in the other, the correlation will come to zero. This is the state of **no correlation or zero correlation**. This is shown in *Fig. 2.9*. Seatter plot A shows no correlation when Y does not respond to changes in X. Similarly, zero correlation occurs in Seatter plot B when X does not respond to changes in Y.

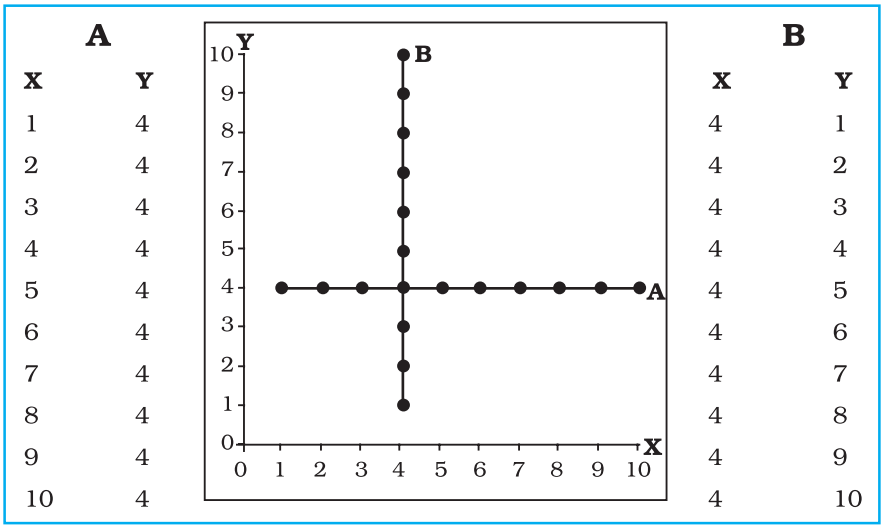


Fig. 2.9 : *Scatter plot showing No Correlation*

Other Correlations

Between the perfect correlations (±1) and zero correlation lies generalised conditions popularly referred to as weak, moderate and strong correlations. These conditions are clearly exhibited in *Figs. 2.10, 2.11* and *2.12* respectively. Notice the spreading or the scattering of the plotted points and the assignment of the terms weak, medium and strong to them (generalised terms having no specific limits). Larger is the scattering, weaker is the correlation. Smaller is the scattering, stronger is the correlation, and when the plotted points fall on a straight line, the correlation is perfect (*Fig. 2.6 and 2.7*).

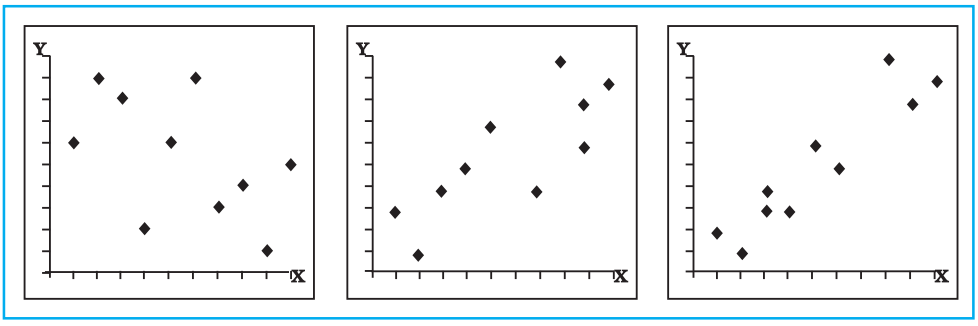


Fig. 2.10 : *Weak Negative Correlation*

Fig. 2.11 : *Moderate Positive Correlation*

Fig. 2.12 : *Strong Positive Correlation*

Methods of Calculating Correlation

There are various methods by which correlation can be calculated. However, under the constraints of time and space, we will discuss the Spearman's Rank Correlation method only.

Spearman's Rank Correlation

Spearman devised a method of computing correlation with the help of ranks. The method is popularly known as Spearman's Rank Correlation symbolised as ρ (the Greek letter **rho**). Spearman's Rank Correlation method is widely used. The computation of the correlation is undertaken in the steps given below:

(i) Copy the data related to X-Y variables given in the exercise and put them in the first and second columns of the table.

(ii) Both the variables are to be ranked separately. The ranks of X-variable are to be recorded in column 3 headed by **XR** (ranks of X). Similarly, the ranks of Y-variable **(YR)** are to be recorded in the fourth column. The highest value in the data is to be awarded rank one, second highest rank two and so on. Suppose the data for X-variable are 4, 8, 2, 10, 1, 9, 7, 3, 0 and 5, the **XR** will be 6, 3, 8, 1, 9, 2, 4, 7, 10 and 5 respectively. Notice that the last rank (10 in this case) equals the number of observations. Assignment of **YR** is also done in the same way.

(iii) Now since both **XR** and **YR** have been obtained, find the difference between the two sets of ranks (disregarding the sign plus or minus) and record it in the fifth column. **The sign of the difference is of no importance**, since, these differences are squared in the next operation.

(iv) Each of these differences is squared and sum of this column of squares is obtained. These values are placed in the sixth column.

(v) Then the computation of the rank correlation is done by the application of the following equation:

$$\sum D_1^2 \quad \frac{6 \quad D^2}{N(N^2 \quad 1)}$$

Where,

ρ = rank correlation

= sum of the squares of the differences between two sets of ranks

N = the number of pairs of X-Y

Example 2.9: Calculate Spearman's Rank Correlation with the help of the following data :

Scores in Economics (X) :	02	08	00	20	12	16	06	18	09	10
Scores in Geography (Y) :	04	12	06	24	16	18	08	20	09	10

Table 2.8 : Computation of Spearman's Rank Correlation

(1) *X*	(2) *Y*	(3) *XR*	(4) *YR*	(5) *D*	(6) D^2
2	4	9	10	1	1
8	12	7	5	2	4
0	6	10	9	1	1
20	24	1	1	0	0
12	16	4	4	0	0
16	18	3	3	0	0
6	8	8	8	0	0
18	20	2	2	0	0
9	9	6	7	1	1
10	10	5	6	1	1
N=10					D^2=8

Calculation:

Where, ρ is Rank Correlation; **D** is difference between the rank of X and Y; and **N** is number of items of x – y

$$1 \quad \frac{6 \quad 8}{10 \quad 10^2 \quad 1}$$

$$1 \quad \frac{48}{10 \quad 100 \quad 1}$$

$$1 \quad \frac{48}{10 \quad 99}$$

$$1 \quad \frac{48}{990}$$

$$1 \quad 0.05$$

$$0.95$$

In rho, we obtain a correlation, which makes a good substitute for other types of correlations, when the number of cases is small. It is almost useless when N is large, because by the time all the data are ranked, other type of correlation could have been calculated.

Excercises

1. Choose the correct answer from the four alternatives given below:

(i) The measure of central tendency that does not get affected by extreme values:

(a) Mean (b) Mean and Mode

(c) Mode (d) Median

(ii) The measure of central tendency always coinciding with the hump of any distribution is:

(a) Median (b) Median and Mode

(c) Mean (d) Mode

(iii) A scatter plot represents negative correlation if the plotted values run from:

(a) Upper left to lower right (b) Lower left to upper right

(c) Left to right (d) Upper right to lower left

2. Answer the following questions in about 30 words:

(i) Define the mean.

(ii) What are the advantages of using mode ?

(iii) What is dispersion ?

(iv) Define correlation.

(v) What is perfect correlation ?

(vi) What is the maximum extent of correlation?

3. Answer the following questions in about 125 words:

(i) Explain relative positions of mean, median and mode in a normal distribution and skewed distribution with the help of diagrams.

(ii) Comment on the applicability of mean, median and mode (*hint: from their merits and demerits).*

(iii) Explain the process of computing Standard Deviation with the help of an imaginary example.

(iv) Which measure of dispersion is the most unstable statistic and why?

(v) Write a detailed note on the degree of correlation.

(vi) What are various steps for the calculation of rank order correlation?

Activity

1. Take an imaginary example applicable to geographical analysis and explain direct and indirect methods of calculating mean from ungrouped data.

2. Draw scatter plots showing different types of perfect correlations.

3 Graphical Representation of Data

You must have seen graphs, diagrams and maps showing different types of data. For example, the thematic maps shown in Chapter 1 of book for Class XI entitled *Practical Work in Geography, Part-I (NCERT, 2006)* depict relief and slope, climatic conditions, distribution of rocks and minerals, soils, population, industries, general land use and cropping pattern in the Nagpur district, Maharashtra. These maps have been drawn using large volume of related data collected, compiled and processed. Have you ever thought what would have happened if the same information would have been either in tabular form or in a descriptive transcript? Perhaps, it would not have been possible from such a medium of communication to draw visual impressions which we get through these maps. Besides, it would also have been a time consuming task to draw inferences about whatever is being presented in non–graphical form. Hence, the graphs, diagrams and maps enhance our capabilities to make meaningful comparisons between the phenomena represented, save our time and present a simplified view of the characteristics represented. In the present chapter, we will discuss methods of constructing different types of graphs, diagrams and maps.

Representation of Data

The data describe the properties of the phenomena they represent. They are collected from a variety of sources (Chapter 1). The geographers, economists, resource scientists and the decision makers use a lot of data these days. Besides the tabular form, the data may also be presented in some graphic or diagrammatic form. The transformation of data through visual methods like graphs, diagrams, maps and charts is called representation of data. Such a form of the presentation of data makes it easy to understand the patterns of population growth, distribution and the density, sex ratio, age–sex composition, occupational structure, etc. within a geographical territory. There is a Chinese proverb that *'a picture is equivalent to thousands of words'*. Hence, the graphic method of the representation of data enhances our understanding, and makes the comparisons easy. Besides, such methods create an imprint on mind for a longer time.

General Rules for Drawing Graphs, Diagrams and Maps

1. Selection of a Suitable Method

Data represent various themes such as temperature, rainfall, growth and distribution of the population, production, distribution and trade of different commodities, etc. These characteristics of the data need to be suitably represented by an appropriate graphical method. For example, data related to the temperature or growth of population between different periods in time and for different countries/states may best be represented using line graphs. Similarly, bar diagrams are suited best for showing rainfall or the production of commodities. The population distribution, both human and livestock, or the distribution of the crop producing areas may suitably be represented on dot maps and the population density using choropleth maps.

2. Selection of Suitable Scale

The scale is used as measure of the data for representation over diagrams and maps. Hence, the selection of suitable scale for the given data sets should be carefully made and must take into consideration entire data that is to be represented. The scale should neither be too large nor too small.

3. Design

We know that the design is an important cartographic task (Refer 'Essentials of Map Making' as discussed in Chapter 1 of the *Practical Work in Geography, Part-I (NCERT, 2006),* a textbook of Class XI). The following components of the cartographic designs are important. Hence, these should be carefully shown on the final diagram/map.

Title

The title of the diagram/map indicates the name of the area, reference year of the data used and the caption of the diagram. These components are represented using letters and numbers of different font sizes and thickness. Besides, their placing also matters. Normally, title, subtitle and the corresponding year are shown in the centre at the top of the map/diagram.

Legend

A legend or index is an important component of any diagram/map. It explains the colours, shades, symbols and signs used in the map and diagram. It should also be carefully drawn and must correspond to the contents of the map/diagram. Besides, it also needs to be properly positioned. Normally, a legend is shown either at the lower left or lower right side of the map sheet.

Direction

The maps, being a representation of the part of the earth's surface, need be oriented to the directions. Hence, the direction symbol, i. e. North, should also be drawn and properly placed on the final map.

Construction of Diagrams

The data possess measurable characteristics such as length, width and volume. The diagrams and the maps that are drawn to represent these data related characteristics may be grouped into the following types:

(i) One-dimensional diagrams such as line graph, poly graph, bar diagram, histogram, age, sex, pyramid, etc.;

(ii) Two-dimensional diagram such as pie diagram and rectangular diagram;

(iii) Three-dimensional diagrams such as cube and spherical diagrams.

It would not be possible to discuss the methods of construction of these many types of diagrams and maps primarily due to the time constraint. We will, therefore, describe the most commonly drawn diagrams and maps and the way they are constructed. These are :

- Line graphs
- Pie diagram
- Bar diagrams
- Wind rose and star diagram
- Flow Charts

Line Graph

The line graphs are usually drawn to represent the time series data related to the temperature, rainfall, population growth, birth rates and the death rates. *Table 3.1* provides the data used for the construction of Fig 3.2.

Construction of a Line Graph

(a) Simplify the data by converting it into round numbers such as the growth rate of population as shown in *Table 3.1* for the years 1961 and 1981 may be rounded to 2.0 and 2.2 respectively.

(b) Draw X and Y-axis. Mark the time series variables (years/months) on the X axis and the data quantity/value to be plotted (growth of population in per cent or the temperature in ^{0}C) on Y axis.

(c) Choose an appropriate scale and label it on Y-axis. If the data involves a negative figure then the selected scale should also show it as shown in *Fig. 3.1.*

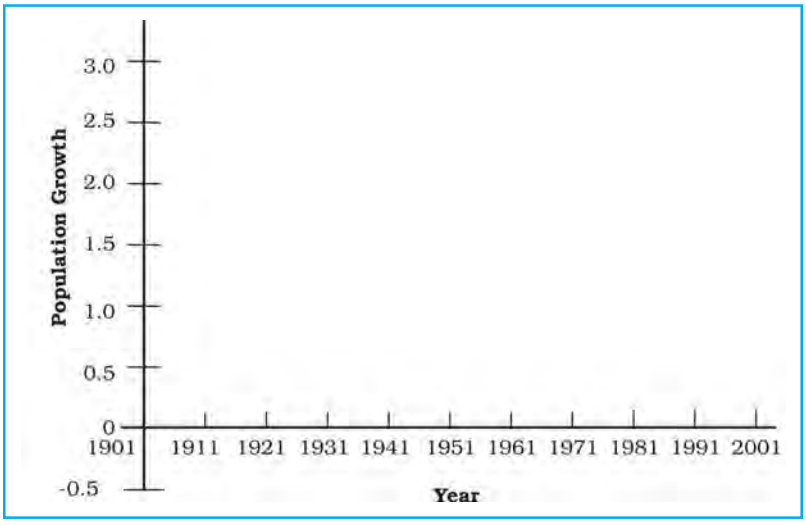


Fig. 3.1 : *Construction of a Line Graph*

(d) Plot the data to depict year/month-wise values according to the selected scale on Y-axis, mark the location of the plotted values by a dot and join these dots by a free hand drawn line.

Example 3.1 : Construct a line graph to represent the data as given in *Table 3.1*:

Table 3.1 : Growth rate of Population in India - 1901 to 2001

Year	*Growth rate in percentage*
1901	-
1911	0.56
1921	-0.30
1931	1.04
1941	1.33
1951	1.25
1961	1.96
1971	2.20
1981	2.22
1991	2.14
2001	1.93

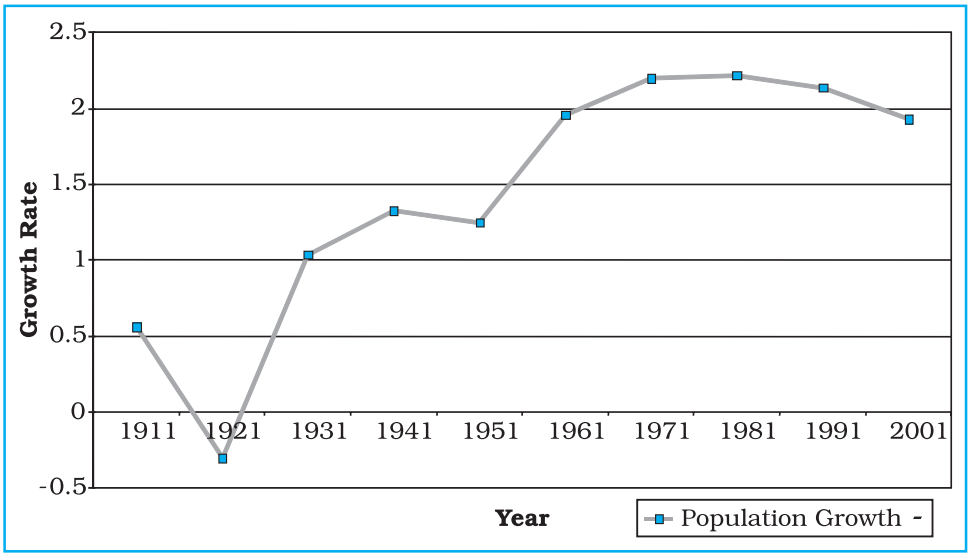


Fig. 3.2 : *Annual Growth of Population in India 1901-2001*

Activity

Find out the reasons for sudden change in population between 1911 and 1921as shown in *Fig. 3.2*.

Polygraph

Polygraph is a line graph in which two or more than two variables are shown by an equal number of lines for an immediate comparison, such as the growth rate of different crops like rice, wheat, pulses or the birth rates, death rates and life expectancy or sex ratio in different states or countries. A different line pattern such as straight line (____), broken line (- - -), dotted line (.....) or a combination of dotted and broken line (-.-.-) or line of different colours may be used to indicate the value of different variables (*Fig 3.3*).

Example 3.2 : Construct a polygraph to compare the growth of sex-ratio in different states as given in the *Table 3.2* :

Table 3.2 : Sex-Ratio (Female per 1000 male) of Selected Sates – 1961-2001

States/UT	*1961*	*1971*	*1981*	*1991*	*2001*
Delhi	785	801	808	827	821
Haryana	868	867	870	**86**	846
Uttar Pradesh	907	876	882	876	898

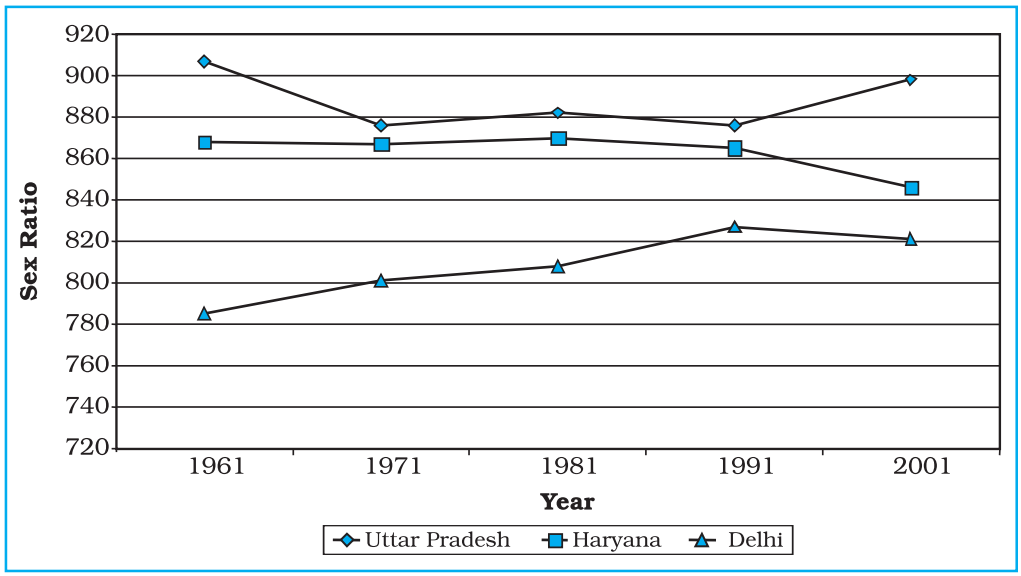


Fig. 3.3 : *Sex-Ratio of Selected States 1961-2001*

Bar Diagram

The bar diagrams are drawn through columns of equal width. It is also called a columnar diagram. Following rules should be observed while constructing a bar diagram:

(a) The width of all the bars or columns should be similar.

(b) All the bars should be placed on equal intervals/distance.

(c) Bars may be shaded with colours or patterns to make them distinct and attractive.

The simple, compound or polybar diagram may be constructed to suit the data characteristics.

Simple Bar Diagram

A simple bar diagram is constructed for an immediate comparison. It is advisable to arrange the given data set in an ascending or descending order and plot the data variables accordingly. However, time series data are represented according to the sequencing of the time period.

Example 3.3 : Construct a simple bar diagram to represent the rainfall data of Tiruvanantapuram as given in *Table 3.3* :

Table 3.3 : Average Monthly Rainfall of Tiruvanantapuram

Months	J	F	M	A	M	J	J	A	S	O	N	D
Rainfall in cm	2.3	2.1	3.7	10.6	20.8	35.6	22.3	14.6	13.8	27.3	20.6	7.5

Construction

Draw X and Y-axes on a graph paper. Take an interval of 5 cm and mark it on Y-axis to plot rainfall data in cm. Divide X-axis into 12 equal parts to represent 12 months. The actual rainfall values for each month will be plotted according to the selected scale as shown in *Fig. 3.4*.

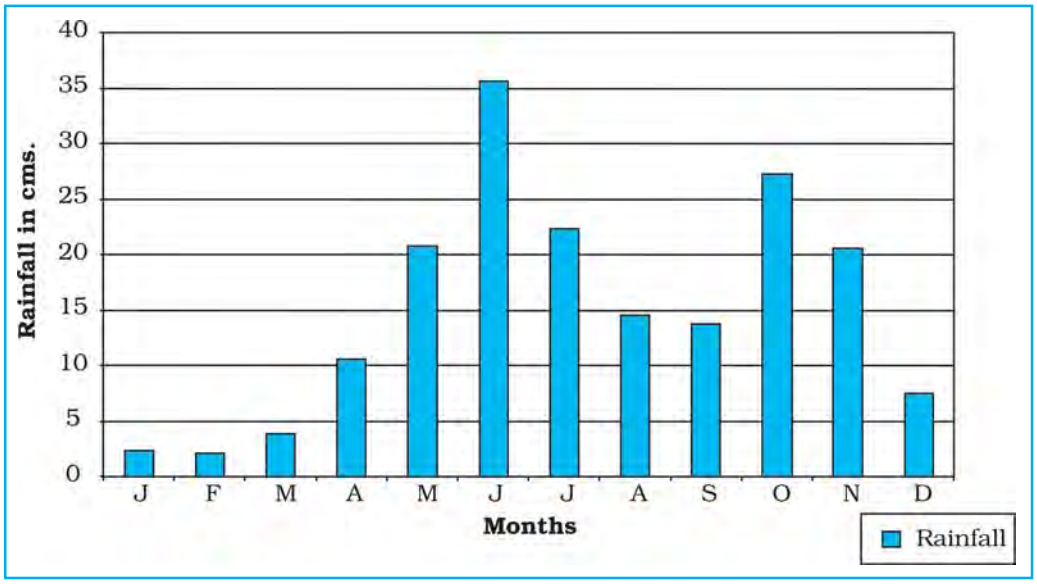


Fig. 3.4 : *Average Monthly Rainfall of Tiruvanantapuram*

Line and Bar Graph

The line and bar graphs as drawn separately may also be combined to depict the data related to some of the closely associated characteristics such as the climatic data of mean monthly temperatures and rainfall. In doing so, a single diagram is drawn in which months are represented on X-axis while temperature and rainfall data are shown on Y-axis at both sides of the diagram.

Example 3.4 : Construct a line graph and bar diagram to represent the average monthly rainfall and temperature data of Delhi as given in *Table 3.4* :

Table 3.4 : Average monthly Temperature and Rainfall in Delhi

Months	***Temp. in C***	***Rainfall in cm.***
Jan.	14.4	2.5
Feb.	16.7	1.5
Mar.	23.30	1.3
Apr.	30.0	1.0
May	33.3	1.8
June	33.3	7.4
Jul.	30.0	19.3
Aug.	29.4	17.8
Sep.	28.9	11.9
Oct.	25.6	1.3
Nov.	19.4	0.2
Dec.	15.6	1.0

Construction

(a) Draw X and Y-axes of a suitable length and divide X-axis into 12 parts to show months in a year.

(b) Select a suitable scale with equal intervals of 5 C or 10 C for temperature data on the Y-axis and label it at its right side.

(c) Similarly, select a suitable scale with equal intervals of 5 cm or 10 cm for rainfall data on the Y-axis and label at its left side.

(d) Plot temperature data using line graph and the rainfall by columnar diagram as shown in *Fig. 3.5*.

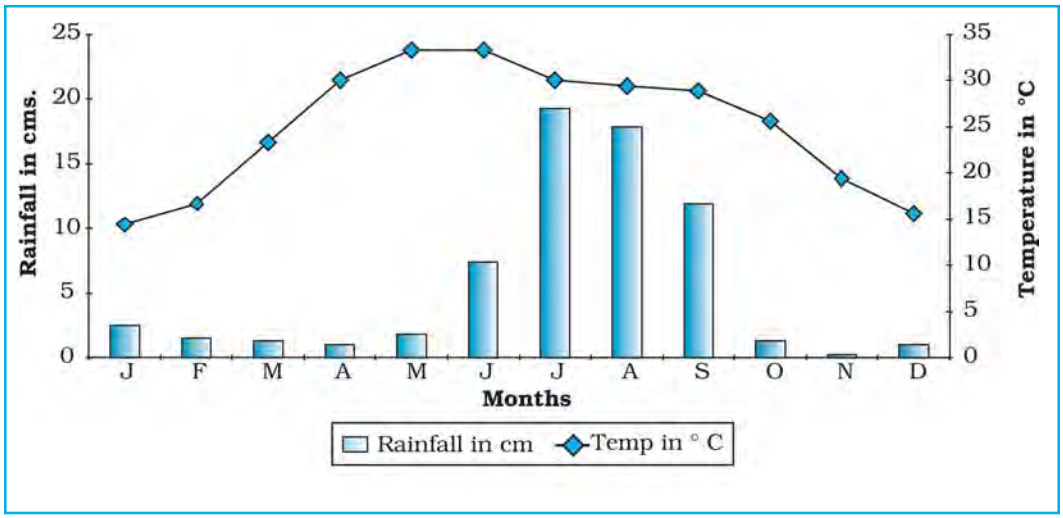


Fig. 3.5 : *Temperature and Rainfall in Delhi*

Multiple Bar Diagram

Multiple bar diagrams are constructed to represent two or more than two variables for the purpose of comparison. For example, a multiple bar diagram may be constructed to show proportion of males and females in the total, rural and urban population or the share of canal, tube well and well irrigation in the total irrigated area in different states.

Example 3.5 : Construct a suitable bar diagram to show decadal literacy rate in India during 1951 – 2001 as given in *Table 3.5* :

Table 3.5 : Literacy Rate in India, 1951-2001 (in %)

Year	Literacy Rate		
	Total population	Male	Female
1951	18.33	27.16	8.86
1961	28.3	40.4	15.35
1971	34.45	45.96	21.97
1981	43.57	56.38	29.76
1991	52.21	64.13	39.29
2001	64.84	75.85	54.16

Construction

(a) Multiple bar diagram may be chosen to represent the above data.

(b) Mark time series data on X-axis and literacy rates on Y-axis as per the selected scale.

(c) Plot the per cent of total population, male and female in closed columns (*Fig 3.6*).

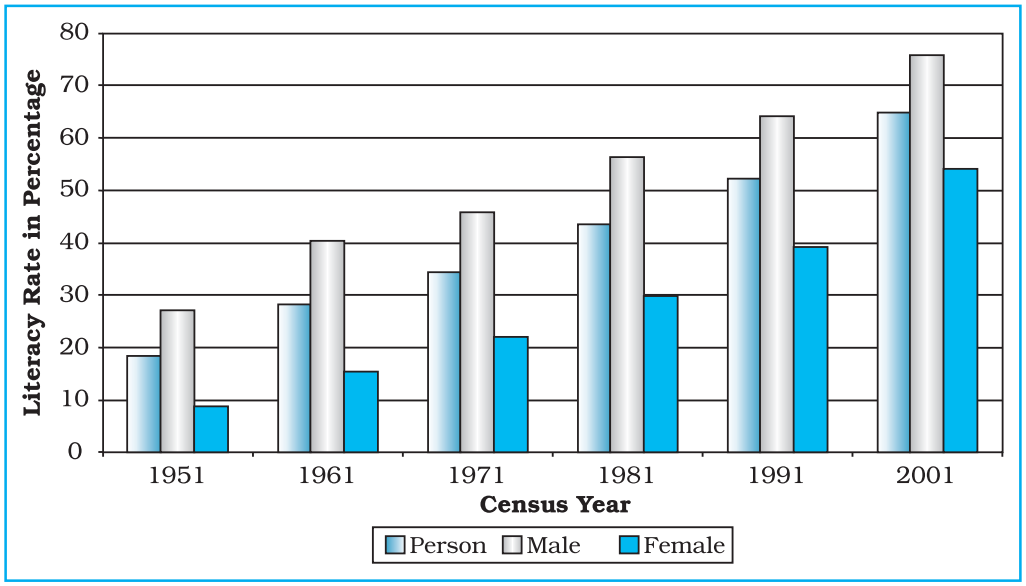


Fig. 3.6 : *Literacy Rate, 1951-2001*

Compound Bar Diagram

When different components are grouped in one set of variable or different variables of one component are put together, their representation is made by a compound bar diagram. In this method, different variables are shown in a single bar with different rectangles.

Example 3.6 : Construct a compound bar diagram to depict the data as shown in *Table 3.6* :

Table 3.6 : Gross Generation of Electricity in India (in Billion KWh)

Year	*Thermal*	*Hydro*	*Nuclear*	*Total*
2001-02	424.4	73.5	19.5	517.4
2002-03	451.0	63.8	19.2	534.0
2003-04	472.1	75.2	17.8	565.1

Construction

(a) Arrange the data in ascending or descending order.

(b) A single bar will depict the gross electricity generation in the given year and the generation of thermal, hydro and nuclear electricity be shown by dividing the total length of the bar as shown in *Fig 3.7*.

Pie Diagram

Pie diagram is another graphical method of the representation of data. It is drawn to depict the total value of the given attribute using a circle. Dividing the circle into corresponding degrees of angle then represent the sub–sets of the data. Hence, it is also called as **Divided Circle Diagram**.

The angle of each variable is calculated using the following formulae.

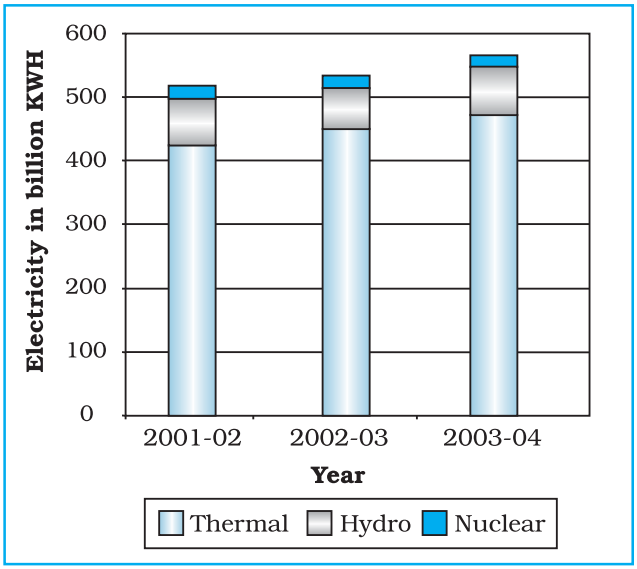


Fig. 3.7 : *Gross Electricity Generation in India*

$$\frac{\text{Value of given State/Region X 360}}{\text{Total Value of All States/Regions}}$$

If data is given in percentage form, the angles are calculated using the given formulae.

$$\frac{\text{Percentage of x X 360}}{100}$$

For example, a pie diagram may be drawn to show total population of India along with the proportion of the rural and urban population. In this case the circle of an appropriate radius is drawn to represent the total population and its sub-divisions into rural and urban population are shown by corresponding degrees of angle.

Example 3.7: Represent the data as given in *Table 3.7 (a)* with a suitable diagram.

Table 3.7 (a) : India's Export to Major Regions of the World in 2004-05

Unit/Region	*% of Indian Export*
European Union	21.1
North America	19.2
Australia	3.7
OPEC	15
African Countries	3.3
Asian Countries	27.6
Others	10.1
Total	100

Calculation of Angles

(a) Arrange the data on percentages of Indian exports in an ascending order.

(b) Calculate the degrees of angles for showing the given values of India's export to major regions/countries of the world, *Table 3.7* (b). It could be done by multiplying percentage with a constant of 3.6 as derived by dividing the total number of degrees in a circle by 100, i. e. 360/100.

(c) Plot the data by dividing the circle into the required number of divisions to show the share of India's export to different regions/countries (*Fig. 3.8*).

Table 3.7 (b) : India's Export to Major Regions of the World in 2004-05

Countries	*%*	*Calculation*	*Degree*
African Countries	3.3	3.3 x 3.6 = 11.88	12º
Australia	3.7	3.7 x 3.6 = 13.32	13º
others	10.1	10.1 x 3.6 = 36.36	36º
OPEC Countries	15	15 x 3.6 = 54	54º
North America	19.2	19.2 x 3.6 = 69.12	69º
European Union	21.1	21.1 x 3.6 = 75.96	76º
Asian Countries	27.6	27.6 x 3.6 = 99.36	100º
Total	100		360º

Construction

(a) Select a suitable radius for the circle to be drawn. A radius of 3, 4 or 5 cm may be chosen for the given data set.

(b) Draw a line from the centre of the circle to the arc as a radius.

(c) Measure the angles from the arc of the circle for each category of vehicles in an ascending order clock-wise, starting with smaller angle.

(d) Complete the diagram by adding the title, sub – title, and the legend. The legend mark be chosen for each variable/category and highlighted by distinct shades/ colours.

Precautions

(a) The circle should neither be too big to fit in the space nor too small to be illegible.

(b) Starting with bigger angle will lead to accumulation of error leading to the plot of the smaller angle difficult.

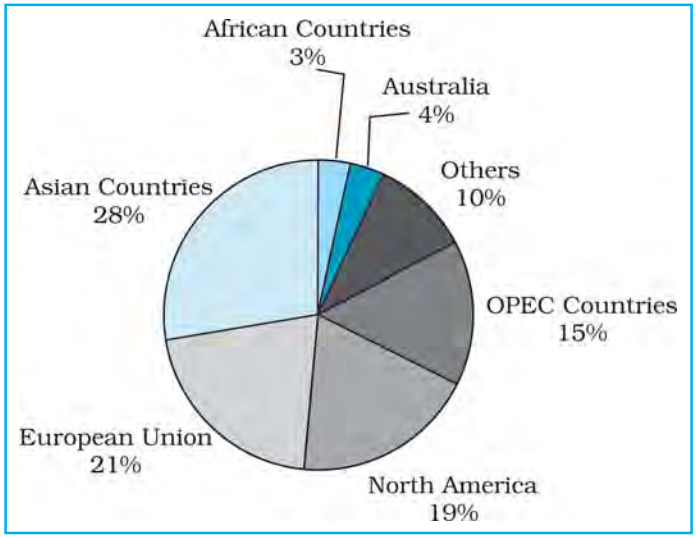


Fig. 3.8 : *Direction of Indian Exports 2004-05*

Flow Maps/Chart

Flow chart is a combination of graph and map. It is drawn to show the flow of commodities or people between the places of origin and destination. It is also called as **Dynamic Map**. Transport map, which shows number of passengers, vehicles, etc., is the best example of a flow chart. These charts are drawn using lines of proportional width. Many government agencies prepare flow maps to show density of the means of transportation on different routes. The flow maps/ charts are generally drawn to represent two the types of data as given below:

1. The number and frequency of the vehicles as per the direction of their movement
2. The number of the passengers and/or the quantity of goods transported.

Requirements for the Preparation of a Flow Map

(a) A route map depicting the desired transport routes along with the connecting stations.

(b) The data pertaining to the flow of goods, services, number of vehicles, etc., along with the point of origin and destination of the movements.

(c) The selection of a scale through which the data related to the quantity of passengers and goods or the number of vehicles is to be represented.

Example 3.10 : Construct a flow map to represent the number of trains running in Delhi and the adjoining areas as given in the *Table 3.8*.

Table 3.8 : No. of trains of selected routes of Delhi and adjoining areas

S. No.	Railway Routes	No. of Trains
1.	Old Delhi – New Delhi	50
2.	New Delhi-Nizamuddin	40
3.	Nizamuddin-Badarpur	30
4.	Nizamuddin-Sarojini Nagar	12
5.	Sarojini Nagar – Pusa Road	8
6.	Old Delhi – Sadar Bazar	32
7.	Udyog Nagar-Tikri Kalan	6
8.	Pusa Road – Pehladpur	15
9.	Sahibabad-Mohan Nagar	18
10.	Old Delhi – Silampur	33
11.	Silampur – Nand Nagari	12
12.	Silampur-Mohan Nagar	21
13.	Old Delhi-Shalimar Bagh	16
14.	Sadar Bazar-Udyog Nagar	18
15.	Old Delhi – Pusa Road	22
16.	Pehladpur – Palam Vihar	12

Construction

(a) Take an outline map of Delhi and adjoining areas in which railway line and the nodal stations are depicted (*Fig.3.9*).

(b) Select a scale to represent the number of trains. Here, the maximum number is 50 and the minimum is 6. If we select a scale of 1cm = 50 trains, the maximum and minimum numbers will be represented by a strip of 10 mm and 1.2 mm thick lines respectively on the map.

(c) Plot the thickness of each strip of route between the given rail route (*Fig. 3.10*).

(d) Draw a terraced scale as legend and choose distinct sign or symbol to show the nodal points (stations) within the strip.

Fig. 3.9 : *Map of Delhi*

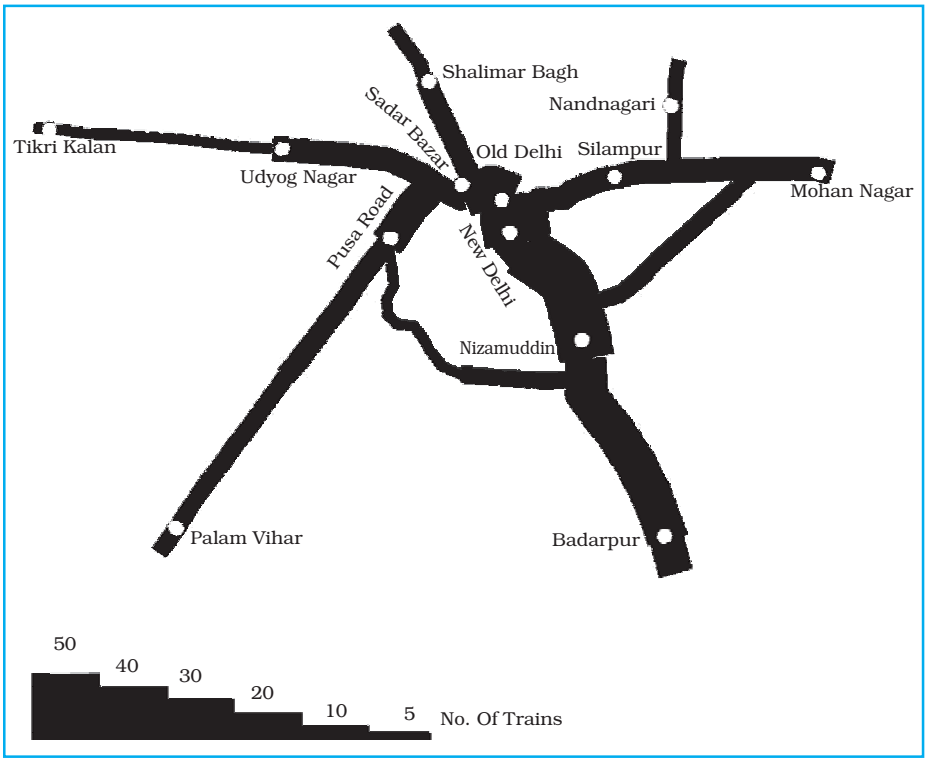


Fig. 3.10 : *Traffic (Railway) Flow Map of Delhi*

Example 3.10 : Construct a water flow map of Ganga Basin as shown in *Fig. 3.11.*

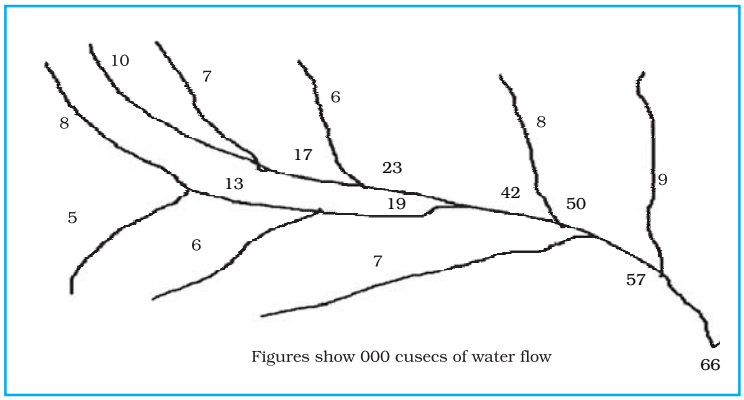


Fig. 3.11 : *Ganga Basin*

Construction

(a) Take a scale as a strip of 1cm width = 50,000 cusecs of water.

(b) Make the diagram as shown in *Fig. 3.12.*

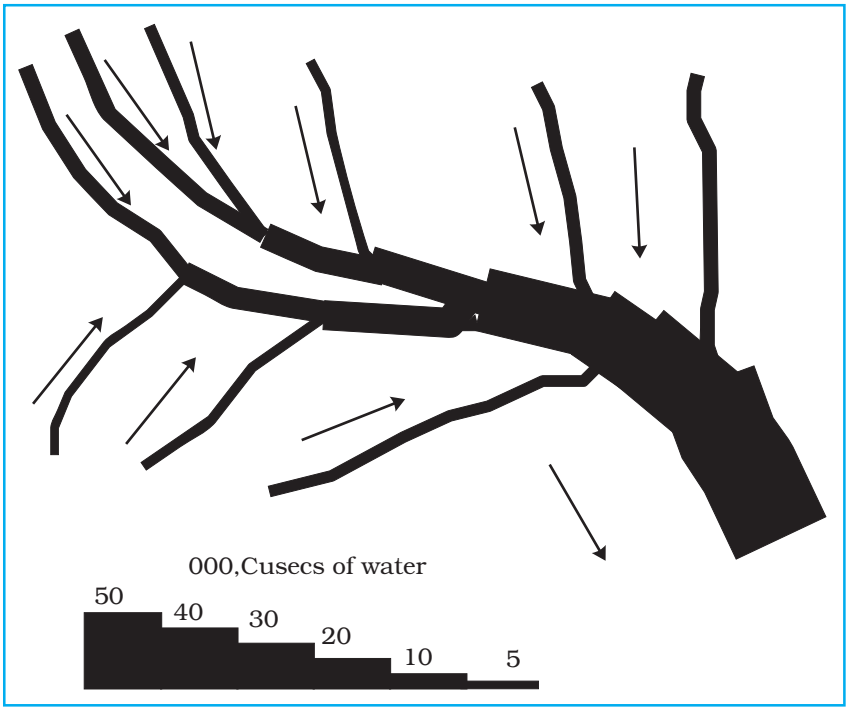


Fig. 3.12 : *Construction of a Flow Map*

Thematic Maps

Graphs and diagrams serve a useful purpose in providing a comparison between the internal variations within the data of different characteristics represented. However, the use of graphs and diagrams, at times, fails to produce a regional perspective. Hence, variety of maps may also be drawn to understand the patterns

of the regional distributions or the characteristics of variations over space. These maps are also known as the **distribution maps**.

Requirements for Making a Thematic Map

(a) State/District level data about the selected theme.

(b) Outline map of the study area alongwith administrative boundaries.

(c) Physical map of the region. For example, physiographic map for population distribution and relief and drainage map for constructing transportation map.

Rules for Making Thematic Maps

(i) The drawing of the thematic maps must be carefully planned. The final map should properly reflect the following components:

a. Name of the area

b. Title of the subject-matter

c. Source of the data and year

d. Indication of symbols, signs, colours, shades, etc.

e. Scale

(ii) The selection of a suitable method to be used for thematic mapping.

Classification of Thematic Maps based on Method of Construction

The thematic maps are generally, classified into quantitative and non-quantitative maps. The quantitative maps are drawn to show the variations within the data. For example, maps depicting areas receiving more than 200 cm, 100 to 200 cm, 50 to 100 cm and less than 50 cm of rainfall are referred as quantitative maps. These maps are also called as statistical maps. The non-quantitative maps, on the other hand, depict the non–measurable characteristics in the distribution of given information such as a map showing high and low rainfall-receiving areas. These maps are also called as qualitative maps. It would not be possible to discuss drawing these different types of thematic maps under the constraint of time. We will, therefore, confine to discuss the methods of the construction of the following types of quantitative maps :

(a) Dot maps

(b) Choropleth maps

(c) Isopleth maps

Dot Maps

The dot maps are drawn to show the distribution of phenomena such as population, cattle, types of crops, etc. The dots of same size as per the chosen scale are marked over the given administrative units to highlight the patterns of distributions.

Requirement

(a) An administrative map of the given area showing state/district/block boundaries.

(b) Statistical data on selected theme for the chosen administrative units, i.e., total population, cattle etc.

(c) Selection of a scale to determine the value of a dot.

(d) Physiographic map of the region especially relief and drainage maps.

Precaution

(a) The lines demarcating the boundaries of various administrative units should not be very thick and bold.

(b) All dots should be of same size.

Example 3.12 : Construct a dot map to represent population data as given in *Table 3.9*.

Table 3.9 : Population of India, 2001

Sl. No.	States/Union Territories	Total Population	No. of dots
1.	Jammu & Kashmir	10,069,917	100
2.	Himachal Pradesh	6,077,248	60
3.	Punjab	24,289,296	243
5.	Uttaranchal	8,479,562	85
6.	Haryana	21,082,989	211
7.	Delhi	13,782,976	138
8.	Rajasthan	56,473,122	565
9.	Uttar Pradesh	166,052,859	1,660
10.	Bihar	82,878,796	829
11.	Sikkim	540,493	5
12.	Arunachal Pradesh	1,091,117	11
13.	Nagaland	1,988,636	20
14.	Manipur	2,388,634	24
15.	Mizoram	891,058	89
16.	Tripura	3,191,168	32
17.	Meghalaya	2,306,069	23
18.	Assam	26,638,407	266
19.	West Bengal	80,221,171	802
20.	Jharkhand	26,909,428	269
21.	Orissa	36,706,920	367
22.	Chhattisgarh	20,795,956	208
23.	Madhya Pradesh	60,385,118	604
24.	Gujarat	50,596,992	506
25.	Maharashtra	96,752,247	968
26.	Andhra Pradesh	75,727,541	757
27.	Karnataka	52,733,958	527
28.	Goa	1,343,998	13
29.	Kerala	31,838,619	318
30.	Tamil Nadu	62,110,839	621

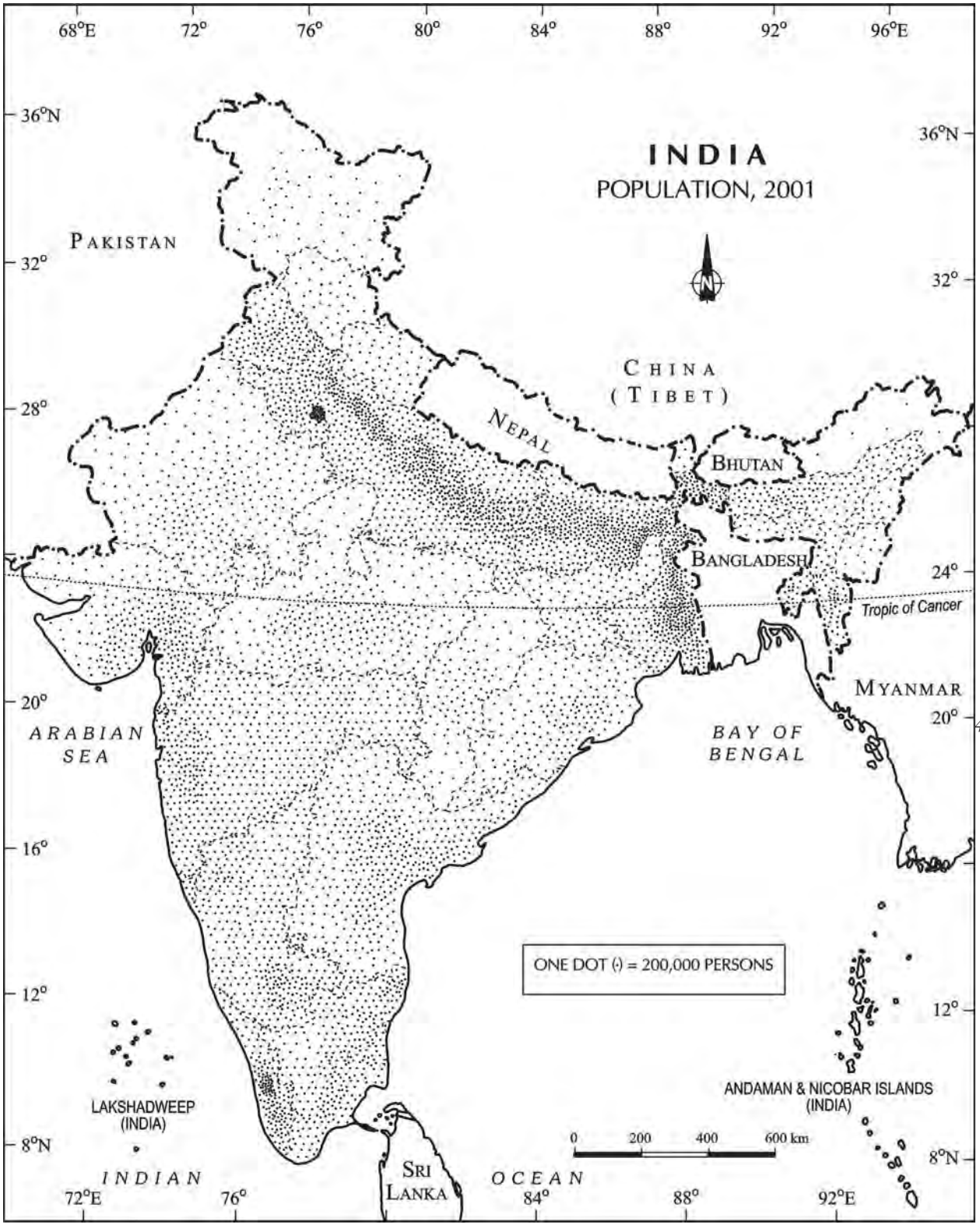


Fig. 3.13 : *Population of India, 2001*

Construction

(a) Select the size and value of a dot.

(b) Determine the number of dots in each state using the given scale. For example, number of dots in Maharashtra will be 9,67,52,247/100,000 = 967.52. It may be rounded to 968, as the fraction is more than 0.5.

(c) Place the dots in each state as per the determined number in all states.

(d) Consult the physiographic/relief map of India to identify mountainous, desert, and/or snow covered areas and mark lesser number of dots in such areas.

Choropleth Map

The choropleth maps are also drawn to depict the data characteristics as they are related to the administrative units. These maps are used to represent the density of population, literacy/growth rates, sex-ratio, etc.

Requirement for drawing Choropleth Map

(a) A map of the area depicting different administrative units.

(b) Appropriate statistical data according to administrative units.

Steps to be followed

(a) Arrange the data in ascending or descending order.

(b) Group the data into 5 categories to represent very high, high, medium, low and very low concentrations.

(c) The interval between the categories may be identified on the following formulae i.e. Range/5 and Range = maximum value – minimum value.

(d) Patterns, shades or colour to be used to depict the chosen categories should be marked in an increasing or decreasing order.

Example 3.13: Construct a Choropleth map to represent the literacy rates in India as given in *Table 3.10*.

Construction

(a) Arrange the data in ascending order as shown above.

(b) Identify the range within the data. In the present case, the states recording the lowest and highest literacy rates are Bihar (47%) and the Kerala (90.9%) respectively. Hence, the range would be 91.0 – 47.0 = 44.0

(c) Divide the range by 5 to get categories from very low to very high. (44.0/5 = 8.80. We can convert this value to a round number, i. e., t 9.0

(d) Determine the number of the categories alongwith range of each category. Add 9.0 to the lowest value of 47.0 as so on. We will finally get following categories :

47 – 56	Very low (Bihar, Jharkhand, Arunachal Pradesh, Jammu and Kashmir)
56 – 65	Low (Uttar Pradesh, Rajasthan, Andhra Pradesh, Meghalaya, Orissa, Assam, Madhya Pradesh, Chhattisgarh)

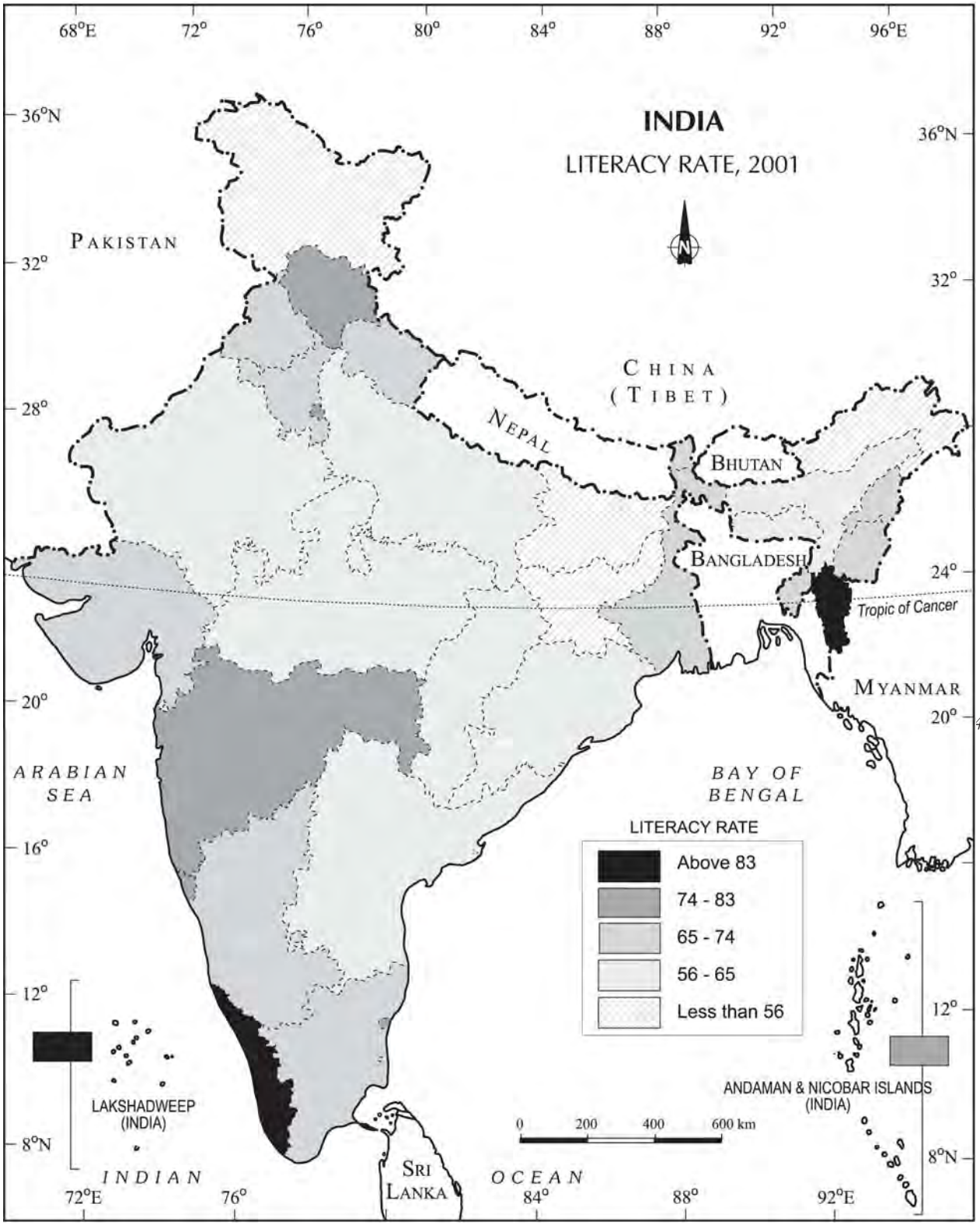


Fig. 3.14 : *Literacy Rate, 2001*

Table 3.10 : Literacy Rate in India, 2001

Original Data on Literacy in India		
S. No.	**States / Union Territories**	**Literacy Rate**
1.	Jammu & Kashmir	55.5
2.	Himachal Pradesh	76.5
3.	Punjab	69.7
4.	Chandigarh	81.9
5.	Uttaranchal	71.6
6.	Haryana	67.9
7.	Delhi	81.7
8.	Rajasthan	60.4
9.	Uttar Pradesh	56.3
10.	Bihar	47.0
11.	Sikkim	68.8
12.	Arunachal Pradesh	54.3
13.	Nagaland	66.6
14.	Manipur	70.5
15.	Mizoram	88.8
16.	Tripura	73.2
17.	Meghalaya	62.6
18.	Assam	63.3
19.	West Bengal	68.6
20.	Jharkhand	53.6
21.	Orissa	63.1
22.	Chhattisgarh	64.7
23.	Madhya Pradesh	63.7
24.	Gujarat	69.1
25.	Daman & Diu	78.2
26.	Dadra & Nagar Haveli	57.6
27.	Maharashtra	76.9
28.	Andhra Pradesh	60.5
29.	Karnataka	66.6
30.	Goa	82.0
31.	Lakshadweep	86.7
32.	Kerala	90.9
33.	Tamil Nadu	73.5
34.	Pondicherry	81.2
35.	Andaman & Nicobar	81.3

Data on Literacy in India as arranged in Ascending order		
S. No.	**States / Union Territories**	**Literacy Rate**
10.	Bihar	47.0
20.	Jharkhand	53.6
12.	Arunachal Pradesh	54.3
01.	Jammu & Kashmir	55.5
9.	Uttar Pradesh	56.3
26.	Dadra & Nagar Haveli	57.6
08.	Rajasthan	60.4
28.	Andhra Pradesh	60.5
17.	Meghalaya	62.6
21.	Orissa	63.1
18.	Assam	63.3
23.	Madhya Pradesh	63.7
22.	Chhattisgarh	64.7
13.	Nagaland	66.6
29.	Karnataka	66.6
06.	Haryana	67.9
19.	West Bengal	68.6
11.	Sikkim	68.8
24.	Gujarat	69.1
03.	Punjab	69.7
14.	Manipur	70.5
05.	Uttaranchal	71.6
16.	Tripura	73.2
33.	Tamil Nadu	73.5
02.	Himachal Pradesh	76.5
27.	Maharashtra	76.9
25.	Daman & Diu	78.2
34.	Pondicherry	81.2
35.	Andaman & Nicobar	81.3
07.	Delhi	81.7
04.	Chandigarh	81.9
30.	Goa	82.0
31.	Lakshadweep	86.7
15.	Mizoram	88.8
32.	Kerala	90.9

65 – 74 Medium (Nagaland, Karnataka, Haryana, West Bengal, Sikkim, Gujarat, Punjab, Manipur, Uttaranchal, Tripura, Tamil Nadu)

74 – 83 High (Himachal Pradesh, Maharashtra, Delhi, Goa)

83 – 92 Very High (Mizoram, Kerala)

(e) Assign shades/pattern to each category ranging from lower to higher hues.

(f) Prepare the map as shown in *Fig. 3.14*.

(g) Complete the map with respect to the attributes of map design.

Isopleth Map

We have seen that the data related to the administrative units are represented using choropleth maps. However, the variations within the data, in many cases, may also be observed on the basis of natural boundaries. For example, variations in the degrees of slope, temperature, occurrence of rainfall, etc. possess characteristics of the continuity in the data. These geographical facts may be represented by drawing the lines of equal values on a map. All such maps are termed as Isopleth Map. The word **Isopleth** is derived from **Iso** meaning equal and **pleth** means lines. Thus, an imaginary line, which joins the places of equal values, is referred as Isopleth. The more frequently drawn isopleths include Isotherm (equal temperature), Isobar (equal pressure), Isohyets (equal rainfall), Isonephs (equal cloudiness), Isohels (equal sunshine), contours (equal heights), Isobaths (equal depths), Isohaline (equal salinity), etc.

Requirement

(a) Base line map depicting point location of different places.

(b) Appropriate data of temperature, pressure, rainfall, etc. over a definite period of time.

(c) Drawing instrument specially French Curve, etc.

Rules to be observed

(a) An equal interval of values be selected.

(b) Interval of 5, 10, or 20 is supposed to be ideal.

(c) The value of Isopleth should be written along the line on either side or in the middle by breaking the line.

Interpolation

Interpolation is used to insert the intermediate values between the observed values of at two stations/locations, such as temperature recorded at Chennai and Hyderabad or the spot heights of two points. Generally, drawing of isopleths joining the places of same value is also termed as interpolation.

Method of Interpolation

For interpolation, follow the following steps:

(a) Firstly, determine the minimum and maximum values given on the map.

(b) Calculate the range of value i.e. Range = maximum value – minimum value.

(c) Based on range, determine the interval in a whole number like 5, 10, 15, etc.

The exact point of drawing an Isopleth is determined by using the following formulae.

$$\text{Point of Isopleth} \quad \frac{\text{Distance between two points in cm}}{\text{Difference between the two values of corresponding points}} \quad \text{Interval}$$

The interval is the difference between the actual value on the map and interpolated value. For example, in an Isotherm map of two places show 28º C and 33º C and you want to draw 30º C isotherm, measure the distance between the two points. Suppose the distance is 1cm or 10 mm and the difference between 28 and 33 is 5, whereas 30 is 2 points away from 28 and 3 points behind 33, thus, exact point of 30 will be

Thus, isotherm of 30ºC will be plotted 4mm away from 28ºC or 6mm ahead of 33ºC.

(d) Draw the isopleths of minimum value first; other isopleths may be drawn accordingly.

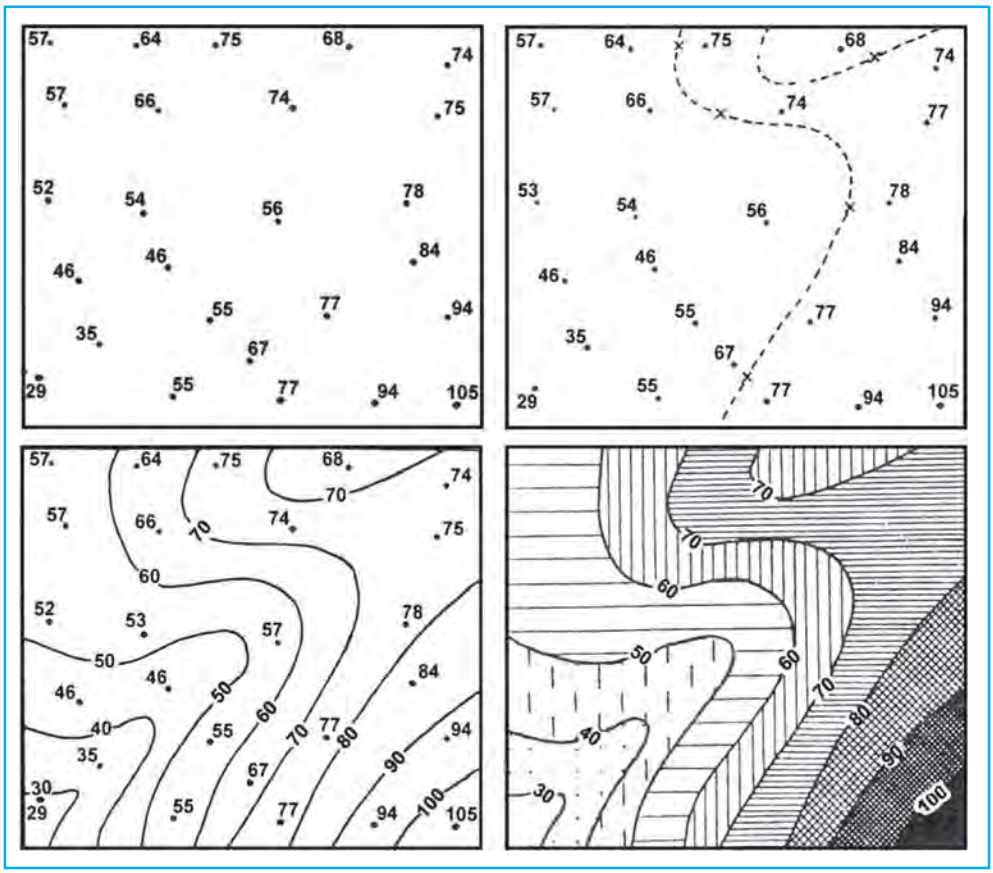


Fig. 3.15 : *Drawing of Isopleths*

Excercises

1. Choose the right answer from the four alternatives given below:

(i) Which one of the following map shows the Population distribution:

(a) Choropleth maps (b) Isopleth maps

(c) Dot maps (d) Square root map

(ii) Which one of the following is best suited to represent the decadal growth of population?

(a) Line graph (b) Bar diagram

(c) Circle diagram (d) Flow diagram

(iii) Polygraph is constructed to represent:
(a) Only one variable (b) Two variables only
(c) More than two variables (d) None of the above

(iv) Which one of the following maps is known as "Dynamic Map"?
(a) Dot map (b) Choropleth
(c) Isopleth (d) Flow map

2. Answer the following questions in about 30 words:

(i) What is a thematic map?

(ii) Differentiate between multiple bar diagram and compound bar diagram.

(iii) What are the requirements to construct a dot map?

(iv) Describe the method of constructing a traffic flow map.

(v) What is an Isopleth map ? How an interpolation is carried out?

(vi) Describe and illustrate important steps to be followed in preparing a choropleth map.

(vii) Discuss important steps to represent data with help of a pie-diagram.

Activity

1. Represent the following data with the help of suitable diagram.

India : Trends of Urbanisation 1901-2001

Year	*Decennial growth (%)*
1911	0.35
1921	8.27
1931	19.12
1941	31.97
1951	41.42
1961	26.41
1971	38.23
1981	46.14
1991	36.47
2001	31.13

2. Represent the following data with the help of suitable diagram.

India : Literacy and Enrolment Ratio in Primary and Upper Primary Schools

Year	***Literacy Ratio***			***Enrolment Ratio Primary***			***Enrolment Ratio Upper Primary***		
	Person	Male	Female	Boys	Girls	Total	Boys	Girls	Total
1950-51	18.3	27.2	8.86	60.6	25	42.6	20.6	4.6	12.7
1999-2000	65.4	75.8	54.2	104	85	94.9	67.2	50	58.8

3. Represent the following data with help of pie-diagram.

India : Land use 1951-2001

	1950-51	1998-2001
Net Sown Area	42	46
Forest	14	22
Not available for cultivation	17	14
Fallow Land	10	8
Pasture and Tree	9	5
Cultruable Waste Land	8	5

4. Study the table given below and draw the given diagrams/maps.

Area and Production of Rice in major States

States	*Area in 000 ha*	*% to total area*	*Production 000 tones*	*% to total production*
West Bengal	5,435	12.3	12,428	14.6
Uttar Pradesh	5,839	13.2	11,540	13.6
Andhra Pradesh	4,028	9.1	12,428	13.5
Punjab	2,611	5.9	9,154	10.8
Tamil Nadu	2,113	4.8	7,218	8.5
Bihar	3,671	8.3	5,417	6.4

(a) Construct a multiple bar diagram to show area under rice in each State.

(b) Construct a pie-diagram to show the percentage of area under rice in each State.

(c) Construct a dot map to show the production of rice in each State.

(d) Construct a Choropleth map to show the percentage of production of rice in States.

5. Show the following data of temperature and rainfall of Kolkata with a suitable diagram.

Months	*Temperature in ° C*	*Rainfall in cm*
Jan.	19.6	1.2
Feb.	22.0	2.8
Mar.	27.1	3.4
Apr.	30.1	5.1
May	30.4	13.4
June	29.9	29.0
Jul.	28.9	33.1
Aug.	28.7	33.4
Sep.	28.9	25.3
Oct.	27.6	12.7
Nov.	23.4	2.7
Dec.	19.7	0.4

4 Use of Computer in Data Processing and Mapping

You have learnt various methods of data processing and representation that you can use to analyse the geographical phenomena in the preceding chapters. You must have observed that these methods are time consuming and tedious. Have you ever thought of a method of data processing and their graphical presentation that can save time and improve efficiency? If you have used a computer for word processing then you must have noticed that the computer is more versatile as it facilitates the onscreen editing of the text, copy and move it from one place to another, or even delete the unwanted text. Similarly, the computer may also be used for data processing, preparation of diagrams/graphs and the drawing of maps, provided you have an access to the related application software. In other words, a computer can be used for a wide range of applications. It must, however, be clearly understood that a computer carries out the instructions it receives from the users. In other words, it cannot perform any function on its own. In the present chapter, we will discuss the use of computers in data processing and mapping.

What can a Computer do?

A computer is an electronic device. It consists of various sub-systems like memory, micro-processor, input system and output system. All these sub-systems work together to make it an integrated system. It is an extremely powerful device, which is apt to have an important effect on the systems of data processing, mapping and analysis. A computer is a fast and versatile machine that can perform simple arithmetic operations, such as, addition, subtraction, multiplication, and division and can also solve complex mathematical formulae. It also performs simple logical operations, distinguishing zero from non-zero and plus from minus and discharge the results. In short, a computer is a data processor that can perform substantial computation, including numerous arithmetic or logical operations, without intervention by a human operator during the run.

Provided that you have the basic conceptual clarity, computer can be used very effectively to represent data through maps and diagrams. It makes your job extremely fast. The following advantages of a computer make it distinct from the manual methods:

1. It substantially increases the speed of the computation and data processing.
2. It can handle huge volume of the data, which is normally not possible manually.
3. It facilitates copy, edit, save and retrieve the data at will.
4. It further enables validation, checking and correction of data easily.
5. Aggregation and analysis of data becomes extremely simple. Computer makes it very easy to perform comparative analysis, whether by drawing maps or graphs.
6. The type of graph or map (i.e. bar/pie or types of shades), heading, indexing and other formats can be changed very easily.

There are many other advantages that a computer offers, that you will observe yourselves while carrying out your practical work using a computer.

Hardware Configuration and Software Requirements

A computer as an aid to data processing and mapping comprises of hardware and software. The hardware configurations comprise of the storage, display, and input and output sub-systems, whereas software are the programs that are made up of electronic codes. The computer–aided data processing and mapping, hence, requires both hardware components and related application software.

Hardware

The hardware components of a computer include :

(a) A Central Processing Unit (CPU) and Storage System
(b) A Graphic Display Sub-system
(c) Input Devices
(d) Output Devices

A Central Processing Unit and Storage System

The core of modern computers consists of a central processing unit (CPU), which facilitates the execution of program instructions for processing data and controlling peripheral equipments. All data together with the operating system and the application programs occupy space in disk storage unit which functions as working memory.

The total storage capacity depends upon the types of activity for which the computer is to be used. The hardware storage capacity for data processing and mapping should be in the range of 1 GB to 4 GB or more and the Random Access Memory (RAM) 32 MB or more. Besides the disk storage, the secondary storage such as floppy disks, CD, pen drives, and magnetic tapes are also used to store permanently large quantities of data that is not actively being processed.

The operating system is a basic program, which administers the internal data processing in a computer. The operating systems like MS-DOS, Windows, and UNIX are in general use, with the Windows being the most preferred one.

A Graphic Display System or Monitor

A graphic display system or monitor serves as the user's prime visual communication medium in all computers. A high resolution display system with a greater range of possible display colours and Look–up Tables (LUT) for rapid alteration of colour patterns is generally preferred in graphic and mapping applications.

Input Devices

The instruction and the statistical data are entered into the computer using the keyboard functions. The **keyboard** is an important input device that resembles with a typewriter. It has various keys for different purposes. While working on a PC you will notice a flash point on the screen. This is known as **cursor**. When you press a key on the keyboard, a character is displayed at the point where the cursor is flashing and the cursor moves one position forward. Besides, scanners and digitisers of different size and capabilities are also used for spatial data entry.

Output Devices

The output devices include a variety of printers such as ink-jet, laser and colour laser printers; and the plotters that are available in different sizes ranging from A3 to A0 size.

Computer Software

Computer software is a written program that is stored in memory. It performs specific functions as per the instructions given by the user. A data processing and mapping software requires the following modules :

- Data Entry and Editing Modules
- Coordinate Transformation and Manipulation Modules
- Data Display and Output Modules

The Data Entry and Editing Modules

These inbuilt modules in the data processing and mapping software facilitate the data entry system interface, database creation, error removal, scale and projection manipulations, their organisation, and maintenance of the data. Any of these and other related data entry, editing and management capabilities might be performed using displayed menus and icons on the screen. The present day commercial packages such as MS Excel/Spread sheet, Lotus 1 – 2 – 3, and d – base provide capabilities for data processing and generation of graphs. On the other hand, Arc View/Arc GIS, Geomedia, possess modules for mapping and analysis.

Coordinate Transformation and Manipulation Modules

The present day softwares provide a wide range of capabilities used to create layers of spatial data, coordinate transformation, editing and linking the spatial data sets with the related non - spatial attributes of data.

Data Display and Output Modules

The data display and output operations vary over a range of functions and are very much dependent on the skills developed in the field of computer graphics. Some of the common capabilities that the present day softwares provide are:

- Zooming/Windowing to display of selected areas and scale change operation
- Colour assignment/change operation
- Three dimensional and perspective display
- Selective display of various themes
- Polygon shading, line styling and point markers display
- Output device interface commands for interfacing with plotter devices/ printers
- Graphic User Interface (GUI) based menu organisation for an easy interface

Computer Software for Your Use

In the preceding paragraphs, a number of data processing softwares have been referred. However, it would be difficult to discuss the capabilities and functions of each one of these softwares under the constraints of time and space. We will, therefore, describe the procedure that is followed in data processing and the preparation of graphs and diagrams using **MS Excel or Spreadsheet** program. The **spreadsheet** enables us to feed data, compute various statistics and represent the raw data or computed statistics through graphical methods.

MS Excel or Spreadsheet

As mentioned earlier, MS Excel, Lotus 1 – 2 – 3, and d – base are some of the important softwares used for data processing, and drawing graphs and diagrams. MS Excel being most widely used and commonly available software program in all parts of the country has been chosen among other software to carry out the data processing. Besides, it is also compatible with map-making software as one can easily feed data in MS Excel and attach it to the map-making software to create maps.

MS Excel is also called a spreadsheet programme. A spreadsheet is a rectangular table (or grid) to store information. The spreadsheets are located in Workbooks or Excel files.

Most of the MS Excel screen is devoted to the display of the worksheet, which consists of rows and columns. The intersection of a row and column is a rectangular area, which is called a **cell**. In other words, a worksheet is made up of cells. A cell can contain a numerical value, a formula (which after calculation provides numerical value) or text. Texts are generally used for labelling numbers entered in the cells. A value entry can either be a number (entered directly) or result of a formula. The value of a formula will change when the components (arguments) of the formula change.

An Excel worksheet contains 16,384 rows, numbered 1 through 1,6384 and 256 columns, represented by default through letters A through Z, AA through AZ, BA through BZ, and continuing to IA through IZ. By default, an Excel workbook consists of three worksheets. If you require, you can insert more, up to 256 worksheets. This means that in the same file/workbook you can store a large number of data and charts. *Fig.4.1* shows how an excel workbook looks like.

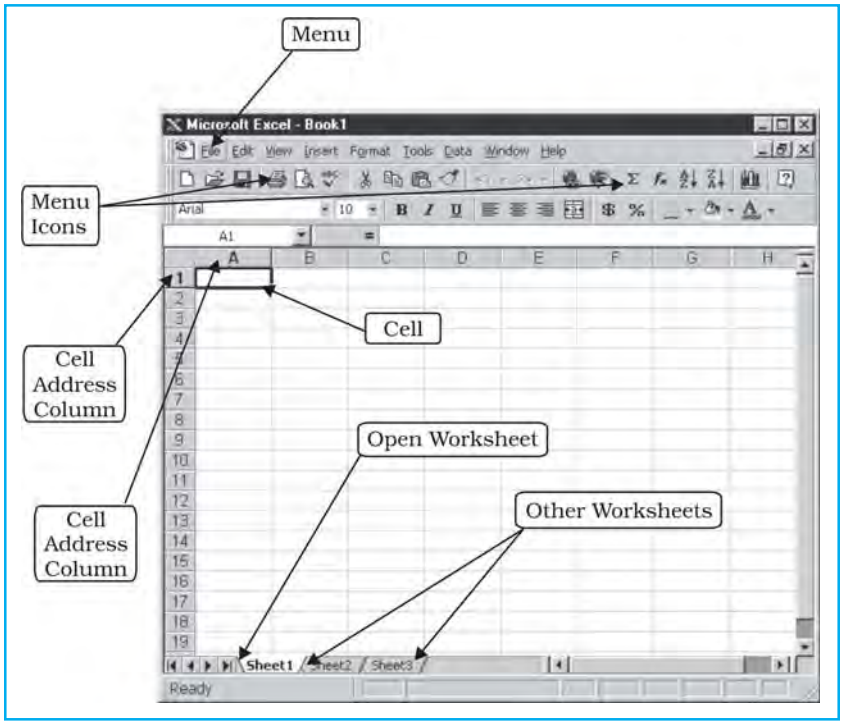


Fig. 4.1 : *MS Excel Workbook*

Data Entry and Storing Procedures in Excel

The data entry and storing procedures are very simple in Excel. You can enter, copy and move any data from one cell to another and save them. You may also delete incorrect or unwanted data entry or a complete file, if it is not required for further use. The elementary functions of Excel that you would require for entering data and storing them are described in *Table 4.1*. You can learn more on your own by exploring other menus and options by yourself. Further, you will find it easier to feed data if you use the **number pad** given on the right side of your keyboard. For entering data column-wise, you need to press 'enter key' or 'down arrow' after typing a number. While row-wise pressing right arrow key after typing a number can enter data.

Data Processing and Computation

Often raw data need to be processed for further use. You can easily add, subtract, multiply, and divide numbers using the keyboard signs of +, -, *, and /, respectively. These signs are known as **operators** and they connect elements in a **formula** or **expression**. For example, if you want to solve the expression 5 + 6 – 8 – 5, then you can easily work it out in steps below :

Step 1 : Click on any cell (with the help of mouse).

Step 2 : Type =, followed by the expression. Thus, the expression becomes = 5 + 6 – 8 – 5.

Step 3 : Press enter key, and you will get the result in the same cell that you had chosen in Step 1.

Note : The numerical operations can only be performed in excel by first typing = sign.

Table 4.1: Important Functions for Entering and Storing Data

S. No.	Function	Instructions	Menu	Secondary Menu (from dropdown list)	Keyboard Shortcuts
1.	For opening a new file		File	New	Ctrl N
	For opening an existing file		File	Open	Ctrl O
2.	Save a file	Give a file name and define where you want to store it (by default, it is c:\....\my documents\)	File	Save	Ctrl S
3.	Copy, move and paste a set of data	Select the set of data by pressing the left mouse button and dragging it over the set of the data you want to select	Edit	Copy	Ctrl C
4.	Cut, move and paste a set of data	Select the set of data by pressing the left mouse button and dragging it over the set of the data you want to select	Edit	Cut	Ctrl X
5.	Paste a set of data	Take the cursor to the cell where you want to paste it	Edit	Paste	Ctrl V
6.	For undoing the last action*		Edit	Undo	Ctrl Z
7.	For redoing the last action*		Edit	Repeat	Ctrl Y

*Note: * You cannot undo or redo any action if you have saved the file after the last action.*

These operators that connect elements in a formula are solved in an order. The expressions enclosed in 'brackets' are solved first and are followed by the 'exponents', 'division', 'multiplication', 'addition' and 'subtraction'. For example, expression/formula within a cell given as =A8/(A9 + A4) will be solved using Excel as under:

It will first add the values entered in cells A9 and A4, and then will divide the value of A8 by the sum.

Further, if you want to supplement your understanding on the percentage share of urban population to the total population, in that case, you have to calculate the percentage of urban population in various states of India. To do so, you will require the data on urban population and total population for each

state of India. The worksheet allows you to easily calculate the percentage of urban population in each state provided you adopt the following steps :

Step 1 : Enter the name of the states in first column (i.e. column A).

Step 2 : In Column B, corresponding to each state, enter the size of urban population.

Step 3 : In Column C, corresponding to respective state enter the size of total population.

Step 4 : In Column D and row 2, type = followed by B2/C2 (that is total urban population of Andhra Pradesh divided by the total population in the same State) and *100 (multiplied by 100). Thus, the expression becomes =B2/C2*100

Step 5 : Press enter key. This will give you solution of the expression, that is, the percentage of urban population in Andhra Pradesh.

Step 6 : Now you need not to write the formula again for calculating percentage of urban population for other states. Simply, click on the cell D2. This will copy the formula of the first state/cell to all the downward cells you have dragged it over.

(Note: *the formula =B2/C2*100 that has been written in cell D2, and becomes B3/C3*100 in cell D3, and so on).*

'*Fig. 4.2* graphically shows steps 1 to 5 as given above, while step 6 is shown in *Fig.4.3*.

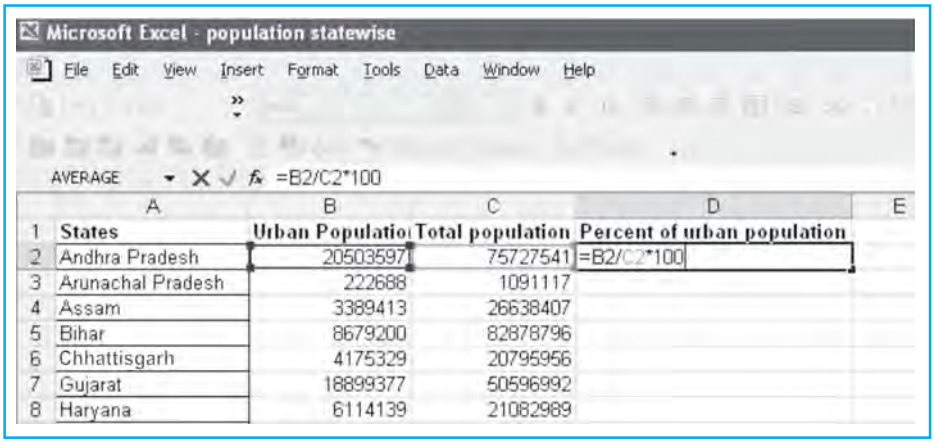


Fig. 4.2 : *Cell Operation in MS Excel*

	A	B	C	D
1	States	Urban Populatio	Total population	Percent of urban population
2	Andhra Pradesh	20503597	75727541	27.08
3	Arunachal Pradesh	222688	1091117	20.41
4	Assam	3389413	26638407	12.72
5	Bihar	8679200	82878796	10.47
6	Chhattisgarh	4175329	20795956	20.08
7	Gujarat	18899377	50596992	37.35
8	Haryana	6114139	21082989	29.00
9	Himachal Pradesh	594881	6077248	9.79
10	Jammu & Kashmir	2505309	10069917	24.88
11	Jharkhand	5986697	26909428	22.25
12	Karnataka	17919858	52733958	33.98
13	Kerala	8267135	31838619	25.97

Fig. 4.3 : *Copying through Dragging over in MS Excel*

You have already been introduced to some basic statistical methods such as measures of central tendency, dispersion and correlation in Chapter 2. You must have understood the concept and rationale behind these techniques. The use of worksheet functions to compute these statistics will be discussed in the subsequent paragraphs.

In MS Excel, there are numerous inbuilt statistical and mathematical functions. These functions are located in **Insert** menu. To use the function, click on the **Insert** menu, and choose **f_x (Function)** from the dropdown list. Note that your cursor should be located in the cell where you want the formula to appear. Some examples of application of statistical functions are given below.

Central Tendencies

Central tendencies are represented by mean, median and mode. Arithmetic mean, also called as average, is a commonly used method for calculating the central tendency. In MS Excel, it is denoted by its popular name **average**. As an example, we shall calculate mean cropping intensity in India during various decades using the average function in Excel. The following steps are to be undertaken :

Step 1 **:** Enter year-wise cropping intensity data in a worksheet, as shown in *Fig.4.4*.

Step 2 **:** Click on cell **B12** using mouse**.**

Step 3 **:** Click on **Insert Menu** and choose **f_x (Function)** from dropdown list, this will open **Insert Function dialogue box**.

Step 4 **:** Select **Statistical** from **select a category** menu on the dialogue box. This will bring forth the statistical functions available in Excel in the box below in the same dialogue box,

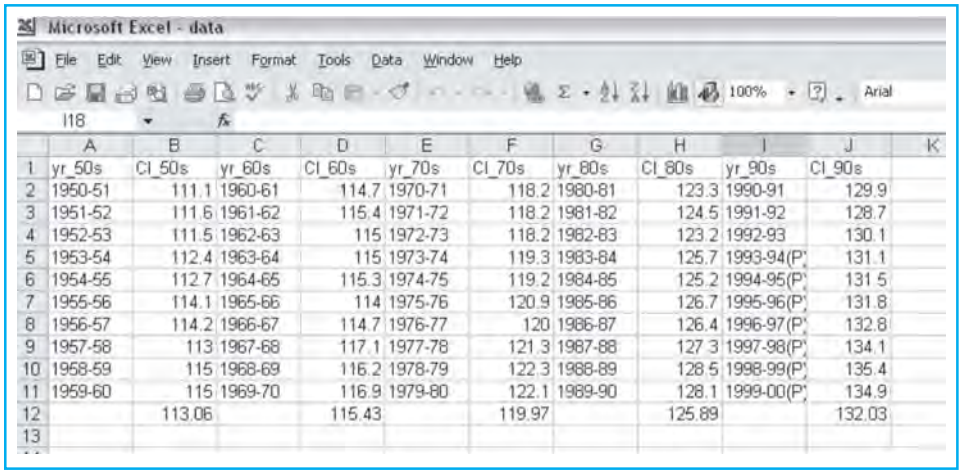
Step 5 **:** In the box **Select a Function,** click on **Average,** and press **OK** button. This will open another dialogue box called **Function Argument**.

Step 6 **:** Either enter the cell range of data of the first decade CI_50s (which shows year wise cropping intensity in 1950s) in the **Number 1 box** on **Function Argument** dialogue box of data, or drag cursor pressing the left button of mouse over the cell range of data.

Step 7 **:** Press OK button on the **Function Argument** dialogue box. This calculates mean cropping intensity for the decade 1950s in cell **B12**, where you had put your cursor in the beginning.

Step 8 **:** Now calculate the mean for other decade either following Steps 1 – 7 given above or dragging cursor right handward in the same row selecting the small square from rectangle of cell B12 or you can copy the cell B12 and paste it on D12, F12, H12 and J12. This will give you mean value of cropping intensity for the decades 1960s, 1970s, 1980s and 1990s, respectively.

These steps are further explained in *Fig. 4.4* through *Fig.4.6*.



	A	B	C	D	E	F	G	H	I	J	K
1	yr_50s	CI_50s	yr_60s	CI_60s	yr_70s	CI_70s	yr_80s	CI_80s	yr_90s	CI_90s	
2	1950-51	111.1	1960-61	114.7	1970-71	118.2	1980-81	123.3	1990-91	129.9	
3	1951-52	111.6	1961-62	115.4	1971-72	118.2	1981-82	124.5	1991-92	128.7	
4	1952-53	111.5	1962-63	115	1972-73	118.2	1982-83	123.2	1992-93	130.1	
5	1953-54	112.4	1963-64	115	1973-74	119.3	1983-84	125.7	1993-94(P)	131.1	
6	1954-55	112.7	1964-65	115.3	1974-75	119.2	1984-85	125.2	1994-95(P)	131.5	
7	1955-56	114.1	1965-66	114	1975-76	120.9	1985-86	126.7	1995-96(P)	131.8	
8	1956-57	114.2	1966-67	114.7	1976-77	120	1986-87	126.4	1996-97(P)	132.8	
9	1957-58	113	1967-68	117.1	1977-78	121.3	1987-88	127.3	1997-98(P)	134.1	
10	1958-59	115	1968-69	116.2	1978-79	122.3	1988-89	128.5	1998-99(P)	135.4	
11	1959-60	115	1969-70	116.9	1979-80	122.1	1989-90	128.1	1999-00(P)	134.9	
12		113.06		115.43		119.97		125.89		132.03	
13											

Fig. 4.4 : *Calculation of Mean Using Statistical Function in MS Excel*

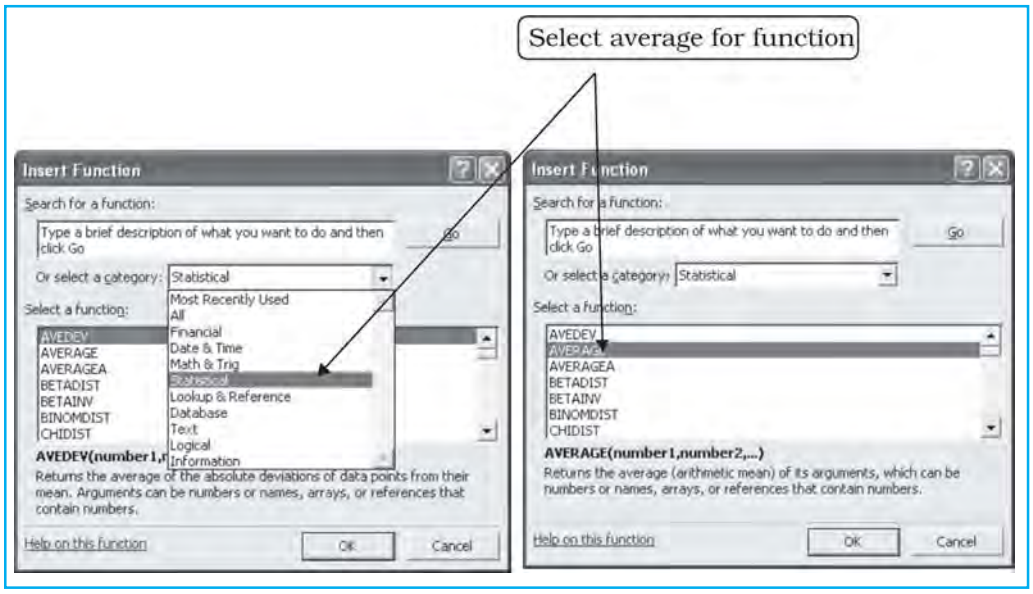


Fig. 4.5 : *Selection of Statistical Function*

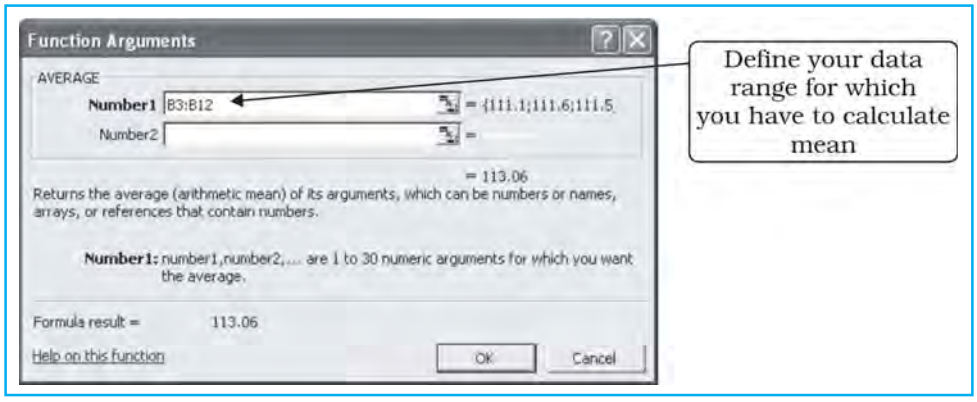


Fig. 4.6 : *Defining Range in Function Arguments dialogue box*

The computation of mean for the given data reveals that there has been an impressive increase in mean decadal cropping intensity over different decades in general, and 1980s onwards in particular. In fact, during 1980s the "Green Revolution" underwent a spatial spread and a tremendous increase in area under tube-well irrigation took place, which facilitated cultivation in the arid regions as well as during the dry seasons.

Using almost the same procedure used for calculating mean, as outlined above, you can calculate median, standard deviation, and correlation. Some hints for this are provided in *Fig.4.7* and *Fig. 4.8*.

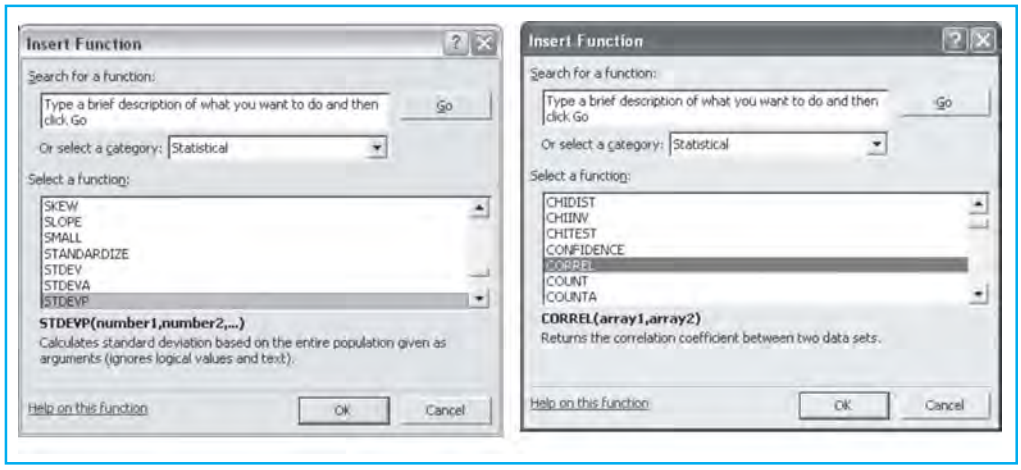


Fig. 4.7 : *Function for Standard Deviation* ***Fig. 4.8 :*** *Function for Correlation*

Construction of Graphs

You know that the data in tabular form, at times make it difficult to draw inferences about whatever is being presented. On the other hand, the representation of the data in graphical form enhances our capabilities to make meaningful comparisons between the phenomena represented, and present a simplified view of the characteristics depicted. In other words, graphs and diagrams make us to decipher the contents of data easily. For example, it would be difficult to make sense of the Cropping Intensity in India if the data for all Fifty years are presented in tabular form. However, through a line graph or bar diagram, we can easily draw meaningful conclusions about the trend in Cropping Intensity in India.

Data Types and Some Suitable Graphical Methods of their Presentation

1. Time series data are represented through line graphs or bar diagram.
2. Bar diagrams and histograms are generally used for showing shares or frequencies of various units.
3. Compound bar diagrams, and pie-charts are used for showing shares of various units.
4. Maps are used for location-wise representation of data. This helps in comprehending spatial patterns in the data.

The selection of a suitable graphical method is very important for the presentation of data. In Chapter 3, you have learnt about graphs and diagrams, and the kind of data suitable for. Here, you will learn how graphs and diagrams are constructed in Excel.

Suppose, you want to represent changes in the share of workers in different industrial categories between 1981 and 2001, the most suitable graphical methods would be **bar diagram** as it shows changes over different years clearly. The construction of bar diagram requires the following steps :

Step 1 : Enter the data in worksheet as shown in *Fig.4.9*.

Step 2 : Select the cells dragging mouse (right button pressed) over the cells.

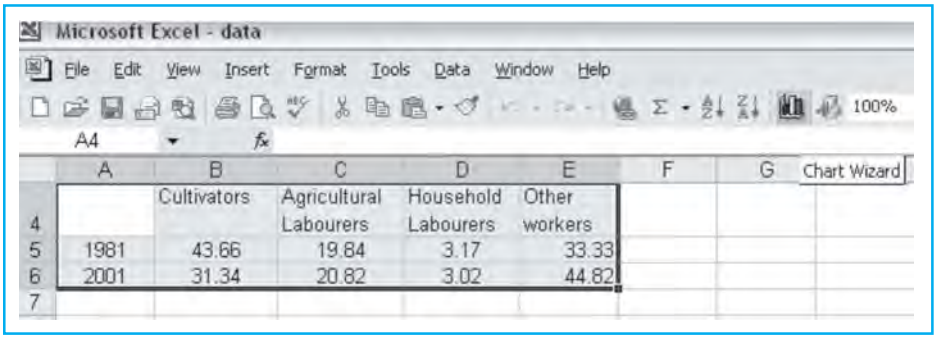


Fig. 4.9 : *Entering data and selecting cells for Construction of Bar Diagram*

Step 3 : Click on **Chart Wizard** (*Fig.4.9*). This will open **Step 1 of 4 of Chart Wizard** (*Fig.4.10*).

Step 4 : Double click on the simple bar diagram in the box '**Chart Sub-type**' (*Fig.4.10*). This will lead you to **Step 2 of 4 of Chart Wizard** (*Fig.4.11*), in which worksheet number and selected data range, and a preview of bar diagram appear. As categories in data are arranged row-wise, therefore, it is row-wise chart construction.

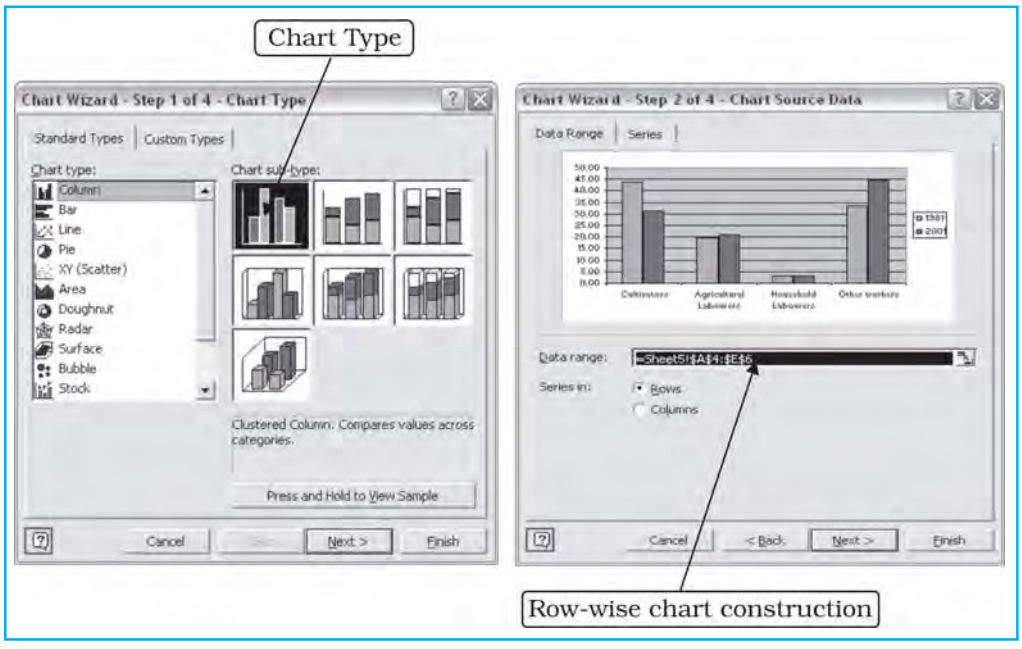


Fig. 4.10 : *Step 1 to 4 of Chart Wizard* ***Fig. 4.11 :*** *Step 2 of 4 of Chart Wizard*

Step 5 : Click on the **Next** radio button, and this will lead you to Step 3 of 4 of **Chart Wizard** (*Fig.4.12*). Here you will find various options for entering 'title' 'name of axes', options for 'grid lines', 'data labels' and

'data table'. Chart Titles and axes name entry are shown in *Fig.4.12*, while options for 'legend placement' are shown in *Fig. 4.13*. Type the axes names as shown in *Fig.4.13* and select the 'placement of legend' as shown in *Fig.4.14*.

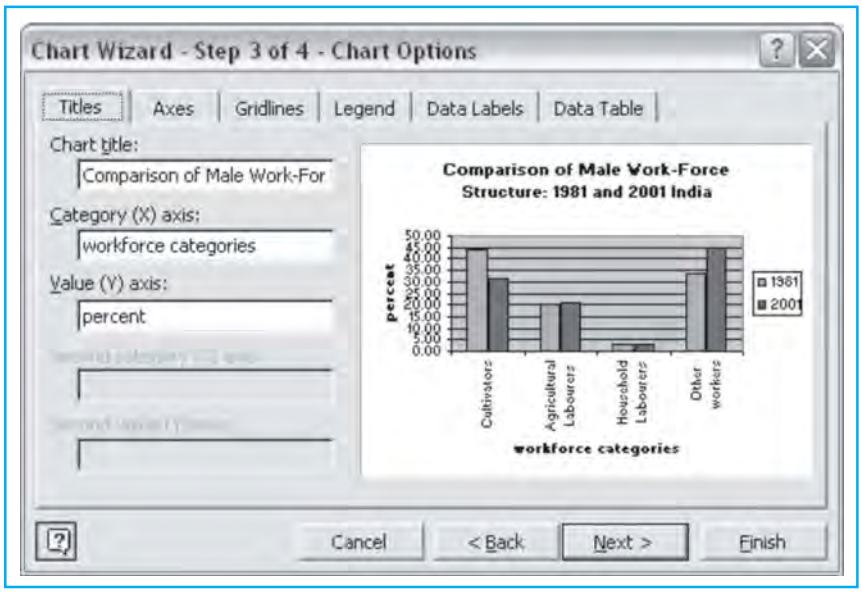


Fig. 4.12 : *Entering names of Axes*

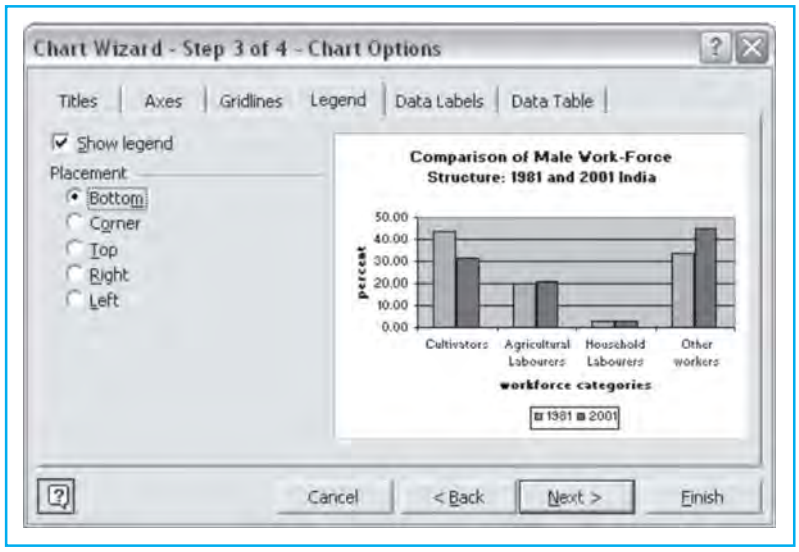


Fig. 4.13 : *Choosing Location of Legend*

Step 6 : When you have finished entering axes titles and legend options, etc., click on **Next** radio button (*Fig. 4.13*). This will lead you to **step 4 of 4 of Chart Wizard**, which will let you choose the location of the constructed bar diagram for the data (*Fig.4.14*). Choose 'As Object in' and select the same sheet you have entered the data, i.e. Sheet 5 (optionally, you can also place your Bar Diagram in a new sheet choosing 'as new sheet').

Step 7 : Press **OK** radio button in *Fig. 4.14*. This will complete the Chart Wizard and your Bar Diagram as shown in *Fig.4.15* will appear in Worksheet 5.

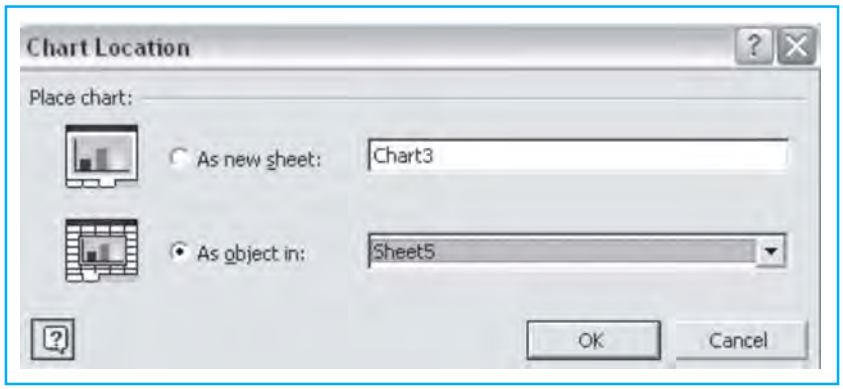


Fig. 4.14 : *Choosing Location of Chart*

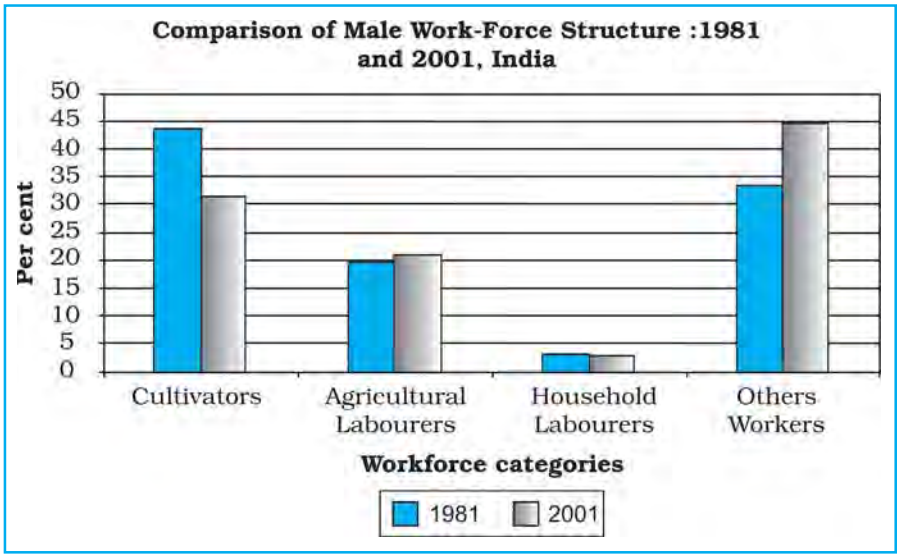


Fig. 4.15 : *The Complete Bar Diagram*

You can change the pattern of bars from colours to shades or vice versa by clicking on the bars. Similarly, you can also change the fonts or gridlines if required.

The above diagram shows that the share of cultivators has declined significantly over the two decades and the share of other workers has appreciably risen and the shares of agricultural and household labourers have largely been the same.

Some Important Norms for Data Representation

1. A figure should have its figure number.
2. It should have a suitable title in which time and space it relates to should also be mentioned.
3. Within title or as sub-title, the unit in which the quantities are shown should be mentioned.
4. The title, sub-title, title of axes, legend and the main presentation should be shown with suitable font size and type so that they occupy space in a balanced manner.

Computer Assisted Mapping

The maps may also be drawn using a combination of computer hardware and the mapping software. The computer assisted mapping essentially requires the creation of a spatial database alongwith its integration with attribute or non – spatial data. It further involves verification and structuring of the stored data. What is most important in this context is that the data must be geometrically registered to a generally accepted and properly defined coordinate system and coded so that they can be stored in the internal data base structure within the computer. Hence, care must be taken while using the computer for mapping purposes.

Spatial Data

The spatial data represent a geographical space. They are characterised by the points, lines and the polygons. The point data represent positional characteristics of the some of the geographical features such as schools, hospitals, wells, tube-wells, towns and villages, etc. on the map. In other words, if we want to present occurrence of the objects on a map in dimensionless scale but with reference to location, we use points. Similarly, lines are used to depict linear features like roads, railway lines, canals, rivers, power and communication lines, etc. Polygons are made up of a number of inter-connected lines bounding a certain area and are used to show area features such as administrative units (countries, districts, states, blocks); land use types (cultivated area, forest lands, degraded/waste lands, pastures, etc.) and features like ponds, lakes, etc.

Non–Spatial Data

The data describing the information about spatial data are called as non-spatial or attribute data. For example, if you have a map showing positional location of your school you can attach the information such as the name of the school, subject stream it offers, number of students in each class, schedule of admissions, teaching and examinations, available facilities like library, labs, equipments, etc. In other words, you will be defining the attributes of the spatial data. Thus, non-spatial data are also known as attribute-data.

Sources of Geographical Data

The geographical data are available in analogue (map and aerial photographs) or digital form (scanned images).

The procedure of creating spatial data in the computer has been discussed in Chapter 6.

Mapping Software and their Functions

There are a number of commercially available mapping softwares such as ArcGIS, ArcView, Geomedia, GRAM, Idrisi, Geometica, etc. There are also a few freely downloadable softwares that can be downloaded with the help of Internet. However, it would be difficult to discuss the capabilities of each one of these softwares under the constraints of time and space. We will, therefore, describe the procedure in general used in choropleth mapping using a mapping software.

A mapping software provides functions for spatial and attribute data input through onscreen digitisation of scanned maps, corrections of errors, transformation of scale and projection, data integration, map design, presentation and analysis.

A digitised map consists of three files. The extensions of these files are shp, shx and dbf. The dbf file is dbase file that contains attribute data and is linked to shx and shp files. The shx and shp files, on the other hand, contain spatial (map) information. The dbf file can be edited in MS Excel.

You can construct a choropleth map using any of the mapping software available to you, provided you follow the steps given in the user manual of the given software. If you experiment with the different options available in the software, you would be able to construct several types of maps using different methods.

Excercises

1. Choose the correct option for the alternatives given below :

(i) What type of graph would you use to represent the following data?

States	***Share of Production of Iron-Ore (in %)***
Madhya Pradesh	23.44
Goa	21.82
Karnataka	20.95
Bihar	16.98
Orissa	16.30
Andhra Pradesh	0.45
Maharashtra	0.04

(a) Line (b) Multiple bar graph
(c) Pie-diagram (d) None of the above

(ii) Districts within a state would be represented in which type of spatial data?
(a) Points (b) Lines
(c) Polygons (d) None of the above

(iii) Which is the operator that is calculated first in a formula given in a cell of a worksheet?
(a) + (b) –
(c) / (d)

(iv) Function Wizard in Excel enables you to:
(a) Construct graphs
(b) Carry out mathematical/ statistical operations
(c) Draw maps
(d) None of the above

2. Answer the following questions in about 30 words:

(i) What are the functions of different parts of a computer?

(ii) What are the advantages of using computer over manual methods of data processing and representation?

(iii) What is a worksheet?

3. Answer the following questions in about 125 words:

(i) What is difference between spatial and non-spatial data? Explain with examples.

(ii) What is the three forms of geographical data?

Activity

1. Carry out the following steps using the given data set:
   (a) Enter the given data in a file and store in 'My Documents' folder (Name the file as rainfall).
   (b) Calculate the standard deviation and mean for the given data set using Function Wizard in Excel spreadsheet.
   (c) Compute coefficient of variation using the results derived in step (b).
   (d) Analyse the results.
2. Represent the data given below using a suitable technique with the help of a computer and analyse the graph.

Cropping Intensity in India

Year_80s	*CI_80s*	*Year_90s*	*CI_90s*
1980-81	123.3	1990-91	129.9
1981-82	124.5	1991-92	128.7
1982-83	123.2	1992-93	130.1
1983-84	125.7	1993-94	131.1
1984-85	125.2	1994-95	131.5
1985-86	126.7	1995-96	131.8
1986-87	126.4	1996-97	132.8
1987-88	127.3	1997-98	134.1
1988-89	128.5	1998-99	135.4
1989-90	128.1	1999-00	134.9

5 Field Surveys

You have studied aspects of Physical Geography of the world as well as of India in Class XI. In the present class, besides the Practical Work in Geography you will also study various aspects of Human Geography. While studying these aspects, you may have observed that issues addressed pertain to global or national level. In other words, the given information helps us to understand the issues at macro level. You may also have observed that the forms, events and processes in your surroundings are similar to what you have studied at macro level. Have you ever thought how would you study some of the aspects at local level? You know that the regional level information is used to analyse different physical and human parameters of a large area. Similarly, information has to be gathered at the local level by conducting primary surveys for generating information. The primary surveys are also called **field surveys**. They are an essential component of geographic enquiry. It is a basic procedure to understand the earth as a home of humankind and are carried out through observation, sketching, measurement, interviews, etc. In the present chapter, we will discuss the procedure involved in carrying out the field surveys.

Why is Field Survey Required ?

Like many other sciences, geography is also a field science. Thus, a geographical enquiry always needed to be supplemented through well –planned field surveys. These surveys enhance our understanding about patterns of spatial distributions, their associations and relationships at the local level. Further, the field surveys facilitate the collection of local level information that is not available through secondary sources. Thus, the field surveys are carried out to gather required information so as the problem under investigation is studied in depth as per the predefined objectives. Such studies also enable the investigator to comprehend the situation and processes in totality and at the place of their occurrence. This is possible through 'Observation', which is a useful method of gathering information and then to derive inferences.

Field Survey Procedure

The field survey is initiated with well-defined procedure. It is performed in the following functionally inter – related stages :

1. Defining the Problem

The problem to be studied should be defined precisely. This can be achieved by way of statements indicating the nature of the problem. This should also be reflected in the title and sub-title of the topic of the survey.

2. Objectives

A further specification of the survey is done by listing the objectives. Objectives provide outline of the survey and in accordance to these, suitable tools of acquisition of data and methods of analysis will be chosen.

3. Scope

Like clearly defined objectives, scope of survey needs to be delimited in terms of geographical area to be covered, time framework of enquiry and if required themes of studies to be covered. This multi-dimensional delimitation of the study is essential in relation to fulfilment of the predefined objectives and limitations of analysis, inferences and their applicability.

4. Tools and Techniques

Field survey is basically conducted to collect information about the chosen problem for which varied types of tools are required. These include secondary information including maps and other data, field observation, data generated by interviewing people through questionnaires.

(i) Recorded and Published Data

These data provide base information about the problem. These are collected and published by different government agencies, organisations and other agencies. This information alongwith cadastral and topographical maps, provides basis to prepare the framework of survey. Listing of households, persons, landholdings in the survey area can be done using the official records or electoral rolls available with the village panchayat or the revenue officials. Similarly, essential physical features like relief, drainage, vegetation, land use and cultural features like settlements, transport and communication lines, irrigation infrastructure, etc. can be traced out from the topographical maps. The field boundaries of land parcels can be marked out from cadastral maps available with land revenue officials. The field survey is conducted either for the entire 'population' or for the 'samples'. These basic informations and maps are required to select the units of observation. The large-scale maps of the survey area also help the investigator to orient and locate him/her on the ground. This initial orientation helps the investigator to insert additional features in the map appropriately.

(ii) Field Observation

The effectiveness of field survey is associated with the investigators capability to collect information about the landscape through observation. The very purpose of field survey is to observe the characteristics and associations of geographic phenomena.

To supplement the observation, certain techniques of acquisition of information are very useful like that of sketching and photography. As you find sketches and photographs provided in your textbooks enhance your comprehension of facts, situations and processes being explained. It is, therefore, essential to learn and apply sketching techniques to capture the prominent features of the landscape to strengthen the explanations. Similarly, landscape scenario can also be captured by photography of the landscape, objects and activities.

At times, when suitable large-scale map is not available, a sketch or a notional map of the survey area can be prepared based on reconnaissance survey. This kind of exercise also helps in getting oneself introduced with the area as each feature needs to be observed carefully for locating them in the sketch.

All the observations in the field are to be noted down for keeping a systematic record. You cannot memorise every thing you see, feel or understand. Thus, using appropriate scheme of categorising of facts one should record relevant characteristic of objects. While taking notes, a brief interaction with the people or with the members of the field party or referring to recorded information is always required for clarifications and unambiguous recording of observations.

(iii) Measurement

Some of field surveys demand on site measurement of objects and events. This is all the more necessary when one wants to present the analysis with precision. It involves use of appropriate equipments, which enables the investigator to measure the characteristics precisely. Thus, the field party should carry with them relevant equipment required to measure the selected features such as measuring tape, weighing machine to weigh soil, pH meter or paper strip to measure the acidity or alkalinity and thermometer.

(iv) Interviewing

In all field surveys, dealing with social issues information is gathered through personal interviews. Experiences and knowledge of each individual about his/her environs as well as about his/her own livings are nothing but information. These experiences, if retrieved efficiently are important sources of information. However, extraction of information through personal interviews is greatly influenced by interviewer's abilities in terms of understanding of the subject and the people to be interviewed, communicative skills and rapport with the people.

(a) *Tools :* Interviewing of people can be done either through pre-structured questionnaires and schedules or through participatory appraisal methods like social and resource mapping and discussions.

(b) *Basic Information :* While conducting interviews as means of data collection, certain information like that of location, socio-economic background of the respondent are to be noted. On the basis of these parameters, investigator categorises and compiles the information for further computations and analysis.

(c) *Coverage :* During field studies, investigator has to decide whether the survey will be conducted in the form of census for the entire population or will be based on selected sample. If the study area is not very large but composed of diverse elements then entire population should be surveyed. In case of large size area, one can limit the study to selected samples representing all segments of the population.

(d) *Units of Study :* Elements of study need to be defined precisely alongwith the decision about census or sample survey. These elements consist of primary unit of observation like households, parcels of land, business units, etc.

(e) *Sample Design :* A framework of sample survey including its size and method of selecting samples is to be decided in relation to objectives of survey, variations in population and cost and time constraints.

(f) *Cautions :* Field interviews or participatory appraisal methods are highly sensitive activities and should be conducted with utmost sincerity and cautions as one is dealing with human groups which always do not share the cultural ethos and practices that of the investigators. As a student of social science, you should be careful of the larger purpose of the study and should not stretch the argument beyond the scope of the study. To get the correct picture your conversation and behaviour should reflect that you are one of them. While conducting the interview ensure that no other person is intervening in your conversation either by his presence or reply in between.

5. Compilation and Computation

You need to organise the information of varied types collected during the fieldwork for their meaningful interpretation and analysis to achieve the set objectives. Notes, field sketches, photographs, case studies, etc. are first organised according to sub-themes of the study. Similarly, questionnaire and schedule based information should be tabulated either on a master sheet or on the spreadsheet. You have already learnt the features and use of spreadsheet. You can even construct indicators and compute descriptive statistics.

6. Cartographic Applications

You have learnt different methods of mapping and drawing of diagrams and graphs and also use of computer in drawing them neatly and accurately. For getting visual impressions of variations in the phenomena, diagrams and graphs are very effective tools. Thus, the description and analysis should be duly supported by these presentations.

7. Presentations

The field study report in concise form should contain all the details of the procedures followed, methods, tools and techniques employed. The major part of the report will be devoted to the interpretation and analysis of information gathered and computed alongwith supportive facts in the form of tables, charts, statistical inferences, maps and references. At the end of the report, you should also provide the summary of the investigation.

On the basis of above outlines, you will select a problem or topic and carry out the fieldwork as a team of investigators in the supervision of your teacher.

Field Survey : Case Studies

You know that the field survey plays a significant role in understanding the forms, processes and events at local level. A field survey may be conducted to study any issue of general concern. However, the selection of a topic for the case study depends upon the nature and character of the area where the survey is to

be carried out. For example, in low rainfall and agriculturally less productive regions, droughts form a major topic of study. On the other hand, in the States like Assam, Bihar and West Bengal, which experience high rainfall conditions and occurrences of frequent floods during rainy season necessitates a survey for the assessment of the damages caused by the floods. Similarly, a case study on air pollution emerges as a major topic near a smog emitting industrial plant or a survey of the changing patterns of agricultural land use in Punjab and western Uttar Pradesh, which has drawn the benefits of the Green Revolution for several years becomes important. In the present chapter, we will discuss how specific case studies on droughts, and poverty are conducted. These have been selected from case studies given in your syllabus. These are :

1. Ground Water Change
2. Environmental Pollution
3. Soil Degradation
4. Poverty
5. Droughts and Floods
6. Energy Issues
7. Land use survey and Change Detection.

A summary of the procedure that could be followed in carrying out the field survey on any of these topics is provided in *Table 5.1*.

Instructions for the Students

The students should prepare a blue print of the field survey in consultation with the class teacher to include details of the area to be visited alongwith a map of the area, if available, clear understanding of the objectives of the survey and the well-structured questionnaire. The teacher should also give a few necessary instructions to the students. These include :

1. Be courteous to the people of the area, you are visiting for the field survey.
2. Develop friendly attitude with the people you meet and establish rapport.
3. Ask questions in comprehendible language.
4. Avoid asking the questions that either may hurt the feelings of the people you are interviewing or those that may irritate them.
5. Do not make any promises with the inhabitants of the area and do not tell lies about your purpose.
6. Record each and every detail as given by the respondent of your querries and show them the recorded version if so asked for.

Field Study of Poverty: Extent, Determinants and Consequences

The Problem

Poverty denotes a state of people in terms of income, assets, consumption or nutrition at a given point of time. It is often understood and conveyed in the context of **poverty line**, which is a critical threshold level of income, consumption, access to productive resources, and services below which people are classed as poor.

The issue of poverty is closely linked with inequality, which is the cause of poverty. Poverty is, thus, not only an absolute but also a relative state. It varies

from region to region. However, there is something absolute about it and despite the variations in regions and diversified society, people require adequate levels of food, clothing and shelter. Poverty can be either a chronic or temporary phenomena. The chronic poverty, which is also known as structural poverty, is more crucial to be understood. Another significant aspect of poverty is that in spite of high rate of economic growth more and more people are identified below the poverty line. It is rampant in both rural and urban areas alike. Thus, the dimensions of poverty and its measures could be studied through a field survey. *Fig. 5.1* and *5.2* provide a glimpse of poverty-ridden families and the villages.

Fig. 5.1 : *A Poverty-ridden family*

Fig. 5.2 : *A Poverty-ridden Village*

The first step to conduct such a survey is listing of its objectives.

Objectives

The study of extent, determinants and consequences of poverty can be carried out with the following objectives in mind:

1. To identify appropriate criteria to measure poverty line.
2. To assess the levels of well-being of people in terms of income, assets, expenditure, nutrition, access to resources and services.
3. To explain the state of poverty in relation to historical and structural conditions of the village and its people.
4. To examine the implications of poverty.

Coverage

The spatial, temporal and thematic aspects of survey be understood clearly.

Spatial

In order to achieve the aforesaid objectives a field study may be conducted in a selected part of the rural or urban settlements. Spatially, it may cover an area of 200 hectares or more and inhabited by about 400 persons or 100 households.

Temporal

If the problem pertains to chronic poverty, the study should be based on average conditions or reflecting responses with references to normal rainfall year for the

village as well as for the surrounding area. In case of temporary poverty, current year situations are to be investigated.

Thematic

Thematically, the study should cover household and individual level aspects like socio-demographic characteristics, permanent and consumer assets, income and expenditure, access to health, educational, transport and power services and infrastructure facilities to capture the targeted issues of status, determinants and implications of poverty.

Tools and Techniques

Secondary Information

Before you proceed for field study, you should go through the literature on poverty and the region in general and the selected village in particular. The conceptual aspects of poverty like its meaning, measurement, criteria, causes, etc. can be understood through published work related to economic development, social changes and economic surveys. Basic population statistics can be obtained from district census handbooks or the village level primary census abstract, agricultural and livestock statistics can be acquired from village revenue official or the *Patwari Lekhpal, Karamchari, Karnain,* etc. Household lists and other village level information can be collected from *Gram Panchayat* office. Similarly, other relevant data are available with respective departments located at tehsil or block headquarters. All these informations are essential to build up the framework of the village resources and economy as well as to develop research design including sample design if survey is not to be based on entire population.

Maps

Topographic details including relief, drainage, water bodies, settlements, means of communications and other topographical features of a village and its surrounding region are to be traced and studied from 1:50,000 or 1:25,000 scale topographical maps. Similarly, the 1:4,000 scale cadastral maps and revenue records of the villages may be obtained from the revenue officials. These maps provide spatial dimension of inequality in land distribution if plotted by ownership of households.

Observations

As a fundamental tool of field survey, much of the details of poverty scenario can be visualised through keen observation. Observations of the routine activities of the poverty ridden people; quality and quantity of the food items; sources of fuel wood and drinking water; state of clothing and shelter human sufferings associated with malnutrition, hunger, sickness, etc.; locational, social and political deprivations due to poverty and other pertinent attributes can be understood. These observations with aids like photography, sketching, audio-visual recordings, etc. or just in the form of notes are valuable source of non-quantifiable information to validate different point of views and to draw conclusions.

Measurement

In some situation, actual measurement need to be taken up. This is required in case of non-availability of data pertaining to quantity of food items consumed

daily or the state of health in terms of height and weight, quality of drinking water or the nutritional value of different food items, availability of living space, etc. Simpler means of measurement are very fruitful in quantifying certain items precisely.

Personal Interview

Most of poverty measures are based on aggregate household conditions. Thus, field data collection through interviewing will be at household level. However, information about the household will have to be extracted either from the head of the household or the more responsive and knowledgeable member of the household. Apart from canvassing questionnaires household data will also be collected interviewing village leaders, service providers, institutional heads, etc. to compute relevant indices.

Survey Design

Survey can be conducted, as census covering all the households of the village if the number of household are manageable with the number of students in the class otherwise a stratified sampling will be appropriate to extract information. Stratification of households can be done on the basis of land holdings classes, social classes, division of settlement into grids or concentric circles. For stratification listing of households alongwith these criteria/attributes and notional map showing the plan of settlement are to be completed as follows :

Table 5.1 : List of Households with Basic Attributes of Stratification of Sampling

S. No.	Head of Household with Father's Name	Social Class/ Category	Land Holding (ha)	Location of House (Grid/ Circle Reference)	Remarks
1.	Mohanlal S/o Sohanlal	Dhaker : OBC	7.2	A2	
2.	Homaji S/o Kaluji	Bheel : ST	0.2	D4	
3.	.....	.....	.....	.....	

Grids or circles in the notional map/plan may be drawn for spatial stratification as shown in *Fig. 5.3*.

Schedule/Questionnaire

Interview, observation and at times, measurement based household information is to be enquired and recorded systematically in the pre-designed questionnaire (Please see Annexure 1 A to H).

Compilation and Computation

Data Entry and Tabulation

After completing the survey in the field collected information need to be compiled for further computation and analysis. Now, this task can be accomplished more conveniently in spreadsheet formats, which you have already practised as part of your computer related practical work. For efficient management of these data follow the following sequence :

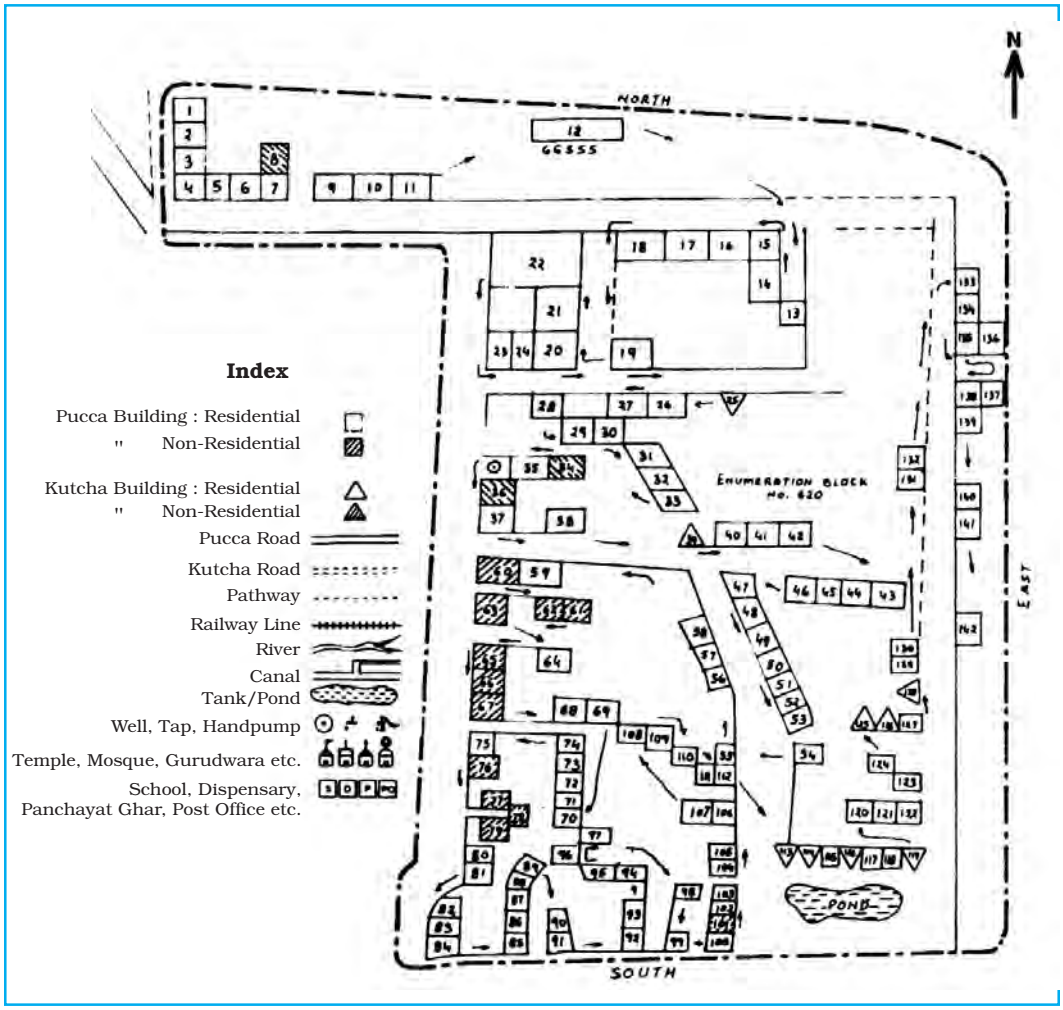


Fig. 5.3 : *Notional Map of the Settlement with Grids for Sampling*

1. Assign unique identity code to each surveyed household.
2. Similarly, each person in the demographic table will also be assigned unique identity code for compilation in separate spreadsheet.
3. It will be more convenient if each type of household level information is compiled on separate sheet.
4. Unique name to be assigned for each attribute in each column.
5. Information on each sheet will be filled according to household code for further processing.

Verification and Consistency Checks

After data entry, random verification of entries is necessary to ascertain the correctness of data. This is further checked by cross total and with the help of maximum and minimum values as well as in the light of related variables.

Computation of Indices

Computation of indices using available value parameters and calculating the ratios is a significant task before analysing the situation of poverty. In this regard,

following set of indices may be computed at household level for further analysis :

1. Indices indicating the state of well-being measured on the basis of total assets, total income, total expenditure, food consumption, nutrition level, etc.
2. Indices explaining the reasons of chronic poverty like social class membership and perpetuating legacies, size of household, type of family, type of occupations, educational levels, size of land holdings and state of irrigation, type of crops cultivated, subsidiary sources of employment, ownership of productive assets, state of gender equality, etc.
3. Indices related to consequences of poverty can be computed on the basis of state of gender discrimination, literacy and educational level among the youths and young ones, employment diversification, productive and consumer assets, crop yields, pattern of expenditure and nutritional intakes.

It is significant to note that many of the causative factors are also resultant facts due to their circular relationship with poverty.

Visual Presentation

Summarised tables, diagrams and graphs as you learnt as part of cartographic work can be employed to represent the salient characteristics of poverty in the village. For this purpose, tables may be prepared according to land holding categories or the social categories of households including the caste based classifications. Similarly, composite indices of productive assets or total expenditure can be used to segregate households for showing their state of well-being. Variations in well-being can also be shown in the form of drawing a poverty line and class-wise distribution of households above and below that line to visualise the poverty-affected sections of the society and their social background. A very significant graphical tool to indicate the inequality is Lorenz curve and it can be drawn to show unequal distribution of assets, income and expenditure among the households of the village.

Thematic Mapping

Spatial distribution of agricultural as well as non-agricultural land within the revenue limits of the village and in the settlement can be shown by chorochromatic maps to assess the extent of control on natural resources of certain social groups as a source of inequality, and one of the important causes of poverty. Poor accessibility in relation to site of houses and location of services can also be visualised with the help of maps.

Statistical Analysis

Simple descriptive statistical methods as well as measures of associations, explanatory relationships and composite indices based on household level indicators can be employed meaningfully to draw inferences. In this regard, simple arithmetic mean can indicate the average situation whereas the coefficient of variation will indicate the extent of relative inconsistency in socio-economic well-being among different groups of households. Similarly, you can measure the intensity of relationship between two indices using the coefficient of correlation and explain the probable causes of perpetuation of poverty or its impact on other socio-economic aspects.

Report Writing

Finally, using all the analysed material, you will present your report in group or individually as instructed by your teacher in the systematic way as you followed in the investigation of the problem. All the details, we discussed till now will be part of your presentation in the same sequence alongwith major conclusions and inferences you have drawn. You will also enrich your presentation with appropriate illustrations including maps, diagrams, graphs, photographs, sketches, etc. The statement in the text will be duly supported by the facts shown in tabular forms as well as references of earlier works.

Field Study of Droughts : A Study of Belgaum District, Karnataka

Some of the regions in India have plenty of water, and shortages are rare. But in many parts of the country, water is scarce and people can never be sure when it will rain next. Droughts happen when for months or even years, the earth's surface loses more water than it collects. In some places of deserts, it almost never rains at all. Droughts can affect many peoples' lives.

Droughts and floods are two adverse factors, which Indian farmers have to face. A specific definition of any one of them is quite difficult. However, qualitatively, agricultural drought can be defined as a prolonged and acute moisture deficiency.

Drought, as commonly understood, is a condition of climatic dryness that is severe enough to reduce soil moisture and water below the minimum limit necessary for sustaining plant, animal and human life (Fig 5.4 and 5.5). It is usually accompanied by hot dry winds and may be followed by damaging floods.

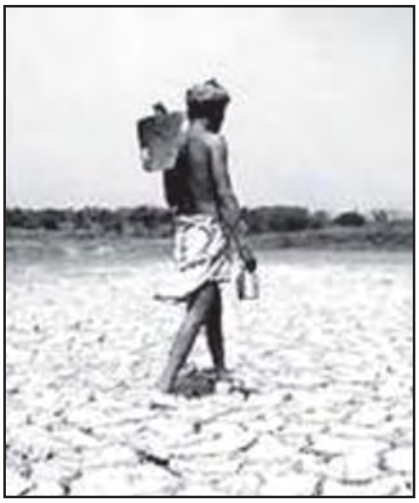


Fig. 5.4 : *Drought Affected Area*

Fig. 5.5 : *Soil Moisture Loss*

Drought has been recognised as one of the main causes of human misery. While generally associated with semiarid or desert conditions, drought can occur in areas that normally enjoy adequate rainfall and moisture levels. In the broadest sense, any lack of water for the normal needs of agriculture, livestock, industry, or human population may be termed a drought. The cause may be lack of supply, pollution of the water, inadequate storage, conveyance facilities, or abnormal demand.

The effects of drought depend on its severity and duration and the size of the affected area. The impact depends on the level of socio-economic development. Societies that are more developed and economically diversified can better adjust to a drought and can recover more quickly. The poor regions, especially those reliant on any crop or pastoral economies, are more severely affected.

The worst effects of drought are the dramatic reduction of surface water and loss of food. Crop failures cause a chain reaction of human suffering (hunger and malnutrition) and economic difficulties. In developing countries, these conditions can culminate in a large number of starvation deaths and farmers' suicides.

Objectives

A field survey for the assessment and magnitude of the droughts can be carried out with the following objectives in mind :

(a) To identify and record areas experiencing recurring drought conditions.

(b) To get the first hand experience of droughts as a natural disaster.

(c) To suggest drought preparedness measures for the people of the area.

Coverage

The aspects related to the spatial, temporal and thematic coverage be understood.

Spatial

In order to achieve the aforesaid objectives, a field study may be conducted of a drought prone area, if it has experienced drought in or around your district.

Temporal

If the problem pertains to recurring droughts, the study should be based on average conditions reflecting responses with references to normal rainfall year for affected area and its surroundings. Besides, the data on agricultural productions for the drought years may be compared with the non–drought year production figures.

Thematic

Thematically the assessments for the agricultural production and crop land use, rainfall variability and vegetation status should be made to understand the magnitude, determinants and implications of the droughts.

Tools and Techniques

Secondary Information

The maps and the data pertaining to the rainfall, crop production and population should be collected for drought affected areas for the drought years from the following government/ quasi-government offices :

(i) *Indian Daily Weather Reports*, Indian Meteorological Department (IMD), Division of Agricultural Meteorology, Pune

(ii) *Crop Weather Calendar*, IMD, Agrimet Division, Pune

(iii) Government of Karnataka, *Belgaum District Gazetteers*, Bangalore 1987

(iv) *Census Handbooks*, Census of India, New Delhi

(v) *District Handbook/Village Directories,* Government of Karnataka

(vi) *Statistical Abstracts*, Bureau of Economics and Statistics, Government of Karnataka, Bangalore.

Maps

1 : 50,000 and large-scale topographical maps of the drought affected areas enable the identification and mapping of the perennial and non-perennial water bodies, settlements, land use, and other physical and cultural features. Besides, the cadastral maps help in collecting the data about land use.

Observation

Observation means looking around and talking to people and noting down the observed information about the shortage of water, crop failures, lack of fodder, starvation deaths, farmer's suicides, if any.

(a) *Targeted Objects and Processes :* A detailed study of the changes in the crop land use pattern of the selected village as well as major rivers, streams, nalla, tanks and wells and irrigation facilities, if any, should be made in the light of the drought situation.

(b) *Photographs and Sketches :* Photographs and sketches of the parched lands, people and livestock can give a qualitative touch to the study if carried out during the field survey.

Measurement

Objects (to be measured)

The village, as a unit, is selected for this type of survey. A cadastral map is obtained from the village *patwari*. This map shows the *Khasra* numbers and boundaries of the fields. Some copies of the outline map are prepared and information filled in. These include the wells, tanks, and streams in terms of depth of water, limits of perennial water in larger streams; sowing in the total number of fields, loss of seeds, harvesting; availability of drinking water facilities; official relief measures, etc.

Interviewing

The questionnaire method involves asking previously framed questions to the person to be interviewed. The surveyor has to ask the question and take down the answer if it is a structured questionnaire. The questions should be related to the drought and economic conditions of the farmers in terms of amount of rainfall received, rainy days, sowing, watering, nature of crops, livestock and fodder, domestic water supply, health care, rural credit and employment and anti-poverty programmes of the government. The degree of feeling of the respondent can be noted on a five–point scale (very good, good, satisfactory, bad and very bad).

Tabulation

The data collected from the primary and secondary sources has to be organised in a systematic manner for easy processing and interpretation. Different methods are used to quantify the data into groups or heading such as the tally mark method.

Presentation of Report

The information gathered during field survey is finally recorded in the form of a detailed report about the cause and magnitude of the drought and its impact on the economy and life of the people.

Excercises

1. Choose the right answer from the four alternatives given below :

(i) Which one of the following helps most in planning for a field survey ?
(a) Personal Interviews (b) Secondary Information
(c) Measurements (d) Experimentation

(ii) Which one of the following is taken up at the conclusion of a field survey ?
(a) Data entry and Tabulation (b) Report Writing
(c) Computation of Indices (d) None of the above

(iii) What is most important at the initial stages of field survey ?
(a) Outlining the Objectives
(b) Collection of Secondary Information
(c) Defining the spatial and thematic coverages
(d) Sample Design

(iv) What level of information is acquired during a field survey ?
(a) Macro level information
(b) Maso level information
(c) Micro level information
(d) All of the above levels of information

2. Answer the following questions in about 30 words :

(i) Why is a field survey required ?
(ii) List the tools and techniques used during a field survey.
(iii) What type of coverages need to defined before undertaking a field survey?
(iv) Describe survey design in brief.
(v) Why is the well-structured questionnaire important for a field survey ?

3. Design a field survey on any one of the following problems :

(a) Environmental Pollution
(b) Soil Degradation
(c) Floods
(d) Energy Issues
(e) Land Use Change Detection

6 Spatial Information Technology

You know that the computers enhance our capabilities in data processing and in drawing graphs, diagrams and maps (Refer Chapter 4 of the present book). The disciplines that deals with the principles and methods of data processing and mapping using a combination of computer hardware and the application software are referred as the **Database Management System** (DBMS) and the **Computer– Assisted Cartography** respectively. However, the role of such computer applications is restricted to merely processing of the data and their graphical presentation. In other words, the data so processed or the maps and diagrams so prepared could not be used to evolve a decision support system. As a matter of fact, there are several questions that we normally encounter in our day-to-day life and look for satisfactory solutions. These questions may be : What is where ? Why is it there ? What will happen if it is shifted to a new location ? Who will be benefited by such a reallocation? Who are expected to loose the benefits if reallocation takes place? In order to, understand these and many other questions we need to capture the necessary data collected from different sources and integrate them using a computer that is supported by geo–processing tools. Herein lays the concept of a **Spatial Information System**. In the present chapter, we will discuss basic principles of the **Spatial Information Technology** and its extension to the Spatial Information System, which is more commonly known as **Geographical Information System**.

What is Spatial Information Technology?

The word **spatial** is derived from **space**. It refers to the features and the phenomena distributed over a geographically definable space, thus, having physically measurable dimensions. We know that most data that are used today have spatial components (location), such as an address of a municipal facility, or the boundaries of an agricultural holdings, etc. Hence, the Spatial Information Technology relates to the use of the technological inputs in collecting, storing, retrieving, displaying, manipulating, managing and analysing the spatial information. It is an amalgamation of Remote Sensing, GPS, GIS, Digital Cartography, and Database Management Systems.

What is GIS (Geographical Information System)?

The advance computing systems available since mid 1970's enable the processing of georeferenced information for the purpose of organising spatial and attribute data and their integration; locating specific information in individual files and executing the computations, performing analysis and evolving a decision support system. A system capable of all such functions is called Geographic Information System (GIS). It is defined as **A system for capturing, storing, checking, integrating, manipulating, analysing and displaying data, which are spatially referenced to the Earth. This is normally considered to involve a spatially referenced computer database and appropriate applications software.** It is an amalgamation of Computer Assisted Cartography and Database Management System and draws conceptual and methodological strength from both spatial and allied sciences such as Computer Science, Statistics, Cartography, Remote Sensing, Database Technology, Geography, Geology, Hydrology, Agriculture, Resource Management, Environmental Science, and Public Administration.

Forms of Geographical Information

As discussed in Chapter 4, two types of the data represent the geographical information. These are spatial and non – spatial data (*Box 6.1*). The spatial data are characterised by their positional, linear and areal forms of appearances (*Fig. 6.1*).

Box 6.1 : Spatial and non-spatial data

Stock Register of a Cycle shop			Literate Population in States 1981		
Part No.	***Quantity***	***Description***	***State***	***% Male***	***% Female***
101435	54	Wheel Spoke	Kerala	75.3	65.7
108943	68	Ball Bearing	Maharashtra	58.8	34.8
105956	25	Wheel Rim	Gujarat	54.4	32.3
123545	108	Tyre	Punjab	47.2	33.7

Geographic Database : A database contains attributes and their value or class. The non-spatial data on the left display cycle parts, which can be located anywhere. The data record on the right is spatial because one of the attributes, the name of different states, which have a definite locations in a map. This data can be used in GIS.

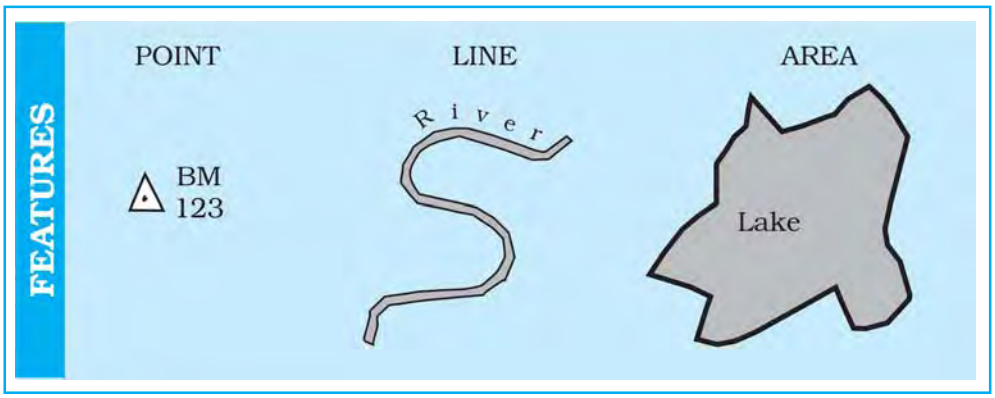


Fig. 6.1 : *The Point, a Line and an Area Feature*

These data forms must be geometrically registered to a generally accepted and properly defined coordinate system and coded so that they can be stored in the internal data base structure of GIS. On the other hand, the data those describe the spatial data are called as **Non–Spatial** or attribute data. The spatial data are the most important pre-requisite in a spatial or geographical information system. In a GIS core, it could be built in several ways. These are :

- Acquire data in digital form from a data supplier
- Digitise existing analogue data
- Carry out one's own surveys of geographic entities.

The choice of a source of geographical data for a GIS application is, however, largely governed by :

- The application area in itself
- The available budget, and
- The type of data structure, i.e. vector/raster.

For many users, the most common source of spatial data is topographical or thematic maps in hard copy (paper) or soft copy form (digital). All such maps are characterised by :

- A definite scale which provides relationship between the map and the surface it represents,
- Use of symbols and colours which defines attributes of entities mapped, and
- An agreed coordinate system, which defines the location of entities on the Earth's surface.

Advantages of GIS over Manual Methods

The maps, irrespective of a graphic medium of communication of geographic information and possessing geometric fidelity, are inherited with the following limitations :

(i) Map information is processed and presented in a particular way.

(ii) A map shows a single or more than one predetermined themes.

(iii) The alteration of the information depicted on the maps require a new map to be drawn.

Contrarily, a GIS possesses inherent advantages of separate data storage and presentation. It also provides options for viewing and presenting the data in several ways. The following advantages of a GIS are worth mentioning :

1. Users can interrogate displayed spatial features and retrieve associated attribute information for analysis.
2. Maps can be drawn by querying or analysing attribute data.
3. Spatial operations (Polygon overlay or Buffering) can be applied on integrated database to generate new sets of information.
4. Different items of attribute data can be associated with one another through shared location code.

Components of GIS

The important components of a Geographical Information System include the followings :

(a) Hardware (b) Software
(c) Data (d) People

The different components of GIS are shown in *Fig. 6.2*.

Hardware

As discussed in Chapter 4 the GIS has three major components :

- Hardware comprising of the processing storage, display, and input and output sub-systems.
- Software modules for data entry, editing, maintenance, analysis, transformation, manipulation, data display and outputs.
- Database management system to take care of the data organisation.

Software

An application software with the following functional modules is important prerequisite of a GIS :

- Software related to data entry, editing and maintenance
- Software related to analysis/transformation/manipulation
- Software related to data display and output.

Data

Spatial data and related tabular data are the backbone of GIS. The existing data may be acquired from a supplier or a new data may be created/collected in-house by the user. The digital map forms the basic data input for GIS. Tabular data related to the map objects can also be attached to the digital data. A GIS will integrate spatial data with other data resources and can even use a DBMS.

People

GIS users have a very wide range from hardware and software engineers to resources and environmental scientists, policy makers, and the monitoring and implementing agencies. These cross-section of people use GIS to evolve a decision support system and solve real time problems.

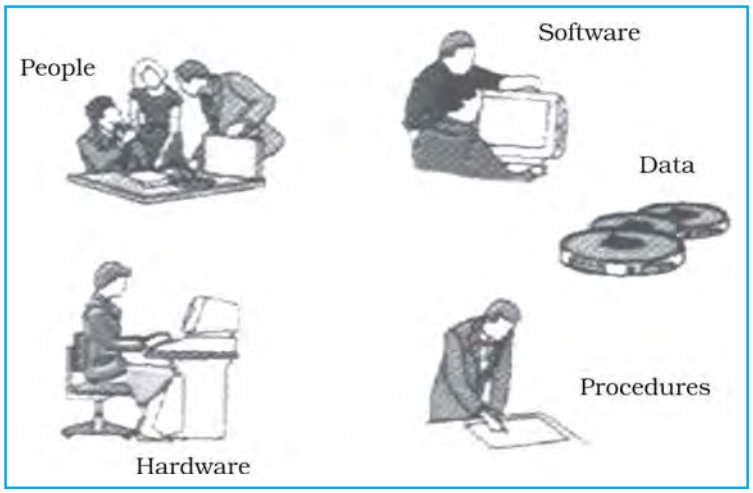


Fig. 6.2 : *Basic Components of GIS*

Spatial Data Formats

The spatial data are represented in raster and vector data formats :

Raster Data Format

Raster data represent a graphic feature as a pattern of grids of squares, whereas vector data represent the object as a set of lines drawn between specific points. Consider a line drawn diagonally on a piece of paper. A raster file would represent this image by sub-dividing the paper into a matrix of small rectangles, similar to a sheet of graph paper called cells. Each cell is assigned a position in the data file and given a value based on the attribute at that position. Its row and column coordinates may identify any individual pixel (*Fig. 6.3*). This data representation allows the user to easily reconstruct or visualise the original image.

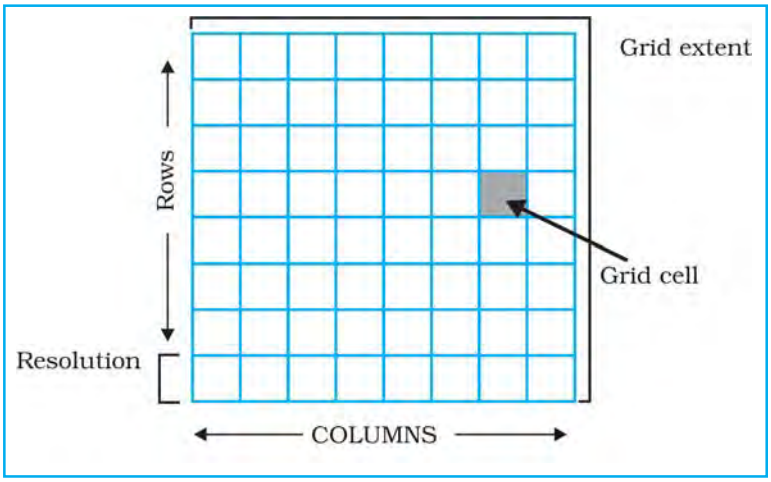


Fig. 6.3 : *Generic Structure for a Grid*

The relationship between cell size and the number of cells is expressed as the **resolution** of the raster. The effect of grid size on data in raster format is explained in *Fig. 6.4*.

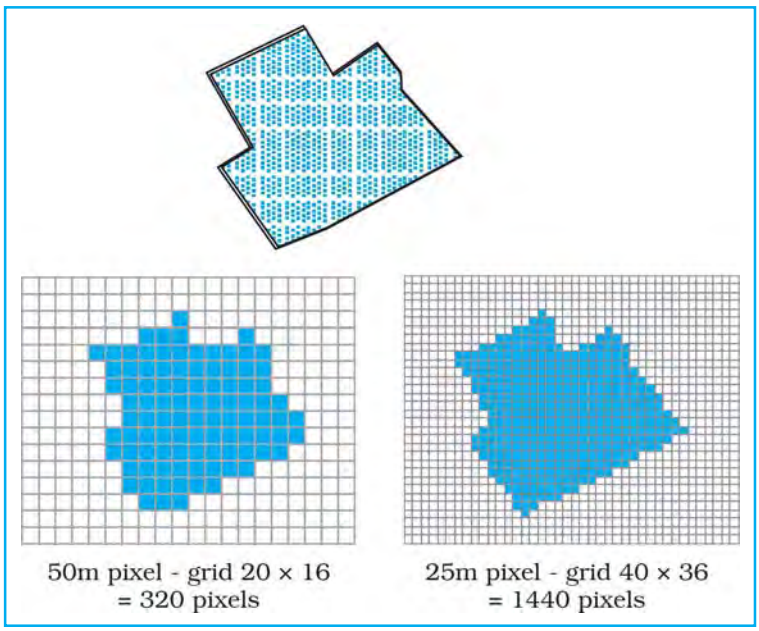


Fig. 6.4 : *Effect of Grid Size on Data in Raster Format*

The Raster file formats are most often used for the following activities :

- For digital representations of aerial photographs, satellite images, scanned paper maps, and other applications with very detailed images.
- When costs need to be kept down.
- When the map does not require analysis of individual map features.
- When "backdrop" maps are required.

Vector Data Format

A vector representation of the same diagonal line would record the position of the line by simply recording the coordinates of its starting and ending points. Each point would be expressed as two or three numbers (depending on whether the representation was 2D or 3D, often referred to as X,Y or X,Y,Z coordinates) (*Fig. 6.5*). The first number, X, is the distance between the point and the left side of the paper; Y, the distance between the point and the bottom of the paper; Z, the point's elevation above or below the paper. Joining the measured points forms the vector.

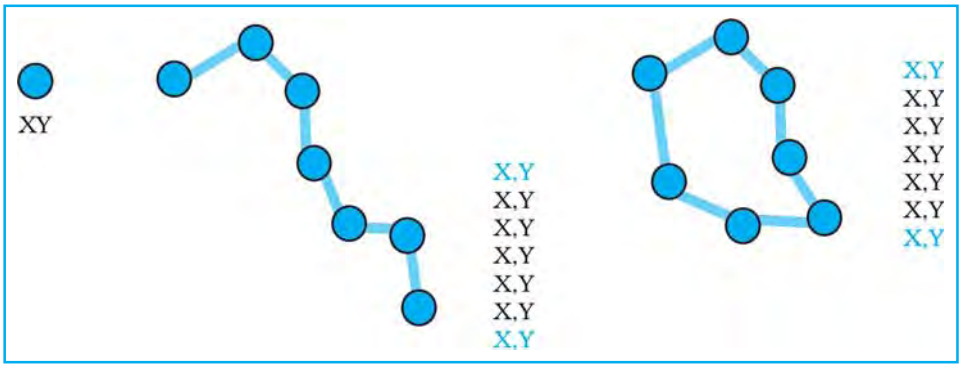


Fig. 6.5 : *The Vector Data Model is based around Coordinate Pairs*

A vector data model uses points stored by their real (earth) coordinates. Here lines and areas are built from sequences of points in order. Lines have a direction to the ordering of the points. Polygons can be built from points or lines. Vectors can store information about topology. Manual digitising is the best way of vector data input.

The Vector files are most often used for :

- Highly precise applications
- When file sizes are important
- When individual map features require analysis
- When descriptive information must be stored

The advantages and the disadvantages of the raster and vector data formats are explained in *Box 6.2*.

Box 6.2 : Comparison of Raster and Vector Data Formats

Raster Model	*Vector Model*
Advantages • Simple data structure • Easy and efficient overlaying • Compatible with RS imagery • High spatial variability is efficiently represented • Simple for own programming • Same grid cells for several attributes	**Advantages** • Compact data structure • Efficient for network analysis • Efficient projection transformation • Accurate map output
Disadvantages • Inefficient use of computer storage • Errors in perimeter and shape • Difficult network analysis • Inefficient projection transformations • Loss of information when using large cells Less accurate (although interactive) maps	**Disadvantages** • Complex data structure • Difficult overlay operations • High spatial variability is inefficiently represented • Not compatible with RS imagery

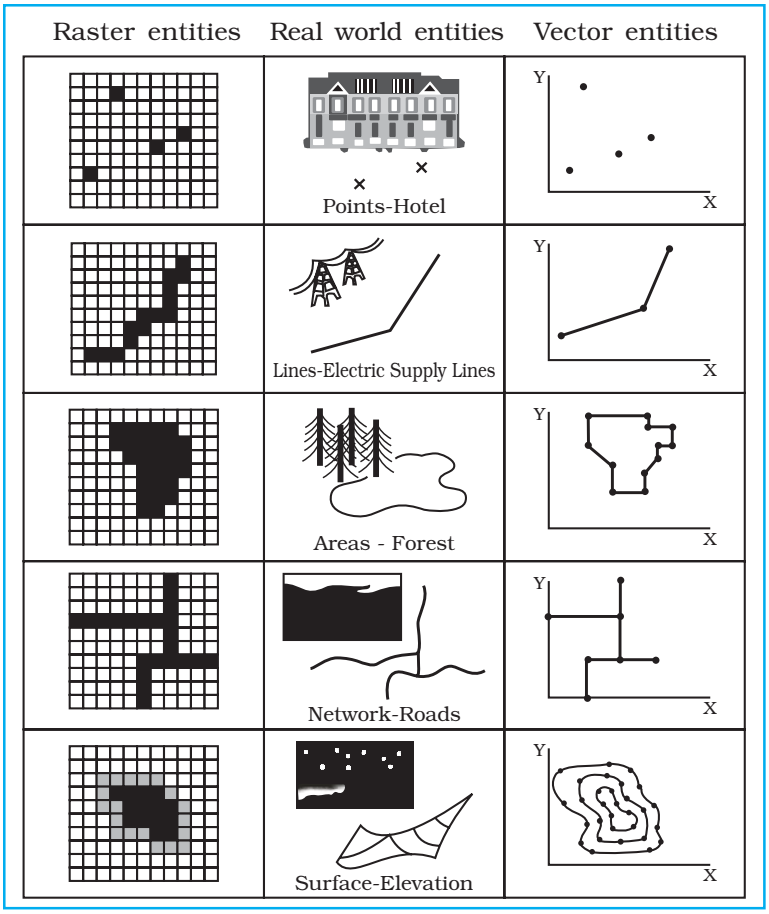


Fig. 6.6 : Representation of Spatial Entities in Raster and Vector Data Formats

Sequence of GIS Activities

The following sequence of the activities are involved in GIS related work :

1. Spatial data input
2. Entering of the attribute data
3. Data verification and editing
4. Spatial and attribute data linkages
5. Spatial analysis

Spatial Data Input

As already mentioned, the spatial database into a GIS can be created from a variety sources. These could be summarised into the following two categories :

(a) *Acquiring Digital Datasets From a Data Supplies*

The present day data supplies make the digital data readily available, which range from small-scale maps to the large-scale plans. For many local governments and private, such data form an essential source and keep such groups of users free from overheads of digitising or collecting their own data. Although, using such existing data sets is attractive and time saving, serious attention must be paid to data compatibility when data from different sources/supplies are combined in one project. The differences in terms of projection, scale, base level and description in attributes may cause problems.

At a practical level, users must consider the following characteristics of the data to ensure that they are compatible with the application:

- The scale of the data
- The geo-referencing system used
- The data collection techniques and sampling strategy used
- The quality of data collected
- The data classification and interpolation methods used
- The size and shape of the individual mapping units
- The length of the record.

It must also be noted that where data are used from a number of sources, and particularly where the area of study crosses administrative boundaries, the difficulties in data integration are caused by different geographical referencing systems, data classification and sampling. Hence, the user needs to be aware of these problems, which are particularly prone when compiling inter province, and inter-district data sets. Once, the compatibility between the data acquired from different suppliers is established, the next stage involves the transfer of data from a medium of transfer to the GIS. The use of DAT tapes, CD ROMS and floppy disks is becoming increasingly common for the purpose. At this stage, the conversion from encoding and structuring system of the source to that of GIS to be used is important.

(b) *Creating digital data sets by manual input*

The manual input of data to a GIS involves four main stages :

- Entering the spatial data.
- Entering the attribute data.
- Spatial and attribute data verification and editing.
- Where necessary, linking the spatial to the attribute data.

The manual data input methods depend on whether the database has a vector topology or grid cell (raster) structure. The most common ways of inputting data in to a GIS are through:

- Digitisers
- Scanners

With the entity model, geographical data are in the form of points, lines and/or polygons (areas)/pixels which are defined using a series of coordinates. These are obtained by referring to the geographical referencing systems of the map or aerial photograph, or by overlaying a graticule or grid onto it. The use of digitisers and the scanners greatly reduce the time and labour involved in writing down coordinates. We shall, briefly, discuss how the spatial data are created in GIS core using a scanner.

Scanners

The scanners are the devices for converting analogue data into digital grid-based images. They are used in spatial data capture to convert a line map to high-resolution raster images which may be used directly or further processed to get vector topology. There are two basic types of scanners :

- Scanners that record data on a step-for -step basis, and
- Those that can scan whole document in one operation.

The first type of scanners incorporates a source of illumination on a movable arm (usually light emitting diodes or a stabilised fluorescent lamp) and a digital camera with high-resolution lamp. The camera is usually equipped with special sensors called Charged Coupled Devices (CCDs) arranged in an array. These are semi-conductor devices that translate the photons of light falling on their surface into counts of electrons, which are then recorded as a digital value.

The movement of either the scanner or the map builds up a digital two-dimensional image of the map. The map to be scanned can be mounted either on a flat bed, or on a rotating drum. With flatbed scanners, the light source is moved systematically up and down over the surface of the document. For large maps, scanners are used which are mounted on a stand and the illumination source and camera array are fixed in a position. The map is moved past by a feeding mechanism. Modern document scanners resemble laser printers in reverse because the scanning surface is manufactured with a given resolution of light sensitive spots that can be directly addressed by the software. There are no moving parts except a movable light source. The resolution is determined by the geometry of the sensor surface and the amount of memory rather than by a mechanical arm.

The scanned image is always far from perfect even with the best possible scanners, as it contains all the smudges and defects of the original map. The excess data, therefore, in a digital image must be removed to make it usable.

Entering the Attribute Data

Attribute data define the properties of a spatial entity that need to be handled in the GIS, but which are not spatial. For example, a road may be captured as a set of contiguous pixels or as a line entity and represented in the spatial part of the GIS by a certain colour, symbol or data location. Information describing the type of road may be included in the range of cartographic symbols. The attribute values associated with the road, such as road width, type of surface, estimated number of traffic and specific traffic regulation may also be stored separately

either as spatial information in the GIS in case of relational databases, or input along with spatial description with the object-oriented data bases.

The attribute data acquired from sources like published record, official censuses, primary surveys or spread sheets can be used as input into GIS database either manually or by importing the data using a standard transfer format.

Data Verification and Editing

The spatial data captured into a GIS require verification for the error identification and corrections so as to ensure the data accuracy. The errors caused during digitisation may include data omissions, and under/over shoots. The best way to check for errors in the spatial data is to produce a computer plot or print of the data, preferably on translucent sheet, at the same scale as the original. The two maps may then be placed over each other on a light table and compared visually, working systematically from left to right and top to bottom of the map. Missing data and locational errors should be clearly marked on the printout. The errors that may arise during the capturing of spatial and attribute data may be grouped as under :

Spatial data are incomplete or double

The incompleteness in the spatial data arises through omissions in the input of points, lines, or polygons/area of manually entered data. In scanned data the omissions are usually in the form of gaps between lines where the raster vector conversion process has failed to join up all parts of a line.

Spatial data at the wrong scale

The digitising at the wrong scale produces input spatial data at a wrong scale. In scanned data, the problems usually arise during the geo-referencing process when incorrect values are used.

Spatial data are distorted

The spatial data may also be distorted if the base maps used for digitising are not scale correct. The aerial photographs, in particular, are characterised by incorrect scale because of the lens distortions, relief and till displacements. In addition, paper maps and field documents used for scanning or digitising may contain random distortions as a result of having been exposed to rain, sunshine and frequent folding. Hence, transformation from one coordinate system to another may be needed if the coordinate system of the database is different from that used in the input document or image.

These errors need corrections through various editing and updating functions as supported directly by most GIS. The process is time-consuming and interactive that can take longer time than the data input itself. The data editing is usually undertaken by viewing the portion of map containing the errors on the computer screen and correcting them through the software using the keyboard, screen cursor controlled by a mouse or a small digitiser tablet.

Minor locational errors in a vector database may be corrected by moving the spatial entity through the screen cursor. In some GIS, computer commands may be used directly to move, rotate, erase, insert, stretch or truncate the graphical entities are required. Where excess coordinates define a line these may be removed using 'weeding' algorithms. Attribute values and spatial errors in raster data

must be corrected by changing the value of the faulty cells. Once, the spatial errors have been corrected, the topology of vector line and polygon networks can be generated.

Data Conversion

While manipulating and analysing data, the same format should be used for all data. When different layers are to be used simultaneously, they should all be in vector or all in raster format. Usually, the conversion is from vector to raster, because the biggest part of the analysis is done in the raster domain. Vector data are transformed to raster data by overlaying a grid with a user-defined cell size.

Sometimes, the data in the raster format are converted into vector format. This is the case especially if one wants to achieve data reduction because the data storage needed for raster data are much larger than for vector data.

Geographic Data : Linkages and Matching

The linkages of spatial and the attribute data are important in GIS. It must, therefore, carefully be undertaken. Linking of attribute data with a non-related spatial data shall lead to chaos in ultimate data analysis. Similarly, matching of one data layer with another is also significant.

Linkages

A GIS typically links different data sets. Suppose, we want to know the mortality rate due to malnutrition among children under 10 years of age in any state. If we have one file that contains the number of children in this age group, and another that contains the mortality rate from malnutrition, we must first combine or link the two data files. Once this is done, we can divide one figure by the other to obtain the desired answer.

Exact Matching

Exact matching means when we have information in one computer file about many geographic features (e.g., towns) and additional information in another file about the same set of features. The operation to bring them together may easily be achieved using a key common to both files, i. e. name of the towns. Thus, the record in each file with the same town name is extracted, and the two are joined and stored in another file.

Hierarchical Matching

Some types of information, however, are collected in more detail and less frequently than other types of information. For example, land use data covering a large area are collected quite frequently. On the other hand, land transformation data are collected in small areas but at less frequent intervals. If the smaller areas adjust within the larger ones, then the way to make the data match of the same area is to use hierarchical matching — add the data for the small areas together until the grouped areas match the bigger ones and then match them exactly.

Fuzzy Matching

On many occasions, the boundaries of the smaller areas do not match with those of the larger ones. The problem occurs more often when the environmental data are involved. For example, crop boundaries that are usually defined by field edges/boundaries rarely match with the boundaries of the soil types. If we want

to determine the most productive soil for a particular crop, we need to overlay the two sets and compute crop productivity for each soil type. This is like laying one map over another and noting the combinations of soil and productivity.

A GIS can carry out all these operations. However, the sets of spatial information are linked only when they relate to the same geographical area.

Spatial Analysis

The strength of the GIS lies in its analytical capabilities. What distinguish the GIS from other information systems are its spatial analysis functions. The analysis functions use the spatial and non-spatial attributes in the database to answer questions about the real world. Geographic analysis facilitates the study of real-world processes by developing and applying models. Such models provide the underlying trends in geographic data and thus, make new possibilities available. The objective of geographic analysis is to transform data into useful information to satisfy the requirements of the decision-makers. For example, GIS may effectively be used to predict future trends over space and time related to variety of phenomena. However, before undertaking any GIS based analysis, one needs to identify the problem and define purpose of the analysis. It requires step – by – step procedures to arrive at the conclusions. The following spatial analysis operation may be undertaken using GIS :

(i) Overlay (ii) Buffer analysis

(iii) Network analysis (iv) Digital Terrain Model

However, under the constraints of time and space only the overlay and buffer analysis operations will be dealt herewith.

Overlay Operations

The hallmark of GIS is overlay operations. An integration of multiple layers of maps using overlay operations is an important analysis function. In other words, GIS makes it possible to overlay two or more thematic layers of maps of the same area to obtain a new map layer (*Fig. 6.7*). The overlay operations of a GIS are

Overlay Operation : x + y = z

where

x = Road map
y = Rail map
z = Communication Map
+ = The spatial overlay operation 'union'

Graphical representation

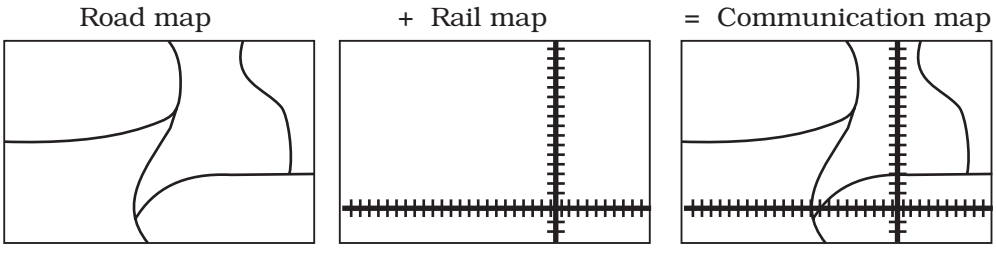


Fig. 6.7 : *Simple Overlay Operation*

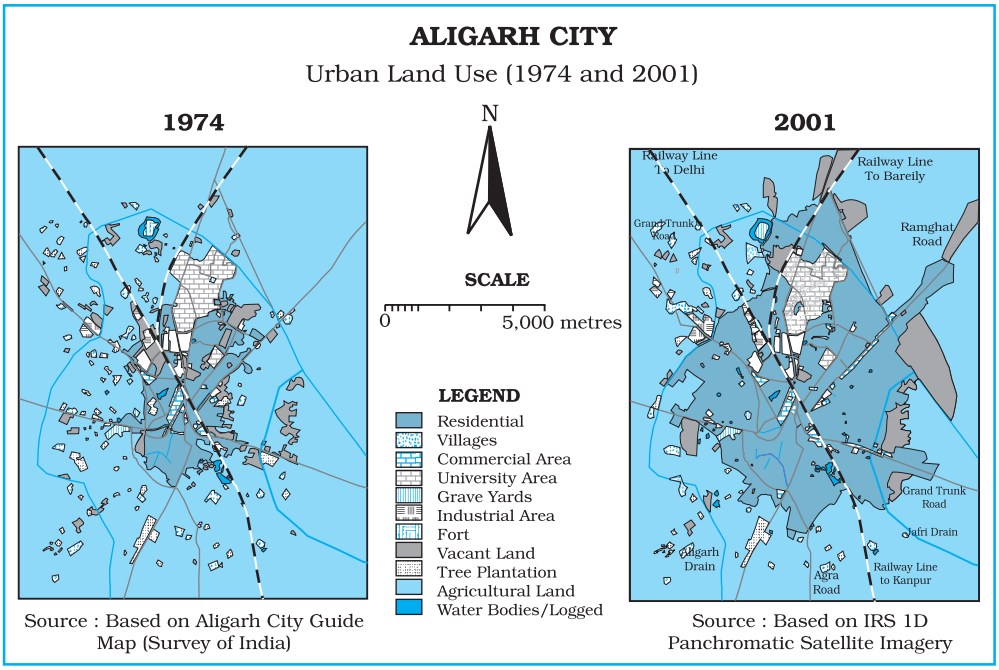


Fig. 6.8 : *Urban Land Use in Aligarh City, Uttar Pradesh during 1974 and 2001*

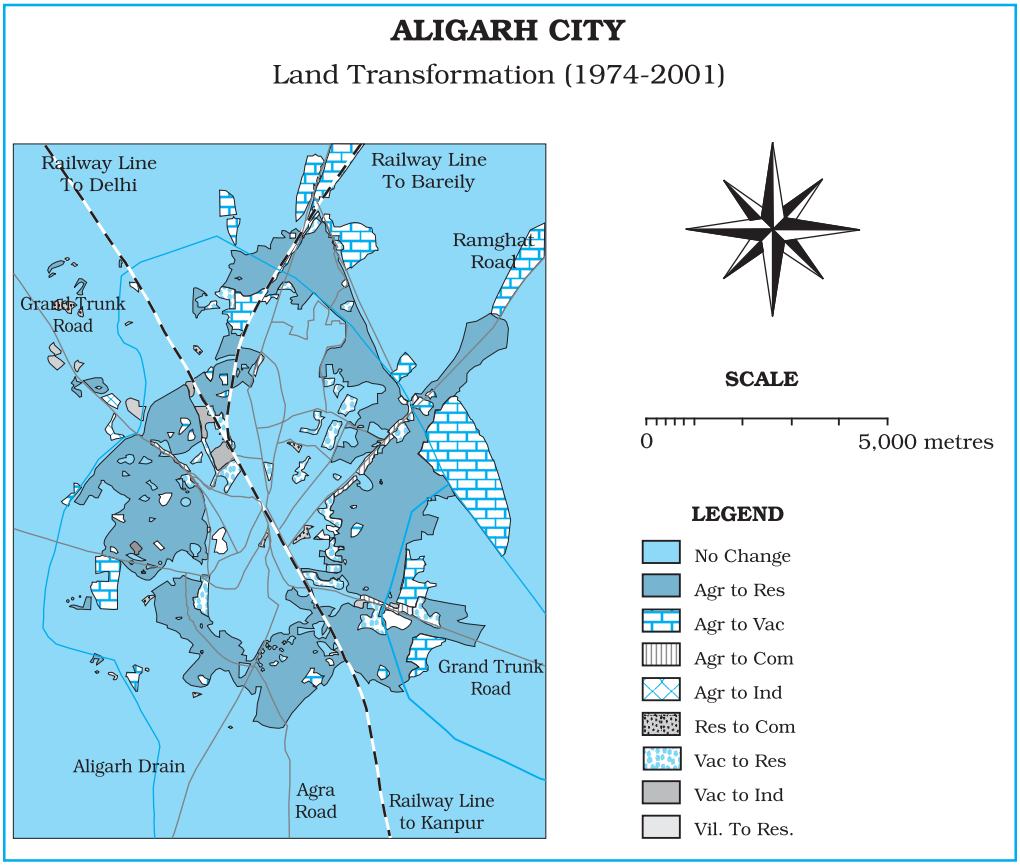


Fig. 6.9 : *Urban Land Transformations in Aligarh City during 1974-2001*

similar to the sieve mapping, i. e. the overlaying of tracing of maps on a light table to make comparisons and obtain an output map.

Map overlay has many applications. It can be used to study the changes in land use/land cover over two different periods in time and analyse the land transformations. For example, *Fig. 6.8* depicts urban land use during 1974 and 2001. When the two maps overlaid, the changes in urban land use have been obtained (*Fig. 6.9*) and the urban sprawl is mapped during the given time period (*Fig. 6.10*). Similarly, overlay analysis is also useful in suitability analysis of the given land use for proposed land uses.

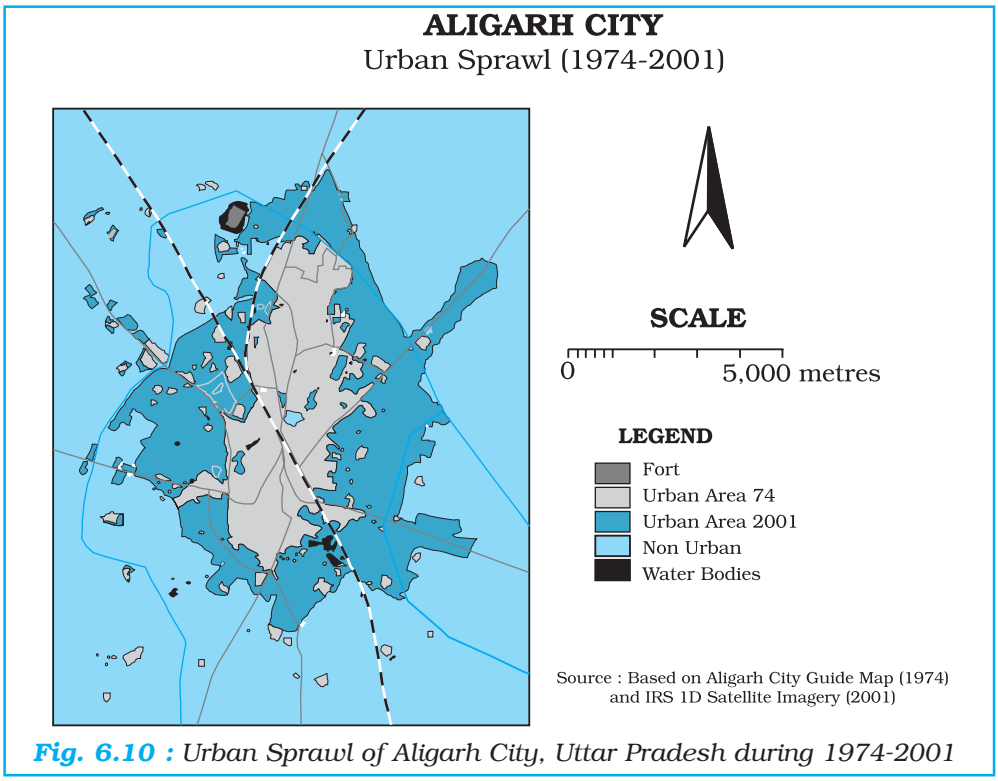


Fig. 6.10 : *Urban Sprawl of Aligarh City, Uttar Pradesh during 1974-2001*

Buffer Operation

Buffer operation is another important spatial analysis function in GIS. A buffer of a certain specified distance can be created along any point, line or area feature (*Fig. 6.11*). It is useful in locating the areas/population benefitted or denied of the facilities and services such as hospitals, medical stores, post office, asphalt roads, regional parks, etc. Similarly, it can also be used to study the impact of point sources of air, noise or water pollution on human health and the size of the population so affected. This kind of analysis is called proximity analysis. The buffer operation will generate polygon feature types irrespective of geographic features and delineates spatial proximity. For example, numbers of household living within one-kilometre buffer from a chemical industrial unit are affected by industrial waste discharged from the unit.

Arc View/ArcGIS, Geomedia and all other GIS software provide modules for buffer analysis along point, line and area features. For example, using appropriate commands of either of the available software one can create buffers of 2, 4, 6, 8, and 10 kilometres around the cities having a major hospital located therein. As a case study, point location of Saharanpur, Muzaffarnagar, Meerut, Ghaziabad,

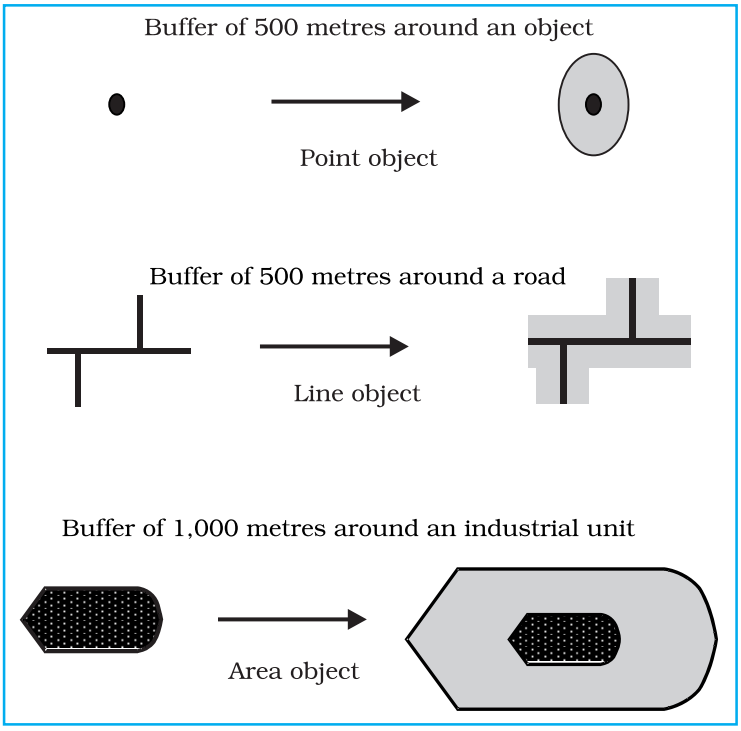


Fig. 6.11 : *Buffers of Constant Width Drawn around a Point, Line and a Polygon*

Gautam Budh Nagar and Aligarh has been mapped (*Fig. 6.12*) and the buffer have been created from the cities where major hospitals are found. One can observe that the areas closer to the cities are better served, people living away from the cities have to travel long distances to utilise the medical services and their areas that are least benefitted (*Fig. 6.13*).

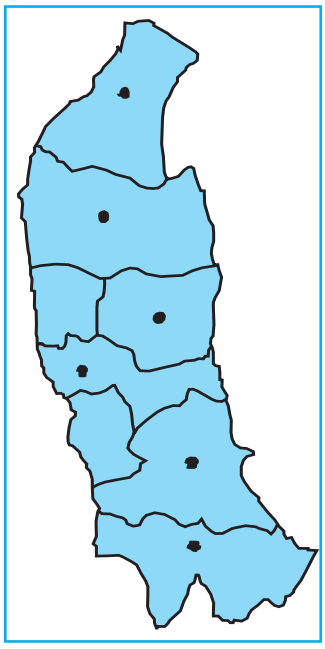


Fig. 6.12 : *Location Map of the Cities of Western Uttar Pradesh*

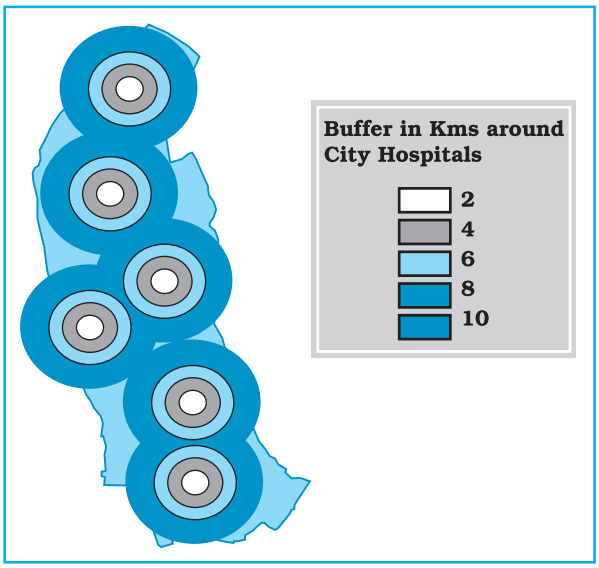


Fig. 6.13 : *Buffers of Specified Distances around Hospitals*

Excercises

1. Choose the right answer from the four alternatives given below :

(i) The spatial data are characterised by the following forms of appearance :
(a) Positional (b) Linear
(c) Areal (d) All the above forms

(ii) Which one of the following operations requires analysis module software?
(a) Data storage (b) Data display
(c) Data output (d) Buffering

(iii) Which one of the following is disadvantage of Raster data format ?
(a) Simple data structure
(b) Easy and efficient overlaying
(c) Compatible with remote sensing imagery
(d) Difficult network analysis

(iv) Which one of the following is an advantage of Vector data format ?
(a) Complex data structure
(b) Difficult overlay operations
(c) Lack of compatibility with remote sensing data
(d) Compact data structure

(v) Urban change detection is effectively undertaken in GIS core using:
(a) Overlay operations
(b) Proximity analysis
(c) Network analysis
(d) Buffering

2. Answer the following questions in about 30 words :

(i) Differentiate between raster and vector data models.
(ii) What is an overlay analysis?
(iii) What are the advantages of GIS over manual methods?
(iv) What are important components of GIS?
(v) What are different ways in which spatial data is built in GIS core?
(vi) What is Spatial Information Technology?

3. Answer the following questions in about 125 words :

(i) Discuss raster and vector data formats. Give example.
(ii) Write an explanatory account of the sequence of activities involved in GIS related work.

Glossary

Bar Graph : A series of columns or bars drawn proportional in length to the quantities they represent. They are drawn on a selected scale. They may be drawn either horizontally or vertically.

Central Tendency : The tendency of quantitative data to cluster around some value.

Choropleth Maps : Maps drawn on quantitative areal basis, calculated as average values per unit of area within specific administrative units, e.g. density of population and percentage of urban to total population. Distribution of a given phenomenon is shown by various shades of a colour or intensity.

Class Intervals : The difference between the lower and upper limits of any class of a frequency distribution is known as its class interval.

Correlation Co-efficient : A measure of the degree and direction of relationship between two variables.

Cumulative Frequency : The measurement of distribution of values in the different class intervals expressed as a percentage of the total frequencies either above or below specified value.

Dispersion : The degree of internal variations in the different values of a variable.

Flow Maps : Maps in which the "flow" or movement of people or commodities is represented by riband whose thickness is proportional to the quantity of goods or the number of people moving along different routes.

Histogram : A graphical representation of a frequency distribution, such as seasonal frequencies of rainfall.

Mean Deviation : A measure of dispersion derives from the average of deviations from some central value. Such deviations are taken absolutely, i.e., their signs are ignored. The central value is generally mean or median.

Median : It is the value which divides the number of observations in such a way that half the value are less than this value and half of them are more. If the values of a variable are arranged in either ascending of descending order, the median is the middle value.

Mode : The mode is that value of a variable which occurs maximum number of times.

Pie Diagram : A circular diagram in which a circle is divided into sectors for presenting data in percentage.

Standard Deviation : The most commonly used measure of dispersion. The standard deviation is the positive square root of the mean of the squares of deviations from the mean.

Tabulation : The process of putting raw data into a systematically arranged tabular form.

Variable : Any characteristic which varies. A quantitative variable is a characteristic which has different values; the differences of which are quantitatively measurable. Rainfall, for example, is a quantitative variables, because the differences in its different values at different places or at different times are quantitatively measurable. A qualitative variable on the other hand, is the characteristic; the different values of which cannot be measured quantitatively. Sex, for example, is a qualitative variable, it can be either male or female. A qualitative variable is also known as an attribute.

Annexure

Household Schedule

Poverty: Extent, Determinants and Consequences
Note: Collected data will be used only for academic exercise and will be kept confidential.

A. Identification

Village/Mohallah _____ Tehsil/City _____ District _____ State _____

Head of Household _____ S/o _____ Caste _____

S. No. of Respondent _____

B. Basic Demographic Information

S. No.	*Relation to Head of Household*	*Gender (M/F)*	*Age (Years)*	*Education Level (Yrs./Qua)*	*Marital Status (Code)*	*Primary Activity Specify with Code*	*Secondary Activity Specify with Code*	*Annual Non-Agricultural Income (Rs.)*
1.								
2.								
3.								
4.								
5.								
6.								
7.								
8.								
9.								
10.								

Activity Codes : None-0; Cultivator-1; Agricultural Labourer-2; Livestock Rearing-3; Mining-4; Household M, Industry-5A; Other M Industry-5B; Construction-6; Trade-7; Transport-8; Other Services -9; Student-10; Unemployed-11.

C. Capital Assets (Own share only)

Asset	*Unit*	*Size/No.*	*Asset*	*Unit/Type*	*No./Size*
1.	Built-up Area		13.	Goat	
2.	Non-Agricultural Land		14.	Sheep	
3.	Un-irrigated Land		15.	Donkey	
4.	Irrigated Land		16.	Other (specify)	
5.	Beed Land		17.	Cart	
6.	Peta Land		18.	Pump Set	
7.	Cow		19.	Tube-Well	
8.	Bullock		20.	Oil Engine	
9.	Calf		21.	Business Establishment	
10.	Buffalo		22.	Manufacturing Unit	
11.	He Buffalo		23.	Tractors/Trucks/Bus/Taxi	
12.	Calf (Buffalo)		24.	Others (Specify)	

D. Consumption Assets

Item	*No.*	*Model*	*Item*	*No.*	*Model*
Two Wheeler Fan/Cooler Others			Sewing Machine Bicycle		

E. Agricultural Production

Kharif			*Rabi*			*Zaid*		
Crop	*Area-Sown (ha)*	*Production (quintals)*	*Crop*	*Area-Sown (ha)*	*Production (quintals)*	*Crop*	*Area-Sown (ha)*	*Production (quintals)*

F. Livestock Production

Stock	*Milk (l/yr)*	*Power (day/yr)*	*Stock*	*Milk (l/yr)*	*Power (day/yr)*	*Wool (kg/yr)*
Cow Bullock Calf Buffalo He Buffalo			Goat Sheep Donkey Calf (Buffalo) Other			

G. Consumption

Item	*Unit*	*Quantity*	*Source*	*Item*	*Unit*	*Quantity*	*Source*
Wheat	Quintal/yr			Firewood	Quintal/ month		
Rice	Quintal/yr			Petrol/Diesel	Rs./month		
Jowar	Quintal/yr			Gas/Kerosene	Rs./month		
Bajra	Quintal/yr			Electricity Bill	Rs/month		
Maize	Quintal/yr			Water Charges	Rs/month		
Oth. Cereals	Quintal/yr			Clothing	Rs./yr		
Pulses	Quintal/yr			Educational	Rs./yr		
Sugar	Kg/month			Medicine, etc.	Rs./yr		
Gur	Kg/month			Others	—		
Coffee/ Tea	Kg/month			Milk	l/day		
Ghee	Kg/month			Meat	Kg/month		
Veg. Oil	Kg/month			Fish	Kg/month		
Vegetables/ fruits	kg/day						

H. Case specific observations of the Interviewer about the state of poverty/ well-being and associated conditions.

_____

_____

_____

_____

_____

_____

_____

Signature and Name of the Interviewer :

Date :

Table 5.1
Guidelines for Field Surveys

Item		***Survey Example***			
		Pollution	***Ground Water***	***Land Use***	***Poverty***
1. Title and Sub-Title		Industrial Effluents: Causes and Impacts – A Study of.....	Institutional Impacts of Ground Water Depletion - A Case Study of	Techno-Economic Changes and State of Land Use-A Study of	Urban Poverty : Causes and Consequences - A Study of.....
2. Objectives					
3. Coverage	(a) Spatial Coverage				
	(b) Temporal Coverage				
	(c) Thematic Coverage				
4. Tools and Techniques	(a) Secondary Information				
	(b) Maps				
	(c) Observations				
	(d) Measurement				
	(e) Interviewing Unit				
	(f) Survey Design				
	(g) Schedule/ Questionnaire				
5. Compilation and Computation	(a) Data Entry and Tabulation				
	(b) Computation of Indices				
	(c) Visual Presentation				
	(d) Thematic Mapping				
	(e) Statistical Analysis				
6. Report writing	(a) Outline				
	(b) Major Findings				

Item		Survey Example			
		Energy Issues	**Soil Degradation**	**Drought**	**Floods**
1. Title and Sub-Title		Pattern of Energy Sources and Consumption - A Study of	Deforestation and State of Soil Degradation - A Study of	Impact of Drought Condition and Coping Strategies - A Study of....	Costs and Benefits of Recurring Floods : A Study of.....
2. Objectives					
3. Coverage	(a) Spatial Coverage				
	(b) Temporal Coverage				
	(c) Thematic Coverage				
4. Tools and Techniques	(a) Secondary Information				
	(b) Maps				
	(c) Observations				
	(d) Measurement				
	(e) Interviewing Unit				
	(f) Survey Design				
	(g) Schedule/ Questionnaire				
5. Compilation and Computation	(a) Data Entry and Tabulation				
	(b) Computation of Indices				
	(c) Visual Presentation				
	(d) Thematic Mapping				
	(e) Statistical Analysis				
6. Report writing	(a) Outline				
	(b) Major Findings